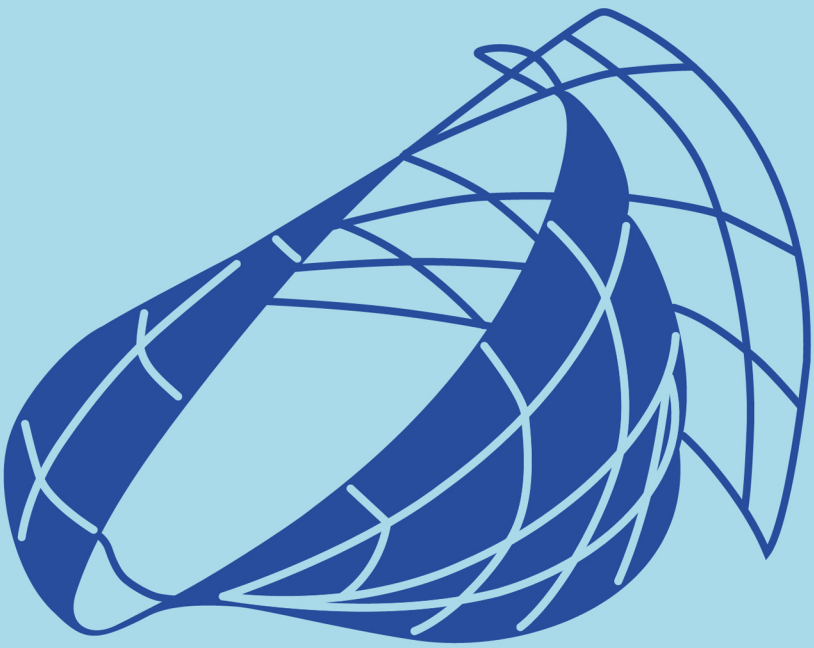
PART A

PROCEEDINGS OF THE SIXTEENTH

MARCEL GROSSMANN MEETING ON GENERAL RELATIVITY



Editors

Remo Ruffini

Gregory Vereshchagin

World Scientific

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

This page intentionally left blank

PART A

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

Proceedings of the MG16 Meeting
on General Relativity
Online, 5–10 July 2021

Editors

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Gregory Vereshchagin
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Series Editor

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Names: Marcel Grossmann Meeting on General Relativity (16th : 2021 : Online) | Ruffini, Remo, editor. | Vereshchagin, Gregory, editor.

Title: The sixteenth Marcel Grossmann meeting on recent developments in theoretical and experimental general relativity, astrophysics, and relativistic field theories : proceedings of the MG16 meeting on general relativity online, 5–10 July 2021 / editors: Remo Ruffini, University of Rome "La Sapienza", Rome, Italy, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy, Gregory Vereshchagin, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy.

Other titles: Proceedings of the MG16 meeting on general relativity online, 5–10 July 2021

Description: New Jersey : World Scientific, [2023] | Includes bibliographical references.

Identifiers: LCCN 2022047088 | ISBN 9789811269769 (set ; hardcover) | ISBN 9789811266584 (v. 1) | ISBN 9789811266591 (v. 2) | ISBN 9789811266607 (v. 3) | ISBN 9789811266614 (v. 4) | ISBN 9789811269776 (set ; ebook)

Subjects: LCSH: General relativity (Physics)--Congresses. | Gravitation--Congresses. | Quantum gravity--Congresses. | Cosmology--Congresses. | Astrophysics--Congresses.

Classification: LCC QC173.6 .M37 2021 | DDC 523.01--dc23/eng/20220929

LC record available at https://lccn.loc.gov/2022047088

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2023 by Editors

All rights reserved.

This is an Open Access volume published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 (CC BY-NC) License. Further distribution of this work is permitted, provided the original work is properly cited.

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/13149#t=suppl

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

THE MARCEL GROSSMANN MEETINGS

Series Editor: REMO RUFFINI

Publications in the Series of Proceedings

Proceedings of the Sixteenth Marcel Grossmann Meeting on General Relativity
(Virtual Meeting, 2021)
Edited by G. Vereshchagin, R. Ruffini
World Scientific, 2022

Proceedings of the Fifteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2018)
Edited by E.S. Battistelli, R.T. Jantzen, R. Ruffini
World Scientific, 2022

Proceedings of the Fourteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2015)
Edited by M. Bianchi, R.T. Jantzen, R. Ruffini
World Scientific, 2017

Proceedings of the Thirteenth Marcel Grossmann Meeting on General Relativity
(Stockholm, Sweden, 2012)
Edited by K. Rosquist, R.T. Jantzen, R. Ruffini
World Scientific, 2015

Proceedings of the Twelfth Marcel Grossmann Meeting on General Relativity
(Paris, France, 2009)
Edited by T. Damour, R.T. Jantzen, R. Ruffini
World Scientific, 2012

Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity
(Berlin, Germany, 2006)
Edited by H. Kleinert, R.T. Jantzen, R. Ruffini
World Scientific, 2007

Proceedings of the Tenth Marcel Grossmann Meeting on General Relativity
(Rio de Janiero, Brazil, 2003)
Edited by M. Novello, S. Perez-Bergliaffa, R. Ruffini
World Scientific, 2005

Proceedings of the Ninth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2000)
Edited by V.G. Gurzadyan, R.T. Jantzen, R. Ruffini
World Scientific, 2002

Proceedings of the Eighth Marcel Grossmann Meeting on General Relativity
(Jerusalem, Israel, 1997)
Edited by T. Piran
World Scientific, 1998

Proceedings of the Seventh Marcel Grossmann Meeting on General Relativity
(Stanford, USA, 1994)
Edited by R.T. Jantzen and G.M. Keiser
World Scientific, 1996

Proceedings of the Sixth Marcel Grossmann Meeting on General Relativity
(Kyoto, Japan, 1991)
Edited by H. Sato and T. Nakamura
World Scientific, 1992

Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity
(Perth, Australia, 1988)
Edited by D.G. Blair and M.J. Buckingham
World Scientific, 1989

Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 1985)
Edited by R. Ruffini
World Scientific, 1986

Proceedings of the Third Marcel Grossmann Meeting on General Relativity
(Shanghai, People's Republic of China, 1982)
Edited by Hu Ning
Science Press – Beijing and North-Holland Publishing Company, 1983

Proceedings of the Second Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1979)
Edited by R. Ruffini
North-Holland Publishing Company, 1982

Proceedings of the First Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1975)
Edited by R. Ruffini
North-Holland Publishing Company, 1977

SPONSORS

International Center for Relativistic Astrophysics Network (ICRANet)
International Center for Relativistic Astrophysics (ICRA)

FREEDOM OF MOVEMENT FOR SCIENTISTS

The Marcel Grossmann Meetings were founded with the premise that scientists of all nations have a right to meet to exchange knowledge independent of national borders.

ACKNOWLEDGEMENTS

We acknowledge the outstanding job done before, during and after the meeting by the ICRANet/ICRA administrative and secretarial staff: Cristina Adamo, Silvia Latorre, Elisabetta Natale, and Cinzia di Niccolo. Finally this meeting and its proceedings could not have functioned without the dedicated IT support of the ICRANet system manager Gabriele Brandolini, with some temporary assistance from Domenico La Selva and Damiano Verzulli. We would like to thank Linda Kwan from World Scientific for valuable assistance during preparation of these proceedings.

ORGANIZING BODIES OF THE SIXTEENTH MARCEL GROSSMANN MEETING

INTERNATIONAL ORGANIZING COMMITTEE

Blair David, Choquet Bruhat Yvonne, Damour Thibault, De Bernardis Paolo, Everitt C. W. Francis, Fryer Chris, Haensch Theodor, Henneaux Marc, Jones Christine, Kerr Roy, Kleinert Hagen, Kunz Jutta, Laemmerzahl Claus, Longair Malcolm, Mirabel Felix, Mirzoyan Razmik, Piran Tsvi, Rueda Jorge, Ruffini Remo (chair), Sasaki Misao, Sato Humitaka, Sunyaev Rashid, 't Hooft Gerard, Weinberg Steven, Yau Shing-Tung, Zhang Bing

LOCAL ORGANIZING COMMITTEE

Adamo Cristina, Bianco Carlo Luciano, Brandolini Gabriele A., di Niccolo Cinzia, Latorre Silvia, La Selva Domenico, Li Liang, Loppini Alessandro, Natale Elisabetta, Verzulli Damiano, Vereshchagin Gregory (chair), Wang Yu

INTERNATIONAL COORDINATING COMMITTEE

ALBANIA: Hafizi M. - ARGENTINA: Arguelles C., Scoccola C., Reula O., Romero G.E. - ARMENIA: Sahakyan N. - AUSTRALIA: Blair D., Ju L., Lun A., Manchester D., Melatos A., Quinn P., Scott S.M., Steele J.D. - AUSTRIA: Aichelburg P.C., Schindler S. - BELARUS: Kilin S., Prakapenia M., Siutsou I. - BELGIUM: Henneaux M. - BOLIVIA: Aguirre C.B. - BOSNIA: Pasic V. - BRAZIL: Barres de Almeida U., Coelho Goulart J., Dalmolin F.T., de Lima Rafael C.R., Guzzo M., Maia C., Malheiro M., Romero Filho C.A., Shellard R.C., Zen Vasconcellos C. - BULGARIA: Yazadjiev S. - CANADA: Singh D., Smolin L., Turok N. - CHILE: Bauer F., Bunster W.C., Giacomini A. - CHINA (MAINLAND): Cai R., Cai Y., Cao Z., Chang J., Chen J., Chen X., Dai Z., Feng L.-L., Han W., Jing Y., Li T.-P., Lin W., Lou Y.-Q., Luo J., Mei J., Tam T., Wang A., Wang Y., Wu X.-P., Wu Y.-L., Yuan Y.-F., Zhang B.-B., Zhang S.-N., Zhao G. - CHINA (TAIWAN): Chen Chiang-Mei, Chen Pisin, Lee Da-Shin, Lee Wo-Lung, Ni Wei-Tou - COLOMBIA: Bargueño de Retes P., Gonzalez G., Higuera Garzon M.A., Núñez L., Romano A.E., Valenzuela Toledo C.A., Zuluaga J.I. - CROATIA: Dominis Prester D., Karlica M., Milekovic M., Smolcic V., Smolic I., Suric T. - CUBA: Perez Martinez A., Pérez Rojas H. - CZECH REPUBLIC: Bicak J., Stuchlik Z. - DENMARK: Naselsky P. - ECUADOR: Contreras E. - EGYPT: Tawfik A.N., Wanas M.I. - ESTONIA: Einasto J., Saar E. - FINLAND: Poutanen J., Volovik G. - FRANCE: Brillet A., Buchert T., Chardonnet P., Coullet P., de Freitas Pacheco J.A., Deruelle N., Iliopoulos J., Lamanna G., Mignard F. - GEORGIA: Lavrelashvili George, Machabeli Giorgi - GERMANY: Biermann P., Blumlein J., Di Piazza A., Fritzsch H., Genzel R., Gilmozzi R., Hasinger G., Hehl F., Keitel C., Kiefer C., Mirzoyan R.,

Neugebauer G., Nicolai H., Renn J., Ringwald A., Ruediger A. - GREECE: Batakis N.A., Cotsakis S., Vagenas E.C. - HUNGARY: Fodor G., Levai P. - ICELAND: Bjornsson G., Jakobsson P. - INDIA: Chakrabarti S.K., Iyer B., Padmanabhan T., Souradeep T. - IRAN: Baghram S., Bavarsad E., Eslam Panah B., Firouzjahi H., Haghighat M., Mansouri R., Mashhoon B., Shakeri S., Sobouti Y., Taghi Mirtorabi M. - IRELAND: O'Murchada N. - ISRAEL: Milgrom M., Nakar E., Pe'er A., Piran T. - ITALY: Belinski V., Bianchi M., Bianco C.L., Cherubini C., Della Valle M., Falciano S., Filippi S., Haardt F., Menotti P., Merafina M., Pani P., Ricci F., Treves A., Vereshchagin G.V., Vitale S., Xue S.- S. - JAPAN: Fujimoto M.-K., Makishima K., Nakamura T., Sato K., Shibata M. - KAZAKHSTAN: Abishev M., Aimuratov Y., Boshkayev K., Mychelkin E.G., Spitaleri C. - KOREA (PYEONGYANG): Kim J.S. - KOREA (SEOUL): Kim S.P., Kim S.-W., Lee H.K., Lee H.-W., van Putten M. - KYRGYZSTAN: Gurovich V.Ts. - LIBYA: Gadri M. - MEXICO: Breton N., Cervantes-Cota J.L., Fraija Cabrera N.I., García-Diaz A.A., Macías Alvarez A., Mielke Eckehard W., Quevedo H., Rodriguez L.F. - NETHERLANDS: Slagter R. - NEW ZEALAND: Visser M., Wiltshire D. - NORWAY: Elgaroy O., Fonseca Mota D., Knutsen H. - PAKISTAN: Qadir A., Qamar S. - PERU: Vargas T. - POLAND: Belczynski K., Demianski M., Lewandowski Jerzy, Nurowski P., Sokolowski L. - PORTUGAL: Costa M., Da Silva A., Lemos J.P.S., Lobo F., Moniz P., Silva L.O. - ROMANIA: Visinescu M. - RUSSIA: Aksenov A., Arkhangelskaja I., Bisnovatyi-Kogan G., Blinnikov S., Chechetikin V.M., Cherepaschuk A.M., Khriplovich I., Lipunov V.M., Lukash V.N., Novikov I., Rudenko V.N., Starobinsky A.A. - SERBIA: Djordjevic G., Jovanovic P., Knezevic Z., Pankov-Hzvojevic M., Popovic L., Prodanovic T., Sijacki D., Simic S. - SLOVAKIA: Balek V. - SLOVENIA: Cadez A., Gomboc A., Zavrtanik D. - SOUTH AFRICA: Larena J., Maharaj S. - SPAIN: Elizalde E., Ibanez J., Perez M.J., Verdaguer E. - SWEDEN: Abramowicz M.A., Marklund M., Ryde F. - SWITZERLAND: Durrer R., Jetzer P. - TURKEY: Aliev A., Gurses M. - UKRAINE: Novosyadlyj B., Zaslavskii O., Zhuk A. - UNITED ARAB EMIRATES: Fernini I. - UNITED KINGDOM: Cruise A.M., Frenk Carlos S., Green M., Mavromatos N., Perry M., Willingale R. - USA: Abel T., Ashtekar A., Bardeen J., Carlstrom J., Cornish N., Dermer C., Fan X., Flanagan E., Fraschetti F., Fryer C., Incera V., Jantzen R.T. (chairperson), Kolb R., Laguna P., Lousto C., Madau Piero, Mathews Grant, Matzner Richard, Melia Fulvio, Mester John, Michelson Peter, Nordtvedt Kenneth, Parker Leonard, Pretorius F., Pullin J., Shapiro I., Shapiro S., Shoemaker D., Smoot G., Stiavelli M., Teukolsky S., van Nieuwenhuizen P., Zhang B. - UZBEKISTAN: Ahmedov B., Zalaletdinov R.M. - VATICAN CITY: Gionti G. - VENEZUELA: Fuenmayor E. - VIETNAM: Long H.N.

MARCEL GROSSMANN AWARDS

Sixteenth Marcel Grossmann Meeting

Institutional Awards

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

Individual Awards

DEMETRIOS CHRISTODOULOU
"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

GERARD 't HOOFT
"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

TSVI PIRAN
"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".

STEVEN WEINBERG
"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Each recipient is presented with a silver casting of the TEST sculpture by the artist A. Pierelli. The original casting was presented to His Holiness Pope John Paul II on the first occasion of the Marcel Grossmann Awards.

15th Marcel Grossmann Meeting
July 2018, Rome, Italy

Institutional Awards

PLANCK SCIENTIFIC COLLABORATION (ESA)
"for obtaining important constraints on the models of inflationary stage of the Universe and level of primordial non-Gaussianity; measuring with unprecedented sensitivity gravitational lensing of Cosmic Microwave Background fluctuations by large-scale structure of the Universe and corresponding B-polarization of CMB, the imprint on the CMB of hot gas in galaxy clusters; getting unique information about the time of reionization of our Universe and distribution and properties of the dust and magnetic fields in our Galaxy"

- presented to Jean-Loup Puget, the Principal Investigator of the High Frequency Instrument (HFI)

HANSEN EXPERIMENTAL PHYSICS LABORATORY AT STANFORD UNIVERSITY
"to HEPL for having developed interdepartmental activities at Stanford University at the frontier of fundamental physics, astrophysics and technology"

- presented to Research Professor Leo Hollberg, HEPL Assistant Director

Individual Awards

LYMAN PAGE
"for his collaboration with David Wilkinson in realizing the NASA Explorer WMAP mission and as founding director of the Atacama Cosmology Telescope"

RASHID ALIEVICH SUNYAEV
"for the development of theoretical tools in the scrutinising, through the CMB, of the first observable electromagnetic appearance of our Universe"

SHING-TUNG YAU
"for the proof of the positivity of total mass in the theory of general relativity and perfecting as well the concept of quasi-local mass, for his proof of the Calabi conjecture, for his continuous inspiring role in the study of black holes physics"

14th Marcel Grossmann Meeting
July 2015, Rome, Italy

Institutional Award

EUROPEAN SPACE AGENCY (ESA)
"for the tremendous success of its scientific space missions in astronomy, astrophysics, cosmology and fundamental physics which have revolutionized our knowledge of the Universe and hugely benefited science and mankind"

- presented to its Director General Johann-Dietrich Woerner

Individual Awards

KEN'ICHI NOMOTO
"for heralding the role of binary systems in the evolution of massive stars"

MARTIN REES
"for fostering Research in black holes, gravitational waves and cosmology"

YAKOV G. SINAI
"for applying the mathematics of chaotic systems to physics and cosmology"

SACHIKO TSURUTA
"for pioneering the physics of hot neutron stars and their cooling"

FRANK C.N. YANG
"for deepening Einstein's geometrical approach to physics in the best tradition of Paul Dirac and Hermann Weyl"

T.D. LEE (award received by Yu-Qing Lou on behalf of Prof. T.D. Lee)
"for his work on white dwarfs motivating Enrico Fermi's return to astrophysics and guiding the basic understanding of neutron star matter and fields"

13th Marcel Grossmann Meeting
July 2012, Stockholm, Sweden

Institutional Award

ALBANOVA
for its innovative status as a joint institute established by Stockholm University and the Royal Institute of Technology and for fostering contributions to cosmology and astrophysics in the profound scientific tradition established by Oskar Klein.

- presented to the Rector of Stockholm University, Prof. Kåre Bremer.

Individual Awards

DAVID ARNETT
for exploring the nuclear physics and yet unsolved problems of the endpoint of thermonuclear evolution of stars, leading to new avenues of research in physics and astrophysics.

VLADIMIR BELINSKI and I.M. KHALATNIKOV
for the discovery of a general solution of the Einstein equations with a cosmological singularity of an oscillatory chaotic character known as the BKL singularity.

FILIPPO FRONTERA
for guiding the Gamma-ray Burst Monitor Project on board the BeppoSAX satellite, which led to the discovery of GRB X-ray afterglows, and to their optical identification.

12th Marcel Grossmann Meeting
July 2009, Paris, France

Institutional Award

INSTITUT DES HAUTES ÉSTUDES SCIENTIFIQUE (IHÉS)
for its outstanding contributions to mathematics and theoretical physics, and notably for having renewed basic geometrical concepts, and having developed new mathematical and physical aspects of spacetime.

- presented to Prof. Jean-Pierre Bourguignon

Individual Awards

JAAN EINASTO
for pioneering contributions in the discovery of dark matter and cosmic web and fostering research in the historical Tartu Observatory.

CHRISTINE JONES
for her fundamental contributions to the X-ray studies of galaxies and clusters tracing their formation and evolution and for her role in collaborations using clusters to study dark matter and in analyzing the effects of outbursts from supermassive black holes on the intracluster gas.

MICHAEL KRAMER
for his fundamental contributions to pulsar astrophysics, and notably for having first confirmed the existence of spin-orbit precession in binary pulsars.

11th Marcel Grossmann Meeting
July 2006, Berlin, Germany

Institutional Award

FREIE UNIVERSITÄT BERLIN
for the successful endeavor of re-establishing — in the spirit of the Humboldt tradition — freedom of thinking and teaching within a democratic society in a rapidly evolving cosmos

- presented to Dr. Dieter Lenzen, President of FUB

Individual Awards

ROY KERR
for his fundamental contribution to Einstein's theory of general relativity: "The gravitational field of a spinning mass as an example of algebraically special metrics."

GEORGE COYNE
for his committed support for the international development of relativistic astrophysics and for his dedication to fostering an enlightened relationship between science and religion.

JOACHIM TRUMPER
for his outstanding scientific contributions to the physics of compact astrophysical objects and for leading the highly successful ROSAT mission which discovered more than 200,000 galactic and extragalactic X-ray sources: a major step in the observational capabilities of X-ray astronomy and in the knowledge of our universe.

10th Marcel Grossmann Meeting
July 2003, Rio de Janeiro, Brazil

Institutional Award

CBPF (Brazilian Center for Research in Physics)
for its role as a teaching and research institution and as a place originating fundamental physics ideas in the exploration of the universe.

- presented to its founders Cesar Lattes, Josè Leite Lopez and Jayme Tiomno

Individual Awards

YVONNE CHOQUET-BRUHAT AND JAMES W. YORK, JR.
for separate as well as joint work in establishing the mathematical framework for proving the existence and uniqueness of solutions to Einstein's gravitational field equations.

YUVAL NE'EMAN
for his contributions to science, epistimology, mathematics and physics from subnuclear to space sciences.

9th Marcel Grossmann Meeting
July 2000, Rome, Italy

Institutional Award

SOLVAY INSTITUTES
for identifying and recording in discussions by the protagonists the crucial developments of physics and astrophysics in the twentieth century.

- presented to Jacques Solvay

Individual Awards

CECILLE AND BRYCE DEWITT
for promoting General Relativity and Mathematics research and inventing the "summer school" concept.

RICCARDO GIACCONI
for opening, five successive times, new highways for exploring the Universe.

ROGER PENROSE
for extending the mathematical and geometrical foundations of General Relativity.

8th Marcel Grossmann Meeting
June 1997, Jerusalem

Institutional Award

HEBREW UNIVERSITY
for its role as a cradle of Science and Humanities and for hosting the manuscripts of Albert Einstein.

- presented to M. Magidor, President of the Hebrew University of Jerusalem

Individual Awards

TULLIO REGGE
for his contributions to the interface between mathematics and physics leading to new fields of research of paramount importance in relativistic astrophysics and particle physics.

FRANCIS EVERITT
for leading the development of extremely precise space experiments utilizing superconducting technology to test General Relativity and the Equivalence Principle.

7th Marcel Grossmann Meeting
June 1994, Stanford, USA

Institutional Award

SPACE TELESCOPE SCIENCE INSTITUTE
for its critical role in the direction and operation of the Hubble Space Telescope, a truly unique international laboratory for the investigation and testing of general relativity in the context of modern astrophysics and cosmology.

- presented to Peter Stockman

Individual Awards

SUBRAHMANYAN CHANDRASEKHAR
for his contributions to the analysis of gravitational phenomena from Newton to Einstein and especially for leading the way to relativistic astrophysics with the concept of critical mass for gravitational collapse.

JIM WILSON
for having built on his experience in nuclear physics, thermonuclear reactions, and extensive numerical simulation to create a new testing ground for the novel concepts of relativistic astrophysics.

6th Marcel Grossmann Meeting
June 1991, Kyoto, Japan

Institutional Award

RITP
for keeping alive first in Hiroshima and them in Kyoto research in relativity, cosmology, and relativistic field theory and the development of a school of international acclaim.

- presented to Professor K. Tomita

Individual Awards

MINORU ODA
for participating in the pioneering work of the early sixties in X-ray astronomy and for his subsequent molding of an agile and diversified Japanese scientific space program investigating the deepest aspects of relativistic astrophysics.

STEPHEN HAWKING
for his contributions to the understanding of spacetime singularities and of the large scale structure of the Universe and of its quantum origins.

5th Marcel Grossmann Meeting
August 1988, Perth, Australia

Institutional Award

THE UNIVERSITY OF WESTERN AUSTRALIA
for its contributions to relativistic astrophysics.

- presented to the Vice Chancellor, Professor Robert Smith

Individual Awards

SATIO HAYAKAWA
for his contributions to research in gamma, X-ray and infrared radiation as well as cosmic rays.

JOHN ARCHIBALD WHEELER
for his contributions to geometrodynamics and Einstein's visions.

4th Marcel Grossmann Meeting
July 1985, Rome, Italy

Institutional Award

THE VATICAN OBSERVATORY
for its contributions to the origin and development of astrophysics.

- presented to His Holiness Pope John Paul II

Individual Awards

WILLIAM FAIRBANK
for his work in gravitation and low temperature physics.

ABDUS SALAM
for his work in unifying fundamental interactions.

Institutional Awards for the Spektrum-Roentgen-Gamma (SRG) mission

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

On Tuesday June 29, 2021, the following 31 astro-ph appeared:

1. https://arxiv.org/abs/2106.14517
2. https://arxiv.org/abs/2106.14518
3. https://arxiv.org/abs/2106.14519
4. https://arxiv.org/abs/2106.14520
5. https://arxiv.org/abs/2106.14521
6. https://arxiv.org/abs/2106.14522
7. https://arxiv.org/abs/2106.14523
8. https://arxiv.org/abs/2106.14524
9. https://arxiv.org/abs/2106.14525
10. https://arxiv.org/abs/2106.14526
11. https://arxiv.org/abs/2106.14527
12. https://arxiv.org/abs/2106.14528
13. https://arxiv.org/abs/2106.14529
14. https://arxiv.org/abs/2106.14530
15. https://arxiv.org/abs/2106.14531
16. https://arxiv.org/abs/2106.14532
17. https://arxiv.org/abs/2106.14533
18. https://arxiv.org/abs/2106.14534
19. https://arxiv.org/abs/2106.14535
20. https://arxiv.org/abs/2106.14536
21. https://arxiv.org/abs/2106.14537
22. https://arxiv.org/abs/2106.14541
23. https://arxiv.org/abs/2106.14542
24. https://arxiv.org/abs/2106.14543
25. https://arxiv.org/abs/2106.14544
26. https://arxiv.org/abs/2106.14545
27. https://arxiv.org/abs/2106.14546
28. https://arxiv.org/abs/2106.14547
29. https://arxiv.org/abs/2106.14548
30. https://arxiv.org/abs/2106.14549
31. https://arxiv.org/abs/2106.14550

S.A. LAVOCHKIN ASSOCIATION
presented to its Designer General **Alexander Shirshakov**

Dr Alexander Shirshakov

S.A. Lavochkin Association created the Navigator space platform carrying German eRosita and Russian ART-XC X-Ray Telescopes, organized the launch of SRG Orbital X-Ray Observatory to the second Lagrangian point of the Sun-Earth system at a distance of 1.5 million km from the Earth and managed the observatory flight and the daily reception of its scientific data on Earth for 23.5 months.

Dr Alexander Shirshakov, Designer General of the S.A. Lavochkin Association, is specialized in design, manufacture, testing, launch and control of S/C for scientific purposes. Among those S/C launched, there are the «Radiostron» Astrophysical Observatory (2011) and the «Spektr-RG» space observatory (2019), while the planned S/C launches are «Luna-25» and «Exomars».

Dr Shirshakov started his career in 1973, working as an engineer of the State Unitary Enterprise «NPO named by S.A. Lavochkin» in Khimki (Russian Federation). Starting from 1989 he has played multiple roles within the Lavochkin Association, been appointed head of the group, head of the sector, head of department, deputy head of the complex, head of the branch, director of the center, deputy head of the Design Bureau, deputy General Designer and deputy General Director.

Dr Shirshakov is an editorial board Member of the reviewed edition of «Vestnik of Lavochkin Association». Since 2017, he is also member of the General Designer council. He has been awarded Honored Mechanical engineer of the Russian Federation as well as Agency-level award of the Russian Federal Space Agency.

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

Professor Peter Predehl

eROSITA is the soft X-ray telescope on-board the Russian-German Spektr-RG mission which was successfully launched from Baikonur on July 13, 2019 and placed in a halo orbit around the L2 point. 30 years after ROSAT, eROSITA performs an all-sky survey with an unprecedented sensitivity, spectral and angular resolution. Clusters of galaxies are the largest collapsed objects in the Universe. Their formation and evolution is dominated by gravity, i.e. Dark Matter, while their large scale distribution and number density depends on the geometry of the Universe, i.e. Dark Energy. X-ray observations of clusters of galaxies provide information on the rate of expansion of the Universe, the fraction of mass in visible matter, and the amplitude of primordial fluctuations which are the origin of clusters of galaxies and the whole structure of the universe. eROSITA has been designed to detect at least 100.000 clusters of galaxies and to detect systematically more than 3 million obscured accreting Black Holes. eROSITA will also allow to study the physics of galactic X-ray source

populations, like pre-main sequence stars, supernova remnants and X-ray binaries. The eROSITA telescope consists of seven identical Wolter-1 mirror modules. A novel detector system has been developed by MPE on the basis of the successful XMM-Newton pn-CCD technology. MPE is the scientific lead institute of eROSITA, responsible for the development of the instrument, the operation, the analysis software and data archive. Peter Predehl led this development as Principal Investigator of eROSITA and German lead scientist of the SRG mission for more than 15 years until the completion of the first of eight surveys in 2020. At this time eROSITA has already discovered more than 1 million X-ray sources, more than all X-ray observatories of the last 50 years together. This demonstrates that the design goals of the mission will easily be fulfilled.

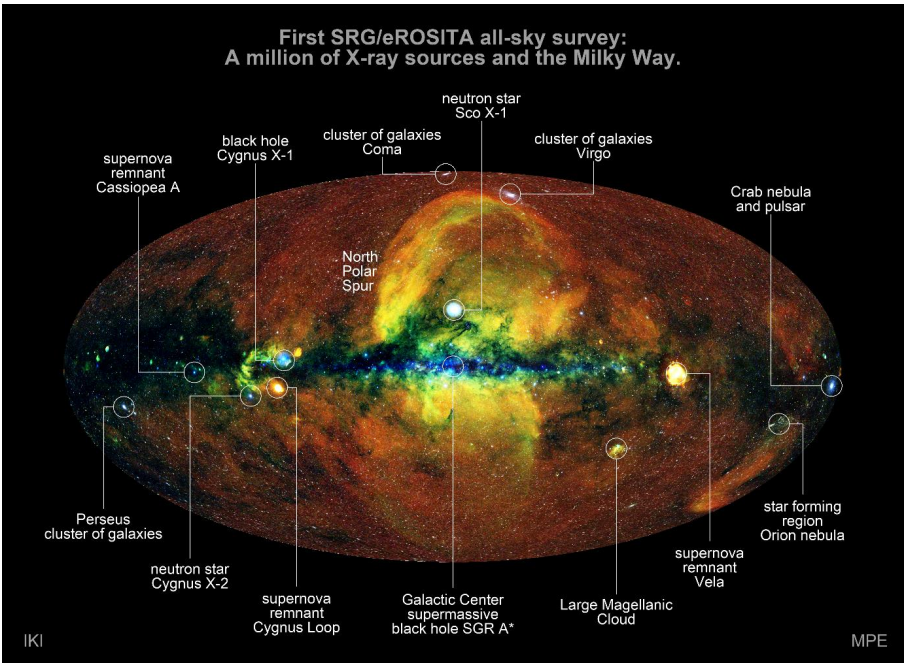
SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
presented to Professor **Rashid Sunyaev**

Professor Rashid Sunyaev

Space Research Institute (IKI) of the Russian Academy of Sciences was responsible for developing the overall concept and scientific program of the SRG Orbital observatory and played a leading role in developing the ART-XC telescope and the entire SRG observatory as part of the Russian space science program carried out by Roskosmos Corporation in the interests of the Russian Academy of Sciences.

During the flight to the L2 point of the Sun-Earth system, SRG with German (eRosita) and Russian (ART-XC named after Mikhail Pavlinsky) X-ray Telescopes aboard performed calibrations and long duration Performance Verification observations of a dozen of targets and deep fields. Starting in the middle of December 2019, the SRG scanned the whole sky three times. During these scans, SRG discovered two million point X-ray sources: mainly quasars, stars with hot and bright coronae, and more than 30 thousand clusters of galaxies. There is a competition and synergy in the search for clusters of galaxies between SRG and the ground-based Atacama Cosmology and South Pole Telescopes, which are searching for clusters of galaxies in microwave spectral band using Sunyaev-Zeldovich effect. SRG provided the X-Ray map of the whole sky in hard and soft bands, the last is now the best among existing. The huge samples of the X-ray selected quasars at the redshifts up to z = 6.2 and clusters of galaxies will be used for well-known cosmological tests and detailed study of the growth of the large scale structure of the Universe during and after reionization. SRG/eRosita is discovering every day several extragalactic objects which increased or decreased their brightness more than 10 times during half of the year after the previous scan of the same one-degree wide strip on the sky. A significant part of these objects has observational properties similar to the Events of Tidal Disruption of a star orbiting in the vicinity of the supermassive black hole. ART-XC discovered a lot of bright galactic and extragalactic transients.

Rashid Sunyaev is the Principal Investigator of SRG mission in Russia, director-emeritus of the Max-Planck Institute for Astrophysics and Maureen and John Hendricks distinguished visiting professor of the Institute for Advanced Study, Princeton.



Individual Awards

Professor **DEMETRIOS CHRISTODOULOU**

"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

Professor Demetrios Christodoulou

It was back in 1967 that Achille Papapetrou mentioned the case of the 16-year-old Demetrios Christodoulou to John Archibad Wheeler. Wheeler interviewed Demetrios in Paris and brought him immediately to Princeton where he was registered as an undergraduate at the university. After one year he entered the graduate school and started collaborating with me. At the time I was working with Wheeler on the effective potential approach to geodesics co-rotating and counter-rotating (see e.g. reference in The Classical Theory of Fields (Landau and Lifshitz, 1980) in the Kerr metric (later renamed as ISCO; see e.g. (Gravitation Misner, Thorne, Wheeler. 1973). In parallel, Frank Zerilli was working on the gravitational radiation emitted by the fall of a test particle in a Schwarzschild black hole (Zerilli 1970). From these limited conceptual arena Charles Misner and later Kip Thorne launched a program for the detection of gravitational waves on the Earth; see e.g. Misner 1974, Abbott et al. 2016, Abbott et al. 2017. See however Davis et al. 1972, Rodriguez et al. 2018 and J.A. Rueda et al. 2018.

A new approach started with the arrival of Demetrios: he was just creating mathematics following his needs. We identified the reversible and irreversible transformations of a Kerr black hole. Wheeler advanced a thermodynamic analogy. I addressed the need of identifying the concept of irreducible mass (from the Italian "irriducibile"), and was Demetrios's contribution to integrate, overnight, the differential equation for infinitesimal reversible transformations which led to the finite mass-energy formula of a Kerr black hole. That evening, while walking back home through IAS woods, I expressed to Wheeler the great relevance of the newly found formula by Demetrios and proposed to let Demetrios be the single author of this article, admiring his great mathematical talent. Wheeler agreed. The Editor of PRL objected since in that two pages article the Fig. 2 by Wheeler and myself was still unpublished. Actually that Fig. 2 followed a discussion I previously had with Penrose in Florence (Penrose 1961) which allowed us to present there, for the first time, a "Penrose Process". Some difficulties in achieving this process were obvious from the example in Fig. 2, which Roger later recognized himself (Penrose & Floyd 1971). The Editor finally agreed on our written request and the paper appeared on September 17, 1970 (Christodoulou, 1970). On January 1971 appeared my article with Johnny introducing the Black Hole (Ruffini & Wheeler, 1971), with the new physics we were developing in Princeton, including the concept of the "ergosphere". On march 1 1971 we submitted the mass formula of the Kerr Newmann metric, including the relation between the surface area of the horizon and the irreducible (Christodoulou & Ruffini, 1971). On March 11, 1971 the same results were independently confirmed by Steven Hawking, extending further the applicability of our equation (Hawking 1971).

The thesis was successfully discussed by a committee including Eugene Wigner (see Fig. 1), one of the closest collaborators of Albert Einstein and David Wilkinson (see Fig. 2), the head of the NASA WMAP mission, and Johnny and myself as supervisors. The new message was clear: Black Holes, far from being a sink of energy, were energy sources emitting "in principle" 50% of their mass energy, being extractable (Christodoulou & Ruffini, 1971).

Fig. 1 and Fig. 2: Demetrios during his thesis presentation with Eugene Wigner (Fig. 1) and David Wilkinson (Fig.2). Johnny and I were supervisors, ready to intervene in case of need, but no need of intervention was necessary! Wigner elaborated the aphorism of Niels Bohr "Interesting = wrong" in the most definite "very interesting if true = totally wrong".

Demetrios turned soon to the study of partial differential equations and mathematical physics, to which he dedicated for the rest of his career and results were published in four monographs: (Christodoulou and Klainerman 1994, Christodoulou 2007, Christodoulou 2009, Christodoulou 2019). In 1968, Johnny proposed to Demetrios the collapse of a "geon" composed of massless scalar field as a second topic for his thesis. It took almost forty years for him to solve this problem, extended by Demetrios to the focusing of gravitational waves leading to black hole formation (Christodoulou 2009).

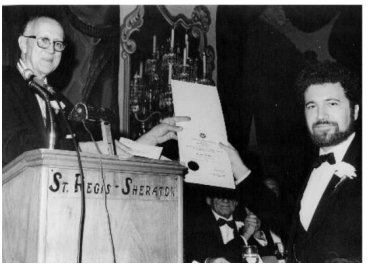


Fig. 3: Prof. Remo Ruffini receiving the Cressy Morrison Award of the New York Academy of Sciences, 1972 for the discovery of the first Black Hole in our galaxy Cygnus X1.

A "long march" started on 12 December 1970 with the launch of the Uhuru satellite by Riccardo Giacconi. Early in 1971 an almost daily conversation with him and Herb Gursky at the Smithsonian Astrophysical Observatory, leading to the discovery of binary X-ray sources. This was soon followedby the announcement of Cygnus X1 identified as the first black hole in our galaxy (Ruffini 1973); see e.g. Gursky & Ruffini 1975, which contained as well the first publicannouncement of the Discovery of Gamma Ray burst, as well as Giacconi & Ruffini 1980, 2009; see Figs. 3 and 4).

Fig. 4: In the second row, from left to right, there are, among others: E. T. Newman, S. Chandrasekhar (Nobel 1983), R. Giacconi (Nobel 2002), R. Ruffini, A. Treves, A. Hewish (Nobel 1974), D. Arnett, J.H. Taylor (Nobel 1993), J. Wilson, R. Penrose (Nobel 2020), as well as J. Bahcall, T. Damour, T. Piran et al.

Today, after fifty years, this "long march" has reached a definite result: through the grandest observational multi-wavelength effort in the history of mankind, from space, ground and underground observatories, we are finally finding evidence that black holes are "alive" and their "extractable energy" in our mass formula (Christodoulou & Ruffini, 1971), is the energy source of the most energetic cosmological sources: gamma ray bursts (GRBs),the active galactic nuclei (AGNs) as well as the ultra-high energy cosmic rays (UHECRs) (Ruffini et al. 2021 and references therein). Their "inner engine", has three independent components: 1) a Kerr black hole which is neither in a stationary state nor in vacuum, 2) a background magnetic field aligned with the black hole rotation axis, and 3) an extremely diluted fully ionized plasma (Moradi et al. 2021).There is no role in this inner engine for ISCO. Indeed a new electro dynamical field equations describe the synchrotron radiation emitted close to the black hole horizon, they point to a discrete and repetitive emission of "blackholic quanta" in the MeV and in the GeV. The magnitudes and the emission time scales of these quanta, for M87 and GRB 130427A, are expressed as a function of the above three parameters (Rueda & Ruffini, 2021). A long lasting GeV emission with a luminosity decreasing as a temporal power law, allows for the first time in GRBs, the determination of the black hole mass and spin as well as their time evolution perfectly fulfilling our mass energy formula (Christodoulou & Ruffini, 1971): a long lasting emission process profoundly different from the traditional process of continued gravitational contraction.

Remo Ruffini

Professor **GERARD 't HOOFT**

"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

Professor Gerard 't Hooft

Prof. Gerard 't Hooft has been a full Professor at the Utrecht University (the Netherlands), since 1977. Nowadays, he is an Emeritus Professor at that University. During his career, he has paid extended scientific visits to CERN (Geneva), Harvard, Stanford, Princeton and Duke University, NC. In 1999, together with M. Veltman, he received the Nobel Prize in Physics, awarded by The Royal Swedish Academy of Sciences, *"For elucidating the quantum structure of electroweak interactions in physics"*.

Prof. 't Hooft's main subjects of research includes:

– Gauge Theories for the sub-atomic particles and forces, various aspects and ingredients of what is now called "The Standard Model of the sub-atomic particles: renormalizability, topological features such as magnetic monopoles and instantons, 1/N expansions.

– Theories for the quantization of the gravitational force and black holes: producing models for the quantum properties of a black hole, as derived from Standard Model and General Relativity alone; its topological features such as antipodal identification.

– Fundamental theories underlying quantum mechanics, in particular returning determinism and reality to the dynamics of the tiniest material entities in his universe.

Prof. 't Hooft has been awarded the Wolf Prize of the State of Israel (1982), the Pius XI Medal (Vatican City, 1983), the Lorentz Medal (KNAW Amsterdam, 1986) as well as the Spinoza Premium (Netherlands Organization for Scientific Research NWO, 1995).

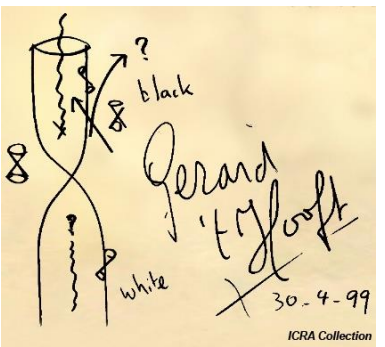


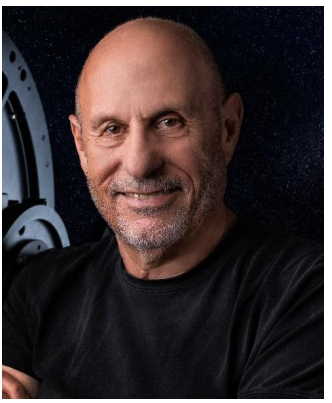
Fig. 2: The signature of Gerard 't Hooft on the wall of ICRA Room 301 (April 4, 1999).

A special event took place at ICRA on April 30, 1999. Prof. Ruffini invited Gerard 't Hooft to Rome to discuss a boundary condition for a quantum field on the black hole horizon, a topic Prof. Ruffini discussed in a previous article "Black-hole evaporation in the Klein-Sauter-Heisenberg-Euler formalism" with Thibault Damour (Phys. Rev. D 14, 332, 1976), but which needed to be examined in more detail. Prof. Ruffini planned to direct Gerard's attention to some specific aspects of this problem. Because we have traditionally been very attentive in spending ICRA travel funds, ICRA offered Gerard to come to Rome on a reduced fare weekend ticket arriving Friday and departing Monday. He had a great relaxing weekend together with Prof. Ruffini following his seminar, which among other things allowed Gerard to sign the wall in our ICRA Room (see Fig. 2), and during this splendid Rome spring weekend he also was able to find a missing factor of 2 in a formula in Prof. Ruffini's 1971 paper with Demetri Christodoulou on the black hole mass formula. The following October, Gerard received the Nobel prize, which meant that we could no longer get away with bringing him to Rome on a cheap ticket! Ever since Gerard has been in our MG IOC helping us with the preparation of the meetings. We are very happy to announce this MG16 Award to Gerard 't Hooft with the motivating phrase *"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon"*.

Remo Ruffini

Professor **TSVI PIRAN**

"for extending relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".



Professor Tsvi Piran

Tsvi Piran is the emeritus Schwartzmann professor at the Hebrew University of Jerusalem. He obtained his PhD in Physics, in 1976 from the Hebrew University working on the collisional Penrose process. Piran returned to the Hebrew University at 1981after being a post doc at Oxford and Texas and a long-term member at the IAS at Princeton. In 1982 he initiated and directed the first ever summer school on Gravitational Waves that took place at Les Houches. Piran was a visiting professor at Harvard, Columbia and New York and a Moore scholar at Caltech.

Piran's research deals with numerous aspects of relativistic astrophysics, ranging from the foundation of numerical relativity to modeling of observer relativistic phenomena and analytic work on the fate of gravitational collapse. Piran's research work focuses mostly on black holes and in particular on gamma-ray bursts. He was among the first to point out their cosmological origin and their association with merging neutron stars and heavy r-process nucleo synthesis. Piran's achievements were recognized in the 2019 EMET prize for Physics.

Professor **STEVEN WEINBERG**

"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Professor Steven Weinberg. Photo courtesy of Matt Valentine.

Steven Weinberg is a member of the Physics and Astronomy Departments at The University of Texas at Austin. His research has covered a broad range of topics in quantum field theory, elementary particle physics and cosmology. He has been honored with numerous awards, including the Nobel Prize in Physics, the National Medal of Science, the Heinemann Prize in Mathematical Physics and in 2020, the Breakthrough Prize. He is a member of the US National Academy of Sciences, Britain's Royal Society, and other academies in the USA and abroad. The American Philosophical Society awarded him the Benjamin Franklin Medal, with a citation that said he is "considered by many to be the preeminent theoretical physicist alive in the world today." His books for physicists include *Gravitation and Cosmology*, the three-volume work *The Quantum Theory of Fields*, *Cosmology* and published in April of 2021, *Foundations of Modern Physics*. Educated at Cornell, Copenhagen, and Princeton, he also holds honorary degrees from sixteen other universities. He taught at Columbia, Berkeley, M.I.T., and Harvard, where he was Higgins Professor of Physics, before coming to Texas in 1982.

Fig. 1: Chuo Pei Yuan and Cheng Ning Yang at MG2 in Trieste, Italy (1979).

The Sixteenth Marcel Grossmann Meeting (MG16) is a very special one in many respects: it will take place during a pandemic and in spite of the many difficulties, we have decided not to postpone it but to organize it as a virtual meeting. As described on the MG series webpage, these meetings started in 1975 with the first meeting at the International Centre for Theoretical Physics (ICTP) in Trieste (Italy) that I organized with Nobel Prize winner Abdus Salam. A second meeting followed in 1979, with a significantly larger participation including Nobel Laurate Cheng Ning Yang and a Chinese delegation led by Chuo Pei Yuan (see Fig. 1), including Fang Li-Zhi who had accompanied me during my entire first visit to China in 1979. The first truly international MG meeting followed in 1982 in Shanghai (China):this represented an especially important step forward both for the meeting and for China. A multi-millennia *"motto"* in China, which was then proclaimed on banners everywhere, read *"Friends from all over the world are welcomed"*.

We were soon at an impasse over the participation of scientists from Israel, since no diplomatic relations existed between China and Israel at that time and the Israeli scientists were not to be allowed to attend the meeting. A long negotiation began. The boundary conditions were clearly set by Steven Weinberg, a member of the present MG16 IOC: no MG meetings on Einstein's theory of general relativity could occur without the participation of Israeli scientists. The intervention of Yuval Ne'emann, also a member of the MG IOC then as well as the Minister of Science of Israel (see Fig.2), proposed a compromise that would admit at least one Israeli scientist. I went to Beijing alone, meeting every morning for a week with 12 Chinese representatives led by Chuo Pei Yuan going over all possible options. I stayed in an isolated villa not far from Tiananmen Square, accompanied by the 3 volumes of Matteo Ricci (RI MA TO) to keep me company. No solution was in sight the entire week. At the last moment, just before my departure, an agreement was finally reached allowing two Israeli scientists into China. The historic compromise would admit Gerard Tauber and Tsvi Piran into China using a special ICRA travel document I had proposed for them to be able to participate in the meeting, accepted by the Chinese Ambassador in Rome. This modified the thousand-year Chinese *"motto"* to read *"Scientists from all over the world are welcomed"*. The event was extremely beneficial for China and signaled the truly international nature of the MG meetings.

Fig. 2: From right to left: Chaim Weizmann, President of Israel; Yuval Ne'emann, Minister of Science of Israel; R. Ruffini.

I kept on meeting Tauber in the years which followed (see Fig. 3). Soon after, Yuval Ne'emann visited China. The development of bilateral relations, including military cooperation and economical tights, grow exponentially until the establishment of normal diplomatic relations between Israel and China in 1992.

Fig. 3: From right to left: Arrigo Finzi, Remo Ruffini, Gerard Tauber and Konrad Bleuler.

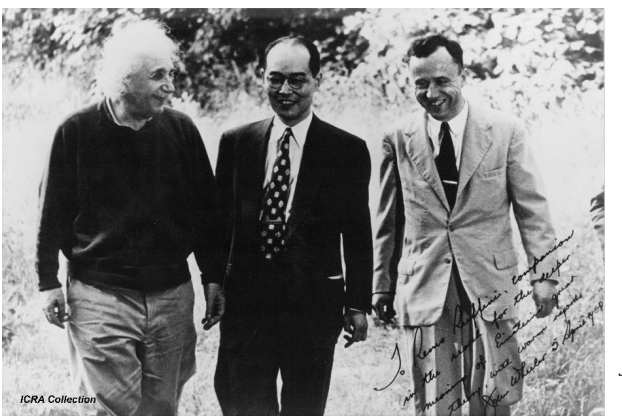


Fig. 4: Albert Einstein, Hideki Yukawa and John. A. Wheeler with a hand-written dedication to Remo Ruffini "To Remo Ruffini, companion in the search for the deeper meaning of Einstein great theory. With warm regards, John Wheeler 5 April 1968".

Given their key role played in the foundations of the MG meetings, I am very happy to propose on behalf of the MG16 IOC, two special Marcel Grossmann Individual Awards: one to Steven Weinberg for *"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory"* and another one to Tsvi Piran, *"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory"*, in the words of John A. Wheeler's photo dedication to me (see Fig. 4).

Remo Ruffini

PREFACE

Since 1975, the Marcel Grossmann Meetings on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories have been organized in order to provide opportunities for discussing recent advances in gravitation, general relativity and relativistic field theories, emphasizing mathematical foundations, physical predictions and experimental tests. The objective of these meetings is to elicit exchange among scientists that may deepen our understanding of spacetime structures as well as to review the status of ongoing experiments aimed at testing Einstein's theory of gravitation either from the ground or from space. Previous meetings have been held in Trieste (MG1: 1975) and (MG2: 1979), Shanghai (MG3: 1982), Rome (MG4: 1985, MG9: 2000), Perth (MG5: 1988), Kyoto (MG6: 1991), Stanford (MG7: 1994), Jerusalem (MG8: 1997), Rio (MG10: 2003), Berlin (MG11: 2006), Paris (MG12: 2009), Stockholm (MG13: 2012), MG14 in 2015 and MG15 in 2018 both in Rome.

Due to the COVID-19 pandemic spreading in the last two years the decision was taken to organize the Sixteenth Marcel Grossmann meeting for the first time in history entirely online. Despite numerous challenges, related to the organization of large worldwide event, MG16 showed the strongest ever interest from the scientific community with a record-breaking number of almost 1200 registered participants and of more than 1000 speakers.

The traditional six-day schedule has been modified to account for different time zones of the speakers and each day the program of the meeting was divided in three blocks with the reference to the Central European Summer Time. The first block was starting at 06:30 in the morning, allowing comfortable time for speakers from Asia and Oceania. The second block was held in the daytime in Europe and Africa. The third block was starting in the afternoon and ending at 19:30 allowing accommodation of the speakers from the Americas. Each day the blocks of plenary sessions were interchanging with the blocks of about 30 parallel sessions each, making this one of the most intense MG meetings ever. All this was possible thanks to recent developments in communication technologies. The Indico open-source software was selected as a web platform for this meeting, while Zoom platform was adopted for the video-conferencing. The meeting was streamed on ICRANet YouTube channel.

The meeting started on Monday July 5 with the Award ceremony. The individual awards went to Demetrios Christodoulou, Tsvi Piran, Gerard 't Hooft and Steven Weinberg, while the Institutional Awards went to the S.A. Lavochkin Association, to the Max Planck Institute for Extraterrestrial Physics – MPE and to the Space Research Institute IKI of the Russian Academy of Sciences. Overall there were 54 plenary talks, 4 public lectures and 5 roundtables and about 90 parallel sessions. The plenary session "Events in Relativistics Astrophysics" on Monday have seen the contributions from Rashid Sunyaev, Michael Kramer, James Miller-Jones, Felix Mirabel. The public lectures were delivered by Razmik Mirzoyan, Asghar Qadir

and Mohammad Bagheri. Plenary talks on Tuesday session "Black holes and the Quantum" by Juan Maldacena, Ahmed Almheiri, Gerard 't Hooft, Mihalis Dafermos, Sergiu Klainerman, Abhay Ashtekar and Frank Wilczek were bracketed by two roundtables on "New results from SRG/eRosita" with the participation of Andrea Merloni, Prof. Rashid Sunyaev, Alexander Lutovinov, Chandreyee Maitra, Esra Bulbul and "Solar neutrinos and Borexino" with the participation of Gianpaolo Bellini and Wick Haxton. Plenary talks on Wednesday in the session "Lambda CDM tensions" by George Efstathiou, Scolnic Daniel, Marc Kamionkowski, Wendy Freedman, Priya Natarajan and Licia Verde were followed by the roundtable "Precision cosmology" with the participation of Licia Verde, Marc Kamionkowski, Piero Rosati, and the public lecture by Francis Halzen. Two blocks of Thursday plenary sessions "Black holes in GRBs" and "Precision tests" included the talks by Roy Kerr, Yuan Ha, Lorenzo Amati, Elena Pian, Carlos Raúl Argüelles, Di Li, Jianglai Liu, Claus Lämmerzahl, Gerhard Heinzel and Ignazio Ciufolini and were followed by the roundtable "GRB 170817A and GRB 190829A" with the participation of Eleonora Troja, Liang Li, Rahim Moradi, Jorge Armando Rueda Hernandez. Two plenary blocks on Friday "Massive stars" and "Physics behind stellar collapse" included the talks by Selma de Mink, Norbert Langer, Jiri Bicak and Tomáš Ledvinka, Ivan De Mitri, Rahim Moradi and Giancarlo Cella. Finally, two plenary blocks on Saturday "Current and future missions" have seen the talks by Shuang-Nan Zhang, Weimin Yuan, Makoto Tashiro, Ruoyu Liu, Jean-Luc Atteia, Jim Hinton and Nicholas White and were followed by the roundtable "What is in our Galactic center" with the participation of Reinhard Genzel, Carlos Raúl Argüelles, Andreas Krut, Jorge Armando Rueda Hernandez, Eduar Becerra Vergara. The program of the meeting can be found at the official website http://www.icra.it/mg/mg16 and at ICRANet Indico website https://indico.icranet.org/event/1/.

These proceedings include about 400 papers containing the results presented at the Sixteenth Marcel Grossmann meeting. The plenary papers from the meeting have been published in International Journal of Modern Physics D as they were submitted. The table of contents includes also the links to YouTube videos with talks given at the meeting and cover plenary talks, public lectures, roundtables and all parallel sessions. The general link to the videos from MG16 is: https://www.youtube.com/watch?v=QFe1lsSid-o&list=PLr5RLbSWSonsaOnZukBDs0qsNIWM8AvRF.

As the editors we would like to express our gratitude to all the chairpersons of the parallel sessions at MG16, who peer-reviewed the papers submitted for these proceedings, as well as to the ICRANet secretariat office and in particular to Cinzia di Niccolo, Elisabetta Natale and Yasmina Di Domizio, as well as to ICRANet system manager Gabriele Brandolini for their help in preparation of this publication.

Remo Ruffini and Gregory Vereshchagin
November 2021

This page intentionally left blank

Contents

Publications in this Series v

Sponsors and Acknowledgements vii

Organizing Committees ix

Marcel Grossmann Awards xi

Preface xxxvi

PART A
PLENARY SESSIONS

New results from testing relativistic gravity with radio pulsars
Michael Kramer 3

Dragging of inertial frames by matter and waves
Jiří Bičák and Tomáš Ledvinka 22

Probes of the progenitors, engines and physics behind stellar collapse
Chris L. Fryer 39

The observation of high-energy neutrinos from the cosmos: Lessons learned for multimessenger astronomy
Francis Halzen 59

The first results of PandaX-4T
Jianglai Liu on behalf of the PandaX Collaboration 85

XRISM: X-ray imaging and spectroscopy mission
Makoto S. Tashiro and the XRISM team 95

The SVOM mission
J.-L. Atteia, B. Cordier and J. Wei on behalf of the SVOM Collaboration 104

Quantum field theory with boundary conditions at the horizons
Gerard 't Hooft 133

The development of general relativity and the cosmological constant
Asghar Qadir 143

The irreducible mass of Christodoulou-Ruffini-Hawking mass formula
Yuan K. Ha 154

Reshaping our understanding on structure formation with the quantum nature of the dark matter
C. R. Argüelles, E. A. Becerra-Vergara, A. Krut et al. 164

First results of LHAASO
Ruo-Yu Liu for the LHAASO Collaboration 180

On the MG16 awards 2021
Remo Ruffini 196

The white dwarf binary merger model of GRB 170817A
J. A. Rueda, R. Ruffini, Liang Li et al. 217

PARALLEL SESSIONS

Accretion

• MHD Processes Near Compact Objects
Chairperson: Sergey Moiseenko
YouTube link: https://youtu.be/2WMTg06ZmV8

A semi-implicit multidimensional unstructured gas dynamical solver for astrophysical applications
Ilya A. Kondratyev and Sergey G. Moiseenko 242

Magnetized neutron stars propagating through a non-uniform ISM
O. D. Toropina, M. M. Romanova and R. V. E. Lovelace 255

Calculation of the kinetic coefficients of arbitrary degenerate electrons in magnetized dense matter
M. V. Glushikhina and G. S. Bisnovatyi-Kogan 264

Modeling of magnetic fields of accretion discs, using no-z- and RZ-approximations
M. V. Pashentseva and E. A. Mikhailov 272

• Accretion Discs and Jets
Chairpersons: Audrey Trova and Shokoufe Faraji
YouTube link: https://youtu.be/29Wj9RCVEKw

Limiting effects in tori clusters
D. Pugliese and Z. Stuchlik 280

Hydrodynamical transport of angular momentum in accretion disks in the presence of nonlinear perturbations due to noise
Subham Ghosh and Banibrata Mukhopadhyay 295

Properties of accretion disc models in the background quadrupole
Shokoufe Faraji 307

Magnetized tori around a uniformly accelerating black hole
Shokoufe Faraji and Audrey Trova 317

Multifrequency behaviour of high mass X-ray binaries (Time lag between optical and X-ray outbursts)
Franco Giovannelli 321

Active Galactic Nuclei

• The Black Hole in M87
Chairpersons: Brian Punsly and Jorge Rueda
YouTube link: https://youtu.be/l1lTgksyJag

Rotation of the crescent image of M87* and polarization of its ESE hotspot
Krzysztof Nalewajko 339

Magnetic reconnection in jet-accretion disk systems
Chandra B. Singh, Elisabete M. de Gouveia Dal Pino, Luis H. S. Kadowaki et al. 344

• Machine Learning in Astronomy: AGN, Transient Events, Cosmology and Others
Chairpersons: Rahim Moradi and Yu Wang
YouTube link: https://youtu.be/ErqrmMZQsBk

Exact fractal model of the universe and possible machine learning methods for the verification of the fractality
A. A. Kirillov, E. P. Savelova and P. O. Vladykina 352

Estimating the photometric redshifts of galaxies and QSOs using regression techniques in machine learning
A. Momtaz, M. H. Salimi and S. Shakeri 368

Deep learning in quasar physics
F. Rastegar Nia, M. T. Mirtorabi, R. Moradi, Y. Wang and A. Vafaei Sadr 382

Cosmological density field emulation and gravitational wave inference based on dimensionality reduction and supervised machine learning
Miguel Conceição, António da Silva and Alberto Krone-Martins 391

Unsupervised photometric detection of galaxy cluster candidates in large surveys
Ana Carvalho, Alberto Krone-Martins and Antonio da Silva 409

• Multiwavelength and Multi-Messenger Observations of Active Galactic Nuclei
Chairpersons: Paolo Giommi and Narek Sahakyan

YouTube link: https://youtu.be/-Hyu2NQsExg

Time-dependent lepto-hadronic modeling of the emission processes in blazar jets
S. Gasparyan, D. Bégué and N. Sahakyan 429

Multiwavelength study of high-redshift blazars
G. Harutyunyan and D. Israyelyan 445

Alternative Theories

• Extended Theories of Gravity and Quantum Cosmology
Chairpersons: Yi-Fu Cai and Wentao Luo

YouTube links: https://youtu.be/ADRr9DfV5zM
https://youtu.be/eOzpiC1cFkU
https://youtu.be/kYzJds_JIp8

Quantum gravity phenomenology from thermodynamics of spacetime
A. Alonso-Serrano and M. Liška 462

Gauge theory of gravity based on the correspondence between the 1^{st} and the 2^{nd} order formalisms
David Benisty 479

$U(1)$ local strings in hybrid metric-Palatini gravity
Tiberiu Harko, Francisco S. N. Lobo and Hilberto M. R. da Silva 485

Inflationary supersymmetric FRLW quantum cosmology
N. E. Martínez-Pérez, C. Ramirez and V. M. Vázquez-Báez 499

Effective $f(R)$ actions for modified loop quantum cosmologies
Ana Rita Ribeiro, Daniele Vernieri and Francisco S. N. Lobo 517

Probing multiverse using gravitational wave observations
Moe Kukihara and Kazuhiro Hayama 531

Operator ordering ambiguity in observables of quantum cosmology
Harkirat Singh Sahota and Kinjalk Lochan 538

Decoupled quark stars in self-interacting Brans-Dicke gravity
M. Sharif and Amal Majid 548

Big-bounce in projectively invariant Nieh-Yan models: The Bianchi I case
Flavio Bombacigno, Simon Boudet, Gonzalo J. Olmo and Giovanni Montani 561

Late time cosmology with derivatives of matter Lagrangian
Shahab Shahidi 576

On the semiclassical and quantum picture of the Bianchi I polymer dynamics
E. Giovannetti, G. Montani and S. Schiattarella 588

Quantum corrections to the Bianchi II transition under local rotational invariance
Sara F. Uria, David Brizuela and Ana Alonso-Serrano 597

• **Mathematical Problems of Relativistic Physics: Classical and Quantum**
Chairpersons: A. Shadi Tahvildar-Zadeh and Michael Kiessling

YouTube links: https://youtu.be/9Dr3M9Kb2jo
https://youtu.be/gMOykpapJ5A

The hypercomplex medium as storage of physical equations
Alexander P. Yefremov 605

The Maxwell-Bopp-Landé-Thomas-Podolsky-Einstein system for a static point source
Érik Amorim 619

On recent developments in the theory of relativistic dissipative fluids
V. Hoang 627

Adiabatic solutions in general relativity as null geodesics on the space of boundary diffeomorphisms
Emine Şeyma Kutluk 635

The point spectrum of the Dirac Hamiltonian on the zero-gravity Kerr-Newman spacetime
M. Kiessling, E. Ling and A. S. Tahvildar-Zadeh 648

Causal fermion systems: Classical gravity and beyond
Felix Finster 661

Newman-Penrose-Debye formalism for fields of various spins in pp-wave backgrounds
Aleksandr Kulitskii and Elena Yu Melkumova 679

Gravitational geometric phase
Banibrata Mukhopadhyay, Tanuman Ghosh and Soumya Kanti Ganguly 689

Retarded potentials and radiation in odd dimensions
D. V. Gal'tsov and M. Khlopunov 699

Wave propagation in the anti-deSitter optical metric
D. García-Peláez and C. S. López-Monsalvo 713

New approaches to constrained dynamics and Hamilton-Jacobi procedures in general relativity
D. Salisbury, J. Renn and K. Sundermeyer 719

Orientability of space from electromagnetic quantum fluctuations
N. A. Lemos and M. J. Rebouças 725

Essential self-adjointness of Dirac operators under the influence of general-relativistic gravity
Michael K.-H. Kiessling, A. Shadi Tahvildar-Zadeh and Ebru Toprak 736

• **Wormholes, Energy Conditions and Time Machines**
Chairpersons: Francisco Lobo and Diego Rubiera-Garcia

YouTube links: https://youtu.be/tu_3Wqcd9Ys
https://youtu.be/NqN1c-2fv8Y

Relic magnetic wormholes as possible source of toroidal magnetic fields in galaxies
A. A. Kirillov and E. P. Savelova 743

Wormhole geometries induced by action-dependent Lagrangian theories
Ismael Ayuso, Francisco S. N. Lobo and José P. Mimoso 756

Gravitational lensing by wormholes in binary systems
S. Pietroni 774

Hyper-fast positive energy warp drives
E. W. Lentz 779

From black-bounce to traversable wormhole, and beyond
Alex Simpson 787

Tractor beams, pressor beams, and stressor beams within the context of general relativity
Matt Visser, Jessica Santiago and Sebastian Schuster 808

A singularity theorem for evaporating black holes
E.-A. Kontou, B. Freivogel and D. Krommydas 822

Circularly symmetric thin-shell wormholes in F(R) gravity with (2+1)-dimensions
Cecilia Bejarano, Ernesto F. Eiroa and Griselda Figueroa-Aguirre 831

Warp drive dynamic solutions considering different fluid sources
Osvaldo L. Santos-Pereira, Everton M. C. Abreu and Marcelo B. Ribeiro 840

Symmetries and geometry of spacetime: Towards a new paradigm
Francisco Cabral, Francisco S. N. Lobo and Diego Rubiera-Garcia 856

• Theories of gravity: Alternatives to the cosmological and particle standard models
Chairpersons: Stefano Bellucci and Orlando Luongo

YouTube links: https://youtu.be/-aQcrYoQfBM
https://youtu.be/WsCwPb5OQhY
https://youtu.be/JMnrUfBgqVU
https://youtu.be/pPhZY-bbsew

Thermodynamics of scalar-tensor gravity: A new approach
Valerio Faraoni 876

Two body dynamics in a quadratic modification of general relativity
Soham Bhattacharyya 883

Alternatives to Λ: Torsion, generalized couplings, and scale invariance
C. J. A. P. Martins, C. M. J. Marques, C. B. D. Fernandes et al. 907

Model-independent test of scalar-tensor gravity theory by reconstructing scalar mode of GW170817
Yuya Gushima and Kazuhiro Hayama 921

Cosmology in the novel scalar-tensor representation of $f(R,T)$ gravity
Tiago B. Gonçalves, João Luís Rosa and Francisco S. N. Lobo 932

On the interaction between electromagnetic, gravitational, and plasma related perturbations on LRS class II spacetimes
P. Semrén 943

Condition for expansion-collapse duality between Einstein and Jordan frames
Dipayan Mukherjee, H. K. Jassal and Kinjalk Lochan 958

The model of dark energy based on the quantum-mechanical uncertainty relation
Yu. V. Dumin 967

• Conformal Dilaton Gravity and Related Issues
Chairperson: Reinoud Jan Slagter

YouTube link: https://youtu.be/A3Ygi3YBs5A

Conformal dilaton gravity, antipodal mapping and black hole physics on a warped spacetime
R. J. Slagter 978

From neutrino masses to the full size of the universe — Some intriguing aspects of the tetron model
B. Lampe 999

Summary parallel session AT5
Reinoud Jan Slagter 1014

• Horava-Lifshitz Gravity
Chairperson: Anzhong Wang

YouTube links: https://youtu.be/vPT1dH1zITE
https://youtu.be/5z7zhpiDOpw

Boundary conditions for the Klein-Gordon field on Lifshitz spacetimes
Lissa de Souza Campos 1017

Dynamical system analysis of Bianchi-I spacetimes in $f(R)$ gravity
Saikat Chakraborty, Kazuharu Bamba and Alberto Saa 1026

Cosmological implications in modified Hořava-Lifshitz gravity
Abdul Jawad, Kazuharu Bamba and Farwa Khurshid 1038

Finite action principle and wormholes
Jan Chojnacki and Jan Kwapisz 1046

Strange quark stars in Hořava gravity
Grigoris Panotopoulos 1054

Shadows of Kerr-like black holes in $4D$ Einstein–Gauss–Bonnet gravity and constraints from EHT observations
Sushant G. Ghosh and Rahul Kumar Walia 1069

Wormhole interaction in 2d Horăva-Lifshitz quantum gravity
Jan Ambjørn, Yuki Hiraga, Yoshiyasu Ito and Yuki Sato 1084

Nature of singularities in vector-tensor theories of gravity
V. H. Satheeshkumar 1095

Hořava-Lifshitz and Einstein-Æther gravity in the light of Event Horizon Telescope observations of M87*
Emmanuel N. Saridakis 1104

Hořava-Lifshitz gravity in $(3 + 1)$ dimensions coupled with anisotropic matter and possible constraints from GRB 170817A
Tao Zhang and Fu-Wen Shu 1112

Summary of the parallel session AT6
Anzhong Wang 1119

• Ghost-Free Models of Modified Gravity: Massive Gravity, Horndeski and DHOST Theories, Other Related Models; Their Properties and Solutions
Chairpersons: Dmitry Gal'tsov and Michael Volkov

YouTube links: https://youtu.be/l8KHUPnT2D8
https://youtu.be/48OHIKpgNqs

Non-local R^2-like inflation, gravitational waves and Non-Gaussianities
K. Sravan Kumar 1124

Palatini kinetic scalar-tensor theory: Analytical and numerical solutions
D. V. Gal'tsov and D. S. Bushuev 1136

PART B

Black Holes: Theory and Observations/Experiments

• Theoretical and Observational Studies of Astrophysical Black Holes
Chairperson: Alexander Zakharov

YouTube links: https://youtu.be/fiv_MH-N2kw
https://youtu.be/GYoOb17GvE8

Reconstruction of a star motion in the vicinity of black hole from the redshift of the electromagnetic spectrum
S. O. Komarov and A. K. Gorbatsievich 1151

Shadows of hairy Kerr black holes and constraints from M87*
Sushant G. Ghosh and Misba Afrin 1167

Displacement memory and BMS symmetries
Shailesh Kumar 1179

Physical black holes in semiclassical gravity
Sebastian Murk and Daniel R. Terno 1196

• Black Hole Thermodynamics
Chairperson: Hernando Quevedo

YouTube links: https://youtu.be/XmZDf5mrXQk
https://youtu.be/amseL2qykfk

Information recovery from evaporating rotating charged black holes
Zhi-Wei Wang, Samuel L. Braunstein and Saurya Das 1212

Black hole thermodynamics from entanglement mechanics
S. Mahesh Chandran and S. Shankaranarayanan 1223

Thermodynamics of charged black hole
M. Sharif and Amjad Khan 1238

Rindler trajectories and Rindler horizons in the Schwarzschild spacetime
Kajol Paithankar and Sanved Kolekar 1250

Linear growth of the two-point function for the Unruh state in $1+1$ dimensional black holes
Paul R. Anderson, Zachary P. Scofield and Jennie Traschen 1255

Stress-energy tensor for a quantized scalar field in a four-dimensional black hole that forms from the collapse of a null shell
Shohreh Gholizadeh Siahmazg, Paul R. Anderson, Raymond D. Clark and Alessandro Fabbri 1265

Microscopic model building for black hole membranes from constraints of symmetry
Swastik Bhattacharya and S. Shankaranarayanan 1275

Einstein-Maxwell-Dilaton-Axion mass formulas for black holes with struts and strings
Dmitri Gal'tsov, Gérard Clément and Igor Bogush 1291

• Black Holes in Alternative Theories of Gravity
Chairpersons: Jutta Kunz and Kamal Hajian

YouTube links: https://youtu.be/FRNGJKhiw7c
https://youtu.be/Tjfmuut1Eo0

Holography for rotating black holes in $f(T)$ gravity
Masoud Ghezelbash 1308

Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity
Semin Xavier, Jose Mathew and S. Shankaranarayanan 1319

Universe in a black hole with spin and torsion
Nikodem Popławski 1327

Asymptotically flat black hole solution in modified gravity
Surajit Kalita and Banibrata Mukhopadhyay 1337

Shadow of a charged black hole surrounded by an anisotropic matter field
Javier Badía and Ernesto F. Eiroa 1343

Constraining modified gravity theories with physical black holes
Sebastian Murk 1351

Penrose suggestion as to pre-Planck-era black holes showing up in present universe data sets discussed, with a possible candidate as to GW radiation which may provide initial CMBR data
A. W. Beckwith 1359

Summary of the parallel session BH3
Kamal Hajian and Jutta Kunz 1372

Binaries

• **Explosive Events Associated with Compact-Object Binary Mergers**
Chairpersons: Chris Belczynski and Jorge Rueda

YouTube links: https://youtu.be/Dwq1ZU3gKrg
https://youtu.be/nw02ylI6R2M

Uncertainties in kilonova modeling
C. L. Fryer, C. J. Fontes, O. Korobkin et al. 1391

• **Post-Newtonian and Post-Minkowskian Corrections for Binary Gravitating Systems**
Chairperson: Johannes Bluemlein

YouTube link: https://youtu.be/wfiLG5r08yE

Tutti-Frutti method: Recent developments in the PN/PM/SF treatment of the gravitational two-body problem
Donato Bini and Andrea Geralico 1405

• **Multichannel Studies of Nonstationary Relativistic Stars**
Chairperson: Vladimir Lipunov

YouTube link: https://youtu.be/usn2PlU_qFA

GRB observations on cubesate satellites in the Universat–SOCRAT project
Sergey I. Svertilov, Michail I. Panasyuk, Vitaly V. Bogomolov et al. 1412

Multiwavelength observations of GRB160625B by MASTER, Lomonosov, Konus-Wind and three stage collapse
V. M. Lipunov, V. A. Sadovnichy, M. I. Panasyuk et al. 1429

The role of the magnetic fields in GRB outflows
N. Jordana-Mitjans, C. G. Mundell, S. Kobayashi et al. 1449

MASTER optical observations of the blazar TXS0506+056 during the IC170922A
V. M. Lipunov, K. Zhirkov, V. G. Kornilov et al. 1467

Boson Stars

• Scalar Fields in Cosmology
Chairpersons: Carlos Herdeiro and Alfredo Macias

YouTube links: https://youtu.be/SyLoguueGKk
https://youtu.be/f9nuo8Jvw-w

A short review on nonlinear perturbation theory of structure formation for modified gravity
Jorge L. Cervantes-Cota and Alejandro Aviles 1474

Testing modified gravity theories with marked statistics
Alejandro Aviles 1494

Dark matter as condensed phase of generic bosons
Elías Castellanos and Jorge Mastache 1513

Cosmic Microwave Background

• Cosmic Backgrounds from Radio to Far-IR
Chairperson: Carlo Burigana

YouTube link: https://youtu.be/4e3Cj5wahck

New Planck tSZ map and its cosmological analysis
H. Tanimura, M. Douspis and N. Aghanim 1527

The CMB dipole: Eppur si muove
R. M. Sullivan and D. Scott 1532

High angular resolution Sunyaev Zel'dovich observations: The case of MISTRAL
E. S. Battistelli, E. Barbavara, P. de Bernardis et al. 1542

Cosmological and astrophysical results exploiting magnification bias with high-z sub-millimetre galaxies
L. Bonavera, M. M. Cueli and J. Gonzalez-Nuevo 1557

The impact of the Lorentz symmetry violation on the CMB polarization
Seddigheh Tizchang, Rohoollah Mohammadi and She-Sheng Xue 1571

Cosmic backgrounds from the radio to the far-infrared: Recent results and perspectives from cosmological and astrophysical surveys
Carlo Burigana, Elia Sefano Battistelli, Laura Bonavera et al. 1579

• **New Horizons in Cosmology with CMB Spectral Distortions**
Chairpersons: Jens Chluba and Andrea Ravenni

YouTube links: https://youtu.be/uBYLO4Smw3o
https://youtu.be/5oAPfzAe35k

CMB μT cross-correlations as a probe of PBH scenarios
Ogan Özsoy and Gianmassimo Tasinato 1609

Theoretical and numerical aspects of CMB spectral distortions from non-thermal energy injections
Sandeep Kumar Acharya, Jens Chluba and Abir Sarkar 1628

BISOU: A balloon project to measure the CMB spectral distortions
B. Maffei, M. H. Abitbol, N. Aghanim et al. 1633

Cosmic microwave background spectral distortions constraints on decaying dark matter particles and axion-like particles using *COBE/FIRAS* and *EDGES*
Boris Bolliet 1645

The COSmic Monopole Observer (COSMO)
S. Masi, E. Battistelli, P. de Bernardis et al. 1654

• **Status of the $\mathbf{H_0}$ and $\boldsymbol{\sigma_8}$ Tensions: Theoretical Models and Model-Independent Constraints**
Chairpersons: Joan Solà Peracaula and Adrià Gómez-Valent

YouTube links: https://youtu.be/VNWZ1Bzjus4
https://youtu.be/zEBOCwvetKE
https://youtu.be/nzgC7qV9H_Y

Measuring the Hubble constant H_0 from gravitational lensing
Liliya L. R. Williams 1672

Extra components consistency in the Hubble tension and BBN
Osamu Seto and Yo Toda 1686

Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology
Nick E. Mavromatos 1693

Early and not so early dark energy. What do cosmological observations tell us about them?
Adrià Gómez-Valent, Ziyang Zheng, Luca Amendola, Valeria Pettorino and Christof Wetterich 1713

Renormalized $\rho_{\rm vac}$ without m^4 terms
Cristian Moreno-Pulido and Joan Solà Peracaula 1733

BD-ΛCDM and running vacuum models: Theoretical background and current observational status
Javier de Cruz Pérez, Joan Solà Peracaula, Adrià Gómez-Valent and Cristian Moreno-Pulido 1752

Cosmological tensions: Hints for a new concordance model?
E. Di Valentino 1770

Solving both H_0 and σ_8 tensions in $f(T)$ gravity
Emmanuel N. Saridakis 1783

Precision Cosmology and Hubble tension in the era of LSS surveys
G. Fanizza 1792

• **Effects of Primordial Perturbations Enhancement: From Black Holes Formation to CMB Anomalies**
Chairpersons: Antonio Enea Romano and Krzysztof Turzynski
YouTube link: https://youtu.be/frjjONXbd1M

Primordial black holes arise when the inflaton falls
Keisuke Inomata 1803

Effects of the modification of gravity on the production of primordial black holes
Sergio Andrés Vallejo Peña 1809

Cosmic Strings

• **Cosmic Strings**
Chairpersons: Reinoud Jan Slagter and Batool Imtiaz
YouTube link: https://youtu.be/ZpaU82ZUHzM

$U(1)$ local strings in generalized hybrid metric-Palatini gravity
Hilberto M. R. da Silva, Tiberiu Harko, Francisco S. N. Lobo and João Luís Rosa 1820

New evidence of the azimuthal alignment of quasars spin vector in Large Quasar Groups and cosmic strings
R. J. Slagter 1835

Summary parallel session cosmic strings I
Reinoud Jan Slagter and Batool Imtiaz 1848

• **From Cosmic Strings to Superstrings**
Chairpersons: Carlos Martins and Ivan Rybak
YouTube link: https://youtu.be/LJFtV_4aSAg

Scaling solutions of wiggly cosmic strings
A. R. R. Almeida and C. J. A. P. Martins 1851

High resolution calibration of string network evolution
J. R. C. C. C. Correia and C. J. A. P. Martins 1871

Radiation from Global Cosmic Strings using adaptive mesh refinement
Amelia Drew and E. P. S. Shellard 1891

Analysing the scaling density of axion strings
A. Lopez-Eiguren 1898

Electroweak axion string and superconductivity
Yu Hamada, Yoshihiko Abe and Koichi Yoshioka 1912

Dark Energy and Large Scale Structure

• **Dark Energy and the Accelerating Universe**
Chairpersons: Alexei Starobinky and David Polarski
YouTube links: https://youtu.be/JZEPRS_rqbE
https://youtu.be/3aG_AT4UzWE
https://youtu.be/ZlTWquG-pFk

Hints for the $H_0 - r_d$ tension in uncorrelated Baryon Acoustic Oscillations dataset
Denitsa Staicova 1923

Observational constraints on nonlinear matter extensions of general relativity
E.-A. Kolonia and C. J. A. P. Martins 1935

Constraining the dark energy-dark matter interaction model using low-redshift observations
Archana Sangwan, Joseph P. J. and S. Shankaranarayanan 1948

On the evolution of inhomogeneous perturbations in the ΛCDM model and $f(R)$ modified gravity theories
T. Schiavone and G. Montani 1961

Soft dark energy and soft dark matter
Emmanuel N. Saridakis 1970

A simple parametrisation for coupled dark energy
Vitor da Fonseca, Nelson J. Nunes and Tiago Barreiro 1979

• **Cosmography with Gravitational Lensing**
Chairpersons: Claudio Grillo and Mimoza Hafizi
YouTube link: https://youtu.be/FXplIAvUDBM

A tale of two double quasars: Hubble constant tension or biases?
L. J. Goicoechea and V. N. Shalyapin 1990

Dark Matter

• **Interacting Dark Matter**
Chairpersons: Nikolaos Mavromatos
YouTube links: https://youtu.be/zOshKJlwD-Y
https://youtu.be/JR44dh2GYik

Dark energy and dark matter unification from dynamical space time: BBN constraints
D. Benisty 2005

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2013

LHC experiments for long-lived particles of the dark sector
Vasiliki A. Mitsou 2029

Constraining the interactions in the dark sector with cosmological data
Adrià Gómez-Valent, Valeria Pettorino and Luca Amendola 2050

Running vacuum interacting with dark matter or with running gravitational coupling. Phenomenological implications
Joan Solà Peracaula 2069

Dark matter properties from the Fornax globular cluster timing: Dynamical friction and cored profiles
D. Blas 2089

Growth of linear perturbations in a universe with superfluid dark matter
S. Banerjee, S. Bera and D. F. Mota 2101

Interacting dark sector in the late Universe: Mapping fields and fluids, and observational signatures
Joseph P. J. and S. Shankaranarayanan 2119

The role of self interactions in the cosmological evolution of warm dark matter
R. Yunis, C. R. Argüelles, D. López Nacir et al. 2127

Interaction energy between a charged medium and its electromagnetic field as a dark matter candidate
Mayeul Arminjon 2139

• Dark Matter Searches with Liquid Xenon and Argon Detectors and Self Gravitating Systems and Dark Matter
Chairpersons: Marco Merafina and Soroush Shakeri and She-Sheng Xue

YouTube link: https://youtu.be/H9oGYnGq9pI

The maximum mass of dilute axion stars
Pierre-Henri Chavanis 2149

A dark matter solution for the XENON1T electron excess and the galactic center 511 keV line
Yasaman Farzan 2174

Preliminary results of rich galaxy clusters' spatial distribution analysis on CfA2 Redshift Survey data: Compact objects or dark matter presence at redshift less 0.032
I. V. Arkhangelskaja, A. M. Galper, L. N. Khanh and D. N. Dorosheva 2189

• Dark Matter: Beyond ΛCDM
Chairpersons: Carlos Argüelles and Andreas Krut

YouTube links: https://youtu.be/hdKeo5L7pYE
https://youtu.be/i0IPHXzmV-s

Probing the nature of dark matter with Milky Way subhaloes
M. R. Lovell 2202

Addressing classical cosmological back-reaction with multiple scales
Yonadav Barry Ginat 2217

Imaging formation process for DM profiles
Omar de J. Cabrera-Rosas and Tonatiuh Matos 2222

The self-gravitating Fermi gas in Newtonian gravity and general relativity
Pierre-Henri Chavanis 2230

• Dark Matter and Rare Processes
Chairpersons: Rita Bernabei and Zurab Berezhiani

YouTube links: https://youtu.be/wheVrbETP_0
https://youtu.be/dVdZ4TIDxt0

The dark matter: DAMA/LIBRA and its perspectives
R. Bernabei, P. Belli, V. Caracciolo et al. 2252

Dark matter directionality approach
R. Bernabei, P. Belli, V. Caracciolo et al. 2272

Collapse models under test by high sensitivity γ-ray and X-ray measurements
C. Curceanu, Kristian Piscicchia, Massimiliano Bazzi et al. 2288

Leptophilic dark matter at linear colliders
P. S. Bhupal Dev 2296

DM6 session: Dark matter and rare processes
R. Bernabei and Z. Berezhiani 2316

• **The Nature of Galactic Halos**
Chairpersons: Francesco De Paolis and Asghar Qadir
YouTube link: https://youtu.be/qIwVnNxi0n0

Primordial black holes as dark matter candidates in the Galactic halo
Lindita Hamolli, Mimoza Hafizi, Francesco De Paolis and Achille A. Nucita 2319

Giant cosmic ray halos around M31 and the Milky Way
S. Recchia, S. Gabici, F. A. Aharonian and V. Niro 2335

A nearly complete census of intergalactic gas using the kinematic Sunyaev-Zel'dovich effect
Chaves-Montero, Jonás 2345

Searching for Intermediate Mass Black Holes in the Milky Way's galactic halo
A. Franco, A. A. Nucita, F. De Paolis, F. Strafella and M. Maiorano . . . 2352

Virial clouds evolution from the last scattering up to the formation of first stars
Noraiz Tahir, Asghar Qadir, Muhammad Sakhi and Francesco De Paolis 2360

Testing Weyl-modified gravity on M31 and Milky Way
Muhammad Bilal and Asghar Qadir 2365

PART C

Education

• **Teaching Einsteinian Physics to School Students**
Chairpersons: David Blair and Matteo Luca Ruggiero
YouTube links: https://youtu.be/W-WV6J8kprg
https://youtu.be/UHYpwKQ09SU

Teaching relativity: A paradigm change
F. Herrmann and M. Pohlig 2371

Teaching relativity: Computer-aided modeling
F. Herrmann and M. Pohlig 2381

Solstices and Equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, and the stellar aberration of Sirius
Costantino Sigismondi and Silvia Pietroni 2388

Positional astrometry at arcsecond accuracy using historical instruments, with light equipment
Costantino Sigismondi and Lorenzo Ricciardi 2398

Daily, seasonal, and equinoctial solar paths on a school soccer field
Costantino Sigismondi 2411

Teaching relativity at the AstroCamp
C. J. A. P. Martins 2415

Sungrazing comets as General Relativistic gravitational probes
Silvia Pietroni and Costantino Sigismondi 2424

The three Summer solstice's markers of 1721 unveiled in the Basilica of Santa Maria degli Angeli in Rome
Costantino Sigismondi 2428

Einstein-First: Bringing children our best understanding of reality
A. Popkova, K. Adams, S. Boublil et al. 2438

Exact Solutions

• **Exact Solutions in Four and Higher Dimensions**
Chairpersons: David Blair and Matteo Luca Ruggiero

YouTube link: https://youtu.be/GHZUS5-4gVQ

Kundt spacetimes in the Einstein–Gauss–Bonnet theory
R. Švarc, J. Podolský and O. Hruška 2453

Exact decoupled solutions in curvature-matter coupled gravity
M. Sharif and Fizza Furqan 2464

Tolman-Oppenheimer-Volkov conditions beyond spherical symmetry
José P. Mimoso, Alan Maciel and Morgan Le Delliou 2479

A spherically symmetric stiff fluid spacetime in light of cosmic structure formation
Daniele Gregoris 2497

• **Exact Solutions (Including Higher Dimensions)**
Chairperson: Susan Scott

YouTube link: https://youtu.be/-hJjJvmmOew

Three-parameter solution for the null-surface formulation in 2+1 dimensions
Tina A. Harriott and J. G. Williams 2510

New exact stationary cylindrical anisotropic fluid solution of GR
M.-N. Célérier 2522

Early Universe

• **Quantum Fields**
Chairperson: Andrei Lebed

YouTube links: https://youtu.be/BLaTp0r0TkQ
https://youtu.be/ZifgaSDy5Vc

Hydrodynamic representation and energy balance for the Dirac and Weyl fermions in curved space-times
Tonatiuh Matos, Omar Gallegos and Pierre-Henri Chavanis 2533

Breakdown of the Equivalence Principle for a composite quantum body
A. G. Lebed 2551

Extended DeWitt-Schwinger subtraction scheme, heavy fields and decoupling
Antonio Ferreiro and Jose Navarro-Salas 2557

Renormalization and decoupling for the Yukawa model in curved spacetime
Sergi Nadal-Gisbert, Antonio Ferreiro and José Navarro-Salas 2562

Trace anomaly and evaporation of spherical black holes
P. Meda 2573

On decay of the false Unruh vacuum
A. Shkerin 2587

Behaviour of noise kernel in de Sitter and FRW space-times
Ankit Dhanuka and Kinjalk Lochan 2600

Breaking Buchdahl: Ultracompact stars in semiclassical gravity
Julio Arrechea, Carlos Barceló, Raúl Carballo-Rubio and Luis J. Garay 2608

Einstein anomaly with tensors of odd order in six dimensional curved space
Kohei Yamamoto and Satoshi Yajima 2619

Quantum memory and BMS symmetries
Sanved Kolekar and Jorma Louko 2623

• Topological Methods, Global Existence Problems, and Spacetime Singularities
Chairperson: Spiros Cotsakis

YouTube link: https://youtu.be/H8Itnc_C10c

Gravitational singularities, scattering maps for bouncing, and structure-preserving algorithms
Philippe G. LeFloch 2630

Brane-world asymptotics in a nonlinear fluid bulk
I. Antoniadis, S. Cotsakis and Ifigeneia Klaoudatou 2645

Primordial synchronization of Mixmaster spatial points
Spiros Cotsakis 2657

• The Early Universe
Chairperson: Stefano Ansoldi

YouTube links: https://youtu.be/m80hHWgOlFs
https://youtu.be/6tyaZ8MMVtw

Quintessential inflation from Lorentzian slow roll
David Benisty 2663

Condensed light, quantum black holes and L-CDM cosmology: Experimentally suggested and tested unified approach to dark matter, dark energy, cosmogenesis and two-stage inflation
Victor Borsevici 2672

Helical magnetic fields lead to baryogenesis
Ashu Kushwaha and S. Shankaranarayanan 2692

Polymer Quantization of the Isotropic Universe: Comparison with the Bounce of Loop Quantum Cosmology
G. Barca, E. Giovannetti, F. Mandini and G. Montani 2700

General relativistic evolution equations for density perturbations in open, flat and closed FLRW universes and the problem of structure formation
Pieter G. Miedema 2708

Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data
M. Berti 2726

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2737

Fundamental Interactions and Stellar Evolution

• Why and How the Sun and the Stars Shine: The Borexino Experiment
Chairpersons: Giampaolo Bellini, Dmitry Naumov, Gioacchino Ranucci, Gemma Testera

YouTube links: https://youtu.be/hh5wDnM8miU
https://youtu.be/TG-HgBf7W4s
https://youtu.be/HLiis2LFeEs

Experimental detection of the CNO cycle
B. Caccianiga, N. Rossi, G. Testera et al. 2753

Borexino detector performances
A. Caminata, M. Agostini, K. Altenmüller et al. 2765

Study of antineutrinos from the Earth and the Cosmos with the Borexino detector
Sandra Zavatarelli, M. Agostini, K. Altenmuller et al. 2774

Unveiling the engine of the Sun: Measurements of the *pp*-chain solar neutrinos with Borexino
D. Guffanti, A. C. Re and O. Smirnov 2785

Electron neutrino survival probability in the energy range 200 keV–15 MeV
Marco Pallavicini on behalf of the Borexino Collaboration 2804

The relevance of pp-chain and CNO-cycle neutrino measurements for solar physics
F. L. Villante and A. M. Serenelli 2815

Role of the CNO cycles in stars
A. Ianni 2835

Geoneutrino observation
Tadao Mitsui for the KamLAND Collaboration 2840

Synthesis of the session: Why and how the Sun and the stars shine
Gianpaolo Bellini 2845

• **Rotation in Stellar Evolution**
Chairperson: Georges Meynet

YouTube link: https://youtu.be/9LNoiwa0nv8

The internal rotation of low-mass stars from solar and stellar seismology
G. Buldgen and P. Eggenberger 2848

The rotation of supermassive stars
L. Haemmerlé 2865

Fast Transients

• **What Can We Learn from a Growing Sample of Fast Radio Bursts?**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/yo4n1SgfUrQ
https://youtu.be/NwGooPauhjU

Cosmology with high-redshift FRBs
A. Fialkov 2880

• **Non Standard Cosmological Probes**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/EEFUgiFeMck
https://youtu.be/Ryb15AINfMs

Closing the cosmological loop with the redshift drift
C. J. A. P. Martins, C. S. Alves, J. Esteves, A. Lapel and B. G. Pereira 2890

Gamma-Ray Bursts as potential cosmological probes
L. Izzo 2906

Surface brightness fluctuations: The method and future applications
Michele Cantiello 2915

Preliminary results of analysis of Ia supernovae redshift distributions on data of the Asiago Supernova and Open Supernova Catalogues
I. V. Arkhangelskaja 2930

• **Photospheric Emission in GRBs**
Chairpersons: Gregory Vereshchagin and Damien Bégué

YouTube links: https://youtu.be/ZifnzoUXIFc
https://youtu.be/n7vWAVhFXiU

Understanding prompt emission: Where do we stand?
Asaf Pe'er 2946

On explaining prompt emission from GRB central engines with photospheric emission model
M. Bhattacharya and P. Kumar 2957

Monte Carlo simulations of photospheric emission in Gamma Ray Bursts
T. M. Parsotan and D. Lazzati 2972

The photosphere emission spectrum of hybrid relativistic outflow for Gamma-Ray Bursts
Yan-Zhi Meng, Jin-Jun Geng and Xue-Feng Wu 2982

On diffusive photospheres in Gamma-Ray Bursts
G. V. Vereshchagin 2989

Summary of the parallel session GB3
G. V. Vereshchagin and D. Bégué 3002

• High and Very High Energy Emission from Gamma-Ray Bursts
Chairpersons: Francesco Longo and Fabian Schüssler

YouTube links: https://youtu.be/ZUkLyyYowaM
https://youtu.be/zwN1mNVeUzA

Synchrotron and synchrotron self-Compton emission components in GRBs detected at very high energies
Jagdish C. Joshi, Vikas Chand and Soebur Razzaque 3009

The VERITAS gamma-ray burst follow-up program
D. Ribeiro for the VERITAS Collaboration 3017

MAGIC view of Gamma-Ray Bursts at very high energies
A. Berti on behalf of the MAGIC Collaboration 3030

Prospects for VHE monitoring of gamma-ray bursts with SWGO
G. La Mura, U. Barres de Almeida, R. Conceição et al. 3041

Theoretical implications on the very high energy emission from GRB 190114C
D. Miceli, A. Berti, Z. Bosnjak et al. on behalf of the MAGIC Collaboration 3052

AGILE and GRBs: 13 years of observations
A. Ursi on behalf of the AGILE Team 3062

Searching for Gamma-Ray Bursts with the High-Altitude Water Cherenkov (HAWC) observatory
K. L. Engel for the HAWC Collaboration 3074

• **Electromagnetic Counterparts of Compact Binary Mergers**
Chairpersons: Jonathan Granot and Paz Beniamini

YouTube link: https://youtu.be/AS7bwaT48us

CALET search for gamma-ray counterparts of gravitational wave events
Masaki Mori for the CALET Collaboration 3084

• **Unusual and New Types of Gamma-Ray Bursts**
Chairperson: Binbin Zhang

YouTube link: https://youtu.be/VrF0iU8Q6us

Off-axis jet scenario for early afterglow emission of low-luminosity Gamma-Ray Burst GRB 190829A
Yuri Sato, Kaori Obayashi, Ryo Yamazaki, Kohta Murase and Yutaka Ohira 3095

• **Gamma-Ray Burst Correlations: Observational Challenges and Theoretical Interpretation**
Chairpersons: Maria Giovanna Dainotti and Liang Li

YouTube links: https://youtu.be/gcEy2h6y1jg
https://youtu.be/m2pCT6RAmI0
https://youtu.be/GTCvcVhBjN4

GRB prompt phase spectra under backscattering dominated model
Mukesh Kumar Vyas, Asaf Pe'er and David Eichler 3101

Applying models of pulsar wind nebulae to explain X-ray plateaux following short Gamma-Ray Bursts
L. C. Strang and A. Melatos 3107

Searching for strange quark planets
Xu Wang, Yong-Feng Huang and Bing Li 3118

Probe the universe by using Gamma-Ray Bursts with X-ray plateaus
Fan Xu and Yong-Feng Huang 3124

A new perspective on cosmology through Supernovae Ia and Gamma Ray Bursts
B. De Simone, V. Nielson, E. Rinaldi and M. G. Dainotti 3130

Theory of plateau phase in Gamma-Ray Bursts
Asaf Pe'er 3141

Exploring the canonical behaviour of long Gamma-Ray Bursts using an intrinsic multi-wavelength afterglow correlation
S. R. Oates, J. L. Racusin, M. De Pasquale et al. 3150

• **GRB 170817A and Binary Models**
Chairpersons: Marica Branchesi and Giulia Stratta
YouTube link: https://youtu.be/FRBvkaLX5WU

Kilonova emission observed so far: A comparison with AT2017gfo
A. Rossi 3162

• **Binary-Driven Hypernovae of Type 1, 2 and 3**
Chairpersons: Carlo Luciano Bianco, Christian Cherubini and Simonetta Filippi
YouTube link: https://youtu.be/8IFLfget3C0

Neutrinos and gamma-ray production from proton-proton interactions in binary-driven hypernovae framework
S. Campion, J. D. Melon Fuksman and J. A. Rueda Hernandez 3172

General relativistic turbulence in spherically symmetric core-collapse supernovae simulations
L. Boccioli, G. J. Mathews and E. P. O'Connor 3184

Gravitational Waves

• **Sources of Gravitational Waves**
Chairperson: Andrew Melatos
YouTube links: https://youtu.be/Lr8b9nKsFm4
https://youtu.be/xuqJOXPDyVY
https://youtu.be/qzOrgojHsMg
https://youtu.be/KCJoWMd71pE

Mountain formation by repeated, inhomogeneous crustal failure in a neutron star
A. D. Kerin and A. Melatos 3194

Gravitational waves from neutrino mass generating phase transitions
Nobuchika Okada and Osamu Seto 3206

Efficiency of registration of chirp bursts and signals of collapsing stars by the Euro-Asian network of GW interferometers
V. N. Rudenko, S. L. Andrusenko, D. P. Krichevskiy and G. D. Manucharyan 3219

Joint analysis method on gravitational waves and low-energy neutrinos to detect core-collapse supernovae
O. Halim, C. Casentini, M. Drago et al. 3228

• **Mid-frequency Gravitational Waves (0.1–10 Hz): Sources and Detection Methods**
Chairperson: Wei-Tou Ni
YouTube link: https://youtu.be/sJ6A7a73Vxw

A cryogenic and superconducting inertial sensor for the Lunar Gravitational–Wave Antenna, the Einstein Telescope and Selene-physics
F. Badaracco, J. V. van Heijningen, E. C. Ferreira and A. Perali 3245

Space gravitational wave antenna DECIGO and B-DECIGO
S. Kawamura and the DECIGO working group 3254

Summary of the parallel session GW2
Dongfeng Gao, Wei-Tou Ni, Jin Wang et al. 3261

• **Numerical Relativity and Gravitational Wave Observations**
Chairperson: Nigel Bishop

YouTube links: https://youtu.be/eZPytAU4Zmk
https://youtu.be/th-3KqkxDnU

Salient features of the optimised PyCBC IMBH search
Koustav Chandra, Archana Pai, V. Villa-Ortega et al. 3277

Matter shells modifying gravitational wave signals
Monos Naidoo, Nigel T. Bishop and Petrus J. van der Walt 3286

Odd-dimensional gravitational waves from a binary system on a three-brane
D. V. Gal'tsov and M. Khlopunov 3301

Developments in numerical relativity and gravitational wave observations
Nigel T. Bishop 3309

High Energy

• **Very High Energy Gamma Rays**
Chairpersons: Razmik Mirzoyan and Alessandro De Angelis

YouTube link: https://youtu.be/jHM1RH20ZyM

Insights into the Galactic Center environment from VHE gamma-ray observations with ground-based facilities
C. Fruck 3316

The TAIGA experiment
M. Tluczykont, I. I. Astapov, A. K. Awad et al. 3324

Science perspectives of the Southern Wide-field Gamma-ray Observatory (SWGO)
K. L. Engel for the SWGO Collaboration 3343

• **Future Missions for High-Energy Astrophysics**
Chairpersons: Filippo Frontera and Shaolin Xiong

YouTube link: https://youtu.be/JgDsZX6RUkU

Laue lenses: Focusing optics for hard X/soft Gamma-ray astronomy
L. Ferro, M. Moita, P. Rosati et al. 3355

ASTENA: A mission concept for a deep study of the transient gamma-ray sky and for nuclear astrophysics
E. Virgilli, F. Frontera, P. Rosati et al. on behalf of the ASTENA Collaboration 3368

Polarimetric prospects of a new hard X-soft gamma-ray space mission for next decades
M. Moita, L. Ferro, F. Frontera et al. 3385

• **The SRG Mission: First Results from eROSITA and ART-XC**
Chairperson: Andrea Merloni
YouTube link: https://youtu.be/l0t1B716UcM

Prospect for WHIM detection in the cosmic web by *SRG/eROSITA*
H. Tanimura and N. Aghanim 3400

• **eXTP – Enhanced X-Ray Timing and Polarimetry Mission**
Chairpersons: Marco Feroci and Fangjun Lu
YouTube link: https://youtu.be/3d08eKmsImI

The role of eXTP in the multi-messenger astronomy era
G. Stratta and Gor Oganesyan 3403

• **Observations of HE and UHE Cosmic Rays**
Chairpersons: Ivan De Mitri and Fabio Gargano
YouTube link: https://youtu.be/l9DZyBMWO5c

CALET on the ISS: The first 5 years
Pier Simone Marrocchesi for the CALET Collaboration 3427

The fluxes of charged cosmic rays as measured by the DAMPE satellite
Paolo Bernardini on behalf of the DAMPE Collaboration 3442

Recent results from the Pierre Auger Observatory
E. Roulet for the Pierre Auger Collaboration 3449

The HERD space mission
F.C.T. Barbato on behalf of the HERD Collaboration 3455

PART D

History of Relativity

• **The "Fall and Rise" of Betelgeuse**
Chairperson: Costantino Sigismondi
YouTube link: https://youtu.be/VmbrE2gYmOM

The observation of the stars in daytime and near the horizon
Costantino Sigismondi and Paolo Ochner 3471

Fall and Rise of Betelgeuse: The summary of HR1 session
Costantino Sigismondi and Paolo Ochner 3475

Photometry of Betelgeuse at daylight
Otmar Nickel 3479

Evidence for dynamical changes in Betelgeuse using multi-wavelength data
Sneha Kachhara, Sandip V. George, Ranjeev Misra and G. Ambika 3485

The curious case of Betelgeuse
Jacco Th. van Loon 3494

Variable stars observed from city sites: The 2500 AAVSO-SGQ database
Costantino Sigismondi and Paolo Ochner 3501

Betelgeuse: An introductory course to observational astronomy
Costantino Sigismondi and Tiziana Pompa 3507

The meridian line of the Vatican obelisk to study the stellar aberration
Costantino Sigismondi and Lorenzo Ricciardi 3513

Betelgeuse, Sirius and the stars in the roman *Settecento*
Costantino Sigismondi 3519

• **History of Relativity, Gravitation and Cosmology**
Chairperson: Luis Crispino

YouTube link: https://youtu.be/RNbPUSp95PQ

On Einstein's last bid to keep a stationary cosmology
Salvador Galindo-Uribarri and Jorge L. Cervantes-Cota 3536

Jayme Tiomno: Relativity, gravity, cosmology, and the Marcel Grossmann Meetings
William D. Brewer 3547

A look inside Feynman's route to gravitation
M. Di Mauro, S. Esposito and A. Naddeo 3563

Towards detecting gravitational waves: A contribution by Richard Feynman
M. Di Mauro, S. Esposito and A. Naddeo 3576

Stellar gravitational collapse, singularity formation and theory breakdown
Kiril Maltsev 3596

The Hamilton-Jacobi analysis by Peter Bergmann and Arthur Komar of classical general relativity
D. Salisbury 3626

• Time and Philosophy in Physics
Chairperson: Shokoufe Faraji

YouTube link: https://youtu.be/986v-V5JJEk

The passage of time and top-down causation
Barbara Drossel 3631

Explaining time's passage
Jonathan J. Dickau 3646

A glimpse to Feynman's contributions to the debate on the foundations of quantum mechanics
M. Di Mauro, S. Esposito and A. Naddeo 3657

Summary of the parallel session HR3
Shokoufe Faraji 3671

Neutron Stars

• Dense Matter in Compact Stars
Chairpersons: Alessandro Drago and Jorge Rueda

YouTube links: https://youtu.be/U5Yr0oDVhqY
https://youtu.be/MvMDFh-1_bc

Massive compact stars in the two-families scenario
P. Char, A. Drago and G. Pagliara 3677

Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond
J. Soldateschi, N. Bucciantini and L. Del Zanna 3684

Screening and elastic properties of the NS crust in the OCP approximation
D. Barba González, C. Albertus Torres and M. A. Pérez-García 3703

Tidal deformability as a probe of dark matter in neutron stars
D. Rafiei Karkevandi, S. Shakeri, V. Sagun and O. Ivanytskyi 3713

Binary neutron star mergers with quark matter equations of state
Atul Kedia, Hee Il Kim, Grant Mathews and In-Saeng Suh 3732

Probing dense matter physics with transiently-accreting neutron stars: The case of source MXB 1659-29
Melissa Mendes, Andrew Cumming, Charles Gale and Farrukh J. Fattoyev 3736

• **Compact Stars as Laboratories for Testing Strong Gravity**
Chairpersons: Aurora Perez Martinez and César Augusto Zen Vasconcellos

YouTube link: https://youtu.be/C1KGecdNEfs

Vacuum properties and astrophysical implications
A. Pérez Martínez, M. Pérez-Garcia, E. Rodríguez Querts and A. Romero Jorge 3756

Testing extended theories of gravity with GRBs
L. Mastrototaro 3762

• **Pulsar Power in Physics and Astrophysics and Pulsars and Pulsar Systems at High Energies**
Chairpersons: Andrea Possenti and Pak-Hin Tam

YouTube links: https://youtu.be/gAG29DZwUbM
https://youtu.be/ytdUBFrViHI

News and views regarding PSR J1757–1854, a highly-relativistic binary pulsar
A. D. Cameron, M. Bailes, V. Balakrishnan et al. 3774

On the origin of the unique isolated X-Ray pulsar 1E 161348-5055 with 6.7 hr. spin period
V. Yu. Kim 3785

Advantages of including globular cluster millisecond pulsars in Pulsar Timing Arrays
M. Maiorano, F. De Paolis, A. A. Nucita and A. Franco 3791

Searching for pulsars in globular clusters with the MeerKAT Radio Telescope
F. Abbate on behalf of the MeerTIME/TRAPUM Collaboration 3799

Precision Tests

• **Gravitational Lensing and Shadows**
Chairpersons: Perlick Volker and Oleg Tsupko

YouTube links: https://youtu.be/6DXXWpMJ3IQ
https://youtu.be/F-K1gXn71_Y
https://youtu.be/B-V8r2HMztw

Gravitational lensing by rotating Simpson–Visser black holes
Sushant G. Ghosh and Shafqat Ul Islam 3812

Killing tensors in foliated spacetimes and photon surfaces
Igor Bogush, Kirill Kobialko and Dmitri Gal'tsov 3827

Decoding black hole metrics from the interferometric pattern of relativistic images
V. Bozza 3839

Symplectic evolution of an observed light bundle
N. Uzun 3844

Shadow of black holes with a plasma environment in 4D Einstein-Gauss-Bonnet gravity
Javier Badía and Ernesto F. Eiroa 3856

Aspects of neutrino mass hierarchy in gravitational lensing
Himanshu Swami 3865

Photon regions in stationary axisymmetric spacetimes and umbilic conditions
K. V. Kobialko and D. V. Gal'tsov 3874

Gravitational lensing by charged accelerating black holes
Torben C. Frost 3885

• Experimental Gravitation
Chairpersons: Angela di Virgilio and Claus Lammerzahl

YouTube link: https://youtu.be/mXcxztQ0nyk
https://youtu.be/tE3gIUBviTc

A manmade experiment aimed to clarify the gravity law in the Solar system
Alexander P. Yefremov and Alexandra A. Vorobyeva 3905

Gravitomagnetic field generation using high permittivity materials in superconducting magnetic energy storage devices
G. V. Stephenson 3910

Large ring laser gyroscopes: Geometry stabilization and laser control
U. Giacomelli, N. Beverini, G. Carelli et al. 3920

Dark gravitomagnetism with LISA and gravitational waves space detectors
A. Tartaglia, M. Bassan, G. Pucacco, V. Ferroni and D. Vetrugno 3929

Light rays in the Solar system experiments: Phases and displacements
Pravin Kumar Dahal and Daniel R. Terno 3942

The Ginger project – preliminary results
C. Altucci, F. Bajardi, A. Basti et al. 3956

• **Variation of the Fundamental Constants, Tests of the Fundamental Symmetries and Probes of the Dark Sector**
Chairpersons: Angela Victor Flambaum and Yevgeny Stadnik

YouTube links: https://youtu.be/NAZA-0tWHak
https://youtu.be/juhDGBJ12Lg
https://youtu.be/zxGgDP2sn60

Varying fundamental constants and dark energy in the ESPRESSO era
C. J. A. P. Martins 3963

• **Dragging is Never Draggy: MAss and CHarge Flows in GR**
Chairperson: Oldrich Semerak

YouTube link: https://youtu.be/ho31IgLNxu8

Testing the general relativistic nature of the Milky Way rotation curve with Gaia DR2
Mariateresa Crosta 3970

Spinning cylinders in general relativity: A canonical form for the Lewis metrics of the Weyl class
L. Filipe O. Costa, José Natário and N. O. Santos 3982

Magnetized black holes: The role of rotation, boost, and accretion in twisting the field lines and accelerating particles
Ondřej Kopáček and Vladimír Karas 3999

Spinning particle: Is Newton-Wigner the only way?
V. Witzany 4010

Gravitomagnetic resonance and gravitational waves
Matteo Luca Ruggiero and Antonello Ortolan 4019

Quantum Gravity

• **Loop Quantum Gravity**
Chairpersons: Marcin Kisielowski and Jerzy Lewandowski

YouTube links: https://youtu.be/VaLPseYWh9E
https://youtu.be/WFIgMqrQ07U

Studying the EPRL spinfoam self-energy
Pietropaolo Frisoni 4026

A spin foam framework for the black-to-white hole transition
Farshid Soltani 4045

Holographic properties of the bulk-to-boundary transmission of information in regions of quantum space
Eugenia Colafranceschi 4062

• **Quantum Gravity Phenomenology**
Chairpersons: Giovanni Amelino-Camelia and Jerzy Kowalski-Glikman

YouTube links: https://youtu.be/vYnVb2zNl0o
https://youtu.be/icK8z80Hm-Y
https://youtu.be/svXLa0yhyYY

Minimal length discretization and properties of modified metric tensor and geodesics
Abdel Nasser Tawfik, Fady T. Farouk, F. Salah Tarabia and Muhammad Maher 4074

The structure of the multiverse from the entanglement entropy
Samuel Barroso Bellido 4082

Effective field theory from relativistic Generalized Uncertainty Principle
Vasil N. Todorinov, Saurya Das and Pasquale Bosso 4088

Stelle gravity as the limit of quantum gravity with maximal momentum
V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das 4107

Baryon asymmetry and minimum length
Saurya Das, Mitja Fridman, Gaetano Lambiase and Elias C. Vagenas . . . 4114

On quantum gravity and quantum gravity phenomenology
Douglas Edmonds, Djordje Minic and Tatsu Takeuchi 4126

WKB approach to the gravity-matter dynamics: A cosmological implementation
G. Maniccia and G. Montani 4146

Natural evidence for fuzzy sphere noncommutative geometry: Super-Chandrasekhar white dwarfs
Surajit Kalita, T. R. Govindarajan and Banibrata Mukhopadhyay 4159

A model of polymer gravitational waves: Theory and some possible observational consequences
Angel Garcia-Chung, James B. Mertens, Saeed Rastgoo, Yaser Tavakoli and Paulo Vargas Moniz 4166

• Loop Quantum Gravity: Cosmology and Black Holes
Chairpersons: Jorge Pullin and Parampreet Singh

YouTube links: https://youtu.be/CAuAK31Ukho
https://youtu.be/JlYnEzRiRR0

Primordial power spectrum from a matter-ekpyrotic scenario in loop quantum cosmology
Bao-Fei Li, Sahil Saini and Parampreet Singh 4178

The primordial power spectra in modified loop quantum cosmology
Bao-Fei Li, Javier Olmedo, Parampreet Singh and Anzhong Wang 4188

Primordial perturbations in kinetically dominated regimes of classical and quantum cosmology
B. Elizaga Navascués, R. Jiménez-Llamas and G. A. Mena Marugán 4193

Revisiting the Hamiltonian formalism of the Ashtekar–Olmedo–Singh black hole model
Alejandro García-Quismondo and Guillermo A. Mena Marugán 4211

A comparison of different choices of clocks in a reduced phase space quantization in loop quantum cosmology with an inflationary potential using effective techniques
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4228

Initial conditions in LQC/mLQCs
Bao-Fei Li, Parampreet Singh and Anzhong Wang 4234

Holonomy corrections in effective midisuperspace models
A. Alonso-Bardaji and D. Brizuela 4239

Infrared signatures of quantum bounce in collapsing geometry
Harkirat Singh Sahota and Kinjalk Lochan 4247

Effective black hole interior and the Raychadhuri equation
Keagan Blanchette, Saurya Das, Samantha Hergott and Saeed Rastgoo . . . 4256

Effect of loop quantization prescriptions on the physics of non-singular gravitational collapse
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4267

Summary of the parallel session QG3
Jorge Pullin and Parampreet Singh 4272

Strong Field

• Strong Electromagnetic and Gravitational Field Physics: From Laboratories to Early Universe

Chairpersons: Sang Pyo Kim and She-Sheng Xue

YouTube links: https://youtu.be/kexqTayqFiU
https://youtu.be/laymGp6x9Hg

Reliable equations of state of viscous strong and electroweak matter
A. Tawfik 4277

Neutral fermion pair production by Sauter-like magnetic step
T. C. Adorno, Zi-Wang He, S. P. Gavrilov and D. M. Gitman 4290

On the magnetic field screening in strong crossed electromagnetic field
S. Campion, J. A. Rueda Hernandez, S.-S. Xue and R. Ruffini 4306

Particle creation by strong fields and quantum anomalies
José Navarro-Salas 4317

Constraints on the non-minimal coupling of electromagnetic fields from astrophysical observations
Susmita Jana and S. Shankaranarayanan 4326

New partial resummation of the QED effective action
Silvia Pla and José Navarro-Salas 4337

Can a detector detect soft photons
Sanved Kolekar and Jorma Louko 4347

Breaking of the adiabatic invariance in the production of particles by strong fields
P. Beltrán-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla 4352

Dynamics of relativistic electrons in non-uniform magnetic fields and its applications in quantum computing and astrophysics
Srishty Aggarwal and Banibrata Mukhopadhyay 4362

Validity of the semiclassical approximation in 1+1 electrodynamics: Numerical solutions to the linear response equation
Ian M. Newsome 4374

On Kerr black hole perfect MHD processes in Doran coordinates
C. Cherubini, S. Filippi, A. Loppini et al. 4387

Tadpole contribution to magnetic photon-graviton conversion
N. Ahmadiniaz, F. Bastianelli, F. Karbstein and C. Schubert 4393

• **The Effects of (Non)Linear Electrodynamics on the Properties of Astrophysical/Gravitational Compact Objects**
Chairperson: Seyed Hossein Hendi

YouTube link: https://youtu.be/SAIAXtQhavE

Correspondence of gamma radiation coming from GRBs and magnetars based on the effects of nonlinear vacuum electrodynamics
Tursynbek Yernazarov, Medeu Abishev and Yerlan Aimuratov 4401

Absorption of massless scalar waves by electrically charged regular black holes
Marco A. A. de Paula, Luiz C. S. Leite and Luís C. B. Crispino 4410

White Dwarfs

• **White Dwarf Explosions**
Chairpersons: Robert Fisher and María Pilar Ruiz Lapuente

YouTube links: https://youtu.be/ndaW6u2xuOo
https://youtu.be/IwjYaQaqJeI

Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass
Friedrich K. Röpke, Florian Lach, Sabrina Gronow et al. 4420

Type Ia supernovae and their explosive nucleosynthesis: Constraints on progenitors
Shing-Chi Leung and Ken'ichi Nomoto 4427

Charged polarized white dwarfs with finite temperature as a possible source of type Ia supernovae
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4447

• **White Dwarfs, Magnetic Compact Stars and Nuclear Astrophysics**
Chairpersons: Manuel Malheiro and Jaziel Goulart Coelho

YouTube link: https://youtu.be/onicFElJQnA

CTCV J2056-3014 and other fast-spinning white dwarfs
C. V. Rodrigues, R. Lopes de Oliveira, A. Bruch et al. 4453

Gravitational waves from fast-spinning white dwarfs
M. F. Sousa, J. C. N. de Araujo and J. G. Coelho 4461

Highly magnetized white dwarfs: Implications and current status
B. Mukhopadhyay, M. Bhattacharya, A. J. Hackett et al. 4475

Electron captures and stability of white dwarfs
N. Chamel, L. Perot, A. F. Fantina et al. 4488

Massive hot white dwarfs: Consequences of finite temperature in the structure and on the onset of instabilities
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4508

A study of the infrared emission of SGR/AXPs in a disk scenario and its implications for their origin
Sarah Villanova Borges 4514

Particle acceleration and high energy emission in the white dwarf binaries AE Aquarii and AR Scorpii
P. J. Meintjes, S. T. Madzime, Q. Kaplan et al. 4522

Study the effects of anisotropy on the highly magnetized white dwarfs
Debabrata Deb, Banibrata Mukhopadhyay and Fridolin Weber 4532

List of Participants 4545

PART A

This page intentionally left blank

New results from testing relativistic gravity with radio pulsars

Michael Kramer

Max-Planck-Institut für Radioastronomie
Auf dem Hügel 69, 53177 Bonn, Germany
mkramer@mpifr-bonn.mpg.de

We experience a golden era in testing and exploring relativistic gravity. Whether it is results from gravitational wave detectors, satellite or lab experiments, radio astronomy plays an important complementary role. Here one can mention the cosmic microwave background, black hole imaging and, obviously, binary pulsars. This talk will concentrate on the latter and new results from studies of strongly self-gravitating bodies with unrivalled precision. This presentation compares the results to other methods, discusses implications for other areas of relativistic astrophysics and will give an outlook of what we can expect from new instruments in the near future.

Keywords: Pulsars; Neutron stars; Black holes; Gravitational-wave astrophysics; Experimental tests of gravitational theories

PACS numbers: 97.60.Gb; 97.60.Jd; 97.60.Lf; 04.30.Tv; 04.80.Cc

1. Introduction

The latest Marcel-Grossmann Meeting has showed once more that gravity can be explored and tested in many different ways. Whether it is via lab experiments on Earth, in the Solar system or in outer space. Whether it is via radio astronomy, optical/infrared astronomy, high-energy observations or gravitational wave detectors. Nearly all experiments use techniques that were not even remotely thinkable at the time, when Einstein conceived general relativity (GR). And yet, all these experiments that the theory has been confronted with, seem to confirm GR's predictions. Obviously, the phenomena of Dark Matter and Dark Energy, or the incompatibility of GR with quantum physics, justify the search for alternatives and the continuing search for experiments that may eventually show a flaw in Einstein's theory.

In this endeavour it is important to complement the different approaches as they test different regimes of gravity, in each of which one may find a deviation for a given theory. This is illustrated in Figure 1. In the solar systems, gravity tests are made in the quasi-stationary weak-field regime, often involving satellite experiments of high precision (e.g. Ref. 1). The quasi-stationary strong-field regime can be probed with experiments such as stellar orbits around Sgr A* (e.g. Ref. 2) or radio images of black holes (Ref. 3). But also pulsar experiments contribute to this regime, and indeed they already allow precision tests of the radiative properties of gravity (e.g. Ref. 4), as we also show in the following. Using radio pulsars

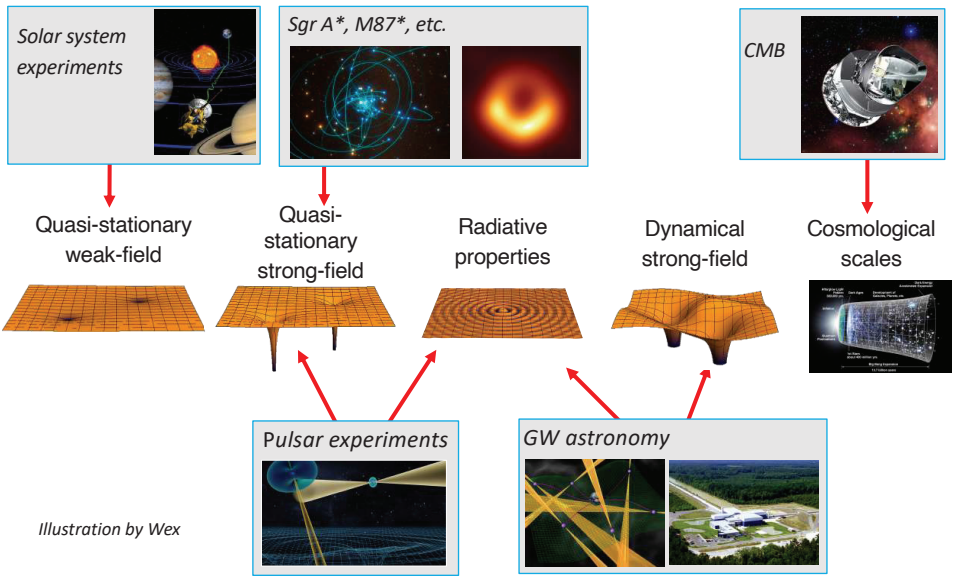


Fig. 1. Different regimes of gravity tests and examples of methods and experiments that allow tests in those.

also as galaxy-sized gravitational wave (GW) detectors (e.g. Refs. 5), one also enters the dynamical strong-field regime, where obviously also the Earth-bound GW detectors are operating (e.g. Ref. 6). Finally, theories should also be tested on the largest scales, which can be done, for instance, with studies of the cosmic microwave background (CMB).

We note that in most of these experimental methods (certainly in those examples given here) radio astronomy plays a crucial role. In the following, we will concentrate on recent results involving radio pulsars.

2. Testing gravity with pulsars

The most common, and most well known way, of testing gravity with pulsars is via a technique called pulsar timing.[7] Less common but still powerful is the usage of pulse structure data that allow us to probe precessional effects when the line-of-sight through the emission beam is changing, modifying the observed pulse properties like width or generally pulse shape.[8] In extreme cases, the pulsar may completely disappear from our view after a while.[9,10])

In pulsar timing one makes use of the often very stable rotation of a pulsar that leads to a regularity in the emitted pulsed signal, caused by a lighthouse effect, that rivals the best atomic clocks on Earth. As an observer on Earth, we can measure the arrival time of the pulses at the telescope and infer from it the motion of the

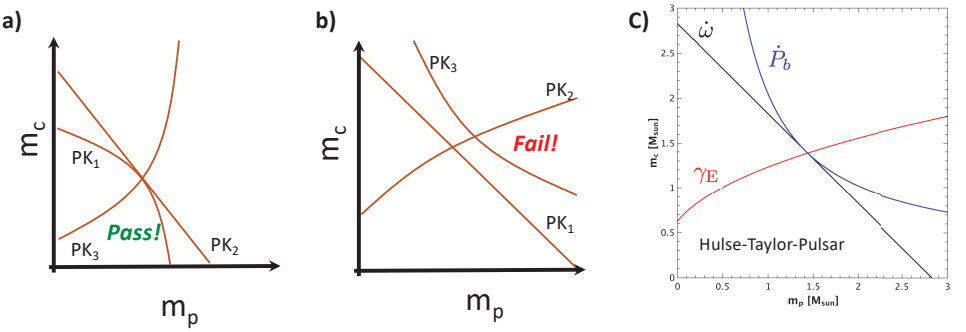


Fig. 2. Gravity tests using binary pulsars performed via the measurements of Post-Keplerian (PK) parameters and their graphical representation in a mass-mass diagram. a) all PK parameter lines insect in a single point: the test is passed. b) the lines do not intersect in a single point: the theory fails the test. c) Example for the Hulse-Taylor pulsar, here shown for general relativity.

pulsar in a binary orbit. The timing precision that is attainable (i.e. better than 100 ns in many cases) is provided by following the rotational phase of the pulsar in a coherent fashion, making measurements more precise as the time baseline increases. In a similar way, tracking the orbital phase reveals periodic and secular deviations from a simple Keplerian orbit, expressed in the so called Post-Keplerian (PK) parameters.[8,11,12] Those can typically be expressed as functions of the Keplerian parameters and the two unknown masses of the binary system. Once two PK parameters can be measured, the two masses can be determined within the framework of a given theory and the value of further PK parameters can be predicted. Once that is done, the predictions can be compared with similar measurements and the theory, hence, can be tested.

One can express such a test in a graphical form, where each PK parameter describes a line in a diagram with the values of two unknown masses as its axes. Two PK parameter lines define an intersection point through which every other PK parameter line must pass. If that is not the case, the theory determining the lines is falsified and needs to be rejected. The measurement uncertainties in the PK parameters is expressed as the width of the PK lines in the mass-mass diagram. See Figure 2.

2.1. *The Double Pulsar*

We demonstrate the power of pulsar timing using the best laboratory currently known, the so called "Double Puslar".[13,14] The system consists of two active radio pulsars – a currently still unique feature of the system – orbiting each other in 147 min. The recycled (old) pulsar (A) in the system has a period of 23 ms, the non-recycled (young) pulsar (B) a period of 2.8 s. An important feature of the system is its edge-on geometry (see Figure 3) that also leads to 30-s long eclipses and our

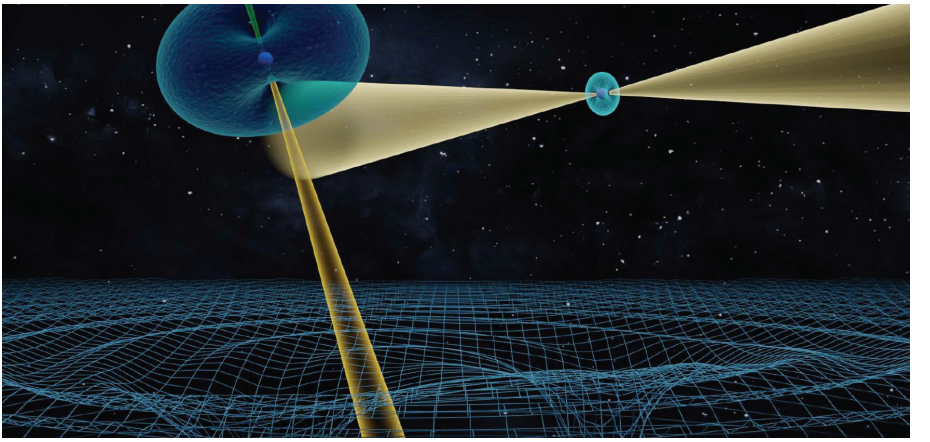


Fig. 3. Artistic impression of the Double Pulsar. Here, two active radio pulsars orbit each other in 147 min in a slightly eccentric orbit ($e = 0.09$). One pulsar is old and recycled ($P = 23$ ms), while the second pulsar is young ($P = 2.8$ s). Figure is not to scale.

ability to measure light-propagation effects (see below). The previous gravity tests with the system made usage of the ratio of the masses of the two pulsars, that can be directly (i.e. theory independently to 1PN order) measured from the two semi-major axes the pulsars' orbit around the common centre of mass.[4] Meanwhile the data have improved very significantly in both precision and length. The latest results were presented by Kramer et al. (Ref. 15), reporting the measurement of, in total, seven PK parameters.

What is important and different to previous experiments, the measurement precision is now so high that for the first time higher-order contributions need to be taken into account for some of the PK parameters. This includes the contribution of the A pulsar's effective mass loss (due to spin-down) to the observed orbital period decay, a relativistic deformation of the orbit, and the effects of the equation of state of super-dense matter on some of the observed PK parameters via relativistic spin-orbit coupling. The system also provides the currently most precise test of general relativity's quadrupolar description of GWs, validating GR's prediction at a level of 1.3×10^{-4} with 95% confidence.[15] As indicated, the impact of the mass loss due to rotational spin-down is taken into account in this result. Interestingly, as the pulsar is losing rotational energy, and while $E = m\,c^2$, the slow-down corresponds to a mass loss of 8.4 Million tons per second or 3.2×10^{-21} of the mass of A per second.

2.1.1. *Light propagation*

The edge-on geometry of the system allows not only a test of relativistic spin-precession as discussed in Section 3, but it also leads to a "Shapiro-delay", i.e. the a delay in the measured arrival time as the photons have to propagate in the curved

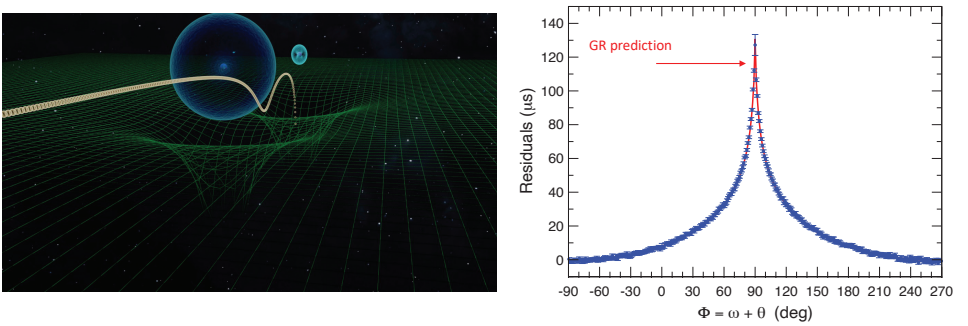


Fig. 4. Shapiro delay measurement in the Double Pulsar. The curvature of spacetime around the companion B (see artistic impression on the left, not to scale) leads to a delay in the arrival times of the pulses of pulsar A (right).

spacetime near the companion pulsar. We illustrate this effect in Figure 4. The data points follow the prediction of GR (red line) remarkable well (at a level of 4×10^{-4} with 95% confidence.[15] This agreement with the data is, however, only achieved when next-to-leading order effects are taking into account: at the 1.5PN limit one has to take into account that the speed of light is finite and that pulsar B moves during the flight time of the photos across the orbit. See Figure 5.

When comparing the Shapiro delay curve with precision data, another effect needs to be considered which relates to the fact the pulsar is indeed a lighthouse rather than a pulsating or oscillating object. This effect is aberration, which for pulsar timing means in principle a delay in arrival time that depends on orbital phase and on the orientation of the pulsar with respect to the orbit and the observer.[12,16] Due to the geometry of the Double Pulsar that does, however, not lead to separately measurable, additional PK parameters.[15] Nevertheless and most interestingly, given the precision obtained in the measurements, the curved spacetime encountered by the photons near superior conjunction leads to a deflection of the signal in the gravitational field of pulsar B. This gravitational signal deflection leads to a small change in the proper angle of emission, which in turn leads to a lensing correction to the classical ("longitudinal") aberration. See Ref. 15 for details. As a result, only those photons reach the observer on Earth that are emitted slightly earlier (or later, respectively) in pulsar rotational phase, before (and after, respectively) superior conjunction. The measurement shown in Figure 6 not only demonstrates the lighthouse character of pulsars but also indicates a prograde alignment of the pulsar spin relative to the total angular momentum vector.

2.1.2. *Orbital deformation, lense-thirring effect and moment-of-inertia*

A further effect that is observed in the Double Pulsar and that is now required to be considered in the data analysis is a relativistic deformation of the orbit. The presence of the effect[11,12] was already seen by Weisberg & Huang (Ref. 17). In

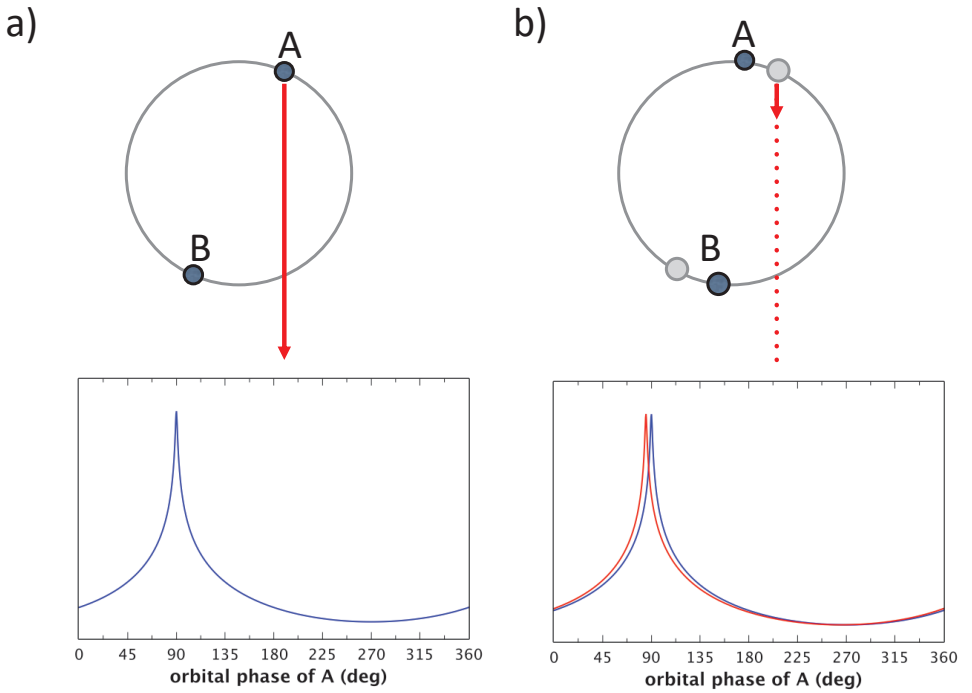


Fig. 5. Illustration of the physical origin and impact of the next-to-leading order (1.5PN) contribution to the Shapiro delay. a) Previously, a "static" Shapiro delay was implemented in timing formula, assuming the position of the companion pulsar B to be fixed during the propagation of A's photons. b) In reality, pulsar B moves during the flight time of the photons, which effectively leads to a small shift in the Shapiro delay curve at the 1.5PN level. Illustration by N. Wex.

this case, ignoring this subtle effect would (wrongly) suggest a tension between the measured "Einstein delay" and its prediction by GR. See Kramer et al. (Ref. 15) for details.

As already indicated, the extreme precision of the experiment now also needs to incooperate the effects of the equation of state (EoS) of super-dense matter on some of the observed post-Keplerian parameters via relativistic spin-orbit coupling. Indeed, the most precisely measured PK parameter, $\dot{\omega}$, representing the precession of the orbit, has a contribution from the Lense-Thirring effect due to the rotation of pulsar A at the 2PN level, which is more than 30 times larger than our measurement error. However, a calculation of the Lense-Thirring contribution requires the knowledge of the Moment-of-Inertia of pulsar A, $I_{\rm A}$, which comes with a significant uncertainty due to our imperfect knowledge of the EoS for neutron star matter. To constrain $I_{\rm A}$ one can use constraints on the EoS obtained from the double neutron star merger GW170817[18] and determine precise measurements for the pulsars. Alternatively, one can ignore any existing constraints on the EoS of neutrons stars to simultaneously determine the pulsar masses, $m_{\rm A}$, $m_{\rm B}$, and $I_{\rm A}$, assuming GR to be

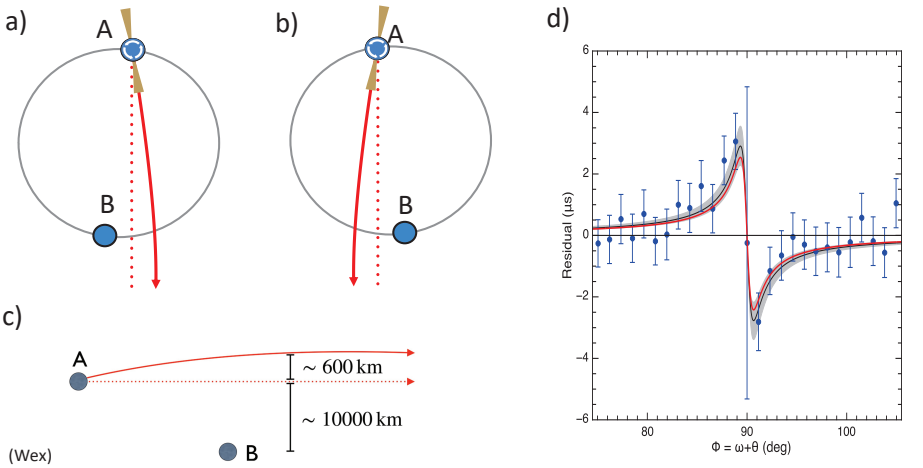


Fig. 6. Illustration and measurement of another next-to-leading order effect that needs to be taken into account for the Shapiro delay. In this case, aberrational light bending due to the curvature of spacetime around pulsar B is considered. a) Before superior conjunction, the photons have to be emitted at a slightly later rotational phase than usual to compensate for the light bending near B to reach the observer on Earth. b) After superior conjunction, the photons have to be emitted slightly earlier, respectively. c) The light bending contributes an additional 600 km to the total impact distance. d) Residuals observed near superior conjunction (i.e. periastron, $\Phi = 0$ deg), when this effect is *not* taken into account. The red curve shows the theoretical prediction arising from this effect, and the black curve corresponds to the best fit, with a 2σ range indicated by the light grey band. This effect allows us to infer the spin direction of pulsar A, indicating a prograde orientation. See Kramer et al. (2021) for details.

the correct theory of gravity. In this case, simply based on pulsar timing, one finds for the the MoI $I_{\rm A} < 3.0 \times 10^{45}\,{\rm g\,cm^2}$ with 90% confidence,[15] which in turn can be directly compared to the results derived from the GW170817 LIGO/Virgo merger and from NICER X-ray timing. Using a universal relation, like the one in Ref. 19, one can convert the obtained probability distribution of $I_{\rm A}$ into a probability distribution for A's radius. With 90% confidence, this gives an upper limit for A's radius of 22 km, a value outside any physically valid EoS but purely based on the arrival times measurements of pulses.[15]

In summary, the Double Pulsar allows to test conservative aspects of the orbital dynamics of two strongly self-gravitating masses up to 2PN order, including a $\sim 1\sigma$ constraint on the Lense-Thirring contribution, which in turn can be used to constrain the MoI of pulsar A under the assumption of GR. The experiments allow to test the propagation of photons in the gravitational field of a strongly self-gravitating (material) body, with next-to-leading-order contributions being clearly present in the timing data, confirmed with a precision of about 10%. We can also test the radiative aspects of GR, yielding a test at 2.5PN level in the equations of motion with a precision of 1.3×10^{-4} (95% C.L.). In terms of overall fractional precision, this is the most precise test of GR's predictions for GW emission

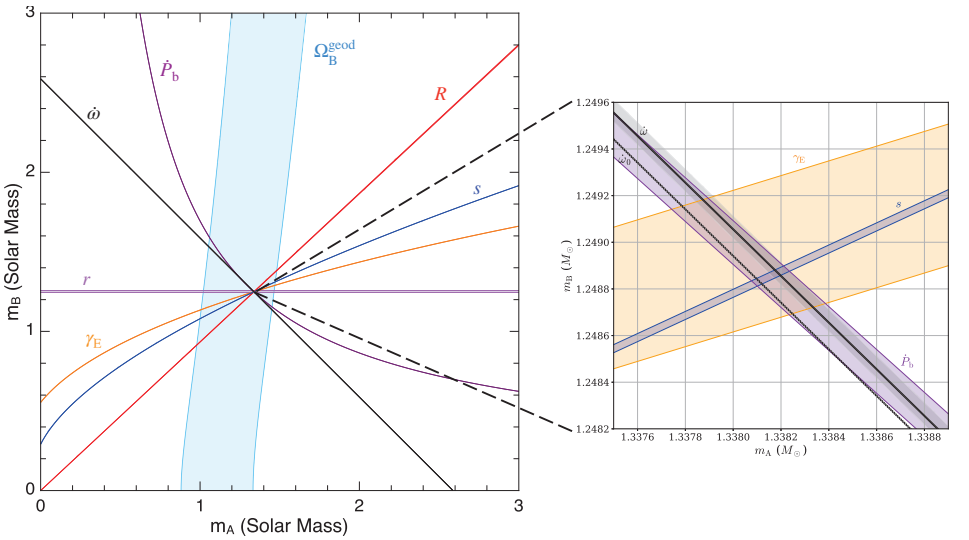


Fig. 7. Mass-mass diagram for the Double Pulsar as presented by Kramer et al. (2021). The right panel is a zoomed-in version of the intersection point of the PK parameter lines in the left panel. The shift of the line for the PK-parameter $\dot{\omega}$ when ignoring the Lense-Thirring effect ($\dot{\omega}_0$) is clearly visible. See text for details.

currently available. These tests are summarized in the mass-mass diagram shown in Figure 7.

3. Relativistic spin-precession

A gyroscope freely falling in curved spacetime suffers a precession with respect to a distant observer. Experiments made in Solar System provide precise weak-field tests and confirm it, e.g. with Lunar Laser Raning or GRAVITY Probe-B. In a binary pulsar system, the pulsar - also being a gyroscope - we observe, strictly speaking, a mixture of different contributions to relativistic spin-orbit coupling. One contribution comes from the motion of the first body around the centre of mass of the system (deSitter-Fokker precession), while the other comes from the dragging of the internal frame at internal frame at the first body due to the translational motion of the companion.[20] This resulting relativistic spin-precession causes nevertheless the expected effect: the pulsar precesses about the total angular momentum, which as a result changes the relative orientation of the pulsar towards Earth with potentially observable consequences as predicted by Damour & Ruffini in Ref. 21 even before the publication of the discovery of the Hulse-Taylor binary pulsar.[22] We note that the orbital angular momentum is expected to be much larger than that of a pulsar in the system, so that the orbital spin practically represents a fixed direction in space. Hence, effectively the pulsar precesses about the orbital angular momentum.

The precession rate, $\Omega_{\rm SO}$, is another PK as introduced above. Assuming GR to be valid, it is given by[23]

$$\Omega_{\rm SO} = T_\odot^{2/3} \left(\frac{2\pi}{P_{\rm b}}\right)^{5/3} \frac{m_{\rm A}(4m_{\rm B}+3m_{\rm A})}{2(m_{\rm B}+m_{\rm A})^{4/3}} \frac{1}{1-e^2}. \quad (1)$$

The effect of relativistic sin precession on pulsar timing is in principle measurable due to a change in the aberration parameters with time.[8] Much more obvious, and hence easier to measure, are the consequences of the changing line-of-sight as the pulsar precesses. We expect changes in the pulse shape (e.g. in simply the measured width, but also more dramatic changes) and especially in the measured linearly polarised emission which is a sensitive probe of the pulsar geometry.[21,24] Such measurements have indeed enabled the detection of spin precession in a number of relativistic binaries. After the first detection in the Hulse-Taylor pulsar B1913+16[9,25] and later in PSR B1534+12,[26,27] it is now possible to convert long-term observations in precise quantitative tests of spin precession and, hence, test of the "effacement" property of a spinning body.

3.1. *Spin precession in the Double Pulsar*

A prime example to study relativistic spin precession is, again, the Double Pulsar. Here, we have two active pulsars, both of which could, in principle show the effects of precession. However, it turns out that the supernova explosion creating the second-born pulsar, B, was a low-kick supernova which left the pre-Supernova orientation of the orbital angular momentum vector unchanged. The spin of pulsar A, having been aligned with the orbital spin during its recycling period, is therefore still aligned with it, so that precession for A is not expected. Indeed, so far, profile changes for A have not been detected, suggesting that A's misalignment angle is less than 3 deg.[28,29]

In contrast, secular changes in the pulse shape and the orbital visibility windows of pulsar B were observed soon after its discovery,[30] providing early evidence for spin precession of B. Around 2008, the pulsar vanished from our view when our line-of-sight moved out of the emission beam.[10] B will eventually re-appear. The exact epoch for this to happen depends on the actual beam shape, but despite a precession period of 71 years (see Table 1), it could happen relatively soon.[10,31]

Meanwhile, it has been possible to track the orientation of the spin axis of pulsar B in a rather fortuitous way: Breton et al. (Ref. 32) studied the time evolution of the $\sim$ 30-s long eclipses of A that are caused by the blocking rotating magnetosphere of B at superior conjunction. This absorption of the background emission, presumably by synchrotron self-absorption,[33] is not complete, but because of the torus-shaped dipolar magnetosphere[a] the light from the background pulsar A is visible every half-turn or full-turn of pulsar B, depending on the orientation of the spin-axis and,

[a]Deviations from a purely dipolar-shaped magnetosphere do exist and are observed. Indeed, about 50% of the magnetosphere is "missing" but for the remaining parts, the torus shape appears more than adequate, as the successful modelling shows. See, for instance Ref. 34.

hence, the magnetosphere. Applying a simple but successful geometrical model,[33] Breton et al. were able to explain the observed modulation of A's lightcurve during the eclipse phase in great detail. Model-fitting indicates a constant magnetic inclination angle and constant misalignment angle, while the azimuthal spin angle is changing with a rate of $\Omega_{\rm SO,B} = 4.77^{+0.66}_{-0.65}{}^{\circ}$ yr^{-1}, in agreement with general relativity[32] (see Table 1).

3.2. *PSR J1906+0746 – The currently best test*

A further observational consequence of relativistic spin precession is its impact on the polarisation properties of binary pulsars. Here, two separate but related aspects are observable.

The first aspect is due to the changing line-of-sight within the pulsar beam. According to the rotating vector model (RVM),[24] we expect an S-like swing of the position angle of the linearly polarised emission component of the studied pulsar. The closer the line-of-sight approaches the magnetic axis, the steeper is the expected S-swing. In general, the shape of the S-swing depends on four quantities: the magnetic inclination angle, α, between the spin and magnetic axes, the impact angle between our line-of-sight and the magnetic axis, β, the pulse phase, ϕ, and, finally, the absolute (reference) position angle, Ψ_0, which is obtained at $\phi = 0$ and identical to the orientation of the pulsar spin-axis projected on the plane of the sky. Such a measurement was first achieved for PSR B1534+12,[26,27] where the impact angle β was observed to change with time.

The second consequence is the change of Ψ_0 due to the changing orientation of the pulsar spin and the resulting projection on the sky.[35] This change of the absolute position angle is periodic, with a period equal to the precessional period. Note that there is also a small contribution due to orbital aberration, which we will ignore here. We note that the measured absolute position angle will be affected by Faraday rotation due to propagation in the magnetised ionised interstellar medium. But this effect can, in principle, be measured and removed.

It is the additional information provided by the change in absolute position angle that allows to break some degeneracies and to potentially determine the true orbital inclination angle, i (rather than $180-i$, as pulsar timing only provides $\sin i$). Modelling the evolution of the position angle swing in the described way requires in total only four free parameters, i.e. the precession rate, $\Omega_{\rm SO}$, the misalignment angle between pulsar spin and total angular momentum vector, δ, the precession reference phase, $\Phi^0_{\rm SO}$, and the magnetic inclination angle of the pulsar, α. A requirement for precession to occur is $\delta > 0$, a measurement of which is very useful to study the formation of the system (see Refs. 34 and 35 for a detailed discussion on this topic).

Both aspects just described are relevant for the beautiful case of PSR J1906+0746. This pulsar is an almost perfect orthogonal rotator, and so this pulsar was discovered with the emission from both magnetic poles being visible

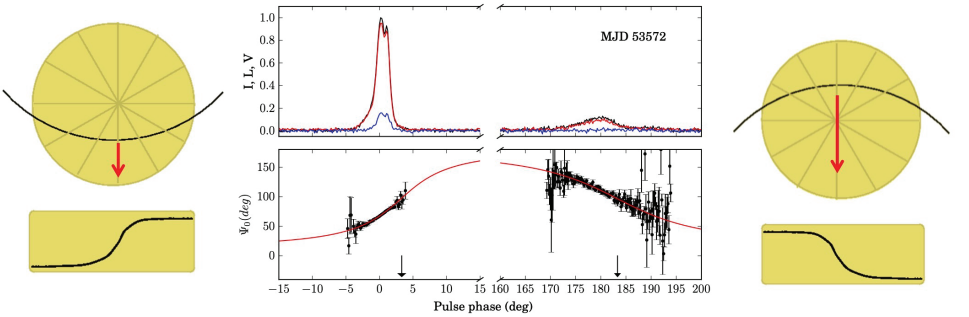


Fig. 8. Pulse profile of PSR J1906+0746 as observed by Desvignes et al. (2019) shown in the middle panel. Both magnetic poles (schematics left and right) were originally visible after discovery. Their observed emission is highly linearly polarised (red line), with a position angle swing (middle lower panel) that resembles a textbook-like rotating vector model. The pulsar also shows a modest amount of circularly polarised emission (middle upper panel, blue line). At the left pole, our line-of-sight moves out of the beam, while at the right pole our line-of-sight crossed the magnetic pole during the long-term monitoring of this source.

(see Figure 8). The emission's high degree of linear polarisation reveals a perfect RVM-like behaviour of the measured position angle that was studied by Desvignes et al. (Ref. 36). Fitting those to the RVM as a function of time using the precessional RVM model[35] described above, allows to measure the precession rate to $\Omega_{\rm SO} = 2.17 \pm 0.11$ deg/yr which is in perfect agreement with the expectation from GR (see Table 1). This test is the currently best test of relativistic spin-precession for strongly self-gravitating bodies.

The study of the position angle reveals even more: Apart from providing measurements of the geometry of the system ($i = 45 \pm 3$ deg, $\alpha = 99.4 \pm 0.2$ deg, $\delta = 104 \pm 9$ deg), the evolution of the position angle swing with time shows how its slope becomes steeper when our lines-of-sight approaches the magnetic pole. It also shows how it swaps sign in slope when crossing the pole, and how it becomes flatter again when moving away from the pole[36] (see right of Figure 9). This is the first unequivocal evidence that the RVM is correct and that the position angle swing of pulsars is in principle determined by the geometry of the pulsar and its fieldlines. Deviation may, however, still occur in cases due to propagation effects or modifications to the field line structure in the emission region.

Furthermore, we can reconstruct the emission beams of the pulsar as projected on the sky and as shown in the left two panels of Figure 9). As the light of sight crosses the two beams, we can obtain "tomographic" images, albeit incomplete ones due to lack of observations in part of the beam. We note from the middle panel, showing the pole that is responsible for the (initially) weaker pulse component in Figure 8, that the emission beam is active on either side of the pole, ruling out models that predict radio emission to be restricted to one side of the pole. Secondly, the emission pattern is not symmetric in the latitudinal direction relative

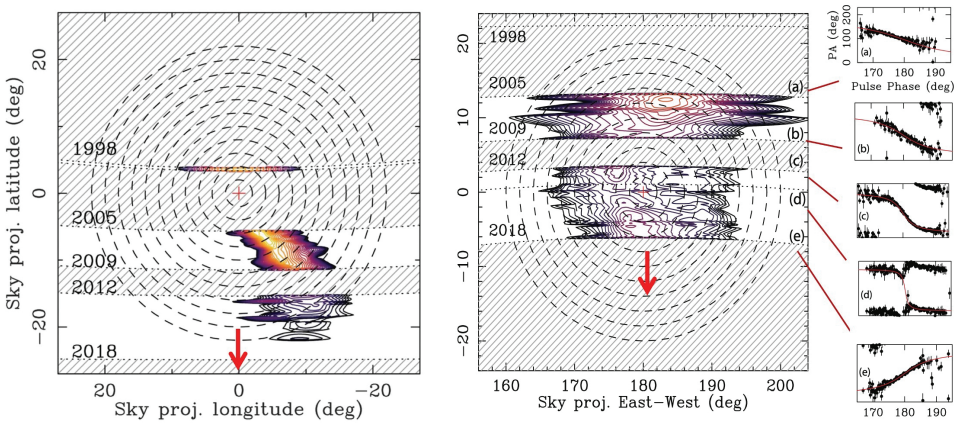


Fig. 9. Beam pattern of the two magnetic poles of PSR J1906+0746 as projected on the sky and observed by Desvignes et al.[36] The beams are reconstructed from observations spanning more than 10 years, when our line-of-sight moved in the indicated direction above both poles. For the right beam, our line-of-sight crossed the magnetic pole. The resulting position angle swing shown on the right evolves with time as expected from the rotating vector model (see text for details). For the left beam, the emission becomes weaker to the beam edge and the initially much stronger component (see Figure 8) is hardly detectable anymore and will soon vanish completely.

to the pole. When looking more closely at our data, we also notice that above the pole itself the emission is reduced, matching theoretical predictions for the current density in the polar cap for an orthogonal rotator.[37] What is also not visible in these total intensity plots shown here (see Ref. 36 for details) is an interesting behaviour of the polarisation characteristics more generally: the handedness of the circularly polarised emission changes when crossing the pole, the fractional linear polarization overall decreases with increasing distance to the beam centre, and finally, emission heights inferred from polarisation data increase from very low emission heights near the pole to emission heights of several hundreds of kilometres at the beam edge, again consistent with theoretical expectations[38] and inferred more indirectly from other pulsars.[39,40]

In summary, somewhat unexpectedly, an experiment designed to test theories of gravity and to tests predictions made by Thibault Damour and Remo Ruffini nearly 50 years ago, leads to unique insight into the emission physics of pulsars.

3.3. *Relativistic spin precession in other pulsars*

Today, spin precession has been routinely observed in all those binary pulsars, where we can expect the pulsar spin to be misaligned with the orbital momentum vector (i.e. systems where we observe the young pulsar or where the last supernova explosion in the system imparted a large kick onto the newly born neutron star) and where the precession rate is high enough due to a sufficient compactness of the orbit. Using the precession rate as given in GR, Eqn. 1, we provide a list of pulsars

where we would expect to observe spin precession. Table 1 gives the corresponding values, sorted by decreasing rate. We note that for the vast majority of sources, spin precession has been detected (see Ref 41). The second highest rate is expected for the highly accelerated binary pulsar J1757−1854 discovered with in the High Time Resolution Universe Survey.[42] As presented by Cameron et al. in this conference, first signs of profile changes may already be present but further data are needed. If this is confirmed, the first top-6 pulsars (and seven out of eight) have all been observed to undergo spin precession. The only other exception in the table is PSR J1756−2251. Here the spin may be aligned with the orbital momentum vector, but studies are in progress to revisit this pulsar with larger precision and longer time baseline.[43]

Table 1. Binary pulsar systems where relativistic spin precession is expected or observed, ordered according to the precession rate as expected from general relativity. Pulsars marked with an asterisk have been identified of showing spin precession.

PSR	P(ms)	$P_{\rm b}$ (d)	x(lt-s)	e	$\Omega_{\rm p}$ (deg yr^{-1})
J0737−3039A/B*	22.7/2770	0.10	1.42/1.51	0.09	4.8/5.1
J1757−1854	22.50	0.18	2.24	0.61	3.1
J1906+0746*	144.1	0.17	1.42	0.09	2.2
B2127+11C*	30.5	0.34	2.52	0.68	1.9
J1141−6545*	394.0	0.20	1.89	0.17	1.4
B1913+16*	59.0	0.33	2.34	0.62	1.2
J1756−2251	28.5	0.32	2.76	0.18	0.8
B1534+12*	37.9	0.42	3.73	0.27	0.5

4. Further recent tests with radio pulsars

In this work, we have so far concentrated on a few prominent, perhaps even spectacular recent examples of using pulsars for testing strong-field gravity. Much more can be studied, including fundamental concepts, such as the Universality of Free Fall (UFF) and the Strong Equivalence Principle (SEP), the possible variation of fundamental constants, or bounds on the graviton mass. We can also test specific alternatives to GR. Here, we refer to recent reviews (see Refs. 44, 45 for more in-depth discussions) and we point to other contributions of the MG16 meeting. Here, we will only highlight some recent examples.

4.1. *Universality of Free Fall (UFF)*

The UFF can be tested in the various regime that we discussed in context of Figure 1. In the solar system, we refer to the superb measurement using test masses aboard the MICROSCOPE satellite that was launched in April 2016 and operated until October 2018. The results indicate that any violation of the weak equivalence principle, Δ, is smaller than $\Delta < 10^{-15}$ (see Ref. 1).

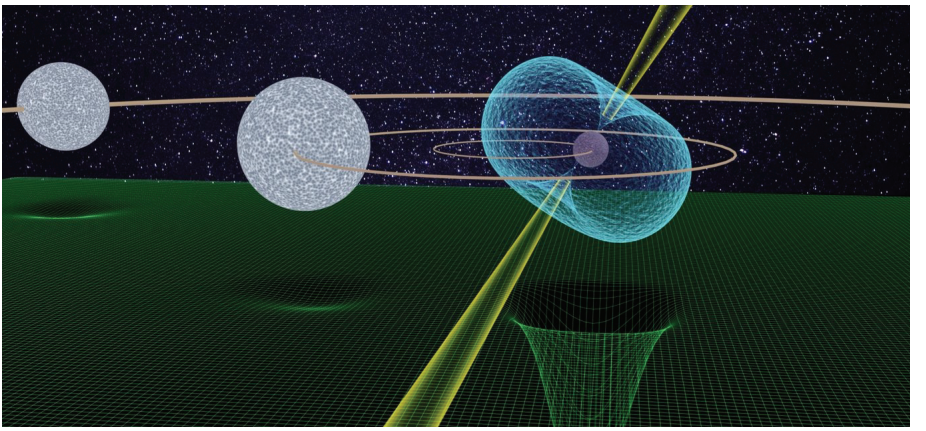


Fig. 10. Artistic impression of the Triple System, where a 2.7-ms pulsar, PSR J0337+1715, and a $0.20 M_\odot$ white dwarf are in an inner 1.63-day orbit, while being orbited themselves by a second $0.41 M_\odot$ white dwarf in 327 days. The figure is not to scale.

The Earth-Moon system allows to test the fall of two bodies with different gravitational self-energy, $\epsilon = E_{\rm grav}/mc^2$, where $E_{\rm grav}$ is the Newtonian gravitational binding energy, in the gravitational field of the Sun. For the Earth $\epsilon_E = -4.6 \times 10^{-10}$ and for the Moon, $\epsilon_M = -0.2 \times 10^{-10}$. The fall can be measured with Lunar Laser Ranging and limits on a deviation are currently at $\Delta < 2 \times 10^{-13}$.[46,47] Comparing this to the gravitational self-energy of the involved bodies, a limit on the violation of the SEP in the weak field of the solar system is $\Delta/\epsilon < 0.04\%$.

Pulsars allow us to test the SEP for strongly-self gravitating bodies, i.e. $\epsilon = -0.13$. The idea was already presented by Damour & Schäfer in 1991 (Ref. 48) who proposed to study the possible variation of the orbits of pulsar-white dwarf systems as they fall in the gravitational potential of the Milky Way. The discovery of a Triple System,[49] a pulsar with two White Dwarf (WD) companions, allows an important variation of this experiment, leading to a much more stringent test (see Figure 10).

The 2.7-ms pulsar, also known as PSR J0377+1715 is in a 1.63-day orbit with a WD of mass $0.2 M_\odot$. This inner binary system is orbited by a second, more massive WD ($0.41 M_\odot$) in a 327-day orbit. As pointed out already by Freire et al. (Ref. 50), in such a case, the pair of bodies with different gravitational self-energy falls in the gravitational potential of the outer WD, providing a much stronger field than that of the Milky Way. Archibald et al. (Ref. 51) presented the result of such an experiment and derived a limit of $\Delta < 2 \times 10^{-6}$. This may appear weaker than the weak-field limit from the solar system, but the SEP should apply for both weak- and strong-field conditions. Hence, when referring to a corresponding $\Delta/\epsilon < 0.002\%$, the importance of this measurement becomes clear.

Even more recently, Voisin et al. (Ref. 52) presented an important improvement on the experiment, using a completely independent data set. The improvements arise from a more uniformly sampled data set but also from an improved theoretical

treatment that incorporates the interesting parameter Δ directly into the analysis of the timing data set. The result is a limit of $\Delta = (+0.5 \pm 1.8) \times 10^{-6}$ at 95% confidence level. While the numerical improvement upon Ref. 51 is 30%, this result is statistics-limited and avoids limitation by systematics as previously encountered.

We complete this section by also referring to recent results and proposals for further experiments to test the UFF also towards dark matter. See Ref. 53 for more details.

4.2. *Alternative theories of gravity*

Despite the successes of GR when being confronted with experimental data, GR may not be our final answer in describing gravity on a macroscopic scale. Given a number of unsolved questions (e.g. dark matter, dark energy, inflation) or its incompatibility with quantum mechanics, is still possible that we encounter an experiment that contradicts GR's predictions. In this case, we would learn important and substantial lessons, and so it is important to explore all regimes presented in Figure 1, probing different aspects of the predictions of GR and alternative theories with different methods. For instance, observations with GW detectors are able to test the highly dynamical strong-field regime and radiative aspects of gravity, but they are not able to test aspects of light-propagation in strong fields. This may be eventually possible with images from super-massive black holes, but the curvature of spacetime is many orders of magnitude smaller than around stellar compact objects. This, on the other hand, and other aspects can be tested with binary pulsars.

The close agreement of the Double Pulsar experiment with the predictions of GR shown in Section 2.1 allows us to place tight constraints on various alternatives to GR. Kramer et al. (Ref. 15) used their recent results to study both "Damour–Esposito-Farése" (DEF) gravity and Bekenstein's tensor-vector-scalar theory (TeVeS).[54] The latter is a MONDian relativistic gravity theory that evades the need for dark matter in galaxies by a modification of GR. Even though this theory has been ruled out already with observations of GWs,[55] Ref. 15 points out that pulsar experiments are still useful to test the aspects of scalar field in the theory.

DEF is a two-parameter mono-scalar-tensor gravity,[56] containing GR as a limit with the two parameters $\alpha_0 = \beta_0 = 0$. As other alternative of GR, DEF violates the SEP, which should result in observable dipolar GWs and a location-dependent gravitational constant. We would measure such phenomena in the form of different functional dependencies of the PK parameters as described in Section 2 (see e.g. Ref. 57). Moreover, in certain regions of the parameter space, DEF gravity shows genuine non-perturbative strong-field effects that are only present in NSs and therefore are not testable in the weak-field regime of the solar system. Hence, one can convert our observational results into constraints for the parameters α_0–β_0. This is done by Ref. 15 as shown in Figure 11. It demonstrates that pulsar tests are complementary to each other as well as to other methods.

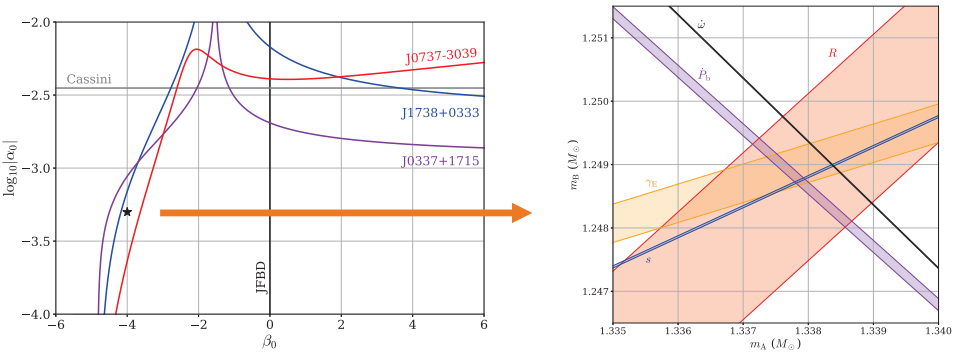


Fig. 11. (Left) Diagram displaying the α_0–β_0 parameter space of DEF gravity. Different experiments provide constraints shown as lines. Parameter space above a line is ruled out. In addition to constraints from the Cassini spacecraft, the figure includes constraints from limits on the existence of dipolar GW emission obtained from PSR J1738+0333, and constraints from the Triple System and the Double Pulsar as described earlier. GR is contained in this parameter space at $(\alpha_0 = 0, \beta_0 = 0)$. The Jordan-Fierz-Brans-Dicke (JFBD) theory is also contained along the vertical $(\beta_0 = 0)$-line. (Right) Every point in the left parameter space corresponds to a specific DEF theory and different realisations of the PK parameters. For a specific choice as an example, indicated by the star in the left figure, one obtains a mass-mass diagram that can be compared to Figure 7. Here, the PK lines fail to intersect in a single point and the specific theory should be rejected. This is already visible by the star's position in the left diagram, left and above the Double Pulsar line. This specific theory, however, would have passed other existing constraints as indicated. See Kramer et al. (2021) for details and references therein.

5. Conclusions

In this contribution, I have tried to summarise some recent key results in probing strong-field gravity using pulsars. In the past three years, the results have not only improved dramatically in precision, but also new effects could be studied. Pulsar continue to play an important role and complement other experiments which have only become possible in the recent years. This includes terrestrial GW detection as well as the imaging of black holes. What is unique about pulsars is the incredible precision that we can obtain in pulsar timing experiments. This is perhaps best demonstrated by the fact that we now have to consider the mass loss of a pulsar spinning down and the interior structure of neutron stars when we one to interpret our results.

Figure 12 attempts to bring the pulsar experiments into context with others. The chosen parameters plotted are a deliberate choice, as they clearly demonstrate that pulsars probe strong-field gravity, as pointed out by many authors in the past (e.g. Ref. 58). It also shows the size of the parameter space to be probed. Future telescopes, like the Square Kilometre Array (SKA), will contribute many more experiments and exciting constraints, as first results with MeerKAT clearly demonstrate.[43] The best is still to come.

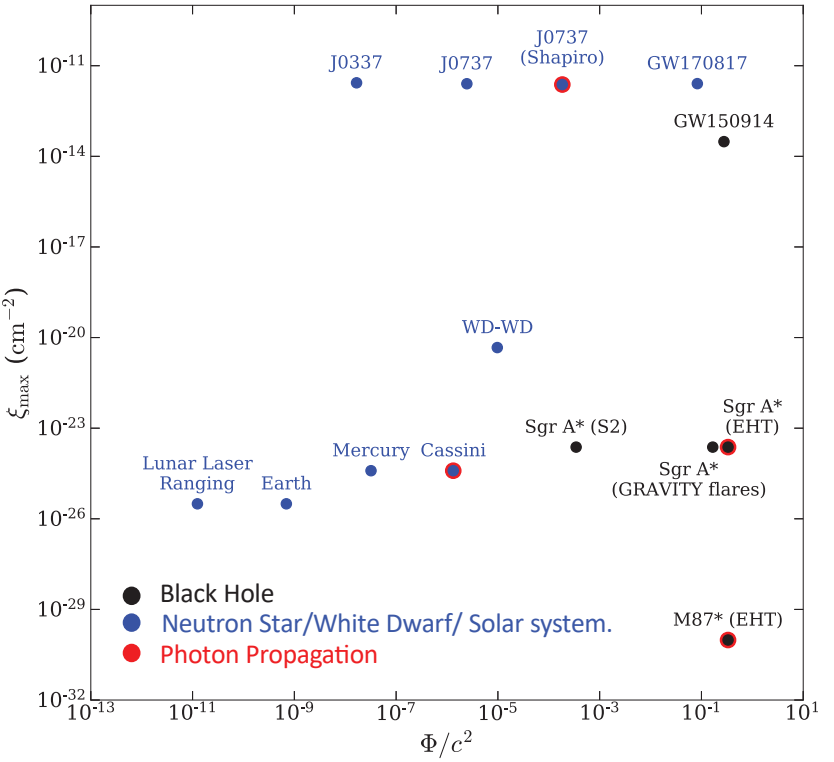


Fig. 12. Parameter space to compare different gravity experiments. While several possible combinations for 2-D diagrams are possible, none is complete in capturing all quantities relevant for gravity tests, and therefore each of them always gives an incomplete comparison. Adopting the plot from Kramer et al. (2021), we show the gravitational potential and the maximum spacetime curvature in a system, distinguishing between weakly and strongly self-gravitating masses amongst material bodies. Experiments that directly probe the coupling between gravitational and electromagnetic fields, i.e. the propagation of photons in a curved spacetime, are highlighted by red circles. We also mark experiments involving black holes as black dots. See Kramer et al. for a discussion of the choice of parameters and further details.

Acknowledgments

I am grateful to the organisers for inviting me to give the presentation. I thank my co-authors in the recent Double Pulsar work and in particular Norbert Wex for providing figures.

References

1. P. Touboul, G. Métris, M. Rodrigues, Y. André, Q. Baghi, et al. *Phys. Rev. Lett.* **119** (December 2017) 231101.
2. Gravity Collaboration,, *A&A* **615** (July 2018) L15.
3. Event Horizon Telescope Collaboration,, *ApJ* **875** (April 2019) L1.

4. M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, et al., R. D., *Science* **314** (2006) 97.
5. S. Chen, R. N. Caballero, Y. J. Guo, A. Chalumeau, K. Liu, et al., *MNRAS* **508** (December 2021) 4970.
6. B. P. Abbott, LIGO Scientific Collaboration and Virgo Collaboration, *Phys. Rev. Lett.* **116** (February 2016) 061102.
7. D. R. Lorimer and M. Kramer, *Handbook of Pulsar Astronomy* (Cambridge University Press, Cambridge, England, 2005).
8. T. Damour and J. H. Taylor, *Phys. Rev. D* **45** (1992) 1840.
9. M. Kramer, *ApJ* **509** (1998) 856.
10. B. B. P. Perera, M. A. McLaughlin, M. Kramer, I. H. Stairs, R. D. Ferdman, et al., *ApJ* **721** (October 2010) 1193.
11. T. Damour and N. Deruelle, *Ann. Inst. H. Poincaré (Physique Théorique)* **43** (1985) 107.
12. T. Damour and N. Deruelle, *Ann. Inst. H. Poincaré (Physique Théorique)* **44** (1986) 263.
13. M. Burgay, N. D'Amico, A. Possenti, R. N. Manchester, A. G. Lyne, et al., *Nature* **426** (December 2003) 531.
14. A. G. Lyne, M. Burgay, M. Kramer, A. Possenti, R. N. Manchester, et al., *Science* **303** (2004) 1153.
15. M. Kramer, I. H. Stairs, R. N. Manchester, N. Wex, A. Deller, W. Coles and et al., *Physical Reviews X* **in press** (December 2021).
16. L. L. Smarr and R. Blandford, *ApJ* **207** (1976) 574.
17. J. M. Weisberg and Y. Huang, *ApJ* **829** (September 2016) 55.
18. LIGO Scientific Collaboration, Virgo Collaboration, et al., *Phys. Rev. Lett.* **119** (October 2017) 161101.
19. J. M. Lattimer, *Universe* **5** (June 2019) 159.
20. G. Boerner, J. Ehlers and E. Rudolph, *A&A* **44** (1975) 417.
21. T. Damour and R. Ruffini, *Academie des Sciences Paris Comptes Rendus Ser. Scie. Math.* **279** (1974) 971.
22. R. A. Hulse and J. H. Taylor, *ApJ* **195** (1975) L51.
23. B. M. Barker and R. F. O'Connell, *ApJ* **199** (1975) L25.
24. V. Radhakrishnan and D. J. Cooke, *Astrophys. Lett.* **3** (1969) 225.
25. J. M. Weisberg, R. W. Romani and J. H. Taylor, *ApJ* **347** (1989) 1030.
26. Z. Arzoumanian, Radio observations of binary pulsars: Clues to binary evolution and tests of general relativity, PhD thesis, Princeton University (1995).
27. I. H. Stairs, S. E. Thorsett, J. H. Taylor and Z. Arzoumanian, Geodetic precession in PSR B1534+12, in *Pulsar Astronomy - 2000 and Beyond, IAU Colloquium 177*, eds. M. Kramer, N. Wex and R. Wielebinski (Astronomical Society of the Pacific, San Francisco, 2000), pp. 121–124.
28. R. N. Manchester, M. Kramer, A. Possenti, A. G. Lyne, M. Burgay, et al., *ApJ* **621** (March 2005) L49.
29. R. D. Ferdman, I. H. Stairs, M. Kramer, R. P. Breton, M. A. McLaughlin, et al., *ApJ* **767** (April 2013) 85.
30. M. Burgay, A. Possenti, R. N. Manchester, M. Kramer, M. A. McLaughlin, et al., *ApJ* **624** (May 2005) L113.
31. A. Noutsos, G. Desvignes, M. Kramer, N. Wex, P. C. C. Freire, et al., *A&A* **643** (November 2020) A143.
32. R. P. Breton, V. M. Kaspi, M. Kramer, M. A. McLaughlin, M. Lyutikov, et al., *Science* **321** (July 2008) 104.

33. M. Lyutikov and C. Thompson, *ApJ* **634** (December 2005) 1223.
34. M. Kramer and I. H. Stairs, *Ann. Rev. Astr. Ap.* **46** (September 2008) 541.
35. M. Kramer and N. Wex, *Class. Quant Grav.* **26** (April 2009) 073001.
36. G. Desvignes, M. Kramer, K. Lee, J. van Leeuwen, I. Stairs, et al, *Science* **365** (September 2019) 1013.
37. S. E. Gralla, A. Lupsasca and A. Philippov, *apj* **851** (December 2017) 137.
38. R. Yuen and D. B. Melrose, *PASA* **31** (October 2014) e039.
39. Y. Gupta and R. T. Gangadhara, *ApJ* **584** (2003) 418.
40. S. Johnston and M. Kramer, *MNRAS* **490** (December 2019) 4565.
41. M. Kramer, *International Journal of Modern Physics D* **23** (December 2014) 1430004.
42. A. D. Cameron, D. J. Champion, M. Kramer, M. Bailes, E. D. Barr, et al., *MNRAS* **475** (March 2018) L57.
43. M. Kramer, I. H. Stairs, V. Venkatraman Krishnan, P. C. C. Freire, F. Abbate, et al., *MNRAS* **504** (June 2021) 2094.
44. L. Shao and N. Wex, *Science China Physics, Mechanics, and Astronomy* **59** (September 2016) 699501.
45. N. Wex and M. Kramer, *Universe* **6** (September 2020) 156.
46. J. G. Williams, S. G. Turyshev and D. H. Boggs, *International Journal of Modern Physics D* **18** (January 2009) 1129.
47. F. Hofmann and J. Müller, *Classical and Quantum Gravity* **35** (February 2018) 035015.
48. T. Damour and G. Schäfer, *Phys. Rev. Lett.* **66** (1991) 2549.
49. S. M. Ransom, I. H. Stairs, A. M. Archibald, J. W. T. Hessels, D. L. Kaplan, et al., *Nature* **505** (January 2014) 520.
50. P. C. C. Freire, M. Kramer and N. Wex, *Classical and Quantum Gravity* **29** (September 2012) 184007.
51. A. M. Archibald, N. V. Gusinskaia, J. W. T. Hessels, A. T. Deller, D. L. Kaplan, et al., *Nature* **559** (2018) 73.
52. G. Voisin, I. Cognard, P. C. C. Freire, N. Wex, L. Guillemot, G. Desvignes, M. Kramer and G. Theureau, *A&A* **638** (June 2020) A24.
53. L. Shao, N. Wex and M. Kramer, *Phys. Rev. Lett.* **120** (June 2018) 241104.
54. J. D. Bekenstein, *Phys. Rev. D* **70** (October 2004) 083509.
55. S. Boran, S. Desai, E. O. Kahya and R. P. Woodard, *Physical Reviews D* **97** (February 2018) 041501.
56. T. Damour and G. Esposito-Farese, *Phys. Rev. Lett.* **70** (1993) 2220.
57. T. Damour and G. Esposito-Farese, *Phys. Rev. D* **54** (1996) 1474.
58. C. M. Will, *Living Reviews in Relativity*, **17** (2014), 4.

Dragging of inertial frames by matter and waves

Jiří Bičák
Tomáš Ledvinka
Institute of Theoretical Physics, Charles University
V Holešovičkách 2, Prague, Czech Republic

We review and analyze four specific general-relativistic problems in which gravitomagnetism plays the important role: the dragging of magnetic fields around rotating black holes, dragging inside a collapsing slowly rotating spherical shell of dust, compared with the dragging by rotating gravitational waves. We demonstrate how the quantum detection of inertial frame dragging can be accomplished by using the Unruh-DeWitt detectors. Finally, we shall briefly show how "instantaneous Machian gauges" can be useful in the cosmological perturbation theory.

Keywords: Einstein equations, Gravitomagnetism, Cosmology

PACS numbers:04.20.-q,98.80.-k

1. Introduction

The relativistic effect of the dragging of inertial frames is associated with the profound criticism of the Newtonian concepts of absolute space and time by Mach. It was this criticism which appears to be one of the most influential for Einstein in the creation of general relativity. Ernst Mach (1838-1916) was born very close to Brno (like Kurt Goedel), today Czech Republic, formerly Austria-Hungary. He was professor at the Charles-Ferdinand University in Prague for 28 years and his influence on Prague physics indirectly also led to the stay of Albert Einstein in Prague.

Let us illustrate Mach's ideas by just one thought from his most influential work:

> [The] investigator ... must feel the need of ... knowledge of the immediate connections, say, of the masses of the universe. There will hover before him as an ideal insight into the principles of the whole matter, from which accelerated and inertial motions will result in the same way *(Science of mechanics)*.

Keeping the Mach tradition, on the 150th anniversary of the birth of Ernst Mach, the international conference was organized in September 1988 at the Charles University in Prague. The conference papers are published in the book [1], including several contributions by some leading scientists, philosophers and historians of science. The meeting was inspiring for Julian Barbour and Herber Pfister who, during the meeting, decided to organize the meeting on just Mach's principle. A comprehensive volume [2], based on the conference at Tübingen in July 1993, includes not

only many contributions by leading experts but also detailed texts recording many discussions.

Prague and Brno historically became attractive places for a number of influential scientists. Here we naturally recall the stay of Albert Einstein at the Karls-Ferdinand University from April 1911 until July 1912. His invitation to Prague was strongly supported by Mach's Prague pupils. Einstein wrote several pioneering papers showing the route to the final version of General Relativity, in particular in his answer to Max Abraham how a future theory of gravity should look like and, in the best known paper from Prague, he forecasted the light bending (we refer, e.g., to [3] for details). From the point of view of dragging, however, it is most interesting that it was in Prague where this phenomenon was first discussed, albeit based on Einstein's Prague preliminary version of general relativity. In his work [4] he considers a shell of matter and its influence on a mass point placed in its center as the shell starts to accelerate. In his words: *This suggests that the entire inertia of a mass point is an effect of the presence of all other masses, which is based on a kind of interaction with the latter. (This is exactly the same point of view that E. Mach advanced in his astute investigations on this subject.)*

Experiments

A nice experiment to measure rotational dragging was suggested by Braginsky, Polnarev and Thorne in 1984 [5]. The plane of a Foucault pendulum at the South Pole will be fixed with respect to "fixed stars" around which on average will not produce the dragging of the pendulum into the rotation. However, slowly rotating "very close" Earth does produce the effect of $\omega_{\rm drag} = 2J/R^3$, where J and R are angular momentum and radius of the Earth—see Fig. 1 left.

Most sophisticated experiment to confirm the dragging of inertial frames by the rotating Earth is, of course, Gravity Probe B.

The idea of placing a gyroscope on a free orbit around the Earth was conceived independently by L. J. Schiff and G. E. Pugh. In fact, the gyros (the "most spherical balls" produced by man) were four. There was also a telescope in the satellite with the gyros which was pointing towards the Guide Star—see Fig. 1. The launch occurred on April 20, 2004, and lasted 16 months. The first results appeared in April 2007 but the complete analysis was finished only in 2015 (see [6]). The measured frame dragging effect, -37.2 ± 7.2 mas y^{-1}, confirmed the general-relativistic prediction -39.2 ± 0.2 mas y^{-1}. The relatively large error was caused primarily by random patches of electric charge on rotors (gyros) and their housing.

A very nice experiment demonstrating the dragging effects on the nodal rates of 3 laser-ranged satellites using the Earth gravity field model produced by space mission GRACE was performed by the group of I. Ciufolini (see his plenary talk at this conference, contribution by D. Lucchesi in the Section PT5; see also the book on Gravitomagnetism by I. Ciufolini and J. A. Wheeler [7]).

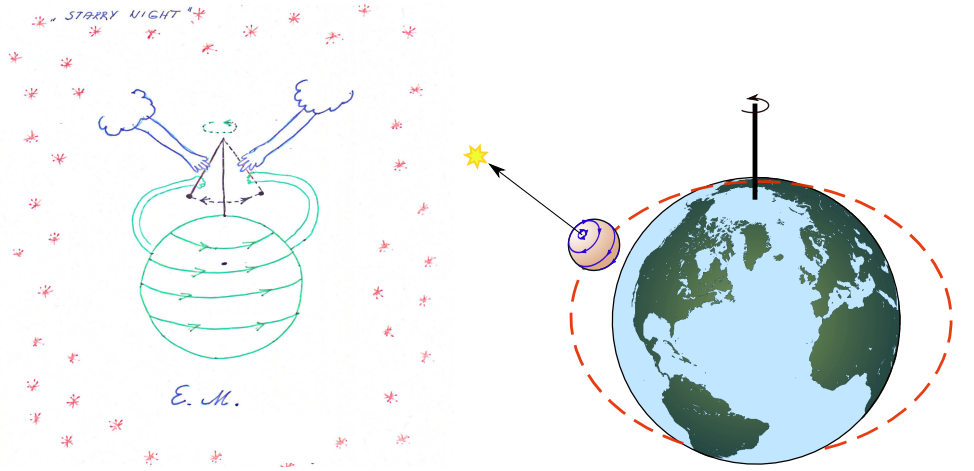


Fig. 1. *Left:* Dragging of the pendulum plane observed at the pole seen as a Machian competition between masses of distant stars and Earth. *Right:* A gyroscope on the polar orbit. Its axis changes direction both due to the geodesic precision ($\approx$ 6600mas y^{-1}) and due to frame dragging induced by Earth rotation ($\approx$ 39mas y^{-1}). This change with respect to a distant "guide" star was measured by the Gravity Probe B space experiment.

2. Magnetic fields, Meissner effect and dragging

Consider first the magnetic test field B_0 which is uniform at infinity and aligned with the hole rotation axis. Solution of Maxwell's equations on the background geometry of a rotating (Kerr) black hole with boundary condition of uniformity at infinity and finiteness at the horizon yields the field components; from these the lines of force are defined as lines tangent to the Lorentz force experienced by test magnetic/electric charges at rest with respect to locally non-rotating frames (preferred by the Kerr background field). The field lines are plotted in Fig. 2 for $a = 0.5M$ and in the extreme case when $a = M$. Notice that only weak expulsion occurs in the former case. There is a simple analytic formula for the flux across the hemisphere of the horizon, see [8–10]:

$$\Phi = B_0 \pi r_+^2 \left(1 - \frac{a^4}{r_+^4}\right), \tag{1}$$

where $r_+ = M + (M^2 - a^2)^{1/2}$. As a consequence of the coupling of magnetic field to frame-dragging effects of the Kerr geometry the electric field of a quadrupolar nature arises. Its field lines are shown in Fig. 3. Again the field lines are expelled: while even with $a = 0.95M$ it is still not very distinct, the expulsion becomes complete in the extreme case. One can demonstrate that total flux expansion takes place for all axisymmetric stationary fields around a rotating black hole [9, 10]. In Fig. 4 the field lines of a current loop in the equatorial plane are shown. The Meissner-type

effect arises also for charged (Reissner-Nordström) black holes as shown in Fig. 4 right.

Although extremely charged black holes ($e^2 = M^2$) are probably not important in astrophysics they may be significant in fundamental physics (as, for example, very special supersymmetric BPS states mass of which does not get any quantum corrections). In the charged case the electromagnetic perturbations are in general coupled to gravitational perturbations, the resulting formalism is involved. Nevertheless, one may construct explicit solutions, at least in stationary cases. From these the magnetic field lines follow as in the Kerr case. The magnetic field lines of a dipole located far away from the hole look like in a flat space, however, when the dipole is close to the horizon, the expulsion in the extreme case is evident (Fig. 4 right). Due to the coupling of perturbations closed field lines appear without any electric current inside; see [11] for details. There exist exact models (exact solutions of the Einstein-Maxwell equations) representing in general rotating, charged black holes immersed in an axisymmetric magnetic field. The expulsion takes place also within this exact framework - see [12–14]. Recently, the Meissner effect was also demonstrated for extremal black-hole solutions in higher dimensions in string theory and Kaluza-Klein theory. The question of the flux expulsion from the horizons of extreme black holes in more general frameworks is not yet understood properly. The authors of [15] "believe this to be a generic phenomenon for black holes in theories with more complicated field content, although a precise specification of the dynamical situations where this effect is present seems to be out of reach."

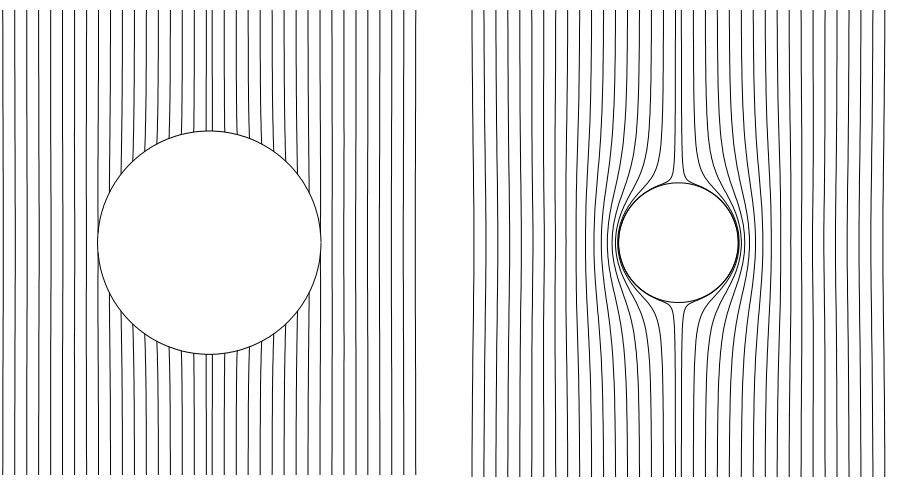


Fig. 2. Field lines of the test magnetic field uniform at infinity and aligned with the hole rotation axis. Two cases with a $= 0.5M$ (left) and $a = M$ (right) are shown.

The flux expulsion does not occur when the configuration is not axially symmetric. The electromagnetic field occurring when a Kerr black hole is placed in an

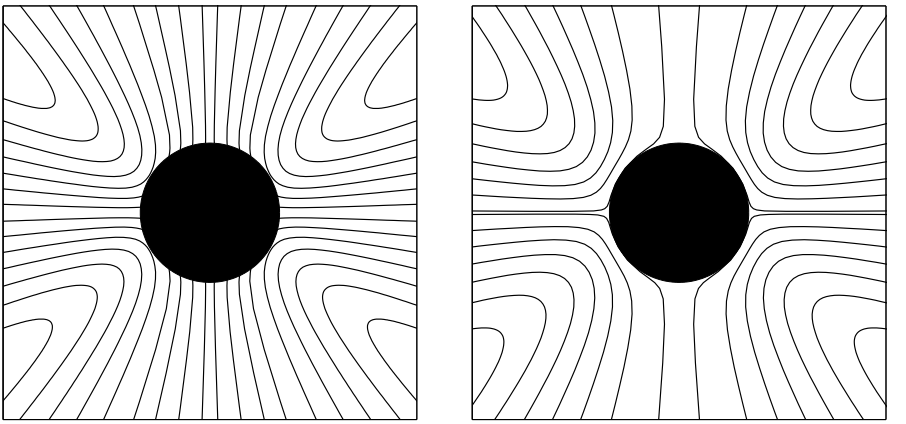


Fig. 3. Field lines of the electric field induced by the "rotating geometry" of Kerr black hole in asymptotically uniform test magnetic field; $a = 0.95M$ (left), and $a = M$ (right).

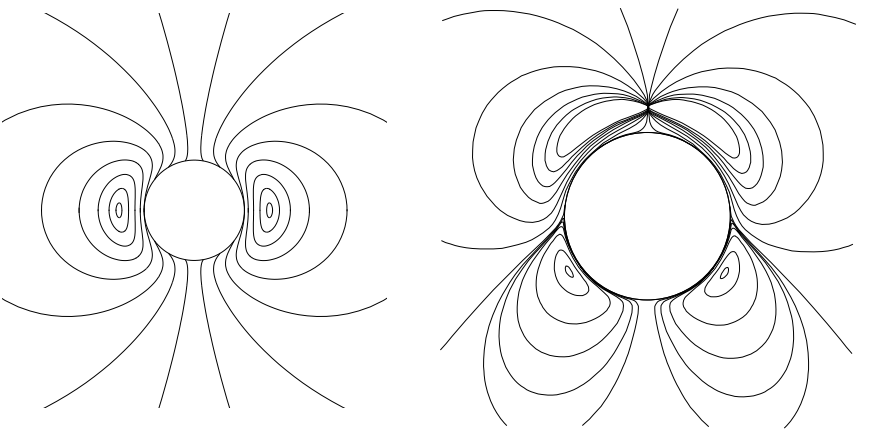


Fig. 4. *Left:* Field lines of the test magnetic field of a current loop in the equatorial plane of the Kerr spacetime with $a = 0.995M$. *Right:* Field lines of the test magnetic field of a magnetic dipole placed near the extreme Rissner-Nordström black hole.

originally uniform magnetic field without assuming the alignment of the direction of the magnetic field and the axis of symmetry of the black hole was first given in [16] (see also [17], [9]).

The properties of these "oblique" fields and their possible astrophysical relevance were already studied in the contribution [12] to the 5th Marcel Grossmann Meeting in Perth. They were then much developed in a number of important papers by V. Karas and his group appearing until today. Here we mention just few results and refer to the paper by V. Karas given in the Session PT5 of this MG16 meeting. One of the effects of the rotation on the fields which are asymptotically uniform and

perpendicular to the rotation axis is the dragging of field lines by rotation and, as a consequence, the appearance of critical points where the field vanishes as seen in Fig. 5.

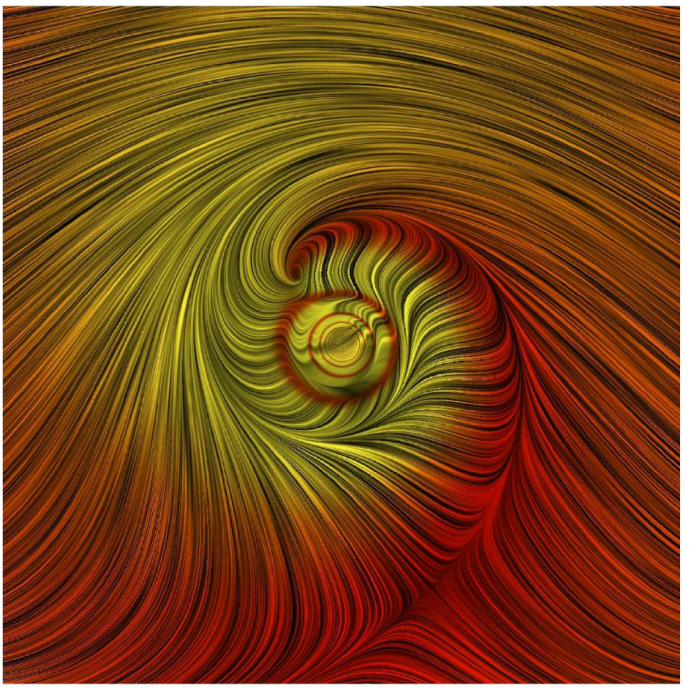


Fig. 5. Field lines in the equatorial plane of the Kerr black hole with color indicating the intensity of the field. Field lines which are asymptotically uniform and perpendicular to the rotation axis are dragged by rotation in vacuum (no conductive medium around). The horizon is in the center as a point. The critical point appears where the field vanishes (approx. at 11 hours). The figure is taken from [18].

In the most recent work, Karas *et al* realized that due to the presence of the plasma in the accretion flows and differential rotation even weak electromagnetic fields are crucial. Although magnetic fields within the accretion flow are turbulent in almost empty funnels around the rotation axis they can be organized on large scales and it is from here where they can accelerate the charged particles and produce collimated jets. Most recently, Karas and Kopáček conclude that inclined field (its oblique component) leads to more efficient acceleration and larger final Lorentz factors of escaping particles; see [19] and number of references therein. For an leading expert view on formation of jets and black-hole shadow in case of M87, see the contribution of R. Blandford to the Session PT5 of this MG16 meeting.

3. Dragging by a slowly rotating, collapsing spherical shell

A spherical shell in slow rotation and collapse (see Fig. 6) produces slightly perturbed Schwarzschild spacetime outside with the metric

$$ds^2 \doteq \left(1-\frac{2M}{r}\right)dt^2 - \left(1-\frac{2M}{r}\right)^{-1} dr^2 - r^2\,d\theta^2 - r^2\,(d\phi-\omega dt)^2\,. \qquad (2)$$

Here ω is the frame dragging potential given by $\omega = 2J/r^3$, J is fixed (small) total angular momentum of the shell. At the shell's surface $r = r_s$ is decreasing as shell collapses and $\omega_s = 2J/r_s^3$ is increasing. Notice (Fig. 6) that $\Omega = d\phi_s/dt$ is the angular velocity of the shell, $r_s^2\Omega^2$ is neglected.

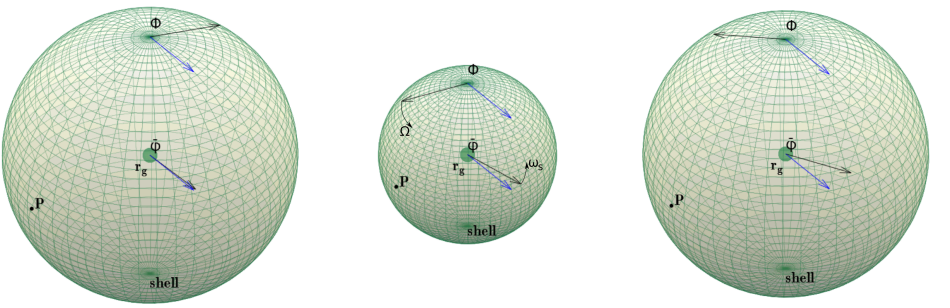


Fig. 6. Dragging of a gyroscope inside a slowly rotating, collapsing and rebouncing thin shell. Angular velocities of the vectors indicating the shell rotation Ω and the central inertial frame rotation ω_s are shown in the middle panel. See [21] for details. (Stills from the animation `https://utf.mff.cuni.cz/~ledvinka/psi/a1.mp4` by W. Barker.)

The spacetime inside the shell is flat in this approximation. Its metric $ds^2 = dt^2 - dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\bar{\phi}^2$ can be joined across the shell to the metric outside. Because $d\bar{\phi} = d\phi - \omega_s\,dt$, the local *inertial* frames (LIFs) inside ($\bar{\phi}$ = const.) all rotate rigidly with the same angular velocity w.r.t. the observers at rest relative to infinity ("static observers" with ϕ = const). Thus $d\bar{\phi}/dt = 0$ implies the time-dependent angular velocity $d\phi/dt = \omega_s(t)$ of the rigid rotation.

As measured *in LIF's* own proper time the rate of rotation is

$$\frac{d\phi}{d\bar{t}} = \bar{\omega}_s = \omega_s \left.\frac{dt}{d\bar{t}}\right|_s . \qquad (3)$$

Static observers inside experience *Euler acceleration* (Coriolis $\sim \omega_s^2$, centrifugal $\sim \omega_s^2$) and the congruence of their *worldlines twists*. Rate of rotation $\bar{\Omega}_\tau$ of the shell itself measured in its proper comoving time τ is

$$\bar{\Omega}_\tau = \frac{3r_s}{4m_s}\omega_s = \frac{3J}{2m_s r_s^2}. \qquad (4)$$

Many details about this system can be found in Refs. [20–22].

4. Quantum detection of inertial frames dragging

Recently, we studied quantum Unruh-DeWitt detectors [23] and their suitability, at least in principle, for the detection of the dragging of inertial frames [24] and for the detection of a conicity of space [25]. We have shown, for the first time as far as we know, that the dragging of inertial frames (as well as conicity) can be observed by a quantum detector. We studied the response function of UdW detector placed in a slowly rotating shell which has flat spacetime inside and slowly rotating Kerr metric outside, as discussed in the previous Section. (Here we assume the shell to be stationary, not collapsing.)

The detector is a two-state system with energy gap Ω and the field-interaction Hamiltonian $\hat{H}(\tau) = \lambda\chi(\tau)\hat{\mu}(\tau) \otimes \hat{\Psi}(x(\tau))$, where $\chi(\tau)$ is switching function of the detector (ensuring that the interaction duration is $\Delta\tau = \pi/k$), $x(\tau)$ its worldline, $\hat{\mu}(\tau)$ its monopole momentum operator and λ is the coupling constant. We assume the detector-field system is in initial state $|0\rangle_D |0\rangle_\Phi$. Then the transition probability P to $|1\rangle_D$ is

$$P = \lambda^2 \mathcal{F} + O(\lambda^4). \tag{5}$$

The response function $\mathcal{F}$ of the detector turns out to be

$$\mathcal{F} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(\tau_1)\chi(\tau_2) e^{-i\Omega(\tau_2-\tau_1)} W(x(\tau_1), x(\tau_2))\, d\tau_1\, d\tau_2, \tag{6}$$

where the Wightman function of the field is

$$W(x(\tau_1), x(\tau_2)) = {}_\Phi\langle 0| \hat{\phi}(x(\tau_2))\hat{\phi}(x(\tau_1)) |0\rangle_\Phi . \tag{7}$$

We show that the response function picks up the presence of rotation even though the spacetime inside the shell is flat and the detector is locally inertial. The detector can distinguish between the static situation when the shell is non-rotating and the stationary case when the shell rotates and the dragging of inertial frames, i.e. gravitomagnetic effects, arise. Moreover, it can do so when the detector is switched on for a finite time interval within which a light signal cannot travel to the shell and back to convey the presence of rotation.

The summary of the results for quantum detection of the dragging of inertial frames is taken from the publication [24]. (See also the contribution of W. Cong in the Session PT5 at MG 16.)

5. Dragging effects of gravitational waves

Rotating gravitational waves can also become a source of the dragging. The situation when the central frame is surrounded by rotating gravitational waves was for the first time modeled assuming the translational symmetry along z-axis in [26, 27]. Although this assumption implies unbounded energy of the gravitational waves and the spacetime is not asymptotically flat, the problem can be treated analytically as the master equation for the single function describing the gravitational wave has

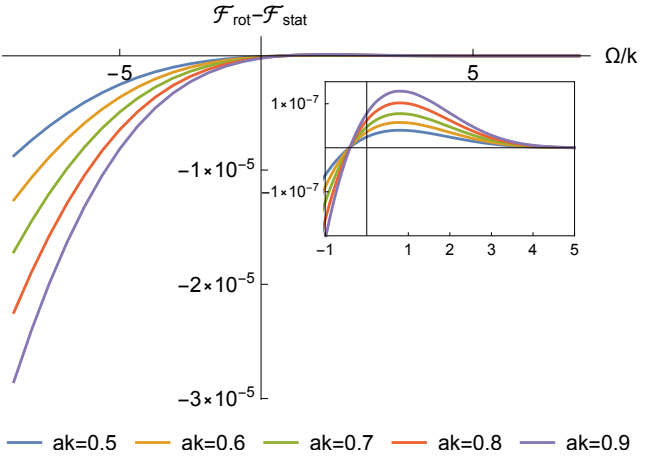


Fig. 7. Comparison with the detector in a static shell. Detector response function inside a slowly rotating shell is plotted for several values of the shell angular momentum $J = Ma$ appearing in the dimensionless parameter ak, where k is the frequency of detector switching function $\chi(\tau)$. The difference $\mathcal{F}_{rot} - \mathcal{F}_{stat}$ is plotted as a function of the energy gap Ω of the detector. Shell mass M and radius R radius satisfy $Mk = 1$, $Rk = 3$, detector is placed at $r_d = 0.5/k$ from the center.

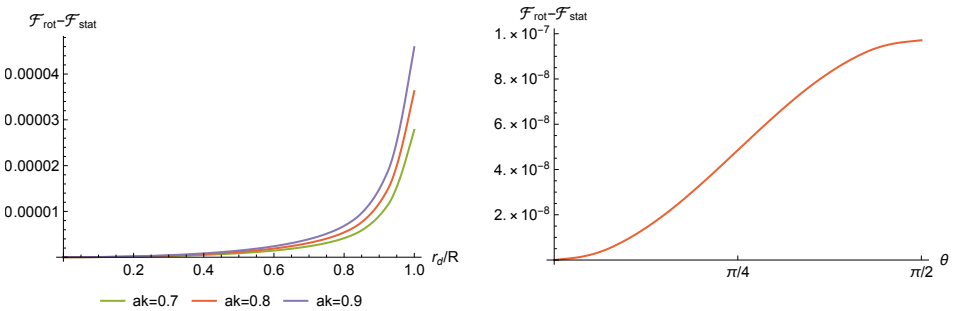


Fig. 8. Detected difference between rotating and static shell depends of the distance of the detector from the center (left) and on the spherical angular coordinate θ of the detector. Remaining parameters are described in Fig. 7.

the form of a flat-space wave equation $\Box\psi(t, \rho, \varphi) = 0$. Given a particular solution to this equation, other metric functions appearing in the line element

$$ds^2 = e^{2\gamma-2\psi}(dt^2 - d\rho^2) - W^2e^{-2\psi}(d\varphi + \omega\, dt)^2 - e^{2\psi}dz^2 \tag{8}$$

can be determined from the Einstein equations. In particular, averaging of their $t-\varphi$ component identifies $\langle\psi_{,t}\psi_{,\phi}\rangle$ as a source of the dragging of the inertial frames on the axis and for ψ in the form of a cylindrical shell the central frame rotation is then found in a closed form. The analogy with angular momentum transport in spiral galaxies is discussed in [27].

A similar problem permitting asymptotic flatness was then studied in [21, 28]. The gravitational waves are assumed to form a spherical shell described again by a single scalar function $\psi(t, \mathbf{x})$ satisfying flat-space wave equations $\Box\psi = 0$ which this time appears only as the first order approximation of the full Einstein equations. The spacetime metric (in which we now use signature $-+++$)

$$g_{\mu\nu} = \eta_{\mu\nu} + h^{(1)}_{\mu\nu} + h^{(2)}_{\mu\nu} + ... \tag{9}$$

is decomposed into a flat Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, r^2, r^2 \sin^2\theta)$ in spherical coordinates t, r, θ, φ, and the first and second-order perturbations $h^{(1)}_{\mu\nu}$ and $h^{(2)}_{\mu\nu}$. Then the first-order metric perturbations due to linearized gravitational waves appear the source of the second order perturbations

$$G^{(1)}_{\mu\nu}[h^{(2)}] = -G^{(2)}_{\mu\nu}[h^{(1)}, h^{(1)}], \tag{10}$$

where $G^{(2)}_{\mu\nu}[h^{(1)}, h^{(1)}]$ contains terms of the Einstein tensor $G_{\mu\nu}$ quadratic in the first-order perturbations.

Assuming the Regge-Wheeler gauge the function ψ then directly determines $h^{(1)}_{t\theta}, h^{(1)}_{t\varphi}, h^{(1)}_{r\theta}$, and $h^{(1)}_{r\varphi}$, with remaining first-order perturbations vanishing, and the effects of the linearized gravitational waves are then determined by the analysis of the second-order terms. To make space approximately flat for the central observer

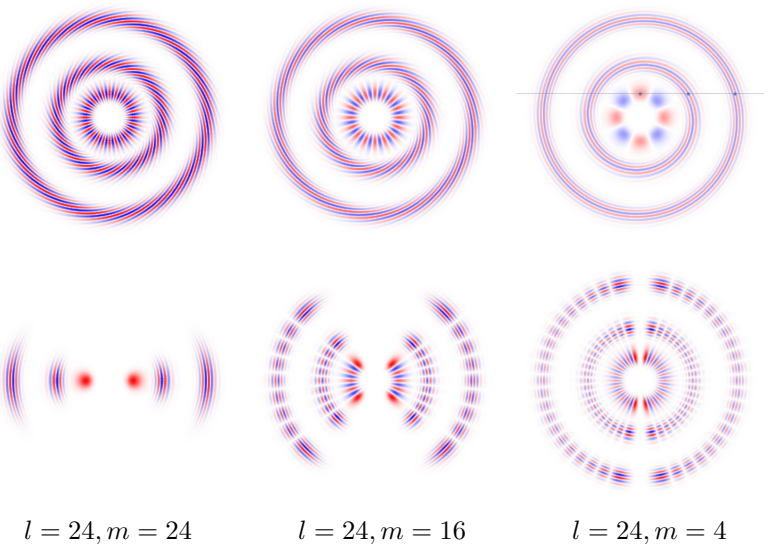


Fig. 9. Snapshots of the function ψ in the equatorial plane $\theta = \pi/2$ (top) and in the meridional plane $\varphi = 0, \pi$ (bottom) at three distinct times $t = 0, 2a, 4a$. The well-known behavior of spherical harmonics $Y_{lm} \sim \sin^{|m|}\theta$ means that for higher m the first order perturbations vanish not only near the center where we study the frame dragging but also along the z axis. The top right plot also shows the position of a null particle with $\mathbf{r} = a\hat{\mathbf{y}} + t\hat{\mathbf{x}}$ at given times to illustrate the localization of the wave at radii $r \approx \sqrt{a^2 + t^2}$. (We denote Cartesian unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}$, red/blue color indicates positive/negative ψ.)

and his inertial frame, $h^{(1)}_{\mu\nu}$ is assumed to vanish near the origin. Assuming a particular gauge, the quantity determining the central frame dragging can be determined from a quantity satisfying an elliptic equation, in a way similar to other situations. The central-frame rotation appears in the perturbation approach as a $l = 1, m = 0$ component in the expansion of $h^{(2)}_{\mu\nu}$ into mutually orthogonal spherical tensor harmonics. Then a projection of Eq. (10) into the relevant $m = 0$ tensor harmonic function resembles an averaging and yields

$$\frac{1}{2}\left[h_0^{(2)\prime\prime} - \frac{l(l+1)}{r^2}h_0^{(2)} - \dot{h}_1^{(2)\prime} - \frac{2}{r}\dot{h}_1^{(2)}\right] = \frac{1}{l(l+1)}\int_0^{2\pi}\int_0^{\pi} G^{(2)}_{t\varphi}[h^{(1)}, h^{(1)}]\partial_\theta Y_{l0}\; d\theta\; d\varphi,$$

where $h_0^{(2)}$ determines $h^{(2)}_{t\theta}, h^{(2)}_{t\varphi}$, and $h_1^{(2)}$ determines $h^{(2)}_{r\theta}, h^{(2)}_{r\varphi}$ components of the second-order metric perturbations. Dots and primes denote the time and radial derivatives. The rotation of the central frame $d\tilde{\varphi} = d\varphi - \omega_0 dt$ appears as $h^{(2)}_{t\varphi} = -\omega_0\, r^2\, \sin^2\theta$, so it is determined by the behavior of $h_0^{(2)}$ at $r = 0$. Using a global change of coordinate $\varphi \to \varphi + \delta\varphi^{(2)}(t, r)$, we can set

$$h_1^{(2)} = 0, \quad \text{i.e.,} \quad h^{(2)}_{r\varphi} = 0. \tag{11}$$

Near the center we then have the Minkowski metric in spherical coordinates with the dominating perturbation corresponding to the slow rigid rotating of the central frame with angular velocity $\omega_0(t)$. Fixing the gauge condition (11), $h^{(2)}_{1\,l=1,m=0} = 0$, prohibits any radial dependence of an additional coordinate transformation $\varphi \to \varphi + \delta\varphi^{(2)}(t, r)$ and the angle φ in the center and thus also the central frame rotation ω_0 is determined *unambiguously* with respect to spatial infinity. We find

$$\omega_0 = \frac{1}{4\pi}\iiint R^{(2)}_{t\varphi}[h^{(1)}, h^{(1)}]\;\frac{\sin\theta}{r} dr\, d\theta\, d\varphi. \tag{12}$$

To investigate further a particular closed-form solution ψ of the wave equation, it has been chosen in the form of a shell of null radiation converging toward the origin, bouncing at the minimal radius $r \approx a$, and then expanding back to infinity (see Fig. 9). This allowed us to evaluate the integral (12) and find the explicit (though lengthy) formula for ω_0. Assuming $l \gg 1$ it simplifies to

$$\omega_0(t) \doteq \frac{\omega_0^{\max}}{\left(1 + \frac{t^2}{a^2}\right)^{3/2}}. \tag{13}$$

In the same limit, we show in [21] that the frame dragging is determined by the angular momentum of the gravitational wave L_z and that the long exact formula can be approximated by $\omega_0^{\max} \doteq 2L_z/a^3$. The angular momentum of the linearized gravitational waves is defined using the effective stress energy tensor

$$L_z = -\int T^{\text{eff}}_{t\varphi}\; d^3x, \qquad T^{\text{eff}}_{t\varphi} = \frac{1}{8\pi}G^{(2)}_{t\varphi}[h^{(1)}, h^{(1)}]. \tag{14}$$

We can see that (12) and (14) differ by a factor r^3 inside the integral. This explains why the approximate relation (13) holds: because for $l \gg 1$ the function ψ is localized around a thin shell with radius $r(t) \doteq \sqrt{a^2 + t^2}$ (see Fig. 9), the factor r^3 can

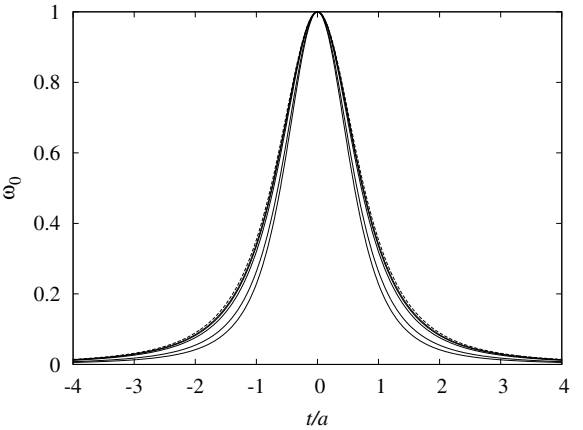


Fig. 10. The dependence of the normalized angular velocity of the central inertial frame $\omega_0(l,1;t)/\omega_0(l,1;t=0)$ on the parameter $l = 2, 3, 10, 20, 30$ (from inside to out). The dependence (13) is shown as a dashed line.

be put in front of the integral. The time dependence of $\omega_0(t)$ on the parameter l is shown in Fig. 10.

In an asymptotically flat spacetime we have two special flat-space worldlines categories — the one of the central observer and that of a cautious observer who slowly retreats to $r \gg a$ so that she never meets significant metric perturbations. The discrepancy between the orientation of the gyroscopes following these worldlines

$$\Delta\varphi_0 = \int_{-\infty}^{\infty} \omega_0(t)\, dt \tag{15}$$

can be seen as an illustration of the dependence of the parallel transport on the chosen worldline. In Fig. 11 we illustrate $\Delta\varphi_0$ as an obvious implication of the spacetime curvature due to the rotating gravitational waves. Thus, although the immediate value of ω_0 involves instantaneous effects, its integral (15) representing the total rotation of the central gyroscope is a well-defined observable quantity. In the approximation $l \gg 1$ we then obtain $\Delta\varphi_0 \doteq 2a\omega_0^{\max} \doteq 4L_z/a^2$. Such a simple relation is not available for dragging by a massive rotating shell, because its dynamics is not as unambiguous as that of gravitational radiation.

6. On the dragging of inertial frames and Mach's principle in cosmology

In our treatment of the dragging of inertial frames in a cosmological context we shall mostly confine ourselves to the linear (cosmological) perturbation theory, rather than to exact models. Our inspiration will be Mach's principle as generally formulated by Hermann Bondi in his classical book [29]: *...all motions, velocities, rotations and accelerations are relative; local inertial frames are determined through*

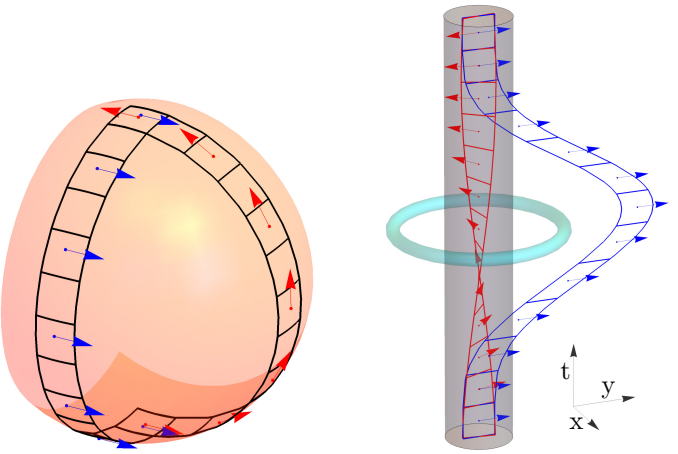


Fig. 11. *Left:* The fundamental dependence of parallel transport on the chosen path is usually demonstrated on a spherical surface naturally embedded in three-dimensional flat Euclidean space. Here we use neighborhoods of two meridians as an example of two approximately flat patches which yield mismatch when vector from the south pole is extended into both patches. *Right:* In our spacetime with rotating gravitational waves we also have two approximately flat patches. The spacetime is asymptotically flat which in the figure is symbolized by the blue "ladder" with arrows indicating "fixed" direction of a gyroscope. Because the gravitational waves do not reach the center, there is also approximately flat region near the center. Its worldtube is depicted as a gray cylinder. The gravitational waves are shown at the moment they are the strongest ($t = 0$) as a blue torus encircling the central observer. The rotation of the central inertial frame (and gyroscopes there) is illustrated by the twist of the red spacetime-coordinate "ladder" and the gyroscope orientation. The mismatch of gyroscope directions at the top demonstrates the meaning of Eq. (15) as the implication of a particular form of spacetime curvature accompanying the rotating gravitational wave.

distributions of energy and momentum in the Universe by some weighted averages of the apparent motions.

We started to realize such a "Machian program" in [30]. We first analysed frame-dragging effects due to slowly, rigidly rotating, but collapsing or expanding spheres in the (inhomogeneous) Lemaître-Tolman-Bondi universes, and we analysed the dragging effects of the vector perturbations of the FLRW universes described in a special gauge such that three (momentum) constraint equations enabled us to determine instantaneously metric perturbations h_{0k} $(k = 1, 2, 3)$ in terms of energy-momentum perturbations δT_{0k} and show how such averages are to be taken. In closed universes a linear combination of six Killing vectors (three rotations plus three quasitranslations) may be added to the h_{0k}. We also obtain the solutions of the three constraint equations when angular momenta corresponding to the three rotations and quasimomenta corresponding to the three quasitranslations of the sources (determined by δT_{0k}) are given. No absolute rotations exist in closed universe, only differences of rotation rates are determinable — in accord with Mach's ideas that 'all motions are relative'. (If the velocities of the bodies, described by perturbations of perfect fluid, are given, the metric perturbations are determined

uniquely.) The last result is related to the fundamental fact that six globally conserved quantities, corresponding to the six Killing vectors in a FLRW universe, must all vanish if considered for the whole closed universe.

It was, among others, an attempt to understand Mach's principle in cosmological perturbation theory, which inspired us to formulate conservation laws even for large perturbations with respect to curved backgrounds [31]. The resulting 'KBL superpotential' (using the designation by Julia and Silva in their profound analysis [32]), was found, after applying certain natural criteria, to be unambigous and most satisfactory in spacetimes with or without a cosmological constant, in any spacetime dimension. It also found applications in various studies of generation of cosmological perturbations (see [33] for references). For the recent generalization to the Horndeski theory, see [34].

In a more recent paper [33] we studied general linear perturbations of the FLRW universes from a 'Machian perspective'. This led us to investigate both rotations and accelerations of local inertial frames in perturbed universes. We first introduced congruences of cosmological observers' worldlines, defined their acceleration, rotation (twist, vorticity), shear and expansion in general, and then considered perturbed FLRW models ($g_{\mu\nu} = g^{\mathrm{FLRW}}_{\mu\nu} + h_{\mu\nu}$). We found that un-accelerated and non-rotating local inertial frames (LIFs) are determined by $h_{00,l}, h_{0l,m}, h_{0l,0}$.

We developed all the perturbed Einstein equations in a general gauge 'ab initio', without assuming harmonic decompositions. Introducing the standard conformal time in FLRW universes, putting tildas over all the perturbation quantities, introducing the traceless part of h^l_k and notation

$$\tilde{h}^l_{Tk} = \tilde{h}^l_k - \tfrac{1}{3}\delta^l_k \tilde{h}^n_n \,, \qquad \mathcal{T}_k = \nabla_l \tilde{h}^l_{Tk}, \qquad \mathcal{K} = \tfrac{3}{2}\dot{a}\tilde{h}_{00} \; + \tfrac{1}{2}a\dot{\tilde{h}}^n_n - \nabla_l \tilde{h}^l_0, \tag{16}$$

where ∇_l is the covariant derivative associated with the spatial FLRW background metric f_{kl}, a is the expansion factor, dot the derivative w.r.t. standard cosmological time t whereas the prime denotes the derivative with respect to conformal time $\eta, adt = d\eta$. Using $\nabla^2 = f^{kl}\nabla_{kl}$, $k = 0, \pm 1$ for the curvature index and $\mathcal{H} = aH$, H being the standard Hubble parameter, we find Einstein's equations for perturbations to obtain the form

$$a^2\kappa\delta\tilde{T}^0_0 = a^2\delta\tilde{G}^0_0 = \tfrac{1}{3}\nabla^2\tilde{h}^n_n + k\tilde{h}^n_n - 2\mathcal{H}\mathcal{K} - \tfrac{1}{2}\nabla_k\mathcal{T}^k, \tag{17}$$

$$a^2\kappa\delta\tilde{T}^0_k = a^2\delta\tilde{G}^0_k = \tfrac{1}{2}\nabla^2\tilde{h}_{k0} + k\tilde{h}_{k0} + \tfrac{1}{6}\nabla_{kl}\tilde{h}^l_0 + \tfrac{2}{3}\nabla_k\mathcal{K} - \tfrac{1}{2}\left(\mathcal{T}_k\right)', \tag{18}$$

$$a^2\kappa\left(\delta\tilde{T}^0_0 - \delta\tilde{T}^n_n\right) = a^2\left(\delta\tilde{G}^0_0 - \delta\tilde{G}^n_n\right) = \nabla^2\tilde{h}_{00} + 3a\left(\frac{1}{a}\mathcal{H}\right)'\tilde{h}_{00} + \frac{2}{a}\left(a\mathcal{K}\right)', \tag{19}$$

and

$$a^2\kappa\left(\delta\tilde{T}^l_k - \tfrac{1}{3}\delta^l_k\delta\tilde{T}^n_n\right) = a^2\delta\tilde{G}^l_{Tk} = \ldots \,. \tag{20}$$

We do not write down fully the last equation since it describes waves and is not important for the determination of LIFs.

To see how the LIFs can be determined by surrounding matter instantaneously on certain time-slices we use some specific gauges which we call the "Machian gauges". We give three examples of such gauges. For example, by putting $\mathcal{T}_k = 0$ and $\mathcal{K} = 0$, the first three equations become (hyper-) elliptic and the quantities determining LIFs can be found instantaneously when the (perturbations of) matter distribution are given. The gauge conditions $\mathcal{T}_k = 0$, fixing spatial coordinates, are associated with the "transverse-traceless" gauges in the linearized gravity and minimal-shear condition in numerical relativity. We assume these conditions to be valid in all three Machian gauges. In the first Machian gauge we choose the time slices to be so that $\mathcal{K} = 0$. This implies the "constant mean curvature slices", and it coincides with Bardeen's uniform-Hubble expansion gauge. In other two Machian gauges, together with the same gauge condition on spatial coordinates, we require "uniform-intrinsic scalar curvature condition" and the "minimal-shear hypersurface condition" (called the Poisson gauge by Bertschinger in 1995). In [33] these gauges are discussed in detail. In particular it is shown that they admit much less residual freedom than the synchronous gauge, frequently used in cosmology.

These Machian gauges have been considered in the group of D. Wiltshire, in particular, in [35], in the review [36] and, most recently, by his students M. Williams [37] and R. Gaur [38] in the context of the Post-Newtonian Cosmology.

We believe that dragging effects and Machian ideas will remain the source of inspiration.

Acknowledgments

We acknowledge the partial support from Grant No. GAČR 21/11268S.

References

1. V. Prosser and J. Folta (eds.), *Ernst Mach and the Development of Physics* (Universitas Carolina Pragensis, Karolinum, Prague, 1991).
2. J. B. Barbour and H. Pfister (eds.), *Mach's Principle: From Newton's Bucket to Quantum Gravity* (Birkhäuser, Boston, 1995).
3. J. Bičák, *Einstein in Prague: Relativity Then and Now*, in *General Relativity, Cosmology and Atrophysics*, Fundamental Theories of Physics 177, eds. J. Bičák and T. Ledvinka (Springer, Cham, 2014), p. 33.
4. A. Einstein, Gibt es eine Gravitationswirkung, die der elecktrodynamischen Induktionswirkung analog ist?, *Vierteljahrsschrift für gerichtliche Medizin und öffentliches Sanitätswesen* **44**, 37–40 (1912).
5. V. B. Braginskii, A. G. Polnarev and K. S. Thorne, Foucault pendulum at the south pole - Proposal for an experiment to detect the Earth's general relativistic gravitomagnetic field, *Phys. Rev. Lett.* **53**, p. 863 (1984).
6. C. W. F. Everitt *et al.*, The Gravity Probe B test of general relativity, *Class. Quantum Grav.* **32**, 224001 (2015).
7. I. Ciufolini and J. A. Wheeler, *Gravitation and Inertia* (Princeton University Press, 1995).
8. A. R. King, J. P. Lasota and W. Kundt, Black holes and magnetic fields, *Phys. Rev. D* **12**, p. 3037 (1975).

9. J. Bičák and V. Janiš, Magnetic fluxes across black holes, *Mon. Not. R. Astron. Soc.* **212**, p. 899 (1985).
10. J. Bičák and T. Ledvinka, Electromagnetic fields around black holes and Meissner effect, *Nuovo Cimento B Serie* **115**, p. 739 (2000).
11. J. Bičák and L. Dvořák, Stationary electromagnetic fields around black holes. III. General solutions and the fields of current loops near the Reissner-Nordström black hole, *Phys. Rev. D* **22**, p. 2933 (1980).
12. J. Bičák and V. Karas, The influence of black holes on uniform magnetic fields, in *Proceedings of the 5th Marcel Grossmann Meeting on General Relativity*, Part B, eds. D. Blair and M. Buckingham (World Scientific, Singapore, 1989), p. 1199.
13. V. Karas and D. Vokrouhlický, On interpretation of the magnetized Kerr-Newman black hole., *Journal of Mathematical Physics* **32**, p. 714 (1991).
14. V. Karas and Z. Budínová, Magnetic Fluxes Across Black Holes in a Strong Magnetic Field Regime, *Physica Scripta* **61**, p. 253 (2000).
15. A. Chamblin, R. Emparan and G. W. Gibbons, Superconducting p-branes and extremal black holes, *Phys. Rev. D* **58**, 084009 (1998).
16. J. Bičák and L. Dvořák, Stationary electromagnetic fields around black holes. II. General solutions and the fields of some special sources near a Kerr black hole., *General Relativity and Gravitation* **7**, p. 959 (1976).
17. M. D. Pollock and W. P. Brinkmann, The Interaction between a Weak Magnetic Field and a Black Hole, *Proceedings of the Royal Society of London Series A* **356**, p. 351 (1977).
18. V. Karas, O. Kopáček and D. Kunneriath, Influence of frame-dragging on magnetic null points near rotating black holes, *Class. Quantum Grav.* **29**, 035010 (2012).
19. V. Karas and O. Kopáček, Near horizon structure of escape zones of electrically charged particles around weakly magnetized rotating black hole: Case of oblique magnetosphere, *Astronomische Nachrichten* **342**, p. 357 (2021).
20. J. Katz, D. Lynden-Bell and J. Bičák, Instantaneous inertial frames but retarded electromagnetism in rotating relativistic collapse, *Class. Quantum Grav.* **15**, p. 3177 (1998).
21. W. Barker, T. Ledvinka, D. Lynden-Bell and J. Bičák, Rotation of inertial frames by angular momentum of matter and waves, *Class. Quantum Grav.* **34**, 205006 (2017).
22. H. Pfister and M. King, *Inertia and Gravitation*, Lecture Notes in Physics 897 (Springer, Cham, 2015).
23. B. S. DeWitt, Quantum gravity: the new synthesis, in *General Relativity: An Einstein centenary survey*, eds. S. W. Hawking and W. Israel (University Press, Cambridge, 1979).
24. W. Cong, J. Bičák, D. Kubizňák and R. B. Mann, Quantum detection of inertial frame dragging, *Phys. Rev. D* **103**, 024027 (2021).
25. W. Cong, J. Bičák, D. Kubizňák and R. B. Mann, Quantum detection of conicity, *Physics Letters B* **820**, 136482 (2021).
26. J. Bičák, J. Katz and D. Lynden-Bell, Gravitational waves and dragging effects, *Class. Quantum Grav.* **25**, 165017 (2008).
27. D. Lynden-Bell, J. Bičák and J. Katz, Inertial frame rotation induced by rotating gravitational waves, *Class. Quantum Grav.* **25**, 165018 (2008).
28. J. Bičák, J. Katz, T. Ledvinka and D. Lynden-Bell, Effects of rotating gravitational waves, *Phys. Rev. D* **85**, 124003 (2012).
29. H. Bondi, *Cosmology* (Cambridge University Press, London, 1960).
30. D. Lynden-Bell, J. Katz and J. Bičák, Mach's principle from the relativistic constraint equations, *Mon. Not. R. Astron. Soc.* **272**, p. 150 (1995). Errata **277**, 1600(E) (1995).

31. J. Katz, J. Bičák and D. Lynden-Bell, Relativistic conservation laws and integral constraints for large cosmological perturbations, *Phys. Rev. D* **55**, p. 5957 (1997).
32. B. Julia and S. Silva, Currents and superpotentials in classical gauge theories: II. Global aspects and the example of affine gravity, *Class. Quantum Grav.* **17**, p. 4733 (2000).
33. J. Bičák, J. Katz and D. Lynden-Bell, Cosmological perturbation theory, instantaneous gauges, and local inertial frames, *Phys. Rev. D* **76**, 063501 (2007).
34. J. Schmidt and J. Bičák, Covariant conserved currents for scalar-tensor Horndeski theory, *Journal of Mathematical Physics* **59**, 042501 (2018).
35. D. L. Wiltshire, Cosmological equivalence principle and the weak-field limit, *Phys. Rev. D* **78**, 084032 (2008).
36. D. L. Wiltshire, What is dust?—Physical foundations of the averaging problem in cosmology, *Class. Quantum Grav.* **28**, 164006 (2011).
37. M. Williams, Post-Newtonian Cosmology, dissertation, University of Canterbury, (New Zealand, 2020).
38. R. Gaur, Perturbation Theory and Averages in Inhomogeneous Cosmologies, Master's thesis, University of Canterbury (2021).

Probes of the progenitors, engines and physics behind stellar collapse

Chris L. Fryer

Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
E-mail: fryer@lanl.gov
https://ccsweb.lanl.gov/astro/index.html

SN are one of the most powerful explosions in the universe and astronomers have invoked the collapse of a stellar core down to a neutron star as a potential power source behind these cosmic blasts. The current paradigm behind core-collapse SN relies on convection in the region just above the newly formed neutron star. This engine was driven and confirmed by observations. We review this observational evidence, and the potential for further observational constraints in this paper.

Keywords: SNe; neutron star; black hole

PACS numbers: 97.60.Bw, 97.60.Jd, 97.60.Lf

1. Core-Collapse SNe (CCSNe)

The SN field was established Baade and Zwicky coined the term "Super-nova" (SN) to explain the most powerful class of novae.[1] Shortly thereafter, Zwicky also proposed what has become one of the standard energy sources behind these cosmic explosions: the collapse of the core of a massive star down to a neutron star.[2] Zwicky argued that this core-collapse SNe (CCSN) mechanism would tap the potential energy released when the core of a massive star collapsed down to a neutron star. The details of how this works has taken nearly 90 years to work out.

Over the past 8-9 decades, scientists have continued to refine the basic model behind CCSN. Stellar models focused on identifying which stars would produce cores that would collapse down to SNe. Stars are powered by nuclear fusion in their cores and the most massive stars can undergo a series of nuclear burning phases, where the ashes of the previous phase are fused in the next burning phase, until an iron core is produced. Iron is maximally compact, its fusion does not generate energy and the contraction of the iron core causes electron capture and the collapse of the core. This collapse only stops when the core forms a proto-neutron star, realizing the engine proposed by Zwicky.

1-dimensional simulations of stellar evolution have allowed astronomers to piece together the detailed evolution of these successive burning stages. This evolution is capture in a diagram developed by Kippenhahn.[3] Figure 1 shows the Kippenhahn diagram describing the evolution of a $22\,M_\odot$ star.[4] These Kippenhahn diagrams contain a wealth of information. The top curve denotes the mass of the star as a

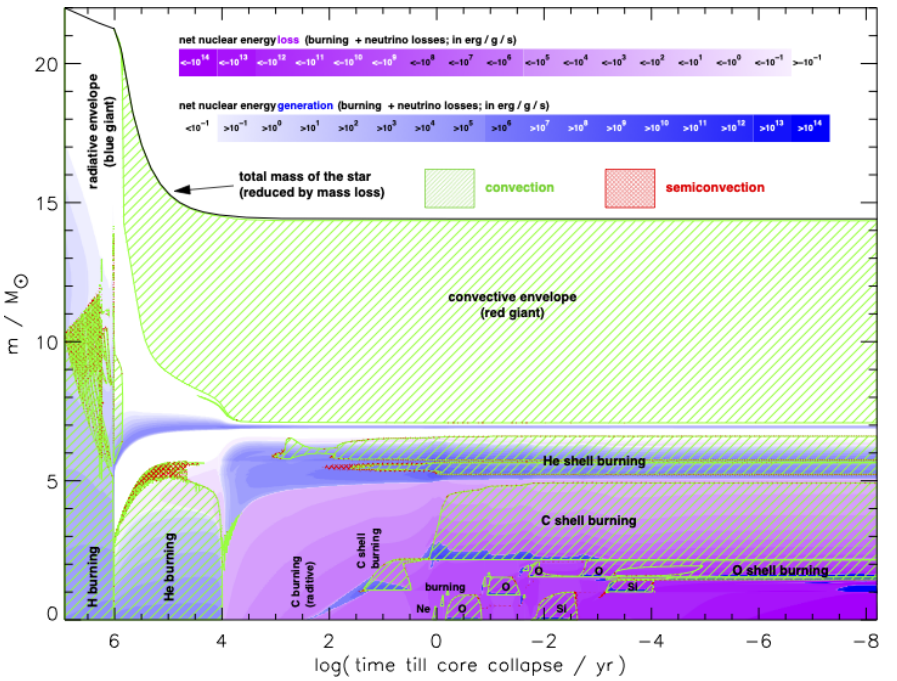


Fig. 1. Enclosed mass vs. time before collapse of a 22 $M_\odot$ star. This diagram encapsulates the entire life of the star. It color codes energy deposition from fusion and energy losses from neutrinos. Hash marks denote convection and the top line denotes mass loss from winds.

function of time. The lower-left corner shows the hydrogen burning core which lasts until a million years prior to collapse. The next phase is the helium core burning phase that lasts until 10,000 y prior to collapse. At the end of helium core burning, the core collapses and helium begins to fuse in a shell layer above this core. In the last 100 y, the star goes through a series of carbon, oxygen and silicon core and shell burning, building up an iron core. The timescales and the evolution of the core can all be learned from this diagram.

This iron core is supported by a combination of thermal and electron-degeneracy pressure. As the mass of this iron core continues to grow, the core contracts. Ultimately, electrons in the core capture onto protons, removing the support pressure and causing a runaway collapse. This collapse proceeds until the core reaches nuclear densities and neutron degeneracy pressure and nuclear forces halt the collapse. The collapse of the iron core to this proto-neutron star releases nearly 100 times the energy needed to drive the observed explosion. When the collapse halts, it can produce a strong bounce and preliminary 1-dimensional calculations of this collapse phase produced weak explosions.[5] These explosions did not match the observed energies and were missing critical pieces of physics, most importantly, the neutrino

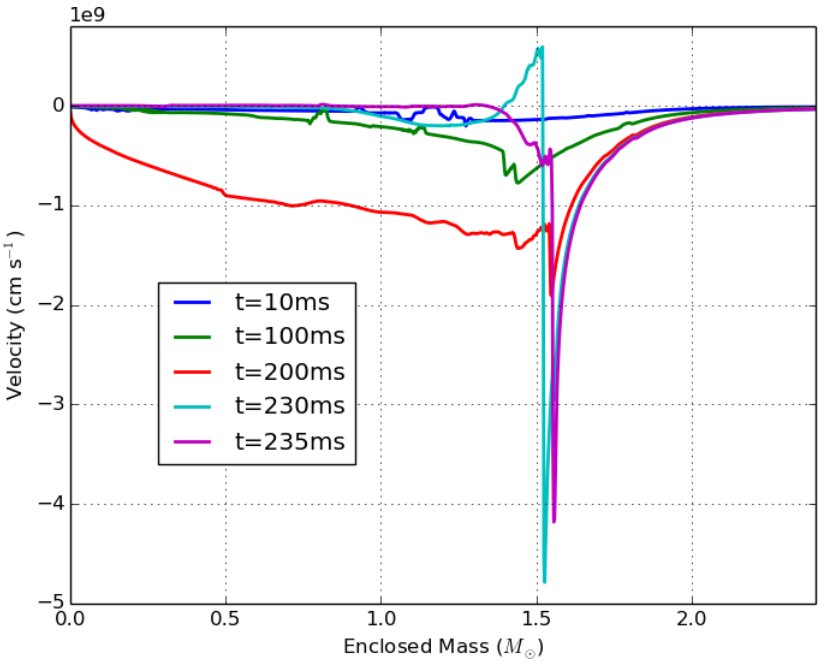


Fig. 2. Velocity vs. enclosed mass of the collapse, bounce and stall of a bounce shock for a 15 $M_\odot$ star at a series of times in the collapse. As the core begins to collapse, the infall accellerates until bounce. Shortly after the bounce, the bounce shock stalls. The different lines show this evolution.

physics. In the bounce, most of the potential energy converted into neutrinos and, as long as the neutrinos are trapped, the bounce shock propagates outward. But when the neutrinos in the outward-moving shock can escape, much of the bounce energy is lost, causing this shock to stall. Figure 2 shows the velocity evolution of a sample core collapse, bounce and bounce shock stall.

Although scientists realized that tapping the neutrinos leaking out of the core would be one way to revive the explosion, the diffusion of neutrinos was too slow to revive the shock. The first mention of convection in CCSN focused on using this convection to transport the neutrinos to the edge of the proto-neutron star, allowing the neutrinos to more quickly deposit their energy into the stalled shock.[6,7] Although this convection can play a role in our models, this is not the convection that has formed the basis for the current paradigm in SNe. The critical convection behind the convection-enhanced neutrino-driven SN engine is convection above the proto-neutron star, in the region between the edge of the proto-neutron star and the stalled shock. This paper focuses on this currently favored convection-enhanced paradigm, both the observational evidence leading to its development (Sec. 2) and the current and ongoing measurements confirming this paradigm (Sec. 3).

2. Observed Asymmetries and the Birth of the Convective Paradigm

Due to the difficulty in including all of the physics in SN models, most SN-engine simulations until the 1990s were 1-dimensional. But, with better observations of SNe and their remnants, astronomers began to demonstrate that these explosions were anything but spherically-symmetric.= In this section, we review the observations the led to the development of the convection-enhanced neutrino-driven SN explosion engine that is the current paradigm for normal SNe.

2.1. *Observed Asymmetries that Drove the Convective Paradigm*

Observations of SN remnants, both ejecta and remnant, have long argued for an asymmetries in the SN engine. Large asymmetries have been observed in many of the observed SN remnants. In spherically symmetric models, the ejecta from the SN is expected to follow the same elemental abundance layered distribution that was produced in the stellar evolution with the iron peak elements in the center with the elements getting lighter as we move out in the ejecta. The distribution in the Cassiopeia A SN remnant not only displays heavier elements further out than lighter elements but also a sizable amount of mixing,[8] suggesting an asymmetry in the explosion mechanism. Determining whether this mixing was due to asymmetries in the explosion or in the blastwave propagation through the surrounding medium,[9] along with limitations in the requirements to observe these abundances (shock-heating requirements, out-of-equilibrium atomic physics), limited the constraints these observations placed on the engine.

Another observational indication that the explosion mechanism in core collapse had large asymmetries is the observed distribution pulsar velocities with some pulsars moving faster than $1000\,\mathrm{km\,s^{-1}}$.[10] The high velocities observed in the pulsar velocity distribution suggests that these "kicks" are imparted onto the neutron stars during formation. One of the leading mechanisms behind these kicks is the asymmetric explosion of a SN engine. In this mechanism, the newly-formed compact remnant is kicked in the opposite direction of the ejecta conserving linear momentum. But a wide range of models have been proposed to explain these kicks (for a review, see[11]) and many don't require an asymmetric SN explosion. Like SN remnants, although suggestive, pulsar velocities were not definitive proof of supernova asymmetries.

Observations of SN 1987A provided this evidence and drove SN engine theorists to begin to actively gain new insight into the CCSN engine. First and foremost, SN 1987A provided essential confirmation of the basic core-collapse idea. Although theorists had argued that a CCSN could be produced by stars above $\sim 10\,\mathrm{M}_\odot$, no definitive proof existed until 1987, when a star in the Large Magellanic Cloud collapsed. Prior to its collapse, this star had been observed and its identification of this progenitor, Sanduleak -69 202, providing the first irrefutable evidence that

at least some SNe are produced through the death of massive stars.[12] In addition, a burst of neutrinos were detected from this nearby SN at a rate that can only be produced with a compact object (either formation or a merger).[13,14] These two observations definitively demonstrated that at least some SNe are produced by the core-collapse to a compact object of a massive star.

Although SN 1987A solidified the foundation of the CCSN engine, it also raised a number of a new issues: the star was a blue, not a red, supergiant as predicted by models (we now believe this to be due to binary effects), the gamma-rays from the decay of radioactive nickel appeared much sooner than expected and the line profiles from these gamma-rays and infra-red lines (of the iron) were both broader than expected as well as being redshifted (instead of blue-shifted). From infra-red to gamma-rays, SN 1987A was showing the limitations of our spherically symmetric model. Scientists immediately realized that by artificially mixing out nickel produced in the innermost ejecta of 1-dimensional models, they could explain the early rise time of gamma-ray signal.[15] However, this artificial mixing did not explain the redshifted lines. In addition, producing the extent of mixing needed to explain the observations was difficult to achieve with simple Rayleigh-Taylor mixing developing in a symmetric shock propagated through the stellar profile.[16,17]

These difficulties led scientists to turn their attention to the engine itself and, instead of focusing on convection in the proto-neutron star, they focused on the region between the edge of the neutron star and the stalled shock. The unstable entropy gradient produced as the shock stalled led to convection in this region.[18,19] After a few preliminary studies, the first 2-dimensional calculations found this convection could change the fizzle produced by a 1-dimensional model into a strong explosion.[20] These instabilities evolved into low-mode convection which could both facilitate the extensive mixing and explain the pulsar kicks.[21] Three-dimensional models followed soon after,[22] including 3-dimensional explosion models that demonstrated that very low-mode convection could also explain the redshifted lines features.[23,24] This initiated a broad set of scientific work studying the SN engine that we will discuss in the next section.

2.2. *Evolution of the SN Engine*

Before we discuss the details of the convective engine, let's review the basic steps behind this leading paradigm:[25]

(i) **Onset of Collapse:** The iron core of a massive star because so massive that it compresses to densities that cause electrons to capture onto protons, removing electron degeneracy pressure. This causes further compression and accelerates the electron capture, driving a runaway collapse (Fig 2).

(ii) **Bounce** This runaway collapse proceeds until the core reaches nuclear densities where nuclear forces and neutron degeneracy pressure halt the collapse, causing

the infalling core to bounce. The bounce shock moves out through the star (Fig 2).

(iii) **Stall of the Shock:** Most of the energy in the shock is stored in neutrinos. When the neutrinos are no longer trapped in the core, the shock loses energy and stalls at ∼100-200 km (Fig 2).

(iv) **Convective Engine:** The region between the compact core and the stalled shock is unstable to a range of instabilities, including Rayleigh-Taylor convection. This convection transports energy deposited at its base and transfers it outward to push against the infalling star. At the same time, it transports mass piling up at the stalled shock and allows it to convect to the core, releasing further potential energy and reducing the pressure that must be overcome at the stalled shock. These two effects, not modeled in 1-dimensional models, make it easier to drive an explosion (Fig 3).

(v) **Shock Launch:** When the pressure in the convective engine can overcome the ram pressure of the infalling star, an explosion is launched. The shock will be highly asymmetric (Rayleigh-Taylor tends to evolve to the largest mode possible in a convective region, Standing Accretion Shock Instabilities also produce low-mode convection). This aspherical shock will allow further accretion after its launch, allowing further energy deposition. But the neutrino-energy deposition will decrease as the density above the proto-neutron star decreases.

Here we focus most of our discussion on the convective engine. Figure 3 shows a 2-dimensional slice of a 3-dimensional simulation. This image shows both the high-entropy upflows and the low-entropy downflows. During the 15-20 s after the first 2-dimensional simulation of this engine, most studies focused on reproducing these results. Additional instabilities were studied, most notably the standing accretion shock instability (SASI).[26] The primary difference between these two mechanisms is the growth time of the instability and the latter interpretation of the convective motions is favored by models where the turbulence growth is slow. Simulations with more rapid convective growth argue for Rayleigh-Taylor instabilities.

The growth time is a key feature in the final result, both in the explosion energy and the remnant mass, so understanding this convection is critical. Unfortunately, a number of numerical artifacts make it difficult to model this convection. We can not resolve the convection and artificial viscosity (either explicitly state in smooth particle hydrodynamics or implicit in the numerical advection of grid-based codes). This growth time can be estimated analytically and we can test our calculations against this analysis. The growth time of this convection is on order of the inverse of the Brunt-Väisäla frequency:

$$T_{\rm growth} = \sqrt{-\omega_{\rm BV}^2} \tag{1}$$

where ω_{BV} is given by:p

$$\omega_{\rm BV}^2 = -g_{\rm eff}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial r} - \frac{1}{\rho c_s^2}\frac{\partial P}{\partial r}\right) \tag{2}$$

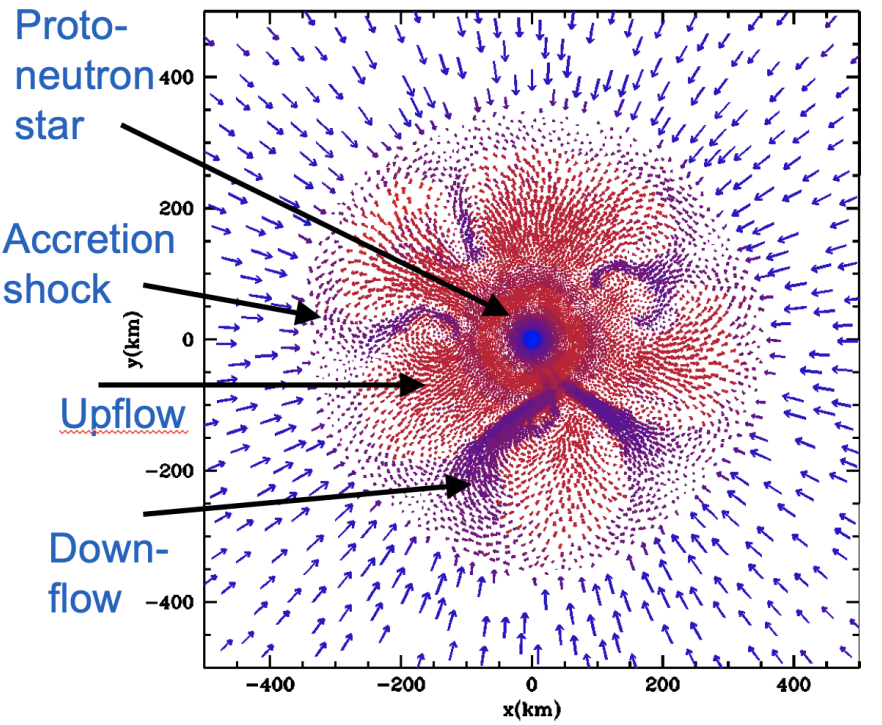


Fig. 3. Slice of a 3-dimensional convective engine. In this simulation, the convective region now extends to 300 km. The color-coding denotes entropy with red being high entropy and blue being low entropy. The vectors denote velocity magnitude (length) and direction. The proto-neutron star is in the inner 30 km. Material piles up at the top of the convective region near the stalled shock and the flows down in streams to the core.

where c_s is the sound speed, P is the pressure and $g_{\rm eff} = -\partial\Phi/\partial r + v_r \partial v_r/\partial r$ is the effective gravity with Φ is the gravitational potential and v_r is the radial velocity. For the region above the proto-neutron star, the growth time can be approximated by:

$$T_{\rm growth} \approx \sqrt{-\frac{GM_{\rm PNS}}{r_{\rm PNS}^2}\frac{\Delta S}{S\Delta r}} \tag{3}$$

where G is the gravitational constant, $M_{\rm PNS}$, $r_{\rm PNS}$ are mass and radius of the newly-formed, proto-neutron star, S is the entropy and ΔS is the change in entropy of a distance Δr in the region between the proto-neutron star and the stalled shock. With these conditions, we find the post-bounce convective growth-time to be on order of a few milliseconds.[25] All current calculations have growth times that are much longer than this prediction.

It could be that this simple analysis underestimates the growth time. We can further test the growth term by comparing our growth times to different prescriptions

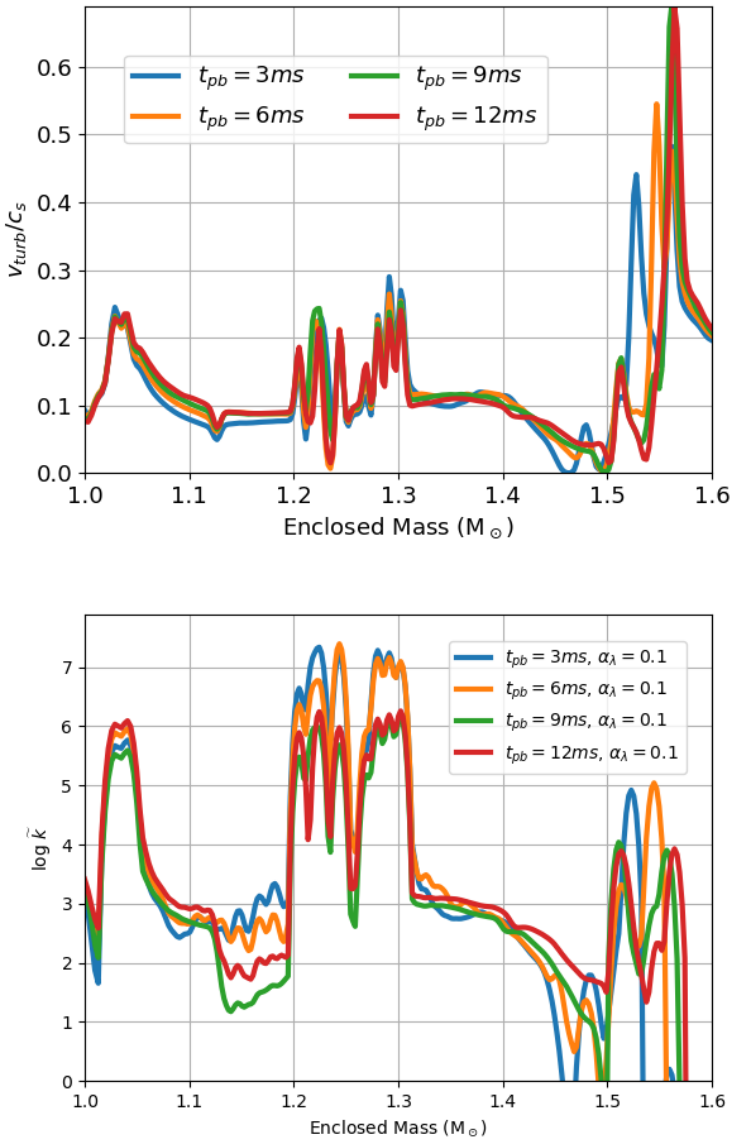


Fig. 4. Growth of turbulence, sound speed from mixing-length theory[27] and turbulent energy from a Reynolds Averaged Navier Stokes solution.[28] Strong turbulence develops in less than 10 ms.

(Fig 4) developed both in astrophysics (mixing-length theory[27]) and engineering (a Reynolds Averaged Navier Stokes solution[28]). Using 1-dimensional collapse models, these prescriptions also predict a very quick growth time for the instabilities: within a few ms, the Rayleigh-Taylor convection is strong with velocities/turbulent energies reaching a sizable fraction of the sound speed limit. Again, current models do not

reproduce these growth times. Until we match these growth times, our simulations will not be quantitatively accurate, but they can help us study many of the trends with respect to different physics in the SN engine. They can also be used to study qualitative features in the SN explosion.

Beyond insufficient resolution, what else can be missing in these convective models. One of the most important is the initial seeds used to drive the explosion. The initial 2- and 3-dimensional models used smooth particle hydrodynamics.[20,22] The initial setup conditions have an intrinsic stochasticity in the density and these initial studies argued that these perturbations were on par with the convective inhomogeneities of the progenitor star. Later studies focused on very extreme instabilities in an attempt to produce the observed neutron star kicks.[29,30] An interesting feature of this neutron-star kick mechanism is that, unlike most other kick mechanisms, it predicts that lower-mass progenitors will produce the strongest neutron star kicks. Observations can potentially confirm or refute this kick mechanism. More recent studies have modeled the stellar convection at the final stages of a star's life, following this burning through collapse to produce more consistent simulations of these convective seeds.[31] In all cases, these models found, as expected by analytic theories, that these seeds accelerated the growth of convection.

Active areas of research in the convective engine includes increased-resolution models to better understand the limitations of the numerical methods modeling this convection. But a number of more physics-informed models are also being developed to better understand convective in the collapsed core. These methods will lead to better implementation of subgrid models that can help capture effects like the turbulent pressure. Much more work must be done to understand convection and core-collapse engine theorists are bringing a wide array of numerical and analysis tools to improve our understanding of this engine.

3. Confirmation and Ongoing Constraints of the Convective Engine

Even with the limitations of the convective-enhanced neutrino-driven engine, we are able to make predictions from this model both to test it as a model and use observational features to help improve our understanding. Here we review current observational comparisons that have solidified this standard paradigm. We also discuss upcoming observations that will further test the engine, providing valuable insight into the explosion engine. Like 1987A, we expect upcoming observations to not only constrain this engine, but perhaps provide the evidence to pursue new insight into this developing engine.

3.1. *Explaining SN Energies*

Prior to the development of the convection-enhanced SN explosion mechanism, explaining why CCSNe tap only $\sim$1% of the total potential energy released in the

collapse of the core was one of the most baffling conundrums of the SN engine. But the convection-enhanced SN engine provided a natural explanation for this observation. The energy in the convective region continues to increase only until its energy is sufficient to overcome the ram pressure of the infalling material. The accretion rate of this infalling material and the time it takes the engine to overcome its ram pressure dictates the explosion energy. Figure 5 shows the energy required to overcome the ram pressure as a function of time in the collapse for a series of different progenitor stars. In all cases, this energy lies between a few tenths to a few times 10^{51}erg[a].

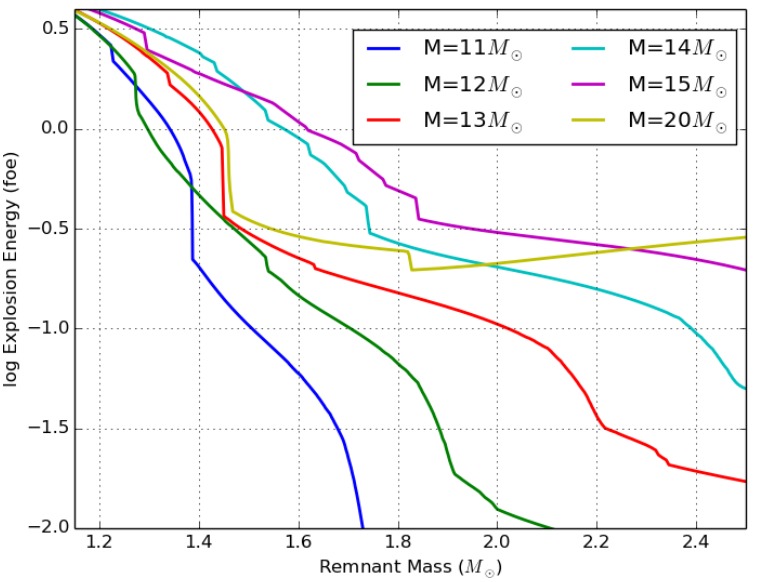


Fig. 5. Explosion energy vs. enclosed mass for different stellar models.

The power of this natural explanation for the explosion energy is often underappreciated in the astronomy community. Let's compare it to the leading alternative engine for SNe, the magnetar-engine scenario. In this scenario, strong magnetic fields tap the rotation energy of the SN to power the explosion. The rotational energy is given by:

$$E_{\rm magnetar} = 2/5 M_{\rm PNS} R_{\rm PNS}^2 \Omega_{\rm PNS}^2 \tag{4}$$

where $M_{\rm PNS}$, $R_{\rm PNS}$, and $\Omega_{\rm PNS}$ are the mass, radius and rotation rate of the proto-magnetar. When the explosion is launched (which must be in the first $\sim$ second after collapse so that the core does not accrete too much to collapse to a black hole), $R_{\rm PNS}$ is roughly 30 km. If the proto-neutron star contracts to form a 10 km

[a]Ten to the Fifty One Erg was termed a "foe" by Hans Bethe. This unit is also known as a Bethe.

neutron star spinning at 100 ms, we'd expect an average SN energy of 5×10^{47} erg instead of the observed $0.1 - 3 \times 10^{51}$ erg. Only the rare, fastest spinning stars would produce strong SNe.

3.2. *Predicting SN Progenitors*

One of the great successes of the convective-engine paradigm is its prediction of which stars produce strong SN explosions versus which collapse to form black holes. Under the convective paradigm, whether or not the convective engine can produce an explosion depends upon whether the energy of the convective engine can grow fast enough to overcome the ram pressure of the infalling material. A successful explosion is a race between the energy deposition timescale and the collapse timescale from accretion. If we ignore stellar winds, massive stars will have larger cores with higher accretion rates. This means that the timescale to collapse decreases and there is less time to develop convection and drive an explosion. In addition, the energy needed to overcome the higher accretion ram pressure is higher. Unless the engine becomes more efficient, these stars will not explode. Assuming minimal winds, the convective engine predicted that stars more massive than $\sim 23\, M_\odot$ would either fail to explode or produce weak explosions.[32]

Shortly thereafter, observations of SN lightcurves argued that most SN had ejecta masses above 15-20 $M_\odot$.[33] The magnetar model could produce explosions with such high ejecta masses, suggesting it might be the better standard SN engine. In this case, however, the analysis of the SN light-curves was wrong. As the sample of observed progenitors increased, it became clear that the vast majority of SN are produced from stars less massive that $20\, M_\odot$,[34] confirming the first prediction (vs. post-diction) of the convective engine and providing another strike for the magnetar model.

3.3. *Compact Remnant Distribution*

For the convective-engine paradigm, the mass of the compact remnant formed in CCSN can be described through a series of phases each of which depend on different properties of the explosion. The compact remnant initially forms during the collapse bounce phase. Typically, the mass of this bounce core is low (roughly 1.1 $M_\odot$[35]). As the convective engine develops, the core continues to accrete mass. Both the progenitor mass (that determines the infalling mass accretion rate with time) and the time it takes for this engine to develop dictates the mass of the core at the launch of the explosion. Finally, the explosion energy determines how much material ultimately falls back onto the compact core. All of these effects are important in determining the final compact remnant mass.

Observations of the compact remnant mass distribution provide insight into these explosion properties. Initially, observations argued for a remnant mass distribution that peaked strongly at two separate masses: a peak at $\sim$1.4 $M_\odot$ and another

peak at $\sim M_{\odot}$.[36] These highly-peaked mass distributions are difficult to reproduce within convective-engine paradigm, especially when including the fallback of material after the launch of the launch.[37] Fallback occurs because the SN blastwave must drive the entire star to explode. As it decelerates, some of the ejecta can decelerate sufficiently that this material can not escape the gravitational potential of the star. This fallback has been studied both analytically and through a series of simulations,[38,39] demonstrating that fallback accretion decreases following a power-law time evolution: $\dot{M}_{\rm fallback} \propto^{-5/3}$. The amount of integrated fallback depends on the explosion energy and the progenitor structure (and hence, its mass).

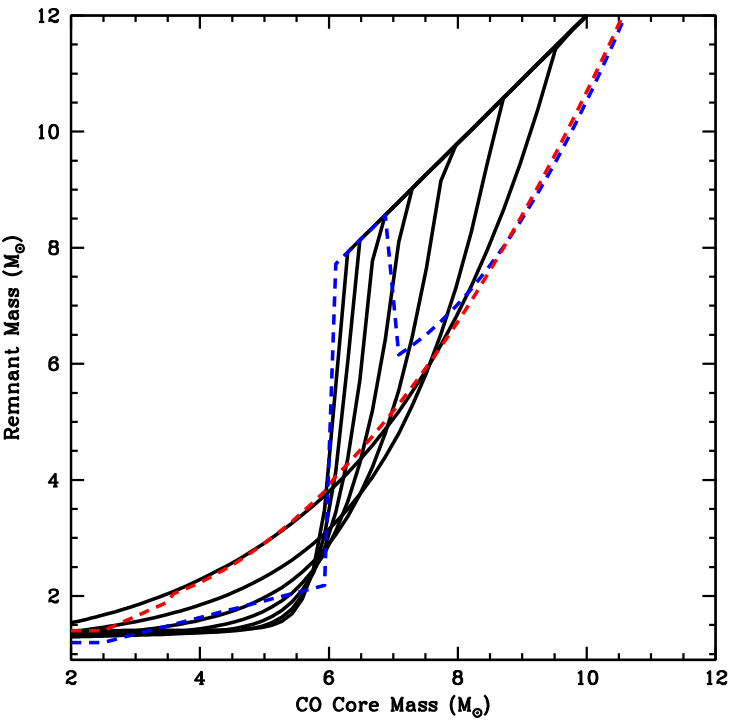


Fig. 6. Remnant mass as a function of CO core mass for rapid/slow convective models[35] along with a parameterized growth time model. The steeper the rise in remnant mass (caused by faster convective growth time), the deeper the mas gap produced.

One of the key observational constraints from the compact remnant mass distribution is the gap between neutron stars and black holes. Observations of neutron stars and black holes (binary pulsars and X-ray binaries) argued for an absence of neutron stars in the compact mass region lying between 2.5 and $4\,M_{\odot}$. Whether this was an observational bias or due to low-number statistics has been a matter of debate for many years. If the gap exists, it places strong constraints on the growth time of the convection.[35] Figure 6 shows the compact remnant mass as a function

of the CO core mass while varying the convective growth time (and, hence, the injection time of the energy). Producing a pristine mass gap requires very rapid and strong explosions to avoid falllback filling in this gap. It is possible that fallback will drive further outflows that then limit further fallback such that, even though fallback occurs, the amount of mass accreted is limited.[40] Until recently, the mass gap favored rapid explosions, affecting both our understanding of the SN engine and predictions for a wide range of SN properties. However, gravitational wave observations suggest that the mass gap region is not as pristine as originally believed[41] and, because of this, the constraints on the convective growth time are less stringent (we will discuss this further in section 3.6).

With all of the current observations, we must consider the observational biases. One of the potential observational biases of the mass gap is that the observed compact mass distribution is limited to systems in close binaries. Any comparison of the predictions of the remnant mass distribution to the observed mass distribution must include calculations of these binaries and are limited by the associated uncertainties of binary population synthesis.[42–44]

3.4. *SN Ejecta Remnants*

The ejecta from CCSNe form another form of remnant as the ejecta in the SN blastwave propagates through the interstellar medium. As we discussed in Section 2.1, the morphology of these ejecta remnants provided the first hints that the SN engine must be aspherical. Although these remnants are highly aspherical, it is often difficult to extract properties of the explosive engine from other effects in the propagation of a SN remnant. As also we discussed in section 2.1, the effects of the propagation through the circumstellar medium and the modeling of the atomic physics in these non-local thermodynamic equilibrium conditions can make it difficult to tie the properties of the remnant to the explosion itself. Nonetheless, remnants can prove powerful probes of the SN explosion and its engine.

First, let's better understand the difficulties through the example of analyses of the Cassiopeia A SN remnant. As the blastwave propagates through the circumstellar medium, it decelerates, causing a reverse shock to progress inward (in mass coordinates) through the remnant, lighting up the material and allowing astronomers to probe properties of the explosion.[45,46] But the morphology of the shock can be shaped by the circumstellar medium[9] and the masses and properties inferred (masses, abundance rations) depend on assumptions in the atomic physics, i.e.[47] Cassiopeia A demonstrates the pitfalls of such an analysis. Images and analysis of the shock-heated material in Cassiopeia A suggests a jet-like explosion, suggesting a magnetar and disk model.[8] To verify this conclusion, astronomers developed another probe of this morphology by, instead of observing shocked material, observing material through the decay of radioactive material produced in the explosion. In particular, the NuStar satellite mapped out the ^{44}Ti production from the SN

producing this remnant. ^{44}Ti is ideally suited as a probe of the central engine. It is produced in the innermost ejecta where ^{56}Ni is produced. It has a long enough half-life to be observed in SN remnants (unlike ^{56}Ni whose half-life is too short) and, because it is unstable (unlike iron), we know the observed ^{44}Ti was produced in the explosion. This innermost material may not yet be affected by the surrounding medium, especially if the reverse shock has yet to reach it. What the ^{44}Ti observations of the Cassiopeia A SN remnant showed was a clumpy distribution mimicking the convective motions in the SN engine.[48,49] This is the most direct observation to date proving the convection-enhance neutrino-driven SN engine. At least for this SN remnant, the engine is not matched by the magnetar model. Although limited to velocity imaging, the ^{44}Ti observations of SN 1987A suggest the same convective engine.[50]

Although these NuStar observations of remnants through the decay photons from radioactive elements provide a direct probe, the number of remnants that can be observed in such a manner is limited. Even with next generation gamma-ray detectors like COSI mission,[51] increasing the number of remnants with resolved images of the ^{44}Ti will be difficult. However, there exists a wealth of data of shock-heated material and combined with infra-red observations (and increasingly sophisticated models of the atomic physics), SN remnants will become an increasingly powerful probe of the SN engine. SN remnants measure key elements in the explosion that probe both the progenitor and the explosion energy. To demonstrate this, figure 7 shows the abundance fraction versus enclosed ejecta mass for two 1-dimensional SN explosions of a $15\,M_\odot$ star.[52] The abundances in this plot span the typically-observed abundances in SN remnants. The different observed abundances probe different aspects of the SN and different regions in the progenitor star. For example, Fe is produced in the innermost ejecta and is very sensitive to the explosion energy. Si, S, Ca, and Ar are produced just above the iron ejecta also probe the explosion energy. O, Mg, and Ne yields are produced even further out and are better probes of the progenitor because their production is much less sensitive to the explosion energy.

3.5. *Nucleosynthetic Yields*

Nucleosynthetic yields provide an excellent probe of the SN engine, especially of the elements produced near the engine (e.g. iron peak and heavier isotopes). As we discussed in Section 3.4, SN remnants provide a measure of the nucleosynthetic yields and their morphology. These remnant measurements probe the progenitor, explosion energy, and morphology of the explosion. ^{44}Ti observations are not only more direct probes (because the physics behind its analysis is more straightforward). The sensitivity of the ^{44}Ti production on the electron fraction of the inner ejecta[53] also means that it can probe the conditions in the convective region (and the properties of the neutrinos, including neutrino oscillations) in this region.

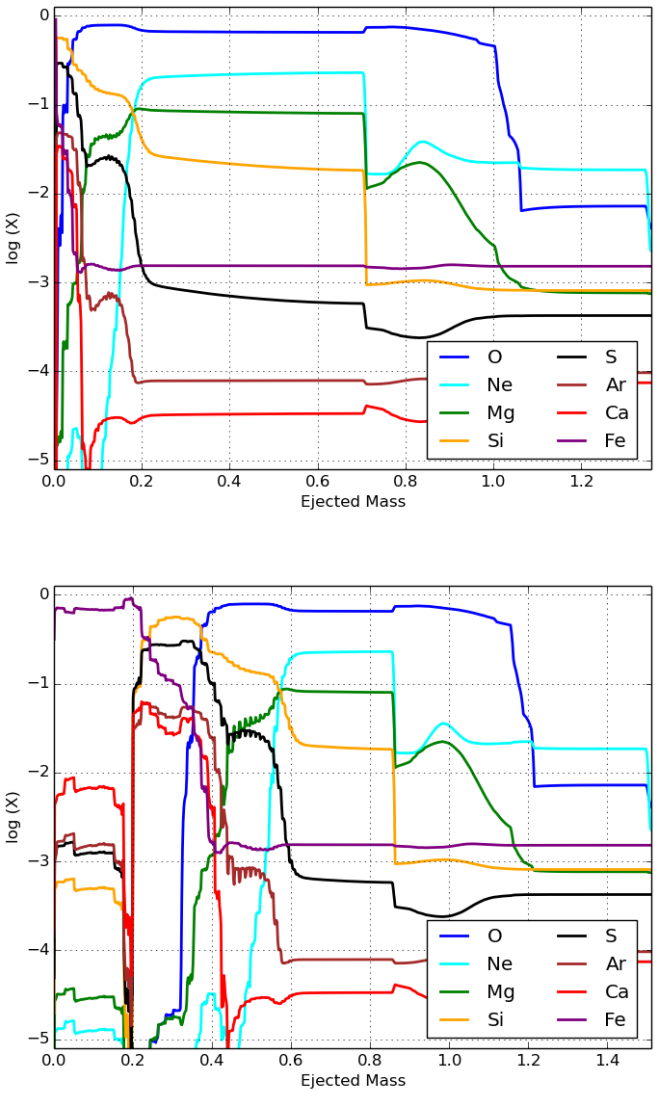


Fig. 7. Nucleosynthetic yiels from an 0.82 (top) and 1.86 (bottom) foe (10^{51} erg) explosion of a a 15 $M_\odot$ star. O, Mg, Ne are produced in stellar evolution and the supernova shock primarily destroys these abundances. Si, S, Ar, Ca are produced at the base of the C/O layer and their production depends somewhat on the SN explosion energy. Iron is produced in the core from shock-driven Si burning. It is very sensitive to the explosion energy, both to drive the nuclear fusion and to avoid falling back onto the proto-neutron star.

But remnants aren't the only way we can measure nucleosynthetic yields. Based on isotopic ratios, scientists have identified a set of pre-solar dust grains produced by SNe. With the detailed isotopic ratios and relative element abundances, scientists can begin to probe the detailed yields from SNe.[54–56] In addition, galactic chemical

evolution, which measures the evolution of the abunadances as a function of metallicity, can be used to both determine which transients form different elements and the variation in their production rates.[57, 58]

3.6. *Neutrinos and Gravitational Waves*

The neutrino detection of SN 1987A provided conclusive evidence that at least some SNe are produced in the formation of a compact remnant. Neutrinos remain one of the most direct probes of the SN engine. With a detailed neutrino spectrum, scientists can not only probe the nature of the proto-neutron star (and the nuclear matter equation of state) and the neutrino physics (including neutrino oscillations). Unfortunately, neutrinos are difficult to detect and, to date, SN 1987A remains the only SN detected in neutrinos. With hyperKamiokande, scientists expect to observer SNe out to 1 Mpc.[59] This can increase the neutrino detection rate, but for the detailed spectra to do most of the neutrino SN science will still require a Galactic SN.

Gravitational waves provide another[53] probe different physics of the inner engine: the rotation and the convection. To measure convection, we need a nearby SN, well within the Milky Way.[60] But, fast rotating collapses can be seen further out (in extreme cases, out to 1 Mpc).[61] With next generation detectors, a Galactic SN would produce a strong enough signal to measure the growth and extent of the SN mixing and the expected rate of fast-spinning collapses is sufficiently high that we will begin to place constraints on the angular momentum in stars (determining the importance of rotation in the SN engine). Next generation detectors will increase the number of mass-constraints on compact objects 100-fold, providing an extremely accurate mass distribution of compact objects (in close binary systems). With these masses, we can constrain the mixing in the SN engine (see Section 3.3).

3.7. *SN Light Curves and Early-Time Emission*

We have already discussed issues with using SN light-curves. One of the primary difficulties in studying the engines and progenitors of SNe from the SN light-curve lies in the fact that most of the emission occurs long after the SN blastwave has broken out of the star and interactions with the surrounding medium blur the features (and in particular the extent of the asymmetries) in the explosion and its progenitor star. Observing the emission from the SN blastwave as it breaks out of the star provides a much more direct probe of the explosion (asymmetries) and progenitor (convection, radii, and mass loss).[62–64] Figure 8 shows the early emission from a SN as it breaks out of the photosphere of a Wolf-Rayet star. This emission is driven by shocks produced as the SN blastwave propagates first through the convective envelope, then down the transition region from stellar edge to wind and then through the clumpy wind. By characterizing this emission, astronomers will probe the properties of the star, in particular, the convection in the stellar

envelope (convection is a key uncertainty of massive star evolution) and stellar mass-loss (which alters the structure of the star and can change its final fate). This emission also prov ides a probe of the SN asymmetries. Ultrasat[65] and SIBEX[66,67] satellite missions are being developed to obtain a large number of this early-time SN emission.

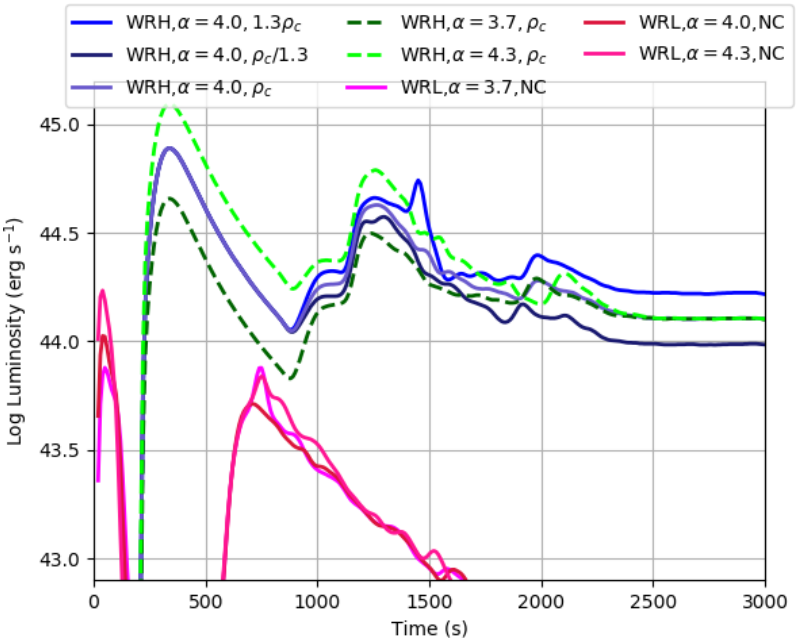


Fig. 8. Shock breakout emission (bolometric) from a high mass (WRH) and low mass (WRL) Wolf-Rayet progenitor star varying properties of the wind (with and without clumps) and the density slope of the transition region ($\rho = \rho_0 r^{-\alpha}$). Signal without clumps (NC) are compared to those with a wind clump density that varies by 30%. This emission is predominantly in the X-ray and UV bands.

4. Conclusion

In this paper, we reviewed the broad set of observations that led to, confirmed, and have constrained the convective-enhanced neutrino-driven SN engine paradigm. We have also outlined the the future observations that will place further constraints. For our understanding of the SN engine, the future is bright.

Acknowledgments

This section should come before the References. Dedications and funding information may also be included here.

References

1. W. Baade and F. Zwicky, *Proceedings of the National Academy of Science* **20** (May 1934) 254.
2. F. Zwicky, *ApJ* **88** (November 1938) 522.
3. R. Kippenhahn, S. Temesváry and L. Biermann, *Zeitschrift für Astrophysik* **46** (January 1958) 257.
4. A. Heger, S. E. Woosley, C. L. Fryer and N. Langer, Massive Star Evolution Through the Ages, in *From Twilight to Highlight: The Physics of Supernovae*, eds. W. Hillebrandt and B. Leibundgut (January 2003), p. 3.
5. S. A. Colgate and R. H. White, *ApJ* **143** (March 1966) 626.
6. R. I. Epstein, *MNRAS* **188** (August 1979) 305.
7. J. R. Wilson and R. W. Mayle, *Physics Reports* **163** (January 1988) 63.
8. U. Hwang, J. M. Laming, C. Badenes, F. Berendse, J. Blondin, D. Cioffi, T. DeLaney, D. Dewey, R. Fesen, K. A. Flanagan, C. L. Fryer, P. Ghavamian, J. P. Hughes, J. A. Morse, P. P. Plucinsky, R. Petre, M. Pohl, L. Rudnick, R. Sankrit, P. O. Slane, R. K. Smith, J. Vink and J. S. Warren, *ApJ Letters* **615** (November 2004) L117.
9. J. M. Blondin, P. Lundqvist and R. A. Chevalier, *ApJ* **472** (November 1996) 257.
10. J. M. Cordes and D. F. Chernoff, *ApJ* **505** (September 1998) 315.
11. C. L. Fryer and A. Kusenko, *ApJ Supp* **163** (April 2006) 335.
12. R. M. West, A. Lauberts, H. E. Jorgensen and H. E. Schuster, *Astronomy and Astrophysics* **177** (May 1987) L1.
13. K. Hirata, T. Kajita, M. Koshiba, M. Nakahata, Y. Oyama, N. Sato, A. Suzuki, M. Takita, Y. Totsuka, T. Kifune, T. Suda, K. Takahashi, T. Tanimori, K. Miyano, M. Yamada, E. W. Beier, L. R. Feldscher, S. B. Kim, A. K. Mann, F. M. Newcomer, R. van, W. Zhang and B. G. Cortez, *Physical Review Letters* **58** (April 1987) 1490.
14. R. M. Bionta, G. Blewitt, C. B. Bratton, D. Casper, A. Ciocio, R. Claus, B. Cortez, M. Crouch, S. T. Dye, S. Errede, G. W. Foster, W. Gajewski, K. S. Ganezer, M. Goldhaber, T. J. Haines, T. W. Jones, D. Kielczewska, W. R. Kropp, J. G. Learned, J. M. Losecco, J. Matthews, R. Miller, M. S. Mudan, H. S. Park, L. R. Price, F. Reines, J. Schultz, S. Seidel, E. Shumard, D. Sinclair, H. W. Sobel, J. L. Stone, L. R. Sulak, R. Svoboda, G. Thornton, J. C. van der Velde and C. Wuest, *Physical Review Letters* **58** (April 1987) 1494.
15. P. A. Pinto and S. E. Woosley, *ApJ* **329** (June 1988) 820.
16. D. Arnett, B. Fryxell and E. Mueller, *ApJ Letters* **341** (June 1989) L63.
17. M. Herant and W. Benz, *ApJ Letters* **370** (April 1991) L81.
18. M. Herant and W. Benz, *ApJ* **387** (March 1992) 294.
19. M. Herant, W. Benz and S. Colgate, *ApJ* **395** (August 1992) 642.
20. M. Herant, W. Benz, W. R. Hix, C. L. Fryer and S. A. Colgate, *ApJ* **435** (November 1994) 339.
21. M. Herant, *Space Science Reviews* **74** (November 1995) 335.
22. C. L. Fryer and M. S. Warren, *ApJ Letters* **574** (July 2002) L65.
23. A. L. Hungerford, C. L. Fryer and M. S. Warren, *ApJ* **594** (September 2003) 390.
24. A. L. Hungerford, C. L. Fryer and G. Rockefeller, *ApJ* **635** (December 2005) 487.
25. C. L. Fryer and P. A. Young, *ApJ* **659** (April 2007) 1438.
26. J. M. Blondin, A. Mezzacappa and C. DeMarino, *ApJ* **584** (February 2003) 971.
27. E. Böhm-Vitense, *Zeitschrift für Astrophysik* **46** (January 1958) 108.
28. D. Livescu, J. R. Ristorcelli, R. A. Gore, S. H. Dean, W. H. Cabot and A. W. Cook, *Journal of Turbulence* **10** (January 2009) 13.
29. A. Burrows and J. Hayes, *Physical Review Letters* **76** (January 1996) 352.
30. C. L. Fryer, *ApJ Letters* **601** (February 2004) L175.

31. C. E. Fields and S. M. Couch, *ApJ* **901** (September 2020) 33.
32. C. L. Fryer, *ApJ* **522** (September 1999) 413.
33. M. Hamuy, *ApJ* **582** (January 2003) 905.
34. S. J. Smartt, *ARA&A* **47** (September 2009) 63.
35. C. L. Fryer, K. Belczynski, G. Wiktorowicz, M. Dominik, V. Kalogera and D. E. Holz, *ApJ* **749** (April 2012) 91.
36. S. E. Thorsett and D. Chakrabarty, *ApJ* **512** (February 1999) 288.
37. C. L. Fryer and V. Kalogera, *ApJ* **554** (June 2001) 548.
38. R. A. Chevalier, *ApJ* **346** (November 1989) 847.
39. C. L. Fryer, *New Astron.* **50** (October 2006) 492.
40. C. L. Fryer, S. A. Colgate and P. A. Pinto, *ApJ* **511** (February 1999) 885.
41. A. et al., *arXiv e-prints* (August 2021) arXiv:2108.01045.
42. C. L. Fryer, S. E. Woosley and D. H. Hartmann, *ApJ* **526** (November 1999) 152.
43. M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel and R. O'Shaughnessy, *ApJ* **759** (November 2012) 52.
44. P. Drozda, K. Belczynski, R. O'Shaughnessy, T. Bulik and C. L. Fryer, *arXiv e-prints* (September 2020) arXiv:2009.06655.
45. J. C. Raymond, *ARA&A* **22** (January 1984) 75.
46. S. P. Reynolds, *ARA&A* **46** (September 2008) 89.
47. J. C. Raymond, B. C. Koo, Y. H. Lee, D. Milisavljevic, R. A. Fesen and I. Chilingarian, *ApJ* **866** (October 2018) 128.
48. B. W. Grefenstette, F. A. Harrison, S. E. Boggs, S. P. Reynolds, C. L. Fryer, K. K. Madsen, D. R. Wik, A. Zoglauer, C. I. Ellinger, D. M. Alexander, H. An, D. Barret, F. E. Christensen, W. W. Craig, K. Forster, P. Giommi, C. J. Hailey, A. Hornstrup, V. M. Kaspi, T. Kitaguchi, J. E. Koglin, P. H. Mao, H. Miyasaka, K. Mori, M. Perri, M. J. Pivovaroff, S. Puccetti, V. Rana, D. Stern, N. J. Westergaard and W. W. Zhang, *Nature* **506** (February 2014) 339.
49. B. W. Grefenstette, C. L. Fryer, F. A. Harrison, S. E. Boggs, T. DeLaney, J. M. Laming, S. P. Reynolds, D. M. Alexander, D. Barret, F. E. Christensen, W. W. Craig, K. Forster, P. Giommi, C. J. Hailey, A. Hornstrup, T. Kitaguchi, J. E. Koglin, L. Lopez, P. H. Mao, K. K. Madsen, H. Miyasaka, K. Mori, M. Perri, M. J. Pivovaroff, S. Puccetti, V. Rana, D. Stern, N. J. Westergaard, D. R. Wik, W. W. Zhang and A. Zoglauer, *ApJ* **834** (January 2017) 19.
50. S. E. Boggs, F. A. Harrison, H. Miyasaka, B. W. Grefenstette, A. Zoglauer, C. L. Fryer, S. P. Reynolds, D. M. Alexander, H. An, D. Barret, F. E. Christensen, W. W. Craig, K. Forster, P. Giommi, C. J. Hailey, A. Hornstrup, T. Kitaguchi, J. E. Koglin, K. K. Madsen, P. H. Mao, K. Mori, M. Perri, M. J. Pivovaroff, S. Puccetti, V. Rana, D. Stern, N. J. Westergaard and W. W. Zhang, *Science* **348** (May 2015) 670.
51. J. A. Tomsick, S. E. Boggs, A. Zoglauer, E. Wulf, L. Mitchell, B. Phlips, C. Sleator, T. Brandt, A. Shih, J. Roberts, P. Jean, P. von Ballmoos, J. Martinez Oliveros, A. Smale, C. Kierans, D. Hartmann, M. Leising, M. Ajello, E. Burns, C. Fryer, P. Saint-Hilaire, J. Malzac, F. Tavecchio, V. Fioretti, A. Bulgarelli, G. Ghirlanda, H.-K. Chang, T. Takahashi, K. Nakazawa, S. Matsumoto, T. Melia, T. Siegert, A. Lowell, H. Lazar, J. Beechert and H. Gulick, *arXiv e-prints* (September 2021) arXiv:2109.10403.
52. S. Andrews, C. Fryer, W. Even, S. Jones and M. Pignatari, *ApJ* **890** (February 2020) 35.
53. G. Magkotsios, F. X. Timmes, A. L. Hungerford, C. L. Fryer, P. A. Young and M. Wiescher, *ApJ Supp* **191** (November 2010) 66.
54. L. R. Nittler, S. Amari, R. S. Lewis, R. M. Walker and E. Zinner, *Meteoritics* **30** (September 1995) 557.

55. A. N. Nguyen and S. Messenger, *ApJ* **784** (April 2014) 149.
56. J. Schulte, M. Bose, P. A. Young and G. S. Vance, *ApJ* **908** (February 2021) 38.
57. B. Côté, M. Eichler, A. Arcones, C. J. Hansen, P. Simonetti, A. Frebel, C. L. Fryer, M. Pignatari, M. Reichert, K. Belczynski and F. Matteucci, *ApJ* **875** (April 2019) 106.
58. M. N. Ishigaki, T. Hartwig, Y. Tarumi, S.-C. Leung, N. Tominaga, C. Kobayashi, M. Magg, A. Simionescu and K. Nomoto, *MNRAS* **506** (October 2021) 5410.
59. K. e. a. Abe, *arXiv e-prints* (January 2019) arXiv:1901.03750.
60. C. L. Fryer and K. C. B. New, *Living Reviews in Relativity* **14** (January 2011) 1.
61. C. L. Fryer, D. E. Holz and S. A. Hughes, *ApJ* **565** (January 2002) 430.
62. E. Waxman and B. Katz, Shock Breakout Theory, in *Handbook of Supernovae*, eds. A. W. Alsabti and P. Murdin 2017, p. 967.
63. C. L. Fryer, C. J. Fontes, J. S. Warsa, P. W. A. Roming, S. X. Coffing and S. R. Wood, *ApJ* **898** (August 2020) 123.
64. C. M. Irwin, I. Linial, E. Nakar, T. Piran and R. Sari, *MNRAS* (September 2021).
65. I. Sagiv, A. Gal-Yam, E. O. Ofek, E. Waxman, O. Aharonson, S. R. Kulkarni, E. Nakar, D. Maoz, B. Trakhtenbrot, E. S. Phinney, J. Topaz, C. Beichman, J. Murthy and S. P. Worden, *AJ* **147** (April 2014) 79.
66. P. W. A. Roming, T. A. Pritchard, J. L. Prieto, C. S. Kochanek, C. L. Fryer, K. Davidson, R. M. Humphreys, A. J. Bayless, J. F. Beacom, P. J. Brown, S. T. Holland, S. Immler, N. P. M. Kuin, S. R. Oates, R. W. Pogge, G. Pojmanski, R. Stoll, B. J. Shappee, K. Z. Stanek and D. M. Szczygiel, *ApJ* **751** (June 2012) 92.
67. P. Roming, C. Fryer, A. Bayless and J. De la Rosa, Shock Breakout and Shock Interaction Mission, in *43rd COSPAR Scientific Assembly. Held 28 January - 4 February*, (January 2021), p. 1640.

The observation of high-energy neutrinos from the cosmos: Lessons learned for multimessenger astronomy

Francis Halzen

Department of Physics, University of Wisconsin–Madison, Madison, WI 53705, US

The IceCube neutrino telescope discovered PeV-energy neutrinos originating beyond our Galaxy with an energy flux that is comparable to that of GeV-energy gamma rays and EeV-energy cosmic rays. These neutrinos provide the only unobstructed view of the cosmic accelerators that power the highest energy radiation reaching us from the universe. We will review the results from IceCube's first decade of operations, emphasizing the measurement of the diffuse multiflavored neutrino flux from the universe and the identification of the supermassive black hole TXS 0506+056 as a source of cosmic neutrinos and, therefore, cosmic rays. We will speculate on the lessons learned for multimessenger astronomy, among them that extragalactic neutrino sources may be a relatively small subset of the cosmic accelerators observed in high-energy gamma rays and that these may be gamma-ray-obscured at the times that they emit neutrinos.

Keywords: Neutrinos; cosmic rays; multimessenger astronomy.

PACS numbers: 95.55.Vj; Neutrino; muon; pion; and other elementary particle detectors; cosmic ray detectors

1. Neutrino Astronomy: A Brief Introduction

The shortest wavelength radiation reaching us from the universe is not radiation at all; it consists of cosmic rays—protons and high-energy nuclei. Some reach us with extreme energies exceeding 10^8 TeV from a universe beyond our Galaxy that is obscured to gamma rays and from which only neutrinos reach us as astronomical messengers.[1] Their origin is still unknown but the identification of a supermassive black hole powering a cosmic-ray accelerator[2,3] represents a breakthrough towards a promising path for resolving the century-old puzzle of the origin of cosmic rays: multimessenger astronomy.

The rationale for searching for cosmic-ray sources by observing neutrinos is straightforward: in relativistic particle flows near neutron stars or black holes, some of the gravitational energy released in the accretion of matter is transformed into the acceleration of protons or heavier nuclei, which subsequently interact with ambient radiation or matter to produce pions and other secondary particles that decay into neutrinos. For instance, when protons interact with intense radiation fields near the source via the photoproduction processes

$$p + \gamma \to \pi^0 + p \quad \text{and} \quad p + \gamma \to \pi^+ + n\,, \tag{1}$$

both neutrinos and gamma rays are produced with roughly equal rates; while neutral pions decay into two gamma rays, $\pi^0 \to \gamma + \gamma$, the charged pions decay into three high-energy neutrinos (ν) and antineutrinos ($\bar{\nu}$) via the decay chain $\pi^+ \to \mu^+ + \nu_\mu$ followed by $\mu^+ \to e^+ + \bar{\nu}_\mu + \nu_e$. The fact that cosmic neutrinos are inevitably accompanied by high-energy photons transforms neutrino astronomy into multimessenger astronomy.

A main challenge of multimessenger astronomy is to separate these photons, which we will refer to as *pionic* photons, from photons radiated by electrons that are accelerated along with the cosmic ray protons. Another challenge is to identify the electromagnetic energy associated with the pionic photons because they do not reach our telescopes with their initial energy. They suffer losses in interactions with the extragalactic background light (EBL), predominantly with microwave photons, via the process $\gamma + \gamma_{cmb} \to e^+ + e^-$. Importantly, they may also lose energy in the source. As is the case when constructing a neutrino beam in a particle physics laboratory, neutrinos require a target that transforms the energy of the proton beam into neutrinos. Powerful neutrino sources within reach of IceCube's sensitivity require a dense target that is likely to be opaque to pionic gamma rays. Additionally losing energy in the source, these may reach Earth with MeV energies or below. We will review the accumulating indications that cosmic neutrinos originate in gamma-obscured sources with their associated multimessenger signals emerging below the detection thresholds of high-energy gamma-ray satellites and ground-based TeV gamma-ray telescopes.[4,5]

2. High-Energy Cosmic Neutrinos

Close to the National Science Foundation's research station located at the geographical South Pole, the IceCube project[6] transformed one cubic kilometer of natural Antarctic ice into a Cherenkov detector. The deep ice of the Antarctic glacier constitutes the detector, forming both support structure and Cherenkov medium. Below a depth of 1,450 meters, a cubic kilometer of glacial ice has been instrumented with 86 cables called "strings," each of which is equipped with 60 optical sensors; see Fig. 1. Each digital optical module (DOM) consists of a glass sphere containing a photomultiplier and the electronics board that captures and digitizes the light signals locally using an onboard computer; see Fig. 2. The digitized signals are given a global time stamp accurate to 2 ns and are subsequently transmitted to the surface. Processors at the surface continuously collect the time-stamped signals from the optical modules, each of which functions independently. The digital messages are sent to a string processor and a global event trigger. They are sorted into the Cherenkov radiation patterns that are emitted by muon tracks produced by muon neutrinos interacting in the ice, or by secondary particle showers initiated by electron and tau neutrinos. These reveal the flavor, energy, and direction of the incident neutrino.[7]

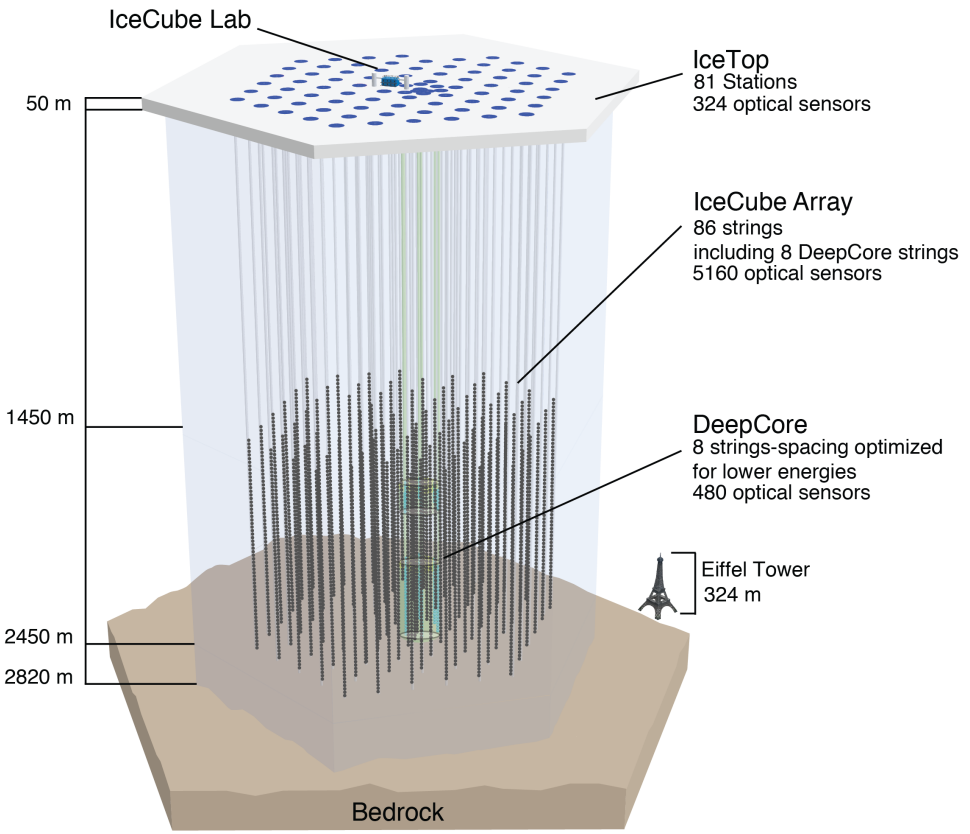


Fig. 1. Architecture of the IceCube observatory.

Constructed between 2004 and 2010, IceCube has now taken 10 years of data with the completed detector.

The arrival direction of a secondary muon track, or of an electromagnetic shower initiated by an electron or tau neutrino, is reconstructed from the arrival times of the Cherenkov photons at the optical sensors, while the number of photons is a proxy for the energy deposited by secondary particles in the detector. Although the detector only records the energy of the secondary muon inside the detector, from Standard Model physics we can infer the energy spectrum of the parent neutrino.

Tracks resulting from muon neutrino interactions can be pointed back to their sources with a $\leq 0.4^\circ$ angular resolution for the highest energy events. In contrast, the reconstruction of cascade directions, in principle possible to within a few degrees, is still in the development stage in IceCube, achieving 8° resolution.[8,9] On the other hand, determining the shower energy from the observed light pool is straightforward, and a resolution of better than 15% can be achieved. For illustration, we contrast

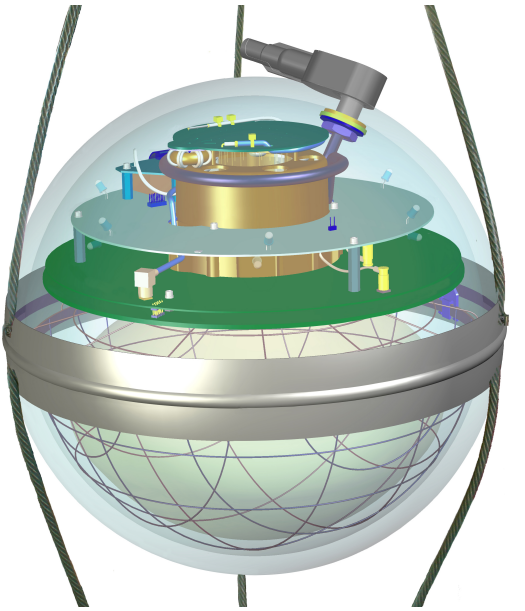


Fig. 2. Digital optical module showing the down-facing 10-inch photomultiplier and the associated electronics that digitize the light signals.

in Fig. 3 the Cherenkov patterns initiated by an electron (or tau) neutrino of 1 PeV energy (top) and a neutrino-induced muon losing 2.6 PeV energy while traversing the detector (bottom).

IceCube identifies cosmic neutrinos in a large background of muons and neutrinos produced by cosmic rays interacting in the atmosphere. Two principal methods are used to separate neutrinos of cosmic origin from the background of atmospheric muons and neutrinos. The first method reconstructs muon tracks reaching the detector from directions below the horizon, the second identifies neutrinos of all flavors that interact inside the instrumented volume of the detector; examples of events are shown in Fig. 3. We will describe these methods in turn.

2.1. *Muon Neutrinos Through the Earth*

Detecting particles from directions below the horizon has the immediate advantage of eliminating the overwhelming background of cosmic-ray muons that reach the detector from above. The tracks are separated from the background of atmospheric muons using the Earth as a filter. IceCube thus collects samples of muon neutrinos with high purity, typically well above 99%, at a rate of more than 100,000 per year. The neutrino energies cover more than six orders of magnitude, from ~ 5 GeV in the highly instrumented inner core, labeled DeepCore in Fig. 1, to extreme energies beyond 10 PeV. IceCube thus measured the atmospheric neutrino flux over more than five orders of magnitude in energy with a result that is consistent with theoretical

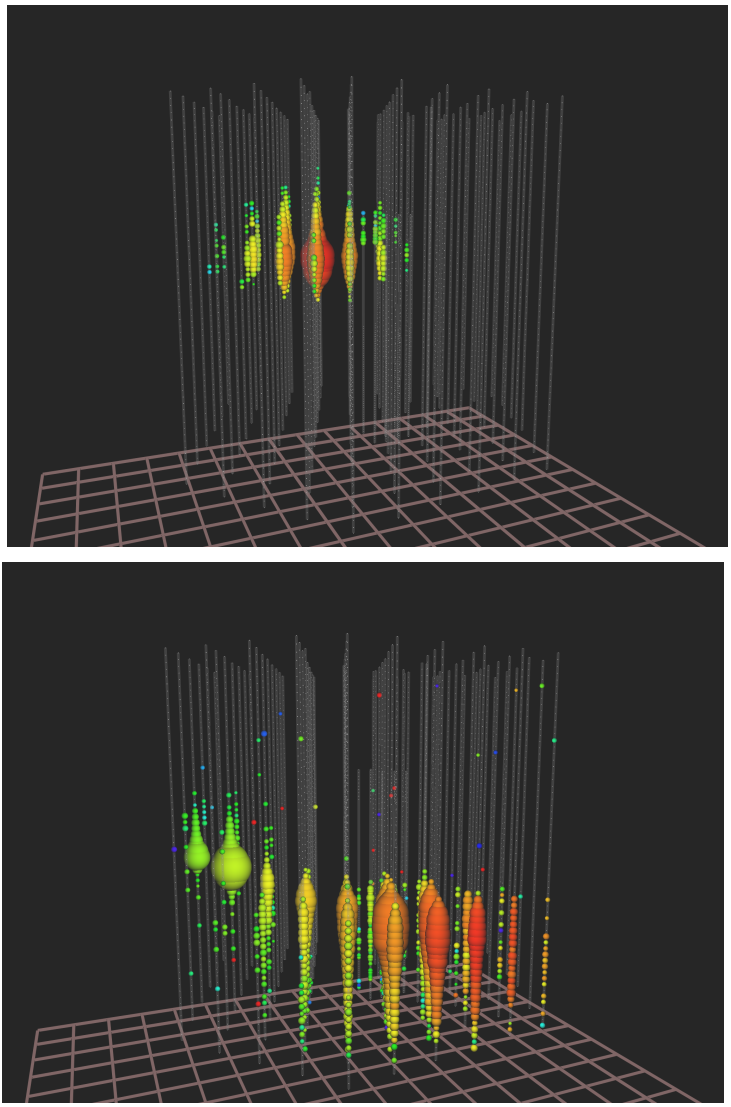


Fig. 3. **Top Panel:** Light pool produced in IceCube by a shower initiated by an electron or tau neutrino of 1.14 PeV, which represents a lower limit on the energy of the neutrino that initiated the shower. White dots represent sensors with no signal. For the colored dots, the color indicates arrival time, from red (early) to purple (late) following the rainbow, and size reflects the number of photons detected. **Bottom Panel:** A muon track coming up through the Earth, traverses the detector at an angle of 11° below the horizon. The deposited energy, i.e., the energy equivalent of the total Cherenkov light of all charged secondary particles inside the detector, is 2.6 PeV.

calculations. Muon neutrinos can be detected even when interacting outside the detector because of the kilometer range of the secondary muons. More importantly, IceCube has observed an excess of neutrino events at energies beyond 100 TeV[10–12]

that cannot be accounted for by the atmospheric flux at the 5.6σ level. A recent measurement of the energy flux covering 9.5 years of data was performed yielding a sample of 650,000 neutrinos with 99.7% purity; see Fig. 4. The excess cosmic neutrino flux (red) over the atmospheric background (purple) is well described by a power law with a spectral index of -2.37 ± 0.09 and a normalization at 100 TeV neutrino energy of $(1.36^{+0.24}_{-0.25}) \times 10^{-18}\,\mathrm{GeV}^{-1}\mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{s}^{-1}$.[13] The residual atmospheric muon background is small (yellow). For more details, see Ref. 14.

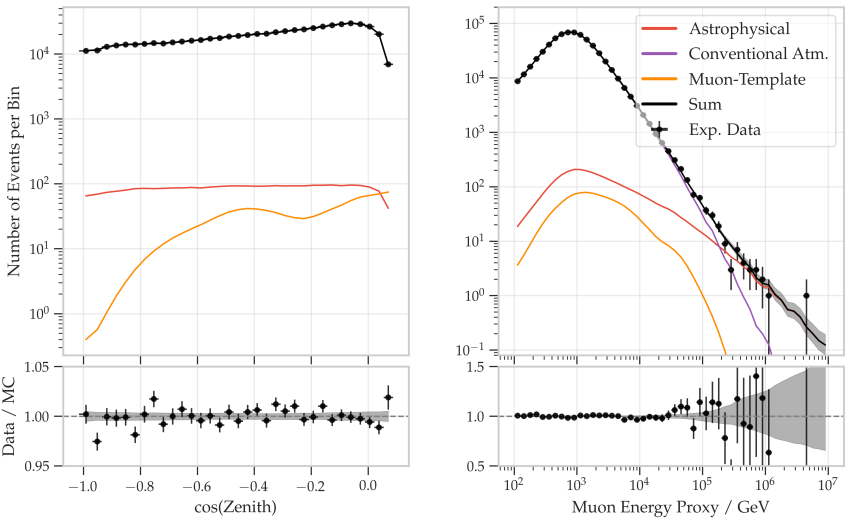


Fig. 4. The distributions of muon tracks arriving from the Northern Hemisphere, i.e., with declination greater than $-5°$, for the period 2010-18,[13] are shown as a function of reconstructed zenith (left) and muon energy (right). The full data set consists of about 650,000 neutrino events with a purity of 99.7%. Best fits to the low-energy atmospheric and high-energy astrophysical components of the neutrino flux are superimposed. Statistical errors are shown as crosses, the grey bands in the ratio plots show an estimate of the systematic error obtained by varying all fit-parameters within their uncertainties.

The measured arrival directions of the astrophysical muon tracks are isotropically distributed over the sky. Surprisingly, there is no evidence for a correlation to nearby sources in the Galactic plane; IceCube observes a diffuse flux of extragalactic sources. Only after collecting 10 years of data[15] did the first evidence emerge at the 3σ level that the neutrino sky is not isotropic. The anisotropy results from four sources—TXS 0506+056 among them (more about that source later)—that emerge as point sources of neutrinos with a p-value of less than 0.01 (pretrial); see Fig. 5. The strongest of these sources is the nearby active galaxy NGC 1068, also known as Messier 77, which, independently, also emerges as the most significant source in a list of about a hundred candidates that had been preselected by the collaboration as targets of interest.

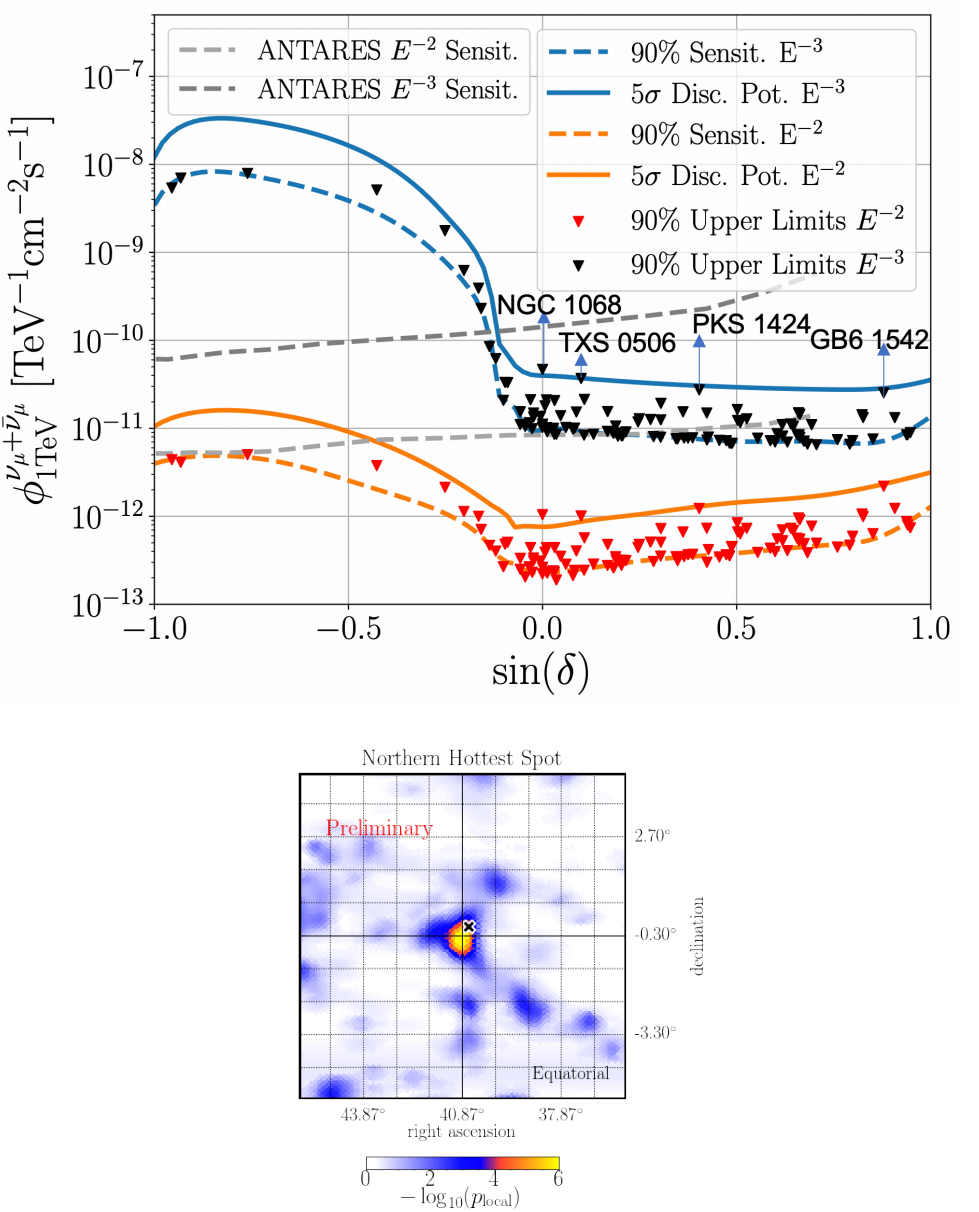


Fig. 5. **Top Panel:** Upper limits on the flux from candidate point sources of neutrinos in 10 years of IceCube data assuming two spectral indices of the flux.[15] Also shown as triangles are limits on a preselected list of about one hundred candidate sources. Four sources exceed the 4σ level (pretrial) and collectively result in a 3σ anisotropy of the sky map. **Bottom Panel:** Association of the hottest source in the sky map as well as in the list of preselected candidate sources with the active galaxy NGC 1068.

2.2. *Neutrinos Interacting Inside the Instrumented Volume*

The second method for separating cosmic from atmospheric neutrinos exclusively identifies high-energy neutrinos interacting inside the instrumented volume of the detector, so-called starting events. After only two years of operation, IceCube used this method to make the initial discovery of an extragalactic flux of cosmic neutrinos[1] with an energy flux, $E^2 dN/dE$, in the local universe that is, surprisingly, similar to that in gamma rays.[16,17] Using this method, one divides the instrumented volume of ice into an outer veto shield and a $\sim$ 500-megaton inner fiducial volume. The advantage of focusing on neutrinos interacting inside the instrumented volume of ice is that the detector functions as a total absorption calorimeter,[8] allowing for a good energy measurement that separates cosmic from lower-energy atmospheric neutrinos. In contrast to the first method, neutrinos from all directions in the sky and of all flavors can be identified, including both muon tracks and secondary showers produced by charged-current interactions of electron and tau neutrinos and neutral current interactions of neutrinos of all flavors. A sample event with a light pool of roughly one hundred thousand photoelectrons extending over more than 500 meters is shown in the top panel of Fig. 3.

The starting event samples revealed the first evidence for neutrinos of cosmic origin.[1,18] Events with PeV energies and with no trace of coincident muons that reveal either the decay products of a parent meson or an accompanying atmospheric shower are highly unlikely to be of atmospheric origin. The present seven-year data set contains a total of 60 neutrino events with deposited energies ranging from 60 TeV to 10 PeV that are likely to be of cosmic origin. The deposited energy and zenith dependence of the high-energy starting events[12,19] is compared to the atmospheric background in Fig. 6. A purely atmospheric explanation of the observation is excluded at 8σ.

The flux of cosmic neutrinos has by now also been characterized with a range of other methods. Their results agree, pointing at extragalactic sources whose flux has equilibrated in the three neutrino flavors after propagation over cosmic distances,[21] with $\nu_e : \nu_\mu : \nu_\tau \sim 1 : 1 : 1$. Fig. 7 shows the measurement of the cosmic neutrino flux of a recent analysis that specializes to showers only. These have been isolated from the atmospheric background down to energies below 10 TeV.[22] The energy spectrum of $E^{-2.5}$ agrees with the measurement using upgoing muons with a spectral index of $E^{-2.4}$ above an energy of $\sim$ 100 TeV.[12] In general, analyses reaching lower energies exhibit larger spectral indices with the updated 7.5 years starting-event sample,[19] yielding a spectral index value of -2.87 ± 0.2 for the 68.3% confidence interval.

2.3. *Cosmic Tau Neutrinos*

We should comment at this point that there is yet another method to conclusively identify neutrinos that are of cosmic origin: the observation of high-energy tau neutrinos. Tau neutrinos are produced in the atmosphere by the oscillations of muon neutrinos into tau neutrinos, but only for neutrino energies well below 100 GeV.

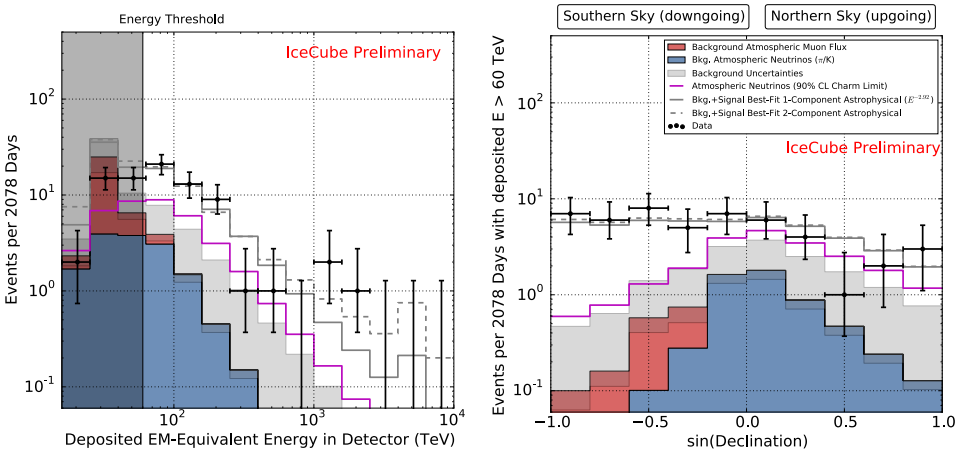


Fig. 6. **Left Panel:** Deposited energies, by neutrinos interacting inside IceCube, observed in six years of data.[12] The grey region shows uncertainties on the sum of all backgrounds. The atmospheric muon flux (blue) and its uncertainty is computed from simulation to overcome statistical limitations in our background measurement and scaled to match the total measured background rate. The atmospheric neutrino flux is derived from previous measurements of both the π, K, and charm components of the atmospheric spectrum.[20] Also shown are two fits to the spectrum, assuming a simple power-law (solid gray) and a broken power-law (dashed gray). **Right Panel:** The same data and models, but now showing the distribution of events with deposited energy above 60 TeV in declination. At the South Pole, the declination angle δ is equivalent to the distribution in zenith angle θ related by the identity, $\delta = \theta - \pi/2$. It is clearly visible that the data is flat in the Southern Hemisphere, as expected from the contribution of an isotropic astrophysical flux.

Above that energy a tau neutrino must be of cosmic origin, produced in cosmic accelerators with a neutrino flux with a tau fraction of approximately one third. Tau neutrinos produce two spatially separated showers in the detector, one from the interaction of the tau neutrino and the second from the decay of the secondary tau produced.[23] The mean decay length of the tau lepton is $\lambda_\tau = (E_\tau/m)\, c\tau \approx 50\ \mathrm{m} \times (E_\tau/\mathrm{PeV})$, where m, τ, and E_τ are the mass, lifetime, and energy of the tau, respectively. Two such candidate events have been identified.[24] An event with a decay length of 17 m and a probability of 98% of being produced by a tau neutrino is shown in Fig. 8. The energies of the two showers are 9 TeV and 80 TeV.

Yet another independent confirmation of the observation of neutrinos of cosmic origin appeared in the form of the Glashow resonance event shown in Fig. 9. The event was identified in a dedicated search for partially contained showers of very high energy.[25] The reconstructed energy of the shower is 6.3 PeV, which matches the laboratory energy for the production of a weak intermediate W^- in the resonant interaction of an electron antineutrino with an atomic electron: $\bar{\nu}_e + e^- \to W^- \to q + \bar{q}$. Given its high energy, the initial neutrino is cosmic in origin; it represents an independent discovery of cosmic neutrinos at the level of 5σ. Assuming the Standard Model cross section, we expect 1.55 events in the data sample searched, assuming an antineutrino:neutrino ratio of 1:1 characteristic of a cosmic source producing

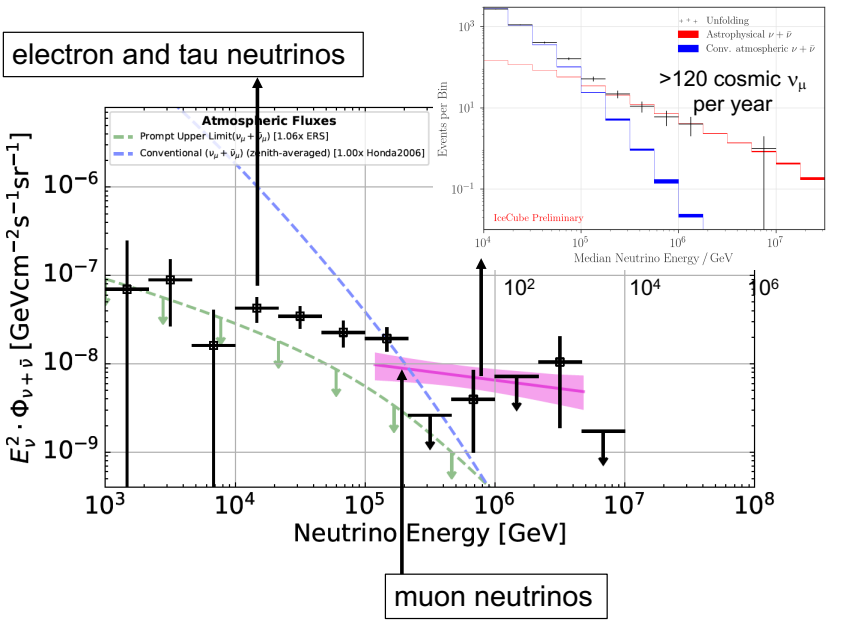


Fig. 7. The flux of cosmic muon neutrinos[12] inferred from the eight-year upgoing-muon track analysis (red solid line) with 1σ uncertainty range (shaded range; from fit shown in upper-right inset) is compared with the flux of showers initiated by electron and tau neutrinos.[22] The measurements are consistent assuming that each neutrino flavor contributes an identical flux to the diffuse spectrum.

an equal number of pions of all three electric charges. Taking into account the detector's energy resolution, the probability that the event is produced off resonance by deep inelastic scattering is only 0.01, assuming a spectrum with a spectral index of $\gamma = -2.5$. Furthermore, the presence of both muons and an electromagnetic shower is consistent with the hadronic decay of a W^- produced on the Glashow resonance.[25]

The observation of a Glashow resonance event indicates the presence of electron antineutrinos in the cosmic neutrino flux. Its unique signature provides a method to disentangle neutrinos from antineutrinos; their ratio distinguishes accelerators that produce neutrinos via pp and $p\gamma$ interactions and is also sensitive to their magnetic field.[25]

Finally, data from the ANTARES experiment are consistent with the observation of a flux of cosmic origin although with limited statistical significance.[26]

3. Multimessenger Astronomy

The most important message emerging from the IceCube measurements may not be apparent yet: the prominent role of neutrinos relative to photons in the extreme

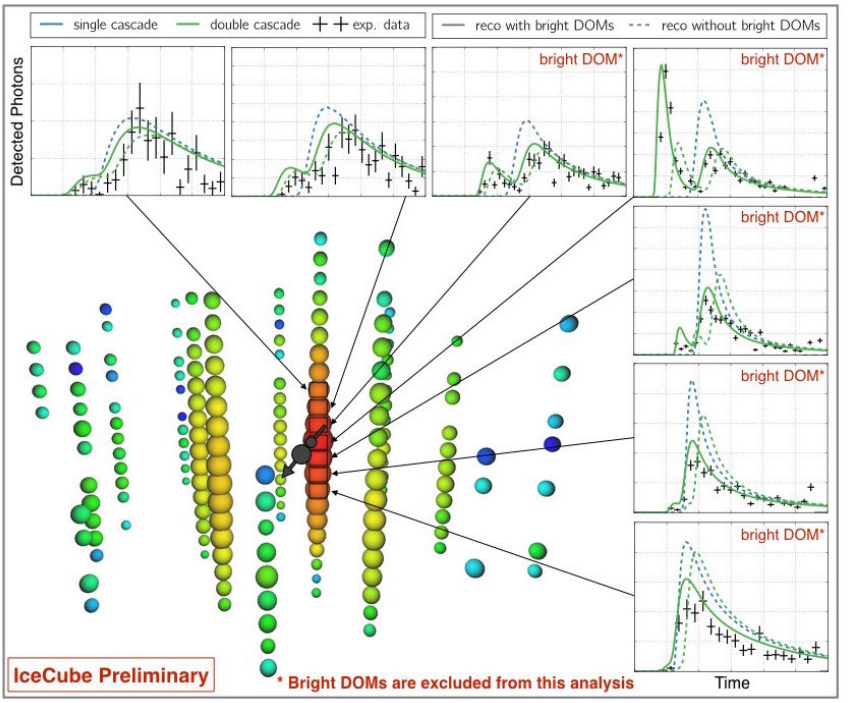


Fig. 8. Event view of a tau neutrino.[24] The Cherenkov photons associated with the production and subsequent decay of the tau neutrino are identified by the double-peaked photon count as a function of time for the bright DOMs, for instance, the one shown in the top-right corner. The best fit (solid line) corresponds to a 17 m decay length and is far superior to fits assuming a single electromagnetic or hadronic shower (dashed lines).

universe. Photons are inevitably produced in association with neutrinos when accelerated cosmic rays produce both neutral and charged pions in interactions with target material in the vicinity of the accelerator. While neutral pions decay into two gamma rays, $\pi^0 \to \gamma + \gamma$, the charged pions decay into three high-energy neutrinos (ν) and antineutrinos ($\bar{\nu}$) via the decay chain $\pi^+ \to \mu^+ + \nu_\mu$ followed by $\mu^+ \to e^+ + \bar{\nu}_\mu + \nu_e$ and the charged-conjugate process. On average, the four final state leptons equally share the energy of the charged pion. With these approximations, gamma rays and neutrinos carry on average 1/2 and 1/4 of the energy of the parent pion.

Modeling of gamma ray observations generally indicated that the flux in pionic photons would be relatively small compared to a dominant flux of photons radiated by electrons in electromagnetic processes, synchrotron radiation, and inverse Compton scattering. These were generally assumed to dominate the energy production of sources in the Galaxy and beyond. IceCube observations show that this is not the case. To illustrate this point, we show in Fig. 10 the energy fluxes, $E^2 dN/dE$, of neutrinos and gamma rays in the universe. Clearly, the energy flux of cosmic

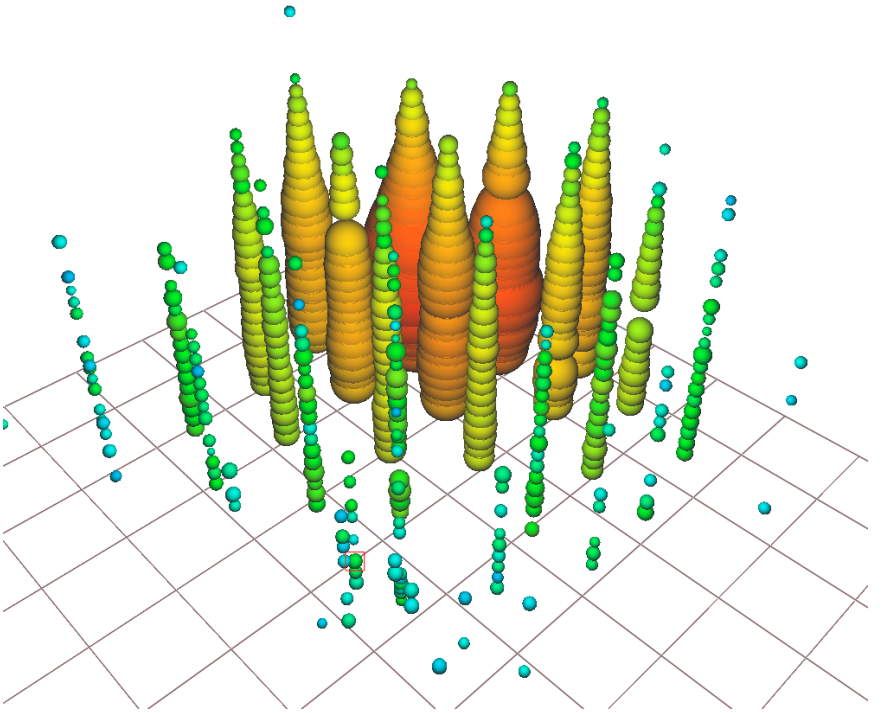


Fig. 9. Particle shower created by the Glashow resonance.[25] Its energy is reconstructed at the resonant energy for the production of a weak intermediate boson W^- in the interaction of an antielectron neutrino with an atomic electron in the ice. The properties of the secondary muons produced in the particle shower are consistent with the hadronic decay of a W^- boson.

neutrinos is comparable to the one for the highest energy gamma rays observed by the NASA Fermi satellite.[16] This already indicates a more prominent role of hadronic processes than routinely anticipated. We will look at this match of the neutrino and gamma-ray energies in more detail.

The neutrino production rate Q_{ν_α} (with typical units $\mathrm{GeV}^{-1}\mathrm{s}^{-1}$ and with the subscript α labeling the neutrino flavor) can be related to the one for charged pions Q_π by

$$\sum_\alpha E_\nu Q_{\nu_\alpha}(E_\nu) \simeq 3\,[E_\pi Q_{\pi^\pm}(E_\pi)]_{E_\pi \simeq 4E_\nu} \ , \tag{2}$$

while, similarly, the production rate of pionic gamma rays is related to the one for neutral pions by

$$E_\gamma Q_\gamma(E_\gamma) \simeq 2\,[E_\pi Q_{\pi^0}(E_\pi)]_{E_\pi \simeq 2E_\gamma} \ . \tag{3}$$

Pion production in the interactions of cosmic rays with photon fields proceeds resonantly via the processes $p + \gamma \to \Delta^+ \to \pi^0 + p$ and $p + \gamma \to \Delta^+ \to \pi^+ + n$.

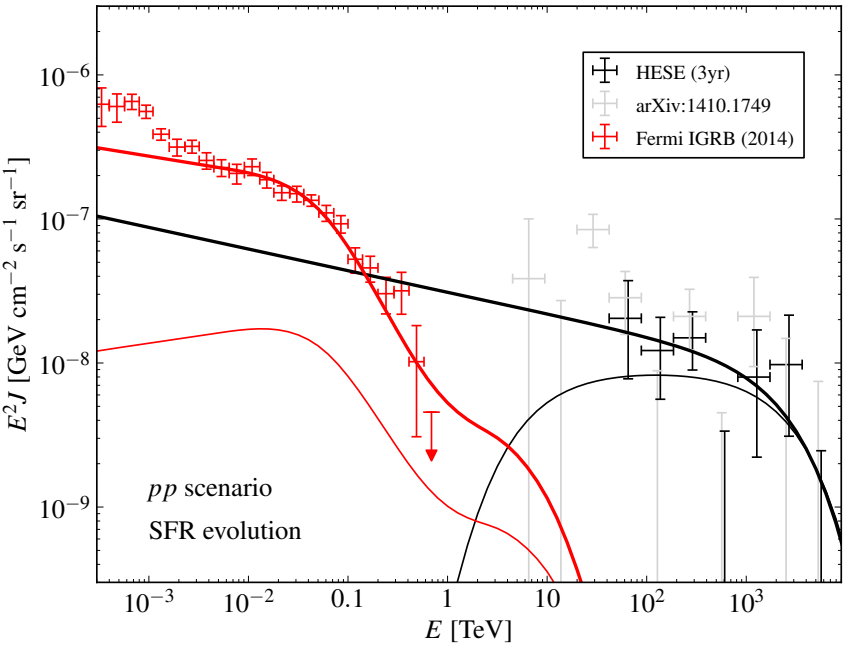


Fig. 10. A calculation illustrating that the photon flux that accompanies the neutrino flux (thick black line) measured by IceCube matches the gamma-ray flux (thick red line) observed by Fermi. We assume a $E^{-1.5}$ energy spectrum, star-formation redshift evolution and, importantly, gamma-ray transparent sources, i.e., pionic photons lose energy in the EBL but not in their sources. The black data points are early IceCube measurements.[27,28] The result suggests that the decay products of neutral and charged pions from pp interactions may be significant components of the nonthermal radiation in the extreme universe.[29] (Introducing the cutoff on the high-energy flux, shown in the figure, does not affect the result.) The calculation is compared to an identical calculation adopting a spectral shape characteristic for the production of cosmic neutrinos on a gamma-ray target in the source (thin lines). While the pionic gamma-ray energy flux is now suppressed relative to the Fermi observations, the neutrino energy spectrum underestimates the IceCube observations and the conclusion that the sources are likely obscured is recovered after correct normalization to the most up to date measurements.[29]

These channels produce charged and neutral pions with probabilities 2/3 and 1/3, respectively. However, the additional contribution of nonresonant pion production changes this ratio to approximately 1/2 and 1/2. In contrast, cosmic rays interacting with matter produce equal numbers of pions of all three charges: $p + p \to n_\pi\,[\,\pi^0 + \pi^+ + \pi^-] + X$, where n_π is the pion multiplicity. We thus obtain a charge ratio $K_\pi = n_{\pi^\pm}/n_{\pi^0} \simeq 2$ and 1 for pp and $p\gamma$ interactions, respectively.

Eqs. 2 and 3 can now be combined to obtain a direct relation between the gamma-ray and neutrino production rates:

$$\frac{1}{3}\sum_\alpha E_\nu^2 Q_{\nu_\alpha}(E_\nu) \simeq \frac{K_\pi}{4}\left[E_\gamma^2 Q_\gamma(E_\gamma)\right]_{E_\gamma = 2E_\nu}\,, \tag{4}$$

where the factor 1/4 accounts for the fact that two gamma rays are produced in the neutral pion decay with twice the energy of the accompanying neutrino, $\langle E_\nu \rangle / \langle E_\gamma \rangle \simeq 1/2$. Note that the relative production rate of gamma rays and neutrinos only depends on the ratio of charged-to-neutral pions produced without any reference to the cosmic-ray beam that initiates their production in the target. This powerful relation follows from the fact that pion production conserves isospin and nothing else.

Before applying this relation to data, one must recall that the universe is transparent to extragalactic neutrinos but not to the accompanying gamma rays. These will interact with microwave photons and other components of the extragalactic background light (EBL) to initiate an electromagnetic cascade that reaches Earth in the form of multiple photons of lower energy. The electromagnetic shower subdivides the initial PeV photon energy into multiple photons with GeV to TeV energies by the time it reaches Earth.[30,31] If the source itself is opaque to gamma rays, the high-energy gamma rays will lose energy even before reaching the EBL to possibly emerge at Earth below the threshold of Fermi, at MeV energies and below.

In order to underscore the power of the multimessenger connection between photons and neutrinos, we first calculate the gamma-ray flux accompanying the diffuse cosmic neutrino flux observed by IceCube, which we describe by a power law with spectral index of -2.15, consistent with the neutrino data above an energy of 100 TeV. The result is shown in Fig. 10 assuming gamma-ray transparent sources and equal multiplicities of all three pion charges, i.e., $K_\pi = 2$. The cascaded gamma-ray energy flux resulting from the pionic photons accompanying the neutrino flux matches the energy flux of extragalactic gamma rays measured by Fermi.

Clearly, in this illustration, the slope and overall normalization of the neutrino spectrum have been adjusted to not exceed the isotropic extragalactic gamma-ray background observed by the Fermi satellite. We conclude that the high-energy cosmic neutrino flux above 100 TeV shown in Fig. 10 saturates the Fermi measurement for the highest photon energies; higher normalization and larger spectral index of the neutrino flux will result in a gamma-ray flux that exceeds the Fermi observations. Fitting the IceCube data with a $E^{-2.5}$ spectral index, closer to the present observations, results in larger neutrino energy fluxes at energies below 100 TeV for both neutrinos and their accompanying photons. After cascading in the EBL, the latter exceeds the Fermi observations. There is no conflict here; in this case, we conclude that the assumption that the sources themselves are transparent to photons is untenable. The resolution is as mentioned above, that the photons lose energy in the source even before entering the EBL and, as a result, reach Earth with energies that are below the detection threshold of the Fermi satellite, at MeV energy or below.

Alternatively, the target for producing the neutrinos may be photons. This changes the value of K_π and, more importantly, the shape of the energy spectrum; for a detailed discussion, see, for instance, Ref. 32. Yielding an energy spectrum that peaks near PeV energies, as shown in Fig. 10, the contribution to the Fermi flux is suppressed at lower energies relative to the power law assumed in Fig. 10. However,

as was the case for *pp* interactions, fits that do not exceed the Fermi data tend not to accommodate the cosmic neutrino spectrum below 100 TeV. This is illustrated in Fig. 10 where we obtain a neutrino spectrum well below the Fermi observations at the price of not fitting the overall normalization of the neutrino data. If sources of the TeV-PeV neutrinos are transparent to gamma rays with respect to two-photon annihilation, tensions with the isotropic diffuse gamma-ray background measured by Fermi seem unavoidable, independently of the production mechanism.[29]

The conclusion is inescapable that the energy fluxes of neutrinos and gamma rays in the extreme universe are qualitatively the same. Furthermore, the IceCube observations point at contributions to the diffuse flux from gamma-obscured sources.

We therefore anticipate that multimessenger studies of gamma-ray and neutrino data will be a powerful tool to identify and study the cosmic ray accelerators that produce cosmic neutrinos. Accordingly, IceCube has developed methods, most promising among them the real-time multiwavelength observations with astronomical telescopes, to identify the sources and build on the discovery of cosmic neutrinos to launch a new era in astronomy.[33,34]

An important lesson for multiwavelength astronomy is that strong high-energy gamma-ray emitters may not be the best candidate neutrino sources. For instance, IceCube does not observe neutrinos from gamma-ray bursts, but this result covers a sample of bursts that are strong gamma-ray emitters.[35] A handful of bursts have been identified that are obscured in gamma rays and may instead provide the key to detecting a neutrino signal. The strongest sources in the 10-year IceCube sky map, the galaxies NGC 1068 and TXS 0506+O56, are gamma-ray-obscured sources at least at the time that they emitted neutrinos. We will discuss this next.

Interestingly, the common energy density of photons and neutrinos is also comparable to that of the ultra-high-energy extragalactic cosmic rays.[36]

4. Identifying Neutrino Sources: The Supermassive Black Hole TXS 0506+056

Phenomenological studies[37,38] and recent data analyses[39–41] have converged on the fact that Fermi's extragalactic gamma-ray flux shown in Fig. 10 is dominated by blazars, AGN with jets pointing at Earth. It is tempting to conclude, based on the matching energy fluxes of photons and neutrinos discussed in the previous section, that the unidentified neutrino sources contributing to the diffuse neutrino flux have already been observed as strong gamma-ray emitters. This is not the case. A dedicated IceCube study[42] correlating the arrival directions of cosmic neutrinos with Fermi blazars shows no evidence of neutrino emission from these sources. The limit leaves room for a contribution of Fermi blazars to the diffuse cosmic neutrino flux below the 10% level. Surprisingly, the multimessenger campaign launched by the neutrino alert IC-170922A[2] identified the first source of cosmic neutrinos as a Fermi "blazar." We will discuss how the multiwavelength data shed light on the apparent contradiction.

Since 2016, the IceCube multimessenger program has grown from issuing Galactic supernova alerts[43] and matching neutrinos with early LIGO/Virgo gravitational wave candidates to a steadily expanding set of automatic filters that select in real time rare, very high energy neutrino events that are likely to be cosmic in origin.[44] Within less than one minute of stopping in the instrumented Antarctic ice, the arrival directions of the neutrinos are reconstructed and automatically sent to the Gamma-ray Coordinate Network for potential follow-up by astronomical telescopes.

4.1. *Observation of a Cosmic Neutrino Source: TXS 0506+056*

On September 22, 2017, the tenth such alert, IceCube-170922A,[45] reported a well-reconstructed muon that deposited 180 TeV inside the detector, corresponding to a most probable energy of the parent neutrino of 290 TeV. Its arrival direction was aligned with the coordinates of a known Fermi blazar, TXS 0506+056, to within 0.06°. The source was "flaring" with a gamma-ray flux that had increased by a factor of seven in recent months. A variety of estimates converged on a probability on the order of 10^{-3} that the coincidence was accidental. The identification of the neutrino with the source reached the level of evidence, but not more. What clinched the association was a series of subsequent observations, culminating with the optical observation of the source switching from an "off" to an "on" state two hours after the emission of IC-170922A, conclusively associating the neutrino with TXS 0506+056.[46] The sequence of observations can be summarized as follows:

- The redshift of the host galaxy, a known blazar, was measured to be $z \simeq 0.34$.[47] It is important to realize that nearby blazars like the Markarian sources are at a distance that is ten times closer, and therefore TXS 0506+056, with a similar flux despite the greater distance, is one of the most luminous sources in the universe. This suggests that it belongs to a special class of sources that accelerate proton beams in dense environments, revealed by the neutrino. That the source is special eliminates any conflict between its observation and the lack of correlation between the arrival directions of IceCube neutrinos and the bulk of the blazars observed by Fermi.[42] Such limits implicitly assume that all sources in an astronomical category are identical, and this is a strong, unstated assumption as underscored by this observation.
- Originally detected by NASA's Swift[48] and Fermi[49] satellites, the alert was followed up by ground-based air Cherenkov telescopes.[50] MAGIC detected the emission of gamma rays with energies exceeding 100 GeV starting several days after the observation of the neutrino.[51] Given its distance, this establishes the source as a relatively rare TeV blazar.
- Informed where to look, IceCube searched its archival neutrino data up to and including October 2017 for evidence of neutrino emission at the location of TXS 0506+056.[3] When searching the sky for point sources of

neutrinos, two analyses have been routinely performed: one that searches for steady emission of neutrinos and one that searches for flares over a variety of timescales. Evidence was found for 19 high-energy neutrino events on a background of fewer than 6 in a burst lasting 110 days. This burst dominates the integrated flux from the source over the last 9.5 years of archival IceCube data, leaving the 2017 flare as a second subdominant feature. We note that this analysis applied a published prescription to data; the chance that this observation is a fluctuation is small.

- Radio interferometric images[52,53] of the source revealed a jet that loses its tight collimation beyond 5 milliarcseconds running into material or intense radiation fields that are likely to be the target for producing the neutrinos. The nature of the target is still a matter of debate. Speculations include the merger with another galaxy that may supply plenty of material to interact with the jet of the dominant galaxy. Alternatively, the jet may interact with the dense molecular clouds of a star-forming region or simply with supermassive stars in the central region of the host galaxy.[52,53] Also, in a so-called structured jet, the accelerated protons may catch up and collide with a slower moving and denser region of jetted photons. Additionally, the VLBA data reveal that the neutrino burst occurs at the peak of enhanced radio emission at 15 GHz, which started five years ago; see Fig. 11. The radio flare may be a signature of a galaxy merger; correlations of radio bursts with the process of merging supermassive black holes have been anticipated.[54]
- The MASTER robotic optical telescope network has been monitoring the source since 2005 and detected its strongest time variation in the last 15 years to occur two hours after the emission of IC170922, with a second variation following the 2014-15 burst.[46] The blazar switches from the "off" to the "on" state two hours after the emission of the neutrino. After an episode of monitoring the uniformity of their observations of the source in the first quarter of 2020, they argue that the time variation detected on September 22, 2017 conclusively associates the source with the neutrino.[46]

Additionally, it is important to note the fact that the high-energy photon and neutrino spectra covering the 2014 burst are consistent with a hard E^{-2} spectrum, which is expected for a cosmic accelerator. In fact, the gamma-ray spectrum shows a hint of flattening beyond E^{-2} during the 110-day period of the 2014 burst.[55,56]

In summary, both the multiwavelength campaign[2] and the observation of an earlier burst of the same source in archival neutrino data provide statistically independent[3] evidence for the association of neutrinos with TXS 0506+056 at the 3 and 3.7σ level, respectively. The significances contributed by the optical and TeV associations on timescales of hours and days are not taken into account, and it is not straightforward to do so because of their a posteriori nature. We conclude, however, that the association of neutrinos with the source summarized above is compelling.

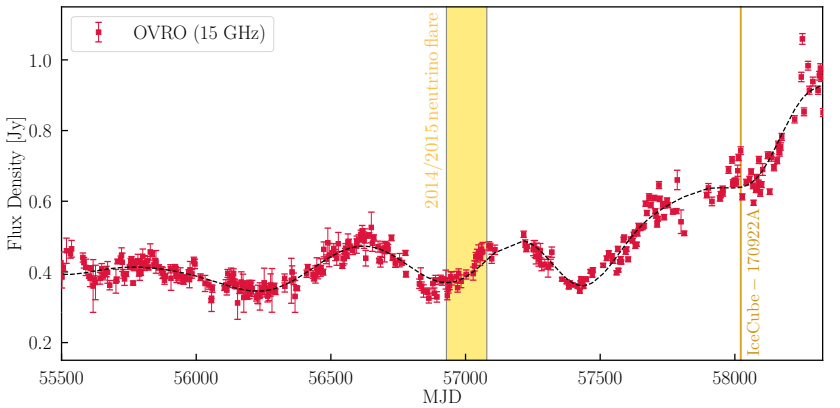


Fig. 11. TXS 0506+056 radio light curve from Owen Valley Radio Observatory (OVRO) at 15 GHz (red). The dashed line illustrates the pattern of the radio flux density. The 2014/15 110-day neutrino flare (yellow band) and the IceCube-170922A episodes are shown.

Other IceCube alerts have triggered intriguing observations. Following up on a July 31, 2016, neutrino alert, the AGILE collaboration, which operates an orbiting X-ray and gamma-ray telescope, reported a day-long blazar flare in the direction of the neutrino one day before the neutrino detection.[57] A tentative but very intriguing association of an IceCube alert[58] has been made with a tidal disruption event, an anticipated source of high-energy neutrino emission. Even before IceCube issued automatic alerts, in April 2016, the TANAMI collaboration argued for the association of the highest energy IceCube event at the time, dubbed "Big Bird," with the flaring blazar PKS B1424-418.[59] Interestingly, the event was produced at a minimum of the Fermi flux,[5] indicating a gamma-ray-obscured source. AMANDA, IceCube's predecessor, observed[60] three neutrinos in coincidence with a rare flare of the blazar 1ES 1959+650 detected by the Whipple telescope in 2002.[61] However, none of these identifications reach the significance of the observations triggered by IC-170922A.

4.2. *The Blueprint of the TXS 0506+056 Beam Dump?*

Blazars are accelerators with their jets pointing at Earth. The gamma-ray community has developed a routine procedure for modeling their spectrum with two components: a lower energy component produced by synchrotron radiation by the electron beam and a high-energy component resulting from the inverse Compton scattering of (possibly the same) photons by accelerated electrons; for a recent discussion, see Ref. 62. Such a source cannot accommodate the TXS observations for two reasons: an electron beam does not produce pions that decay into neutrinos, and, even with the presence of protons in the beam, a target is required to produce the parent pions. It is also evident that a source that emits high-energy gamma rays is transparent to $\gamma\gamma$ absorption and unlikely to host the target material to

produce neutrinos. The opacity for $\gamma\gamma$ interactions to absorb photons is typically two orders of magnitude larger than the one for $p\gamma$ interactions to produce pions and neutrinos.[63]

With the opacity to photons about two orders of magnitude larger, the target producing the neutrinos is unlikely to be transparent to TeV photons, only to photons with tens of GeV energy, and that is indeed what is observed by Fermi at the time of the 2014 flare. There is also evidence that, temporarily, TXS 0506+056 was a gamma-ray-obscured source at the time the IC-170922A neutrino was emitted. The optical observations show a dramatic transition of the blazar from the "off" to the "on" state two hours after the emission of the neutrino, resulting additionally in the doubling of its total optical luminosity.[46] Also at the time of IC-170922 the ground-based atmospheric gamma-ray telescopes observed rapid variations in the flux around the time of the neutrino emission,[2] with the gamma-ray emission observed by MAGIC only emerging after several days.[51]

A more direct indication that strong neutrino emitters are gamma-ray-obscured is indicated by a more recent alert, IC-190730A, sent by IceCube on July 30, 2019. A well-reconstructed 300-TeV muon neutrino was observed in spatial coincidence with the blazar PKS 1502+106.[64] With a reconstructed energy just exceeding that of IC-170922A, it is the highest energy neutrino alert so far. OVRO radio observations[65] show that the neutrino is coincident with the peak flux density of a flare at 15 GHz that started five years prior,[66] matching the similar long-term radio outburst of TXS 0506+056 at the time of IC-170922A; see Fig. 11. Even more intriguing is the fact that the gamma-ray flux observed by Fermi shows a clear minimum at the time that the neutrino is emitted; see Fig. 12. We infer that at this time the jet meets the target that produces the neutrino. Inevitably, the accompanying high-energy gamma rays will be absorbed and the bulk of their electromagnetic energy will cascade down to energies below the Fermi threshold, i.e., MeV or below. Accumulating evidence indicates that cosmic neutrinos are produced by temporarily gamma-suppressed blazars or some other category of AGN.

The coincidence of the two highest energy alerts to date with extended periods of radio emission has led to the speculation that all,[67] or at least some,[68] cosmic neutrinos originate in radio-bright galaxies. The significance of the correlation and its physical origin, if any,[69] are debated.

We suggest that neutrino sources are special, with properties unlikely to match any astronomical classification; astronomical observations catalog accelerators, not neutrino sources. If the neutrino flux observed from TXS 0506+056 over the last decade were typical of all blazars, these would overproduce the total diffuse flux observed by IceCube by well over an order of magnitude.[63] The diffuse neutrino flux limits the density of such sources in the universe to $\sim 5 \times 10^{-10}\,\mathrm{per\,Mpc^3}$, i.e., about 5% of all blazars, and more likely an even smaller fraction of all AGN.[4] Attempts by IceCube and others to find a correlation between the directions of high-energy neutrinos and *all* Fermi blazars must inevitably be unsuccessful, as was the case in Ref. 42.

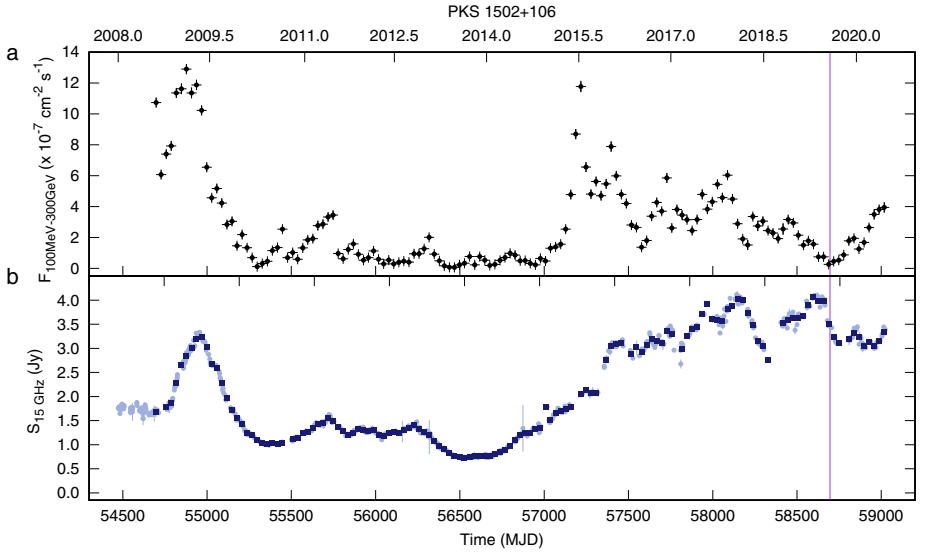


Fig. 12. Temporal variation of the γ-ray and radio brightness of PKS 1502+106.[5] **Top Panel:** Fermi-LAT likelihood light curve integrated between 100 MeV and 300 GeV (marked by black dots with error bars). **Bottom Panel:** OVRO flux density curve of PKS 1502+106 plotted with light blue dots, which is superimposed by the radio flux density curve binned to the Fermi-LAT light curve (marked with dark blue squares). The detection time of the neutrino IC-190730A is labeled by a vertical purple line.

The nature of this special class of sources has not been settled. One straightforward explanation could be that a subclass of AGN, selected by redshift evolution, are powerful proton accelerators producing neutrinos in the past and no longer active today. This assumption accommodates the relatively large redshift of TXS 0506+056, which would be the closest among a set of sources that only accelerated cosmic rays at early redshifts.[70]

Alternatively, in merging galaxies there is plenty of material for accelerated cosmic rays to interact with the jet of the dominant galaxy. Merger activity in active galaxies in not uncommon. The fresh material provides optically thick environments and allows for rapid variation of the Lorentz factors. A cursory review of the literature on the production of neutrinos in galaxy mergers is sufficient to conclude that they can indeed accommodate the observations of both the individual sources discussed above and the total flux of cosmic neutrinos.[71–73] Besides mergers, some form of structured jet where the accelerated protons collide with a slower moving and denser region of jetted photons is a possibility. The jet could also interact with dense molecular clouds of a star-forming region or simply with supermassive stars in the central region of the host galaxy.[52,53]

The observation that the energy flux in neutrinos and very high energy cosmic rays are similar supports the fact that cosmic rays must be highly efficient at

producing neutrinos, requiring a large target density that renders them opaque to high-energy gamma rays. A consistent picture emerges with the source opacity $\tau_{p\gamma}$ exceeding a value of 0.8,[4,63] resulting in a gamma-ray cascade where photons lose energy in the source before cascading to yet lower energies in the extragalactic background light. Some of their energy emerges below the Fermi threshold by the time they reach Earth. This is consistent with the discussion in the previous section that the multimessenger relation between neutrinos and gamma rays points at obscured sources.

We previously mentioned the evidence emerging from 10 years of IceCube data that the arrival directions of cosmic neutrinos are no longer isotropic.[74] The anisotropy results from four sources—TXS 0506+056 among them—that show evidence for clustering above the 4σ level (pretrial). The strongest of these sources is the nearby active galaxy NGC 1068. There is evidence for shocks near the core and for molecular clouds with column density reaching $\sim 10^{25}\,\mathrm{cm}^{-2}$.[75] Similar to TSX 0506+056, a merger onto the black hole is observed—either with a satellite galaxy or, more likely, with a star-forming region[76] accounting for the molecular clouds. This major accretion event may be the origin of the increased neutrino emission.

Although it is obviously challenging to provide a final conclusion on the origin of neutrinos and cosmic rays, we should not lose sight of the fact that high-energy neutrino astronomy exists and that IceCube has demonstrated it has the tools to reveal the extreme universe with more data, or, more realistically, with a larger detector.

5. From Discovery to Astronomy: Larger Telescopes with Better Angular Resolution

Neutrino astronomy has reached a stage that reminds us of the time when Trevor Weekes's 11-meter ground-based Cherenkov telescope had established one convincing TeV source: the Crab. Even today we do not know how the Crab produces gamma rays that track the pulsations of the pulsar at TeV energy. We have to take the lead on TeV astronomy: building more and better telescopes.

Following the pioneering work of DUMAND,[77] several neutrino telescope projects were initiated in the Mediterranean Sea and in Lake Baikal in the 1990s.[78–81] In 2008, the construction of the ANTARES detector off the coast of France was completed. It demonstrated the feasibility of neutrino detection in the deep sea and has provided a wealth of technical experience and design solutions for deep-sea components. An international collaboration has started construction of a multi-cubic-kilometer neutrino telescope in the Mediterranean Sea, KM3NeT.[82] KM3NeT in its second phase[82] will consist of two units for astrophysical neutrino observations, each consisting of 115 strings carrying more than 2,000 optical modules. Since April 2021, six are operational and taking data.[83]

A parallel effort is underway in Lake Baikal with the construction of the deep underwater neutrino telescope Baikal-GVD (Gigaton Volume Detector).[84] The first

GVD cluster was upgraded in the spring of 2016 to its final size: 288 optical modules, a geometry of 120 meters in diameter and 525 meters high, and an instrumented volume of 6 Mton. Each of the eight strings consists of three sections with 12 optical modules. At this time, 7 of the 14 clusters have been deployed, reaching a sensitivity close to the diffuse cosmic neutrino flux observed by IceCube.

IceCube itself is deploying seven new strings at the bottom of the detector array that have been designed as an incremental extension of the DeepCore detector and as a test bed for the technologies of a next-generation detector. The new instrumentation will dramatically boost IceCube's performance at the lowest energies, increasing the samples of atmospheric neutrinos by a factor of ten. New calibration devices will advance our understanding of the response of the light sensors in both current and new strings, resulting in improved reconstructions of cascade events, better identification of tau neutrinos, and an enhanced pointing resolution of muon neutrinos that could approach the 0.1 degree level for the highest energy events of cosmic origin. The improved calibration of the existing sensors will also enable a reanalysis of more than ten years of archival data and significantly increase the discovery potential for neutrino sources before the construction of a second-generation instrument.

Further progress requires a larger instrument. Therefore, as a next step, IceCube proposes to instrument $10\,\mathrm{km}^3$ of glacial ice at the South Pole, capitalizing on the large absorption length of light in ice to thereby increase IceCube's sensitive volume by an order of magnitude.[85] This large gain is made possible by the unique optical properties of the Antarctic glacier revealed by the construction of IceCube. Exploiting the extremely long photon absorption lengths in the deep Antarctic ice, the spacing between strings of light sensors will be increased from 125 to close to 250 meters without significant loss of performance of the instrument at TeV energies and above. The instrumented volume can therefore grow by one order of magnitude while keeping the instrumentation and its budget at the level of the current IceCube detector. The new facility will increase the rates of cosmic events from hundreds to thousands over several years. The superior angular resolution of the longer muon tracks will allow for the discovery of cosmic neutrino sources, currently seen at the $\sim 3\sigma$ level in the 10-year sky map; see Fig. 5.

Acknowledgements

Discussion with collaborators inside and outside the IceCube Collaboration, too many to be listed, have greatly shaped this presentation. Thanks. I would like to single out Markus Ahlers, Qinrui Liu and Ali Kheirandish for contributing to aspects of this manuscript. This research was supported in part by the U.S. National Science Foundation under grants PLR-1600823 and PHY-1913607 and by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation.

References

1. IceCube Collaboration (M. G. Aartsen *et al.*), *Science* **342** (2013) 1242856, arXiv:1311.5238 [astro-ph.HE].
2. Liverpool Telescope, Swift/NuSTAR, MAGIC, H.E.S.S., AGILE, Kiso, VLA/17B-403, INTEGRAL, Kapteyn, Subaru, HAWC, Fermi-LAT, ASAS-SN, VERITAS, Kanata, IceCube Collaboration (M. G. Aartsen *et al.*), *Science* **361** (2018) eaat1378, arXiv:1807.08816 [astro-ph.HE].
3. IceCube Collaboration (M. G. Aartsen *et al.*), *Science* **361** (2018) 147, arXiv:1807.08794 [astro-ph.HE].
4. F. Halzen and A. Kheirandish, *Front. Astron. Space Sci.* **6** (2019) 32.
5. E. Kun, I. Bartos, J. B. Tjus, P. L. Biermann, F. Halzen and G. Mező, *The Astrophysical Journal Letters* **911** (Apr 2021) L18.
6. IceCube Collaboration (M. G. Aartsen *et al.*), *JINST* **12** (2017) P03012, arXiv:1612.05093 [astro-ph.IM].
7. F. Halzen, *Eur. Phys. J.* **C46** (2006) 669, arXiv:astro-ph/0602132 [astro-ph].
8. IceCube Collaboration Collaboration (M. Aartsen *et al.*), *JINST* **9** (2014) P03009, arXiv:1311.4767 [physics.ins-det].
9. T. Yuan, *TeV Particle Astrophysics 2017 (TeVPA2017), Columbus, Ohio, USA* (2017).
10. IceCube Collaboration (M. G. Aartsen *et al.*), *Phys. Rev. Lett.* **115** (2015) 081102, arXiv:1507.04005 [astro-ph.HE].
11. IceCube Collaboration (M. G. Aartsen *et al.*), *Astrophys. J.* **833** (2016) 3, arXiv:1607.08006 [astro-ph.HE].
12. IceCube Collaboration (M. G. Aartsen *et al.*), *PoS* **ICRC2017** (2017) arXiv:1710.01191 [astro-ph.HE].
13. IceCube Collaboration (J. Stettner), *PoS* **ICRC2019** (2020) 1017, arXiv:1908.09551 [astro-ph.HE].
14. J. B. Stettner, Measurement of the energy spectrum of astrophysical muon-neutrinos with the IceCube Observatory, dissertation, RWTH Aachen University (2021). pp. 1 Online–Ressource : Illustrationen, Diagramme. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2021.
15. IceCube Collaboration (M. G. Aartsen *et al.*), *Phys. Rev. Lett.* **124** (2020) 051103, arXiv:1910.08488 [astro-ph.HE].
16. Fermi-LAT Collaboration (M. Ackermann *et al.*), *Astrophys. J.* **799** (2015) 86, arXiv:1410.3696 [astro-ph.HE].
17. K. Fang and K. Murase, *Nature Phys.* **14** (2018) 396, arXiv:1704.00015 [astro-ph.HE], [Nature Phys.14,no.4,396(2018)].
18. IceCube Collaboration Collaboration (M. Aartsen *et al.*), *Phys.Rev.Lett.* **111** (2013) 021103, arXiv:1304.5356 [astro-ph.HE].
19. R. Abbasi, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, M. Ahrens, C. Alispach, A. Alves, N. Amin, K. Andeen and et al., *Physical Review D* **104** (Jul 2021).
20. IceCube Collaboration Collaboration (M. Aartsen *et al.*), *Phys.Rev.* **D89** (2014) 102001, arXiv:1312.0104 [astro-ph.HE].
21. IceCube Collaboration (M. G. Aartsen *et al.*), *Phys. Rev. Lett.* **114** (2015) 171102, arXiv:1502.03376 [astro-ph.HE].
22. IceCube Collaboration (M. G. Aartsen *et al.*), *Phys. Rev. Lett.* **125** (2020) 121104, arXiv:2001.09520 [astro-ph.HE].
23. J. G. Learned and S. Pakvasa, *Astropart.Phys.* **3** (1995) 267, arXiv:hep-ph/9405296 [hep-ph].

24. IceCube Collaboration (R. Abbasi *et al.*), *to be published in Phys. Rev. Lett.* (11 2020) arXiv:2011.03561 [hep-ex].
25. IceCube Collaboration (M. G. Aartsen *et al.*), *Nature* **591** (2021) 220.
26. A. Albert, M. André, M. Anghinolfi, G. Anton, M. Ardid, J.-J. Aubert, J. Aublin, T. Avgitas, B. Baret, J. Barrios-Martí and et al., *The Astrophysical Journal* **853** (Jan 2018) L7.
27. IceCube Collaboration (M. Aartsen *et al.*), *Phys.Rev.Lett.* **113** (2014) 101101, arXiv:1405.5303 [astro-ph.HE].
28. IceCube Collaboration (M. G. Aartsen *et al.*), *Phys. Rev.* **D91** (2015) 022001, arXiv:1410.1749 [astro-ph.HE].
29. K. Murase, M. Ahlers and B. C. Lacki, *Phys.Rev.* **D88** (2013) 121301, arXiv:1306.3417 [astro-ph.HE].
30. R. J. Protheroe and T. Stanev, *Monthly Notices of the Royal Astronomical Society* **264** (09 1993) 191.
31. M. Ahlers, L. Anchordoqui, M. Gonzalez-Garcia, F. Halzen and S. Sarkar, *Astropart.Phys.* **34** (2010) 106, arXiv:1005.2620 [astro-ph.HE].
32. S. Yoshida and K. Murase, *Physical Review D* **102** (Oct 2020).
33. IceCube, MAGIC, VERITAS Collaboration (M. G. Aartsen *et al.*), *JINST* **11** (2016) P11009, arXiv:1610.01814 [hep-ex].
34. IceCube Collaboration (M. G. Aartsen *et al.*), *Astropart. Phys.* **92** (2017) 30, arXiv:1612.06028 [astro-ph.HE].
35. IceCube Collaboration (M. G. Aartsen *et al.*), *Astrophys. J.* **843** (2017) 112, arXiv:1702.06868 [astro-ph.HE].
36. Pierre Auger Collaboration, A. Aab *et al.*, The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015), in *Proceedings, 34th International Cosmic Ray Conference (ICRC 2015): The Hague, The Netherlands, July 30-August 6, 2015*, (2015). arXiv:1509.03732 [astro-ph.HE].
37. M. Ajello *et al.*, *Astrophys. J.* **751** (2012) 108, arXiv:1110.3787 [astro-ph.CO].
38. M. Di Mauro, F. Donato, G. Lamanna, D. A. Sanchez and P. D. Serpico, *Astrophys. J.* **786** (2014) 129, arXiv:1311.5708 [astro-ph.HE].
39. Fermi-LAT Collaboration (M. Ackermann *et al.*), *Phys. Rev. Lett.* **116** (2016) 151105, arXiv:1511.00693 [astro-ph.CO].
40. H.-S. Zechlin, A. Cuoco, F. Donato, N. Fornengo and A. Vittino, *Astrophys. J. Suppl.* **225** (2016) 18, arXiv:1512.07190 [astro-ph.HE].
41. M. Lisanti, S. Mishra-Sharma, L. Necib and B. R. Safdi, *Astrophys. J.* **832** (2016) 117, arXiv:1606.04101 [astro-ph.HE].
42. IceCube Collaboration (M. G. Aartsen *et al.*), *Astrophys. J.* **835** (2017) 45, arXiv:1611.03874 [astro-ph.HE].
43. IceCube Collaboration (R. Abbasi *et al.*), *Astron. Astrophys.* **535** (2011) A109, arXiv:1108.0171 [astro-ph.HE], [Erratum: Astron.Astrophys. 563, C1 (2014)].
44. IceCube Collaboration (R. Abbasi *et al.*), *Astrophys. J.* **910** (2021) 4, arXiv:2012.04577 [astro-ph.HE].
45. C. Kopper and E. Blaufuss, *GRB Coordinates Network, Circular Service, No. 21916, #1 (2017)* **21916** (2017).
46. V. M. Lipunov, V. G. Kornilov, K. Zhirkov, E. Gorbovskoy, N. M. Budnev, D. A. H. Buckley, R. Rebolo, M. Serra-Ricart, R. Podesta, N. Tyurina, O. Gress, Y. Sergienko, V. Yurkov, A. Gabovich, P. Balanutsa, I. Gorbunov, D. Vlasenko, F. Balakin, V. Topolev, A. Pozdnyakov, A. Kuznetsov, V. Vladimirov, A. Chasovnikov, D. Kuvshinov, V. Grinshpun, E. Minkina, V. B. Petkov, S. I. Svertilov, C. Lopez, F. Podesta,

H. Levato, A. Tlatov, B. Van Soelen, S. Razzaque and M. Böttcher, *ApJLett* **896** (June 2020) L19, arXiv:2006.04918 [astro-ph.HE].
47. S. Paiano, R. Falomo, A. Treves and R. Scarpa, *Astrophys. J.* **854** (2018) L32, arXiv:1802.01939 [astro-ph.GA].
48. P. A. E. A. Keivani, J. A. Kennea, D. B. Fox, D. F. Cowen, J. P. Osborne, F. E. Marshall and Swift-IceCube Collaboration, *The Astronomer's Telegram* **10792** (September 2017).
49. Y. T. Tanaka, S. Buson and D. Kocevski, *The Astronomer's Telegram* **10791** (September 2017).
50. R. Mirzoyan, *The Astronomer's Telegram* **10817** (October 2017).
51. MAGIC Collaboration (S. Ansoldi *et al.*), *Astrophys. J. Lett.* **863** (2018) L10, arXiv:1807.04300 [astro-ph.HE].
52. S. Britzen, C. Fendt, F. Böttcher, F. Jaron, I. N. Pashchenko, A. Araudo, O. Karas and O. Kurtanidze, *Mem. Soc. Ast. It.* **90** (2019) 77.
53. E. Kun, P. L. Biermann and L. Á. Gergely, *Monthly Notices of the Royal Astronomical Society: Letters* **483** (Nov 2018) L42–L46.
54. L. Á. Gergely and P. L. Biermann, *The Astrophysical Journal* **697** (May 2009) 1621.
55. P. Padovani, P. Giommi, E. Resconi, T. Glauch, B. Arsioli, N. Sahakyan and M. Huber, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 192, arXiv:1807.04461 [astro-ph.HE].
56. S. Garrappa *et al.*, *8th International Fermi Symposium* (2018).
57. F. Lucarelli *et al.*, *Astrophys. J.* **846** (2017) 121, arXiv:1707.08599 [astro-ph.HE].
58. R. Stein, S. v. Velzen, M. Kowalski, A. Franckowiak, S. Gezari, J. C. A. Miller-Jones, S. Frederick, I. Sfaradi, M. F. Bietenholz, A. Horesh and et al., *Nature Astronomy* **5** (Feb 2021) 510–518.
59. M. Kadler, F. Krauß, K. Mannheim, R. Ojha, C. Müller, R. Schulz, G. Anton, W. Baumgartner, T. Beuchert, S. Buson and et al., *Nature Physics* **12** (Apr 2016) 807–814.
60. IceCube Collaboration (E. Bernardini), *Nucl. Instrum. Meth. A* **567** (2006) 418.
61. VERITAS Collaboration (M. K. Daniel *et al.*), *Astrophys. J.* **621** (2005) 181, arXiv:astro-ph/0503085 [astro-ph].
62. J. Biteau, E. Prandini, L. Costamante, M. Lemoine, P. Padovani, E. Pueschel, E. Resconi, F. Tavecchio, A. Taylor and A. Zech, *Nature Astronomy* **4** (January 2020) 124–131.
63. F. Halzen, A. Kheirandish, T. Weisgarber and S. P. Wakely, *Astrophys. J.* **874** (2019) L9, arXiv:1811.07439 [astro-ph.HE].
64. V. Lipunov, E. Gorbovskoy, V. Kornilov, N. Tyurina, F. Balakin, V. Vladimirov, P. Balanutsa, A. Kuznetsov, D. Vlasenko, I. Gorbunov, A. Pozdnyakov, D. Zimnukhov, V. Senik, A. Chasovnikov, V. Grinshpun, T. Pogrosheva, R. Rebolo, M. Serra, R. Podesta, C. Lopez, F. Podesta, C. Francile, H. Levato, D. Buckley, O. Gress, N. M. Budnev, O. Ershova, A. Tlatov, D. Dormidontov, V. Yurkov, A. Gabovich and Y. Sergienko, *The Astronomer's Telegram* **12971** (July 2019) 1.
65. S. Kiehlmann, T. Hovatta, M. Kadler, W. Max-Moerbeck and A. C. S. Readhead, *The Astronomer's Telegram* **12996** (August 2019) 1.
66. V. Karamanavis, L. Fuhrmann, T. P. Krichbaum, E. Angelakis, J. Hodgson, I. Nestoras, I. Myserlis, J. A. Zensus, A. Sievers and S. Ciprini, *Astronomy and Astrophysics* **586** (February 2016) A60, arXiv:1511.01085 [astro-ph.HE].
67. A. V. Plavin, Y. Y. Kovalev, Y. A. Kovalev and S. V. Troitsky, *Astrophys. J.* **908** (2021) 157, arXiv:2009.08914 [astro-ph.HE].
68. T. Hovatta *et al.*, *Astron. Astrophys.* **650** (2021) A83, arXiv:2009.10523 [astro-ph.HE].

69. B. Zhou, M. Kamionkowski and Y.-f. Liang, *Phys. Rev. D* **103** (2021) 123018, arXiv:2103.12813 [astro-ph.HE].
70. A. Neronov and D. Semikoz, *Journal of Experimental and Theoretical Physics* **131** (Aug 2020) 265–272.
71. K. Kashiyama and P. Mészáros, *Astrophys.J.* **790** (2014) L14, arXiv:1405.3262 [astro-ph.HE].
72. C. Yuan, P. Mészáros, K. Murase and D. Jeong, *Astrophys. J.* **857** (2018) 50, arXiv:1712.09754 [astro-ph.HE].
73. C. Yuan, K. Murase and P. Mészáros, *Astrophys. J.* **878** (2019) 76, arXiv:1810.04155 [astro-ph.HE].
74. IceCube Collaboration, T. Carver, Ten years of All-sky Neutrino Point-Source Searches, in *36th International Cosmic Ray Conference (ICRC 2019) Madison, Wisconsin, USA, July 24-August 1, 2019*, (2019). arXiv:1908.05993 [astro-ph.HE].
75. A. Marinucci *et al.*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) L94, arXiv:1511.03503 [astro-ph.HE].
76. S. García-Burillo, F. Combes, A. Usero, S. Aalto, M. Krips, S. Viti, A. Alonso-Herrero, L. K. Hunt, E. Schinnerer, A. J. Baker, F. Boone, V. Casasola, L. Colina, F. Costagliola, A. Eckart, A. Fuente, C. Henkel, A. Labiano, S. Martín, I. Márquez, S. Muller, P. Planesas, C. Ramos Almeida, M. Spaans, L. J. Tacconi and P. P. van der Werf, *Astronomy and Astrophysics* **567** (July 2014) A125, arXiv:1405.7706 [astro-ph.GA].
77. DUMAND Collaboration Collaboration (J. Babson *et al.*), *Phys.Rev.* **D42** (1990) 3613.
78. BAIKAL Collaboration (I. A. Belolaptikov *et al.*), *Astropart. Phys.* **7** (1997) 263.
79. NESTOR Collaboration Collaboration (G. Aggouras *et al.*), *Astropart.Phys.* **23** (2005) 377.
80. ANTARES Collaboration Collaboration (J. Aguilar *et al.*), *Astropart.Phys.* **26** (2006) 314, arXiv:astro-ph/0606229 [astro-ph].
81. E. Migneco, *J.Phys.Conf.Ser.* **136** (2008) 022048.
82. KM3NeT Collaboration (S. Adrian-Martinez *et al.*), *J. Phys.* **G43** (2016) 084001, arXiv:1601.07459 [astro-ph.IM].
83. A. Sinopoulou, R. Coniglione, R. Muller and E. Tzamariudaki, Atmospheric neutrinos with the first detection units of km3net/arca (2021).
84. A. D. Avrorin *et al.*, *Phys. Part. Nucl.* **46** (2015) 211.
85. IceCube-Gen2 Collaboration (M. G. Aartsen *et al.*), *J. Phys. G* **48** (2021) 060501, arXiv:2008.04323 [astro-ph.HE].

The first results of PandaX-4T

Jianglai Liu*

School of Physics & Astronomy and Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai, 200240, China

E-mail: jianglai.liu@sjtu.edu.cn

**On behalf of the PandaX Collaboration*

Dark matter, an invisible substance which constitutes 85% of the matter in the observable universe, is one of the greatest puzzles in physics and astronomy today. Dark matter can be made of a new type of fundamental particle, not yet observed due to its feeble interactions with visible matter. In this talk, we present the first results of PandaX-4T, a 4-ton-scale liquid xenon dark matter observatory, searching for these dark matter particles from deep underground. We will briefly summarize the performance of PandaX-4T, introduces details in the data analysis, and present the latest search results on dark matter-nucleon interactions.

Keywords: Dark Matter; WIMPs; Liquid xenon; PandaX-4T.

1. Introduction

The detection of the weakly-interacting-massive-particles (WIMPs) has been outlined by Goodman and Witten in 1985.[1] Yet, in the direct detection of dark matter, we have not seen convincing signals. At the large mass region (approximately 100 GeV/c^2 and above), the front runners are xenon-based experiments.[2–4] The sensitivities of these experiments, roughly speaking, correspond to a few events per tonne per year.

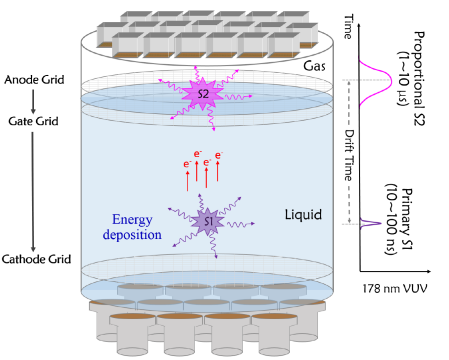


Fig. 1. Working principle of a two-phase TPC.

All of these xenon experiments are based on a tank of liquid xenon cooled at negative one hundred degrees. The detector is the so-called Time Projection

Chamber (TPC), capable of measuring both the energy deposition and its location, utilizing the detection of prompt scintillation photons ($S1$) and delayed electroluminescence signals due to ionized electrons ($S2$). See Fig. 1. The interaction of dark matter and xenon nucleus produces nuclear recoils (NR), and the background gamma rays produce electron recoils (ER). The ratio of $S1$ and $S2$ is a natural discriminant for ER background.

The three dark matter experiments with xenon TPC are LUX-ZEPLIN (LZ), XENONnT and PandaX-4T. LZ is located at the 4850 feet level of the Sanford Underground Research Facility in Lead, SD, containing seven active tonnes of liquid xenon. At the Laboratori Nazionali del Gran Sasso (LNGS) of INFN in Italy, the XENONnT detector is an upgrade to XENON1T, with an active mass of six tonnes. Both LZ and XENONnT are under active preparation. The experiment I am going to focus here, PandaX-4T, with 3.7 tonnes of active mass, is located in China Jinping Underground Laboratory (CJPL), southwest of China, 2400 meters underneath a rock mountain.

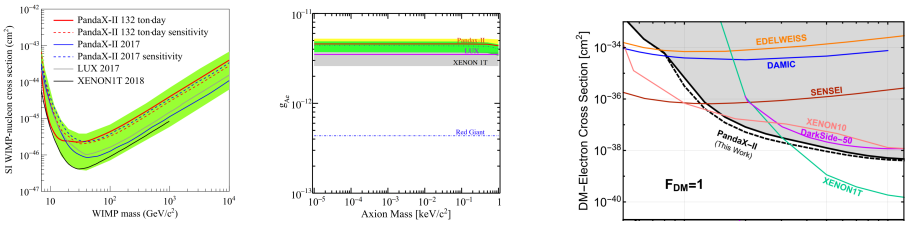
CJPL is the deepest underground laboratory in the world. CJPL-I has a small space and only holds two experiments, PandaX and CDEX (China Dark matter Experiment), to detect dark matter with different methods. CJPL-II has a much expanded lab space, with eight experimental halls, each with a dimension of 65 m long and 14×14 m^2 cross section. Currently, there are three experimental halls in occupation, PandaX, CDEX and JUNA (Jingping Underground lab for Nuclear Astrophysics).

2. The PandaX Experiment

The PandaX collaboration is formed in 2009. Up to now, we have undergone two completed phases. The first phase, PandaX-I, had 120 kg of xenon target. The second, PandaX-II, has 580 kg of xenon target. In particularly, PandaX-II has produced several leading results on the direct detection of dark matter. Since the August of 2019, the new generation of the experiment, PandaX-4T, was officially moved into CJPL-II.

2.1. *Results of PandaX-II*

In 2019, we finished the full exposure of 130 ton·day of PandaX-II. As shown in Fig. 2(a), the data set was increased by a factor of 2.5 in sensitivity for WIMPs,[5] in comparison with the results in 2017.[8] But the upper limit is worsened due to background fluctuation. For ER signals, we are looking directly for the solar axion signal and the anonymous neutrino magnetic moment from the solar neutrinos. In light of the recent XENON1T excess, in this region, the data are contaminated by the tritium background, with limited sensitivity to confirm or refute XENON1T (Fig. 2(b)).[6] Fig. 2(c) gives a new search using an extremely low threshold data by giving up $S1$ and only taking $S2$. This pushes the energy threshold down to 80 eV.



(a) (b) (c)

Fig. 2. (a) The 90% confidence level upper limits vs. WIMP mass for the spin-independent WIMP-nucleon elastic scattering cross sections from the full exposure data of PandaX-II, overlaid with that from PandaX-II 2017, LUX 2017, and XENON1T 2018;[5] (b) A search of solar axion-electron coupling constant g_{Ae} in PandaX-II, in comparison with other experiments;[6] (c) 90% C.L. upper limits (solid: the constant model, dashed: NEST 2.0) on WIMP-electron scattering cross section from PandaX-II data for a point-like interaction.[7]

Therefore, we can look for the direct scatterings of WIMPs and electrons. Because of the low threshold, we have the best sensitivity around 25 MeV/c^2 in WIMP mass for a point-like interaction, which was unexplored previously.[7]

2.2. *Overview of PandaX-4T*

For the newly deployed PandaX-4T experiment, a 900 m^3 water tank is acting as the shielding of the central detector. The detector resides in a platform at the middle of this tank. At the center of the detector, there is a TPC with a 1.2 m height and 1.2 m diameter, holding 3.7 tonnes liquid xenon. On the top and bottom, it is covered by roughly 400 channels of 3-in PMTs.

The preparation of PandaX-4T started in April, 2018, after the permission from the CJPL management. The installation of the entire apparatus was completed on May, 2020. The commissioning run extended from Nov. 28, 2020 to Apr. 16, 2021, of which the data will be reported here.

Most of our instrumentation severing the detector is located in a 10000-class clean room. All the TPC assemblies were carried out in a 1000-class clean room. To avoid the contamination of radon, the air going into the 1000-class clean room were purified by a radon removal system.

3. Performance of PandaX-4T Commissioning

3.1. *Cryogenics*

The cryogenics is very stable during the commissioning. In fact, the heat load of the system is only 50 W, much smaller than the maximum cooling power 580 W by design. We have two circulation loops trying to clean out the electronegative outgassing from the detector with a rate of 40 kg/h. Through out the stable running period, the pressure and the temperature are stable within 0.5% and 0.1 K, respectively.

3.2. *Electronics and DAQ*

We utilize a triggerless electronics system. Every channel is read out separately by a faster digitizer whenever the signal is above 1/3 photonelectron (PE), with a sampling rate of 250 MS/s. The average single photon efficiency is around 95%, and the maximum bandwidth of the DAQ is 450 MB/s.

3.3. *TPC operation conditions*

As illustrated in Fig. 1, between the anode and the gate is a 10 mm amplification region for the ionized electrons. Between the gate and the cathode is the drifting region, with a maximum drift distance of 1.2 m. Below the cathode, we have a 10 cm shielding region to suppress the background from bottom PMTs. During the operations, the HVs on the electrodes were set at a few different values to avoid excessive discharges.

3.4. *Data taking history*

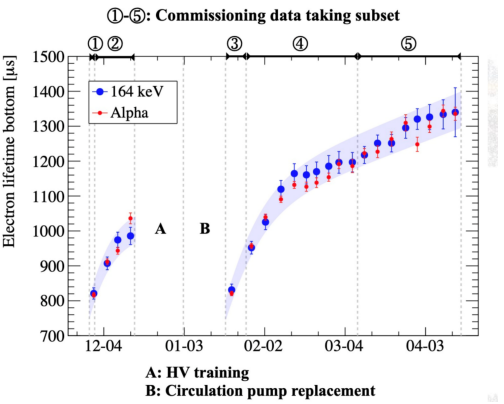


Fig. 3. Electron lifetime for different commissioning data taking subset.

Electronegative impurities tend to attach ionized electrons, leading to an exponential attenuation of the $S2$ signals vs. drift time. The purity of the liquid is characterized by the mean attenuation time, i.e. the electron lifetime. Using internal gamma or alpha peaks, the extracted electron lifetime also gives the vertical correction for $S2$. In total, we have a stable data running period of 95 days, and as shown in Fig. 3, the electron life time is kept around 1000 μs. In comparison, the maximum drift time in the TPC is about 840 μs.

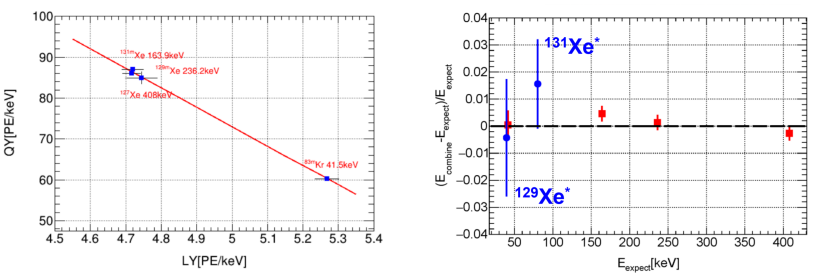
3.5. *Calibration*

Calibration is one of the most important aspects. PandaX-4T adopted three main methods. One is the gas source injection, where we can load in radioactive ^{83m}Kr and ^{220}Rn into the detector. They are short-lived, but uniformly distributed in the detector. Second is the external calibration loops, where a source can go into these loops at three different height of the TPC. Lastly, we have a Deuteron-Deuteron (D-D) fusion generator for neutrons. It produces semi-monoenergetic neutrons, which are then beamed directly into the detector.

- Uniformity correction with ^{83m}Kr. When ^{83m}Kr radioactive sources are loaded into the detector, they uniformly distribute in the detector with a characteristic 41.5 keV energy, based on which the uniformity of $S1$ and $S2$ can be calibrated. For $S2_b$, the RMS variation in the fiducial volume is 12% and for $S1$, the RMS variation 20%. The measured energy resolution is 6.8%.
  For the energy reconstruction, event-by-event we use a linear combination of $S1$ and $S2$ according to Eq. 1.

$$\mathrm{E} = 13.7\mathrm{eV} \times \left(\frac{\mathrm{S1}}{\mathrm{PDE}} + \frac{\mathrm{S2_b}}{\mathrm{EEE} \times \mathrm{SEG}} \right), \tag{1}$$

  in which PDE, EEE and SEG are corresponding to the photon detection efficiency, electron extraction efficiency and single electron gain, respectively.



(a) (b)

Fig. 4. (a) Charge yield vs. light yield in PE/keV and the linear fit. (b) Fractional differences between the reconstructed and expected energy. Blue squares represent uniformly distributed electron recoil peaks, ^{131m}Xe(164 keV), ^{129m}Xe(236 keV),^{127}Xe(408 keV) and ^{83m}Kr(41.5 keV). Blue points in (b) represent de-excitation gammas from ^{129}Xe*(39.6 keV) and ^{131}Xe*(80.2 keV) after neutron activation, with the NR components subtracted from $S1$ and $S2$, which are not used in the linear fit.

SEG can be obtained using the smallest $S2$ signals identified in the data. The uniform distribution to the detector is only 13% for the RMS variation. The top and bottom ratio for these low energy is also very stable. Then with various ER peaks in the data, we can fit for PDE and EEE using Eq. 1, as shown in Fig. 4(a). Just for the check, the reconstruction energy, after applying these two

fitted parameters, shows on Fig. 4(b). The red points have been used in the fit. We use two more independent data points produced by illuminating xenon with neutrons and looking at de-excitation gamma rays. They are also at the zero line, but with a large uncertainty because the de-excitation energy contains some of the nuclear energy component.

- Nuclear recoil calibration with D-D neutrons.

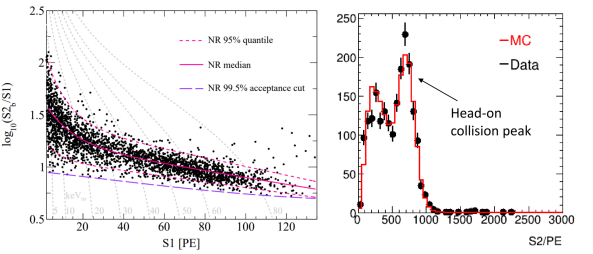


Fig. 5. D-D neutron calibration. The left picture shows $\log_{10} S2/S1$ over $S1$; The right plot gives the one-dimensional projection of $S2$, which is in agreement with MC expectation.

The other aspect of the calibration is to determine the low energy response of the detector, particularly relevant to the dark matter search. If dark matter interacts with xenon, and produces a low energy nuclear recoil, the best calibration should come from the neutrons. Fig. 5 summarizes the neutron recoil calibration with the D-D neutron band. D-D neutron band can generate a monoenergetic neutron around 2.2 MeV. The data set is used in combination with the AmBe data set, another neutron source, to tune the parameters, such as the light yield, charge yield, as well as fluctuations in the signal model.

4. ER calibration with ^{220}Rn

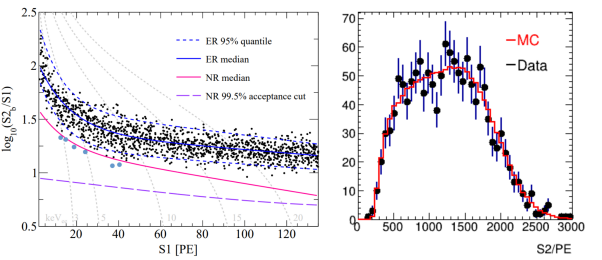


Fig. 6. ^{220}Rn calibration. In the left plot, purple solid line is the medium of neutron calibration. The right plot gives the one-dimensional projection of $S2$, which is in agreement with the MC data.

ER calibration is critical for us to understand the majority of the background. The ER response is calibrated by injecting ^{220}Rn into the detector, which only have 55 s life time. One of the decay progeny ^{212}Pb is a beta emitter, producing a flat energy spectrum at low energy. We characterize the rejection power by simply counting how many events are leaked underneath the nuclear recoil medium line. As shown in Fig. 6, the full rejection power is 0.43%, with roughly 50% uncertainty.

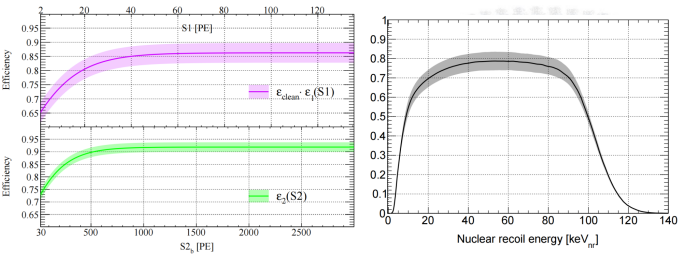


Fig. 7. Selection efficiency. The left picture shows the data quality cut efficiency for $S1$ an $S2_b$; The right plot gives the efficiency vs. nuclear recoil energy.

Using the same electron or nuclear recoil low energy data set, we can determine the efficiency for both $S1$ and $S2$. As illustrated in Fig. 7, at the plateau region, we have an efficiency of around 78%@40 keV.

4.1. *Backgrounds*

The major background components considered in the dark matter analysis are the following:

The first is the material background from the detector. If we bin our detector into several regions, the top, the bottom, the outer and the center, the background component is different. Comparing the expectation from the data, the overall agreement is around 14%. After converting all these high energy events into the low energy in the dark matter region, it is a minor background with 33 events expected.

^{222}Rn background, distributed uniformly into a detector, is important. They can come out of the surface of an emanation material. However, ^{222}Rn can be identified easily through their decay progeny. There are alphas that one can identify, as well as α-β coincidence signals. We find that the rate of radon is about 5 μBq/kg. This number is improved by six times from PandaX-II. The extracted background is estimated to be 347$\pm$190 events.

^{85}Kr, another noble gas, is used to be a major problem for the earlier generation of PandaX. Using a delayed coincidence technique, the krypton level is measured to be 0.33 ppt, improved about twenty times from PandaX-II. The residual ^{85}Kr background is estimated to 53$\pm$34 events.

^{127}Xe normally do not exist in the xenon. It is produced when xenon is exposed too longer on the surface. Through the characterized gamma and X-rays in the electron-capture process, we can track them in the detector. The half live of ^{127}Xe is only 36 days, and will decay soon. The background in the dark matter region is the 5 keV K-shell X-rays. Actually, we can tag these X-rays through the 33 keV L-shell X-rays. The expected events of ^{127}Xe background is estimated to be 9±2.

Tritium background is unfortunately identified in our data. They are likely originated from a calibration at the end of PandaX-II, where we loaded tritium into the detector, trying to calibrate the detector and see how well we can remove them. Unfortunately, the removal is probably inefficient, leaving some residuals in PandaX-4T. In our fit, we left this contaminant to be floated, resulting into about 5×10^{-24} mol/mol concentration, a value that one cannot see with direct assay measurement.

Surface background is related to radon. Radon progeny can attach to the surface of the detector. These events can be clearly identified from the data by looking for αs with abnormally small $S2$. Because they are closed to the PTFE reflector, we can get both the position distribution and the $S2$ distributions to estimate how much of these backgrounds could leak underneath the dark matter search window. The expected background for these wall surface is estimated to be 0.6±0.2.

Neutron background generates directly nuclear recoil. For these neutron background, once they enter the detector, they could multi-scatter, and eventually get capture onto xenon, producing high energy capture signals. Both processes are bench-marked by neutron source calibration. In the end, the neutron background is estimated to be 1.3±0.6.

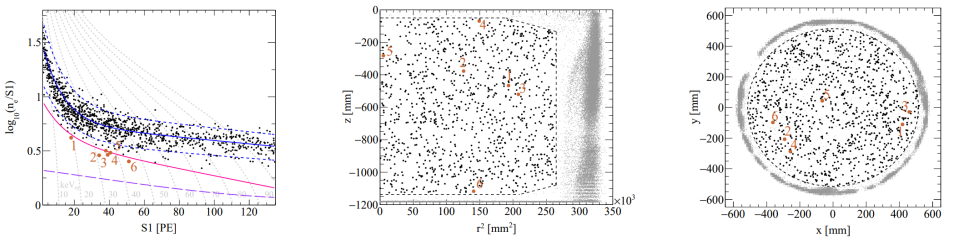
Accidental background is formed by random coincidence between uncorrelated $S1$ and $S2$. Therefore, cuts imposing the diffusion relations observed in data are effective in removing them. The expected accidental background was suppressed by two times in comparison with PandaX-II. It is estimated to be 2.4±0.5.

Fiducial volume determination utilizes the expected background from detector materials, as well as a combination of the radon background and the tritium background that is measured *in-situ*, to maximize a figure-of-merit related to the search sensitivity. The fiducial mass is settled at 2.67 tonnes.

5. Results

Fig. 8 shows the dark matter candidates within the fiducial volume and dark matter selection window. There are 1058 candidates in total. Six events are identified below the NR median curve. Candidates are uniformly distributed in the fiducial volume, which is expected if backgrounds are dominated by the tritium and radon. A fit to the data is consistent with background-only hypothesis, with a goodness-of-fit p-value of 0.7.

Based on the fit, we therefore set the bound on the spin-independent WIMP-nucleon scattering, summarized in Fig. 9. With a 0.63 tonne·year, we have a



(a) $\log_{10}(n_e/S1)$ vs. $S1$ (b) z vs. r^2 (c) y vs. x

Fig. 8. Distributions of the final dark matter candidates in $\log_{10}(n_e/S1)$ vs. $S1$ (a), z vs. r^2 (b), and y vs. x (c). n_e is the number of electrons derived from $S2$ so that different run sets under different gate conditions can be plotted on the same graph. In (a), the solid blue and red lines are the ER and NR medians, respectively, and the dashed blue lines are the corresponding 95% quantiles of ER events. The dashed violet line represents the 99.5% NR acceptance cut. The nuclear recoil energy in keV_{nr} are indicated with the grey dashed lines. The six ER events located below the NR median line are highlighted in brown. In (b) and (c), the dashed lines are projections of the fiducial volume, and black (light grey) dots represent events inside (outside).

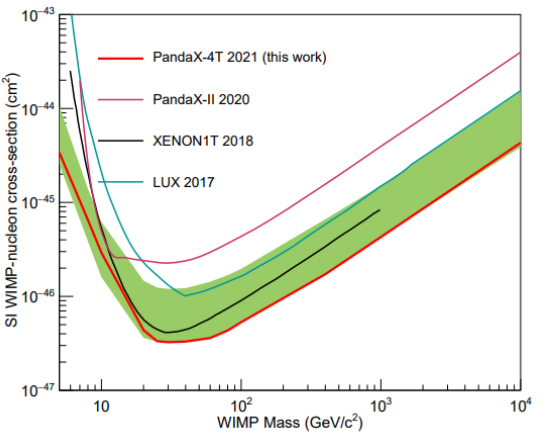


Fig. 9. The 90% C.L. upper limit vs. m_χ for the SI WIMP-nucleon elastic cross section from PandaX-4T commissioning data (red), overlaid with that from LUX 2017,[4] XENON1T 2018[3] and PandaX-II 2020.[5] The green band represents the $\pm 1\sigma$ sensitivity band.

sensitivity which is improved from the previous PandaX-II final analysis by 2.8 times at 40 GeV/c^2. Comparing the results (red line) with XENON1T, we are slightly edged, for example, at 40 GeV/c^2 where our exclusion limit is 1.3 times stronger. With this data, we can also dive into previously unexplored territory, such as approaching the low energy neutrino floor, and getting deeper into the SUSY parameter space.

6. Conclusion and Outlook

As mentioned, we have completed the first commissioning run of PandaX-4T with a 0.63 tonne·year exposure. PandaX-4T has produced the strongest WIMP-nucleon interaction constraint.[2] This demonstrates the physics potential of a highly sensitive multi-ton liquid xenon detector. PandaX-4T is currently performing an offline tritium removal campaign, aiming to reduce the electron recoil background by at least two-fold. In parallel, the collaboration is developing the plan for the next generation xenon experiment at CJPL.

References

1. M. W. Goodman and E. Witten, Detectability of certain dark-matter candidates, *Phys. Rev. D* **31**, 3059 (Jun 1985).
2. Y. Meng, Z. Wang, Y. Tao *et al.*, Dark Matter Search Results from the PandaX-4T Commissioning Run (2021).
3. E. Aprile *et al.*, Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, *Phys. Rev. Lett.* **121**, p. 111302 (2018).
4. D. S. Akerib *et al.*, Results from a search for dark matter in the complete LUX exposure, *Phys. Rev. Lett.* **118**, p. 021303 (2017).
5. Q. Wang, A. Abdukerim, W. Chen *et al.*, Results of dark matter search using the full PandaX-II exposure, *Chinese Physics C* **44**, p. 125001 (Nov 2020).
6. X. Zhou, X. Zeng, X. Ning, *et al.*, A Search for Solar Axions and Anomalous Neutrino Magnetic Moment with the Complete PandaX-II Data, *Chinese Physics Letters* **38**, p. 011301 (Jan 2021).
7. C. Cheng, P. Xie, A. Abdukerim *et al.*, Search for Light Dark Matter–Electron Scattering in the PandaX-II Experiment, *Physical Review Letters* **126** (May 2021).
8. X. Cui, A. Abdukerim, W. Chen *et al.*, Dark matter results from 54-ton-day exposure of pandax-ii experiment, *Physical Review Letters* **119**, p. 181302 (Oct 2017).

XRISM: X-ray imaging and spectroscopy mission

Makoto S. Tashiro
The Institute of Space and Astronautical Science, JAXA
Tsukuba, Ibaraki, Kanagawa, 305-8505, Japan
Department of Physics, Saitama University
Sakura-ku, Saitama, Saitama, 338-8570, Japan
tashiro@mail.saitama-u.ac.jp

and the XRISM team

The X-Ray Imaging and Spectroscopy Mission (XRISM) is a JAXA/NASA X-ray observatory with collaboration from ESA and several institutes and academic institutions worldwide. It is proposed to fulfill the promise of high-resolution X-ray spectroscopy with imaging once realized but unexpectedly terminated by a mishap of ASTRO-H/Hitomi. XRISM carries two sets of X-ray Mirror Assemblies and is equipped on the focal plane with a 6×6 pixelized X-ray micro-calorimeter array and an aligned X-ray CCD camera. With the combination of high-resolution spectroscopy imaging and the broader field of view, XIRSM is expected to pioneer a new horizon of the Universe in X-ray astrophysics. Aiming to launch the satellite in the Japanese Fiscal Year 2022, we fabricate the instruments and test for the satellite integration starting at the beginning of 2022. The paper reports the development status, reviewing the science objectives and the operation plan.

Keywords: X-ray astronomy; observation satellite; spectroscopy; imaging

1. Overview of XRISM

The X-ray imaging and spectroscopy mission — XRISM[1, 2] is a recovery mission of ASTRO-H Hitomi[3] launched on February 17, 2016. After the launch and following the commissioning of the onboard instruments, Hitomi started observation. However, due to a series of mishaps in the attitude control system, Hitomi was forced to stop operation on March 26. The observation period was only one month, but the results clearly show the X-ray microcalorimeter epoch-making capability. The observation was terminated before opening the gate valve. The valve was equipped to protect the sensor assembly from the atmosphere on the ground, but the gate valve limited the bandpass below ~ 2 keV. Even with the limitation in the low energy band, the unprecedented energy resolution enables us to measure the plasma turbulence velocity and detect weak lines of rare metals with a few ks exposure. The obtained spectrum is a highlight obtained from the Perseus cluster of galaxies.[4]

The central concept of XRISM is to maximize the X-ray micro-calorimeter science since it is an entirely new scientific field. We chose a conservative combination

with the CCD camera to strengthen the science. Table 1 summarizes of required specifications of the instruments of XRISM.

Table 1. XRISM mission instruments specifications

Parameters	Resolve	Xtend
Feild of view	$\geq 2.9' \times 2.9'$	$\geq 22' \times 22'$
Angular resolution†	$\leq 1.7'$	$\leq 1.7'$
Effective area at 6 keV	≥ 210 cm^2	≥ 300 cm^2
Imaging pixel format	6×6	1280×1280
Energy range	$0.3 - 12$ keV	$0.4 - 13$ keV
Energy resolution‡	≤ 7 eV (goal 5 eV)	≤ 250 eV*

†: The half power diameter
‡: The full width half maximum at 6 keV
*: At the end of mission life.

1.1. *Resolve*

The critical device of XRISM — the X-ray microcalorimeter realizes the high-resolution spectroscopy with imaging by the X-ray Mirror Assembly (XMA). The XMA is a Wolter-I X-ray mirror of their reflectors' figure in conical approximation. The reflectors of the mirror are made of heat formed aluminum sbstrate followed by epoxy replication on gold-sputtered smooth Pyrex clindrical mandrels. The XMA is developed and fabricated by NASA/Goddard Space Flight Center (GSFC). The imaging spectroscopy telescope of XRISM is named "Resolve."[5] The device measures phonons induced by a penetrated X-ray photon. The X-ray photon heats the mercury tellurized absorber by a few mK, and the temperature rise is measured by the attached silicon thermometer in the accuracy of μK. All the pixelized detectors are cooled down to 50 mK to reduce both heat capacity and thermal noise. The equipped adiabatic de-magnetized refrigerator (ADR) and liquid helium realize the extremely low temperature at the sensor in the vacuum dewar, whose thermal shields are cooled by Joule-Tomson coolers and two-stage Stirring coolers. The detector system is placed on the focal plane of an X-ray mirror assembly, and the 6×6 pixels realize a moderate imaging capability and unprecedented energy resolution. The Resolve telescope is emphasized with the high energy resolution of 7 eV FWHM at 6 keV. The energy resolution allows us to measure the velocity dispersion of the iron line in the accuracy of 100 km s^{-1}.

The detector assembly and the ADR cooling system are developed by NASA/GSFC, while JAXA and Sumitomo Heavy Industry (SHI) develop and provide the vacuum dewar and the mechanical cooling systems.

1.2. *Xtend*

The wide-field X-ray CCD imager with the identical XMA constitutes the "Xtend" telescope, which covers the same energy band of Resolve with a 200 μm thick back-illuminated CCDs.[6] The X-ray CCD, operated at the temperature of -110°C utilizing the one-stage Stirling cooler, covers almost the same energy band from oxygen lines to highly ionized irons. The device was developed for the Hitomi/SXI and employed by XRISM with a slight modification. The CCD is produced by Hamamatsu Photonics, is calibrated by the Xtend team of XRISM, and is installed in a camera body electronics by Mitsubishi Heavy Industry (MHI).

The spacecraft of 2.3 tons is 7.9 m long, 9.2 m wide, and the baseplate diameter is about 3 m. It is designed to work for three years with cryogen of Resolve. After three years, we can conduct cryogen-free operation with installed mechanical coolers.

2. Science Cases

The high energy resolution and weak line detection capability of Resolve exceed the previous observatories above 0.8 keV. The Xtend is excellent in grasp capability, which is the product of effective area and the solid angle of field of view. With the combination of a narrow field of view, Resolve, Xtend is useful for producing reliable scientific outputs with the brand new X-ray micro-calorimeter system. The science cases with X-ray microcalorimeter are summarized in ASTRO-H White Papers.[7–23]

2.1. *Structure formation of cluster of galaxies*

The precision of cluster cosmology depends on the accuracy of the mass measurements. One of the most important ways of measuring cluster masses is based on hydrostatic equilibrium — the assumption that the thermal pressure alone supports the hot gas against the cluster's gravity. However, numerical simulations show that the macroscopic motions of the intra-cluster medium (ICM) may provide significant non-thermal pressure support, up to 25 – 30% at the radius at which the cluster mass density is 200 times the critical density of the universe (r_{200}).

Hitomi observed the Perseus core, which is dominated by the central massive galaxy NGC 1275, to reveal the macroscopic motion of the ICM (Fig. 1. By measuring the line doppler broadening, the Hitomi succeeded in showing that the turbulent velocity dispersion is 164 ± 10 km s^{-1} as mentioned.[4] This kinetic pressure is less than 10% of the thermal pressure on a scale of 60 kpc. XRISM will continue Hitomi's legacy by measuring velocities of gas motions and the non-thermal pressure contribution with an accuracy of a few percent in many relaxed, bright galaxy clusters. Velocity measurements up to r_{2500}, for even a small sample of nearby relaxed clusters, will significantly improve the constraints on dark energy and other cosmological parameters.[17]

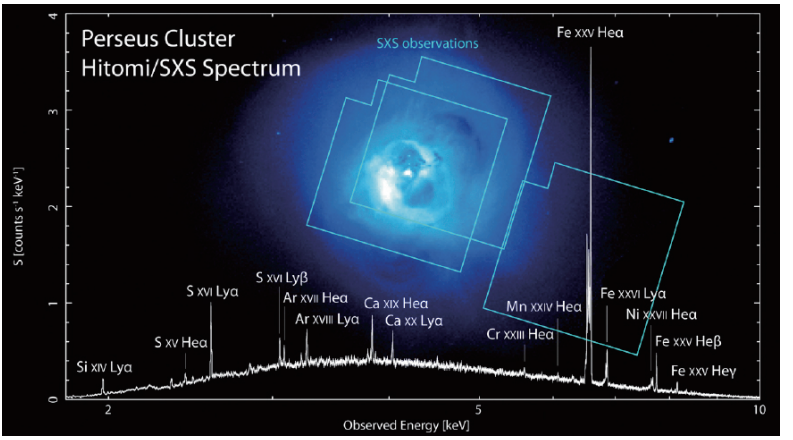


Fig. 1. Chandra image[24] and the Hitomi SXS spectrum from Perceus cluster of galaxies.[4,25] The blue boxes in the image indicates the field of view of Hitmoi SXS. Hitomi observed the centeral region and obtained the high resolution spectrum. Evaluating the Fe-K lines width, the Hitomi collaboration[4] reveals the turbulent velocity dispersion of 164 ± 10 km s^{-1}.[a]

2.2. *Transportation of energy and matters*

XRISM will conduct spatially and spectrally investigation on the Active Galactic Nucleus (AGN) feedback.[19] Figure 2 left panel shows Chandra X-ray image of the core region of the Virgo cluster of galaxies dominated by the central galaxy M 87, which highlights the complexity of structures in the gas. We focus that the 1 keV gas indicates uplifted gas produced by the AGN feedback. These structures will be resolved by the Resolve utilizing the imaging spectroscopy capability. Each white box (6×6) shows the field of view of Resolve, and by mapping, XRISM will observe both the uplifted gas structure and surrounding ICM. Figure 2 right shows a simulated Resolve spectrum of one of the regions with the expected outflows in M 87 assuming a 100 ks exposure. The spectrum resolves the contributions from the different thermal components to the Fe-L line complex and allows us to distinguish uplifted cool gas from the ambient hot ICM.

The chemical evolution of the universe reflects the history of billions of supernovae. Resolve will detect weak lines from rare elements in the ICM, such as Na, Al, Cr, and Mn. Al and Na are sensitive tracers of the metallicity of the underlying stellar population. Also, Cr and Mn can be used to probe the characteristics of SN Ia progenitors, resulting in information about the chemical enrichment history of the ICM and the universe.

[a]Credit: ISAS/JAXA https://www.isas.jaxa.jp/feature/forefront/190924.html.

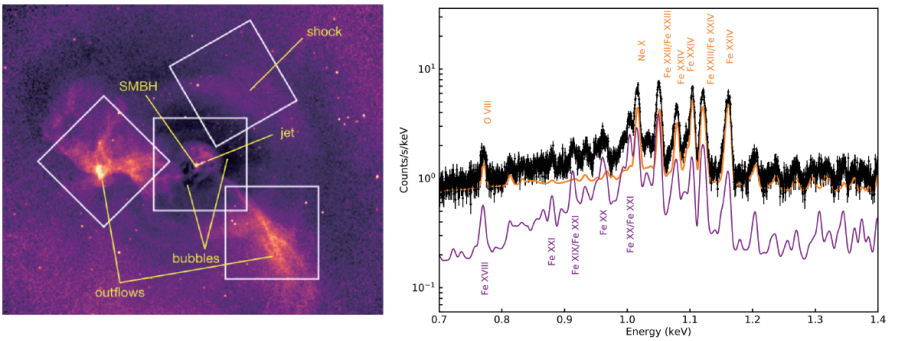


Fig. 2. Spatially and spectrally resolving AGN feedback with XRISM. Reprinted by courtesy of the authors from Figure 5 of "Science of XRISM"[30] which is updated from an ASTRO-H White Paper.[17] **Left:** Chandra X-ray image of the core region of the Virgo cluster, including the central galaxy M 87, divided by a spherically symmetric model of surface brightness. The image highlights the complexity of structures in the gas produced by the AGN feedback.[28,29] Each white box ($2.9' \times 2.9'$) shows the field of view of Resolve. The two regions point at the cool and bright structures, such as the shock and the center of M 87. **Right:** a 100 ks simulated Resolve spectrum of one of the regions with outflows (left box) in M 87. The spectrum shows the contributions from the different thermal components to the Fe-L line complex.

2.3. *Relativistic objects*

XRISM will revolutionize our understanding of accreting SMBHs, starting from how cold gas flows in from the interstellar medium (ISM) of the galaxy down to the inner accretion flow where matter is affected by the strong gravity of the black hole.[18]

Relativistic Doppler broadening and gravitational redshifts in the accretion disks surrounding black holes have been investigated for decades.[26] In particular, the ASCA observation of MCG–6-30-15 opened our observational view of a broad Fe K line in AGN X-ray spectra.[27] The implication of the broad line is immense, as it may provide a way of measuring an elusive fundamental parameter of a black hole — spin. As such, the broad Fe-K line has ignited debate and continued interest.

Rapid time variability shown in the following observations, including the X-ray reverberation mapping or quasar microlensing, improved our understandings. They indicate that the broad Fe K line is produced by reprocessing the inner accretion disk within $\sim 10R_g$. However, the details of the inner disk radius and kinematics of the flow are still highly debated.

By disentangling the expected contributions from other spectral features — narrow emission from the outer disk and torus or absorption through ionized outflows along our line of sight —, XRISM probes properties of the inner accretion flow to estimate the black hole spin. Thus, we will clearly understand how much material is making it to the black hole and how much material is removed through massive outflows.

Figure 3 left chart shows Fe-K line prole of MCG–6-30-15 as seen in 300 ks by XMM-Newton pn (blue) and Chandra HETGS (green), along with a Resolve simulation in red. The red XRISM spectrum alone shows clear spectral features in this Fe-K energy band. The middle chart shows The ratio of the XMM pn spectrum to a power-law model t between 2 and 4 keV, and 7.5 and 10 keV. The pn spectrum can be well described either by pure relativistic reflection with an inclination of 5 deg (blue) or by relativistic reflection with an inclination of 75 deg with additional emission and absorption, including an ultrafast outflow of $v_{\rm out} = 0.13\ c$ (red). The right panel shows that XRISM will easily distinguish between these two scenarios since it resolves the narrow emission and absorption features.

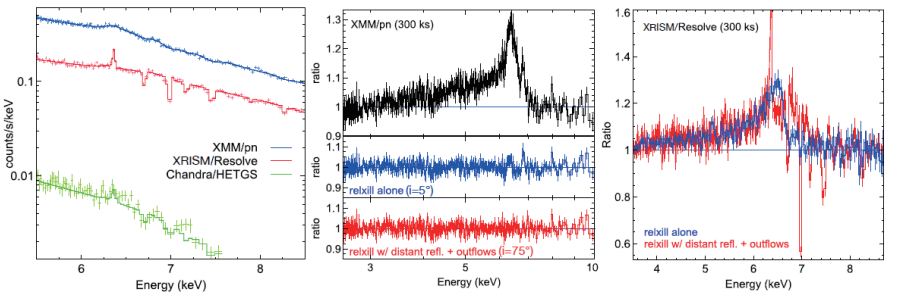


Fig. 3. Reprinted by courtesy of the authors from Figure 6 of "Science of XRISM." **Left**: Fe-K line pro le of MCG–6-30-15 as seen in 300 ks by XMM-Newton pn (blue) and Chandra HETGS (green), along with a Resolve simulation in red. **Middle:** The ratio of the XMM pn spectrum to a power-law model fit between 2–4 keV and 7.5 – 10 keV. The pn spectrum can be equally well described by pure relativistic reflection with an inclination of 5 deg (shown in blue) or by relativistic reflection with an inclination of 75 deg with additional emission and absorption, including an ultrafast outflow of $v_{\rm out} \sim 0.13\ c$ (shown in red). **Right:** XRISM will easily distinguish between these two scenarios, as the narrow emission and absorption will be resolved.[30]

The relativistic redshift on a compact object surface is also one of the scientific targets of XRISM. Here we show an example of the symbiotic star detected by Swift/BAT T Coronae Borealis (T CrB), which is a recurrent nova.[9] This fact and the hard BAT spectrum indicate that its white dwarf is exceptionally massive, around 1.35 solar mass.

T CrB exhibits a strong 6.4 keV Fe-K line, partly from the reflection of the white dwarf surface. Therefore, T CrB is an ideal target to investigate the compact object's mass and radius by measuring the gravitational redshift. This simulated spectrum is guided using the 46 ks Suzaku observation obtained in 2006.[31] The XIS data were reproduced well with a model absorbed by a partial covering absorber, with a single, narrow Gaussian at 6.396 keV. If it is, the inset shows that the XRISM Resolve has the statistical quality necessary to detect the expected gravitational redshift of 4 eV of the neutral iron-K line (Fig. 4).

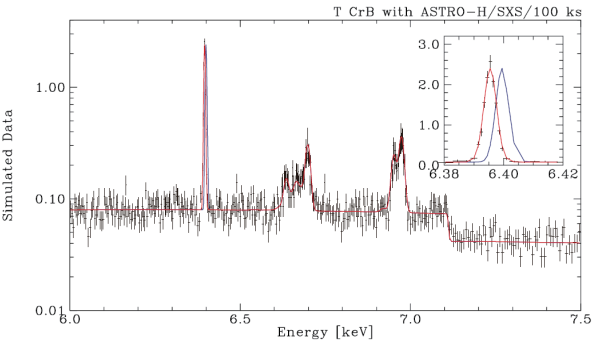


Fig. 4. Simulated 100 ks observation of T Cr B in the Fe-Kα region. The inset showing a close-up spectrum around the energy of 6.4 keV. It shows the Resolve has a capability to detect the expected gravitational redshift of 4 eV (see text). Reprinted by courtesy of the authors from Figure 2 of "ASTRO-H White Paper – White Darf"[9]

3. Mission Status and Observation Plan

All the instruments produced under the international collaboration are being delivered to the JAXA Tsukuba Space Center. The Resolve calorimeter sensor insert was delivered and installed in the dewar produced in SHI, Japan. The integrated dewar system equipped with the mechanical coolers was transferred to Tsukuba Space Center with the cooler drivers. The XMA is under calibration and is planned to be delivered to Japan in January 2022. All CCD chips for the Xtend were screened in Osaka University and were integrated into the camera system at MHI. The camera was transferred to Tsukuba Space Center. In 2022, these mission instruments will be installed on the satellite in the first half. Then we conduct the proto-flight test. XRISM will be launched in the Japanese Fiscal Year 2022 or the early calendar year 2023 from Tanegashima Space Center.

Table 2. XRISM observation phases

Initial phase	3 months
critical operation comissioning	$\sim$ 1 week until 3 months after launch
Nominal phase	until 3 years after launch
initial calibration & test observations calibration & performance verification observation nominal observation (Guest observations)	1 month 6 months 26 months

Observation phases defined in the operation plan are summarized in Table 2. At the beginning of the nominal observation phase, the performance verification (PV)

observations are conducted. The initial observation phase is planned to verify the performance to show the XRISM capability to the astrophysics community. The PV observation target objects list was recently released on the XRISM website for researchers.[b] The XRISM Science Team consists of contributors to develop instruments, data processing software, and calibration and observation plan to carry out the PV observations. In addition to the XRISM Science Team, unlike Hitomi, XRISM introduced the XRISM Guest Scientist Program to the PV phase observation. The XGSs will be solicited by the three agencies — JAXA, NASA, and ESA independently — in 2022 and contribute to PV observation planning, data analysis, and produce science output with the XRISM Science Team. The guest observer phase is open 10 months after the launch. The announcement of opportunity for the guest observations is planned to be submitted at the end of the commissioning.

The approved nominal mission term is 3 years. After the 3 years, the XRISM project is to have a mission completion review. At the moment, if the satellite, including the mechanical cooling system, functions well and the mission extension review approve, the latter phase operation will be conducted in the latter half of the 2020s.

Acknowledgments

XRISM is a JAXA mission joint with NASA in collaboration with ESA. The XRISM Science Team studies the science cases described in this paper based on the previous research by the ASTRO-H Science Working Group.

References

1. M. S. Tashiro et al, Concept of X-ray Astronomy Recovery Mission (XARM), in *Proc. SPIE: Space Telescopes and Instrumentations 2018: Ultraviolet and Gamma-Ray*, **10699** eds. J-W. A. den Herder, S. Nikzad, (2018) id. 1069922.
2. M. S. Tashiro et al, Status of X-ray Imaging and Spectroscopy Mission (XRISM), in *Proc. SPIE: Space Telescopes and Instrumentations 2020: Ultraviolet and Gamma-Ray*, **11444** eds. J-W. A. den Herder, S. Nikzad, K. Nakazawa (2020) id. 1144422.
3. T. Takahashi et al., *Journal of Astronomical Telescopes, Instruments, and Systems*, **4** (2018) id. 021402.
4. Hitomi Collaboration, *Nature*, **535** (2016) 117.
5. Y. Ishisaki et al., Status of resolve instrument for X-ray imaging telescope (Xtend) onboard X-ray Astronomy Recovery Mission (XARM), in *Proc. SPIE: Space Telescopes and Instrumentations 2018: Ultraviolet and Gamma-Ray*, **10699** eds. J-W. A. den Herder, S. Nikzad, (2018) id. 1069924.
6. K. Hayashida et al, Soft X-ray imaging telescope (Xtend) onboard X-ray Astronomy Recovery Mission (XARM), in *Proc. SPIE: Space Telescopes and Instrumentations 2018: Ultraviolet and Gamma-Ray*, **10699** eds. J-W. A. den Herder, S. Nikzad, (2018) id. 1069923.

[b] https://xrism.isas.jaxa.jp/research/

7. T. Takahashi et al., ASTRO-H Space X-ray Observatory White Paper, *arXiv*, (2014) 1412.2351.
8. Y. Tsuboi et al., ASTRO-H Space X-ray Observatory White Paper - Stars – Accretion, Shocks, Charge Exchange and Magnetic Phenomena, *arXiv*, (2014) 1412.1162.
9. K. Mukai et al., ASTRO-H Space X-ray Observatory White Paper - White Dwarf, *arXiv*, (2014) 1412.1163.
10. C. Done et al., ASTRO-H Space X-ray Observatory White Paper - Low-mass X-ray Binaries, *arXiv*, (2014) 1412.1164.
11. S. Kitamoto et al., ASTRO-H Space X-ray Observatory White Paper - Accreting Pulsars and Related Sources, *arXiv*, (2014) 1412.1165.
12. J. Miller et al., ASTRO-H Space X-ray Observatory White Paper - Stellar-Mass Black Holes, *arXiv*, (2014) 1412.1173.
13. J. P. Hughes et al., ASTRO-H Space X-ray Observatory White Paper - Young Supernova Remnants, *arXiv*, (2014) 1412.1169.
14. K. S. Long et al., ASTRO-H Space X-ray Observatory White Paper - Old Supernova Remnants and Pulsar Wind Nebulae, *arXiv*, (2014) 1412.1166.
15. K. Koyama et al., ASTRO-H Space X-ray Observatory White Paper - Plasma Diagnostic and Dynamics of the Galactic Center Region, *arXiv*, (2014) 1412.1170.
16. F. Paerels et al., ASTRO-H Space X-ray Observatory White Paper - High Resolution Spectroscopy of Interstellar and Circumgalactic Gas in the Milky Way and Other Galaxies, *arXiv*, (2014) 1412.1174.
17. T. Kitayama et al., ASTRO-H Space X-ray Observatory White Paper - Clusters of Galaxies and Related Sources, *arXiv*, (2014) 1412.1176.
18. C. Reynolds et al., ASTRO-H Space X-ray Observatory White Paper - AGN Reflection, *arXiv*, (2014) 1412.1177.
19. J. S. Kaastra et al., ASTRO-H Space X-ray Observatory White Paper - AGN Reflection, *arXiv*, (2014) 1412.1171.
20. R. K. Smith et al., ASTRO-H Space X-ray Observatory White Paper - New Spectral Features, *arXiv*, (2014) 1412.1172.
21. F. Aharonian et al., ASTRO-H Space X-ray Observatory White Paper - Shock and Acceleration, *arXiv*, (2014) 1412.1175.
22. P. Coppi et al., ASTRO-H Space X-ray Observatory White Paper - Broad-band Spectroscopy and Polarimetry, *arXiv*, (2014) 1412.1190.
23. M. S. Tashiro et al., ASTRO-H Space X-ray Observatory White Paper - Chemical Evolution in High-z Universe, *arXiv*, (2014) 1412.1179.
24. I. Zhuravleva, et al.,, NASA/CXC/Stanford.
25. Hitomi Collaboration, *Nature*, **551** (2017) 478.
26. A. C Fabian, et al., *MNRAS* **238** (1989) 729.
27. Y. Tanaka et al., *Nature* **6533** (1995) 659.
28. W, Forman et al.., *ApJ.* **665** (2007) 1057.
29. A. Simionescu et al., *A&A* **482** (2008) 97.
30. XRISM Science Team, Science with the X-ray Imaging and Spectroscopy Mission (XRISM), *arXiv*, (2020) 2003.04962.
31. G. J. M. Luna et al., RS Ophiuchi (2006) and the Recurrent Nova Phenomenon ASP Conference Series, **401** (2007) 342.

The SVOM mission

On behalf of the *SVOM* collaboration
J.-L. Atteia[1], B. Cordier[2], J. Wei[3]
[1]*IRAP, Université de Toulouse, CNRS, CNES, UPS, (Toulouse), France*
E-mail: jean-luc.atteia@irap.omp.eu

[2]*Lab AIM - CEA, CNRS, Université Paris-Saclay, Université de Paris, 91191, Gif-sur-Yvette, France*
E-mail: bertrand.cordier@cea.fr

[3]*Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*
E-mail: wjy@nao.cas.cn

The Sino-French space mission *SVOM* is mainly designed to detect, localize and follow-up Gamma-Ray Bursts and other high-energy transients. The satellite, to be launched mid 2023, embarks two wide-field gamma-ray instruments and two narrow-field telescopes operating at X-ray and optical wavelengths. It is complemented by a dedicated ground segment encompassing a set of wide-field optical cameras and two 1-meter class follow-up telescopes. In this contribution, we describe the main characteristics of the mission and discuss its scientific rationale and some original GRB studies that it will enable.

Keywords: *SVOM*; High-Energy Astrophysics; Multi-Messenger Astrophysics; Gamma-Ray Bursts.

1. Why a New Gamma-Ray Burst Mission?

SVOM (Space-based multi-band astronomical Variable Objects Monitor) is a Sino-French mission devoted to the study of gamma-ray bursts (GRBs) to be launched in 2023.[1] Its main goals are the detection and multi-wavelength study of high-energy transients, and the search for the electromagnetic counterparts of non-photonic transient sources detected in gravitational waves or neutrinos. Before describing *SVOM* in Section 2, we briefly discuss in this section the rationale for a new GRB mission in the coming years.

1.1. *High-energy Astrophysics*

The astrophysics of high-energy transients is essential for the study of several major questions of modern astrophysics.

High-energy transients offer a unique view at extreme stellar explosions with relativistic jets. High-energy observations reveal the presence of a 'central engine' able to launch relativistic jets, and as such they are essential to clarify key issues like the complex zoology of GRBs (which encompasses long and short duration GRBs,

low-luminosity GRBs and ultra-long GRBs), the nature of the GRB-supernova (SN) connection or the distribution of GRB beaming angles. The jets themselves raise several unanswered questions, like their composition and geometry, the nature of the central engine (a magnetar or a black hole?) and the mechanism of their acceleration, the physical processes explaining their high radiative efficiency, and their role in the acceleration of Very High-Energy cosmic rays. The jets have their origin in the universal accretion/ejection mechanism that takes place around accreting compact objects. However, GRB jets being extremely relativistic with Lorentz factors of the order of one hundred, they probe a regime that is not accessible with other phenomena.

Sometimes, high-energy transients are associated with the birth of stellar mass black holes (BHs). BH astrophysics has emerged as a major topic after the discovery of dozens of binary black hole mergers by the LIGO Scientific Collaboration and Virgo Collaboration (LVC).[2] Black hole demography and zoology, their mass & spins distribution and their birth places are key questions which are now actively studied. In this context, GRBs are of high interest because they offer a complementary view on the birth of stellar mass BHs across the history of the universe.

Gamma-ray bursts are so intense that some of them are detectable out to high redshift ($z \geq 5$).[3] They illuminate regions of the young universe (at z=6 the universe is about 950 Myr old), pinpointing galaxies that would otherwise remain undetectable and measuring their chemical enrichment and physical state, as well as the properties of the intergalactic medium at such high redshifts. Additionally, the history of GRB formation provides crucial insight into the formation rate of massive stars in the early universe, and pushing the redshift limit a bit further may permit the identification of GRBs from the first generation of stars (usually called population III stars) giving us a glimpse at the formation of the first stellar mass black holes. The current GRB samples probably contain a small fraction of very distant GRBs that have not been recognized in the absence of NIR spectra of their afterglows, this emphasizes the need for fast NIR followup of GRB afterglows.

Neutron stars (NSs) are prolific sources of high-energy transients, which can be emitted by very different mechanisms in accreting neutron stars, magnetars and in mergers of neutron stars. Fundamental questions like the equation of state of neutron star, the mass gap between neutrons stars and black holes, the radiation process of magnetars are active fields of research which require continued observations of violent phenomena involving neutron stars.

Finally, GRBs are powerful tools to test the Lorentz invariance over very large distances and they may also become a privileged tool for cosmography if we manage to standardize them.

1.2. *Multi-messenger Astrophysics*

The detection of transient gravitational wave signals from cosmic explosions has opened a new window on violent phenomena involving compact objects like neutron

stars and black holes. The complementarity of gravitational wave (GW) astrophysics and HE astrophysics has been wonderfully illustrated by the joint detection of GW170817 and GRB 170817A[4] followed by the discovery of the kilonova AT 2017gfo. These discoveries opened the era of multi-messenger astrophysics. This new context raises new questions for HE astrophysicists, like the geometry of the jets of short GRB, the compared demography of BNS mergers and short GRBs, the compared demography of BBH mergers and long GRBs, the masses of BHs involved in galactic HE transients, in BBH mergers and in GRBs, the photonic emission of BNS mergers, BHNS mergers, and BBH mergers, the delay between a merger and the emergence of a GRB, or the comparison of the speed of gravitational waves with the speed of light. This non-comprehensive list illustrates a renewed interest for transient HE observations in the multi-messenger era.

1.3. *A Rich Instrumental Panorama*

Additional motivations to develop new instruments for the exploration of the transient high-energy sky are the diversity of sources that emit bursts of HE photons, the fact that some of them are rare events that do not repeat, and the rapid evolution of the instrumental panorama.

The transient manifestations of black holes at high energies involve galactic and extragalactic X-ray binaries, GRBs of all types, Active Galactic Nuclei (AGNs), Relativistic Tidal Disruption Events (TDEs), and possibly BHNS mergers. Neutron stars, on the other hand, manifest themselves as galactic and extragalactic X-ray binaries, magnetars or soft gamma repeaters eventually associated with FRBs, and of course as BNS mergers. The diversity of these manifestations calls for different observing strategies, in terms of energy range, sensitivity, pointing strategy, follow-up, etc., which cannot be accommodated with a single mission.

The construction of novel missions is further justified by the fact that the coming years are expected to be a golden age for transient sky astronomy and for multi-messenger astrophysics, with a number of large facilities surveying the sky for photons at all wavelengths and for non-photonic messengers. In the electromagnetic domain, we can mention the Cherenkov Telescope Array (CTA[5]), the High-Altitude Water Cherenkov Observatory (HAWC[6]) and the Large High Altitude Air Shower Observatory (LHAASO[7]) for Very High-Energy gamma-rays. In X-rays, eROSITA[8] will establish a reliable map of the X-ray sky, allowing the fast identification of new sources. In the visible the already operating Zwicky Transient Facility (ZTF[9]) and the future Vera Rubin Observatory[10] will map the visible transient sky every night allowing meaningful comparisons with the transient high-energy sky. In the meantime, new spectrometers on large telescopes, like SOXS[11] or NTE[a] will increase our capacity to measure the redshifts of extragalactic transients. In the radio domain, the counterparts of HE transients will be observed with very good sensitivity with

[a] https://nte.nbi.ku.dk/

the precursors of SKA, like ASKAP[12] or MEERKAT,[13] while instruments like the Canadian Hydrogen Intensity Mapping Experiment (CHIME[14]) or the Commensal Real-time ASKAP Fast Transients Survey (CRAFT[15]) will monitor the transient radio sky for Fast Radio Bursts (FRBs).

The true revolution, however, comes from the instruments looking for non-photonic messengers from violent events, in particular the gravitational waves interferometers LIGOâ VIRGO, GEO and KAGRA, which have opened the exploration of the transient gravitational wave universe.[16] As shown with the detection of GW170817 and GRB 170817A, the joint operation of gravitational wave detectors and HE detectors brings crucial information on transient gravitational wave sources. The field of neutrino astronomy is also evolving very quickly and the large neutrinos observatories like ICECUBE[17] or KM3NeT[18] have started to constrain the physics of GRB jets.[19]

After this short introduction, we now present the SVOM mission (Section 2) and some of its science capabilities for the study of GRBs (Section 3).

2. The *SVOM* Mission

The *SVOM* mission (Figure 1) is the result of a bilateral collaboration between France (CNES) and China (CAS, CNSA).[20,21] It involves several research institutes from these two countries and contributions from the University of Leicester, the Max Planck Institut für Extraterrestische Physik and the Universidad Nacional Autónoma de México. *SVOM* is led by J.Y. Wei from NAOC in China and B. Cordier from CEA in France. The mission has been designed to survey the high-energy sky and to follow-up cosmic transients at optical and X-ray wavelengths. It encompasses a space segment, with four instruments embarked onboard a low Earth orbit satellite and a strong ground segment[22] with two sets of wide-field optical cameras, two 1-meter class ground follow-up telescopes, and a network of $\sim$ 45 VHF receiving stations distributed along the footprint of the orbit. The launch of *SVOM* is scheduled mid 2023 with three years of nominal operations and a possible extension of two years.

The main science drivers of *SVOM* are high-energy transient astrophysics with a focus on gamma-ray bursts, multi-messenger astrophysics, and time domain astronomy. The "core program" is dedicated to GRB detection and follow-up, aiming to improve our understanding of these phenomena. Between GRBs, *SVOM* will carry out other programs driven by the narrow-field instruments: a target of opportunity program focused on multi-messenger astrophysics and a general program focused on multi-wavelength astronomy. A detailed description of the *SVOM* mission and its science objectives can be found in Wei et al.[1] and in the website of the mission[b]. We now briefly describe the space and ground instruments developed for *SVOM*.

[b] http://www.svom.fr/

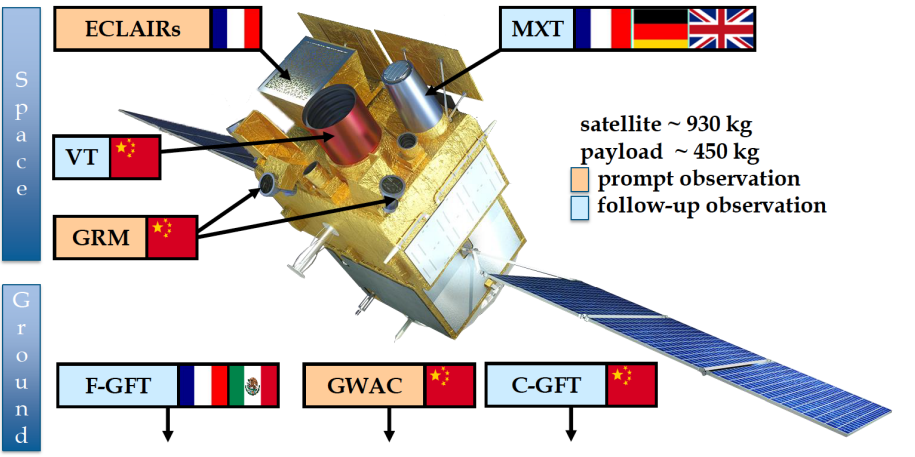
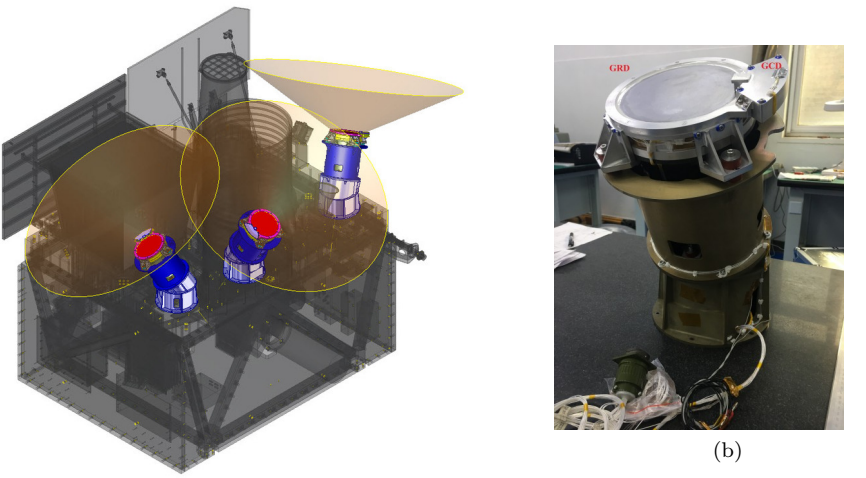


Fig. 1. Outline of the *SVOM* mission, showing the satellite and the space and ground instruments.

2.1. *Space-based Instruments*

2.1.1. *Gamma-Ray Monitor – GRM*



(a) (b)

Fig. 2. The GRM instrument. (a) Schematic drawing showing the position and orientation of the three GRD detectors on the science platform. (b) One of three GRD detection modules of the GRM.

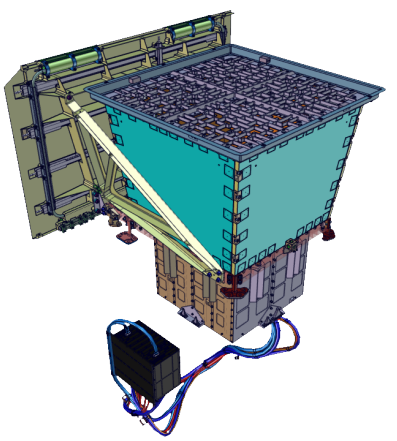
The Gamma-ray monitor (GRM) is a wide-field gamma-ray spectrometer.[23–25] It encompasses three NaI-based detection modules (called GRDs) looking at different regions of the sky. The pointing directions of the three GRDs are 30° from the

pointing axis of the satellite and they are separated by 120° in azimuth (Figure 2). The main characteristics of the instrument are given in Table 1. The GRM allows the detection of GRBs with a good sensitivity within a large field of view (2.6 sr), a feature which is especially important for the detection of short/hard GRBs in the era of transient gravitational wave detectors. It will also nicely complement ECLAIRs to measure the GRB spectral parameters over a large energy range (4 keV–5 MeV).[26]

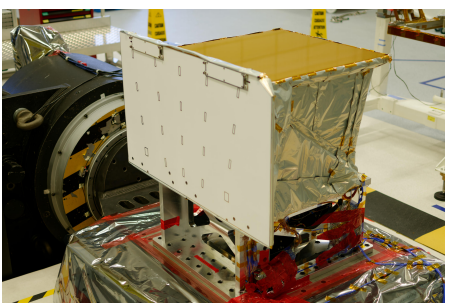
Table 1. GRM main characteristics.

Energy range (keV)	15 – 5000
Number of GRD modules	3
Detection area (cm^2, 1 of 3 modules)	200
Field of View (sr, all 3 modules)	2.6
Expected GRB rate (yr^{-1})	90 ($\sim$ 17% short[27])
Additional features	14 detection times from 5 ms to 40 seconds in 5 energy bands
	6 mm plastic scintillator to monitor particles & reject particle events
	Crude localization capability

2.1.2. *ECLAIRs*



(a)



(b)

Fig. 3. The ECLAIRs instrument. (a) Schematic drawing showing the radiator on the left, the coded mask (dark blue), the shield (light blue), the detection plane electronics (grey) and the onboard calculator (black). The detection plane itself is not visible, it is located 45 cm below the coded mask at the bottom of the shield. (b) ECLAIRs during vibration tests.

ECLAIRs is the hard X-ray imager and trigger of SVOM. The instrument is based on a coded-mask placed in front of a CdTe detection plane with 6400 4x4 mm^2

pixels,[28,29] its main characteristics are summarized in Table 2. An interesting feature of ECLAIRs is its low energy threshold at 4 keV. Reaching this threshold was a significant technological challenge for the detectors and their readout electronics, but also for the mask, whose holes have to be transparent at these energies, and for the trigger software which has to deal with non-GRB transient sources (e.g. X-ray bursters), which are numerous at energies of a few keV. When it detects an excess in a sky image, ECLAIRs sends an alert to the satellite in order to start a slew towards the source (if the conditions are met) and to start followup with the onboard narrow-field instruments MXT & VT. In parallel, the alert is broadcast by the onboard VHF antennas allowing to trigger followup observations from the ground. All the photons detected by ECLAIRs are sent to the ground, allowing to look for faint transients with improved detection methods after a typical delay of few hours.

Table 2. ECLAIRs main characteristics.

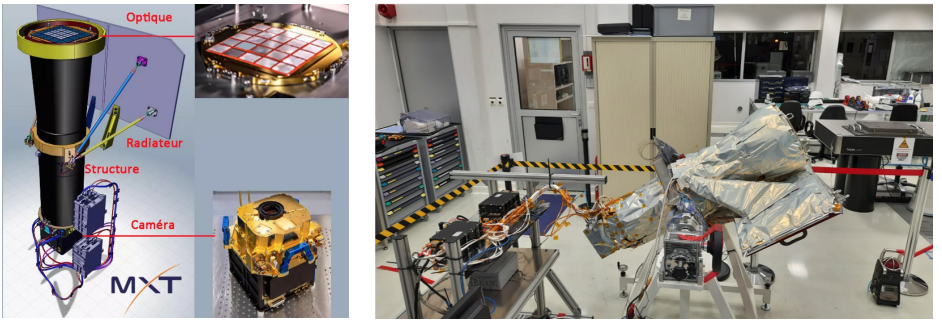
Energy range (keV)	4.0 – 150.0
Detection area (cm^2)	1000
Number of CdTe detectors	6400
Median FWHM at 60 keV (keV)	1.3
Field of View (total, sr)	2.0
Mask open fraction (%)	40
Localisation accuracy at detection limit (arcminute)	12.0
Expected GRB rate (yr^{-1})	65
Additional features	18 detection times from 10 ms to 20 min. in 3 energy bands

2.1.3. *Microchannel X-ray Telescope – MXT*

The Microchannel X-ray Telescope (MXT) is a focusing X-ray telescope using an innovative focusing "Lobster-Eye" micro-pores optics and a fully-depleted frame-store silicon pnCCD camera read out by two ASICs.[30,31] Its main characteristics are summarized in Table 3. The MXT will observe the X-ray afterglow promptly, improving ECLAIRs localization by a factor $\geq 10^3$.

Table 3. MXT main characteristics.

Energy range (keV)	0.2 – 10.0
Focal length (cm)	100
Field of View (arcminute)	64 x 64
Effective area (cm^2 at 1 keV)	27
Energy resolution at 1.5 keV (eV)	80
Localisation accuracy (arcsecond)	13.0
(for 50% of GRBs, within 5 minutes of the trigger)	
Additional features	Micro-pores "Lobster Eye" optics with square 40 μm pores
	Silicon pnCCD based camera

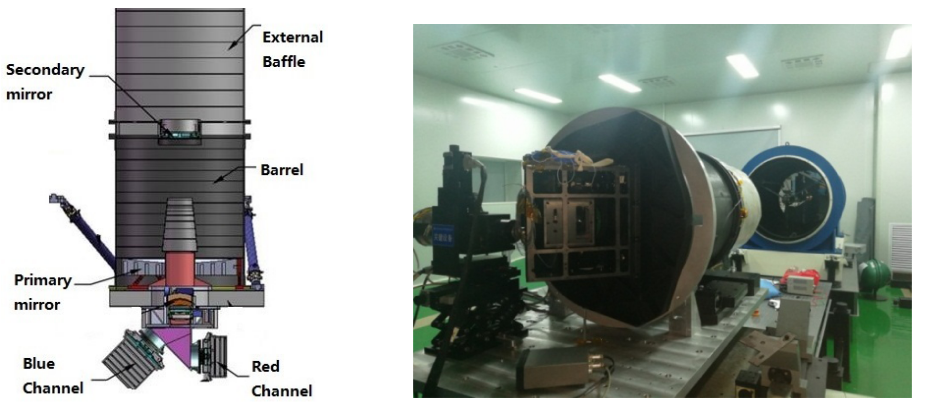


(a) (b)

Fig. 4. The MXT instrument. (a) Schematic drawing showing the radiator in light blue, the lobster eye optics (top inset), the shield (black), the camera (bottom inset), and the onboard calculator. (b) MXT during tests.

2.1.4. *Visible Telescope – VT*

The Visible Telescope (VT) is a Ritchey-Chretien optical telescope with two channels, able to detect GRBs up to redshift z $\sim$ 6.5.[32,33] It will observe SVOM GRBs quickly (in less than 5 minutes), providing an arcsecond localization of the afterglow or a deep limit ($M_v = 22.5$) very quickly after the burst.



(a) (b)

Fig. 5. The VT instrument. (a) Schematic drawing showing the various components except the onboard calculator. (b) VT during tests.

Table 4. VT main characteristics.

Diameter (mm)	400
Focal length (mm)	3600
Two channels	blue (400-650 nm) and red (650-1000 nm)
Field of View (arcminute)	26 x 26
Localisation accuracy (arcsecond)	≤ 1.0
Sensitivity (300 s integration)	$M_v = 22.5$
Additional features	Ritchey-Chretien telescope
	Two 2k x 2k CCD detectors
	Will detect 80% of ECLAIRs GRBs
	Fully covers ECLAIRs error boxes

2.2. *Ground-based Instruments*

2.2.1. *Ground Wide Angle Cameras – GWAC*

The Ground Wide Angle Cameras (GWAC)[34] is a set of several units, each made of 4 wide-angle optical cameras (called JFoV) and a small photographic camera (called FFoV) on a single fast-moving mount (Fig. 6). Each unit covers instantaneously about 12% of the ECLAIRs field of view, five of them have been installed at Xinglong Observatory (China) and 4 more will be installed at Muztagh Ata Observatory (China). The characteristics of JFoV and FFoV cameras are given in Table 5, the FFoV extends the optical flux coverage by ~ 6 mag in R-band, at the bright end. GWACs will scan the entire accessible sky each night to detect optical transients, offering the unique capability to detect precursor optical emission for *SVOM* GRBs. Three robotic telescopes on the same site (two 60 cm diameter and one 30 cm diameter) can quickly follow interesting events detected by GWAC. The GWACs are in commissioning phase, they have already participated in the search for counterparts of gravitational wave events during the O2 and O3 runs of LIGO/VIRGO.[35]

Fig. 6. Four GWAC units at the Xinglong Observatory (on the right of the picture).

Table 5. GWAC main characteristics.

Number of units	5 (Xinglong) + 4 (Muztagh Ata)
Number of telescopes per unit	4 (JFoV) + 1 (FFoV)
Diameter (mm)	180 (JFoV) – 35 (FFoV)
Detector	4k x 4k CCD (JFoV) – 3k x 3k CCD (FFoV)
Field of View (degrees)	12.8 x 12.8 (JFoV) – 30 x 30 (FFoV)
Sensitivity (σ)	R $\sim$ 16 (JFoV) – R $\sim$ 12 (FFoV)
Location	Xinglong and Muztagh Ata Observatories (China)
Additional features	Self triggering capability The FFoV performs guiding.

2.2.2. *Ground Followup Telescopes – GFT*

Two Ground Followup Telescopes (GFT, Fig. 7) aim at observing quickly the high-energy transients detected by *SVOM*, to precisely measure their light curves and their positions. The two GFTs are respectively located in China and Mexico, about 120° apart in longitude, to maximise the probability of immediate observations in response to *SVOM* alerts. Upon reception of an alert, the GFT which is in the night (if any) will start observing the source - if it is above the horizon - in less than 1 minute. The GFTs will also be used to perform joint observations with the satellite as well as non-*SVOM* observations. The main characteristics of the Chinese GFT (CGFT[c]) and of the French-Mexican GFT (Colibri[36]) are given in Table 6.

Table 6. GFTs main characteristics.

	C-GFT	Colibri
Diameter (mm)	1200	1300
Focal ratio	8	3.75
Number of channels	3 (g ; r ; i)	3 (g/r/i ; z/y ; J/H)
Field of View (arcminute)	21 x 21	26 x 26 (grizy) ; 21 x 21 (JH)
Sensitivity (r channel, 300 s, 10σ)	$m_{AB} \approx 20$	$m_{AB} \approx 22$

2.3. *The Mission*

2.3.1. *Orbit and Pointing Strategy*

The *SVOM* satellite will be launched by a Long March 2C launch vehicle, on a 29° inclined orbit, at an altitude of 630 km. The pointing strategy has been designed to favour the detection of extragalactic transients and their immediate observability by ground based telescopes. This calls for a nearly anti-solar pointing (Fig. 8a) with some excursions to avoid SCO-X1 and the galactic plane.[37] In the following this pointing strategy is called "the B1 law". The resulting 1 year sky coverage of ECLAIRs is shown in Fig. 8b, it favours the galactic poles, with exposure times

[c]https://www.svom.eu/en/#filter=.instrumentsarticle.portfolio217load

(a) (b)

Fig. 7. The two GFTs. (a) Chinese GFT at Jilin observatory. (b) French-Mexican GFT (aka Colibri) under test at the "Observatoire de Haute Provence" in France, before its installation at the Mexican National Astronomical Observatory in San Pedro Mártir.

reaching 3 to 4 Ms. The B1 law will be the rule during the nominal part of the mission (the first three years), with three exceptions: to point GRBs detected onboard, to point targets of opportunity, and to observe few remarkable sources as part of the general program (cf. Section 2.3.2).

(a) (b)

Fig. 8. *SVOM* pointing strategy. (a) Illustration of the anti-solar pointing that allows telescopes on the night side of the Earth to observe SVOM alerts promptly. (b) ECLAIRs one year exposure map in galactic coordinates.

2.3.2. *Observing Programs*

The observing time of *SVOM* is divided into three components: the *Core Program* (CP), which is the observation of HE transients detected onboard, the *General*

Program (GP), which is the observation of pre-selected sources close to the B1 law with the narrow field instruments, and the *Target of Opportunity* program (ToO) to observe active sources upon request from the ground.[1] The GP is selected by a TAC every six months and uploaded every week for the next two weeks. The ToOs are approved by the PIs of the mission and uploaded as soon as possible (see Sec. 2.3.3). Figure 9 shows the time allocated to each of the three programs during the nominal mission and during the extended mission.

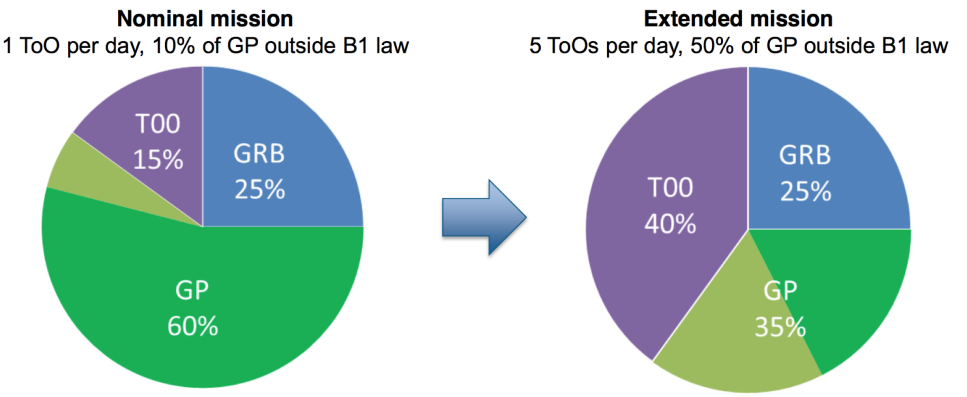


Fig. 9. Fraction of observing time attributed to the three observing programs of SVOM. Left: during the nominal 3-yr observing phase. Right: during the 2-yr extended phase. Light green indicates the time devoted to the General Program outside the B1 law.

2.3.3. *Alerts and ToOs*

SVOM alerts will be distributed promptly to the community through the GCN network. Alerts can report a count rate excess or the detection of a new source or both. The detection of a count rate excess triggers the construction of a sky image and the search for a new source. A catalog of known sources onboard permits differentiating new transients from known sources,[38] the slew threshold for known sources is set very high in order to slew towards a known source only when it is in a truly exceptional state. Alerts validated onboard will be broadcast by two onboard VHF antennas, hopefully detected by one of the $\sim$ 45 VHF receiving stations distributed around the world (Fig. 10) and forwarded to the French Science Center, which will format them as GCN Notices and VOEvents and transmit them immediately to the GCN.[39]

On the way up, commands can be sent quickly to *SVOM* with the Beidou short messages communication service.[40] This will be used to program target of opportunity observations towards active high-energy sources or towards neutrinos or gravitational waves transient sources.

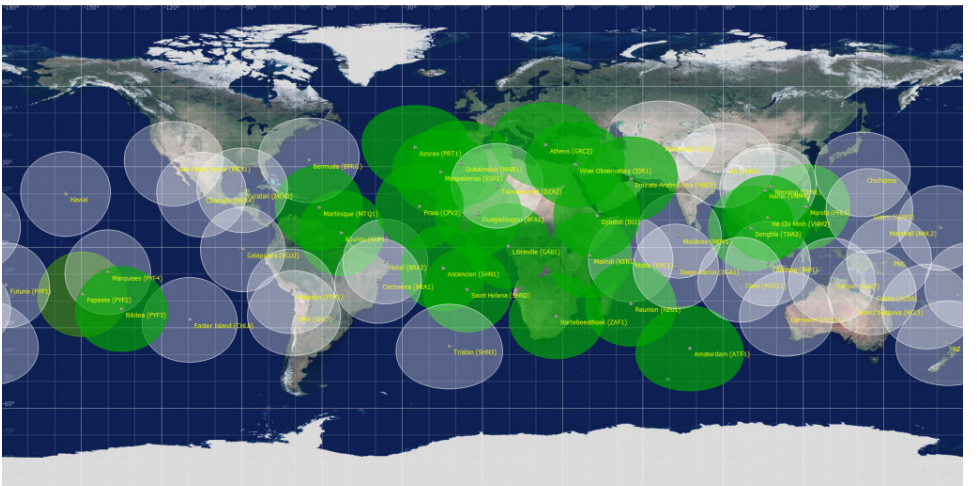


Fig. 10. *SVOM* network of VHF receiving stations. In green the stations installed and fully operational at the end of 2021.

2.3.4. *Burst Advocates and Instrument Specialists*

Like other GRB missions (e.g. *Swift* or *Fermi*), the management of *SVOM* alerts will be assigned to *Burst Advocates* who will survey all activities connected with the alert, act as points of contact for the non-*SVOM* world, and eventually take decisions about the follow-up of the source. Considering the large number of space and ground instruments involved in the detection and follow-up of transients and the number of burst advocates needed to ensure the proper follow-up of the alerts, it is needed to complement the pool of burst advocates with a few *Instrument Specialists* (2 or 3 per instrument) that can be contacted at any time by a burst advocate who has questions about the details of an instrument's operation.

2.3.5. *Mission Status*

All *SVOM* instruments have been constructed, they have undergone environment testing and full characterization, which have demonstrated their nominal performance. They are ready to be shipped to Shanghai to be integrated on the satellite in the first half of 2022. The full satellite will then undergo complete testing, before the launch planned mid 2023. The complex *SVOM* system (satellite operations and programming, observation sequences with and without triggers, satellite slews, alert reception and distribution, data management, etc.) has also been tested during extensive "system tests" lasting several days, that reproduced realistic operating conditions. These tests have demonstrated the capability of the *SVOM* collaboration to manage all aspects of this complex mission.

3. *SVOM* & GRB Science

We conclude this presentation with a subjective list of GRB science topics that could make good progress in the coming years thanks to *SVOM*. A presentation of the complete list of scientific objectives of the mission can be found in Wei et al.[1]

3.1. *GRB detection*

Thanks to its low energy threshold and a thorough onboard image processing, ECLAIRs will be particularly sensitive to long and soft GRBs. As such, it provides a unique opportunity to increase the sample of X-Ray Flashes[41,42] (XRFs) that may hold the key to the connection between SNIbc supernovae and classical long GRBs. This is also the guarantee to detect several high-redshift GRBs ($z \geq 5$) during the nominal mission,[1] despite the relatively small size of ECLAIRs. This is important because SVOM has the potential to quickly identify high-z GRBs as explained below.

The pointing strategy of the satellite will guarantee long periods of stable pointing, allowing the detection of long, faint transients like the ultra-long GRBs. A study by Dagoneau et al.[43] shows that *SVOM* may double the number of detected UL-GRBs (Figure 11b), allowing detailed multi-WL studies of these mysterious events and their host galaxies.

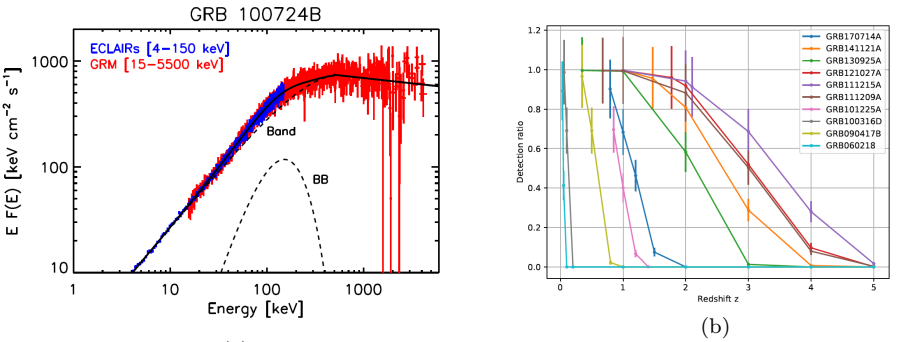
Finally, all ECLAIRs and GRM photons are sent to the ground, allowing the construction of various delayed off-line triggers that will benefit of more computation power than available onboard and a much better knowledge of the context (solar activity, particles, pre- and post- count-rate evolution...), and could be used to search for classical HE transients with better sensitivity as well as new types of HE transients.

3.2. *Physics of the Prompt Emission*

The unique combination of observations with GRM, ECLAIRs and GWAC for some GRBs will provide a unique coverage of the prompt emission. The systematic broadband time-resolved spectroscopy of the prompt GRB emission with ECLAIRs+GRM (and occasionally GWAC) will offer a renewed view of this phase, allowing to disentangle the role of the photosphere from internal shocks and other dissipation processes[26] (Figure 11a). For some GRBs, the prompt coverage may be extended to VHE gamma-rays with Fermi/LAT and CTA, allowing crucial diagnostic on the physics of GRB relativistic jets.[44,45]

3.3. *Physics of the Afterglow*

SVOM has two features that will permit to study the physics of GRB afterglows systematically with more detailed data. The first one is a good coverage of the prompt-afterglow transition for GRBs longer than 2-3 minutes, in X-rays thanks to



(a) (b)

Fig. 11. Predicted *SVOM* performance. (a) Joint reconstruction of GRB spectra with ECLAIRs and GRM.[26] (b) ECLAIRs sensitivity to ultra-long GRBs.[43]

the overlap of ECLAIRs and MXT energy ranges, and in the optical thanks to the combination of GWAC + GFT + VT. The second is the multi-wavelength coverage of the afterglow with MXT, VT and GFT, with a good sensitivity balance between VT & MXT, ensuring that $\sim 70\%$ of the afterglows of *SVOM* GRBs will be detected by both instruments. In addition, it is expected that most *SVOM* GRBs will have their redshift measured (see below), giving us access to the energetics of the bursts.

3.4. *Redshifts of GRBs*

SVOM will provide arcsecond GRB localizations in less than 5 minutes with the GFTs and the VT, and the majority of *SVOM* GRBs will be immediately observable with telescopes in the night hemisphere. These localizations will permit fast spectroscopic observations with mid- or large-size telescopes with the goal of measuring the redshifts of a large fraction ($\geq 2/3$) of *SVOM* GRBs. *SVOM* will also provide quick indications of intermediate redshift GRBs[46] ($3 \leq z \leq 6$), and of high-z GRBs candidates, when the afterglow is undetected in the VT. In all cases, dark bursts in the VT will indicate GRBs of interest, deserving additional observations with powerful telescopes.

3.5. *SN-less GRBs*

The detection of two SN-less GRBs with *Swift* in 2006[47–50] raised a lot of attention, since the association of long GRBs with SNIbc was by then considered secure. In 2011, another SN-less GRB was detected by *Swift*: GRB 111005A[51] at 57 Mpc (z=0.01326). These events raise the question of the nature of SN-less GRBs: are they the result of the explosion of the core of a massive star or a merger? Since GRB 111005A was well within the detection volume of GW detectors for NSNS or BHNS mergers and easily detectable with ECLAIRs (Figure 12), the occurrence of

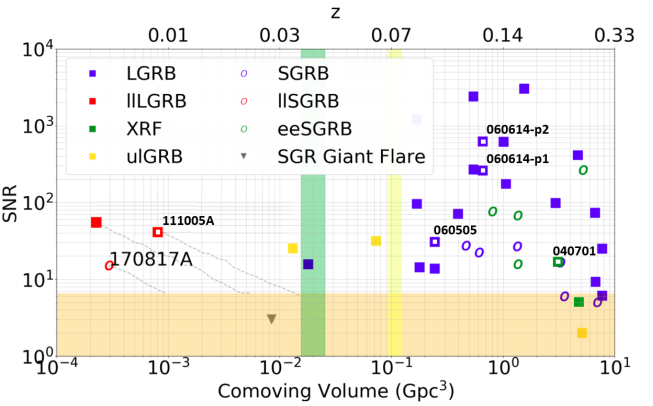


Fig. 12. ECLAIRs on-axis count SNR for GRBs in the local universe ($z \leq 0.3$). The orange horizontal band represents the detection limit of ECLAIRs at SNR = 6.5. The green and yellow bands represent the O4 LIGO detection limits for NS-NS and BH-NS mergers respectively. The light grey trails represent the evolution of the on-axis count SNR with the redshift.[52]

another event of this type within the field of view of ECLAIRs during the operation of GW interferometers would provide crucial data to solve this mystery.[52]

4. Conclusion

The operation of *SVOM* simultaneously with powerful instruments in the 2020's decade will certainly bring new crucial data for our understanding of GRBs. We have already mentioned the diagnostic brought by gravitational waves for nearby events, like short GRBs, sub-luminous GRBs or XRFs. We can also mention the search for GRBs coincident in time and direction with afterglow candidates that will be detected by the Vera Rubin Observatory. The operation of *SVOM* simultaneously with large neutrino detectors like ICECUBE or KM3NeT may also bring its share of surprises. Last but not least, the discovery of VHE gamma-ray emission from GRBs by HESS and MAGIC raises the hope to increase the number of such detections with CTA in the SVOM era, revealing key information about the composition and physical processes at work in ultrarelativistic jets.

Observing GRBs, AGNs, TDEs and transient sources of gravitational waves, SVOM will become a key player in the fields of High-Energy Astrophysics, Time Domain Astronomy and Multi-Messenger Astrophysics, after its launch mid-2023.

References

1. J. Wei, B. Cordier, S. Antier, P. Antilogus, J. L. Atteia, A. Bajat, S. Basa, V. Beckmann, M. G. Bernardini, S. Boissier, L. Bouchet, V. Burwitz, A. Claret, Z. G. Dai, F. Daigne, J. Deng, D. Dornic, H. Feng, T. Foglizzo, H. Gao, N. Gehrels, O. Godet, A. Goldwurm, F. Gonzalez, L. Gosset, D. Götz, C. Gouiffes, F. Grise, A. Gros,

J. Guilet, X. Han, M. Huang, Y. F. Huang, M. Jouret, A. Klotz, O. La Marle, C. Lachaud, E. Le Floch, W. Lee, N. Leroy, L. X. Li, S. C. Li, Z. Li, E. W. Liang, H. Lyu, K. Mercier, G. Migliori, R. Mochkovitch, P. O'Brien, J. Osborne, J. Paul, E. Perinati, P. Petitjean, F. Piron, Y. Qiu, A. Rau, J. Rodriguez, S. Schanne, N. Tanvir, E. Vangioni, S. Vergani, F. Y. Wang, J. Wang, X. G. Wang, X. Y. Wang, A. Watson, N. Webb, J. J. Wei, R. Willingale, C. Wu, X. F. Wu, L. P. Xin, D. Xu, S. Yu, W. F. Yu, Y. W. Yu, B. Zhang, S. N. Zhang, Y. Zhang and X. L. Zhou, The Deep and Transient Universe in the SVOM Era: New Challenges and Opportunities - Scientific prospects of the SVOM mission, *arXiv e-prints* , p. arXiv:1610.06892 (October 2016).

2. R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, A. Allocca, P. A. Altin, A. Amato, S. Anand, A. Ananyeva, S. B. Anderson, W. G. Anderson, S. V. Angelova, S. Ansoldi, J. M. Antelis, S. Antier, S. Appert, K. Arai, M. C. Araya, J. S. Areeda, M. Arène, N. Arnaud, S. M. Aronson, K. G. Arun, Y. Asali, S. Ascenzi, G. Ashton, S. M. Aston, P. Astone, F. Aubin, P. Aufmuth, K. AultONeal, C. Austin, V. Avendano, S. Babak, F. Badaracco, M. K. M. Bader, S. Bae, A. M. Baer, S. Bagnasco, J. Baird, M. Ball, G. Ballardin, S. W. Ballmer, A. Bals, A. Balsamo, G. Baltus, S. Banagiri, D. Bankar, R. S. Bankar, J. C. Barayoga, C. Barbieri, B. C. Barish, D. Barker, P. Barneo, S. Barnum, F. Barone, B. Barr, L. Barsotti, M. Barsuglia, D. Barta, J. Bartlett, I. Bartos, R. Bassiri, A. Basti, M. Bawaj, J. C. Bayley, M. Bazzan, B. R. Becher, B. Bécsy, V. M. Bedakihale, M. Bejger, I. Belahcene, D. Beniwal, M. G. Benjamin, T. F. Bennett, J. D. Bentley, F. Bergamin, B. K. Berger, G. Bergmann, S. Bernuzzi, C. P. L. Berry, D. Bersanetti, A. Bertolini, J. Betzwieser, R. Bhandare, A. V. Bhandari, D. Bhattacharjee, J. Bidler, I. A. Bilenko, G. Billingsley, R. Birney, O. Birnholtz, S. Biscans, M. Bischi, S. Biscoveanu, A. Bisht, M. Bitossi, M. A. Bizouard, J. K. Blackburn, J. Blackman, C. D. Blair, D. G. Blair, R. M. Blair, O. Blanch, F. Bobba, N. Bode, M. Boer, Y. Boetzel, G. Bogaert, M. Boldrini, F. Bondu, E. Bonilla, R. Bonnand, P. Booker, B. A. Boom, R. Bork, V. Boschi, S. Bose, V. Bossilkov, V. Boudart, Y. Bouffanais, A. Bozzi, C. Bradaschia, P. R. Brady, A. Bramley, M. Branchesi, J. E. Brau, M. Breschi, T. Briant, J. H. Briggs, F. Brighenti, A. Brillet, M. Brinkmann, P. Brockill, A. F. Brooks, J. Brooks, D. D. Brown, S. Brunett, G. Bruno, R. Bruntz, A. Buikema, T. Bulik, H. J. Bulten, A. Buonanno, R. Buscicchio, D. Buskulic, R. L. Byer, M. Cabero, L. Cadonati, M. Caesar, G. Cagnoli, C. Cahillane, J. Calderón Bustillo, J. D. Callaghan, T. A. Callister, E. Calloni, J. B. Camp, M. Canepa, K. C. Cannon, H. Cao, J. Cao, G. Carapella, F. Carbognani, M. F. Carney, M. Carpinelli, G. Carullo, T. L. Carver, J. Casanueva Diaz, C. Casentini, S. Caudill, M. Cavaglià, F. Cavalier, R. Cavalieri, G. Cella, P. Cerdá-Durán, E. Cesarini, W. Chaibi, K. Chakravarti, C. L. Chan, C. Chan, K. Chandra, P. Chanial, S. Chao, P. Charlton, E. A. Chase, E. Chassande-Mottin, D. Chatterjee, D. Chattopadhyay, M. Chaturvedi, K. Chatziioannou, A. Chen, H. Y. Chen, X. Chen, Y. Chen, H. P. Cheng, C. K. Cheong, H. Y. Chia, F. Chiadini, R. Chierici, A. Chincarini, A. Chiummo, G. Cho, H. S. Cho, M. Cho, S. Choate, N. Christensen, Q. Chu, S. Chua, K. W. Chung, S. Chung, G. Ciani, P. Ciecielag, M. Cieślar, M. Cifaldi, A. A. Ciobanu, R. Ciolfi, F. Cipriano, A. Cirone, F. Clara, E. N. Clark, J. A. Clark, L. Clarke, P. Clearwater, S. Clesse, F. Cleva, E. Coccia, P. F. Cohadon, D. E. Cohen, M. Colleoni, C. G. Collette, C. Collins, M. Colpi, M. Constancio, L. Conti, S. J. Cooper, P. Corban, T. R. Corbitt, I. Cordero-Carrión, S. Corezzi, K. R. Corley, N. Cornish, D. Corre, A. Corsi, S. Cortese, C. A. Costa, R. Cotesta, M. W. Coughlin, S. B. Coughlin, J. P. Coulon, S. T. Countryman, B. Cousins, P. Couvares, P. B. Covas, D. M.

Coward, M. J. Cowart, D. C. Coyne, R. Coyne, J. D. E. Creighton, T. D. Creighton, M. Croquette, S. G. Crowder, J. R. Cudell, T. J. Cullen, A. Cumming, R. Cummings, L. Cunningham, E. Cuoco, M. Curyło, T. D. Canton, G. Dálya, A. Dana, L. M. DaneshgaranBajastani, B. D'Angelo, B. Danila, S. L. Danilishin, S. D'Antonio, K. Danzmann, C. Darsow-Fromm, A. Dasgupta, L. E. H. Datrier, V. Dattilo, I. Dave, M. Davier, G. S. Davies, D. Davis, E. J. Daw, R. Dean, D. DeBra, M. Deenadayalan, J. Degallaix, M. De Laurentis, S. Deléglise, V. Del Favero, F. De Lillo, N. De Lillo, W. Del Pozzo, L. M. DeMarchi, F. De Matteis, V. D'Emilio, N. Demos, T. Denker, T. Dent, A. Depasse, R. De Pietri, R. De Rosa, C. De Rossi, R. DeSalvo, O. de Varona, S. Dhurandhar, M. C. Díaz, M. Diaz-Ortiz, N. A. Didio, T. Dietrich, L. Di Fiore, C. DiFronzo, C. Di Giorgio, F. Di Giovanni, M. Di Giovanni, T. Di Girolamo, A. Di Lieto, B. , GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run, *Physical Review X* **11**, p. 021053 (April 2021).

3. G. Vedrenne and J.-L. Atteia, *Gamma-Ray Bursts* 2009.
4. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, G. Allen, A. Allocca, P. A. Altin, A. Amato, A. Ananyeva, S. B. Anderson, W. G. Anderson, S. V. Angelova, S. Antier, S. Appert, K. Arai, M. C. Araya, J. S. Areeda, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, M. Ast, S. M. Aston, P. Astone, D. V. Atallah, P. Aufmuth, C. Aulbert, K. AultONeal, C. Austin, A. Avila-Alvarez, S. Babak, P. Bacon, M. K. M. Bader, S. Bae, P. T. Baker, F. Baldaccini, G. Ballardin, S. W. Ballmer, S. Banagiri, J. C. Barayoga, S. E. Barclay, B. C. Barish, D. Barker, K. Barkett, F. Barone, B. Barr, L. Barsotti, M. Barsuglia, D. Barta, S. D. Barthelmy, J. Bartlett, I. Bartos, R. Bassiri, A. Basti, J. C. Batch, M. Bawaj, J. C. Bayley, M. Bazzan, B. Bécsy, C. Beer, M. Bejger, I. Belahcene, A. S. Bell, B. K. Berger, G. Bergmann, J. J. Bero, C. P. L. Berry, D. Bersanetti, A. Bertolini, J. Betzwieser, S. Bhagwat, R. Bhandare, I. A. Bilenko, G. Billingsley, C. R. Billman, J. Birch, R. Birney, O. Birnholtz, S. Biscans, S. Biscoveanu, A. Bisht, M. Bitossi, C. Biwer, M. A. Bizouard, J. K. Blackburn, J. Blackman, C. D. Blair, D. G. Blair, R. M. Blair, S. Bloemen, O. Bock, N. Bode, M. Boer, G. Bogaert, A. Bohe, F. Bondu, E. Bonilla, R. Bonnand, B. A. Boom, R. Bork, V. Boschi, S. Bose, K. Bossie, Y. Bouffanais, A. Bozzi, C. Bradaschia, P. R. Brady, M. Branchesi, J. E. Brau, T. Briant, A. Brillet, M. Brinkmann, V. Brisson, P. Brockill, J. E. Broida, A. F. Brooks, D. A. Brown, D. D. Brown, S. Brunett, C. C. Buchanan, A. Buikema, T. Bulik, H. J. Bulten, A. Buonanno, D. Buskulic, C. Buy, R. L. Byer, M. Cabero, L. Cadonati, G. Cagnoli, C. Cahillane, J. Calderón Bustillo, T. A. Callister, E. Calloni, J. B. Camp, M. Canepa, P. Canizares, K. C. Cannon, H. Cao, J. Cao, C. D. Capano, E. Capocasa, F. Carbognani, S. Caride, M. F. Carney, J. Casanueva Diaz, C. Casentini, S. Caudill, M. Cavaglià, F. Cavalier, R. Cavalieri, G. Cella, C. B. Cepeda, P. Cerdá-Durán, G. Cerretani, E. Cesarini, S. J. Chamberlin, M. Chan, S. Chao, P. Charlton, E. Chase, E. Chassande-Mottin, D. Chatterjee, K. Chatziioannou, B. D. Cheeseboro, H. Y. Chen, X. Chen, Y. Chen, H. P. Cheng, H. Chia, A. Chincarini, A. Chiummo, T. Chmiel, H. S. Cho, M. Cho, J. H. Chow, N. Christensen, Q. Chu, A. J. K. Chua, S. Chua, A. K. W. Chung, S. Chung, G. Ciani, R. Ciolfi, C. E. Cirelli, A. Cirone, F. Clara, J. A. Clark, P. Clearwater, F. Cleva, C. Cocchieri, E. Coccia, P. F. Cohadon, D. Cohen, A. Colla, C. G. Collette, L. R. Cominsky, J. Constancio, M., L. Conti, S. J. Cooper, P. Corban, T. R. Corbitt, I. Cordero-Carrión, K. R. Corley, N. Cornish, A. Corsi, S. Cortese, C. A. Costa, M. W. Coughlin, S. B. Coughlin, J. P. Coulon, S. T. Countryman, P. Couvares, P. B. Covas,

E. E. Cowan, D. M. Coward, M. J. Cowart, D. C. Coyne, R. Coyne, J. D. E. Creighton, T. D. Creighton, J. Cripe, S. G. Crowder, T. J. Cullen, A. Cumming, L. Cunningham, E. Cuoco, T. Dal Canton, G. Dálya, S. L. Danilishin, S. D'Antonio, K. Danzmann, A. Dasgupta, C. F. Da Silva Costa, V. Dattilo, I. Dave, M. Davier, D. Davis, E. J. Daw, B. Day, S. De, D. DeBra, J. Degallaix, M. De Laurentis, S. Deléglise, W. Del Pozzo, N. Demos, T. Denker, T. Dent, R. De Pietri, V. Dergachev, R. De Rosa, R. T. DeRosa, C. De Rossi, R. DeSalvo, O. de Varona, J. Devenson, S. Dhurandhar, M. C. Díaz, L. Di Fiore, M. Di Giovanni, T. Di Girolamo, A. Di Lieto, S. Di Pace, I. Di Palma, F. Di Renzo, Z. Doctor, V. Dolique, F. Donovan, K. L. Dooley, S. Doravari, I. Dorrington, R. Douglas, M. Dovale Álvarez, T. P. Downes, M. Drago, C. Dreissigacker, J. C. Driggers, Z. Du, M. Ducrot, P. Dupej, S. E. Dwyer, T. B. Edo, M. C. Edwards, A. Effler, P. Ehrens, J. Eichholz, S. S. Eikenberry, R. A. Eisenstein, R. C. Essick, D. Estevez, Z. B. Etienne, T. Etzel, M. Evans, T. M. Evans, M. Factourovich, V. Fafone, H. Fair, S. Fairhurst, X. Fan, S. Farinon, B. Farr, W. M. Farr, E. J. Fauchon-Jones, M. Favata, M. Fays, C. Fee, H. Fehrmann, J. Feicht, M. M. Fejer, A. Fernandez-Galiana, I. Ferrante, E. C. Ferreira, F. Ferrini, F. Fidecaro, D. Finstad, I. Fiori, D. Fiorucci, M. Fishbach, R. P. Fi, Multi-messenger Observations of a Binary Neutron Star Merger, *ApJ* **848**, p. L12 (October 2017).

5. Cherenkov Telescope Array Consortium, B. S. Acharya, I. Agudo, I. Al Samarai, R. Alfaro, J. Alfaro, C. Alispach, R. Alves Batista, J. P. Amans, E. Amato, G. Ambrosi, E. Antolini, L. A. Antonelli, C. Aramo, M. Araya, T. Armstrong, F. Arqueros, L. Arrabito, K. Asano, M. Ashley, M. Backes, C. Balazs, M. Balbo, O. Ballester, J. Ballet, A. Bamba, M. Barkov, U. Barres de Almeida, J. A. Barrio, D. Bastieri, Y. Becherini, A. Belfiore, W. Benbow, D. Berge, E. Bernardini, M. G. Bernardini, M. Bernardos, K. Bernlöhr, B. Bertucci, B. Biasuzzi, C. Bigongiari, A. Biland, E. Bissaldi, J. Biteau, O. Blanch, J. Blazek, C. Boisson, J. Bolmont, G. Bonanno, A. Bonardi, C. Bonavolontà, G. Bonnoli, Z. Bosnjak, M. Böttcher, C. Braiding, J. Bregeon, A. Brill, A. M. Brown, P. Brun, G. Brunetti, T. Buanes, J. Buckley, V. Bugaev, R. Bühler, A. Bulgarelli, T. Bulik, M. Burton, A. Burtovoi, G. Busetto, R. Canestrari, M. Capalbi, F. Capitanio, A. Caproni, P. Caraveo, V. Cárdenas, C. Carlile, R. Carosi, E. Carquín, J. Carr, S. Casanova, E. Cascone, F. Catalani, O. Catalano, D. Cauz, M. Cerruti, P. Chadwick, S. Chaty, R. C. G. Chaves, A. Chen, X. Chen, M. Chernyakova, M. Chikawa, A. Christov, J. Chudoba, M. Cieślar, V. Coco, S. Colafrancesco, P. Colin, V. Conforti, V. Connaughton, J. Conrad, J. L. Contreras, J. Cortina, A. Costa, H. Costantini, G. Cotter, S. Covino, R. Crocker, J. Cuadra, O. Cuevas, P. Cumani, A. D'Aì, F. D'Ammando, P. D'Avanzo, D. D'Urso, M. Daniel, I. Davids, B. Dawson, F. Dazzi, A. De Angelis, R. de Cássia dos Anjos, G. De Cesare, A. De Franco, E. M. de Gouveia Dal Pino, I. de la Calle, R. de los Reyes Lopez, B. De Lotto, A. De Luca, M. De Lucia, M. de Naurois, E. de Oña Wilhelmi, F. De Palma, F. De Persio, V. de Souza, C. Deil, M. Del Santo, C. Delgado, D. della Volpe, T. Di Girolamo, F. Di Pierro, L. Di Venere, C. Díaz, C. Dib, S. Diebold, A. Djannati-Ataï, A. Domínguez, D. Dominis Prester, D. Dorner, M. Doro, H. Drass, D. Dravins, G. Dubus, V. V. Dwarkadas, J. Ebr, C. Eckner, K. Egberts, S. Einecke, T. R. N. Ekoume, D. Elsässer, J. P. Ernenwein, C. Espinoza, C. Evoli, M. Fairbairn, D. Falceta-Goncalves, A. Falcone, C. Farnier, G. Fasola, E. Fedorova, S. Fegan, M. Fernandez-Alonso, A. Fernández-Barral, G. Ferrand, M. Fesquet, M. Filipovic, V. Fioretti, G. Fontaine, M. Fornasa, L. Fortson, L. Freixas Coromina, C. Fruck, Y. Fujita, Y. Fukazawa, S. Funk, M. Füßling, S. Gabici, A. Gadola, Y. Gallant, B. Garcia, R. Garcia López, M. Garczarczyk, J. Gaskins, T. Gasparetto, M. Gaug, L. Gerard, G. Giavitto, N. Giglietto, P. Giommi, F. Giordano, E. Giro, M. Giroletti, A. Giuliani,

J. F. Glicenstein, R. Gnatyk, N. Godinovic, P. Goldoni, G. Gómez-Vargas, M. M. González, J. M. González, D. Götz, J. Graham, P. Grandi, J. Granot, A. J. Green, T. Greenshaw, S. Griffiths, S. Gunji, D. Hadasch, S. Hara, M. J. Hardcastle, T. Hassan, K. Hayashi, M. Hayashida, M. Heller, J. C. Helo, G. Hermann, J. Hinton, B. Hnatyk, W. Hofmann, J. Holder, D. Horan, J. Hörandel, D. Horns, P. Horvath, T. Hovatta, M. Hrabovsky, D. Hrupec, T. B. Humensky, M. Hütten, M. Iarlori, T. Inada, Y. Inome, S. Inoue, T. Inoue, Y. Inoue, F. Iocco, K. Ioka, M. Iori, K. Ishio, Y. Iwamura, M. Jamrozy, P. Janecek, D. Jankowsky, P. Jean, I. Jung-Richardt, J. Jurysek, P. Kaaret, S. Karkar, H. Katagiri, U. Katz, N. Kawanaka, D. Kazanas, B. Khélifi, D. B. Kieda, S. Kimeswenger, S. Kimura, S. Kisaka, J. Knapp, J. Knödlseder, B. Koch, K. Kohri, N. Komin, K. Kosack, M. Kraus, M. Krause, F. Krauß, H. Kubo, G. Kukec Mezek, H. Kuroda, J. Kushida, N. La Palombara, G. Lamanna, R. G. Lang, J. Lapington, O. Le Blanc, S. Leach, J. P. Lees, J. Lefaucheur, M. A. Leigui de Oliveira, J. P. Lenain, R. Lico, M. Limon, E. Lindfors, T. Lohse, S. Lombardi, F. Longo, M. López, R. López-Coto, C. C. Lu, F. Lucarelli, P. L. Luque-Escamilla, E. Lyard, M. C. Maccarone, G. Maier, P. Majumdar, G. Malaguti, D. Mandat, G. Maneva, M. Manganaro, S. Mangano, A. Marcowith, J. Marín, S. Markoff, J. Martí, P. Martin, M. Martínez, G. Martínez, N. Masetti, S. Masuda, G. Maurin, N. Maxted, D. Mazin, C. Medina, A. Melandri, S. Mereghetti, M. Meyer, I. A. Minaya, N. Mirabal, R. Mirzoyan, A. Mitchell, T. Mizuno, R. Moderski, M. M, *Science with the Cherenkov Telescope Array* 2019.

6. A. Albert, R. Alfaro, C. Alvarez, J. R. A. Camacho, J. C. Arteaga-Velázquez, K. P. Arunbabu, D. Avila Rojas, H. A. Ayala Solares, V. Baghmanyan, E. Belmont-Moreno, S. Y. BenZvi, C. Brisbois, K. S. Caballero-Mora, T. Capistrán, A. Carramiñana, S. Casanova, U. Cotti, S. Coutiño de León, E. De la Fuente, R. Diaz Hernandez, L. Diaz-Cruz, B. L. Dingus, M. A. DuVernois, M. Durocher, J. C. Díaz-Vélez, R. W. Ellsworth, K. Engel, C. Espinoza, K. L. Fan, K. Fang, M. F. Alonso, H. Fleischhack, N. Fraija, A. Galván-Gámez, D. Garcia, J. A. García-González, F. Garfias, G. Giacinti, M. M. González, J. A. Goodman, J. P. Harding, S. Hernandez, J. Hinton, B. Hona, D. Huang, F. Hueyotl-Zahuantitla, P. Hüntemeyer, A. Iriarte, A. Jardin-Blicq, V. Joshi, D. Kieda, A. Lara, W. H. Lee, H. León Vargas, J. T. Linnemann, A. L. Longinotti, G. Luis-Raya, J. Lundeen, R. López-Coto, K. Malone, V. Marandon, O. Martinez, I. Martinez-Castellanos, J. Martínez-Castro, J. A. Matthews, P. Miranda-Romagnoli, J. A. Morales-Soto, E. Moreno, M. Mostafá, A. Nayerhoda, L. Nellen, M. Newbold, M. U. Nisa, R. Noriega-Papaqui, L. Olivera-Nieto, N. Omodei, A. Peisker, Y. Pérez Araujo, E. G. Pérez-Pérez, Z. Ren, C. D. Rho, C. Rivière, D. Rosa-González, E. Ruiz-Velasco, H. Salazar, F. Salesa Greus, A. Sandoval, M. Schneider, H. Schoorlemmer, F. Serna, G. Sinnis, A. J. Smith, R. W. Springer, P. Surajbali, K. Tollefson, I. Torres, R. Torres-Escobedo, T. N. Ukwatta, F. Ureña-Mena, T. Weisgarber, F. Werner, E. Willox, A. Zepeda, H. Zhou, C. de León, J. D. Álvarez and HAWC Collaboration, 3HWC: The Third HAWC Catalog of Very-high-energy Gamma-Ray Sources, *ApJ* **905**, p. 76 (December 2020).
7. X. Bai, B. Y. Bi, X. J. Bi, Z. Cao, S. Z. Chen, Y. Chen, A. Chiavassa, X. H. Cui, Z. G. Dai, D. della Volpe, T. Di Girolamo, G. Di Sciascio, Y. Z. Fan, J. Giacalone, Y. Q. Guo, H. H. He, T. L. He, M. Heller, D. Huang, Y. F. Huang, H. Jia, L. T. Ksenofontov, D. Leahy, F. Li, Z. Li, E. W. Liang, P. Lipari, R. Y. Liu, Y. Liu, S. Liu, X. Ma, O. Martineau-Huynh, D. Martraire, T. Montaruli, D. Ruffolo, Y. V. Stenkin, H. Q. Su, T. Tam, Q. W. Tang, W. W. Tian, P. Vallania, S. Vernetto, C. Vigorito, J. . C. Wang, L. Z. Wang, X. Wang, X. Y. Wang, X. J. Wang, Z. X. Wang, D. M. Wei, J. J. Wei, D. Wu, H. R. Wu, X. F. Wu, D. H. Yan, A. Y. Yang, R. Z. Yang,

Z. G. Yao, L. Q. Yin, Q. Yuan, B. Zhang, B. Zhang, L. Zhang, M. F. Zhang, S. S. Zhang, X. Zhang, Y. Zhao, X. X. Zhou, F. R. Zhu and H. Zhu, The Large High Altitude Air Shower Observatory (LHAASO) Science White Paper, *arXiv e-prints* , p. arXiv:1905.02773 (May 2019).

8. P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, H. Böhringer, A. Bogomolov, T. Boller, K. Borm, W. Bornemann, H. Bräuninger, M. Brüggen, H. Brunner, M. Brusa, E. Bulbul, M. Buntov, V. Burwitz, W. Burkert, N. Clerc, E. Churazov, D. Coutinho, T. Dauser, K. Dennerl, V. Doroshenko, J. Eder, V. Emberger, T. Eraerds, A. Finoguenov, M. Freyberg, P. Friedrich, S. Friedrich, M. Fürmetz, A. Georgakakis, M. Gilfanov, S. Granato, C. Grossberger, A. Gueguen, P. Gureev, F. Haberl, O. Hälker, G. Hartner, G. Hasinger, H. Huber, L. Ji, A. v. Kienlin, W. Kink, F. Korotkov, I. Kreykenbohm, G. Lamer, I. Lomakin, I. Lapshov, T. Liu, C. Maitra, N. Meidinger, B. Menz, A. Merloni, T. Mernik, B. Mican, J. Mohr, S. Müller, K. Nandra, V. Nazarov, F. Pacaud, M. Pavlinsky, E. Perinati, E. Pfeffermann, D. Pietschner, M. E. Ramos-Ceja, A. Rau, J. Reiffers, T. H. Reiprich, J. Robrade, M. Salvato, J. Sanders, A. Santangelo, M. Sasaki, H. Scheuerle, C. Schmid, J. Schmitt, A. Schwope, A. Shirshakov, M. Steinmetz, I. Stewart, L. Strüder, R. Sunyaev, C. Tenzer, L. Tiedemann, J. Trümper, V. Voron, P. Weber, J. Wilms and V. Yaroshenko, The eROSITA X-ray telescope on SRG, *A&A* **647**, p. A1 (March 2021).
9. E. C. Bellm, S. R. Kulkarni, M. J. Graham, R. Dekany, R. M. Smith, R. Riddle, F. J. Masci, G. Helou, T. A. Prince, S. M. Adams, C. Barbarino, T. Barlow, J. Bauer, R. Beck, J. Belicki, R. Biswas, N. Blagorodnova, D. Bodewits, B. Bolin, V. Brinnel, T. Brooke, B. Bue, M. Bulla, R. Burruss, S. B. Cenko, C.-K. Chang, A. Connolly, M. Coughlin, J. Cromer, V. Cunningham, K. De, A. Delacroix, V. Desai, D. A. Duev, G. Eadie, T. L. Farnham, M. Feeney, U. Feindt, D. Flynn, A. Franckowiak, S. Frederick, C. Fremling, A. Gal-Yam, S. Gezari, M. Giomi, D. A. Goldstein, V. Z. Golkhou, A. Goobar, S. Groom, E. Hacopians, D. Hale, J. Henning, A. Y. Q. Ho, D. Hover, J. Howell, T. Hung, D. Huppenkothen, D. Imel, W.-H. Ip, Ž. Ivezić, E. Jackson, L. Jones, M. Juric, M. M. Kasliwal, S. Kaspi, S. Kaye, M. S. P. Kelley, M. Kowalski, E. Kramer, T. Kupfer, W. Landry, R. R. Laher, C.-D. Lee, H. W. Lin, Z.-Y. Lin, R. Lunnan, M. Giomi, A. Mahabal, P. Mao, A. A. Miller, S. Monkewitz, P. Murphy, C.-C. Ngeow, J. Nordin, P. Nugent, E. Ofek, M. T. Patterson, B. Penprase, M. Porter, L. Rauch, U. Rebbapragada, D. Reiley, M. Rigault, H. Rodriguez, J. van Roestel, B. Rusholme, J. van Santen, S. Schulze, D. L. Shupe, L. P. Singer, M. T. Soumagnac, R. Stein, J. Surace, J. Sollerman, P. Szkody, F. Taddia, S. Terek, A. Van Sistine, S. van Velzen, W. T. Vestrand, R. Walters, C. Ward, Q.-Z. Ye, P.-C. Yu, L. Yan and J. Zolkower, The Zwicky Transient Facility: System Overview, Performance, and First Results, *PASP* **131**, p. 018002 (January 2019).
10. Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, J. R. P. Angel, G. Z. Angeli, R. Ansari, P. Antilogus, C. Araujo, R. Armstrong, K. T. Arndt, P. Astier, É. Aubourg, N. Auza, T. S. Axelrod, D. J. Bard, J. D. Barr, A. Barrau, J. G. Bartlett, A. E. Bauer, B. J. Bauman, S. Baumont, E. Bechtol, K. Bechtol, A. C. Becker, J. Becla, C. Beldica, S. Bellavia, F. B. Bianco, R. Biswas, G. Blanc, J. Blazek, R. D. Blandford, J. S. Bloom, J. Bogart, T. W. Bond, M. T. Booth, A. W. Borgland, K. Borne, J. F. Bosch, D. Boutigny, C. A. Brackett, A. Bradshaw, W. N. Brandt, M. E. Brown, J. S. Bullock, P. Burchat, D. L. Burke, G. Cagnoli, D. Calabrese, S. Callahan, A. L. Callen, J. L. Carlin, E. L. Carlson, S. Chandrasekharan, G. Charles-Emerson, S. Chesley, E. C. Cheu, H.-F. Chiang, J. Chiang, C. Chirino, D. Chow, D. R. Ciardi, C. F. Claver, J. Cohen-Tanugi, J. J.

Cockrum, R. Coles, A. J. Connolly, K. H. Cook, A. Cooray, K. R. Covey, C. Cribbs, W. Cui, R. Cutri, P. N. Daly, S. F. Daniel, F. Daruich, G. Daubard, G. Daues, W. Dawson, F. Delgado, A. Dellapenna, R. de Peyster, M. de Val-Borro, S. W. Digel, P. Doherty, R. Dubois, G. P. Dubois-Felsmann, J. Durech, F. Economou, T. Eifler, M. Eracleous, B. L. Emmons, A. Fausti Neto, H. Ferguson, E. Figueroa, M. Fisher-Levine, W. Focke, M. D. Foss, J. Frank, M. D. Freemon, E. Gangler, E. Gawiser, J. C. Geary, P. Gee, M. Geha, C. J. B. Gessner, R. R. Gibson, D. K. Gilmore, T. Glanzman, W. Glick, T. Goldina, D. A. Goldstein, I. Goodenow, M. L. Graham, W. J. Gressler, P. Gris, L. P. Guy, A. Guyonnet, G. Haller, R. Harris, P. A. Hascall, J. Haupt, F. Hernandez, S. Herrmann, E. Hileman, J. Hoblitt, J. A. Hodgson, C. Hogan, J. D. Howard, D. Huang, M. E. Huffer, P. Ingraham, W. R. Innes, S. H. Jacoby, B. Jain, F. Jammes, M. J. Jee, T. Jenness, G. Jernigan, D. Jevremović, K. Johns, A. S. Johnson, M. W. G. Johnson, R. L. Jones, C. Juramy-Gilles, M. Jurić, J. S. Kalirai, N. J. Kallivayalil, B. Kalmbach, J. P. Kantor, P. Karst, M. M. Kasliwal, H. Kelly, R. Kessler, V. Kinnison, D. Kirkby, L. Knox, I. V. Kotov, V. L. Krabbendam, K. S. Krughoff, P. Kubánek, J. Kuczewski, S. Kulkarni, J. Ku, N. R. Kurita, C. S. Lage, R. Lambert, T. Lange, J. B. Langton, L. Le Guillou, D. Levine, M. Liang, K.-T. Lim, C. J. Lintott, K. E. Long, M. Lopez, P. J. Lotz, R. H. Lupton, N. B. Lust, L. A. MacArthur, A. Mahabal, R. Mandelbaum, T. W. Markiewicz, D. S. Marsh, P. J. Marshall, S. Marshall, M. May, R. McKercher, M. McQueen, J. Meyers, M. Migliore, M. Miller, D. J. Mills, C. Miraval, J. Moeyens, F. E. Moolekamp, D. G. Monet, M. Moniez, S. Monkewitz, C. Montgomery, C. B. Morrison, F. Mueller, G. P. Muller, F. Muñoz Arancibia, D. R. Neill, S. P. Newbry, J.-Y. Nief, A. Nomerotski, M. Nordby, P. O'Connor, J. Oliver, S. S. Olivier, K. Olsen, W. O'Mullane, S. Ortiz, S. Osier, R. E. Owen, R. Pain, P. E. Palecek, J. K. Parejko, J. B. Parsons, N. M. Pease, J. M. Peterson, J. R. Peterson, D. L. Petravick, M. E. Libby Petrick, C. E. Petry, F. Pierfederici, S. Pietrowicz, R. Pike, P. A. Pinto, R. Plante, S. Plate, J. P. Plutchak, P. A. Price, M. Prouza, V. Radeka, J. Rajagopal, A. P. Rasmussen, N. Regnault, K. A. Reil, D. J. Reiss, M. A. Reuter, S. T. Ridgway, V. J. Riot, S. Ritz, S. Robinson, W. Roby, A. Roodman, W. Rosing, C. Roucelle, M. R. Rumore, S. Russo, A. Saha, B. Sassolas, T. L. Schalk, P. Schellart, R. H. Schindler, S. Schmidt, D. P. Schneider, M. D. Schneider, W. Schoening, G. Schumacher, M. E. Schwamb, J. Sebag, B. Selvy, G. H. Sembroski, L. G. Seppala, A. Serio, E. Serrano, R. A. Shaw, I. Shipsey, J. Sick, N. Silvestri, C. T. Slater, J. A. Smith, R. C. Smith, S. Sobhani, C. Soldahl, L. Storrie-Lombardi, E. Stover, M. A. Strauss, R. A. Street, C. W. Stubbs, I. S. Sullivan, D. Sweeney, J. D. Swinbank, A. Szalay, P. Takacs, S. A. Tether, J. J. Thaler, J. G. Thayer, S. Thomas, A. J. Thornton, V. Thukral, J. Tice, D. E. Trilling, M. Turri, R. Van Berg, D. Vanden Berk, K. Vetter, F. Virieux, T. Vucina, W. Wahl, L. Walkowicz, B. Walsh, C. W. Walter, D. L. Wang, S.-Y. Wang, M. Warner, O. Wiecha, B. Willman, S. E. Winters, D. Wittman, S. C. Wolff, W. M. Wood-Vasey, X. Wu, B. Xin, P. Yoachim and H. Zhan, LSST: From Science Drivers to Reference Design and Anticipated Data Products, *ApJ* **873**, p. 111 (March 2019).

11. P. Schipani, R. Claudi, S. Campana, A. Baruffolo, S. Basa, S. Basso, E. Cappellaro, E. Cascone, R. Cosentino, F. D'Alessio, V. De Caprio, M. Della Valle, A. de Ugarte Postigo, S. D'Orsi, R. Franzen, J. Fynbo, A. Gal-Yam, D. Gardiol, E. Giro, M. Hamuy, M. Iuzzolino, D. Loreggia, S. Mattila, M. Munari, G. Pignata, M. Riva, S. Savarese, B. Schmidt, S. Scuderi, S. Smartt and F. Vitali, The new SOXS instrument for the ESO NTT, **9908**, p. 990841 (August 2016).
12. S. Johnston, R. Taylor, M. Bailes, N. Bartel, C. Baugh, M. Bietenholz, C. Blake, R. Braun, J. Brown, S. Chatterjee, J. Darling, A. Deller, R. Dodson, P. Edwards,

R. Ekers, S. Ellingsen, I. Feain, B. Gaensler, M. Haverkorn, G. Hobbs, A. Hopkins, C. Jackson, C. James, G. Joncas, V. Kaspi, V. Kilborn, B. Koribalski, R. Kothes, T. Landecker, E. Lenc, J. Lovell, J. P. Macquart, R. Manchester, D. Matthews, N. McClure-Griffiths, R. Norris, U. L. Pen, C. Phillips, C. Power, R. Protheroe, E. Sadler, B. Schmidt, I. Stairs, L. Staveley-Smith, J. Stil, S. Tingay, A. Tzioumis, M. Walker, J. Wall and M. Wolleben, Science with ASKAP. The Australian square-kilometre-array pathfinder, *Experimental Astronomy* **22**, 151 (December 2008).

13. R. S. Booth, W. J. G. de Blok, J. L. Jonas and B. Fanaroff, MeerKAT Key Project Science, Specifications, and Proposals, *arXiv e-prints* , p. arXiv:0910.2935 (October 2009).
14. CHIME/FRB Collaboration, M. Amiri, K. Bandura, P. Berger, M. Bhardwaj, M. M. Boyce, P. J. Boyle, C. Brar, M. Burhanpurkar, P. Chawla, J. Chowdhury, J. F. Cliche, M. D. Cranmer, D. Cubranic, M. Deng, N. Denman, M. Dobbs, M. Fandino, E. Fonseca, B. M. Gaensler, U. Giri, A. J. Gilbert, D. C. Good, S. Guliani, M. Halpern, G. Hinshaw, C. Höfer, A. Josephy, V. M. Kaspi, T. L. Landecker, D. Lang, H. Liao, K. W. Masui, J. Mena-Parra, A. Naidu, L. B. Newburgh, C. Ng, C. Patel, U. L. Pen, T. Pinsonneault-Marotte, Z. Pleunis, M. Rafiei Ravandi, S. M. Ransom, A. Renard, P. Scholz, K. Sigurdson, S. R. Siegel, K. M. Smith, I. H. Stairs, S. P. Tendulkar, K. Vanderlinde and D. V. Wiebe, The CHIME Fast Radio Burst Project: System Overview, *ApJ* **863**, p. 48 (August 2018).
15. J.-P. Macquart, M. Bailes, N. D. R. Bhat, G. C. Bower, J. D. Bunton, S. Chatterjee, T. Colegate, J. M. Cordes, L. D'Addario, A. Deller, R. Dodson, R. Fender, K. Haines, P. Halll, C. Harris, A. Hotan, S. Johnston, D. L. Jones, M. Keith, J. Y. Koay, T. J. W. Lazio, W. Majid, T. Murphy, R. Navarro, C. Phillips, P. Quinn, R. A. Preston, B. Stansby, I. Stairs, B. Stappers, L. Staveley-Smith, S. Tingay, D. Thompson, W. van Straten, K. Wagstaff, M. Warren, R. Wayth, L. Wen and CRAFT Collaboration, The Commensal Real-Time ASKAP Fast-Transients (CRAFT) Survey, *PASA* **27**, 272 (June 2010).
16. B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, T. Akutsu, G. Allen, A. Allocca, M. A. Aloy, P. A. Altin, A. Amato, A. Ananyeva, S. B. Anderson, W. G. Anderson, M. Ando, S. V. Angelova, S. Antier, S. Appert, K. Arai, K. Arai, Y. Arai, S. Araki, A. Araya, M. C. Araya, J. S. Areeda, M. Arène, N. Aritomi, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, Y. Aso, S. M. Aston, P. Astone, F. Aubin, P. Aufmuth, K. Aultoneal, C. Austin, V. Avendano, A. Avila-Alvarez, S. Babak, P. Bacon, F. Badaracco, M. K. M. Bader, S. W. Bae, Y. B. Bae, L. Baiotti, R. Bajpai, P. T. Baker, F. Baldaccini, G. Ballardin, S. W. Ballmer, S. Banagiri, J. C. Barayoga, S. E. Barclay, B. C. Barish, D. Barker, K. Barkett, S. Barnum, F. Barone, B. Barr, L. Barsotti, M. Barsuglia, D. Barta, J. Bartlett, M. A. Barton, I. Bartos, R. Bassiri, A. Basti, M. Bawaj, J. C. Bayley, M. Bazzan, B. Bécsy, M. Bejger, I. Belahcene, A. S. Bell, D. Beniwal, B. K. Berger, G. Bergmann, S. Bernuzzi, J. J. Bero, C. P. L. Berry, D. Bersanetti, A. Bertolini, J. Betzwieser, R. Bhandare, J. Bidler, I. A. Bilenko, S. A. Bilgili, G. Billingsley, J. Birch, R. Birney, O. Birnholtz, S. Biscans, S. Biscoveanu, A. Bisht, M. Bitossi, M. A. Bizouard, J. K. Blackburn, C. D. Blair, D. G. Blair, R. M. Blair, S. Bloemen, N. Bode, M. Boer, Y. Boetzel, G. Bogaert, F. Bondu, E. Bonilla, R. Bonnand, P. Booker, B. A. Boom, C. D. Booth, R. Bork, V. Boschi, S. Bose, K. Bossie, V. Bossilkov, J. Bosveld, Y. Bouffanais, A. Bozzi, C. Bradaschia, P. R. Brady, A. Bramley, M. Branchesi, J. E. Brau, T. Briant, J. H. Briggs, F. Brighenti, A. Brillet, M. Brinkmann, V. Brisson,

P. Brockill, A. F. Brooks, D. A. Brown, D. D. Brown, S. Brunett, A. Buikema, T. Bulik, H. J. Bulten, A. Buonanno, D. Buskulic, C. Buy, R. L. Byer, M. Cabero, L. Cadonati, G. Cagnoli, C. Cahillane, J. C. Bustillo, T. A. Callister, E. Calloni, J. B. Camp, W. A. Campbell, M. Canepa, K. Cannon, K. C. Cannon, H. Cao, J. Cao, E. Capocasa, F. Carbognani, S. Caride, M. F. Carney, G. Carullo, J. Casanueva Diaz, C. Casentini, S. Caudill, M. Cavaglià, F. Cavalier, R. Cavalieri, G. Cella, P. Cerdá-Durán, G. Cerretani, E. Cesarini, O. Chaibi, K. Chakravarti, S. J. Chamberlin, M. Chan, M. L. Chan, S. Chao, P. Charlton, E. A. Chase, E. Chassande-Mottin, D. Chatterjee, M. Chaturvedi, K. Chatziioannou, B. D. Cheeseboro, C. S. Chen, H. Y. Chen, K. H. Chen, X. Chen, Y. Chen, Y. R. Chen, H. P. Cheng, C. K. Cheong, H. Y. Chia, A. Chincarini, A. Chiummo, G. Cho, H. S. Cho, M. Cho, N. Christensen, H. Y. Chu, Q. Chu, Y. K. Chu, S. Chua, K. W. Chung, S. Chung, G. Ciani, A. A. Ciobanu, R. Ciolfi, F. Cipriano, A. Cirone, F. Clara, J. A. Clark, P. Clearwater, F. Cleva, C. Cocchieri, E. Coccia, P. F. Cohadon, D. Cohen, R. Colgan, M. Colleoni, C. G. Collette, C. Collins, L. R. Cominsky, M. Constancio, L. Conti, S. J. Cooper, P. Corban, T. R. Corbitt, I. Cordero-Carrión, K. R. Corley, N. Cornish, A. Corsi, S. Cortese, C. A. Costa, R. Cotesta, M. W. Coughlin, S. B. Coughlin, J. P. Coulon, S. T. Countryman, P. Couvares, P. B. Covas, E. E. Cowan, D. M. Coward, M. J. Cowart, D. C. Coyne, R. Coyne, J. D. E. Creighton, T. D. Creighton, J. Cripe, M. Croquette, S. G. Crowder, T. J. Cullen, A. Cumming, L. Cunningham, E. Cuoco, T. Dal Canton, G. Dálya, S. L. Danilishin, S. D'Antonio, K. Danzmann, A. Dasgupta, C. F. da Silva Costa, L. E. H. Datrier, V. Dattilo, I. Dave, M. Davier, D. Davis, E. J. Daw, D. Debra, M. Deenadayalan, J. Degallaix, M. de Laurentis, S. Deléglise, W. D. Pozzo, L. M. Demarchi, N. Demos, T. Dent, R. de Pietri, J. Derby, R. De Rosa, C. de Rossi, R. Desalvo, O. de Varona, S. Dhurandhar, M. C. Díaz, T. Dietrich, L. di Fiore, M. di Giovanni, T. di Girolamo, A. di Lieto, B. Ding, S. di Pace, I. di Palma, F. di Renzo, A. Dmitriev, Z. Doctor, K. Doi, F. Donovan, K. L. Dooley, S. Doravari, I. Dorrington, T. P. Downes, M. Drago, J. C. Driggers, Z. Du, J. G. Ducoin, P. Dupej, S. E. Dwyer, P. J. Easter, T. B. Edo, M. C. Edwards, A. Effler, S. Eguchi, P. Ehrens, J. Eichholz, S. S. Eikenberry, M. Eisenmann, R. A. Eisenstein, Y. Enomoto, R. C. Essick, H. Estelles, D. Estevez, Z. B. Etienne, T. Etzel, M. Evans, T. M. Evan, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Reviews in Relativity* **23**, p. 3 (September 2020).

17. IceCube Collaboration, Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector, *Science* **342**, p. 1242856 (November 2013).

18. S. Adrián-Martínez, M. Ageron, F. Aharonian, S. Aiello, A. Albert, F. Ameli, E. Anassontzis, M. Andre, G. Androulakis, M. Anghinolfi, G. Anton, M. Ardid, T. Avgitas, G. Barbarino, E. Barbarito, B. Baret, J. Barrios-Martí, B. Belhorma, A. Belias, E. Berbee, A. van den Berg, V. Bertin, S. Beurthey, V. van Beveren, N. Beverini, S. Biagi, A. Biagioni, M. Billault, M. Bondì, R. Bormuth, B. Bouhadef, G. Bourlis, S. Bourret, C. Boutonnet, M. Bouwhuis, C. Bozza, R. Bruijn, J. Brunner, E. Buis, J. Busto, G. Cacopardo, L. Caillat, M. Calamai, D. Calvo, A. Capone, L. Caramete, S. Cecchini, S. Celli, C. Champion, R. Cherkaoui El Moursli, S. Cherubini, T. Chiarusi, M. Circella, L. Classen, R. Cocimano, J. A. B. Coelho, A. Coleiro, S. Colonges, R. Coniglione, M. Cordelli, A. Cosquer, P. Coyle, A. Creusot, G. Cuttone, A. D'Amico, G. De Bonis, G. De Rosa, C. De Sio, F. Di Capua, I. Di Palma, A. F. Díaz García, C. Distefano, C. Donzaud, D. Dornic, Q. Dorosti-Hasankiadeh, E. Drakopoulou, D. Drouhin, L. Drury, M. Durocher, T. Eberl, S. Eichie, D. van Eijk, I. El Bojaddaini, N. El Khayati, D. Elsaesser, A. Enzenhöfer, F. Fassi, P. Favali, P. Fermani, G. Ferrara, C. Filippidis, G. Frascadore, L. A. Fusco, T. Gal, S. Galatà, F. Garufi, P. Gay, M. Gebyehu,

V. Giordano, N. Gizani, R. Gracia, K. Graf, T. Grégoire, G. Grella, R. Habel, S. Hallmann, H. van Haren, S. Harissopulos, T. Heid, A. Heijboer, E. Heine, S. Henry, J. J. Hernández-Rey, M. Hevinga, J. Hofestädt, C. M. F. Hugon, G. Illuminati, C. W. James, P. Jansweijer, M. Jongen, M. de Jong, M. Kadler, O. Kalekin, A. Kappes, U. F. Katz, P. Keller, G. Kieft, D. Kießling, E. N. Koffeman, P. Kooijman, A. Kouchner, V. Kulikovskiy, R. Lahmann, P. Lamare, A. Leisos, E. Leonora, M. L. Clark, A. Liolios, C. D. Llorens Alvarez, D. Lo Presti, H. Löhner, A. Lonardo, M. Lotze, S. Loucatos, E. Maccioni, K. Mannheim, A. Margiotta, A. Marinelli, O. Mariş, C. Markou, J. A. Martínez-Mora, A. Martini, R. Mele, K. W. Melis, T. Michael, P. Migliozzi, E. Migneco, P. Mijakowski, A. Miraglia, C. M. Mollo, M. Mongelli, M. Morganti, A. Moussa, P. Musico, M. Musumeci, S. Navas, C. A. Nicolau, I. Olcina, C. Olivetto, A. Orlando, A. Papaikonomou, R. Papaleo, G. E. Păvălaş, H. Peek, C. Pellegrino, C. Perrina, M. Pfutzner, P. Piattelli, K. Pikounis, G. E. Poma, V. Popa, T. Pradier, F. Pratolongo, G. Pühlhofer, S. Pulvirenti, L. Quinn, C. Racca, F. Raffaelli, N. Randazzo, P. Rapidis, P. Razis, D. Real, L. Resvanis, J. Reubelt, G. Riccobene, C. Rossi, A. Rovelli, M. Saldaña, I. Salvadori, D. F. E. Samtleben, A. Sánchez García, A. Sánchez Losa, M. Sanguineti, A. Santangelo, D. Santonocito, P. Sapienza, F. Schimmel, J. Schmelling, V. Sciacca, M. Sedita, T. Seitz, I. Sgura, F. Simeone, I. Siotis, V. Sipala, B. Spisso, M. Spurio, G. Stavropoulos, J. Steijger, S. M. Stellacci, D. Stransky, M. Taiuti, Y. Tayalati, D. Tézier, S. Theraube, L. Thompson, P. Timmer, C. Tönnis, L. Trasatti, A. Trovato, A. Tsirigotis, S. Tzamarias, E. Tzamariudaki, B. Vallage, V. Van Elewyck, J. Vermeulen, P. Vicini, S. Viola, D. Vivolo, M. Volkert, G. Voulgaris, L. Wiggers, J. Wilms, E. de Wolf, K. Zachariadou, J. D. Zornoza and J. Zúñiga, Letter of intent for KM3NeT 2.0, *Journal of Physics G Nuclear Physics* **43**, p. 084001 (August 2016).

19. N. Senno, K. Murase and P. Mészáros, Choked jets and low-luminosity gamma-ray bursts as hidden neutrino sources, *Phys. Rev. D* **93**, p. 083003 (April 2016).
20. F. Gonzalez and S. Yu, SVOM: a French/Chinese cooperation for a GRB mission, **10699**, p. 1069920 (July 2018).
21. K. Mercier, F. Gonzalez, M. Jouret-Perl, J.-L. Atteia, P. Mandrou, R. Pons, S. Basa, B. Cordier, D. Götz, F. Pinsard, S. Schanne, C. Lachaud, J. Wei and S. Zhang, The French payload on-board the SVOM French-Chinese mission, **9144**, p. 914422 (July 2014).
22. L. Chaoul, V. Mousset and G. Quenouille, Svom french ground segment : Mission, science & instrument centres.
23. Y. Dong, B. Wu, Y. Li, Y. Zhang and S. Zhang, SVOM gamma ray monitor, *Science China Physics, Mechanics, and Astronomy* **53**, 40 (January 2010).
24. D.-H. Zhao, B.-B. Wu, L.-M. Song, Y.-W. Dong, S. Schanne, B. Cordier and J.-T. Liu, Onboard GRB trigger algorithms of SVOM-GRM, *Research in Astronomy and Astrophysics* **13**, 1381 (November 2013).
25. X. Wen, J. Sun, J. He, R. Song, E. Wang, P. Zhou, Y. Dong, X. Liu, J. Liu, L. Li, H. Shi, R. Wang, B. Wu, L. Zhang and S. Zhang, Calibration study of the Gamma-Ray Monitor onboard the SVOM satellite, *Nuclear Instruments and Methods in Physics Research A* **1003**, p. 165301 (July 2021).
26. M. G. Bernardini, F. Xie, P. Sizun, F. Piron, Y. Dong, J. L. Atteia, S. Antier, F. Daigne, O. Godet, B. Cordier and J. Wei, Scientific prospects for spectroscopy of the gamma-ray burst prompt emission with SVOM, *Experimental Astronomy* **44**, 113 (October 2017).
27. A. von Kienlin, C. A. Meegan, W. S. Paciesas, P. N. Bhat, E. Bissaldi, M. S. Briggs, E. Burns, W. H. Cleveland, M. H. Gibby, M. M. Giles, A. Goldstein, R. Hamburg,

C. M. Hui, D. Kocevski, B. Mailyan, C. Malacaria, S. Poolakkil, R. D. Preece, O. J. Roberts, P. Veres and C. A. Wilson-Hodge, The Fourth Fermi-GBM Gamma-Ray Burst Catalog: A Decade of Data, *ApJ* **893**, p. 46 (April 2020).
28. O. Godet, G. Nasser, J. . Atteia, B. Cordier, P. Mandrou, D. Barret, H. Triou, R. Pons, C. Amoros, S. Bordon, O. Gevin, F. Gonzalez, D. Götz, A. Gros, B. Houret, C. Lachaud, K. Lacombe, W. Marty, K. Mercier, D. Rambaud, P. Ramon, G. Rouaix, S. Schanne and V. Waegebaert, The x-/gamma-ray camera ECLAIRs for the gamma-ray burst mission SVOM, **9144**, p. 914424 (July 2014).
29. S. Schanne, N. Dagoneau, F. Château, H. Le Provost, F. Daly, S. Anvar, S. Antier, A. Gros and B. Cordier, The SVOM ECLAIRs gamma-ray burst trigger, *Mem. Soc. Astron. Italiana* **90**, p. 267 (January 2019).
30. D. Götz, J. Osborne, B. Cordier, J. Paul, P. Evans, A. Beardmore, A. Martindale, R. Willingale, P. O'Brien, S. Basa, C. Rossin, O. Godet, N. Webb, J. Greiner, K. Nandra, N. Meidinger, E. Perinati, A. Santangelo, K. Mercier and F. Gonzalez, The microchannel x-ray telescope for the gamma-ray burst mission SVOM, **9144**, p. 914423 (July 2014).
31. K. Mercier, F. Gonzalez, D. Götz, M. Boutelier, N. Boufracha, V. Burwitz, M. C. Charmeau, P. Drumm, C. Feldman, A. Gomes, J. M. Le Duigou, N. Meidinger, A. Meuris, P. O'Brien, J. Osborne, P. Pasqier, L. Perraud, J. F. Pearson, F. Pinsard, E. Raynal and R. Willingale, MXT instrument on-board the French-Chinese SVOM mission, **10699**, p. 1069921 (July 2018).
32. C. Wu, Y. L. Qiu and H. B. Cai, SVOM Visible Telescope: Performance and Data Process Scheme, **279**, 421 (September 2012).
33. X. Fan, G. Zou, Y. Qiu, Z. Pang, H. Zhao, Q. Chen, Y. Pan and H. Yuan, Optical design of the visible telescope for the SVOM mission, *Appl. Opt.* **59**, p. 3049 (April 2020).
34. X. Han, Y. Xiao, P. Zhang, D. Turpin, L. Xin, C. Wu, H. Cai, W. Dong, L. Huang, Z. Kang, N. Leroy, H. Li, Z. Li, X. Lu, Y. Qiu, B. E. Stahl, J. Wang, X. Wang, Y. Xu, Y. Yang, Y. Zhao, R. Zhang, W. Zheng, Y. Zheng and J. Wei, The Automatic Observation Management System of the GWAC Network. I. System Architecture and Workflow, *PASP* **133**, p. 065001 (June 2021).
35. D. Turpin, C. Wu, X.-H. Han, L.-P. Xin, S. Antier, N. Leroy, L. Cao, H.-B. Cai, B. Cordier, J.-S. Deng, W.-L. Dong, Q.-C. Feng, L. Huang, L. Jia, A. Klotz, C. Lachaud, H.-L. Li, E.-W. Liang, S.-F. Liu, X.-M. Lu, X.-M. Meng, Y.-L. Qiu, H.-J. Wang, J. Wang, S. Wang, X.-G. Wang, J.-Y. Wei, B.-B. Wu, Y.-J. Xiao, D.-W. Xu, Y. Xu, Y.-G. Yang, P.-P. Zhang, R.-S. Zhang, S.-N. Zhang, Y.-T. Zheng and S.-C. Zou, The mini-GWAC optical follow-up of gravitational wave alerts - results from the O2 campaign and prospects for the upcoming O3 run, *Research in Astronomy and Astrophysics* **20**, p. 013 (January 2020).
36. D. Corre, S. Basa, A. Klotz, A. M. Watson, M. Ageron, P. Ambert, F. Ángeles, J. L. Atteia, P. E. Blanc, O. Boulade, J. Boy, E. Cadena, B. Cordier, S. Cuevas Cardona, F. Dolon, D. Dornic, D. Dubreuil, A. S. Farah, L. Figueroa, J. Floriot, J. Fuentes-Fernández, P. Gallais, E. Jiménez-Bailón, R. Langarica Lebre, J. F. Le Borgne, W. H. Lee, A. Le Van Suu, M. Marcos, R. Mathon, C. Meessen, J. L. Ochoa, E. Pallier, J. Platzer, F. Quirós, S. Ronayette, D. Turpin, H. Valentin and S. D. Vergani, End-to-end simulations for COLIBRI, ground follow-up telescope for the SVOM mission., **10705**, p. 107051R (July 2018).
37. B. Cordier, F. Desclaux, J. Foliard and S. Schanne, SVOM pointing strategy: how to optimize the redshift measurements?, **1000**, 585 (May 2008).

38. N. Dagoneau, S. Schanne, J. Rodriguez, J. L. Atteia and B. Cordier, Onboard catalogue of known X-ray sources for SVOM/ECLAIRs, *A&A* **645**, p. A18 (January 2021).
39. S. D. Barthelmy, P. Butterworth, T. L. Cline, N. Gehrels, F. Marshall, T. Takeshima, V. Connaughton, R. M. Kippen, C. Kouveliotou and C. R. Robinson, The GRB coordinates network (GCN): A status report, **428**, 99 (May 1998).
40. G. Li, S. Guo, J. Lv, K. Zhao and Z. He, Introduction to global short message communication service of beidou-3 navigation satellite system, *Advances in Space Research* **67**, 1701 (2021).
41. C. Barraud, J. F. Olive, J. P. Lestrade, J. L. Atteia, K. Hurley, G. Ricker, D. Q. Lamb, N. Kawai, M. Boer, J. P. Dezalay, G. Pizzichini, R. Vanderspek, G. Crew, J. Doty, G. Monnelly, J. Villasenor, N. Butler, A. Levine, A. Yoshida, Y. Shirasaki, T. Sakamoto, T. Tamagawa, K. Torii, M. Matsuoka, E. E. Fenimore, M. Galassi, T. Tavenner, T. Q. Donaghy, C. Graziani and J. G. Jernigan, Spectral analysis of 35 GRBs/XRFs observed with HETE-2/FREGATE, *A&A* **400**, 1021 (March 2003).
42. T. Sakamoto, D. Q. Lamb, N. Kawai, A. Yoshida, C. Graziani, E. E. Fenimore, T. Q. Donaghy, M. Matsuoka, M. Suzuki, G. Ricker, J. L. Atteia, Y. Shirasaki, T. Tamagawa, K. Torii, M. Galassi, J. Doty, R. Vanderspek, G. B. Crew, J. Villasenor, N. Butler, G. Prigozhin, J. G. Jernigan, C. Barraud, M. Boer, J. P. Dezalay, J. F. Olive, K. Hurley, A. Levine, G. Monnelly, F. Martel, E. Morgan, S. E. Woosley, T. Cline, J. Braga, R. Manchanda, G. Pizzichini, K. Takagishi and M. Yamauchi, Global Characteristics of X-Ray Flashes and X-Ray-Rich Gamma-Ray Bursts Observed by HETE-2, *ApJ* **629**, 311 (August 2005).
43. N. Dagoneau, S. Schanne, J.-L. Atteia, D. Götz and B. Cordier, Ultra-Long Gamma-Ray Bursts detection with SVOM/ECLAIRs, *Experimental Astronomy* **50**, 91 (July 2020).
44. MAGIC Collaboration, V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González, W. Bednarek, L. Bellizzi, E. Bernardini, A. Berti, J. Besenrieder, W. Bhattacharyya, C. Bigongiari, A. Biland, O. Blanch, G. Bonnoli, Ž. Bošnjak, G. Busetto, A. Carosi, R. Carosi, G. Ceribella, Y. Chai, A. Chilingaryan, S. Cikota, S. M. Colak, U. Colin, E. Colombo, J. L. Contreras, J. Cortina, S. Covino, G. D'Amico, V. D'Elia, P. da Vela, F. Dazzi, A. de Angelis, B. de Lotto, M. Delfino, J. Delgado, D. Depaoli, F. di Pierro, L. di Venere, E. Do Souto Espiñeira, D. Dominis Prester, A. Donini, D. Dorner, M. Doro, D. Elsaesser, V. Fallah Ramazani, A. Fattorini, A. Fernández-Barral, G. Ferrara, D. Fidalgo, L. Foffano, M. V. Fonseca, L. Font, C. Fruck, S. Fukami, S. Gallozzi, R. J. García López, M. Garczarczyk, S. Gasparyan, M. Gaug, N. Giglietto, F. Giordano, N. Godinović, D. Green, D. Guberman, D. Hadasch, A. Hahn, J. Herrera, J. Hoang, D. Hrupec, M. Hütten, T. Inada, S. Inoue, K. Ishio, Y. Iwamura, L. Jouvin, D. Kerszberg, H. Kubo, J. Kushida, A. Lamastra, D. Lelas, F. Leone, E. Lindfors, S. Lombardi, F. Longo, M. López, R. López-Coto, A. López-Oramas, S. Loporchio, B. Machado de Oliveira Fraga, C. Maggio, P. Majumdar, M. Makariev, M. Mallamaci, G. Maneva, M. Manganaro, K. Mannheim, L. Maraschi, M. Mariotti, M. Martínez, S. Masuda, D. Mazin, S. Mićanović, D. Miceli, M. Minev, J. M. Miranda, R. Mirzoyan, E. Molina, A. Moralejo, D. Morcuende, V. Moreno, E. Moretti, P. Munar-Adrover, V. Neustroev, C. Nigro, K. Nilsson, D. Ninci, K. Nishijima, K. Noda, L. Nogués, M. Nöthe, S. Nozaki, S. Paiano, J. Palacio, M. Palatiello, D. Paneque, R. Paoletti, J. M. Paredes, P. Peñil, M. Peresano, M. Persic, P. G. Prada Moroni, E. Prandini, I. Puljak, W. Rhode, M. Ribó, J. Rico, C. Righi, A. Rugliancich, L. Saha, N. Sahakyan, T. Saito, S. Sakurai, K. Satalecka, K. Schmidt, T. Schweizer, J. Sitarek,

I. Šnidarić, D. Sobczynska, A. Somero, A. Stamerra, D. Strom, M. Strzys, Y. Suda, T. Surić, M. Takahashi, F. Tavecchio, P. Temnikov, T. Terzić, M. Teshima, N. Torres-Albà, L. Tosti, S. Tsujimoto, V. Vagelli, J. van Scherpenberg, G. Vanzo, M. Vazquez Acosta, C. F. Vigorito, V. Vitale, I. Vovk, M. Will, D. Zarić and L. Nava, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).

45. H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E. O. Angüner, M. Arakawa, C. Arcaro, C. Armand, H. Ashkar, M. Backes, V. Barbosa Martins, M. Barnard, Y. Becherini, D. Berge, K. Bernlöhr, E. Bissaldi, R. Blackwell, M. Böttcher, C. Boisson, J. Bolmont, S. Bonnefoy, J. Bregeon, M. Breuhaus, F. Brun, P. Brun, M. Bryan, M. Büchele, T. Bulik, T. Bylund, M. Capasso, S. Caroff, A. Carosi, S. Casanova, M. Cerruti, T. Chand, S. Chandra, A. Chen, S. Colafrancesco, M. Curyło, I. D. Davids, C. Deil, J. Devin, P. deWilt, L. Dirson, A. Djannati-Ataï, A. Dmytriiev, A. Donath, V. Doroshenko, J. Dyks, K. Egberts, G. Emery, J. P. Ernenwein, S. Eschbach, K. Feijen, S. Fegan, A. Fiasson, G. Fontaine, S. Funk, M. Füßling, S. Gabici, Y. A. Gallant, F. Gaté, G. Giavitto, L. Giunti, D. Glawion, J. F. Glicenstein, D. Gottschall, M. H. Grondin, J. Hahn, M. Haupt, G. Heinzelmann, G. Henri, G. Hermann, J. A. Hinton, W. Hofmann, C. Hoischen, T. L. Holch, M. Holler, D. Horns, D. Huber, H. Iwasaki, M. Jamrozy, D. Jankowsky, F. Jankowsky, A. Jardin-Blicq, I. Jung-Richardt, M. A. Kastendieck, K. Katarzyński, M. Katsuragawa, U. Katz, D. Khangulyan, B. Khélifi, J. King, S. Klepser, W. Kluźniak, N. Komin, K. Kosack, D. Kostunin, M. Kreter, G. Lamanna, A. Lemière, M. Lemoine-Goumard, J. P. Lenain, E. Leser, C. Levy, T. Lohse, I. Lypova, J. Mackey, J. Majumdar, D. Malyshev, V. Marandon, A. Marcowith, A. Mares, C. Mariaud, G. Martí-Devesa, R. Marx, G. Maurin, P. J. Meintjes, A. M. W. Mitchell, R. Moderski, M. Mohamed, L. Mohrmann, C. Moore, E. Moulin, J. Muller, T. Murach, S. Nakashima, M. de Naurois, H. Ndiyavala, F. Niederwanger, J. Niemiec, L. Oakes, P. O'Brien, H. Odaka, S. Ohm, E. de Ona Wilhelmi, M. Ostrowski, I. Oya, M. Panter, R. D. Parsons, C. Perennes, P. O. Petrucci, B. Peyaud, Q. Piel, S. Pita, V. Poireau, A. Priyana Noel, D. A. Prokhorov, H. Prokoph, G. Pühlhofer, M. Punch, A. Quirrenbach, S. Raab, R. Rauth, A. Reimer, O. Reimer, Q. Remy, M. Renaud, F. Rieger, L. Rinchiuso, C. Romoli, G. Rowell, B. Rudak, E. Ruiz-Velasco, V. Sahakian, S. Sailer, S. Saito, D. A. Sanchez, A. Santangelo, M. Sasaki, R. Schlickeiser, F. Schüssler, A. Schulz, H. M. Schutte, U. Schwanke, S. Schwemmer, M. Seglar-Arroyo, M. Senniappan, A. S. Seyffert, N. Shafi, K. Shiningayamwe, R. Simoni, A. Sinha, H. Sol, A. Specovius, M. Spir-Jacob, Ł. Stawarz, R. Steenkamp, C. Stegmann, C. Steppa, T. Takahashi, T. Tavernier, A. M. Taylor, R. Terrier, D. Tiziani, M. Tluczykont, C. Trichard, M. Tsirou, N. Tsuji, R. Tuffs, Y. Uchiyama, D. J. van der Walt, C. van Eldik, C. van Rensburg, B. van Soelen, G. Vasileiadis, J. Veh, C. Venter, P. Vincent, J. Vink, H. J. Völk, T. Vuillaume, Z. Wadiasingh, S. J. Wagner, R. White, A. Wierzcholska, R. Yang, H. Yoneda, M. Zacharias, R. Zanin, A. A. Zdziarski, A. Zech, A. Ziegler, J. Zorn, N. Żywucka, F. de Palma, M. Axelsson and O. J. Roberts, A very-high-energy component deep in the γ-ray burst afterglow, *Nature* **575**, 464 (November 2019).
46. J. Wang, Y.-L. Qiu and J.-Y. Wei, A pilot study of catching high-z GRBs and exploring circumburst environment in the forthcoming SVOM era, *Research in Astronomy and Astrophysics* **20**, p. 124 (August 2020).
47. M. Della Valle, G. Chincarini, N. Panagia, G. Tagliaferri, D. Malesani, V. Testa, D. Fugazza, S. Campana, S. Covino, V. Mangano, L. A. Antonelli, P. D'Avanzo, K. Hurley, I. F. Mirabel, L. J. Pellizza, S. Piranomonte and L. Stella, An enigmatic

long-lasting γ-ray burst not accompanied by a bright supernova, *Nature* **444**, 1050 (December 2006).

48. J. P. U. Fynbo, D. Watson, C. C. Thöne, J. Sollerman, J. S. Bloom, T. M. Davis, J. Hjorth, P. Jakobsson, U. G. Jørgensen, J. F. Graham, A. S. Fruchter, D. Bersier, L. Kewley, A. Cassan, J. M. Castro Cerón, S. Foley, J. Gorosabel, T. C. Hinse, K. D. Horne, B. L. Jensen, S. Klose, D. Kocevski, J.-B. Marquette, D. Perley, E. Ramirez-Ruiz, M. D. Stritzinger, P. M. Vreeswijk, R. A. M. Wijers, K. G. Woller, D. Xu and M. Zub, No supernovae associated with two long-duration γ-ray bursts, *Nature* **444**, 1047 (December 2006).
49. A. Gal-Yam, D. B. Fox, P. A. Price, E. O. Ofek, M. R. Davis, D. C. Leonard, A. M. Soderberg, B. P. Schmidt, K. M. Lewis, B. A. Peterson, S. R. Kulkarni, E. Berger, S. B. Cenko, R. Sari, K. Sharon, D. Frail, D. S. Moon, P. J. Brown, A. Cucchiara, F. Harrison, T. Piran, S. E. Persson, P. J. McCarthy, B. E. Penprase, R. A. Chevalier and A. I. MacFadyen, A novel explosive process is required for the γ-ray burst GRB 060614, *Nature* **444**, 1053 (December 2006).
50. N. Gehrels, J. P. Norris, S. D. Barthelmy, J. Granot, Y. Kaneko, C. Kouveliotou, C. B. Markwardt, P. Mészáros, E. Nakar, J. A. Nousek, P. T. O'Brien, M. Page, D. M. Palmer, A. M. Parsons, P. W. A. Roming, T. Sakamoto, C. L. Sarazin, P. Schady, M. Stamatikos and S. E. Woosley, A new γ-ray burst classification scheme from GRB060614, *Nature* **444**, 1044 (December 2006).
51. M. J. MichałowskI, D. Xu, J. Stevens, A. Levan, J. Yang, Z. Paragi, A. Kamble, A.-L. Tsai, H. Dannerbauer, A. J. van der Horst, L. Shao, D. Crosby, G. Gentile, E. Stanway, K. Wiersema, J. P. U. Fynbo, N. R. Tanvir, P. Kamphuis, M. Garrett and P. Bartczak, The second-closest gamma-ray burst: sub-luminous GRB 111005A with no supernova in a super-solar metallicity environment, *A&A* **616**, p. A169 (September 2018).
52. B. Arcier, J. L. Atteia, O. Godet, S. Mate, S. Guillot, N. Dagoneau, J. Rodriguez, D. Gotz, S. Schanne and M. G. Bernardini, Detection of short high-energy transients in the local universe with SVOM/ECLAIRs, *Ap&SS* **365**, p. 185 (December 2020).

Quantum field theory with boundary conditions at the horizons*

Gerard 't Hooft

Faculty of Science, Department of Physics,
Institute for Theoretical Physics,
Princetonplein 5, 3584 CC Utrecht, The Netherlands
g.thooft@uu.nl

A procedure to derive a unitary evolution law for a quantised black hole has been proposed by the author. The proposal implies several assumptions, which seem almost unavoidable to this author. We start off with the question how to describe the energy eigenstates of a black hole. The background metric required for this cannot be the Vaidya metric let alone the metric proposed by Hawking, who included the effect of the final evaporation of the black hole. This however leads to the formation if firewalls at both the future and the past event horizon, unless one anticipates the effects that the firewalls have. These effects can be handled as new boundary conditions at the horizons, describing the flow of the participating particles. It is subsequently explained how these boundary conditions must involve the antipodes of the outside world. Imposing unitarity and continuity then automatically leads to a unique, unitary evolution operator. We exhibit the resulting, quite coherent picture.

Keywords: Quantum black hole; past and future event horizon; firewall; Penrose diagram; eternal black hole; time slices; entanglement; vector representation; Shapiro shift; Cauchy surface; boundary condition; antipodal identification.

1. Introduction

There is a general agreement that a theoretical study of black holes in a regime, where quantum mechanical effects play a role, is important for a more complete understanding of general relativity and/or some modifications of this theme, in its relation to quantum mechanics.[1,2] There are several problems and disagreements, however. What is the classical limit from where one should start adding quantum corrections? The first thing many investigators assume is that there is a classical metric describing a black hole from beginning to end, and after that, the dynamical variables should be subject to some quantum mechanical replacement.[3] This should give the "quantum black hole".

However, there is a difficulty concerning the way these quantum variables would evolve. First, in accordance with the no-hair theorem, the evolution of these

*Based on a talk presented at the Sixteenth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories, online, July 2021.

"quantum corrections" would quickly fade away, together with all other quantum field theoretical degrees of freedom. There seems to be just a single state that survives: the Hartle–Hawking state.[4] We do have evolution in this Hartle–Hawking state in a quantum mechanical manner, which is why investigators did continue along such directions, but, entanglement or not, only a single state seems to survive, and this appears to be incompatible with unitarity.

Second, if there is a quantum mechanical evolution law, it should also apply both during the initial implosion and the final puff of the evaporation process. These depend on quantum processes, and demanding their tie evolution to be fixed destroys unitarity.

Third, the quantum states will be entangled with particles entering or leaving the scene far from the black hole.

There is a way out of these difficulties, which is to consider more carefully how ingoing particles assemble near the past event horizon, and how the emitted Hawking particles are prepared near the future event horizon. They form what has become to be known as *firewalls*.[5–7] This is the subject of the present investigation. Already years ago, the author reported about what seems to be considerable progress along these lines,[8] but there continue to be communication barriers that should be overcome.

Particle states emerging from the past event horizon will be shown here to generate a rich spectrum of different possible quantum states of out-going particles, and, considering the equally rich spectrum of in-going particles, the quantum mechanical interactions between the two horizons can be seen to restore unitarity.

There will be limitations to what we can do presently, as will be briefly explained later.

2. The Background Metric

The Penrose diagram[9,10] for the eternal black hole is depicted in Fig. 1(a). This diagram is the unique analytic extension of the Schwarzschild metric that solves Einstein's equations without matter effects added. In line of arguments just sketched above, and clarified further below, this is the only metric that we believe can be used for the quantum theory.

The metric often considered in quantum speculations for black holes, is the Vaidya metric of the kind sketched in Fig. 1(b). It was suspected to be superior since it includes the black hole formation process, so it may seem to contain more freely adjustable parameters that can be subject to quantisation.

Adding to that the gravitational effects of the Hawking evaporation, Hawking proposed to substitute this by Fig. 1(c). This diagram,[11] like the others, has the singularity at $r = 0$ absorbing in-going material, which is problematic since unitarity would demand information to be returned in the entire process. Anyway, one should

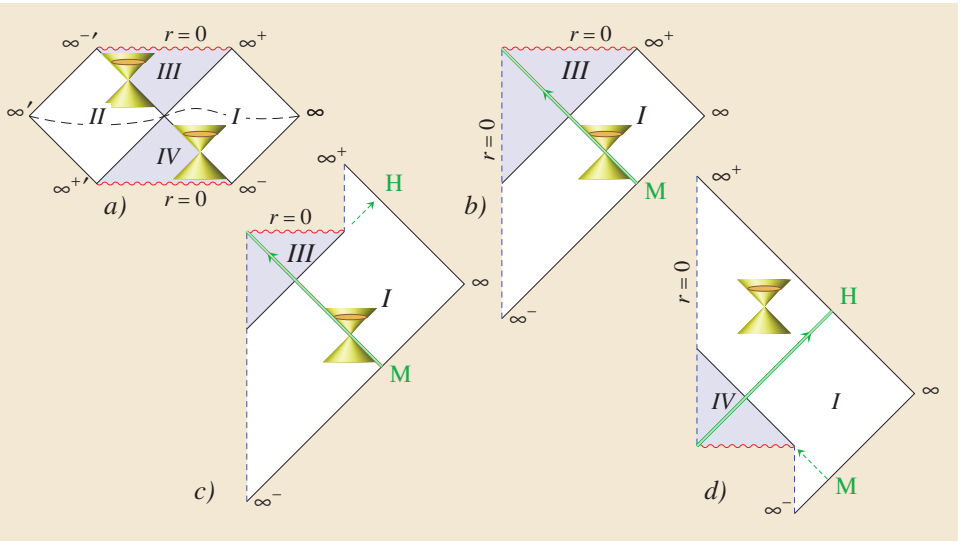


Fig. 1. Different Penrose diagrams for the Schwarzschild black hole. M stands for matter moving in, H for Hawking radiation going out. See text.

realise that the underlying equations, both for general relativity and for quantum mechanics, are invariant under time reversal, so that Fig. 1(d) should at least be considered as possibility as well.

We shall later conclude that only Fig. 1(a) can be used for the quantum process, but the arguments will be more convincing after we have treated the details of our theory. At this point, let us emphasise here that Fig. 1(a) for the eternal black hole is the most suitable one to use for the description of the stationary modes of a black hole, in particular its energy eigenstates, which make use of beams of in-going and out-going particles asymptotically far away, so as to make up for the evaporation.

To set up our program, we momentarily return to Schwarzschild coordinates, we divide the Schwarzschild time coordinate t in equal slices, each lasting for at most one period $\Delta t \approx M_{\rm BH}$ in Planck units, see Fig. 2. Within each of these time slices, the physical variables are defined to consist of all standard model particles and the slightly hidden degrees of freedom such as dark matter and gravitons. These particles are following their (quantum) trajectories, but, and this is essential for our procedure, particles that venture too close to either the future or the past horizon, will *not* be included. The idea is that such particles will become more manifestly observable in future or past slices (where they belong).

This paper is about stitching these slices back together in order to get the complete picture. Particularly at the horizons, this is a nontrivial procedure.

There may be some misunderstanding as to the quantum state of the Hawking particles. The Hawking particles are seen to be entangled, and as such it may appear

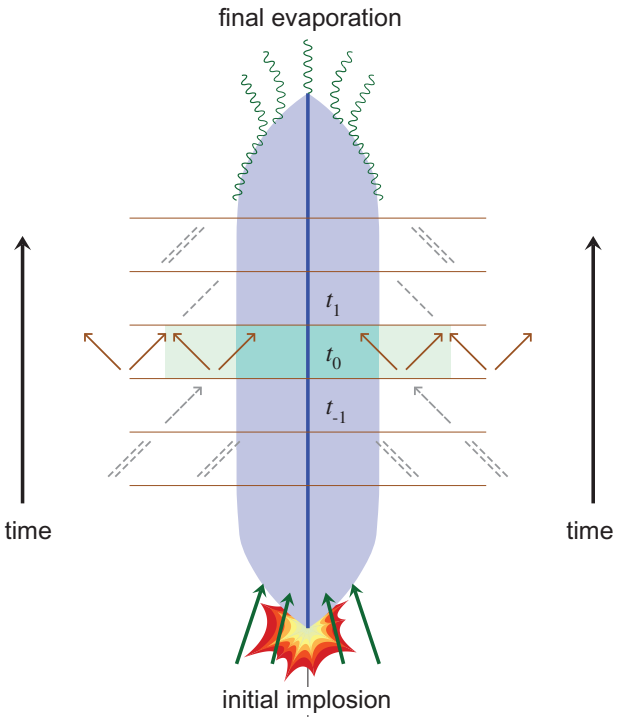


Fig. 2. Slicing the Schwarzschild coordinates in sequences of short periods. The quantum variables can be defined on each slice, lasting not longer than $\mathcal{O}(M_{\rm BH})$ in Planck units.

that they can only form a single quantum state. This happens only, however, if none of the Hawking particles are subject to observation, and if nothing is thrown into the black hole for large stretches in time. This will become more clear below.[a]

Quantum field theory (QFT), projected on this background of time-slices, will reveal how particles propagate and interact. At any safe distance (of the order of a couple of Planck lengths), the outcomes of QFT calculations will be decisive as to what happens. Up to this point, it should be clear that the metric of an eternal black hole serves just fine as a background.

However, particles inevitably will approach the horizons more closely. There are particles entering the future event horizon, and particles emerging from the past event horizon. A theory is needed that *dictates* how particles appear out of the past horizon, and how, if at all, this is determined by the in-going ones.

At this point, it seems that Hawking's original calculation[1] cannot be the entire story since it puts all out-particles into just one state, the Hartle–Hawking state.

[a]For a better understanding of these aspects of quantum mechanics, it may be helpful to view quantum mechanics itself as a *vector representation* of physically observable objects, see the author's ideas about this, in Refs. 12 and 13.

In contrast, we expect that different states of particles entering the future horizon should lead to different states of out-particles.

It suffices to produce a prescription that holds just for a few Planck lengths away from the intersection point of future and past event horizons. There, we expect "something" that causes in-particles to bounce against, or turn into, out-particles. So-far, in this treatment, we have only used the physically accessible or observable part of the Schwarzschild metric. Quantum gravity has been used only in its perturbative regime, so it only enters in the form of some stray gravitons. But now, this is going to change.

What has been found is that *interactions* between in- and out-particles can do the job. We need interactions that become sufficiently strong only at tiny distances. Which force acts more strongly at close distances than the renormalizable interactions in the Standard Model? Of course, the answer is: the gravitational force.

3. The In–Out Interaction Caused by Gravity

Indeed, the gravitational effect of an in-going particle on an out-going one is that the wave function of the out-particle is modified. This is illustrated in Fig. 3: the dashed arrowline marked "in", going through the future event horizon very early in time, carries a large momentum δp^-. Coming from region (IV) in the diagram, a Hawking particle was on its way out (wiggled line marked "out"). Due to δp^-, a shock wave[14] from the in-particle drags the out-particle along, displacing it by a distance δu^-, which is proportional to δp^- and a Green function $F(\theta)$ depending on the transverse, angular, distance θ between the particles. This function $F(\theta)$ has been calculated.[15]

This effect is closely related to the Shapiro effect,[16] the fact that a light or radio signal from an astrophysical source or a spacecraft, grazing the Sun, experiences a delay due the solar gravitational field. All Hawking particles are delayed in a different way, depending on their transverse positions, and so it happens that the entire quantum state describing the out-particles is modified.

The modification operator is easy to calculate, in principle

$$U = e^{i\delta p^-\ \sum_{\rm out} p^+_{\rm out} GF(\theta_{\rm out})} = e^{i\ \sum_{\rm in,out} p^+_{\rm out} p^-_{\rm in} GF(\theta_{\rm out},\varphi_{\rm out},\theta_{\rm in},\varphi_{\rm in})}, \tag{3.1}$$

where G is Newton's constant, and furthermore, in the second part we replaced δp^- by $\sum_{\rm in} p^-_{\rm in}$ since *all* in-particles have this effect on *all* out-particles.

It must be emphasised that the dependence of U on the in-particle momenta increases exponentially with Schwarzschild time, because, the later the time t where we study the particles in the background metric, the closer the older in-particles squeeze against the past horizon, so the larger their influence will be. The very first in-particles in fact were the particles that formed the black hole by imploding. Since they came in very early, their exponentiated in-momenta p^- are gigantic, in terms

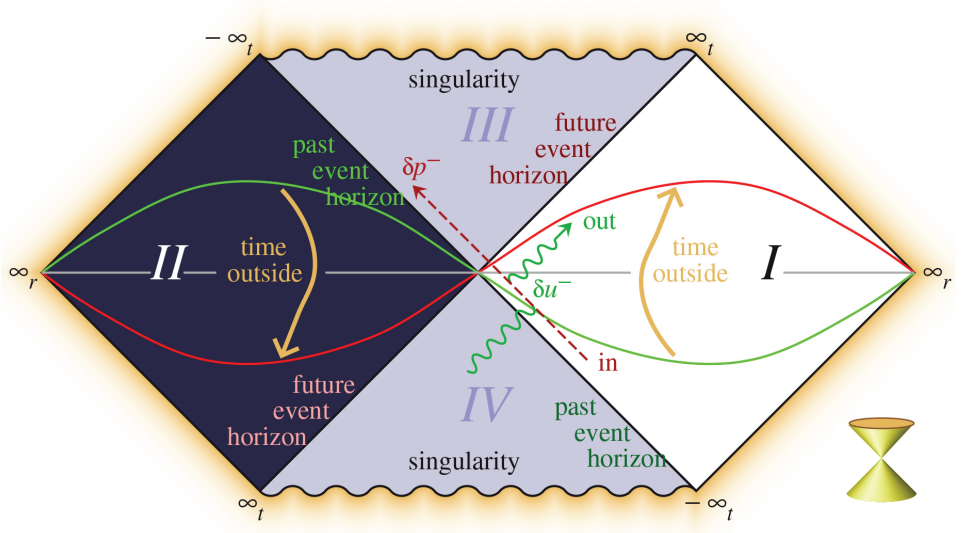


Fig. 3. The Penrose diagram for the eternal black hole, see text. Light cones everywhere in the diagram are oriented the same way as the one shown.

of the Penrose coordinates for the particles later in time, *but this is nothing to worry about*: they caused the wave functions of the early out particles to be separated by equally large separation variables $u^-_{\rm out}$.

The increase of u^- changes from exponential to linear in time as soon as the out-particles move out of the direct gravitational influence of the black hole. They are joining the other distant out-particles that no longer play any role for us. Our complete quantum states of course include the early out-going particles, and our calculations show that they are entangled with the later particles, but from this moment onwards we can just re-arrange the basis of quantum states for the out-particles, and describe their wave functions any way we please.

The reader may have felt something strange in the emerging picture: the time dependence of all particle states is such that the outgoing particles move away exponentially from the past and future event horizons, but there are *two* regions in which they can escape. The Penrose diagram used so-far is twice the size of the universe that we usually consider as the home universe of this black hole. We labelled these two regions (I) and (II) in Fig. 3. What happens to a particle that makes its way out in region (II)? If we ignore it, we make a mistake. The operator U derived in Eq. (3.1) is the operator for the Fourier transform (apart from the Green function F that will be treated later). *The Fourier transformation process is unitary*, which means that we can use it to describe a unitary quantum evolution of the black hole, *but it is only unitary if we consider in- and out-wave functions to cover the entire line* $[-\infty, +\infty]$, and not just the half-lines that form the boundaries of region (I) or region (II) of a black hole.

Indeed, the Shapiro effect can easily transport the geodesic of a particle across the other horizon. This is the real reason why we are destined to use the entire, double, Penrose diagram for our considerations. Before disclosing the *physical* role of region (II), we sketch the picture obtained so-far.

A time boost for an outside observer, $t \Rightarrow t+\Delta t$, corresponds to a *Lorentz boost*, $(u^+,\, u^-) \Rightarrow (u^+ e^{-t/4GM_{\rm BH}},\ u^- e^{t/4GM_{\rm BH}})$ for an observer close to the origin of this diagram. This implies that Cauchy surfaces for the particles in the regions (I) and (II) take the shape of the green curve in Fig. 3, that evolves into and beyond the red curve as time continues.

The Cauchy surfaces never enter into the regions (III) or (IV) in the Penrose diagram. They pivot around the origin. This is where our procedure contrasts with other work.[3]

However, we do have to pay for this — in an interesting way. Our general finding is that, when particles go into and out of the black hole, they cause shifts along the horizons. During black hole formation, these shifts are large. Effectively, we are using the eternal Penrose diagram, but the *boundary conditions*, by which the horizons connect different domains of the spacetime metric, are strongly affected by the gravitational force: the Shapiro effect.

Now, we have to address the question what region (II) stands for physically. It cannot just be ignored. It also cannot just be there to describe another black hole. That other black hole would not only be strongly entangled with the first one, as some investigators argue, but the two black holes exchange information. Entangled particles do *not* exchange information without violating locality.

Mathematically, it is easy to "cure" our theory. We have to find an isometry transformation in spacetime, whose square is the identity. We can write it as $\mathbb{Z}_2$. One such transformation is the identity itself, or, regions (I) and (II) are then assumed to represent the same spacetime. Our problem is then that this does not work since the Shapiro shift would not commute with such a mapping, or, it would generate cusp singularities near the origin. However, such cusp singularities do not arise if we accompany the $\mathbb{Z}_2$ mapping with an antipodal mapping:

$$(\theta,\, \varphi) \leftrightarrow (\pi - \theta,\ \varphi + \pi). \tag{3.2}$$

In flat Minkowski spacetime, this would also generate a cusp singularity at the origin, but for Schwarzschild (and its rotating and charged relatives), it does not. This is because in the entire spacetime that we are using, $r \geq 2\mathrm{GM_{BH}}$ so that angular positions never come close to their antipodes. Only deep inside the regions (III) and (IV), $r \to 0$, and points coincide with their antipodes. We never venture more than infinitesimal distances into these unphysical regions.

Nevertheless, at the beginning of this paragraph, we put the word "cure" in quotation marks. The fact that a solution looks mathematically pleasing does not automatically mean that it embodies the physical solution.

This antipodal mapping procedure may help us to avoid the "quantum no-cloning" condition, but does it? We do see a potential danger. In region (II),

compared with the time variable for a local observer, the arrow of time for the distant observer will flip over. *Time runs backwards there!* This means that, if, near the antipodes, an observer sees a particle, the same particle will appear to move backwards in time, or, have negative energy, when viewed by an observer in region (II).

Our resolution to that problem is to redefine energy by a constant. A particle in region (II) has energy $E_{\rm II} = E^{\max} - E_{\rm I}$, where $E^{\max}$ is the *energy of the "anti-vacuum" state.* The antivacuum is the state completely filled with particles, having the largest possible amount of energy.

Better understanding is obtained by adhering to expansions in partial harmonics, item 9 in the next section, but this we leave for future considerations.

4. On the Complete Theory

We hope the above can serve as a more elaborate explanation as to the various ingredients for a consistent theory of quantum black holes. It can be viewed as a "package deal": there are several postulates that by now sound natural and unavoidable by this author, but seem to meet with considerable skepticism from other researchers:

(1) Up to a few Planck lengths from the horizon, matter particles may be handled using conventional quantum field theory, only during small time steps, each lasting for the "scrambling time", $T_{\rm scrambl} = \mathcal{O}(M_{\rm BH})$. This is far too short to include collapsing matter or late evaporation. These would act as projection operators that ruin unitarity. In practice, we can only handle modest amounts of matter going in and/or out during any of these short time spans.
(2) One has to use the metric of the eternal black hole as background.
(3) The metric must be divided by $\mathbb{Z}_2$, to restore a single asymptotic region.
(4) The only way to do that is by *antipodal identification.*
(5) Time outside runs *backwards* in region (II). This implies that the local states to be compared with the black hole states must approach the *antivacuum* in region (II). Creating a particle near the black hole corresponds to creating a particle in (I) or annihilating a particle in (II).
(6) During the time stretch around time $t \approx t_0$ for the distant observer, in-particles at $t \ll t_0$ are seen to generate a Shapiro shift along the past event horizon; out-particles at $t \gg t_0$ generate Shapiro shift along future event horizon.
These form *firewalls* that can be removed by re-arranging the corrections at the horizons (cut-and-paste procedures that effectively modify the boundary conditions there).
(7) This links the *positions* of the out particles to the momenta on the in-particles, and *vice versa.* That opens the way to define a unitary evolution law.
(8) Initial imploding matter and the final evaporating matter have decisive effects at the horizons; rather than merely shifting the geodesics, these may *initiate*

the links between the antipodes there. Initial and final black holes are mere *seeds*, to be described as gravitational *instantons*. This is as yet a suspicion; it is not known how to substantiate that with calculations.

(9) To do explicit calculations, it is advisable to expand the in-going momenta and the out-going momenta or positions at the horizons in *spherical harmonics* for optimal use of linearity of the equations. The Green function $F(\theta, \varphi, \theta', \varphi')$ then diagonalises, much like Schrödinger's equation for the hydrogen atom.
Note that these spherical harmonics are not to be confused; in our case, we are dealing with *operators*. Since these operators involve all of the in-going and out-going particles, it is not appropriate to try to "second quantise" them. At every (ℓ, m), we have only one position operator and one momentum operator.

The mathematics works correctly only when these views are all combined. In particular, we may *only* compare states in a background metric to states as seen from the outside, if the mapping on the Cauchy surfaces is 1 to 1.

This enforced the antipodal mapping (it cannot be avoided).

Note that this mapping keeps all information visible on the Cauchy surfaces; we have no information loss.

The procedure described above can lead to an accurately defined scattering matrix linking every in-state to an out state in a unitary manner. Unitarity may follow because, at every (ℓ, m) mode, where ℓ must be odd because of the antipodal identification, the Fourier transformation is a unitary operation.

It is important to emphasise that the Shapiro effect only acts on energy and momentum, not on other quantum numbers. This means that further analysis is needed to establish how data other than positions and momenta are transported across horizons.

Thus, we are not yet ready to consider the complete Standard Model to describe the interactions in the immediate vicinity of the black hole. This is somewhat reminiscent to the situation in string theory. This theory also requires first to characterise all its excited states in terms of momentum and/or position operators, while its relation with the Standard Model can only be established by extensive studies of the relevant symmetry groups and algebras.

A new energetic generation of young students and ambitious postdocs is asked for to continue research along these lines.

References

1. S. W. Hawking, *Commun. Math. Phys.* **43** (1975) 199.
2. G. 't Hooft, *Nucl. Phys. B* **256** (1985) 727.
3. S. B. Giddings, *Phys. Lett. B* **754** (2016) 39, https://doi.org/10.1016/j.physletb.2015.12.076.
4. J. B. Hartle and S. W. Hawking, *Phys. Rev. D* **13** (1976) 2188.
5. A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *J. High Energy Phys.* **1302** (2013) 062, arXiv:1207.3123 [hep-th].
6. D. Marolf, *Rept. Prog. Phys.* **80** (2017) 092001, arXiv:1703.02143 [gr-qc].

7. J. Polchinski, arXiv:1609.04036 [hep-th].
8. G. 't Hooft, *Foundations Phys.* **47** (2017) 1503, arXiv:1612.08640 [gr-qc].
9. R. Penrose, *Phys. Rev. Lett.* **10** (1963) 66.
10. S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge University Press, Cambridge, 1973).
11. J. B. Hartle, *Generalized Quantum Theory and Black Hole Evaporation*, in *Black Holes and High Energy Astrophysics*, eds. H. Sato and N. Sugiyama (Universal Academy Press, Tokyo, 1998), arXiv:grqc/9808070.
12. G. 't Hooft, *The Cellular Automaton Interpretation of Quantum Mechanics*, Fundamental Theories of Physics, Vol. 185 (Springer International Publishing, 2016), http://arxiv.org/abs/1405.1548.
13. G. 't Hooft, arXiv:2103.04335[quant-ph].
14. P. C. Aichelburg and R. U. Sexl, *Gen. Relativ. Gravit.* **2** (1971) 303.
15. T. Dray and G. 't Hooft, *Nucl. Phys. B* **253** (1985) 173.
16. I. I. Shapiro, *Phys. Rev. Lett.* **13** (1964) 789.

The development of general relativity and the cosmological constant

Asghar Qadir

Abdus Salam School of Mathematical Sciences
Government College University Lahore, Pakistan
E-mail: asgharqadir46@gmail.com

There is much talk of a "mysterious form of energy", called "dark energy" that forms the bulk of the energy content of the Universe. Perhaps the most mysterious aspect of it is why it should be regarded as mysterious in the first place. This question is discussed in the context of the development of relativity and relativistic cosmology. It will be argued that there is no good reason to treat it as other than Einstein's cosmological constant.

Keywords: Cosmological constant; dark energy; geodesic equation; geodesic deviation; Einstein equations; Planck units; vacuum energy.

1. Introduction

It is often taken that there is no need to hark back to a subject and one only needs to consider current developments. However, people like Julian Barbour[1] point out that the past debates and the process of the development of a subject might provide the correct direction in which to proceed. He was talking of the problem of combining quantum theory and general relativity, which has not been achieved despite the best efforts of the best scientists for about a century. This problem is at the base of, or contributes significantly to, many other important problems that Physics is facing today. I would agree with Barbour that when we are so badly stuck, we should look back to see where we took a wrong turn. Very often we do not realize that some hidden assumptions have been made and proceed forward in various erroneous directions. We should look back lest, in our headlong rush forward, we miss the lessons lying in the past. This is especially true if it is argued that we cannot require experimental verification because it is beyond our current technology, and must instead rely on the concept of "naturalness" — reminiscent of Aristotle's "self-evident truths". This argument is appealed to for the attempted merger of the two major theories, which have so far resisted a union — the so-called "quantum-gravity problem". The problem is exacerbated by the current trend to publish as much as possible, which encourages wild speculation, instead of scientific progress based on adequate experimental evidence. One might fruitfully apply the same caution, to look back, for other recent developments.

I am, here, concerned with the so-called "dark-energy problem". In an attempt to see significant effects of his general theory of relativity, Albert Einstein applied it to the Universe in the large.[2] The equations in his seminal paper on general

relativity[3] implied a non-static Universe, and he firmly believed that it should be static (see[4] for a more detailed explanation of why he believed it). In this paper he added a constant to his Lagrangian and obtained a static Universe model. However, in 1929 Edwin Hubble found that further away galaxies are redder. Treating the reddening as a Doppler shift, he deduced the Hubble Law,[5] that galaxies are generally receding from us at a rate proportional to their distance. (Einstein called his modification "the greatest blunder of my life".) This fitted with Friedmann's solutions for Einstein's field equations,[6] which provided three universe models: two expanding forever; and one contracting later; but all expanding at more or less the same decreasing rate currently. Thus, if the scaling factor for the distance between galaxies is given by $a(t)$, the Hubble parameter that relates the speed of expansion with the distance is $H = \dot{a}(t)/a(t)$. As the scale is increasing this is positive. However, the dimensionless acceleration $q = a(t)\ddot{a}(t)/\dot{a}^2(t)$, would be negative. For the model which contracts later, H would become zero later, and then negative.

Observations strongly supported the Friedmann models, but the fact that the Universe consists mainly of hydrogen and helium and all other elements are essentially "cooked" in stars, seemed to be a stumbling block in the theory till George Gamow[7] suggested that there must have been radiation in the early Universe that stopped nucleosynthesis but allowed some deuterium, tritium, helium$_3$ and normal helium to form. Later observations further supported the Friedmann open model.[8] Gamow's "hot big bang model predicted that there would then be a remnant of that radiation still left over as the Universe cooled. A rough estimate was of 3°K. It was confirmed in 1965 when Penzias and Wilson discovered a microwave background radiation,[9] which turned out be in a Planck spectrum, as predicted. Fitting the observed primordial abundance of elements with the radiation gave very precise estimates of the amount of baryonic matter in the early Universe.

In 1997 new observations of the further reaches of the Universe, obtained by studying type Ia supernovae, indicated that the Universe is expanding at an exponential rate.[10,11] As it happens, much earlier in 1927, Georges Lemaître[12] had independently derived the Friedmann solutions, but with an extra feature allowing for an exponential expansion in later times. He had derived the solution using Einstein's equations with a cosmological constant. This was what Perlmutter and Reiss[10,11] had referred to. Nevertheless, a flurry of papers appeared in the journals about evidence for a "mysterious form of dark energy", despite the fact that Riess' paper explicitly mentions the cosmological constant. In this paper I will try to demonstrate that it is because of not looking back to the past, that theory is still blundering down *dark* blind alleys, instead of accepting the obvious solution. It is that, and *the dark side of the force*, forcing many people to publish many papers. The current standard model of cosmology[13] is called the ΛCDM model, the CDM stands for "Cold Dark Matter" — and the "Λ" is Einstein's "greatest blunder"!

2. Generalizing Special Relativity

Soon after Einstein published his paper on special relativity,[14] his one-time teacher Hermann Minkowski (whose classes he had skipped), wrote the theory in terms of a four-dimensional spacetime (a space with an indefinite metric).[15] Einstein was not pleased with the development at the time, regarding the physics of the theory as being obfuscated by the mathematical formalism. However, the theory was restricted to uniform linear motion, and he needed to generalize it. On considering small accelerations along the direction of motion he was able to derive the physical consequence of a deflection of light due to gravity, which amounts to a bending of a straight line by acceleration. He considered uniform circular motion and found that it entailed a change in the geometry implying that one needed non-Euclidean geometry (see for example[16] for an explanation of the reasoning).

Isaac Newton[17] had defined the world-view that held sway before Einstein, according to which there is an absolute space and an absolute time, whereby space exists in itself independent of any matter or any observer in it. God was explicitly given the role of the Universal observer to provide reality to space. Time was taken to "flow at a constant rate". What sense would that phrase make? A stream can flow at a constant rate, meaning that in equal intervals of time the particles in the stream would flow equal distance. How could points of *time* flow at a constant rate in equal units of *time*? When Einstein had challenged this so-called "common sense" view with his special theory, non-physicists found it impossible to come to terms with the new view. However, while he challenged the absolute time more directly, and allowed lengths to decrease for moving observers, he had not altered the basic picture of a rigid grid. With a change of geometry that picture would become untenable. Before the Newtonian revolution, the world-view of Aristotle (c. 350 BC) held sway. In that view, space only existed as the distance between bodies. To give it reality, Aristotle had postulated an all-pervading aether. Further, time only gave a chronology of events, and precise measures of time were not assumed. If Newton's absolute grid had to go, then we should expect space to be defined by its matter content. Measures of space and time would depend on the procedures specified to measure them and would not be automatically defined.

Einstein had another problem, this time with Newton's law of gravitation. According to it, if a body moves away from its position, another body will *instantaneously* feel the change in the gravitational force between the two bodies. However, this militates against the basic precept of his special theory, as *all* information must obey the speed of light limit. Somehow, Aristotle's aether would be needed. However, he had managed to get rid of the luminiferous aether. As such, the spacetime would have to carry the information across at a speed limited by that of light. Instead of being Newton's rigid grid, or a rotated version of it (causing distances to appear shrunk), Einstein needed a more malleable spacetime that could be distorted by its matter content. Over time it became clear to him that he needed to use the geometry taught by Minkowski that he had skipped.

3. The Einstein-Grossmann Collaboration

Einstein had already arrived at the key to his generalization by, what he had called "the happiest thought of my life". This was the realization that gravity would not be felt in a freely falling frame, since all objects would fall together, as Galileo had shown. Hence, in that frame a ball released would not fall but float. This led him to a prediction of the bending of light by gravity,[18] as an experimenter in a fixed frame would see light sent across a falling lift hit the other side lower than if the lift is falling.[16] The effect turns out to be half what he later derived from the full general theory. He proposed that the effect could be seen by looking at the bending of the light from a star near the edge of the solar disc, seen during a full solar eclipse. (Fortunately for him, World War I broke out and prevented the expedition to observe the solar eclipse in South America. By the time it finally became possible, he had the correct answer.) In that paper he also obtained the gravitational red-shift, which is unaltered in the full theory. However, the effect is so small that it was not till 1956 (after Einstein's death) that the experiment could actually be performed and not till 1959 that the effect was finally verified.[19]

His friend, Marcel Grossmann (who is commemorated in these meetings) had not only attended Minkowski's lectures, but was proficient in non-Euclidean geometry, so Einstein asked him to help. Though Einstein had tried other non-Euclidean extensions of Geometry, Grossmann taught him Riemann's extension using differential calculus — what is nowadays called differential, or Riemannian, geometry. This turned out to be especially fruitful. Though "the happiest thought" had come earlier, in many ways "the happiest *development*" came in the first paper of their collaboration,[20] where Newtonian's scalar gravitational potential is replaced by the geometrical metric tensor. Thus *the spacetime itself becomes the medium to carry the gravitational field*, participating in dynamics as was needed. This set the base for the theory. In it, the gravitational "force" is nothing but the curvature of the four-dimensional spacetime. Paths are bent due to the warping of the spacetime, rather than a mysterious force acting instantaneously from a distance. In the second paper of the collaboration,[21] the basic field equations were set up and shown to be covariant. This was well before the derivation of the equations from a variational principle. In it, the Einstein tensor was defined by using the divergence-free condition for the conservation of the stress-energy tensor and the second Bianchi identities. However, they limited themselves to those transformations that left the determinant of the metric tensor invariant. In other words, instead of the symmetry group being $GL(4)$, $SL(4)$ was used. Since that reduces one of the degrees of freedom, the field equations were not correctly identified there.

There remains a crucial step in obtaining the field equations that was not made explicit in their derivation. How do we suddenly jump to relating the curvature tensor to the stress-energy tensor, without it coming from the bending of the path due to the presence of matter-energy as the warping of the spacetime? I will derive them the way I think Einstein must have done, but unfortunately left hidden.

4. Derivation of the Einstein Field Equations by Using Geodesic Deviation

To follow the argument, one needs to think of one of a family of observers looking at another observer of the same family, i.e. both in the same frame. I hope the reader will bear with me while I go through basic differential geometry, so that the reasoning can be seen in full clarity. The observers will follow *geodesics*, the straightest available paths. If the tangent vector to the path is $\mathbf{t}$, treating it as a derivative operator, $\mathbf{t}(\mathbf{t}) = 0$, so that the direction does not change, i.e. $d\mathbf{t}/ds = 0$. This is the *geodesic equation*. In terms of the components of $\mathbf{t}$, $t^a = dx^a/ds$, where x^a is the position vector of the observer, the geodesic equation is $t^b t^a_{;b} = 0$, where the summation convention that repeated indices are summed over, has been used and ";" is the covariant derivative, $t^a_{;b} = t^a_{,b} + \Gamma^a_{bc} t^c$, "," being the partial derivative $\partial/\partial x^a$. Thus, in this index notation, the geodesic equation becomes

$$\ddot{x}^a + \Gamma^a_{bc} \dot{x}^b \, \dot{x}^c = 0 \ , \tag{1}$$

where Γ^a_{bc} is the Christoffel symbol, given in terms of the metric tensor, g_{bc}, by

$$\Gamma^a_{bc} = \Gamma^a_{cb} = \frac{1}{2} g^{ad} (g_{bd,c} + g_{cd,b} - g_{bc,d}) \ , \tag{2}$$

and g^{ad} is the inverse of g_{bd}, so that $g_{bd} g^{ad} = \delta^a_b$.

Let $\mathbf{p}$ represent the position vector of observer B, as seen by observer A. If the geodesics are radial lines on a plane and A is an ant moving outward at a constant rate on one line at angle θ_A, looking at an ant B on another line at θ_B, $\mathbf{p}$ will increase at a constant rate as the two move out, so $\ddot{\mathbf{p}} = 0$. Instead, let A be a jet plane flying due South on longitude ϕ_A and B another on longitude ϕ_B. For the sake of definiteness, let them both start at a latitude θ_1 in the northern hemisphere. Even if the two jets fly at the same speed $\mathbf{p}$ will increase, $\dot{\mathbf{p}} > 0$, as θ decreases, but the rate of increase of $\dot{\mathbf{p}}$ will *decrease* at a constant rate, i.e. $\ddot{\mathbf{p}} < 0$. As the rate of change of position is the velocity, $\dot{\mathbf{p}}$ is the velocity and *its* rate of change, $\ddot{\mathbf{p}}$, is the acceleration, $\mathcal{A}$. Thus curvature is seen to be related to acceleration in kinematic terms. To see precisely how the two are related, one needs to require that the position vector $\mathbf{p}$ be "carried along" by the tangent vector to the geodesic, $\mathbf{t}$. We say that $\mathbf{p}$ is *Lie transported* along $\mathbf{t}$. In the above example, if A starts at point P and goes on to point Q, then the position vector of B at Q will be given by $\mathbf{t}(\mathbf{p})$. Further, if B starts at R and ends at S, then S is given by $\mathbf{p}(\mathbf{t})$. But the position of S relative to Q is also given by $\mathbf{t}(\mathbf{p})$. Hence they commute, i.e. $\mathbf{t}(\mathbf{p}) = \mathbf{p}(\mathbf{t})$.

The curvature of a space over a small area is measured by carrying a vector field, v^a along one direction in the locally flat coordinate system and then the other, and subtracting the result from carrying it in the reverse order. (Carrying along the map, rather than on the curved space itself, is called *parallel transport*.) This is seen most easily on the surface of the Earth with one direction along a line of longitude and the other a line of latitude. Carrying equal *distances* and not angles (which would be the case for Lie transport), as defined in the local coordinate system of a map,

the rectangle carrying the vector will not close, because the cosine of the latitude will enter into the distance. Also the final direction of the vector will be different for the two, as it gets "twisted around more at a higher latitude than a lower one. Thus the Riemann curvature tensor is defined in index notation by

$$2v^{a}_{\ ;[c;d]} = R^{a}_{\ bcd}v^{b} \ , \tag{3}$$

where $S_{[ab]} = (S_{ab} - S_{ba})/2!$.

For our purposes we need to use a four-dimensional spacetime. Since the geodesic of A is the path of a test particle, $\mathbf{t}$ must be a timelike vector, and since $\mathbf{p}$ is the position vector of a neighbouring test particle seen at some time in the rest-frame of A, it must be a spacelike vector. I use the metric signature convention $(+,-,-,-)$, so that $\mathbf{t}.\mathbf{t} > 0 > \mathbf{p}.\mathbf{p}$. Thus the physical acceleration of B relative to A is

$$\mathcal{A} = \ddot{\mathbf{p}} = \mathbf{t}[\mathbf{t}(\mathbf{p})] \ . \tag{4}$$

In the geometrical context it is called the *geodesic deviation*. We now use the Lie transport and geodesic equations to get the relation between the curvature tensor and the acceleration.[4] First, we use the Lie transport for the acceleration, written in index notation, by commuting the expression inside the bracket to get

$$\mathcal{A}^{\mu} = t^{\rho}(t^{\nu}p^{\mu}_{\ ;\nu})_{;\rho} = t^{\rho}(p^{\nu}t^{\mu}_{\ ;\nu})_{;\rho} \ , \tag{5}$$

and then expand the bracket to obtain

$$\mathcal{A}^{\mu} = t^{\rho}p^{\nu}_{\ ;\rho}t^{\mu}_{\ ;\nu} + t^{\rho}p^{\nu}t^{\mu}_{\ ;\nu;\rho} \ . \tag{6}$$

Next, we use the Lie transport equation again for the first term to write it as

$$\mathcal{A}^{\mu} = p^{\rho}t^{\nu}_{\ ;\rho}t^{\mu}_{\ ;\nu} + t^{\rho}p^{\nu}t^{\mu}_{\ ;\nu;\rho} \ . \tag{7}$$

Notice that

$$p^{\rho}(t^{\nu}t^{\mu}_{\ ;\nu})_{;\rho} = p^{\rho}t^{\nu}_{\ ;\rho}t^{\mu}_{\ ;\nu} + p^{\rho}t^{\nu}t^{\mu}_{\ ;\nu;\rho} \ , \tag{8}$$

so we and add and subtract $p^{\rho}t^{\nu}t^{\mu}_{\ ;\nu;\rho}$, using the positive part to rewrite the first term and obtain

$$\mathcal{A}^{\mu} = p^{\rho}(t^{\nu}t^{\mu}_{\ ;\nu})_{;\rho} - p^{\rho}t^{\nu}t^{\mu}_{\ ;\nu;\rho} + t^{\rho}p^{\nu}t^{\mu}_{\ ;\nu;\rho} \ . \tag{9}$$

Now use the geodesic equation to get rid of the first term. Interchanging dummy indices and using the definition of the Riemann tensor, we finally arrive at

$$\mathcal{A}^{\mu} = -R^{\mu}_{\ \nu\rho\pi}t^{\nu}p^{\rho}t^{\pi} \ . \tag{10}$$

If the curvature is due to gravity, then this is the tidal acceleration, or the acceleration of one test particle as seen by an observer on the other. If the mass of the test particle is m, $m\mathcal{A}^{\mu}$ is the tidal force, which is the gravitational force equivalent to the curvature, which has been called the pseudo-Newtonian, or ψN-, force.[4,22]

Since the acceleration is taken to be physically "caused" by the distribution of matter and energy, represented by the stress-energy tensor, $T^{\mu\nu}$, and geometrically by the Riemann tensor, $R^{\mu}{}_{\nu\rho\pi}$, the two must be functionally related. Thus

$$T^{\mu\nu} = f^{\mu\nu}(R^{\alpha}{}_{\beta\gamma\delta}) \ . \tag{11}$$

The problem is that the former is a second rank tensor while the latter is a fourth rank tensor. The trace of the Riemann tensor, called the *Ricci* tensor, $R_{\beta\delta} = R^{\alpha}{}_{\beta\alpha\delta}$, is a second rank tensor which could be used in place of the Riemann tensor. Further, for mass-energy conservation the stress-energy tensor must be divergence-free, i.e. $T^{\mu\nu}{}_{;\nu} = 0$, which requires that $f^{\mu\nu}{}_{;\nu} = 0$. Now the Riemann tensor satisfies the second Bianchi identities (for a proof of which, see for example[4])

$$R^{\alpha}{}_{\beta\gamma\delta;\epsilon} + R^{\alpha}{}_{\beta\epsilon\gamma;\delta} + R^{\alpha}{}_{\beta\delta\epsilon;\gamma} = 0 \ , \tag{12}$$

contracting which yields the divergence-free *Einstein* tensor

$$\mathcal{E}_{\beta\delta} = R_{\beta\delta} - \frac{1}{2} g_{\beta\delta} R \ , \tag{13}$$

$R = g^{\mu\nu} R_{\mu\nu}$ being the *Ricci scalar*. Integrating $f^{\mu\nu}{}_{;\nu} = 0$, the simplest function $f^{\mu\nu}$ would be a constant, which means a constant times the covariantly constant metric tensor, $g^{\mu\nu}$, but that would "throw the baby out with the bath-water". The simplest meaningful expression is to add in a constant times the Einstein tensor, which yields the relation between curvature and matter-energy as

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu} \ , \tag{14}$$

where Λ is a constant of integration and the overall proportionality constant, κ, is put in with the matter-energy tensor, rather than the geometrical tensor.

5. What Difference Does It Make?

So what have we achieved? Just an alternate derivation of the Einstein field equations with a cosmological constant? I believe that this derivation had already been arrived at by Einstein. Undoubtedly, Einstein first *published* his field equations by using the variational principle. If he had got them earlier by this method, why did he not publish that derivation? I think that he felt that the Physics community was not so readily adopting the geometrical work of Grossmann and him, and wanted a derivation that would convince the community, and so waited till he had the field theoretic derivation worked out. Why would he not, then, have inserted the cosmological constant in there? There seem to be some letters between Hilbert and him that indicate he *had* done so in an earlier draft, but removed it later. The reason would be that the jewel in the crown of relativity was the explanation of the perihelion shift of Mercury by a theory with no free parameters. Had he put in the cosmological constant, there *would* be a free parameter. He was convinced that the theory was right. In this derivation the cosmological constant would *have to be there.*

In the derivation from the variational principle, it would have to be inserted by hand and so Hilbert was not getting it.

Even if this was the original derivation by Einstein, what difference does it make in any case? The point is that if we accept that Einstein inserted the constant later, by hand for his paper on cosmology a couple of years later,[2] it appears artificial and not part of the original theory, that had been verified. I have made much of the constant of integration coming with the geometry and not with the (physical) stress-energy tensor. It could, apparently, equally well have been inserted on the right side of the equation. In that case it would be an all-pervasive vacuum energy and not part of the geometry. Thus, it would need to be included in quantization. It is then assumed that the renormalization effect would come at the cut-off where quantum gravity takes over. On dimensional arguments one can argue that neither quantum mechanics nor relativity can be neglected when $\hbar, G, c^{-1}$ are non-negligible. Using them to make dimensional quantities, we get *Planck units*. It is further argued that "hence quantum gravity will set in *at* the Planck scale". Using the Planck energy density, which has the same units as the cosmological constant, *one gets a value about 126 orders of magnitude off from observation!* Hawking called this "the worst prediction in the history of Physics", and one cannot but agree with him. However, the original argument says that it must be invoked *by* the Planck scale, not *at* it. There is no reason that it should not set in *well before* the Planck scale. As I mentioned earlier, there is no theory of quantum gravity in any case, so one can only say what a valid theory for it *might* say. In that context Husain and Qureshi[23] have pointed out that the claimed problem does not arise. They point out that the argument is based on quantization of energy that is well defined in Minkowski space but is ambiguous in an arbitrarily curved spacetime. *There is no vacuum instability problem.*

So, what difference did the history make, if Einstein *had* got the cosmological constant earlier? It would have been perceived as part of the geometry, as a constant curvature, and the question of quantizing it would not arise till we had a "quantum geometrical" description of spacetime. One could then try to see what it would say, if anything, about the matter. The point is that it may be just our perception of what is "natural" that leads to the problems.

6. Conclusion and Discussion

In this paper it has been argued that it makes a huge difference whether the cosmological constant is regarded as part of the geometry or part of the stress-energy tensor. In the former role, as shown by Willem de Sitter,[24] it yields a static, constant curvature spacetime. Though this universe model does not change with time, an observer looking at the geodesic of a neighbouring test particle sees it accelerating *away* and so sees an expansion of this universe. Transforming to a cosmological time, using which one could talk of what is happening in a distant part of this universe "at the same time", the expansion would seem to be exponential (see for example[4]).

Put with the stress-energy tensor it would be an all-pervasive vacuum energy which gives a repulsive force for positive Λ.

Which way *should* it be looked at? Well, it was the expression involving the curvature that had to be determined, and so it is natural to keep the constant on the left side of the equation with the geometry. There is then no energy field causing the expansion — which appears as a kinematic effect. As such it cannot be considered for quantization without quantizing the rest of the spacetime. In that case the renormalization of the quantization of the rest of the curvature might, for example, balance that of the cosmological constant, so that there would be no net renormalization effect. I am not claiming that this *must* be so, only that we do not know and it could even be so. In fact, Maxwell had pointed out that his electromagnetic force, and hence energy, is opposite to the gravitational. Thus all gravitational effects should be expected to behave opposite to electromagnetic effects. Now the net mass-energy, material and gravitational, of the Universe is zero. That is what made it possible for Brout et al.[25] to propose that the Universe may be a vacuum fluctuation. In fact, have argued that there is reason to believe that any quantum gravitational effect should act opposite to other quantum fields.[26]

The current speculation is that the exponential expansion is due to a "mysterious form of dark energy" with thoroughly counter-intuitive properties. In fact, in some ways it is reminiscent of the luminiferous aether with any number of properties inserted by hand, with no evidence for them. This introduces any number of free parameters, reminiscent of Aristotle's epicycles. Regardless of these problems, let us go ahead and see what the claim is. The argument goes that if it were due to a cosmological constant, it would be orders of magnitude too small — *hence it must be zero*. You might well wonder how zero would be closer to the Planck value than the small observed value? Surely, that should be infinitely many orders of magnitude off! Why should it be zero in any case? If there is a constant appearing in the equations, should one not naturally look to the observational data to provide the value? The response normally given is that "it must be zero for some hitherto unknown reason, which will emerge later, but in the meantime we should assume that it is zero."

The standard model of cosmology is called the "Λ-CDM model", of which the latter stands for "cold dark matter" and the former for the cosmological constant. Thus, within observational errors, the data gives a behaviour of the Universe identical with what would be expected with a cosmological constant. To argue that it is a dynamical field seems perverse. Alternatively, we can say that the Almighty has set it to behave that way just to confuse us — transferring the perversity to Him (as we tend to do for most of our sins). Perhaps He is busting a rib up in Heaven, laughing at how "He has confused these arrogant blighters!" Why do we not simply accept the obvious explanation of the cosmological constant and determine it, just as we do Newton's constant of gravity? I am forced to conclude that it is because we would then eliminate a whole spew of papers and students supervised, based on wild speculations. In this conference "the Hubble tension" has been taken to imply

that the "dark energy" underwent a phase transition and started to behave like the cosmological constant.[27] The essence of the argument was because "the ΛCDM is correct, and if it does not fit at the very early stage then the Λ cannot be constant", despite the fact, as pointed out,[28] that the accuracy of the measurements for it are not yet sufficient for any definite conclusion. As we have no idea of the Physics of the CDM, why should there not be a phase transition for *it*? That would appear to be far more natural as matter *does* undergo phase transitions. Because people have got wedded to "the mysterious dark energy" — the dark side of the force.

Acknowledgments

I am most grateful for very useful comments to Francesco De Paolis, Viqar Husain, Babar Qureshi and Khalid Saifullah. I am also most grateful to the Organizing Committee of MG 16 for providing me a Registration fee waiver to participate in the Virtual Meeting.

References

1. J.B. Barbour, *The Discovery of Dynamics: A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories* (Oxford University Press 2001).
2. A. Einstein, "Kosmologische betrachtungen zur allgemeinen relativitätstheorie", *Preuss. Akad. der Wissen. Sitz.* **1917 (part 1)** 142–152.
3. A. Einstein, "Zur allgemeinen relativitätstheorie", *Preuss. Akad. der Wissen. Sitz.* **1915 (part 2)** 844–847; "Grundlage der allgemeinen relativitätstheorie", *Annalen der Physik* **49** (1915) 769–822.
4. A. Qadir, *Einstein's General Theory of Relativity* (Cambridge Scholars Publishing 2020).
5. E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae", *Proc. Nat. Acad. Sci.* **15** (1929) 168–173.
6. A. Friedman, "Über die krümmung des raumes", *Z. Phys.* **10** (1922) 377–386.
7. G. Gamow, "The evolution of the Universe", *Nature* **162** (1970) 680–682.
8. A. Sandage, "Distances to galaxies, the Hubble constant, the Friedmann time and the edge of the World", *Proc. Symp. on the Galaxy and Distance Scale, in Quart. Jour. Roy. Astron. Soc.* **13** (1972) 282–296; "Observational tests of world models", *Ann. Rev. Astron. & Astrophys.* **26** (1988) 561–630.
9. A.A. Penzias and R.W. Wilson, "A measurement of excess antenna temperature at 4080 Mc/s", *Astrophys. Jour.* **142** (1965) 419–421.
10. S. Perlmutter et al, "Measurements of the cosmological parameters Ω and Λ from the first seven supernovae at $z \geq 0.35$", *Astrophys. Jour.* **483** (1997)m565–581.
11. A.G. Reiss et al., "Observational evidence from supernovae for an accelerating Universe and a cosmological constant", *Astron. Jour.* **116** (1998) 1009–10038.
12. G. Lemâitre, "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques", *Annales de la Société Scientifique de Bruxelles* **47** (1927) 49–59.
13. P.J.E. Peebles, *Principles of Physical Cosmology* (Princeton University Press 2019).
14. A. Einstein, "Zur elektrodynamik bewegter körper", *Annalen der Physik* **17** (1905) 891–921.

15. H. Minkowski, "Das relativitätsprinzip", *Annalen der Physik* **352** (1907) 927–938; "Die grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern", *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1908) 53–111; "Raum und Zeit", *Jahresbericht der Deutschen Mathematiker-Vereinigung* **18** (1909) 75–88.
16. A. Qadir, *Relativity: An Introduction to the Special Theory* (World Scientific 1989).
17. I. Newton, *Philosophiae Naturalis Principia Mathematica* S. Pepys Reg. Soc. Praeses 1686.
18. A. Einstein, "Relativitätsprinzip und die aus demselben gezogenen folgerungen", *Jahrbuch der Radioaktivität* **4** (1907) 411–462.
19. R.V. Pound and G.A. Rebka Jr., "Gravitational red-shift in nuclear resonance", *Phys. Rev. Lett.* **3** 439–441.
20. A. Einstein and M. Grossmann, "Entwurf einer verallgemeinerten relativitätstheorie und eine theorie der gravitation: I. Physikalischer teil von A. Einstein; II. Mathematischer teil von M. Grossmann", *Zeitschrift für Mathematik und Physik*, **62** (1913) 225–261.
21. A. Einstein and M. Grossmann, "Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie", *Zeitschrift für Mathematik und Physik*, **63** (1914) 215–225.
22. A. Qadir, "General relativity in terms of forces", *Proc. Third Regional Conference on Mathematical Physics*, eds. F. Hussain and A. Qadir (World Scientific 1990) pp. 481–490; "The gravitational force in general relativity", *M.A.B. Bég Memorial Volume*, eds. A. Ali and P.A. Hoodbhoy (World Scientific 1991) pp.159–178.
23. V. Husain and B. Qureshi, "Ground state of the Universe and the cosmological constant. A nonperturbative analysis", *Phys. Rev. Lett.* **116** (2016) 061302 (6 pages).
24. W. de Sitter, "On Einstein's theory of gravitation and its astronomical consequences", *Mon. Not. Roy. Astron. Soc.* **78** (1917) 3–28.
25. R. Brout, F. Englert and E. Gunzig, "The creation of the Universe as a quantum phenomenon", *Ann. Phys.* **115** (1978) 78–106.
26. A. Qadir, "On the reality of Hawking radiation", *Int. J. Mod. Phys. D* **28** (2019) 2040001 (11 pages).
27. M. Kamionkowski, "The Hubble tension", talk on Wednesday 07/07 at 16.40.
28. W. Freedman, "The Hubble constant tension", talk on Wednesday 07/07 at 17.15.

The irreducible mass of Christodoulou-Ruffini-Hawking mass formula

Yuan K. Ha

Department of Physics, Temple University
Philadelphia, Pennsylvania 19122 U.S.A.
E-mail: yuanha@temple.edu

We reveal three new discoveries in black hole physics previously unexplored in the Hawking era. These results are based on the remarkable 1971 discovery of the irreducible mass of the black hole by Christodoulou and Ruffini, and subsequently confirmed by Hawking.
1. The Horizon Mass Theorem shows that the mass at the event horizon of any black hole: neutral, charged, or rotating, depends only on twice its irreducible mass observed at infinity.
2. The External Energy Conjecture proposes that the electrostatic and rotational energy of a general black hole exist completely outside the horizon due to the nature of the irreducible mass.
3. The Moment of Inertia Property shows that every Kerr black hole has a moment of inertia. When the rotation stops, there is an irreducible moment of inertia as a result of the irreducible mass.
Thus after 50 years, the irreducible mass has gained a new and profound significance. No longer is it just a limiting value in energy extraction, it can also determine black hole dynamics and structure. What is believed to be a black hole is a physical body with an extended structure. Astrophysical black holes are likely to be massive compact objects from which light cannot escape.

Keywords: Black holes; quasi-local energy; irreducible mass; Kerr metric; moment of inertia.

1. 50th Anniversary 1971–2021

This article is to celebrate the 50th anniversary of the discovery of the mass-energy formula of a Kerr-Newman black hole in 1971 by introducing three new results recently found in black hole physics. Surprisingly, these results all invove the concept of the irreducible mass.

First, congratulations to Demetrios Christodoulou and Remo Ruffini for their remarkable discovery of the *irreducible mass*[1] of the black hole and confirmed by Hawking,[2] one of the most important concepts in black hole physics.

This year is also the 50th anniversary of the renormalization of Yang-Mills theory[3] in 1971. Congratulations to Gerard 't Hooft and the late Martinus Veltman for their elucidation of the quantum structure of electroweak interactions, one of the great achievements in 20th Century physics.

The irreducible mass formula discovered by Christodoulou and Ruffin in 1971 is the following:[1]

$$M^2 = \left(M_{irr} + \frac{Q^2}{4GM_{irr}}\right)^2 + \frac{J^2c^2}{4G^2M_{irr}^2}. \quad (1)$$

Here M is the total mass of the Kerr-Newman black hole, M_{irr} is the irreducible mass; Q is the electric charge and J is the angular momentum. All quantities are reckoned according to the distant observer. When Q and J are zero, the irreducible mass is the mass of a Schwarzschild black hole. 50 years later, the irreducible mass has gained unexpected new and profound significance besides energy extraction. It can also determine black hole dynamics and structure.

It is especially appropriate to explain the many definitions of a black hole in physics.[4] The mathematical black hole in general relativity has a singularity hidden by a horizon. However, neither singularity nor horizon has been observed. Compact objects like the one at the center of our galaxy are also called black holes in common usuage, even though their nature is still unknown. This is pointed out in the 2020 Nobel Prize in Physics citation. To Roger Penrose, the citation is 'for the discovery that black hole formation is a robust prediction of the general theory of relativity'. To Reinhard Genzel and Andrea Ghez, the citatation is 'for the discovery of a supermassive compact object at the center of our galaxy'. The term black hole is avoided. Strictly speaking, the black hole has not been discovered, but only a black hole-like object has been observed in astyrophysics. The Laplace 'dark star' introduced the concept of the black hole as a massive body from which light cannot escape due to its strong gravity.

Between 1965–1985, several important theorems on classical black holes were gradually discovered. They are known as:

(1) Singularity Theorem (1965),[5]
(2) Area Non-decrease Theorem (1972),[6]
(3) Uniqueness Theorem (1975),[7]
(4) Positive Energy Theorem (1983).[8]

These theorems have been well discussed for many years in general relativity and accepted as basic properties of the classical black hole. In recent years, three new results on black holes previously unexplored in the Hawking era are found. They were developed using the quasi-local energy approach and angular momentum consideration. Remarkably, they all contain the irreducible mass of Christodoulou and Ruffini. They are:

(5) Horizon Mass Theorem (2005),[9]
(6) External Energy Conjecture (2017),[10]
(7) Moment of Inertia Property (2018).[11]

These results are derived completely within general relativity and therefore legitimate. They add new properties to the classical black hole with potential to resolve several long-standing paradoxes in black hole physics.

2. Horizon Mass Theorem

Theorem. *For neutral and charged black holes, the horizon mass is always twice the irreducible mass observed at infinity.*

For rotating black holes, the horizon mass is found to be extremely close to twice the irreducible mass for all rotations. It is conjectured that a rigorous proof will eventually show that the horizon mass is exactly twice the irreducible mass.

In notation, it is simply

$$M(r_+) = 2M_{irr} \tag{2}$$

where r_+ is the horizon radius of the black hole. The theorem relates the mass of a black hole observed at the event horizon to its irreducible mass observed at infinity. The irreducible mass does not contain electrostatic and rotational energy. The Horizon Mass Theorem is the final outcome of quasi-local mass applied to black holes.

The quasi-local energy is one of the most important concepts in general relativity after decades of searching for a consistent definition of gravitational energy. It was finally obtained in 1993. The Brown and York expression[12] for quasi-local energy is given in terms of the total mean curvature of a surface bounding a volume for a gravitational system in four-dimensional spacetime. The total energy E , including binding energy, is given in the form of an integral,

$$E = \frac{c^4}{8\pi G}\int_{^2B} d^2x\sqrt{\sigma}(k - k^0), \tag{3}$$

where σ is the determinant of the metric defined on the two-dimensional surface ^2B ; k is the trace of the extrinsic curvature of the surface, and k_0 , the trace of the curvature of a reference space. For asymptotically flat reference spacetime, k^0 is taken to be zero. The expression in Eq.(3) is the basis for establishing the Horizon Mass Theorem.

For a Schwarzschild black hole, the total energy contained in a sphere enclosing the black hole at a coordinate distance r is calculated,[12–14]

$$E(r) = \frac{rc^4}{G}\left[1 - \sqrt{1 - \frac{2GM}{rc^2}}\right]. \tag{4}$$

At the Schwarzschild radius, $r = r_+ = 2GM/c^2$, the above equation reduces to

$$E(r) = \left(\frac{2GM}{c^2}\right)\frac{c^4}{G} = 2Mc^2, \tag{5}$$

giving the first case of the Horizon Mass Theorem in Schwarzschild spacetime, i.e. $M(r_+) = 2M$.

For a Reissner-Nordström black hole enclosed within a radius at coordinate r, the total energy calculated is now [9],

$$E(r) = \frac{rc^4}{G}\left[1 - \sqrt{1 - \frac{2GM}{rc^2} + \frac{GQ^2}{r^2c^4}}\right]. \tag{6}$$

Here M is the total mass of the black hole including electrostatic energy observed at infinity, and Q is the electric charge.

The irreducible mass is the final mass of the black hole when its charge is neutralized by adding oppositely charged particles, extracting energy from the black hole. It is defined as in electrostatics,

$$M = M_{irr} + \frac{Q^2}{4GM_{irr}}. \tag{7}$$

Inverting the equation to solve for M_{irr}, we find

$$M_{irr} = \frac{M}{2} + \frac{M}{2}\sqrt{1 - \frac{Q^2}{GM^2}}. \tag{8}$$

The horizon radius of a Reissner-Nordström black hole is known to be

$$r_+ = \frac{GM}{c^2} + \frac{GM}{c^2}\sqrt{1 - \frac{Q^2}{GM^2}}. \tag{9}$$

Combining Eq.(6) and Eq.(9), we find the mass contained within the horizon to be

$$\frac{E(r_+)}{c^2} = M(r_+) = M + M\sqrt{1 - \frac{Q^2}{GM^2}}, \tag{10}$$

i.e. $M(r_+) = 2M_{irr}$. This is the second case of the Horizon Mass Theorem in Reissner-Nordström spacetime. It is seen that the horizon mass of the charged black hole depends only on the energy of the black hole when it is neutral.

We proceed next to the case for a slowly rotating black hole with mass M and angular momentum J . The total energy contained within a sphere of radius r can only be given by an approximate expression. This is due to the complexity of the Kerr metric, and more importantly, due to the fact that the Kerr metric only has axial symmetry instead of spherical symmetry. At the horizon, it is found that,[15]

$$\begin{aligned} E(r) = {} & \frac{rc^4}{G}\left[1 - \sqrt{1 - \frac{2GM}{rc^2} + \frac{\alpha^2}{r^2}}\right] \\ & + \frac{\alpha^2 c^4}{6rG}\left[2 + \frac{2GM}{rc^2} + \left(1 + \frac{2GM}{rc^2}\right)\sqrt{1 - \frac{2GM}{rc^2} + \frac{\alpha^2}{r^2}}\right] + \cdots \end{aligned} \tag{11}$$

where $\alpha = J/Mc$ is the angular momentun length paraneter. The leading term of the expression is similar to the energy expression in the Reissner-Nordström case,

suggesting that the mass at the horizon is twice the irreducible mass. The next term depends on α^2, adding a small contribution to the leading term.

The irreducible mass of a Kerr black hole is the final mass when its rotational energy is completely extracted by adding external particles, such as in the Penrose process.[16] It is given in the form

$$M_{irr}^2 = \frac{M^2}{2} + \frac{M^2}{2}\sqrt{1 - \frac{J^2c^2}{G^2M^4}}. \tag{12}$$

The horizon radius in this case is,

$$r_+ = \frac{GM}{c^2} + \frac{GM}{c^2}\sqrt{1 - \frac{J^2c^2}{G^2M^4}}. \tag{13}$$

An approximate relation for the horizon energy is therefore found,

$$E(r_+) \approx 2M_{irr} + O(\alpha^2). \tag{14}$$

The conclusion is that there is very little rotational energy insiide the Kerr black hole.

It is natural to extend the quasi-local energy investigation to include higher rotations, and logically, all rotations. However, a severe challenge appeared at this stage and progress on black hole rotation in this approach stopped. The calculations became extremely difficult to perform. No analytical expression or numerical evaluation could achieve an *exact* expression for the horizon mass of the Kerr black hole. An analysis of the horizon mass in the teleparallel equivalent formulation of general relativity[17] reveals that it is strikingly close to twice the irreducible mass $2M_{irr}$ for all range of the parameter $0 \leq \alpha < GM/c^2$. The tiny discrepancy is likely due to evaluating method and describing the spherical horizon region in a system with intrinsic axial symmetry. A general principle based on equipartition of energy at the horizon also suggests the horizon mass result for the Kerr black hole[18] by invoking one-half of the horizon mass for compensating the negative gravitational potential energy and the other half for supplying the irreducible mass.

We give a heuristic argument for the Horizon Mass Theorem with the area concept of a black hole.[9] It has been known from the Kerr metric that the area at the event horizon of a Kerr black hole for all rotations is[1]

$$A = 4\pi(r_+^2 + \alpha^2) = \frac{16\pi G^2 M_{irr}^2}{c^4}, \tag{15}$$

and the area of a Schwarzschild black hole of mass M_S and radius R_S is

$$A = 4\pi R_S^2 = 4\pi\left(\frac{2GM_S}{c^2}\right)^2 = \frac{16\pi G^2 M_S^2}{c^4}. \tag{16}$$

The two areas can be related by invoking Hawking's Area Non-decrease Theorem in the energy extraction process. The theorem asserts that the area of a Kerr black hole is the same as the area of the final Schwarzschild black hole when rotational energy is extracted in a smooth and reversible process. Since the horizon mass of the

Schwarzschild black hole is proven to be twice its asymptotic mass M_S, the horizon mass of the Kerr black hole in this process is therefore $2M_{irr}$. The result applies to all rotations. It is believed that a rigorous mathematical proof will eventually show that the horizon mass is *exactly* twice the irreducible mass.

The Horizon Mass Theorem is crucial for understanding processes occuring near the horizon, such as the merging of two black holes,[19] and quantum emission of Hawking radiation.[20]

3. External Energy Conjecture

Proposition. *The electrostatic energy and rotational energy of a general black hole exist completely outside the horizon.*

The conjecture is a direct consequence of the irreducible mass in the Horizon Mass Theorem.

By definition, the irreducible mass does not contain rotational energy or electrostatic energy. A rotating black hole does not have rotational energy inside the horizon; therefore rotational energy exists outside the surface. Similarly, an electrically charged black hole does not have electrostatic energy inside. Electrostatic energy exists only outside, like that of a conductor. When quantum particles carrying electric charges and spins reach the black hole, they are forbidden to enter inside. They can only stay outside or at the surface. Since all matter particles in Nature are quantum particles, this makes the interior of the black hole completely hollow. Classical particles do not exist in Nature; they are a tool in classical mechanics.

We may generalize the External Energy Conjecture to include other energies of a black hole and introduce a new paradigm.[11]

External Energy Paradigm:

> *All energies of a black hole are external quantities. They include: constituent mass, gravitational energy, electrostatic energy, magnetic energy, rotational energy, heat energy, etc.*

The validity of this paradigm will be demonstrated in the next section in which the moment of inertia of a black hole is presented.

4. Moment of Inertia Property

Statement. *A black hole with an angular momentum and an angular velocity at the event horizon has a moment of inertia given by:*

$$\textit{coefficient} \times \textit{Kerr mass} \times \textit{(ergosphere radius)}^2 \ .$$

When rotation stops, there is an irreducible moment of inertia given by:

$$\textit{irreducible mass} \times \textit{(Schwarzschild radius)}^2 \ .$$

The result of this statement is derived solely from the Kerr metric and is therefore a *bona fide* property of general relativity. Every black hole has a moment of inertia, even when it is not rotating. Moment of inertia indicates structure of the black hole.

The Kerr metric[21] discovered in 1963 ushered in a new epoch in general relativity and in astrophysics. It is absolutely indispensable for the study of rotating black holes. We demonstrate the existence of moment of inertia uniquely from the Kerr metric, using angular momentum and angular velocity consideration. We present the Kerr metric in an *explicit* form of the Boyer-Lindquist coordinates[22] (t, r, θ, ϕ) so that the metric coefficients can be readily extracted for calculation. It contains two constants $\alpha = J/Mc$ and $m = GM/c^2$ for the stationary case,

$$\begin{aligned} ds^2 = &\left(\frac{r^2 + \alpha^2 cos^2\theta - 2mr}{r^2 + \alpha^2 cos^2\theta}\right) c^2 dt^2 + \left(\frac{4m\alpha r sin^2\theta}{r^2 + \alpha^2 cos^2\theta}\right) c d\phi dt \\ &- \frac{[(r^2 + \alpha^2)(r^2 + \alpha^2 cos^2\theta) + 2mr\alpha^2 sin^2\theta]}{r^2 + \alpha^2 cos^2\theta} sin^2\theta d\phi^2 \\ &-(r^2 + \alpha^2 cos^2\theta)\left(d\theta^2 + \frac{dr^2}{r^2 + \alpha^2 - 2mr}\right). \end{aligned} \tag{17}$$

The Kerr spacetime rotates with different angular velocities at different locations. The angular velocity at a point is defined as the change in azimuthal angle ϕ with respect to the change in coordinate time t. It can be expressed in terms of the metric coefficients as

$$\Omega = \frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}}. \tag{18}$$

where $g_{t\phi} = g_{\phi t}$. At the equatorial region, $\theta = 90^\circ$, the angular velocity expression at a distance r can be written as

$$\Omega = \frac{2m\alpha rc}{(r^2 + \alpha^2)r^2 + 2mr\alpha^2}. \tag{19}$$

Further simplification can be achieved at the event horizon $r = r_+$, using the identity $r_+^2 + \alpha^2 = 2mr_+$, i.e.

$$\Omega_+ = \frac{\alpha c}{r_+^2 + \alpha^2}. \tag{20}$$

In terms of actual physical quantities, we have an exact algebraic relation for the angular velocity of the Kerr black hole,

$$\Omega_+(J) = \frac{\dfrac{J}{M}}{\dfrac{2G^2M^2}{c^4}\left[1 + \sqrt{1 - \dfrac{J^2c^2}{G^2M^4}}\right]}. \tag{21}$$

Given an angular momentun J and a Kerr mass M determined by a distant observer, the angular velocity at the event horizon can be obtained in radians/sec.

Equivalently, we can express the angular momentum J in terms of the angular velocity Ω through the moment of inertia $I(\Omega)$ in the form

$$J(\Omega) = I(\Omega) \cdot \Omega. \tag{22}$$

We find, from Eq.(21), after substantial algebra,

$$J(\Omega_+) = \frac{M\left(\frac{2GM}{c^2}\right)^2}{1+\left(\frac{2GM}{c^2}\right)^2 \frac{\Omega_+^2}{c^2}} \cdot \Omega_+. \tag{23}$$

The crucial step here is to extract a radius factor in the above relation. It is recognized that the factor $2GM/c^2$ is the ergosphere radius at the equator, since M is the Kerr mass which includes rotational energy. Accordingly, the moment of inertia of the Kerr black hole is

$$I(\Omega_+) = \frac{M\left(\frac{2GM}{c^2}\right)^2}{1+\left(\frac{2GM}{c^2}\right)^2 \frac{\Omega_+^2}{c^2}}. \tag{24}$$

It is further recognized that the entire denominator in the above expression becomes a dimensionless number and acts as a numerical coefficient. The structure of the Kerr black hole is therefore,

$$\textit{coefficient} \times \textit{Kerr mass} \times \textit{(ergosphere radius)}^2.$$

As angular momentum is continually reduced in the energy extraction process, a slowly rotating black hole is formed. In the static limit, the quantity $\Omega_+^2 \to 0$ first, the Kerr mass becomes the irreducible mass, $M \to M_{irr}$ and the coefficient becomes exactly equal to 1. The moment of inertia of a Schwarzschild black hole is derived. It is the limiting value of the moment of inertia of the Kerr black hole in Eq.(24), given by

$$I = M_S R_S^2. \tag{25}$$

There is an irreducible moment of inertia of the Kerr black hole as a result of the irreducible mass. It is the rotatonal analogue of the rest mass of a moving body $E = mc^2$.

A further observation that leads quickly to the irreducible moment of inertia is by considering the angular momentum definition in the Kerr metric $J = M\alpha c$ and the angular velocity Ω_+ at the horizon found in Eq.(20). We find,

$$I = \frac{J}{\Omega_+} = M\left(r_+^2 + \alpha^2\right) = \sqrt{M_{irr}^2 + M_{rot}^2}\left(r_+^2 + \alpha^2\right). \tag{26}$$

As the rotational parameter $\alpha \to 0$, the rotational mass $M_{rot} \to 0$; while the horizon radius $r_+ \to R_S$. In the limit, $I = M_{irr} R_S^2 = M_S R_S^2$. If one directly puts $J = 0$ and

$\Omega_+ = 0$ in the definition $I = J/\Omega_+$, one would get an undefined result 0/0. The static limit of the Kerr black hole hole has the structure,

$$\textit{irreducible mass} \times \textit{(Schwarzschild radius)}^2.$$

A Schwarzschild black hole does not have an axis of rotation. Introducing an axis destroys its spherical symmetry. The center is therefore the point of symmetry. The moment of inertia $I = M_S R_S^2$ is a statement about the mass distribution of a body with respect to the center in that the total mass is to be located at the Schwarzschild radius. A natural interpretation is that the static black hole is a hollow massive shell. This would go against the Equivalence Principle and the Singularity Theorem. It is possible that the static black hole resulting from the energy extraction process of a Kerr black hole is fundamentally different from the original Schwarzschild black hole. The static black hole derived from a Kerr black hole may be a quasi-black hole. A quasi-black hole has the same exterior spacetime as that of the Schwarzschild black hole but without the conceptual difficulties associated with the latter.[23] Singularity does not exist because particles are forbidden to cross the horizon by the very presence of the moment of inertia. The black hole firewall is such a scenario.[24] The quasi-black hole would also provide a physical surface where electric charges can stay instead of hovering without cause in the case of the charged black hole. In addition, if the surface area is identified as the entropy of a black hole according to Bekenstein[25] and Hawking,[6] then it is logical to expect that all mass of the quasi-black hole is at the surface. The moment of inertia is a new property of the Kerr black hole.

5. Epilogue

After 50 years, the irreducible mass has progressed from the gedanken Penrose process to quantum energy extraction in gamma-ray bursts and active galactic nuclei emissions.[26] Instead of being a cold and inert body, the real black hole is a highly active object. The irreducible mass further plays a central role in newly discovered black hole properties. These results may resolve some of the long-standing paradoxes in black hole physics.

References

1. D. Christodoulou and R. Ruffini, *Phys. Rev. D* **4**, 3552 (1971).
2. S. W. Hawking, *Phys. Rev. Lett.* **26**, 1344 (1971).
3. G. 't Hooft and M. Veltman, *Nucl. Phys. B* **44**, 189 (1972).
4. E. Curiel, *Nature Astronomy* **3**, 27 - 34 (2019).
5. R. Penrose, *Phys. Rev. Lett.* **14**, 57 (1965).
6. S. W. Hawking, *Commun. Math. Phys.* **25**, 152 (1972).
7. D. C. Robinson, *Phy. Rev. Lett.* **34**, 905 (1975).
8. R. Schoen and S. T. Yau, *Commun. Math. Phys.* **65**, 45 (1979).

9. Y. K. Ha, *Int. J. Mod. Phys. D* **14**, 2219 (2005).
10. Y. K. Ha, *Mod. Phys. Lett. A* **32**, 1730021 (2017).
11. Y. K. Ha, *Int. J. Mod. Phys. A* **33**, No. 31, 1844025 (2018).
12. J. D. Brown and J. W. York, Jr., *Phys. Rev. D* **47**, 1407 (1993).
13. J. W. Maluf, *J. Math. Phys.* **36**, 4242 (1995).
14. Y. K. Ha, *Gen. Relativ. Gravit.* **35**, 2045 (2003).
15. E. A. Martinez, *Phys. Rev. D* **50**, 4920 (1994).
16. R. M. Floyd and R. Penrose, *Nature Phys. Sci.* **229**, 177 (1971).
17. J. W. Maluf, E. F. Martins and A. Kneip, *J. Math. Phys.* **37**, 6302 (1996).
18. N. Dadhich, *Curr. Sci.* **76**, 831 (1999).
19. B. P. Abbott *el al.*, *Phys. Rev. Lett.* **116**, 061102 (2016).
20. S. W. Hawking, *Commun. Math. Phys.* **43**, 199 (1975).
21. R. P. Kerr, *Phys. Rev. Lett.* **11**, 237 (1963).
22. R. M. Boyer and R. W. Lindquist, *J. Math. Phys.* **8**, 265 (1967).
23. J. V. Narlikar and T. Padmanabhan, *Found. Phys.* **18**, No. 6, 639 (1988).
24. A. Almheiri, D. Maroff, J. Polchinski and J. Sully, *J. High Energy Phys.* **2013**, 62, (2013).
25. J. Bekenstein, *Lett. Nuovo. Cim.* **4**, 737 (1972).
26. R. Moradi, J. A. Rueda, R. Ruffini, Y. Wang, *A&A* **649**, A75 (2021)

Reshaping our understanding on structure formation with the quantum nature of the dark matter

C. R. Argüelles,[1,2,*] E. A. Becerra-Vergara,[2,3,4] A. Krut,[2] R. Yunis,[2,3] J. A. Rueda,[2,3,5,6] R. Ruffini[2,3,6]

[1] *Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, B1900FWA La Plata, Argentina*

[2] *International Center for Relativistic Astrophysics Network – ICRANet, Piazza della Repubblica 10, 65122 Pescara, Italy*

[3] *ICRA, Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy*

[4] *GIRG, Escuela de Física, Universidad Industrial de Santander, 680002, Cra 27 calle 9, Bucaramanga, Colombia*

[5] *ICRANet-Ferrara, Dip. di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, 44122 Ferrara, Italy*

[6] *INAF, Istituto de Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, 00133 Rome, Italy*

[*] *carguelles@fcaglp.unlp.edu.ar*

We study the non-linear structure formation in cosmology accounting for the quantum nature of the dark matter (DM) particles in the initial conditions at decoupling, as well as in the relaxation and stability of the DM halos. Differently from cosmological N-body simulations, we use a thermodynamic approach for collisionless systems of self-gravitating fermions in General Relativity, in which the halos reach the steady state by maximizing a coarse-grained entropy. We show the ability of this approach to provide answers to crucial open problems in cosmology, among others: the mass and nature of the DM particle, the formation and nature of supermassive black holes in the early Universe, the nature of the intermediate mass black holes in small halos, and the *core-cusp* problem.

Keywords: Dark Matter; Galaxies: Super Massive Black Holes - Halos; Self-gravitating Systems: fermions

1. Introduction

In the particle DM paradigm, different particle candidates (e.g. WIMPS, axions, sterile neutrinos, etc.) have different early Universe histories that, in turn, translates into different evolution of primordial perturbations in the cosmic plasma which seeds the formation of galaxies at later stages.[1] The effects of the DM microphysics is, in these cases, often quantified only by the (linear) matter power spectrum: a measure of the distribution of density perturbations before non-linear processes are involved. Such non-linear processes typically include the gravitational collapse of primordial structures, key to the formation of the observed structures in the late Universe, such as galaxies and clusters.

From the point of view of structure formation, the quantum nature of particle DM has been traditionally involved only in the initial conditions before self-gravity comes into play, neglecting any role it may have in the dynamics of the formed systems. However, whether or not this DM microphysics affects the overall structure of DM halos has been questioned in recent works. That is, our ignorance about its potential (quantum) effects through the innermost regions of the halos and their stability, may represent a hole in our knowledge of these systems.

Indeed, the limited inner spatial resolution of DM halos inherent of cosmological N-body simulations implies huge uncertainties on the DM central distribution. It includes the open questions on how the DM is correlated with baryonic matter on such inner scales, and ultimately, how is the relation with the supermassive black hole (SMBH) at the center of large galaxies, and how they can co-exist in a steady state. Moreover, the use of classical point masses as the building blocks of matter (without considering the quantum nature of the particles) in traditional N-body simulations within the LCDM paradigm does not allow for any new possible sources of quantum-pressure in DM halos (e.g. as recently shown within 3D numerical simulations for bosons 2 or fermions 3). The arising of these quantum effects at the center of equilibrium systems of self-gravitating quantum particles (either bosons or fermions) is a long-known physical phenomenon.[4]

Motivated by these open issues, we here present an alternative mechanism of structure formation in cosmology based on thermodynamics of self-gravitating fermions. In such a scenario, the quantum nature of the DM particles is duly taken into account in the physics of decoupling in the early Universe (with corresponding effects in the linear power spectrum), as well as at DM halo formation and subsequent stability and morphology.

2. DM halo formation: A thermodynamic approach

The thermodynamic approach for DM halo formation used here is motivated by the seminal work of *Lynden-Bell* on violent relaxation.[5] This approach was originally applied for collisionless stellar systems, and few decades later extended to indistinguishable particles such as massive neutrinos in Ref. 6. Relevant extensions to Lynden-Bell theory allowing for out of equilibrium effects (i.e particle evaporation), were performed in Refs. 7, 8 using the framework of kinetic theory. Quasi-stationary solutions to such equations (see Eq. (2) below), lead to a most-probable, coarse-grained distribution function (DF) at (violent) relaxation, of more realistic applications than the original Lynden-Bell's version: a tidally truncated Fermi–Dirac DF of the form

$$\bar{f}(\epsilon \leq \epsilon_c) = \frac{1 - e^{(\epsilon-\epsilon_c)/kT}}{e^{(\epsilon-\mu)/kT} + 1}, \qquad \bar{f}(\epsilon > \epsilon_c) = 0\,, \tag{1}$$

where $\epsilon = \sqrt{c^2p^2 + m^2c^4} - mc^2$ is the particle kinetic energy, μ is the chemical potential with the particle rest-energy subtracted off, T is the effective temperature, k is the Boltzmann constant, c is the speed of light, and m is the DM fermion mass. The full set of dimensionless-parameters of the model are defined by the temperature, degeneracy and cutoff parameters, $\beta = kT/(mc^2)$, $\theta = \mu/(kT)$ and $W = \epsilon_c/(kT)$, respectively (when these parameters are evaluated at the center of the configuration, they will be written with a subscript 0).

DM halos built upon such a fermionic DF in a cosmological framework, naturally lead to finite-sized halos which can be stable, extremely long-lived, and holding key implications to supermassive BH formation from DM as shown here (and detailed in Ref. 3).

In the limiting case of $\epsilon_c \gg \epsilon$, there is no particle escape, and (1) takes the form of the traditional (coarse-grained) Fermi-Dirac DF obtained in the original work of Lynden-Bell. Indeed, this DF is a stationary solution of the Vlasov-Poisson (VP) equation governing collisionless systems,[7] that is a solution of Eq. (2) shown below, with $\mathbf{J} = 0$, usually called 'VP dynamically-stable' solutions. The more realistic "tidally-truncated" DF (1) is another kind of quasi-stationary solution of the more general *convection-diffusion* kinetic equation (2), as demonstrated in Refs. 7, 9. This accounts for the 'microphysics' of the long-range particle-particle interactions[a] involved in the diffusion current $\mathbf{J}$, caused mainly by the rapid fluctuations of the gravitational potential (proper of violent relaxation processes):

$$\frac{\partial \bar{f}}{\partial t} + \mathbf{v}\frac{\partial \bar{f}}{\partial \mathbf{r}} - \nabla\Phi\frac{\partial \bar{f}}{\partial \mathbf{v}} = -\frac{\partial \mathbf{J}}{\partial \mathbf{v}}. \tag{2}$$

When the gravitational fluctuations die down at the end of relaxation, the diffusion current approaches zero (i.e. no more evaporation), and the DF (1) remains as a stationary solution of the VP equation. Now, these dynamically-stable VP solutions can be easily used to build DM halos in (hydrostatic) equilibrium able to reproduce galaxy rotation curves (as shown in Refs. 10, 11). However, such analysis does not explain neither *how* such a collisionless self-gravitating system reaches the required steady state, nor if they maximize the coarse-grained entropy S of such a fermion-gas, nor if they are just transitional (unstable/unreachable) states. This stems from the fact that VP dynamical stability *does not* necessarily imply thermodynamical stability, therefore some dynamically stable solutions for DM halos in hydrostatic equilibrium can be more likely to occur in Nature than others!

[a]For isolated systems made of fermions, such a current can make (2) to be of the form of a fermionic Landau equation, conserving mass and energy, and continuously increase the Fermi-Dirac entropy functional (i.e. fulfilling H-theorem).[9]

In order to find dynamically and thermodynamically stable configurations of collisionless self-gravitating fermions in General Relativity (GR), able to model realistic DM halos, it is needed to solve the problem of finding solutions that maximize the global entropy $S = \int_0^R s(r) e^{\lambda/2} 4\pi r^2 \mathrm{d}r$ (at fixed total energy and particle number), where $s(r)$ is the entropy density given by the Gibbs-Duhem relation

$$s(r) = \frac{P(r) + \rho(r) - \mu(r) n(r)}{T(r)}, \tag{3}$$

$e^{\lambda/2} 4\pi r^2 dr$ is the proper volume, being e^{λ} the spatial component of the spherically-symmetric spacetime metric: $\mathrm{d}s^2 = e^{\nu} c^2 \mathrm{d}t^2 - e^{\lambda} \mathrm{d}r^2 - r^2 \mathrm{d}\Theta^2 - r^2 \sin^2\Theta \mathrm{d}\phi^2$, with $(r,\, \Theta,\, \phi)$ the spherical coordinates, and ν and λ depend only on r.

The mass-energy density $\rho(r)$, pressure $P(r)$ and particle number density $n(r)$ are directly obtained as the corresponding integrals of $\bar{f}(p)$ over momentum space (bounded from above by $\epsilon \leq \epsilon_c$) as detailed in Ref. 3. This makes a four-parametric equation of state $(\rho(r),\, P(r))$ depending on (β, θ, W, m) according to the dimensionless parameters in (1). The stress-energy tensor is the one of a perfect fluid with the above density and pressure, so the self-gravitating configurations of fermions are solutions of the hydrostatic equilibrium equation, with the local $T(r)$ and $\mu(r)$ fulfilling, respectively, the Tolman and Klein equilibrium relations (see Ref. 3 for details). We shall analyze which of such solutions are thermodynamically stable and of astrophysical interest regarding realistic DM halos in cosmology.

The above stability analysis can be solved in an elegant manner via the 'Katz criterion'.[12] This allows to find the full set of thermodynamically-stable solutions (i.e. maxima of global entropy S) along the full series of equilibrium solutions of self-gravitating fermions. This is a powerful and rigorous method which relies only in the derivatives of the caloric curve given by total energy $E = c^2 M(R)$ vs. $1/T_{\infty}$ (like in Fig. 1), and without the need of calculating explicitly the (rather involving) second order entropy variations $\delta^2 S$. This is a commonly used method to calculate the stability of self-gravitating systems (see e.g. Refs. 9, 3). We work in the microcanonical ensemble applied for isolated systems since, as explained in detail in Ref. 9, it is the most appropriate for astrophysical applications. The relevant thermodynamic potential to be extremized in this ensemble is the coarse-grained entropy S.

In order for such a maximization entropy problem to be well defined, the self-gravitating solutions with given particle number $N = \int_0^R n(r) e^{\lambda/2} 4\pi r^2 \mathrm{d}r$ and total energy $E = c^2 M(R)$ must be bounded in radius, as occurring when working with the tidally-truncated DF (1). We recall this is not the case for the standard Fermi-Dirac DF (i.e coarse-grained solution of the VP equation), which instead leads to unbounded systems. Therefore, our approach is well posed and without the need to invoke unrealistic bounding boxes.[3]

3. Astrophysical & cosmological applications

We use the following method to assure that the fixed boundary conditions — such as N and corresponding $M(R)$ required to maximize the global entropy S — do correspond to values of a given fermionic halo in a realistic cosmological framework: **(i)** calculate the linear matter power spectrum for such fermionic (quantum) candidates; **(ii)** apply the corresponding Press-Schechter formalism to obtain the virial halo mass, $M_{\rm vir} \equiv M(R \equiv R_{\rm vir})$, with associated redshift $z_{\rm vir}$ for such a particle mass; and then **(iii)** use the the most probable coarse-grained DF (1) obtained at the end of violent relaxation (valid for that redshift), to obtain the full families of DM (equilibrium) profiles in agreement with above virial constraints. Having done this, we can analyze the (thermodynamic) stability of such halos using the Katz criterion in a given cosmological framework, and further check their astrophysical applications.

This strategy was recently applied in Ref. 3 within a WDM cosmology for fermion masses of $\mathcal{O}(10)$ keV, and exemplified here for an average DM halo of mass $M_{\rm vir} \approx 5 \times 10^{10} M_{\odot}$, which started to form at $z_{\rm vir} \approx 10$ according to the Press-Schechter formalism in our cosmological setup.

We show in Fig. 1 the caloric curve in dimensionless variables, $-\hat{M}$ vs. $1/\hat{T}_{\infty}$, for an average DM halo (with corresponding $\hat{N} = 76.25$, and $W_0 - \theta_0 = 24$ typical of average-size galaxies[3]) and for a particle mass of $m = 48$ keV. The stability analysis is based on the Katz criterion and follows the global procedure **(i-iii)** mentioned above. A practical *rule of thumb* to understand how the Katz criterion works is: (a) an unstable mode (coming from stable ones) arises when the caloric curve *rotates clockwise* (i.e. when the negative slope in the $(-\hat{M},\, \hat{T}_{\infty}^{-1})$ curve becomes infinite for then turning into positive), (b) when the same curve *rotates counterclockwise* a stable mode has been re-gained (the later implying that a positive slope turned into negative just after becoming vertical). In this sense, once in a given unstable branch of the caloric curve, it is necessary as many counterclockwise turns of the curve as clockwise passed, to regain the thermodynamic stability (see Appendix A in Ref. 3 for details).

In Fig. 1, we display the full family of thermodynamically-stable (i.e. maxima of S) solutions in continuous-blue line, and the thermodynamically unstable ones (i.e. minimum or saddle points of S) in dotted-violet. Interestingly, there are two spiral features in this caloric curve. The upper one arises due to the fermion (quantum) degeneracy which avoids the *gravothermal catastrophe* typical of Boltzmannian distributions (as first shown for classical systems in Ref. 13). The second spiral feature is of relativistic origin and lead to the turning point criterion for gravitational collapse discussed below.

We now highlight the main conclusions of this stability analysis, and its cosmological and astrophysical consequences in three different subsections.

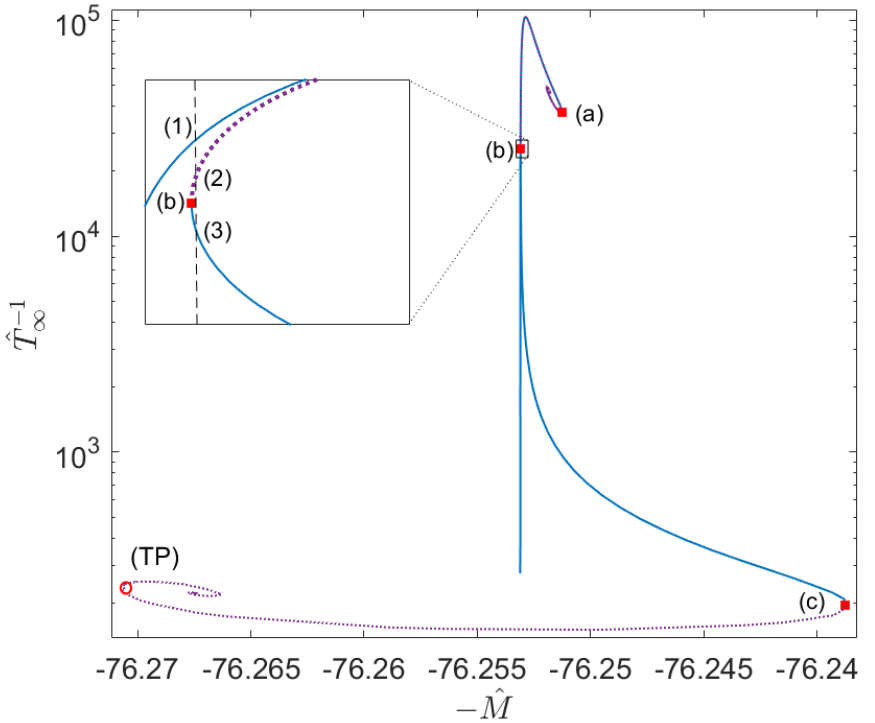


Fig. 1. Series of equilibrium solutions along the caloric curve for tidally-truncated configurations of $mc^2 = 48$ keV fermions, for typical DM halos of $M_{\rm vir} \approx 5 \times 10^{10} M_\odot$. The equilibrium states within the continuous-blue branches are thermodynamically (and dynamically) stable, while the dotted-violet branches — between (a) and (b) and after (c) — are unstable, according to the Katz criterion. Figure taken from Ref. 3.

3.1. *The shape of the fermionic DM halos*

We analyze from the caloric curve of Fig. 1 all the different kinds of density profiles for a conveniently fixed value of the total energy $\hat{M}$ (see vertical black-dashed line in the inset-box of Fig. 1, leading to the associated profiles of Fig. 2). There are two different families of stable and astrophysical DM profiles in a $\mathcal{O}(10)$ keV cosmology, arising from our (thermodynamic) halo-formation approach. They are either King-like, as the solution (1) in Fig. 1 (resembling the Burkert profile), or they develop a *core-halo* morphology (like the solution (3) in Fig. 1). Importantly, both kinds of solutions can agree with the observational DM surface density relation of galaxies (the latter $\propto \rho(r_{pl})\, r_h$, with $\rho(r_{pl})$ the density at plateau, and r_h the one-halo scale-length of such profiles), as demonstrated in Ref. 3. In the first case, the fermions are in the dilute regime (i.e. $\theta_0 \ll -1$) and correspond to a global maxima of entropy, while in the second case, the degeneracy pressure (i.e. Pauli principle) is holding the quantum core against gravity, and correspond to a local

maxima of entropy. This kind of fermionic *core-halo* solutions are usually referred as the Ruffini-Argüelles-Rueda (RAR) model.[10,14] They are extremely long-lived, meta-stable states as demonstrated in Ref. 3, with a lifetime $\propto e^N$, and more likely to arise in Nature than the former as argued in Ref. 9.

Remarkably, solutions like (3) in Fig. 1 can fit the rotation curve in galaxies, such as in the case of the Milky Way as shown in Fig. 3, while their dense quantum core can work as an alternative to the SMBH in non-active galaxies.[10,11,15,16] Indeed, this latter statement was analyzed in the case of our own Galaxy in Ref. 15 within the extended RAR model for $mc^2 = 56$ keV. It has been there shown that the gravitational potential of the DM dense core of such a configuration leads to better fit of the orbits of the S2 and G2 objects around Sgr A*, with respect to the one obtained by the central BH model (see Figs. 4 and 5). Furthermore, a geodesic in the gravitational field of the DM profile of the extended RAR model naturally predicts the post-pericenter slowing down phenomenon observed in G2 (see Fig. 5), result that contrasts with the BH hypothesis where it has been proposed that G2 is being slowed down by a drag force caused by an accretion flow onto the massive BH over which G2 should move.[17,18] Finally, strengthening this DM interpretation of the Galactic Center, in Ref. 16, this result has been extended to the up-to-date astrometry data of the 17 best-resolved S-stars orbiting Sgr A*, achieving to explain the dynamics of the S-stars with similar (and some cases better) accuracy compared to a central BH model (see Fig. 6), strengthen the alternative nature of Sgr A*.

Another key consequence of such a thermodynamic approach for DM halo formation is that, on inner halo scales, the $\mathcal{O}(10)$ keV fermionic density profiles develops an extended plateau (similar to Burkert profiles, with or without the innermost quantum core depending on the degree of central degeneracy), thus not suffering from the core-cusp problem associated to the standard ΛCDM cosmology.[19] Interestingly, while in N-body simulations (within the CDM paradigm) it is possible to obtain 'cored' profiles when including for DM self-interactions (see e.g. Ref. 20), here the flatness on inner halo scales are obtained due to the fulfillment of an entropy maximization. The later can be achieved thanks to the long range particle-particle interactions involved in the diffusion-current term of equation (2) (see also footnote a) as shown here, or complementarily it could be enhanced by (short-range) DM self-interactions as well.

3.2. *Gravitational core-collapse of DM and SMBH formation*

Another striking consequence of the present DM halo formation approach is that it predicts a novel SMBH formation mechanism from DM. Namely, a galactic quantum core made of DM (like in solution (3) of Fig. 2) might become so densely packed that, above a threshold critical mass, the quantum pressure can not support it any longer against its own weight, leading to its gravitational collapse and forming a SMBH. Such threshold solution correspond to the last-stable configuration depicted

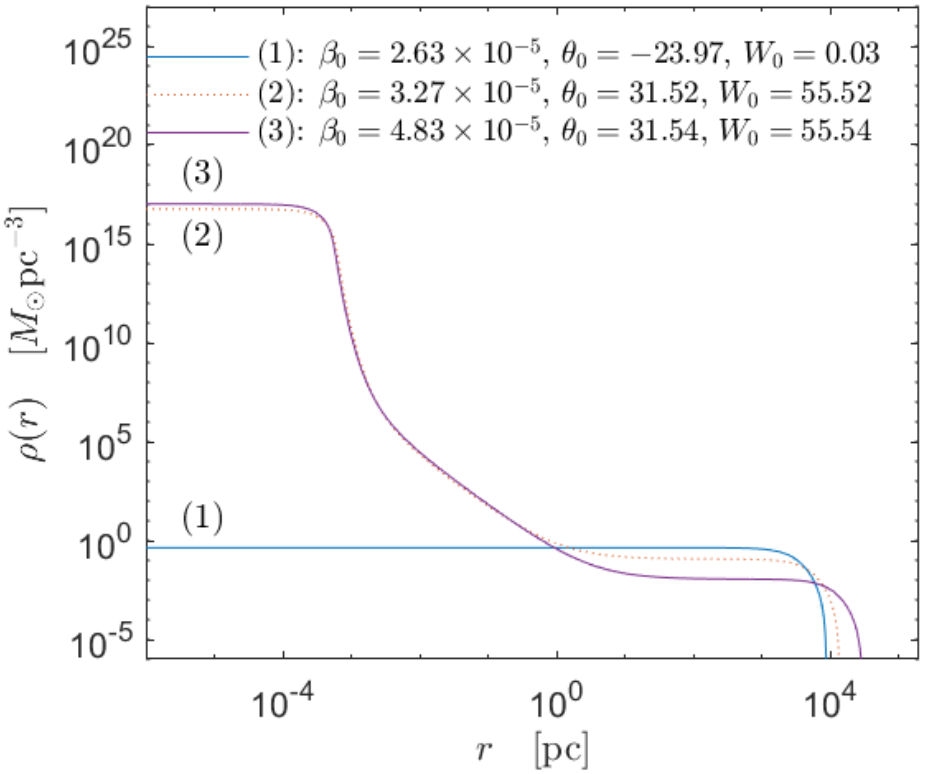


Fig. 2. Density profiles for $mc^2 = 48$ keV corresponding with three different solutions along the caloric curve in Fig. 1 at a fixed energy. Only the profiles (1) (resembling a King distribution) and the *core-halo* one (3) are stable, while profile (2) is thermodynamically unstable. Solutions like (3) are successfully applied to explain the Milky Way rotatoin curve (see Fig. 3). They are stable, extremely long-lived and fulfill with different observational probes. Figure taken from Ref. 3.

by the point (c) in the caloric curve of Fig. 1, below which there is no possible accessible state (see Ref. 3 for details). Moreover, such a collapse occurs *prior* to the turning point in the (traditional) *total mass* vs. *central density* diagram (see Fig. 7). This result, occurring here for self-gravitating systems of neutral fermions at finite T_∞ in GR, agrees with similar findings within rotating perfect-fluid stars,[25] and generalizes the traditional result applied to compact objects (such as neutron stars) based on full degeneracy ($T \to 0$).

This kind of DM core-collapse can only occur when the total particle number exceeds the so-called Oppenheimer-Volkoff limit (i.e. $\hat{N} > \hat{N}_{\mathrm{OV}} \approx 0.4$, as fulfilled in the case here exemplified of $\hat{N} = 76.25$). However, as explained in Sec. 3.1, it can also exist a stable branch of *core-halo* solutions for such $\hat{N} > \hat{N}_{\mathrm{OV}}$ as exemplified through the solution (3) in Fig. 1, before the last stable one (c). This fact is characteristic of self-gravitating solutions of fermions at finite T (as in the RAR model), which

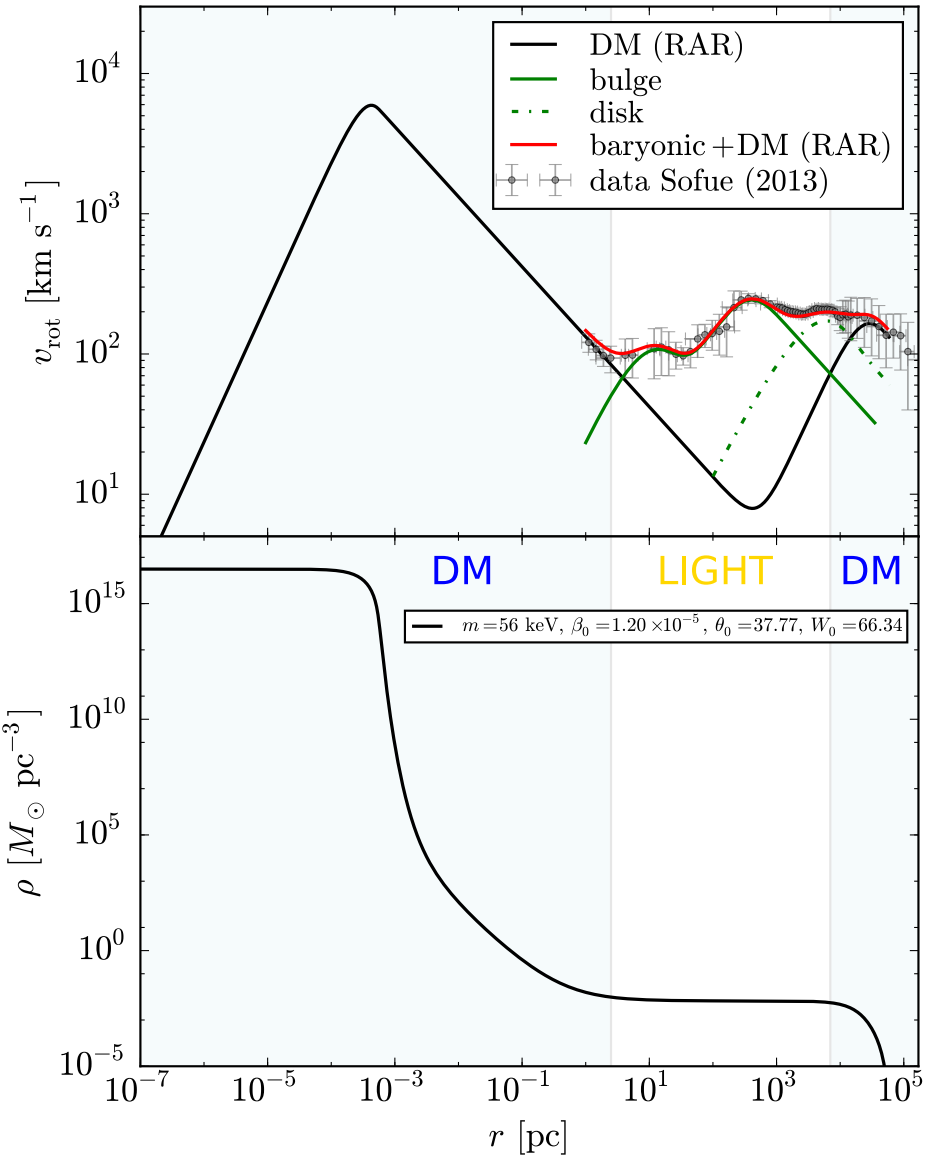


Fig. 3. Milky Way rotation curve and DM density profile within the core-halo DM RAR profiles for $mc^2 = 56$ keV, with a DM quantum core of mass of $M_c = M(r_c) = 3.5 \times 10^6 M_\odot$. Top: DM (black) and baryonic (bulge + disk) contribution to the rotation curve $v_{\rm rot}$ (total in red). Bottom: corresponding DM density profile. Figure taken from Ref. 15.

differ from the traditional $T \to 0$ limit where the system must necesarily collapse when $\hat{N} > \hat{N}_{\rm OV}$ (see Ref. 3 for further details).

In cases when $\hat{N} > \hat{N}_{\rm OV}$ as the one shown here, and for $mc^2 \approx 50$ keV, it is possible to form a SMBH of $M_c^{\rm cr} \approx 2 \times 10^8 M_\odot$ at the center of a realistic DM halo:

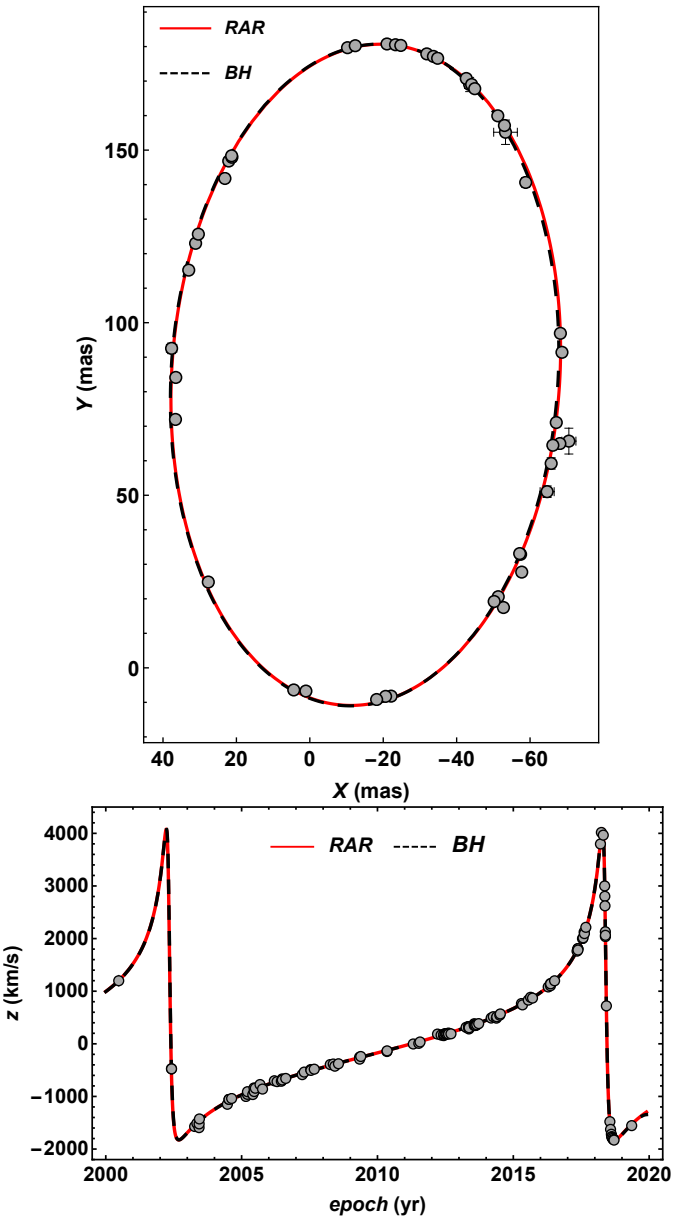


Fig. 4. Theoretical and observed orbit and radial velocity (z) of S2 around Sgr A* (wiht data taken from Ref. 21). Top panel: shows the projected orbit on the sky, X (Right Ascension) vs. Y (Declination). Bottom panel: shows the line–of–sight radial velocity as a function of time. The theoretical models are calculated by solving the equations of motion of a test particle in the gravitational field of: (i) a Schwarzschild BH of $4.075 \times 10^6\ M_\odot$ (*black dashed curve*), and (ii) the fermionic DM distribution obtained from the core-halo RAR model with $mc^2 = 56$ keV fermions (*red curves*). Figure adapted from Ref. 15.

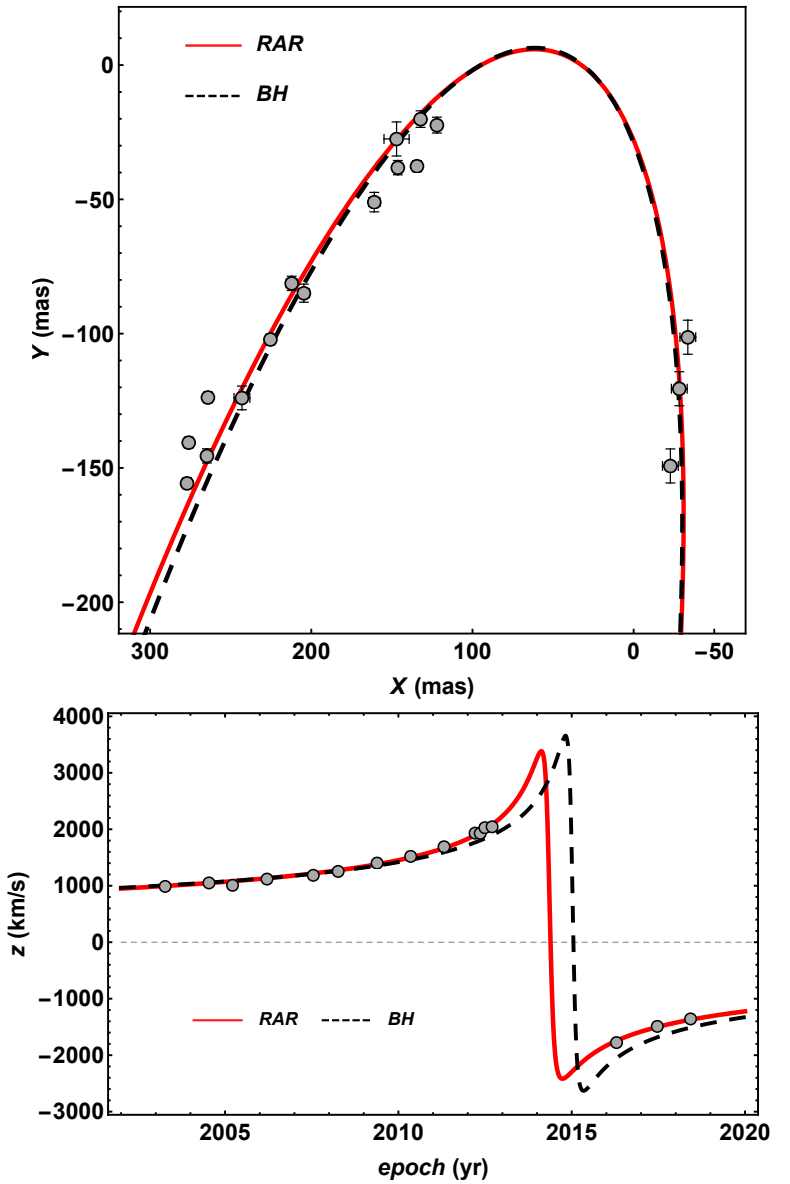


Fig. 5. Theoretical and observed orbit and line-of-sight radial velocity of G2 around Sgr A*. *Top panel*: shows the projected orbit on the sky, X (Right Ascension) vs. Y (Declination). *Bottom panel*: shows the redshift function z as a function of time. The theoretical models are calculated by solving the equations of motion of a test particle in the gravitational field of: (1) a Schwarzschild BH of $4.075 \times 10^6\,\mathrm{M}_\odot$ (*black dashed curve*), and (2) the DM distribution obtained from the extended RAR model for 56 keV fermions (*red curves*). The mass of the quantum core in the RAR model is $3.5 \times 10^6\,\mathrm{M}_\odot$. The observational data are taken from Refs. 17,18,22. Figure adapted from Ref. 15.

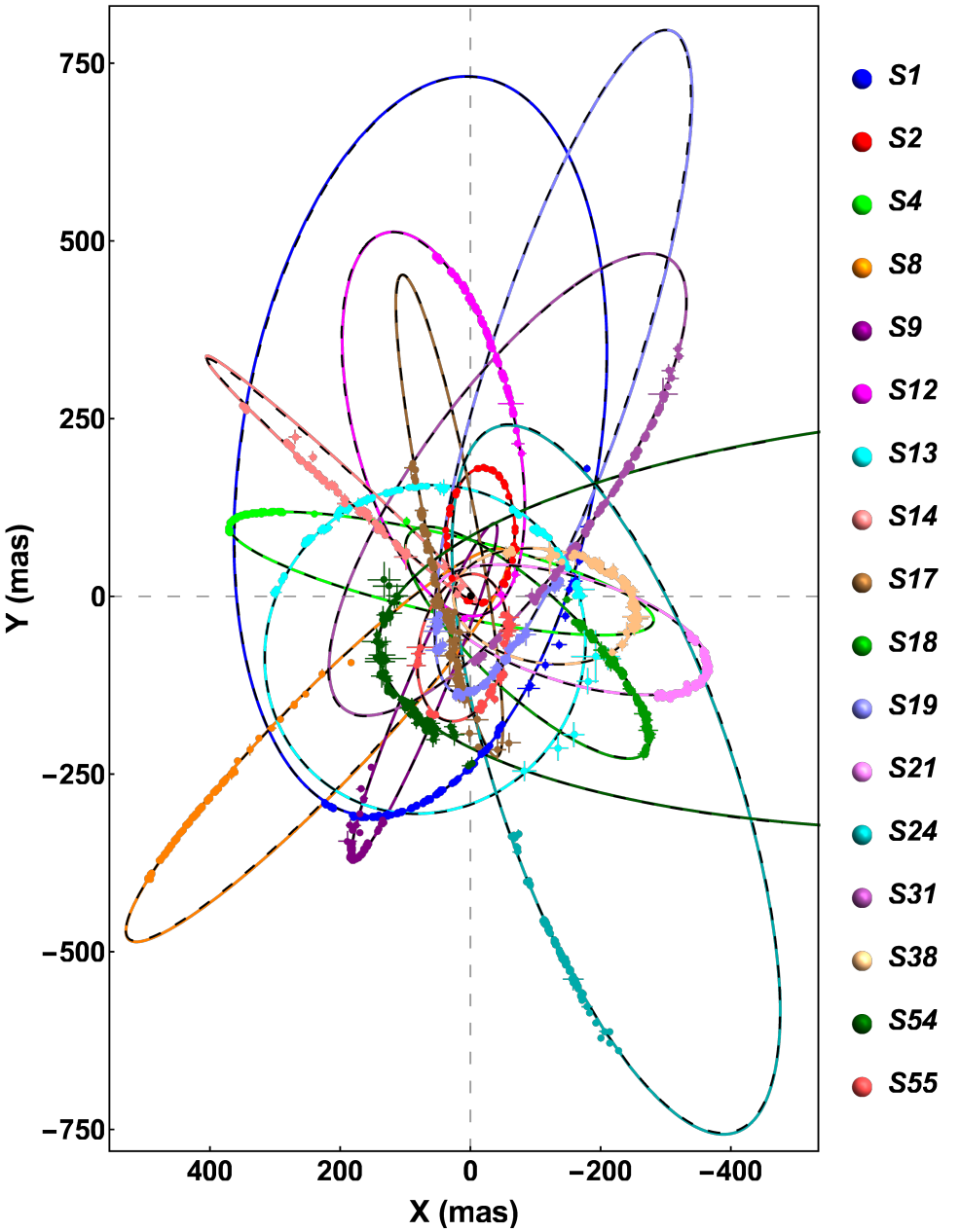


Fig. 6. Best-fit orbits for the 17 best-resolved S-star orbiting Sgr A*. It shows the projected orbit on the sky, X vs. Y, where X is right ascension and Y is declination. The *black dashed curves* correspond to the BH model and the *colored curves* to the RAR model for $m \approx 56$ keV fermions. The astrometric data was taken from Refs. 21, 23, 24. Figure taken from Ref. 16.

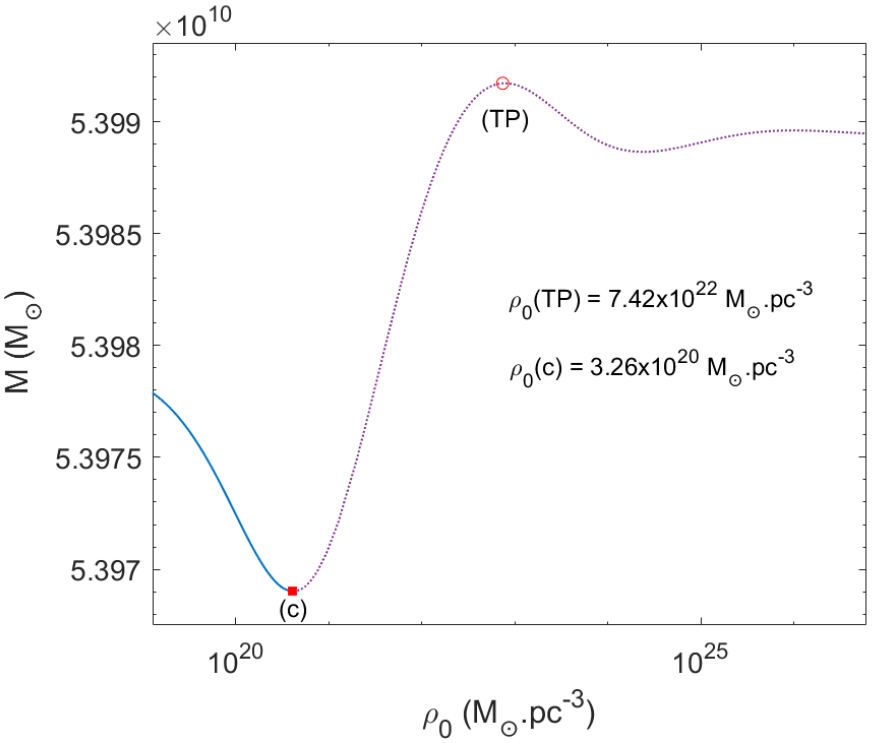


Fig. 7. Series of equilibrium states with $N > N_{\rm OV}$ are shown along a ρ_0 vs. M curve, in correspondence with the caloric curve of Fig. 1. The last stable configuration at the onset of the core-collapse occurs at the minimum of this curve and *prior* to the turning point instability (corresponding with point (c) in Fig. 1). Such a critical solution has a critical core mass $M_c^{\rm cr} \approx 2 \times 10^8 M_\odot$, and thus a forming a SMBH from DM core-collapse. Figure taken from Ref. 3.

that is, the last stable configuration at the onset of the core-collapse (point (c) in Fig. 1) correspond to a critical core mass $M_c^{\rm cr} = 2.17 \times 10^8 M_\odot$, and thus a forming a SMBH from DM core-collapse (see Ref. 3 for more details). This is an interesting number, since the majority of the SMBHs are comprised within $\sim 10^8 M_\odot$,[26] in some way separating active from non-active galaxies. Moreover, the relevance of such a DM core-collapse scenario, is that it can occur within the high $z \sim 10$ Universe when such halos start to form, without the need of prior star formation or any BH seed mechanisms involving (likely unrealistic) super-Eddington accretion rates. This may provide the long-sought solution to the problem of SMBH formation in the early Universe, which deserves further detailed investigations.

Such a critical value triggering core-collapse can be written in a more familiar way in terms of the Planck mass m_{pl} and the fermion mass m as $N_{\rm OV} \approx 0.398\, m_{pl}^3/m^3$, corresponding to a critical (total) mass $M_{\rm OV} \approx 0.384\, m_{pl}^3/m^2$ which for $m \approx 50$ keV yields $M_{\rm OV} \sim 10^8 M_\odot$. Therefore, equilibrium configurations of fermions at finite T_∞ with $M_{\rm vir} < 10^8 M_\odot$ can not undergo core-collapse towards a

SMBH within the present scenario. Indeed, the caloric curves in such cases never develop a last-stable solution (like (c) in Fig. 1), nor the second spiral of relativistic origin as shown in Ref. 3.

The above numbers imply another intriguing consequence for this new theory: the critical mass for collapse into a massive BH can not be reached for smaller DM halos, for example those surrounding typical dwarf galaxies with $M_{\rm vir} \lesssim 10^8 M_\odot$. Instead, this $\hat{N} < \hat{N}_{\rm OV}$ scenario would leave smaller dwarf galaxies with a central DM nucleus with a mass typically of $M_c \sim 10^5 M_\odot$ (or less) as shown in Ref. 11, thus offering a natural explanation for the so called 'intermediate-mass BHs'. Such a DM quantum-core can still mimic the gravitational signatures of a central BH, whilst the dark matter outer halo could also explain the observed galaxy rotation curves.[11]

4. Conclusions

We have shown that mechanisms of DM halo formation based on entropy maximization in which the quantum (fermionic) nature of the particles is accounted for, may provide answers to crucial open problems in cosmology. In particular they allow: to constraint the mass of the DM particles, to provide a novel mechanisms of supermassive BH formation from DM core-collapse, to understand the nature of the intermediate mass black holes in small halos, and to provide a solution to the *core-cusp* problem.

The approach applied here (and detailed in[3]) for DM halo formation in terms of self-gravitating fermions is self-consistent: the nature and mass of the DM particle involved in the linear matter power spectrum calculations (as obtained in Ref. 3 within a CLASS code for WDM cosmology of $\mathcal{O}(10)$ keV), are the same building blocks at the basis of virialized DM configurations with its inherent effects in the *core-halo* profiles. Importantly, such a fermion mass would produce the same ΛCDM power-spectrum on scales down to $\mathcal{O}$(Mpc), hence providing the expected large-scale structure (see Appendix B in Ref. 3). Moreover, for such typical fermion masses the number of sub-halos in the corresponding WDM cosmology is not in tension with the number of Milky Way satellites,[27] while being consistent with Lyman-α forest observations.[28] Another relevant consequence is that the maximum entropy approach for structure formation applying here for $\mathcal{O}(10)$ keV particles, can naturally avoid the so called 'catch-22' tension arising for fermionic halos within (N-body simulation-based) WDM cosmologies.[29] That is, the resulting fermionic profiles within our approach allow for 'cored' density profiles in dwarfs with right sizes and masses, and at the same time in agreement with Lyman-α bounds.[11]

To conclude, we believe the results shown here open new insights in the formation and stability of DM halos. Moreover, the quantum degenerate core at the center of the stable DM configurations, and its eventual core-collapse, may play crucial roles in helping to understand the formation of SMBHs in the early Universe, or in mimicking its effects without the need of a singularity at all. The astrophysical

consequences of the analysis here shown for $\mathcal{O}(10)$ keV — together with the recent results in Refs. 3, 10, 11, 15 — strongly suggest that such DM *core-halo* morphology are a very plausible scenario within the late stages of non-linear structure formation.

References

1. H. Mo, F. Van den Bosch and S. White, *Galaxy formation and evolution* (Cambridge University Press, 2010).
2. H.-Y. Schive, M.-H. Liao, T.-P. Woo, S.-K. Wong, T. Chiueh, T. Broadhurst and W. Y. P. Hwang, *Phys. Rev. Lett.* **113** (December 2014) 261302, arXiv:1407.7762 [astro-ph.GA].
3. C. R. Argüelles, M. I. Díaz, A. Krut and R. Yunis, *Mon. Not. R. Astron. Soc.* **502** (April 2021) 4227, arXiv:2012.11709 [astro-ph.GA].
4. R. RUFFINI and S. BONAZZOLA, *Phys. Rev.* **187** (Nov 1969) 1767.
5. D. Lynden-Bell, *Mon. Not. R. Astron. Soc.* **136** (January 1967) 101.
6. A. Kull, R. A. Treumann and H. Boehringer, *Astrophys. J.l* **466** (July 1996) L1, arXiv:astro-ph/9606057 [astro-ph].
7. P.-H. Chavanis, *Mon. Not. R. Astron. Soc.* **300** (November 1998) 981.
8. P.-H. Chavanis, *Physica A Statistical Mechanics and its Applications* **332** (February 2004) 89, arXiv:cond-mat/0304073 [cond-mat.stat-mech].
9. P.-H. Chavanis, M. Lemou and F. Méhats, *Phys. Rev. D* **92** (December 2015) 123527, arXiv:1409.7840 [astro-ph.CO].
10. C. R. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, *Physics of the Dark Universe* **21** (September 2018) 82, arXiv:1810.00405 [astro-ph.GA].
11. C. R. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, *Physics of the Dark Universe* **24** (March 2019) 100278.
12. J. Katz, *Mon. Not. R. Astron. Soc.* **189** (December 1979) 817.
13. P. H. Chavanis and J. Sommeria, *Mon. Not. R. Astron. Soc.* **296** (May 1998) 569.
14. R. Ruffini, C. R. Argüelles and J. A. Rueda, *Mon. Not. R. Astron. Soc.* **451** (July 2015) 622, arXiv:1409.7365 [astro-ph.GA].
15. E. A. Becerra-Vergara, C. R. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, *Astron. & Astrophys.* **641** (September 2020) A34, arXiv:2007.11478 [astro-ph.GA].
16. E. A. Becerra-Vergara, C. R. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, *Mon. Not. R. Astron. Soc.* **505** (July 2021) L64, arXiv:2105.06301 [astro-ph.GA].
17. P. M. Plewa, S. Gillessen, O. Pfuhl, F. Eisenhauer, R. Genzel, A. Burkert, J. Dexter, M. Habibi, E. George, T. Ott, I. Waisberg and S. von Fellenberg, *Astrophys. J.* **840** (May 2017) 50, arXiv:1704.05351.
18. S. Gillessen, P. M. Plewa, F. Widmann, S. von Fellenberg, M. Schartmann, M. Habibi, A. Jimenez Rosales, M. Bauböck, J. Dexter, F. Gao, I. Waisberg, F. Eisenhauer, O. Pfuhl, T. Ott, A. Burkert, P. T. de Zeeuw and R. Genzel, *Astrophys. J.* **871** (January 2019) 126.
19. J. S. Bullock and M. Boylan-Kolchin, *Annu. Rev. Astron. Astrophys.* **55** (August 2017) 343, arXiv:1707.04256 [astro-ph.CO].
20. R. Davé, D. N. Spergel, P. J. Steinhardt and B. D. Wandelt, *Astrophys. J.* **547** (February 2001) 574, arXiv:astro-ph/0006218 [astro-ph].
21. T. Do, A. Hees, A. Ghez, G. D. Martinez, D. S. Chu, S. Jia, S. Sakai, J. R. Lu, A. K. Gautam, K. K. O'Neil, E. E. Becklin, M. R. Morris, K. Matthews, S. Nishiyama, R. Campbell, S. Chappell, Z. Chen, A. Ciurlo, A. Dehghanfar, E. Gallego-Cano, W. E. Kerzendorf, J. E. Lyke, S. Naoz, H. Saida, R. Schödel, M. Takahashi, Y. Takamori,

G. Witzel and P. Wizinowich, *Science* **365** (August 2019) 664, arXiv:1907.10731 [astro-ph.GA].

22. K. Phifer, T. Do, L. Meyer, A. M. Ghez, G. Witzel, S. Yelda, A. Boehle, J. R. Lu, M. R. Morris, E. E. Becklin and K. Matthews, *Astrophys. J.l* **773** (August 2013) L13, arXiv:1304.5280 [astro-ph.GA].
23. S. Gillessen, P. M. Plewa, F. Eisenhauer, R. Sari, I. Waisberg, M. Habibi, O. Pfuhl, E. George, J. Dexter, S. von Fellenberg, T. Ott and R. Genzel, *Astrophys. J.* **837** (March 2017) 30, arXiv:1611.09144 [astro-ph.GA].
24. S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexander, R. Genzel, F. Martins and T. Ott, *Astrophys. J.* **692** (February 2009) 1075, arXiv:0810.4674 [astro-ph].
25. K. Takami, L. Rezzolla and S. Yoshida, *Mon. Not. R. Astron. Soc.* **416** (September 2011) L1, arXiv:1105.3069 [gr-qc].
26. K. Gültekin, D. O. Richstone, K. Gebhardt, T. R. Lauer, S. Tremaine, M. C. Aller, R. Bender, A. Dressler, S. M. Faber, A. V. Filippenko, R. Green, L. C. Ho, J. Kormendy, J. Magorrian, J. Pinkney and C. Siopis, *Astrophys. J.* **698** (June 2009) 198, arXiv:0903.4897 [astro-ph.GA].
27. E. J. Tollerud, J. S. Bullock, L. E. Strigari and B. Willman, *Astrophys. J.* **688** (November 2008) 277, arXiv:0806.4381 [astro-ph].
28. V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, S. Cristiani, G. D. Becker, V. D'Odorico, G. Cupani, T.-S. Kim, T. A. M. Berg, S. López, S. Ellison, L. Christensen, K. D. Denney and G. Worseck, *Phys. Rev. D* **96** (July 2017) 023522, arXiv:1702.01764 [astro-ph.CO].
29. A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider and B. Moore, *Mon. Not. R. Astron. Soc.* **424** (August 2012) 1105, arXiv:1202.1282 [astro-ph.CO].

First results of LHAASO

Ruo-Yu Liu* for the LHAASO Collaboration
School of Astronomy and Space Science, Nanjing University, Nanjing, 210023, China
**E-mail: ryliu@nju.edu.cn*

The Large High Altitude Air Shower Observatory (LHAASO) has recently published the first results, including the discovery of 12 ultrahigh-energy gamma-ray sources (with emission above 100 TeV) above 7σ confidence level and a detailed analysis of Crab Nebula. This contribution gives a brief introduction to the LHAASO experiment and its recent discoveries.

Keywords: Gamma rays; Cosmic rays; Pulsar Wind Nebula

1. Introduction

The Large High Altitude Air Shower Observatory (LHAASO) is one of the key national science and technology infrastructure facilities of China, located in Mt. Haizi (29°21'27.6" N, 100°08'19.6" E, 4410 m above the sea level), Sichuan Province of China. It is a new generation instrument aiming for a deep investigation of the so-called "non-themral Universe'' at extreme energies. The observatory consists of three detectors: the Water Cherenkov Detector Array (WCDA) for gamma rays in 0.5 – 10 TeV, the Kilometer Square Array (KM2A) for gamma rays in 10 TeV – 1 PeV, and the Wide Field Chenrenkov Telescope Array (WFCTA) for cosmic rays from 1 TeV – 1 EeV.

KM2A comprises an array of 5195 electron detectors (ED) made by scintillation tiles, and an array of 1188 underground muon detectors (MD) each of which is a pure water Cherenkov detector enclosed within a cylindrical concrete tank. Both arrays cover an area of 1 km^2 with an additional 0.3 km^2 skirt array used to identify showers falling outside the central region. Once charged particles in an air shower enter an ED, they lose energy and excite the scintillation medium to produce a large amount of scintillation photons, which will be recorded by photomultiplier tubes and used to determine the number of electromagnetic particles N_e in a shower. Each MD is enveloped by a steel lid underneath soil. The thickness of the overburden soil is 2.5 m to absorb the secondary electrons/positrons and gamma rays in showers. Thus particles that can reach the water inside and produce Cherenkov signals are almost exclusively muons, although a few very energetic electrons/positrons or gamma rays in the central part of showers may have a chance to penetrate through the screening soil layer. To reduce the possible pollution from the high energy electromagnetic particles near the shower core, only MDs with a distance farther than 15 m from the core are used to count the number of muons N_μ in a shower.

Each air shower event is composed of many ED and MD hits with timing and charge information. In the reconstruction of primary particle event, only the ED hits are used for direction, core location, and energy reconstruction. Both ED hits and MD hits are used for composition discrimination. Gamma-ray-induced showers are muon-poor while cosmic-ray-induced showers are muon-rich. Therefore, the ratio between the measured muons and electrons can be used to discriminate primary gamma rays from cosmic protons/nuclei, which is defined as

$$R = \log\left(\frac{N_\mu + 0.0001}{N_e}\right). \tag{1}$$

0.0001 in the numerator is used to show the case with zero muon. Generally, a simple cut on the parameter R is used in the discrimination, where the detailed value of R depends on the energy of the reconstructed event and can be adjusted to maximize the detection efficiency.[1] This process is extremely important for the gamma ray detection because cosmic rays constitute the majority of the recorded events by KM2A and become a background for the gamma-ray observations. The rejection power of KM2A can be better than 10^4 above 100 TeV (i.e., reduce the background cosmic-ray flux by a factor higher than 10^4). The large collective area and the high rejection power make LHAASO the most sensitive detector for ultrahigh-energy (UHE) gamma rays with energy above 100 TeV.

Detection of UHE gamma-ray emission is crucial to understand the origin of Galactic cosmic rays. The measured cosmic-ray spectrum roughly maintains a power-law behaviour up to about 1 PeV. After that it shows a spectral break feature known as the "knee". This indicates the existence of astrophysical objects in our Galaxy capable of accelerating protons up to at least PeV energies (which is also known as proton "PeVatrons"). When these high-energy protons interact with matter inside or close to the accelerators, neutral pions are generated, which further decay into gamma-ray photons. These gamma-ray emission reveal the location of the cosmic-ray accelerators. Since each produced photon takes about 10% of the energy of the parent proton, we need to search for UHE gamma-ray sources in order to find out PeV cosmic-ray sources. In the rest of this artcile, we will introduce the first results of LHAASO, which are mainly based on[2] and.[3]

2. Discovery of 12 Ultrahigh-energy Gamma-ray Sources

The first results of LHAASO are mainly discovered by the half KM2A during the operation of 11 months in the year 2020. By reconstructing showers with zenith angle $< 50°$, we have detected 12 sources with emission above 100 TeV at the statistical significance exceeding 7σ level and the basic information of the sources are listed in Table. 1. All of these sources are located in the Galactic plane and hence very likely Galactic sources. Most of them show extended spatial structures with angular extensions up to 1°. Generally, there are multiple astrophysical objects proposed as potential PeVatron candidates in spatial association with these sources.

Table 1. UHE γ-ray sources. The celestial coordinates (R.A.,dec), the statistical significance of detection above 100 TeV (using a point-like template for the Crab Nebula and LHAASO J2108+5157 and 0.3° extension templates for other sources) are shown together with the corresponding differential photon fluxes at 100 TeV in unit of the Crab flux (1 CU=6.1×10^{-17} photons TeV^{-1} cm^{-2} s^{-1}) and the detected highest photon energies.

Source Name	R.A. (°)	dec (°)	Significance (σ) above 100 TeV	E_{Max} (PeV)	Flux (± error) (CU) at 100 TeV
LHAASO J0534+2202	83.55	22.05	17.8	0.88 ±0.11	1.00(0.14)
LHAASO J1825-1326	276.45	-13.45	16.4	0.42 ±0.16	3.57(0.52)
LHAASO J1839-0545	279.95	-5.75	7.7	0.21 ±0.05	0.70(0.18)
LHAASO J1843-0338	280.75	-3.65	8.5	$0.26\ ^{+0.16}_{-0.10}$	0.73(0.17)
LHAASO J1849-0003	282.35	-0.05	10.4	0.35 ±0.07	0.74(0.15)
LHAASO J1908+0621	287.05	6.35	17.2	0.44 ±0.05	1.36(0.18)
LHAASO J1929+1745	292.25	17.75	7.4	$0.71\ ^{+0.16}_{-0.07}$	0.38(0.09)
LHAASO J1956+2845	299.05	28.75	7.4	0.42 ±0.03	0.41(0.09)
LHAASO J2018+3651	304.75	36.85	10.4	0.27 ±0.02	0.50(0.10)
LHAASO J2032+4102	308.05	41.05	10.5	1.42 ±0.13	0.54(0.10)
LHAASO J2108+5157	317.15	51.95	8.3	0.43 ±0.05	0.38(0.09)
LHAASO J2226+6057	336.75	60.95	13.6	0.57 ±0.19	1.05(0.16)

Among these astrophysical objects, most of them are supernova remnants, pulsar wind nebulae and young massive star clusters.

The top-three brightest sources at 100 TeV are LHAASO J1825-1326, LHAASO J1908+0621 and LHAASO J2226+6057. The regions of the three sources have been extensively studied in the previous literature (e.g.,[4–27]). All of them are associated with both energetic pulsars and supernova remnants. Dense molecular clouds also appear in their region. LHAASO's observations extend their spectra up to several hundred TeV energies, demonstrating there exist efficient particle accelerators acting as PeV particle factories. The spectra present gradual softening towards high energies. The log-parabola function can describe the spectra better than power-law function, as can be seen in Fig. 1. However, we cannot tell whether these multi-hundred-TeV gamma rays arise from the leptonic process such as the inverse Compton (IC) scattering of electrons, or hadronic process such as the proton-proton (pp) collisions, based on the present data. Deep observations of multiwavelength instruments are needed to elucidate their natures.

We'd like to highlight that the energy of the most energetic photons detected from these sources is 1.42 ± 0.13 PeV in association with the Cygnus Cocoon, which is a energetic gamma-ray emitter at GeV – TeV band. It has been proposed that Cygnus Cocoon is powered by the O/B star association, Cygnus OB2, possibly via collisions of the stellar winds.[28–30] LHAASO's discovery may be considered as the evidence of Cygnus OB2 as a proton PeVatron.

It is also worth mentioning that among the 12 detected sources, there is a new source, LHAASO J2108+5157, without association with any of the previously detected very-high-energy sources. The position of its gamma-ray emission is correlated with a giant molecular cloud, which makes the hadronic origin of the

gamma-ray emission possible. Detection of this source also implies that there could be much more new UHE gamma-ray sources hiding in our Galaxy than we previously expected. Future observations of LHAASO with longer exposure can hopefully discovered more UHE gamma-ray sources or PeVatron candidates. We stress that the 12 sources reported here are UHE gamma-ray sources with significant emission above 100 TeV. There are much more sources with significantly detected by KM2A above 20 TeV, such as LHAASO J0621+3755[31] and LHAASO J0341+5258.[32]

Note that the only point-like source discovered in this survey, i.e., LHAASO J0534+2202, which is also the only identified source, is the Crab Nebula. The highest-energy photon from Crab Nebula reaches 0.88 ± 0.11 PeV, demonstrating it as an electron PeVatron. We will show more detailed analysis on Crab Nebula in the next section.

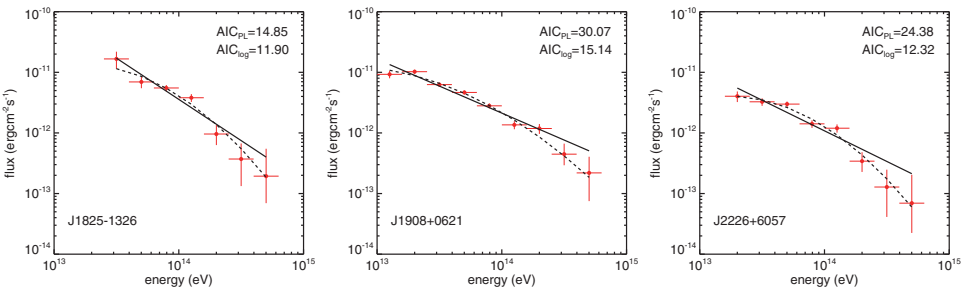


Fig. 1. The spectral energy distributions and significance maps of LHAASO J1825-1326 (panel a), LHAASO J1908+0621 (panel b), and LHAASO J2226+6057 (panel c) The spectral fits with a log-parabola function (solid lines) in the form of $(E/10\,\mathrm{TeV})^{[-a-b\log(E/10\mathrm{TeV})]}$ are compared with the power-law fits, $E^{-\Gamma}$: (a) $a = 0.92, b = 1.19$ vs $\Gamma = 3.36$; (b) $a = 2.27, b = 0.46$ vs $\Gamma = 2.89$, and (c) $a = 1.56, b = 0.88$ vs $\Gamma = 3.01$. The comparison of the two models with the Akaike Information Criterion[33] (AIC) gives: $\mathrm{AIC_{LOG}} = 11.6$ and $\mathrm{AIC_{PL}} = 14.8$ for LHAASO J1825-1326; $\mathrm{AIC_{LOG}} = 15.1$ and $\mathrm{AIC_{PL}} = 30.1$ for LHAASO J1908+0621; $\mathrm{AIC_{LOG}} = 12.3$ and $\mathrm{AIC_{PL}} = 24.4$ for LHAASO J2226+6057.

3. PeV Gamma rays from the Crab Nebula

The Crab Nebula was born from a supernova explosion occurring at Jul 4, 1054 A.D. It is the brightest pulsar wind nebula known so far, powered by the ultrarelativistic electron-positron wind launched by the Crab pulsar. It is a very important source in gamma-ray astronomy and considered as the standard candle. Previous observations revealed gamma-ray emission beyond 100 TeV.[34,35] LHAASO measured the spectrum of Crab Nebula in TeV – PeV range. The spectrum from 0.5 TeV to 13 TeV is measured using the WCDA pond, from September 2019 to October 2020. The KM2A data consists of that collected at the half array stage, which is the same as that used for the discovery of the twelve sources, and the data collected by three quarters of the array in about 3 months. The 0.88 PeV photon, as mentioned in the previous section, was detected on 11 January 2020. The shower of the event has

also triggered WCDA. Coincidently, the event occurred after local midnight and was recorded by WFCTA. The derived energy is $0.92^{+0.28}_{-0.20}$ PeV which is well consistent with that derived from KM2A's measurement, and demonstrates that LHAASO can cross check the results using multiple detection techniques, ensuring reliable and accurate results. In the stage of the 3/4 KM2A, the highest-energy photon detected from the Crab Nebula has an energy of 1.12 PeV, arriving at 4 January 2021. The spectrum between 0.5 TeV to 1.1 PeV can be described by a log-parabola function, where the local slope softens from 2.5 around 1 TeV to 3.7 around 1 PeV.

The gradual softening of the spectrum is consistent with the prediction of IC radiation of electrons where the Klein-Nishina effect operates at the highest energies and softens the spectrum. For gamma-ray emission above 100 TeV, cosmic microwave background is the dominant target radiation field for the IC radiation. The relation between the average energy of upscattered IC photons (from 2.7K CMB) and the energy of the parent electrons can be given by

$$E_{\rm e} \simeq 2.15(E_\gamma/1\ {\rm PeV})^{0.77}\ {\rm PeV}\ . \tag{2}$$

To account for 1.1 PeV photon via the IC radiation of electrons, the energy of electron needs to be about 2.3 PeV. The acceleration rate of electrons can be given by $\dot{E}_e = \eta eBc$,[36] where η is the ratio of the projection of the electric field to the mean magnetic field. Its value cannot exceed unity under the classical electrodynamics and ideal magnetohydrodynamics, and can be also regarded as the acceleration efficiency. On the other hand, electrons will lose energy via the synchrotron radiation when it propagates in the magnetic field. The maximum energy of electrons is then is determined by the balance between the acceleration and the energy loss rates, i.e.,

$$E_{e,\rm max} = 5.8\eta^{-1/2}\,(B/100\mu{\rm G})^{-1/2}\ {\rm PeV} \tag{3}$$

On the other hand, in the framework of the standard one-zone model, fitting to the multiwavelength data from X-ray to PeV gamma-ray energies results in a magnetic field of $B = 112^{+15}_{-13}\,\mu$G, as shown in Fig. 2. According to the relation between E_γ and $E_{\rm e}$ given by Eq.(2), we find the electron acceleration in the Crab Nebula must proceed at an extreme efficiency

$$\eta = 0.16(B/112\mu{\rm G})(E_\gamma/1.1\ {\rm PeV})^{1.54}\ {\rm PeV}\ . \tag{4}$$

in the framework of the standard one-zone model. For comparison, at the diffusive shock acceleration in young supernova remnants,[37] η is smaller by at least 3 orders of magnitude.

Note that although the standard one-zone model can generally give an acceptable fit to the overall spectrum of the Crab Nebula, the presence of the intense flare at MeV/GeV band (also known as "Crab flare",[40,41]) implies fast acceleration and synchrotron cooling of sub-PeV/PeV electrons in compact (with size < 0.01 pc) highly magnetized (with $B \gtrsim 1$ mG) regions inside the nebula, and is thus inconsistent with the standard one-zone picture. Actually, in the spectral fitting of the KM2A data in the standard one-zone scenario, we find that the one-zone model does not agree very well with the measured flux in the energy range of $60-500$ TeV,

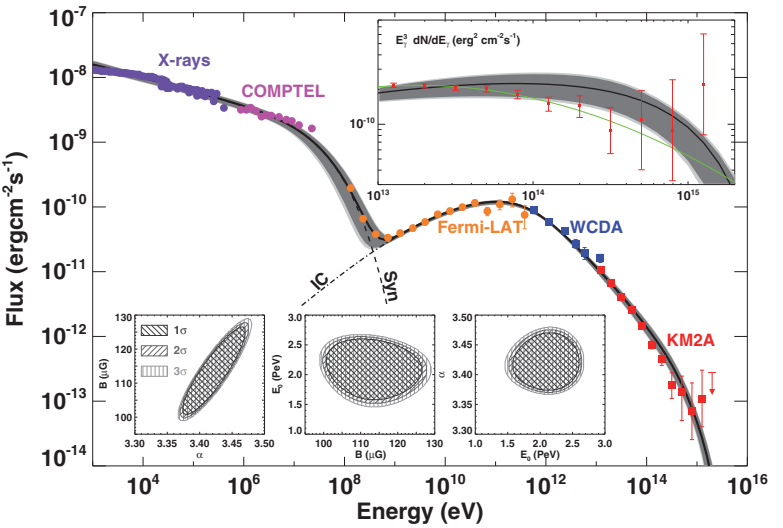


Fig. 2. **The Spectral Energy Distribution of the Crab Nebula.** The black curves represent the fluxes of the Synchrotron and IC components of radiation of an electron population calculated within the one-zone model. The electron spectrum above 1 TeV is assumed to be a power-law function terminated by an super-exponential cutoff, $E^{-\alpha}\exp[-(E/E_0)^2]$. The model's best fit parameters are: $\alpha = 3.42 \pm 0.05$, $E_0 = 2.15^{+0.55}_{-0.65}$ PeV, $B = 112^{+15}_{-13}\,\mu$G. The total energy in electrons above 1 TeV is $W_e = 7.7 \times 10^{47}$ erg. A break in the electron spectrum at $E_b = 0.76$ TeV is assumed to provide a consistency with the GeV γ-ray and low-frequency synchrotron data. The dark-grey and light-grey shaded regions show the 1σ and 3σ error bands, respectively. The insets in the bottom exhibit the 2-dimensional projected parameter spaces of the free parameters α, B and E_0. The green curve presents the log-parabola spectral fitting. The purple and the magenta circles show the X-ray and the MeV emission of the Crab Nebula.[38] The orange circles represent the Crab observations by Fermi-LAT in the non-flare state.[39] The blue and red squares represent WCDA and KM2A measurements reported in.[3] The inserted figure in the top right corner zooms the fluxes above 10 TeV in the presentation of $E^3\mathrm{d}N/\mathrm{d}E$.

nor it reproduces the spectral hardening above 500 TeV although the latter is not statistically sound at the present time. The fitting can be ameliorated by introducing a second spectral component, either produced by an extra population of PeV electrons or protons, which are accelerated presumably in different sites by different mechanisms. Based on the present observation, we expect that the full LHAASO array can detect $1-2$ PeV photons from the Crab Nebula per year. Hopefully, we can accumulate sufficient statistics to clarify whether such a spectral hardening truly exists in 2 years.

4. Conclusion

LHAASO is a dual-purpose complex of particle detectors for the measurements of gamma rays in sub-TeV to PeV energy range and cosmic rays in TeV to EeV energy range. It is designed to be operating continuously and steadily for many years in order to accumulate a large data sets. The first results of LHAASO introduced here are probably only the tip of the iceberg, but they demonstrate LHAASO's potential

to provide a deep insight into the mystery of cosmic-ray origin and related physics, especially when combined with the multiwavelength data of other instruments. The full LHAASO array has started to collect data since the end of July 2021. We anticipate more interesting discoveries by LHAASO in the coming years.

References

1. F. Aharonian, Q. An, Axikegu, L. X. Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Z. Cao, Z. Cao, J. Chang, J. F. Chang, X. C. Chang, B. M. Chen, J. Chen, L. Chen, L. Chen, L. Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. Della Volpe, B. D. Piazzoli, X. J. Dong, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, Q. Gao, W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, Q. B. Gou, M. H. Gu, J. G. Guo, X. L. Guo, Y. Q. Guo, Y. Y. Guo, Y. A. Han, H. H. He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, X. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, D. Kuleshov, K. Levochkin, B. B. Li, C. Li, C. Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, J. Li, K. Li, W. L. Li, X. Li, X. Li, X. R. Li, Y. Li, Y. Z. Li, Z. Li, Z. Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, D. Ruffolo, V. Rulev, A. Sáiz, L. Shao, O. Shchegolev, X. D. Sheng, J. R. Shi, H. C. Song, Y. V. Stenkin, V. Stepanov, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, W. Wang, W. Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. D. Wang, Y. J. Wang, Y. P. Wang, Z. Wang, Z. Wang, Z. H. Wang, Z. X. Wang, D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. Zhang, L. Zhang, L. X. Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. Zhang, Y. Zhang, Y. F. Zhang, Y. L. Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu, X. Zuo and (Lhaaso Collaboration), Observation of the Crab Nebula with LHAASO-KM2A - a performance study, *Chinese Physics C* **45**, p. 025002 (February 2021).
2. Z. Cao, F. A. Aharonian, Q. An, L. X. Axikegu, Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Z. Cao, J. Chang, J. F. Chang, X. C. Chang, B. M. Chen, J. Chen, L. Chen, L. Chen, L. Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. della Volpe, B. D'Ettorre Piazzoli, X. J. Dong, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, Q. Gao,

W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, Q. B. Gou, M. H. Gu, J. G. Guo, X. L. Guo, Y. Q. Guo, Y. Y. Guo, Y. A. Han, H. H. He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, X. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, D. Kuleshov, K. Levochkin, B. B. Li, C. Li, C. Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, J. Li, K. Li, W. L. Li, X. Li, X. Li, X. R. Li, Y. Li, Y. Z. Li, Z. Li, Z. Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, D. Ruffolo, V. Rulev, A. Sáiz, L. Shao, O. Shchegolev, X. D. Sheng, J. R. Shi, H. C. Song, Y. V. Stenkin, V. Stepanov, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, W. Wang, W. Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. D. Wang, Y. J. Wang, Y. P. Wang, Z. Wang, Z. Wang, Z. H. Wang, Z. X. Wang, D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. Zhang, L. Zhang, L. X. Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. Zhang, Y. Zhang, Y. F. Zhang, Y. L. Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu and X. Zuo, Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources, *Nature* **594**, 33 (January 2021).

3. Z. Cao, F. A. Aharonian, Q. An, L. X. Axikegu, Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Z. Cao, J. Chang, J. F. Chang, X. C. Chang, B. M. Chen, J. Chen, L. Chen, L. Chen, L. Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. della Volpe, B. D'Ettorre Piazzoli, X. J. Dong, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, Q. Gao, W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, Q. B. Gou, M. H. Gu, J. G. Guo, X. L. Guo, Y. Q. Guo, Y. Y. Guo, Y. A. Han, H. H. He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, X. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, D. Kuleshov, K. Levochkin, B. B. Li, C. Li, C. Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, J. Li, K. Li, W. L. Li, X. Li, X. Li, X. R. Li, Y. Li, Y. Z. Li, Z. Li, Z. Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, D. Ruffolo, V. Rulev, A. Sáiz, L. Shao, O. Shchegolev, X. D. Sheng, J. R. Shi, H. C. Song, Y. V. Stenkin, V. Stepanov, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, W. Wang, W. Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. D. Wang,

Y. J. Wang, Y. P. Wang, Z. Wang, Z. Wang, Z. H. Wang, Z. X. Wang, D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. Zhang, L. Zhang, L. X. Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. Zhang, Y. Zhang, Y. F. Zhang, Y. L. Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu and X. Zuo, Peta–electron volt gamma-ray emission from the crab nebula, *Science* **373**, 425 (2021).

4. A. J. B. Downes, T. Pauls and C. J. Salter, G 40.5-0.5 : A previously unrecognised supernova remnant in Aql., *A&A* **92**, 47 (December 1980).
5. J. Yang, J.-L. Zhang, Z.-Y. Cai, D.-R. Lu and Y.-H. Tan, Molecular Gas Distribution around the Supernova Remnant G40.5 0.5, *CJA&A* **6**, 210 (April 2006).
6. F. Aharonian, A. G. Akhperjanian, G. Anton, U. Barres de Almeida, A. R. Bazer-Bachi, Y. Becherini, B. Behera, W. Benbow, K. Bernlöhr, C. Boisson, A. Bochow, V. Borrel, I. Braun, E. Brion, J. Brucker, P. Brun, R. Bühler, T. Bulik, I. Büsching, T. Boutelier, S. Carrigan, P. M. Chadwick, A. Charbonnier, R. C. G. Chaves, A. Cheesebrough, L. M. Chounet, A. C. Clapson, G. Coignet, M. Dalton, M. K. Daniel, B. Degrange, C. Deil, H. J. Dickinson, A. Djannati-Ataï, W. Domainko, L. O'C. Drury, F. Dubois, G. Dubus, J. Dyks, M. Dyrda, K. Egberts, D. Emmanoulopoulos, P. Espigat, C. Farnier, F. Feinstein, A. Fiasson, A. Förster, G. Fontaine, M. Füßling, S. Gabici, Y. A. Gallant, L. Gérard, B. Giebels, J. F. Glicenstein, B. Glück, P. Goret, D. Hauser, M. Hauser, S. Heinz, G. Heinzelmann, G. Henri, G. Hermann, J. A. Hinton, A. Hoffmann, W. Hofmann, M. Holleran, S. Hoppe, D. Horns, A. Jacholkowska, O. C. de Jager, I. Jung, K. Katarzyński, U. Katz, S. Kaufmann, E. Kendziorra, M. Kerschhaggl, D. Khangulyan, B. Khélifi, D. Keogh, N. Komin, K. Kosack, G. Lamanna, J. P. Lenain, T. Lohse, V. Marandon, J. M. Martin, O. Martineau-Huynh, A. Marcowith, D. Maurin, T. J. L. McComb, M. C. Medina, R. Moderski, E. Moulin, M. Naumann-Godo, M. de Naurois, D. Nedbal, D. Nekrassov, J. Niemiec, S. J. Nolan, S. Ohm, J. F. Olive, E. de Oña Wilhelmi, K. J. Orford, M. Ostrowski, M. Panter, M. Paz Arribas, G. Pedaletti, G. Pelletier, P. O. Petrucci, S. Pita, G. Pühlhofer, M. Punch, A. Quirrenbach, B. C. Raubenheimer, M. Raue, S. M. Rayner, M. Renaud, O. Reimer, F. Rieger, J. Ripken, L. Rob, S. Rosier-Lees, G. Rowell, B. Rudak, C. B. Rulten, J. Ruppel, V. Sahakian, A. Santangelo, R. Schlickeiser, F. M. Schöck, R. Schröder, U. Schwanke, S. Schwarzburg, S. Schwemmer, A. Shalchi, J. L. Skilton, H. Sol, D. Spangler, Ł. Stawarz, R. Steenkamp, C. Stegmann, G. Superina, P. H. Tam, J. P. Tavernet, R. Terrier, O. Tibolla, C. van Eldik, G. Vasileiadis, C. Venter, L. Venter, J. P. Vialle, P. Vincent, M. Vivier, H. J. Völk, F. Volpe, S. J. Wagner, M. Ward, A. A. Zdziarski and A. Zech, Detection of very high energy radiation from HESS J1908+063 confirms the Milagro unidentified source MGRO J1908+06, *A&A* **499**, 723 (June 2009).
7. B. Bartoli, P. Bernardini, X. J. Bi, C. Bleve, I. Bolognino, P. Branchini, A. Budano, A. K. Calabrese Melcarne, P. Camarri, Z. Cao, R. Cardarelli, S. Catalanotti, C. Cattaneo, S. Z. Chen, T. L. Chen, Y. Chen, P. Creti, S. W. Cui, B. Z. Dai, G. D'Alí Staiti, Danzengluobu, M. Dattoli, I. De Mitri, B. D'Ettorre Piazzoli, T. Di Girolamo, X. H. Ding, G. Di Sciascio, C. F. Feng, Z. Feng, Z. Feng, F. Galeazzi, E. Giroletti, Q. B.

Gou, Y. Q. Guo, H. H. He, H. Hu, H. Hu, Q. Huang, M. Iacovacci, R. Iuppa, I. James, H. Y. Jia, Labaciren, H. J. Li, J. Y. Li, X. X. Li, G. Liguori, C. Liu, C. Q. Liu, J. Liu, M. Y. Liu, H. Lu, X. H. Ma, G. Mancarella, S. M. Mari, G. Marsella, D. Martello, S. Mastroianni, P. Montini, C. C. Ning, A. Pagliaro, M. Panareo, B. Panico, L. Perrone, P. Pistilli, X. B. Qu, F. Ruggieri, P. Salvini, R. Santonico, P. R. Shen, X. D. Sheng, F. Shi, C. Stanescu, A. Surdo, Y. H. Tan, P. Vallania, S. Vernetto, C. Vigorito, B. Wang, H. Wang, C. Y. Wu, H. R. Wu, B. Xu, L. Xue, Y. X. Yan, Q. Y. Yang, X. C. Yang, Z. G. Yao, A. F. Yuan, M. Zha, H. M. Zhang, J. Zhang, J. Zhang, L. Zhang, P. Zhang, X. Y. Zhang, Y. Zhang, Zhaxiciren, Zhaxisangzhu, X. X. Zhou, F. R. Zhu, Q. Q. Zhu, G. Zizzi and Argo-YBJ Collaboration, Observation of the TeV Gamma-Ray Source MGRO J1908+06 with ARGO-YBJ, *ApJ* **760**, p. 110 (December 2012).

8. E. Aliu, S. Archambault, T. Aune, B. Behera, M. Beilicke, W. Benbow, K. Berger, R. Bird, J. H. Buckley, V. Bugaev, J. V. Cardenzana, M. Cerruti, X. Chen, L. Ciupik, E. Collins-Hughes, M. P. Connolly, W. Cui, J. Dumm, V. V. Dwarkadas, M. Errando, A. Falcone, S. Federici, Q. Feng, J. P. Finley, H. Fleischhack, P. Fortin, L. Fortson, A. Furniss, N. Galante, D. Gall, G. H. Gillanders, S. Griffin, S. T. Griffiths, J. Grube, G. Gyuk, D. Hanna, J. Holder, G. Hughes, T. B. Humensky, P. Kaaret, M. Kertzman, Y. Khassen, D. Kieda, F. Krennrich, S. Kumar, M. J. Lang, A. S. Madhavan, G. Maier, A. J. McCann, K. Meagher, J. Millis, P. Moriarty, R. Mukherjee, D. Nieto, A. O'Faoláin de Bhróithe, R. A. Ong, A. N. Otte, D. Pandel, N. Park, M. Pohl, A. Popkow, H. Prokoph, J. Quinn, K. Ragan, J. Rajotte, G. Ratliff, L. C. Reyes, P. T. Reynolds, G. T. Richards, E. Roache, J. Rousselle, G. H. Sembroski, K. Shahinyan, F. Sheidaei, A. W. Smith, D. Staszak, I. Telezhinsky, K. Tsurusaki, J. V. Tucci, J. Tyler, A. Varlotta, V. V. Vassiliev, S. Vincent, S. P. Wakely, J. E. Ward, A. Weinstein, R. Welsing and A. Wilhelm, Investigating the TeV Morphology of MGRO J1908+06 with VERITAS, *ApJ* **787**, p. 166 (June 2014).
9. A. U. Abeysekara, A. Albert, R. Alfaro, J. R. Angeles Camacho, J. C. Arteaga-Velázquez, K. P. Arunbabu, D. Avila Rojas, H. A. Ayala Solares, V. Baghmanyan, E. Belmont-Moreno, S. Y. BenZvi, C. Brisbois, K. S. Caballero-Mora, T. Capistrán, A. Carramiñana, S. Casanova, U. Cotti, J. Cotzomi, S. Coutiño de León, E. De la Fuente, C. de León, S. Dichiara, B. L. Dingus, M. A. DuVernois, J. C. Díaz-Vélez, R. W. Ellsworth, K. Engel, C. Espinoza, H. Fleischhack, N. Fraija, A. Galván-Gámez, D. Garcia, J. A. García-González, F. Garfias, M. M. González, J. A. Goodman, J. P. Harding, S. Hernandez, J. Hinton, B. Hona, D. Huang, F. Hueyotl-Zahuantitla, P. Hüntemeyer, A. Iriarte, A. Jardin-Blicq, V. Joshi, S. Kaufmann, D. Kieda, A. Lara, W. H. Lee, H. León Vargas, J. T. Linnemann, A. L. Longinotti, G. Luis-Raya, J. Lundeen, R. López-Coto, K. Malone, S. S. Marinelli, O. Martinez, I. Martinez-Castellanos, J. Martínez-Castro, H. Martínez-Huerta, J. A. Matthews, P. Miranda-Romagnoli, J. A. Morales-Soto, E. Moreno, M. Mostafá, A. Nayerhoda, L. Nellen, M. Newbold, M. U. Nisa, R. Noriega-Papaqui, A. Peisker, E. G. Pérez-Pérez, J. Pretz, Z. Ren, C. D. Rho, C. Rivière, D. Rosa-González, M. Rosenberg, E. Ruiz-Velasco, F. Salesa Greus, A. Sandoval, M. Schneider, H. Schoorlemmer, G. Sinnis, A. J. Smith, R. W. Springer, P. Surajbali, E. Tabachnick, M. Tanner, O. Tibolla, K. Tollefson, I. Torres, R. Torres-Escobedo, L. Villaseñor, T. Weisgarber, J. Wood, T. Yapici, H. Zhang, H. Zhou and HAWC Collaboration, Multiple Galactic Sources with Emission Above 56 TeV Detected by HAWC, *PRL* **124**, p. 021102 (January 2020).

10. J. Li, R.-Y. Liu, E. de Oña Wilhelmi, D. F. Torres, Q.-C. Liu, M. Kerr, R. Bühler, Y. Su, H.-N. He and M.-Y. Xiao, Investigating the Nature of MGRO J1908+06 with Multiwavelength Observations, *ApJL* **913**, p. L33 (June 2021).
11. G. G. Pavlov, O. Kargaltsev and W. F. Brisken, Chandra Observation of PSR B1823-13 and Its Pulsar Wind Nebula, *ApJ* **675**, 683 (Mar 2008).
12. H. Uchiyama, H. Matsumoto, T. G. Tsuru, K. Koyama and A. Bamba, Suzaku Observation of HESS J1825-137: Discovery of Largely-Extended X-Rays from PSR J1826-1334, *PASJ* **61**, p. S189 (January 2009).
13. A. Van Etten and R. W. Romani, Multi-zone Modeling of the Pulsar Wind Nebula HESS J1825-137, *ApJ* **742**, p. 62 (Dec 2011).
14. F. Voisin, G. Rowell, M. G. Burton, A. Walsh, Y. Fukui and F. Aharonian, ISM gas studies towards the TeV PWN HESS J1825-137 and northern region, *MNRAS* **458**, 2813 (May 2016).
15. D. Khangulyan, A. V. Koldoba, G. V. Ustyugova, S. V. Bogovalov and F. Aharonian, On the Anomalously Large Extension of the Pulsar Wind Nebula HESS J1825-137, *ApJ* **860**, p. 59 (June 2018).
16. H. E. S. S. Collaboration, H. Abdalla, F. Aharonian, F. Ait Benkhali, E. O. Angüner, M. Arakawa, C. Arcaro, C. Armand, M. Arrieta, M. Backes, M. Barnard, Y. Becherini, J. Becker Tjus, D. Berge, K. Bernlöhr, R. Blackwell, M. Böttcher, C. Boisson, J. Bolmont, S. Bonnefoy, P. Bordas, J. Bregeon, F. Brun, P. Brun, M. Bryan, M. Büchele, T. Bulik, T. Bylund, M. Capasso, S. Caroff, A. Carosi, S. Casanova, M. Cerruti, N. Chakraborty, T. Chand, S. Chandra, R. C. G. Chaves, A. Chen, S. Colafrancesco, B. Condon, I. D. Davids, C. Deil, J. Devin, P. deWilt, L. Dirson, A. Djannati-Ataï, A. Dmytriiev, A. Donath, V. Doroshenko, L. O. C. Drury, J. Dyks, K. Egberts, G. Emery, J. P. Ernenwein, S. Eschbach, S. Fegan, A. Fiasson, G. Fontaine, S. Funk, M. Füßling, S. Gabici, Y. A. Gallant, F. Gaté, G. Giavitto, D. Glawion, J. F. Glicenstein, D. Gottschall, M. H. Grondin, J. Hahn, M. Haupt, G. Heinzelmann, G. Henri, G. Hermann, J. A. Hinton, W. Hofmann, C. Hoischen, T. L. Holch, M. Holler, D. Horns, D. Huber, H. Iwasaki, A. Jacholkowska, M. Jamrozy, D. Jankowsky, F. Jankowsky, L. Jouvin, I. Jung-Richardt, M. A. Kastendieck, K. Katarzyński, M. Katsuragawa, U. Katz, D. Kerszberg, D. Khangulyan, B. Khélifi, J. King, S. Klepser, W. Kluźniak, N. Komin, K. Kosack, M. Kraus, G. Lamanna, J. Lau, J. Lefaucheur, A. Lemière, M. Lemoine-Goumard, J. P. Lenain, E. Leser, T. Lohse, R. López-Coto, I. Lypova, D. Malyshev, V. Marandon, A. Marcowith, C. Mariaud, G. Martí-Devesa, R. Marx, G. Maurin, P. J. Meintjes, A. M. W. Mitchell, R. Moderski, M. Mohamed, L. Mohrmann, C. Moore, E. Moulin, T. Murach, S. Nakashima, M. de Naurois, H. Ndiyavala, F. Niederwanger, J. Niemiec, L. Oakes, P. O'Brien, H. Odaka, S. Ohm, M. Ostrowski, I. Oya, M. Panter, R. D. Parsons, C. Perennes, P. O. Petrucci, B. Peyaud, Q. Piel, S. Pita, V. Poireau, A. Priyana Noel, D. A. Prokhorov, H. Prokoph, G. Pühlhofer, M. Punch, A. Quirrenbach, S. Raab, R. Rauth, A. Reimer, O. Reimer, M. Renaud, F. Rieger, L. Rinchiuso, C. Romoli, G. Rowell, B. Rudak, E. Ruiz-Velasco, V. Sahakian, S. Saito, D. A. Sanchez, A. Santangelo, M. Sasaki, R. Schlickeiser, F. Schüssler, A. Schulz, H. Schutte, U. Schwanke, S. Schwemmer, M. Seglar-Arroyo, M. Senniappan, A. S. Seyffert, N. Shafi, I. Shilon, K. Shiningayamwe, R. Simoni, A. Sinha, H. Sol, A. Specovius, M. Spir-Jacob, Ł. Stawarz, R. Steenkamp, C. Stegmann, C. Steppa, T. Takahashi, J. P. Tavernet, T. Tavernier, A. M. Taylor, R. Terrier, L. Tibaldo, D. Tiziani, M. Tluczykont, C. Trichard, M. Tsirou, N. Tsuji, R. Tuffs, Y. Uchiyama, D. J. van der Walt, C. van Eldik, C. van Rensburg, B. van Soelen, G. Vasileiadis, J. Veh, C. Venter, P. Vincent, J. Vink, F. Voisin, H. J. Völk, T. Vuillaume, Z. Wadiasingh, S. J. Wagner, R. M. Wagner, R. White, A. Wierzcholska,

R. Yang, H. Yoneda, D. Zaborov, M. Zacharias, R. Zanin, A. A. Zdziarski, A. Zech, F. Zefi, A. Ziegler, J. Zorn and N. Żywucka, Particle transport within the pulsar wind nebula HESS J1825-137, *A&A* **621**, p. A116 (January 2019).

17. R.-Y. Liu and H. Yan, On the unusually large spatial extent of the TeV nebula HESS J1825-137: implication from the energy-dependent morphology, *MNRAS* **494**, 2618 (April 2020).
18. G. Principe, A. M. W. Mitchell, S. Caroff, J. A. Hinton, R. D. Parsons and S. Funk, Energy dependent morphology of the pulsar wind nebula HESS J1825-137 with Fermi-LAT, *A&A* **640**, p. A76 (August 2020).
19. A. Albert, R. Alfaro, C. Alvarez, J. R. A. Camacho, J. C. Arteaga-Velázquez, K. P. Arunbabu, D. A. Rojas, H. A. Ayala Solares, V. Baghmanyan, E. Belmont-Moreno, S. Y. BenZvi, C. Brisbois, T. Capistrán, A. Carramiñana, S. Casanova, U. Cotti, J. Cotzomi, E. D. l. Fuente, R. D. Hernandez, B. L. Dingus, M. A. DuVernois, M. Durocher, J. C. Díaz-Vélez, K. Engel, C. Espinoza, K. Fang, H. Fleischhack, N. Fraija, A. Galván-Gámez, D. Garcia, J. A. García-González, F. Garfias, G. Giacinti, M. M. González, J. A. Goodman, J. P. Harding, B. Hona, D. Huang, F. Hueyotl-Zahuantitla, P. Hüntemeyer, A. Iriarte, A. Jardin-Blicq, V. Joshi, G. J. Kunde, A. Lara, W. H. Lee, H. L. Vargas, J. T. Linnemann, A. L. Longinotti, G. Luis-Raya, J. Lundeen, K. Malone, V. Marandon, O. Martinez, J. Martínez-Castro, J. A. Matthews, P. Miranda-Romagnoli, E. Moreno, M. Mostafá, A. Nayerhoda, L. Nellen, M. Newbold, M. U. Nisa, R. Noriega-Papaqui, N. Omodei, A. Peisker, Y. P. Araujo, E. G. Pérez-Pérez, C. D. Rho, D. Rosa-González, H. Salazar, F. S. Greus, A. Sandoval, M. Schneider, F. Serna, R. W. Springer, K. Tollefson, I. Torres, R. Torres-Escobedo, F. Ureña-Mena, L. Villaseñor, E. Willox, H. Zhou and C. d. León, Evidence of 200 TeV Photons from HAWC J1825-134, *ApJL* **907**, p. L30 (February 2021).
20. J. P. Halpern, E. V. Gotthelf, K. M. Leighly and D. J. Helfand , A Possible X-Ray and Radio Counterpart of the High-Energy Gamma-Ray Source 3EG J2227+6122, *ApJ* **547**, 323 (January 2001).
21. R. Kothes, B. Uyaniker and S. Pineault, The Supernova Remnant G106.3+2.7 and Its Pulsar-Wind Nebula: Relics of Triggered Star Formation in a Complex Environment, *ApJ* **560**, 236 (October 2001).
22. V. A. Acciari, E. Aliu, T. Arlen, T. Aune, M. Bautista, M. Beilicke, W. Benbow, D. Boltuch, S. M. Bradbury, J. H. Buckley, V. Bugaev, Y. Butt, K. Byrum, A. Cannon, A. Cesarini, Y. C. Chow, L. Ciupik, P. Cogan, W. Cui, R. Dickherber, T. Ergin, S. J. Fegan, J. P. Finley, P. Fortin, L. Fortson, A. Furniss, D. Gall, G. H. Gillanders, E. V. Gotthelf, J. Grube, R. Guenette, G. Gyuk, D. Hanna, J. Holder, D. Horan, C. M. Hui, T. B. Humensky, P. Kaaret, N. Karlsson, M. Kertzman, D. Kieda, A. Konopelko, H. Krawczynski, F. Krennrich, M. J. Lang, S. LeBohec, G. Maier, A. McCann, M. McCutcheon, J. Millis, P. Moriarty, R. Mukherjee, R. A. Ong, A. N. Otte, D. Pandel, J. S. Perkins, M. Pohl, J. Quinn, K. Ragan, L. C. Reyes, P. T. Reynolds, E. Roache, H. J. Rose, M. Schroedter, G. H. Sembroski, A. W. Smith, D. Steele, S. P. Swordy, M. Theiling, J. A. Toner, V. V. Vassiliev, S. Vincent, R. G. Wagner, S. P. Wakely, J. E. Ward, T. C. Weekes, A. Weinstein, T. Weisgarber, D. A. Williams, S. Wissel, M. Wood and B. Zitzer, Detection of Extended VHE Gamma Ray Emission from G106.3+2.7 with Veritas, *ApJL* **703**, L6 (September 2009).
23. Y. Xin, H. Zeng, S. Liu, Y. Fan and D. Wei, VER J2227+608: A Hadronic PeVatron Pulsar Wind Nebula?, *ApJ* **885**, p. 162 (November 2019).
24. S. Liu, H. Zeng, Y. Xin and H. Zhu, Hadronic versus Leptonic Models for γ-Ray Emission from VER J2227+608, *ApJL* **897**, p. L34 (July 2020).

25. Tibet ASγ Collaboration, M. Amenomori, Y. W. Bao, X. J. Bi, D. Chen, T. L. Chen, W. Y. Chen, X. Chen, Y. Chen, S. W. Cirennima, Cui, L. K. Danzengluobu, Ding, J. H. Fang, K. Fang, C. F. Feng, Z. Feng, Z. Y. Feng, Q. Gao, Q. B. Gou, Y. Q. Guo, Y. Y. Guo, H. H. He, Z. T. He, K. Hibino, N. Hotta, H. Hu, H. B. Hu, J. Huang, H. Y. Jia, L. Jiang, H. B. Jin, K. Kasahara, Y. Katayose, C. Kato, S. Kato, K. Kawata, W. Kihara, Y. Ko, M. Kozai, G. M. Labaciren, Le, A. F. Li, H. J. Li, W. J. Li, Y. H. Lin, B. Liu, C. Liu, J. S. Liu, M. Y. Liu, W. Liu, Y. Q. Lou, H. Lu, X. R. Meng, K. Munakata, H. Nakada, Y. Nakamura, H. Nanjo, M. Nishizawa, M. Ohnishi, T. Ohura, S. Ozawa, X. L. Qian, X. B. Qu, T. Saito, M. Sakata, T. K. Sako, J. Shao, M. Shibata, A. Shiomi, H. Sugimoto, W. Takano, M. Takita, Y. H. Tan, N. Tateyama, S. Torii, H. Tsuchiya, S. Udo, H. Wang, H. R. Wu, L. Xue, Y. Yamamoto, Z. Yang, Y. Yokoe, A. F. Yuan, L. M. Zhai, H. M. Zhang, J. L. Zhang, X. Zhang, X. Y. Zhang, Y. Zhang, Y. Zhang, Y. Zhang, S. P. Zhao and X. X. Zhaxisangzhu, Zhou, Potential PeVatron supernova remnant G106.3+2.7 seen in the highest-energy gamma rays, *Nature Astronomy* **5**, 460 (January 2021).
26. C. Ge, R.-Y. Liu, S. Niu, Y. Chen and X.-Y. Wang, Revealing a peculiar supernova remnant G106.3+2.7 as a petaelectronvolt proton accelerator with X-ray observations, *The Innovation* **2**, p. 100118 (May 2021).
27. Y. Fujita, A. Bamba, K. K. Nobukawa and H. Matsumoto, X-Ray Emission from the PeVatron-candidate Supernova Remnant G106.3+2.7, *ApJ* **912**, p. 133 (May 2021).
28. C. J. Cesarsky and T. Montmerle, Gamma-Rays from Active Regions in the Galaxy - the Possible Contribution of Stellar Winds, *SSR* **36**, 173 (October 1983).
29. F. Aharonian, R. Yang and E. de Oña Wilhelmi, Massive stars as major factories of Galactic cosmic rays, *Nature Astronomy* **3**, 561 (March 2019).
30. A. M. Bykov, A. Marcowith, E. Amato, M. E. Kalyashova, J. M. D. Kruijssen and E. Waxman, High-Energy Particles and Radiation in Star-Forming Regions, *SSR* **216**, p. 42 (April 2020).
31. F. Aharonian, Q. An, L. X. Axikegu, Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Z. Cao, Z. Cao, J. Chang, J. F. Chang, X. C. Chang, B. M. Chen, J. Chen, L. Chen, L. Chen, L. Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. Della Volpe, B. D'Ettorre Piazzoli, X. J. Dong, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, Q. Gao, W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, Q. B. Gou, M. H. Gu, J. G. Guo, X. L. Guo, Y. Q. Guo, Y. Y. Guo, Y. A. Han, H. H. He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, X. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, D. Kuleshov, K. Levochkin, B. B. Li, C. Li, C. Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, J. Li, K. Li, W. L. Li, X. Li, X. Li, X. R. Li, Y. Li, Y. Z. Li, Z. Li, Z. Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, D. Ruffolo, V. Rulev, A. Sáiz, L. Shao, O. Shchegolev, X. D. Sheng, J. R. Shi, H. C. Song, Y. V. Stenkin, V. Stepanov, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, W. Wang, W. Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. D. Wang, Y. J. Wang, Y. P. Wang, Z. Wang, Z. Wang, Z. H. Wang, Z. X. Wang,

D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. Zhang, L. Zhang, L. X. Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. Zhang, Y. Zhang, Y. F. Zhang, Y. L. Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu, X. Zuo, LHAASO Collaboration and X. Y. Huang, Extended Very-High-Energy Gamma-Ray Emission Surrounding PSR J 0622 +3749 Observed by LHAASO-KM2A, *PRL* **126**, p. 241103 (June 2021).

32. Z. Cao, F. Aharonian, Q. An, Axikegu, L. X. Bai, Y. X. Bai, Y. W. Bao, D. Bastieri, X. J. Bi, Y. J. Bi, H. Cai, J. T. Cai, Z. Cao, J. Chang, J. F. Chang, B. M. Chen, E. S. Chen, J. Chen, L. Chen, L. Chen, L. Chen, M. J. Chen, M. L. Chen, Q. H. Chen, S. H. Chen, S. Z. Chen, T. L. Chen, X. L. Chen, Y. Chen, N. Cheng, Y. D. Cheng, S. W. Cui, X. H. Cui, Y. D. Cui, B. D'Ettorre Piazzoli, B. Z. Dai, H. L. Dai, Z. G. Dai, Danzengluobu, D. d. Volpe, X. J. Dong, K. K. Duan, J. H. Fan, Y. Z. Fan, Z. X. Fan, J. Fang, K. Fang, C. F. Feng, L. Feng, S. H. Feng, Y. L. Feng, B. Gao, C. D. Gao, L. Q. Gao, Q. Gao, W. Gao, M. M. Ge, L. S. Geng, G. H. Gong, Q. B. Gou, M. H. Gu, F. L. Guo, J. G. Guo, X. L. Guo, Y. Q. Guo, Y. Y. Guo, Y. A. Han, H. H. He, H. N. He, J. C. He, S. L. He, X. B. He, Y. He, M. Heller, Y. K. Hor, C. Hou, H. B. Hu, S. Hu, S. C. Hu, X. J. Hu, D. H. Huang, Q. L. Huang, W. H. Huang, X. T. Huang, X. Y. Huang, Z. C. Huang, F. Ji, X. L. Ji, H. Y. Jia, K. Jiang, Z. J. Jiang, C. Jin, T. Ke, D. Kuleshov, K. Levochkin, B. B. Li, C. Li, C. Li, F. Li, H. B. Li, H. C. Li, H. Y. Li, J. Li, K. Li, W. L. Li, X. R. Li, X. Li, X. Li, Y. Li, Y. Z. Li, Z. Li, Z. Li, E. W. Liang, Y. F. Liang, S. J. Lin, B. Liu, C. Liu, D. Liu, H. Liu, H. D. Liu, J. Liu, J. L. Liu, J. S. Liu, J. Y. Liu, M. Y. Liu, R. Y. Liu, S. M. Liu, W. Liu, Y. Liu, Y. N. Liu, Z. X. Liu, W. J. Long, R. Lu, H. K. Lv, B. Q. Ma, L. L. Ma, X. H. Ma, J. R. Mao, A. Masood, Z. Min, W. Mitthumsiri, T. Montaruli, Y. C. Nan, B. Y. Pang, P. Pattarakijwanich, Z. Y. Pei, M. Y. Qi, Y. Q. Qi, B. Q. Qiao, J. J. Qin, D. Ruffolo, V. Rulev, A. Sáiz, L. Shao, O. Shchegolev, X. D. Sheng, J. Y. Shi, H. C. Song, Y. V. Stenkin, V. Stepanov, Y. Su, Q. N. Sun, X. N. Sun, Z. B. Sun, P. H. T. Tam, Z. B. Tang, W. W. Tian, B. D. Wang, C. Wang, H. Wang, H. G. Wang, J. C. Wang, J. S. Wang, L. P. Wang, L. Y. Wang, R. N. Wang, W. Wang, W. Wang, X. G. Wang, X. J. Wang, X. Y. Wang, Y. Wang, Y. D. Wang, Y. J. Wang, Y. P. Wang, Z. H. Wang, Z. X. Wang, Z. Wang, Z. Wang, D. M. Wei, J. J. Wei, Y. J. Wei, T. Wen, C. Y. Wu, H. R. Wu, S. Wu, W. X. Wu, X. F. Wu, S. Q. Xi, J. Xia, J. J. Xia, G. M. Xiang, D. X. Xiao, G. Xiao, H. B. Xiao, G. G. Xin, Y. L. Xin, Y. Xing, D. L. Xu, R. X. Xu, L. Xue, D. H. Yan, J. Z. Yan, C. W. Yang, F. F. Yang, J. Y. Yang, L. L. Yang, M. J. Yang, R. Z. Yang, S. B. Yang, Y. H. Yao, Z. G. Yao, Y. M. Ye, L. Q. Yin, N. Yin, X. H. You, Z. Y. You, Y. H. Yu, Q. Yuan, H. D. Zeng, T. X. Zeng, W. Zeng, Z. K. Zeng, M. Zha, X. X. Zhai, B. B. Zhang, H. M. Zhang, H. Y. Zhang, J. L. Zhang, J. W. Zhang, L. X. Zhang, L. Zhang, L. Zhang, P. F. Zhang, P. P. Zhang, R. Zhang, S. R. Zhang, S. S. Zhang, X. Zhang, X. P. Zhang, Y. F. Zhang, Y. L. Zhang, Y. Zhang, Y. Zhang, B. Zhao, J. Zhao, L. Zhao, L. Z. Zhao, S. P. Zhao, F. Zheng, Y. Zheng, B. Zhou, H. Zhou, J. N. Zhou, P. Zhou, R. Zhou, X. X. Zhou, C. G. Zhu, F. R. Zhu, H. Zhu, K. J. Zhu and X. Zuo, Discovery of a New Gamma-Ray Source, LHAASO J0341+5258, with Emission up to 200 TeV, *ApJL* **917**, p. L4 (August 2021).

33. H. Akaike, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control* **19**, 716 (January 1974).
34. M. Amenomori, Y. W. Bao, X. J. Bi, D. Chen, T. L. Chen, W. Y. Chen, X. Chen, Y. Chen, Cirennima, S. W. Cui, Danzengluobu, L. K. Ding, J. H. Fang, K. Fang, C. F. Feng, Z. Feng, Z. Y. Feng, Q. Gao, Q. B. Gou, Y. Q. Guo, H. H. He, Z. T. He, K. Hibino, N. Hotta, H. Hu, H. B. Hu, J. Huang, H. Y. Jia, L. Jiang, H. B. Jin, F. Kajino, K. Kasahara, Y. Katayose, C. Kato, S. Kato, K. Kawata, M. Kozai, Labaciren, G. M. Le, A. F. Li, H. J. Li, W. J. Li, Y. H. Lin, B. Liu, C. Liu, J. S. Liu, M. Y. Liu, Y. Q. Lou, H. Lu, X. R. Meng, H. Mitsui, K. Munakata, Y. Nakamura, H. Nanjo, M. Nishizawa, M. Ohnishi, I. Ohta, S. Ozawa, X. L. Qian, X. B. Qu, T. Saito, M. Sakata, T. K. Sako, Y. Sengoku, J. Shao, M. Shibata, A. Shiomi, H. Sugimoto, M. Takita, Y. H. Tan, N. Tateyama, S. Torii, H. Tsuchiya, S. Udo, H. Wang, H. R. Wu, L. Xue, K. Yagisawa, Y. Yamamoto, Z. Yang, A. F. Yuan, L. M. Zhai, H. M. Zhang, J. L. Zhang, X. Zhang, X. Y. Zhang, Y. Zhang, Y. Zhang, Y. Zhang, Zhaxisangzhu, X. X. Zhou and Tibet AS γ Collaboration, First Detection of Photons with Energy beyond 100 TeV from an Astrophysical Source, *PRL* **123**, p. 051101 (August 2019).
35. MAGIC Collaboration, V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González, W. Bednarek, L. Bellizzi, E. Bernardini, A. Berti, J. Besenrieder, W. Bhattacharyya, C. Bigongiari, A. Biland, O. Blanch, G. Bonnoli, Ž. Bošnjak, G. Busetto, R. Carosi, G. Ceribella, Y. Chai, A. Chilingaryan, S. Cikota, S. M. Colak, U. Colin, E. Colombo, J. L. Contreras, J. Cortina, S. Covino, V. D'Elia, P. da Vela, F. Dazzi, A. de Angelis, B. de Lotto, M. Delfino, J. Delgado, D. Depaoli, F. di Pierro, L. di Venere, E. Do Souto Espiñeira, D. Dominis Prester, A. Donini, D. Dorner, M. Doro, D. Elsaesser, V. Fallah Ramazani, A. Fattorini, G. Ferrara, D. Fidalgo, L. Foffano, M. V. Fonseca, L. Font, C. Fruck, S. Fukami, R. J. García López, M. Garczarczyk, S. Gasparyan, M. Gaug, N. Giglietto, F. Giordano, N. Godinović, D. Green, D. Guberman, D. Hadasch, A. Hahn, J. Herrera, J. Hoang, D. Hrupec, M. Hütten, T. Inada, S. Inoue, K. Ishio, Y. Iwamura, L. Jouvin, D. Kerszberg, H. Kubo, J. Kushida, A. Lamastra, D. Lelas, F. Leone, E. Lindfors, S. Lombardi, F. Longo, M. López, R. López-Coto, A. López-Oramas, S. Loporchio, B. Machado de Oliveira Fraga, C. Maggio, P. Majumdar, M. Makariev, M. Mallamaci, G. Maneva, M. Manganaro, K. Mannheim, L. Maraschi, M. Mariotti, M. Martínez, D. Mazin, S. Mićanović, D. Miceli, M. Minev, J. M. Miranda, R. Mirzoyan, E. Molina, A. Moralejo, D. Morcuende, V. Moreno, E. Moretti, P. Munar-Adrover, V. Neustroev, C. Nigro, K. Nilsson, D. Ninci, K. Nishijima, K. Noda, L. Nogués, S. Nozaki, S. Paiano, J. Palacio, M. Palatiello, D. Paneque, R. Paoletti, J. M. Paredes, P. Peñil, M. Peresano, M. Persic, P. G. Prada Moroni, E. Prandini, I. Puljak, W. Rhode, M. Ribó, J. Rico, C. Righi, A. Rugliancich, L. Saha, N. Sahakyan, T. Saito, S. Sakurai, K. Satalecka, K. Schmidt, T. Schweizer, J. Sitarek, I. Šnidarić, D. Sobczynska, A. Somero, A. Stamerra, D. Strom, M. Strzys, Y. Suda, T. Surić, M. Takahashi, F. Tavecchio, P. Temnikov, T. Terzić, M. Teshima, N. Torres-Albà, L. Tosti, V. Vagelli, J. van Scherpenberg, G. Vanzo, M. Vazquez Acosta, C. F. Vigorito, V. Vitale, I. Vovk, M. Will and D. Zarić, MAGIC very large zenith angle observations of the Crab Nebula up to 100 TeV, *A&A* **635**, p. A158 (March 2020).
36. F. A. Aharonian, A. A. Belyanin, E. V. Derishev, V. V. Kocharovsky and V. V. Kocharovsky, Constraints on the extremely high-energy cosmic ray accelerators from classical electrodynamics, *PRD* **66**, p. 023005 (July 2002).
37. M. A. Malkov and L. O. Drury, Nonlinear theory of diffusive acceleration of particles by shock waves, *Reports on Progress in Physics* **64**, 429 (April 2001).

38. L. Kuiper, W. Hermsen, G. Cusumano, R. Diehl, V. Schönfelder, A. Strong, K. Bennett and M. L. McConnell, The Crab pulsar in the 0.75-30 MeV range as seen by CGRO COMPTEL. A coherent high-energy picture from soft X-rays up to high-energy gamma-rays, *A&A* **378**, 918 (November 2001).
39. M. Arakawa, M. Hayashida, D. Khangulyan and Y. Uchiyama, Detection of Small Flares from the Crab Nebula with Fermi-LAT, *ApJ* **897**, p. 33 (July 2020).
40. A. A. Abdo, M. Ackermann, M. Ajello, A. Allafort, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, R. D. Blandford, E. D. Bloom, E. Bonamente, A. W. Borgland, A. Bouvier, T. J. Brand t, J. Bregeon, A. Brez, M. Brigida, P. Bruel, R. Buehler, S. Buson, G. A. Caliandro, R. A. Cameron, A. Cannon, P. A. Caraveo, J. M. Casand jian, Ö. Çelik, E. Charles, A. Chekhtman, C. C. Cheung, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, L. Costamante, S. Cutini, F. D'Ammando, C. D. Dermer, A. de Angelis, A. de Luca, F. de Palma, S. W. Digel, E. do Couto e Silva, P. S. Drell, A. Drlica-Wagner, R. Dubois, D. Dumora, C. Favuzzi, S. J. Fegan, E. C. Ferrara, W. B. Focke, P. Fortin, M. Frailis, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Gehrels, S. Germani, N. Giglietto, F. Giordano, M. Giroletti, T. Glanzman, G. Godfrey, I. A. Grenier, M. H. Grondin, J. E. Grove, S. Guiriec, D. Hadasch, Y. Hanabata, A. K. Harding, K. Hayashi, M. Hayashida, E. Hays, D. Horan, R. Itoh, G. Jóhannesson, A. S. Johnson, T. J. Johnson, D. Khangulyan, T. Kamae, H. Katagiri, J. Kataoka, M. Kerr, J. Knödlseder, M. Kuss, J. Lande, L. Latronico, S. H. Lee, M. Lemoine-Goumard, F. Longo, F. Loparco, P. Lubrano, G. M. Madejski, A. Makeev, M. Marelli, M. N. Mazziotta, J. E. McEnery, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, C. Monte, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, T. Nakamori, M. Naumann-Godo, P. L. Nolan, J. P. Norris, E. Nuss, T. Ohsugi, A. Okumura, N. Omodei, J. F. Ormes, M. Ozaki, D. Paneque, D. Parent, V. Pelassa, M. Pepe, M. Pesce-Rollins, M. Pierbattista, F. Piron, T. A. Porter, S. Rainò, R. Rando, P. S. Ray, M. Razzano, A. Reimer, O. Reimer, T. Reposeur, S. Ritz, R. W. Romani, H. F. W. Sadrozinski, D. Sanchez, P. M. S. Parkinson, J. D. Scargle, T. L. Schalk, C. Sgrò, E. J. Siskind, P. D. Smith, G. Spand re, P. Spinelli, M. S. Strickman, D. J. Suson, H. Takahashi, T. Takahashi, T. Tanaka, J. B. Thayer, D. J. Thompson, L. Tibaldo, D. F. Torres, G. Tosti, A. Tramacere, E. Troja, Y. Uchiyama, J. Vandenbroucke, V. Vasileiou, G. Vianello, V. Vitale, P. Wang, K. S. Wood, Z. Yang and M. Ziegler, Gamma-Ray Flares from the Crab Nebula, *Science* **331**, p. 739 (February 2011).
41. M. Tavani, A. Bulgarelli, V. Vittorini, A. Pellizzoni, E. Striani, P. Caraveo, M. C. Weisskopf, A. Tennant, G. Pucella, A. Trois, E. Costa, Y. Evangelista, C. Pittori, F. Verrecchia, E. Del Monte, R. Campana, M. Pilia, A. De Luca, I. Donnarumma, D. Horns, C. Ferrigno, C. O. Heinke, M. Trifoglio, F. Gianotti, S. Vercellone, A. Argan, G. Barbiellini, P. W. Cattaneo, A. W. Chen, T. Contessi, F. D'Ammand o, G. DeParis, G. Di Cocco, G. Di Persio, M. Feroci, A. Ferrari, M. Galli, A. Giuliani, M. Giusti, C. Labanti, I. Lapshov, F. Lazzarotto, P. Lipari, F. Longo, F. Fuschino, M. Marisaldi, S. Mereghetti, E. Morelli, E. Moretti, A. Morselli, L. Pacciani, F. Perotti, G. Piano, P. Picozza, M. Prest, M. Rapisarda, A. Rappoldi, A. Rubini, S. Sabatini, P. Soffitta, E. Vallazza, A. Zambra, D. Zanello, F. Lucarelli, P. Santolamazza, P. Giommi, L. Salotti and G. F. Bignami, Discovery of Powerful Gamma-Ray Flares from the Crab Nebula, *Science* **331**, p. 736 (February 2011).

On the MG16 awards 2021

Remo Ruffini

ICRA and Dipartimento di Fisica, Università di Roma La Sapienza
Piazzale Aldo Moro 5, I-00185 Roma, Italy
ruffini@icra.it

ICRANet
Piazza della Repubblica 10, I-65122 Pescara, Italy
ruffini@icra.it

INAF
Viale del Parco Mellini 84, 00136 Rome, Italy
ruffini@icra.it

On Monday July 5, 2021 took place the official ceremony for the attribution of the 4 MG16 Individual Awards to Demetrios Christodoulou, Gerard 't Hooft, Tsvi Piran and Steven Weinberg, as well as for the attribution of the 3 MG16 Institutional Awards to the Spektrum-Roentgen-Gamma (SRG) mission of the Max Planck Institute for Extraterrestrial Physics (MPE), of the Space Research Institute (IKI) of the Russian Academy of Sciences and of the S.A. Lavochkin Association. The MG Awards consist in a silver casting of the TEST sculpture by the artist Attilio Pierelli. The MG16 Awards were presented to the Awardees by Prof. Roy P. Kerr.

Keywords: Black Holes; mass energy formula; Horizons; International scientific cooperation; Scientific space missions

1. The MG Awards

Since 1985, the MG Awards have been attributed on the occasion of the Marcel Grossmann meetings. The MG Award consist in a silver casting of the TEST sculpture by the artist Attilio Pierelli.

On the occasion of the MG4 meeting (July 1985, Rome, Italy) the Institutional Award went to the Vatican Observatory (presented to His Holiness Pope John Paul II), while the Individual Awards went to William Fairbank and to Abdus Salam. On the occasion of the MG5 meeting (August 1988, Perth, Australia) the Institutional Award went to the University of Western Australia (presented to the then Vice Chancellor, Prof. Robert Smith) while the Individual Awards went to Satio Hayakawa and to John Archibal Wheeler. On the occasion of the MG6 meeting (June 1991, Kyoto, Japan) the Institutional Award went to the RITP (presented to Prof. K. Tomita) while the Individual Awards went to Minoru Oda and to Stephen Hawking. On the occasion of the MG7 meeting (June 1994, Stanford, USA) the

Institutional Award went to the Space Telescope Institute (presented to Prof. Peter Stockman) while the Individual Awards went to Subrahmanyan Chandrasekhar and to Jim Wilson. On the occasion of the MG8 meeting (June 1997, Jerusalem) the Institutional Award went to the Hebrew University (presented to the then President M. Magidor), while the Individual Awards went to Tullio Regge and to Francis Everitt. On the occasion of the MG9 meeting (July 2000, Rome, Italy) the Institutional Award went to the Solvay Institutes (presented to Jacques Solvay), while the Individual Awards went to Cecile and Bryce Dewitt, Riccardo Giacconi and Roger Penrose. On the occasion of the MG10 meeting (July 2003, Rio de Janeiro, Brazil) the Institutional Award went to the Brazilian Center for Research in Physics - CBPF (presented to its founders Cesar Lattes, Josè Leite Lopez and Jayme Tiomno), while the Individual Awards went to Yvonne Choquet-Bruhat, to James W. York, Jr and to Yuval Ne'eman. On the occasion of the MG11 meeting (July 2006, Berlin, Germany) the Institutional Award went to the Freie Universitaet Berlin (presented to the then President Dr. Dieter Lenzen), while the Individual Awards went to Roy Kerr, George Coyne and Joachim Trumper. On the occasion of the MG12 meeting (July 2009, Paris, France) the Institutional Award went to the Institut des hautes études scientifiques - IHES (presented to Prof. Jean-Pierre Bourguignon), while the Individual Awards went to Jaan Einasto, to Christine Jones and to Michael Kramer. On the occasion of the MG13 meeting (July 2012, Stockholm, Sweden) the Institutional Award went to Albanova (presented to the Rector of the f Stockholm University, Prof. Kåre Bremer), while the Individual Awards went to David Arnett, to Vladimir Belinski and I. M. Khalatnikov and to Filippo Frontera. On the occasion of the MG14 meeting (July 2015, Rome, Italy) the Institutional Award went to the European Space Agency - ESA (presented to its Director General Johann-Dietrich Woerner), while the Individual Awards went to Ken'ichi Nomoto, to Martin Rees, to Yakov G. Sinai and to Sachiko Tsuruta. The 2 additional MG14 Individual Awards to Frank C.N. Yang and to T.D. Lee have been delivered on the occasion of the International Conference on Gravitation and Cosmology/the fourth Galileo-Xu Guangqi meeting in Beijing on May 4, 2015. On the occasion of the MG15 meeting (July 2018, Rome, Italy) the Institutional Award went to the Planck scientific collaboration - ESA (presented to the Principal Investigator of the High Frequency Instrument (HFI) Jean-Loup Puget) and to the Hansen experimental physics laboratory at Stanford University (presented to its Assistant Director Leo Hollberg), while the Individual Awards went to Lyman Page, to Rashid Sunyaev and to Shing-Tung Yau.

2. MG16 Individual Awardees

The MG16 Individual Awards went to Demetrios Christodoulou, Gerard 't Hooft, Tsvi Piran and Steven Weinberg.

2.1. *Demetrios Christodoulou*

Motivation: Professor **Demetrios Christodoulou** received the MG16 Award *"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."*

Scientific and historical background: It was back in 1967 that Achille Papapetrou mentioned the case of the 16 year old Demetrios Christodoulou to John Archibad Wheeler. Wheeler interviewed Demetrios in Paris and brought him immediately to Princeton where he was registered as an undergraduate at the university. After only one year he entered the graduate school and started collaborating with me. At the time I was working with Wheeler on the effective potential approach

to co-rotating and counter-rotating geodesics (see e.g. reference in The Classical Theory of Fields (Landau and Lifshitz, 1980) in the Kerr metric (later renamed as ISCO; see e.g. (Gravitation Misner, Thorne, Wheeler. 1973). In parallel, Frank Zerilli was working on the gravitational radiation emitted by a test particle falling into a Schwarzschild black hole (Zerilli 1970). From this limited conceptual arena Charles Misner and later Kip Thorne launched a program for the detection of gravitational waves on the Earth; see e.g. Misner 1974, Abbott et al. 2016, Abbott et al. 2017; see however Davis et al. 1972, Rodriguez et al. 2018 and J.A. Rueda et al. 2018. A new approach started with the arrival of Demetrios: he was just creating mathematics following his needs. We identified the reversible and irreversible transformations of a Kerr black hole. Wheeler advanced a thermodynamic analogy. I addressed the need to identify the concept of irreducible mass (from the Italian "irriducibile"), and was Demetrios's contribution to integrate, overnight, the differential equation for infinitesimal reversible transformations which led to the finite mass-energy formula for a Kerr black hole. That evening, while walking back home through the IAS woods, I expressed to Wheeler the great relevance of the newly found formula by Demetrios and proposed to let Demetrios be the single author of this article, admiring his great mathematical talent. Wheeler agreed. The Editor of PRL objected since in that two page article Fig. 2 by Wheeler and myself was still unpublished. Actually that Fig. 2 followed a discussion I previously had had with Penrose in Florence (Penrose 1969) which allowed us to present the "Penrose process" there. Some difficulties in achieving this process were obvious from the example in Fig. 2, which Roger later recognized himself (Penrose & Floyd 1971). The Editor finally agreed to our written request and the paper appeared on Sep 17, 1970 (Christodoulou, 1970). In January 1971 my article with Johnny *Introducing the Black Hole* (Ruffini & Wheeler, 1971) appeared with the new physics we were developing in Princeton, including the concept of the "ergosphere". On March 1, 1971 we submitted an article giving the mass formula for the Kerr-Newman metric, including the relation between the surface area of the horizon and the irreducible mass (Christodoulou & Ruffini, 1971) . On March 11, 1971 the same results were independently confirmed by Steven Hawking, extending further the applicability of our formula (Hawking 1971).

The thesis of Demetrios was successfully discussed by a committee including Eugene Wigner (see Fig. 1a), one of the closest collaborators of Albert Einstein and David Wilkinson (See Fig. 1b), the head of the NASA WMAP mission and Johnny and myself as supervisors. The new message was clear: black holes, far from being energy sinks, were energy sources able to emit "in principle" up to 50% of their mass energy as being extractable (Christodoulou & Ruffini, 1971).

Demetrios turned soon to the study of partial differential equations and mathematical physics, to which he dedicated for the rest of his career and results were published in four monographs: (Christodoulou and Klainerman 1994, Christodoulou 2007, Christodoulou 2009, Christodoulou 2019). In 1968, Johnny proposed to Demetrios the collapse of a "geon" composed of massless scalar field as a second

Fig. 1. Demetrios during his thesis presentation with Eugene Wigner (above) and David Wilkinson (below). Myself and Johnny were supervisors, ready to intervene in case of need, but no need of intervention was necessary! Wigner elaborated the aphorism of Niels Bohr "Interesting = wrong" in the most definite "very interesting, if true = totally wrong"

topic for his thesis. It took almost forty years for him to solve this problem, which he extended to the more difficult case of focusing gravitational waves leading to black hole formation (Christodoulou 2009).

A “long march” started on December 12, 1970 with the launch of the Uhuru satellite by Riccardo Giacconi. Early in 1971 an almost daily conversation with him and Herb Gursky at the Smithsonian Astrophysical Observatory led to the discovery of binary X-ray sources. This was soon followed by the announcement of Cygnus X1 identified as the first black hole in our galaxy (Ruffini 1973); see e.g. Gursky & Ruffini 1975 which also contained the first public announcement of the discovery of gamma ray bursts, as well as Giacconi & Ruffini 1980, 2009; see Fig. 2).

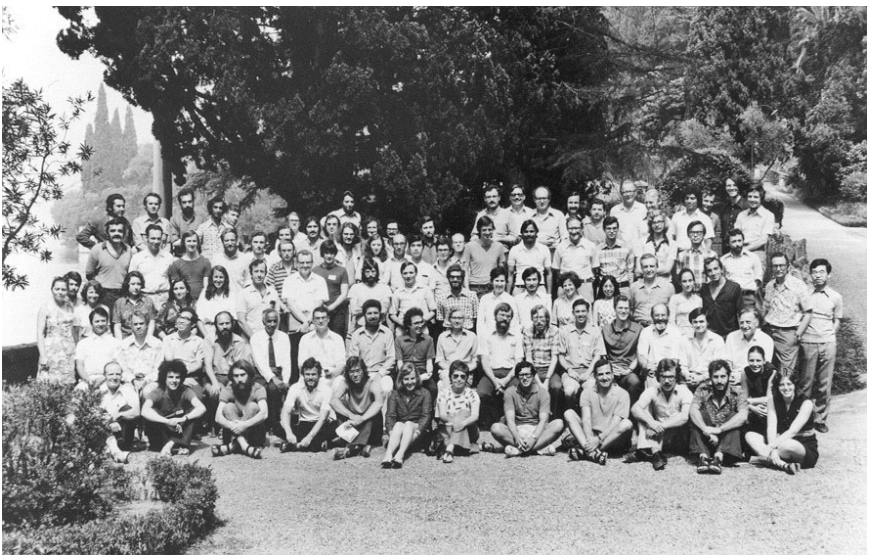


Fig. 2. In the second row, from left to right, there are, among others: E. T. Newman, S. Chandrasekhar, R. Giacconi, R. Ruffini, A. Treves, A. Hewish, D. Arnett, G. I. Taylor, J. Wilson, R. Penrose, as well as J. Bahcall, T. Damour, T. Piran et al.

Today, after fifty years this “long march” has reached a definite result: through the grandest observational multi-wavelength effort in the history of mankind, from space, ground and underground observatories, we are finally finding evidence that black holes are “alive” and their “extractable energy” in our mass formula (Christodoulou & Ruffini, 1971) is the energy source of the most energetic sources in our universe: gamma ray bursts (GRBs), active galactic nuclei (AGNs) as well as ultra-high energy cosmic rays (UHECRs) (Ruffini et al. 2021 and references therein). Their “inner engine” has three independent components: 1) a Kerr black hole which is neither in a stationary state nor in vacuum, 2) a background magnetic field aligned with the black hole rotation axis, and 3) an extremely diluted fully ionized plasma (Moradi et al. 2021). There is no role in this inner engine for ISCOs. Indeed new electrodynamical field equations describe the synchrotron radiation emitted close to the black hole horizon, and they point to a discrete and repetitive emission of “blackholic quanta” in the MeV and in the GeV. The magnitudes and the emission

time scales of these quanta, for M87 and GRB 130427A, are expressed as a function of the above three parameters (Rueda & Ruffini, 2021). A long lasting GeV emission with a luminosity decreasing as a temporal power law allows for the first time for GRBs the determination of the black hole mass and spin as well as their time evolution perfectly fulfilling our mass energy formula (Christodoulou & Ruffini, 1971): a long lasting emission process profoundly different from the traditional process of continued gravitational contraction.

2.2. *Gerard 't Hooft*

Motivation: Professor **Gerard 't Hooft** received the MG16 Individual Award *"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon"*.

Scientific and historical background:Prof. Gerard 't Hooft has been a full Professor at the Utrecht University (the Netherlands), since 1977. Nowadays, he is an Emeritus Professor at that University. During his career, he has paid extended scientific visits to CERN (Geneva), Harvard, Stanford, Princeton and Duke University, NC. In 1999, together with M. Veltman, he received the Nobel Prize in Physics, awarded by The Royal Swedish Academy of Sciences, *"For elucidating the quantum structure of electroweak interactions in physics"*. Prof. 't Hooft main subjects of research includes:

- Gauge Theories for the sub-atomic particles and forces, various aspects and ingredients of what is now called "The Standard Model" of the sub-atomic particles: renormalizability, topological features such as magnetic monopoles and instantons, 1/N expansions,
- Theories for the quantization of the gravitational force and black holes: producing models for the quantum properties of a black hole, as derived from Standard Model and General Relativity alone; its topological features such as antipodal identification,
- Fundamental theories underlying quantum mechanics, in particular returning determinism and reality to the dynamics of the tiniest material entities in his universe, Prof. 't Hooft has been awarded the Wolf Prize of the State of Israel (1982), the Pius XI Medal (Vatican city, 1983), the Lorentz Medal (KNAW Amsterdam, 1986) as well as the Spinoza Premium (Netherlands Organization for Scientific Research NWO, 1995).

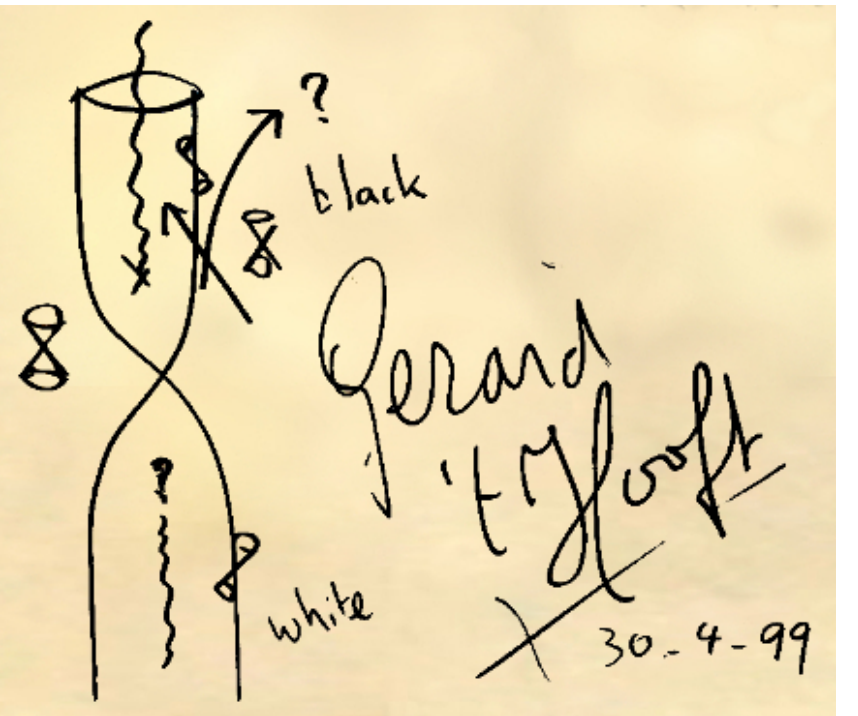


Fig. 3. The signature of Gerard 't Hooft on the wall of ICRA Room 301 (April 4, 1999).

A special event took place at ICRA on April 30, 1999. Prof. Ruffini invited Gerard 't Hooft to Rome to discuss a boundary condition for a quantum field on the black hole horizon, a topic Prof. Ruffini discussed in a previous article "Black-hole evaporation in the Klein-Sauter-Heisenberg-Euler formalism" with Thibault Damour

(Phys. Rev. D 14, 332, 1976), but which needed to be examined in more detail. Prof. Ruffini planned to direct Gerard's attention to some specific aspects of this problem. Because we have traditionally been very attentive in spending ICRA travel funds, ICRA offered Gerard to come to Rome on a reduced fare weekend ticket arriving Friday and departing Monday. He had a great relaxing weekend together with Prof. Ruffini following his seminar, which among other things allowed Gerard to sign the wall in our ICRA Room (see Fig. 3), and during this splendid Rome spring weekend he also was able to find a missing factor of 2 in a formula in Prof. Ruffini's 1971 paper with Demetri Christodoulou on the black hole mass formula. The following October, Gerard received the Nobel prize, which meant that we could no longer get away with bringing him to Rome on a cheap ticket! Ever since Gerard has been in our MG IOC helping us with the preparation of the meetings. We are very happy to announce this MG16 Award to Gerard 't Hooft with the motivating phrase "*for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon*".

2.3. *Tsvi Piran*

Motivation: Professor **Tsvi Piran** received the MG16 Individual Award *"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory"*.

Scientific and historical background: Tsvi Piran is the emeritus Schwartzmann professor at the Hebrew University of Jerusalem. He obtained his PhD in

Physics, in 1976 from the Hebrew University working on the collisional Penrose process. Piran returned to the Hebrew University at 1981after being a post doc at Oxford and Texas and a long term member at the IAS at Princeton. In 1982 he initiated and directed the first ever summer school on Gravitational Waves that took place at Les Houches. Piran was a visiting professor at Harvard, Columbia and New York and a Moore scholar at Caltech. Piran's research deals with numerous aspects of relativistic astrophysics, ranging from the foundation of numerical relativity to modeling of observer relativistic phenomena and analytic work on the fate of gravitational collapse. Piran's research work focuses mostly on black holes and in particular on gamma-ray bursts. He was among the first to point out their cosmological origin and their association with merging neutron stars and heavy r-process nucleo synthesis. Piran's achievements were recognized in the 2019 EMET prize for Physics.

2.4. *Steven Weinberg*

Motivation: Professor **Steven Weinberg** received the MG16 Individual Award *"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory"*.

Scientific and historical background: Steven Weinberg is a member of the Physics and Astronomy Departments at The University of Texas at Austin. His research has covered a broad range of topics in quantum field theory, elementary

particle physics and cosmology. He has been honored with numerous awards, including the Nobel Prize in Physics, the National Medal of Science, the Heinemann Prize in Mathematical Physics and in 2020, the Breakthrough Prize. He is a member of the US National Academy of Sciences, Britain's Royal Society, and other academies in the USA and abroad. The American Philosophical Society awarded him the Benjamin Franklin Medal, with a citation that said he is "considered by many to be the preeminent theoretical physicist alive in the world today." His books for physicists include *Gravitation and Cosmology*, the three-volume work *The Quantum Theory of Fields, Cosmology* and published in April of 2021, *Foundations of Modern Physics*. Educated at Cornell, Copenhagen, and Princeton, he also holds honorary degrees from sixteen other universities. He taught at Columbia, Berkeley, M.I.T., and Harvard, where he was Higgins Professor of Physics, before coming to Texas in 1982.

2.5. *Scientific and historical background for the Tsvi Piran and Steven Weinberg MG16 Awards*

The Sixteenth Marcel Grossmann Meeting (MG16) is a very special one in many respects: it will take place during a pandemic and in spite of the many difficulties, we have decided not to postpone it but to organize it as a virtual meeting. As described on the MG series webpage, these meetings started in 1975 with the first meeting at the International Centre for Theoretical Physics (ICTP) in Trieste (Italy) that I organized with Nobel Prize winner Abdus Salam. A second meeting followed in 1979, with a significantly larger participation including Nobel Laurate Cheng Ning Yang and a Chinese delegation led by Chuo Pei Yuan (see Fig. 4), including Fang Li-Zhi who had accompanied me during my entire first visit to China in 1979. The first truly international MG meeting followed in 1982 in Shanghai (China):this represented an especially important step forward both for the meeting and for China. A multi-millennia *"motto"* in China, which was then proclaimed on banners everywhere, read *"Friends from all over the world are welcomed"*.

We were soon at an impasse over the participation of scientists from Israel, since no diplomatic relations existed between China and Israel at that time and the Israeli scientists were not to be allowed to attend the meeting. A long negotiation began. The boundary conditions were clearly set by Steven Weinberg, a member of the present MG16 IOC: no MG meetings on Einstein's theory of general relativity could occur without the participation of Israeli scientists. The intervention of Yuval Ne'emann, also a member of the MG IOC then as well as the Minister of Science of Israel (see Fig. 5), proposed a compromise that would admit at least one Israeli scientist.

I went to Beijing alone, meeting every morning for a week with 12 Chinese representatives led by Chuo Pei Yuan going over all possible options. I stayed in an isolated villa not far from Tiananmen Square, accompanied by the 3 volumes of Matteo Ricci (RI MA TO) to keep me company. No solution was in sight the

Fig. 4. Chuo Pei Yuan and Cheng Ning Yang at MG2 in Trieste, Italy (1979).

entire week. At the last moment, just before my departure, an agreement was finally reached allowing two Israeli scientists into China. The historic compromise would admit Gerard Tauber and Tsvi Piran into China using a special ICRA travel document I had proposed for them to be able to participate in the meeting, accepted by the Chinese Ambassador in Rome. This modified the thousand year Chinese *"motto"* to read *"Scientists from all over the world are welcomed"*. The event was extremely beneficial for China and signaled the truly international nature of the MG meetings.

I kept on meeting Tauber in the years which followed (see Fig. 6). Soon after, Yuval Ne'emann visited China. The development of bilateral relations, including military cooperation and economical tights, grow exponentially until the establishment of normal diplomatic relations between Israel and China in 1992.

Given their key role played in the foundations of the MG meetings, I am very happy to propose on behalf of the MG16 IOC, two special Marcel Grossmann Individual Awards: one to Steven Weinberg for *"for unwavering support for the MG*

Fig. 5. From right to left: Chaim Weizmann, President of Israel; Yuval Ne'emann, Minister of Science of Israel; R. Ruffini.

meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory" and another one to Tsvi Piran, "*for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory*", in the words of John A. Wheeler's photo dedication to myself (see Fig. 7).

3. MG16 Institutional Awardees

The MG16 Institutional Awards to the Spektrum-Roentgen-Gamma (SRG) mission *"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance"* went to the S.A. Lavochkin Association (presented to its Designer General Alexander Shirshakov), to the Max Planck Institute for Extraterrestrial Physics - MPE (presented to Professor Peter Predehl, Principal Investigator of eROSITA) and to the Space Research Institute (IKI) of the Russian Academy of Sciences (presented to Professor Rashid Sunyaev, Principal Investigator of SRG Observatory in Russia).

Fig. 6. From right to left: Arrigo Finzi, Remo Ruffini, Gerard Tauber and Konrad Bleuler.

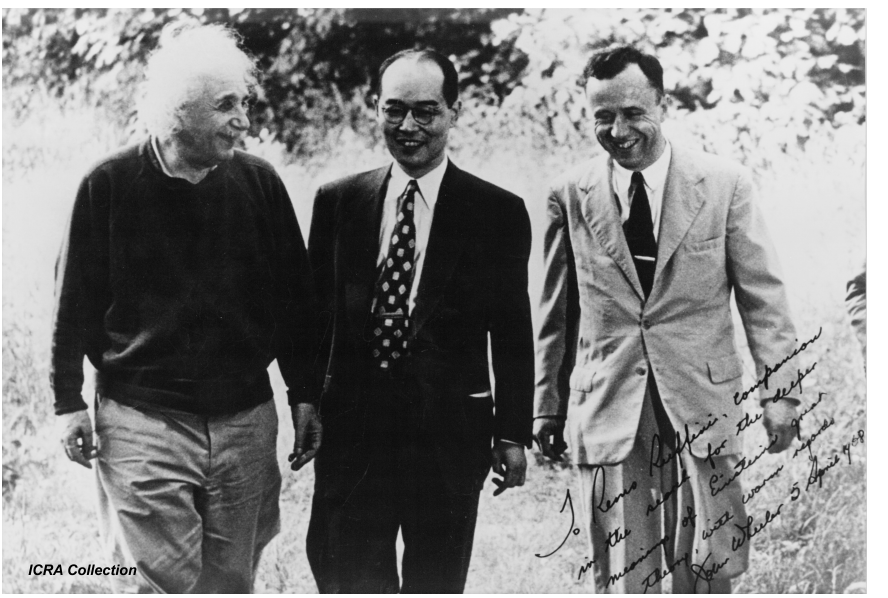


Fig. 7. Albert Einstein, Hideki Yukawa and John. A. Wheeler with a hand-written dedication to Remo Ruffini "To Remo Ruffini, companion in the search for the deeper meaning of Einstein great theory. With warm regards, John Wheeler 5 April 1968".

On Tuesday June 29, 2021, the following 31 Astro-Ph appeared:

(1) https://arxiv.org/abs/2106.14517
(2) https://arxiv.org/abs/2106.14518
(3) https://arxiv.org/abs/2106.14519
(4) https://arxiv.org/abs/2106.14520
(5) https://arxiv.org/abs/2106.14521
(6) https://arxiv.org/abs/2106.14522
(7) https://arxiv.org/abs/2106.14523
(8) https://arxiv.org/abs/2106.14524
(9) https://arxiv.org/abs/2106.14525
(10) https://arxiv.org/abs/2106.14526
(11) https://arxiv.org/abs/2106.14527
(12) https://arxiv.org/abs/2106.14528
(13) https://arxiv.org/abs/2106.14529
(14) https://arxiv.org/abs/2106.14530
(15) https://arxiv.org/abs/2106.14531
(16) https://arxiv.org/abs/2106.14532
(17) https://arxiv.org/abs/2106.14533
(18) https://arxiv.org/abs/2106.14534
(19) https://arxiv.org/abs/2106.14535
(20) https://arxiv.org/abs/2106.14536
(21) https://arxiv.org/abs/2106.14537
(22) https://arxiv.org/abs/2106.14541
(23) https://arxiv.org/abs/2106.14542
(24) https://arxiv.org/abs/2106.14543
(25) https://arxiv.org/abs/2106.14544
(26) https://arxiv.org/abs/2106.14545
(27) https://arxiv.org/abs/2106.14546
(28) https://arxiv.org/abs/2106.14547
(29) https://arxiv.org/abs/2106.14548
(30) https://arxiv.org/abs/2106.14549
(31) https://arxiv.org/abs/2106.14550

3.1. *S.A. Lavochkin Association*

The MG16 Institutional Award to the S.A. Lavochkin Association has been presented to its Designer General, **Alexander Shirshakov**.

S.A. Lavochkin Association created the Navigator space platform carrying German eRosita and Russian ART-XC X-Ray Telescopes, organized the launch of SRG Orbital X-Ray Observatory to the second Lagrangian point of the Sun-Earth system at a distance of 1.5 million km from the Earth and managed the observatory flight and the daily reception of its scientific data on Earth for 23.5 months.

Dr Alexander Shirshakov, Designer General of the S.A. Lavochkin Association, is specialized in design, manufacture, testing, launch and control of S/C for scientific purposes. Among those S/C launched, there are the *Radiostron* Astrophysical Observatory (2011) and the *Spektr-RG* space observatory (2019), while the planned S/C launches are *Luna-25* and *Exomars*.

Dr Shirshakov started his career in 1973, working as an engineer of the State Unitary Enterprise *NPO named by S.A. Lavochkin* in Khimki (Russian Federation). Starting from 1989 he has played multiple roles within the Lavochkin Association, been appointed head of the group, head of the sector, head of department, deputy head of the complex, head of the branch, director of the center, deputy head of the Design Bureau, deputy General Designer and deputy General Director.

Dr Shirshakov is an editorial board Member of the reviewed edition of *Vestnik of Lavochkin Association*. Since 2017, he is also member of the General Designer council. He has been awarded Honored Mechanical engineer of the Russian Federation as well as Agency-level award of the Russian Federal Space Agency.

3.2. *Max Planck Institute for Extraterrestrial Physics (MPE)*

The MG16 Institutional Award to the Max Planck Institute for Extraterrestrial Physics (MPE) has been presented to Professor Peter Predehl, the Principal Investigator of eROSITA.

eROSITA is the soft X-ray telescope on-board the Russian-German Spektr-RG mission which was successfully launched from Baikonur on July 13, 2019 and placed

in a halo orbit around the L2 point. 30 years after ROSAT, eROSITA performs an all-sky survey with an unprecedented sensitivity, spectral and angular resolution. Clusters of galaxies are the largest collapsed objects in the Universe. Their formation and evolution is dominated by gravity, i.e. Dark Matter, while their large scale distribution and number density depends on the geometry of the Universe, i.e. Dark Energy. X-ray observations of clusters of galaxies provide information on the rate of expansion of the Universe, the fraction of mass in visible matter, and the amplitude of primordial fluctuations which are the origin of clusters of galaxies and the whole structure of the universe. eROSITA has been designed to detect at least 100.000 clusters of galaxies and to detect systematically more than 3 Million obscured accreting Black Holes. eROSITA will also allow to study the physics of galactic X-ray source populations, like pre-main sequence stars, supernova remnants and X-ray binaries. The eROSITA telescope consists of seven identical Wolter-1 mirror modules. A novel detector system has been developed by MPE on the basis of the successful XMM-Newton pn-CCD technology. MPE is the scientific lead institute of eROSITA, responsible for the development of the instrument, the operation, the analysis software and data archive. Peter Predehl led this development as Principal Investigator of eROSITA and German lead scientist of the SRG mission for more than 15 years until the completion of the first of eight surveys in 2020. At this time eROSITA has already discovered more than 1 Million X-ray sources, more than all X-ray observatories of the last 50 years together. This demonstrates, that the design goals of the mission will easily be fulfilled.

3.3. *Space Research Institute (IKI) of the Russian Academy of Sciences*

The MG16 Institutional Award to the Space Research Institute (IKI) of the Russian Academy of Sciences has been presented to Professor Rashid Sunyaev, the Principal Investigator of SRG Observatory in Russia.

The Space Research Institute (IKI) of the Russian Academy of Sciences was responsible for developing the overall concept and scientific program of the SRG Orbital observatory and played a leading role in developing the ART-XC telescope and the entire SRG observatory as part of the Russian space science program carried out by Roskosmos Corporation in the interests of the Russian Academy of Sciences.

During the flight to the L2 point of the Sun-Earth system, SRG with German (eRosita) and Russian (ART-XC named after Mikhail Pavlinsky) X-Ray Telescopes aboard performed calibrations and long duration Performance Verification observations of a dozen of targets and deep fields. Starting in the middle of December 2019, the SRG scanned the whole sky three times. During these scans, SRG discovered two million point X-Ray sources: mainly quasars, stars with hot and bright coronae, and more than 30 thousand clusters of galaxies. There is a competition and synergy in the search for clusters of galaxies between SRG and the ground-based Atacama Cosmology and South Pole Telescopes, which are searching for clusters of galaxies in microwave spectral band using Sunyaev-Zeldovich effect.

SRG provided the X-Ray map of the whole sky in hard and soft bands, the last is now the best among existing. The huge samples of the X-ray selected quasars at the redshifts up to z=6.2 and clusters of galaxies will be used for well-known cosmological tests and detailed study of the growth of the large scale structure of the Universe during and after reionization.

SRG/eRosita is discovering every day several extragalactic objects which increased or decreased their brightness more than 10 times during half of the year after the previous scan of the same one-degree wide strip on the sky. A significant part of these objects has observational properties similar to the Events of Tidal Disruption of a star orbiting in the vicinity of the supermassive black hole. ART-XC discovered a lot of bright galactic and extragalactic transients.

Rashid Sunyaev is the Principal Investigator of SRG mission in Russia, director-emeritus of the Max-Planck Institute for Astrophysics and Maureen and John Hendricks distinguished visiting professor of the Institute for Advanced Study, Princeton.

Fig. 8. The SRG space mission.

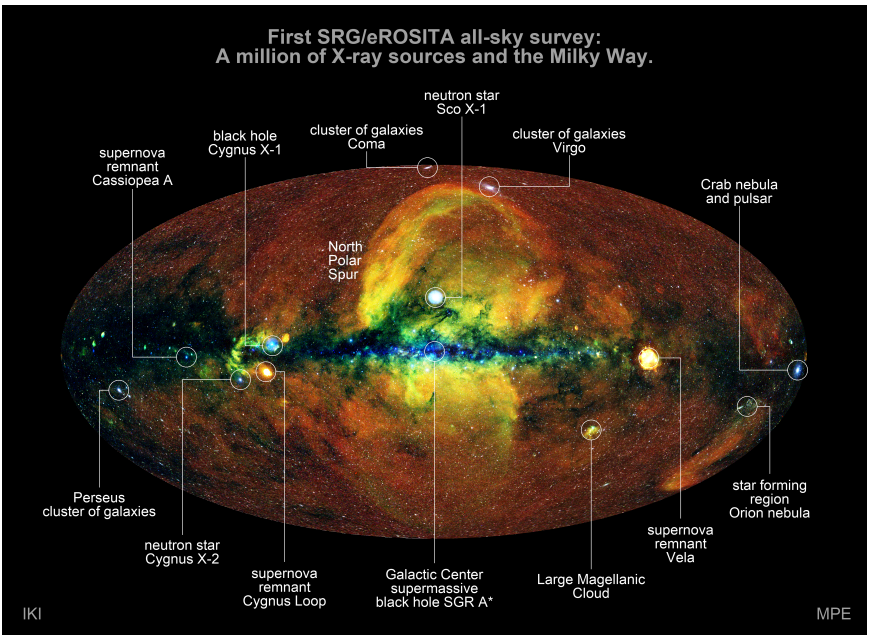


Fig. 9. First SRG/eROSITA all-sky survey, a million of X-ray sources and the Milky Way.

Acknowledgments

It gives me a great pleasure to thank the members of the International Organising Committee (IOC) of the MG meetings as well as members of the scientific and academic communities at large, for discussion and advices.

References

1. Landau, L. D., Lifshitz, E. M., The Classical Theory of Fields, 4th Edition, *Butterworth-Heinemann*, 1980
2. Misner, Charles W., Thorne, Kip S., Wheeler, John A., Gravitation, *San Francisco: W.H. Freeman and Co.*, 1973
3. F. Zerilli, Gravitational Field of a Particle Falling in a Schwarzschild Geometry Analyzed in Tensor Harmonics, *Phys. Rev. D 2, 2141*, 1970
4. Misner, C. W., Mechanisms for the Emission and Absorption of Gravitational Radiation, *Proceedings of the IAU Symposium, No. 64* 1974
5. Abbott B.P. et al., Improved Analysis of GW150914 Using a Fully Spin-Precessing Waveform Model, *Phys. Rev. X 6, 041014*, 2016
6. Abbott B.P. et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett. 119, 161101*, 2017
7. Davis M. et al., Can Synchrotron Gravitational Radiation Exist?, *Phys. Rev. Lett. 28, 1352*, 1972
8. Rodriguez et al., On the Final Gravitational Wave Burst from Binary Black Holes Mergers, *Astron. Rep. 62, 940–952*, 2018

9. Rueda J.A. et al., GRB 170817A-GW170817-AT 2017gfo and the observations of NS-NS, NS-WD and WD-WD mergers, *JCAP10 006*, 2018
10. Penrose R., Gravitational Collapse: the Role of General Relativity, *Rivista del Nuovo Cimento, Numero Speziale I, 252*, 1969
11. Penrose, R., Floyd, R. Extraction of Rotational Energy from a Black Hole, *Nature Physical Science 229*, 1971
12. Christodoulou D., Reversible and Irreversible Transformations in Black-Hole Physics, *Phys. Rev. Lett. 25, 1596*, 1970
13. Ruffini R., Wheeler J. A., Introducing the black hole, *Physics Today 24, 1, 30*, 1971
14. Christodoulou D., Ruffini R., Reversible Transformations of a Charged Black Hole, *Phys. Rev. D 4, 3552*, 1971
15. Hawking S.W., Gravitational Radiation from Colliding Black Holes, *Phys. Rev. Lett. 26, 1344*, 1971
16. Christodoulou D., Klainerman S., The Global Nonlinear Stability of the Minkowski Space (PMS-41), *Princeton Legacy Library*, 1994
17. Christodoulou D., The Formation of Shocks in 3-Dimensional Fluids, *EMS Monographs in Mathematics, European Mathematical Society*, 2007
18. Christodoulou D., The Formation of Black Holes in General Relativity, *EMS Monographs in Mathematics, European Mathematical Society*, 2009
19. Christodoulou D., The Shock Development Problem, *EMS Monographs in Mathematics, European Mathematical Society*, 2019
20. Ruffini R., Neutron stars and Black Holes in our Galaxy, *New York Academy of Sciences*, 1973
21. Gursky, H., Ruffini, R., Neutron Stars, Black Holes and Binary X-Ray Sources, *Springer Netherlands*, 1975
22. Giacconi, R.; Ruffini, R., Physics and astrophysics of neutron stars and black holes, *Bologna, Societa Italiana di Fisica; Amsterdam, North Holland Publishing Co.*, 1980
23. Rees M., Ruffini R., Wheeler J.A., Black Holes, Gravitational Waves and Cosmology, *CRC*, 2004
24. Ruffini R. et al., The morphology of the X-ray afterglows and of the jetted GeV emission in long GRBs, *Monthly Notices of the Royal Astronomical Society, Volume 504, Issue 4*, 2021
25. Rueda, J.A., Ruffini, R. The blackholic quantum, *Eur. Phys. J. C 80, 300*, 2021

The white dwarf binary merger model of GRB 170817A*

J. A. Rueda[†,‡,§,¶,††,¶¶], R. Ruffini[†,‡,‖,**], Liang Li[†,‡,‡‡], R. Moradi[†,‡,‡‡], N. Sahakyan[§§] and Y. Wang[†,‡,‡‡]

[†]*ICRANet, Piazza della Repubblica 10, I-65122 Pescara, Italy*

[‡]*ICRA, Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy*

[§]*ICRANet-Ferrara, Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, I–44122 Ferrara, Italy*

[¶]*Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, I–44122 Ferrara, Italy*

[‖]*Université de Nice Sophia-Antipolis, Grand Château Parc Valrose, Nice, CEDEX 2, France*

[**]*INAF, Viale del Parco Mellini 84, 00136 Rome, Italy*

[††]*INAF, Istituto di Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, 00133 Rome, Italy*

[‡‡]*INAF – Osservatorio Astronomico d'Abruzzo, Via M. Maggini snc, I-64100 Teramo, Italy*

[§§]*ICRANet-Armenia, Marshall Baghramian Avenue 24a, Yerevan 0019, Republic of Armenia*

[¶¶]*jorge.rueda@icra.it*

Following the GRB 170817A prompt emission lasting a fraction of a second, 10^8 s of data in the X-rays, optical, and radio wavelengths have been acquired. We here present a model that fits the spectra, flux, and time variability of all these emissions, based on the thermal and synchrotron cooling of the expanding matter ejected in a binary white dwarf merger. The $10^{-3} M_\odot$ of ejecta, expanding at velocities of $10^9\,\mathrm{cm\,s^{-1}}$, are powered by the newborn massive, fast rotating, magnetized white dwarf with a mass of $1.3 M_\odot$, a rotation period of $\gtrsim 12$ s, and a dipole magnetic field $\sim 10^{10}$ G, born in the merger of a $1.0 + 0.8 M_\odot$ white dwarf binary. Therefore, the long-lasting mystery of the GRB 170817A nature is solved by the merger of a white dwarf binary that also explains the prompt emission energetics.

Keywords: Gamma-ray bursts; white dwarfs; white dwarf mergers.

1. Introduction

GRB 170817A is a short gamma-ray burst (GRB) whose prompt emission lasts less than a second, as was detected by the gamma-ray burst monitor (GBM) onboard

*Based on a talk presented at the Sixteenth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories, online, July 2021.

the NASA Fermi Gamma-ray Space Satellite,[1,2] and confirmed by *INTEGRAL*.[3] It was subsequently associated with GW170817, a gravitational wave signal reported by the LIGO/Virgo Collaboration about 40 min after the Fermi-GBM circular.[4] These initial data were then associated with the optical-infrared-ultraviolet source AT 2017gfo, which started to be observed about 12 h ($\approx 4 \times 10^4$ s) after the GRB trigger.[5–8] Further data of GRB 170817A have been in the mean time acquired in the X-rays and in the radio from 10^6 s after the GRB trigger, and still ongoing.

It has been well established that short GRBs are produced by neutron star binary (NS-NS) mergers.[9–11] Therefore, it is not surprising that GRB 170817A was labeled as such from the very beginning,[1,4,12] despite the fact that it had been soon recognized that GRB 170817A was observationally very different from typical short GRBs.[2] Indeed, a comparison of GRB 170817A in the gamma-rays, X-rays and in the optical with typical short GRBs led[13] to suggest that GRB 170817A looks more like a white dwarf binary (WD-WD) merger rather than an abnormal, special or unique NS-NS merger.[14] Identified additional sources similar to GRB 170817A and have proposed an alternative interpretation of them as WD-WD mergers.

In the mean time, 10^8 s of data of GRB 170817A have been acquired in the X-rays, in the optical, and in the radio wavelengths, besides just the MeV radiation of the prompt emission. These observations have indeed led to alternative explanations. In fact

- The NS-NS merger interprets the associated optical counterpart AT 2017gfo as a nuclear *kilonova* produced by the decay of r-process, which yields in the matter ejected in the merger.[5–8]
- The experimental confirmation of the nuclear *kilonova* needs a univocal spectroscopic identification of the atomic species present in the ejecta.[15–19] This has not been achievable in view of lack of available accurate models of atomic spectra, the nuclear reaction network, density profile, and details of the radiative transport (opacity). Other mechanisms can also explain the photometric properties of AT 2017gfo, for instance the cooling of the expanding ejecta of a WD-WD merger.[13,20] We will further elaborate this scenario in this paper.
- The NS-NS merger leading to a jet propagating throughout the ejected matter appears in conflict with recent data by the *Chandra* X-ray Telescope at 10^7–10^8 s after the GRB trigger.[21,22]

In view of all the above, we here explore further and extend the suggestion by Ref. 13 of GRB 170817A being the product of a WD-WD merger, adding new observations all the way up to 10^8 s.

- The possibly observed re-brightening in the X-ray afterglow of GRB 170817A at 1000 days agrees with the predicted appearance of the pulsar-like activity of the newborn WD from a WD-WD merger.[13,20]

- The rate of GRB 170817A-like events is well explained by the rate of WD-WD mergers.[13,20]
- Interestingly, the host galaxy of GRB 170817, NGC 4993 distant at about 40 Mpc, is an old elliptical galaxy.[1] Elliptical old galaxies are amply recognized as preferred sites of type Ia supernovae produced by the so-called double-degenerate scenario, namely, by WD-WD mergers.[23,24]

The aim of this paper is to extend the treatment of Ref. 20 on WD-WD mergers, and exploit the analogy with the synchrotron emission in the X-rays, optical and radio bands in the afterglow of long GRBs[25–28] to determine the emission of WD-WD mergers across the electromagnetic spectrum. Then, we apply the above considerations to the luminosity in the X-rays, optical and radio wavelengths observed in the afterglow of GRB 170817A.

We here show the prominent role of rotation and its effect on the synchrotron emission from the interaction of the newborn rotating object with the ejected matter in the merger. This process is energetically predominant and has been neglected in traditional simulations of these merging systems. The ejected matter expands in the magnetic field of the newborn fast rotating WD, which injects rotational and accretion energy into the expanding ejecta. While expanding, the ejecta radiate energy across the electromagnetic spectrum due to thermal cooling and synchrotron emission. We evidence that the newborn WD becomes observable as a pulsar when the synchrotron radiation fades off. The amount of mass ejected, the mass, rotation period, and strength of the magnetic field of the newborn WD are the most important features that determine the electromagnetic emission of the system.

We show that the above process leads to a hard-to-soft evolution of the emitted radiation with specific decreasing luminosities that approach a distinct power-law behavior. The late-time luminosity is dominated by the pulsar activity of the newborn object, therefore the asymptotic power-law gives information on the parameters of the newborn central object. The total energy radiated during the whole evolution is dominated by the energy injected and radiated from the central WD, so it is covered by its rotational energy.[29] Energy and angular momentum conservation allow to infer, for instance, the spin and magnetic field of the newborn WD directly from the light-curve of the source, prior to any detailed fit of the observational data with the theoretical model (see Refs. 26–28, for the case of long GRBs).

We apply the above considerations to GRB 170817A and show the agreement of the WD-WD merger scenario with all the available observational multiwavelength data from the gamma-rays all the way down to the radio wavelengths. This paper is organized as follows. In Sec. 2, we formulate the general physical conditions of the WD-WD coalescence that constrain the parameters of the newborn WD formed at merger. Section 3 presents an estimate of a possible mechanism leading to a gamma-ray prompt emission in these mergers, and how it compares with GRB

1780817A. Section 4 is devoted to the analysis of the WD-WD post-merger early optical-infrared-ultraviolet emission by thermal cooling, and how it compares with AT 2017gfo. In Sec. 5, we present the theoretical model of the synchrotron emission powered by the newborn WD, and how it leads to a multiwavelength emission (from the radio to the gamma-rays). A comparison with the emission of GRB 170817A $t \gtrsim 10^6\,\mathrm{s}$ is presented. We outline our conclusions in Sec. 6. We use cgs units throughout.

2. Merging Binary and Post-Merger Remnant

The fate of the central remnant of a WD-WD merger with a total mass near (below or above) the Chandrasekhar mass limit can be one of the three possibilities: (i) a stable newborn WD, (ii) a type Ia supernova, or (iii) a newborn neutron star. Sub-Chandrasekhar remnants can lead either to (i) and (ii), while super-Chandrasekhar remnants produce either (ii) or (iii). Super-Chandrasekhar remnants are supported by angular momentum, so they are less dense and metastable objects whose final fate is delayed until the excess of angular momentum is loss, e.g. via magnetic braking, inducing its compression.[30,31]

We are here interested in WD-WD mergers leading to stable, massive, sub-Chandrasekhar newborn WDs with a mass $\gtrsim 1.0 M_\odot$. These WDs can have rotation periods as short as $\sim 0.5\,\mathrm{s}$ (see Ref. 32) and can also avoid the trigger of unstable burning leading to type Ia supernova providing its central density is kept under some critical value of a few $10^9\,\mathrm{g\,cm^{-3}}$.[31]

Numerical simulations of WD-WD mergers show that the merged configuration has in general three distinct regions:[31,33–39] a rigidly rotating, central WD, on top of which there is a hot, convective corona with differential rotation, surrounded by a rapidly rotating Keplerian disk. Roughly, half of the mass of the secondary star, which is totally disrupted, goes to the corona while the other half goes to the disk. The above implies that little mass is ejected in the merger. Numerical simulations show that the amount of expelled mass is approximated by Ref. 39

$$m_{\rm ej} \approx h(q)\, M, \quad h(q) = \frac{0.0001807}{-0.01672 + 0.2463q - 0.6982q^2 + q^3}, \tag{1}$$

where

$$M = m_1 + m_2 = \left(\frac{1+q}{q}\right) m_2, \tag{2}$$

is the total binary mass, and $q \equiv m_2/m_1 \leq 1$ is the binary mass ratio. Equation (1) tells us that for a fixed total binary mass, the larger the mass symmetry, the smaller the mass that is ejected. Thus, for a fully symmetric mass ratio, $q = 1$, the amount of expelled matter becomes $m_{\rm ej} \approx 3.4 \times 10^{-4} M$.

WD-WD merger simulations show two important ingredients for our model. First, the central remnant (the newborn WD) is degenerate, namely, massive

($\gtrsim 1.0 M_\odot$), fast rotating, and magnetized.[31] Second, although the amount of expelled matter is negligible with respect to the total mass of the system, the ejecta are crucial for the electromagnetic emission in the post-merger evolution.

We start with a double WD with components of mass m_1 and m_2, with corresponding radii R_1 and R_2. We shall make use of the analytic mass-radius relation[40]

$$\frac{R_i}{R_\odot} = \frac{0.0225}{\bar{\mu}} \frac{\sqrt{1-(m_i/M_{\rm crit})^{4/3}}}{(m_i/M_{\rm crit})^{1/3}}, \tag{3}$$

where $\bar{\mu} \approx 2$ is the molecular weight, and

$$M_{\rm crit} \approx \frac{5.816 M_\odot}{\bar{\mu}^2} \approx 1.4\ M_\odot, \tag{4}$$

is the critical mass of (carbon) WDs, $M_\odot$ and $R_\odot$ are the solar mass and radius. Since little mass is expelled, we estimate the newborn WD mass as

$$m_{\rm wd} \approx M - m_d = m_1 + m_2 - m_d \approx \left(\frac{2+q}{2\,q}\right) m_2, \tag{5}$$

where we have approximated the disk mass by $m_d \approx m_2/2$, according to numerical simulations. Combining Eqs. (2) and (5), we obtain

$$M \approx 2 \left(\frac{1+q}{2+q}\right) m_{\rm wd}, \tag{6}$$

and using Eqs. (1) and (6), we obtain

$$m_{\rm wd} \approx \left(\frac{2+q}{1+q}\right) \frac{m_{\rm ej}}{2\,h(q)}. \tag{7}$$

As we shall see in Sec. 6, the above equations allow us to infer, from the inferred mass of the ejecta from the fit of the multiwavelength data of GRB 170817A, the parameters of the merging components and of the newborn WD.

3. The Prompt γ-Ray Emission

GRB 170817 was first detected by the GBM on board the *Fermi* satellite.[2] The gamma-ray emission was confirmed by *INTEGRAL*.[3]

GRB 170817A is as a short burst with a duration (T_{90}) of 2.048 s, as reported in the NASA/HEASARC database.[a] We performed a Bayesian spectral analysis of the *Fermi*-GBM data by using the Multi-Mission Maximum Likelihood Framework (3ML, see Ref. 41), and the best model is selected by comparing the deviance information criterion (DIC, see Refs. 42, 43). We first fit the data with a single power-law function, and obtained a DIC value of 3138. We then compare this model to the blackbody (Planck) spectrum over the same time interval, and obtained a DIC value of 3146. We also fit the data with a Comptonized (i.e. a cutoff power-law,

[a] https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html.

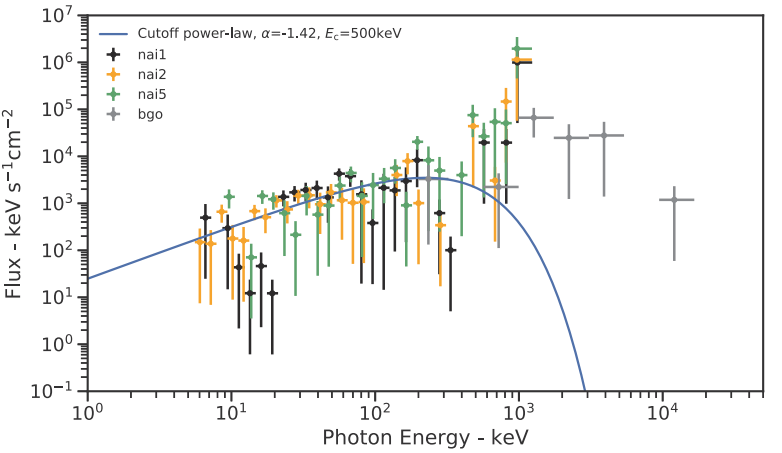


Fig. 1. Spectral fits of νF_ν spectrum for the entire pulse (-0.320 to $1.984\,$s) of the *Fermi*-GBM observation of GRB 170817A. This time interval is the best fit with a Comptonized function, with a cutoff energy $E_c = 500 \pm 317\,$keV, $\alpha = -1.42 \pm 0.18$, and time-averaged flux is $(1.84 \pm 0.82) \times 10^{-7}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ (see Sec. 3 for details of the data analysis).

hereafter CPL) function, and obtained a DIC value of 3128. The CPL model leads to a DIC improvement of 10 with respect to the power-law model, and of 18 with respect to the blackbody model, which suggests the CPL as the model that best fits the data. We refer to Refs. 44–48 for a detailed Bayesian analysis of the data and the reduction procedure applied to GRBs.

As discussed above, the entire pulse (-0.320 to $1.984\,$s) is best fitted by a CPL with a cutoff energy $E_c = 500 \pm 317\,$keV and power-law index $\alpha = -1.42 \pm 0.18$ (see Fig. 1). The time-averaged flux is $(1.84 \pm 0.82) \times 10^{-7}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$. With the measured cosmological redshift of $z = 0.009783$, corresponding to a source distance of $\approx 43\,$Mpc, the isotropic energy released in this time interval is estimated to be $(4.16^{+3.15}_{-1.84}) \times 10^{46}\,$erg. The nonthermal energy released at energies above 1 MeV corresponds to only 2.82% of the emission corresponding to $\approx 1.17 \times 10^{45}\,$erg. Therefore, most of the energy is released below MeV energies, which corresponds to $\approx 4.04 \times 10^{46}\,$erg.

We here advance the possibility that the γ-ray prompt emission of GRB 170817A occurs from activity in the merged magnetosphere. We could think of the WD pulsar magnetosphere in an analogous way as the NS pulsar magnetosphere, therefore the presence of the strong magnetic field and rotation produces the presence of a electric field by Faraday (unipolar) induction.[49] Numerical simulations show that the merger forms a transient hot corona with temperatures 10^8–$10^9\,$K that cools down rapidly mainly by neutrino emission.[31] Therefore, thermal production of e^+e^- pairs can occur for short time before it cools below the pair formation energy threshold. The charged particles are accelerated by the electric field to then follow the magnetic field lines generating both curvature and synchrotron photons. Since the magnetic

field lines are curved, photon–photon collisions occur roughly in all directions, so the majority of the photons with energy in excess of $m_e c^2$ can decay into pairs again and generate a thermal plasma. A minority of photons escape along the rotation axis (see below), leading to the observed nonthermal emission above 1 MeV.

The cross-section of the $\gamma\gamma \to e^- e^+$ process is given by

$$\sigma_{\gamma\gamma} = \frac{3\sigma_T}{16}(1-\bar{\beta}^2)\left[2\bar{\beta}(\bar{\beta}^2-2) + (3-\bar{\beta}^4)\ln\left(\frac{1+\bar{\beta}}{1-\bar{\beta}}\right)\right], \tag{8}$$

where $\sigma_T \approx 6.65 \times 10^{-25}\,\mathrm{cm}^2$ is the Thomson cross-section, and $\bar{\beta}$ is the e^- (or e^+) velocity (in units of c) in the center of momentum frame

$$\bar{\beta} \equiv \sqrt{1 - \frac{2}{\bar{\epsilon}_{\rm inc}\bar{\epsilon}_{\rm tgt}(1-\cos\theta)}}, \tag{9}$$

being $\bar{\epsilon}_{\rm inc,tgt} \equiv \epsilon_{\rm inc,tgt}/(m_e c^2)$ the normalized energy of the incident and target photons which collide making an angle θ measured in laboratory frame.

Photons emitted along the curved magnetic field lines are expected to be absorbed since they will be radiated nearly isotropically. In this case, $\langle\cos\theta\rangle \sim 0$ and the cross-section becomes maximal at $\bar{\epsilon}_{\rm inc}\bar{\epsilon}_{\rm tgt} \approx 4$, and $\sigma_{\gamma\gamma} \approx \sigma_T/4$. Under these conditions, the $\gamma\gamma$ optical depth is

$$\tau_{\gamma\gamma} \approx n_{\rm tgt}\sigma_{\gamma\gamma} r \approx \frac{L_{\rm tgt}\sigma_{\gamma\gamma}}{4\pi\, r\, c\, \epsilon_{\rm tgt}} \approx \frac{L_{\rm tgt}\,\bar{\epsilon}_{\rm inc}\,\sigma_T}{64\pi\, r\, m_e c^3}, \tag{10}$$

where r is the source size and $n_{\rm tgt}$ is the density of target photons, which we have estimated as $n_{\rm tgt} \approx L_{\rm tgt}/(4\pi\, r^2\, c\, \epsilon_{\rm tgt})$, where $L_{\rm tgt}$ is the luminosity emitted at energies larger than the target photon energy.

For a transient hot corona, most photons are emitted at energies around the peak of the Planck spectrum, which for a temperature of a few 10^9 K implies $\epsilon_{\rm inc} \sim \epsilon_{\rm tgt} \sim 3{\rm kT} \sim 1\,{\rm MeV}$. Assuming a source size $r \sim R_{\rm wd} \sim 10^9\,{\rm cm}$, and a target luminosity $L_{\rm tgt} \sim 4\pi R_{\rm wd}^2 \sigma T^4 \sim 10^{51}\,{\rm erg\,s^{-1}}$, the optical depth (10) $\tau_{\gamma\gamma} \sim 10^{10}$.

The above conditions imply that most photons interact generating an optically thick pair plasma which explains the dominant blackbody component observed by *Fermi*-GBM. The observed nonthermal component is explained if $\approx 1\%$ of the photons escape from the system, which can occur near the rotation axis of the WD. There, the interaction angle could approach values as small as $\cos\theta \sim 1$, thereby reducing drastically the photon–photon cross-section.

4. Thermal Cooling of the Ejecta as Origin of the Kilonova

The second observed emission associated with GRB 170817A is the optical counterpart at about 0.5 d after the *Fermi*-GBM trigger, i.e. AT 2017gfo.[7,8,50,51] For the modeling of this thermal emission of the expanding ejecta, we must take into account that in a nonhomogeneous distribution of matter, the layers reach transparency at

different times. For simplicity, we consider the ejected matter as a spherically symmetric distribution extending at radii $r_i \in [R_*, R_{\rm max}]$, with corresponding velocities $v_i \in [v_*, v_{\rm max}]$, in self-similar expansion

$$r_i(t) = r_{i,0}\hat{t}^n, \quad v_i(t) = n\frac{r_i(t)}{t} = v_{i,0}\hat{t}^{n-1}, \tag{11}$$

where $\hat{t} \equiv t/t_*$, being $t_* \equiv nR_{*,0}/v_{*,0}$ the characteristic expansion timescale, which is the same for all layers in view of the condition of self-similarity. Here, $r_{i,0}$ and $v_{i,0}$ are the initial radius and velocity of the layer (so at times $t \ll t_*$ close to the beginning of the expansion. The case $n = 1$ corresponds to a uniform expansion.

The density at the position $r = r_i$ is given by

$$\rho(r_i) = \frac{(3-m)}{4\pi}\frac{m_{\rm ej}}{R_{*,0}^3}\left[\left(\frac{R_{\rm max}}{R_*}\right)^{3-m} - 1\right]^{-1}\left(\frac{r_i}{R_*}\right)^{-m}\hat{t}^{-3n}, \tag{12}$$

where $m_{\rm ej}$ is the total mass of the ejecta, and m is a positive constant. The distribution and time evolution given by Eq. (12) ensure that at any time the total mass of the ejecta, i.e. the volume integral of the density, is always equal to $m_{\rm ej}$.

We divide the ejecta into N shells defined by the $N + 1$ radii

$$r_{i,0} = R_{*,0} + i\frac{(R_{\rm max,0} - R_{*,0})}{N}, \quad i = 0, 1, ..., N, \tag{13}$$

so the width and mass of each shell are, respectively, $\Delta r = (R_{\rm max,0} - R_{*,0})/N$, and

$$m_i = \int_{r_i}^{r_{i+1}} 4\pi r^2\rho(r)dr \approx \frac{4\pi}{m-3}r_i^2\rho(r_i)\Delta r, \tag{14}$$

so in view of the decreasing density with distance, the inner layers are more massive than the outer layers. The number of shells to be used must be chosen to satisfy the constraint that the sum of the shells mass gives the total ejecta mass, i.e.

$$\sum_{j=1}^{N} m_j = m_{\rm ej}, \tag{15}$$

where we have introduced the discrete index $j = i+1$ to differentiate the counting of the shells from the counting of radii given by Eq. (13). In this work, we use $N = 100$ shells which ensures that Eq. (15) is satisfied with 99% of accuracy.

Under the assumption that the shells do not interact with each other, we can estimate the evolution of the ith shell from the energy conservation equation

$$\dot{E}_i = -P_i\,\dot{V}_i - L_{{\rm cool},i} + H_{{\rm inj},i}, \tag{16}$$

where $V_i = (4\pi/3)r_i^3$, E_i, and P_i are the volume, energy, and pressure of the shell, while $H_{{\rm inj},i}$ is the power injected into the shell, and

$$L_{{\rm cool},i} \approx \frac{cE_i}{r_i(1+\tau_{{\rm opt},i})}, \tag{17}$$

is the bolometric luminosity radiated by the shell, being $\tau_{\rm opt,i}$ the optical depth.

Assuming a spatially constant, gray opacity throughout the ejecta, the optical depth of the radiation emitted by the ith layer is given by

$$\tau_{\rm opt,\it i} = \int_{\infty}^{r_i} \kappa\rho(r)dr = \int_{R_{\rm max}}^{r_i} \kappa\rho(r)dr = \tau_{i,0}\hat{t}^{-2n}, \tag{18}$$

$$\tau_{i,0} \equiv \frac{m-3}{m-1}\frac{\kappa m_{\rm ej}}{4\pi R_{*,0}^2}\frac{\left[\left(\frac{R_*}{r_i}\right)^{m-1} - \left(\frac{R_*}{R_{\rm max}}\right)^{m-1}\right]}{\left[1-\left(\frac{R_*}{R_{\rm max}}\right)^{m-3}\right]}, \tag{19}$$

where we have used Eq. (12), and κ is the opacity.

We adopt a radiation-dominated equation of state for the ejecta and, improving with respect to Ref. 20, accounting for the radiation pressure, i.e.

$$E_i = 3P_i\,V_i + L^{\rm abs}_{{\rm cool},i}\frac{r_i}{c}. \tag{20}$$

The power injected into the ejecta originates from the newborn central WD.[20] This energy is absorbed and thermalized becoming a heating source for the expanding matter. The power-law decreasing density (12) suggests that the inner the layer the more radiation it should absorb. In order to account for this effect, we weigh the heating source for each shell using the mass fraction, i.e.

$$H_{{\rm inj},i} = \frac{m_i}{m_{\rm ej}}H_{\rm inj}, \tag{21}$$

where m_i is the shell's mass, and adopts the following form for the heating source:

$$H_{\rm inj} = H_0\left(1+\frac{t}{t_c}\right)^{-\delta}, \tag{22}$$

where H_0 and δ are model parameters. According to Ref. 20, power from fallback accretion with $H_0 \sim 10^{45}\,{\rm erg\,s^{-1}}$, $\delta \sim 1.3$, and $t_c \sim t_*$ (see Table 1), dominates the energy release from the newborn WD at these early-times.

The photospheric radius at a time t is given by the position of the shell that reaches transparency at that time. Namely, it is given by the position of the shell whose optical depth fulfills $\tau_{{\rm opt},i}[r_i(t)] = 1$. Using Eq. (18), we obtain

$$R_{\rm ph} = \frac{R_{{\rm max},0}\hat{t}^n}{\left[1+\frac{m-1}{m-3}\frac{4\pi R_{*,0}^2\left[1-\left(\frac{R_*}{R_{\rm max}}\right)^{m-3}\right]}{\kappa m_{\rm ej}}\left(\frac{R_{\rm max}}{R_*}\right)^{m-1}\hat{t}^{2n}\right]^{\frac{1}{m-1}}}. \tag{23}$$

Equation (23) shows that when the entire ejecta is optically thick, $R_{\rm ph} = R_{\rm max}$. Then, the transparency reaches the inner shells all the way to the instant over which

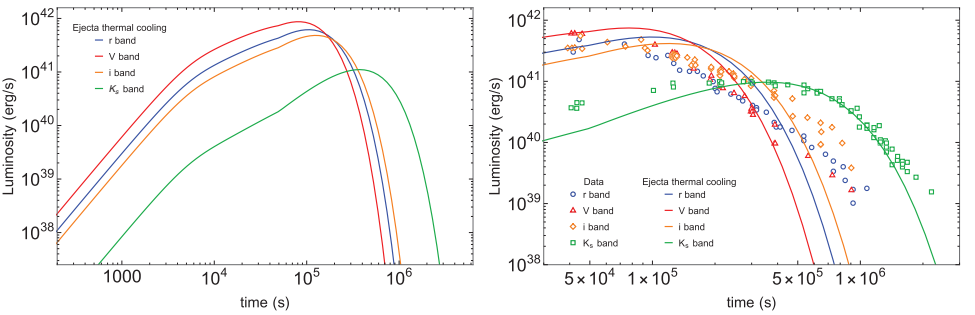


Fig. 2. Left: emission from the expanding, cooling ejecta at early-times in the visible (r and V) and in the infrared (i and K_s) bands, following the theoretical treatment of Sec. 6. Right: zoomed view of the left panel figure at the times relevant for the comparison with the observational data of AT 2017gfo.[7, 8, 50, 51]

$R_{\rm ph} = R_*$, reached at $t = t_{{\rm tr},*}$, when the entire ejecta is transparent. The time $t_{{\rm tr},*}$ is found from the condition $\tau_{{\rm opt},*}[R_*(t_{{\rm tr},*})] = 1$, and is given by

$$\hat{t}_{{\rm tr},*} = \left\{ \frac{m-3}{m-1} \frac{\kappa m_{\rm ej}}{4\pi R_{*,0}^2} \left(\frac{R_*}{R_{\rm max}}\right)^{m-1} \frac{\left[\left(\frac{R_{\rm max}}{R_*}\right)^{m-1} - 1\right]}{\left[1 - \left(\frac{R_*}{R_{\rm max}}\right)^{m-3}\right]} \right\}^{\frac{1}{2n}} . \tag{24}$$

At $t < t_{{\rm tr},*}$, the photospheric radius evolves as $R_{\rm ph} \propto t^{\frac{n(m-3)}{m-1}}$, while at later times, $R_{\rm ph} \propto t^n$. For the parameters of our system, $t_{{\rm tr},*} \sim 10^5$ s (see Fig. 2).

The bolometric luminosity is given by the sum of the luminosity of the shells

$$L_{\rm bol} = \sum_{j=1}^{N} L_{{\rm cool},j}, \tag{25}$$

so the effective temperature of the thermal blackbody radiation, T_s, can be obtained from the Stefan–Boltzmann law, i.e.

$$T_s = \left(\frac{L_{\rm bol}}{4\pi R_{\rm ph}^2 \sigma}\right)^{1/4} , \tag{26}$$

where σ is the Stefan–Boltzmann constant. The power per unit frequency, per unit area, is given by Planck's spectrum

$$B_\nu(\nu, t) = \frac{2\pi h\nu^3}{c^2} [e^{\frac{h\nu}{k_b T_s(t)}} - 1]^{-1}, \tag{27}$$

where ν is the radiation frequency, h and k_b are the Planck and Boltzmann constants. Most of the thermal cooling is radiated in the visible, infrared and ultraviolet wavelengths, which we refer to as optical. Therefore, the spectral density

Table 1. Numerical values of the theoretical model parameters that determine the thermal cooling of the expanding ejecta which fits the data of AT 2017gfo shown in Fig. 2.

Parameter	Value
n	1.22
m	9.00
$m_{\rm ej}$ ($10^{-3} M_\odot$)	1.00
$R_{*,0}$ (10^{11} cm)	4.00
$v_{*,0}$ (10^{9} cm s^{-1})	1.00
κ (cm^2 g^{-1})	0.20
H_0 (10^{45} erg s^{-1})	8.16
δ	1.30
t_c/t_*	1.00

(power per unit frequency) given by the thermal cooling at a frequency ν is

$$J_{\rm cool}(\nu, t) = 4\pi R_{\rm ph}^2(t) B_\nu(\nu, t), \tag{28}$$

and the luminosity radiated in the frequency range $[\nu_1, \nu_2]$ can be then obtained as

$$L_{\rm cool}(\nu_1, \nu_2; t) = \int_{\nu_1}^{\nu_2} J_{\rm cool}(\nu, t) d\nu. \tag{29}$$

Figure 2 shows the luminosity in the r, V, i, and K_s energy bands obtained from Eq. (29), and compares them with the corresponding observations of AT 2017gfo. For the fit of these data, we have set the parameters, as shown in Table 1.

The value of the parameter $v_{\rm max,0}$ does not have any appreciable effect in the evolution, so it cannot be constrained from the data. This happens because most of the mass is concentrated in the innermost layers, so they dominate the thermal evolution. For self-consistency of the model, we have set $v_{\rm max,0} = 2v_{*,0}$, a value that keeps the outermost shell velocity well below the speed of light at any time in the evolution. As for the initial value of the internal energy of the shells, $E_i(t_0)$, we have set them to the initial kinetic energy of each layer, $E_i = (1/2) m_i v_i(t_0)^2$.

There is a general agreement of the model with the observations, although it cannot catch any detailed observational feature. There are some extensions to the present model that can increase its accuracy. For instance, we can abandon the assumption of spherical expansion allowing the layers to have a latitude-dependent velocity. Such a detailed treatment goes beyond our present scope that is to show the broad agreement of a WD-WD merger model with the multiwavelength data but not a dedicated model of AT 2017gfo.

5. Synchrotron and WD Pulsar Radiation

We have shown above that the expanding matter reaches full transparency at about 10^5 s. After this time, the emission originated from the newborn WD as well as the one originated in the ejecta itself, become observable. We here follow the treatment

in Ref. 28 for the explanation of the X-ray afterglow of long GRBs as originating from a newborn spinning NS powering the expanding SN. Here, we simulate the emission generated in the X-rays, in the optical, and in the radio by the synchrotron emission of electrons accelerated in the expanding magnetized ejecta, together with the emission of the newborn spinning WD pulsar.

We show below that synchrotron radiation originating in the merger ejecta dominates the emission up to nearly 10^8 s. We find evidence of the newborn WD pulsar emission, owing to magnetic dipole braking, in the X-ray luminosity at approximately 10^6 s, when the synchrotron radiation was not fully overwhelming yet, and then at times 10^8 s, when the synchrotron luminosity sufficiently decreased for the WD pulsar emission to be fully observed (see Fig. 4 for details).

5.1. *Synchrotron emission by the expanding ejecta*

In this model, a fraction of the kinetic energy of the merger ejecta is used to accelerate electrons that, owing to the presence of the magnetized medium provided by the newborn WD, convert their kinetic energy into synchrotron radiation. The electrons are continuously injected from the newborn WD into the ejecta. The magnetic field threading every ejecta layer evolves as

$$B_i(t) = B_{i,0}\left[\frac{r_{i,0}}{r_i(t)}\right]^{\mu} = \frac{B_{i,0}}{\hat{t}^{\mu n}}, \tag{30}$$

where $B_i^{(0)}$ is the magnetic field strength at $r = r_{i,0}$, and μ gives the spatial dependence of the field at large distance from the newborn WD.

Because the electrons lose their energy very efficiently by synchrotron radiation (see details below), we can simplify our calculation by adopting that the radiation originates from the innermost layer of the ejecta, which we will denote to as R_*. The evolution of this layer, following Eq. (11), is given by $R_*(t) = R_{*,0}\hat{t}^n$, $v_*(t) = v_{*,0}\hat{t}^{n-1}$, $t_* = nR_{*,0}/v_{*,0}$, and the magnetic field at its varying position decreases with time as $B_*(t) = B_{*,0}\hat{t}^{-n}$.

The evolution of the distribution of radiating electrons is determined by the kinetic equation accounting for the particle energy losses[52]

$$\frac{\partial N(E,t)}{\partial t} = -\frac{\partial}{\partial E}[\dot{E}\,N(E,t)] + Q(E,t), \tag{31}$$

where $Q(E,t)$ is the number of injected electrons per unit time, per unit energy, and $\dot{E}$ is the electron energy loss rate.

In our case, we assume electrons are subjected to adiabatic losses by expansion and synchrotron radiation losses, i.e.

$$-\dot{E} = \frac{E}{\tau_{\rm exp}} + \beta B_*(t)^2 E^2, \tag{32}$$

where $\beta = 2e^4/(3m_e^4c^7)$, $B(t)$ is the magnetic field, and

$$\tau_{\rm exp} \equiv \frac{R_*}{v_*} = \frac{t}{n} = \frac{t_*}{n}\hat{t}, \tag{33}$$

is the characteristic timescale of expansion.

In order to find the solution to the kinetic equation (31), we follow the treatment of Ref. 53, adapted to our specific physical situation. We consider a distribution of the injected particles following a power-law behavior, i.e.

$$Q(E,t) = Q_0(t)E^{-\gamma}, \quad 0 \leq E \leq E_{\rm max}, \tag{34}$$

where γ and $E_{\rm max}$ are parameters to be determined from the observational data, and $Q_0(t)$ can be related to the power released by the newborn WD and injected into the ejecta. We assume that the injected power has the form

$$L_{\rm inj}(t) = L_0 \left(1 + \frac{t}{t_q}\right)^{-k}, \tag{35}$$

where L_0, t_q, and k are model parameters. We have not chosen arbitrarily the functional form of Eq. (35), actually, both the powers released by magnetic dipole braking and by fallback accretion (see Eq. (22)) obey this sort of time evolution.

Therefore, the function $Q_0(t)$ can be found from

$$L_{\rm inj}(t) = \int_0^{E_{\rm max}} E\, Q(E,t) dE = \int_0^{E_{\rm max}} Q_0(t) E^{1-\gamma} dE = Q_0(t) \frac{E_{\rm max}^{2-\gamma}}{2-\gamma}, \tag{36}$$

which using Eq. (35) leads to

$$Q_0(t) = q_0 \left(1 + \frac{t}{t_q}\right)^{-k}, \tag{37}$$

where $q_0 \equiv (2-\gamma)L_0/E_{\rm max}^{2-\gamma}$.

Having specified the evolution of the ejecta by Eq. (11) and the magnetic field by Eq. (30), as well as the rate of particle injection given by Eqs. (34) and (37), we can now proceed to the integration of the kinetic equation (31).

First, we find the evolution of a generic electron injected at time $t = t_i$ with energy E_i. Integration of Eq. (32) leads to the energy evolution

$$E = \frac{E_i\,(t_i/t)^n}{1 + \mathcal{M} E_i t_i^n [\frac{1}{\hat{t}_i^{n(1+2\mu)-1}} - \frac{1}{\hat{t}^{n(1+2\mu)-1}}]}, \tag{38}$$

where we have introduced the constant

$$\mathcal{M} \equiv \frac{\beta B_{*,0}^2 t_*^{1-n}}{n(1+2\mu)-1}, \tag{39}$$

which have units of $1/(\text{energy} \times \text{time}^n)$. In the limit $t/t_* \gg 1$ and $n = 1$, Eq. (38) reduces to Eq. (3.3) of Ref. 53, and in the limit $t_* \to \infty$, reduces to the solution presented in Sec. 3 of Ref. 52 for synchrotron losses in a constant magnetic field.

The solution of Eq. (31) is given by

$$N(E,t) = \int_E^\infty Q[E_i, t_i(t, E_i, E)] \frac{\partial t_i}{\partial E} dE_i, \tag{40}$$

where the relation $t_i(t, E_i, E)$ is obtained from Eq. (38).

We can write $N(E,t)$ as a piecewise function of time, separating it into different time intervals that allow simplifications and approximations depending upon the physical situation at work, and on the behavior of the energy injection given by Eq. (37). All the observational data of GRB 170817A are contained in the time interval $t < t_b$ and at electron energies int he range $E_b < E < E_{\max}$ (see definition of t_b and E_b below) where synchrotron losses are dominant. Under these conditions, the solution of Eq. (40) is well approximated by

$$N(E,t) \approx \begin{cases} \dfrac{q_0}{\beta B_{*,0}^2(\gamma-1)} \hat{t}^{2\mu n} E^{-(\gamma+1)}, & t < t_q, \\ \dfrac{q_0}{\beta B_{*,0}^2(\gamma-1)} \left(\dfrac{t_q}{t_*}\right)^k \hat{t}^{2\mu n-k} E^{-(\gamma+1)}, & t_q < t < t_b, \end{cases} \tag{41}$$

and we have defined

$$E_b = \frac{\hat{t}^{2\mu n-1}}{\mathcal{M} t_*^n}, \quad t_b = t_* (\mathcal{M} t_*^n E_{\max})^{\frac{1}{2\mu n-1}}. \tag{42}$$

With the knowledge of $N(E,t)$, we can proceed to estimate the synchrotron spectral density (energy per unit time, per unit frequency) from $J_{\rm syn}(\nu,t)d\nu = P_{\rm syn}(\nu,E)N(E,t)dE$, where $P_{\rm syn}(\nu,E)$ is the synchrotron power per unit frequency ν, radiated by a single electron of energy E. Most of the synchrotron radiation is emitted in a narrow range of frequencies around the so-called photon critical frequency, $\nu_{\rm crit}$. Thus, we can assume electrons emit the synchrotron radiation at

$$\nu \approx \nu_{\rm crit} \approx \alpha B_* E^2, \tag{43}$$

where $\alpha = 3e/(4\pi m_e^3 c^5)$. This gives a relation between the electron energy and the radiation frequency, and $P_{\rm syn}(\nu,E)$ can be approximated to the bolometric power

$$P_{\rm syn}(\nu,E) \approx P_{\rm syn}(\nu) = \beta B_*^2 E^2(\nu) = \frac{\beta}{\alpha} B_* \nu. \tag{44}$$

Within this approximation, the spectral density is

$$J_{\rm syn}(\nu,t) \approx P_{\rm syn}(\nu) N(E,t) \frac{dE}{d\nu}. \tag{45}$$

It can be seen from Eq. (41) that in each time and frequency interval we can write

$$N(E,t) = \eta\, \hat{t}^l E^{-p}, \tag{46}$$

where η and the power-law indexes l and p are known constants from Eq. (41). With this, the spectral density (45) becomes

$$J_{\rm syn}(\nu,t) = \frac{\beta}{2} \alpha^{\frac{p-3}{2}} \eta B_{*,0}^{\frac{p+1}{2}} \hat{t}^{\frac{2l-\mu n(p+1)}{2}} \nu^{\frac{1-p}{2}}. \tag{47}$$

The synchrotron luminosity in the frequencies $[\nu_1, \nu_2]$ can be then obtained as

$$L_{\rm syn}(\nu_1, \nu_2; t) = \int_{\nu_1}^{\nu_2} J_{\rm syn}(\nu, t) d\nu, \tag{48}$$

which in a narrow frequency band from $\nu_1 = \nu$ to $\nu_2 = \nu + \Delta\nu$ where $\Delta\nu/\nu \ll 1$, can be well approximated as

$$L_{\rm syn}(\nu, t) \approx \nu J_{\rm syn}(\nu, t) = \frac{\beta}{2} \alpha^{\frac{p-3}{2}} \eta B_{*,0}^{\frac{p+1}{2}} \hat{t}^{\frac{2l-\mu n(p+1)}{2}} \nu^{\frac{3-p}{2}}, \tag{49}$$

where we have used Eq. (47).

5.2. *WD evolution and pulsar emission*

The central WD emits also pulsar-like radiation. We adopt a dipole + quadrupole magnetic field model.[54] In this model, the total luminosity of spindown is

$$\begin{aligned} L_{\rm sd} &= L_{\rm dip} + L_{\rm quad} \\ &= \frac{2}{3c^3} \Omega^4 B_{\rm dip}^2 R_{\rm wd}^6 \sin^2 \chi_1 \left(1 + \xi^2 \frac{16}{45} \frac{R_{\rm wd}^2 \Omega^2}{c^2} \right), \end{aligned} \tag{50}$$

where the parameter ξ defines the quadrupole to dipole strength ratio as

$$\xi \equiv \sqrt{\cos^2 \chi_2 + 10 \sin^2 \chi_2} \frac{B_{\rm quad}}{B_{\rm dip}}, \tag{51}$$

and the modes can be separated: $\chi_1 = 0$ and any value of χ_2 for the $m = 0$ mode, $(\chi_1, \chi_2) = (90°, 0°)$ for the $m = 1$ mode, and $(\chi_1, \chi_2) = (90°, 90°)$ for the $m = 2$ mode.

The WD evolution is obtained from the energy balance equation

$$-(\dot{W} + \dot{T}) = L_{\rm tot} = L_{\rm inj} + L_{\rm sd}, \tag{52}$$

where W and T are, respectively, the gravitational and rotational energies of the newborn WD. We can obtain an analytic, sufficiently accurate solution of Eq. (52) by noticing the following. The power injected in electrons $L_{\rm inj}$ is larger than $L_{\rm sd}$ and has a shorter timescale with respect to the spindown timescale (see Eq. (35) and Fig. 5), so at $t < t_q$, we have $L_{\rm tot} \approx L_{\rm inj}$. At later times, $L_{\rm tot} \approx L_{\rm sd}$, so the luminosity should approach the spindown luminosity

$$L_{\rm sd} = L_{\rm sd,0} \left(1 + \frac{t}{\tau_{\rm sd}} \right)^{-s}, \tag{53}$$

where $s = (n_b + 1)/(n_b - 1)$, being n_b the so-called braking index ($n_b = 3$ for a pure dipole and $n_b = 5$ for a pure quadrupole), and $\tau_{\rm sd}$ is the spindown timescale

$$\tau_{\rm sd} = \frac{1}{2\mathcal{A}\Omega_0^2}, \tag{54}$$

being $\mathcal{A} = (2/3)(B_{\rm dip}^2 R_{\rm wd}^6)/(c^3 I)$, and Ω_0 the initial angular velocity of the WD.

With the above, Eq. (52) is integrated analytically accounting for changes in the WD structure. We describe the WD as an effective Maclaurin spheroid,[55] so the angular velocity, Ω, is related to the spheroid eccentricity, e, by

$$\Omega^2 = 2\pi G\rho g(e), \quad g(e) = \frac{\left(3-2e^2\right)\left(1-e^2\right)^{1/2}\arcsin(e)}{e^3} - \frac{3\left(1-e^2\right)}{e^2}, \tag{55}$$

where $\rho = 3m_{\rm wd}/(4\pi R_{\rm wd}^3)$ is the density of the sphere with the same volume of the spheroid, being $m_{\rm wd}$ and $R_{\rm wd}$ the corresponding values of the mass and radius of the WD. The total energy of the spheroid is also a function of the eccentricity as

$$E = T + W = \pi G\rho I_0 \mathcal{F}(e), \tag{56}$$

where $I_0 = (2/5)m_{\rm wd}R_{\rm wd}^2$, and

$$\mathcal{F}(e) = -2 + \frac{3(1-e^2)^{2/3}}{e^2} + \frac{(4e^2-3)(1-e^2)^{1/6}}{e^3}\arcsin(e) \approx -\frac{4e^2}{15}, \tag{57}$$

being the last line a series expansion of the function $\mathcal{F}$ which is accurate enough for low values of the eccentricity and which allows to give an analytic solution for the eccentricity as a function of time.

Then, integrating Eq. (52) and using Eqs. (56) and (57), we obtain

$$e(t) \approx \sqrt{\frac{15\Theta(t)}{4}}, \quad \Omega \approx \sqrt{2\pi G\rho\Theta(t)}, \tag{58}$$

where

$$\Theta(t) = -\mathcal{F}(e_0) + \mathcal{G}(t), \tag{59}$$

$$\mathcal{G}(t) = \frac{L_0 t_q}{\pi G\rho I_0(k-1)}\left[\left(1+\frac{t}{t_q}\right)^{1-k} - 1\right] + \frac{L_{\rm sd,0}\tau_{\rm sd}}{\pi G\rho I_0(s-1)}\left[\left(1+\frac{t}{\tau_{\rm sd}}\right)^{1-s} - 1\right], \tag{60}$$

where e_0 is the initial value of the spheroid eccentricity, and we have used that the function $g(e)$ in Eq. (55) satisfies $g(e) = -\mathcal{F}(e)$, at the order of our approximation. We recall that the moment of inertia changes with the eccentricity as $I = I_0(1-e^2)^{-1/3} \approx I_0(1+e^2/3)$. The corresponding parameters of the model that explains the afterglow emission at different wavelengths are presented in the next section.

6. Model Parameters from the Multiwavelength Data

We proceed to determine the model parameters that best fit the GRB 170817A afterglow. We list in Table 2 the value adopted for each parameter of the present model to fit the multiwavelength data of GRB 170817A shown in Fig. 3. We did not consider here data at MeV energies because it is only present in the prompt emission that we have already discussed in Sec. 3 and is explained by a different mechanism from the synchrotron radiation. There are observations in the 30 MeV-10 GeV energy band by AGILE[67] which give upper limits $\sim 10^{44}$–$10^{45}\,{\rm erg\,s^{-1}}$ in

Table 2. Numerical values of the theoretical model of synchrotron radiation of Sec. 5 that fit the multiwavelength observational data of GRB 170817A as shown in Fig. 3.

Parameter	Value
γ	1.13
k	2.70
μ	1.50
L_0 (10^{39} erg s^{-1})	1.80
$B_{*,0}$ (10^9 G)	1.00
$E_{\rm max}$ (10^6 $m_e c^2$)	1.00
t_q (10^7 s)	1.22
ξ	0.00
$B_{\rm dip}$ (10^{10} G)	1.30
P (s)	12.21

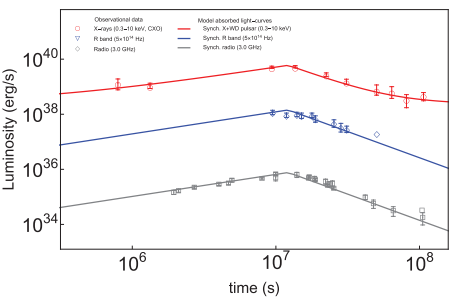


Fig. 3. Comparison of the theoretical (solid curves) light-curves with the observational data (points) of GRB 170817A, in selected energy bands from the radio to the gamma-rays. The radio data at 3 GHz have been taken from Refs. 56–63; the infrared (F606W HST band) data points are retrieved from Refs. 63–64; the X-ray (0.3–10 keV) data from CXO are taken from Ref. 66.

the time interval $\sim 10^3$–10^6 s. For the parameters of Table 2, no emission is indeed expected at these energies because the maximum synchrotron radiation frequency obtained from Eq. (43) falls below 10 GeV before $\sim 10^4$ s. The synchrotron luminosity vanishes at these energies at longer times.

Having discussed the gamma-rays, we turn now to the X-rays, optical and radio emission. Figure 3 compares the absorbed luminosity predicted by the model (see Sec. 4), as a function of time, in selected energy bands, with the corresponding observational data of GRB 170817A. We have here included the X-ray data the 0.3–10 keV energy band from the *Chandra* X-ray Observatory (CXO) including the latest observations,[66] the infrared data from the HST at $\approx 5 \times 10^{14}$ Hz,[63–65] and the radio data at 3 GHz.[56–63, 66]

The model shows a satisfactory fit of the data in the X-rays, optical and radio data, both where the luminosity rises, at times $t \sim 10^6$–10^7 s, and where it fades off, at $t \gtrsim 10^7$ s. We show a closer view in Fig. 4 of the X-rays, optical, and radio luminosities around the time of the peak luminosity. The synchrotron luminosity

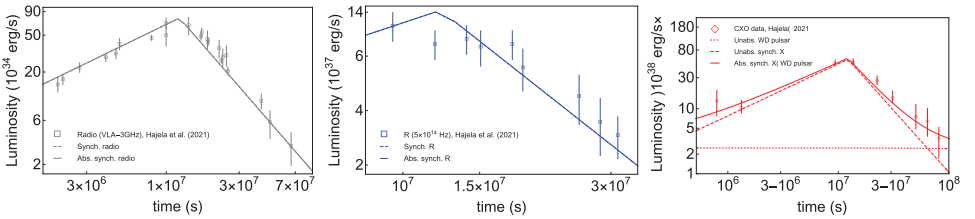


Fig. 4. Zoomed views of the radio (left), optical (center), and X-ray (right) luminosities around transparency. The dashed curves represent the unabsorbed luminosities. The dotted curve in the right panel shows the contribution from the newborn WD pulsar which causes the deviation from the pure synchrotron power-law luminosity at times $\gtrsim 3 \times 10^7$ s.

rises as a power-law while the energy injection is constant, i.e. up to $t \approx t_q = 1.2 \times 10^7$ s, while it decreases as a power-law at later times. Probably the most interesting feature that can be seen from these zoomed views appears in the X-ray emission, where we can see in addition to the synchrotron luminosity, evidence of the WD pulsar emission owing to the magnetic dipole braking. The contribution from the pulsar emission is seen first at $t \sim 10^6$ s when the synchrotron radiation is rising but is still comparable with the pulsar spindown luminosity. Then, the synchrotron luminosity takes over, reaches a peak at approximately 10^7 s, and then decreases. While the optical and radio counterparts continue to fade with time as dictated by the synchrotron radiation, the accuracy of the X-ray data of the CXO presented in Ref. 66 allows to identify a clear deviation in the X-rays at a few 10^7 s from such a power-law behavior. This is again the signature of the emergence of the WD pulsar emission.

We have used the entity of this deviation to constrain the WD pulsar parameters. Since the pulsar emission depends on the WD radius (see Sec. 5.2), we first estimated the mass of the newborn WD. To accomplish this task, we must apply the considerations of Sec. 2. From the inferred mass of the ejecta, $m_{\rm ej} = 10^{-3} M_\odot$ (see Table 1), we obtain an upper limit to the binary mass ratio via Eq. (7), by requesting that the newborn object be a stable, sub-Chandrasekhar WD, i.e. $m_{\rm wd} \lesssim 1.4 M_\odot$, which leads to $q \lesssim 0.87$. According to this maximum mass ratio and the ejecta mass value, Eq. (5) constraints the secondary mass to the range $m_2 \lesssim 0.85 M_\odot$. With the knowledge of q and m_2, Eq. (2) constrains the total binary mass to $M \lesssim 1.82 M_\odot$. Thus, the primary component must satisfy $m_1 \lesssim 0.97 M_\odot$.

We assume that the newborn WD is stable, therefore it might have a mass close but not equal to the Chandrasekhar mass, since some mass will be accreted via matter fallback. Hereafter, we shall adopt in our estimates $m_{\rm wd} \approx 1.3 M_\odot$, so a radius $R_{\rm wd} \approx 3.4 \times 10^8$ cm. With these WD structure parameters, we can proceed to constrain the magnetic field strength and rotation period.

The X-ray emission data show that deviation from the pure synchrotron emission behavior starts at $\approx 3 \times 10^7$ s, and extends up to when we have data, namely, up to $\approx 10^8$ s (see Fig. 4). This would suggest to chose this time for the spindown timescale

$\tau_{\rm sd}$, but at the moment it is only a lower limit to $\tau_{\rm sd}$ because the luminosity did not reach yet the power-law given by the pulsar luminosity.

Hereafter, we assume a pure dipole (i.e. $\xi = 0$) because the fit of the X-ray emission does not require at the moment the quadrupole component (see Fig. 4). By eliminating the rotation angular velocity between the pulsar luminosity (50) and the spindown timescale (54), we can express the magnetic field strength as

$$B_{\rm dip} = \frac{3^{1/2} c^{3/2} I}{2^{3/2} L_{\rm sd}^{1/2} R_{\rm wd}^3 \tau_{\rm sd}}, \tag{61}$$

where I is the moment of inertia. We can use Eq. (61) to give an upper limit to $B_{\rm dip}$ by setting as values of $L_{\rm sd}$ and $\tau_{\rm sd}$, the values of the latest value of the X-ray luminosity data, i.e. $L_{\rm sd} = L_X \approx 4.87 \times 10^{38}\,{\rm erg\,s^{-1}}$, and $\tau_{\rm sd} \approx 10^8\,{\rm s}$. With this, we obtain an upper value $B_{\rm dip,max} \approx 7.46 \times 10^{11}$ G. To this upper value of $B_{\rm dip}$, it corresponds an upper value of the initial rotation period which can be obtained by calculating $P_0 = 2\pi/\Omega_0$ from Eq. (50), i.e.

$$P_0 = 2\pi \left(\frac{2 B_{\rm dip}^2 R_{\rm wd}^6}{c^3 L_{\rm sd}} \right)^{1/4}, \tag{62}$$

from which we obtain $P_{0,\rm max} \approx 75.25\,{\rm s}$. We can further constrain the rotation period by seeking for values of the magnetic field strength and rotation period in agreement with the model presented in Sec. 3 for the prompt emission. Such a mechanism is expected to release magnetic energy stored in the magnetosphere, i.e.

$$E_B \approx \frac{1}{6} B_{\rm dip}^2 R_{\rm wd}^3, \tag{63}$$

so we need a dipole magnetic field strength

$$B_{\rm dip} = \left(\frac{6 E_B}{R_{\rm wd}^3} \right)^{1/2} \approx \left(\frac{6 E_{\rm prompt}}{R_{\rm wd}^3} \right)^{1/2}. \tag{64}$$

If we assume that the entire energy of the prompt emission, $E_{\rm prompt} \approx 4.16 \times 10^{46}\,{\rm erg}$ (see Sec. 3) is paid by the magnetosphere energy, we obtain a magnetic field $B_{\rm dip} = 9.61 \times 10^{10}$ G. If we require the magnetic field energy to cover only the nonthermal component of the prompt, i.e. $1.17 \times 10^{45}\,{\rm erg}$ (see Sec. 3), then the dipole magnetic field becomes 1.30×10^{10} G. For the above magnetic field values, Eq. (62) gives, respectively, $P_0 \approx 30\,{\rm s}$, and $P_0 \approx 12\,{\rm s}$. The WD pulsar luminosity shown in Figs. 3 and 5 corresponds to the latter case.

The energy released (and injected into the ejecta) by the fallback accretion phase is $E_{\rm fb} = H_0 t_*/(\delta - 1) \approx 3.34 \times 10^{48}\,{\rm erg}$. Energy and angular momentum are transferred to the newborn WD during this phase, and since the rotational to gravitational energy ratio of a uniformly rotating WD is of the order of 10^{-2},[32] the newborn WD has gained about a few $10^{46}\,{\rm erg}$ of rotational energy during this phase. This might produce at a rotation period decrease of the order of a second, which confirms that the WD must be already fast rotating at birth.

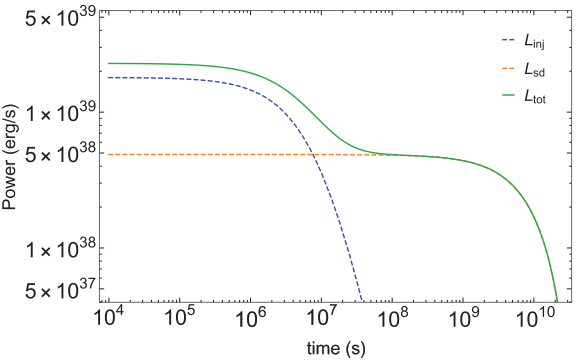


Fig. 5. Power injected into the ejecta by the newborn WD and the pulsar emission as given, respectively, by Eqs. (35) and (50). See the main text for further details.

In Fig. 5, we plot the power injected in energetic electrons from the WD, $L_{\rm inj}$, and the luminosity due to magnetic dipole braking, $L_{\rm sd}$. Both components release an energy of the order of 10^{46} erg. From the inferred rotation period of 12.21 s, the initial eccentricity turns out to be $e_0 \approx 0.39$, so the moment of inertia is about 5% bigger than the one of the equivalent spherical configuration. Using the evolution equations (58), we obtain that the moment of inertia, for instance from 10^4 s to 10^8 s, changes in about 0.03%. This small change in the structure of the WD, and the associated change in the rotational and gravitational energy, are sufficient to pay for the energy released by the ongoing magnetospheric phenomena responsible for the injection of particles into the ejecta and for the pulsar emission; see Sec. 5.2.

7. Conclusions

We have here addressed a self-consistent explanation of GRB 170817A, including its associated optical emission AT 2017gfo, based on a WD-WD merger. The most recent data of *Chandra* of the X-ray emission of GRB 170817A at $\sim 10^8$ s (~ 1000 d) after the GRB trigger,[21,22] indicate an X-ray re-brightening. This is explained by the emergence of the pulsar-like activity of the newborn WD (see Figs. 3 and 4), as predicted in Refs. 13, 20. We have here inferred that the newborn object is consistent with a massive ($\sim 1.3\,M_\odot$), fast rotating ($P \gtrsim 12$ s), highly magnetized ($B \sim 10^{10}$ G) WD, formed in a $1.0 + 0.8$ WD-WD merger (see Secs. 2 and 6).

The post-merger emission at different wavelengths is explained as follows. The prompt gamma-ray emission detected by the *Fermi*-GBM, with a luminosity of $\sim 10^{47}\,{\rm erg\,s^{-1}}$ and observed duration of $\lesssim 1$ s, can be explained by the transient hot corona produce at the merger. The high temperature produces photons that undergo $e^- e^+$ pair creation, the pairs are accelerated by the electric field induced by the 10^{10} G magnetic field an the WD rotation, thereby producing photons. The system is highly opaque to these photons (see Sec. 3) to the $\gamma\gamma$ pair production process. Only a small percentage of photons is expected to be able to escape from

the system along the polar axis, leading to the small amount of nonthermal emission observed above 1 MeV, while the rest is expected to form a nearly thermal plasma.

The ejecta expand with velocities $\sim 10^9\,\mathrm{cm\,s^{-1}}$, and release energy by thermal cooling (see Sec. 4) and synchrotron radiation (see Sec. 5), powered by the newborn WD at the merger (see Sec. 5.2). Fallback accretion onto the newborn WD injects energy into the ejecta at early-times, heating up the ejecta. The ejecta is optically thick up to nearly 10^5, so the ejecta cool by diffusion while it expands. The thermal radiation is in agreement with the data of the early optical counterpart AT 2017gfo (see Fig. 2). This explanation is markedly different from the *nuclear kilonova* from decay of r-process synthesized heavy nuclei in an NS-NS merger ejecta.

The signature of the synchrotron radiation is identified from nearly 10^6 s, which explains the rising and decreasing luminosities with the same power-law slopes in the X-ray, optical and radio emissions (see Fig. 3 for details).

The X-ray data are essential for identifying the emergence of the newborn WD as a pulsar. We have shown evidence of the pulsar emission around 10^6 s and at late-times 10^8 s, causing the X-ray luminosity to deviate from the power-law emission of a pure synchrotron emission (see Figs. 3 and 4). These data reveal a rotation period $\gtrsim 12$ s, and magnetic field of $\sim 10^{10}$ G. The follow-up of the GRB 170817A X-ray emission in the next months/years to come is crucial to confirm this prediction.

Summarizing, GRB 170817A/AT 2017gfo are explained by a WD-WD merger. The $10^{-3} M_\odot$ expelled in the merger expand and radiates via thermal and synchrotron cooling. The former explains AT 2017gfo and the latter the late-time X-rays, optical and radio emission. In this line, the association of GW170417A with GRB 170817A[1] is not confirmed in our treatment based on the new data in the X-rays, optical, and in the radio up to 10^8 s. Therefore, we indicate the necessity to further inquire on the spacetime sequence of the early part of these events.

Indeed, WDs of parameters approaching the present ones have been already identified, e.g. the WD in V1460 Her with $P \approx 39$ s,[68] and the most recent observation of the WD in LAMOST J024048.51+195226.9 with $P \approx 25$ s.[69] WDs of similar properties have been proposed as a model of SGRs and AXPs.[70–73] Therefore, the newborn WD pulsar in GRB 170817A could show itself in the near future as an SGR/AXP in the GRB 170817A sky position.

References

1. B. P. Abbott *et al.*, *Astrophys. J. Lett.* **848** (2017) L12, arXiv:1710.05833 [astro-ph.HE].
2. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, *Astrophys. J. Lett.* **848** (2017) L14, arXiv:1710.05446 [astro-ph.HE].
3. V. Savchenko, C. Ferrigno, E. Kuulkers, A. Bazzano, E. Bozzo, S. Brandt, J. Chenevez, T. J. L. Courvoisier, R. Diehl, A. Domingo, L. Hanlon, E. Jourdain, A. von Kienlin,

P. Laurent, F. Lebrun, A. Lutovinov, A. Martin-Carrillo, S. Mereghetti, L. Natalucci, J. Rodi, J. P. Roques, R. Sunyaev and P. Ubertini, *Astrophys. J. Lett.* **848** (2017) L15, arXiv:1710.05449 [astro-ph.HE].
4. B. P. Abbott *et al.*, *Astrophys. J. Lett.* **848** (2017) L13, arXiv:1710.05834 [astro-ph.HE].
5. E. Pian *et al.*, *Nature* **551** (2017) 67, arXiv:1710.05858 [astro-ph.HE].
6. I. Arcavi, G. Hosseinzadeh, D. A. Howell, C. McCully, D. Poznanski, D. Kasen, J. Barnes, M. Zaltzman, S. Vasylyev, D. Maoz and S. Valenti, *Nature* **551** (2017) 64, arXiv:1710.05843 [astro-ph.HE].
7. P. S. Cowperthwaite *et al.*, *Astrophys. J. Lett.* **848** (2017) L17, arXiv:1710.05840 [astro-ph.HE].
8. M. Nicholl *et al.*, *Astrophys. J. Lett.* **848** (2017) L18, arXiv:1710.05456 [astro-ph.HE].
9. B. Paczynski, *Astrophys. J. Lett.* **308** (1986) L43.
10. D. Eichler, M. Livio, T. Piran and D. N. Schramm, *Nature* **340** (1989) 126.
11. R. Narayan, B. Paczynski and T. Piran, *Astrophys. J. Lett.* **395** (1992) L83, arXiv:astro-ph/9204001.
12. L. Li and Z.-G. Dai, *Astrophys. J.* **918** (2021) 52, arXiv:2106.04788 [astro-ph.HE].
13. J. A. Rueda, R. Ruffini, Y. Wang, Y. Aimuratov, U. Barres de Almeida, C. L. Bianco, Y. C. Chen, R. V. Lobato, C. Maia, D. Primorac, R. Moradi and J. F. Rodriguez, *J. Cosmol. Astropart. Phys.* **10** (2018) 006, arXiv:1802.10027 [astro-ph.HE].
14. C. Yue, Q. Hu, F.-W. Zhang, Y.-F. Liang, Z.-P. Jin, Y.-C. Zou, Y.-Z. Fan and D.-M. Wei, *Astrophys. J. Lett.* **853** (2018) L10, arXiv:1710.05942 [astro-ph.HE].
15. R. T. Wollaeger, C. L. Fryer, E. A. Chase, C. J. Fontes, M. Ristic, A. L. Hungerford, O. Korobkin, R. O'Shaughnessy and A. M. Herring, *Astrophys. J.* **918** (2021) 10, arXiv:2105.11543 [astro-ph.HE].
16. O. Korobkin, R. T. Wollaeger, C. L. Fryer, A. L. Hungerford, S. Rosswog, C. J. Fontes, M. R. Mumpower, E. A. Chase, W. P. Even, J. Miller, G. W. Misch and J. Lippuner, *Astrophys. J.* **910** (2021) 116, arXiv:2004.00102 [astro-ph.HE].
17. Y. L. Zhu, K. A. Lund, J. Barnes, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower and R. Surman, *Astrophys. J.* **906** (2021) 94, arXiv:2010.03668 [astro-ph.HE].
18. J. Barnes, Y. L. Zhu, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower and R. Surman, *Astrophys. J.* **918** (2021) 44, arXiv:2010.11182 [astro-ph.HE].
19. B. Côté, C. L. Fryer, K. Belczynski, O. Korobkin, M. Chruślińska, N. Vassh, M. R. Mumpower, J. Lippuner, T. M. Sprouse, R. Surman and R. Wollaeger, *Astrophys. J.* **855** (2018) 99, arXiv:1710.05875 [astro-ph.GA].
20. J. A. Rueda, R. Ruffini, Y. Wang, C. L. Bianco, J. M. Blanco-Iglesias, M. Karlica, P. Lorén-Aguilar, R. Moradi and N. Sahakyan, *J. Cosmol. Astropart. Phys.* **2019** (2019) 044, arXiv:1807.07905 [astro-ph.HE].
21. E. Troja, B. O'Connor, G. Ryan, L. Piro, R. Ricci, B. Zhang, T. Piran, G. Bruni, S. B. Cenko and H. van Eerten, arXiv e-prints (2021), arXiv:2104.13378 [astro-ph.HE].
22. E. Troja, H. van Eerten, B. Zhang, G. Ryan, L. Piro, R. Ricci, B. O'Connor, M. H. Wieringa, S. B. Cenko and T. Sakamoto, *Mon. Not. R. Astron. Soc.* **498** (2020) 5643, arXiv:2006.01150 [astro-ph.HE].
23. M. Livio and P. Mazzali, *Phys. Rep.* **736** (2018) 1, arXiv:1802.03125 [astro-ph.SR].
24. M. Della Valle and L. Izzo, *Astron. Astrophys. Rev.* **28** (2020) 3, arXiv:2004.06540 [astro-ph.SR].
25. R. Ruffini, R. Moradi, J. A. Rueda, L. Li, N. Sahakyan, Y. C. Chen, Y. Wang, Y. Aimuratov, L. Becerra, C. L. Bianco, C. Cherubini, S. Filippi, M. Karlica, G. J. Mathews,

M. Muccino, G. B. Pisani and S. S. Xue, *Mon. Not. R. Astron. Soc.* **504** (2021) 5301, arXiv:2103.09142 [astro-ph.HE].
26. J. A. Rueda, R. Ruffini, M. Karlica, R. Moradi and Y. Wang, *Astrophys. J.* **893** (2020) 148, arXiv:1905.11339 [astro-ph.HE].
27. Y. Wang, J. A. Rueda, R. Ruffini, L. Becerra, C. Bianco, L. Becerra, L. Li and M. Karlica, *Astrophys. J.* **874** (2019) 39, arXiv:1811.05433 [astro-ph.HE].
28. R. Ruffini, M. Karlica, N. Sahakyan, J. A. Rueda, Y. Wang, G. J. Mathews, C. L. Bianco and M. Muccino, *Astrophys. J.* **869** (2018) 101, arXiv:1712.05000 [astro-ph.HE].
29. L. Li, X.-F. Wu, W.-H. Lei, Z.-G. Dai, E.-W. Liang and F. Ryde, *Astrophys. J. Suppl.* **236** (2018) 26, arXiv:1712.09390 [astro-ph.HE].
30. L. Becerra, K. Boshkayev, J. A. Rueda and R. Ruffini, *Mon. Not. R. Astron. Soc.* **487** (2019) 812, arXiv:1812.10543 [astro-ph.SR].
31. L. Becerra, J. A. Rueda, P. Lorén-Aguilar and E. García-Berro, *Astrophys. J.* **857** (2018) 134, arXiv:1804.01275 [astro-ph.SR].
32. K. Boshkayev, J. A. Rueda, R. Ruffini and I. Siutsou, *Astrophys. J.* **762** (2013) 117, arXiv:1204.2070 [astro-ph.SR].
33. W. Benz, A. G. W. Cameron, W. H. Press and R. L. Bowers, *Astrophys. J.* **348** (1990) 647.
34. J. Guerrero, E. García-Berro and J. Isern, *Astron. Astrophys.* **413** (2004) 257.
35. P. Lorén-Aguilar, J. Isern and E. García-Berro, *Astron. Astrophys.* **500** (2009) 1193.
36. R. Longland, P. Lorén-Aguilar, J. José, E. García-Berro and L. G. Althaus, *Astron. Astrophys.* **542** (2012) A117, arXiv:1205.2538 [astro-ph.SR].
37. C. Raskin, E. Scannapieco, C. Fryer, G. Rockefeller and F. X. Timmes, *Astrophys. J.* **746** (2012) 62, arXiv:1112.1420 [astro-ph.HE].
38. C. Zhu, P. Chang, M. H. van Kerkwijk and J. Wadsley, *Astrophys. J.* **767** (2013) 164, arXiv:1210.3616 [astro-ph.SR].
39. M. Dan, S. Rosswog, M. Brüggen and P. Podsiadlowski, *Mon. Not. R. Astron. Soc.* **438** (2014) 14, arXiv:1308.1667 [astro-ph.HE].
40. M. Nauenberg, *Astrophys. J.* **175** (1972) 417.
41. G. Vianello, R. J. Lauer, P. Younk, L. Tibaldo, J. M. Burgess, H. Ayala, P. Harding, M. Hui, N. Omodei and H. Zhou, arXiv e-prints (2015) arXiv:1507.08343 [astro-ph.HE].
42. D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. Van Der Linde, *J. R. Stat. Soc. Series B* **64** (2002) 583.
43. E. Moreno, F. J. Vazquez-Polo and C. P. Robert, arXiv e-prints (2013), arXiv:1310.2905 [stat.ME].
44. L. Li, *Astrophys. J. Suppl.* **242** (2019) 16, arXiv:1810.03129 [astro-ph.HE].
45. L. Li, *Astrophys. J. Suppl.* **245** (2019) 7.
46. L. Li, *Astrophys. J.* **894** (2020) 100, arXiv:1908.09240 [astro-ph.HE].
47. L. Li, F. Ryde, A. Pe'er, H.-F. Yu and Z. Acuner, *Astrophys. J. Suppl.* **254** (2021) 35, arXiv:2012.03038 [astro-ph.HE].
48. L. Li and B. Zhang, *Astrophys. J. Suppl.* **253** (2021) 43, arXiv:2101.04325 [astro-ph.HE].
49. P. Goldreich and W. H. Julian, *Astrophys. J.* **157** (1969) 869.
50. E. Troja *et al.*, *Nature* **551** (2017) 71, arXiv:1710.05433 [astro-ph.HE].
51. D. Lazzati, D. López-Cámara, M. Cantiello, B. J. Morsony, R. Perna and J. C. Workman, *Astrophys. J. Lett.* **848** (2017) L6, arXiv:1709.01468 [astro-ph.HE].
52. N. S. Kardashev, *Sov. Astron.* **6** (1962) 317.
53. F. Pacini and M. Salvati, *Astrophys. J.* **186** (1973) 249.

54. J. Pétri, *Mon. Not. R. Astron. Soc.* **450** (2015) 714, arXiv:1503.05307 [astro-ph.HE].
55. S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (1969).
56. K. D. Alexander, E. Berger, W. Fong, P. K. G. Williams, C. Guidorzi, R. Margutti, B. D. Metzger, J. Annis, P. K. Blanchard, D. Brout, D. A. Brown, H. Y. Chen, R. Chornock, P. S. Cowperthwaite, M. Drout, T. Eftekhari, J. Frieman, D. E. Holz, M. Nicholl, A. Rest, M. Sako, M. Soares-Santos and V. A. Villar, *Astrophys. J. Lett.* **848** (2017) L21, arXiv:1710.05457 [astro-ph.HE].
57. G. Hallinan, A. Corsi, K. P. Mooley, K. Hotokezaka, E. Nakar, M. M. Kasliwal, D. L. Kaplan, D. A. Frail, S. T. Myers, T. Murphy, K. De, D. Dobie, J. R. Allison, K. W. Bannister, V. Bhalerao, P. Chandra, T. E. Clarke, S. Giacintucci, A. Y. Q. Ho, A. Horesh, N. E. Kassim, S. R. Kulkarni, E. Lenc, F. J. Lockman, C. Lynch, D. Nichols, S. Nissanke, N. Palliyaguru, W. M. Peters, T. Piran, J. Rana, E. M. Sadler and L. P. Singer, *Science* **358** (2017) 1579, arXiv:1710.05435 [astro-ph.HE].
58. S. Kim *et al.*, *Astrophys. J. Lett.* **850** (2017) L21, arXiv:1710.05847 [astro-ph.HE].
59. K. P. Mooley, E. Nakar, K. Hotokezaka, G. Hallinan, A. Corsi, D. A. Frail, A. Horesh, T. Murphy, E. Lenc, D. L. Kaplan, K. de, D. Dobie, P. Chandra, A. Deller, O. Gottlieb, M. M. Kasliwal, S. R. Kulkarni, S. T. Myers, S. Nissanke, T. Piran, C. Lynch, V. Bhalerao, S. Bourke, K. W. Bannister and L. P. Singer, *Nature* **554** (2018) 207, arXiv:1711.11573 [astro-ph.HE].
60. K. P. Mooley, D. A. Frail, D. Dobie, E. Lenc, A. Corsi, K. De, A. J. Nayana, S. Makhathini, I. Heywood, T. Murphy, D. L. Kaplan, P. Chandra, O. Smirnov, E. Nakar, G. Hallinan, F. Camilo, R. Fender, S. Goedhart, P. Groot, M. M. Kasliwal, S. R. Kulkarni and P. A. Woudt, *Astrophys. J. Lett.* **868** (2018) L11, arXiv:1810.12927 [astro-ph.HE].
61. K. D. Alexander, R. Margutti, P. K. Blanchard, W. Fong, E. Berger, A. Hajela, T. Eftekhari, R. Chornock, P. S. Cowperthwaite, D. Giannios, C. Guidorzi, A. Kathirgamaraju, A. MacFadyen, B. D. Metzger, M. Nicholl, L. Sironi, V. A. Villar, P. K. G. Williams, X. Xie and J. Zrake, *Astrophys. J. Lett.* **863** (2018) L18, arXiv:1805.02870 [astro-ph.HE].
62. D. Dobie, D. L. Kaplan, T. Murphy, E. Lenc, K. P. Mooley, C. Lynch, A. Corsi, D. Frail, M. Kasliwal and G. Hallinan, *Astrophys. J. Lett.* **858** (2018) L15, arXiv:1803.06853 [astro-ph.HE].
63. L. Piro, E. Troja, B. Zhang, G. Ryan, H. van Eerten, R. Ricci, M. H. Wieringa, A. Tiengo, N. R. Butler, S. B. Cenko, O. D. Fox, H. G. Khandrika, G. Novara, A. Rossi and T. Sakamoto, *Mon. Not. R. Astron. Soc.* **483** (2019) 1912, arXiv:1810.04664 [astro-ph.HE].
64. J. D. Lyman *et al.*, *Nature Astronomy* **2** (2018) 751, arXiv:1801.02669 [astro-ph.HE].
65. G. P. Lamb, J. D. Lyman, A. J. Levan, N. R. Tanvir, T. Kangas, A. S. Fruchter, B. Gompertz, J. Hjorth, I. Mandel, S. R. Oates, D. Steeghs and K. Wiersema, *Astrophys. J. Lett.* **870** (2019) L15, arXiv:1811.11491 [astro-ph.HE].
66. A. Hajela *et al.*, arXiv e-prints (2021), arXiv:2104.02070 [astro-ph.HE].
67. F. Verrecchia *et al.*, *Astrophys. J. Lett.* **850** (2017) L27, arXiv:1710.05460 [astro-ph.HE].
68. I. Pelisoli, T. R. Marsh, R. P. Ashley, P. Hakala, A. Aungwerojwit, K. Burdge, E. Breedt, A. J. Brown, K. Chanthorn, V. S. Dhillon, M. J. Dyer, M. J. Green, P. Kerry, S. P. Littlefair, S. G. Parsons, D. I. Sahman, J. F. Wild and S. Yotthanathong, *Mon. Not. R. Astron. Soc.* **507** (2021) 6132, arXiv:2109.00553 [astro-ph.SR].

69. I. Pelisoli, T. R. Marsh, V. S. Dhillon, E. Breedt, A. J. Brown, M. J. Dyer, M. J. Green, P. Kerry, S. P. Littlefair, S. G. Parsons, D. I. Sahman and J. F. Wild, arXiv e-prints (August 2021), arXiv:2108.11396 [astro-ph.SR].
70. J. G. Coelho, D. L. Cáceres, R. C. R. de Lima, M. Malheiro, J. A. Rueda and R. Ruffini, *Astron. Astrophys.* **599** (2017) A87.
71. J. A. Rueda, K. Boshkayev, L. Izzo, R. Ruffini, P. Lorén-Aguilar, B. Külebi, G. Aznar-Siguán and E. García-Berro, *Astrophys. J. Lett.* **772** (2013) L24, arXiv:1306.5936 [astro-ph.SR].
72. K. Boshkayev, L. Izzo, J. A. Rueda Hernandez and R. Ruffini, *Astron. Astrophys.* **555** (2013) A151, arXiv:1305.5048 [astro-ph.SR].
73. M. Malheiro, J. A. Rueda and R. Ruffini, *PASJ* **64** (2012) 56, arXiv:1102.0653 [astro-ph.SR].

A semi-implicit multidimensional unstructured gas dynamical solver for astrophysical applications

Ilya A. Kondratyev*, Sergey G. Moiseenko

Space Research Institute RAS, Profsoyuznaya st., 84-32, Moscow, Russia, 117485
** E-mail: mrkondratyev95@gmail.com*

National Research University "Higher School of Economics", Staraya Basmannaya st., 21/4 s. 5, Moscow, Russia, 105066

Astrophysical problems such as modelling of core-collapse supernovae, collapse of protostellar clouds as well as other processes, involving collapsing matter, deal with regions (e.g. protostars, protoneutron stars), where a speed of sound has much larger values, than in remaining parts of a computational domain. A time-step in explicit numerical schemes, thus, has to be bounded by acoustic Courant-Friedrichs-Lewy condition, due to high speed of sound in these compact regions. In some cases, this condition can be very restrictive, and (semi-) implicit numerical schemes may outperform the explicit ones. We propose a semi-implicit solver on a collocated mesh for self-gravitating gas dynamical flows, in which only acoustic waves are treated implicitly. We use an operator-difference approach to construct difference analogues of vector differential operators on unstructured meshes in two and three dimensions, which allows us to save the conjugacy properties of the operators. A Rusanov-type dissipation was used to get monotonic flow profiles and usual linear flux reconstruction to improve an order of spatial approximation. Results of test calculations are presented.

Keywords: Gas dynamics; semi-implicit schemes; self-gravity.

1. Introduction

Numerical simulations of astrophysical fluid flows is one of the key instruments, which allow us to explain a wide range of astronomical phenomena on the qualitative level, as well as to compare the theoretical predictions to observational data quantitatively. An astrophysics is rich for different types of gas and plasma flows in the case, when one can use ideal compressible fluid approximation without dissipation (e.g. Euler equations). The ideal flow is described in terms of hyperbolic systems of partial differential equations, for which the explicit schemes, such as Godunov-type finite volume,[1] can be successfully applied (see also textbooks[2,3]). Explicit SPH solvers are broadly used by astrophysical community as well.[4] But there are the problems, such as, for example, star formation from protostellar clouds, collapses of stars' iron cores and supernovae explosions, where the implicit or semi-implicit schemes may possibly outperform the explicit solvers. It is a common situation in astrophysical modelling, when the compact high-density regions of a collapsed matter with high speed of sound (such as protostars, protoneutron stars) appear in a computational domain, where the velocity of the flow is very low compared

to the acoustic speed, so that we deal with low Mach number flows. At the same time, in other parts of the domain, the flow can be considered as compressible, strong shocks and other different flow discontinuities may form. The latter situation is common for e.g. a post-bounce evolution of core-collapse supernovae.[5] It also stands for magnetorotational supernove simulations, when we have small acoustic timescale and large magnetic field evolution timescale.[6,7] The mesh size during such astrophysical simulations is usually refined in the regions of collapsed objects up to several orders of magnitude, compared to the cell size on the outer boundary. The explicit solvers are restricted by Courant-Friedrichs-Lewy (CFL) condition. In the mentioned types of astrophysical flows, the global CFL condition comes mostly from the high sound speed from resolved collapsed objects and may be very restrictive, which makes such kind of simulations computationally expensive and challenging. Thus, the implicit treatment of acoustic waves in hydrodynamical schemes allows to increase timestep significantly, and hence, to reduce the computational cost of such astrophysical simulations.

Implicit evaluation of the pressure field is a common approach in incompressible flow simulations.[8] The idea to extend the semi-implicit to compressible flows was proposed in mid-80's by Casulli and Greenspan,[9] but only quite recently this approach was used to construct the so-called *All-Mach number solvers* for gas dynamics (see e.g. works[10–13] on structured and[14,15] on unstructured grids, and references therein) and for MHD equations,[16,17] written in conservative form. They are the pressure-based solvers by its nature, but they can work as high resolution shock-capturing methods as well in presence of discontinuities. The usage of semi-implicit solvers results to the solution of an elliptic equation with a Laplace-type operator for the pressure on each time-step, and it is very important to construct a consistent difference analogue of the solving differential problem, which provides the cheaply solvable linearized systems. To obtain a symmetric sign-definite matrices for the linear elliptic problems, we use a Support (basic) Operators Method (operator-difference approach)[19] to construct the difference analogues of differential operators in three dimensions. The support operators method has also many common features with mimetic finite difference[18] approach.

The idea of the support operators method is that the difference operators should be derived in the way to fulfil the same properties as the continuous operators. Thus, the constructed operators should satisfy the difference analogues of integral relations such as Green formula. This approach allows to obtain the so-called completely conservative fully implicit finite-difference schemes for gas dynamics and MHD equations[19] on unstructured grids. For 2D simulations of magnetorotational supernova explosion this method was used e.g. in[20] with a cell-nodal approximation[21,22] in the framework of Newtonian MHD with self-gravity.

The paper is organized as follows. In the next subsection we derive the nodal difference operators of gradient and divergence in three dimensions. In the 3-rd subsection we provide the description of the solving gravitational gas-dynamics

system and derive a semi-implicit numerical scheme for these equations, and in the last subsections we present and discuss the results of numerical simulations of several benchmark problems in low and high Mach number regimes.

2. Difference analogues of differential operators

In the Support operators method the difference analogues of differential operators are built in pairs. Further we are going to consider only the nodal gradient and divergence operators, while the other difference analogues can be obtained in a similar way. To introduce difference GRAD and DIV operators one should define finite dimensional linear grid spaces and scalar products of grid functions in them. The first operator can be obtained using any numerical differencing technique, while the second one has to satisfy the grid analogue of the Green formula in corresponding grid space. Thus, the resulting operators are conjugated to each other. The GRAD operator can be derived in two different ways: either using the finite element approach with linear basis functions inside the tetrahedra (for 2D case see e.g.[23]) or with a finite volume approach using piecewise-constant basis functions on the nodal edge-based (median-dual) cells, as in Godunov-type methods on unstructured grids (see e.g.[24]). The grid gradient operator reads (for 3D case):

$$(\nabla_\times p)_j = -\frac{1}{12W_j}\sum_{i\in G_j}\sum_{k=1}^{4}\left(S_k\mathbf{n}_k p_k\right)_i, \tag{1}$$

where p is a nodal scalar function, G_j is a set of tetrahedral cells adjoined to sought j-th node, i is a cell index, k is a node index inside the i-th cell, $\mathbf{n}_k$ and S_k are unit normal and area to the tetrahedron's triangular face, which does not contain the node k in its vertexes, and $W_j = \frac{1}{4}\sum_{i\in G_j} V_i$ is a "node" volume (the volume of a median-dual cell), V_i is a volume of i-th tetrahedron.

To obtain the consistent approximation of the divergence operator, the Green formula and its difference analogue should be used:

$$\begin{gathered}\int_V p\nabla\cdot\mathbf{v}dV + \int_V(\nabla p)\cdot\mathbf{v}dV = \oint_{\partial V} p\mathbf{v}\cdot\mathbf{dS},\\ \sum_{l=1}^{N_n}(\nabla_\times p)_l\cdot\mathbf{v}_l W_l + \sum_{k=1}^{N_n} p_k(\nabla_\times\cdot\mathbf{v})_k W_k = (\Phi\cdot\mathbf{v}, p),\end{gathered} \tag{2}$$

where ∇ and $\nabla\cdot$ are the gradient and the divergence operators correspondingly, $\mathbf{v}$ is a vector function, N_n is a total number of the nodes, and V is a volume of the computational domain. The surface integral in the right hand side of the formulae (2) correspond to the so-called *Boundary operator* Φ in a formalism of the support operators method. Its evaluation can be found in.[21,22,25] The boundary operator is used to formulate the boundary-value problems in the operator-difference form.

We do not consider it here, because in the solver we use ordinary finite volume boundary conditions via ghost cells approach.

To derive the DIV operator in the interior nodes of the computational domain, the Green formula in (2) is written without the integral in the right hand side. By substituting the grid gradient operator (1) in the analogue of the Green formula (2) and rearranging the terms, the following approximation for the divergence operator can be obtained

$$(\nabla_\times \cdot \mathbf{v})_j = -\frac{1}{12W_j} \sum_{i \in G_j} \sum_{k=1}^{4} \left(S_k \mathbf{n}_k \cdot \mathbf{v}_k\right)_i. \tag{3}$$

We note, that the latter operator has the same form, as the initially derived gradient one (1) due to the usage of purely nodal approximation, while its form and stencil is different from the cell-nodal DIV operator. The divergence operator (3) is conjugated to (1), and thus, the grid Laplacian transforms into the symmetric and sign-definite matrix. The usage of the support operators technique allows to derive the full set of vector calculus operators in any geometry, which satisfy the conjugacy properties.

Finally, the difference analogues of differential operators can be written in a finite-volume form[26]

$$(\nabla_\times^{fw} p)_j = \frac{1}{W_j} \sum_{n \in H_j} \left(\frac{S_{nj} \mathbf{n}_{nj} (p_n + p_j)}{2} \right), \tag{4}$$

where H_j is a set of neighbouring nodes to the sought node with index j, S_{nj} and $\mathbf{n}_{nj}$ are the area and the unit normal to the n-th face of the node-centred median-dual control volume. This approximation leads to exactly the same expression as (1).[26]

3. Semi-implicit numerical scheme for the equations of gas dynamics with gravitation

3.1. *Euler equations with self-gravity*

Gas dynamical equations with self-gravity can be written in the following conservative formulation:

$$\begin{gathered} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \\ \frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + \nabla P = -\rho \nabla \Psi, \\ \frac{\partial \rho E}{\partial t} + \nabla \cdot \left(\rho \mathbf{v} \big(E + \frac{P}{\rho}\big) \right) = -\rho \mathbf{v} \cdot \nabla \Psi, \\ \nabla \cdot \nabla \Psi = 4\pi G \rho. \end{gathered} \tag{5}$$

In the system above, the first three equations without source terms in right hand sides correspond to ideal Euler equations, while the last one is the Poisson equation for a gravitational potential Ψ. The source terms correspond to presence of an acceleration by gravity, G is a gravitational constant. In ideal Euler equations (5) ρ is a mass density, $\mathbf{v}$ is a flow velocity; P is a pressure, $E = \frac{\mathbf{v}^2}{2} + E_{th}$ is a total energy density, E_{th} is an internal energy density. A thermodynamic equation of state is needed to close the system of equations (5). In this paper we consider only a perfect gas equation of state with an adiabatic index γ:

$$P = (\gamma - 1)\rho E_{th}. \tag{6}$$

Extension of a semi-implicit solver (see next subsection) to an arbitrary equation of state is straightforward and can be found in e.g.[11,12]

3.2. *Poisson solver*

The Poisson equation should be solved on each time-step to calculate the value of the gravitational potential, and hence, a gravitational force. The support operators method allows to construct a discrete Laplace operator in the form of a symmetric and sign-definite matrix. Purely nodal approximation requires the stencil, which consists of the second-order neighbours for each node. The boundary conditions are projected to the boundary nodes and included in the operator-difference form of the solving problem. For example, the Dirichlet boundary-value problem reads[21,22]

$$(I - \delta)\nabla_\times \cdot \nabla_\times (I - \delta)\Psi + (I - \delta)\nabla_\times \cdot \nabla_\times \delta\Psi_\gamma = 4\pi G(I - \delta)\rho, \tag{7}$$

where Ψ_γ is a boundary value of the gravitational potential, I is a unit operator, δ is an operator, which is equal to unity in the boundary nodes, and zero - in the interior ones.[22] For other types of boundary conditions (e.g. an absence of the force on the boundary) we solve the Poisson using the ghost cells approach. The resulting linear system can be solved with a matrix-free conjugate gradient method. The gravitational acceleration then can be calculated as a nodal gradient of the potential.

In Fig. 1 the solution of the Poisson equation is shown for the spherically-symmetric Gaussian-law solution $\Psi = \exp\big(-\frac{x^2+y^2+z^2}{2}\big)$ given by the function $f = \exp\big(-\frac{x^2+y^2+z^2}{2}\big)(x^2 + y^2 + z^2 - 3)$ in the right-hand side of the equation.

The cell-node approximation can also be used for the solution of the Poisson equation procedure. It allows to use a smaller stencil for the matrix, while the gradient of the potential still can be found with a purely nodal operator (4).

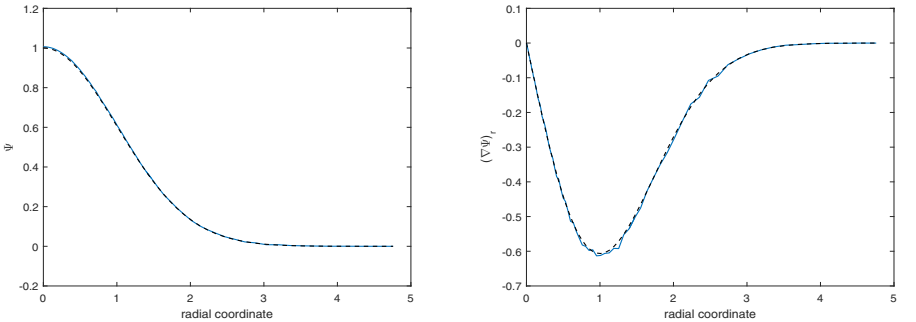


Fig. 1. The solution of the Poisson equation (blue lines, 1D plots are given for the X-axis) for the potential (left) and its gradient (right) together with an analytical solution (black lines). The computational domain is cube ($0 < x, y, z < 5$) with no force conditions on the symmetry boundaries and Dirichlet conditions on outer ones.

3.3. *Hydro-solver*

The difference gas dynamical system of equations can be written in the following form on a collocated mesh

$$\begin{gathered}
\frac{\rho^{n+1}-\rho^n}{\tau} + \widetilde{\nabla}_\times \cdot (\rho^n \mathbf{v}^n) = 0, \\
\frac{\rho^{n+1}\mathbf{v}^{n+1}-\rho^n\mathbf{v}^n}{\tau} + \widetilde{\nabla}_\times \cdot (\rho^n \mathbf{v}^n \otimes \mathbf{v}^n) + \nabla_\times P^{n+1} = \mathbf{S}_v, \\
\frac{\rho^{n+1}E^{n+1}-\rho^n E^n}{\tau} + \widetilde{\nabla}_\times \cdot \left(\rho^n \mathbf{v}^n \frac{(\mathbf{v}^2)^n}{2}\right) + \nabla_\times \cdot \left(h^{n+1}\rho^{n+1}\mathbf{v}^{n+1}\right) = \mathrm{S}_e,
\end{gathered} \tag{8}$$

where $h = E_{th} + \frac{P}{\rho} = \frac{\gamma P}{(\gamma-1)\rho}$ is a thermodynamic enthalpy. This system describes the flow evolution from time $t = t^n$ to $t = t^{n+1} = t^n + \tau$. The last terms in the left parts of the second and third equations in (8) should be evaluated implicitly according to,[9,11] which allows to treat implicitly only the acoustic waves in the gas dynamical system (5). The source terms are given by the formulae $\mathbf{S}_v = -\rho^{n+1}\nabla_\times\Psi$, $S_e = -\rho^{n+1}\mathbf{v}^n \cdot \nabla_\times\Psi$. The gravitational potential is solved for the density distribution on the $(n+1)$-th time-step. Note, that such separate treatment of acoustic waves is equivalent to the Toro-Vazquez flux splitting approach.[27]

The difference operator with tilde $\widetilde{\nabla}_\times\cdot$ in explicitly evaluated advective fluxes corresponds to an upwind monotonic scheme for advection. It is written together with a Rusanov-type numerical dissipation:

$$(\widetilde{\nabla}_\times \cdot \mathbf{F})_j = (\nabla_\times \cdot \mathbf{F})_j - \frac{1}{2W_j}\sum_{i\in H_j}\left(\left|\frac{\partial \mathbf{F}}{\partial \mathbf{u}}\right||\mathbf{n}_{ij}|S_{ij}(u_i - u_j)\right), \tag{9}$$

where dissipation is written in a finite-volume form (4), in which $\mathbf{u}$ and $\mathbf{F}$ are the vector of conservative variables and the flux vector correspondingly:

$$\begin{aligned} \mathbf{u} &= (\rho, \rho\mathbf{v}, \rho E) \\ \mathbf{F} &= (\rho\mathbf{v}, \rho\mathbf{v} \otimes \mathbf{v}, \rho\mathbf{v}E). \end{aligned} \tag{10}$$

We use $|\frac{\partial \mathbf{F}}{\partial \mathbf{u}}| = \max(|\mathbf{v}|_i, |\mathbf{v}|_j)$, thus, dissipation coefficient is independent on the speed of sound.[11] The explicit advective scheme is implemented in the same manner, as the edge-based Godunov-type finite volume solvers on median-dual grids,[28] so that the advective fluxes $\mathbf{F}$ in (9) are calculated at the centres of the edges between the nodes.

The numerical scheme is conditionally stable with a milder CFL condition, which depend only on the flow velocity and not on the speed of sound:

$$\tau < \tau_{adv} = \min \frac{\Delta X}{2 \mid \mathbf{v} \mid}, \tag{11}$$

where ΔX is a characteristic local mesh size.

By substituting the momentum from the momentum equation in (8) to the energy equation, we can obtain the following nonlinear equation:

$$\begin{aligned} &\frac{P^{n+1}}{\gamma - 1} - \tau^2 \nabla_\times \cdot \left(h^{n+1} \nabla_\times P^{n+1} \right) \\ &= \rho^n E^n - \rho^{n+1} \frac{(\mathbf{v}^2)^{n+1}}{2} - \tau \widetilde{\nabla}_\times \cdot \left(\rho^n \mathbf{v}^n \frac{(\mathbf{v}^2)^n}{2} \right) + \tau S_e \\ &-\tau \nabla_\times \cdot \left(h^{n+1} (\rho^n \mathbf{v}^n - \tau \widetilde{\nabla}_\times \cdot (\rho^n \mathbf{v}^n \otimes \mathbf{v}^n) + \tau \mathbf{S}_v) \right), \end{aligned} \tag{12}$$

The latter system is nonlinear, and the usual approach to solve it is a usage of nested Newton-type iterations (r-iterations, Picard iterations),[11,12] which converge rapidly (usually 2-3 iterations are enough). The resulting linearised system reads:

$$\begin{aligned} &\frac{P^{n+1,r+1}}{\gamma - 1} - \tau^2 \nabla_\times \cdot \left(h^{n+1,r} \nabla_\times P^{n+1,r+1} \right) \\ &= \rho^n E^n - \rho^{n+1} \frac{(\mathbf{v}^2)^{n+1,r}}{2} - \tau \widetilde{\nabla}_\times \cdot \left(\rho^n \mathbf{v}^n \frac{(\mathbf{v}^2)^n}{2} \right) + \tau S_e \\ &-\tau \nabla_\times \cdot \left(h^{n+1,r} (\rho^n \mathbf{v}^n - \tau \widetilde{\nabla}_\times \cdot (\rho^n \mathbf{v}^n \otimes \mathbf{v}^n) + \tau \mathbf{S}_v) \right), \end{aligned} \tag{13}$$

$$h^{n+1,r} = \frac{\gamma P^{n+1,r}}{(\gamma - 1)\rho^{n+1}}.$$

To obtain the solution from n-th time level to $n+1$-th, the following procedures should be done:

(I) solving the mass conservation equation in (8) with the explicit scheme and obtaining ρ^{n+1}

(II) solving the Poisson equation (7) with the density ρ^{n+1} for the gravitational potential and calculating the forces in the source terms of gas-dynamical system (8)
(III) defining $P^{n+1,1} = P^n$, $h^{n+1,1} = h^n$ and $\mathbf{v}^{n+1,1} = \mathbf{v}^n$
(IV) r-iterations, $r = \overline{1, r_p}$

  (a) Solving (13) and obtaining $P^{n+1,r+1}$
  (b) recalculating $h^{n+1,r+1}$ and calculating $\mathbf{v}^{n+1,r+1}$ from the momentum equation (8) using $P^{n+1,r+1}$

(V) defining $P^{n+1} = P^{n+1,r_p}$, $h^{n+1} = h^{n+1,r_p}$ and $\mathbf{v}^{n+1} = \mathbf{v}^{n+1,r_p}$
(VI) calculating the energy using a semi-implicit equation from the system (8).

This procedures allow to obtain a numerical solution of the gas-dynamical system (5) with a first order in space and time. In this study we use $r_p = 3$ Picard iterations.

To extend this scheme to higher order in space, we use the same approach as in explicit finite volume Godunov-type schemes on unstructured meshes, as it is done in e.g.[28] (see also[29] for a different approach). We apply a piecewise-linear reconstruction of conservative gas-dynamical variables in advection upwind fluxes (9) with Barth-Jespersen slope limiters,[24] thus, the explicit advective part of the solver is calculated with the second order in space. For this reason, the resulting order of spatial approximation of the scheme is higher, than one. Note, that the difference operators like (4) provide the second order in space on the mesh, consisting on equilateral tetrahedra (or equilateral triangles in 2D), because these operators are analogous to the usual central differences on the uniform grid, so that on such kind of meshes the solver has the second approximation order in space. High order in time approximation can be achieved either by employing explicit high order methods for the advective parts of the solving system, as was done in e.g.,[15] or by using the implicit-explicit Runge-Kutta methods.[13]

The linearised problem (13) has a matrix of a Laplace-type second order difference operator and, in addition, a diagonal, filled with the positive elements. The conjugated difference operators allow to obtain an easily reversible symmetric and positive definite matrix. We use a matrix-free conjugate gradient method to solve the system (13), which is very efficient.

It is also possible to make a discretization of the system (5) in time, substitute the expression for $\rho^{n+1}\mathbf{v}^{n+1}$ in the energy equation, and further discretize the energy equation in space.[13] The cell-node approximation and the Support operators method can be applied for the Laplace-type operator in (13), thus, making the matrix stencil more compact. Preliminary calculations show, that this approach is slightly more effective, as well as it provides a collocated pressure system, which is similar in properties to the staggered grid approach.[13]

4. Test computations

The code was developed in two variants: 2D and 3D Cartesian versions. We have tested our code in several benchmark problems in high and low Mach number regimes as well as simulated a test problem of a dust collapse to check the code's ability to simulate the flows with self-gravity.

In Fig. 2 the Sod shock tube problem solution is shown for the 3D (the first three pictures) and 2D versions of our code at the time $T = 0.45$ for the mesh with $N_x = 200$ nodes in X-direction and $N_y = N_z = 10$ ones in Y- and Z-directions. The initial conditions are as usual: $\rho(x < 1) = 0.125, P(x < 1) = 0.1$ and $\rho(x \geq 1) = 1, P(x \geq 1) = 1$, while the velocity is zero everywhere. The shock profile is about 4-5 cells thick. The solution is obtained by the conservative scheme, thus, the discontinuities have physically correct amplitudes and velocities.

In Fig. 3 the density profile of a slowly moving contact discontinuity is presented. The initial conditions are the following: $\mathbf{v} = (0.01, 0, 0), P = 3$ everywhere, while the density $\rho(x < 0.25) = 50, \rho(x \geq 0.25) = 1$. The problem is solved until the final time $T = 50$, which corresponds to the contact discontinuity located at $x = 0.75$.

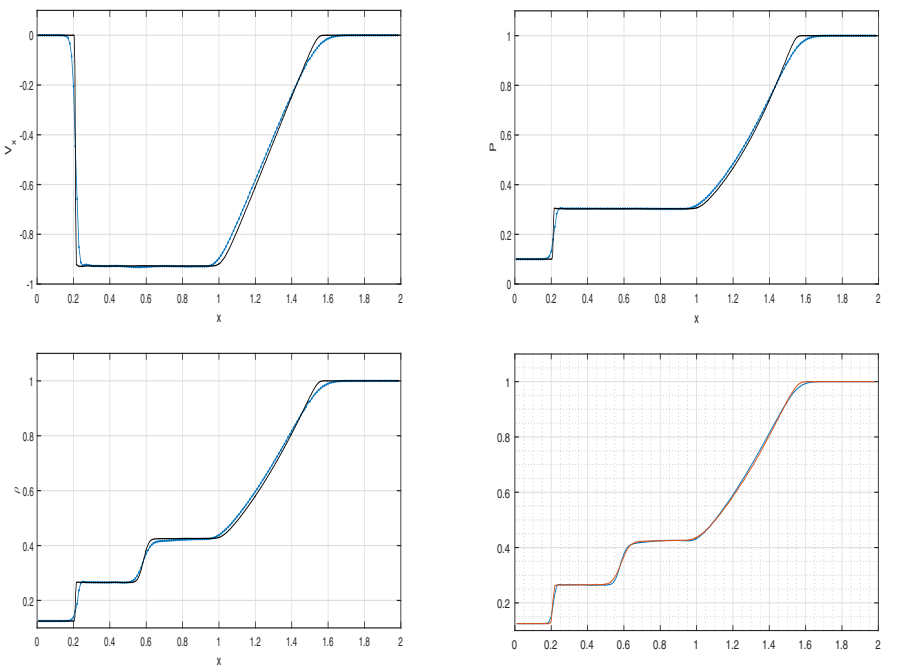


Fig. 2. Velocity, pressure and density fields (the first three figures, blue lines) of the gas in the Sod shock tube at time $T = 0.45$, compared to the reference solution given by a second-order explicit second-order local Lax-Friedrichs solver with Barth-Jespersen slope limiters (black lines) on a finer mesh with $N_x = 500$ nodes. On the last picture the density plot is shown for 2D explicit (red line) and 2D semi-implicit (blue line) solvers with the same mesh resolution.

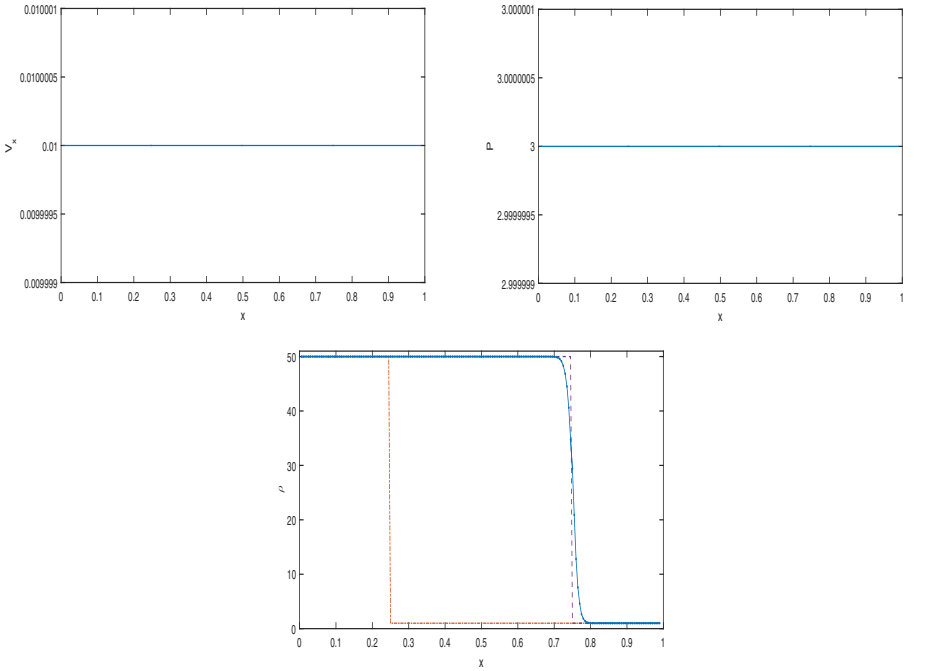


Fig. 3. Velocity, pressure and density fields of the gas in a slowly moving contact discontinuity problem at time $T = 50$. The blue lines correspond to our numerical solution, while the red dash-dotted line and dashed purple line are initial and final location of the contact discontinuity correspondingly.

The profile is smeared over ~ 10 grid points. On this problem, the semi-implicit scheme greatly outperforms the explicit solver by a factor of ~ 40. The time step τ_{adv} (11) is equal to $\sim 200\tau_{CFL}$.

The next test is the Sedov-Taylor planar blast wave solution, which was performed in a 2D cylindrical geometry. The shock front location is $R \approx \sqrt{T}$. Fig. 4 shows the density and pressure profiles at $T = 0.2$. The 2D mesh, used in the simulations, is refined at the centre of explosion by a factor of 10. It has 8756 nodes. The numerical solution shows a good agreement with analytics. In the refined central part ($r = \sqrt{x^2 + y^2} \ll 1$) of the computational domain, the sound speed is greatly exceeds the velocity, while at $r \lesssim 1$ the strong high Mach number shock wave is pronounced. At the end of the simulation, the advective time-step was about 50 times higher, than the CFL time-step τ_{CFL}. In real astrophysical conditions this ratio of time-steps could be even higher.

The latter situation is more pronounced in 3D Sedov-Taylor blast wave problem, where the semi-implicit calculation (see Fig. 5) was at least 10 times faster, than the explicit solver, due to restrictive acoustic time-step in the centre of computational domain.

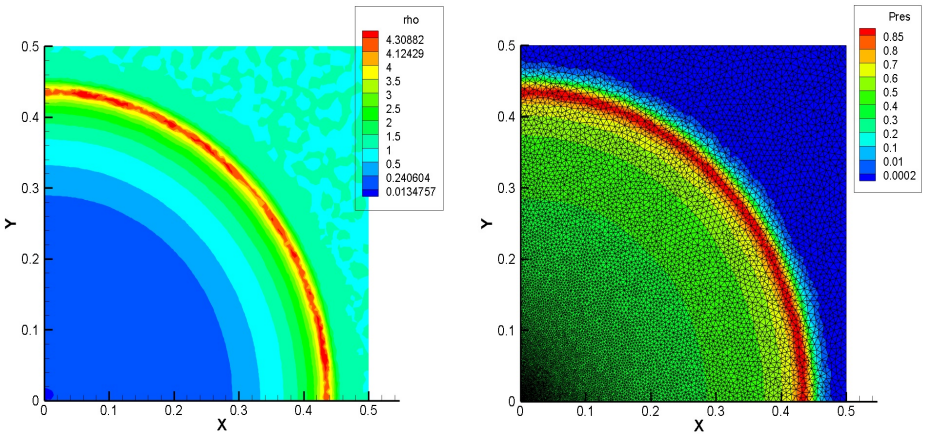


Fig. 4. Density (left) and pressure (right, together with the 2D mesh) of the gas in a planar Sedov-Taylor blast wave problem at time $T = 0.2$.

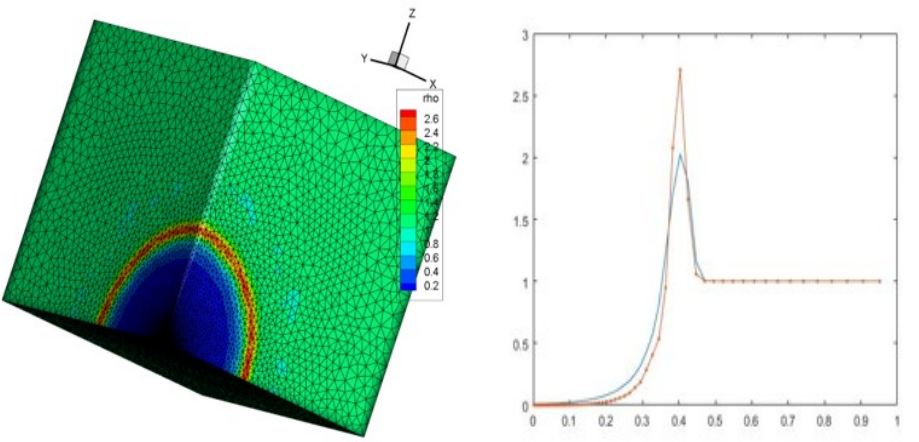


Fig. 5. Density (left) and its 1D slice (right, red line, together with the explicit solver calculation) of the gas in a 3D Sedov-Taylor blast wave problem at time $T = 0.1$.

We have tested our code including the self-gravity module and simulated a well-known problem on the spherical collapse of a pressureless star (dust collapse), which was originally considered in the paper by Colgate and White.[30] The octant-symmetric 3D results for the density at dimensionless time $T = 1.791$ are presented in Fig. 6. We use the system of units, which removes the multiplier $4\pi G$ in the Poisson equation for the gravitational potential (7). In these units the free-fall time $t_{ff} = \sqrt{\frac{3}{2}}\frac{\pi}{2} \approx 1.9238$. The initial cloud radius and density are equal to unity. We use the mesh, which is refined to the centre of the computational domain. The snapshot

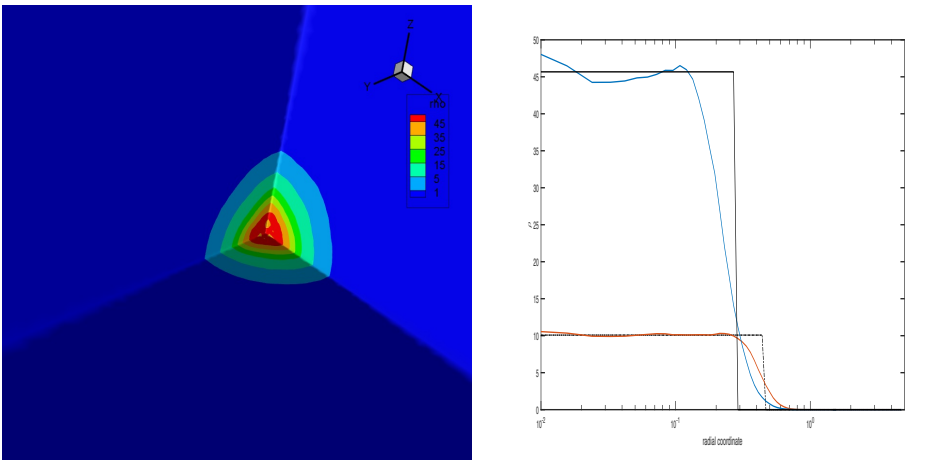


Fig. 6. Density of the gas in the dust collapse problem at time $T = 1.791$. Left panel correspond to 3D solution, while the right panel corresponds to 1D spherical plot (blue line - numerical solution, black dashed line - analytics) together with the solution at the time $T = 1.618$ (red line - numerical solution, black dotted line - analytics).

of our simulation in Fig. 6 shows a physically correct evolution of collapsing matter, although at the developed stage of the collapse the density field has a discrepancy from the analytical solution (blue line in Fig. 6, right) due to the high spatial and temporal steps. The initial spherically-symmetric conditions are also disturbed by unstructured mesh, and these disturbances tend to slight deviation from spherical symmetry.

5. Conclusion

In this paper we have constructed the semi-implicit nodal (edge-based) pressure-based solver for gravitational gas dynamics on unstructured tetrahedral meshes. The usage of the Support operators technique results to linear systems with symmetric and sign-definite matrices for the pressure system in semi-implicit gas-dynamical solver as well as for the Poisson equation for the gravitational potential. The code works efficiently both in low and high Mach number regimes, and its spatial accuracy is close to the second order. In future, we plan to apply this code to simulation of the post-bounce evolution of core-collapse supernovae. Further development of the code will be devoted to improvement of temporal accuracy, massive parallelization and inclusion of magnetic fields.

References

1. Godunov S.K., Mat. Sbornik, 47(89):3, 271 (1959)
2. Bisikalo D.V., Zhilkin A.G. and Boyarchuk A.A. “Gas dynamics of close binary stars”, Moscow-Fiz.Mat.Lit. (in Russian) (2015)

3. Toro E.F., "Riemann Solvers and numerical methods for fluid dynamics: A practical introduction", Springer (2009)
4. Monaghan J.J., Annu. Rev. Astron. Astrophys., 30:543-74 (1992)
5. B. Mueller, Living reviews in computational astrophysics, 6, 3 (2020)
6. Ardeljan N.V., Bisnovatyi-Kogan G.S. and Moiseenko S.G., Mon. Not. R. Astron. Soc., 359, 333 (2005)
7. Obergaulinger M., Aloy M.Á., Mon. Not. R. Astron. Soc., 492 (4), 4613 (2020)
8. Harlow F.H. and Welch J.E., Physics of Fluids, 8, 2182 (1965)
9. Casulli V. and Greenspan D., International Journal for Numerical Methods in Fluids, 4, 11, 1001 (1984)
10. Cordier F., Degond P. and Kumbaro A., Journal of Computational Physics, 231, 5685 (2012)
11. Dumbser M., Casulli V., Applied Mathematics and Computation, 272, 479 (2016)
12. Boscheri W., Dimarco G., Loub'ere R., Tavelli M. and Vignal M.H., Journal of Computational Physics, 415, 109486 (2020)
13. Boscheri W. and Pareschi L., Journal of Computational Physics, 434, 110206 (2021)
14. Tavelli M. and Dumbser M., Journal of Computational Physics, 341, 341 (2017)
15. Busto S., Río-Martín L., Vázquez-Cendón M.E. and Dumbser M., Applied Mathematics and Computation, 402, 126117 (2021)
16. Dumbser M., Balsara D.S., Tavelli M. and Fambri F., International Journal for Numerical Methods in Fluids, 89 (1-2) 16 (2019)
17. Fambri F., arXiv:2012.11218 (2021)
18. Lipnikov K., Manzini G. and Shashkov M., Journal of Computational Physics, 257, 1163 (2014)
19. Samarskii A.A. and Popov Yu.P., "Difference methods for solving problems of gas dynamics" Moscow, Izdatel'stvo Nauka (in Russian) (1992)
20. Moiseenko S.G., Ardeljan N.V. and Bisnovatyi-Kogan G.S., Monthly Notices of the Royal Astronomical Society, 370 (1) 501 (2006)
21. Ardeljan N.V. and Kosmachevskii K.V., "Problems of Construction and Research of Conservative Difference Schemes for Magneto-gas-dynamics" Moscow - Moscow State University (in Russian) (1987)
22. Ardeljan N.V. and Kosmachevskii K.V., Comput. Math. Modelling,6, 209 (1995)
23. Ardelyan N.V. and Sablin M.N., Comput. Math. Modelling, 14, 217 (2003)
24. Barth, T.J. and Jespersen D.C., The design and application of upwind schemes on unstructured meshes, AIAA Paper, 89-0366 (1989)
25. Kondratyev I.A. and Moiseenko S.G., Journal of Physics: Conf. Series, 1163, 012069 (2019)
26. Barth T.J., Numerical Aspects of Computing Viscous High Reynolds Number Flows on Unstructured Meshes, AIAA paper, 91-0721 (1991)
27. Toro E.F., Vázquez-Cendón M.E., Computers and fluids, 70, 1 (2012)
28. Barth T.J., A 3-D Upwind Euler Solver for Unstructured Meshes, AIAA paper, 91-1548CP (1991)
29. Kulikov I. Vorobyov E., Chernykh I. and Elbakyan V., Journal of Physics: Conf. Series, 1640, 012003 (2020)
30. Colgate S.A., White R.H., Astrophys. J., 143, 626 (1966)

Magnetized neutron stars propagating through a non-uniform ISM

O. D. Toropina

Space Research Institute, Russian Academy of Sciences
Profsoyuznaya 84/32, Moscow 117997, Russia
E-mail: toropina@iki.rssi.ru

M. M. Romanova and R. V. E. Lovelace

Department of Astronomy, Cornell University,
NY, 14853, Ithaca, USA

Many neutron stars propagate through the interstellar medium with supersonic velocities, and their magnetospheres interact with the interstellar medium (ISM), forming bow shocks and magnetotails. Using numerical MHD simulations, we investigated the propagation of a magnetized neutron stars through a non-uniform ISM, the interaction of the magnetospheres with the ISM and the influence of ISM density on the shape of the magnetosphere tail. We consider the interaction of magnetized neutron stars with small-scale and large-scale inhomogeneities in the ISM. We conclude that the inhomogeneities in the ISM can change the shapes of the bow shocks and magnetotails at different values of the magnetization.

Keywords: Neutron stars; magnetic field; PWN; ISM; MHD.

1. Introduction

The most dramatic bow shock nebulae are associated with the neutron stars. Pulsars emit winds of relativistic particles and magnetic fields, and are frequently surrounded by the pulsar wind nebulae. Most pulsars propagate supersonically through the interstellar medium, and their PWNs interact with the ISM, forming bow shocks and magnetotails. The bow shocks are often observed in the H_α spectral line.[2] Many interesting structures (bow shocks, very long tails, and jet-like features) are observed in the X-ray[5] and radio wavebands. A remarkable example of PWN is the Guitar Nebula discovered in 1992. It's powered by an ordinary NS, PSR B2224+65. This pulsar travels at an extraordinarily high velocity, about 1600 km/sec. It leaves behind a "tail" in the ISM, that looks like a guitar. The high-resolution observations in the H_α line show that the shape of the Nebula's head becomes wider with time. The variation in shape may be connected with the variation in the density of the ISM.[4]

Many of the PWN show irregularities in their shapes in the X-ray band, as well as very long tails (see review by Kargaltsev et al.[7]). Longer tails are observed in the radio band.[8] Two examples of PWN, observed in the X-ray and radio bands are presented on the Figure 1: the PWN associated with the pulsar PSR J1509-5850

and the Mouse Nebula, powered by the pulsar PSR J1747-2958.[6] These PWNs have extended tails in the X-ray band (red lines), and longer tails in the radio-band (blue lines). One can see that both the heads of the PWN and their tails have different unevenness in shapes.

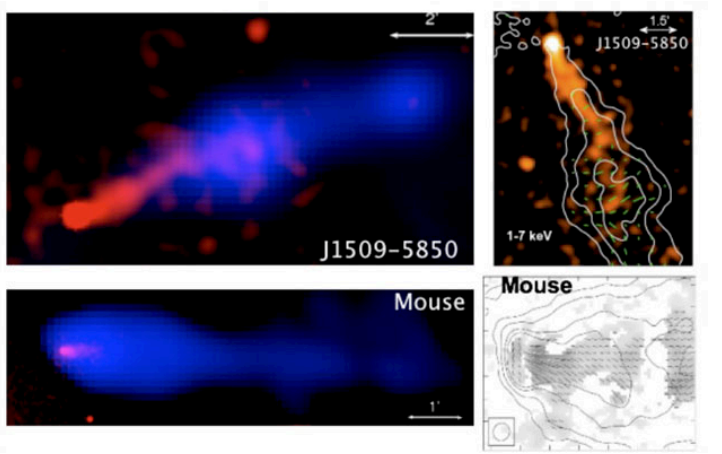


Fig. 1. X-ray and radio images of the very long pulsar tails, by Kargaltsev and Pavlov.[6] Right panels show the radio contours and the direction of the magnetic field. The red and blue colours in the left panels correspond to X-ray and radio, respectively.

When a pulsar or an isolated neutron star propagates supersonically through the ISM, then the PWN or magnetosphere of isolated star interacts with the ISM, forming a bow shock and magnetotails. In the bow shock, the energy of accelerated particles may dominate over magnetic energy. However, in the magnetotails, the magnetic energy may be comparable to or larger than the energy of the particles. Long and magnetically dominated magnetotails can to form in the PWN. They may be visible, if the accelerated particles propagate into the magnetotails, or invisible. In the simulation of the supersonic PWN, it is important to take into account both, the matter and the magnetic field components of the PWN.

Supersonic propagation of pulsars and isolated neutron stars through the ISM has been studied in a number of axisymmetric non-relativistic and relativistic hydrodynamic simulations (e.g., Bucciantini,[3] Barkov[1]). Most of the these simulations of the bow shocks propagating through an inhomogeneous ISM medium were performed in hydrodynamic approximation. Early we performed MHD simulation of supersonic propagation of magnetized isolated neutron stars in the homogeneous ISM.[9,10] Then we started MHD simulation of the bow shock PWN propagating through a non-uniform ISM.[12] Our aim is to consider this problem in axisymmetric MHD simulations at different magnetization values of the pulsar wind.

2. Model

We performed MHD simulations to investigated the supersonic propagation of magnetized isolated neutron star and PWNs through the non-uniform ISM. We used an axisymmetric, resistive MHD code, that incorporates the methods of local iterations and flux-corrected-transport.[13] The flow is described by the resistive MHD equations:

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\ \mathbf{v}) = 0\ ,$$

$$\rho\frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \frac{1}{c}\mathbf{J}\times\mathbf{B} + \mathbf{F}^g\ ,$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}) + \frac{c^2}{4\pi\sigma}\nabla^2\mathbf{B}\ ,$$

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \nabla\cdot(\rho\varepsilon\mathbf{v}) = -p\nabla\cdot\mathbf{v} + \frac{\mathbf{J}^2}{\sigma}\ . \tag{1}$$

We assume axisymmetry ($\partial/\partial\phi = 0$), but calculate all three components of $\mathbf{v}$ and $\mathbf{B}$. The equation of state for an ideal gas, $p = (\gamma-1)\rho\varepsilon$, where $\gamma = 5/3$ is the specific heat ratio and ε is the specific internal energy of the gas. The equations incorporate Ohm's law $\mathbf{J} = \sigma(\mathbf{E}+\mathbf{v}\times\mathbf{B}/c)$, where σ is the electric conductivity. The associated magnetic diffusivity, $\eta_{\rm m} \equiv c^2/(4\pi\sigma)$, is assumed to be a constant. The gravitational force, $\mathbf{F}^g = -GM\rho\mathbf{R}/R^3$, is due to the star because the self-gravity of the accreting gas is negligible.

We use a cylindrical, inertial coordinate system (r, ϕ, z) with the $z-$axis parallel to the star's dipole moment μ and rotation axis $\mathbf{\Omega}$. The vector potential $\mathbf{A}$ is calculated so that automatically $\nabla\cdot\mathbf{B} = 0$ at all times. The intrinsic magnetic field of the star is taken to be an aligned dipole, with vector potential $\mathbf{A} = \mu\times\mathbf{R}/R^3$. A description of numerical model in detail can be found in.[11,12]

We measure length in units of the Bondi radius $R_B \equiv GM/c_s^2$, with c_s the sound speed at infinity, density in units of the density of the interstellar medium ρ, and magnetic field strength in units of B_0 which is the field at the pole of the numerical star. We measure time in units of $t_0 = (Z_{max} - Z_{min})/v_\infty$, which is the crossing time of the computational region with the star's velocity, v_∞.

3. Propagation of magnetized isolated neutron stars through non-uniform ISM

First, we consider a supersonic propagation of magnetized, isolated old neutron stars (IONS) through the non-uniform ISM, in particular through the a dense and cold cloud. Simulations were done for a number of values of the star's magnetic moments μ, the velocity of the star v and the density of the ISM cloud ρ. We received pictures

of the flow of matter for different parameters μ, v and ρ. For all used parameters a magnetospheric radius of the star is much larger than accretion radius $R_m >> R_{acc}$, so that matter cannot accretes onto neutron star surface. This case corresponds to the georotator regime for isolated neutron stars.

When a neutron star moving through an uniform medium with the density ρ_0, incoming matter forms a conical shock wave around the star. The magnetic field lines stretches in the direction of flow of matter out of the magnetospheric radius R_m and form a long tail of the magnetosphere. A picture of the matter flow one can see on top panel of the Figure 2. We observed a reconnection of magnetic field lines in the tail. Then a star passes through the non-uniform medium which is a cold and dense cloud, and the shape of the bow shock and the tail of the magnetosphere varies depending on the density and temperature of the ISM. The shape of the bow shock changes its width and the shape of the magnetotail gets multiple "bumps".

A typical example of the matter flow is represented on Figure 2. In this case Mach numbers $\mathcal{M} = v/c_s = 6$ and density of the cloud $\rho/\rho_0 = 6$. A central region of calculation at moments $t_1 = 0$, $t_2 = 0.33t_0$, $t_3 = 0.66t_0$ and $t_4 = 1.0t_0$ is shown, when t_0 is the crossing time of the computational region. On the left panel the background represents the logarithm of the density and solid lines are magnetic field lines. The length of the arrows is proportional to the poloidal velocity of the matter. We can see that with increasing density of the cloud the width of the bow shock and width of the tail decreases. Then with decreasing density of of the cloud the magnetosphere recovers its shape. Magnetic field lines are compressed by the dense incoming flow and form new reconnections in the tail. On the right panel the background represents the logarithm of the temperature. We can see that with decreasing temperature of the cloud the width of the bow shock and width of the tail decreases. Then with increasing temperature of of the cloud the magnetosphere recovers its shape.

Left panel on Figure 3 shows a density distribution across the magnetotail on the distance $z = 0.5$ and at the same time $t = 0.33t_0$, A solid line represents an uniform medium with constant density ρ_0. A maximum of the density correspond to the distance $r = 0.37$ approximately. A dotted line represents the cloud density $\rho/\rho_0 = 3$. A maximum of the density correspond to $r = 0.26$. A dashed line represents the cloud density $\rho/\rho_0 = 6$. A maximum of the density correspond to $r = 0.20$. We can see that with an increase in the density of the ISM the tail of the magnetosphere becomes more narrow and hollow.

Right panel on Figure 3 shows a temperature distribution across the magnetotail on same distance $z = 0.5$ and at the same time $t = 0.33t_0$. The solid line represents an uniform medium with constant density ρ_0. The dotted line represents the cloud density $\rho/\rho_0 = 3$ and a dashed line represents the cloud density $\rho/\rho_0 = 6$. For all three values of the density, the temperature is maximum at the Z axis, then decreases along R axis and forms a "step", and then becomes a constant. We can see that with a decrease in the temperature of the ISM the tail of the magnetosphere becomes more narrow and hollow.

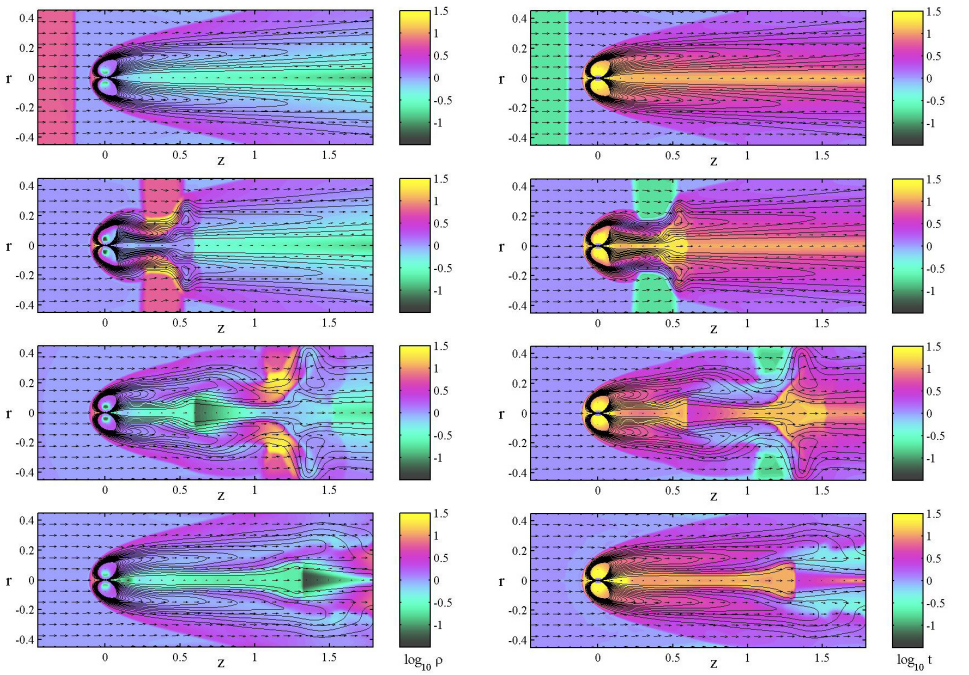


Fig. 2. The propagation of a neutron star with a magnetic field through the dense and cold cloud of interstellar medium. Mach numbers $\mathcal{M} = 6$ and dencity of the cloud $\rho/\rho_0 = 6$. Central region of matter flow at moments $t_1 = 0$, $t_2 = 0.33t_0$, $t_3 = 0.66t_0$ and $t_4 = 1.0t_0$ is shown. On the left panel the background represents the logarithm of the density the background represents the logarithm of density. Solid lines are magnetic field lines. The length of the arrows is proportional to the poloidal velocity. On the right panel the background represents the logarithm of the temperature.

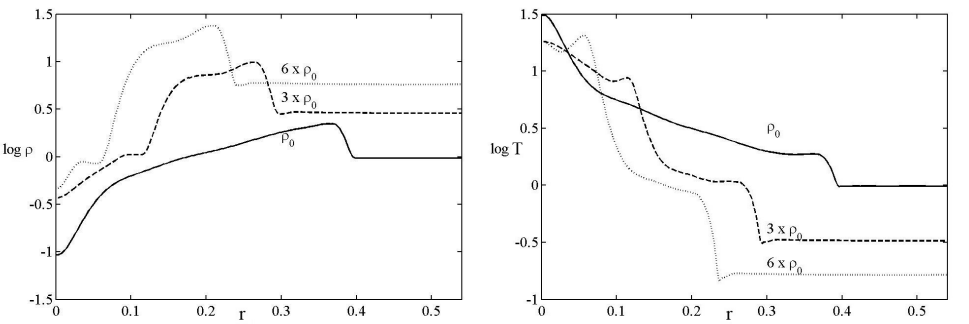


Fig. 3. Left panel: Distribution of the density across the magnetotail at the same time $t = 0.33t_0$ on the distance $z = 0.5$. Right panel: Distribution of the temperature across the magnetotail at the same time $t = 0.33t_0$ on the distance $z = 0.5$.

4. Propagation of PWN through the non-uniform ISM

We investigate the propagation of a PWN through an ISM with an inhomogeneous matter distribution. Observations point to large and small-scale inhomogeneities, and therefore we consider two types of imhomogeneities: (1) large-scale clouds, which are much larger than the size of the bow shock, (2) small-scale clouds, whose size is comparable with the width of the bow shock.

4.1. *Large-scale cloud*

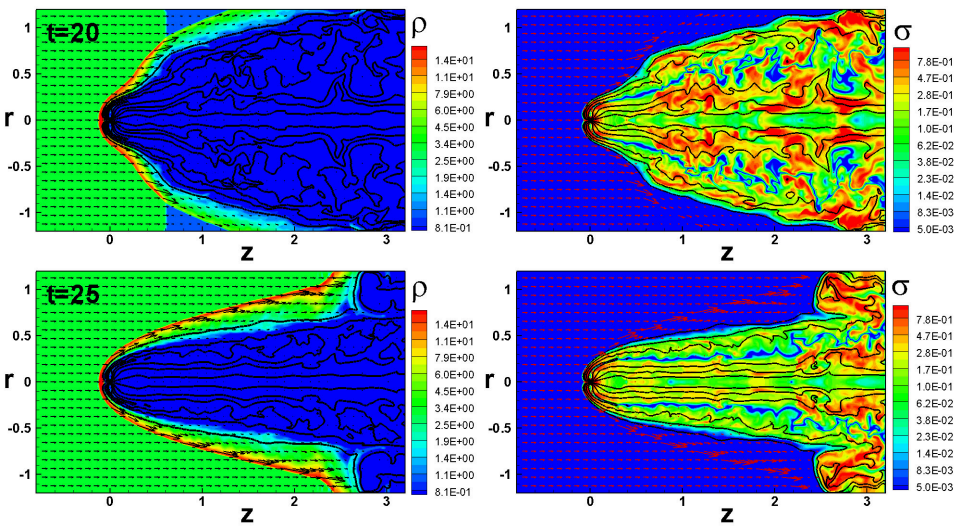


Fig. 4. *Left panels:* Propagation of a PWN of moderate magnetization through a large cloud of density $\rho_{\rm cloud} = 3\rho_0$ at two moments in time. The background represents the logarithm of density. The solid lines are magnetic field lines. *Right panels:* Same, but the background shows magnetization σ.

Fig. 4 shows an example of propagation of a large cloud of density $\rho_{\rm cloud} = 3\rho_0$ through the PWN in the model with a moderate magnetization. The left panels show that the cloud compresses the bow shock and the bow shock has a smaller opening angle (the Mach cone). The cloud is in pressure balance with the rest of the ISM, so that the sound speed in the cloud is three times lower than in the rest of the ISM. Therefore, the Mach number of the star inside the cloud is $M_{\rm new} = v_p/c_s = 3M_{\rm old} = 60$. This is why we observe a smaller opening angle of the Mach cone. The bottom left panel shows that after entering the cloud, the KH instability develops in the region of CD. This is an expected result: at larger Mach number, there is a larger difference between velocities of matter from two sides of the contact discontinuity. The right panels of Fig. 4 show that the region of high magnetization becomes more narrow after the passage of the cloud. Also, the region

of high magnetization becomes less homogeneous (more patchy) due to the action of the KH instability. These simulations demonstrate that propagation through a large-scale cloud leads to a reconstruction of the magnetic field in space and to change in the value of the magnetic field. The tangled magnetic field may reconnect and the magnetic energy can be converted to the energy of accelerated particles and radiate. This may be a possible reason for the re-brightening of PWN magnetotails observed in X-ray.[7]

4.1.1. *Small-scale inhomogeneities*

Our simulations show that the KH instability frequently forms at the CD and produces some wavy structure of the bow shock. However, a wavy structure could also result from the interaction of the bow shock with the small-scale inhomogeneities in the ISM. To model the interaction of the small-scale inhomogeneities with the bow shock, we took models where the KH instability is weak and does not modify the shape of the bow shock.

We model small-scale inhomogeneities as a set of small clouds with Gaussian density distribution, with the maximum density of $\rho_{\max}$ and half-width of $\Delta r = 0.2$. To keep the pressure balance between the clouds and the rest of the ISM, we take the temperature in the cloud to be $T_{\rm cloud} = T_{\rm ISM} * \rho_0/\rho_{\rm cloud}$. We experimented with clouds of different densities, and found that the shape of the bow shock varies significantly if the density in the cloud is $\rho_{\rm cloud} \gtrsim 2\rho_0$. At $\rho_{\rm cloud} = 2\rho_0$, a wavy structure starts to become visible, while at $\rho_{\rm cloud} = 3\rho_0$, the bow shock changes its local shape significantly.

Fig. 5 shows a wavy variation of density in the bow shocks that appears after the propagation of clouds of density $\rho_{\rm cloud} = 3\rho_0$ in the models with low and moderate magnetizations. Such an interaction with small-scale inhomogeneities may explain the wiggles in the shape of the bow shock observed in PSR J0742-2822 (see Fig. 1). In PSR J0742-2822, the estimated stand-off distance is $R_{\rm sd} \approx 4.2 \times 10^{16} d/2{\rm kpc}$ cm. The size of clouds in our model $R_{\rm cl} \approx (2-4)R_{\rm sd} \approx (0.8-1.6) \times 10^{17} d/2{\rm kpc}$ cm. It may correspond to one of scales in turbulent ISM matter. We should note that an alternative explanation of these small-scale features in bow shocks can be connected with the KH instability. Currently, we cannot distinguish between these two mechanisms.

5. Conclusions

We performed MHD simulations of the supersonic propagation of magnetized neutron star and bow shock PWN propagating through a uniform and non-uniform ISM at different levels of magnetization. Our main findings are the following:

The fluctuations in the density or temperature leads to a change in shape of bow shock and the tail of the magnetosphere structure of the neutron star. With an increase in the density of the ISM the bow shock nebulae and the tail of the

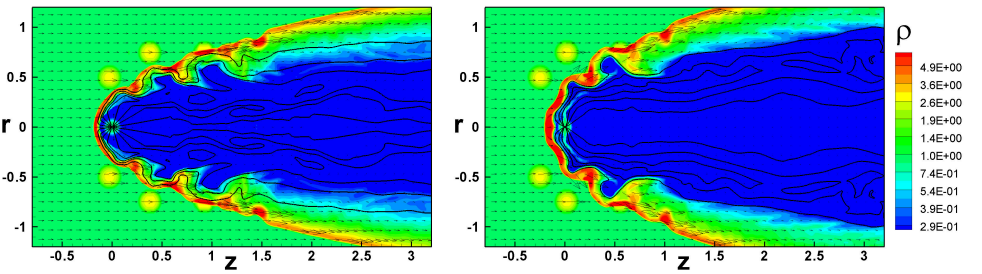


Fig. 5. Interaction of the bow shock with small-scale clouds of maximum density $\rho_{\rm cloud} = 3\rho_0$ in the model with low σ (left panel) and in the model with medium σ (right panel). The background represents the logarithm of density. The solid lines are magnetic field lines.

magnetosphere become narrow and vice versa. A temperature dependence is opposite. With an increase in the temperature of the ISM the bow shock nebulae and the tail of the magnetosphere become wider. A presence of "bubbles" or "bumps" in the shape of bow shock and the magnetotail most likely indicates on the changing density or temperature of the ISM.

The interaction of the bow shock with a large-scale, dense cloud leads to the compression of the bow shock and the formation of a new bow shock with a smaller opening angle. The level of compression (the ratio between the original and final widths of the bow shock) is approximately the same for PWN of different magnetization.

The interaction of the bow shock with the small-scale inhomogeneities in the ISM leads to a wavy structure in the bow shock. The amplitude of the "waves" increases with the ratio $\rho_{\rm cloud}/\rho_0$. For example, variation in the shape of the bow shock of pulsar PSR J0742-2822 can be explained by the propagation through a series of clouds with a density ratio of $\rho_{\rm cloud}/\rho_0 \approx 3$. Small-scale clouds produce wavy shapes of the bow shocks in PWN of different magnetization.

Acknowledgements

We are grateful to V.V. Saveliev (IPM RAS) for the original version of MHD simulation program. This work was partially supported by RFBR grants 18-29-21021, 20-02-00455.

References

1. Barkov, M., & Lyutikov, M. 2018, MNRAS, eprint arXiv:1804.07327
2. Brownsberger, S., Romani, R.W., 2014, ApJ, Volume 784, Issue 2, article id. 154, 14 pp.
3. Bucciantini, N., 2002, A&A, 387, 1066
4. Chatterjee S. & Cordes J. M., 2004, ApJ, Volume 600, Issue 1, pp. L51
5. Kargaltsev, O., & Pavlov, G.G. 2010, AIPC, 1248, 25

6. Kargaltsev, O., Durant, M., Pavlov, G., Garmire, G. 2012, ApJ Supplement, Volume 201, Issue 2, article id. 37, 7 pp.
7. Kargaltsev, O., Pavlov, G.G., Klingler, N., & Rangelov, B. 2017, J. Plasma Phys., Vol. 83, pp. 1-25
8. Ng, C.-Y. et al. 2010, ApJ, 712, 596
9. Toropina, O.D., Romanova, M.M., Toropin, Yu.M.,& Lovelace, R. V. E. 2001, ApJ, 561, 964
10. Toropina O.D., Romanova M.M., Lovelace R.V.E., 2007, MNRAS 371, 569–576 (2006)
11. Toropina, O.D., Romanova, M.M., & Lovelace, R. V. E., 2012, MNRAS, Volume 420, Issue 1, pp. 810-816
12. Toropina O.D., Romanova M.M., Lovelace R.V.E., 2019, MNRAS, Volume 484, Issue 2, P. 1475–1486,
13. Zhukov, V. T., Zabrodin, A. V., & Feodoritova, O. B. 1993, Comp. Maths. Math. Phys., 33, No. 8, 1099

Calculation of the kinetic coefficients of arbitrary degenerate electrons in magnetized dense matter

M. V. Glushikhina*,‡ and G. S. Bisnovatyi-Kogan*,†,§

*Space Research Institute of Russian Academy of Sciences (IKI RAN),
Profsoyuznaya str., 84/32, Moscow, 117997, Russia
†National Research Nuclear University (MEPhI),
Kashirskoe Shosse, 31, Moscow, 115409, Russia
‡E-mail: m.glushikhina@iki.rssi.ru
§E-mail: gkogan@iki.rssi.ru
www.iki.cosmos.ru

A solution of the Boltzmann equation is obtained for a magnetized plasma with arbitrary degenerate electrons and nondegenerate nuclei. For the arbitrary and non-degenerate electrons kinetic coefficients are obtained by solving Boltzmann equation by Chapman-Enskog method of successive approximations. The expressions have a considerably more complicated dependence on magnetic field than analogous dependences derived in previous publications on this subject.

Keywords: Neutron stars; magnetic field; kinetics.

1. Introduction

The transfer of heat and charge in a magnetized nondegenerate plasma plays an important role in describing its behavior both under laboratory conditions and in the structure and evolution of stars. Kinetic coefficients such as thermal conductivity, diffusion, thermal diffusion and diffusional thermal effect determine heat fluxes and current densities. Knowing the distribution of heat and current, one can calculate the magneto-thermal evolution, the distribution of the magnetic field and temperature over the surface of stars, or describe the behavior of plasma obtained and accelerated in laboratory conditions.

Here is presented results of solving Boltzmann equation by two methods. For arbitrary degeneracy case we used the Chapman–Enskog method of successive approximations. Tensors of thermal diffusion, diffusion and diffusional thermal effect are found by using an expansion of three polynomials.[1,2] In Lorentz approximation Boltzmann equation has an exact solution. It is shown by using Lorentzian gas as an example that the method has good convergence to the exact solution.

2. General equations

We use the Boltzmann equation for arbitrary-degenerate electrons in a magnetic field, taking into account the interaction of electrons with ions and with each other. Boltzmann equation describing the change in the electron distribution function f

over time in the presence of electric and magnetic fields, written as[3,4]

$$\frac{\partial f}{\partial t} + c_i \frac{\partial f}{\partial r_i} - \frac{e}{m_e}\left(E_i + \frac{1}{c}\varepsilon_{ikl} c_k B_l\right)\frac{\partial f}{\partial c_i} + J = 0. \quad (1)$$

Here $(-e), m_e$ — charge (negative) and electron mass, E_i, B_i — electric field strength and magnetic induction, ε_{ikl} completely antisymmetric Levi - Civita tensor, c — the speed of light. The collision integral J for nondegenerate electrons and singly charged ions, according to,[5–8] is written in the form:

$$J = J_{ee} + J_{eN} = \int [f f_1 - f' f'_1] g_{ee} b \, db \, d\varepsilon dc_{1i} + \int [f f_I - f' f'_I] g_{eI} b \, db \, d\varepsilon dc_{Ii}. \quad (2)$$

Here the impact parameter b and ε are the geometric parameters of the collision of particles with relative speeds g_{ee}, g_{eI}.

Integration in the electronic part of the collision integral in (2) is performed over the phase space incident particles (dc_{1i}) and their physical space ($b \, db \, d\varepsilon$).[5] Post-collision velocity functions are marked with strokes.

The transfer equations for the electron concentration total momentum and energy of electrons in a two-component mixture of electrons and kernels can be obtained in the usual way from Boltzmann equations in quasineutral plasma[3–5,9]

Here and in what follows, we assume the average mass velocity to be equal to the average ion velocity $c_{0i} = \langle c_{Ii} \rangle$, we also consider electric current and heat flux only from electrons.

The Boltzmann equation can be solved by using the Chapman-Enskog method of successive approximations.[5] This method is used when the distribution functions are close to their values in the state of thermodynamic equilibrium, and deviations are considered in a linear approximation. In a first approximation, we look for the distribution function f in the form $f = f_0[1 + \chi(1 - f_0)]$

The χ function allows the solution to be represented as:[1]

$$\chi = -A_i \frac{\partial \ln T}{\partial r_i} - n_e D_i d_i \frac{G_{5/2}}{G_{3/2}}, \quad (3)$$

where diffusion vector d_i:

$$d_i = \frac{\rho_N}{\rho}\frac{\partial \ln P_e}{\partial r_i} - \frac{\rho_e}{P_e}\frac{1}{\rho}\frac{\partial P_N}{\partial r_i} + \frac{e n_e}{P_e}\left(E_i + \frac{1}{c}\varepsilon_{ikl} c_{0k} B_l\right). \quad (4)$$

$G_n = \frac{1}{\Gamma(n)} \int_0^\infty \frac{x^{n-1} dx}{1 + exp(x - x_0)}$, $x_0 = \frac{\mu}{kT}$, here G_n is Fermi integral and μ is chemical potential. The functions A_i and D_i determine the heat flux and diffusion.

The collision integral similar to J_{ee} from (2) for strongly degenerate neutrons in nuclear matter is given in,[8] see also.[11] In the presence of non-degenerate heavy nuclei and highly degenerate neutrons, the contribution of collisions between them to the heat transfer and diffusion coefficients is negligible compared to neutron-nucleus

collisions. The same situation holds for highly degenerate electrons. Therefore, for strongly degenerate electrons, the Lorentz approximation, taking into account collisions between light and heavy particles, is asymptotically accurate. Therefore, we can neglect J_{ee} compared to J_{eN}, and we can equate $J = J_{eN}$ in (2). In this case Boltzmann equation is solved in Lorentz approximation as it was shown in[1,12] and the components of four tensors of kinetic coefficients are calculated.

General expressions for the heat flux q_i, and the average directional (diffusion) electron velocity $\langle v_i \rangle$ are given by

$$q_i = -\lambda_{ij} \frac{\partial T}{\partial x_j} - n_e \nu_{ij} d_j = q_i^{(A)} + q_i^{(D)}, \tag{5}$$

$$\langle v_i \rangle = -\mu_{ij} \frac{\partial T}{\partial x_j} - n_e \eta_{ij} d_j = \langle v_i^{(A)} \rangle + \langle v_i^{(D)} \rangle, \tag{6}$$

where λ_{ij} and ν_{ij} are the tensors of thermal conductivity and diffusional thermal effect, and μ_{ij} and η_{ij} — thermal diffusion and diffusional tensors, respectively.[11,13] The indices (A) and (D) determine the heat fluxes and diffusion velocities of electrons determined by the temperature gradient $\partial T/\partial x_j$, and the diffusion vector d_j, respectively.

The method for finding the coefficients of the thermal conductivity tensor λ_{ij} for arbitrary degeneracy is described in detail in the work,[1] where analytical expressions are found for them. For strongly degenerate electrons, the coefficients of thermal diffusion, diffusion and diffusion thermal effect in the Lorentz approximation are found in.[12]

3. Exact solution in Lorentz approximation

The Lorentz approximation for solving the kinetic equation is used when the mass of light particles much less than the mass of heavy particles, and electron-electron collisions can be neglected. So we can neglect J_{ee} in comparison with J_{eN}, and we can equate $J = J_{eN}$ in the (2). In this approximation, the linearized Boltzmann equation has exact solution. This approximation works well for transport in metal, where strong electron degeneracy makes it possible to neglect electron-electron collisions. The Lorentz approximation can be used to check the approximate polynomial solution, since it makes it possible to trace the convergence of the approximate solution to the exact one with increasing degree of the polynomials. Solution in Lorentz's approximation was considered in different approaches,[5,16,17] see also.[15] For thermal conductivity, the convergence of a polynomial solution to an exact one was considered in,[1] if we consider only the heat flux connected with the temperature gradient, at zero value of the diffusive vector d_i from (5), (6) than we obtain the expression

for the heat flux for the case $B = 0$ from[15] as:

$$q_i^A = -\frac{640k}{\Lambda}\frac{m_e(kT)^4}{n_N Z^2 e^4 h^3}\left(G_5 - \frac{1}{2}\frac{G_{5/2}}{G_{3/2}}G_4\right)\frac{\partial T}{\partial r_\mathrm{i}}. \tag{7}$$

In the limiting cases the coefficient in (7) is reduces to

$$\begin{aligned}\lambda_e^l &= \frac{40\sqrt{2}}{\pi^{3/2}\Lambda}k\frac{n_e}{n_N}\frac{(kT)^{5/2}}{\mathrm{e}^4 Z^2\sqrt{m_e}} = \frac{320}{3\pi}\frac{k^2 T n_e}{m_e}\tau_{nd} \quad (ND)\\ &= \frac{5}{64\Lambda}\frac{k^2 T n_\mathrm{e}^2 h^3}{m_\mathrm{e}^2 n_N Z^2 e^4} = \frac{5\pi^2}{6}\frac{k^2 T n_e}{m_e}\tau_d \quad (D).\end{aligned} \tag{8}$$

Average frequency of electron-ion collisions ν_{ei} in[14] in the limiting case of non-degenerate electrons is written as

$$\begin{aligned}\nu_{ei} &= \frac{4}{3}\sqrt{\frac{2\pi}{m_e}}\frac{Z^2\mathrm{e}^4 n_N\Lambda}{(kT)^{3/2}} \qquad (ND)\\ &= \frac{32\pi^2}{3}m_e\frac{Z^2\mathrm{e}^4\Lambda n_N}{h^3 n_\mathrm{e}} \qquad (D).\end{aligned} \tag{9}$$

Here indices (ND) and (D) determine non-degenerate and degenerate cases respectively.

Using the same approach as in,[1] we can find an exact solution in explicit form for the Lorentz approximation for the case $B = 0$:

$$\langle v_i^A\rangle = -\frac{128k}{\Lambda}\frac{m_e(kT)^3}{n_N n_e Z^2 e^4 h^3}\left(G_4 - \frac{5}{8}\frac{G_{5/2}}{G_{3/2}}G_3\right)\frac{\partial T}{\partial r_\mathrm{i}}, \tag{10}$$

$$\langle v_i^D\rangle = -n_e\frac{G_{5/2}}{G_{3/2}}\eta_{ij}d_j = \frac{32}{\Lambda}\frac{m_e(kT)^4}{n_e n_N Z^2 e^4 h^3}\frac{G_{5/2}}{G_{3/2}}G_3 d_i, \tag{11}$$

$$q_i^D = -n_e\frac{G_{5/2}}{G_{3/2}}\nu_{ij}d_j = \frac{128}{\Lambda}\frac{m_e(kT)^5}{n_N Z^2 e^4 h^3}\frac{G_{5/2}}{G_{3/2}}G_4\frac{\partial T}{\partial r_\mathrm{i}}. \tag{12}$$

In limiting cases, the coefficient in (10) reduces to

$$\mu_e^l = \frac{12k}{\pi^{3/2}\Lambda n_N}\frac{(kT)^{3/2}}{\mathrm{e}^4 Z^2\sqrt{2m_e}} = \frac{16k}{m_e\pi}\tau_{nd} \quad (ND), \tag{13}$$

and the coefficients in (11), (12) are reduced to

$$\eta_e^l = -\frac{32}{3\pi}\frac{kT}{n_e m_e\pi}\tau_{nd}, \qquad \nu_e^l = -\frac{128(kT)^2}{3\pi m_e}\tau_{nd}. \tag{14}$$

The exact formulas in the Lorentz model are used by[5] to estimate the accuracy of the polynomial approximation. The contribution of electron-electron collisions to the thermal diffusion coefficient for different Z can be estimated from the graph

of normalized three-polynomial thermal diffusion coefficients in the direction perpendicular to the magnetic field, introducing the quantity $Q_\perp^{(3l)}$, $U_\perp^{(3l)}$ and $W_\perp^{(3l)}$ defined as:

$$Q_\perp^{(3l)} = \frac{\mu_{nd}^{(13)}}{\mu_{e,nd}^{l}}, \qquad U_\perp^{(3l)} = \frac{\eta_{nd}^{(13)}}{\eta_{e}^{l}}, \qquad W_\perp^{(3l)} = \frac{\nu_{nd}^{(13)}}{\nu_{e}^{l}}. \tag{15}$$

Here $\mu_{e,nd}^{l}$ is taken from the top line in (13), $\nu_{e,nd}^{l}$ and $\eta_{e,nd}^{l}$ are taken from (14). Curves for different Z, including $Z = \infty$, related to Lorentz approximation are shown in Fig. (1a) for thermodiffusion, on Fig. (1b) for diffusional thermal effect, on Fig. (2a) for diffusion and on Fig. (2b) thermal conductivity plots[1] are presented for comparison.

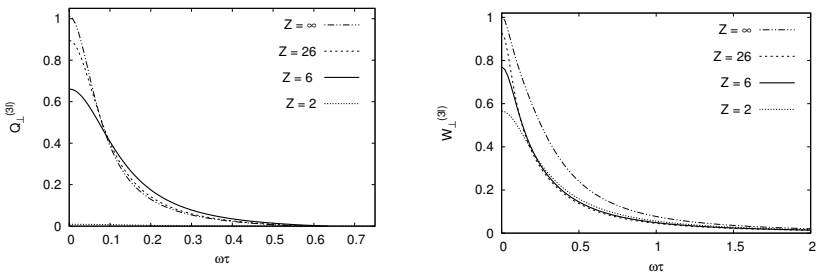


Fig. 1. On the (a) figure are presented the plots of the values of $Q_\perp^{(3l)}$ as functions of $\omega\tau$ in the three-polynomial approximation for nondegenerate plasma of helium ($Z = 2$), carbon ($Z = 6$), iron ($Z = 26$) in comparison with Lorentz plasma, formally corresponding to $Z = \infty$. The deviation from the Lorentzian plasma is due to the contribution of electron-electron collisions. On the right-side figure are plots of the values of the diffusional thermal effect $W_\perp^{(3l)}$ as a function of $\omega\tau$.

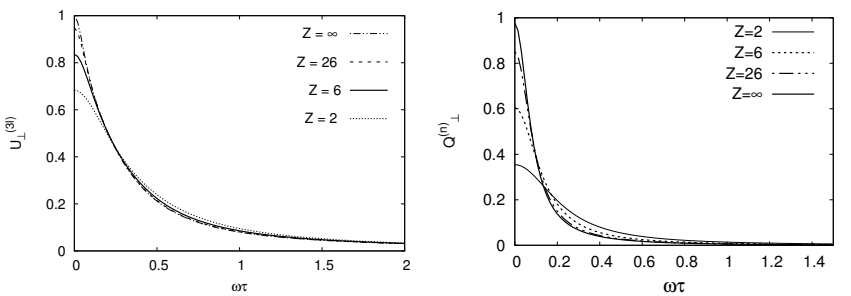


Fig. 2. On tthe (a) figure are plots of diffusion values of $U_\perp^{(3l)}$ as functions of $\omega\tau$ in the three-polynomial approximation for nondegenerate plasma of helium (Z=2), carbon (Z=6), iron (Z=26), for comparison with the Lorentz plasma, formally corresponding to $Z = \infty$. On the (b) figure are plots for heat conductivity coefficient from[1] as a function of $\omega\tau$, here: $Q_\perp^{(3l)} = \lambda_\perp^{(3)}/\lambda_\parallel^{lor}$.

4. Partially degenerate electrons

To examine the effect of degeneracy on the convergence of the polynomial approximation to the exact value, we compare the coefficients calculated by the method of successive approximations with the coefficients obtained as an exact solution in the Lorentz approximation. In a partially degenerate plasma, it is enough to simply calculate collision integral for electron-ion at $\mu = 0$ - chemical potential with the degeneracy level $DL = \epsilon_{fe}/kT = 1.011$. When compared with the exact solutions for heat conductivity and thermaldiffusion, obtained by the Lorentz method, the 3-polynomial solution for thermal conduction differs from the exact solution by about 13%, for thermal diffusion - by 7%. For non-degenerate electrons in the 3-polynomial approximation, the coefficient of thermal conductivity differs by 2.2%, and thermal diffusion by 1.3%.

For partial degeneracy, it is found that the 3-polynomial solution for diffusion differs from the exact one by 2.6%. For diffusion thermal effect: 3-polynomial by about 8%. For the nondegenerate case, the 3-polynomial approximation differs by 0.14% from the exact one for diffusion, and by 0.38% for the diffusion thermal effect. Detailed description of the calculations can be found in.[1,2] The accuracy of approximation by a series of orthogonal functions decreases with an increase in the degree of degeneracy.

5. Discussion

In this work, the tensors of the kinetic coefficients of diffusion, thermal diffusion, and diffusion thermal effect are found for nondegenerate electrons in a nonquantizing magnetic field. Tensors are obtained from the solution the kinetic Boltzmann equation by the classical Chapman–Enskog method using the expansion into Sonin polynomials and taking into account two and three terms of the expansion. Electron-ion and electron-electron collisions are taken into account. Tensors are written for an arbitrary local direction magnetic field and temperature gradient in the Cartesian coordinate system according to.[4]

Using the example of the Lorentz approximation, it was shown that the accuracy of the approximation by a series of orthogonal functions, similar to Sonin's polynomials, decreases with increasing degree of degeneracy.

For non-degenerate electrons, disregarding electron-electron collisions, the value of the thermal diffusion coefficient in the two-polynomial approximation underestimates the exact solution obtained in the Lorentz approximation at $B = 0$ by 26%, and in the three-polynomial approximation it overestimates by approximately 1%. For partially degenerate electrons, at $\mu = 0$, the two-polynomial solution underestimates the exact one by 28%, and the three-polynomial one underestimates the exact one by 7%. It should be noted that electron-electron collisions further decrease the coefficient values.

The diffusion coefficient value obtained from the two-polynomial approximation underestimates the exact solution by about 4%, and the three-polynomial solution

underestimates the exact solution by 0.14% for nondegenerate electrons. In the case of partial degeneracy, the diffusion obtained from the two-polynomial solution underestimates the exact one by 0.4%, and the one obtained from the three-polynomial solution overestimates the exact one by 3%.

The diffusion thermal effect coefficient for nondegenerate electrons in the two-polynomial approximation underestimates the exact solution by 12%, and in the three-polynomial approximation overestimates the exact solution by 0.4%. With partial degeneracy of electrons, the two-polynomial solution underestimates the exact solution by 2.5%, and the three-polynomial solution overestimates the exact solution by 8%.

The Chapman–Enskog method can be used for a sufficiently dense gas (plasma), where the time between particle collisions is the smallest value among other characteristic times. In the presence of a magnetic field, in addition to the lifetime of the system and the characteristic time of changes, in the plasma parameters rotation time in the Larmor circle is added $\tau_L = 2\pi/\omega$. This time must be much less than τ, which leads to a condition under which the Chapman–Enskog method can be used, in the form $\omega\tau \ll 2\pi$. Therefore, this work can be successfully applied for $\omega\tau \lesssim 1$, and for large $\omega\tau$ only qualitative estimates can be obtained.

The coefficients can be used for calculation of heat fluxes and electric current in white dwarfs, on the surface of magnetized neutron star, as well as in the magnetized matter accreting to the magnetized neutron star or for describing plasma obtained in laboratory conditions.

References

1. G. S. Bisnovatyi-Kogan, M. V. Glushikhina, Plasma Physics Reports, **44**, 405 (2018).
2. M. V. Glushikhina, Plasma Physics Reports, **46**, 152 (2020).
3. G. S. Bisnovatyi-Kogan, Transport properties of partially ionized two-temperature plasma and isotropic corrections to Maxwellian distribution function (in Russian), Diploma Thesis, Moscow Institute of Physics and Technology (Soviet Union, 1964).
4. G. S. Bisnovatyi-Kogan, J. Appl. Mech. Tech. Phys., No. 3, 43 (1964).
5. *Chapmen S., Cowling T.G.* // Mathematical Theory of Nonuniform Gases, Cambrige, (1952).
6. E. Uehling, G. Uhlenbeck., Phys. Rev., **43**, 552 (1933).
7. E. Uehling, Phys. Rev., **46**, 917 (1934).
8. S. Z. Tomonaga, Z. Phys. **110**, 573 (1938).
9. W. Marshall, The kinetic theory of an ionized gas, At. Energy Res Estable, N T/R, 2419 (1960).
10. *Landau L. D., Lifshitz E. M.* // Statistical Physics. **5**, 3rd ed., Butterworth-Heinemann, 1980.
11. G. S. Bisnovatyi-Kogan, M. M. Romanova, JETP, **56**, 243 (1983).
12. G. S. Bisnovatyi-Kogan, M. V. Glushikhina, Plasma Physics Reports, **44**, 971 (2018).
13. *Kalikhman L. E.* // Elements of magnetohydrodynamics, Atomizdat, Moscow, 1964.

14. *Ginzburg V. A., Rukhadze A. A.* // Waves in Magnetoactive Plasma, M.: Nauka, 1970.
15. *Bisnovatyi-Kogan G.S* // Stellar Physics I: Fundamental Concepts and Stellar Equilibrium. Springer, 2001.
16. A. A. Wyller, Astrophisica Norvegica. **9**, 79 (1971).
17. A. A. Wyller, Astrophys. J., **184**, 517 (1973).

Modeling of magnetic fields of accretion discs, using no-z- and RZ-approximations

M. V. Pashentseva[1], E. A. Mikhailov[1,2]
[1] *Faculty of Physics, Lomonosov Moscow State University, Moscow, Russia*
[2] *P.N.Lebedev Physical Institute, Russia*

The evolution of the accretion discs surrounding different compact objects (such as white dwarfs, neutron stars and black holes) is closely connected with the magnetic fields and their features. It is highly likely that the magnetic field generation is connected with the dynamo action. As for the dynamo, we have a lot of specific models which have been constructed for different astrophysical objects. For accretion discs we can take the approaches that were used for galactic discs which have nearly the same shape. There are two main models. The first one is connected with no-z approximation which is based on the fact that the discs is quite thin. RZ-model describes the discs with large half-thickness and can be used for the objects where we should study the vertical structure of the magnetic field. Here we show the results obtained for the accretion discs which used both of these approaches and compare the results.

Keywords: Magnetic fields, accretion discs, dynamo theory, no-z model, RZ-model.

1. Introduction

The accretion discs surrounding different compact objects (such as black holes, neutron stars and white dwarfs), play an important role in different branches of astrophysics. Their structure and features is closely connected with the magnetic fields, which should have a strong influence on the motions in the disc.[1] Nowadays there are different approaches that describe the origin of magnetism of such objects. First of all, we should mention the transition of the field which is connected with the accreting matter.[2] Also there are some models which describe the field as a result of interaction between the central object and the accretion disc. However, observational data and theoretical models show that the field is highly likely to be generated by the dynamo mechanism.[3–5]

The dynamo mechanism is widely used to study the magnetic field evolution in different objects such as the Sun, stars, galaxies etc. The equations of the dynamo are quite difficult to be solved in explicit form. So there are different specific two-dimensional models which take into account geometrical features of the celestial body and allow us to simplify the problem. As for the galaxies, there is so-called no-z approximation,[6] which is based on the fact that the galactic disc is quite thin. So some of the partial derivatives can be changed by the algebraic expressions, and the z-component of the field can be taken from the solenoidality condition. This approach can be used to study magnetism of the accretion discs,[7,8] because their shape is quite close to the one of the galaxies. As for thick galactic discs, we can

take the RZ-model, which uses another basic assumptions.[9] In this paper we study the magnetic field using this model, too. After that we compare the given results.

2. Dynamo

The dynamo is connected with transition of the energy of turbulent motions to the energy of the magnetic field.[10] If we are speaking about the large-scale magnetic field, it is based on two main effects. First of all, it is the alpha-effect which is connected with the helicity of the turbulent motions. It is connected with the rotation of the magnetic field. The second process is associated with the differential rotation which is caused by changing angular velocity (most of celestial bodies rotate with a non-solid law). They compete with the dissipative effects which destroy the large-scale structures of the field. So, dynamo is a threshold process: the field can grow only if the alpha-effect and differential rotation are more intensive than the dissipation. Otherwise the field will decay and the dynamo won't be able to generate significant fields.

The evolution of large-scale component of the field is described by the Steenbeck – Krause – Rädler equation, which has been obtained by averaging the magnetohydrodynamics equations by scales comparable with size the turbulent cells:[11]

$$\frac{\partial \vec{B}}{\partial t} = \nabla \left(\alpha \vec{B}\right) + \nabla \left[\vec{V}, \vec{B}\right] + \eta \Delta \vec{B},$$

This equation is rather difficult to be solved. So usually the researchers use specific two-dimensional models for every type of the objects: such as Parker dynamo for stars of torus dynamo for outer rings of galaxies.

2.1. *No-z approximation for accretion discs*

While studying the magnetic fields in the accretion discs, it is quite convenient to use the models which have been developed for the objects that have nearly the same geometrical shape (of course, taking into account sufficiently different lengthscales). As for the galactic discs, so-called no-z approximation is widely used. It studies the evolution of two components of the field (which lie in the equatorial plane), which makes the problem much easier. The z-component of the field is taken using the solenoidality condition. Also it is quite convenient to assume that the field is axisymmetric (as for the accretion discs it is even more reasonable than for the galactic ones).

If we study the accretion disc, we should take the keplerian law for the rotation velocity Ω[1]:

$$\Omega = \frac{(GM)^{\frac{1}{2}}}{r^{\frac{3}{2}}},$$

where M is the mass of the black hole, neutron star or white dwarf; r is the distance from the central object.

The equations should be solved for values $r_{in} < r < R$, where r_{in} is the inner radius and R is the outer one.

It is quite convenient to use the dimensionless units: all parameters are measured in their typical values. As for the magnetic field, some dimensionless parameters characterizing the dynamo regime are introduced.

R_α shows the intensity of the alpha-effect:

$$R_\alpha = \frac{\Omega(R) l^2}{\eta},$$

where Ω is the angular velocity of the disc and l is the typical size of the turbulent cell.

R_ω is connected with the differential rotation:

$$R_\omega = \frac{\Omega(R) h^2}{\eta},$$

λ corresponds to the ration between the lengthscales of the disc (half-thickness h and radius R):

$$\lambda = \frac{h}{R} = 0.037 r^{\frac{1}{8}} \left(1 - \sqrt{\frac{r_{in}}{r}}\right)^{\frac{3}{20}}.$$

Taking these assumptions into account, the equations for the magnetic field in dimensionless form look like this:

$$\frac{\partial B_r}{\partial t} = -R_\alpha B_\varphi \left(1 - \frac{B^2}{B^{*2}}\right) - \frac{\pi^2 B_r}{4} + \lambda^2 \left(\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r B_r\right)\right)\right) - \frac{V_0}{r\left(1 - \sqrt{\frac{r_{in}}{r}}\right)}\left(\frac{\partial B_r}{\partial r}\right), \tag{1}$$

$$\frac{\partial B_\varphi}{\partial t} = -\frac{3 R_\omega}{2 r^{\frac{3}{2}}} B_r - \frac{\pi^2 B_\varphi}{4} + \lambda^2 \left(\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r B_\varphi\right)\right)\right) - \frac{V_0}{r\left(1 - \sqrt{\frac{r_{in}}{r}}\right)}\left(\frac{\partial B_\varphi}{\partial r}\right). \tag{2}$$

The equipartition field is assumed to be the following:[7]

$$B^* = 420 r^{\frac{-21}{16}} \left(1 - \sqrt{\frac{r_{in}}{r}}\right)^{\frac{17}{40}}.$$

It is included in the equations using the non-linearity coefficient $(1 - \frac{B^2}{B^{*2}})$.

As for the boundary conditions, we use the following ones:

$$\frac{\partial B_r}{\partial r}(R) = 0,$$

$$\frac{\partial B_\varphi}{\partial r}(R) = 0.$$

$$B_r\left(r_{in}\right) = 0,$$

$$B_\varphi\left(r_{in}\right) = 0.$$

Also we can take the zero derivative conditions on the inner boundary:[8]

$$\frac{\partial B_r}{\partial r}\left(r_{in}\right) = 0,$$

$$\frac{\partial B_\varphi}{\partial r}\left(r_{in}\right) = 0.$$

However, it has been shown that the results for the magnetic field will be nearly the same as for the model with zero field boundary condition.

The magnetic field for different inner radius is shown on figure 1. It can be seen that larger inner radius moves the field from the center. Also the time dependence is shown on figure 2.

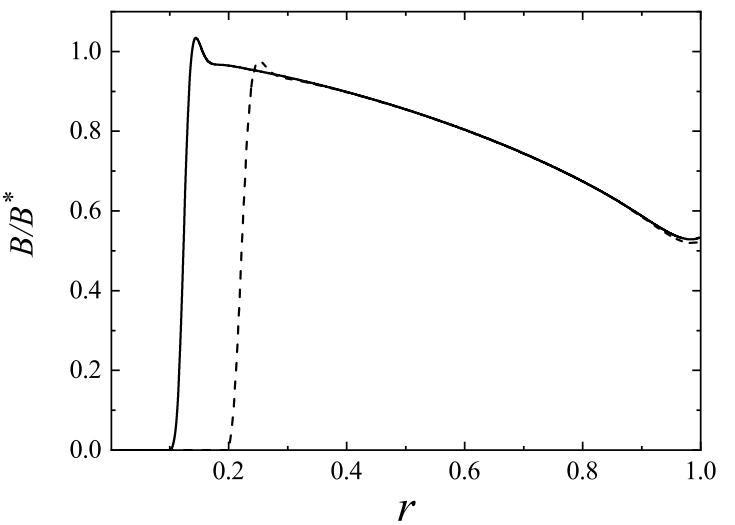


Fig. 1. The dependence of the magnetic field on the distance to the center of the accretion disk for different inner radius with zero boundary condition ($D_{eff} = 15$). Solid line shows the case, when $r_{in} = 0.1$ and the dotted line – the case $r_{in} = 0.2$.

3. RZ-model

No-z approximation has a lot of advantages for a wide spectra of objects (from galaxies to accretion discs). Unfortunately, it cannot be use for the objects where the vertical structure is more complicated than a standard law which is taken into

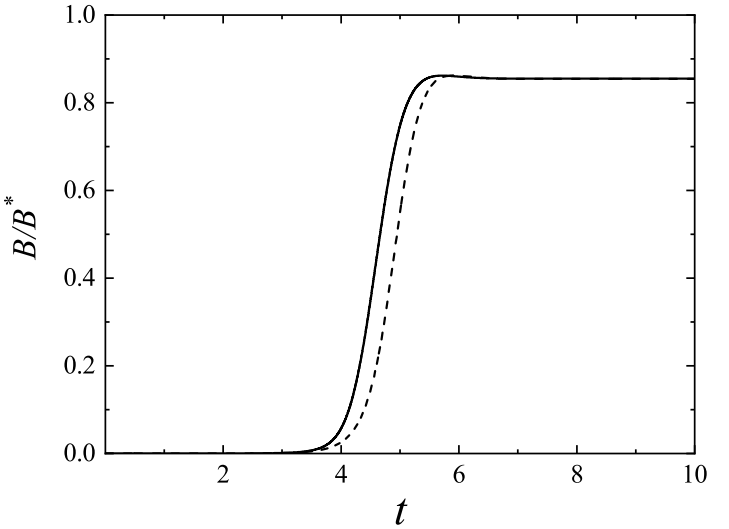


Fig. 2. Time dependence of the magnetic field for different r_{in} with zero boundary condition. Solid line shows the case, when $r_{in} = 0.1$ and the dotted line – the case $r_{in} = 0.2$.

account there. So, for thick discs we should take the RZ-approximation which describes the field using another approaches. Firstly these ideas have been used for the torus dynamo in outer rings of galaxies[12] and now there are results which can describe the whole disc.[9]

The magnetic field $\vec{B}$ can be presented as:

$$\vec{B} = B\vec{e_\varphi} + \nabla\left(A\vec{e_\varphi}\right),$$

where B is the azimuthal magnetic field and A is the azimuthal part of the vector potential of the magnetic field.

If we use this model, the Steenbeck – Krause – Rädler equation will take the following approximate form:

$$\frac{\partial A}{\partial t} = R_\alpha z B\left(1 - B^2\right) + \lambda^2 \Delta A - V\frac{\partial A}{\partial r},$$

$$\frac{\partial B}{\partial t} = R_\omega r\left(-\frac{3}{2}r^{\frac{5}{2}}\frac{\partial A}{\partial z}\right) + \lambda^2 \Delta B - V\frac{\partial B}{\partial r}.$$

All main dimensionless parameters here are the same as for the previous model. The equations are solved for $-h < z < h$, $r_{in} < r < R$, and the boundary conditions are:

$$A|_{t=0} = 0$$

$$B|_{t=0} = B_0 sin(\pi r) cos\Big(\frac{\pi z}{h}\Big)$$

$$\frac{\partial A\,(r,-h)}{\partial\ z} = 0.$$

The comparative results are shown on Fig 3. Here we can see both the field modeled using no-z approximation and the RZ-model. We can see that the field for the RZ-model reaches a bit larger values and the typical dependence is different, too.

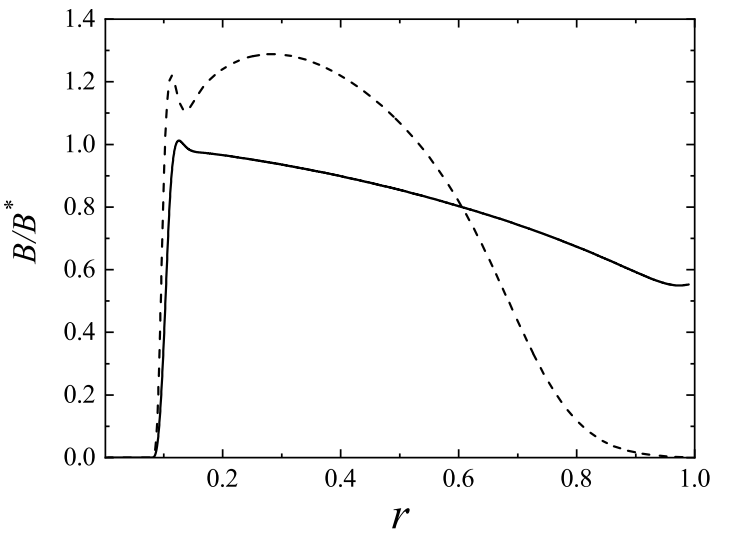


Fig. 3. Comparison of two models. Dashed line shows RZ-model and solid line shows no-z model.

Figure 4 shows the 2D-structure of the field. Here we can see the maximum of the magnetic field quite close to the inner boundary, and near the outer boundary the field decreases.

Also we have studied the connection between the dynamo number $D = R_\alpha R_\omega$ and the r-dependence of the field. It can be seen that for lower D the field is smaller, too.

4. Conclusion

In this work we have studied evolution of the magnetic field in accretion discs. We have used two different models taking into account different approaches to the dynamo. First of all, we studied the evolution of the magnetic field using the no-z approximation which has been developed for thin discs.[6] After that, we passed to

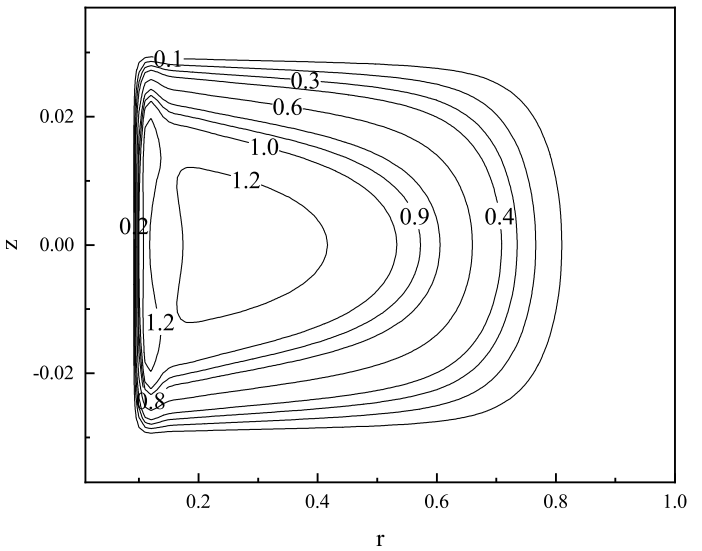


Fig. 4. The dependence of the magnetic field on r and z coordinates for RZ-model.

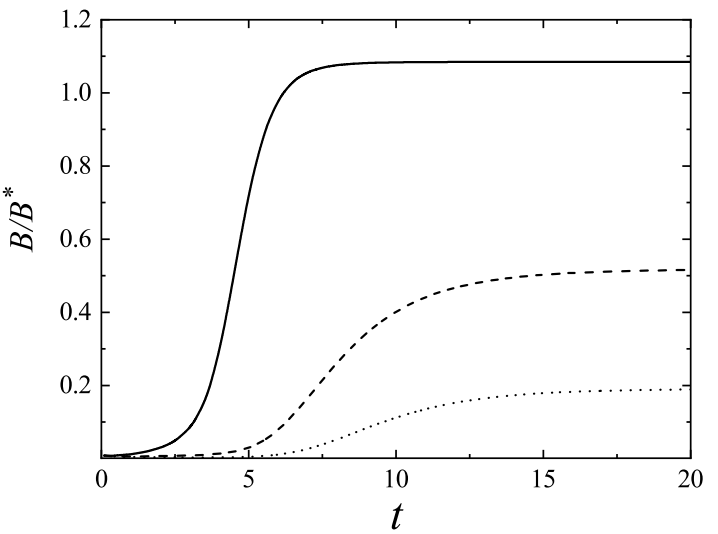


Fig. 5. Time dependence of the magnetic field for RZ-model for different D_{eff}. Solid line shows the case, when $D_{eff} = 15$, dashed line shows the case, when $D_{eff} = 9$, dotted line shows the case, when $D_{eff} = 7$.

the RZ-model, which takes into account z-dependence of the magnetic field, but it is more difficult to be solved from the numerical point of view. We can conclude that if we study thick discs, it is better to take the RZ-model, and for thin one we can take no-z approximation.

Work of M.P. was supported by a grant from the Fund for the Development of Theoretical Physics and Mathematics "BASIS", project 21-1-1-4-4.

References

1. N. I. Shakura, R. A. Sunyaev, Astron. Astrophys, 24, 337 (1973).
2. Okuzumi S., Takeuchi T., Muto T., ApJ, 785, 127 (2014).
3. Brandenburg A., Donner K.J. MNRAS, 288, L29 (1997).
4. von Rekowski B., Brandenburg A., Astronomy & Astrophysics, 420, 1 (2004).
5. Rüdiger R., Elstner D., Stepinsky T. F., Astron. Astrophys., 298, 934 (1995).
6. Moss D., Monthly Notices of the Royal Astronomical Society, 275, 1, 191 (1995).
7. Moss D., Sokoloff D., Suleimanov V., Astronomy & Astrophysics, 588, A18 (2016).
8. D. V. Boneva, E. A. Mikhailov, M. V. Pashentseva, D. D. Sokoloff, Astronomy and Astrophysics., 652, A38 (2021).
9. Mikhailov E. A., Pushkarev V. V., Research in Astronomy and Astrophysics, 21, 56 (2021).
10. Sokoloff D., Physics-Uspekhi 58, 601 (2015).
11. Krause F., Rädler K.-H., Berlin (1980).
12. Mikhailov E. A., Astrophysics., 61, 2, 147 (2018).

Limiting effects in tori clusters

D. Pugliese and Z. Stulchik

Research Centre for Theoretical Physics and Astrophysics
Institute of Physics, Silesian University in Opava,
Bezručovo náměstí 13, CZ-74601 Opava, Czech Republic
E-mail:d.pugliese.physics@gmail.com

We consider agglomerates of misaligned tori orbiting a supermassive black hole. The aggregate of tilted tori is modeled as a single orbiting configuration by introducing a leading function governing the distribution of toroids (and maximum pressure points inside the disks) around the black hole attractor. The orbiting clusters are composed by geometrically thick, pressure supported, perfect fluid tori. This analysis places constraints on the existence and properties of tilted tori and more general aggregates of orbiting disks. We study the constraints on the tori collision emergence and the instability of the agglomerates of tori with general relative inclination angles, the possible effects of the tori geometrical thickness and on the oscillatory phenomena. Some notes are discussed on the orbiting ringed structure in dependence of the dimensionless parameter ξ representing the (total) BH rotational energy extracted versus the mass of the BH, associating ξ to the characteristics of the accretion processes.

Keywords: Accretion disks; Accretion– jets; Black hole physics; Hydrodynamics

1. Introduction

We study agglomerates of tori orbiting one central Kerr super-massive black hole (SMBH), in two macro-configuration models: (i) as an equatorial-**R**inged **A**ccretion **D**isk (eRAD), where the toroids equatorial and symmetry planes coincide with the central Kerr BH equatorial plane;[1–9] (ii) as **R**inged **A**ccretion **D**isk (RAD), where the agglomerate is composed by misaligned (tilted) tori.[10–12]

The eRAD tori rotation orientation is a well defined quantity, and a couple of eRAD toroids can be both corotating or counter-rotating with respect to the central spinning BH, otherwise tori can have a relative alternate rotation orientation (with an inner corotating toroid and an outer counter-rotating toroid with respect to the central attractor or *viceversa*). Tori with different relative rotation orientation constitute an interesting case from the viewpoint of the constrains posed by the eRAD or RAD model, for example in the case of double tori accretion phase or also tori collision which depends generally on the spin-mass ratio of the central BH. It is clear that an orbiting tori aggregate model could be constructed considering a different disk model for each toroidal component. However in this analysis we mostly consider a Polish doughnut (PD) model. This is a well known geometrically thick disk model, widely used in literature in a variety of different applications.[1,13] Polish doughnut shows a remarkably good fitting of main morphological characteristics of thick disks also in comparison with more refined dynamical GRHD or

GRMHD models.[13] RAD is a "constraining-models", providing initial configurations for dynamical (GRMHD) situations. In general aggregates of toroidal structures orbiting one central BH attractor can result from different accreting phases of the SMBHs growing, where the infalling materials, having diverse angular momentum, may trace back the BH story.[2,5,14–23]

This paper is structured in two parts: In the first part, Sec. (2), we discuss the eRAD model, focusing on the model parameters and distribution of pressure and density critical points. We introduce the concept of leading function for the agglomeration describing the tori distribution around the central attractor.

The second part, Sec. (3), focuses on aggregates of misaligned tori (RAD), discussing explicit solutions of its inner structure and an adapted parametrization for the toroidal components.

2. The Equatorial Ringed Accretion Disks

2.1. *Leading function and geodesic structure*

A key step in the modelization of the orbiting agglomerate is to individuate an adapted "leading function", representing the tori location around the central attractor. For large part of this analysis, we can identify the leading function with the definition of fluid specific angular momentum ℓ. Different choices are the agglomerate leading functions also possibles as we see some examples in Sec. (2.4) and Sec. (3.1). In the Kerr spacetime there is

$$\begin{aligned} \ell &\equiv \frac{L}{E} = -\frac{U_\phi}{U_t} = -\frac{g_{\phi\phi}U^\phi + g_{\phi t}U^t}{g_{tt}U^t + g_{\phi t}U^\phi} = -\frac{g_{t\phi} + g_{\phi\phi}\Omega}{g_{tt} + g_{t\phi}\Omega}, \\ \Omega &\equiv \frac{U^\phi}{U^t} = -\frac{Eg_{\phi t} + g_{tt}L}{Eg_{\phi\phi} + g_{\phi t}L} = -\frac{g_{t\phi} + g_{tt}\ell}{g_{\phi\phi} + g_{t\phi}\ell}, \end{aligned} \tag{1}$$

where $g_{\alpha\beta}$ are metric components in the Boyer-Lindquist coordinates, U^α is the fluid four velocity, Ω is the fluid relativistic angular velocity and (E, L) are constants of motions–see for example.[1,2] In general, we may interpret E, for timelike geodesics, as representing the total energy of the test particle coming from radial infinity, as measured by a static observer at infinity, and L as the axial component of the angular momentum of the particle. (For the PD tori orbiting in a Kerr spacetime the set of results known as Von Zeipel theorem holds, therefore the fluid is barotropic and the surfaces of constant pressure coincide with the surfaces of constant density. In these spacetimes, the family of von Zeipel's surfaces does not depend on the particular rotation law of the fluid, $\Omega = \Omega(\ell)$, but on the background spacetime only.[24–28])

The leading function provides the distribution of the possible maximum points of pressure and density in the fluids surrounding the BH, which are identified as the RAD-"rings seeds" , coincident with torus centers r_{center}, and eventually the minimum points of pressure of the fluids orbiting around the central attractor. The

minimum points of pressure are associated to the cusps $r_\times$ of the PD torus Roche lobe and are regulated by the geodesic structure of the background, composed by the marginally stable circular orbit (r_{mso}), the marginally bounded circular orbit (r_{mbo}) and the marginally circular orbit (r_γ) which is also a photon orbit. (The cusp location in the PD model, located in the range $]r_{mbo}, r_{mso}[$, could be related to the inner edge of an accreting torus.) In the RAD and eRAD models, the location of the maximum points of the pressure around the attractor are interpreted as "rings seeds", and are regulated by a set of radii, associated to the geodesics structure, located in the "stability region" at $r > r_{mso}$ and defined by

$$
\begin{aligned}
& r^\pm_{(\mathrm{mbo})} : \ \ell^\pm(r^\pm_{\mathrm{mbo}}) = \ell^\pm(r^\pm_{(\mathrm{mbo})}) \equiv \ell^\pm_{\mathrm{mbo}}, \quad \text{with} \quad r^\pm_{(\gamma)} : \ell^\pm(r^\pm_{\gamma}) = \ell^\pm(r^\pm_{(\gamma)}) \equiv \ell^\pm_{\gamma}, \\
& \text{where} \quad r^\pm_{\gamma} < r^\pm_{\mathrm{mbo}} < r^\pm_{\mathrm{mso}} < r^\pm_{(\mathrm{mbo})} < r^\pm_{(\gamma)} \\
& \text{and} \quad r^\pm_{\mathcal{M}} : \partial_r \partial_r \ell^\pm = 0, \quad \text{with} \quad r^\pm_{(\mathcal{M})} : \ \ell^\pm(r^\pm_{(\mathcal{M})}) = \ell^\pm(r^\pm_{\mathcal{M}})) \equiv \ell^\pm_{\mathcal{M}}.
\end{aligned} \tag{2}
$$

In here and in the following with $\mathcal{Q}^\pm$ we indicate quantities $\mathcal{Q}$ associated to counter-rotating ($\ell a < 0$) or corotating ($\ell a > 0$) structures respectively, with respect to the central BH spin a/M. In the following, for any quantity $\mathbf{Q}$ and radius $r_\bullet$ we adopt the notation $\mathbf{Q}_\bullet \equiv \mathbf{Q}(r_\bullet)$. Where more conveniently, we use dimensionless quantities where $r \to r/M$ and $a \to a/M$.

The leading function, solution of $\partial_r^2 \ell^\pm(r, \theta; a) = 0$ (where (r, θ, ϕ, t) are the Boyer-Lindquist coordinates of the Kerr metric), provides the point of maximum density of the rings seeds distribution around the BH.

The corotating and counter-rotating tori cusps orbital regions are shown in the Figs 1 for different BH attractors. In the Figs 1 the orbital regions for the ring seeds locations (tori centers) are also shown. The union of these regions provides the maximum rage of the location of the rings seeds and the maximum extension of the disks inner part (the region $[r_{inner}, r_{center}]$ where $r_\times \leq r_{inner}$ is the torus cusp and r_{inner} is the inner edge of the quiescent (i.e. not cusped torus)). The orbital strips in Figs 1 relative to the corotating and counter-rotating fluids, cross in different points depending on the BH spin–mass ratio and particularly for slower spinning attractors.

2.2. *Ideal GRMHD and GRHD*

Polish doughnuts have been realized in different ideal GRMHD and GRHD setups, where for the ideal GRMHD (infinitely conductive plasma) case there is

$$
\begin{aligned}
& U_\alpha \nabla^\alpha \rho + (p + \rho) \nabla^\alpha U_\alpha = 0, \\
& (p + \rho) U^\alpha \nabla_\alpha U^\gamma - \epsilon h^{\beta\gamma} \nabla_\beta p - \epsilon (\nabla^\alpha F_{\alpha\delta}) F_\beta{}^\delta h^{\beta\gamma} = 0, \\
& U^\alpha \nabla_\alpha s = 0.
\end{aligned}
$$

(F^β_α is the Faraday tensor and ϵ is a quantity related to the metric signature[29,30]). The electric field does not affect the continuity equation or the equation for the

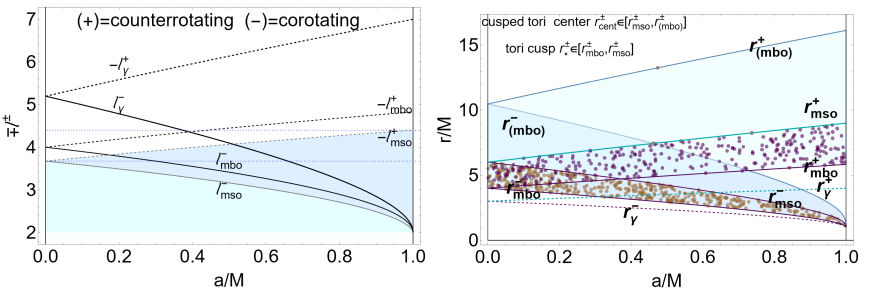


Fig. 1. Left: Fluid specific angular momentum $\ell^{\pm}$ of the eRAD tori, for corotating $(-)$, and counterrotating $(+)$, fluids, versus SMBH dimensionless spin a/M. Tori can form for $\mp\ell^{\pm} > \mp\ell^{\pm}_{mso}$ respectively. There is $\ell_{\bullet} \equiv \ell(r_{\bullet})$ for $r_{\bullet} = \{r_{mso}, r_{mbo}, r_{\gamma}\}$, r_{mso} is the marginally stable circular orbit, r_{mbo} is the marginally bounded orbit and r_{γ} the last circular (photon) orbit. Right: radii r_{mbo} and r_{mso} and the pair $r_{(mbo)}$ and $r_{(mso)}$ as functions of the dimensionless spin of the BH–see Eqs (2).

entropy. The entropy per particles s is conserved along the flow U^{α}, defined for *each* torus of the aggregate, and h^{β}_{α} is the metric on the 3-sheet orthogonal to flow direction U^{α}, defining the projector tensor (there is $\nabla_{\alpha} g_{\beta\gamma} = 0$). The inner ringed structure is defined by the boundary conditions determining the tori (edges)—see for more details.[9,29,31]

To simplify our discussion we consider in this analysis the GRHD scenario (see for example[1,13]). With the barotropic equation of state, $p = p(\rho)$, the set of GRHD equations for the PD model, reduces to the only constrain equation for the pressure (Euler equation):

$$\frac{\nabla_{\mu} p}{p+\rho} = -\nabla_{\mu} W + \frac{\Omega \nabla_{\mu} \ell}{1 - \Omega \ell};$$

where $W = W(a; \ell; r, \theta)$ is an effective potential, function of (r, θ) for the Kerr metric (in the Boyer-Lindquist coordinates). The equation for the pressure critical points can be further simplified by considering ℓ as a model parameter, constant in each torus of the aggregate. In this setup the extreme points of pressure are the extremes of the potential W, and the leading function is given by the radial derivative of the effective potential. More precisely the tori parameters are the couple (ℓ, K) with $\ell = constant$ and $V_{eff} = K = constant$ for each torus, where $W = \ln V_{eff}$.

In this context, the eRAD leading function is $\ell(r) : \partial_r V_{eff} = 0$. The tori (Boyer) surfaces are equipressure surfaces (also surfaces of constant (p, ρ, ℓ, Ω)) and the fluids fill every equipressure surfaces.[32]

2.2.1. *The energy function and tori energetics*

The "energy function" $K(r) = V_{eff}(\ell(r))$ regulates, with the leading function $\ell(r)$, the RAD aggregate, where $K : K(r)$ =constant for each torus. More precisely,

function $K(r,\theta;a) \equiv V_{eff}(\ell,r,\theta;a)|_{\ell(r,\theta;a)}$ determines the flow (and the torus) geometrical thickness, the tori extension on the equatorial plane, and it is uniquely identified by $\ell(r)$ in the case of cusped tori. The relation between the geometrical maxima (defined by K) and the density maxima (fixed by ℓ) is provided by the extreme of the leading function $\ell(r,\theta;a)$.[12] We can relate $K(r)$ to certain features of the tori energetics, evaluating some characteristics related to the flow thickness, as mass accretion rate or cusp luminosity as listed in Table (1). It is clear that these quantities depends on the details of the different tori models, but this analysis can provide an estimation of these quantities with respect to the flow co-rotation or counter-rotation, the dimensionless BH rotational energy (or dimensionless spin) and the tori location in the aggregate.[33]

Table 1. There is $\varpi = n+1$, with $\gamma = 1/n+1$ being the polytropic index, κ is the polytropic constant. Ω is the relativistic angular velocity. $W = \ln V_{eff}$ is the value of the equipotential surface, which is taken with respect to the asymptotic value, $W_\times = \ln K_\times$ is W at the cusp $r_\times$, while $W_s \geq W_\times$ and r_s is related to the accreiting flow thickness. $\mathcal{L}$ representing the total luminosity, $\dot{M}$ the total accretion rate where, for a stationary flow, $\dot{M} = \dot{M}_\times$, $\eta \equiv \mathcal{L}/\dot{M}c^2$ the efficiency, $\mathcal{D}(n,\kappa), \mathcal{C}(n,\kappa), \mathcal{A}(n,\kappa), \mathcal{B}(n,\kappa)$ are functions of the polytropic index and the polytropic constant. $\mathcal{L}_\times/\mathcal{L}$ is the fraction of energy produced inside the flow and not radiated through the surface but swallowed by central BH–see Figs 2

Quantities $\mathcal{O}(r_\times, r_s, n) \equiv q(n,\kappa)(W_s - W_\times)^{d(n)}$	**Quantities** $\mathcal{P} \equiv \frac{\mathcal{O}(r_\times, r_s, n) r_\times}{\Omega(r_\times)}$
$\mathcal{R}$**-quantities:** $\mathcal{R}_\times \equiv (W(r_s) - W_\times)^{\varpi}$	$\mathcal{N}$**-quantities:** $\mathcal{N}_\times \equiv \frac{r_\times (W(r_s) - W_\times)^{\varpi}}{\Omega(r_\times)}$
Enthalpy-flux: $\mathcal{D}(n,\kappa)(W_s - W_\times)^{n+3/2}$, **Mass-Flux:** $\mathcal{C}(n,\kappa)(W_s - W_\times)^{n+1/2}$ $\frac{\mathcal{L}_\times}{\mathcal{L}} = \frac{\mathcal{B}(n,\kappa)}{\mathcal{A}(n,\kappa)} \frac{W_s - W_\times}{\eta c^2}$	**torus-accretion-rate:** $\dot{m} = \frac{\dot{M}}{\dot{M}_{Edd}}$ **Mass-accretion-rates:** $\dot{M}_\times = \mathcal{A}(n,\kappa) r_\times \frac{(W_s - W_\times)^{n+1}}{\Omega(r_\times)}$ **Cusp-luminosity:** $\mathcal{L}_\times = \mathcal{B}(n,\kappa) r_\times \frac{(W_s - W_\times)^{n+2}}{\Omega(r_\times)}$

Evaluation of these quantities defined in Table (1) are in Figs 2 as functions of the BH dimensionless rotational energy for different models, fixed according to selected values of the fluid specific angular momentum ℓ (fixing the cusp location and the center of maximum density) and the $K_s \in]K_\times, 1[$, fixing the flow thickness at the cusp.[34]

2.3. *The BH rotational energy*

The BH rotational energy is related to the BH geometrical features through its irreducible mass M_{irr}. From the definition of irreducible mass $M_{irr}^2 = \left(M^2 + \sqrt{M^4 - J^2}\right)/2$, where M is the BH total (ADM) mass and the $J = aM$,

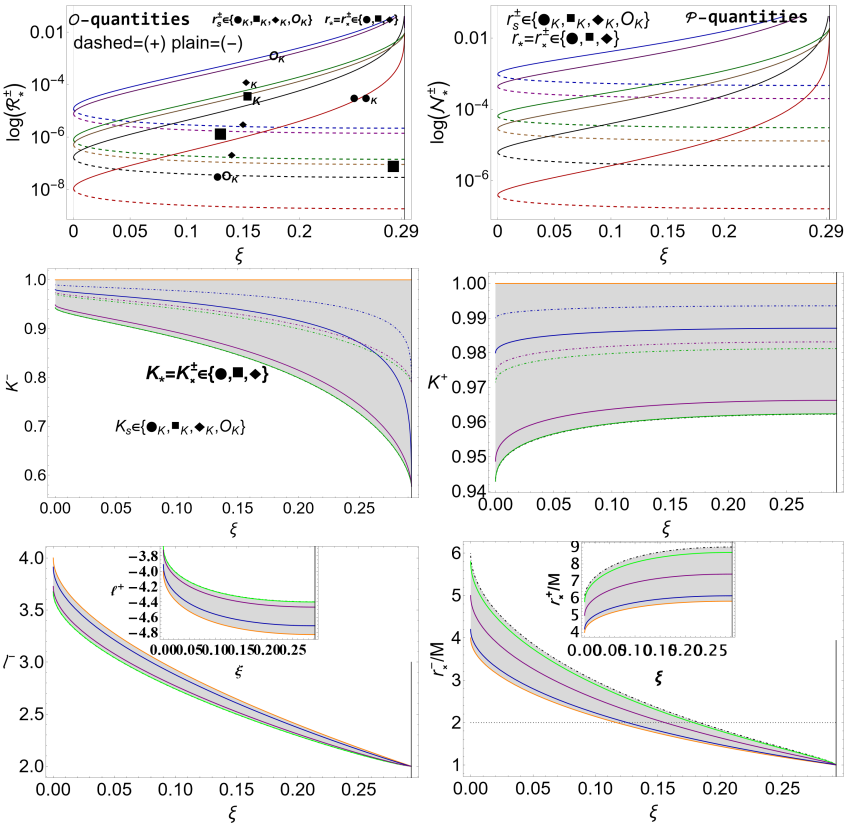


Fig. 2. Plots of $\mathcal{N}_\times^\pm \equiv r_\times(W^\pm(r_s) - W_\times^\pm)^\kappa(\Omega(r_\times^\pm))^{-1}$ for $\mathcal{P}$-quantities analysis and $\mathcal{R}_\times^\pm \equiv (W^\pm(r_s) - W_\times^\pm)^\kappa$ for $\mathcal{O}$-quantities analysis defined in Table (1) for corotating ((-)–continuum curves) and counterrotating ((+)–dashed curves) tori for different values of the cusps $r_* = r_\times^\pm \in \{\bullet, \blacksquare, \blacklozenge\}$ and radii $r_s \in \{\bullet_K, \blacksquare_K, \blacklozenge_K, \mathrm{O_K}\}$, related to thickness of the accreting matter flow, where $\varpi = n+1$, with $\gamma = 1/n+1$ is the polytropic index, κ is the polytropic constant. Radii $(r_\times, r_s)$ and the associated angular momentum ℓ and K parameters are shown with $\{\bullet, \blacksquare, \blacklozenge, \bullet_K, \blacksquare_K, \blacklozenge_K, \mathrm{O_K}\}$. Ω is the relativistic angular velocity. ξ is the dimensionless BH rotational energy.

the dimensionless rotational energy ξ is:

$$\xi \equiv \frac{M_{rot}}{M(0)} = 1 - \frac{\sqrt{1+\sqrt{1-\frac{J(0)^2}{M(0)^4}}}}{\sqrt{2}}, \qquad (3)$$
$$\xi_\pm^\mp = 1 \pm \sqrt{\frac{r_\mp}{2}}, \quad \frac{\delta M_{irr}}{M_{irr}} = \frac{\delta M - \delta J \omega_H^+}{\sqrt{M(0)^2 - \frac{J(0)^2}{M(0)^2}}},$$

where $\xi \equiv \xi_-^+$, M_{rot} is the rotational mass, $r_\pm$ are the outer and inner BH horizons, ω_H^+ is the BH relativistic frequency (light-like limiting circular frequency evaluated at the BH outer horizon r_+). Thee BH dimensionless spin is $a(\xi) = \mathcal{A} \equiv 2\sqrt{-(\xi-2)(\xi-1)^2\xi}$. The rotational energy ξ is governed by the constraint

$\delta M_{irr} \geq 0$ thus $(\delta M - \delta J \omega_H^+) \geq 0$, where $\xi = 1 - M_{irr}/M$, with quantities evaluated at an initial state (0). It is $\xi \in [0, \xi_\ell]$ where $\xi_\ell \equiv \frac{1}{2}\left(2 - \sqrt{2}\right)$, limiting the total rotational energy extracted to a $\approx 29\%$ of the total mass M for a process leading an extreme Kerr BH to the static Schwarzschild BH.

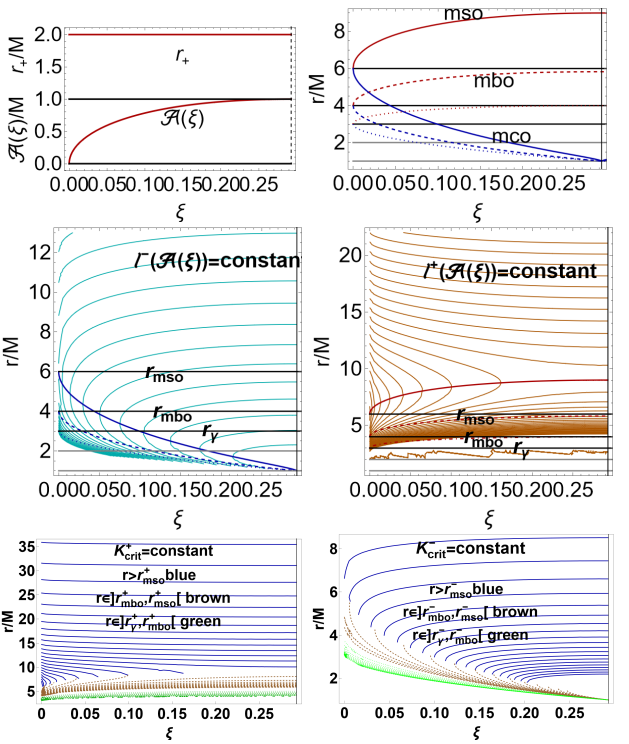


Fig. 3. Upper left panel: the BH outer horizon r_+ and the dimenionless BH spin function $\mathcal{A}(\xi)$, as functions of the dimensionless BH rotational energy ξ. Extreme Kerr BH corresponds to $\xi \approx 0.29$. Upper right panel: marginally bounded orbit *mbo* (dashed curves), marginally stable orbit *mso* (plain curves), marginally circular orbit *mco* (dotted curves) for corotating motion (blue curves) and counter-rotating motion (red curves) as functions of the parameter ξ. Center Bottom panels: the tori specific fluid angular momenta $\ell^{\mp}$ =constant for corotating (center left panel) and counter-rotating (center right panel) tori, the RAD energy function $K^{\pm}(r)$ =constant for the counter-rotating (bottom left panel) and corotating fluids (bottom right panel) as functions of ξ, where $r_\gamma^{\pm}$ is the last circular circular orbit (photon orbit) for counter-rotating (+) and corotating orbits (-).

In Figs 2 and Figs 3 the tori aggregates energetic characteristics at the state 0, prior a possible process involving the BH and its environment, are related to the BH rotational energy. The energetic parameter ξ and the angular relativistic velocity of the BH determine the BH state prior the transition-.[34–36] This analysis turns particularly relevant for the RAD model which, because of its inner discrete

structure, can be characterized by phases of enhanced accretion rates. Furthermore, as each torus of the aggregate is a geometrically thick disk, each component can contribute with super-Eddington accretion rates, with the possibility of further effects as runaway instability,[37,38] or in the case of RAD composed by misaligned tori, the Bardeen-Petterson effect.[39]

2.4. *Magnetized tori: alternative leading function*

As mentioned in Sec. (2.1), the leading function is not necessary the fluid specific angular momentum. An example providing a different aggregate leading function is the case of orbiting agglomerates composed by magnetized tori with a "Komissarov" toroidal magnetic field.[40] In this case the eRAD aggregate leading function is a function $\mathcal{S}$, defined by the magnetic field parameters and proving the rings seeds locations, the minima of pressure (according to the conditions for the cusp formation), the maximum values of the $\mathcal{S}$-parameter for the formation of a tori couple and constrains on the eRAD inner structure in dependence on the tori relative rotation orientation. More precisely, let us consider the toroidal magnetic field: $B^{\phi}=\sqrt{2p_B/A}$, where $A\equiv \ell^2 g_{tt}+2\ell g_{t\phi}+g_{\phi\phi}$, and $p_B=\mathcal{M}\left(g_{t\phi}g_{t\phi}-g_{tt}g_{\phi\phi}\right)^{q-1}\hat{\omega}^q$ is the magnetic pressure, $\hat{\omega}$ is the fluid enthalpy, q and $\mathcal{M}$ are constant; The Euler equation can be written as

$$\frac{\partial_* p}{\rho+p}=G^{(f)}_*+G^{(em)}_*,\quad G^{\natural}_*=-\frac{\partial}{\partial *}W^{\natural}_*;\quad *=\{r,\theta\},\quad \natural=\{(em),(f)\},\tag{4}$$

$$\partial_\mu(W^{(f)}+W^{(em)})=\partial_\mu\left[\ln V_{eff}+\mathcal{G}\right],\ \mathcal{G}(r,\theta)=\mathcal{S}\left(\mathcal{A}V^2_{eff}\right)^{q-1},\ \mathcal{S}\equiv\frac{q\mathcal{M}\hat{\omega}^{q-1}}{q-1}.$$

The RAD leading $\mathcal{S}$–function is

$$\mathcal{S}_{crit}\equiv-\frac{\Delta^{-\mathcal{Q}}}{\mathcal{Q}}f(a,\ell;r)\quad(\mathcal{Q}\equiv q-1),$$

where $f(a,\ell;r)$ is a function of the BH spin and fluid angular momentum[9,31] (there is $\Delta\equiv r^2-2Mr+a^2$).

3. Misaligned Tori

In this section we discuss the tori aggregates containing misaligned tori. To simply our discussion we consider misaligned tori orbiting a central static BH. In this case we can use different leading functions for the description of the aggregate introduced in Sec. (3.1). Tori geometrical thickness is discussed in Sec. (3.2). Frequency models in tori aggregates are considered in Sec. (3.2.1) and tori geometrical characteristics are deepened in Sec. (3.3) for quiescent (not cusped) tori in Sec. (3.3.1) and for cusped tori analyzed in Sec. (3.3.2).

3.1. *Leading functions*

In this section we introduce three definitions of leading functions for the tori aggregates–see.[11]

The critical $\bar{r}(r_i)$ and $r^{\varepsilon}_{\times}$ functions:

$$\bar{r}(r_i) = \frac{2r_i\left[\sqrt{2r_i - r_\gamma} + r_i - 1\right]}{(r_i - r_+)^2}, \quad r^{\varepsilon}_{\times} \equiv \frac{r_{center}\left(\sqrt{2r_{center} - r_\gamma} + 1\right)^2}{(r_{center} - r_+)^2}, \quad (5)$$

(solutions of $\ell(r) = \ell(r_p)$ for two orbits (r, r_p)). In this case the leading function is a relation between the extremes of pressure inside the tori $\bar{r}(r_i)$, or the cusp $r^{\varepsilon}_{\times}$ as function of the center of maximum pressure. The distance $r_{center} - r_\times$ increases with the tori distance in the aggregates from the central BH attractor.

Leading function $\ell^o_{crit}(K)$:

There is

$$\ell^o_{crit}(K) \equiv \sqrt{\frac{27K^4 - K\left(9K^2 - 8\right)^{3/2} - 36K^2 + 8}{2K^2\left(K^2 - 1\right)}}, \quad (6)$$

$$\ell^i_{crit}(K) \equiv \sqrt{\frac{27K^4 + K\left(9K^2 - 8\right)^{3/2} - 36K^2 + 8}{2K^2\left(K^2 - 1\right)}}, \quad (7)$$

$$\text{where} \quad \ell^o_{crit}(K) > \ell^i_{crit}(K) > \ell_{mso}, \quad \ell^i_{crit}(K) \in [\ell_{mso}, \ell_\gamma[.$$

In this case, notably the leading function, relating the two tori parameters ℓ and K, bas been split in function $\ell^o_{crit}(K)$, the leading function providing parameters (ℓ, K) at the tori centers (rings seeds), and function $\ell^i_{crit}(K)$ for the tori cusps.

Similarly we introduce the following alternative functions.

Tori critical radius $r^o_{crit}(K)$: There is

$$r^o_{crit}(K) \equiv -\frac{8}{K\left(\sqrt{9K^2 - 8} + 3K\right) - 4}, \; r^i_{crit}(K) \equiv \frac{8}{K\left(\sqrt{9K^2 - 8} - 3K\right) + 4},$$
$$r^i_{crit}(K_\times) = r_\times, \; r^o_{crit}(K_{center}) = r_{center} \quad (8)$$

similarly to $\ell^o_{crit}(K)$, the leading function is $r^o_{crit}(K)$, relating the center of maximum pressure (ring seed) and the K parameter at the torus center.

3.2. *Geometrical thickness*

Disk geometrical thickness is an important characteristic for the RAD tori. The eRAD is a geometrical thin disk composed by geometrical thick tori with an inner articulated ringed structure, combining some features of geometrical thick disks, inherited by its components and features typical of the geometrical thin disks in its global structure. Geometrical thickness is a relevant parameter in the comparison with other disks model, in the assessment of the torus vertical structure and the influence of a possible poloidal magnetic field, for the accretion mechanism and the study of tori oscillations. The definition of geometrical thickness adopted here coincides with the thickness **S** of the outer Roche lobe section of the PD torus. For large part of the (ℓ, K) range, cusped tori can be considered geometrically thin i.e. $\mathbf{S} < 1$. There are classes of toroidal components with equal thickness. For example, in the "reference" case $\mathbf{S} = 1$, distinguishing geometrically thin and geometrically

thick disks, where there are couples of toroids with equal energy parameter K, regulating also the flow thickness.[10,11]

A further parameter for the evaluation of tori geometrical thickness is the dimensionless β_{crit}

$$\beta_{crit} = \frac{(r_{center}-2)^2(r_{center}-r_\times)\sqrt{r_{center}r_\times - 2(r_{center}+2r_\times)}}{r_{center}\sqrt{r_{center}-3}r_\times\sqrt{r_\times-2}},$$

emerging from the analysis of cusped tori oscillation.[41–43] Similarly to the leading function $\bar{r}(r_i)$ of Eq. (5), β_{crit} depends on the distance between the maximum and minimum point of pressure in the tori, which increases with the distance from the central attractor. For small β_{crit} ($\beta_{crit} \geq 0$), tori may be considered geometrically thin for radial and vertical oscillation, and can be described by the radial and vertical epicyclical frequencies from the hypothesis of thin (slender) tori, coincident therefore with the respective circular orbit frequencies.[42,44,45] The conditions for geometrical thin components according to definition $\mathbf{S} < 1$ and conditions for geometrical thin tori according to β_{crit} coincide only for special conditions on ℓ and K parameters (therefore depending on the tori location in the agglomerate and their dimension), having tori with combined characteristics typical of geometrical thin and thick disks.[10]

3.2.1. *Frequency models in tori aggregates*

In the conditions where $\beta_{crit} \geq 0$, we can consider the circular orbit approximation for the oscillation frequencies. In[10] different frequency models are applied to the RAD structure, interpreted as a frame for the high–frequency (HF) Quasi-Periodic Oscillations (QPOs), assuming the geodesic (nearly circular geodesic motion) frequencies

$$\nu_r(r) = \nu_K(r)\sqrt{1-\frac{r_{mso}}{r}}, \quad \nu_\theta(r) = \nu_K(r) \equiv \frac{1}{r^{3/2}}, \tag{9}$$

determined by the tori constraints. The frequencies (9) are combined for the fitting of resonance ratios, identifying the upper ν_U and lower ν_L frequencies from different oscillation models and assuming $(\nu_r(r), \nu_K(r))$ evaluated at different points r of the tori surfaces. Therefore in[10] we used the frequency models (**TD**,**RP**,**RE**,**WD**) listed Table (2), evaluated in different tori models. The twin peak quasi-periodic oscillations with resonant frequency ratios $\nu_U/\nu_L = \{3:2, 4:3, 5:4, 2:1, 3:1\}$ have been analyzed.[10]

Different components of the aggregates fit different frequency models, according to tori location in the aggregate with respect to the central attractor distinguishing therefore the toroidal components and the different torus active parts.[10] In the models of Table (2), the torus inner edge has been considered the active part of the emission process, the frequencies being evaluated at $r_{inner} > r_\times$ (for quiescent tori) or $r_{inner} = r_\times$ for cusped tori, and as ($\beta_{crit} \gtrapprox 0, \mathbf{S} \ll 1$), the maximum of

Table 2. Frequency models and related tori models for the analysis of Sec. (3.2.1). Frequencies (ν_θ, ν_r) are in Eq. (9), the relativistic frequency $\nu_K = \Omega$, coincident with ν_θ in the static spacetime is in Eq. (1). The fluid specific angular momentum is ℓ and K is the energy parameter, r_{center} is the center of maximum density and pressure in the torus, r_{out} is the torus outer edge. The relativistic-precession model coincides in static BH also with the total precession models **TP**. Radius $r^{\times}_{\rm max}$ is the cusped torus geometrical maximum of Eq. (13), where $r^{o}_{\rm max}(r) \equiv r^{\times}_{\rm max}$ as functions of ℓ. Radii $\bar{r}(r_i)$ and $r^{\epsilon}_{\times}$ are the aggregate leading functions of Eqs (5); function $\bar{r}(r_i)$ relates the critical points of pressure in each toroidal component of the aggregate, and function $r^{\epsilon}_{\times}$ provides the torus cusp as function of the torus center–see Eq. (5).

Frequency models	**Tori models**
(RP): $\nu_U = \nu_K$, $\nu_L = \nu_{per} \equiv \nu_K - \nu_r$;	[**(a)-model**]: function of r/M
(relativistic-precession model)	[**(b)-model**]: $r = r_{center}(\ell)$ function of $\ell \in [\ell_{mso}, \ell_{mbo}]$
(RE): $\nu_U = \nu_\theta$, $\nu_L = \nu_r$;	[**(c)-model**]: for $r = r_{out}(\ell)$ functions of ℓ
(resonance epicyclic models)	[**(d)-model**]: in $r = r^{\times}_{\rm max}(\ell)$ function of ℓ
(**TD**): $\nu_L = \nu_K$, $\nu_U = (\nu_K + \nu_r)$;	[**(e)-model**]: for $r = r^{\epsilon}_{\times}$ function of r/M torus cusp or center
(tidal distortion model)	[**(f)-model**]: function of $\bar{r}(r_i)$
(**WD**): $\nu_L = 2(\nu_K - \nu_r)$, $\nu_U = (2\nu_K - \nu_r)$;	[**(g)-model**]: for $r = r_{center}(K)$ function of K
(warped disk model)	

pressure point, the outer edge and the geometrical maximum point have been also considered.[46–57]

3.3. *Tori geometrical characteristics*

Frequencies models of Table (2) have been evaluated on the outer r_{out} and the inner $r_\times$ tori edges, the tori geometrical maximum $r_{\rm max}$ and tori center r_{center}. The evaluation of the tori geometrical characteristics is relevant in the determination of inner ringed structure and tori collision. In this section we provide $(r_\times, r_{out}, r_{center}, r_{\rm max})$, the torus height h and the inner Roche lobe maximum high. These quantities are functions of tori parameters (ℓ, K) for quiescent (not cusped) tori, considered in Sec. (3.3.1) and ℓ *or* K (or alternately the critical pressure points $r_\times$ and r_{center}) for cusped tori analyzed in Sec. (3.3.2).

3.3.1. *Roche lobes in quiescent and cusped tori*

For quiescent and cusped tori, we provide below the outer and inner torus edge and the tori elongations λ on the tori symmetry plane–see[11]

$$\textbf{Torus outer edge:} \quad r_{out} \equiv \frac{2\left[1 + \mathbb{K}\tau\cos\left(\frac{1}{3}\cos^{-1}(\alpha)\right)\right]}{3\mathbb{K}}, \tag{10}$$

$$\textbf{Tori elongation:} \quad \lambda \equiv \frac{2\tau\cos\left(\frac{1}{6}\left[2\cos^{-1}(\alpha) + \pi\right]\right)}{\sqrt{3}},$$

$$\textbf{Torus inner edge:} \quad r_{inner} \equiv \frac{2\left[1 - \mathbb{K}\tau\sin\left(\frac{1}{3}\sin^{-1}(\alpha)\right)\right]}{3\mathbb{K}}.$$

where $\mathbb{K}:\ K\equiv\sqrt{1-\mathbb{K}}$, $\mathcal{Q}\equiv\ell^2$ and $(\alpha,\mathbb{K},\tau)$ are functions of (ℓ,K)–see.[11] (Note in the case of cusped tori $r_{inner}=r_\times$).

For the inner Roche lobe, the inner edge an the elongation of the lobe on the symmetry plane are

$$r_{inner}^{BH}\equiv\frac{2\left[\frac{1}{\mathbb{K}}-\tau\sin\left(\frac{1}{6}\left[2\cos^{-1}(\alpha)+\pi\right]\right)\right]}{3},$$
$$\lambda_{inner}^{BH}\equiv\frac{2}{3}\tau\left[\sin\left(\frac{1}{6}\left[2\cos^{-1}(\alpha)+\pi\right]\right)-\sin\left[\frac{1}{3}\sin^{-1}(\alpha)\right]\right]. \tag{11}$$

The geometrical maximum for the outer $r^o_{\max}$ and inner Roche lobes $r^i_{\max}$ are

$$\text{Outer lobe}:\quad r^o_{\max}(K,\ell)\equiv\sqrt{\frac{K^2\mathcal{Q}}{K^2-1}+4\sqrt{\frac{2}{3}}\psi\cos\left[\frac{1}{3}\cos^{-1}(\psi_\pi)\right]},$$
$$\text{Inner lobe}:\quad r^i_{\max}(K,\ell)\equiv\sqrt{\frac{K^2\mathcal{Q}}{K^2-1}-4\sqrt{\frac{2}{3}}\psi\sin\left[\frac{1}{3}\sin^{-1}(\psi_\pi)\right]},$$

and the torus height is

$$h^o_{\max}(K,\ell)\equiv\sqrt{\frac{K^2\mathcal{Q}}{1-K^2}+\mathbf{Z}-4\sqrt{\frac{2}{3}}\psi\cos\left[\frac{1}{3}\cos^{-1}(\psi_\pi)\right]},$$

where $(\psi,\psi_\pi,\mathbf{Z})$ are functions of (ℓ,K).[11]

3.3.2. *Cusped tori*

In this section we specialize the analysis of Sec. (3.3.1) for cusped misaligned tori, which are described by one only independent parameter K or ℓ (or equivalently $r_\times$ or r_{center}).

We can express the critical points of pressure in the tori in terms of the parameter ℓ. The torus center and the point of minimum density (and hydrostatic pressure) are

$$r_{center}(\ell)\equiv\frac{1}{3}\left[\mathcal{Q}+2L_\ell\cos\left(\frac{1}{3}\iota_a\right)\right],\quad r_\times(\ell)\equiv\frac{1}{3}\left[\mathcal{Q}-2L_\ell\cos\left(\frac{1}{3}\left[\iota_a+\pi\right]\right)\right],$$

where $L_\ell, L_{\updownarrow\updownarrow}$ are functions of ℓ and $\mathcal{Q}\equiv\ell^2$.[11] The cusped torus outer edge is located at

$$r^\times_{out}\equiv\frac{2\ell^2\hat{\psi}_2}{3\ell^2\hat{\psi}_2+\hat{\psi}_0^2(6-\hat{\psi}_0)}+2\hat{\psi}_4\sqrt{\frac{\ell^2\left[\hat{\psi}_0^2\left(3\ell^2\hat{\psi}_2(6-\hat{\psi}_0)+\hat{\psi}_0^2(\hat{\psi}_0^2-12\hat{\psi}_0+36)\right)+12\ell^2\hat{\psi}_2^2\right]}{3\left(3\ell^2\hat{\psi}_2+\hat{\psi}_0^2(6-\hat{\psi}_0)\right)^2}},$$

(where $(\hat{\psi}_0, \hat{\psi}_2)$ are functions of ℓ). Similarly to functions (ℓ^o, ℓ^i) of Eqs (6), the leading function can be expressed in terms of the energy function K as

$$K_{center}(\ell) \equiv \sqrt{\frac{\left[\mathcal{Q} + 2L_\ell \cos\left(\frac{\iota_a}{3}\right) - 6\right]\left[\mathcal{Q} + 2L_\ell \cos\left(\frac{\iota_a}{3}\right)\right]^2}{3\mathcal{Q}\left[3\ell^4 + 2\left(2\mathcal{Q} - 15\right)L_\ell \cos\left(\frac{\iota_a}{3}\right) - 39\mathcal{Q} + 2L_\ell^2 \cos\left(\frac{2\iota_a}{3}\right) + 54\right]}},$$

$$K_\times(\ell) \equiv \sqrt{\frac{\left[\mathcal{Q} - 2L_\ell \sin\left(\frac{\iota_b}{3}\right) - 6\right]\left(\mathcal{Q} - 2L_\ell \sin\left(\frac{\iota_b}{3}\right)\right)^2}{3\mathcal{Q}\left[3\ell^4 + 2\left(15 - 2\mathcal{Q}\right)L_\ell \sin\left[\frac{\iota_b}{3}\right] - 39\mathcal{Q} - 2L_\ell^2 \cos\left(\frac{2\iota_b}{3}\right) + 54\right]}},$$

$$\iota_a \equiv \cos^{-1}(L_{\updownarrow\updownarrow}), \quad \iota_b \equiv \sin^{-1}(L_{\updownarrow\updownarrow}) \tag{12}$$

(here $L_\updownarrow$ is a function of the momentum ℓ). Function $K_{center}(\ell)$ describes the rings seeds, while $K_\times(\ell)$ refers the tori cusps ($\mathcal{Q} \equiv \ell^2$).

The cusped tori outer edge can be expressed as function of the cusp $r_\times$ as follows:

$$r_{out}^\times(r_\times) = \frac{2}{3}\left[\sqrt{\frac{(r_\times - r_{mso})^2 r_\times^2}{(r_\times - r_{mbo})^2}}\ \cos\left[\frac{1}{3}\cos^{-1}\mathbf{X}\right] + \frac{r_\times}{r_\times - r_{mbo}} + r_\times\right]$$

where $\mathbf{X}$ is a function of $r_\times$, and the the geometrical maxima of cusped tori for the outer (o) and inner (i) Roche lobes are

$$r_{\max}^o(r) = \sqrt{4\sqrt{\frac{2}{3}}\psi_\lambda \cos\left[\frac{1}{3}\cos^{-1}\left(-\frac{3}{4}\sqrt{\frac{3}{2}}\psi_\lambda\psi_\sigma^2\right)\right] + \frac{r^2}{(r - r_\gamma)\psi_\sigma}}$$

$$r_{\max}^i(r) = \sqrt{\frac{r^2}{(r - r_\gamma)\psi_\sigma} - 4\sqrt{\frac{2}{3}}\psi_\lambda \cos\left[\frac{1}{3}\left(\cos^{-1}\left[-\frac{3}{4}\sqrt{\frac{3}{2}}\psi_\lambda\psi_\sigma^2\right] + \pi\right)\right]},$$

(here r is the fluid pressure critical point where $\psi_\lambda, \psi_\sigma$ are functions of r).

The cusped torus height is

$$h_{\max}^o(r_\times) = \left(-2\sqrt{6}\sqrt{\frac{(r_\times - r_\gamma)(r_\times - r_+)^2 r_\times^4}{(r_\times - r_{mbo})^3}}\sec\left[\frac{1}{3}\cos^{-1}(\psi_\rho)\right] + \right.$$

$$\left.\frac{9(r_\times - r_+)^2 r_\times^2 \sec^2\left[\frac{1}{3}\cos^{-1}(\psi_\rho)\right]}{8(r_\times - r_{mbo})(r_\times - r_\gamma)} + \frac{(r_\times - r_+)(5r_\times - 18)r_\times^2}{(r_\times - r_{mbo})^2}\right)^{1/2},$$

(ψ_σ is a function of $r_\times$.) Finally we can express the cusped tori geometrical thickness $\mathbf{S}_\times = 2h_\times/(\lambda_\times)$ in terms of the pressure critical points where $\lambda_\times$ is the cusped torus elongation on its symmetry plane and $h_\times$ the cusped torus height.

4. Conclusion

We explored models of tori clusters orbiting around a central SMBH, detailing the morphological characteristics of the toroidal components. Configurations considered here can be used as initial data for dynamical (time-dependent, evolutive) GRMHD analysis. A "leading function" as been used to constraint the tori distribution around the central attractor, together with the energy function $K(r)$ regulating the agglomerate stability (cusp emergence and tori collision), the flow thickness, mass accretion rate and cusp luminosity.

From the observational viewpoint the inner ringed structure offers several interesting scenarios arising from the unstable states associated to its inner activity, as the presence of multiple accretion points and inter disk shells of multiple jets. Eventually observational evidence of the RAD and the associated inter disk activity could be found in the obscuration of the X-ray emission spectrum, as a track of the agglomerate inner composition. An indication of the presence of multiple orbiting tori could be seen in an increasing BH accretion mass rate and the presence of interrupted phases of BH accretion, or in the emission associated to oscillation tori modes as in HF QPOs. The establishment of runaway instability and the tori self-gravity can be relevant further factors for eRAD tori agglomerate around SMBHs[37,38] and the Bardeen–Petterson effect is main relevant in the misaligned tori case.[39]

References

1. D. Pugliese & G. Montani, Phys. Rev. D, 91, 083011 (2015).
2. D. Pugliese & Z. Stuchlik, Astrophys. J. Suppl., 221, 2, 25 (2015).
3. D. Pugliese & Z.Stuchlík, Astrophys. J. Suppl., **223**, 2, 27 (2016).
4. D. Pugliese & Z. Stuchlik, JHEAp, 17, 1 (2018).
5. D. Pugliese & Z. Stuchlík,Class. Quant. Grav. **35**, 18, 185008 (2018).
6. D. Pugliese & Z. Stuchlík, Astrophys. J. Suppl., **229**, 2, 40 (2017).
7. D. Pugliese & Z. Stuchlik, Eur. Phys. J. **C 79** 4, 288, (2019).
8. D. Pugliese & Z. Stuchlík, Class. Quant. Grav., **35** 10, 105005 (2018)
9. D. Pugliese and G. Montani, Mon. Not. Roy. Astron. Soc., 476, 4, 4346 (2018).
10. D. Pugliese & Z. Stuchlik, Mon. Not. Roy. Astron. Soc., 493, 3, 4229–4255(2020).
11. D. Pugliese & Z. Stuchlik Class. Quant. Grav., 37 19, 195025 (2020).
12. D. Pugliese & Z. Stuchlik, to appear in PASJ (2021).
13. M. A. Abramowicz & P. C. Fragile, Living Rev. Relativity, 16, 1 (2013).
14. C. M. Violette Impellizzeri et al., Astrophys. J. L, 884, L28 (2019).
15. S. Dyda, R. V. E. Lovelace, et al., Mon. Not. Roy. Astron. Soc., **446**, 613-621 (2015).
16. C. Alig, M. Schartmann, A. Burkert, K. Dolag, Astrophys. J., 771, 2, 119 (2013).
17. J. M. Carmona-Loaiza, M. Colpi, M. Dotti et al., Mon. Not. Roy. Astron. Soc., 453, 1608 (2015).
18. R. V. E. Lovelace & T. Chou, Astrophys. J., 468, L25 (1996).
19. R. V. E. Lovelace, M. M. Romanova, P. Lii, et al. Comp. Astroph. Cosmology, 1-3, (2014).
20. C. Nixon, A. King, & D. Price, Mon. Not. Roy. Astron. Soc., 434, 1946 (2013).
21. S. Doğan, C. Nixon, A. King, *et al.*, Mon. Not. Roy. Astron. Soc., 449, 1251 (2015).
22. C. Bonnerot, E. M. Rossi, G. Lodato, Mon. Not. Roy. Astron. Soc., 455, 2, 2253 (2016).

23. H. Aly, W. Dehnen, C. Nixon & A. King, Mon. Not. Roy. Astron. Soc., 449, 1, 65 (2015).
24. O. Zanotti & D. Pugliese, Gen. Rel. Grav., 47, 4, 44 (2015).
25. M. Kozłowski, M. Jaroszyński, M. A. Abramowicz, Astron. Astrophys., 63, 209 (1998).
26. M. A. Abramowicz, Acta. Astron., 21, 81, (1971).
27. S. K. Chakrabarti, Mon. Not. Roy. Astron. Soc., 245, 747 (1990).
28. S. K. Chakrabarti, Mon. Not. Roy. Astron. Soc., 250, 7 (1991).
29. D. Pugliese and G. Montani, Gen. Rel. Grav., 53, 5, 51 (2021).
30. D. Pugliese and J. A. Valiente Kroon, Gen.Rel.Grav. 44, 2785-2810 (2012).
31. D. Pugliese and G. Montani, EPL, 101, 1, 19001 (2013).
32. R. H. Boyer, Proc. Camb. Phil. Soc., **61**, 527 (1965).
33. M. A. Abramowicz, Astronomical Society of Japan, 37, 4, 727–734 (1985).
34. D. Puglese & Z. Stuchlik, Class. Quant. Grav., 38, 14, 145014 (2021).
35. R. A. Daly, APJ, 691, L72-L76, 1 (2009).
36. R. A. Daly and T. B. Sprinkle, Mon. Not. Roy. Astron. Soc., 438, 3233-3242 (2014).
37. M. A. Abramowicz, M. Calvani, L. Nobili, Nature, 302, 597 (1983).
38. J. A. Font and F. Daigne, Mon. Not. Roy. Astron. Soc. **334**, 383 (2002).
39. J. M. Bardeen, J. A. Petterson, Astrophys. J., 195, L65 (1975).
40. S. S. Komissarov, Mon. Not. Roy. Astron. Soc., 368, 993 (2006).
41. G. Török, K. Goluchová, J. Horák, et al., Mon. Not. Roy. Astron. Soc., **457**, L19 (2016).
42. O. Straub & E. Sramkova, Class. Quantum Grav., **26**, 055011 (2009).
43. M. A. Abramowicz, O. M. Blaes, J. Horak et al., Class. Quantum Grav., **23**, 1689 (2006).
44. K. Goluchova, J. Horak et al., MNRAS, 457, L19 (2016).
45. K. Goluchova, M. Urbanec et al., Astrophys. J., 833, 273 (2016).
46. Z. Stuchlìk, J. Schee, E. Šràmkovà & G. Török, Acta Astron. **67**, 181 (2017).
47. Z. Stuchlík & M. Kološ, A&A, **586**, A130 (2016).
48. Z. Stuchlík, P. Slaný & G. Török, A&A, **463**, 807 (2007).
49. G. Török & Z. Stuchlík, A&A, **437**, 775 (2005).
50. A. Kotrlová, E. Šrámková, G. Török et al, A&A, **607**, A69 (2017).
51. G. Török, K. Goluchová, M. Urbanec et al., Astrophys. J., **833**, 273 (2016).
52. E. Šrámková, G. Török, A. Kotrlová, et al., A&A, **578**, A90 (2015).
53. G. Török, A. Kotrlová, E. Šrámková et al., A&A, **531**, A59 (2011).
54. Z. Stuchlík, A. Kotrlová & G. Török, A&A, **525**, A82 (2011).
55. A. Kotrlová, Z. Stuchlík & G. Török, Class. Quantum Grav., **25**, 225016 (2008).
56. Z. Stuchlík, P. Slaný & P. Török, A&A, **470**, 401 (2007).
57. Z. Stuchlik, A. Kotrlová, G. Török, A&A, 552, A10 (2013).

Hydrodynamical transport of angular momentum in accretion disks in the presence of nonlinear perturbations due to noise

Subham Ghosh* and Banibrata Mukhopadhyay†
Department of Physics, Indian Institute of Science, Bangalore, Karnataka, 560012, India
*subham@iisc.ac.in
†bm@iisc.ac.in

The origin of hydrodynamical instability and turbulence in the Keplerian accretion disk is a long-standing puzzle. The flow therein is linearly stable. Here we explore the evolution of perturbation in this flow in the presence of an additional force. Such a force, which is expected to be stochastic in nature hence behaving as noise, could result from thermal fluctuations (however small be), grain–fluid interactions, feedback from outflows in astrophysical disks, etc. We essentially establish the evolution of nonlinear perturbation in the presence of Coriolis and external forces, which is the modified Landau equation. We obtain that even in the linear regime, under suitable forcing and Reynolds number, the otherwise least stable perturbation evolves to a very large saturated amplitude, leading to nonlinearity and plausible turbulence. Hence, forcing essentially leads a linear stable mode to unstable. We further show that nonlinear perturbation diverges at a shorter time-scale in the presence of force, leading to a fast transition to turbulence. Interestingly, the emergence of nonlinearity depends only on the force but not on the initial amplitude of perturbation, unlike the original Landau equation-based solution.

Keywords: Accretion, accretion disks – hydrodynamics – instabilities – turbulence

1. Introduction

Origin of instability and plausible turbulence in Rayleigh stable flows, e.g., the Keplerian accretion flow, is a long-standing problem. While such flows are evident to be turbulent, they are linearly stable for any Reynolds number (Re). Although people[1,2] described the underlying flow in accretion disks to be turbulent with an effective turbulent viscosity, the origin of turbulence therein was not clear until Balbus and Hawley[3] proposed the idea of magnetorotational instability (MRI) following Velikhov[4] and Chandrasekhar.[5] MRI is a Magnetohydrodynamical (MHD) instability, and it operates due to the coupling between the weak magnetic field and the rotation of the fluid parcel. Although MRI is a popular instability mechanism, it poses several problems. There are astrophysical bodies where the ionization fraction is tiny as they are cold. In such systems, MRI gets suppressed.[6] Apart from that, beyond a certain value of the toroidal component of the magnetic field for compressible plasma, MRI gets suppressed.[7,8] There are several other cites too where MRI seizes to work or becomes suppressed.[9,10] We, therefore, venture for a hydrodynamical origin of nonlinearity and hence plausible turbulence in the accretion disk. Our emphasis is the conventional linear instability when perturbation grows

exponentially, unlike the case of transient growth. We particularly consider here an extra force,[11] to fulfill our purpose. The examples[12] of the origin of such force in astrophysical context, particularly in accretion disks, could be: the interaction between the dust grains and fluid parcel in protoplanetary disks, back reactions of outflow/jet to accretion disks.

2. Formalism of the problem

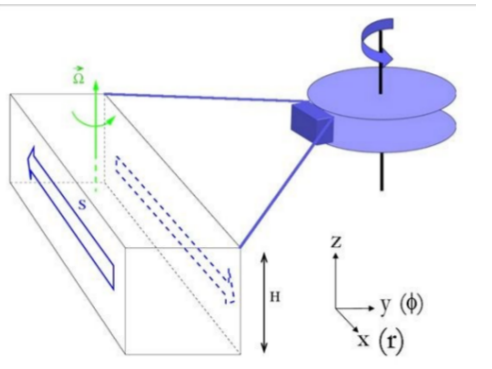


Fig. 1. Model picture of local cubical box in accretion disk where we perform the analysis (http://ipag.obs.ujf-grenoble.fr/ longarep/astrophysics.html). Within the box, the Cartesian coordinate x is along the radial cylindrical coordinate r (with respect to the center of the accretion disk), y is along ϕ, and z is same in both the systems.

To formulate the problem, we have considered a local box at a local patch in the accretion disk, as shown in the Fig. 1. Although the accretion disk possesses cylindrical geometry, In the local box, we study the motion of the local fluid element using the Cartesian coordinates. The one to one correspondence between the Cartesian (x, y, z) and cylindrical (r, ϕ, z) coordinates are shown in the Fig. 1. The Cartesian coordinates x, y, z are along the cylindrical coordinates r, ϕ, z. The details of the local formulation are in Mukhopadhyay et al.,[13,14] and Ghosh and Mukhopadhyay.[15] As we confine ourselves in the local region of the accretion disks, the corresponding flow has been considered incompressible.[9,13,16–18] The governing equations are the ensemble-averaged Orr-Sommerfeld and Squire equations in the presence of Coriolis force and an additional stochastic force,[12,19] which is delta correlated with nonzero mean. The Orr-Sommerfeld and Squire equations are obtained by eliminating the pressure term from the corresponding Navier-Stokes equations and utilizing the continuity equation for the incompressible flow. The governing equations are given by,

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial y}\right)\nabla^2 u - U''\frac{\partial u}{\partial y} + \frac{2}{q}\frac{\partial \zeta}{\partial z} - \frac{1}{Re}\nabla^4 u + \Gamma_1 = NL^u, \tag{1}$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial y}\right)\zeta - U'\frac{\partial u}{\partial z} - \frac{2}{q}\frac{\partial u}{\partial z} - \frac{1}{Re}\nabla^2\zeta + \Gamma_2 = NL^{\zeta}, \tag{2}$$

where u and ζ are respectively the x-component of the velocity and vorticity perturbations after ensemble averaging, U the y-component of background velocity which for the present purpose of plane shear in the dimensionless units is $-x$, q the rotation parameter with $\Omega(r) \propto 1/r^q$, $\Omega(r)$ being the angular frequency of the fluid parcel at radius r, Γ-s are the corresponding constant means of stochastic forces (white noise with nonzero mean due to gravity making the system biased[11]) in the system described below, and NL-s are the non-linear terms due to the perturbation. The x-component of vorticity and the non-linear terms are given by

$$\zeta = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}, \tag{3}$$

$$NL^u = -\nabla^2\{(\mathbf{u}'\cdot\nabla)u\} + \frac{\partial}{\partial x}\nabla\cdot\{(\mathbf{u}'\cdot\nabla)\mathbf{u}'\}, \tag{4}$$

$$NL^{\zeta} = -\frac{\partial}{\partial y}\{(\mathbf{u}'\cdot\nabla)w\} + \frac{\partial}{\partial z}\{(\mathbf{u}'\cdot\nabla)v\}, \tag{5}$$

where $\mathbf{u}' = (u, v, w)$, which is the perturbed velocity vector. For the complete description of the flow, equations 1 and 2 are supplemented with the equation of continuity, which is given by

$$\nabla\cdot\mathbf{u}' = 0, \tag{6}$$

To solve the governing equations, we have no-slip boundary conditions along x-direction,[12,13,20] i.e. $u = v = w = 0$ at $x = \pm 1$ or equivalently

$$u = \frac{\partial u}{\partial x} = \zeta = 0, \text{ at } x = \pm 1. \tag{7}$$

2.1. *Linear theory*

In the evolution of linear perturbation, let the linear solutions be

$$u = \hat{u}(x,t)e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{8}$$

$$\zeta = \hat{\zeta}(x,t)e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{9}$$

with $\mathbf{k} = (0, k_y, k_z)$ and $\mathbf{r} = (0, y, z)$. Substitute these in equations (1) and (2), neglecting non-linear terms, we obtain

$$\frac{\partial}{\partial t}Q + i\mathcal{L}Q + \Gamma = 0, \tag{10}$$

where

$$Q = \begin{pmatrix}\hat{u}\\ \hat{\zeta}\end{pmatrix}, \ \mathcal{L} = \begin{pmatrix}\mathcal{L}_{11} & \mathcal{L}_{12}\\ \mathcal{L}_{21} & \mathcal{L}_{22}\end{pmatrix}, \ \mathcal{D} = \frac{\partial}{\partial x}, \tag{11}$$

$$\mathcal{L}_{11} = (\mathcal{D}^2 - k^2)^{-1}\Big[k_y U\left(\mathcal{D}^2 - k^2\right) - k_y U'' - \frac{1}{iRe}\left(\mathcal{D}^2 - k^2\right)^2\Big],$$
$$\mathcal{L}_{12} = \left(\mathcal{D}^2 - k^2\right)^{-1}\frac{2k_z}{q},$$
$$\mathcal{L}_{21} = -\left(U' + \frac{2}{q}\right)k_z,$$
$$\mathcal{L}_{22} = k_y U - \frac{1}{iRe}\left(\mathcal{D}^2 - k^2\right),$$

and

$$\Gamma = e^{-i\mathbf{k}\cdot\mathbf{r}}\begin{pmatrix}(\mathcal{D}^2 - k^2)^{-1}\Gamma_1\\ \Gamma_2\end{pmatrix}. \tag{12}$$

Let us subsequently assume the trial solution of the equation (10) be

$$Q = AQ_x e^{-i\sigma t} - \frac{1}{\mathcal{D}_t + i\mathcal{L}}\Gamma, \tag{13}$$

where σ is the eigenvalue corresponding to the particular mode and it is complex having real (σ_r) and imaginary (σ_i) parts,

$$Q_x = \begin{pmatrix}\phi^u(x)\\ \phi^\zeta(x)\end{pmatrix} \tag{14}$$

and $\mathcal{D}_t$ stands for $\partial/\partial t$. Q_x is the eigenfunction corresponding to the homogeneous part of the equation (10), i.e., Q_x satisfies $\mathcal{L}Q_x = \sigma Q_x$. The first term of the right-hand side of the equation (13) is due to the homogeneous part of the equation (10), and the second term is due to the inhomogeneous part, i.e., the presence of Γ, of the same equation. Hence, Q is influenced by the force Γ. However, the typical eigenspectra for the Keplerian flow ($q = 1.5$), constant angular momentum flow ($q = 2$) and plane Couette flow ($q \to \infty$) for $Re = 2000$ and $k_y = k_z = 1$ are shown in the Fig. 2. Since $\mathcal{L}_{12}$ and $\mathcal{L}_{21}$ in the equation (11) are zero for the plane Couette flow and constant angular momentum flow, respectively, we obtain the same eigenspectra for both plane Couette, and constant angular momentum flows. We perform the whole analysis for the least stable modes for the respective flows, and these least stable modes are shown in the dotted box in Fig. 2.

2.2. *Nonlinear theory*

For the non-linear solution, following similar work but in the absence of force,[20–22] we assume the series solution for velocity and vorticity, i.e.

$$\psi = \sum_{n\to-\infty}^{\infty}\psi_n = \sum_{n\to-\infty}^{\infty}\bar{\psi}_n(t,x)e^{in(\mathbf{k}\cdot\mathbf{r}-\sigma_r t)}, \tag{15}$$

when obviously $\bar{\psi}_{-n} = \bar{\psi}_n^*$ and ψ could be any one of u and ζ.

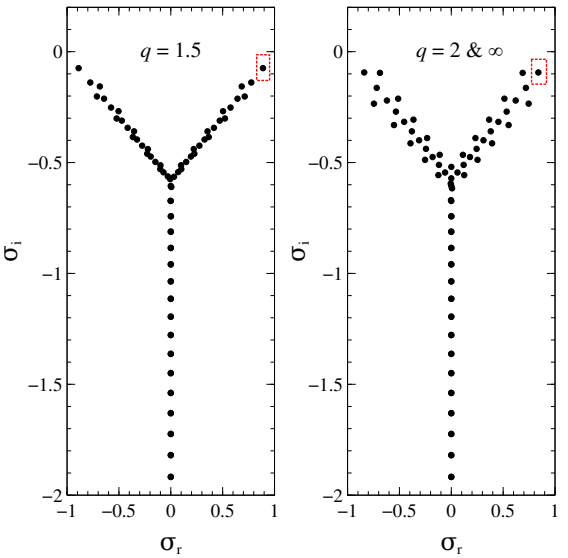


Fig. 2. Variation of σ_i with σ_r for $Re = 2000$ and $k_y = k_z = 1$ for the Keplerian flow ($q = 1.5$), constant angular momentum flow ($q = 2$) and plane Couette flow ($q \to \infty$). The latter two eigenspectra are identical. The dotted box represents the least stable mode for the respective cases.

We substitute these in equations (1) and (2) and collect the coefficients of the term $e^{i(\mathbf{k}\cdot\mathbf{r}-\sigma_r t)}$, to capture least nonlinear effect following, e.g.,,[12,20,22] from both sides and obtain

$$\frac{\partial Q_1}{\partial t} - i\sigma_r Q_1 + i\mathcal{L}Q_1 = NL_1, \tag{16}$$

where $NL_1 = \begin{pmatrix} \left(\mathcal{D}^2 - k^2\right)^{-1} NL_1^u \\ NL_1^\zeta \end{pmatrix}$ and $Q_1 = \begin{pmatrix} \bar{u}_1(x,t) \\ \bar{\zeta}_1(x,t) \end{pmatrix}$. We assume the solution for Q_1 to be

$$Q_1 = \sum_{m=1}^{\infty} A_{t,m} Q_{x,m} - \frac{1}{\mathcal{D}_t + i\mathcal{L}}\Gamma, \tag{17}$$

where m stands for various eigenmodes.

However, to the first approximation, our interest is in the least stable mode. Similar descriptions in two dimensions[20] without Γ and three-dimensional Keplerian disks[22] without Γ are already there in the in the literature and we mostly follow their formalism. We, therefore, omit the summation, and subscript m in the equation (17), and obtain

$$Q_1 = A_t Q_x - \frac{1}{\mathcal{D}_t + i\mathcal{L}}\Gamma. \tag{18}$$

We then substitute the equation (18) in the equation (16) and using bi-orthonormality between Q_x and its conjugate function $\tilde{Q}_x$, we obtain

$$\frac{dA_t}{dt} - \sigma_i A_t + \mathcal{N} = p|A_t|^2 A_t, \tag{19}$$

where

$$\mathcal{N} = \int_{-1}^{1} dx \tilde{Q}_x^{\dagger} \Gamma' \tag{20}$$

and

$$p = \int_{-1}^{1} dx \tilde{Q}_x^{\dagger} \mathcal{S}, \tag{21}$$

where

$$\begin{aligned} \Gamma' &= -\Gamma + i\sigma_r \Big(\frac{1}{\mathcal{D}_t + i\mathcal{L}} \Big) \Gamma \\ &= -\Gamma + i\sigma_r (t - i\mathcal{L}t^2)(1 + \mathcal{L}^2 t^2)^{-1} \Gamma, \end{aligned} \tag{22}$$

$\mathcal{S}$ is the spatial contribution from the nonlinear terms. For the greater details, see Ghosh and Mukhopadhyay.[12]

Throughout the paper, Γ from the equation (12) has been decomposed as $\Gamma \rightarrow \Gamma \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ by adjusting Γ_1 and Γ_2, as they are only the free parameters.

The evolution of $\mathcal{N}$ has been shown in Fig. 3 for the parameters mentioned in the figure. It shows that $\mathcal{N}$ becomes saturated beyond a certain time, and the saturation depends on Re. As Re increases, the saturation increases. Fig. 4 shows the variation of the least stable (the one in the dotted box in Fig. 2) as a function of Re in the case of the Keplerian flow. It tells us that the least stable mode approaches zero as Re is increased.

3. Results

Equation (19), which is a nonlinear equation, tells us about the evolution of the amplitude of the perturbation. To explore the evolution of linear and nonlinear perturbations, we use the equation (19) accordingly. For the evolution of the linear perturbation, we get rid of the nonlinear term in equation (19).

3.1. *Evolution of $|A|_t$ in the linear regime*

In the linear regime, the equation (19) becomes

$$\frac{dA_t}{dt} = \sigma_i A_t - \mathcal{N}. \tag{23}$$

This equation tells that at large t, $|A_t|$ becomes $|\mathcal{N}|/|\sigma_i|$.[12]

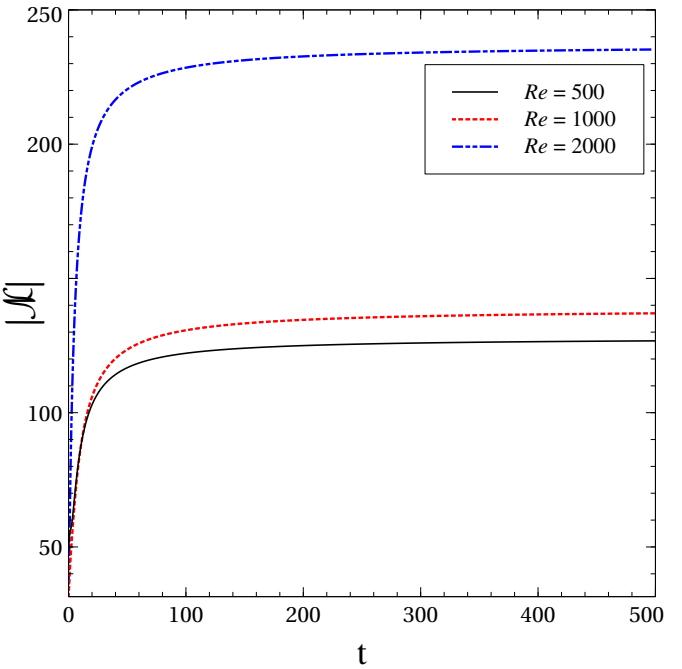


Fig. 3. Variation of $\mathcal{N}$ as a function of t for $Re = 500,\ 1000$ and 2000, for $\Gamma = 10^4$ and $k_y = k_z = 1$, corresponding to the respective least stable modes.

Fig. 5 shows the variation of $|A_t|$ as a function of t for various values of Re and Γ. Fig. 5 also suggests the scaling relation between saturated $|A_t|$ and Γ to be

$$|A_t| \propto \Gamma \tag{24}$$

for a fixed Re. Apart from that, we notice that for a fixed Γ, $|A|_t$ becomes saturated at $|\mathcal{N}|/|\sigma_i|$ beyond a certain time, and the saturation increases as Re increases. This saturation is independent of the initial condition.[12] This is because at a fixed Γ, $|\mathcal{N}|$ becomes saturated (see Fig. 3) at a fixed value, and the saturated value increases as Re increases, as well as $|\sigma_i|$ increases as Re increases (see Fig. 4). Now, in the accretion disk, Re is huge[19] ($\gtrsim 10^{14}$). The smaller Γ, therefore, can bring nonlinearity in the accretion disk, as for huge Re, $|\sigma_i|$ is infinitesimally small. Hence, $|\mathcal{N}|/|\sigma_i|$ becomes enormous. This huge saturation of $|A_t|$ could bring nonlinearity and hence plausibly turbulence in the underlying Keplerian flow.

3.2. *Evolution of $|A_t|$ in the nonlinear regime*

In the nonlinear regime, the evolution of $|A_t|$ is described by the equation (19). Fig. 6 describes the variation of $|A_t|$ as a function of t for nonlinear perturbation in the Keplerian flow. In the figure, $RE(A_t)$ and $IM(A_t)$ indicate the real and

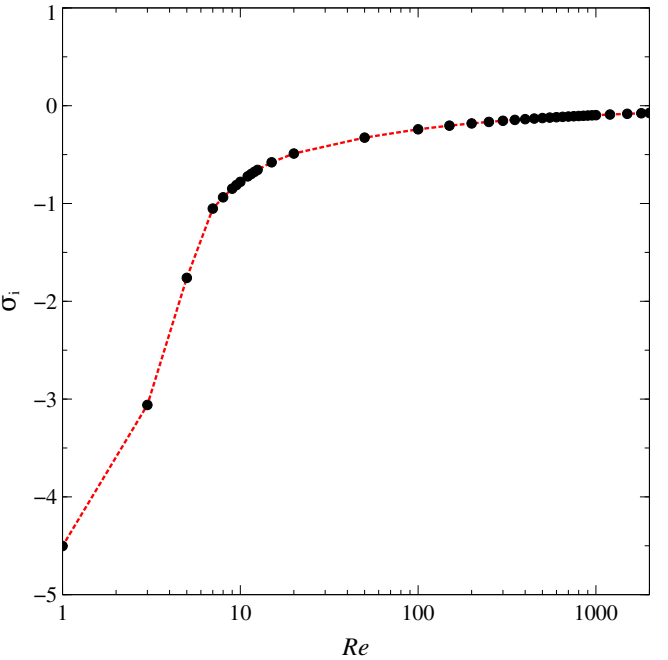


Fig. 4. Variation of the least stable σ_i as a function of Re for $k_y = k_z = 1$ in the Keplerian flow.

imaginary parts of initial A_t, respectively. It tells us that for $\Gamma = 0$, depending on the initial amplitude, $|A_t|$ may diverge (green line) or $|A_t|$ may decay (continuous black line) to zero from the initial amplitude. When there is no noise, i.e., $\Gamma = 0$, the diverging solution of $|A_t|$ occurs when the initial amplitude is beyond a critical amplitude.[12] It is apparent that the onset of the nonlinearity depends on the initial amplitude of perturbation in the absence of the force, but it does not depend on the same in the presence of force.[12] The divergence of $|A_t|$ and hence the onset of nonlinearity and plausible turbulence depends only on the strength of the force, as shown by dashed (magenta) line, compared to dot-dashed (blue) and dotted (red) lines, in Fig. 6.

4. Conclusion

If there is no extra force involved in the system, then the equation (19) becomes the usual Landau equation, which is

$$\frac{dA_t}{dt} = \sigma_i A_t + p|A_t|^2 A_t, \tag{25}$$

which can be further recast to

$$\frac{d|A|^2}{dt} = k_1|A|^2 + k_2|A|^4, \tag{26}$$

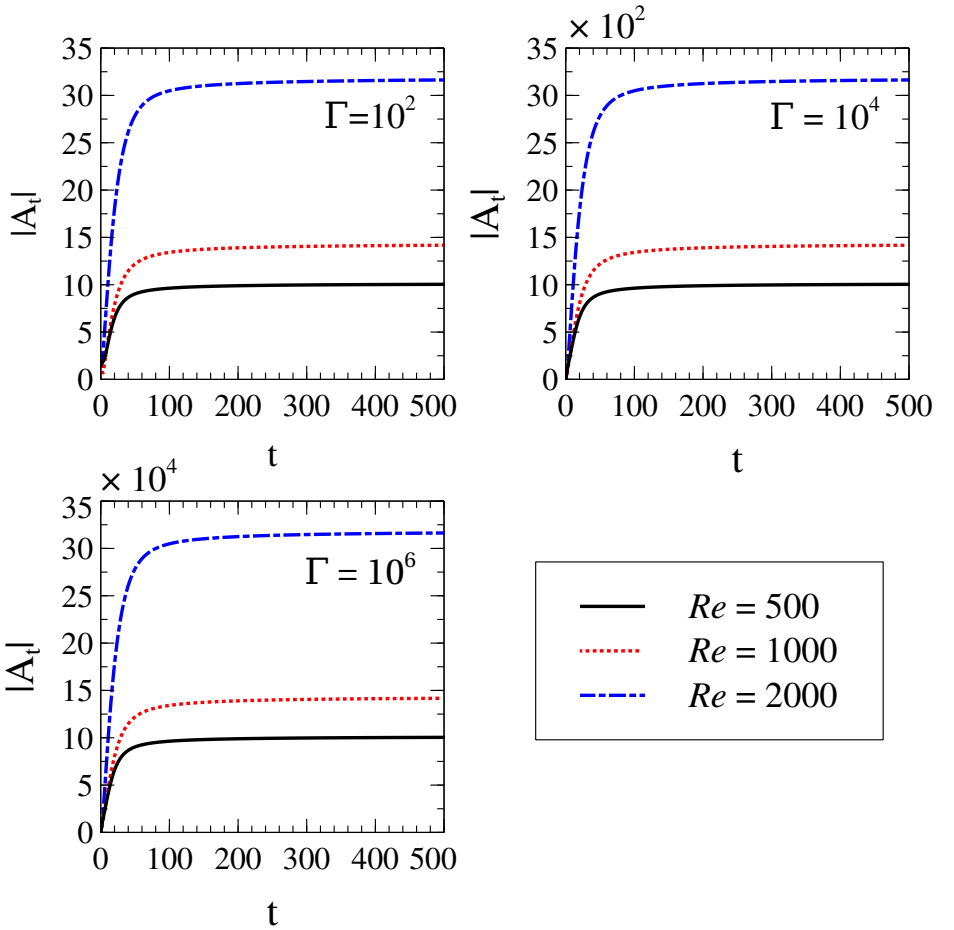


Fig. 5. Variation of $|A_t|$ as a function of t for three sets of Re and Γ with $k_y = k_z = 1$ for linear analysis in the Keplerian flow ($q = 1.5$).

where A is the amplitude of the nonlinear perturbations for the corresponding system, k_1 is $2\sigma_i$ and k_2 is the real part of $2p$, i.e., $2p_r$. Its solution is

$$|A|^2 = \frac{A_0^2}{-\frac{k_2}{k_1}A_0^2 + \left(1 + \frac{k_2}{k_1}A_0^2\right)e^{-k_1 t}}, \tag{27}$$

where A_0 is the initial amplitude. Depending on the sign (positive/negative) of k_1 and k_2, there are four[21,23] different possible evolutions of $|A|$. These given below.

- In the present context of shear flows, k_1 (i.e., σ_i) is negative, but k_2 is positive. Therefore, there will be a threshold for initial amplitude A_i, determining the growth of perturbation. If the initial amplitude $A_0 < A_i$,

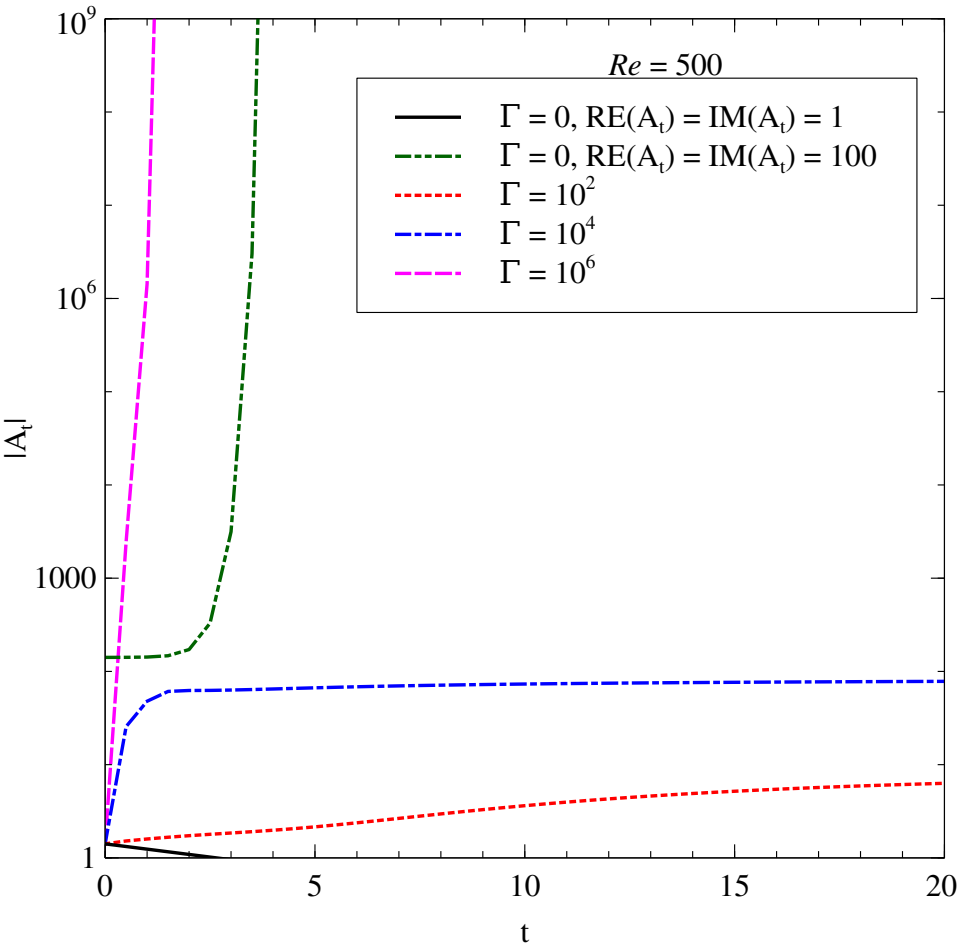


Fig. 6. Variation of $|A_t|$ as a function of t with $k_y = k_z = 1$ for nonlinear analysis in the Keplerian flow for $Re = 500$ and four different Γ. For $\Gamma \neq 0$, initial condition is $\mathrm{RE}(A_t) = \mathrm{IM}(A_t) = 1$.

then

$$|A|^2 \sim \frac{A_i^2 A_0^2 e^{k_1 t}}{A_i^2 - A_0^2} \tag{28}$$

at a large t. Therefore, $|A|^2 \to 0$ for $A_0 < A_i$ at $t \to \infty$. However, if $A_0 > A_i$, then $|A|^2 \to \infty$ at $t \to \ln(1 - A_i^2/A_0^2)/k_1$.

- If both k_1 and k_2 would be positive, $|A|^2$ blows up after a finite time. Hence, there will be a fast transition to turbulence.
- On the other hand, if $k_1 > 0$ but $k_2 < 0$, then $|A|^2 \to k_1/|k_2|$ at $t \to \infty$. In this case, $|A|^2$ at a large t does not depend on A_0.
- Obviously, for k_1 and k_2 both negative, $|A|^2$ decays fast.

However, we have shown that the saturation in $|A_t|$ is $|\mathcal{N}|/|\sigma_i|$ in the linear regime. We have also argued that depending on Re and Γ, the system may already be in the nonlinear regime. The evolution of $|A_t|$ at the linear regime in our case, i.e., with extra force, is similar to that of $|A|$ from the equation (26), i.e. without force, for $k_1 > 0$ and $k_2 < 0$. In the Keplerian flow, k_1, i.e. σ_i, is negative, but k_2, i.e. p_r, could be positive. In the presence of extra force, the Landau equation modifies in such a way that the solution in the linear regime itself mimics the Landau equation without force (i.e., equation 26), however, with $k_1 > 0$ and $k_2 < 0$. Further, in the nonlinear regime, the amplitude A_t (i.e., with extra force included) diverges beyond a certain time, depending on Re and Γ. In the nonlinear regime, the Landau equation in the presence of extra force but negative k_1 (σ_i) is, therefore, mimicking the Landau equation without force but with positive k_1 and k_2. Essentially, the extra force effectively changes the sign of k_1 (i.e., σ_i) for the Landau equation without force. Speaking in another way, the very presence of extra force destabilizes the otherwise stable system. Thus, the presence of force plays an important role in developing nonlinearity and turbulence in the case of Rayleigh stable flows.

Acknowledgment

S.G. acknowledges DST India for INSPIRE fellowship. This work is partly supported by a fund of Department of Science and Technology (DST-SERB) with research Grant No. DSTO/PPH/BMP/1946 (EMR/2017/001226).

References

1. N. I. Shakura and R. A. Sunyaev, Black holes in binary systems. Observational appearance, *Astronomy & Astrophysics* **24**, 337 (1973).
2. D. Lynden-Bell and J. E. Pringle, The Evolution of Viscous Discs and the Origin of the Nebular Variables, *Monthly Notices of the Royal Astronomical Society* **168**, 603 (09 1974).
3. S. A. Balbus and J. F. Hawley, A powerful local shear instability in weakly magnetized disks. I - Linear analysis. II - Nonlinear evolution, *The Astrophysical Journal* **376**, 214 (July 1991).
4. E. Velikhov, Stability of an ideally conducting liquid flowing between rotating cylinders in a magnetic field, *Zhur. Eksptl'. i Teoret. Fiz.* **36** (05 1959).
5. S. Chandrasekhar, The Stability of Non-Dissipative Couette Flow in Hydromagnetics, *Proceedings of the National Academy of Science* **46**, 253 (February 1960).
6. X.-N. Bai, Wind-driven Accretion in Protoplanetary Disks. II. Radial Dependence and Global Picture, *The Astrophysical Journal* **772**, p. 96 (Aug 2013).
7. M. E. Pessah and D. Psaltis, The Stability of Magnetized Rotating Plasmas with Superthermal Fields, *The Astrophysical Journal* **628**, 879 (August 2005).
8. U. Das, M. C. Begelman and G. Lesur, Instability in strongly magnetized accretion discs: A global perspective, *Monthly Notices of the Royal Astronomical Society* **473**, 2791 (2018).
9. S. K. Nath and B. Mukhopadhyay, Origin of nonlinearity and plausible turbulence by hydromagnetic transient growth in accretion disks: Faster growth rate than magnetorotational instability, *Physical Review E* **92**, p. 023005 (August 2015).

10. T. Singh Bhatia and B. Mukhopadhyay, Exploring non-normality in magnetohydrodynamic rotating shear flows: Application to astrophysical accretion disks, *Physical Review Fluids* **1**, p. 063101 (October 2016).
11. S. K. Nath and B. Mukhopadhyay, A Pure Hydrodynamic Instability in Shear Flows and Its Application to Astrophysical Accretion Disks, *The Astrophysical Journal* **830**, p. 86 (October 2016).
12. S. Ghosh and B. Mukhopadhyay, Hydrodynamical instability with noise in the Keplerian accretion discs: modified Landau equation, *Monthly Notices of the Royal Astronomical Society* **496**, 4191 (August 2020).
13. B. Mukhopadhyay, N. Afshordi and R. Narayan, Bypass to Turbulence in Hydrodynamic Accretion Disks: An Eigenvalue Approach, *The Astrophysical Journal* **629**, 383 (August 2005).
14. B. Mukhopadhyay, R. Mathew and S. Raha, Growing pseudo-eigenmodes and positive logarithmic norms in rotating shear flows, *New Journal of Physics* **13**, p. 023029 (February 2011).
15. S. Ghosh and B. Mukhopadhyay, Forced linear shear flows with rotation: rotating couette-poiseuille flow, its stability and astrophysical implications (2021).
16. Yecko, P. A., Accretion disk instability revisited - transient dynamics of rotating shear flow, *Astronomy & Astrophysics* **425**, 385 (2004).
17. N. Afshordi, B. Mukhopadhyay and R. Narayan, Bypass to Turbulence in Hydrodynamic Accretion: Lagrangian Analysis of Energy Growth, *The Astrophysical Journal* **629**, 373 (August 2005).
18. F. Rincon, G. I. Ogilvie and C. Cossu, On self-sustaining processes in Rayleigh-stable rotating plane Couette flows and subcritical transition to turbulence in accretion disks, *Astronomy & Astrophysics* **463**, 817 (Mar 2007).
19. B. Mukhopadhyay and A. K. Chattopadhyay, Stochastically driven instability in rotating shear flows, *Journal of Physics A Mathematical General* **46**, p. 035501 (January 2013).
20. T. Ellingsen, B. Gjevik and E. Palm, On the non-linear stability of plane Couette flow, *Journal of Fluid Mechanics* **40**, 97 (1970).
21. P. Schmid, D. Henningson and D. Jankowski, Stability and Transition in Shear Flows. Applied Mathematical Sciences, *Applied Mechanics Reviews* **55**, p. B57 (2002).
22. S. R. Rajesh, Weakly non-linear stability of a hydrodynamic accretion disc, *Monthly Notices of the Royal Astronomical Society* **414**, 691 (June 2011).
23. P. G. Drazin and W. H. Reid, *Hydrodynamic Stability* September 2004.

Properties of accretion disc models in the background quadrupole

Shokoufe Faraji

University of Bremen, Center of Applied Space Technology and Microgravity (ZARM), 28359 Germany

We consider a static and axially symmetric metric containing two quadrupole parameters. In the present contribution, we study the quadrupole moments constraints on the properties of the relativistic accretion disc models, also explore the relation of oscillatory frequencies of charged particles to the frequencies of the twin high-frequency quasi-periodic oscillations observed in some microquasars. We also compare the results with Schwarzschild and Kerr metrics.

Keywords: Accretion discs; Quadrupole moment; Black hole physics; QPOs

1. Introduction

In this paper, we are interested in studying how the presence of quadrupole moments in the metric affects the nature of astrophysical properties. This metric may link the observable effects to the system due to taking these parameters as the new dynamical degrees of freedom.

The first static and axially symmetric solution of Einstein equation containing arbitrary quadrupole moment were described in.[1] Later, Zipoy and Voorhees[2,3] found an equivalent transformation which can be treated analytically. Later on, with introducing a new parameter is known as q-metric.[4] In this regard, quadrupoles can be seen as perturbation parameters of the Schwarzschild spacetime from a dynamical point of view. This metric represents the exterior gravitational field of an isolated static axisymmetric mass distribution. Of course, there are different ways of including quadrupole in the metric, but we consider ones that can be treated analytically. One simple and straightforward extension of q-metric involves introducing an external field in addition to the dynamical metric. This class of theories is referred to as generalized q-metric. The astrophysical motivation for choosing such fields is the possibility to constitute a reasonable model for a more realistic situation occurring in vicinity of compact objects.

This work studies thin and magnetized thick accretion disc models,[5] and quasi-periodic oscillations (QPOs)[6] in this static and axisymmetric background.

2. q-metric with an external source

Most of the astrophysical studies considered the Kerr black hole. However, here, we choose to work on the generalized q-metric that considers the existence of a static and axially symmetric external distribution of matter in its vicinity up to

quadrupole like adding a magnetic surrounding.[7] In fact, due to their strong gravitational field, compact objects are not necessarily isolated or spherically symmetric as a response to their environments. This metric describes the outer of a deformed compact object up to quadrupole and reads as[8]

$$\begin{aligned} \mathrm{d}s^2 = &-\left(\frac{x-1}{x+1}\right)^{(1+\alpha)} e^{2\hat{\psi}} dt^2 + M^2(x^2-1)e^{-2\hat{\psi}} \\ &\times \left(\frac{x+1}{x-1}\right)^{(1+\alpha)} \left[\left(\frac{x^2-1}{x^2-y^2}\right)^{\alpha(2+\alpha)} e^{2\hat{\gamma}} \right. \\ &\left. \times \left(\frac{dx^2}{x^2-1} + \frac{dy^2}{1-y^2}\right) + (1-y^2)d\phi^2\right], \end{aligned} \tag{1}$$

where $t \in (-\infty, +\infty)$, $x \in (1, +\infty)$, $y \in [-1, 1]$, $\phi \in [0, 2\pi)$, and in principle $\alpha \in (-1, \infty)$. Metric function $\hat{\psi}$ plays the role of gravitational potential, and once we have it by an integration of its explicit form, $\hat{\gamma}$ is obtained. These are given by

$$\hat{\psi} = -\frac{\beta}{2}\left[-3x^2y^2 + x^2 + y^2 - 1\right], \tag{2}$$

$$\begin{aligned} \hat{\gamma} = &-2x\beta(1-y^2) \\ &+ \frac{\beta^2}{4}(x^2-1)(1-y^2)(-9x^2y^2 + x^2 + y^2 - 1). \end{aligned} \tag{3}$$

By its construction, this metric is valid locally.[9] This metric has two parameters aside from the mass; deformation parameter α and distortion parameter β, which are not independent of each other. These are chosen to be relatively small and connected to the q-metric and the surrounding external mass distribution, respectively. It can easily be checked that for vanishing β we recover the q-metric, and in the limits $\alpha = 0$ and $\beta \neq 0$ distorted Schwarzschild written in the prolate spheroidal coordinates,[9] and for $\alpha = 0$ and $\beta = 0$ we recover the Schwarzschild solution. The relation between the prolate spheroidal coordinates (t, x, y, ϕ), and the Schwarzschild coordinates (t, r, θ, ϕ) is

$$x = \frac{r}{M} - 1\,, \quad y = \cos\theta. \tag{4}$$

To the end, we explore this background by analyzing the thick and thin accretion disc models, and Quasi-periodic oscillations.

3. Accretion discs

In this analysis, we consider two well known Thin and Thick analytical models of accretion onto a compact object.

3.1. *Thin model*

The standard thin disc model has been used to explain a variety of observations where gas is cold and neutral. Therefore, the coupling between the magnetic field and

gas is negligible. In general, observations provide the luminosity and the maximum temperature of the disc that fit the data model.

In this model, the steady axisymmetric fluid configuration is assumed. In addition, as a result of the geometrically thin assumption, effectively, all physical quantities only depend on the vertical distance from the equatorial plane and the radial distance to the central object so that we can consider vertically integrated quantities. Therefore, the two-dimensional disc structure can be decoupled to one-dimensional configurations. One is responsible for the radial quasi-Keplerian flow another for the vertical hydrostatic structure.[10,11]

Three fundamental equations govern the radial structure of the thin disc model

$$(\rho u^{\mu})_{;\mu} = 0, \tag{5}$$

$$u_{\mu} T^{\mu\nu}{}_{;\nu} = 0, \tag{6}$$

$$h_{\mu\sigma}(T^{\sigma\nu})_{;\nu} = 0. \tag{7}$$

where u^{μ} is the four-velocity of the fluid and ρ is the rest mass density, $T^{\sigma\nu}$ is the stress-energy tensor containing non-vanishing shear part, and $h^{\mu\nu}$ is the projection tensor. These equations, together with the radial velocity, which is stated in terms of surface density and mass accretion rate, alpha viscosity prescription and heat flow, govern the radial structure of the disc.[10,11]

For the vertical profile in the comoving fluid frame, the force due to vertical pressure gradient is balanced with gravity, the centrifugal force, and vertically Euler force. The pressure is the sum of gas pressure from nuclei and the radiation pressure, which in practical is derived from the relativistic Euler equation.[12]

Finally, by solving this system of linear equations, one obtains a unique solution for considering positive temperature and pressure. Besides quadrupoles, three parameters describe a thin disc solution in this background, namely M, mass accretion rate $\dot{M}$, and viscosity parameter α. Consequently, the geometric configuration of an accretion disc located around this space-time depends explicitly on the value of the quadrupole parameter. An analysis shows the astrophysical quantities possess almost the same pattern. Figure 1 shows the temperature and flux profiles for different values of quadrupole parameters. As it shows, the intensity of these quantities is much higher closer to the central object and for negative values than Schwarzschild and positive values. This links to the location of ISCO, which is considered the inner edge of the thin accretion disc. For negative quadrupoles, ISCO is closer to the central object, pushing away for positive ones. Therefore, for negative values, the inner part of the disc is located closer to the central object, which causes more intense characteristics in all quantities. It means the behaviour of positive quadrupole moments is smoother than negative ones in general. We may compare this result with the Kerr solution if we consider the rotation parameter a kind of quadrupole counterpart in this background. In the Kerr background, the intensity of these quantities for co-rotating is also smaller than the counter-rotating

case. However, the main difference in the Kerr background is this pattern is valid everywhere, not only closer to the inner part of the disc. Therefore, in general, there are distinguishable differences in both cases regarding observational data.[13]

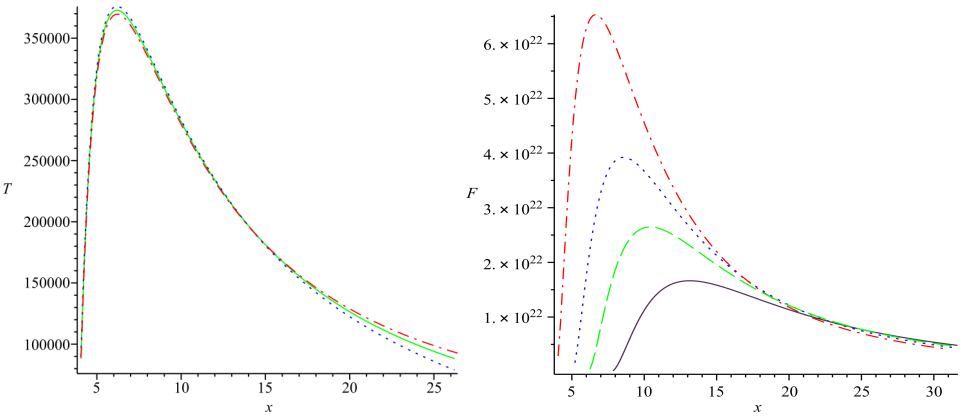


Fig. 1. The left plot is temperaturee T profile for $\alpha = -0.2$; blue line shows $\beta = -0.00001$, green one $\beta = 0$ (Schwarzschild), and red one $\beta = -0.00001$. The right plot is flux profile F for vanishing β (q-metric) and different values of α; from top $\alpha = -0.2$, $\alpha = 0$ (Schwarzschild), $\alpha = 0.2$, $\alpha = 0.5$.

3.2. *Thick model*

The thick disc model describes a general method of constructing perfect fluid equilibria in an axially symmetric and stationary background, which is the simplest analytical model of discs with no accretion flow and is radiatively inefficient. In this work, we considered its magnetic version. The evolution of an ideal magnetised fluid is described by the baryon conservation, energy-momentum conservation, and induction equation.[14,15] The final well known equation in this model reads as[16]

$$W - W_{\rm in} + \frac{\kappa}{\kappa - 1}\frac{p}{w} + \frac{\eta}{\eta - 1}\frac{p_m}{w} = \int_{\ell_{in}}^{\ell} \frac{\Omega \mathrm{d}\ell}{1 - \Omega\ell}, \tag{8}$$

where $W = \ln |u_t|$, magnetic pressure and gas pressure are p_m and p, η and κ are parameters, and w is enthalpy. Therefore, one can construct this model by specifying Ω or ℓ to fix the geometry of the equipotential surfaces. In the constant case, the disc surface is fully determined by the choice of the potential in the inner part of the disc, $W_{\rm in}$, independent of the magnetic field.[17] Thus, the value of angular momentum ,ℓ_0, determines the total potential.[16]

In this work we also considered two more angular momentum distributions; the power-law,[18] and the trigonometric function.[19] In the power-law distribution it is

assumed that

$$\Omega(\ell) = c\ell^n. \tag{9}$$

where by using this equation, the right hand side of the equation (8) can be calculated easily.[18] In the trigonometric function distribution, it is assumed that[19]

$$\ell(x,y) = \begin{cases} \ell_0 \left(\frac{\ell_K(x)}{\ell_0}\right)^{\sigma} (1-y^2)^{\delta}, & x \geq x_{\rm ms}, \\ \ell_0(\zeta)^{-\sigma}(1-y^2)^{\delta}, & x < x_{\rm ms}, \end{cases} \tag{10}$$

where $\ell_0 = \zeta \ell_K(x_{\rm ms})$, and ℓ_K is the Keplerian angular momentum in the equatorial plane, η, σ, and δ are parameters, and $x_{\rm ms}$ is the place of marginally stable orbit. In this model, one only needs to solve function F to obtain the solution[19]

$$\frac{\partial_x p}{\partial_y p} = \frac{\partial_x g^{tt} + \ell^2 \partial_x g^{\phi\phi}}{\partial_y g^{tt} + \ell^2 \partial_y g^{\phi\phi}} := -F(x,y). \tag{11}$$

Analysis of results shows that changing parameter β_c has a strong effect on the location and amplitude of the rest-mass density and spreading the matter through the disc for any chosen value of quadrupoles.

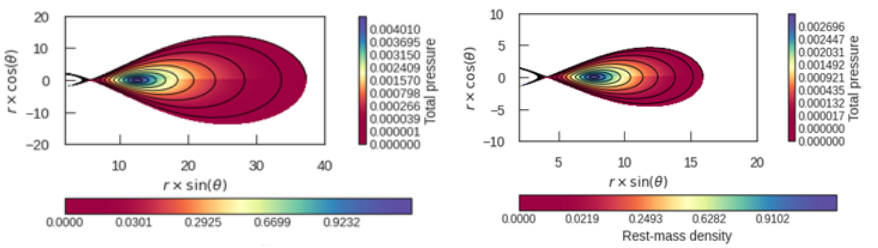


Fig. 2. The left plot is for $(\alpha, \beta) = (0.4, 0.000001)$, the right one is for $(\alpha, \beta) = (-0.1, 0.000001)$ with constant angular momentum.

In fact, a higher magnetic pressure causes the matter to concentrate more in the inner part of the disc. The rest-mass density maximum and its location also depend on the quadrupoles values. Although the effect of quadrupoles correlates with the other parameters, we can conclude that quadrupoles are more responsible for shifting the position of the disc and have a positive contribution to its radial extension. However, this effect for β is not easy to see, as the values of quadrupole parameters that we chose in our models are minimal.[20,21] Moreover, there is a possibility of having two tori in a row for any value of β but negative values for the parameter $\alpha \in (-0.5, -0.553]$. In Figure 3 this situation is plotted with constant angular momentum. However, with two other angular momentum distributions discussed in the paper it is also possible to have two tori.[20]

Furthermore, in the trigonometric distribution, the σ and δ parameters influence each other. However, one can say that σ is more responsible for the location and

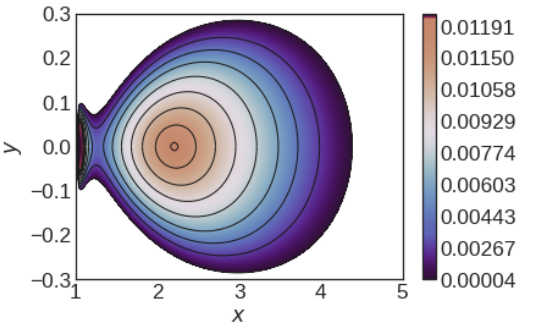


Fig. 3. Possibility to have two tori in a row for $\alpha \in (-0.5, -0.553]$ regardless of the β values.

amplitude of the rest-mass density, and δ affects the distribution of equidensity surfaces. In the case of the power-law distribution, steeper angular momentum tends to shrink the disc and decrease the amplitude of the rest-mass density and pushes away the location of its maximum simultaneously.

4. Quasi-periodic oscillations

Quasi-periodic oscillations (QPOs) of X-ray power spectral density have been observed at low (Hz) and high (kHz) frequencies.[22,23] There have been many models to explain QPOs in the past years. Among them is the Relativistic Precession Model (RPM), which relates the twin-peak QPOs to the Keplerian and periastron precession frequency on an orbit located in the inner part of the accretion disc.[24] Regarding, the high frequencies QPOs (the two-picks in Fourier power density spectra) are considered as the resonances between oscillation modes. Within the ratio 3 : 2 epicyclic resonance, the resonant frequencies are identified with frequencies of radial and vertical epicyclic axisymmetric modes.[25–27]

In this work, we are interested in the dynamics that occur when a static, axisymmetric central compact object is embedded in a weak uniform magnetic field of strength B aligned with the central body's symmetry axis and described by the ϕ-component of the vector potential[28]

$$A_\phi = \frac{1}{2} B(x^2-1)e^{-2\hat{\psi}} \left(\frac{x+1}{x-1}\right)^{(1+\alpha)}. \tag{12}$$

In addition, the specific energy and angular momentum of the particle read as[29]

$$E = \left(\frac{x-1}{x+1}\right)^{(1+\alpha)} e^{2\hat{\psi}} \frac{dt}{ds}, \tag{13}$$

$$L = (x^2-1)e^{-2\hat{\psi}} \left(\frac{x+1}{x-1}\right)^{(1+\alpha)} \left(\frac{d\phi}{ds} + Q\right), \tag{14}$$

where $Q := \frac{qB}{2m}$ is magnetic parameter. Further, the effective potential is

$$V_{\rm Eff} = \left(\frac{x-1}{x+1}\right)^{(\alpha+1)} e^{2\hat{\psi}}\Bigg[\epsilon + e^{-2\hat{\psi}}\left(\frac{x+1}{x-1}\right)^{\alpha}(1-y^2) \times \left(\frac{Le^{2\hat{\psi}}}{(x+1)(1-y^2)}\left(\frac{x-1}{x+1}\right)^{\alpha} - Q(x+1)\right)^2\Bigg], \tag{15}$$

The second term corresponds to the central force potential and electromagnetic potential energy. In what follows, we investigate the stability of circular motion in this background in the presence of a homogeneous magnetic field. We followed the method described in[30,31] and extend the result to this background. The results are in good agreement with these papers regarding vanishing quadrupoles, also rotation in their works.

We need to describe the more general class of orbits slightly deviated from the circular orbits in the equatorial plane x^μ using the perturbation expansion, $x'^\mu = x^\mu + \xi^\mu$. By considering this relation into the geodesic equation with non-vanishing external force and making the integration, for the t and ϕ components, up to linear order in ξ^μ, and we obtain[30,31]

$$\omega_x^2 = \partial_x U^x - \gamma^x{}_\eta \gamma^\eta{}_x, \tag{16}$$

$$\omega_y^2 = \partial_y U^y. \tag{17}$$

These equations describe radial and vertical oscillations of the charged particle about circular orbits. The sign of frequencies determines the dynamic ω_x^2 and ω_y^2 so that if they have a positive sign, we have a stable orbit; otherwise, even a minimal perturbation can cause a substantial deviation from the unperturbed orbit.

In Schwarzschild solution, the vertical epicyclic frequency is a monotonically decreasing function of x, and we have $\omega_x^2 < \omega_y^2 = \Omega^2$. On the contrary to the Schwarzschild case, the interesting situation in this background is to have various orderings among frequencies and thus different possibilities to reproduce the ratio of $3:2$, also other ratios via different combinations of parameters[29] which is not the case in either Schwarzschild or in q-metric.

Among the various models of the resonance of accretion disc oscillations, we consider the group of QPO models (WD, TD, RP, Ep, Kp, RP1, RP2) considered in[32] and examined them in this background and compared with observational data of three microquasars XTE 1550-564, GRS 1915+105, and GRO 1655-40.[29,33] However, the physical details of these models are different and depend explicitly on the time evolution of the desired system. Indeed, the case of relatively slowly rotating XTE 1550-564 source is more compatible with our static set-up. However, different sets of parameters can fit the data even for the two other fast-rotating sources. An analysis shows for chosen parameters Q, α, and β, the best fit almost corresponds to RP2, RP, and RP1 models in low spin cases. Besides, results reveal that even the regular orbits for some combinations of parameters of metric turn to behave chaotically with a higher magnetic field.

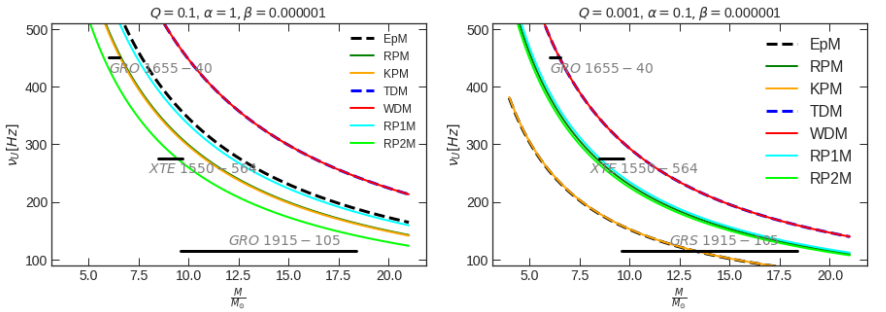


Fig. 4. The upper oscillation frequencies compared to the mass limits obtained from observations of three microquasars for some chosen parameters.

5. Conclusion

In this paper, we mentioned astrophysical features of two accretion disc models in the background of a compact object containing quadrupole moments. We have shown that the presence of quadrupoles changes the accretion disc's geometric properties in both models drastically. In particular, in the Thin accretion disc model, the spectral features would be significantly different and distinguishable in the observation from a black hole. The structure and the shape of the magnetized Thick accretion disc model are also influenced strongly by these parameters.

In addition, the dependency of these two parameters reflects into epicyclic frequencies of particles that cost substantial deviation from the correspondence quantities in the Schwarzschild case. In fact, the resonant phenomena of the radial and vertical oscillations at their frequency ratio 3 : 2 depending on chosen parameters, also in this background can be well related to the frequencies of the twin HF QPOs 3 : 2 observed in the microquasars. Although the fitting depends on the combination of parameters, we still can expect to have a better fitting for choosing a larger quadrupole moment in the combination.

Directions for the future could be adding the strong magnetic field which also influences the metric itself, considering rotation, studying the self-gravity of the Thin disc model, and studying the stability of this solution to extend this work. Further, this theoretical model can be served as the initial data for numerical simulations in the astrophysical setting.

References

1. H. Weyl, Zur gravitationstheorie, *Annalen der Physik* **359**, 117 (1917).
2. D. M. Zipoy, Topology of some spheroidal metrics, *Journal of Mathematical Physics* **7**, 1137 (1966).
3. B. H. Voorhees, Static axially symmetric gravitational fields, *Phys. Rev. D* **2**, 2119 (Nov 1970).

4. H. Quevedo, Mass Quadrupole as a Source of Naked Singularities, *International Journal of Modern Physics D* **20**, 1779 (Jan 2011).
5. M. A. Abramowicz and P. C. Fragile, Foundations of Black Hole Accretion Disk Theory, *Living Reviews in Relativity* **16**, p. 1 (January 2013).
6. A. R. Ingram and S. E. Motta, A review of quasi-periodic oscillations from black hole X-ray binaries: Observation and theory, *New Astronomy Reviews* **85**, p. 101524 (September 2019).
7. F. J. Ernst, Black holes in a magnetic universe, *Journal of Mathematical Physics* **17**, 54 (January 1976).
8. S. Faraji, A generalization of q-metric, *arXiv e-prints* , p. arXiv:2010.15723 (October 2020).
9. R. Geroch and J. B. Hartle, Distorted black holes., *Journal of Mathematical Physics* **23**, 680 (1982).
10. I. D. Novikov and K. S. Thorne, Astrophysics of black holes., 343 (1973).
11. D. N. Page and K. S. Thorne, Disk-Accretion onto a Black Hole. Time-Averaged Structure of Accretion Disk, *ApJ* **191**, 499 (July 1974).
12. M. A. Abramowicz, A. Lanza and M. J. Percival, Accretion Disks around Kerr Black Holes: Vertical Equilibrium Revisited, *ApJ* **479**, 179 (April 1997).
13. S. Faraji, Structure of the thin disk in the background of a distorted naked singularity versus a distorted static black hole, *arXiv e-prints* , p. arXiv:2011.06634 (November 2020).
14. A. M. Anile, *Relativistic fluids and magneto-fluids : with applications in astrophysics and plasma physics* (Oxford Univ. Press, 1989).
15. W. G. Dixon, *Special relativity: the foundation of macroscopic physics.* (Oxford Univ. Press, 1978).
16. S. S. Komissarov, Magnetized tori around kerr black holes: analytic solutions with a toroidal magnetic field, *Monthly Notices of the Royal Astronomical Society* **368**, p. 993–1000 (Apr 2006).
17. R. Okada, J. Fukue and R. Matsumoto, A model of astrophysical tori with magnetic fields, *PASJ* **41**, 133 (January 1989).
18. M. Wielgus, P. C. Fragile, Z. Wang and J. Wilson, Local stability of strongly magnetized black hole tori, *MNRAS* **447**, 3593 (March 2015).
19. L. Qian, M. A. Abramowicz, P. C. Fragile, J. Horák, M. Machida and O. Straub, The Polish doughnuts revisited. I. The angular momentum distribution and equipressure surfaces, *Astron Astroph* **498**, 471 (May 2009).
20. S. Faraji and A. Trova, Magnetised tori in the background of a deformed compact object, *Astron Astroph* **654**, p. A100 (October 2021).
21. S. Faraji and A. Trova, Effect of an external mass distribution on the magnetized accretion disk, *Phys. Rev.* **104**, p. 083006 (October 2021).
22. M. van der Klis, Millisecond Oscillations in X-ray Binaries, *Annu. Rev. Astron. Astrophys.* **38**, 717 (January 2000).
23. J. E. McClintock and R. A. Remillard, *Black hole binaries*, in *Compact stellar X-ray sources*, (Oxford, 2006), pp. 157–213.
24. L. Stella and M. Vietri, Lense-Thirring Precession and Quasi-periodic Oscillations in Low-Mass X-Ray Binaries, *ApJ* **492**, L59 (January 1998).
25. M. A. Abramowicz and W. Kluźniak, A precise determination of black hole spin in GRO J1655-40, *Astron Astroph* **374**, L19 (August 2001).
26. W. Kluzniak and M. A. Abramowicz, Strong-Field Gravity and Orbital Resonance in Black Holes and Neutron Stars — kHz Quasi-Periodic Oscillations (QPO), *Acta Physica Polonica B* **32**, p. 3605 (November 2001).

27. M. A. Abramowicz, D. Barret, M. Bursa, J. Horák, W. Kluźniak, P. Rebusco and G. Török, The correlations and anticorrelations in QPO data, *Astronomische Nachrichten* **326**, 864 (November 2005).
28. R. M. Wald, Black hole in a uniform magnetic field, *Phys. Rev. D* **10**, 1680 (Sep 1974).
29. S. Faraji and A. Trova, Dynamics of charged particles and QPO of related disk models in the background of a distorted, deformed compact object- I. embedded in a uniform magnetic field, *arXiv e-prints* , p. arXiv:2103.03229 (March 2021).
30. A. N. Aliev and D. V. Galtsov, Radiation from relativistic particles in nongeodesic motion in a strong gravitational field., *General Relativity and Gravitation* **13**, 899 (October 1981).
31. A. N. Aliev, D. V. Galtsov and V. I. Petukhov, Negative Absorption Near a Magnetized Black-Hole - Black-Hole Masers, *Astrophys. Space Sci.* **124**, 137 (1986).
32. G. Török, A. Kotrlová, E. Šrámková and Z. Stuchlík, Confronting the models of 3:2 quasiperiodic oscillations with the rapid spin of the microquasar GRS 1915+105, *Astron Astroph* **531**, p. A59 (July 2011).
33. S. Faraji and A. Trova, Quasi-periodic oscillatory motion of particles orbiting a distorted deformed compact object, *arXiv e-prints* , p. arXiv:2102.11871 (February 2021).

Magnetized tori around a uniformly accelerating black hole

Shokoufe Faraji, Audrey Trova

University of Bremen, Center of Applied Space Technology and Microgravity (ZARM), 28359 Germany

We generalize the relativistic accretion thick disc model to the background of a spinning charged accelerating black hole described by the C-metric to study the effects of this background on the disc model. We show the properties of this accretion disc model and its dependence on the initial parameters. This background can be distinguishable from the Kerr space-time by analyzing the observing features of accretion discs.

Keywords: C-metric; Magnetic field; Accretion disc; Black hole physics

1. Introduction

In this paper we investigate the spinning charged C-metric, with the aim of studying the properties of the well known magnetised Thick accretion disc model and the morphology of the equipotential surfaces.

The family of C-metric has accelerating nature and is considered as describing an accelerating black hole.[1] The spinning charged C-metric in Boyer-Lindquist-type coordinates[2] reads as

$$ds^2 = \frac{1}{\Omega^2}\left(-\frac{f}{\Sigma}\left[dt - a\sin^2\theta\frac{d\varphi}{K}\right]^2 + \frac{\Sigma}{f}dr^2 + \Sigma r^2\frac{d\theta^2}{g} + \frac{g\sin^2\theta}{\Sigma r^2}\left[adt - (r^2+a^2)\frac{d\varphi}{K}\right]^2\right), \tag{1}$$

where

$$\Omega = 1 + \alpha r\cos\theta, \tag{2}$$

$$f(r) = \left(1-\alpha^2 r^2\right)\left(1 - \frac{2m}{r} + \frac{e^2+a^2}{r^2}\right), \tag{3}$$

$$g(\theta) = (e^2+a^2)\alpha^2\cos^2\theta + 2m\alpha\cos\theta + 1, \tag{4}$$

$$\Sigma(r,\theta) = \frac{a^2}{r^2}\cos^2\theta + 1, \tag{5}$$

$$\xi = \alpha^2(e^2+a^2) + 1, \tag{6}$$

$$K = \xi + 2m\alpha, \tag{7}$$

where $t \in (-\infty, +\infty)$, $\theta \in (0, \pi)$, $r \in (0, +\infty)$. The metric has four independent parameters: the mass m, the electric charge e, the rotation a, and the so-called acceleration parameter α. In this metric $r = 0$ is the curvature singularity, and there

is a conical singularity on the θ axis. The parameter K affects the distribution of conical defects in the space-time and allows φ to be 2π-periodic (see for example[3–5]).

Almost all studies with this metric are revolved around the coordinate ranges, which are dictated by the metric functions and their root configurations. The main constrains are

$$r_b = -\frac{1}{\alpha \cos\theta}, \tag{8}$$

$$e^2 + a^2 \leq m^2, \tag{9}$$

and

$$2m\alpha \leq \begin{cases} 2\sqrt{\xi - 1} & \xi > 2, \\ \xi & 0 < \xi \leq 2. \end{cases} \tag{10}$$

2. Properties of the disc model

The Thick accretion disc is an analytical model that provides a general method to build equilibrium configurations of the perfect fluid matter orbiting around a stationary and axially symmetric black hole.[6,7] In this procedure, we considered the approach of Komissarov[8] to attach a dynamically toroidal magnetic field to the model.

An analysis shows[9] that by increasing the acceleration parameter α, almost the possibility of having solutions decreases dramatically. On the contrary, the charge parameter e positively contributes to having solutions; mainly, its effect manifests more for a relatively large α. Furthermore, the possibility of having solutions for larger α depends on having large values for e, so the effect of higher α on the disc could be neutralized partially with the higher charge.

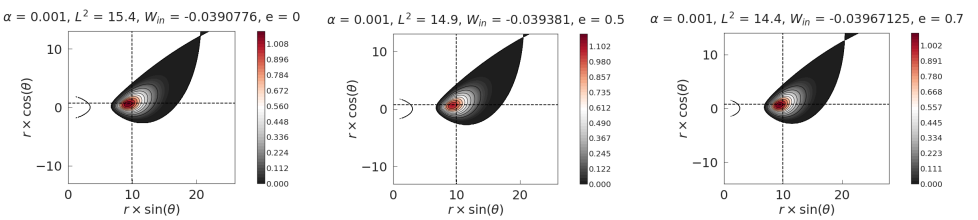


Fig. 1. Contour map of the rest-mass density of magnetised disc. The dashed lines point the center of the disc at $r_c = 10$. The discs are all highly magnetised disc.

The effect of rotation parameter a on having solutions is not strong compared to α but more substantial than the charge parameter e. In fact, parameter a, like α, has a negative effect on having solutions. In general, as e increases, we expect the matter is concentrated closer to the inner edge of the disc. On the contrary, the higher values of a spread the matter more through the disc. Moreover, Increasing a

decreases the size of the disc and change the distribution of matter. In addition, it shifts the disc farther from the compact object, contrary to an increase in e, which shift the disc closer to the central object.

For a fixed value of acceleration parameter α and vanishing rotation, the magnetization parameter does not influence the geometry of the disc; however, it changes the distribution of matter inside the disc and shifts the location of the rest-mass density maximum. In addition, we have a larger oriented disc for larger values of e.

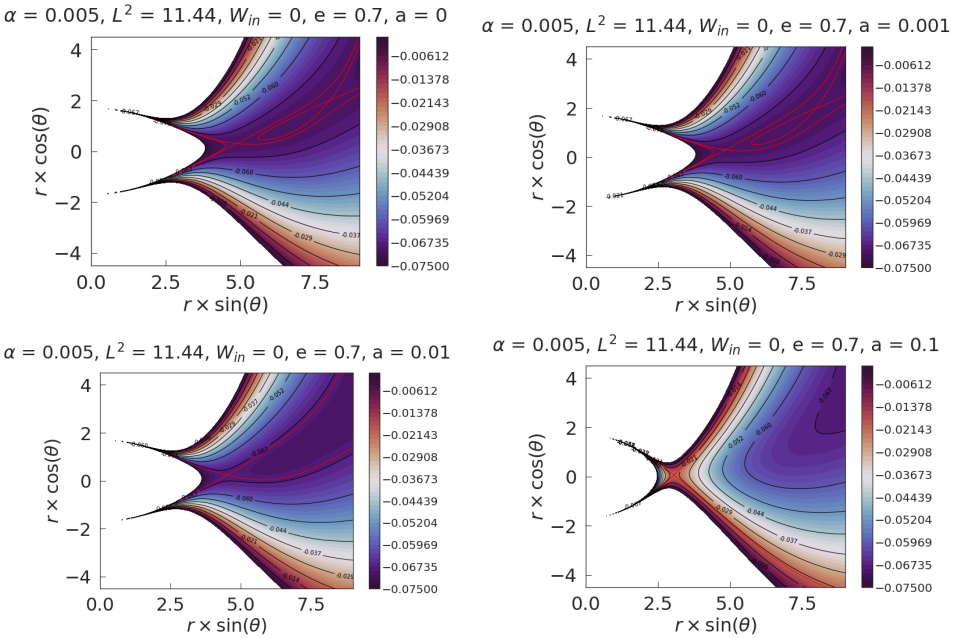


Fig. 2. Contour map of the equipotential surfaces. no solution for $\alpha = 0.005$ unless for high e and very low a. Show no solution for higher α and rotation.

3. Conclusion

In this work, we discussed the influence of the magnetisation parameter, charge e, rotation a, and accelerating parameter α in the spinning charged C-metric on the structure of the magnetised thick disc model.

In general, we can have the Thick disc model for relatively small values of α, while by increasing α, the disc structure becomes smaller and gradually vanishes. The same is true for a. On the contrary to these two parameters, an increase in e increases the disc size and possibility of having a solution. However, we should mention that the strength of the parameters are not the same, as α has the strongest and e has the weaker effect on the disc structure, in comparison.

A further step of this work can be a study on the oscillation of the disc in this setup which is in progress. In addition, this is interesting to examine the Thick disc model with different angular momentum profiles in this space-time.

References

1. W. Kinnersley and M. Walker, Uniformly accelerating charged mass in general relativity, *Phys. Rev. D* **2**, 1359 (Oct 1970).
2. K. Hong and E. Teo, A new form of the rotating C-metric, *Classical and Quantum Gravity* **22**, 109 (January 2005).
3. F. J. Ernst, Generalized c-metric, *Journal of Mathematical Physics* **19**, 1986 (1978).
4. J. Bičák and V. Pravda, Spinning C metric: Radiative spacetime with accelerating, rotating black holes, *Phys. Rev.* **60**, p. 044004 (August 1999).
5. J. B. Griffiths, P. Krtous and J. Podolský, Interpreting the C-metric, *Classical and Quantum Gravity* **23**, 6745 (December 2006).
6. A. M. Anile, *Relativistic fluids and magneto-fluids : with applications in astrophysics and plasma physics* (Oxford Univ. Press, 1989).
7. W. G. Dixon, *Special relativity: the foundation of macroscopic physics.* (Oxford Univ. Press, 1978).
8. S. S. Komissarov, Magnetized tori around kerr black holes: analytic solutions with a toroidal magnetic field, *Monthly Notices of the Royal Astronomical Society* **368**, p. 993–1000 (Apr 2006).
9. S. Faraji and A. Trova, Magnetised relativistic accretion disc around a spinning charged accelerating black hole, *arXiv e-prints*, p. arXiv:2108.07070 (August 2021).

Multifrequency behaviour of high mass X-ray binaries (Time lag between optical and X-ray outbursts)

Franco Giovannelli
INAF - Istituto di Astrofisica e Planetologia Spaziali, Roma, Italy
Email: franco.giovannelli@iaps.inaf.it

In this work, without any claim to completeness, I will review the zoo of binary systems emitting X radiation, the multi-frequency behaviors of high-mass X-ray systems (HMXBs), with particular emphasis on the X-ray/Be system A 0535 + 26/HDE 245770, which for a favorable series of concomitant causes it is the most studied system. I will also discuss the time lag between events occurring in the high-energy and low-energy bands in galactic accreting systems.

Keywords: X-ray binaries; High mass X-ray binaries; individual: A 0535+26/HDE 245770; accretion, accretion disks

1. Introduction

Among X-ray emitters, X-ray binaries are cauldrons of fundamental physical processes which appear along practically the whole electromagnetic spectrum. The sub-class of X-ray transient sources show multifrequency behaviour which deserve particular attention in order to understand the causing physics. These binary systems consist of a compact star and an optical star, therefore there is a mutual influence between these two stars that drive the low energy (LE) (i.e. radio, IR, optical) and high energy (HE) (i.e. UV, X-ray, γ-ray) processes. The LE processes are produced mostly on the optical star and the HE processes mostly on the compact star, typically a neutron star. Thus it appears evident that through the study of LE processes it is possible to understand also the HE processes and vice versa. In this paper we will discuss this problem starting from the experimental evidence of a delay between LE and HE processes detected for the first time in the X-ray/Be system A0535+26/HDE245770 (e.g. Giovannelli & Sabau-Graziati;[40] Giovannelli, Bisnovatyi-Kogan & Klepnev[41] (here after GBK13); Giovannelli et al.[43]). This delay is common in cataclysmic variables (CVs) and other binary systems with either a neutron star or a black hole.

Giovannelli[30] published an extensive review about X-ray binary systems. Briefly we can summarize the main characteristics of these systems.

2. The zoo of X-ray binary systems

The trivial definition of X-ray binaries (XRBs) is that they are binary systems emitting X-rays. However it has been largely demonstrated that X-ray binary systems

emit energy in IR, Optical, UV, X-ray, Gamma-ray and sometimes they show also valuable radio emission. They can be divided in sub-classes, namely:

1. High Mass X-ray Binaries (HMXB) in which the optical companion is an early type giant or supergiant star and the collapsed object is a neutron star or a black hole. They are concentrated around the galactic plane. The mass transfer is usually occurring via stellar wind; they show hard pulsed X-ray emission (from 0.069 to 1413 s) with KT $\geq$ 9 keV; typical X-ray luminosity is ranging from 10^{34} to 10^{39} erg s^{-1}, and the ratio of X-ray to optical luminosity is $\sim 10^{-3}$–10. The HMXBs can be divided in two sub-classes:

1.1. Hard Transient X-ray Sources (HXTS) in which the neutron star is eccentrically (e $\sim$ 0.2–0.5) orbiting around a V-III luminosity-class Be star ($P_{orb} >$ 10 days); they show strong variable pulsed hard X-ray emission ($L_{Xmax}/L_{Xmin} >$ 100) with KT $\geq$ 17 keV, and P_{spin} ranging from 0.069 to 1413 s; $L_X = 10^{34} - 10^{39}$ erg s^{-1}.

1.2. Permanent X-ray Sources in which the neutron star or black hole is circularly orbiting (e $\sim$ 0) around a giant or supergiant OB star ($P_{orb} <$ 10 days); they show an almost steady permanent pulsed hard X-ray emission ($L_{Xmax}/L_{Xmin} \ll$ 100), and P_{spin} ranging from 0.069 to 1413 s; $L_X \sim 10^{37}$ erg s^{-1}.

1.3. Supergiant X-ray Binaries (SGXBs): obscured sources, which display huge amount of low energy absorption produced by the dense wind of the supergiant companion, surrounded by a weakly magnetized neutron star.

1.4. Supergiant Fast X-ray transients (SFXT), a subclass of SGXBs and a new subclass of transients in which the formation of transient accretion discs could be partly responsible for the flaring activity in systems with narrow orbits. They show $L_{Xpeak} \approx 10^{36}$ erg s^{-1}, and $L_{Xquiecence} \approx 10^{32}$ erg s^{-1}.

2. Low Mass X-ray Binaries (LMXB) in which the optical companion is a low-mass-late-type star and the collapsed object is a neutron star or a black hole (P_{orb} from 41 min to 11.2 days). They are concentrated in the globular clusters, and in the halo around the galactic center. The mass transfer in these systems is usually occurring via Roche lobe overflow. Their emission in soft X-ray range is usually not pulsed with KT $\leq$ 9 keV. Their X-ray luminosity is ranging from 10^{36} to 10^{39} erg s^{-1} and $L_X/L_{opt} \sim 10^2$–10^4; many LMXBs show Quasi Periodic Oscillations (QPOs) between 0.02 and 1000 seconds and few of them also pulsed X-ray emission, such as Her X1, 4U 1626-27 and GX 1+4.

Many LMXB show transient behaviour in the form irregular X-ray bursts, when their luminosity increases several tens or hundreds times. During these luminous stages, steady periodical signals, with millisecond periods, have been observed in several of them. In few of them ms X-ray pulsars have been discovered in quiescent stages between bursts. The ms X-ray pulsars in LMXB form a link between binary X-ray sources and recycled binary radio pulsars, with ms periods and low

magnetic fields (Bisnovatyi-Kogan & Komberg[5]), which are formed on the place of these LMXB after ceasing of accretion, due to evolution of the companion star, transforming into low mass white dwarf, or a giant degenerate planet.

3. Cataclysmic Variables (CVs) in which the optical companion is a low-mass-late-type star and the compact object is a white dwarf. The detected CVs are spread roughly around the solar system at distance of 200-300 pc. Orbital periods are ranging from tens of minutes to about ten hours. The mass transfer is occurring either via Roche lobe overflow or via accretion columns or in an intermediate way depending on the value of the magnetic field. Typical X-ray luminosity is ranging from 10^{32} to 10^{34} erg s^{-1}. Updated reviews about CVs are those by Giovannelli[29]) and Giovannelli & Sabau-Graziati.[42]

4. RS Canum Venaticorum (RS CVn) type systems, in which no compact objects are present and the two components are a F or G hotter star and a K star. Typical X-ray luminosity is ranging from 10^{30} to 10^{31} erg s^{-1}. Usually in the current literature they are excluded from the class of X-ray binaries since historically they were discovered as X-ray emitters only with the second generation of X-ray experiments.

In binary systems there are essentially two ways for accreting matter from one star to the other: via accretion disk or via stellar wind (Giovannelli & Sabau-Graziati,[37] adapted from Blumenthal & Tucker[7]) (left panel of Fig 1). But in some cases there is a third way which is a mixture between the two, as for instance in eccentric binary systems close to the periastron passage where a temporary accretion disk can be formed around the neutron star (e.g. Giovannelli & Ziółkowski[35]), like shown in the right panel of Fig. 1 (Giovannelli & Sabau-Graziati,[37] after Nagase[64]).

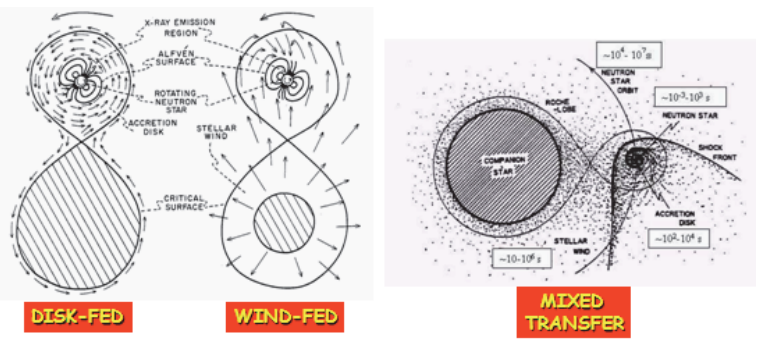


Fig. 1. Left panel: accretion in X-ray binary systems disk-fed and wind-fed (adopted from Giovannelli & Sabau-Graziati,[37] adapted from Blumenthal & Tucker[7]). Right panel: mixed transfer (adopted from Giovannelli & Sabau-Graziati,[37] after Nagase[64]).

XRBs are the best laboratory for the study of accreting processes thanks to their relative high luminosity in a large part of the electromagnetic spectrum. For

this reason, multifrequency observations are fundamental in understanding their morphology and the physics governing their behaviour.

Because of the strong interactions between the optical companion and collapsed object, low and high energy processes are strictly related.

Often, it is easier to perform observations of low energy processes (e.g. in radio, near-infrared (NIR) and optical bands) since the experiments are typically ground-based, on the contrary to observations of high energy processes, for which experiments are typically space-based.

2.1. *High mass X-ray binaries*

Among the X-ray binaries, the class of High Mass X-ray Binaries (HMXBs) constitutes an important group for studying the interactions either via stellar wind either via accretion disk between the optical and the compact companions.

2.2. *X-ray/Be systems*

The X-ray/Be binaries are the most abundant group of massive X-ray binaries in the galaxy, with a total inferred number of between 10^3 and 10^4. The ones which do occasionally flare-up as transient X-ray/Be systems are only the "tip" of this vast "iceberg" of systems (van den Heuvel & Rappaport[46]). The mass loss processes are due to the rapid rotation of the Be star, the stellar wind and, sporadically, to the expulsion of casual quantity of matter essentially triggered by gravitational effects close to the periastron passage of the neutron star. The long orbital period (> 10 days) and a large eccentricity of the orbit (> 0.2) together with transient hard X-ray behavior are the main characteristics of these systems. Among the whole sample of galactic systems containing 114 X-ray pulsars (Liu, van Paradijs & van den Heuvel[56]), only few of them have been extensively studied. Among these, the system A 0535+26/HDE 245770 – HDE 245770 was nicknamed Flavia star by Giovannelli & Sabau-Graziati[36]) – is the best known thanks to concomitant favorable causes, which rendered possible forty six years of coordinated multifrequency observations, most of them discussed in the past by e.g. Giovannelli & Sabau-Graziati,[36] Burger et al.[9]), Piccioni et al.,[68] and later by Giovannelli & Sabau-Graziati[40] and Giovannelli et al.[43, 44]). Accretion powered X-ray pulsars usually capture material from the optical companion via stellar wind, since this primary star generally does not fill its Roche lobe. However, in some specific conditions (e.g. the passage at the periastron of the neutron star) and in particular systems (e.g. A 0535+26/HDE 245770), it is possible the formation of a temporary accretion disk around the neutron star behind the shock front of the stellar wind. This enhances the efficiency of the process of mass transfer from the primary star onto the secondary collapsed star, as discussed by Giovannelli & Ziolkowski[35] and by Giovannelli et al.[38] in the case of A 0535+26.

Optical emission of HMXBs is dominated by that of the optical primary component, which is not, in general, strongly influenced by the presence of the X-ray

source. The behavior of the primary stars can be understood in the classical (or almost) frame-work of the astrophysics of these objects, i.e. by the study of their spectra which will provide indications on mass, radius, and luminosity. Both groups of HMXBs (transient and permanent) differ because of the different origin of the mass loss process: in the first, the mass loss process occurs via a strong stellar wind and/or because of an incipient Roche lobe over-flow; in the second group, the mass transfer is probably partially due to the rapid rotation of the primary star and partially to stellar wind and sporadically to expulsions of a casual quantity of matter, essentially triggered by gravitational effects because of periastron passage where the effect of the secondary collapsed star is more marked. A relationship between orbital period of HMXBs and the spin period of the X-ray pulsars is shown in Fig. 2 (updated from Giovannelli & Sabau-Graziati[37] and from Corbet[20,21]). It allows to recognize three kinds of systems, namely disk-fed, wind-fed [$P_{pulse} \propto (P_{orb})^{4/7}$], and X-ray/Be systems [$P_{pulse} \propto (P_{orb})^{2}$].

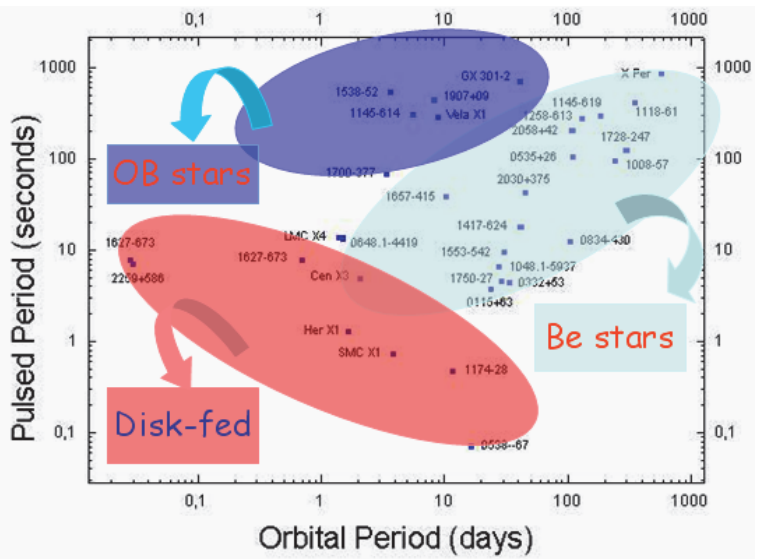


Fig. 2. Spin period vs orbital period for X-ray pulsars. Disk–fed systems are clearly separated by systems having as optical counterparts either OB stars or Be stars (adopted from Giovannelli & Sabau-Graziati,[37] after Corbet[20,21]).

Most of the systems having a Be primary star are hard X-ray (KT $>$ 10 KeV) transient sources (HXTS). They are concentrated on the galactic plane within a band of $\sim$ 3.9°. The orbits are quite elliptic and the orbital periods large (i.e. A 0538-66: e = 0.7, P_{orb} = 16.6 days (Skinner et al.[76]); A 0535+26: e = 0.47 (Finger, Wilson & Hagedon[23]), P_{orb} = 111.0 days (Priedhorsky & Terrell[69]). The X-ray flux during outburst phases is of order 10-1000 times greater than during

quiescent phases. For this reason, on the contrary, the stars belonging to the class of permanent X-ray sources, which do not present such strong variations in X-ray emission, can be also named "standard" high mass X-ray binaries. In X-ray/Be systems, the primary Be star is relatively not evolved and is contained within its Roche lobe. The strong outbursts occur almost periodically in time scales of the order of weeks-months. Their duration is shorter than the quiescent phases. During X-ray outbursts, spin-up phenomena in several systems have been observed (i.e. A 0535+26 and 4U 1118-61 (Rappaport & Joss[70]). The observed spin-up rates during the outbursts are consistent with torsional accretion due to an accretion disk (e.g. Ghosh[26]). So, the formation of a temporary accretion disk around the collapsed object should be possible during outburst phases (e.g. Giovannelli & Ziółkowski[35]).

3. Old & news from the transient X-ray/Be system A0535+26/HDE245770

The most studied HMXB system, for historical reasons and due to concomitant favourable causes, is the X-ray/Be system A 0535+26/HDE 245770. By means of long series of coordinated multifrequency measurements, very often simultaneously obtained, it was possible to:

- identify the optical counterpart HDE 245770 of the X-ray pulsar;
- identify various X-ray outbursts triggered by different states of the optical companion and influenced by the orbital parameters of the system;
- identify the presence of a temporary accretion disc around the neutron star at periastron.

Multifrequency observations of A 0535+26 started soon after its discovery as an X-ray pulsar by the Ariel-5 satellite on April 14, 1975 (Coe et al.[17]). The X-ray source was in outburst with intensity of ~ 2 Crab and showed a pulsation of ~ 104 s (Rosenberg et al.[73]). The hard X-ray spectrum during the decay from the April 1975 outburst became softer, so that the 19 May spectrum had $E^{-0.8}$ and the 1 June spectrum $E^{-1.1}$ (Ricketts et al.[71]). Between 13 and 19 April, 1975, as the nova brightened, the spectra showed some evidence of steepening. The best fit of the experimental data between roughly 27 and 28 April was compatible with an 8 keV black-body curve (Coe et al.[17]). The X-ray source decayed from the outburst with an *e*-folding time of 19 days in the energy range of 3-6 keV (Kaluzienski et al.[51]).

In the X-ray error box of the X-ray source A 0535+26, detected by Ariel V, were present 11 stars up to 23^{rd} magnitude and one of them (HDE 245770) of magnitude around 9 showed the H_α and H_β in emission, H_γ filled in with emission, and H_δ, H_ϵ,..., H_{10} in absorption (Margon et al.[59]). A priori probability of finding a 9 mag star in such a field is 0.004, thus HDE 245770 was considered as the probable optical counterpart of A0535+26.

But in order to really associate this star with the X-ray pulsar, it was necessary to find a clear signature proving that the two objects would belong to the same binary system. This happened thanks to a sudden insight of the author, Franco Giovannelli (FG), who predicted the fourth X-ray outburst of A 0535+26 around mid December 1977. For this reason, Giovannelli's group was observing in optical HDE 245770 around the predicted period for the X-ray outburst of A 0535+26. Figure 3 shows the X-ray flux intensity of A 0535+26 as deduced by various measurements available at that time, with obvious meaning of the symbols used (Giovannelli[28]). FG's intuition was sparked by looking at the rise of the X-ray flux (red line) and at the 24th May 1977 measurement (red asterisk): he assumed that the evident rise of the X-ray flux would have produced an outburst similar to the first one, which occurred in 1975. Then with a simple extrapolation he predicted the fourth outburst, similar to the second: and this happened!

Optical photoelectric photometry of HDE 245770 showed significant light enhancement of the star relative to the comparison star BD +26 876 between Dec. 17 and Dec 21 (here after 771220-E) and successive fading up to Jan. 6 (Bartolini et al.[1]), whilst satellite SAS-3 was detecting an X-ray flare (Chartres & Li[16]). The observed enhancement of optical emission followed by the flare-up of the X-ray source gave a direct argument strongly supporting the identification of HDE 245770 – later nicknamed Flavia' star by Giovannelli & Sabau-Graziati[36] – with A 0535+26.

Soon after, with spectra taken at the Loiano 152 cm telescope with a Boller & Chivens 26767 grating spectrograph (831 grooves/mm II-order grating: 39 Å mm^{-1}) onto Kodak 103 aO plates, it was possible to classify HDE 245770 as O9.7IIIe star. This classification was so good that it survives even to the recent dispute attempts made with modern technology. The mass and radius of the star are 15 $M_\odot$ and 14 $R_\odot$, respectively; the distance to the system is 1.8 ± 0.6 kpc (Giangrande et al.[27]).

UV spectra taken with the IUE enabled the reddening of the system to be determined as E(B-V) = 0.75 ± 0.05 mag, the rotational velocity of the O9.7IIIe star ($v_{rot} \sin i = 230 \pm 45$ km s^{-1}), the terminal velocity of the stellar wind ($v_\infty \simeq 630$ km s^{-1}), the mass loss rate ($\dot{M} \sim 10^{-8}$ $M_\odot$ yr^{-1} in quiescence (Giovannelli et al.[33]). During the October 1980 strong outburst, the mass loss rate was $\dot{M} \sim 7.7 \times 10^{-7}$ $M_\odot$ yr^{-1} (de Martino et al.[61]).

Complete reviews of this system can be found in Giovannelli et al.,[34] Giovannelli & Sabau-Graziati,[36] and Burger et al.[9]

Briefly, the properties of this system, placed at a distance of 1.8 ± 0.6 kpc (Giangrande et al.[27]), can be summarized as follows: a hard X-ray transient, long-period X-ray pulsar – the secondary star – is orbiting around the primary O9.7IIIe star. The masses are of $\sim 1.5 \pm 0.3$ $M_\odot$ (Joss & Rappaport;[50] Thorsett et al.;[79] van Kerkwijk, van Paradijs, J. & Zuiderwijk[52]), and 15 $M_\odot$ (Giangrande et al.[27]) for the secondary and primary star, respectively. The eccentricity is e = 0.47 (Finger, Wilson & Hagedon[23]). Usually the primary star does not fill its Roche lobe (de Loore et al.[58]). However, the suggestion that there might be a temporary accretion disk

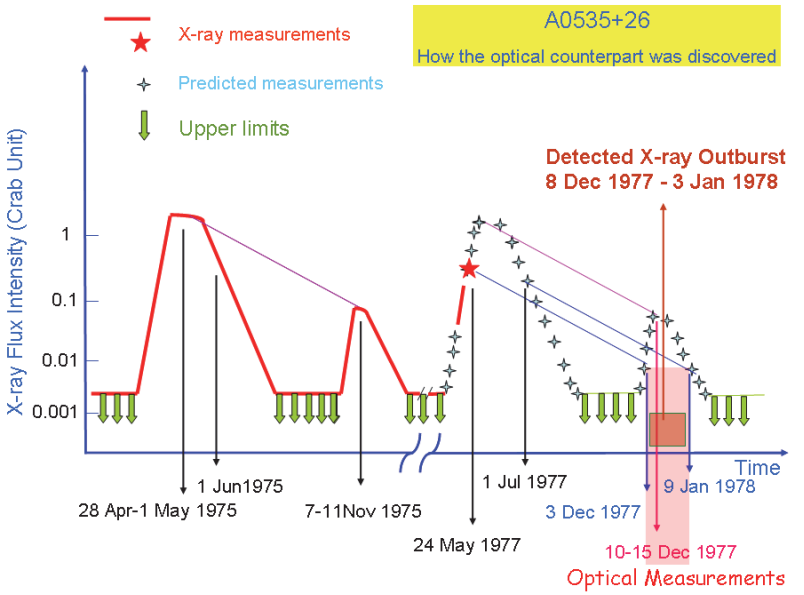


Fig. 3. X-ray flux versus time of A 0535+26. X-ray measurements are reported with red lines and asterisk, upper limits with green arrows, and predicted fluxes with light blue stars. Periods of real detected X-ray outburst and optical measurements are also marked (adopted from Giovannelli[28]).

around the X-ray pulsar when it approaches periastron (Giovannelli & Ziółkowski[35]) was confirmed by the X-ray measurements of Finger, Wilson & Harmon[24] and was discussed by Giovannelli et al.[38]

The first suggestion of Bartolini et al.[2] about the value of the orbital period ($P_{orb} = 110.856 \pm 0.002$ days), allowed Giovannelli & Sabau-Graziati[40] to discover a systematic delay (~ 8 days) of the X-ray outbursts with respect to the periastrons passages of the neutron star. Just a little before or simultaneously to the periastron, the system experiences an optical brightening ranging from ≈ 0.02 to ≈ 0.2 magnitudes.

The trigger of this discovery was the optical flare occurred at JD 2,444,944 (5th December 1981) (hereafter 811205-E; E stands for event) (Giovannelli et al.[34]) and followed by a short X-ray outburst (811213-E) (Nagase et al.[65]) – predicted by a private communication of Adriano Guarnieri (member of Giovannelli's group) to the team of Hakucho Japanise X-ray satellite. Unfortunately in the following years simultaneous optical and X-ray measurements were not always obtained around periastron passage. However, the available data were sufficient for showing the aforementioned systematic delay.

In order to explain and describe the 8-day delay between periastron passage and X-ray outbursts, GBK13 constructed a model adopting the orbital period determined by Priedhorsky & Terrell[69] from X-ray data ($P_{orb} = 111.0 \pm 0.4$ days), and

the ephemeris $JD_{opt-outb} = JD_0(2{,}444{,}944) \pm n(111.0 \pm 0.4)$ days; the 111-day orbital period agrees within the error bars with many other determinations reported in the literature: from optical data (e.g. Guarnieri et al.;[45] de Martino et al.;[60] Hutchings;[47] Janot-Pacheco, Motch & Mouchet[48]), and from X-ray data (e.g. Nagase et al.;[65] Priedhorsky & Terrell;[69] Motch et al.;[63] Finger, Wilson & Harmon;[24] Coe et al.;[18] Finger et al.[25]).

4. Time delay between optical and X-ray outbursts in A 0535+26/HDE 245770

A description of the time–delay among many optical and X-ray events occurring around the periastron passages in the system A 0535+26/HDE 245770 have been presented in the papers by GBK13 and Giovannelli et al.[43,44] However, just to remark the importance of the experimental evidence of such a time–delay I will present a few more examples that in my opinion definitively support the validity of the model developed in GBK13. Briefly, the model is the following: in the vicinity of periastron the mass flux $\dot{M}$ increases (depending on the activity of the Be star) between $\approx 10^{-8}$ and $\approx 10^{-7}$ $M_\odot$ yr^{-1}. The outer part of the accretion disk – geometrically thin and optically thick without advection (Shakura & Sunyaev;[75] Bisnovatyi-Kogan[4]) – becomes hotter, therefore the optical luminosity (L_{opt}) increases. Due to large turbulent viscosity, the wave of the large mass flux is propagating toward the neutron star, thus the X-ray luminosity (L_x) increases due to the appearance of a hot accretion disk region and due the accretion flow channeled by the magnetic field lines onto magnetic poles of the neutron star. The time–delay τ is the time between the optical and X-ray flashes appearance. Figure 4 shows a sketch of this model.

By using the ephemerides given by GBK13, namely:

$$JD_{opt-outb} = JD_0(2{,}444{,}944) \pm n(111.0 \pm 0.4) \text{ days}$$

that fixed the reference point at the date 5th December 1981 (811205-E), it was possible to explain the behaviour of the system during the year 2014 (Giovannelli et al.[43]). It was possible not only to predict the arrival time of the X-ray outbursts following the optical flashes, but also the intensity I_x of the X-ray flares, thanks to the relationship I_x versus ΔV_{mag}, where ΔV_{mag} is the relative variations of the V magnitude of the Be star around the periastron passage with respect to the level before and after such a passage. This relationship is shown in Fig. 5 (adapted from Giovannelli et al.[44]).

An impressive strong optical event have been detected on March 19, 2010 (JD 2,455,275). On the basis of such a strong optical activity – especially H_γ in emission – Giovannelli, Gualandi & Sabau-Graziati[39] predicted the incoming X-ray outburst of A 0535+26, which actually occurred (Caballero et al.[11]). The X-ray intensity reached was 1.18 Crab on April 3, 2010 in the range 15–50 keV of BAT/SWIFT

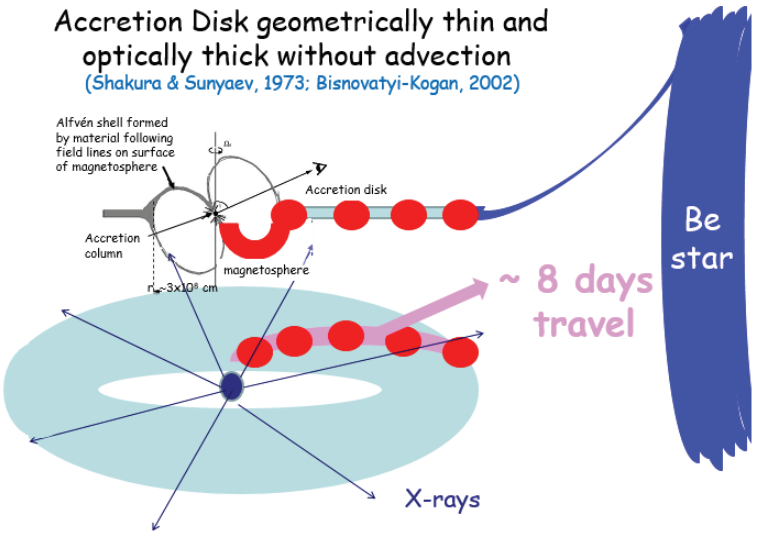


Fig. 4. Sketch of the viscous accretion disk model for explaining the time-delay between X-ray and optical flashes (adopted from:[41] GBK13).

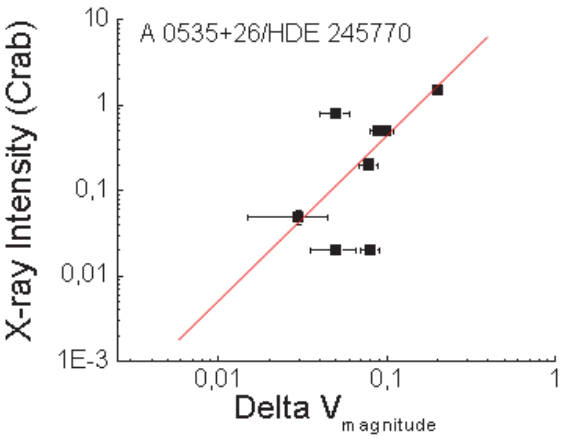


Fig. 5. Intensity of the X-ray flare of A 0535+26 versus the variation of V magnitude of HDE 245770 around the periastron passage (adapted from Giovannelli et al.[44]).

(Caballero et al.;[10–13] Caballero et al.[14]). Figure 6 shows the March–April 2010 event. The X-ray flare started about 8 days after the 93th periastron passage after the 811205-E, just when optical spectroscopy was performed by Giovannelli, Gualandi & Sabau-Graziati,[39] and reached the maximum about 12 days later and decayed in about 20 days roughly as occurred in 1975 when A0535+26 was discovered by the Ariel V satellite.

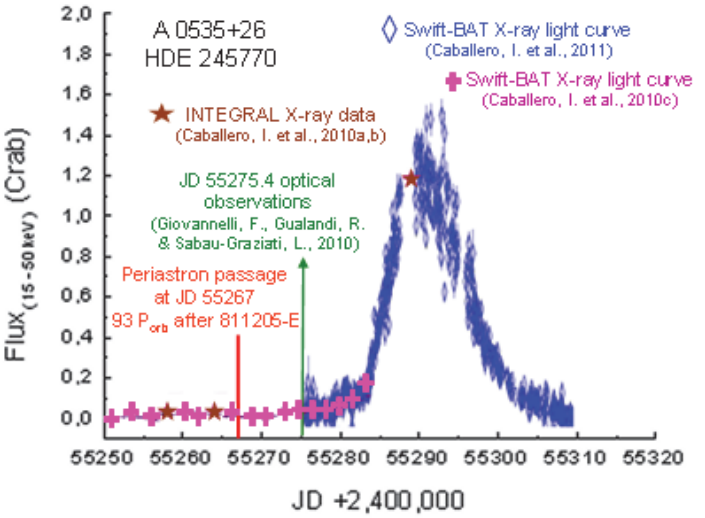


Fig. 6. The predicted March–April 2010 X-ray outburst of A 0535+26 (Giovannelli, Gualandi & Sabau-Graziati[39]) after the 93th passage at the periastron after 811205-E (Caballero et al.;[10–13] Caballero et al.[14]

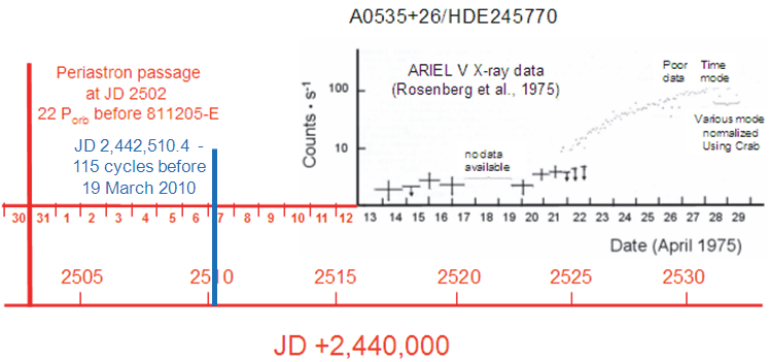


Fig. 7. The Periastron passage at the 22nd cycle before 811205-E (JD 2502) (red line) precedes of $\sim$ 14 days the X-ray outburst of A 0535+26 which starts approximatively on JD 2516 (after Rosenberg et al.[73]). The vertical blue line indicates the day April 7, 1975 (JD 2,442,510.4) just 115 cycles before the strong optical spectroscopic activity detected on March 19, 2010 (JD 2,455,275.4) (after:[41] GBK13).

The astonishing fact that definitively demonstrate the goodness of GBK13's ephemerides and the mechanism triggering the X-ray outburst with a time–delay with respect to the optical flare around the periastron passage is reported in Fig. 7. Indeed, in this figure, the measurements of the first detection of A 0535+26 by Ariel V satellite are reported. Unfortunately in 1975, around the time of the discovery of A 0535+26 no optical measurements are available for obvious reasons. The vertical red line indicates the time of the periastron passage following GBK13's ephemeris.

This passage occurred at the 22nd cycle before the 811205-E. The vertical blue line indicates the day April 7, 1975 (JD 2,442,510.4) just 115 cycles before the strong optical spectroscopic activity detected on March 19, 2010 (JD 2,455,275.4), that preceded the strong X-ray outburst reported in Fig. 6.

The similarity between the first X-ray outburst and that of March–April 2010 is evident, and the separation of the two events is exactly 115 cycles.

5. General model of time lag between optical and X-ray outbursts in binary accreting sources

In LMXBs (Low-Mass X-ray Binaries) the compact object can be either a neutron star or a black hole and the optical companion is a low mass star. The exchange of matter occurs via Roche lobe overflow, like shown in the left panel of Fig. 1. In HMXBs (High-Mass X-ray Binaries) the compact object can be either a neutron star or a black hole and the optical companion is a high mass star: giant or supergiant. The exchange of matter occurs mainly via stellar wind since usually the optical star does not fill its Roche lobe (Fig. 1, left panel). However, sometimes, the exchange of matter can occur in a mixed way because of the formation of an accretion disk around the compact object around the periatron passage (e.g. Giovannelli & Ziółkowski[35]) (Fig. 1, right panel).

As well known, X-ray/Be systems are formed by a compact star and an optical star. Obviously there is a mutual influence between the two stars. Low-energy (LE) processes influence high-energy (HE) processes and vice versa. Never confuse the effect with the cause. There is a general law in the Universe: **Cause and Effect**. The *Cause* generates an *Effect* and NOT vice versa!

It is right to remind that the mechanism proposed by GBK13 for explaining the X-ray-optical delay in A 0535+26/HDE 245770 is based on an enhanced mass flux propagation through the viscous accretion disk. This mechanism, known as UV-optical delay (the delay of the EUV flash with respect to the optical flash) was observed and modeled for cataclysmic variables (e.g. Smak;[77] Lasota[55]). Time delays have been detected also in several other X-ray transient binaries. This is the reason that urged Bisnovatyi-Kogan & Giovannelli[6] (BKG17) to generalize the aforementioned model, developed for the particular case of A 0535+26/HDE 245770 (Flavia' star). This general model provides the formula (1) of the time delay between the optical and X-ray flashes appearance in transient cosmic accreting sources:

$$\tau = 6.9 \frac{\mathrm{m}^{2/3} \dot{\mathrm{m}}^{1/15}}{\alpha^{4/5} \left(\mathrm{T}_4\right)^{28/15}} \tag{1}$$

where:

$\mathrm{m} = \mathrm{M}/\mathrm{M}_\odot$; $\dot{\mathrm{m}} = \dot{\mathrm{M}}/(10^{-8}\ \mathrm{M}_\odot/\mathrm{yr})$; $\mathrm{T}_4 = \mathrm{T}_0/10^4$ K; α = viscosity, and T_0 = maximum temperature in optics.

By using this formula it is possible to obtain an excellent agreement between the experimental and theoretical delays found in:

- X-ray/Be system A0535+26/HDE245770: $\tau_{\rm exp} \simeq 8$ days (GBK13); $\tau_{\rm th} \simeq 8$ days;
- Cataclysmic variable SS Cygni; $\tau_{\rm exp} = 0.9$–1.4 days (Wheatley, Mauche & Mattei[81]); $\tau_{\rm th} \simeq 1.35$ days;
- Low-mass X-ray binary Aql X-1/V1333 Aql: $\tau_{\rm exp} \sim 3$ days (Shahbaz et al.[74]); $\tau_{\rm th} \simeq 3.2$days
- Black hole X-ray transient GRO J1655-40: $\tau_{\rm exp} \sim 6$ days (Orosz et al.[67]); $\tau_{\rm th} \simeq 6.5$ days.

In this general formula the α-viscosity parameter plays an important role, and usually it is hard to be determined. However, if the other parameters are known, because experimentally determined, the formula (1) can be used for determining α, taking into account the experimental delay measured in a certain source.

Over the last couple of decades we have witnessed the discovery of a multitude of highly ionized absorbers in high-resolution X-ray spectra from both BH and NS XRBs. The first detections were obtained thanks to ASCA on the BH binaries GROJ1655-40 and GRS 1915+105. Narrow absorption lines in the spectra of these systems identified as Fe XXV and Fe XXVI indicated the first of many discoveries of photo-ionized plasmas in LMXBs (Chandra, XMM-Newton and Suzaku). Black hole hot accretion flows occur in the regime of relatively low accretion rates and are operating in the nuclei of most of the galaxies in the universe. One of the most important progress in recent years in this field is about the wind or outflow. This progress is mainly attributed to the rapid development of numerical simulations of accretion flows, combined with observations on, e.g., Sgr A*, the SMBH in the Galactic center. The mass loss from a BH via wind is related to the mass accretion rate onto the BH as (Yuan[82]):

$$\dot{\rm M}_{\rm wind}({\rm r}) = \dot{\rm M}_{\rm BH} \times ({\rm r}/20{\rm r_g})^{\rm s} \quad \text{with} \quad {\rm s} \approx 1 \quad \text{and} \quad {\rm r_g} = \frac{2{\rm GM_{BH}}}{{\rm c}^2} \tag{2}$$

At this point it is useful to make a sort of summary about the number of XRBSs, including CVs. Liu, van Paradijs & van den Heuvel[56,57] and Ziółkowski[84] report 315 galactic XRBs: 197 LMXBs (63%) and 118 HMXBs (37%), 72 of which are Be/X-ray systems; moreover there are 62 BH candidates. Coleiro & Chaty[19] report that in the Milky way there are $\geq$ 200 HMXBs. Ritter & Kolb[72] catalogue, in the 7.20 (Dec. 2013) version, reports 1166 CVs. Buckley[8] reports about the discoveries of 530 new CVs from MASTER-Network and 855 CVs from Catalina Real Time Survey (CRTS) (http://nesssi.cacr.caltech.edu/DataRelease/). Ferrario, de Martino & Gänsicke[22] report the number of MCVs as $\approx$ 250, and $\sim$ 60 of which IPs, and considering those systems for which the magnetic field intensity has not yet been determined, their number is of $\approx$ 600. Table 1 shows the content of XRBSs, including CVs,

Table 1. Comparison of numbers of different classes of X-ray Binary Systems in the Milky Way and in the Magellanic Clouds (Ziółkowski;[84] Ferrario, de Martino & Gänsicke;[22] Buckley[8]).

Name of the Class	Milky Way	LMC	SMC
Total mass of the galaxy (in M_{SMC} units)	100	10	1
High Mass X-ray Binaries	118	26	83
in this Be/X-ray	72	19	79
Low Mass X-ray Binaries	197	2	-
Black Hole Candidates	62	2	-
Cataclysmic Variables	≈ 2000	-	-
in this MCVs	≈ 250	-	-
IPs	~ 60	-	-
B not yet determined	≈ 600	-	-

in the Galaxy, and in the LMC and SMC (Ziółkowski;[84] Ferrario, de Martino & Gänsicke;[22] Buckley[8]). The mass are expressed in unit of SMC.

Some attempts of evaluation of the number of SBHs in the Galaxy have been done. For instance, taking into account the γ-ray emissivity of the Galaxy (1.3×10^{43} s^{-1} for E > 100 MeV) measured by the SAS II satellite (Strong, Wolfendale & Worral[78]) and the processes of disk-fed accretion onto black holes, Giovannelli, Karakuła & Tkaczyk[31,32] – considering a spherical accretion flow with a constant Mach number, corresponding to the adiabatic power $\gamma = 5/3$ – found a possible upper limits to the number of black holes ($M \sim 10\ M_\odot$ and $\dot{M} \approx 10^{-8}\ M_\odot\ yr^{-1}$) of 10^{-5} - 10^{-4} of the total star population of the Galaxy for Mach's number 1 and 2, respectively.(*)

There is a class of intermediate-mass black holes (IMBHs), with masses > 100 $M_\odot$ up to $\approx 10^5\ M_\odot$. It contains a dozen systems, as listed in Johnstone.[49] However, black holes with masses of several hundred to a few thousand solar masses remain elusive, as reported in a review by Casares & Jonker[15] where a deeply discussion about the mass measurements of SBHs and IMBHs is contained.

Supermassive black holes (SMBHs) are 10^6–10^{10} times more massive than our Sun and are found in the centers of galaxies (see the exhaustive review by Kormendy & Ho[54]). Most galaxies, and maybe all of them, harbor such a black hole. So in our region of the Universe, there are some 10^{11} SMBHs. The nearest one resides in the center of our Milky Way galaxy. The most distant one we know of resides in a quasar

(*)The Mach number is given by the ratio of the velocity of the gas to the local sound speed. In order to evaluate the temperature, concentration and velocity of the plasma near the black hole it is necessary to solve the system of equations describing the plasma motion (Michel[62]) taking into account the distance from the black hole in units of gravitational radius and the u = R-component of four velocity. Mach's number 1 and 2 correspond to different values of u_0^2 (1.0266 and 2.1213, respectively) at a distance r_0 from the black hole (Giovannelli, Karakuła & Tkaczyk[31,32]).

galaxy billions of lightyears away. SMBHs grow in size as they gorge on surrounding matter.

A list of BH candidates has been reported by Johnston[49] and provides the input for constructing the map of sky locations of BH candidates. Source list includes results reported in Kormendy & Gebhardt,[53] Orosz,[66] Tremaine et al.,[80] and Ziółkowski.[83]

6. Conclusions

The zoo of X-ray binary systems has been presented together with the main characteristics of the single constituent element. A particular emphasis has been devoted to the individual X-ray/Be system A 0535+26/HDE 245770 for which I discussed its history since its discovery by the Ariel-V satellite. Thanks to the long term survey of this system a delay between X-ray and optical flares has been recognized. This fact triggered the work for understanding the physical reasons of this phenomenon that allowed Bisnovatyi-Kogan & Giovannelli[6] to develop a model of time lag between optical and X-ray flashes for close-binary galactic sources with accretion disks.

The time lag in disk-accreting galactic close-binary sources is based on a sudden increase in the accretion flow that starts at the disk periphery and is related to the optical maximum. The massive accretion layer propagates to the central compact source as a result of the turbulent viscosity. The X-ray flash occurs when this massive layer reaches the inner hot regions of the accretion disk and falls into the central compact object. The matter in the accretion disk moves inside with a speed that is determined by the turbulent viscosity. They described this model quantitatively and derived an analytic formula that determines the value of the time lag. This formula gives results that agree well with observational values.

Acknowledgments

This research has made use of the NASA's Astrophysics Data System.

References

1. Bartolini, C., Guarnieri, A., Piccioni, A., Giangrande, A., Giovannelli, F.: 1978, IAU Circ. No. 3167.
2. Bartolini, C., Bianco, G., Guarnieri, A., Piccioni, A., Giovannelli, F.: 1983, *Hvar Obs. Bull.* **7(1)**, 159.
3. Bisnovatyi-Kogan, G.S.: 2002, in *Black Hole Astrophysics*, Proc. of the Sixth APCTP Winter School, Hyun Kyu Lee & Myeong-Gu Park (Eds.), World Scientific Publishing Co. Pte. Ltd., ISBN 9789812776761, pp. 187-206.
4. Bisnovatyi-Kogan, G.S.: 2011, *Stellar Physics 2: Stellar Evolution and Stability*, Berlin-Heidelberg: Springer-Verlag.
5. Bisnovatyi-Kogan, G.S., Komberg, B.V.: 1974, Sov. Astron. 18, 217.
6. Bisnovatyi-Kogan, G.S., Giovannelli, G.: 2017, A&A 599, A55, 7 pp. (BKG17).

7. Blumenthal, G.R., Tucker, W.H.: 1974, Ann. Rev. A&A 12, 23-46.
8. Buckley, D., 2015, talk at the Palermo Workshop on "*The Golden Age of Cataclysmic Variables and Related Objects - III*".
9. Burger, M., van Dessel, E.L., Giovannelli, F., Sabau-Graziati, L., Bartolini, C. et al.: 1996, in *Multifrequency Behaviour of High Energy Cosmic Sources*, F. Giovannelli & L. Sabau-Graziati (eds.), Mem. SAIt. 67, 365.
10. Caballero, I., Lebrun, F., Rodriguez, J., Soldi, S., Mattana, F., et al.: 2010a, ATel. 2496.
11. Caballero, I., Santangelo, A., Pottschmidt, K., Klochkov, D., Rodriguez, J., et al.: 2010b, ATel. 2541.
12. Caballero, I., Pottschmidt, K., Barragán, L., Ferrigno, C., Klochkov, D., et al.: 2010c, Talk at CRSF Meeting, Tübingen 2010.
13. Caballero, I., Kretschmar, P., Pottschmidt, K., Santangelo, A., Wilms, J., et al.: 2010d, AIPC 1248, 147.
14. Caballero, I., Pottschmidt, K., Santangelo, A., Barragán, L., Klochkov, D., et al.: 2011, arXiv:1107.3417.
15. Casares, J., Jonker, P.G.: 2014, SSRv 183, 223-252.
16. Chartres, M., Li, F.: 1977, IAU Circ. No. 3154.
17. Coe, M. J., Carpenter, G.F., Engel, A.R., Quenby, J.J.: 1975, Nature 256, 630.
18. Coe, M.J., Reig, P., McBride, V.A., Galache, J.L., Fabregat, J.: 2006, MNRAS 368, 447.
19. Coleiro, A., Chaty, S.: 2013, ApJ 764, 185.
20. Corbet, R.H.D.: 1984, A&A 141, 91.
21. Corbet, R.H.D.: 1986, MNRAS 220, 1047.
22. Ferrario, L., de Martino, D., Gänsicke, B.T.: 2015, SSRv 191, 111-169.
23. Finger, M.H., Wilson, R.B., Hagedon, K.S. 1994 IAU Circ. No. 5931.
24. Finger, M.H., Wilson, R.B., Harmon, B.A.: 1996, ApJ 459, 288.
25. Finger, M.H., Camero-Arranz, A., Kretschmar, P., Wilson, C., and Patel, S.: 2006, BAAS 38, 359.
26. Ghosh, P.: 1994, in *The Evolution of X-Ray Binaries*, S.S. Holt & C.S. Day (Eds.), AIP Conf. Proc. 308, 439.
27. Giangrande, A., Giovannelli, F., Bartolini, C., Guarnieri, A., Piccioni, A.: 1980, A&A Suppl. Ser. 40, 289.
28. Giovannelli, F.: 2005, *The Impact of Multifrequency Observations in High Energy Astrophysics*, Ph. D. Thesis, University of Barcelona, Spain.
29. Giovannelli, F.: 2008, Ch. J. A&A Suppl. 8, 237-258
30. Giovannelli, F.: 2016, in *Proc. of the 4th Ann. Conf. on High Energy Astrophysics in Southern Africa (HEASA 2016)*. Online at http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=275, id. 31.
31. Giovannelli, F., Karakuła, S., Tkaczyk, W.: 1981, in *Origin of Cosmic Rays*, Setti, G., Spada, G. & Wolfendale, A.W. (Eds.), IAU Symp. 94, p. 335.
32. Giovannelli, F., Karakuła, S., Tkaczyk, W.: 1982, Acta Astron. 32, 121.
33. Giovannelli, F., de Loore, C., Bartolini, C., Burger, M., Ferrari-Toniolo, M., et al. 1982 in *Proc. of the Third European IUE Conference*, ESA SP-176, 233.
34. Giovannelli, F., Ferrari-Toniolo, M., Persi, P., Golynskaya, I.M., Kurt, V.G., et al.: 1985 in *Multifrequency Behaviour of Galactic Accreting Sources*, F. Giovannelli (ed.), Edizioni Scientifiche SIDEREA, Roma, p. 284.
35. Giovannelli, F., Ziółkowski, J.: 1990, AcA 40, 95-103.
36. Giovannelli, F., Sabau-Graziati, L.: 1992, SSRv 59, 1-81.
37. Giovannelli, F., Sabau-Graziati, L.: 2001, Ap&SS, 276, 67-80.

38. Giovannelli, F., Bernabei, S., Rossi, C., Sabau-Graziati, L.: 2007, A&A, 475, 651.
39. Giovannelli, F., Gualandi, R., Sabau-Graziati, L.: 2010, ATel. 2497.
40. Giovannelli, F., Sabau-Graziati, L.: 2011, Acta Polytechnica Vol. 51, No. 2., p. 21.
41. Giovannelli, F., Bisnovatyi-Kogan, G.S., Klepnev, A.S.: 2013, A&A 560, id.A1, 11 pp (GBK13).
42. Giovannelli, F., Sabau-Graziati, L.: 2015, in *The Golden Age of Cataclysmic Variables and Related Objects - II*, F. Giovannelli & L. Sabau-Graziati (Eds.), Acta Polytechnica, CTU Proc. ISSN 2336-5382, Vol. 2 No. 1, 3-20.
43. Giovannelli, F., Bisnovatyi-Kogan, G.S., Bruni, I., Corfini, G., Martinelli, F., Rossi, C.: 2015a, AcA 65, 107-116.
44. Giovannelli, F., Rossi, C., Bisnovatyi-Kogan, G., Bruni, I., Fasano, A., Salas Procas, J.: 2015b, Proc. of the *XI Multifrequency Behaviour of High Energy Cosmic Sources* Workshop (MULTIF15). Online at http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=246, id.39
45. Guarnieri, A., Bartolini, C., Piccioni, A., Giovannelli, F.: 1985, in *Multifrequency Behaviour of Galactic Accreting Sources*, F. Giovannelli (ed.), Frascati: CNR, Istituto di Astrofisica, Edizioni Scientifiche SIDEREA, Roma, p. 310.
46. van den Heuvel, E.P.J., Rappaport, S.: 1987, in *Physics of Be Stars*, A. Slettebak & T.P. Snow (eds.), Cambridge and New York, Cambridge University Press, Proc. of the IAU Coll. N. 92, p. 291.
47. Hutchings, J.B.: 1984, PASP 96, 312.
48. Janot-Pacheco, E., Motch, C., Mouchet, M.: 1987, A&A 177, 91.
49. Johnstone, Wm. R.: 2004, http://www.johnstonsarchive.net/relativity/bhctable.html
50. Joss, P.C., Rappaport, S.A.: 1984, Ann. Rev. Astron. Astrophys. 22, 537.
51. Kaluzienski, L.J., Holt, S.S., Boldt, E.A. Serlemitsos, P.J.: 1975, Nature 256, 633.
52. van Kerkwijk, M.H., van Paradijs, J., Zuiderwijk, E.J.: 1995, A&A 303, 497.
53. Kormendy, J., Gebhardt, K.: 2001, arXiv:astro-ph/0105230v1.
54. Kormendy, J., Ho, L.C.: 2013, Ann. Rev. A&A 51, 511-653.
55. Lasota, J.-P.: 2001, New Astron. Rev. 45, 449.
56. Liu, Q.Z., van Paradijs, J., van den Heuvel, E.P.J.: 2006, A&A 455, 1165.
57. Liu, Q.Z., van Paradijs, J., van den Heuvel, E.P.J.: 2007, A&A 469, 807.
58. de Loore, C., Giovannelli, F., van Dessel, E.L., Bartolini, C., Burger, M., et al.: 1984 A&A 141, 279.
59. Margon, B., Nelson, J., Chanan, G., Bowyer, S., Thorstensen, J.R.: 1977, ApJ 216, 811.
60. de Martino, D., Vittone, A., Giovannelli, F., Ciatti, F., Margoni, R., et al.: 1985, in *Multifrequency Behaviour of Galactic Accreting Sources*, Proceedings of the 1984 Frascati Workshop, Franco Giovannelli (ed.). Frascati: CNR, Istituto di Astrofisica, Ed. Scientifiche SIDEREA, Roma, p. 326.
61. de Martino, D., Waters, L.B.F.M., Giovannelli, F., Persi, P.: 1989, in *The 23rd ESLAB Symposium on Two Topics in X Ray Astronomy*, Volume 1 *X-Ray Binaries*, ESA SP 296, 519-520.
62. Michel, F.C.: 1972, Ap&SS 15, 153.
63. Motch, C., Stella, L., Janot-Pacheco, E., Mouchet, M.: 1991, ApJ 369, 490.
64. Nagase, F.: 1989, PASJ 41, no. 1, 1-79.
65. Nagase, F., Hayakawa, S., Kunieda, H., Makino, F., Masai, K., et al.: 1982, ApJ 263, 814.
66. Orosz, J.A.: 2002, arXiv:astro-ph/0209041v1.
67. Orosz, J.A., Remillard, R.A., Bailyn, C.D., McClintock, J.E.: 1997, ApJL 478, L83.

68. Piccioni, A., Bartolini, C., Bernabei, S., Guarnieri, A., Tarozzi, F., Valentini, G.: 1999, in *Frontier Objects in Astrophysics and Particle Physics*, F. Giovannelli & G. Mannocchi (Eds.), SIF, Bologna, Italy, 65, 195.
69. Priedhorsky, W.C., Terrell, J.: 1983, Nature 303, 681.
70. Rappaport, S., Joss, P.C.: 1981, in *X-Ray Astronomy with the Einstein Satellite*, R. Giacconi (ed.), D. Reidel Publ. Co., Dordrecht, Holland, p. 123.
71. Ricketts, M.J., Turner, M.J.L., Page, C.G., Pounds, K.A.: 1975, 256, 631-633.
72. Ritter, H., Kolb, U.: 2003, A&A 404, 301.
73. Rosenberg, F.D., Eyles, C.J., Skinner, G.K., Willmore, A.P.: 1975, Nature 256, 628.
74. Shahbaz, T., Bandyopadhyay, R.M., Charles, P.A., Wagner, R.M., Muhli, P., et al.: 1998, MNRAS 300, 1035.
75. Shakura, N.I., Sunyaev, R.A.: 1973, A&A 24, 337.
76. Skinner, G.K.,, Bedford, D.K., Elsner, R.F., Leahy, D., Weisskopf, M.C., Grindlay, J.: 1982, Nature, 297, 568-570.
77. Smak, J.: 1984, PASP 96, 5.
78. Strong, A.W., Wolfendale, A.W., Worral, D.M.: 1976, MNRAS 175, 23.
79. Thorsett, S.E., Arzoumanian, Z., McKinnon, M.M., Taylor, J.H.: 1993, ApJ 405, L29.
80. Tremaine, S., Gebhardt, K., Bender, R., Bower, G., Dressler, A. et al.: 2002, ApJ 574, 740-753.
81. Wheatley, P.J., Mauche, C.W., Mattei, J.A.: 2003, MNRAS 345, 49.
82. Yuan, F.: 2016, in *Astrophysics of Black Holes*, C. Bambi (Ed.), ApSSLibrary 440, 152-168.
83. Ziółkowski, J.: 2003, arXiv:astro-ph/0307307v1.
84. Ziółkowski, J.: 2013, Acta Polytechnica Vol 53, Suppl., 665.

Rotation of the crescent image of M87* and polarization of its ESE hotspot

Krzysztof Nalewajko
Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences
Bartycka 18, 00-716 Warsaw, Poland
E-mail: knalew@camk.edu.pl

The first image of the black hole (BH) M87* obtained by the Event Horizon Telescope (EHT) has the shape of a crescent extending from the E to WSW position angles, with a tentative 'ESE hotspot'. Assuming that the BH spin vector is aligned with both the inner accretion axis and the projected direction of the kpc-scale relativistic jet, the position of the ESE hotspot is inconsistent with the axisymmetric accretion flow. Recent polarimetric EHT images of M87* show that the ESE hotspot is essentially unpolarized, which strongly supports its distinct origin. If the hotspot emission is due to the synchrotron radiation, its depolarization requires either isotropically tangled magnetic fields or an additional Faraday dispersion measure. The 6-day EHT observing campaign in April 2017 allowed in principle to detect orbital motions advancing by up to $\sim 60°$. The apparent rotation rate of the major axis of the EHT crescent image is consistent with the rotation rate of the Faraday-corrected polarization angle measured by the ALMA. However, the counterclockwise (CCW) sense of these rotations is opposite to the clockwise (CW) rotation of the plasma flows implied by the N-S brightness asymmetry, which might indicate that accretion in M87 is retrograde.

Keywords: Black hole physics; Galaxies:active; Galaxies:individual:M87; Relativistic processes.

1. Introduction

The Event Horizon Telescope (EHT) resolved at 1.3 mm the core M87* of the nearby (distance of $D \simeq 16.8$ Mpc) radio galaxy Virgo A (M87) into a crescent image of angular diameter of $\simeq 42$ μas, which corresponds to $\simeq 11$ gravitational radii $R_{\rm g} = GM_{\rm BH}/c^2$ for the black hole (BH) mass of $M_{\rm BH} \simeq 6.5 \times 10^9 M_\odot$ (G is the gravitational constant, c is the speed of light, $M_\odot$ is the mass of the Sun; $R_{\rm g}/c \simeq 8.9$ h).[1–6] M87* is thought to be the base of a relativistic jet observed in radio,[7–9] optical[10] and X-ray[11] bands, propagating in the WNW direction (position angle $\simeq 288°$) beyond the kpc scales. The inclination angle of the jet axis is estimated at $\simeq 17°$.[9] If the jet is powered by the Blandford-Znajek mechanism,[12] its direction of propagation should be aligned with the BH spin. Accreting plasma in the vicinity of spinning black hole would be forced to swirl with trans-relativistic velocities preferentially along the BH equatorial plane. The resulting difference in the Doppler beaming of emission should result in a roughly N-S brightness asymmetry of the crescent image. It is in fact observed by the EHT that the S side of the M87* crescent image is significantly brighter than the N side.

The angular extent of the M87* crescent is roughly from the E to WSW position angles. There is a hint of substructures in the angular brightness profiles that depends significantly on the image reconstruction method. The observed crescent may consist of the *proper crescent* extending from the SSE to WSW position angles and the *ESE hotspot*. Using simple axisymmetric models for the geometry of emitting regions, we have previously suggested that the proper crescent can be explained by emitting regions symmetric with respect to the BH spin/jet WNW axis, and that the ESE hotspot (opposite to the WNW jet) would be rather a localized perturbation of the accretion flow.[13] Such azimuthal perturbations are commonly observed in the results of GRMHD numerical simulations of magnetized BH accretion flows.[5,14,15]

2. Polarization of the ESE hotspot

More recently, the EHT Collaboration presented the first resolved polarimetric images of M87*.[16,17] These images show linear polarization at the level of $\sim 15\%$. In our interpretation, this polarization is largely limited the proper crescent, while the ESE hotspot is essentially unpolarized. Such a striking difference between the polarization properties of these two regions strongly supports their distinct origin.

The very low polarization degree of the ESE hotspot calls into question whether it is produced by the same emission mechanism (presumably synchrotron), whether it is optically thin (or a localized millimeter photosphere), or whether it involves ordered or tangled magnetic fields. It could also result from depolarization by an additional localized Faraday screen. Unresolved ALMA polarimetry simultaneous to the EHT campaign was performed in two bands — 1.3 mm and 3 mm — which allowed to estimate the simultaneous rotation measures as $\sim \pm 10^5$ rad/m^2.[18] Since depolarizing a 1.3 mm signal by factor e requires a dispersion measure of $\simeq 4 \times 10^5$ rad/m^2, depolarization of the ESE hotspot may require an additional dispersion measure of order $\sim 10^6$ rad/m^2. Such dispersion measures are plausible in the advection dominated accretion flow (ADAF) models[19] for the mass accretion rate of $\dot{M} \gtrsim 2 \times 10^{-3} M_\odot$/yr.[20]

3. Rotation of the crescent image and net polarization vector

The 2017 EHT observing campaign spanned 6 days, resolved images were obtained from data recorded on April 5th, 6th, 10th and 11th. This time scale can be compared with the orbital period of a source propagating along the innermost stable circular orbit (ISCO) in Schwarzschild[a] metric: $P_{\rm ISCO} \simeq 92 R_{\rm g}/c \simeq 34$ days. Hence, during the 6-day campaign one could expect to observe an orbital advance by $\sim 60^\circ$. Note that the estimated viewing angle of the M87 jet is sufficiently small that equatorial accretion flows would be well aligned with the plane of the sky.

[a]The effect of BH spin is not included in this basic analysis, but it would be an obvious next level of sophistication.

For comparison, in the case of Sgr A* ($R_{\rm g}/c \simeq 20$ s; $P_{\rm ISCO} \simeq 31$ min) the VLTI/GRAVITY Collaboration detected orbital motion at the period of $P_{\rm GRAVITY} \simeq 40$ min $\simeq 1.3 P_{\rm ISCO}$,[21] which corresponds to a circular Schwarzschild geodesic of radius $R_{\rm circ} \simeq 7.1 R_{\rm g}$.

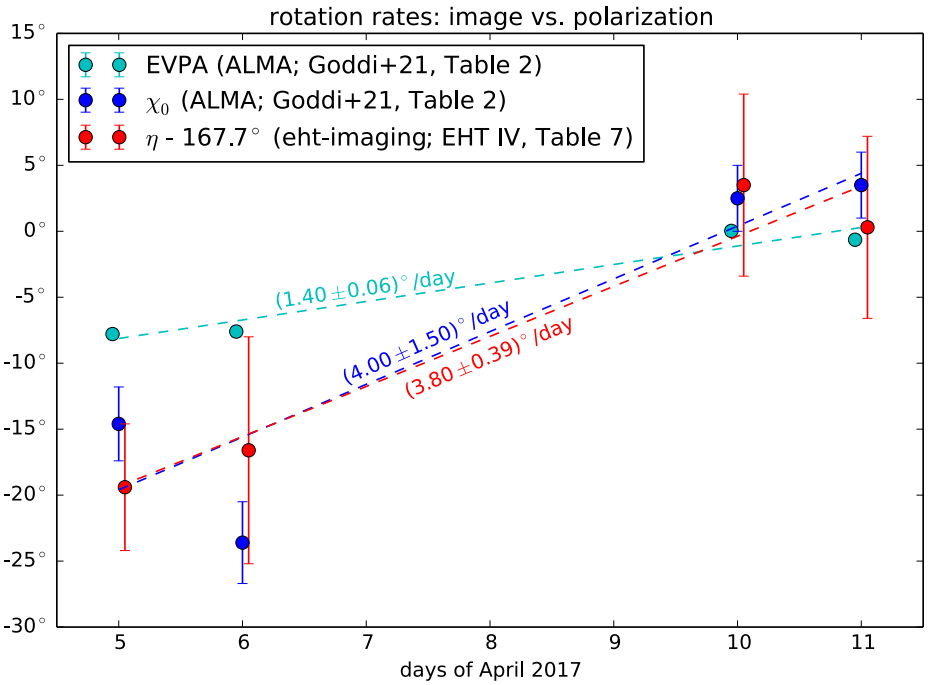


Fig. 1. Time evolution of the orientation angles of the total-intensity crescent image and net polarization. See the main text for details.

In Figure 1, we compare two kinds of orientation angles measured independently for each day of the M87 campaign. First, we consider the mean orientation angles η of the crescent image according to the `eht-imaging` algorithm from the Table 7 of EHT Paper IV.[4] This angle increases by $\simeq +23°$ between April 5th and April 10th, and the mean rotation rate is $(3.8 \pm 0.6)°$/day. Second, we consider the net polarization angles from the Table 2 of the ALMA Paper:[18] the observed angles EVPA, and the angles χ_0 corrected for Faraday rotation using the simultaneous rotation measure values. The EVPA increases by $\simeq +7.8°$ between April 5th and April 10th, while χ_0 increases by $\simeq +27°$ between April 6th and April 11th. The mean rotation rate is $(1.4 \pm 0.3)°$/day for the EVPA and $(4.0 \pm 1.2)°$/day for the χ_0. The rotation rate of the Faraday-corrected polarization angle is thus consistent with the rotation rate of the crescent. The rate of 3.9°/day corresponds to the orbital

period of $P \simeq 92$ days $\simeq 2.7 P_{\rm ISCO}$. The apparent rotation rate of the crescent image and net polarization would correspond to $R_{\rm circ} \simeq 11.6\, R_{\rm g}$.

The trend of increasing position angle measured from N to E corresponds to the counterclockwise (CCW) rotation on the sky for the relatively outer accretion flow. On the other hand, since the S side of the M87* image is clearly brighter than the N side, should this be attributed to the trans-relativistic Doppler beaming of the inner accretion flow (funnel wall), in light of the roughly W direction of the projected jet, the inner accretion flow (and the BH itself) should rotate in the clockwise (CW) direction on the sky. This situation of opposite outer and inner accretion flows corresponds to the bottom left panel of Figure 5 in EHT Paper V.[5] This would suggest that accretion in M87 is retrograde. And this in turn would constrain the range of GRMHD models listed in Table 2 of EHT Paper V[5] to two categories: the high-spin ($a_* = -0.94$) SANE models, and the medium-spin ($a_* = -0.5$) MAD models.

Acknowledgments

I thank Brian Punsly for motivating this contribution by inviting the author to speak at the Sixteenth Marcel Grossmann Meeting. I also thank Agata Różańska and Marek Sikora for discussions. This work was partially supported by the Polish National Science Centre grant 2015/18/E/ST9/00580.

References

1. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *ApJ* **875**, p. L1 (April 2019).
2. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. II. Array and Instrumentation, *ApJ* **875**, p. L2 (April 2019).
3. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. III. Data Processing and Calibration, *ApJ* **875**, p. L3 (April 2019).
4. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole, *ApJ* **875**, p. L4 (April 2019).
5. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring, *ApJ* **875**, p. L5 (April 2019).
6. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole, *ApJ* **875**, p. L6 (April 2019).
7. W. Junor, J. A. Biretta and M. Livio, Formation of the radio jet in M87 at 100 Schwarzschild radii from the central black hole, *Nature* **401**, 891 (October 1999).
8. C. C. Cheung, D. E. Harris and Ł. Stawarz, Superluminal Radio Features in the M87 Jet and the Site of Flaring TeV Gamma-Ray Emission, *ApJ* **663**, L65 (July 2007).
9. R. C. Walker P. E. Hardee, F. B. Davies, C. Ly and W. Junor, The Structure and Dynamics of the Subparsec Jet in M87 Based on 50 VLBA Observations over 17 Years at 43 GHz, *ApJ* **855**, p. 128 (March 2018).
10. J. A. Biretta, W. B. Sparks and F. Macchetto, Hubble Space Telescope Observations of Superluminal Motion in the M87 Jet, *ApJ* **520**, 621 (August 1999).
11. A. S. Wilson and Y. Yang, Chandra X-Ray Imaging and Spectroscopy of the M87 Jet and Nucleus, *ApJ* **568**, 133 (March 2002).

12. R. D. Blandford and R. L. Znajek, Electromagnetic extraction of energy from Kerr black holes., *MNRAS* **179**, 433 (May 1977).
13. K. Nalewajko, M. Sikora and A. Różańska, Orientation of the crescent image of M 87*, *A&A* **634**, p. A38 (February 2020).
14. O. Porth, Y. Mizuno, Z. Younsi and C. M. Fromm, Flares in the Galactic Centre - I. Orbiting flux tubes in magnetically arrested black hole accretion discs, *MNRAS* **502**, 2023 (April 2021).
15. N. Scepi, J. Dexter and M. C. Begelman, Sgr A* X-ray flares from non-thermal particle acceleration in a magnetically arrested disc, *arXiv e-prints* , p. arXiv:2107.08056 (July 2021).
16. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. VII. Polarization of the Ring, *ApJ* **910**, p. L12 (March 2021).
17. Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. VIII. Magnetic Field Structure near The Event Horizon, *ApJ* **910**, p. L13 (March 2021).
18. C. Goddi, I. Martí-Vidal, H. Messias, G. C. Bower, A. E. Broderick and et al., Polarimetric Properties of Event Horizon Telescope Targets from ALMA, *ApJ* **910**, p. L14 (March 2021).
19. D. P. Marrone, J. M. Moran, J.-H. Zhao and R. Rao, Interferometric Measurements of Variable 340 GHz Linear Polarization in Sagittarius A*, *ApJ* **640**, 308 (March 2006).
20. M. Mościbrodzka, J. Dexter, J. Davelaar and H. Falcke, Faraday rotation in GRMHD simulations of the jet launching zone of M87, *MNRAS* **468**, 2214 (June 2017).
21. Gravity Collaboration, Detection of orbital motions near the last stable circular orbit of the massive black hole SgrA*, *A&A* **618**, p. L10 (October 2018).

Magnetic reconnection in jet-accretion disk systems

Chandra B. Singh

South-Western Institute for Astronomy Research, Yunnan University, University Town, Chenggong District, Kunming 650500, People's Republic of China
E-mail: chandrasingh@ynu.edu.cn

Elisabete M. de Gouveia Dal Pino, Luis H.S. Kadowaki and Tania E. Medina-Torrejón

Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, São Paulo, 05508-090, Brasil

Yosuke Mizuno

Tsung-Dao Lee Institute and School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

David Garofalo

Department of Physics, Kennesaw State University, Marietta GA 30060, USA

Grzegorz Kowal

Escola de Artes, Ciências e Humanidades - Universidade de São Paulo, Av. Arlindo Béttio, 1000 – Vila Guaraciaba, São Paulo, 03838-000, Brasil

Black hole surroundings and relativistic jets host magnetically dominated regions and fast magnetic reconnection are likely to play an important role concerning astrophysical phenomena associated with such regions. In this contribution, we highlight the works related to turbulence-driven reconnection processes. These processes have been studied by us using analytical as well as numerical methods which showed that fast reconnection processes are powerful ways to giving rise to relativistic particles and associated non-thermal emissions around stellar-mass and supermassive black holes. The power released from the reconnection can even compete with those of extraction from black hole spin.

Keywords: Instabilities; Jets; Acceleration of Particles; Magnetic Reconnection; Magnetohydrodynamics; Numerical Methods

1. Introduction

Black holes are in a range of masses like 5-10 solar masses for stellar-mass to $10^6 - 10^{10}$ for supermassive black holes. Accretion disks, as well as relativistic jets, are quite common features for those black holes. Usually, relativistic jets can achieve high Lorentz factors and propagate several orders of magnitude in distances. Those are likely to be powered by the energy available from accretion disk or black hole spin. The base of the jets is supposed to be magnetically dominated which later becomes kinetically dominated so an efficient conversion is needed to explain

this picture. So magnetic reconnection can be a viable mechanism to make it happen. Another issue that requires explanation is regarding very high emissions (up to TeV) from highly under luminous radio galaxies which might occur close to the base of the jet and from a very compact region in magnetized surroundings around black holes. As standard processes in sites like shocks are likely to be weak, so promising candidate is magnetic reconnection processes and requires due attention.

In the presence of turbulence, the first-order Fermi process in the fast reconnection sites or current sheets is likely to accelerate particles to high energies as shown in analytical[1] and numerical works.[2,3] It was shown that the particles trapped within current sheets were found to undergo exponential growth in energy with time and the corresponding energy spectrum was power-law with spectral indices $\sim$ 1-2. Such a mechanism is well-suited for addressing very high energy phenomena seen in a range of systems containing black holes like galactic black holes or microquasars(GBHs), active galactic nuclei (AGNs), and gamma-ray bursts(GRBs). In this contribution, we discuss further works showing the role of fast magnetic reconnection in accretion disks and jets in such systems containing black holes.

2. Magnetic reconnection around black holes

de Gouveia Dal Pino and Lazarian[1] proposed a model showing that the first-order Fermi process in the coronal region around a black hole can accelerate particles due to the power extraction from fast magnetic reconnection between lines anchored in the accretion disk and black hole horizon. That model could address the origin of radio flares in GBHs and AGNs and later extended to gamma emissions and larger samples as well.[4] Further, the fundamental plane black hole activity which correlates the radio and X-ray emissions could be easily explained in the framework of fast magnetic reconnection model. In those models, a standard disk with corona as a central source around the black hole was used which is suitable for close to Eddington sources. Later the scenario was revisited for taking care of sub-Eddington sources as well by considering thick magnetically dominated accretion flow (MDAF).[5]

Figure 1 shows the calculated turbulence-driven fast magnetic reconnection power and the observed radio and gamma-ray luminosities for more than 270 sources including XRBs, low luminous AGNs (LLAGNs), and GRBs.[5] The diamonds in red and green colours represent the radio emissions in the case of LLAGNs and XRBs respectively. Some XRBs (Cyg X-1 and Cyg X-3) and LLAGNs (Cen A, IC 310, Per A, and M87) are highlighted which have been studied extensively in multi-wavelength studies. The reconnection power is much greater than the observed radio luminosities which means only a small fraction of this power is enough to explain the radio emission for most of the sources. Observed gamma-ray luminosities for 23 sources are included, represented by red and green circles for LLAGNs and XRBs respectively. The circles represent the emissions of 4 radio galaxies, 2 XRBs, and several Seyfert galaxies (most with upper limits in GeV). It is seen that the released

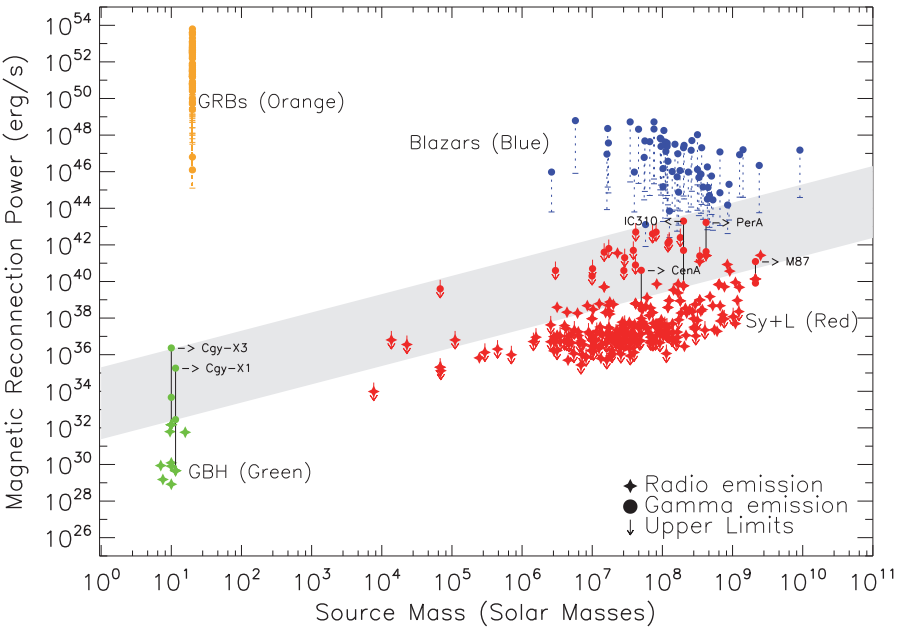


Fig. 1. Tubulent-driven magnetic power against black hole source mass compared to observed emissions of LLAGNs, GBHs, GRBs and blazars. In few cases, maximum and minimum values of gamma-ray luminosity are connected with a vertical black line that extends down to the radio emission. The inverted arrows for some sources indicate that the gamma-ray emission is an upper limit only. The gamma-ray emissions for blazars and GRBs are represented by blue and orange circles respectively. The vertical dashed lines correct the observed emissions by Doppler boosting effects (adopted from ref.[5]).

magnetic energy power is sufficient enough to produce gamma-ray emissions. The gamma-ray and radio emissions seem to be correlated and originating from the same core region around black holes and produced by the same acceleration mechanisms. However, the emissions of blazars and GRBs seem to originate along the jet rather than from the base of the jets. This is likely because the emissions from the core region of these sources are usually screened as the jets point towards our line of sight.

In another work,[6] we performed a semi-analytical comparison between black hole spin and fast magnetic reconnection power for accreting black holes based on fiducial astrophysical parameter space. Besides the rotational energy of the black hole, the fast reconnection power may play a similar role to power relativistic jets. Frame dragging effects of rotating black holes can lead to complicated magnetic structures and enhances the possibility of polarity inversion of the field lines leading to reconnection events and high energy emissions. The reconnection power is

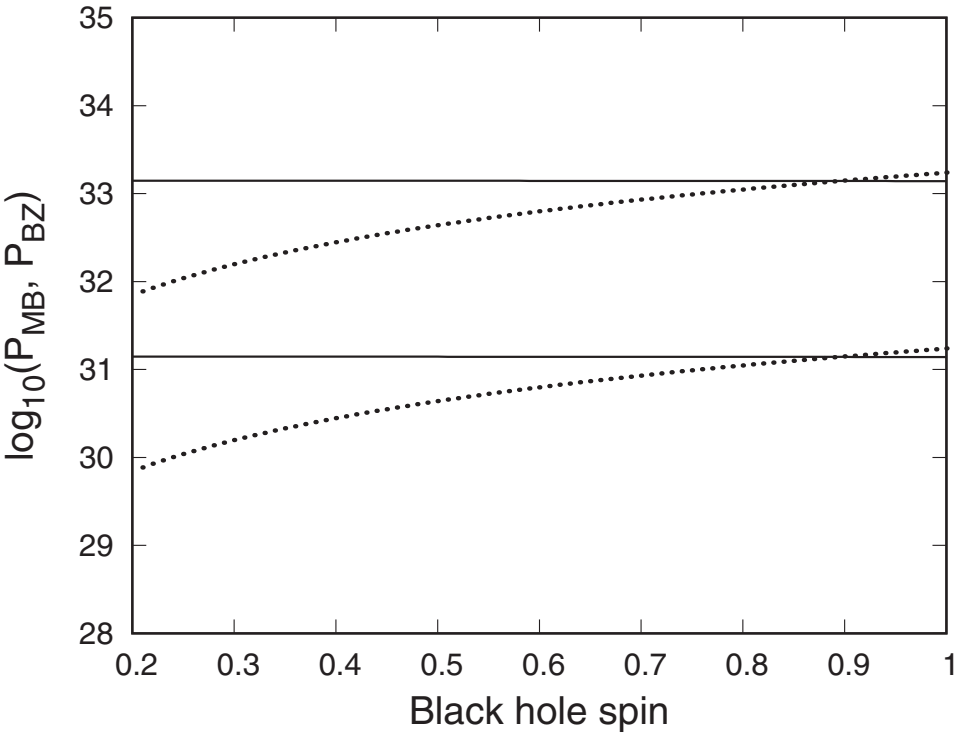


Fig. 2. Magnetic reconnection power P_{MB} and black hole spin power P_{BZ} ploted against black hole spin (adopted from ref.[6]).

a minimum close to the black hole and near the spin axis. It is likely to increase in magnitude with increase in the mass accretion rate. It may compete or even dominate the black hole spin power as shown in Fig. 2. It shows the maximum and minimum values for black hole spin and reconnection power as a function of black hole spin. The power available from the black hole spin and reconnection is comparable for rapidly rotating black holes and the same range of mass accretion rates. The black hole spin power obviously depends on the black hole rotation magnitude while the reconnection process seems to be independent of the black hole spin. The black hole spin-driven process gets suppressed in presence of reconnection. This may occur due to the destruction of large-scale magnetic fields following reconnection events. So state transitions from hard state to soft state can be triggered by this process.

3. Magnetic reconnection in relativistic jets

Using the three-dimensional relativistic magnetohydrodynamic code RAISHIN, we investigated the influence of radial density profile on the spatial development of the current-driven kink instability (CDKI) along magnetized rotating, relativistic jets.[7]

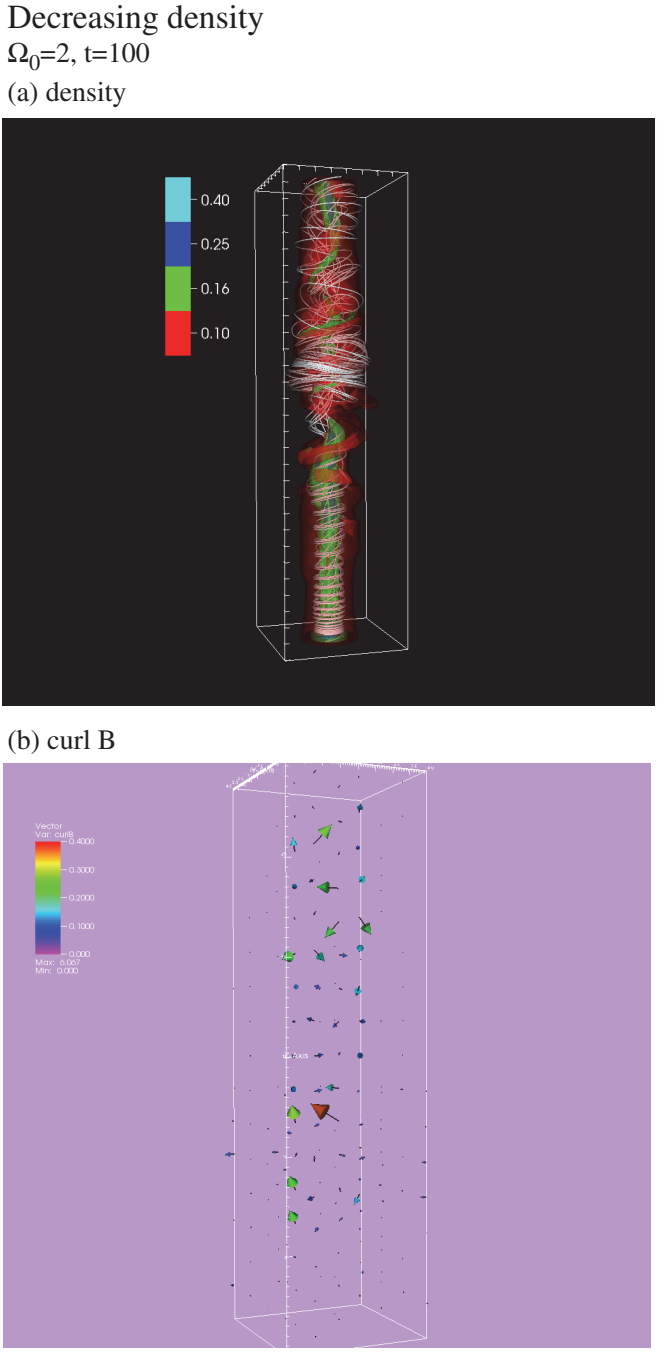


Fig. 3. Current driven kink instability and fast magnetic reconnection in a relativistic magnetized jet for decreasing density profile (heavier jet) with angular velocity amplitude of $\Omega_o = 2$. The upper panel shows 3D density isosurfaces, and bottom panel the locations of maximum current density as tracer of reconnection sites at t = 100 (adopted from ref.[7]).

We studied light as well as heavy jets with respect to the environment depending on the density profile. Different angular velocity amplitudes have been also tested. The results show the propagation of a helically kinked structure along the jet and a relatively stable configuration for the lighter jets. The jets appear to be collimated by the magnetic field and the flow is accelerated due to the conversion of electromagnetic into kinetic energy. We also identified the regions of high current density in filamentary current sheets, indicative of magnetic reconnection, which are associated with the kink unstable regions (Fig. 3) and correlated to the decrease of the sigma parameter of the flow. We discuss the implications of our findings for Poynting-flux dominated jets in connection with the magnetic reconnection process. We find that fast magnetic reconnection may be driven by the kink-instability turbulence and govern the transformation of magnetic into kinetic energy thus providing an efficient way to power and accelerate particles in AGN and gamma-ray-burst relativistic jets.

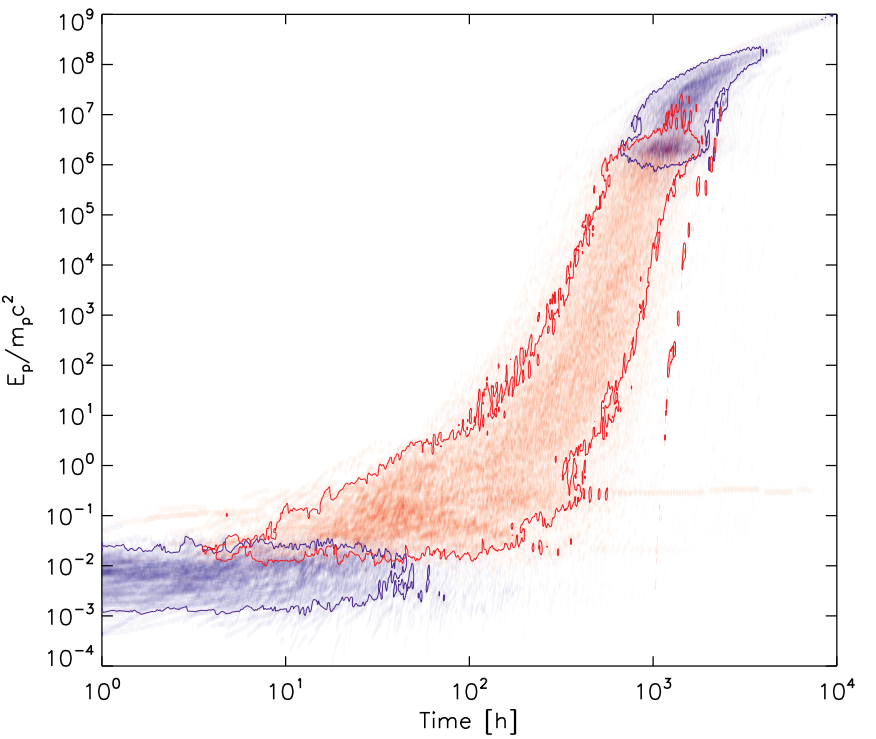


Fig. 4. Kinetic energy evolution, normalized by the proton rest mass energy, for the particles injected. Colours indicate which velocity component is being accelerated (red or blue for parallel or perpendicular component to the local magnetic field, respectively (adopted from ref.[8]).

Further, we also investigated the acceleration of particles injected in the jet subject to CDKI, which drives turbulence and fast magnetic reconnection.[8] Test protons injected in the nearly stationary snapshots of the jet, experience an exponential acceleration up to maximum energy. For a background magnetic field of B $\sim$ 0.1 G, this saturation energy is $\sim 10^{16}$ eV, while for B $\sim$ 10 G it is $\sim 10^{18}$ eV. The simulations also reveal a clear association of the accelerated particles with the regions of fast reconnection. In the early stages of the development of the nonlinear growth of CDKI in the jet, when there are still no sites of fast reconnection, injected particles are also efficiently accelerated, but by magnetic curvature drift in the wiggling jet spine. However, they have to be injected with initial energy much larger than that required for particles to accelerate in reconnection sites. Fig. 4 shows which component of the velocity of the particles is being predominantly accelerated (red for the parallel and blue for the perpendicular component to the direction of the local magnetic field). We note that in the exponential regime, there is a clear dominance of the parallel component, characterizing an effective electric field mostly parallel to the reconnection layers, as expected, though the stochastic nature of the whole process also allows for the acceleration of the perpendicular component. The slower drift acceleration regimes, both in the beginning and after the exponential growth regime, are dominated by the acceleration in the vertical direction. We also obtained from the simulations an acceleration time due to reconnection with a weak dependence on the particles energy E, $t_A \propto E^{0.1}$. The energy spectrum of the accelerated particles develops a high energy tail with a power-law index p $\sim$ -1.2 at the beginning of the acceleration, in agreement with earlier works. Our results provide an appropriate multi-dimensional framework for exploring this process in real systems and explain their complex emission patterns, especially in the very high energy bands and the associated neutrino emission recently detected in some blazars.

4. Summary and Conclusions

In this presentation, we discussed the role of magnetic reconnection in accretion-ejection phenomena around black holes based on results obtained in our theoretical and numerical studies. Our main conclusions can be summarized as follows:

1. Fast magnetic reconnection can play an important role in accretion/jet systems for particle acceleration, dissipation of magnetic energy, and conversion from magnetically dominated region to kinetically dominated region.
2. Power released by the fast magnetic reconnection can explain radio as well as gamma-ray emissions of XRBs and non-blazar AGNs as coming from the core around black holes with masses that span 10 orders of magnitude.
3. Black hole rotation and magnetic reconnection processes can drive jets of comparable power and even compete with each other.

4. The fast turbulent reconnection in magnetically dominated relativistic jets can be triggered by CDKI, possibly drive Fermi acceleration and lead to gamma-ray emissions as required by GRB and AGN blazar jets.

Acknowledgments

C.B.S. is supported by the National Natural Science Foundation of China under grant No. 12073021.

References

1. E. M. de Gouveia dal Pino and A. Lazarian, Production of the large scale superluminal ejections of the microquasar GRS 1915+105 by violent magnetic reconnection, *A&A*, **441**, 845 (2005).
2. G. Kowal, E. M. de Gouveia Dal Pino and A. Lazarian, Particle Acceleration in Turbulence and Weakly Stochastic Reconnection, *PRL*, **108**, 241102 (2012).
3. M. V. del Valle, E. M. de Gouveia Dal Pino and G. Kowal, Properties of the first-order Fermi acceleration in fast magnetic reconnection driven by turbulence in collisional magnetohydrodynamical flows, *MNRAS*, **463**, 4331 (2016).
4. L. H. S. Kadowaki, E. M. de Gouveia Dal Pino and C. B. Singh, The Role of Fast Magnetic Reconnection on the Radio and Gamma-ray Emission from the Nuclear Regions of Microquasars and Low Luminosity AGNs, *ApJ*, **802**, 113 (2015).
5. C. B. Singh, E. M. de Gouveia Dal Pino and L. H. S. Kadowaki, On the Role of Fast Magnetic Reconnection in Accreting Black Hole Sources, *ApJL*, **799**, L20 (2015).
6. C. B. Singh, D. Garofalo and E. M. de Gouveia Dal Pino, Magnetic reconnection and Blandford-Znajek process around rotating black holes, *MNRAS*, **478**, 5404 (2018).
7. C. B. Singh, Y. Mizuno and E. M. de Gouveia Dal Pino, Spatial Growth of Current-driven Instability in Relativistic Rotating Jets and the Search for Magnetic Reconnection, *ApJ*, **824**, 48 (2016).
8. T. E. Medina-Torrejón, E. M. de Gouveia Dal Pino, L. H. S. Kadowaki, G. Kowal, C. B. Singh and Y. Mizuno, Particle Acceleration by Relativistic Magnetic Reconnection Driven by Kink Instability Turbulence in Poynting Flux-Dominated Jets, *ApJ*, **908**, 193 (2021).

Exact fractal model of the universe and possible machine learning methods for the verification of the fractality

A. A. Kirillov, E. P. Savelova, P. O. Vladykina*

Department of Mathematical Modeling, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation

**E-mail: vladykinapo@yandex.ru*

We discuss the possibility that the topological structure of the Universe may possess fractal properties. Relic wormholes and their fractal distribution are predicted in a natural way by lattice quantum gravity models. This gives a new approach to some long-standing problems. Those are the nature of dark matter phenomena, the origin of Faber-Jackson and Tully-Fisher relations, and the observed deficit of baryons. We consider open Friedman model and construct an exact fractal model by means of a factorization of the space over a discrete subgroup of the group of motion. We derive some basic features of the resulting fractal space and discuss applications of machine learning methods for the verification of the fractal properties.

Keywords: Dark matter; wormholes; fractal distribution; baryon deficit.

1. Introduction

The nature of dark matter represents one of the most challenging problems of modern astrophysics. Particle physics suggests many possible candidates. The most promising are WIMPs (weakly interacting massive particles) which can be both new particles beyond the standard model, or particles composed from already known. However, such particles can hardly solve the problem. Indeed, experiments at LHC do not show the presence of particles beyond the standard model. Moreover, such particles should possess very exotic properties. Firstly, in the early Universe at the moment of recombination such particles have a rather high density but do not interact with baryons or radiation, otherwise they essentially change the observed CMB power spectrum of $\Delta T/T$. Secondly, in the modern Universe in centers of galaxies dark matter particles have much more modest densities but they should possess a sufficiently strong interaction with baryons to form the observed cored distribution of dark matter $\rho_{DM}(r) \sim const$ as $r \to 0$, where r is the distance to the center of a galaxy.[1–3] We recall that numerical investigations based on DM particles predict the presence of a cusp $\rho_{DM}(r) \sim 1/r$ [4,5] in centers of galaxies which is the unavoidable consequence of the potential nature of initial velocities and the fact that DM particles interact via gravity only.

In principle, such a difficulty of DM models based on particles may be overcome in some of modified theories of gravity[6] which gain the popularity during the last decade.[7,8] However, any modification should be approved first from particle physics standpoint and most of such modifications can be rejected already at laboratory

scales. In other words, most of such theories have pure academic interest and have no astrophysical applications.

One of the possible solutions of the problem is based on the fact that local topology of space may possess a rather complex or fractal structure.[9,10] It has two basic advantages. First, it does not require any modification of general relativity or the presence of additional dark matter component. And second, it gives the natural explanation to the empirical Tully-Fisher and Faber-Jackson relations[11,12] for rotation curves in galaxies. We point out that except the exotic scheme suggested by Milgrom,[13] where such relations were incorporated from the very beginning, the fractal topology gives the only natural explanation to such relations.[14]

The origin of fractal topological structure comes out directly from predictions of lattice quantum gravity models.[15–17] Indeed lattice models predict that at Planckian scales space-time has a foamy structure which possesses fractal properties. In particular, numerical investigations[17] show that spectral dimension of space-time runs from the value $D = 3/2$ at sub-Planckian scales to the value $D = 4$ on macroscopic scales. At very small scales the space-time foam can be described by a distribution of virtual wormholes.[18] At laboratory scales vacuum is stable which means that all virtual wormholes appear and disappear permanently. However if we believe that the early Universe underwent through an inflationary phase, we should expect that the exponential expansion very rapidly stretches all scales and some virtual wormholes do not decay. They temper and form initial conditions for the topological structure of space. The subsequent evolution of wormholes is the pure classical process. Spherically symmetric wormholes rapidly collapse, while toroidal and more complex wormholes may survive till the present days.[19,20] In other words, the present topological structure of space is an instant print of the space-time foam picture existing on Planck scales. Since the vacuum state possesses homogeneity and isotropy properties, the topology of our Universe inherits the same properties upon the freezing out on the inflationary stage.

The fractal structure of space can be described as follows. Consider an arbitrary point x in space and fix the geodesic distance R. Then the volume $V_x(R)$ of space which gets inside the geodesic ball, i.e., all points y which obey the inequality $\ell(y, x) < R$ (where $\ell(y, x)$ is the geodesic distance between x and y), scales with R as $V_x \sim R^d$, where $d = D - 1$. The homogeneity and isotropy of space requires that the function $V_x(R)$ and the dimension D do not depend on the initial point x.

The initial point x and omnidirectional geodesic lines, issued from the point x, define the extrapolated reference system which is used in astrophysics. The coordinate volume of such a system has always the dimension $D = 4$ which reflects the fact that locally space is Euclidean with the dimension $d = 3$ (i.e., R^3). We point out that the dependence on the redshift (cosmological evolution) and the choice of the background metric (open, flat, or closed model) somewhat changes the behavior of $V(R)$ and the definition of the dimension d, but rough features and the value $d = 3$ remains valid.

The situation changes when the space is filled with a distribution of wormholes. A part of geodesics go through wormhole throats and return to the same physical region of space or get into a remote region. Therefore, the behavior of $V(R)$ and the dimension d may change essentially. From the pure mathematical standpoint this signals to the fact that the extrapolated coordinate system stops to work (the same point may have multiple images), which requires to introduce additional atlases. In general the dimension evaluated in such a manner may either exceed or be less than the coordinate dimension $d = 3$. What kind of the dimension is realized for physical space depends on properties and the distribution of wormholes.

It is curious that the distribution of galaxies gives roughly the analogous picture. In a homogeneous space the number of galaxies reflects sufficiently well the behavior of the physical volume of space. The galaxy counts show the fractal behavior $N(R) \sim R^d$ with the dimension $d \approx 2$ up to the distances $R \sim 200Mpc$[21, 22] where $N(R)$ is the number of galaxies within the distance R. The maximal scale changes with the increase of precision. First estimates[23] for the scale upon which the distribution of galaxies switches to the homogeneity have the order $\sim 200Mpc$. However more recent observations of SDSS[24] have found the existence of wall-type structures (Sloan Great Wall, etc.) on scales up to $0.1R_H$, where R_H is the Hubble radius, see also.[25] Such super-large structures have posed a serious problem for the standard scenario of the structure formation. Recall that according to latter the largest scale on which inhomogeneities may develop is $300Mpc$. Moreover, such structures seem to be inconsistent with the cosmological principle. We consider such data as an indirect evidence for the fractal (or at least complex) topological structure of space and the existence of relic wormholes. The development of perturbations and the formation of structures in the presence of relic wormholes is a rather complex problem which requires the further investigation.

We point out that in the flat space the light when passing through a wormhole throat is too scattered and may contribute only to the diffuse background radiation.[26] Most likely the scattered signal is too dim to observe a separate galaxy through wormhole throats. The exception may be only gigantic throats compared to the void and larger size. This means that the observed dimension counted by galaxies is always less than three $d < 3$. In particular, this may also explain the observed deficit of baryons.[27]

The fractal topological structure can be easily obtained in open Friedman models by means of a specific factorization of the negative curvature (Lobachevsky) space over a discrete subgroup of the group of motion of space.[19, 20] We point out that the fractal distribution of baryonic matter does not necessarily indicate the existence of a fractal topological structure. However such a distribution requires specific initial conditions and, to be consistent with the observed homogeneity of the background, it requires the presence of dark matter which is dually distributed in space. The fractal topological structure, in turn, is the direct prediction of lattice quantum gravity models and, therefore, it has the most rigorous theoretical ground.

At first glance the basic disadvantage of the factorization is the presence of multiple images. Why do we not see images of the same galaxy? To find the answer is not difficult. First, our Universe has flat space. For small throats of wormholes (of the galaxy size or less) the additional images are very small and cannot be directly observed. Indeed, as it was demonstrated[30] in the flat model the intensity of an image I have the order $I \sim \frac{R^2}{L^2} I_0$, where R is the characteristic radius of the throat and L is the distance between the throat center and the original source and the ratio $\frac{R^2}{L^2} \leq 1$. It reaches the order of $\frac{R^2}{L^2} \sim 1$ only for very large throats. Moreover, the position of the image will be seen only in the second entrance into the wormhole and, therefore the image may have an arbitrary distance from the original. We point out that for fractal distribution over distances between entrances into the same wormhole the mean distance between entrances diverges.

Therefore, different images contribute only to the scattered background radiation. For gigantic wormholes all such images are too far from each other and we may see the same galaxy at different stages of its evolution. Therefore, it is impossible to recognize that we indeed see the same galaxy.

Thus the basic disadvantage is that the factorization requires the negative curvature space only, while the position of Doppler picks in the CMB spectrum favor for the flat space. In flat space such a factorization is impossible, while a fractal structure can be obtained by inserting fractal distribution of wormholes by onset or by means of deforming the metric of the factorized Lobachevsky space. We point out that such a deformation is not an unusual phenomenon. For example, if the inflation starts in an open model, then the factorization takes actually place at sub-Planck scales, while the subsequent particle production may change the value of the mean spatial curvature and produce flat or even closed space which depends on the rate of particle production.

In the flat space the presence of wormholes was shown to produce a specific modification of the standard Newton's law.[28,29] If we interpret this as dark matter phenomenon, then we find the rigid relation between distribution of visible and dark components in a galaxy in the form $\rho_{DM}(k) = b(k)\rho_{vis}(k)$, where $\rho(k)$ is the Fourier components for dark and visible matter densities and $b(k)$ corresponds to the correction which depends on the distribution of wormholes around the galaxy. This gives the direct method for the determining of the correction $b(k)$ from observations and the subsequent modeling of the distribution of relic wormholes.

If our Universe is filled with relic wormholes, then how can we recognize this? To give an answer on such a question is not so easy. The investigation of possible independent effects related to the scattering of radiation on wormholes has revealed that almost all effects are well hidden under analogous effects related to the ordinary matter.[30] Here we see the three most promising possibilities. The first possibility relates to recently observed unexpected circular radio objects[31] and the possible role of torus-shaped wormholes in their origin.[32] The second relates to the gravitational wave Astronomy. In particular, several groups have claimed evidence for repeating echoes in the LIGO/Virgo observations of binary black hole mergers.[33–36] And such

echoes can be generated by the near-horizon or other strong field effects related to black holes or wormholes.[37–39] We stress that the scattering of GWs by wormholes produces much stronger effects as compared to the scattering by ordinary matter objects, e.g., see discussions in,[40] which allows to disentangle such effects and detect wormholes.

The third and the most promising possibility lies in the application of machine leaning methods for detecting unusual structures produced by wormholes on the sky maps. The structures we should look for are the ring-type structures, which in general case have elliptical or even rectangular form. The latter may appear in regions where the mean curvature is negative, see Fig. 2. Besides, one should also look for multiple images of galaxies, etc..

2. Stable wormholes and factorization of the open model

As it was discussed above, relic wormholes are remnants of the inflationary stage. A-priori such wormholes may have an arbitrary form and geometric structure. Stable spherically symmetric wormholes cannot exist without an essential modification of general relativity. In general relativity static spherical wormholes require the presence of exotic matter which does not show up in physical systems and laboratory experiments. The exotic matter can be replaced by a proper modification of theory, e.g., see also a recent paper[41] and references therein, but any modification requires first a verification from particle physics point of view. Any deviation from general relativity is not detected yet, while quantum corrections may stabilize only wormholes of a Planck size.[42] Therefore, spherical wormholes collapse very soon upon the inflationary stage and form couples of primordial black holes.

Situation changes when we consider less symmetric configurations. It was demonstrated[19,20] that in open Friedman models a collection of stable wormholes can be constructed simply by means of a factorization of space over a discrete subgroup of the group of motions for the Lobachevsky space. The simplest wormhole here has the throat in the form of a torus and in such a scheme spherical wormholes do not appear at all. Such wormholes expand in the agreement with the expansion of space and, in other words, in the expanding co-moving reference frame they are static. This does not mean that they cannot collapse as the spherical configurations do. In particular, the development of density perturbations may enforce to collapse a whole region of space which contains such a wormhole throat. In what follows we briefly describe the factorization scheme.

2.1. *Factorization of the open Friedman model*

Consider the open Friedman model with the metric

$$ds^2 = dt^2 - a^2(t)dl^2, \tag{1}$$

where the space-like metic corresponds to the Lobachevsky space

$$dl^2 = d\chi^2 + \sinh^2\chi\left(\sin^2\theta d\varphi^2 + d\theta^2\right). \tag{2}$$

For the sake of illustration and simplicity we demonstrate the factorization for 2D sections of the space which correspond to a plane $d\varphi = 0$. Such a plane corresponds to two values of the polar angle $\varphi = \varphi_0$ and $\varphi = \varphi_0 + \pi$ which formally can be covered by the angles $\theta \in [0, 2\pi]$ where the range $\theta \in [\pi, 2\pi]$ corresponds to the half-plane $\varphi = \varphi_0 + \pi$. Using the coordinate transformation as

$$z = x + iy = i\frac{1-w}{1+w} \tag{3}$$

where $w = \tanh\frac{\chi}{2}e^{i\theta}$, we get the realization of the section in the form of the upper complex half-plane $Im\, z > 0$ which is the Klein model of the Lobachevsky plane $dl^2 = \frac{dx^2+dy^2}{y^2}$. Consider two particular geodesic lines on the plane which correspond to semicircles $|z| = 1/R$ and $|z| = R$ with $R > 1$. Then the factorization can be determined by the map $z' = R^2 z$ which represents one of possible motions of the plane. The factorization means that any two points of the plane represent the same point, if they are connected by the relation

$$z_n = R^{2n} z,\ n = 0, \pm 1, \pm 2, ... \tag{4}$$

The above factorization corresponds to the cut along those two geodesic lines and subsequent gluing along them. The resulting space is a wormhole which connects two independent Lobachevsky planes (regions E_+ and E_-, see Fig. 1).

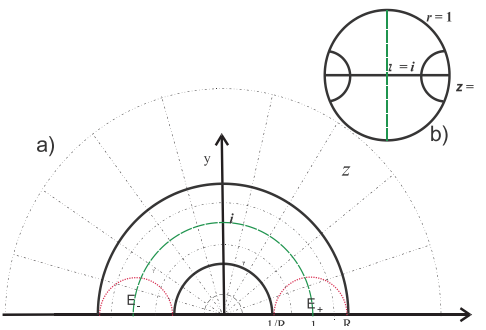


Fig. 1. a) The Lobachevsky plane. Dashed lines give the polar frame. The stripe between the two geodesic lines $r = 1/R$ and $r = R$ corresponds to the wormhole region. Regions below red dashed geodesics (semi-circles) correspond to two unrestricted Lobschevsky spaces $E_\pm$. The point $y = i$ corresponds to the center point of the wormhole throat ($\chi = 0$). The line $y = 0$ corresponds to infinity $\chi \to \infty$ for the plane. b) The same plane in coordinates $w = re^{i\theta}$, where $r = \tanh\frac{\chi}{2}$.

To understand how such a factorization looks in three dimensions let us return to the initial coordinates χ, θ, and φ. The geodesic lines $|z| = R$ and $|z| = 1/R$ are determined by

$$\tan^2\theta = \frac{\cosh^2\chi - \cosh^2\chi_0}{\cosh^2\chi \sinh^2\chi_0}, \tag{5}$$

where $R = e^{\chi_0}$ and χ_0 gives the shortest distance to the geodesic lines from the center point $\chi = 0$. The above equation corresponds to an arbitrary section $d\varphi = 0$ and describes two lines which can be seen in opposite directions $\theta = 0$ and $\theta = \pi$. Here the point $\chi = 0$ corresponds to the center of the throat (the point $z = i$ on Fig. 1). Using the symmetry with respect to the change of φ we find that in 3D space the above equation determines two semi-spheres (rigorously speaking they become semi-spheres in terms of the radial variable $r = \tanh\frac{\chi}{2}$ in which the infinity $\chi \to \infty$ becomes the sphere $r = 1$), while the factorization has the same form determined by (3) and (4). In this case the throat is not restricted. It is important that from the central point $\chi = 0$ such semi-spheres occupy the solid angles $\Delta\Omega = 2\pi\Delta\theta$ where $\Delta\theta$ can be found from (5) in the limit $\chi \to \infty$ which gives the asymptotic angles $\Delta\theta = 2\arctan(1/\sinh\chi_0)$. To get a restricted throat we have to place two more analogous semi-spheres e.g., in the orthogonal to $\theta = 0, \pi$ directions at a distance χ_1, see fig. 2. Moreover, the distances $\chi_{0,1}$ should be chosen in such a way that all

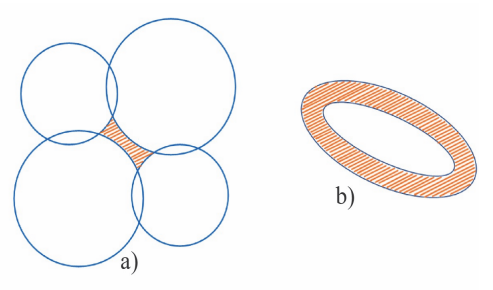


Fig. 2. a) The form of a throat as it is seen on sky in the open model. It represents the colored curved rectangle. Regions behind the circles are copies of the physical region. All additional images lay within the circles. b) The form of the throat as it is seen in flat models upon the deformation of the metric from the negative curvature space. The throat has the torus-like form and only part of the throat surface is seen. Additional images lay inside the surface of the torus.

four semi-spheres restrict a finite minimal section. It is also important that every semi-sphere envelopes a non-physical space, since every point within it represents an image of a point from the corresponding physical space.

Now let us consider how such a wormhole looks for a remote from the center of the throat observer. To this end we should simply shift the central point $\chi = 0$ from the middle of the throat. From qualitative standpoint the picture remains the same, every semi-sphere is determined by the same equation (5) but the distance to it χ_0 changes. Four such semi-spheres group and if they restrict a finite solid angle they form a remote wormhole. The angle they occupy on sky is $\Delta\theta \sim 2\exp(-\chi_0)$ and the throats (minimal sections) are seen as curved rectangles. We point out that from the topological standpoint throats are tori but in open model they will be seen on sky as rectangles. Such wormholes restore the torus-like form of throats only in

the case when the space is additionally deformed or curved in such a way that the mean curvature becomes zero or positive one.

Above we have described the factorization which forms a wormhole that connects two independent (connected only through the throat) Lobachevsky spaces. To get a wormhole whose both entrances lay in the same region requires one more analogous map as described in.[19] In this case we get couples of such throats (groups of semi-spheres) on sky. Now to get a pure fractal picture we may construct a Cantor set on sky by placing such groups of semi-spheres of smaller and smaller solid angles.

2.2. *Stability of wormholes*

The factorization described previously requires the negative curvature space which corresponds to an open Friedman model, while the position of Doppler picks in the power CMB spectrum favor for the flat space. In flat space such a factorization is impossible. However a fractal structure can be obtained by inserting fractal distribution of wormholes or by deforming the metric of a factorized Lobachevsky space. The deformation of the Lobachevsky space can be obtained rather naturally during the inflationary stage. Indeed, if the space had initially a negative curvature, then during the inflation stage production of wormholes and particles took place. Therefore, particle production process could essentially change the value of spatial curvature and essentially deform the picture described above. In this sense we may expect to see a distribution of analogous wormholes even in flat space. We point out that contrary to the open model in flat space upon the deformation throat sections will look as torus- like shapes. This poses the problem how such wormholes may evolve in flat space. Spherical wormholes are unstable and collapse very quickly. The stability of torus-like wormholes is not investigated yet. There are only some encouraging results. For example, the existence and stability of static cylindrical wormholes without exotic matter was demonstrated recently in Refs.[43, 44] We point out that the toroidal wormhole reduces to the cylindrical one in the limit when on of radii tends to infinity. Though the rigorous investigation of the behavior of toroidal wormholes during the cosmological evolution still is not completed, we may expect that such wormholes collapse much slower, than spherical wormholes do and may survive till the present days. As a result we get open or flat Friedman model filled with toroidal wormholes.

When we consider wormhole throats as objects of a finite size (e.g., they occupy a finite solid angle on sky) we may average out over possible random orientations of the throats in space. We point out that there are no strong physical arguments in favor for some particular characteristic size of wormholes, moreover, fractal structure of space-time at Planck scales predicts analogous fractal distribution over the throat sizes. In the case of the statistical averaging over possible orientations the spherical symmetry of the wormhole configuration restores and, therefore, estimates based on spherical wormholes should give correct predictions, at least to the leading order. In what follows we discuss the open Friedman model with almost zero value of the

space curvature where wormholes can be considered simply as quasi-static objects expanding only due to the scale factor $a(t)$. The cosmological evolution of $a(t)$ follows the standard Friedman equations

$$H^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho,$$

where $H = \dot{a}/a$ is the Hubble constant, $k = -1$, and ρ is the matter density (baryons plus radiation). The metric has the standard form $ds^2 = dt^2 - a^2(t)dl^2$, where the space-like part dl^2 corresponds to the Lobachevsky space.

The behavior of small perturbations in such a model is determined by the standard Lifshitz theory[45] with one exception. The basic equations for perturbations have local character and, therefore, do not depend on the topological structure of space. However possible solutions do. In particular, in comparison to the ordinary Lobachevsky space the space with wormholes represents a space with a specific factorization. For example in flat space the factorization over translations in one of directions gives the space which is simply the cylinder and one of coordinates becomes a periodic one. This means that corresponding modes are classified by a discrete wave numbers instead of continues ones and, therefore, some modes are absent. Analogously, the factorization of the open model selects only particular solutions among all possible solutions for the unrestricted space. Therefore, from qualitative and quantitative standpoint the evolution of perturbations will show essential deviations from the behavior in standard Friedman models. Analogous deviations are indeed observed and are interpreted as the presence of dark matter, or a modification of general relativity. According to Ref.[17] the Hausdorff dimension of space should be $1 < d \leq 3$. At laboratory scales we see that $d = 3$ and no deviations from the standard Newton's law are observed. The number of galaxies counts show that starting from galactic scales ($R \sim 5Kpc$) at least up to $200Mpc$ the Hausdorff dimension is $d \approx 2$ and at those scales we observe the phenomenon of dark matter.

Indeed, the simplest estimate for the Newton's law gives $F \propto \frac{GM}{S(R)}$, where M is the mass within the volume $V(R)$ (ball of the radius R) and $S(R) = \frac{dV(R)}{dR}$ is the square of the sphere surrounding the volume $V(R)$. In the case of a fractal law $V \sim R^d$ and $S \sim R^{d-1}$. The value $d \approx 2$ is in the very good agreement with observations of dark matter in the range of scales $5Kpc < R < 200Mpc$. The value $d \approx 1$ gives too strong the gravitational bind and such structures should decouple the cosmological expansion very early, or it even should be suppressed during the inflationary stage.

From one standpoint, stable toroidal wormholes do surely exist in negative curvature spaces (in the open model), while in the flat model their stability requires additional study. From other point, the position of the Doppler picks in the CMB ($\Delta T/T$) spectrum points out to the fact that our Universe is very close to the flat space, which means that the matter density is close to the critical value. Here, however, one should be especially careful in making estimates. First, the positions of picks are evaluated on the basis of the standard Friedman models when the possible

factorization of the space is not taken into account (i.e., the possible presence of wormholes). Secondly, the presence of fluctuations $\Delta T/T \sim 10^{-5}$ means that at the moment of recombination analogous fluctuations are present in the metric and matter density $\Delta\Omega = \Omega - 1 \sim 10^{-5}$, where $\Omega = \rho/\rho_{cr}$ and $\rho_{cr} = 3H^2/(8\pi G)$ is the critical density. In other words, the space is split in regions of a positive ($\Omega > 1$, $k/a^2 \approx 10^{-5}H^2$) and negative ($\Omega < 1$, $k/a^2 \approx -10^{-5}H^2$) curvature. If in regions where $\Omega > 1$ the presence of stable wormholes requires the additional study, in the regions where $\Omega < 1$ stable wormholes do exist. Moreover, in the fractal Universe the observed mean value of the spatial curvature is somewhat different from the actual locall value. It contains the additional factor $V_{ph}(R)/V_{coor}(R)$, where $V_{ph}(R) \sim R^{D-1}$ is the actual physical volume of space, while $V_{coor}(R) \sim R^3$ is the coordinate, or extrapolated volume. In this estimate one should take the Hubble radius $R = R_H$, or the radius of the sphere of last scattering. In accordance to Ref.[6] the estimate $V_{ph}(R)/V_{coor}(R) \approx 10^{-5}$ is in the agreement with observations which means that at the moment of recombination the local value of the spatial curvature and the Hubble constant may have the same order ($k/a_{rec}^2 \approx H_{rec}^2$). We point out that the same value 10^{-5} characterizes the portion of actually observed baryons with respect to the value predicted by nucleosynthesis. Here we assume estimates of the number of baryons based only on counts of galaxies and discrete sources, while the diffuse background (e.g., the X-ray background) gives much higher estimate for such a ratio ~ 0.4.[46–48] In fractal picture the diffuse background does not rigorously reflects the actual number of baryons. It reflects merely the mean homogeneity and isotropy of space and the local density of baryons, while the actual number of baryons is proportional to the physical volume of space $N(R) \sim V_{ph}(R) \ll V_{coor}(R)$. In other words, rigorous estimates require further investigation.

3. Fractal distribution of wormholes and modification of Newton's law

The fractality of the topological structure of space requires the analogous fractal distribution of wormholes. At first glance the fractality is not consistent with the observed homogeneity and isotropy of the Universe. This however is not so, e.g., the diffuse background which appears due to the scattering on wormholes remains homogeneous and isotropic.[30] The basic feature of the fractal distribution is a scale - dependence of different mean values, such as mean density of wormhole throats n, the mean distance between entrances $\langle X \rangle$, the mean radius of the throat $\langle R \rangle$, etc.

Consider the simplest model suggested by us in Ref.,[28] which explicitly demonstrates that non-trivial topology can model effects of dark matter. The model is based on spherical wormholes, while actual stable wormholes should have torus like throats. We assume that throats are averaged over orientations then the spherical symmetry restores and such wormholes can be considered in the leading order as spherical ones.

In the case of weak gravitational fields and potential perturbations the Einstein's equations reduce to the Laplace equation for Newton's potential ($R_0^0 - \frac{1}{2}R = 8\pi T_0^0$)

$$\frac{1}{a^2}\Delta\phi = 4\pi G\left(\delta\rho + \frac{3}{c^2}\delta P\right),$$

where a is the scale factor, $\delta\rho$ and δP are perturbations in density and pressure respectively. The behavior of perturbations are determined by the Green function (Newton's law)

$$\Delta G\left(x, x'\right) = 4\pi\delta\left(x - x'\right).$$

In the Euclidean space the Newton's law gives $G_0 = -1/r$ (in the Fourier representation it gives $G_0 = -4\pi/k^2$). In the presence of wormholes the Green function changes.[28] It is convenient to describe corrections which appear due to wormholes by the presence of additional sources (fictitious sources). Then formally we may think that the space remains Euclidean, while proper periodic conditions on throats are accounted by the bias of sources

$$\Delta G\left(x, x'\right) = 4\pi\left[\delta\left(x - x'\right) + b\left(x - x'\right)\right].$$

Here the bias function $b\left(x\right)$ describes corrections to the Newton's law which appear due to non-trivial topology of space. Indeed the presence of the bias $b(x - x')$ gives rise to the transformation

$$\frac{1}{r^2} \to \frac{2}{\pi}\int_0^\infty [1 + b(k)]\frac{\sin(kr) - kr\cos(kr)}{kr^2}dk.$$

The asymptotically flat rotation curves in galaxies require that the correction to the Newton's potential should have the logarithmic behavior, i.e. the gravitational acceleration should switch from $1/r^2$ to $1/r$. The above expression implies that at galaxy scales the bias behaves as $b(k) \sim 1/k^\alpha$ with $\alpha \sim 1$ and $b(x-x') \sim 1/(x-x')^2$. When considering a galaxy with surface mass density distribution $\rho_L(x)$ the bias $b(x - x')$ forms an effective dark matter halo in the form

$$\rho_{DM}(x) = \int b(x - x')\rho_L(x')d^3x'.$$

In particular, for $b(k) \sim 1/k^\alpha$ and an infinitely thin disc $\rho_L(x) = \sigma e^{-r/R_D}\delta(z)$ this expression gives the pseudo-isothermal halo $\rho_{DM} = \rho_0 R_C^2/(R_C^2 + r^2)$, where R_C is the radius of the core, e.g., see.[6,14]

For a homogeneous distribution of wormholes the bias function $b\left(x\right)$ was first considered in[28] and more general case was constructed in.[18] The bias is expressed via the distribution of wormholes which in the homogeneous and isotropic case is a function $F = nf\left(X, R\right)$, where n is the density of wormhole throats in space, X is the distance between entrances into a wormhole, and R is the throat radius. In the Fourier representation ($b(k) = \int b\left(r\right)e^{-i\mathbf{kr}}d^3r$) it has the form[29]

$$b\left(k\right) = n\overline{R}\left(k\right)\frac{4\pi}{k^2}\left(\nu\left(k\right) - 1\right),$$

where we define

$$\overline{R}\,(k) = \frac{1}{k}\int \widetilde{f}(0,R)\sin(kR)\,dR = -\frac{1}{k}Im\int e^{-ikR}w(R)dR, \tag{6}$$

$$\nu\,(k) = \frac{1}{k\overline{R}\,(k)}\int \widetilde{f}(k,R)\sin(kR)\,dR, \tag{7}$$

and $\widetilde{f}(k,R) = \int f\,(X,R)\,e^{-i\mathbf{kX}}d^3X$ is the Fourier transform for the distribution function over distances. Here $k = |\mathbf{k}|$ and the dependence of the above functions only on k reflects the isotropy of the distribution f. The physical meaning of these two functions $\overline{R}\,(k)$ and $\nu\,(k)$ become quite clear in the limit $kR \ll 1$, when throat radii assumed to be small as compared to the characteristic scales. Then we may set $\sin(kR) \approx kR$ and we find $\overline{R} \approx \int Rw(R)dR = \int Rf\,(X,R)\,d^3X dR$ which is simply the mean value of the throat radius. We point out that $\overline{R}\,(k)$ is proportional to the imaginary part of the characteristic function for the distribution over radii $w(R) = \widetilde{f}(0,R)$ (6) (which is $w(k) = \int e^{-ikR}w(R)dR$). Analogously the function $\nu(k) \approx (1/\overline{R})\int R\widetilde{f}(k,R)dR$ is the characteristic function for the distribution over the distances between throat entrances. In the most general case the effective radius $\overline{R}\,(k)$ has the dependence on scales and it admits both signs. Thus, the presence of a distribution of wormholes gives rise to the specific modification of the Newton's law and the true Green function becomes

$$G(k) = -\frac{4\pi}{k^2}\,(1 + b(k))\,.$$

Consider the simplest particular situation when all distances between throats have the same value X_0 and all throat radii be R_0, then we find the homogeneous and isotropic distribution in the form

$$f(X,R) = \frac{1}{4\pi X_0^2}\delta(|X| - X_0)\delta(R - R_0),$$

and the Fourier transform is

$$\widetilde{f}(k,R) = \delta(R - R_0)\frac{\sin kX_0}{kX_0}.$$

The effective radius and the characteristic function which correspond to such a distribution are

$$\overline{R}\,(k) = R_0\frac{\sin(kR_0)}{kR_0}, \quad \nu\,(k) = \frac{\sin kX_0}{kX_0},$$

and the bias is

$$b\,(k) = -\frac{4\pi}{k^2}nR_0\frac{\sin(kR_0)}{kR_0}\left(1 - \frac{\sin kX_0}{kX_0}\right).$$

We see that the bias $b(k)$ (and, therefore, corrections to the Newton's potential) alternates the sign with the change of scales. It is completely determined by three typical scales related to wormhole distribution, which are the density $n \sim 1/L_w^3$, the typical throat radius R_0, and the typical distance between entrances into throats X_0.

In the general case (of the isotropic and homogeneous distributions) the above expressions can be rewritten as follows

$$\overline{R}\left(k\right)=\frac{\left\langle \sin\left(kR\right)\right\rangle }{k},\quad \nu\left(k\right)=\frac{1}{\overline{R}\left(k\right)}\left\langle \frac{\sin\left(kR\right)}{k}\frac{\sin kX}{kX}\right\rangle ,$$

where the mean value is determined as $\langle g(X,R)\rangle=\int g(X,R)f(X,R)d^{3}XdR$. For normal distributions, when all momenta are finite, we may expand the above expressions in series and find

$$b\left(k\right)=\lambda\sum_{N=1}^{\infty}\left(-1\right)^{N}\left(kL_{N}\right)^{2(N-1)}\approx-\lambda\left(1-\left(L_{2}k\right)^{2}+\left(L_{3}k\right)^{4}+...\right), \qquad (8)$$

where $\lambda=\frac{4\pi n}{6}\left\langle RX^{2}\right\rangle$ and the characteristic scales L_N are determined via mean values as follows

$$\left(L_{N}\right)^{2(N-1)}=\frac{6}{\left\langle RX^{2}\right\rangle }\sum_{l=0}^{N-1}\frac{\left\langle R^{2l+1}X^{2(N-l)}\right\rangle }{\left(2l+1\right)!\left(2\left(N-l\right)+1\right)!}.$$

The first term here determines the re-normalization of the gravitational constant, while next terms determine higher order corrections. The above analysis shows that using only normal distributions it is hardly possible to get the desirable form of corrections (observed in galaxies). This is confirmed by estimates presented in Introduction. Indeed, Gaussian distribution of wormholes is described by three typical scales those are $L_w\sim n^{-1/3}$ which characterizes the absolute value of corrections, R_0, and X_0. Starting from a typical scale $R>X_0$ $V_{phys}(R)\simeq V_{Coor}(R)$ and all corrections reduce to a some renormalization of gravitational constant, while at smaller intermediate scales ($R_0<R<X_0$) corrections have more complex behavior (partial screening, e.g., see[28]). This indicates that we should consider more general fractal distributions.

In the case of fractal distributions such an expansion does not work, since all momenta may diverge, while the density of wormholes n is badly determined (its value also depends on scales). In this case we should directly work with characteristic functions $w(k)$ ($\overline{R}\left(k\right)=-Im\, w(k)/k$) and $\nu(k)$. Indeed, consider the simplest example when $w(k)$ and $\nu(k)$ are independent. Then every function can be expressed as

$$\nu(k)=\exp\left\{-\left(kL_{1}\right)^{\alpha_{1}}\left(1-i\beta_{1}\tan\left(\frac{\alpha_{1}\pi}{2}\right)\right)\right\}, \qquad (9)$$

where $\beta_1\in[-1,1]$ determines an asymmetry of the distribution e.g., see Ref.[49] And analogously

$$w(k)=\exp\left\{-\left(kL_{2}\right)^{\alpha_{2}}\left(1-i\beta_{2}\tan\left(\frac{\alpha_{2}\pi}{2}\right)\right)\right\}, \qquad (10)$$

with its own parameters L, β, and α. The last equation defines the effective radius as

$$\overline{R}\,(k) = -Im\,\frac{w(k)}{k} = -L_2\frac{\exp\left(-\left(kL_2\right)^{\alpha_2}\right)}{kL_2}\sin\left[\left(kL_2\right)^{\alpha_2}\beta_2\tan\left(\frac{\alpha_2\pi}{2}\right)\right]. \qquad (11)$$

In the case $\alpha = 2$ the above characteristic functions corresponds to the standard Gaussian distribution, while for $\alpha < 2$ the dispersion which corresponds to such a distribution diverges as

$$\sigma = \frac{d^2}{dk^2}\ln\nu\,(k)\Big|_{k=0} \propto k^{\alpha-2} \to \infty$$

as well as all higher momenta. The parameters α_i relates to the Hausdorff dimension of space D and it characterizes the fractal structure of $f(X,R)$. To avoid misunderstanding we also point out that the scales L_i do not characterize the mean value between entrances into throats, or the mean radius, since for fractal distributions such moments diverge. In such a case the decomposition in series presented above (8) does not work. In this case we should expand directly the characteristic functions (9), (11) themselves. In the region of sufficiently small scales $kL_i \gg 1$ (i.e., at very small distances) the function $b(k)$ quickly oscillates and vanishes $b(k) \to 0$. In this region of scales the standard Newton's law does work. At sufficiently big scales, i.e., in the limit $kL_i \ll 1$, we find the first leading term as

$$b\,(k) \approx 4\pi\chi\frac{\left(kL_2\right)^{\alpha_2-1}\left(kL_1\right)^{\alpha_1}}{\left(kL_3\right)^2} + ... = 4\pi\chi\left(k\widetilde{L}\right)^{\alpha_1+\alpha_2-3} + ...\,, \qquad (12)$$

where $L_3 = 1/\sqrt{nL_2}$ and

$$\chi = \beta_2\tan\left(\frac{\alpha_2\pi}{2}\right)\left(1 - i\beta_1\tan\left(\frac{\alpha_1\pi}{2}\right)\right).$$

Corrections to the empirical Green function $b(k)$ can be restored directly from observations and by relations (6), (7) they determine the distribution of wormholes. In the present paper we do not analyze any actual data, since it requires an independent study. However some rough estimates can be obtained. Indeed, the first correction for the empirical Green function which follows from rotation curves[14] behaves roughly as $G_{emp} = -\frac{4\pi}{k^2}(1 + (k\widetilde{L})^{-\alpha})$ with $\alpha \simeq 1$ which determines parameters in (12) as $\chi \simeq 1$ ($\beta_1 = 0$), $\alpha_1 + \alpha_2 \simeq 2$, and $\widetilde{L} \sim 5Kpc$. As it was discussed in Ref.[14] the value of the parameter $\widetilde{L}$ may somewhat change from a galaxy to a galaxy which reflects the presence of initial inhomogeneities of the metric at the moment of recombination.

As it was demonstrated in[14] the bias in the form (12) and the homogeneity of total matter (baryon plus dark matter components)[6] straightforwardly lead to the origin of Tully-Fisher and Faber-Jackson relations. Therefore, this can be also considered as indirect indication of the fractal topological structure of space. Moreover, the recent SDSS data[24,25] show the traces of the fractal distribution of baryons up to scales $\sim 0.1R_H$.

4. Concluding remarks

As concluding remarks we point out to basic results of this paper. First, we have demonstrated that the most natural origin of dark matter phenomena relates to the existence of locally complex or fractal topological structure of our Universe. At extremely small (sub-Planckian) distances such a fractal structure is predicted by lattice quantum gravity models. During the inflationary period the topological fractal structure tempers and forms the initial conditions for the actual structure of space in Friedman models. One may imagine such a process as an instant enormous increase of a small (of the Planckian order) portion of space filled with virtual wormholes up to macroscopic scales. The stretched wormholes become real and form the initial topological structure, while their subsequent evolution is already governed by classical gravity. Spherical wormholes collapse and form primordial black holes, while torus-like wormholes may survive till the present days.

Secondly, the fractal structure is consistent with the apparent homogeneity and isotropy of space. The simplest rigorous model is the Lobachevsky space factorized over a discrete subgroup of the group of motion. The resulting space has a constant spatial curvature and it locally cannot be distinguished from Friedman model. However it is not a homogeneous space any more and in general it possesses a complex topological structure. The flat model can be obtained by a deformation (due to particle production) of the factorized Lobachevsky space. In this case the fractal topological structure will correspond to a specific distribution of relic wormholes which possesses fractal properties.

And finally, we have demonstrated that when the distribution of wormholes possesses fractal properties, corrections to the standard Newton's law have the necessary form to imitate carefully the presence of dark matter in galaxies. The empirical Green function restored from the analysis of rotation curves in galaxies allows directly restrict some of parameters of the fractal distribution of wormholes.

The rigorous verification of the fractal topological structure requires the use of machine leaning methods. Such methods allow for detecting unusual structures produced by wormholes on the sky maps. Those are the ring-type structures with a specific temperature profile, e.g., if a wormhole throat rotates, then due to the kinematic Sunaev-Zeldovich effect this produces a specific profile of T in CMB. In general case rings may have elliptical or rectangular form. Besides, the fractal structure predicts that galaxies may have multiple images given at different moments of their evolution. Such features cannot be detected without the use of machine leaning methods.

References

1. de Blok, W. J. G., Bosma A. *Astron. and Astrophys.* **385**, 816 (2002).
2. Gentile G., et al. *Mon. Not. RAS.* **351**, 903 (2004).
3. Salucci P., *Astron. Astrophys. Rev.* **27** 2 (2019).
4. Navarro J. F., Frenk C. S., White S. D. M. *The Astrophys. J.* **462** 563 (1996).

5. Diemand J., et. al. *Mon. Not. RAS.* **364** 665 (2005).
6. Kirillov A.A. *Phys. Lett. B* **632** 453 (2006).
7. Amendola L., et. al. *Living Rev. Relativ.* **21** 2 (2018).
8. Harko, T., & Lobo, F. Extensions of $f(R)$ Gravity: Curvature-Matter Couplings and Hybrid Metric-Palatini Theory. (Cambridge Monographs on Math. Phys., (2018)).
9. Kirillov A. A. and Savelova E. P. *Gravit. Cosmol.* **14** 256 (2008).
10. Kirillov A.A., Turaev D. *Phys. Lett. B* **656** 1 (2007).
11. Tully R.B., Fisher J.R., *Astron. and Astrophys.* **54** 661 (1977).
12. Faber S.M. and Jackson R.E., *Astrophys. J.* **206** 668 (1976).
13. Milgrom M. *Astrophys. J.* **270**, 365 (1983).
14. Kirillov A.A., Turaev D. *Mon. Not. RAS. L* **371**, 31 (2006).
15. Knizhnik V., Polyakov A., Zamolodchikov A. *Mod. Phys. Lett. A* **3** 819 (1988).
16. Kawai H., Kawamoo N., Mogami T., Watabiki Y. *Phys. Lett. B* **306** 19 (1993).
17. Ambjorn J., Jurkiewicz J., Loll R. *Phys. Rev. Lett.* **95**, 171301 (2005).
18. Savelova E.P. *Grav. Cosmol.* **21**, 48 (2015).
19. Kirillov A.A., Savelova E.P. *Int. J. Mod. Phys. D* **25**, 1650075 (2016).
20. Kirillov A.A., Savelova E.P. *Eur. Phys. J. C* **80**, 45 (2020).
21. Coleman P.H., Pietronero L. *Phys. Rep.* **213** 311 (1992).
22. Baryshev Yu. Teerikorpi P. *Discovery of Cosmic Fractals.* WS, (2002), Pages: 408.
23. Geller M. J. Huchra J. P. *Science* **246**, 897 (1989).
24. SDSS collaboration. *Astrophys. Journ. Suppl.*, **233** 25 (2017).
25. Horvath I. et al. *Astron. & Astrophys.*, **584** , A48 (2015).
26. Planck collaboration, Planck 2018 results - IV. *Astron. & Astrophys.* **641**, A4 (2020).
27. Fukugitai M., Hogan C.J., Peebles P.J.E. *Astrophys. J.* **503**, 518 (1998).
28. Kirillov A.A., Savelova E.P. *Phys. Lett. B* **660**, 93 (2008).
29. Kirillov A. A. and Savelova E. P. *Gravit. Cosmol.* **24**, 337 (2018).
30. Kirillov A. A. and Savelova E. P. *Universe* **4**, 35 (2018).
31. Norris, et al *Publications of the Astron. Soc. of Australia*, **38**, E003 (2021).
32. Kirillov A.A., Savelova E.P. *Eur. Phys. J. C* **80**, 810 (2020).
33. Abedi J., Dykaar H. and Afshordi N., *Phys. Rev. D,* **96**, 082004 (2017).
34. Conklin R. S., Holdom B. and Ren J., *Phys. Rev. D*, *98*, 044021 (2018).
35. Westerweck J. et al., *Phys. Rev. D*, **97**, 124037 (2018).
36. Abedi J. and Afshordi N., *JCAP*, **11**, 010 (2019).
37. Cardoso V, Franzin E, and Pani P, *Phys. Rev. Lett.* **116**, 171101 (2016).
38. Cardoso V., et al., *Phys. Rev. D*, **94**, 084031 (2016).
39. Cardoso V. and Pani P., *Living Rev. Rel.*, **22** 4 (2019).
40. Kirillov A.A., Savelova E.P., Lecian O.M. *Eur. Phys. J. C* **81** 263 (2021).
41. De Falco, V., Battista, E., Capozziello, S. et al. *Eur. Phys. J. C* **81**, 157 (2021).
42. Khatsymovsky V. *Phys. Lett. B* **320** 234 (1994).
43. Bronnikov K. A.& Krechet V. G. *Phys. Rev. D* **99** 084051 (2019).
44. Bronnikov K. A., Bolokhov S. V., and Skvortsova M. V., *Int. J. Mod. Phys. D* **28**, 1941008 (2019).
45. E. Lifshitz, *J. Phys. (USSR)* **10**, 116–129 (1946).
46. Nicarstro F., et al. *Astrophys. J.* **629**, 700 (2005).
47. Tripp T.M., et al. *Astrophys. Journal. Suppl.*, **177**, 39 (2008).
48. Buote D.A., et al *Astrophys. J.*, **695**, 1351, (2009).
49. Samorodmitzky G., Taggy M.S. *Stable non-Gaussian random processes* - New York: Chapman & Hall, (1994).

Estimating the photometric redshifts of galaxies and QSOs using regression techniques in machine learning

A. Momtaz, M. H. Salimi and S. Shakeri*
Department of Physics, Isfahan University of Technology, Isfahan, 84156-8311, Iran
** E-mail: s.shakeri@iut.ac.ir*

Measuring distances of cosmological sources such as galaxies, stars and quasars plays an increasingly critical role in modern cosmology. Obtaining the optical spectrum and consequently calculating the redshift as a distance indicator could instantly classify these objects. As long as spectroscopic observations are not available for many galaxies and the process of measuring the redshift is time-consuming and infeasible for large samples, machine learning (ML) approaches could be applied to determine the redshifts of galaxies from different features including their photometric colors. In this paper, by using the flux magnitudes from the Sloan Digital Sky Survey (SDSS) catalog, we develop two ML regression algorithms (decision tree and random forest) for estimating the redshifts taking color indices as input features. We find that the random forest algorithm produces the optimum result for the redshift prediction, and it will be further improved when the dataset is limited to a subset with $z \leq 2$ giving the normalised standard deviation $\overline{\Delta Z}_{\rm norm} = 0.005$ and the standard deviation $\sigma_{\Delta z} = 0.12$. This work shows great potential of using the ML approach to determine the photometric redshifts of distant sources.

Keywords: Spectroscopy; Photometry; Machine Learning; Random Forest; Decision Tree

1. Introduction

Spectroscopy is usually applied as a valuable technique to determine the redshift of extragalactic sources. However, its high wavelength resolution limits its accuracy, a problem that can only partially be solved with more observation time.[1]

In recent years, a growing number of studies have relied on less precise statistical but more efficient estimates of redshifts based on broadband photometry.[2–4] A lot of key projects for the upcoming survey telescopes will involve these photometric redshifts. The photometric redshift is a tool used by multiple fields of astronomy to estimate the distances between objects in the sky. Since the process of redshift estimation is of great importance in various endeavors such as astronomical transient events, galaxy clustering, the mass function of the galaxy and the weak-lensing approach through constraining the presence of Dark Energy, several methods have been developed to choose an optimum technique.[5] These methods include leveraging training sets,[6] utilizing template spectra for comparisons[7] and mostly template fitting methods.[8]

Additionally, for over a decade, cosmologists have used ML techniques based on neural networks and regression algorithms to determine photometric redshifts.[9] In an attempt to establish a quantitative approach towards the performance of Random

Forests, Carliles et al. 2010 concludes that in contrast to other regression techniques, Random Forest regression overcomes several vital weaknesses.[10] Random Forests algorithm as a non-parametric procedure does not use a statistical model to describe the underlying data. The performance of parametric methods depends on how well the model fits the underlying distribution of data.[11] However, with highly skewed noise distributions, the Random Forest still provides reliable estimate of the error distribution, and this behavior is backed up by strong theoretical support.[12]

In Hoyle et al. (2015), an analysis of feature importance selection applied to the photometric redshift estimation using the ML architecture decision trees was presented. Such an investigation in this regard remarked in five optical frequency bands known as u-g-r-i-z, the flux magnitudes available from the SDSS, as a determinative input feature that encodes the most information about the redshifts of galaxies or quasars.[13]

Over the last 20 years, the SDSS has made a map of the universe. SDSS measurements of the galaxies, quasars and intergalactic gas structure have contributed significantly to tests of the standard cosmological model that describes our understanding of the history and future of the universe. Data Release 16 of SDSS[14] includes infrared, extragalactic and integral field spectra for nearby galaxies. The survey has mapped 4846156 useful spectra, 2863635 of which are galaxies and 960678 are quasi-stellar objects or the so-called quasars.

In this paper, we focus on Data Release 16 of SDSS and design a decision tree and a random forest algorithm to obtain an accurate estimation of the redshifts and then investigated their evaluation procedures to obtain optimum algorithms. We create a dataset of color indices acting as an approximation for the spectrum and as our input features. A noteworthy aspect of this work is the effort put into incorporating more information than a simple point estimate of the redshift. We determine the uncertainty associated with redshift estimates and calculate a posterior distribution.[15]

The rest of the paper is organized as follows. In Sec. 2 we explain ML algorithms including decision trees and random forests. In Sec. 3 the last data release of SDSS, DR16, and our input features will be presented besides our ML methodology for training and learning. The analysis and results are presented in Sec. 4 where we apply optimizing process for Decision Tree and Random Forest algorithms and compare the results of these algorithms. Finally, we discuss our results and conclude in Sec. 5.

2. Machine Learning Algorithms

2.1. *Decision Trees*

A decision trees is a ML algorithm which is used for both classification and regression learning tasks.[16] Based on a set of input features, decision trees generates its corresponding output targets. This is accomplished by a series of single decisions, each representing a node or branching of the tree. Following the training data, the

decision tree learning algorithm determines which decision should be made at each branch. Various metrics are employed by each algorithm in order to determine what is the best way to split the data (e.g. Gini impurity or information gain).

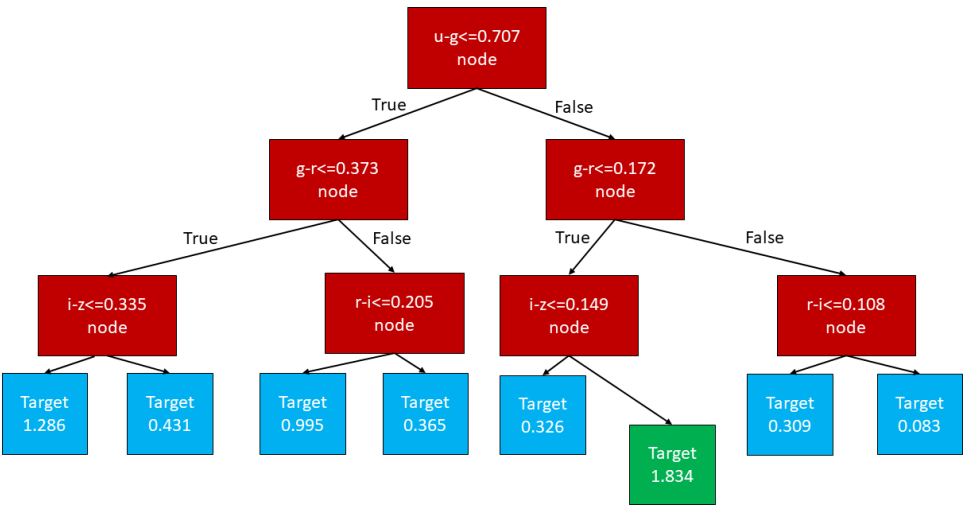


Fig. 1. Schematic of the decision tree algorithm.

The input of the calculated color indices into a series of decision nodes is shown above (Fig. 1), and the target redshift is calculated as output through the series of decision nodes. To implement ML, we will use Python's scikit-learn library. Using scikit-learn decision tree regression, a set of input features and target values are taken into account and then the model is constructed to adapt to new data.

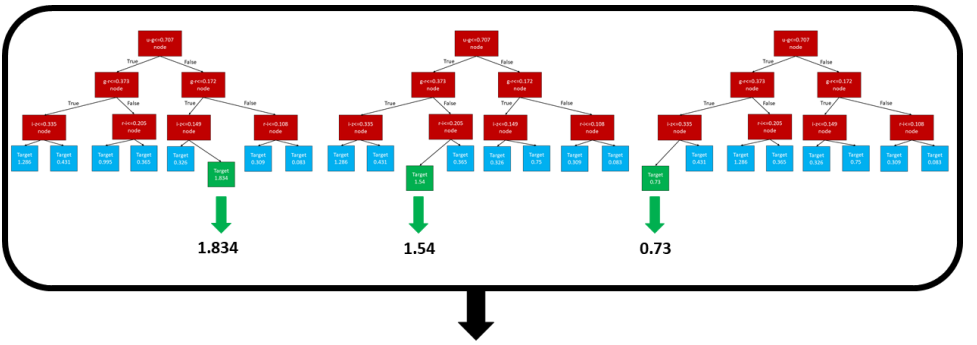
2.1.1. *Important Hyperparameters of Decision Tree*

The most significant hyperparameter of Decision Trees is the max-depth. In all other cases, the nodes are expanded until all leaves contain less than min-samples-split samples, or until all leaves are pure. In the case of a continuous value, as it is in this study, the most commonly used criteria to determine split locations are Mean Square Error (MSE), Poisson deviance as well as Mean Absolute Error (MAE).

2.2. *Random Forests*

Random forests can be viewed as an ensemble of decision trees in which random subsamples of the input attributes are used to create each tree (Fig. 2).[17] In this case, the trees will only learn a portion of the input attribute pattern and it will be poor classifiers.

Additionally, the trees learn only a part of the data, so they cannot learn artificial structures or be influenced by correlated attributes the same way as a neural network can. As final output, they give an average result based on the trained trees, which indicates the probability of each object belonging to one of the specified categories. This method gathers an attractive package of advantages. For instance, the model resists overfitting and is robust against correlated input attributes. In most cases, Random Forest produces excellent results. Moreover, it is one of the most common algorithms, owing to its simplicity and versatility.



$$\hat{Y}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{y}_b\,(x) = \frac{1.834 + 1.54 + 0.73}{3} = 1.36$$

Fig. 2. Schematic of the random forest algorithm.

2.2.1. *Important Hyperparameters of Random Forest*

First and foremost, there is the n-estimators hyperparameter, which is essentially the number of trees the algorithms build before taking the maximum voting or calculating the average prediction. It is generally true that a higher number of trees leads to better estimation and helps to predict more accurately, but it can also slow down the calculation speed.

3. Data

In this study, SDSS Data Release 16[14] is used as the data source. Observations for the SDSS have been carried out from Apache Point Observatory (APO) since 1998 (using the 2.5m Sloan Foundation Telescope[18] and from Las Campanas Observatory (LCO) since 2017 (using the du Pont 2.5m Telescope).

The newest DR16 contains the final sets of spectra collected as part of the main eBOSS observing program. SDSS's DR16 survey, then, ends a twenty-year stretch of performing a large-scale survey of the universe's structure at redshifts. SDSS has produced the largest catalog of spectroscopic redshifts of galaxies compare to any

other program over this time period. It is worth mentioning that DR16 provides spectra with usable redshifts for around 2.6 million unique galaxies. The photometric properties of around 100 million galaxies were also measured in this data set.

DR16 includes SDSS data products that are freely available through several channels. A large number of galaxy photometrically selected galaxies with spectroscopic redshifts are present in the SDSS that lend themselves well to the analysis presented in this paper to use as training, cross-validation and test samples.

3.1. *Input Features*

Using the color indices as our input features and the photometric redshift as the output, we have constructed a decision tree and a random forest. The data that we use for training are collected through accurate spectroscopic measurements of SDSS. These color indices have been created from flux magnitudes which are the total flux received in five frequency bands known as u-g-r-i-z (Fig. 3). Also, an astronomical color is derived from the difference in magnitudes of two filters, i.e. u - g. An object's color index provides an approximation to its spectrum, allowing it to be classified into various types.

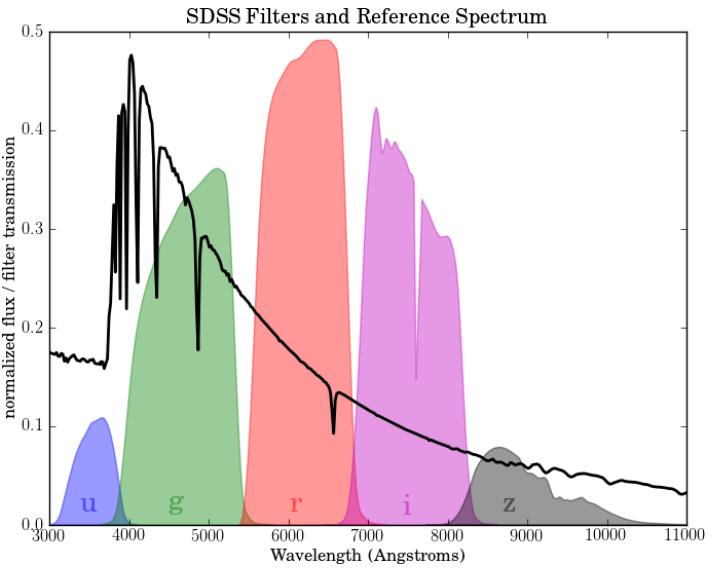


Fig. 3. SDSS filters and reference spectrum.[19]

We can get information about the physical processes taking place in distant astronomical sources by precise analysis of high-resolution spectra. Since the process of obtaining these spectra is very time-consuming and costly, astrophysicists usually observe objects through broadband filters and record them using the magnitude

system. As an example, we can define the magnitude in the u-band filter above as follows:

$$u = m_{\text{ref}} - 2.5 \log 10 \left[\int_0^\infty F(\lambda) S(\lambda) d\lambda \right] \tag{1}$$

where $F(\lambda)$ is the star's flux at wavelength λ, $S(\lambda)$ is a sensitivity function describes the fraction of the star's flux that is detected at a specific wavelength and m_{ref} is apparent bolometric magnitude.

3.2. *Training and Learning*

Following conventional ML methodology, the galaxy catalog is then subdivided into training and testing samples, with portions of 80 and 20, respectively. For each architecture and hyperparameter set, the ML system is trained using a training sample. Testing the learned machine's generalization ability based on a test sample is necessary to determine whether it truly generalizes to new data sets. Furthermore, k-fold cross-validation allows us to test the accuracy of our model. Our model is trained k times, with each training test recording the accuracy. We train the model every time using a different combination of k-1 subsets, and we test it with the final kth subset. Then, the overall accuracy of the model is calculated by taking the average of the k accuracy measurements.

4. Analysis and Results

In the current section, we start to analyze the final results of our structured algorithms. First we make a contour map of the redshifts based on a combination of color indices (Fig. 4), the results indicate that relatively well-defined regions of

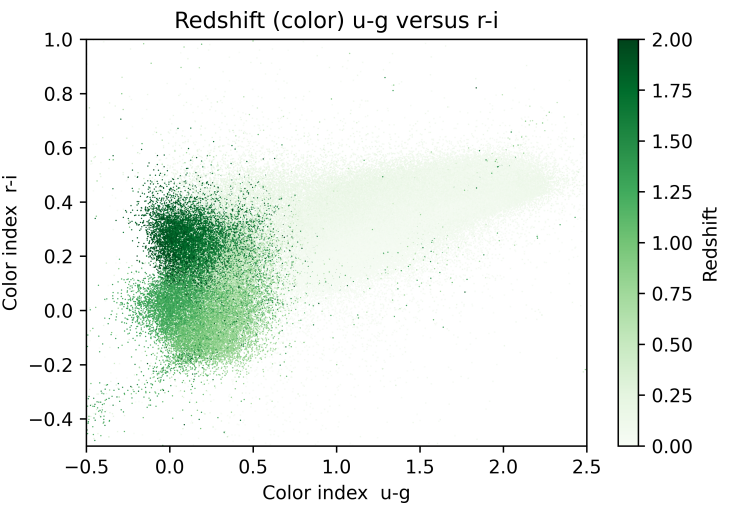
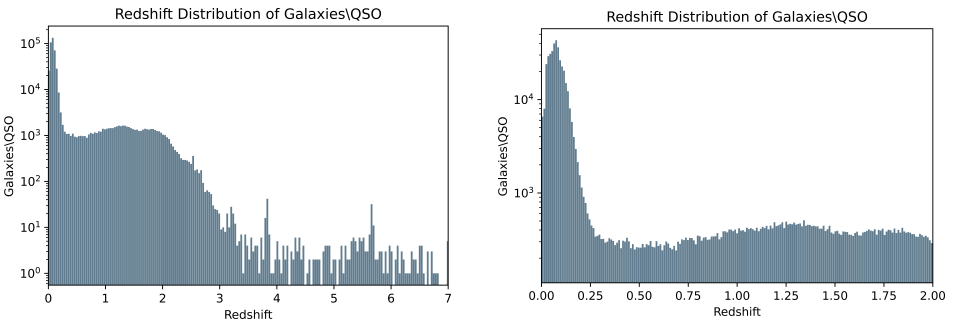


Fig. 4. Contour map of the redshifts based on a combination of color indices.

similar redshifts can be extracted and therefore the redshift of a new data point can be inferred from the color indices. So, this gesture underlines the fact that color indices are, indeed, appropriate input features for our ML algorithms.

In Fig. 5, the redshift distribution of galaxies and quasars are presented. In our dataset there are more galaxies in the dataset than quasars, and galaxies usually have lower redshifts.



(a) (b)

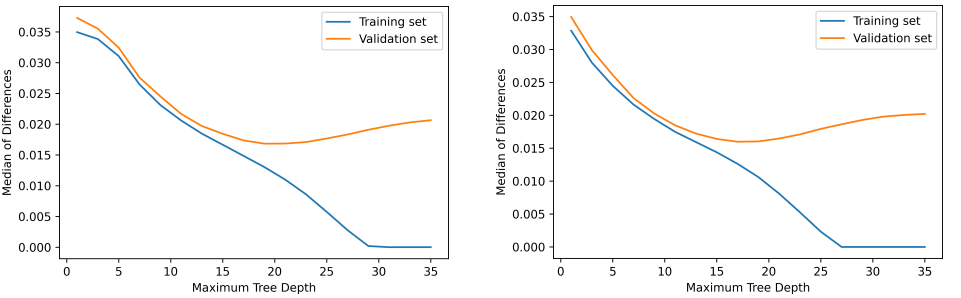
Fig. 5. Redshift distribution of Galaxies/QSOs. (a) Total range of the dataset. (b) limited range ($z \leq 2$) of the SDSS dataset.

We divide Our dataset into two ranges, i) contains all the measured redshifts ii) includes those below value 2. The reason behind this procedure is to achieve better performance of the algorithms, which will be discussed in detail in the following sections.

4.1. *Optimizing Process for Decision Tree*

There are some limitations for decision trees algorithm, including an over-fitting tendency. This means that it would potentially create a tree that is too complicated and does not address the statistical outliers in the data. In the process, general trends may not be accurately characterized.

Among the reasons for the over-fitting is the algorithm works by trying to optimize each node's decision locally. During our analysis, we will examine the impact of constraining the number of decision node rows (tree depth) on predictions. Various tree depths can be used to investigate whether the tree is over-fitting or not. We are particularly interested in comparing the algorithm performance on test data with its performance on training data (Fig. 6).



(a) (b)

Fig. 6. Analysis of the maximum tree depth (a) Total range of the dataset. (b) limited range (z≤2) of the SDSS dataset.

4.2. *Optimizing Process for Random Forest*

Creating a parameter grid prior to fitting is the first step in optimizing the Random Forest algorithm. A new combination of features is selected on each iteration, especially the n-estimators parameter, which represents the number of trees. In our Random Forest algorithm, we see that 600 trees will give the best result, as measured by the median value of the residual of measured and predicted redshifts. By setting the number of trees parameter to 600, we investigate the optimum amount of each tree's depth as in our prior grid search in the algorithm.

4.3. *Illustrations of the Models Performance*

Once the maximum tree depths have been determined, the algorithms are ready to take on the two different sets of data extracted from DR 16. We have used 450000 of the catalog's data based on our system's performance. There were 400000 data under value 2 in the parts of the paper where calculations had been executed on the filtered dataset. Fig. 7 and Fig. 8 present the measured redshifts from the survey versus the algorithms predicted redshifts. The color bars indicate the density of the galaxies or quasars in the datasets. A straight line with an angle of 45 degrees illustrates the success of the algorithms where measured redshifts is equal to the predicted ones. The overall performance of each algorithm was evaluated using three metrics:[20] model accuracy, the Root Mean Square Error (RMSE), and the normalized standard deviation. The normalized standard deviation is defined by:

$$\Delta z(\text{norm}) \equiv \frac{z_{\text{spec}} - z_{\text{phot}}}{z_{\text{spec}} + 1} \tag{2}$$

And, we calculate the standard deviation of the photometric redshifts from the spectroscopic redshift or namely the RMSE:

$$\sigma_{\Delta z} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \Delta z^2} \tag{3}$$

Fig. 7 illustrates how the Decision Tree algorithm has performed on both subsets of the dataset. Each datum's predicted redshift is plotted as a function of its spectroscopically measured value. The left panels show the the results for the whole range of the redshift data, while the right panels limited to the results with redshifts below value 2. Moreover, the error distributions of the predicted redshifts appear under each graph. The mathematical criteria for each model's performance are provided in the corresponding tables, however, it can obviously be inferred from both of the diagrams that filtering the redshift to below 2 values has indeed enhanced the accuracy of results.

As the Table 1 shows, filtering data into lower redshifts leads to a significant improvement in the model's performance. The accuracy of the algorithm has increased by 15%, the median difference, the normalized error as well as the standard deviation were all decreased by corresponding amounts. The explanation which underlies this improvement could be the fact that by splitting the dataset, there are no more considerable biases in the values of target redshifts and the algorithm would not experience a scattered distribution of values in the training procedure. Thereby, there should be no excessive miscalculation or poor estimation in the testing mode.

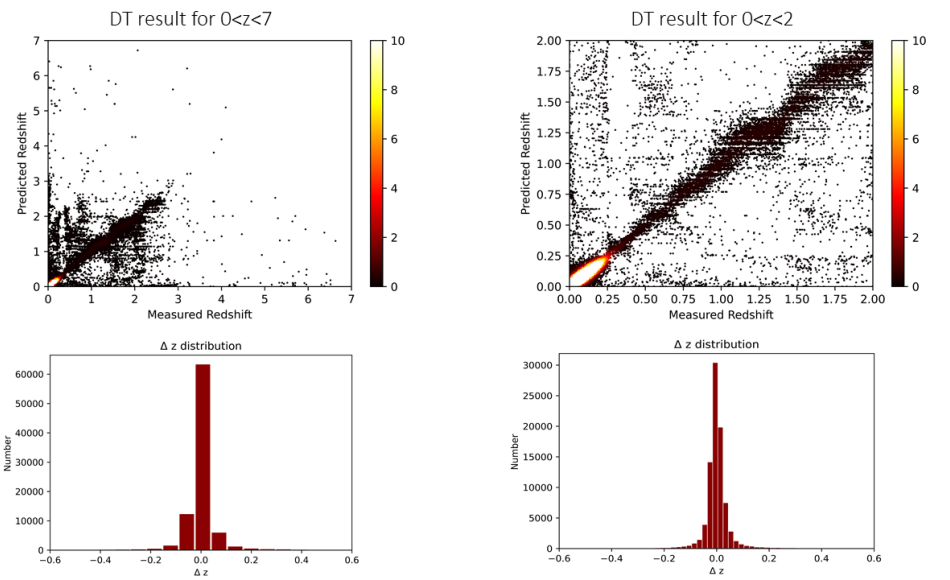


Fig. 7. The predictions of the Decision Tree algorithm for two ranges of redshifts in SDSS dataset.

Table 1. Analytical overview of statistical parameters for the Decision Tree algorithm for two ranges of redshifts z≤7 and z≤2.

Parameters	Decision Tree (z≤7)	Decision Tree (z≤2)
Accuracy	70.17%	85.26%
Max Depth	19	17
Median Difference	0.017	0.0156
$\overline{\Delta Z}_{\mathrm{norm}}$	0.0135	0.005
$\sigma_{\Delta z}$	0.28	0.16

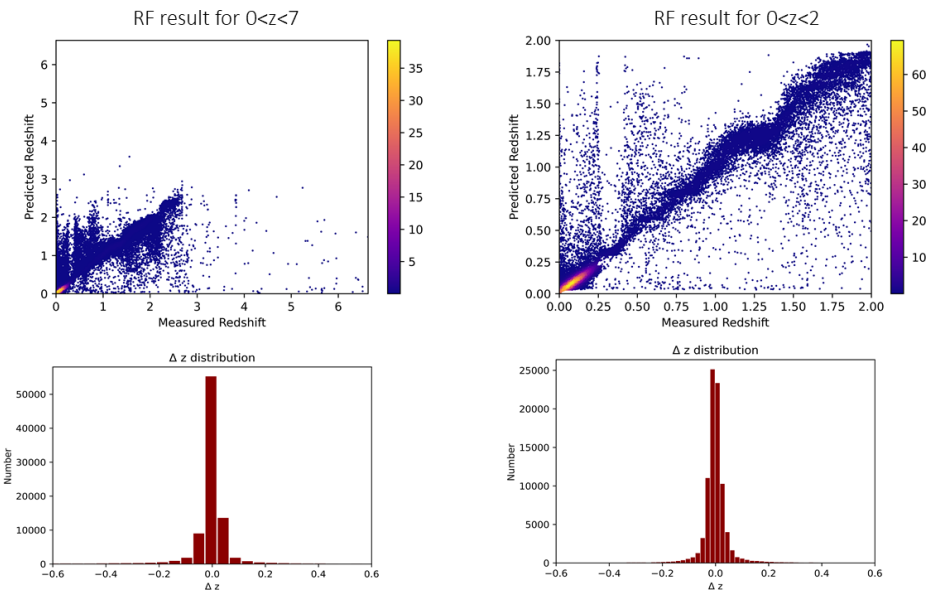


Fig. 8. The predictions of the Random Forest algorithm for two ranges in SDSS dataset.

Contemplating what was discussed in section 2.2, as the number of trees grow, the accuracy of the predicted value increase. Thus, it is expected that the Random Forests would have a better overall performance in terms of making more accurate predictions. As shown in Fig 8, both in the scattered plot and the distribution plot, the amount of precise estimations has increased. Since the color bar range indicates the density of data in the diagram and it includes a greater range of numbers than the Decision Tree algorithm. We conclude that the Random Forest algorithm has better accuracy. This fact is also presented in terms of statistical parameters showing the accuracy in Table 2 such as the standard deviation.

Table 2. Analytical overview of statistical parameters for the Random Forest algorithm for two ranges of redshifts z≤7 and z≤2.

Parameters	Random Forest (z≤7)	Random Forest (z≤2)
Accuracy	81.02%	91.00%
Max Depth	25	13
Number of Trees	600	600
Median Difference	0.0164	0.0154
$\overline{\Delta Z}_{\rm norm}$	0.013	0.005
$\sigma_{\Delta z}$	0.23	0.12

5. Conclusion and Remarks

The current research addresses two leading ML regression algorithms namely the Decision Tree and the Random Forest, that were structured for estimating the redshifts of distant objects such as galaxies and quasars. Since there exist vital limitations in obtaining spectroscopic measurements, the photometric optical band data are widely investigated in this realm. This paper uses the 16 Data Release of the SDSS, where photometric colors are used as input features of the mentioned models. After developing the Decision tree and the Random Forest algorithms and evaluating the final results, it was concluded that the Random Forest algorithm will significantly perform better in this case (Table 3). The Random Forest algorithm leads to much better accuracy and marginally better standard deviation and median difference as seen in Table 3.

Another noteworthy improvement was filtering the dataset to redshifts below value 2. The significant effect of redshift splitting was illustrated in section 4.3 and we extensively discussed about all the possible reasons behind this effect.

Table 3. Comparison between statistical parameters of the Decision Tree and the Random forest algorithms for z≤2.

Parameters	Decision Tree (z≤2)	Random Forest (z≤2)
Accuracy	85.26%	91.00%
Median Difference	0.0156	0.0154
$\overline{\Delta Z}_{\rm norm}$	0.005	0.005
$\sigma_{\Delta z}$	0.16	0.12

A few recent attempts for redshift estimation applying similar developed ML algorithms are addressed in Table 4. The crucial factor affecting their evaluations is the amount of data in the training sets. Remarkably, in this work taking 400000

Table 4. Comparison among results obtained using similar methodologies by different collaborations and our results in this work.

Refrence Articles	Data Release	Training Set	$\overline{\Delta Z}_{\rm norm}$	ML Algorithm
Beck(2016)	12	1,976,978	5.84×10^{-5}	Local Linear Regression
Paul(2018)[1]	12	20,000	2×10^{-3}	Random-Forest
Baldeschi(2021)[12]	16	1,251,249	1×10^{-3}	Random-Forest
This work	16	320,000	5×10^{-3}	Random-Forest

data into consideration for the training set, we achieved nearly equivalent accuracy to the previous works.

Our results show the great potential of the ML methods for redshift estimation of distant sources using color index features. It is worth mentioning that applying ML methods is unavoidable when a large number of astrophysical data will be obtained from the next generation of sky surveys as the era of big data has started.

References

1. N. Paul, N. Virag and L. Shamir, A catalog of photometric redshift and the distribution of broad galaxy morphologies, *Galaxies* **6**, 1 (2018).
2. D. Wilson, H. Nayyeri, A. Cooray and B. Häußler, Photometric Redshift Estimation with Galaxy Morphology Using Self-organizing Maps, *The Astrophysical Journal* **888**, p. 83 (2020).
3. J. v. Vugt, Photometric Redshift Estimation of Distant Quasars (2016).
4. B. Mobasher, P. Capak, N. Z. Scoville, T. Dahlen, M. Salvato, H. Aussel, D. J. Thompson, R. Feldmann, L. Tasca, O. Lefevre, S. Lilly, C. M. Carollo, J. S. Kartaltepe, H. McCracken, J. Mould, A. Renzini, D. B. Sanders, P. L. Shopbell, Y. Taniguchi, M. Ajiki, Y. Shioya, T. Contini, M. Giavalisco, O. Ilbert, A. Iovino, V. Le Brun, V. Mainieri, M. Mignoli and M. Scodeggio, Photometric Redshifts of Galaxies in COSMOS, *The Astrophysical Journal Supplement Series* **172**, 117 (2007).
5. M. Salvato, O. Ilbert and B. Hoyle, The many flavours of photometric redshifts, *Nature Astronomy* **3**, 212 (2019).
6. A. Collister and O. Lahav, ANN z : Estimating Photometric Redshifts Using Artificial Neural Networks, *Publications of the Astronomical Society of the Pacific* **116**, 345 (2004).
7. N. Benitez, Bayesian Photometric Redshift Estimation, *The Astrophysical Journal* **536**, 571 (2000).
8. G. Bruzual and S. Charlot, Stellar population synthesis at the resolution of 2003, *Monthly Notices of the Royal Astronomical Society* **344**, 1000 (2003).
9. K. Zhang, D. J. Schlegel, B. H. Andrews, J. Comparat, C. Schäfer, J. A. Vazquez Mata, J.-P. Kneib and R. Yan, Machine-learning Classifiers for Intermediate Redshift Emission-line Galaxies, *The Astrophysical Journal* **883**, p. 63 (2019).
10. S. Carliles, T. Budavári, S. Heinis, C. Priebe and A. S. Szalay, Random forests for photometric redshifts, *Astrophysical Journal* **712**, 511 (2010).
11. S. Carliles, T. Budavári, S. Heinis, C. Priebe and A. Szalay, Photometric Redshift Estimation on SDSS Data Using Random Forests, **XXX**, 1 (2007).
12. A. Baldeschi, M. Stroh, R. Margutti, T. Laskar and A. Miller, Photometric redshift estimation of galaxies in the Pan-STARRS 3π survey- I. Methodology (2021).

13. B. Hoyle, M. Michael Rau, R. Zitlau, S. Seitz and J. Weller, Feature importance for machine learning redshifts applied to SDSS galaxies, *Monthly Notices of the Royal Astronomical Society* **449**, 1275 (2015).
14. R. Ahumada, C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, E. Armengaud, M. Aubert, S. Avila, V. Avila-Reese, C. Badenes, C. Balland, K. Barger, J. K. Barrera-Ballesteros, S. Basu, J. Bautista, R. L. Beaton, T. C. Beers, B. I. T. Benavides, C. F. Bender, M. Bernardi, M. Bershady, F. Beutler, C. M. Bidin, J. Bird, D. Bizyaev, G. A. Blanc, M. R. Blanton, M. Boquien, J. Borissova, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, M. Bureau, A. Burgasser, E. Burtin, M. Cano-Díaz, R. Capasso, M. Cappellari, R. Carrera, S. Chabanier, W. Chaplin, M. Chapman, B. Cherinka, C. Chiappini, P. Doohyun Choi, S. D. Chojnowski, H. Chung, N. Clerc, D. Coffey, J. M. Comerford, J. Comparat, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, K. Cunha, G. d. S. Ilha, Y. S. Dai, S. B. Damsted, J. Darling, J. W. Davidson, R. Davies, K. Dawson, N. De, A. de la Macorra, N. De Lee, A. B. d. A. Queiroz, A. Deconto Machado, S. de la Torre, F. Dell'Agli, H. du Mas des Bourboux, A. M. Diamond-Stanic, S. Dillon, J. Donor, N. Drory, C. Duckworth, T. Dwelly, G. Ebelke, S. Eftekharzadeh, A. Davis Eigenbrot, Y. P. Elsworth, M. Eracleous, G. Erfanianfar, S. Escoffier, X. Fan, E. Farr, J. G. Fernández-Trincado, D. Feuillet, A. Finoguenov, P. Fofie, A. Fraser-McKelvie, P. M. Frinchaboy, S. Fromenteau, H. Fu, L. Galbany, R. A. Garcia, D. A. García-Hernández, L. A. G. Oehmichen, J. Ge, M. A. G. Maia, D. Geisler, J. Gelfand, J. Goddy, V. Gonzalez-Perez, K. Grabowski, P. Green, C. J. Grier, H. Guo, J. Guy, P. Harding, S. Hasselquist, A. J. Hawken, C. R. Hayes, F. Hearty, S. Hekker, D. W. Hogg, J. A. Holtzman, D. Horta, J. Hou, B.-C. Hsieh, D. Huber, J. A. S. Hunt, J. I. Chitham, J. Imig, M. Jaber, C. E. J. Angel, J. A. Johnson, A. M. Jones, H. Jönsson, E. Jullo, Y. Kim, K. Kinemuchi, C. C. Kirkpatrick IV, G. W. Kite, M. Klaene, J.-P. Kneib, J. A. Kollmeier, H. Kong, M. Kounkel, D. Krishnarao, I. Lacerna, T.-W. Lan, R. R. Lane, D. R. Law, J.-M. Le Goff, H. W. Leung, H. Lewis, C. Li, J. Lian, L. Lin, D. Long, P. Longa-Peña, B. Lundgren, B. W. Lyke, J. Ted Mackereth, C. L. MacLeod, S. R. Majewski, A. Manchado, C. Maraston, P. Martini, T. Masseron, K. L. Masters, S. Mathur, R. M. McDermid, A. Merloni, M. Merrifield, S. Mészáros, A. Miglio, D. Minniti, R. Minsley, T. Miyaji, F. G. Mohammad, B. Mosser, E.-M. Mueller, D. Muna, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, P. Nair, K. Nandra, J. C. do Nascimento, R. J. Nevin, J. A. Newman, D. L. Nidever, C. Nitschelm, P. Noterdaeme, J. E. O'Connell, M. D. Olmstead, D. Oravetz, A. Oravetz, Y. Osorio, Z. J. Pace, N. Padilla, N. Palanque-Delabrouille, P. A. Palicio, H.-A. Pan, K. Pan, J. Parker, R. Paviot, S. Peirani, K. P. Ramírez, S. Penny, W. J. Percival, I. Perez-Fournon, I. Pérez-Ràfols, P. Petitjean, M. M. Pieri, M. Pinsonneault, V. J. Poovelil, J. T. Povick, A. Prakash, A. M. Price-Whelan, M. J. Raddick, A. Raichoor, A. Ray, S. B. Rembold, M. Rezaie, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, A. Roman-Lopes, C. Román-Zúñiga, B. Rose, A. J. Ross, G. Rossi, K. Rowlands, K. H. R. Rubin, M. Salvato, A. G. Sánchez, L. Sánchez-Menguiano, J. R. Sánchez-Gallego, C. Sayres, A. Schaefer, R. P. Schiavon, J. S. Schimoia, E. Schlafly, D. Schlegel, D. P. Schneider, M. Schultheis, A. Schwope, H.-J. Seo, A. Serenelli, A. Shafieloo, S. J. Shamsi, Z. Shao, S. Shen, M. Shetrone, R. Shirley, V. S. Aguirre, J. D. Simon, M. F. Skrutskie, A. Slosar, R. Smethurst, J. Sobeck, B. C. Sodi, D. Souto, D. V. Stark, K. G. Stassun, M. Steinmetz, D. Stello, J. Stermer, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, A. Stutz, G. Suárez, J. Sun, M. Taghizadeh-Popp, M. S. Talbot, J. Tayar, A. R. Thakar, R. Theriault, D. Thomas, Z. C. Thomas, J. Tinker, R. Tojeiro, H. H. Toledo, C. A. Tremonti, N. W. Troup, S. Tuttle, E. Unda-Sanzana,

M. Valentini, J. Vargas-González, M. Vargas-Magaña, J. A. Vázquez-Mata, M. Vivek, D. Wake, Y. Wang, B. A. Weaver, A.-M. Weijmans, V. Wild, J. C. Wilson, R. F. Wilson, N. Wolthuis, W. M. Wood-Vasey, R. Yan, M. Yang, C. Yèche, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, C. Zhao, G. Zhao, Z. Zheng, Z. Zheng, G. Zhu and H. Zou, The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra, *The Astrophysical Journal Supplement Series* **249**, p. 3 (2020).

15. H. Oyaizu, M. Lima, C. E. Cunha, H. Lin and J. Frieman, Photometric Redshift Error Estimators, *The Astrophysical Journal* **689**, 709 (2008).
16. L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone, *Classification and regression trees* (Chapman and Hall/CRC, 1984).
17. L. Breiman, Random Forests, *Machine Learning* **45**, 5 (2001).
18. J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger, L. N. Carey, G. R. Knapp, D. G. York, W. N. Boroski, S. M. Kent, R. H. Lupton, C. M. Rockosi, M. L. Evans, P. Waddell, J. E. Anderson, J. Annis, J. C. Barentine, L. M. Bartoszek, S. Bastian, S. B. Bracker, H. J. Brewington, C. I. Briegel, J. Brinkmann, Y. J. Brown, M. A. Carr, P. C. Czarapata, C. C. Drennan, T. Dombeck, G. R. Federwitz, B. A. Gillespie, C. Gonzales, S. U. Hansen, M. Harvanek, J. Hayes, W. Jordan, E. Kinney, M. Klaene, S. J. Kleinman, R. G. Kron, J. Kresinski, G. Lee, S. Limmongkol, C. W. Lindenmeyer, D. C. Long, C. L. Loomis, P. M. McGehee, P. M. Mantsch, E. H. Neilsen, Jr., R. M. Neswold, P. R. Newman, A. Nitta, J. Peoples, Jr., J. R. Pier, P. S. Prieto, A. Prosapio, C. Rivetta, D. P. Schneider, S. Snedden and S.-i. Wang, The 2.5 m Telescope of the Sloan Digital Sky Survey, *The Astronomical Journal* **131**, 2332 (2006).
19. Ž. Ivezić, A. Connolly, J. Vanderplas and A. Gray, *Statistics, Data Mining and Machine Learning in Astronomy* (Princeton University Press, 2014).
20. S. J. Curran, J. P. Moss and Y. C. Perrott, Qso photometric redshifts using machine learning and neural networks, *Monthly Notices of the Royal Astronomical Society* **503**, p. 2639–2650 (Feb 2021).

Deep learning in quasar physics

F. Rastegar Nia

Physics Department, Alzahra University Vanak, 1993891176, Tehran, Iran
ICRA and Dipartimento di Fisica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, I-00185 Roma, Italy
ICRANet, Piazza della Repubblica 10, I-65122 Pescara, Italy
E-mail: f.rastegarnia@alzahra.ac.ir

M. T. Mirtorabi

Physics Department, Alzahra University Vanak, 1993891176, Tehran, Iran
E-mail: torabi@alzahra.ac.ir

R. Moradi* and Y. Wang†

ICRANet, Piazza della Repubblica 10, Pescara,I–65122, Italy,
ICRA, Dipartimento di Fisica, Sapienza Universita' di Roma, Rome, I–00185, Italy and INAF, Rome, 00136, Italy
**E-mail: ruffini@icra.it*
**E-mail: rahim.moradi@icranet.org*
†*E-mail: yu.wang@icranet.org*

A. Vafaei Sadr

Departement de Physique Theorique and Center for Astroparticle Physics, University of Geneva
E-mail: Alireza.VafaeiSadr@unige.ch

In view of increasing data volume of existing and upcoming telescopes/detectors we here apply the 1–dimensional convolutional neural network (CNN) to estimate the redshift of (high-)redshifts quasars in Sloan Digital Sky Survey IV (SDSS-IV) quasar catalog from DR16 of eBOSS. Our CNN takes the flux of the quasars as an array and their redshift as labels. We here evidence that new structure of the network, and augmenting the training set, provide a high precision result in estimating the redshift of quasars.

Keywords: Quasar, Deep learning, CNN, SDSS

1. Introduction

Quasars are the most luminous active galactic nuclei(AGN) which are powered by accretion disk around supermassive black holes at centers of their host galaxies. Thanks to their high luminosity, they can be observed across the universe in wide range of redshift from z = 0 to z ∼ 7. Thus, quasars give us important information about the early universe, the structure formation and evolution.[1,2]

Nowadays, astronomy and astrophysics have been brought into the big data era through the construction and development of ground-based and space telescope. Since huge amount of data, ranging from gamma-ray, x-ray, ultraviolet, optical,

infrared and radio bands of quasars is available in sky surveys, various challenges and opportunities are created for scientific discoveries.[3,6]

In recent years, machine learning (ML) and deep learning (DL) have been utilized in astronomy and astrophysics in order to deal with the big data surveys and as well extract the new physical understanding. The goal in ML and DL is to diagnose, by optimization, common characteristics and features in data.[10] These interesting branch of computer science have been used in estimating the physical parameters and classification of celestial objects; for example the morphological classification of galaxies,[15,16,20] estimation of photometric redshift,[12,25,33,34] classification of star/galaxy,[4,27,32] spectral classification of stars.[5,7,17,35]

This article is mainly dedicated to train a CNN to predict the redshift of Quasi-stellar radio source (quasars) or quasi stellar objects (QSO) in Sloan Digital Sky Survey IV (SDSS-IV) quasar catalog from Data Release 16 (DR16) of the extended Baryon Oscillation Spectroscopic Survey (eBOSS), which is the most comprehensive catalog of spectroscopically collected quasars to date.[26]

2. Dataset

In this paper, we exploit dataset from the quasar spectra obtained by the Sloan Digital Sky Survey-4(SDSS-4). They are provided by the Sixteenth Data Release (DR16) of SDSS extended Baryon Oscillation Spectroscopic Survey (eBOSS).[14] According to the technical details of eBOSS, namely the wavelength coverage in range of 361nm$-$1014nm with the resolution $R = 2000$, where $R = \frac{\lambda}{\Delta\lambda}$, more than 700,000 quasar spectra have been detected in the redshift range $0 \leq z \leq 7.1$.

The distributions of detected quasars redshift in the DR16 catalog is shown in Fig. 1. As can be seen, the number of quasars reaches its maximum around $z \approx 2.5$; at earlier epochs i.e., higher redshifts, they are comparatively rare.

2.1. *Redshift*

The observed quasar spectra contain broad emission lines, often time-variable flux both in the continuum and in the emission lines and UV flux; see e.g.[11] The shift and deformation of the various lines and specific characteristic of different fluxes in the spectra of quasars, due to the cosmological redshift, makes quasars a proper tool to invest the deep learning method and extract the common features hidden in their spectra (flux).

In DR16Q catalog, redshifts obtained for quasars by different methods, such as principal component analysis (PCA) and QuasarNET are reported. In this work we use "*primary*" redshift or best redshift for quasars which has been selected from the available visual inspection redshifts,[8] or, alternatively, the SDSS automated pipeline redshift;, see[23,24] and references therein for more information. The best option is exploiting only the visual inspection redshifts which will be presented in the forthcoming papers.

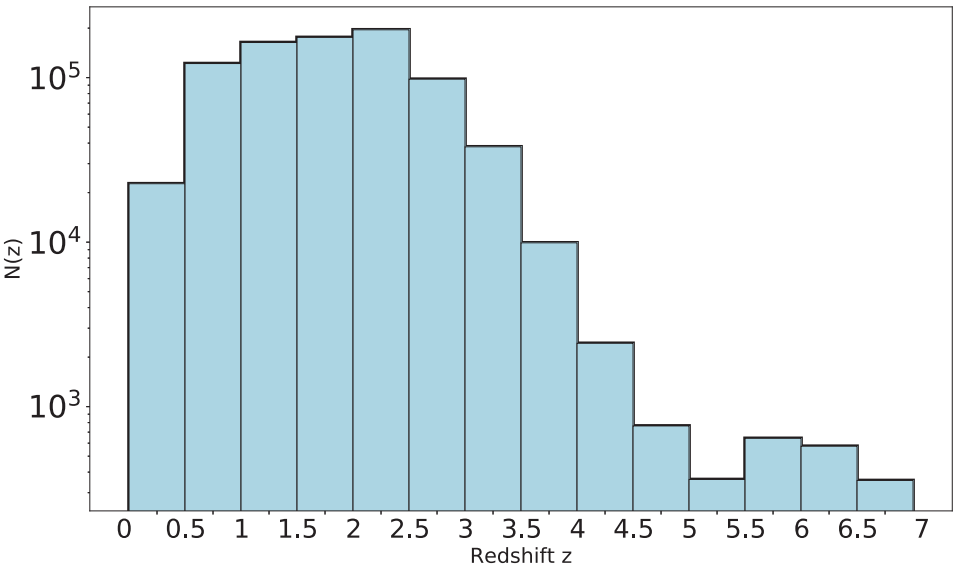


Fig. 1. The distributions of "*primary*" redshift of quasars in the DR16 catalog.

3. Convolutional neural network (CNN)

3.1. *CNN architecture*

Since the spectra of quasars are represented by flux versus the wavelength, namely 1+1 dimensional set of data, they can be considered as as time-series. For this reason, the neural network model in this work is designed to be a 1-dimensional CNN. It operates like a combination of some mathematical functions, which optimizes through several filters, and transforms the input data; here flux of quasars, to the output data, here redshift of quasars, by extracting the hidden features from the quasars' fluxes.

The physical information, such as redshift are hidden in the observed flux, and CNNs, as a regressor, are supposed to estimate the redshift through the training process. The CNN in the current work is a combination of convolutional and fully-connected layers. The convolutional layers are initialized with He Normal initializer.[22] If the model requires to access the non-linear modes in data he Rectified Linear Unit (ReLU) activation function is implemented.[36]

The filters in each convolutional layer scans each row (flux) and extract the prominent features from the raw input for the specific redshift. The feature-extraction layers have a repeating pattern of the convolutional layers whereas the pooling layers reduce the dimensions and concentrate only on the most prominent features. This process recognizes the non-linear correlation within the dataset. Finally, the fully-connected feed-forward layers connect and assign the extracted

features to the output layer (redshift). All free parameters in the model change dynamically as the algorithm finds the best solution, achieved by the back-propagation learning algorithm.

Many free parameters, such as the number of layers, layer specifications, and their arrangement can be changed while one constructs a neural network model; known as model architecture. The architecture selection affects optimization and quality of the performance; this process is called hyperparameter optimization.[18] The hyper-parameter optimization process is necessary in order to ensure that neither underfits nor overfits of the training set happens.

In the hyper-parameter optimization process, the number and length of filters of convolutional layers, the number of nodes in fully-connected layers, and the kernel size of the Maxpooling layer are considered and finally the "Mean Squared Error" (MSE), as the loss function, is utilized in our CNN.

We also use Adam optimizer to optimize the loss function, which is an algorithm for the first-order gradient-based optimization of stochastic objective functions. We here set the learning rate as 0.0001 and the coefficients for computing the averages of gradient and its square as 0.5 and 0.9, the weight decay of L2 penalty is set as 0.[28, 29]

Figure 2 demonstrates the CNN pipeline of this work. It takes a quasar spectrum as a 1-dimensional array and predicts the redshift. We have tested different samples of training set, and the addition of more layers to this configuration does not enhance the prediction accuracy presented in this paper for different redshift intervals.

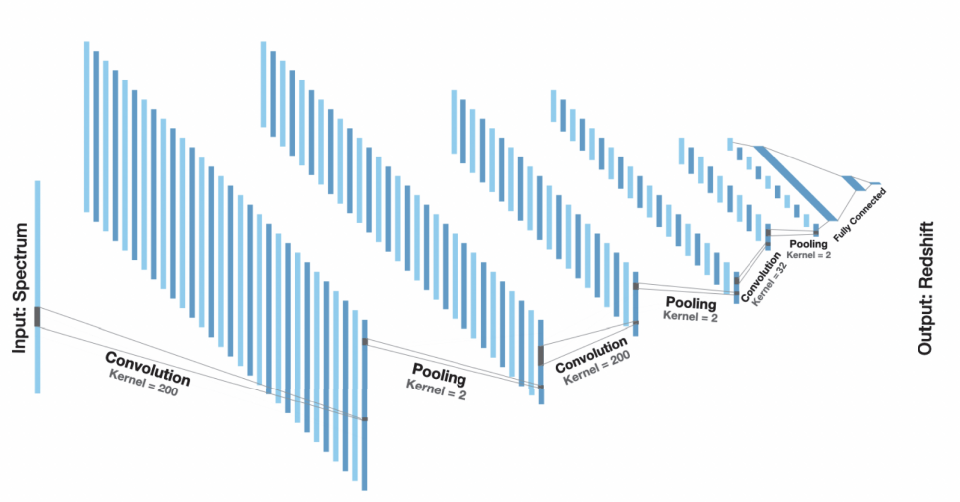


Fig. 2. Structure of 1-dimensional CNN developed in this work to learn higher-order features hidden in the input spectrum. The CNN goes through the spectrum via a convolutional layer of kernel size = 200, 200, 32, respectively in order to search for the global and local pattern. The fully connected layers output the redshift.

3.2. *Pre-processing the data*

Data pre–processing is crucial in two aspects: (1) to provide an understandable dataset for DL networks, and (2) to increase the speed and accuracy of processing. As the first step of the pr–eprocessing procedure in the present work, a two-dimensional matrix of the dataset is created. In the matrix each row represents the quasar flux and each column relates to a flux a certain wavelength. At the second step, the flux of every spectrum is normalized via the Zero-Mean Normalization method. Moreover, the normalized fluxes are stored in a Numpy array and applied as features dataset, their corresponding redshifts are collected for labels dataset as well.

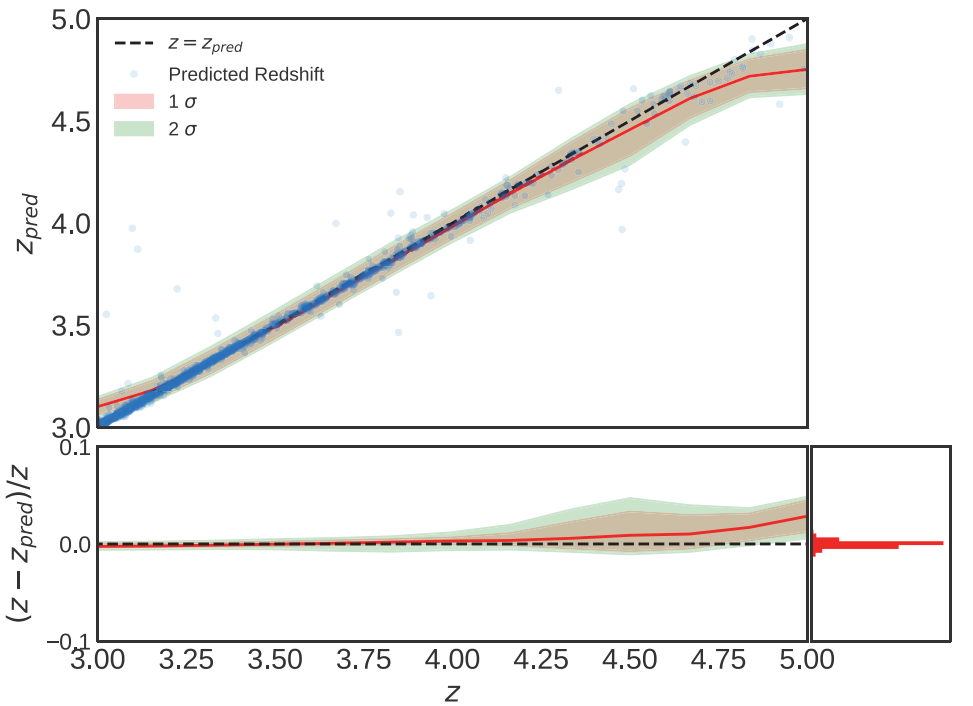


Fig. 3. Redshift predicted by our CNN for $3 \leq z \leq 5$. Top: The predicted redshift vs. the redshift reported in the DR catalog. Bottom: The relative error and their distribution. As it can be seen the distribution of relative error follows a Gaussian distribution with mean, $\mu = 0.0042$ and a standard deviation of $\sigma = 0.014$.

4. Prediction in the $3 \leq z \leq 5$ interval

As an example, we use the quasars in the $3 \leq z \leq 5$ interval as sources to test our CNN. In general, there are $51,100$ quasars in this interval,75% of them, namely 38832 spectra are taken as the sample in order to train our CNN (training set) and

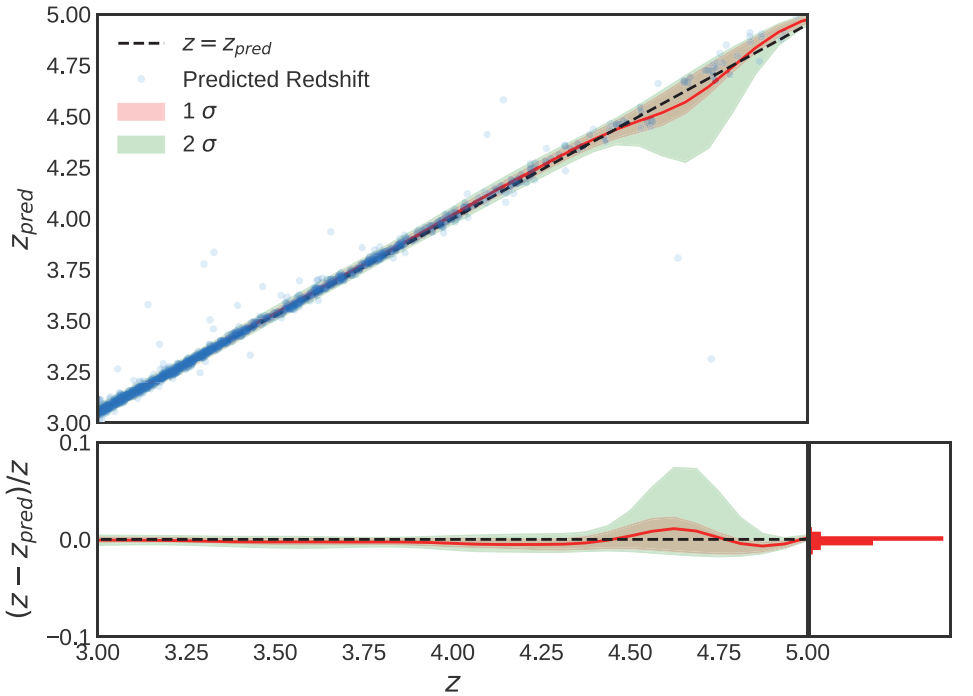


Fig. 4. Redshift predicted by our CNN for $3 \leq z \leq 5$ with training set in $2.9 \leq z \leq 5.2$ interval. Top: The predicted redshift vs. the redshift reported in the DR catalog. Bottom: The relative error and their distribution. The distribution of relative error follows a Gaussian distribution with mean, $\mu = 0.0038$ and a standard deviation of $\sigma = 0.012$.

25%, namely 12776 as sources to testify our CNN (test set). In the whole paper, the test set has not been used to train the CNN. Moreover, no limit on the S/N ratio of spectra is imposed.

The predicted redshift by our CNN for $3 \leq z \leq 5$ together with their relative errors distribution and the best fit is shown in Fig. 3. The distribution of relative error, RE,

$$\mathrm{RE} \equiv \frac{\mathrm{z} - \mathrm{z}_{\mathrm{pred}}}{\mathrm{z}}, \tag{1}$$

which z is the redshift reported by SDSS and z_{pred} is the redshift predicted by our CNN, follows a Gaussian distribution with mean, $\mu = 0.0042$ and a standard deviation of $\sigma = 0.014$. The method used for smoothing the scatterplot is locally weighted regression (LWR).[13]

As it can be seen in Fig. 3, the fit in the lower and upper edges of the prediction, namely $4.5 \lesssim z \leq 5$ and $3 \lesssim z \leq 3.1$ intervals, is not as significant as other time intervals.

This problem can be due to:

(1) the lack of data in $z < 3$ and $z > 5$ and consequently the failed LWR in the aforementioned intervals.
(2) the deficiency of training sample in $4.5 \lesssim z \leq 5$.

In order to solve the first issue, we extend our training set to $2.9 \leq z \leq 5.2$. By adding this extra data to the train set we make sure that there is no bias in the LWR fitting presented in Fig. 3 at its lower and upper part; see Fig. 4.

In order to accommodate the second issue and increase the number of spectra of the training set in $4.5 \lesssim z \leq 5$ we first convert the observed spectra in $4 \leq z \leq 4.5$ interval, where there are enough spectra to produce a reliable training set, into their cosmological rest-frame by dividing their wavelength by 1+z then we convert data set to $4.5 \leq z \leq 5$ by multiplying the wavelengths by a factor of 1+z+0.5.

Figure 5 represents the results when extra samples from $2.9 \leq z \leq 3$ and $5 \leq z \leq 5.2$ as well as the redshifted sample from $4 \leq z \leq 4.5$ to $4.5 \leq z \leq 5$ have been added to the original training set. The distribution of relative error follows a

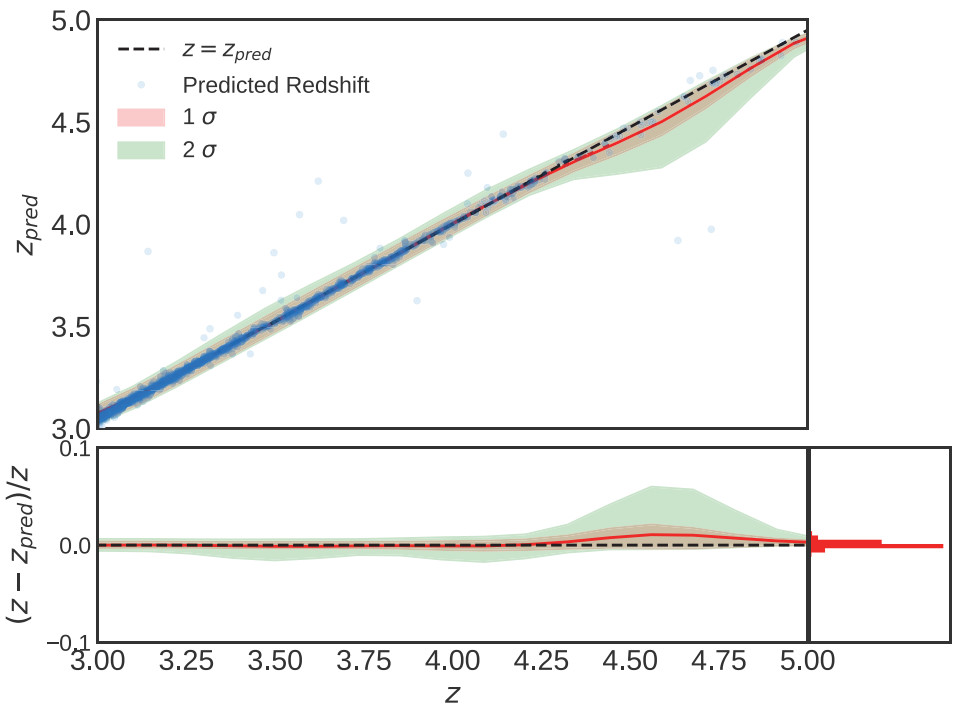


Fig. 5. Redshift predicted by our CNN for $3 \leq z \leq 5$ with training set in $2.9 \leq z \leq 5.2$ interval together with the converted observed spectra in $4 \leq z \leq 4.5$ interval into $4.5 \leq z \leq 5$. Top: The predicted redshift vs. the redshift reported in the DR catalog. Bottom: The relative error and its distribution which follows a Gaussian distribution with mean, $\mu = 0.0030$ and a standard deviation of $\sigma = 0.010$.

Gaussian distribution with mean, $\mu = 0.003$ and a standard deviation of $\sigma = 0.01$. The prediction is clearly improved in $4.5 \leq z \leq 5$ comparing to initial result shown in Fig. 3.

5. Conclusions

Deep learning, especially CNN[31, 37] is promising in the future Astrophysical studies. CNN can be utilized, owing to its several convolutional and fully connected layers to find the common deep hidden patterns in the spectrum, in provision for improving our astrophysical knowledge of distant objects. In this regard, we showed that exploiting a CNN leads to a statistically significant prediction of the redshift of quasars. We also showed that augmentation of dataset, which depends on the statistical and physical features of samples improves the prediction of redshift. Since CNN finds the physical characteristics of the spectra and these characteristics are deformed by the cosmological redshift the results presented here can be extended to the other energy bands like, X-ray, Infrared, UV, Gamma-ray and etc.

References

1. Nina Hernitschek et al 2016 ApJ 817 73, DOI: https://doi.org/10.3847/0004-637X/817/1/73.
2. A. Lupi, M. Volonteri, R. Decarli, S. Bovino, J. Silk and J. Bergeron, Mon. Not. Roy. Astron. Soc. **488** (2019) no.3, 4004-4022 doi:10.1093/mnras/stz1959 [arXiv:1901.02464 [astro-ph.GA]].
3. Allen, G., Andreoni, I., Bachelet, E., et al. 2019, arXiv preprint arXiv:1902.00522.
4. Bai, Y., Liu, J., Wang, S., & Yang, F. 2018, The Astronomical Journal, 157, 9.
5. Bailer-Jones, C. A., Irwin, M., & Von Hippel, T. 1998, Monthly Notices of the Royal Astronomical Society, 298, 361.
6. Ball, N. M., & Brunner, R. J. 2010, International Journal of Modern Physics D, 19, 1049.
7. Bialek, S., Fabbro, S., Venn, K. A., et al. 2019, arXiv preprint arXiv:1911.02602.
8. Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144, doi:10.1088/0004-6256/144/5/144.
9. Busca, N., & Balland, C. 2018, arXiv e-prints, arXiv:1808.09955.
10. Carleo, G., Cirac, I., Cranmer, K., et al. 2019, Rev. Mod. Phys., 91, 045002, doi:10.1103/RevModPhys.91.045002.
11. Carroll, B. W., & Ostlie, D. A. 1996, An Introduction to Modern Astrophysics.
12. Cavuoti, S., Brescia, M., Tortora, C., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 3100.
13. Cleveland, W. S. 1979, Journal of the American Statistical Association, 74, 829, doi:10.1080/01621459.1979.10481038.
14. Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, The Astronomical Journal, 151, 44.
15. De La Calleja, J., & Fuentes, O. 2004, Monthly Notices of the Royal Astronomical Society, 349, 87.
16. Dobrycheva, D., Vavilova, I., Melnyk, O., & Elyiv, A. 2017, arXiv preprint arXiv:1712.08955.

17. Fabbro, S., Venn, K., O'Briain, T., et al. 2018, Monthly Notices of the Royal Astronomical Society, 475, 2978.
18. Feurer, M., & Hutter, F. 2019, in Automated Machine Learning (Springer, Cham), 3–33.
19. Fiorentin, P. R., Bailer-Jones, C., Lee, Y. S., et al. 2007, Astronomy & Astrophysics, 467, 1373.
20. Gauci, A., Adami, K. Z., & Abela, J. 2010, arXiv preprint arXiv:1005.0390.
21. Glazebrook, K., Offer, A. R., & Deeley, K. 1998, The Astrophysical Journal, 492, 98.
22. He, K., Zhang, X., Ren, S., & Sun, J. 2015, in Proceedings of the IEEE international conference on computer vision, 1026–1034.
23. Hewett, P. C., & Wild, V. 2010, MNRAS, 405, 2302, doi:10.1111/j.1365-2966.2010.16648.x.
24. Higley, A. N., Lyke, B. W., Myers, A. D., et al. 2020, in American Astronomical Society Meeting Abstracts, Vol. 235, American Astronomical Society Meeting Abstracts #235, 219.03.
25. Hoyle, B. 2016, Astronomy and Computing, 16, 34.
26. Hutchinson, T. A., Bolton, A. S., Dawson, K. S., et al. 2016, AJ, 152, 205, doi:10.3847/0004-6256/152/6/205.
27. Kim, E. J., & Brunner, R. J. 2016, Monthly Notices of the Royal Astronomical Society, stw2672.
28. Kingma, D. P., & Ba, J. 2014, arXiv preprint arXiv:1412.6980.
29. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278.
30. Li, X.-R., Pan, R.-Y., & Duan, F.-Q. 2017, Research in Astronomy and Astrophysics, 17, 036.
31. Liu, C.-L., Hsaio, W.-H., & Tu, Y.-C. 2018, IEEE Transactions on Industrial Electronics, 66, 4788.
32. Odewahn, S. C., Stockwell, E., Pennington, R., Humphreys, R. M., & Zumach, W. 1992, in Digitised Optical Sky Surveys (Springer), 215–224.
33. Pasquet-Itam, J., & Pasquet, J. 2018, A& A, 611, A97, doi:10.1051/0004-6361/201731106.
34. Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, Publications of the Astronomical Society of the Pacific, 128, 104502.
35. Sharma, K., Kembhavi, A., Kembhavi, A., et al. 2020, Monthly Notices of the Royal Astronomical Society, 491, 2280.
36. Xu, B., Wang, N., Chen, T., & Li, M. 2015, arXiv preprint arXiv:1505.00853.
37. Xu, L., Ren, J. S., Liu, C., & Jia, J. 2014, Advances in neural information processing systems, 27, 1790.

Cosmological density field emulation and gravitational wave inference based on dimensionality reduction and supervised machine learning

Miguel Conceição

Instituto de Astrofísica e Ciências do Espaço,
Departamento de Física, Faculdade de Ciências, Universidade de Lisboa
Lisbon, 1749-016 Campo Grande, Portugal
E-mail: fc46798@alunos.fc.ul.pt

António da Silva

Instituto de Astrofísica e Ciências do Espaço,
Departamento de Física, Faculdade de Ciências, Universidade de Lisboa
Lisbon, 1749-016 Campo Grande, Portugal

Alberto Krone-Martins

Department of Informatics, University of California, Irvine
Irvine CA 92697, USA

Two major challenges in modern cosmology involve understanding the origin and growth of Cosmic structure and the progenitors of Gravitational Waves. Both scenarios currently require heavy computational resources to perform simulations and inference. In this work, we adopt simple Machine Learning methods to alleviate these requirements, to enable significantly faster sampling and inference. We show that using Dimensionality Reduction and simple Supervised Learning methods, it is possible to generate high-precision emulations of density fields given a set of parameters (such as the Dark Matter density parameter and redshift). Our method provides orders of magnitude improvement of CPU run times and much less computational resources when compared with N-Body simulations or more complex supervised learning approaches. We also show that it is possible to generate fast inference of gravitational wave parameters (such as the Chirp Mass) from Binary Black Hole systems using the same method. This method provides a promising approach to fast emulation and parameter inference to further explore in the context of upcoming large surveys like Euclid, LSST/Rubin, and LISA.

Keywords: N-body Simulations, Gravitational Waves, Supervised Machine Learning, Dimensionality Reduction, Emulation, Inference.

1. Introduction

Two main challenges in modern Cosmology involve understanding the origin and growth of Large Scale Structure (LSS) and the physics behind the progenitors of Gravitational Waves (GW). To study these phenomena, usual approaches rely on the generation of mock data for comparison with observations and consequent inference of parameters concerning the underlying preferred models.

Regarding Structure Formation, mock data is usually generated with N-body simulations, which are the most powerful method to study the non-linear evolution of large scale structure, within a huge variety of cosmological scenarios. With upcoming surveys like Euclid,[1] the urgency to generate large databases of these simulations for statistical analysis becomes even higher. One serious limitation is that simulations take large amounts of computational resources and time, making unfeasible their direct adoption in tasks that require the realization of multiple simulations and wide explorations of parameter spaces. To overcome these limitations, we propose the application of simple Machine Learning methods to generate high accuracy emulations of N-body density fields in a minimal amount of time, given a chosen set of cosmological parameters, more specifically the Dark Matter Density (Ω_{dm}) and redshift (z).

Similarly to the N-body case, current approaches for parameter inference from gravitational waveforms also rely on computationally heavy and sometimes model-dependent methods. With GW detectors like LIGO[2] and the upcoming LISA[3] mission there is also the urgency for the development of methods that can efficiently extract information from this type of data. Consequently, we aim at generalizing our proposed N-body approach to enable fast and accurate inferences of astrophysical parameters of the waveforms progenitors, namely, the Chirp Mass (M_{chirp}).

To achieve both goals we combine a simple data projection method called Principal Component Analysis (PCA)[4–6] with Supervised Machine Learning techniques, requiring only an initial dataset of N-body simulations and approximated waveforms. PCA enables us to take advantage of the redundancy intrinsic to the data by compressing the dimensional size of input features before providing them as an input to the Machine Learning methods, which learn the relationship between the compressed representation, i.e. the Principal Component coefficients, and the desired cosmological/astrophysical parameters for emulation/inference.

After convergence is achieved, we can use the models to infer or emulate previously unknown regions of the parameter space, which can be compared with ground truth examples not included in the training data, to gauge the performance of the methods. To test different approaches and discriminate which was more appropriate for our analysis we explored four different machine learning algorithms, Random Forest,[7] Extremely Randomized Trees,[8] Support Vector Machine[9] and Neural Networks.[10] Additionally, we also explore the possibility of using *Functional Principal Component Analysis* (FPCA) to perform the emulation of N-body density fields, a method that combines the capabilities of both the PCA and the Supervised Learning approach, providing better accuracy at the cost of higher computational memory.

2. Methodology

This Section presents the method that we developed and the approaches that we used to generate our training and test data sets. To ensure that our methods are effective even in infrastructures with limited resources, all work presented here was created, run, and tested on a regular desktop computer with an Intel Core i5-4460

(4 cores, 4 threads), with 16Gb RAM at 2400 Mhz, running the Linux Ubuntu 18.04LTS operating system, without discrete high-performance GPU cards.

The whole pipeline includes codes in different programming languages. These include `Fortran`, for generation the N-body simulations and power spectrum computation from 3-dimensional density volumes; `Python`,[11] to generate GW waveforms, noise models, and to compute density field bispectra; and, finally, the `R` programming language[12] for the new statistical learning pipeline that we propose. This also includes R libraries for dimensionality reduction, regression models and parameter estimation, that are written in `C`, `C++` and `Fortran`.

In the following Sections, we provide a detailed description of our pipelines.

2.1. *Adopted Machine Learning Algorithms and Optimization*

To build our pipeline, we tested four classical machine learning algorithms, Random Forest,[7] Extremely Randomized Trees,[8] Support Vector Machine[9] and Neural Networks.[10] Each algorithm is governed by a set of hyper-parameters that define the properties of the learning process and the structure of the model. Consequently, we optimized the combination of these hyper-parameters using three different optimization approaches, Grid-Search, Bootstrap and K-fold Cross-Validation, ultimately choosing the one which provided the best results.

For the Random Forest algorithm we used the `randomForest` package,[13] while optimizing the number of trees in the ensemble (`ntree`), the number of features tried at each split (`mtry`) and the minimum size of the terminal nodes (`nodesize`).

For the Extremely Randomized Trees, our implementation used the package `extraTrees` while optimizing the number of random cuts at each split (`numRandomCuts`) and similarly to the previous case, the number of trees in the ensemble and features tried at each split.

The Support Vector Machine models were built using the package `e1078`[14] and the hyper-parameters optimized were the kernel `kernel`, the margin/error `cost` and the amount of curvature allowed in the decision boundary of the Gaussian Radial Basis Function Kernel (`gamma`).

Finally, for the Neural Networks, our implementation used the package `nnet`,[15] and we optimized the number of perceptrons in the single layer (`size`) and the decay of the weights (`decay`).

For the particular case of Gravitational Wave parameter inference, since the Principal Components are used as independent variables, they were also optimized. Here we used a simple Grid-Search.

2.2. *The N-body Emulation Pipeline*

In the first half of our work, we attempted to estimate N-body dark matter density fields using our proposed compression-based machine learning methods.

We consider two scenarios. In the first scenario, we attempt to estimate the density fields considering just one cosmological parameter as the independent variable

used to build the regression models, the Dark Matter density (Ω_{dm}). In the second scenario, we extend the pipeline to include the redshift (z) as a second cosmological variable and attempt to use both as independent variables to estimate the density field for different values of Ω_{dm} and at a continuous range of z.

2.2.1. *Generation of the Training and Test Datasets*

We generated the N-body simulations using a public version of the Hydra code,[16] which solves the Newtonian equations for a set of mass particles of Dark Matter. Every simulation has the same initial conditions, starting from the first snapshot at $z = 49$ until reaching $z = 0$, summing up to a total of 10 snapshots. We generated the fields for 160^3 particles of dark matter initially positioned on a regular cubic grid of comoving length $L = 100\ h^{-1}$Mpc on the side, and every snapshot contains positions and velocities for each one of them.

The cosmological model assumed was the Λ-*CDM* model with $\Omega_k = 0$, $\Omega_m = 1 - \Omega_\Lambda$, $\sigma_8 = 0.809$, and $h = 0.7$, where σ_8 the stardand deviation of density fluctuations when convoluted with a Kernel at a scale of $8\ h^{-1}$Mpc and h is the Hubble parameter (i.e. $h = \mathrm{H}_0/100$).

After calculating the positions and velocities for each particle using the Hydra code, the data is compiled by another code called *darkdens* which calculates the densities of the particles using the *Smooth Particle Hydrodynamics* (SPH) formalism.[17] Finally, the data goes through yet another fundamental piece of code which computes the density field in 128^3 cubic voxels, in units of $10^{27}\ Kg\ m^{-3}$.

In the first scenario of emulation from a single free-parameter, as a demonstration of the validity of our method, it was enough to generate a small set of 24 simulations, setting one of them apart to evaluate the performance of the models. The 23 simulations in the training set were generated with Ω_m varying in the interval $\Omega_m \in [0.05, 0.6]$ with $\Delta\Omega_m = 0.025$, while our test simulation was generated with the density value from the Planck collaboration 2015,[18] corresponding to $\Omega_m = 0.309$.

For the scenario of multiple parameter (Ω, z) density field emulation, the training set we adopted in this work comprises 23 Ω_m density fields spanning four redshifts, $z = [0, 0.25, 0.75, 1]$, while $z = 0.5$ for $\Omega_m = 0.309$ was held for testing. Consequently, the final training dataset will consist of a total of 92 density cubes.

2.2.2. *The Density Field Emulation*

After creating the N-body simulations, we then proceed to load our training set into a matrix to perform dimensionality reduction using PCA, while forcing the data to be scaled and centered, and apply it to our training matrix in its logarithmic form. For that purpose we use the function `prcomp` from the package `stats`.

As a result of the PCA analysis, we have five objects. A vector containing the standard deviations of each Principal Component. Two vectors containing respectively the values used in the centring and scaling of our original data matrix. The

matrix containing the *loadings* (in each column), holding the correlations between the original data and the PCs. And the scores matrix containing the PC scores, or in other words, our data projected in the principal directions. If we consider the loading and scores matrices to be given by $\mathbf{L}$ and $\mathbf{S}$, the scale and centre vectors to be given by $\mathbf{s}$ and $\mathbf{c}$, to obtain our original data representation the scores must be matrix multiplied by the loadings with a posterior sum of the centre vector and multiplication of the scaling one, yielding:

$$\boldsymbol{M_{orig}} = 10^{(\boldsymbol{SL^T s^T + c^T})} \tag{1}$$

Once we have our projected quantities, we use them to optimize and train the regression models. Since we aim to estimate the field itself, in the case of single free-parameter emulations the training is performed using the PCs as dependent variables and Ω_m's as independent ones. Once trained, the models are used to predict the new set of PCs for $\Omega_m = 0.309$ (or any other Ω_m), which can be de-projected into the final estimated cube. To achieve this, we created an iteration loop wherein each iteration our four different algorithms build a regression model accounting for that specific PC and the same set of Ω_m values.

To obtain the final PC predictions using these models, we rely on the function `predict` from the R package `stats`. This function accepts as inputs the regression model and the predictor variables, which in this case correspond to one of the four algorithms and $\Omega_m = 0.309$, respectively.

For the case of the two free-parameter regressions, it is just a matter of adding the redshift as a second independent variable, providing the correspondent training redshift vector to the supervised learning algorithms and the new test redshift (z=0.5) plus the trained models to the `predict` function. Keep in mind that in this case, we are training with 4 distinct redshifts, and thus 4×23 simulations, as mentioned in the previous section. After having the new PCs, the four final estimated cubes are obtained, by de-projecting them using Eq. 1.

2.2.3. *Testing the Models*

To test the emulation performance of our methods we need to define an appropriate metric for comparison against the ground-truth N-body simulation. Consequently we chose to use both the Power Spectrum and Bispectrum for this comparison.

The power spectrum corresponds to the Fourier Transform of the two-point correlation function defined by:

$$\xi(r,t) = \frac{V}{(2\pi)^3} \int \langle |\delta_k(\vec{k},t)|^2 \rangle e^{-i\vec{k}\cdot\vec{r}} d^3k \tag{2}$$

where V is the considered volume and $\delta_k(\vec{k},t)$ corresponds to the over-density field. Thus the power spectrum yields:

$$P(k,t) = \langle |\delta_k(\vec{k},t)|^2 \rangle = \frac{1}{V} \int \xi(r,t) e^{i\vec{k}\cdot\vec{r}} d^3r \tag{3}$$

this quantity gives a complete statistical description of a Gaussian Field. The Bispectrum corresponds to the Fourier Transform of the three-point correlation function and, as an higher order correlation function, it is sensitive to departures from Gaussianity. It is usually defined as B, yielding:

$$\langle \delta_{\vec{k}_1} \delta_{\vec{k}_3} \delta_{\vec{k}_3} \rangle = \delta_D(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) B(k_1, k_2, k_3) \tag{4}$$

where δ_D is the Kronecker delta function.

Since Dark Matter density contrast fields are considered to be well approximated by Gaussian Random Fields at the extreme of large scales (linear regime), and as we travel to smaller scales we begin to observe departures from Gaussianity (non-linear regime), this two quantities are quite appropriate to be used as testing metrics in the context of this work.

During the optimization process, each explored range and optimization scheme will provide us a different optimal model, in order to discriminate between the different models we compared their estimated outputs against our test simulation using the following quantity:

$$PSD_{RD} = \sum_{i=1}^{n} \left| 1 - \frac{y^i_{est}}{y^i_{sim}} \right| \tag{5}$$

where $\mathbf{y_{est}}$ and $\mathbf{y_{sim}}$ are the power spectrum vector of our estimated field and the ground truth, respectively, n corresponds to the power spectrum vector size, y^i_{est} and y^i_{sim} are the ith power spectrum vector cell value for both the estimated and simulated vectors.

Consequently, to have an estimation as close to the ground truth as possible, we need this quantity as close to zero as possible, and thus, given a set of different optimal models, the best performing will be the one with a minimal PSD_{RD}.

2.2.4. *FPCA: An Alternative Approach*

In an attempt to increase the efficiency of our pipeline and explore less traditional approaches, we also demonstrated an application of *Functional Principal Component Analysis*, in the single free-parameter emulation context.

This method provides a continuous function relating the Principal Components with our desired free-parameter, thus combining the compression capabilities of PCA with the predictive usefulness of the Machine Learning methods, and consequently, enabling us to perform a "short-cut" in our pipeline by discarding the Supervised Machine Learning step. Here we adopt the R package `fda`[19] in a process divided into three steps, as follows.

In the first step we build a *Functional Data Object* from the training data. To do so we first need to specify the set of basis functions used to compute the FPCA. To follow our application of standard PCA, we defined 23 b-spline basis functions.

After having the basis functions, we use them to convert the training data (in logarithmic scaling) into the final *Functional Data Object* where the harmonics/FPCA's will be evaluated.

Once the harmonics are obtained, we apply them to compute the Principal Components (PCs) for $\Omega_{dm} = 0.309$, following the de-projection into the original density cube by multiplying each PC by each column of the scores matrix returned by the FPCA decomposition.

One important remark is that this method is highly expensive in terms of memory. Thus, we evaluated it in lower-resolution simulations of 64^3 density cells, instead of the 128^3 density cell simulations used in the previously other pipelines.

2.3. *The Gravitational Wave Inference Pipeline*

In this Section, we describe the methodology we adopted for the Gravitational Wave parameter inference. We demonstrate our implementation in Frequency Domain waveforms. To probe the generalization capabilities of our pipeline, all algorithms were trained in waveforms *without noise* but tested for Chirp Mass inferences in a dataset of waveforms *imbued with noise*, with decreasing S/N ratios.

In the following Subsections, we will be describing the process of waveform and noise generation, the GW parameter inference pipeline and finally, the performance evaluation of our pipeline.

2.3.1. *Generation of the Training and Test Datasets*

We generated a training set of 256 waveforms and a test set of 255 waveforms, varying the chirp mass in the ranges $M_{chirp} = [9.13; 25.80]\ M_{\odot}$ and $M_{chirp} = [9.19; 21.79]\ M_{\odot}$, respectively.

One crucial aspect of our work is that we tweak the S/N ratio by varying the distance to the source, consequently we generated 3 different test sets for the mentioned test Chirp Mass range, while only varying the distance parameter (eg. 1 Mpc, 100 Mpc and 1 Gpc).

Apart from the distance and masses, it is also necessary to provide high and low frequency cutoffs, the frequency resolution, and the frequency domain waveform approximator. To follow the LIGO detector parameters, the low and high frequency cutoffs were chosen to be, respectively, 10 Hz and 10kHz. Finally, we set the frequency resolution to $\Delta f = \frac{1}{16}$, and choose the *TaylorF2*[20] approximant to generate the simulated waveforms.

2.3.2. *Noise Generation and Data Pre-Processing*

To colour the waveforms with noise, we need to generate an analytical power spectral distribution (PSD) following the expectations for the Advanced Ligo detector. To do so, we use an analytical Gaussian Noise simulation, as implemented by the function `analytical.aLIGOZeroDetHighPower` from the package `pycbc.psd`. We

set the length of the PSD (l_{PSD}), the frequency resolution ($\Delta f = \frac{1}{16}$) and the low-frequency cutoff ($f_{low} = 10\ Hz$).

The noise is then generated from the defined PSD using the `pycbc.noise` package and its `frequency_noise_from_psd` function. After obtaining the noise vector, we add it to the strain obtained as from the waveform generation.

Finally, we resize the simulated waves by cropping all the strain values for frequencies above 51 Hz, which corresponds to the limiting frequency domain of the shortest wave. Thus, each frequency domain waveform results into 656 data points.

2.3.3. *The Inference Pipeline*

Once we generate all waveforms, they are organized to form a training matrix composed of 256 instances and a test matrix with 255 instances, each corresponding to a matrix row. Similarly to the cosmological density fields case, the next step consists of performing dimensionality reduction. This data projection results into a compact representation, where each waveform is represented by 256 PCs.

In this case we are dealing with smaller matrices than in the N-body case. Thus the PCA projection step is significantly faster. However, as now a higher number of instances is considered (e.g 256 instead of 23), the final projection will result in a higher number of Principal Components.

After producing the compressed data representation, we proceed to the machine learning regressions. Different from the N-body emulation scenario, here we are not attempting to emulate waveforms, but we are instead attempting to infer parameters concerning those waveforms given their strain vectors.

Consequently, instead of using the PCs as dependent variables, now they are used as independent variables, while the chirp masses are the dependent variables for prediction. That implies a large simplification in our pipeline since we are not forced to build a regression model for every single PC.

Also, in this part of the work, we excluded the neural networks. Since in this scenario we have 256 dimensions/independent variables, as opposed to the N-body context where we have a maximum of 2 independent variables (e.g Ω_m and z), here the single-layer neural network is expected to highly under-perform.

After we train and optimize the models using the training set, they are used to infeer the parameters for the not previously seen waveforms. To do so we need to first project the test waveforms in the PC basis of the training set, were the same centering and scaling vectors as learnt from the training set are used, obtaining the PC vector for a specific waveform as:

$$\boldsymbol{PC_{new}} = \boldsymbol{L^T}\left(\frac{\boldsymbol{t_{wf}} - \boldsymbol{c}}{\boldsymbol{s}}\right) \tag{6}$$

where, as in the Density fields case, $\boldsymbol{L^T}$ corresponds to the loadings (or rotation) matrix provided by the PCA, $\boldsymbol{s}$ the PCA scaling vector, $\boldsymbol{c}$ the PCA centering vector and $\boldsymbol{t_{WF}}$ to the test waveform vector for chirp mass inference.

After having our 255 inferred chirp masses we proceed to evaluate the algorithms.

2.3.4. *Testing the Models*

In this case of Chirp Mass inferences, since we are dealing with discrete instances represented by a real number, our choice for evaluation metric was the simple Root Mean Squared Error, given by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y_i} - y_i)^2}{n}} \tag{7}$$

where $\hat{y_i}$ and y_i correspond, respectively, to the ith estimated and ground-truth chirp masses and n is the total number of instances.

Here, we use this metric to discriminate between the different models generated in the optimization process and also as the final evaluation measure, used to discriminate between the different algorithms explored in this approach.

3. Results

In this Section, we present the results of our work. We first show the results concerning the N-body dark density field estimations, and following that, the results regarding the GW parameter inference.

We present our final optimized regressions, and compare the performance of the algorithms against the ground-truth using the evaluation metrics described in the previous section.

3.1. *N-body Density Fields*

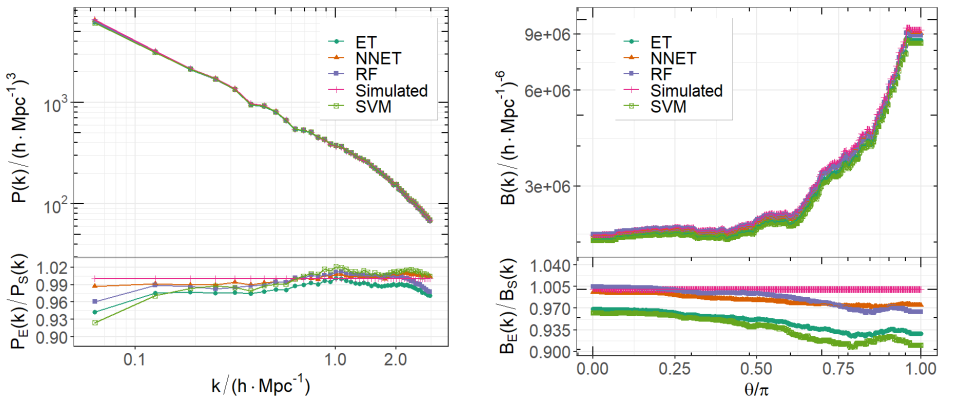
We begin by showing the results concerning our N-body Density Field Emulations.

First we present the results concerning our main pipeline, coupling PCA with the four supervised learning methods in the two emulation scenarios, with one and two free parameters. Lastly, we show the results regarding the FPCA emulations, implemented in the reduced size dataset.

3.1.1. Ω_{dm} *Emulations*

In Figure 1 we show the main results from the four supervised learning emulations. On the top panels, left-hand side we see the Power Spectra of our estimators against the ground-truth while on the right-hand side we see the Bispectra. The lower panels show the ratio between the spectra of the emulations and the ground-truth.

Regarding the Power Spectra, we achieved an accuracy within 99% throughout all of the k domain using the Single-Layered Neural Network, which was the best performing algorithm. This is impressive, because these results show that it is possible to attain very low errors even at extreme scales as $k \geq 2$ hMpc^{-1} and nearing $k \sim 0.3$ hMpc^{-1}, while using one of the most simple types of neural networks, a single-layer network. We note, however, that this is only possible because of the compression step that was adopted here.



(a) (b)

Fig. 1. Results for the single free-parameter emulation scenario implemented at z=0. (a) Power Spectrum Comparison. (b) Bispectrum Comparison.

Looking at the lower panel, it is possible to observe that generally, the algorithms seem to have a sharp decrease in their performance at the large scale extreme ($k \leq 0.2$ hMpc^{-1}), which can be attributed to the lack of independent modes in the training dataset.

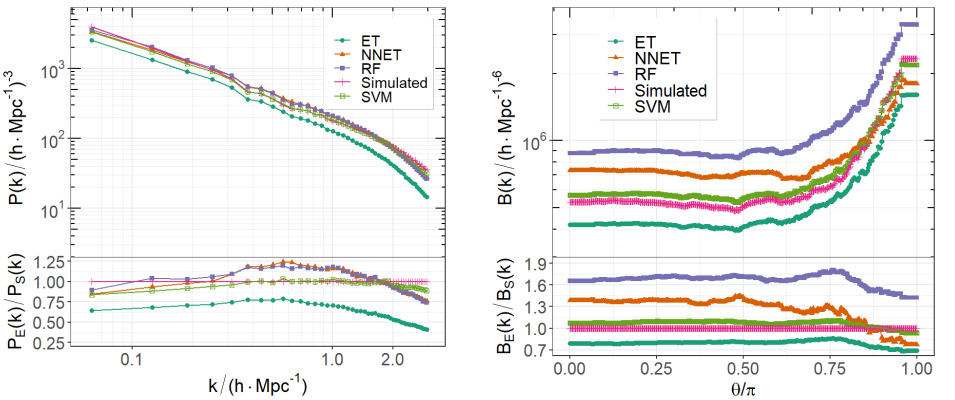
On the opposing small scale extreme, we can also observe a slight decrease in performance, this time due to leakage of the Nyquist modes, as we approach the resolution limit of the simulations.

Regarding the Bispectra, we obtained an accuracy within 97% throughout all of the θ domain using the Neural Networks. Here we see Random Forest excelling at lower θ's ($\frac{\theta}{\pi} \leq 0.75$), but its accuracy eventually falls below 97% at the large θ extreme. The reason why we observe such an overall decrease in performance in the Bispectra case, compared to the Power Spectra can be attributed to the fact that the Bispectrum is non-linear in nature, giving a description of departures from Gaussianity, as opposed to the Power Spectrum. Since the Bispectrum deals with the non-linear aspects of the density fields, we should expect an higher challenge in reproducing it faithfully, since the non-linear features should be harder to compress via a simple PCA method and thus, later to predict via Supervised Learning. Non-linear compression methods would allow these results to be improved significantly.

3.1.2. (Ω_{dm}, z) *Emulations*

Figure 2 shows the results for the two-free parameter emulation scenario.

In the case of the two-free parameter emulation scenario we observe a decrease in the performance of the algorithms from the previous scenario. This can be expected, as now an increased dimensionality is faced. Also, the simple neural network is no



(a) (b)

Fig. 2. Results for the two free-parameter emulation scenario. (a) Power Spectrum Comparison. (b) Bispectrum Comparison.

longer the best performing algorithm, and its adoption results in errors reaching 25% at around $k = 0.6$ hMpc^{-1}.

It is however remarkable how well the SVM method performed in this scenario. Here, the SVM achieves an accuracy of more than 95% for $0.3 \leq k \leq 2$ hMpc^{-1}, decreasing to about 90% at the low scale extreme and $\sim 85\%$ at the large scale one.

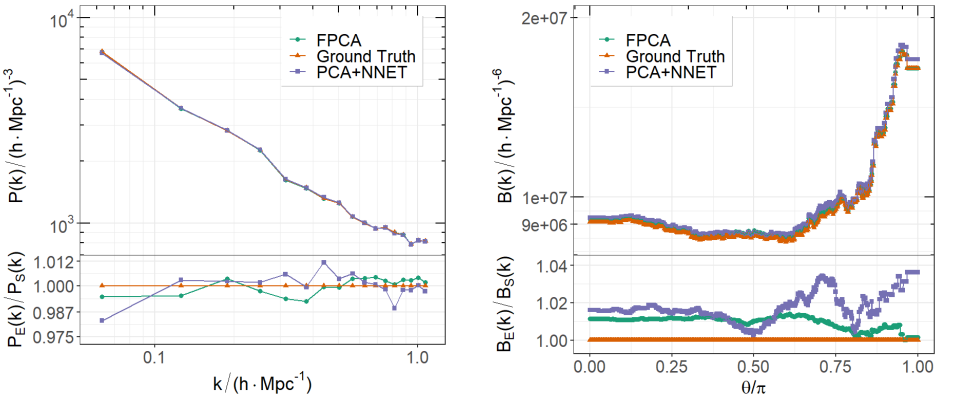
Finally, the observed decrease in overall accuracy between the previous scenario to the present, multidimensional one, can be explained due to the so-called *Curse of Dimensionality*. To reach similar performances compared to lower-dimensional cases, a required increase in the size of the training set can be exponential, when the number of dimensions in the feature space increases.

Regarding the Bispectrum, the results further indicate that SVMs are the best method among those studied in this work, since as we can see, even looking at higher order correlations it still out-performs the remaining algorithms. It is the only algorithm able to maintain an accuracy below 15% for the entire θ domain.

Finally, one additional interesting feature we see in this result, is the fact that the relation between the error and the θ domain seems to be inverted. While in previous, one dimensional case, the errors tend to increase for larger angles (corresponding to larger scales), here they tend to decrease. This could possibly be due to the increased size of our training dataset. By increasing the amount of full simulations in our training data we are effectively increasing the number of large scale modes, and consequently, improving the predictions of the algorithms at such scales.

3.1.3. *FPCA Results*

Finally, we close the N-body section by showing the results regarding our FPCA demonstration. The Power Spectra and Bispectra are shown in Figure 3,



(a) (b)

Fig. 3. FPCA emulation results at z=0. (a) Power Spectrum Results. (b) Bispectrum Results.

compared against an optimized Single-Layered Neural Network and the ground-truth simulation.

We can observe from Figure 3 that even when compared against the best performing algorithm of our previous results, FPCA seems to attain an slightly better performance, particularly at the large scale regime, indicating a strong capacity of the algorithm for generalization.

Additionally, FPCA seems to be more stable than the neural network, in the sense that we do not observe the frequent oscillations in the power ratio that we can observe in the neural network case. This could be due to the smoothing process of the b-spline basis functions that takes place within the FPCA.

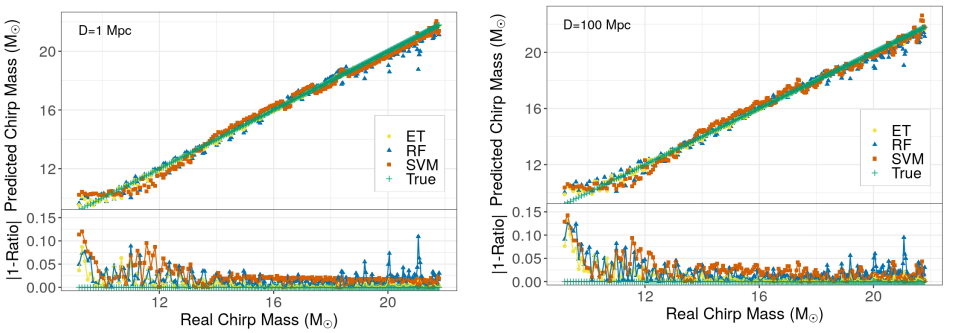
Finally, one interesting thing to notice when looking at the error in terms of overestimation and underestimation of the power is the fact that the neural network and the FPCA seem to roughly exchange roles. When one underestimates, the other overestimates. This suggest that it would be possible to further improve the results by merging these two methods into an ensemble or a stacked generalization.[21] Since, roughly, their average results seem to overlap with the simulated Power Spectrum.

3.2. *Gravitational Waves*

Here we present the results from our Gravitational Wave inference pipeline. We show results concerning 255 Chirp Mass inferences from waveforms emitted at increasing distances and consequently, decreasing signal-to-noise ratios.

3.2.1. *Frequency Domain Inferences*

In Figure 4 we can see the results regarding inferences from waveforms emitted at 1 Mpc (left hand-side) and 100 Mpc (right hand-side). The top panels show our



(a) (b)

Fig. 4. Comparison between the estimated chirp masses and the ground truth in Frequency Domain. (a) For waveforms emitted at 1 Mpc (b) For waveforms emitted at 100 Mpc.

predictions against the ground-truth while the lower panels show a measure of the ratios.

These results show that up to 100 Mpc and for $M_{chirp} \geq 14$ we are able to obtain an accuracy within 98% using our best performing model which in this context was the Extremely Randomized Trees.

The Support Vector Machine also manages to obtain stable results throughout the mentioned domain, with an accuracy within 97%. On the other side, Random Forest was the worst performing algorithm, with its accuracy falling below 95%, even in this expected well behaved region of the Chirp Mass domain.

The large increase in error observed at the lower Chirp Mass extreme ($M_{chirp} < 14$) is a consequence of the fact that lower chirp masses result in lower amplitude Strains, which get highly diluted in the noise.

Turning our attention now to inferences in very low signal to noise ratio regimes, Figure 5 represent the results for inferences from waveforms emitted at much higher distances, at 1 Gpc.

Here we see a decrease in the quality of the results, with the previous best performing algorithm (Extremely Randomized Trees) having its accuracy falling to values close to 85% in the interval $14 \leq M_{chirp} \geq 16$ and to 90% for $16 \leq M_chirp \geq 20$.

We can also observe a shift in the "well behaved" domain, now $M_{chirp} > 16$, which is a natural consequence of increasing the distance to the progenitors of the waveforms, translating in a generalized decrease in the signal to noise ratio.

On the other side, the Support Vector Machine manages to excel compared to the other algorithms, achieving an accuracy $\sim 95\%$ for $M_{chirp} > 16$. The results from these simulated experiments reinforce the strong generalization capabilities of SVMs, and are particularly impressive considering that the algorithms were trained in a dataset of waveforms without noise while they are being tested on noisy ones. In part this is due to the data compression step, since by its very nature, it is also working as a denoising procedure.

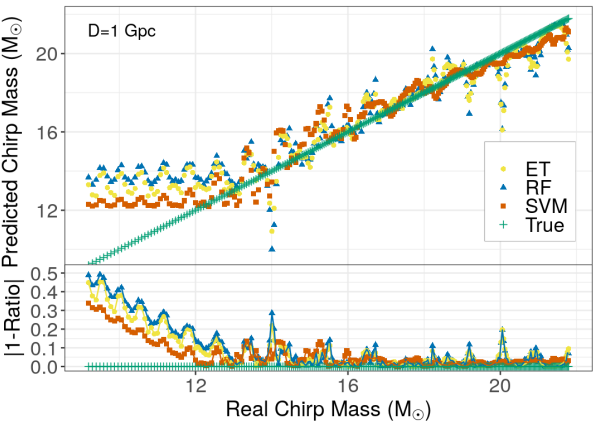


Fig. 5. Comparison between the estimated chirp masses and the ground truth in Frequency Domain for waveforms emitted at 1 Gpc.

4. Discussion

We showed that combining PCA with simple supervised learning methods allows fast and accurate estimation (regression) of cosmological density fields. We also demonstrated that the same method can be used, in the inverse way, to perform parameter inference, applied to the problem of GW source parameter determination.

Indeed, the fast increase in the demand for larger and more diverse sets of cosmological simulations, to study large-scale structure mechanisms and their dependence on Cosmology, demands fast and accurate methods to produce multiple realizations of density volumes. Our results show that our approach achieves this objective. On the other hand, our results demonstrate that our proposed method also provides a fast way of performing GW parameter inference without requiring additional prior assumptions on top of those that were already adopted to build the set of approximated waveforms.

Here we discuss the main results of our work. We comment on the relevance of our method for further developments in these fields, addressing its gains of performance and potential to explore other situations of interest. We also address the method's present constraints and discuss ways to mitigate them, proposing solutions for future implementation.

4.1. *Efficiency*

Here we focus on the efficiency of our method. In other words, we will discuss the amount of CPU time and computational resources required to produce a final result. We begin by showing the CPU run times for the N-body emulations, in both scenarios of single and two free-parameter emulation, and following that we show

the CPU run times for the GW inference pipelines in both low and high signal to noise ratio regimes.

4.1.1. *Cosmological Density Fields*

In Table 1, we show the total time in order to perform the emulations. The first row regards the single free-parameter emulation scenario, where our best performing algorithm was the Neural Network (NNET), while the second row regards the two free-parameter emulation scenario, where the best performing algorithm was the Support Vector Machine (SVM).

Table 1. CPU run times for the N-body Emulation Case.

Emulation	**PC Estimation**	**De-Projection**	**Total Emulation Time**
$\boldsymbol{\Omega_{dm}}$	0.077 s	0.719 s	**0.796 s**
$\boldsymbol{(\Omega_{dm}, z)}$	0.082 s	4.380 s	**4.462 s**

We managed to achieve an high accuracy for the density field estimation, in a total of ~ 0.796 s for the one-dimensional case and ~ 4.462 s for the two-dimensional case. As a comparison, it takes ~ 84 min to perform an N-body simulation of the same scale. Our results, thus, correspond to an improvement of three orders of magnitude in the time required to generate the density field. Moreover, the computational resources required to implement our pipeline are far smaller than those required to perform the N-body simulations. Additionally, these emulations reproduced the real Power Spectra to less than a 1% (NNET in Figure 1 at $0.4 \leq k \leq 3$ and SVM in Figure 2 at $0.4 \leq k \leq 2$).

Finally, with respect to our work on Gravitational Wave inferences, we present the CPU running times of our pipeline in Table 2. The first row regards the high signal-to-noise ratio regime, where the best performing algorithm was the Extremely Randomized Trees (ET), while the second row regards the low signal-to-noise ratio regime, where the best performing algorithm was the Support Vector Machine (SVM).

Table 2. CPU run times for the GW Inference Case.

GW Inference (FD)	**Projections into the PC Basis**	**Inferences**	**Total Time**
D=1 Mpc	0.142 s	0.041 s	**0.182 s**
D=1 Gpc	0.252 s	0.004 s	**0.256 s**

Once again, we achieve a highly significant efficiency in inferring the Chirp Mass of the waveforms, obtain inference times down to 0.183 s for high signal-to-noise

ratio regime. These results are particularly appealing, considering this is the time required to perform 255 simultaneous waveform projections and M_{chirp} inferences.

5. Conclusions

In this work, we presented a simple and computationally light methodology to enable fast sampling and inference of datasets in two gravity dominated contexts, Structure Formation and Gravitational Waves. To achieve that we implemented supervised Machine Learning methods coupled with Principal Component Analysis, on training datasets of full 3D N-body simulations and approximated Gravitational Waveforms.

In the context of structure formation, we provided contributions to the field using our methodology for fast and accurate emulations of full N-body dark matter density fields of 128^3 density cells in a box of 100 h^{-1} Mpc, given a choice of cosmological parameters. We applied four supervised Machine Learning Algorithms, Random Forest, Extremely Randomized Trees, Support Vector Machine and Neural Networks, in two different emulation scenarios. In the first scenario, we performed emulations solely based on the Dark Matter Density (Ω_{dm}) parameter. We managed to reproduce the Power Spectrum of the corresponding simulated field to a less than a 1% difference for $k > 0.3$ hMpc^{-1} in all redshifts, and reproducing the Bispectrum for all θ domain to less than a 3% difference for $z = 0$ and less than a 2% difference for the remaining redshifts, using the best-performing algorithm (Neural Networks). We additionally studied Functional Principal Component Analysis in the context of structure formation. We achieved even better results than the supervised methods for a reduced size dataset of simulations with 64^3 density cells, managing to reproduce the Power Spectrum and Bispectrum to less than a 2% and 5% difference from the ground truth, at all redshifts. In this scenario we trained our algorithms with 23 N-body simulations in the range $\Omega_{dm} = [0.05; 0.6]$ with $\Delta\Omega_{dm} = 0.025$, and tested the emulations for $\Omega_{dm} = 0.309$.

In the second scenario, we introduced an additional dimension, adding the redshift as free-parameter. Using the best-performing algorithm in this scenario (Support Vector Machine), we managed to reproduce the Power Spectrum to less than a 5% difference for $0.3 \leq k \leq 2$ hMpc^{-1} and the Bispectrum to less than a 15% difference in all θ domain. Here we trained the algorithms with simulations in the same Ω_{dm} range as in the first scenario, but in four distinct redshift snapshots ($z = 0$, $z = 0.25$, $z = 1$, $z = 10$), increasing the training dataset by a factor of four. We tested the emulations against an $\Omega_{dm} = 0.309$ simulation at redshift $z = 0.5$, although the emulator can be used for any value of Ω_{dm} and z.

We thus demonstrate that percent level accuracies can be reached using these emulators, and that moreover emulations can be performed in 0.796 s in the first scenario and in 4.462 s for the second scenario. This corresponds to an improvement of three orders of magnitude when compared to the time required to perform an N-body simulation of the same scale.

In the context of Gravitational Waves, we inverted the pipeline to perform inference of the Chirp Masses for a set of Frequency Domain simulated waveforms. We managed to achieve a high generalization capacity in our models by training them with 256 denoised waveforms, and testing on 255 waveforms imbued in theoretical LIGO detector noise, for three different datasets at decreasing signal-to-noise ratios. In this case we achieved differences in the inferences of around 2% for the higher signal-to-noise contexts and for $M_{chirp} > 14\ M_{\odot}$, while around 5% for lower signal-to-noise ratio contexts and for $M_{chirp} > 16\ M_{\odot}$. For both high and low signal-to-noise ratios, the parallel inference of 255 independent waveforms is performed in less than 0.3 seconds. The main success in these Gravitational Wave parameter inference results, however, reside in the generalization capabilities of our proposed methodology, since while it was trained on noise-free training sets, it was tested on highly noisy test sets.

Acknowledgments

The authors acknowledge the support of the Portuguese Fundação para a Ciência e a Tecnologia (FCT) for funding under the following Projects and Research Contracts with References: EXPL/FIS-AST/1368/2021, UID/FIS/00099/2019, CERN/FIS-PAR/0037/2019, PTDC/FIS-OUT/29048/2017, PTDC/FIS-AST/31546/2017, IF/01135/2015.

References

1. R. Laureijs et al., Euclid Definition Study Report, *arXiv e-prints*, p. arXiv:1110.3193 (October 2011).
2. LIGO Scientific Collaboration, Advanced LIGO, *Classical and Quantum Gravity* **32**, p. 074001 (April 2015).
3. P. Amaro-Seoane et al., Laser Interferometer Space Antenna, *arXiv e-prints*, p. arXiv:1702.00786 (February 2017).
4. K. P. F.R.S., Liii. on lines and planes of closest fit to systems of points in space, *Philosophical Magazine Series 1* **2**, 559.
5. H. Hotelling, Analysis of a complex of statistical variables into principal components., *Journal of Educational Psychology* **24**, 498 (1933).
6. I. Jolliffe and Springer-Verlag, *Principal Component Analysis*, 2 edn. (Springer, 2002).
7. L. Breiman, Random forests, *Machine Learning* **45**, 5 (10 2001).
8. P. Geurts, D. Ernst and L. Wehenkel, Extremely randomized trees, *Machine Learning* **63**, 3 (04 2006).
9. C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning* **20**, 273 (09 1995).
10. W. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biology* **52**, 99 (1990).
11. G. Rossum V. and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
12. R. C. Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2021).

13. A. Liaw and M. Wiener, Classification and regression by randomforest, *R News* **2**, 18 (2002).
14. D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, (2021). R package version 1.7-8.
15. W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, fourth edn. (Springer, New York, 2002).
16. P. A. T. H. M. P. Couchman and F. R. Pearce, Hydra: an adaptive-mesh implementation of p 3m-sph, *The Astrophysical Journal* **452**, p. 797 (1995).
17. L. B. Lucy, A numerical approach to the testing of the fission hypothesis, *Astronomical Journal* **82**, 1013 (1977).
18. Planck Collaboration, Planck 2015 results. XIII. Cosmological parameters, *Astronomy & Astrophysics* **594**, p. A13 (10 2016).
19. J. O. Ramsay, S. Graves and G. Hooker, *fda: Functional Data Analysis*, (2020). R package version 5.1.5.1.
20. F. Messina, R. Dudi, A. Nagar and S. Bernuzzi, Quasi-5.5PN TaylorF2 approximant for compact binaries: Point-mass phasing and impact on the tidal polarizability inference, *Physical Review D* **99**, p. 124051 (June 2019).
21. D. Wolpert, Stacked generalization, *Neural Networks* **5**, 241 (12 1992).

Unsupervised photometric detection of galaxy cluster candidates in large surveys

Ana Carvalho

Instituto de Astrofísica e Ciencias do Espaço,
Faculdade de Ciências, Universidade de Lisboa
Lisbon, 1749-016 Campo Grande, Portugal
E-mail: fc46443@alunos.fc.ul.pt

Alberto Krone-Martins

Department of Informatics, University of California, Irvine
Irvine CA 92697, USA

Antonio da Silva

Instituto de Astrofísica e Ciencias do Espaço,
Departamento de Física, Faculdade de Ciências, Universidade de Lisboa
Lisbon, 1749-016 Campo Grande, Portugal

Galaxy Clusters are essential to study galaxy evolution and are sensitive probes of cosmology and the dynamics of the Dark sector. Large galaxy surveys, such as Euclid, DES, LSST/Rubin will detect many new clusters. For example, the Euclid mission survey may reveal more than 6×10^4 clusters with S/N>3 up to $z \sim 2$, representing a whole new era for cluster cosmology. A large fraction of these clusters will be unknown high-redshift cluster candidates, lacking spectroscopic information. Thus, a major challenge for cluster detection is the identification of possible member galaxies from photometry alone, and ideally without strong assumptions of what a cluster is. Here we present the first results for the detection of galaxy clusters of a modified version of the UPMASK (Unsupervised Photometric Membership Assignment) method. The method, created to study star clusters, uses heuristics and statistical analysis to separate cluster candidate galaxies from other field galaxies without assuming cluster profiles or any strong theoretical cluster priors. We show that the method operates in a fully unsupervised way and it can even work with minimal amounts of astrometry and photometry information, using Euclid-like galaxy survey simulations. We then use Pan-STARRS data to assess the performance of the method to identify Planck clusters and present possible detections of optical counterparts for cluster candidates in the second PlanckSZ data release catalog. Finally, we compare our findings with other Planck cluster candidate follow-up efforts.

Keywords: Methods: data analysis, methods: statistical, galaxies: clusters: general.

1. Introduction

The distribution of galaxies in the Universe is not uniform. They permeate an immense cosmic web that has evolved from an initial density field, continuously sculpted by gravity and cosmological expansion, where galaxy groups and clusters emerge at the interceptions of filamentary structures. The study of these objects, their abundance, and spatial correlations provides unique ways to test cosmological

models, including those featuring dark matter, dark energy, or modified gravity theories (see e.g. Refs. 1, 2). Regarding clusters, their use as cosmological probes in galaxy surveys critically depends on the definition of "cluster" and the separation of their member galaxies from other galaxies in the field. Spectroscopic information can significantly improve cluster identification, but present and near-future surveys lack spectroscopic information for a large fraction of detected objects. Additionally, most cluster identification methods adopt strong *a priori* models and explicitly search for the existence of a Red Sequence (see e.g. Ref. 3). Moreover, some of these *a priori* assumptions provide only the expected behavior of idealized systems, which are sometimes poor approximations to a more complex reality.

2. The method

UPMASK[4] is a method designed to study stellar clusters, that was successfully applied to discover multiple new objects of this type.[5,6] This method was created considering only the availability of minimal data, like photometry and astrometry. However it is also possible to use any other input feature that can help to discriminate objects. The method uses processes like the Principal Component Analysis[7] (PCA), clustering algorithms, Kernel Density Estimation[8,9] (KDE), data resampling, and an iterative process. In this proceeding we report our testings of UPMASK for a membership study in galaxy clusters. We also report on tests done with modified versions of the method (see Sec. 3), aiming at improving the processing speed of the original version, and looking towards the development of a future galaxy cluster search application for large datasets.

The method will firstly re-sample the measured data according to their error distributions and then apply PCA to the photometric parameters. It will cluster the data with k-means[10] in the space defined by the Principal Components. Each mathematical cluster is analyzed in the astrometric space with a KDE and then compared with the KDEs of a uniform random distribution. The method eliminates the KDEs compatible with those of a uniform distribution KDE and iterates. This way, UPMASK can discern clustered objects from field objects, by purifying the cluster data at each iteration. Finally, running this routine multiple times, UPMASK accounts for observational errors/uncertainties, and provides a score (equivalent to a frequentist membership probability) for each object in the dataset.

3. Modifying UPMASK

This method, in particular the KDE step, is computationally time-consuming. Thus, its application to large areas of the sky requires significant computational resources. We studied alternatives to replace the KDE step of the method. We have used tools such as the Voronoi tesselation, Anderson-Darling test and regression functions, as described next.

3.1. *Voronoi selection procedure with Anderson-Darling Test*

Here we apply a Voronoi[11] diagram to the K-means clusters in the astrometry space and then compare the polygon areas distributions of these objects with those obtained from a random uniform distribution. For an illustration of the method, we show in Fig. 1 the Voronoi Tesselations and the distribution of sky areas of the polygons of a random normal distribution and a random uniform distribution. Note that the first is naturally clustered at the center, whereas the latter shows a typical uniform distribution of points. The type of point distribution, therefore, affects the distribution of the areas of the Voronoi polygons. We have adopted a two-sample Anderson-Darling test[12] to perform the comparison between the resulting distributions of polygon areas. We use this test twice: testing the area distribution of the k-mean mathematical cluster against the area distribution of a random uniform realization, and testing the area distribution of random uniform realizations against each other – the later test has performed a maximum of 100 times to account for statistical fluctuations. Each test gives us a p-value that is used to discern between the two cases. A K-means group is classified as a clustered group if

$$p_{kmeans-uniform} \leq \langle p_{uniform-uniform} \rangle - T \times \sigma_{uniform-uniform} \tag{1}$$

where $p_{kmeans-uniform}$ is the p-value of the Anderson-Darling test between the distribution of Voronoi areas of the k-means cluster and the distribution of areas of a random uniform realization, $p_{uniform-uniform}$ is the mean p-value of the Anderson-Darling test of two distributions of areas between 2 random uniform realizations, $\sigma_{uniform-uniform}$ is the standard deviation of p-values from the latter case and T is a threshold level. We have set $T = 3$ as default, but one can modify this parameter as desired. We improve the computational processing speed of this step by including a lookup table for quantities that can be pre-computed before running the method.

3.2. *Voronoi selection procedure with simple statistic comparison*

We have studied another modification, also using Voronoi diagrams. This alternative has a simple statistical comparison between tesselation area distributions instead of the Anderson-Darling Test. As in the previous section, Voronoi tessellation is performed on the K-means clustering points to compute the mean of the distribution of tessellation areas. The same thing is done for a set of random uniform realizations in identical conditions (same total area and number of points) and an average and standard deviation of polygon areas are also obtained. The K-means mathematical group is then assigned as an identified cluster if:

$$\langle N_{kmeans} \rangle > \langle N_{uniform} \rangle + T \times \sigma_{uniform} \tag{2}$$

in which $\langle N_{kmeans} \rangle$ is the mean of the density of the polygons from the Voronoi Tesselation applied to the K-means group, $\langle N_{uniform} \rangle$ is the mean of all the uniform realizations average density of the polygons from the Voronoi Tesselation, $\sigma_{uniform}$

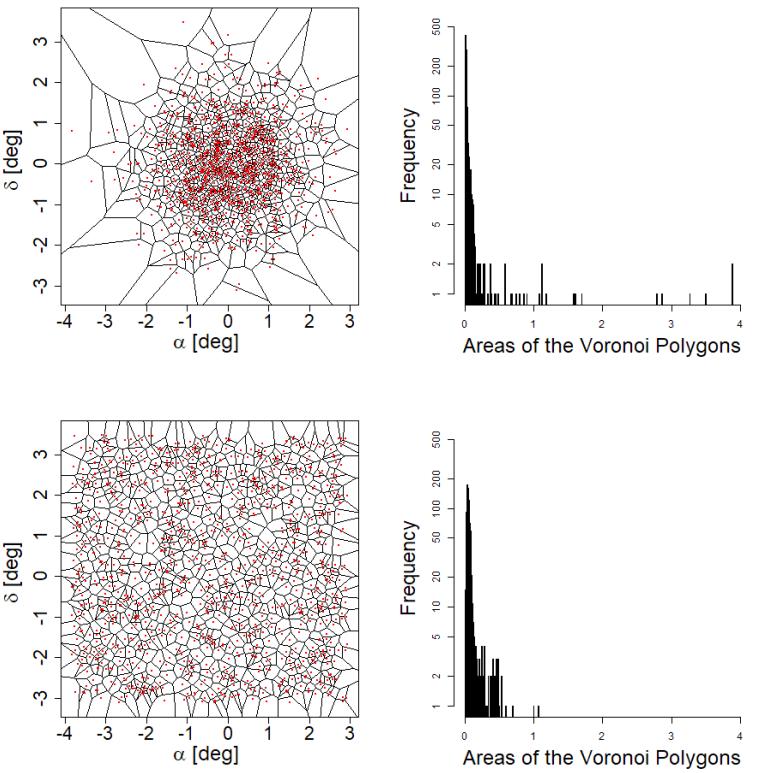


Fig. 1. Top: Left - Voronoi Tesselation (black lines) of a random normal distribution (red points). Right - Histogram of the distribution of polygon areas that resulted from the Voronoi Tesselation on the left. Bottom: Left - Voronoi Tesselation (black lines) of a random uniform distribution (red points). Right - Histogram of the distribution of polygon areas that resulted from the Voronoi Tesselation on the left.

is the standard deviation of all the uniform realizations average density of the polygons from the Voronoi Tesselation and T is a threshold level that can be adjusted by the user. As in the previous subsection, $T = 3$ is chosen to be the default case.

3.3. *Grid selection procedure*

Another approach to replace the KDE part in UPMASK is to create a new version of the method, implementing a "contingency table" as a very simplistic and discrete approximation of the KDE. This approach works as follows. Silverman's rule of thumb[13] is used to provide the width of grid divisions to distribute the K-means cluster points in the astrometry space. Then, a squared grid is generated and the number of points that fall inside each grid cell is counted. The counts are interpreted as a non-normalized frequency. Next, the method computes the distance between the maximum and the mean frequency, normalized to the standard deviation of the count distribution. This is then compared with the average of the same quantity

for multiple random uniform realizations, using the same expression in the original UPMASK version (second equation in section 2 of Ref 4).

3.4. *Grid selection procedure with fast thresholding*

The computation of statistics for the random uniform distributions only depends on the number of points of the realizations and therefore does not depend directly on the input data. As such, and to reduce iteration times, we have applied regression methods to pre-compute a function that provides the threshold according to the number of K-means clustering points. This is obtained by performing 1000 random uniform realizations and computing threshold levels as a function of size as in subsection 3.3. The number, n, of considered points is in the range 1 – 1000 and are more concentrated between 1 – 100, motivated on one hand by the traditional galaxy cluster definition in which a cluster is expected to have more than 50 members and on the other hand by the fact that the data will be divided by the k-means into subsamples. The longer interval accounts for possible group samples that fall inside the 1–100 range. The generated threshold values are represented in Figure 2 and can be fitted by the following, two-branch, polynomial regression function:

$$f(x) = \begin{cases} a_0 + a_1 x + a_2 x^2 & n < 90 \\ b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4 + b_5 x^5 + b_6 x^6 + b_7 x^7 & n \geq 90 \end{cases} \tag{3}$$

The function coefficient values are presented in Table 1. It is worth adding that the creation of this function and its points were generated outside the UPMASK method. As such, the kernel will simply use this function and it does not need to compute the values needed, improving the computational time.

Table 1. Fitting function coefficients of Eq. (3) for the Grid method.

a_0	1.149227	a_1	0.200402	a_2	-0.001189		
b_0	1.351×10	b_1	-7.293×10^{-2}	b_2	4.462×10^{-4}	b_3	-1.443×10^{-6}
b_4	2.529×10^{-9}	b_5	-2.422×10^{-12}	b_6	1.187×10^{-15}	b_7	-2.315×10^{-19}

3.5. *KDE selection procedure with a fitting function*

Following the same procedure and reasoning as in subsection 3.4, we have replaced the KDE random uniform part with a regression function created based on the KDE random uniform realization threshold output for the same range of n used in subsection 3.4. The generated threshold values are represented in Fig. 2, and can be fitted by the following regression function:

$$f(x) = \begin{cases} c_0 + c_1 x + c_2 x^2 + c_3 x^3 & n < 40 \\ d_0 + d_1 x + d_2 x^2 + d_3 x^3 & n \geq 40 \end{cases} \tag{4}$$

where the polynomial coefficients are listed in Table 2.

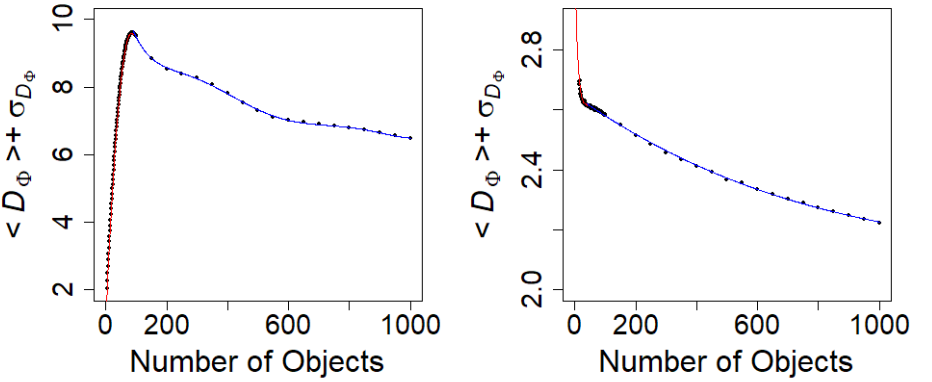


Fig. 2. Fitting function for the Grid (left) and KDE (right) method. In both cases, the red line corresponds to the first branch of the function and the blue line corresponds to the second branch.

Table 2. Fitting function coefficients of Eq. (4) for the KDE method.

c_0	3.0532634	c_1	-0.0381186	c_2	0.0011377	c_3	-0.0000115
d_0	2.648	d_1	-7.205×10^{-4}	d_2	3.781×10^{-7}	d_3	-7.897×10^{-11}

To validate UPMASK and all the proposed modifications of the method, we adopted state-of-the-art galaxy mock catalogues from N-body simulations. The simulated catalogue contains photometric information and is described in the following section. The study performed is presented in Sec. 5.

4. Data

In this section we describe the data that was used in this project. First, we applied UPMASK and its modifications (Refs Sec. 2 and Sec. 3) to simulated data, to validate and inspect their performance. Afterwards, we applied the method to real data, looking to re-identify the PLANCKSZ2 clusters from Pan-STARRS data.

4.1. *Simulated data*

We used simulated galaxy survey catalogues that contain estimated photometry for DES[14] and Euclid[15] bands to test the application of UPMASK and its modifications in the context of galaxy clusters. We used the MICE catalogue version 2,[16–20] obtained from the CosmoHub portal.[21,22] This data was obtained by selecting the most massive cluster that was present inside a one square degree area ($30.5° \leq \delta \leq 31.5°$, $0.5° \leq \alpha \leq 1.5°$) and its surrounding objects. The resulting

subset contained a total of 8774 galaxies, 171 belonging to the most massive cluster that lies at $z = 1.09$. We selected the following catalogue entries: the DES[14] filters g, r, i, z, y, the Euclid[15] filters RIZ, Y, J and H, astrometry (α, δ), redshift, halo identifier and mass, and the galaxy identifier. In Sec. 5 we present the performance of UPMASK and its aforementioned modified versions using only this subset of galaxies and entries from the MICE catalogue.

4.2. *Planck 2nd Sunyaev-Zel'dovich source catalogue*

Many clusters that were detected via the Sunyaev-Zel'dovich (SZ) effect are still lacking optical counterparts as of the present time. After testing UPMASK and modifications with simulations, we search for the optical counterparts of Planck detected clusters. PLANCKSZ2 - Planck 2nd Sunyaev-Zel'dovich Source Catalogue[23] contains detections of galaxy clusters candidates, obtained during 29 months of observations. 1653 of the total objects in this catalogue, 1203 ($\approx 73\%$) are confirmed clusters. Here we use PLANCKSZ2 sources that were detected by the three detection methods in Ref. 23. We used the object identifier, the astrometric information, position errors and the integrated SZ intensity distortion, Y_{5R500}, from this catalogue. With this data we searched for galaxy clusters with UPMASK, in the PanSTARRS[24] catalogue as described below.

4.3. *Pan-STARRS*

As mentioned in the previous subsection, we decided to use the recent observations from the Pan-STARRS Panoramic Survey Telescope and Rapid Response System) survey[24] as the input to re-detect the PLANCKSZ2 galaxy cluster candidates. We used the second data release of the Pan-STARRS catalogue[25] published in 2019. Our method is as follows. Around each PLANCKSZ2 object, we extracted the following astrometric (α, δ) and photometric information: the mean PSF (Point Spread Function) magnitude of each filter (g - $486.6\,nm$, r - $621.5\,nm$, i - $754.5\,nm$, z - $867.9\,nm$ and y - $963.3\,nm$) and its standard deviation, and the mean Kron magnitude[26] of each filter and its standard deviation. We tailored the area queried for each PLANCKSZ2 catalogue entry using the sky coordinates and the position uncertainty – the latter used to determine the radius of the Pan-STARRS query. Thus we obtain multiple subsets, each containing the field that corresponds to the PLANCKSZ2 cluster coordinates. This catalogue contains mostly stars and galaxies, however, no flag is provided to discern between the two. To perform a simple separation, we adopted the difference between the PSF magnitude and the Kron magnitude of a filter (e.g. see Ref 27). To obtain a final galaxy catalogue we selected the objects that have a PSF-Kron magnitude > 0.05 in all of the selected filters. Finally, we applied UPMASK. We describe our findings in Sec. 6.

5. Validation of the method

This section is dedicated to a viability study of applying UPMASK to detect optical counterparts of galaxy clusters in photometric galaxy surveys. The validation process was carried out using state-of-the-art N-body simulations containing photometric and galaxy-dark-matter halo membership information that allows determining which galaxies are part of true cluster halos or are field galaxies. These simulations are described in Sec. 4. Below we present a detailed description of the tests and method yields for all versions of UPMASK under investigation, followed by a discussion of how the different modifications work and compare when varying method parameters.

5.1. *UPMASK and its modifications*

We applied the original UPMASK method as well as its modifications to the subset of simulated galaxies described in the previous section, using Euclid filters RIZ, Y, J, and H and also their respective colors, as well as the astrometry information. To compare performances between different method versions, we choose to set the number of runs equal to 10, *nruns*= 10, in all versions. This choice is a good compromise between having relatively short CPU running times and a fair membership frequentist probability, with a 10% resolution for all method versions. To perform a quantitative comparison, we compute the completeness and purity metrics as:

$$\text{Purity} = \frac{N_{UPMASK,cluster}}{N_{UPMASK,detec}}, \tag{5}$$

$$\text{Completeness} = \frac{N_{UPMASK,cluster}}{N_{cluster}}, \tag{6}$$

where $N_{UPMASK,cluster}$ is the number of known objects belonging to the cluster above a certain UPMASK probability threshold; $N_{cluster}$ is the number of known objects belonging to the simulated cluster and $N_{UPMASK,detec}$ is the number of objects above a certain UPMASK probability threshold. The results of this test are represented in Fig. 3, and their resulting CPU running times are shown in Table 3.

Table 3. CPU running times for the different UPMASK versions. The last column shows the relative CPU running time percentage of each version with respect to the original UPMASK method.

UPMASK Version	CPU running time (s)	%
KDE	274	100
KDE + Fitting Function	262	96
Voronoi + Anderson-Darling Test	190	69
Voronoi + Mean Comparison	148	54
Grid	303	110
Grid + Fitting Function	289	105

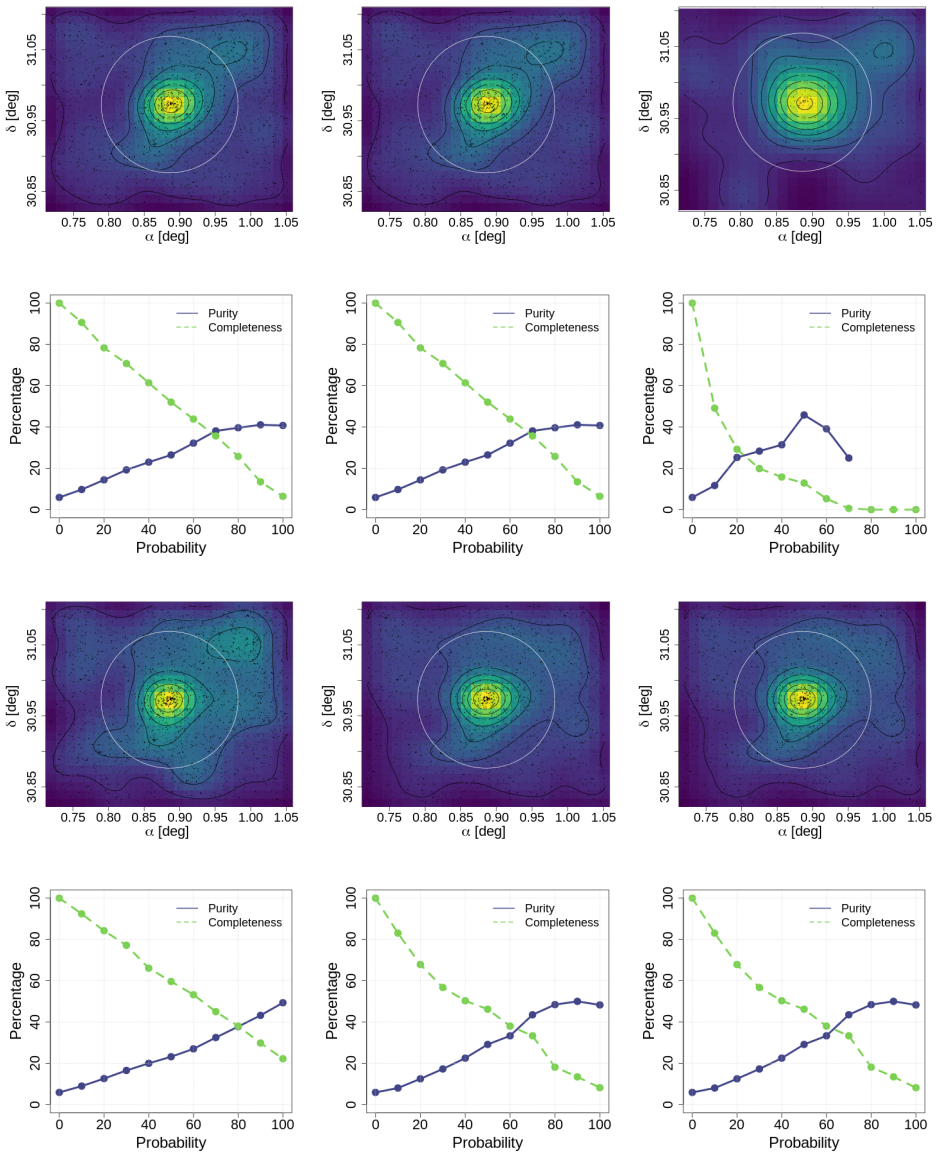


Fig. 3. The detection results for the original and all modified UPMASK versions. The results are displayed in groups of two rows, with each column corresponding to a version. The first and third rows show the density maps of UPMASK detected galaxies for the following versions: KDE, KDE + Fitting Function, Voronoi + Anderson-Darling Test, Voronoi + Mean Comparison, Grid, and Grid + Fitting Function. The color scale was set the same for all the field panels, with a color scheme where lower (higher) densities are in blue (yellow). The white circles indicate the cluster dark-matter halo radius. The second and fourth rows show the Completeness (green dashed) and Purity (blue solid) functions obtained from the unsupervised UPMASK classification, corresponding to the field above each case.

As expected when analyzing Fig. 3 and Table 3, the fitting function versions give the same results and improved in processing time, when comparing to their base UPMASK version. Voronoi+Anderson-Darling implementation is the one that improved the CPU time the most ($\sim 30\%$). However, the application of this version results in a less concentrated field (when comparing with KDE), as well as a more incomplete sample. This is due to the fact that the Anderson-Darling test is a stricter test since it compares the whole distribution of two samples, while the other implementations depend only on simple statistical calculations as the mean and standard deviation. When the Anderson-Darling test is replaced by a comparison of the mean of the Voronoi distributions. we were able to improve the completeness. On the other hand, the Grid implementation looks overall similar to the KDE version, although it results in a slightly purer sample when comparing to the latter. It is however unclear why the grid versions take more time to run than a KDE implementation, as the latter needs to perform more CPU-heavy tasks. We hypothesize that this difference is purely due to the implementation of the prototype code adopted here.

In the following subsections, we study the impact of the UPMASK parameters – number of Principal Components, number of objects per K-means cluster and threshold level T – have on the results, namely in the purity and completeness.

Table 4. Principal Components of the photometric information using simulated data (Sec. 4) and the Standard Deviation for each PC

Principal Component	Standard Deviation
PC_1	2.604481
PC_2	1.792770
PC_3	9.108574×10^{-1}
PC_4	3.631163×10^{-1}
PC_5	2.028331×10^{-1}
PC_6	4.036392×10^{-15}
PC_7	2.559386×10^{-15}
PC_8	2.356076×10^{-15}
PC_9	1.541463×10^{-15}
PC_{10}	1.157850×10^{-15}
PC_{11}	7.081957×10^{-16}

5.2. *Dependency in the number of principal components*

Here we show how the number of principal components affects the purity and completeness of the sample, as well as the UPMASK method run times. Principal components are ordered according to the variance of the original data projected on the PCA basis. Thus, by selecting the first few n principal components, one can grasp the most important information about the data. UPMASK calculates each PC from the provided photometry and colors. The standard deviation of each PC is shown in Table 4. This shows that the first five principal components are those that

contribute the most to the data representation and all the others can be neglected. We have studied the case of using 2, 3, 4, or 5 PCs. For each of these scenarios, we show purity and completeness curves represented in Fig. 4, and the CPU run times in Table 5.

Table 5. CPU run time as a function of the number of Principal Components used in UPMASK.

Nº of Principal Components	2	3	4	5
Time (s)	1073	1193	1224	1249

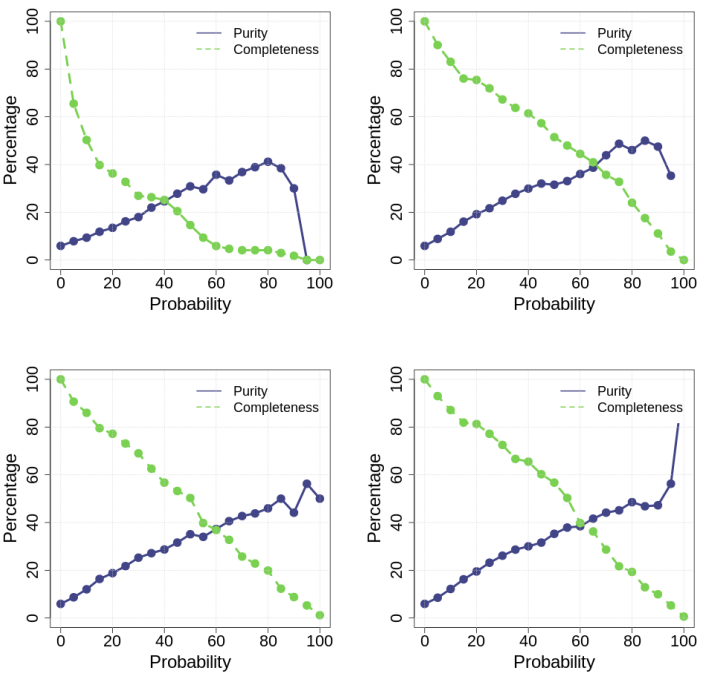


Fig. 4. Completeness (green dashed) and Purity (blue solid) of the unsupervised UPMASK classification (applied to the simulated data of Sec. 4) for different number of PCs being used in the method: 2 PCs (top left), 3 PCs (top right), 4 PCs (bottom left) and 5 PCs (bottom right).

Taking into account the completeness and purity in Fig. 4 as well as the computational time in Table 5, we have chosen to use the first 4 principal components for all of the applications below. The usage of 5 principal components includes more information, however, it is not significant when comparing with the study with 4 principal components. As such, it does not justify the additional run time.

5.3. *Dependency on the number of k-means objects "N"*

The next step is to test how the number of k-means objects "N" affect the results of the method, in all versions. We have performed this study for 20, 35, 50, 65, 80, 95, 110, 125, 140, 165, and 180 objects per K-means cluster. In Fig. 5 we show only some of the results, due to economy of space. For a complete grasp, we refer the reader to the appendix A of Ref 28.

Changing "N" impacts completeness and purity. Figure 5 shows that the completeness is increasing with the number of objects being clustered by k-means. On the other hand, the purity has the contrary behavior and decreases. This is because including more objects, these are not ignored by the method and the completeness will naturally increase. However, we are also contaminating more and as such, the purity is seen to be decreasing as expected.

5.4. *Dependency on the threshold level "T"*

Next, we study the impact of changing the threshold level T in UPMASK. This is a parameter that controls the detection limit comparison between the data and the uniform realization distributions. T can be adjusted in all method versions, except for those with thresholds provided by the fitting function procedures described earlier. We applied UPMASK and its modifications, testing different values for T, $T = \{0.5, 1, 1.5, 2, 2.5, 3\}$. The completeness and purity curves are shown in Fig. 6 for $T = \{0.5, 1.5, 3\}$. For complete results, we refer the reader to the Appendix A of Ref 28. From Fig. 6, we see a general behavior – a higher T leads to higher purity, but lower completeness. When increasing this value, the k-means cluster needs to achieve a higher threshold to be classified as a cluster. Thus, more groups will happen to be classified as following a uniform distribution and the completeness will decrease. On the other hand, k-means clusters that are without a doubt clustered, will be classified as such and kept in the UPMASK kernel.

Finally, the original implementation of the method seems to be a promising alternative for membership determination in galaxy clusters and cluster finding, as well as the newly implemented versions although there is room for accelerating UPMASK methodology, specially when making stronger assumptions about the cluster member distributions.[29]

6. Optical identification of Planck clusters with UPMASK

In this section, we apply UPMASK to search for optical counterparts of the PLANCKSZ2 cluster candidates in the Pan-STARRS catalogue. We are particularly interested in clusters not yet confirmed by other surveys. For each subset (which were obtained as described in Sec. 4) we applied UPMASK using the coordinates and the photometric data of each object. Taking into account the discussion in Sec. 5, here we used 4PCs. Some fields have a small number of objects after the star-galaxy separation, therefore we used 10 objects per k-means cluster for all the

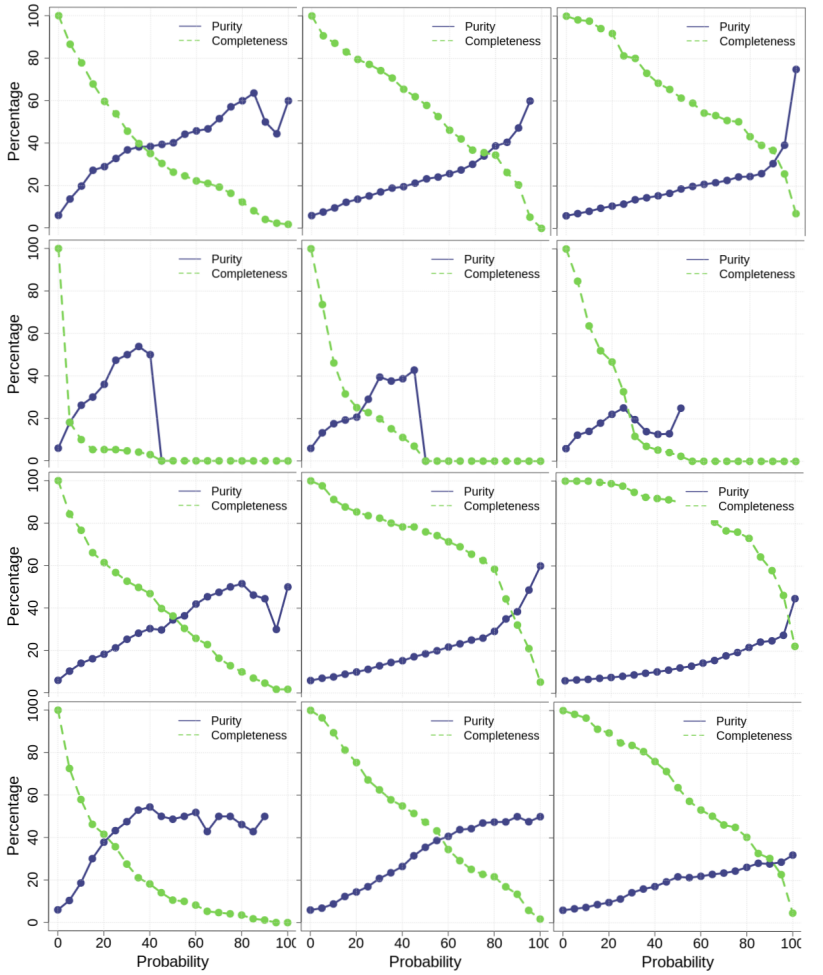


Fig. 5. Completeness (green dashed) and purity (blue solid) of UPMASK classification, applied to the simulated data of Sec. 4 for different N. We change this value along the lines ($N = \{20, 60, 140\}$), for the different methods. The lines are an application of the KDE version (original), Voronoi+Anderson-Darling, Voronoi+Mean and finally the Grid version.

fields. Since we are only interested in detect clusters with optical parameters, a resolution of 10% in the UPMASK score (frequentist probability) is a reasonable choice. The classification of a cluster is entirely up to the user of the method. Ours will depend on the number of objects each field has, above a certain probability. In Fig. 7 we show how this classification behaves when changing the cut in probability.

Thus, for the rest of this proceeding, we have chosen to classify a cluster as the field that has more than 10 members above 50% frequentist membership probability. In Fig. 8 we show the dependency of our findings with the redshift. The first plot shows the dependency of the measured SZ flux with the redshift. The figure

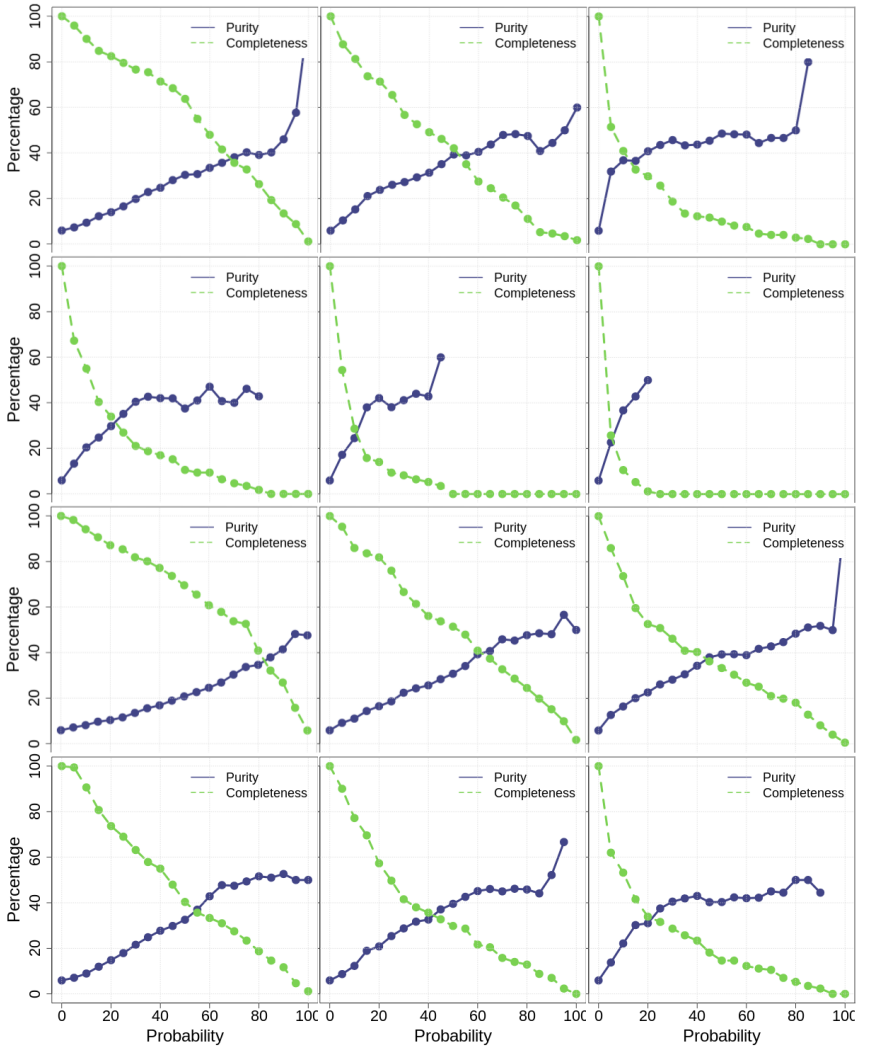


Fig. 6. Completeness (green dashed) and purity (blue solid) of UPMASK classification, applied to the simulated data of Sec. 4 for different T. We change this value along the lines ($T = \{0.5, 1.5, 3.0\}$), for the different methods. The lines are an application of the KDE version (original), Voronoi+Anderson-Darling, Voronoi+Mean and finally the Grid version.

shows three distributions: all PSZ2 sources (grey points), PSZ2 sources inside the Pan-STARRS sky-coverage (blue circles) and finally the UPMASK detections (blue dots). One can conclude that the detections do not seem to depend on the redshift (the dependency is only present due to the Planck selection function). In the second plot, we show the count dependency on the redshift for the same distributions described previously. Finally, the third plot shows the ratio of the detected objects over the total PSZ2 sources present in the Pan-STARRS sky coverage. This

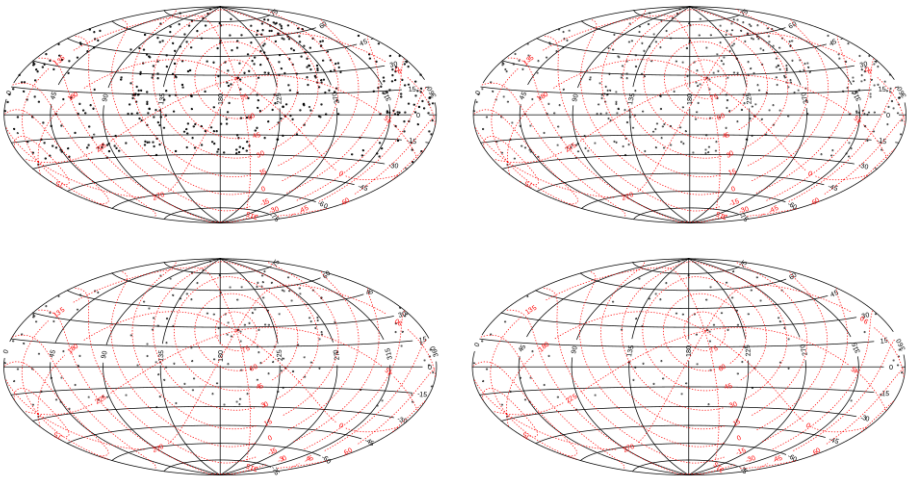


Fig. 7. Projection of the distribution of Planck Clusters in the sky detected by UPMASK for Clusters that have more than 10 objects with probability above of 50, 70, 90 and 100% (Left to right, top to bottom). Each point indicates one galaxy cluster from the PLANCKSZ2 catalogue. The red lines are the Galactic coordinate lines.

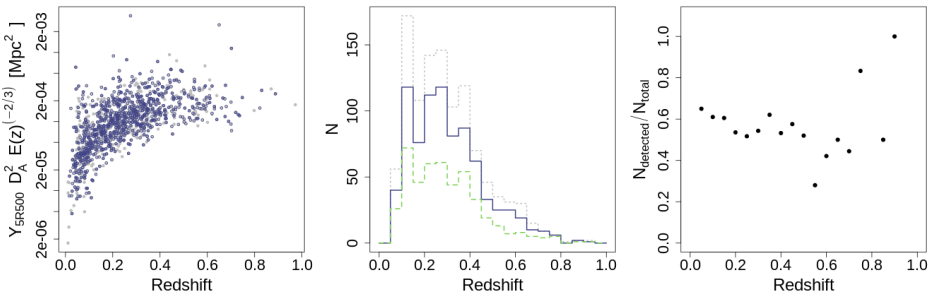


Fig. 8. Variation of PSZ2 sources properties with redshift. Left panel - Dependency of the measured Planck SZ flux with redshift. The grey points represent all the PSZ2 sources, the blue circles represent the PSZ2 sources that are included in the Pan-STARRS sky coverage and the blue points represent the sources that UPMASK detected (here, we established the detection criteria as the fields that contain more than 10 members above 50% membership probability. This detection criterion is the same in all panels of the figure). Center panel - Variation of PS2 source counts with redshift, for all sources (dotted grey), for sources included in the Pan-STARRS sky coverage (solid blue), and for the UPMASK detections (dashed green). Right panel - Fraction of the detected sources as a function redshift.

distribution shows a very weak dependency until redshift 0.5. Above this redshift, the ratio fluctuates significantly due to the small number of objects.

It is also particularly interesting to look for clusters in the PLANCKSZ2 survey catalogue that were not confirmed by an external counterpart at the moment of the

Table 6. Comparison table between the results of Refs 30 and 31 and UPMASK. The "True" and "False" fields are the unconfirmed PSZ2 sources that were found or not by the respective methods. The results are normalized to the total number of unconfirmed sources in the Pan-STARRS sky coverage.

	Streblyanska et al., Barahona et al. False	Streblyanska et al., Barahona et al. True
UPMASK False	36 (12%)	49 (16%)
UPMASK True	78 (26%)	138 (46%)

Table 7. Comparison table between the results of UPMASK and other authors (see Refs. 30–35). Here we used SIMBAD[36] to investigate which among the 216 UPMASK clusters had previous optical counterparts.

	Found (Other Authors)	Not Found (Other Authors)
UPMASK Found	67	149

release of the PLANCKSZ2 catalogue. Following the same classification, out of the 301 galaxy cluster candidates with no external validation, UPMASK was able to rediscover 72% (216) clusters with optical information. Furthermore, we have compared our results with other works in the literature (see Refs 30 and 31) in Table 6. The authors report that there are some spurious detections (ie, no obvious cluster observed in the photometric data) and therefore we have not included these detections in this table. The "False" and "True" fields refer to the fields that were not found in the list of PSZ2 sources without an optical counterpart, and those that are a clear detection, respectively. Overall we have a general accordance with the results from the aforementioned articles, while our method shows a more optimistic result (26% UPMASK clusters when compared with the "False" results). This optimistic behavior emerges naturally from the fact that UPMASK makes no assumptions about the cluster model, or requires direct redshift information from the galaxies in the data. However, more fortuitous detections might also be returned from the method in these conditions, and this should not be considered a drawback since any selected candidate still need to be spectroscopically confirmed. Here we were especially interested in the UPMASK candidates that were not previously confirmed by other authors, Refs. 30–35. In Table 7 we show that we found 67 cluster candidates in other recent efforts but without present optical confirmation. Fig. 9 shows an example of 4 of these detected clusters by UPMASK. Each panel is a KDE image with iso-contours obtained from the spatial distribution of galaxies that have membership probabilities $\geq 50\%$. Our forthcoming work will be dedicated to perform follow-up spectroscopy observations aiming to confirm or rule-out some of these new galaxy cluster candidates.

7. Conclusion

The study and characterization of galaxy clusters requires their detection and the identification of their member galaxies. In this proceeding, we modified and applied

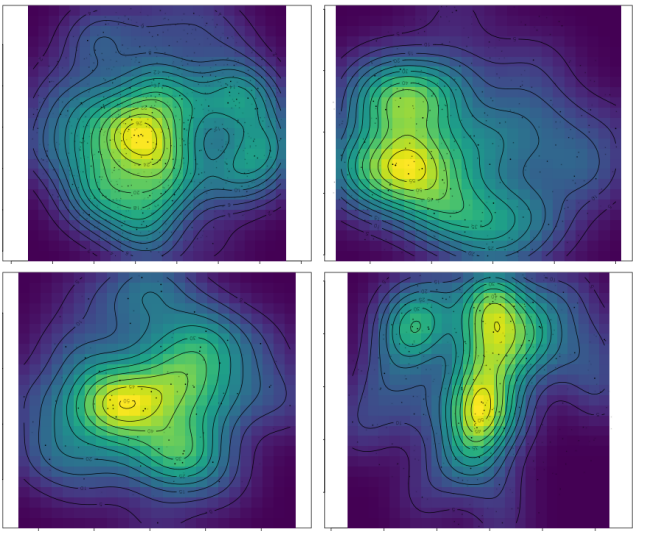


Fig. 9. Examples of UPMASK detections in Pan-STARRS fields around PSZ2 sources. The plots show a KDE image representation with iso-contours (color map: the brighter, the denser) of the most likely members (members with a probability above 50%)

UPMASK, a method originally designed to analyse stars in stellar clusters, to galaxy clusters. We show that the method is effective in this context, adopting only photometric and astrometric data as expected from upcoming surveys. Additionally, our study suggests that the performance of the method seems to be redshift independent.

In section Sec. 3 we describe the modifications proposed in this work: KDE+Fitting Function; Voronoi+Anderson-Darling test; Voronoi+Mean Comparison; Grid; Grid+Fitting Function. These modifications and the method itself were validated in Sec. 5, with simulated data from the MICE version 2 galaxy survey catalogue. In that same section we studied how the results behave with varying parameters, as the number of objects per K-means group, the number of Principal Components, and the threshold level. We conclude that the fastest modification is the Voronoi+Mean Comparison. On the other hand, the Grid implementation is the method that provided the highest improvement in purity. We also studied the optimal number of principal components to use. Overall, we were able to reduce dimensionality to 4 Principal Components. This number was chosen because additional components result in similar outcomes, with the penalty of increasing CPU time. Moreover, we have seen that using more objects per K-means cluster, results into higher completeness, while decreased purity. Also, when using a smaller threshold level, the completeness decreases, while the purity has an ascending behavior. In the end, we conclude that the UPMASK modifications and this parameter study shows that this is a very versatile tool for galaxy cluster cosmology, as the user can adjust the method and the parameters according to their goals: to use the method to quickly find galaxy cluster candidates or, to obtain membership probabilities.

In Sec. 6 we look for optical counterparts of Planck SZ clusters. We start with a list of candidates from the 2nd Sunyaev-Zel'dovich cluster catalogue (PSZ2) and the data from the Pan-STARRS catalogue collected from the Pan-STARRS survey. For each field obtained from Pan-STARRS, which corresponds to a certain PSZ2 candidate, we apply UPMASK. We show in this work (Fig. 8) that UPMASK can detect clusters independently of redshift. Additionally, we were also able to discover optical counterparts for about 72% of the 301 clusters without previous external validation. In the future, it can be interesting to apply the methods developed here to search for the new clusters directly from deep optical surveys, as the Rubin/Large Synoptic Survey Telescope and the Euclid space mission. For the cluster candidates detected in this work, we now plan to perform dedicated spectroscopic observation to determine their redshifts and to confirm or rule-out their physical existence.

Finally, thanks to the methods we developed here and to the adoption of the most modern optical surveys, we performed the detection of the optical counterparts for 67 galaxy cluster candidates previously identified by the ESA Planck space mission. This builds up the evidence for the reality of these candidates, which are at the backbones of the large-scale structure of the Universe.

Acknowledgments

AC acknowledges the support from the Portuguese Fundação para a Ciência e a Tecnologia (FCT) through the PhD Fellowship Ref: 2020.06644.BD. AKM acknowledges the support from FCT through the Portuguese Strategic Programme UID/FIS/00099/2019 for CENTRA and through the grant PTDC/FIS-AST/31546/2017. AdS acknowledges the support from FCT through the research contract No. IF/01135/2015 and Projects No. EXPL/FIS-AST/1368/2021, No. UID/FIS/04434/2020, No. CERN/FIS-PAR/0037/2019 and No. PTDC/FIS-OUT/29048/2017.

References

1. M. Roos, Astrophysical and Cosmological Probes of Dark Matter, *Journal of Modern Physics* **3**, 1152 (January 2012).
2. S. W. Allen, A. E. Evrard and A. B. Mantz, Cosmological Parameters from Observations of Galaxy Clusters, *Annual Review of Astronomy and Astrophysics* **49**, 409 (September 2011).
3. M. D. Gladders and H. K. C. Yee, A New Method For Galaxy Cluster Detection. I. The Algorithm, *The Astronomical Journal* **120**, 2148 (October 2000).
4. A. Krone-Martins and A. Moitinho, UPMASK: unsupervised photometric membership assignment in stellar clusters, *Astronomy and Astrophysics* **561**, p. A57 (January 2014).
5. T. Cantat-Gaudin, C. Jordi, A. Vallenari *et al.*, A Gaia DR2 view of the open cluster population in the Milky Way, *Astronomy and Astrophysics* **618**, p. A93 (October 2018).

6. T. Cantat-Gaudin, A. Krone-Martins, N. Sedaghat *et al.*, Gaia DR2 unravels incompleteness of nearby cluster population: new open clusters in the direction of Perseus, *Astronomy and Astrophysics* **624**, p. A126 (April 2019).
7. H. Hotelling, Analysis of a complex of statistical variables into principal components., *Journal of Educational Psychology* **24**, 417 (1933).
8. E. Parzen, On estimation of a probability density function and mode, *Annals of Mathematical Statistics* **33**, 1065 (1962).
9. M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**, 832 (1956).
10. E. W. Forgy, Cluster analysis of multivariate data: efficiency versus interpretability of classifications, *Biometrics* **21**, 768 (1965).
11. G. Voronoi, Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites., *Journal für die reine und angewandte Mathematik (Crelles Journal)* **1908**, 97 (1908).
12. T. W. Anderson and D. Darling, Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Annals of Mathematical Statistics* **23**, 193 (1952).
13. B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London, 1986).
14. K. Honscheid and D. L. DePoy, The Dark Energy Camera (DECam), *arXiv e-prints*, p. arXiv:0810.3600 (October 2008).
15. R. Laureijs, J. Amiaux, S. Arduini *et al.*, Euclid Definition Study Report, *arXiv e-prints*, p. arXiv:1110.3193 (October 2011).
16. P. Fosalba, M. Crocce, E. Gaztañaga and F. J. Castander, The MICE grand challenge lightcone simulation – I. Dark matter clustering, *Monthly Notices of the Royal Astronomical Society* **448**, 2987 (03 2015).
17. M. Crocce, F. J. Castander, E. Gaztañaga *et al.*, The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues, *Monthly Notices of the RAS* **453**, 1513 (October 2015).
18. P. Fosalba, E. Gaztañaga, F. J. Castander and M. Crocce, The MICE Grand Challenge light-cone simulation – III. Galaxy lensing mocks from all-sky lensing maps, *Monthly Notices of the Royal Astronomical Society* **447**, 1319 (12 2014).
19. J. Carretero, F. J. Castander, E. Gaztañaga *et al.*, An algorithm to build mock galaxy catalogues using MICE simulations, *Monthly Notices of the RAS* **447**, 646 (February 2015).
20. K. Hoffmann, J. Bel, E. Gaztañaga *et al.*, Measuring the growth of matter fluctuations with third-order galaxy correlations, *Monthly Notices of the RAS* **447**, 1724 (February 2015).
21. P. Tallada, J. Carretero, J. Casals *et al.*, Cosmohub: Interactive exploration and distribution of astronomical data on hadoop, *Astronomy and Computing* **32**, p. 100391 (2020).
22. J. Carretero, P. Tallada, J. Casals *et al.*, CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform, *PoS* **EPS-HEP2017**, p. 488 (2017).
23. Planck Collaboration, P. A. R. Ade, N. Aghanim *et al.*, Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources, *Astronomy and Astrophysics* **594**, p. A27 (Sep 2016).
24. K. C. Chambers, E. A. Magnier, N. Metcalfe *et al.*, The Pan-STARRS1 Surveys, *arXiv e-prints* , p. arXiv:1612.05560 (December 2016).

25. H. A. Flewelling, E. A. Magnier, K. C. Chambers *et al.*, The Pan-STARRS1 Database and Data Products, *arXiv e-prints* , p. arXiv:1612.05243 (Dec 2016).
26. R. G. Kron, Photometry of a complete sample of faint galaxies., *Astrophysical Journal, Supplement* **43**, 305 (Jun 1980).
27. D. J. Farrow, S. Cole, N. Metcalfe *et al.*, Pan-STARRS1: Galaxy clustering in the Small Area Survey 2, *Monthly Notices of the RAS* **437**, 748 (Jan 2014).
28. A. S. C. Carvalho, Exploration of unsupervised machine learning methods to study galaxy clustering (2019).
29. M. S. Pera, G. I. Perren, A. Moitinho *et al.*, pyUPMASK: an improved unsupervised clustering algorithm, *Astronomy and Astrophysics* **650**, p. A109 (June 2021).
30. A. Streblyanska, A. Aguado-Barahona, A. Ferragamo *et al.*, Optical validation and characterization of Planck PSZ2 sources at the Canary Islands observatories. I. First year of LP15 observations, *Astronomy and Astrophysics* **628**, p. A13 (Aug 2019).
31. A. Aguado-Barahona, R. Barrena, A. Streblyanska *et al.*, Optical validation and characterization of Planck PSZ2 sources at the Canary Islands observatories. II. Second year of LP15 observations, *arXiv e-prints* , p. arXiv:1909.06235 (Sep 2019).
32. R. Barrena, A. Streblyanska, A. Ferragamo *et al.*, Optical validation and characterization of Planck PSZ1 sources at the Canary Islands observatories. I. First year of ITP13 observations, *Astronomy and Astrophysics* **616**, p. A42 (August 2018).
33. S. Boada, J. P. Hughes, F. Menanteau *et al.*, High Confidence Optical Confirmations among the High Signal-to-noise Planck Cluster Candidates, *The Astrophysical journal* **871**, p. 188 (Feb 2019).
34. A. Aguado-Barahona, R. Barrena, A. Streblyanska *et al.*, Optical validation and characterization of Planck PSZ2 sources at the Canary Islands observatories. II. Second year of LP15 observations, *Astronomy and Astrophysics* **631**, p. A148 (November 2019).
35. H. Zohren, T. Schrabback, R. F. J. van der Burg *et al.*, Optical follow-up study of 32 high-redshift galaxy cluster candidates from Planck with the William Herschel Telescope, *Monthly Notices of the RAS* **488**, 2523 (Sep 2019).
36. M. Wenger, F. Ochsenbein, D. Egret *et al.*, The SIMBAD astronomical database. The CDS reference database for astronomical objects, *Astronomy and Astrophysics, Supplement* **143**, 9 (April 2000).

Time-dependent lepto-hadronic modeling of the emission processes in blazar jets

S. Gasparyan[1], D. Bégué[2,3] and N. Sahakyan[1,4,5]

[1] *ICRANet-Armenia, Marshall Baghramian Avenue 24a, Yerevan 0019, Armenia*
[2] *Department of Physics, Bar Ilan University, Ramat-Gan 52900, Israel* *
[3] *Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching, Germany*
[4] *ICRANet, P.zza della Repubblica 10, 65122 Pescara, Italy*
[5] *ICRA, Dipartimento di Fisica, Sapienza Universita' di Roma, P.le Aldo Moro 5, 00185 Rome, Italy*
* *E-mail: sargisgyan@gmail.com*

The recent associations of neutrino events with blazars (e.g. TXS 0506+056, 3HSP J095507.9+355101) provided a unique opportunity to study the possible physical connection between the multiwavelength electromagnetic and neutrino emissions. We present *SOPRANO*, a new conservative implicit kinetic code which follows the time evolution of the isotropic distribution functions of protons, neutrons and the secondaries produced in photo-pion and photo-pair interactions, alongside with the evolution of photon and electron/positron distribution functions. In the current work, we apply *SOPRANO* to model the broadband spectrum of TXS 0506+056, 3HSP J095507.9+355101 and 3C 279 blazars. It was possible to constrain main physical parameters within both a pure hadronic and lepto-hadronic scenarios.

Keywords: Radiation mechanisms; non-thermal–quasars; individual; TXS 0506+056, 3HSP J095507.9+355101 and 3C 279 – galaxies; jets – gamma-rays; galaxies

1. Introduction

In recent years, the discovery of very high energy neutrinos of astrophysical origin by the IceCube experiment[1–3] has opened a new window on very high energy sources such as gamma-ray bursts (hereafter GRBs), active galactic nuclei (AGNs) and tidal disruption events (TDEs). In 2017, the very high energy neutrino IceCube 170922A[4] and its further association (confidence level of 3.5σ) with the blazar TXS 0506+056[5] made clear that high energy protons, neutron and even possibly nucleons have an important role to play in the dynamics and the radiation of relativistic jets.[6–8]

In the forthcoming years, it is expected that many more crucial data will become available for high energy sources at all wavelengths, but more importantly at very high energy. For instance, the MAGIC collaboration recently reported the first observations of a GRB at energies above TeV with its observation of 190114C,[9] and several more are expected to be observed with CTA.[10,11] In addition, upgrades of existing neutrino facilities, with e.g. IceCube-Gen2,[12] as well as new experiments such as P-one,[13] KM3Net[14] and the Baikal-GVD,[15] will allow for increasingly more

*Current address.

sensitive searches of the origin of very high energy neutrinos.[16] Those experiments should also help to establish more firm coincidental observations of neutrino bursts with identified electromagnetic counterparts, thus allowing for models to be tested, constrained, challenged and eventually ruled out.

In order to understand observations and constrain models of emission able to explain very high energy neutrinos, it is necessary to perform self-consistent simulations of the time evolution of hadrons distribution functions, together with following the evolution of the photon and lepton distribution functions. This is a challenging task since 1- the number of distributions is much larger (from two to fourteen although some are trivial), 2- all equations describing the time evolution of particle distribution functions are coupled in a non-trivial (and non-linear) way by many complex processes that 3- have different time scales, requiring an implicit time discretization. The high number of distribution functions comes from the fact that we also wish to follow the cooling and emission of charged secondaries (pion and muons), preventing us from using semi-analytical expressions for the production rate of neutrinos, see *e.g.*[17]

Over the years, several groups have developed numerical models to estimate leptons, hadron and photon distribution functions in order to make predictions to explain observations, either under the steady state approximation[18–21] or for the fully time dependent equations either for leptons only[22] or adding protons.[23–28] In this paper, we present our approach to the problem: a modular code solving the time dependent isotropic kinetic equations which preserves the total energy of the system as well as the number of particles where needed. The code structure is modular such that processes can be added easily (or removed). It is also implicit so numerical stability is achieved at all time.

The paper is organised as follows. In Section 2, we detail the processes included in our numerical code and the associated kinetic equation. In Section 3, we provide elements of the numerical discretizations in energy and time. In Section 4 we present examples of the theoretical modelings obtained by *SOPRANO*. In particular, we model the spectra of TSX 0506+056, 3HSP J095507.9+355101 and of 3C 279. The conclusion follows.

2. *SOPRANO*: Simulator of Processes in Relativistic AstroNomical Objects

In its current version, the code evolves in time the isotropic kinetic equations for photons, electrons and positrons (considered as one species, see below), protons, neutrons, charged and neutral pions, muons, neutrino and anti-neutrino of all flavors. We summarize here all terms appearing in the kinetic equation for all particle species. For the cross-sections used in SOPRANO see.[29] In the photon distribution function, we assign $n_{\rm ph}$ to be the number of photons per unit volume per hertz. We further define $N_{\rm i}$ to be the number of particles of species i per unit volume per unit Lorentz factor of particle i. Here i can be all leptons and all hadrons. Then, we

define N_{ν_i} the number of neutrinos of flavour i per units volume per GeV. Finally, for the sink and source terms we assign Q and S, respectively. The contribution of inverse Compton scattering is noted R_{IC} for the photons and is a cooling term for the leptons.

- *Photons :* they are created by synchrotron radiation from all charged particles, and by π_0 decay. They are absorbed by pair production and redistributed in energy by inverse Compton scattering. We neglect the absorption of photons in the photo-pion and photo-pair processes. The resulting kinetic equation takes the form

$$\frac{\partial n_{\rm ph}}{\partial t} = -S_{\gamma\gamma\to e^+e^-} + Q_{\pi_0} + R_{\rm IC} + \sum_i Q^i_{\rm synch} \tag{1}$$

- *Leptons :* electron and positrons are considered a single species. They are created by muon decay μ, by photo-pair production and photon pair production. The also undergo synchrotron cooling such that the final kinetic equation reads

$$\frac{\partial N_{\rm e^\pm}}{\partial t} = Q_\mu + Q_{p\gamma\to e^+e^-} + Q_{\gamma\gamma\to e^+e^-} C_{\rm IC} + C_{\rm synch} \tag{2}$$

- *Protons :* protons undergo cooling via synchrotron emission, Bethe-Heitler photo-pair production and via photo-pair production. They are produced by photo-pair interaction with neutron, and disappear for a fraction of photo-pair interaction. The kinetic equation takes the form

$$\frac{\partial N_p}{\partial t} = C_{p\gamma\to p\pi} + C_{p\gamma\to e^+e^-} + C_{\rm synch} - S_{\gamma p\to n\pi} + Q_{\gamma n\to p\pi} \tag{3}$$

- *Neutron :* Neutron are produced in photo-pion production and disappear in photo-pion interactions. The kinetic equation takes the form

$$\frac{\partial N_n}{\partial t} = -S_{n\gamma\to p\pi} + Q_{p\gamma\to n\pi} + C_{n\gamma\to n\pi} \tag{4}$$

- *Charged pions :* π_+ and π_- are produced by photo-pair production. Then, they cool via synchrotron emission and decay. The kinetic equation for both species takes the form

$$\frac{\partial N_{\pi_\pm}}{\partial t} = Q_{p\gamma\to\pi} + Q_{n\gamma\to\pi} - S_\pi + C_{\rm synch} \tag{5}$$

  We note that we solve the kinetic equation independently for both species, since the branching ratio in photo-pion production are different for negative and positive pions.
- *Neutral pions :* the π_0 kinetic equation is similar to that of charged pions but without synchrotron cooling.

- *Muons :* Muons are produced by the decay of charged pions. They are also cooled by the synchrotron process and they decay. Therefore, the kinetic equation is

$$\frac{\partial N_\mu}{\partial t} = Q_{\pi_\pm} - S_\mu + C_{\text{synch}} \tag{6}$$

- *Neutrinos and anti-neutrinos:* neutrinos and anti-neutrinos are produced in the decay of pions and of muons. We consider all 2 flavours ζ

$$\frac{\partial N_{\nu,\zeta}}{\partial t} = Q_{\pi_\pm} + Q_\mu \tag{7}$$

3. Numerical Discretization

The SOPRANO follows the evolution of distribution functions as a function of time under the following simplificative assumptions: i) space is infinite and homogeneous, ii) distribution functions are isotropic. For disretization, SOPRANO uses a flavour of finite volume approximation based on the discontinuous Galerkin method. This method fits nicely for solving integro-differential equations.

3.1. *Energy discretization*

In order to numerically solve the kinetic equations presented in Section 2, a numerical grid for the energy of all particles is introduced. The grid element are equally space in logarithmic of the energy.[a] The grid characteristics for each types of particle are given by Table 1. We use the approach of the discontinuous Galerkin method. On each energy cell I, we approximate the distribution function by a polynomial. In the current version we restraint to first order and consider the Legendre polynomial as basis function. Therefore, on each energy cell I, the distribution function is approximated by

$$N_{\text{i}}^I(x) = N_{\text{i},0}^I L_0^I(x) \tag{8}$$

where the first order Legendre polynomial on the energy cell I is

$$L_0^I = \frac{1}{\sqrt{x_{I+\frac{1}{2}} - x_{I-\frac{1}{2}}}} \equiv \frac{1}{\sqrt{||I||}} \tag{9}$$

Here $x_{I\pm(1/2)}$ are the boundaries of energy cell I and where we introduced the additional notation $||I|| = (x_{I+1/2} - x_{I-1/2})$. An important properly of the Legendre polynomial is that

$$\int_I L_0^I dx = 1. \tag{10}$$

[a]Note that our numerical method do not require to have a uniform grid. Because it is based on finite volume, we can refine the grid in one or several energy bands of interest. In this way, we can provide more detailed results in those specific bands, while the rest of the domain is coarse for faster numerical estimation.

Table 1. Characteristics of the numerical grids used by SOPRANO for the numerical models of this work. The cells are equally space in logarithmic scale.

Particle	Number of cells	Minimum energy	Maximum energy
Photons	150	$\nu = 10^{-2}$Hz	$\nu = 10^{30}$ Hz
Leptons	130	$\gamma_{e\pm} = 1.2$	$\gamma_{e\pm} = 5 \times 10^{13}$
Hadrons	100	$\gamma_h = 1.2$	$\gamma_h = 10^{11}$
Neutrinos	100	$E_\nu = 10^{-3}$ GeV	$E_\nu = 10^{11}$ Gev

In the following we will use interchangeably $N_i^I \equiv N_{i,0}^I$. We seek the weak formulation of all kinetic equations presented below on all energy interval I. For this, we multiply both sides of the equation by L_0 and integrate over I. After simplification, we obtain a system of differential equations for all $N_{i,0}^I$. This specific discretization and the structure of the kinetic equation allows us to retrieve a numerical method which conserves energy and the number of particles when they are conserved.

3.2. *Temporal discretization*

Before discussing the temporal discretization, we note that the equation are non-linear in the distribution function for Compton scattering, pair production, photo-pion and photo-pair processes. We have decided to linearize the kinetic equation of photo-pion and photo-pair processes by assuming that the target photon-field is equal to the one at the previous time step, effectively making those process linear in the distribution function. For all leptonic processes,[b] we preserve the non-linearity of the kinetic equations and solve at each time step a non-linear system via Newton-Raphson iteration. The temporal evolution of the distribution function is performed with the implicit Euler.

The computation of one time step is done in the following form

(1) solve the linear kinetic equation for the protons since for photo-pion and photo-pair processes the photon distribution function is assumed to be known and given by its value at time t. The pairs created in the photo-pair processes are saved to be use as a source term in the leptonic computation.
(2) compute the decay and cooling (when required) of pions and muons. The pairs and photons created in the muon and pion decay are used as source terms in the leptonic computation.
(3) perform the fully implicit leptonic computation with the source term computed in the two previous steps.

4. Modeling of Blazar SEDs

The code *SOPRANO*, described in Section 2, is used to model the multiwavelength SEDs of TXS 0506+056, 3HSP J095507.9+355101 and 3C 279. Two of these sources,

[b]This is true for all process but synchrotron which is linear in the distribution function.

TXS 0506+056 and 3HSP J095507.9+355101, coincide in space and time with the IceCube 170922A[4] and IceCube 200107A[30] events, respectively. The other source, 3C 279, shows a prominent flare in the γ-ray band. In all modelings, the emission region assumed to be a sphere with a radius of R' which is moving with a bulk Lorentz factor of $\Gamma \sim \delta$, carries a magnetic field with an intensity of B' and a population of relativistic particles. The magnetic jet luminosity is defined as

$$L_{\rm B} = \pi c R'^2 \delta^2 \frac{B'^2}{8\pi}, \tag{11}$$

while for protons and electrons as follows:

$$L_{\rm p} = \pi R'^2 \delta^2 m_{\rm p} c^3 \int \gamma_{\rm p} Q'_{\rm p}(\gamma_{\rm p}) \gamma_{\rm p}, \tag{12}$$

$$L_{\rm e} = \pi R'^2 \delta^2 m_{\rm e} c^3 \int \gamma_{\rm e}\, Q'_{\rm e}(\gamma_{\rm e})\, d\gamma_{\rm e}, \tag{13}$$

where c is the speed of light, $m_{\rm p}$ and $m_{\rm e}$ are the proton and electron masses, while $Q'_{\rm p}$ and $Q'_{\rm e}$ refer to proton and electron injection spectra, with the following forms:

$$Q'_{\rm p}(\gamma_{\rm p}) = Q'_{0,p} \gamma_{\rm p}^{-\alpha_{\rm p}} \qquad \gamma_{\rm p} < \gamma_{\rm p,max}. \tag{14}$$

$$Q'_{\rm e}(\gamma_{\rm e}) = \begin{cases} Q'_0 \gamma_{\rm e}^{-\alpha_{\rm e}} \exp\left(-\dfrac{\gamma_{\rm e}}{\gamma_{\rm e,cut}}\right) & \gamma_{\rm e,min} \le \gamma_{\rm e} \le \gamma_{\rm e,max}, \\ 0 & \text{otherwise}, \end{cases} \tag{15}$$

where $\gamma_{\rm e,min}$ is the minimum injection Lorentz factor. All primed quantities refer to the comoving frame. The distribution functions of protons and electrons evolve via cooling and via interaction with photons, producing different signatures in the broadband spectrum. Our aim is to identify those signatures and use them to constrain the emission mechanism within the framework of different scenarios. We also assume that the injection power-law indexes are such that $\alpha_{\rm e} = \alpha_{\rm p}$. Once injected in the emitting region, particles interact with the magnetic field and with the photons, producing secondary particles, which themselves interact, radiate and decay, shaping the broadband SED. The low energy component is interpreted as the synchrotron emission of the primary electrons while the HE component is modelled either within a pure hadronic model(hereafter HM) or a hybrid lepto-hadronic (hereafter hybrid) model e.g.[31,32] For the HM, the high energy peak on SED is formed due to the proton synchrotron emission, however in the hybrid model, the high energy peak is explained by leptonic processes and proton synchrotron emission is required to be subdominant. The proton content is obtained by maximizing the neutrino flux at PeV energies, which is constrained by the radiation from the secondaries produced by the Bethe-Heithler and photo-pion processes. Indeed, it has long been speculated that efficient neutrino production is associated with efficient Bethe-Heithler process, creating a population of HE pairs, which can over-shine the tight constraints in the X-ray band e.g., order of minutes.[33] The system of kinetic equations is evolved for

one dynamical time scale $t'_{\rm dyn} \sim R'/c$ considering the magnetic field to be constant, and taking into account all relevant processes for particles interactions. Having estimated the model parameters of each blazar SEDs, the corresponding neutrino flux can be derived. The flux of muon neutrino, $F_{\nu_\mu}(E_{\nu_\mu})$, in all considered scenarios is shown by the light blue line in Figures 1-3. When available, the neutrino flux is compared with the limit imposed by the IceCube detector. This flux can be transformed to the expected observed number of neutrinos in the IceCube detector using its averaged effective area $A_{\rm eff}(\mathrm{E}_\mu)$, which is mostly a function of the incident neutrino energy. For 3HSP J095507.9+355101 and 3C 279, the average area from[34] was considered, while for TXS 0506+056 we used the area released after the observation of IceCube-170922A.[c] The effective area increases with energy and reaches its maximal value for energies above several hundreds of PeV. The expected number of muon neutrinos and anti-neutrinos is computed through

$$N_{\nu_\mu+\bar{\nu}_\mu} = t_{\rm exp} \int_{E_{\rm min,\nu_\mu}}^{E_{\rm max,\nu_\mu}} F_{\nu_\mu}(E_{\nu_\mu})\, A_{\rm eff}(\mathrm{E}_{\nu_\mu})\, dE_{\nu_\mu} \tag{16}$$

where the minimum and maximum energy of the neutrinos are $E_{\rm min,\nu_\mu} = 100$ GeV and $E_{\rm max,\nu_\mu} = 10^9$ GeV, respectively, chosen to correspond the limits for the effective area. The expected number of neutrino events depends on the duration of the source activity, $t_{\rm exp}$, over which the neutrinos are emitted. The neutrino oscillation, within the quasi-two neutrino oscillation assumption, is taken into account by

$$N^{obs}_{\nu_\mu} = 0.575 N_{\nu_\mu} + 0.425 N_{\nu_e}, \tag{17}$$

where $N^{obs}_{\nu_\mu}$ is the observable distribution of muon neutrinos, while, N_{ν_μ} and N_{ν_e} are the emitted muon and electron neutrino distributions.[35]

5. Results

The results of the SEDs modeling are shown in Fig. 1-3 with the Table 2 and 3.

TXS 0506+056: Two different periods are chosen for modelling: the single neutrino detection period,[4] and the historical neutrino flaring period.[5] The modeling results are shown in Figure 1 and Table 2. Panels a) and b) of Figure 1 represent the modeling of SED of the single neutrino event period, while panel c) and d) show flare for historical neutrino flare SED modelings. The multiwavelength data from[4] are modeled within a HM scenario in panel a). The corresponding model parameters are given in Table 2. The sum of all components, represented by the blue line in the top left panel of Figure 1, satisfactorily explains the observed data. The model overpredicts the radio data, but by taking into account the synchrotron self-absorption effect,[29] the model is in agreement with the data. The modeling parameters given in the first column of Table 1 are in the range of similar estimations for blazars in general and for TXS 0506+056 in particular. A Doppler factor $\delta = 20$ and a radius

[c] https://icecube.wisc.edu/science/data-releases/

$R' = 2.5 \times 10^{15}$ cm were used in our modeling. This is in agreement with the limits on the variability time of 10^5s presented in.[36] We note that when $\delta = 10$ or 15, the data can also be well reproduced by the model. The initial injection power-law index of the emitting electrons is $\alpha_e = 2.1$, a value that can be formed by shock accelerations, *e.g.*[37] Due to the high magnetic field, $B = 80$ G, electrons are in the fast cooling regime and their distribution function is a power-law with index $\alpha_e + 1$. The initial electron distribution extends up to $\gamma_{\rm cut} = 2.4 \times 10^3$ (~ 1 GeV) which is representative of the acceleration and cooling time scales. Instead, protons cool less efficiently and they could be accelerated up to much higher energies, *i.e.* $\gamma_{\rm max} = 10^9$ (9.4×10^{17} eV), see.[29]

As an alternative model to HM or leptonic models, the hybrid model is also applied to the SED of TXS 0506+056, shown in panel b) of Figure 1. The model parameters are given in the second column of Table 2. The blue dashed lines represent the time evolution of the spectrum in selected numerical steps, which builds and forms the overall SED, represented by the solid blue line after one dynamical time scale. The synchrotron component peaking between 1-10 eV is up-scattered by the relativistic electrons to produce the HE and VHE component. In contrast to HM scenario, the magnetic field in the emitting region is $B = 0.57$G significantly lower, therefore, electrons with Lorentz factor $\gamma_{\rm e,min} = 10^3$ are not substantially cooled within one dynamical time scale. For this hybrid model, protons do not directly contribute to the observed SED, instead the contribution of secondary electron-positrons dominates in the X-ray band, which puts constraints on the proton luminosity and as a consequence the neutrino luminosity. For example, if one increases by 1.5 times the proton luminosity, the model would overshoot the X-ray data, as shown by the dotted-dashed blue line in panel b) of Figure 1.

Further the results of the SED modeling obtained during the historical neutrino flare of TXS 0506+056 is presented. Unfortunately, when 13 ± 5 neutrinos were observed between October 2014 and March 2015,[5] the multiwavelength coverage is scarce. Yet, the flux upper limit of $F < 9.12 \times 10^{-12}$erg cm^{-2} s^{-1} derived from Swift BAT observations[38] introduces substantial difficulties for a one-zone modeling. Indeed, the predicted number of neutrino events cannot be matched to the IceCube observations[38] and[39] have shown that only few neutrino events could be detected under different optimistic considerations for the emitting region and for the target photon field (internal or external to the jet). Matching together the observed multiwavelength data and the neutrino data seems to require two zone models with more free parameters.[38,39]

To accommodate the X-ray limit and try to account for the neutrino flux during this flare, two different assumptions on the proton distribution function are made. On the one hand, radiation from the secondaries can be constrained to be dominant in the MeV band, in which there are no observational constraint. On the other hand, radiation from the secondaries could be dominant in the GeV band and produce the second HE hump. The SEDs of these two models are respectively shown in panels c) and d) of Figure 1, with data from.[39] Those two models lead to two very

different sets of parameters for the emitting region, see column 3 and 4 of Table 2. The first model requires a large radius $R' = 10^{17}$cm and a slowly moving jet with Doppler factor $\delta = 10$, while the second model necessitates those parameters to be $R' = 10^{15}$cm and $\delta = 15$. The required magnetic field also significantly differs between these two models with $B = 35$G for the first model, to be compared to $B = 0.65$G for the second one. The first model tends to reproduce the neutrino flux, albeit produces the peak at lower energies. The second model puts the neutrino peak at larger energy, but is not able to reproduce the observed neutrino number. In both interpretations, it is clear that the upper limit in the X-ray band imposes strong constraints on the photon spectrum, which in turn limits the proton content in the jet. Considering larger proton luminosity would lead to over-estimate both the observed γ-ray flux and the X-ray upper limit.

Table 2. Parameter sets used for modeling the SEDs of TXS 0506+056, observed in 2017 and during the neutrino flare in 2014-2015. The electon, proton and magnetic luminosities are also given.

	TXS 0506+056: 2017		TXS 0506+056: 2014-2015	
	Hadronic	Lepto-hadronic	Hadronic	Lepto-hadronic
δ	20	20	15	10
$R/10^{15}$ cm	2.5	10	1	100
$B[G]$	80	0.57	35	0.65
$\gamma_{\rm e,min}$	100	1000	2×10^2	9×10^3
$\gamma_{\rm e,cut}$	2.4×10^3	4.5×10^4	10^4	$=\gamma_{\rm e,max}$
$\gamma_{\rm e,max}$	3×10^4	6×10^4	8×10^4	8×10^4
$\alpha_{\rm e}$	2.1	2.0	2.0	2.0
$\alpha_{\rm p} = \alpha_{\rm e}$	2.1	2.0	2.0	2.0
$\gamma_{\rm p,min}$	1	1	1	1
$\gamma_{\rm p,max}$	10^9	10^6	2×10^8	1.2×10^5
$L_{\rm e}$ (erg s^{-1})	2.2×10^{44}	9.3×10^{44}	2.8×10^{44}	5.3×10^{44}
$L_{\rm B}$ (erg s^{-1})	6.0×10^{46}	4.9×10^{43}	10^{45}	1.6×10^{45}
$L_{\rm p}$ (erg s^{-1})	2.1×10^{47}	2.6×10^{50}	3.4×10^{47}	4.9×10^{52}

From the jet energetic perspective, the total jet luminosity in the case of TXS 0506+056, for the proton synchrotron model, shown on panel a) of Figure 1, is $L_{\rm tot} = L_{\rm p} + L_{\rm e} + L_{\rm B} = 2.7 \times 10^{47}\,{\rm erg\,s^{-1}}$, which exceeds by one order of magnitude the Eddington luminosity of $\simeq 4 \times 10^{46}{\rm erg\,s^{-1}}$ for a black hole mass of $3 \times 10^8\,M_\odot$, as estimated for TXS 0506+056 using the absolute R-band magnitude.[40] This is in agreement with previous studies suggesting that in the case of proton synchrotron models or models producing a high neutrino flux, the required jet luminosity exceeds that of the Eddington limit.[41] Within the hybrid model, matching the neutrino flux with the neutrino event of TXS 0506+056, displayed on panel b) of Figure 1, requires the jet luminosity to be $\sim 10^{50}\,{\rm erg\,s^{-1}}$, significantly exceeding that of the Eddington luminosity. Although the Eddington luminosity is not a strict limit and super-Eddington luminosities were previously reported,[42] this value is extremely large. On the other hand, for the neutrino flare in 2014-2015, when assuming that

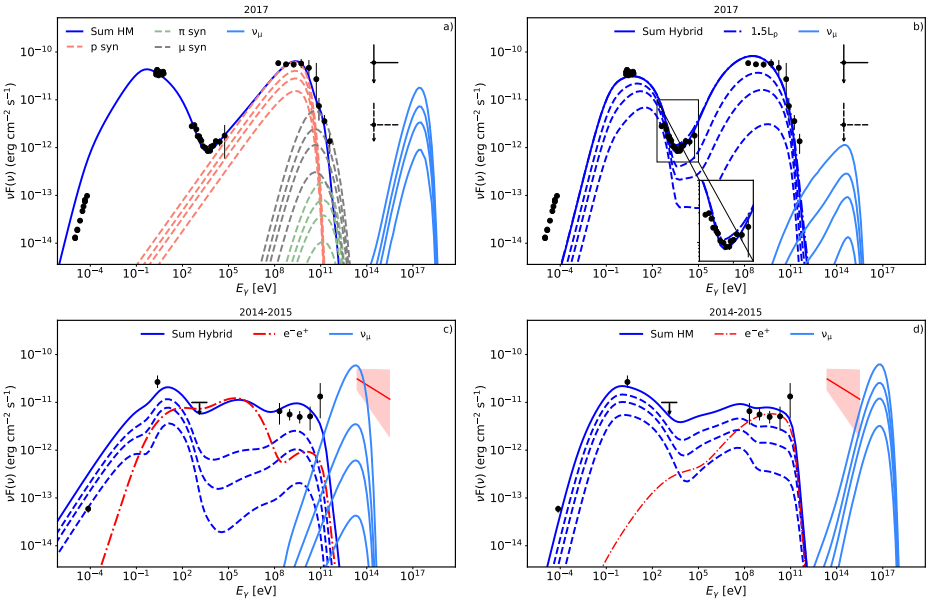


Fig. 1. The multiwavelength SED of TXS 0506+056 during the neutrino emission in 2017 (upper panels) and during the neutrino flare in 2014-2015 (lower panels) modeled within the hadronic and lepto-hadronic hybrid scenarios. The solid blue line in all plots represents the sum of all components which has been corrected for EBL absorption considering the model of.[43]

the emission from the secondary pairs solely dominates in the X-ray and γ-ray bands, an unrealistically high luminosity of $\sim 10^{52}$erg s^{-1} is obtained. Indeed, matching the high neutrino flux with the large radius (10^{17} cm) imposed by the modeling in this case, requires a large protons density, hence the too large proton luminosity. In the alternative interpretation, when the emission from the secondary pairs dominates in the GeV band, a modest luminosity of 3.4×10^{47} erg s^{-1} is estimated.

The expected numbers of neutrinos during the 6 months flare of TXS 0506+056 are 0.43 and 0.23 for the hadronic and hybrid scenarios, respectively. During the 2014-2015 neutrino flare, our most optimistic models predict 3.0-3.3 neutrinos for a 6 months exposure time (note however that the IceCube observational window was $\sim$ 110 days). Our results are in agreement with previous estimations for TXS 0506+056 and confirm that within a one-zone scenarios, 13 ± 5 events from the direction of TXS 0506+056 cannot be explained.[38,39]

3HSP J095507.9+355101: It is the first time that the jet of an extreme blazar is associated with a neutrino event,[30] straightening the assumption that the jets of this blazar type are potential sites for cosmic rays and even ultra-high energy cosmic ray acceleration.[44]

The SED of 3HSP J095507.9+355101 is shown in Figure 2, where the multiwavelength data are from.[45] Optical, UV and X-ray data were acquired on the 8$^{\text{th}}$, 10$^{\text{th}}$ and 11$^{\text{th}}$ of January. However, since the data taken on the 8$^{\text{th}}$ and 11$^{\text{th}}$ of January

seem to have the same flux and spectral shape,[46] we only model the data from the $8^{\rm th}$. The lack of available multiwavelength data does not allow to constrain the low and high energy peaks, which hardens the estimation of the model free parameters. A hint of a $20-30$ minutes variability has been found in the NICER and NuSTAR data, but only at the $\sim 3.5\,\sigma$ level.[47] Therefore, the compactness of the emitting region cannot be constrained. In order to keep the generality, the SED of 3HSP J095507.9+355101 is modeled for two different parameter configurations. For the HM, we consider $R' = 3\times10^{14}$ cm and $\delta = 15$, while for the hybrid model we assume $R' \simeq 10^{16}$ cm and $\delta = 30$. The all estimated parameters are shown in Table 3.

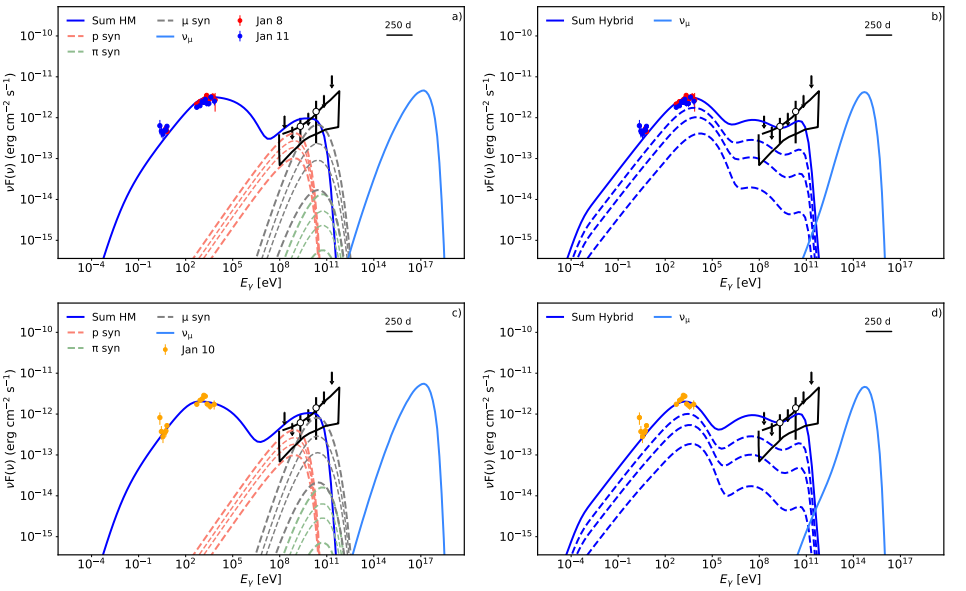


Fig. 2. The multiwavelength SEDs of 3HSP J095507.9+355101 on the the $8^{\rm th}$ and the $11^{\rm th}$ of January (panels a) and b) on the top raw) and on the $10^{\rm th}$ of January (panels c) and d) on the bottom raw). The data are taken from.[45,46] The observed spectrum including all processes is shown by the solid blue line. All models (solid blue lines) have been corrected for EBL absorption considering the model of.[43]

The estimated muon neutrino rate for 3HSP J095507.9+355101, is within $6\times10^{-4} - 4.8\times10^{-3}$ per day. This implies that under this rate of emission, the expected number of neutrinos to be detected by IceCube in a time corresponding to the duration of the flare is very low. Such low neutrino estimation was obtained also by the authors of,[46] where they concluded that in the most promising scenarios, there is a $\sim 1\%$ to $\sim 3\%$ to observe one neutrino over the time characteristic of the long-term emission of 3HSP J095507.9+355101 (years).

3C 279: The broadband emission from the powerful FSRQs 3C 279 at redshift $z = 0.536$ has characterized by high amplitude variability almost in all energy

bands e.g., order of minutes,[48] and in particular in the HE γ-ray band, which present the fastest variability. On the 16$^{\rm th}$ of June 2015, Fermi LAT observations showed that 3C 279 was in an exceptionally bright state. The flux increased up to 3.6×10^{-5} photon cm^{-2} s^{-1} with a flux doubling time on the order of 5 minutes.[49] IceCube performed a time-dependent neutrino signal search correlated with this γ-ray flare but no evidence for a signal was found.[50] We apply *SOPRANO* to model the SED of 3C 279 during its flare in the framework of the HM, when the second peak dominates by proton synchrotron emission. The parameters of our modeling are given in Table 3. Figure 3 shows the multiwavelength SED of 3C 279 taken from,[51] alongside with the results of our modeling. During the brightening, the X-ray emission of the source appears with a hard photon index < 1.50, smoothly connecting with the INTEGRAL data, described by a power-law with index 1.08.[51] In the HE γ-ray band, the spectrum presents a power-law with photon index 2.21 with a turn over.[52] We make the hypothesis that the HE component is produced from a single mechanism. In our modeling it is interpreted as proton synchrotron emission, represented by the red dashed lines in Figure 3. This interpretation requires that protons are accelerated up to $\gamma_{\rm p,max} = 2.1 \times 10^8$, see Table 3.

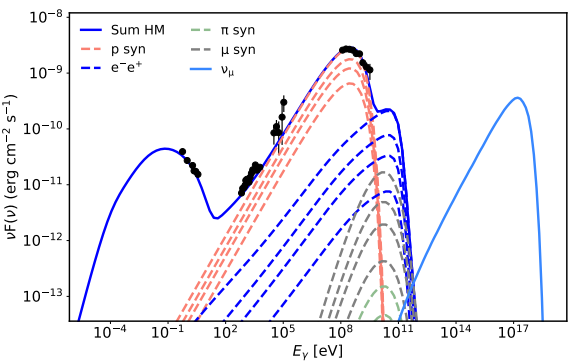


Fig. 3. The multiwavelength SED of 3C 279 during the exceptional flaring activity in 2015. The contribution of different particle emission is shown by dashed lines whereas the thick solid line represents the observed spectrum, corrected for EBL absorption considering the model of.[43]

The compactness of the emitting region implies a high efficiency for photo-pion and photo-pair interactions, which inject energetic secondary pairs. The contribution of these pairs dominates above ~ 10 GeV and peak at ~ 100 GeV, as can be seen by the blue dashed line in Figure 3. The synchrotron radiation of the primary electrons peaks at ~ 0.1 eV and its HE tail accounts for the observed optical/UV data. These data constrain the cut-off energy to be relatively low, $\gamma_{\rm e,cut} = 2.4 \times 10^2$, otherwise, for $B = 70$ G and $\delta = 55$, the synchrotron radiation would overshoot the observed flux in optical and UV bands. A similar hadronic modeling for this flare is presented in[51] and in.[53] The neutrino rate for the considered flaring period

is as high as 0.15 per day is estimated. However, for a relatively short period of the source activity, from minutes to one day, no neutrino events in the IceCube detector are expected, in agreement with.[50]

Table 3. Parameters used to model the multiwavelength SEDs of 3HSP J095507.9+355101 and 3C 279. The electron, proton and magnetic luminosity is also displayed.

	3HSP J095507.9+355101				3C 279
	January 8th		January 10th		
	Hadronic	Lepto-hadronic	Hadronic	Lepto-hadronic	Hadronic
δ	15	30	15	30	55
$R/10^{15}$ cm	0.3	10	0.3	10	0.32
$B[G]$	45	0.11	45	0.08	70
$\gamma_{\rm e,min}$	10^4	100	5×10^3	100	1
$\gamma_{\rm e,cut}$	6×10^5	2×10^6	2×10^5	7×10^5	2.4×10^2
$\gamma_{\rm e,max}$	9×10^5	6×10^6	5×10^5	6×10^6	4×10^2
$\alpha_{\rm e}$	1.9	2.0	1.9	2	1.8
$\alpha_{\rm p}=\alpha_{\rm e}$	1.9	2.0	1.9	2	1.8
$\gamma_{\rm p,min}$	1	1	1	1	1
$\gamma_{\rm p,max}$	9×10^8	10^6	9×10^8	10^6	2.1×10^8
$L_{\rm e}$ (erg s^{-1})	1.2×10^{44}	1.6×10^{44}	7.3×10^{43}	2.1×10^{44}	1.9×10^{44}
$L_{\rm B}$ (erg s^{-1})	1.5×10^{44}	4.1×10^{42}	1.5×10^{44}	2.2×10^{42}	5.7×10^{45}
$L_{\rm p}$ (erg s^{-1})	3.2×10^{46}	8.0×10^{50}	3.2×10^{46}	1.8×10^{51}	1.3×10^{49}

6. Conclusion

We have presented a new kinetic model of photo-hadronic and leptonic interactions aiming at studying the emission of optically thin (for Compton scattering) scenarios of relativistic sources (e.g., AGNs and GRBs). Our numerical solution of the kinetic equations for protons, neutrons, pions, muons, neutrinos, pairs, and photons conserves the total energy of the system as well as the number of particles where required. The code takes as an input the spectral injection rate of the particles (e.g., electrons and/or protons), and compute the time evolution of all relevant particles, including the secondaries, as they interact and cool, allowing the computation of the broadband emission spectrum at any given period. In this paper, we have applied *SOPRANO* to model the SEDs of three blazars, two of which are potentially associated to neutrino emission observed by IceCube. We have assumed different models for the production of the HE component and compute in all cases the expected number of muon neutrinos. The time-dependent nature of the code allowed to follow the evolution of all particles in one dynamical time scale and then assess the proton content in the jet by using the radiative spectrum of either secondaries or initial particles. This is necessary for the estimation of the expected number of neutrinos. Such time-dependent treatment of the particle evolution enabled us to constrain different scenarios of neutrino production by using the limits imposed by the observations in different bands.

References

1. IceCube Collaboration, Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector, *Science* **342**, p. 1242856 (November 2013).
2. M. G. Aartsen, et al., First Observation of PeV-Energy Neutrinos with IceCube, *Physical Review Letters* **111**, p. 021103 (July 2013).
3. M. G. Aartsen, et al., Characteristics of the Diffuse Astrophysical Electron and Tau Neutrino Flux with Six Years of IceCube High Energy Cascade Data, *Phys. Rev. Lett.* **125**, p. 121104 (Sep 2020).
4. IceCube Collaboration, et al., Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A, *Science* **361**, p. eaat1378 (July 2018).
5. IceCube Collaboration, et al., Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert, *Science* **361**, 147 (July 2018).
6. M. C. Begelman, et al., Consequences of relativistic proton injection in active galactic nuclei, *The Astrophysical Journal* **362**, 38 (October 1990).
7. M. Sikora, et al., Radiation drag in relativistic active galactic nucleus jets, *Monthly Notices of the RAS* **280**, 781 (June 1996).
8. A. M. Atoyan and C. D. Dermer, Neutral Beams from Blazar Jets, *The Astrophysical Journal* **586**, 79 (Mar 2003).
9. MAGIC Collaboration, et al., Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).
10. J. Kakuwa, et al., Prospects for detecting gamma-ray bursts at very high energies with the Cherenkov Telescope Array, *Monthly Notices of the RAS* **425**, 514 (September 2012).
11. Gilmore Rudy C., et al., IACT observations of gamma-ray bursts: prospects for the Cherenkov Telescope Array, *Experimental Astronomy* **35**, 413 (2013).
12. The IceCube-Gen2 Collaboration, et al., IceCube-Gen2: The Window to the Extreme Universe, *arXiv e-prints*, p. arXiv:2008.04323 (August 2020).
13. M. Agostini, et al., The Pacific Ocean Neutrino Experiment, *Nature Astronomy* **4**, 913 (September 2020).
14. S. Adrián-Martínez, et al., Letter of intent for KM3NeT 2.0, *Journal of Physics G Nuclear Physics* **43**, p. 084001 (August 2016).
15. C. Baikal-GVD, et al., Neutrino telescope in Lake Baikal: Present and Future, in *36th International Cosmic Ray Conference (ICRC2019)*, International Cosmic Ray Conference Vol. 36July 2019.
16. M. G. Aartsen, et al., Time-Integrated Neutrino Source Searches with 10 Years of IceCube Data, *Physical Review Letters* **124**, p. 051103 (February 2020).
17. S. Kelner and F. Aharonian, Energy spectra of gamma rays, electrons, and neutrinos produced at interactions of relativistic protons with low energy radiation, *Physical Review D* **78**, p. 034013 (2008).
18. K. Asano and S. Inoue, Prompt GeV-TeV Emission of Gamma-Ray Bursts Due to High-Energy Protons, Muons, and Electron-Positron Pairs, *The Astrophysical Journal* **671**, 645 (December 2007).
19. M. Böttcher, et al., Leptonic and Hadronic Modeling of Fermi-detected Blazars, *The Astrophysical Journal* **768**, p. 54 (May 2013).
20. M. Cerruti, et al., A hadronic origin for ultra-high-frequency-peaked BL Lac objects, *Monthly Notices of the RAS* **448**, 910 (March 2015).
21. A. Zech, et al., Expected signatures from hadronic emission processes in the TeV spectra of BL Lacertae objects, *Astronomy and Astrophysics* **602**, p. A25 (June 2017).

22. Y. G. Zheng, et al., A time-dependent particle acceleration and emission model: understanding particle spectral evolution and blazar flares, *Monthly Notices of the RAS* **499**, 1188 (October 2020).
23. A. Mastichiadis and J. G. Kirk, Self-consistent particle acceleration in active galactic nuclei., *Astronomy and Astrophysics* **295**, p. 613 (March 1995).
24. A. Pe'er and E. Waxman, Time-dependent Numerical Model for the Emission of Radiation from Relativistic Plasma, *The Astrophysical Journal* **628**, 857 (August 2005).
25. S. Dimitrakoudis, et al., The time-dependent one-zone hadronic model. First principles, *Astronomy and Astrophysics* **546**, p. A120 (October 2012).
26. C. Diltz, et al., Time Dependent Hadronic Modeling of Flat Spectrum Radio Quasars, *The Astrophysical Journal* **802**, p. 133 (April 2015).
27. S. Gao, et al., On the Direct Correlation between Gamma-Rays and PeV Neutrinos from Blazars, *The Astrophysical Journal* **843**, p. 109 (July 2017).
28. D. Kantzas, et al., A new lepto-hadronic model applied to the first simultaneous multiwavelength data set for Cygnus X-1, *Monthly Notices of the RAS* (October 2020).
29. S. Gasparyan, et al., Time-dependent lepto-hadronic modeling of the emission from blazar jets with SOPRANO: the case of TXS 0506+056, 3HSP J095507.9+355101 and 3C 279, *arXiv e-prints*, p. arXiv:2110.01549 (October 2021).
30. IceCube Collaboration, IceCube-200107A: IceCube observation of a high-energy neutrino candidate event, *GRB Coordinates Network* **26655**, p. 1 (January 2020).
31. M. Cerruti, et al., Leptohadronic single-zone models for the electromagnetic and neutrino emission of TXS 0506+056, *Monthly Notices of the RAS* **483**, L12 (February 2019).
32. S. Gao, et al., Modelling the coincident observation of a high-energy neutrino and a bright blazar flare, *Nature Astronomy* **3**, 88 (January 2019).
33. M. Petropoulou and A. Mastichiadis, Bethe-Heitler emission in BL Lacs: Filling the gap between X-rays and γ-rays, *Monthly Notices of the RAS* **447**, 36 (February 2015).
34. M. G. Aartsen, et al., Search for steady point-like sources in the astrophysical muon neutrino flux with 8 years of IceCube data, *European Physical Journal C* **79**, p. 234 (March 2019).
35. G. Fantini, et al., The formalism of neutrino oscillations: an introduction, *arXiv e-prints*, p. arXiv:1802.05781 (February 2018).
36. A. Keivani, et al., A Multimessenger Picture of the Flaring Blazar TXS 0506+056: Implications for High-energy Neutrino Emission and Cosmic-Ray Acceleration, *The Astrophysical Journal* **864**, p. 84 (September 2018).
37. R. Blandford and D. Eichler, Particle acceleration at astrophysical shocks: A theory of cosmic ray origin, *Physics Reports* **154**, 1 (October 1987).
38. A. Reimer, et al., Cascading Constraints from Neutrino-emitting Blazars: The Case of TXS 0506+056, *The Astrophysical Journal* **881**, p. 46 (August 2019).
39. X. Rodrigues, et al., Leptohadronic Blazar Models Applied to the 2014-2015 Flare of TXS 0506+056, *Astrophysical Journal, Letters* **874**, p. L29 (April 2019).
40. P. Padovani, et al., TXS 0506+056, the first cosmic neutrino source, is not a BL Lac, *Monthly Notices of the RAS* **484**, L104 (March 2019).
41. R. Xue, et al., On the Minimum Jet Power of TeV BL Lac Objects in the p-γ Model, *The Astrophysical Journal* **871**, p. 81 (January 2019).
42. Y.-F. Jiang, et al., Super-Eddington Accretion Disks around Supermassive Black Holes, *The Astrophysical Journal* **880**, p. 67 (August 2019).
43. A. Domínguez, et al., Extragalactic background light inferred from AEGIS galaxy-SED-type fractions, *Monthly Notices of the RAS* **410**, 2556 (February 2011).

44. P. Padovani, et al., Extreme blazars as counterparts of IceCube astrophysical neutrinos, *Monthly Notices of the RAS* **457**, 3582 (April 2016).
45. P. Giommi, et al., 3HSP J095507.9+355101: A flaring extreme blazar coincident in space and time with IceCube-200107A, *Astronomy and Astrophysics* **640**, p. L4 (August 2020).
46. M. Petropoulou, et al., Comprehensive Multimessenger Modeling of the Extreme Blazar 3HSP J095507.9+355101 and Predictions for IceCube, *The Astrophysical Journal* **899**, p. 113 (August 2020).
47. V. S. Paliya, et al., Multifrequency Observations of the Candidate Neutrino-emitting Blazar BZB J0955+3551, *The Astrophysical Journal* **902**, p. 29 (October 2020).
48. J. Aleksić, et al., MAGIC observations and multifrequency properties of the flat spectrum radio quasar 3C 279 in 2011, *Astronomy and Astrophysics* **567**, p. A41 (July 2014).
49. M. Ackermann, et al., Minute-timescale >100 MeV γ-Ray Variability during the Giant Outburst of Quasar 3C 279 Observed by Fermi-LAT in 2015 June, *Astrophysical Journal, Letters* **824**, p. L20 (June 2016).
50. R. Abbasi, et al., A Search for Time-dependent Astrophysical Neutrino Emission with IceCube Data from 2012 to 2017, *The Astrophysical Journal* **911**, p. 67 (April 2021).
51. E. Bottacini, et al., 3C 279 in Outburst in 2015 June: A Broadband SED Study Based on the INTEGRAL Detection, *The Astrophysical Journal* **832**, p. 17 (November 2016).
52. V. S. Paliya, Fermi-Large Area Telescope Observations of the Exceptional Gamma-Ray Flare from 3C 279 in 2015 June, *Astrophysical Journal, Letters* **808**, p. L48 (August 2015).
53. M. Petropoulou, et al., A hadronic minute-scale GeV flare from quasar 3C 279?, *Monthly Notices of the RAS* **467**, L16 (May 2017).

Multiwavelength study of high-redshift blazars

G. Harutyunyan and D. Israyelyan
ICRANet Armenia, Marshall Baghramian Avenue 24a, Yerevan 0019, Republic of Armenia

High-redshift blazars are among the most powerful objects in the Universe. The spectral and temporal properties of 33 distant blazars ($z > 2.5$) detected in the high-energy γ-ray band are discussed using *Fermi*-LAT and *Swift* Ultraviolet and Optical Telescope/X-ray Telescope (UVOT/XRT) data accumulated during 2008-2018. The properties of the jets of these blazar obtained by modeling the multiwavelength spectral energy distributions within a one-zone leptonic scenario are presented and discussed.

Keywords: Galaxies; active –Gamma rays; galaxies–radiation mechanisms; non-thermal.

1. Introduction

Blazars are a type of active galactic nuclei(AGN) which have relativistic jet that is directed towards to the Earth.[1] Blazars characteristic features are high degree of polarization, short and strong non-thermal emission variations both in time and amplitude. Based on the absence or presence of emission lines in their optical spectra blazars are divided in two large classes, namely FSRQs and BL Lacs.For FSRQs the equivalent width of emission line is >5 $\mathring{A}$ meanwhile BL Lacs have weak emission lines <5 $\mathring{A}$.[1]

The spectral energy distribution (SED) of blazars has a double-peaked structure. Low energy peak, from radio to X-ray is due to synchrotron radiation of relativistic electrons. Instead, the high energy peak is most probably due to the Inverse Compton scattering of either synchrotron photons(Synchrotron-Self Compton(SSC)[2–4]) or photons field external to the jet (external Inverse Compton).[5–7] The external target fields can be photons of the accretion disc, broad-line region and dusty torus.

The dominant class of extragalactic γ-ray sources are blazars. During its ten year survey, Fermi-LAT has observed $\sim$ 5000 sources in γ-ray band from which more than 2863 are AGNs detected above the 5σ.[8] Only 33(1.18%) out of 2863 are distant blazars $z > 2.5$ (Table 1) from which 26 are FSRQs, 2 are BL Lacs and 5 are blazar candidate sources. Studying of these sources can shed light on some questions such as, cosmological evolution of blazars and black holes, evolution of relativistic jets across different cosmic epochs, accretion disc-jet connection, extragalactic Background Light density limit, etc.[9]

In this paper were investigated multiwavelength emissions from high red shift blazars($z > 2.5$) using data accumulated by *Fermi*-LAT and Swift telescopes.

Table 1. List of γ-ray emitting blazars at $z > 2.5$. The results of observation of distant blazars from August 4, 2008, to August 4, 2018 by *Fermi*-LAT are presented in the right part of the table.

Object	4FGL name	Class	RA	Dec	F_γ [1]	Photon index	TS	$E_{\rm max}$ [2]	Probability	z
GB 1508+5714	J1510.1+5702	FSRQ	227.54	57.04	0.76 ± 0.15	2.95 ± 0.15	63.50	8.76	0.8268	4.31
PKS 1351-018	J1354.3-0206	FSRQ	208.50	-2.11	1.11 ± 0.15	2.69 ± 0.09	98.49	1.30	0.7010	3.72
MG3 J163554+3629	J1635.6+3628	BCU	248.92	36.48	1.46 ± 0.27	2.84 ± 0.12	123.95	3.74	0.8413	3.65
NVSS J121915+365718	J1219.0+3653	BCU	184.77	36.89	0.22 ± 0.08	2.20 ± 0.14	40.56	2.67	0.8273	3.53
PKS 0335-122	J0337.8-1157	FSRQ	54.47	-11.96	0.72 ± 0.16	2.69 ± 0.13	49.47	10.19	0.9245	3.44
PKS 0537-286	J0539.9-2839	FSRQ	84.99	-28.66	4.38 ± 0.18	2.72 ± 0.03	1694.80	7.60	0.9619	3.10
TXS 0800+618	J0805.4+6147	FSRQ	121.36	61.80	2.23 ± 0.14	2.82 ± 0.05	475.17	4.83	0.8441	3.03
S4 1427+543	J1428.9+5406	FSRQ	217.23	54.11	0.59 ± 0.15	2.62 ± 0.14	67.10	10.26	0.7831	3.01
GB6 J0733+0456	J0733.8+0455	FSRQ	113.47	4.93	1.14 ± 0.15	2.39 ± 0.07	197.26	17.65	0.9854	3.01
B2 0743+25	J0746.4+2546	FSRQ	116.60	25.77	2.06 ± 0.19	2.87 ± 0.07	280.99	1.52	0.7449	2.99
PKS 0347-211	J0349.8-2103	FSRQ	57.47	-21.06	3.25 ± 0.15	2.47 ± 0.03	1372.81	5.83	0.9690	2.94
S4 1124+57	J1127.4+5648	FSRQ	171.87	56.80	0.96 ± 0.13	2.75 ± 0.09	147.30	7.09	0.8495	2.89
MRSS 291-081526	J2313.9-4501	BCU	348.49	-45.02	0.89 ± 0.20	2.80 ± 0.13	81.12	2.25	0.7550	2.88
PKS 0438-43	J0440.3-4333	FSRQ	70.09	-43.55	2.24 ± 0.20	2.60 ± 0.05	574.84	2.00	0.9225	2.85
S4 2015+65	J2015.4+6556	FSRQ	303.86	65.95	0.39 ± 0.15	2.37 ± 0.15	23.77	4.34	0.7257	2.84
87GB 214302.1+095227	J2145.5+1006	BL Lac	326.38	10.12	0.048 ± 0.014	1.71 ± 0.19	40.46	67.15	0.9874	2.83
MG2 J174803+3403	J1748.0+3403	FSRQ	267.01	34.06	0.97 ± 0.13	2.31 ± 0.07	284.21	31.63	0.9934	2.76
PKS 0834-20	J0836.5-2026	FSRQ	129.13	-20.45	1.93 ± 0.19	2.94 ± 0.08	171.15	1.15	0.5796	2.75
TXS 0222+185	J0224.9+1843	FSRQ	36.24	18.72	1.72 ± 0.22	3.05 ± 0.12	101.08	2.59	0.5327	2.69
OD 166	J0242.3+1102	FSRQ	40.59	11.05	1.94 ± 0.20	2.59 ± 0.06	252.83	6.91	0.8651	2.68
CRATES J233930+024420	J2339.6+0242	BCU	354.90	2.71	1.00 ± 0.21	2.58 ± 0.11	94.50	6.17	0.8537	2.66
TXS 0907+230	J0910.6+2247	FSRQ	137.67	22.80	1.17 ± 0.14	2.37 ± 0.06	262.58	5.38	0.9578	2.66
PMN J1441-1523	J1441.6-1522	FSRQ	220.41	-15.38	0.65 ± 0.23	2.32 ± 0.13	72.02	11.77	0.9535	2.64
CRATES J105433+392803	J1054.2+3926	BCU	163.56	39.44	0.27 ± 0.09	2.30 ± 0.16	34.05	4.79	0.9000	2.63
MG1 J154930+1708	J1549.6+1710	BL Lac	237.41	17.18	0.17 ± 0.08	2.01 ± 0.16	44.75	17.03	0.9759	2.62
TXS 1448+093	J1450.4+0910	FSRQ	222.62	9.18	0.76 ± 0.13	2.35 ± 0.08	130.24	10.58	0.9599	2.61
PMN J0226+0937	J0226.5+0938	FSRQ	36.63	9.63	0.48 ± 0.12	2.18 ± 0.10	96.49	56.42	0.9523	2.61
PKS 0451-28	J0453.1-2806	FSRQ	73.29	-28.11	5.83 ± 0.17	2.66 ± 0.02	3118.20	10.77	0.9815	2.56
B3 0908+416B	J0912.2+4127	FSRQ	138.06	41.46	1.51 ± 0.14	2.42 ± 0.05	539.57	12.10	0.9903	2.56
TXS 1616+517	J1618.0+5139	FSRQ	244.52	51.67	0.69 ± 0.13	2.68 ± 0.12	72.65	12.38	0.8408	2.56
B3 1624+414	J1625.7+4134	FSRQ	246.45	41.57	1.38 ± 0.14	2.49 ± 0.06	395.12	9.69	0.9787	2.55
B3 1343+451	J1345.5+4453c	FSRQ	206.39	44.88	15.01 ± 0.16	2.25 ± 0.008	34652.79	24.25	0.9994	2.53
PKS 2107-105	J2110.2-1021c	FSRQ	317.56	-10.36	0.88 ± 0.18	2.66 ± 0.13	53.29	5.46	0.6500	2.50

Note: [1] Integrated γ-ray flux in units of $\times 10^{-8}$photon cm^{-2} s^{-1}.
Note: [2] Photon energy in GeV.

Table 2. Spectral parameters of the sources modeled with logparabola.

Object	F_γ [1]	α	β
PKS 1351-018	$(6.86 \pm 1.66) \times 10^{-9}$	2.20 ± 0.23	0.63 ± 0.23
PKS 0537-286	$(4.15 \pm 0.20) \times 10^{-8}$	2.66 ± 0.04	0.10 ± 0.03
TXS 0800+618	$(2.04 \pm 0.16) \times 10^{-8}$	2.67 ± 0.08	0.17 ± 0.06
B2 0743+25	$(1.72 \pm 0.23) \times 10^{-8}$	2.53 ± 0.14	0.40 ± 0.13
PKS 0347-211	$(2.69 \pm 0.17) \times 10^{-8}$	2.32 ± 0.05	0.19 ± 0.03
PKS 0438-43	$(1.15 \pm 0.23) \times 10^{-8}$	2.35 ± 0.12	0.48 ± 0.11
S4 2015+65	$(1.66 \pm 1.00) \times 10^{-9}$	2.42 ± 0.35	0.42 ± 0.31
PKS 0834-20	$(1.76 \pm 0.20) \times 10^{-8}$	2.62 ± 0.16	0.29 ± 0.13
OD 166	$(1.45 \pm 0.24) \times 10^{-8}$	2.43 ± 0.11	0.24 ± 0.08
TXS 0907+230	$(8.68 \pm 1.77) \times 10^{-8}$	2.28 ± 0.09	0.12 ± 0.06
PKS 0451-28	$(5.56 \pm 0.18) \times 10^{-8}$	2.56 ± 0.04	0.09 ± 0.02
B3 1343+451	$(1.41 \pm 0.02) \times 10^{-7}$	2.18 ± 0.01	0.07 ± 0.006

Note: [1] Integrated γ-ray flux in units of photon $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

2. Fermi LAT analysis

To investigate considered sources in the γ-ray band the data accumulated by *Fermi*-LAT were used. In the current paper, the 100 MeV to 500 GeV data from the region with 12° radius were downloaded and analysed. The data had been accumulated during ten years since launch of *Fermi*-LAT (4 August 2008 to 4 August 2018). *gtselect* and *gtmktime* tools were applied to select good time intervals when data were valid. The maximum zenith angle was 90° to avoid γ-rays which are due to Earth's limb. Events were binned into 16.9×16.9 ROI (region of interest) and 37 equal logarithmically spaced energy bins using *gtbin* tool. The model file of each source includes all the 4FGL sources falling within the ROI+5° region and it was generated using 4FGL-DR2 catalog which is the second release of 4FGL (based on ten years operation). For galactic and extragalactic diffuse emission components *gll_iem_v07* and *iso_P8R3_SOURCE_V2_v1* models were used. With the *gtlike* tool spectral analysis of considered sources using binned maximum likelihood method was performed. The spectral parameters of sources located within the ROI are set to free. Detection significance of the sources estimated using TS (Test Statistic)[10] which is defined as $TS{=}2 \times (logL_1 - logL_0)$ where L_1 is the likelihood of the data with a point source at the given position and L_0 without the source. For each source the same γ-ray spectral model as in 4FGL catalog was used and additionally analysis was performed using power-law spectral model. SEDs of the sources were created by running *gtlike* for smaller energy intervals equal in logarithmic space. The γ-ray light curves were computed with *gtlike* tool applying unbinned likelihood analysis method which is suitable for shorter periods. The spectra of the considered source was modeled using a power-law which provide good results for small interval. The normalization and photon index of the chosen sources are free to vary except photon index of all sources which are lying within the ROI. For the fitting the events

with energy 0.1-300 GeV were considered. To investigate γ-ray variability initially we calculated 30 day bins light curves for all the sources. There are statistically significant ($\geq 5\sigma$) sources in the sample which have flaring periods and for that periods the light curves with small bins such as day or week were calculated. For some powerful sources such as B3 1343+451, PKS 0537-286, PKS 0347-211, PKS 0451-28 additionally the adaptively binned method has been applied to generate the light curves. In this method time intervals are adapt to provide bins with constant flux uncertainty(with 15%) above the E_0 which is a powerful method to study flux variation in γ-ray band.[11–17]

3. Swift observations

In the current paper the data from the two of the instruments onboard Swift XRT[18] and UVOT[19] have been analysed. Although the primary scientific goal of the Swift telescope is the observation of GRBs,[20] also, due to its wide frequency coverage, it is suitable for studies of blazars. Except 4 sources from the Table 1 the rest of the sources at least once have been observed by Swift, some of them (e.g. B2 0743+25, PKS 0438-43, TXS 0222+185, TXS 1448+093, PMN J0226+0937, TXS 1616+517 and PKS 2107-105) had been observed more than ten times. The Swift XRT data analysis was available for twenty-two sources, and only twenty-six sources were detected in at least one of the optical-UV Swift-UVOT filters (see Table 3).

The Swift UVOT performs measurements at 6 wavelengths: UVW2 (188 nm), UVM2 (217 nm), UVW1 (251 nm), U (345 nm), B (439 nm), and V (544 nm). The photometry analysis of the sample was performed using the standard UVOT software distributed within the HEAsoft package (v6.25) and the calibration included in the CALDB (v.20170922). The source counts were extracted from a circular region with a 5″ radius around the source, while the background one is a region with 20″ radius not being contaminated with any signal from the nearby sources. If in the observation data from the different filters exists, the same models were used, after checking background region contamination by nearby sources. For converting source counts to magnitudes *uvotsource* tool and the conversion factors provided by[21] was used. The fluxes were corrected for extinction using the reddening coefficient $E(B-V)$ from the Infrared Science Archive[a].

In the X-ray band the Swift XRT data observed both in photon counting and window timing modes were used. The data were analyzed with the standard XRTDAS tool distributed within the HEAsoft package (v6.25), applying standard procedures, filtering, and screening criteria. As a source region was selected a circular region with a 20-pixel (47″) centered on the coordinates of each source, while as a background region was selected an annulus, with the same center as the source region and inner and outer radii of 51 (120″) and 85 pixels (200″), respectively. For fitting the 0.3-10 kev X-ray spectrum was used XSPEC12.10.1 tool, adopting

[a] `http://irsa.ipac.caltech.edu/applications/DUST/`

Table 3. Summary of Swift-UVOT observations of the considered sources. For the sources, when several observations were available, the fluxes in each band are computed from the summed images. Averaged flux in each band is in units of erg cm^{-2} s^{-1}.

Object	V	B	U	$W1$	$M2$	$W2$
PKS 1351-108	–	–	$(4.54 \pm 2.53) \times 10^{-14}$	–	–	
MG3 J163554 +3629	–	–	$(6.93 \pm 2.72) \times 10^{-14}$	–	–	$(7.50 \pm 4.11) \times 10^{-14}$
PKS 0537-286	$(3.78 \pm 5.86) \times 10^{-15}$	$(4.53 \pm 1.56) \times 10^{-14}$	$(3.54 \pm 0.42) \times 10^{-13}$	$(4.54 \pm 2.54) \times 10^{-14}$	$(6.84 \pm 0.75) \times 10^{-14}$	$(4.61 \pm 0.55) \times 10^{-13}$
TXS 0800+618	$(7.69 \pm 0.12) \times 10^{-15}$	–	$(2.39 \pm 0.6) \times 10^{-13}$	$(5.68 \pm 0.18) \times 10^{-15}$	–	$(6.58 \pm 0.13) \times 10^{-14}$
S4 1427+543	–	$(2.52 \pm 2.03) \times 10^{-14}$	$(2.53 \pm 0.92) \times 10^{-13}$	–	$(3.88 \pm 1.4) \times 10^{-14}$	$(1.50 \pm 0.02) \times 10^{-15}$
GB6 J0733+0456	–	$(2.97 \pm 2.55) \times 10^{-14}$	–	$(9.40 \pm 0.25) \times 10^{-15}$	–	–
B2 0743+25	–	$(4.92 \pm 0.82) \times 10^{-14}$	$(3.39 \pm 0.59) \times 10^{-13}$	–	$(1.12 \pm 0.22) \times 10^{-13}$	$(4.76 \pm 0.76) \times 10^{-13}$
PKS 0347-211	–	–	$(2.80 \pm 0.52) \times 10^{-13}$	–	$(1.69 \pm 0.5) \times 10^{-13}$	$(2.03 \pm 1.09) \times 10^{-13}$
S4 1124+57	$(1.16 \pm 0.85) \times 10^{-14}$	–	–	–	–	–
PKS 0438-43	$(7.90 \pm 0.1) \times 10^{-15}$	–	$(3.45 \pm 0.38) \times 10^{-13}$	–	–	$(8.02 \pm 0.69) \times 10^{-13}$
S4 2015+65	$(1.63 \pm 0.02) \times 10^{-15}$	–	–	$(6.91 \pm 0.33) \times 10^{-14}$	–	–
87 GB 214302.1 +095227	$(1.01 \pm 0.48) \times 10^{-13}$	$(2.49 \pm 0.36) \times 10^{-13}$	–	$(2.78 \pm 0.72) \times 10^{-13}$	$(3.49 \pm 0.63) \times 10^{-13}$	–
MG2 J174803 +3403	$(1.92 \pm 0.32) \times 10^{-13}$	$(2.53 \pm 0.25) \times 10^{-13}$	$(4.68 \pm 0.87) \times 10^{-13}$	$(2.40 \pm 0.33) \times 10^{-13}$	$(3.73 \pm 0.61) \times 10^{-13}$	$(3.20 \pm 1.52) \times 10^{-13}$
PKS 0834-20	–	–	$(7.82 \pm 0.54) \times 10^{-13}$	$(3.84 \pm 1.72) \times 10^{-14}$	$(2.39 \pm 0.36) \times 10^{-13}$	$(7.29 \pm 0.91) \times 10^{-13}$
TXS 0222+185	–	–	$(9.36 \pm 0.89) \times 10^{-13}$	–	$(8.18 \pm 1.11) \times 10^{-13}$	$(1.12 \pm 0.21) \times 10^{-12}$
OD 166	–	–	–	–	–	$(9.19 \pm 0.36) \times 10^{-14}$
TXS 0907+230	$(9.53 \pm 4.27) \times 10^{-14}$	$(9.66 \pm 3.17) \times 10^{-14}$	$(5.28 \pm 2.66) \times 10^{-13}$	$(8.77 \pm 3.31) \times 10^{-14}$	$(2.09 \pm 1.55) \times 10^{-13}$	–
PMN J1441-1523	–	–	$(6.33 \pm 4.81) \times 10^{-14}$	–	–	–
CRATES J05433 +392803	–	–	–	$(1.90 \pm 0.88) \times 10^{-16}$	–	–
TXS 1448+093	–	–	$(6.53 \pm 2.36) \times 10^{-14}$	–	–	$(1.35 \pm 0.35) \times 10^{-13}$
PMN J0226+0937	$(9.95 \pm 4.17) \times 10^{-14}$	$(5.13 \pm 0.9) \times 10^{-13}$	$(1.22 \pm 0.06) \times 10^{-12}$	$(4.39 \pm 1.53) \times 10^{-13}$	$(1.30 \pm 0.06) \times 10^{-12}$	$(1.26 \pm 0.08) \times 10^{-12}$
PKS 0451-28	–	$(9.52 \pm 1.21) \times 10^{-14}$	–	–	–	–
B3 0908+416B	$(3.45 \pm 0.11) \times 10^{-15}$	$(3.46 \pm 2.22) \times 10^{-14}$	$(2.08 \pm 0.66) \times 10^{-13}$	$(4.94 \pm 1.95) \times 10^{-14}$	$(6.87 \pm 1.91) \times 10^{-14}$	$(2.67 \pm 1.2) \times 10^{-13}$
TXS 1616+517	–	–	$(1.96 \pm 0.27) \times 10^{-13}$	–	–	$(1.25 \pm 0.34) \times 10^{-13}$
B3 1343+451	–	$(3.12 \pm 2.77) \times 10^{-14}$	$(4.54 \pm 0.55) \times 10^{-13}$	$(7.41 \pm 0.23) \times 10^{-15}$	$(2.52 \pm 0.29) \times 10^{-13}$	$(4.47 \pm 0.65) \times 10^{-13}$
PKS 2107-105	–	–	$(1.74 \pm 0.1) \times 10^{-12}$	–	–	$(1.44 \pm 0.13) \times 10^{-12}$

an absorbed power-law model with a neutral hydrogen column density fixed to its Galactic value in each direction.

As these sources are not observed with the Swift XRT instrument too long, for most of the sources the count rates were below 20, which would prevent spectral fitting, some cases several observations of the same source were merged if they were available for increasing the photon statistics. The merging was performed using the tool available from the UK Swift Science Data Center[b].[22] Again for the fitting of the 0.3-10 keV spectrum XSPEC was used, applying both the absorbed power-law and log-parabola models. Also, light curves have been generated for the investigation of the flux variation in each band.

4. Results of data analysis

In Table 1 the results of γ-ray data analysis(γ-ray flux, index, TS, etc.) obtained after power-law fitting are presented. Detection significance(TS) for all sources was TS $>$ 34.05, only S4 2015+65 detected with TS=23.77.Also, fit with log-parabola spectral model was performed for sources which spectral model is log-parabola in 4FGL(Table 2). Results show that fluxes estimated using log-parabola spectral model are not much different from corresponding fluxes estimated from the power-law fitting. The results presented in Table 1 are shown in Fig. 1. The γ-ray photon index versus flux is shown in he upper left panel.The photon index estimated in the 0.1-500 GeV ranges from 1.71 to 3.05 with a mean of 2.54. The flux of considered sources ranges from 4.84×10^{-10} to $1.5 \times 10^{-7} photon\ cm^{-2}s^{-1}$. The two BL Lacs, 87 GB 214302.1+095227 and MG1 J154930+1708, have the lowest flux $(4.84 \pm 1.37) \times 10^{-10} photon\ cm^{-2}s^{-1}$ and $(1.66 \pm 0.79) \times 10^{-9} photon\ cm^{-2}s^{-1}$. The γ-ray flux of BCUs included in the sample ranges from 2.19×10^{-9} to $1.46 \times 10^{-8} photon\ cm^{-2}s^{-1}$ with Γ_γ within 2.20-2.84. The FSRQs occupy the region of high flux $\geq 3.89 \times 10^{-9} photon\ cm^{-2}s^{-1}$ with a mean $2.11 \times 10^{-8} photon\ cm^{-2}s^{-1}$. The highest γ-ray flux of $1.50 \times 10^{-7} photon\ cm^{-2}s^{-1}$ was observed from the bright FSRQ B3 1343+451. The luminosity of each source was computed using the observed flux(F_γ) and photon index(Γ_γ). In upper right panel is shown the distribution of the considered sources in Γ_γ-L_γ plane. The γ-ray luminosity of the considered sources ranges from $1.01 \times 10^{47} erg\ s^{-1}$ to $5.54 \times 10^{48} erg\ s^{-1}$. The lowest luminosity of $1.01 \times 10^{47} erg\ s^{-1}$ has been estimated for CRATES J105433+392803 which is of the same order with that of the two BL Lacs included in the sample $(1.42 \pm 0.98) \times 10^{47} erg\ s^{-1}$ for 87GB 214302.1+095227 and $(1.24 \pm 0.59) \times 10^{47} erg\ s^{-1}$ for MG1 J154930+1708. The luminosity of these BL Lacs corresponds to the highest end of the luminosity distribution of BL Lacs included in the fourth catalog of AGNs detected by *Fermi*-LAT.[23] The γ-ray luminosity of PKS 0347-211, PKS 0451-28, PKS 0537-286 and B3 1343+451 $\geq 10^{48} erg\ s^{-1}$. X-ray data analysis results (Table 3) show that flux ranges from $\simeq 5 \times 10^{-14} erg\ cm^{-2}s^{-1}$ to

[b] `https://www.swift.ac.uk/user\_objects/`

$\simeq 10^{-14} erg\ cm^{-2} s^{-1}$, the highest flux of $(1.19 \pm 0.04) \times 10^{-11} erg\ cm^{-2} s^{-1}$ being observed for TXS 0222+185. In the lower left panel X-ray flux versus the photon index was plotted. Because of small number of detected counts flux and photon index were estimated with large uncertainties. X-ray photon indexes estimated for these sources is <2 which means that X-ray spectra has a rising shape. The hardest X-ray spectrum detected for B3 0908+416B with $\Gamma_x = 1.01 \pm 0.46$. PKS 0451-28, TXS 0222+185, PKS 0834-20, PKS 0537-286, TXS 0800+618, B2 0743+25 and PKS 0438-43 have a comparable high X-ray flux $\geq 2.13 \times 10^{-12} erg\ cm^{-2} s^{-1}$. In the lower right panel are compared γ-ray and X-ray fluxes. Sources which are bright in γ-ray band such as PKS 0537-286 and PKS 0451-28 appear to be,also a bright sources in X-ray band.

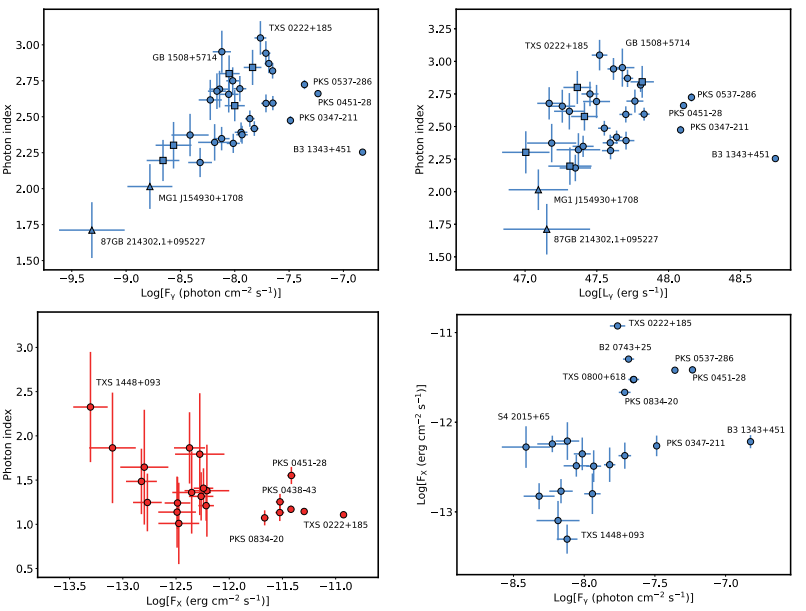


Fig. 1. The γ-ray flux (> 100 MeV) and luminosity of considered sources versus the photon index are shown in the upper panels. BL lacs, FSRQs and BCUs are shown with triangles, circles, and squares, respectively. *Lower left panel:* The X-ray flux ($0.3 - 10$ keV) versus the photon index. *Lower right panel:* Comparison of γ-ray and X-ray (if available) fluxes.

4.1. *Variability*

Multiple observations of some of these sources with the Swift telescope allow us to investigate their X-ray flux variability. After processing and analyzing data from each observation there were only eight objects with enough counts for constraining the flux and the photon index in a single observation. The results showed that the

X-ray emission from the considered sources was relatively constant except for PKS 0438-43, B2 0743+25, and TXS 0222+185.

The X-ray flux of TXS 0222+185 is mostly around $(0.96 - 1.1) \times 10^{-11}$ erg cm^{-2} s^{-1} which increased to $(1.62 \pm 0.13) \times 10^{-11}$ erg cm^{-2} s^{-1} and $(1.37 \pm 0.16) \times 10^{-11}$ erg cm^{-2} s^{-1} on 24 December 2014 and 31 July 2006, respectively. In case of PKS 0438-43, it was in a bright X-ray state on 15 December 2016 with a flux of $(1.09 \pm 0.16) \times 10^{-11}$ erg cm^{-2} s^{-1} as compared with the flux of $(1.30 \pm 0.31) \times 10^{-11}$ erg cm^{-2} s^{-1} in the quiescent state. Similarly, the X-ray flux of B2 0743+25 in the high state on 01 January 2006 was $(1.06 \pm 0.56) \times 10^{-11}$ erg cm^{-2} s^{-1}. The variation of the 0.3-10 keV X-ray flux of PKS 0438-43, B2 0743+25 and TXS 0222+185 is shown in Fig. 4. No variation of X-ray photon index was found, due to the large uncertainties in its estimation.

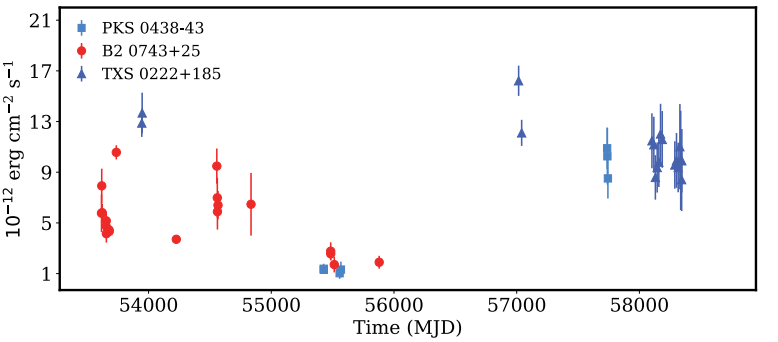


Fig. 2. The 0.3-10 keV X-ray fluxes of PKS 0438-43, B2 0743+25 and TXS 0222+185 measured by Swift-XRT.

A detailed investigation of γ-ray flux variation of the considered sources during the ten years was possible due to the continuous observation by *Fermi*-LAT. In some cases, when source detection significance is $< 10\sigma$, the data are not enough for variability searches in month scales, in these cases the variability has been checked for longer scales (6 to 12 months).[24] If the source detection significance is $> 37\,\sigma$, it is possible to use the adaptively binned intervals for computations. The adaptively binned light curves were possible to compute only for B3 1343+451, PKS 0451-28, PKS 0347-211 and PKS 0537-286, which are shown in Fig. 3. For the other sources included in Table 1 the variability on week and month scales are investigated.

The adaptively binned light curve of B3 1343+451, which is the brightest and the most variable source in the sample is shown in 3. During the bright γ-ray flaring periods: around MJD 55100, MJD 55890, MJD 56170, MJD 56640, and MJD 57060 the flux changes in sub-day scales. The width of 83 bins was shorter than 12 hours, which were observed during the flares around MJD 55890 (24 bins) and MJD 56170 (35 bins). There are about 200 bins with a width shorter than a day, and the

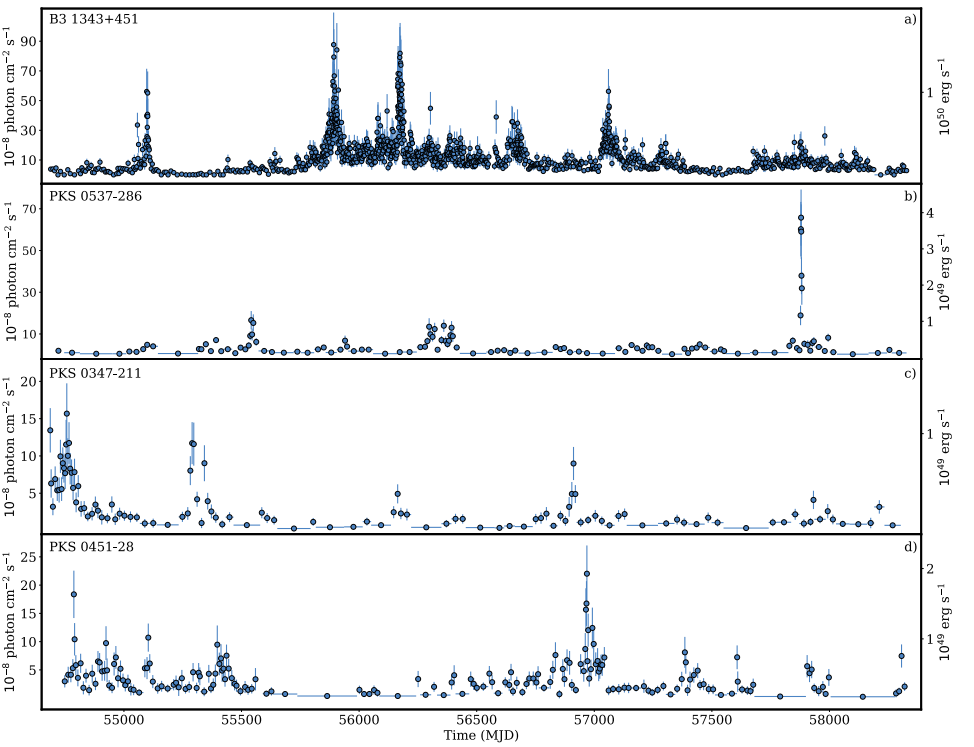


Fig. 3. γ-ray light curve of B3 1343+451, PKS 0537-286, PKS 0347-211, and PKS 0451-28 for the period from August 2008 to August 2018 calculated using adaptively-binned timescales.

minimum adaptively binned time width is 6.33 hours observed on MJD 56176.34 when the source flux was $(7.52\pm1.85)\times10^{-7}$photon cm^{-2} s^{-1}. In the quiescent state, the γ-ray flux of the source is $\simeq(1-5)\times10^{-8}$photon cm^{-2} s^{-1} with $\Gamma_\gamma = 2.48\pm0.29$ was observed on MJD 55891.7 with a detection significance of 9.1σ within a time bin having a width of 8.2 hours, which is 36.4 times higher than the γ-ray flux of the source in the quiescent state but the Γ_γ is within the uncertainties of the value given in Table 1. During the hardest γ-ray emission period, $\Gamma_\gamma = 1.45\pm0.21$ was detected on MJD 56432 with a significance of 7.2σ, which is unusual for this source.

PKS 0537-286 ($z = 3.10$) is the furthest source showing γ-ray flux variability in short time scales (Fig. 3b). The adaptive light curve showed variability in intra-day scales: 20.45, 15.96, and 18.90 hours. First about flaring activity of PKS 0537-286 was reported in[25] and the rapid γ-ray flux variations in.[24] γ-ray flux of PKS 0537-286 was increased from its 10 year average value, which was $(4.38\pm0.18)\times10^{-8}$ photon cm^{-2} s^{-1}, on MJD 55528-55553, MJD 56264-56400 and MJD 57878.0-57883.4, with the maximum value of $(6.58\pm1.35)\times10^{-7}$ photon cm^{-2} s^{-1} observed on MJD 57879.2 with a detection significance of 10.57σ. But in this periods the Γ_γ

is not significantly different from its average value in Table 1, which means that the emission is dominated by the MeV photons.

In the case of PKS 0347-211 and PKS 0451-28 (Fig. 3c and d) the adaptively binned light curves show several periods of γ-ray brightening. For example, for PKS 0347-211 the shortest γ-ray flux increase time interval is 2.65 days, and for PKS 0451-28 it is 1.56 days. The light curves of both sources show multiple periods when γ-ray flux increased. For example, on MJD 54757.04 ± 2.71 the γ-ray flux of PKS 0347-211 was $(1.57 \pm 0.41) \times 10^{-7}$ photon cm^{-2} s^{-1}. In the case of PKS 0451-28, the peak γ-ray flux of $(2.20 \pm 0.50) \times 10^{-7}$ photon cm^{-2} s^{-1} was observed on MJD 56968.60 ± 0.79 with 9.64σ. During this period, Γ_γ was 2.06 ± 0.19.

For finding out whether the γ-ray emission of other sources is variable or not,a simple $\chi 2$ test was performed;[26] the flux measured in each interval was fitted by a constant flux and the reduced $\chi 2$ and the probability of the flux being constant are computed. The $\chi 2$ fitting showed that the γ-ray emission variability of these sources varies from week to month scales. MG3 J163554+3629 at $z = 3.65$ is the most distant γ-ray flaring blazar; this source was reported to be in active state during May-July 2018: for example on July 7 2018 its daily averaged γ-ray flux peak value was $(6.4 \pm 1.15) \times 10^{-7}$ photon cm^{-2} s^{-1}.[27] During June-July 2018 this source was also bright and the maximum daily averaged γ-ray flux was observed $(1.18 \pm 0.23) \times 10^{-7}$ photon cm^{-2} s^{-1}.[27]

For the considered sources also investigated the γ-ray photon index evolution in time has been. This may reveal crucial information, because the photon index is defined by the processes responsible for particle acceleration and cooling and its changes are directly linked with the processes inside the jet. For the searching the photon index variation are used the γ-ray light curves produced in 30 and 7 days, in the case of B3 1343+451 3 day bins, and for escaping large uncertainties only those periods when the source detection was $TS > 25$ were considered, because adaptively binned light curves have narrow width, and photon indexes would be estimated with big uncertainties, which would would not allow to make definite conclusions. The $\chi 2$ test shows that the photon index of B3 1343+451, PKS 0451-28, B3 0908+416B and TXS 0907+230 only varies in time and the evolution was shown in Fig. 4 where the horizontal line corresponds to the averaged photon index estimated in ten years (Table 1).

The most interesting photon index variability has B3 1343+451, there are 60 periods when $\Gamma_\gamma < 2.0$. The hardest spectrum of B3 1343+451 was observed on 58089.16 ± 1.5 with $\Gamma_\gamma = 1.73 \pm 0.24$, also it is interesting, that the γ-ray spectrum was hard in a brightly γ-ray flaring state on MJD 56172.16 ± 1.5 with 1.95 ± 0.07. In the 30-day binned light curve of PKS 0451-28 are observed $\Gamma_\gamma = 2.06 \pm 0.07$ and $\Gamma_\gamma = 2.17 \pm 0.15$ on MJD 56977.66 ± 15 and 58297.66 ± 15, respectively. In the case of B3 0908+416B we can see remarkable hardening on MJD 57517.66 ± 15 when the Γ_γ changed to 1.84 ± 0.25 compared to the average of 10 years 2.42 ± 0.05. And in the case of the most distant object where γ-ray spectrum hardening was

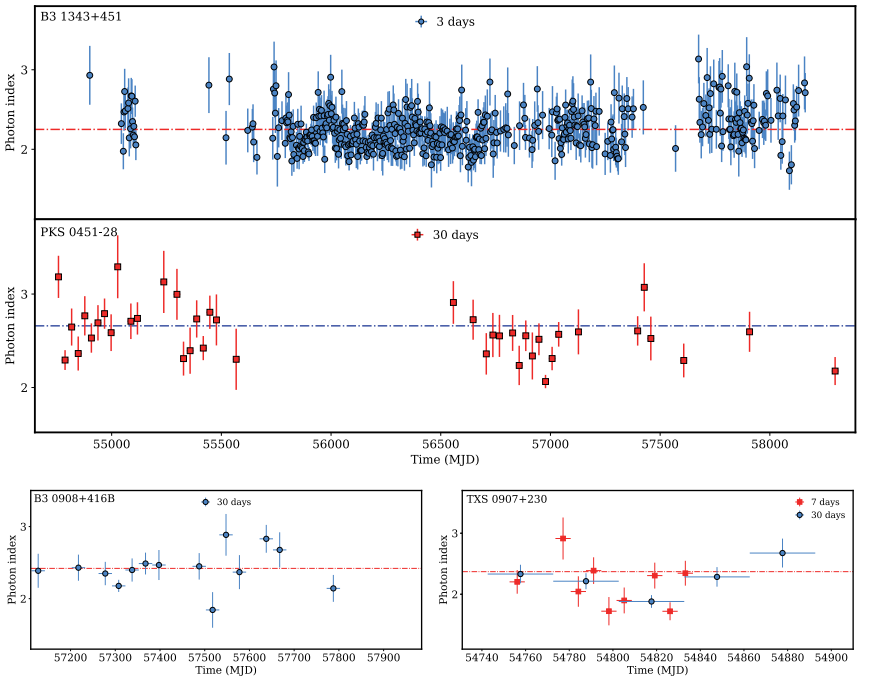


Fig. 4. The evolution of the γ-ray photon indexes of B3 1343+451, PKS 0451-28, B3 0908+416B and TXS 0907+230 in time.

observed is TXS 0907+230 with $z = 2.66$. It's 7 day γ-ray light curve shows that on MJD 54798.16 ± 3.5, 54805.16 ± 3.5 and 54826.16 ± 3.5 its γ-ray photon index appeared $\Gamma_\gamma = 1.72 \pm 0.23$, 1.90 ± 0.21 and 1.72 ± 0.15, respectively. Such a hard γ-ray spectrum is more typical for BL Lacs, but it has also been occasionally observed for the FSRQs during the flares.[12, 15, 28–31]

5. Theoretical modelling and results

The multiwavelength emission from blazars is due interaction of accelerated electrons[2–7] or protons[32–39] with magnetic and photon fields. The currently proposed models can not explain dynamical changes of the radiation and flares which are most likely caused by changes of radiating particles or the emission region.[40] Origin of flares can be investigated with contemporaneous multiwavelength data. From considered sources only for PKS 0438-43 and TXS 0800+618 are available contemporaneous Swift data. Our goal is to constrain the main physical parameters of distant blazars jets and because of lack of multiwavelength data in the modeling will be used only the averaged multiwavelength SEDs which presents the typical state of the source. To investigate broad band emission from considered sources we have used a one-zone leptonic model. The emission region is assumed to be a spherical blob with a radius of R which is filled with uniformly tangled magnetic field

B and with a homogeneous population of relativistic electrons. Non-thermal energy distribution of electrons is described by a power-law with exponential cut-off:

$$N(\gamma) \backsim \gamma^{-\alpha} exp(-\gamma/\gamma_{cut})\ \gamma > \gamma_{min}, \tag{1}$$

where γ is the Lorentz factor of electrons in the blob rest frame and α is the power-law index which defines the properties of acceleration mechanism.[41–44] γ_{min} and γ_{cut} are the minimum and cut-off energy of the electron in the emission region correspondingly. γ_{cut} allows constraining the cooling process of the particles and the state of plasma in the jet.[29,45] In a simple one zone leptonic model emitting region parameters are the radius, Doppler factor(Γ=δ), and magnetic field. Magnetic field is a free parameter and its density is $B^2/4\pi$. In the modeling it is assumed that from radio to optical/UV emission is due to synchrotron emission of relativistic electrons, meanwhile high energy peak is due to the Inverse Compton scattering of both internal[2–4] or external[5–7] photons.

The previous studies show that high energy peak is well explained when assumed external photons for Inverse Compton scattering.[15,46–49] Depending on the position of emission region(blob) in the jet different photon field can be assumed.[50] In case of these sources it is assumed that emission region is outside of broad-line region(BLR) and in this position dominant photon field is the infrared emission from dusty torus.[51] IR radiation from dusty torus has a black body spectrum which luminosity is $L_{IR} = 0.6L_{disc}$,[46] where L_{disc} is the accretion disc luminosity. Directly emission from accretion disk showing a UV excess in the SED and fitting of this excess allows us to estimate photons temperature and luminosity.[52] SEDs modeling results are shown in Table 3 and SEDs of sources for which were generated adaptively binned light curves are shown in Figure 4. The archival data from the Space Science Data Center are in gray, while the optical/UV, X-ray, and γ-ray data, obtained here, are shown in cyan, red, and blue, respectively. The radio data are not included in the SEDs fits, but are considered only as upper limits. The observed radio emission is assumed to originate from a different and extended region. The hard γ-ray emission spectra of B3 1343+451, PKS 0451-28 are in magenta, showing that the γ-ray flux increased and their spectra extend to higher energies.

The SEDs in Fig. 5 contain enough data from radio to HE γ-ray bands to shape both low and high energy peaks. The electron power-law index is defined by the X-ray data and depending on whether the SSC or EIC component is dominating in the X-ray band, different values for α are obtained. When the X-ray spectrum is hard and the SSC component is dominating, the energy distribution of the emitting electrons has a hard spectrum as well. For example for PKS 0537-286 $\alpha = 1.33 \pm 0.07$. The HE tails of both synchrotron and IC components are well defined by the optical/UV and γ-ray data, respectively, allowing precise estimation of $\gamma_{cut} = (1.01 - 15.73) \times 10^3$.$\gamma_{cut}$ strongly depends on α and its highest values is $(15.73 \pm 1.60) \times 10^3$ which was estimated when $\alpha = 2.80 \pm 0.04$. Meanwhile, when $\alpha = 2.2 - 2.5$, the highest γ_{cut} is $(1.01 - 15.73) \times 10^3$ for B3 1343+451. The estimated Doppler factor is from $\delta = 10.00$ to $\delta = 27.42$ with a mean of $\delta = 19.09$ which is higher

Table 4. Parameters obtained from the modeling of multiwavelength SEDs. [1]: object name. [2]: Doppler factor. [3]: Slope of electron energy distribution. [4] and [5]: the Lorentz factors corresponding to the minimum and cutoff energy of the electron distribution. [6]: Magnetic field in units of G. [7] Radius of the emitting region in units of 10^{16} cm. [8] and [9]: Electron and magnetic field energy densities. [10]: accretion disk luminosity in units of 10^{46} erg s^{-1}. [11] and [12]: The power of the jet in the form of the relativistic electrons (L_e) and magnetic field (L_B) in units of 10^{45} erg s^{-1} and 10^{43} erg s^{-1}, respectively.

Sources	δ	α	$\gamma_{\rm min}$	$\gamma_{\rm cut}$	B	R	$U_{\rm e}$	$U_{\rm B}$	$L_{\rm d}$	$L_{\rm e}$	$L_{\rm B}$
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
GB 1508+5714	15.72 ± 1.29	1.17 ± 0.07	26.90 ± 2.88	1.30 ± 0.10	0.19 ± 0.02	2.28	0.50	1.49	3.02	2.43	0.73
PKS 1351-018	20.47 ± 2.49	2.16 ± 0.11	2.68 ± 0.36	4.99 ± 0.71	0.20 ± 0.02	2.29	0.54	1.66	4.04	2.68	0.82
PKS 0537-286	11.50 ± 0.57	1.33 ± 0.07	15.70 ± 1.49	2.45 ± 0.16	0.28 ± 0.02	1.14	5.93	3.21	3.44	7.31	0.40
TXS 0800+618	14.04 ± 0.56	2.75 ± 0.04	13.98 ± 0.86	2.20 ± 0.02	0.26 ± 0.01	15.5	0.06	2.73	1.65	13.49	61.84
S4 1427+543	10.00 ± 0.37	2.04 ± 0.10	29.00 ± 2.55	2.79 ± 0.27	0.53 ± 0.04	1.90	0.63	10.02	1.83	2.14	3.75
GB6 J0733+0456	16.28 ± 1.36	2.80 ± 0.04	47.90 ± 3.42	15.73 ± 1.60	0.16 ± 0.03	2.98	0.12	6.22	3.40	0.98	5.20
PKS 0347-211	26.00 ± 1.02	2.79 ± 0.02	23.09 ± 1.12	2.62 ± 0.16	0.20 ± 0.01	8.15	0.03	1.61	1.99	1.72	10.08
B2 0743+25	10.02 ± 0.45	1.13 ± 0.19	7.66 ± 0.22	2.03 ± 0.08	0.36 ± 0.01	0.70	16.90	0.003	3.58	7.80	0.24
S4 1124+57	22.17 ± 1.37	2.78 ± 0.04	22.92 ± 1.45	1.28 ± 0.10	0.22 ± 0.01	5.15	0.14	1.95	1.69	3.39	4.87
PKS 0438-43	18.17 ± 1.29	2.78 ± 0.04	23.13 ± 1.54	7.19 ± 0.58	0.34 ± 0.02	5.42	0.12	4.52	3.91	3.33	12.52
S4 2015+65	17.85 ± 1.32	2.73 ± 0.05	20.63 ± 1.64	2.75 ± 0.29	0.46 ± 0.03	13.50	0.07	8.41	4.55	1.26	14.44
MG2 J174803+3403	24.50 ± 2.06	2.87 ± 0.06	14.67 ± 1.58	1.40 ± 0.49	1.45 ± 0.09	7.41	0.004	83.98	8.87	0.23	434.69
PKS 0834-20	27.42 ± 0.97	2.70 ± 0.06	20.57 ± 1.51	2.13 ± 0.18	0.37 ± 0.02	6.83	0.02	5.44	5.51	0.70	23.92
TXS 0222+185	10.03 ± 0.28	1.62 ± 0.05	19.56 ± 0.98	2.38 ± 0.10	0.35 ± 0.02	1.07	10.28	5.04	2.71	11.10	0.54
OD 166	19.02 ± 0.84	1.96 ± 0.04	2.58 ± 0.15	1.01 ± 0.03	1.15 ± 0.05	4.32	0.01	0.52	0.53	0.25	92.38
TXS 0907+230	21.66 ± 1.66	2.23 ± 0.11	20.26 ± 1.84	1.44 ± 0.10	0.31 ± 0.02	5.23	0.03	3.72	1.09	0.70	5.96
PMN J1441-1523	17.01 ± 1.50	2.19 ± 0.07	2.86 ± 0.29	3.31 ± 0.46	1.68 ± 0.14	1.47	0.07	112.38	0.17	0.14	22.85
TXS 1448+093	17.90 ± 1.13	1.52 ± 0.15	49.64 ± 5.82	0.84 ± 0.06	0.70 ± 0.05	7.96	0.003	19.44	0.56	0.17	115.93
PMN J0226+0937	25.02 ± 1.98	2.41 ± 0.04	5.37 ± 0.59	8.32 ± 0.62	1.74 ± 0.12	3.09	0.03	119.86	10.94	0.03	107.98
PKS 0451-28	26.14 ± 1.27	2.90 ± 0.28	21.93 ± 1.3‘	2.19 ± 0.01	0.45 ± 0.03	5.90	0.11	8.01	7.20	3.59	26.32
B3 0908+416B	23.22 ± 1.72	1.31 ± 0.25	6.41 ± 0.72	1.11 ± 0.10	0.39 ± 0.03	1.76	0.09	6.10	0.43	0.26	1.78
TXS 1616+517	10.11 ± 0.31	2.09 ± 0.09	93.28 ± 4.15	4.34 ± 0.36	0.52 ± 0.02	3.59	0.06	10.79	0.35	0.70	13.11
B3 1343+451	26.55 ± 1.04	2.48 ± 0.04	16.49 ± 1.31	8.67 ± 0.48	0.10 ± 0.01	4.16	0.11	0.42	0.48	1.76	0.68
PKS 2107-105	27.32 ± 1.34	2.30 ± 0.06	7.45 ± 0.90	3.63 ± 0.39	0.67 ± 0.05	9.48	0.0006	17.73	8.30	0.05	150.16

Note: [1] Integrated γ-ray flux in units of $\times 10^{-8}$photon cm^{-2} s^{-1}.
Note: [2] Photon energy in GeV.

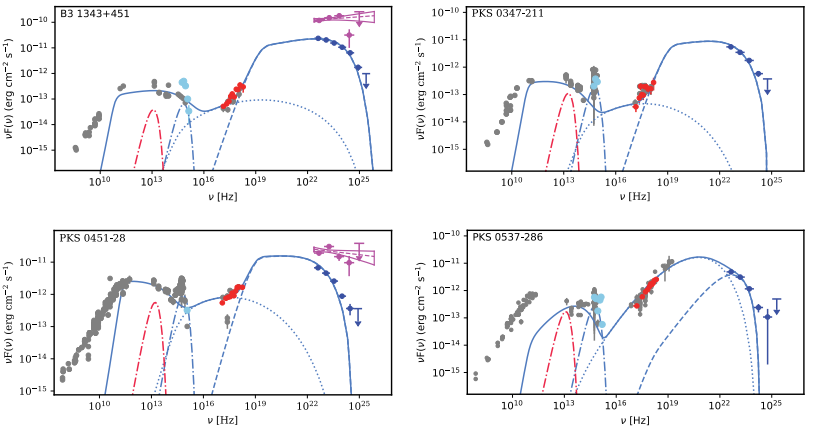


Fig. 5. The evolution of the γ-ray photon indexes of B3 1343+451, PKS 0347-211, PKS 0451-28, PKS 0537-286 in time.

than the average values estimated for FSRQs.[53,54] The emitting region size is within $R = (0.70-9.48)\times10^{16}$ cm. Results obtained for R suggest that the multiwavelength emission is produced in the sub-parsec scale regions of the jet. Modeling results allow us to estimate energetics of the considered sources. For seventeen sources out of thirty three was possible to estimate discs luminosity assuming that disc has a black body spectrum. Estimations shows that $L_d \simeq (1.09 - 10.94) \times 10^{46}$ erg s^{-1} and highest value estimated for PMN J0226+0937 is $L_d \simeq 1.09 \times 10^{47}$ erg s^{-1}. Jet power was calculated as

$$L = \pi R^2 c \Gamma^2 U_i \tag{2}$$

where U_i is electron or magnetic field energy density.Corresponding values from Table 4 show that $L_e \simeq (0.03 - 13.49) \times 10^{45}$ erg s^{-1} and $L_B \simeq (0.24 - 434.69) \times 10^{44}$ erg s^{-1}.

6. Summary

The main results after analysis of Fermi-LAT data, accumulated in 2008-2018 and Swift data, observed in fifteen years are summarized. The γ-ray photon index of all the considered sources is between 2.18-3.05 except for two BL Lacs. Swift XRT has detected significant X-ray emission only from FSRQs considered here and X-ray flux ranges from 5×10^{-14} erg cm^{-2} s^{-1} to 10^{-11} erg cm^{-2} s^{-1}. The γ-ray variability of the considered sources has shown emission on short and long time scales from sub day to month scales. During γ-ray flares the luminosity riches $(10^{49} - 10^{50})$ erg s^{-1}. The most luminous source in the sample is B3 1343+451 which luminosity during flare increases 36.4 times and corresponds to $L_\gamma = 1.5 \times 10^{50}$ erg s^{-1}. For SED modeling one zone leptonic model considering SSC of synchrotron and IC of dusty

torus photons was used. Magnetic field ranges from $0.10-1.74$ G and Doppler factor is within $10.00-27.42$. Estimated black hole masses are within $(1.69-5.35)\times10^9 M_\odot$.

References

1. C. M. Urry and P. Padovani, Unified Schemes for Radio-Loud Active Galactic Nuclei, *Publications of the ASP* **107**, p. 803 (September 1995).
2. G. Ghisellini, et al., Inhomogeneous synchrotron-self-Compton models and the problem of relativistic beaming of BL Lac objects, *Astronomy and Astrophysics* **146**, 204 (May 1985).
3. S. D. Bloom and A. P. Marscher, An Analysis of the Synchrotron Self-Compton Model for the Multi–Wave Band Spectra of Blazars, *The Astrophysical Journal* **461**, p. 657 (April 1996).
4. L. Maraschi, et al., A jet model for the gamma-ray emitting blazar 3C 279, *Astrophysical Journal, Letters* **397**, L5 (September 1992).
5. M. Błażejowski, et al., Comptonization of Infrared Radiation from Hot Dust by Relativistic Jets in Quasars, *The Astrophysical Journal* **545**, 107 (December 2000).
6. M. Sikora, et al., Comptonization of diffuse ambient radiation by a relativistic jet: The source of gamma rays from blazars?, *The Astrophysical Journal* **421**, 153 (January 1994).
7. G. Ghisellini and F. Tavecchio, Canonical high-power blazars, *Monthly Notices of the RAS* **397**, 985 (August 2009).
8. S. Abdollahi, et al., Fermi Large Area Telescope Fourth Source Catalog, *Astrophysical Journal, Supplement* **247**, p. 33 (March 2020).
9. M. Volonteri, Formation of supermassive black holes, **18**, 279 (July 2010).
10. J. R. Mattox, et al., The Likelihood Analysis of EGRET Data, *The Astrophysical Journal* **461**, p. 396 (April 1996).
11. B. Rani, et al., Constraining the location of rapid gamma-ray flares in the flat spectrum radio quasar 3C 273, *Astronomy and Astrophysics* **557**, p. A71 (Sep 2013).
12. N. Sahakyan and S. Gasparyan, High energy gamma-ray emission from PKS 1441+25, *Monthly Notices of the RAS* **470**, 2861 (September 2017).
13. N. Sahakyan, et al., Fermi-LAT observation of nonblazar AGNs, *Astronomy and Astrophysics* **614**, p. A6 (June 2018).
14. V. Baghmanyan, et al., Rapid Gamma-Ray Variability of NGC 1275, *The Astrophysical Journal* **848**, p. 111 (October 2017).
15. S. Gasparyan, et al., On the Multiwavelength Emission from CTA 102, *The Astrophysical Journal* **863**, p. 114 (August 2018).
16. M. S. Dutka, et al., Multiband Observations of the Quasar PKS 2326-502 during Active and Quiescent Gamma-Ray States in 2010-2012, *The Astrophysical Journal* **835**, p. 182 (February 2017).
17. D. Zargaryan, et al., Comparing 3C 120 jet emission at small and large scales, *Astronomy and Astrophysics* **608**, p. A37 (December 2017).
18. D. N. Burrows, et al., The Swift X-Ray Telescope, **120**, 165 (October 2005).
19. P. W. A. Roming, et al., The Swift Ultra-Violet/Optical Telescope, **120**, 95 (October 2005).
20. N. Gehrels, et al., The Swift Gamma-Ray Burst Mission, *The Astrophysical Journal* **611**, 1005 (August 2004).
21. T. S. Poole, et al., Photometric calibration of the Swift ultraviolet/optical telescope, *Monthly Notices of the RAS* **383**, 627 (January 2008).

22. P. A. Evans, et al., Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs, *Monthly Notices of the RAS* **397**, 1177 (August 2009).
23. M. Ajello, et al., The Fourth Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope, *The Astrophysical Journal* **892**, p. 105 (April 2020).
24. S. Li, et al., Fast γ-Ray Variability in Blazars beyond Redshift 3, *The Astrophysical Journal* **853**, p. 159 (February 2018).
25. C. C. Cheung, Fermi LAT detection of a GeV flare from High-redshift Blazar PKS 0537-286, *The Astronomer's Telegram* **10356**, p. 1 (May 2017).
26. A. A. Abdo, et al., Gamma-ray Light Curves and Variability of Bright Fermi-detected Blazars, *The Astrophysical Journal* **722**, 520 (October 2010).
27. V. S. Paliya, et al., Fermi-LAT detection of high gamma-ray activity from the z=3.6 quasar NVSS J163547+362930, *The Astronomer's Telegram* **11847**, p. 1 (July 2018).
28. L. Pacciani, et al., Exploring the Blazar Zone in High-energy Flares of FSRQs, *The Astrophysical Journal* **790**, p. 45 (August 2014).
29. N. Sahakyan, Investigation of the γ-ray spectrum of CTA 102 during the exceptional flaring state in 2016-2017, *Astronomy and Astrophysics* **635**, p. A25 (March 2020).
30. V. S. Paliya, et al., Detection of a Gamma-Ray Flare from the High-redshift Blazar DA 193, *The Astrophysical Journal* **871**, p. 211 (February 2019).
31. R. Angioni, et al., The large gamma-ray flare of the flat-spectrum radio quasar PKS 0346-27, *Astronomy and Astrophysics* **627**, p. A140 (July 2019).
32. A. Dar and A. Laor, Hadronic Production of TeV Gamma-Ray Flares from Blazars, *Astrophysical Journal, Letters* **478**, L5 (March 1997).
33. J. H. Beall and W. Bednarek, On the Hadronic Beam Model for Gamma-Ray Production in Blazars, *The Astrophysical Journal* **510**, 188 (January 1999).
34. W. Bednarek and R. J. Protheroe, Gamma-rays from interactions of stars with active galactic nucleus jets, *Monthly Notices of the RAS* **287**, L9 (May 1997).
35. K. Mannheim, High-energy neutrinos from extragalactic jets, *Astroparticle Physics* **3**, 295 (May 1995).
36. K. Mannheim and P. L. Biermann, Photomeson production in active galactic nuclei, *Astronomy and Astrophysics* **221**, 211 (September 1989).
37. K. Mannheim, The proton blazar, *Astronomy and Astrophysics* **269**, 67 (March 1993).
38. A. Mücke and R. J. Protheroe, A proton synchrotron blazar model for flaring in Markarian 501, *Astroparticle Physics* **15**, 121 (March 2001).
39. A. Mücke, et al., BL Lac objects in the synchrotron proton blazar model, *Astroparticle Physics* **18**, 593 (March 2003).
40. A. Paggi, et al., Flaring Patterns in Blazars, *The Astrophysical Journal* **736**, p. 128 (August 2011).
41. J. G. Kirk and A. F. Heavens, Particle acceleration at oblique shock fronts, *Monthly Notices of the RAS* **239**, 995 (Aug 1989).
42. D. C. Ellison, et al., First-Order Fermi Particle Acceleration by Relativistic Shocks, *The Astrophysical Journal* **360**, p. 702 (Sep 1990).
43. E. J. Summerlin and M. G. Baring, Diffusive Acceleration of Particles at Oblique, Relativistic, Magnetohydrodynamic Shocks, *The Astrophysical Journal* **745**, p. 63 (Jan 2012).
44. L. O. Drury, REVIEW ARTICLE: An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas, *Reports on Progress in Physics* **46**, 973 (Aug 1983).
45. M. G. Baring, et al., Probing acceleration and turbulence at relativistic shocks in blazar jets, *Monthly Notices of the RAS* **464**, 4875 (Feb 2017).

46. M. L. Ahnen, et al., Very High Energy γ-Rays from the Universe's Middle Age: Detection of the z = 0.940 Blazar PKS 1441+25 with MAGIC, *Astrophysical Journal, Letters* **815**, p. L23 (December 2015).
47. A. U. Abeysekara, et al., Gamma-Rays from the Quasar PKS 1441+25: Story of an Escape, *Astrophysical Journal, Letters* **815**, p. L22 (December 2015).
48. MAGIC Collaboration, et al., Detection of persistent VHE gamma-ray emission from PKS 1510-089 by the MAGIC telescopes during low states between 2012 and 2017, *Astronomy and Astrophysics* **619**, p. A159 (November 2018).
49. M. Hayashida, et al., Rapid Variability of Blazar 3C 279 during Flaring States in 2013-2014 with Joint Fermi-LAT, NuSTAR, Swift, and Ground-Based Multiwavelength Observations, *The Astrophysical Journal* **807**, p. 79 (July 2015).
50. M. Sikora, et al., Constraining Emission Models of Luminous Blazar Sources, *The Astrophysical Journal* **704**, 38 (October 2009).
51. M. Sikora, et al., On the Nature of MeV Blazars, *The Astrophysical Journal* **577**, 78 (September 2002).
52. G. Ghisellini, et al., Jet and accretion power in the most powerful Fermi blazars, *Monthly Notices of the RAS* **399**, 2041 (November 2009).
53. G. Ghisellini and F. Tavecchio, Fermi/LAT broad emission line blazars, *Monthly Notices of the RAS* **448**, 1060 (April 2015).
54. V. S. Paliya, et al., General Physical Properties of CGRaBS Blazars, *The Astrophysical Journal* **851**, p. 33 (December 2017).

Quantum gravity phenomenology from thermodynamics of spacetime

A. Alonso-Serrano

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut)
Am Mühlenberg 1, 14476 Potsdam, Germany
E-mail: ana.alonso.serrano@aei.mpg.de

M. Liška

Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University
V Holešovičkách 2, 180 00 Prague 8, Czech Republic,
Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut)
Am Mühlenberg 1, Potsdam, Germany
E-mail: liska.mk@seznam.cz

In this review, we discuss a derivation of effective low energy quantum gravitational dynamics from thermodynamics. The derivation is based on the formalism developed in semiclassical thermodynamics of spacetime that allows to obtain Einstein equations from the proportionality of entropy to the area. We first introduce the relevant ingredients of semiclassical thermodynamics of spacetime, paying special attention to the various concepts of entropy involved and their relations. We then extend the semiclassical formalism by considering low energy quantum gravity effects which imply a modified entropy formula with an additional term logarithmic in the area. Upon discussing the derivation of effective gravitational dynamics from this modified entropy, we comment on the most important features of our proposal. Moreover, we show its physical implications on a simple cosmological model and show that it suggests the replacement of the Big Bang singularity by a regular bounce.

Keywords: Quantum gravity phenomenology, Thermodynamics of spacetime

1. Introduction

The search for a consistent theory unifying gravity and quantum physics has been an important direction of research in the last decades. While promising candidates have been put forward, none of them presently provides a complete and consistent final theory. In the absence of a fully developed theoretical framework, quantum gravitational phenomenology offers a way to gain information about possible low energy dynamical effects of quantum gravity.[1–5] Phenomenological models concerning strong gravitational fields are mainly available for physics in the vicinity of classical singularities.[6–10] However, these models face limitations coming from the simplicity of the studied geometries and their results cannot be directly applied in more general settings. Motivated by overcoming these issues, the authors have employed thermodynamics of spacetime to study quantum gravitational phenomenology and propose effective equations of motion applicable in generic spacetimes.[11] As we will see, thermodynamic methods allow us to look for model-independent

low energy phenomenological dynamics of quantum gravity without choosing any specific background spacetime. One can then particularise the resulting dynamics to any case of interest. Furthermore, the relevant thermodynamic predictions are common for most of the candidate theories of quantum gravity and can even be obtained by model independent considerations.

Thermodynamics of spacetime presents a useful tool for understanding gravitational dynamics. Following the seminal developments of black hole thermodynamics,[12–14] the framework has been extended both to more general spacetimes[15] and beyond general relativity.[16] A key step forward has been the derivation of Einstein equations from thermodynamics of local Rindler horizons.[17]The original approach has since been improved and generalised to work for different constructions of local horizons,[18–22] certain modified theories of gravity[20–26] and even to derive semiclassical gravitational equations of motion.[19,21,22]

Taking a step further, the authors have included quantum gravity effects on the thermodynamics of local causal horizons, obtaining equations governing low energy effective dynamics of quantum gravity.[11] To get sufficiently robust and general results, we concentrated on the leading order quantum correction to Bekenstein entropy, a term logarithmic in the horizon area. This is advantageous, since its presence is predicted by many different approaches to quantum gravity, including loop quantum gravity (LQG),[27,28] string theory[29,30] and AdS/CFT correspondence,[31] non-local effective field theory,[32] entanglement entropy calculations,[33] as well as model independent considerations based on generalised uncertainty principle (GUP) phenomenology[7] and analysis of statistical fluctuations.[34,35] Thus, the effective dynamics we propose is relevant to most of the main approaches to quantum gravity.

While we are chiefly concerned with low energy quantum gravity effects, thermodynamics of spacetime provides interesting insights already on the semiclassical level. Furthermore, a complete understanding of the semiclassical regime is necessary to extend the thermodynamic formalism to the realm of quantum gravity. Therefore, we include a detailed review of the concepts considered in semiclassical thermodynamics of spacetime, especially various notions of entropy. Since derivations of gravitational dynamics rely on equilibrium conditions for local causal horizons which involve Bekenstein, entanglement and Clausius entropy, we further discuss to what extend are these entropies equivalent.

In the present work, we review the main features of our proposal, paying special attention to the various concepts of entropy involved as well as to the unimodular nature of the resulting gravitational dynamics.[36] Section 2 introduces the necessary entropy definitions and discusses their relations both in the semiclassical and low energy quantum gravity settings. In section 3, we first briefly recap derivation of the effective equations. Then we discuss their properties and illustrate their physical implications on a simple cosmological toy model (finding that the Big Bang singularity can be resolved). Lastly, section 4 sums up our results and presents possible directions for future research.

Throughout the paper, we work in four spacetime dimensions and use metric signature $(-,+,+,+)$. Definitions of the curvature-related quantities follow.[37] We use lower case Greek letters to denote abstract spacetime indices and lower case Latin letters for spatial indices with respect to a (local) Cartesian basis. Unless otherwise explicitly stated, we use the SI units.

2. Entropy in thermodynamics of spacetime

Thermodynamic derivations of gravitational dynamics are based on the observation that gravitational dynamic are encoded in the horizon area equilibrium condition for maximal entropy, $\delta S = 0$, if it holds for all local causal horizons. This condition sums together variations of entropy of the horizon, often interpreted in terms of entanglement (von Neumann) entropy due to quantum correlations across the horizon,[17,19,26] and of entropy of the matter. Since the latter is usually described in terms of thermodynamic Clausius entropy,[17,18] it is not obvious that one can combine both entropies to define a meaningful equilibrium condition (it has actually been suggested that this fails in the case of local Rindler wedges[38]). This combination requires that fluxes of Clausius and von Neumann entropy of matter across the horizon are equal with sufficient precission. In the following, we first introduce our implementation of local causal horizons: geodesic local causal diamonds (GLCD). Then we present the definitions of all the relevant entropies associated with GLCD's and review our previous argument for the interchangeability of Clausius and von Neumann entropy.[36]

2.1. *Geodesic local causal diamonds*

We begin by briefly introducing the construction and most important properties of GLCD's. More detailed description of GLCD's can be found, e.g. in.[39–42]

In an arbitrary spacetime point P choose any unit timelike vector $n(P)$. In every direction orthogonal to n send out of P geodesics of parameter length l. These form a spatial geodesic 3-ball, Σ_0, and the region causally determined by Σ_0 constitutes a GLCD (see figure 1). The boundary, $\mathcal{B}$, of Σ_0 is approximately a 2-sphere. Its area reads[19]

$$\mathcal{A} = 4\pi l^2 - \frac{4\pi}{9} l^4 G_{00}(P) + O\left(l^5\right), \tag{1}$$

where $G_{00} = G_{\mu\nu} n^\mu n^\nu$. The GLCD possesses an approximate (up to $O\left(l^3\right)$ curvature dependent terms) conformal Killing vector[19]

$$\zeta = C\left(\left(l^2 - t^2 - r^2\right)\frac{\partial}{\partial t} - 2rt\frac{\partial}{\partial r}\right), \tag{2}$$

where C denotes an arbitrary normalisation constant. It has been argued that one can assign Hawking temperature to the GLCD's conformal horizon, $T_H = \hbar\kappa/2\pi k_B c$, where $\kappa = C/2l$ is the surface gravity corresponding to ζ.[19,42]

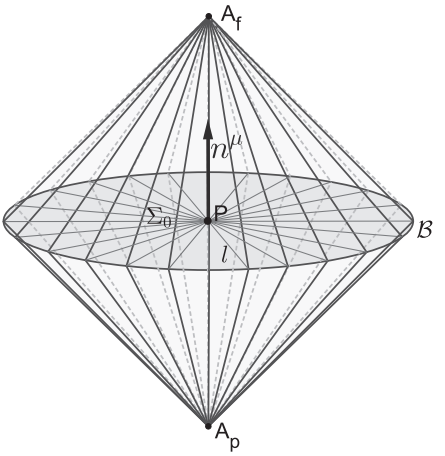


Fig. 1. A sketch of a GLCD with the origin in point P (the angular coordinate θ is suppressed). Σ_0 is a spatial geodesic ball of radius l (several of the geodesics forming it are depicted as grey lines), its boundary an approximate 2-sphere $\mathcal{B}$. Unit timelike vector n^μ is a normal of Σ_0. The tilted lines from the past apex A_p ($t = -l/c$) to the future apex A_f ($t = l/c$) represent the null geodesic generators of the GLCD boundary. The diamond's base Σ_0 is the spatial cross-section of both the future domain of dependence of A_p and the past domain of dependence of A_f at $t = 0$.

2.2. *Bekenstein entropy*

To derive the Einstein equations from thermodynamics of GLCD's (or, conversely, to interpret the first law of GLCD dynamics implied by Einstein equations in thermodynamic terms[42]), one must assume that entropy associated with its horizon obeys the Bekenstein formula[19,22]

$$S_{BH} = \frac{k_B \mathcal{A}}{4 l_P^2}, \tag{3}$$

where $\mathcal{A}$ denotes the horizon area. However, it is far from clear to what types of horizons can one assign Bekenstein entropy. Since its interpretation in terms of quantum entanglement allows a natural extension of the Bekenstein formula to any causal horizon, it is often assumed in the context of thermodynamics of spacetime (although it was originally proposed in a different context[43,44]). To understand the entanglement interpretation of Bekenstein entropy, consider two causally separated spacetime regions. An observer in one region cannot measure vacuum fluctuations in the other one. Since the fluctuations are correlated between regions, some information is inaccessible to the observer and non-zero entanglement entropy appears. Its value is proportional to the area of causal horizon, $S_e = \eta \mathcal{A}$.[33,43,44] The proportionality constant η is infinite unless one introduces a cutoff. Then η becomes finite and depends on the cutoff length, the number and type of quantum fields considered in the theory and even on the position in spacetime.[26,33] To recover the Bekenstein entropy, a universal value $\eta = k_B/4l_P^2$ is necessary (see, e.g.[33,45] for discussions of feasibility and shortcomings of this assumption). Let us stress that

the entanglement interpretation of Bekenstein entropy is in no way necessary to derive gravitational dynamics from thermodynamics. Any microscopic interpretation of black hole entropy which also applies to observer-dependent causal horizons would work just as well. Nevertheless, here we focus on the entanglement entropy as it is currently the most developed proposal with this property.

Beyond this semiclassical picture, many different models indicate that the leading order quantum gravity correction to Bekenstein entropy is a non-local term logarithmic in horizon area[7, 9, 27–35, 46–49]

$$S = \frac{k_B \mathcal{A}}{4 l_P^2} + k_B \mathcal{C} \ln\left(\frac{\mathcal{A}}{\mathcal{A}_0}\right), \tag{4}$$

where $\mathcal{C}$ is a real number and $\mathcal{A}_0$ a constant with dimensions of area. The values of $\mathcal{C}$ and $\mathcal{A}_0$ are model dependent. Considering the entanglement entropy interpretation allows us to calculate modified entropy even for causal horizons in a flat spacetime.[33] Interestingly, a logarithmic term then appears for spherical horizons,[50] but not for planar ones[33] (due to different Euler characteristics). This is the reason we consider GLCD's rather than local Rindler wedges to find effective low energy dynamics of quantum gravity from thermodynamics[11] (for arguments against the Rindler wedges even in the semiclassical setting see, e.g.[20, 38]).

While the presence of a logarithmic term seems to be a very general prediction of quantum gravity, the value and even the sign of $\mathcal{C}$ differ in various approaches. Generally speaking, two types of corrections to Bekenstein entropy have been studied in the literature. On one hand, microcanonical corrections appear due to more precise counting of the microstates at fixed horizon area, which reduces our uncertainty, leading to negative contribution to entropy. On the other hand, canonical corrections stem from thermal fluctuations at fixed temperature, which are an additional source of uncertainty increasing the entropy. As we discuss in section 3, the sign of $\mathcal{C}$ determines the physical implications of the modified dynamics we propose, allowing the avoidance of spacetime singularities only for $\mathcal{C} > 0$. Therefore, the presently unknown overall sign of $\mathcal{C}$ is crucial for the interpretation of our results.

2.3. *Clausius entropy of the matter*

Apart from Bekenstein entropy, derivation of gravitational dynamics from thermodynamics also requires a way to account for the entropy of matter. The simplest approach is to consider the thermodynamic Clausius entropy, $\mathrm{d}S_C = \delta Q/T$. The heat flux δQ across an arbitrary timelike hypersurface Σ reads[18]

$$\delta Q = -\int_\Sigma T_{\mu\nu} V^\nu N^\mu \mathrm{d}^3\Sigma, \tag{5}$$

where V^ν and N^μ are timelike and spacelike unit normals of Σ. If V^ν corresponds to velocity of an eternal, uniformly accelerating observer with acceleration a, we can define the corresponding Unruh temperature, $T_U = \hbar a/2\pi k_B c$. This definition holds with sufficient precission even for observers with slowly varying acceleration

or finite lifetime, as long as the proper length λ of the trajectory with an approximately constant acceleration is large enough, i.e. $\lambda \gg c^2/a$.[51] For the special case of uniformly accelerated observers travelling inside causal diamonds one also finds an equivalent result due to conformal mapping of a Rindler wedge to a causal diamond.[52,53] Therefore, as long as we consider a sufficiently large a, we can use the standard Unruh temperature together with the expression for heat flux to define Clausius entropy flux across a timelike hypersurface

$$S_C \equiv \frac{\delta Q}{T_U} = -\int_\Sigma \frac{2\pi k_B c}{\hbar a} T_{\mu\nu} V^\nu N^\mu \mathrm{d}^3\Sigma. \tag{6}$$

The timelike surface Σ coincides with the causal horizon perceived by the uniformly accelerating observer in the limit $a \to \infty$. Notably, this limit is well defined both for arbitrary causal horizons in flat spacetime and for sufficiently small horizons in curved spacetimes[18] (in that case, one simply considers Riemann normal coordinate expansion to the leading order). Before moving on to the special case of causal diamonds, several further remarks are in order. First, the definition of Clausius entropy flux across null surfaces is completely independent of gravitational dynamics or any requirements on symmetries of the spacetime. Second, the construction of the entropy is semiclassical as it explicitly depends on quantum field theory by invoking the Unruh effect. And, lastly, in contrast with the nature of Clausius entropy in non-relativistic thermodynamics, the entropy flux is manifestly observer dependent.[18]

For a GLCD, the resulting expression for Clausius entropy flux from $t = 0$ (geodesic ball Σ_0) to $t = l/c$ (future apex A_f) takes form[36]

$$\Delta S_C = \int_0^{l/c} \frac{2\pi k_B c}{\hbar} t \left(\int_{\mathcal{S}(t)} T_{\mu\nu} k^\mu k^\nu \mathrm{d}^2\mathcal{A} \right) \mathrm{d}t + O\left(l^5\right), \tag{7}$$

where k^μ denotes the future pointing null normal to the GLCD's boundary. Performing the integration and some straightforward manipulations then yields

$$\Delta S_C = -\frac{8\pi^2 k_B l^4}{9\hbar c} \left(T_{00}(P) + \frac{1}{4} T(P) \right) + O\left(l^5\right), \tag{8}$$

where we explicitly stress that the energy momentum tensor is evaluated at the diamond's origin, P.

When one wishes to take into account low energy quantum gravity effects, defining the Clausius entropy flux becomes more complicated. The subtleties in its definition come from the need to consider the Unruh effect, which requires that the ground state of quantum fields is locally well approximated by Minkowski vacuum. This amounts to invoking Einstein equivalence principle:[26] "Fundamental non-gravitational test physics is not affected, locally and at any point of spacetime, by the presence of a gravitational field".[54] However, the status of the equivalence principle in the low energy quantum gravity regime is far from clear. For instance, possible violations of weak equivalence principle resulting from GUP phenomenology have been explored in a number of works with very different conclusions.[55–58]

A simple way to deal with possible influence of the equivalence principle violations on the Unruh effect is to consider heuristic modifications to Unruh temperature due to GUP, previously suggested in the literature.[59,60] All the proposals for modified Unruh temperature obey

$$T_{mod} = \frac{\hbar a \left(1 + \psi \frac{l_P^2}{c^4} a^2\right)}{2\pi k_B c} + O\left(\frac{l_P^4 a^5}{c^8}\right), \tag{9}$$

where ψ denotes a model dependent real number that is expected to be of the order of unity.[59,60] A similar prescription has also been earlier proposed for modified Hawking temperature.[7,9,33] To our best knowledge, these modifications to temperature have not been confirmed by any rigorous method. However, since GUP phenomenologically implements effects of a minimal resolvable length that arises in a number of approaches to quantum gravity,[1,2] the idea of modified Unruh temperature is worth considering. Hence, to incorporate leading order quantum gravity corrections in thermodynamics of spacetime, we need to show that the previously outlined construction of Clausius entropy flux works with this modified temperature. Note that the previous reasoning does not apply directly, since in the limit $a \to \infty$ correction terms in the modified temperature formula become dominant. Instead, one must consider acceleration much larger than c^2/l but much smaller than $c^2/\sqrt{\psi} l_P$ (this can be satisfied for $l \gg l_P$, i.e., whenever the GLCD is much larger than the Planck scale). The construction of Clausius entropy flux under this assumption has been carried out by the authors,[11] yielding

$$\Delta S_{C,mod} = -\left(1 - \psi \frac{l_P^2 a^2}{c^4}\right) \frac{2\pi k_B c}{\hbar} \int_0^{l/c} t \left(\int_{\mathcal{S}(t)} T_{\mu\nu} k^\mu k^\nu \mathrm{d}^2\mathcal{A} \right) \mathrm{d}t + O\left(l^5, \frac{l_P^4 a^4}{c^8}, \frac{1}{a^2}\right). \tag{10}$$

The only difference with respect to the semiclassical formula are the a-dependent sub-leading terms. We can recover it by setting $\psi = 0$ and then taking the limit $a \to \infty$. Let us note that while we assumed finite acceleration simply for mathematical convenience, the limit $a \lesssim c^2/\sqrt{\psi} l_P$ agrees quite well with the proposal of an upper limit to acceleration a_M set by quantum gravity effects, $a_M = c^2/l_P$.[61]

2.4. *Von Neumann entropy of the matter*

To be consistent with the entanglement interpretation of the entropy of GLCD's horizon, one should consider quantum von Neumann entropy of the matter rather than the thermodynamic Clausius one. For small variations from vacuum, this entropy can be obtained from the vacuum state density operator, $\rho = e^{-K/k_B T_\zeta}/Z$, where Z is the partition function, T_ζ denotes the previously defined temperature associated with conformal Killing vector ζ^μ and K is known as the modular Hamiltonian.[42] While variations of K can in general be non-local, for conformal fields they correspond to variations of the local matter Hamiltonian[42]

$$\delta H = \int_{\Sigma_0} \delta \langle T_{\mu\nu} \rangle \xi^\mu n^\nu \mathrm{d}^2\Omega \mathrm{d}r, \tag{11}$$

where $n = \partial/\partial t$. Then, the variation of entropy reads

$$\delta S_m = \frac{1}{T}\delta K = \frac{2\pi k_B}{\hbar c}\frac{4\pi l^4}{15}\delta\langle T_{00}\rangle + O\left(l^5\right). \tag{12}$$

For non-conformal field theories with a UV fixed point, the von Neumann entropy variation is modified only by the presence of an additional l-dependent spacetime scalar, δX,[19,62,63]

$$\delta S_m = \frac{2\pi k_B}{\hbar c}\frac{4\pi l^4}{15}\left(\delta\langle T_{00}\rangle + \delta X\right). \tag{13}$$

In our study concerning the quantum gravity corrections, we have gone beyond the semiclassical case just by including modifications to the temperature[11] (in a future work we plan to analyse possible corrections to the modular Hamiltonian). Implementing the leading order quantum gravity modification to temperature gives rise to the following modified variation of matter von Neumann entropy

$$\delta S_{m,mod} = \frac{2\pi k_B c}{\hbar}\frac{4\pi}{15}l^4\left(\delta\langle T_{00}\rangle + \delta X\right)\left(1 - \psi l_P^2\kappa^2\right) + O\left(l^5\right). \tag{14}$$

Note that for the particular case $\psi = 0$ we straightforwardly recover the semiclassical expression.

2.5. *Entropy equivalence*

Assuming the entanglement interpretation of Bekenstein entropy, the natural equilibrium condition for the GLCD involves von Neumann entropy of the matter. To consider the Clausius entropy instead, its equivalence with the von Neumann entropy is necessary. This condition requires that fluxes of Clausius and von Neumann entropy of matter across the horizon are equal with sufficient precission. Since the von Neumann entropy is defined for the geodesic ball Σ_0 corresponding to $t = 0$, the meaningful quantities to compare are time derivatives of both entropies evaluated at $t \ll l/c$ (at precisely $t = 0$ the entropy fluxes vanish). We start by considering conformal fields. For the time derivative Clausius entropy we have[36]

$$\frac{\mathrm{d}S_C}{\mathrm{d}t}(t) \approx -\frac{32\pi^2 k_B c}{3\hbar}tl^2 T_{00}, \tag{15}$$

while the time derivative of the von Neumann entropy reads[36]

$$\frac{\mathrm{d}S_m}{\mathrm{d}t} \approx -\frac{32\pi^2 k_B c}{3\hbar}tl^2\delta\langle T_{00}\rangle. \tag{16}$$

Both expressions are indeed equivalent, making the use of Clausius entropy in thermodynamics of spacetime justified for conformal fields. Note that the argument in no way involves gravitational dynamics.

The situation becomes more complicated for non-conformal matter as the previously outlined approach cannot be directly applied to formulas for the von Neumann

entropy.[62,63] Nevertheless, semiclassical gravitational dynamics derived from Clausius and von Neumann entropy are equivalent even in this case,[17,19,36] suggesting that both entropies are at least strongly related.

When low energy quantum gravity effects on the temperature are taken into account, we find for conformal fields

$$\frac{\mathrm{d}S_C}{\mathrm{d}t}(t) \approx -\frac{32\pi^2 k_B c}{3\hbar}\left(1-\psi\frac{l_P^2}{c^4}a^2\right) tl^2 T_{00}, \tag{17}$$

and

$$\frac{\mathrm{d}S_m}{\mathrm{d}t} \approx -\frac{32\pi^2 k_B c}{3\hbar}\left(1-\psi l_P^2\kappa^2\right) tl^2 \delta\langle T_{00}\rangle. \tag{18}$$

Both formulas are equivalent only for a special value of the surface gravity, $\kappa = a/c^2$. Since we are not aware of any way to motivate a specific choice of κ, and a needs to be very large, $a \gg c^2/l$, we cannot say anything conclusive about the equivalence of both entropies. We will return to this question after we introduce the modified gravitational dynamics.

3. Effective low energy quantum gravity dynamics

Upon presenting the various notions of entropy involved, we proceed to discuss our proposal for the effective low energy quantum gravitational dynamics. Details of the derivation can be found in the original paper of the authors;[11] here we just briefly recall the main points. For the sake of comparison, we employed two different thermodynamic derivations. The first one, known as the physical process approach, keeps track of the Clausius entropy flux across the horizon and the corresponding changes in its Bekenstein entropy. Equilibrium condition for this process reads $\Delta S_{e,q} + \Delta S_C = 0$, where the Clausius entropy flux contribution, ΔS_C, was discussed in the previous section. The Bekenstein entropy part, $\Delta S_{e,q}$, can be evaluated from the formula for area of GLCD's boundary cross-section introduced in subsection 2.1. After a series of calculations (see[11] for details), we obtain the modified equations of motion

$$S_{\mu\nu} - \frac{\mathcal{C}l_P^2}{18\pi} S_{\mu\lambda} S^{\lambda}{}_{\nu} + \frac{\mathcal{C}l_P^2}{72\pi}\left(R_{\kappa\lambda}R^{\kappa\lambda} - \frac{1}{4}R^2\right) g_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{4}T g_{\mu\nu}\right), \tag{19}$$

where $S_{\mu\nu} = R_{\mu\nu} - R g_{\mu\nu}/4$ denotes the traceless part of Ricci tensor. The derivation works for GLCD's length scale l much larger than the Planck length but much smaller than the curvature length scale (square root of the inverse of the largest eigenvalue of the Riemann tensor).

The second method to derive the modified equations of motion starts from the entanglement equilibrium hypothesis: "When the geometry and quantum fields are simultaneously varied from maximal symmetry, the entanglement entropy in a small geodesic ball is maximal at fixed volume".[19] For a small variation from a maximally symmetric spacetime the hypothesis implies $\delta S_{e,q} + \delta S_m = 0$, where δS_m denotes the

variation of matter von Neumann entropy. From here, calculations proceed similarly as for the physical process approach, yielding[11]

$$S_{\mu\nu} - \frac{\mathcal{C} l_{Pl}^2}{30\pi} S_{\mu\lambda} S^{\lambda}{}_{\nu} + \frac{\mathcal{C} l_{Pl}^2}{120\pi} \left(R_{\kappa\lambda} R^{\kappa\lambda} - \frac{1}{4} R^2 \right) g_{\mu\nu} = \frac{8\pi G}{c^4} \left(\delta \langle T_{\mu\nu} \rangle - \frac{1}{4} \delta \langle T \rangle g_{\mu\nu} \right). \tag{20}$$

Notably, in neither derivation does the form of modified temperature affect the resulting equations. This is expected as definitions of both Hawking and Unruh temperature (even modified ones) are purely kinematic with no dependence on gravitational dynamics.[64] The correction terms are thus fully determined by the logarithmic contribution to Bekenstein entropy and they depend only on one presently unknown parameter, $\mathcal{C}$.

Comparing the results of both approaches, we can see that the equations have the same form, although coefficients in front of the terms quadratic in curvature differ. We can write both equations in a common form

$$S_{\mu\nu} - D l_P^2 S_{\mu\lambda} S^{\lambda}{}_{\nu} + \frac{D}{4} l_P^2 \left(R_{\kappa\lambda} R^{\kappa\lambda} - \frac{1}{4} R^2 \right) g_{\mu\nu} = \frac{8\pi G}{c^4} \left(\langle T_{\mu\nu} \rangle - \frac{1}{4} \langle T \rangle g_{\mu\nu} \right), \tag{21}$$

with $D = \mathcal{C}/18\pi$ for the physical process approach and $D = \mathcal{C}/30\pi$ for the entanglement equilibrium hypothesis one. In both cases, the coefficient D has the same sign and differs only by a factor 3/5. The difference might be due to the details of the variation considered in the entanglement equilibrium hypothesis derivation, which is performed at fixed spatial volume.[11,19] Since, for modified theories of gravity, a more complicated quantity known as generalised volume needs to be held fixed instead,[21] finding the appropriate generalised volume in our case could account for the difference in D between both approaches. Alternatively, it might indicate a failure on the equivalence between Clausius and von Neumann entropy of the matter at this level (although, given the similarity of the results, both entropies would have to remain closely related).

Upon reviewing the two derivations of the modified equations of motion, we proceed to discus their properties (for a more detailed treatment, see[11]). First, note that while the value and even the sign of $\mathcal{C}$ is model dependent, any possible value can be viewed as being of the order of unity compared to the squared Planck length, $l_P^2 \approx 2.6 \times 10^{-70}\,\mathrm{m}^2$. Therefore, the correction terms become relevant only when the curvature length scale nears the Planck length (although it needs to remain significantly larger than l_P, otherwise the assumptions of our derivation will break down).

As a second property, it is easy to check that the modified equations are traceless. Hence, they do not imply local energy-momentum conservation. If we want to assume it, we must impose divergence-free energy-momentum tensor, $\langle T_{\mu}{}^{\nu} \rangle_{;\nu} = 0$, as an additional condition. Satisfying it requires

$$\frac{1}{4} R_{;\mu} - D l_P^2 \left(S^{\lambda\nu} S_{\mu\lambda} \right)_{;\nu} + \frac{D l_P^2}{2} \left(R^{\kappa\lambda} R_{\kappa\lambda;\mu} - \frac{1}{4} R R_{;\mu} \right) = -\frac{2\pi G}{c^4} \langle T \rangle_{;\mu}. \tag{22}$$

This condition does not have a general solution for T. Therefore, in general, the equations of motion cannot be recast in a form directly implying a divergence-free energy-momentum tensor. Nevertheless, the condition can be solved for spacetimes with a vanishing Weyl tensor. In that case, the cosmological constant, absent in the traceless equations of motion, appears as an arbitrary integration constant.

Third, while any terms containing higher than fourth derivatives of the metric or more than quadratic in curvature tensors are likely suppressed by higher powers of l_P, the higher derivative terms known from quadratic gravity should appear at the same order as the corrections we propose. They implicitly appear on the right hand side of modified equations in the quantum expectation value of the energy-momentum tensor.[65] In principle, one might also find higher derivative contributions to the left hand side by considering higher orders in the Riemann normal coordinate expansion of the metric. However, these corrections are ambiguous as they depend on shape deformations of the GLCD's horizon.[40] Without a physically motivated way to resolve these ambiguities, it is not possible to determine higher derivative terms contributing to the left hand side of the equations. Since these terms are anyway contained with undetermined coefficients on the right hand side, in the energy-momentum tensor expectation value, their omission on the left hand side does not change the resulting dynamics in any significant way.

Traceless equations of motion and the status of cosmological constant as an arbitrary integration constant both point out to the modified dynamics being a generalisation of unimodular gravity (or, more precisely, Weyl transverse gravity[66]). In fact, even the semiclassical gravitational dynamics derived from thermodynamics has more in common with unimodular gravity than general relativity.[25,36,67] We can partially understand the emergence of unimodular gravity from thermodynamics by noting that only the difference between entropy of two states is relevant for deriving the gravitational dynamics. In other words, vacuum contribution to matter entanglement entropy does not affect the conditions for thermodynamic equilibrium. Consequently, vacuum energy naturally does not couple to gravity, leading to the behaviour of the cosmological constant characteristic for unimodular gravity. Furthermore, thermodynamics of spacetime do not imply local energy-momentum conservation, which always needs to be postulated as an additional condition, again in agreement with unimodular gravity.

3.1. *Application to a cosmological toy model*

To illustrate the physical consequences of the modified dynamics, we examine a simple cosmological toy model, a homogeneous, isotropic, spatially flat universe filled with a perfect fluid with the equation of state $p = (\gamma - 1)\, c^2 \rho$ for some $\gamma \in [1, 2]$ (the limits correspond to dust and stiff matter, respectively). Due to previously noted unimodular character of the equations, we consider a unimodular form of the

Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$\mathrm{d}s^2 = -\frac{c^2}{a\left(t\right)^6}\mathrm{d}t^2 + a\left(t\right)^2\left(\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2\right), \tag{23}$$

where $a\left(t\right)$ denotes the scale factor. Because of the symmetries of the metric, the modified equations of the motion yield only one non-trivial condition

$$\dot{H} - D\frac{l_P^2\dot{H}^2}{c^2} = -4\pi\gamma G\rho, \tag{24}$$

where $H = \dot{a}/a$ is the Hubble parameter and the dot denotes a time derivative. This condition corresponds to a modified Raychaudhuri equation, with a correction term, $-Dl_P^2\dot{H}^2/c^2$, non-linear in second time derivatives of the scale factor. To simplify it, we assume that the Hubble parameter corresponds to its classical value H_0 up to $O\left(l_P^2\right)$ terms (validity of this assumption is extensively discussed in[11]),

$$H = H_0 + l_P^2 H_1 + O\left(l_P^4\right), \tag{25}$$

and rewrite the modified Raychaudhuri equations in the following way

$$\dot{H} = -4\pi\gamma G\rho\left(1 - 4\pi D\gamma\frac{\rho}{\rho_P}\right), \tag{26}$$

where $\rho_P = \sqrt{c^5/G^2\hbar}$ denotes the Planck density. If we further assume local energy-momentum conservation, we can integrate the modified Raychaudhuri equation to obtain the modified Friedmann equation

$$H^2 = \frac{8\pi G\rho}{3}\left(1 - \frac{2\pi D\gamma\rho}{\rho_P}\right) + \tilde{\Lambda}, \tag{27}$$

where $\tilde{\Lambda}$ is an arbitrary integration constant corresponding to the cosmological term ($\tilde{\Lambda} = \Lambda c^2/3$). We see that the appearance of cosmological constant is indeed consistent with unimodular gravity.

Notably, for $D > 0$, our results correspond to the effective description of loop quantum cosmology (LQC).[68] Since effective dynamics of LQC replaces the Big Bang singularity by a non-singular quantum bounce, the same conclusion holds for our modified equations of motion assuming $D > 0$. The bounce corresponds to a critical density, $\rho_{crit} = \rho_P/2\pi D\gamma$. In other words, if the logarithmic correction to Bekenstein entropy of a GLCD is positive, the resulting modifications of the dynamics are already sufficient to remove the cosmological singularity. On the other hand, the case of $D < 0$ appears to not only preserve the singularity, but actually strengthen the gravitational attraction responsible for it. These preliminary results warrant a more detailed analysis of cosmological models and possible values for parameter D.

4. Discussion

In the present work we have first reviewed the issue of entropy equivalence in thermodynamics of spacetime. For conformal fields in the semiclassical regime, Clausius and von Neumann entropy turn out to be equivalent in the semiclassical regime independently of the gravitational dynamics, although a similar clear result is not available for non-conformal fields.

Then, we proceeded to review our proposal for a new phenomenological perspective on the effects of quantum gravity at low energies. The effective low energy quantum gravitational dynamics we have proposed is based on a single assumption: leading order quantum gravity correction to entropy associated with spherical local causal horizons is logarithmic in the horizon area. Since a number of conceptually different approaches predict such a logarithmic term (at the very least in the case of black hole entropy), our conclusions are fairly robust and relevant to many candidate theories of quantum gravity, e.g. LQG, string theory, AdS/CFT correspondence and GUP phenomenology.

Let us remark that the equations we found are a generalisation of the classical equations of motion of unimodular gravity and cannot even be restated as some generalised Einstein equations. This agrees with our previous results concerning the relation between thermodynamics of spacetime and unimodular gravity in the semiclassical setting. It does appear that the corrections we propose break the equivalence of unimodular gravity and general relativity that holds on the level of classical dynamics.

Our results also extend the semiclassical equivalence between Clausius and entanglement entropies. While they suggest a possible breaking of the exact equivalence on the quantum level, both entropies remain strongly related. Their precise relation should be analysed carefully in a future work. Importantly, a more systematic inclusion of quantum gravity effects on matter entropy will be necessary to completely resolve this issue.

Note that our approach still requires further development. First, issues of diffeomorphism invariance, equivalence principle and locality of the modified dynamics previously discussed by the authors[11] need to be addressed in greater detail. Moreover, the apparent breaking of local energy conservation requires a physically motivated explanation. Relating our proposal to other modified gravity theories might shed some light on all these questions. In particular, our effective equations of motion resemble the 4-dimensional version of Einstein-Gauss-Bonnet gravity[69, 70] and non-local effective field theory of gravity.[71] The latter approach even leads to logarithmic term in Bekenstein entropy consistent with our assumptions.[32]

Another possible further development of our results lies in a careful inclusion of higher order Riemann normal coordinate corrections in the analysis in order to obtain higher derivative terms on the left hand side. Similarly, accounting for low energy quantum gravity effects on the entropy of matter might lead to interesting conclusions.

Furthermore, we have seen that the modified dynamics resolve the Big Bang singularity only if the logarithmic correction to entropy of a GLCD is positive, i.e., $D > 0$. To find which approaches to quantum gravity support this sign and what are the bounds on the magnitude of D they imply would help to constrain the physical implications of our proposal.

Lastly, the emergence of unimodular (Weyl transverse) gravity from thermodynamics needs to be better understood even in the semiclassical case. The first steps in this direction include developing rigorous formalism for thermodynamics of spacetime in the Weyl transverse gravity and connecting Weyl invariance of gravitational dynamics with properties of entropy. All these issues will be addressed in future works.

Acknowledgments

AA-S is supported by the ERC Advanced Grant No. 740209. ML is supported by the Charles University Grant Agency project GAUK 297721.

References

1. L. J. Garay, "Quantum gravity and minimum length," Int. J. Mod. Phys. A **10** (1995) 145 [gr-qc/9403008].
2. S. Hossenfelder, "Minimal Length Scale Scenarios for Quantum Gravity," Living Rev. Relativ. **16** (2013) 2 [gr-qc/1203.6191].
3. L. Smolin, "Four principles for quantum gravity," Fundam. Theor. Phys. **187** (2017) 427 [gr-qc/1610.01968].
4. S. Chakraborty, D. Kothawala and A. Pesci, "Raychaudhuri equation with zero point length," Phys. Lett. B **797** (2019) 134877 [gr-qc/1904.09053].
5. l. Buoninfante, G. Luciano, L. Petruzziello and F. Scardigli, "Bekenstein bound and uncertainty relations," [hep-th/2009.12530].
6. J. D. Bekenstein and V. F. Mukhanov, "Spectroscopy of the quantum black hole," Phys. Lett. B **360** (1995) 7 [gr-qc/9505012].
7. R. J. Adler, P. Chen and D. I. Santiago, "The Generalized Uncertainty Principle and Black Hole Remnants," Gen. Rel. Grav. **33** (2001) 2101 [gr-qc/0106080].
8. A. Awad and A. F. Ali, "Minimal Length, Friedmann Equations and Maximum Density," JHEP **06** (2014) 093 [gr-qc/1404.7825].
9. A. Alonso-Serrano, M. P. Dabrowski and H. Gohar, "Generalized uncertainty principle impact onto the black holes information flux and the sparsity of Hawking radiation," Phys. Rev. D **97** (2018) 044029 [gr-qc/1801.09660].
10. R. Casadio, "Quantum black holes and resolution of the singularity," [gr-qc/2103.00183].
11. A. Alonso-Serrano and M. Liška, "Quantum phenomenological gravitational dynamics: A general view from thermodynamics of spacetime," JHEP **2020** (2020) [gr-qc/2009.03826].
12. J. D. Bekenstein, "Black Holes and Entropy," Phys. Rev. D **7** (1973) 2333.
13. J. D. Bardeen, B. Carter and S. W. Hawking, "The Four Laws of Black Hole Mechanics," Commun. Math. Phys. **31** (1973) 161.
14. S. W. Hawking, "Particle Creation by Black Holes," Commun. Math. Phys. **43** (1975) 199.

15. T. Jacobson and R. Parentani, "Horizon Entropy," Found. Phys. **33** (2003) 323 [gr-qc/0302099].
16. R. M. Wald, "Black Hole Entropy is Noether Charge," Phys. Rev. D **48** (1993) 3427 [gr-qc/9307038].
17. T. Jacobson, "Thermodynamics of space-time: The Einstein equation of state," Phys. Rev. Lett. **75** (1995) 1260 [gr-qc/9504004].
18. V. Baccetti and M. Visser, "Clausius entropy for arbitrary bifurcate null surfaces," Class. Quant. Grav. **31** (2014) 035009 [gr-qc/1303.3185].
19. T. Jacobson, "Entanglement Equilibrium and the Einstein Equation," Phys. Rev. Lett. **116** (2016) 201101 [gr-qc/1505.04753].
20. M. Parikh and A. Svesko, "Einstein's Equations from the Stretched Future Light Cone," Phys. Rev. D **98** (2018) 026018 [hep-th/1712.08475].
21. P. Bueno, V. S. Min, A. J. Speranza and M. R. Visser, "Entanglement equilibrium for higher order gravity," Phys. Rev. D **95** (2017) 046003 [hep-th/1612.04374].
22. A. Svesko, "Equilibrium to Einstein: Entanglement, Thermodynamics, and Gravity," Phys. Rev. D **99** (2019) 086006 [hep-th/1810.12236].
23. R. Guedens, T. Jacobson and S. Sarkar, "Horizon entropy and higher curvature equations of state," Phys. Rev. D **85** (2012) 064017 [gr-qc/1112.6215].
24. C. Eling, R. Guedens and T. Jacobson "Non-equilibrium Thermodynamics of Spacetime," Phys. Rev. Lett. **96** (2006) 121301 [gr-qc/0602001].
25. T. Padmanabhan, "Thermodynamical Aspects of Gravity: New insights," Rep. Prog. Phys. **73** (2010) 046901 [gr-qc/0911.5004].
26. G. Chirco, S. Liberati, "Non-equilibrium Thermodynamics of Spacetime: the Role of Gravitational Dissipation," Phys. Rev. D **81** (2010) 024016 [gr-qc/0909.4194v1].
27. R. K. Kaul and P. Majumdar, "Logarithmic Correction to the Bekenstein-Hawking Entropy," Phys. Rev. Lett. **84** (2000) 5255 [gr-qc/0002040].
28. K. A. Meissner, "Black hole entropy in Loop Quantum Gravity," Class. Quant. Grav. **21** (2004) 5245 [gr-qc/0407052].
29. A. Sen, "Logarithmic Corrections to Schwarzschild and Other Non-extremal Black Hole Entropy in Different Dimensions," JHEP **04** (2013) 156 [hep-th/1205.0971].
30. S. Banerjee, R. K. Gupta, I. Mandal and A. Sen, "Logarithmic Corrections to N=4 and N=8 Black Hole Entropy: A One Loop Test of Quantum Gravity," JHEP **11** (2011) 143 [hep-th/1106.0080].
31. T. Faulkner, A. Lewkowycz and J. Maldacena, "Quantum corrections to holographic entanglement entropy," JHEP **2013** (2013) 074 [hep-th/1307.2892].
32. Y. Xiao and Y. Tian, [gr-qc/2104.14902].
33. S. Solodukhin, "Entanglement entropy of black holes," Living Rev. Rel. **14** (2011) 8 [hep-th/1104.3712].
34. S. Das, P. Majumdar and R. K. Bhaduri, "General Logarithmic Corrections to Black Hole Entropy," Class. Quant. Grav. **19** (2002) 2355 [hep-th/0111001].
35. G. Gourd and A. J. M. Medved, "Thermal Fluctuations and Black Hole Entropy," Class. Quant. Grav. **20** (2003) 3307 [gr-qc/0305018].
36. A. Alonso-Serrano and M. Liška, "New Perspective on thermodynamics of spacetime: The emergence of unimodular gravity and the equivalence of entropies," Phys. Rev. D **102** (2020) 104056 [gr-qc/2008.04805].
37. C. W. Misner, K. S. Thorne and J. A. Wheeler, "Gravitation," Princeton University Press, Princeton, NJ, U.S.A. (2017).
38. S. M. Carroll and G. N. Remmen, "What is the Entropy in Entropic Gravity?," Phys Rev D **93** (2016) 124052 [hep-th/1601.07558].

39. G. W. Gibbons and S. Solodukhin, "The Geometry of Small Causal Diamonds," Phys. Lett. B **649** (2007) 317 [hep-th/0703098].
40. T. Jacobson, J. M. M. Senovilla and A. J. Speranza, "Area deficits and the Bel-Robinson tensor," Class. Quant. Grav. **35** (2018) 085005 [gr-qc/1710.07379].
41. J. Wang, "Geometry of small causal diamonds," Phys. Rev. D **100** (2019) 064020 [gr-qc/1904.01034].
42. T. Jacobson and M. R. Visser, "Gravitational Thermodynamics of Causal Diamonds in (A)dS," SciPost Phys **7** (2019) 079 [hep-th/1812.01596].
43. L. Bombelli, R. K. Koul, J. Lee, R. D. Sorkin, "Quantum source of entropy for black holes," Phys. Rev. D **34** 373 (1986).
44. M. Srednicki, "Entropy and area," Phys. Rev. Lett. **71** (1993) 666 [hep-th/9303048v2].
45. T. Jacobson, "Black Hole Entropy and Induced Gravity," [gr-qc/9404039].
46. A. Davidson, "From Planck area to graph theory: Topologically distinct black hole microstates," Phys Rev D **100** (2019) 081502 [hep-th/1907.03090].
47. S. Hod, "High-Order Corrections to the Entropy and Area of Quantum Black Holes," Class. Quant. Grav. **21** (2004) L97 [hep-th/0405235].
48. A. J. M. Medved and E. C. Vagenas, "When conceptual worlds collide: The GUP and the BH entropy," Phys. Rev. D **70** (2004) 124021 [hep-th/0411022].
49. A. J. M. Medved, "A follow-up to 'Does Nature abhor a logarithm?' (and apparently she doesn't)," Class. Quant. Grav. **22** (2005) 5195 [gr-qc/0411065].
50. S. Solodukhin, "Entanglement entropy of round spheres," Phys. Lett. B **693** (2010) 605 [hep-th/1008.4314].
51. L. C. Barbado and M. Visser, "Unruh-DeWitt detector event rate for trajectories with time-dependent acceleration," Phys. Rev. D **86** (2012) 084011 [gr-qc/1207.5525].
52. P. Martinetti and C. Rovelli, "Diamonds's Temperature: Unruh effect for bounded trajectories and thermal time hypothesis," Class. Quant. Grav. **20** (2003) 4919-4932 [gr-qc/0212074].
53. M. Arzano, "Conformal quantum mechanics of causal diamonds," JHEP **2020** (2020) [gr-qc/2002.01836].
54. E. di Casola, S. Liberati, S. Sonego, "Nonequivalence of equivalence principles," Am. J. Phys. **83** (2015) 39 [gr-qc/1310.7426].
55. S. Ghosh, "Quantum Gravity Effects in Geodesic Motion and Predictions of Equivalence Principle Violation," Class. Quant. Grav. **31** (2014) 025025 [gr-qc/1303.1256].
56. V. M. Tkachuk, "Effect of the generalized uncertainty principle on Galilean and Lorentz transformations," Found. Phys. **46** (2016) 1666 [quant-ph/1310.6243].
57. S. Pramanik, "Implication of Geodesic Equation in Generalized Uncertainty Principle framework," Phys. Rev. D **90** (2014) 024023 [hep-th/1404.2567].
58. R. Casadio and F. Scardigli, "Generalized Uncertainty Principle, Classical Mechanics, and General Relativity," Phys. Lett. B **598** (2020) 135558 [gr-qc/2004.04076].
59. F. Scardigli, M. Blasone, G. Luciano and R. Casadio, "Modified Unruh effect from Generalized Uncertainty Principle," Eur. Phys. J. C **78** (2018) 728 [hep-th/1804.05282].
60. G. G. Luciano and L. Petruzziello, "GUP parameter from Maximal Acceleration," Eur. Phys. J. C **79** (2019) 283 [hep-th/1902.07059].
61. E. R. Caianiello, "Maximal acceleration as a consequence of Heisenberg's uncertainty relations," Lett. Nuovo Cimento **41** (1984) 370.
62. A. J. Speranza, "Entanglement entropy of excited states in conformal perturbation theory and the Einstein equation," JHEP **2016** (2016) [hep-th/1602.01380].
63. H. Casini, D. A. Galante, R. C. Myers, "Comments on Jacobson's "Entanglement equilibrium and the Einstein equation"," JHEP **2016** (2016) [hep-th/1601.00528].

64. M. Visser, "Essential and inessential features of Hawking radiation," Int. J. Mod. Phys. D **12** (2003) 649 [hep-th/0106111].
65. R. M. Wald, "Quantum field theory in curved spacetime and black hole thermodynamics," The University of Chicago Press, Chicago, IL, U.S.A. and London, U.K. (1994).
66. C. Barceló, L. J. Garay and R. Carballo-Rubio, "Absence of cosmological constant problem in special relativistic field theory of gravity," Ann. Phys. **398** (2018) 9 [gr-qc/1406.7713].
67. S. C. Tiwari, "Thermodynamics of Spacetime and Unimodular Relativity," (2006) [gr-qc/0612099].
68. A. Ashtekar and B. Gupt, "Generalized effective description of loop quantum cosmology," Phys. Rev. D **92** (2015) 084060 [gr-qc/1509.08899].
69. D. Glavan and C. Lin, "Einstein-Gauss-Bonnet gravity in 4-dimensional space-time," Phys. Rev. Lett. **124** (2019) 081301 [gr-qc/1905.03601].
70. J. Arrechea, A. Delhom and A. Jiménez-Cano, "Inconsistencies in four-dimensional Einstein-Gauss-Bonnet gravity," CPC **45** (2021) 013107 [gr-qc/2004.12998].
71. J. F. Donoghue, "General relativity as an effective field theory: The leading quantum corrections," Phys. Rev. D **50** (1994) 3784 [gr-qc/405057].

Gauge theory of gravity based on the correspondence between the 1^{st} and the 2^{nd} order formalisms

David Benisty

DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

A covariant canonical gauge theory of gravity free from torsion is studied. Using a metric conjugate momentum and a connection conjugate momentum, which takes the form of the Riemann tensor, a gauge theory of gravity is formulated, with form-invariant Hamiltonian. By the metric conjugate momenta, a correspondence between the Affine-Palatini formalism and the metric formalism is established. For, when the dynamical gravitational Hamiltonian $\tilde{H}_{Dyn}$ does not depend on the metric conjugate momenta, a metric compatibility is obtained from the equation of motions, and the equations of motion correspond to the solution is the metric formalism.

Keywords: Gauge-Theory; Gravity; Hamiltonian formulation.

1. Introduction

General Relativity is one of the well tested theories in physics, with many excellent predictions. A search of a rigorous derivation of General Relativity on the basis of the action principle and the requirement that the description of any system should be form-invariant under general space time transformations has been constructed in the framework of the Covariant Canonical Gauge theory of Gravity.

The Covariant Canonical Gauge theory of Gravity[1,2] is formulated within the framework of the covariant Hamiltonian formalism of classical field theories. The latter ensures by construction that the action principle is maintained in its form requiring all transformations of a given system to be canonical. The imposed requirement of invariance of the original action integral with respect to local transformations in curved space time is achieved by introducing additional degrees of freedom, the gauge fields. In the basis of the formulation there are two independent fields: the metric $g^{\alpha\beta}$, which contains the information about lengths and angles of the space time, and the connection $\gamma^{\lambda}_{\alpha\beta}$, which contains the information how a vector transforms under parallel displacement. In this formulation, these two fields are assumed to be independent dynamical quantities in the action and referred to as the Affine-Palatini formalism (or the 1^{st} order formalism).

Using the structure of metric and the affine connection independently, with their conjugate momenta, yields a new formulation of gauge theories of gravity. In[1] the discussion was with the presence of torsion, and here we discuss about the formulation with no torsion from the beginning, that has a link between the metric affine and the metric formalism.[3]

2. A basic formulation

The Covariant Canonical Gauge theory of Gravity is a well defined formulation derived from the canonical transformation theory in the covariant Hamiltonian picture of classical field theories.[1] It identifies two independent fundamental fields, which form the basis for a description of gravity: the metric $g^{\alpha\beta}$ and the connection $\gamma^{\lambda}_{\alpha\beta}$. In the Hamiltonian description, any fundamental field has a conjugate momentum: the metric conjugate momentum is $\tilde{k}^{\alpha\lambda\beta}$ and the connection conjugate momentum is $\tilde{q}^{\alpha\xi\beta}_{\eta}$:

$$S = \int_R \left(\tilde{k}^{\,\alpha\lambda\beta}\, g_{\alpha\lambda,\beta} - \frac{1}{2}\tilde{q}^{\alpha\xi\beta}_{\eta}\gamma^{\eta}_{\alpha\xi,\beta} - \tilde{\mathcal{H}}_0 \right) d^4x \tag{1}$$

where the "tilde" sign denotes a tensor density, which multiplies the tensor with $\sqrt{-g}$. As the conjugate momentum components of the fields are the duals of the complete set of the derivatives of the field, the formulation is referred to as "covariant canonical". A closed description of the coupled dynamics of fields and space-time geometry has been derived in,[1] where the gauge formalism yields:

$$S = \int_R \left(\tilde{k}^{\,\alpha\lambda\beta}\, g_{\alpha\lambda;\beta} - \frac{1}{2}\tilde{q}^{\alpha\xi\beta}_{\eta}R^{\eta}_{\alpha\xi\beta} - \mathcal{H}(\tilde{q},\tilde{k},g) \right) d^4x \tag{2}$$

As a result of the gauge procedure, all partial derivatives of tensors in Eq. (1) reappear as covariant derivatives. The partial derivative of the (non-tensorial) connection changes into the tensor $R^{\eta}_{\alpha\xi\beta}$, which was shown to be the Riemann-Christoffel curvature tensor:

$$R^{\eta}_{\alpha\xi\beta} = \frac{\gamma^{\eta}_{\alpha\beta}}{x^{\xi}} - \frac{\gamma^{\eta}_{\alpha\xi}}{x^{\beta}} + \gamma^{\tau}_{\alpha\beta}\gamma^{\eta}_{\tau\xi} - \gamma^{\tau}_{\alpha\xi}\gamma^{\eta}_{\tau\beta}. \tag{3}$$

The "dynamics" Hamiltonian $\tilde{H}_{Dyn}$ — which is supposed to describe the dynamics of the free (uncoupled) gravitational field — is to be built from a combination of the metric conjugate momentum, the connection conjugate momentum, and the metric itself.

3. A correspondence between the 1^{st} and the 2^{nd} order formalism

In addition to the foundations of the gauge theory of gravity, it turned out that the part of the action: $\tilde{k}^{\alpha\beta\gamma}g_{\alpha\beta;\gamma}$, which contains the metric conjugate momentum, has a strong impact as a connector between the affine-Palatini formalism (or the 1^{st} order formalism) and the metric formalism (or the 2^{nd} order formalism):

$$\mathcal{L}(g,\gamma)_{\,1^{\mathbf{order}}} + k^{\alpha\beta\gamma}g_{\alpha\beta;\gamma} \Leftrightarrow \mathcal{L}(g)_{\,2^{\mathbf{order}}} \tag{4}$$

In the 1^{st} order formalism, one assumes that there are two independent fields: the metric $g^{\mu\nu}$ and the connection $\gamma^{\mu}_{\alpha\beta}$. In contrast to that, in the 2^{nd} order formalism the connection is assumed to be the Levi Civita or Christoffel symbol:

$$\gamma^{\rho}_{\mu\nu} = \left\{ \begin{matrix} \rho \\ \mu\nu \end{matrix} \right\} = \frac{1}{2}g^{\rho\lambda}(g_{\lambda\mu,\nu} + g_{\lambda\nu,\mu} - g_{\mu\nu,\lambda}) \tag{5}$$

and appears in the action directly in this way. In general, only for Lovelock theories, which includes Einstein Hilbert action, both formulations will yield the same equations of motion and the connection will be in both cases the Christoffel symbol. In Ref,[3] it was proved that for any general action which starts in the 1^{st} order formalism in addition to the term $k^{\alpha\beta\gamma}g_{\alpha\beta;\gamma}$ the energy momentum tensor will be the same as it would be calculated in the 2^{nd} order formalism. The main reason for that correspondence is the metric compatibility constraint. The variation with respect to $k^{\alpha\beta\gamma}$ gives the metricity condition:

$$g_{\alpha\beta;\gamma} = 0 \quad \Rightarrow \quad \gamma^{\rho}_{\mu\nu} = \left\{ \begin{array}{c} \rho \\ \mu\nu \end{array} \right\}, \tag{6}$$

which cause the connection to be the Christoffel symbol. The variation with respect to the connection gives the tensors:

$$\frac{\delta}{\delta\gamma^{\rho}_{\mu\nu}} k^{\alpha\beta\gamma} g_{\alpha\beta;\gamma} = -k^{\alpha\mu\nu} g_{\rho\alpha} - k^{\alpha\nu\mu} g_{\rho\alpha} \tag{7}$$

with a symmetrization between the components μ and ν. The variation with respect to the metric is:

$$\frac{\delta}{\delta g_{\mu\nu}} k^{\alpha\beta\gamma} g_{\alpha\beta;\gamma} = -k^{\mu\nu\lambda}_{;\lambda}. \tag{8}$$

Because of the new contribution to the field equation $k^{\mu\nu\lambda}_{;\lambda}$, the complete field equation will contains additional terms which make the first order field equations to be equivalent to the field equation under the second order formalism. Indeed, isolating the tensor $k^{\mu\nu\lambda}$ and inserting it back into Eq. (8) gives the relation:

$$\frac{\partial \mathcal{L}(\kappa)}{\partial g_{\sigma\nu}} = \frac{1}{2}\nabla_{\mu}\left(g^{\rho\sigma}\frac{\partial \mathcal{L}(\kappa)}{\partial \gamma^{\rho}_{\mu\nu}} + g^{\rho\nu}\frac{\partial \mathcal{L}(\kappa)}{\partial \gamma^{\rho}_{\mu\sigma}} - g^{\rho\mu}\frac{\partial \mathcal{L}(\kappa)}{\partial \gamma^{\rho}_{\nu\sigma}} \right) \tag{9}$$

where $\mathcal{L}(\kappa) = k^{\alpha\beta\gamma}g_{\alpha\beta;\gamma}$. The terms in the right hand side represents the additional terms that appear in the second order formalism. One option for obtain the contributions into the field equation is to solve $k^{\alpha\beta\gamma}$. The direct way is by using this equation, that gives the new contributions for the second order formalism into the field equation, from the variation with respect to the connection $\gamma^{\rho}_{\mu\nu}$. An application for this correspondence is from the Covariant Canonical Gauge theory of gravity action (2).

4. Path to Gauge theories

From the correspondence between the 1^{st} and the 2^{nd} order formalisms theorem, we obtain a basic link between the dependence of the $\mathcal{H}_{\mathbf{Dyn}}$ with the metric conjugate momentum $\tilde{k}^{\alpha\beta\gamma}$ and the structure of the metric energy momentum tensor. In the first case, $\mathcal{H}_{\mathbf{Dyn}}$ does not depend on the metric conjugate momentum $\tilde{k}^{\alpha\beta\gamma}$. A variation with respect to the metric conjugate momentum $\tilde{k}^{\alpha\beta\gamma}$ gives the metric compatibility condition. According to the theorem (4) the gravitational energy momentum tensor is the same as the gravitational energy momentum tensor in the

second order formalism. In the second case $\mathcal{H}_{\mathbf{Dyn}}$ does depend on the metric conjugate momentum $\tilde{k}^{\alpha\beta\gamma}$. A variation with respect to the metric conjugate momentum $\tilde{k}^{\alpha\beta\gamma}$ breaks the metric compatibility condition. This basic framework is not a special feature only for the Covariant Canonical Gauge Theory of Gravity, but leads to a fundamental correlation for many options of $\mathcal{H}_{Dyn}$.

In analogy to the definition of the metric energy-momentum tensor density of the given system Hamiltonian, the metric energy-momentum tensor density is being define as the variation of the $\tilde{\mathcal{L}}_m$ with respect to the metric:

$$T^{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\partial\tilde{\mathcal{L}}_m}{\partial g_{\mu\nu}} \tag{10}$$

Therefore the complete action takes the form:

$$\tilde{\mathcal{L}} = \tilde{k}^{\alpha\beta\gamma}g_{\alpha\beta;\gamma} - \frac{1}{2}\tilde{q}^{\alpha\beta\gamma}_{\lambda}R^{\lambda}_{\alpha\beta\gamma} - \tilde{\mathcal{H}}_{\mathbf{dyn}}(\tilde{q},\tilde{k},g) + \tilde{\mathcal{L}}_m \tag{11}$$

The variation with respect to the metric conjugate momenta:

$$g_{\alpha\beta;\gamma} = \frac{\partial\tilde{H}_{\mathbf{Dyn}}}{\partial\tilde{k}^{\alpha\beta\gamma}} \tag{12}$$

which presents the existence of non-metricity if $\mathcal{H}_{Dyn}$ depends on $k^{\alpha\beta\gamma}$. The second variation is the variation with respect to the connection:

$$-\left(\tilde{k}^{\alpha\mu\nu} + \tilde{k}^{\alpha\nu\mu}\right)g_{\alpha\rho} = \frac{1}{2}\nabla_{\beta}\left(\tilde{q}^{\mu\beta\nu}_{\rho} + \tilde{q}^{\nu\beta\mu}_{\rho}\right), \tag{13}$$

which contracts the relation between the momenta of the metric and the connection. The third variation is with respect to the connection conjugate momentum $\tilde{q}^{\mu\nu\rho}_{\sigma}$, which gives:

$$\frac{\partial\tilde{H}_{\mathbf{Dyn}}}{\partial\tilde{q}^{\mu\nu\rho}_{\sigma}} = -\frac{1}{2}R^{\sigma}_{\mu\nu\rho} \tag{14}$$

If $\tilde{H}_{\mathbf{Dyn}}$ is not depend on $\tilde{q}$, the Riemann tensor will be zero. Therefore the contribution for the stress energy tensor comes from the Dynamical Hamiltonian and from the metric conjugate momenta:

$$T^{\mu\nu} = g^{\mu\nu}(k^{\alpha\beta\gamma}g_{\alpha\beta;\gamma} - \frac{1}{2}q^{\alpha\beta\gamma}_{\lambda}R^{\lambda}_{\alpha\beta\gamma}) - 2k^{\mu\nu\gamma}_{;\gamma} + \frac{2}{\sqrt{-g}}\frac{\partial\tilde{\mathcal{H}}_{\mathbf{Dyn}}}{\partial g_{\mu\nu}} \tag{15}$$

From the variation with respect the connection (13), the value of the momentum $\tilde{k}^{\alpha\beta\gamma}$. As an example, we consider a Dynamical Hamiltonian which has no dependence with the metric conjugate momentum.

5. Sample $\tilde{\mathcal{H}}_{\mathbf{dyn}}$ without breaking metricity

Our starting point is a dynamical Hamiltonian with the connection conjugate momentum up to the second order, without a dependence on the metric conjugate momentum:

$$\tilde{\mathcal{H}}_{\mathbf{dyn}} = \frac{1}{4g_1}\tilde{q}^{\alpha\epsilon\beta}_{\eta}q^{\eta\tau\lambda}_{\alpha}g_{\epsilon\tau}g_{\beta\lambda} - g_2\,\tilde{q}^{\alpha\tau\beta}_{\eta}g_{\alpha\beta}\delta^{\eta}_{\tau} + g_3\sqrt{-g} \tag{16}$$

This Hamiltonian was investigated in[4] under the original formalism for non-zero torsion (which is finally set to zero), and led to resolving the cosmological constant problem. In our case, assuming that there is no torsion, the formalism demands that the energy momentum tensor is covariantly conserved, as is supposed to be in the second order formalism.

The variation with respect to the metric conjugate momenta $\tilde{k}^{\alpha\beta\gamma}$ give the the metricity condition. The variation with respect to the connection conjugate momenta $\tilde{q}^{\mu\nu\rho}_{\sigma}$ gives

$$q_{\eta\alpha\epsilon\beta} = g_1 \left(R_{\eta\alpha\epsilon\beta} - \hat{R}_{\eta\alpha\epsilon\beta} \right) \tag{17}$$

where:

$$\hat{R}_{\eta\alpha\epsilon\beta} = g_2 \left(g_{\eta\epsilon} g_{\alpha\beta} - g_{\eta\beta} g_{\epsilon\alpha} \right) \tag{18}$$

refers to the ground state geometry of space-time which is the de Sitter (dS) or the anti-de Sitter (AdS) space-time for the positive or the negative sign of g_2, respectively. The last variation is with respect to the metric. In order to isolate the tensor $k^{\mu\nu\gamma}$ one can use the following process: First, we multiply by the metric $g^{\rho\sigma}$ and sum over the index σ:

$$-\tilde{k}^{\sigma\mu\nu} - \tilde{k}^{\sigma\nu\mu} = \frac{1}{2} \nabla_\alpha \left(\tilde{q}^{\ \sigma\mu\alpha\nu} + \tilde{q}^{\ \sigma\nu\alpha\mu} \right) \tag{19}$$

Switching the indices $\sigma \leftrightarrow \nu$ and the indices $\mu \leftrightarrow \nu$ gives a new combination of the $k^{\alpha\beta\gamma}$ tensor. By summing the Eqs. the isolated value gives:

$$-\tilde{k}^{\sigma\nu\mu} = \frac{1}{2} \nabla_\alpha \left(\tilde{q}^{\sigma\mu\alpha\nu} + \tilde{q}^{\nu\mu\alpha\sigma} \right) \tag{20}$$

Therefore, the contribution for the stress energy momentum comes from the covariant derivative of (20):

$$T^{\mu\nu} = -\frac{1}{2} g^{\mu\nu} q^{\alpha\beta\gamma}_{\lambda} R^{\lambda}_{\alpha\beta\gamma} + \nabla_\gamma \nabla_\alpha \left(q^{\mu\gamma\nu\alpha} + q^{\nu\gamma\mu\alpha} \right) + \frac{2}{\sqrt{-g}} \frac{\partial \tilde{\mathcal{H}}_{\mathbf{Dyn}}}{\partial g_{\mu\nu}} \tag{21}$$

Plugging in the value of the tensor $q_{\alpha\beta\gamma\delta}$ from Eq. (17) gives the result:

$$\begin{aligned} T^{\mu\nu} &= \frac{1}{8\pi G} G^{\mu\nu} + g^{\mu\nu} \Lambda \\ &+ g_1 \left(R^{\mu\alpha\beta\gamma} R^{\nu}_{\alpha\beta\gamma} - \frac{1}{4} g^{\mu\nu} R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta} + (\nabla_\alpha \nabla_\beta + \nabla_\beta \nabla_\alpha) R^{\mu\alpha\nu\beta} \right) \end{aligned} \tag{22}$$

where $G^{\mu\nu}$ is the Einstein tensor. The coupling constants relate to the physical quantities with the relations:

$$g_1 g_2 = \frac{1}{16\pi G}, \quad 6 g_1 g_2^2 + g_3 = \frac{\Lambda}{8\pi G} \tag{23}$$

This stress energy momentum tensor is exactly the same metric energy momentum tensor if our starting point was the effective Lagrangian:

$$\mathcal{L} = g_1 R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta} - \frac{1}{16\pi G} (R - 2\Lambda) \tag{24}$$

and the stress energy momentum tensor is the same stress energy momentum tensor for this Lagrangian in the second order formalism. One from the big benefits of this formulation is common in many gauge theories of gravity,[5] where the starting point is with additional variables with no higher derivatives in the action, and the equations of motion are equivalent to actions with higher derivatives of metric. In this specific case, the starting point is with the quartic momentum q and at the end is equivalent to an action with quadratic Riemann term.

6. Discussion

In this paper we investigated the formulation of the covariant canonical gauge theory of gravity free from torsion. Diffeomorphisms appear as canonical transformations. A tensor field which plays the role of the canonical conjugate of the metric is introduced. It enforces the metricity condition provided that the "Dynamics" Hamiltonian does not depend on this field. The resulting theory has a direct correspondence with our recent work concerning the correspondence between the first order formalism and the second order formalism through the introduction of a Lagrange multiplier field which in this case corresponds with the field that is used to provide the metric with a canonically conjugate momentum. The procedure is exemplified by using a "Dynamics" Hamiltonian which consists of a quadratic term of the connection conjugate momentum. The effective stress energy momentum tensor that emerged from the canonical equations of motion were equivalent to Einstein Hilbert tensor in addition to quadratic term.[6] derives the complete combination for the quadratic theories of gravity with R^2 and $R_{\mu\nu}R^{\mu\nu}$.[7] derives inflation from fermios based on the Covariant Canonical Gauge theories of Gravity approach to spinors.[8]

References

1. J. Struckmeier, J. Muench, D. Vasak, J. Kirsch, M. Hanauske and H. Stoecker, *Phys. Rev. D* **95** (2017) 124048, arXiv:1704.07246 [gr-qc].
2. J. Struckmeier, J. Muench, P. Liebrich, M. Hanauske, J. Kirsch, D. Vasak, L. Satarov and H. Stoecker, *Int. J. Mod. Phys. E* **28** (2019) 1950007, arXiv:1711.10333 [gr-qc].
3. D. Benisty and E. I. Guendelman, *Phys. Rev. D* **98** (2018) 044023, arXiv:1805.09667 [gr-qc].
4. D. Vasak, J. Struckmeier, J. Kirsch and H. Stoecker (2 2018) arXiv:1802.07137 [gr-qc].
5. M. H. Lagraa, M. Lagraa and N. Touhami, *Class. Quant. Grav.* **34** (2017) 115010, arXiv:1606.06918 [gr-qc].
6. D. Benisty, E. I. Guendelman, D. Vasak, J. Struckmeier and H. Stoecker, *Phys. Rev. D* **98** (2018) 106021, arXiv:1809.10447 [gr-qc].
7. D. Benisty, E. I. Guendelman, E. N. Saridakis, H. Stoecker, J. Struckmeier and D. Vasak, *Phys. Rev. D* **100** (2019) 043523, arXiv:1905.03731 [gr-qc].
8. J. Struckmeier, D. Vasak, A. Redelbach, P. Liebrich and H. Stöcker (12 2018) arXiv:1812.09669 [gr-qc].

$U(1)$ local strings in hybrid metric-Palatini gravity

Tiberiu Harko

Astronomical Observatory, 19 Ciresilor Street, 400487 Cluj-Napoca, Romania
Faculty of Physics, Babes-Bolyai University, 1 Kogalniceanu Street, 400084 Cluj-Napoca, Romania
School of Physics, Sun Yat-Sen University, Xingang Road, 510275 Guangzhou, People's Republic of China

Francisco S. N. Lobo

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisbon, Portugal

Hilberto M. R. da Silva

Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal and Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, PT4150-762 Porto, Portugal

In this work we made use of a general static cillindrically symmetric metric to find $U(1)$ local cosmic string solutions in the context of the hybrid metric-Palatini theory of gravity in it's scalar-tensor representation. After finding the dynamical equations for this particular case, we imposed boost invariance along t and z directions, which simplified the equations of motions, leaving only one single metric tensor component, $W^2(r)$. For an arbitrary potential $V(\phi)$, the solutions obtained can be put in a closed parametric form, with ϕ taken as a parameter. Several particular cases of the potential were studied, some yielding simple mathematical forms, others with only numerical solutions. In this way, we obtain a large class of novel stable stringlike solutions in the context of hybrid metric-Palatini gravity, in which the basic parameters, such as the scalar field, metric tensor components, and string tension, depend essentially on the initial values of the scalar field, and of its derivative, on the $r(0)$ circular axis.

Keywords: Cosmic Strings, Modified Gravity, hybrid metric-Palatini theory

1. Introduction

The recent observation of the accelerated expansion of the Universe[1] brought an identity crisis upon the standard theory of gravity, General Relativity. Albeit being a well established theory, with several important experimental confirmations, there is now a wide agreement that General Relativity may be, at the very least, an incomplete theory.[2] Not only due to the observations that demand the introduction of a Dark Sector to the energy budget of the Universe, but also due to theoretical difficulties, namely the existence of mathematical singularities and the non-quantization of the theory.[3] All of these issues paved the way to more general theories for gravitation, focusing on solving some of these problems, or at least to be used to falsify the theory. Lovelock's theorem[4] is a crucial roadmap to this quest,

as it establishes the GR equations as the only possible solution (minus constant factors), if one assumes the same conditions Einstein followed when he proposed the theory. In order to produce a viable modified gravity theory, at least one of the assumptions must be broken.

The hybrid metric-Palatini theory was proposed by Capozzielo, Harko, Koivisto, Lobo and Olmo in 2011[5] in order to overcome the difficulties faced by $f(R)$ theories of gravity and successfully unifies the late-time cosmic acceleration period with the weak-field solar system dynamics without the need for chameleon mechanisms.[5] For a more in-depth review of the theory, we refer the reader to the review article.[6] In this theory, the connection is considered an independent dynamical variable, adding to the metric.

Cosmic strings are one of the possible topological defects formed after spontaneous symmetry breaking (SSB) during phase transitions in the history of the Universe. In fact, many grand unification theories (GUTs) postulate that the universe, as it cooled, underwent a series of phase transitions associated with SSB, meaning that at sufficiently high temperatures there was invariance under a more general group of symmetries. Each of these phase transitions may have left behind a network of topological defects, through the Kibble-Zurek mechanism. For a more in depth review of topological defects, we refer the reader to Ref. 7.

The type of strings to be considered in this work are local $U(1)$ cosmic strings, which are an extension of the global $U(1)$ strings to include gauge fields. Local strings differ from the global cosmic strings in what concerns the symmetry that is effective above the spontaneous breaking scale; in the case of local strings, the lagrangian remains invariant under local transformations of the type $\phi(x) \longrightarrow e^{i\alpha(x)}\phi(x)$.

The study of the properties and dynamics of cosmic strings in the context of modified theories of gravity is crucial in the advent of powerful observatories, such as LISA, as it may allow us to constrain both Modified Gravity theories and Grand Unified theories.

2. Hybrid metric-Palatini gravity

In this hybrid regime, the metric and Palatini approaches are combined, by adding a new term, $f(\mathcal{R})$ to the Einstein-Hilbert action:

$$S = \frac{1}{2\kappa^2}\int \sqrt{-g}\; d^4x\;\left[R + f(\mathcal{R}) + \mathcal{L}_m\right]. \tag{1}$$

where $\mathcal{R}$ is the Ricci-Palatini scalar, the geometrical analogous to the Ricci scalar, constructed in terms of an independent connection, $\hat{\Gamma}$:

$$\mathcal{R}_{\mu\nu} = \partial_\alpha \hat{\Gamma}^\alpha_{\mu\nu} - \partial_\nu \hat{\Gamma}^\alpha_{\mu\alpha} + \hat{\Gamma}^\alpha_{\alpha\beta}\hat{\Gamma}^\beta_{\mu\nu} - \hat{\Gamma}^\alpha_{\mu\beta}\hat{\Gamma}^\beta_{\alpha\nu}. \tag{2}$$

$$\mathcal{R} = g^{\mu\nu}\mathcal{R}_{\mu\nu} \tag{3}$$

In this theory we have two dynamical variables, the metric and the (independent) connection. And so, we apply variation calculations of the action with respect to both.

Varying (1) with respect to the metric, $g_{\mu\nu}$ we obtain

$$G_{\mu\nu} + \frac{df}{d\mathcal{R}}\mathcal{R}_{\mu\nu} - \frac{1}{2}f(\mathcal{R})g_{\mu\nu} = \kappa^2 T_{\mu\nu} \tag{4}$$

where $T_{\mu\nu}$ is the energy-momentum tensor defined, as usual, as:

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\mathcal{L}_m)}{\delta(g^{\mu\nu})} \tag{5}$$

When we perform the variation of the action (1) with respect to the independent connection, $\hat{\Gamma}$, the resulting equation is:

$$\hat{\nabla}_\alpha = \left(\sqrt{-g}\frac{df}{d\mathcal{R}}g^{\mu\nu}\right) \tag{6}$$

which has an interesting feature: if we consider a metric conformally related to $g_{\mu\nu}$ by a $\frac{df}{d\mathcal{R}}$ factor, $h_{\mu\nu} = \frac{df}{d\mathcal{R}}g_{\mu\nu}$, equation (6) implies that the independent connection $\hat{\Gamma}$ is the Levi-Civita connection of such metric.

One of the useful features of the hybrid metric-Palatini theory is that it admits a scalar-tensor representation, which simplifies the analysis of the dynamics of the theory.

Let's introduce an auxiliary field, A, so that the action (1) becomes

$$S = \frac{1}{2\kappa^2}\int \sqrt{-g}\; d^4x \left[\Omega_A R + f(A) + \frac{df}{dA}(\mathcal{R} - A) + \mathcal{L}_m\right]. \tag{7}$$

where the coupling constant Ω_A is introduced for generality. If we further define $\phi \equiv \frac{df}{dA}$ and $V(\phi) \equiv A\frac{df}{dA} - f(A)$, the action becomes:

$$S = \frac{1}{2\kappa^2}\int \sqrt{-g}\; d^4x \left[\Omega_A R + \phi\mathcal{R} - V(\phi) + \mathcal{L}_m\right]\ . \tag{8}$$

Since we now have three dynamical variables, the metric $g_{\mu\nu}$, the independent connection, $\hat{\Gamma}$ and the scalar field ϕ, we perform the variation of the action (8) with respect to each of the dynamical variables, resulting in the following equations of motion:

$$\Omega_A R_{\mu\nu} + \phi\mathcal{R}_{\mu\nu} - \frac{1}{2}(\Omega_A R + \phi\mathcal{R} - V(\phi))g_{\mu\nu} = \kappa^2 T_{\mu\nu} \tag{9}$$

$$\mathcal{R} - V_\phi = 0 \tag{10}$$

$$\hat{\nabla}_\alpha = \left(\sqrt{-g}\phi g_{\mu\nu}\right) \tag{11}$$

As we've seen before, the last equation implies that the independent connection is the Levi-Civita connection to a conformal related metric to $g_{\mu\nu}$: $h_{\mu\nu} = \phi g_{\mu\nu}$,

allowing us to write the Ricci-Palatini tensor from the Ricci tensor and the conformal factor ϕ:

$$\mathcal{R}_{\mu\nu} = R_{\mu\nu} - \frac{1}{\phi}\left(\nabla_\mu\nabla_\nu\phi + \frac{1}{2}g_{\mu\nu}\Box\phi\right) + \frac{3}{2\phi^2}\partial_\mu\phi\partial_\nu\phi \tag{12}$$

$$R = \mathcal{R} + \frac{3}{\phi}\Box\phi - \frac{3}{2\phi^2}\partial_\mu\phi\partial^\mu\phi \tag{13}$$

where $\Box \equiv \nabla_\mu\nabla^\mu$ in the D'Alembert operator. Eq.(13) can be used to recast the action (8) into the following form:

$$S = \frac{1}{2\kappa^2}\int \sqrt{-g}\; d^4x \left[(\Omega_A + \phi)R + \frac{3}{2\phi}\partial_\mu\phi\partial^\mu\phi - V(\phi) + \mathcal{L}_m\right] . \tag{14}$$

We can note now three different cases for the value of the coupling constant Ω_A, for $\Omega_A = 1$ we have the original hybrid threory, for $\Omega_A = 0$ we have the Palatini-$f(\mathcal{R})$ gravity and for $\Omega_A \to \infty$ we recover the metric-$f(R)$ theory.

Using eqs.(2) and (10), we can rewrite eq.(9) in the scalar-tensor representation for the original hybrid theory as:

$$(1+\phi)R_{\mu\nu} = \kappa^2\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right) + \frac{1}{2}g_{\mu\nu}(V + \Box\phi) + \nabla_\mu\nabla_\nu\phi - \frac{3}{2\phi}\partial_\mu\phi\partial_\nu\phi \tag{15}$$

In this equation we can conclude that the curvature of spacetime is due not only to the matter distribution, but also to the presence of the scalar field.

Taking the trace of equation (9) with $g^{\mu\nu}$, and using equation (10), we get

$$2V - \phi V_\phi = \kappa^2 T + \Omega_A R \tag{16}$$

If we now rewrite eq. (16) using eqs. (13) and (10), we get an effective Klein-Gordon equation for the scalar field:

$$-\Box\phi + \frac{1}{2\phi}\partial_\mu\phi\partial^\mu\phi + \frac{\phi\left[2V - (1+\phi)V_\phi\right]}{3} = \frac{\phi\kappa^2}{3}T \tag{17}$$

Which demonstrates that, unlike the purely Palatini case, the scalar field is dynamic and the theory is therefore not affected by the instabilities found in the Palatini gravity.[8]

3. Dynamical equations of local strings in Hybrid metric-palatini using Vilenkin's approximation

For the study of local $U(1)$ string in the context of the hybrid metric-Palatini gravity, we will start by defining the energy-momentum tensor of a straight, infinite cosmic string along the z-direction. Following Vilenkin's prescription,[9] one possible energy-momentum tensor is

$$T^r_r = T^t_t = -\sigma(r) \tag{18}$$

where $\sigma(r)$ is the linear energy density of the string.

We consider a general cylindrically symmetric static metric

$$ds^2 = -e^{2(K-U)}dt^2 + e^{2(K-U)}dr^2 + e^{-2U}W^2 d\theta^2 + e^{2U}dz^2 \tag{19}$$

Where t, r, θ and z denote the time, radial, angular and axial cylindrical coordinates respectively, and K, U and W are functions of r alone.

Inserting the metric (19) into the gravitational field equation (15) provides the following components

$$(1+\phi)\left(-U'^2 + K'\frac{W'}{W} - \frac{W''}{W}\right) = \phi'' - \frac{3}{4\phi}\phi'^2 - \left(K' - U' - \frac{W'}{W}\right)\phi' + \left(\kappa^2\sigma + \frac{1}{2}V\right)e^{2(K-U)}\,, \tag{20}$$

$$(1+\phi)\left(-U'^2 + K'\frac{W'}{W}\right) = -\frac{3}{4\phi}\phi'^2 - \left(K' - U' + \frac{W'}{W}\right)\phi' - \frac{1}{2}Ve^{2(K-U)}\,, \tag{21}$$

$$(1+\phi)\left(U'^2 + K''\right) = -\phi'' + \frac{3}{4\phi}\phi'^2 - U'\phi' - \frac{1}{2}Ve^{2(K-U)}\,, \tag{22}$$

$$(1+\phi)\left(U'^2 + K'' - 2U'' - 2U'\frac{W'}{W} + \frac{W''}{W}\right) = -\phi'' + \frac{3}{4\phi}\phi'^2 + \left(U' - \frac{W'}{W}\right)\phi' - \left(\kappa^2\sigma + \frac{1}{2}V\right)e^{2(K-U)}. \tag{23}$$

Additionally, we can use Eq. (17) to determine the effective Klein-Gordon equation for the scalar field ϕ:

$$e^{-2(K-U)}\left(-\phi'' - \frac{W'}{W}\phi' + \frac{\phi'^2}{2\phi}\right) + \frac{\phi}{3}\left[2V - (\phi+1)V_{,\phi}\right]\frac{2\phi\kappa^2\sigma}{3} = 0\,. \tag{24}$$

Since in this model the matter field couples minimally with curvature, it's possible to show that the energy conservation equation still holds, i.e.,

$$\nabla_\mu T^\mu{}_\nu = 0 \tag{25}$$

which yields $K'\sigma = 0$, and apart from the trivial vacuum solution, $\sigma = 0$, this implies that $K' = 0$. Thus, we consider from now on that $e^K = 1$, so that Eqs. (20)–(23) simplify to the following relations

$$(1+\phi)\left(-U'^2 - \frac{W''}{W}\right) = \phi'' - \frac{3}{4\phi}\phi'^2 + \left(U' + \frac{W'}{W}\right)\phi' + \left(\kappa^2\sigma + \frac{1}{2}V\right)e^{-2U}\,, \tag{26}$$

$$(1+\phi)U'^2 = \frac{3}{4\phi}\phi'^2 + \left(-U' + \frac{W'}{W}\right)\phi' + \frac{1}{2}Ve^{-2U}\,, \tag{27}$$

$$(1+\phi)U'^2 = -\phi'' + \frac{3}{4\phi}\phi'^2 - U'\phi' - \frac{1}{2}Ve^{-2U}\,, \tag{28}$$

$$(1+\phi)\left(U'^2-2U''-2U'\frac{W'}{W}+\frac{W''}{W}\right)=-\phi''+\frac{3}{4\phi}\phi'^2 + \left(U'-\frac{W'}{W}\right)\phi'-\left(\kappa^2\sigma+\frac{1}{2}V\right)e^{-2U}\,, \tag{29}$$

With the effective Klein-Gordon equation for the scalar field ϕ reducing to

$$e^{2U}\left(-\phi''-\frac{W'}{W}\phi'+\frac{\phi'^2}{2\phi}\right)+\frac{\phi}{3}\left[2V-(\phi+1)V_{,\phi}\right]+\frac{2\phi\kappa^2\sigma}{3}=0\,. \tag{30}$$

If we further consider that local gauge strings preserve boost invariance along the t and z directions,[9] meaning, in this case, that $U=0$, the gravitational field equations simplify considerably:

$$(1+\phi)\left(-\frac{W''}{W}\right)=\phi''-\frac{3}{4\phi}\phi'^2+\frac{W'}{W}\phi'+\kappa^2\sigma+\frac{1}{2}V\,, \tag{31}$$

$$0=\frac{3}{4\phi}\phi'^2+\frac{W'}{W}\phi'+\frac{1}{2}V\,, \tag{32}$$

$$0=-\phi''+\frac{3}{4\phi}\phi'^2-\frac{1}{2}V\,, \tag{33}$$

$$(1+\phi)\frac{W''}{W}=-\phi''+\frac{3}{4\phi}\phi'^2-\frac{W'}{W}\phi'-\kappa^2\sigma-\frac{1}{2}V\,. \tag{34}$$

where we can see now that Eqs. (31) and (34) become redundant. Combining Eqs. (32) and (33) yields the following relation for the potential V:

$$V=-\phi''-\frac{W'}{W}\phi'\,, \tag{35}$$

substituting into the Klein-Gordon equation (30), the latter reduces to:

$$V\,(3+2\phi)-V_\phi\phi\,(\phi+1)+2\kappa^2\sigma\phi+\frac{3\phi'^2}{2\phi}=0\,. \tag{36}$$

Additionally, we can further deduce:

$$\kappa^2\sigma=\frac{1}{W}\left[(1+\phi)W'\right]'\,, \tag{37}$$

and

$$\frac{\left[(1+\phi)W\right]''}{W}=-(V+\kappa^2\sigma)\,. \tag{38}$$

An important physical parameter characterizing the cosmic string properties is the mass per unit length of the string, which is defined as

$$\begin{aligned} m(r)&=\int_0^{2\pi}d\theta\int_0^{R_s}\sigma(r)W(r)dr \\ &=2\pi\int_0^{R_s}\sigma(r)W(r)dr, \end{aligned} \tag{39}$$

where R_s is the string radius.

The set of the previous equations allows us to write the gravitational equations of a cosmic string in hybrid metric-Palatini gravity in an exact(closed) form, with all the geometric and physical quantities expressed in a parametric form, in ϕ taken as the parameter.

By taking into account Eq. (33), the field equations (31) and (34) reduce to the form

$$(1+\phi)\frac{W''}{W} = -\frac{W'}{W}\phi' - \kappa^2\sigma. \tag{40}$$

3.1. *Parametric form of the dynamical equations*

From a mathematical point of view Equation (33) is independent of the metric tensor coefficient W and it represents a second order nonlinear differential equation. In order to solve it we first rescale the radial coordinate r according to the transformation $r = \beta\xi$. Hence Eq. (33) takes the form

$$\frac{d^2\phi}{d\xi^2} - \frac{3}{4\phi}\left(\frac{d\phi}{d\xi}\right)^2 + \frac{1}{2}\beta^2 V(\phi) = 0. \tag{41}$$

In order to solve Eq. (41) we introduce the transformations

$$\frac{d\phi}{d\xi} = u, \quad \frac{d^2\phi}{d\xi^2} = \frac{du}{d\xi} = \frac{du}{d\phi}\frac{d\phi}{d\xi} = u\frac{du}{d\phi} = \frac{1}{2}\frac{d}{d\phi}u^2, \tag{42}$$

and

$$u^2 = v, \tag{43}$$

respectively. Then Eq. (41) becomes a first order linear differential equation of the form

$$\frac{dv}{d\phi} - \frac{3}{2\phi}v + \beta^2 V(\phi) = 0, \tag{44}$$

with the general solution given by

$$v(\phi) = \phi^{3/2}\left[C - \beta^2\int \phi^{-3/2}V(\phi)d\phi\right], \tag{45}$$

where C is an arbitrary constant of integration. We immediately obtain

$$u(\phi) = \phi^{3/4}\sqrt{\left[C - \beta^2\int \phi^{-3/2}V(\phi)d\phi\right]}, \tag{46}$$

and

$$\xi + C_0 = \int \frac{\phi^{-3/4}d\phi}{\sqrt{\left[C - \beta^2\int \phi^{-3/2}V(\phi)d\phi\right]}}, \tag{47}$$

respectively, where C_0 is an arbitrary constant of integration.

Equation (32) can be successively transformed as

$$\frac{1}{W}\frac{dW}{d\xi}\frac{d\phi}{d\xi}=-\frac{3}{4\phi}\left(\frac{d\phi}{d\xi}\right)^2-\frac{\beta^2}{2}V(\phi), \tag{48}$$

and

$$\begin{aligned}\frac{1}{W}\frac{dW}{d\phi}=&-\frac{3}{4\phi}-\frac{\beta^2}{2}\frac{\phi^{-3/2}V(\phi)}{\left[C-\beta^2\int\phi^{-3/2}V(\phi)d\phi\right]}\\ =&-\frac{3}{4\phi}+\frac{1}{2}\frac{d}{d\phi}\ln\left[C-\beta^2\int\phi^{-3/2}V(\phi)d\phi\right],\end{aligned} \tag{49}$$

yielding

$$W(\phi)=W_0\phi^{-3/4}\sqrt{C-\beta^2\int\phi^{-3/2}V(\phi)d\phi}, \tag{50}$$

where W_0 is an arbitrary constant of integration.

As a last step we need to obtain the expression of σ. With the use of Eq. (32), then Eq. (40) can be rewritten as

$$(1+\phi)\frac{1}{W}\frac{d^2W}{d\xi^2}=\frac{3}{4\phi}\left(\frac{d\phi}{d\xi}\right)^2+\frac{1}{2}\beta^2V(\phi)-\beta^2\kappa^2\sigma. \tag{51}$$

With the use of the mathematical identities

$$\frac{dW}{d\xi}=\frac{dW}{d\phi}\frac{d\phi}{d\xi}=\frac{dW}{d\phi}u, \tag{52}$$

$$\frac{d^2W}{d\xi^2}=\frac{d^2W}{d\phi^2}v+\frac{1}{2}\frac{dW}{d\phi}\frac{dv}{d\phi}, \tag{53}$$

Eq. (51) takes the form

$$(1+\phi)\left(\frac{1}{W}\frac{d^2W}{d\phi^2}v+\frac{1}{2}\frac{1}{W}\frac{dW}{d\phi}\frac{dv}{d\phi}\right)=\frac{3}{4\phi}v+\frac{1}{2}\beta^2V(\phi)-\beta^2\kappa^2\sigma. \tag{54}$$

Finally, after some simple calculations we obtain

$$\begin{aligned}\kappa^2\sigma(\phi)=&\frac{1}{4\phi}\left\{\left[2\phi(\phi+1)V'(\phi)+3\sqrt{\phi}\int\frac{V(\phi)}{\phi^{3/2}}\,d\phi\right.\right.\\ &\left.\left.-2(2\phi+3)V(\phi)\right]-3\left(C/\beta^2\right)\sqrt{\phi}\right\}.\end{aligned} \tag{55}$$

Equation (47), (50) and (55) give the complete solution of the field equations describing the geometry of a cosmic string in hybrid metric-Palatini gravity. The solution is obtained in a parametric form, with ϕ taken as a parameter. It also contains three arbitrary integration constants ξ_0, C, and W_0 which must be obtained from the initial or boundary conditions imposed on the cosmic string configuration.

As for the mass of the string, in the dimensionless variable ξ it can be obtained as

$$m(\xi) = 2\pi\beta \int_0^{\xi_s} \sigma(\xi) W(\xi) d\xi, \tag{56}$$

where $\xi_s = R_s/\beta$.

In order to study specific cosmic string solutions in the hybrid metric-Palatini framework, we need now to specify the form of the potential $V(\phi)$.

4. Solutions to the dynamical equations with specific potentials

In this section we will investigate the application of the set of parametric equations deduced on the previous section to different potential configurations, for a more complete set of possible potential configurations, we refer the reader to the original article Ref. 10.

4.1. *Power law potential*

The gravitational field equations describing a cosmic string in hybrid metric-Palatini gravity admit an exact solution, corresponding to the power law type scalar field potential $V(\phi) = V_0\phi^{3/4}$. We rescale the radial coordinate r by imposing the condition $\beta^2 V_0 = 1$, which gives $r = \xi/\sqrt{V_0}$. With these choices from Eq. (47) we obtain explicitly the scalar field as a function of ξ, given by

$$\phi(\xi) = \frac{\left(\xi^2 {\phi_0}^{3/4} \pm 2\xi\phi_0' \pm 8\phi_0\right)^4}{4096{\phi_0}^3}, \tag{57}$$

where we have used the usual initial conditions $\phi(0) = \phi_0$ and $\phi'(0) = \phi_0'$, respectively. For the metric tensor component W we obtain

$$W(\xi) = \frac{W_0}{\left(\xi^2 {\phi_0}^{3/4} \pm 2\xi\phi_0' \pm 8\phi_0\right)^3 \sqrt{2\xi \pm 2\phi_0'/{\phi_0}^{3/4}}}. \tag{58}$$

where W_0 is an arbitrary constant of integration. On the string axis, i.e., $\xi = 0$, we obtain $W^2(0) = \pm {W_0}^2/524288\,{\phi_0}^{21/4}\phi_0'$. Since the metric tensor component W^2 must be positive for all $\xi \geq 0$, it follows that the physical solution for the string configuration is the one with the positive sign. Hence in the case of the $V(\phi) = V_0\phi^{3/4}$ potential, the solutions of the field equations describing a cosmic string in hybrid metric-Palatini gravity are

$$\begin{aligned} \phi(\xi) &= \frac{\left(\xi^2 {\phi_0}^{3/4} + 2\xi\phi_0' + 8\phi_0\right)^4}{4096{\phi_0}^3}, \\ W^2(\xi) &= \frac{{W_0}^2}{\left(\xi^2 {\phi_0}^{3/4} + 2\xi\phi_0' + 8\phi_0\right)^6 \left(2\xi + 2\phi_0'/{\phi_0}^{3/4}\right)}, \end{aligned} \tag{59}$$

respectively, with ${W_0}^2 = 524288\, W^2(0){\phi_0}^{21/4}\phi_0'$, a condition that implies $\phi_0 > 0$ and $\phi_0' > 0$. For the string tension as a function of ϕ and ξ we obtain the expressions

$$\kappa^2\sigma(\phi) = V_0\frac{-6C - 5(\phi-3)\sqrt[4]{\phi}}{8\sqrt{\phi}}, \tag{60}$$

and

$$\kappa^2\sigma(\xi) = \frac{V_0\phi_0^{3/4}}{\left(\xi^2\phi_0^{3/4}+2\xi\phi_0'+8\phi_0\right)^2}\left\{-48C\phi_0^{3/4} - 5\left(\xi^2\phi_0^{3/4}+2\xi\phi_0'+8\phi_0\right)\times \left[\frac{\left(\xi^2\phi_0^{3/4}+2\xi\phi_0'+8\phi_0\right)^4}{4096\phi_0^3} - 3\right]\right\}, \tag{61}$$

respectively.

In this case, the scalar field is a monotonically increasing function of the radial distance from the string axis and tends to infinity for $\xi \to \infty$. On the other hand, the metric tensor component decreases monotonically from a finite value on the string axis to zero at infinity. For $\xi = 0$, the string tension takes the finite value

$$\sigma(0) = V_0\left[-48C\phi_0^{3/4} - 40\left(\phi_0 - 3\right)\phi_0\right]/64\phi_0^{5/4} \tag{62}$$

while $\lim_{\xi\to\infty}\sigma(\xi) = -\infty$, indicating that σ is a monotonically decreasing function of the radial coordinate. In the first order of approximation we obtain for the mass of the string of radius ξ_s the expression

$$m = \frac{\pi\beta W_0\xi_s}{8192\phi_0^{33/8}{\phi_0'}^{3/2}}\left\{6C\xi_s\phi_0^{7/4} + 15\xi_s{\phi_0'}^2\left(C - 2\sqrt[4]{\phi_0}\right) - 4\phi_0\phi_0'\left[6C + 5\left(\phi_0 - 3\right)\sqrt[4]{\phi_0}\right] + 5\xi_s\left(\phi_0 - 3\right)\phi_0^2\right\}. \tag{63}$$

In this approximation the mass monotonically increases with the string radius.

4.2. *The Higgs-type potential*

Next we consider the case where the scalar field potential is of the Higgs type, given by

$$V(\phi) = \pm\frac{\bar{\mu}^2}{2}\phi^2 + \frac{\nu}{4}\phi^4, \tag{64}$$

where $\bar{\mu}^2$ and ν are constants. In the following we will investigate only the case with $\bar{\mu}^2 < 0$, meaning we will adopt the minus sign in the definition of the potential. By following the standard approach in elementary particle physics, we assume that the constant $\bar{\mu}^2$ is related to the mass of the scalar field particle as $m_\phi^2 = 2\xi v^2 = 2\bar{\mu}^2$, where $v^2 = \bar{\mu}^2/\xi$ gives the minimum value of the potential. The Higgs self-coupling constant ν can be obtained, in the case of strong interactions, from the determination of the mass of the Higgs boson in laboratory experiments, and its

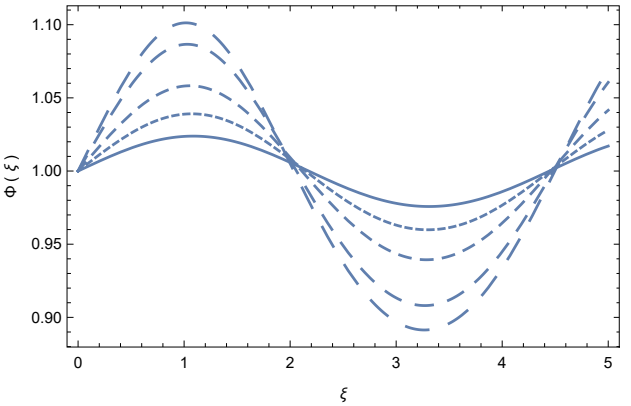


Fig. 1. Variation of the scalar field for the cosmic string configuration in the presence of a Higgs-type potential $V(\Phi) = -\Phi^2 + \Phi^4$ for $\Phi(0) = \Phi_0 = 1$, and for different values of Φ_0': $\Phi_0' = 0.034$ (solid curve), $\Phi_0' = 0.056$ (dotted curve), $\Phi_0' = 0.084$ (short dashed curve), $\Phi_0' = 0.126$ (dashed curve), and $\Phi_0' = 0.148$ (long dashed curve), respectively.

numerical value is of the order of $\nu \approx 1/8$.[11] By rescaling the radial coordinate and the scalar field according to

$$r = \sqrt{2}\bar{\mu}\xi, \qquad \phi = \frac{\Phi}{(\nu\bar{\mu})^{1/3}}, \tag{65}$$

then Eq. (47) provides the profile of the scalar field in the following form

$$\frac{d^2\Phi}{d\xi^2} - \frac{3}{4\phi}\left(\frac{d\Phi}{d\xi}\right)^2 - \Phi^2 + \Phi^4 = 0. \tag{66}$$

The general solution for this equation is given in a closed form by

$$\xi + C_0 = \int \frac{1}{\Phi^{3/4}\sqrt{C + \frac{2}{21}\left(7 - 3\Phi^2\right)\Phi^{3/2}}}\, d\Phi. \tag{67}$$

However, this solution cannot be expressed in an analytical form in terms of known functions. In the first order approximation, we obtain

$$\xi + C_0 \approx \frac{4\sqrt[4]{\Phi}}{\sqrt{C}} - \frac{4\Phi^{7/4}}{21C^{3/2}} + O\left(\Phi^{9/4}\right), \tag{68}$$

but this representation is not particularly useful from the point of view of concrete calculations.

The variation of the scalar field with Higgs potential supporting a string configuration in hybrid metric-Palatini gravity is represented in Fig. 1. There is a significant qualitative difference between this string model and the constant, simple power law or exponential potentials. Note that the scalar field for the Higgs-type potential shows a basically periodic structure, changing between successive maxima and minima. There are singularities in the field. Its behavior is strongly affected by the initial conditions on the string axis, and the field extends to infinity.

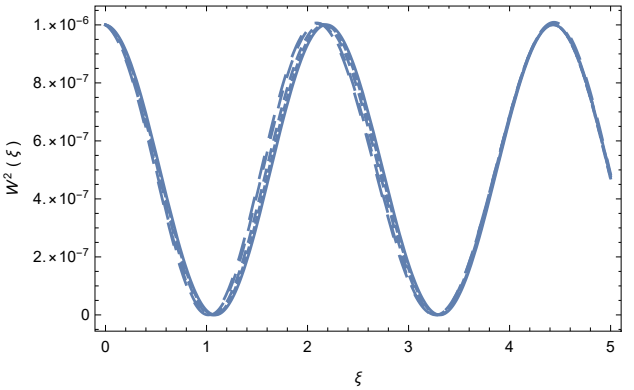


Fig. 2. Variation of the metric tensor component $W^2(\xi)$ of the cosmic string configuration in the presence of a Higgs type potential $V(\Phi) = -\Phi^2 + \Phi^4$ for $\Phi(0) = 1$, $W(0) = 10^{-3}$, and for different values of Φ_0': $\Phi_0' = 0.034$ (solid curve), $\Phi_0' = 0.056$ (dotted curve), $\Phi_0' = 0.084$ (short dashed curve), $\Phi_0' = 0.126$ (dashed curve), and $\Phi_0' = 0.148$ (long dashed curve), respectively.

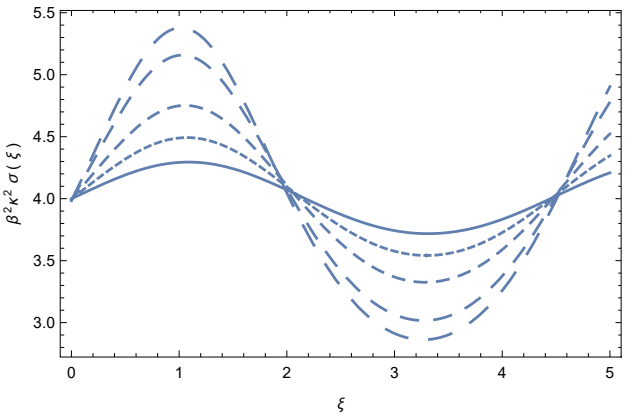


Fig. 3. Variation of the string tension $\kappa^2\sigma(\xi)$ of the cosmic string configuration in the presence of a Higgs type potential $V(\Phi) = -\Phi^2 + \Phi^4$ for $\Phi(0) = 1$, $W(0) = 10^{-3}$, and for different values of Φ_0': $\Phi_0' = 0.034$ (solid curve), $\Phi_0' = 0.056$ (dotted curve), $\Phi_0' = 0.084$ (short dashed curve), $\Phi_0' = 0.126$ (dashed curve), and $\Phi_0' = 0.148$ (long dashed curve), respectively.

The variation of the metric tensor component $W^2(\xi)$ in the presence of a Higgs potential is represented in Fig. 2. The same oscillatory pattern can also be observed in the case of the metric tensor component W^2. However, there is a difference in the phase of these to quantities. When the field reaches its maximum at $\xi \approx 1$, the metric tends to zero, $W^2(1) \approx 0$. Then, while the scalar field decreases, the metric tensor increases, reaching its maximum at the minimum of the field, corresponding to $\xi \approx 2$. This pattern is repeated up to infinity.

The variation of the string tension with respect to the radial coordinate in the presence of the Higgs potential is depicted in Fig. 3. The variation of σ is in

phase with that of the scalar field, and both quantities reach their maxima and minima at the same position. The string tension also has an oscillatory behaviour, which is a general property of all physical and geometrical parameters of the string configurations supported by scalar fields with a Higgs-type potential.

5. Conclusions

In this work we studied the existence and physical properties of cosmic strings in the context of the hybrid metric-Palatini gravity. The theory is an extension to General Relativity, combining both metric and Palatini formalism. A main success of the theory is the possibility to generate long-range forces that pass the classical local tests of gravity at the Solar System level, thus avoiding some problematic features of the standard $f(R)$ theories. Another interesting advantage of the theory is that it admits an equivalent scalar-tensor representation, simplifying greatly the dynamical equations. The type of strings studied in this work are local gauge strings, using an approximation to the Vilenkin-prescribed energy-momentum tensor and different potential configurations. For the case of the potential of the form $V = V_0\phi^{3/4}$. The requirement of the positivity of the metric tensor for $r = 0$ imposes the condition that the metric tensor and the string tension are decreasing functions of r, since both ϕ_0 and ϕ_0' are required to be positive. In this case, the scalar field becomes singular at infinity. However, it is possible to construct finite string configurations: defining a string radius by introducing an effective cuttoff length, ξ_{co} for both the metric and the scalar field. This radius would allow us to get finite values for the scalar field, the metric tensor components and the string tension. For this type of potential the choice of ξ_{co} should be made based on empirical considerations, such as consistency with observational data. The case of the Higgs-type potential is quite different, since the string tension does not vanish for any value of the radial coordinate but it does reaches its minimum value at $\xi \approx 3.5$, where the scalar field is also at its minimum and $W^2(\xi)$ is singular and tends to zero. Another possible choice for the string radius could be $\xi \approx 1$, the first zero of the metric tensor component; in this case both the string tension and the scalar field are at their maxima. Or we could, as in the exponential potential, introduce a cutoff radius to be determined by confrontation with observations, matching the solution with the cosmological background or another GR string solution, thus achieving a well behaved structure throughout the radius of the string.

Acknowledgments

F. S. N. L. acknowledges support from the Fundação para a Ciência e a Tecnologia Scientific Employment Stimulus contract with reference CEECIND/04057/2017 and the research Grants No. UID/FIS/04434/2019 and No. PTDC/FIS-OUT/29048/2017.

References

1. A. G. Riess *et al.* [Supernova Search Team], "Observational evidence from supernovae for an accelerating universe and a cosmological constant," Astron. J. **116** (1998), 1009-1038 doi:10.1086/300499 [arXiv:astro-ph/9805201 [astro-ph]].
2. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, "Modified Gravity and Cosmology," Phys. Rept. **513** (2012), 1-189 doi:10.1016/j.physrep.2012.01.001 [arXiv:1106.2476 [astro-ph.CO]].
3. R. Utiyama and B. S. DeWitt, "Renormalization of a classical gravitational field interacting with quantized matter fields," J. Math. Phys. **3** (1962), 608-618 doi:10.1063/1.1724264
4. D. Lovelock, J. Math. Phys. **12** (1971), 498-501 doi:10.1063/1.1665613
5. T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, "Metric-Palatini gravity unifying local constraints and late-time cosmic acceleration," Phys. Rev. D **85** (2012), 084016 doi:10.1103/PhysRevD.85.084016 [arXiv:1110.1049 [gr-qc]].
6. T. Harko and F. S. N. Lobo, "Beyond Einstein's General Relativity: Hybrid metric-Palatini gravity and curvature-matter couplings," Int. J. Mod. Phys. D **29** (2020) no.13, 2030008 doi:10.1142/S0218271820300086 [arXiv:2007.15345 [gr-qc]].
7. A. Vilenkin and E. P. S. Shellard, "Cosmic Strings and Other Topological Defects,"
8. G. J. Olmo, "Palatini Approach to Modified Gravity: f(R) Theories and Beyond," Int. J. Mod. Phys. D **20** (2011), 413-462 doi:10.1142/S0218271811018925 [arXiv:1101.3864 [gr-qc]].
9. A. Vilenkin, "Gravitational Field of Vacuum Domain Walls and Strings," Phys. Rev. D **23** (1981), 852-857 doi:10.1103/PhysRevD.23.852
10. T. Harko, F. S. N. Lobo and H. M. R. da Silva, Phys. Rev. D **101** (2020) no.12, 124050 doi:10.1103/PhysRevD.101.124050 [arXiv:2003.09751 [gr-qc]].
11. G. Aad *et al.* [ATLAS and CMS], Phys. Rev. Lett. **114** (2015), 191803 doi:10.1103/PhysRevLett.114.191803 [arXiv:1503.07589 [hep-ex]].

Inflationary supersymmetric FRLW quantum cosmology

N.E. Martínez-Pérez* and C. Ramirez†

Facultad de Ciencias Físico-Matemt́icas, Benemérita Universidad Autónoma de Puebla, Puebla, México P.O. Box 165, 72000

** E-mail: nephtali.martinezper@alumno.buap.mx*

† E-mail: cramirez@fcfm.buap.mx

V.M. Vázquez-Báez

Facultad de Ingeniería, Benemérita Universidad Autónoma de Puebla, Puebla, México 72000

E-mail: manuel.vazquez@correo.buap.mx

We consider inflationary scenarios of the supersymmetric quantum cosmology of FRLW models with a scalar field. We use the superfield formalism with a superpotential for the scalar superfield. The probability amplitude solution of the supersymmetric Wheeler-DeWitt equation, gives a probability density from which we can compute mean trajectories that can be parametrized by the scalar. By suitable choices of the superpotential, the resulting evolutions of the scale factor correspond to consistent inflationary scenarios. We show the acceleration, the resulting e-folds and the horizon for several superpotentials.

Keywords: FRW metric; supersymmetry; quantum cosmology; inflation.

1. Introduction

We can think of the universe as originating from a homogeneous early phase, that lies presumably in the Planck scale, based on the numerous observations available nowadays from worldwide experiments and satellite missions which tell us about of the past convergnece of matter, the homogeneity and isotropy of the CMB, light elements abundance, the distribution and content of large scale structures, etc. To attack directly this early phase one would desire to be armed with a theory of the quantum spacetime, which we lack today. Nonetheless at classical level we have descriptions of the early universe with a FRLW metric, see e.g.,[1] which can be allowed to incorporate inhomogeneity perturbations of the metric and scalar fields, which upon quantization are believed to seed the structure we observe today.[2] However, this picture does not account for the very known problems we are confronted by the observations, such as the horizon problem, the homogeneity of the CMB and the flatness problem. The most accepted solution to these problems is inflation, see e.g.,[3] a mechanism that produces enough fluctuation growth as well as the suitable causal relations on the universe, from which, by means of a semiclassical quantum gravity treatment, one can obtain the seeds of structure at the ending of the homogeneous phase, making natural to think that the early homogenous phase should be treated by quantum mechanics. In this line of thought, quantum cosmology[4] gives

a canonical quantization of cosmology, resulting in a time independent Schrödinger equation for the wave function of the universe, the Wheeler-deWitt equation. Given that the hamiltonian operator acting on the wave function vanishes, this is a timeless theory, but time can be reinstated in some way for the subsequent classical universe.[5–7]

In,[8] we are given a way to obtain a time, where a scalar field is chosen to be time, and a time dependent effective wave function is defined by the conditional probability of measuring a value of the scale factor for a given value of the scalar field, this allows to compute mean values (dependent on time) from the wave function. On the other hand, the knowledge of the fermionic and bosonic constituents of matter, as well as the required symmetry breaking, has led to the search and construction of unified theories, including fermions and bosons related and interacting in a non trivial way. Supersymmetry,[9] lies well in the category of such theories and it has given the way to supergravity and string theory. In this view, supersymmetric quantum cosmology[10–12] is a good option to study the early universe and attains its goals by the extension of spacetime diffeomorphisms to diffeomorphisms in a Grassmann, extended spacetime, called superspace. This supersymmetry-superspace has additional fermionic coordinates, and the fields become superfields, power expansions on the grassmann variables. Thus, supergravity is formulated as general relativity theory on a supermanifold.[9,18] There are several formulations for supersymmetric extensions of homogeneous models.[11,12] The type of formulation we consider here, is obtained by the extension of time general reparametrizations, to general reparametrizations on a "time superspace".[13,14] Another option is dimensional reduction which involves fields depending only on time, and the spatial coordinates are integrated out.[10] In the simple cases this system has exact solutions,[13–15] and the wave functions have several independent components. In[16] a semiclassical WKB analysis has been made, from which follow a classical hamiltonian, and the corresponding classical equations of motion, for the two relevant spinorial components of the wave function.

In this work we consider the supersymmetric quantum cosmology model of,[8] to explore inflationary scenarios. In sections 2 and 3, we review the supersymmetric model and its component formulation. In section 4 we review the quantization of the model from.[8] In section 5 we make a discussion of the problem of time. Following,[8] a time dependent, effective wave function, can be given. We obtain inflationary behavior from suitable exponential superpotentials, as shown in section 6. We illustrate the inflationary behavior for two types of decaying superpotentials. Finally, in section 7, we make a short discussion with some remarks and future work perspectives.

2. Supersymmetric FRLW model with a scalar field

We begin with a model consisting of a FRLW metric with a scalar field, a setting that can account for inflation, primordial matter generation and structure formation, and

possibly dark energy, thus our model is

$$I=\frac{1}{\kappa^2}\int\left\{-\frac{3}{c^2}N^{-1}a\dot{a}^2+3Nka-Na^3\Lambda+\kappa^2a^3\left[\frac{1}{2c^2}N^{-1}\dot{\phi}^2-NV\left(\phi\right)\right]\right\}dt, \quad (1)$$

where $\kappa^2=\frac{8\pi G}{c^4}$.

2.1. *Supersymmetric cosmology*

Supersymmetric cosmology can be obtained from one dimensional supergravity.[13] Here we review the derivation of the supersymmetric Wheeler-DeWitt equation following,[8,14] with several minor modifications. In particular we introduce SI dimensional constants into the action and the ensuing equations, and correct typos. In these works, we have formulated it as general relativity on supersymmetry-superspace, $t\rightarrow z^M=(t,\Theta,\bar{\Theta})$, where Θ and $\bar{\Theta}$ are anticommuting coordinates. Hence, under $z^M\rightarrow z'^M=z^M+\zeta^M(z)$, the superfields, see e.g.,[9,18] transform as $\delta_\zeta\Phi(z)=-\zeta^M(z)\partial_M\Phi(z)$, and their covariant derivatives are $\nabla_A\Phi=\nabla_A^{\ M}(z)\partial_M\Phi$. $\nabla_A^{\ M}(z)$ is the superspace vielbein, whose superdeterminant gives the invariant superdensity $\mathcal{E}=\mathrm{Sdet}\nabla_M^{\ A}$, $\delta_\zeta\mathcal{E}=(-1)^m\partial_M(\zeta^M\mathcal{E})$. For the supersymmetric extension of the FRLW metric, in the covariant Wess-Zumino gauge,[18] we have $\mathcal{E}=-e-\frac{i}{2}(\Theta\bar{\Psi}+\bar{\Theta}\Psi)$.[14] In this formulation, to the scale factor and the scalar field correspond real scalar superfields[9,13]:

$$\mathcal{A}\left(t,\Theta,\bar{\Theta}\right)=a\left(t\right)+\Theta\lambda\left(t\right)-\bar{\Theta}\bar{\lambda}\left(t\right)+\Theta\bar{\Theta}B\left(t\right), \quad (2)$$

$$\Phi\left(t,\Theta,\bar{\Theta}\right)=\phi\left(t\right)+\Theta\eta\left(t\right)-\bar{\Theta}\bar{\eta}(t)+\Theta\bar{\Theta}G\left(t\right). \quad (3)$$

Thus the superfield extension of the model (1), for $k=0,1$, is $I=I_G+I_M$, where I_G is the supergravity action and I_M is the matter term[8,13,21]

$$I_G=\frac{3}{\kappa^2}\int\mathcal{E}\left(\mathcal{A}\nabla_{\bar{\Theta}}\mathcal{A}\nabla_\Theta\mathcal{A}-\sqrt{k}\mathcal{A}^2\right)d\Theta d\bar{\Theta}dt, \quad (4)$$

$$I_M=\int\mathcal{E}\mathcal{A}^3\left[-\frac{1}{2}\nabla_{\bar{\Theta}}\Phi\nabla_\Theta\Phi+W\left(\Phi\right)\right]d\Theta d\bar{\Theta}dt. \quad (5)$$

W is the superpotential.

3. Component formulation

Upon component expansion of the superfields in (4) and (5), once we have performed the integration of the ferminic coordinates, we have[8]

$$L_G=\frac{1}{\kappa^2}\left[-\frac{3a}{c^2e}\dot{a}^2+\frac{3a}{ce}\dot{a}\left(\psi\lambda-\bar{\psi}\bar{\lambda}\right)+\frac{3i}{c}a\left(\lambda\dot{\bar{\lambda}}+\bar{\lambda}\dot{\lambda}\right)+eaB\left(6\sqrt{k}-3B\right)\right.$$
$$\left.+e\left(6\sqrt{k}-3B\right)\lambda\bar{\lambda}+ia\left(3\sqrt{k}-\sqrt{3\Lambda}a\right)\left(\bar{\psi}\bar{\lambda}+\psi\lambda\right)-\frac{3a}{2e}\lambda\bar{\lambda}\psi\bar{\psi}\right], \quad (6)$$

$$
\begin{aligned}
L_M = & \frac{a^3}{2c^2e}\dot{\phi}^2 + \frac{a^3}{2ce}\dot{\phi}\left(\bar{\psi}\bar{\eta} - \psi\eta\right) + \frac{3ia^2}{2c}\dot{\phi}\left(\lambda\bar{\eta} + \bar{\lambda}\eta\right) \\
& - \frac{ia^3}{2c}\left(\eta\dot{\bar{\eta}} + \bar{\eta}\dot{\eta}\right) - 3ea^2BW - ea^3GW' + \frac{ea^3}{2}G^2 \\
& - 6eaW\lambda\bar{\lambda} - ea^2\left(\frac{3B}{2} + aW''\right)\eta\bar{\eta} - \frac{3ia^2}{2}W\left(\psi\lambda + \bar{\psi}\bar{\lambda}\right) \\
& + 3ea^2\left(\frac{1}{2}G - W'\right)\left(\lambda\bar{\eta} - \bar{\lambda}\eta\right) - \frac{ia^3}{2}W'\left(\psi\eta + \bar{\psi}\bar{\eta}\right) \\
& + \frac{a^3}{4e}\eta\bar{\eta}\psi\bar{\psi} - 3ea\eta\bar{\eta}\lambda\bar{\lambda},
\end{aligned}
\tag{7}
$$

where $W \equiv W(\phi)$, $W' \equiv \partial_\phi W(\phi)$, and $W'' \equiv \partial^2_\phi W(\phi)$. We note the presence of auxiliary fields B and G, whose equations of motion are purely algebraic and can be solved straigthforward, $B = \sqrt{k} - \frac{\lambda\bar{\lambda}}{2a} - \frac{a\kappa^2}{2}(W - \frac{1}{2}\eta\bar{\eta})$ and $G = W' + \frac{3}{2a}(\bar{\lambda}\eta - \lambda\bar{\eta})$.

Then, substituting the previous equations in (6) and (7), and making the redefinitions $\lambda \to a^{1/2}\lambda$, $\bar{\lambda} \to a^{1/2}\bar{\lambda}$, $\eta \to a^{3/2}\eta$, $\bar{\eta} \to a^{3/2}\bar{\eta}$, we end up with the following lagrangian

$$
\begin{aligned}
L_{Tot} = & -\frac{3a\dot{a}^2}{c^2e\kappa^2} + \frac{a^3\dot{\phi}^2}{2c^2e} + \frac{3kea}{\kappa^2} + 3\sqrt{k}ea^2W + \frac{3e\kappa^2}{4}a^3W^2 \\
& - \frac{1}{2}a^3eW'^2 + \frac{3i}{c\kappa^2}\left(\lambda\dot{\bar{\lambda}} + \bar{\lambda}\dot{\lambda}\right) - \frac{i}{2c}\left(\eta\dot{\bar{\eta}} + \dot{\eta}\bar{\eta}\right) + \frac{3\sqrt{a}\dot{a}}{c\kappa^2e}\left(\psi\lambda - \bar{\psi}\bar{\lambda}\right) \\
& - \frac{a^{3/2}\dot{\phi}}{2ce}\left(\psi\eta - \bar{\psi}\bar{\eta}\right) + \frac{3i\dot{\phi}}{2c}\left(\lambda\bar{\eta} + \bar{\lambda}\eta\right) + 3e\left(\frac{\sqrt{k}}{\kappa^2 a} - \frac{3}{2}W\right)\lambda\bar{\lambda} \\
& + e\left(-\frac{3\sqrt{k}}{2a} + \frac{3\kappa^2}{4}W - W''\right)\eta\bar{\eta} + 3ia^{3/2}\left(\frac{\sqrt{k}}{\kappa^2 a} - \frac{1}{2}W\right)\left(\psi\lambda + \bar{\psi}\bar{\lambda}\right) \\
& - \frac{ia^{3/2}}{2}W'\left(\psi\eta + \bar{\psi}\bar{\eta}\right) - \frac{3eW'}{2}\left(\lambda\bar{\eta} - \bar{\lambda}\eta\right) - \frac{3}{2e\kappa^2}\psi\bar{\psi}\lambda\bar{\lambda} + \frac{1}{4e}\psi\bar{\psi}\eta\bar{\eta}.
\end{aligned}
$$

The conjugated momenta are $\pi_a = -\frac{6a\dot{a}}{c^2\kappa^2e} + \frac{3\sqrt{a}}{c\kappa^2e}\left(\psi\lambda - \bar{\psi}\bar{\lambda}\right)$, $\pi_\phi = \frac{a^3\dot{\phi}}{c^2e} - \frac{a^{3/2}}{2ce}\left(\psi\eta - \bar{\psi}\bar{\eta}\right) + \frac{3i}{2c}\left(\lambda\bar{\eta} + \bar{\lambda}\eta\right)$, $\pi_\lambda = -\frac{3i}{c\kappa^2}\bar{\lambda}$, $\pi_{\bar{\lambda}} = -\frac{3i}{c\kappa^2}\lambda$, $\pi_\eta = \frac{i}{2c}\bar{\eta}$, $\pi_{\bar{\eta}} = \frac{i}{2c}\eta$. The Dirac Brackets are $\{a, \pi_a\} = \{\phi, \pi_\phi\} = 1$, $\{\lambda, \bar{\lambda}\}_+ = \frac{c\kappa^2}{6}$, $\{\eta, \bar{\eta}\}_+ = -c$, where $\{,\}_+$ are fermionic Dirac brackets. Therefore, the Hamiltonian is $H = NH_0 + \frac{1}{2}\psi S - \frac{1}{2}\bar{\psi}\bar{S}$, where the hamiltonian and supersymmetric constraints are

$$
\begin{aligned}
H_0 = & -\frac{c^2\kappa^2\pi_a^2}{12a} + \frac{c^2\pi_\phi^2}{2a^3} - \frac{3ic\pi_\phi}{2a^3}\left(\lambda\bar{\eta} + \bar{\lambda}\eta\right) - \frac{3ak}{\kappa^2} + 3\sqrt{k}a^2W \\
& - \frac{3\kappa^2}{4}a^3W^2 + \frac{1}{2}a^3W'^2 + \frac{9}{2}W\lambda\bar{\lambda} - \frac{3\kappa^2}{4}W\eta\bar{\eta} + \frac{3}{2}W'\left(\lambda\bar{\eta} - \bar{\lambda}\eta\right) \qquad (8) \\
& + W''\eta\bar{\eta} - \frac{3\sqrt{k}}{\kappa^2 a}\lambda\bar{\lambda} + \frac{3\sqrt{k}}{2a}\eta\bar{\eta} - \frac{9\lambda\bar{\lambda}\eta\bar{\eta}}{4a^3}, \qquad (9)
\end{aligned}
$$

$$S = \frac{c\pi_a \lambda}{\sqrt{a}} + \frac{c\pi_\phi \eta}{a^{3/2}} - \frac{6i\sqrt{k}}{\kappa^2}\sqrt{a}\lambda + 3ia^{\frac{3}{2}}W\lambda + ia^{\frac{3}{2}}W'\eta + \frac{3i}{2}a^{-3/2}\lambda\eta\bar{\eta}, \tag{10}$$

$$\bar{S} = \frac{c\pi_a \bar{\lambda}}{\sqrt{a}} + \frac{c\pi_\phi \bar{\eta}}{a^{3/2}} + \frac{6i\sqrt{k}}{\kappa^2}\sqrt{a}\bar{\lambda} - 3ia^{3/2}W\bar{\lambda} - ia^{3/2}W'\bar{\eta} - \frac{3i}{2a^{3/2}}\bar{\lambda}\eta\bar{\eta}, \tag{11}$$

they satisfy the Dirac fermionic brackets

$$\{S, \bar{S}\}_+ = -2H_0, \tag{12}$$

$$\{H_0, S\} = \{H_0, \bar{S}\} = 0. \tag{13}$$

Further, the scalar potential in the hamiltonian H_0 is

$$V_S = \frac{3\sqrt{k}}{a}W - \frac{3\kappa^2}{4}W^2 + \frac{1}{2}W'^2. \tag{14}$$

4. Quantization

There are several well known problems of the quantization of this system, which are address in.[19] In particular the time problem, a consequence of the invariance under time reparametrizations. It has been argued that time is an internal property that can be determined by a the choice of a clock.[5] On the other side, the observed universe is classical,[19] hence its description is given by mean values of the quantum operators. We further discuss the time problem in section 5.

We follow,[8,20] for the derivation of the Wheeler-DeWitt equation, which follows from the Hamiltonian constraint. In the supersymmetric case, the supercharges (10) and (11) give first order differential equations, from which the Hamiltonian constraint follows (12). For consistency it is required that the Hamiltonian is hermitian, hence the supercharges must satisfy $\bar{S} = S^\dagger$ and $S = \bar{S}^\dagger$. In fact, the constraints (9), (10), (11), and the algebra (12), (13) are symmetric under complex conjugation, with $\lambda^* = \bar{\lambda}$ and $\eta^* = \bar{\eta}$. As the momenta of the fermionic variables satisfy second class constraints, Dirac brackets must be implemented. It follows that the non-zero brackets are

$$[a, \pi_a] = [\phi, \pi_\phi] = i\hbar, \qquad \{\lambda, \bar{\lambda}\}_+ = \frac{4\pi}{3}l_p^2, \qquad \{\eta, \bar{\eta}\}_+ = -\hbar c. \tag{15}$$

$l_p^2 = \frac{\hbar G}{c^3}$ is the Planck length. For the quantization, we redefine the fermionic degrees of freedom as $\lambda = \sqrt{\frac{\hbar c\kappa^2}{6}}\alpha$, $\bar{\lambda} = \sqrt{\frac{\hbar c\kappa^2}{6}}\bar{\alpha}$, $\eta = \sqrt{\hbar c}\beta$ and $\bar{\eta} = \sqrt{\hbar c}\bar{\beta}$. Hence the anticommutators are

$$\{\alpha, \bar{\alpha}\}_+ = 1, \qquad \{\beta, \bar{\beta}\}_+ = -1. \tag{16}$$

as well as $\alpha^2 = \beta^2 = \bar{\alpha}^2 = \bar{\beta}^2 = 0$. The bosonic momenta are represented by derivatives, α and β are annihilation operators, and $\bar{\alpha}$ and $\bar{\beta}$ are creation operators. We fix the ordering ambiguities in the first and last terms of (10) and (11) by Weyl

ordering, which for fermions is antisymmetric. Hence

$$\begin{aligned}\frac{1}{\sqrt{\hbar c}}S &= \frac{c\kappa}{2\sqrt{6}}\left(a^{-\frac{1}{2}}\pi_a + \pi_a a^{-\frac{1}{2}}\right)\alpha + ca^{-\frac{3}{2}}\pi_\phi\beta + \frac{3i\kappa}{\sqrt{6}}a^{\frac{3}{2}}W\alpha \\ &\quad + ia^{\frac{3}{2}}W'\beta - i\frac{\sqrt{6k}}{\kappa}a^{\frac{1}{2}}\alpha - \frac{i\sqrt{3}}{4\sqrt{2}}\hbar c\kappa a^{-\frac{3}{2}}\alpha[\bar\beta,\beta], \qquad (17)\\ \frac{1}{\sqrt{\hbar c}}\bar S &= \frac{c\kappa}{2\sqrt{6}}\left(a^{-\frac{1}{2}}\pi_a + \pi_a a^{-\frac{1}{2}}\right)\bar\alpha + ca^{-\frac{3}{2}}\pi_\phi\bar\beta - \frac{3i\kappa}{\sqrt{6}}a^{\frac{3}{2}}W\bar\alpha \\ &\quad - ia^{\frac{3}{2}}W'\bar\beta + i\frac{\sqrt{6k}}{\kappa}a^{\frac{1}{2}}\bar\alpha + \frac{i\sqrt{3}}{4\sqrt{2}}\hbar c\kappa a^{-\frac{3}{2}}\bar\alpha[\bar\beta,\beta]. \qquad (18)\end{aligned}$$

The anticommutator of these operators is $\{S,\bar S\}_+ = -2\hbar c H_0$, hence the quantum hamiltonian is[a]

$$\begin{aligned}H_0 &= -\frac{c^2\kappa^2}{24}\left(a^{-1}\pi_a^2 + \pi_a^2 a^{-1}\right) + \frac{c^2}{2}a^{-3}\pi_\phi^2 - \frac{\sqrt{k}}{4}\hbar c a^{-1}[\alpha,\bar\alpha] \\ &\quad - \frac{\sqrt{3}i}{2\sqrt{2}}\hbar c^2\kappa a^{-3}\pi_\phi(\alpha\bar\beta + \bar\alpha\beta) + \frac{3\sqrt{k}}{4}\hbar c a^{-1}[\beta,\bar\eta] \\ &\quad - \frac{3k}{\kappa^2}a - \frac{3\kappa^2}{4}a^3W^2 + 3\sqrt{k}a^2W + \frac{1}{2}a^3W'^2 \qquad (19)\\ &\quad + \frac{3}{8}\hbar c\kappa^2 W[\alpha,\bar\alpha] - \frac{3}{8}\hbar c\kappa^2 W[\beta,\bar\beta] + \frac{1}{2}\hbar c W''[\beta,\bar\beta] \\ &\quad + \frac{\sqrt{3}}{2\sqrt{2}}\hbar c\kappa W'(\alpha\bar\beta - \bar\alpha\beta) + \frac{3}{16}(\hbar c\kappa)^2 a^{-3}\left(\bar\alpha\alpha\beta\bar\beta + \alpha\bar\alpha\bar\beta\beta\right).\end{aligned}$$

Thus, $\bar S = S^\dagger$, and H_0 is self-adjoint. The Hilbert space is generated from the vacuum state $|1\rangle$, which satisfies $\alpha|1\rangle = \beta|1\rangle = 0$. Hence, there are four orthogonal states

$$|1\rangle, \quad |2\rangle = \bar\alpha|1\rangle, \quad |3\rangle = \bar\beta|1\rangle \quad \text{and} \quad |4\rangle = \bar\alpha\bar\beta|1\rangle, \qquad (20)$$

which have norms $\langle 2|2\rangle = \langle 1|1\rangle$, $\langle 3|3\rangle = -\langle 1|1\rangle$ and $\langle 4|4\rangle = -\langle 1|1\rangle$. Further, from the constraint equation $S|\Psi\rangle = 0$, we get, up to nonvanishing factors[8]

$$a\left(\partial_a - \frac{3}{\hbar c}a^2W + \frac{6\sqrt{k}}{\hbar c\kappa^2}a + \frac{1}{2}a^{-1}\right)\psi_2 - \frac{\sqrt{6}}{\kappa}\left(\partial_\phi - a^3W'\right)\psi_3 = 0, \qquad (21)$$

$$\left(\partial_a - \frac{3}{\hbar c}a^2W + \frac{6\sqrt{k}}{\hbar c\kappa^2}a - a^{-1}\right)\psi_4 = 0 \quad \text{and} \quad \left(\partial_\phi - \frac{1}{\hbar c}a^3W'\right)\psi_4 = 0, \qquad (22)$$

[a]Note that the ordering differs from the one in[8]

while from $\bar{S}\Psi = 0$, we get

$$a\left(\partial_a + \frac{3}{\hbar c}a^2 W - \frac{6\sqrt{k}}{\hbar c\kappa^2}a + \frac{1}{2}a^{-1}\right)\psi_3 - \frac{\sqrt{6}}{\kappa}\left(\partial_\phi + a^3 W'\right)\psi_2 = 0, \quad (23)$$

$$\left(\partial_a + \frac{3}{\hbar c}a^2 W - \frac{6\sqrt{k}}{\hbar c\kappa^2}a - a^{-1}\right)\psi_1 = 0, \quad \text{and} \quad \left(\partial_\phi + \frac{1}{\hbar c}a^3 W'\right)\psi_1 = 0, \quad (24)$$

4.1. *Solutions*

Since the Wheeler-DeWitt equation is second order, its solutions require boundary conditions. However, in the supersymmetric theory the equations are first order (21)-(24), and have unique solutions, which can be fixed by the requirement of normalization.[8,17,20] The equations for ψ_1 and ψ_4 can be straightforwardly solved, yielding the, up to constant factors, unique solutions

$$\psi_1(a,\phi) = a\exp\left[-\frac{1}{\hbar c}\left(a^3 W(\phi) - \frac{3\sqrt{k}a^2}{\kappa^2}\right)\right], \quad (25)$$

$$\psi_4(a,\phi) = a\exp\left[\frac{1}{\hbar c}\left(a^3 W(\phi) - \frac{3\sqrt{k}a^2}{\kappa^2}\right)\right]. \quad (26)$$

Following[8,20] and the arguments stated there, we choose as the only physical solutions $|\Psi\rangle = C_1\,\psi_1(a,\phi)\,|1\rangle + C_4\,\psi_4(a,\phi)\,|4\rangle$, where the factors are arbitrary constants.

In the following, unless otherwise stated, we will consider $k = 0$, in this case we can write (25) and (26), together as

$$\Psi(a,\phi) = Ca\exp\left[-\frac{1}{\hbar c}a^3|W(\phi)|\right], \quad (27)$$

where the normalization constant is

$$|C|^2 = \frac{6}{c\hbar}\left\{\int_{-\infty}^{\infty}\frac{d\phi}{|W(\phi)|}\right\}^{-1}. \quad (28)$$

In[20] the corresponding expressions for $k = 1$ are given.

5. Time

A brief but comprehensive discussion of the problem of time in quantum cosmology is given in,[20] for the moment let us give some plausibility coments in the same line of thought. The wave function of a quantum system should give high probabilities on paths around classical trajectories in superspace.[19] Hence, it is possible to identify mean trajectories along regions around maxima of the wave function, and it should be possible to parameterize these trajectories, to give a time. Strictly speaking, measurements should give random values around these mean values. Therefore,

this time would correspond itself to a sort of mean value. We consider the standard interpretation of quantum mechanics for the solutions of the Wheeler-DeWitt equation, i.e. the square module of the wave function is the probability density of measuring a certain three-geometry. Thus, the wave function must give probabilities for all possible three geometries of the universe.

In our case, the configuration space is given by the scale factor and the scalar field. If the probability density has crests, we could speak of "trajectories" in the (a, ϕ) plane. From (27) we see, that for ϕ constant, we have a bell curve for (27), with a maximum at $a_{\max}(\phi) = \left[\frac{c\hbar}{3W(\phi)}\right]^{1/3}$, and $\psi_{\max}(\phi) = \psi(a_{\max}, \phi) = e^{-1/3} a_{\max}(\phi)$. As an example, we set $W(\phi) = \phi^{-1}(e^{-\phi^2} + 1)$, with the probability density given in figure 2, and taking the scalar field to lie in $\phi \in [0, \infty)$, then the most probable value for the universe would be somewhere at $\phi \to \infty$; hence, the system is not localized. But, if the universe is at some value ϕ_0, it will be less probable to find it for $\phi < \phi_0$, and more probable to find it for $\phi > \phi_0$, around $a_{\max}(\phi)$, this behaviour suggest that the field ϕ can take the role of time,[6,8,20] as far as it is causal. This corresponds to the comoving gauge, where the scalar field is constant on the spacial slices.

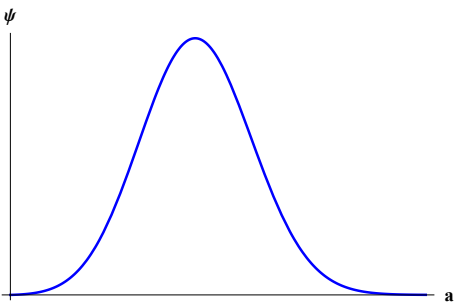


Fig. 1. Profile of $\psi(a, \phi)$, for ϕ constant.

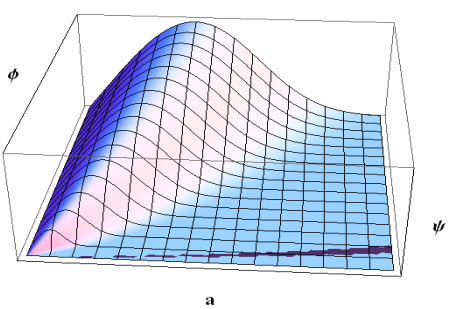


Fig. 2. Probability density for $W(\phi) = 1/\phi$.

Hence we set $\phi \to t$, with a probability amplitude[8]

$$\Psi(a,t) = \frac{1}{\sqrt{\int_0^\infty da\,|\psi(a,t)|^2}}\,\psi(a,t) = \sqrt{\frac{6|W(t)|}{\hbar c}}\,a\exp\left[-\frac{a^3|W(t)|}{\hbar c}\right], \tag{29}$$

normalized as $\int_0^\infty da\,|\Psi(a,t)|^2 = 1$, where the resulting probability density[8]

$$|\Psi(a,t)|^2 = \frac{|\psi(a,t)|^2}{\int_0^\infty da\,|\psi(a,t)|^2}, \tag{30}$$

is the conditional probability of the universe being at a and $t \sim \phi$, if the universe is at ϕ regardless of a. It satisfies a conservation equation as stated in.[8,20]

Since we do not have the possibility of recreating the universe in order to give a meaning to probabilities, what we can give a meaning is to mean values. Therefore, under the preceding ansatz, the classical setup arises from mean values with the probability amplitude (29), upon this line we get a time dependent scale factor

$$a(t) = \int_0^\infty a|\Psi(a,t)|^2 da = \Gamma\left(4/3\right)\left[\frac{\hbar c}{2|W(t)|}\right]^{1/3}. \tag{31}$$

with the standard deviations

$$(\Delta a)^2 = \left[\Gamma\left(5/3\right) - \Gamma\left(4/3\right)^2\right]\left[\frac{\hbar c}{2|W|}\right]^{2/3}, \tag{32}$$

and

$$(\Delta\pi_a)^2 = \hbar^2\,\Gamma\left(1/3\right)\left[\frac{2|W|}{\hbar c}\right]^{2/3}, \tag{33}$$

from which follows the uncertainty relation holds

$$\Delta a\Delta\pi_a = \sqrt{\Gamma\left(1/3\right)\left[\Gamma\left(5/3\right) - \Gamma\left(4/3\right)^2\right]}\,\hbar \approx 0.53\,\hbar. \tag{34}$$

6. Inflationary scenarios

In this section we will consider several superpotentials that lead to cosmological scenarios. The scalar field has real values, from $-\infty$ to ∞, and as mentioned in 4.1 and,[20] we require that the superpotential is a symmetric function and always positive. It is convenient to choose time beginning at $t = 0$. Hence, these superpotentials correspond to bouncing universes. With these settings, we are considering universes with a defined origin at t=0, described by homogeneous quantum cosmology from its very beginning. We will consider two types of superpotentials, depending on the initial conditions. The first type is suitable for $a(0) = 0$, hence $\lim_{t\to 0} W(t) = \infty$ (31), and the wave function (29) will vanish at $t = 0$, $\Psi(a,0) = 0$, similar to figure 1. Hence, the universe will arise for $t > 0$ from nothing. Otherwise, if $W(0)$ takes some

finite value, then $a(0) = \Gamma(4/3)\left[\frac{\hbar c}{2|W(0)|}\right]^{1/3}$. We can obtain superpotentials of the first type by $W(\phi) \to \phi^{-2\alpha} W(\phi)$, with α a positive number.

From (31) we get

$$\ddot{a}(t) = \frac{a(t)}{9}\left[4\frac{\dot{W}^2(t)}{W^2(t)} - 3\frac{\ddot{W}(t)}{W(t)}\right]. \tag{35}$$

Hence the acceleration is positive if $\dot{W}^2(t) > \frac{3}{4}W(t)\ddot{W}(t)$. We consider inflation starting when the scale factor acceleration becomes positive, at $t = t_i$, until the exit when it becomes zero again $\ddot{a}(t_e) = 0$, with $N = \ln\frac{a(t_e)}{a(t_i)} = \frac{1}{3}\ln\frac{W(t_i)}{W(t_e)}$ e-folds. Note that in this way N is independent of the time parametrization.

The feasibility of inflation has an indicative in the Hubble horizon

$$(aH)^{-1} = -\frac{3 \times 2^{1/3}}{\Gamma(4/3)(c\hbar)^{1/3}}\frac{W(t)^{4/3}}{W'(t)}. \tag{36}$$

For simplicity, in the following we will consider an adimensional time $t \to \tau = t/t_p$, where $t_p = \sqrt{\hbar G/c^5}$ is the Planck time.

6.1. *Superpotentials*

We consider two types of superpotentials[20]

Gaussian superpotentials

$$W(\tau) = \frac{c^4 M_p^3}{\hbar^2}\tau^{-2\alpha}\left(e^{-\tau^2} + \lambda\right), \tag{37}$$

and Step superpotentials

$$W(\tau) = \frac{c^4 M_p^3}{\hbar^2}\tau^{-2\alpha}\left(\frac{1}{\tau^2 + \lambda} + 1\right), \tag{38}$$

where $M_p = \sqrt{\frac{\hbar c}{8\pi G}}$ is the reduced Planck mass, $\alpha \geq 0$, and λ is a positive constant. We consider representative cases which reproduce an inflationary period with $N \sim 60$ e-folds, evaluated considering the beginning and exit of inflation from the time interval where acceleration is positive, and a consistent Hubble horizon. As we will see, a very small λ is required for enough e-folds, and we set $\lambda = e^{-\nu}$. Regarding other parameters, as the tensor-scalar ratio, an estimate considering an effective one scalar FRLW cosmology leads to inconsistent results, a more precise evaluation is needed.

The general behaviour of the system under these superpotentials, (37) and (38), is as follows: the scale factor has an initial value $a(0) > 0$ with a very small positive acceleration for $\alpha = 0$, and $a(0) = 0$, with a vanishing acceleration for $\alpha \geq 1$. In order to see if there is a consistent inflation, we evaluate the values of the scale factor at $\tau = 0$, and the corresponding acceleration. If $a(0) > 0$, then inflation

begins at the time $\tau = \tau_i$, when the acceleration becomes positive, until it becomes negative, at $\tau = \tau_f$. If the acceleration is positive already at $\tau = 0$, inflation begins right away. Further, if $a(0) = 0$ and $\ddot{a}(0) \geq 0$, and afterwards $\ddot{a} > 0$, then there will be exit, but there are infinite e-folds, unless we discard some initial period, for any reason. Otherwise, if $a(0) = 0$ and $\ddot{a}(0) < 0$, then inflation will begin as soon as the acceleration becomes positive, at $\tau = \tau_i$, and finish when acceleration becomes negative again, at $\tau = \tau_f$.

6.2. *Gaussian superpotentials*

The scale factor for (37) is

$$a(\tau) = \frac{\Gamma(4/3)\hbar\tau^{\alpha/3}}{2^{1/3}cM_p\left(e^{-\tau^2}+\lambda\right)^{1/3}}. \tag{39}$$

In the following, we will consider unities $c = 1$, $\hbar = 1$, and $\kappa = 1$. The acceleration vanishes around the value at which the superpotential (37) becomes nearly constant, at $e^{-\tau^2} \approx \lambda$, i.e. $\tau \approx \sqrt{\nu}$. Thus, for $\tau > \sqrt{\nu}$, the scale factor behaves as $a \gtrsim (\sqrt{\nu})^{\alpha/3}\lambda^{-1/3}$. Hence to have $a \lesssim 1$ for late times, it is convenient to rescale (39) by $\lambda^{1/3}$, such that

$$a(\tau) = \frac{\Gamma(4/3)\hbar\tau^{\alpha/3}}{2^{1/3}cM_p\left(\lambda^{-1}e^{-\tau^2}+1\right)^{1/3}} \tag{40}$$

Therefore, for $\alpha = 0$, we have

$$a_0 = a(0) = \frac{\Gamma(4/3)\hbar}{2^{1/3}cM_p}\left(\frac{\lambda}{1+\lambda}\right)^{1/3} \approx \frac{\Gamma(4/3)\hbar\lambda^{1/3}}{2^{1/3}cM_p}, \tag{41}$$

with a positive acceleration

$$\ddot{a}(0) \approx \frac{2^{2/3}\Gamma(4/3)\hbar\lambda^{1/3}}{3cM_p}. \tag{42}$$

On the other side, for large times, $\tau > \sqrt{\nu}$, the gaussian in (37) can be neglected, and the scale factor is almost constant

$$a(\tau) \approx a_e = \frac{\Gamma(4/3)\hbar}{2^{1/3}cM_p}, \tag{43}$$

see figure 4. Thus, for $\alpha = 0$, the e-fold number satisfies

$$N \lesssim \ln\frac{a_e}{a_0} = -\frac{1}{3}\ln\lambda = \frac{\nu}{3}. \tag{44}$$

It turns out that a value $\lambda \sim 10^{-79}$ is necessary for ~ 60 e-folds. From (41), we see that there is a very small initial value $a(0) \sim 3 \times 10^{-27}$, with an initial velocity

$\dot{a}(0) = 0$. The initial acceleration is also very small $\ddot{a}(0) \sim 10^{-27}$, but inflation begins right away, until the acceleration slows and becomes negative at $\tau_f \approx 13.4$, see figure 3. We get $N = \ln \frac{a(\tau_f)}{a_0} \approx 60.1$, consistently with (44). A plot for the horizon is given in figure 5.

For $\alpha > 0$, $a(0) = 0$, and at early times, $\tau \ll 1$, the scale factor can be approximated by $a(\tau) \approx \frac{\Gamma(4/3)\hbar\lambda^{1/3}}{2^{1/3}cM_p}\tau^{\alpha/3}e^{\frac{\tau^2}{3}}$, and $\ddot{a}(\tau) \approx \frac{\Gamma(4/3)\kappa(c\hbar)^{1/2}\lambda^{1/3}}{9\times 2^{1/3}}[\alpha(\alpha-3)+2(2\alpha+3)\tau^2]\tau^{\alpha/3-2}$. The acceleration at late times is shown in figure 3 for $\alpha = 1, 2, 3$, for higher values it is similar. Further, for early times, we have: For $\alpha = 1, 2$, $\ddot{a}(0) = -\infty$, grows and becomes positive. For $\alpha = 3$, $\ddot{a}(0) = 0$, and then grows. For $\alpha = 4, 5$, $\ddot{a}(0) = \infty$, decreases almost to zero for $0 < \tau \lesssim 0.5$, and then grows again. For $\alpha = 6$, $\ddot{a}(0) = \frac{2^{2/3}\Gamma(4/3)\hbar}{cM_p} \sim 3 \times 10^{-27}$, and grows. Finally, the cases $\alpha > 6$ are similar to $\alpha = 3$.

For $\alpha = 1$ a meaningful e-fold number can be obtained. Namely, for $\lambda \sim 10^{-77}$, the acceleration becomes positive at $\tau_i \sim 0.4$, with $N = \ln a(\tau_f)/a(\tau_i) \sim 60.2$. For $\alpha = 2$, the initial acceleration is infinite; hence as $a(0) = 0$, N is unbounded. However, the initial acceleration is very quickly decreasing, see figure 11, and we could discard the initial, decelerating period, which presumably would require full quantum gravity. In this case, if we consider inflation starting at $\tau \sim 0.3$, where acceleration becomes practically zero and begins to grow, we get similar values as for $\alpha = 1$.

Thus, the previous three cases $\alpha = 0, 1, 2$, are consistent with inflation, see figures 3 and 4. Even if the previous results are alike, the event horizons, plotted in figure 5, are quite different. With the previous results, we can compute the effective matter density and pressure for a perfect fluid. It turns out that the resulting potential differs from the scalar potential following from the hamiltonian (14). Moreover, for the gaussian potential and $k = 0$, in the analysed cases, the fluid has phantom energy. The effective potentials for $k = 0$ and $k = 1$ are given in figures 6 and 7. Note that the second case seems to correspond to a tunneling.

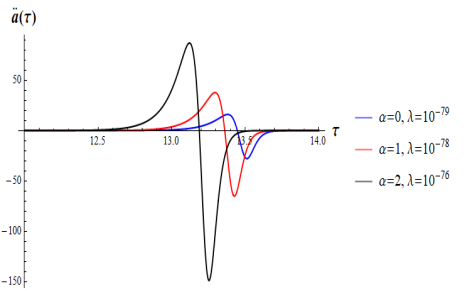


Fig. 3. Acceleration $\ddot{a}(\tau)$ for gaussian superpotentials.

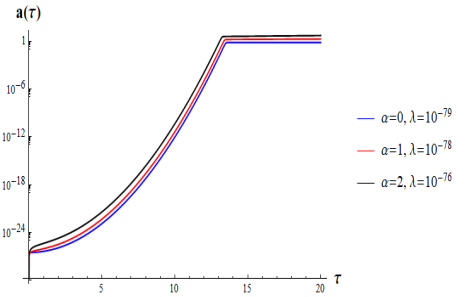


Fig. 4. Scale factor $a(\tau)$ for gaussian superpotentials, the graphics are logarithmic for clarity.

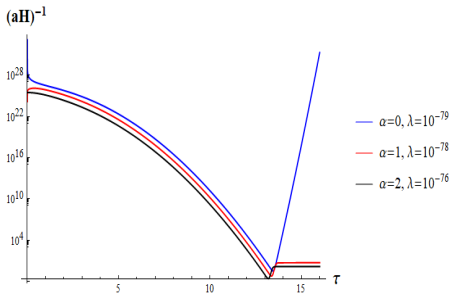


Fig. 5. Comoving Hubble radius $(aH)^{-1}$ for gaussian superpotentials, the graphics are logarithmic for clarity.

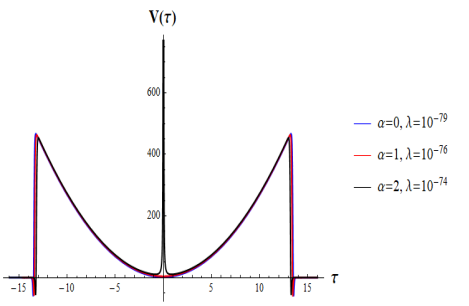


Fig. 6. Effective potentials from Friedmann equations, for $k = 0$, for $k = 1$, for $\alpha = 1, 2$ the potential becomes ∞ at $\tau = 0$.

6.3. *Step superpotentials*

For the step superpotentials (38), we have

$$a(\tau) = \frac{\Gamma(4/3)\hbar\tau^{\alpha/3}}{2^{1/3}cM_p\left[(\tau^2+\lambda)^{-1}+1\right]^{1/3}} \tag{45}$$

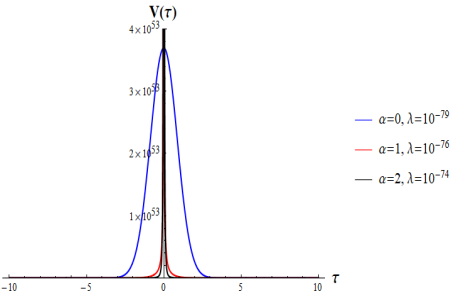


Fig. 7. Effective potentials from Friedmann equations, for $k = 1$, for $\alpha = 2$ the potential becomes ∞ at $\tau = 0$.

And the behaviour of the system is the same as for gaussian superpotentials, if $\alpha = 0$ the scale factor initial value is (41), and for large time values, $\tau \gg 1$, the scale factor tends to (43). Hence, the e-folds satisfy as well (44). The initial velocity and acceleration are zero, $\dot{a}(0) = 0$, $\ddot{a}(0) = 0$, and inflation begins from this moment, and ends at $\tau \approx 0.51$, when the acceleration becomes negative. For $N \sim 60$ a value $\lambda \sim 10^{-80}$ is required.

In the case $\alpha \geq 1$ the situation is also similar to the one of the preceding section, the initial values are practically the same. The acceleration initial behavior is similar, but after a short time the differences become notable, as can be seen from figures 11 and 12. We give also the corresponding profiles for the event horizons, shown in figures 5 and 10. For $\alpha = 1$, the 60 e-folds are reached for $\lambda \sim 10^{-53}$.

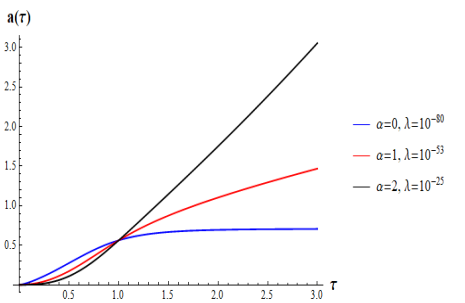


Fig. 8. Scale factor $a(\tau)$ for step superpotentials.

In this case the effective potentials, from the Friedmann equations, look qualitatively similar to the previous ones.

7. Discussion

Quantum cosmology gives a canonical quantization of general relativity in the Schrödinger picture. The invariance under time reparametrizations leads to the

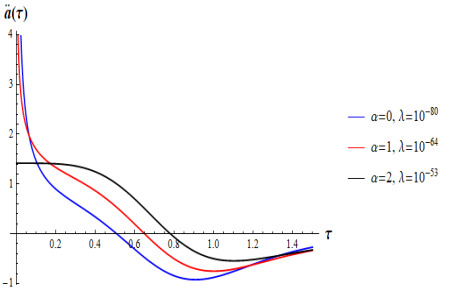


Fig. 9. Acceleration $\ddot{a}(\tau)$ for step superpotentials.

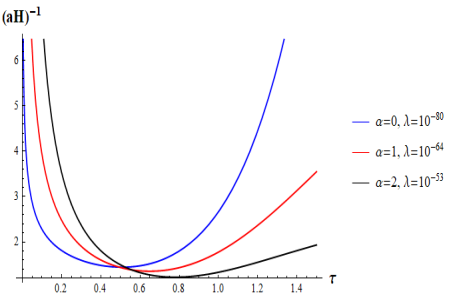


Fig. 10. Comoving Hubble radius $(aH)^{-1}$ for step superpotentials.

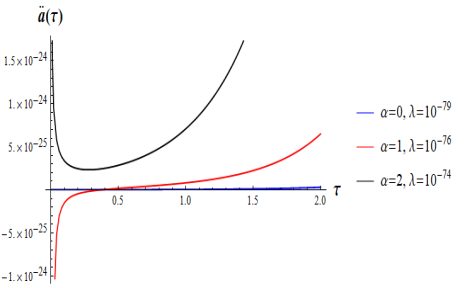


Fig. 11. Initial acceleration for gaussian superpotentials.

hamiltonian constraint that corresponds to a zero energy time independent Schrödinger equation, i.e. the Wheeler-DeWitt equation. There is no time dependent Schrödinger equation that generates a time dependence for the wave function. Thus, the wave function encodes the probability amplitudes for all possible configurations at any time. The observed universe is classical and has time, and classical physics should follow from quantum physics, hence a time evolution should follow from the

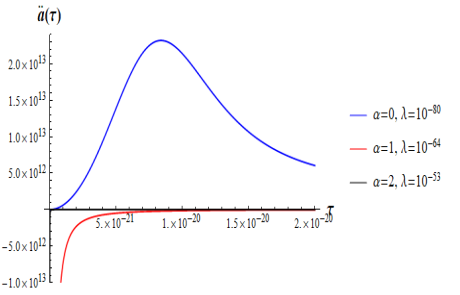


Fig. 12. Initial acceleration for step superpotentials.

quantum description. The Wheeler-DeWitt equation in the supersymmetric case is a system of first order equations, whose solutions are spinorial wave functions, see e.g.,[11,12] whose meaning has not been established. In general, there are only two nonvanishing components, see e.g.,[15] given by real exponentials with opposite exponent sign. Here we have considered a homogeneous theory with a minimally coupled scalar field. In this case, the solution of the supersymmetric Wheeler-DeWitt equation (21)-(24) has four components, two of them have analytic expressions (25),(26), and only one is physically consistent. As the equations are homogeneous and two components decouple, the other ones can be taken as trivial. The form of the solution, (27), suggests that in a certain gauge, time can be given by the scalar field and trajectories for the observables can be given by their mean values. We compute an evolution for the scale factor $a(t)$, and time reparametrizations can be performed. It remains the general question of how far can we get classical trajectories from a quantum state in this way.

We have considered several inflationary scenarios with this approach, with superpotentials (37) and (38). These superpotentials have a factor $\tau^{-2\alpha}$ that, for suitable values, gives a wave function that vanishes for $\tau = 0$ and produces a path of increasing probabilities, we can interpret this as giving the direction of time. The exit of inflation is implemented at the level of the superpotential through a constant $\lambda \sim e^{-3N}$, where N are the observationally required e-folds. Three cases were given in plots for each type of superpotential, figures 3-12. These results indicate that this kind of models could be realistic. The inclusion of inhomogeneous perturbations and the evolution of the fluctuations are part of future work. Further, more complex models, e.g. with several scalars could be considered, as in.[14]

Acknowledgments

We thank VIEP-BUAP, PRODEP-SEP and PROFEXCE-SEP for the support.

References

1. R. D'Inverno, *Introducing Einstein's Relativity* (Oxford University Press, Oxford, England, 1998).
2. V.F. Mukhanov and G.V. Chibisov, Quantum Fluctuations and a Nonsingular Universe, *JETP Lett.* **33** 532 (1981).
3. D. Baumann, *TASI Lectures on Inflation*, arXiv:0907.5424 [hep-th] (2012).
4. J. B. Hartle and S. W. Hawking, Wave function of the Universe, *Phys. Rev.* D **28** 2960 (1983).
5. K. V. Kuchař, Time and interpretations of quantum gravity, *Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics* edited by Kunstatter G, Vincent D and Williams J *Int. J. Mod. Phys* D20 3 (2011).
6. C. J. Isham, Canonical quantum gravity and the problem of time, *Proceeding of the NATO Advanced Study Institute "Recent Problems in Mathematical Physics"* arXiv:gr-qc/9210011 (1992).
7. E. Anderson, Problem of time in quantum gravity, *Ann. Phys.* **524** 12 757-786 (2012).
8. C. Ramírez and V.M. Vázquez-Báez, Quantum supersymmetric FRW cosmology with a scalar field, *Phys. Rev.* D **93** 043505 (2016).
9. J. Wess and J. Bagger, *Supersymmetry and Supergravity*, 2nd edn (Princeton University, Princeton, NJ, 1992).
10. A. Macías, O. Obregón and M. P. Ryan, Quantum cosmology: The supersymmetric square root, *Class. Quantum Grav.* **4** 1477 (1987).
11. P. D. D'Eath, *Supersymmetric quantum cosmology* (Cambridge University Press, Cambridge, England, 1996).
12. P. V. Moniz, *Quantum Cosmology - The Supersymmetric Perspective - Vol. 1: Fundamentals* (Springer, Berlin, Germany,2010).
    Moniz P V 2010 *Quantum Cosmology - The Supersymmetric Perspective - Vol. 2: Advanced Topics* (Springer, Berlin, Germany,2010).
13. O. Obregón, J.J. Rosales, J. Socorro and V. I. Tkach, Supersymmetry breaking and a normalizable wavefunction for the FRW ($k = 0$) cosmological model *Class. Quantum Grav.* **16** 2861 (1999).
14. G. García-Jiménez, C. Ramírez and V. Vázquez-Báez, Tachyon potentials from a supersymmetric FRW model *Phys. Rev.* D **89**, 043501 (2014).
15. O. Obregón and C. Ramírez, Dirac like formulation of quantum supersymmetric cosmology *Phys. Rev.* D **57** 1015 (1998).
16. C. Escamilla-Rivera, O. Obregon and L. A. Urena-Lopez, Supersymmetric classical cosmology *JCAP* **12**, 011 (2010).
17. C. Ramírez and V. M. Vázquez-Báez, Quantum cosmology of quadratic $f(R)$ theories with a FRW metric. *Adv. Math. Phys.* **2017** 1056514 (2017).
18. C. Ramírez, The realizations of local supersymmetry *Ann. Phys.* **186** 43 (1988).
19. J. J. Halliwell, Decoherence in quantum cosmology *Phys. Rev. D* **39** 2912 (1989).
20. N. E. Martínez-Pérez, C. Ramírez and V. Vázquez-Báez, arXiv:2104.12914v2 (2021).
21. O. Obregón, J. J. Rosales and V. I. Tkach, Superfield description of the FRW universe *Phys. Rev.* D **53** R1750 (1996).
22. M. Ryan, *Hamiltonian Cosmology* (Springer, Berlin, Germany, 1972).
23. V. I. Tkach, J. J. Rosales and O. Obregón, Supersymmetric action for Bianchi type models *Class. Quantum Grav.* **13** 2349 (1996).
24. A. Linde, Fast-Roll inflation *JHEP* **11** 052 (2001).
25. G. Felder, A. Frolov, L. Kofman and A. Linde, Cosmology with negative potentials *Phys. Rev.* D **66** 023507 (2002).

26. A. Csordás and R. Graham, Hartle-Hawking state in supersymmetric minisuperspace *Phys. Lett.* B **373** 51 (1996).
27. J. Bene and R. Graham, Supersymmetric homogeneous quantum cosmologies coupled to a scalar field *Phys. Rev.* D **49** 799 (1994).
28. B. S. DeWitt, Dynamical Theory in Curved Spaces. I. A Review of the Classical and Quantum Action Principles *Rev. Mod. Phys.* **29** 377 (1957).
A. Vilenkin, Interpretation of the wave function of the Universe *Phys. Rev.* D **39** 1116 (1989).
29. H. García-Compeán, O. Obregón and C. Ramírez, Noncommutative Quantum Cosmology *Phys. Rev. Lett.* **88** 161301 (2002).
30. M. Bojowald, *Quantum Cosmology A fundamental Description of the Universe* (Springer, Berlin,Germany, 2011).
31. J. D. Brown and K. Kuchař, Dust as a standard of space and time in canonical quantum gravity *Phys. Rev.* D 51 5600 (1995).

Effective $f(R)$ actions for modified loop quantum cosmologies

Ana Rita Ribeiro

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016, Lisbon, Portugal
E-mail: anaritaribeiro95@hotmail.com

Daniele Vernieri

Department of Physics "E. Pancini", University of Naples "Federico II", Naples, Italy and INFN Sez. di Napoli, Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Naples, Italy
E-mail: daniele.vernieri@unina.it

Francisco S. N. Lobo

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016, Lisbon, Portugal and
Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisbon, Portugal
E-mail: fslobo@fc.ul.pt

Although General Relativity (GR) is an extremely successful theory, at least for weak gravitational fields, it breaks down at very high energies. For example, extrapolating the expansion of the Universe backwards in time yields an infinite energy density, which is referred to as the initial singularity problem. Quantum Gravity is expected to provide a solution to this open question. In fact, one alternative scenario to the Big Bang, that avoids the singularity, is offered by Loop Quantum Cosmology (LQC), which predicts that the Universe undergoes a collapse to an expansion through a bounce. In this work we use metric f(R) gravity to reproduce the modified Friedmann equations, which have been obtained in the context of modified loop quantum cosmologies (mLQC). Using a order reduction method, we obtain covariant effective actions that lead to a bounce, for specific models of mLQC, considering a massless scalar field.

Keywords: Initial singularity, cyclic cosmology, loop quantum cosmology, $f(R)$ gravity, covariant order reduction.

1. Introduction

1.1. *Motivations: Loop quantum cosmology and its modifications*

A quantum theory of gravity is expected to provide insight on the behaviour of gravity at quantum scales,[1] and consequently in addressing fundamental open questions,[2] such as the initial singularity problem. Loop Quantum Gravity (LQG)[3–5] is one of the main candidates we have today. Indeed, through an effective Hamiltonian description, LQC, which is described as a symmetry-reduced model of LQG,[6] yields a modified Friedmann equation that dictates that the big bang singularity is

replaced by a quantum bounce:[7–13]

$$H^2 = \frac{1}{3}\kappa\rho\left(1 - \frac{\rho}{\rho_c}\right), \tag{1}$$

where $H \equiv \dot{a}/a$ is the Hubble function, given in terms of the scale factor $a(t)$, and $\rho_c = \sqrt{3}/(32\pi^2\gamma^3)$ is the critical energy density, where $\gamma \approx 0.2375$ is the Barbero–Immirzi parameter.[14] Throughout this work, we will be using the geometrized system of natural units in which $c = \hbar = G = 1$ and, accordingly, $\kappa \equiv 8\pi$. It is clear from Eq. (1) that the Universe behaves in a cyclic way and that the quantum bounce occurs at the critical density, when $H = 0$ and $\ddot{a} > 0$.

Nevertheless, the procedure used in LQC, to obtain the Hamiltonian constraint which then leads to Eq. (1), differs from the one used in the full theory of LQG. In LQC the Lorentzian and Euclidean terms are treated as proportional to each other, contrary to LQG, leading to different Hamiltonian constraints in the process of quantization.[9] One of the first attempts to construct a Hamiltonian constraint that is similar to the construction in LQG was undertaken in Ref. 15, in which Thiemann's regularization was used. Even so, there is still no systematic way of deriving LQC as a cosmological sector of LQG and different paths to obtain LQG-like Hamiltonians from LQC followed.[16–19] Given these quantization ambiguities in constructing the effective Hamiltonian constraint operator, in isotropic LQC, there is a need to understand how different quantization methods affect the physical predictions. More specifically, there are still great possibilities for the expanding universe to recollapse due to quantum gravity effects.

In this work, we focus on two specific modified loop quantum cosmology (mLQC) models,[15,20] which differ from standard LQC in the way the Lorentzian term in the Hamiltonian constraint is treated.[21]

1.2. *Modified loop quantum cosmology – I: mLQC-I*

This model was derived in Ref. 20, by treating the Lorentzian term in the Hamiltonian constraint separately and using Thiemann's regularization,[15] yielding a fundamental difference between LQC and mLQC-I, in that in the latter the evolution of the universe is described by two different branches, b_+ and b_-, which coincide at the bounce.

The modified Friedmann equation for the b_- branch takes the form[20,21]

$$H_-^2 = \frac{\kappa\rho}{3}\left(1 - \frac{\rho}{\rho_c^I}\right)\left[1 + \frac{\gamma^2\rho/\rho_c^I}{(\gamma^2+1)\left(1+\sqrt{1-\rho/\rho_c^I}\right)^2}\right], \tag{2}$$

where $\rho_c^I \equiv \rho_c/[4(1+\gamma^2)]$ is the critical density at which the quantum bounce occurs. Notice how, for $\rho/\rho_c^I \ll 1$, the standard Friedmann equation $H^2 \approx \kappa\rho/3$ is recovered.

Concerning the b_+ branch, on the other hand, the modified Friedmann equation takes the form[20,21]

$$H_+^2 = \frac{\kappa\alpha\rho_\Lambda}{3}\left(1-\frac{\rho}{\rho_c^I}\right)\left[1+\frac{\rho\left(1-2\gamma^2+\sqrt{1-\rho/\rho_c^I}\right)}{4\gamma^2\rho_c^I\left(1+\sqrt{1-\rho/\rho_c^I}\right)}\right], \tag{3}$$

where $\alpha \equiv (1-5\gamma^2)/(\gamma^2+1)$, $\rho_\Lambda \equiv 3/[\kappa\alpha\lambda^2(1+\gamma^2)^2]$, and $\lambda^2 \equiv 4\sqrt{3}\pi\gamma$. The limit of Eq. (3) when $\rho/\rho_c \ll 1$, is given by $H^2 \approx 8\pi G_\alpha(\rho+\rho_\Lambda)/3$, where $G_\alpha := \alpha G$. Comparing with the standard Friedmann equation, this limit shows that quantum geometric effects are still present for low energy densities in the form of a modified Newtonian constant, G_α, and an emergent positive cosmological constant, ρ_Λ. If this branch lies in the post-bounce universe, in which we live, these pose phenomenological problems. It can be shown that the classical energy conservation law still holds in this modified model of LQC, concerning both branches.[20]

1.3. *Modified loop quantum cosmology – II: mLQC-II*

The second model that we consider in this work was derived by using Thiemann's regularization, similarly to the previous model, while also using the proportionality of between the Ashtekar's connection and the extrinsic curvature.[15,22] In doing so, yields [15]

$$H^2 = \frac{2\kappa\rho}{3}\left(1-\frac{\rho}{\rho_c^{II}}\right)\left[\frac{1+4\gamma^2\left(\gamma^2+1\right)\rho/\rho_c^{II}}{1+2\gamma^2\rho/\rho_c^{II}+\sqrt{1+4\gamma^2\left(1+\gamma^2\right)\rho/\rho_c^{II}}}\right], \tag{4}$$

where the quantum bounce occurs at the critical density $\rho_c^{II} \equiv 4(\gamma^2+1)\rho_c$. Unlike in mLQC-I, the bounce in this model is perfectly symmetric, as in LQC. Similarly to the mLQC-I model, the dynamics in the Planck regime are less trivial and the classical limit is given by the usual Friedmann equation.[22] Moreover, the energy-momentum conservation law still holds without any change in the properties of the barotropic equation of state (EoS).

1.4. *Covariant order reduction*

The only covariant action that leads to second-order equations, under metric variation, is the Einstein–Hilbert (EH) action which leads to the usual Friedmann equation. In this work, we are interested in asking if the modified Friedmann Eqs. (2)–(4) can be obtained from different actions. An interesting analysis was considered in the Palatini approach of $f(R)$ gravity in Ref. 23, where R is the Ricci scalar and f some function of it. In this case, the modified Friedmann equation does not involve higher than second-order derivatives of the geometrical quantities and is just a function of the matter sources.[24]

In this work, however, we have used metric $f(R)$ gravity,[25–31] where the full field equations are of fourth-order. Given that our current theory of gravity is a

second-order one, we have considered our chosen theory as an effective one, by applying the method of covariant order reduction, in which only the solutions that are perturbatively close to GR are considered as physical ones and the rest considered spurious.[32–34]

This technique has been applied within the context of bouncing cosmologies, for different modifications of GR. In Ref. 35, bouncing solutions in a modified Gauss–Bonnet gravity theory were explored, of the type $R + f(G)$, where G is the Gauss–Bonnet term.[35] This analysis was then extended in the context of $f(R, G)$ gravity.[36] The same approach has also been implemented in the context of $f(Q)$ symmetric teleparallel gravity theories, where Q is the non-metricity scalar.[37] Indeed, in all of these works, a covariant gravitational action, that leads to the quantum bounce observed in LQC, was found by demanding that the Friedmann equation derived within such gravity theories coincides with the one emerging from LQC.

2. Reduced equations in $f(R)$ gravity

In metric $f(R)$ gravity, the corresponding action is given by

$$S = \frac{1}{2\kappa}\int_{\mathcal{V}} d^4x\sqrt{-g}f(R) + S_M(g_{\mu\nu}, \Psi)\,, \tag{5}$$

where g is the determinant of the metric tensor $g_{\mu\nu}$, S_M is the matter action defined as $S_M = \int_{\mathcal{V}} d^4x\sqrt{-g}\,\mathcal{L}_M(g_{\mu\nu}, \Psi)$, being $\mathcal{L}_M$ the matter Lagrangian density, in which matter is minimally coupled to the metric $g_{\mu\nu}$, and Ψ collectively denotes the matter fields. Without loss of generality, we parametrize the Lagrangian density $f(R)$ as the sum of the gravitational Lagrangian of GR and a deviation term, as follows

$$f(R) = R + 2\Lambda + \epsilon\varphi(R)\,, \tag{6}$$

where Λ is the cosmological constant, ϵ is a dimensionless parameter, and $\epsilon\varphi(R)$ represents the deviation from GR. Substituting Eq. (6) in Eq. (5), and varying the action with respect to the metric, yields the following field equations

$$G_{\mu\nu} - g_{\mu\nu}\Lambda + \epsilon\left[-\frac{1}{2}g_{\mu\nu}\varphi(R) + \varphi'(R)R_{\mu\nu} - (\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box)\,\varphi'(R)\right] = \kappa T_{\mu\nu}\,, \tag{7}$$

where the energy-momentum tensor $T_{\mu\nu}$ is defined as

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\,\frac{\delta(\sqrt{-g}\mathcal{L}_M)}{\delta g^{\mu\nu}}\,. \tag{8}$$

In simple terms, applying the order reduction technique to Eq. (7) amounts to replacing the Ricci scalar and the Ricci tensor, in the terms of order ϵ, by the expression we get for them, from the $\epsilon = 0$ version of the same equations. Thus, setting $\epsilon = 0$ in Eq. (7), yields the following

$$R^T_{\mu\nu} = \frac{1}{2}g_{\mu\nu}(R_T + 2\Lambda) + \kappa T_{\mu\nu}\,, \tag{9}$$

as the reduced expression of the Ricci tensor, where R_T is the order reduced Ricci scalar. Taking into account the trace of Eq. (9), the latter provides $R_T = -\kappa T - 4\Lambda$, and after substituting it into Eq. (9), we arrive at

$$R^T_{\mu\nu} = \kappa T_{\mu\nu} - \frac{1}{2} g_{\mu\nu}\kappa T - \Lambda g_{\mu\nu} \,. \tag{10}$$

Replacing R and $R_{\mu\nu}$, in the ϵ-order terms in Eq. (7), with the expressions for R_T and $R^T_{\mu\nu}$ respectively, yields the reduced field equations

$$\begin{aligned} G_{\mu\nu} - g_{\mu\nu}\Lambda + \epsilon \Bigg[&- \frac{1}{2} g_{\mu\nu}\varphi(R_T) + \varphi'(R_T)\left(\kappa T_{\mu\nu} - \frac{1}{2} g_{\mu\nu}\kappa T - \Lambda g_{\mu\nu}\right) \\ &- \left(\nabla_\mu \nabla_\nu - g_{\mu\nu}\Box\right)\varphi'(R_T)\Bigg] = \kappa T_{\mu\nu} \,, \end{aligned} \tag{11}$$

which are valid for

$$|\epsilon\varphi(R)| \ll |R| \,. \tag{12}$$

Throughout this work, we consider the FLRW metric with a $(-,+,+,+)$ signature, which in spherical coordinates is given by

$$ds^2 = -dt^2 + a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right], \tag{13}$$

where $a(t)$ is the scale factor and k is the curvature of the universe. We assume a perfect fluid description for the content of the universe, given by the energy-momentum tensor, $T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}$, where U^μ is the four-velocity field of an observer comoving with the fluid, defined in such a way that $U_\mu U^\mu = -1$, and $\rho = \rho(t)$ and $p = p(t)$ are the fluid's energy density and isotropic pressure, respectively.

Now, by taking into account a EoS $p = w\rho$, with $-1 \leq w \leq 1$, and taking into account Eq. (13), then Eq. (11) provides the following modified Friedmann equation in $f(R)$ gravity[38]

$$\begin{aligned} H^2 = \frac{1}{3}\kappa\rho - \frac{k}{a^2} - \frac{\Lambda}{3} - \frac{\epsilon}{3}\Bigg[&\frac{1}{2}(3w+1)\varphi'(R_T)\kappa\rho + \varphi'(R_T)\Lambda + \frac{1}{2}\varphi(R_T) \\ &- \left(3\kappa\rho - \frac{9k}{a^2} - 3\Lambda\right)\varphi''(R_T)(1+w)(1-3w)\kappa\rho\Bigg], \end{aligned} \tag{14}$$

where $\varphi_T = \varphi(R_T)$. It is clear that this equation is effectively a second-order one and that we recover the standard Friedmann equation for $\epsilon = 0$. With this result, we aim to find a function $\varphi(R_T)$, such that Eq. (14) is the same as some Friedmann equation with a modified source.

3. Strategy to obtain effective actions

Here, we describe the procedure for finding effective actions, which was used in Ref. 38, in order to determine an effective action for LQC. In this context, we consider a spatially flat universe ($k = 0$) with a zero cosmological constant ($\Lambda = 0$).

In order to be in agreement with the approach leading to Eq. (1), we consider a massless scalar field ($w = 1$) so that

$$R_T = -2\kappa\rho\,. \tag{15}$$

Taking these aspects into account, then Eq. (14) becomes

$$H^2 = \frac{1}{3}\kappa\rho - \frac{\epsilon}{3}\left[2\varphi'(R_T)\kappa\rho + \frac{1}{2}\varphi(R_T) + 12\kappa^2\rho^2\varphi''(R_T)\right], \tag{16}$$

In the following sections, our goal is to determine $\varphi(R_T)$ such that Eq. (16) is the same as the modified Friedmann Eqs. (2)–(4), considered in the modified loop quantum cosmologies. We conveniently write Eqs. (2)–(4) in a general way as

$$H^2 = \frac{1}{3}\kappa\rho + \Psi(\rho)\,, \tag{17}$$

where $\Psi(\rho)$ depends on the model. Using Eq. (15) and comparing Eq. (17) with Eq. (16), the requirement is that $\varphi(R_T)$ satisfies

$$-\frac{\epsilon}{3}\left[\frac{1}{2}\varphi(R_T) - R_T\varphi'(R_T) + 3R_T^2\varphi''(R_T)\right] = \Psi(R_T)\,. \tag{18}$$

In finding the deviation term $\varphi(R_T)$, a metric $f(R)$ action of the form

$$f(R) = R + \epsilon\varphi(R) \tag{19}$$

is found, that leads to the quantum bounce which characterizes the models we are considering.

4. Effective actions for modified loop quantum cosmology models

Independently of the model of LQC, the solution to the non-homogeneous Eq. (18) consists of the sum of the solution to the corresponding homogeneous equation and its particular solution, $\varphi(R_T) = \varphi_h(R_T) + \varphi_p(R_T)$. The homogeneous equation is a second order Cauchy–Euler equation, with a characteristic equation given by $m^2 - 4m/3 + 1/6 = 0$. Solving for m, we get two real roots, $m = \frac{1}{6}(4 \pm \sqrt{10})$, meaning that the solution to the homogeneous equation is given by

$$\varphi_h(R) = c_1 R^{\frac{1}{6}(4-\sqrt{10})} + c_2 R^{\frac{1}{6}(4+\sqrt{10})}\,. \tag{20}$$

This solution is not locally given by a convergent power series. For this reason, and also for the fact that $\varphi_h(R)$ does not contribute to the full Eq. (18), we set $c_1 = c_2 = 0$, without loss of generality.[38] As such, the solutions to Eq. (18) are simply given by the particular solution in question, i.e. $\varphi(R_T) = \varphi_p(R_T)$, depending on $\Psi(R_T)$.

4.1. *Effective actions for mLQC-I*

As described in the Introduction, mLQC-I is divided in two branches. As such, we consider each of them separately and find an effective action for both.

4.1.1. *Effective action for the b_- branch*

Considering Eq. (2) for the b_- branch, the definition for $\Psi(\rho)$ given by Eq. (17) and Eq. (15), $\Psi(R_T)$ is given by

$$\Psi(R_T) = -\frac{R_T^2}{12\rho_c^I\kappa} + \frac{R_T}{6}\left(1+\frac{R_T}{2\kappa\rho_c^I}\right)\frac{\gamma^2 R_T/(2\kappa\rho_c^I)}{(\gamma^2+1)\left(1+\sqrt{1+R_T/(2\kappa\rho_c^I)}\right)^2}\,. \tag{21}$$

Therefore, in this case, Eq. (18) is given by

$$-\frac{\epsilon}{3}\left[\frac{1}{2}\varphi_T - R_T\varphi_T' + 3R_T^2\varphi_T''\right] = -\frac{R_T^2}{12\rho_c^I\kappa} + \frac{R_T}{6}\frac{\left(1+\frac{R_T}{2\kappa\rho_c^I}\right)\gamma^2 R_T/(2\kappa\rho_c^I)}{(\gamma^2+1)\left(1+\sqrt{1+R_T/(2\kappa\rho_c^I)}\right)^2}\,, \tag{22}$$

This equation was solved for $\varphi(R_T)$ using *Wolfram Mathematica*. Thus a metric $f(R)$ action, Eq. (19), was found which, when treated as an effective action, leads to the modified Friedmann Eq. (3). This effective action is given by

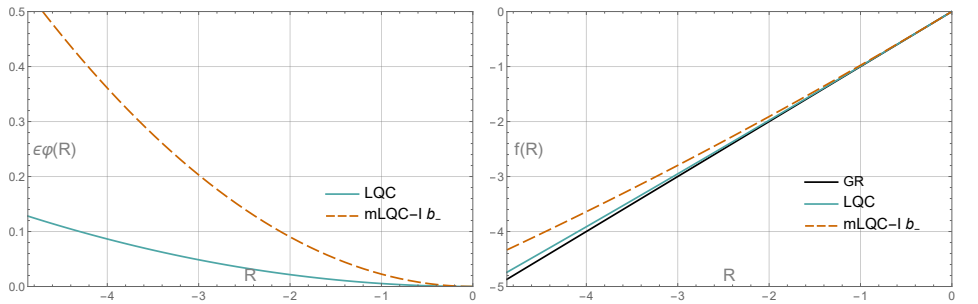
$$\begin{aligned}
f(R) = R + \frac{1}{90(\gamma^2+1)}\Bigg\{ & 5\left(-72\gamma^2\kappa\rho_c^I + \frac{R^2}{\kappa\rho_c^I} + 54\gamma^2 R\right) \\
& + \frac{18\gamma^2}{\sqrt{10}+10}\Bigg[3\left(4\left(5-4\sqrt{10}\right)\kappa\rho_c^I + \left(\sqrt{10}+10\right)R\right) \\
& \quad {}_2F_1\left[-\frac{1}{2},\frac{1}{6}\left(-\sqrt{10}-4\right);\frac{1}{6}\left(2-\sqrt{10}\right);-\frac{R}{2\kappa\rho_c^I}\right] \\
& + 3\left(12\left(2\sqrt{10}+5\right)\kappa\rho_c^I + \left(\sqrt{10}+10\right)R\right) \\
& \quad {}_2F_1\left[-\frac{1}{2},\frac{1}{6}\left(\sqrt{10}-4\right);\frac{1}{6}\left(\sqrt{10}+2\right);-\frac{R}{2\kappa\rho_c^I}\right] \\
& + (2\kappa\rho_c^I + R)\Bigg(9\sqrt{10}\,{}_2F_1\left[\frac{1}{2},\frac{1}{6}\left(-\sqrt{10}-4\right);\frac{1}{6}\left(2-\sqrt{10}\right);-\frac{R}{2\kappa\rho_c^I}\right] \\
& - \left(11\sqrt{10}+20\right)\,{}_2F_1\left[\frac{1}{2},\frac{1}{6}\left(\sqrt{10}-4\right);\frac{1}{6}\left(\sqrt{10}+2\right);-\frac{R}{2\kappa\rho_c^I}\right]\Bigg)\Bigg]\Bigg\}\,.
\end{aligned} \tag{23}$$

where we have dropped the subscript T since, to ϵ order, $f(R)$ and $f(R_T)$ are the same. This action contains hypergeometric functions, denoted by ${}_2F_1$, and its domain is given by $-2\kappa\rho_c^I \leq R$. Since it is only physically meaninful to consider $R \leq 0$, given Eq. (15), we have that our physical solution is defined for the range $-2\kappa\rho_c^I \leq R \leq 0$.

Using *Wolfram Mathematica*, a Taylor expansion of Eq. (23), about $R = 0$ and up to third order, was also determined and is given by

$$f(R) = R + \epsilon\varphi(R) = R + \frac{(4+3\gamma^2)R^2}{72(1+\gamma^2)\kappa\rho_c^I} + \frac{\gamma^2 R^3}{992(1+\gamma^2)(\kappa\rho_c^I)^2} + \cdots\,. \tag{24}$$

In order to visualize our solution, we provide two different plots in Fig. 1 regarding Eq. (23). In Fig. 1a we present the deviation terms of Eq. (23) and the deviation term of the effective action for LQC, given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$ according to Ref. 38. In Fig. 1b, we compare Eq. (23) with the effective $f(R)$ function for the case of LQC and the Lagrangian density of GR, with $\Lambda = 0$.



(a) (b)

Fig. 1. Figure 1a presents the comparison between the deviation, from GR, of the effective Lagrangian for the case of LQC (solid blue line), given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$, and the deviation terms of Eq. (23), for the b_- branch of mLQC-I (dashed dark orange line), for $-2\kappa\rho_c^I \leq R \leq 0$. Although the deviation is higher for the case of the b_- branch of this modified model, both of them are dominated by the quadratic term. Figure 1b illustrates the differences between the Lagrangian density of GR (solid black line), with $\Lambda = 0$, and the effective $f(R)$ functions for the case of LQC (solid blue line) and Eq. (23), for the b_- branch of mLQC-I (dashed dark orange line), for the interval $-2\kappa\rho_c^I \leq R \leq 0$.

4.1.2. *Effective action for the b_+ branch*

Considering Eq. (2) for the b_+ branch, the definition for $\Psi(\rho)$ given by Eq. (17) and Eq. (15), $\Psi(R_T)$ is given by

$$\Psi(R_T) = \frac{R_T}{6} + \frac{1}{3}\alpha\kappa\rho_\Lambda\left(1 + \frac{R_T}{2\kappa\rho_c^I}\right)\left[1 - \frac{R_T\left(1 - 2\gamma^2 + \sqrt{1 + R_T/(2\kappa\rho_c^I)}\right)}{8\gamma^2\kappa\rho_c^I\left(1 + \sqrt{1 + R_T/(2\kappa\rho_c^I)}\right)}\right]. \tag{25}$$

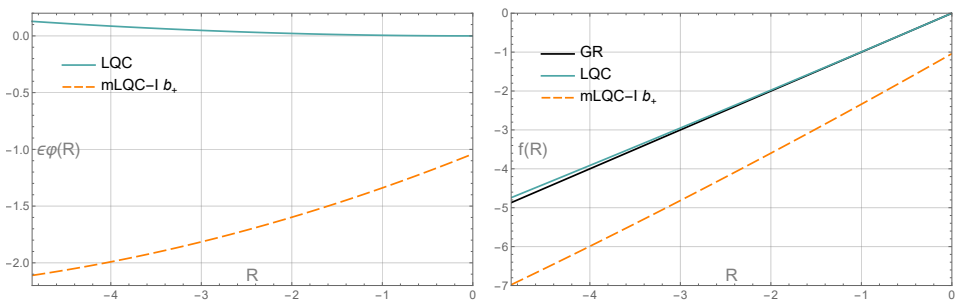
Therefore, in this case, Eq. (18) becomes

$$-\frac{\epsilon}{3}\left[\frac{\varphi_T}{2} - R_T\varphi_T' + 3R_T^2\varphi_T''\right] = \frac{R_T}{6} + \frac{\alpha\kappa\rho_\Lambda}{3}\left[1 + \frac{R_T}{2\kappa\rho_c^I}\right]\left[1 - \frac{R_T\left(1 - 2\gamma^2 + \sqrt{1 + \frac{R_T}{2\kappa\rho_c^I}}\right)}{8\gamma^2\kappa\rho_c^I\left(1 + \sqrt{1 + \frac{R_T}{2\kappa\rho_c^I}}\right)}\right], \tag{26}$$

As before, this equation was solved for $\varphi(R)$ using *Wolfram Mathematica*, and the corresponding effective action, that leads to the modified Friedmann Eq. (3), is

given by

$$
\begin{aligned}
f(R) = R - \alpha\kappa\rho_\Lambda + R + \frac{\alpha\rho_\Lambda R\left(18\left(2\gamma^2-1\right)\kappa\rho_c^I + R\right)}{72\gamma^2\kappa(\rho_c^I)^2} \\
+ \frac{\alpha\rho_\Lambda}{20\rho_c^I}\Bigg\{ -3\left[R - 2\left(\sqrt{10}-2\right)\kappa\rho_c^I\right] \\
{}_2F_1\left[-\frac{1}{2}, \frac{1}{6}\left(-\sqrt{10}-4\right); \frac{1}{6}\left(2-\sqrt{10}\right); -\frac{R}{2\kappa\rho_c^I}\right] \\
-3\left(2\left(\sqrt{10}+2\right)\kappa\rho_c^I + R\right) {}_2F_1\left[-\frac{1}{2}, \frac{1}{6}\left(\sqrt{10}-4\right); \frac{1}{6}\left(\sqrt{10}+2\right); -\frac{R}{2\kappa\rho_c^I}\right] \\
+ (2\kappa\rho_c^I + R)\Bigg(-\left(\sqrt{10}-1\right) {}_2F_1\left[\frac{1}{2}, \frac{1}{6}\left(-\sqrt{10}-4\right); \frac{1}{6}\left(2-\sqrt{10}\right); -\frac{R}{2\kappa\rho_c^I}\right] \\
+ \left(\sqrt{10}+1\right) {}_2F_1\left[\frac{1}{2}, \frac{1}{6}\left(\sqrt{10}-4\right); \frac{1}{6}\left(\sqrt{10}+2\right); -\frac{R}{2\kappa\rho_c^I}\right]\Bigg)\Bigg\}. \qquad (27)
\end{aligned}
$$



(a) (b)

Fig. 2. Figure 2a. presents the comparison between the deviation, from GR, for the case of LQC (solid blue line), given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$, and the deviation terms of Eq. (27), for the b_+ branch of mLQC-I (dashed light orange line), for $-2\kappa\rho_c^I \leq R \leq 0$. Contrary to the previous case, for this branch the deviation is predominantly negative since it is dominated by the linear term. Figure 2b. illustrates the differences between the Lagrangian density of GR (solid black line), with $\Lambda = 0$, and the effective $f(R)$ functions for the case of LQC (solid blue line) and Eq. (27), for the b_+ branch of mLQC-I (dashed light orange line), for the interval $-2\kappa\rho_c^I \leq R \leq 0$.

For the same reasons as in the previous branch, the physical domain of this solution is also given by $-2\kappa\rho_c^I \leq R \leq 0$. The Taylor expansion of Eq. (27) up to third order, about $R = 0$, was computed using *Wolfram Mathematica*, resulting in

$$
f(R) = -2\alpha\kappa\rho_\Lambda + \left(\frac{\alpha\left(5\gamma^2-1\right)\rho_\Lambda}{4\gamma^2\rho_c^I} + 2\right)R + \frac{\alpha\left(4-3\gamma^2\right)\rho_\Lambda R^2}{288\gamma^2\kappa(\rho_c^I)^2} + \frac{\alpha\rho_\Lambda R^3}{3968\kappa^2(\rho_c^I)^3} + \ldots . \qquad (28)
$$

We observe that, for this particular branch, $f(0) \neq 0$. We recall that ρ_Λ was interpreted as the energy density of an emergent positive cosmological constant in Ref. 20. Finally we provide similar plots for this branch, in Fig. 2.

4.2. *Effective action for mLQC-II*

We now consider the mLQC-II model, in which the modified Friedmann equation is given by Eq. (4). Considering the definition for $\Psi(\rho)$ given by Eq. (17) and Eq. (15), $\Psi(R_T)$ is given by

$$\Psi(R_T) = \frac{R_T}{6} - \frac{R_T\left(1 + R_T/(2\kappa\rho_c^{II})\right)\left(1 - 2\gamma^2\left(\gamma^2+1\right)R_T/(\kappa\rho_c^{II})\right)}{3\left(1 - \gamma^2 R_T/(\kappa\rho_c^{II}) + \sqrt{1 - 2\gamma^2\left(\gamma^2+1\right)R_T/(\kappa\rho_c^{II})}\right)}. \tag{29}$$

Therefore, in this case, Eq. (18) is given by

$$-\frac{\epsilon}{3}\left[\frac{1}{2}\varphi_T - R_T\varphi_T' + 3R_T^2\varphi_T''\right] = \frac{R_T}{6} - \frac{R_T\left(1 + \frac{R_T}{2\kappa\rho_c^{II}}\right)\left(1 - \frac{2\gamma^2\left(\gamma^2+1\right)R_T}{\kappa\rho_c^{II}}\right)}{3\left(1 - \frac{\gamma^2 R_T}{\kappa\rho_c^{II}} + \sqrt{1 - \frac{2\gamma^2(\gamma^2+1)R_T}{\kappa\rho_c^{II}}}\right)}. \tag{30}$$

This equation was solved for $\varphi(R)$ using *Wolfram Mathematica*. The corresponding metric $f(R)$ action which, when treated as an effective one, leads to the modified Friedmann Eq. (4) is given by

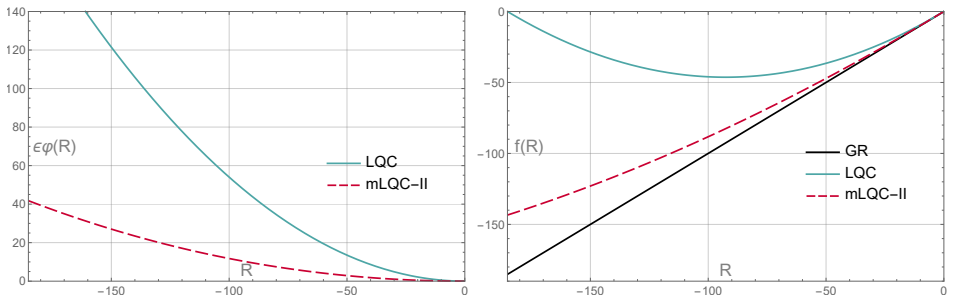
$$\begin{aligned}
f(R) = R + \frac{1}{90\gamma^4}\Bigg\{ & 270\left(\gamma^4+\gamma^2\right)R + 90\kappa\rho_c^{II} + \frac{20\left(\gamma^6+\gamma^4\right)R^2}{\kappa\rho_c^{II}} \\
& + 9\Bigg[3\left(\left(\sqrt{10}-2\right)\kappa\rho_c^{II} + 2\left(\gamma^4+\gamma^2\right)R\right) \\
& \quad {}_2F_1\left[-\frac{1}{2}, \frac{1}{6}\left(-\sqrt{10}-4\right); \frac{1}{6}\left(2-\sqrt{10}\right); \frac{2R\gamma^2\left(\gamma^2+1\right)}{\kappa\rho_c^{II}}\right] \\
& + \left(6\left(\gamma^4+\gamma^2\right)R - 3\left(\sqrt{10}+2\right)\kappa\rho_c^{II}\right) \\
& \quad {}_2F_1\left[-\frac{1}{2}, \frac{1}{6}\left(\sqrt{10}-4\right); \frac{1}{6}\left(\sqrt{10}+2\right); \frac{2R\gamma^2\left(\gamma^2+1\right)}{\kappa\rho_c^{II}}\right] \\
& + \left(2\left(\gamma^4+\gamma^2\right)R - \kappa\rho_c^{II}\right)\Bigg(\left(\sqrt{10}-1\right) \\
& \quad {}_2F_1\left[\frac{1}{2}, \frac{1}{6}\left(-\sqrt{10}-4\right); \frac{1}{6}\left(2-\sqrt{10}\right); \frac{2R\gamma^2\left(\gamma^2+1\right)}{\kappa\rho_c^{II}}\right] \\
& - \left(\sqrt{10}+1\right) {}_2F_1\left[\frac{1}{2}, \frac{1}{6}\left(\sqrt{10}-4\right); \frac{1}{6}\left(\sqrt{10}+2\right); \frac{2R\gamma^2\left(\gamma^2+1\right)}{\kappa\rho_c^{II}}\right]\Bigg)\Bigg]\Bigg\}.
\end{aligned} \tag{31}$$

This solution also contains hypergeometric functions but, for this model, they present a different argument. According to this argument the domain of Eq. (31) is given by $R \leq \kappa\rho_c^{II}/\left(2\gamma^2(\gamma^2+1)\right)$. As such, since the Ricci scalar must be less or equal to zero for a massless sclar field scenario, in this case the solution only has physical meaning for the interval $R \leq 0$.

Using *Wolfram Mathematica*, a Taylor expansion of Eq. (31), about $R = 0$ and up to third order, was computed and is given by

$$f(R) = R + \frac{\left(-3\gamma^4 - 2\gamma^2 + 1\right) R^2}{18\kappa\rho_c^{II}} - \frac{\gamma^2\left(\gamma^2+1\right)^3 R^3}{62(\kappa\rho_c^{II})^2} + \dots\,, \tag{32}$$

The solution for the mLQC-II model is depicted in Fig. 3. In Fig. 3a, we present the deviation terms of Eq. (31) and the deviation term of the effective action for LQC, given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$ according to Ref. 38. In Fig. 3b, we compare Eq. (31) with the effective $f(R)$ function for the case of LQC and the Lagrangian density of GR, with $\Lambda = 0$. As a summary, a combination of our results is presented in Fig. 4.



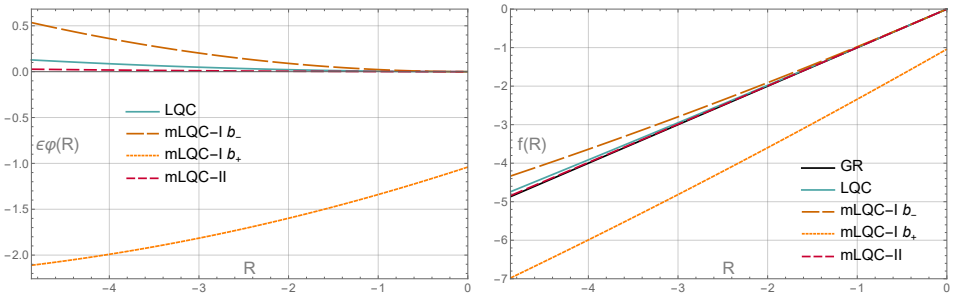
(a) (b)

Fig. 3. Figure 3a. presents the comparison between the deviation, from GR, of the effective Lagrangian for the case of LQC (solid blue line), given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$, and the deviation terms of Eq. (31), for the case of mLQC-II (dashed pink line), for $-18\kappa\rho_c \leq R \leq 0$. Both of the deviations are dominated by the quadratic term. Figure 3b. illustrates the differences between the Lagrangian density of GR (solid black line), with $\Lambda = 0$, and the effective $f(R)$ functions, for the case of LQC (solid blue line), and Eq. (31), for mLQC-II (dashed pink line), for the interval $-18\kappa\rho_c \leq R \leq 0$.

5. Conclusions

In this work, we have addressed the initial singularity problem that is present in the ΛCDM model and considered a possible resolution to this question, as proposed by LQC, in which the Big Bang is replaced by a quantum bounce. Since this result comes from a different field in physics, it is relevant to ask if it is possible to replicate it in the framework of GR and its modifications. In this context we considered the class of metric $f(R)$ gravity. Motivated by Ref. 38, we have used the same procedure with two modifications of standard LQC, that also yield a quantum bounce, which are two attempts to bridge the gap between LQG and LQC, by means of similar treatments of the Hamiltonian constraint.

We have applied the covariant order reduction method to obtain a reduced version of the full field equations in $f(R)$ gravity. The corresponding modifed



(a) (b)

Fig. 4. Figure 4a. presents the comparison between the deviation, from GR, of the effective Lagrangian for the case of LQC (solid blue line), given by $\epsilon\varphi(R) = R^2/(18\kappa\rho_c)$, and the deviation terms of Eq. (23), for the b_- branch of mLQC-I (dashed dark orange line), Eq. (27), for the b_+ branch of mLQC-I (dashed light orange line) and Eq. (31), for mLQC-II (dashed pink line), for $-2\kappa\rho_c^I \leq R \leq 0$. Figure 4b. presents the differences between the Lagrangian density of GR (solid black line), with $\Lambda = 0$, and the effective $f(R)$ functions for the case of LQC (solid blue line), Eq. (23), for the b_- branch of mLQC-I (dashed dark orange line), Eq. (27), for the b_+ branch of mLQC-I (dashed light orange line) and Eq. (31), for mLQC-II (dashed pink line), for $-2\kappa\rho_c^I \leq R \leq 0$.

Friedmann equation depends on the $f(R)$ function, which we chose to parametrize as $f(R) = R + \epsilon\varphi(R)$. In this framework, we have determined functions $\varphi(R)$, such that Eqs. (2)–(4) are the same as Eq. (18). In other words, we have determined effective covariant $f(R)$ actions which lead to Eqs. (2)–(4). Moreover, these effective actions allow for a positive effective gravitational coupling constant and avoid the Dolgov–Kawasaki instability.[25] According to Eq. (12), the approximation that we use is valid for $|R| \ll \rho_c \sim l_p^{-2}$, therefore, breaking down at the bounce, since there $|R| \sim \rho_c \sim l_p^{-2}$. Thus, the argument that the effective Lagrangians deduced in this work yield Eqs. (2)–(4) is not quite a rigorous statement. However, our claim is that, at stages close enough to the bounce, the solutions found in this work may effectively, and correctly, describe the collapsing universe.

If a quantum theory of gravity is found, a proper description of the beginning of the Universe could eventually be reached. In this regard, the ΛCDM model is incomplete and, for this reason, there is motivation for attempting modifications of GR, in order to accommodate scenarios in which the Big Bang singularity is non-existent, such as the covariant effective actions determined in this work. In a forthcoming paper we will extend our analysis for mLQC-II to arbitrary values of w.

Acknowledgements

DV acknowledges support from the *Istituto Nazionale di Fisica Nucleare* (INFN) (*iniziativa specifica* TEONGRAV). FSNL acknowledges support from the Fundação para a Ciência e a Tecnologia (FCT) Scientific Employment Stimulus contract with reference CEECINST/00032/2018, and funding from the research grants

No. UID/FIS/04434/2020, No. PTDC/FIS-OUT/29048/2017 and No. CERN/FIS-PAR/0037/2019.

References

1. G. Amelino-Camelia, Quantum-Spacetime Phenomenology, *Living Rev. Rel.* **16**, p. 5 (2013).
2. S. Carlip, Quantum gravity: A Progress report, *Rept. Prog. Phys.* **64**, p. 885 (2001).
3. C. Rovelli, Loop quantum gravity, *Living Rev. Rel.* **1**, p. 1 (1998).
4. T. Thiemann, Modern canonical quantum general relativity (2001).
5. A. Ashtekar and J. Lewandowski, Background independent quantum gravity: A Status report, *Class. Quant. Grav.* **21**, p. R53 (2004).
6. M. Bojowald, Loop quantum cosmology, *Living Rev. Rel.* **8**, p. 11 (2005).
7. M. Bojowald, Absence of singularity in loop quantum cosmology, *Phys. Rev. Lett.* **86**, 5227 (2001).
8. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang: An Analytical and Numerical Investigation. I., *Phys. Rev. D* **73**, p. 124038 (2006).
9. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang: Improved dynamics, *Phys. Rev. D* **74**, p. 084003 (2006).
10. V. Taveras, Corrections to the Friedmann Equations from LQG for a Universe with a Free Scalar Field, *Phys. Rev. D* **78**, p. 064072 (2008).
11. K. Banerjee, G. Calcagni and M. Martin-Benito, Introduction to loop quantum cosmology, *SIGMA* **8**, p. 016 (2012).
12. J. De Haro and J. Amorós, Bouncing cosmologies via modified gravity in the ADM formalism: Application to Loop Quantum Cosmology, *Phys. Rev. D* **97**, p. 064014 (2018).
13. P. Singh, Loop cosmological dynamics and dualities with Randall-Sundrum braneworlds, *Phys. Rev. D* **73**, p. 063508 (2006).
14. K. A. Meissner, Black hole entropy in loop quantum gravity, *Class. Quant. Grav.* **21**, 5245 (2004).
15. J. Yang, Y. Ding and Y. Ma, Alternative quantization of the Hamiltonian in loop quantum cosmology II: Including the Lorentz term, *Phys. Lett. B* **682**, 1 (2009).
16. A. Dapor and K. Liegener, Cosmological Effective Hamiltonian from full Loop Quantum Gravity Dynamics, *Phys. Lett. B* **785**, 506 (2018).
17. A. Dapor and K. Liegener, Cosmological coherent state expectation values in loop quantum gravity I. Isotropic kinematics, *Class. Quant. Grav.* **35**, p. 135011 (2018).
18. E. Alesci, A. Barrau, G. Botta, K. Martineau and G. Stagno, Phenomenology of Quantum Reduced Loop Gravity in the isotropic cosmological sector, *Phys. Rev. D* **98**, p. 106022 (2018).
19. J. Bilski and A. Marcianò, Critical Insight into the Cosmological Sector of Loop Quantum Gravity, *Phys. Rev. D* **101**, p. 066026 (2020).
20. B.-F. Li, P. Singh and A. Wang, Towards Cosmological Dynamics from Loop Quantum Gravity, *Phys. Rev. D* **97**, p. 084029 (2018).
21. B.-F. Li, P. Singh and A. Wang, Genericness of pre-inflationary dynamics and probability of the desired slow-roll inflation in modified loop quantum cosmologies, *Phys. Rev. D* **100**, p. 063513 (2019).
22. B.-F. Li, P. Singh and A. Wang, Qualitative dynamics and inflationary attractors in loop cosmology, *Phys. Rev. D* **98**, p. 066016 (2018).
23. G. J. Olmo and P. Singh, Effective Action for Loop Quantum Cosmology a la Palatini, *JCAP* **01**, p. 030 (2009).

24. G. J. Olmo, Palatini Approach to Modified Gravity: f(R) Theories and Beyond, *Int. J. Mod. Phys. D* **20**, 413 (2011).
25. T. P. Sotiriou and V. Faraoni, f(R) Theories Of Gravity, *Rev. Mod. Phys.* **82**, 451 (2010).
26. F. S. N. Lobo, The Dark side of gravity: Modified theories of gravity (7 2008).
27. A. De Felice and S. Tsujikawa, f(R) theories, *Living Rev. Rel.* **13**, p. 3 (2010).
28. S. Nojiri and S. D. Odintsov, Unified cosmic history in modified gravity: from F(R) theory to Lorentz non-invariant models, *Phys. Rept.* **505**, 59 (2011).
29. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, Modified Gravity and Cosmology, *Phys. Rept.* **513**, 1 (2012).
30. S. Capozziello and M. De Laurentis, Extended Theories of Gravity, *Phys. Rept.* **509**, 167 (2011).
31. T. Harko and F. S. N. Lobo, *Extensions of f(R) Gravity: Curvature-Matter Couplings and Hybrid Metric-Palatini Theory* (Cambridge University Press, 11 2018).
32. L. Bel and H. S. Zia, Regular reduction of relativistic theories of gravitation with a quadratic Lagrangian, *Phys. Rev. D* **32**, 3128 (1985).
33. J. Z. Simon, Higher Derivative Lagrangians, Nonlocality, Problems and Solutions, *Phys. Rev. D* **41**, p. 3720 (1990).
34. J. Z. Simon, No Starobinsky inflation from selfconsistent semiclassical gravity, *Phys. Rev. D* **45**, 1953 (1992).
35. I. Terrucha, D. Vernieri and J. P. S. Lemos, Covariant action for bouncing cosmologies in modified Gauss–Bonnet gravity, *Annals Phys.* **404**, 39 (2019).
36. B. J. Barros, E. M. Teixeira and D. Vernieri, Bouncing cosmology in $f(R, \mathcal{G})$ gravity by order reduction, *Annals Phys.* **419**, p. 168231 (2020).
37. F. Bajardi, D. Vernieri and S. Capozziello, Bouncing Cosmology in f(Q) Symmetric Teleparallel Gravity, *Eur. Phys. J. Plus* **135**, p. 912 (2020).
38. T. P. Sotiriou, Covariant Effective Action for Loop Quantum Cosmology from Order Reduction, *Phys. Rev. D* **79**, p. 044035 (2009).

Probing multiverse using gravitational wave observations

Moe Kukihara* and Kazuhiro Hayama†

Department of Applied Physics, Fukuoka University, Fukuoka, 814-0180, Japan

**E-mail: sd211003@cis.fukuoka-u.ac.jp*

†*E-mail: hayama@fukuoka-u.ac.jp*

From the theory of the multiverse cosmology, it is possible that our universe collides with other universes locally in its history, which may result in local changes of the curvature of the spacetime. In this paper, we propose a method to probe the multiverse using gravitational wave observations for the first time. Our method firstly makes triangles using two detected gravitational wave sources and the Sun, and then measures the curvature of the triangles. We use 11 gravitational wave sources detected by LIGO and Virgo during O1 and O2, and make 55 triangles by combining them to measure their curvature. The curvature is measured by comparing the distance between two gravitational wave sources estimated by the gravitational wave observations with the one obtained with assumption of a simple model of the cosmological evolution.

As a result, we found that, for 43 of 55 triangles, the distances estimated by the model are greater than the ones obtained by the gravitational wave observations. This indicates a negative curvature, which may be due to the simplification of the cosmological evolution. For the rest 12, the distances are not determined because of uncertainty of the parameters of the gravitational wave observations. Further gravitational wave observations and more sophisticated model of the cosmological evolution is essential to test the multiverse cosmology observationally.

Keywords: Gravitational wave observations, multiverse, spacetime geometry, cosmological evolution

1. Introduction

Since E. Hubble discovered the expansion of the universe in 1929,[1] it has been considered of the possibility that the beginning of the universe was a small point. G. Gamow proposed the Big Bang theory, from which the beginning of the universe was a fireball. The Big Bang theory was supported by observations such as the first observation of the cosmic microwave background radiation in 1965[2] and the discovery of temperature fluctuations originating in the early universe by the cosmic background radiation explorer COBE in 1989.[3] However, the Big Bang theory has problems such as the homogeneity and the flatness problem. Inflation theory proposed to solve these problems,[4–6] but inflation theory also has following problems: The average value of the vacuum energy density predicted by the field theory is $\rho_\Lambda \sim 10^{113}$ J/m^3, whereas the currently observed vacuum energy density is, $\rho_\Lambda \sim 10^{-9}$ J/m^3. There is a difference of about 120 digits in size between the theoretical value and the observed value.[7]

Multiverse cosmology may solve these problems. In the multiverse cosmology, each universe is allowed to have different vacuum energy density, and our universe happened to have current vacuum energy densities. Thus, in the multiverse cosmology, the formation of multiple universes is naturally predicted. It is possible that our universe collides with other universes locally and there is a local change of curvature at the collision surface.[8] The measurement of the curvature performed by the cosmic microwave background radiation observation satellite WMAP[9,10] showed that the curvature of the entire universe was flat. But there were uncertainties that the vortex-shaped polarization component were not discovered by CMB polarization observations, and therefore the determination of the curvature needs to be further refined in future observations.

In order to find whether there has been collisions between our universe and other universes in the past, There are studies that explore the temperature fluctuations that occur at the collisions surface. This temperature fluctuation appears as a ring pattern on the cosmic microwave background radiation. Several radio observations were conducted to see the properties,[11,12] and as a result, four sky areas were reported from the WMAP observations that were consistent with the ring patterns that appeared due to collision between the universes.[13,14] It should be noted that as the universe is expanded with time, the effects of collisions tend to be diluted as much as the structure of the initial universe.

In this paper, we proposed for the first time a method for probing the multiverse using gravitational wave observations, which is independent of radio wave observations performed so far.[15] Chapter 2 describes how to measure the local curvature of the universe, and Chapter 3 describes the results and discussion.

2. Measurement of the local curvatures of the universe

We will explain how we estimate local curvatures of the universe.The gravitational waves GW150914, GW151012, GW151226, GW170104, GW170608, GW170729, GW170809, GW170814, GW170817, GW170818, GW170823 were detected by the LIGO and Virgo observations.[16] Two of the 11 gravitational wave sources and the sun are used to form a triangle, and the distance between the two gravitational wave sources is estimated by gravitational wave observations.The distances are also estimated by simple assumption that the gravitational wave sources were born in the early universe and the distance between the two sources is the size of the flat universe. The local curvature of the universe was measured by comparing these values.

2.1. *Distance between gravitational waves by gravitational wave observations*

The sky locations of the almost all detected gravitational waves are highly uncertain. The summary plots in the paper B. P. Abbott, *et al.*[16] which show 50% and 90% credible area for the sky locations of detected gravitational wave events and most

of the events have error regions with several hundred square degrees. But the error regions are estimated to be reduced in future observations. In this paper we regard the sky locations of the gravitational waves as the positions with the highest probabilities in the contours.

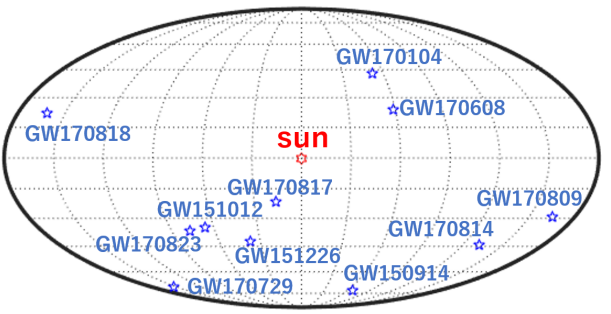


Fig. 1. Most probable sky locations of the detected gravitational waves in 2015-2017 in the Mollweide coordinates.[16]

index	combination
area 1	GW150914,GW151012
area 2	GW150914,GW151226
area 3	GW150914,GW170104
area 4	GW150914,GW170608
area 5	GW150914,GW170729
area 6	GW150914,GW170809
area 7	GW150914,GW170814
area 8	GW150914,GW170817
area 9	GW150914,GW170818
area 10	GW150914,GW170823
area 11	GW151012,GW151226
area 12	GW151012,GW170104
area 13	GW151012,GW170608
area 14	GW151012,GW170729
area 15	GW151012,GW170809
area 16	GW151012,GW170814
area 17	GW151012,GW170817
area 18	GW151012,GW170818
area 19	GW151012,GW170823
area 20	GW151226,GW170104
area 21	GW151226,GW170608
area 22	GW151226,GW170729
area 23	GW151226,GW170809
area 24	GW151226,GW170814
area 25	GW151226,GW170817
area 26	GW151226,GW170818
area 27	GW151226,GW170823
area 28	GW170104,GW170608

index	combination
area 29	GW170104,GW170729
area 30	GW170104,GW170809
area 31	GW170104,GW170814
area 32	GW170104,GW170817
area 33	GW170104,GW170818
area 34	GW170104,GW170823
area 35	GW170608,GW170729
area 36	GW170608,GW170809
area 37	GW170608,GW170814
area 38	GW170608,GW170817
area 39	GW170608,GW170818
area 40	GW170608,GW170823
area 41	GW170729,GW170809
area 42	GW170729,GW170814
area 43	GW170729,GW170817
area 44	GW170729,GW170818
area 45	GW170729,GW170823
area 46	GW170809,GW170814
area 47	GW170809,GW170817
area 48	GW170809,GW170818
area 49	GW170809,GW170823
area 50	GW170814,GW170817
area 51	GW170814,GW170818
area 52	GW170814,GW170823
area 53	GW170817,GW170818
area 54	GW170817,GW170823
area 55	GW170818,GW170823

Fig. 2. 55 combinations of two gravitational sources for making triangles including the Sun.

Figure 1 shows the sky positions of the gravitational wave sources used here and the sun. we constructed triangles using the sun and two of the gravitational wave sources. The distances between the two gravitational wave sources were estimated using the estimated distances and sky locations of the gravitational wave sources in the triangle. Figure 2 shows 55 combinations of two gravitational sources for making triangles including the Sun.

2.2. *Distance between gravitational waves by a flat universe model*

We consider of the following four-dimensional metric that describes a uniform and isotropic expansion universe.

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 sin^2\theta d\phi^2)\right] \tag{1}$$

c: speed of light, a(t): Scale factor at time t, r: Radius in 3D spherical coordinates, θ, ϕ: Declination in 3D spherical coordinates, K: constant

Using the natural unit system, the metric of the flat universe model is,

$$ds^2 = -dt^2 + a^2(t)[dr^2 + r^2(d\theta^2 + sin^2\theta d\phi^2)] \tag{2}$$

When a gravitational wave with a redshift z is observed, the relation of the scale factor and z is as follows:

$$\frac{a(t)}{a(t_0)} = \frac{1}{1+z} \tag{3}$$

Here, when we estimate the distance between two gravitational wave sources in the flat universe model, we put simple assumption that the gravitational wave sources occurred in the early universe and the evolution of the universe are not affected by gravity from other objects such as dark matter and galaxies.[17,18] z is the average of the redshifts of the two gravitational wave sources, and the distances between the two points are estimated to be the size of the universe at z. Hereafter we refer this as the simple universe model.

2.3. *Comparison of the estimated distances for 55 triangular areas*

Figure 3 shows the distances between the sources by gravitational wave observations and the distances between the sources by the simple universe model for each triangular area. The blue cross marks show the value distances estimated from the gravitational wave observations. The error bars reflect the uncertainties of the distances of the gravitational wave sources and the uncertainties of the sky locations of the gravitational wave sources. The red asterisks are the distances between the two gravitational wave sources in the simple universe model.

Figure 4 shows the curvature of each area. The curvature of each area was estimated by comparing the distances estimated by the two ways. If the two estimated

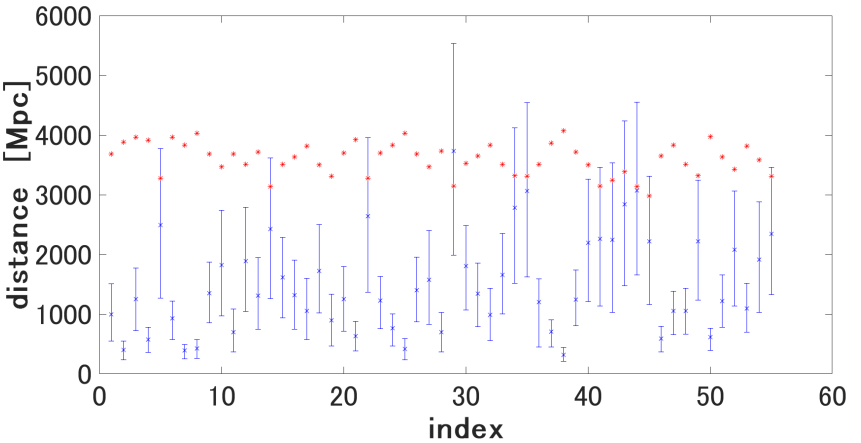


Fig. 3. Blue cross marks show the distances between gravitational wave sources estimated by the gravitational wave observations with uncertainty. Red asterisks show the distances calculated using a simple model of the cosmological evolution.

index	curvature
area 1	negative
area 2	negative
area 3	negative
area 4	negative
area 5	Indistinguishable
area 6	negative
area 7	negative
area 8	negative
area 9	negative
area 10	negative
area 11	negative
area 12	negative
area 13	negative
area 14	Indistinguishable
area 15	negative
area 16	negative
area 17	negative
area 18	negative
area 19	negative
area 20	negative
area 21	negative
area 22	Indistinguishable
area 23	negative
area 24	negative
area 25	negative
area 26	negative
area 27	negative
area 28	negative

index	curvature
area 29	Indistinguishable
area 30	negative
area 31	negative
area 32	negative
area 33	negative
area 34	Indistinguishable
area 35	Indistinguishable
area 36	negative
area 37	negative
area 38	negative
area 39	negative
area 40	negative
area 41	Indistinguishable
area 42	Indistinguishable
area 43	Indistinguishable
area 44	Indistinguishable
area 45	Indistinguishable
area 46	negative
area 47	negative
area 48	negative
area 49	negative
area 50	negative
area 51	negative
area 52	negative
area 53	negative
area 54	negative
area 55	Indistinguishable

Fig. 4. Curvatures of the 55 triangles.

distances are equal each other, then the local curvature of the universe is zero. If the distance estimated by gravitational wave observations is larger than the distance estimated by the simple universe model, the positive curvature is suggested, and if smaller, the negative curvature is suggested.

In the 43 areas among 55 areas, the distances estimated by the simple universe model were larger than the distances estimated by gravitational wave observations, so negative curvatures were suggested. On the other hands area 5, area 14, area 22, area 29, area 34, area 35, area 41, area 42, area 43, area 44, area 45, area 55 in this 12 areas, the estimated distances estimated by the simple universe model were in the error bars of the distances estimated by gravitational wave observations, so the curvatures were not estimated from all at the gravitational wave observations.

3. Summary

In this study, we proposed for the first time a method to probe the multiverse cosmology using gravitational wave observations. We estimated the local curvatures of various areas in the universe by measuring the sums of the internal angles of the 55 triangles formed by the sun and the all combinations of 11 gravitational wave sources.

In the 43 areas among 55 areas, the distances estimated by the simple universe model were larger than the distances estimated by gravitational wave observations, so negative curvatures were suggested. This is due to the simplification of the universe model, and it is necessary to sophisticate the model of the universe by including the effects of other astronomical objects, dark matter, dark energy, and so on. On the other hands in the area 5, area 14, area 22, area 29, area 34, area 35, area 41, area 42, area 43, area 44, area 45, area 55 in this 12 areas, the distances estimated by the simple universe model were in the error bar of the distances estimated by gravitational wave observations, so the curvatures could not be estimated at the current gravitational wave observations. In the future, it will be important to reduce the error in order to limit the sign of the curvature.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K03896. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (`https://www.gw-openscience.org`) , a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionaledella Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

References

1. E. Hubble, PNAS, **15**, 168-173 (1929).
2. A. A. Penzias, R. W. Willson, Astrophysical Journal, **142**, 419-421 (1965).
3. G. F. Smoot, *et al.*, Astrophysical Journal Letters, **396**, p.L1 (1992).
4. K. Sato, Physics Letters., **99**, 1, 66-70 (1981).
5. K. Sato, Monthly Notices of the Royal Astronomical Society, **195**, 467-479 (1981).
6. B. Einhorn Martin, K. Sato, Nuclear Physics B, **180**, 3, 385-404 (1981).
7. S. Weinberg, The cosmological constatnt problems, Red. Mod. Phys. **61**, 1 (1989).
8. A. Lue, G. Starkman, Physical Review D, **67**, 6, 064002 (2003).
9. A. Riess, *et al.*, Astrophys. J., **560**, 49 (2001).
10. D. N. Spergel, *et al.*, The Astrophysical Journal Supplement Series, **148**, 1, 175-194 (2003).
11. Spencer Chang, *et al.*, Journal of Cosmology and Astroparticle Physics, **32**, 4, 025 (2009).
12. M. Kleban, T. S. Levi, K. Sigurdson, Physical Review D, **87**, 4, 041301 (2013).
13. S. M. Feeney, *et al.*, Physical Review Letters, **107**, 7, 071301 (2011).
14. S. M. Feeney, *et al.*, Physical Review D, **84**, 4, 043507 (2011).
15. A. Gupta, D. Fox, B. S. Sathyaprakash and B. F. Schutz, The Astrophys. J., **886**, 1 (2019).
16. B. P. Abbott, *et al.*, Physical Review X, **9**, 3, 031040 (2019).
17. V. Bromm, N. Yoshida, C. McKee and L. Hernquist, Nature, **459**, 49 (2009).
18. V. Bromm and N. Yoshida, Annual Reviews of Astronomy and Astrophysics, **49**, 373 (2011).

Operator ordering ambiguity in observables of quantum cosmology

Harkirat Singh Sahota*
Kinjalk Lochan†
Department of Physical Sciences, Indian Institute of Science Education & Research (IISER) Mohali, Sector 81 SAS Nagar, Manauli PO 140306 Punjab India.
**ph17078@iisermohali.ac.in, †kinjalk@iisermohali.ac.in*

We discuss the status of observables and operator ordering ambiguity in the quantum cosmology model with Brown-Kuchař dust as the matter field. In order to study the dynamics of the FLRW universe, Hubble parameter and Ricci scalar are expressed as a function of phase space variables. As these functions exhibit operator ordering ambiguity, we write several Hermitian extensions corresponding to these observables. For the unitarily evolving semiclassical wave packet constructed in,[1] we have computed the expectation value of these observables, which shows that very early in the collapsing branch and very late in the expanding branch, the expectation values of the Hubble parameter and the Ricci scalar matches the classically obtained results irrespective of the operator ordering chosen. The expectation value of the Hubble parameter vanishes, and Ricci scalar attains an extremum at the point of classical singularity for all orderings, showing a robust singularity resolution. The signature of the operator ordering ambiguity is most pronounced at the classical singularity. For Weyl ordering, the expectation value of the Ricci scalar becomes negative for certain parameter values. We have computed the expectation value of other curvature invariants as well, which follows the trend.

Keywords: FLRW Model, Wheeler-DeWitt equation, Operator ordering ambiguity, Quantum Cosmology, Brown-Kuchař dust

1. Introduction

Any consistent quantization scheme for gravity must address the longstanding issues that plague the theory, like observables in quantum gravity or operator ordering ambiguity. In this work we will address the issue of operator ordering ambiguity in the observables of quantum gravity, in the context of flat FLRW model with Brown-Kuchař dust.[2]

General relativity is an example of singular systems, which has diffeomorphisms and time reparametrizations as gauge freedom.[3] A consistent (Dirac) observable in the general relativity must be invariant under diffeomorphisms and time reparametrizations. Now the systems with time reparametrization symmetry are tricky to handle, as the Hamiltonian of such systems itself is a constraint. Which would means that physical observables in the theory would be frozen in time or are constant of motion. To circumvent this issue, a possible resolution is proposed by Kuchař,[4] a phase space function does not need to have weakly vanishing Poisson bracket with Hamiltonian constraint to be an observable of general relativity.

The special objects which have vanishing Poisson bracket with Hamiltonian constraint are called a perennials.

Upon quantization, one would ideally like to obtain a self-adjoint extension of all the relevant observables appearing in the theory and study their spectral properties. But if one is interested in the expectation value of the operators in a wave packet, Hermiticity of the operator is sufficient to ensure the reality of the expectation value. For that reason we will concern ourselves with the Hermitian extensions of the observables and their expectation value for the wave packets constructed in the quantum model.

In this analysis, we will follow Kuchař's prescription[4] of observables in the quantum gravity models. For the FLRW model we will write the Hermitian extension of the operators that correspond to the Hubble parameter and the curvature invariants. We will find the expectation value of these operators and show the robustness of the singularity resolution claimed in the aforementioned quantum cosmology model. Apart from that, we will also address the operator ordering ambiguity in these observables and discuss at what stage this ambiguity will play a role.

2. FLRW Model with Brown-Kuchař Dust

We will start with the canonical formulation of the flat FLRW model coupled to Brown-Kuchař dust. We will first write the Hamiltonian constraint for this system and then write the phase space expression for the observables we are interested in.

The line element and the Ricci scalar for a homogeneous and isotropic flat FLRW spacetime is

$$ds^2 = -\mathcal{N}^2(t)dt^2 + a^2(t)d\mathbf{x}^2, \tag{1}$$

$$\mathcal{R} = \frac{6}{\mathcal{N}^2}\left[-\frac{\dot{\mathcal{N}}\dot{a}}{\mathcal{N}a} + \frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right], \tag{2}$$

where $a(\tau)$ is the scale factor and $\mathcal{N}$ is the lapse function. The action for this model with Brown-Kuchař dust[2,5] written in ADM form is,

$$S = S_G + S_m = \int \left(p\dot{a} + p_\tau\dot{\tau} - \mathcal{N}(H^G + H^D)\right) dt. \tag{3}$$

The Hamiltonian constraint for this model is given by,

$$\mathsf{H} = H^G + H^D = -\frac{\kappa^2}{6V_0}\frac{p_a^2}{2a} + p_\tau \approx 0. \tag{4}$$

In the further analysis, we will choose $\kappa^2 = 6V_0$. The solution of the Friedmann's equations with dust as matter is $a(t) \propto t^{2/3}$. The dust proper time τ and the coordinate time t are related as $\dot{\tau} = 1 \implies \tau(t) = t + \text{const}$. This model exhibits curvature singularity and has two disjointed solutions, representing a universe expanding from a Big Bang singularity and a universe collapsing to a Big Crunch singularity.

2.1. *Hubble Parameter and Ricci Scalar*

We are interested in the canonical expression for the Hubble parameter and the Ricci scalar. Using defining equation for the momentum conjugate to scale factor and the Hamilton's equations of motion, we get

$$\dot{a} = -\frac{p_a \mathcal{N}}{a}, \tag{5}$$

$$\ddot{a} = -\frac{p_a \dot{\mathcal{N}}}{a} - \frac{p_a^2 \mathcal{N}^2}{2a^3}. \tag{6}$$

Since we are working in comoving gauge $\mathcal{N} = 1$, the Hubble parameter takes the form

$$\mathbb{H} = \frac{\dot{a}}{a} = -a^{-2} p_a. \tag{7}$$

From (2), (5) and (6), the Ricci scalar is given by,

$$\mathcal{R} = \frac{3p_a^2}{a^4}. \tag{8}$$

Classically for dust as matter source, the Ricci scalar is $\mathcal{R} = 4/3t^2$. The flat FLRW model with dust as matter has curvature singularity at $t = 0$ when $a \sim t^{2/3} \to 0$ and $\mathcal{R} \to \infty$.

2.2. *Riemann and Kretschmann Scalar*

Riemann Scalar for this model is

$$\mathcal{R}ie = R_{\mu\nu} R^{\mu\nu} = 9\frac{(\dot{a}\dot{\mathcal{N}} - \mathcal{N}\ddot{a})^2}{a^2\mathcal{N}^6} + 3\frac{(a\mathcal{N}\ddot{a} - a\dot{a}\dot{\mathcal{N}} + 2\mathcal{N}\dot{a}^2)^2}{a^4\mathcal{N}^6} = \frac{3p_a^4}{a^8}. \tag{9}$$

Kretschmann Scalar for this model is,

$$K = R_{\mu\nu\alpha\beta} R^{\mu\nu\alpha\beta} = 12\left(\frac{(\dot{a}\dot{\mathcal{N}} - \mathcal{N}\ddot{a})^2}{\mathcal{N}^6 a^2} + \frac{\dot{a}^4}{a^4\mathcal{N}^4}\right) = \frac{15p_a^4}{a^8} = 5\mathcal{R}ie. \tag{10}$$

Classically for dust as matter, Riemann and Kretschmann scalar takes the form $\mathcal{R}ie = 16/27t^4$, and $K = 80/27t^4$. These curvature invariants also diverge at $t = 0$.

3. Quantum Cosmology

Quantization of this system is done by implementation of the constraint $\mathsf{H} \approx 0$ as supplementary condition on wave functions. Brown-Kuchař dust appears as natural clock for the quantum theory and we are essentially working with dust proper time. Wheeler Dewitt equation for this model takes the form of Schrödinger equation,

$$i\frac{\partial \Psi(R,\tau)}{\partial \tau} = \hat{H}\Psi(a,\tau), \tag{11}$$

$$\hat{H} = \hbar^2 a^{-1+p+q} \frac{d}{da} a^{-p} \frac{d}{da} a^{-q}. \tag{12}$$

Since the Hamiltonian is a product of non-commuting variables, therefore, it does not have a unique quantum counterpart and the model exhibits operator ordering ambiguity. The parameters p and q represent our freedom to choose operator ordering. The eigenvalue of Hamiltonian can be interpreted as Misner-Sharp mass which on-shell is related to the energy of the dust[a]. The self-adjoint extensions of Hamiltonian (12) are discussed in.[1]

To make the momentum operator $\hat{P}$ Hermitian on the real half line $\mathbb{R}^+$, we choose measure a^2 and the momentum operator that is symmetric with this measure $\hat{p}_a = -ia^{-1}\partial_a a$. Therefore, to accommodate hermiticity of the momentum, we have to impose the constraint $1 - p - 2q = 2$ or $p + 2q = -1$ on operator ordering parameters. Using this constraint, we can eliminate one of these parameters. It is shown in,[6] the expectation values of a general observables in a general wave packet constructed in this model is independent of parameter q and it appears as a free parameter in the model. Therefore, the above constraint does not put any restriction on the physical content of the theory. The detailed discussion about the Hermiticity and the self-adjointness of momentum operator on real half-line can be found in.[6] The positive energy stationary states with this choice are,

$$\phi^1_E(a) = \frac{2}{\sqrt{3}} E^{\frac{1}{4}} J_{\frac{2|q|}{3}}\left(\frac{2}{3}\sqrt{2E}a^{\frac{3}{2}}\right). \tag{13}$$

Which form an orthogonal set under the scalar product we have chosen, thus making them suitable for the construction of wave packet. From the positive energy modes, a unitarily evolving wave packet is constructed by choosing a normalized Poisson-like distribution

$$\psi(R,\tau) = \int_0^\infty d\sqrt{E}\phi^1_E(R)e^{iE\tau}A(\sqrt{E}), \tag{14}$$

$$A(\sqrt{E}) = \frac{\sqrt{2}\lambda^{\frac{1}{2}(\kappa+1)}}{\sqrt{\Gamma(\kappa+1)}}\sqrt{E}^{\kappa+\frac{1}{2}}e^{-\frac{\lambda}{2}E}, \tag{15}$$

where $\kappa \geq 0$ and $\lambda > 0$ are real parameters with κ being dimensionless and λ has dimensions of length or inverse of energy. For this choice of distribution, expectation value of the Hamiltonian is inversely proportional to λ. With this distribution and constraint $\kappa = 2|q|/3$, the wave packet takes the form,

$$\psi(a,\tau) = \sqrt{3}\frac{a^{|q|}}{\sqrt{\Gamma(\frac{2}{3}|q|+1)}}\left(\frac{\frac{\sqrt{2\lambda}}{3}}{\frac{\lambda}{2}-i\tau}\right)^{\frac{2}{3}|q|+1} e^{-\frac{2a^3}{9\left(\frac{\lambda}{2}-i\tau\right)}}. \tag{16}$$

Taking this constraint on the parameter makes the distribution a function of the operator ordering parameter. We can not be sure if the dependency of an observable

[a]Misner-Sharpe mass for spherically symmetric system $ds^2 = g_{ab}(z)dz^a dz^b + R^2(z)d\Omega^2$ is $M_{MS} = R(z)\left(1 - g^{ab}\partial_a R(z)\partial_b R(z)\right)/2$. For the case of FLRW mode, Misner-Sharp mass is $M_{MS} = a\dot{a}^2 r^3/2 = (4\pi r^3/3)\rho a^3 G$. The gravitational Hamiltonian is given by $H = -(3V/8\pi G)a\dot{a}^2 = -V\rho a^3$, since ρa^3 is a constant of motion and the Hamiltonian represents energy associated with the dust.

on parameter q is coming from the its dependence on the shape of the distribution or is it a genuine artifact of the operator ordering ambiguity. The signature of this parameter on the observables is discussed in[6] and we are not interested in operator ordering ambiguity in the Hamiltonian operator here. We are interested in the operator ordering ambiguity in the observables of this theory. It is shown in,[1] this model avoids singularity following DeWitt's criteria and in the current context, it represents a bouncing cosmology model.

4. Observables

In this analysis, we will follow Kuchař's proposal,[4,7,8] the observables in a time reparametrization invariant systems need not to commute with Hamiltonian constraint. Moreover in the quantum domain, we will require the Hermitian extension of operators to ensure the reality of expectation value of these observables. We will write the Hermitian extension of phase space functions that we are interested in and compute their expectation values in the wave packet (16).

4.1. *Hubble Parameter*

Corresponding to the canonical form of the Hubble parameter (7), several operator orderings for which the Hubble parameter is Hermitian can be written as,

$$F.O.1 \rightarrow \hat{\mathbb{H}}_1 = -a^{-1}\hat{p}_a a^{-1}, \tag{17}$$

$$F.O.2 \rightarrow \hat{\mathbb{H}}_2 = -\frac{1}{2}\left(a^{n-2}\hat{p}_a a^{-n} + a^{-n}\hat{p}_a a^{n-2}\right), \tag{18}$$

$$\hat{\mathbb{H}}_1\psi = \hat{\mathbb{H}}_2\psi = i\, a^{-2}\frac{\partial\psi}{\partial a} = \bar{\mathbb{H}}. \tag{19}$$

To ensure the Hermiticity of the operator, the boundary term $i\left[\psi^*\chi\right]_0^\infty$ has to vanish. It is achieved by the wavefunctions that vanishes when $a \rightarrow 0$ or $a \rightarrow \infty$. For the wave packets (16), it is the case provided $q \neq 0$.

The expectation value of the Hubble parameter for the wave packet (16) is,

$$\bar{\mathbb{H}}_1 = \langle\psi|\hat{\mathbb{H}}_1|\psi\rangle = i\int_0^\infty \psi^*(a,\tau)\frac{\partial\psi(a,\tau)}{\partial a}\, da = \frac{8\tau}{3(\lambda^2+4\tau^2)}. \tag{20}$$

In the large τ limit, i.e. $\tau^2 >> \lambda^2$, we recovers the classical value of Hubble parameter for the flat FLRW model with dust as matter.

$$\bar{\mathbb{H}}|_{\lambda\rightarrow 0} = \frac{2}{3\tau}. \tag{21}$$

The expectation value of the Hubble parameter is plotted in Fig. 1.

The Hubble parameter $\bar{\mathbb{H}}$ has global maximum at $\tau = \lambda/2$ and global minimum at $\tau = -\lambda/2$. At the point of classical singularity the Hubble parameter vanishes thus signifying a robust singularity resolution.

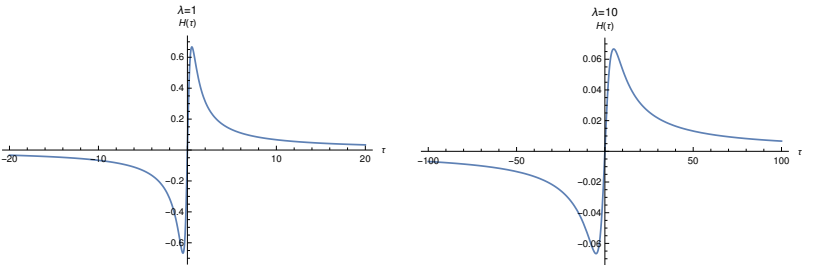


Fig. 1. Expectation value of the Hubble Parameter for different λ values.

4.2. *Ricci Scalar*

We will write general operator orderings that will make the operator corresponding to the phase space function given in Eq. (8) Hermitian with the given measure. There are two choices, normal symmetric operator ordering and Weyl ordering.

$$\hat{\mathcal{R}}_1 = 3a^{-j}\hat{p}_a a^{2j-4}\hat{p}_a a^{-j}, \tag{22}$$

$$\hat{\mathcal{R}}_2 = -\frac{3}{2}\left(a^{-j}\hat{p}_a a^{-k}\hat{p}_a a^{j+k-4} + a^{j+k-4}\hat{p}_a a^{-k}\hat{p}_a a^{-j}\right). \tag{23}$$

Here j and k are parameters that encapsulate the operator ordering ambiguity and describe our freedom in choosing operator ordering corresponding to the Ricci scalar.

For the operators to be Hermitian, the boundary term

$$-3\left[a^{-2}\left(\psi^*\frac{\partial\chi}{\partial a} - \frac{\partial\psi^*}{\partial a}\chi\right)\right]_0^\infty = 0 \tag{24}$$

has to be satisfied. For the case of the wave packet (16), the boundary term vanishes when $q > 3/2$. The expectation value of the Ricci scalar for this wave packet is,

$$\bar{\mathcal{R}}_1 = \frac{16\left(\lambda^2(3|q| + 2(j-4)(j-1)) + 4|q|(2|q|-3)\tau^2\right)}{3|q|(2|q|-3)\left(\lambda^2+4\tau^2\right)^2}, \tag{25}$$

$$\bar{\mathcal{R}}_2 = \frac{16\left(\lambda^2(3|q| - 2j(j+k-4) + 5k - 12) + 4|q|(2|q|-3)\tau^2\right)}{3|q|(2|q|-3)\left(\lambda^2+4\tau^2\right)^2}. \tag{26}$$

We see, the expectation value is well behaved regular function in the domain of parameters that ensure the Hermiticity. Again for large $|\tau|$ i.e. $\tau^2 >> \lambda^2$, we recover the classical expression for the Ricci scalar irrespective of the operator ordering chosen,

$$\bar{\mathcal{R}} = \frac{4}{3\tau^2}. \tag{27}$$

For the limit $q >> j, k$ there is no signature of the operator ordering parameters j and k in the expectation value of the Ricci scalar.

$$\bar{\mathcal{R}}_1 = \bar{\mathcal{R}}_2 = \frac{16\left(3|q|\lambda^2 + 4|q|(2|q|-3)\tau^2\right)}{3|q|(2|q|-3)\left(\lambda^2+4\tau^2\right)^2}. \tag{28}$$

The Ricci scalar in this regime has local minima at $\tau = 0$ and global maxima at $\tau = \pm\lambda/2$. Surprisingly in the large q regime, the Ricci scalar has maxima at the dust proper time, where the Hubble parameter has extrema.

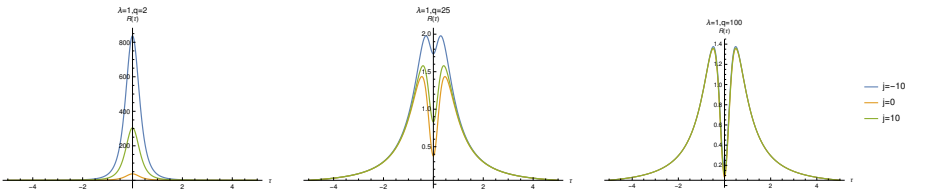


Fig. 2. Expectation value of the Ricci scalar $\bar{\mathcal{R}}_1$ for different operator orderings.

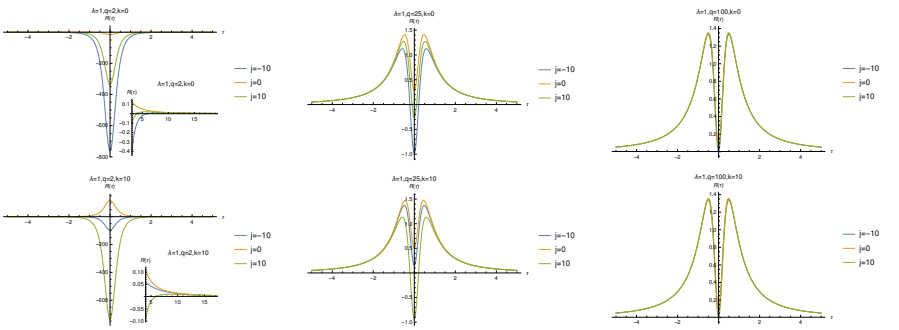


Fig. 3. Expectation value of Ricci scalar with Weyl ordering $\bar{\mathcal{R}}_2$ for different operator ordering parameters.

We have plotted the Ricci scalar for normal operator ordering in Fig. 2 and for the Weyl ordering in Fig 3. We see that the signature of the operator ordering parameters is most pronounced at the point of classical singularity i.e. at $\tau = 0$. Moreover, as we keep on increasing parameter q, the expectation value of the Ricci scalar for different operator ordering keeps on merging and for very large q, there is no signature of the operator ordering parameters. For the case of Weyl ordered Ricci scalar, the expectation value of the operator has negative values for certain parameter range.

4.3. *Riemann and Kretschmann Scalar*

Since the Kretschmann scalar is proportional to the Riemann scalar, we will write the Hermitian extensions of one of them, say, Kretschmann scalar and find the expectation value of the operator with given wave packet (16).

$$F.O.1 \to \hat{K}_1 = 15a^{-4}\hat{p}_a^4 a^{-4}, \tag{29}$$

$$F.O.2 \to \hat{K}_2 = 15\hat{p}_a^2 a^{-8}\hat{p}_a^2, \tag{30}$$

$$F.O.3 \to \hat{K}_3 = \frac{15}{2}\left(\hat{p}_a^4 a^{-8} + a^{-8}\hat{p}_a^4\right), \tag{31}$$

$$F.O.4 \to \hat{K}_4 = \frac{15}{2}\left(a^{-2}\hat{p}_a^4 a^{-6} + a^{-6}\hat{p}_a^4 a^{-2}\right). \tag{32}$$

The boundary term which needs to vanish for the Hermiticity of these operators is,

$$\Big[a^{-6}\big(\psi^*\chi''' - \psi'^*\chi'' + \psi''^*\chi' - \psi'''^*\chi - 6a^{-1}(\psi^*\chi'' - \psi''^*\chi\big) + 30a^{-2}(\psi^*\chi' - \psi'^*\chi))\Big]_0^\infty = 0 \implies |q| > \frac{9}{2} \tag{33}$$

The expectation value of the Kretschamann scalar for various operator orderings is,

$$\bar{K}_1(\tau) = \frac{1280\left(\lambda^2+4\tau^2\right)^{-4}}{27|q|(2|q|-9)(2|q|-3)}\Big(8(2|q|-9)(9|q|-20)\lambda^2\tau^2 + 3(9|q|-40)\lambda^4 + 16|q|(2|q|-9)(2|q|-3)\tau^4\Big), \tag{34}$$

$$\bar{K}_2(\tau) = \frac{1280\left(\lambda^2+4\tau^2\right)^{-4}}{27|q|(2|q|-9)(2|q|-3)(|q|-3)}\Big(3(|q|(9|q|+29)+504)\lambda^4 + 8(|q|-3)(2|q|-9)(9|q|+28)\lambda^2\tau^2 + 16(|q|-3)|q|(2|q|-9)(2|q|-3)\tau^4\Big), \tag{35}$$

$$\bar{K}_3(\tau) = \frac{1280\left(\lambda^2+4\tau^2\right)^{-4}}{27(|q|-3)|q|(2|q|-9)(2|q|-3)}\Big(3(|q|(9|q|-259)+1272)\lambda^4 + 8(|q|-3)(2|q|-9)(9|q|-116)\lambda^2\tau^2 + 16(|q|-3)|q|(2|q|-9)(2|q|-3)\tau^4\Big), \tag{36}$$

$$\bar{K}_4(\tau) = \frac{1280\left(\lambda^2+4\tau^2\right)^{-4}}{27(|q|-3)|q|(2|q|-9)(2|q|-3)}\Big(3(|q|-8)(9|q|-43)\lambda^4 + 8(|q|-3)(2|q|-9)(9|q|-44)\lambda^2\tau^2 + 16(|q|-3)|q|(2|q|-9)(2|q|-3)\tau^4\Big). \tag{37}$$

Again, the expectation value of these observables is a well behaved regular function in the domain of parameters that ensures the Hermiticity of this observable. Classical expression is recovered irrespective of the operator ordering chosen i.e., when $\tau^2 >> \lambda^2$, $\bar{K} = 80/27\tau^4$.

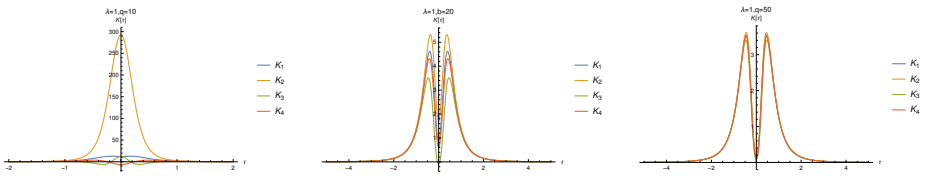


Fig. 4. Expectation value of Kretschmann scalar for various operator orderings.

We have plotted the expectation value of the Kretschmann scalar for different operator orderings. Here as well, we can see the signature of operator ordering is most pronounced at the point of classical singularity and Weyl ordered operator attains negative values for certain parameter range. Thus, the Kretschmann scalar also shows the features that are observed in the case of the Ricci scalar.

5. Discussion and Conclusion

We have studied the operator ordering ambiguity in the observables of a quantum model of cosmology with Brown-Kuchař dust as the matter source. The quantum model is singularity-free following DeWitt's criteria and it is shown that the expectation value of the scale factor is regular function with non-zero minima at the point of classical singularity and it follows classical trajectory, far away from the singularity. Thus, the quantum model represents a bouncing cosmological model which mimics classical behavior away from singularity.

We have studied the dynamics of the quantum FLRW model via the observables of the theory, Hubble parameter and curvature invariants. After writing the phase space functions for these observables, we have constructed the Hermitian extension of these observables. As these functions exhibit operator ordering ambiguity, there exists infinite many Hermitian extensions. For unitarily evolving wave packet, we have computed the expectation value of these observables. Very early in the collapsing phase and late in the expanding phase, these expectation values mimic the classical behavior irrespective of the operator ordering chosen. At the point of classical singularity, expectation value of the Hubble parameter vanishes and the Ricci scalar attains an extremum for all operator orderings, thus showing a robust singularity resolution.

This extremum value of the Ricci scalar is sensitive to the operator ordering parameters and the signature of these parameters is most pronounced at the point of singularity. The expectation value of the Ricci scalar is insensitive to this parameter away from singularity. Furthermore, in the large q regime, there is no signature of operator ordering parameter and the two maxima of Ricci scalar are at the time where the extrema of the Hubble parameter are located. For certain parameter range, the expectation value of the Weyl ordered Ricci scalar attains negative values. These features are present for the case of other curvature invariants as well.

Acknowledgments

HSS would like to acknowledge the financial support from University Grants Commission, Government of India, in the form of Junior Research Fellowship (UGC-CSIR JRF/Dec- 2016/503905). Research of KL is partially supported by the Startup Research Grant of SERB, Government of India (SRG/2019/002202).

References

1. C. Kiefer and T. Schmitz, Singularity avoidance for collapsing quantum dust in the Lemaître-Tolman-Bondi model, *Physical Review D* **99**, p. 126010 (June 2019).
2. J. D. Brown and K. V. Kuchar, Dust as a Standard of Space and Time in Canonical Quantum Gravity, *Physical Review D* **51**, 5600 (May 1995).
3. C. Kiefer, *Quantum gravity*International series of monographs on physics, no. 155 in International series of monographs on physics, third edition edn. (Oxford University Press, Oxford, 2012).
4. K. Kuchař, Canonical Quantum Gravity, *arXiv:gr-qc/9304012* (April 1993).
5. H. Maeda, Unitary evolution of the quantum universe with a Brown-Kuchar dust, *Classical and Quantum Gravity* **32**, p. 235023 (December 2015).
6. H. S. Sahota and K. Lochan, Infrared signatures of quantum bounce in collapsing geometry, *arXiv:2110.06247 [gr-qc]* (Oct 2021).
7. J. Barbour and B. Z. Foster, Constraints and gauge transformations: Dirac's theorem is not always valid *arXiv:0808.1223 [gr-qc]* (Aug 2008).
8. M. Bojowald, P. A. Höhn and A. Tsobanjan, Effective approach to the problem of time: General features and examples, *Physical Review D* **83** (Jun 2011).

Decoupled quark stars in self-interacting Brans-Dicke gravity

M. Sharif* and Amal Majid
Department of Mathematics, University of the Punjab, Quaid-e-Azam Campus, Lahore-54590, Pakistan.
**E-mail: msharif.math@pu.edu.pk*
www.pu.edu.pk

In this paper, we generate an anisotropic solution for a static sphere filled with quark matter in the framework of self-interacting Brans-Dicke theory. For this purpose, we add an anisotropic source in the seed distribution and decouple the field equations through deformation in the radial metric function. As a result of this transformation, the field equations are disintegrated into two systems which separately include the effects of isotropic and anisotropic sources. The system related to the additional source is solved via the MIT bag model equation of state. We consider Tolman V spacetime to formulate a solution for the isotropic sector which is extended to the anisotropic domain via decoupling technique. The junction conditions at the boundary determine the unknown parameters in terms of mass and radius of the spherical object. We investigate the viability and stability of the constructed strange star model in the presence of massive scalar field corresponding to the strange star candidate PSR J1614-2230. It is concluded that the anisotropic extension is well-behaved as it fulfills the necessary requirements of a physically acceptable model.

Keywords: Brans-Dicke theory; Quark star; Gravitational decoupling.

1. Introduction

Astronomical systems (such as stellar structures, galaxies and planets) are scattered throughout the universe in the form of a cosmic web. These self-gravitating objects contain significant clues regarding the history and structure of the vast universe. Strange quark stars intriguing cosmic objects that are smaller than neutron stars and possess extremely dense cores. Researchers believe that the enormous release of radiation and energy in extremely luminous supernovae can be explained through quark stars.[1] A model that adequately describes the interactions of quarks at high density levels is still under investigation. However, in the absence of the best fit, researchers have used MIT bag model equation of state (EoS) to examine different properties of strange quark stars.[2] The statistical data obtained from GW170817[3] and GW190425[4] was used to predict the compactness and mass of quark stars. These predictions match the estimates made via the MIT bag model which strengthens the choice of this EoS. Bhar[5] used the MIT bag model to compute a solution representing strange stars in the background of Krori and Barua spacetime. The bag model has also been used to investigate the stability of quark stars by introducing radial perturbations.[6] Deb et al.[7] adopted the EoS to construct a charged stellar structure.

The task of obtaining well-behaved spacetimes corresponding to realistic models is difficult due to the non-linearity of field equations. Recently, Ovalle simplified the extraction of solutions by proposing the gravitational decoupling approach via minimal geometric deformation (MGD).[8] In this scheme, an extra source is successively added to a seed distribution (vacuum or isotropic) to incorporate intricate features of physical systems. The degree of freedom in the field equations is reduced by splitting the field equations into two arrays via a linear transformation in the radial metric function. These sets exclusively include the effects of seed and additional gravitational sources. Moreover, no energy is transferred between the two matter distributions as they are individually conserved. The respective solutions of these systems are combined to formulate a new solution. An extension of this technique has also been proposed which induces transformations in temporal as well as radial metric functions.[9]

The efficient technique of gravitational decoupling has successfully been employed to generate anisotropic solutions from well-known ansatz. The anisotropic extensions of Tolman IV have been obtained in braneworld[10] as well as general relativity (GR)[11] via MGD scheme. Isotropic configurations corresponding to Durgapal-Fuloria,[12] Krori-Barua,[13] Heintzmann[14] and Schwarzschild[15] interior solutions have been decoupled to evaluate their corresponding anisotropic versions. Contreras[16,17] deformed the metric component of (2+1)- dimensional spacetime to decouple the field equations with a cosmological constant. Singh et al.[18] derived an anisotropic spacetime by incorporating the embedding class-one condition in the decoupling approach. Hensh and Stuchlik[19] constructed an anisotropic counterpart of Tolman VII by the method of gravitational decoupling. Sharif and Ama-Tul-Mughani[20] implemented the MGD scheme to extend charged solutions to anisotropic domain in a cloud of strings. Anisotropic cosmological solutions have also been obtained by applying the MGD technique to FLRW as well as Kantowski-Sachs spacetimes.[21] Recently, Contreras et al.[22] studied axially symmetric black holes and showed that gravitational decoupling can be adopted beyond spherical symmetry.

The general theory of relativity is modified to incorporate the effects of cosmic expansion. In this regard, the family of scalar-tensor theories has gained attention. Brans-Dicke (BD) theory (a prototype of scalar-tensor theories) includes a spacetime varying scalar field $\vartheta = \frac{1}{G}$. (G is Newton's gravitational constant) along with the metric tensor to describe the effects of gravitational field.[23] A BD coupling parameter (ω_{BD}) couples the matter distribution and the massless scalar field. Different values of the tunable coupling parameter explain different cosmological scenarios. However, a conflict occurs when large values of ω_{BD} in the weak-field limit[24] do not coincide with the small values required to describe the inflationary era.[25] In order to remove the discrepancy, a potential function $V(\varsigma)$ is introduced which allots a mass to the scalar field (ς). This modification leads to self-interacting BD (SBD) theory[26] which admits all values of the coupling parameter greater than $-3/2$ for $m_\varsigma > 2 \times 10^{-25} GeV$ (m_ς denotes mass of the scalar field).[27] Recently,

MIT bag model was employed to examine the behavior of strange quark stars in the presence of a massive scalar field.[28] Some well-behaved solutions in SBD gravity as well as modified theories have been computed through gravitational decoupling technique.[29]

In this paper, we formulate two anisotropic strange star models by implementing MGD scheme on isotropic distributions in the context of SBD gravity. For this purpose, we adopt Tolman V ansatz and use the MIT bag model to investigate the viability and stability of the constructed structures. We evaluate the field equations under the effect of the additional source in section **2**. In section **3**, we deform the radial metric function to split the field equations. The anisotropic extension is checked for viability and stability in section **4**. In section **5**, the main results are discussed.

2. Self-interacting Brans-Dicke Theory

The SBD gravity is described by the action

$$S=\int\sqrt{-g}(\mathcal{R}\varsigma-\frac{\omega_{BD}}{\varsigma}\nabla^{\gamma}\nabla_{\gamma}\varsigma-V(\varsigma)+L_m+\sigma L_{\Theta})d^4x, \tag{1}$$

where relativistic units have been used. Here, $\mathcal{R}$ and L_m denote the Ricci scalar and matter lagrangian whereas $g=|g_{\gamma\delta}|$. The contribution of the anisotropic source (Θ) via the lagrangian L_{Θ} is controlled by the decoupling parameter σ. The additional gravitational source can be categorized as a scalar, vector or tensor field. The field and wave equations derived from the action (1) are written as

$$G_{\gamma\delta}=T^{(\mathrm{eff})}_{\gamma\delta}=\frac{1}{\varsigma}(T^{(m)}_{\gamma\delta}+\sigma\Theta_{\gamma\delta}+T^{\varsigma}_{\gamma\delta}), \tag{2}$$

$$\Box\varsigma=\frac{T^{(m)}+\Theta}{3+2\omega_{BD}}+\frac{1}{3+2\omega_{BD}}(\varsigma\frac{dV(\varsigma)}{d\varsigma}-2V(\varsigma)), \tag{3}$$

where $\Box\varsigma=\frac{1}{\sqrt{-g}}(\sqrt{-g}g^{\gamma\delta}\varsigma_{,\gamma})_{,\delta}$, $T^{(m)}=g^{\gamma\delta}T^{(m)}_{\gamma\delta}$ and $\Theta=g^{\gamma\delta}\Theta_{\gamma\delta}$. We consider an isotropic seed source whose energy-momentum tensor has the form

$$T^{(m)}_{\gamma\delta}=\rho u_{\gamma}u_{\delta}+p(u_{\gamma}u_{\delta}-g_{\gamma\delta}), \tag{4}$$

where p, u_{γ} and ρ represent pressure, four-velocity and energy density, respectively. The influence of massive scalar field on the physical aspects of static sphere are governed by the following energy-momentum tensor

$$T^{\varsigma}_{\gamma\delta}=\varsigma_{,\gamma;\delta}-g_{\gamma\delta}\Box\varsigma+\frac{\omega_{BD}}{\varsigma}(\varsigma_{,\gamma}\varsigma_{,\delta}-\frac{g_{\gamma\delta}\varsigma_{,\alpha}\varsigma^{,\alpha}}{2})-\frac{V(\varsigma)g_{\gamma\delta}}{2}. \tag{5}$$

The anisotropic extension is obtained by describing the internal geometry of the static sphere as

$$ds^2=e^{\phi(r)}dt^2-e^{\psi(r)}dr^2-r^2d\theta^2-r^2\sin^2\theta d\varphi^2. \tag{6}$$

The field equations, formulated in the presence of the anisotropic source, are expressed as

$$\frac{1}{r^2}-e^{-\psi}\left(\frac{1}{r^2}-\frac{\psi'}{r}\right)=\frac{1}{\varsigma}(\rho+\sigma\Theta_0^0+T_0^{0\varsigma}), \tag{7}$$

$$-\frac{1}{r^2}+e^{-\psi}\left(\frac{1}{r^2}+\frac{\phi'}{r}\right)=\frac{1}{\varsigma}(p-\sigma\Theta_1^1-T_1^{1\varsigma}), \tag{8}$$

$$\frac{e^{-\psi}}{4}\left(2\phi''+\phi'^2-\psi'\phi'+2\frac{\phi'-\psi'}{r}\right)=\frac{1}{\varsigma}(p-\sigma\Theta_2^2-T_2^{2\varsigma}), \tag{9}$$

where

$$T_0^{0\varsigma}=e^{-\psi}\left[\varsigma''+\left(\frac{2}{r}-\frac{\psi'}{2}\right)\varsigma'+\frac{\omega_{BD}}{2\varsigma}\varsigma'^2-e^{\psi}\frac{V(\varsigma)}{2}\right],$$

$$T_1^{1\varsigma}=e^{-\psi}\left[\left(\frac{2}{r}+\frac{\phi'}{2}\right)\varsigma'-\frac{\omega_{BD}}{2\varsigma}\varsigma'^2-e^{\psi}\frac{V(\varsigma)}{2})\right],$$

$$T_2^{2\varsigma}=e^{-\psi}\left[\varsigma''+\left(\frac{1}{r}-\frac{\psi'}{2}+\frac{\phi'}{2}\right)\varsigma'+\frac{\omega_{BD}}{2\varsigma}\varsigma'^2-e^{\psi}\frac{V(\varsigma)}{2}\right].$$

Here differentiation with respect to the radial coordinate is denoted by $'$. Moreover, the wave equation (3), containing information about the evolution of the massive scalar field, takes the form

$$\begin{aligned}\Box\varsigma&=-e^{-\psi}\left[\left(\frac{2}{r}-\frac{\psi'}{2}+\frac{\phi'}{2}\right)\varsigma'+\varsigma''\right]\\&=\frac{1}{3+2\omega_{BD}}\left[T^{(m)}+\Theta+\left(\varsigma\frac{dV(\varsigma)}{d\varsigma}-2V(\varsigma)\right)\right].\end{aligned} \tag{10}$$

The system of field equations (7)-(9) correspond to an anisotropic fluid distribution for $\Theta_1^1\neq\Theta_2^2$.

3. Gravitational Decoupling

Currently, the system (7)-(9) involves unknown metric potentials, state determinants, massive scalar field and anisotropic source. Thus, we require additional constraints to reduce the degrees of freedom and close the system. In this regard, the MGD technique is an efficient method for decoupling the complex field equations into simpler sets by transforming the radial metric component as

$$e^{-\psi(r)}\mapsto\lambda(r)+\sigma\nu(r), \tag{11}$$

where $\nu(r)$ is the deformation function that governs the translation in the radial metric component. The temporal metric potential is not transformed and remains unchanged. It must be noted that the linear mapping does not affect the spherical geometry of the structure. Plugging the transformation in Eqs. (7)–(9) and setting

$\sigma = 0$ yields the first set as

$$\rho = -\frac{1}{2r^2\varsigma(r)}[r^2\omega\lambda(r)\varsigma'^2(r) - r^2\varsigma(r)V(\varsigma) + r\varsigma(r)(r\lambda'(r)\varsigma'(r) + 2r\lambda(r)\varsigma''(r) + 4\lambda(r)\varsigma'(r)) + 2\varsigma^2(r)(r\lambda'(r) + \lambda(r) - 1)], \tag{12}$$

$$p = \frac{1}{r^2}[\varsigma(r)(r\lambda(r)\phi'(r) + \lambda(r) - 1)] + \frac{1}{2r\varsigma(r)}[\lambda(r)\varsigma'(r)(\varsigma(r)(r\phi'(r) + 4) - r\omega_{BD}\varsigma'(r))] - \frac{V(\varsigma)}{2}, \tag{13}$$

$$p = \frac{1}{4r\varsigma(r)}[\varsigma(r)\lambda'(r)(\varsigma(r)(r\phi'(r) + 2) + 2r\varsigma'(r)) + \lambda(r)(2\varsigma(r)\varsigma'(r) \times ((r\phi'(r) + 2) + 2r\varsigma''(r)) + \varsigma^2(r)(2r\phi''(r) + r\phi'^2(r) + 2\phi'(r)) + 2r\omega_{BD}\varsigma'^2(r)) - 2r\varsigma(r)V(\varsigma)]. \tag{14}$$

The above system corresponds to the perfect fluid configuration exclusively as the anisotropic effects are excluded. The isotropic source observes the following conservation equation

$$T_1^{1'(\text{eff})} - \frac{\phi'(r)}{2}(T_0^{0(\text{eff})} - T_1^{1(\text{eff})}) = 0. \tag{15}$$

The second set, encompassing the terms related to the additional source, is expressed as

$$\Theta_0^0 = \frac{-1}{2r^2\varsigma(r)}[r\varsigma(r)\nu'(r)(r\varsigma'(r) + 2\varsigma(r)) + \nu(r)(r^2\omega_{BD}\varsigma'^2(r) + 2r\varsigma(r) \times (r\varsigma''(r) + 2\varsigma'(r)) + 2\varsigma^2(r))], \tag{16}$$

$$\Theta_1^1 = \frac{-1}{2r^2\varsigma(r)}[\nu(r)(-r^2\omega_{BD}\varsigma'(r)^2 + r\varsigma(r)(r\phi'(r) + 4)\varsigma'(r) + 2\varsigma^2(r) \times (r\phi'(r) + 1))], \tag{17}$$

$$\Theta_2^2 = \frac{-1}{4\varsigma(r)}[2\varsigma(r)(r\nu'(r)\varsigma'(r) + \nu(r)((r\phi'(r) + 2)\varsigma'(r) + 2r\varsigma''(r))) + \varsigma^2(r)(\nu'(r)(r\phi'(r) + 2) + \nu(r)(2r\phi''(r) + r\phi'^2(r) + 2\phi'(r))) + 2r\omega_{BD}\nu(r)\varsigma'^2(r)]. \tag{18}$$

The energy and momentum of the anisotropic source is conserved as

$$\Theta_1^{1'(\text{eff})} - \frac{\phi'(r)}{2}(\Theta_0^{0(\text{eff})} - \Theta_1^{1(\text{eff})}) - \frac{2}{r}(\Theta_2^{2(\text{eff})} - \Theta_1^{1(\text{eff})}) = 0, \tag{19}$$

where

$$\Theta_0^{0(\text{eff})} = \frac{1}{\varsigma}\left(\Theta_0^0 + \frac{1}{2}\nu'(r)\varsigma'(r) + \nu(r)\varsigma''(r) + \frac{\omega_{BD}\nu(r)\varsigma'^2(r)}{2\varsigma(r)} + \frac{2\nu(r)\varsigma'(r)}{r}\right),$$

$$\Theta_1^{1(\text{eff})} = \frac{1}{\varsigma}\left(\Theta_1^1 + \frac{1}{2}\nu(r)\phi'(r)\varsigma'(r) - \frac{\omega_{BD}\nu(r)\varsigma'^2(r)}{2\varsigma(r)} + \frac{2\nu(r)\varsigma'(r)}{r}\right),$$

$$\Theta_2^{2(\text{eff})} = \frac{1}{\varsigma}\left(\Theta_2^2 + \frac{1}{2}\nu'(r)\varsigma'(r) + \frac{1}{2}\nu(r)\phi'(r)\varsigma'(r) + \nu(r)\varsigma''(r) + \frac{\omega_{BD}\nu(r)\varsigma'^2(r)}{2\varsigma(r)} + \frac{\nu(r)\varsigma'(r)}{r}\right).$$

The MGD scheme restricts the exchange of matter or energy between the two sources (seed and additional) by conserving them individually. As a consequence of disintegrating the field equations via MGD, the two systems can be solved separately. A well-behaved isotropic spacetime can specify the system (12)-(14) which minimizes the degrees of freedom as the second system contains the deformation function and components of $\Theta_{\gamma\delta}$ as undetermined variables only. In the next section, we attain a solution for the anisotropic setup by assuming a well-known ansatz along with an additional constraint.

4. Anisotropic Solution

Tolman V ansatz (one of the eight solutions proposed by Tolman) describes a spherical structure with infinite density and pressure at the center.[30] Zubair and Azmat[31] evaluated the anisotropic counterpart of Tolman V via the MGD method. Recently, Jasim et al.[32] employed this spherical solution to study charged strange stars. Tolman V spacetime is defined as

$$ds^2 = B^2 r^{2n} dt^2 - \left(\frac{1+2n-n^2}{1-(1+2n-n^2)\left(\frac{r}{F}\right)^W}\right) dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\varphi^2, \quad (20)$$

where n, F and B are unknown constants with $W = \frac{2(1+2n-n^2)}{1+n}$. We set $\lambda = \frac{1+2n-n^2}{1-(1+2n-n^2)\left(\frac{r}{F}\right)^W}$ and determine the isotropic sector via the above spacetime. For this purpose, we evaluate the unknown parameters through the continuity of first and second fundamental forms at the junction (Σ) of interior and exterior geometries. The exterior of a static uncharged celestial object is defined via Schwarzschild metric given as

$$ds^2 = \frac{r-2M}{r} dt^2 - \frac{r}{r-2M} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\varphi^2, \quad (21)$$

where M is the total mass of the compact structure. The boundary conditions at $r = R$ (R is the radius of compact object)

$$(g_{\gamma\delta}^-)_\Sigma = (g_{\gamma\delta}^+)_\Sigma, \quad (p_r)_\Sigma = 0,$$
$$(\varsigma^-(r))_\Sigma = (\varsigma^+(r))_\Sigma, \quad (\varsigma'^-(r))_\Sigma = (\varsigma'^+(r))_\Sigma.$$

are utilized to determine the constants as

$$B=\frac{\sqrt{M}R^{\frac{R}{4M-2R}}}{\sqrt{-\frac{M}{2M-R}}}, \tag{22}$$

$$n=\frac{M}{R-2M}, \tag{23}$$

$$W=-\frac{2\left(M^2+2MR-R^2\right)}{2M^2-3MR+R^2}, \tag{24}$$

$$F=R\left(\frac{M^2(R-2M)}{R\left(-M^2-2MR+R^2\right)}\right)^{-\frac{1}{W}} \tag{25}$$

$$\begin{aligned}\omega_{BD}=\frac{2M-R}{8M^2((n-2)n-1)}\Bigg(&-((n-2)n-1)(2M-R)\left(\frac{R}{F}\right)^W(m_\varsigma R\\&\times(2M-R)-4)-4m_\varsigma M^2R+4m_\varsigma MR^2-m_\varsigma R^3+8Mn^3\\&-24Mn^2+8Mn+16M-8n^3R+12n^2R+16nR\Bigg).\end{aligned} \tag{26}$$

Here, the corresponding scalar field has been calculated by following the technique in.[33]

The energy density and pressure of the anisotropic analogue of Tolman V are expressed as

$$\begin{aligned}\rho=\frac{-1}{2r^2\varsigma}\Bigg(&r\varsigma\left(r\varsigma'\left(\sigma\nu'(r)+\frac{\zeta_1}{r}\right)+2\sigma\nu(r)\left(r\varsigma''+2\varsigma'\right)+2\zeta_2\left(r\varsigma''\right.\right.\\&\left.+\ 2\varsigma'\right))+2\varsigma^2\left(\sigma r\nu'(r)+\sigma\nu(r)+\zeta_1+\zeta_2-1\right)+r^2\omega_{BD}\varsigma'^2\left(\sigma\nu(r)\right.\\&\left.+\ \zeta_2\right)-r^2\varsigma V(\varsigma)\Big),\end{aligned} \tag{27}$$

$$\begin{aligned}p_r=\frac{\varsigma}{r^2}\left(\nu(r)(\sigma+2\sigma n)+\zeta_2(2n+1)-1\right)+\frac{\varsigma'\left(2(n+2)\varsigma-r\omega_{BD}\varsigma'\right)}{2r\varsigma}\\\times\left(\sigma\nu(r)+\zeta_2\right)-\frac{V(\varsigma)}{2},\end{aligned} \tag{28}$$

$$\begin{aligned}p_\perp=\frac{\varsigma}{2r^2}\Bigg(&\sigma(n+1)r\nu'(r)+2\sigma n^2\nu(r)+\frac{1}{\zeta_3^2}((n-2)n-1)\left(2n-n^2+1\right)\\&\times\ \left(2n^2-(n+1)W\right)\left(\frac{r}{F}\right)^W-2n^2\Bigg)\Bigg)+\left(\sigma\nu(r)+\zeta_2\right)\Bigg(\frac{\varsigma'}{2}\\&\times\left(\frac{\sigma\nu'(r)+\frac{\zeta_1}{r}}{\sigma\nu(r)+\zeta_2}+\frac{2n}{r}+\frac{2}{r}\right)+\varsigma''+\frac{\omega_{BD}\varsigma'^2}{2\varsigma}\Bigg)-\frac{V(\varsigma)}{2},\end{aligned} \tag{29}$$

where $\zeta_1=\frac{\left(n^2-2n-1\right)^2W\left(\frac{r}{F}\right)^W2}{\zeta_3}$, $\zeta_2=\frac{-n^2+2n+1}{\zeta_3}$ and $\zeta_3=\Big(\left(n^2-2n-1\right)\left(\frac{r}{F}\right)^W+1\Big)$ with anisotropy $\Delta=p_\perp-p_r$. The matter variables are completely specified if the deformation function is known. For this purpose, a constraint on the deformation function or appropriate EoS is applied to the anisotropic structure.

In this work, we develop an anisotropic model for strange quark stars by implementing the MIT bag model. It is hypothesized that the core of a quark star is

composed of three flavors of quarks: up (u), strange (s), and down (d). We assume that the quarks are massless and non-interacting in nature. According to the bag model, the pressure and density of quark matter read

$$p_r = \sum_f p^f - \mathcal{B}, \quad \rho = \sum_f \rho^f + \mathcal{B}, \quad f = u,\ d,\ s, \tag{30}$$

where p^f and ρ^f correspond to the pressure and density of each flavor, respectively. Kapusta[34] formulated the EoS for strange quark matter through the relation $\rho^f = 3p^f$ as

$$3p_r = \rho - 4\mathcal{B}. \tag{31}$$

Utilizing Eqs.(27) and (28) in the bag model, we obtain the following differential equation

$$\begin{aligned}&(\sigma r\nu'(r)\varsigma' + \sigma\nu(r)\,(2(3n+8)\varsigma' + 2r\varsigma'') + 2\zeta_2 r\varsigma'' + \zeta_1\varsigma' + 6n\zeta_2\varsigma'\\ &+ 16\zeta_2\varsigma' - 4rV(\varsigma) + 8r\mathcal{B}) + \frac{2\varsigma}{r}\,(\sigma r\nu'(r) + \sigma(6n+4)\nu(r) + \zeta_1\\ &+(6n+4)\zeta_2 - 4) - 2r^2\omega_{BD}\varsigma'^2\,(\sigma\nu(r) + \zeta_2) = 0,\end{aligned} \tag{32}$$

which is numerically solved along with the wave equation subject to central conditions ($\varsigma(0) = 0.2,\ \varsigma'(0) = 0$ and $\nu(0) = \nu_c$) for $V(\varsigma) = \frac{1}{2}m_\varsigma^2\varsigma^2$.

In SBD gravity, the values of coupling parameter greater than $-3/2$ are allowed for $m_\varsigma > 10^{-4}$ (in dimensionless units). Therefore, the wave equation is solved numerically for $m_\varsigma = 0.001$. Furthermore, the constants B, F, n remain unchanged for the anisotropic sphere whereas ω_{BD} is an undetermined free parameter. Thus, we choose the value of ω_{BD} as given in Eq.(26). The constants are computed by utilizing the observed mass and radii of three quark star candidates (PSR J1614-2230, Her X1 and 4U 1608-52) for $\sigma = 0.2,\ 0.9$ and $\mathcal{B} = 60MeV/fm^3$. The values of ω_{BD} and ν_c for the considered stars are given in Table **1**. It is noted that ω_{BD} attains values greater than $-3/2$ for all stellar candidates. The physical properties of the anisotropic stellar model are examined in the next section.

Table 1. Values of different parameters corresponding for $m_\varsigma = 0.001$ and $\mathcal{B} = 60MeV/fm^3$.

	PSR J1614-2230	Her X-1	4U 1608-52
$M\ (M_\odot)$	1.97	0.88	1.74
$R\ (km)$	13	7.7	9.3
ω_{BD}	9.750	15.353	6.593
$\nu_c\ (\sigma = 0.2)$	-5.1	-3.6	-6.8

4.0.1. *Physical Features of Anisotropic Model*

The metric potentials of a well-behaved solution are positive and increase monotonically.[35] The state parameters must observe a monotonically decreasing behavior

away from the center. The graphical representations of Eqs.(27)-(29) in Figure 1 reveals that the strange star model has maximum density and radial/tangential pressures in the core. Furthermore, radial pressure is zero at the surface for the considered stellar objects. It is observed from the plots of anisotropy that $p_\perp > p_r$ for 4U 1608-52 whereas anisotropy related to the remaining stars is negative, i.e., $p_\perp < p_r$. It is worthwhile to mention here that the GR analog exhibits similar behavior.[31]

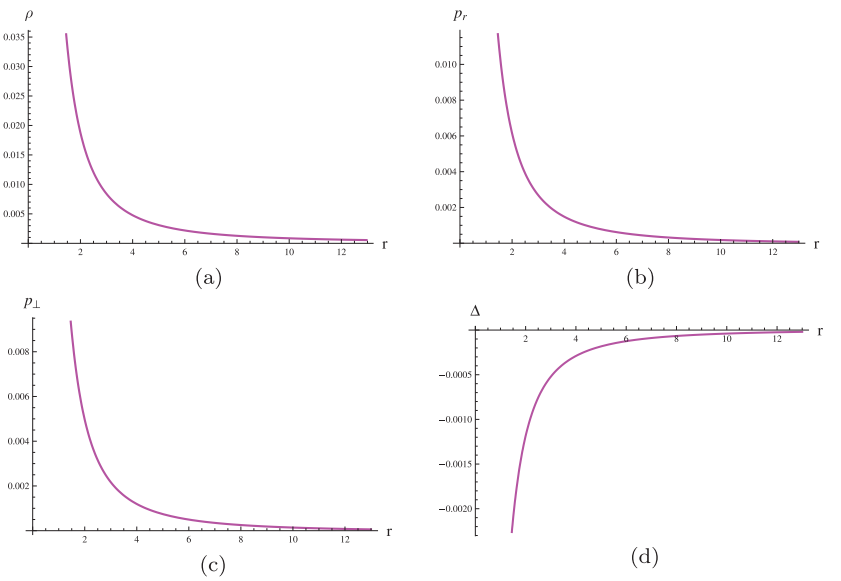


Fig. 1. Plots of (a) ρ, (b) p_r, (c) $p_\perp$ (in km^{-2}) and (d) Δ of anisotropic Tolman V solution for $\sigma = 0.2$.

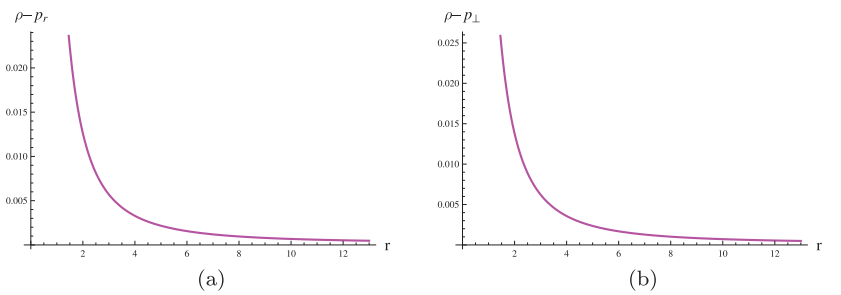


Fig. 2. (a), (b) DEC for extended Tolman V solution for $\sigma = 0.2$.

As the interior of compact stellar structures consists of normal matter, therefore the energy-momentum tensor describing the internal configuration must obey the

four energy bounds[36]

$$\begin{aligned}
&\text{null energy condition:} \quad \rho + p_r \geq 0, \quad \rho + p_\perp \geq 0,\\
&\text{weak energy condition:} \quad \rho \geq 0, \quad \rho + p_r \geq 0, \quad \rho + p_\perp \geq 0,\\
&\text{strong energy condition:} \quad \rho + p_r + 2p_\perp \geq 0,\\
&\text{dominant energy condition:} \quad \rho - p_r \geq 0, \quad \rho - p_\perp \geq 0.
\end{aligned}$$

The first three conditions require that energy density and pressure components remain positive throughout the setup. As the matter variables presented in Figure 1 fulfil this criterion, the anisotropic Tolman V is consistent with null, weak and strong energy constraints. Figure 2 indicates that the decoupled model obeys the fourth condition as well. Thus, the constructed model represents a viable quark star.

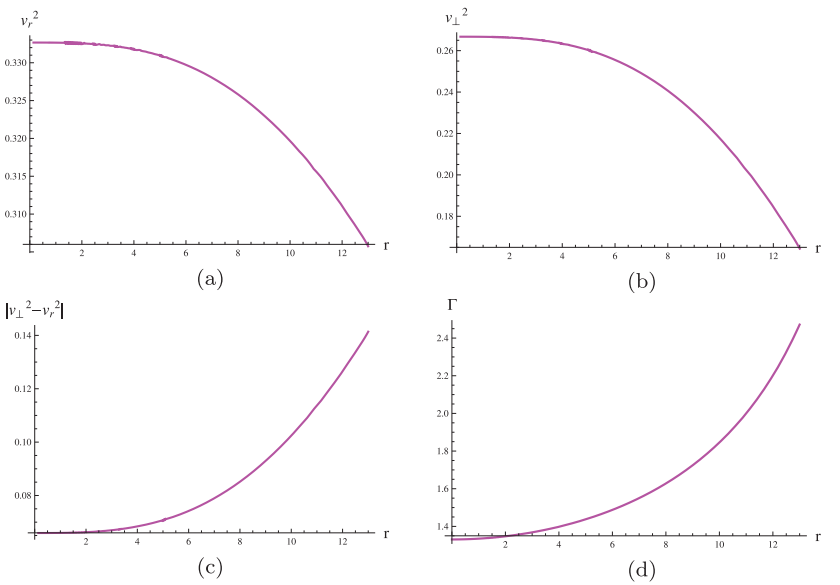


Fig. 3. Plots of (a)radial/(b) tangential velocities, (c) $|v_\perp^2 - v_r^2|$ and (d) adiabatic index of extended Tolman V solution for $\sigma = 0.2$.

The extended version of Tolman V solution is stable if it complies with the causality condition which states that velocity of light always exceeds the rate at which sound travels, i.e., $0 < v_\perp^2 < 1$ and $0 < v_r^2 < 1$ ($v_\perp^2 = \frac{dp_\perp}{d\rho}$ and $v_r^2 = \frac{dp_r}{d\rho}$ represent transverse and radial components of sound, respectively). It must be noted that the propagation of sound waves changes with a change in the medium. Figure 3 depicts that the extended solution is consistent with the causality criterion for higher as well as lower values of decoupling parameter. In order to inspect the anisotropic system for potential stability, we implement Herrera's cracking criterion. According

to this approach, if a disturbance in the equilibrium of the system leads to radial forces of different signs, the configuration cracks. The condition $0 < |v_{\perp}^2 - v_r^2| < 1$ ensures that no cracking appears within the compact sphere. This condition holds true for the current setup as shown in Figure 3.

If a stellar model obeys a stiff EoS, it is compact and difficult to compress as pressure increases greatly in response to a small change in density. Adiabatic index measures the stiffness of the EoS as

$$\Gamma = \frac{p_r + \rho}{p_r}\frac{dp_r}{d\rho} = \frac{p_r + \rho}{p_r}v_r^2.$$

The value of adiabatic index greater than $4/3$ corresponds to a stiff matter distribution.[37] The adiabatic index for $\sigma = 0.2$ is displayed in Figure 3. The generated extension of Tolman V observes the required restraint under the influence of massive scalar field.

5. Conclusions

In this work, we have extended a known isotropic solution to the anisotropic domain through the method of gravitational decoupling. Anisotropy has been induced in the isotropic matter configuration by means of an additional source. The two sources (seed and additional) have been decoupled through a linear transformation of the radial metric component. As a consequence, we have obtained two systems of differential equations. The first system includes the effects of the seed distribution only while the influence of the additional source is restricted to the second set. We have considered Tolman V spacetime to represent the first array. The system related to the anisotropic source has been solved by imposing MIT bag model on state variables of the anisotropic solution. We have numerically evaluated the wave equation for $V(\varsigma) = \frac{1}{2}m_{\varsigma}^2\varsigma^2$, $m_{\varsigma} = 0.001$ and $\sigma = 0.2$. Finally, the physical behavior of the extended model has been examined for $\mathcal{B} = 60MeV/fm^3$ through energy constraints as well as causality and cracking approaches for the star PSR J1614-2230.

The graphical analysis has shown that the state parameters observe a monotonically decreasing trend towards the surface and attain the maximum value at $r = 0$. Moreover, the anisotropy corresponding to PSR J1614-2230 is negative. Thus, radial pressure is more than the transverse component indicating a presence of attractive force within the spherical object. The extended model is composed of viable normal matter as it is consistent with the energy conditions. We have established that the constructed setup is stable as radial/tangential components of speed agree with causality and cracking criteria. The adiabatic parameter is greater than $4/3$ which indicates that the static model holds itself against inward gravitational pull for the considered values of the parameters. Thus, the strange star structure formulated via decoupling is physically realistic and stable in the presence of a massive scalar field. It is interesting to mention here that all the obtained results reduce to GR for $\varsigma =$ constant and $\omega_{BD} \to \infty$.

References

1. E. O. P. Ofek, et al., SN 2006gy: An extremely luminous supernova in the galaxy NGC 1260, *Astrophys. J.* **659**, L13 (2007); R. Ouyed, D. Leahy and P. Jaikumar, P., Predictions for signatures of the quark-nova in superluminous supernovae, arXiv:0911.5424.
2. G. H. Bordbar and A. R. Peivand, Computation of the structure of a magnetized strange quark star, *Res. Astron. Astrophys.* **11**, 851 (2011).
3. B. P. Abbott et al., GW170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
4. The LIGO Scientific Collaboration, the Virgo Collaboration, B. P. Abbott et al., GW190425: Observation of a compact binary coalescence with total mass $3.4M_{\odot}$, *Astrophys. J. Lett.* **892**, L3 (2020).
5. P. Bhar, A new hybrid star model in Krori-Barua spacetime, *Astrophys. Space Sci.* **357**, 46 (2015).
6. J. D. V. Arbañil and M. Malheiro, Stability of charged strange quark stars, *AIP Conf. Proc.* **1693**, 030007 (2015).
7. D. Deb, Anisotropic strange stars in the Einstein–Maxwell spacetime, *Eur. Phys. J. C* **78**, 465 (2018).
8. J. Ovalle, Searching exact solutions for compact stars in braneworld: A conjecture, *Mod. Phys. Lett. A* **23** 3247 (2008).
9. J. Ovalle, Decoupling gravitational sources in general relativity: The extended case, *Phys. Lett. B* **788**, 213 (2019).
10. J. Ovalle and F. Linares, Tolman IV solution in the Randall-Sundrum braneworld, *Phys. Rev. D* **88**, 104026 (2013).
11. J. Ovalle et al., Anisotropic solutions by gravitational decoupling, *Eur. Phys. J. C* **78**, 122 (2018).
12. L. Gabbanelli, A. Rincon and C. Rubio, Gravitational decoupled anisotropies in compact stars, *Eur. Phys. J. C* **78**, 370 (2018).
13. M. Sharif and S. Sadiq, Gravitational decoupled charged anisotropic spherical solutions, *Eur. Phys. J. C* **78**, 410 (2018).
14. M. Estrada and F. Tello-Ortiz, A new family of analytical anisotropic solutions by gravitational decoupling, *Eur. Phys. J. C* **133**, 453 (2018).
15. L. Gabbanelli et al., A causal Schwarzschild-de Sitter interior solution by gravitational decoupling, *Eur. Phys. J. C* **79**, 486 (2019).
16. E. Contreras, Gravitational decoupling in 2+1-dimensional spacetimes with cosmological term, *Class. Quantum Grav.* **36**, 095004 (2019).
17. E. Contreras and P. Bargueño, Extended Gravitational Decoupling in 2+1 Dimensional Spacetimes, *Class. Quantum Grav.* **36**, 215009 (2019).
18. K. N. Singh et al., Minimally deformed anisotropic model of class one space-time by gravitational decoupling, *Eur. Phys. J. C* **79**, 851 (2019).
19. S. Hensh and Z. Stuchlik, Anisotropic Tolman VII solution by gravitational decoupling, *Eur. Phys. J. C* **79**, 834 (2019).
20. M. Sharif and Q. Ama-Tul-Mughani, Study of (1+2)-dimensional charged string cloud with minimal geometric deformation, *Int. J. Geom. Methods Mod. Phys.* **16**, 1950187 (2019); Gravitational Decoupled Solutions of Axial String Cosmology *Mod. Phys. Lett. A* **35**, 2050091 (2020).
21. F. X. L. Cedeño and E. Contreras, Gravitational decoupling in cosmology, *Phys. Dark Universe* **28**, 100543 (2020).
22. E. Contreras, J. Ovalle and R. Casadio, Gravitational decoupling for axially symmetric systems and rotating black holes, *Phys. Rev. D* **103**, 044020 (2021).

23. C. Brans and R. H. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev.* **124**, 3 (1961).
24. C. M. Will, The confrontation between general relativity and experiment, *Living Rev. Rel.* **4**, 4 (2001).
25. E. J. Weinberg, Some problems with extended inflation, *Phys. Rev. D* **40**, 3950 (1989).
26. J. Khoury and A. Weltman, Chameleon Cosmology, *Phys. Rev. D* **69**, 044026 (2004).
27. L. Perivolaropoulos, PPN parameter γ and solar system constraints of massive brans-dicke theories, *Phys. Rev. D* **81**, 047501 (2010).
28. M. Sharif, and A. Majid, Anisotropic strange stars through embedding technique in massive Brans–Dicke gravity, *Eur. Phys. J. Plus* **135**, 558 (2020); Quark stars in massive Brans–Dicke gravity with tolman–kuchowicz spacetime, *Universe* **6**, 8 (2020).
29. M. Sharif and S. Saba, Gravitational decoupled anisotropic solutions in $f(\mathcal{G})$ gravity, *Eur. Phys. J. C* **78**, 921 (2018); M. Sharif and A. Waseem, Anisotropic spherical solutions by gravitational decoupling in $f(\mathcal{R})$ gravity, *Ann. Phys.* **405**, 14 (2019); M. Sharif and A. Majid, Anisotropic compact stars in self-interacting Brans-Dicke gravity, *Astrophys. Space Sci.* **365**, 42 (2020); Decoupled anisotropic spheres in self-interacting Brans-Dicke gravity, *Chin. J. Phys.* **68**, 406 (2020); Extended gravitational decoupled solutions in self-interacting Brans–Dicke theory, *Phys. Dark Universe* **30**, 100610 (2020).
30. R. C. Tolman, Static solutions of einstein's field equations for spheres of fluid, *Phys. Rev.* **55**, 364 (1939).
31. M. Zubair and H. Azmat, Anisotropic Tolman V solution by minimal gravitational decoupling approach, *Ann. Phys.* **420**, 168248 (2020).
32. M. K. Jasim et al., Charged strange stellar model describing by Tolman V metric, *Results Phys.* **20**, 103648 (2021).
33. W. F. Bruckman and E. Kazes, Properties of the solutions of cold ultradense configurations in the Brans-Dicke theory, *Phys. Rev. D* **16**, 2 (1977).
34. J. Kapusta, *Finite-Temperature Field Theory* (Cambridge University Press, Cambridge, 1994).
35. K. Lake, All static spherically symmetric perfect-fluid solutions of einstein's equations, *Phys. Rev. D* **67**, 104015 (2003).
36. Y. Fujii and K. Maeda, *The Scalar-Tensor Theory of Gravitation* (Cambridge University Press, Cambridge, 2003).
37. H. Heintzmann and W. Hillebrandt, Neutron Stars with an Anisotropic Equation of State: Mass, Redshift and Stability, *Astron. Astrophys.* **24**, 51 (1975).

Big-bounce in projectively invariant Nieh-Yan models: The Bianchi I case

Flavio Bombacigno

Departament de Física Teòrica and IFIC, Centro Mixto Universitat de València - CSIC, Universitat de València, Burjassot 46100, València, Spain
E-mail: flavio.bombacigno@ext.uv.es

Simon Boudet

Dipartimento di Fisica, Università di Trento, Via Sommarive 14, I-38123 Povo (TN), Italy
and
Trento Institute for Fundamental Physics and Applications (TIFPA)-INFN, Via Sommarive 14, I-38123 Povo (TN), Italy
E-mail: simon.boudet@unitn.it

Gonzalo J. Olmo

Departament de Física Teòrica and IFIC, Centro Mixto Universitat de València - CSIC, Universitat de València, Burjassot 46100, València, Spain
E-mail:gonzalo.olmo@uv.es

Giovanni Montani

Physics Department, "Sapienza" University of Rome, P.le Aldo Moro 5, 00185 (Roma), Italy
and
ENEA, Fusion and Nuclear Safety Department, C. R. Frascati, Via E. Fermi 45, 00044 Frascati (Roma), Italy
E-mail: giovanni.montani@enea.it

We show that the Nieh-Yan topological invariant breaks projective symmetry and loses its topological character in presence of non vanishing nonmetricity. The notion of the Nieh-Yan topological invariant is then extended to the generic metric-affine case, defining a generalized Nieh-Yan term, which allows to recover topologicity and projective invariance, independently. As a concrete example a class of modified theories of gravity is considered and its dynamical properties are investigated in a cosmological setting. In particular, bouncing cosmological solutions in Bianchi I models are derived. Finite time singularities affecting these solutions are analysed, showing that the geodesic completeness and the regular behavior of scalar perturbations in these space-times are not spoiled.

Keywords: Nieh-Yan, metric-affine, Bianchi I, big-bounce

1. Introduction

The theory of General Relativity (GR)[1,2] relies on the geometric interpretation of the gravitational field, described in terms of a metric tensor and a connection on

a pseudo-Riemannian manifold. Both GR and many alternative theories of gravity are based on a metric formulation, in which the connection is given by the symmetric and metric compatible Levi-Civita connection, which is completely determined by the metric and its derivatives. An alternative formulation for geometric theories of gravity consists in adopting the metric-affine paradigm, in which the metric tensor and the connection are considered as independent variables. In this approach, symmetry and metric compatibility of the connection are not imposed a priori, resulting in the presence of torsion and nonmetricity, respectively. Well known examples of metric-affine theories are Ricci based gravity,[3,4] Palatini $f(R)$ theory,[5] quadratic gravity,[6] Born-Infeld-type models,[7] general teleparallel models,[8] generalized hybrid metric-Palatini gravity[9–13] and metric-affine extension of higher order theories.[14–17]

The metric-affine approach plays a crucial role also in one of the current attempts to quantize gravity, i.e. loop quantum gravity (LQG),[18,19] where GR is reformulated in terms of a gauge $SU(2)$ connection (Ashtekar-Barbero-Immirzi connection) and its conjugate momentum, the densitized triad.[20–23] This formulation, indeed, can be derived[24] by including an additional contribution to the first order (Palatini) action of GR, namely the Nieh-Yan (NY) topological invariant[25,26] (the Holst term[27] can be used as well). The NY term was discovered in the context of Riemann-Cartan theory (where nonmetricity is set to zero) and its main property is topologicity: it reduces to a boundary term without affecting the field equations at all. This additional term is driven by the so called Immirzi parameter β,[28,29] which concurs in the definition of the Ashtekar variables and is related to a quantization ambiguity.[30] Attempts to address this issue led to the proposal of considering the Immirzi parameter as a new fundamental field,[31–33] an idea that has been later developed within several different contexts.[31,33,33–40,40–46] The promotion of such constant parameter to a dynamical field is usually pursued "by hand", substituting $\beta \to \beta(x)$ in the Lagrangian and possibly adding a potential term $V(\beta)$.

More recently, beside LQG the NY term has been studied in the context of teleparallel gravity[47] and in condensed matter physics.[48–51]

Another important property we will focus on, is projective invariance,[52,53] which has recently been shown to be of crucial importance in metric-affine theories since the breaking of this symmetry can give rise to dynamical instabilities.[54] In this regard, we want to stress that the NY term breaks this symmetry. This feature has always been neglected in literature and a revision of previous formulations seems necessary. Moreover, as will be shown in the following, the topological character of the NY term is also lost when nonmetricity is included.

The approach followed in this note is grounded on the choice of recovering these features from the very beginning in the action, without imposing any restriction on the affine connection. After a formal discussion, we will implement the gravitational model in a cosmological setting. In particular, we investigate Bianchi I models,[55–57] focusing our attention on the emergence of a classical bouncing cosmology.[42,58–64]

2. The role of nonmetricity in the Nieh-Yan term

In Einstein-Cartan theory the NY term[25,26] is explicitly defined as

$$NY \equiv \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\left(\frac{1}{2}T^{\lambda}{}_{\mu\nu}\,T_{\lambda\rho\sigma} - R_{\mu\nu\rho\sigma}\right)\,, \tag{1}$$

where

$$R^{\rho}{}_{\mu\sigma\nu} = \partial_{\sigma}\Gamma^{\rho}{}_{\mu\nu} - \partial_{\nu}\Gamma^{\rho}{}_{\mu\sigma} + \Gamma^{\rho}{}_{\tau\sigma}\Gamma^{\tau}{}_{\mu\nu} - \Gamma^{\rho}{}_{\tau\nu}\Gamma^{\tau}{}_{\mu\sigma}, \tag{2}$$

is the Riemann tensor built with the independent connection and

$$T^{\mu}{}_{\nu\rho} = \Gamma^{\mu}{}_{\nu\rho} - \Gamma^{\mu}{}_{\rho\nu} \tag{3}$$

is the torsion tensor. The starting point of our discussion is the observation that for a non-vanishing nonmetricity tensor $Q_{\mu\nu\rho} = -\nabla_{\mu}g_{\nu\rho}$, the NY term (1) is spoilt of its topological character. Indeed, extracting the nonriemmanian part of the Riemann tensor leads to[65]

$$NY = -\frac{1}{2}\bar{\nabla}\cdot S - \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}T^{\lambda}{}_{\mu\nu}\,Q_{\rho\sigma\lambda}, \tag{4}$$

where $\bar{\nabla}_{\mu}$ is built with the Levi-Civita connection $\bar{\Gamma}^{\mu}{}_{\nu\rho}$ and

$$S_{\mu} \equiv \varepsilon_{\mu\nu\rho\sigma}T^{\nu\rho\sigma}. \tag{5}$$

Therefore, when $Q_{\rho\mu\nu} \neq 0$, the Nieh-Yan term does not simply reduce to the divergence of a vector, and the appearance of nonmetricity spoils the topologicity. Let us now consider the behavior of (1) under projective transformations of the connection, namely

$$\tilde{\Gamma}^{\rho}{}_{\mu\nu} = \Gamma^{\rho}{}_{\mu\nu} + \delta^{\rho}{}_{\mu}\xi_{\nu}, \tag{6}$$

It can easily be shown that (1) is also not invariant under projective transformations, since

$$\frac{1}{4}\varepsilon^{\mu\nu\rho\sigma}\tilde{T}^{\lambda}{}_{\mu\nu}\,\tilde{T}_{\lambda\rho\sigma} - \frac{1}{4}\varepsilon^{\mu\nu\rho\sigma}T^{\lambda}{}_{\mu\nu}\,T_{\lambda\rho\sigma} = -S_{\mu}\xi^{\mu}. \tag{7}$$

Now, by looking at (4), we point out that a newly topological Nieh-Yan term can be recovered by simply setting

$$NY^{*} \equiv NY + \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}T^{\lambda}{}_{\mu\nu}\,Q_{\rho\sigma\lambda}. \tag{8}$$

We note that projective invariance is now enclosed as well, since

$$\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\tilde{T}^{\lambda}{}_{\mu\nu}\,\tilde{Q}_{\rho\sigma\lambda} - \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}T^{\lambda}{}_{\mu\nu}\,Q_{\rho\sigma\lambda} = +S_{\mu}\xi^{\mu}, \tag{9}$$

which exactly cancels out (7). We stress, however, that projective invariance is not strictly related to topologicity, and suitable generalizations of (2) breaking up only with the latter can be actually formulated. Let us consider, for instance, the following modified Nieh-Yan term

$$NY_{gen} \equiv \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\left(\frac{\lambda_1}{2}T^{\lambda}{}_{\mu\nu}\,T_{\lambda\rho\sigma} + \lambda_2\,T^{\lambda}{}_{\mu\nu}\,Q_{\rho\sigma\lambda} - R_{\mu\nu\rho\sigma}\right), \tag{10}$$

where we introduced the real parameters λ_1, λ_2. In this case the term (10) transforms under a projective transformation as

$$NY_{gen} \to NY_{gen} - (\lambda_1 - \lambda_2)\xi^\mu S_\mu, \tag{11}$$

so that by setting $\lambda_1 = \lambda_2$ we can recover again projective invariance, despite topologicity being in general violated if $\lambda_1 = \lambda_2 \neq 1$, since

$$NY_{gen} = -\frac{1}{2}\bar{\nabla}\cdot S + \frac{(\lambda_1 - 1)}{4}\varepsilon^{\mu\nu\rho\sigma}T^\lambda{}_{\mu\nu}T_{\lambda\rho\sigma} + \frac{(\lambda_2 - 1)}{2}\varepsilon^{\mu\nu\rho\sigma}T^\lambda{}_{\mu\nu}Q_{\rho\sigma\lambda}. \tag{12}$$

In the following, we will consider the general form (10), which by a suitable choice of the parameters $\lambda_{1,2}$ can reproduce all known actions usually studied in literature, as the Holst ($\lambda_1 = \lambda_2 = 0$) or the standard Nieh-Yan (4) ($\lambda_1 = 1$, $\lambda_2 = 0$) terms.

3. Generalized Nieh-Yan models

As a specific gravitational model featuring the generalized NY term we consider an action defined by a general function of two arguments, the Ricci scalar and the generalized NY term (10):

$$S_g = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,F(R, NY_{gen}), \tag{13}$$

Now, performing the transformation to the Jordan frame leads to the scalar tensor representation

$$S_g = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\left(\phi R + \beta NY_{gen} - W(\phi,\beta)\right), \tag{14}$$

with $\phi \equiv \frac{\partial F}{\partial R}$, $\beta \equiv \frac{\partial F}{\partial NY_{gen}}$ and $W \equiv \phi R(\phi,\beta) + \beta NY_{gen}(\phi,\beta) - F(\phi,\beta)$.

The scalar field β can be identified with the Immirzi field, which acquires in this way a dynamical character without the need of introducing this feature by hand in the action. Moreover, this formulation offers a viable mechanism to produce an interaction term $W(\phi,\beta)$ as well. Now, the field equation for the connection are obtained varying (14) with respect to $\Gamma^\mu{}_{\nu\rho}$. For the full set of equations the reader may cosnult,[65] while here we are interest in the following contraction

$$\delta^\lambda_\nu \frac{\delta S_g}{\delta \Gamma^\lambda{}_{\nu\mu}} = 0 \tag{15}$$

which leads to

$$(\lambda_1 - \lambda_2)S^\mu = 0. \tag{16}$$

This implies that the features of the solutions depend on the parameters λ_1 and λ_2, and when projective invariance is broken ($\lambda_1 \neq \lambda_2$) one is compelled to set $S_\mu = 0$. In this case, (10) can be re-expressed as

$$NY_{gen} = -\frac{1}{2}\bar{\nabla}\cdot S - \frac{(1-\lambda_1)}{3}S^\mu T^\rho{}_{\mu\rho} - \frac{(1-\lambda_2)}{2}S^\mu Q^\rho{}_{\mu\rho}, \tag{17}$$

implying that the generalized Nieh-Yan term (10) is identically vanishing on half-shell. In other words, terms violating projective invariance are harmless along the dynamics. This can be further appreciated deriving the effective scalar tensor action stemming from (14), when the solutions of the full set of connection field equations are plugged in it. Explicit calculations (see[65] for details) lead to

$$S = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\left(\phi\bar{R} + \frac{3}{2\phi}\bar{\nabla}_\mu\phi\bar{\nabla}^\mu\phi - W(\phi,\beta)\right), \tag{18}$$

where $\bar{R}$ is the Ricci scalar of the Levi-CIvita connection. This resembles the form of a Palatini $f(R)$ theory, with a potential depending on the Immirzi field as well. The equation for the latter, i.e.

$$\frac{\partial W(\phi,\beta)}{\partial\beta} = 0, \tag{19}$$

fixes its form in terms of the remaining scalar field: $\beta = \beta(\phi)$. Then, using the trace of the metric field equations, variation of (18) with respect to ϕ results in the usual structural equation featuring Palatini $f(R)$ theories,[5] i.e.

$$\left[2W(\phi,\beta) - \phi\frac{\partial W(\phi,\beta)}{\partial\phi}\right]_{\beta=\beta(\phi)} = \kappa T, \tag{20}$$

where T is the trace of the stress energy tensor of matter. This implies that the dynamics of the scalaron ϕ is frozen as well, and completely determined by T. Conditions (19) and (20) then guarantee that the scalar fields ϕ, β are not propagating degrees of freedom, and reduce to constants in vacuum, where the theory is stable and the breaking of projective invariance does not lead to ghost instabilities, in contrast to.[54]

If $\lambda_1 = \lambda_2 \equiv \lambda$, instead, the projective invariance of the model can be used to get rid of one vector degree of freedom, which can be set to zero properly choosing the vector ξ_μ. A convenient choice consists in setting $\xi_\mu = -\frac{1}{2}Q^\rho{}_{\mu\rho}$, which allows to deal only with torsion in the connection field equations. The effective action stemming from (14) then reads (see[65])

$$S = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\left(\phi\bar{R} + \frac{3}{2\phi}\bar{\nabla}_\mu\phi\bar{\nabla}^\mu\phi - \frac{3\phi}{2}\frac{1}{\phi^{2\lambda}+(1-\lambda)^2\psi^2}\bar{\nabla}_\mu\psi\bar{\nabla}^\mu\psi - V(\phi,\psi)\right), \tag{21}$$

where we used the transformation $\psi \equiv \beta\phi^{\lambda-1}$ and redefined the potential as $V(\phi,\psi) = W(\phi,\psi\phi^{1-\lambda})$. In general, the Immirzi field is expected to be a well-behaved dynamical degree of freedom, since in the Einstein frame action, defined by the conformal rescaling $\tilde{g}_{\mu\nu} = \phi\, g_{\mu\nu}$, the kinetic term for the Immirzi field takes the form

$$-\frac{3}{2}\frac{\tilde{g}^{\mu\nu}\nabla_\mu\psi\nabla_\nu\psi}{\phi^{2\lambda}+(1-\lambda)^2\psi^2}. \tag{22}$$

Since $\phi^{2\lambda}+(1-\lambda)^2\psi^2 > 0$ for every value of ϕ, ψ and λ, (22) has always the correct sign and no ghost instability arise.[66]

Let us end this section with a remark on how previous results with vanishing nonmetricity can properly be recovered. In particular, the Riemann-Cartan structure of[32,33,36,37,46,67–69] can be replicated by inserting in (13) the condition of vanishing nonmetricity with a Lagrange multiplier, i.e. adding to the Lagrangian a term $l^{\rho\mu\nu}Q_{\rho\mu\nu}$, with $l^{\rho\mu\nu} = l^{\rho\nu\mu}$. Then, results of[32,33,36,37,46,67–69] are simply obtained by setting $\lambda_1 = 1$. The fact that the usual Einstein-Cartan NY invariant and related models are recovered in this way, supports the correctness of our generalization of the NY term, with respect to other possible generalizations.

4. Big bounce in Bianchi I cosmology

In this section we consider dynamical models, i.e. those described by (21), which are characterized by a dynamical Immirzi field and look for cosmological solutions in Bianchi I spacetimes. In particular, we will be interested in obtaining solutions characterized by a bouncing behavior for the universe volume, thanks to which the big bang singularity is regularized in favour of a big bounce scenario.

Let us start from the equations of motion for the metric and scalar fields which are given by

$$\bar{G}_{\mu\nu} = \frac{\kappa}{\phi}T_{\mu\nu} + \frac{1}{\phi}\left(\bar{\nabla}_\mu\bar{\nabla}_\nu - g_{\mu\nu}\bar{\Box}\right)\phi - \frac{3}{2\phi^2}\bar{\nabla}_\mu\phi\bar{\nabla}_\nu\phi + \frac{3}{2}\frac{\bar{\nabla}_\mu\psi\bar{\nabla}_\nu\psi}{\phi^{2\lambda}+(1-\lambda)^2\psi^2} + \frac{1}{2}g_{\mu\nu}\left(\frac{3(\bar{\nabla}\phi)^2}{2\phi^2} - \frac{3}{2}\frac{(\bar{\nabla}\psi)^2}{\phi^{2\lambda}+(1-\lambda)^2\psi^2} - \frac{V(\phi,\psi)}{\phi}\right), \tag{23}$$

$$2V(\phi,\psi) - \phi\frac{\partial V(\phi,\psi)}{\partial\phi} + \frac{3\lambda\phi^{2\lambda+1}}{\left(\phi^{2\lambda}+(1-\lambda)^2\psi^2\right)^2}(\bar{\nabla}\psi)^2 = \kappa T, \tag{24}$$

$$\bar{\Box}\psi - \frac{(1-\lambda)^2\psi}{\phi^{2\lambda}+(1-\lambda)^2\psi^2}(\bar{\nabla}\psi)^2 + \left(1 - \frac{2\lambda\phi^{2\lambda}}{\phi^{2\lambda}+(1-\lambda)^2\psi^2}\right)\bar{\nabla}_\mu\ln\phi\bar{\nabla}^\mu\psi = \frac{\partial V(\phi,\psi)}{3\partial\psi}. \tag{25}$$

As will be shown, cosmological solutions can be found for projective invariant models ($\lambda = 1$), and restricting to potentials of the form $V(\phi,\psi) = V(\phi)$. To this end, we consider the metric for a Bianchi I flat spacetime, i.e.

$$ds^2 = -dt^2 + a(t)^2dx^2 + b(t)^2dy^2 + c(t)^2dz^2, \tag{26}$$

which represents a homogeneous but anisotropic spacetime, with three different scale factors $a(t), b(t), c(t)$. We include matter in the form of a perfect fluid with stress energy tensor

$$T_{\mu\nu} = \mathrm{diag}(\rho, a^2p, b^2p, c^2p), \tag{27}$$

where ρ is the energy density and p the pressure. Assuming the equation of state $p = w\rho$, conservation of the stress energy tensor implies

$$\rho(t) = \frac{\mu^2}{(abc)^{w+1}}, \tag{28}$$

where μ^2 is a constant. Now, considering a Starobinsky quadratic potential[70]

$$V(\phi) = \frac{1}{\alpha}\left(\frac{\phi - 1}{2}\right)^2, \tag{29}$$

it can be shown that the equations for the scalar fields yield[a]

$$\dot{\psi} = \frac{k_0 \phi}{v}, \qquad \phi = \frac{v^2 f(v)}{6\alpha k_0^2 + v^2}, \tag{30}$$

in terms of the volume-like variable $v \equiv abc$ and the function $f(v) \equiv 1 - 2\alpha\kappa(3w - 1)\rho(v)$, while k_0 is an integration constant. Regarding the metric field equations, lengthy computations[65] allow to rewrite the tt component in the form of a modified Friedmann equation:

$$H^2 \equiv \left(\frac{\dot{v}}{3v}\right)^2 = \frac{\frac{\kappa}{3}\left(\frac{\mu_I^2}{v^2} + \frac{\rho}{\phi} + \frac{\mu_{AN}^2}{\phi^2 v^2}\right) + \frac{V(\phi)}{6\phi}}{\left(1 + \frac{3v}{2}\frac{d}{dv}\ln\phi\right)^2}, \tag{31}$$

where μ_A^2 and μ_I^2 are constants, representing the anisotropy density parameter and the energy density parameter of the Immirzi field, respectively. We note that the r.h.s is a rational function of the volume alone.

In the following we will take into account dust and radiation as matter contributions, specified by the choices $w = 0$, $\mu = \mu_D$ and $w = 1/3$, $\mu = \mu_R$ in (28), respectively. Then, bouncing solutions can be derived for $\alpha < 0$ integrating Eq. (31) for $v(t)$, which then yields the scalar fields behavior via (30).

4.1. *Vacuum case*

It is convenient to first focus on the vacuum case, where $f(v) = 1$ and $\mu_D = \mu_R = 0$. In this case we found the big bounce in the volume variable depicted in Fig. 1, where also the behavior of each scale factor is shown. We note that the volume is affected by a future finite-time *singularity*[71,72] where the Hubble function diverges, while the scale factors are always finite and nonvanishing. Such singular points will be carefully investigated in the next section studying the geodesic completeness and scalar perturbations across them. Here we just note that, in general, quantum effects of particle creation[73–76] can produce additional effective terms in the Friedman equation, able to regularize singularities of the Hubble function.

Finally, the scalar fields behaviour is shown in Fig. 2 where the field ϕ asymptotically reaches unity as $t \to \infty$, while the Immirzi field relaxes to a constant Immirzi parameter.

4.2. *Radiation and dust*

The above analysis can be extended to the non vacuum case, including the energy density of radiation and dust. A first difference concerns the late time region, since

[a] A dot denotes derivatives with respect to time.

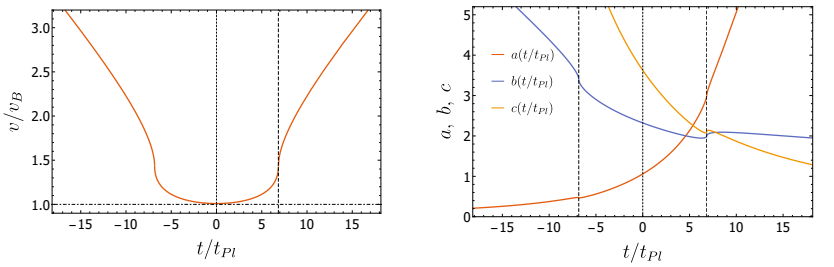


Fig. 1. Numerical solutions for $\alpha = -5/3$, $\mu_I = \sqrt{3}$, $\mu_A = 0.2\mu_I$ as a function of t/t_{Pl}. Dotted and dashed lines represent where bounce and future time singularity happen, respectively. The bounce is centered at the origin of time for convenience, and the values of the parameters are chosen in order to yield graphs that display features in a clear fashion. (a) Universe volume normalised to v_B. (b) Scale factors.

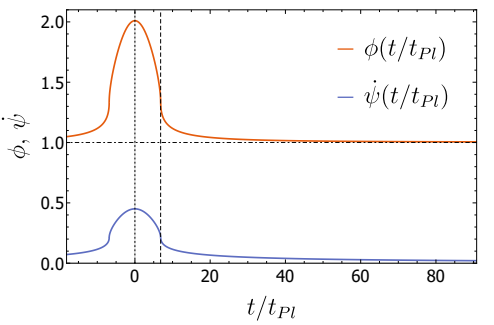


Fig. 2. Scalaron ϕ and Immirzi field derivative $\dot{\psi}$ for $\alpha = -5/3$, $\mu_I = \sqrt{3}$, $\mu_A = 0.2\mu_I$ as a function of t/t_{Pl}.

both radiation and dust are able to provide an isotropization mechanism for the universe, as is apparent studying the anisotropy degree

$$A(t) = \frac{\left(H_A^2 + H_B^2 + H_C^2\right)}{3H^2} - 1, \tag{32}$$

whose behavior is shown in Fig. 3. Regarding the universe in its early phase, instead, two different scenarios may occur, depending on the value of the parameter α. If $\bar{\alpha} < \alpha < 0$, where $\bar{\alpha} = -2\mu_I^2/\mu_D^4$, the solutions are qualitatively equivalent to the vacuum case. If $\bar{\alpha} > \alpha$, instead, we still obtain a bouncing behavior for the volume (See Fig. 4), which however is now devoid of singularities. However, now the scale factors either diverge or vanish at some critical time t_c.

5. Physical implications of curvature divergences

As described in the previous section there are two classes of solutions, characterized by a singularity in the Hubble function and non vanishing and finite scale factors

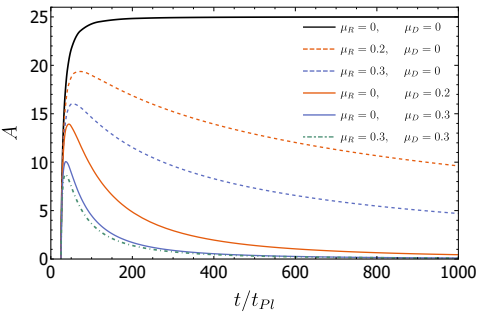
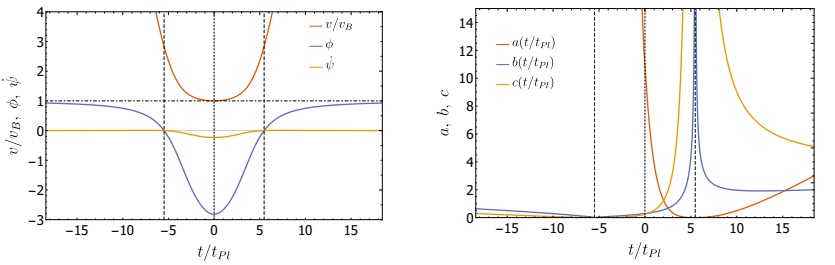


Fig. 3. Asymptotic behavior of the anisotropy degree A as a function of t/t_{Pl} after the finite time singularity for various values of μ_R, μ_D and $\alpha = -5/3$, $\mu_I = \sqrt{3}$, $\mu_A = 0.2\mu_I$.



(a) (b)

Fig. 4. Numerical solutions as a function of t/t_{Pl} for $\mu_I = 0.057$, $\mu_A = 2.4$, $\mu_D = 0.365$, $\mu_R = 1.56$ and $\alpha = -8.42 < \bar{\alpha}$. The dashed lines represent where the scalaron vanishes. (a) Volume normalised to v_B, ϕ and derivative of the Immirzi field. (b) Scale factors.

(Fig. 1) or with regular Hubble function but vanishing or divergent scale factors (Fig. 4). In this section we will analyse such singularities studying the behavior of null geodesics and of scalar perturbations near the critical time t_c, at which the singularity is located.

Regarding null geodesics with tangent vector $u^\alpha = dx^\alpha/ds$, it can be shown that they admit a first integral of the form[77,78]

$$x' = \frac{k_a}{a^2}, \qquad y' = \frac{k_b}{b^2}, \qquad z' = \frac{k_c}{c^2}, \qquad t' = \left(\frac{k_a^2}{a^2} + \frac{k_b^2}{b^2} + \frac{k_c^2}{c^2}\right)^{1/2} + C_0, \tag{33}$$

where prime denotes derivative with respect to the affine parameter s and k_a, k_b, k_c and C_0 are integration constants. It follows that if $a(t)$, $b(t)$, and $c(t)$ are continuous and non-vanishing, as in Fig. 1, the tangent vector to the geodesics will be unique and well defined. Therefore, such cases are geodesically complete, a result that holds both in the anisotropic and in the isotropic case.[79]

In the other class of solutions (Fig. 4), instead, we see that the volume remains finite despite the vanishing/divergence of some scale factors. The divergence of individual scale factors does not affect the geodesics, but the vanishing of some of

them may spoil the continuity and lead to the impossibility of a unique extension across t_c. Indeed, let us consider the case in which one scale factor vanishes at some affine parameter s_c. In particular, suppose that

$$a(s) = a_0(s - s_c)^\gamma, \tag{34}$$

with $\gamma > 0$. Then, integrating the relevant equations yields

$$x(s) = x_c + \frac{k_a(s - s_c)^{1-2\gamma}}{a_0^2(1 - 2\gamma)}, \tag{35}$$

$$t(s) = t_c + C_0(s - s_c) + \frac{k_a(s - s_c)^{1-\gamma}}{a_0(1 - \gamma)}, \tag{36}$$

which are smooth if $0 < \gamma < 1/2$ and $0 < \gamma < 1$, respectively. Note that if $1/2 < \gamma < 1$, then $x(s) \xrightarrow{s \to s_c} \pm\infty$, which would imply reaching infinity in finite coordinate time. Conversely, if $0 < \gamma < 1/2$, then the geodesic path will span the range $\{t, x\} \in (-\infty, \infty)$, and the geodesics would be complete, despite the vanishing of some scale factors. However, this result is dependent on how rapidly the zero is reached, i.e. we have to determine the value of γ relative to the solutions in question. The results are shown in Fig. 5 and prove that the solution approaches zero too rapidly, corresponding to a value of γ larger than $1/2$. We are thus forced to conclude that the example shown in Fig. 4 does represent a geodesically incomplete space-time.

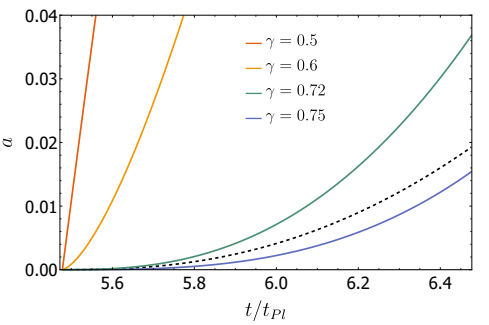


Fig. 5. Outcomes of null geodesics test for $\alpha < \bar{\alpha}$. Scale factor $a(s(t))$ for different values of γ. The dashed-black line represent the numerical solution $a(t)$ reported in Fig. 4.

We now turn the attention to the behavior of scalar field perturbations. For a scalar mode given by $\sigma_{\vec{k}}(t, \vec{x}) = \Theta(t)e^{i\vec{k}\cdot\vec{x}}$, one finds[65] an equation of the form

$$\ddot{\Theta} + h(v)\frac{\dot{v}}{v}\dot{\Theta} + \left(\frac{k_x^2}{a^2} + \frac{k_y^2}{b^2} + \frac{k_z^2}{c^2}\right)\Theta = 0\ , \tag{37}$$

where $h(v)$ represents some regular function of the volume v and $\vec{k} = (k_x, k_y, k_z)$ represents a set of constants. From this expression, we see that scalar modes are sensible to the presence of the individual scale factors a, b, and c, and of the Hubble

function $H = \dot{v}/3v$. If the scalar factors do not vanish anywhere, then any potential problems should come only from the term involving the Hubble function, which diverges in t_c. In particular, in vacuum one finds[65] the following approximate solution near the singularity

$$\Theta(t) \approx \Theta_c + \frac{\dot{\Theta}_c}{2\tilde{h}_c^2} e^{\mp 2\tilde{h}_c|t-t_c|^{1/2}} \left(1 \pm 2\tilde{h}_c|t-t_c|^{1/2}\right) , \tag{38}$$

with $\Theta_c, \dot{\Theta}_c, \tilde{h}_c$ constants, from which we see that scalar field perturbations remain bounded around t_c despite the divergence in the Hubble function. A similar result holds in presence of dust and radiation, since

$$\begin{aligned}\Theta \approx \Theta_c - \frac{\dot{\Theta}_c}{32\tilde{h}_c^4} e^{\mp 4\tilde{h}_c|t-t_c|^{1/4}} \Big(3 \pm 12\tilde{h}_c|t-t_c|^{1/4} \\ +24\tilde{h}_c^2|t-t_c|^{1/2} \pm 32\tilde{h}_c^3|t-t_c|^{3/4}\Big) ,\end{aligned} \tag{39}$$

which is easy to see to be again bounded. On the other hand, for a regular Hubble function but vanishing scale factors, equation (37) describes a harmonic oscillator with a time dependent frequency,

$$\ddot{\Theta}(t) + \frac{k_x^2}{a^2(t)}\Theta(t) \approx 0, \tag{40}$$

which diverges as $a(t) \to 0$. Therefore, for the values of $a(t)$ obtained numerically in the previous section neither geodesics nor scalar perturbations are well behaved.

6. Conclusion

We proposed a generalization of the Nieh-Yan term to metric-affine gravity, by including an additional term featuring nonmetricity and inserting two parameters (λ_1, λ_2), which allow to recover the projective invariance and the topological character, otherwise lost in presence of nonmetricity. In particular, projective invariance can be independently obtained by setting $\lambda_1 = \lambda_2 = \lambda$, whereas topologicity is only guaranteed for $\lambda = 1$.

As an explicit example, we considered a model with Lagrangian $F(R, NY_{gen})$, which re-expressed in the Jordan frame features two scalar fields. We identified these scalar degrees as the $f(R)$-like scalaron ϕ and the Immirzi field β and showed that the latter acquires dynamical character and a potential term in a more natural way than in previous treatments, where these features are introduced by hand in the action. Depending on the values of λ_1 and λ_2, we found two different scalar tensor theories. Models with $\lambda_1 \neq \lambda_2$, are non-dynamical, and the scalar fields are frozen to constant values in vacuum. In the projective invariant case ($\lambda_1 = \lambda_2$), instead, the theory is endowed with one additional dynamical degree of freedom, the Immirzi field, while the field ϕ is algebraically related to the latter via a modified structural equation.

Then, we considered in more detail the dynamical models, looking for cosmological solutions in Bianchi I spacetimes. We found numerical solutions in which the big bang singularity is replaced by a big bounce scenario, in which the universe volume undergoes a contraction up to a minimum value and then bounces back, re-expanding in another branch. The scalaron and the Immirzi field reach a maximum during the bounce and relax to constant values at later times, where the standard LQG picture, with $\phi = 1$ and a constant Immirzi parameter $\beta = \beta_0$, is recovered. Moreover, the inclusion of dust and radiation turns out to provide an isotropization mechanism at late times, a feature that is absent in the vacuum case.

Such solutions are characterized by future finite time singularities after the bounce, either in the Hubble function or in the individual scale factors. In the former case, we showed that null geodesics are still well behaved and scalar perturbations bounded, which allows us to conclude that the solution is physically acceptable. In the latter case, instead, the study of null geodesics shows that they cannot be extended across the singular point, where also scalar perturbations grow in time, leading us to regard such solutions as unphysical.

References

1. Wald R. M., *General Relativity* (University of Chicago Press, 1984).
2. T. K. S. W. J. A. Misner C. W. and K. D.I., *Gravitation* (Princeton University Press, 2017).
3. V. I. Afonso, G. J. Olmo and D. Rubiera-Garcia, Mapping Ricci-based theories of gravity into general relativity, *Phys. Rev. D* **97**, p. 021503 (2018).
4. V. I. Afonso, G. J. Olmo, E. Orazi and D. Rubiera-Garcia, Correspondence between modified gravity and general relativity with scalar fields, *Phys. Rev. D* **99**, p. 044040 (2019).
5. G. J. Olmo, PALATINI APPROACH TO MODIFIED GRAVITY: f(R) THEORIES AND BEYOND, *International Journal of Modern Physics D* **20**, 413 (2011).
6. F. S. N. Lobo, G. J. Olmo and D. Rubiera-Garcia, Semiclassical geons as solitonic black hole remnants, *JCAP* **07**, p. 011 (2013).
7. J. Beltran Jimenez, L. Heisenberg, G. J. Olmo and D. Rubiera-Garcia, Born–Infeld inspired modifications of gravity, *Phys. Rept.* **727**, 1 (2018).
8. J. Beltrán Jiménez, L. Heisenberg, D. Iosifidis, A. Jiménez-Cano and T. S. Koivisto, General teleparallel quadratic gravity, *Physics Letters B* **805**, p. 135422 (2020).
9. J. a. L. Rosa, F. S. N. Lobo and G. J. Olmo, Weak-field regime of the generalized hybrid metric-Palatini gravity (4 2021).
10. T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Metric-Palatini gravity unifying local constraints and late-time cosmic acceleration, *Phys. Rev. D* **85**, p. 084016 (2012).
11. S. Capozziello, T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Hybrid metric-Palatini gravity, *Universe* **1**, 199 (2015).
12. F. Bombacigno, F. Moretti and G. Montani, Scalar modes in extended hybrid metric-Palatini gravity: weak field phenomenology, *Phys. Rev. D* **100**, p. 124036 (2019).
13. K. A. Bronnikov, Spherically symmetric black holes and wormholes in hybrid metric-Palatini gravity, *Grav. Cosmol.* **25**, 331 (2019).

14. M. Borunda, B. Janssen and M. Bastero-Gil, Palatini versus metric formulation in higher curvature gravity, *JCAP* **11**, p. 008 (2008).
15. B. Janssen, A. Jiménez-Cano and J. A. Orejuela, A non-trivial connection for the metric-affine Gauss-Bonnet theory in $D = 4$, *Phys. Lett. B* **795**, 42 (2019).
16. B. Janssen and A. Jiménez-Cano, On the topological character of metric-affine Lovelock Lagrangians in critical dimensions, *Phys. Lett. B* **798**, p. 134996 (2019).
17. R. Percacci and E. Sezgin, New class of ghost- and tachyon-free metric affine gravities, *Phys. Rev. D* **101**, p. 084040 (2020).
18. C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, England, 2004).
19. T. Thiemann, *Modern Canonical Quantum General Relativity* (Cambridge University Press, Cambridge, England, 2007).
20. A. Ashtekar, New Variables for Classical and Quantum Gravity, *Phys. Rev. Lett.* **57**, 2244 (nov 1986).
21. A. Ashtekar, New Hamiltonian formulation of general relativity, *Physical Review D* **36**, 1587 (sep 1987).
22. A. Ashtekar, J. D. Romano and R. S. Tate, New variables for gravity: Inclusion of matter, *Physical Review D* **40**, 2572 (oct 1989).
23. A. Ashtekar and C. J. Isham, Representations of the holonomy algebras of gravity and nonAbelian gauge theories, *Classical and Quantum Gravity* **9**, 1433 (jun 1992).
24. G. Date, R. K. Kaul and S. Sengupta, Topological interpretation of Barbero-Immirzi parameter, *Phys. Rev. D* **79**, p. 44008 (feb 2009).
25. H. T. Nieh and M. L. Yan, An identity in Riemann–Cartan geometry, *Journal of Mathematical Physics* **23**, 373 (1982).
26. H. T. Nieh, A TORSIONAL TOPOLOGICAL INVARIANT, *International Journal of Modern Physics A* **22**, 5237 (2007).
27. S. Holst, Barbero's Hamiltonian derived from a generalized Hilbert-Palatini action, *Phys. Rev. D* **53**, 5966 (may 1996).
28. G. Immirzi, Real and complex connections for canonical gravity, *Class. Quant. Grav.* **14**, L177 (1997).
29. G. Immirzi, Real and complex connections for canonical gravity, *Classical and Quantum Gravity* **14**, L177 (oct 1997).
30. C. Rovelli and T. Thiemann, The Immirzi parameter in quantum general relativity, *Phys. Rev. D* **57**, 1009 (1998).
31. V. Taveras and N. Yunes, Barbero-Immirzi parameter as a scalar field: K-inflation from loop quantum gravity?, *Phys. Rev. D* **78**, p. 64070 (sep 2008).
32. G. Calcagni and S. Mercuri, Barbero-Immirzi field in canonical formalism of pure gravity, *Phys. Rev. D* **79**, p. 84004 (apr 2009).
33. F. Bombacigno, F. Cianfrani and G. Montani, Big-bounce cosmology in the presence of Immirzi field, *Phys. Rev. D* **94**, p. 64021 (sep 2016).
34. O. J. Veraguth and C. H. Wang, Immirzi parameter without Immirzi ambiguity: Conformal loop quantization of scalar-tensor gravity, *Physical Review D* **96**, p. 084011 (oct 2017).
35. C. H. Wang and D. P. Rodrigues, Closing the gaps in quantum space and time: Conformally augmented gauge structure of gravitation, *Physical Review D* **98**, p. 124041 (dec 2018).
36. F. Bombacigno and G. Montani, Implications of the Holst term in a f(R) theory with torsion, *Phys. Rev. D* **99**, p. 64016 (mar 2019).
37. F. Bombacigno and G. Montani, f(R) gravity with torsion and the Immirzi field: Signature for gravitational wave detection, *Phys. Rev. D* **97**, p. 124066 (jun 2018).

38. C. H.-T. Wang and M. Stankiewicz, Quantization of time and the big bang via scale-invariant loop gravity, *Physics Letters B* **800**, p. 135106 (jan 2020).
39. D. Iosifidis and L. Ravera, Parity Violating Metric-Affine Gravity Theories (9 2020).
40. F. Bombacigno, S. Boudet and G. Montani, Generalized ashtekar variables for palatini f(r) models, *Nuclear Physics B* **963**, p. 115281 (2021).
41. M. Långvik, J.-M. Ojanperä, S. Raatikainen and S. Räsänen, Higgs inflation with the Holst and the Nieh–Yan term, *Phys. Rev. D* **103**, p. 083514 (2021).
42. F. Bombacigno and G. Montani, Big bounce cosmology for Palatini R^2 gravity with a Nieh-Yan term, *The European Physical Journal C* **79**, p. 405 (may 2019).
43. S. Boudet, F. Bombacigno, G. Montani and M. Rinaldi, Superentropic black hole with immirzi hair, *Phys. Rev. D* **103**, p. 084034 (Apr 2021).
44. M. Lattanzi and S. Mercuri, A solution of the strong *cp* problem via the peccei-quinn mechanism through the nieh-yan modified gravity and cosmological implications, *Phys. Rev. D* **81**, p. 125015 (Jun 2010).
45. O. Castillo-Felisola, C. Corral, S. Kovalenko, I. Schmidt and V. E. Lyubovitskij, Axions in gravity with torsion, *Phys. Rev. D* **91**, p. 085017 (Apr 2015).
46. S. Mercuri, Peccei-Quinn Mechanism in Gravity and the Nature of the Barbero-Immirzi Parameter, *Phys. Rev. Lett.* **103**, p. 81302 (aug 2009).
47. M. Li, H. Rao and D. Zhao, A simple parity violating gravity model without ghost instability, *Journal of Cosmology and Astroparticle Physics* **2020**, 023 (nov 2020).
48. J. Nissinen and G. E. Volovik, Thermal nieh-yan anomaly in weyl superfluids, *Phys. Rev. Research* **2**, p. 033269 (Aug 2020).
49. J. Nissinen and G. E. Volovik, On Thermal Nieh-Yan Anomaly in Topological Weyl Materials, *JETP Letters* **110**, 789 (2019).
50. Z.-M. Huang, B. Han and M. Stone, Nieh-yan anomaly: Torsional landau levels, central charge, and anomalous thermal hall effect, *Phys. Rev. B* **101**, p. 125201 (Mar 2020).
51. C.-X. Liu, Phonon helicity and nieh-yan anomaly in the kramers-weyl semimetals of chiral crystals (2021).
52. V. I. Afonso, C. Bejarano, J. Beltran Jimenez, G. J. Olmo and E. Orazi, The trivial role of torsion in projective invariant theories of gravity with non-minimally coupled matter fields, *Class. Quant. Grav.* **34**, p. 235003 (2017).
53. D. Iosifidis, Linear Transformations on Affine-Connections, *Class. Quant. Grav.* **37**, p. 085010 (2020).
54. J. Beltrán Jiménez and A. Delhom, Ghosts in metric-affine higher order curvature gravity, *The European Physical Journal C* **79**, p. 656 (2019).
55. G. Montani, M. V. Battisti, R. Benini and G. Imponente, *Primordial cosmology* (World Scientific, Singapore, 2009).
56. A. Corichi and P. Singh, A Geometric perspective on singularity resolution and uniqueness in loop quantum cosmology, *Phys. Rev. D* **80**, p. 044024 (2009).
57. A. Y. Kamenshchik, E. O. Pozdeeva, A. A. Starobinsky, A. Tronconi, G. Venturi and S. Y. Vernov, Induced gravity, and minimally and conformally coupled scalar fields in Bianchi-I cosmological models, *Phys. Rev. D* **97**, p. 023536 (2018).
58. G. Montani and R. Chiovoloni, A Scenario for a Singularity-free Generic Cosmological Solution (3 2021).
59. G. Montani, A. Marchi and R. Moriconi, Bianchi I model as a prototype for a cyclical Universe, *Phys. Lett. B* **777**, 191 (2018).
60. F. Cianfrani, A. Marchini and G. Montani, The picture of the Bianchi I model via gauge fixing in Loop Quantum Gravity, *EPL* **99**, p. 10003 (2012).
61. R. Moriconi and G. Montani, Behavior of the Universe anisotropy in a big-bounce cosmology, *Phys. Rev. D* **95**, p. 123533 (2017).

62. E. Giovannetti, G. Montani and S. Schiattarella, Semiclassical and quantum features of the Bianchi I cosmology in the polymer representation (5 2021).
63. C. Barragan and G. J. Olmo, Isotropic and Anisotropic Bouncing Cosmologies in Palatini Gravity, *Phys. Rev. D* **82**, p. 084015 (2010).
64. C. Barragan, G. J. Olmo and H. Sanchis-Alepuz, Bouncing Cosmologies in Palatini f(R) Gravity, *Phys. Rev. D* **80**, p. 024016 (2009).
65. F. Bombacigno, S. Boudet, G. J. Olmo and G. Montani, Big bounce and future time singularity resolution in bianchi i cosmologies: The projective invariant nieh-yan case, *Phys. Rev. D* **103**, p. 124031 (Jun 2021).
66. G. J. Olmo, Post-Newtonian constraints on f(R) cosmologies in metric and Palatini formalism, *Phys. Rev. D* **72**, p. 083505 (2005).
67. S. Mercuri, Fermions in the Ashtekar-Barbero connection formalism for arbitrary values of the Immirzi parameter, *Phys. Rev. D* **73**, p. 84016 (apr 2006).
68. S. Mercuri, From the Einstein-Cartan to the Ashtekar-Barbero canonical constraints, passing through the Nieh-Yan functional, *Physical Review D - Particles, Fields, Gravitation and Cosmology* **77**, p. 024036 (jan 2008).
69. S. Mercuri and V. Taveras, Interaction of the Barbero-Immirzi field with matter and pseudoscalar perturbations, *Phys. Rev. D* **80**, p. 104007 (nov 2009).
70. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Physics Letters B* **91**, 99 (1980).
71. S. Nojiri, S. D. Odintsov and S. Tsujikawa, Properties of singularities in (phantom) dark energy universe, *Phys. Rev. D* **71**, p. 063004 (2005).
72. S. D. Odintsov and V. K. Oikonomou, Dynamical Systems Perspective of Cosmological Finite-time Singularities in $f(R)$ Gravity and Interacting Multifluid Cosmology, *Phys. Rev. D* **98**, p. 024013 (2018).
73. G. Montani, Influence of the particles creation on the flat and negative curved FLRW universes, *Class. Quant. Grav.* **18**, 193 (2001).
74. K. Dimopoulos, Is the Big Rip unreachable?, *Phys. Lett. B* **785**, 132 (2018).
75. L. H. Ford, Gravitational Particle Creation and Inflation, *Phys. Rev. D* **35**, p. 2955 (1987).
76. F. Contreras, N. Cruz, E. Elizalde, E. González and S. Odintsov, Linking little rip cosmologies with regular early universes, *Phys. Rev. D* **98**, p. 123520 (2018).
77. P. Singh, Curvature invariants, geodesics and the strength of singularities in Bianchi-I loop quantum cosmology, *Phys. Rev. D* **85**, p. 104011 (2012).
78. K. Nomura and D. Yoshida, Past extendibility and initial singularity in Friedmann-Lemaître-Robertson-Walker and Bianchi I spacetimes (5 2021).
79. J. Beltran Jimenez, R. Lazkoz, D. Saez-Gomez and V. Salzano, Observational constraints on cosmological future singularities, *Eur. Phys. J. C* **76**, p. 631 (2016).

Late time cosmology with derivatives of matter Lagrangian

Shahab Shahidi

School of Physics, Damghan University, Damghan, 36716-41167, Iran
E-mail: s.shahidi@du.ac.ir

A class of modified gravity theories with higher order derivative terms of a function of the matter Lagrangian $f(L_m)$ is considered. We will consider the Newtonian limit of the theory and show that the model predicts the standard Poisson equation for a massive test particle due to the higher order nature of the derivative matter coupling. Generally the energy momentum tensor is not conserved, leading to the fifth force similar to $f(R,T)$ theories. We will however show that in the FRW background the energy-momentum tensor is conserved. Cosmological implications of this model with different functions of the matter Lagrangian f will be investigated in details and we will show that current observational data can be satisfied. Evolution of the matter density perturbation in the longitudinal gauge is also considered for dust matter sources and we will show that the observational data can be satisfied in this model.

Keywords: Modified gravity, Late time cosmology.

1. Introduction

The accelerated expansion of the universe is one of the most interesting findings of modern cosmology which is confirmed by observations.[1] Our first and standard treatment to explain this expansion is to introduce a cosmological constant Λ to the standard Einstein Hilbert action so that the Lagrangian of the gravitational theory becomes

$$\mathcal{L} = \sqrt{-g}(R - 2\Lambda + L_m), \tag{1}$$

where R is the Ricci scalar and L_m is the matter Lagrangian. Together with cold dark matter which takes care of local celestial motions, this theory could explain almost all of current observations and is well-known as the ΛCDM model for cosmology. However, the standard ΛCDM model suffers from some phenomenological and also observational issues such as the cosmological constant problem,[2] the Hubble and σ_8 tensions,[3] etc. In fact there are many attempts to explain these problems in the context of ΛCDM model. For example in,[4] the authors have tried to explain why the σ_8 tension is not a problem at all.

One of the solutions to all of the aforementioned problems is to dismiss the cosmological constant in the Lagrangian and search for a new method to describe the dynamics of accelerated universe at late times. This can be done by introducing extra degrees of freedom to the theory such as scalar-tensor[5] or vector-tensor theories,[6] or by generalizing the gravitational action itself, such as $f(R)$ gravity theories[7] or massive gravity theories.[8] Another interesting idea among these modified

gravity theories, is to introduce some non-standard couplings between gravity and matter fields. These theories include minimal couplings such as energy-momentum powered gravity[9] or non-minimal couplings such as $f(R,T)$ theories and their generalizations.[10] In figure (1) we have summarized the most important generalizations of gravitational action. The general property in non standard matter coupling theories is that the energy-momentum tensor is no longer conserved. This implies that we have a fraction of matter creation from geometry in these classes of theories which could be tested by observations.

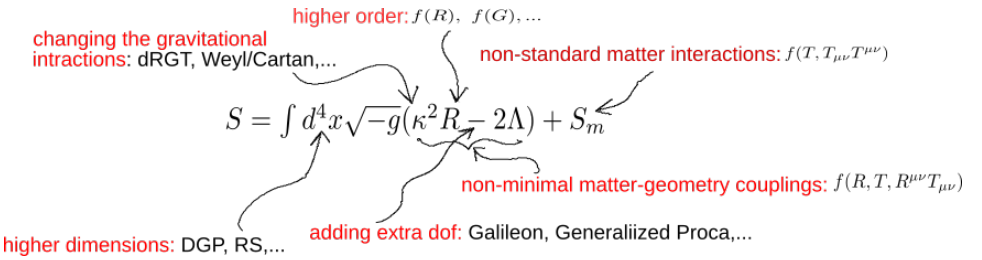


Fig. 1. Summary of modified gravity theories from the Einstein-Hilbert action.

2. The model

In this paper we introduce a new method to generalize the coupling between matter and geometry. Let us consider the evolution of the non-relativistic matter energy density with respect to the redshift z which is plotted in figure (2). One can see from this figure that the matter density is nearly flat around $z \approx 0$. This implies that the derivatives of matter density is nearly constant around present time which can mimic the cosmological constant (CC). This brings the idea that maybe derivatives of the

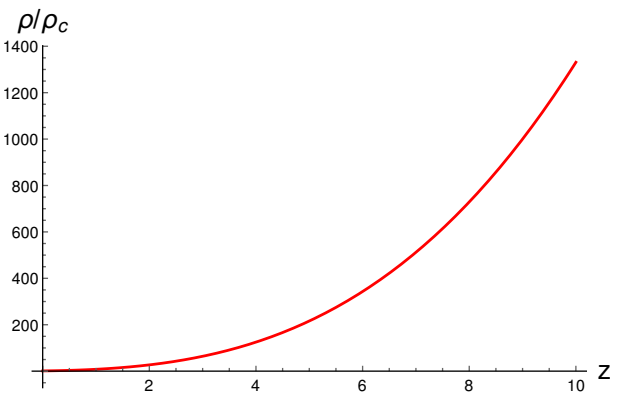


Fig. 2. The evolution of the pressure-less energy density as a function of redshift z.

matter density or matter Lagrangian could play the role of CC and be responsible for late time accelerated expansion of the universe. To explore this possibility, let us consider the action functional[11]

$$S = \int d^4x\sqrt{-g}\Big[\kappa^2(R-2\Lambda) + \alpha\,\nabla_\mu f\,\nabla^\mu f\Big] + S_m, \tag{2}$$

where $f = f(L_m)$ is an arbitrary function of the matter Lagrangian with mass dimension 1 and α is a dimensionless constant. It should be noted that the function f is a scalar field. As a result, one can consider the α-term in the action, as a canonical kinetic term for f. So, the theory (2) is a scalar tensor theory (or a simplest class of Galileon theory) where the scalar field is substituted by an energy-density valued scalar field. As we will see in the following, this rich structure allows us to explain the late time accelerated expansion of the universe. We should also note that the theory (2) is different from the so-called $f(R, L_m)$ gravity theories,[10] since in $f(R, L_m)$ theories, we have non-minimal couplings between matter and geometry but here we have a minimal derivative interaction of the matter Lagrangian.

In order to obtain the equation of motion of this model, note that the definition of the energy-momentum tensor is given by

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}L_m)}{\delta g^{\mu\nu}}, \tag{3}$$

which implies that

$$\delta L_m = \frac{1}{2}\big(L_m g_{\mu\nu} - T_{\mu\nu}\big)\delta g^{\mu\nu}. \tag{4}$$

As a result, one can obtain the equation of motion of the derivative matter coupling (DMC) model (2) as

$$\begin{aligned}\kappa^2(G_{\mu\nu} + \Lambda g_{\mu\nu}) + \alpha\, f'\Box f\,(T_{\mu\nu} - g_{\mu\nu}L_m)\\ + \alpha\left(\nabla_\mu f\nabla_\nu f - \frac{1}{2}g_{\mu\nu}\nabla_\alpha f\nabla^\alpha f\right) = \frac{1}{2}T_{\mu\nu},\end{aligned} \tag{5}$$

Here, prime denotes derivative with respect to the argument. As we mentioned above, the term proportional to the coupling constant α is the second Galileon equation of motion for the scalar field f.

The general behavior of the coupled matter-geometry theories is that the energy-momentum tensor is no longer conserved. In our model, one can easily obtain

$$\nabla^\mu T_{\mu\nu} = 2(T_{\mu\nu} - L_m g_{\mu\nu})\nabla^\mu \mathrm{Log}\big[1-\alpha f'\Box f\big]. \tag{6}$$

One can see that in the case $\alpha = 0$, the energy-momentum tensor becomes conserved. Here, we have to note an important property of the model. Let us assume that the matter content of the universe could be explained by a perfect fluid with Lagrangian $L_m = -\rho$ and energy-momentum tensor

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}. \tag{7}$$

In this case, the non-conservation of the energy-momentum tensor becomes

$$\nabla^{\mu}T_{\mu\nu} = -2(\rho + p)h_{\mu\nu}\,\nabla^{\mu}\mathrm{Log}\big[1-\alpha f'\,\Box f\big]. \tag{8}$$

where $h_{\mu\nu} = u_{\mu}u_{\nu} + g_{\mu\nu}$ is the projection tensor. On top of FRW space-time, the right hand side of the conservation equation (8) vanishes and we recover the conservation of the energy-momentum tensor in our model. This is however a new feature in the context of matter-geometry coupling theories.

3. Newtonian limit

Let us first consider the geodesics of the DMC model. We assume that the cosmological constant could be ignored in the low energy limit. One can easily prove that the conservation of the energy-momentum tensor (8) for a perfect fluid results in the geodesic equation of the form

$$\frac{d^2x^{\lambda}}{ds^2} + \Gamma^{\lambda}{}_{\mu\nu}u^{\mu}u^{\nu} = f^{\lambda}, \tag{9}$$

where

$$f^{\lambda} = -2h^{\lambda\nu}\nabla_{\nu}\log\Big[1 - \alpha f'\Box f\Big] - \frac{h^{\lambda\nu}\nabla_{\nu}p}{\rho + p}, \tag{10}$$

where the new vector field f^{μ} represents an extra force inserted to the particles moving in the geodesic lines. This will imply that a test particle move along a non-geodesic path. This is actually a general behavior of theories with matter-geometry coupling. In the low energy limit of the theory the metric takes the form

$$ds^2 = -(1 + 2\Phi(\vec{x}))dt^2 + (1 - 2\Phi(\vec{x}))d\vec{x}^2, \tag{11}$$

where Φ is the Newtonian potential. In this limit, only the (00) component of the energy-momentum tensor, namely ρ, is non-vanishing. If we assume that the function f in the Lagrangian is a smooth funciton and could be expanded as a power series in ρ, one obtains

$$f = \sum_{i=0}\alpha_i\rho^i.$$

As one can see from the field equation (5), f appears with at least two derivatives. As a result the derivative matter coupling terms could not contribute in the linear Newtonian approximation. We then recover the standard Poisson equation as

$$\vec{\nabla}^2\Phi(\vec{x}) = 4\pi G\rho(\vec{x}). \tag{12}$$

One should note that this result is in contrast to the general discussion on the form of Poisson equation in matter-geometry coupling theories,[12] since in that work the coupling is assumed to have a linear contribution to the field equations.

4. Cosmology

In this section, we consider the cosmological implications of the DMC model. Let us assume that the universe can be explained by a spatially flat FRW space-time with line element

$$ds^2 = a^2(t)(-dt^2 + dx^2 + dy^2 + dz^2), \tag{13}$$

where we have chosen the conformal time t with scale factor a. From the field equation (5), one can obtain the Friedmann and Raychaudhuri equations as

$$6\kappa^2 H^2 = a^2\rho - \alpha f'^2\,\dot{\rho}^2 + 2\kappa^2 a^2\Lambda, \tag{14}$$

and

$$\begin{aligned} -4\kappa^2\dot{H} - 2\kappa^2 H^2 = {} & pa^2 - 2\kappa^2 a^2\Lambda \\ & + 2\alpha f'(\rho + p)(f''\dot{\rho}^2 - f'\ddot{\rho}) \\ & - 4\alpha H(\rho + p)f'^2\dot{\rho} - 2\alpha f'^2\dot{\rho}^2. \end{aligned} \tag{15}$$

In this paper, we will consider a power-law case for the function f as $f = \gamma(-L_m/\rho_c)^n$. Defining the following dimensionless parameters as

$$\begin{aligned} & H = H_0 h, \quad \bar{\rho}_m = \frac{\rho_m}{\rho_c}, \quad \bar{\rho}_r = \frac{\rho_r}{\rho_c}, \quad \Omega_{\Lambda 0} = \frac{\Lambda}{3H_0^2}, \\ & r \equiv \bar{\rho}_m + \bar{\rho}_r, \quad \rho_c = 6\kappa^2 H_0^2, \qquad \beta = \frac{\gamma^2\alpha}{6\kappa^2}, \end{aligned} \tag{16}$$

where H_0 is the current value of the Hubble parameter and ρ_c is the critical density, the Friedmann equation reduces to

$$h^2 = a^2(r + \Omega_{\Lambda 0}) + \frac{1}{9}\beta r^{2(n-1)}\dot{r}^2. \tag{17}$$

We note that the matter sources on top of FRW universe is conserved. As a result, we obtain

$$\bar{\rho}_m = \frac{\Omega_{m0}}{a^3}, \qquad \bar{\rho}_r = \frac{\Omega_{r0}}{a^4}.$$

Transforming to the redshift coordinate, defined as

$$1 + z = \frac{1}{a},$$

and using the above equations, one obtains the dimensionless Hubble parameter as

$$(1+z)^2 h^2 = \frac{(1+z)^3\Omega_{m0} + \Omega_{\Lambda 0}}{1 + \beta(1+z)^{6n}\,\Omega_{m0}^{\;2n}}. \tag{18}$$

We should note that the dimensionless Hubble parameter h has a property $h(z = 0) = 1$. In the following, we will consider two different cases, corresponding to the vanishing CC and non-vanishing CC. In the case of vanishing CC, one obtains from the constraint $h(0) = 1$

$$\beta = (\Omega_{m0} - 1)\Omega_{m0}^{\;-2n}.$$

In the case of non-vanishing CC, one obtains

$$\Omega_{\Lambda 0} = 1 - \Omega_{m0} + \beta \Omega_{m0}^{2n}.$$

In order to obtain the best estimate on the model parameters, H_0, Ω_{m0}, n and β, we will obtain the best fit values of these parameters by confronting the model with observational data on the Hubble parameter in the redshift range $(0, 2)$.[13] In order to do this, we use the maximum likelihood analysis of the model based on the data on H_0. The likelihood function in the case of n independent data points can be defined as

$$L = L_0 e^{-\chi^2/2}, \tag{19}$$

where L_0 is the normalization constant, and χ^2 is

$$\chi^2 = \sum_i \left(\frac{O_i - T_i}{\sigma_i} \right)^2, \tag{20}$$

where i indicates the number of data, O_i are the observational values, T_i are the theoretical values and σ_i are the observational errors associated with the ith data. In the present model we have

$$L = L_0 \exp \left[-\frac{1}{2} \sum_i \left(\frac{O_i - H_0 T_i}{\sigma_i} \right)^2 \right]. \tag{21}$$

By maximizing the likelihood function, one can find the best fit values of the parameters. In the presence of the CC, one obtains

$$\begin{aligned} H_0 &= 65.91^{+3.23}_{-3.45}, \\ \Omega_{m0} &= 0.41^{+0.11}_{-0.08}, \\ \beta &= 0.004^{+0.003}_{-0.002}, \end{aligned} \tag{22}$$

where the $\pm$ values are the 1σ confidence intervals. In this case, the CC parameter $\Omega_{\Lambda 0}$ could be obtained as

$$\Omega_{\Lambda 0} = 0.59. \tag{23}$$

In figure (3), we have plotted the triangle diagram for this case. It should be noted that the parameter values for the Hubble parameter H_0 is in agreement with the ΛCDM model, indicating that the present model could not make the Hubble tension any better. We can see from the estimate (22) that the matter density Ω_0 is more than the ΛCDM value by 10 percent. Also, the CC abundance of the universe reduces up to 10 percent in this model. This in fact shows that the CC could be partially substituted by a simple derivative matter coupling. In fact, more complicated couplings could alleviate the contribution of the CC to the accelerated expansion of the universe even more.[11]

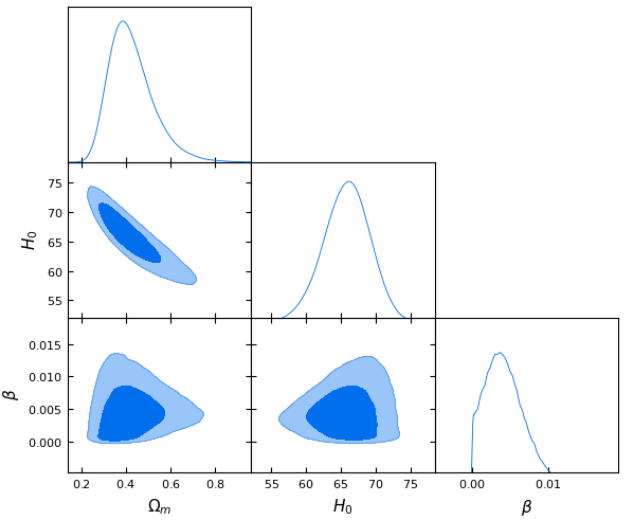


Fig. 3. The triangle diagram corresponding to 1σ intervals of the model paratmer in the case of non-vanishing CC.

Let us now consider the case of vanishing CC. In this case, all the accelerated expansion of the universe should be fully addressed by derivatives of the matter Lagrangian. The maximum likelihood analysis in this case gives

$$\begin{aligned} H_0 &= 69.32^{+3.85}_{-4.82}, \\ \Omega_{m0} &= 0.29^{+0.07}_{-0.09}, \\ n &= -0.21^{+0.11}_{-0.15}. \end{aligned} \tag{24}$$

In this case on obtains $\beta = -0.42$. In figure (4), we have plotted the triangle diagram for this case. Also, in figure (5), we have depicted the evolution of the Hubble and deceleration parameters for the best fit values of the parameters. One can see from the figures that the DMC model satisfies the observational data. In fact the Hubble parameter is a little larger than the ΛCDM value for redshifts bigger that $z \approx 0.5$. This implies that the DMC model predicts smaller universe respect to the ΛCDM model. Also, the transition from deceleration to acceleration phase is occurred later than ΛCDM model, which indicates that the universe remains in dust dominated universe more than the ΛCDM model. However, the late time observations is fully addressed in this model, thanks to the presence of derivative of matter Lagrangian.

5. Dynamical system analysis

In order to see whether the DMC model could explain the cosmic history of the universe well or not, let us consider the dynamical system analysis of the model.

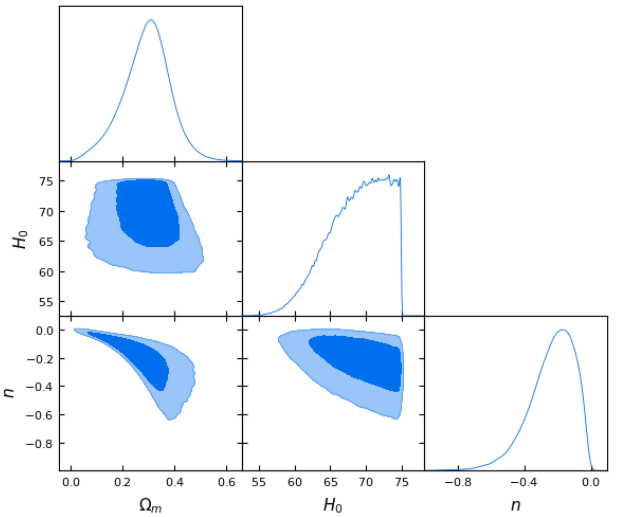


Fig. 4. The triangle diagram corresponding to 1σ intervals of the model paratmer in the case of vanishing CC.

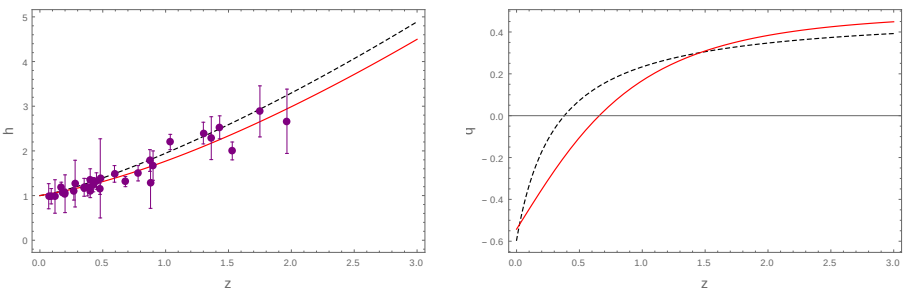


Fig. 5. The evolution of the Hubble (left) and deceleration (right) parameters as a function of redshift z for the best fit values of model parameters. The red solid line indicates the ΛCDM curve. Also, the error bars are the observational values together with their errors.[13]

We define the dynamical variables as

$$\Omega_m = \frac{\bar{\Omega}_m a^2}{h^2}, \quad \Omega_r = \frac{\bar{\Omega}_r a^2}{h^2}, \quad \Omega_h = \frac{h^2}{a^2}, \quad \Omega_\Lambda = \frac{\Omega_{\Lambda 0} a^2}{h^2}. \tag{25}$$

In the case of vanishing CC, one obtains from the Friedmann equation

$$\Omega_h = \left[\sqrt{\frac{1-\Omega_m-\Omega_r}{-\beta}} \frac{3}{(\Omega_m+\Omega_r)^{n-1}(3\Omega_m+4\Omega_r)} \right]^{\frac{1}{n}}. \tag{26}$$

In this case, we have two dynamical degrees of freedom. Also, in the case of non-vanishing CC, one obtains

$$\Omega_\Lambda = 1 - \Omega_m - \Omega_r + \frac{1}{9}\beta\,\Omega_h^{2n}(\Omega_m + \Omega_r)^{2(n-1)}(3\Omega_m + 4\Omega_r)^2, \tag{27}$$

indicating that the model has three dynamical degrees of freedom. In figure (6), we have plotted the evolution of the dust and radiation abundances and dark energy abundance defined as $\Omega_{dark} \equiv 1 - \Omega_m - \Omega_r$, as a function of $N \equiv \ln a$. The dashed lines represent the behavior of the abundances in ΛCDM model, the dotted lines indicating the DMC model with CC and the solid lines indicate the DMC model with vanishing CC. One can see from this figure, that both cases could fully describe the history of the universe as in the ΛCDM model.

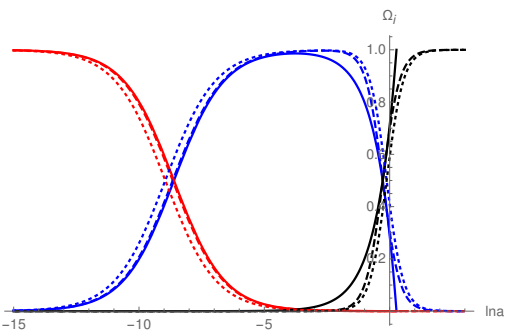


Fig. 6. The evolution of the dust (blue), radiation (red) and dark energy (black) abundances in ΛCDM model (dahsed), DMC model with CC (dotted) and DMC model with vanishing CC (solid).

In the case of vanishing CC, we have two fixed points corresponding to the radiation and dust dominated universe. In table (1), we have summarized these two fixed points. The stream plot of the evolution of the dynamical system could be seen in figure (7). In this case, there is no stable fixed point corresponds to the de Sitter acceleration of the universe. However, we have a stable manifold passing through the origin. As we get closer to the origin, the effective parameter of state of the model takes negative and larger values $\omega_{eff} < 0$ and we have more acceleration. In figure (7), we have plotted the $\omega_{eff} = -1$ curve as a solid line, indicating the

Table 1. The fixed points together with their properties for the DMC model without CC.

	(Ω_m, Ω_r)	Type	ω_{eff}	Status
R	$(0, 1)$	Radiation	$1/3$	unstable
M	$(1, 0)$	matter	0	saddle

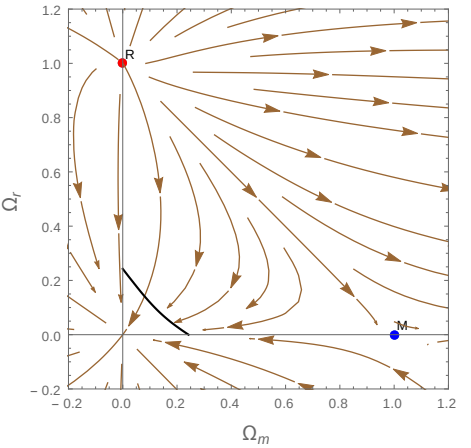


Fig. 7. The phase portrait of the dynamical system of the DMC model in the case of vanishing CC. The fixed points can also be seen in the figure. The black solid line indicates the line of de Sitter expansion.

de Sitter expansion. In figure (6), one can see that the dark energy abundance blows up at $N \approx 0$, indicating the aforementioned property of the model in this case. As a result, in this case, the universe could begin from an unstable radiation dominated fixed point, fall into the saddle dust dominated point and end at some accelerated expanding phase.

In the case of non-vanishing CC, we have three fixed points as we have summarized in table (2). In this case, we have all the standard fixed points of the ΛCDM model. However, as one can see from the table, the stable de Sitter fixed point is in fact a fixed line, passing through the origin. We have plotted the $\Omega_h = 0$ plane of the phase space in figure (8). In this case, the history of the universe is the same as the standard ΛCDM model. This can also be seen in figure (6).

Table 2. The fixed points together with their properties for the DMC model with CC.

	$(\Omega_m, \Omega_r, \Omega_h)$	Type	ω_{eff}	Status
R	$(0, 1, 0)$	Radiation	$1/3$	unstable
M	$(1, 0, 0)$	matter	0	saddle
Λ	$(0, 0, x)$	de Sitter	-1	stable

6. Conclusion

In this paper, we have considered a new possibility in the class of matter-geometry coupled theories namely, adding derivative interactions of the matter Lagrangian. We have used an arbitrary function f of the matter Lagrangian L_m which is a scalar

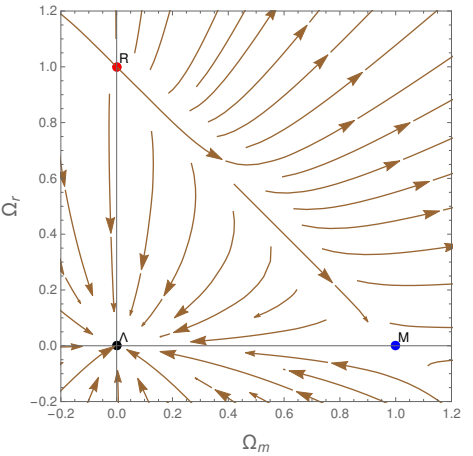


Fig. 8. The phase portrait of the dynamical system of the DMC model in the case of non-vanishing CC. The fixed points can also be seen in the figure.

field in nature, and add a canonical kinetic term of f to the action. In fact more general Galileon interactions could also be added to the theory which makes the late time behavior more reliable.

The cosmological implications of the theory in the case of vanishing and non-vanishing CC was studied in detail and we have found that CC could in part or fully be substituted by derivatives of matter Lagrangian. In the case of vanishing CC however, the cosmic history of the universe does not have a stable de Sitter fixed point and we need some higher order Galileon invariants to make the theory reliable. This was done in.[11]

In summary, we have shown in this paper that derivatives of matter Lagrangian could be considered more seriously as a reliable modified gravity theory for late times.

References

1. A. G. Riess et al., Astron. J. 116 (1998) 1009, arXiv:astro-ph/9805201; S. Perlmutter et al., Astrophys. J. 517(1999) 565, arXiv:astro-ph/9812133.
2. S. Weinberg, Rev. Mod. Phys. 61 (1989) 1; L. Lombriser, arXiv:1901.08588 [gr-qc].
3. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, J. Silk, arXiv:2103.01183 [astro-ph.CO].
4. R. C. Nunes and S. Vagnozzi, Mon. Not. Roy. Astron. Soc. 505 (2021) 5427.
5. I. Quiros, Int. J. Mod. Phys. D 28 (2019) 1930012; T. Kobayashi, Rept. Prog. Phys. 82 (2019) 086901; C. Deffayet, D. A Steer, Class. Quantum Grav. 30 (2013) 214006.
6. R. Kimura, A. Naruko, D. Yoshida, JCAP 01 (2017) 002; L. Heisenberg, JCAP 05 (2014) 015; L. Heisenberg, R. Kase, S. Tsujikawa, Phys. Lett. B 760 (2016) 617; Z. Haghani, T. Harko, H. R. Sepangi, S. Shahidi, Eur. Phys. J. C 77 (2017) 137; S. Shahidi, EPJC 79 (2019) 448; S. Shahidi, Phys. Rev. D 98 (2018) 084004; A. Amado, Z. Haghani, A. Mohammadi, S. Shahidi, Phys. Lett. B 772 (2017) 141.

7. T. Clifton, P. G. Ferreira, A. Padilla, C. Skordis, Physics Reports 513 (2012) 1; K. Koyama, Rept. Prog. Phys. 79 (2016) 046902; S. Nojiri, S. Odintsov, Int. J. Geom. Methods in Mod. Phys. 04 (2007) 115.
8. C. de Rham, Living Rev. Relativity 17 (2014), 7; K. Hinterbichler, Rev. Mod. Phys. 84 (2012) 671.
9. M. Roshan and f. Shojai, Phys. Rev. D 94 (2016) 044002; , J. D. Barrow and C. Board, Phys. Rev. D 96 (2017) 123517; O. Akarsu, N. Katirci and S. Kumar, Phys. Rev. D 97 (2018) 024011.
10. T. Harko, F. S. N. Lobo, S. Nojiri and S. D. Odintsov, Phys. Rev. D 84, 024020 (2011); T. Harko and F. S. N. Lobo, Eur. Phys. J. C 70, 373 (2010); S. D. Odintsov and D. Saez-Gomez, Phys. Lett. B 725 (2013) 437; O. Bertolami, C.G. Boehmer, T. Harko, f.S.N. Lobo, Phys. Rev. D 75 (2007) 104016; Z. Haghani, T. Harko, f.S.N. Lobo, H.R. Sepangi, S. Shahidi, Phys. Rev. D 88 (2013) 044023.
11. Z. Haghani, S. Shahidi, PDU (2020) 100683; Z. Haghani, S. Shahidi, EPJ plus 135 (2020) 509.
12. B. T. P. Sotiriou and V. Faraoni, Rev. Mod. Phys. 82 (2010) 451.
13. H. Boumaza and K. Nouicer, Phys. Rev. D 100 (2019) 124047.

On the semiclassical and quantum picture of the Bianchi I polymer dynamics

E. Giovannetti*

Department of Physics, La Sapienza University of Rome, Rome, Italy
**Speaker. E-mail: eleonora.giovannetti@uniroma1.it*

G. Montani

Department of Physics, La Sapienza University of Rome, Rome, Italy
Fusion and Nuclear Safety Department, ENEA, Frascati (RM), Italy

S. Schiattarella

Department of Physics, La Sapienza University of Rome, Rome, Italy

We analyze the Bianchi I cosmology in the presence of a massless scalar field and describe its dynamics via a semiclassical and quantum polymer approach. We investigate the morphology of the emerging Big Bounce by adopting three different sets of configurational variables: the natural Ashtekar connections, the Universe volume plus two anisotropy coordinates and a set of anisotropic volume-like coordinates (the latter two sets of variables would coincide in the case of an isotropic Universe). In the semiclassical analysis we demonstrate that the Big Bounce emerges in the dynamics for all the three sets of variables. Moreover, when the Universe volume itself is considered as a configurational variable, we have derived the polymer-modified Friedmann equation and demonstrated that the Big Bounce has a universal nature, *i.e.* the total critical energy density has a maximum value fixed by fundamental constants and the Immirzi parameter only. From a pure quantum point of view, we investigate the Bianchi I dynamics only in terms of the Ashtekar connections. In particular, we apply the Arnowitt–Deser–Misner (ADM) reduction of the variational principle and then we quantize the system. We study the resulting Schrödinger dynamics, stressing that the wave packet peak behavior over time singles out common features with the semiclassical trajectories.

Keywords: Polymer Quantum Mechanics; Polymer Cosmology; Bianchi I Universe; Ashtekar variables.

1. INTRODUCTION

The emergence of a Big Bounce in the quantum dynamics of the isotropic Universe is one of the most relevant result achieved in Loop Quantum Cosmology (LQC) as a phenomenological implication of Loop Quantum Gravity (LQG)[1,2] when applied to the cosmological sector.[3–9] Although, many criticisms have been expressed towards the LQC framework, especially regarding the lack of a real quantum description of the Big Bounce properties, the unclear equivalence between the so-called improved scheme with the original LQC formulation in the Ashtekar variables and the process through which the symmetry reduction is performed in order to properly derive LQC as the cosmological sector of LQG.[10] In this context, Polymer Quantum Mechanics (PQM)[11,12] represents a convenient mathematical instrument through which deeply

analyze bouncing cosmologies, since it is able to reproduce LQC results at least at an effective level and without entering the issues of LQG and LQC. In this work we apply PQM to the Bianchi I model in the presence of a massless scalar field and we explore both its semiclassical and quantum dynamics in different sets of configurational variables,[12] in particular the natural Ashtekar connections and two different sets of volume-like variables (the anisotropic volume-like coordinates and the Universe volume itself plus two anisotropies), following the LQC formulation in.[13,14]

The paper is structured as follows. In Sec. 2 the theory of PQM is introduced, while the original part of the paper is developed in Secs. 3–5. In particular, in Secs. 3 and 4 it is solved the semiclassical polymer dynamics of the Bianchi I model in the Ashtekar variables and in the two sets of volume-like variables respectively, then in Sec. 5 a full quantum treatment in the Ashtekar variables is developed using a Scrödinger-like formalism. Finally, in Sec. 6 some concluding remarks are commented.

2. POLYMER QUANTUM MECHANICS

PQM is a non-equivalent representation of quantum mechanics with respect to the standard Schrödinger one[12] and is based on the assumption that one or more variables of the phase space are discretized.

Firstly, we consider a set of abstract kets $|\mu\rangle$ with $\mu \in \mathbb{R}$ with the inner product $\langle\mu|\,\nu\rangle = \delta_{\mu\nu}$, where $\delta_{\mu\nu}$ is a Kronecker delta. This procedure defines the non-separable Hilbert space H_{poly} where two fundamental operators can be introduced: the *label* operator $\hat{\epsilon}$, whose action on the kets is given by $\hat{\epsilon}\,|\mu\rangle = \mu\,|\mu\rangle$, and the *shift* operator $\hat{s}(\lambda)$ ($\lambda \in \mathbb{R}$), where $\hat{s}(\lambda) = |\mu + \lambda\rangle$. The action of $\hat{s}(\lambda)$ is discontinuous since the kets are orthonormal $\forall\lambda$. In this sense, the set of kets indexed by μ is said to be discrete.

Now, we consider a one-dimensional system identified by the coordinates (q, p) and we suppose that the coordinate q has a discrete character. In the p-polarization the wave function writes as ($\hbar = 1$) $\psi_\mu(p) = \langle p|\mu\rangle = e^{ip\mu}$, the action of the operator $\hat{q}$ is differential and in particular coincident with the label operator. Its eigenvalues are precisely the μ parameters and constitute a discrete set. On the other hand, in this polarization the shift operator acts in a multiplicative way as

$$\hat{s}(\lambda)\psi_\mu(p) = e^{i\lambda p}e^{i\mu p} = \psi_{(\mu+\lambda)}(p)\,. \tag{1}$$

As already mentioned, its action is discontinuous; therefore, we must conclude that the operator $\hat{p}$ does not exist since it is not possible to obtain a discontinuous operator from the exponentiation of a Hermitian one. Similar conclusions are reached by analyzing the structure of the q-polarization. Indeed, defining the variable q as discrete implies the non-existence of the operator $\hat{p}$, which must be regularized in order to deal with a well-defined dynamics. In order to overcome this issue, we introduce a regular graph $\gamma_{\mu_0} = \{q \in \mathbb{R}\,|\,q = n\mu_0, \forall n \in \mathbb{Z}\}$, *i.e.* a numerable set of

equidistant points whose spacing is given by the scale μ_0, and we restrict the action of the shift operator $e^{i\lambda p}$ by imposing $\lambda = n\mu_0$ in order to remain in the lattice. Its action is well-defined, so we use it to approximate any function of p in the following way:

$$p \approx \frac{1}{\mu_0} \sin(\mu_0 p) = \frac{1}{2i\mu_0} \left(e^{i\mu_0 p} - e^{-i\mu_0 p} \right) . \tag{2}$$

We notice that this approximation is good for $\mu_0 p \ll 1$. Under this hypothesis, the action of the regularized $\hat{p}$ operator is

$$\hat{p}_{\mu_0} |\mu_n\rangle = -\frac{i}{2\mu_0} (|\mu_{n+1}\rangle - |\mu_{n-1}\rangle) . \tag{3}$$

In the original part of the paper we will apply this picture to the configurational variables describing the Bianchi I cosmology. The interest we have in using a representation of quantum mechanics characterized by a discrete configurational variable relies in investigating cut-off effects on geometrical quantities and hence on the cosmological dynamics.

3. SEMICLASSICAL POLYMER DYNAMICS OF THE BIANCHI I MODEL IN THE ASHTEKAR VARIABLES

In this section we firstly investigate the semiclassical polymer dynamics of the Bianchi I model in terms of the Ashtekar variables. The phase space is six-dimensional and it is expressed through the canonical couple (c_i, p_j), which link with the scale factors a_i and their velocities is

$$p_i = |\epsilon_{ijk} a_j a_k| \text{sign}(a_i) , \qquad c_i = \gamma \dot{a}_i , \tag{4}$$

with $i = 1, 2, 3$ and $\{c_i, p_j\} = \kappa\gamma\delta_{ij}$. In order to implement the polymer paradigm, we proceed by imposing the polymer substitution for the Ashtekar connections

$$c_i \to \frac{1}{\mu_i} sin(\mu_i c_i) , \tag{5}$$

so the polymer Hamiltonian constraint with a massless scalar field takes the form

$$\mathcal{H}_{poly} = -\frac{1}{\kappa\gamma^2 V} \sum_{i \neq j} \frac{\sin(\mu_i c_i) p_i \sin(\mu_j c_j) p_j}{\mu_i \mu_j} + \frac{p_\phi^2}{2V} = 0 , \tag{6}$$

where $i, j = 1, 2, 3$ and $V = \sqrt{p_1 p_2 p_3}$. This way, we have chosen to define the variables p_i on three independent polymer lattices (with spacing μ_i respectively) due to their geometrical character. We also choose ϕ as the internal time, so N is fixed by the gauge $N = \frac{\sqrt{p_1 p_2 p_3}}{p_\phi}$, so that the equations of motion in the polymer representation are:

$$\begin{cases} \dfrac{dp_i}{d\phi} = -\dfrac{p_i \cos(\mu_i c_i)}{\gamma p_\phi} \left[\dfrac{p_j}{\mu_j} \sin(\mu_j c_j) + \dfrac{p_k}{\mu_k} \sin(\mu_k c_k) \right] \\ \dfrac{dc_i}{d\phi} = \dfrac{\sin(\mu_i c_i)}{\gamma \mu_i p_\phi} \left[\dfrac{p_j}{\mu_j} \sin(\mu_j c_j) + \dfrac{p_k}{\mu_k} \sin(\mu_k c_k) \right] \end{cases} \tag{7}$$

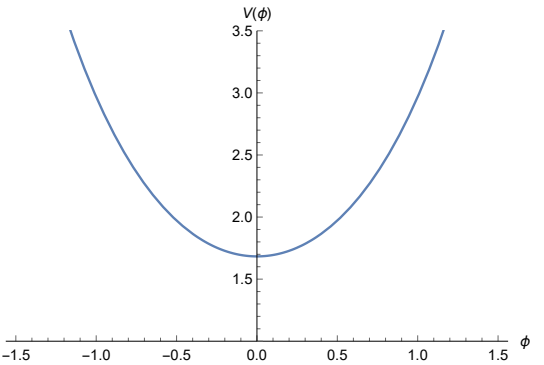


Fig. 1. Polymer trajectory of the Universe volume $V = \sqrt{p_1 p_2 p_3}$ in function of time ϕ: the Big Bounce replaces the initial singularity of the Bianchi I model.

for $i, j, k = 1, 2, 3$, $i \neq j \neq k$. By identifying proper constants of motion, the equations in (7) can be decoupled and the system can be analytically solved. In particular, in Fig. 1 it is shown the behavior of the Universe volume $V(\phi) = \sqrt{p_1(\phi)p_2(\phi)p_3(\phi)}$ as function of ϕ. The resulting trajectory highlights that a semi-classical Big Bounce replaces the classical Big Bang in the polymer framework.

4. SEMICLASSICAL POLYMER DYNAMICS OF THE BIANCHI I MODEL IN THE VOLUME-LIKE VARIABLES

In this section we study the dynamics of the Bianchi I Universe for a new choice of variables. In particular, the anisotropic character of the Bianchi I model leads to the possibility of considering two different sets of volume-like variables, that coincide in the case of the isotropic model.

4.1. *Analysis in the anisotropic volume-like variables:* (V_1, V_2, V_3)

Firstly, we consider three equivalent generalized coordinates which correspond to the proper Universe volume in the isotropic limit only:[14]

$$V_i = \mathrm{sign}(p_i)|p_i|^{\frac{3}{2}}\,, \qquad \eta_i = \frac{2c_i}{3\sqrt{|p_i|}}\,, \tag{8}$$

where η_i for $i = 1, 2, 3$ are the conjugate momenta and the new symplectic structure for the system is characterized by the Poisson brackets $\{\eta_i, V_j\} = \kappa\gamma\delta_{ij}$.

After that the canonical transformation has been performed, we obtain the polymer Hamiltonian constraint in the new variables by using the polymer substitution for the momenta η_i, since the variables V_i live on the corresponding polymer lattices. It reads as

$$\mathcal{H}_{poly} = -\frac{9}{4\kappa\gamma^2 V}\sum_{i\neq j}\frac{V_i \sin(\mu_i\eta_i)V_j \sin(\mu_j\eta_j)}{\mu_i\mu_j} + \frac{p_\phi^2}{2V} = 0\,, \tag{9}$$

where $i,j=1,2,3$ and $V=(V_1V_2V_3)^{\frac{1}{3}}$. Analogously, we impose $N=\frac{V}{p_\phi}$ due to the choice of ϕ as relational time, so the Hamilton equations describing the dynamics are

$$\begin{cases} \dfrac{dV_i}{d\phi} = -\dfrac{9V_i\cos(\mu_i\eta_i)}{4\gamma p_\phi}\left[\dfrac{V_j}{\mu_j}\sin(\mu_j\eta_j)+\dfrac{V_k}{\mu_k}\sin(\mu_k\eta_k)\right] \\ \dfrac{d\eta_i}{d\phi} = \dfrac{9\sin(\mu_i\eta_i)}{4\gamma\mu_1 p_\phi}\left[\dfrac{V_j}{\mu_j}\sin(\mu_j\eta_j)+\dfrac{V_k}{\mu_k}\sin(\mu_k\eta_k)\right] \end{cases} \tag{10}$$

for $i\neq j\neq k$. In analogy with the previous treatment, we can identify proper constants of motion that decouple the system (10) along the three directions.

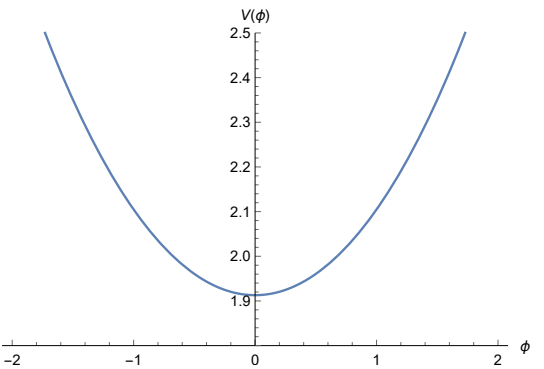


Fig. 2. Semiclassical polymer trajectory of the Universe volume $V=(V_1V_2V_3)^{1/3}$ as function of ϕ.

By taking general initial conditions according to (9), analytical solutions for the anisotropic volume coordinates can be gained. In particular, the Universe volume as function of ϕ reads as $V(\phi)=(V_1(\phi)V_2(\phi)V_3(\phi))^{1/3}$ and its behavior is shown in Fig. 2. The Big Bounce clearly appears as a polymer regularization effect instead of the classical Big Bang.

4.2. *Analysis in the volume variables:* (v,λ_1,λ_2)

In the new set of volume variables ($\lambda_i=\mathrm{sign}(p_i)\sqrt{|p_i|}$, $v=\lambda_1\lambda_2\lambda_3$) the Universe volume itself v is considered and $\eta_{1,2}=2\sqrt{p_{1,2}}c_{1,2}$, $\eta_3=2\sqrt{\frac{p_3}{p_1p_2}}c_3$ are the conjugate momenta. The Poisson brackets are conserved ($\{\eta_i,\lambda_j\}=\kappa\gamma\delta_{ij}$ for $i\neq j$, $i,j=1,2$ and $\{\eta_3,v\}=\kappa\gamma$) and the semiclassical polymer Hamiltonian takes the form

$$\mathcal{H}_{poly}=-\frac{1}{4\kappa\gamma^2V}\Big(\sum_{i=1,2}\frac{\lambda_i\sin(\mu_i\eta_i)v\sin(\mu_3\eta_3)}{\mu_i\mu_3}+\frac{\lambda_1\sin(\mu_1\eta_1)\lambda_2\sin(\mu_2\eta_2)}{\mu_1\mu_2}\Big)+\frac{p_\phi^2}{2V}=0, \tag{11}$$

where the canonical transformation has been performed before the implementation of the semiclassical polymer paradigm on the variables η_i, as in the previous subsection. Analogously, we derive the Hamilton equations for the couple of variables (v, η_3)

$$\begin{cases} \dfrac{dv}{d\phi} = -\dfrac{v\cos(\mu_3\eta_3)}{4\gamma p_\phi}\left[\dfrac{\lambda_1}{\mu_1}\sin(\mu_1\eta_1) + \dfrac{\lambda_2}{\mu_2}\sin(\mu_2\eta_2)\right] \\ \dfrac{d\eta_3}{d\phi} = \dfrac{\sin(\mu_3\eta_3)}{4\gamma\mu_3 p_\phi}\left[\dfrac{\lambda_1}{\mu_1}\sin(\mu_1\eta_1) + \dfrac{\lambda_2}{\mu_2}\sin(\mu_2\eta_2)\right] \end{cases} \tag{12}$$

and also for the conjugate variables (λ_1, η_1), (λ_2, η_2)

$$\begin{cases} \dfrac{d\lambda_i}{d\phi} = -\dfrac{\lambda_i\cos(\mu_i\eta_i)}{4\gamma p_\phi}\left[\dfrac{v}{\mu_3}\sin(\mu_3\eta_3) + \dfrac{\lambda_j}{\mu_j}\sin(\mu_j\eta_j)\right] \\ \dfrac{d\eta_i}{d\phi} = \dfrac{\sin(\mu_i\eta_i)}{4\gamma\mu_i p_\phi}\left[\dfrac{v}{\mu_3}\sin(\mu_3\eta_3) + \dfrac{\lambda_j}{\mu_j}\sin(\mu_j\eta_j)\right] \end{cases} \tag{13}$$

where we have used $N = \frac{V}{p_\phi}$ in order to derive the dynamics of the model in function of the relational time ϕ.

Once fixed the initial conditions on the variables (λ_1, η_1), (λ_2, η_2), (v, η_3) according to (11), we can solve this system analytically since the 3D-motion is decoupled in three one-dimensional trajectories, thanks to the use of constants of motion analogous to those ones of the previous cases. In particular, the analytical solution for the Universe volume $v(\phi)$ clearly resembles a bouncing behavior, as shown in Fig. 3. Also, by choosing a particular shape for these constants of motion, we find the following convenient form for the polymer-modified Friedmann equation of the Bianchi I model in the proper volume variables:

$$H^2 = \frac{\kappa}{54}\frac{p_\phi^2 + \bar{\mathcal{K}}^2}{v^2}\left[1 - \frac{4\kappa\gamma^2\mu^2}{3}\left(\frac{p_\phi^2 + \bar{\mathcal{K}}^2}{2v^2}\right)\right], \quad \bar{\mathcal{K}} = \frac{\mathcal{K}}{\sqrt{2\kappa\gamma^2}}, \tag{14}$$

where we have fixed $\mu_1 = \mu_2 = \mu_3 = \mu$ without loss of generality. In (14), the additional term $\bar{\mathcal{K}}^2/2v^2$ reasonably mimics the energy-like contribution of the

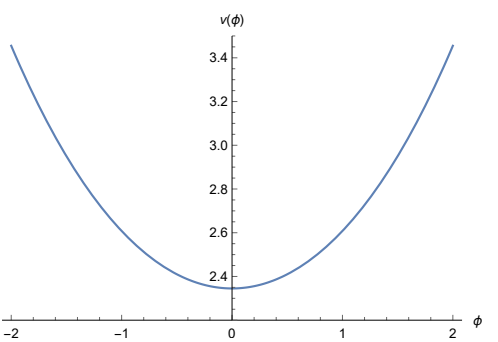


Fig. 3. Semiclassical polymer trajectory of the Universe volume $v(\phi)$.

anisotropies to the total energy density ($\mathcal{K}$ is an arbitrary constant) and the total critical energy density results to be

$$\rho_{crit}^{tot} = \frac{3}{4\kappa\gamma^2\mu^2}\,, \tag{15}$$

so it has universal features since it is independent from the initial conditions on the motion. By concluding, thanks to equation (14) we have inferred the physical properties of the critical point, showing that taking the Universe volume itself as a configurational variable makes the Big Bounce acquire universal physical properties.

5. QUANTUM ANALYSIS

The purpose of this section is studying the Bianchi I model in the Ashtekar variables at a quantum level by applying the Dirac quantization. Firstly, we perform an ADM reduction of the variational principle at a semiclassical level in order to deal with a Schrödinger-like formalism. More specifically, after choosing the scalar field ϕ as the temporal parameter, we derive the ADM-Hamiltonian by solving the scalar constraint (6) with respect to the momentum associated to the scalar field:

$$p_\phi \equiv \mathcal{H}_{ADM} = \sqrt{\Theta}\,, \quad \Theta = \frac{2}{\kappa\gamma^2}\sum_{i\neq j}\frac{\sin(\mu_i c_i)p_i\sin(\mu_j c_j)p_j}{\mu_i\mu_j}\,, \tag{16}$$

where $i,j = 1,2,3$. We choose the positive root in order to guarantee the positive character of the lapse function (see Sec. 3). Now we promote the ADM-Hamiltonian to a quantum operator, obtaining the following Schrödinger-like equation:

$$-i\partial_\phi\Psi = \sqrt{\hat{\Theta}}\Psi\,, \quad \sqrt{\hat{\Theta}} = \Big[\frac{2}{\kappa^2\gamma^2}\Big(\partial_{x_1}\partial_{x_2} + \partial_{x_1}\partial_{x_3} + \partial_{x_2}\partial_{x_3}\Big)\Big]^{1/2}\,, \tag{17}$$

where we have used the substitution $x_i = \ln\left[\tan\left(\frac{\mu c_i}{2}\right)\right] + \bar{x}_i$. The associated probability density is $\mathcal{P}(\vec{x},\phi) = \Psi^*(\vec{x},\phi)\Psi(\vec{x},\phi)$, where

$$\begin{aligned}\Psi(\vec{x},\phi) = \int_{-\infty}^{\infty} dk_1\,dk_2\,dk_3\prod_{i=1}^{3}\exp\left[-\frac{(k_i-\mu_{k_i})^2}{2\sigma_{k_i}^2}\right] \\ \times\, e^{i(k_1x_1+k_2x_2+k_3x_3+\sqrt{2|k_1k_2+k_1k_3+k_2k_3|}\phi)}\,.\end{aligned} \tag{18}$$

Now we can analyze the quantum dynamics of (18) by following the peak of the probability density $\mathcal{P}$ over time, in order to verify the consistency between the information carried by the quantum wave packet of the Bianchi I model and the semiclassical solutions provided in Sec. 3 when the Ashtekar variables are considered. In Fig. 4 some different sections of the probability density $\mathcal{P}$ at different values of ϕ are shown. As we can see, their spreading behavior over time is evident, as well as their gaussian-like shape. Also, in Fig. 5 the position of the peaks of $\mathcal{P}$ are represented by the red dots and have also been fitted by means of a linear interpolation. The resulting fitting functions are represented by the red dashed straight lines, while the semiclassical trajectories by the continuous straight ones. In particular,

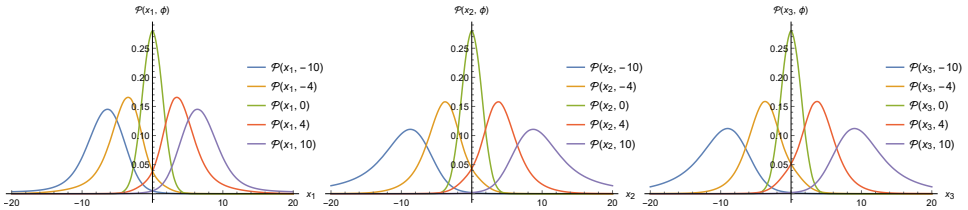


Fig. 4. The normalized sections $\mathcal{P}(x_i, \phi)$ are shown in sequence for $i = 1, 2, 3$ respectively at different times (here $\mu_{k_1} = \mu_{k_2} = \mu_{k_3} = 0, \sigma_{k_1} = \sigma_{k_2} = \sigma_{k_3} = 1/2$). Their spreading behavior over time is evident together with the gaussian-like shape.

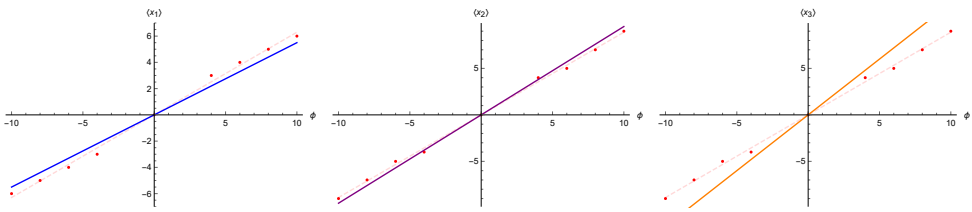


Fig. 5. The three pictures show the position of the peaks of $\mathcal{P}(x_i, \phi)$ for $i = 1, 2, 3$ respectively in function of time ϕ (red dots). The resulting fitting functions (red dashed straight lines) overlap the semiclassical trajectories (continuous lines) with a confidence level of three standard deviations.

the slope of the fitting straight lines is consistent with the semiclassical one with a confidence level of three standard deviations for all the three coordinates. Therefore, we conclude that there is a good correspondence between the quantum behavior of the Universe wave function and the solutions of the semiclassical dynamics.

6. CONCLUDING REMARKS

In this paper we analyzed the polymer semiclassical and quantum dynamics of the Bianchi I model with a massless scalar field in the Ashtekar variables and in two sets of volume-like ones. In the polymer semiclassical analysis we obtained a bouncing behavior in all the three sets of variables. Also, a polymer-modified Friedmann equation of the Bianchi I model has been derived when the Universe volume itself is considered as a configurational variable. This way, the proper expression of the total critical energy density has been provided, giving a complete picture of the Big Bounce and showing its universal features, *i.e.* its dependence from fundamental constants and parameters only. Finally, by performing an ADM reduction of the variational problem, we passed to a Schrödinger-like formalism and then we implemented the canonical quantization in the Ashtekar variables. In particular, the study of the probability density behavior has outlined a good correspondence between the dynamics of the quantum Universe wave packet and the corresponding semiclassical trajectories. In conclusion, due to the role of the Bianchi I model in constructing the general cosmological solution, this study acquires a great relevance in order to understand the properties of the primordial Universe.

References

1. C. Rovelli and F. Vidotto, *Covariant Loop Quantum Gravity: an Elementary Introduction to Quantum Gravity and Spinfoam Theory* (Cambridge University Press, 2014).
2. F. Cianfrani, O. M. Lecian, M. Lulli and G. Montani, *Canonical Quantum Gravity* (World Scientific, 2014).
3. A. Ashtekar, M. Bojowald, J. Lewandowski *et al.*, Mathematical Structure of Loop Quantum Cosmology, *Advances in Theoretical and Mathematical Physics* **7** (2003).
4. A. Ashtekar, Gravity and the Quantum, *New Journal of Physics* **7** (2005).
5. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the big bang: An analytical and numerical investigation, *Physical Review D* **73** (Jun 2006).
6. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the big bang: Improved dynamics, *Physical Review D* **74** (Oct 2006).
7. A. Ashtekar, A. Corichi and P. Singh, Robustness of key features of Loop Quantum Cosmology, *Physical Review D* **77** (Jan 2008).
8. A. Ashtekar and P. Singh, Loop Quantum Cosmology: a Status Report, *Classical and Quantum Gravity* **28** (2011).
9. M. Bojowald, Loop Quantum Cosmology, *Living Reviews in Relativity* **11** (2008).
10. M. Bojowald, Critical evaluation of common claims in loop quantum cosmology, *Universe* **6**, p. 36 (Feb 2020).
11. A. Ashtekar, Polymer geometry at Planck scale and quantum Einstein equations, *International Journal of Modern Physics D* **05**, p. 629–648 (Dec 1996).
12. J. A. Z. Alejandro Corichi, Tatjana Vukasinac, Polymer Quantum Mechanics and its Continuum Limit, *Phys.Rev.* (2007).
13. A. Ashtekar and E. Wilson-Ewing, Loop Quantum Cosmology of Bianchi type I models, *Physical Review D* **79** (2009).
14. Ł. Szulc, Loop Quantum Cosmology of Diagonal Bianchi type I model: Simplifications and Scaling Problems, *Physical Review D* **78** (2008).

Quantum corrections to the Bianchi II transition under local rotational invariance

Sara F. Uria* and David Brizuela†

Fisika Saila, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Barrio Sarriena s/n, 48940 Leioa, Spain

**E-mail: sara.fernandezu@ehu.eus*

†E-mail: david.brizuela@ehu.eus

Ana Alonso-Serrano

Max-Planck-Institut für Gravitationsphysik, Am Mühlenberg 1, 14476 Potsdam, Germany

E-mail: ana.alonso.serrano@aei.mpg.de

A quantum state in a Bianchi II model is studied as it approaches the cosmological singularity, by means of the evolution of its moments. Classically this system presents a transition between two Bianchi I models. This phenomenon is described by a very specific and well-known transition law, which is derived based on the conservation of certain physical quantities. In the quantum theory fluctuations, as well as higher-order quantum moments, of the different variables arise. Consequently, these constants of motion are modified and hence also the transition rule. We focus on the so-called locally rotationally symmetric and vacuum case, as a first step towards a more complete study. Indeed, the future goal of this research line is to generalize this analysis to the Bianchi IX spacetime, which can be seen as a succession of Bianchi II models. Ultimately, these results will shed light on the role played by quantum effects in the BKL conjecture.

Keywords: Bianchi II; local rotational symmetry (LRS); quantum Bianchi spacetime; Bianchi transition.

1. Introduction

Close to a spacelike singularity, it is conjectured that the classical dynamics of any universe follows a chaotic behavior, according to Belinski-Khalatnikov-Lifshitz (BKL).[1] In this scenario, the dynamics of each point is decoupled from the rest and can be described by a Bianchi IX spacetime, which can be understood as a succession of Bianchi II models.[1] This model is much easier to analyze than the full Bianchi IX dynamics, and therefore provides a good first step to study the full BKL scenario. More precisely, as it was described in Misner's seminal work,[2] in the Bianchi II model the system undergoes a single transition between two Bianchi I models. However, close to the singularity quantum effects are expected to become relevant, and thus, in the present work we will analyze how they modify this classical transition.

2. Classical Model

A Bianchi II model is a type of spatially homogeneous but anisotropic four-dimensional spacetime. If we follow the usual 3+1 decomposition in order to describe

the model canonically, the vacuum Hamiltonian constraint of this system is found to be

$$\mathcal{C} = \frac{1}{2} e^{-3\alpha} \left(-p_\alpha^2 + p_-^2 + p_+^2 \right) + e^{\alpha - 8\beta_+} = 0, \tag{1}$$

where α describes the spatial volume and $\beta_\pm$ are the shape parameters that encode the spatial anisotropy of the spacetime. These are the so-called Misner variables and p_α and $p_\pm$ are their corresponding conjugate momenta. If α is chosen as the internal time, the physical Hamiltonian of the system takes the subsequent form:

$$H := \ -p_\alpha = \left(p_+^2 + p_-^2 + 2e^{4\alpha - 8\beta_+} \right)^{1/2}. \tag{2}$$

This is the most general vacuum description but, in order to further simplify the problem, we will remove one of the shape parameters by imposing $\beta_- = 0 = p_-$, which is known as the *locally rotationally symmetric* (LRS) case. Consequently, the Hamiltonian (2) is reduced to:

$$H = \left(p^2 + 2e^{4\alpha - 8\beta} \right)^{1/2}, \tag{3}$$

where, for compactness, $\beta := \beta_+$ and $p := p_+$ have been defined.

This model presents a singularity, even in the particular LRS case, precisely when $\alpha \to -\infty$. Hence, we will be interested in studying the evolution towards this limit. For this purpose, the first step is to obtain the equations of motion:

$$\dot{\beta} = \frac{p}{H}, \tag{4}$$

$$\dot{p} = \frac{8}{H} e^{4\alpha - 8\beta}, \tag{5}$$

where the dot represents the derivative with respect to the internal time α. From (4)-(5) we note that, when the exponential term $e^{4\alpha - 8\beta}$ is negligible, the equations of motion can easily be solved. Under such an assumption, p would be a constant of motion and, in this sense, the system would follow a free dynamics, whereas the shape-parameter β would be a linear function of α:

$$\beta = \text{sign}(p)\alpha + c, \tag{6}$$

with an integration constant c. During this period the system is equivalent to a LRS vacuum Bianchi I –or Kasner– solution. This Kasner regime is completely characterized by the values of the constants of motion p and c. As illustrated in Fig. 1 and Fig. 2, if we begin the evolution towards the singularity in one of these regimes, with $p > 0$ and at large values of α, the shape parameter β follows the linear behavior as given in (6) until the exponential term $e^{4\alpha - 8\beta}$ ceases to be negligible. At that point, a transition happens and the system enters in another Kasner regime.

In order to study this transition, a useful procedure is to exploit the conserved quantities of the system. The mentioned constants of motions are found to be the following:

$$R_1 := 2H - p, \tag{7}$$

$$R_2 := e^{2(\alpha+\beta)}(H - p). \tag{8}$$

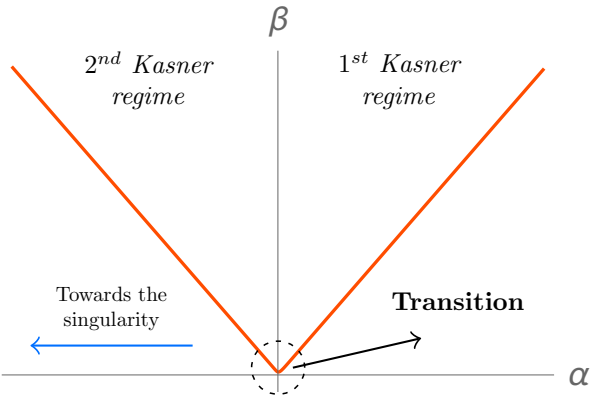


Fig. 1. Classical evolution of the variable β with respect to α, where $p > 0$ for large values of α and β.

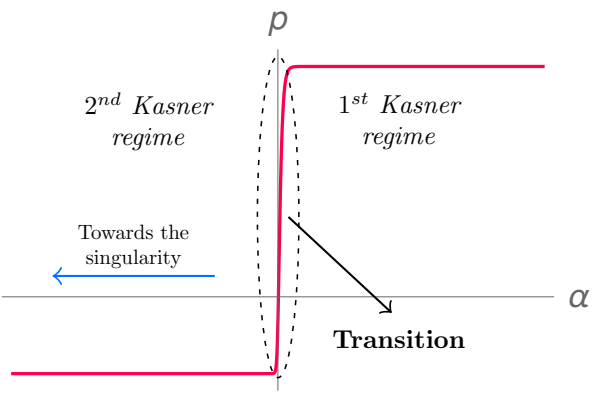


Fig. 2. Classical evolution of the variable p with respect to α, where $p > 0$ for large values of α and β.

If we evaluate these quantities in the initial and final states and equate them, we obtain a system of two equations, which relate the parameters of the first Kasner epoch $(\overline{c}, \overline{p})$ with those of the second one $(\widetilde{c}, \widetilde{p})$. Thus, after solving it we obtain the law that describes the transition between the two Kasner regimes, namely:

$$\widetilde{p} = -\frac{1}{3}\overline{p}, \tag{9}$$

$$\widetilde{c} = -3\overline{c} - \frac{1}{2}\ln\left(\frac{2}{3}\overline{p}^3\right). \tag{10}$$

3. Quantum Transition Rules

Once that we have provided a canonical description of the classical model, we are ready to develop the quantum analysis. The quantum dynamics is governed by the

following Hamiltonian operator, which is defined from the classical one and chosen to be Weyl-ordered, that is, with a totally symmetric ordering of the basic operators:

$$\hat{H}(\hat{p},\hat{\beta}) = \left(\hat{p}^2 + 2e^{4\alpha-8\hat{\beta}}\right)^{1/2}_{\text{Weyl}}. \tag{11}$$

Then, instead of studying the evolution of the wave function itself, we follow a formalism developed for quantum cosmology and quantum mechanics in general, which is based on the moment decomposition of the wave function.[4] Thus, according to this formalism, the system is fully described by the expectation values of the basic variables,

$$\beta := \langle\hat{\beta}\rangle, \quad p = \langle\hat{p}\rangle, \tag{12}$$

as well as the following infinite set of quantum moments:

$$\Delta(\beta^i p^j) := \langle(\hat{\beta}-\beta)^i(\hat{p}-p)^j\rangle_{\text{Weyl}} \quad (i,j\in\mathbb{N}). \tag{13}$$

The sum of the indices $i+j$ will be referred as the order of the corresponding moment. The evolution of these variables is ruled by a quantum effective Hamiltonian H_Q, that is defined as the expectation value of $\hat{H}$. By performing a Taylor expansion around the expectation values of $\hat{\beta}$ and $\hat{p}$, it can be written as the following infinite series:

$$\begin{aligned} H_Q &= \left\langle \hat{H}(\hat{p},\hat{\beta})\right\rangle \\ &= H(\beta,p) + \sum_{i+j=2}^{\infty} \frac{1}{i!j!}\frac{\partial^{i+j}H(\beta,p)}{\partial\beta^i\partial p^j}\Delta(\beta^i p^j), \end{aligned} \tag{14}$$

where $H(\beta,p)$ is the classical Hamiltonian (3). This effective Hamiltonian H_Q encodes the complete dynamical information of the system. Indeed, the equations of motion for the variables of the system –β, p and $\Delta(\beta^i p^j)$– are obtained by computing their Poisson brackets with H_Q:

$$\frac{d\beta}{d\alpha} = \{\beta, H_Q\} = \frac{\partial H_Q}{\partial p}, \tag{15}$$

$$\frac{dp}{d\alpha} = \{p, H_Q\} = -\frac{\partial H_Q}{\partial \beta}, \tag{16}$$

$$\begin{aligned} \frac{d\Delta(\beta^i p^j)}{d\alpha} &= \{\Delta(\beta^i p^j), H_Q\} \\ &= \sum_{m,n=0}^{+\infty} \frac{1}{m!n!}\frac{\partial^{m+n}H(\beta,p)}{\partial\beta^m\partial p^n}\{\Delta(\beta^i p^j), \Delta(\beta^m p^n)\}, \end{aligned} \tag{17}$$

where the Poisson brackets between expectation values are defined in the usual way in terms of the commutator, that is, $\{\langle\hat{A}\rangle, \langle\hat{B}\rangle\} := -i\langle[\hat{A},\hat{B}]\rangle/\hbar$. Let us emphasize

that this evolution of the variables is completely equivalent to that given by the Schrödinger picture in terms of the wave function.

We observe that the classical Hamiltonian H defined in (3) is not polynomial, which means that the series in (14) is indeed infinite and it is not truncated at certain order. Thus, the equations of motion (15)-(17) form an infinite system of coupled equations. Therefore, in order to study and compute the dynamics explicitly, a cut-off must be introduced. That is, we will consider all the moments of order higher than a certain value N to be vanishing:

$$\Delta(\beta^i p^j) \approx 0 \quad \text{for} \quad i + j > N. \tag{18}$$

In fact, such a truncation corresponds to considering a semiclassical regime, where the quantum state is peaked on a classical trajectory and hence the high-order moments are negligible.

However, even if we apply a cut-off, instead of solving the coupled system of equations for any value of α, we will focus on the so-called Kasner regimes, as we have done in the classical analysis. During these periods, the equations of motion take the form

$$\frac{d\beta}{d\alpha} \approx \text{sign}(p), \quad \frac{dp}{d\alpha} \approx 0, \quad \frac{d\Delta(\beta^i p^j)}{d\alpha} \approx 0, \tag{19}$$

which are immediate to solve:

$$\beta \approx \text{sign}(p)\alpha + c, \tag{20}$$

$$p \approx const., \tag{21}$$

$$\Delta(\beta^i p^j) \approx const., \tag{22}$$

with an integration constant c. Comparing these results with the classical ones, we observe that the behavior of p and β in these states remains identical in the quantum analysis. Moreover, the moments hold a constant value. In summary, the parameters that characterize each quantum Kasner epoch will be, as in the classical case, the constants c and p, along with the infinite constant set of moments $\Delta(\beta^i p^j)$. Then, in order to analyze the transition between two of these quantum Kasner regimes, we will once again exploit the constants of motion. In this quantum scenario, the conserved quantities are found to be the expectation values $\langle \hat{R}_1 \rangle$ and $\langle \hat{R}_2 \rangle$, where $\hat{R}_1$ and $\hat{R}_2$ are the operator counterpart of the classical conserved quantities (7) and (8). In fact, any of the following combinations is also a constant of motion:

$$\langle \hat{R}_i^n \hat{R}_j^m \rangle = const., \quad \text{for} \quad m, n \in \mathbb{N} \text{ and } i, j \in \{1, 2\}. \tag{23}$$

Nonetheless, as a technical subtlety, it must be mentioned that the choice of ordering of these operators is far from trivial: one must find the proper ordering so that they are conserved. Furthermore, at higher orders in moments, there may appear some terms that go with some explicit powers of $\hbar$. A more detailed discussion about these issues can be found in.[3]

Based on the classical results, we will consider that we have an initial Kasner regime characterized by the parameters $(\overline{c}, \overline{p}, \overline{\Delta(\beta^i p^j)})$ and that, as we approach the singularity, another such regime is reached described by the parameters $(\widetilde{c}, \widetilde{p}, \widetilde{\Delta(\beta^i p^j)})$. Hence, if we evaluate the constants of motion in these two states and equate them, we obtain a system of algebraic equations that relate the different parameters. After solving this system, we get the quantum transition law, which takes the following form:

$$\widetilde{p} = -\frac{1}{3}\overline{p}, \tag{24}$$

$$\widetilde{c} = -3\overline{c} - \frac{1}{2}\ln\left(\frac{2\overline{p}^2}{3}\right) + \sum_{n=2}^{N} \frac{(-1)^n}{n\overline{p}^n}\overline{\Delta(p^n)}, \tag{25}$$

$$\widetilde{\Delta(\beta^m p^n)} = \sum_{l=0}^{N-(m+n)} \sum_{k=0}^{m} \frac{a_{mnkl}}{\overline{p}^{m+l-k}} \overline{\Delta(\beta^k p^{m+n+l-k})} \tag{26}$$

$$+ \sum_{r=2}^{N-2} \sum_{k=0}^{m-1} \sum_{l=l_{min}}^{l_{max}} \frac{b_{mnklr}}{\overline{p}^{m+r+l-k}} \overline{\Delta(p^r)}\ \overline{\Delta(\beta^k p^{m+n+l-k})} + \sum_{k=1}^{\lfloor N/4 \rfloor} \frac{c_{mnk}}{\overline{p}^{2k-n}} \hbar^{2k},$$

where N denotes the order of the truncation. On the one hand, as can be seen from (24), the transition law for p remains identical as the classical one. However, we do obtain some quantum corrections for the transition law of β, as can be seen by comparing (10) with (25). Nevertheless, as expected, if all the moments are set to zero we recover the classical transition law. On the other hand, the transition law for the moments can be written in the compact form (26), with certain numerical coefficients a_{mnkl}, b_{mnklr} and c_{mnk}. Up to fifth order in moments, we have obtained the explicit expression of these coefficients and we refer the reader to[3] for the detailed expressions. If we analyze this formula in detail, we observe that it is linear in all of the moments except for some quadratic terms that contain pure moments of p. Moreover, regardless of the order of the truncation, there is a strong dependence on the index m of the shape parameter β of the corresponding moment. In particular, this index determines which initial moments will directly affect the value of a given final moment. For instance, the value of $\widetilde{\Delta(p^n)}$ (which corresponds to $m = 0$) only depends on the initial pure p-moments $\overline{\Delta(p^n)}$. Nevertheless, the value of $\widetilde{\Delta(\beta p^n)}$, that is $m = 1$, depends not only on pure p-moments, but also on $\overline{\Delta(\beta p^n)}$. In general, the value of the final moment $\widetilde{\Delta(\beta^m p^n)}$ will depend on the initial moments $\overline{\Delta(\beta^j p^i)}$ with $j \leq m$.

4. Quantum Dynamics

Once we have studied the asymptotic characteristics of this model in the Kasner regimes, we complete it by performing a numerical analysis to examine in detail its

dynamical evolution. For this purpose, we have chosen an initial Gaussian state, namely

$$\overline{\Delta(\beta^{2m}p^{2n})} = 2^{-2(m+n)}\hbar^{2m}\sigma^{2(n-m)}\frac{(2m)!(2n)!}{m!n!}, \tag{27}$$

for $\forall m, n \in \mathbb{N}$, and vanishing otherwise. As expected, the numerical study shows that starting from high values of α and evolving towards lower ones, the system follows the Kasner dynamics until it undergoes a transition, which happens near $\alpha = 0$. This transition takes place in a very small period of time and then the system reaches its final equilibrium value –the final Kasner regime– very quickly.

Regarding the evolution of the expectation values p and β, we remark that they remain almost identical as in the classical case, that is, as depicted in Fig. 1 and Fig. 2. In fact, only in a very close zoom into the transition (near $\alpha = 0$) we can appreciate a slight modification in their evolution: quantum effects slightly increase the value of p and β during the transition.

Moreover, the evolution of the moments show that during the initial Kasner regime they hold a constant value, at certain point they begin to perform strong oscillations and then quickly relax to their final constant value, which characterizes the coherent state during the final Kasner regime. The precise dynamics of the moments are quite complicate but, if we look into the transition in more detail, some general features can be detected. For instance, we remark that the higher the index of p, the earlier the moments reach their final equilibrium value. Furthermore, we observe that the lower the order of a given moment, the sooner it starts to oscillate and with a larger amplitude, which is in agreement with the semiclassical hierarchy of moments we are assuming. More specifically, at each given order the pure fluctuations of p show the biggest amplitudes. In addition, moments with at least one odd index, *i.e.*, the initially vanishing ones, present more oscillations.

It is particularly interesting to observe that the moments that were vanishing in the initial Kasner epoch –due to the choice of the Gaussian state– are not vanishing in the final one, due to the mentioned oscillations. This is indeed an expected outcome according to the transition laws (26). However, it means that even when we choose an initial state where there are no correlations between β and p, the transition will generate correlations for the final state. In particular, all the initial vanishing moments are activated in a similar way: as they approach the transition they experience an excitation and start to grow exponentially, until they begin to oscillate.

5. Conclusions

In summary, we have described the quantum LRS Bianchi II model making use of a decomposition of the wave function into its infinite set of moments. For the classical case, we have obtained the explicit constants of motion for the full dynamics, and thus we have completely integrated the equations of motion. For the quantum case,

making use of the conserved quantities (23), we have been able to obtain the exact quantum transition law for every variable (24)-(26) up to fifth-order in moments. Furthermore, by performing a numerical analysis of the model, we have studied the dynamical part of the evolution. In particular we have focused on the region where the transition takes place, and we have derived some general features of the behavior of the different variables. The follow-up of this work is to remove the LRS symmetry and to study the most general Bianchi II model. Then, the next step will be to extend the analysis to the full Bianchi IX model, and ultimately, to obtain the quantum corrections to the BKL transitions.

Acknowledgments

SFU acknowledges financial support from an FPU fellowship of the Spanish Ministry of Universities. AA-S is supported by the ERC Advanced Grant No. 740209. This work is funded by Projects FIS2017-85076-P and PID2020-118159GB-C44 (MICINN/AEI/FEDER, UE), and by Basque Government Grant No. IT956-16.

References

1. V. A. Belinskii, I. M. Khalatnikov and E. M. Lifschitz, A general solution of the Einstein equations with a time singularity, *Adv. Phys.* **13**, 639 (1982).
2. C. W. Misner, Mixmaster Universe, *Phys. Rev. Lett.* **22**,1071 (1969).
3. A. Alonso-Serrano, D. Brizuela and S. F. Uria, Quantum Kasner transition in a locally rotationally symmetric Bianchi II universe, *Phys. Rev. D* **104**, 024006 (2021).
4. M. Bojowald and A. Skirzewski, Effective equations of motion for quantum system, *Rev. Math. Phys.* **18**, 713 (2006).

The hypercomplex medium as storage of physical equations

Alexander P. Yefremov

Institute of Gravitation and Cosmology, RUDN University,
6 Miklukho-Maklay Street, Moscow, 117198 Russian Federation,
E-mail: efremov-ap@rudn.ru
www.rudn.ru

Using the quaternion framework (Q-math), we show that the specific mathematical equations born in "quaternion medium" in physical units become known physical laws. In particular, it is shown how one can discover immanently hidden "geometric physical laws": Cartesian frames, equations of electrodynamics, Q-vector formulation of the relativity theory. One can also find the linked logical chain between laws of quantum, classical, and relativistic mechanics.

Keywords: Hypercomplex numbers; Quaternions; Fundamental equations.

1. Introduction: Hypercomplex Numbers

We draw attention to the mathematics of hypercomplex (HC) numbers, which, by analogy with differential geometry so fruitful for modern physics, can serve as an exceptionally rich environment. The HC math set comprises, first of all, four "good" algebras of real, complex, quaternion (Q) and octonion numbers; the Frobenius-Hurwitz theorem states that these are the fundamental division algebras admitting the identities of squares (Table 1).

Table 1. Classical "good" HC algebras

Algebra of numbers	Symbol	Number of units (real, imaginary)	Coefficients at units	Multiplication	Division, identity of squares
Real	R	$2^0 = 1$ (1, 0)	real	commutative, associative	+
Complex	C	$2^1 = 2$ (1, 1)	real	commutative, associative	+
Quaternion	Q, H	$2^2 = 4$ (1, 3)	real	non-commut., associative	+
Octonion	O	$2^3 = 8$ (1, 7)	real	non-commut., non-assoc.	+

"Close in spirit" algebras of "not good" HC numbers: double (perplex, split-complex), dual, biquaternion numbers (all having zero dividers) must be also mentioned here because their units can be structured similarly to the units of "good" numbers.

These sets of HC numbers, together with their adjacent domains (spinor elements and related groups), deserve to be considered not only as suitable or helpful mathematical instruments but also as an environment in which physical laws (once discovered empirically or heuristically) turn out to be immanently contained.

Without loss of generality here we make an accent on the still associative but non-commutative algebra of quaternion (Q-) numbers. Nonetheless, under the term "Q-math set" we imply the Q-numbers, biquaternion (BQ) numbers, specific Q-spinor set, and relevant invariance groups. Due to Heaviside and Gibbs with their palliative vector algebra, the Q-math set was unfairly abandoned in XX century though rarely emerging just as an exotic math tool (see e.g. Ref. 1). However, firstly incidental findings and later systematic investigations demonstrated that the Q-set contains many geometric images and equations closely related to fundamental physics. Here is a brief review of the revealed Q-math – physical coincidences.

2. First Physical Findings in Q-Math

Discovered by W.R.Hamilton in 1843, Q-numbers are built on one real scalar unit 1 (normally omitted in formulas), and three non-commuting imaginary vector units $\mathbf{i},\ \mathbf{j},\ \mathbf{k} \equiv \mathbf{q}_n$; coefficients at the units a, b_n are real

$$q = a + b_1 i + b_2 j + b_3 k \equiv a + b_n \mathbf{q}_n. \tag{1}$$

The multiplication law for the units has the form

$$\mathbf{q}_k \mathbf{q}_n = -\delta_{kn} + \varepsilon_{knj}\mathbf{q}_j, \quad 1 \cdot \mathbf{q}_n = \mathbf{q}_n \cdot 1, \tag{2}$$

$\delta_{kn}, \varepsilon_{knj}$ are 3D Kronecker and Levi-Civita tensors, index summation rule is implied. This algebra appeared to be near as "good" as that of complex numbers. With the definition of conjugation $\bar{q} = a - b_n \mathbf{q}_n$ a Q-number may be endowed the norm $\|q\| \equiv q\bar{q} = \bar{q}q = a^2 + b_n b_n$, then the algebra admits division $(q_2/q_1)_{left} = \bar{q}_1 q_2 / \|q_1\|$ (right and left as multiplication) and comprises the identity of four squares $\|q_1 q_2\| = \|q_1\|\,\|q_2\|$.

2.1. *Cartesian frame*

Hamilton was the first to realize that Q-imaginary units automatically behave like three vectors initializing an orthogonal frame which was empirically postulated by Descartes two hundred years before; in fact, this observation appeared to be the first Q-math – Physics coincidence. However, the scalar unit until recently had no clear geometric image (see below).

2.2. *Electrodynamics*

The next finding occurs a century later. Nearly abandoned quaternions (due to the predominance of the "vector algebra"), nonetheless had their faithful supporters. One of them Rudolf Fueter considered ways to build a theory of functions of quaternion variables. In 1937 he published a series of papers (e.g., Ref. 2) where a Cauchi-Riemann-type differentiability condition

$$d_u \mathbf{F} \equiv \left(\frac{i}{c}\frac{\partial}{dt} - \mathbf{q}_n \frac{\partial}{dx_n} \right) \mathbf{F} = 0 \tag{3}$$

for a vector BQ-function $\mathbf{F}(u) = (H_k + iE_k)\mathbf{q}_k$ of a BQ-variable $u = -ict + x_n\mathbf{q}_n$ was postulated. It was a great surprise to reveal that Eq. (3) written in the form of "imaginary" and "real" parts (in the C-algebra sense)

$$d_u\mathbf{F} = \underbrace{\frac{i}{c}\frac{\partial H_k}{\partial t}\mathbf{q}_k - i\frac{\partial E_k}{\partial x_k} + \frac{\partial E_k}{\partial x_n}\varepsilon_{knj}\mathbf{q}_j}_{\text{"imaginary" part}} + \underbrace{\frac{1}{c}\frac{\partial E_k}{\partial t}\mathbf{q}_k + \frac{\partial H_k}{\partial x_k} - \frac{\partial H_k}{\partial x_n}\varepsilon_{knj}\mathbf{q}_j}_{\text{"real" part}}$$

eventually (in the Q-algebra sense) decays into four real equations which exactly coincide with the vacuum Maxwell equations of electrodynamics.

2.3. *Pauli basis for Q-algebra*

Here we must recall one more event concerning Q-math non-directly, but very important. In 1927 Pauli suggested his mechanical equation for a charged quantum particle in an exterior magnetic field. To describe spin-magnetic field interaction he heuristically introduced three traceless 2×2-matrices with determinants equal to -1. Later analysis showed that these matrices multiplied by $-i$ together with the unit 2×2-matrix form a set of Q-units identically obeying the product rule (2)

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \mathbf{q}_1 = -i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \mathbf{q}_2 = -i\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \mathbf{q}_3 = -i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (4)$$

It is important to note that representation (4) is the simplest one but not unique.

3. Form-Invariance of Q-Multiplication and Q-Spaces

Within the last three decades, the interest in Q-math has grown greatly; in current databases, one finds hundreds of relevant pure math investigations as well as Q-number links with physics (see e.g. Ref. 3). The first natural aspect of this research concerns the stability of the multiplication law (Eq. (2)) under changers of the imaginary (vector) Q-units $\mathbf{q}_n$, i.e., under changes of the geometry of the Q-frames. A simple analysis shows[4] that Eq. (2) keeps its form under transformations of the vector Q-units by the vector group SO(3,C) or by $1:2$ isomorphic to it the spinor (reflection) group SL(2,C)

$$\mathbf{q}_{n'} = O_{n'm}\mathbf{q}_m \to O \in SO(3,C); \quad \mathbf{q}_{n'} = U\mathbf{q}_n U^{-1} \to U \in SL(2,C). \qquad (5)$$

In general case each group has 6 independent parameters, implying natural isomorphism with the Lorentz group. Three real parameters just rotate the frame in the 3D space; three imaginary parameters perform respective hyperbolic rotations (though specific since only the 3D frame's vectors are involved).

The Q-units derivatives by the parameters Φ_ξ can be defined as

$$d\mathbf{q}_k(\Phi_\xi) = \omega_{\xi\,kn}\mathbf{q}_n\,d\Phi_\xi, \qquad (6)$$

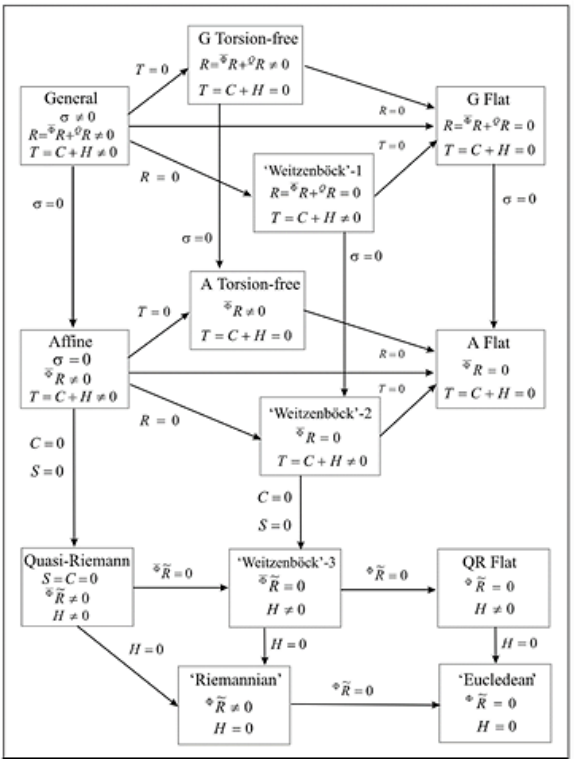


Fig. 1. Q-space of general type with full affine connection (Riemann + contortion + non-metricity) and full Q-connection (proper + Q-nonmetricity) generates at least 13 variants of different Q-spaces.

with $\omega_{\xi\,kn}$ being some coefficients skew in vector indices. If the parameters are functions of space-time coordinates x^{α}, then the Q-unit derivatives acquire the form

$$\frac{d\mathbf{q}_k}{dx^{\alpha}} = \frac{d\Phi_{\xi}}{dx^{\alpha}}\omega_{\xi\,kn}\mathbf{q}_n \equiv \Omega_{\alpha\,kn}\mathbf{q}_n; \tag{7}$$

$\Omega_{\alpha kn}$ can be associated with a connection of a specific vector space each point of which is an initial point of a Q-frame. This additional geometric property can be endowed to any space-time (e.g., already curved), thus making it a Q-enriched differential manifold with supplementary components of affine connection (therefore, curvature).[5]

For instance, apart from "standard" Riemann, Cartan, and non-metricity terms, the Q-space geometry can include respective proper Q-connection (giving zero curvature), Q-torsion, and Q-non-metricity. This gives a greatly wider scheme of the space geometry classification, hence greater possibilities to formulate generalized versions of relativistic theories of gravity comprising models with distributed polarized sources (Fig. 1).

4. Equations of Mechanics in Quaternion Frames

4.1. *Representation of Newton's dynamic equation in rotating Q-frames*

For real parameters $\Phi_\xi(t)$ of the subgroup $SO(3,R) \in SO(3,C)$ dependent only on time provide one gets a special type of the Q-connection components which in this case are physically equivalent to angular velocities of the Q-frame space rotations $\Omega_{kn} \equiv \dot{\Phi}_\xi \omega_{\xi\, kn}$. Together with the Q-vector form-invariance conditions $\mathbf{r} \equiv x_{\tilde{k}} \mathbf{q}_{\tilde{k}} = x_n\, \mathbf{q}_n$, $\mathbf{f} \equiv f_{\tilde{k}} \mathbf{q}_{\tilde{k}} = f_n\, \mathbf{q}_n$(specific force), this helps to rewrite "inertial" Newton's dynamic equation $\frac{d^2}{dt^2}\,(x_{\tilde{k}} \mathbf{q}_{\tilde{k}}) = f_{\tilde{k}}\, \mathbf{q}_{\tilde{k}}$ in an arbitrarily rotating Q-frame $\mathbf{q}_n(t)$

$$\ddot{\mathbf{r}} = \mathbf{f}. \tag{8}$$

The Q-vector components of Eq. (8) are

$$\ddot{x}_k + 2\dot{x}_n \Omega_{nk} + x_n \dot{\Omega}_{nk} + x_n \Omega_{nm} \Omega_{mk} = f_k; \tag{9}$$

its form in the traditional vector notations is

$$\ddot{\mathbf{r}} + 2[\mathbf{\Omega} \times \mathbf{v}] + [\dot{\mathbf{\Omega}} \times \mathbf{r}] + [\mathbf{\Omega} \times [\mathbf{r} \times \mathbf{\Omega}]] = \mathbf{f}.$$

Equation (8) fits for any arbitrarily complex rotations of the reference system; in particular, it describes very well the dynamics of bodies observed from the so-called chasing frames.[4]

4.2. *BQ-vector version of the theory of relativity*

Full group $SO(3,C)$ preserve form-invariance of the following "time + space" BQ-vector

$$dz = (idt\, e_k + dx_k)\, \mathbf{q}_k, \quad e_k e_k = 1, \quad e_k dx_k = 0, \tag{10}$$

which can be regarded as a specific "quaternion square root" from the space-time interval of special relativity $dz^2 = dt^2 - dr^2$. This gives a chance to build a specific BQ-vector relativity theory (Fig. 2), with the basic correlation given by Eq. (10) and the generalized "rotational equation"

$$\Sigma' = O\, \Sigma, \quad \Sigma \equiv \{\mathbf{q}_k\}, \quad O \in SO(3,C);$$

this "rotational relativity" gives all effects of Einstein's Special Relativity, but a different model of the universe, which in this case is (3+3)D. However, the most suitable (and fruitful) theory is for solving problems with non-inertial frames, directly giving solutions for the so-called hyperbolic motion, Thomas precession, linear relativistic oscillator, and relativistic shift of planets' satellites.[6,7] As well, this theory straightly admits a format of Newton-type dynamic equation for relativistic particles.[8]

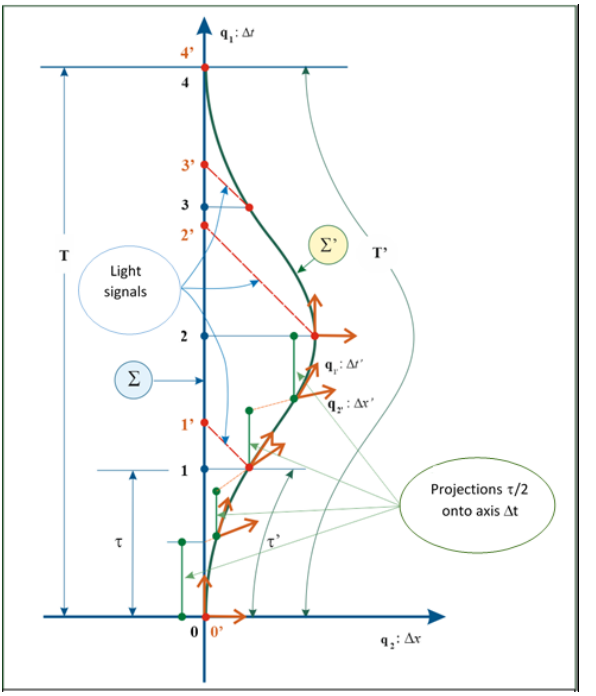


Fig. 2. Minkowski diagram for relativistic harmonic oscillator (1 cycle).

5. Q-Math Version of Pregeometry and a Matrix Image of the Complex Number

In search for the nature of quantum physics, J. A. Wheeler heuristically suggested a concept of pregeometry which "breaks loose all mention of geometry and distance".[9]

5.1. *Spectral theorem and fractal surface "under 3D Q-geometry"*

Within Q-math this concept of pregeometry can be in a way associated with a specific fractal surface (with dimensionality 1/2), being a type of subspace of 3D geometric space generated by Q-triad $\mathbf{q}_n$ (Fig. 3).

An area of the fractal surface (2D-cell) is formed by two vectors, a dyad, which originates as a bi-orthogonal basis of vectors $\{\psi^+,\ \psi^-\}$ and convectors $\{\varphi^+,\ \varphi^-\}$ within the framework of the spectral theorem since vector Q-units can be represented as simple non-degenerate matrices. In particular, the fractal dyad vectors may appear as right and left eigenvectors of one vector Q-unit. Unexpectedly it turned out that all four quaternion units are, in fact, structured, since they can be built from only one set of a Q-unit's eigenvectors (and co-vectors) as bilinear tensor

products

$$\mathbf{q}_1 = -i(\psi^+\varphi^+ + \psi^-\varphi^-), \quad \mathbf{q}_2 = \psi^+\varphi^- - \psi^-\varphi^+,$$
$$\mathbf{q}_3 = i(\psi^+\varphi^+ - \psi^-\varphi^-); \quad 1 = \psi^+\varphi^+ + \psi^-\varphi^-. \tag{11}$$

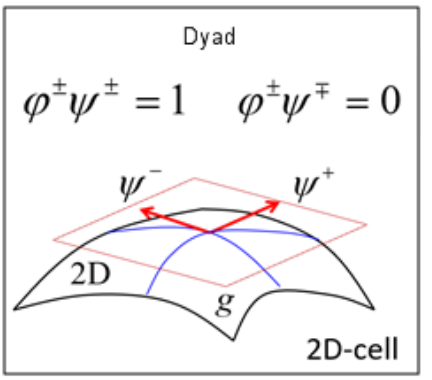


Fig. 3. A fractal surface domain.

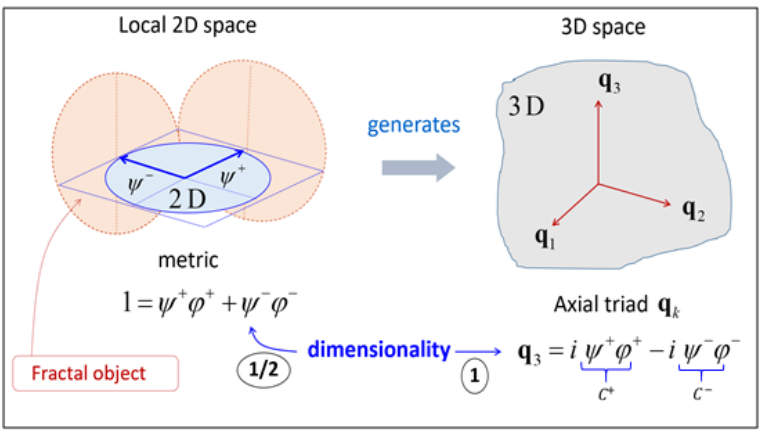


Fig. 4. Elements of a fractal surface generate volume of 3D space.

The first three equalities of Eq. (11) state that a domain of the 3D space formed by axial vector Q-triad $\mathbf{q}_n$ is a composed structure whose elements are vectors $\{\psi^+, \psi^-\}$ and convectors $\{\varphi^+, \varphi^-\}$ forming a local area of the fractal surface. Thus, the local volume of three-dimensional space associated with the triad of unit quaternionic axial vectors is, in fact, a derived geometric object, the true elements

of which are two vectors (co-vectors) of the fractal dyad. It is possible to say that an invisible 2D fractal space (a pre-geometric surface) generates a 3D geometric space (Fig. 4).

In its turn, the last equality of Eq. (11) demonstrates that the scalar unit, in this case, the unit diagonal 2×2-matrix (in Cartesian components)

$$\psi^{+(A)}\psi^{+}{}_{(B)} + \psi^{-(A)}\psi^{-}{}_{(B)} = \delta^{A}_{B}$$
$$\Rightarrow \begin{pmatrix}0\\1\end{pmatrix}(0\ 1) + \begin{pmatrix}1\\0\end{pmatrix}(1\ 0) = \begin{pmatrix}0\ 0\\0\ 1\end{pmatrix} + \begin{pmatrix}1\ 0\\0\ 0\end{pmatrix} = \begin{pmatrix}1\ 0\\0\ 1\end{pmatrix} = 1,$$

is nothing but metric tensor of the fractal surface; thus, the scalar unit may be said to acquire its not geometric but pregeometric meaning.

Another significant observation in this context is that it is precisely due to the structural decomposition of Q-units in Eq. (11) that the law of multiplication of the Q-algebra keeps its form under $SL(2, C)$ transformations. In this case, each dyad vector (co-vector) is subject to a direct $\psi' = U\psi$ (inverse $\varphi' = \varphi U^{-1}$) transformation of this group or its subgroup.

5.2. *Conic-gearing model of the complex number*

One of the pure math consequences of the existence of the Q-unit's structure is the appearance of an original matrix model of the complex number previously imaged on the complex plane or the Riemann sphere.[10]

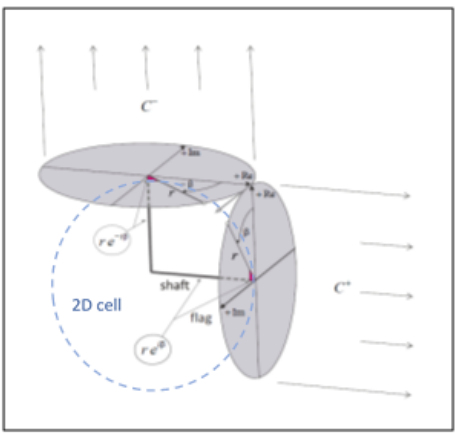


Fig. 5. C-number as a conic-gearing couple on a 2D-cell.

Indeed, if only two units are in the game (e.g., 1, $\mathbf{q}_3$), then any complex number can be represented as $z = x + iy \equiv x \cdot 1 + y \cdot \mathbf{q}_3$, $a, b \in R$; this number belongs to the C-algebra commutative in multiplication. Applying to the units the decomposition

in Eq. (11) one gets

$$z = (x + iy)\psi^+\varphi^+ + (x - iy)\psi^-\varphi^-.$$

This number can be rewritten in the polar coordinates

$$z = r\left(e^{i\beta}C^+ + e^{-i\beta}C^-\right), \tag{12}$$

here $C^\pm \equiv \psi^\pm\varphi^\pm$ are the projectors, idempotent, and mutually orthogonal matrices. Equation (12) has a clear geometric image. It models a type of conic-gearing couple, one gear orthogonal to another touching in a point (Fig. 5). Each gear has the shaft equal to r; a rotation of one gear by angle β forces another one to rotate by the angle $-\beta$. This picture conformally changes with the change of the number's modulus. Besides, its projectors C^+ and C^- determine the same orthogonal directions as the dyad vectors ψ^+ and ψ^-, so, in a way this model can be used to visualize the behavior of a 2D-cell under variations of the phase angle. Below we use it in this sense.

6. "General Theory of Particle Mechanics"

6.1. *Q-algebra stability under transformations of the fractal structure*

The next question to answer is which transformations of a fractal surface leave the algebra validity. The simplest mapping is a multiplication of the initially constant dyad by an arbitrary complex number

$$\Psi^+ = \lambda\,\psi^+, \quad \Psi^- = \lambda^*\psi^-; \quad \lambda \equiv \sigma e^{i\alpha} \tag{13}$$

thus, introducing the phase change α and the 2D-cell's conformal stretching factor σ (Fig. 6). Under mapping (13) new 3D vectors (and the scalar too) built according to the pattern (11) lose their property of unity, e.g., $\Psi^+\Phi^+ + \Psi^-\Phi^- = \lambda\lambda^*(\psi^+\varphi^+ + \psi^-\varphi^-) \neq 1$.

To cure the algebra, we, first, assume that the mapping parameters are functions of a scalar θ and coordinates ξ_n of an abstract math phase $\alpha(\theta, \xi)$, $\sigma(\theta, \xi)$, then define a normalizing integral over a volume of the space

$$f(\theta) \equiv \int_V \lambda\lambda^*\, d^3\xi = 1, \tag{14}$$

and finely, renormalize new vectors and the scalar, e.g., as $1' = f(\theta)\,(\psi^+\varphi^+ + \psi^-\varphi^-) = 1$. Further on, we demand that this normalization last "forever" (in the sense of parameter θ)

$$\partial_\theta \int_{V_\xi} \lambda\lambda^* dV_\xi = 0; \tag{15}$$

this gives a differential condition of the algebra stability (a continuity-type equation)

$$\partial_\theta(\lambda\lambda^*) + \nabla(\lambda\lambda^*\mathbf{k}) = 0; \tag{16}$$

with $\mathbf{k}$ being a 3D vector. We can choose this vector arbitrarily.

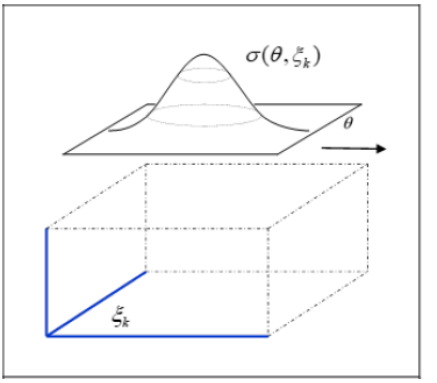


Fig. 6. σ-factor over an abstract space.

6.2. *Quantum mechanics*

If the free vector indicates the direction of the phase increase (behaves as a propagator) $k_n = \partial_n \alpha$, then the algebra stability condition given by Eq. (16) fractalizes and acquires the form

$$\left[\partial_\theta - \frac{i}{2}\left(\partial_n \partial_n - 2W\right)\right]\lambda \; = 0, \tag{17}$$

where W is an arbitrary scalar function. All magnitudes and operators entering Eq. (17) are pure mathematical (abstract), hence, measured in no units. It is natural to introduce micro-world units' standards, e.g., the Compton length $\varepsilon \equiv \hbar/(mc)$ and respective time $\tau \equiv \hbar/(mc^2)$. Then the abstract coordinates and scalar parameter are transformed as $\xi_n \to x_n/\varepsilon,\ \theta \to t/\tau$ and the fractal algebra stability condition in Eq. (17) takes the precise form of the Schrodinger equation

$$\left(i\hbar\,\partial_t + \frac{\hbar^2}{2m}\nabla^2 - U(x_n,t)\right)\lambda = 0 \tag{18}$$

the free function acquiring the meaning of the potential energy $U \equiv mc^2 W$.

In a more complex case, the propagator can depend also on a vector field $k_n = \partial_n \alpha + A_n$, this requires the use of the matrix dyad (e.g., of positive parity $\varphi,\ \psi$) and the Clifford-type metric $g_{mn} \equiv (\mathbf{p}_m \mathbf{p}_n + \mathbf{p}_n \mathbf{p}_m)/2$, $\mathbf{p}_k \equiv i\mathbf{q}_k$. Then the continuity-type stability condition takes the form $\partial_\theta(\varphi\psi) + g_{mn}\partial_m[\varphi\psi\,(\partial_n\alpha + A_n)] = 0$, while its fractal mathematical version is

$$\left[i\partial_\theta - \frac{1}{2}(-i\partial_k + A_k)(-i\partial_k + A_k) - \frac{1}{2}\mathbf{p}_k B_k - W\right]\psi = 0, \quad B_n \equiv \varepsilon_{kmn}\partial_k A_m; \tag{19}$$

written in physical units (on the base of the micro-world units' standards) Eq. (19) becomes exactly the Pauli equation for quantum-mechanical charged particle in the exterior magnetic field.[11] Thus, investigation of fractal pregeometry helps to deduce the basic quantum mechanical equations purely out of mathematical considerations.

6.3. *Non-static solution for H-atom*

The suggested approach hints to look for non-standard solutions to the quantum mechanical equation. As an example, we mention a non-static solution for the hydrogen atom where instead of Schrodinger's static ansatz for the wave function $\lambda = \sigma(r)e^{-\alpha(t)}$ a stationary function is assumed: $\lambda = \sigma(r,t)e^{i(k\varphi-\omega t)}$, with the same Coulomb potential $U = q^2/r$. In this case, the quantization follows from the normalization condition (so that no special cutting of a series is needed), while the result of the quantization precisely gives formulas of the Bohr H-atom heuristic model $m\,\Omega_n r_n^2 = n\hbar$, $\Omega_n = 2\omega/n = mq^4/(2n^2\hbar^3)$. In this stationary solution, the electron is not a point; its density is harmonically distributed along a circular orbit, and this configuration revolves around the nucleus with a constant angular velocity.[12]

6.4. *From quantum to analytical mechanics*

Decomposition of the math Schrodinger equation (Eq. (17)) into real and imaginary parts leads to a couple of the Bohm-type equations,[13] the first one providing conservation of the conformal factor $\partial_\theta\sigma + \partial_k\sigma\,\partial_k\alpha + \sigma\,\partial_k\partial_k\alpha = 0$, the second (more complicated) dealing with the "phase dynamics"

$$\partial_\theta\alpha + \frac{1}{2}\,(\partial_n\alpha)\,(\partial_n\alpha) + W_{\rm ex} + W_{\rm int} - \frac{1}{2}\partial_n\partial_n\sigma/\sigma = 0. \tag{20}$$

Separation functions of Eq. (20) into fast (σ, $W_{\rm int}$) and slow ones (α, $W_{\rm ex}$) gives a Helmholtz-type equation for the interior distribution of the σ-factor "density" onto the 2D-cell surface, while the "slow" equation

$$\partial_\theta\alpha + \frac{1}{2}\,(\partial_n\alpha)\,(\partial_n\alpha) + W_{\rm ex} = 0 \tag{21}$$

in the physical units and under assumption $S = \alpha\,\hbar$ takes the exact form of the Hamilton-Jacobi equation of the analytical mechanics

$$\partial_t S + \frac{(\partial_n S)(\partial_n S)}{2m} + U_{\rm ex} = 0, \quad U_{\rm ex} = mc^2 W_{\rm ex}. \tag{22}$$

Thus, the classical action function can be shown directly linked to the fractal 2D-cell's phase measured in values of Planck constant. Then, using evident math steps, we automatically arrive at the Lagrangian format (kinematic energy minus potential energy) heuristically handpicked for the Newtonian dynamics.[14] In this approach, an elementary object of mechanics, a particle, turns out to have two images (Fig. 7), a fractal one – a 2D-cell loaded with a fractal mass density and oscillating within its real and complex sectors, and a geometric one – a point-like massive particle rotating about one 3D-axis with the angular velocity twice greater in modulus than the angular frequency of the 2D-cell's oscillation.

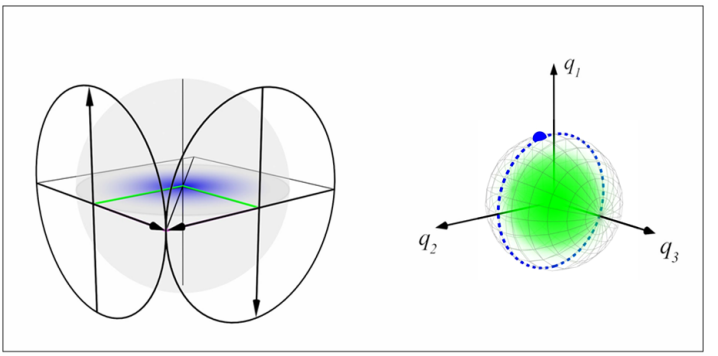


Fig. 7. Particle's 2D and 3D models.

6.5. *From analytical to relativistic mechanics*

The last model leads to an original derivation of the action function for the classical relativistic particle. We consider a point-like object of the characteristic size ε (length standard) moving in 3D space with velocity u and rotating about one axis with angular velocity Ω (twice as great as its 2D-cell analog's oscillation frequency), see Fig. 8.

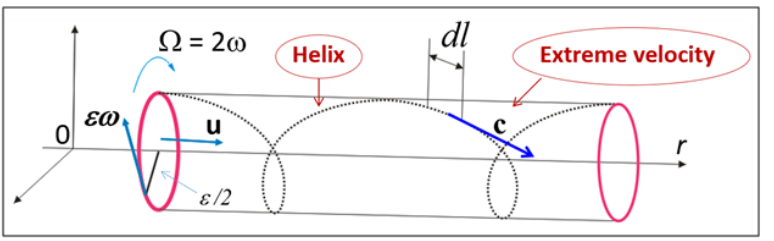


Fig. 8. Derivation of the action function for a relativistic particle.

A bordering point of this object, hence, depicts in the space a helix line; the only assumption for this model is that the velocity modulus of the bordering point is always equal to the fundamental velocity c. We compute a square of the helix' line element

$$dl^2 = c^2 dt^2 = \varepsilon^2 d\alpha^2 + u^2 dt^2$$

and find from here the phase change for the time interval $[t_1, t_2]$

$$\alpha = \pm \frac{c}{\varepsilon} \int_{t_1}^{t_2} \sqrt{1 - \frac{u^2}{c^2}}\, dt, \tag{23}$$

the signs indicate the right or left helicity. Choosing the sign minus (left helicity) and taking into account the correlation $\alpha = S/\hbar$ we rewrite Eq. (23) in the physical units to obtain

$$S = \alpha\,\hbar = -mc^2 \int_{t_1}^{t_2} \sqrt{1 - \frac{u^2}{c^2}}\, dt, \tag{24}$$

which precisely coincides with the mechanical action of the relativistic particle (previously introduced heuristically). It can be shown that reduction of Eq. (24) (written in differential form) to the non-relativistic case automatically establishes relations between classical and quantum quantities, thus linking the free particle 2D model as a De Broglie wave, the particle's rest energy linked to the permanent oscillation of the 2D cell.

6.6. *The map of the general theory of particle mechanics*

Thus, a profound investigation of the Q-math environment helps to distinguish in it a series of pure mathematical equations similar in the form to those merging from the physical experiment or found by great researchers "in a flash of a genius". Written in proper physical units these equations become the undisputed laws of quantum, classical and relativistic mechanics. The logical chain of this "general theory of particle mechanics" can be represented in the form of the following schematic map (see Fig. 9).

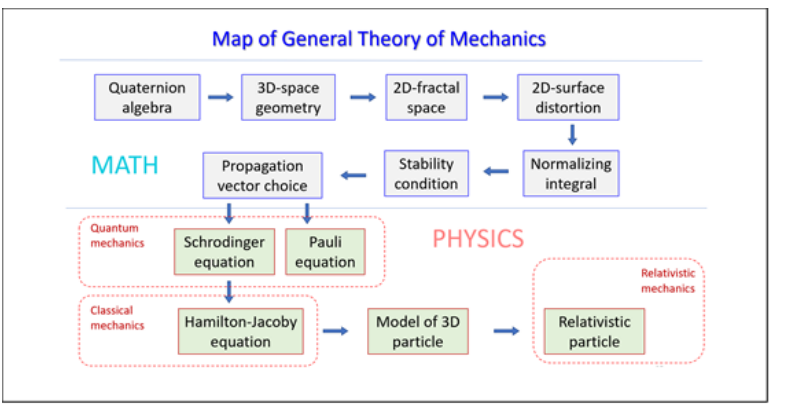


Fig. 9. Map of general theory of mechanics.

7. Conclusion and Perspectives

We demonstrated that the specific mathematical equations born in "quaternion medium" in physical units become known physical laws; briefly we remind the results. Q-math environment comprises 3D geometric and 2D pre-geometric (fractal)

domains. In the geometric domain, one discovers immanently hidden "geometric physical laws": Cartesian frames, equations of electrodynamics, Q-vector version of the relativity theory ("Rotational Relativity") giving the same effects as the special relativity. In the fractal domain, one finds the linked logical chain laws of quantum mechanics, classical mechanics, and relativistic mechanics. These coincidences are hardly incidental; moreover, they suggest new ways to move the theory ahead.

Among possible directions, one can pay attention to the following (at least three) aspects. (i) Coupling affine differential manifold with the rotational relativity gives a chance to build a new type of Q-vector gravity. (ii) Such a gravity model is known to be naturally split into spinor-type structures[15] which can reconcile relativistic gravity with quantum physics. (iii) Gauge fields, in particular, electromagnetic and Yang-Mills potentials seem to be promising objects to investigate their geometric and fractal structures within Q-math medium approaches.

References

1. P. Rastall, Quaternions in relativity, *Rev. Mod. Phys.* **36**, 820 (1964).
2. R. Fueter Über die analytische Darstellung der regulären Funktionen einer Quaternionenvariablen. *Commentarii Mathematici Helvetici* **8**, 371 (1935).
3. A. Gsponer and J.-P. Hurni, *Quaternions in mathematical physics (2): analytical bibliography* (Independent Sci. Research Inst. Oxford, OX4 4YS, England, 2005), `arXiv:math-ph/0510059`.
4. A. P. Yefremov, Physical theories in hypercomplex geometric description, *Int J. Geom. Methods Mod. Phys.* **11**, 1450062 (2014).
5. A. P. Yefremov, Vector quaternionic spaces: geometry and classification, *Grav. Cosmol.* **9**, 319 (2003).
6. A. P. Yefremov, Six-Dimensional rotational relativity, *Acta Phys. Hung.* **11**, 147 (2000).
7. A. P. Yefremov, Quaternion model of relativity: solutions for non-inertial motions and new effects, *Adv. Sci. Lett.* **1**, 179 (2008).
8. A. P. Yefremov, Newton-type dynamics in quaternion model of relativity, *Adv. Sci. Lett.* **2**, 81 (2009).
9. J. A. Wheeler, Pregeometry: motivations and prospects, in *Quantum theory and gravitation*, ed. A. R. Marlow (Academic Press, New York, 1980).
10. A. P. Yefremov, The conic-gearing image of a complex number and a spinor-born surface geometry, *Grav. Cosmol.* **17**, 1 (2011).
11. A. P. Yefremov, The fractal structure of space entails origin of Pauli's equation, *Grav. Cosmol.* **25**, 305 (2019).
12. A. P. Yefremov, Bohr-Schrödinger hydrogen truce, *Quantum Matter* **6**, 510 (2014).
13. D. Bohm, A suggested interpretation of the quantum theory in terms of hidden variables. I, *Phys. Rev.* **85**, 166 (1952).
14. A. P. Yefremov, General theory of particle mechanics arising from a fractal surface, *Grav. Cosmol.* **21**, 19 (2015).
15. A. P. Yefremov, Theory of relativity in quaternion spinors, *Grav. Cosmol.* **22**, 97 (2016).

The Maxwell-Bopp-Landé-Thomas-Podolsky-Einstein system for a static point source

Érik Amorim
Department Mathematik, Universität zu Köln,
Cologne, Germany
E-mail: edeamori@math.uni-koeln.de

We discuss the existence of a static, spherically symmetric spacetime that is the solution of the Einstein field equations coupled with an electric field obeying the equations of electromagnetism of Maxwell-Bopp-Landé-Thomas-Podolsky for a static point charge. Contrary to what happens with the Reissner-Weyl-Nordström spacetime, the electric field energy is finite, just as for this same theory on a background flat spacetime.

Keywords: Bopp-Podolsky, Field self-energy

1. Introduction

The problem of the *self-force* in electrodynamics consists of finding an expression for the force that the electromagnetic field generated by a charged point particle exerts on the particle itself. The possibility to write well-posed classical systems of equations for the joint evolution of electromagnetic (EM) fields and their sources, without resorting to *ad hoc* field averaging or bare mass renormalization at point charges, requires working with laws of electromagnetism such that the energy-momentum density of the electromagnetic field generated by charged particles is locally integrable, which is not the case for the usual Maxwell equations. There are generalized EM theories that were proposed specifically to address this problem, one of them being the so-called *Bopp-Podolsky* theory, which we will call **Bopp-Landé-Thomas-Podolsky** or **BLTP** theory. Using this theory, Kiessling and Tahvildar-Zadeh[1] have recently shown how to formulate a well-posed system for the joint evolution of point particles and their EM fields in *flat-space*, that is, the Minkowski spacetime of Special Relativity. Local integrability of the field energy-momentum at the location of the particles is essential in their work, hence why it is important to work with generalizations of the usual Maxwell equations. But the non-linearities in the Einstein field equations pose serious obstacles to generalizing this study to the theory of General Relativity.

The present work is a small first step towards extending this framework to General Relativity. It reports on a recent result by the author from his PhD thesis (Ref. 2) concerning the existence of a finite-energy solution to the Einstein equations for the spacetime of a single, static point charge whose electric field obeys the equations in Bopp-Podolsky theory.

2. Problem description

The usual Maxwell equations or **Maxwell-Maxwell system** for the EM fields $\boldsymbol{E}, \boldsymbol{B}, \boldsymbol{D}, \boldsymbol{H}$ of a static point particle of charge Q sitting at position $\boldsymbol{x} = 0$ in flat-space consists of the **pre-metric Maxwell equations**

$$\begin{cases} \nabla \cdot \boldsymbol{B} = 0 \\ \nabla \times \boldsymbol{E} + \dfrac{1}{c}\dfrac{\partial \boldsymbol{B}}{\partial t} = 0 \end{cases} , \qquad \begin{cases} \nabla \cdot \boldsymbol{D} = 4\pi Q \delta_0 \\ \nabla \times \boldsymbol{H} - \dfrac{1}{c}\dfrac{\partial \boldsymbol{D}}{\partial t} = 0 \end{cases} , \tag{1}$$

where δ_0 is the Dirac Delta, together with the **Maxwell vacuum law**, which specifies $\boldsymbol{D} = \boldsymbol{E}$ and $\boldsymbol{H} = \boldsymbol{B}$. Upon solving this system, one finds $\boldsymbol{B} = \boldsymbol{H} = \boldsymbol{0}$ and

$$\boldsymbol{D}(\boldsymbol{r}) = \boldsymbol{E}(\boldsymbol{r}) = -\varphi'(r)\boldsymbol{e}_{\boldsymbol{r}} = \frac{Q}{r^2}\boldsymbol{e}_{\boldsymbol{r}} \ ,$$

for $r = \|(x, y, z)\| = \|\boldsymbol{r}\|$ and $\boldsymbol{e}_{\boldsymbol{r}} = r^{-1}\boldsymbol{r}$, and where the *electric potential* φ is the so-called *Coulomb potential*

$$\varphi(r) = \frac{Q}{r} \ .$$

But plugging these fields into the standard integral formula for the electric field energy (set $\varkappa^{-2} = 0$ in (4) and (5) ahead) produces an integral that diverges at the location $\boldsymbol{x} = 0$ of the point charge.

In the 1940's, a modification of the above system was proposed. We call it the **Maxwell-BLTP** system, in honor of its original proponents Bopp,[3] Landé and Thomas,[4] and Podolsky.[5] The pre-metric equations (1) remain the same, but the **BLTP vacuum law** stipulates a modified relation between the electric and magnetic fields:

$$\begin{cases} \boldsymbol{D} = \boldsymbol{E} - \varkappa^{-2}\Box \boldsymbol{E} \\ \boldsymbol{H} = \boldsymbol{B} - \varkappa^{-2}\Box \boldsymbol{B} \end{cases} \tag{2}$$

where $\Box = -c^{-2}\partial_t + \Delta$ is the wave operator and $\varkappa > 0$ is a postulated large parameter of Nature with dimension of inverse length. Note that $\varkappa = \infty$ recovers the Maxwell-Maxwell system. The solution of (1) and (2) has $\boldsymbol{B} = \boldsymbol{H} = \boldsymbol{0}$ and

$$\boldsymbol{D}(\boldsymbol{r}) = \frac{Q}{r^2}\boldsymbol{e}_{\boldsymbol{r}} \ , \quad \boldsymbol{E}(\boldsymbol{r}) = -\varphi'(r)\boldsymbol{e}_{\boldsymbol{r}} \ ,$$

for the electric potential being

$$\varphi(r) = \frac{Q(1 - e^{-\varkappa r})}{r} \ . \tag{3}$$

We remark that φ is continuous at $r = 0$ and everywhere bounded. The exponential term in it is a small correction to the Coulomb potential for large values of r, while also ensuring that $\varphi(r)$ remains bounded for small r. The field energy density, which in this static case works out to be

$$\varepsilon_{\text{BLTP}} = \frac{1}{8\pi}\left(\boldsymbol{E} \cdot \boldsymbol{D} - \frac{1}{2}(|\boldsymbol{E}|^2 - \varkappa^{-2}(\nabla \cdot \boldsymbol{E})^2)\right) \ , \tag{4}$$

yields a finite value for the total field energy $\mathcal{E}$ when integrated:

$$\mathcal{E} := \int_{\mathbb{R}^3} \varepsilon_{\mathrm{BLTP}} \, \mathrm{d}V = \frac{Q\varphi(0)}{2} = \frac{Q^2 \varkappa}{2} \ . \tag{5}$$

Now suppose we "switch on gravity," that is, we seek solutions of the Einstein field equations of General Relativity for a static, spherically symmetric spacetime with a naked timelike singularity representing a single point charge at rest. The underlying manifold is $\mathbb{R}^4$ minus a line, with a global spherical coordinate system (ct, r, θ, ϕ). This universe is devoid of matter away from the particle, but the latter generates electromagnetic fields that can be calculated at any point and contribute to the stress-energy tensor $T_{\mu\nu}$, which is the source term in the Einstein equations. In the case of EM theories arising from a Lagrangian, which is the case for both the Maxwell-Maxwell and Maxwell-BLTP theories, one calculates $T_{\mu\nu}$ via the standard procedure of considering variations of the matter action with respect to the metric. A globally defined electric potential $\varphi(r)$ satisfying $\lim_{r\to\infty} \varphi(r) = 0$ enters this long calculation and must be considered an unknown of the system. We refer the reader to Ref. 2 for the details. The other two unknowns of the problem appear in the metric, which can be written in the general form

$$\mathrm{d}s^2 = -c^2 \frac{\psi(r)}{\zeta(r)} \mathrm{d}t^2 + \zeta(r) \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2) \ ,$$

where r is the area-radius coordinate and $\psi(r), \zeta(r)$ are unknown. In order for the spacetime to be asymptotically flat away from the singularity, we look for solutions satisfying

$$\lim_{r\to\infty} \psi(r) = \lim_{r\to\infty} \zeta(r) = 1 \ . \tag{6}$$

In place of φ, we will consider as the third unknown of the problem the function

$$w(r) := \frac{r^2 \varphi'(r)}{\psi(r)} + Q \ . \tag{7}$$

One can think of w as measuring how much the electric potential deviates from the Coulomb potential. Physically meaningful solutions should satisfy

$$\lim_{r\to\infty} w(r) = 0 \tag{8}$$

so that the laws of EM away from the point charge become the usual Maxwell laws.

Using the Lagrangian of the Maxwell-Maxwell theory in the above framework, it is well-known that the solution of the Einstein equations is the *Reissner-Weyl-Nordström solution*, which we abbreviate as **RWN solution**:

$$\psi^{(\mathrm{RWN})}(r) = 1 \ , \quad \zeta_M^{(\mathrm{RWN})}(r) = \left(1 - \frac{2GM}{c^2 r} + \frac{GQ^2}{c^4 r^2}\right)^{-1} \ , \quad w^{(\mathrm{RWN})}(r) = 0 \ ,$$

with G being the gravitational constant and M a free parameter that turns out to be the *ADM mass* of the spacetime. However, just like in flat-space, it can be calculated from the stress-energy tensor that the electric field energy integral (set

$\varkappa^{-2} = 0$ in (11) ahead) diverges around the singularity. This means that the field self-energy of the particle is infinite.

Using instead the Lagrangian of the Maxwell-BLTP theory, the Einstein equations reduce to the **Maxwell-BLTP-Einstein system**. In units where $c = Q = \varkappa = 1$, it reads

$$\begin{cases} \psi' = -\dfrac{\varepsilon\psi}{r^3}(w')^2 \ , \\ \\ \zeta' = \dfrac{\zeta(1-\zeta)}{r} + \dfrac{\varepsilon}{r^3}\left(\zeta^2 - \zeta^2 w^2 - (w')^2\zeta\right) \ , \\ \\ w'' = \left(\dfrac{3-\zeta}{r} + \dfrac{\varepsilon\zeta}{r^3}(1-w^2)\right) w' + \zeta w \ . \end{cases} \tag{9}$$

where the dimensionless parameter ε takes the role of Newton's G, that is, it stands for $\varepsilon := GQ^2\varkappa^2/c^4$. Even if $\varkappa$ is as large as physically reasonable — the inverse of the Planck length, — we find that ε is of order 10^{-20} for Q being the charge of an electron. Therefore it makes sense to study the solutions of system (9) as a perturbation of those corresponding to the $\varepsilon = 0$ case, which we call the **flat-space solution**:

$$\psi^{(\mathrm{flat})}(r) = 1 \ , \quad \zeta^{(\mathrm{flat})}(r) = 1 \ , \quad w^{(\mathrm{flat})}(r) = (1+r)e^{-r} \ , \tag{10}$$

where $w^{(\mathrm{flat})}$ was obtained from (7) and (3). The integral for the electric field energy, reinstating the constants $c, Q, \varkappa$, can be calculated to be

$$\mathcal{E} = \frac{1}{2}\int_0^\infty \frac{\psi}{r^2}\left(Q^2 - w^2 - \frac{1}{\varkappa^2}\cdot\frac{(w')^2}{\zeta}\right)\mathrm{d}r \tag{11}$$

(performed on any constant-time hypersurface).

The goal of this work is to prove that a finite-energy solution to system (9) exists and to study the nature of its singularity at $r = 0$. For the lack of a better name, the ensuing spacetime will be referred to as the **M-BLTP spacetime**.

Remark 2.1. For any value of $\varkappa \leq \infty$, the RWN solution (2) also solves the version of system (9) with the constants $c, G, \varkappa$ reintroduced, but it is an undesirable solution because of the divergence of the energy integral. A recent result by Cuzinatto[6] et al. shows that RWN is in fact the only physically meaningful exterior solution of this system in case the spacetime contains a black hole — that is, if the coordinate system is only valid on a domain of the form $r \in (r_0, \infty)$ for $r_0 > 0$. So it is remarkable that another solution can exist if the domain is taken to be $r \in (0, \infty)$, as it is for us.

3. Main result

The main theorem proved in Ref. 2 can be paraphrased as follows:

Theorem 3.1. There exists $\varepsilon_* \in (0,1)$ such that, for all $\varepsilon \in [0, \varepsilon_*]$, the Maxwell-BLTP-Einstein system (9) of a static point charge admits a solution $(\psi_\varepsilon, \zeta_\varepsilon, w_\varepsilon)$ for $r \in (0, \infty)$ satisfying the asymptotic properties (6), (8) and the condition $\mathcal{E} < \infty$, where $\mathcal{E}$ is given in (11). Furthermore, the behavior of the solution for large r is

$$\psi_\varepsilon(r) = 1 + O(e^{-r}) \ , \quad \zeta_\varepsilon(r) = \zeta_M^{(\mathrm{RWN})}(r) + O(e^{-r}) \ , \quad w_\varepsilon(r) = O(e^{-r}) \ ,$$

and that for small r is

$$\psi_\varepsilon(r) = O(e^{-Y\varepsilon}) \ , \quad \zeta_\varepsilon(r) = X + O(r^2) \ , \quad w_\varepsilon(r) = 1 + O(r^2) \ ,$$

where the parameters M, X, Y are determined by ε, with $X \in [1, 1.1)$ and $Y \in (0.9, 1]$. Additionally, pointwise at any $r > 0$, we have

$$\lim_{\varepsilon \to 0} (\psi_\varepsilon(r), \zeta_\varepsilon(r), w_\varepsilon(r)) = (\psi^{(\mathrm{flat})}(r), \zeta^{(\mathrm{flat})}(r), w^{(\mathrm{flat})}(r)) \ . \tag{12}$$

We give an outline of the proof:

- First it is shown that, for any fixed $r_0 > 0$, a family of solutions exists on $[r_0, \infty)$ satisfying the asymptotic conditions (6), (8) as long as $\varepsilon > 0$ is small enough. This family is parametrized by (and continuous in) two arbitrary real parameters μ and α, the first one being related to the ADM mass. The method consists of writing the solution as a power series in ε starting from the $\varepsilon = 0$ flat-space solution, and then employing a novel technique for estimating the growth of the coefficients in order to get a lower bound for the radius of convergence. Also note that conditions (6), (8) already imply the finiteness of the energy integral (11) away from $r = 0$.
- Next a suitable rewriting of the system is found that casts its study, for small r, as a first-order, 4-dimensional autonomous dynamical system around a hyperbolic equilibrium point P. The 2-dimensional unstable manifold $\mathcal{W}$ of P is analytic and consists of a continuous 1-parameter family (we call σ the parameter) of solutions which, in the original variables, corresponds to solutions satisfying the condition

  $$\lim_{r \to 0} \frac{\psi(r)(w(r) - 1)}{r^2} \in \mathbb{R} \ . \tag{13}$$

  This is proven to imply $\mathcal{E} < \infty$. More than that, it also implies $\mathcal{E} = Q\varphi(0)/2$, as in flat-space. The estimation technique mentioned above, this time adapted for power series in two real variables, enables the computation of points in $\mathcal{W}$ to arbitrarily small error.
- With the above work, we can construct, in a suitable 4-dimensional space $\{(x, y, z, r)\}$, a 3-dimensional hypersurface $\mathcal{Z}$ corresponding to solutions satisfying the desirable asymptotic conditions for large r, and a 2-dimensional surface $\mathcal{W}$ corresponding to solutions satisfying sufficient asymptotic conditions for small r to guarantee a finite $\mathcal{E}$. To complete the proof, one needs to show that $\mathcal{W}$ and $\mathcal{Z}$ intersect. To this end we consider the 3-dimensional

hypersurface $r = r_0$ of (x, y, z, r)-space. Intersected with it, $\mathcal{W}$ becomes a continuous curve $\mathcal{W}_0$ and $\mathcal{Z}$ becomes a continuous surface $\mathcal{Z}_0$. By looking at the x, y, z coordinates of points on $\mathcal{W}_0$ and $\mathcal{Z}_0$ as continuous functions of their parameters σ, μ, α, and by considering that a point intersection exists when $\varepsilon = 0$, the *Poincaré-Miranda theorem* can be applied to ensure that this point intersection still exists for small $\varepsilon > 0$. It corresponds to one solution over $r \in (0, \infty)$ for the chosen value of ε.

4. Properties of the solution

We conclude by analyzing some important properties of the M-BLTP spacetime and reporting on current work in progress related to recent results in the same vein.

4.1. *Bare mass*

There is another well-known modification of the Maxwell-Maxwell equations which also deals with the problem of infinite field energy-momentum of a point particle in flat-space. Originally proposed by Born,[7] it is part of what nowadays is commonly called *Born-Infeld electrodynamics*. It was then first observed by Hoffmann[8] that, under this formulation of EM, the singularity of the static, spherically symmetric spacetime of a resting point charge is milder than that of the RWN spacetime, in the sense that the blowup of certain curvature scalars at $r = 0$ is less severe in it. In Ref. 9, a class of electrostatic, spherically symmetric spacetimes that generalize that of Hoffmann is studied in much the same way as we did here for the BLTP laws. We call these *Hoffmann spacetimes*. An important quantity considered in that work is the so-called *mass-function* $m(r)$, which is defined by its relation to the second metric coefficient:

$$\zeta(r) = \left(1 - \frac{2Gm(r)}{c^2 r}\right)^{-1} . \tag{14}$$

The ADM mass is given by $m(\infty) := \lim_{r\to\infty} m(r)$, while the number $m(0) := \lim_{r\to 0} m(r)$ is called the **bare mass** of the spacetime. It is observed that $m(0) \in (-\infty, 0)$ for the class of Hoffmann spacetimes, and that this finite, negative bare mass contributes to the electric field energy to provide the entire mass/energy content of the spacetime:

$$\mathcal{E} = c^2(m(\infty) - m(0)) .$$

But the methods of that work are fundamentally distinct from ours. The metric coefficients of the Hoffmann spacetimes can be explicitly solved by quadrature, partly due to the helpful fact that they satisfy $g_{tt}g_{rr} = -1$. This is not the case for the M-BLTP spacetime, turning even the question of existence of a solution into a completely different and more challenging task for us. Nevertheless, one can prove that the curvature blowup of the M-BLTP spacetime is just as mild as that of the

Hoffmann spacetimes, but, contrary to the latter, the following relation holds in the former:

$$\mathcal{E} > c^2(m(\infty) - m(0)) \ .$$

This suggests that $m(0)$ doesn't have a simple interpretation in the M-BLTP spacetime as a bare mass located at the point charge.

4.2. *Equations of motion*

In Ref. 10, a weak form of the twice-contracted second Bianchi identity for static, spherically symmetric spacetimes with a "singularity worldline" is studied. Sufficient conditions for it to hold are found for when $\psi \equiv 1$ and $m(0) < 0$. Considering how the classical Bianchi identities imply the local conservation laws $\nabla_\mu T^{\mu\nu} = 0$ (equations of motion of the matter/fields), the importance of this work to the scope of our project is that it provides a rigorous formulation of an equation of motion for a spacetime singularity, such as a point charge in the way that it is modeled in our spacetime. However, we find that the M-BLTP spacetime satisfies $\psi(r) > 1$ and $m(0) = 0$, which means that this work would not be directly applicable to study the motion of point-charge-singularities in GR under the BLTP laws of electromagnetism. More work in this topic will be needed when considering spacetimes containing two or more point charges that obey the BLTP laws.

4.3. *Further solutions*

Our existence proof obtains a solution whose ADM mass parameter $M = m(\infty)$ is determined by the value of ε. Physically it should be the case that M can be chosen *a priori*, independently from ε, so that point charges of arbitrary mass can be modeled. So the question arises whether we are missing more solutions to system (9). One point where our existence proof has potentially missed other solutions is clear: we have imposed condition (13), which is sufficient but not necessary for a finite energy. It also turns out that this condition implies $m(0) = 0$, but, considering what happens in the Hoffmann spacetimes, it is not unreasonable to expect that other solutions with $m(0) < 0$ and a finite energy might exist. If this turns out to be true, then, in the proof outline above, the curve $\mathcal{W}_0$ corresponding to suitable solutions for small r could potentially be turned into a surface with parameters σ and $m(0)$, which would intersect the surface $\mathcal{Z}_0$ in more than just one point. That is, one would expect a whole 1-parameter family of solutions over $r \in (0, \infty)$, allowing for a free parameter such as M. This is currently being studied by the author.

Acknowledgments

The author would like to thank his PhD advisors Michael Kiessling and Shadi Tahvildar-Zadeh for all the valuable discussions throughout his course of studies at Rutgers University and for the opportunity to present this work during the 16th Marcel Grossmann Meeting.

References

1. M. Kiessling and S. Tahvildar-Zadeh, The Einstein–Infeld–Hoffmann legacy in mathematical relativity I: The classical motion of charged point particles, *International Journal of Modern Physics D* **28** (2019).
2. E. Amorim, Two Problems in Mathematical Physics, PhD Thesis, Rutgers University (NJ, USA, 2021), `https://doi.org/doi:10.7282/t3-czsg-4t42`.
3. F. Bopp, Eine lineare Theorie des Elektrons, *Annalen Phys.* **430** (1940).
4. A. Landé and L. H. Thomas, Finite self-energies in radiation theory. Part II, *Phys. Rev.* **60** (1941).
5. B. Podolsky, A generalized electrodynamics. Part I: Non-quantum, *Phys. Rev.* **62** (1942).
6. R. R. Cuzinatto, C. A. M. de Melo, L. G. Medeiros, B. M. Pimentel, P. J. Pompeia, Bopp–Podolsky black holes and the no-hair theorem, *Eur. Phys. J. C* **78** (2018).
7. M. Born, Modified field equations with a finite radius of the electron, *Nature* **132** (1933).
8. B. Hoffmann, Gravitational and electromagnetic mass in the Born-Infeld electrodynamics, *Phys. Rev.* **47** (1935).
9. S. Tahvildar-Zadeh, On the static spacetime of a single point charge, *Rev. Math. Phys.* **23** (2011).
10. A. Burtscher, M. Kiessling, S. Tahvildar-Zadeh, Weak second Bianchi identity for spacetimes with timelike singularities, preprint (2021), `https://arxiv.org/abs/1901.00813`.

On recent developments in the theory of relativistic dissipative fluids

V. Hoang
Department of Mathematics, University of Texas at San Antonio,
One UTSA Circle
San Antonio, Texas 78249 (USA)
E-mail: duynguyenvu.hoang@utsa.edu

I give an overview about recent results on the well-posedness and breakdown of solutions for relativistic fluid equations.

Keywords: Relativistic dissipation, fluids, quark-gluon plasma, neutron star merger, causality, blowup of solutions

1. Relativistic fluids and heavy-ion collisions

The study of the quark-gluon plasma is provides a window into fundamental physics 10^{-6} seconds after the big bang.[1] On Earth, this state of matter is produced in collisions between heavy nuclei at relativistic speeds, creating a strongly interacting mixture of quarks and gluons. Current state-of-the art modeling of the quark-gluon plasma[1] is based on relativistic fluid equations, showing that the plasma behaves very much like an expanding liquid drop. Dissipative effects[1–3] have to be taken into account to make theoretical predictions match experimental data. Dissipative effects may also be relevant for the study of neutron star mergers.[4]

In Einstein's theory of relativity, a Lorentzian manifold (M, g) is taken as a model for spacetime. M can be thought of as the set of all events in the universe, characterized by *where and when* they happen. For simplicity consider a flat spacetime $M = \mathbb{R}^4 = \{x^\alpha := (x^0, \dots, x^3)\}$, where $x^0 = ct$ is the time coordinate and c the speed of light. The Minkowski metric $g = \mathrm{diag}(-1, 1, 1, 1)$ describes distances of events between one another. We use standard tensor notation throughout (repeated greek indices summed from 0 to 3, latin indices from 1 to 3). For a textbook introduction, see Refs. 5–7.

The basic fields in a relativistic fluid description are (e, n, u^α), where the total energy density e, the particle number density n and the four-velocity $u^\alpha = \frac{1}{\sqrt{1-|\vec{v}|^2/c^2}}\left(1, \frac{\vec{v}}{c}\right)$ all depend on x^α. The four-velocity replaces the ordinary three-dimensional velocity $\vec{v}$ of a fluid and is a time-like vector satisfying $g_{\mu\nu}u^\mu u^\nu = -1$. We first review the perfect fluid equations:

$$\begin{aligned}
&\text{relativistic Euler/momentum equation:} \quad (e+p)\dot{u}^\beta = -(g^{\beta\alpha} + u^\beta u^\alpha)\nabla_\alpha p \\
&\text{conservation of energy:} \quad \dot{e} + (e+p)\nabla_\alpha u^\alpha = 0 \\
&\text{conservation of particle number:} \quad \nabla_\alpha(nu^\alpha) = 0 \\
&\text{equation of state/pressure equation:} \quad p = p(e, n)
\end{aligned} \tag{1}$$

where a dot $\dot{f} = u^\alpha \nabla_\alpha f$ denotes the relativistic material derivative. The momentum equation is analogous to Newton's second law $m \cdot a = F$, as $(e+p)$ corresponds to relativistic mass density, $a^\beta := u^\alpha \nabla_\alpha u^\beta = \dot{u}^\beta$ is the acceleration and the gradient of the pressure is a force. Additionally the projection operator $h^{\beta\alpha} := g^{\beta\alpha} + u^\beta u^\alpha$ appears in the relativistic version. In the formal limit $c \to \infty$, the velocity field $\vec{v}$ will satisfy the non-relativistic Euler equation for perfect fluids with rest mass density ρ

$$\rho \left(\partial_t v^i + v^j \partial_j v^i \right) = -g^{ij} \partial_j p. \tag{2}$$

The usual way to derive the fluid equations (1) is to define the energy momentum tensor

$$T^{\mu\nu} = (e+p) u^\mu u^\nu + p g^{\mu\nu} \tag{3}$$

and to postulate the conservation of energy and momentum $\nabla_\mu T^{\mu\nu} = 0$. This formulation goes back to the work of Einstein and Schwarzschild.[8,9] The first and second equation of (1) are equivalent to $\nabla_\mu T^{\mu\nu} = 0$. While the relativistic Euler equations are a rich source of mathematical problems, they constitute a highly idealized description of a fluid. A physically more complete description includes dissipative processes such as viscosity, diffusion and heat conduction.

To model dissipative effects of viscosity and heat conduction, the energy-momentum tensor is extended to

$$T^{\mu\nu} = (e + p + \Pi) u^\mu u^\nu + (p + \Pi) g^{\mu\nu} + \pi^{\mu\nu} + q^\mu u^\nu + q^\nu u^\mu, \tag{4}$$

where q^ν is the heat-flux vector, Π the bulk viscosity, and $\pi^{\mu\nu}$ the tensor of shear viscosity. These quantities satisfy the orthogonality and symmetry constraints

$$q^\nu u_\nu = 0, \quad \pi^{\mu\nu} u_\mu = 0, \quad \pi^{\mu\nu} = \pi^{\nu\mu}, \quad \pi^{\mu\nu} g_{\mu\nu} = 0. \tag{5}$$

The equations of motion that result from conservation of energy, momentum and particle number are now

$$\begin{aligned}
&\text{momentum:} \quad (e+p+\Pi)\dot{u}_\mu + h^\nu_\mu \nabla_\nu (p+\Pi) + h^\alpha_\mu \nabla_\nu \pi^\nu_\alpha \\
&\qquad\qquad + h_{\nu\mu} \dot{q}^\nu + \left(\nabla_\nu u_\mu + \nabla_\alpha u^\alpha h_{\mu\nu} \right) q^\nu = 0 \\
&\text{energy:} \quad \dot{e} + (e+p+\Pi) \nabla_\mu u^\mu + q_\mu \dot{u}^\mu + \nabla_\mu q^\mu + \pi^{\mu\nu} \sigma_{\mu\nu} = 0 \\
&\text{particles:} \quad \nabla_\alpha (n u^\alpha) = 0 \\
&\text{equation of state:} \quad p = p(e, n)
\end{aligned} \tag{6}$$

To complete the system, additional dynamical equations for $(\Pi, q^\nu, \pi^{\mu\nu})$ have to be postulated. A large body of literature (see Ref. 7 and references therein) has been devoted to finding suitable dynamical equations for these quantities.

For a fluid with a single species of particles, i.e. excluding mixtures and chemical reactions, these evolution equations in the Israel-Stewart model take the form

$$\begin{aligned} \tau_0\dot{\Pi} + \Pi + \zeta\nabla_\alpha u^\alpha + \left[\frac{1}{2}\zeta T\nabla_\alpha\left(\frac{\tau_0}{\zeta T}u^\alpha\right)\Pi\right] &= 0 \\ \tau_1 h^\lambda_\nu \dot{q}_\lambda + q_\nu = -\kappa T(h^\lambda_\nu\nabla_\lambda \ln T + a_\nu) - \left[\frac{1}{2}\kappa T^2\nabla_\alpha\left(\frac{\tau_1}{\kappa T^2}u^\alpha\right)q_\nu\right] \\ \tau_2 h^\alpha_\mu h^\beta_\nu \dot{\pi}_{\alpha\beta} + \pi_{\mu\nu} = -2\eta\sigma_{\mu\nu} - \left[\frac{1}{2}\eta T\nabla_\gamma\left(\frac{\tau_2}{\eta T}u^\gamma\right)\pi_{\mu\nu}\right] \end{aligned} \tag{7}$$

The appearance of the relaxation times $\tau_0, \tau_1, \tau_2 > 0$ is crucial for the stability and well-posedness of the model. The physical meaning of variables and various transport coefficients is as follows:

- Π bulk viscous pressure (isotropic force from fluid element shrinking/expanding).
- $\pi^{\mu\nu}$ shear viscous stresses (anisotropic force from transverse fluid layer movement).
- $p = p(e, n)$ hydrostatic pressure (isotropic gas pressure)
- e total energy density (rest energy density + internal energy density)
- n particle density (measures particles per volume)
- T absolute temperature
- $\zeta > 0, \eta > 0$ bulk/shear viscosity coefficient
- $\kappa > 0$ heat conductivity coefficient $\tau_0, \tau_1, \tau_2 > 0$ characteristic relaxation times for $\Pi, q^\nu, \pi^{\mu\nu}$
- $\sigma_{\mu\nu}$ kinematic shear tensor $\sigma_{\mu\nu} = h^\alpha_\mu h^\beta_\nu \frac{1}{2}(\nabla_\beta u_\alpha + \nabla_\alpha u_\beta)$, distortion of fluid element

I will refer to the set of equations (6), (7) as the full Israel-Stewart (IS) model. To see the mathematical structure of the equations (7) more clearly, the *truncated IS model*[7] (p. 302) is introduced. In (7) the divergence-type terms are dropped, leading to

$$\begin{aligned} \tau_0\dot{\Pi} + \Pi &= -\zeta\nabla_\alpha u^\alpha, \\ \tau_1 h^\lambda_\nu \dot{q}_\lambda + q_\nu &= -\kappa T(h^\lambda_\nu\nabla_\lambda \ln T + a_\nu), \\ \tau_2 h^\alpha_\mu h^\beta_\nu \dot{\pi}_{\alpha\beta} + \pi_{\mu\nu} &= -2\eta\sigma_{\mu\nu} \end{aligned} \tag{8}$$

A formal non-relativistic limit $c \to \infty$, $\tau_i \to 0$ of (6), (8) recovers the classical compressible Navier-Stokes equations.[7] The equations (7) can be derived using a systematic procedure starting from the second law to thermodynamics (entropy production) (see Ref. 7). The full IS model can also be motivated by considerations from kinetic theory, which aims to derive macroscopic fluid equations by a moment expansion process (see Refs. 10–14).

The first dissipative relativistic fluid theories were formulated by Eckart[15] and Landau-Lifshitz.[16] These models are more direct covariant analogues of the non-relativistic Navier-Stokes equations. Consequently the equations of motion have a

parabolic character and information can propagate at infinite speed, violating the most basic requirement of relativity.[17] In addition, the Eckart model can exhibit catastrophic instabilities.[18–20] As a reaction to these undesired properties, the IS model was developed in Refs. 21–25, with an initial contribution by Müller.[26] Partial differential operators with the property of finite speed of propagation are called *hyperbolic*[27] and the IS model is conjectured to have the even stronger property of *causality*. Causality requires both finite speed of propagation and that changes in initial data propagate at sub-light speeds, i.e. that the values of $(e, n, \Pi, u^\mu, q^\nu, \pi^{\mu\nu})$ at any point (t_0, x_0) depend only on the initial data given on the domain of dependence, i.e. inside the past-directed light cone.

The following heuristic argument illustrates why (7) should lead to finite speed of propagation: Consider the following modification of Fourier's law of heat conduction

$$\tau_0 \partial_t \vec{q} + \vec{q} = -\kappa \nabla T \tag{9}$$

with $\vec{q}(t, \cdot) : \mathbb{R}^3 \to \mathbb{R}^3$ being the three-dimensional heat flux. Assuming that the temperature T is proportional to the internal energy of the fluid and using conservation of energy leads to

$$\tau_0 \partial_t^2 T + \partial_t T - \kappa \Delta T = 0, \tag{10}$$

a generalized form of the heat equation, which is recovered upon setting $\tau_0 = 0$. (10) for $\tau_0 > 0$ is the telegraph equation[28] which has finite speed of propagation. The idea is therefore that the presence of relaxation times τ_i in (7) will have a similar effect and guarantee finite speed of propagation.

2. Local-wellposedness

The local well-posedness problems for the full and truncated systems (6),(7) (or (6), (8)) is still open. In Ref. 29, local-wellposedness and causality was shown for a related system used in heavy-ion collisions:

$$\begin{aligned}
u^\alpha \nabla_\alpha e + (e + P + \Pi) \nabla_\alpha u^\alpha + \pi^\alpha_\mu \nabla_\alpha u^\mu &= 0, \\
(e + P + \Pi) u^\beta \nabla_\beta u_\alpha + c_s^2 h^\beta_\alpha \nabla_\beta e + h^\beta_\alpha \nabla_\beta \Pi + h^\beta_\alpha \nabla_\mu \pi^\mu_\beta &= 0, \\
\tau_\Pi u^\mu \nabla_\mu \Pi + \Pi &= -\zeta \nabla_\mu u^\mu - \delta_{\Pi\Pi} \Pi \nabla_\mu u^\mu - \lambda_{\Pi\pi} \pi^{\mu\nu} \sigma_{\mu\nu}, \\
\tau_\pi h^{\mu\nu}_{\alpha\beta} u^\lambda \nabla_\lambda \pi^{\alpha\beta} + \pi^{\mu\nu} &= -2\eta \sigma^{\mu\nu} - \delta_{\pi\pi} \pi^{\mu\nu} \nabla_\alpha u^\alpha - \tau_{\pi\pi} \pi^{\langle\mu}_\alpha \sigma^{\nu\rangle\alpha} - \lambda_{\pi\Pi} \Pi \sigma^{\mu\nu},
\end{aligned} \tag{11}$$

which introduces new relaxation times and system coefficients $\tau_\pi, \lambda_{\pi\Pi}, \ldots$ etc. This system contains shear/bulk viscosity but no heat conduction. The equations for shear and bulk viscosity differ somewhat from the Israel-Stewart model, but follow the same general idea of introducing a relaxation mechanism.

Theorem 2.1. *Consider the Cauchy problem for equations* (11) *in Minkowski space, with initial data* $\mathring{\Psi} = (\mathring{e}, \mathring{u}^\nu, \mathring{\Pi}, \mathring{\pi}^{\mu\nu})$ *given on* $\{t = 0\}$. *There is an open set of inital data satisfying the constraints and that* $\mathring{\Psi} \in \mathcal{G}^\delta(\{t = 0\})$, *where* $1 \leq \delta < 20/19$. *Then, there exist a* $T > 0$ *and a unique* $\Psi = (e, u^\nu, \Pi, \pi^{\mu\nu})$ *defined on* $[0, T) \times \mathbb{R}^3$

such that Ψ *is a solution in* $[0,T)\times\mathbb{R}^3$ *and* $\Psi = \mathring{\Psi}$ *on* $\{t=0\}$. *Moreover, the solution* Ψ *is causal (i.e. domain of influence of initial data is contained in forward light-cones).*

The set of admissible initial data is physically reasonable and can be described by a set of inequalities involving ε, Π and the eigenvalues of π^μ_ν. The notation $\mathcal{G}^\delta(\{t=0\})$ refers to a Gevrey-type space, and we refer to Ref. 29 for more details.

3. Breakdown of solutions

Solutions to nonlinear partial differential equations may have a finite lifespan, the most elementary example being Burgers' equation.[30] For suitable C^∞ initial data, solutions develop a shock, i.e. a singularity in the first derivative after a finite time. A deeper question can also be asked: What precisely happens at breakdown time? Do derivatives or the fields themselves become infinite? For perfect relativistic fluids there exists a well-developed theory that proves shock formation[31–33] giving a detailed picture of the quantities that actually blow up. These results are very deep and rely on hidden geometric-analytic structures in the perfect fluid equations, which allow their formulation as quasilinear second-order covariant wave equations. It is a hard unsolved problem how to generalize these delicate hidden structures to dissipative fluids. Considering the available mathematical technology, asking for a detailed description of the singularity in the dissipative case seems to be out of reach for now. The investigation of these matters is clearly in its beginning stages, and in Ref.[34] we establish a first blowup result for a viscous dissipative fluid with only bulk viscosity (no shear viscosity or heat conduction). The basic ideas are based on the work by Sideris[35] and for relativistic fluids based on Guo-Tahvildar-Zadeh,[36] who developed a robust technique to show breakdown of smooth solutions for compressible fluids using the fundamental conservation $\nabla_\mu T^{\mu\nu} = 0$. Novel aspects of this method continue being discovered.[34,37] At the core of this technique lies a type of virial identity involving the second moment of the energy density, from which differential inequalities are derived, leading to the blowup via a contradiction argument.

Theorem 3.1. *Consider the viscous equations with bulk viscosity (without shear viscosity) with given smooth constitutive functions* p, ζ, τ_0. *Assume that*

$$\mathcal{C} := \int_0^\infty \frac{1}{n} \sup_{\rho\geq 0} \frac{\zeta(\rho,n)}{\tau_0(\rho,n)}\, dn < \infty$$

and make further (mild) technical assumptions of the equation of state $p = p(e,n)$. *There exists a smooth initial data* $(\mathring{\rho}, \mathring{n}, \mathring{\Pi}, \mathring{\mathbf{u}})$ *with large velocities such that the local smooth solution breaks down in finite time.*

A detailed discussion of the model and the assumptions on equations of state and bulk viscosity can be found in Ref. 34. The viscous case requires novel estimates

for Π. From its evolution, we have $\dot{\Pi} \sim \nabla_\alpha u^\alpha$, but the quantity $\nabla_\alpha u^\alpha$ is very hard to control in a non-constructive blowup proof. Instead, we use a hidden structure in the equations, namely an auxiliary transport equation in ρ, n, Π allowing us to get estimates for Π.

4. Conclusion

Including dissipation into models of relativistic fluids is necessary for a complete macroscopic description, as effects such as viscosity and heat conduction are ubiquitous in nature. Investigating dissipative effects in relativity has a long history, yet there is a gap in knowledge considering many fundamental questions about existence, uniqueness, blowup and decay properties of solutions. First breakthroughs have been made but many important problems remain open, for example to show well-posedness and singularity formation for the Israel-Stewart model (6), (7).

References

1. W. Busza, K. Rajagopal and W. van der Schee, Heavy ion collisions: The big picture and the big questions, *Annual Review of Nuclear and Particle Science* **68**, p. 339–376 (2018).
2. U. Heinz and R. Snellings, Collective flow and viscosity in relativistic heavy-ion collisions, *Ann. Rev. Nucl. Part. Sci.* **63**, 123 (2013).
3. The 2015 Nuclear Science Advisory Committee, *Reaching for the Horizon. The 2015 Long Range Plan for Nuclear Science* (US Department of Energy and the National Science Foundation, 2015), `https://science.energy.gov/~/media/np/nsac/pdf/2015LRP/2015_LRPNS_091815.pdf`.
4. M. G. Alford, L. Bovard, M. Hanauske, L. Rezzolla and K. Schwenzer, Viscous dissipation and heat conduction in binary neutron-star mergers, *Phys. Rev. Lett.* **120**, p. 041101 (2018).
5. L. P. Hughston and K. P. Tod, *An introduction to general relativity* (Cambridge University Press, 2006).
6. E. Poisson and C. M. Will, *Gravity: Newtonian, post-Newtonian, relativistic* (Cambridge University Press, 2014).
7. L. Rezzolla and O. Zanotti, *Relativistic hydrodynamics* (Oxford University Press, New York, 2013).
8. A. Einstein, The formal foundation of the general theory of relativity, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1914**, 1030 (1914).
9. K. Schwarzschild, On the gravitational field of a sphere of incompressible fluid according to Einstein's theory, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916**, 424 (1916).
10. S. Bhattacharyya, V. E. Hubeny, S. Minwalla and M. Rangamani, Nonlinear fluid dynamics from gravity, *JHEP* **02**, p. 045 (2008).
11. M. P. Heller, A. Kurkela, M. Spalinski and V. Svensson, Hydrodynamization in kinetic theory: Transient modes and the gradient expansion, *Phys. Rev.* **D97**, p. 091503 (2018).
12. R. Baier, P. Romatschke, D. T. Son, A. O. Starinets and M. A. Stephanov, Relativistic viscous hydrodynamics, conformal invariance, and holography, *JHEP* **04**, p. 100 (2008).

13. G. S. Denicol and J. Noronha, Divergence of the Chapman-Enskog expansion in relativistic kinetic theory (2016), https://arxiv.org/abs/1608.07869.
14. G. S. Denicol, H. Niemi, E. Molnar and D. H. Rischke, Derivation of transient relativistic fluid dynamics from the Boltzmann equation, *Phys. Rev.* **D85**, p. 114047 (2012), [Erratum: Phys. Rev.D91,no.3,039902(2015)].
15. C. Eckart, The thermodynamics of irreversible processes III. Relativistic theory of the simple fluid, *Physical Review* **58**, 919 (1940).
16. L. D. Landau and E. M. Lifshitz, *Fluid Mechanics - Volume 6 (Course of Theoretical Physics)*, second edn. (Butterworth-Heinemann, 1987).
17. W. A. Hiscock and L. Lindblom, Linear plane waves in dissipative relativistic fluids, *Phys. Rev. D* **35** (1987).
18. W. A. Hiscock and L. Lindblom, Stability and causality in dissipative relativistic fluids, *Annals of Physics* **151**, 466 (1983).
19. W. A. Hiscock and L. Lindblom, Generic instabilities in first-order dissipative fluid theories, *Phys. Rev. D* **31**, 725 (1985).
20. W. A. Hiscock and L. Lindblom, Nonlinear pathologies in relativistic heat-conducting fluid theories, *Physics Letters A* **131**, 509 (1988).
21. W. Israel, Nonstationary irreversible thermodynamics: A causal relativistic theory, *Ann. Phys.* **100**, 310 (1976).
22. W. Israel and J. M. Stewart, Thermodynamics of nonstationary and transient effects in a relativistic gas, *Phys. Lett. A* **58**, 213 (1976).
23. J. M. Stewart, On transient relativistic thermodynamics and kinetic theory, *Proc. R. Soc. London, Ser. A* **357**, 59 (oct 1977).
24. W. Israel and J. M. Stewart, On transient relativistic thermodynamics and kinetic theory. II, *Proc. R. Soc. London, Ser. A* **365**, 43 (feb 1979).
25. W. Israel and J. M. Stewart, Transient relativistic thermodynamics and kinetic theory, *Ann. Phys.* **118**, 341 (1979).
26. I. Mueller, Zum Paradox der Wärmeleitungstheorie, *Zeit. fur Phys* **198**, 329 (1967).
27. P. D. Lax, *Hyperbolic partial differential equations* (Courant Institute of Mathematical Sciences, 2006).
28. G. B. Nagy, O. E. Ortiz and O. A. Reula, The behavior of hyperbolic heat equations' solutions near their parabolic limits, *Journal of Mathematical Physics* **35**, p. 4334–4356 (1994).
29. F. S. Bemfica, M. M. Disconzi, V. Hoang, J. Noronha and M. Radosz, Nonlinear constraints on relativistic fluids far from equilibrium, *Physical Review Letters* **126**, p. 222301 (2021), https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.126.222301.
30. L. C. Evans, *Partial differential equations*, Graduate Studies in Mathematics, Vol. 19, second edn. (American Mathematical Society, Providence, RI, 2010).
31. D. Christodoulou, The Euler equations of compressible fluid flow, *Bull. Amer. Math. Soc. (NS)* **44**, 581 (2007).
32. J. Speck, *Shock formation in small-data solutions to 3D quasilinear wave equations*, Mathematical Surveys and Monographs, Vol. 214 (American Mathematical Society, Providence, RI, 2016).
33. M. M. Disconzi and J. Speck, The Relativistic Euler Equations: Remarkable Null Structures and Regularity Properties, *Ann. Henri Poincaré* **20**, 2173 (2019).
34. M. Disconzi, V. Hoang and M.Radosz, Breakdown of smooth solutions to the Müller-Israel-Stewart equations of relativistic viscous fluids https://arxiv.org/abs/2008.03841.

35. T. C. Sideris, Formation of singularities in three-dimensional compressible fluids, *Communications in Mathematical Physics* **101**, p. 475–485 (1985).
36. Y. Guo and A. S. Tahvildar-Zadeh, Formation of singularities in relativistic fluid dynamics and in spherically symmetric plasma dynamics, *Nonlinear Partial Differential Equations* **238**, p. 151–161 (1999).
37. Z. Jin and Y. Zhou, Formation of finite-time singularities for nonlinear hyperbolic systems with small initial disturbances, *Journal of Mathematical Physics* **61**, p. 071510 (2020).

Adiabatic solutions in general relativity as null geodesics on the space of boundary diffeomorphisms

Emine Şeyma Kutluk

Physics Department, Middle East Technical University, Ankara, 06800, Turkey
E-mail: ekutluk@metu.edu.tr
www.metu.edu.tr

We use Weinberg's trick for adiabatic modes, in a Manton approximation for general relativity on manifolds with spatial boundary. This results in a description of the slow-time dependent solutions as null geodesics on the space of boundary diffeomorphisms, with respect to a metric we prove to be composed solely of the boundary data. We show how the solutions in the bulk space is determined with the constraints of general relativity.

To give our description a larger perspective, we furthermore identify our resulting Lagrangian as a generalized version of the covariantized Lagrangian for continuum mechanics. We study the cases of 3+1 and 2+1 dimensions and show for the solutions we propose, the Hamiltonian constraint becomes the real homogeneous Monge-Ampere equation in the special case of two spatial dimensions.

Keywords: Adiabatic Solutions; Large Gauge Symmetries; Manton Approximation; Soft Modes; Moduli Space; Geometric Continuum Mechanics

1. Introduction and Summary

Asymptotic symmetries and soft theorems has been a topic that drew much attention in the last decade. Starting from 2013, Strominger have shown[1–3] that soft theorems arise as Ward identities of asymptotic space-time and/or gauge symmetries. Since then many new versions of soft theorems have been suggested.

Independently of this line of research, in quantum cosmology people have been studying similar Ward identities.[4–6] These studies heavily relies on Weinberg's argument for the existence of adiabatic modes on a generic cosmological background.[7]

We use a similar argument for general relativity in a Manton approximation.[8,9] Manton approximation tells one slow time trajectories on the space of static and minimum energy solutions, called the vacua, are good approximate solutions. In our approach we take the vacua to be gauge transformations of a reference metric; then introduce the time dependency in a way that the resulting trajectory is no longer a redundant gauge transformation-as was the case in Weinberg's argument. We study what becomes of the Lagrangian and the equations of motion for these trajectories, and see

- Lagrangian solely depends on the quantities on the boundary of spatial slices,
- constraints of general relativity serves to determine the bulk metric from its boundary values,

- solutions are geodesics on the space of vacua, which are further required to be null due to the Hamiltonian constraint.

Trajectories on diffeomorphism groups were previously used to describe the motion of ideal fluids[10] (see also Ref. 11), where they are described as geodesics with respect to a metric. We covariantize this description,[9] and show that our aforementioned solutions correspond to a specific choice of a generalized metric.

As particular examples we study solutions produced from a flat metric on 3-dimensional ball and also general two dimensional enclosed surface. We show how Hamiltonian constraint enforces coupling of two independent towers in the former and how it reduces down to homogeneous-real Monge Ampere equation in the latter case.

2. A Manton Approximation for General Relativity

2.1. *Setup*

We start by reviewing the Manton approximation[12] (see also Refs. 13, 14). Let us have a set of fields $\phi^I(t,x)$, on a spacetime described by the coordinates (t,x). Let us describe our system with a *natural* action

$$S = \int \left(\frac{1}{2} g_{IJ}(\phi)\, \dot{\phi}^I \dot{\phi}^J - V(\phi) \right) dV^{d+1} \tag{1}$$

where dV^{d+1}, with a slight abuse of notation, denotes the Riemannian volume form for a metric on the $d+1$ dimensional space-time manifold. Let us consider the configuration space, i.e. the space of static fields, called $\mathcal{F}$.

Assume that the potential have a continuous set of minima called $\mathcal{E}$; more specifically, in the configuration space let there be a submanifold such that

$$(\partial_I V)(\varphi^I) = 0 \quad \text{where} \quad \varphi^I \in \mathcal{E} \ . \tag{2}$$

Note that these static fields are solutions to equations of motion for the type of theory we are considering, since the equation of motion is given by

$$g_{IK}\left(\ddot{\phi}^I + \Gamma^I_{MJ} \dot{\phi}^M \dot{\phi}^J \right) = -\partial_K V \ . \tag{3}$$

Now the metric g_{IJ} will induce a metric on $\mathcal{E}$. The Manton approximation tells us that a *geodesic* on $\mathcal{E}$ with respect to this metric, that is slow in time, is a good approximate solution and the motion off $\mathcal{E}$ can be ignored.

Now we want to apply this procedure to general relativity. For this we need a Lagrangian for general relativity that is of the natural form. This can be achieved by considering a spacetime manifold in the form of $\mathbb{R} \times M$ with the ADM action, and then choosing Gaussian normal coordinates (GNC) where $g_{00} = -1, g_{0i} = 0$. Then, renaming the spatial part of the metric as $h_{ij} = g_{ij}$ the action we have is

$$S = \frac{1}{2} \int \mathrm{g}_h(\dot{h}_{ij}, \dot{h}_{kl})\, dt - \int V[h]\, dt \tag{4}$$

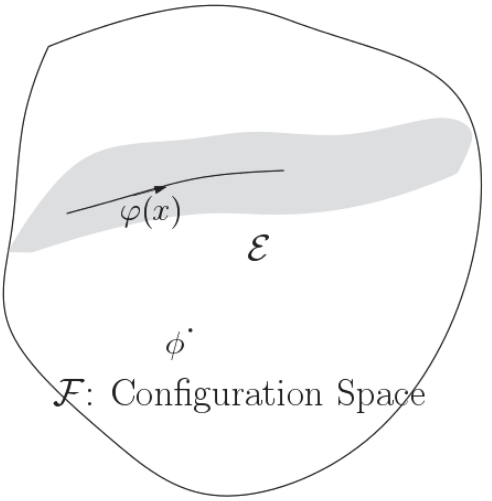


Fig. 1. Configuration space and the space of minima

where

$$\mathrm{g}_h(\delta_1 h, \delta_2 h) = \frac{1}{2}\int_M \left(h^{ik}h^{jl} - h^{ij}h^{kl}\right)\delta_1 h_{ij}\delta_2 h_{kl}\, dV^d \ , \tag{5}$$

$$V[h] = -\int_M R(h)\, dV^d \ . \tag{6}$$

Here dV^d is the Riemannian volume form and $R(h)$ Ricci scalar for h.g_h will be the metric on the configuration space. Note that since g_{00}, g_{0i}, generally known as lapse and shift, are Lagrange multipliers which we have fixed; constraints of GR

$$\nabla^i\left(\dot{h}_{ij} - h^{kl}\dot{h}_{kl}h_{ij}\right) = 0\,, \tag{7}$$

$$R(h) + \frac{1}{2}h^{ij}h^{kl}\dot{h}_{i[j}\dot{h}_{k]l} = 0\,, \tag{8}$$

needs to be imposed externally from now on.

To implement the Manton approximation, we now need to identify the set of static solutions that extremize the theory. Note that the potential is not bounded from below, so there will be no minimal energy. However it will have an extrema:

$$\delta V|_{\bar{h}} = 0 \to \quad R_{ij}(\bar{h}) = 0 \ . \tag{9}$$

So the Ricci flat metrics $\bar{h}$ and their gauge transformations will compose the space of static set of extrema of the action. For our investigation we focus on a single reference metric and its gauge transformations, i.e. we define the static extrema $\mathcal{E}$ to be composed of metrics

$$\hbar = \psi \cdot \bar{h} \ , \tag{10}$$

i.e. a Ricci flat metric and its gauge transformations. Remember we have chosen a gauge, thus the remaining gauge transformations are not full set of diffeomorphisms. They turn out to be two sets that can be written as

$$\psi_\chi = \left(0, \chi^i(x)\right) \ , \tag{11}$$

$$\psi_\alpha = \left(\alpha(x), \partial_j \alpha(x) \int_0^t h^{ij}(t', x) dt'\right) . \tag{12}$$

First of these are simply the spatial diffeomorphisms, whereas the second type are a form of local boost- they are the transformations that take a choice of "initial" hypersurface to another one when choosing GNC, see Ref. 14 for more details. Local boosts are field-dependent transformations, and thus they do not form a proper algebra- reflecting the fact that constraints of general relativity do not form a proper algebra.[a] Moreover they are not static as what a Manton description tells a vacuum ought to be. Because of these reasons we restrict our attention to spatial diffeomorphisms and define $\mathcal{E}$ as the space of metrics produced by spatial gauge transformations of the flat metric.

So we take $\mathcal{E}$ to be the set of spatial metrics that are -spatial- gauge transformations of a spatial reference metric, i.e.

$$h = \phi^* \bar{h} \ . \tag{13}$$

Written in a coordinate basis this reads as

$$h^z_{ij}(x) = \frac{\partial \phi^{\underline{k}}(x, z)}{\partial x^i} \frac{\partial \phi^{\underline{m}}(x, z)}{\partial x^j} \bar{h}_{\underline{km}} \tag{14}$$

where z parametrizes the gauge transformations. The dynamics is then described as a motion on this space:

$$h^z_{ij}(x, t) = \frac{\partial \phi^{\underline{k}}(x, z(t))}{\partial x^i} \frac{\partial \phi^{\underline{m}}(x, z(t))}{\partial x^j} \bar{h}_{\underline{km}} \ , \tag{15}$$

or

$$h(t) = \phi_t^* \bar{h} \ . \tag{16}$$

Note here the "trickiness": Once we introduce the time dependency this way, what we are doing is not a coordinate/gauge transformation in the full spacetime.

Let us illustrate this with a simple example in 3+1 dimensions. Take reference spatial metric to be the flat metric $\bar{h}_{ij} = \delta_{ij}$. Consider the transformations

$$\phi^i_t = a(t) x^i \ . \tag{17}$$

For these transformations our ansatz will produce the spacetime metric

$$ds^2 = -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) \ . \tag{18}$$

[a] Commutators of constraints of general relativity gives what is called an algebroid, see Ref. 15 and Ref. 16 for further discussion.

This illustrates very clearly that our ansatz gives us physically new solutions, for this case the spatially flat FRW spacetime.[b]

2.2. *Motion as a Geodesic on the Space of Boundary Diffeomorphisms*

In the last subsection we have proposed solutions to Einstein equations in GNC where

$$h(t) = \phi_t^* \bar{h} \ . \tag{19}$$

Now we would like to check whether these solve the Einstein equations or if they should satisfy further conditions. Note that our free variable is the diffeomorphism $\phi(t, x)$ here. We start by first checking the constraint equations that we have started imposing externally after we have eliminated the Lagrange multipliers. First we observe for our solutions Eq. (19) we have

$$\dot{h}_{ij} = \nabla_i \xi_j + \nabla_j \xi_i \qquad \text{where} \qquad \xi^i := \dot{\phi}^{\underline{k}} \frac{\partial x^i}{\partial \phi^{\underline{k}}} \qquad \text{so that} \qquad K_{ij} = \nabla_{(i} \xi_{j)} \ . \tag{20}$$

Here ∇ is the covariant derivative belonging to h. Using this and the fact $R_{ij}(h) = R_{ij}(\bar{h}) = 0$, momentum constraint becomes

$$\mathcal{P}_j(\xi) = \nabla^i \nabla_{[i} \xi_{j]} = 0 \ , \tag{21}$$

whereas Hamiltonian constraint is

$$\mathcal{H}(\xi) = (\nabla_i \xi^i)^2 - \nabla_{(i} \xi_{j)} \nabla^i \xi^j = 0 \ . \tag{22}$$

Furthermore the dynamics will be determined with the reduced Lagrangian, where one plugs Eq. (19) in Eq. (4). For our solutions the potential is zero whereas the kinetic term is the length of $\dot{h}$ with respect to the metric

$$\mathrm{g}_h(\delta_1 h, \delta_2 h) = \frac{1}{2} \int_M \left(\nabla_{(i} \xi_{1j)} \nabla^i \xi_2^j - (\nabla_i \xi_1^i)(\nabla_i \xi_2^i) \right) dV^d \ . \tag{23}$$

This is the metric Eq. (5) on the configuration space pulled back to the space of vacua. As we have noted, solutions to the dynamical equation of the above action will be geodesics on the space of vacua with respect to this metric. Then we see *Hamiltonian constraint restricts these geodesics to be null.*

Now note that each spatial slice is diffeomorphic to the reference slice, so everything can be re-expressed in coordinates where the spatial metric is equal to reference metric i.e. defining

$$\sigma := \phi_{t*} \xi \tag{24}$$

[b]This will not solve vacuum Einstein equations, but this is not relevant for the point made here.

constraint equations and the metric on the vacua space can be rewritten as

$$\mathcal{P}_j(\sigma) = \bar{\nabla}^i \bar{\nabla}_{[i}\sigma_{j]} = 0 \ , \tag{25}$$

$$\mathcal{H}(\sigma) = (\bar{\nabla}_i \sigma^i)^2 - \bar{\nabla}_{(i}\sigma_{j)} \bar{\nabla}^i \sigma^j = 0 \ , \tag{26}$$

$$\mathrm{g}_h(\delta_1 h, \delta_2 h) = \frac{1}{2} \int_M \left(\bar{\nabla}_{(i}\sigma_{1j)} \bar{\nabla}^i \sigma_2^j - (\bar{\nabla}_i \sigma_1^i)(\bar{\nabla}_i \sigma_2^i) \right) dV^d(\bar{h}). \tag{27}$$

We will use this coordinate frame for ease of calculation. Note that coordinate transformation to this frame will be different for each spatial slice.

Now we proceed to showing that the metric on the space of vacua, Eq. (27), is composed of the boundary data. For this we would like to decompose things into parts tangent and normal to the boundary ∂M, so we simply write everything in the basis $\{n, e^a\}$ where n is the unique normalized vector field orthogonal to ∂M and e^a is a orthogonal basis tangent to ∂M.[c] Now if one uses the momentum constraint Eq. (25) within Eq. (27), and assume vector field σ to be boundary preserving i.e.

$$n_i \sigma^i \big|_{\partial M} = 0 \ , \tag{28}$$

one can see that

$$\mathrm{g}_h(\sigma_{(1)}, \sigma_{(2)}) = \oint_{\partial M} \left(\sigma^a_{(1)} D^\perp \sigma^{(2)}_a - \mathcal{K}_{ab} \sigma^a_{(1)} \sigma^b_{(2)} \right) dV^{d-1}(\bar{k}) \tag{29}$$

where $\mathcal{K}$ is the extrinsic curvature of ∂M in M, σ^a is the part of σ^i that is tangential to ∂M, $\bar{k}$ is the induced metric on the boundary and we have defined a normal derivative

$$D^\perp \chi^a \equiv n^i \partial_i \chi^a + \Gamma^a_b \chi^b \quad \text{where} \quad \Gamma^b_a = n^i e^b_j \bar{\nabla}_i e^j_a \ . \tag{30}$$

Note that

- $D^\perp$ is $\bar{k}$-compatible.
- For a basis where $[n, e_a] = 0$ one has $\Gamma^a_b = K^a_b$.

Looking back at the metric Eq. (29) we see that it is completely composed of the boundary data, except the term with the normal derivative. Remember that σ still need to satisfy the constraint equations Eq. (25) and Eq. (26) in the bulk. If we can show given a vector field on the boundary, there exist a solution to these equations with unique normal derivative $D^\perp \sigma$ on the boundary, we can say we have reduced our action to boundary data. Let us examine these equations and their solutions.

We will start with the momentum constraint: Using the theory of differential forms on manifolds with boundary, especially the *Hodge-Morrey-Friedrichs decomposition* following theorem can be proven.[8]

[c] We take this to be true also slightly off the ∂M.

Theorem 2.1. *On a homologically trivial manifold with boundary, given a boundary vector field ζ^a, there exists a vector field σ such that*

$$P(\sigma) = 0, \quad n_i \sigma^i\big|_{\partial M} = 0, \quad \sigma^a = \zeta^a \tag{31}$$

and unique up to the gradients of scalars that vanish on the boundary. More specifically σ can be written as

$$\sigma = (d\alpha + *d * \beta)^\sharp \quad \textit{where} \quad \mathrm{n}\beta = 0 \ . \tag{32}$$

This is true in arbitrary dimensions, and in 3d it takes the form of the Helmholtz theorem. We conjecture that the arbitrariness in α will be fixed by the Hamiltonian constraint, though we were not able to prove this; please see Ref. 8 for more details.

However to show that the metric on the space of vacua is solely determined by boundary data one will not need to show this. This is because for the exact part where arbitrariness lies, one can show

$$D^\perp \partial_a \alpha\big|_{\partial M} = -\mathcal{K}_a^b \partial_b \alpha \, . \tag{33}$$

With this, we have shown that the metric on the space of vacua is solely composed of the data on the spatial boundary.

Because of this conclusion, we can consider this metric as an inner product on the space of boundary vector fields now:

$$\mathrm{g}_h(\sigma_1, \sigma_2) = \langle \zeta_{(1)}, \zeta_{(2)} \rangle_h \tag{34}$$

$$\equiv \oint_{\partial M} \left(\zeta^a_{(1)} D^\perp \zeta^{(2)}_a - \mathcal{K}_{ab} \zeta^a_{(1)} \zeta^b_{(2)} \right) dV^{d-1}(\bar{k}) \ . \tag{35}$$

We conclude this section by noting that the space of vacua now can be described as

$$\mathcal{V} \cong \frac{\mathrm{Diff}(\partial M)}{\mathrm{biso}(\bar{h})}. \tag{36}$$

We quotient by boundary preserving isometries of the reference metric since these will disappear in g_h, see the original expression Eq. (23). This quotient can be performed precisely because of our conjecture that for a given arbitrary boundary diffeomorphism there exists a unique solution to constraint equations implies $\mathrm{Diff}(\partial M) \cong \mathrm{BDiff}_C(M)$, boundary preserving diffeomorphisms of M that satisfies constraint equations.

One can show that $\mathcal{V}$ is a (pseudo-)Riemannian homogeneous space: i.e. the metric $\mathrm{g}_{\bar{h}}$ we have defined is $\mathrm{Diff}(\partial M)$ invariant. This metric will have a mixed signature. Indeed one can see $\langle , \rangle$ is zero for boundary vector fields such that

$$D^\perp \zeta^a - \mathcal{K}^a_b \zeta^b = \mathbb{D}\zeta = 0 \ . \tag{37}$$

i.e. if $\zeta \in \ker \mathbb{D}$. One can show that isometries are necessarily in $\ker \mathbb{D}$. If the reverse is not also true the metric will be degenerate on the space of vacua.

3. Geometric Continuum Mechanics Interpretation

The idea of describing motion as a trajectory on a diffeomorphism group is well-known in studies of "ideal continuum motion",[11] which are generalizations of Arnold's work on putting motion of an incompressible-ideal fluid in the Lagrangian and Hamiltonian framework.[10] In the following we summarize a covariantized generalization of ideal continuum motion of Ref. 11, as presented in Ref. 9. Later on we will see how our formalism of previous section fits within this generalization.

Consider a manifold M with a flow $\phi(x,t)$ on it. Derivative with respect to t, at $t=0$ of this flow is a vector field on M. What about at other times? At other times hardship arises since $\dot{\phi}(x,t) := \partial_t \phi(x,t)$ is the velocity at the point $\phi(x,t)$, not x. However one can define

$$\chi(x;t) := \dot{\phi}(y,t)\Big|_{y=\phi^{-1}(x;t)}, \tag{38}$$

then $\chi(x;t)$ will be a proper vector field. Now assume there exists a metric g on this manifold M, and consider the action

$$S = \frac{1}{2}\int_I \int_M g(\chi, \mathbb{D}\chi) dV_g \, dt \tag{39}$$

for some interval I of t, where dV_g is the Riemannian volume form for the metric g and $\mathbb{D} : \mathfrak{X}(M) \to \mathfrak{X}(M)$ is a self-adjoint, positive definite operator. We want to find the equation of motion for this action. Since the operator $\mathbb{D}$ is self-adjoint

$$\delta S = \int_I \int_M g(\delta\chi, \mathbb{D}\chi) dV_g \, dt. \tag{40}$$

Now let us find $\delta\chi$ in terms of $\delta\phi$. We see

$$\delta\chi^i(x;t) = \delta^i \dot{\phi}(y;t) - \partial_j \dot{\phi}(y;t) \partial_k \left(\phi^{-1}\right)^j \delta\Phi^k(x;t) \tag{41}$$

where we set $y = \phi^{-1}(x;t)$ and defined

$$\delta\Phi^i(z;t) = \delta\phi^j\left(\phi^{-1}(z;t);t\right) . \tag{42}$$

With further modifications we get

$$\delta\chi(x;t) = \partial_t \delta\Phi(x;t) + \mathcal{L}_\chi \delta\Phi , \tag{43}$$

so that

$$\delta S = \int_I \int_M g\left(\partial_t \delta\Phi + \mathcal{L}_\chi \delta\Phi, \mathbb{D}\chi\right) dV_g dt . \tag{44}$$

To get variational principle working we should integrate by parts to collect all the operators to one side. For this we note

$$\int_M g\left(\mathcal{L}_\chi \eta, \alpha\right) dV_g = \int_M \alpha_i \left(\chi^j \nabla_j \eta^i - \eta^j \nabla_j \chi^i\right) dV_g , \tag{45}$$

$$= \int_{\partial M} \alpha_i \eta^i \chi^\perp - \int_M \eta^i \left(\nabla_j \alpha_i \chi^j + \alpha_i \nabla_j \chi^j + \alpha_j \nabla_i \chi^j\right) dV_g , \tag{46}$$

where ∇ is the unique metric compatible and torsion free connection for g. We will assume M to have no boundary so that the first term drops. Defining $\mathcal{L}_\chi^\dagger$ as

$$\int_M g\left(\mathcal{L}_\chi \eta, \alpha\right) dV_g = \int_M g\left(\eta, \mathcal{L}_\chi^\dagger \alpha\right) dV_g \tag{47}$$

we see that

$$\mathcal{L}_\chi^\dagger \alpha = -\left(\nabla \cdot \chi\, \alpha + \mathcal{L}_\chi \alpha + \alpha \lrcorner \mathcal{L}_\chi g\right) \ , \tag{48}$$

where $\mathcal{L}_\chi g_{ij} = 2\nabla_{(i}\chi_{j)}$ and $\left(\alpha \lrcorner \omega\right)_j := \alpha^i \omega_{ij}$. Using this then, we see that the equation of motion becomes

$$\left(\partial_t - \mathcal{L}_\chi^\dagger\right) \mathbb{D}\chi = 0 \ . \tag{49}$$

Writing down explicitly $\mathcal{L}^\dagger$ and expressing in a basis we get

$$(\partial_t + \zeta^b \nabla_b)\mathbb{D}\zeta^a + (\nabla^a \zeta_b + \delta^a_b \nabla_c \zeta^c)\mathbb{D}\zeta^b = 0 \ . \tag{50}$$

One can retrieve some classic examples for certain choices of $\mathbb{D}$:

- $\mathbb{D} = \mathrm{id}$ gives Euler's equation for the free motion of an incompressible fluid.
- $\mathbb{D} = \mathrm{id} - \nabla^2$ gives Camassa-Holm equation for shallow water waves.

As it can be easily identified we get the system of previous subsection for the choice of $\mathbb{D} = D^\perp - \mathcal{K}^a_b$.

4. Some Results in 3+1 and 2+1 Dimensions

Having described our solutions as null geodesics on the space of boundary diffeomorphisms we now proceed to studying some specific examples.

4.1. *On Round Ball*

We start by considering the case where we take Euclidean 3d ball with round boundary as reference. Let us first look at the momentum constraint. Expressed in vector calculus form it turns out to be

$$\vec{\nabla} \times \vec{\nabla} \times \vec{\sigma} = 0 \ . \tag{51}$$

Solution to this in spherical coordinates will be

$$\vec{\sigma} = a_{\ell m} r^\ell \vec{r} \times \vec{\nabla} Y_{\ell m}(\theta, \varphi) + \vec{\nabla}\left(f_{\ell m}(r) Y_{\ell m}(\theta, \varphi)\right) \ . \tag{52}$$

Note that this satisfies our theorem: σ is composed of parts that are exact and coexact with zero normal on the boundary. We conjectured that the exact part should be fixed by the Hamiltonian constraint. However we were not able to solve it even for this simple case.

Plugging in the solution, we get for the metric on the space of vacua

$$\mathrm{g}(\sigma^{(1)}, \sigma^{(2)}) = R^3 \sum_{\ell,m} \ell(\ell+1)\left((\ell-1) a^{(1)}_{\ell m} a^{(2)}_{\ell m} - 2 f^{(1)}_{\ell m}(R) f^{(2)}_{\ell m}(R)\right) \ . \tag{53}$$

Notice that even though we do not know $f_{\ell m}(r)$ in the bulk, indeed its bulk value does not appear in the metric, as we have proposed. Moreover Hamiltonian equations forces $\mathrm{g}(\sigma,\sigma)=0$, thus we see that single towers of $a_{\ell m}$ or $f_{\ell m}$ cannot be solutions, there has to be mixing.

On the boundary our solutions have the form

$$\sigma^a = \epsilon^{ab}\partial_b\tau + \bar{k}^{ab}\partial_b\kappa \ . \tag{54}$$

Note that this is basically the Hodge decomposition of a one form on S^2, thus our solution space is set of all diffeomorphisms on S^2. Except the isotropy group: boundary preserving isometries of $\bar{h}_{ij}=\delta_{ij}$, these will be rotations. For this case the isotropy group will be equal to $\ker\mathbb{D}$ since

$$\ker\mathbb{D} = \{\sigma^a \,|a_{\ell m}=0 \ \ell>1, f_{\ell m}(R)=0 \ \forall \ell,m\} = \{\sigma^a = \sum_m a_{1m}\, R\vec{R}\times\vec{\nabla}Y_{1m}\} \ . \tag{55}$$

Thus our physical vacua space is

$$\mathcal{V} = \mathrm{Diff}(S^2)/\mathrm{SO}(3) \tag{56}$$

and our metric on this is non-degenerate.

4.2. *Case of 2 Dimensions*

Since we were not able to solve the Hamiltonian constraint to find solutions in the bulk, we now check if things simplify in $2+1$ dimensions. We will consider a flat reference metric with an arbitrary boundary, and then focus on some specific choices for the boundary. In two spatial bulk dimensions momentum equation gets solved when

$$\vec{\sigma} = c\frac{\partial}{\partial\theta} + \vec{\nabla}\alpha \tag{57}$$

where c and α is arbitrary. Hamiltonian equation then becomes

$$(\partial_x\partial_y\alpha)^2 - \partial_x^2\alpha\partial_y^2\alpha = 0 \ , \tag{58}$$

the well-known *homogeneous real Monge-Ampere equation (HRMA)*.[17] Most generalized solution to this is known in a parametric form, together with some family of explicit solutions. However we find that these mostly do not survive the boundary preservingness condition we impose for simple choices of the boundary.

Let us first consider some specific examples and start with the case of a round boundary, where $r=R$. The boundary preservingness condition become

$$\partial_r\alpha|_{r=R} = 0 \ . \tag{59}$$

We find no solution that satisfy this condition. Indeed looking at the metric we see

$$\mathrm{g}(\chi,\chi) = -\frac{2}{R^2}\int \partial_\theta\alpha(R,\theta)\partial_\theta\alpha(R,\theta)d\theta \ . \tag{60}$$

Remember that the metric is derived from the Hamiltonian condition, and thus since it cannot be made to be zero for a non trivial α, there exists no non-trivial solution to this problem. (Note that the coexact part is an isometry and will be quotiented out.)

Next simplest example is the case of an elliptic boundary. We consider an ellipse that is written as

$$r = \frac{\rho}{2} \frac{\sinh(2\Upsilon)}{\sqrt{\cosh^2 \Upsilon - \cos^2 \varphi}} \ , \tag{61}$$

where Υ and ρ are parameters of the ellipse. In this case the boundary preservingness condition is

$$\partial_u \alpha|_{\partial M} = -\frac{\rho^2 c}{4} \sin(2v) \tag{62}$$

where u, v are elliptic coordinates such that

$$x = \rho \cosh u \cos v \ , \tag{63}$$

$$y = \rho \sinh u \sin v \ . \tag{64}$$

For this case we find a solution to the HRMA equation

$$\alpha(x, y) = -\frac{\rho^2 c}{2 \cosh \Upsilon} \operatorname{sgn}(x) \ y \ F\left(\frac{y}{x}\right) \ , \tag{65}$$

where F is a hypergeometric function. However this solution has a singularity at the origin. We note here that the HRMA equation is pseudo-elliptic; its elliptic version where the right hand side is positive is more studied and results on existence exists, see e.g. Ref. 18. Because of this we think this type of singularity may not be a surprise in this case. We note that this singularity will be transferred to the resulting metric. We leave further investigation of whether a sensible solution can be obtained from this to the future.

5. Conclusions

In this note we summarized the results of Refs. 8, 9 where using Manton approximation we have proposed some adiabatic solutions of GR produced from diffeomorphisms. We have concluded that an adiabatic solution to GR can be described as a geodesic on the space of vacua

$$\mathcal{V} \sim \mathrm{Diff}(\partial M)/\mathrm{biso}(\bar{h}) \tag{66}$$

that are null with respect to the metric

$$\left\langle \zeta^1, \zeta^2 \right\rangle = \int_{\partial M} \zeta^1_a \mathbb{D} \zeta^a_2 \tag{67}$$

that is solely composed of the data on the boundary of spatial slices. We remind that along the way we have used the conjecture that the Hamiltonian constraint solves for the exact part uniquely. We argued $\mathcal{V}$ is a "pseudo"-Riemannian homogeneous space with respect to this metric. The metric has mixed but constant signature, and is degenerate if $\ker \mathbb{D} \neq \text{biso}(\bar{h})$. We showed our formulation is akin to examples in continuum mechanics in which choice of $\mathbb{D}$ is different.

For the specific example of 3 spatial dimensions with round boundary; we explicitly saw our theorems to be holding, we found our vector field only up to exact parts since Hamiltonian equation seems hard to solve. However we saw any non-trivial solution of it has to have both exact and coexact parts. We also have realized $\ker \mathbb{D} = \text{biso}(\bar{h})$ for this case. In 2 spatial dimensions things got easier in solving the Hamiltonian constraint, since it becomes the Monge-Ampere equation whose solutions are known. For round boundary we found that there exists no non-trivial solution for the vector field in the bulk that is boundary preserving, and for elliptic boundary we were only able to find a solutions that is singular at the origin.

We conclude with a quick overview of the remaining issues. First of all let us list some technical questions still to be worked out:

(1) Does Hamiltonian equation completely fixes the exact part?
(2) Is $\ker \mathbb{D} = \text{biso}(\bar{h})$?
(3) For 2d case can HRMA be solved for some boundary?

Note that only after we complete solve the bulk vector field, we can construct first σ in the bulk of each spatial slice, then we can construct ξ by making the proper transformation for each slice, and only after that one can find $h_{ij}(x,t)$ and the full space-time metric for the adiabatic solutions we propose.

Analysis thus far might be suggesting the set of solutions we propose do not survive for vacuum Einstein equations, so next logical step is to perform the analysis for Einstein equations with matter, for example with a cosmological constant. Similarly one can also consider non-compact spatial slices with some fall-off conditions at infinities and study the infrared effects. This type of study would link our work to the studies in asymptotic symmetries and soft theorems and shed a different light to the subject.

Acknowledgments

This article summarizes and slightly expands the works in Refs. 8, 9, 14 which was produced in collaboration with Dieter Van den Bleeken and Ali Seraj. I am supported by Research Fund of the Middle East Technical University, Project Number DOSAP-B-105-2021-10763.

References

1. A. Strominger, Asymptotic Symmetries of Yang-Mills Theory, *JHEP* **07**, p. 151 (2014).
2. T. He, V. Lysov, P. Mitra and A. Strominger, BMS supertranslations and Weinberg's soft graviton theorem, *JHEP* **05**, p. 151 (2015).
3. T. He, P. Mitra, A. P. Porfyriadis and A. Strominger, New Symmetries of Massless QED, *JHEP* **10**, p. 112 (2014).
4. P. Creminelli, J. Noreña and M. Simonović, Conformal consistency relations for single-field inflation, *Journal of Cosmology and Astroparticle Physics* **2012**, p. 052–052 (Jul 2012).
5. K. Hinterbichler, L. Hui and J. Khoury, An infinite set of ward identities for adiabatic modes in cosmology, *Journal of Cosmology and Astroparticle Physics* **2014**, p. 039–039 (Jan 2014).
6. M. Mirbabayi and M. Simonović, Weinberg Soft Theorems from Weinberg Adiabatic Modes (2 2016).
7. S. Weinberg, Adiabatic modes in cosmology, *Physical Review D* **67** (Jun 2003).
8. E. S. Kutluk, A. Seraj and D. Van Den Bleeken, Strolling along gravitational vacua, *JHEP* **01**, p. 184 (2020).
9. E. S. Kutluk and D. V. d. Bleeken, Geodesic motion on the group of boundary diffeomorphisms from Einstein's equations (6 2021).
10. V. Arnold, Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, *Annales de l'Institut Fourier* **16**, 319 (1966).
11. D. Holm, T. Schmah and C. Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions* 01 2009.
12. N. Manton and P. Sutcliffe, *Topological Solitons* (Cambridge University Press, 2004).
13. D. M. A. Stuart, Analysis of the adiabatic limit for solitons in classical field theory, *Proc. Roy. Soc. Lond. A* **463**, 2753 (2007).
14. E. Şeyma Kutluk, Adiabatic solutions in general relativity and boundary symmetries (2021).
15. C. Isham and K. Kuchar, Representations of spacetime diffeomorphisms. i. canonical parametrized field theories, *Annals of Physics* **164**, 288 (1985).
16. M. Bojowald, Symmetries of spacetime, *International Journal of Modern Physics D* **25**, p. 1644007 (Oct 2016).
17. V. F. Zaitsev and A. D. Polyanin, *Handbook of Nonlinear Partial Differential Equations: Exact Solutions, Methods, and Problems* (Chapman and Hall/CRC, 2012).
18. N. Trudinger and X.-J. Wang, The monge-ampère equation and its applications, *Handbook of Geometric Analysis* **7** (01 2008).

The point spectrum of the Dirac Hamiltonian on the zero-gravity Kerr-Newman spacetime

M. Kiessling, E. Ling, and A. S. Tahvildar-Zadeh
Math Department, Rutgers University,
New Brunswick, New Jersey 08854, USA

In this short paper, we review the Dirac equation on the zero-gravity Kerr-Newman spacetime. Our main objective is to provide a correspondence between the classification of the bound states for the zGKN spectrum and the usual hydrogenic states $1s_{1/2}$, $2s_{1/2}$, etc. of the Hydrogen atom.

Keywords: Kerr-Newman, Dirac, Hamiltonian, Hydrogen, Point spectrum.

1. The purpose of this paper

The zero-gravity Kerr-Newman (zGKN) spacetime has been studied extensively.[8,9,13] In reference 9 it was shown that the discrete spectrum of the Dirac Hamiltonian on zGKN is nonempty. In an upcoming paper[10] we classify the discrete spectrum and show that the spectrum is indexed by three integers. See Theorem 2.1 below. It was conjectured[8,9] that the discrete spectrum of the Dirac Hamiltonian on zGKN should converge to the Bohr-Sommerfeld spectrum of the usual Hydrogen problem on Minkowski spacetime with a Coulomb potential in the limit as the ring radius of zGKN approaches 0. This problem remains open but a first step in solving this problem is to determine which states in the zGKN spectrum should correspond to which states in the usual hydrogenic spectrum. For example, which states should correspond to $1s_{1/2}$, $2s_{1/2}$, $2p_{1/2}$, etc.? The purpose of this paper is to provide this correspondence.

2. The zGKN spacetime

2.1. *The zGKN spacetime and the Dirac equation*

The zero-gravity Kerr-Newman (zGKN) spacetime[8] is obtained by formally taking Newton's gravitational constant $G \to 0$ in the Kerr-Newman spacetime. In Boyer-Lindquist coordinates, the resulting spacetime has line element

$$ds^2_{\mathbf{g}} = c^2dt^2 - (r^2 + a^2)\sin^2\theta d\varphi^2 - \frac{r^2 + a^2\cos^2\theta}{r^2 + a^2}\big(dr^2 + (r^2 + a^2)d\theta^2\big). \quad (1)$$

The zGKN spacetime is static and its orthogonal slices have the topology of two copies of $\mathbb{R}^3$ glued along a disc in the $z = 0$ plane; this topology is known as the

Zipoy topology.[18] Specifically the spacetime manifold is given by $\mathbb{R} \times \Sigma$ where

$$\Sigma \;=\; \big\{(r,\theta,\varphi) \mid r \in \mathbb{R},\, \theta \in [0,\pi], \varphi \in [0,2\pi)\big\} \setminus R \tag{2}$$

where R denotes the ring $R = \{r = 0, \theta = \pi/2, \varphi \in [0, 2\pi)\}$.

The electromagnetic fields on the Kerr-Newman spacetime $\mathbf{F}_{\mathrm{KN}} = d\mathbf{A}_{\mathrm{KN}}$ do not depend on Newton's gravitational constant and so they survive the $G \to 0$ limit. Hence the zGKN spacetime comes already decorated with the same electromagnetic fields. The four-potential reads

$$\mathbf{A}_{\mathrm{KN}} = -\frac{r}{r^2 + a^2\cos^2\theta}\left(\mathrm{Q}dt - \frac{\mathrm{Q}a}{c}\sin^2\theta\, d\varphi\right). \tag{3}$$

The field $\mathbf{F}$ is thus singular on the same ring $\{r = 0, \theta = \pi/2, \varphi \in [0, 2\pi)\}$ as the metric; for r very large and positive it exhibits an electric monopole of strength Q and a magnetic dipole moment of strength $\mathrm{Q}a$, while for r very negative it exhibits an electric monopole of strength $-\mathrm{Q}$ and a magnetic dipole moment of strength $-\mathrm{Q}a$.

Equipped with the Kerr-Newman electromagnetic fields, one can interpret the ring bounding the disc as an elementary particle. With this interpretation, it is natural to consider the Dirac equation

$$\tilde{\gamma}^{\mu}\left(-i\hbar\nabla_{\mu} + eA_{\mu}\right)\Psi + mc^2\Psi = 0, \tag{4}$$

on the zGKN spacetime and compare its spectral properties to that of the usual Hydrogen problem (i.e. the Dirac equation on Minkowski spacetime with a Coulomb potential centered at the origin in $\mathbb{R}^3$ within the Born-Oppenheimer approximation).

In equation (4), m is the mass of an electron, $-e$ is its fundamental charge, and A_{μ} is the 1-form electromagnetic potential for $\mathbf{F}_{\mathrm{KN}}$ related via $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. The Dirac matrices $(\tilde{\gamma}^{\mu})_{\mu=0}^{3}$ satisfy $\tilde{\gamma}^{\mu}\tilde{\gamma}^{\nu} + \tilde{\gamma}^{\nu}\tilde{\gamma}^{\mu} = 2g^{\mu\nu}$, with $g^{\mu\nu}$ the (inverse) metric coefficients of the zGKN metric.

At this point, one should recognize that if one takes the limit $a \to 0$, then (1) becomes the Minkowski metric, (3) becomes the usual Coulomb potential of a charge Q located at the origin, and (4) becomes the usual Dirac equation on Minkowski spacetime. These statements hold only formally as there are issues with the domain of the Dirac Hamiltonian when trying to make these limits precise. In particular, note that the domain of the r coordinate is $\mathbb{R}$ in zGKN while it's just $(0, \infty)$ for Minkowski spacetime. Nevertheless, the $a \to 0$ limit approach of the usual Hydrogen problem suggests a tantalizing mathematical problem.

2.2. *Separation of variables*

Using Cartan's frame method[1] with a frame well-adapted to oblate spheroidal coordinates, Chandrasekhar,[3,4] Page,[12] and Toop[16] transformed the Dirac equation (4) into an equation for a bispinor that allows a clear separation of the t, r, θ, and φ derivatives. The same transformation works for zGKN as well.[9] By introducing an

explicit diagonal matrix $\mathfrak{D} = \mathfrak{D}(r, \theta, \varphi)$ and defining $\hat{\Psi} = \mathfrak{D}^{-1}\Psi$, the Dirac equation becomes

$$(\hat{R} + \hat{A})\hat{\Psi} = 0, \tag{5}$$

where $\hat{R}$ and $\hat{A}$ are given by (with $\hbar = c = 1$ units from now on)

$$\hat{R} := \begin{pmatrix} imr & 0 & D_- - ie\mathrm{Q}\frac{r}{\varpi} & 0 \\ 0 & imr & 0 & D_+ - ie\mathrm{Q}\frac{r}{\varpi} \\ D_+ - ie\mathrm{Q}\frac{r}{\varpi} & 0 & imr & 0 \\ 0 & D_- - ie\mathrm{Q}\frac{r}{\varpi} & 0 & imr \end{pmatrix}, \tag{6}$$

$$\hat{A} := \begin{pmatrix} -ma\cos\theta & 0 & 0 & -L_- \\ 0 & -ma\cos\theta & -L_+ & 0 \\ 0 & L_- & ma\cos\theta & 0 \\ L_+ & 0 & 0 & ma\cos\theta \end{pmatrix}, \tag{7}$$

where

$$D_\pm := \pm\varpi\partial_r + \left(\varpi\partial_t + \frac{a}{\varpi}\partial_\varphi\right), \qquad L_\pm := \partial_\theta \pm i\left(a\sin\theta\partial_t + \csc\theta\partial_\varphi\right), \tag{8}$$

and

$$\varpi := \sqrt{r^2 + a^2}.$$

Once a solution $\hat{\Psi}$ to (5) is found, the bispinor $\Psi := \mathfrak{D}\hat{\Psi}$ solves the original Dirac equation (4). The explicit form of $\mathfrak{D}$ can be found in reference 9, but it's not needed for this paper.

Separation of variables is now achieved with the Ansatz that a solution $\hat{\Psi}$ of (5) is of the form

$$\hat{\Psi} = e^{-i(Et - \kappa\varphi)} \begin{pmatrix} R_1 S_1 \\ R_2 S_2 \\ R_2 S_1 \\ R_1 S_2 \end{pmatrix}, \tag{9}$$

with E a yet to be found energy eigenvalue of the Dirac hamiltonian, $\kappa \in \mathbb{Z} + \frac{1}{2}$, and with R_k being complex-valued functions of r alone, and S_k being real-valued functions of θ alone. Let

$$\vec{R} := \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}, \qquad \vec{S} := \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}. \tag{10}$$

Plugging the Chandrasekhar Ansatz (9) into (5) one easily finds that there must be $\lambda \in \mathbb{C}$ such that

$$T_{rad}\vec{R} = E\vec{R}, \tag{11}$$

$$T_{ang}\vec{S} = \lambda\vec{S}, \tag{12}$$

where

$$T_{rad} := \begin{pmatrix} d_- & m\frac{r}{\varpi} - i\frac{\lambda}{\varpi} \\ m\frac{r}{\varpi} + i\frac{\lambda}{\varpi} & -d_+ \end{pmatrix} \tag{13}$$

$$T_{ang} := \begin{pmatrix} ma\cos\theta & l_- \\ -l_+ & -ma\cos\theta \end{pmatrix} \tag{14}$$

The operators $d_\pm$ and $l_\pm$ are ordinary differential operators in r and θ respectively, with coefficients that depend on the unknown E, and parameters a, κ, and $e\mathrm{Q}$:

$$d_\pm := -i\frac{d}{dr} \pm \frac{-a\kappa + e\mathrm{Q}r}{\varpi^2} \tag{15}$$

$$l_\pm := \frac{d}{d\theta} \pm (aE\sin\theta - \kappa\csc\theta)\,. \tag{16}$$

2.3. *The coupled spectral problems for T_{rad} and T_{ang}*

The angular operator T_{ang} in (12) is easily seen to be essentially self-adjoint on $(C_c^\infty((0,\pi), \sin\theta d\theta))^2 \subset (L^2((0,\pi), \sin\theta d\theta))^2$, and its self-adjoint extension (also denoted T_{ang}) has a purely discrete spectrum $\lambda = \lambda_n(am, aE, \kappa) \in \mathbb{R}$, $n \in \mathbb{Z}\setminus 0$ (see references 2, 14).

With $\lambda \in \mathbb{R}$ it then follows that the radial operator T_{rad} is essentially self-adjoint on $(C_c^\infty(\mathbb{R}, dr))^2 \subset (L^2(\mathbb{R}, dr))^2$; its self-adjoint extension will also be denoted T_{rad}. Moreover, we can take $R_1 = R_2^*$ without loss of generality.[a] Thus, we can set

$$R_1 = \frac{1}{\sqrt{2}}(u - iv), \qquad R_2 = \frac{1}{\sqrt{2}}(u + iv) \tag{17}$$

for real funcions u and v. This brings the radial system (11) into the following standard (Hamiltonian) form

$$(H_{rad} - E)\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{18}$$

where

$$H_{rad} := \begin{pmatrix} m\frac{r}{\varpi} + \frac{\gamma r + a\kappa}{\varpi^2} & -\partial_r + \frac{\lambda}{\varpi} \\ \\ \partial_r + \frac{\lambda}{\varpi} & -m\frac{r}{\varpi} + \frac{\gamma r + a\kappa}{\varpi^2} \end{pmatrix}, \tag{19}$$

and

$$\gamma := -e\mathrm{Q} < 0. \tag{20}$$

Equation (19) should be compared with equation (7.105) in reference 15. Specifically, note that as $a \to 0$, H_{rad} approaches the radial Hamiltonian of the usual Hydrogen problem on Minkowski spacetime with a Coulomb potential.

[a] This follows by multiplying the rows of equation (11) by either R_1^* and R_2^* and adding the equations to conclude that $|R_1| = |R_2|$. Therefore $R_1 = Re^{i\phi_1}$ and $R_2 = Re^{i\phi_2}$. Again, by multiplying by conjugates, one can show that $\frac{d}{dr}(R_1/R_2^*) = 0$ which implies $\frac{d}{dr}(\phi_1 + \phi_2) = 0$.

Using techniques of Weidmann[17] it is straightforward to show that the essential spectrum of H_{rad} consists of values $E \in (-\infty, 1] \cup [1, \infty)$, and its interior is purely absolutely continuous; see reference 9. The remaining task is to characterize the discrete spectrum in the gap, i.e. the eigenvalues $E \in (-1, 1)$. In reference 9 it was shown that the spectrum is symmetric about 0, hence it suffices to consider $E > 0$.

2.4. *The Prüfer transformed system*

Following reference 9, we transform the equations (18) and (12) for the four unknowns (u, v) and (S_1, S_2) by defining four new unknowns (R, Ω) and (S, Θ) via the Prüfer transform

$$u = \sqrt{2}R\cos\frac{\Omega}{2}, \quad v = \sqrt{2}R\sin\frac{\Omega}{2}, \quad S_1 = S\cos\frac{\Theta}{2}, \quad S_2 = S\sin\frac{\Theta}{2}. \tag{21}$$

Thus

$$R = \frac{1}{2}\sqrt{u^2+v^2}, \quad \Omega = 2\tan^{-1}\frac{v}{u}, \quad S = \sqrt{S_1^2+S_2^2}, \quad \Theta = 2\tan^{-1}\frac{S_2}{S_1}. \tag{22}$$

As a result, $R_1 = Re^{-i\Omega/2}$ and $R_2 = Re^{i\Omega/2}$. Hence $\hat{\Psi}$ can be re-expressed in terms of the Prüfer variables as

$$\hat{\Psi}(t, r, \theta, \varphi) = R(r)S(\theta)e^{-i(Et-\kappa\varphi)} \begin{pmatrix} \cos(\Theta(\theta)/2)e^{-i\Omega(r)/2} \\ \sin(\Theta(\theta)/2)e^{i\Omega(r)/2} \\ \cos(\Theta(\theta)/2)e^{i\Omega(r)/2} \\ \sin(\Theta(\theta)/2)e^{-i\Omega(r)/2} \end{pmatrix}, \tag{23}$$

and we obtain the following equations for the new unknowns, first

$$\frac{d}{dr}\Omega = 2\frac{mr}{\varpi}\cos\Omega + 2\frac{\lambda}{\varpi}\sin\Omega + 2\frac{a\kappa+\gamma r}{\varpi^2} - 2E, \tag{24}$$

$$\frac{d}{dr}\ln R = \frac{mr}{\varpi}\sin\Omega - \frac{\lambda}{\varpi}\cos\Omega, \tag{25}$$

and second,

$$\frac{d}{d\theta}\Theta = -2ma\cos\theta\cos\Theta + 2\left(aE\sin\theta - \frac{\kappa}{\sin\theta}\right)\sin\Theta + 2\lambda, \tag{26}$$

$$\frac{d}{d\theta}\ln S = -ma\cos\theta\sin\Theta - \left(aE\sin\theta - \frac{\kappa}{\sin\theta}\right)\cos\Theta. \tag{27}$$

Note that the Ω-equation (24) is decoupled from R, and the Θ-equation (26) is decoupled from S. Thus the pair (24), (26) can be solved together independently of equations (25), (27), which in turn can be integrated subsequently by direct quadrature.

We can further simplify the analysis of these systems and reduce the number of parameters involved by noting that by defining the constants $a' = ma$, $E' = E/m$, and changing to the variable $r' = mr$, we eliminate m from the system. Henceforth we therefore set $m = 1$.

2.5. *Transformation onto a coupled dynamical system on cylinders*

Equations (24) and (26) exhibit the independent and the dependent variables explicitly. It is more convenient to transform them to a parametrically coupled pair of autonomous two-dimensional dynamical systems, by introducing a new independent variable τ, as follows.

Equation (26) can be written as a smooth dynamical system in the (θ, Θ) plane by introducing τ such that $\frac{d\theta}{d\tau} = \sin\theta$. Then, with dot representing differentiation in τ, we have,

$$\begin{cases} \dot{\theta} = \sin\theta \\ \dot{\Theta} = -2a\sin\theta\cos\theta\cos\Theta + 2aE\sin^2\theta\sin\Theta - 2\kappa\sin\Theta + 2\lambda\sin\theta \end{cases} \tag{28}$$

Identifying the line $\Theta = \pi$ with $\Theta = -\pi$, this becomes a dynamical system on a closed finite cylinder $\mathcal{C}_1 := [0, \pi] \times \mathbb{S}^1$. The only equilibrium points of the flow are on the two circular boundaries: Two on the left boundary: $S^- = (0, 0)$, $N^- = (0, \pi)$; two on the right: $S^+ = (\pi, -\pi)$ and $N^+ = (\pi, 0)$.

For $\kappa > 0$, the linearization of the flow at the equilibrium points reveals that S^- and S^+ are hyperbolic saddle points (with eigenvalues $\{1, -2\kappa\}$ and $\{-1, 2\kappa\}$ respectively), while N^- is a source node (with eigenvalues 1 and 2κ) and N^+ is a sink node (with eigenvalues -1 and -2κ). The situation with $\kappa < 0$ is entirely analogous, with the two critical points on each boundary switching their roles.

Similarly, the Ω equation (24) can be rewritten as a smooth dynamical system on a cylinder, in this case by setting $\tau := \frac{r}{a}$ as new independent variable, as well as introducing a new dependent variable

$$\xi := \tan^{-1}\frac{r}{a} = \tan^{-1}\tau \tag{29}$$

Then, with dot again representing differentiation in τ, (24) is equivalent to

$$\begin{cases} \dot{\xi} = \cos^2\xi \\ \dot{\Omega} = 2a\sin\xi\cos\Omega + 2\lambda\cos\xi\sin\Omega + 2\gamma\sin\xi\cos\xi + 2\kappa\cos^2\xi - 2aE \end{cases} \tag{30}$$

Once again, identifying $\Omega = -\pi$ with $\Omega = \pi$ turns this into a smooth flow on the closed finite cylinder $\mathcal{C}_2 := [-\frac{\pi}{2}, \frac{\pi}{2}] \times \mathbb{S}^1$. The only equilibrium points of the flow are on the two circular boundaries. For $E \in (0, 1)$ there are two equilibria on each: $S_E^- = (-\frac{\pi}{2}, -\pi + \cos^{-1}E)$ and $N_E^- = (-\frac{\pi}{2}, \pi - \cos^{-1}E)$ on the left boundary, and $S_E^+ = (\frac{\pi}{2}, -\cos^{-1}E)$ and $N_E^+ = (\frac{\pi}{2}, \cos^{-1}E)$ on the right boundary. $S_E^\pm$ are non-hyperbolic (degenerate) saddle-nodes, with eigenvalues 0 and $\pm 2a\sqrt{1-E^2}$, while N_E^- is a degenerate source-node and N_E^+ a degenerate sink-node (see Theorem 2.19(iii) in reference 6).

In reference 9, it was shown that E is an energy eigenvalue of the Dirac Hamiltonian and the corresponding Ψ is a bound state if and only if there exists a $\lambda \in \mathbb{R}$ such that each of the two dynamical systems above possesses a *saddles connector,* i.e. an orbit on $\mathcal{C}_1$ connecting the two saddle-nodes S^- and S^+ in the Θ-system (28) and an orbit on $\mathcal{C}_2$ connecting the two saddle-nodes S_E^- and S_E^+ in the Ω-system (30).

Given a dynamical system on a cylinder, there corresponds an integer known as the *winding number* which describes how many times an orbit in the dynamical system winds around the cylinder before terminating at an equilibrium point. See reference 9 and the upcoming paper 10. For the Ω system, saddles connectors with different winding numbers correspond to different energy values (with energy increasing as the winding number increases). In reference 9 it was shown that a bound state Ψ exists corresponding to winding number $N_\Theta = 0$ for the Θ system and winding number $N_\Omega = 0$ for the Ω system (and $\kappa = \frac{1}{2}$). In the upcoming paper 10, we improve on this result by fully classifying the spectrum. Specifically, we prove the following theorem.

Theorem 2.1. *Set* $a_{\max} = 1 - \frac{1}{\sqrt{2}}$ *and* $\gamma_{\min} = -\frac{1}{2}$. *Fix* $a \in (0, a_{\max})$, $\gamma \in (\gamma_{\min}, 0)$, *and* $\kappa \in \mathbb{Z} + \frac{1}{2}$. *Assume* Ψ *is of the form* (23) *constructed from solutions of* (24)–(27).

- *Suppose* $N_\Theta \geq 0$ *is an integer. For all integers* $N_\Omega \geq 0$, *there is a bound state* Ψ *such that the* Θ *system and* Ω *system have winding numbers* N_Θ *and* N_Ω, *respectively. There are no bound states with* $N_\Omega \leq -1$.
- *Suppose* $N_\Theta \leq -1$ *is an integer. For all integers* $N_\Omega \geq 1$, *there is a bound state* Ψ *such that the* Θ *system and* Ω *system have winding numbers* N_Θ *and* N_Ω, *respectively. There are no bound states with* $N_\Omega \leq 0$.

Our conditions on a and γ are used to ensure that there are no bound states with $N_\Omega \leq -1$ for $N_\Theta \geq 0$ and no bound states with $N_\Omega \leq 0$ for $N_\Theta \leq -1$, which, as we will see in the next section, occurs in an analogous way for the familiar hydrogenic Dirac operator on a Minkowski background with a Coulomb potential.

When restoring units and mass m, we have $a_{\max} = (1 - \frac{1}{\sqrt{2}})\frac{\hbar}{mc}$ and $\gamma_{\min} = -\frac{1}{2}\hbar c$. If, with the hydrogenic problem in mind, we set $\gamma = -Ze^2$, then $\gamma \in (\gamma_{\min}, 0)$ implies $\frac{Ze^2}{\hbar c} < \frac{1}{2}$, that is, $Z < \frac{137.036}{2}$. Now let's compare these conditions with the conditions on Z for the familiar hydrogenic Dirac operator on a Minkowski background with a Coulomb potential, for which essential self-adjointness breaks down for $Z > 118$. For the Dirac operator on zGKN, there is no condition for essential self-adjointness.[9] So our conditions are probably not optimal.

3. Relating the zGKN bound states to the usual hydrogenic states

In this section we relate the zGKN bound states found in Theorem 2.1 to the usual hydrogenic states of the Dirac problem on Minkowski spacetime with a Coulomb potential. The main objective is to relate the winding numbers which appear in Theorem 2.1 to the usual spectroscopic notation $n\ell_j$ of hydrogenic states.

3.1. *The correspondence between winding numbers and the usual spectroscopic notation of hydrogenic states*

Based on the above results, for fixed a and γ, the discrete spectrum of our Dirac Hamiltonian is indexed by *three* integers: N_Ω, N_Θ, and 2κ. By contrast, the energy spectrum of special relativistic Hydrogen, i.e., the Dirac Hamiltonian for a point-like electron in ordinary Minkowski space interacting with a Coulomb point charge at the origin, is indexed by *two* integers only, namely the main (or Bohr's) quantum number, often denoted by n, and the spin-orbit quantum number[b], i.e., the set of eigenvalues of the spin-orbit operator $K = \beta(2\mathbf{S}\cdot\mathbf{L} + 1)$.

In the limit $a \to 0$, the angular Hamiltonian (14), takes the simple form

$$\mathfrak{a}_\kappa := \lim_{a\to 0} T_{ang} = i\sigma_2\partial_\theta + \frac{\kappa}{\sin\theta}\sigma_1. \tag{31}$$

From reference 2, if λ is an eigenvalue of T_{ang}, then $k := \lim_{a\to 0}\lambda$ is an eigenvalue of $\mathfrak{a}_\kappa$, also the limit defining k exists since λ is analytic in a. Note that $\mathfrak{a}_\kappa$ is independent of E unlike T_{ang}.

In the limit $a \to 0$, the formal limit of the radial Hamiltonian (19) coincides with the radial Hamiltonian arising in the special relativistic Hydrogen problem (e.g. see reference 15, eq. (7.105)):

$$\mathfrak{h}_k := \lim_{a\to 0} H_{rad} = \begin{pmatrix} m + \dfrac{\gamma}{r} & -\partial_r + \dfrac{k}{r} \\ \partial_r + \dfrac{k}{r} & -m + \dfrac{\gamma}{r} \end{pmatrix}. \tag{32}$$

Therefore k can be identified with the *spin-orbit coupling*.

The spectrum of $\mathfrak{a}_\kappa$ is completely understood.[2] In particular, for all half-integers $\kappa \in Z + \frac{1}{2}$, the operator $\mathfrak{a}_\kappa$ is essentially self-adjoint and has a discrete spectrum indexed by a nonzero integer which we call N:

$$k = -\mathrm{sgn}(N)\left(|N| + |\kappa| - \frac{1}{2}\right), \tag{33}$$

as well as a complete set of eigenvectors $\vec{S}_{N,\kappa}$ that are explicitly known and can be expressed in terms of Jacobi polynomials.

$$\vec{S}_{N,\kappa}(\theta) := \sin^{\kappa+\frac{1}{2}}\theta \begin{pmatrix} -\sqrt{\cot\frac{\theta}{2}}\,P^{\kappa-\frac{1}{2},\kappa+\frac{1}{2}}_{|N|-1}(\cos\theta) \\ \mathrm{sgn}(N)\sqrt{\tan\frac{\theta}{2}}\,P^{\kappa+\frac{1}{2},\kappa-\frac{1}{2}}_{|N|-1}(\cos\theta) \end{pmatrix}. \tag{34}$$

Note that the above holds for $\kappa > 0$. To find the eigenvectors for $\kappa < 0$, recognize that if $\vec{S}$ is an eigenvector of $\mathfrak{a}_\kappa$, then $i\sigma_2\vec{S}$ is an eigenvector of $\mathfrak{a}_{-\kappa}$.

[b]The spin-orbit quantum number is called κ_j in reference;[15] it should not be confused with our κ, which is the eigenvalue of the z-component of angular momentum, for which Thaller uses the notation m_j.

From the above and the definition of Θ it follows that the saddles connectors of the Θ-system that correspond to the eigenvectors of the angular Hamiltonian are given explicitly by the formula

$$\Theta_{N,\kappa}(\theta) = -\mathrm{sgn}(N)\left\{2\tan^{-1}\left(\frac{P_{|N|-1}^{|\kappa|+\frac{1}{2},|\kappa|-\frac{1}{2}}(\cos\theta)}{P_{|N|-1}^{|\kappa|-\frac{1}{2},|\kappa|+\frac{1}{2}}(\cos\theta)}\tan\frac{\theta}{2}\right) - \pi\boldsymbol{\chi}_{\kappa<0}\right\}. \tag{35}$$

Here the branch of $\tan^{-1}$ needs to be chosen in such a way that $\Theta_{N,\kappa}$ is continuous on $[0,\pi]$.

The above formula implies that we can fix the initial value of the saddles connectors to be

$$\Theta_{N,\kappa}(0) = \begin{cases} 0 & \kappa > 0 \\ \mathrm{sgn}(N)\pi & \kappa < 0. \end{cases} \tag{36}$$

and that for the final value of those connectors, we need

$$\Theta_{N,\kappa}(\pi) = \begin{cases} -\mathrm{sgn}(N)\pi & \kappa > 0 \\ 0 & \kappa < 0 \end{cases} \qquad [\mathrm{mod} 2\pi] \tag{37}$$

in agreement with what we have already observed about the boundary values of the saddles connectors for the Θ-system (28).

Furthermore, from the properties of Jacobi polynomials, it follows that for $\Theta_{N,\kappa}$ to be a continuous function of θ on $[0,\pi]$, we need

$$\Theta_{N,\kappa}(\pi) = -2\pi\mathrm{sgn}(N)(|N|-1) + \begin{cases} -\mathrm{sgn}(N)\pi & \kappa > 0 \\ 0 & \kappa < 0 \end{cases} \tag{38}$$

Thus we establish a correspondence between the integer N and the winding number N_Θ of the Θ-saddles connectors in the case $a = 0$.

$$N_\Theta \sim \begin{cases} N-1 & N \geq 1 \\ N & N \leq -1, \end{cases} \tag{39}$$

where '$\sim$' simply means a correspondence.

From the relationship (33) between the eigenvalue k and the number N it follows that instead of N we can equally well label the angular eigenstates by the integer k, in which case

$$N \;=\; -\mathrm{sgn}(k)\left(|k| - |\kappa| + \frac{1}{2}\right). \tag{40}$$

We now turn our attention to the radial Hamiltonian $\mathfrak{h}_k = \lim_{a\to 0} H_{rad}$. This operator is closely related to the radial Hamiltonian h_k of the special-relativistic Hydrogen problem, as formulated by Dirac:[5]

$$h_k = -i\sigma_2\partial_r + m\sigma_3 + \frac{k}{r}\sigma_1 + \frac{\gamma}{r}I_{2\times 2}. \tag{41}$$

The only difference between h_k and $\mathfrak{h}_k$ is their domains: Since H_{rad} is defined on the double-sheeted Sommerfeld space, this is inherited by its $a \to 0$ limit $\mathfrak{h}_k$, which

is still defined on two copies of Minkowski space glued together along a timelike line. In particular the r variable in $\mathfrak{h}_k$ has the range $(-\infty, \infty)$. By contrast, the r variable in h_k goes from 0 to ∞. The eigenvalue problem for h_k was shown to be exactly solvable by Gordon,[7] who proved the discrete spectrum to coincide exactly with the Bohr-Sommerfeld spectrum, and found explicit formulas for the eigenfunctions in terms of generalized Laguerre polynomials.

Consider first the restriction of $\mathfrak{h}_k$ to functions supported in the sheet $r > 0$. As described in detail in section 7.4 of Thaller's book,[15] this operator, for all $k \in \mathbb{Z}\backslash\{0\}$, is essentially self-adjoint on $C_c^\infty((0,\infty))$, has a discrete spectrum in $(0,1)$ and a complete set of eigenvectors, if $-\sqrt{3}/2 < \gamma < 0$. The discrete spectrum is indexed by two integers, $n \geq 1$ and $k = -n, \ldots, -1, 1, \ldots, n-1$. Let $M := n - |k|$. Then we have

$$E_{M,k} \;=\; \frac{m}{\sqrt{1 + \left(\frac{\gamma^2}{M + \sqrt{k^2 - \gamma^2}}\right)^2}} \tag{42}$$

Note that the case $k > 0$ and $M = 0$ is excluded.

Gordon[7] computed the corresponding eigenfunctions in terms of generalized Laguerre polynomials: Let (ϕ_1, ϕ_2) be defined by

$$u = \sqrt{1 + E_{M,k}}(\phi_1 + \phi_2), \qquad v = \sqrt{1 - E_{M,k}}(\phi_1 - \phi_2) \tag{43}$$

where (u, v) is an eigenfunction of $\mathfrak{h}_k$ with eigenvalue $E_{M,k}$ (and we have set $m = 1$). Then, using the abbreviations

$$\rho := \sqrt{k^2 - \gamma^2}, \qquad \eta := \sqrt{1 - E_{M,k}^2} \tag{44}$$

we have that for all nonnegative integers M and real constants c_1, c_2 such that

$$\mu := \frac{c_1}{c_2} = \frac{M}{k + \frac{\gamma}{\eta}}. \tag{45}$$

we have

$$\phi_1(r) \;=\; c_1 e^{-\eta r} r^\rho F(-M+1, 2\rho+1, 2\eta r) \tag{46}$$

$$\phi_2(r) \;=\; c_2 e^{-\eta r} r^\rho F(-M, 2\rho+1, 2\eta r) \tag{47}$$

where F denotes Gauss's *confluent hypergeometric function*

$$F(\alpha, \beta, x) \;=\; 1 + \frac{\alpha}{1!\beta}x + \frac{\alpha(\alpha+1)}{2!\beta(\beta+1)}x^2 + \ldots \tag{48}$$

Note that when α is a negative integer, the above series terminates, and F is a polynomial of degree $-\alpha$, which (up to a numerical factor) is the generalized Laguerre polynomial $L_{-\alpha}^{(\beta-1)}(x)$.

Accordingly, since $\Omega = 2\tan^{-1}\frac{v}{u}$, the corresponding solution to the $a \to 0$ limit of the Ω-equation (24) will be

$$\Omega(r) = 2\tan^{-1}\left(\sqrt{\frac{1-E}{1+E}}\;\frac{\mu F(-M+1, 2\rho+1, 2\eta r) - F(-M, 2\rho+1, 2\eta r)}{\mu F(-M+1, 2\rho+1, 2\eta r) + F(-M, 2\rho+1, 2\eta r)}\right) \tag{49}$$

The values of Ω at $r = 0$ and $r = \infty$ can thus be calculated (modulo 2π):

$$\Omega(0) \;=\; 2\tan^{-1}\left(\sqrt{\frac{1-E}{1+E}}\;\frac{\mu-1}{\mu+1}\right) = \sin^{-1}(\frac{-\gamma}{k}) \tag{50}$$

$$\Omega(\infty) \;=\; -2\tan^{-1}\left(\sqrt{\frac{1-E}{1+E}}\right) \;=\; -\cos^{-1}E \tag{51}$$

which is in agreement with the analysis of the equilibrium points of the corresponding dynamical system. We choose our principal branch of $\sin^{-1}$ depending on the sign of k:

$$\Omega(0) \;=\; \begin{cases} \sin^{-1}(\frac{-\gamma}{k}) & \text{if } k < 0. \\ -\pi - \sin^{-1}(\frac{-\gamma}{k}) & \text{if } k > 0 \end{cases} \tag{52}$$

On the other hand, from the properties of Laguerre polynomials, it follows that the denominator of the rational function in (49) has M zeros in $[0, \infty)$ when $k < 0$ and $M - 1$ zeros when $k > 0$ (note that this explains why $k \leq n - 1$). Thus in order for Ω to be a continuous function of r, the branch of $\tan^{-1}$ needs to be chosen such that -2π gets added to the value every time r crosses one of those poles, which implies that

$$\Omega(\infty) \;=\; -2\pi M - \cos^{-1}E \tag{53}$$

which holds for both $k < 0$ and $k > 0$ (this is why we chose $\Omega(0) = -\pi - \sin^{-1}(\frac{-\gamma}{k})$ for $k > 0$). Thus we establish the following correspondence between the integer M and the winding number N_Ω for the Ω-saddles connectors in the case $a = 0$.

$$N_\Omega \;\sim\; M. \tag{54}$$

Now consider the case $a > 0$ (i.e. the Dirac Hamiltonian on zGKN). The bound states are indexed by three integers N_Θ, N_Ω, and 2κ appearing in Theorem 2.1. Using (54) and (39), we can define a correspondence between the integers $N_\Theta, N_\Omega, 2\kappa$ and the usual spectroscopic notation $n\ell_j$ of hydrogenic states. Given N_Θ and N_Ω, define N and M via

$$N \;:=\; \begin{cases} N_\Theta + 1 & N_\Theta \geq 0 \\ N_\Theta & N_\Theta \leq -1, \end{cases} \qquad \text{and} \qquad M \;:=\; N_\Omega. \tag{55}$$

Then n, ℓ, and j appearing in $n\ell_j$ are given by

$$k \;:=\; -N - \operatorname{sgn}(N)\big(|\kappa| - \frac{1}{2}\big) \tag{56}$$

$$n \;:=\; M + |k| \tag{57}$$

$$j \;:=\; |k| - \frac{1}{2} \tag{58}$$

$$\ell \;:=\; j + \operatorname{sgn}(k)\frac{1}{2} \tag{59}$$

$$m_j \;:=\; \kappa. \tag{60}$$

This establishes the desired correspondence between the winding numbers appearing in Theorem 2.1 and the usual spectroscopic notation $n\ell_j$ of hydrogenic states. For example, suppose we ask: which state in zGKN corresponds to $2p_{1/2}$ with $m_{1/2} = -\frac{1}{2}$? Well $\ell = 1$ implies $1 = \frac{1}{2} + \mathrm{sgn}(k)\frac{1}{2}$. Therefore k is positive and so $j = \frac{1}{2}$ implies $k = 1$. Therefore $N = -1$ implies $N_\Theta = -1$. Lastly, $n = 2$ implies $M = 1$ and so $N_\Omega = 1$. Thus the state $2p_{1/2}$ with $m_{1/2} = -\frac{1}{2}$ corresponds to $N_\Theta = -1$, $N_\Omega = 1$, and $\kappa = -\frac{1}{2}$.

3.2. *Breaking degeneracies*

It is well known that the energy levels of the special-relativistic Hydrogen Hamiltonian are independent of m_j, the eigenvalue of the z-component of total angular momentum, while the eigenstates of it do depend on m_j. As a result, all energy levels are degenerate in that case, i.e. have multiplicity at least two. In the case of our Hamiltonian H_{rad}, this degeneracy is broken by the appearance of $a\kappa$ in equation (19). This yields a hyperfine-like splitting of the spectral lines, which in the standard setting is thought of as being a consequence of the nucleus having a magnetic dipole moment. In our setting, it can be thought of as following from the nonzero $d\varphi$ term in equation (3) which produces a magnetic dipole moment.

In addition to the above symmetry-related degeneracy, Dirac's original model for special-relativistic Hydrogen also has *accidental* degeneracies; the energy eigenvalues do not depend on the *sign* of the spin-orbit quantum number k, only on its magnitude, while the eigenstates do depend on the sign of k. The celebrated experiment of Lamb[11] showed that this degeneracy is only in Dirac's model, not in nature, i.e., the measured energy levels of orbitals corresponding to k and $-k$ are slightly different. This difference is known as the *Lamb shift*. In our model, these accidental degeneracies are also broken producing a small but observable Lamb shift-like effect.

Recall that hyperfine splitting and the Lamb shift are calculated using perturbative techniques. It is important to note that our model does not use perturbation theory to calculate energy differences.

In the upcoming paper,[10] we will report on numerical results of the zGKN spectrum showing the breaking of these degeneracies for various values of a.

References

1. D. R. Brill and J. M. Cohen, *Cartan frames and the general relativistic Dirac equation*, J. Math. Phys. **7** (1966).
2. D. Batic, H. Schmid, and M. Winklmeier, *On the eigenvalues of the Chandrasekhar-Page angular equation*, J. Math. Phys. **46** (2005).
3. S. Chandrasekhar, *The solution of Dirac's equation in Kerr geometry*. Proc. Roy. Soc. London Ser. A **349** (1976).
4. S. Chandrasekhar, *Errata: The solution of Dirac's equation in Kerr geometry*. Proc. Roy. Soc. London Ser. A **350** (1976).
5. P. A. M. Dirac, *The quantum theory of the electron I*. Proc. Roy. Soc. London Ser. A **117** (1928).

6. F. Dumortier, J. Llibre, and J. C. Artés, *Qualitative theory of planar differential systems*, Universitext, Springer-Verlag, Berlin, (2006).
7. W. Gordon, *Die Energieniveaus der Wassenstoffatoms nach der Diracschen Quantentheorie des Elektrons*, Z.
8. M.K.-H. Kiessling and A. S. Tahvildar-Zadeh, *A novel quantum-mechanical interpretation of the Dirac equation*, J. Phys. A. **49** (2016).
9. M.K.-H. Kiessling and A. S. Tahvildar-Zadeh, *The Dirac point electron in zero-gravity Kerr-Newman spacetime*, J. Math. Phys. **56** (2015).
10. M.K.-H. Kiessling, E. Ling, and A. S. Tahvildar-Zadeh, *On the discrete Dirac spectrum of a point electron in the zero-gravity Kerr-Newman spacetime*, in preparation.
11. W. E. Lamb, *Fine structure of the hydrogen atom*, Nobel Lecture. (1955)
12. D. Page, *Dirac equation around a charged, rotation black hole*, Phys. Rev. D **14** (1976).
13. A. S. Tahvildar-Zadeh, *On a zero-gravity limit of the Kerr-Newman spacetimes and their electromagnetic fields*, J. Math. Phys. **56** (2015).
14. K. G. Suffern, E. D. Fackerell, and C.M. Cosgrove, *Eigenvalues of the Chanrasekhar-Page angular functions*, J. Math. Phys. **24** (1983).
15. B. Thaller, *The Dirac Equation*. Springer-Verlag, Berlin, (1992).
16. N. Toop, *The Thermal Radiation of Electrons from a Charged Spinning Black Hole in a Cosmological background* (DAMTP, Cambridge, 1976).
17. J. Weidmann, *Absolut stetiges Spektrum bei Sturm-Liouville-Operatoren und Systemen*. Math. Z. **180** (1982).
18. D. M. Zipoy, *Topology of Some Spheroidal Metrics*. J. Math. Phys. **7** (1996).

Causal fermion systems: Classical gravity and beyond

Felix Finster
Fakultät für Mathematik, Universität Regensburg, D-93040 Regensburg, Germany
E-mail: finster@ur.de
www.mathematik.ur.de/Mat1

In this short review, we explain how and in which sense the causal action principle for causal fermion systems gives rise to classical gravity and the Einstein equations. Moreover, methods are presented for going beyond classical gravity, with applications to a positive mass theorem for static causal fermion systems, a connection between area change and matter flux and the construction of a quantum state.

Keywords: Causal fermion system, causal action principle, globally hyperbolic spacetimes, Einstein equations, surface layer integrals.

1. Introduction

The theory of *causal fermion systems* is a recent approach to fundamental physics (see a few basics in Section 2, the reviews,[8,10] the textbooks[7,14] or the website[1]). In this approach, spacetime and all objects therein are described by a measure ρ on a set $\mathcal{F}$ of linear operators on a Hilbert space $(\mathcal{H}, \langle .|. \rangle_{\mathcal{H}})$. The physical equations are formulated via the so-called *causal action principle*, a nonlinear variational principle where an action $\mathcal{S}$ is minimized under variations of the measure ρ.

The purpose of this survey article is to report on the present status of the theory with regards to the gravitational interaction. It has been shown that, in a well-defined limiting case, the so-called *continuum limit*, the causal action principles gives classical gravity and the Einstein equations, with the gravitational coupling constant determined by the regularization scale. Without taking the continuum limit, the causal action principle describes a novel gravitational theory which is presently under investigation. We here give a survey on different approaches towards unraveling the structure and the properties of this gravitational theory.

The paper is structured as follows. After giving the necessary background on causal fermion systems (Section 2), we explain how and in which sense classical gravity and the Einstein equations are obtained in the continuum limit (Section 3). In Section 4 we report on the approaches for going beyond classical gravity. After a few comments on the intrinsic geometric structures of a causal fermion system (Section 4.1), we introduce surface layer integrals as the main object for the subsequent analysis (Section 4.2). We finally report on a positive mass theorem for static causal fermion systems (Section 4.3), a relation between area change and matter flux (Section 4.4) and the construction of a quantum state (Section 4.5).

2. A Brief Introduction to Causal Fermion Systems

This section provides the necessary abstract background on causal fermion systems.

2.1. *Causal Fermion Systems and the Causal Action Principle*

We begin with the general definitions.

Definition 2.1. (causal fermion systems) *Given a separable complex Hilbert space $\mathcal{H}$ with scalar product $\langle .|. \rangle_{\mathcal{H}}$ and a parameter $n \in \mathbb{N}$ (the* spin dimension*), we let $\mathcal{F} \subset L(\mathcal{H})$ be the set of all symmetric operators on $\mathcal{H}$ of finite rank, which (counting multiplicities) have at most n positive and at most n negative eigenvalues. On $\mathcal{F}$ we are given a positive measure ρ (defined on a σ-algebra of subsets of $\mathcal{F}$). We refer to $(\mathcal{H}, \mathcal{F}, \rho)$ as a* causal fermion system*.*

A causal fermion system describes a spacetime together with all structures and objects therein. In order to single out the physically admissible causal fermion systems, one must formulate physical equations. To this end, we impose that the measure ρ should be a minimizer of the causal action principle, which we now introduce. For any $x, y \in \mathcal{F}$, the product xy is an operator of rank at most $2n$. However, in general it is no longer symmetric because $(xy)^* = yx$, and this is different from xy unless x and y commute. As a consequence, the eigenvalues of the operator xy are in general complex. We denote the non-trivial eigenvalues counting algebraic multiplicities by $\lambda_1^{xy}, \ldots, \lambda_{2n}^{xy} \in \mathbb{C}$ (more specifically, denoting the rank of xy by $k \leq 2n$, we choose $\lambda_1^{xy}, \ldots, \lambda_k^{xy}$ as all the non-zero eigenvalues and set $\lambda_{k+1}^{xy}, \ldots, \lambda_{2n}^{xy} = 0$). We introduce the Lagrangian and the causal action by

$$\textit{Lagrangian:} \qquad \mathcal{L}(x,y) = \frac{1}{4n} \sum_{i,j=1}^{2n} \left(\left|\lambda_i^{xy}\right| - \left|\lambda_j^{xy}\right| \right)^2 \tag{1}$$

$$\textit{causal action:} \qquad \mathcal{S}(\rho) = \iint_{\mathcal{F}\times\mathcal{F}} \mathcal{L}(x,y)\, d\rho(x)\, d\rho(y)\,. \tag{2}$$

The *causal action principle* is to minimize $\mathcal{S}$ by varying the measure ρ under the following constraints,

$$\textit{volume constraint:} \qquad \rho(\mathcal{F}) = \text{const} \tag{3}$$

$$\textit{trace constraint:} \qquad \int_{\mathcal{F}} \mathrm{tr}(x)\, d\rho(x) = \text{const} \tag{4}$$

$$\textit{boundedness constraint:} \qquad \iint_{\mathcal{F}\times\mathcal{F}} |xy|^2\, d\rho(x)\, d\rho(y) \leq C\,, \tag{5}$$

where C is a given parameter, tr denotes the trace of a linear operator on $\mathcal{H}$, and the absolute value of xy is the so-called spectral weight,

$$|xy| := \sum_{j=1}^{2n} \left|\lambda_j^{xy}\right|.$$

This variational principle is mathematically well-posed if $\mathcal{H}$ is finite-dimensional. For the existence theory and the analysis of general properties of minimizing measures we refer to[3,5] or [14, Chapter 12]. In the existence theory one varies in the class of regular Borel measures (with respect to the topology on $\mathrm{L}(\mathcal{H})$ induced by the operator norm), and the minimizing measure is again in this class. With this in mind, here we always assume that

$$\rho \text{ is a regular Borel measure}\,.$$

2.2. *Spacetime and Causal Structure*

Let ρ be a *minimizing* measure. *Spacetime* is defined as the support of this measure,

$$M := \operatorname{supp} \rho \;\subset\; \mathcal{F}\,.$$

Thus the spacetime points are symmetric linear operators on $\mathcal{H}$. On M we consider the topology induced by $\mathcal{F}$ (generated by the operator norm on $\mathrm{L}(\mathcal{H})$). Moreover, the measure $\rho|_M$ restricted to M gives a volume measure on spacetime. This gives spacetime the structure of a *topological measure space*.

The operators in M contain a lot of information which, if interpreted correctly, gives rise to spacetime structures like causal and metric structures, spinors and interacting fields (for details see [7, Chapter 1]). All the resulting objects are *inherent* in the sense that we only use information already encoded in the causal fermion system. Here we restrict attention to those structures needed in what follows. We begin with the following notion of causality:

Definition 2.2. (causal structure) *For any $x, y \in \mathcal{F}$, we again denote the non-trivial eigenvalues of the operator product xy (again counting algebraic multiplicities) by $\lambda_1^{xy}, \ldots, \lambda_{2n}^{xy}$. The points x and y are called* spacelike *separated if all the λ_j^{xy} have the same absolute value. They are said to be* timelike *separated if the λ_j^{xy} are all real and do not all have the same absolute value. In all other cases (i.e. if the λ_j^{xy} are not all real and do not all have the same absolute value), the points x and y are said to be* lightlike *separated.*

Restricting the causal structure of $\mathcal{F}$ to M, we get causal relations in spacetime.

The Lagrangian (1) is compatible with the above notion of causality in the following sense. Suppose that two points $x, y \in M$ are spacelike separated. Then the eigenvalues λ_i^{xy} all have the same absolute value. As a consequence, the Lagrangian (1) vanishes. Thus pairs of points with spacelike separation do not enter the action. This can be seen in analogy to the usual notion of causality where points with spacelike separation cannot influence each other. This is the reason for the notion "causal" in *causal* fermion system and *causal* action principle.

Moreover, a causal fermion system distinguishes a *direction of time*. To this end, we let π_x be the orthogonal projection in $\mathcal{H}$ on the subspace $x(\mathcal{H}) \subset \mathcal{H}$ and introduce the functional

$$\mathcal{C} \;:\; M \times M \to \mathbb{R}\,, \qquad \mathcal{C}(x,y) := i \operatorname{tr}\big(y\, x\, \pi_y\, \pi_x - x\, y\, \pi_x\, \pi_y\big)\,.$$

Obviously, this functional is anti-symmetric in its two arguments, making it possible to introduce the notions

$$\begin{cases} y \text{ lies in the } \textit{future} \text{ of } x & \text{if } \mathcal{C}(x,y) > 0 \\ y \text{ lies in the } \textit{past} \text{ of } x & \text{if } \mathcal{C}(x,y) < 0\,. \end{cases}$$

2.3. *The Euler-Lagrange Equations*

A minimizer of a causal variational principle satisfies the following *Euler-Lagrange (EL) equations*. For a suitable value of the parameter $\mathfrak{s} > 0$, the function $\ell : \mathcal{F} \to \mathbb{R}_0^+$ defined by

$$\ell(x) := \int_M \mathcal{L}_\kappa(x,y)\, d\rho(y) - \mathfrak{s} \tag{6}$$

is minimal and vanishes on spacetime $M := \operatorname{supp} \rho$,

$$\ell|_M \equiv \inf_{\mathcal{F}} \ell = 0\,. \tag{7}$$

Here the *κ-Lagrangian* $\mathcal{L}_\kappa$ is defined by

$$\mathcal{L}_\kappa \;:\; \mathcal{F} \times \mathcal{F} \to \mathbb{R}\,, \qquad \mathcal{L}_\kappa(x,y) := \mathcal{L}(x,y) + \kappa\, |xy|^2 \tag{8}$$

with a non-negative parameter κ, which can be thought of as the Lagrange parameter corresponding to the boundedness constraint. Likewise, the parameter $\mathfrak{s} \geq 0$ in (6) is the Lagrange parameter corresponding to the volume constraint. For the derivation and further details we refer to [16, Section 2] or [14, Chapter 7].

2.4. *Spinors and Physical Wave Functions*

A causal fermion system also gives rise to spinorial wave functions in spacetime, as we now explain. For every $x \in \mathcal{F}$ we define the *spin space* S_x by $S_x = x(\mathcal{H})$; it is a subspace of $\mathcal{H}$ of dimension at most $2n$. It is endowed with the *spin inner product* $\prec .|. \succ_x$ defined by

$$\prec u|v\succ_x = -\langle u|xv\rangle_{\mathcal{H}} \qquad (\text{for all } u, v \in S_x)\,. \tag{9}$$

A *wave function* ψ is defined as a function which to every $x \in M$ associates a vector of the corresponding spin space,

$$\psi \;:\; M \to \mathcal{H} \qquad \text{with} \qquad \psi(x) \in S_x M \quad \text{for all } x \in M\,.$$

It is an important observation that every vector $u \in \mathcal{H}$ of the Hilbert space gives rise to a unique wave function denoted by ψ^u. It is obtained by projecting the vector u to the corresponding spin spaces,

$$\psi^u \;:\; M \to \mathcal{H}\,, \qquad \psi^u(x) := \pi_x u \in S_x M\,. \tag{10}$$

We refer to ψ^u as the *physical wave function* of the vector $u \in \mathcal{H}$. Choosing an orthonormal basis (e_i) of $\mathcal{H}$, we obtain a whole family of physical wave functions (ψ^{e_i}).

This ensemble of wave functions is crucial for the understanding of what a causal fermion systems is about. In fact, all spacetime structures (like for example the causal structure in Definition 2.2) can be recovered from this ensemble. Moreover, one can construct concrete examples of causal fermion systems by choosing the physical wave functions more specifically as the quantum mechanical wave functions in a classical Lorentzian spacetime. In the next section we explain this construction in more detail.

3. The Limiting Case of Classical Gravity

In this section we outline how and in which sense the causal action principle gives rise to classical gravity. For more details we refer to[7] and the review article.[8]

3.1. *Describing a Lorentzian Spacetime by a Causal Fermion System*

We now explain how a classical curved spacetime is described by a causal fermion system. Our starting point is Lorentzian spin geometry. Thus we let $(\mathscr{M}, g)$ be a smooth, globally hyperbolic, time-oriented Lorentzian spin manifold of dimension four. For the signature of the metric we use the convention $(+,-,-,-)$. We denote the corresponding spinor bundle by $S\mathscr{M}$. Its fibers $S_x\mathscr{M}$ (with $x \in \mathscr{M}$) are endowed with an inner product $\prec .|. \succ_x$ of signature $(2,2)$. Clifford multiplication is described by a mapping γ which satisfies the anti-commutation relations,

$$\gamma \,:\, T_x\mathscr{M} \to \mathrm{L}(S_x\mathscr{M}) \qquad \text{with} \qquad \gamma(u)\,\gamma(v) + \gamma(v)\,\gamma(u) = 2\,g(u,v)\,\mathbb{1}_{S_x(\mathscr{M})}\,.$$

We also write Clifford multiplication in components with the Dirac matrices γ^j. The metric connections on the tangent bundle and the spinor bundle are denoted by ∇. The sections of the spinor bundle are also referred to as wave functions.

We denote the smooth sections of the spinor bundle by $C^\infty(\mathscr{M}, S\mathscr{M})$. The Dirac operator $\mathcal{D}$ is defined by

$$\mathcal{D} := i\gamma^j \nabla_j \,:\, C^\infty(\mathscr{M}, S\mathscr{M}) \to C^\infty(\mathscr{M}, S\mathscr{M})\,.$$

Given a real parameter $m \in \mathbb{R}$ (the *mass*), the Dirac equation reads

$$(\mathcal{D} - m)\,\psi = 0\,.$$

We mainly consider solutions in the class $C^\infty_{\mathrm{sc}}(\mathscr{M}, S\mathscr{M})$ of smooth sections with spatially compact support (i.e. wave functions whose restriction to any Cauchy surface is compact). On such solutions, one has the scalar product

$$(\psi|\phi)_m = 2\pi \int_{\mathscr{N}} \prec \psi \,|\, \gamma(\nu)\,\phi \succ_x d\mu_{\mathscr{N}}(x)\,,$$

where $\mathscr{N}$ denotes any Cauchy surface and ν its future-directed normal (due to current conservation, the scalar product is in fact independent of the choice

of $\mathcal{N}$; for details see [20, Section 2]). Forming the completion gives the Hilbert space $(\mathcal{H}_m, (.|.)_m)$.

Next, we choose a closed subspace $\mathcal{H} \subset \mathcal{H}_m$ of the solution space of the Dirac equation. The induced scalar product on $\mathcal{H}$ is denoted by $\langle .|. \rangle_{\mathcal{H}}$. There is the technical difficulty that the wave functions in $\mathcal{H}$ are in general not continuous, making it impossible to evaluate them pointwise. For this reason, we need to introduce an *ultraviolet regularization* on the length scale ε, described mathematically by a linear

$$\textit{regularization operator} \qquad \mathfrak{R}_\varepsilon \;:\; \mathcal{H} \to C^0(\mathcal{M}, S\mathcal{M})\,.$$

In the simplest case, the regularization can be realized by a convolution on a Cauchy surface or in spacetime (for details see [20, Section 4] or [7, Section §1.1.2]). For us, the regularization is not merely a technical tool, but it realizes the concept that we want to change the geometric structures on the microscopic scale. With this in mind, we always consider the regularized quantities as those having mathematical and physical significance. Different choices of regularization operators realize different microscopic spacetime structures.

Evaluating the regularization operator at a spacetime point $x \in \mathcal{M}$ gives the *regularized wave evaluation operator* $\Psi^\varepsilon(x)$,

$$\Psi^\varepsilon(x) = \mathfrak{R}_\varepsilon(x) \;:\; \mathcal{H} \to S_x\mathcal{M}\,. \tag{11}$$

We also take its adjoint (with respect to the Hilbert space scalar product $\langle .|. \rangle_{\mathcal{H}}$ and the spin inner product $\prec .|. \succ_x$),

$$\big(\Psi^\varepsilon(x)\big)^* \;:\; S_x\mathcal{M} \to \mathcal{H}\,.$$

Multiplying $\Psi^\varepsilon(x)$ by its adjoint gives the operator

$$F^\varepsilon(x) := -\big(\Psi^\varepsilon(x)\big)^*\, \Psi^\varepsilon(x) \;:\; \mathcal{H} \to \mathcal{H}\,, \tag{12}$$

referred to as the *local correlation operator* at the spacetime point x. The local correlation operator is also characterized by the relation

$$(\psi \,|\, F^\varepsilon(x)\, \phi) = -\prec (\mathfrak{R}_\varepsilon \psi)(x) | (\mathfrak{R}_\varepsilon \phi)(x) \succ_x \qquad \text{for all } \psi, \phi \in \mathcal{H}\,. \tag{13}$$

Taking into account that the inner product on the Dirac spinors at x has signature $(2,2)$, it is a symmetric operator on $\mathcal{H}$ of rank at most four, which (counting multiplicities) has at most two positive and at most two negative eigenvalues. Varying the spacetime point, we obtain a mapping

$$F^\varepsilon \;:\; \mathcal{M} \to \mathcal{F} \subset \mathrm{L}(\mathcal{H})\,,$$

where $\mathcal{F}$ denotes all symmetric operators of rank at most four with at most two positive and at most two negative eigenvalues. Finally, we introduce the measure ρ on $\mathcal{F}$ by taking the push-forward of the volume measure on $\mathcal{M}$ under the mapping F^ε,

$$\rho := (F^\varepsilon)_* \mu_{\mathcal{M}} \tag{14}$$

(thus $\rho(\Omega) := \mu_{\mathcal{M}}((F^\varepsilon)^{-1}(\Omega))$). The resulting structure $(\mathcal{H}, \mathcal{F}, \rho)$ is a causal fermion system of spin dimension two.

We conclude with a few comments on the significance of this construction. We first point out that the construction uses all the structures of Lorentzian spin geometry as well as the properties of the Dirac wave functions in $\mathcal{H}$. We thus obtain a *very specific class of examples* of causal fermion systems describing classical spacetimes. In general, the measure ρ defined by (14) will not be a minimizer of the causal action principle. But it is an approximate minimizer if the Einstein equations are satisfied. Before making this point precise in the next section (Section 3.2), we now explain the physical picture behind the Dirac wave functions in $\mathcal{H}$. These wave functions have the interpretation as being those Dirac wave functions which are realized in the physical system under consideration. If we describe for example a system of one electron, then the wave function of the electron is contained in $\mathcal{H}$. Moreover, $\mathcal{H}$ includes all the wave functions which form the so-called Dirac sea (for an explanation of this point see for example[6]). All the Dirac wave functions in $\mathcal{H}$ can be identified with the ensemble of *physical wave functions* of the causal fermion system as introduced in (10). This "identification" is made mathematically precise by also identifying the objects of Lorentzian spin geometry with corresponding inherent objects of the causal fermion system (for details see [14, Section 1.2]). The name causal *fermion* system is motivated by the fact that Dirac particles are fermions. According to (13), the local correlation operator $F^\varepsilon(p)$ describes densities and correlations of the physical wave functions at the spacetime point p. Working exclusively with the local correlation operators and the corresponding push-forward measure ρ means in particular that the geometric structures are encoded in and must be retrieved from the physical wave functions. Since the physical wave functions describe the distribution of matter in spacetime, one can summarize this concept by saying that *matter encodes geometry*.

3.2. *Classical Gravity in the Continuum Limit*

The construction of the causal fermion system in the previous section involved the Lorentzian metric g. But this Lorentzian metric did not need to satisfy the Einstein equation. Instead of postulating the Einstein equations, our strategy is to *derive* these equations from the causal action principle. To this end, we need to evaluate the EL equations (6) for the causal fermion system $(\mathcal{H}, \mathcal{F}, \rho)$ constructed in the previous section in the limit $\varepsilon \searrow 0$ when the ultraviolet regularization is removed. This analysis, referred to as the *continuum limit*, is carried out in detail in [7, Chapter 4]. Before giving a brief outline of how this analysis works, we state the main result in the context of gravitational fields: The EL equations (6) are satisfied asymptotically for small $\varepsilon > 0$ only if the Lorentzian metric satisfies the Einstein equations, up to possible higher order corrections in curvature (which scale in powers of $(\delta^2 \,\mathrm{Riem})$, where δ is the Planck length and Riem is the curvature tensor), i.e. (see [7, Theorems 4.9.3 and 5.4.4])

$$R_{jk} - \frac{1}{2}\, R\, g_{jk} + \Lambda\, g_{jk} = G\, T_{jk} + \mathcal{O}\big(\delta^4\, \mathrm{Riem}^2\big)\,. \tag{15}$$

Moreover, it is shown that the gravitational coupling constant G is determined by the length scale of the microscopic spacetime structures and has the scaling

$$G \sim \delta^2 .$$

The cosmological constant Λ, however, is not determined by our method. In order to avoid confusion, we finally note that we carefully distinguish the Planck length δ from the regularization length ε. The reason is that, although it seems natural to assume that these length scales coincide, this does not necessarily need to be the case. In fact, the are indications that ε should be chosen even much smaller than δ (for a detailed discussion of the length scales see for example [4, Appendix A]).

We now briefly explain the general procedure in the analysis of the continuum limit. Given a causal fermion system $(\mathcal{H}, \mathcal{F}, \rho)$ describing a globally hyperbolic spacetime $(\mathcal{M}, g)$ (as constructed in Section 3.1), our task is to evaluate the EL equations (7). The first step is to compute the eigenvalues $\lambda_1^{xy}, \ldots, \lambda_{2n}^{xy}$ of the operator product $F^\varepsilon(x)\, F^\varepsilon(y)$ for given $x, y \in \mathcal{M}$. Using (12) together with the fact that the non-zero eigenvalues of a matrix product as well as the corresponding algebraic multiplicities do not change when the matrices are cyclically commuted, one can just as well compute the eigenvalues of the *closed chain* A_{xy} defined by

$$A_{xy} := P^\varepsilon(x,y)\, P^\varepsilon(y,x) \;:\; S_x\mathcal{M} \to S_x\mathcal{M} , \tag{16}$$

where the *regularized kernel of the fermionic projector* $P^\varepsilon(x,y)$ is defined by

$$P^\varepsilon(x,y) = -\Psi^\varepsilon(x)\big(\Psi^\varepsilon(y)\big)^* \;:\; S_y\mathcal{M} \to S_x\mathcal{M} .$$

In this way, it suffices to compute the eigenvalues of a linear endomorphism of S_xM, which can be represented by a 4×4-matrix. Moreover, the regularized kernel of the fermionic projector can be analyzed explicitly using the *regularized Hadamard expansion*. In general terms, $P^\varepsilon(x,y)$ is a smooth function which in the limit $\varepsilon \searrow 0$ converges to a distribution which has singularities on the lightcone (i.e. for lightlike separation of x and y). As a consequence, the pointwise product in (16) is ill-defined in this limit. In the continuum limit analysis, one deals with this issue by studying the Lagrangian and other composite expressions in $P^\varepsilon(x,y)$ asymptotically for small $\varepsilon > 0$. Evaluating the resulting expressions in the EL equations (7), one gets equations involving the gravitational field and the Dirac wave functions. In models containing neutrinos, these equations imply the Einstein equations (15).

The detailed computations can be found in,[7] where the continuum limit analysis is carried out in Minkowski space. In this context, the Hadamard expansion is referred to as the *light cone expansion*. The reader interested in the general construction of the regularized Hadamard expansion in curved spacetime is referred to.[18]

4. Going Beyond Classical Gravity

The above derivation of the Einstein equations in the continuum limit has two disadvantages. First, it is rather technical and thus does not give a good intuitive

understanding of the underlying mechanisms. Second and more importantly, the Einstein equations are obtained only for the special class of examples of causal fermion systems constructed in Section 3.1 and hold only in the continuum limit. But the methods do not give an insight into the geometric meaning of the EL equations (7) for general causal fermion systems describing more general "quantum" spacetimes. This raises the following question:

$$\text{Given a general causal fermion system } (\mathcal{H}, \mathcal{F}, \rho), \text{ how do the EL equations (7) relate matter to the geometry of spacetime?} \tag{17}$$

For a general causal fermion system, we cannot work with tensor fields, making it impossible to formulate the Einstein equations or modifications thereof. Therefore, we need to go beyond the mathematical setting of Lorentzian geometry. In the next sections, we explain step by step how this can be done and mention a few results.

4.1. *A Lorentzian Quantum Geometry*

In[9] it was shown that a causal fermion system gives rise to geometric structures in spacetime. The general strategy for obtaining these geometric structures is as follows. Given two spacetime points $x, y \in M$, the corresponding spin space S_x and S_y (see Section 2.4) are subspaces of the underlying Hilbert space $\mathcal{H}$. Denoting the orthogonal projection to the spin spaces S_x again by

$$\pi_x : \mathcal{H} \to S_x ,$$

the *kernel of the fermionic projector* is introduced by

$$P(x,y) = \pi_x y|_{S_y} \;:\; S_y \to S_x .$$

Being an operator from one spin space to another, it gives relations between the spacetime points. In other words, the kernel of the fermionic projector induces additional structures in spacetime. One important structure is the *spin connection* $D_{x,y}$, being a unitary mapping between the spin spaces,

$$D_{x,y} \;:\; S_y M \to S_x M \qquad \text{unitary} \tag{18}$$

(unitary with respect to the spin inner product (9)). A first idea for construction $D_{x,y}$ is to take a polar decomposition of $P(x,y)$. This idea needs to be refined in order to also obtain a metric connection and to arrange that the different connections are compatible. Here for brevity we omit the details and refer to[9] or the review.[8] In general terms, it turns out that there is a canonical spin connection (18), provided that the operators x and y satisfy certain conditions, which are subsumed in the notion that the spacetime points be *spin-connectable*. *Curvature* $\mathfrak{R}$ can be defined as the holonomy of the spin connection. Thus, in the simplest case, for three points $x, y, z \in M$ which are mutually spin-connectable, one sets

$$\mathfrak{R}(x,y,z) = D_{x,y}\, D_{y,z}\, D_{z,x} \;:\; S_x M \to S_x M .$$

In[9] also the correspondence to Lorentzian spin geometry is established, in the sense that for causal fermion systems describing a globally hyperbolic Lorentzian manifold (as constructed in Section 3.1) and taking a suitable limit $\varepsilon \searrow 0$, the spin connection (18) goes over to the spinorial Levi-Civita connection of the Lorentzian manifold.

While these structures give a good understanding of the geometry of a causal fermion system, so far they have not been fruitful for unraveling the form of the gravitational interaction as described by the causal action principle. The basic problem is that the EL equations (7) cannot be formulated in terms of these geometric structures. In other words, it does not seem possible to rewrite the causal action principle as a geometric variational principle involving $D_{x,y}$ and $\mathfrak{R}$. With this in mind, we now move on to other structures which again have a geometric meaning, but harmonize better with the EL equations (7).

4.2. *Surface Layer Integrals*

Coming back to our question (17), we follow another path for getting a connection between the geometric structures and the EL equations (7). This method is inspired by the fact that the effects of gravity can also be captured by considering the volume and area of surfaces in spacetime, and by analyzing how this area changes under flows of the surfaces. Typical examples for this connection are Huisken's isoperimetric mass (see for example[22]) and Jacobson's connection between area change and matter flux.[4] Thinking along these lines, the first obvious question is how a surface integral can be defined in the setting of causal fermion systems. Once this question has been answered, one can analyze the area of families of surfaces and make the above analogies more precise.

As a typical example, suppose we want to define the analog of an integral over a Cauchy surface $\mathcal{N}$ in a globally hyperbolic spacetime, i.e. symbolically

$$\int_{\mathcal{N}} (\cdots)\, d\mu_{\mathcal{N}}(x)\,, \tag{19}$$

where $d\mu_{\mathcal{N}}$ is the induced volume measure on $\mathcal{N}$. Here one can think of $(\cdots)$ as a density like for example the inner product $\nu_i J^i$ of a current vector field with the future-directed normal ν. In the setting of causal fermion system, surface integrals like (19) are undefined. Instead, one considers so-called *surface layer integrals*. In general terms, a surface layer integral is a double integral of the form

$$\int_{\Omega} \left(\int_{M \setminus \Omega} (\cdots)\, \mathcal{L}_\kappa(x,y)\, d\rho(y) \right) d\rho(x)\,, \tag{20}$$

where one variable is integrated over a subset $\Omega \subset M$, and the other variable is integrated over the complement of Ω. Here $(\cdots)$ is the analog of the corresponding factor in (19), but now having the mathematical structure of being a differential operator acting on the Lagrangian.

In order to explain the basic concept, let us assume for a moment that the Lagrangian is of *short range* in the following sense. We let $d \in C^0(M \times M, \mathbb{R}_0^+)$ be a suitably chosen distance function on M. Then the assumption of short range can be quantified by demanding that $\mathcal{L}_\kappa$ should vanish on distances larger than l, i.e.

$$d(x,y) > l \quad \Longrightarrow \quad \mathcal{L}_\kappa(x,y) = 0\,. \tag{21}$$

Under this assumption, the surface layer integral (20) only involves pairs (x,y) of distance at most l, where x lies in Ω, whereas y lies in the complement $M \setminus \Omega$. As a consequence, the integral only involves points in a layer around the boundary of Ω of width l, i.e.

$$x, y \in B_l\big(\partial\Omega\big)\,.$$

Therefore, a double integral of the form (20) can be regarded as an approximation of a surface integral on the length scale l, as shown in Figure 1. In most applications, the Lagrangian is *not* of short range in the strict sense (21). But it decays on the Compton scale $l \sim 1/m$ (where m denotes again the mass of the Dirac particles), so that the qualitative picture in Figure 1 still applies.

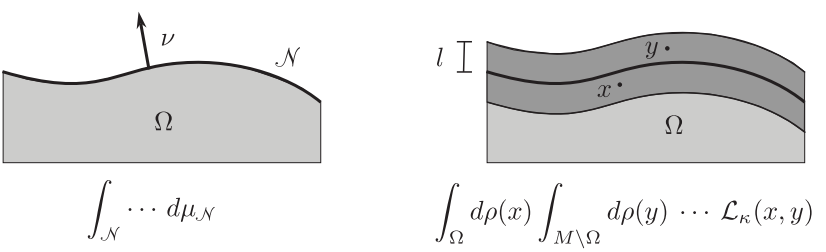


Fig. 1. A surface integral and a corresponding surface layer integral.

Surface layer integrals were first introduce in[15] in the context of Noether-like theorems. The analysis of this paper revealed that, in contrast to the geometric structures in Section 4.1, the structure of a surface layer integral does fit nicely to the EL equations (7), giving rise to very useful conservation laws. This connection was analyzed further and more systematically in.[16,17] Moreover, in[11] another variant of a surface layer integral was introduced, which will be importance for what follows. This new surface layer integral can be regarded as a generalization or "nonlinear version" of (20), as can be understood as follows. The differential operator $(\cdots)$ in the integrand can be regarded as describing first or second variations of the measure ρ. Instead of considering variations of ρ, we now consider an additional measure $\tilde{\rho}$ which can be thought of as a finite perturbation of the measure ρ. Consequently, we also have two spacetimes

$$M := \operatorname{supp} \rho \qquad \text{and} \qquad \tilde{M} := \operatorname{supp} \tilde{\rho}\,.$$

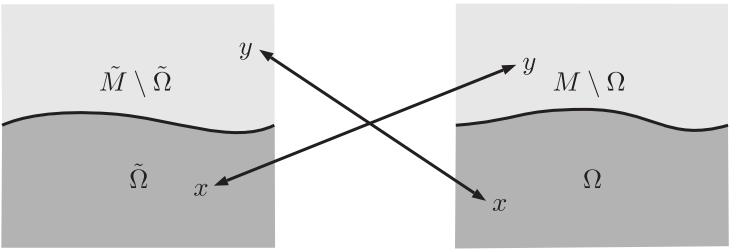


Fig. 2. The nonlinear surface layer integral.

Choosing two compact subsets $\Omega \subset M$ and $\tilde{\Omega} \subset \tilde{M}$ of the corresponding spacetimes, we form the *nonlinear surface layer integral* by

$$\gamma^{\tilde{\Omega},\Omega}(\tilde{\rho}, \rho) := \int_{\tilde{\Omega}} d\tilde{\rho}(x) \int_{M \setminus \Omega} d\rho(y)\, \mathcal{L}_\kappa(x,y) - \int_{\Omega} d\rho(x) \int_{\tilde{M} \setminus \tilde{\Omega}} d\tilde{\rho}(y)\, \mathcal{L}_\kappa(x,y)\,. \quad (22)$$

Note that one argument of the Lagrangian is in M, whereas the other is in $\tilde{M}$. Moreover, one argument lies inside the set Ω respectively $\tilde{\Omega}$, whereas the other argument lies outside this set. In this way, the nonlinear surface layer integral "compares" the two spacetimes near the boundaries of Ω and $\tilde{\Omega}$, as is illustrated in Figure 2.

4.3. *The Total Mass of a Static Causal Fermion System*

Combining the general structure of Huisken's isoperimetric mass with the concept of a surface layer integral, the question arises whether notions like the total mass and total energy can be introduced for a class of causal fermion systems which generalize asymptotically flat Lorentzian manifolds. This question has been answered in the affirmative in[19] in the static setting. We now outline a few constructions and results of this paper.

Static causal fermion systems are introduced as usual by demanding a one-parameter group of symmetries:

Definition 4.1. Let $(\mathcal{U}_t)_{t \in \mathbb{R}}$ be a strongly continuous one-parameter group of unitary transformations on the Hilbert space $\mathcal{H}$ (i.e. $s\text{-}\lim_{t' \to t} \mathcal{U}_{t'} = \mathcal{U}_t$ and $\mathcal{U}_t \mathcal{U}_{t'} = \mathcal{U}_{t+t'}$). The causal fermion system $(\mathcal{H}, \mathcal{F}, \rho)$ is **static with respect to** $(\mathcal{U}_t)_{t \in \mathbb{R}}$ if it has the following properties:

(i) Spacetime $M := \operatorname{supp} \rho \subset \mathcal{F}$ is a topological product,

$$M = \mathbb{R} \times N\,.$$

We write a spacetime point $x \in M$ as $x = (t, \mathbf{x})$ with $t \in \mathbb{R}$ and $\mathbf{x} \in N$.

(ii) The one-parameter group $(\mathcal{U}_t)_{t \in \mathbb{R}}$ leaves the measure ρ invariant, i.e.

$$\rho\big(\mathcal{U}_t\, \Omega\, \mathcal{U}_t^{-1}\big) = \rho(\Omega) \qquad \text{for all } \rho\text{-measurable } \Omega \subset \mathcal{F}\,.$$

Moreover,

$$\mathcal{U}_{t'}\,(t,\mathbf{x})\,\mathcal{U}_{t'}^{-1} = (t+t',\mathbf{x})\,.$$

As a consequence of (ii), the measure ρ has the product form

$$d\rho = dt\, d\mu\,,$$

where μ is a measure on $\mathcal{G} := \mathcal{G} = \mathcal{F}^{\text{reg}}/\mathbb{R}$.

Restricting attention to static causal fermion systems and varying within the class of measures which are invariant under the action of a given one-parameter group $(\mathcal{U}_t)_{t\in\mathbb{R}}$, one obtains a corresponding *static causal action principle*, where one minimizes the static action

$$\mathcal{S}_{\text{static}}(\mu) = \int_{\mathcal{G}} d\mu(\mathbf{x}) \int_{\mathcal{G}} d\mu(\mathbf{y})\, \mathcal{L}_{\text{static}}(\mathbf{x},\mathbf{y})$$

under variations of the measure μ on $\mathcal{G}$, keeping the total volume $\mu(\mathcal{G})$ fixed (*volume constraint*). The static Lagrangian is obtained by integrating one argument of the κ-Lagrangian (8) over time,

$$\mathcal{L}_{\text{static}}(\mathbf{x},\mathbf{y}) := \int_{-\infty}^{\infty} \mathcal{L}_\kappa\big((0,\mathbf{x}),(t,\mathbf{y})\big)\, dt\,.$$

The *total mass* $\mathfrak{M}$ of a static causal fermion system described by the measure $\tilde{\mu}$ is defined by comparing $\tilde{\mu}$ asymptotically near infinity with a measure μ describing the vacuum. To this end, one exhausts the supports N and $\tilde{N}$ of these measures by sequences of compact subsets $(\Omega_n)_{n\in\mathbb{N}}$ and $(\tilde{\Omega}_n)_{n\in\mathbb{N}}$, respectively, and takes the limit of a nonlinear surface layer integral of structure similar to (22),

$$\begin{aligned}\mathfrak{M}(\tilde{\mu},\mu) = &\lim_{\Omega_n \nearrow N,\, \tilde{\Omega}_n \nearrow \tilde{N} \text{ with } \mu(\Omega_n)=\tilde{\mu}(\tilde{\Omega}_n)<\infty} \\ &\times \left(\int_{\tilde{\Omega}_n} d\tilde{\mu}(\mathbf{x}) \int_{N\setminus\Omega_n} d\mu(\mathbf{y})\, \mathcal{L}_{\text{static}}(\mathbf{x},\mathbf{y}) - \int_{\Omega_n} d\mu(\mathbf{x}) \int_{\tilde{N}\setminus\tilde{\Omega}_n} d\tilde{\mu}(\mathbf{y})\, \mathcal{L}_{\text{static}}(\mathbf{x},\mathbf{y}) \right). \quad (23)\end{aligned}$$

In order for this expression to be well-defined, one needs to assume that the spacetimes are *asymptotically flat* (for brevity, we do not give the precise definition, which can be found in [19, Definition 1.5]). As a consequence of the EL equations, the mass does not depend on the choice of the exhaustions, except that the volume condition $\mu(\Omega_n) = \tilde{\mu}(\tilde{\Omega}_n)$ must hold. Moreover, a *positive mass theorem* is proved which states that the total mass $\mathfrak{M}(\tilde{\mu},\mu)$ is non-negative if a suitable *local energy condition* holds (for details see [19, Definition 1.8 and Theorem 1.9]).

The correspondence to the classical positive mass theorem (see for example[23]) is established by showing that for a static causal fermion systems describing the Schwarzschild spacetime (constructed again as explained in Section 3.1), the total mass $\mathfrak{M}(\tilde{\mu},\mu)$ coincides (up to an irrelevant prefactor) with the ADM mass[2] (see [19, Theorem 1.10]). For brevity, we cannot enter the proof and the detailed constructions. But we make one remark which will be important in connection with the construction of the quantum state in Section 4.5: When comparing causal

fermion systems describing the Schwarzschild spacetime and Minkowski space, one needs to take into account that the resulting causal fermion systems are defined on two different Hilbert spaces, namely a Hilbert space $\tilde{\mathcal{H}}$ formed of Dirac solutions in Schwarzschild and a Hilbert space $\mathcal{H}$ formed of Dirac solutions in Minkowski space. Before we can make sense of the nonlinear surface layer integral (23), these Hilbert spaces must be identified by a unitary transformation V,

$$V \,:\, \mathcal{H} \to \tilde{\mathcal{H}}\,. \tag{24}$$

This identification is *not* canonical but leaves us with the freedom of choosing a unitary transformation. Therefore, in order to make sense of (23), one must prove that the total mass is independent of the choice of the unitary transformation V (for details see [19, Section 4.3]).

4.4. *Connection Between Area Change and Matter Flux*

In 1995, Ted Jacobson gave a derivation of the Einstein equations from thermodynamic principles.[21] At the heart of his argument is the formula

$$\frac{d}{d\tau} A(S_\tau) = c\, F(S_\tau) \tag{25}$$

which states that the area change of a family of two-surfaces S_τ propagating along a null Killing direction is proportional to the matter flux $F(S_\tau)$ across these surfaces (with c a universal constant). In,[4] this formula is derived in the setting of causal fermion systems from the EL equations (7) (without referring to thermodynamics). This result gives an alternative derivation of the Einstein equations from the causal action principle (again with an undetermined cosmological constant). Compared to the continuum limit analysis described in Section 3.2, this alternative derivation has the advantage that it is more conceptual and thus gives a more direct understanding of why and how the Einstein equations arise. Moreover, the procedure in[4] has the benefit that it does not rely on tensor calculus. Therefore, it goes beyond Lorentzian geometry and applies to more general "quantum" spacetimes.

We now explain the general procedure and a few constructions from.[4] The first question is how to define the area A of a two-dimensional surface S in the setting of causal fermion systems. It is most convenient to describe $S \subset M$ as

$$S = \partial\Omega \cap \partial V\,,$$

where Ω can be thought of as being the past of a Cauchy surface, and V describing a spacetime cylinder. This description has the advantage that the resulting surface layer integrals are well-defined even in situations when spacetime is singular or discrete, in which case the boundaries $\partial\Omega$ and ∂V are no longer a sensible concept. The most natural way of introducing a surface layer integral localized in a neighborhood of S is a double integral of the form

$$\int_{\Omega\cap V} \left(\int_{M\setminus(\Omega\cup V)} (\cdots)\, \mathcal{L}_\kappa(x,y)\, d\rho(y) \right) d\rho(x) \tag{26}$$

(where $(\cdots)$ stands again for a differential operator acting on the Lagrangian). If the Lagrangian has short range, we only get contributions to this surface layer integral if both x and y are close to the two-dimensional surface S (see the left of Figure 3).

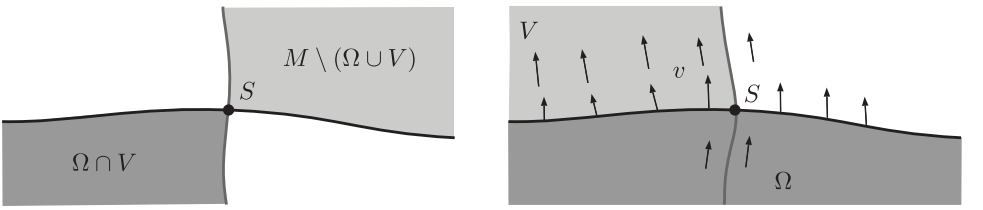


Fig. 3. Two-dimensional surface layer integrals.

The disadvantage of this method is that, similar as explained in Section 4.1 for the structures of the Lorentzian quantum geometry and the EL equations, the surface layer integral (26) does not fit together with the notion of matter flux. Therefore, it is preferable to define the two-dimensional area as follows. We need to assume that M has a smooth manifold structure, and that the measure ρ is absolutely continuous and smooth with respect to the Lebesgue measure in every chart (x, U) of M, i.e.

$$d\rho = h(x)\, d^4x \qquad \text{with } h \in C^\infty(U, \mathbb{R}^+)\,. \tag{27}$$

Also assuming that the boundaries of Ω and V are smooth, we can choose a smooth vector field v which is transverse to the hypersurface $\partial\Omega$ and tangential to ∂V (see the right of Figure 3). Under these additional assumptions, one can introduce an integration measure $d\mu$ on the hypersurface $\partial\Omega$ by

$$d\mu(v, x) = h\, \epsilon_{ijkl}\, v^i\, dx^j dx^k dx^l\,,$$

where ϵ_{ijkl} is the totally anti-symmetric Levi-Civita symbol (normalized by $\epsilon_{0123} = 1$). Now we can define the two-dimensional area by

$$A := \int_{\partial\Omega \cap V} d\mu(v, x) \int_{M \setminus V} d\rho(y)\, \mathcal{L}_\kappa(x, y)\,. \tag{28}$$

We next describe a symmetry by a notion of a Killing field. This notion will also give rise to a notion of matter flux.

Definition 4.2. A vector field u on M is called **Killing field** of the causal fermion system if the following conditions hold:

(i) The divergence of u vanishes, i.e.

$$\frac{1}{h}\, \partial_j \big(h\, u^j\big) = 0$$

(where h is again the weight function in (27)).

(ii) The directional derivative of the Lagrangian is small in the sense that

$$\big(D_{1,u} + D_{2,u}\big)\mathcal{L}_\kappa(x,y) \lesssim \frac{m^4}{\varepsilon^4\, \delta^4} \tag{29}$$

(where again ε is the regularization length, and δ is the Planck length).

The reason why the right side in (29) is non-zero is that the derivative of the Lagrangian involves contributions from both the geometry and the matter fields. The condition (29) means that the geometric contributions vanish (which are typically much larger). The fact that the matter contribution remains makes it possible to define the matter flux by

$$F(S) := \int_{\partial\Omega\cap V} d\mu(v,x) \int_{M\setminus V} d\rho(y)\, \big(D_{1,u} + D_{2,u}\big)\mathcal{L}_\kappa(x,y)\,, \tag{30}$$

where we assume that the Killing field u is tangential to Ω (see Figure 4). In the limiting case of lightlike propagation, the vector fields u and v coincide, making it possible to recover the matter flux (30) as the derivative of the area (28) in the direction of the Killing field u. This gives the desired relation (25).

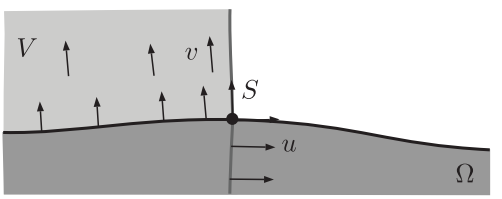


Fig. 4. Matter flux through S.

4.5. *The Quantum State of a Causal Fermion System*

We finally give an outlook on quantum field theory and quantum gravity. So far, the connection to quantum field theory has been established only for causal fermion systems describing Minkowski space.[11,12] But most of the methods apply just as well to more general causal fermion systems describing curved spacetimes. Moreover, one should keep in mind that in a causal fermion system, all the bosonic interactions are described in a unified way with the same mathematical structures, giving hope that methods developed for the electromagnetic interaction in Minkowski space can be adapted to the gravitational field. But, as is usually the case in mathematics, the connections and analogies are rather subtle and need to be explored carefully and in depth. We plan to do so in the near future.

Coming back to causal fermion systems describing Minkowski space, in[12] a *quantum state* ω^t of a causal fermion system $(\tilde{\mathcal{H}}, \tilde{\mathcal{F}}, \tilde{\rho})$, which describes an interacting

spacetime, is constructed. Here by a "quantum state" we mean a positive linear functional on the $*$-algebra of observables $\mathscr{A}$, i.e.

$$\omega^t : \mathscr{A} \to \mathbb{C} \quad \text{complex linear} \qquad \text{and} \qquad \omega^t\big(A^* A\big) \geq 0 \quad \text{for all } A \in \mathscr{A} .$$

The observables are formed of linearized solutions of the EL equations. The general idea for constructing the quantum state is to compare $(\tilde{\mathcal{H}}, \tilde{\mathcal{F}}, \rho)$ at time t with a causal fermion system $(\mathcal{H}, \mathcal{F}, \rho)$ describing the vacuum. Similar to the procedure for the total mass (23), the "comparison" is again performed with a nonlinear surface layer integral (22), but now with $\tilde{\Omega}$ and Ω chosen as the past of the respective Cauchy surfaces (similar as shown in Figure 2). The unitary freedom in identifying the Hilbert spaces already mentioned in the connection of the positive mass theorem in (24) is now taken care of by integrating over a group $\mathcal{G}$ of unitary operators. This leads to a formula for the quantum state of the symbolic form

$$\omega^t(A) := \frac{1}{Z^t\big(\beta, \tilde{\rho}\big)} \int_{\mathcal{G}} (\cdots)\, e^{\beta\, \gamma^t(\tilde{\rho}, \mathcal{U}\rho)}\, d\mu_{\mathcal{G}}(\mathcal{U}) , \tag{31}$$

where $Z^t\big(\beta, \tilde{\rho}\big)$ is the partition function defined by

$$Z^t\big(\beta, \tilde{\rho}\big) := \int_{\mathcal{G}} e^{\beta\, \gamma^t(\tilde{\rho}, \mathcal{U}\rho)}\, d\mu_{\mathcal{G}}(\mathcal{U}) ,$$

and β is a real parameter. The factor $(\cdots)$ in (31) consists of a product of surface layer integrals formed of the linearized solutions contained in the observable A (for details see [12, Section 4]).

The significance of this construction is that it becomes possible to describe the interacting causal fermion system in the familiar language of quantum field theory. Consequently, also the dynamics as described by the EL equations (7) can be rewritten as a time evolution of the state ω^t. The detailed form of the resulting quantum dynamics is presently under investigation.[13]

Acknowledgments

I would like to thank Michael Kiessling and Shadi Tahvildar-Zadeh for the kind invitation to the Marcel Grossmann Meeting. I am grateful to Christoph Krpoun for helpful comments on the manuscript.

References

1. *Link to web platform on causal fermion systems: www.causal-fermion-system.com.*
2. R. Arnowitt, S. Deser, and C.W. Misner, *Energy and the criteria for radiation in general relativity*, Phys. Rev. (2) **118** (1960), 1100–1104.
3. Y. Bernard and F. Finster, *On the structure of minimizers of causal variational principles in the non-compact and equivariant settings*, arXiv:1205.0403 [math-ph], Adv. Calc. Var. **7** (2014), no. 1, 27–57.
4. E. Curiel, F. Finster, and J.M. Isidro, *Two-dimensional area and matter flux in the theory of causal fermion systems*, arXiv:1910.06161 [math-ph], Internat. J. Modern Phys. D **29** (2020), 2050098.

5. F. Finster, *Causal variational principles on measure spaces*, arXiv:0811.2666 [math-ph], J. Reine Angew. Math. **646** (2010), 141–194.
6. ———, *A formulation of quantum field theory realizing a sea of interacting Dirac particles*, arXiv:0911.2102 [hep-th], Lett. Math. Phys. **97** (2011), no. 2, 165–183.
7. ———, *The Continuum Limit of Causal Fermion Systems*, arXiv:1605.04742 [math-ph], Fundamental Theories of Physics, vol. 186, Springer, 2016.
8. ———, *Causal fermion systems: A primer for Lorentzian geometers*, arXiv: 1709.04781 [math-ph], J. Phys.: Conf. Ser. **968** (2018), 012004.
9. F. Finster and A. Grotz, *A Lorentzian quantum geometry*, arXiv:1107.2026 [math-ph], Adv. Theor. Math. Phys. **16** (2012), no. 4, 1197–1290.
10. F. Finster and M. Jokel, *Causal fermion systems: An elementary introduction to physical ideas and mathematical concepts*, arXiv:1908.08451 [math-ph], Progress and Visions in Quantum Theory in View of Gravity (F. Finster, D. Giulini, J. Kleiner, and J. Tolksdorf, eds.), Birkhäuser Verlag, Basel, 2020, pp. 63–92.
11. F. Finster and N. Kamran, *Complex structures on jet spaces and bosonic Fock space dynamics for causal variational principles*, arXiv:1808.03177 [math-ph], Pure Appl. Math. Q. **17** (2021), no. 1, 55–140.
12. ———, *Fermionic Fock spaces and quantum states for causal fermion systems*, arXiv:2101.10793 [math-ph], Ann. Henri Poincaré **23** (2022), no. 4, 1359–1398.
13. F. Finster, N. Kamran, and M. Reintjes, *The quantum dynamics of a causal fermion system: Holographic mixing and bosonic loop diagrams*, in preparation.
14. F. Finster, S. Kindermann, and J.-H. Treude, *An Introductory Course on Causal Fermion Systems*, in preparation, www.causal-fermion-system.com/intro-public.pdf.
15. F. Finster and J. Kleiner, *Noether-like theorems for causal variational principles*, arXiv:1506.09076 [math-ph], Calc. Var. Partial Differential Equations **55:35** (2016), no. 2, 41.
16. ———, *A Hamiltonian formulation of causal variational principles*, arXiv:1612.07192 [math-ph], Calc. Var. Partial Differential Equations **56:73** (2017), no. 3, 33.
17. ———, *A class of conserved surface layer integrals for causal variational principles*, arXiv:1801.08715 [math-ph], Calc. Var. Partial Differential Equations **58:38** (2019), no. 1, 34.
18. F. Finster and M. Kraus, *The regularized Hadamard expansion*, arXiv:1708.04447 [math-ph], J. Math. Anal. Appl. **491** (2020), no. 2, 124340.
19. F. Finster and A. Platzer, *A positive mass theorem for static causal fermion systems*, arXiv:1912.12995 [math-ph], Adv. Theor. Math. Phys. **25** (2021), no. 7, 1735–1818.
20. F. Finster and M. Reintjes, *A non-perturbative construction of the fermionic projector on globally hyperbolic manifolds I – Space-times of finite lifetime*, arXiv:1301.5420 [math-ph], Adv. Theor. Math. Phys. **19** (2015), no. 4, 761–803.
21. T. Jacobson, *Thermodynamics of spacetime: The Einstein equation of state*, Phys. Rev. Lett. **75** (1995), 1260–1263.
22. J.L. Jauregui and D.A. Lee, *Lower semicontinuity of mass under C^0 convergence and Huisken's isoperimetric mass*, arXiv:1602.00732 [math.DG], J. Reine Angew. Math. **756** (2019), 227–257.
23. R. Schoen and S.-T. Yau, *On the proof of the positive mass conjecture in general relativity*, Commun. Math. Phys. **65** (1979), no. 1, 45–76.

Newman-Penrose-Debye formalism for fields of various spins in pp-wave backgrounds

Aleksandr Kulitskii

Faculty of Physics, Lomomosov Moscow State University, Moscow, Russia
E-mail: av.kulitskiy@yandex.ru

Elena Yu Melkumova

Faculty of Physics, Lomomosov Moscow State University, Moscow, +74992492878, Russia
E-mail: EYM@srdlan.npi.msu.su

Using Newman-Penrose formalism in tetrad and spinor forms, we perform separation of variables in the wave equations for massless fields of various spins s=1/2, 1, 3/2, 2 on the background of exact plane-fronted gravitational wave metrics. Then, applying Wald's method of conjugate operators, we derive equations for Debye potentials and we find the back-projection operators expressing multicomponent fields in terms of these potentials. For shock wave backgrounds, as a special case of the non-vacuum pp-waves, the exact solutions for Debye potentials are constructed explicitly. The possibilities of generalization to the case of massive fields are discussed, in particular, construction of exact solutions of the Dirac and Proca equations. These results can be used in various supergravity problems on the pp-wave backgrounds, including holographic applications.

Keywords: PP-waves, Newman-Penrose formalism, Debye potentials, shock waves

1. Introduction

Wave equations for fields of different spins in curved space have been extensively studied since the early 1970s. Solutions describing propagation of fields on curved backgrounds are in demand in various contexts including the problem of radiation, supergravity and superstrings theories. Generically, considering fields of higher spins in curved space-time, difficulties arise with splitting the systems of coupled equations into separate ones for certain combinations of the field components. Two efficient methods to do this are known. The first is the application of the Newman-Penrose (NP) formalism,[1] in which the intrinsic symmetries of the masseless field equations with respect to the Lorentz group are conveniently incorporated. The second efficient tool is the method of Debye potentials,[2] which allows to reconstruct the multicomponent fields in terms of solutions of some scalar equations. Debye potentials are complex functions which incorporate two degree of freedom of the massless field of any spin.

Formalism of Debye potentials was suggested in gravity theory in the works by Cohen[3] and Kegeles[4] and then effectively used by Chrzanowksi[5] in Kerr spacetime of type D, according to the Petrov classification,[6] in combination with the Teukolsky

equation.[7] In the Kerr metric Debye representations were found for electromagnetic field and gravitational perturbations. Later, formalism of Debye potentials was discussed in detail for Rarita-Schwinger spin 3/2 field by Torres del Castilho.[8]

In this work we apply the Newman-Penrose-Debye formalism to the case of metrics of type N. More specifically, we consider the case of pp-wave metrics, which have numerous applications both in astrophysics and in various theoretical aspects, including supergravity and holography.

2. PP-waves

These metrics were introduced by Brinkmann[9] in 1925 and later interpreted as representing the plane-fronted gravitational waves (pp-waves). They are exact solutions of Einstein's equations of the following form

$$ds^2 = dudv - H(u,\zeta,\bar{\zeta})dudu - d\zeta d\bar{\zeta}, \tag{1}$$

where $\zeta = x + iy$ is a complex transverse coordinate, H is arbitrary nonlinear real profile function of ζ, $\bar{\zeta}$, which specifies the nature of the wave. The determinant $g = 1/16$ is coordinate-independent. The scalar curvature turns out to be zero. For $H = 0$ the metric is Minkowski. These metrics describe plane waves with parallel rays belonging to the class of solutions that admits isotropic nonexpanding congruences without shear and twist, as well as the existence of an isotropic Killing vector.

A natural choice of the null tetrad basis for them is

$$\mathbf{l} = \sqrt{2}\partial_v, \qquad \mathbf{n} = \sqrt{2}(\partial_u + H\partial_v), \qquad \mathbf{m} = \sqrt{2}\partial_\zeta, \qquad \bar{\mathbf{m}} = \sqrt{2}\partial_{\bar{\zeta}}. \tag{2}$$

We will also use the NP covariant derivatives along the null tetrad vectors:

$$D = l^\mu\nabla_\mu, \qquad \Delta = n^\mu\nabla_\mu, \qquad \delta = m^\mu\nabla_\mu, \qquad \bar{\delta} = \bar{m}^\mu\nabla_\mu \tag{3}$$

and satisfying, in the case of a plane wave metrics, the following commutation relations

$$[\Delta, D] = 0, \qquad [\delta, D] = 0, \qquad [\delta, \Delta] = -\bar{\nu}D, \qquad [\bar{\delta}, \delta] = 0. \tag{4}$$

The nonzero tetrad projections of the traceless part of the Ricci tensor, the Weyl scalar, and the only nonzero spin coefficient ν are

$$\begin{aligned} &\Phi_{22} = \frac{1}{2}n^\alpha n^\beta R_{\alpha\beta} = -H_{,\zeta\bar{\zeta}}, && \nu = -\bar{m}^\mu\delta n_\mu = -H_{,\bar{\zeta}}, \\ &\Psi_4 = -n^\alpha\bar{m}^\beta n^\sigma\bar{m}^\tau C_{\alpha\beta\sigma\tau} = -H_{,\bar{\zeta}\bar{\zeta}}, && \Lambda = \frac{1}{24}R = 0, \end{aligned} \tag{5}$$

where $C_{\alpha\beta\sigma\tau}$ is the Weyl tensor. In the vacuum case, the gravitational field equations $\Phi_{22} = 0$ reduce to the two-dimensional Laplace equation

$$H_{,\zeta\bar{\zeta}} = 0.$$

The d'Alembert operator for the massless scalar field $\Box = \nabla_\mu\nabla^\mu$ reads:

$$\Box \equiv \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}g^{\mu\nu}\partial_\nu\right) = 2(D\Delta - \delta\bar{\delta}). \tag{6}$$

3. Debye potentials

Here we briefly recall the Wald's procedure[10] for constructing a solution of the field equations of higher spins, which we then apply to the metrics of plane waves. Consider some multicomponent field f (tensor or spinor) satisfying the field equation $\mathcal{E}(f) = 0$, where $\mathcal{E}$ is an appropriate matrix linear differential operator. This is generically a non-quadratic matrix $n \times m$ taking the m-component field into the column of n differential equation. To solve this system of equations one has to disentangle it, which in general is not possible in the closed form. However, it can be possible to decouple a separate equation (or several equations) for some combination φ of the components of the initial field f by the action of another linear partial differential operator $\mathcal{T}$, defining a scalar $\varphi = \mathcal{T}(f)$. Then one can define a pair of linear operators $\mathcal{S}$ and $\mathcal{O}$ such, that

$$\mathcal{S}\mathcal{E}(f) = \mathcal{O}\mathcal{T}(f) = \mathcal{O}(\varphi).$$

The problem of finding such operators is facilitated if one knows the source terms in the inhomogeneous equations both for an initial multicomponent field $\mathcal{E}(f) = J$, and in the decoupled scalar equation $O(\varphi) = \mathcal{S}(J)$. The number of decoupled equations depends on the number of principal null direction of the metric.

The final step is the construction of the adjoint operator $S^\dagger$ with respect to a suitably defined functional scalar product such that, in the matrix form,

$$\int \phi^{A_n} M_{A_n B_m} \psi^{B_m} = \int \psi^{B_n} M^\dagger_{B_n A_m} \phi^{A_m} \ ,$$

where the indices A_n, B_m take n and m values respectively and the integration measure is supressed. Note that complex conjugation is not used.

Now it can be verified that the solution of the homogeneous field equation $\mathcal{E}(f) = 0$ will be guaranteed if

$$f = \mathcal{S}^\dagger \psi, \qquad\qquad \mathcal{O}^\dagger \psi = 0. \tag{7}$$

The last equation is the Debye potential equation, the relevant operator is therefore the adjoint to one used in the equation for a decoupled scalar. It there are several such decoupled equations (e.g. two in type D metrics), one will have two different representations for the initial multicomponent field which are usually related by some tetrad transformation. In type N case the Debye potential is unique.

The complex Debye representation for for real-valued massless fields reflects the existence of two independent polarizations, which can be obtained as the real and imaginary parts of the complex f in the Debye form.

In what follows we apply this procedure to the pp-wave background with the only one nonzero spin coefficient ν. Clearly, the adjoint to the product of two operators will be the product of the adjoints to each of them in reverse order. So to construct an adjoint of some polynomial product it will suffice to know the adjoint to basic operators. The adjoint NP differential operators will read

$$D^\dagger = -D, \qquad \Delta^\dagger = -\Delta, \qquad \delta^\dagger = -\delta, \qquad \bar{\delta}^\dagger = -\bar{\delta}. \tag{8}$$

4. Maxwell field

Maxwell's equations in the Newman-Penrose formalism are written in terms of the projections of the electromagnetic field tensor onto the bivector constructed of null tetrad as follows:

$$\Phi_0 = F_{lm}, \qquad 2\Phi_1 = (F_{ln} + F_{m\bar{m}}), \qquad \Phi_2 = F_{\bar{m}n}. \tag{9}$$

A sourceless decoupled equation for the scalar Φ_0 in the case of vacuum pp-waves was obtained in the Refs.[11] and.[12] According to above procedure, we have to construct the source term to this equation. For this, one should act with an appropriate NP differential operators on the first pair of the system of equations and exclude the variable Φ_1 from them using the commutation relations. This gives

$$\Box\Phi_0 = 2\pi J_0, \qquad J_0 = 2\left[\delta J_l - D J_m\right]. \tag{10}$$

Since metrics of type N have only one principal null direction (contrary to the type D, where there are two), the NP component Φ_0 will be the only one for which the decoupling can be made. Further, having written down both adjoint operators, one can easily construct a solution for vector A^μ, satisfying the Lorentz gauge condition, in terms of the Debye potential:

$$A^\mu = \left[\bar{m}^\mu D - l^\mu \bar{\delta}\right]\psi, \qquad \Box\psi = 0. \tag{11}$$

It is easy to see that the contraction of the Ricci tensor with the expression for the vector potential gives zero; therefore, the solution is valid also for the non-vacuum pp-waves.

5. Weyl field

For the Weyl equations of spin 1/2, one uses the two-component spinor version of the NP formalism. In this case, we will no longer deal with a tetrad, but with a spinor dyad. The massless spin 1/2 equation has the following form:

$$\nabla^A{}_{B'}\chi_A = 0. \tag{12}$$

Writing this in components and carrying out transformations similar those for the vector field, we obtain the decoupled equation with the source term:

$$\Box\chi_0 = N_0, \qquad N_0 = 2\left[D j_{1'} - \delta j_{0'}\right], \tag{13}$$

where the spinor source is denoted as $j_{A'}$. Then performing the conjugation, we construct a solution in terms of the Debye potential, which, in turn, satisfies the d'Alembert equation:

$$\chi_0 = -D\psi, \qquad \chi_1 = -\bar{\delta}\psi, \qquad \Box\psi = 0. \tag{14}$$

6. Rarita-Schwinger field

The spin 3/2 field is described by the Rarita-Schwinger equation for the spinor-vector ψ_μ, which satisfies the equations

$$\gamma^\mu(\nabla_\mu\psi_\nu - \nabla_\nu\psi_\mu) = 0, \qquad \gamma^\rho\psi_\rho = 0. \tag{15}$$

This system is consistent only in the case of the vacuum metrics, the non-vacuum case refers to the supergravity context.

The field of an arbitrary spin s in the two-component spinor formalism is described by a totally symmetric spinor of valence 2s satisfying the equation of motion

$$\nabla^{AB'}\phi_{AB\ldots C} = 0.$$

If $s \geqslant 3/2$ the Buchdahl consistency constraint[13]

$$\Psi^{ABC}{}_{(D}\phi_{EF\ldots)ABC} = 0$$

must be satisfyed, where Ψ_{SABC} denotes the Weyl spinor. But there exists also an alternative approach, developed in the work,[8] where the consistency constraint is satisfied automatically. This method can be easily adapted to the case of the pp-wave backgrounds.

In the spinor equivalent of the Eq. (15)

$$\nabla_{AB'}\psi^A{}_{CD'} - \nabla_{CD'}\psi^A{}_{AB'} = 0, \tag{16}$$

one has to pass from field $\psi^A{}_{CD'}$ to the symmetric rank three spinor arriving at the modified equations of motion

$$\nabla^{AB'}\phi_{ABC} = \Psi^S{}_{ABC}\psi^A{}_S{}^{B'}, \quad \phi^A{}_{BC} \equiv \nabla_{(B|R'|}\psi^A{}_{C)}{}^{R'}.$$

We write down the complete system of equations with the sources for the symmetric spinor field ϕ_{ABC} in the component form, applicable to the type N metrics:

$$\begin{aligned}
&\bar\delta\phi_{000} - D\phi_{100} = \delta J_{0'0'0} - DJ_{1'0'0}; \qquad \Delta\phi_{000} - \delta\phi_{100} = \delta J_{0'1'0} - DJ_{1'1'0};\\
&\bar\delta\phi_{100} - D\phi_{110} = \frac{1}{2}\big[\Delta J_{0'0'0} + \delta J_{0'0'1} - \bar\delta J_{0'1'0} - DJ_{1'0'1}\big];\\
&\Delta\phi_{100} - \delta\phi_{110} - \nu\phi_{000} = \frac{1}{2}\big[\Delta J_{1'0'0} + \delta J_{0'1'1} - \bar\delta J_{1'1'0} - DJ_{1'1'1} - \bar\nu J_{0'0'0}\big];\\
&\bar\delta\phi_{110} - D\phi_{111} = \Psi_4\psi_{000'} + \Delta J_{0'0'1} - \frac{1}{2}\bar\delta(J_{1'0'1} + J_{0'1'1}) - \nu J_{0'0'0};\\
&\Delta\phi_{110} - \delta\phi_{111} - 2\nu\phi_{100} = \Psi_4\psi_{001'} - \bar\delta J_{1'1'1} + \frac{1}{2}\Delta(J_{0'1'1} + J_{1'0'1}) - \bar\nu J_{0'0'1} - \nu J_{1'0'0}.
\end{aligned} \tag{17}$$

One can exclude the component ϕ_{100} from the first pair of equations, obtaining the decoupled equation for ϕ_{000} with the source term:

$$\Box\phi_{000} = K_0, \qquad K_0 = 2\big[D\delta(J_{0'1'0} + J_{1'0'0}) - D^2J_{1'1'0} - \delta^2J_{0'0'0}\big]. \tag{18}$$

Using this, one can further represent a spin-vector in terms of the Debye potential as follows:

$$\psi_\mu = \begin{pmatrix} \bar{m}_\mu D\bar{\delta} - l_\mu \bar{\delta}^2 \\ l_\mu D\bar{\delta} - \bar{m}_\mu D^2 \end{pmatrix} \psi, \qquad \Box\psi = 0. \tag{19}$$

The resulting expression satisfies the system (15). Note that the equation for the Debye potential is the same for the vector field, the Weyl field and, as we will see in the next section, also for the tensor field.

7. Gravitational perturbations

Starting with the Einstein's equations and expanding the metric on the background

$$g_{\mu\nu} = g^{(0)}_{\mu\nu} + h_{\mu\nu},$$

one derives the Lichnerowitz equation for the spin 2 field in curved space-time:[14]

$$\nabla^\alpha \nabla_\alpha \psi_{\mu\nu} + 2R^\sigma{}_\mu{}^\tau{}_\nu{}^{(0)} \psi_{\sigma\tau} - 2R^\sigma{}_{(\nu}{}^{(0)} \psi_{\mu)\sigma} = 0, \qquad \nabla_\mu \psi^{\mu\nu} = 0, \tag{20}$$

where

$$\psi_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} g_{\mu\nu} h, \quad h = g^{\mu\nu} h_{\mu\nu}, \quad \psi = g^{\mu\nu} \psi_{\mu\nu}.$$

In the case of non-vacuum background, the additional gauge fixing condition $\psi = 0$ should be imposed.

To apply the NP formalism, similar splitting has to be performed in the null tetrad vectors, spin coefficients, Weyl scalars and Ricci tensor deviators, equipping the first order perturbations with an index one. The complete set of gravitational perturbation equations in the Newman-Penrose formalism is obtained by linearizing the Bianchi identities. We present here two equations from the resulting system, which are relevant for pp-waves:

$$\bar{\delta}\Psi_0^{(1)} - D\Psi_1^{(1)} = 4\pi\left[\delta T_{ll}^{(1)} - DT_{lm}^{(1)}\right], \qquad \Delta\Psi_0^{(1)} - \delta\,\Psi_1^{(1)} = 4\pi\left[\delta T_{lm}^{(1)} - DT_{mm}^{(1)}\right]. \tag{21}$$

After some manipulations using the perturbed Ricci identities in the NP formalism, we obtain the decoupled equation for the perturbation of Ψ_0 with the corresponding source term:

$$\Box\Psi_0^{(1)} = 4\pi T_0^{(1)}, \qquad T_0^{(1)} = 2\left[2D\delta T_{lm}^{(1)} - D^2 T_{mm}^{(1)} - \delta^2 T_{ll}^{(1)}\right]. \tag{22}$$

Then using the conjugate operator we can write down the solution in terms of the Debye potential:

$$\psi^{\mu\nu} = \left[2l^{(\mu}\bar{m}^{\nu)}\bar{\delta}D - \bar{m}^\mu \bar{m}^\nu D^2 - l^\mu l^\nu \bar{\delta}^2\right]\psi, \qquad \Box\psi = 0. \tag{23}$$

This expression satisfies the Lichnerowitz equation with the gauge condition (20).

As in the case of the vector field, this construction will be also valid for the non-vacuum pp-waves. This can be easily verified by the direct substitution.

8. Shock wave backgrounds

An important subclass of pp-waves is generated via boosting the black hole metrics to an infinite-momentum frame. The line element of the resulting metrics (parametrized by the real transverse coordinates x_i $i = 1, 2$) reads

$$ds^2 = \delta(u) f(\mathrm{x}) du^2 + du dv - dx_i^2. \tag{24}$$

This is an exact solution of the Einstein equations representing the gravitational shock wave. When $u \neq 0$, the space-time is flat, but for $u = 0$ it has a delta-like singularity. But the field equations on such a background are still meaningful due to lack of singularities in the metric determinant. The function $f(\mathrm{x})$ describes the gravitational wave profile. In particular, for the case of the boosted Schwarzschild solution (the Aichelburg-Sexl metric[15])

$$f(\mathrm{x}) = -8p_m \ln \rho, \quad \rho = \sqrt{x_1^2 + x_2^2},$$

where p_m denotes the energy of shock wave. For the case of a boosted Einstein-Maxwell-dilaton solution[16]

$$f(\mathrm{x}) = -8p_m \ln \rho + \frac{(3 - 4a^2)}{(1 - a^2)} \frac{\pi p_e}{\rho},$$

whetre a - the dilaton coupling constant, p_e - the electric charge. For boosted Taub-NUT[17]

$$f(\mathrm{x}) = -8p_n \tan^{-1} \frac{x_1}{x_2},$$

where p_n is the NUT charge. Also known in the literature are the boosted Kerr- Newman solution,[18] the boosted Schwarzschild-anti-de Sitter[19] space and some other metrics.

Here we present the solution of the massive Klein-Gordon equation on the background of the singular shock-wave metrics:

$$(\Box + m^2)\phi = 2(D\Delta - \delta\bar{\delta})\psi + m\psi = \left(4\partial_u \partial_v + 4\,\kappa\,\delta(u) f(\mathrm{x}) \partial_v^2 - \partial_i^2 + m^2\right) \phi = 0. \tag{25}$$

In spite of presence of the singularity, there exists an exact solution of this equation containing the Heaviside function discontinuity only:

$$\phi = \int \exp[i\,\mathcal{W}]\, d\mathcal{G}, \qquad d\mathcal{G} = \frac{d\boldsymbol{q}}{(2\pi)^2}\, d\boldsymbol{x},$$

$$\mathcal{W} = -\frac{k_v}{2}\big[v - \kappa\,\theta(u) f(\mathrm{x}')\big] - \frac{(\boldsymbol{k} - \boldsymbol{q})^2 + m^2}{2k_v} u + \boldsymbol{k}\boldsymbol{x} + \boldsymbol{q}(\boldsymbol{x}' - \boldsymbol{x}), \tag{26}$$

where we have introduced the notation $\boldsymbol{k} = (k_{x_1}, k_{x_2})$, $\boldsymbol{q} = (q_{x_1}, q_{x_2})$, $\boldsymbol{x} = (x_1, x_2)$ for the two-dimensional transverse vectors. If we put the mass equal to zero, then we obtain the solution of the equation for the Debye potential.

9. Massive field with spin $\frac{1}{2}$

Now consider other massive fields, starting from the spin 1/2. In the two-component notation, the Dirac bispinor equation $i\gamma^\mu\nabla_\mu\psi - \mu\,\psi = 0$ with $\psi = \left(\xi^A, \eta_{A'}\right)$, splits into two equations

$$\nabla_{AB'}\xi^A - i\mu\,\eta_{B'} = 0, \qquad\qquad \nabla^{AB'}\eta_{B'} - i\mu\,\xi^A = 0. \tag{27}$$

Their solution in the case of shock-wave backgrounds is a generalization of the previously obtained expression (14) for the massless field and can be written in the following form

$$\eta_{1'} = 2\int([k_{x_1} - q_{x_1}] + i[k_{x_2} - q_{x_2}] - \mu)\exp[i\,\mathcal{W}]\;d\mathcal{G}, \quad \eta_{0'} = -\int k_v \exp[i\,\mathcal{W}]\;d\mathcal{G},$$
$$\xi^0 = 2\int\Big([k_{x_1} - q_{x_1}] - i[k_{x_2} - q_{x_2}] + \mu\Big)\exp[i\,\mathcal{W}]\;d\mathcal{G}, \quad \xi^1 = \int k_v \exp[i\,\mathcal{W}]\;d\mathcal{G}, \tag{28}$$

where $\mathcal{W}$ defined in (26).

10. Proca equation

For the massive spin-1 field the gauge invariance of Maxwell's field is broken by the mass term. Instead of the gauge fixing condition, we deal with the Lorentz-like dynamical constraint, so we have two equations:

$$\nabla_\mu F^{\mu\nu} + m^2 A^\nu = 0, \qquad\qquad \nabla_\mu A^\mu = 0. \tag{29}$$

The massive vector field is no longer transverse, possessing three physical degrees of freedom. The first two polarizations in the case of shock wave background are realized by modified expressions with Debye potentials (the real and imaginary parts of the solution):

$$A^\mu_{(1,2)} = \int \mathcal{K}^\mu_{(1,2)} \exp[i\,\mathcal{W}]\;d\mathcal{G}, \quad \mathcal{K}^\mu_{(1,2)} = \Big\{0, 2[k_{x_1} - q_{x_1}] - 2i[k_{x_2} - q_{x_2}], k_v, -ik_v\Big\}. \tag{30}$$

For the third polarization we solve the dynamical constraint acting by the covariant derivative on the massive scalar field, obtaining

$$A^\mu_{(3)} = \int \mathcal{K}^\mu_{(3)} \exp[i\,\mathcal{W}]\;d\mathcal{G} + c.c,$$

$$\mathcal{K}^\mu_{(3)} = \Big\{k_v,\ \frac{(-\boldsymbol{q})^2 - m^2}{k_v} + k_v\delta(u)f(\mathrm{x}),\ \ k_{x_1} - q_{x_1},\ k_{x_1} - q_{x_1}\Big\}. \tag{31}$$

It can be seen that this expression satisfies the constraint indeed. But this expression gives us only one additional polarization, because of real multipliers in front of the exponent. Therefore the complete solution of the Proca equation in AS metric is a sum $A^\mu = A^\mu_{(1,2)} + A^\mu_{(3)}$. To write down the solution for massless electromagnetic field, it is sufficient to set the mass to zero.

11. Conclusions

This work is devoted to some novel applications of the Newman-Penrose formalism and the method of Debye potentials. Previously this technique was successfully used to construct solutions of equations for massless fields of different spins on the background of vacuum black hole solutions of Petrov type D. Here it was applied to solutions of type N. Unlike the case D, where the metric has two principal null directions and, accordingly, two decoupled equations for NP projections can be derived, in the metrics of type N only one decoupled equation exists. Namely, one can decouple the χ_0-equation for the Weyl field, the ϕ_{000}-equation for the Rarita-Schwinger field, and the equations for perturbations of Φ_0 and Ψ_0 of the vector and tensor fields respectively. It is still enough to construct the Debye representation for the solutions obtaining the universal equation for the Debye potential for all spins. We also managed to generalize our construction to the case of massive fields on the background of shock-wave metrics. Our formulas can be used for quantization in shock wave backgrounds and in some holographic applications.

Acknowledgments

The work was supported by the Russian Foundation for Basic Research on the project 20-52-18012, and the Scientific and Educational School of Moscow State University "Fundamental and Applied Space Research".

References

1. E. Newman and R. Penrose, "An Approach to gravitational radiation by a method of spin coefficients," J. Math. Phys. **3** (1962), 566-578.
2. P. Debye, Ann. Phys. (Leipz.) **30**, 57 (1909).
3. J. M. Cohen and L. S. Kegeles, "Electromagnetic fields in curved spaces - A constructive procedure," Phys. Rev. D **10** (1974), 1070-1084.
4. L. S. Kegeles and J. M. Cohen, "Constructive procedure for perturbations of spacetimes," Phys. Rev. D **19** (1979), 1641-1664.
5. P. L. Chrzanowski, "Vector Potential and Metric Perturbations of a Rotating Black Hole," Phys. Rev. D **11** (1975), 2042-2062.
6. A. Z. Petrov, "The Classification of Spaces Defining Gravitational Fields," Gen. Relativ. Gravit. **32** 1665-1685 (2000).
7. S. A. Teukolsky, "Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations," Astrophys. J. **185** (1973), 635-647.
8. G. F. Torres Del Castillo, "Debye Potentials for Rarita-schwinger Fields in Curved Space-times," J. Math. Phys. **30** (1989), 1323-1328.
9. H. W. Brinkmann, "Einstein spapces which are mapped conformally on each other," Math. Ann. **94** (1925) 119.
10. R. M. Wald, "Construction of Solutions of Gravitational, Electromagnetic, Or Other Perturbation Equations from Solutions of Decoupled Equations," Phys. Rev. Lett. **41** (1978) 203.
11. C. Sanchez-Trujillo and G. F. Torres del Castillo, "Petrov type N space-times admitting a Killing spinor," Revista Mexicana de Fisica **36** (Suppl1) (1990) 167-171.

12. K. Duztas and I. Semiz, "The decoupling problem of the Proca equation; and treatment of Dirac, Maxwell and Proca fields on the resulting pp-wave spacetimes," Gen. Rel. Grav. **48** (2016) no.7, 99.
13. H. A. Buchdahl, "On the compatibility of relativistic wave equations for particles of higher spin in the presence of a gravitational field," Nuovo Cim. **10** (1958) 96.
14. A. Lichnerowicz, in "*Relativity, Groups and Topology*", edited by C. DeWitt and B. DeWitt (Gordon and Breach Science Publishers, Inc., New York, 1964), p. 827.
15. P. C. Aichelburg and R. U. Sexl, "On the Gravitational field of a massless particle," Gen. Rel. Grav. **2** (1971), 303.
16. R. G. Cai, J. Y. Ji and K. S. Soh, "Ultrarelativistic limits of boosted dilaton black holes," Nucl. Phys. B **528** (1998), 265-282.
17. R. Argurio, F. Dehouck and L. Houart, "Boosting Taub-NUT to a BPS NUT-wave," JHEP **01** (2009), 045.
18. C. O. Lousto and N. G. Sanchez, "The Ultrarelativistic limit of the boosted Kerr-Newman geometry and the scattering of spin 1/2 particles," Nucl. Phys. B **383** (1992), 377-394.
19. J. Podolsky and J. B. Griffiths, "Impulsive gravitational waves generated by null particles in de Sitter and anti-de Sitter backgrounds," Phys. Rev. D **56** (1997), 4756-4767.

Gravitational geometric phase

Banibrata Mukhopadhyay
Department of Physics, Indian Institute of Science, Bangalore 560012, India
E-mail: bm@iisc.ac.in

Tanuman Ghosh
Raman Research Institute, Bangalore, 560080, India
E-mail: tanuman@rri.res.in

Soumya Kanti Ganguly
Ongil Private Limited, Tidel Park, Rajiv Gandhi IT Expressway, Chennai 600113, India
E-mail: soumya09ganguly@gmail.com

We show that spinors propagating in curved gravitational background acquire an interaction with spacetime curvature, which leads to a quantum mechanical geometric effect. This is similar to what happens in the case of magnetic fields, known as Pancharatnam-Berry phase. As the magnetic and gravitational fields have certain similar properties, e.g. both contribute to curvature, this result is not difficult to understand. Interestingly, while spacetime around a rotating black hole offers Aharonov-Bohm and Pancharatnam-Berry both kinds of geometric effect, a static spacetime offers only the latter. In the bath of primordial black holes, such gravity induced effects could easily be measured due to their smaller radius.

Keywords: Geometric phase; Dirac equation; semi-classical theory; classical black holes; field theory; curved spacetime.

1. Introduction

In recent works,[1,2] we showed that spinors propagating in a curved background acquire a geometric phase (GP) similar to the case of GP in electromagnetic field. The study of GP in electromagnetic field has been widely explored. Two most important notions of GPs are: Aharonov-Bohm (AB) effect[3] and Pancharatnam-Berry (PB)[4,5] phase. AB effect and PB phase occur in two different physical scenarios. In presence of a varying magnetic field, PB phase is acquired by a spinor, whereas AB effect can be acquired even without a varying field. In fact, AB effect is one of the signature properties of magnetic potential influencing a particle's quantum state. These effects/phases in case of electromagnetic fields originate from the effective curvature of magnetic potential and field in the (flat) Minkowski spacetime. We showed[2] that in a curved spacetime, a spinor in the vicinity of a gravitating body can acquire similar geometric phases which originate from the effective curvature of the spacetime itself.

Analogous effects for gravitational and magnetic fields are well known. Some of the examples are: curvature generated in both scenarios; electromagnetic radiation from accelerated charge and gravitational radiation from accelerated mass (only quadrupolar in nature), both propagating with the speed of light; energy splitting of spin-half particles in presence of both fields.[6,7] With the continuation of these analogous results, it is expected that the quantum effects, which generate the GPs in case of magnetic field, similarly should be present in case of gravitational field. Some studies have already explored these effects in local coordinates.[1,7,8] There are many astrophysical and cosmological phenomena like baryogenesis, neutrino emissions from supernovae and active galactic nuclei (AGNs), neutrino oscillation, neutrino dominated accretion disks, Fermi degenerate gas in compact objects, etc., where study of spinor propagation and evolution is crucial to understand the physical nature of these phenomena.[9–16]

When it comes to the study of gravity with quantum mechanics and understand various effects of quantum systems in curved spacetime, semi-classical formalism is a very effective tool to perform such study. In fact, there are plenty of available works in the literatures which have explored the dynamics of spinors in curved spacetime. These have been done with the Lagrangian[9,17] and Hamiltonian[6,18] formalisms. It is also well known fact that in the Hamiltonian formalism, the Dirac Hamiltonian of a spinor in curved spacetime has a non-hermiticity and uniqueness problem. Different authors tried to implement different approaches to address these issues.[6,18–22] Non-hermitian biorthogonal quantum mechanics[23] provides tool for solving the non-hermiticity issue of Dirac Hamiltonian. One approach is to define a relativistically invariant scalar product (Parker scalar product)[6,18] and use that scalar product to define the expectation values of all operators. In case of curved spacetime, this scalar product is different from the standard flat spacetime scalar product because of the presence of curved spacetime metric and vierbeins. The other approach is known as $\eta-$ representation or pseudo-Hermitian approach,[21,22] where the Hamiltonian and eigenstates of the particle are modified by the metric and vierbeins in such a way that effectively one can use the standard flat scalar product to define all the operators in the Hilbert space. In fact, these two approaches are shown to be equivalent.[21,22]

In this work, we explore the Dirac Hamiltonian in curved spacetime background, especially in Kerr and Schwarzschild geometries, using the pseudo-Hermitian approach or the η-representation and explore the possibilities of appearance of GPs in spinors traversing in curved spacetime. In section 2, we recapitulate the quantum mechanics in η-representation. Then in section 3, we derive the Dirac Hamiltonian in the Kerr geometry and further derive its non-relativistic counterpart for slowly moving particles around a weakly rotating black hole and its analogy with that in electromagnetic fields in section 4. In section 5, we find GPs in curved spacetime. Finally, in section 6, we summarize our findings.

2. Quantum Mechanics in the η-Representation

We briefly discuss the quantum mechanics in η-representation.[21,22,24,25] If an operator η satisfies the relation

$$\rho = \eta^\dagger \eta, \tag{1}$$

where ρ is an invertible operator which assures the relation to be satisfied as

$$\rho H \rho^{-1} = H^\dagger, \tag{2}$$

then the Hamiltonian in η-representation turns out to be

$$H_\eta = \eta H \eta^{-1} = H_\eta^\dagger. \tag{3}$$

The relation between the wave functions in the η–representation and in the standard representation is given by

$$\Psi = \eta\psi, \tag{4}$$

where these two wave functions satisfy the following wave equations:

$$i\frac{\partial\psi}{\partial t} = H\psi, \tag{5}$$

$$i\frac{\partial\Psi}{\partial t} = H_\eta\Psi. \tag{6}$$

As we mentioned above, the Parker scalar product,[6,18] which is a modification of standard scalar product, in Hilbert space is defined as

$$(\phi,\psi)_\rho = \int d^3x(\phi^\dagger\rho\psi), \tag{7}$$

whereas in the pseudo-Hermitian approach, the scalar product takes the standard form of flat space scalar product since the Hamiltonian and the wave functions are already modified there, which is

$$(\Phi,\Psi) = \int (\Phi^\dagger\Psi)d^3x. \tag{8}$$

We can easily verify that

$$(\phi,\psi)_\rho = (\Phi,\Psi). \tag{9}$$

3. Dirac Hamiltonian in the Kerr and Schwarzschild Metrices

We use the convention of natural unit system with $\hbar = G = c = 1$ and metric signature of $(+,-,-,-)$.

We can start from the Dirac equation in curved spacetime which is,

$$(i\gamma^\mu D_\mu - m)\psi(x) = 0. \tag{10}$$

where the covariant derivative $D_\mu = \partial_\mu + \Gamma_\mu$. Here Γ_μ is the spinorial affine connection.[6,18] m is the mass of the Dirac particle, ψ is the four-component column bispinor. γ^μ are the Dirac matrices in curved spacetime which satisfies the relation

$$\gamma^\alpha \gamma^\beta + \gamma^\beta \gamma^\alpha = 2g^{\alpha\beta} I_4, \tag{11}$$

where I_4 is the 4×4 identity matrix and $g^{\alpha\beta}$ is the contravariant metric tensor of curved spacetime. The adjoint spinor is defined as $\bar{\psi} = \psi^\dagger \gamma^0$.

The gamma matrices in curved spacetime, also known as global gamma matrices (γ^α), are related to the local flat spacetime gamma matrices (γ^a) by the relation

$$\gamma^\alpha = e^\alpha_a \gamma^a, \tag{12}$$

where e^α_a-s are the tetrads defined by

$$g_{\mu\nu} = e^a_\mu e^b_\nu \eta_{ab}. \tag{13}$$

We then derive the global Dirac Hamiltonian operator from equation (10), which is

$$\begin{aligned} H = &-i\Gamma_t - i(g^{tt})^{-1}\gamma^t \left[\gamma^r(\partial_r + \Gamma_r)\right. \\ &\left. + \gamma^\theta(\partial_\theta + \Gamma_\theta) + \gamma^\phi(\partial_\phi + \Gamma_\phi)\right] + (g^{tt})^{-1}\gamma^t m. \end{aligned} \tag{14}$$

This Hamiltonian is self-adjoint under the Parker scalar product given in equation (7). Nevertheless, our purpose is to find the Dirac Hamiltonian in η-representation which will be useful for working in standard flat Hilbert space. We choose astrophysically important metrices to work with and hence we shall find the Dirac Hamiltonian in η-representation for the Kerr metric, which reduces to the Schwarzschild metric when the Kerr parameter $a = 0$.

3.1. *Kerr metric*

The Kerr metric in the Boyer-Lindquist coordinates is

$$\begin{aligned} ds^2 = &\left(1 - \frac{2Mr}{\rho^2}\right) dt^2 + \frac{4Mra\sin^2\theta}{\rho^2} dt d\phi - \frac{\rho^2}{\Delta} dr^2 \\ &- \rho^2 d\theta^2 - \left[(r^2 + a^2)\sin^2\theta + \frac{2Mra^2\sin^4\theta}{\rho^2}\right] d\phi^2, \end{aligned} \tag{15}$$

where $\rho^2 = r^2 + a^2\cos^2\theta$, $\Delta = r^2 - 2Mr + a^2$, M is the mass of black hole and a is the Kerr parameter (angular momentum per unit mass of the black hole). We choose the Schwinger gauge of tetrad,[26–28] given by

$$\begin{aligned} e^t_0 = \sqrt{g^{tt}}, e^r_1 = \frac{\sqrt{\Delta}}{\rho}, e^\theta_2 = \frac{1}{\rho}, \\ e^\phi_3 = \frac{1}{\sin\theta\sqrt{\Delta}\sqrt{g^{tt}}}, e^\phi_0 = \frac{2Mar}{\rho^2\Delta\sqrt{g^{tt}}}. \end{aligned} \tag{16}$$

The Dirac Hamiltonian in the η-formalism in Kerr metric then turns out to be[29]

$$\begin{aligned}H_\eta = &\frac{m}{\sqrt{g^{tt}}}\gamma^0 - i\frac{\sqrt{\Delta}}{\rho\sqrt{g^{tt}}}\left(\frac{\partial}{\partial r}+\frac{1}{r}\right)\gamma^0\gamma^1\\ &- i\frac{1}{\rho\sqrt{g^{tt}}}\left(\frac{\partial}{\partial\theta}+\frac{1}{2}\cot\theta\right)\gamma^0\gamma^2 - i\frac{1}{g^{tt}\sqrt{\Delta}\sin\theta}\frac{\partial}{\partial\phi}\gamma^0\gamma^3\\ &- i\frac{2Mar}{g^{tt}\rho^2\Delta}\frac{\partial}{\partial\phi} - i\frac{1}{2}\frac{\partial}{\partial r}\left(\frac{\sqrt{\Delta}}{\rho\sqrt{g^{tt}}}\right)\gamma^0\gamma^1 - i\frac{1}{2}\frac{\partial}{\partial\theta}\left(\frac{1}{\rho\sqrt{g^{tt}}}\right)\gamma^0\gamma^2\\ &+ i\frac{\sqrt{g^{tt}}\Delta Ma\sin\theta}{2\rho}\left(\frac{\partial}{\partial r}\left(\frac{r}{g^{tt}\rho^2\Delta}\right)\gamma^3\gamma^1\right.\\ &\left.+\frac{1}{\sqrt{\Delta}}\frac{\partial}{\partial\theta}\left(\frac{r}{g^{tt}\rho^2\Delta}\right)\gamma^3\gamma^2\right).\end{aligned} \tag{17}$$

We can write this Hamiltonian as

$$H_\eta = (\sqrt{g^{tt}})^{-1}\left[\gamma^0 m + \gamma^0\gamma^j(p_j - iA_j) + i\gamma^0\gamma^j\gamma^5 k_j + e_0^\phi p_\phi\right], \tag{18}$$

where $A_1 = \frac{\sqrt{\Delta}}{\rho r} + \frac{\sqrt{g^{tt}}}{2}\frac{\partial}{\partial r}\left(\frac{\sqrt{\Delta}}{\rho\sqrt{g^{tt}}}\right)$, $A_2 = \frac{\cot\theta}{2\rho} + \frac{\sqrt{g^{tt}}}{2}\frac{\partial}{\partial\theta}\left(\frac{1}{\rho\sqrt{g^{tt}}}\right)$ and $A_3 = 0$; $k_1 = \frac{ig^{tt}Ma\sqrt{\Delta}\sin\theta}{2\rho}\frac{\partial}{\partial\theta}\left(\frac{r}{g^{tt}\rho^2\Delta}\right)$, $k_2 = -\frac{ig^{tt}Ma\Delta\sin\theta}{2\rho}\frac{\partial}{\partial r}\left(\frac{r}{g^{tt}\rho^2\Delta}\right)$ and $k_3 = 0$. Here we use the fact that $\gamma^3\gamma^2$ and $\gamma^3\gamma^1$ can be respectively written as $i\gamma^0\gamma^1\gamma^5$ and $-i\gamma^0\gamma^2\gamma^5$.

Now using the relations $H_\eta = i\frac{\partial}{\partial t} = p_t$ and $\partial_0 = e_0^t\partial_t + e_0^\phi\partial_\phi$, we rearrange the equation (18) to

$$\left[p_0 - \gamma^0\{m + \gamma^j(p_j - iA_j) + i\gamma^j\gamma^5 k_j\}\right]\Psi = 0, \tag{19}$$

where we use the tetrad transformation relation between global and local 4-momenta as

$$p_j = e_j^\mu p_\mu, \tag{20}$$

where Greek indices represent global coordinates (t, r, θ, ϕ) and roman indices represent flat coordinates $(0, 1, 2, 3)$.

Now we need to identify the terms in these equations with the analogy of magnetic field scenario which we discussed in section 1. A_js are analogous to the magnetic vector potential and k_js are "pseudo-vector" potential, which in this case of Kerr metric appears due to the chirality in the system owing to the rotation of spacetime.

If we put $a = 0$, the equations in the Kerr metric reduce to those in the Schwarzschild metric. The Hamiltonian then turns out to be

$$H_\eta = (\sqrt{g^{tt}})^{-1}\left[\gamma^0 m + \gamma^0\gamma^j(p_j - iA_j^s)\right], \tag{21}$$

where $A_1^s = \frac{1}{r}\sqrt{1-\frac{2M}{r}} + \frac{\sqrt{g^{tt}}}{2}\frac{\partial}{\partial r}\left(\frac{\sqrt{1-\frac{2M}{r}}}{\sqrt{g^{tt}}}\right)$, $A_2^s = \frac{\cot\theta}{2r}$ and $A_3^s = 0$.

We can also rearrange the Hamiltonian for $a = 0$ to write the Dirac equation in a compact form as

$$\left[p_0 - \gamma^0\{m + \gamma^j(p_j - iA_j^s)\}\right]\Psi = 0. \tag{22}$$

Here A_j^ss are the "gravito-magnetic potential" in the Schwarzschild geometry similar to the A_js in Kerr metric.

4. Nonrelativistic Approximation of Dirac Hamiltonian

In the standard Dirac representation, we write the Dirac equation in Kerr metric from equation (19) as

$$\begin{pmatrix} -p_0 + i\vec{\sigma}\cdot\vec{k} + m & \vec{\sigma}\cdot\vec{\Pi}_A \\ \vec{\sigma}\cdot\vec{\Pi}_A & -p_0 + i\vec{\sigma}\cdot\vec{k} - m \end{pmatrix}\Psi = 0, \tag{23}$$

where

$$\begin{aligned} \vec{k} &= (k_1, k_2, 0), \\ \vec{\Pi}_A &= \vec{p} - i\vec{A}. \end{aligned} \tag{24}$$

Following the standard process of deriving the non-relativistic limit of the Hamiltonian (see reference 30) we can write the coupled equations as

$$(\vec{\sigma}\cdot\vec{\Pi}_A)\Psi_B = (E - i\vec{\sigma}\cdot\vec{k} - m)\Psi_A, \tag{25a}$$

$$(\vec{\sigma}\cdot\vec{\Pi}_A)\Psi_A = (E - i\vec{\sigma}\cdot\vec{k} + m)\Psi_B. \tag{25b}$$

where E is the energy eigenvalue, Ψ_A and Ψ_B are the two components of wavefunction Ψ and we take into account the fact that $\vec{A}$ and $\vec{k}$ are time-independent.

We now assume slowly rotating spacetime and non-relativistic particles such that the particle velocity $v << c$, hence

$$E \sim m, \qquad |\vec{k}| << m \tag{26}$$

and define the non-relativistic particle energy,

$$E^{NR} = E - m. \tag{27}$$

It can easily be derived that by keeping only leading order terms by combining equations (25a) and (25b), we obtain

$$\frac{1}{2m}(\vec{\sigma}\cdot\vec{\Pi}_A)(\vec{\sigma}\cdot\vec{\Pi}_A)\Psi_A = (E^{NR} - i\vec{\sigma}\cdot\vec{k})\Psi_A, \tag{28}$$

which can be further written as

$$\left[\frac{\Pi^2}{2m} + \sigma\cdot(i\vec{k} - \frac{i}{2m}\vec{B}_g)\right]\Psi_A = E^{NR}\Psi_A, \tag{29}$$

where $\vec{B}_g = \vec{\nabla}\times\vec{A}$ is the effective "gravito-magnetic field" and $\vec{A}$ is "gravito-magnetic potential". This implies that

$$H^{NR}\Psi_A = \left[\frac{\Pi^2}{2m} + \vec{\sigma}\cdot\vec{B}_g^{kerr}\right]\Psi_A, \tag{30}$$

where $\vec{B}_g^{kerr} = (i\vec{k} - \frac{i}{2m}\vec{B}_g)$ which involves both field analogue term and potential analogue term.

Similarly for $a = 0$ (Schwarzschild) case, we can find that

$$H^{sNR}\Psi_A = \left[\frac{\Pi^{s\,2}}{2m} + \vec{\sigma}\cdot\vec{B}_g^{sch}\right]\Psi_A, \tag{31}$$

where $\vec{B}_g^{sch} = -\frac{i}{2m}\vec{B}_g^s$ where $\vec{B}_g^s = \vec{\nabla}\times\vec{A}^s$.

Note importantly that $\vec{B}_g^{kerr}$ involves the field $(\vec{B}_g)$ and potential $(\vec{k})$ both, whereas $\vec{B}_g^{sch}$ involves only field.

4.1. *Analogy with electromagnetism*

In the presence of electromagnetic field, the Dirac equation is given by

$$\left[i\gamma^\mu\left(\partial_\mu - ieA_\mu\right) - m\right]\psi = 0, \tag{32}$$

where e is the electric charge and A_μ is the electromagnetic covariant 4-vector potential. For the non-trivial solution for ψ, the energies/Hamiltonians of the spin-up and spin-down particles are given by

$$(H + eA_0)^2 = (\hat{p} - e\vec{A})^2 + m^2 + e\vec{\sigma}\cdot\vec{B}, \tag{33}$$

where A_0 is the temporal component of A_μ which is basically the Coulomb potential and $\hat{p}$ is the quantum mechanical momentum operator $-i\nabla$. In the non-relativistic limit, equation (33) reduces to[1]

$$H = -eA_0 \pm \left[\frac{(\hat{p} - e\vec{A})^2}{2m} + m + \frac{e\vec{\sigma}\cdot\vec{B}}{2m}\right]. \tag{34}$$

Clearly equation (34) is very similar to equation (31) and, hence, whatever effects have been proposed for electromagnetism, the same are expected to suffice for gravitation.

Apart from the split due to the positive and negative energy solutions, clearly there is an additional split in the respective energy levels in equations (31) and (34). This is basically the Zeeman-splitting (or Zeeman-like for gravitation) governed by the term with Pauli's spin matrix, in the up- and down-spinors for the positive and negative energy spinors induced by magnetic/gravitational fields, whether we choose relativistic or non-relativistic regimes.

5. Geometric Phase

To find the geometric effects, we construct Poincaré sphere by $\vec{B}_g^{kerr}$. Essentially, the interaction between the spinor and the gravitational background comes from the interaction Hamiltonian of

$$H_{int} = \vec{\sigma}\cdot\vec{B}, \tag{35}$$

where $\vec{B} = \vec{B}_g^{kerr}$. We can write this Hamiltonian as

$$H_{int} = |\vec{B}| \begin{pmatrix} \cos\zeta & \sin\zeta \exp(-i\xi) \\ \sin\zeta \exp(+i\xi) & -\cos\zeta \end{pmatrix}, \tag{36}$$

where ζ and ξ are the latitude and azimuthal angles of spherical polar coordinates respectively of the parameter space constructed by the vector $\vec{B}$. Here ρ is the radial coordinate in this system.

Now, using the standard definition of phase (Φ_B) and connection ($\vec{A}_B$) (see reference 2 for details) we find that the connection is

$$\vec{A}_B = \frac{(1-\cos\zeta)}{2\rho\sin\zeta}\hat{\xi}, \tag{37}$$

and the phase is

$$\Phi_B = \frac{\tilde{\xi}}{2}(1-\cos\zeta) = \frac{\Omega}{2}, \tag{38}$$

where $\tilde{\xi}$ is the total integrated azimuthal coordinate and Ω is the integrated solid angle.

Thus the GP in curved spacetime around Kerr (and Schwarzschild) geometry takes the similar form as revealed in the case of magnetic field but the angle coordinates and phases are fixed by the spacetime geometry. It is important to note that in the Kerr metric case, the GP is a combination of AB effect which appears due to the potential like term $\vec{k}$ and PB phase which appears due to the field like term $\vec{\nabla} \times \vec{A}$. In case of Schwarzschild metric however only PB contributes to the GP.

5.1. *Possible measurement*

Previous studies[31,32] have proven the relevance of semi-classical effects in gravitational background for massive particles. In this paper, we have shown that origin of GPs in spinors traversing in curved spacetime is theoretically possible. However, the question remains that how effectively these effects can be measured. For that, we need to estimate some of the length scales of the systems. Usually, quantum effects become prominent when length scale (l) of the system becomes comparable to or less than the de-Broglie wavelength of the particle $\lambda = \hbar/p$, where p is the momentum. In the case of gravity, this l is typically the radius of the gravitational body. We find that although the appearance of GPs is theoretically possible in all conditions, it may be detected only in selective cases like, e.g. primordial black holes (see reference 2 for details). The qualitative reason is simple. Larger the size of the gravitating body, smaller the field $\vec{B}$ (or $\vec{A}$ or $\vec{k} \sim 1/r$) is and thus smaller the geometric effects are. Hence, for smaller black holes these effects are prominent, but not for astrophysical black holes.

6. Summary

It is found that the Dirac particles interacting with spacetime curvature give rise to effective spin-orbit coupling which eventually manifests geometric phases/effects: Aharonov-Bohm (AB) effect and Pancharatnam-Berry (PB) phase. AB term originates from the spin of the background geometry which is responsible for the chirality in the system and this effect goes off when black hole spin becomes zero. PB phase appears in the Kerr metric as well as in spherically symmetric static spacetime like Schwarzschild geometry. Although theoretically these effects can appear in all cases, detection of such semi-classical effects is observationally possible, at least for non-relativistic particles, in a scenario where a nonrelativistic massive particle moves around a primordial black hole. Future mission by *Fermi* satellite can prove the existence of primordial black holes by detecting small interference pattern within gamma-ray bursts. Although the results of this work are outcome for a special case of nonrelativistic particles and weakly rotating black holes, it is expected that a covariant formalism will provide similar results in a general gravitational background from the analogy of covariant formalism of study of GPs in magnetic fields.[33]

References

1. B. Mukhopadhyay and S. K. Ganguly, *Universe*, **6(10)**, 160 (2020), arXiv:1802.10377 [gr-qc]
2. T. Ghosh and B. Mukhopadhyay, *IJMPD*, **30**, 12 (2021), arXiv:2010.05945 [gr-qc]
3. Y. Aharonov and D. Bohm, *Phys. Rev.*, **115**, 485 (1959)
4. S. Pancharatnam, *Proceedings of the Indian Academy of Sciences-Section A*, **Vol. 44** (Springer, 1956) pp. 398–417
5. M. V. Berry, *Proc. Roy. Soc. Lon. A*, **392**, 45 (1984)
6. L. Parker, *Phys. Rev. D*, **22**, 1922 (1980)
7. B. Mukhopadhyay, *Exploring the Universe: From Near Space to Extra-Galactic*, **Vol. 53**, edited by B. Mukhopadhyay and S. Sasmal (2018) p. 3
8. K. Dixit, J. Naikoo, B. Mukhopadhyay, and S. Banerjee, *Phys. Rev. D*, **100**, 055021 (2019), arXiv:1903.05664 [hep-ph]
9. B. Mukhopadhyay, *MPLA*, **20**, 2145 (2005), arXiv:astro-ph/0505460 [astro-ph]
10. B. Mukhopadhyay, *Class. Quantum Grav.*, **24**, 1433 (2007), arXiv:gr-qc/0702062 [gr-qc]
11. V. Alan Kostelecký and M. Mewes, *Phys. Rev. D*, **69**, 016005 (2004), arXiv:hep-ph/0309025 [hep-ph]
12. D. Píriz, M. Roy, and J. Wudka, *Phys. Rev. D*, **54**, 1587 (1996), arXiv:hep-ph/9604403 [hep-ph]
13. W.-X. Chen and A. M. Beloborodov, *Astrophys. J.* **657**, 383 (2007), arXiv:astro-ph/0607145 [astro-ph]
14. C. Y. Cardall, M. Prakash, and J. M. Lattimer, *Astrophys. J.*, **554**, 322 (2001), arXiv:astro-ph/0011148 [astro-ph]
15. U. Das and B. Mukhopadhyay, *JCAP*, **06**, 050, arXiv:1404.7627 [astro-ph.SR]
16. G. B. Cook, S. L. Shapiro, and S. A. Teukolsky, *Astrophys. J.*, **424**, 823 (1994)
17. J. Schwinger *Particles, sources and fields, Vol. 1*, Redwood City, California (Addison-Wesley Publishing Company, INC. 1973)
18. X. Huang and L. Parker, *Phys. Rev. D*, **79**, 024020 (2009)

19. Y. N. Obukhov, *Phys. Rev. Lett.*, **86**, 192 (2001)
20. Y. N. Obukhov, A. J. Silenko, and O. V. Teryaev, *Phys. Rev. D*, **80**, 064044 (2009)
21. M. V. Gorbatenko and V. P. Neznamov, *Phys. Rev. D*, **82**, 104056 (2010)
22. M. V. Gorbatenko and V. P. Neznamov, *Phys. Rev. D*, **83**, 105002 (2011)
23. D. C. Brody, *J. Phys. A: Math. and Th.*, **47**, 035305 (2013)
24. C. M. Bender, D. C. Brody, and H. F. Jones, *Phys. Rev. Lett.*, **89**, 270401 (2002)
25. A. Mostafazadeh, *J. Math. Phys.*, **43**, 205 (2002), arXiv:math-ph/0107001 [math-ph]
26. J. Schwinger, *Phys. Rev.*, **130**, 800 (1963)
27. M. V. Gorbatenko and V. P. Neznamov, *A Modified Method for Deriving Self-Conjugate Dirac Hamiltonians in Arbitrary Gravitational Fields and Its Application to Centrally and Axially Symmetric Gravitational Fields*, 2011, arXiv:1107.0844 [gr-qc]
28. V. P. Neznamov and V. E. Shemarulin, *Gravitation and Cosmology*, **24**, 129 (2018), arXiv:1806.03835 [gr-qc]
29. M. V. Gorbatenko and V. P. Neznamov, *Annalen der Physik*, **526**, 491 (2014), https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.201400035
30. J. Sakurai, *Advanced Quantum Mechanics*, Always learning (Pearson Education, Incorporated, 1967)
31. Y. Wu, and G. Meng, *Chinese Phys. Lett.*, **12**, 193 (1995)
32. R. Colella, A. W. Overhauser, and S. A. Werner, *Phys. Rev. Lett.*, **34**, 1472 (1975)
33. M. Stone, V. Dwivedi, and T. Zhou, *Phys. Rev. D*, **91**, 025004 (2015)

Retarded potentials and radiation in odd dimensions

D. V. Gal'tsov* and M. Khlopunov*,†

*Faculty of Physics, Lomonosov Moscow State University, Moscow, 119899, Russia

†Institute of Theoretical and Mathematical Physics, Lomonosov Moscow State University, Moscow 119991, Russia

E-mail: *galtsov@phys.msu.ru, †khlopunov.mi14@physics.msu.ru

Free massless fields of any spin in flat D-dimensional spacetime propagate at the speed of light. But the retarded fields produced by the corresponding point-like moving sources share this property only for even D. Since the Green's functions of the d'Alembert equation are localized on the light cone in even-dimensional spacetime, but not in odd dimensions, extraction of the emitted part of the retarded field in odd D requires some care. We consider the wave equations for spins 0, 1, and 2 in five-dimensional spacetime and analyze the fall-off conditions for the retarded fields at large distances. It is shown that the farthest part of the field contains a component propagating at the speed of light, while the non-derivative terms propagate with all velocities up to that of light. The generated radiation will contain a radiation tail corresponding to the complete prehistory of the source's motion preceding the retarded moment of time. We also demonstrate that dividing the Green's function into a part localized on the light cone and another part that is not zero inside the light cone gives separately the divergent terms in the Coulomb field of a point source. Their sum, however, is finite and corresponds to the usual power-law behaviour.

Keywords: Extra dimensions, radiation, scalar field, electromagnetic field, gravitational waves

1. Introduction

The recent interest in the theory of radiation in spacetime dimensions other than four is mostly related to the development of the theories with extra dimensions of spacetime. While the superstring theory, pretending to the status of fundamental theory, predicts the existence of extra dimensions, there is a number of phenomenological multi-dimensional gravity theories[1–4] solving some problems of elementary particles physics and cosmology. However, the characteristics of extra dimensions, such as their number, geometry and size, vary widely from one theory to another.

The actively developing gravitational-wave astronomy is the one of the most promising tools to probe the extra dimensions. So, the first constraints on the characteristics of extra dimensions have already been obtained by use of the GW170817/GRB170817A event data.[5–7] The possibility of using the future gravitational-wave observatories, such as LISA, to constrain the extra dimensions on cosmological scales has also been discussed.[8,9] Also, it is worth to note the recent advances in the constraining extra dimensions by the photograph of the supermassive black hole M87* shadow.[10,11]

However, in most of the literature, only the radiation in even-dimensional spacetimes is considered,[12–16] while the odd dimensions have been mainly discussed in the context of radiation reaction force.[17–20] It is mostly due to the Huygens principle violation in odd dimensions known since the classical works of Hadamard, Courant and Hilbert, Ivanenko and Sokolov.[21–23] In any dimensions, the signal from the instant flash of the source reaches an observer in the interval of time required for it to propagate at the speed of light. However, in odd dimensions, the endless tail signal decaying with time is observed after that, which is not the case in even dimensions. Mathematically, the Huygens principle violation consists in odd-dimensional retarded Green's functions being localised not only on the light cone, as they are in even dimensions, but also inside it. As a result, the retarded fields in odd dimensions propagate in space with all velocities up to that of light. However, free massless fields propagate exactly at the speed of light in any dimensions. Therefore, there is the apparent mismatch as the radiation being the free field far from the source is determined by its retarded field.

In this paper, we demonstrate that despite the Huygens principle violation in odd dimensions the radiation can be computed by the integration of the energy-momentum flux in the wave zone. We use the Rohrlich-Teitelboim radiation definition[24–26] (see, also,[27–29]) based on the Lorentz-invariant decomposition of the on-shell energy-momentum tensor. Considering the radiation of spin-0, spin-1 and spin-2 fields in five spacetime dimensions we show that the emitted part of the field energy-momentum propagates in space exactly at the speed of light, while it depends on the entire history of the source motion preceding the retarded time, in contrast with the four-dimensional theory.

The paper is organised as follows. In second section we consider the scalar radiation from the point charge in five spacetime dimensions. We briefly recall the recurrent relation for the odd-dimensional retarded Green's functions and the Rohrlich-Teitelboim approach to radiation. Based on the latter, we compute the emitted part of the on-shell energy-momentum tensor. In Sec. 3, we consider the analogous problem for the five-dimensional electromagnetic field. Section 4 is devoted to the calculation of the gravitational radiation from the point particle moving along an arbitrary world line in five dimensions. In the last section we briefly discuss our results.

2. Scalar radiation in five dimensions

Action of the massless scalar field $\varphi(x)$ interacting with the massive point particle moving along an arbitrary world line $z^\mu(\tau)$ in the five-dimensional Minkowski spacetime is written as

$$S = \frac{1}{4\pi^2}\int d^{4+1}x\,\partial^\mu\varphi(x)\partial_\mu\varphi(x) - \int d\tau\sqrt{\dot z^\alpha \dot z_\alpha}(m + g\varphi(z)), \quad \dot z^\mu = \frac{dz^\mu}{d\tau}, \qquad (1)$$

where m is the particle's mass, g its scalar charge, and Minkowski metric is $\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1,-1)$. We assume that particle's motion is governed by

the external forces and is not affected by the radiation of scalar field, so here the world line variables are non-dynamical.

The action (1) leads to the scalar field equation of motion in form

$$\Box\varphi(x) = -2\pi^2 j(x), \tag{2}$$

$$j(x) = g\int d\tau\sqrt{\dot{z}^\alpha \dot{z}_\alpha}\,\delta^{(4+1)}(x-z), \tag{3}$$

where $j(x)$ is the scalar current, and $\Box = \partial^\mu\partial_\mu$ is the d'Alembert operator on the five-dimensional Minkowski space. To determine the radiation energy-momentum flux carried by the scalar field, we use its canonical energy-momentum tensor

$$T_{\mu\nu}(x) = \frac{1}{2\pi^2}\left(\partial_\mu\varphi\partial_\nu\varphi - \frac{1}{2}\eta_{\mu\nu}\partial^\alpha\varphi\partial_\alpha\varphi\right). \tag{4}$$

The retarded solution of the Eq. (2) is given as

$$\varphi(x) = -2\pi^2\int d^{4+1}x'\, j(x')G^{4+1}_{\rm ret}(x-x'), \tag{5}$$

where $G^{4+1}_{\rm ret}(x)$ is the retarded Green's function of the five-dimensional d'Alembert equation. It is defined by the equation

$$\Box G^{4+1}_{\rm ret}(x) = \delta^{(4+1)}(x), \tag{6}$$

$$G^{4+1}_{\rm ret}(x) = 0,\ x^0 < 0. \tag{7}$$

In the odd-dimensional Minkowski spacetimes, retarded Green's functions are determined by the following recurrent relation[23] (see, also,[30])

$$G^{2\nu+1}_{\rm ret}(x) = \frac{(-1)^{\nu-1}}{(2\pi)^\nu}\frac{d^{\nu-1}}{(rdr)^{\nu-1}}\frac{\theta(t)\,\theta(t^2-r^2)}{\sqrt{t^2-r^2}},\ \nu\in\mathbb{N}, \tag{8}$$

where $t = x^0$ and $r = |\mathbf{x}|$. Considering the expression under derivatives in Eq. (8) as a product of separate distributions and taking into that $d\theta(x)/dx = \delta(x)$, we find the five-dimensional retarded Green's function as

$$G^{4+1}_{\rm ret}(x) = \frac{\theta(t)}{2\pi^2}\left[\frac{\delta(t^2-r^2)}{(t^2-r^2)^{1/2}} - \frac{1}{2}\frac{\theta(t^2-r^2)}{(t^2-r^2)^{3/2}}\right]. \tag{9}$$

As discussed above, it is localised not only on the light cone, but also inside it, leading to the propagation of the retarded field in space with all velocities up to that of light. Also, retarded field depends on the entire history of the source's motion preceding the retarded time and is given by the sum of separately divergent on the light cone $t^2 - r^2 = 0$ terms. However, one can show that the divergences contained in each of the terms in Eq. (9) mutually cancel out and the resulting field is finite.

2.1. *Field of a static charge*

Let us consider the case of static particle, to demonstrate the absence of divergences in the retarded field. In our calculations, we will use the "finite lifetime" trick to show the cancellation of divergences between the two terms.

We assume that the particle is at the origin and it is "switched on" for a finite interval of the proper time $a < \tau < b$, where $a < 0$ and $b > 0$ for concreteness. For this interval its world line has the form $z^\mu = [\tau, 0, 0, 0, 0]$. Then, by use of Eqs. (5), (3) and (9) the scalar field is written as

$$\varphi(x) = \frac{g}{2}\int_a^b d\tau \left[\frac{\theta(t-\tau-r-\epsilon)}{[(t-\tau)^2-r^2]^{3/2}} - \frac{\delta(t-\tau-r-\epsilon)}{r[(t-\tau)^2-r^2]^{1/2}}\right], \tag{10}$$

where we introduced the regularising parameter $\epsilon \to +0$ into the delta and Heaviside functions to shift the divergences from the light cone. Preforming the integration we obtain

$$\varphi(x) = \frac{g}{2}\begin{cases} 0, \ t < a + r, \\ -\dfrac{t-a}{r^2[(t-a)^2-r^2]^{1/2}}, \ a + r \le t < b + r, \\ \dfrac{t-b}{r^2[(t-b)^2-r^2]^{1/2}} - \dfrac{t-a}{r^2[(t-a)^2-r^2]^{1/2}}, \ t \ge b + r. \end{cases} \tag{11}$$

In the limit of eternal particle, we arrive at the finite Coulomb-like field

$$\lim_{a,b\to\pm\infty} \varphi(x) = -\frac{g}{2r^2}, \tag{12}$$

with the power-law behaviour corresponding to the increased dimensionality of space. Similar cancellation of divergences has been shown to take place, also, for the moving particle.[30]

2.2. *The Rohrlich-Teitelboim radiation definition*

The structure of the odd-dimensional Green's functions (8) makes the extraction of the emitted part of the retarded field in a standard manner non-trivial and requires a more sophisticated approach.

Such an approach was suggested by Rohrlich[24,25] and Teitelboim[26] (see, also,[27–29]). It is based on the use of certain covariantly defined quantities, so we briefly recall their definitions. Let us consider the point particle moving along the world line $z^\mu(\tau)$ with velocity $v^\mu = dz^\mu/d\tau$ in the D-dimensional spacetime. The observation point coordinates are denoted as x^μ. Assume the observation point to be a top of the light cone in the past and denote the intersection point of this light cone with the particle's world line as $z^\mu(\hat{\tau}) \equiv \hat{z}^\mu$. The corresponding moment of proper time $\hat{\tau}$ is called the retarded proper time and is defined by equation

$$(x^\mu - \hat{z}^\mu)^2 = 0, \ x^0 \ge \hat{z}^0. \tag{13}$$

Further, all the hatted quantities correspond to that moment. Then, we introduce three spacetime vectors: a lightlike vector $\hat{X}^\mu = x^\mu - \hat{z}^\mu$ directed from the retarded

point of the world line to the observation point, a spacelike vector $\hat{u}^\mu$ orthogonal to the particle's velocity at the retarded moment of proper time, and a lightlike vector $\hat{c}^\mu = \hat{u}^\mu + \hat{v}^\mu$ aligned with the vector $\hat{X}^\mu$. Introduced vectors have the following properties

$$\hat{X}^2 = 0, \quad (\hat{u}\hat{v}) = 0, \quad \hat{u}^2 = -\hat{v}^2 = -1, \quad \hat{c}^2 = 0, \tag{14}$$

where $(\hat{u}\hat{v}) \equiv \hat{u}^\alpha \hat{v}_\alpha$. Using these vectors we, also, introduce the Lorentz-invariant distance $\hat{\rho}$ being the scalar product of two spacetime vectors

$$\hat{\rho} \equiv (\hat{v}\hat{X}), \quad \hat{X}^\mu = \hat{\rho}\hat{c}^\mu. \tag{15}$$

It is equal to the spatial distance in the Lorentz frame comoving with the particle at the retarded proper time. If the particle moves inside the compact region of space, then the Lorentz-invariant distance $\hat{\rho}$ is equivalent to the spatial distance

$$\hat{\rho} \xrightarrow{r \gg |\mathbf{\hat{z}}|} r. \tag{16}$$

In accordance with the Rohrlich-Teitelboim approach, it is the Lorentz-invariant distance $\hat{\rho}$ that is used in the long-range expansion of tensors and definition of the wave zone.

In the Rohrlich-Teitelboim approach, the radiation is determined by the most long-range part of the on-shell energy-momentum tensor expansion in the inverse powers of the Lorentz-invariant distance $\hat{\rho}$. In D dimensions, the retarded field's on-shell energy-momentum tensor is expanded as[12, 16, 26–29]

$$T^{\mu\nu} = T^{\mu\nu}_{\text{Coul}} + T^{\mu\nu}_{\text{mix}} + T^{\mu\nu}_{\text{rad}} \tag{17}$$

$$T^{\mu\nu}_{\text{Coul}} \sim \frac{A^{\mu\nu}}{\hat{\rho}^{2D-4}}, \quad T^{\mu\nu}_{\text{mix}} \sim \frac{B^{\mu\nu}}{\hat{\rho}^{2D-5}} + \ldots + \frac{C^{\mu\nu}}{\hat{\rho}^{D-1}}, \quad T^{\mu\nu}_{\text{rad}} \sim \frac{D^{\mu\nu}}{\hat{\rho}^{D-2}}. \tag{18}$$

Here, the first term $T^{\mu\nu}_{\text{Coul}}$ is the energy-momentum tensor of the deformed Coulomb-like part of the retarded field. The second one is the mixed part, which consists of more than one term for $D > 4$ and is absent in $D = 3$. The most long-range part $T^{\mu\nu}_{\text{rad}}$ of the on-shell energy-momentum tensor expansion has the properties allowing to associate it with the radiation energy-momentum:

- It is separately conserved $\partial_\mu T^{\mu\nu}_{\text{rad}} = 0$, corresponding to its dynamical independence from the other parts;
- It is proportional to the direct product of two null vectors $T^{\mu\nu}_{\text{rad}} \sim \hat{c}^\mu \hat{c}^\nu$, corresponding to its propagation exactly with the speed of light $\hat{c}_\mu T^{\mu\nu}_{\text{rad}} = 0$;
- It falls down as $T^{\mu\nu}_{\text{rad}} \sim 1/r^{D-2}$ and gives positive definite energy-momentum flux through the distant $(D-2)$-dimensional sphere of area $\sim r^{D-2}$.

Therefore, the radiation power in D-dimensions can be computed as the energy flux associated with $T^{\mu\nu}_{\text{rad}}$ through the distant $(D-2)$-dimensional sphere of radius r

$$W_D = \int T^{0i}_{\text{rad}}\, n^i\, r^{D-2}\, d\Omega_{D-2}; \quad i = \overline{1, D-1}, \tag{19}$$

where $n^i = x^i/r$ is the unit spacelike vector in the direction of observation, and $d\Omega_{D-2}$ is the angular element on the $(D-2)$-dimensional sphere. This structure holds in both even and odd dimensions with the only difference that in odd dimensions the emitted part of the energy-momentum tensor depends on the entire history of the particle's motion preceding the retarded proper time $\hat{\tau}$.

Note that, usually, the energy-momentum tensor is just the bilinear form of the field derivatives, as, e.g., in Eq. (4). Then, one can define the emitted part of the retarded field derivative, by analogy with that of the energy-momentum tensor,

$$(\partial_\mu \Phi)^{\rm rad} \sim 1/\hat{\rho}^{D/2-1}. \tag{20}$$

This definition is valid in the cases of scalar and electromagnetic fields and linearised gravity considered below.

2.3. *Emitted part of the scalar field*

Now we turn to the calculation of the emitted part of the retarded scalar field derivative. The computations below are similar to that in,[30] where we have considered the scalar synchrotron radiation in three and five spacetime dimensions.

Using the Eq. (5) together with Eqs. (9) and (3) we obtain the retarded scalar field of the moving particle as

$$\varphi(x) = -g \int d\tau\, \theta(X^0(\tau)) \left[\frac{\delta(X^2(\tau))}{(X^2(\tau))^{1/2}} - \frac{1}{2} \frac{\theta(X^2(\tau))}{(X^2(\tau))^{3/2}} \right], \tag{21}$$

where we introduced the vector $X^\mu(\tau) = x^\mu - z^\mu(\tau)$. In what follows, we omit its dependence on the proper time for brevity.

Its derivative is found to have the form

$$\partial_\mu \varphi(x) = -2g \int d\tau\, \theta(X^0) \left[\frac{3}{4} \frac{\theta(X^2)}{(X^2)^{5/2}} + \frac{\delta'(X^2)}{(X^2)^{1/2}} - \frac{\delta(X^2)}{(X^2)^{3/2}} \right] X_\mu, \tag{22}$$

where $\delta'(x) = d\delta(x)/dx$. Integrating by parts the term containing derivative of delta function by use of the relation

$$\frac{dX^2}{d\tau} = -2(vX), \tag{23}$$

we obtain the scalar field derivative in form

$$\begin{aligned} \partial_\mu \varphi(x) = -g \int d\tau\, \theta(X^0) \Bigg[& \frac{3}{2} \frac{\theta(X^2)}{(X^2)^{5/2}} X_\mu - \frac{\delta(X^2)}{(X^2)^{3/2}} X_\mu - \\ & - \frac{\delta(X^2)}{(vX)^2 (X^2)^{1/2}} \left[(aX) - 1 \right] X_\mu - \frac{\delta(X^2)}{(vX)(X^2)^{1/2}} v_\mu \Bigg], \end{aligned} \tag{24}$$

where we introduced the acceleration vector $a^\mu = d^2 z^\mu / d\tau^2$.

To obtain its emitted part, we extract the leading $\hat{\rho}$-asymptotic of the Eq. (24). We transform the products of delta and Heaviside functions by use of the relation for the delta function of complex argument

$$\theta(X^0)\delta(X^2) = \frac{\delta(\tau - \hat{\tau})}{2\hat{\rho}}. \tag{25}$$

We, also, rewrite the vector X^μ in terms of the Lorentz-invariant distance

$$X^\mu = Z^\mu + \hat{\rho}\hat{c}^\mu, \quad Z^\mu = \hat{z}^\mu - z^\mu. \tag{26}$$

Then, by use of Eqs. (25) and (26) we arrive at the emitted part of the scalar field derivative in form

$$(\partial_\mu \varphi(x))^{\rm rad} = -\frac{g\hat{c}_\mu}{2^{5/2}\hat{\rho}^{3/2}} \int_{-\infty}^{\hat{\tau}} d\tau \left[\frac{3}{2}\frac{1}{(Z\hat{c})^{5/2}} - \frac{\delta(\tau-\hat{\tau})}{(Z\hat{c})^{3/2}} - \frac{2(a\hat{c})\delta(\tau-\hat{\tau})}{(v\hat{c})^2(Z\hat{c})^{1/2}}\right], \tag{27}$$

where we have taken into account that up to the leading order $X^2 \sim 2\hat{\rho}(Z\hat{c})$ and $(vX) \sim \hat{\rho}(v\hat{c})$. Note that each integral in the Eq. (27) diverges on the upper integration limit $\tau \to \hat{\tau}$, as $(Z\hat{c}) \to (\hat{v}\hat{c})(\hat{\tau}-\tau) = (\hat{\tau}-\tau)$. However, one can show that the last two terms of the integrand do not carry physical information concerning the field in the wave zone and just eliminate the divergences contained in the first one. To do this, we introduce the regularising parameter $\epsilon \to +0$ into the argument of delta function and perform integration by use of it, obtaining the divergent result

$$\int_{-\infty}^{\hat{\tau}} d\tau \, \frac{\delta(\tau-\hat{\tau}+\epsilon)}{(Z\hat{c})^{3/2}} = \frac{1}{\epsilon^{3/2}}, \tag{28}$$

which can be rewritten as

$$\frac{1}{\epsilon^{3/2}} = \frac{3}{2}\int_{-\infty}^{\hat{\tau}-\epsilon} \frac{d\tau}{(\hat{\tau}-\tau)^{3/2}}. \tag{29}$$

By analogy, the second term with the delta function transforms to

$$\int_{-\infty}^{\hat{\tau}} d\tau \, \frac{(a\hat{c})\delta(\tau-\hat{\tau})}{(v\hat{c})^2(Z\hat{c})^{1/2}} = \frac{1}{2}\int_{-\infty}^{\hat{\tau}-\epsilon} d\tau \frac{(\hat{a}\hat{c})}{(\hat{\tau}-\tau)^{3/2}}. \tag{30}$$

Then, omitting the regularising parameter in the upper integration limit we find the emitted part of the scalar field derivative as

$$(\partial_\mu \varphi(x))^{\rm rad} = -\frac{g\hat{c}_\mu}{2^{5/2}\hat{\rho}^{3/2}} \int_{-\infty}^{\hat{\tau}} d\tau \left[\frac{3}{2}\frac{1}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{1}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{31}$$

In practice, calculating the radiation from the particle moving along some trajectory one needs to make the convergence of the integral in Eq. (31) explicit. To achieve it, one has to reintroduce the regularising parameter $\epsilon \to +0$ into the upper integration limit and perform some transformation of the first term in the integrand: usually, it is the integrations by parts, which extract the divergences from it and cancel them out by use of the remaining two integrals. The above structure of the emitted part of the field holds in any odd spacetime dimensions.[30]

Substituting the emitted part of the scalar field derivative into the Eq. (4) we find the radiated part of the energy-momentum tensor

$$T^{\rm rad}_{\mu\nu}(x) = \frac{g^2\hat{c}_\mu\hat{c}_\nu}{64\pi^2\hat{\rho}^3}\mathcal{A}^2(x), \tag{32}$$

$$\mathcal{A}(x) = \int_{-\infty}^{\hat{\tau}} d\tau \left[\frac{3}{2}\frac{1}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{1}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{33}$$

In accordance with the Rohrlich-Teitelboim approach, it is proportional to the direct product of two null vectors $\hat{c}_\mu \hat{c}_\nu$ corresponding to the propagation of this part of field's energy-momentum exactly with the speed of light. Also, as was discussed above, the emitted part of the energy-momentum tensor depends on the entire history of the source's motion preceding the retarded proper time $\hat{\tau}$, as well as the radiation power determined by it.

3. Electromagnetic radiation in five dimensions

Having found the emitted part of the five-dimensional scalar field and the corresponding radiation energy-momentum tensor, let us now briefly discuss the electromagnetic and gravitational radiation in five dimensions. We start with the former.

3.1. *The setup*

Action of the electromagnetic field $A_\mu(x)$ interacting with the massive point particle moving along an arbitrary world line $z^\mu(\tau)$ in five dimensions is analogous to that of the scalar field (1) and has the form

$$S = -\frac{1}{8\pi^2}\int d^{4+1}x\, F^{\mu\nu}F_{\mu\nu} - e\int d\tau\, \dot{z}^\mu A_\mu(z), \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{34}$$

where e is particle's electric charge. Here, we omit the particle's kinetic term in the action, given that its motion is completely governed by the external forces and, thus, the world line variables are non-dynamical.

The action (34) yields the standard equation of motion of the electromagnetic field

$$\Box A^\mu(x) = 2\pi^2 j^\mu(x), \tag{35}$$

$$j^\mu(x) = e\int d\tau\, \dot{z}^\mu\, \delta^{(4+1)}(x - z). \tag{36}$$

where we imposed on the field the Lorentz gauge condition to fix its gauge symmetry

$$\partial_\mu A^\mu = 0. \tag{37}$$

Note that this condition requires the current to be conserved

$$\partial_\mu j^\mu = 0, \tag{38}$$

and it is, when the observation point is off the world line, as can be easily seen from the Eq. (36).

To determine the energy flux carried by the electromagnetic radiation we use symmetric energy-momentum tensor of the electromagnetic field

$$\Theta_{\mu\nu} = \frac{1}{2\pi^2}\left[F_{\mu\cdot}^{\ \alpha}F_{\alpha\nu} + \frac{1}{4}\eta_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right]. \tag{39}$$

By analogy with that of the scalar field, it is the bilinear functional of the field derivatives $\Theta \sim \partial A \partial A$ and, thus, admits defining the emitted part of the electromagnetic field derivative.

3.2. *Emitted part of the electromagnetic field*

By use of the Eqs. (36) and (9) we obtain the retarded electromagnetic field of the point charge in form

$$A_\mu(x) = e\int d\tau\, \dot{z}_\mu\, \theta(X^0)\left[\frac{\delta(X^2)}{(X^2)^{1/2}} - \frac{1}{2}\frac{\theta(X^2)}{(X^2)^{3/2}}\right]. \tag{40}$$

We calculate its derivative using the relation (23)

$$\partial_\nu A_\mu(x) = e\int d\tau\, \theta(X^0)\left[\frac{3}{2}\frac{\theta(X^2)}{(X^2)^{5/2}}v_\mu X_\nu - \frac{\delta(X^2)}{(X^2)^{3/2}}v_\mu X_\nu - \right.$$
$$\left. - \frac{\delta(X^2)}{(vX)^2(X^2)^{1/2}}[(aX)-1]v_\mu X_\nu + \frac{\delta(X^2)}{(vX)(X^2)^{1/2}}[a_\mu X_\nu - v_\mu v_\nu]\right]. \tag{41}$$

Using the relations (25) and (26), and transforming the terms with delta functions by analogy with the Eqs. (28–30) we obtain the emitted part of the retarded electromagnetic field as the long-range part of its derivative with respect to the Lorentz-invariant distance

$$(\partial_\nu A_\mu)^{\rm rad} \;=\; \frac{e\hat{c}_\nu}{2^{5/2}\hat{\rho}^{3/2}}\int_{-\infty}^{\hat{\tau}} d\tau\left[\frac{3}{2}\frac{v_\mu}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{\hat{v}_\mu}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})\hat{v}_\mu - \hat{a}_\mu}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{42}$$

By analogy with the scalar field, in Eq. (42) all the physical information concerning the electromagnetic field in the wave zone is contained in the first term of the integrand, while the remaining ones just subtract the divergences contained in it at the upper integration limit.

Substituting the Eq. (42) into the symmetric energy-momentum tensor (39) we arrive at the five-dimensional electromagnetic radiation energy-momentum tensor

$$\Theta^{\rm rad}_{\mu\nu} = -\frac{e^2\hat{c}_\mu\hat{c}_\nu}{64\pi^2\hat{\rho}^3}\mathcal{A}^2_\alpha(x), \tag{43}$$

$$\mathcal{A}_\alpha(x) = \int_{-\infty}^{\hat{\tau}} d\tau\left[\frac{3}{2}\frac{v_\alpha}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{\hat{v}_\alpha}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})\hat{v}_\alpha - \hat{a}_\alpha}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{44}$$

The obtained tensor structure of the radiated part of the energy-momentum tensor corresponds to the propagation of the associated energy flux exactly with the speed of light. Also, by analogy with the scalar field (32), it depends on the history of charge's motion preceding the retarded proper time.

4. Five-dimensional gravitational radiation

Now we turn to the calculation of the gravitational radiation produced by the point particles. If interaction between them is also gravitational, the problem is not described by the linearized approximation, and the quadratic terms in expansion of the Einstein tensor are required to take into account the contribution of field stresses to radiation.[32] If the dominant forces are non-gravitational, the corresponding field stresses are also required, making the calculation rather difficult. Here we calculate

only the local contribution of particles themselves, so the result is incomplete. We give it just to reveal difference with the scalar and electromagnetic cases and to show how to deal with polarisations.

4.1. *The setup*

We start with the generally covariant Einstein-Hilbert action for the five-dimensional gravity interacting with massive point particle and some external field governing its motion

$$S = \frac{1}{2\kappa_5}\int d^{4+1}x\sqrt{-g}R - m\int d\tau\sqrt{g_{\alpha\beta}(z)\dot{z}^\alpha\dot{z}^\beta} + S_{\rm F}, \tag{45}$$

where κ_5 is the five-dimensional gravitational constant, g is the determinant of metric tensor, and $S_{\rm F}$ is the action for the external field moving point particle.

We linearise the Einstein's equation over the background Minkowski metric

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x), \quad |h_{\mu\nu}| \ll 1. \tag{46}$$

Introducing the reduced metric perturbations

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \quad h = \eta^{\alpha\beta}h_{\alpha\beta}, \tag{47}$$

we arrive at the linearised Einstein equation

$$\Box\bar{h}_{\mu\nu}(x) = 2\kappa_5\left(T^{\rm P}_{\mu\nu}(x) + T^{\rm F}_{\mu\nu}(x)\right), \tag{48}$$

$$T^{\rm P}_{\mu\nu}(x) = m\int d\tau\,\dot{z}_\mu\dot{z}_\nu\,\delta^{(5)}(x-z), \quad T^{\rm F}_{\mu\nu}(x) = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\rm F}}{\delta g^{\mu\nu}}\bigg|_{g=\eta} \tag{49}$$

where $T^{\rm P}_{\mu\nu}(x)$ is the energy-momentum of point particle moving on the flat Minkowski background and interacting with some external field, which energy-momentum tensor is given by $T^{\rm F}_{\mu\nu}(x)$. To obtain the Eq. (48) we, also, imposed the Lorentz gauge condition on the metric perturbations

$$\partial^\mu\bar{h}_{\mu\nu} = 0, \tag{50}$$

to fix the gauge symmetry of the system. Such a condition requires the particle's and external field's energy momentum-tensors to be jointly conserved

$$\partial^\mu\left(T^{\rm P}_{\mu\nu} + T^{\rm F}_{\mu\nu}\right) = 0, \tag{51}$$

which is assumed to be valid on-shell.

To determine the energy-momentum flux carried by the gravitational radiation we use the effective energy-momentum tensor of metric perturbations, by analogy with,[31]

$$t_{\mu\nu} = \frac{1}{4\kappa_5}\left\langle\partial_\mu\bar{h}^{\rm TT}_{ij}\,\partial_\nu\bar{h}^{\rm TT}_{ij}\right\rangle, \quad \bar{h}^{\rm TT}_{ij}(x) = \Lambda_{ij,kl}(\mathbf{n})\bar{h}_{kl}(x), \tag{52}$$

$$\Lambda_{ij,kl}(\mathbf{n}) = P_{ik}P_{jl} - \frac{1}{3}P_{ij}P_{kl}, \quad P_{ij} = \delta_{ij} - n_in_j, \quad n^i = \frac{x^i}{r}, \tag{53}$$

where we assume the particle's motion to be periodic, and $\langle \dots \rangle$ is the averaging over the period. Here, we turned into the transverse-traceless gauge by the contraction of metric perturbations (52) with the projector (53) defined by analogy with.[31] Note that differentiation of the projector would increase the fall-off asymptotic of the field, so its derivative can be neglected when one computes the emitted part of the gravitational field in the transverse-traceless gauge.

4.2. *Emitted part of the gravitational field (incomplete)*

Using the Eqs. (9) and (49) we find the retarded gravitational field of the point particle as

$$\bar{h}^{\rm P}_{\mu\nu}(x) = \frac{m\kappa_5}{\pi^2}\int d\tau\, \dot{z}_\mu \dot{z}_\nu\, \theta(X^0)\left[\frac{\delta(X^2)}{(X^2)^{1/2}} - \frac{1}{2}\frac{\theta(X^2)}{(X^2)^{3/2}}\right]. \tag{54}$$

We compute its derivative integrating by parts by use of the relation (23) arriving at

$$\partial_\alpha \bar{h}^{\rm P}_{\mu\nu} = \frac{m\kappa_5}{\pi^2}\int d\tau\, \theta(X^0)\left[\frac{3}{2}\frac{\theta(X^2)}{(X^2)^{5/2}} v_\mu v_\nu X_\alpha - \frac{\delta(X^2)}{(X^2)^{3/2}} v_\mu v_\nu X_\alpha - \right.$$
$$\left. - \frac{\delta(X^2)}{(vX)^2(X^2)^{1/2}}[(aX)-1] v_\mu v_\nu X_\alpha + \frac{\delta(X^2)}{(vX)(X^2)^{1/2}}[2a_{(\mu}v_{\nu)}X_\alpha - v_\mu v_\nu v_\alpha]\right], \tag{55}$$

where $a^\mu = d^2 z^\mu/d\tau^2$ is the acceleration vector, and we define the symmetrisation over two indices as $A_{(\mu}B_{\nu)} = (A_\mu B_\nu + A_\nu B_\mu)/2$.

Extracting the long-range part of the gravitational field derivative by use of the relations (25) and (26) and transforming the terms with delta functions by analogy with Eqs. (28–30) we obtain the emitted part of the gravitational field as

$$\left(\partial_\alpha \bar{h}^{\rm P}_{\mu\nu}\right)^{\rm rad} = \frac{m\kappa_5 \hat{c}_\alpha}{2^{5/2}\pi^2\hat{\rho}^{3/2}}\int_{-\infty}^{\hat{\tau}} d\tau \left[\frac{3}{2}\frac{v_\mu v_\nu}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{\hat{v}_\mu\hat{v}_\nu}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})\hat{v}_\mu\hat{v}_\nu - 2\hat{a}_{(\mu}\hat{v}_{\nu)}}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{56}$$

By analogy with the scalar (31) and electromagnetic (42) fields, it consists of one integral determining the properties of the gravitational field in the wave zone and two integral being counter-terms eliminating the divergences from the first one.

Substituting the obtained expression for the emitted part of the retarded gravitational field (56) into the Eq. (52) we find the energy-momentum tensor of gravitational radiation generated by the point particle

$$\left(t^{\rm P}_{\mu\nu}\right)^{\rm rad} = \frac{m^2\kappa_5 \hat{c}_\mu \hat{c}_\nu}{128\pi^4\hat{\rho}^3}\mathcal{A}^{\rm TT}_{ij}(x)\mathcal{A}^{\rm TT}_{ij}(x), \quad \mathcal{A}^{\rm TT}_{ij}(x) \equiv \Lambda_{ij,kl}(\mathbf{n})\mathcal{A}_{kl}(x), \tag{57}$$

$$\mathcal{A}_{ij}(x) = \int_{-\infty}^{\hat{\tau}} d\tau \left[\frac{3}{2}\frac{v_i v_j}{(Z\hat{c})^{5/2}} - \frac{3}{2}\frac{\hat{v}_i\hat{v}_j}{(\hat{\tau}-\tau)^{5/2}} - \frac{(\hat{a}\hat{c})\hat{v}_i\hat{v}_j - 2\hat{a}_{(i}\hat{v}_{j)}}{(\hat{\tau}-\tau)^{3/2}}\right]. \tag{58}$$

As in the cases of scalar (32) and electromagnetic (43) radiation, the emitted part of the gravitational field energy-momentum tensor has the tensor structure corresponding to the propagation of the associated energy-momentum flux exactly at the

speed of light and depends on the entire history of the particle's motion preceding the retarded proper time $\hat{\tau}$.

5. Conclusions

In this short article, we have demonstrated that despite the Huygens principle violation in odd spacetime dimensions, radiation can be computed by the integration of the energy-momentum flux in the wave zone. However, it requires the Lorentz-invariant modification of the radiation and wave zone definitions, in accordance with the Rohrlich-Teitelboim approach. Also, due to the Huygens principle violation, the emitted part of the field depends on the history of the source's motion preceding the retarded proper time, in contrast with even dimensions, where properties of radiation at given moment of time are determined only by the state of the source at the retarded time. Another feature of the odd dimensions is that the retarded field is given by the sum of separately divergent integrals. Nevertheless, these divergence mutually cancel out and the resulting field is finite.

Based on the Rohrlich-Teitelboim approach, we have considered the radiation of the scalar, electromagnetic and gravitational fields by point particle in five dimensions. We computed the emitted parts of the fields' energy-momentum tensors. The obtained expressions have the tensor structures corresponding to the propagation of the radiated energy-momentum in space exactly with the speed of light. Also, they are given by the sum of separately divergent integrals over the history of particle's motion, one of which carries all the physical information concerning the field in the wave zone, while the remaining ones are just the counter-terms subtracting the divergences from the former. To make the convergence of the resulting tensor explicit, one has to integrate the first term by parts.

In our previous work,[30] we have considered the scalar synchrotron radiation in three and five dimensions. The results were checked by the calculation of the spectral decompositions of the radiation power, which are indifferent to the dimensionality of the spacetime. Based on this, we assume that the obtained similar expressions for the emitted parts of the electromagnetic and gravitational fields should, also, be valid.

Acknowledgements

The work of M. Kh. was supported by the "BASIS" Foundation Grant No. 20-2-10-8-1. The work of D.G. was supported by the Russian Foundation for Basic Research on the project 20-52-18012, and the Scientific and Educational School of Moscow State University "Fundamental and Applied Space Research".

References

1. N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, The Hierarchy problem and new dimensions at a millimeter, *Phys. Lett. B* **429**, 263 (1998).

2. L. Randall and R. Sundrum, A Large mass hierarchy from a small extra dimension, *Phys. Rev. Lett.* **83**, 3370 (1999).
3. L. Randall and R. Sundrum, An Alternative to compactification, *Phys. Rev. Lett.* **83**, 4690 (1999).
4. G. R. Dvali, G. Gabadadze and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Phys. Lett. B* **485**, 208 (2000).
5. L. Visinelli, N. Bolis and S. Vagnozzi, Brane-world extra dimensions in light of GW170817, *Phys. Rev. D* **97**, no.6, 064039 (2018).
6. K. Pardo, M. Fishbach, D. E. Holz and D. N. Spergel, Limits on the number of spacetime dimensions from GW170817, *JCAP* **07**, 048 (2018).
7. K. Chakravarti, S. Chakraborty, K. S. Phukon, S. Bose and S. SenGupta, Constraining extra-spatial dimensions with observations of GW170817, *Class. Quant. Grav.* **37**, no.10, 105004 (2020).
8. C. Deffayet and K. Menou, Probing Gravity with Spacetime Sirens, *Astrophys. J. Lett.* **668**, L143 (2007).
9. M. Corman, C. Escamilla-Rivera and M. A. Hendry, Constraining extra dimensions on cosmological scales with LISA future gravitational wave siren data, *JCAP* **02**, 005 (2021).
10. S. Vagnozzi and L. Visinelli, Hunting for extra dimensions in the shadow of M87*, *Phys. Rev. D* **100**, no.2, 024020 (2019).
11. I. Banerjee, S. Chakraborty and S. SenGupta, Silhouette of M87*: A New Window to Peek into the World of Hidden Dimensions, *Phys. Rev. D* **101**, no.4, 041301 (2020).
12. B. P. Kosyakov, Exact solutions of classical electrodynamics and the Yang-Mills-Wong theory in even-dimensional space-time, *Theor. Math. Phys.* **119**, 493 (1999).
13. V. Cardoso, M. Cavaglia and J. Q. Guo, Gravitational Larmor formula in higher dimensions, *Phys. Rev. D* **75**, 084020 (2007).
14. V. Cardoso, O. J. C. Dias and J. P. S. Lemos, Gravitational radiation in D-dimensional space-times, *Phys. Rev. D* **67**, 064026 (2003).
15. A. Mironov and A. Morozov, Radiation beyond four space-time dimensions, *Theor. Math. Phys.* **156**, 1209 (2008).
16. B. P. Kosyakov, Electromagnetic radiation in even-dimensional spacetimes, *Int. J. Mod. Phys. A* **23**, 4695 (2008).
17. D. V. Galtsov, Radiation reaction in various dimensions, *Phys. Rev. D* **66**, 025016 (2002).
18. P. O. Kazinski, S. L. Lyakhovich and A. A. Sharapov, Radiation reaction and renormalization in classical electrodynamics of point particle in any dimension, *Phys. Rev. D* **66**, 025017 (2002).
19. Y. Yaremko, Self-force in 2+1 electrodynamics, *J. Phys. A* **40**, 13161 (2007).
20. E. Shuryak, H. U. Yee and I. Zahed, Self-force and synchrotron radiation in odd space-time dimensions, *Phys. Rev. D* **85**, 104007 (2012).
21. J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Dover Publications, New York, 2014).
22. R. Courant and D. Hilbert, *Methods of Mathematical Physics: Partial Differential Equations*, Wiley Classics Library (Wiley, New York, 2008).
23. D. Ivanenko and A. Sokolov, *Sov. Phys. Dokl.* **36**, 37 (1940); *Classical Field Theory* (State Publishing House of Technical and Theoretical Literature, Moscow, 1951) (in Russian) [*Klassische Feldtheorie* (Akademie-Verlag, Berlin, 1953)].
24. F. Rohrlich, The definition of electromagnetic radiation, *Nuovo Cim* **21**, 811 (1961).
25. F. Rohrlich, *Classical Charged Particles* (World Scientific, Singapore, 2007).

26. C. Teitelboim, Splitting of the maxwell tensor - radiation reaction without advanced fields, *Phys. Rev. D* **1**, 1572 (1970) [erratum: *Phys. Rev. D* **2**, 1763 (1970)].
27. D. V. Gal'tsov and P. Spirin, Radiation reaction reexamined: Bound momentum and Schott term, *Grav. Cosmol.* **12**, 1 (2006).
28. B. P. Kosyakov, Radiation in electrodynamics and in Yang-Mills theory, *Sov. Phys. Usp.* **35**, 135 (1992).
29. P. A. Spirin, Massless field emission in the space-time of extra dimensions, *Grav. Cosmol.* **15**, 82 (2009).
30. D. V. Gal'tsov and M. Khlopunov, Synchrotron radiation in odd dimensions, *Phys. Rev. D* **101**, no.8, 084054 (2020).
31. M. Maggiore, *Gravitational Waves: Volume 1: Theory and Experiments*, Gravitational Waves (OUP, Oxford, 2008).
32. S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972).

Wave propagation in the anti-deSitter optical metric

D. García-Peláez

Universidad Autónoma Metropolitana Azcapotzalco,
Avenida San Pablo Xalpa 180, Azcapotzalco, Reynosa Tamaulipas, 02200 Ciudad de México, México

Universidad Panamericana,
Tecoyotitla 366. Col. Ex Hacienda Guadalupe Chimalistac, 01050 México D.F., México
E-mail: dgarciap@up.edu.mx

C. S. López-Monsalvo

Conacyt-Universidad Autónoma Metropolitana Azcapotzalco,
Avenida San Pablo Xalpa 180, Azcapotzalco, Reynosa Tamaulipas, 02200 Ciudad de México, México
E-mail: cslopezmo@conacyt.mx

In this work, we use the fact that kinematics of light propagation in a *non-dispersive medium* associated with a bi-metric spacetime is expressed by means of a 1-parameter family of contact transformations. We present a general technique to find such transformations and explore some explicit examples for Minkowski and anti-deSitter spacetimes geometries.

Keywords: Contact geometry; Wave propagation; Anti-deSitter.

1. Introduction

Soon after the advent of general relativity, it was observed that the gravitational field *bends* the path followed by a light ray. This was confirmed by Eddington after noticing that the apparent position of the stars are shifted from their expected position in the sky when observed during a solar eclipse. This *bending* phenomenon is analogous to the deviation of light rays while traveling in a medium whose refractive index changes from one place to the other. In this sense, it was shown that a class of optical media can be modeled by means of a metric tensor encoding its electromagnetic properties (Refs. 1–3). Furthermore, this analogy has evolved into the very active field of transformation optics, where the techniques and tools of differential geometry – that have been insightful in gravitational physics – have found its way into the more applied area of material science (Refs. 4–7). In addition, this has also been used in modeling analogue gravitational spacetimes such as black holes and cosmological solutions (Refs. 8–11).

Fermat's principle poses a well known problem in the calculus of variations. In the non-relativistic setting – where the notion of time is *absolute* and *universal* – it states that light travels between two given *spatial* locations following a path

minimizing a time functional. Such a perspective is clearly untenable in the context of General Relativity, where it is commonly replaced by the assumption that the path taken by light in traveling from one location to another corresponds to a null geodesic. However, albeit it remains a variational problem, its precise formulation is far more elaborate (see Theorem 7.3.1 in Ref. 12).

Similarly (cf. Theorem 7.1.2 in Ref.12), Huygens' principle is centered in the idea of light emissions being instantaneous for different observers at each instant in time (Ref. 13). In this way, light emission is described by well localized wavefronts, as they travel through three dimensional space. In the relativistic setting, the electromagnetic field on a precise event on spacetime should depend only on initial conditions originated its past null cone (Ref. 14). This has led to explore different wave phenomena where Huygens' principle is not fulfilled, most of which are due to dissipation processes where the wave distribution has not converging tails. As a consequence, the wavefronts cannot be located in exactly at a point of spacetime. With the Hadamard conjecture, this phenomenon has been related to the dimensions and curvature of spacetime (Refs. 15, 16).

One of the simplest and of most interesting curved spacetimes for theoretical physicists, is the anti-deSitter spacetime (AdS). This spacetime model has a very deep relationship with hyperbolic geometry, where its boundary at spacelike infinity leads to have strange peculiar like closed timelike curves and achornal surfaces (Refs. 17,18). Nevertheless, AdS has gained more interest in recent years, as it has showed its value in topics like the *holographic principle*, the ground state of gauged supergravity theories and its central role in high energy physics in the correspondence with the Conformal Field Theory (AdS/CFT).

In this work, we use the fact that an optical medium can be represented by a Riemannian manifold $(\mathcal{B}, g)$ where $\mathcal{B}$ is considered to be the physical space and g the optical spatial metric. The geodesic flow in its unitary tangent bundle can be represented by a contact transformation acting on its space of contact elements. This fact, allows us to describe the wavefront evolution in an optical medium solely in terms of the contact transformation and to reconstruct the geodesic flow from the Reeb vector field. This provides us with a way to construct wavefronts in optical media without directly solving the wave equation. Particularly, we use this powerful tool to reconstruct wave propagation, along with with its corresponding light rays in a $(2+1)$-dimensional AdS spacetime.

This manuscript is structured as follows. In section 2 we give the mathematical formalism to construct the technique. In section 3 we use this technique to reconstruct wavefronts and ray lights in a $(2+1)$-dimensional anti-deSitter spacetime. We find explicitly the 1-parameter family of contact transformations and solve numerically to reconstruct the wave propagation in optical media with interfaces of different refractive index.

2. Mathematical background

In this work, an optical medium will be a Riemannian manifold $(\mathcal{B}, g)$ submersed in a bi-metric Lorentzian spacetime $(M, g_0, \tilde{g})$. Where g_0 is the spacetime metric, used to low and rise index, the metric $\tilde{g}$ is the optical metric introduced by Gordon (Ref. 1), from which the effective speed of light in a medium can be read off.The manifold $\mathcal{B}$ along with his metric g, is called *the material manifold*, The metric g can be decomposed in terms of the material manifold Riemannian metric and an observer's four velocity.

The cotangent bundle of the material manifold carries a natural symplectic structure, endowed by a non-degenerated closed 2-form ω. In this sense, the Liouville's vector filed, which preserves the symplectic 2-form in terms of its Lie derivative, defines a 1-form $\lambda \in T^*(T^*\mathcal{B})$ which generates the symplectic structure. As defined by Perlick (see Definition 5.1.1 and Proposition 5.1.1 in Ref. 12) the punctured cotangent bundle $PT^*\mathcal{B}$, corresponds to a *ray-optical structure* $\mathcal{N}$. Using the inclusion map from this ray-optical structure to the cotangent bundle, the 1-form λ induces a contact 1-form η, where the pair $(\mathcal{N}, \mathcal{D})$ is a contact manifold, known as the *space of contact elements* or the *contact bundle* of $\mathcal{B}$.

Let us use the defintion of a *wave front* centered at the point $b \in \mathcal{B}$ as the hypersurface (Ref. 19)

$$F_b(t) = \{b_i \in \mathcal{B} \,|\gamma(0) = b, \gamma(t) = b_i\} \tag{1}$$

where γ is a geodesic on $(\mathcal{B}, g)$. For every point in the wave front, the contact element of $\mathcal{B}$ associated to that point is tangent to the wave front. If we consider a geodesic flow on the unitary tangent bundle of $\mathcal{B}$, there exists a unique contact element, which is perpendicular to. So the Reeb vector filed, associated to the contact 1-form of the ray-optical structure is dual to the geodesic flow in the unitary tangent bundle of $\mathcal{B}$. The Reeb vector field is also, the infinitesimal generator of a family of 1-parameter strict contact transformations $\phi_t : \mathcal{N} \to \mathcal{N}$, which preserves the contact form and thus the contact elements on $\mathcal{B}$, which are tangent to the wave fronts. In this sense, the family of transformations ϕ_t can evolve the geodesic flow preserving the wave fronts in each point.

The method is to solve the Reeb's vector field flow, associated to contact 1-form in the ray-optical structure to reconstruct numerically from it, the wave fronts and the ray lights propagating in an optical medium. When interfaces are present between media with different refractive index, no further adaptions are needed for the technique to reconstruct wave fronts and light rays while refracted. If this technique is used in a Minkowski spacetime geometry, Snell's law can be deduced and total inner reflection is obtained when the source of light is settle to refract to another medium with smaller refractive index.

3. Wave propagation in (2 + 1)-dimensional anti-deSitter spacetime

If we consider now a (2+1)-dimensional anti-deSitter spacetime with spatial optical metric given by (Ref. 7)

$$g = \left(\frac{n}{y}\right)^2 \sum_{i=1}^{2} \mathrm{d}x^i \otimes \mathrm{d}x^i. \tag{2}$$

we find the associated 1-parameter family of strict contact transformations, generated by the Reeb's flow, for which we obtained

$$\phi_t = \Bigg[x = \frac{(x\sin\varphi + y\cos\varphi - x)e^{\frac{2t}{n}} - x\sin\varphi - y\cos\varphi - x}{(\sin\varphi - 1)e^{\frac{2t}{n}} - \sin\varphi - 1}, \tag{3}$$

$$y = -\frac{2ye^{\frac{t}{n}}}{(\sin\varphi - 1)e^{\frac{2t}{n}} - \sin\varphi - 1}\ ,$$

$$\varphi = \arctan\left(-\frac{(-\sin\varphi + 1)e^{\frac{2t}{n}} - \sin\varphi - 1}{2e^{\frac{t}{n}}\cos\varphi}\right).\Bigg]$$

With this transformation, it is possible to observe that the projection of the flows corresponds to semi-circles centered in the x-axis, as expected by the hyperbolic geometry associated to the anti-deSitter spacetime. For interfaces between media with different refractive index, refraction can be observed and no total inner reflection is observed when projected to the Poincaré disc. As it is known, the AdS spacetime is nos globally hyperbolic, so it admits closed timelike curves and achronal surfaces can be observed where light rays are closer to the infinity circle boundary of the disc. It is also observed that light rays can intersect more than one time the same wave front (figure 1 Ref. 20).

4. Final remarks

The fact that geodesic flows can be written in terms of a 1-parameter family of contact transformations, allows us to reconstruct trajectories and wavefronts of light while propagating through an optical medium, just from the Reeb flow. With this technique, we recreate the wave fronts as they propagate in an optical medium with interfaces of different refractive index in a (2 + 1)-dimensional Minkowski spacetime and (2 + 1)-dimensional anti-deSitter spacetime. In the first example, we could reproduce Snell's law of refraction and the total inner reflexion phenomenon. In the second example, we observe refraction when mapped to the Poincaré disc. No inner total reflection was observed when the light source is in a medium with larger refractive index. Achronal surfaces can be observed in this configuration, as light rays closer to the circle boundary intersects more than once the same wavefront.

The formalism, is robust enough to deal with interfaces without further adaptations. This technique can be used to recreate wave propagation in analogue gravitational spacetimes, such as Schwarzschild's blackholes, or to study interface between

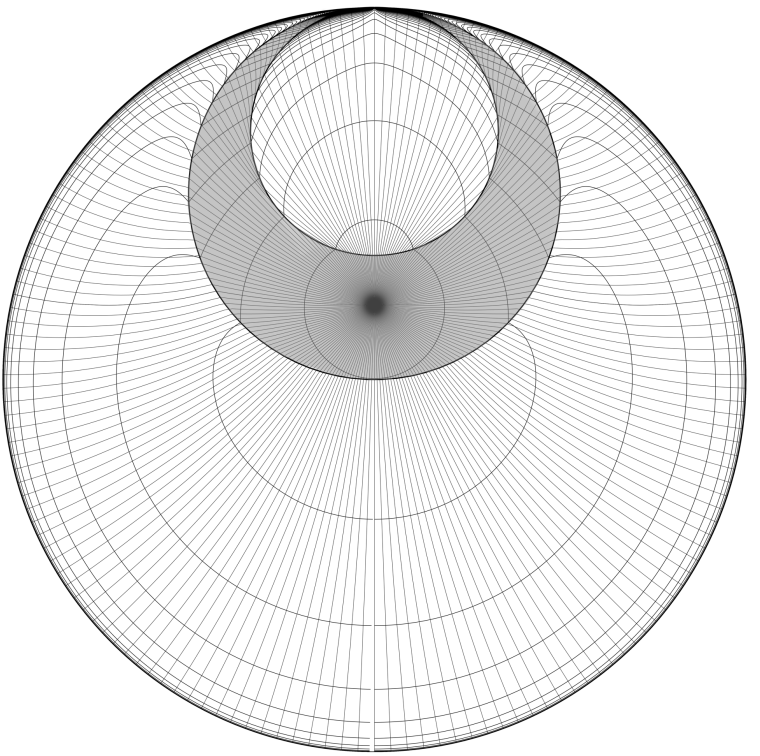


Fig. 1. Light rays and wavefronts emitted by a source of light refracting through a horizontal interface mapped to the Poincaré disc. The refractive index in the darker layer (where the light source is settled) is larger than the one in the whiter space.

vacuum and plasma. In recent years, this type of techniques can be important in the development of metamaterials, which can mimic some spacetime properties impossible to find in nature.

Acknowledgments

DGP was funded by a CONACYT Scholarship with CVU 425313.

References

1. W. Gordon, Zur lichtfortpflanzung nach der relativitätstheorie, *Annalen der Physik* **377**, 421 (1923).
2. F. de Felice, On the gravitational field acting as an optical medium, *General Relativity and Gravitation* **2**, 347 (1971).
3. J. Ehlers, F. A. Pirani and A. Schild, Republication of: The geometry of free fall and light propagation, *General Relativity and Gravitation* **44**, 1587 (2012).
4. J. B. Pendry, D. Schurig and D. R. Smith, Controlling electromagnetic fields, *science* **312**, 1780 (2006).

5. U. Leonhardt and T. G. Philbin, General relativity in electrical engineering, *New Journal of Physics* **8**, p. 247 (2006).
6. H. Chen, C. T. Chan and P. Sheng, Transformation optics and metamaterials, *Nature materials* **9**, 387 (2010).
7. C. Lopez-Monsalvo, D. Garcia-Pelaez, A. Rubio-Ponce and R. Escarela-Perez, The geometry of induced electromagnetic fields in moving media, *Annals of Physics* **420**, p. 168270 (2020).
8. H. Chen, S. Tao, J. Bělín, J. Courtial and R.-X. Miao, Transformation cosmology, *Phys. Rev. A* **102**, p. 023528 (Aug 2020).
9. S. Schuster and M. Visser, Electromagnetic analogue space-times, analytically and algebraically, *Classical and Quantum Gravity* (2019).
10. D. Faccio, F. Belgiorno, S. Cacciatori, V. Gorini, S. Liberati and U. Moschella, *Analogue gravity phenomenology: analogue spacetimes and horizons, from theory to experiment* (Springer, 2013).
11. S. Schuster and M. Visser, Bespoke analogue space-times: Meta-material mimics, *General Relativity and Gravitation* **50**, 1 (2018).
12. V. Perlick, *Ray optics, Fermat's principle, and applications to general relativity* (Springer Science & Business Media, 2000).
13. Y. Y. Berest and A. P. Veselov, Huygens' principle and integrability, *Russian Mathematical Surveys* **49**, 5 (1994).
14. A. I. Harte, Tails of plane wave spacetimes: Wave-wave scattering in general relativity, *Physical Review D* **88**, p. 084059 (2013).
15. R. McLenaghan, An explicit determination of the empty space-times on which the wave equation satisfies huygens' principle, in *Mathematical Proceedings of the Cambridge Philosophical Society*, (1)1969.
16. T. W. Noonan, Huygens' principle in conformally flat spacetimes, *Classical and Quantum Gravity* **12**, p. 1087 (1995).
17. G. W. Gibbons, Anti-de-sitter spacetime and its uses, in *Mathematical and quantum aspects of relativity and cosmology*, (Springer, 2000) pp. 102–142.
18. U. Moschella, The de sitter and anti-de sitter sightseeing tour, in *Einstein, 1905–2005*, (Springer, 2006) pp. 120–133.
19. H. Geiges, *An Introduction to Contact Topology* (Cambridge University Press, 2008).
20. D. García-Peláez, C. López-Monsalvo and A. R. Ponce, The duality of fermat and huygens principles through contact transformations, *arXiv preprint arXiv:2108.06044* (2021).

New approaches to constrained dynamics and Hamilton-Jacobi procedures in general relativity

D. Salisbury

Physics Department, Austin College,
Sherman, Texas 75090, USA
E-mail: dsalisbury@austincollege.edu
www.austincollege.edu

J. Renn and K. Sundermeyer

Max Planck Institute for the History of Science, Boltzmannstrasse 22,
14195 Berlin, Germany

Élie Cartan's invariant integral formalism is extended to gauge field theory, including general relativity. This constitutes an alternative procedure that is equivalent to the Rosenfeld, Bergmann, Dirac algorithm. Also a new derivation of the generator of diffeomorphism-induced canonical transformations is given that proceed's from the Poincaré-Cartan form. In addition, a Hamilton-Jacobi formalism is developed for constructing explicit phase space functions in general relativity that are invariant under the full four-dimensional diffeomorphism group. These identify equivalence classes of classical solutions of Einstein's equations. Each member is dependent on intrinsic spatial coordinates and also undergoes non-trivial evolution in intrinsic time. The intrinsic coordinates are determined by the spacetime geometry in terms of Weyl scalars. The implications of this analysis for an eventual quantum theory of gravity are profound.

Keywords: General relativity; Hamilton-Jacobi formalism.

1. Introduction

We have found some new profound implications of Élie Cartan's theory, as presented in his 1922 work, Ref. 1. We show how his invariant integral program can be extended to gauge field theories, including general relativity. Our procedure in fact constitutes a new approach to constrained Hamiltonian dynamics that is equivalent to the Rosenfeld-Bergmann-Dirac procedure. Secondly, we show that the Poincaré-Cartan form that arises in this context can be utilized to deduce the generator of diffeomorphism-induced canonical transformations that have previously been derived using group theoretical arguments. We show how the required metric field dependence of the diffeomorphism-induced transformation group comes into play. Finally, employing the related symplectic form, we undertake canonical transformations to intrinsic spacetime coordinates, thus isolating the true gravitational degrees of freedom in a formalism in which all objects are invariant under the action of the full four-dimensional diffeomorphism group. The following is a summary of some results. Full details will appear in Ref. 2.

2. Cartan invariant integral — extension to gauge field theory

In 1922 Élie Cartan, Ref. 1, introduced an alternative approach to the classical dynamics of a configuration variable $q^i(t)$ in which he demanded that a certain closed integral evaluated at a fixed time t be independent of the chosen time.[a] He proposed that one consider a tube of trajectories parameterized by s with the property that as s ranges varies over its full range one returns to the original trajectory. Thus he demanded that the integral I over s, employing the parameterized set $q^i(t(s);s)$, $v^j(t(s);s)$, $t(s)$, be independent of t where

$$\begin{aligned} I &= \oint ds \left(\frac{\partial L}{\partial v^i} \left. \frac{\partial q^i}{\partial s} \right|_{t(s)} - \left(\frac{\partial L}{\partial v^i} v^i - L \right) \frac{dt}{ds} \right) \\ &=: \oint \left(\frac{\partial L}{\partial v^i} dq^i - \left(\frac{\partial L}{\partial v^i} v^i - L \right) dt \right) \end{aligned} \tag{1}$$

We consider the free relativistic particle as a first example of an extension of this procedure to a gauge model. In this case the required invariant integral is

$$I_p = \oint \left(N^{-1} v_\mu dq^\mu - \frac{1}{2} \left(N^{-1} v^2 + N m^2 \right) dt \right). \tag{2}$$

We require that its variation vanish under independent δ variations. Integrating by parts over the closed interval we get

$$\begin{aligned} \delta I_p = \oint \Bigg[&\left(-N^{-2} \delta N v_\mu + N^{-1} \delta v_\mu \right) dq^\mu + \left(\frac{1}{2} N^{-2} \delta N v^2 - N^{-1} v_\mu \delta v^\mu \right. \\ &\left. \frac{1}{2} \delta N m^2 \right) dt - \left(\frac{1}{2} N^{-2} dN v^2 - N^{-1} v_\mu dv^\mu - \frac{1}{2} dN m^2 \right) \delta t \Bigg] = 0. \end{aligned} \tag{3}$$

The required vanishing coefficient of δv^μ yields $v^\mu = \frac{dq^\mu}{dt} := \dot{q}^\mu$. Similarly, the vanishing coefficient for δN, gives $N = m^{-1} \left(-\dot{q}^2 \right)^{1/2}$. The result for δt is redundant, yielding $-m^2 \dot{N} = N^{-1} \dot{q} \cdot \ddot{q}$. Finally, the coefficient of δq^μ yields the reparameterization covariant equation of motion $\frac{d}{dt} \left(N^{-1} \dot{q}^\mu \right) = 0$.

It is especially instructive to undertake a Legendre transformation, but with the understanding that the analogues of the velocities that we represent as v^μ and N_v represent independent functions. Thus $\pi = \frac{\partial L_p}{\partial N_v} = 0$ is a primary constraint. The required invariant integral becomes

$$I_p = \oint \left[p_\mu dq^\mu + \pi dN - \left(\frac{N}{2} \left(p^2 + m^2 \right) + \pi N_v \right) dt \right], \tag{4}$$

[a] A translation into English by D. H. Delphenich, Ref. 3, is available.

resulting in

$$dI_p = \oint \Big[\delta p_\mu \left(dq^\mu - Np^\mu dt\right) - \delta q^\mu dp_\mu + \delta N \left(-d\pi - \frac{1}{2}(p^2+m^2)dt\right) + \delta\pi\left(dN - \dot{N}dt\right) + \delta t\left(\left(\frac{1}{2}(p^2+m^2)\right)dN + Np^\mu\delta p_\mu + \delta\pi\dot{N} + \pi\delta\dot{N}\right) + \delta\dot{N}\pi dt\Big] = 0. \tag{5}$$

The required vanishing of the first and second coefficients result in $\frac{dq^\mu}{dt} = Np^\mu$ and $\frac{dp^\mu}{dt} = 0$. The third term yields the secondary constraint $p^2 + m^2 = 0$, and the forth $N_v = \frac{dN}{dt}$. The remaining terms are redundant.

One can in fact carry out a similar variation with an analogous general relativistic invariant integral, namely

$$I_{ADM} = \oint d^3x \left[p^{ab}dg_{ab} + P_\mu dN^\mu - \mathcal{H}_{ADM}dt - P_\mu N_v^\mu dt\right], \tag{6}$$

where the Arnowittt-Deser-Misner (ADM) Hamiltonian

$$\mathcal{H}_{ADM} = \frac{N}{\sqrt{{}^3g}}\left(p_{ab}p^{ab} - \left(p^a{}_a\right)^2 - N\sqrt{{}^3g}{}^3R - 2N^a p^b{}_{a|b}\right), \tag{7}$$

is derived from the ADM Lagrangian

$$\mathcal{L}_{ADM} = \left({}^3g\right)^{1/2} N\left({}^3R + K_{ab}K^{ab} - K^2\right), \tag{8}$$

with $K_{ab} = \frac{1}{2N}\left(v_{ab} - N_{a|b} - N_{b|a}\right)$. Note that the v_{ab} are assumed to be independent spacetime functions. This Lagrangian is independent of the analogue velocity N_v^μ and there is therefore a primary constraint $P_\mu = 0$. The independent variations of g_{ab}, K_{ab}, N^μ, and N_v^μ. yields the correct secondary constraints and the correct Einstein Hamiltonian equations. Details are given in Ref. 2. These two systems are examples of a general new invariant integral approach to constrained Hamiltonian dynamics, equivalent to the Rosenfeld-Bergmann-Dirac prodedure.

3. Derivation of diffeomorphism generator

In the free relattivistic particle model take ds in the Poincaré-Cartan increment to correspond to the change in solutions that results from an infinitesimal reparameterization $t' = t - ds\epsilon(t)$. Represent the change in the variables as $\bar{d}q^\mu = ds\dot{q}^\mu\epsilon$ and $\bar{d}N = ds(\dot{\epsilon}N + \epsilon\dot{N})$, recognizing that N transforms as a scalar density under the reparamterization. (These are actually the Lie derivatives with respect to ϵ.)

Now use the identity which is the statement that the particle Lagrangian transforms as a scalar density under reparameterizations to deduce the existence of a vanishing Noether charge C_p:

$$\frac{dC_p}{dt} = \frac{d}{dt}\left[\frac{N\epsilon}{2}\left(p^2+m^2\right) + \pi\left(\dot{N}\epsilon + N\dot{\epsilon}\right)\right] = 0. \tag{9}$$

But since this depends on $\dot{N}$ this is clearly not projectable under the Legendre transformation to the particle co-tangent space. We must assume a variable-dependent reparameterizations $\epsilon(t) = N^{-1}\xi(t)$ to get Legendre-projectable reparameterizations. The general proof for the required dependence is given in.[4] In this case we have

$$\begin{aligned} C_p &= \frac{\xi}{2}\left(p^2 + m^2\right) + \pi\left(\dot{N}\epsilon + N\dot{\epsilon}\right) \\ &= \frac{\xi}{2}\left(p^2 + m^2\right) + \pi\left(\dot{N}N^{-1}\xi - NN^{-2}\dot{N}\xi + \dot{\xi}\right) \\ &= \frac{\xi}{2}\left(p^2 + m^2\right) + \pi\dot{\xi}. \end{aligned} \tag{10}$$

This is the phase space generator of the N-dependent reparameterizations.

It is shown in Ref. 2 that the analogous substitution into the ADM vanishing Noether charge yields the Pons-Salisbury-Shepley generator $\mathcal{C}_{PSS}$, Ref. 4, of diffeomorphism-induced general coordinate transformations $\epsilon^\mu(x) = \left(N^{-1}\xi^0(x), -N^{-1}N^a\xi^0(x) + \xi^a(x)\right)$,

$$\begin{aligned} \mathcal{C}_{PSS} &= P_\mu\dot{\xi}^\mu + \mathcal{H}_\mu\xi^\mu + P_0(\xi^0 - N^a\xi^0_{,a} + N_{,a}\xi^a) \\ &\quad + P_a(N_{,b}e^{ab}\xi^0_{,b} - N\xi^0_{,b}e^{ab} + N_{,a}\xi^a + N^a_{,b}\xi^b - N^b\xi^a_{,b}). \end{aligned} \tag{11}$$

4. Intrinsic coordinates and diffeomophism invariants

The non-vanishing contributions to the integrand in Cartan's invariant integral are of course now recognized as the symplectic potential one-form. Perhaps less fully appreciated is the fact that Cartan justified canonical transformations to new phase space coordinates through the recognition that exact differential generator did not contribute to the invariant integral. We examine first what we believe is novel new approach to the introduction of a proper time parameterization of the relativistic free particle. This is an example of a choice of intrinsic time. As proved in general in Ref. 5 the chosen intrinsic canonical coordinate must transform as a scalar under general coordinate transformations - in this case, reparamterizations.

Thus we now choose as the intrinsic particle time evolution variable the scalar $T = -mq^0/p_0$. This is the proper time. Then we perform a canonical transformation in the Poincaré-Cartan increment such that

$$p_0 dq^0 = PdT + \frac{\partial G}{\partial q^0}dq^0 + \frac{\partial G}{\partial T}dT. \tag{12}$$

The required non-trivial generator is

$$G = -\frac{m}{2T}\left(q^0\right)^2 - mT + mT\ln(T), \tag{13}$$

with the resulting canonical conjugate

$$P = -m\ln\left(-\frac{mq^0}{p_0}\right) - \frac{p_0^2}{2m}. \tag{14}$$

It is important to recognize that the system is still reparameterization covariant – and now with a new generator

$$C_p = \xi \left[-P - m \ln(T) + \frac{1}{2} \left(p^a p_a + m^2 \right) \right] + \pi \dot{\xi} \tag{15}$$

But we can now make a proper time gauge choice by setting the evolution parameter equal to the intrinsic variable T, i.e., set $t = T$. And given particle solutions in any parameterization the gauge generator can be employed to transform to solutions satisfying the gauge condition. The remaining canonical variables are invariants under reparameterizations. The general procedure is described in Ref. 5.

We can indeed, in principle, carry out a similar procedure in general relativity. Choose appropriate spacetime scalars $X^\mu \left(g_{ab}, p^{cd} \right)$, constructed with the use of Weyl scalars, as intrinsic spacetime coordinates. An initial effort in this direction is described in Refs. 6 and 7. Fuller details with corrections and revisions will appear in Ref. 2. Perform canonical transformations in the non-vanishing portion of the Poincaré-Cartan increment so that the non-vanishing contribution becomes

$$\begin{aligned} dI_{ADM} &= \int d^3x \, p^{ab} dg_{ab} \\ &= \int d^3x \left(\pi_\mu dX^\mu + p^A dg_A + \frac{\delta G}{\delta g_{ab}} dg_{ab} + \frac{\delta G}{\delta \alpha_A} d\alpha_A + \frac{\delta G}{\delta X^\mu} dX^\mu \right). \end{aligned} \tag{16}$$

Now make the gauge choices $x^\mu = X^\mu \left(g_{ab}, p^{cd} \right)$. The diffeomorphism generator C_{PSS} can be employed to explicitly display the corresponding diffeomorphism invariants, as shown in Ref. 5.

5. Quantum outlook

The results that we describe here were largely inspired by the work of Peter G. Bergmann and his associates. An extensive analysis of the work of his school and the conflicts that arose with the geometrical dynamical approach advocated by the John A. Wheeler can be found in Ref. 8. Of particular relevance is work performed by Bergmann in collaboration with Arthur Komar, analyzed in detail in Ref. 9. A revision and further elaboration is in preparation. The fundamental disagreement with the conventional geometrodynamical approach is in the treatment of the full four-dimensional diffeomorphism invariance. The classical approach we describe here works exclusively with the $2 \times \infty^3$ invariant fundamental gravitational degrees of freedom. It remains to be seen how these invariants may be employed in an eventual quantum theory of gravity. Work is now underway. A central challenge is to find a scheme that takes into account the abundance of physically distinct gauge choices, each of which leads to a distinct classical intrinsic time evolution.

References

1. E. Cartan, *Leçons sur les Invariants Intégraux* (Librairie Scientifique A. Hermann et Fils, 1922).

2. D. Salisbury, J. Renn and K. Sundermeyer, Rediscovering cartan, *in preparation* (2021).
3. E. Cartan, *Lessons on Integral Invariants* (Hermann, 1922).
4. J. Pons, D. Salisbury and L. Shepley, Gauge transformations in the Lagrangian and Hamiltonian formalisms of generally covariant theories, *Physical Review D* **55**, 658 (1997).
5. J. Pons, D. Salisbury and K. Sundermeyer, Revisiting observables in generally covariant theories in light of gauge fixing methods, *Physical Review D* **80**, 084015 (2009).
6. D. Salisbury, J. Renn and K. Sundermeyer, Restoration of four-dimensional diffeomorphism covariance in canonical general relativity: An intrinsic Hamilton-Jacobi approach, *International Journal of Modern Physics A* **31**, 1650014 (2016).
7. D. Salisbury, J. Renn and K. Sundermeyer, Erratum. Restoration of four-dimensional diffeomorphism covariance in canonical general relativity: An intrinsic Hamilton Jacobi approach, *International Journal of Modern Physics A* **33**, p. 1892001 (2018).
8. D. Salisbury, Toward a quantum theory of gravity: Syracuse 1949-1962, in *The Renaissance of General Relativity in Context*, eds. A. Blum, R. Lalli and J. Renn (Birkhäuser, 2020) pp. 221–255.
9. D. Salisbury, Observables and Hamilton-Jacobi approaches to general relativity. I. Early history, *arXiv:2106.10774[gr-qc]* (2021).

Orientability of space from electromagnetic quantum fluctuations

N. A. Lemos

Instituto de Física, Universidade Federal Fluminense, Av. Litorânea, S/N, 24210-340 Niterói – RJ, Brazil
E-mail: nivaldolemos@id.uff.br

M. J. Rebouças

Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro – RJ, Brazil
E-mail: reboucas.marcelo@gmail.com

Whether the 3−space where we live is a globally orientable manifold M_3, and whether the local laws of physics require that M_3 be equipped with a canonical orientation, are among the important unsettled questions in cosmology and quantum field theory. It is often assumed that a test for spatial orientability requires a global journey across the whole 3−space to check for orientation-reversing closed paths. Since such a global expedition is not feasible, physically motivated theoretical arguments are usually offered to support the choice of canonical time orientation for the 4−dimensional spacetime manifold, and space orientation for 3−space. One can certainly take advantage of such theoretical arguments to support these assumptions on orientability, but the ultimate answer should rely on cosmological observations or local experiments, or can come from a topological fundamental theory of physics. In a recent paper we have argued that it is potentially possible to locally access the the 3−space orientability of Minkowski empty spacetime through physical effects involving point-like 'charged' objects under vacuum quantum electromagnetic fluctuations. More specifically, we have studied the stochastic motions of a charged particle and an electric dipole subjected to these fluctuations in Minkowski spacetime, with either an orientable or a non-orientable 3−space topology, and derived analytical expressions for a statistical orientability indicator in these two flat topologically inequivalent manifolds. For the charged particle, we have shown that it is possible to distinguish the two topologies by contrasting the evolution of their respective indicators. For the point electric dipole we have found that a characteristic inversion pattern exhibited by the curves of the orientability indicator is a signature of non-orientability, making it possible to locally probe the orientability of Minkowski 3−space in itself. Here to shed some additional light on the spatial orientability, we briefly review these results, and also discuss some of its features and consequences. The reviewed results might be seen as opening the way to a conceivable experiment involving quantum vacuum electromagnetic fluctuations to look into the spatial orientability of Minkowski empty spacetime.

Keywords: Spatial topology of Minkowski spacetime; Orientability of Minkowsk space; Quantum fluctuations of electromagnetic field; Motion of changed particle and electric dipole under electromagnetic fluctuations

1. Introduction

In the framework of general relativity the Universe is described as a four-dimensional differentiable manifold locally endowed with a spatially homogeneous and isotropic

Friedmann–Lemaître–Robertson–Walker (FLRW) metric. Geometry is a local attribute that brings about intrinsic curvature, whereas topology is a global feature of a manifold related, for example, to compactness and orientability. The FLRW spatial geometry constrains but does not specify the topology of the spatial sections, M_3, of the space-time manifold, which we assume to be of the form $\mathscr{M}_4 = \mathbb{R} \times M_3$. In the FLRW description of the Universe, two diverse sets of fundamental questions are related to, first, the 3−geometry; second, to the topology of spatial sections M_3. Regarding the latter, at the cosmological level it is expected that one should be able to detect the spatial topology through cosmic microwave background radiation (CMB) or (and) stochastic primordial gravitational waves.[1,2] However, so far, direct searches for a nontrivial topology of M_3, using CMB data from Wilkinson Microwave Anisotropy Probe (WMAP) and Planck collaborations, have found no convincing evidence of nontrivial topology below the radius of the last scattering surface.[3–9] This absence of evidences does not exclude the possibility of a FLRW universe with a detectable nontrivial spatial topology.[10–12]

It is well-known that the topological properties of a manifold antecede its geometrical features and the differential tensor structure with which the physical theories are formulated. Thus, it is relevant to determine whether, how and to what extent physical results depend upon or are somehow affected by a nontrivial topology. Since the net role played by the spatial topology is more properly examined in static space-times, the dynamical degrees of freedom of which are frozen, here we focus on the static Minkowski space-time, whose spatial geometry is Euclidean. However, rather than the general topology of the spatial sections M_3 of Minkowski space-time, in this work we investigate its topological property called *orientability*, which is related to the existence of orientation-reversing closed paths on the spatial section M_3. Questions as to whether the 3−space of Minkowski space-time, which is the standard arena of quantum field theory, is necessarily an orientable manifold, or to what extent the known *laws of physics require a canonical spatial orientation* are among the underlying primary concerns of this work.

To be more precise regarding the setting of the present work, let us first briefly provide some mathematical results. The spatial section M_3 of the Minkowski space-time manifold $\mathscr{M}_4 = \mathbb{R} \times M_3$ is usually taken to be the simply-connected Euclidean space $\mathbb{E}^3$, but it is known mathematical result that it can also be any one of the possible 17 topologically distinct quotient manifolds $M_3 = \mathbb{E}^3/\Gamma$, where Γ is a discrete group of isometries or holonomies acting freely on the covering space $\mathbb{E}^3$.[a] The action of Γ tiles the covering manifold $\mathbb{E}^3$ into identical cells which are copies of what is known as fundamental domain (FD) fundamental cell or polyhedron (FC or FP). So, the multiple-connectedness gives rise to periodic conditions (repeated cells) in the simply-connected covering manifold $\mathbb{E}^3$ that are defined by the action

[a]For recent accounts on the classification of three-dimensional Euclidean spaces the reader is referred to Refs. 13 –18.

of the group Γ on $\mathbb{E}^3$. Different groups Γ give rise to different periodicities on the covering manifold $\mathbb{E}^3$, which in turn define different Euclidean spatial topologies M_3 for Minkowski spacetime. These mathematical results make it explicit that besides the simply-connected $\mathbb{E}^3$ there is a variety of topological possibilities (17 classes) for the spatial section of Minkowski spacetime. The potential consequences of multiple-connectedness for physics come about when, for example, one takes into account that in a manifold with periodic boundary conditions only certain modes of fields can exist. In this way, a nontrivial topology may leave its signature on the expectation values of local physical quantities.[19] So, for example, the energy density for a scalar field in Minkowski space-time with nontrivial spatial section is shifted from the corresponding result for the Minkowski space-time with trivial spatial topology. This is the Casimir effect of topological origin.[20–25]

Regarding orientability, the central issue of this work, three important points should be emphasized. First, we mention that it is widely assumed, implicitly or explicitly, that a four-dimensional manifold $\mathscr{M}_4 = \mathbb{R} \times M_3$ that models the physical world is spacetime orientable and, additionally, that it is separately time and space orientable. Second, eight out of the above-mentioned 17 quotient flat 3−manifolds are non-orientable.[13–18] A non-orientable 3−space is then a concrete mathematical possibility among the quotient manifolds $M_3 = \mathbb{E}^3/\Gamma$, and it comes about when the holonomy group Γ contains at least a flip or reflexion as one of its elements. Finally, it is generally assumed that, being a global property, the 3−space orientability cannot be tested locally. In this way, to disclose the spatial orientability of our physical world one would have to make trips along some specific closed paths around 3−space to check, for example, whether one returns with left- and right-hand sides exchanged.

Since such a global journey across the whole 3−space is not feasible one might think that spatial orientability cannot be probed. In this way, one would have either to answer the orientability question through cosmological observations or local experiments. Hence, assuming that spatial orientability is a falsifiable property of 3−space, a question that arises is whether spatial orientability can be subjected to local tests. Our main goal in this work is to present a way to tackle this question. To this end, we have investigated[26] stochastic motions of a charge particle and an electric dipole under vacuum quantum fluctuations of the electromagnetic field in Minkowski spacetime with two inequivalent spatial topologies, namely the non-orientable *slab with flip* and the orientable *slab*, which are often denoted by the symbols E_{17} and E_{16}, respectively.[17,18] Manifolds endowed with these topologies turned out to be appropriate to identify orientability or non-orientability signatures through the stochastic motions of point-like particles in Minkowski spacetime. In the next section, to shed some additional light on spatial orientability, we briefly review the most significant results of our article, Ref. 26, and discuss some of their consequences.

2. Main results and their interpretation

The general idea underlying our work is to perform a comparative study of the time evolution of physical systems in Minkowski spacetime with orientable and non-orientable spatial sections. To this end, the physical systems chosen are a point charged particle and a point electric dipole under vacuum quantum electromagnetic fluctuations. We shall describe the results for the dipole only because they are more significant — see Ref. 26 for the point charge case.

2.1. *The setting and the physical system*

We begin by recalling that simply-connected spacetime manifolds are necessarily orientable. On the other hand, the product of two manifolds is simply-connected if and only if the factors are. Thus, the space-orientability of Minkowski spacetime manifold $\mathscr{M}_4 = \mathbb{R} \times M_3$ reduces to orientability of the 3−space M_3. In this paper, we shall consider the topologically nontrivial spaces E_{16} and E_{17}. The slab space E_{16} is constructed by tessellating $\mathbb{E}^3$ by equidistant parallel planes, so it has only one compact dimension associated with a direction perpendicular to those planes. Taking the x-direction as compact, one has that, with $n_x \in \mathbb{Z}$ and $a > 0$, points (x, y, z) and $(x + n_x a, y, z)$ are identified in the case of the slab space E_{16}. The slab space with flip E_{17} involves an additional inversion of a direction orthogonal to the compact direction, that is, one direction in the tessellating planes is flipped as one moves from one plane to the next. Letting the flip be in the y-direction, the identification of points (x, y, z) and $(x + n_x a, (-1)^{n_x} y, z)$ defines the E_{17} topology. The slab space E_{16} is orientable whereas the slab space with flip E_{17} is non-orientable.

We consider a nonrelativistic point electric dipole with mass m and electric dipole moment $\mathbf{p}$ which is locally subject to vacuum fluctuations of the electric field $\mathbf{E}(\mathbf{x}, t)$ in a topologically nontrivial spacetime manifold equipped with the Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$. The spatial section is usually taken to be $\mathbb{E}^3$, but here we take for M_3 either E_{16} or E_{17}.

Locally, the nonrelativistic motion of the dipole is determined by the equation of motion

$$m\frac{d\mathbf{v}}{dt} = \mathbf{p} \cdot \boldsymbol{\nabla}\mathbf{E}(\mathbf{x}, t)\,, \tag{1}$$

where $\mathbf{v}$ is the dipole's velocity and $\mathbf{x}$ its position at time t. We assume that on the time scales of interest the dipole practically does not move, that is, it has a negligible displacement, so we can ignore the time dependence of $\mathbf{x}$. Thus, the dipole's position $\mathbf{x}$ is taken as constant in what follows.[27, 28] Assuming the dipole is initially at rest, integration of Eq. (1) yields

$$\mathbf{v}(\mathbf{x}, t) = \frac{1}{m} p_j \int_0^t \partial_j \mathbf{E}(\mathbf{x}, t')\, dt' \tag{2}$$

with $\partial_j = \partial/\partial x_j$ and summation over repeated indices implied.

The mean squared speed in each of the three independent directions $i = x, y, z$ is given by

$$\left\langle \Delta v_i^2 \right\rangle = \frac{p_j p_k}{m^2} \int_0^t \int_0^t \left\langle \big(\partial_j E_i(\mathbf{x}, t')\big)\big(\partial_k E_i(\mathbf{x}, t'')\big) \right\rangle dt' dt'' \,, \tag{3}$$

which can be conveniently rewritten as

$$\left\langle \Delta v_i^2 \right\rangle = \lim_{\mathbf{x}' \to \mathbf{x}} \frac{p_j p_k}{m^2} \int_0^t \int_0^t \partial_j \partial'_k \left\langle E_i(\mathbf{x}, t') E_i(\mathbf{x}', t'') \right\rangle dt' dt'' \tag{4}$$

where $\partial'_k = \partial/\partial x'_k$ and there is no summation over i. Following Yu and Ford,[27] we assume that the electric field is a sum of classical $\mathbf{E}_c$ and quantum $\mathbf{E}_q$ parts. Because $\mathbf{E}_c$ is not subject to quantum fluctuations and $\langle \mathbf{E}_q \rangle = 0$, the two-point function $\langle E_i(\mathbf{x}, t) E_i(\mathbf{x}', t') \rangle$ in equation (4) involves only the quantum part of the electric field.[27]

In Minkowski spacetime with a topologically nontrivial spatial section, the spatial separation r^2 that enters the electric field correlation functions takes a different form that captures the periodic boundary conditions imposed on the covering space $\mathbb{E}^3$ by the covering group Γ, which characterize the spatial topology. In consonance with Ref. 22, the spatial separations for E_{16} and E_{17} are

$$E_{16}: \qquad r^2 = (x - x' - n_x a)^2 + (y - y')^2 + (z - z')^2, \tag{5}$$

$$E_{17}: \qquad r^2 = (x - x' - n_x a)^2 + (y - (-1)^{n_x} y')^2 + (z - z')^2. \tag{6}$$

2.2. *Orientability indicator*

The orientability indicator that we will consider is defined by replacing the electric field correlation functions in Eq. (4) by their renormalized counterparts.[26] For a dipole oriented along the y-axis the dipole moment is $\mathbf{p} = (0, p, 0)$ and we have

$${}^{(y)}I^{E_{17}}_{v_i^2}(\mathbf{x}, t) = \lim_{\mathbf{x}' \to \mathbf{x}} \frac{p^2}{m^2} \int_0^t \int_0^t \partial_y \partial_{y'} \left\langle E_i(\mathbf{x}, t') E_i(\mathbf{x}', t'') \right\rangle^{E_{17}}_{ren} dt' dt'' \,, \tag{7}$$

where the left superscript within parentheses indicates the dipole's orientation.

The renormalized correlation functions are given by[26]

$$\left\langle E_i(\mathbf{x}, t) E_i(\mathbf{x}', t') \right\rangle_{ren} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x'_i} D_{ren}(\mathbf{x}, t; \mathbf{x}', t') - \frac{\partial}{\partial t} \frac{\partial}{\partial t'} D_{ren}(\mathbf{x}, t; \mathbf{x}', t') \tag{8}$$

where

$$D_{ren}(\mathbf{x}, t; \mathbf{x}', t') = \sum_{n_x = -\infty}^{\infty}{}' \frac{1}{4\pi^2(\Delta t^2 - r^2)} \tag{9}$$

in which $\sum'$ indicates that the Minkowski contribution term $n_x = 0$ is excluded from the summation, $\Delta t = t - t'$, and the spatial separation for E_{17} is given by Eq. (6). The Hadamard function $D(\mathbf{x}, t; \mathbf{x}', t')$ for the multiply-connected space is

defined by including the term with $n_x = 0$ in the summation (9). Thus, $D_{ren} = D - D_0$, where D_0 is the Hadamard function for the simply-connected space.

Before presenting the results for some components of the orientability indicator (7), we point out that from equations (4) and (7)–(9) a general definition of the orientability indicator reads

$$I^{MC}_{v_i^2} = \left\langle \Delta v_i^2 \right\rangle^{MC} - \left\langle \Delta v_i^2 \right\rangle^{SC}, \tag{10}$$

where $\left\langle \Delta v_i^2 \right\rangle$ is the mean square velocity dispersion, and the superscripts MC and SC stand for multiply- and simply-connected manifolds, respectively. The right-hand side of (10) is defined by first taking the difference of the two terms with $\mathbf{x}' \neq \mathbf{x}$ and then setting $\mathbf{x}' = \mathbf{x}$.

The components of the orientability indicator for the dipole in E_{17} are[26]

$${}^{(y)}I^{E_{17}}_{v_x^2}(\mathbf{x},t) = -\frac{4p^2}{\pi^2 m^2} {\sum_{n_x=-\infty}^{\infty}}' (-1)^{n_x} \left\{ 2I_1 + 3(r^2 - r_x^2 + 6r_y^2)I_2 + 24(r^2 - r_x^2) r_y^2 I_3 \right\}, \tag{11}$$

$${}^{(y)}I^{E_{17}}_{v_y^2}(\mathbf{x},t) = -\frac{2p^2}{\pi^2 m^2} {\sum_{n_x=-\infty}^{\infty}}' (-1)^{n_x} \Big\{ (5 - 3(-1)^{n_x})I_1 + 6[r^2 + (7 - 6(-1)^{n_x}) r_y^2] I_2 + 48[r^2 - (-1)^{n_x} r_y^2] r_y^2 I_3 \Big\} \tag{12}$$

and

$${}^{(y)}I^{E_{17}}_{v_z^2}(\mathbf{x},t) = -\frac{4p^2}{\pi^2 m^2} {\sum_{n_x=-\infty}^{\infty}}' (-1)^{n_x} \left\{ 2I_1 + 3(r^2 + 6r_y^2)I_2 + 24 r^2 r_y^2 I_3 \right\}, \tag{13}$$

where, with $\Delta t = t' - t''$,

$$I_1 = \int_0^t \int_0^t \frac{dt' dt''}{(\Delta t^2 - r^2)^3} = \frac{t}{16} \left[\frac{4t}{r^4 (t^2 - r^2)} + \frac{3}{r^5} \ln \frac{(r-t)^2}{(r+t)^2} \right], \tag{14}$$

$$I_2 = \int_0^t \int_0^t \frac{dt' dt''}{(\Delta t^2 - r^2)^4} = \frac{1}{6r} \frac{\partial I_1}{\partial r} = \frac{t}{96} \left[\frac{4t(9r^2 - 7t^2)}{r^6 (t^2 - r^2)^2} - \frac{15}{r^7} \ln \frac{(r-t)^2}{(r+t)^2} \right], \tag{15}$$

$$\begin{aligned} I_3 &= \int_0^t \int_0^t \frac{dt' dt''}{(\Delta t^2 - r^2)^5} = \frac{1}{8r} \frac{\partial I_2}{\partial r} \\ &= \frac{t}{768} \left[\frac{4t(57t^4 - 136 r^2 t^2 + 87 r^4)}{r^8 (t^2 - r^2)^3} + \frac{105}{r^9} \ln \frac{(r-t)^2}{(r+t)^2} \right]. \end{aligned} \tag{16}$$

In Eqs. (11) to (16) one must put

$$\begin{aligned} r &= \sqrt{n_x^2 a^2 + 2(1 - (-1)^{n_x}) y^2}, \\ r_x^2 &= n_x^2 a^2, \qquad r_y^2 = 2(1 - (-1)^{n_x}) y^2. \end{aligned} \tag{17}$$

For the slab space E_{16} the components of the dipole orientability indicator are obtained from those for E_{17} by setting $r_x^2 = r^2, r_y = 0$, and replacing $(-1)^{n_x}$ by 1 everywhere. Therefore, we have

$$ {}^{(y)}I^{E_{16}}_{v_x^2}(\mathbf{x},t) = -\frac{8p^2}{\pi^2 m^2} \sum_{n_x=-\infty}^{\infty}{}' I_1, \tag{18} $$

$$ {}^{(y)}I^{E_{16}}_{v_y^2}(\mathbf{x},t) = -\frac{4p^2}{\pi^2 m^2} \sum_{n_x=-\infty}^{\infty}{}' (I_1 + 3r^2 I_2), \tag{19} $$

$$ {}^{(y)}I^{E_{16}}_{v_z^2}(\mathbf{x},t) = -\frac{4p^2}{\pi^2 m^2} \sum_{n_x=-\infty}^{\infty}{}' (2I_1 + 3r^2 I_2), \tag{20} $$

in which $r = |n_x|a$.

2.3. *Analysis of the results*

Since equations (11)–(13) and (18)–(20) are too complicated to allow a straightforward interpretation, we plot figures for the components of the orientability indicator. Figures 1 and 2 arise from Eqs. (11)–(13) as well as (18)–(20), with the topological length $a = 1$ and $n_x \neq 0$ ranging from -50 to 50. In both figures the solid lines stand for the orientability indicator curves for the dipole in Minkowski spacetime

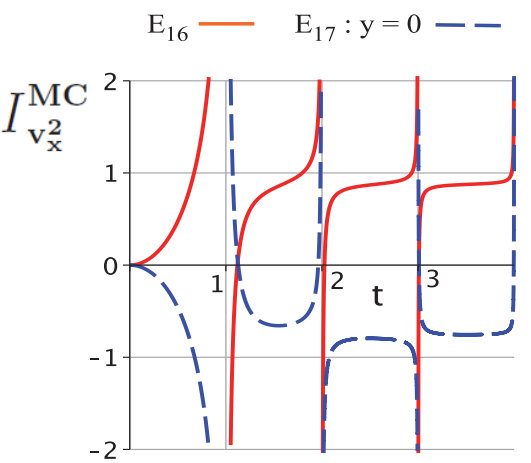


Fig. 1. Time evolution of the x-component of the orientability indicator in units of p^2/m^2 for a point electric dipole oriented in the flip y-direction in Minkowski spacetime with orientable E_{16} and non-orientable E_{17} spatial topologies, both with compact length $a = 1$. The solid and dashed lines stand, respectively, for the indicator curves for a dipole in 3−space with E_{16} and E_{17} topologies. For the globally inhomogeneous topology E_{17} the dipole is at $P_0 = (x, 0, z)$, thus freezing out the topological inhomogeneity. Both orientability indicator curves show a periodicity, but the curve for E_{17} exhibits a different kind of periodicity characterized by a distinctive inversion pattern. Non-orientability is responsible for this pattern of successive inversions, which is absent in the indicator curve for the orientable E_{16}.

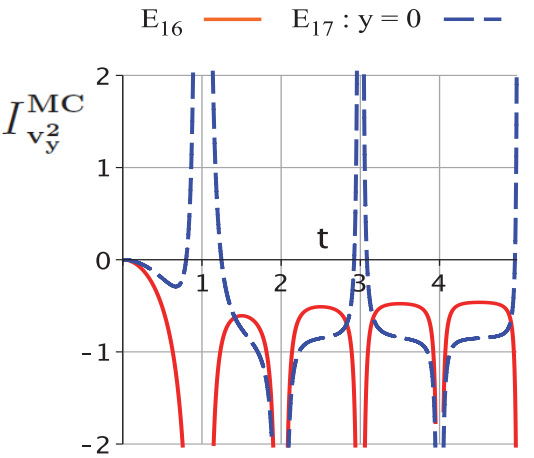


Fig. 2. Time evolution of the y-component of the orientability indicator for a point electric dipole oriented in the y-direction under the same conditions as those of Fig. 1. The orientability indicator curve for E_{17} also displays a characteristic inversion pattern but which is different from the one for the x-component shown in Fig. 1. For the y-component of the orientability indicator the signature of non-orientability can be recognized in the pattern of successive upward and downward "horns" formed by the dashed curve.

with E_{16} orientable spatial topology, whereas the dashed lines represent orientability indicator curves for the dipole located at $P_0 = (x, 0, z)$ in a 3−space with E_{17} non-orientable topology.

In the case of the x-component, the time evolution curves of the orientability indicator for E_{16} and E_{17}, shown in Fig. 1, present a common periodicity but with distinguishable patterns. The orientability indicator curve for E_{17} displays a distinctive sort of periodicity characterized by an inversion pattern. Non-orientability gives rise to this pattern of consecutive inversions, which is not present for the orientable E_{16}.

The differences become more striking when one considers the y-component of the orientability indicator, shown in Fig. 2. The non-orientability of E_{17} is disclosed by an inversion pattern whose structure is more striking than the one for the x-component. The orientability indicator curves for E_{17} form a pattern of alternating upward and downward "horns", making the non-orientability of E_{17} unmistakably identifiable. The z-component of the indicator behaves the same way.[26]

The characteristic inversion pattern exhibited by the dipole indicator curves makes it possible to identify the non-orientability of E_{17} in itself, without having to make a comparison with the indicator curves for its orientable counterpart, which is necessary in the point charge case.[26]

This discussion suggests that it may be possible to unveil a presumed spatial non-orientability by local means, namely by the stochastic motions of charged objects caused by quantum electromagnetic vacuum fluctuations. If the motion of a point

electric dipole is taken as probe, non-orientability can be intrinsically discerned by the inversion pattern of the dipole's orientability indicator curves.

3. Concluding remarks

In general relativity and quantum field theory spacetime is modeled as a differentiable manifold, which is a topological space equipped with a differential structure. It is usually assumed that the spacetime manifold is orientable and, additionally, that it is separately time and space orientable. The theoretical arguments usually offered to assume orientability combine the space-and-time universality of local physical experiments with physically well-defined (thermodynamically, for example) local arrow of time, violation of charge conjugation and parity (CP violation) and CPT invariance.[42–44] One can certainly use such reasonings in support of the standard assumptions on the global structure of spacetime.[b] Nevertheless, it is reasonable to expect that the ultimate answer to questions regarding the orientability of spacetime should rely on cosmological observations or local experiments, or might come from a fundamental theory of physics.

In the physics at daily and even astrophysical length and time scales, we do not find any sign or hint of non-orientability. This being true, the remaining open question is whether the physically well-defined local orientations can be extended continuously to cosmological and/or microscopic scales.

At the cosmological scale, one would think at first sight that to disclose spatial orientability one would have to make a trip around the whole 3−space to check for orientation-reversing paths. Since such a global journey across the Universe is not feasible one might think that spatial orientability cannot be probed globally. However, a determination of the spatial topology through the so-called circles in the sky,[47] for example, would bring out as a bonus an answer to the 3−space orientability problem at the cosmological scale.

We have addressed the question as to whether electromagnetic quantum vacuum fluctuations can be used to bring out the spatial orientability of Minkowski spacetime. We have found that there exists a characteristic inversion pattern exhibited by the orientability indicator curves for a dipole in the case of E_{17}, signaling that the non-orientability of E_{17} might be identified per se. The inversion pattern of the orientability indicator curves for the dipole is a signature of the reflection holonomy, and ought to be present in the orientability indicator curves for the dipole in all remaining seven non-orientable topologies with flip, namely the four Klein spaces (E_7 to E_{10}) and those in the chimney-with-flip class (E_{13} to E_{15}).

Observation of physical phenomena and controlled experiments are essential to our understanding of nature. Our results indicate that it might be possible to locally unveil spatial non-orientability through stochastic motions of point-like objects

[b]See Ref. 45 for a dissenting point of view, and also the related Ref. 46.

under electromagnetic quantum vacuum fluctuations. The present paper may be seen as a hint of a possible way to locally probe the spatial orientability of Minkowski spacetime.

Acknowledgements

M.J. Rebouças acknowledges the support of FAPERJ under a *CNE E-26/202.864/2017* grant, and thanks CNPq for the grant *306104/2017-2* under which this work was carried out. We also thank A.F.F. Teixeira and C.H.G. Bessa for fruitful discussions.

References

1. G.F.R. Ellis, Gen. Rel. Grav. **2**, 7 (1971); M. Lachièze-Rey and J.P. Luminet, Phys. Rep. **254**, 135 (1995); G.D. Starkman, Class. Quantum Grav. **15**, 2529 (1998); J. Levin, Phys. Rep. **365**, 251 (2002); M.J. Rebouças and G.I. Gomero, Braz. J. Phys. **34**, 1358 (2004); M.J. Rebouças, A Brief Introduction to Cosmic Topology, in *Proc. XIth Brazilian School of Cosmology and Gravitation*, eds. M. Novello and S.E. Perez Bergliaffa (Americal Institute of Physics, Melville, NY, 2005) AIP Conference Proceedings vol. **782**, p 188, also: arXiv:astro-ph/0504365; J.P. Luminet, Universe 2016, 2(1), 1.
2. M.J. Rebouças, *Detecting cosmic topology with primordial gravitational waves*, in preparation (2022).
3. P.A.R. Ade *et al.* (Planck Collaboration 2013), Astron. Astrophys. **571**, A26 (2014).
4. P.A.R. Ade *et al.* (Planck Collaboration 2015), Astron. Astrophys. **594**, A18 (2016).
5. N.J. Cornish, D.N. Spergel, G.D. Starkman and E. Komatsu, Phys. Rev. Lett. **92**, 201302 (2004).
6. J.S. Key, N.J. Cornish, D.N. Spergel, and G D. Starkman, Phys. Rev. D **75**, 084034 (2007).
7. P. Bielewicz and A.J. Banday, M.N.R.A.S. **412**, 2104 (2011).
8. P.M. Vaudrevange, G.D. Starkman, N.J. Cornish, and D.N. Spergel, Phys. Rev. D **86**, 083526 (2012).
9. R. Aurich and S. Lustig, M.N.R.A.S. **433**, 2517 (2013).
10. A. Bernui, C.P. Novaes, T.S. Pereira, G.D. Starkman, arXiv:1809.05924 [astro-ph.CO].
11. R. Aurich, T. Buchert, M.J. France, and F. Steiner, arXiv:2106.13205 [astro-ph.CO].
12. G. Gomero, B. Mota, and M.J. Rebouças, Phys. Rev. D **94**, 043501 (2016).
13. J.A. Wolf, *Spaces of Constant Curvature* (McGraw-Hill, New York, 1967).
14. W.P. Thurston, *Three-Dimensional Geometry and Topology. Vol.1*, Edited by Silvio Levy (Princeton University Press, Princeton, 1997).
15. C. Adams and J. Shapiro, American Scientist **89**, 443 (2001).
16. B. Cipra, *What's Happening in the Mathematical Sciences* (American Mathematical Society, Providence, RI, 2002).
17. A. Riazuelo, J. Weeks, J.P. Uzan, R. Lehoucq, and J.P. Luminet, Phys. Rev. D **69**, 103518 (2004).
18. H. Fujii and Y. Yoshii, Astron. Astrophys. **529**, A121 (2011).
19. C.H.G. Bessa and M.J. Rebouças, Class. Quantum Grav. **37**, 125006 (2020).
20. B.S. DeWitt, C.F. Hart and C.J. Isham, Physica **96A**, 197 (1979).
21. J.S. Dowker and R. Critchley, J. Phys. A **9**, 535 (1976).

22. P.M. Sutter and T. Tanaka, Phys. Rev. D **74**, 024023 (2006).
23. M.P. Lima and D. Muller, Class. Quant. Grav. **24**, 897 (2007).
24. D. Muller, H.V. Fagundes, and R. Opher, Phys. Rev. D **66**, 083507 (2002).
25. D. Muller, H.V. Fagundes, and R. Opher, Phys. Rev. D **63**, 123508 (2001).
26. N. A. Lemos and M. J. Rebouças, Eur. Phys. J. C **81**, 618 (2021).
27. H. Yu and L.H. Ford, Phys. Rev. **D 70**, 065009 (2004); 062501 (2009).
28. V.A. De Lorenci, C.C.H. Ribeiro, and M. M. Silva, Phys. Rev. D **94**, 105017 (2016).
29. M. Seriu and C.H. Wu, Phys. Rev. A **77**, 022107 (2008).
30. G. Gour and L. Sriramkumar, Found. Phys. **29**, 1917 (1999).
31. M. T. Jaekel and S. Reynaud, Quant. Opt. **4**, 39 (1992).
32. J.R. Weeks, *The Shape of Space*, 3rd ed. (CRC Press, Boca Raton, FL, 2020).
33. N.D. Birrel and P.C.W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).
34. J. Chen and H.W. Yu, Chin. Phys. Lett. **37**, 2362 (2004).
35. R. Lehoucq, M. Lachièze-Rey & J.-P. Luminet, *Astron. Astrophys.* **313**, 339 (1996).
36. G.I. Gomero, M.J. Rebouças, and A.F.F. Teixeira, Class. Quantum Grav. **18**, 1885 (2001).
37. G.I. Gomero, A.F.F. Teixeira, M.J. Rebouças, and A. Bernui, Int. J. Mod. Phys. D **11**, 869 (2002).
38. M.J. Rebouças Int. J. Mod. Phys. D **9** 561 (2000).
39. L.S Brown and G.J. Maclay, Phys. Rev. **184**, 1272 (1969).
40. V.A. De Lorenci, and C.C.H. Ribeiro, JHEP **1904**, 072 (2019).
41. O. Boada, A. Celi, J. Rodríguez-Laguna, J. Latorre, M. Lewenstein, New J. Phys. **17**, 045007 (2015).
42. Ya.B. Zeldovich and I.D. Novikov, JETP Letters **6**, 236 (1967).
43. S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).
44. R. Geroch and G.T. Horowitz, *Global structure of spacetimes* in General relativity: An Einstein centenary survey, pp. 212-293, Eds. S. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979).
45. M. Hadley, Testing the Orientability of Time. Preprints 2018, 2018040240.
46. M. Hadley, Class. Quantum Grav. **19**, 4565 (2002).
47. N.J. Cornish, D. Spergel, and G. Starkman, Class. Quantum Grav. **15**, 2657 (1998).

Essential self-adjointness of Dirac operators under the influence of general-relativistic gravity

Michael K.-H. Kiessling
Mathematics Department, Rutgers University,
Piscataway, NJ, 08854, USA
E-mail: miki@math.rutgers.edu

A. Shadi Tahvildar-Zadeh
Mathematics Department, Rutgers University,
Piscataway, NJ, 08854, USA
E-mail: shadit@math.rutgers.edu

Ebru Toprak
Mathematics Department, Yale University,
New Haven, CT, 06520, USA
E-mail: ebru.toprak@yale.edu

Physical reasoning gives expressions for the hamiltonian of a system of quantum-mechanical particles. These hamiltonians are often differential operators that are symmetric in a densely-defined domain. However, to study the dynamics of the unitary group corresponding to a hamiltonian, it is required that the hamiltonian be self-adjoint or essentially self-adjoint. This study analyzes the effect of the static non-linear electromagnetic-vacuum spacetime of a point nucleus on the self-adjointness and the spectrum of the general–relativistic Dirac hamiltonian for a test electron.

Keywords: Dirac hamiltonian, Reissner–Weyl–Nordström spacetime, self-adjointness

1. Introduction

In non-relativistic physics, whether Newtonian mechanics or quantum mechanics, the gravitational and the electrical attraction between a point electron and a point nucleus obey the same mathematical law and only their coupling strengths differ. Indeed, as the gravitational force is too small compared to the electrical force, it is common to ignore its effect in calculations of realistic hydrogenic spectra.

By contrast, if one models gravity the Newtonian way in the special-relativistic setting, the situation is more subtle. While Coulomb and Newton potential still have the same mathematical form, differing only in their coupling constants, they appear paired up with different Dirac matrices in the Dirac hamiltonian, viz.[a]

$$H := -i\alpha \cdot \nabla + \Big(m - \frac{GmM}{r}\Big)\beta - \frac{Qe}{r}$$

[a] In this note we take $c = \hbar = 1$.

where the hermitian matrices α_j satisfy $\alpha_j\alpha_k + \alpha_k\alpha_j = 2\delta_{jk}1_{\mathbb{C}^4}$, $0 \le j,k \le 3$, with $\alpha_0 = \beta$. Here, (m,e) are the mass and charge of the electron and (M,Q) are the mass and charge of the nucleus. It is known, see [6], that H is essentially self-adjoint if and only if

$$Q^2e^2 - G^2m^2M^2 \le \frac{3}{4}.$$

The self-adjoint extension has $\sigma_{ess}(H) = (-\infty,-m] \cup [m,\infty)$ with $\sigma_{ac}(H) = \sigma_{ess}(H)$ and $\sigma_{sc}(H) = \emptyset$. One can conclude that Newtonian gravity interaction has no significant effect on realistic hydrogenic spectra also in otherwise relativistic settings, as $G^2M^2m^2$ is an exceedingly tiny number[b] compared to Q^2e^2. For instance, one has $\frac{Gmm_p}{e^2} \approx 4.5 \cdot 10^{-40}$, where m_p is the mass of the proton.

On the other hand, the relative contributions of gravitation and electricity to the atomic spectra might be dramatically different, if one considers relativistic gravity. Einstein's gravity refines Newton's law of universal gravitation by providing a unified description of gravity as a geometric property of space and time, and it is formulated by Einstein's field equations. Coupling Einstein's field equations to Maxwell's electromagnetic field equations, we obtain the Einstein-Maxwell system of PDE's

$$\mathbf{R}_{\mu\nu} - \frac{1}{2}R\mathbf{g}_{\mu\nu} = \frac{8\pi G}{c^4}\mathbf{T}_{\mu\nu} \tag{1}$$

$$d\mathbf{F} = 0, \;\; d\mathbf{M} = 0, \tag{2}$$

which describes the geometry of a spacetime endowed with an electromagnetic field. Here, $\mathbf{R}$ is the Ricci curvature tensor, R is the scalar curvature of the metric $\mathbf{g}$, $\mathbf{F}$ is the Faraday tensor and $\mathbf{M}$ is the Maxwell tensor. Furthermore, $\mathbf{T}$ is the electromagnetic energy-momentum-stress tensor associated to $\mathbf{F}$. In order to measure the electrical and gravitational effects on the Dirac spectrum in the relativistic settings, one needs to analyze the relativistic Dirac equation

$$\gamma^\mu(-i\nabla_\mu + eA_\mu)\psi + mI_{4\times4}\psi = 0 \tag{3}$$

in a background that is a solution to (1), (2). Here, ∇_μ is the covariant derivative and A_μ is the electromagnetic 4-potential. In a curved spacetime, $\beta = \gamma^0$, $\beta\alpha^k = \gamma^k$ are related to the metric tensor by $\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}I_{4\times4}$. Tahvildar-Zadeh[5] showed the following: spherically symmetric, static, asymptotically flat solutions to (1), (2) that are topologically identical to $\mathbb{R}^{1,3}$ minus a timelike line, are covered by a single global chart and have a metric whose line element in spherical coordinates $(t,r,\theta,\phi) \in \mathbb{R}\times\mathbb{R}_+\times[0,\pi]\times[0,2\pi)$ reads

$$ds^2 = -f^2(r)dt^2 + f^{-2}(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{4}$$

$$f^2(r) = 1 - 2G\frac{m(r)}{r}, \quad m(r) = M - \mathcal{E}(r). \tag{5}$$

[b]Gravity could be significant for those nuclei satisfying $GM^2 > Q^2$, with mass M and electric charge Q, however, this inequality is not reachable for known nuclei.

Here r is the so-called area radius of a spherical orbit,[c] $m(r)$ is the radial mass function, and $\mathcal{E}(r)$ is the electrostatic field energy outside a ball of surface area $4\pi r^2$. Moreover, M is the ADM mass.

There are well-known examples of such solutions. In particular, $\mathcal{E}(r) = 0$ gives the Schwarzschild metric, and $\mathcal{E}(r) = \frac{Q^2}{2r}$ gives the Reissner–Weyl–Nordström (RWN) metric. In [2], Cohen and Powers considered the relativistic Dirac equation in the RWN spacetime. Note that the structure of the RWN spacetimes depends on the ratio of $\frac{Q^2}{GM^2}$. If $\frac{Q^2}{GM^2} \leq 1$, then $f^2(r)$ has zeros given by

$$r_\pm = GM\left(1 \pm \sqrt{1 - \tfrac{Q^2}{GM^2}}\right). \tag{6}$$

Also, in this case, the RWN spacetime features a dynamical black hole region in $r_- < r < r_+$.[d] Otherwise, $f^2(r)$ is always positive, and the manifold has a naked singularity at $r = 0$. In [2], it was obtained that in the naked singularity sector, the Dirac hamiltonian $\mathcal{H}$ associated with the relativistic Dirac equation is not essentially self-adjoint.

The RWN spacetime of a nucleus has some undesirable features.[7] It has a very strong curvature singularity at its center associated with a negative infinite bare mass, arising from $\mathcal{E}(r)$. In fact, the electrostatic energy density is not integrable about $r = 0$, whereas the system has finite ADM mass, interpreted as the total energy of the spacetime. In Section 2 we answer the question whether a more mildly singular $f^2(r)$, which is compatible with (4), suffices to guarantee an essentially self-adjoint $\mathcal{H}$ for a test electron.

In RWN settings, we say that $f^2(r)$ yields a *subextremal* black hole if $r_- < r_+$; and $f^2(r)$ yields an *extremal* black hole if $r_- = r_+ =: r_0$. In [2], Cohen and Powers also investigated (3) in the static region outside the event horizon of a subextremal (and extremal) RWN black hole, i.e. when $r \geq r_+$. They found that $\mathcal{H}$ is essentially self-adjoint, that its self-adjoint extension has essential spectrum equal to the real line, without a gap, and that it has no eigenvalues at all. Thus, the electron is expected to be swallowed by the black hole unless it escapes to ∞. Therefore, the absence of bound states of an electron outside of the event horizon of a RWN black hole is not a surprising result. In Section 3 we complement this study by considering the static region inside the Cauchy horizon of the subextremal RWN black hole spacetime, i.e., when $0 \leq r \leq r_-$, respectively inside the event horizon of the extremal RWN black hole spacetime, i.e., when $0 \leq r \leq r_0$.

[c]Every point in the stipulated spacetime is an element of a unique orbit under a Killing vector flow corresponding to the $SO(3)$ symmetry, and this orbit is a scaled copy of $\mathbb{S}^2$ with area $4\pi r^2$, *defining* $r > 0$.

[d]Note that $r = r_+$ and $r = r_-$ are the boundaries of the dynamical region and not themselves dynamical. In particular, in the extremal black hole case, there is no dynamical part of the spacetime.

2. Electrostatic spacetimes with negative bare mass and no horizon

Due to the spherical symmetry and static character of the spacetimes, the Dirac operator $\mathcal{H}$ of a test electron in the curved space whose line element ds^2 is given by (4) separates in the spherical coordinates and their default spin frame.[2] More precisely, we study the family of radial Dirac operators H_k, $k \in \mathbb{Z}\backslash\{0\}$, with

$$H_k := \begin{bmatrix} f(r) - e\phi(r) & \frac{k}{r}f(r) - f^2(r)\partial_r \\ \frac{k}{r}f(r) + f^2(r)\partial_r & -f(r) - e\phi(r) \end{bmatrix}, \tag{7}$$

acting on $g(r) := \big(g_1(r), g_2(r)\big)^T$, with a weighted L^2 norm given by

$$\|g\|^2 := \int_0^\infty \frac{1}{f^2(r)}\Big(|g_1(r)|^2 + |g_2(r)|^2\Big)dr. \tag{8}$$

Recall that we want to consider $f^2(r)$ having singularity at zero that is milder than the RWN spacetime has. For this purpose, we make the following assumptions on $m(r)$:

- $m(r)$ is continuous for $r > 0$;
- $m(r)/r < 1/2G$;
- $m(r) \sim -C_\alpha r^{-\alpha}$ as $r \searrow 0$, where $\alpha \geq 0$ and $C_\alpha > 0$;
- $m(r) \to M > 0$ as $r \to \infty$.

Our first two assumptions on $m(r)$ are equivalent to ruling out black holes in spacetimes with the line element (4). So our spacetimes feature a charged naked singularity. By the third and fourth assumptions on $m(r)$, the naked singularity has a negative *bare mass* ($\lim_{r\downarrow 0} m(r)$), which is finite only for $\alpha = 0$, in which case $m(0) = -C_0$.

The function $\phi(r)$ is the potential of the electrostatic field generated by the nucleus, and we have $\mathcal{E}(r) = \frac{1}{2}\int_r^\infty |\phi'(s)|^2 s^2 ds$. We also assume

- $\phi(r)$ is continuously differentiable;
- $\phi(r) \sim Ze/r$ as $r \to \infty$;
- $\phi(r) \sim C''_\beta + C'_\beta r^{-\beta}$ as $r \searrow 0$, with $\beta \leq 1$.

Before we give the main theorem, we need to define the unitary operator U. Let $r(x)$ be the solution to the differential equation

$$\frac{d}{dx}r(x) = f^2(r(x)) \tag{9}$$

subject to the conditions that $r(x) \to 0^+$ as $x \to 0^+$. We now define U as

$$(Ug_i)(x) = g_i(r(x))f(r(x))^{1/2} \tag{10}$$

for (g_1, g_2) in the subspace induced by the norm (8). Let also $\widetilde{H}_k = UH_kU^{-1}$. Then, we have the following theorem. Interested readers can find the details of the proof in [3].

Theorem 2.1. *Under the stated assumptions on $m(r)$ and $\phi(r)$ the operator $\widetilde{H}_k$, and, therefore H_k, for $k \in \mathbb{Z}\backslash\{0\}$, has uncountably many self-adjoint extensions. The extensions are given by $\widetilde{H}_{k:\theta} = \widetilde{H}_k|_{D_\theta}$ where*

$$D_\theta = \Big\{(f_1, f_2) \in D(\widetilde{H}_k^*); \quad f_1(0) = -f_2(0)\tan\theta, \ -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}\Big\}.$$

Moreover, for each θ

(a) The essential spectrum $\sigma_{ess}(\widetilde{H}_{k:\theta}) = (-\infty, -m] \cup [m, \infty)$;
(b) $\widetilde{H}_{k:\theta}$ has purely absolutely continuous spectrum in $(-\infty, -m) \cup (m, \infty)$;
(c) the singular continuous spectrum $\sigma_{sc}(\widetilde{H}_{k:\theta}) = \emptyset$.

Note that this theorem concludes that a milder singularity with strictly negative bare mass than that of RWN does not yield an essentially self-adjoint H_k. On the contrary, it appears that the self-adjointness might require a stronger singularity such as when $\beta > \frac{\alpha+3}{2}$. However, these values of α and β do not match with the asymptotics of the mass functions for the electrostatic spacetimes in (4).

It turns out that there is a way to restore the essential self-adjointness of the Dirac operator, namely by including a sufficiently large anomalous magnetic moment of the electron. This has been carried out in the naked RWN geometry in [1].

Assuming a nonzero anomalous magnetic moment, the radial partial-wave Dirac operator is given by

$$H_{\mu_a,k} := H_k + \begin{bmatrix} 0 & -\mu_a\phi'(r)f(r) \\ -\mu_a\phi'(r)f(r) & 0 \end{bmatrix}. \tag{11}$$

Regarding $H_{\mu_a,k}$, we have the following theorem:

Theorem 2.2. *Let $m(r)$ be as in the Theorem 3.1 with $\alpha \geq 0$ and $\phi(r) = C''_\beta + C'_\beta r^{-\beta} + O_1(r^{\frac{3}{2}-\beta})$ for $\beta \leq 1$, where $f \in O_k(g)$ indicates $\frac{d^j}{dr^j}f = O(\frac{d^j}{dr^j}g)$ for $j = 0, 1, ..k$. Then the operator $H_{\mu_a,k}$ is essentially self-adjoint if either $\beta > \frac{1+\alpha}{2}$, or $\beta = \frac{1+\alpha}{2}$ and $|\mu_a| \geq \frac{2+\alpha}{1+\alpha}\sqrt{2GC_\alpha}/|C'_\beta|$. On the other hand, if $\beta < \frac{1+\alpha}{2}$, then $H_{\mu_a,k}$ has multiple self-adjoint extension.*

An important consequence of this theorem is the fact that the Dirac operator in the Hoffmann spacetime is not self-adjoint unless $m(0) = 0$; in the negative bare mass case the anomalous magnetic moment cannot restore essential self-adjointness. Recall, the electric potential in the Hoffmann spacetime is given as

$$\phi(r) = Q\int_r^\infty \zeta'\Big(\frac{Q^2}{2s^4}\Big)\frac{ds}{s^2}, \ \ \zeta(\mu) = \sqrt{1+2\mu} - 1.$$

3. Point nucleus as singularity in static interior of black hole spacetime

Recall that in the RWN spacetime, the line element $f^2(r)$ in (4) arises as

$$f^2(r) = 1 - 2\frac{GM}{r} + \frac{GQ^2}{r^2} = \frac{1}{r^2}(r - r_-)(r - r_+) \tag{12}$$

where r_- and r_+ are given as in (6). We consider the operators H_k given in (7) when $0 \leq r \leq r_-$. Note that, in the extremal case, we have $r_- = r_+ = r_0$.

As in the previous section, we define $\widetilde{H}_k = UH_kU^{-1}$, where U is defined as in (10). We use (12) in the change of variable (9), with $x = 0$ when $r = 0$, which for the subextremal sector yields

$$x = r + \frac{r_+^2}{r_+ - r_-} \ln\left(1 - \frac{r}{r_+}\right) - \frac{r_-^2}{r_+ - r_-} \ln\left(1 - \frac{r}{r_-}\right); \qquad r < r_-; \tag{13}$$

in the extremal limit $r_- \nearrow r_0$ & $r_+ \searrow r_0$ this becomes

$$x = r_0 \left[\frac{1}{1 - \frac{r}{r_0}} + 2\ln\left(1 - \frac{r}{r_0}\right) - \left(1 - \frac{r}{r_0}\right)\right], \qquad r < r_0. \tag{14}$$

It is clear from (13) and (14), one has $r - r_- \sim e^{-ax}$ as $x \to \infty$ in the subextremal case, whereas $r - r_0 \sim x^{-1}$ as $x \to \infty$ in the extremal case. This difference between the asymptotics result in differences in the spectrum of the corresponding Dirac operators, see [4] for more details. We have,

Theorem 3.1. *The operator $\widetilde{H}_k$ has uncountably many self-adjoint extensions for both the subextremal and the extremal black hole sector. The extensions are given by $\widetilde{H}_{k:\theta} = \widetilde{H}_k|_{D_\theta}$ where*

$$D_\theta = \left\{(f_1, f_2) \in D(\widetilde{H}_k^*); \quad f_1(0) = -f_2(0)\tan\theta, \ -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}\right\}.$$

Moreover for each θ,

- *(a) The essential spectrum $\sigma_{ess}(\widetilde{H}_{k:\theta}) = \mathbb{R}$ in both subextremal and the extremal case.*
- *(b) $\widetilde{H}_{k:\theta}$ has absolutely continuous spectrum in $\mathbb{R}$ in both subextremal and the extremal case. Moreover, $\widetilde{H}_{k:\theta}$ has no eigenvalue in the subextremal case.*
- *(c) The singular continuous spectrum $\sigma_{sc}(\widetilde{H}_{k:\theta}) = \emptyset$ in both subextremal and the extremal case.*

In the extremal case the only possible eigenvalue is $-\frac{Q}{r_0}$, and whether it is an eigenvalue or not is not determined in this study. However, we believe that the answer should depend on the values of Q and r_0.

Similar to our previous section, we also considered the effect of an anomalous magnetic moment. Consider (11) with $\phi(r) = Q/r$. We have

Theorem 3.2. *In both the subextremal and the extremal case, the operator $H_{\mu_a,k}$ is essentially self-adjoint iff $|\mu_a| \geq \frac{3}{2}\sqrt{G}$. Moreover, in both subextremal and extremal cases, the self-adjoint operators have $\sigma_{ess}(H_{\mu_a,k}) = \sigma_{ac}(H_{\mu_a,k}) = \mathbb{R}$, $\sigma_{sc}(H_{\mu_a,k}) = \emptyset$, and $\sigma_p(H_{\mu_a,k}) = \emptyset$.*

References

1. Belgiorno, F., Martellini, M., and Baldicchi, M., *Naked Reissner–Nordström singularities and the anomalous magnetic moment of the electron field*, Phys. Rev. D **62**:084014, (2000).
2. Cohen, J.M. and Powers, R.T., *The general-relativistic hydrogen atom*, Commun. Math. Phys. **86**:96–86 (1982).
3. Kiessling, M.K.-H., Tahvildar-Zadeh, A.S., and Toprak, E., *On general-relativistic hydrogen and hydrogenic ions*, J. Math. Phys. 61, 092303 (2020).
4. Kiessling, M.K.-H., Tahvildar-Zadeh, A.S., and Toprak, E., *On the Dirac operator for a test electron in a Reissner–Weyl–Nordström black hole spacetime*, General Relativity and Gravitation (2021), 53 (1).
5. Tahvildar-Zadeh, A.S., *On the static spacetime of a single point charge*, Rev. Math. Phys. **23**:309–346 (2011).
6. Thaller, B., *The Dirac equation*, Springer, Berlin – Heidelberg (1992).
7. Weyl, H., *Space, Time, Matter*, 4th. ed., Dover (1952).

Relic magnetic wormholes as possible source of toroidal magnetic fields in galaxies

A. A. Kirillov* and E. P. Savelova

Department of Mathematical Modeling, Bauman Moscow State Technical University, Moscow, 105005, Russian Federation

**E-mail: kirillov@bmstu.ru*

We present the hypothesis that some of ring galaxies were formed by relic magnetic torus - shaped wormholes. In the primordial plasma before the recombination magnetic fields of wormholes trap baryons whose energy is smaller than a threshold energy. They work as the Maxwell's demons collecting baryons from the nearest (horizon size) region and thus forming clumps of baryonic matter which have the same torus-like shapes as wormhole throats. Such clumps may serve as seeds for the formation of ring galaxies and smaller objects having the ring form. Upon the recombination torus-like clumps may decay and merge. Unlike galaxies, such objects may contain less or even no dark matter in halos. However, the most stringent feature of such objects is the presence of a large - scale toroidal magnetic field. We show that there are threshold values of magnetic fields which give the upper and lower boundary values for the baryon clumps in such protogalaxies.

Keywords: Wormholes; cosmic rays; ring galaxies; galactic magnetic fields.

1. Introduction

One of challenges of modern astrophysics is to finding possible traces of relic wormholes. It is expected that relic cosmological wormholes were created from virtual wormholes on quantum stage, or during the inflationary period of evolution of the Universe. Virtual wormholes exist at Planckian scales and they represent the most natural objects to compose spacetime foam picture.[1,2] Their existence is straightforwardly predicted by lattice quantum gravity models, e.g., see Refs.[3–6] and references therein.

Spherically symmetric relic wormholes are highly instable and collapse very rapidly. To be stable they require the presence of exotic matter at throats or an essential modification of general relativity.[7,8] This means that all primordial spherical wormholes have collapsed long ago and at present they cannot probably be distinguished from black holes. However less symmetric configurations can be made stable without exotic matter or any modification of GR. First rigorous example was presented in Ref.[9] It was demonstrated that in the open Friedmann model a stable wormhole can be obtained simply by the factorization of space over a discrete subgroup of the group of motions of space. To illustrate such a factorization one may consider flat plane in which one of coordinates, say x, becomes periodic $x = R\varphi/(2\pi)$ with the angle $0 < \varphi < 2\pi$. Then the space becomes a cylinder with

the radius R. In the open model (on space of a constant negative curvature) any parallel geodesics ($\varphi = 0$ and $\varphi = 2\pi$) diverge and the distance between them changes $R(\ell)$, where ℓ is a parameter along the geodesic line. Therefore, if we move along the geodesic line from the point where the distance is the shortest R_{min}, the space opens out $R \to \infty$ and becomes unrestricted.[10] Here the wormhole configuration coincides with a cylinder on which the metric is specified to produce a constant negative curvature. The simplest 3D wormhole obtained by the factorization has also the topology of a cylinder, while its throat has the topology of the 2D- torus.[9,10] In flat space such a configuration corresponds to a wormhole whose throat has the shape of a doughnut. Such wormholes are not static but expand in agreement with the expansion rate of the Universe. It can be considered as having been frozen into space and, therefore, they are static in the co-moving coordinates. It is important that the factorization allows to get an arbitrary number of such wormholes in space.

In flat space the torus - like wormholes become dynamical objects and evolve.[9] Whether they are static, expand, or collapse, depends on surrounding matter and peculiar motions of throats. The shape of throats of such wormholes resembles a doughnut and is characterized by two radii R_w and r_w. In the limit $R_w \gg r_w$ it can be approximately described by the cylindrical (axial) configuration. It turns out that static and stationary cylindrical wormhole solutions do exist and it was found that asymptotically flat wormhole configurations do not require exotic matter violating the weak energy condition.[11,12] This shows that such objects, as doughnut - shaped wormholes, have all chances to be observed in astrophysical systems. We point out that the investigation of the evolution of doughnut-shaped wormholes in flat space is a rather complex problem.

The possibility to directly observe wormholes attracts the more increasing attention, e.g., see Refs.[13–15] At first glance the most promising are collective effects produced by a distribution of wormholes in space. However, our previous investigation have shown that observational effects of a distribution of wormholes are very well hidden under analogous effects produced by ordinary matter e.g., see Refs.[16,17] The only exclusion may be the noise (stochastic background) produced by the scattering of emitted by binaries gravitational waves on wormholes.[18]

In general a single wormhole produces much less noticeable effects (lensing, cosmic ray scattering, etc.). However, wormholes may possess non-trivial magnetic fields as vacuum solutions. In this case possible imprints of wormholes in the present picture of the Universe may be rather considerable. In particular, when such a magnetic wormhole gets close to a galaxy, it starts to work as an accelerator of charged particles[10,19] which is capable of explaining the origin of high-energy cosmic-ray particles.[20] The idea that the observed cosmic rays require magnetic fields for their creation was first suggested by Fermi.[21] In voids a relic magnetic wormhole works simply as a generator of synchrotron radiation and can be detected via the magnetic field.[22,23] The primordial magnetic fields[24] in turn may form small-scale non-linear clumps of baryonic matter[25,26] and as it was recently discussed in Ref.[27] they may

allow to solve the existing Hubble tension. In other words, magnetic wormholes should leave a clear imprint on the sky. In particular, relic magnetic wormholes may play also the key role in formation of ring type baryonic structures, analogous to the ring galaxies without involving dark matter.[28] Moreover, such wormholes may also form ring - type structures even in the present epoch which may explain the origin of the recently found unexpected class of astronomical ring-type objects.[29]

2. Vacuum magnetic fields of wormholes

Nontrivial topology of space which contains a static wormhole allows us to get additional nontrivial solutions of static vacuum Maxwell equations.[30] This means that already in the absence of real sources (charged particles, electric currents) space may possess nontrivial quasi-static magnetic and electric fields. The physical mechanism is rather clear, during the quantum period of the Universe when the wormhole forms, it may capture some portion of closed magnetic or electric lines. Such lines cannot simply leave the wormhole. Electric fields perform work and decay very rapidly in the primordial plasma. Magnetic fields do not perform work and may survive till the present days. For exact solutions which involve magnetic fields and wormholes see e.g.,[31,32]

In this section for the sake of simplicity we assume that sufficiently far from the wormhole entrances the space is flat. Generalization to the curved spacetime and consideration of magnetic fields in the Friedmann model can be found, e.g., in.[33] Consider first the simplest genus $n = 0$ wormhole. Then the space-time metric can be taken as

$$dl^2 = c^2dt^2 - h^2(r)\left(dx^2 + dy^2 + dz^2\right). \tag{1}$$

We shall use the Ellis–Bronnikov massless wormhole[34,35] when the scale function is simply $h = 1 + \frac{R^2}{r^2}$. This model of a wormhole possesses two asymptotic Euclidean spaces E_+ as $r \gg R$ and E_- as $r \ll R$. In the region $r \ll R$ the transformation $\widetilde{r} = R^2/r$ reduces the above metric to the standard Euclidean form. The Maxwell equations for magnetic field take the standard form

$$rot\mathbf{B} = div\mathbf{B} = 0, \tag{2}$$

where $B_i = \varepsilon_{ijk}F^{jk}$. These equations possess a nontrivial solution in the form

$$\mathbf{B} = -\frac{Q}{r^2h}\mathbf{n} \tag{3}$$

with an arbitrary constant value Q, where $\mathbf{n} = \mathbf{r}/r$ is the unit vector. Indeed, vacuum magnetic field can be expressed via the magnetic scalar potential $\mathbf{B} = -\nabla\phi$ which obeys the equation

$$\Delta\phi = 0. \tag{4}$$

In the spherically symmetric case this equation reduces to

$$\frac{1}{r^2h^3}\partial_r r^2 h \partial_r \phi = 0 \tag{5}$$

and has the solution as

$$\phi = \int_0^r \frac{Q}{r^2 h} dr + \phi_0. \tag{6}$$

This defines the magnetic field

$$\mathbf{B} = -\nabla\phi = -\frac{Q}{r^2 h}\nabla r = -\frac{Q}{r^2 h}\mathbf{n}. \tag{7}$$

This solution works for any spherically symmetric wormhole with an arbitrary scale function $h(r)$ in (1).

In the region E_+ ($r \gg R$) it describes the field of the magnetic charge Q homogeneously distributed over the sphere $r = R$ (Coulomb law). In the region E_- ($r \ll R$) the transformation $\widetilde{r} = R^2/r$ interchanges the inner and outer regions of the sphere $r = R$ and we get the same field with the magnetic charge $-Q$.

In the case when both entrances are in the same space this transforms to the dipole field. Let the positions of the two spheres are $\mathbf{x}_+$ and $\mathbf{x}_-$ then the field can be taken as ($h_\pm = h(r_\pm) \sim 1$ as $r_\pm = |\mathbf{x} - \mathbf{x}_\pm| \gg R$)

$$\mathbf{B}(\mathbf{x}) = \frac{Q\,(\mathbf{x} - \mathbf{x}_+)}{h_+\,|\mathbf{x} - \mathbf{x}_+|^3} - \frac{Q\,(\mathbf{x} - \mathbf{x}_-)}{h_-\,|\mathbf{x} - \mathbf{x}_-|^3}. \tag{8}$$

Consider now the genus $n = 1$ wormhole. In the flat space such a wormhole can be constructed by means of cutting two solid tori and gluing along their surfaces. In this case we have two different new kinds of solutions. First is obtained by placing an arbitrary magnetic charge density $\rho(x)$ in the internal region of one torus and the opposite density $-\rho(x)$ in the internal region of the second torus. Then we solve the system

$$rot\mathbf{B} = 0, \quad div\mathbf{B} = 4\pi\rho. \tag{9}$$

Recall that internal regions of tori correspond to fictitious points, while $\rho(x) = 0$ as x lies outside the tori. Therefore, such a system coincides exactly with (2). In this case the exact form of the magnetic field is rather complicated. However averaging over orientations of the tori we restore the spherical symmetry of the wormhole and get exactly the same solution as (8). In this sense for the sake of simplicity we may always restrict to the spherically symmetric wormholes (i.e., consider $n = 0$ wormhole as the first order approximation).

Solutions of the second kind can be obtained by placing an arbitrary current density $\mathbf{j}(x)$ within the surface of torus/throat and solving the system

$$rot\mathbf{B} = \frac{4\pi}{c}\mathbf{j}, \quad div\mathbf{B} = 0. \tag{10}$$

Again here the non-vanishing current density corresponds only to fictitious points, while in physical regions of space such a system represents the same system (2).

The system (10) corresponds to the field generated by an electric loop which produces the toroidal magnetic field. Such solutions cannot be reduced to the spherically symmetric case, since upon the averaging over orientations of the loop the field vanishes.

In conclusion of this section we point out that the new two classes of vacuum solutions reflect the topological non-triviality of space. Indeed, according to the Stocks theorem the system (2) implies $\oint \mathbf{Bdl} = 0$ for any loop which can be pulled to a point. In the case of a non-trivial topology of space there appear new classes of loops Γ_a which cannot be contracted to a point and, therefore, to fix the unique solution we have to fix additional boundary data $\oint_{\Gamma_a} \mathbf{Bdl} = C_a$. In general $C_a \neq 0$. In the case of genus $n = 0$ wormhole there is only one such a non-trivial loop which goes through the wormhole throat. In the case of genus $n = 1$ wormhole we have already two such loops, one goes through the throat (it corresponds to the system (9)) and one additional crosses the torus (the system (10)).

3. Wormhole as an accelerator

In galaxies charged particles undergo an acceleration when interacting with galactic radiation. For definiteness we shall speak of electrons. Indeed, by means of the Compton scattering photons transmit part of their momentum to electrons. Upon the scattering the momentum obtained by the electron from the incident photon is

$$\mathbf{\Delta p} = \mathbf{p}' - \mathbf{p} = \frac{h\nu}{c}\left(\mathbf{n} - \frac{\nu'}{\nu}\mathbf{n}'\right) \tag{11}$$

where $\mathbf{n}$ and $\mathbf{n}'$ are the direction of the photon before and after scattering and

$$\frac{\nu'}{\nu} = \frac{\left(1 - \frac{c}{E}\mathbf{pn}\right)}{\left(1 + \frac{h\nu}{E}\left(1 - \mathbf{nn}'\right) - \frac{c}{E}\mathbf{pn}'\right)}. \tag{12}$$

Upon averaging over possible orientations of $\mathbf{n}'$ we get for the average momentum transmitted from the incident photon

$$\mathbf{\Delta p} = \frac{1}{c}\beta_\nu(\mathbf{p}) h\nu \mathbf{n}. \tag{13}$$

Here the spectral coefficient $\beta_\nu(\mathbf{p}) = 1 - \left\langle \frac{\nu'}{\nu}\mathbf{nn}' \right\rangle$ is given by

$$\beta_\nu(\mathbf{p}) = 1 - \frac{1}{4\pi}\int \left(\frac{\nu'}{\nu}\mathbf{nn}'\right) d\Omega', \tag{14}$$

where $\frac{\nu'}{\nu}$ is determined by (12). In the case when $\mathbf{p} = p\mathbf{n}$, it reduces to the form

$$\beta_\nu = 1 + \frac{1}{2x}\left[2 - \left(1 + \frac{1}{x}\right)\ln\left(1 + 2x\right)\right],$$

where $x = \left(h\nu + cp\right)\left(E + cp\right)/m^2c^4$. Now multiplying (13) on the number density of photons with the frequency ν and on the cross section we get the spectral force

which accelerates the electron in the form

$$\mathbf{f}_\nu = \frac{\Delta \mathbf{p}}{\Delta t} = c\sigma_T N_\nu \frac{1}{c}\beta_\nu(p) h\nu \mathbf{n} = \frac{\sigma_T \beta_\nu(p)}{c}\mathbf{P}_\nu \tag{15}$$

where $\mathbf{P}_\nu$ is the spectral component of the Poynting vector $\mathbf{P}_\nu = \frac{c}{4\pi}\mathbf{E}_\nu \times \mathbf{B}_\nu$ and σ_T is the Thomson cross section. The total force is given by $\mathbf{F} = \int \mathbf{f}_\nu d\nu$.

It is important that the force is determined by the Poynting's vector $\mathbf{P}_\nu$. In a quite (quasi-stationary or steady state) galaxy both the Poynting's vector and the force have the potential character, i.e., they can be presented as $\mathbf{P}_\nu = -\nabla\Psi_\nu$. Non-stationary processes in active galactic nuclei may produce some additional acceleration, e.g., the stochastic Fermi acceleration, etc., which we do not discus here. For the quasi-stationary galaxy the Poynting's theorem gives the discontinuity equation

$$div\mathbf{P}_\nu = \ell_\nu \tag{16}$$

where $\ell_\nu(x)$ is the spectral density of sources of radiation (stars, hot gas, dust, etc.) or the radiative capability of a unite volume in the galaxy. If the topology is simple, then sufficiently far from the galaxy we get

$$\mathbf{P}_\nu = \frac{M_\nu}{4\pi r^2}\mathbf{l}, \tag{17}$$

where $\mathbf{l} = \mathbf{r}/r$, $\mathbf{r}$ is the distance from the center of the galaxy, and M_ν is the total spectral energy emitted by the galaxy in the unit time. Observations show that the intergalactic medium possesses a magnetic field.[22,23] Therefore, the electron may have a closed trajectory. It is easy to verify that the total energy obtained by the electron from the galactic radiation during the cycle is exactly zero, i.e., $\oint \mathbf{f}_\nu \mathbf{dl} \equiv 0$. This means that in the case when topology is simple, the only possible mechanism of the electron acceleration relates to high-energy non-stationary processes (jets, shock waves, supernovae explosions, active galactic nuclei, etc.).

The situation changes when the galaxy is accompanied with a wormhole. In the presence of the wormhole the Poynting's field also admits non-trivial solutions of (16). For the sake of simplicity we consider the spherically symmetric (genus $n = 0$) wormhole. For a more general wormhole the rough picture remains the same, at least from the qualitative standpoint. We point out that the genus $n \geq 1$ wormholes are more preferred from the astrophysical standpoint, since they may work as accelerators even in the case when a wormhole is not traversable (e.g., when the length of the trajectories which go through the throat are too big).

Indeed, the scattering of the radiation on the wormhole (e.g., see[16,17]) produces an additional field in the dipole form

$$\delta\mathbf{P}_\nu = \frac{\delta M_\nu}{4\pi r_-^2}\mathbf{n}_- - \frac{\delta M_\nu}{4\pi r_+^2}\mathbf{n}_+ \tag{18}$$

where $\mathbf{n}_\pm = \mathbf{r}_\pm/r_\pm$, $\mathbf{r}_\pm = \mathbf{r} - \mathbf{x}_\pm$, $\mathbf{x}_\pm$ are positions of the wormhole entrances (we assume that $r_+ \ll r_-$), $\delta M_\nu \simeq \frac{M_\nu \pi R^2}{4\pi r_+^2}$ is the portion of the spectral energy absorbed

by the closest entrance into the wormhole throat and R is the radius of the throat. If the wormhole possesses a magnetic field in the form (8), it forms a magnetic trap for the electron, while the Poynting's vector field $\delta\mathbf{P}_\nu$ in form (18) forms the accelerating force which acts exactly along the magnetic lines. On every cycle the electron will gain the energy from radiation $A = \int A_\nu d\nu > 0$, where $A_\nu = \oint \delta\mathbf{f}_\nu \mathbf{dl}$, whose exact value depends on the length of the trajectory of the electron and on all the rest parameters of the wormhole (distance to the galaxy, throat size, etc.). Some part of this energy will be spent on the synchrotron radiation of electrons and, therefore, there is a competition between the acceleration produced by the galactic radiation and loss of energy on the reradiation. The reradiation can be accounted for by adding the standard force of the radiation friction.

4. Magnetic wormholes as baryon traps

Consider a single wormhole whose throat has the shape of a torus (the genus - 1 wormhole by the classification suggested in Ref.[10]). In the presence of such a wormhole Maxwell's equations possess two additional classes of non-trivial vacuum solutions. Indeed, according to the Stocks theorem the system of vacuum Maxwell equations implies $\oint \mathbf{Bdl} = 0$ for any loop which can be pulled to a point (where $\mathbf{B}$ is the magnetic field). In the case of a non-trivial topology of space[a] there appear new classes of loops Γ_a which cannot be contracted to a point and, therefore, to fix the unique solution we have to fix additional boundary data $\oint_{\Gamma_a} \mathbf{Bdl} = \frac{4\pi}{c} I_a$ and in general $I_a \neq 0$. The constants I_a depend only on time and they can be viewed as fictitious currents[b] which intersect the loops Γ_a. In the case of genus $n = 0$ (spherical) wormhole there is only one such a non-trivial loop which goes through the wormhole throat. In the case of genus $n = 1$ wormhole (doughnut - shaped throat) we have already two such loops, one goes through the throat and one additional goes through the hole in the center of the doughnut and surrounds the throat.

The first class produces the magnetic field of a wormhole which can be described by magnetic poles placed in two different entrances into the throat. The two entrances have opposite magnetic poles. If the distance between the entrances is big enough the resulting field is very weak and when crossing such a field high-energy charged particles only slightly change the direction of propagation. The field can be strong only very close to the throat entrances. However, since charged particles can

[a]The simplest scheme to get a general genus - n wormhole can be described as follows. We take a couple of equal spheres with n handles in space, remove the internal regions (insides of the spheres), and glue along their surfaces (the so-called Hegor diagrams). If we take two simple spheres without handles the resulting space corresponds to a spherical wormhole (throat is the sphere). The sphere with a handle is the torus. A couple of toruses corresponds to the doughnut shaped wormhole, etc.
[b]We point out that the currents are fictitious, for from the point of view of the Hegor diagrams they take place in inner regions of spheres which are removed, i.e., in fictitious regions.

freely propagate along the field lines (which are roughly orthogonal to entrances), the particles captured by the field are distributed in the whole region between the entrances. Such fields have the long-range character and can be used to explain the origin of long-correlated magnetic fields in voids[22,23] and, more generally, of primordial magnetic fields.[24]

The situation changes when the wormhole possesses also the field of the second class. The second class corresponds to the field produced by a single loop of a current (the loop of the corresponding fictitious current goes inside of the surface of the doughnut). In this case the field lines of force repeat the shape of the entrance (the shape of a doughnut) which roughly corresponds to the fields observed in spiral galaxies.[36–38] It is necessary to point out that the magnetic fields in the galaxies cannot exactly repeat the shape of a doughnut. There are a number of well-known processes in galaxies which naturally lead to the generation of galactic magnetic fields[37] and the actual field does not reduce to such a simple model. It is known that the pitch angle (between the toroidal field and radial one) is usually between 4 and 17 degrees.

The most strong field is close to the entrance and particles captured by the field remain always near the entrance. In the primordial plasma before the recombination such a wormhole traps all baryons[c] propagating near it and thus forms a primeval structure of a galaxy (a protogalaxy). Such a scheme works only for wormholes which possess sufficiently strong magnetic fields, since high-energy baryons cannot be captured by the wormhole. The mean energy of baryons is determined by the temperature which depends on the redshift. The intensity of the magnetic field also depends on the redshift. While baryons are relativistic particles the threshold value of the fictitious (or equivalent) current does not depend on time.

Indeed, only particles below the threshold energy are captured by the wormhole magnetic field[10] which is given by

$$E = 3kT < E_{th} = eBR_w,$$

where e is the electron charge and R_w is the biggest radius of the doughnut - shaped throat of the wormhole. The energy of relativistic baryons behaves with the redshift as $T = T_\gamma(1+z)$, where T_γ is the present day temperature of CMB radiation. The intensity of the magnetic field can be estimated by the value in the hole of the doughnut. We take it as

$$B = \frac{\kappa I}{cR_w},$$

where $\kappa = 2\pi$ in the center of the doughnut hole, while close to the throat (surface of the doughnut), where the field reaches the maximum value, in the approximation $r_w/R_w \ll 1$ we get the estimate $\kappa \simeq 2R_w/r_w$. The field depends on the parameter I

[c]We present here estimates for baryons only, since leptons are much lighter than baryons and in the primordial plasma leptons simply follow baryons (bounded by the Coulomb potential).

(which is the fictitious or equivalent current) and the big radius R_w of the doughnut which also depends on the redshift as $R_w = R_0/(1+z)$.

The constant I behaves with the redshift as $I = I_0(1+z)$. Indeed the invariant characteristics is the number of magnetic lines captured by the throat (which go through the internal hole of the doughnut). This gives $\int_S \mathbf{Bds} = const = \Phi$, where S an arbitrary surface whose boundary contour γ (dual to Γ) lays on the surface of the throat and cannot be contracted to a point, so that all magnetic lines intersect S only once. Taking the minimal surface S we find $\Phi \sim \frac{2\pi^2}{c} R_w I$ which gives the behavior $I \sim 1/R_w \sim 1/a$, where $a = a_0/(1+z)$ is the scale factor of the Universe. We point out that the same dependence on the redshift z follows from the fact that the energy density of the magnetic field $\rho_B \sim B^2/4\pi$ decreases with the scale factor as $\rho_B \sim 1/a^4$. This gives the threshold value for the equivalent current which defines the intensity of the magnetic field as

$$I_0 > I_{th} = \frac{ce}{\kappa r_\gamma} \sim \frac{3.2}{\kappa} \times 10^{-5} A, \tag{19}$$

where $r_\gamma = \frac{e^2}{3kT_\gamma}$. It is convenient to express r_γ as follows $r_\gamma = r_p(1+z_r)$, where $r_p = \frac{e^2}{m_p c^2}$ is the classical radius of the proton and $1+z_r = \frac{m_p c^2}{3kT_\gamma} \sim 10^{12}$ is the redshift at which baryons become relativistic particles. It is curious that the threshold value I_{th} is extremely small. It does not depend on the absolute size of the wormhole (which is given by the big radius R_w) but only on the ratio of the wormhole radii $\kappa = 2R_w/r_w$. All wormholes with the present day values $I_0 > I_{th}$ strongly interact with baryons. We may say that they are frozen into baryons and, therefore, peculiar motions of baryons repeat peculiar motions of such wormholes. Wormholes with smaller magnetic fields $I_0 < I_{th}$ slightly interact with baryons and can be considered as free objects. They may participate in independent from baryons motions. As we shall see on the early stage of the evolution of the Universe, before the recombination, they cannot capture baryons and therefore do not form an enhancement in the baryon density. On latter stages upon reheating they may capture charged particles and form cosmic rays and compact sources of synchrotron radiation. The relativistic stage for baryons finishes at the redshift z_r upon which (e.g., for $z < z_r$) baryons become non-relativistic and the above consideration brakes.

At redshifts $z_r > z > z_{rec}$ baryons are non-relativistic, while upon the recombination z_{rec} baryons form neutral Hydrogen atoms and do not interact with magnetic fields of all wormholes. They again start to interact with wormholes only at the epoch of re-ionization when first stars have fired.

On the stage $z_r > z > z_{rec}$ for every particular wormhole it is possible to define the critical redshift z_0 when it captures baryons (for $z > z_0$). The critical value z_0 can be estimated as follows. The mean energy of baryons is $m_p V^2/2 = 3kT/2$ with $T = T_\gamma(1+z)$. Then the critical redshift can be found from the inequality $r_B = \frac{V}{\omega_B} = \frac{V m_p c}{eB} < r_d$, where r_B and ω_B are Larmor radius and frequency respectively for protons, while r_d denotes the width of the baryon cloud around the wormhole

throat. The width r_d should be smaller, than the proton diffusion length and it defines the critical value of the magnetic field. This gives

$$B = \frac{\kappa I}{cR_w} > \frac{m_p c}{e r_d}\sqrt{\frac{3kT}{m_p}},$$

or equivalently

$$I_0(1+z) > \frac{\alpha e}{\kappa r_p}\sqrt{\frac{3kT}{m_p}} = \alpha I_{th}(1+z_r)\sqrt{\frac{(1+z)}{(1+z_r)}},$$

where the ratio $\alpha = \frac{R_w}{r_d} \gg 1$. This defines the critical redshift z_0 at which the wormhole starts to trap baryons as

$$(1+z) > (1+z_0) = \alpha^2 \frac{I_{th}^2}{I_0^2}(1+z_r).$$

The relation between z_0 and I_0 can be rewritten as

$$I_0^2 = \frac{(1+z_r)}{(1+z_0)}\alpha^2 I_{th}^2. \tag{20}$$

When $z_0 = z_r$ we get $I_0 = \alpha I_{th}$ and if we take $\alpha = 1$, this will give the absolute threshold of the field. We point out that the value $\alpha = 1$ may have sense only for sufficiently small wormholes with the radius $R_w \lesssim 10^{-5} R_{gal}$, e.g., see estimates in the next section. At redshifts $z > z_r$ we get into the epoch where baryons are relativistic particles and wormholes with $I_0 < I_{th}$ do not bound baryons at all. There is one more critical value I_{rec} which corresponds to $z_0 = z_{rec}$.

$$I_{rec}^2 = \alpha^2 I_{th}^2 \frac{(1+z_r)}{(1+z_{rec})} \gg I_{th}^2. \tag{21}$$

All fields with $I_0 > I_{rec}$ are strong enough to capture baryons during the whole evolution $z > z_{rec}$.

5. The number of baryons in traps

The efficiency of wormhole traps can be described by the number of baryons collected. The number of baryons collected around magnetic wormholes depends on the proton diffusion length $\ell(z)$, the big radius a wormhole throat $R_w(z)$, and the radius of the baryon cloud $r_{cl}(z) \ll R_w$. For estimates one may use $r_{cl} \sim r_w$. This number can be estimated as the increase of the effective volume of the torus-shaped throat

$$\Delta N = < n_b > \left(V(R_w + \ell, r_{cl} + \ell) - V(R_w, r_{cl})\right),$$

where $< n_b >$ is the mean density of baryons and $V = 2\pi^2 R_w r_{cl}^2$ is the throat volume. We define the parameter $\delta_b = \Delta N/(V < n_b >)$ which depends on the position in space. Close to wormhole throats $\delta_b > 0$, while sufficiently far from the wormhole $\delta_b < 0$ since baryons from those regions have captured by the wormhole.

The value $< \delta_b^2 >= b$, where brackets define the averaging over the space, relates to the baryon clumping factor $b = (< n_b^2 > - < n_b >^2)/ < n_b >^2$, e.g.,[25,26]

In the case $\ell(z) \ll r_{cl}(z)$ we get

$$\delta_b(z) \sim \frac{\ell(z)}{R_w(z)} + \frac{2\ell(z)}{r_{cl}(z)} \ll 1. \tag{22}$$

In the intermediate case $r_{cl}(z) \ll \ell(z) \ll R_w(z)$ we find the estimate

$$\delta_b(z) \sim \frac{\ell^2(z)}{r_{cl}^2(z)} \left(1 + \frac{\ell(z)}{R_w(z)}\right) \gg 1. \tag{23}$$

And in the case $\ell(z) \gg R_w(z)$ the estimate has the order

$$\delta_b(z) \sim \frac{\ell^3(z)}{R_w(z) r_{cl}^2(z)} \gg 1. \tag{24}$$

On the stage $z > z_r$ protons are relativistic particles, plasma is degenerate, and the length of the proton propagation can be estimated by the value of the horizon size $\ell(z) \sim l_h = c/H(z)$. At the redshift $z = z_r$ it is extremely small and has the order $\ell(z_r) \sim (7 \div 8) \times 10^{-15} pc$. Consider a wormhole throat with the big radius $R_w(0) \sim 15 kpc$ which corresponds to a galaxy size,[41] while the radius of the baryon cloud has the order $r_{cl} \sim 0.2 R_w$. Then we find $R_w(z_r) \sim 15 \times 10^{-9} pc \gg \ell(z_r)$ and from (22) we get $\delta_b(z_r) \sim (0.5 \div 0.6) \times 10^{-5}$. Such a value is too small to form a galaxy without additional means (e.g., a dark matter clump).

At the recombination $z_{rec} = 1100$ the proton diffusion length has the co-moving value of the order $\ell(z_{rec}) \sim 0.4 \div 1 pc$, while $R_w(z_{rec}) \sim 13.6 pc$ and we still may consider $\ell(z_{rec}) \ll R_w$ and $\ell \lesssim r_{cl}$. Therefore we again may use (22) and find $\delta_b(z) \sim 0.32 \div 0.8$. Such a big value shows that the respective protogalaxy forms immediately after the recombination and it is already in the non-linear regime. We should expect that wormholes with such strong clumps of baryons depart the Hubble expansion very soon and form rather small objects. Smaller wormholes form too strong inhomogeneities before recombination and probably collapse to blackholes. We point out that this may give a new mechanism of blackhole formation with huge masses.

Galaxies are observed up to $z \simeq 11$ (e.g., the most distant GN-z11 is observed at $z = 11.09$[39]). Therefore, to be consistent with the present day size of a typical ring galaxy the wormhole radius should be at least two orders bigger $R_w(0) \lesssim 1 Mpc$, which gives already $R_w(z_{rec}) \lesssim 900 pc$ and $\delta_b(z_{rec}) \sim 4,8 \times 10^{-3}$. Such a clump departs the Hubble expansion already at $z \sim 100$ and gives a ring of the order $R \sim 10 Kpc$.

6. Conclusions

In conclusion we point out two important facts. First one is that the wormholes whose size $R(z_{rec})$ exceeds the value $\ell(z_{rec})$ more than on the factor 10^3 do not form a sufficient enhancement in the baryon number density and, therefore, cannot

form ring galaxies. They however may form ring-type structures in the future, e.g. see the recently reported findings of unexpected class of astronomical ring-type objects in.[29]

The second fact is that upon the recombination $z < z_{rec}$ the doughnut-shaped wormholes do not interact with baryons and evolve. Therefore, they may leave the ring clump formed. They either expand or collapse forming a magnetized black hole in the middle. If the wormhole collapses, it should also draw some portion of the baryon clump and form a bulge in the center of the ring. In this case it may form the ideal symmetric structure similar to the Hoag's object. Additional rotational perturbations of clumps may however lead to irregular structures. If the wormhole expands further, the center part of the ring remains to be empty. We may expect that in such a case the mean value of the magnetic field within the ring retains. We recall that different active processes in galaxies generate magnetic fields which in general have a turbulent character. Therefore, by the measuring the mean flow of the magnetic field in a ring galaxy one may at least estimate the present day value of the big radius $R = \frac{\kappa I}{cB}$ of such a wormhole and estimate its present day position. This may be used in the direct search for wormhole traces in the Universe.

At first glance the most direct indication on the possible role of relic magnetic wormholes in the formation of some ring galaxies should be the discrepancy between the observed amount of dark matter in such galaxies and the predictions of the standard theory.[40,42,43] However there exist some extensions of general relativity which are capable of reproducing dark matter effects in galaxies without dark matter particles, e.g., see[44,45] and references therein. Therefore, we think that the only rigorous indication on the presence of relic magnetic wormholes should be large-scale toroidal magnetic fields.

References

1. Hawking, S.W. *Nucl. Phys. B* **114**, 349 (1978).
2. Savelova, E.P. *Grav. Cosmol.* **21**, 48–56 (2015).
3. Knizhnik V., Polyakov A., Zamolodchikov A. *Mod. Phys. Lett. A,* **3**, p. 819 (1988).
4. Kawai H., Kawamoto N., Mogami T., Watabiki Y. Phys. Lett. B, **306**, pp. 19–26 (1993).
5. Ambjorn, J. Jurkiewicz, and R. Loll, *Phys. Rev. Lett.* **95** 17 (2005).
6. Ambjorn J. et al, *Eur. Phys. J. C* **81**:708 (2021).
7. Amendola, et al. *Living Rev. Relativ.* **21** 2 (2018).
8. Harko, T.; Lobo, F. *Extensions of f (R) Gravity: Curvature-Matter Couplings and Hybrid Metric-Palatini Theory*, Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK, (2018).
9. Kirillov A.A. & Savelova E.P., *Int. J. Mod. Phys. D* **25**, 1650075 (2016).
10. Kirillov A.A. & Savelova E.P., *Eur. Phys. J. C* **80** 45 (2020).
11. Bronnikov K.A., Krechet V.G., and Lemos J.P.S., *Phys. Rev. D* **87** 084060 (2013).
12. Bronnikov K.A.& Krechet V.G., *Phys. Rev. D* **99** 084051 (2019).
13. Shaikh, R. et al., *Phys. Lett. B* **789** 270 (2019).
14. Shaikh, R., Banerjee, P., Paul, S., and Sarkar, T., *JCAP* 1907, 028 (2019).

15. Safonova, M., Torres, D. F., and Romero, G. E., *Phys. Rev. D* **65** 023001 (2002).
16. Kirillov A.A. & Savelova E.P., *Universe* **4** 35 (2018).
17. Kirillov A.A. & Savelova E.P., *Astr. Sp. Sc.* **364** 1 (2019).
18. Kirillov A.A., Savelova E.P., Lecian O.M., *Eur. Phys. J. C* **81**: 263 (2021).
19. Kirillov A.A., Savelova E.P., *Int J Mod Phys A,* **35** 2040026 (2020).
20. DAMPE Collaboration, *Nature* **552** 24475 (2017).
21. Fermi, E.: *Phys. Rev.* **75**, 1169 (1949).
22. Neronov, A.; Vovk, I., Science **328** 73(2010).
23. Takahashi, K.; Mori, M.; Ichiki, K.; Inoue, S.; Takami, H., *Astrophys. J. Lett.* **771** L42 (2013).
24. Subramanian K., *Rep. Prog. Phys.* **79** 076901 (2016).
25. Jedamzik, K. & Abel, T., *JCAP*, 1310, 050 (2013).
26. Jedamzik, K. & Saveliev, A., *Phys. Rev. Lett.* **123** 021301 (2019).
27. Jedamzik, K. & Pogosian, L., Relieving the Hubble tension with primordial magnetic fields, arXive: 2004.09487 (2020).
28. Kirillov A.A. & Savelova E.P., *Eur. Phys. J. C* **80** 810 (2020).
29. Norris, R. P. et al., Unexpected Circular Radio Objects at High Galactic Latitude, *Publications of the Astronomical Society of Australia,* **38** E003 (2021).
30. Eguchi, T., Gilkey, P.B., Hanson, A., *Phys. Rep.* **66**(6) 213-393 (1980).
31. Arellano A. V. B., Lobo, F.S.N., *Class. Quant. Grav.* **23** 5811 (2006).
32. Bronnikov K. A., *Int. J. Mod. Phys. D* **27** 1841005 (2018).
33. Subramanian K., *Rep. Prog. Phys.* **79**, 076901 (2016).
34. Bronnikov K. A., *Acta Physica Polonica A* **4** 251 (1973).
35. Ellis H. G., *J. Math. Phys.* **14** 104 (1973).
36. Jones, T. J. et al. *The Astrophysical Journal Letters* **870** L9 (2019).
37. Beck, R., *Astron Astrophys Rev.* **24**: 4 (2016).
38. Moss D., Mikhailov E., Sokoloff D., Silchenko O., Horellou C., Beck R. *Astron. Astrophys.* **592** A44 (2016).
39. Oesch, P. A. et al. *Astrophys. J.* **819** 129 (2016).
40. White, S. & Rees, M. *Mon. Not. R. Astron. Soc.* **183** 341 (1978).
41. Yuan, T. et al. *Nature Astronomy Letters.* (2020). https://doi.org/10.1038/ s41550-020-1102-7.
42. Steinmetz, M. & Navarro, J. F. *New Astronomy* **7** 155 (2002).
43. Persic M., Salucci P., Stel F., *Mon. Not. Roy. Astron. Soc.* **281** 27 (1996).
44. Kirillov, A. A. and D. Turaev, *Mon. Not. Roy. Astron. Soc.* **371** L31 (2006).
45. Kirillov A. A., *Phys. Lett. B,* **632** 453 (2006).

Wormhole geometries induced by action-dependent Lagrangian theories

Ismael Ayuso*, Francisco S. N. Lobo† and José P. Mimoso‡

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, Lisbon, P-1749-016, Portugal

*E-mail: iayuso@fc.ul.pt

†E-mail: fslobo@fc.ul.pt

‡E-mail: jpmimoso@fc.ul.pt

In this work, we explore wormhole geometries in a recently proposed modified gravity theory arising from a non-conservative gravitational theory, tentatively denoted action-dependent Lagrangian theories. The generalized gravitational field equation essentially depends on a background four-vector λ^μ, that plays the role of a coupling parameter associated with the dependence of the gravitational Lagrangian upon the action, and may generically depend on the spacetime coordinates. Considering wormhole configurations, by using "Buchdahl coordinates", we find that the four-vector is given by $\lambda_\mu = (0, 0, \lambda_\theta, 0)$, and that the spacetime geometry is severely restricted by the condition $g_{tt}g_{uu} = -1$, where u is the radial coordinate. We find a plethora of specific asymptotically flat, symmetric and asymmetric, solutions with power law choices for the function λ, by generalizing the Ellis-Bronnikov solutions and the recently proposed black bounce geometries, amongst others. We show that these compact objects possess a far richer geometrical structure than their general relativistic counterparts.

Keywords: General relativity, alternative theories of gravity; wormhole; action-dependent Lagrangian theories; energy conditions.

1. Introduction

A key ingredient in traversable wormhole geometries is the flaring-out condition,[1,2] which in General Relativity (GR) entails the violation of the null energy condition (NEC). The latter is defined as $T_{\mu\nu}k^\mu k^\nu \geq 0$, for *any* null vector k^μ,[3,4] and matter violating the NEC has been denoted as *exotic matter*. However, it has been shown that these violations may be minimized using several procedures, such as the cut-and-paste techniques in the thin-shell formalism, where the exotic matter is concentrated at the junction interface.[5–14] In fact, the problem is improved with evolving traversable wormholes, where it has been demonstrated that these time-dependent geometries may satisfy the energy conditions in arbitrary finite intervals of time,[15,16] and recently specific dynamical four-dimensional solutions were presented that satisfy the null and weak energy conditions everywhere and everywhen.[17,18] In fact, modified theories of gravity is an interesting avenue of research to explore traversable wormholes, where these compact objects possess a richer geometrical structure than their general relativistic counterparts. In this context, it has been shown that the NEC can be satisfied for normal matter threading the

wormhole throat, where it is the higher order curvature terms that sustain the wormhole.[19–27]

In this work, we will be interested in studying wormhole geometries in a recently proposed modified gravity theory arising from a non-conservative gravitational theory, tentatively denoted action-dependent Lagrangian theories.[28] The latter are obtained through an action principle for action-dependent Lagrangians by generalizing the Herglotz variational problem[29,30] for several independent variables. The novel feature when comparing with previous implementations of dissipative effects in gravity is the possible arising of such phenomena from a least action principle, so they are of a purely geometric nature. Applications to this model have also been explored, namely, in cosmology,[31] braneworld gravity,[32] cosmic string configurations,[33] the late-time cosmic accelerated expansion and large scale structure,[34] and static spherically symmetric stellar solutions,[35] amongst others.

The complete set of field equations considered in this action-dependent Lagrangian theory[28] is based on the following total Lagrangian

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_m = (R - \lambda_\mu s^\mu) + \mathcal{L}_m, \tag{1}$$

where the Einstein-Hilbert Lagrangian is extended with the geometrical sector dealing with the additional dissipative term $\lambda_\mu s^\mu$ while $\mathcal{L}_m$ is the Lagrangian of the matter fields. In general, the background four-vector λ^μ depends on the spacetime coordinates, however it can be assumed to be constant. The field s^μ is an action-density field which disappears after the variation of the action such that the modification to the GR counterpart is given by the four-vector λ^μ only. Note that λ^μ may be considered a background four-vector, that plays the role of a coupling parameter associated with the dependence of the gravitational Lagrangian upon the action. In the majority of the works considered above it is assumed to be constant, however, in a more general scenario, one may assume it to be a coordinate-dependent four-vector.

Thus, the field equations are given by

$$G_{\mu\nu} + Z_{\mu\nu} = \kappa^2 T_{\mu\nu}, \tag{2}$$

where we have defined $\kappa^2 = 8\pi$, $G_{\mu\nu}$ is the Einstein tensor, and for notational simplicity, we have defined $Z_{\mu\nu}$ as

$$Z_{\mu\nu} = K_{\mu\nu} - \frac{1}{2} g_{\mu\nu} K. \tag{3}$$

The symmetric geometric structure $K_{\mu\nu}$ is defined as

$$K_{\mu\nu} = \lambda_\alpha \Gamma^\alpha_{\mu\nu} - \frac{1}{2}(\lambda_\mu \Gamma^\alpha_{\nu\alpha} + \lambda_\nu \Gamma^\alpha_{\mu\alpha}), \tag{4}$$

which is constructed from the particular combination of the four-vector λ_μ and the Christofell symbols

$$\Gamma^\alpha_{\mu\nu} = \frac{g^{\alpha\beta}}{2}\left(g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}\right). \tag{5}$$

The quantity $K_{\mu\nu}$ (and its trace K) represents the geometric structure behind the dissipative nature of the theory. Note that the limit of a vanishing λ_μ restores the dissipationless feature of GR.

Thus, motivated by the existence of static spherically-symmetric compact objects analysed in,[35] we extend this analysis to the context of wormhole physics. This work is outlined in the following manner: In Sec. 2, we present the most general restrictions on static and spherically symmetric wormhole geometries imposed by the geometrical structure of the action-dependent Lagrangian theory. In Sec. 3, we consider a plethora of specific solutions of action-dependent Lagrangian induced wormhole geometries. Finally, in Sec. 4, we summarize our results and conclude.

2. General restrictions on wormhole geometries

Consider the general static and spherically symmetric metric given by

$$ds^2 = -f(u)dt^2 + g(u)du^2 + R^2(u)d\Omega^2, \tag{6}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ is the linear element of the unit sphere, and the metric functions $f(u)$, $g(u)$ and $R(u)$ are functions of the radial coordinate u. The coordinate choices used in metric (6) are often called "Buchdahl coordinates".[36–38] Note that one possesses a freedom in choosing the radial coordinate, consequently allowing one to fix the form of one of the metric functions $f(u)$, $g(u)$ or $R(u)$, which will be considered below. Here the radial coordinate lies in the range $u \in (-\infty, +\infty)$, so that two asymptotically flat regions exist, i.e., $u \to \pm\infty$, and are connected by the throat. The function $R(u)$ possesses a global positive minimum at the wormhole throat $u = u_0$, which one can set at $u_0 = 0$, without a loss of generality. Thus, the wormhole throat is defined as $R_0 = \min\{R(u)\} = R(0)$.

In order to avoid event horizons and singularities throughout the spacetime, one imposes that the metric functions $f(u)$ and $g(u)$ are positive and regular everywhere. Taking into account these restrictions, namely, the necessary conditions for the minimum of the function imposes the flaring-out conditions, which are given by

$$R_0' = 0\,, \qquad R_0'' > 0\,. \tag{7}$$

In this work, we consider an anisotropic distribution of matter threading the wormhole described by the following stress-energy tensor $T_{\mu\nu}$

$$T_{\mu\nu} = (\rho + p_t)U_\mu U_\nu + p_t\, g_{\mu\nu} + (p_r - p_t)\chi_\mu\chi_\nu\,, \tag{8}$$

where U^μ is the four-velocity, χ^μ is the unit spacelike vector in the radial direction, i.e., $\chi^\mu = g^{-1/2}(u)\delta^\mu_u$; $\rho(u)$ is the energy density, $p_r(u)$ is the radial pressure measured in the direction of χ^μ, and $p_t(u)$ is the transverse pressure measured in the orthogonal direction to χ^μ.

Now, rather than write out the full gravitational field equations (2) for the metric (6), we note that the only non-zero components of the Einstein and the stress-energy tensor are the diagonal terms, so that the non-diagonal part of the additional

tensor $Z_{\mu\nu}$, defined by Eq. (3), also provides additional information on the geometrical structure of the solutions of the theory, namely, that $Z_{\mu\nu} = 0$ for $\mu \neq \nu$. More specifically, the non-diagonal components of the symmetric tensor $Z_{\mu\nu}$ places restrictions on the form of the four-vector λ_μ.

The independent components of the tensor $Z^\mu_\nu = K^\mu_\nu - \frac{1}{2}\delta^\mu_\nu K$ are given by

$$Z^u_t = -\frac{\lambda_t[fg'R - g(f'R - 4fR')]}{4fg^2R}, \quad Z^u_u = \frac{\lambda_\theta}{R^2}\cot\theta, \tag{9}$$

$$Z^u_\theta = -\frac{\lambda_\theta(fg)'}{4fg^2} - \frac{\lambda_u}{2g}\cot\theta, \qquad Z^u_\phi = -\frac{\lambda_\phi(fg)'}{4fg^2}, \tag{10}$$

$$Z^\theta_\theta = Z^\phi_\phi = \frac{\lambda_u(Rf' + 2fR')}{2fgR}, \tag{11}$$

$$Z^\theta_\phi = \frac{\lambda_\phi}{2R^2}\cot\theta, \qquad Z^\theta_t = -\frac{\lambda_t}{2R^2}\cot\theta, \tag{12}$$

$$Z^t_t = \frac{2\lambda_u R'}{gR} + \frac{\lambda_\theta}{R^2}\cot\theta, \qquad Z^t_\phi = 0\,. \tag{13}$$

From $Z^\theta_\phi = Z^\theta_t = 0$, one readily extracts the restrictions $\lambda_\phi = \lambda_t = 0$. Taking into account the assumption of the static and spherical symmetric character of the spacetime, the field equations should only depend on the radial coordinate, so that from the diagonal Z^u_u component, one readily verifies that $\lambda_\theta \propto (\cot\theta)^{-1}$, or more specifically $\lambda_\theta = \lambda(u)/(\cot\theta)$ (one may consider the simple case $\lambda(u) = \lambda_0 = \text{const}$). Note that if one were to consider $\lambda_\theta = 0$, then the condition $Z^u_\theta = 0$ would impose that $\lambda_u = 0$, taking us trivially back to GR. Analogously, in order for Z^t_t to only depend on the radial coordinate, from $Z^u_\theta = 0$ this imposes that $\lambda_u = 0$ and consequently places a further constraint on the metric functions, namely, $(fg)' = 0$.

Thus, the additional information on the geometrical structure of the theory, which imposes that the non-diagonal components of the symmetric tensor $Z_{\mu\nu}$ vanish, imposes the following condition on the four-vector λ_μ:

$$\lambda_\mu = \left(0, 0, \frac{\lambda(u)}{\cot\theta}, 0\right), \tag{14}$$

and the additional geometric tensor Z^μ_ν takes the diagonal form $Z^\mu_\nu = (\lambda(u)/R^2)\,\mathrm{diag}(1,1,0,0)$. Furthermore, from the constraint on the metric functions $(fg)' = 0$, we can consider, without a loss of generality, the following choice:

$$g(u) = f^{-1}(u) = A(u). \tag{15}$$

3. Specific solutions of action-dependent Lagrangian induced wormhole geometries

The analysis outlined in the previous section imposes that the static and spherical symmetric configuration (6) in the theory (1), can be written as

$$ds^2 = -A(u)dt^2 + A^{-1}(u)du^2 + R^2(u)d\Omega^2, \tag{16}$$

where as before the wormhole throat is defined as $R_0 = \min\{R(u)\} = R(0)$, and in order to avoid event horizons and singularities throughout the spacetime, one imposes that the function $A(u)$ is positive and regular everywhere. These restrictions imposes the flaring-out conditions, translated by Eq. (7).

As the metric function $A(u)$ is positive and regular for $\forall u$, it is useful to analyse its derivatives at the throat $u = 0$. In particular, the sign of A_0'' determines the type of extrema of $A(u)$, i.e., it is a minimum if $A_0'' > 0$ and a maximum if $A_0'' < 0$. This implies that the maximum (minimum) of $A(u)$ corresponds to a maximum (minimum) of the gravitational potential, so that in the vicinity of a maximum (minimum) the gravitational force is repulsive (attractive). Thus, the wormhole throat possesses a repulsive or an attractive nature that depends on the sign of A_0''.

Now, taking into account the modified Einstein equation (2), the spacetime metric (16) and the stress-energy tensor (8), the gravitational field equations are finally given by:

$$8\pi\rho = -\frac{2ARR'' + AR'^2 + A'RR' - 1}{R^2} - \frac{\lambda(u)}{R^2}\,, \tag{17}$$

$$8\pi p_r = \frac{AR'^2 + A'RR' - 1}{R^2} + \frac{\lambda(u)}{R^2}\,, \tag{18}$$

$$8\pi p_t = \frac{A''R + 2AR'' + 2A'R'}{2R}\,. \tag{19}$$

Adding Eqs. (17) and (18), yields the following relation

$$R''|_{R_0} = -\frac{4\pi R}{A}(\rho + p_r)|_{R_0}\,, \tag{20}$$

and using the condition at the throat $R_0'' > 0$, one verifies that in these specific action-dependent Lagrangian theories the NEC is generically violated at the throat, i.e., $(\rho + p_r)|_{R_0} < 0$.

Taking into account the field equations (17)–(19), one has three independent equations with six unknown functions of the radial coordinate u, namely, $\rho(u)$, $p_r(u)$, $p_t(u)$, $A(u)$, $R(u)$ and $\lambda(u)$. There are several strategies that one may now follow. More specifically, one may consider specific choices for the components of the stress-energy tensor, and then solve the field equations to determine the metric functions and $\lambda(u)$; one may also take into account a plausible stress-energy tensor profile by imposing equations of state $p_r = p_r(\rho)$ and $p_t = p_t(\rho)$, and close the system by adequately choosing the energy density, or any of the metric functions. In alternative to this approach, one may use the reverse philosophy usually adopted in wormhole physics by simple choosing specific choices for the metric functions and $\lambda(u)$, and through the field equations determine the stress-energy profile responsible for sustaining the wormhole geometry. In the following section, we will adopt several strategies outlined above, and a mixture thereof, to obtain specific exact solutions of wormhole spacetimes induced by these action-dependent Lagrangian theories.

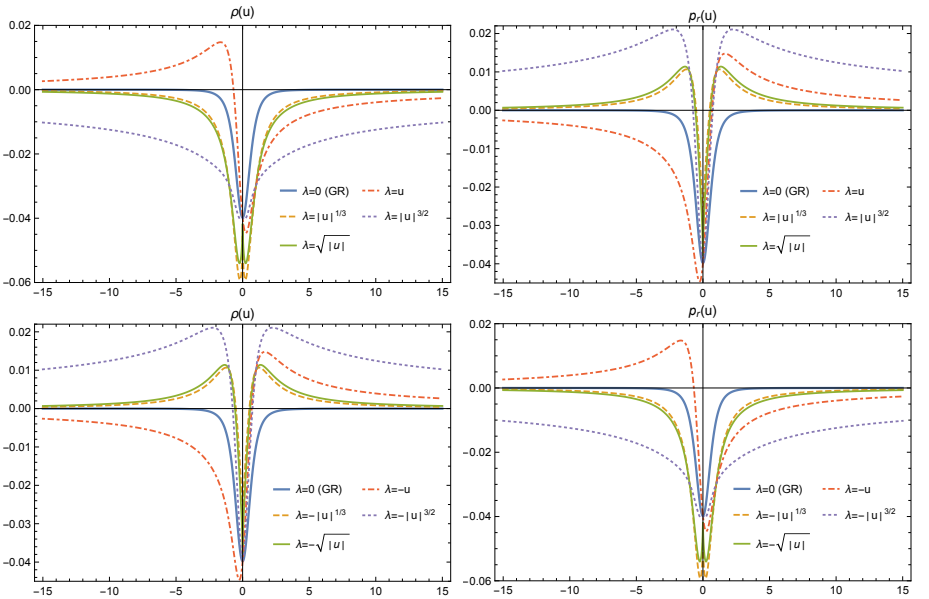


Fig. 1. The plots depict the specific case of the Ellis-Bronnikov wormhole configuration with $a = 1$, for convenience, and for different choices of the function $\lambda(u)$. Depending on the sign of $\lambda(u)$, one obtains a plethora of specific symmetric or asymmetric solutions. Note that for $\lambda(u) \sim \pm u$, one obtains asymmetric solutions, where the energy density is negative at the throat and in the positive (negative) branch of u, but becomes positive in the negative (positive) branch; the radial pressure exhibits the inverse qualitative behavior. We refer the reader to the text for more details.

3.1. *Specific wormhole solutions: Ellis-Bronnikov solution*

Using an appropriate parametrization, we can present a solution by taking into account the reverse philosophy of solving the modified field equations, as follows[39]

$$R(u) = e^{-\alpha(u)}\sqrt{u^2 + a^2}, \qquad A(u) = e^{2\alpha(u)}, \tag{21}$$

with the factor $\alpha(u)$ defined as

$$\alpha(u) = \left(\frac{m}{a}\right) \arctan\left(\frac{u}{a}\right), \tag{22}$$

where m and a are two free parameters. Thus, the spacetime metric is given by

$$ds^2 = -e^{2\alpha}dt^2 + e^{-2\alpha}\left[du^2 + (u^2 + a^2)d\Omega^2\right]. \tag{23}$$

Following the previous definition of the wormhole throat, which is situated at $u_0 = 0$, we readily obtain $R'(0) = -m/a$, so that the condition $R'(0) = 0$ imposes $m = 0$. Note that these conditions imply that the solution reduces to the well-known Ellis-Bronnikov wormhole spacetime.[40–42] This does indeed simplify the analysis below, such that $\alpha(u) = 0$, $A(u) = 1$, and

$$R''(u_0) = \frac{1}{|a|} > 0 \tag{24}$$

In addition to this, Eqs. (17)–(19) yield the following stress-energy profile:

$$\rho(u) = -\frac{\left(a^2+u^2\right)\lambda(u)+a^2}{8\pi\left(a^2+u^2\right)^2}, \tag{25}$$

$$p_r(u) = \frac{\left(a^2+u^2\right)\lambda(u)-a^2}{8\pi\left(a^2+u^2\right)^2}, \tag{26}$$

$$p_t(u) = \frac{a^2}{8\pi\left(a^2+u^2\right)^2}\,. \tag{27}$$

For the specific case of $\lambda = 0$, where the four-vector λ^μ vanishes, this solution simply reduces to the general relativistic Ellis-Bronnikov stress-energy components. However, for the general case, one still needs to impose one more condition to close the system, and in the following we consider specific choices for the function $\lambda(u)$. Equations (25)–(26) yield the following relation:

$$\rho(u) + p_r(u) = -\frac{a^2}{4\pi\left(a^2+u^2\right)^2}\,, \tag{28}$$

which states that the NEC is violated throughout the entire spacetime, and is independent of the function $\lambda(u)$.

We are only interested in asymptotically flat solutions, so that taking into account the limit of Eq. (25), one finds

$$\lim_{u\to\infty}\rho(u) \sim -\lim_{u\to\infty}\frac{\lambda(u)}{u^2}\,. \tag{29}$$

For instance, assuming a power law solution for $\lambda(u) \sim u^\alpha$, the asymptotic flatness condition imposes that $\alpha < 2$, and from the regularity of the stress-energy components we have $\alpha \geq 0$, so that the parameter lies in the range $0 \leq \alpha < 2$. One may perform a similar analysis with the radial pressure $p_r(u)$, but with a change in the sign for the limit. Note that the tangential pressure, $p_t(u)$, is independent of $\lambda(u)$, possesses a maximum value at the throat, $p_t(u=0) = (8\pi a^2)^{-1}$, and tends to zero with increasing u.

Several choices for the function $\lambda(u)$ are depicted in Fig. 1. Depending on the sign of $\lambda(u)$, one obtains a plethora of specific symmetric or asymmetric solutions. More specifically, for the case of $\lambda(u) \sim \pm u$, one obtains asymmetric solutions where the energy density is negative at the throat and in the positive (negative) branch of u, but becomes positive in the negative (positive) branch, while the radial pressure possesses the inverse qualitative behavior, as is transparent from Fig. 1. Thus, it is possible to alleviate the negative energy densities needed to thread this wormhole configurations, relative to GR. The wormhole solutions obtained with $\lambda \neq 0$ possess a richer structure that their general relativistic counterparts.

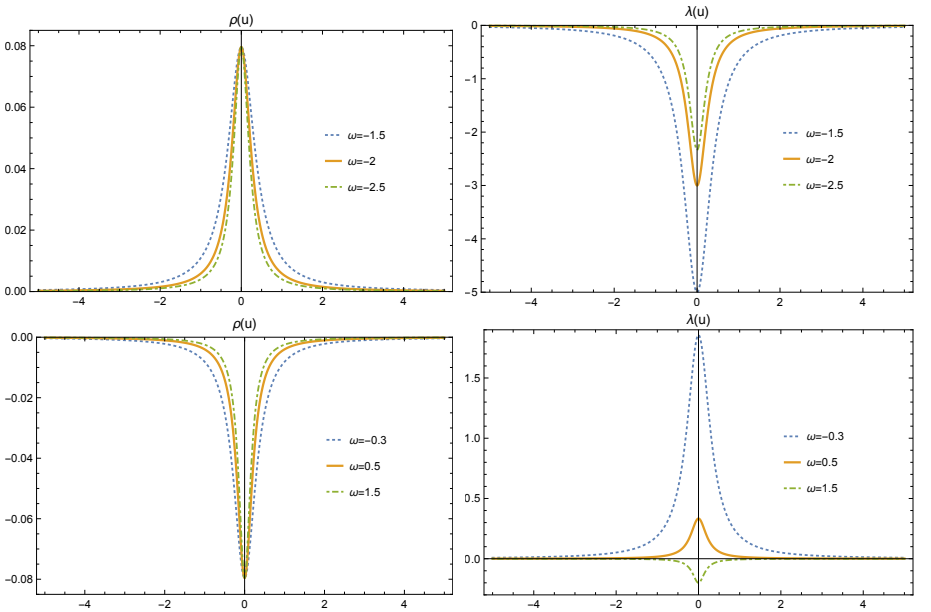


Fig. 2. Results for the specific stress-energy profile given by (30) and (31) with $a^2 = -1/(4\pi\rho_0(1+\omega))$ and $\alpha = 2$. The upper plots are for the case $\rho_0 = 1$ ($\omega < -1$), and the lower plots for $\rho_0 = -1$ ($\omega > -1$).

3.2. *Specific stress-energy profile*

We now consider the strategy of specifying the profile of the energy density and radial pressure given by:

$$\rho(u) = \rho_0 \left(\frac{a^2}{a^2+u^2}\right)^{\alpha}, \tag{30}$$

$$p_r(u) = p_0 \left(\frac{a^2}{a^2+u^2}\right)^{\alpha}, \tag{31}$$

with $\alpha > 0$, so that both components tend to zero at spatial infinity. In addition to this, we close the system by considering the specific choice for the metric function

$$R(u) = \sqrt{a^2+u^2}\,. \tag{32}$$

Note that Eqs. (30) and (31) can be written as $p_r = \omega\rho$, with $\omega = p_0/\rho_0$. Thus, this case is formally equivalent to choosing Eq. (32), one of Eqs. (30) or (31), and the equation of state $p_r(u) = \omega\rho(u)$.

The gravitational field equations (17)–(19) provide the following solutions:

$$A(u) = -4\pi a^2\rho_0(\omega+1)\left(\frac{a^2}{a^2+u^2}\right)^{\alpha-2}, \tag{33}$$

$$\lambda(u) = 1 + 4\pi\rho_0\left[2a^2\omega - (2\alpha-5)u^2(\omega+1)\right]\left(\frac{a^2}{a^2+u^2}\right)^{\alpha-1}. \tag{34}$$

As before, we impose the asymptotic flatness condition, namely,

$$\lim_{u\to\pm\infty} A(u) \to 1\,, \tag{35}$$

and taking into account that $A(u)$ should be positive and regular $\forall u$, implies the following two stringent restrictions:

$$\alpha = 2, \qquad \text{and} \qquad 4\pi a^2\rho_0(1+\omega) = -1, \tag{36}$$

where the second condition imposes:

$$\rho_0(1+\omega) < 0. \tag{37}$$

This implies two specific cases: (i) $\rho_0 > 0$ and $\omega < -1$, so that taking into account the equation of state $\omega = p_0/\rho_0$, implies a negative radial pressure at the throat; or (ii) $\rho_0 < 0$ and $\omega > -1$, so that $p_0 > 0$ for $-1 < \omega < 0$, and $p_0 < 0$ for $\omega > 0$. Specific cases are depicted in Fig. 2. Relative to the analysis at the throat, note that the wormhole conditions are satisfied, namely, $R'(u_0) = 0$ and $R''(0) = 1/|a| > 0$. In addition to this, for the imposition of the asymptotic flatness condition, namely, $\alpha = 2$, we readily obtain $A(u) = 1$, so that $A'(u_0) = A''(u_0) = 0$.

3.3. *Black bounce solutions*

Recently, a number of novel regular "black-bounce" spacetimes were explored.[43,44] These are specific geometries where the "area radius" always remains non-zero, thereby leading to a "throat" that is either timelike (corresponding to a traversable wormhole), spacelike (corresponding to a "bounce" into a future universe), or null (corresponding to a "one-way wormhole"). The regularity, the energy conditions, and the causal structure of these models were analysed in detail in Refs.[43,44] The main results are several new geometries with two or more horizons, with the possibility of an extremal case. Motivated by these novel solutions, in this subsection we shall analyse specific generalized "black-bounce" wormhole geometries induced by action-dependent Lagrangian theories.

3.3.1. *Simpson–Visser black-bounce spacetime*

In this section, we consider a specific black bounce geometry, which we denote as the Simpson-Visser solution.[13,43,44] Consider the following parameters, which were presented in Ref.[43]:

$$R(u) = \sqrt{u^2+a^2}, \qquad A(u) = 1 - \frac{2m}{\sqrt{u^2+a^2}}\,. \tag{38}$$

Note that the Schwarzschild solution is recovered, if we take the limit $a \to 0$.

This spacetime possesses several interesting properties.[43] First, for $a > 0$ the geometry is everywhere regular, which can verified as $R(u)$ is never zero, and is regular, as is $A(u)$. Now, one has several cases, for instance: (i) if $0 < a < 2m$, two horizons exist, namely, $u_\pm = \pm\sqrt{(2m)^2 - a^2}$, where u_+ is positive and u_- is

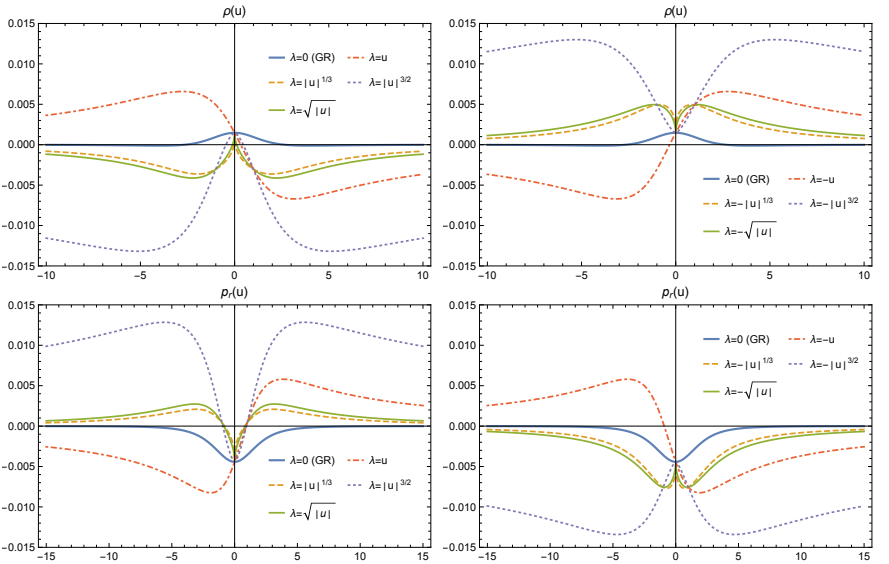


Fig. 3. The plots depict the Simpson-Visser black bounce solution, for the specific choices $m = 1$, $a = 3$. Note that one may obtain wormhole configurations with an entirely positive energy density throughout the spacetime, for negative values of the function $\lambda(u)$; and positive radial pressures in the negative branch of the u-axis. As before, it is transparent from the plots that these compact objects possess a richer geometrical structure than their general relativistic counterparts. See the text for more details.

negative. This solution corresponds to a regular black hole spacetime, where the core consists of a bounce located at $u = 0$; (ii) if $a = 2m$, a wormhole exists with a throat located at $u = 0$. This is an extremal null throat, which can only be crossed from one region to another, so that the wormhole is only one-way traversable; (iii) finally, if $a > 2m$, a two-way traversable wormhole exists, that possesses a timelike throat located at $u = 0$. Thus, only the case (c) for $a > 2m$ interests us here. We refer the reader to Refs.[43,44] for specific details.

Taking into account the choices for the metric functions (38), the field equations (17)–(19) provide the following stress-energy profile:

$$\rho(u)_{SV} = -\frac{1}{8\pi}\left[\frac{a^2\left(\sqrt{a^2+u^2}-4m\right)}{\left(a^2+u^2\right)^{5/2}} + \frac{\lambda(u)}{a^2+u^2}\right], \tag{39}$$

$$p_r(u)_{SV} = \frac{\left(a^2+u^2\right)\lambda(u) - a^2}{8\pi\left(a^2+u^2\right)^2}, \tag{40}$$

$$p_t(u)_{SV} = \frac{a^2\left(\sqrt{a^2+u^2}-m\right)}{8\pi\left(a^2+u^2\right)^{5/2}}, \tag{41}$$

respectively. The asymptotic limits of the energy density and the radial pressure

are given by:

$$\lim_{u\to\pm\infty}\rho(u)_{SV}\sim-\lim_{u\to\pm\infty}p_r(u)_{SV}\sim-\lim_{u\to\pm\infty}\frac{\lambda(u)}{a^2+u^2}. \tag{42}$$

As before, if we assume a power law for $\lambda(u)\sim u^{\alpha}$, the asymptotic flatness condition and the regularity of the stress-energy components, as before, imposes that $0\leq\alpha<2$; note that the tangential pressure $p_t(u)\to 0$ for $u\to\pm\infty$, and possesses a maximum at the wormhole throat, i.e., $p_t(u=0)=(a-m)/(8\pi a^3)$, and is positive throughout the spacetime as we are only considering the condition $a>2m$.

Several choices for the function are depicted in Fig. 3. Depending on the sign of $\lambda(u)$, one obtains a plethora of specific symmetric or asymmetric solutions. Note that these compact objects possess a richer geometrical structure than their general relativistic counterparts. It is transparent from Fig. 3 that one may obtain wormhole configurations with an entirely positive energy density throughout the spacetime, for negative values of the function $\lambda(u)$; for instance, for the latter it is also possible to obtain positive radial pressures in the negative branch of the u-axis. However, the NEC is always violated at the wormhole throat.

3.3.2. *Black bounce II*

Another black bounce spacetime that exhibits interesting properties, is given by the following specific metric functions:[44]

$$R(u)=\sqrt{u^2+a^2},\qquad A(u)=1-\frac{2mu^2}{(u^2+a^2)^{3/2}}\,. \tag{43}$$

Note that by solving for the roots of the function $A(u)=0$, we have: i) for $a<a_{\rm ext}=4m/(3\sqrt{3})$, there are four real solutions, which are symmetrical to each other, namely, $(r_+,r_C,-r_C,-r_+)$, (where u_+ corresponds to the event horizon and u_C to a Cauchy horizon); ii) for $a=a_{\rm ext}$, we have two real solutions $(u_+,-u_+)$; iii) and for $a>a_{\rm ext}$, no real value exists. We refer the reader to[44] for more details. Thus, in order to have a traversable wormhole solution, where only the case of $A(u)>0$ is satisfied, only the specific case for $a>a_{\rm ext}$ interests us here.

For this case, the gravitational field equations (17)–(19) yield the stress-energy profile, given by the following relations:

$$\rho(u)=-\frac{1}{8\pi}\left[\frac{a^2\left(a^2+u^2\right)^{3/2}-8ma^2u^2}{\left(a^2+u^2\right)^{7/2}}+\frac{\lambda(u)}{a^2+u^2}\right], \tag{44}$$

$$p_r(u)=\frac{1}{8\pi}\left[-\frac{a^2\left(a^2+u^2\right)^{3/2}+4ma^2u^2}{\left(a^2+u^2\right)^{7/2}}+\frac{\lambda(u)}{a^2+u^2}\right], \tag{45}$$

$$p_t(u)=\frac{a^2u^2\left(\sqrt{a^2+u^2}+5m\right)+a^4\left(\sqrt{a^2+u^2}-2m\right)}{8\pi\left(a^2+u^2\right)^{7/2}}, \tag{46}$$

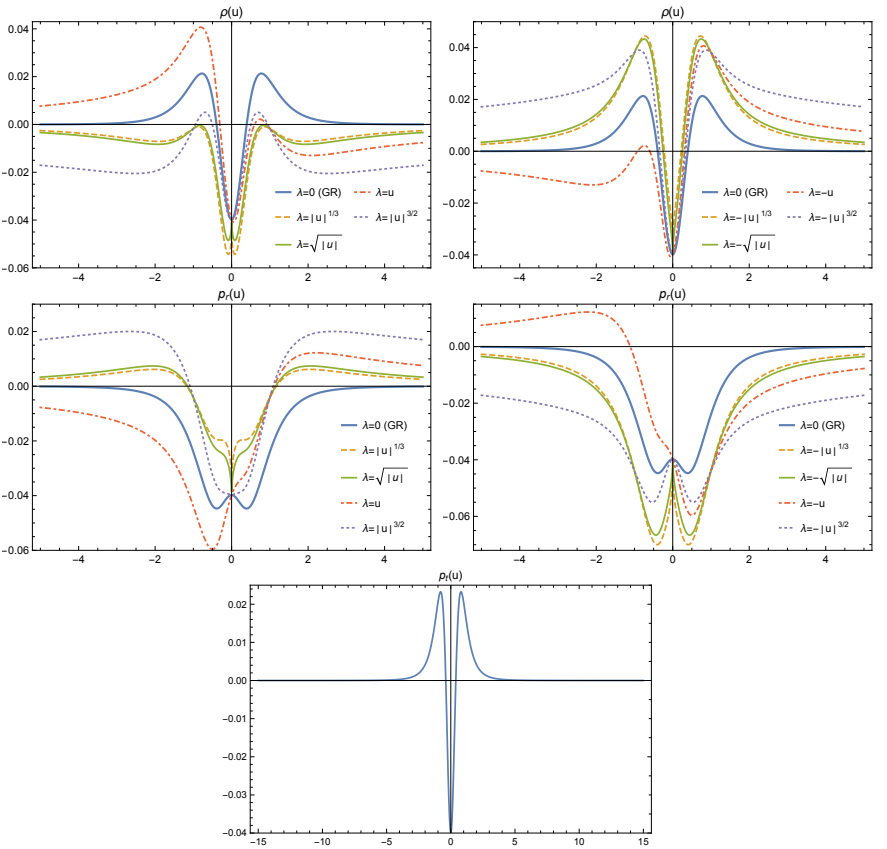


Fig. 4. The plots depict the stress-energy profile for the black bounce II solution, given by the metric function (43), where for numerical convenience we have assumed the following choices for the parameters: $m = 1, a = 1$. Recall that the parameters are restricted by the condition $a > 4m/(3\sqrt{3})$. As before these wormhole configurations possess a far richer internal structure than their general relativistic counterparts, and depending on the sign of $\lambda(u)$, specific symmetric or asymmetric solutions are obtained. However, here the tangential pressure at the wormhole throat, $p_t(u=0) = (a-2m)/(8\pi a^3)$, takes negative values for the parameter range $4m/(3\sqrt{3}) < a < 2m$, and positive values for $a > 2m$. We refer the reader to the text for more details.

respectively, which are depicted in Fig. 4 for specific choices of the model parameters.

Assume, once again, a power law for $\lambda(u) \sim u^{\alpha}$, we have that $0 \leq \alpha < 2$ by the asymptotic flatness condition and the regularity of the stress-energy components, as before. As before these wormhole geometries induced by action-dependent Lagrangian theories possesses a far richer internal structure than their general relativistic counterparts, and depending on the sign of $\lambda(u)$, specific symmetric or asymmetric solutions are obtained. We refer the reader to Fig. 4 for a qualitative behaviour of the stress-energy profile; recall that taking into account the parameter range, we consider the condition $a > 4m/(3\sqrt{3})$. Here the tangential pressure at

the wormhole throat is given by $p_t(u=0) = (a-2m)/(8\pi a^3)$, and takes negative values for $4m/(3\sqrt{3}) < a < 2m$, possessing a minimum at the throat, and positive values for $a > 2m$; note that $p_t(u) \to 0$ for $u \to \pm\infty$.

3.3.3. *Black bounce III*

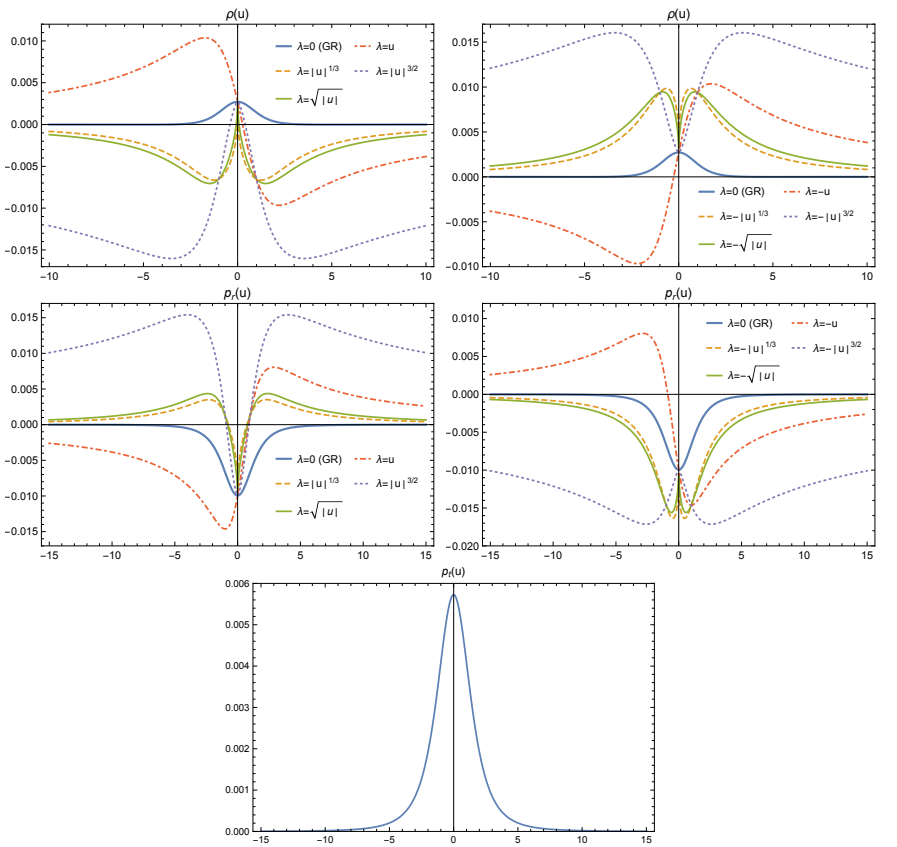


Fig. 5. The plots depict the stress-energy profile for the black bounce III solution, given by the functions (47) and (49), for the case $n = 1$, where $a > a_{\text{ext}}$, so that there are no event horizons. We have chosen the following values for the parameters: $m = 1$ and $a = 2$. Note that for these solutions, the negative energy densities are improved, and one may also obtain positive radial pressures for positive values of the function $\lambda(u)$. See the text for more details.

Finally, we consider another black bounce solution explored in,[44] that also exhibits interesting properties, given by

$$R(u) = \sqrt{u^2 + a^2}\,. \tag{47}$$

and the mass function:

$$M(u) = m\left(\frac{R(u)}{u}\right)\left(\frac{2}{\pi}\right)^n \arctan^n\left(\frac{u}{a}\right), \tag{48}$$

so that the metric function $A(u)$ is given by

$$A(u) = 1 - \frac{2M(u)}{R(u)} = 1 - \frac{2m}{u}\left(\frac{2}{\pi}\right)^n \arctan^n\left(\frac{u}{a}\right). \tag{49}$$

In the limit $(a, n) \to 0$ we regain the Schwarzschild solution. However, one can fix n and regulate the presence of horizons by adjusting a. For instance, consider $n = 1$, where the extreme case is given by $a_{\rm ext} = 4m/\pi$.[44]

The causal structure, for the specific case of $n = 1$, is given by: (i) for $a > a_{\rm ext}$, this corresponds to the traditional two-way traversable wormhole; (ii) for $a = a_{\rm ext}$, we have a one-way wormhole geometry with an extremal null throat; (iii) for $a < a_{\rm ext}$, we have one horizon located in each universe, where one may propagate through this event horizon, located at $u = u_+$ in order to reach the spacelike "bounce" hypersurface at $u = 0$, before "bouncing" into a future version of our own universe. We refer the interested reader to Ref.[44] for more details. Thus, we are only interested in the case $a > a_{\rm ext}$, for $n = 1$, where there are no event horizons.

Taking into account the metric functions (47) and (49), the gravitational field equations (17)–(19) provide the following stress-energy components:

$$\rho(u) = -\frac{1}{8\pi}\left\{\frac{a^2}{\left(a^2+u^2\right)^2}\left[1 - \frac{2m}{ua}\left(\frac{2}{\pi}\right)^n\left(\arctan\left(\frac{u}{a}\right)\right)^{n-1}\left(a\arctan\left(\frac{u}{a}\right) + nu\right)\right] + \frac{\lambda(u)}{a^2+u^2}\right\}, \tag{50}$$

$$p_r(u) = \frac{1}{8\pi}\left\{\frac{2ma}{u\left(a^2+u^2\right)^2}\left(\frac{2}{\pi}\right)^n\left(\arctan\left(\frac{u}{a}\right)\right)^{n-1}\left[a\arctan\left(\frac{u}{a}\right) - nu\right] + \frac{\left(a^2+u^2\right)\lambda(u) - a^2}{(a^2+u^2)^2}\right\}, \tag{51}$$

$$p_t(u) = \frac{a^2}{8\pi(a^2+u^2)^2}\left\{1 - \frac{m}{au^3}\left(\frac{2}{\pi}\right)^n\left(\arctan\left(\frac{u}{a}\right)\right)^{n-2} \times\left[2\left(a^3+2au^2\right)\left(\arctan\left(\frac{u}{a}\right)\right)^2 - 2nu\left(a^2+u^2\right)\arctan\left(\frac{u}{a}\right) + a(n-1)nu^2\right]\right\}. \tag{52}$$

respectively. Here, we will only consider, for simplicity, the specific case of $n = 1$ and $a > a_{\rm ext}$.

If we consider, as before, a power law for $\lambda(u) \sim u^\alpha$, the asymptotic flatness and regularity conditions impose $0 \leq \alpha < 2$. Once again, one obtains a wide variety of solutions, both symmetric and asymmetric, which are depicted in Fig. 5. The

advantage of these solutions consists essentially in that they ameliorate the negative energy densities for negative values of the function $\lambda(u)$. However, positive values of the function $\lambda(u)$ allow positive radial pressures as is transparent in Fig. 5. The tangential pressure tends to zero at spatial infinity, i.e., $p_t(u) \to 0$ for $u \to \pm\infty$, and possesses a maximum at the throat, as depicted in Fig. 5.

4. Conclusions

In this work, we have explored wormhole geometries in the recently proposed action-dependent Lagrangian theories,[28] that are obtained through an action principle for action-dependent Lagrangians by generalizing the Herglotz variational problem for several independent variables. An interesting feature of these theories as compared with previous implementations of dissipative effects in gravity is the possible arising of such phenomena from a least action principle, so they are of a purely geometric nature. It was shown that the generalized gravitational field equation essentially depends on a background four-vector λ^μ, that plays the role of a coupling parameter associated with the dependence of the gravitational Lagrangian upon the action, and may generically depend on the spacetime coordinates. In the context of wormhole configurations, we have used the "Buchdahl coordinates", and found that the four-vector is given generically by $\lambda_\mu = (0, 0, \lambda_\theta(u, \theta), 0)$. In addition to this restriction the spacetime geometry is also severely constrained by the condition $g_{tt}g_{uu} = -1$, where u is the radial coordinate.

More specifically, we have shown that the field equations (17)–(19), impose a system of three independent equations with six unknown functions of the radial coordinate u, namely, $\rho(u)$, $p_r(u)$, $p_t(u)$, $A(u)$, $R(u)$ and $\lambda(u)$. Thus, one possesses several strategies to solve the system of equations. For instance, one may consider a plausible stress-energy tensor profile by imposing equations of state $p_r = p_r(\rho)$ and $p_t = p_t(\rho)$, and close the system by adequately choosing the energy density, or a specific metric function. However, one may also adopt the reverse philosophy approach usually used in wormhole physics by simple choosing specific choices for the metric functions and $\lambda(u)$, and through the field equations determine the stress-energy profile responsible for sustaining the wormhole geometry. Here, we have found a plethora of specific asymptotically flat, symmetric and asymmetric, solutions with power law choices for the function λ, for instance, by generalizing the Ellis-Bronnikov solutions and the recently proposed black bounce geometries, amongst other solutions. We have shown that these compact objects possess a far richer geometrical structure than their general relativistic counterparts. It would be interesting to investigate time-dependent spacetimes as outlined in[17,18] in order to explore the energy conditions, and work along these lines is presently underway.

Acknowledgments

IA is funded by the Fundação para a Ciência e a Tecnologia (FCT, Portugal) grant No. PD/BD/114435/2016 under the IDPASC PhD Program. FSNL acknowledges

support from the FCT Scientific Employment Stimulus contract with reference CEECIND/04057/2017. The authors also acknowledge funding from FCT Projects No. UID/FIS/04434/2020, No. CERN/FIS-PAR/0037/2019 and No. PTDC/FIS-OUT/29048/2017.

References

1. M. S. Morris and K. S. Thorne, "Wormholes in space-time and their use for interstellar travel: A tool for teaching general relativity," Am. J. Phys. **56**, 395 (1988).
2. M. S. Morris, K. S. Thorne and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition," Phys. Rev. Lett. **61**, 1446 (1988).
3. M. Visser, "Lorentzian wormholes: From Einstein to Hawking," Woodbury, USA: AIP (1995).
4. F. S. N. Lobo, "Wormholes, Warp Drives and Energy Conditions", Fundam. Theor. Phys. **189**, pp. (2017), (formerly Lecture Notes in Physics), Springer Nature Switzerland AG.
5. M. Visser, "Traversable wormholes from surgically modified Schwarzschild spacetimes", Nucl. Phys. B **328** (1989) 203 [arXiv:0809.0927 [gr-qc]].
6. M. Visser, "Traversable wormholes: Some simple examples", Phys. Rev. D **39** (1989) 3182 [arXiv:0809.0907 [gr-qc]].
7. E. Poisson and M. Visser, "Thin shell wormholes: Linearization stability", Phys. Rev. D **52**, 7318 (1995) [arXiv:gr-qc/9506083].
8. F. S. N. Lobo and P. Crawford, "Linearized stability analysis of thin shell wormholes with a cosmological constant," Class. Quant. Grav. **21** (2004), 391-404 [arXiv:gr-qc/0311002 [gr-qc]].
9. F. S. N. Lobo, "Energy conditions, traversable wormholes and dust shells," Gen. Rel. Grav. **37** (2005), 2023-2038 [arXiv:gr-qc/0410087 [gr-qc]].
10. F. S. N. Lobo and P. Crawford, "Stability analysis of dynamic thin shells," Class. Quant. Grav. **22** (2005), 4869-4886 [arXiv:gr-qc/0507063 [gr-qc]].
11. N. Montelongo-García, F. S. N. Lobo and M. Visser, "Generic spherically symmetric dynamic thin-shell traversable wormholes in standard general relativity", Phys. Rev. D **86** (2012) 044026 [arXiv:1112.2057 [gr-qc]].
12. M. Bouhmadi-López, F. S. N. Lobo and P. Martín-Moruno, "Wormholes minimally violating the null energy condition," JCAP **11** (2014), 007 [arXiv:1407.7758 [gr-qc]].
13. F. S. N. Lobo, A. Simpson and M. Visser, "Dynamic thin-shell black-bounce traversable wormholes," Phys. Rev. D **101** (2020) no.12, 124035 [arXiv:2003.09419 [gr-qc]].
14. T. Berry, F. S. N. Lobo, A. Simpson and M. Visser, "Thin-shell traversable wormhole crafted from a regular black hole with asymptotically Minkowski core," Phys. Rev. D **102** (2020) no.6, 064054 [arXiv:2008.07046 [gr-qc]].
15. S. Kar, "Evolving wormholes and the weak energy condition," Phys. Rev. D **49**, 862-865 (1994).
16. S. Kar and D. Sahdev, "Evolving Lorentzian wormholes," Phys. Rev. D **53**, 722 (1996) [arXiv:gr-qc/9506094 [gr-qc]].
17. M. Kord Zangeneh, F. S. N. Lobo and H. Moradpour, "Evolving traversable wormholes satisfying the energy conditions in the presence of pole dark energy," [arXiv:2008.04013 [gr-qc]].
18. M. Kord Zangeneh and F. S. N. Lobo, "Dynamic wormhole geometries in hybrid metric-Palatini gravity," [arXiv:2011.01745 [gr-qc]].

19. F. S. N. Lobo, "General class of wormhole geometries in conformal Weyl gravity", Class. Quant. Grav. **25**, 175006 (2008) [arXiv:0801.4401 [gr-qc]].
20. F. S. N. Lobo, M. A. Oliveira, "Wormhole geometries in $f(R)$ modified theories of gravity", Phys. Rev. **D80**, 104012 (2009) [arXiv:0909.5539 [gr-qc]].
21. N. Montelongo-García, F. S. N. Lobo, "Wormhole geometries supported by a nonminimal curvature-matter coupling", Phys. Rev. **D82**, 104018 (2010) [arXiv:1007.3040 [gr-qc]].
22. N. Montelongo-García, F. S. N. Lobo, "Nonminimal curvature-matter coupled wormholes with matter satisfying the null energy condition", Class. Quant. Grav. **28**, 085018 (2011) [arXiv:1012.2443 [gr-qc]].
23. C. G. Boehmer, T. Harko and F. S. N. Lobo, "Wormhole geometries in modified teleparralel gravity and the energy conditions", Phys. Rev. D **85** (2012) 044033 [arXiv:1110.5756 [gr-qc]].
24. T. Harko, F. S. N. Lobo, M. K. Mak and S. V. Sushkov, "Modified-gravity wormholes without exotic matter", Phys. Rev. D **87** (2013) no.6, 067504 [arXiv:1301.6878 [gr-qc]].
25. M. R. Mehdizadeh, M. Kord Zangeneh and F. S. N. Lobo, "Einstein-Gauss-Bonnet traversable wormholes satisfying the weak energy condition", Phys. Rev. D **91**, no. 8, 084004 (2015) [arXiv:1501.04773 [gr-qc]].
26. M. Kord Zangeneh, F. S. N. Lobo and M. H. Dehghani, "Traversable wormholes satisfying the weak energy condition in third-order Lovelock gravity", Phys. Rev. D **92**, no. 12, 124049 (2015) [arXiv:1510.07089 [gr-qc]].
27. J. L. Rosa, J. P. S. Lemos and F. S. N. Lobo, "Wormholes in generalized hybrid metric-Palatini gravity obeying the matter null energy condition everywhere," Phys. Rev. D **98**, no. 6, 064054 (2018) [arXiv:1808.08975 [gr-qc]].
28. M. J. Lazo, J. Paiva, J. T. S. Amaral and G. S. F. Frederico, "Action principle for action-dependent Lagrangians toward nonconservative gravity: Accelerating universe without dark energy," Phys. Rev. D **95** (2017) no. 10, 101501 [arXiv:1705.04604 [gr-qc]].
29. G. Herglotz, Berührungstransformationen, Lectures at the University of Götingen, Göttingen, (1930).
30. R. B. Guenther, C. M. Guenther and J. A. Gottsch, The Herglotz Lectures on Contact Transformations and Hamiltonian Systems, Lecture Notes in Nonlinear Analysis, Vol. 1, Juliusz Schauder Center for Nonlinear Studies, Nicholas Copernicus University, Torún, (1996).
31. J. C. Fabris, H. Velten, T. R. P. Caramês, M. J. Lazo and G. S. F. Frederico, "Cosmology from a new nonconservative gravity," Int. J. Mod. Phys. D **27** (2018) no.06, 1841006 [arXiv:1711.06206 [gr-qc]].
32. J. C. Fabris, T. R. P. Caramês and J. M. Hoff Da Silva, "Braneworld gravity within non-conservative gravitational theory," Eur. Phys. J. C **78** (2018) no.5, 402 [arXiv:1804.01072 [gr-qc]].
33. E. A. F. Bragança, T. R. P. Caramês, J. C. Fabris and A. d. Santos, "Some effects of non-conservative gravity on cosmic string configurations," Eur. Phys. J. C **79** (2019) no.2, 162 [arXiv:1808.01653 [gr-qc]].
34. T. R. P. Caramês, H. Velten, J. C. Fabris and M. J. Lazo, "Dark energy with zero pressure: Accelerated expansion and large scale structure in action-dependent Lagrangian theories," Phys. Rev. D **98** (2018) no. 10, 103501 [arXiv:1808.02798 [gr-qc]].
35. J. C. Fabris, H. Velten and A. Wojnar, "Existence of static spherically-symmetric objects in action-dependent Lagrangian theories," Phys. Rev. D **99** (2019) no.12, 124031 [arXiv:1903.12193 [gr-qc]].
36. M. R. Finch and J. E. F. Skea, "A review of the relativistic static fluid sphere", 1998, unpublished.

37. P. Boonserm and M. Visser, "Buchdahl-like transformations for perfect fluid spheres", Int. J. Mod. Phys. D **17** (2008), 135–163 [arXiv:0707.0146 [gr-qc]].
38. P. Boonserm and M. Visser, "Buchdahl-like transformations in general relativity", Thai Journal of Mathematics **5 # 2** (2007) 209–223.
39. S. V. Sushkov and R. Korolev, "Scalar wormholes with nonminimal derivative coupling," Class. Quant. Grav. **29** (2012), 085008 [arXiv:1111.3415 [gr-qc]].
40. H. G. Ellis, "Ether flow through a drainhole: A particle model in general relativity," *J. Math. Phys.* **14**, 104 (1973).
41. H. G. Ellis, "The evolving, flowless drain hole: a nongravitating particle model in general relativity theory," *Gen. Rel. Grav.* **10**, 105-123 (1979).
42. K. A. Bronnikov, "Scalar-tensor theory and scalar charge," *Acta Phys. Pol. B* **4**, 251 (1973).
43. A. Simpson and M. Visser, "Black-bounce to traversable wormhole," JCAP **02** (2019), 042 [arXiv:1812.07114 [gr-qc]].
44. F. S. N. Lobo, M. E. Rodrigues, M. V. d. S. Silva, A. Simpson and M. Visser, "Novel black-bounce geometries," [arXiv:2009.12057 [gr-qc]].

Gravitational lensing by wormholes in binary systems

S. Pietroni

Dipartimento di Fisica "E.R. Caianiello", Università di Salerno, Via Giovanni Paolo II 132, I-84084 Fisciano, Italy
Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Via Cintia, 80126, Napoli, Italy
E-mail: spietroni@unisa.it

We investigate binary lenses with $1/r^n$ potentials in the asymmetric case with two lenses with different indexes n and m. These kinds of potentials have been widely used in several contexts, ranging from galaxies with halos described by different power laws to lensing by wormholes or exotic matter.

Keywords: Gravitational lensing; black holes; wormholes; galaxies

1. Introduction

Space-time is curved by the presence of massive bodies and this curvature influences the motion of the bodies themselves: this leads to a geometry in constant evolution. One of the consequences is that even light, supposed to be massless, bends its trajectory while passing close to a massive body.[1]

Gravitational lensing is an important tool in astrophysics and in cosmology widely used to study both populations of compact objects (including exoplanets, black holes and other stellar remnants) , and extended objects, such as galaxies, clusters of galaxies and large-scale structures . Since most of the mysteries of our Universe do not show up in observations based on electromagnetic interactions, gravitational lensing is more and more employed to study the dark side of the Universe, including dark matter, dark energy, and any kind of exotic matter (such as wormholes) conjectured by theorists.[2]

2. Lensing by Wormholes

The metric of the Ellis wormhole falls down asymptotically as $1/r^2$ and its deflection angle goes as the inverse square of the impact parameter $1/u^2$.

Metrics falling as $1/r^n$ were investigated also by Kitamura et al.[3] who found out that the deflection angle falls down with the same exponent as the metric, $\hat{\alpha} \sim 1/u^n$ with $n > 1$, and that a demagnification of the total lensed images could appear at $\beta > \frac{2}{n+1}$ (in units of θ_E, the Einstein radius, and under a large-n approximation, β is the source position). This demagnification effect may be evidence of an Ellis wormhole and it might be used for hunting the search for exotic matter. Particular attention was posed on the study of caustics of $1/r^n$ binary lenses by Bozza and Melchiorre in Ref.[4]

The implications for the energy-momentum tensor supporting this kind of metrics falling as $1/r^n$ were investigated by Bozza and Postiglione,[5] we remind that $0 < n < 1$ may describe galactic halos, $n > 1$ would be the signature of a violation of the weak energy condition and so the existence of exotic matter, $n = 2$ correspond to the Ellis wormhole; n is the ratio between tangential and radial pressure, $n = -2p_t/p_r$.

3. The Lens Equation, Critical Curves and Caustics

In a binary system composed by two lenses (binary lenses, A and B), the lens equation can be written as follows

$$\vec{\beta} = \vec{\theta} - \frac{\vec{\theta} - \vec{\theta_A}}{|\vec{\theta} - \vec{\theta_A}|^{n+1}} - \gamma^{m+1} \frac{\vec{\theta} - \vec{\theta_B}}{|\vec{\theta} - \vec{\theta_B}|^{m+1}}, \tag{1}$$

where $\gamma = \theta_{E,B}/\theta_{E,A}$ is the "strength ratio".

The lens equation allows us to find images, given the source position and a lens configuration.

The number of images formed at a given source β depends on the source position and these regions with a different number of images are delimited by caustics.

The condition $J(z) = 0$ (the Jacobian determinant of the lens map) defines the critical curves on the lens plane; and by applying the lens map on critical points we find the corresponding points on the source plane, which form the caustics. When a source crosses a caustic, a new pair of images is created on the corresponding point in the critical curve.

4. Cases and Topology Regimes

We investigated 3 cases: the equal-strength binary with $\gamma = 1$, the unequal-strength binary with $\gamma = \sqrt{0.1}$ (the bigger lens is the standard one with fixed $n = 1$), and the reversed unequal-strength binary with $\gamma = \sqrt{0.1}$ (the standard lens is the smaller one).

For the standard binary Schwarzschild lens in the equal-strength case, we know that three topologies exist:

- close separation, for $s < s_{CI}$;
- intermediate separation, for $s_{CI} < s < s_{IW}$;
- wide separation, for $s > s_{IW}$;

and the two transitions are $s_{CI} = 1$ and $s_{IW} = 2\sqrt{2}$ in our units.

The three topologies exist for any value of m and n.

Our model contains 4 parameters: the indexes of the two potentials n, m, the separation between the two lenses s, and the ratio of the two Einstein radii γ.

5. Critical Curves and Caustics: Wide, Intermediate and Close Separations

Here we show the equal-strength case with $\gamma = 1$, the full research can be found in: *Bozza V., Pietroni S., Melchiorre C., Universe 2020, 6(8), 106*.[6]

The red curve is for the standard Schwarzschild case $n = m = 1$, we keep $n = 1$ for the first lens and we see what happens when m varies in the second lens

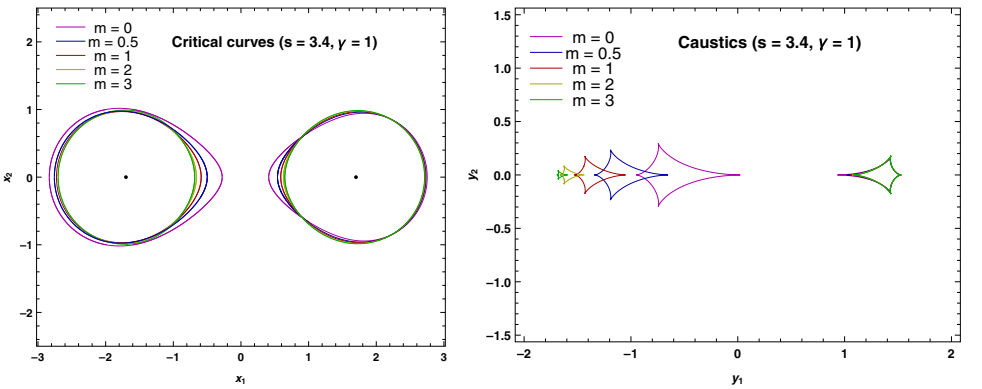


Fig. 1. Critical curves and caustics in the equal-strength binary, wide separation with fixed $n = 1$ and variable m: $m = 0$ is the singular isothermal sphere already investigated by Shin and Evans, $m = 0.5$ is the galactic halo, $m = 2$ is the Ellis wormhole and $m = 3$ is for exotic matter.

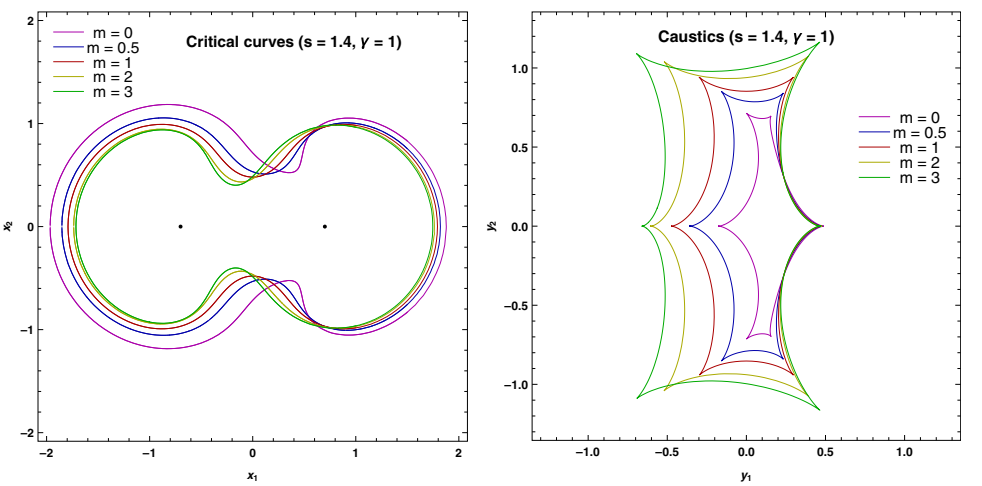


Fig. 2. Critical curves and caustics in the equal-strength binary, intermediate separation with fixed $n = 1$ and variable m. The red curve is for the standard Schwarzschild case $n = m = 1$.

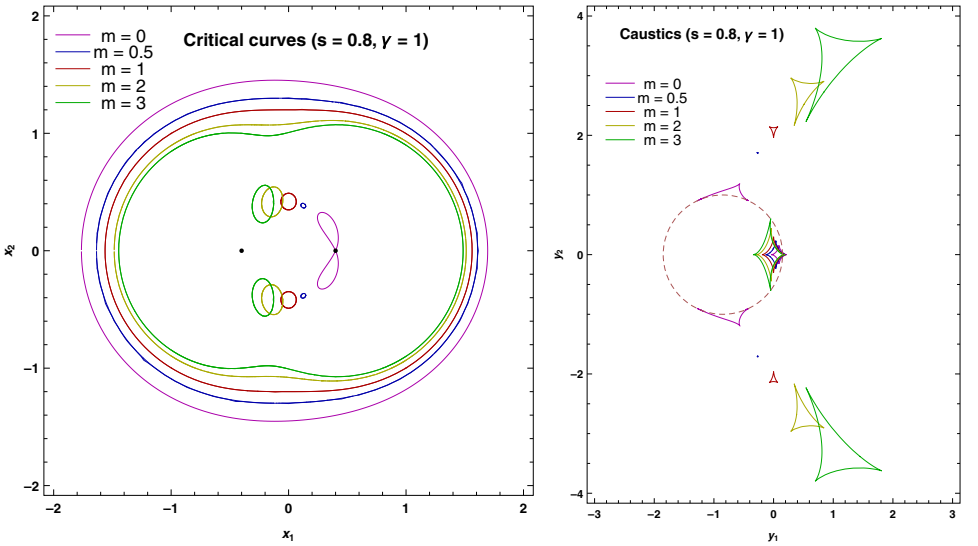


Fig. 3. Critical curves and caustics in the equal-strength binary, close separation with fixed $n = 1$ and variable m. The red curve is for the standard Schwarzschild case $n = m = 1$. Dashed magenta circle indicates the pseudocaustic for $m = 0$.

6. The Elliptic Umbilic and the Extremely Unequal-Strength Ratio Limit

In the range $0 \leq m < 1$ an *elliptic umbilic catastrophe* exists in the close separation. The value of s at which the catastrophe happens is

$$s_{euc} = \left(\frac{1-mn}{m+1}\right)^{\frac{1}{n+1}} \sqrt{1 + \frac{\gamma^2 (m+1)^{\frac{2}{n+1}}}{(n+1)^{\frac{2}{m+1}} (1-mn)^{\frac{2(m-n)}{(m+1)(n+1)}}}}. \tag{2}$$

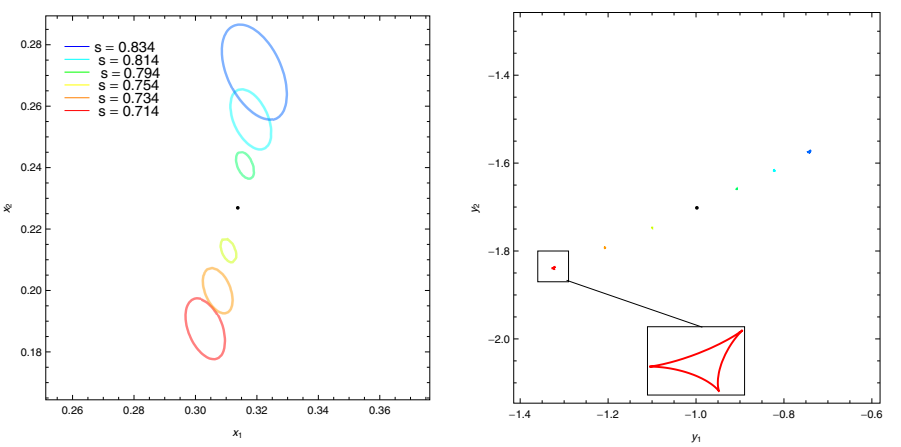


Fig. 4. The elliptic umbilic catastrophe for $n = 2$, $m = 0.25$. The separation at which the catastrophe occurs is $s_{euc} = 0.774$.

For any n, m, γ we found the boundaries for s_{CI}, given only numerically, and the analytical expression for s_{IW}.

The caustic evolution in the extreme limit $\theta_{E,B} << \theta_{E,A}$, in the case of two Schwarzschild objects ($n = m = 1$), is the so-called "planetary" limit. For the caustics of the perturbing object in the wide case we have an extension of the caustics in the parallel and in the vertical direction:

$$\Delta\zeta_{||,wide} = 2(n+1)\frac{\gamma}{s^{\frac{m(n+1)}{m+1}}\left(s^{n+1}-1\right)^{\frac{1}{m+1}}} \tag{3}$$

$$\Delta\zeta_{\perp,wide} = 2(n+1)\frac{\gamma}{s^{\frac{m(n+1)}{m+1}}\left(s^{n+1}+n\right)^{\frac{1}{m+1}}}. \tag{4}$$

7. Conclusions

These mixed binary lenses are important from the astrophysical point of view in the investigation of pairs of galaxies with different halos ($n, m < 1$), in the opening of a new channel in the search for wormholes when they appear in a non-isolated environment ($n = 1$, $m = 2$), in the case in which one object is made up of exotic matter and the other one is a normal star ($n \geq 2$, $m = 1$).

References

1. Einstein, *Entwurf einer verallgemeinerten Relativitätstheorie und eine Theorie der Gravitation* (Leipzig, Berlin: Teubner, 1913).
2. V. Bozza, *Gravitational Lensing by Black Holes* (Gen. Rel. and Grav., 2010).
3. K. A. H. Kitamura, T.; Nakajima, *Demagnifying gravitational lenses toward hunting a clue of exotic matter and energyr* (Phys. Rev. D, 2013).
4. C. M. Bozza, V., *Gravitational Lensing by Black Holes Caustics of* $1/r^n$ *binary gravitational lenses: from galactic haloes to exotic matter* (JCAP, 2016).
5. P. A. Bozza, V., *Gravitational Lensing by Black Holes Alternatives to Schwarzschild in the weak field limit of General Relativity.* (JCAP, 2015).
6. M. C. Bozza V., Pietroni S., *Caustics in gravitational lensing by mixed binary systems* (Universe, 2020).

Hyper-fast positive energy warp drives

E. W. Lentz
Pacific Northwest National Laboratory, Richland, WA 99354 USA
E-mail: erik.lentz@pnnl.gov

Solitons in space–time capable of transporting time-like observers at superluminal speeds have long been tied to violations of the weak, strong, and dominant energy conditions of general relativity. This trend was recently broken by a new approach that identified soliton solutions capable of superluminal travel while being sourced by purely positive energy densities. This is the first example of hyper-fast solitons satisfying the weak energy condition, reopening the discussion of superluminal mechanisms rooted in conventional physics. This article summarizes the recent finding and its context in the literature. Remaining challenges to autonomous superluminal travel, such as the dominant energy condition, horizons, and the identification of a creation mechanism are also discussed.

Keywords: Warp Drive; Energy Conditions; Weak Energy Condition; Superluminal Travel.

1. Introduction

One of the most prominent critiques of superluminal travel within Einstein's general relativity (GR) is that any geometry that facilitates such travel must be largely sourced by a form of negative energy density.[1–18] Other concerns include the immense (magnitude) energy requirements to create a soliton, the difficulty associated with constructing a soliton from a nearly flat spacetime up to the superluminal phase, where the transported central observers become surrounded by a horizon, and the equal difficulties of driving the superluminal phase back the nearly flat spacetime.

There has been a recent uptick in interest regarding warp drives due to a set of papers made available in 2020 and 2021 claiming the construction of solutions that do not require sources with negative energy density,[19–22] obeying the weak energy condition (WEC). These increasingly physical warp drives were the focus of a mini-session at the 16^{th} Marcel Grossmann meeting that covered the history of warp drive research in academia, the recent positive energy warp drive research papers, and two articles regarding their reception.[23,24] This article will concern only the paper written by the author,[19] summarizing its findings and discussing its standing in the literature. Mathematical notation will largely follow that of the original paper.

2. Natário Class Spacetimes

The class of relativistic spacetime metrics describing warp drive geometries in the literature are decomposed according to "3+1" Arnowitt-Deser-Misner (ADM)

formalism[25]

$$ds^2 = -\left(N^2 - N^i N_i\right) dt^2 - 2N_i dx^i dt + h_{ij} dx^i dx^j, \tag{1}$$

where the time coordinate t stratifies spacetime into space-like hypersurfaces, the space metric components h_{ij} evaluated at t provide the intrinsic geometry of that hypersurface, and the similarly-evaluated shift vector components N^i at t provide the coordinate three-velocity of the hypersurface's normal. The time-like unit normal one-form is therefore proportional to the coordinate time element $\mathbf{n}^* = N dt$, and the unit normal vector $\mathbf{n}$ to the hypersurface has components

$$n^\nu = \left(\frac{1}{N}, \frac{N^i}{N}\right). \tag{2}$$

For simplicity, we will use natural units $G = c = 1$.

The majority of previous warp drive papers including Lentz 2021[19] set the lapse function N to unity and the hypersurface metric to be flat under Cartesian coordinates $h_{ij} = \delta_{ij}$. The non-flat geometry is therefore encoded in the three-component shift vector, N_i. The class of spacetimes described in this way have been coined as "Natário spacetimes".[20,23,24]

The projection of the Einstein equation onto the hypersurface normal gives the Hamiltonian constraint of a Natário spacetime

$$G^{\mu\nu} n_\mu n_\nu = 8\pi T^{\mu\nu} n_\mu n_\nu, \tag{3}$$

with the projected stress-energy being referred to as the Eulerian energy density

$$T^{\mu\nu} n_\mu n_\nu = T^{00} = E. \tag{4}$$

The geometric side of this energy constraint equation can be expressed in terms of the extrinsic curvature's trace $K = K^i_i$ and its quadratic hypersurface scalar $K^i_j K^j_i$

$$8\pi E = \frac{1}{2}\left(-K^i_j K^j_i + K^2\right). \tag{5}$$

The combination of extrinsic curvatures expanded in terms of the shift vector components take the form

$$\begin{aligned} K^2 - K^i_j K^j_i = {} & 2\partial_x N_x \partial_y N_y + 2\partial_x N_x \partial_z N_z + 2\partial_z N_z \partial_y N_y \\ & - \frac{1}{2}\left(\partial_x N_y + \partial_y N_x\right)^2 - \frac{1}{2}\left(\partial_x N_z + \partial_z N_x\right)^2 - \frac{1}{2}\left(\partial_z N_y + \partial_y N_z\right)^2. \end{aligned} \tag{6}$$

The warp drive solution of Alcubierre[1] set the precedent for WEC violation by requiring negative Eulerian energy throughout. Specifically, utilizing only a single component of the shift vector in the direction of motion, here taken to be along the positive z-axis, produces the renowned toroid of negative energy density about the soliton bubble of N_z, here displayed in Cartesian coordinates,

$$E_{\text{Alc}} = \frac{-1}{32\pi}\left(\left(\partial_x N_z\right)^2 + \left(\partial_y N_z\right)^2\right). \tag{7}$$

The expansionless ($K = -1/2(\partial_x N_x + \partial_y N_y + \partial_z N_z) = 0$) elliptic relation of Natário 2002[11] restricted the energy form to the negative definite square of the extrinsic curvature

$$E_{\text{Nat}} = \frac{-1}{16\pi} K^i_j K^j_i. \tag{8}$$

Parabolic and hyperbolic relations remained to be explored at the start of 2020.

3. Positive Energy Warp Drives

The soliton geometry of Lentz 2021[19] distinguishes itself from the previous literature in that it satisfies the WEC, even when moving at superluminal speeds. The WEC states that the energy of a spacetime is nowhere negative for any time-like observer. Mathematically, this means that for any time-like vector field X, the projection with the stress-energy tensor T must be non-negative

$$X_\mu T^{\mu\nu} X_\nu \geq 0. \tag{9}$$

The positive-energy soliton was identified through the construction of a set of rules sufficient to define geometries with everywhere non-negative energy. A brief presentation of the rules is given below.

Recall the expansion of the Hamiltonian constraint of Eqn. 6. Observe that the last three elements of the above expression are negative definite, while the first three are of indeterminant type. These first three terms provide opportunity for the Eulerian energy density function to be non-negative under particular configurations, so long as they are everywhere dominate over the first three terms. The next steps focus on such configurations.

The first rule is to reduce the 3D shift vector field to a single potential function, a real-valued function ϕ with spatial gradient relating the shift vector components

$$N_i = \partial_i \phi, \tag{10}$$

satisfying a linear wave equation over the spatial coordinates

$$\partial_x^2 \phi + \partial_y^2 \phi - \frac{2}{v_h^2} \partial_z^2 \phi = \rho, \tag{11}$$

where $v_h/\sqrt{2}$ is the dimensionless wave front 'speed' on the hypersurface, and ρ is the wave equation source function, not to be confused with mass or energy density. This step provides nearly all the structure needed to find the first example positive energy drives.

Two more simplifications are used to set sufficient rules for positive Eulerian energy. The energy functional is reduced to a two-coordinate form (z, x) by restricting ρ and ϕ to be parameterized in the (x, y) plane by the l_1 norm $s = |x| + |y|$,

$$E = \frac{1}{16\pi} \left(2\partial_z^2 \phi \left(\rho + \frac{2}{v_h^2} \partial_z^2 \phi \right) - 4 \left(\partial_z \partial_x \phi \right)^2 \right), \tag{12}$$

which can be bounded from below by

$$16\pi E \geq 2\rho \times \partial_z^2 \phi$$
$$= \rho \times \frac{1}{2v_h} \int_{-\infty}^{\infty} dx' \partial_r \rho(r, |x'| + |y|)|_{r=z-|\Delta x|/v_h}, \tag{13}$$

where the Green's expressions for the shift vector is used in the last expression. From the lower bound expression, the last rule is formed to ensure the Eulerian energy density is everywhere non-negative: the energy function will be non-negative for configurations such that the local source density and the z-component source density gradient integrated along the intersecting 'past' wave trajectories are of the same sign. In other words, the two factors in Expr. 13 must have the same sign.

It is from this sequence of rules that the Eulerian energy can be constrained to be non-negative. Demonstrating the fullness of the WEC takes several additional steps to understand the contributions of the Eulerian momentum and stress components, and is covered in detail in the original publication,[19] but are omitted here in the interest of space. The rules invoked are not strictly necessary to positive-energy warp drives. The solution space of physical warp drives is expected to be much larger and more diverse.

The shift vector of the positive-energy soliton created in Lentz 2021[19] is given in Fig. 1. The soliton moves along the positive z axis at a speed set by the value of the shift vector at the origin of the co-moving coordinate coordinates in Fig. 1, which may be given arbitrary positive value. The transport logistics of the solitons are then similar to that of the Alcubierre solution. The solitons are constructed to contain a central region with minimal tidal forces, where proper time coincides with asymptotic coordinate time, and any Eulerian observer – which in this case is free falling and whose velocity matches the shift vector – within the central region would remain stationary with respect to the soliton. This is the region where a spacecraft would be placed.

The total energy requirements of the positive-energy solitons closely follow that of Pfenning & Ford 1997[3] as applied to the Alcubierre solution

$$E_{tot} = \int E\sqrt{-g} d^3x. \tag{14}$$

For solitons where the radial extent of the central region R is much larger than the thickness of the energy-density laden boundary shell w ($w \ll R$), the energy is estimated to be

$$E_{tot} \sim C v_s^2 \frac{R^2}{w} \tag{15}$$

where C is a form factor typically of order unity. The required energy for a positive-energy soliton with central region mean radius $R = 100$ m and average source thickness along the z-axis $w = 1$ m approaches a mass equivalent of $E_{tot} \sim$ (few) $\times$ $10^{-1} M_{\odot} v_s^2$, which is of the same magnitude as the estimate of an Alcubierre solution of the same dimensions.

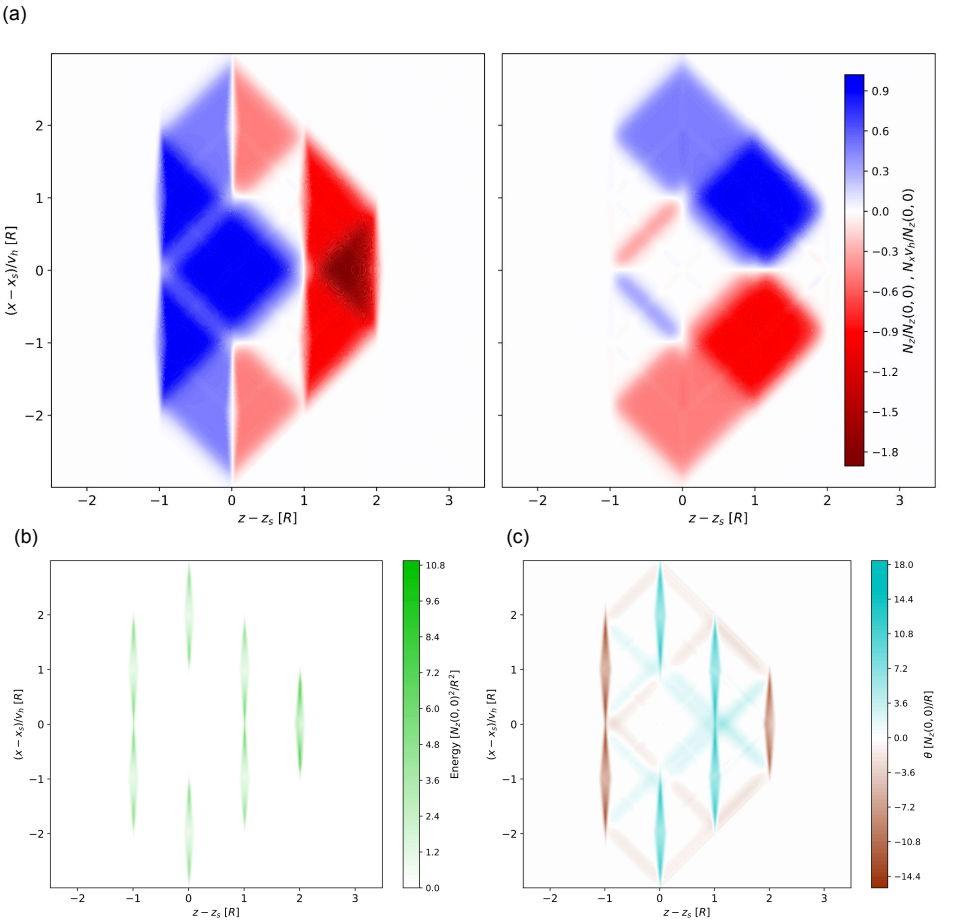


Fig. 1. (a) Projection of the shift vector components N_z (left) and N_x (right) along $(x, 0, z)$. Propagation direction of the soliton is from left to right along the z-axis. The multi-compartment structure is a distinct departure from the single top-hat soliton found in Alcubierre 1994[1] and Natário 2002.[11] Total integrated shift in each direction is 0. Note that the shift vector components are normalized with respect to the value of N_z a the co-moving origin. (b) Projection of the local energy density. The energy density is dominated by those regions containing hyperbolic source ρ, but also extends weakly to the boundaries of the wavefronts. The energy density is everywhere positive for Eulerian observers. (c) Projection of the local volume expansion factor θ. Positive and negative expansion factor are largely associated with negative and positive hyperbolic sources respectively. Non-zero expansion factor also exist in the spaces in-between hyperbolic sources along the hyperbolic wavefronts. Total integrated expansion factor is 0. These plots are taken from Lentz 2021.[19]

4. Addressing the Literature

The findings of Lentz 2021[19] run against the common wisdom of the warp drive literature to that point and the proofs set forth in Olum 1999[6] and Lobo & Crawford

2003[13] stating that any superluminal spacetime must violate the WEC via violations of the null energy condition (NEC). The proofs are both based on an analysis of the Raychaudhuri equation for null geodesics, confined to spacetimes with only a single fastest (superluminal) causal path between two space-like 2-surfaces. The pre-conditions of these proofs are very local in nature and appear analogous to collapsing the interior of a warp drive soliton to a point in order to produce a single fastest causal path. The solitons of the early literature, such as Alcubierre 1994[1] and Natário 2002,[11] have simple structures and can survive this limit. The example positive energy warp drive of Lentz 2021[19] cannot undergo this limit without being destroyed. This drive therefore does not meet the pre-conditions of the proofs and exists outside their scope, implying that the proofs are not applicable.

Several warp drive papers have addressed the findings of Lentz 2021[19] since an early manuscript of it was made publicly available.[20,22,23,23,26] Of particular note, the papers of[23,24] have made several assertions claiming that the solution of Lentz 2021[19] cannot satisfy the WEC. Follow-up correspondence with the authors as well as discussion captured at the recent Marcel Grossmann meeting have demonstrated that these papers did not adequately analyze the contents of Lentz 2021.[19] To summarize the discussions, Santiago et al.[23,24] argue that the Eulerian energy density of a soliton in a Natário class spacetime can be written as the sum of a total divergence and a negative definite term

$$E = \frac{1}{16\pi}\left(\partial_i\left(N_i\partial_j N_j - N_j\partial_j N_i\right) - \frac{1}{2}\omega_i\omega_i\right), \tag{16}$$

where $\omega_i = \epsilon_{ijk}\partial_k N_j$ is the shift vector vorticity. The divergence term is then argued to produce zero net energy if the warp drive is finite in size due to an application of the divergence theorem on the hypersurface where the integral's volume boundary is extended towards infinity where the divergence kernel quickly vanishes, implying that the total Eulerian energy of a Natário spacetime is non-positive. This argument does not hold in the case of Lentz 2021[19] as the Eulerian energy density of the example positive energy soliton is smooth save for the boundaries $x = 0$ and $y = 0$ where stress-energy sources are only continuous, while a requirement of the divergence theorem is that the total divergence be at least first order smooth everywhere. The integral volume boundary therefore cannot be separated from the soliton and instead must be applied in a patchwork,[27] with some boundaries running adjacent to the 2-surfaces $x = 0$ and $y = 0$, where the divergence kernel is non-vanishing. Expansion beyond Natário class of spacetimes may smooth the geometry and sources of positive energy warp drives further.

5. Further Challenges and Future Prospects

There are still numerous challenges between the current state of physical warp drive research and a functioning prototype. I list here several of the more near-term challenges and give my perspective as to how research in these areas may be approached.

The most glaring challenge is the astronomical energy cost of even a modest warp drive, currently measured in solar masses where kilograms is closer to the threshold of human technology. Extreme energy savings is going to be necessary – tens of orders of magnitude – to bring the energy required for a warp drive down to a level that can be tested in a laboratory setting let alone be considered a viable transportation technology.

There exist numerous techniques for reducing the energy requirements of the Alcubierre solution, several of which have been very successful in reducing the (magnitude) energy requirements of the system in excess of thirty orders of magnitude.[7, 10, 16, 18, 28] Unfortunately, each one of these methods in their presented forms require negative energies themselves. One possible approach to uncovering significant energy savings is to modify one of these existing techniques to obey the WEC.

If the required energy can be sufficiently reduced, the next hurdle to approach is modeling the full life cycle of a physical warp drive (creation, acceleration, inertial motion, deceleration, and diffusion). Every previous publication in the field of warp drives has either assumed inertial motion (constant velocity) or has produced an accelerating/decelerating drive that violates the law of covariant conservation of stress-energy-momentum

$$\nabla \cdot \mathbf{T} = \mathbf{0}, \tag{17}$$

that accompanies the Einstein equation. Deriving mechanisms for creation and acceleration is crucial to any experimental test.

The last hurdle I will mention is the full characterization of the sourcing fields, whether it be a plasma or other state of matter and energy. As stated by Matt Visser in the Q&A of my talk, the specification of the drive geometry only is an incomplete description of the full solution. Stress-energy sources must be specified to close the system. In the hypothetical plasma of Lentz 2021,[19] the stress-energy governing equations include the Maxwell equations for the electric and magnetic fields, the equations of motion for each species of matter, and various constituent equations governing the state of the Einstein-Maxwell-matter system. The total system is expected to be far too complex to provide analytical solutions, requiring numerical simulation as the primary means to specify each field of a soliton at any point in its life cycle.

References

1. M. Alcubierre, The warp drive: hyper-fast travel within general relativity, *Classical and Quantum Gravity* **11**, L73 (May 1994).
2. A. E. Everett, Warp drive and causality, *Phys. Rev. D* **53**, 7365 (Jun 1996).
3. M. J. Pfenning and L. H. Ford, The unphysical nature of 'warp drive', *Classical and Quantum Gravity* **14**, 1743 (Jul 1997).
4. W. A. Hiscock, Quantum effects in the alcubierre warp-drive spacetime, *Classical and Quantum Gravity* **14**, L183 (Nov 1997).
5. S. V. Krasnikov, Hyperfast travel in general relativity, *Phys. Rev. D* **57**, 4760 (Apr 1998).

6. K. D. Olum, Superluminal travel requires negative energies, *Phys. Rev. Lett.* **81**, 3567 (Oct 1998).
7. C. V. D. Broeck, A 'warp drive' with more reasonable total energy requirements, *Classical and Quantum Gravity* **16**, 3973 (Nov 1999).
8. M. G. Millis, NASA breakthrough propulsion physics program, *Acta Astronautica* **44**, 175 (1999), Missions to the Outer Solar System and Beyond.
9. M. Visser, B. Bassett and S. Liberati, Superluminal censorship, *Nuclear Physics B - Proceedings Supplements* **88**, 267 (2000).
10. F. Loup, D. Waite and J. Halerewicz, E., Reduced Total Energy Requirements for a Modified Alcubierre Warp Drive Spacetime, *arXiv e-prints*, p. arXiv:grqc/0107097 (July 2001).
11. J. Natário, Warp drive with zero expansion, *Classical and Quantum Gravity* **19**, 1157 (Mar 2002).
12. C. Gauthier, P. Gravel and J. Melanson, New Lower Bounds for Warp Drive Energy, *International Journal of Modern Physics A* **17**, p. 2761 (January 2002).
13. F. Lobo and P. Crawford, *Weak Energy Condition Violation and Superluminal Travel*, in *Current Trends in Relativistic Astrophysics*, eds. L. Fernández-Jambrina and L. M. González-Romero 2003, p. 277.
14. F. S. N. Lobo and M. Visser, Fundamental limitations on 'warp drive' spacetimes, *Classical and Quantum Gravity* **21**, 5871 (Nov 2004).
15. F. S. N. Lobo, Exotic solutions in General Relativity: Traversable wormholes and 'warp drive' spacetimes, *Classical and Quantum Gravity Research* , 1 (October 2008).
16. R. K. Obousy and G. Cleaver, Warp Drive - A New Approach, *Journal of the British Interplanetary Society* **61**, 364 (January 2008).
17. S. Finazzi, S. Liberati and C. Barceló, Semiclassical instability of dynamical warp drives, *Phys. Rev. D* **79**, p. 124017 (Jun 2009).
18. H. White, Warp Field Mechanics 101, *Journal of the British Interplanetary Society* **66**, 242 (January 2013).
19. E. W. Lentz, Breaking the warp barrier: hyper-fast solitons in einstein–maxwell-plasma theory, *Classical and Quantum Gravity* **38**, p. 075015 (Mar 2021).
20. A. Bobrick and G. Martire, Introducing physical warp drives, *Classical and Quantum Gravity* **38**, p. 105009 (Apr 2021).
21. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, Fluid dynamics in the warp drive spacetime geometry, *The European Physical Journal C* **81** (Feb 2021).
22. S. D. B. Fell and L. Heisenberg, Positive energy warp drive from hidden geometric structures, *Classical and Quantum Gravity* **38**, p. 155020 (Jul 2021).
23. J. Santiago, S. Schuster and M. Visser, Generic warp drives violate the null energy condition, *arXiv e-prints* , p. arXiv:2105.03079 (May 2021).
24. J. Santiago, S. Schuster and M. Visser, Tractor Beams, Pressor Beams and Stressor Beams in General Relativity, *Universe* **7**, p. 271 (July 2021).
25. R. Arnowitt, S. Deser and C. W. Misner, Dynamical structure and definition of energy in general relativity, *Phys. Rev.* **116**, 1322 (Dec 1959).
26. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, Perfect fluid warp drive solutions with the cosmological constant, *The European Physical Journal Plus* **136** (Sep 2021).
27. T. Dray and C. Hellaby, The patchwork divergence theorem, *Journal of Mathematical Physics* **35**, p. 5922–5929 (Nov 1994).
28. S. Krasnikov, Quantum inequalities do not forbid spacetime shortcuts, *Phys. Rev. D* **67**, p. 104013 (May 2003).

From black-bounce to traversable wormhole, and beyond

Alex Simpson

School of Mathematics and Statistics, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
E-mail: alex.simpson@sms.vuw.ac.nz

Key results from the literature pertaining to a class of nonsingular black hole mimickers are explored. The family of candidate spacetimes is for now labelled the 'black-bounce' family, stemming from the original so-called 'Simpson–Visser' spacetime in static spherical symmetry. All model geometries are analysed through the lens of standard general relativity, are globally free from curvature singularities, pass all weak-field observational tests, and smoothly interpolate between regular black holes and traversable wormholes. The discourse is segregated along geometrical lines, with candidate spacetimes each belonging to one of: static spherical symmetry, spherical symmetry with dynamics, and stationary axisymmetry.

Keywords: Black-bounce, regular black hole, traversable wormhole, Simpson–Visser, Vaidya, thin-shell, black hole mimic.

1. Introduction

In classical general relativity (GR), solutions to the Einstein equations representing black holes typically contain curvature singularities at their cores. These singularities occur at a distance scale that only a complete theory of quantum gravity could adequately describe. In the absence of such a theory, or at least in the absence of one that has any phenomenological verification, one desirable approach is to appeal to the recent advances made in observational and gravitational wave astronomy. In view of the propagation of gravitational waves emanating from an astrophysical source being directly observed in the recent LIGO/VIRGO merger events,[1,2] it is well-motivated to explore mathematically tractable candidate spacetimes, which are curvature-singularity-free alternatives to classical black holes, through the lens of standard GR. The community hopes that LIGO/VIRGO (or more likely LISA[3]) will eventually be able to provide phenomenological evidence that allows us to delineate between specific candidate spacetimes based on their astrophysical accuracy. Theoretical physicists would then have experimentally informed clues as to which modifications to the Einstein equations are most desirable in working towards a so-called 'theory of everything' (or, indeed, whether an entirely different framework is required). Discussion regarding the extraction of astrophysical observables for nonsingular sample spacetimes in several frameworks can be found in references 4–14.

One such family of candidate geometries which models alternatives to classical black holes is the family of 'black-bounce' spacetimes.[15–20] All of these black hole mimickers are globally free from curvature singularities, pass all weak-field

observational tests of standard GR, and belong either to the class of regular black holes or traversable wormholes. By 'regular black hole', one means in the sense of Bardeen,[21] with regularity defined *via* enforcing global finiteness on nonzero curvature tensor components and Riemann curvature invariants. Regular black holes have a well-established lineage in the historical literature.[21–27] By 'traversable wormhole', one means in the sense of Morris and Thorne;[28] a horizon-free geometry with a centralised throat connecting two asymptotically Minkowski regions of spacetime and satisfying the 'flare-out' condition for the area function: $A''(r_{throat}) > 0$. Various intriguing wormhole spacetimes have been developed and explored in references 28–31.

For exposition, it is prudent to separate the discourse surrounding the family of black-bounce spacetimes along geometrical lines. Broadly speaking, there are three relevant geometrical categories: static spherical symmetry, spherical symmetry with dynamics, and stationary axisymmetry (this also *somewhat* follows the chronological development of the literature). Unless otherwise stated, all candidate spacetimes discussed have metric signature $(-,+,+,+)$ outside any would-be horizons.

2. Static spherical symmetry

2.1. *Simpson–Visser*

The so-called 'Simpson–Visser' (SV) spacetime, initially presented in reference 15, is represented by the line element

$$\mathrm{d}s^2 = -\left(1-\frac{2m}{\sqrt{r^2+\ell^2}}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\left(1-\frac{2m}{\sqrt{r^2+\ell^2}}\right)} + \left(r^2+\ell^2\right)\mathrm{d}\Omega_2^2\,. \tag{1}$$

It should be noted that the parameter 'ℓ' was in fact labelled 'a' in the original article.[15] This minor alteration is performed for consistency with the remainder of the discourse herein, as when in the axisymmetric environment of § 4, it is prudent to use ℓ in order to avoid confusion with the spin parameter from Kerr spacetime, a.

The original motivation for the construction of the metric specified by Eq. (1) was to minimally modify the Schwarzschild solution in common curvature coordinates such that the resulting candidate spacetime was globally nonsingular. It was considered that the most straightforward way to achieve this was to introduce a new scalar parameter to the line element in a tightly controlled manner; this is ℓ in Eq. (1). By minimally modifying Schwarzschild, it was hoped the result would have a high degree of mathematical tractability. When viewed as a modification of Schwarzschild, there are the following two alterations:

- $m \mapsto m(r) = \frac{mr}{\sqrt{r^2+\ell^2}}$;
- The coefficient of $\mathrm{d}\Omega_2^2$ is modified from $r^2 \to r^2+\ell^2$.

Analysis of the nonzero curvature tensor components and Riemann curvature invariants concludes that for $|\ell| > 0$, the resulting candidate spacetime is globally

regular. It was also noticed that Eq. (1) has very neat limiting behaviour. In the limit as $m \to 0$, one obtains

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}r^2 + \left(r^2 + \ell^2\right)\, \mathrm{d}\Omega_2^2 \,. \tag{2}$$

This is precisely the two-way traversable wormhole solution as presented in Morris and Thorne's aforementioned seminal paper[28] (and arguably the most straightforward of all traversable wormhole geometries). In the limit as $\ell \to 0$, Eq. (1) becomes the Schwarzschild solution in the usual curvature coordinates. The newly introduced scalar parameter ℓ can hence be viewed as quantifying the extent of the deviation away from Schwarzschild, and it invokes a rich, ℓ-dependent horizon structure. The causal structure is characterised by

$$r_H = \sqrt{(2m)^2 - \ell^2} \,, \tag{3}$$

and the candidate spacetime neatly interpolates between the following qualitatively different geometries:

- $\ell = 0$ corresponds to the Schwarzschild black hole;
- $|\ell| \in (0, 2m)$ corresponds to a regular black hole in the sense of Bardeen;
- $|\ell| = 2m$ corresponds to a one-way wormhole with an extremal null throat;
- $|\ell| > 2m$ corresponds to a two-way traversable wormhole geometry in the canonical sense of Morris and Thorne.

The Carter–Penrose diagrams for the two lesser known cases are worth closer examination; this is when $|\ell| \in (0, 2m)$ (see Fig. 1), and when $|\ell| = 2m$ (see Fig. 2).

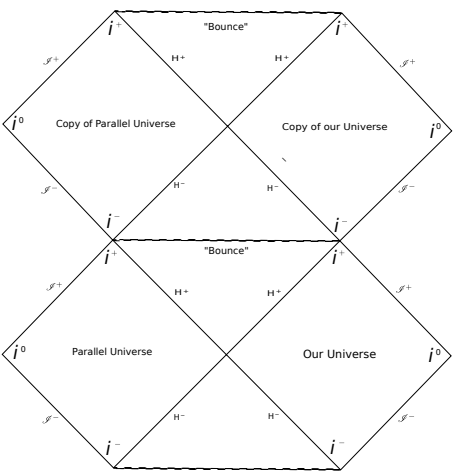


Fig. 1. Carter–Penrose diagram for the maximally extended spacetime when $|\ell| \in (0, 2m)$. In this example one 'bounces' through the $r = 0$ hypersurface in each black hole region into a future copy of the universe *ad infinitum*.

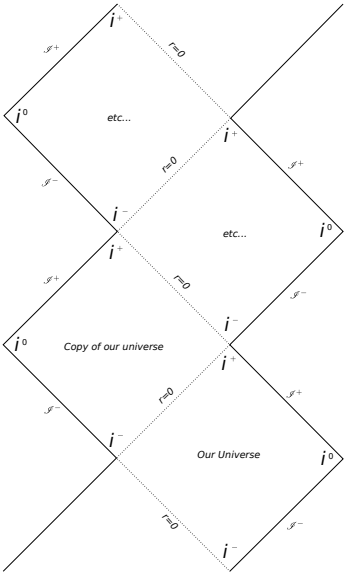


Fig. 2. Carter–Penrose diagram for the maximally extended spacetime in the case when $|\ell| = 2m$. In this example one has a one-way wormhole geometry with an extremal null throat.

Analysing the satisfaction/violation of the standard point-wise energy conditions of GR, one concludes that when $|\ell| > 0$, the radial 'null energy condition' (NEC) is manifestly violated in the region $\sqrt{r^2 + \ell^2} > 2m$. This is *outside* any would-be horizons. In the context of static spherical symmetry, this is sufficient to conclude that all of the standard point-wise energy conditions shall be similarly violated. If horizons are present, then the geometry has surface gravity

$$\kappa = \frac{\sqrt{(2m)^2 - \ell^2}}{8m^2} = \kappa_{Sch.} \sqrt{1 - \frac{\ell^2}{(2m)^2}} \,, \tag{4}$$

and hence associated Hawking temperature

$$T_H = \frac{\hbar\sqrt{(2m)^2 - \ell^2}}{16\pi\, k_B\, m^2} = T_{H,Sch.} \sqrt{1 - \frac{\ell^2}{(2m)^2}} \,. \tag{5}$$

SV spacetime is amenable to the extraction of astrophysical observables, and the coordinate locations of the photon sphere for null orbits and the ISCO for timelike orbits are straightforward, given by

$$r_\gamma = \sqrt{(3m)^2 - \ell^2} \,; \qquad r_{ISCO} = \sqrt{(6m)^2 - \ell^2} \,. \tag{6}$$

Further research analysing SV spacetime has been performed in a plethora of other papers; please see reference 32 for details. These analyses include discussion of the quasi-normal modes and associated ringdown,[33,34] calculations pertaining to shadows and gravitational lensing effects,[35–40] as well as discourse surrounding precession phenomena.[41]

2.2. *Black-bounce Reissner–Nordström*

The extension of SV spacetime which represents the black-bounce analog to Reissner–Nordström (RN) spacetime was constructed and analysed in reference 20. There is a simple regularisation procedure which can be applied to any spherically symmetric or axisymmetric geometry in possession of a curvature singularity at $r = 0$ in the standard (t, r, θ, ϕ) curvature coordinates. The procedure outputs a candidate geometry which is globally nonsingular, and maintains the manifest symmetries. It is as follows:

- Leave the object $\mathrm{d}r$ in the line element undisturbed;
- Whenever the metric components $g_{\mu\nu}$ have an explicit r-dependence, replace the r-coordinate with $\sqrt{r^2 + \ell^2}$, where ℓ is some length scale, performing the same holistic role as the ℓ parameter in SV spacetime (such parameters are often identified with the Planck scale; $\ell \sim m_p$).

It should be emphasised that this procedure is *not* a coordinate transformation. Application of this procedure to the usual RN spacetime in standard curvature coordinates yields the following line element, the 'black-bounce Reissner–Nordström' (bbRN) spacetime:

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{\sqrt{r^2+\ell^2}} + \frac{Q^2}{r^2+\ell^2}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1 - \frac{2m}{\sqrt{r^2+\ell^2}} + \frac{Q^2}{r^2+\ell^2}} + (r^2+\ell^2)\,\mathrm{d}\Omega_2^2 \ . \quad (7)$$

For $|\ell| > 0$, this spacetime neatly interpolates between nonsingular electrovac black holes and traversable wormholes in standard GR, depending on the value of the charge parameter Q and the 'bounce' parameter ℓ. When compared with standard RN spacetime, the domain for the r coordinate is extended from $r \in [0, +\infty)$ to $r \in (-\infty, +\infty)$. The presence of the additional charge term invokes a richer causal structure than that of SV spacetime. Horizons are characterised by

$$r_H = S_1\sqrt{(m + S_2\sqrt{m^2 - Q^2})^2 - \ell^2} \ . \quad (8)$$

Here $S_1 = S_2 = \pm 1$, with S_1 fixing which universe one is in, and S_2 determining whether it is an outer ($S_2 = +1$) or inner ($S_2 = -1$) horizon. There are several qualitatively different geometries:

- $|Q| > m$: There are no horizons, and the geometry models a two-way traversable wormhole in the sense of Morris and Thorne;
- $|Q| \leq m$ but $|\ell| > m \pm \sqrt{m^2 - Q^2}$: first the inner, and then the outer horizons vanish, leaving a traversable wormhole;
- $|Q| = m$ and $|\ell| \leq m$: one finds extremal horizons at $r_H = \pm\sqrt{m^2 - \ell^2}$;
- $|Q| \leq m$ and $|\ell| < m - \sqrt{m^2 - Q^2}$: one observes the 'standard' causal structure of both an inner and outer horizon present in each universe.

If horizons are present, the surface gravity of the outer horizon is given by

$$\kappa = \frac{\sqrt{(m+\sqrt{m^2-Q^2})^2-\ell^2}\sqrt{m^2-Q^2}}{(m+\sqrt{m^2-Q^2})^3} = \kappa_{\rm RN}\sqrt{\frac{r_H^2}{r_H^2+\ell^2}}\ . \tag{9}$$

The following astrophysical observables are amenable to extraction; firstly the location of the photon sphere

$$r_\gamma = \sqrt{\frac{m}{2}(9m+3\sqrt{9m^2-8Q^2})-2Q^2-\ell^2}\ , \tag{10}$$

and also the location of the ISCO for timelike particles

$$r_{ISCO} = \frac{\sqrt{9m^4Q^4-6m^2Q^2(A^2+2Am^2+4m^4)+(A^2+2Am^2+4m^4)^2-A^2\ell^2m^2}}{mA},$$

$$A = \left[2m^2Q^4+m^2(\pm B-9m^2)Q^2+8m^6\right]^{\frac{1}{3}}\ , \quad B=\sqrt{4Q^4-9m^2Q^2+5m^4}\ . \tag{11}$$

In standard GR, the stress-energy tensor takes the following form outside the outer horizon (or inside the inner horizon):

$$\frac{1}{8\pi}G^{\hat\mu}{}_{\hat\nu} = T^{\hat\mu}{}_{\hat\nu} = [T_{\rm bb}]^{\hat\mu}{}_{\hat\nu} + [T_Q]^{\hat\mu}{}_{\hat\nu} = {\rm diag}(-\rho,p_r,p_t,p_t)\ , \tag{12}$$

whilst in between the two horizons one has

$$\frac{1}{8\pi}G^{\hat\mu}{}_{\hat\nu} = T^{\hat\mu}{}_{\hat\nu} = [T_{\rm bb}]^{\hat\mu}{}_{\hat\nu} + [T_Q]^{\hat\mu}{}_{\hat\nu} = {\rm diag}(p_r,-\rho,p_t,p_t)\ , \tag{13}$$

where $[T_{\rm bb}]^{\hat\mu}{}_{\hat\nu}$ is the gravitational stress-energy tensor (identical to that of SV spacetime), and $[T_Q]^{\hat\mu}{}_{\hat\nu}$ is the charge-dependent contribution, corresponding to the electromagnetic stress-energy tensor. Isolating the charge-dependent components, it is straightforward to make the following decomposition

$$[T_Q]^{\hat\mu}{}_{\hat\nu} = \frac{Q^2r^2}{8\pi(r^2+\ell^2)^3}\left[{\rm diag}(-1,-1,1,1)+{\rm diag}(\frac{2\ell^2}{r^2},0,0,0)\right]$$

$$= [T_{\rm Maxwell}]^{\hat\mu}{}_{\hat\nu} + \Xi\, V^{\hat\mu}V_{\hat\nu}\ , \tag{14}$$

where $[T_{\rm Maxwell}]^{\hat\mu}{}_{\hat\nu}$ is the usual stress-energy associated with standard Maxwell electromagnetism, and $\Xi\, V^{\hat\mu}V_{\hat\nu}$ is interpreted as the stress-energy of "charged dust", with $V^{\hat\mu}$ being the normalised time-translation Killing vector. The density of the charged dust, Ξ, is given by

$$\Xi = -\frac{1}{4\pi}\frac{Q^2\ell^2}{(r^2+\ell^2)^3}\ , \tag{15}$$

while the electric field strength is determined *via*

$$[T_Q]^{\hat t}{}_{\hat t} = -\rho_{\rm em} = [T_{\rm Maxwell}]^{\hat t}{}_{\hat t} - \Xi = -\frac{1}{8\pi}E^2 - \Xi\ ;$$

$$\implies \quad E = \frac{Qr}{(r^2+\ell^2)^{\frac{3}{2}}} = E_{\rm RN}\left[\frac{r^3}{(r^2+\ell^2)^{\frac{3}{2}}}\right]\ , \tag{16}$$

where $E_{\rm RN}$ is the electric field strength for a usual RN black hole. The electromagnetic potential is readily obtained *via* integration of Eq. (16),

$$A_\mu = (\Phi_{\rm em}(r), 0, 0, 0) = -\frac{Q}{\sqrt{r^2+\ell^2}}(1,0,0,0)\ , \tag{17}$$

which is simply the RN electromagnetic potential under the mapping $r \to \sqrt{r^2+\ell^2}$. All told, one has:

$$[T_Q]^{\hat\mu}{}_{\hat\nu} = \frac{1}{4\pi}\left[-F^{\hat\mu}{}_{\hat\alpha}F^{\hat\alpha}{}_{\hat\nu} - \frac{1}{4}\delta^{\hat\mu}{}_{\hat\nu}F^2\right] - \frac{1}{4\pi}\frac{Q^2\ell^2}{(r^2+\ell^2)^3}\,V^{\hat\mu}V_{\hat\nu}\ ,$$

$$\text{with}\quad F_{\hat\mu\hat\nu} = \nabla_{\hat\mu}A_{\hat\nu} - \nabla_{\hat\nu}A_{\hat\mu}\ . \tag{18}$$

Consequently, in standard GR the bbRN spacetime possesses stress-energy components which are interpreted as everywhere-SV in the gravitational sector, coupled to standard Maxwell's electromagnetism in the presence of charged dust in the electromagnetic sector.

2.3. *Generalised extensions to Simpson–Visser*

Further extensions to the SV line element in the context of static spherical symmetry were explored in reference 16, with some useful general theorems also being presented. The first of these theorems is general to all static spacetimes, and is concerned with expediting the 'usual' test for curvature-regularity; examination of the finiteness of all nonzero components of the Riemann curvature tensor in an orthonormal basis. It is as follows, with the proof provided in reference 16:

Theorem 2.1. *For any static spacetime, in the strictly static region, the Kretschmann scalar is positive semi-definite, being a sum of squares which involves all of the nonzero components* $R^{\hat\mu\hat\nu}{}_{\hat\alpha\hat\beta}$. *Then if this scalar is finite, all of the orthonormal components of the Riemann curvature tensor must also be finite. Consequently, for static candidate spacetimes, confirmation of the global finiteness of the Kretschmann scalar is sufficient to conclude as to curvature-regularity in the sense of Bardeen.*

It is by now well-known that the most general static spherically symmetric line element can always locally be put into the following form, typically known as "Buchdahl" coordinates[42–46] (for § 2.3 only, the metric signature $(+,-,-,-)$ is adopted outside horizons for consistency with the discourse in reference 16)

$$\mathrm{d}s^2 = f(r)\,\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{f(r)} - \Sigma^2(r)\,\mathrm{d}\Omega_2^2\ . \tag{19}$$

Now constrained to Buchdahl coordinates, it is straightforward to establish the following list of sufficient conditions for curvature regularity in the sense of Bardeen (note these constraints exist as an independent result from Theorem 2.1; which of the techniques is preferable/applicable should be determined from context):

- $\Sigma(r) \neq 0$ globally;
- $\Sigma'(r)$ and $\Sigma''(r)$ must be globally finite;
- $f(r)$, $f'(r)$, and $f''(r)$ must be globally finite.

With respect to Buchdahl coordinates, another theorem was presented in reference 16. Concerning satisfaction/violation of the standard point-wise energy conditions of GR, it states:

Theorem 2.2. *For any static anisotropic fluid sphere with line element as in Eq. (19), all of the standard point-wise energy conditions are violated whenever* $f(r) \neq 0$, $\Sigma(r) > 0$, *and* $\Sigma''(r) > 0$.

In the same article,[16] further decomposition of Eq. (19) *via* $f(r) = 1 - 2M(r)/\Sigma(r)$ was performed to allow for a two-parameter extension of SV spacetime. One fixes

$$\Sigma(r) = \sqrt{r^2 + \ell^2}\ , \qquad M(r) = \frac{m r^k \sqrt{r^2 + \ell^2}}{\left(r^{2n} + \ell^{2n}\right)^{\frac{k+1}{2n}}}\ , \tag{20}$$

with the form for $M(r)$ mathematically inspired by the Fan–Wang mass function for regular black holes.[27] The resulting class of geometries has the following properties:

- SV spacetime is recovered *via* fixing $n = 1, k = 0$;
- $\forall\ n, k \in \mathbb{Z}^+$, Schwarzschild in the usual curvature coordinates is recovered as $\ell \to 0$;
- Enforcing $|\ell| > 0$, $\forall\ n, k \in \mathbb{Z}^+$ the candidate geometry is globally regular.

Several models of interest possessing vibrant causal structures can be explored using Eq. (20), each corresponding to various fixed values of both n and k. All models in some fashion smoothly interpolate between regular black holes and traversable wormholes. Several other model geometries were also explored in reference 16, *via* certain modifications to the $M(r)$ present in Eq. (20), all within the geometric context of static spherical symmetry in Buchdahl coordinates.

3. Spherical symmetry with dynamics

3.1. *Vaidya black-bounce*

The SV metric was elevated to the regime of dynamical spherical symmetry in reference 17. One first rewrites the line element from Eq. (1) in Eddington–Finkelstein coordinates, before playing a Vaidya-like 'trick' by allowing the mass parameter m to be a function of the null time coordinate w. The resulting candidate spacetime is given by the following line element

$$\mathrm{d}s^2 = -\left(1 - \frac{2m(w)}{\sqrt{r^2 + \ell^2}}\right) \mathrm{d}w^2 - (\pm 2\,\mathrm{d}w\mathrm{d}r) + (r^2 + \ell^2)\,\mathrm{d}\Omega_2^2\ , \tag{21}$$

where $w = \{u, v\}$ denotes the *outgoing/ingoing* null time coordinate, representing *retarded/advanced* time respectively. In the limit as $m(w) \to m$, SV spacetime is

recovered (in Eddington–Finkelstein coordinates). Introducing dynamics in this specific manner allows for one to discuss simple phenomenological models of an evolving regular black hole/traversable wormhole geometry, either *via* net accretion or net evaporation, whilst still keeping the discourse mathematically tractable. Analysis of the radial null curves yields a dynamical horizon location

$$r_H(w) = \pm\sqrt{[2m(w)]^2 - \ell^2}\ . \tag{22}$$

While this permits analysis of numerous phenomenological models, the most qualitatively interesting are:

- Increasing $m(v)$ crossing the $\ell/2$ limit describing the conversion of a traversable wormhole into a regular black hole *via* the accretion of null dust. This qualitative scenario is depicted in Fig. 3;
- Decreasing $m(u)$ crossing the $\ell/2$ limit describes the evaporation of a regular black hole leaving a traversable wormhole remnant. The causal structure is depicted in Fig. 4. This phenomena is causally equivalent to a regular black hole transmuting into a traversable wormhole *via* the accretion of phantom energy (though mathematically this scenario would instead correspond to increasing $m(v)$ crossing the $\ell/2$ limit).

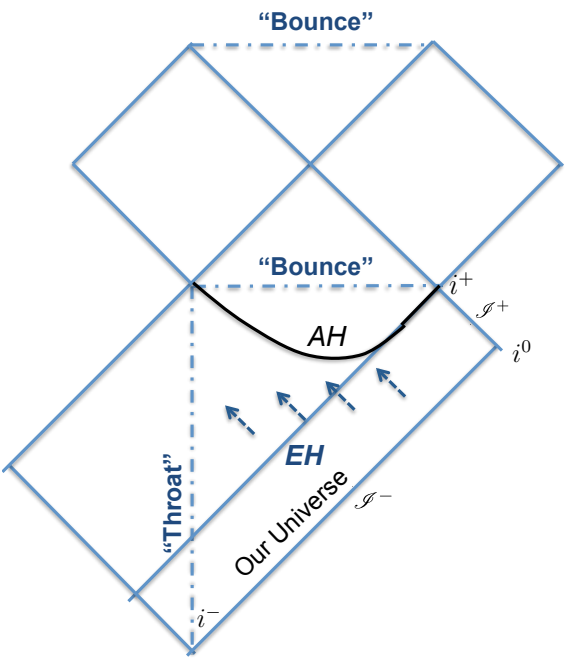


Fig. 3. Carter–Penrose diagram for a wormhole to black-bounce transition *via* accretion of null dust.

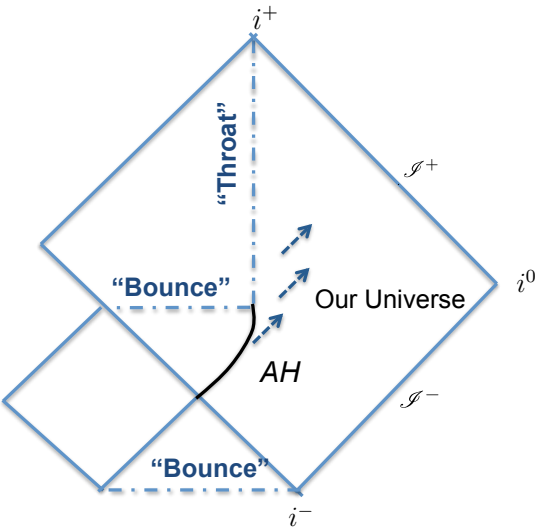


Fig. 4. Carter–Penrose diagram for a black-bounce to wormhole transition due to the emission of positive energy. This is essentially the back reaction of the Hawking evaporation for a semi-classical regular black hole; in this case leaving a wormhole remnant.

General discussion concerning evaporation/accretion models involving both regular black holes and traversable wormholes can be found in references 47–54. It is worth emphasising that one is able to classically describe the transmutation of a wormhole into a regular black hole, or *vice versa*, only because the curvature-regularity of the black hole implies there is no topology change.

In what has become a typical result for nonsingular black hole mimickers in standard GR, the candidate spacetime is found to be in global violation of the NEC. For the 'Vaidya black-bounce' spacetime, the existence of the radial null vector $k^\mu = (0, 1, 0, 0)$ implies that one has

$$T_{\mu\nu}k^\mu k^\nu \propto G_{\mu\nu}k^\mu k^\nu = -\frac{2\ell^2}{(r^2+\ell^2)^2} < 0\ . \tag{23}$$

For a full analysis of the nonzero curvature tensor components and Riemann curvature invariants for the Vaidya black-bounce geometry, as well as a more detailed discussion surrounding energy condition satisfaction/violation and various phenomenological models, please see reference 17.

3.2. *Thin-shell black-bounce traversable wormhole*

A thin-shell traversable wormhole variant of SV spacetime was constructed and analysed in reference 18. The thin-shell surgery is performed *via* a 'cut and paste' procedure which has its mathematical roots in the Darmois–Israel formalism.[55] Physically, the geometrodynamics are kept as close to standard GR as possible

via examination of the Lanczos equations (please see references 56–58 for details pertaining to thin-shell surgery).

In the thin-shell construction,[18] one analyses both the stability and the evolution of an SV thin-shell traversable wormhole under linearized radial perturbation of the throat. The 'bounce' parameter ℓ induces a nonzero flux term arising from the conservation identity due to the second contracted Gauss–Codazzi–Mainardi equation. Consequently, enforcing a nonzero ℓ leads to an additional constraint on the stability analysis. Several qualitative examples from the parameter space are explored, and the viable stability regions in each case are readily extracted.

4. Stationary axisymmetry

4.1. *Black-bounce Kerr–Newman*

The first entry from the family of 'black-bounce' spacetimes belonging to the geometrical regime of stationary axisymmetry was constructed and explored by Liberati *et al.* in reference 19. This spacetime was constructed *via* application of the Newman–Janis procedure to standard SV spacetime, and acts as the black-bounce analog to Kerr spacetime. Given that in any real, astrophysical setting, one shall have both rotation and the condition that $|Q|/m \ll 1$, this is the most astrophysically relevant member of the family of black-bounce spacetimes. The black-bounce analog to Kerr–Newman spacetime was then discovered in reference 20; this was obtained *via* application of the regularisation procedure outlined in § 2.2, applied to the Kerr–Newman geometry in standard Boyer–Lindquist (BL) coordinates. The line element is given by

$$\mathrm{d}s^2 = -\frac{\Delta}{\rho^2}(a\sin^2\theta\,\mathrm{d}\phi-\mathrm{d}t)^2+\frac{\sin^2\theta}{\rho^2}[(r^2+\ell^2+a^2)\,\mathrm{d}\phi-a\,\mathrm{d}t]^2+\frac{\rho^2}{\Delta}\,\mathrm{d}r^2+\rho^2\,\mathrm{d}\theta^2\,, \quad (24)$$

with the usual Kerr quantities modified to be

$$\rho^2 = r^2+\ell^2+a^2\cos^2\theta\,, \qquad \Delta = r^2+\ell^2+a^2-2m\sqrt{r^2+\ell^2}+Q^2\,. \quad (25)$$

In the astrophysically relevant limit as $Q \to 0$, the Kerr-analog explored in reference 19 is recovered precisely. For specifics pertaining to 'black-bounce Kerr' (bbK), please consult reference 19. Generally speaking, if one takes the $Q \to 0$ limit for the relevant results obtained for the 'black-bounce Kerr–Newman' (bbKN) spacetime (Eq. (24)), the analogous result for bbK is recovered precisely. In the limit as the spin parameter $a \to 0$, bbKN reduces to the bbRN geometry from § 2.2.

One can see, somewhat trivially, that the natural domains for the angular and temporal coordinates are unaffected when comparing Eq. (24) with standard Kerr–Newman spacetime in BL coordinates. In contrast, the domain for the radial coordinate extends from $r \in [0,+\infty)$ to $r \in (-\infty,+\infty)$, in exactly the same fashion as was seen for bbRN spacetime in § 2.2. Examination of the Riemann curvature invariants and orthonormal Riemann tensor components reveals that they are globally finite when $|\ell| > 0$, hence the geometry is free from curvature singularities. Consequently,

bbKN qualitatively interpolates between charged rotating regular black holes and charged rotating traversable wormholes. Both the Ricci and Einstein tensors are diagonal in an orthonormal basis.

The causal structure is nontrivial, with horizons characterised by the roots of Δ. This gives the following horizon location(s):

$$r_H = S_1\sqrt{(m + S_2\sqrt{m^2 - Q^2 - a^2})^2 - \ell^2}\ . \tag{26}$$

As it was in § 2.2, $S_1 = S_2 = \pm 1$, with S_1 fixing which universe one is in, and S_2 determining whether it is an outer ($S_2 = +1$) or inner ($S_2 = -1$) horizon. The qualitatively different possible causal structures are the same as explored in § 2.2, with the quantitative aspects slightly shifted due to the presence of the spin parameter a. Notably, in the standard Kerr–Newman geometry in BL coordinates one demands $Q^2 + a^2 \leq m^2$ in order to avoid naked singularities. In the case for bbKN, arbitrary values of both spin and charge may be considered, as in the absence of any horizons one has a traversable wormhole geometry with a rotating throat. Therefore, if one uses bbKN to model astrophysical objects, the weak cosmic censorship conjecture is trivially satisfied.

If horizons are present, their surface gravity is given by

$$\kappa_{S_2} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{\Delta}{r^2 + \ell^2 + a^2}\right)\bigg|_{r=r_H} = \kappa^{\mathrm{KN}}_{S_2}\sqrt{\frac{r_H^2}{r_H^2 + \ell^2}}\ , \tag{27}$$

where S_2 fixes whether one is describing the surface gravity of an outer or inner horizon. The ergosurface is characterised by $g_{tt} = 0$, giving

$$r_{\mathrm{erg}} = S_1\sqrt{(m + S_2\sqrt{m^2 - Q^2 - a^2\cos^2\theta})^2 - \ell^2}\ . \tag{28}$$

An informative result pertaining to circular orbits in the equatorial plane for the bbKN spacetime is as follows: if standard Kerr–Newman in BL coordinates has an equatorial circular orbit at $r_{\mathrm{KN},0}$, then bbKN possesses an analogous equatorial circular orbit at coordinate location $r_0 = \sqrt{r^2_{\mathrm{KN},0} - \ell^2}$, provided that $r^2_{\mathrm{KN},0} \geq \ell^2$.

The candidate spacetime possesses a nontrivial Killing tensor; upon definition of the objects

$$l^\mu = \left(\frac{r^2 + \ell^2 + a^2}{\Delta}, 1, 0, \frac{a}{\Delta}\right)\ ,\quad \text{and } n^\mu = \frac{1}{2\rho^2}(r^2 + \ell^2 + a^2, -\Delta, 0, a)\ , \tag{29}$$

one finds that the following is a Killing tensor, *i.e.*, $K_{(\mu\nu;\lambda)} = 0$:

$$K_{\mu\nu} = \rho^2(l_\mu n_\nu + l_\nu n_\mu) + (r^2 + \ell^2)g_{\mu\nu}\ . \tag{30}$$

In this context, the objects l^μ and n^μ are a pair of geodesic null vectors belonging to a generalised Kinnersley tetrad (see reference 19 for details). The Killing tensor from Eq. (30) has diagonal matrix representation. The existence of a nontrivial Killing tensor directly implies the existence of a generalised Carter constant $\mathcal{C}$, and together with the conserved quantities arising from the temporal and azimuthal

Killing vectors (energy and angular momentum per unit mass), as well as the conserved quantity arising from the metric itself (which is a *trivial* Killing tensor), this gives four constants of the motion. This, in principle, means that the geodesics governing the motion of test particles are integrable on the bbKN spacetime. It should be noted that the resulting integral formulae may or may not be analytic.

In reference 59, a refinement was made to Proposition 1.3 from reference 60. It should also be noted that this result is implicitly present in reference 61. The result is best summarised by the following theorem:

Theorem 4.1. *Let $(\mathcal{M}, g_{\mu\nu})$ be a Lorentzian manifold in possession of a nontrivial Killing tensor $K_{\mu\nu}$. Then upon definition of the Carter operator as: $\mathcal{K}\Phi = \nabla_\mu (K^{\mu\nu}\nabla_\nu\Phi)$, and the D'Alembertian scalar wave operator as: $\Box\Phi = \nabla_\mu (g^{\mu\nu}\nabla_\nu\Phi)$, there is the following result:*

$$[\mathcal{K}, \Box]\Phi = \frac{2}{3}\left(\nabla_\mu [R, K]^\mu{}_\nu\right)\nabla^\nu\Phi \ . \tag{31}$$

A sufficient condition for this operator commutator to vanish is therefore the commutativity of the Ricci and Killing tensors via *matrix multiplication; $R^\mu{}_\alpha K^\alpha{}_\nu = K^\mu{}_\alpha R^\alpha{}_\nu \implies [R, K]^\mu{}_\nu = 0$. Hence for any candidate spacetime with a nontrivial Killing tensor, commutativity of the Ricci and Killing tensors* via *matrix multiplication is sufficient to conclude that the massive minimally coupled Klein–Gordon equation is separable on the candidate geometry.*

In view of Theorem 4.1, given that both the Ricci and nontrivial Killing tensors for the bbKN geometry have a diagonal matrix representation, one may conclude that the scalar wave equation is separable on bbKN spacetime.

Given the existence of a nontrivial Killing tensor, it is natural to ponder the existence of the full 'Killing tower'[62] in bbKN spacetime. One finds a 'would-be' Killing–Yano tensor, of the form

$$\begin{aligned} f_{\mu\nu} = &\begin{pmatrix} 0 & -a\cos\theta & 0 & 0 \\ a\cos\theta & 0 & 0 & -a^2\cos\theta\sin^2\theta \\ 0 & 0 & 0 & 0 \\ 0 & a^2\cos\theta\sin^2\theta & 0 & 0 \end{pmatrix} \\ &+ \sqrt{r^2+\ell^2}\sin\theta \begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 0 & 0 & 0 \\ -a & 0 & 0 & (r^2+\ell^2+a^2) \\ 0 & 0 & -(r^2+\ell^2+a^2) & 0 \end{pmatrix}. \end{aligned} \tag{32}$$

It is not difficult to verify that $f_{\mu\alpha}f^\alpha{}_\nu = K_{\mu\nu}$, and as such this is a genuine 'square root' of the Killing tensor, however it fails to be a genuine Killing–Yano tensor because $f_{\mu(\nu;\alpha)} \neq 0$. It is a known result that a candidate spacetime only possesses a closed conformal Killing–Yano tensor if the spacetime is of Petrov type D.[62] However, it was proven by Liberati *et al.* in reference 19 that the bbK spacetime is Petrov type I, and it is a direct implication of this proof that the bbKN spacetime

is also not algebraically special. Consequently, it is an expected result that one can not find a genuine Killing–Yano tensor on bbKN spacetime.

Now coupling the geometry to the Einstein equations of standard GR, the fact that the Einstein tensor is diagonal in an orthonormal basis proves that the stress-energy tensor for bbKN spacetime is Hawking–Ellis type I.[63–66] The radial NEC is violated in view of

$$\varepsilon + p_r = -\frac{\ell^2 \Delta}{8\pi\rho^6} \ , \tag{33}$$

where ε and p_r are energy density and radial pressure respectively.

The stress-energy tensor can be decomposed into gravitational and electromagnetic contributions (similar to bbRN spacetime in § 2.2):

$$\frac{1}{8\pi} G^{\hat{\mu}}{}_{\nu} = T^{\hat{\mu}}{}_{\nu} = [T_{\text{bbK}}]^{\hat{\mu}}{}_{\hat{\nu}} + [T_Q]^{\hat{\mu}}{}_{\hat{\nu}} \ , \tag{34}$$

where

$$[T_Q]^{\hat{\mu}}{}_{\hat{\nu}} = \frac{1}{8\pi} \frac{Q^2\left(\rho^2 - \ell^2\right)}{\rho^6} \left[\text{diag}\left(-1,-1,1,1\right) + \frac{2\ell^2}{\rho^2 - \ell^2} \text{diag}\left(1,0,0,0\right) \right] \ . \tag{35}$$

When comparing Eq. (35) directly with the charge-dependent contribution to the stress-energy from bbRN spacetime, one simply makes the substitutions

$$\frac{Q^2 r^2}{(r^2+\ell^2)^3} \longleftrightarrow \frac{Q^2\left(\rho^2-\ell^2\right)}{\rho^6} \qquad \text{and} \qquad \frac{2\ell^2}{r^2} \longleftrightarrow \frac{2\ell^2}{\rho^2-\ell^2}. \tag{36}$$

Structurally, the first term in Eq. (35) is of the form of the Maxwell stress-energy tensor. The second term appears to be structurally of the form of charged dust. However, the interpretation of the stress-energy is far less straightforward than the somewhat similar picture present in bbRN spacetime (§ 2.2). If one performs the substitution $r \to \sqrt{r^2+\ell^2}$ to the electromagnetic potential from Kerr–Newman spacetime, one finds

$$A_{\hat{\mu}} = -\frac{Q\sqrt{r^2+\ell^2}}{\sqrt{\rho^2|\Delta|}} \left(1,0,0,0\right) \ , \tag{37}$$

giving the following for the electromagnetic field strength tensor $F_{\hat{\mu}\hat{\nu}}$:

$$F_{\hat{t}\hat{r}} = -F_{\hat{r}\hat{t}} = -\frac{Q}{\rho^4}\sqrt{\frac{r^2}{r^2+\ell^2}}(r^2+\ell^2-a^2\cos^2\theta) \ , \tag{38}$$

$$F_{\hat{\theta}\hat{\phi}} = -F_{\hat{\phi}\hat{\theta}} = \frac{2aQ\cos\theta\sqrt{r^2+\ell^2}}{\rho^4} \ . \tag{39}$$

Immediately it is clear that $F_{[\hat{\mu}\hat{\nu},\hat{\sigma}]} = 0$. For the inhomogeneous Maxwell equation, defining $z = \sqrt{r^2+\ell^2}$, one has

$$\nabla^{\hat{\mu}} F_{\hat{\mu}\hat{\nu}} = J_{\hat{\nu}} = \frac{Q\ell^2}{\rho^7 z} \left(-\frac{\Delta\left(\rho^4 + 2\rho^2 z^2 - 4z^4\right)}{z^2\sqrt{|\Delta|}}, 0, 0, 2a\sin\theta\left(\rho^2 - 2z^2\right) \right). \tag{40}$$

Interpreting this as an effective electromagnetic source, one has $E_{\hat{r}} = F_{\hat{t}\hat{r}}$ and $B_{\hat{r}} = F_{\hat{\theta}\hat{\phi}}$, giving

$$[T_{\text{Maxwell}}]^{\hat{\mu}}{}_{\hat{\nu}} = \frac{E_{\hat{r}}^2 + B_{\hat{r}}^2}{8\pi} \text{diag}\,(-1,-1,1,1) \ , \tag{41}$$

with

$$E_{\hat{r}}^2 + B_{\hat{r}}^2 = \frac{Q^2}{\rho^4} \frac{r^2}{r^2+\ell^2} + \frac{4Q^2\ell^2 a^2 \cos^2\theta}{\rho^8} \ . \tag{42}$$

Comparison with Eq. (35) reveals that a more subtle interpretation of the electromagnetic stress-energy than standard Maxwell's is required. Presented herein are two alternatives.

Nonlinear electrodynamics One can decompose the electromagnetic stress-energy tensor as follows

$$[T_Q]^{\hat{\mu}}{}_{\hat{\nu}} = \mathcal{A}\,[T_{\text{Maxwell}}]^{\hat{\mu}}{}_{\hat{\nu}} + \Xi\, V^{\hat{\mu}} V_{\hat{\nu}} \ . \tag{43}$$

The $\Xi\, V^{\hat{\mu}} V_{\hat{\nu}}$ term is once again appropriate to interpret as charged dust. Comparing Eq's. (41) and (43) allows one to solve for the multiplicative factor $\mathcal{A}$:

$$\mathcal{A} = \frac{Q^2(\rho^2-\ell^2)}{\rho^6\,(E_{\hat{r}}^2 + B_{\hat{r}}^2)} = 1 - a^2\ell^2 \frac{\cos^2\theta[4(r^2+\ell^2)-\rho^2]}{\rho^4 r^2 + 4a^2\ell^2\cos^2\theta(r^2+\ell^2)} \ . \tag{44}$$

It is clear that in both the limit as $\ell \to 0$ (Kerr–Newman), and the limit as $a \to 0$ (bbRN), $\mathcal{A} \to 1$, recovering standard Maxwell's as expected. The fact that $\mathcal{A}$ is independent of the charge Q indicates it is appropriate to interpret this term through the lens of nonlinear electrodynamics (NLED), where generally one has $[T_{\text{NLED}}]^{\hat{\mu}}{}_{\hat{\nu}} \propto [T_{\text{Maxwell}}]^{\hat{\mu}}{}_{\hat{\nu}}$. NLED in the context of regular black holes is prevalent throughout the literature.[22,67–73] In the large distance limit one has

$$\mathcal{A} = 1 - \frac{3\ell^2 a^2 \cos^2\theta}{r^4} + \mathcal{O}(r^{-6}) \ , \tag{45}$$

indicating that standard Maxwell's + charged dust is a safe approximation as $r \to +\infty$. Conversely, for small r one has

$$\mathcal{A} = \frac{\ell^2 + a^2\cos\theta^2}{4\ell^2} + \mathcal{O}(r^2) \ , \tag{46}$$

which allows one to interpret a rescaling of the Maxwell stress-energy deep in the core of the black-bounce. This holistic picture is common in NLED.

Anisotropic fluid One can instead ask that

$$[T_Q]_{\hat{\mu}\hat{\nu}} - [T_{\text{Maxwell}}]_{\hat{\mu}\hat{\nu}} = \text{diag}\,(\varepsilon_f, -p_f, p_f, p_f) \,, \tag{47}$$

which would imply that (recall $z = \sqrt{r^2+\ell^2}$)

$$\varepsilon_f = \frac{Q^2\ell^2}{z^2\rho^8}\left(4z^4 - 7z^2\rho^2 + \rho^4\right), \quad p_f = \frac{Q^2\ell^2}{z^2\rho^8}\left(4z^4 - 5z^2\rho^2 + \rho^4\right). \tag{48}$$

Consequently, $[T_Q]^{\hat{\mu}}{}_{\hat{\nu}}$ can be interpreted as standard Maxwell's in the presence of an anisotropic fluid. The anisotropic fluid can then be written as

$$\varepsilon_f V_{\hat{\mu}} V_{\hat{\nu}} + \frac{p_f}{3}\left(g_{\hat{\mu}\hat{\nu}} + V_{\hat{\mu}} V_{\hat{\nu}}\right) + \pi_{\hat{\mu}\hat{\nu}} \ , \tag{49}$$

with $V^{\hat{\mu}} = (1,0,0,0)$ being the velocity of the fluid, and

$$\pi_{\hat{\mu}\hat{\nu}} = \frac{2p_f}{3}\mathrm{diag}\,(0,-2,1,1) \tag{50}$$

being the (traceless) anisotropic shear.[63]

5. Discussion

The family of black-bounce spacetimes contains many different candidate geometries which smoothly interpolate between regular black holes and traversable wormholes in standard GR, both with and without rotation, both with and without electrical charge, and both with and without dynamics. The high degree of mathematical tractability of the black-bounce spacetimes means that all models are amenable to the straightforward extraction of astrophysical observables, which *may* be falsifiable/verifiable by the experimental community.

Future research options abound. Given the knowledge of Klein–Gordon separability on bbKN spacetime, one should also test for Maxwell separability on the bbKN spacetime. The knowledge of Klein–Gordon separability already implies that bbKN is amenable to a standard spin zero quasi-normal modes analysis, and if Maxwell's equations also separate then this would extend to a spin one analysis. Further investigation into the spin two axial and polar modes may eventually lead to a measurable ringdown signal able to be captured by LIGO/VIRGO (or LISA) and compared with the analogous signal from Kerr-like mergers. Alternative lines of inquiry for future research could involve attempting to provide a physical mechanism for the bounce, or possibly discovering precisely which mathematical modifications to the Einstein equations would be required such that SV is the unique solution in static spherical symmetry in the new framework.

Acknowledgements

AS was supported by a Victoria University of Wellington PhD scholarship, and was also indirectly supported by the Marsden Fund, *via* a grant administered by the Royal Society of New Zealand.

AS would also like to acknowledge the following list of authors, all of whom have contributed in meaningful ways towards the development of the family of black-bounce spacetimes: Matt Visser, Prado Martín–Moruno, Francisco S. N. Lobo, Manuel E. Rodrigues, Marcos V. d. S. Silva, Stefano Liberati, Edgardo Franzin, and Jacopo Mazza.

References

1. See `https://www.ligo.caltech.edu/page/detection-companion-papers` for a collection of detection papers from LIGO.
   See also `https://pnp.ligo.org/ppcomm/Papers.html` for a complete list of publications from the LIGO Scientific Collaboration and Virgo Collaboration.
2. See, for example, wikipedia.org/List_of_gravitational_wave_observations for a list of current (May 2020) gravitational wave observations.
3. E. Barausse, E. Berti, T. Hertog, S. A. Hughes, P. Jetzer, P. Pani, T. P. Sotiriou, N. Tamanini, H. Witek and K. Yagi, *et al.*, "Prospects for Fundamental Physics with LISA", Gen. Rel. Grav. **52** (2020) no. 8, 81, doi:10.1007/s10714-020-02691-1, [arXiv:2001.09793 [gr-qc]].
4. E. F. Eiroa and C. M. Sendra, "Gravitational lensing by a regular black hole", Class. Quant. Grav. **28** (2011), 085008, doi:10.1088/0264-9381/28/8/085008, [arXiv:1011.2455 [gr-qc]].
5. A. Flachi and J. P. S. Lemos, "Quasinormal modes of regular black holes", Phys. Rev. D **87** (2013) no. 2, 024034, doi:10.1103/PhysRevD.87.024034, [arXiv:1211.6212 [gr-qc]].
6. A. Abdujabbarov, M. Amir, B. Ahmedov and S. G. Ghosh, "Shadow of rotating regular black holes", Phys. Rev. D **93** (2016) no. 10, 104004, doi:10.1103/PhysRevD.93.104004, [arXiv:1604.03809 [gr-qc]].
7. R. Carballo-Rubio, F. Di Filippo, S. Liberati, C. Pacilio and M. Visser, "On the viability of regular black holes", JHEP **07** (2018), 023, doi:10.1007/JHEP07(2018)023, [arXiv:1805.02675 [gr-qc]].
8. R. Carballo-Rubio, F. Di Filippo, S. Liberati and M. Visser, "Opening the Pandora's box at the core of black holes", Class. Quant. Grav. **37** (2020) no. 14, 14, doi:10.1088/1361-6382/ab8141, [arXiv:1908.03261 [gr-qc]].
9. R. Carballo-Rubio, F. Di Filippo, S. Liberati and M. Visser, "Geodesically complete black holes", Phys. Rev. D **101** (2020), 084047, doi:10.1103/PhysRevD.101.084047, [arXiv:1911.11200 [gr-qc]].
10. T. Berry, A. Simpson and M. Visser, "Photon spheres, ISCOs, and OSCOs: Astrophysical observables for regular black holes with asymptotically Minkowski cores", Universe **7** (2020) no. 1, 2, doi:10.3390/universe7010002, [arXiv:2008.13308 [gr-qc]].
11. R. Carballo-Rubio, F. Di Filippo and S. Liberati, "Hearts of Darkness: the inside out probing of black holes", IJMPD (in press), [arXiv:2106.01530 [gr-qc]].
12. K. A. Bronnikov, R. A. Konoplya and T. D. Pappas, "General parametrization of wormhole spacetimes and its application to shadows and quasinormal modes", Phys. Rev. D **103** (2021) no. 12, 124062, doi:10.1103/PhysRevD.103.124062, [arXiv:2102.10679 [gr-qc]].
13. M. S. Churilova, R. A. Konoplya, Z. Stuchlik and A. Zhidenko, "Wormholes without exotic matter: quasinormal modes, echoes and shadows", [arXiv:2107.05977 [gr-qc]].
14. A. Simpson, "Ringing of the regular black hole with asymptotically Minkowski core", [arXiv:2109.11878 [gr-qc]].
15. A. Simpson and M. Visser, "Black-bounce to traversable wormhole", JCAP **02** (2019), 042, doi:10.1088/1475-7516/2019/02/042 [arXiv:1812.07114 [gr-qc]].
16. F. S. N. Lobo, M. E. Rodrigues, M. V. d. S. Silva, A. Simpson and M. Visser, "Novel black-bounce spacetimes: wormholes, regularity, energy conditions, and causal structure", Phys. Rev. D **103** (2021) no. 8, 084052, doi:10.1103/PhysRevD.103.084052, [arXiv:2009.12057 [gr-qc]].

17. A. Simpson, P. Martín-Moruno and M. Visser, "Vaidya spacetimes, black-bounces, and traversable wormholes", Class. Quant. Grav. **36** (2019) no. 14, 145007, doi:10.1088/1361-6382/ab28a5, [arXiv:1902.04232 [gr-qc]].
18. F. S. N. Lobo, A. Simpson and M. Visser, "Dynamic thin-shell black-bounce traversable wormholes", Phys. Rev. D **101** (2020) no. 12, 124035, doi:10.1103/PhysRevD.101.124035, [arXiv:2003.09419 [gr-qc]].
19. J. Mazza, E. Franzin and S. Liberati, "A novel family of rotating black hole mimickers", JCAP **04** (2021), 082, doi:10.1088/1475-7516/2021/04/082, [arXiv:2102.01105 [gr-qc]].
20. E. Franzin, S. Liberati, J. Mazza, A. Simpson and M. Visser, "Charged black-bounce spacetimes", JCAP **07** (2021), 036, doi:10.1088/1475-7516/2021/07/036, [arXiv:2104.11376 [gr-qc]].
21. J. M. Bardeen, "Non-singular general relativistic gravitational collapse", Abstracts of the 5th international conference on gravitation and the theory of relativity (GR5), eds. V. A. Fock *et al.* (Tbilisi University Press, Tblisi, Georgia, former USSR, 1968), pages 174—175.
22. E. Ayon-Beato and A. Garcia, "The Bardeen model as a nonlinear magnetic monopole", Phys. Lett. B **493** (2000), 149–152, doi:10.1016/S0370-2693(00)01125-4, [arXiv:gr-qc/0009077 [gr-qc]].
23. S. A. Hayward, "Formation and evaporation of regular black holes", Phys. Rev. Lett. **96** (2006), 031103, doi:10.1103/PhysRevLett.96.031103, [arXiv:gr-qc/0506126 [gr-qc]].
24. C. Bambi and L. Modesto, "Rotating regular black holes", Phys. Lett. B **721** (2013), 329–334, doi:10.1016/j.physletb.2013.03.025, [arXiv:1302.6075 [gr-qc]].
25. V. P. Frolov, "Information loss problem and a 'black hole' model with a closed apparent horizon", JHEP **05** (2014), 049, doi:10.1007/JHEP05(2014)049, [arXiv:1402.5446 [hep-th]].
26. J. C. S. Neves and A. Saa, "Regular rotating black holes and the weak energy condition", Phys. Lett. B **734** (2014), 44–48, doi:10.1016/j.physletb.2014.05.026, [arXiv:1402.2694 [gr-qc]].
27. Z. Y. Fan and X. Wang, "Construction of Regular Black Holes in General Relativity", Phys. Rev. D **94** (2016) no. 12, 124027, doi:10.1103/PhysRevD.94.124027, [arXiv:1610.02636 [gr-qc]].
28. M. S. Morris and K. S. Thorne, "Wormholes in space-time and their use for interstellar travel: A tool for teaching general relativity", Am. J. Phys. **56** (1988), 395–412, doi:10.1119/1.15620.
29. M. S. Morris, K. S. Thorne and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition", Phys. Rev. Lett. **61** (1988), 1446–1449, doi:10.1103/PhysRevLett.61.1446.
30. M. Visser, "Traversable wormholes: Some simple examples", Phys. Rev. D **39** (1989), 3182–3184, doi:10.1103/PhysRevD.39.3182, [arXiv:0809.0907 [gr-qc]].
31. M. Visser, "Traversable wormholes from surgically modified Schwarzschild spacetimes", Nucl. Phys. B **328** (1989), 203–212, doi:10.1016/0550-3213(89)90100-4, [arXiv:0809.0927 [gr-qc]].
32. See inspirehep.net for a full list of pre-published and published papers citing reference 15.
33. K. A. Bronnikov and R. A. Konoplya, "Echoes in brane worlds: ringing at a black hole–wormhole transition", Phys. Rev. D **101** (2020) no. 6, 064004, doi:10.1103/PhysRevD.101.064004, [arXiv:1912.05315 [gr-qc]].
34. M. S. Churilova and Z. Stuchlik, "Ringing of the regular black-hole/wormhole transition", Class. Quant. Grav. **37** (2020) no. 7, 075014, doi:10.1088/1361-6382/ab7717, [arXiv:1911.11823 [gr-qc]].

35. S. U. Islam, J. Kumar and S. G. Ghosh, "Strong gravitational lensing by rotating Simpson–Visser black holes", [arXiv:2104.00696 [gr-qc]].
36. N. Tsukamoto, "Gravitational lensing by two photon spheres in a black-bounce spacetime in strong deflection limits", Phys. Rev. D **104** (2021) no. 6, 064022, doi:10.1103/PhysRevD.104.064022, [arXiv:2105.14336 [gr-qc]].
37. M. Guerrero, G. J. Olmo, D. Rubiera-Garcia and D. S. C. Gómez, "Shadows and optical appearance of black bounces illuminated by a thin accretion disk", JCAP **08** (2021), 036, doi:10.1088/1475-7516/2021/08/036, [arXiv:2105.15073 [gr-qc]].
38. H. C. D. Lima, Junior., L. C. B. Crispino, P. V. P. Cunha and C. A. R. Herdeiro, "Can different black holes cast the same shadow?", Phys. Rev. D **103** (2021) no. 8, 084040, doi:10.1103/PhysRevD.103.084040, [arXiv:2102.07034 [gr-qc]].
39. N. Tsukamoto, "Gravitational lensing in the Simpson-Visser black-bounce spacetime in a strong deflection limit", Phys. Rev. D **103** (2021) no. 2, 024033, doi:10.1103/PhysRevD.103.024033, [arXiv:2011.03932 [gr-qc]].
40. J. R. Nascimento, A. Y. Petrov, P. J. Porfirio and A. R. Soares, "Gravitational lensing in black-bounce spacetimes", Phys. Rev. D **102** (2020) no. 4, 044021, doi:10.1103/PhysRevD.102.044021, [arXiv:2005.13096 [gr-qc]].
41. T. Y. Zhou and Y. Xie, "Precessing and periodic motions around a black-bounce/traversable wormhole", Eur. Phys. J. C **80** (2020) no. 11, 1070, doi:10.1140/epjc/s10052-020-08661-w.
42. M. S. R. Delgaty and K. Lake, "Physical acceptability of isolated, static, spherically symmetric, perfect fluid solutions of Einstein's equations", Comput. Phys. Commun. **115** (1998), 395–415, doi:10.1016/S0010-4655(98)00130-1, [arXiv:gr-qc/9809013 [gr-qc]].
43. M. R. Finch and J. E. F. Skea, "A review of the relativistic static fluid sphere", 1998, unpublished.
44. P. Boonserm and M. Visser, "Buchdahl-like transformations for perfect fluid spheres", Int. J. Mod. Phys. D **17** (2008), 135–163, doi:10.1142/S0218271808011912, [arXiv:0707.0146 [gr-qc]].
45. P. Boonserm and M. Visser, "Buchdahl-like transformations in general relativity", Thai Journal of Mathematics **5 #2** (2007) 209–223.
46. İ. Semiz, "On the (non)genericity of the Kiselev spacetime", doi:10.1088/2633-1357/aba1f5, [arXiv:2001.06310 [gr-qc]].
47. E. Babichev, V. Dokuchaev and Y. Eroshenko, "Black hole mass decreasing due to phantom energy accretion", Phys. Rev. Lett. **93** (2004), 021102, doi:10.1103/PhysRevLett.93.021102, [arXiv:gr-qc/0402089 [gr-qc]].
48. E. Babichev, V. Dokuchaev and Y. Eroshenko, "Dark energy cosmology with generalized linear equation of state", Class. Quant. Grav. **22** (2005), 143–154, doi:10.1088/0264-9381/22/1/010, [arXiv:astro-ph/0407190 [astro-ph]].
49. P. Martín-Moruno, J. A. J. Madrid and P. F. González-Díaz, "Will black holes eventually engulf the universe?", Phys. Lett. B **640** (2006), 117–120, doi:10.1016/j.physletb.2006.07.067, [arXiv:astro-ph/0603761 [astro-ph]].
50. P. F. González-Díaz and P. Martín-Moruno, "Wormholes in the accelerating universe", doi:10.1142/9789812834300_0358, [arXiv:0704.1731 [astro-ph]].
51. P. Martín-Moruno, "On the formalism of dark energy accretion onto black- and worm-holes", Phys. Lett. B **659** (2008), 40–44, doi:10.1016/j.physletb.2007.10.083, [arXiv:0709.4410 [astro-ph]].
52. J. A. J. Madrid and P. Martín-Moruno, "On accretion of dark energy onto black- and worm-holes", (2010), [arXiv:1004.1428 [astro-ph.CO]].

53. F. S. N. Lobo, J. Martinez-Asencio, G. J. Olmo and D. Rubiera-Garcia, "Planck scale physics and topology change through an exactly solvable model", Phys. Lett. B **731** (2014), 163–167, doi:10.1016/j.physletb.2014.02.038, [arXiv:1311.5712 [hep-th]].
54. S. Chakrabarti and S. Kar, "Wormhole geometry from gravitational collapse", Phys. Rev. D **104** (2021) no. 2, 024071, doi:10.1103/PhysRevD.104.024071, [arXiv:2106.14761 [gr-qc]].
55. W. Israel, "Singular hypersurfaces and thin shells in general relativity", Nuovo Cim. B **44S10** (1966), 1 [erratum: Nuovo Cim. B **48** (1967), 463], doi:10.1007/BF02710419.
56. M. Visser, "Traversable wormholes from surgically modified Schwarzschild spacetimes", Nucl. Phys. B **328** (1989), 203–212, doi:10.1016/0550-3213(89)90100-4, [arXiv:0809.0927 [gr-qc]].
57. E. Poisson and M. Visser, "Thin shell wormholes: Linearization stability", Phys. Rev. D **52** (1995), 7318–7321, doi:10.1103/PhysRevD.52.7318, [arXiv:gr-qc/9506083 [gr-qc]].
58. M. Visser, "Lorentzian wormholes: From Einstein to Hawking".
59. J. Baines, T. Berry, A. Simpson and M. Visser, "Killing tensor and Carter constant for Painleve-Gullstrand form of Lense-Thirring spacetime", [arXiv:2110.01814 [gr-qc]].
60. E. Giorgi, "The Carter tensor and the physical-space analysis in perturbations of Kerr-Newman spacetime", [arXiv:2105.14379 [gr-qc]].
61. S. Benenti, C. Chanu, and G. Rastelli, "Remarks on the connection between the additive separation of the Hamilton–Jacobi equation and the multiplicative separation of the Schrödinger equation. I. The completeness and Robertson conditions", Journal of Mathematical Physics, **43** (2002) no. 11, 5183–5222, American Institute of Physics, doi:10.1063/1.1506180.
62. V. Frolov, P. Krtous and D. Kubiznak, "Black holes, hidden symmetries, and complete integrability", Living Rev. Rel. **20** (2017) no. 1, 6, doi:10.1007/s41114-017-0009-9, [arXiv:1705.05482 [gr-qc]].
63. S. W. Hawking and G. F. R. Ellis, "The Large Scale Structure of Space-Time", doi:10.1017/CBO9780511524646.
64. P. Martín-Moruno and M. Visser, "Generalized Rainich conditions, generalized stress-energy conditions, and the Hawking-Ellis classification", Class. Quant. Grav. **34** (2017) no. 22, 225014, doi:10.1088/1361-6382/aa9039, [arXiv:1707.04172 [gr-qc]].
65. P. Martín-Moruno and M. Visser, "Essential core of the Hawking–Ellis types", Class. Quant. Grav. **35** (2018) no. 12, 125003, doi:10.1088/1361-6382/aac147, [arXiv:1802.00865 [gr-qc]].
66. P. Martín-Moruno and M. Visser, "Hawking-Ellis classification of stress-energy tensors: Test fields versus backreaction", Phys. Rev. D **103** (2021) no. 12, 124003, doi:10.1103/PhysRevD.103.124003, [arXiv:2102.13551 [gr-qc]].
67. K. A. Bronnikov, "Regular magnetic black holes and monopoles from nonlinear electrodynamics", Phys. Rev. D **63** (2001), 044005, doi:10.1103/PhysRevD.63.044005, [arXiv:gr-qc/0006014 [gr-qc]].
68. E. Ayon-Beato and A. Garcia, "Four parametric regular black hole solution", Gen. Rel. Grav. **37** (2005), 635, doi:10.1007/s10714-005-0050-y, [arXiv:hep-th/0403229 [hep-th]].
69. F. S. N. Lobo and A. V. B. Arellano, "Gravastars supported by nonlinear electrodynamics", Class. Quant. Grav. **24** (2007), 1069–1088, doi:10.1088/0264-9381/24/5/004, [arXiv:gr-qc/0611083 [gr-qc]].
70. S. V. Bolokhov, K. A. Bronnikov and M. V. Skvortsova, "Magnetic black universes and wormholes with a phantom scalar", Class. Quant. Grav. **29** (2012), 245006, doi:10.1088/0264-9381/29/24/245006, [arXiv:1208.4619 [gr-qc]].

71. L. Balart and E. C. Vagenas, "Regular black holes with a nonlinear electrodynamics source", Phys. Rev. D **90** (2014) no. 12, 124045, doi:10.1103/PhysRevD.90.124045, [arXiv:1408.0306 [gr-qc]].
72. K. A. Bronnikov, "Nonlinear electrodynamics, regular black holes and wormholes", Int. J. Mod. Phys. D **27** (2018) no. 06, 1841005, doi:10.1142/S0218271818410055, [arXiv:1711.00087 [gr-qc]].
73. M. E. Rodrigues and M. V. d. Silva, "Bardeen Regular Black Hole With an Electric Source", JCAP **06** (2018), 025, doi:10.1088/1475-7516/2018/06/025, [arXiv:1802.05095 [gr-qc]].

Tractor beams, pressor beams, and stressor beams within the context of general relativity

Matt Visser* and Jessica Santiago†

School of Mathematics and Statistics, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
** E-mail: matt.visser@sms.vuw.ac.nz*
† E-mail: jessica.santiago@sms.vuw.ac.nz

Sebastian Schuster

Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Prague 8, Czech Republic
E-mail: sebastian.schuster@utf.mff.cuni.cz

Both traversable wormholes and warp drives, concepts originally developed within the context of science fiction, have now (for some 30 odd years) been studied, debated, and carefully analyzed within the framework of general relativity. An overarching theme of the general relativistic analysis is unavoidable violations of the classical point-wise energy conditions. Another science fiction trope, now over 80 years old, is the tractor beam and/or pressor beam. We shall discuss how to formulate both tractor beams and/or pressor beams, and a variant to be called a stressor beam, within the context of reverse engineering the spacetime metric. (While such reverse engineering is certainly well beyond our civilization's current capabilities, we shall be more interested in asking what an arbitrarily advanced civilization might be able to accomplish.) We shall see that tractor beams and/or pressor beams can be formulated by suitably modifying the notion of warp drives, and that, as for wormholes and warp drives, violations of the classical point-wise energy conditions are utterly unavoidable.

Keywords: Traversable wormholes; warp drives; tractor/pressor/stressor beams

1. Introduction

Traversable wormholes have now been extensively studied for some 33 years.[1–19] The general theoretical framework has been to precisely define, and then reverse engineer, a physically interesting spacetime metric, using the Einstein equations to determine what matter sources some arbitrarily advanced civilization might need to employ in order to generate such interesting spacetime geometries. Perhaps the key feature of reverse engineering is that adopting this point of view reduces the problem of solving the Einstein equations to a relatively simple exercise in differentiation, instead of the much harder problem of integrating the underlying nonlinear PDEs.

A key feature of traversable wormhole spacetimes, at least within the usual context of standard general relativity, is the utterly unavoidable violation of the classical energy conditions.[1–19]

Even if one moves beyond the framework of standard general relativity there are unavoidable purely geometric constraints, violations of both the null convergence condition and also violations of the timelike convergence condition, that place very strong constraints on any "effective" stress-energy tensor that one might use as a source to generate the spacetime.[9]

Similarly, "warp drives" have now been extensively studied for some 27 years.[20–31] As is the case for traversable wormholes once one defines a suitable class of warp field spacetimes it is easy to see that reverse engineering the spacetime metric, at least within the context of standard general relativity, quite unavoidably leads to significant violations of the classical energy conditions.[20–31] Again, even when stepping outside the framework of standard general relativity there are still unavoidable but purely geometric constraints, violations of the null convergence condition and violations of the timelike convergence condition, that place very strong constraints on any "effective" stress-energy tensor.[20–31] (There have recently been several misguided counterclaims regarding these issues which we shall address more fully below, see references 32, 33, 34. For somewhat related but distinct analyses, with somewhat different flaws, see also the discussion in references 35, 36, 37, 38, and in reference 39.)

Much more recently, the last 6 months have seen the current authors reformulate yet another science fiction trope, tractor beams, pressor beams, and stressor beams, within the framework of general relativity.[40] Again we resort to reverse engineering, and again the classical energy conditions are violated.[40] Even if one steps outside the framework of standard general relativity there are still purely geometric constraints, violations of the null convergence condition, and also violations of the timelike convergence condition, that make the underlying physics "exotic" in some purely technical sense.[40]

Of course, in all of these cases one must first precisely define what one means by a traversable wormhole, or a warp drive, or a tractor/pressor/stressor beam. One also needs to carefully assess the extent to which the energy conditions (or more geometrically, the null and timelike convergence conditions) are to be taken as truly fundamental physics.[9, 31, 40, 41]

2. Energy conditions — generalities

It should be emphasized that the energy conditions are not fundamental physics.[41] The energy conditions are instead warning flags: They tell you "here be unusual physics". Violations of the energy conditions are not an absolute prohibition; but they are an indication that one should think very carefully about the underlying physics.

For instance:

- Observational cosmology (either in the current epoch of the accelerating universe, or during the epoch of cosmological inflation) violates the strong energy condition (SEC).[42,43]
- Speculative theoretical cosmologies (asymptotically AdS spacetimes) with a negative cosmological constant violate the weak energy condition (WEC).
- Gravitational vacuum polarization (in the test field limit) can violate all the energy conditions, null, weak, strong, and dominant (NEC, WEC, SEC, and DEC).[44–48]
- Hawking radiation violates all the energy conditions (NEC, WEC, SEC, and DEC).[9]
- Quantum scale anomalies (trace anomalies) violate even the averaged null energy condition (ANEC).[49]

But these known energy condition violating scenarios do not necessarily imply that it is possible to concentrate energy condition violating physics in a sufficiently small region to enable "interesting" effects — such as wormholes, warp drives, and the tractor/pressor/stressor beams.[9,13,16] However, in view of these known energy condition violating scenarios, neither would the necessity of energy condition violations necessarily prohibit "interesting" physics.[50–69]
We emphasise that to verify that the energy conditions are *satisfied*:

- One needs to check *all* relevant "observers" — both *all* physical timelike observers, and in the null limit, *all* null tangent vectors.
- One needs at least *some* information concerning *all* of the stress-energy components.

In contrast, to demonstrate the energy conditions are *violated*:

- It is sufficient to check that *one* timelike observer, (or *one* null curve), violates the energy conditions.
- It is sufficient to check that *some* combination of stress-energy components violates the energy conditions.

3. Generic warp metric and tractor/pressor/stressor metric

A sufficiently generic spacetime metric, (suitable for investigating both warp drives and tractor/pressor/stressor beams), is to consider:[20–31]

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \delta_{ij}\,\left(\mathrm{d}x^i - v^i(t,\vec{x})\,\mathrm{d}t\right)\,\left(\mathrm{d}x^j - v^j(t,\vec{x})\,\mathrm{d}t\right). \tag{1}$$

This class of metrics has the properties that it is:

- Spatially flat.
- Unit lapse.
- The gradients of the flow vector $v^i(t,\vec{x})$ must satisfy suitable asymptotic fall-off conditions.
  - (Otherwise you do not have a localized warp-field.)
  - (You do not want to accelerate the entire universe.)

  In particular, the spacetime must be asymptotically flat.
- Three special cases are commonly considered:
  - Alcubierre: $\vec{v}(t,\vec{x}) = v(t,z)\,\hat{z}$.
  - Zero-expansion: $\boldsymbol{\nabla}\cdot\vec{v}(t,\vec{x}) = 0$.
  - Zero-vorticity: $\boldsymbol{\nabla}\times\vec{v}(t,\vec{x}) = 0$.
- For tractor/pressor/stressor beam spacetimes one will add extra conditions, such as cylindrical symmetry, and simplify the form of the flow vector $\vec{v}(t,\vec{x})$ by introducing suitable "envelope" and "profile" functions.
- Further generalizations are in principle possible, but very quickly become calculationally intractable.

To start specific computations it is convenient to first pick a co-tetrad:[31,40]

- Take the timelike leg to be:

$$(e^{\hat{0}})_a = (1;0,0,0)_a = n_a. \tag{2}$$

- Take the spatial co-triad to be:

$$(e^{\hat{1}})_a = (-v_x;1,0,0)_a, \quad (e^{\hat{2}})_a = (-v_y;0,1,0)_a, \quad (e^{\hat{3}})_a = (-v_z;0,0,1)_a. \tag{3}$$

This *choice* serves to keep computations more-or-less tractable.

The corresponding tetrad is then fixed:[31,40]

- The timelike leg is:

$$(e_{\hat{0}})^a = (1;v^x,v^y,v^z)^a = n^a. \tag{4}$$

- The spatial triad is:

$$(e_{\hat{1}})^a = (0;1,0,0)^a, \quad (e_{\hat{2}})^a = (0;0,1,0)^a, \quad (e_{\hat{3}})^a = (0;0,0,1)^a. \tag{5}$$

With this choice of tetrad the orthonormal components of the extrinsic curvature are now particularly simple:[31,40]

$$K_{\hat{a}\hat{b}} = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & K_{ij} \end{array}\right], \tag{6}$$

where

$$K_{ij} = v_{(i,j)}. \tag{7}$$

(Effectively one can just quietly drop the hats from covariant spatial components of any tensor in this tetrad basis.[31,40])

Brute force computation now yields two easy results, and one somewhat messier result.[31,40]

- Easy (Gauss–Coddazzi):

$$G_{nn} = \frac{1}{2}\left(K^2 - \mathrm{tr}\big(K^2\big)\right), \tag{8}$$

$$G_{ni} = K_{ij,j} - K_{,i}. \tag{9}$$

- Somewhat messier (specialization of ADM formalism):

$$G_{ij} = \mathcal{L}_n K_{ij} + K K_{ij} - 2(K^2)_{ij} - \left(\mathcal{L}_n K + \frac{1}{2}K^2 + \frac{1}{2}\,\mathrm{tr}\big(K^2\big)\right)\delta_{ij}. \tag{10}$$

In order to claim that the energy conditions are satisfied (or claim that the null and timelike convergence conditions are satisfied) one would need to compute and analyze *all* the components of the Einstein tensor. (This is one specific place, there are others, at which the claims made in references 32, 33, 34 simply fail. See reference 31 for an extensive discussion of these issues. The analyses of references 35, 36, 37, 38, and reference 39 exhibit somewhat different pathologies; once the Einstein equations are fully solved, many of those specific examples are actually Riemann flat.)

Standard general relativity (the Einstein equations) now yields the stress-energy:

- Eulerian energy density:

$$\rho = \frac{G_{nn}}{8\pi} = \frac{1}{16\pi}\left\{\boldsymbol{\nabla}\cdot\{\vec{v}\,K - (\vec{v}\cdot\boldsymbol{\nabla})\vec{v}\} - \frac{1}{2}\,(\vec{\omega}\cdot\vec{\omega})\right\}. \tag{11}$$

  Note: (comoving density) = (3-divergence) + (negative semidefinite term).

- Eulerian flux:

$$f_i = \frac{G_{ni}}{8\pi} = \frac{1}{8\pi}(K_{ij,j} - K_{,i}) = \frac{1}{16\pi}\,(\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\vec{v}))_i\,. \tag{12}$$

  Note: curl of a curl. (So this can sometimes be set to zero.)

- Eulerian spatial stresses

$$T_{ij} = \frac{G_{ij}}{8\pi} = \frac{1}{8\pi}\left(\mathcal{L}_n K_{ij} + K K_{ij} - 2(K^2)_{ij} - \left(\mathcal{L}_n K + \frac{1}{2}K^2 + \frac{1}{2}\,\mathrm{tr}\big(K^2\big)\right)\delta_{ij}\right). \tag{13}$$

The Eulerian density and Eulerian flux are easy, the Eulerian spatial stresses are more complicated. A key quantity of interest is the average pressure:

$$\bar{p} = \frac{1}{3}\,T_{ij}\,\delta^{ij} = \frac{1}{24\pi}\left(-2\mathcal{L}_n K - \frac{1}{2}K^2 - \frac{3}{2}\,\mathrm{tr}\big(K^2\big)\right). \tag{14}$$

This can also be written as:

$$\bar{p} = \frac{1}{3}\, T_{ij}\, \delta^{ij} = -\frac{1}{24\pi}\left(2\nabla_a(Kn^a) - \frac{3}{2}K^2 + \frac{3}{2}\,\mathrm{tr}\big(K^2\big)\right). \tag{15}$$

Equivalently:

$$\bar{p} = \rho - \frac{1}{12\pi}\,\nabla_a(Kn^a). \tag{16}$$

Furthermore, consider the two combinations:

$$\rho + 3\bar{p} = -\frac{1}{4\pi}\left(\mathcal{L}_n K + \mathrm{tr}\big(K^2\big)\right), \tag{17}$$

$$\rho + \bar{p} = \frac{1}{24\pi}\left(-2\mathcal{L}_n K + K^2 - 3\,\mathrm{tr}\big(K^2\big)\right). \tag{18}$$

These two quantities are relevant to investigating violations of the SEC and NEC, respectively.[31]

4. Forces: Tractor/pressor/stressor beams

For tractor/pressor/stressor beam spacetimes we now make a specific, concrete choice for the form of the generic warp metric. Let us consider a beam pointed in the $\hat{z}$ direction, and specialize the flow vector to be of the form:

$$\begin{aligned} v_x(t,x,y,z) &= k(t,z)\, x\, h(x^2+y^2),\\ v_y(t,x,y,z) &= k(t,z)\, y\, h(x^2+y^2),\\ v_z(t,x,y,z) &= v(t,z)\, f(x^2+y^2). \end{aligned}$$

Note the introduction of two envelope functions, $k(t,z)$ and $v(t,z)$; and two profile functions, $h(x^2+y^2)$ and $f(x^2+y^2)$. Note that for simplicity we have also imposed cylindrical symmetry (in Cartesian coordinates).

Now consider the forces such a tractor/pressor/stressor beam spacetime would exert on a target with cross-section $\mathbb{S}$ that is at time t placed at position z.

$$F(t,z) = \pm \int_{\mathbb{S}} T_{zz}(t,x,y,z)\,\mathrm{d}x\,\mathrm{d}y. \tag{19}$$

There are two special cases to consider depending on whether the beam is wide or narrow compared to the cross-sectional area $A(\mathbb{S})$ of the tatget.

- Narrow beam:

$$F(t,z) = \pm \int_{\mathbb{R}^2} T_{zz}(t,x,y,z)\,\mathrm{d}x\,\mathrm{d}y. \tag{20}$$

- Wide beam:

$$F(t,z) = \pm T_{zz}(t,0,0,z)\, A(\mathbb{S}). \tag{21}$$

Let us now calculate, setting $u = x^2 + y^2$ for conciseness.

- Narrow beam:

$$F(t,z) = -\frac{1}{2} v(t,z)^2 \int_0^\infty u[f'(u)]^2 \mathrm{d}u + \frac{1}{8}\int_0^\infty u\left[v(t,z) f'(u) + \frac{1}{2}\partial_z k(t,z) h(u)\right]^2 \mathrm{d}u.$$

  This is of indefinite sign, depending delicately on the envelope functions and profile functions, potentially allowing either tractor/pressor behaviour.

- Wide beam:

$$F(t,z) = -\frac{h(0)}{8\pi}\left\{2f(0)\; v(t,z)\partial_z k(t,z) + 2\;\partial_t k(t,z) + 3h(0)\; k(t,z)^2\right\} A(\mathbb{S}). \tag{22}$$

  This is of indefinite sign, depending delicately on the envelope functions, potentially allowing either tractor/pressor behaviour.

There are very many special cases to study, depending on one's choice of envelope and profile functions. See reference 40 for extensive details and discussion.

Generic Natário	**Modified Alcubierre**	**Zero Expansion**	**Zero Vorticity**
envelope $k(t,z)$	0	$-\partial_z v$	Φ
envelope $v(t,z)$	v	v	$\partial_z \Phi$
profile $h(u)$	0	h	$2f'$
profile $f(u)$	f	$2(h + uh')$	f

One can easily tune the resulting forces in a wide variety of ways. In particular we can define a *stressor* beam to be one where the *net* force is zero, but forces alternate between attractive and repulsive as one moves across the cross section of the target, thereby stressing the target in a (in principle) controllable manner.

5. Energy conditions — specifics

In this context, all of the classical point-wise energy conditions are violated. The arguments are fully generic to tractor/pressor/stressor beam spacetimes, and, with suitable subsidiary conditions, apply also to warp drive configurations. (In the warp drive context, many special cases have been well known for over 20 years.)

Generically for either tractor/pressor/stressor or warp drive configurations we have:

$$\rho = (\text{3-divergence}) - \frac{1}{32\pi}\,(\vec{\omega}\cdot\vec{\omega}). \tag{23}$$

Then for all tractor/pressor/stressor beams (where by construction all fields fall off sufficiently rapidly at spatial infinity):

$$\int \rho \, \mathrm{d}^3x = -\frac{1}{32\pi} \int (\vec{\omega} \cdot \vec{\omega}) \, \mathrm{d}^3x \leq 0. \tag{24}$$

This kills off the WEC for all tractor/pressor/stressor beams. For warp drives one must sometimes be more careful, see the Appendix, and to prove or disprove the WEC one must check both the Eulerian energy density ρ *and* the NEC ($\rho + \bar{p} \geq 0$).

Generic to either tractor/pressor/stressor or warp drive configurations we have:

$$\bar{p} = \rho - \frac{1}{12\pi} \nabla_a (K n^a). \tag{25}$$

$$\int \bar{p} \, \mathrm{d}^3x = \int \rho \, \mathrm{d}^3x - \frac{1}{12\pi} \int \partial_t K \, \mathrm{d}^3x. \tag{26}$$

$$\int (\rho + \bar{p}) \, \mathrm{d}^3x = -\frac{1}{16\pi} \int (\vec{\omega} \cdot \vec{\omega}) \, \mathrm{d}^3x - \frac{1}{12\pi} \int \partial_t K \, \mathrm{d}^3x. \tag{27}$$

For any tractor/pressor/stressor/beam configuration (and for any suitably localized warp bubble):

$$\int (\rho + \bar{p}) \, \mathrm{d}^3x = -\frac{1}{16\pi} \int (\vec{\omega} \cdot \vec{\omega}) \, \mathrm{d}^3x - \frac{1}{12\pi} \partial_t \int K \, \mathrm{d}^3x. \tag{28}$$

Furthermore, for any tractor/pressor/stressor/beam configuration (and for any suitably localized warp bubble) we also have:

$$\int K \, \mathrm{d}^3x = 0. \tag{29}$$

Therefore, for any tractor/pressor/stressor/beam configuration (and for any suitably localized warp bubble) we conclude:

$$\int (\rho + \bar{p}) \, \mathrm{d}^4x = -\frac{1}{16\pi} (\vec{\omega} \cdot \vec{\omega}) \, \mathrm{d}^4x \leq 0. \tag{30}$$

This kills off the NEC.

That is, all of the classical point-wise energy conditions are violated. This certainly holds for any tractor/pressor/stressor/beam configuration, and for any suitably localized warp bubble, will also work for warp drive spacetimes. An example of a more delocalized warp bubble is considered in the Appendix.

6. Conclusions

- Traversable wormholes, warp drives, and tractor/pressor/stressor beams can all usefully be "reverse engineered", debated, analyzed, and studied within the framework of standard general relativity.
- Reverse engineering a specified interesting spacetime reduces the task of solving the Einstein equations to a relatively straightforward exercise of differentiation, rather than the more complicated task of integration.

- Traversable wormholes, warp drives, and tractor/pressor/stressor beams all violate all of the standard point-wise energy conditions.
- Even in various modified theories of gravity, wormholes, warp drives, and tractor/pressor/stressor beams violate the null convergence condition and timelike convergence condition.
- Violation of the classical energy conditions is not an absolute prohibition on "interesting" physics — it is merely an invitation to think very carefully about the underlying physics.

Appendix A. Falloff conditions at spatial infinity

For tractor/pressor/stressor beams suitable falloff conditions at spatial infinity are inherent to the definition, and utterly automatic. Similarly, for Alcubierre warp drives and zero-expansion warp drives suitable falloff conditions at spatial infinity are also inherent to the definition and utterly automatic. For zero vorticity warp drives one must be just a little more careful.

As an example, consider the general framework

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \delta_{ij}\, \left(\mathrm{d}x^i - v^i(t,\vec{x})\,\mathrm{d}t\right)\, \left(\mathrm{d}x^j - v^j(t,\vec{x})\,\mathrm{d}t\right), \tag{A.1}$$

and for one's warp field take the specific case

$$v^i(t,\vec{x}) = -\sqrt{\frac{2m}{r}}\,\hat{r}^i + v^i_\infty(t). \tag{A.2}$$

This corresponds to a warp bubble, centred at the origin, in a coordinate system that is moving with the warp bubble. The presence of the warp velocity term, $v^i_\infty(t)$, is essential, since otherwise the warp bubble is not moving with respect to spatial infinity, (the "fixed stars"), which would make the warp field trivial.

For the specific warp field (A.2) it is easy to calculate

$$K = \boldsymbol{\nabla}\cdot\vec{v} = -\frac{3}{2}\frac{\sqrt{2m}}{r^{3/2}}, \tag{A.3}$$

and see that this is independent of the warp velocity $v^i_\infty(t)$.

Then for the Eulerian energy density and flux one has the particularly simple results:

$$\rho = 0; \qquad \text{and} \qquad \vec{f} = 0. \tag{A.4}$$

This does *not* mean that the WEC is satisfied since one still has to check $\rho + \bar{p}$.

Specifically, in this particular warp field we have

$$\rho + \bar{p} = \bar{p} = -\frac{1}{12\pi}\,\nabla_a(K\,n^a) = -\frac{1}{12\pi}\,\boldsymbol{\nabla}\cdot(K\,\vec{v}(t,\vec{x})). \tag{A.5}$$

That is

$$\rho + \bar{p} = -\frac{1}{12\pi}\,\boldsymbol{\nabla}\cdot\left(\frac{3}{2}\,\frac{2m}{r^2}\,\hat{r} + K\vec{v}_\infty\right) = -\frac{1}{12\pi}\,\vec{v}_\infty(t)\cdot\boldsymbol{\nabla}K. \tag{A.6}$$

Finally

$$\rho + \bar{p} = -\frac{3}{16\pi} \frac{\sqrt{2m}}{r^{5/2}} \ \vec{v}_\infty(t) \cdot \hat{r}. \tag{A.7}$$

But the factor $\vec{v}_\infty(t) \cdot \hat{r}$ explictly changes sign depending on whether one is "in front of" or "behind" the warp bubble. Thus this warp bubble explicitly violates the NEC, and so violates the WEC as well — this is a specific and explicit example of why merely checking the non-negativity of the Eulerian density is simply insufficient for checking the WEC.

More generally, note that an *asymptotic* version of this argument will still apply to any warp field that is asymptotically of the form (A.2):

$$v^i(t, \vec{x}) = -\sqrt{\frac{2m}{r}} \ \{1 + o(1/r)\} \ \hat{r}^i + v^i_\infty(t). \tag{A.8}$$

This now applies to any warp bubble that asymptotes to Schwarzschild spacetime at large distances.

Acknowledgments

MV was directly supported by the Marsden Fund, via a grant administered by the Royal Society of New Zealand.

JS acknowledges indirect financial support via the Marsden fund, administered by the Royal Society of New Zealand.

SS acknowledges support from both the technical and administrative staff at the Charles University, and specific financial support via OP RDE project number CZ.02.2.69/0.0/0.0/18_053/0016976 (International mobility of research).

References

1. M. S. Morris and K. S. Thorne, "Wormholes in space-time and their use for interstellar travel: A tool for teaching general relativity", Am. J. Phys. **56** (1988), 395-412 doi:10.1119/1.15620.
2. M. S. Morris, K. S. Thorne and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition", Phys. Rev. Lett. **61** (1988), 1446-1449 doi:10.1103/PhysRevLett.61.1446.
3. M. Visser, "Traversable wormholes: Some simple examples", Phys. Rev. D **39** (1989), 3182-3184 doi:10.1103/PhysRevD.39.3182 [arXiv:0809.0907 [gr-qc]].
4. M. Visser, "Traversable wormholes from surgically modified Schwarzschild spacetimes", Nucl. Phys. B **328** (1989), 203-212 doi:10.1016/0550-3213(89)90100-4 [arXiv:0809.0927 [gr-qc]].
5. T. A. Roman, "Inflating Lorentzian wormholes", Phys. Rev. D **47** (1993), 1370-1379 doi:10.1103/PhysRevD.47.1370 [arXiv:gr-qc/9211012 [gr-qc]].
6. J. G. Cramer, R. L. Forward, M. S. Morris, M. Visser, G. Benford and G. A. Landis, "Natural wormholes as gravitational lenses", Phys. Rev. D **51** (1995), 3117-3120 doi:10.1103/PhysRevD.51.3117 [arXiv:astro-ph/9409051 [astro-ph]].
7. S. Kar, "Evolving wormholes and the weak energy condition", Phys. Rev. D **49** (1994), 862-865 doi:10.1103/PhysRevD.49.862.

8. S. Kar and D. Sahdev, "Restricted class of traversable wormholes with traceless matter", Phys. Rev. D **52** (1995), 2030-2035 doi:10.1103/PhysRevD.52.2030.
9. M. Visser, "Lorentzian wormholes: From Einstein to Hawking", (AIP Press, now Springer, New York, 1995).
10. L. H. Ford and T. A. Roman, "Quantum field theory constrains traversable wormhole geometries", Phys. Rev. D **53** (1996), 5496-5507 doi:10.1103/PhysRevD.53.5496 [arXiv:gr-qc/9510071 [gr-qc]].
11. E. Teo, "Rotating traversable wormholes", Phys. Rev. D **58** (1998), 024014 doi:10.1103/PhysRevD.58.024014 [arXiv:gr-qc/9803098 [gr-qc]].
12. S. A. Hayward, "Dynamic wormholes", Int. J. Mod. Phys. D **8** (1999), 373-382 doi:10.1142/S0218271899000286 [arXiv:gr-qc/9805019 [gr-qc]].
13. M. Visser, S. Kar and N. Dadhich, "Traversable wormholes with arbitrarily small energy condition violations", Phys. Rev. Lett. **90** (2003), 201102 doi:10.1103/PhysRevLett.90.201102 [arXiv:gr-qc/0301003 [gr-qc]].
14. J. P. S. Lemos, F. S. N. Lobo and S. Quinet de Oliveira, "Morris-Thorne wormholes with a cosmological constant", Phys. Rev. D **68** (2003), 064004 doi:10.1103/PhysRevD.68.064004 [arXiv:gr-qc/0302049 [gr-qc]].
15. T. A. Roman, "Some thoughts on energy conditions and wormholes", doi:10.1142/9789812704030_0236 [arXiv:gr-qc/0409090 [gr-qc]].
16. S. Kar, N. Dadhich and M. Visser, "Quantifying energy condition violations in traversable wormholes", Pramana **63** (2004), 859-864 doi:10.1007/BF02705207 [arXiv:gr-qc/0405103 [gr-qc]].
17. T. Muller, "Exact geometric optics in a Morris-Thorne wormhole spacetime", Phys. Rev. D **77** (2008), 044043 doi:10.1103/PhysRevD.77.044043.
18. S. A. Hayward, "Wormhole dynamics in spherical symmetry", Phys. Rev. D **79** (2009), 124001 doi:10.1103/PhysRevD.79.124001 [arXiv:0903.5438 [gr-qc]].
19. O. James, E. von Tunzelmann, P. Franklin and K. S. Thorne, "Visualizing Interstellar's Wormhole", Am. J. Phys. **83** (2015), 486 doi:10.1119/1.4916949 [arXiv:1502.03809 [gr-qc]].
20. M. Alcubierre, "The Warp drive: Hyperfast travel within general relativity", Class. Quant. Grav. **11** (1994), L73-L77 doi:10.1088/0264-9381/11/5/001 [arXiv:gr-qc/0009013 [gr-qc]].
21. R. J. Low, "Speed limits in general relativity", Class. Quant. Grav. **16** (1999), 543-549 doi:10.1088/0264-9381/16/2/016 [arXiv:gr-qc/9812067 [gr-qc]].
22. C. Clark, W. A. Hiscock and S. L. Larson, "Null geodesics in the Alcubierre warp drive space-time: The View from the bridge", Class. Quant. Grav. **16** (1999), 3965-3972 doi:10.1088/0264-9381/16/12/313 [arXiv:gr-qc/9907019 [gr-qc]].
23. F. Lobo and P. Crawford, "Weak energy condition violation and superluminal travel", Lect. Notes Phys. **617** (2003), 277-291 [arXiv:gr-qc/0204038 [gr-qc]].
24. F. S. N. Lobo and M. Visser, "Fundamental limitations on 'warp drive' spacetimes", Class. Quant. Grav. **21** (2004), 5871-5892 doi:10.1088/0264-9381/21/24/011 [arXiv:gr-qc/0406083 [gr-qc]].
25. F. S. N. Lobo and M. Visser, "Linearized warp drive and the energy conditions", [arXiv:gr-qc/0412065 [gr-qc]].
26. F. S. N. Lobo, "Exotic solutions in General Relativity: Traversable wormholes and 'warp drive' spacetimes", [arXiv:0710.4474 [gr-qc]].
27. S. Finazzi, S. Liberati and C. Barceló, "Semiclassical instability of dynamical warp drives", Phys. Rev. D **79** (2009), 124017 doi:10.1103/PhysRevD.79.124017 [arXiv:0904.0141 [gr-qc]].

28. B. K. Tippett and D. Tsang, "The Blue Box White Paper", [arXiv:1310.7983 [physics.pop-ph]].
29. M. Alcubierre and F. S. N. Lobo, "Wormholes, Warp Drives and Energy Conditions", Fundam. Theor. Phys. **189** (2017), pp. 1–279 doi:10.1007/978-3-319-55182-1 [arXiv:2103.05610 [gr-qc]].
30. M. Alcubierre and F. S. N. Lobo, "Warp Drive Basics", Fundam. Theor. Phys. **189** (2017), 257–279 doi:10.1007/978-3-319-55182-1_11.
31. J. Santiago, S. Schuster and M. Visser, "Generic warp drives violate the null energy condition", [arXiv:2105.03079 [gr-qc]].
32. E. W. Lentz, "Breaking the warp barrier: hyper-fast solitons in Einstein–Maxwell-plasma theory", Class. Quant. Grav. **38** (2021) no.7, 075015 doi:10.1088/1361-6382/abe692 [arXiv:2006.07125 [gr-qc]].
33. A. Bobrick and G. Martire, "Introducing Physical Warp Drives", Class. Quant. Grav. **38** (2021) no.10, 105009 doi:10.1088/1361-6382/abdf6e [arXiv:2102.06824 [gr-qc]].
34. S. D. B. Fell and L. Heisenberg, "Positive energy warp drive from hidden geometric structures", Class. Quant. Grav. **38** (2021) no.15, 155020 doi:10.1088/1361-6382/ac0e47 [arXiv:2104.06488 [gr-qc]].
35. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, "Dust content solutions for the Alcubierre warp drive spacetime", Eur. Phys. J. C **80** (2020) no.8, 786 doi:10.1140/epjc/s10052-020-8355-2 [arXiv:2008.06560 [gr-qc]].
36. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, "Fluid dynamics in the warp drive spacetime geometry", Eur. Phys. J. C **81** (2021) no.2, 133 doi:10.1140/epjc/s10052-021-08921-3 [arXiv:2101.11467 [gr-qc]].
37. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, "Charged dust solutions for the warp drive spacetime", Gen. Rel. Grav. **53** (2021) no.2, 23 doi:10.1007/s10714-021-02799-y [arXiv:2102.05119 [gr-qc]].
38. O. L. Santos-Pereira, E. M. C. Abreu and M. B. Ribeiro, "Perfect fluid warp drive solutions with the cosmological constant," Eur. Phys. J. Plus **136** (2021) no.9, 902 doi:10.1140/epjp/s13360-021-01899-7 [arXiv:2108.10960 [gr-qc]].
39. W. Béatrix-Drouhet, "Exotic Fluids Matching the Stress-Energy Tensor of Alcubierre Warp Drive Spacetimes", [arXiv:2012.09941 [gr-qc]].
40. J. Santiago, S. Schuster and M. Visser, "Tractor beams, pressor beams, and stressor beams in general relativity", Universe **7** (2021) no.8, 271 doi:10.3390/universe7080271 [arXiv:2106.05002 [gr-qc]].
41. C. Barceló and M. Visser, "Twilight for the energy conditions?", Int. J. Mod. Phys. D **11** (2002), 1553-1560 doi:10.1142/S0218271802002888 [arXiv:gr-qc/0205066 [gr-qc]].
42. M. Visser, "General relativistic energy conditions: The Hubble expansion in the epoch of galaxy formation", Phys. Rev. D **56** (1997), 7578-7587 doi:10.1103/PhysRevD.56.7578 [arXiv:gr-qc/9705070 [gr-qc]].
43. M. Visser and C. Barceló, "Energy conditions and their cosmological implications", doi:10.1142/9789812792129_0014 [arXiv:gr-qc/0001099 [gr-qc]].
44. M. Visser, "Gravitational vacuum polarization. 1: Energy conditions in the Hartle-Hawking vacuum", Phys. Rev. D **54** (1996), 5103-5115 doi:10.1103/PhysRevD.54.5103 [arXiv:gr-qc/9604007 [gr-qc]].
45. M. Visser, "Gravitational vacuum polarization. 2: Energy conditions in the Boulware vacuum", Phys. Rev. D **54** (1996), 5116-5122 doi:10.1103/PhysRevD.54.5116 [arXiv:gr-qc/9604008 [gr-qc]].
46. M. Visser, "Gravitational vacuum polarization. 3: Energy conditions in the (1+1) Schwarzschild space-time", Phys. Rev. D **54** (1996), 5123-5128 doi:10.1103/PhysRevD.54.5123 [arXiv:gr-qc/9604009 [gr-qc]].

47. M. Visser, "Gravitational vacuum polarization. 4: Energy conditions in the Unruh vacuum", Phys. Rev. D **56** (1997), 936-952 doi:10.1103/PhysRevD.56.936 [arXiv:gr-qc/9703001 [gr-qc]].
48. M. Visser, "Gravitational vacuum polarization", MG8: 8th Marcel Grossmann Meeting, Jerusalem, July 1997, [arXiv:gr-qc/9710034 [gr-qc]].
49. M. Visser, "Scale anomalies imply violation of the averaged null energy condition", Phys. Lett. B **349** (1995), 443-447 doi:10.1016/0370-2693(95)00303-3 [arXiv:gr-qc/9409043 [gr-qc]].
50. V. Baccetti, P. Martín-Moruno and M. Visser, "Null Energy Condition violations in bimetric gravity", JHEP **08** (2012), 148 doi:10.1007/JHEP08(2012)148 [arXiv:1206.3814 [gr-qc]].
51. P. Martín-Moruno and M. Visser, "Semiclassical energy conditions for quantum vacuum states", JHEP **09** (2013), 050 doi:10.1007/JHEP09(2013)050 [arXiv:1306.2076 [gr-qc]].
52. P. Martín-Moruno and M. Visser, "Classical and quantum flux energy conditions for quantum vacuum states", Phys. Rev. D **88** (2013) no.6, 061701 doi:10.1103/PhysRevD.88.061701 [arXiv:1305.1993 [gr-qc]].
53. D. Hochberg, C. Molina-París and M. Visser, "Tolman wormholes violate the strong energy condition", Phys. Rev. D **59** (1999), 044011 doi:10.1103/PhysRevD.59.044011 [arXiv:gr-qc/9810029 [gr-qc]].
54. P. Martín-Moruno and M. Visser, "Classical and semi-classical energy conditions", Fundam. Theor. Phys. **189** (2017), 193-213 doi:10.1007/978-3-319-55182-1_9 [arXiv:1702.05915 [gr-qc]].
55. M. Visser, B. Bassett and S. Liberati, "Perturbative superluminal censorship and the null energy condition", AIP Conf. Proc. **493** (1999) no.1, 301-305 doi:10.1063/1.1301601 [arXiv:gr-qc/9908023 [gr-qc]].
56. P. Martín-Moruno and M. Visser, "Semi-classical and nonlinear energy conditions", doi:10.1142/9789813226609_0126 [arXiv:1510.00158 [gr-qc]].
57. J. Friedman, M. S. Morris, I. D. Novikov, F. Echeverria, G. Klinkhammer, K. S. Thorne and U. Yurtsever, "Cauchy problem in space-times with closed timelike curves", Phys. Rev. D **42** (1990), 1915-1930 doi:10.1103/PhysRevD.42.1915.
58. S. W. Hawking, "The Chronology protection conjecture", Phys. Rev. D **46** (1992), 603-611 doi:10.1103/PhysRevD.46.603.
59. S. W. Kim and K. S. Thorne, "Do vacuum fluctuations prevent the creation of closed timelike curves?", Phys. Rev. D **43** (1991), 3929-3947 doi:10.1103/PhysRevD.43.3929.
60. F. Echeverria, G. Klinkhammer and K. S. Thorne, "Billiard balls in wormhole spacetimes with closed timelike curves: Classical theory", Phys. Rev. D **44** (1991), 1077-1099 doi:10.1103/PhysRevD.44.1077.
61. M. Visser, "From wormhole to time machine: Comments on Hawking's chronology protection conjecture", Phys. Rev. D **47** (1993), 554-565 doi:10.1103/PhysRevD.47.554 [arXiv:hep-th/9202090 [hep-th]].
62. M. Visser, "Hawking's chronology protection conjecture: Singularity structure of the quantum stress energy tensor", Nucl. Phys. B **416** (1994), 895-906 doi:10.1016/0550-3213(94)90560-6 [arXiv:hep-th/9303023 [hep-th]].
63. B. S. Kay, M. J. Radzikowski and R. M. Wald, "Quantum field theory on space-times with a compactly generated Cauchy horizon", Commun. Math. Phys. **183** (1997), 533-556 doi:10.1007/s002200050042 [arXiv:gr-qc/9603012 [gr-qc]].
64. S. Krasnikov, "Quantum field theory and time machines", Phys. Rev. D **59** (1999), 024010 doi:10.1103/PhysRevD.59.024010 [arXiv:gr-qc/9802008 [gr-qc]].

65. M. Visser, "The Quantum physics of chronology protection", [arXiv:gr-qc/0204022 [gr-qc]].
66. J. L. Friedman, K. Schleich and D. M. Witt, "Topological censorship", Phys. Rev. Lett. **71** (1993), 1486-1489 [erratum: Phys. Rev. Lett. **75** (1995), 1872] doi:10.1103/PhysRevLett.71.1486 [arXiv:gr-qc/9305017 [gr-qc]].
67. T. Jacobson and S. Venkataramani, "Topology of event horizons and topological censorship", Class. Quant. Grav. **12** (1995), 1055-1062 doi:10.1088/0264-9381/12/4/012 [arXiv:gr-qc/9410023 [gr-qc]].
68. G. J. Galloway, K. Schleich, D. M. Witt and E. Woolgar, "Topological censorship and higher genus black holes", Phys. Rev. D **60** (1999), 104039 doi:10.1103/PhysRevD.60.104039 [arXiv:gr-qc/9902061 [gr-qc]].
69. J. L. Friedman and A. Higuchi, "Topological censorship and chronology protection", Annalen Phys. **15** (2006), 109-128 doi:10.1002/andp.200510172 [arXiv:0801.0735 [gr-qc]].

A singularity theorem for evaporating black holes

E.-A. Kontou* and B. Freivogel
ITFA and GRAPPA, Universiteit van Amsterdam
Science Park 904, Amsterdam, the Netherlands
**E-mail: e.a.kontou@uva.nl*

D. Krommydas
Instituut-Lorentz, Universiteit Leiden
P.O. Box 9506, 2300 RA Leiden, The Netherlands

The classical singularity theorems of General Relativity rely on energy conditions that are easily violated by quantum fields. Here, we provide motivation for an energy condition obeyed in semiclassical gravity: the smeared null energy condition (SNEC), a proposed bound on the weighted average of the null energy along a finite portion of a null geodesic. Using SNEC as an assumption we proceed to prove a singularity theorem. This theorem extends the Penrose singularity theorem to semiclassical gravity and has interesting applications to evaporating black holes.

Keywords: Quantum fields, gravity, energy conditions, quantum inequalities, singularities

1. Introduction

The classical or pointwise energy conditions are bounds on components of the stress-energy tensor and were introduced early on in the history of general relativity. Their purpose is to encode properties of what is considered "reasonable" matter, and predict the evolution of gravitational systems in a model-independent way.

The energy condition of interest in this work is the null energy condition (NEC). The NEC states that the stress-energy tensor contracted with two null vectors is non-negative everywhere. Using the Einstein Equation, we get the geometric form of the NEC or the null convergence condition. Using a perfect fluid stress-energy tensor, we can give a physical interpretation of the NEC: the sum of energy density and pressure cannot be negative. The three forms are summarized in Table 1.

Table 1. The different forms of the NEC. Here ℓ^μ is a null vector.

Physical form	Geometric form	Perfect fluid
$T_{\mu\nu}\ell^\mu\ell^\nu \geq 0$	$R_{\mu\nu}\ell^\mu\ell^\nu \geq 0$	$\rho + P \geq 0$

*Presenting author.

The NEC is obeyed by minimally coupled scalar fields but as with all pointwise energy conditions, it is violated by quantum fields.[1] Ford[2] was the first to introduce *quantum energy inequalities* (QEIs), restrictions on the possible magnitude and duration of any negative energy densities within a quantum field theory.

QEIs have since been derived for flat and curved spacetimes, bosonic and fermionic fields (see Ref. 3 and 4 for recent reviews). Those bounds are for averages over timelike curves or worldvolumes. As an example, the renormalized null energy of the quantum massless minimally coupled scalar field in Minkowski spacetime averaged over a smooth timelike curve γ obeys the following QEI[5]

$$\int dt \langle :T_{\mu\nu}:\ell^\mu \ell^\nu \rangle_\omega f^2(t) \geq -\frac{(v^\mu \ell_\mu)}{12\pi^2} \int dt f''(t)^2 \,, \tag{1}$$

for all Hadamard states ω and any smooth, real-valued compactly supported function f. Here v^μ is the timelike vector tangent to γ. For f a normalized Gaussian with zero mean and t_0 variance the right hand side of (1) becomes

$$\int dt \langle :T_{\mu\nu}:\ell^\mu \ell^\nu \rangle_\omega f^2(t) \geq -\frac{(v^\mu \ell_\mu)}{64\pi^2 t_0^4} \,. \tag{2}$$

Then we can see the physical interpretation of the QEI: the longer the timescale t_0, the less negative null energy is allowed.

Important classical relativity results such as the Penrose singularity theorem[6] have the NEC in their hypotheses. If one wants to apply such theorems in a semiclassical setting it is necessary to replace the pointwise energy condition with a condition obeyed by quantum fields, namely a QEI. As the Penrose theorem proves null geodesic incompleteness, the relevant QEI would be a null averaged one.

The purpose of this contribution is to motivate a null QEI, the smeared null energy condition (SNEC) and use it to prove a semiclassical singularity theorem for null geodesic incompleteness. This theorem is applicable to the case of evaporating black holes. We begin with a description of the challenges to develop a null QEI and motivation for SNEC in Sec. 2. In Sec. 3 we state the singularity theorem of Ref. 7 and show that SNEC can be used as an assumption. In Sec. 4 we apply the theorem to a toy model of evaporating black holes. We conclude in Sec. 5 with a summary and discussion of future work.

2. Null quantum energy inequalities

2.1. *The Fewster-Roman counterexample*

In the expression of Eq. (1) the renormalized null energy is averaged over a timelike curve. A similar expression integrated over a null geodesic has been derived two-dimensions. In particular, Fewster and Hollands[8] showed that

$$\int_{-\infty}^{+\infty} f(\lambda) \langle :T_{\mu\nu}:\ell^\mu \ell^\nu \rangle_\omega \geq -\frac{c}{48\pi} \int_{-\infty}^{+\infty} \frac{(f')^2}{f} d\lambda \,, \tag{3}$$

holds for a class of interacting quantum fields, namely the unitary, positive energy conformal field theories (CFTs) with stress-energy tensor in Minkowski spacetime. Here c is the central charge of the theory. We recently generalized that result for a large class of curved backgrounds.[9]

The situation is different in more than two dimensions. Fewster and Roman[5] showed using an explicit construction, that the renormalized null energy averaged over a null geodesic is unbounded from below for the massless minimally coupled scalar field. So there are no null QEIs in four-dimensional Minkowski space.

Their construction was a sequence of vacuum-plus-two-particle states. Then they allowed the three-momenta of excited modes to become increasingly parallel to the spatial part of the null vector ℓ^μ. As the three momenta grows, the lower bound of the inequality diverges to negative infinity.

2.2. *The smeared null energy condition*

To overcome the problem encountered by Fewster and Roman, Freivogel and Krommydas proposed the smeared null energy condition (SNEC).[10] The main concept behind it is that in quantum filed theory there often exists an ultraviolet cutoff $\ell_{\rm UV}$. It was shown[9,10] that the existence of a cutoff restricts the three momenta of the excited modes in the Fewster-Roman counterexample, leading to a finite lower bound.

The SNEC can be written as

$$\int_{-\infty}^{+\infty} d\lambda g^2(\lambda)\langle T_{kk}(x^\mu(\lambda))\rangle \geq -\frac{4B}{G_N}\int_{-\infty}^{+\infty} d\lambda \left(g'(\lambda)\right)^2 , \tag{4}$$

where $x^\mu(\lambda)$ is a null geodesic, $g(\lambda)$ is a differentiable 'smearing function' that controls the region where the null energy is averaged, B is a constant and G_N is the Newton constant. In four-dimensional field theory we can write

$$NG_N \lesssim \ell_{\rm UV}^2 , \tag{5}$$

where N is the number of fields. This relationship means the SNEC provides a finite lower bound even for a large number of fields. This is particularly useful for applications where the negative energy arises from multiple fields with small negative energy fluctuations (see e.g. 11).

To have B be an order one number, we need to saturate that inequality. This is the case for the induced gravity proof of 12 where they derived $B = 1/32\pi$. However, it is reasonable to consider a $B \ll 1$ since (5) is typically not saturated in controlled constructions.

SNEC has been proven to hold for free fields on Minkowski spacetime.[9,13] The proof utilizes the fact that free field theory factorizes on the lightsheet in a collection of two-dimensional CFTs. For each of those CFTs the two-dimensional null QEI of Eq. (3) holds leading to a proof for the higher dimensional theory.

2.3. *The double smeared null energy condition*

It is unclear if the proof of SNEC described in the previous subsection can be generalized to curved spacetimes and interacting fields. Additionally, the bound diverges fro $\ell_{\rm UV} \to 0$, with the ultraviolet cutoff depending on the theory.

Those disadvantages of SNEC led to the proposal of a different bound, the double smeared null energy condition (DSNEC). The idea is to average the renormalized null energy density in both null directions, denote + and −. Schematically the DSNEC can be written as

$$\int d^2x^{\pm} g^2(x^+, x^-) \langle :T_{--}: \rangle_\omega \geq -\frac{\mathcal{N}}{\delta_+ \delta_-^3}\,, \tag{6}$$

where $\mathcal{N}$ depends on the number of fields and the smearing function, and $\delta_\pm$ is the smearing length in each of the null directions.

For a massless scalar in Minkowski the DSNEC can be explicitly written as

$$\int d^2x^{\pm} g^2(x^+, x^-) \langle :T:_{--} \rangle_\omega \geq -A \left[\int dx^+ (g''_+(x^+))^2\right]^{1/4} \left[\int dx^- (g''_-(x^-))^2\right]^{3/4}\,, \tag{7}$$

where A is a number and we assumed that the smearing function factorizes as $g^2 = g_+(x^+)^2 g_-(x^-)^2$.

The DSNEC was motivated in 13 and its rigorous proof will appear in future work.[14] The proof of DSNEC can straightforwardly generalized for curved spacetimes as it is derived from a general QEI valid in spacetimes with curvature.[15] It includes no theory dependent cutoff and the smearing in each direction can be controlled. However, it is still unclear if the DSNEC can be used to prove singularity theorems. The main obstacle is that the usual proofs of those theorems require bounds on single geodesics.

3. The singularity theorem

3.1. *The Penrose singularity theorem*

In general relativity a spacetime is singular if it possesses at least one incomplete and inextendible geodesic. This definition does not give us information about the nature of the singularity (e.g. if curvature scalars diverge) but it allowed for the first model-independent theorems, the singularity theorems of Penrose[6] and Hawking.[16]

Most singularity theorems have the same three types of hypotheses: the energy condition, the initial or boundary condition and the causality condition. In the case of the Penrose theorem the energy condition is the NEC or more accurately the geometric form of the NEC, the null convergence condition. The boundary condition is the existence of a trapped surface, a co-dimension two spacelike submanifold which has two null normals with negative expansions. Equivalently, a trapped surface has negative null normal curvature everywhere. Finally, the causality condition is the existence of a non-compact Cauchy hypersurface. The conclusion is that the spacetime is future null geodesically incomplete.

Schematically, singularity theorems work in the following way: the initial condition establishes the convergence of a congruence of geodesics. The energy condition guarantees that the convergence will continue and a focal point will form. Finally, the causality condition does not allow the formation of focal points leading to a contradiction that proves the geodesic incompleteness.

3.2. *Singularity theorems with weaker conditions*

As quantum fields violate all pointwise energy conditions, a semiclassical singularity theorem is required to have a weaker energy condition. Examples of singularity theorems with such conditions include Refs. 17, 18 and 19 but none of them address the case of a condition obeyed by quantum fields. First Ref. 20 proved singularity theorems with energy conditions inspired by QEIs. Ref. 7 proved singularity theorems with similar conditions using index form methods. Utilizing these results Ref. 21 proved the first semiclassical singularity theorem for timelike geodesic incompleteness. Here we follow Ref. 21 theorem for null geodesic incompleteness.

To state the theorem we first need to fix a parametrization of the affine parameter of the null geodesic. For a manifold M let P be submanifold of co-dimension 2 with mean normal curvature vector field $H^\mu = H\hat{H}^\mu$ where $\hat{H}^\mu$ is a future-pointing timelike unit vector. Then let γ be a future-directed null geodesic emanating normally from P. Then $\hat{H}_\mu$ is extended by parallel transporting along γ. Now we can choose an affine parameter λ on γ, such that $\hat{H}_\mu d\gamma^\mu/d\lambda = 1$. Now we can state the energy condition from 7

$$\int_0^\ell g(\lambda)^2 R_{\mu\nu}\ell^\mu\ell^\nu d\lambda \geq -Q_m(\gamma)\|g^{(m)}\|^2 - Q_0(\gamma)\|g\|^2\,, \tag{8}$$

where Q_m and Q_0 are unknown constants dependent on the choice of γ and m a positive integer. The notation $\|\cdot\|$ denotes the L^2 norm.

The bound required by the singularity theorem is a geometric assumption, while SNEC is an assumption on the renormalized stress-energy tensor. Classically, the Einstein equation connects curvature to the stress-energy tensor. Semiclassically, the semiclassical Einstein equation (SEE) equates the expectation value of the stress-energy tensor with the classical Einstein tensor

$$8\pi G_N\langle T_{\mu\nu}\rangle_\omega = G_{\mu\nu}\,. \tag{9}$$

Using the SEE the bound of Eq. (4) can be written as

$$\int_{-\infty}^{\infty} g(\lambda)^2 R_{kk} d\lambda \geq -32\pi B\|g'(\lambda)\|^2\,. \tag{10}$$

Then this is a bound of the form of Eq. (8) with $m = 1$, $Q_1 = 32\pi B$ and $Q_0 = 0$. Using the SEE assumes that we have a self-consistent solution, which includes a state ω and a metric $g_{\mu\nu}$.

In addition to the energy condition the theorem of Ref. 7 has an assumption on the pointwise null energy density for a finite affine parameter. In particular, there

are two scenarios to describe all possible initial conditions: in scenario 1, initially the NEC is satisfied for an affine length ℓ_0, short compared to the one for the formation of a focal point ℓ. In scenario 2 this requirement is dropped and instead conditions are imposed on the null contracted Ricci tensor for small negative values of the affine parameter. Here we focus on scenario 2.

We first extend γ to $\gamma : [-\ell_0, \ell] \to M$ and assume that Eq. (10) holds on the extended geodesic. Then we define $\rho_{\max} = \max_{[-\ell_0,0]} \rho$ and we can use Lemma 4.7 of Ref. 7 with $m = 1$, $Q_0 = 0$, $A_1 = 1/3$, $B_1 = C_1 = 1$. If we additionally assume that $\rho_{\max} < 0$ we have

Lemma 3.1. *For ρ satisfying Eq. (10) on $[-\ell_0, \ell]$ if*

$$-2H \geq \frac{Q_1 + 2}{\ell} + \frac{Q_1}{\ell_0} + \frac{1}{3}\rho_{\max}\ell_0 \,, \tag{11}$$

then there is a focal point to P along γ in $[0, \ell]$.

Negative null energy in $[-\ell_0, 0]$ region leads to smaller required initial contraction because this negative energy must be over-compensated by positive energy. This effect has been studied and it is known as "quantum interest".[22]

4. Application to evaporating black holes

Penrose proved the first singularity theorem which applies to a classical black hole spacetime. However, this theorem cannot be applied in an evaporating black hole spacetime, where the NEC is violated. Here we apply Lemma 3.1 in a toy model of an evaporating black hole spacetime.

First, we assume that the metric is approximated by Schwarzschild geometry near the classical horizon

$$ds^2 = \left(\frac{R_s}{r} - 1\right) dt^2 - \left(\frac{R_s}{r} - 1\right)^{-1} dr^2 + r^2 d\Omega^2 \,, \tag{12}$$

where R_s is the Schwarzschild radius. We focus on spherically symmetric hypersurfaces P, so that the hypersurface is defined by Schwarzschild coordinates (t_p, r_p) where the mean normal curvature vector field is purely in the r direction.

Inside the horizon, the mean normal curvature H of our surfaces P is given by[23]

$$H(r_P) = -\frac{1}{r_P}\sqrt{\frac{R_s}{r_P} - 1} \,. \tag{13}$$

Since we assumed $\rho_{\max} < 0$ we can drop the last term of Eq. (11)

$$H < -\frac{Q_1}{2\ell} - \frac{1}{\ell} - \frac{Q_1}{2\ell_0} \,. \tag{14}$$

H depends on two parameters, the maximum affine parameter for the formation of the singularity ℓ and the length of the affine parameter that the NEC is violated ℓ_0.

We define the dimensionless parameter x

$$R_s - r_P \equiv x R_s \,, \qquad 0 < x < 1 \,, \tag{15}$$

and y by demanding that the affine distance ℓ is a coordinate distance yR_s. We can consider the case that $y \to \infty$ meaning we have no information about the location of the singularity.

The idea is that if the mean normal curvature of the hypersurfaces P is smaller than the one required by Lemma 3.1 we have a singularity. So we equate the expressions (13) and (14) to find the location of the first hypersurface inside the horizon for which we can apply the Lemma. We want P to be as close to the classical horizon as possible. The setup is shown in Fig. 1.

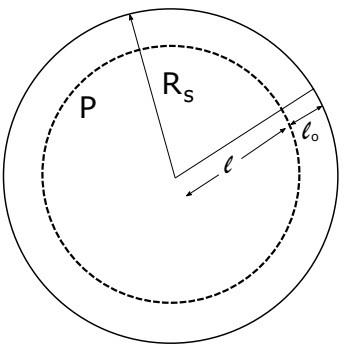


Fig. 1. Schematic representation of a Schwarzschild black hole and the parameters. The dashed circle is constant r and t hypersurface P. Distance ℓ_0 is from the point where the NEC starts being violated, and distance ℓ is from P to the singularity (pictured here at $r = 0$).

A plot of x for different values of y is shown in Fig. 2 for two different values of Q_1. The Ref. 12 value of $B = 1/32\pi$ translates to $Q_1 = 1$. Using this value for Q_1, we find that the minimum x is $1/3$. As discussed earlier, there is also strong motivation to use a value of $B \ll 1$ and so $Q_1 \ll 1$. For small Q_1, we have a singularity theorem for spheres P with

$$R_s - r_P \gtrapprox R_s \frac{Q_1}{2} \qquad \text{for } Q_1 \ll 1 \,. \tag{16}$$

5. Conclusions

In this work we provided motivation for both the smeared null energy condition (SNEC) and the double smeared null energy condition (DSNEC). We proved a semiclassical singularity theorem using SNEC and applied this theorem to establish that spacetimes that approximate the Schwarzschild solution near the horizon must contain a singularity.

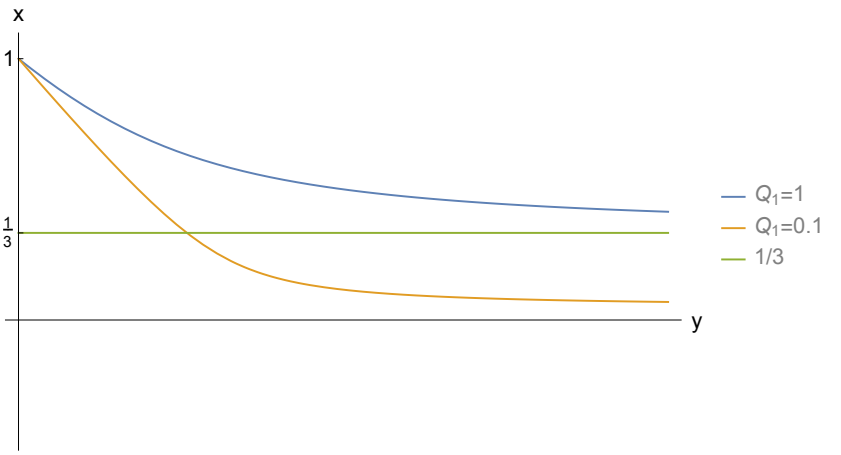


Fig. 2. Required value of x to have a singularity for different values y. For $Q_1 = 0.1$ the minimum value is much smaller compared to $Q_1 = 1$.

As the version of SNEC we use for the singularity theorem has only been proven for Minkowski spacetime, an important future direction is a theorem with a condition that incorporates curvature. There are different ways to approach that. One is to attempt to prove SNEC for spacetimes with curvature. Another is to use the new bound of DSNEC instead. This would require novel concepts as the current singularity theorem proofs are for bounds on single geodesics.

Acknowledgments

BF and E-AK are supported by the ERC Consolidator Grant QUANTIVIOL. This work is part of the Δ ITP consortium, a program of the NWO that is funded by the Dutch Ministry of Education, Culture and Science (OCW).

References

1. H. Epstein, V. Glaser and A. Jaffe, Nonpositivity of energy density in Quantized field theories, *Nuovo Cim.* **36**, p. 1016 (1965).
2. L. Ford, Quantum Coherence Effects and the Second Law of Thermodynamics, *Proc. Roy. Soc. Lond. A* **364**, 227 (1978).
3. E.-A. Kontou and K. Sanders, Energy conditions in general relativity and quantum field theory, *Class. Quant. Grav.* **37**, p. 193001 (2020).
4. C. J. Fewster, *Quantum Energy Inequalities*, in *Wormholes, Warp Drives and Energy Conditions*, ed. F. S. N. Lobo (Springer International Publishing, Cham, 2017), Cham, pp. 215–254.
5. C. J. Fewster and T. A. Roman, Null energy conditions in quantum field theory, *Phys. Rev.* **D67**, p. 044003 (2003), [Erratum: Phys. Rev.D80,069903(2009)].
6. R. Penrose, Gravitational collapse and space-time singularities, *Phys. Rev. Lett.* **14**, 57 (1965).

7. C. J. Fewster and E.-A. Kontou, A new derivation of singularity theorems with weakened energy hypotheses, *Class. Quant. Grav.* **37**, p. 065010 (2020).
8. C. J. Fewster and S. Hollands, Quantum energy inequalities in two-dimensional conformal field theory, *Rev. Math. Phys.* **17**, p. 577 (2005).
9. B. Freivogel, E.-A. Kontou and D. Krommydas, The Return of the Singularities: Applications of the Smeared Null Energy Condition (12 2020).
10. B. Freivogel and D. Krommydas, The Smeared Null Energy Condition, *JHEP* **12**, p. 067 (2018).
11. J. Maldacena, A. Milekhin and F. Popov, Traversable wormholes in four dimensions (7 2018).
12. S. Leichenauer and A. Levine, Upper and Lower Bounds on the Integrated Null Energy in Gravity, *JHEP* **01**, p. 133 (2019).
13. J. R. Fliss and B. Freivogel, Semi-local Bounds on Null Energy in QFT (8 2021).
14. F. B. Fliss, Jackson R. and E.-A. Kontou, In preparation. (2021).
15. C. J. Fewster and C. J. Smith, Absolute quantum energy inequalities in curved spacetime, *Annales Henri Poincare* **9**, 425 (2008).
16. S. W. Hawking, The Occurrence of singularities in cosmology, *Proc. Roy. Soc. Lond.* **A294**, 511 (1966).
17. F. J. Tipler, Energy conditions and spacetime singularities, *Phys. Rev.* **D17**, 2521 (1978).
18. C. Chicone and P. Ehrlich, Line integration of Ricci curvature and conjugate points in Lorentzian and Riemannian manifolds, *Manuscripta Math.* **31**, 297 (1980).
19. A. Borde, Geodesic focusing, energy conditions and singularities, *Class. Quant. Grav.* **4**, 343 (1987).
20. C. J. Fewster and G. J. Galloway, Singularity theorems from weakened energy conditions, *Class. Quant. Grav.* **28**, p. 125009 (2011).
21. C. J. Fewster and E.-A. Kontou, A semiclassical singularity theorem (8 2021).
22. L. Ford and T. A. Roman, The Quantum interest conjecture, *Phys. Rev. D* **60**, p. 104018 (1999).
23. T. Pilkington, A. Melanson, J. Fitzgerald and I. Booth, Trapped and marginally trapped surfaces in Weyl-distorted Schwarzschild solutions, *Class. Quant. Grav.* **28**, p. 125018 (2011).

Circularly symmetric thin-shell wormholes in F(R) gravity with (2+1)-dimensions

Cecilia Bejarano*, Ernesto F. Eiroa† and Griselda Figueroa-Aguirre‡
Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), Casilla de Correo 67, Sucursal 28, 1428, Buenos Aires, Argentina
**E-mail: cbejarano@iafe.uba.ar*
†*E-mail: eiroa@iafe.uba.ar*
‡*E-mail: gfigueroa@iafe.uba.ar*

Within the framework of $F(R)$ theories of gravity with (2+1)-dimensions and constant scalar curvature R, we construct a family of thin-shell wormholes with circular symmetry and we analyze the stability of the static configurations under radial perturbations. We show an example of asymptotically anti-de Sitter thin-shell wormholes with charge, finding that stable configurations with normal matter are possible for a suitable range of the parameters.

Keywords: Wormholes; F(R) gravity; Thin shells.

1. Introduction

One of the main goals of alternative theories of gravity is focused on dealing with some fundamental issues in the current cosmological picture, from the early-time inflation scenario to the late-time cosmic accelerated expansion. Basically, the intention is to modify the geometrical side of the gravitational field equations without adding any unknown matter-energy content. Certainly, any alternative approach also intends to provide a geometrical explanation for the spacetime singularities as well as a cornerstone for a quantum version of gravity theory. There is a plethora of different modified gravity models. In particular, $F(R)$ gravity[1–3] is one of the most straightforward modifications to General Relativity (GR), since the Einstein-Hilbert action is replaced by

$$S = \frac{1}{2\kappa}\int d^4x\sqrt{|g|}\left(F(R)+\mathcal{L}_m\right), \tag{1}$$

with $\kappa = 8\pi$ and R the scalar curvature (we adopt units such that $c = G = 1$, with c the speed of light and G the gravitational constant). In the metric formalism, in which the metric tensor is the only dynamical variable to perform the variation of the action, the field equations read

$$F'(R)R_{\mu\nu} - \frac{g_{\mu\nu}}{2}F(R) - \nabla_\mu\nabla_\nu F'(R) + g_{\mu\nu}\nabla_\rho\nabla^\rho F'(R) = \kappa T_{\mu\nu}, \tag{2}$$

where ∇ is the covariant derivative and $T_{\mu\nu}$ is the energy-momentum tensor.

The junction formalism[4–7] in (3+1)-dimensions has been applied within $F(R)$ gravity, for instance, to study thin layers of matter surrounding vacuum (namely, bubbles)[8,9] and black holes.[9,10] Bubbles made of a pure double layer in quadratic

$F(R)$ have been also studied a few years ago.[11] Recently, thin-shell wormholes in $F(R)$ gravity have been explored.[12–16] In this work, we present charged thin-shell wormholes in (2+1)-dimensional $F(R)$ gravity with and we analyze the stability of the static solutions under radial perturbations.

2. Junction conditions

The cut-and-paste formalism is commonly used to construct some wormhole geometries. It is based on the junction conditions of the corresponding theory, which allows to match different solutions across a hypersurface Σ. For convenience, we adopt the usual notation that the jump of any quantity Υ across Σ is $[\Upsilon] \equiv (\Upsilon^2 - \Upsilon^1)|_\Sigma$.

2.1. *General relativity*

In GR, the junction conditions are expressed in the Darmois-Israel formalism,[4,5] which can be sketched as follows:

- Two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ are glued at a hypersurface Σ to obtain a geodesically complete manifold $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$.
- The first fundamental form (the induced metric) at Σ is

$$h_{ij}^{1,2} = g_{\mu\nu}^{1,2} \frac{\partial X_{1,2}^{\mu}}{\partial \xi^i} \frac{\partial X_{1,2}^{\nu}}{\partial \xi^j}\bigg|_\Sigma , \tag{3}$$

  where $X_{1,2}^{\mu}$ are, respectively, the coordinates that describe $\mathcal{M}_{1,2}$, and ξ^a are the intrinsic local coordinates at the hypersurface, hence $X_{1,2}^{\mu} = X_{1,2}^{\mu}(\xi^a)$.
- The second fundamental form (the extrinsic curvature) at Σ is given by

$$K_{ij}^{1,2} = -n_\gamma^{1,2} \left(\frac{\partial^2 X_{1,2}^{\gamma}}{\partial \xi^i \partial \xi^j} + \Gamma_{\alpha\beta}^{\gamma} \frac{\partial X_{1,2}^{\alpha}}{\partial \xi^i} \frac{\partial X_{1,2}^{\beta}}{\partial \xi^j} \right)\bigg|_\Sigma , \tag{4}$$

  with $n_\gamma^{1,2}$ the unit normal ($n^\gamma n_\gamma = 1$) pointing from $\mathcal{M}_1$ to $\mathcal{M}_2$.
- The matching between $\mathcal{M}_1$ and $\mathcal{M}_2$ at Σ, should fulfill the so-called junction conditions:
  - The line element should be continuous across Σ:

$$[h_{\mu\nu}] = 0. \tag{5}$$

  - The field equations at Σ (Lanczos equations) reduce to

$$\kappa S_{\mu\nu} = -[K_{\mu\nu}] - h_{\mu\nu} K^{\gamma}{}_{\gamma}, \quad n^\mu S_{\mu\nu} = 0 \tag{6}$$

    where $S_{\mu\nu}$ is the energy-momentum tensor at the gluing hypersurface.

The hypersurface Σ is called a *boundary surface* if $S_{\mu\nu} = 0$, or a *thin shell* of matter if $S_{\mu\nu} \neq 0$.

2.2. *F(R) gravity*

The formalism for the junction conditions in $F(R)$ gravity was developed in recent years.[6,7] In this theory, there are two separate cases to take into account:

- General (non-quadratic) case: $F'''(R) \neq 0$
  - At the matching hypersurface Σ, the following continuity relations should be satisfied

$$[h_{\mu\nu}] = 0, \quad [K^{\mu}{}_{\mu}] = 0, \quad [R] = 0, \tag{7}$$

  - The field equations at Σ result

$$\kappa S_{\mu\nu} = -F'(R)[K_{\mu\nu}] + F''(R)[n^{\rho}\nabla_{\rho}R]h_{\mu\nu}, \quad n^{\mu}S_{\mu\nu} = 0. \tag{8}$$

- Quadratic case: $F'''(R) = 0$
  - There is one less restriction in this case, so that the discontinuity of R at Σ is allowed

$$[h_{\mu\nu}] = 0, \quad [K^{\mu}{}_{\mu}] = 0. \tag{9}$$

  - If we write $F(R) = \alpha R^2 + R - 2\Lambda$, the field equations at Σ read

$$\kappa S_{\mu\nu} = -[K_{\mu\nu}] + 2\alpha([n^{\gamma}\nabla_{\gamma}R]h_{\mu\nu} - [RK_{\mu\nu}]), \quad n^{\mu}S_{\mu\nu} = 0. \tag{10}$$

  - In order to guarantee a divergence-free energy-momentum tensor (locally conservation) at Σ, it is necessary to add the following extra contributions[17,18]
    - *The external energy flux vector*:

$$\kappa\mathcal{T}_{\mu} = -2\alpha\nabla_{\mu}[R], \quad n^{\mu}\mathcal{T}_{\mu} = 0,$$

    - *The external scalar pressure/tension*:

$$\kappa\mathcal{T} = 2\alpha[R]K^{\gamma}{}_{\gamma},$$

    - *The double layer contribution*:

$$\kappa\mathcal{T}_{\mu\nu} = \nabla_{\gamma}\left(2\alpha[R]h_{\mu\nu}n^{\gamma}\delta^{\Sigma}\right),$$

      with δ^{Σ} the Dirac delta on Σ (which is analogous to the dipole distributions in classical electrodynamics). It is easy to verify that all these contributions vanish if the scalar curvature is continuous at Σ.

The constraint $[K^{\mu}{}_{\mu}] = 0$ of $F(R)$ gravity, which is not present in GR, has important consequences, as we shall see in the next sections.

3. Charged solutions with constant curvature

$F(R)$ gravity coupled to a nonlinear electromagnetic field is described in (2+1)-dimensions by the action[19]

$$S = \frac{1}{2\kappa}\int d^3x\sqrt{|g|}(F(R) + (-\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu})^s), \tag{11}$$

where s is a positive parameter ($s \neq 1/2$) and $\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu$ is the electromagnetic tensor. The corresponding energy-momentum tensor becomes traceless for $s = 3/4$, so this value is adopted in what follows.

The dynamical equations given by Eq. (2) for the gravitational sector are completed with the electromagnetic field part

$$\partial_\mu \left(\sqrt{-g}F^{\mu\nu}(-F_{\alpha\beta}F^{\alpha\beta})^{-1/4}\right) = 0. \tag{12}$$

For constant $R = R_0$, the spherically symmetric geometry[19]

$$ds^2 = -A(r)dt^2 + A(r)^{-1}dr^2 + r^2 d\theta^2, \tag{13}$$

with

$$A(r) = -M - \frac{\left(2\mathcal{Q}^2\right)^{3/4}}{2F'(R_0)r} - \frac{R_0 r^2}{6}, \tag{14}$$

is solution of the theory represented by the action (11), where $R_0 = 6\Lambda_{eff}$ can be thought in terms of an effective cosmological constant, M is the mass and $\mathcal{Q}$ is the charge. Notice that this spacetime is asymptotically anti-de Sitter for $R_0 < 0$. The vacuum static BTZ geometry of GR is obtained if $\mathcal{Q} = 0$, so it is also a solution in $F(R)$ gravity. We can define an effective charge with its sign determined by the sign of $F'(R_0)$

$$Z \equiv \frac{\left(2\mathcal{Q}^2\right)^{3/4}}{2F'(R_0)}. \tag{15}$$

The different values of Z establish the horizon structure of the geometry given by Eq. (14).

In the next Section, we use this spacetime to construct thin-shell wormholes by applying the junction formalism in $F(R)$ gravity, and we study the stability of the static configurations by following the standard potential approach.[20]

4. Wormhole geometry: Construction and stability

We proceed to build a class of thin-shell wormholes by cutting the outer parts of the spherically symmetric spacetimes with metrics of the form

$$ds^2_{1,2} = -A_{1,2}(r)dt^2_{1,2} + A_{1,2}(r)^{-1}dr^2 + r^2 d\theta^2, \tag{16}$$

and glueing them on the circle Σ defined by $G(r) = r - a = 0$. In this way, we create the manifold $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, where the regions at the sides of Σ are given

by $\mathcal{M}_{1,2} = \{X^\alpha = (t_{1,2}, r, \theta)/r \geqslant a\}$ and with the metrics $A_{1,2}$, respectively. The only angular coordinate is $0 \leq \theta \leq 2\pi$ and the circle length $2\pi r$ is minimal at Σ, so $\mathcal{M}$ represents a wormhole connecting two regions with a throat of radius a, i.e. the flare-out condition is satisfied there. In order to obtain a traversable wormhole, the radius a is taken large enough to remove the presence of possible event horizons and singularities in both original manifolds. We take $\xi^i = (\tau, \theta)$ as the coordinates on Σ, being τ the proper time. The radius of Σ is taken as a function of time, i.e. $a = a(\tau)$, in order to analyze the stability of the static shells. The proper time should coincide at both sides of Σ, so that $dt_{1,2}/d\tau = \sqrt{A_{1,2}(a) + \dot{a}^2}/A_{1,2}(a)$, with the overdot representing $da/d\tau$ (the free signs are fixed provided that the times $t_{1,2}$ and τ all run into the future).

In the orthonormal basis $\{e_{\hat{\tau}} = e_\tau, e_{\hat{\theta}} = a^{-1}e_\theta\}$, at the hypersurface Σ, the first fundamental form results

$$h^{1,2}_{\hat{i}\hat{j}} = \mathrm{diag}(-1, 1), \tag{17}$$

and the non-null components of the second fundamental form for static configurations are

$$K^{1,2}_{\hat{\tau}\hat{\tau}} = \pm\frac{A'_{1,2}(a_0)}{2\sqrt{A_{1,2}(a_0)}}, \tag{18}$$

$$K^{1,2}_{\hat{\theta}\hat{\theta}} = \mp\frac{1}{a_0}\sqrt{A_{1,2}(a_0)}, \tag{19}$$

where the prime represents derivative with respect to r. In our construction, the equation $[h_{\hat{i}\hat{j}}] = 0$ is automatically satisfied, while the condition $[K^{\hat{i}}{}_{\hat{i}}] = 0$ gives the following relation

$$\frac{A'_2(a_0)}{2\sqrt{A_2(a_0)}} + \frac{A'_1(a_0)}{2\sqrt{A_1(a_0)}} + \frac{1}{a_0}\left(\sqrt{A_1(a_0)} + \sqrt{A_2(a_0)}\right) = 0, \tag{20}$$

that should be always fulfilled by the throat radius a_0.

We can obtain the dynamics of the thin shell from the time dependent version of $[K^{\hat{i}}{}_{\hat{i}}] = 0$, which takes the form

$$\frac{2\ddot{a} + A'_2(a)}{2\sqrt{A_2(a) + \dot{a}^2}} + \frac{2\ddot{a} + A'_1(a)}{2\sqrt{A_1(a) + \dot{a}^2}} + \frac{1}{a}\left(\sqrt{A_1(a) + \dot{a}^2} + \sqrt{A_2(a) + \dot{a}^2}\right) = 0. \tag{21}$$

Then, we can analyze the stability of the configuration by rewriting this dynamical equation of the throat in terms of the potential

$$\dot{a}^2 = -V(a), \tag{22}$$

where

$$V(a) = -\frac{a^2\left(A_1(a) - A_2(a)\right)^2}{4a_0^2\left(\sqrt{A_1\left(a_0\right)} + \sqrt{A_2\left(a_0\right)}\right)^2} - \frac{a_0^2\left(\sqrt{A_1\left(a_0\right)} + \sqrt{A_2\left(a_0\right)}\right)^2}{4a^2} + \frac{A_1(a) + A_2(a)}{2}. \tag{23}$$

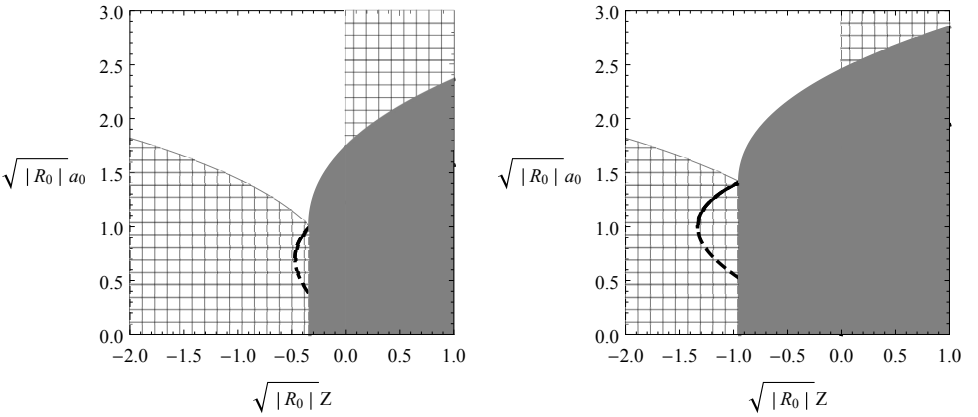


Fig. 1. Wormholes in general $F(R)$ gravity symmetric across the throat. Solid lines represent stable solutions while dashed ones unstable solutions. Meshed zones represent matter satisfying WEC. Grey zones have no physical meaning. Left plot: $M = 0.5$; right one: $M = 1$.

It is straightforward to see that $V(a_0) = 0$ and $V'(a_0) = 0$. Then, the static solutions are stable under radial perturbations when $V''(a_0) > 0$, otherwise they are unstable.

In the following examples, we adopt for our construction the charged solution with constant scalar curvatue introduced in the previous section, so the metric functions read

$$A_i(r) = -M_i - \frac{(2\mathcal{Q}_i^2)^{3/4}}{2F'(R_i)r} - \frac{R_i r^2}{6}, \quad i = 1, 2 \tag{24}$$

where R_i, M_i, and $\mathcal{Q}_i$ are the values of the different parameters in each region $\mathcal{M}_i$.

4.1. *Wormholes in general $F(R)$ with $[R] = 0$*

In the first example, we adopt a symmetric construction across the throat, i.e. $R_1 = R_2 = R_0$, $M_1 = M_2 = M$, and $\mathcal{Q}_1 = \mathcal{Q}_2 = \mathcal{Q}$ so that $A_1(r) = A_2(r) = A(r)$. The throat radius a_0 should fulfill Eq. (20). Then, from Eq. (8), by using that in the orthonormal basis $S_{\hat{i}\hat{j}} = \mathrm{diag}(\sigma_0, p_0)$, we get the surface energy density and the pressure for the static solutions

$$\sigma_0 = \frac{F'(R_0)}{\kappa} \frac{A'(a_0)}{\sqrt{A(a_0)}}, \tag{25}$$

$$p_0 = -2\frac{F'(R_0)}{a_0\kappa}\sqrt{A(a_0)}. \tag{26}$$

If $\sigma_0 \geq 0$ and $\sigma_0 + p_0 \geq 0$, we say that the matter at the throat is normal, because it satisfies the weak energy condition (WEC), otherwise it is exotic. By using Eq. (20), we can see that the matter should satisfy the equation of state $\sigma_0 - 2p_0 = 0$.

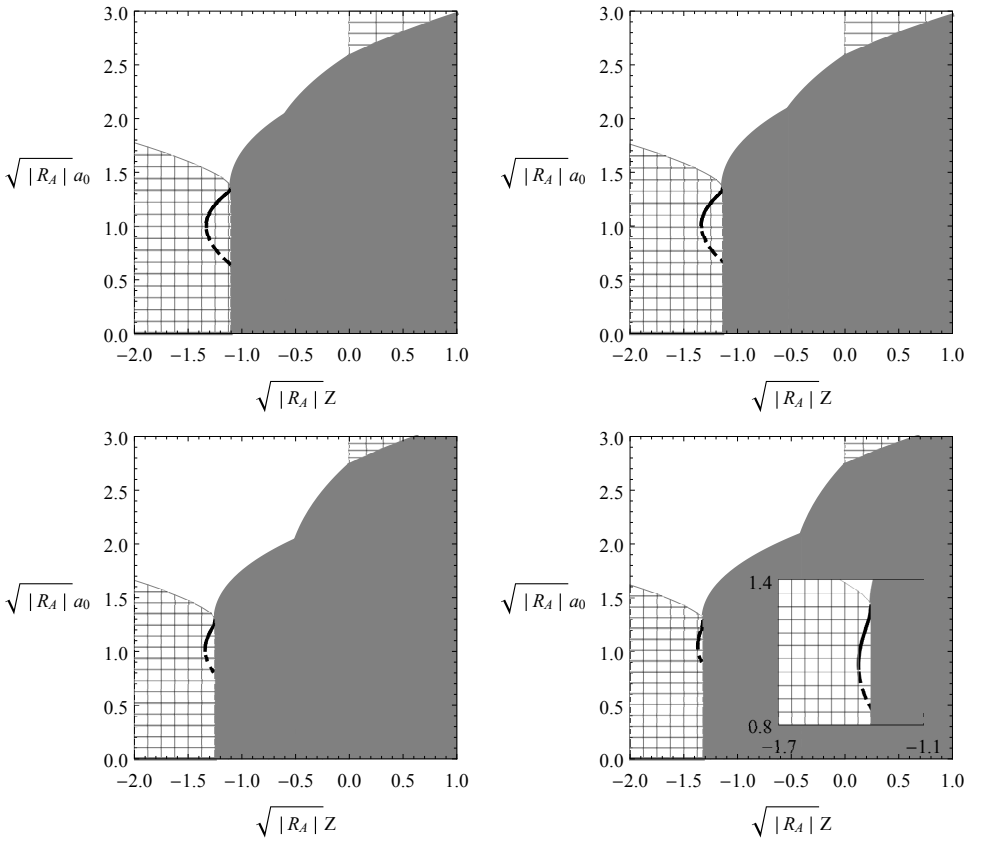


Fig. 2. Wormholes in quadratic $F(R)$ gravity with $R_1 \neq R_2$ at each side of the throat. Solid lines represent stable solutions while dashed ones unstable solutions. Meshed zones represent matter satisfying WEC. Grey zones have no physical meaning. For all plots: $M = 1$, $R_A = (R_1 + R_2)/2$. Upper row: $R_1 = 0.9R_A$ and $R_2 = 1.1R_A$; lower row: $R_1 = 0.8R_A$ and $R_2 = 1.2R_A$. Left column: $\alpha R_A = -0.1$ for $Z > 0$ and $\alpha R_A = -0.9$ for $Z < 0$; right column: $\alpha R_A = -0.2$ for $Z > 0$ and $\alpha R_A = -0.8$ for $Z < 0$.

From Fig. 1, we see that there are two solutions constituted of normal matter for a short range of $\sqrt{|R_0|}Z$ with $Z < 0$. Hence it requires the presence of ghost fields, i.e. $F'(R_0) < 0$. The solution with the largest radius is stable while the other one is unstable under radial perturbations. The qualitative behavior of the solutions is not affected by the different values of M which only represent a change of scale.

4.2. *Wormholes in quadratic $F(R)$ with $[R] \neq 0$*

In the second example, we adopt $M_1 = M_2 = M$, and $\mathcal{Q}_1 = \mathcal{Q}_2 = \mathcal{Q}$ but $R_1 \neq R_2$, so we are restricted to quadratic $F(R)$ gravity. The throat radius a_0 should satisfy Eq. (20). From Eq. (10) the energy density and pressure at the throat in the static

configuration are

$$\sigma_0 = \frac{1+2\alpha R_2}{\kappa}\left(\frac{A_2'(a_0)}{2\sqrt{A_2(a_0)}}\right) + \frac{1+2\alpha R_1}{\kappa}\left(\frac{A_1'(a_0)}{2\sqrt{A_1(a_0)}}\right), \tag{27}$$

$$p_0 = -\frac{1+2\alpha R_2}{\kappa}\left(\frac{\sqrt{A_2(a_0)}}{a_0}\right) - \frac{1+2\alpha R_1}{\kappa}\left(\frac{\sqrt{A_1(a_0)}}{a_0}\right). \tag{28}$$

In this case, there exists an external scalar pressure/tension $\mathcal{T}_0$ that satisfies the equation of state $\sigma_0 - p_0 = \mathcal{T}_0$ which results by using Eq. (20). The other extra contributions are: $\mathcal{T}_\mu^{(0)} = 0$ and $\mathcal{T}_{\mu\nu}^{(0)} \propto (2\alpha[R]/\kappa)h_{\mu\nu}$.

From Fig. 2, we see that there are two solutions made of normal matter for a short range of $\sqrt{|R_A|}Z$ with $R_A = (R_1 + R_2)/2$; since $Z < 0$, ghost fields appear. The solution with the largest radius is stable; the other is unstable. The larger the difference between R_1 and R_2, the smaller the range of $\sqrt{|R_A|}Z$ where solutions exist. The same qualitative behaviour is obtained when α decreases. As in the previous example, the different values of M does not affect the qualitative behavior of the solutions and it only represents a change of scale.

5. Summary

We have constructed a family of circularly symmetric thin-shell wormholes in (2+1)-dimensional $F(R)$ gravity with constant scalar curvature R at each side of the throat. We have studied their stability under perturbations preserving the symmetry. The junction conditions in $F(R)$ gravity are more restrictive than in GR. In fact, the extra condition $[K^\mu{}_\mu] = 0$ forces the type of equation of state: for general $F(R)$, it should be $\sigma_0 - 2p_0 = 0$, while for quadratic $F(R)$, it results $\sigma_0 - 2p_0 = \mathcal{T}_0$. We have analyzed two examples: (i) symmetric wormholes ($R_1 = R_2 = R_0$; $M_1 = M_2 = M$; $\mathcal{Q}_1 = \mathcal{Q}_2 = \mathcal{Q}$) across the throat in general $F(R)$ and (ii) wormholes asymmetric in the scalar curvature ($R_1 \neq R_2$; $M_1 = M_2 = M$ and $\mathcal{Q}_1 = \mathcal{Q}_2 = \mathcal{Q}$) only in quadratic $F(R)$. In both cases, we have found two solutions made of normal matter (i.e., satisfying WEC) for a range of values with the effective charge $Z < 0$. However, for these solution the presence of ghost fields is always required since $Z < 0$ implies $F'(R_0) < 0$. We have found that the solution with the larger radius is stable while the other is unstable. The qualitative behavior of the solutions is not altered by M, resulting only in a change of scale. In particular, for quadratic $F(R)$ as long as the difference between the constant (negative) scalar curvatures R_1 and R_2 grows, the range of the squared charge $\mathcal{Q}^2$ where stable solutions exist becomes smaller; similar behavior can be observed when α decreases. Another example and more references can be found in our recently published work.[20]

Acknowledgments

This work has been supported by CONICET and Universidad de Buenos Aires. C. B. thanks the partial support of the John Templeton Foundation.

References

1. T. P. Sotiriou and V. Faraoni, F(R) theories of gravity, *Rev. Mod. Phys.* **82**, 451 (2010).
2. A. De Felice and S. Tsujikawa, F(R) theories, *Living Rev. Relativ.* **13**, 3 (2010).
3. S. Nojiri, S. D. Odintsov, and V. K. Oikonomou, Modified gravity theories on a nutshell: Inflation, bounce and late-time evolution, *Phys. Rep.* **692**, 1 (2017).
4. G. Darmois, Mémorial des Sciences Mathématiques, Fascicule XXV, Chap. V (Gauthier-Villars, Paris, 1927).
5. W. Israel, Singular hypersurfaces and thin shells in general relativity, *II Nuovo Cimento B* **44**, 1 (1966); **48**, 463 (E) (1967).
6. N. Deruelle, M. Sasaki, and Y. Sendouda, Junction conditions in F(R) theories of gravity, *Prog. Theor. Phys.* **119**, 237 (2008).
7. J. M. M. Senovilla, Junction conditions for F(R) gravity and their consequences, *Phys. Rev. D.* **88**, 064015 (2013).
8. E. F. Eiroa and G. Figueroa-Aguirre, Spherical thin shells in F(R) gravity: construction and stability, *Eur. Phys. J. C* **78**, 54 (2018).
9. E. F. Eiroa and G. Figueroa-Aguirre, Thin shells in F(R) gravity with non-constant scalar curvature, *Eur. Phys. J. Plus* **135**, 774 (2020).
10. E. F. Eiroa and G. Figueroa-Aguirre, Thin shells surrounding black holes in F(R) gravity, *Eur. Phys. J. C* **79**, 171 (2019).
11. E. F. Eiroa, G. Figueroa-Aguirre, and J. M. M. Senovilla, Pure double-layer bubbles in quadratic F(R) gravity, *Phys. Rev. D* **95**, 124021 (2017).
12. E. F. Eiroa and G. Figueroa-Aguirre, Thin-shell wormholes with charge in F(R) gravity, *Eur. Phys. J. C* **76**, 132 (2016).
13. E. F. Eiroa and G. Figueroa-Aguirre, Thin-shell wormholes with a double layer in quadratic F(R) gravity, *Phys. Rev. D* **94**, 044016 (2016).
14. M. Zaeem-ul-Haq Bhatti, A. Anwar, S. Ashraf, Construction of thin shell wormholes from metric F(R) gravity, *Mod. Phys. Lett. A* **32**, 1750111 (2017).
15. S. Habib Mazharimousavi, A note on thin-shell wormholes with charge in F(R)-gravity, *Eur. Phys. J. C* **78**, 612 (2018).
16. S. Habib Mazharimousavi, M. Halilsoy, K. Kianfar, Thin-shell wormholes in F(R)-gravity coupled with nonlinear electrodynamics, *Eur. Phys. J. Plus* **135**, 440 (2020).
17. J. M. M. Senovilla, Gravitational double layers, *Class. Quant. Grav.* **31**, 072002 (2014).
18. B. Reina, J. M .M. Senovilla, and R. Vera, Junction conditions in quadratic gravity: thin shells and double layers, *Class. Quant. Grav.* **33**, 105008 (2016).
19. S. H. Hendi, B. Eslam Panah, and R. Saffari, Exact solutions of three-dimensional black holes: Einstein gravity versus F(R) gravity, *Int. J. Mod. Phys. D* **23**, 1450088 (2014).
20. C. Bejarano, E. F. Eiroa, and G. Figueroa-Aguirre, Thin-shell wormholes in (2+1)-dimensional F(R) theories, *Eur. Phys. J. C* **81**, 668 (2021).

Warp drive dynamic solutions considering different fluid sources

Osvaldo L. Santos-Pereira[1,*], Everton M. C. Abreu[2,3,4,†] and Marcelo B. Ribeiro[1,4,‡]

[1] *Physics Institute, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil,*
[2] *Physics Department, Universidade Federal Rural do Rio de Janeiro, Seropédica, Brazil,*
[3] *Physics Department, Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil*
[4] *Applied Physics Graduate Program, Physics Institute, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil*
E-mails: * *olsp@if.ufrj.br,* † *evertonabreu@ufrrj.br,* ‡ *mbr@if.ufrj.br*

Alcubierre proposed in 1994 that the well known special relativistic limitation that particles cannot travel with velocities bigger than the light speed can be bypassed when such trips are considered globally within specific general relativistic frameworks. Although initial results indicated this scenario as being unphysical, since it would seem to require negative mass-energy density, recent theoretical analyses suggest that such an unphysical situation may not always be necessarily true. In this paper, we present solutions of the Einstein equations using the original Alcubierre warp drive metric endowed with various matter-energy sources, namely dust, perfect fluid, anisotropic fluid, and perfect fluid within a cosmological constant spacetime. A connection of some of these solutions featuring shock waves described by the Burgers equation is also shown.

Keywords: Warp drive solutions, perfect fluid, anisotropic fluid, shock waves, electromagnetic field, cosmological constant

1. Introduction

Alcubierre[1] advanced a model based on a specific general relativistic spacetime geometry in which massive particles can travel at superluminal speeds if they are located inside a specially designed spacetime distortion. The physics of this possible propulsion method, named as *warp drive* (WD) after science fiction literature, consists of a special spacetime metric forming a spacetime distortion, called *warp bubble*, such that it generates an expansion behind the distortion and a contraction in front of it. That makes it possible for a massive particle to be in a local sub-luminal speed inside the bubble, as required by special relativity, whereas outside the particle is propelled at superluminal speeds. Therefore, employing of this spacetime distortion a massive particle travels between two points in spacetime with apparent time less than the time a light particle with zero mass would require to travel between the same two points.

The WD metric, as originally proposed by Alcubierre, is based on a general metric that uses the 3+1 formalism.[2,3] It consists of a boost in the direction of one of the spatial coordinates, described by the shift vector, a function of spacetime coordinates and the product of two functions, the velocity of the center of the warp bubble and the shape function that controls the shape of the bubble. Although

Alcubierre did not solve the Einstein equations with his proposed spacetime geometry, he gave an example of a shape function in the form of a hat function.[1] One of the main caveats of the Alcubierre WD spacetime is the requirement of negative mass-energy density and the violation of the dominant and weak energy conditions.

Ford and Roman[4] derived quantum inequalities for free massless quantized scalar fields, electromagnetic fields, and massive scalar fields in four-dimensional Minkowski spacetime and concluded that these constraints have the form of an uncertainty principle limitation on the magnitude and duration of negative mass-energy densities and that the exotic solutions of Einstein equations such as wormholes and WD would have significant limitations in their viability. Pfenning and Ford[5] followed this last work and calculated the upper bound limits necessary for the WD viability, concluding then that the energy required to create a warp bubble is ten orders of magnitude greater than the total mass of the entire visible universe, also negative.

Krasnikov[6] approached the hyperfast interstellar travel problem in general relativity by discussing the possibility of whether or not a mass particle can reach a remote point in spacetime and return sooner than a photon would. Krasnikov argued that it is not possible for a mass particle to win this race under reasonable assumptions for globally hyperbolic spacetimes. He discussed in detail the specific spacetime topologies, but with the constraint that tachyons would be a requirement for superluminal travel to occur. He conjectured a spacetime modification device that can be used to make superluminal travel possible without tachyons. Such spacetime was named as *Krasnikov tube* by Everet and Roman,[7] and they generalized the metric proposed by Krasnikov by hypothesizing a tube along the path of the particle connecting Earth to a distant star.

The Krasnikov metric cannot shorten the time for a one-way trip from Earth to a distant star, but it can make the time for a round trip arbitrarily short for clocks on Earth. However, the Everet and Roman[7] extension of the Krasnikov metric has the property that inside the tube the spacetime is flat and the lightcones are opened in such a way that they allow the superluminal travel in one direction. They also mentioned that even though the Krasnikov tube does not involve closed timelike curves it is possible to construct a time machine with a system of two non-overlapping tubes, and demonstrated that the Krasnikov tube also requires thin layers of negative energy density and large total negative energies.

Lobo and Crawford[8] discussed the Krasnikov metric in detail, addressed the violation of the weak energy condition and the viability of superluminal travel. These authors concluded that with the imposition of the weak energy condition the Olum theorem[9] prohibits superluminal travel and pointed out the necessity of further research on spacetimes with closed timelike curves and the need for a precise definition of superluminal travel. They argued that one can construct metrics that allow superluminal travel but are flat Minkowski spacetimes. The quantum inequalities, brought from quantum field theory,[7] were also discussed. Van de Broeck[10] showed

how a minor modification of the Alcubierre WD geometry can reduce the total energy required for the creation of a warp bubble. He presented a modification of the original WD metric where the total negative mass would be of the order of just a few solar masses.

Natário[11] proposed a new version of the WD theory with zero expansion and questioned the effects that occur in the WD in opposition to his newly proposed metric, namely, the zero expansion Natário WD metric. Natário discussed the nature of the WD spacetime symmetries with a series of propositions and corollaries such as the proof that the WD spacetime is flat whenever the tangent vector to the Cauchy surfaces is a Killing vector field for the Euclidean metric despite being time-dependent and as a particular case. In addition, it is also flat wherever the tangent vector is spatially constant. He also showed that nonflat WD spacetime violate both the *weak energy condition* (WEC) and the *strong energy condition* (SEC), as well as that the WD metric can be obtained from the Natário metric by a particular choice of coordinates. The spherical coordinates were the choice of charts for the zero expansion of the Natário WD spacetime.

Lobo and Visser[12] pointed out that the WD theory is an example of reverse engineering of the solutions of the Einstein equations where one defines a specific spacetime metric and then one finds the matter distribution responsible for the respective geometry. These authors verified that the class of WD spacetimes necessarily violate the classical energy conditions even for low warp bubble velocity. Hence, this is the case of a geometric choice and not of superluminal properties. They proposed a more realistic WD by applying linearized gravity to the weak WD with nonrelativistic warp bubble velocities and argued that for the Alcubierre WD and its version proposed by Natário[11] the center of the bubble must be massless. They found that even for low velocities the negative energy stored in the warp fields must be a significant fraction of the particle's mass at the center of the warp bubble.

White[13,14] described how a warp field interferometer could be implemented at the Advanced Propulsion Physics Laboratory with the help of the original[1] WD ideias. He also pointed out that the expansion behind the warp bubble and contraction in front of it is due to the nature of the WD functions, and that the distortion of spacetime may be interpreted as a kind of Doppler effect or stress and strain on spacetime. Lee and Cleaver[15] argued that external radiation might affect the WD, and that the warp field interferometer proposed by White[13,14] could not detect spacetime distortions. Mattingly et al.[16] discussed curvature invariants in the Natário WD.

Bobrick and Martire[17] proposed a general WD spacetime that encloses all WD definitions removing any alleged issues with the original Alcubierre WD. They presented a general subluminal model with spherical symmetry and positive energy solutions that satisfies the quantum energy inequalities such that it reduces two orders of magnitude the WD requirement for negative energy density. They claimed

that any type of WD, including the original one, is a place of regular or exotic material moving in inertial form with a certain speed.

Lentz[18] claimed that the original warp bubble proposed by Alcubierre[1] can be physically interpreted as hyper-fast gravitational solitons and presented the first solution for superluminal solitons in general relativity satisfying the WEC and the momentum conditions for conventional sources of stress, like energy and momentum, that do not require large amounts of negative energy. The basis for his work is to assume that the shift vector components obey a kind of wave equation giving rise to a positive energy geometry. Fell and Heisenberg[19] also addressed the WD as gravitational solitons and shed light on the Eulerian energies and their relation to the WEC, raising the possibility for superluminal spacetimes with viable amounts of energy density.

Quarra[20] established that within the scope of general relativity certain gravitational waveforms can result in geodesics that arrive at distant points earlier than light signals in flat spacetime, and presented an example waveform that can be used to manifest superluminal behavior.

Santiago et al.[21] claimed that generic WD metrics violate the *null energy conditions* (NEC) and discussed how Eulerian observers are privileged, meaning that these observers may perceive positive energy densities causing the impression of viable WD. They also argue that for the WD to become a possibility it would require that all timelike observers observe positive energy densities. They stated that any WD spacetime will unavoidably violate the energy conditions. In a subsequent work[22] they claimed that exotic spacetimes, such as the WD one will always violate energy conditions. They also provided other examples of such spacetimes relating them to wormholes, tractor beams and stress beams.

For a detailed discussion on the basics of WD theory, the reader is referred to Alcubierre and Lobo.[23] For a nice exposition on WD theory one is referred to the lecture notes of the course given by Shoshany.[24] It is particularly noteworthy the recent results concerning the possibility of WD features in negative energy density distribution of an experimental Casimir cavity,[25] results that open the striking possibility of ingenious experimental avenues on how to create the WD phenomena in a laboratory environment.

In this work, we present a summary of the results we have recently obtained when solving the Einstein equations with the assumption of simple fluid distributions embedded in the original Alcubierre WD spacetime geometry. New results such as the vacuum solutions connecting the WD to shock waves via the Burgers equation are presented.[26] We used the perfect fluid energy-momentum tensor[27] (EMT) which disclosed new exact solutions of the Einstein equations. We also demonstrated that starting with simple forms of mass and energy sources, vacuum solutions for the WD spacetime arise when we impose that the WEC is trivially satisfied by a choice of gauge on the shift vector and the spacetime coordinates.[27] Further results of the Alcubierre WD metric endowed with a charged dust EMT that considers

an electromagnetic tensor in curved spacetime with the cosmological constant are mentioned.[28] Finally, adding the cosmological constant to the Einstein equations coupled with the Alcubierre WD spacetime having perfect fluid as a source has the effect of leading the energy density to possibly becoming positive depending on both the value and sign of the cosmological constant.[29]

The paper is organized as follows. Section 2 reviews very brief the Alcubierre WD metric. Section 3 depicts the perfect and anisotropic fluid cases[27] and Section 4 discusses the relationship between the WD spacetime, vacuum solutions, and shock waves described by the Burgers equations.[26] Section 5 presents a discussion on the role of the cosmological constant in the Einstein equation solutions for the WD spacetime.[29] Section 6 presents the conclusions.

2. The Alcubierre Warp Drive Spacetime Metric

Alcubierre originally used[1] the Arnowitt-Deser-Misner (ADM) approach of general relativity,[2,3] a formalism where the spacetime is described by a foliation of space-like hypersurfaces of constant time coordinate. The general form of the WD metric in this formalism is described by the following equation,

$$\begin{aligned} ds^2 = -\,d\tau^2 &= g_{\alpha\beta} dx^\alpha dx^\beta \\ &= -\left(\alpha^2 - \beta_i \beta^i\right) dt^2 + 2\beta_i\, dx^i\, dt + \gamma_{ij}\, dx^i\, dx^j \ , \end{aligned} \tag{1}$$

where $d\tau$ is the lapse of proper time, α is the lapse function, β^i is the space-like shift vector, and γ_{ij} is the spatial metric for the hypersurfaces. The Greek indices range from 0 to 3, whereas the Latin ones indicate the space-like hypersurfaces and range from 1 to 3. The lapse function α and the shift vector β^i are functions to be determined, γ_{ij} is a positive-definite metric on each of the space-like hypersurfaces, for all values of time, a feature that makes the spacetime globally hyperbolic. The lapse function $d\tau$ is a measure of proper time between two adjacent hypersurfaces by observers moving along the normal direction to the hypersurfaces (Eulerian observers). The shift vector is a tangent vector to the hypersurfaces that relates the spatial coordinate systems on different hypersurfaces.

Alcubierre assumed the following ad hoc particular choices for the parameters of the general ADM metric,[1]

$$\alpha = 1, \tag{2}$$

$$\beta^1 = -v_s(t) f\big[r_s(t)\big], \tag{3}$$

$$\beta^2 = \beta^3 = 0, \tag{4}$$

$$\gamma_{ij} = \delta_{ij}. \tag{5}$$

Rewriting Eq. (1) with these definitions yields the *Alcubierre WD spacetime metric* below,

$$ds^2 = -\left(1 - \beta^2\right) dt^2 - \beta\, dx\, dt + dx^2 + dy^2 + dz^2. \tag{6}$$

3. Einstein Equations Solutions for Simple Fluids

We to find solutions to the Einstein equations with the Alcubierre WD metric as the underlying geometry by coupling simple matter-energy distributions, namely the dust of particles, the perfect fluid and the anisotropic fluid with off-diagonal metric terms to see if those EMTs could lead to a viable WD bubble.

Starting with incoherent fluid, or dust, vacuum solutions of the Einstein equations for the WD metric (6) were recovered, which led to the connection of the WD with shock waves via Burgers equation.[26] Vacuum solutions mean that all the energy conditions are trivially satisfied, but the possibility of vacuum shock waves that could physically represent warp bubbles indicated that other known matter-energy distributions could result in more complex solutions of Einstein equations. Under this perception we proposed two other energy-momentum tensors, the perfect and "parametrized" fluids, where the latter is, in fact, an anisotropic fluid with heat flux.[27]

The EMT for an anisotropic and dissipative fluid may be written by the following equation,

$$T^{\alpha\beta} = \mu u^{\alpha} u^{\beta} + p h^{\alpha\beta} + u^{\alpha} q^{\beta} + u^{\beta} q^{\alpha} + \pi^{\alpha\beta}, \tag{7}$$

where

$$h_{\alpha\beta} = g_{\alpha\beta} + u_a \, u_b \tag{8}$$

projects tensors onto hypersurfaces orthogonal to u^{α}, μ is the matter density, p is the fluid static pressure, q^{α} is the heat flux vector and $\pi^{\alpha\beta}$ is the viscous shear tensor. The world lines of the fluid elements are the integral curves of the four-velocity vector u^{α}. The heat flux vector and the viscous shear tensor are transverse to the world lines, that is,

$$q_a \, u^a = 0, \qquad \text{and} \qquad \pi_{ab} \, u^b = 0 \, . \tag{9}$$

For simplicity we depicted the EMT of the anisotropic fluid in the following matrix form,

$$T_{\alpha\sigma} = \begin{pmatrix} \mu + \beta^2 p & -\beta D & 0 & 0 \\ -\beta D & A & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & C \end{pmatrix}, \tag{10}$$

where p, A, B and C are anisotropic static pressures, D is the momentum density parameter that represents the heat flux for this fluid, and μ is the particles' matter density. Notice that if we choose all pressures and momentum density as being

equal to the static pressure p we recover the perfect fluid form shown below,

$$T_{\alpha\sigma} = \begin{pmatrix} \mu + \beta^2 p & -\beta p & 0 & 0 \\ -\beta p & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} . \tag{11}$$

The EMT for a perfect fluid can be written in terms of the tensor notation as follows,

$$T_{\alpha\beta} = (\mu + p)\, u_\alpha u_\beta + p\, g_{\alpha\beta} \,, \tag{12}$$

where μ is the matter density, p is the fluid static pressure, $g_{\alpha\beta}$ is the metric tensor and u_α is the 4-velocity of an observer inside the fluid. Perfect fluids have no shear stress, rotation, heat conduction or viscosity. To recover the EMT for dust of particles, we have to choose the static fluid pressure to be zero, and the matrix form for the dust is given by the following equation,

$$T_{\alpha\sigma} = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , \tag{13}$$

which can be written in tensor notation as follows,

$$T_{\alpha\beta} = \mu\, u_\alpha u_\beta \,, \tag{14}$$

where μ is the matter density represented by a scalar function of the spacetime coordinates, and u^α are the observers 4-velocity components.

Table 1 lists the solutions of the Einstein equations for the anisotropic fluid defined by Eq. (10),[27] where the two solutions subsets 1b and 2b are considered unphysical because the resulting EMT leads to the anisotropic static pressures A and D equal to zero, whereas $B = C$. The energy density in the T_{00} component of the EMT is equal to zero, resulting in the following equation of state relating the matter density μ, the shift vector β and the static pressure p,

$$\mu = -\beta^2 p \,. \tag{15}$$

We must point out that we made a notational change on the shift vector definition β^1 so that it has the negative sign of the original parameter defined by Alcubierre,[1] yielding the shift vector β used in here as being given as below,

$$\beta = -\beta^1 = v_s(t) f\big[r_s(t)\big] \,. \tag{16}$$

Finally, one should also notice that solutions 1b and 2b brought about the following Burgers equation in dissipative form,

$$\frac{\partial}{\partial x}\left[\frac{\partial \beta}{\partial t} + \frac{1}{2}\frac{\partial}{\partial x}(\beta^2)\right] = -64\pi B \,. \tag{17}$$

Table 1. WD solutions of the Einstein equations with anisotropic fluid as source.

Case	Conditions	Results
1) $\dfrac{\partial\beta}{\partial z}=0$	1a) $\dfrac{\partial\beta}{\partial x}=0$	$\mu=\beta^2(2D-A-p)+\dfrac{A}{3}$ $\beta=\beta(t,y)$ $B=-C=\dfrac{A}{3}$ $\left(\dfrac{\partial\beta}{\partial y}\right)^2=32\pi C,$ $\dfrac{\partial^2\beta}{\partial y^2}=16\pi\beta(A-D)$
	1b) $\dfrac{\partial\beta}{\partial y}=0$	$\mu=-\beta^2 p$ $\beta=\beta(t,x)$ $B=C$ $A=D=0$ $\dfrac{\partial}{\partial x}\left[\dfrac{\partial\beta}{\partial t}+\dfrac{1}{2}\dfrac{\partial}{\partial x}(\beta^2)\right]=-64\pi B$ $\rightarrow$ solution *dismissed* as unphysical
2) $\dfrac{\partial\beta}{\partial y}=0$	2a) $\dfrac{\partial\beta}{\partial x}=0$	$\mu=\beta^2(2D-A-p)+\dfrac{A}{3}$ $\beta=\beta(t,z)$ $B=-C=\dfrac{A}{3}$ $\left(\dfrac{\partial\beta}{\partial z}\right)^2=32\pi C$ $\dfrac{\partial^2\beta}{\partial z^2}=16\pi\beta(A-D)$
	2b) $\dfrac{\partial\beta}{\partial z}=0$	$\mu=-\beta^2 p$ $\beta=\beta(t,x)$ $B=C$ $A=D=0$ $\dfrac{\partial}{\partial x}\left[\dfrac{\partial\beta}{\partial t}+\dfrac{1}{2}\dfrac{\partial}{\partial x}(\beta^2)\right]=-64\pi B$ $\rightarrow$ solution *dismissed* as unphysical

Perfect fluid solutions have equations of state given by the expression below,[30]

$$p = p(\mu) = (\gamma - 1)\mu \ , \tag{18}$$

where $\gamma =$ is a constant which for ordinary fluids can be approximated by $1 \leq \gamma \leq 2$. The incoherent matter, or dust, corresponds to $\gamma = 1$, and radiation corresponds to $\gamma = \frac{4}{3}$. For solutions 1b and 2b in Table 1 it is clear that γ is a function of the shift vector given by the following equation,

$$\gamma = 1 - \frac{1}{\beta^2} \ . \tag{19}$$

Notice that since the shift vector is a function of the warp velocity and it is also the regulating function, we are facing a problem with a discontinuity when $\beta = 0$ since outside the warp bubble the regulating function approaches to zero, but inside the warp bubble $f(r_s) = 1$. So Eq. (19) takes the following form,

$$\gamma = 1 - \frac{1}{v_s^2} \, , \tag{20}$$

where $v_s = v_s(t)$ is the warp bubble velocity. Solutions 1a and 2a shown in Table 1 are very similar, except that for solution 1a the shift vector is a function of spacetime coordinates (t, y), whereas solution 2a is a function of (t, z) coordinates. However, in both of these cases, the shift vector may be a complex-valued function depending on the values of the static pressure C. For these solutions the equations of state relating the matter density μ and the fluid pressures are identical and given by the following expression,

$$\mu = \beta^2(2D - A - p) + \frac{A}{3} \, . \tag{21}$$

The perfect fluid solutions listed in Table 2 are a special case of the anisotropic fluid if we chose all pressures to be equal to p. Table 3 presents the special dust case $(p = 0)$, which leads to vacuum.

4. Warp Drive and Shock Waves

In previous studies[26, 27] it was found an intrinsic relationship between the WD concept and shock waves via the Burgers equation arising from vacuum solutions of Einstein equations for the Alcubierre WD metric. Such a result may mean that the warp bubble could be interpreted as shock waves in vacuum. In other words, the Burgers equation can be seen as a vacuum solution of Einstein equations connecting the Alcubierre WD metric to shock waves.

In the case of the anisotropic fluid, we considered as unphysical the solution which led to a Burgers equation (17) because the parameter values resulted in an EMT with only two static pressures (diagonal terms) and no energy density. Nevertheless, for the perfect fluid and dust EMTs we found the Burgers equation

Table 2. WD solutions of the Einstein equations with perfect fluid as source.

Case	Condition	Results
1) $\frac{\partial\beta}{\partial z}=0$	$1a)$ $\frac{\partial\beta}{\partial x}=0$	$p=3\mu$ $\beta=\beta(y,t)$ $\frac{\partial\beta}{\partial y}=\pm\sqrt{-32\pi\mu}$
	$1b)$ $\frac{\partial\beta}{\partial y}=0$	$p=3\mu=0$ $\beta=\beta(x,t)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$
2) $\frac{\partial\beta}{\partial y}=0$	$2a)$ $\frac{\partial\beta}{\partial x}=0$	$p=3\mu$ $\beta=\beta(z,t)$ $\frac{\partial\beta}{\partial z}=\pm\sqrt{-32\pi\mu}$ $\frac{\partial\beta}{\partial z}=\pm\sqrt{\pm 96\pi\mu}$
	$2b)$ $\frac{\partial\beta}{\partial z}=0$	$p=3\mu=0$ $\beta=\beta(x,t)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$

Table 3. WD solutions of the Einstein equations with dust source (leads to vacuum).

Case	Consequence	Results
1) $\frac{\partial\beta}{\partial z}=0$	$\frac{\partial\beta}{\partial y}=0$	$\mu=0$ $\beta=\beta(t,x)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$
2) $\frac{\partial\beta}{\partial y}=0$	$\frac{\partial\beta}{\partial z}=0$	$\mu=0$ $\beta=\beta(t,x)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$

in its viscous form describing a dissipative system with the right-hand side being a function of time coordinate only, according to the expression below,

$$\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)\ . \tag{22}$$

Here $h = h(t)$ is an arbitrary function to be determined by the boundary conditions. If $h(t) = 0$ Eq. (22) takes its homogeneous form describing a conservative system known as inviscid Burgers equation.

The Burgers or Bateman-Burgers, equation is a well-known partial differential equation that models several physical systems such as gas dynamics, traffic flows, and even stock market, since, for the latter application, it is connected to the symmetries of the Black-Scholes equation. Its most notorious system description is, however, the phenomena arising from conservation laws and formation of shock waves, that is, discontinuities that appears after a finite time and then propagates in a regularly. The physics behind the field solutions of the Burgers equation can be seen as a current density.

The WD is depicted by the shift vector $\beta = v_s(t)f(r_s)$, where v_s is the bubble velocity, $f(r_s)$ is the regulating function of the warp bubble shape and the inviscid Burgers in Eq. (22) can be interpreted and representing a conservation law for this current density. Analyzing each term of Eq. (22), the first one on the l.h.s., $\partial\beta/\partial t$ can be interpreted as a force per unit mass, i.e., the time derivative of momentum. The second term on the l.h.s., $\frac{1}{2}\,\partial(\beta^2)/\partial x$, can be interpreted as the divergence of the total energy, which is entirely kinetic. Physically the result can be understood by considering the WD metric as conservation of both energy and momentum in the direction of the wave propagation. Calculating the divergence of the parametrized, or anisotropic, perfect fluid energy-momentum tensor, and demanding that it should be zero, one arrives at the following system,[27]

$$
\begin{aligned}
&-\frac{\partial\beta}{\partial x}(D+\mu)-\frac{\partial\mu}{\partial t}-\beta\left[\frac{\partial D}{\partial x}+\frac{\partial\mu}{\partial x}+\frac{\partial\beta}{\partial t}(2p+A-3D)\right]\\
&\qquad+\beta^2\left[\frac{\partial D}{\partial t}-\frac{\partial p}{\partial t}+3\frac{\partial\beta}{\partial x}(D-p)\right]+\beta^3\left(\frac{\partial D}{\partial x}-\frac{\partial p}{\partial x}\right)=0, \qquad (23)\\
&\frac{\partial A}{\partial x}+\frac{\partial\beta}{\partial t}(D-A)+\beta\left[3\frac{\partial\beta}{\partial x}(D-A)+\frac{\partial D}{\partial t}-\frac{\partial A}{\partial t}\right]\\
&\qquad+\ \beta^2\left(\frac{\partial D}{\partial x}-\frac{\partial A}{\partial x}\right)=0, \qquad (24)\\
&\frac{\partial B}{\partial y}+\beta\frac{\partial\beta}{\partial y}(D-A)=0, \qquad (25)\\
&\frac{\partial C}{\partial z}+\beta\frac{\partial\beta}{\partial z}(D-A)=0. \qquad (26)
\end{aligned}
$$

The perfect fluid zero divergence is recovered by letting

$$p = A = B = C = D\,, \qquad (27)$$

hence, Eqs. (23) to (26) become,

$$-\frac{\partial\mu}{\partial t}-\beta\left(\frac{\partial p}{\partial x}+\frac{\partial\mu}{\partial x}\right)=0\,, \tag{28}$$

$$\frac{\partial p}{\partial x}=\frac{\partial p}{\partial y}=\frac{\partial p}{\partial z}=0\,, \tag{29}$$

which means that the static pressure p of the perfect fluid does not depend on the spatial coordinates, and Eq. (28) reduces to the expression below,

$$\frac{\partial\mu}{\partial t}+\beta\frac{\partial\mu}{\partial x}=0\,, \tag{30}$$

which is a continuity equation, μ is the fluid density, and β is the flow velocity vector field. Notice that for constant density the fluid has an incompressible flow. Therefore, all partial derivatives of β in terms of the spatial coordinates vanish, and the flow velocity vector field has zero divergence, this being a classical fluid dynamics scenario. The local bubble volume expansion rate is zero, and the WD metric becomes the Minkowski metric in this scenario.

If we want to find these results for the dust of particles as a particular case, we can see that the partial time derivative of the matter-density is zero and we have

$$\mu\frac{\partial\beta}{\partial x}=0\,, \tag{31}$$

which is immediately satisfied since the matter density is zero for the dust EMT and the Einstein equations solution for the WD is the vacuum one. This is also true for the sets of solutions 1b and 2b found for the perfect fluid, as shown in Table 2. This result for the vacuum solutions of the Einstein equations with the WD metric as a background suggests that the necessary energy to create the associate shock wave is purely geometrical. It is interesting evidence that the WD metric can be understood as a spacetime motion equivalent to a shock wave moving in a fluid, where spacetime itself plays the role of the fluid.

Considering all these results one can speculate about vacuum energy, quantum fluctuations and dark matter as fuel for the feasibility of the warp bubble, but one has to overcome possible difficulties caused by the event horizon of the bubble that would require particles with imaginary mass like tachyons.[31]

5. The Cosmological Constant

Bearing in mind the results above led to the next step of adding the cosmological constant in the Einstein equation to seek solutions for Alcubierre WD spacetime geometry.[29] The motivation came from Eq. (19) in the original Alcubierre work,[1] which was obtained by applying the WEC considering Eulerian observers, contracting the EMT and writing the final expression $G_{\mu\nu}=\kappa T_{\mu\nu}$ in the Einstein equation, whose result reads as follows,

$$T^{\mu\nu}u_\mu u_\nu=\alpha^2T^{00}=G^{00}=-\frac{v_s^2\rho^2}{4r_s^2}\left(\frac{df}{dr_s}\right)^2. \tag{32}$$

Here u^μ and u_μ are respectively the contravariant and covariant components of the four-vector velocity of the Eulerian observers given by

$$u^\mu = \frac{1}{\alpha}(1, \beta^1, \beta^2, \beta^3) \quad , \quad u_\mu = -(\alpha, 0, 0, 0)\,, \tag{33}$$

where α is the lapse time between the constant hypersurfaces of the ADM-formalism and β^i are the three-components of the shift vector $\boldsymbol{\beta}$. Using the original parameters from the $3+1$-formalism, its general metric and the chosen chart of coordinates we can write Eq. (32) as below,

$$T_{\alpha\beta}\, u^\alpha u^\beta = -\frac{1}{4}\left[\left(\frac{\partial\beta}{\partial y}\right)^2 + \left(\frac{\partial\beta}{\partial z}\right)^2\right]. \tag{34}$$

The above expression is clearly a non-positive term everywhere, violating the WEC in a nontrivial way. We then showed that if we begin with simple energy and momentum sources and impose Eq. (34) to be as identically zero we arrive at vacuum solutions for the WD spacetime.[26,27] However, if we include the cosmological constant in the Einstein equations the expression (34) is changed, yielding,

$$T_{\alpha\beta}\, u^\alpha u^\beta = \Lambda - \frac{1}{4}\left[\left(\frac{\partial\beta}{\partial y}\right)^2 + \left(\frac{\partial\beta}{\partial z}\right)^2\right] \ . \tag{35}$$

For positive and large enough values of the cosmological constant, it may be possible to construct a WD that does not violate the WEC, because the original negative mass-energy density necessary to create the warp bubble could become positive.

The solutions for the WD metric which include the cosmological constant in the Einstein equations[29] are presented in Table 4. Cases 1b and 2b are just the vacuum solutions connecting the WD spacetime to shock waves via the Burgers equations, and they also require a vanishing cosmological constant, since both the static fluid pressure (p) and the matter density (μ) also vanish. The other two solutions subsets 1a and 2a are also similar to the equally labeled ones in Table 2, but now they include the cosmological constant and the solutions are no longer required to be complex if the following conditions are satisfied,

$$\Lambda - 8\pi\mu \geq 0\,, \tag{36}$$

$$\Lambda - 8\pi p \geq 0\,. \tag{37}$$

Taking all these results into account imply that creating a warp bubble for superluminal travel of massive particles seem to require more complex forms of matter than the dust for stable solutions, and that the requirement of negative mass-energy may not be as strict as originally thought for superluminal spacetime travel with warp speeds. The shift vector in the direction of the warp bubble movement creates a coupling in Einstein equations that requires the off-diagonal source terms in the energy-momentum tensor that represent momentum densities.

Previous authors investigated the WD theory from the geometrical point of view[6,7,10–12] and came up with other types of WD mechnics, but it now seems clear that it is also necessary to consider both the energy and momentum tensors to propose a superluminal propelling system in the WD theory.

Table 4. WD solutions of the Einstein equations with Λ and perfect fluid.

Case	Condition	Results
1) $\frac{\partial\beta}{\partial z}=0$	1a) $\frac{\partial\beta}{\partial x}=0$	$\Lambda=6\pi\left(\mu-\frac{p}{3}\right)$ $\beta=\beta(y,t)$ $\frac{\partial\beta}{\partial y}=\pm\sqrt{4(\Lambda-8\pi\mu)}$ $\frac{\partial\beta}{\partial y}=\pm\sqrt{\frac{4}{3}(\Lambda-8\pi p)}$ $\beta\frac{\partial\mu}{\partial x}+\frac{\partial\mu}{\partial t}=0$ (null divergence)
	1b) $\frac{\partial\beta}{\partial y}=0$	$\Lambda=8\pi\mu=8\pi p=0$ $\beta=\beta(x,t)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$ Null divergence is trivially satisfied This is the solution found in Ref.[26]
2) $\frac{\partial\beta}{\partial y}=0$	2a) $\frac{\partial\beta}{\partial x}=0$	$\Lambda=6\pi\left(\mu-\frac{p}{3}\right)$ $\beta=\beta(y,t)$ $\frac{\partial\beta}{\partial z}=\pm\sqrt{4(\Lambda-8\pi\mu)}$ $\frac{\partial\beta}{\partial z}=\pm\sqrt{\frac{4}{3}(\Lambda-8\pi p)}$ $\beta\frac{\partial\mu}{\partial x}+\frac{\partial\mu}{\partial t}=0$ (null divergence)
	2b) $\frac{\partial\beta}{\partial z}=0$	$\Lambda=8\pi\mu=8\pi p=0$ $\beta=\beta(x,t)$ $\frac{\partial\beta}{\partial t}+\frac{1}{2}\frac{\partial}{\partial x}(\beta^2)=h(t)$ Null divergence is trivially satisfied This is the solution found in Ref.[26]

6. Conclusions

In this work, we described new exact solutions of the Einstein equations endowed with the Alcubierre warp drive (WD) spacetime geometry having simple energy-momentum tensor (EMT) distributions of matter and energy. The incoherent matter, or dust, the simplest fluid source EMT used with the WD metric, led to vacuum solutions and the weak energy condition being satisfied with null energy density. In addition, we showed that the dynamics of the vacuum WD spacetime are governed by the Burgers equation, which links the WD bubble to shock waves.[26] Both the perfect fluid EMT and an anisotropic fluid recover the Burgers equation as vacuum solutions, but some of these new sets of solutions were considered unphysical, and others showed that the shift vector can be a complex-valued function.[27] The Einstein equations with the cosmological constant were used with the WD metric having the perfect fluid as the source, whose results showed that if the cosmological constant is big enough it would be possible to attain positive energy density for the WD spacetime.[29] This result is in line with recent research activity in WD theory by other authors, who provide examples of different WD spacetimes geometries,[17–19] which suggests that it may be possible to create warp bubbles with positive energy densities.

References

1. M. Alcubierre, *The warp drive: hyper-fast travel within general relativity*. Class. Quant. Grav. 11 (1994) L73, arXiv:gr-qc/0009013.
2. R. Arnowitt, S. Deser and C.W. Misner. *Dynamical Structure and Definition of Energy in General Relativity*, Phys. Rev. 116 (1959) 1322.
3. R. Arnowitt, S. Deser and C.W. Misner. *The Dynamics of General Relativity. Gravitation: An introduction to current research*, L. Witten, (ed.), Wiley, NY, 1962. Reprinted in arXiv:gr-qc/0405109.
4. L.H. Ford and T.A. Roman, *Quantum Field Theory Constrains Traversable Wormhole Geometries*, Phys. Rev. D 53 (1996) 5496, arXiv:gr-qc/9510071.
5. M.J. Pfenning and L.H. Ford *The unphysical nature of Warp Drive*, Class. Quant. Grav. 14 (1997) 1743, arXiv:gr-qc/9702026.
6. S.V. Krasnikov *Hyperfast Interstellar Travel in General Relativity*, Phys. Rev. D 57 (1998) 4760, arXiv:gr-qc/9511068.
7. A. Everett and T.A. Roman, *A Superluminal Subway: The Krasnikov Tube*, Phys. Rev. D 56 (1997) 2100, arXiv:gr-qc/9702049.
8. F.S.N. Lobo and P. Crawford, *Weak Energy Condition Violation and Superluminal Travel*, Lect. Notes Phys. 617 (2003) 277, arXiv:gr-qc/0204038.
9. K. Olum, *Superluminal travel requires negative energy density*, Phys. Rev. Lett, 81 (1998) 3567, arXiv:gr-qc/9805003.
10. C. Van Den Broeck, *A warp drive with more reasonable total energy*, Class. Quant. Grav. 16 (1999) 3973, arXiv:gr-qc/9905084.
11. J. Natário, *Warp Drive With Zero Expansion*, Class. Quant. Grav. 19 (2002) 1157, arXiv:gr-qc/0110086.
12. F.S.N. Lobo and M. Visser, *Linearized warp drive and the energy conditions*, arXiv:gr-qc/0412065.

13. H.G. White, *A Discussion of Space-Time Metric Engineering*, Gen. Relat. Grav. 35 (2003) 2025.
14. H.G. White, *Warp Field Mechanics 101*, J. Brit. Interplanetary Society 66 (2011) 242.
15. J. Lee and G. Cleaver, *Effects of External Radiation on an Alcubierre Warp Bubble*, Physics Essays 29 (2016) 201.
16. B. Mattingly, A. Kar, M. Gorban, W. Julius, C. Watson, M.D. Ali, A. Baas, C. Elmore, J. Lee, B. Shakerin, E. Davis, and G. Cleaver, *Curvature Invariants for the Accelerating Natário Warp Drive*, Particles 3 (2020) 642, arXiv:2008.03366.
17. A. Bobrick and G. Martire, *Introducing physical warp drives*, Class. Quant. Grav. 38 (2021) 105009, arXiv:2102.06824.
18. E.W. Lentz, *Breaking the warp barrier: Hyper-fast solitons in Einstein-Maxwell-plasma Theory*, Class. Quant. Grav. 38 (2021) 075015, arXiv:2006.07125.
19. S.D.B. Fell and L. Heisenberg, *Positive energy warp drive from hidden geometric structures*, Class. Quant. Grav. 38 (2021) 155020, arXiv:2104.06488.
20. C.J. Quarra, *Creating spacetime shortcuts with gravitational waveforms*, arXiv:1602.01439.
21. J. Santiago, S. Schuster and M. Visser, *Generic warp drives violate the null energy condition*, arXiv:2105.03079.
22. J. Santiago, S. Schuster and M. Visser, *Tractor beams, pressor beams and stressor beams in General Relativity*, Universe 7 (2021) 271, arXiv:2106.05002.
23. M. Alcubierre and F.S.N. Lobo, *Wormholes, Warp Drives and Energy Conditions*, Fundam. Theor. Phys. 189 (2017) 279, arXiv:2103.05610.
24. B. Shoshany, *Lectures on Faster-than-Light Travel and Time Travel*, SciPost Phys. Lect. Notes, 10 (2019), arXiv:1907.04178.
25. H. White, J. Vera, A. Han. A.R. Bruccoleri and J. MacArthur, Eur. Phys. J. C 81 (2021) 677.
26. O.L. Santos-Pereira, E.M.C. Abreu and M.B. Ribeiro, *Dust content content solutions for the Alcubierre warp drive spacetime*, Eur. Phys. J. C 80 (2020) 786, arXiv:2008.06560.
27. O.L. Santos-Pereira, E.M.C. Abreu and M.B. Ribeiro, *Fluid dynamics in the warp drive spacetime geometry*, Eur. Phys. J. C 81 (2021) 133, arXiv:2101.11467.
28. O.L. Santos-Pereira, E.M.C. Abreu and M.B. Ribeiro, *Charged dust solutions for the warp drive spacetime*, Gen. Relat. Gravit. 53 (2021) 23, arXiv:2102.05119.
29. O.L. Santos-Pereira, E.M.C. Abreu and M.B. Ribeiro, *Perfect fluid warp drive solutions with the cosmological constant*, Eur. Phys. J. Plus 136 (2021) 902, arXiv:2108.10960.
30. G.F.R. Ellis and H. van Elst, *Cosmological Models*, (Cargése Lectures 1998), NATO Adv. Study Inst. Ser. C. Math. Phys. Sci. 541 (1999) 1, arXiv:gr-qc/9812046.
31. M. Alcubierre, talk at the 16th Marcel Grossmann Meeting: AT3-Wormholes, Energy Conditions and Time Machines (7 July 2021). Available at `https://youtu.be/NqN1c-2fv8Y?t=784`.

Symmetries and geometry of spacetime: Towards a new paradigm

Francisco Cabral[1]

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, P-1749-016 Lisbon, Portugal
[1] *Email:ftcabral@fc.ul.pt*

Francisco S. N. Lobo[2]

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, P-1749-016 Lisbon, Portugal
[2] *Email:fslobo@fc.ul.pt*

Diego Rubiera-Garcia[3]

Departamento de Física Teórica and IPARCOS, Universidad Complutense de Madrid, E-28040 Madrid, Spain
[3] *Email:drubiera@ucm.es*

In this work the geometrical methods and symmetry principles in gravitation are explored motivating a new perspective into the spacetime paradigm. The effects of post-Riemann spacetime geometries with torsion are briefly studied in applications to fundamental fermionic and bosonic fields, cosmology and gravitational waves. The physical implications and related phenomenological considerations are addressed, and the fundamental ideas related to spacetime physics, motivated by geometrical methods and symmetry principles, are also briefly discussed in the context of the possible routes towards a new spacetime paradigm in gravitation and unified field theories

Keywords: Riemann-Cartan spacetime, torsion, fermionic systems, gravitational waves

1. Introduction

The success of Einstein's General Theory of Relativity (GR) to describe the gravitational interaction is quite remarkable. GR has passed all tests performed so far: Solar System observations and binary pulsars,[1] stellar orbits around the central galactic black hole,[2] gravitational waves (GWs) from coalescing compact objects (black holes and neutron stars),[3–6] or the indirect observation of the black hole horizon with the Event Horizon Telescope,[7] among others. At the same time, GR provides us with the observationally valid framework for the standard cosmological paradigm when supplemented with the (cold) dark matter and dark energy hypothesis.[8]

Soon after GR theory, Weyl (1918) introduced the notion of gauge transformations, in an attempt to unify gravity and electromagnetism.[9] By extending the Local Lorentz group to include scale transformations (dilatations) he was led to assume what we call a Riemann-Weyl spacetime geometry, a post-Riemann geometry with (the trace-vector part of) non-metricity beyond curvature. The theory was then

abandoned and Weyl (1929) clarified that the electromagnetic field is intimately related to local internal symmetries, under the U(1) group that act on the 4-spinor fields of charged matter.[10] In the mid 1950's Yang (1954) and Mills (1954)[11] further explored the notion of gauge symmetries in field theories going beyond the $U(1)$ group to include non-abelian Lie groups ($SU(2)$), in order to address nuclear physics, while Utiyama (1956)[12] extended the gauge principle to all semi-simple Lie groups including the Lorentz group. The gauge principle is based on the localization of the rigid, global symmetry group of a field theory, introducing a new interaction described by the gauge potential. The latter is a compensating field that makes it possible for the matter Lagrangian to be locally invariant under the symmetry group and is included in the covariant derivative of the theory. There is a clear geometrical interpretation of the gauge potential as the connection of the fiber bundle, which is the manifold obtained from the base spacetime manifold and the set of all fibers. These are attached at each spacetime point and are the (vector, tensor or spinor) spaces of representation of the local symmetries. In the geometrical interpretation, the imposition of local symmetries implies that the geometry of the fiber bundle is non-Euclidean, and the gauge field strengths are the curvatures of such a manifold.

The gauge formulation of gravity was resumed through the works of Kibble (1961)[13] and Sciama[14] (1964), who gauged the (rigid) Poincaré group of Minkowski spacetime symmetries. This can be viewed as the starting point of a self-consistent gauge theory of gravity. They arrived at what is now known as a Riemann-Cartan (RC) geometry, and to the corresponding Einstein-Cartan-Sciama-Kibble (ECSK) gravity, with non-vanishing torsion and curvature. This is a natural extension of GR, which is able to successfully incorporate the intrinsic spin of fermions as a source of gravity, while passing all weak-field limit tests. The theory has no free parameters but introduces a new scale given by Cartan's density, which yields many relevant applications in cosmology and astrophysics.[15–23] The ECSK is the simplest of all Poincaré gauge theories of gravity (PGTG) in RC spacetime. Beyond the Poincaré group we have, for instance, the Weyl and the conformal groups, which live on a subset of a general metric-affine geometry, with non-vanishing curvature, torsion and non-metricity (for a detailed analysis and reviews on several topics of the gauge approach to gravity see the remarkable works in[24–26]). *By extending the gauge symmetry group of gravity, one is naturally led to extend the spacetime geometry paradigm as well.*

Just as quantum field theories of the standard model are gauge theories with (local) internal symmetries, similarly, classical theories of gravity can indeed be formulated as gauge theories, leading to non-Euclidean geometries. Therefore, the gauge principle (with its geometrical methods) is a convenient formalism in order to address the possible avenues towards unified field theories of matter, gravity and spacetime geometry. The spacetime paradigm changes when we extend the gauge group of gravity and the history of Physics have shown that changing the fundamental ideas about space and time is at the basis of major breakthroughs. Gauge

methods in gravity and post-Riemann geometries deserve further developments with potential applications in astrophysical and cosmological scenarios with very strong gravitational fields. Above all, they are a vital part of the effort to understand the nature of spacetime and gravitation.

In this paper we review the fundamental geometrical objects and its relation to spacetime symmetries in the gauge formulation of gravity, and consider specific applications. We will consider the structure of gauge theories of gravity and have a special focus on PGTG, in particular we will consider some of its predictions and implications including interactions with fermions and bosons, cosmological and gravitational wave applications.

2. The gauge approach to gravity

The gauge approach to the gravitational interaction is based on the localization of the global (rigid) spacetime symmetries of the matter field Lagrangian, under a certain group of spacetime coordinate transformations. In this procedure, the gauge potentials, the field strengths and the Noether currents are clearly identified. The gauge potentials are then interpreted as the geometrical degrees of freedom representing the gravitational interaction, while the corresponding field strengths correspond to well identified objects from differential geometry such as curvature, torsion or non-metricity. The Noether currents associated to the spacetime symmetries represent the sources of the gravitational field and provide also valuable information on the physical properties of matter that couple to gravity, clarifying the appropriate test matter for the probing of spacetime geometry. One of the main lessons from the gauge approach to gravity is the fact that the requirement of local, non-rigid symmetries in a matter field theory under spacetime transformations, leads naturally to non-rigid spacetime geometries, meaning general non-Euclidean geometries with curvature, torsion and non-metricity.

2.1. *From symmetries to spacetime geometry*

The gauge approach to gravity broadens our study of the deep relation between symmetry principles (group theory) and geometrical methods. In this framework, the Metric-Affine Gravity (MAG) formalism,[27] where the metric and the affine connection are both independent degrees of freedom, becomes clearly formulated. In fact, the MAG approach includes a general (metric-affine) geometry with curvature, torsion and non-metricity, that arises naturally from the localization (gauging) of the Affine group A(4,$\Re$) and taking into consideration also a metric structure (although the metric structure can be considered not to be "fundamental" in the so-called pre-metric approaches to field theories).

The Affine group A(4,$\Re$) is the (semi-direct) product of the group of 4-translations T(4) with the "General Linear group" GL(4,$\Re$). The gauge potentials

of T(4) are the vector valued 1-form tetrads (or covector frames)

$$\theta^a = \theta^a{}_\mu dx^\mu, \tag{1}$$

where $\{dx^0, dx^1, dx^2, dx^3\}$ is the holonomic co-vector frame basis, with the vector valued 2-form torsion as the field strength

$$T^a = D\theta^a = d\theta^a + \Gamma^a{}_b \wedge \theta^b, \tag{2}$$

while the gauge potentials of GL(4,$\Re$) are the 2-tensor valued linear connection 1-form fields

$$\Gamma^a{}_b = \Gamma^a{}_{b\mu} dx^\mu, \tag{3}$$

and the corresponding field strength is the curvature, 2-tensor valued 2-form, given by

$$R^a{}_b = d\Gamma^a{}_b + \Gamma^a{}_c \wedge \Gamma^c{}_b. \tag{4}$$

In these expressions D is the (gauge) covariant exterior derivative, d is the exterior derivative, a kind of curl operator that raises the degree of any p-form, $\wedge$ is the wedge product. The torsion 2-form has $4 \times 6 = 24$ independent components

$$T^a{}_{\mu\nu} = 2\partial_{[\mu}\theta^a{}_{\nu]} + 2\Gamma^a{}_{c[\mu}\theta^c{}_{\nu]}, \tag{5}$$

while the curvature 2-form has $16 \times 6 = 96$ independent components

$$R^a{}_{b\mu\nu} = 2\partial_{[\mu}\Gamma^a{}_{b|\nu]} + 2\Gamma^a{}_{c[\mu}\Gamma^c{}_{b|\nu]}. \tag{6}$$

The presence of the linear connection, which is the gauge potential associated to curvature, in the last term of the expression for the torsion in (2) is of high importance in the context of the Poincaré gauge gravity and MAG, as will be discussed. In MAG the Noether currents (the sources of gravity) are the canonical energy-momentum (related to T(4)) and the hyper-momentum (related to GL(4,$\Re$)) which includes a spin part s_{ab}, a term related to dilatations and another connected to shear type of "deformations" $\bar{\Delta}_{ab}$. The hypermomentum can then be decomposed according to

$$\Delta_{ab} = s_{ab} + \frac{1}{4} g_{ab} \Delta^c{}_c + \bar{\Delta}_{ab}\ , \tag{7}$$

including the spin $s_{ab} = -s_{ba}$, the dilatation $\Delta^c{}_c$, and the shear $\bar{\Delta}_{ab}$ currents. The Noether currents are not only the fundamental sources of gravity in MAG, but also represent the physical properties of the appropriate test matter to probe the spacetime geometry.

We can also introduce the Lorentzian metric as the (0,2) tensor $g = g_{ab}\theta^a \otimes \theta^b$, where $a, b = 0, 1, 2, 3$ are anholonomic indices, which is the metric (with constant components) of the tangent space, while the spacetime metric $g_{\mu\nu}$ gives the inner products $g(u, v) = g_{\alpha\beta} u^\alpha v^\beta = u_\alpha v^\beta$ and is given by $g_{\mu\nu} = \theta^a{}_\mu \theta^a{}_\nu g_{ab}$. Therefore, the spacetime metric can be seen as the deformation of the Lorentzian (tangent space)

metric according to $g_{\mu\nu} = \Omega^{ab}{}_{\mu\nu} g_{ab}$, with $\Omega^{ab}{}_{\mu\nu} \equiv \theta^a{}_\mu \theta^a{}_\nu$ being a deformation tensor.

Besides the linear co-frame (tetrads) and the linear connection, the 0-form metric g_{ab} can then be introduced as a kind of potential and the corresponding field strength is the tensor valued 1-form $Q_{ab} = Q_{ab\mu} dx^\mu$, given by

$$Q_{ab} = D g_{ab} = d g_{ab} + \Gamma^c{}_a \wedge g_{cb} + \Gamma^c{}_b \wedge g_{ac} = 2\Gamma_{(ab)}, \tag{8}$$

with $10 \times 4 = 40$ independent components $Q_{ab\mu}$. This field strength corresponds to the non-metricity tensor valued 1-form and one concludes that if the connection is non-Lorentzian ($\Gamma^{ab} \neq -\Gamma^{ba}$), then the non-metricity is non-vanishing. The linear connection 1-form can be decomposed according to

$$\Gamma_{ab} = \tilde{\Gamma}_{ab} + N_{ab} = \tilde{\Gamma}_{ab} + N_{[ab]} + \frac{1}{2} Q_{ab} \ , \tag{9}$$

where $\tilde{\Gamma}_{ab}$ is the Levi-Civita part of the connection and $N^a{}_b$ is the so-called distortion 1-form characterizing the post-Riemannian geometries. In particular, one finds that $Q_{ab} = 2N_{(ab)}$ and $T^a = N^a{}_b \wedge \theta^b$. In the most general case, the connection is non-Lorentzian (which implies the presence of non-metricity in the spacetime geometry) and is linked to gauge symmetry groups that extend the Poincaré group, leading therefore to the breaking of Lorentz invariance and to the physics beyond the standard model. If one restricts the symmetry group to the case of PGTG, then the linear connection is Lorentzian (a spin connection) and the hypermomentum is restricted to the spin current. Therefore the presence of non-metricity is related to the breaking of Lorentz invariance, non-Lorentzian connections and hypermomentum with non-trivial currents beyond the spin part.

2.2. *Metric-Affine gravity revisited*

Let us now summarize the fundamental structures of spacetime and their relation to symmetry groups. We have the fundamental 1-forms, the linear co-frame θ^a and the linear connection $\Gamma^a{}_b$, which are the potentials for the 4-dimensional translations $T(4)$ and the general linear $GL(4, \Re)$ groups, respectively. The corresponding field strengths correspond to well known objects from differential geometry, namely, the torsion $T^a = D\theta^a = d\theta^a + \Gamma^a{}_b \wedge \theta^b$ and curvature $R^a{}_b = d\Gamma^a{}_b + \Gamma^a{}_c \wedge \Gamma^c{}_b$ 2-forms, respectively. The metric is introduced as the 0-form potential with the corresponding field strength being the non-metricity 1-form $Q_{ab} = D g_{ab}$. The linear frame establishes a link between the (symmetries on the) local tangent fibers and the spacetime manifold, while the linear connection can be viewed as a guidance field reflecting the inertial character for matter fields propagating on the spacetime manifold. Finally, the metric allows the determination of spatial and temporal distances and angles. The spacetime geometry with all these structures is called a *metric-affine geometry*, with non-vanishing torsion, curvature and non-metricity, and its fundamental

local group of spacetime symmetries is the affine group $A(4,\Re) = T(4) \rtimes GL(4,\Re)$. As one can see from the expressions for the field strengths, torsion and curvature, in (2) and (4), respectively, the connection potential also enters in the expression for the field strength of the co-frame potential. This term is unavoidably present and is due to the semi-direct product[a] structure of the Poincaré group (or the affine group) and, therefore, curvature and torsion are somehow intertwined.

In the self-consistent metric-affine formalism, the gravitational interaction is described as a gauge theory of the affine group $A(4,\Re)$, together with the assumption of a metric, with the potentials (θ^a, Γ^{ab}) coupled to the corresponding Noether currents $(\tau^a, \Delta^a{}_b)$. The latter are the vector-valued canonical energy-momentum $\tau^a = \delta\mathcal{L}_{mat}/\delta\theta_a$ and the tensor-valued hypermomentum $\Delta^a_b = \delta\mathcal{L}_{mat}/\delta\Gamma_a{}^b$ 3-form currents[b], respectively, while the metric g_{ab} couples to the symmetric (Hilbert) energy momentum $T_{ab} = 2\delta\mathcal{L}_{mat}/\delta g^{ab}$.

A truly independent connection in MAG can be written as

$$\Gamma^{ab} = \Gamma^{[ab]} + \frac{1}{4} g^{ab}\Gamma^c{}_c + \left(\Gamma^{(ab)} - \frac{1}{4} g^{ab}\Gamma^c{}_c\right), \tag{10}$$

including the Lorentzian piece $\Gamma^{[ab]}$, the trace part $\frac{1}{4} g^{ab}\Gamma^c{}_c$ and the shear part $\Gamma^{(ab)} - \frac{1}{4} g^{ab}\Gamma^c{}_c$. The Lorentzian connection couples to the spin current and the trace and shear parts couple to the dilatation and shear currents, respectively. As previously said it is the non-Lorentzian part of the connection that imply the non-vanishing of non-metricity, therefore, the dilatation and shear hypermomentum currents are intimately related to the non-metricity of metric affine spacetime geometry.

The variational principle is applied to the action of this theory (including the gravitational part and the matter Lagrangian) by varying it with respect to the gauge potentials of the affine group, (θ^a, Γ^{ab}). This leads to two sets of dynamical equations, while a third set of equations is obtained by varying the action with respect to the metric potential g_{ab}. At the end, the dynamics is described only via two sets of equations, since the gravitational equation obtained by varying with respect to the metric potential or the one obtained by variation with respect to the translational potential can be dropped out, as long as the other gravitational equation derived from variation with respect to the linear connection is fulfilled (for details we recommend the reader to see[25]). This procedure and the fundamental quantities and relations here exposed summarizes the basics of the MAG formalism.

---

[a]The semi direct product implies that the generators of $T(4)$ and $GL(4,\Re)$ (or $SO(1,3)$) do not commute.

[b]These currents can be represented as 1-forms or as 3-forms. In fact, in the gauge approach to gravity they emerge naturally from Noether equations as 3-forms, being natural objects for integration over volumes. As 1-forms one can write $\tau^a = \tau^a_\mu dx^\mu$ and $\Delta^a{}_b = \Delta^a{}_{b\mu} dx^\mu$.

2.3. *Poincaré gauge theories and its extensions*

The description of the gravitational interaction as a gauge theory opens the door for its inclusion in unifying gauge field theories. Indeed, the analogies with gauge Yang-Mills theories can be easily established, (see[28]). Particularly relevant is of course the PGTG that constitutes a valid and quite promising class of models for an appropriate description of classical gravity including post-Einstein strong gravity predictions. The PGTG arises from restricting the symmetry group to the Poincaré group $P(1,3) = T(4) \rtimes SO(1,3)$, and this leads directly to RC spacetime geometry. If only the translational part is taken as symmetry group, one gets the translational gauge theories (TGTG) that include non-vanishing torsion but zero curvature and non-metricity, while both Weyl(-Cartan) gauge gravity (WGTG) and conformal gauge gravity (CGTG) include curvature, torsion and non-metricity. For the 15-parametric conformal group $C(1,3)$ a generalization of Kibble's gauge procedure is required, since although locally $C(1,3)$ is isomorphic to $SO(2,4)$ its realization in M_4 (Minkowski spacetime) is non-linear.[29] The PGTG can further be extended into the de Sitter or anti-de Sitter (A)dS gauge theories of gravity by localizing the $SO(1,4)$ or $SO(2,3)$ groups, respectively. Another important class of extensions requires going beyond the Lie algebra (by considering also anticommutators), in order to arrive at the super-Poincaré group[30] proper supersymmetry (SUSY) transformations between fermions and bosons. The gauging of these (and other) super-algebras lead directly supergravity (SUGRA), that need to take into account post-Riemannian spacetime geometries.

In PGTG the tetrads and the spin connection 1-forms are the gauge potentials, associated with translations, $T(4)$, and Lorentz rotations, $SO(1,3)$, respectively. Torsion and the curvature 2-forms are the respective field strengths. There are 6 generators in the Lorentz group with the corresponding potentials ($\Gamma^{ab} = -\Gamma^{ba}$) and 6 Noether (spin) currents ($s^{ab} = -s^{ba}$). Analogously, there are 4 generators in the group of spacetime translations, which entail 4 gauge potentials θ^a and 4 conserved Noether (canonical energy-momentum) currents τ^a. By constructing the gravitational Lagrangian with the curvature and torsion invariants, the potentials are coupled to the Noether currents via $24 + 16 = 40$ second order field equations. Let us briefly summarize the PGTG formalism. We consider a physical system, represented by the matter fields Lagrangian $\mathcal{L}_m = \mathcal{L}_m(g_{ab}, \theta^c, D\Psi)$. The covariant derivative with respect to the Riemann-Cartan connection $D\Psi$ allows the Lagrangian to be invariant under local Poincaré spacetime transformations. The canonical energy-momentum tensor density (equivalent to the dynamical tetrad energy-momentum density $\tau_a \equiv \delta\mathcal{L}_m/\delta\theta^a$) and the canonical spin density (equivalent to the dynamical spin density $s_{ab} \equiv \delta\mathcal{L}_m/\delta\Gamma^{ab}$) couple to the gravitational potentials, acting as sources of gravity and obey generalized conservation equations. The gravity sector in the action is constructed with the gauge-invariant gravitational field strengths in the kinetic part associated with the dynamics of the gravitational

degrees of freedom. The total Lagrangian density thus reads

$$\mathcal{L} = \mathcal{L}_G(g_{ab}, \theta^a, T^a, R^{ab}) + \mathcal{L}_m(g_{ab}, \theta^a, D\Psi) \ . \tag{11}$$

By varying this action with respect to the gauge fields of gravity (θ^a, Γ^{ab}) and the matter fields Ψ, we get the corresponding field equations.

2.4. *Applications in PGTG*

Without going into details we will briefly mention here some results from simple applications of PGTG, namely in the case of the ECSK theory with fermions and $(s = 1)$ bosons and some of its extensions. The reader can find details in the references.[31–33]

2.4.1. *Einstein-Cartan-Dirac-Maxwell dynamics*

Considering in the ECSK action

$$S_{\rm EC} = \frac{1}{2\kappa^2}\int d^4x\sqrt{-g}R(\Gamma) + \int d^4x\sqrt{-g}\,\mathcal{L}_m \tag{12}$$

the matter Lagrangian

$$\mathcal{L}_m = \mathcal{L}_{\rm D} + \mathcal{L}_{\rm M} + j^\mu A_\mu, \tag{13}$$

where $\mathcal{L}_{\rm D}$ and $\mathcal{L}_m$ are the Lagrangians for a Dirac spinor field and a bosonic $(s = 1)$ field minimaly coupled to the spacetime torsion, respectively, and the Ricci scalar is computed from the connection of RC spacetime geometry, the full dynamics for the matter and geometrical degrees of freedom can be derived. This results in the non-linear dynamics in the matter sector with both the non-linearities and non-minimal couplings between fermions and bosons induced by torsion, and in the generalized Einstein and Cartan equations for the metric and torsion/contorsion variables. For example, varying the Dirac Lagrangian minimally coupled to torsion[c] (in these expressions all tilde quantities correspond to the case of pseudo-Riemann geometry)

$$\mathcal{L}_{\rm Dirac} = \tilde{\mathcal{L}}_{\rm Dirac} + \frac{i\hbar}{4}K_{\alpha\beta\mu}\bar{\psi}\gamma^{[\mu}\gamma^{\alpha}\gamma^{\beta]}\psi, \tag{14}$$

and the bosonic Lagrangian minimally coupled to torsion,

$$\mathcal{L}_{\rm Max} = \tilde{\mathcal{L}}_{\rm Max} + \lambda\left(K^{\lambda[\mu\nu]}K^{\gamma}{}_{[\mu\nu]}A_\gamma + K^{\lambda[\mu\nu]}\tilde{F}_{\mu\nu}\right)A_\lambda, \tag{15}$$

with respect to the corresponding spinor and vector fields, and substituting the torsion as a function of the matter fields (Cartan equations), one obtains the generalized Dirac and Maxwell dynamics

$$\begin{aligned} i\hbar\gamma^\mu\tilde{D}_\mu\psi + (q\gamma^\mu A_\mu - m)\,\psi = f(A)(\bar{\psi}\gamma^\nu\gamma^5\psi)\gamma_\nu\gamma^5\psi + \alpha^\lambda{}_\alpha(A)(\bar{\psi}\gamma^\alpha\gamma^5\psi)\gamma_\lambda\gamma^5\psi \\ +\beta^\lambda(A,\tilde{F})\gamma_\lambda\gamma^5\psi \ , \end{aligned} \tag{16}$$

[c]We consider the torsion tensor as $T^\lambda{}_{\alpha\beta} \equiv \Gamma^\lambda_{[\alpha\beta]}$ and the contortion given by $K^\lambda{}_{\mu\nu} \equiv T^\lambda{}_{\mu\nu} - 2T_{(\mu}{}^\lambda{}_{\nu)} = T^\lambda{}_{\mu\nu} + 2T_{(\mu\nu)}{}^\lambda$.

and

$$\tilde{\nabla}_\mu \tilde{F}^{\mu\nu} = \lambda^{-1}(j^\nu + J^\nu) \ , \tag{17}$$

where

$$J^\nu \equiv \frac{\partial \mathcal{L}_{\rm M}^{corr}}{\partial A_\nu} - \tilde{\nabla}_\mu \left(\frac{\partial \mathcal{L}_{\rm M}^{corr}}{\partial(\tilde{\nabla}_\mu A_\nu)} \right) ,$$

with $\mathcal{L}_{\rm M}^{corr}$ the correction term to the usual Maxwell Lagrangian in (15), and we have defined

$$\begin{aligned} f(A) &\equiv \frac{3\kappa^2\hbar^2}{8} + \frac{\lambda\kappa^4\hbar^2}{4}A^2, \\ \alpha^{\sigma\varepsilon}(A) &\equiv -\lambda\kappa^4\hbar^2 A^\sigma A^\varepsilon, \\ \beta^\lambda(A,\tilde{F}) &\equiv -\frac{\lambda\kappa^2\hbar}{2}\epsilon^{\lambda\alpha\beta\gamma}A_{[\alpha}\tilde{F}_{\beta\gamma]} \ . \end{aligned} \tag{18}$$

Due to the presence of the cubic terms, this Dirac equation changes under the actions of C (charge conjugation) transformations, resulting in a different dynamics for anti-fermions. This dynamics breaks the U(1) symmetry explicitly and might be relevant in the very early Universe. These non-linear dynamics in the matter fields can be simplified for cases where certain approximations are valid as in cosmological applications for homogeneous fields and random spin distributions.

2.4.2. *Cosmological applications*

The application of this theory, the Einstein-Cartan-Dirac model coupled to ($s = 1$) bosons with explicit $U(1)$ symmetry breaking, to homogeneous and isotropic cosmologies leads to the dynamics of the scale factor given by the generalized Einstein equations (assuming the FLRW metric and isotropic pressure)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\kappa^2}{3}\left(\rho + \rho^{\rm corr}\right) - \frac{k}{a^2}, \tag{19}$$

$$\frac{\ddot{a}}{a} = -\frac{\kappa^2}{6}\left[3(p + p^{\rm corr}) + (\rho + \rho^{\rm corr})\right], \tag{20}$$

where dots over functions represent time derivatives. Effective densities can be defined $\rho^{\rm eff} = \rho + \rho^{\rm corr}$, with the corrections to GR, $\rho^{\rm corr} = \rho^s + \rho^{s-A} + \rho^A$, corresponding to the spin-spin interaction energy, the non-minimal interactions between fermionic spin and the bosonic four-potential, and self-interactions in the bosonic sector, respectively and the same apply to the pressure contributions. Different families of solutions can be derived for the dynamics of the scale factor including cyclic models with early bounce, early acceleration, deceleration and late-bounce, and also cases with early bounce, early acceleration, deceleration and late-time acceleration. Some of these are illustrated in the figures 1 and 2.

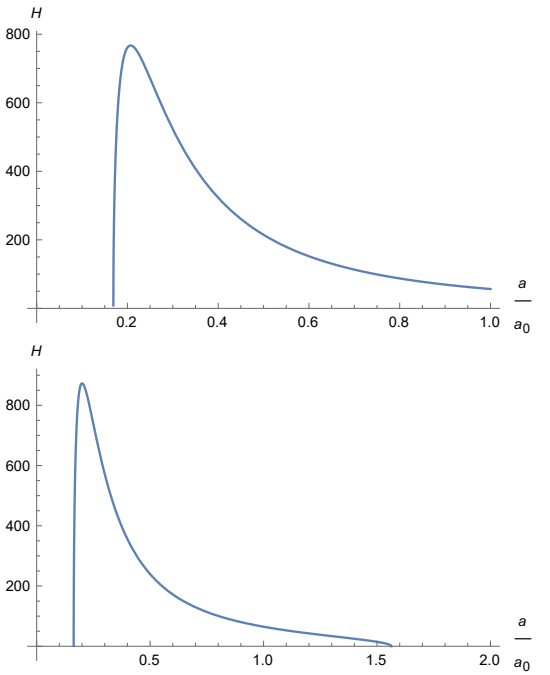


Fig. 1. Evolution of the Hubble parameter $H(a)$ with the scale factor a/a_0 for the ECDM model without (ECSK model, top) and with (bottom) the non-minimal couplings in the matter fields induced by torsion. These corrections to the effective energy density $\rho^{s-A} \sim \rho^s f(A) \sim a^b$ give raise to late-time effects, whereas $\rho^s \sim -\kappa^2 \breve{s}^2$ is the spin-spin interaction term that is responsible for the non-singular behaviour in the early Universe. The plot on the bottom shows a typical solution with a future bounce, a non-singular behaviour at the minimum of the scale factor, and a period of early accelerated expansion.

2.4.3. *Gravitational waves in Riemann-Cartan spacetime*

By considering the quadratic models in PGTG, where the Lagrangian is quadratic in the curvature and torsion, one obtains two sets of field equations for the metric and the contorsion, (or equivalently for the tetrads and the spin connection), where both fields are dynamically coupled and propagate in vacuum. Gravitational waves with torsion modes coupled to the metric modes are then expected and the possibility of measuring these effects with GW astronomy is a relevant open question.

In cosmological backgrounds, one possible parametrization for GW, encompassing a wide class of metric theories of gravity, is provided in the following wave equation

$$h''_{ij} + (2+\nu)\mathcal{H}h'_{ij} + (c_g^2 k^2 + m_g^2 a^2)h_{ij} = \tilde{\Pi}_{ij}. \tag{21}$$

In this equation the parameters (ν, c_g, m_g) and the source term represent deviations from GR and the derivatives are taken with respect to conformal time. The parameter ν represents an additional "friction" term, affecting the decay induced by the

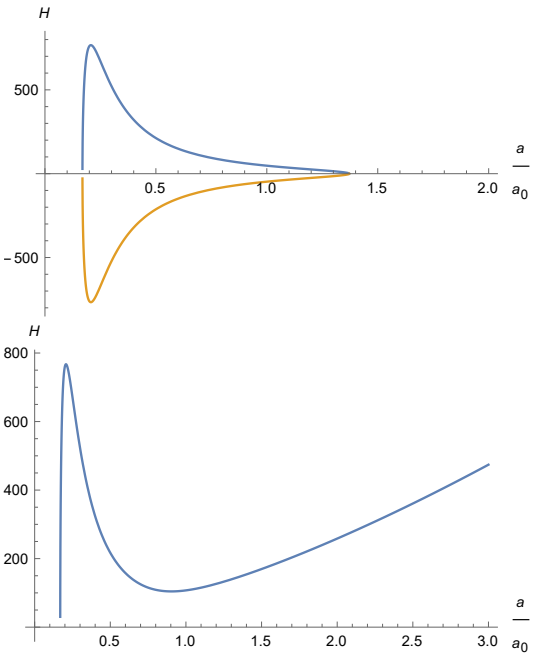


Fig. 2. On the top figure we can see the cyclic behaviour of ECDM model explicitly, with the two branches in $H(a) = \pm\sqrt{\rho^{\text{eff}}(a) - k/a^2}$ smoothly joined together at the bounces. A period of early accelerated expansion is followed by decelerated expansion, bounce and accelerated contraction, decelerated contraction and again the bounce at the minimum of the scale factor, with the repetition of the cosmological cycle. On the bottom we have the relevant case where a late-time accelerated phase is also present.

cosmological expansion. The effect of a propagation speed different than the speed of light is represented by c_g, while m_g stands for an effective mass for the graviton ($s = 2$) field. In GR $\nu = 0$, $c_g = 1$ and $m_g = 0$ and the right-hand side of (21) is equal to zero. The tensor $\tilde{\Pi}_{ij}$ represents an extra "source" term due to additional propagating tensor fields.

In the PGTG, the situation is generalized to include in the dynamics another set of equations. Through the variational principle, the two sets of field equations corresponding to the dynamics of the two gravitational potentials $(\theta^a_\mu, \Gamma^a_{\ b\nu})$, can also be expressed in terms of the equivalent set of dynamical variables $(g_{\alpha,\beta}, K^\alpha_{\ \mu\nu})$. Since in quadratic PGTG models the torsion in general propagates, the contorsion is an extra propagating tensor field with respect to GR and therefore, the wave equation for the metric in the first-order perturbation theory in cosmological backgrounds will have a term as $\tilde{\Pi}_{ij}$ in (21) due to the propagating torsion (or contorsion) tensor degrees of freedom.

One simple approach for the study of GW with torsion modes is to consider the linear perturbation theory in RC spacetime. The background spacetime in general is assumed to have the following metric and connection $(\mathring{g}_{\mu\nu}, \mathring{\Gamma}^\alpha_{\lambda\beta})$. The perturbed

quantities are

$$g_{\mu\nu} = \mathring{g}_{\mu\nu} + h_{\mu\nu}, \quad \Gamma^{\alpha}_{\ \lambda\beta} = \mathring{\Gamma}^{\alpha}_{\ \lambda\beta} + \gamma^{\alpha}_{\ \lambda\beta}, \tag{22}$$

where the perturbations $h_{\mu\nu} \equiv \delta g_{\mu\nu}$ and $\gamma^{\alpha}_{\ \lambda\beta} \equiv \delta\Gamma^{\alpha}_{\ \lambda\beta}$, are assumed to be small. The RC connection is given by $\Gamma^{\alpha}_{\ \beta\nu} = \tilde{\Gamma}^{\alpha}_{\ \beta\nu} + K^{\alpha}_{\ \beta\nu}$, expression also valid at the background level. It can be shown that the perturbation in the connection includes a perturbation in contorsion and another part related to the background metric and metric perturbation. In summary, the relevant quantities in this perturbative approach can be chosen to be

$$g_{\mu\nu} = \mathring{g}_{\mu\nu} + h_{\mu\nu}, \quad K^{\alpha}_{\ \lambda\beta} = \mathring{K}^{\alpha}_{\ \lambda\beta} + \kappa^{\alpha}_{\ \lambda\beta}. \tag{23}$$

From this the generalized curvature tensor, Ricci tensor and Ricci scalar of the linearly perturbed RC spacetime geometry are given by

$$\begin{aligned} R^{\alpha}_{\ \beta\mu\nu} = \tilde{R}^{\alpha}_{\ \beta\mu\nu} + 2\tilde{\nabla}_{[\mu}\mathring{K}^{\alpha}_{\ \beta|\nu]} + 2\tilde{\nabla}_{[\mu}\kappa^{\alpha}_{\ \beta|\nu]} + \mathring{K}^{\alpha}_{\ \lambda\mu}\mathring{K}^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\mu} \\ + \mathring{K}^{\alpha}_{\ \lambda\mu}\kappa^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\kappa^{\lambda}_{\ \beta\mu} + \kappa^{\alpha}_{\ \lambda\mu}\mathring{K}^{\lambda}_{\ \beta\nu} - \kappa^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\mu}, \end{aligned} \tag{24}$$

$$\begin{aligned} R_{\beta\nu} = \tilde{R}_{\beta\nu} + 2\tilde{\nabla}_{[\alpha}\mathring{K}^{\alpha}_{\ \beta|\nu]} + 2\tilde{\nabla}_{[\alpha}\kappa^{\alpha}_{\ \beta|\nu]} + \mathring{K}^{\alpha}_{\ \lambda\alpha}\mathring{K}^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\alpha} \\ + \mathring{K}^{\alpha}_{\ \lambda\alpha}\kappa^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\kappa^{\lambda}_{\ \beta\alpha} + \kappa^{\alpha}_{\ \lambda\alpha}\mathring{K}^{\lambda}_{\ \beta\nu} - \kappa^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\alpha}, \end{aligned} \tag{25}$$

and

$$\begin{aligned} R = \tilde{R} - 2\tilde{\nabla}^{\lambda}\mathring{K}^{\alpha}_{\ \lambda\alpha} - 2\tilde{\nabla}^{\lambda}\kappa^{\alpha}_{\ \lambda\alpha} + g^{\beta\nu}(\mathring{K}^{\alpha}_{\ \lambda\alpha}\mathring{K}^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\alpha}) \\ + g^{\beta\nu}(\mathring{K}^{\alpha}_{\ \lambda\alpha}\kappa^{\lambda}_{\ \beta\nu} - \mathring{K}^{\alpha}_{\ \lambda\nu}\kappa^{\lambda}_{\ \beta\alpha}) + g^{\beta\nu}(\kappa^{\alpha}_{\ \lambda\alpha}\mathring{K}^{\lambda}_{\ \beta\nu} - \kappa^{\alpha}_{\ \lambda\nu}\mathring{K}^{\lambda}_{\ \beta\alpha}), \end{aligned} \tag{26}$$

respectively.

For a torsionless background and at first order in the metric and connection perturbations, we get

$$R^{\alpha}_{\ \beta\mu\nu} = \tilde{R}^{\alpha}_{\ \beta\mu\nu} + 2\tilde{\mathring{\nabla}}_{[\mu}\kappa^{\alpha}_{\ \beta|\nu]} \tag{27}$$

$$R_{\beta\nu} = \tilde{R}_{\beta\nu} + 2\tilde{\mathring{\nabla}}_{[\alpha}\kappa^{\alpha}_{\ \beta|\nu]}, \tag{28}$$

and

$$R = \tilde{R} - 2\tilde{\mathring{\nabla}}^{\lambda}\kappa^{\alpha}_{\ \lambda\alpha}, \tag{29}$$

respectively. These quantities are useful to derive the GW equations for PGTG in the linear perturbation framework. As an example and without going into details, by choosing the quadratic model given by[26]

$$\mathcal{L} = \frac{1}{2\kappa^2}\left(R - 2\alpha T^{\alpha}_{\ \mu\nu}T_{\alpha}^{\ \mu\nu}\right) - \frac{1}{8\lambda}R^{\alpha\beta}_{\ \ \mu\nu}R_{\alpha\beta}^{\ \ \mu\nu} + \mathcal{L}_m, \tag{30}$$

where α and λ are free parameters, one can derive the following coupled wave equations

$$\frac{l_0^2}{2}\left(\Box\partial_{(\alpha}\bar{h}_{\beta)}{}^{\mu}+2\partial_\nu\partial^{[\mu}\kappa_{\alpha\beta}{}^{\nu]}\right)+3\eta^{\mu\nu}\kappa_{[\alpha\beta\nu]}+\delta^{\mu}_{[\alpha}\kappa_{\beta]}=0, \tag{31}$$

$$-\Box\bar{h}_{ij}=2\Lambda^{TT}_{ij}, \tag{32}$$

with $\kappa_\beta \equiv \kappa_{\beta\alpha}{}^{\alpha}$, $\Box \equiv \partial^\alpha\partial_\alpha$, we choose $\alpha = 1$ and define $l_0 \equiv \kappa^2/\lambda$, $\bar{h}_{ij}$ is the metric perturbation in the TT Lorentz gauge, $\eta^{\mu\nu}$ is the Minkowski metric, δ^μ_α is the Kronecker symbol and

$$\Lambda_{\mu\nu}\equiv\partial_{(\nu}\kappa_{\mu)}-\eta_{\mu\nu}\partial_\lambda\kappa^\lambda. \tag{33}$$

In principle one can apply Greens functions methods to get

$$\bar{h}_{ij}(t,\vec{x})\sim\int d^3y\frac{\Lambda^{TT}_{ij}(t_R,\vec{y})}{|\vec{x}-\vec{y}|}. \tag{34}$$

This reinforces the notion that metric and torsion perturbations can be interconvertible. Moreover, the equation (32) can be used to replace in (31) to obtain a second order partial diferential equation for the torsion modes

$$\frac{l_0^2}{2}\left(-2\partial_{(\alpha}\Lambda_{\beta)}{}^{\mu}+2\partial_\nu\partial^{[\mu}\kappa_{\alpha\beta}{}^{\nu]}\right)+3\eta^{\mu\nu}\kappa_{[\alpha\beta\nu]}+\delta^{\mu}_{[\alpha}\kappa_{\beta]}=0, \tag{35}$$

with all quantities calculated in the TT gauge. Now, taking into account the RC curvature tensor in (24) for a RC background and in (27) for a torsionless background (to first order), respectively, one expects generalizations to the geodesic deviation equations. This implies a generalization to the GR effects of the propagating GW in test-masses. For example, in Minkowski backgrounds, useful to compute the weak, linearised GW effects in detectors, we have

$$R_{\alpha\beta\mu\nu}=\tilde{R}_{\alpha\beta\mu\nu}+2\partial_{[\mu}\kappa_{\alpha\beta|\nu]}, \tag{36}$$

and using the TT gauge, the non-zero components of the Riemann part of the RC curvature are given by

$$\tilde{R}_{i0j0}=-\frac{1}{2}\ddot{h}_{ij}. \tag{37}$$

If we take the corresponig components in (36),

$$R_{i\beta j\nu}=-\frac{1}{2}\ddot{h}_{ij}\delta^0_\beta\delta^0_\nu+2\partial_{[j}\kappa_{i\beta|\nu]}, \tag{38}$$

where δ^0_β are defined as Kronecker symbols, and we see explicit extra terms due to the presence of the propagating contorsion perturbation.

The FRWL torsionless background geometry can be taken to illustrate the effects of torsion GW modes propagating over cosmological distances. For simplicity the spatially flat FRWL metric background can be taken. In a more general approach the perturbed versions of the field equations would be computed over a RC

backgroud geometry with a FRWL metric and a background contorion obeying the cosmological principle. Notice also that in curved backgrounds, terms such as $\tilde{\mathring{R}}^2$, $\tilde{\mathring{R}}\kappa$ and $\tilde{\mathring{R}}\tilde{\mathring{\nabla}}\kappa$ would also apear in the field equations, therefore, a far more richer phenomenology is expected than the one derived from a flat, Minkowski, torsionless background. In principle, it should be possible to obtain a set of two wave equations in cosmological backgrounds, similar to and generalizing the expression in (21). From such equations, the general beyond-GR effects can be recognized in the amplitudes, dispersion relation and due to the presence of extra degrees of freedom (for example contorsion d.o.f.). Then these direct consequences can be applied for the general expressions for the GW forms (in the weak source limit) due to binaries, in order to encompass the effects of a propagation over cosmological distances. We will not perform such procedure here, but emphasize the relevance of it in order to search for signatures of propagating torsion modes in the GW signals.

3. Probing post-Riemann geometries

It is known that torsion can give rise to precession effects in systems with intrinsic spin, for example, elementary particles such as the electron, or baryons such as the neutron.[34–36] Moreover, as we will see, torsion effects on the dynamics and energy levels of spinors are expected. Also, the spin-spin contact interactions of the ECSK theory, or the propagating spin-spin interactions mediated by gravitational gauge ($s = 0, 1, 2$) bosons of quadratic PGTG might be tested/constrained in laboratory experiments and cosmological GW probes.

Regarding non-metricity, the gauge approach to gravity clearly shows that the hypermomentum currents, such as the dilatations or the shear currents, couple to the trace and shear parts of the connection (10), respectively and therefore to non-metricity. If torsion can be measured by the spin precession of test matter with intrinsic spin, then the non-metricity of spacetime can be measured by pulsations (mass quadrupole excitations) of test matter with (non-trivial) hypermomentum currents. In order to be "sensitive" to non-Riemann geometries, test matter should carry dilatation, shear, or spin currents, whether macroscopic or at the level of fundamental fields/particles. From the gauge approach to gravity one conclusion is very clear: if matter has neither (intrinsic) spin, nor dilatation/shear currents, then it follows the Riemannian (extremal length) geodesics, regardless of the geometry of spacetime or of the form of the Lagrangian in metric-affine geometry.

3.1. *Effects on fundamental fermionic systems*

From the Dirac Lagrangian of a fermion minimally coupled to the background RC geometry, one can predict torsion effects on the energy levels of quantum systems.[37] Similarly, if non-minimal couplings between the torsion trace vector and Dirac axial vector and/or between torsion axial vector and Dirac vector are present, then the parity symmetry is broken and the corresponding energy level corrections due to

torsion will contain the signatures of those parity breaking interactions. Therefore, tests with advanced spectrographs might be able to probe torsion effects on quantum systems. Without going into details, we present here some simple results concerning predicted effects on the energy levels of fermionic spinors coupled to background torsion quantities, for cases of minimal and non-minimal couplings (including parity breaking effects).[37]

For the minimal coupling case, which can be expressed by the Lagrangian

$$\mathcal{L}_{\rm Dirac} = \tilde{\mathcal{L}}_{\rm Dirac} + 3\breve{T}^{\lambda}\breve{s}_{\lambda}, \tag{39}$$

where

$$\breve{s}^{\lambda} \equiv \frac{\hbar}{2}\bar{\psi}\gamma^{\lambda}\gamma^{5}\psi \tag{40}$$

is the Dirac axial spin vector current and

$$\breve{T}^{\lambda} \equiv \frac{1}{6}\epsilon^{\lambda\alpha\beta\gamma}T_{\alpha\beta\gamma} \tag{41}$$

is the axial vector part of torsion, we obtain the Dirac equation,

$$i\hbar\gamma^{\mu}\tilde{D}_{\mu}\psi - m\psi = -\frac{3\hbar}{2}\breve{T}^{\lambda}\gamma_{\lambda}\gamma^{5}\psi\,, \tag{42}$$

for spinors. From this, energy level corrections due to torsion can be derived, depending on the alignment or anti-alignment between the spin and the spacetime torsion. For a constant background torsion $\breve{T}^3$, the two corresponding energy levels are

$$E^2 = p^2 + \left(m \mp \frac{3\hbar}{2}\breve{T}^3\right)^2\,, \tag{43}$$

for the spin down/up states, respectively.

Let us denote by $m_{\breve{T}}$ the mass correction due to the spin-torsion interaction, and consider the two possible energy levels E_1 and E_2, with $E_2 > E_1$. Then in the reference frame of the particle, the corresponding energy transition is given by

$$h\nu = \frac{1}{2}m_{\breve{T}} = \frac{3\hbar}{4}\breve{T}\,. \tag{44}$$

Reinserting the speed of light in vacuum, we get

$$\nu = \frac{3c}{8\pi}\breve{T}. \tag{45}$$

If we consider, for instance, $\breve{T} \sim 10^{-16}m^{-1}$, then we end up with the prediction of a transition in the $\nu \sim$nHz regime. In the case of a spherically symmetric torsion, using perturbation theory to first order, where

$$E \simeq E_{(0)} + < \psi_{(0)}|\hat{U}_{torsion-spin}|\psi_{(0)} >, \tag{46}$$

and $\psi_{(0)}$ are the eigen states of the unperturbed Hamiltonian associated to the eigen value $E_{(0)}$, we get a similar result with

$$\nu = \frac{\delta E}{h} = \frac{3E_{(0)}}{\pi}F(r)\,, \tag{47}$$

where $F(r)$ represents a geometrical term coming from the integration of the background spherical torsion over some volume.

To include parity breaking terms in the model we also consider the couplings $T \cdot a$ and $\breve{T} \cdot j$ in the Lagrangian

$$\mathcal{L}_{\text{fermions}} = \tilde{\mathcal{L}}_{\text{Dirac}} + (\alpha_1 T + \beta_2 \breve{T}) \cdot j + (\alpha_2 \breve{T} + \beta_1 T) \cdot a. \tag{48}$$

The terms $\beta_1 T^\lambda a_\lambda$ and $\beta_2 \breve{T}^\mu j_\mu$ break the parity invariance. Here j and a represent the Dirac vector and axial vectors, respectively, and T is the trace torsion vector. The corresponding extended Dirac equation is given by

$$i\hbar\gamma^\mu \tilde{D}_\mu \psi - m\psi = -\left(\alpha_1 T^\lambda + \beta_2 \breve{T}^\lambda\right) \gamma_\lambda \psi - \left(\alpha_2 \breve{T}^\lambda + \beta_1 T^\lambda\right) \gamma_\lambda \gamma^5 \psi \,. \tag{49}$$

In this case, and considering constant torsion quantities, and using the definitions

$$t^\nu \equiv \alpha_1 T^\nu + \beta_2 \breve{T}^\nu, \qquad \tau^\lambda \equiv \alpha_2 \breve{T}^\lambda + \beta_1 T^\lambda \,, \tag{50}$$

and taking $t^\mu = (0, t^1, t^2, t^3)$, and $\tau^\mu = (0, 0, 0, \tau)$, for simplicity, we obtain the following energy level corrections $E_2^2 = p_{\text{eff}}^2 + (m + \tau)^2$ and $E_1^2 = p_{\text{eff}}^2 + (m - \tau)^2$, corresponding to the energy transition

$$h\nu = E_2 - E_1 = \frac{4m\tau}{E_1 + E_2} \,, \tag{51}$$

which in the reference frame of the particle reads

$$h\nu = \frac{4m\tau}{\left[\vec{t}^2 + (m + \tau)^2\right]^{1/2} + \left[\vec{t}^2 + (m - \tau)^2\right]^{1/2}} \,, \tag{52}$$

and $p_{\text{eff}}^2 \equiv (p_1 + t^1)^2 + (p_2 + t^2)^2 + (p_3 + t^3)^2$. The parity symmetry breaking is one of the conditions usually considered in order to explain some early Universe particle physics that can provide a matter/anti-matter asymmetry. The torsion of the RC spacetime can induce parity breaking effects in the fermionic sector via the non-minimal couplings explored here. These effects include the prediction of well-defined frequencies that a free fermion can absorb or emit in order to make transitions between the predicted two energy levels that arise depending on the spin orientation with respect to external torsion quantities. The signature of parity breaking (chirality) might also be present in the radiated field itself, in the form of polarized light, and these type of parity-breaking spin-torsion interactions might be probed in the very early Universe with GW astronomy, where the signatures of chirality in the GW signals in principle require non-planar detectors and analysis with 3-point correlation functions.

4. Open questions and concluding remarks

In special and general relativity theories the metric is fundamental, providing the relevant spacetime geometry. In special relativity the metric provides a fixed background, while in GR it is described by a tensor field theory being one of the

fundamental dynamical fields together with the matter fields. In its essence GR is a background independent theory, where the metric (the local geometry) has to be derived by solving the Einstein field equations, for each particular physical/ astrophysical scenario. It is the fundamental geometrical field, solution to the Einstein gravity equations, from which all the aspects of pseudo-Riemannian geometry can be derived. It is therefore natural that due to epistemological historical reasons the metric has gained a fundamental importance in gravity and in the description of physical fields propagating in some spacetime geometry.

In spite of this, developments of geometrical methods in field theories led to the notion that the spacetime metric is no longer to be considered as a fundamental field. One can say that both the pre-metric foundations of electrodynamics and the gauge approach to gravity point in the same direction. The pre-metric approach to electrodynamics as expressed in differential forms give completely general, coordinate-free covariant inhomogeneous and homogeneous equations from charge conservation and magnetic flux conservation respectively. No metric is involved and therefore, electrodynamics is not linked to Minkowski spacetime at a fundamental level. Moreover, the spacetime metric up to a conformal factor, and therefore, the conformal-causal spacetime structure, can be derived from local and linear electrodynamics.[38] The resulting lightcone or causal-electromagnetic structure is a conformal geometry with local conformal symmetries associated to the lightcone at each spacetime point. The group of conformal transformations preserve the light-cone (causal) structure but the metric can change, with the set of all conformally related metrics associated to the local light-cone representing different equally valid metrics. There isn't a specific metric which is more fundamental or that can be considered to be more "physical" than the others. Moreover, in such geometry, under the assumption of locality, the parallel transport of a light cone from a given point to a neighbouring point gives rise to a deformation of the local causal/light-cone structure according to the non-metricity tensor. And since, as we saw, the tensor-valued non-metricity 1-form is linked to the existence of a non-Lorentzian linear connection, from a gauge point of view, this alone inevitably leads to Lorentz symmetry breaking. Therefore, on one hand, and in spite of the historical reasons relating Maxwell's theory with special relativity and Minkowski spacetime, electrodynamics is fundamentally connected to the conformal geometrical structure and the conformal group, and not to Minkowski spacetime, nor the Poincaré or Lorentz group. On the other hand, the notion of local symmetries, the basis of gauge theories, as applied to spacetime leads to gravitational theories with post-Riemann geometries, and non-metricity is directly related to non-Lorentzian connections, i.e, to Lorentz symmetry breaking. *If both the gauge theories of gravity and the premetric formulation of electrodynamics (with its spacetime constitutive relations) are jointly analysed, then we are led to consider that the conformal structure is more fundamental than the metric, post-Riemann geometries with non-metricity and to conformal gauge theories of gravity. Symmetry breaking mechanisms are plausible scenarios for bringing both gravity and electrodynamics into the phenomenological regime of Poincaré symmetries and related spacetime*

paradigm. It should be mentioned that the analogies between Yang-Mills fields and gauge theories of gravity in the exterior calculus of forms, clearly show that gravitational dynamics can also be described in a pre-metric way, just as electrodynamics, and its the gravitational spacetime constitutive relations that bring the conformal causal structure.

The exploration of the gauge theories of gravity, including its connection to spacetime symmetries and post-Riemann geometries, can lead to a vast number of predictions relevant for relativistic astrophysics, cosmology and GWs and also to interesting avenues towards unified gauge field theories. Moreover the connection between the gauge formalism, the Lagrangian and Hamiltonian formalisms, the pre-metric foundations, the constitutive relations (between the field strengths and its conjugate momenta, the excitations) and the causal structure of spacetime, might also contribute towards relevant non-perturbative, background-independent methods for quantum field theory and quantum gravity. The gauge approach can clarify which gravitational geometrical degrees of freedom should be taken as canonical conjugate variables for quantization. In its turn, this can relax the traditional emphasis that is usually put into the metric degrees of freedom, for historical reasons, within attempts at reconciling GR (or more general metric theories) with quantum theory.

The gauge theories of gravity are built upon a self-consistent approach that clarifies the fundamental connection between spacetime symmetries and spacetime geometries. In this context, the metric-affine formalism is clearly formulated within general spacetime geometries with curvature, torsion and non-metricity. In this work, the geometrical methods and symmetry principles in gravitation were explored motivating a new perspective into the spacetime paradigm. The effects of post-Riemann spacetime geometries with torsion were briefly exemplified in applications to fundamental fermionic and bosonic fields, cosmology and gravitational waves. Besides the brief exposition on some phenomenological considerations within PGTG, the fundamental ideas related to spacetime physics, motivated by geometrical methods and symmetry principles, have been discussed in the context of the possible routes towards a new spacetime paradigm in gravitation and unified field theories.

Acknowledgments

FC acknowledges the support by Fundação para a Ciência e a Tecnologia (FCT, Portugal) through the grant No. PD/BD/128017/2016 and the project PTDC/FIS-OUT/29048/2017. FSNL acknowledges support from the FCT Scientific Employment Stimulus contract with reference CEECIND/04057/2017. DRG is funded by the *Atracción de Talento Investigador* programme of the Comunidad de Madrid (Spain) No. 2018-T1/TIC-10431, and acknowledges further support from the Ministerio de Ciencia, Innovación y Universidades (Spain) project No. PID2019-108485GB-I00/AEI/10.13039/501100011033. The authors also acknowledge funding

from the FCT projects No.UID/FIS/04434/2019 and also the organizers of the MG16 conference for the opportunity to share this research.

References

1. C. M. Will, "The Confrontation between General Relativity and Experiment," Living Rev. Rel. **17** (2014) 4 [arXiv:1403.7377 [gr-qc]].
2. A. Hees, T. Do, A. M. Ghez, G. D. Martinez, S. Naoz, E. E. Becklin, A. Boehle, S. Chappell, D. Chu and A. Dehghanfar, *et al.* "Testing General Relativity with stellar orbits around the supermassive black hole in our Galactic center," Phys. Rev. Lett. **118** (2017) 211101 [arXiv:1705.07902 [astro-ph.GA]].
3. B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], "Observation of Gravitational Waves from a Binary Black Hole Merger," Phys. Rev. Lett. **116** (2016), 061102 [arXiv:1602.03837 [gr-qc]].
4. B. P. Abbott *et al.* [LIGO Scientific and Virgo], "GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence," Phys. Rev. Lett. **119** (2017) no. 14, 141101 [arXiv:1709.09660 [gr-qc]].
5. B. P. Abbott *et al.* "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," Phys. Rev. Lett. **119**, 161101 (2017).
6. B. P. Abbott *et al.* [LIGO Scientific and Virgo], "Tests of general relativity with GW150914," Phys. Rev. Lett. **116** (2016) no. 22, 221101 [erratum: Phys. Rev. Lett. **121** (2018) no. 12, 129902] [arXiv:1602.03841 [gr-qc]].
7. K. Akiyama *et al.* [Event Horizon Telescope], "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," Astrophys. J. **875** (2019) L1 doi:10.3847/2041-8213/ab0ec7 [arXiv:1906.11238 [astro-ph.GA]].
8. P. Bull, Y. Akrami, J. Adamek, T. Baker, E. Bellini, J. Beltran Jimenez, E. Bentivegna, S. Camera, S. Clesse and J. H. Davis, *et al.* "Beyond ΛCDM: Problems, solutions, and the road ahead", Phys. Dark Univ. **12** (2016) 56.
9. H. Weyl, "A New Extension of Relativity Theory," Annalen Phys. **59** (1919) 101.
10. H. Weyl, "Gravitation and the electron", Proc. Nat. Acad. Sci. **15** (1929) 323.
11. C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," Phys. Rev. **96** (1954) 191.
12. R. Utiyama, "Invariant theoretical interpretation of interaction," Phys. Rev. **101** (1956) 1597.
13. T. W. B. Kibble, "Lorentz invariance and the gravitational field," J. Math. Phys. **2** (1961) 212.
14. D. W. Sciama, "The Physical structure of general relativity," Rev. Mod. Phys. **36** (1964), 463. [Erratum: Rev. Mod. Phys. **36** (1964) 1103].
15. C. G. Boehmer and J. Burnett, "Dark spinors with torsion in cosmology," Phys. Rev. D **78** (2008) 104001 [arXiv:0809.0469 [gr-qc]].
16. "Nonsingular, big-bounce cosmology from spinor-torsion coupling", Phys. Rev. D **85** (2012) 107502 [arXiv:1111.4595 [gr-qc]].
17. N. J. Poplawski, "Big bounce from spin and torsion", Gen. Rel. Grav. **44** (2012) 1007 [arXiv:1105.6127 [astro-ph.CO]].
18. B. Vakili and S. Jalalzadeh, "Signature transition in Einstein-Cartan cosmology," Phys. Lett. B **726** (2013), 28-32 [arXiv:1308.2877 [gr-qc]].
19. K. A. Bronnikov and A. M. Galiakhmetov, "Wormholes and black universes without phantom fields in Einstein-Cartan theory," Phys. Rev. D **94** (2016) no. 12, 124006 [arXiv:1607.07791 [gr-qc]].

20. A. N. Ivanov and M. Wellenzohn, "Einstein–cartan Gravity with Torsion Field Serving as an Origin for the Cosmological Constant or Dark Energy Density," Astrophys. J. **829** (2016) no. 1, 47 [arXiv:1607.01128 [gr-qc]].
21. J. A. R. Cembranos, J. Gigante Valcarcel and F. J. Maldonado Torralba, "Singularities and n-dimensional black holes in torsion theories," JCAP **1704**, 021 (2017) [arXiv:1609.07814 [gr-qc]].
22. J. A. R. Cembranos and J. Gigante Valcarcel, "Extended Reissner–Nordström solutions sourced by dynamical torsion," Phys. Lett. B **779** (2018), 143-150 [arXiv:1708.00374 [gr-qc]].
23. G. Unger and N. Popławski, "Big bounce and closed universe from spin and torsion," Astrophys. J. **870** (2019) no. 2, 78 [arXiv:1808.08327 [gr-qc]].
24. M. Blagojevic and F. W. Hehl, "Gauge Theories of Gravitation,", Imperial College Press, London, April, 2013. arXiv:1210.3775 [gr-qc].
25. M. Blagojević, "Gravitation and Gauge symmetries" Institute of Physics Publishing, 2002.
26. V. N. Ponomarev, A. O. Barvinsky, Y. N. Obukhov, "Gauge approach and quantization methods in gravity theory" Nauka, 2017.
27. F. W. Hehl, J. D. McCrea, E. W. Mielke and Y. Ne'eman, "Metric affine gauge theory of gravity: Field equations, Noether identities, world spinors, and breaking of dilation invariance," Phys. Rept. **258** (1995) 1 [arXiv:gr-qc/9402012 [gr-qc]].
28. F. Cabral, F. S. N. Lobo and D. Rubiera-Garcia, "Fundamental Symmetries and Spacetime Geometries in Gauge Theories of Gravity: Prospects for Unified Field Theories," Universe **6** (2020), 238 [arXiv:2012.06356 [gr-qc]].
29. E. A. Lord and P. Goswami, "Gauging the conformal group," Pramana **25** (1985) 635.
30. N. Berkovits, "SuperPoincaré invariant superstring field theory," Nucl. Phys. B **450** (1995) 90 [Erratum: Nucl. Phys. B **459** (1996) 439] [arXiv:hep-th/9503099].
31. F. Cabral, F. S. N. Lobo and D. Rubiera-Garcia, "Einstein-Cartan-Dirac gravity with U(1) symmetry breaking," Eur. Phys. J. C **79**, 1023 (2019).
32. F. Cabral, F. S. N. Lobo and D. Rubiera-Garcia, "Cosmological bounces, cyclic universes, and effective cosmological constant in Einstein-Cartan-Dirac-Maxwell theory," Phys. Rev. D **102** (2020) no. 8, 083509 [arXiv:2003.07463 [gr-qc]].
33. F. Cabral, F. S. N. Lobo and D. Rubiera-Garcia, "The cosmological principle in theories with torsion: The case of Einstein-Cartan-Dirac-Maxwell gravity," JCAP **10** (2020), 057 [arXiv:2004.13693 [gr-qc]].
34. H. Rumpf, "Quasiclassical limit of the Dirac equation and the equivalence principle in the Riemann-Cartan geometry," NATO Sci. Ser. B **58** (1980) 93.
35. J. Audretsch, "Dirac Electron in Space-times With Torsion: Spinor Propagation, Spin Precession, and Nongeodesic Orbits," Phys. Rev. D **24** (1981) 1470.
36. J. Nitsch and F. W. Hehl, "Translational Gauge Theory of Gravity: Postnewtonian Approximation and Spin Precession," Phys. Lett. B **90** (1980) 98.
37. F. Cabral, F. S. N. Lobo, D. Rubiera-Garcia, "Imprints from a Riemann-Cartan space-time on the energy levels of Dirac spinors", Class. Quant. Grav. **38** (2021) 19 [arXiv:2102.02048 [gr-qc]].
38. F. W. Hehl and Y. N. Obukhov, "Spacetime metric from local and linear electrodynamics: A New axiomatic scheme," Lect. Notes Phys. **702** (2006) 163 [arXiv:gr-qc/0508024].

Thermodynamics of scalar-tensor gravity: A new approach

Valerio Faraoni
Dept. of Physics & Astronomy, Bishop's University
Sherbrooke, Québec, Canada J1M 1Z7
E-mail: vfaraoni@ubishops.ca

We propose a new approach to the thermodynamics of scalar-tensor gravity and its possible "diffusion" toward general relativity, previously regarded as an equilibrium state in spacetime thermodynamics. The main idea is describing scalar-tensor gravity as an effective dissipative fluid and applying Eckart's first order thermodynamics to it. This gives explicit effective quantities: heat current density, "temperature of gravity", viscosity coefficients, entropy density, plus an equation describing the "diffusion" to Einstein gravity. These quantities, otherwise missing in spacetime thermodynamics, pop out with minimal assumptions.

Keywords: Scalar-tensor gravity, temperature of gravity theories, limit to general relativity.

1. Introduction

The thermodynamics of spacetime proposal became the subject of great interest when Jacobson derived the Einstein equation as an equilibrium equation of state from the thermodynamical properties of vacuum[1] and, later, the field equation of (metric) $f(R)$ gravity was obtained in a non-equilibrium thermodynamical setting.[2] These works have generated a large literature and suggest that general relativity (GR) could be a sort of lowest energy state of gravity, while alternative theories of gravity correspond to excited states. Indeed, modifying gravity with respect to GR introduces extra degrees of freedom in addition to the usual massless spin two field of GR. Therefore, theories containing GR as a limit correspond to excitations of these extra degrees of freedom, which are quiescent or absent in GR. Ultimately, if modified gravity is an excited state of GR, one should quantify the distance of an excited state from the GR "ground state" and identify some order parameter and equations describing the relaxation of gravity to GR. This order parameter, a sort of "temperature of gravity theories", is a crucial piece missing from this picture. Here we restrict ourselves to scalar-tensor[3,4] and $f(R)$ gravity (the latter is a subclass of the former[5–7]) and we identify a possible "temperature of gravity" candidate. The identification should be made in some relativistic thermodynamical context. Here we start from a very simple fact: the field equations of scalar-tensor gravity can be rewritten as effective Einstein equations by grouping the terms generated by the extra scalar field in the right hand side and noticing that they assume the form of the stress-energy tensor of an *imperfect* effective fluid. Next, we ask ourselveves what we know about dissipative fluids. The simplest description of dissipation and

heat flux in GR is given by Eckart's first order thermodynamics.[8] This theory is notoriously plagued by non-causality and instabilities but is still widely used as an approximation. By describing scalar-tensor gravity as an effective imperfect fluid and using the constitutive equations of Eckart's theory (which are minimal assumptions and could be imposed in different models as well), we identify a "temperature of gravity" in the space of scalar-tensor theories,[3,4] which contain GR as the zero temperature limit.[9]

Let us recall the basics of scalar-tensor gravity, the prototypical alternative to Einstein's GR. The Jordan frame action is[a3]

$$S_{\rm ST} = \frac{1}{16\pi}\int d^4x\sqrt{-g}\left[\phi R - \frac{\omega(\phi)}{\phi}\,\nabla^c\phi\nabla_c\phi - V(\phi)\right] + S^{(\rm m)}\,, \tag{1}$$

where $\phi > 0$ is the Brans-Dicke scalar (approximately equivalent to the inverse of the effective gravitational coupling strength), $\omega(\phi)$ is the "Brans-Dicke coupling", $V(\phi)$ is a scalar field potential, and $S^{(\rm m)}$ is the matter action. Varying (1) produces the field equations

$$R_{ab} - \frac{1}{2}\,g_{ab}R = \frac{8\pi}{\phi}\,T^{(\rm m)}_{ab} + \frac{\omega}{\phi^2}\left(\nabla_a\phi\nabla_b\phi - \frac{1}{2}\,g_{ab}\nabla_c\phi\nabla^c\phi\right) + \frac{1}{\phi}\left(\nabla_a\nabla_b\phi - g_{ab}\Box\phi\right)$$
$$-\frac{V}{2\phi}\,g_{ab}\,, \tag{2}$$

$$\Box\phi = \frac{1}{2\omega+3}\quad\left(\frac{8\pi g^{ab}T^{(\rm m)}_{ab}}{\phi} + \phi\,\frac{dV}{d\phi} - 2V - \frac{d\omega}{d\phi}\nabla^c\phi\nabla_c\phi\right)\,, \tag{3}$$

where $T^{(\rm m)}_{ab}$ is the matter stress-energy tensor.

2. The temperature of scalar-tensor gravity

An effective fluid description of the field equations (2) has been proposed in[11] for non-minimally coupled scalar fields and in[12] for general scalar-tensor gravity, and completed in Ref. 13. The fluid 4-velocity is

$$u^a = \frac{\nabla^a\phi}{\sqrt{-\nabla^e\phi\nabla_e\phi}} \tag{4}$$

when the gradient $\nabla^c\phi$ is timelike. The $3+1$ splitting of spacetime into the 3-space "seen" by the comoving observers of the fluid and their time direction u^a follows, with $h_{ab} \equiv g_{ab} + u_a u_b$ and $h_{ab}u^a = h_{ab}u^b = 0$, $h^a{}_b\,h^b{}_c = h^a{}_c$. The fluid 4-acceleration is

$$\dot{u}_a = (-\nabla^e\phi\nabla_e\phi)^{-2}\,\nabla^b\phi\Big[(-\nabla^e\phi\nabla_e\phi)\nabla_a\nabla_b\phi + \nabla^c\phi\nabla_b\nabla_c\phi\nabla_a\phi\Big]\,, \tag{5}$$

[a]We follow the notation of Ref. 10, using units in which Newton's constant and the speed of light are unity.

and the effective stress-energy tensor of the Brans-Dicke-like field

$$8\pi T^{(\phi)}_{ab} = \frac{\omega}{\phi^2}\left(\nabla_a\phi\nabla_b\phi - \frac{1}{2}\,g_{ab}\nabla^c\phi\nabla_c\phi\right) + \frac{1}{\phi}\left(\nabla_a\nabla_b\phi - g_{ab}\Box\phi\right) - \frac{V}{2\phi}\,g_{ab} \tag{6}$$

can be written as the imperfect fluid stress-energy tensor

$$T_{ab} = \rho u_a u_b + q_a u_b + q_b u_a + \Pi_{ab}\,, \tag{7}$$

where[12,13]

$$\rho = T_{ab}u^a u^b\,, \qquad q_a = -T_{cd}\,u^c h_a{}^d\,, \qquad \Pi_{ab} \equiv P h_{ab} + \pi_{ab} = T_{cd}\,h_a{}^c\,h_b{}^d\,, \tag{8}$$

$$P = \frac{1}{3}\,g^{ab}\Pi_{ab} = \frac{1}{3}\,h^{ab}T_{ab}\,, \qquad \pi_{ab} = \Pi_{ab} - P h_{ab}\,, \tag{9}$$

are, respectively, the effective energy density, heat flux density, stress tensor, isotropic pressure, and anisotropic stresses in the comoving frame, while $q_c u^c = \Pi_{ab}u^b = \pi_{ab}u^b = \Pi_{ab}u^a = \pi_{ab}u^a = \pi^a{}_a = 0$. These comoving frame fluid quantities read[12,13]

$$8\pi\rho^{(\phi)} = -\frac{\omega}{2\phi^2}\,\nabla^e\phi\nabla_e\phi + \frac{V}{2\phi} + \frac{1}{\phi}\left(\Box\phi - \frac{\nabla^a\phi\nabla^b\phi\nabla_a\nabla_b\phi}{\nabla^e\phi\nabla_e\phi}\right), \tag{10}$$

$$8\pi q^{(\phi)}_a = \frac{\nabla^c\phi\nabla^d\phi}{\phi\left(-\nabla^e\phi\nabla_e\phi\right)^{3/2}}\left(\nabla_d\phi\nabla_c\nabla_a\phi - \nabla_a\phi\nabla_c\nabla_d\phi\right), \tag{11}$$

$$8\pi\Pi^{(\phi)}_{ab} = \left(-\frac{\omega}{2\phi^2}\,\nabla^c\phi\nabla_c\phi - \frac{\Box\phi}{\phi} - \frac{V}{2\phi}\right)h_{ab} + \frac{1}{\phi}\,h_a{}^c h_b{}^d\nabla_c\nabla_d\phi\,, \tag{12}$$

$$8\pi P^{(\phi)} = -\frac{\omega}{2\phi^2}\,\nabla^e\phi\nabla_e\phi - \frac{V}{2\phi} - \frac{1}{3\phi}\left(2\Box\phi + \frac{\nabla^a\phi\nabla^b\phi\nabla_b\nabla_a\phi}{\nabla^e\phi\nabla_e\phi}\right), \tag{13}$$

$$\begin{aligned} 8\pi\pi^{(\phi)}_{ab} = {} & \frac{1}{\phi\nabla^e\phi\nabla_e\phi}\left[\frac{1}{3}\left(\nabla_a\phi\nabla_b\phi - g_{ab}\nabla^c\phi\nabla_c\phi\right)\left(\Box\phi - \frac{\nabla^c\phi\nabla^d\phi\nabla_d\nabla_c\phi}{\nabla^e\phi\nabla_e\phi}\right)\right. \\ & \left. + \nabla^d\phi\left(\nabla_d\phi\nabla_a\nabla_b\phi - \nabla_b\phi\nabla_a\nabla_d\phi - \nabla_a\phi\nabla_d\nabla_b\phi + \frac{\nabla_a\phi\nabla_b\phi\nabla^c\phi\nabla_c\nabla_d\phi}{\nabla^e\phi\nabla_e\phi}\right)\right]. \end{aligned} \tag{14}$$

The comparison of Eqs. (11) and (5) yields

$$q^{(\phi)}_a = -\frac{\sqrt{-\nabla^c\phi\nabla_c\phi}}{8\pi\phi}\,\dot{u}_a\,, \tag{15}$$

which brings us to our main point. The relation (15) has a physical consequence in Eckart's first order thermodynamics,[8] in which the heat flux density is related to the temperature $\mathcal{T}$ by the generalized Fourier law (a constitutive relation)[8]

$$q_a = -K\left(h_{ab}\nabla^b\mathcal{T} + \mathcal{T}\dot{u}_a\right), \tag{16}$$

where K is the thermal conductivity. Comparing Eqs. (15) and (16) leads to the result that, in the comoving frame, the spatial temperature gradient vanishes and the heat flow is caused solely by the inertia of energy (the acceleration term in (16)). The positive-definite temperature of the effective fluid is[9, 13]

$$\mathcal{T} = \frac{\sqrt{-\nabla^c\phi\nabla_c\phi}}{8\pi K\phi}\,. \tag{17}$$

According to Eq. (17) GR, which is the limit $\phi = $ const. of scalar-tensor gravity,[3, 4] corresponds to zero temperature $\mathcal{T}$, while in scalar-tensor spacetimes the scalar degree of freedom ϕ is excited and $\mathcal{T} > 0$.[9] The "temperature of gravity" is higher when the gravitational coupling $G_{\rm eff} \sim \phi^{-1}$ is larger. In the space of scalar-tensor theories, gravity (viewed as a dissipative system) left to itself can "diffuse" toward GR, as described in.[9] Let us confront this idea with examples from gravitational physics.

3. Examples

3.1. *f(R) gravity*

Metric $f(R)$ gravity (see[5–7] for reviews) is the most popular alternative to Einstein theory to explain the current acceleration of the universe without invoking an *ad hoc* dark energy. This theory is equivalent to an $\omega = 0$ Brans-Dicke theory with dynamical[b] scalar field $\phi = f'(R)$ and potential $V(\phi) = \phi R - f(R)\Big|_{R=R(\phi)}$,[5–7] hence

$$\mathcal{T} = \frac{f''(R)\sqrt{-\nabla^c R\nabla_c R}}{8\pi K f'(R)}\,. \tag{18}$$

In order for the graviton to carry positive energy, it must be $f'(R) > 0$, while $f''(R) > 0$ is required for stability,[14] which guarantees that $K\mathcal{T} \geq 0$. This inequality is saturated by GR which, with $f(R) = R - 2\Lambda$, corresponds to $\mathcal{T} = 0$. It is significant that the requirement $\mathcal{T} > 0$ corresponds to local stability.

3.2. *Friedmann-Lemaître-Robertson-Walker universes*

In Friedmann-Lemaître-Robertson-Walker (FLRW) universes, spatial isotropy implies that the spatial heat flux of the effective fluid vanishes. For the same reason, the spatial temperature gradient $h_{ab}\nabla^b\mathcal{T} = 0$, while $\dot{a}_c \neq 0$ (in general, the ϕ-fluid does not obey the dust equation of state). Then, Eq. (16) implies that FLRW universes have "gravity temperature" $\mathcal{T} = 0$ in the comoving frame, whether they are solutions of scalar-tensor gravity or GR. Equivalently, in any FLRW space the imperfect effective fluid reduces to a perfect one for which heat flux and anisotropic

[b] In the Palatini version of $f(R)$ gravity, instead, ϕ is not dynamical and this theory is not an excitation of GR (cf.[23]).

stresses vanish identically (this property is valid also for Lovelock and $f(R,\mathcal{G})$ theories, where $\mathcal{G} \equiv R^2 - 4R_{ab}R^{ab} + R_{abcd}R^{abcd}$ is the Gauss-Bonnet term, in arbitrary dimension[15]). This conclusion goes against the idea (see, *e.g.*, Ref.[16]) that an attractor mechanism to GR operates in scalar-tensor cosmology. The situation of FLRW and Bianchi cosmologies with regard to the temperature of gravity is under examination and will be reported elsewhere.

3.3. *Vacuum, asymptotically flat, stationary black holes*

It is well known since an early theorem by Hawking that vacuum, asymptotically flat black holes in Brans-Dicke gravity are the same as in GR,[17] a result later generalized to scalar-tensor theories[18,19] provided that $V(\phi)$ has a minimum for the scalar ϕ to sit in stable equilibrium (the exceptions are maverick solutions with ϕ diverging on the horizon, *e.g.*,[20]). The key ingredient of Hawking's theorem consists of proving that ϕ is constant outside the horizon, in which case the theory reduces to GR in that region, corresponding to zero "theory temperature" $\mathcal{T}$. The interpretation is that these extreme objects "freeze" the extra dynamical degrees of freedom outside their horizons. Then, the no-hair theorems state that the GR black holes are the states of "lowest temperature" in the space of scalar-tensor black holes.

3.4. *Stealth scalar fields*

Clearly, Minkowski spacetime should correspond to zero temperature and, indeed, it is a constant ϕ solution of the field equations (2) and (3). However, Minkowski spaces with non-gravitating stealth scalar fields are also solutions of scalar-tensor gravity (*e.g.*,[21,22]). For example, the Ayon-Beato stealth solution $g_{ab} = \eta_{ab}$, $\phi = a_1 x^\mu x_\mu + p_\mu x^\mu + a_2$ (where $a_{1,2}$ and p_μ are constants)[21] correspond to "theory temperature"

$$\mathcal{T} = \frac{\sqrt{4a_1\left(\phi - a_2\right) + p_\alpha p^\alpha}}{8\pi K \phi}\,. \tag{19}$$

This interpretation makes sense where $\phi > 0$ and $a_1(\phi - a_2) + p_\mu p^\mu/4 > 0$. Even though the stress-energy tensor of the effective imperfect fluid disappears, the scalar degree of freedom is excited and the temperature of gravity is non-zero.

In another stealth solution due to Robinson,[22] $V \equiv 0$, $\omega = \text{const.}$, the geometry is Minkowskian, and the scalar is the wave

$$\phi = \begin{cases} \phi_0\, \mathrm{e}^{c_1 u + c_2} & \text{if} \quad \omega = -1\,, \\ \phi_0\left(c_1 u + c_1\right)^{\frac{1}{\omega+1}} & \text{if} \quad \omega \neq -1\,, \end{cases} \tag{20}$$

where u is the retarded time and $c_{1,2}$ are constants. In this case, a massless wave-like excitation of the scalar degree of freedom with $\nabla_c\phi\nabla^c\phi = 0$ gives $\mathcal{T} = 0$. A ϕ-wave as a dissipative system was regarded as a problem in[23] because classical gravitational theories are time-reversal invariant, but this state cannot decay, therefore this is not a truly dissipative system.

4. Conclusions

We have introduced a candidate to the role of "temperature of gravity" $\mathcal{T}$ using the effective imperfect fluid description of scalar-tensor gravity and Eckart's thermodynamics.[9,13] In this picture, the approach to GR is a diffusion-like process. Altough our proposal is intriguing at first sight, a deeper analysis is needed before it can be taken seriously. The specific examples discussed (plus those in[9]) corroborate the proposed Eq. (17), but our proposal remains preliminary. In any case, ultimately it will suffer from the same limitations of Eckart's theory. Nevertheless, the explicit expression of the "temperature of gravity" in the wide class of scalar-tensor theories (the prototypical alternative to GR, containing the $f(R)$ subclass extremely popular in cosmology) provides a working scheme to investigate the relaxation to GR.

Alternative approaches are possible: one can trade temperature with chemical potential, assigning to the effective ϕ-fluid zero temperature and entropy but nonzero chemical potential as done, *e.g.*, in.[24] These approaches and a more detailed analysis of our proposal for $\mathcal{T}$ are under study and will be reported elsewhere.

Acknowledgments

I am grateful to Andrea Giusti and Jeremy Côté, who contributed to parts of the research presented here. This work is supported by the Natural Sciences & Engineering Research Council of Canada (Grant No. 2016-03803) and by Bishop's University.

References

1. T. Jacobson, Phys. Rev. Lett. **75**, 1260 (1995).
2. C. Eling, R. Guedens, and T. Jacobson, Phys. Rev. Lett. **96**, 121301 (2006).
3. C. Brans and R.H. Dicke, Phys. Rev. **124**, 925 (1961).
4. P.G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968); R.V. Wagoner, Phys. Rev. D **1**, 3209 (1970); K. Nordvedt, Astrophys. J. **161**, 1059 (1970).
5. T.P. Sotiriou and V. Faraoni, Rev. Mod. Phys. **82**, 451 (2010).
6. A. De Felice and S. Tsujikawa, Living Rev. Relativ. **13**, 3 (2010).
7. S. Nojiri and S.D. Odintsov, Phys. Rept. **505**, 59 (2011).
8. C. Eckart, Phys. Rev. **58**, 919 (1940).
9. V. Faraoni and A. Giusti, Phys. Rev. D **103**, no.12, L121501 (2021).
10. R.M. Wald, *General Relativity* (Chicago University Press, Chicago, 1987).
11. M.S. Madsen, Class. Quantum Grav. **5**, 627 (1988).
12. L.O. Pimentel, Class. Quantum Grav. **6**, L263 (1989).
13. V. Faraoni and J. Coté, Phys. Rev. D **98**, 084019 (2018).
14. V. Faraoni, Phys. Rev. D **74**, 104017 (2006).
15. M. Gurses and Y. Heydarzade, Eur. Phys. J. C **80**, 1061 (2020).
16. T. Damour and K. Nordtvedt, Phys. Rev. Lett. **70**, 2217 (1993).
17. S.W. Hawking, Commun. Math. Phys. **25**, 167 (1972).
18. J.D. Bekenstein, 1996, in *Second International A.D. Sahkarov Conference on Physics*, Moscow, 20-23 May 1996, edited by I.M. Dremin and A.M. Semikhatov (World Scientific, Singapore), p. 21 (arXiv:9605059 [gr-qc]).

19. T.P. Sotiriou and V. Faraoni, Phys. Rev. Lett. **108**, 081103 (2012).
20. N.M. Bocharova, K.A. Bronnikov, and V.N. Melnikov, Vestn. Mosk. Univ. Fiz. Astron. **6**, 706 (1970); J.D. Bekenstein, Ann. Phys. (N.Y.) **82**, 535 (1974).
21. E. Ayon-Beato, C. Martinez, R. Troncoso, and J. Zanelli, Phys. Rev. D **71**, 104037 (2005).
22. D.C. Robinson, Gen. Relativ. Gravit. **38**, 153 (2006).
23. G. Chirco, C. Eling, and S. Liberati, Phys. Rev. D **83**, 024032 (2011).
24. O. Pujolas, I. Sawicki, and A. Vikman, J. High Energy Phys. **1111**, 156 (2011); R. Akhoury, C. Gauthier, and A. Vikman, J. High Energy Phys. **0903**, 082 (2009).

Two body dynamics in a quadratic modification of general relativity

Soham Bhattacharyya
Max Planck Institute for Gravitational Physics (Albert Einstein Institute)
Leibniz Universität Hannover

It is shown in this study that deviations from the Einstein-Hilbert action at the quadratic level using a proper analyses and suitable dynamical variables lead to a tiny modification to the post Newtonian equations of motion, and non-GR like behavior at very short length scales.

Keywords: Modified gravity; two-body dynamics; post-Newtonism formalism.

1. Introduction

General theory of Relativity (GR) is a field theory, one of the first of its kind, that is classified under classical physics. The validity of the theory ranges from solar system length scales to cosmological length scales. It has passed most of its predictions with flying colors, starting from the extra perihelion precession of Mercury in the inner solar system,[1] to gravitational lensing,[2] to the generation of gravitational waves (GW),[3–8] and finally to the existence of objects called black holes (BH).[9]

One of the first calculations that tested the predictive power of GR was the perihelion precession of Mercury, which is to say that GR gave predictions of all the possible orbital dynamics that was possible to be observed in the early 20th century, and its predictive power remains the same to this date. GR predicted that the *gravitational force* will not always be inverse squared of the distance from the source, and that at shorter length scales, the behavior changes. To be precise, a perturbative treatment of GR, also known as the post Newtonian (PN) formalism,[10–16] reveals that a series of shorter (than inverse squared) ranged velocity dependent forces can have a prominent effect on the dynamics of two objects trapped in their mutual gravitational field (like the Sun and Mercury system). Hence, the universality of Newtonian gravity stops being universal when one goes beyond the *Newtonian bubble*, broadly speaking.

However, an uncomfortable question pops up when one considers the PN forces of GR, and keeps on continuing the PN series to shorter and shorter length scales. Or when one considers the collapse of a self-gravitating mass of fluid at the extreme scenario. At every point along the collapse, the gravitational force, including the PN forces, are attractive in nature. Hence, if the fermionic pressure, or the force that makes matter solid, is unable to balance the increasingly powerful (but shorter ranged) forces of gravity, a singularity may occur. This is because in the tug of war of all the forces present in one or multiple gravitating systems, gravity seems to win

in some extreme conditions. Such conditions can occur in ultra dense self gravitating systems like the end stages of a star collapsing under its own gravity. If there is no other force that can stop the unyielding pull of gravity, matter eventually collapses fully in on itself. But long before such a full collapse, or singularity, can happen, a special surface forms around the eventual singularity. Such a surface is known as the event horizon, or the surface the forms the termination of all *events* in the Universe that lies outside of the surface. Here *event* refers to the special relativistic notation of space-time events, which are, in the sense of mathematical modelling, points in a four dimensional object known as the manifold, which in turn refers to the complete past, present, and future of the Universe that is being modelled.

However, the universality of attraction of gravity may not be the case when one goes beyond the Einstein-Hilbert (EH) action, referred to as GR in literature. In order to demonstrate that, a quadratic deviation to the EH action is considered, which is a subclass of post-Einsteinian theories known as $f(R)$ theories of gravity.[17,18] Such theories are characterized by an extra massive scalar degree of freedom,[19–21] in addition to the two massless tensor degrees of freedom.

A PN treatment of the densitized field equations of a quadratic extension to GR will be performed in this study that resembles the Landau-Lifshitz formulation of GR. It will be shown that the equations of motion have a rather good limit to GR (almost indistinguishable in fact), and only deviates from the predictions of GR at very high energy scales (where energy scales refer to really large velocities and very high spatial curvatures combined).

Geometrized units, $c = G = 1$ will be used in this study unless explicitly mentioned otherwise. The notations of[22] will be followed except a few changes in variable and index labeling. The mostly minus sign convention will be followed for the metric, that is $(+,-,-,-)$.

2. Layout

2.1. *The densitized formulation of GR*

The gothic metric/Landau-Lifshitz/densitized formulation of GR may be motivated from a particular observation regarding the structure of the second-most condensed form of the Einstein field equations in its covariant form (or the canonical form), that is

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \propto T_{\mu\nu} \tag{1}$$

where the first tensor quantity on the LHS of the above is the Ricci tensor, and the scalar multiplied to the metric tensor in the second term of the LHS is the contraction of the Ricci tensor with the metric, known as the Ricci scalar. The RHS is the classical matter energy-momentum-stress density tensor, which contains certain information (although non-exhaustive) about the content that is separate from the space-time. One notices that the net content of the RHS of (1) is essentially a trace reversed form of the Ricci tensor. Trace reversal implies that the trace of

the new tensor that was introduced in the beginning of the 20th century has a trace that is negative of the trace of the Ricci tensor, and is known as the Einstein tensor, which leads to the most condensed form of the field equations of GR, or

$$G_{\mu\nu} \propto T_{\mu\nu} \tag{2}$$

where all components of the tensor $G_{\mu\nu}$ has a one-to-one relationship with all of the corresponding components of the matter energy-momentum-stress density tensor. However, the cost of condensation is the loss of information and usability. As it is, Eq. (2) does not tell one about how planets and stars move and how GWs are radiated. To be able to predict such phenomenon mathematically, one needs to 'dirty their hands'.

In classical GR, $T_{\mu\nu}$ can contain fluids (serving as a model for stars, dust, etc.), or it can contain the energy-momentum-stress density equivalent of electromagnetic fields and charges. The Ricci tensor/scalar are obtained from geometric principles governing the nature of surfaces, volumes, and higher dimensional generalization; which are in turn obtained by contracting the tensor quantity that serves as the model for the space-time 'fabric'[a], or the Riemann tensor $R_{\mu\sigma\rho\nu}$. The Riemann tensor in GR is a function of the metric and its derivatives (till second). It has certain symmetries that are given as follows

$$R_{\mu\sigma(\nu\rho)} = R_{(\mu\sigma)\nu\rho} = 0 \tag{3}$$

$$R_{\mu[\sigma\nu\rho]} = 0 \tag{4}$$

$$R_{\mu\rho\nu\sigma} = R_{\nu\sigma\mu\rho} \tag{5}$$

$$R_{\mu\rho[\nu\sigma;\alpha]} = 0. \tag{6}$$

The first, third of the identities, that is Eqs. (3) and (5), are called skew and interchange symmetries respectively. Whereas Eqs. (4) and (6) are known as the first and second Bianchi identities respectively. The second Bianchi identity is particularly useful since it leads to the following identity on the Einstein tensor $G_{\mu\nu}$

$$\nabla^{\mu} G_{\mu\nu} = 0, \tag{7}$$

where $\nabla^{\mu} \equiv g^{\mu\rho} \nabla_{\rho}$ is the index-raised covariant derivative associated with the metric $g_{\alpha\beta}$. The above, combined with the proportionality condition (2), lead to the following

$$\nabla^{\mu} T_{\mu\nu} = 0 \tag{8}$$

which is a covariant conservation law on the classical matter (or electromagnetic fields) energy-momentum-stress density tensor, and using which equation of motion of classical matter itself can be deduced. Hence the famous saying by J.A. Wheeler

[a] Although the term fabric is used, it is put under quotation to differentiate from an actual material object, which is something that the physics community thought of as 'Aether' before the famous Michelson-Morley experiment. In conventional notion, material or fabric can be torn or broken; however, a similar 'tearing' of space-time is something that does not seem to happen. Hence one must take the concept of 'fabric' with a pinch of salt.

about Eq. (2): *Spacetime tells matter how move; matter tells spacetime how to curve*, is incomplete without throwing in Eq. (8) into the fold.

With the above motivations one can define a new tensor given by[23–25]

$$\mathfrak{g}^{\mu\nu} = \sqrt{-g}\, g^{\mu\nu} \tag{9}$$

which utilizes the contravariant metric, or the 'dual' of the covariant metric, whose components actually define the space-time structure. Now the point of taking the dual is to simplify the equations that need to be solved, by replacing the problem at hand with another problem which relates to the original problem via transformation and/or reparametrizations. One also notices the presence of the determinant of the metric tensor through $\sqrt{-g}$ multiplying $g^{\mu\nu}$. Under the transformation (9), Eq. (1) or (2) can be framed as a slightly less condensed (but more useful) form given as follows

$$\frac{1}{2\,(-g)}\,\partial_{\alpha\,\beta} H^{\mu\alpha\nu\beta} \;-\; 8\,\pi\, t^{\mu\nu}_{LL} \;=\; 8\,\pi\, T^{\mu\nu}_{m} \tag{10}$$

$$H^{\mu\alpha\nu\beta} = \mathfrak{g}^{\alpha\beta}\mathfrak{g}^{\mu\nu} - \mathfrak{g}^{\alpha\nu}\mathfrak{g}^{\beta\mu} \tag{11}$$

$$\begin{aligned} 16\,\pi\,(-g)\,t^{\alpha\beta}_{\mathrm{LL}} = {} & \partial_\alpha \mathfrak{g}^{\mu\nu}\partial_\beta \mathfrak{g}^{\alpha\beta} - \tfrac{1}{4}\mathfrak{g}^{\alpha\beta}\mathfrak{g}_{\gamma\zeta}\mathfrak{g}_{\theta\iota}\mathfrak{g}^{\mu\nu}\partial_\alpha \mathfrak{g}^{\gamma\theta}\partial_\beta \mathfrak{g}^{\zeta\iota} \\ & + \tfrac{1}{2}\mathfrak{g}_{\gamma\zeta}\mathfrak{g}_{\theta\iota}\mathfrak{g}^{\mu\alpha}\mathfrak{g}^{\nu\beta}\partial_\alpha \mathfrak{g}^{\gamma\theta}\partial_\beta \mathfrak{g}^{\zeta\iota} + \tfrac{1}{8}\mathfrak{g}^{\alpha\beta}\mathfrak{g}_{\gamma\zeta}\mathfrak{g}_{\theta\iota}\mathfrak{g}^{\mu\nu}\partial_\alpha \mathfrak{g}^{\gamma\zeta}\partial_\beta \mathfrak{g}^{\theta\iota} \\ & - \tfrac{1}{4}\mathfrak{g}_{\gamma\zeta}\mathfrak{g}_{\theta\iota}\mathfrak{g}^{\mu\alpha}\mathfrak{g}^{\nu\beta}\partial_\alpha \mathfrak{g}^{\gamma\zeta}\partial_\beta \mathfrak{g}^{\theta\iota} - \partial_\alpha \mathfrak{g}^{\mu\alpha}\partial_\beta \mathfrak{g}^{\nu\beta} + \mathfrak{g}^{\alpha\beta}\mathfrak{g}_{\gamma\zeta}\partial_\alpha \mathfrak{g}^{\mu\gamma}\partial_\beta \mathfrak{g}^{\nu\zeta} \\ & + \tfrac{1}{2}\mathfrak{g}_{\alpha\beta}\mathfrak{g}^{\mu\nu}\partial_\gamma \mathfrak{g}^{\beta\zeta}\partial_\zeta \mathfrak{g}^{\alpha\gamma} - \mathfrak{g}_{\beta\gamma}\mathfrak{g}^{\nu\alpha}\partial_\alpha \mathfrak{g}^{\gamma\zeta}\partial_\zeta \mathfrak{g}^{\mu\beta} - \mathfrak{g}_{\beta\gamma}\mathfrak{g}^{\mu\alpha}\partial_\alpha \mathfrak{g}^{\gamma\zeta}\partial_\zeta \mathfrak{g}^{\nu\beta}. \end{aligned} \tag{12}$$

The definition of the Landau-Lifshitz pseudo-tensor $t^{\mu\nu}_{LL}$ deviates slightly from the usual definition found in literature since all the metrics that are used to raise or lower the indices have been transformed into the gothic metric. The covariant gothic metric is defined as

$$\mathfrak{g}_{\mu\nu} = \frac{g_{\mu\nu}}{\sqrt{-g}}. \tag{13}$$

One notices that the rank 4 tensor $H^{\mu\alpha\nu\beta}$ has the same symmetries as the Riemann tensor, and is proportional to the Riemann tensor corresponding to the maximally symmetric Riemann tensor. The maximally symmetric Riemann tensor is a solution of the field equations of GR when the space-time is homogeneous and isotropic everywhere. Deviation from such a space-time with the maximum number of symmetries is marked by the appearance of a coordinate or a gauge dependent tensor (or a pseudo-tensor) $t^{\mu\nu}_{LL}$, also known as the Landau-Lifshitz energy-momentum-stress density pseudo-tensor. $t^{\mu\nu}_{LL}$ is a quantification of the energy, momentum, and stress that is inherent in the space-time, and leads to a solutions that are less symmetric than the completely homogeneous and isotropic space-time one started off with. Such solutions are perturbative in nature and formed the early basis for the Post-Newtonian formalism in GR.[10–16] The perturbation is defined about the Minkowski space-time as

$$\mathfrak{h}^{\mu\nu} = \eta^{\mu\nu} \;-\; \mathfrak{g}^{\mu\nu}. \tag{14}$$

The above tensor is projected on a Cartesian coordinate frame of a freely falling observer by choosing the following gauge condition

$$\partial_\mu \mathfrak{h}^{\mu\nu} = 0 \tag{15}$$

which not only simplifies the field equations but also puts it in a form that offers a more intimate insight into the nature of space-time and its relationship to classical matter. They are given by the following

$$\Box \mathfrak{h}^{\mu\nu} = -16\,\pi\,\Lambda^{\mu\nu} \tag{16}$$

$$\Lambda^{\mu\nu} = \Theta^{\mu\nu} + \partial_{\alpha\,\beta}\chi^{\mu\nu\alpha\beta} \tag{17}$$

$$\Theta^{\mu\nu} = (-g)\,(T^{\mu\nu} + t^{\mu\nu}_{LL} - t^{\mu\nu}_{H}) \tag{18}$$

$$\chi^{\mu\nu\alpha\beta} = \frac{1}{16\,\pi}\left(\mathfrak{h}^{\alpha\nu}\,\mathfrak{h}^{\beta\mu} - \mathfrak{h}^{\alpha\beta}\,\mathfrak{h}^{\mu\nu}\right) \tag{19}$$

where a proportionality constant of $16\,\pi$ has now been used, which is the case when one uses geometrized units ($G = c = 1$). The rank 4 tensor $\chi^{\mu\alpha\nu\beta}$ with Riemann symmetry, being a total derivative term, can be turned into a boundary term using Gauss's law, and can be discarded as the source of the wave equation (16) when a particular choice of the Green's function is made for the flat-space Laplacian/wave operator $\Box$. One immediately notices that the dynamical variable now follows the well studied massless wave equation, which has standard solutions. It is also to be noted that if one chooses the effective energy-momentum-stress density tensor to be the sum of both the matter and the pseudo part, the dynamical variable has a one-to-one relationship with the matter density. The Landau-Lifshitz or the gothic metric formalism is hence often called as the densitized formalism, as it allows the introduction of a variable that can be taken as the space-time density itself, in contrast to the regular metric tensor which has a very complicated relationship with the energy-momentum-stress density tensor. The covariant conservation identity in Eq. (8) in this 'reduced' formalism, when added to the Landau-Lifshitz pseudo-tensor, along with the harmonic gauge condition (15), reduces to the following

$$\partial_\nu\,\Lambda^{\mu\nu} = 0. \tag{20}$$

Now to come back to the motivation for mentioning trace-reversal at the beginning of the section; one may choose to study perturbative GR by simply linearizing the metric tensor about the Minkowski space-time itself, that is

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} \tag{21}$$

At the linear order the gothic metric perturbation and the metric perturbation are related as follows (using Jacobi's identity)

$$\mathfrak{h}^{\mu\nu} = h^{\mu\nu} - \frac{1}{2}\,\eta^{\mu\nu}\,h \tag{22}$$

$$\mathfrak{h} = -h \tag{23}$$

where h and $\mathfrak{h}$ are the trace of the metric perturbation and the gothic metric perturbation respectively. Eq. (22) is used as a standard textbook solution for modeling the propagation of gravitational waves in empty space-times. Multiplying $\sqrt{-g}$ to the dual of the metric tensor is also similar to the multiplication of the same factor to the Ricci scalar (other than making the volume element covariant) to make the Einstein-Hilbert Lagrangian density, which has the same dimensions as any other matter Lagrangian density which may appear at the action level along with the Einstein-Hilbert Lagrangian density, that is

$$S = \int_{\mathscr{M}} \sqrt{-g}\, d^4x \left(\frac{R}{16\,\pi} + \mathscr{L}_{matter} + \mathscr{L}_{electromagnetism} + ... \right) \tag{24}$$

2.2. *The densitized formulation of $f(R)$ theories of gravity*

Like any other field theory, the class of theories known as $f(R)$ theories of gravity can be derived from the following action

$$S = \int_{\mathscr{M}} \sqrt{-g}\, d^4x \left(\frac{f(R)}{16\,\pi} + \mathscr{L}_{matter} + \mathscr{L}_{electromagnetism} + ... \right) \tag{25}$$

However, in the current study for simplicity, we will not be including electromagnetic fields or any other 'exotic' fields. The field equations can be derived by varying the action with respect to the metric, and its higher derivatives, and by enforcing all variations vanish at the boundaries of the integration, which is well suited to study the classical problem at hand. The field equations so obtained are as follows[21]

$$f'(R)\, R_{\mu\nu} - \frac{1}{2} f(R)\, g_{\mu\nu} - [\nabla_\mu \nabla_\nu - g_{\mu\nu} \Box]\, f'(R) = 8\,\pi\, T_{\mu\nu} \tag{26}$$

$$f'(R) = \frac{df(R)}{dR} \tag{27}$$

where ∇_μ, as usual, is the covariant derivative associated with the metric $g_{\mu\nu}$. The above field equations can be written in two different forms (if one intends to use the machinery of GR): one modifies the Newtonian constant (which has been put to unity in the current study), and the other preserves it. The former is written as[21]

$$G_{\mu\nu} = \frac{8\,\pi}{f'} \left(T_{\mu\nu} + T^{eff}_{\mu\nu} \right) \tag{28}$$

$$T^{eff}_{\mu\nu} \equiv \frac{1}{2} \left(f - R\, f' \right) g_{\mu\nu} + \left(\nabla_\mu \nabla_\nu - g_{\mu\nu} \Box \right) f' \tag{29}$$

where $\Box \equiv g^{\alpha\beta}\, \nabla_\alpha \nabla_\beta$. The latter version has the following form

$$G_{\mu\nu} = 8\,\pi\, T_{\mu\nu} + \tilde{T}^{eff}_{\mu\nu} \tag{30}$$

$$\tilde{T}^{eff}_{\mu\nu} = (1 - f')\, R_{\mu\nu} + \frac{1}{2} (R - f)\, g_{\mu\nu} + \left(\nabla_\mu \nabla_\nu - g_{\mu\nu} \Box \right) f'. \tag{31}$$

In this study, only the former, that is Eqs. (28)-(29) will be considered because of convenience, and the extra terms are less in number. There is also another differential equation that is satisfied by the scalar function f', which is independent of whatever form is chosen for the effective energy-momentum-stress density tensor, and is obtained by taking the trace of either Eqs. (28) or (30). It's given by

$$3\, g^{\mu\nu}\, \nabla_\mu\, \nabla_\nu\, f' \,+\, f'\, R \,-\, 2\, f = 8\, \pi\, T \tag{32}$$

and can be regarded as the dynamics of some effective scalar field. The transition of f' from an 'effective' to an actual scalar degree of freedom will be made clear in a few paragraphs.

Before trying to obtain an equivalent of Eq. (10) one must be made aware of a particular conformal property of $f\,(R)$ theories of gravity, that is under a general conformal transformation of the metric tensor

$$\tilde{g}_{\mu\nu} = \Omega^2\, g_{\mu\nu} \tag{33}$$

$f\,(R)$ theories can be written as an *Einstein + massive scalar* system. The $f\,(R)$ action without matter, for example transforms as[26–30]

$$S = \int_{\mathscr{M}} d^4x\, \sqrt{-g}\, \left(\frac{\tilde{R}}{16\, \pi} \,+\, \frac{1}{2}\, \tilde{g}^{\mu\nu}\, \phi_{,\mu}\, \phi_{,\nu} \,-\, V\,(\phi) \right) \tag{34}$$

$$\phi \equiv \sqrt{\frac{3}{16\, \pi}}\, \ln \Omega \tag{35}$$

$$V\,(\phi) \equiv \frac{1}{16\, \pi}\, \frac{f \,-\, R\, f'}{\left(f'\right)^2} \tag{36}$$

$$\Omega \equiv f' \tag{37}$$

Where $\tilde{R}$ is the Ricci scalar of the conformally transformed metric $\tilde{g}_{\mu\nu}$. This confirms the presence of an extra massive scalar degree of freedom in $f\,(R)$ theories of gravity. In the literature of $f\,(R)$ and scalar-tensor theories of gravity, the 'vanilla' metric $g_{\mu\nu}$ is known as the Jordan frame, whereas the conformally transformed metric $\tilde{g}_{\mu\nu}$ is called the Einstein frame. From this motivation, one can define a new gothic metric (in the Jordan frame) $\tilde{g}^{\mu\nu}$ given by

$$\tilde{\mathfrak{g}}^{\mu\nu} = f'\,(R)\, \sqrt{-g}\, g^{\mu\nu} \tag{38}$$

which frames the field equations (28) in the following form

$$\partial_{\alpha\,\beta} \tilde{H}^{\alpha\mu\beta\nu} = -16\, \pi\, (-g)\, f'\,(R)\, \left(T^{\mu\nu} \,+\, t^{\mu\nu}_{eff} \,+\, \tilde{t}^{\mu\nu}_{LL} \right), \tag{39}$$

$$\begin{aligned} 16\, \pi\, (-g)\, f'\, t^{\mu\nu}_{eff} = \sqrt{-g}\, \left(f \,-\, R\, f' \right)\, \tilde{\mathfrak{g}}^{\mu\nu} \,+\, \frac{3}{\left(f'\right)^2}\, \left(\tilde{\mathfrak{g}}^{\mu\alpha}\, \tilde{\mathfrak{g}}^{\nu\beta} \,-\, \frac{1}{2}\, \tilde{\mathfrak{g}}^{\mu\nu}\, \tilde{\mathfrak{g}}^{\alpha\beta} \right) \\ \times \quad \partial_\alpha f'\, \partial_\beta f' \end{aligned} \tag{40}$$

where $\tilde{H}^{\alpha\mu\beta\nu}$ and $\tilde{t}^{\mu\nu}_{LL}$ has the same form as that of $H^{\alpha\mu\beta\nu}$ and $t^{\mu\nu}_{LL}$, respectively, but with $\mathfrak{g}$ replaced by $\tilde{\mathfrak{g}}$. The newly appeared tensor $t^{\mu\nu}_{eff}$ contains products of only quadratic forms of first derivatives of f', as seen in Eq. (40). Owing to the conformal transformation, all the double derivatives of f' in the tensor field equations cancel away, making $t^{\mu\nu}_{eff}$ a perfectly usable source for the perturbative post Newtonian expansion. Curiously, the definition Eq. (38) appears as an identity of Palatini formalism of $f(R)$ theories of gravity as well, for example in Eq. (17) of.[31]

Now before a post Newtonian sequence of solutions can be found, one needs to choose a particular form of the function $f(R)$, which consequently fixes the form of the function $f'(R)$ as well, and can be taken for simplicity as

$$f(R) = R + \frac{f''(0)}{2} R^2 \tag{41}$$

$$f'(R) = 1 + f''(0)\, R \tag{42}$$

where the coefficients of expansion $f'(0)$ is taken to be unity to recover GR at the $R = 0$ limit, and $f''(0)$ will be taken to be negative for the course of this article, following.[32–35] One defines a tensor field $\tilde{\mathfrak{h}}^{\mu\nu}$, similar to Eq. (14), that propagates on the Minkowski background. One can define that as follows

$$\tilde{\mathfrak{h}}^{\mu\nu} = \eta^{\mu\nu} - \tilde{\mathfrak{g}}^{\mu\nu}. \tag{43}$$

In order to obtain wave like equations of motion where the coordinate variable functions comprising the coordinate 4-vectors follow the conditions

$$\Box\, x^{\mu} = 0, \tag{44}$$

one must choose the following gauge/coordinate condition similar to Eq. (15)

$$\partial_{\mu}\, \tilde{\mathfrak{h}}^{\mu\nu} = 0 \tag{45}$$

which will be called as the conformal harmonic gauge. Now similar to the conformally transformed new metric density $\tilde{\mathfrak{g}}^{\mu\nu}$, one must also define a scalar variable corresponding to a scalar density, say $\mathfrak{R}$, that follows a massive wave equation (to model the dynamics of the massive scalar degree of freedom), and maintains the $(-g)$ relationship with the trace of the net energy-momentum-stress density tensor $\tilde{\Lambda}^{\mu\nu}$ as well, as in Eq. (16). One may define that as follows

$$\mathfrak{R} = \sqrt{-g}\, R. \tag{46}$$

Substituting Eq. (43), (46), and (45) in the reduced field equations (39), one obtains the following

$$\Box\tilde{\mathfrak{h}}^{\mu\nu} = -16\,\pi\,\tilde{\Lambda}^{\mu\nu} \tag{47}$$

$$\tilde{\Lambda}^{\mu\nu} = (-g)\left[T^{\mu\nu}_{m}\left(1 - \frac{f''\,\mathfrak{R}}{\sqrt{-g}}\right) + t^{\mu\nu}_{LL} + t^{\mu\nu}_{H} + t^{\mu\nu}_{eff}\right]. \tag{48}$$

$t^{\mu\nu}_{eff}$ is comprised of various products of $\tilde{\mathfrak{h}}^{\mu\nu}$, $\mathfrak{R}$, and first derivatives of $\tilde{\mathfrak{h}}^{\mu\nu}$ and $\mathfrak{R}$; whose truncated form till the quadratic order of $f''(0)$ is given as follows

$$16\,\pi\,(-g)\,t^{\mu\nu}_{eff} = \frac{f''\,\mathfrak{R}^2}{\sqrt{-g}}\,\eta^{\mu\nu} + \frac{3\,(f'')^2}{4\,(-g)}\Big[2\,\big(\eta^{\mu\alpha}\,\eta^{\nu\beta} + \eta^{\mu\beta}\,\eta^{\nu\alpha} - \eta^{\mu\nu}\,\eta^{\alpha\beta}\big)\,\mathfrak{R}\,\partial_\alpha\mathfrak{R}\,\partial_\beta\tilde{\mathfrak{h}} + \Big(\eta^{\mu\alpha}\,\eta^{\nu\beta} - \frac{1}{2}\,\eta^{\mu\nu}\,\eta^{\alpha\beta}\Big)\Big(4\,\partial_\alpha\mathfrak{R}\,\partial_\beta\mathfrak{R} + \mathfrak{R}^2\,\partial_\alpha\tilde{\mathfrak{h}}\,\partial_\beta\tilde{\mathfrak{h}}\Big)\Big] + \mathscr{O}\left[\left(\frac{f''}{\sqrt{-g}}\right)^3\right] \tag{49}$$

$$\tilde{\mathfrak{h}} \equiv \eta_{\mu\nu}\,\tilde{\mathfrak{h}}^{\mu\nu} \tag{50}$$

with the following conservation law being satisfied by $\tilde{\Lambda}^{\mu\nu}$

$$\partial_\mu\,\tilde{\Lambda}^{\mu\nu} = 0\,, \tag{51}$$

whereas the new scalar dynamical variable $\mathfrak{R}$ satisfies the following, as obtained by substituting Eq. (43), (46), and (45) in Eq. (32)

$$\Box\mathfrak{R} + \sqrt{-g}\,\gamma^2\,\mathfrak{R} = -8\,\pi\,(-g)\,\gamma^2\,\tilde{\Lambda} \tag{52}$$

$$\gamma^2 \equiv -\frac{1}{3\,f''(0)} \tag{53}$$

$$\tilde{\Lambda} = T_m + f''(0)\left\{\frac{T_m\,\mathfrak{R}}{3\,\sqrt{-g}} + \frac{1}{8\,\pi\,(-g)}\left(\frac{\mathfrak{R}^2}{3} - \partial_\mu\mathfrak{R}\,\partial^\mu\tilde{\mathfrak{h}} - \frac{1}{2}\,\mathfrak{R}\,\partial_\mu\tilde{\mathfrak{h}}^{\alpha\beta}\,\partial^\mu\tilde{\mathfrak{h}}_{\alpha\beta} - \frac{1}{4}\,\mathfrak{R}\,\partial_\mu\tilde{\mathfrak{h}}\,\partial^\mu\tilde{\mathfrak{h}} - \frac{1}{2}\,\mathfrak{R}\,\Box\tilde{\mathfrak{h}}\right)\right\} + \mathscr{O}\left([f''(0)]^2\right) \tag{54}$$

$$\tag{55}$$

$$T_m = \eta_{\mu\nu}\,T^{\mu\nu}_m. \tag{56}$$

Before moving on to finding the equations of motion of compact objects, and to conclude this section one may take the LHS of Eq. (28) to be some modified Einstein tensor $\mathscr{G}_{\mu\nu}$, that is

$$\mathscr{G}_{\mu\nu} \equiv f'(R)\,R_{\mu\nu} - \frac{1}{2}\,f(R)\,g_{\mu\nu} - [\nabla_\mu\,\nabla_\nu - g_{\mu\nu}\,\Box]\,f'(R)\,, \tag{57}$$

it can be shown that[32] the modified tensor $\mathscr{G}_{\mu\nu}$ satisfies the same covariant conservation condition as the Einstein tensor $G_{\mu\nu}$, that is

$$\nabla^\mu\,\mathscr{G}_{\mu\nu} = 0 \tag{58}$$

which further implies

$$\nabla^\mu\,T_{\mu\nu} = 0. \tag{59}$$

The above leads to the conclusion that the covariant conservation law holds for $f(R)$ theories of gravity in general, and may help in finding a more generalized second Bianchi identity. However, from what is known in the literature (for example in[32,36])

about the consequence of the covariant conservation law is a peculiar identity on the function $f'(R)$, given by

$$(\Box \nabla_\nu - \nabla_\nu \Box)\, f' = R_{\mu\nu} \nabla^\mu f' \tag{60}$$

which is similar to the appearance of the Riemann curvature tensor due to covariant derivative of a vector field not commuting, implying curvature or non-flatness of the space-time. However, the above commuting relation is on a scalar field, implying that some unconventional geometry might be at large here.

3. Equations of motion from the covariant conservation law

References 22 and 37 showed that it is possible to obtain the post Newtonian sequence of solutions for beyond the weak field scenario, that is when the post Newtonian parameter squared ϵ^2 is proportional to the compactness (or equivalently, the potential at the surface of an ultra-compact object like Neutron stars and small black holes being proportional to the kinetic energy of either bodies). To be precise, in non-geometrized units when

$$\epsilon^2 \sim \frac{G\,M}{c^2\,a} \tag{61}$$

where M and a are the total mass and radius of the compact object. In order that strong surface gravity (post Minkowskian) can be incorporated within the post Newtonian (relativistic) framework, a series of coordinate transformations can be made that requires the integration of wave equations like (16) or (47) and (52) to be performed in the so-called *body-zone coordinate system*, while not worrying about the effects of strong internal gravity.

In order to maintain the compactness of the body to unity at all times, some temporal+spatial scalings are made, with the temporal scaling defining a proper time s (τ in[37]) given by

$$s = \epsilon\, t \tag{62}$$

where t is the coordinate time of the inertial observer. Under the above transformation, the Minkowski metric and the corresponding metric determinant scale as follows

$$\eta_{\mu\nu} = diag\left(\epsilon^{-2}, -1, -1, -1\right) \tag{63}$$

$$-\det\left(\eta_{\mu\nu}\right) = \epsilon^{-2} \tag{64}$$

which is singular for $\epsilon \to 0$. The spatial scalings that are going to be used for the integration in the body zone of Eq. (47) are given by

$$X^i_L \equiv \frac{x^i - z^i_L\,(s)}{\epsilon^2}. \tag{65}$$

The scalings in Eq. (62) and (65) lead to any rank two tensor, or in our case, the energy-momentum-stress density tensor, transforming as

$$T_L^{s\,s} = \epsilon^2\, T^{t\,t} \sim \mathscr{O}\left(\epsilon^{-2}\right) \tag{66}$$

$$T_L^{s\,i'} = \epsilon^{-1}\, T_L^{t\,i} \sim \mathscr{O}\left(\epsilon^{-5}\right) \tag{67}$$

$$T_L^{i'\,j'} = \epsilon^{-4}\, T_L^{i\,j} \sim \mathscr{O}\left(\epsilon^{-8}\right), \tag{68}$$

with the un-primed coordinates representing the unscaled coordinates. Now the above scalings, i.e. Eqs. (62)-(65), are made in a coordinate system that is stationary with respect to the origin or the center of mass of the $L-th$ compact body, denoted by an external observer with the position 3-vector $z_L^i\,(s)$, and the 3-velocity v_L^i defined as

$$v_L^i = \frac{d\, z_L^i\,(s)}{ds}. \tag{69}$$

Hence the next (and final) transformation requires the definition of a co-moving frame (but not co-rotating, since boost and rotation in the wrong order may lead to unnecessary Thomas precession), moving along an ultra compact object whose spin can be ignored. The energy-momentum-stress density tensor of just the matter content of the compact object transforms from an inertial frame to the near-zone or the co-moving frame as follows

$$T_L^{s\,s} = T_L^{s\,s} \tag{70}$$

$$T_L^{s\,i} = \epsilon^2 T_L^{s\,i'} + v_L^i T_L^{s\,s} \tag{71}$$

$$T_L^{i\,j} = \epsilon^4 T_L^{i'\,j'} + 2\epsilon^2 v_L^{(i}\, T_L^{j')s} + v_L^i\, v_L^j T_L^{s\,s}. \tag{72}$$

Then at the leading order the post Newtonian solutions for both GR and $f\,(R)$ theories of gravity, as seen by the inertial observer, are given by

$$\mathfrak{h}_{\mathscr{B}}^{s\,s} = 4\,\epsilon^4 \sum_{L=1,\,2} \left(\frac{P_L^s}{r_L} + \epsilon^2\, \frac{D_L^k\, r_L^k}{r_L^3} \right) + \mathscr{O}\left(\epsilon^8\right) \tag{73}$$

$$\mathfrak{h}_{\mathscr{B}}^{s\,i} = 4\,\epsilon^4 \sum_{L=1,\,2} \left(\frac{P_L^i}{r_L} + \epsilon^2\, \frac{J_L^{k\,i} r_L^k}{r_L^3} \right) + \mathscr{O}\left(\epsilon^8\right) \tag{74}$$

$$\mathfrak{h}_{\mathscr{B}}^{i\,j} = 4\,\epsilon^2 \sum_{L=1,\,2} \left(\frac{Z_L^{i\,j}}{r_L} + \epsilon^2\, \frac{Z_L^{k\,i\,j}\, r_L^k}{r_L^3} \right) + \mathscr{O}\left(\epsilon^6\right) \tag{75}$$

In the current formalism, the above components of the tensor potential are sufficient to obtain equations of motion of the two sources till the first PN order. The monopole and dipole moments, respectively, of various components of $\Lambda^{\mu\nu}$ (or $\tilde{\Lambda}^{\mu\nu}$) are defined

as follows

$$P_L^s = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{s\,s} \tag{76}$$

$$P_L^i = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{s\,i} \tag{77}$$

$$Z_L^{i\,j} = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{i\,j} \tag{78}$$

$$D_L^i = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{s\,s} \, X_L^i \tag{79}$$

$$J_L^{i\,j} = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{s\,i} X_L^j \tag{80}$$

$$Z_L^{i\,j\,k} = \lim_{\epsilon \to 0} \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{i\,j} \, X_L^k \tag{81}$$

where $\mathscr{B}_L$ implies the body zone/co-moving+scaled coordinate system of the $L-th$ body zone.

One defines a 4-momentum, by putting together the scalar in Eq. (76) and the 3-vector in Eq. (77), as[22,37–39]

$$P_L^\mu (s) = \epsilon^2 \int_{\mathscr{B}_L} d^3 X_L \, \Lambda^{s\,\mu} \tag{82}$$

which along with the conservation law (20) or (51), leads to the following 4-force or the 4-momentum evolution formula[22,37,38]

$$\frac{dP_L^\mu}{ds} = -\epsilon^{-4} \oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{k\,\mu} + \epsilon^{-4} \, v_L^k \oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{s\,\mu}. \tag{83}$$

When one defines the bare quasi-local mass of the body zone L to be Eq. (76), the 3-momentum to 3-velocity relation is given at the leading order by

$$P_L^i = P_L^s \, v_L^i + Q_L^i + \mathscr{O}\left(\epsilon^2\right) \tag{84}$$

$$Q_L^i = \epsilon^{-4} \oint_{\partial \mathscr{B}_L} dS_k \left(\Lambda^{s\,k} - v_L^k \, \Lambda^{s\,s}\right) X_L^i. \tag{85}$$

By choosing the origin of the coordinate system to be the center of mass of the body zone L (which is a choice that has already been made), Q_L^i can be shown to vanish at the leading (Newtonian) order, that is $\mathscr{O}(1)$. Therefore, one obtains the 3-velocity evolution relation (or the 3-force on the body zone L) as

$$\begin{aligned} P_L^s \, \frac{dv_L^i}{ds} = & -\epsilon^{-4} \oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{k\,i} + \epsilon^{-4} \, v_L^k \oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{s\,i} + \epsilon^{-4} \, v_L^i \left(\oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{k\,s} \right. \\ & \left. - v_L^k \oint_{\partial \mathscr{B}_L} dS_k \, \Lambda^{s\,s} \right) + \mathscr{O}\left(\epsilon^2\right). \end{aligned} \tag{86}$$

All of the relations in this section (as mentioned along the way) holds both for GR and $f(R)$ theories of gravity, and can be utilized to obtain the 2-body equations of motion of both theories, simply owing to the fact that the second Bianchi identity (or its generalization) holds for both theories of gravity. In GR for example, the equations of motion till 1 PN or $\mathscr{O}\left(\epsilon^2\right)$ of Eq. (86) for one of the bodies (say body zone 1) is given by

$$M_1 \frac{dv_1^i}{ds} = -\frac{M_1 M_2}{r_{1\,2}^2} n^i + \epsilon^2 \frac{M_1 M_2}{r_{1\,2}^2} \left[\left(-v_1^2 - 2\, v_2^2 + \frac{3}{2} \left(\hat{n} \cdot \mathbf{v}_2 \right)^2 + 4 \left(\mathbf{v}_1 \cdot \mathbf{v}_2 \right) \right.\right. \\ \left.\left. + \frac{5\, M_1}{r_{1\,2}} + \frac{4\, M_2}{r_{1\,2}} \right) n^i + \left\{ 4 \left(\hat{n} \cdot \mathbf{v}_1 \right) - 3 \left(\hat{n} \cdot \mathbf{v}_2 \right) \right\} \left(v_1^i - v_2^i \right) \right] + \mathscr{O}\left(\epsilon^4\right). \tag{87}$$

4. Post Newtonian sequence of solutions and equations of motion for the quadratic deviation to the Einstein-Hilbert action

4.1. *The curious case of the homogeneous solution*

One may be tempted to immediately go for the particular solution of the differential equation (52), that is by solving for the following inhomogeneous problem using the retarded Green's function $G_\gamma \left(x^\mu,\, x^{\mu'} \right)$

$$\left(\Box + \sqrt{-g}\, \gamma^2 \right) G_\gamma \left(x^\mu, x^{\mu'} \right) = \delta \left(x^\mu - x^{\mu'} \right). \tag{88}$$

The support of the corresponding integral problem (that is the limit for $x^{\mu'}$ in the Green's function integral) of the above lies in the past light-cone of the event x^μ. However, one must remember that the homogeneous solution of GR leads to the BH solutions. Hence, one must also consider the homogeneous solution of the Klein-Gordon equation, which must be solved simultaneously along with the homogeneous solution of GR.

The homogeneous Klein-Gordon equation of the current theory under study, under the transformation (62), has an ϵ dependence given as follows

$$\Box \mathfrak{R}_{hom} + \frac{\gamma^2}{\epsilon} \mathfrak{R}_{hom} = 0. \tag{89}$$

While the radiative solutions of the above, under the choice of no-incoming scalar radiation from past null infinity, can be put to zero, the time independent part of the homogeneous solution cannot be ignored. The time independent differential equation is given by

$$\nabla \mathfrak{R}_{hom} - \frac{\gamma^2}{\epsilon} \mathfrak{R}_{hom} = 0 \tag{90}$$

where ∇ is the Laplace operator in Minkowski space-time. In spherical symmetry, for example, the LHS of Eq. (90) is given by

$$\frac{1}{r^2}\left[r^2\,\left(\mathfrak{R}_{hom}\right)_{,r}\right]_{,r} - \frac{\gamma^2}{\epsilon}\,\mathfrak{R}_{hom} = 0 \tag{91}$$

which has a solution which is both regular at $r = 0$ and $r = \infty$, and is given by

$$\mathfrak{R}_{hom} = \frac{C\,e^{-\frac{\gamma\,r}{\sqrt{\epsilon}}}}{r} \tag{92}$$

where C is a real constant of integration yet to be fixed. It is to be noted that the homogeneous solution, which may be part of an extended homogeneous solution of the combined scalar+tensor theory under study (like the Schwarzschild and Kerr solutions of GR), for vanishing post Newtonian or post Minkowskian parameter ϵ, leads to an infinite mass of the scalar field. Which also implies for low energies, the Ricci scalar or the Ricci scalar density considered here is practically un-excitable.

However, it is interesting to look at the effect of $\mathfrak{R}$ on the tensor wave equation (39), and how it modifies the Newtonian order solution, or more specifically, the Newtonian potential. The time-time component of the conformally transformed metric density deviation satisfies the flat-space wave equation with a source, following Eq. (47)

$$\Box\tilde{\mathfrak{h}}^{t\,t} = 16\,\pi\,(-g)\,T^{t\,t}\,(1 - \epsilon\,f''\,\mathfrak{R}) \tag{93}$$

where the leading order, or only the terms in Eq. (48) that are proportional to the classical or matter energy-momentum-stress density tensor were used, along with Eq. (64). The above can be solved by using the retarded Green's function for the massless wave operator, with $\tilde{\mathfrak{h}}^{s\,s}$ given by the body zone integral

$$\tilde{\mathfrak{h}}^{s\,s} = 4\,\epsilon^4 \sum_{L=1,\,2}\int_{\mathscr{B}_L} d^3X_L\,\frac{T^{s\,s}\,(1 - \epsilon\,f''\,\mathfrak{R})}{\left|Z^i_L - \epsilon^2\,X^i_L\right|} \tag{94}$$

Eq. (64) and (66) was used to obtain the above from Eq. (93). The first integral, and its ϵ expansion of the series is given in Eq. (73), and is the usual Newtonian potential. However, a more curious thing occurs when one considers the extra term coming due to the quadratic modification to the gravitational action

$$\tilde{\mathfrak{h}}^{s\,s}_{extra} = -4\,\epsilon^5\,f'' \sum_{L=1,\,2}\int_{\mathscr{B}_L} d^3X_L\,\frac{T^{s\,s}\,\mathfrak{R}_{hom}}{\left|Z^i_L - \epsilon^2\,X^i_L\right|}. \tag{95}$$

Considering now the particular form of the homogeneous Ricci density (92), one obtains the following integral using the scaled coordinate (65)

$$\tilde{\mathfrak{h}}^{s\,s}_{extra} = -4\,\epsilon^3\,f'' \sum_{L=1,\,2} C_L\int_{\mathscr{B}_L} d^3X_L\,\frac{T^{s\,s}}{\left|Z^i_L\,(s) - \epsilon^2\,X^i_L\right|\,\left|X^i_L\right|}\,e^{-\gamma\,\epsilon^{\frac{3}{2}}\,\left|X^i_L\right|} \tag{96}$$

which seems to preceed the order at which the gothic metric deviation first appears, that is at $\mathscr{O}\left(\epsilon^3\right)$ compared to $\mathscr{O}\left(\epsilon^4\right)$. However, expanding the extra term about $\epsilon \to 0$ leads to the following

$$\tilde{\mathfrak{h}}^{s\,s}_{extra} = -4\,\epsilon^3\,f'' \sum_{L=1,\,2} \frac{C_L}{\left|Z^i_L\right|} \int_{\mathscr{B}_L} d^3X_L\,\frac{T^{s\,s}}{\left|X^i_L\right|} \tag{97}$$

$$\tilde{\mathfrak{h}}^{s\,s}_{extra} = -4\,\epsilon^3\,f'' \sum_{L=1,\,2} \frac{C_L\,N_L}{\left|X^i_L\right|} \tag{98}$$

which is a Newtonian potential like term with negative one multipole moment of the source, denoted by N_L. Therefore, the modified leading order metric deviation $\tilde{\mathfrak{h}}^{s\,s}$ can be written in the following form

$$\tilde{\mathfrak{h}}^{s\,s} = 4\,\epsilon^4 \sum_{L=1,\,2} \left[\frac{\tilde{P}^s_L}{\left|Z^i_L\right|} + \mathscr{O}\left(\epsilon^2\right)\right] \tag{99}$$

$$\tilde{P}^s_L = P^s_L - \lim_{\epsilon\,\to\,0} \frac{f''\,C_L}{\epsilon} \int_{\mathscr{B}_L} d^3X_L\,\frac{T^{s\,s}}{\left|X^i_L\right|} \tag{100}$$

where P^s_L was defined in Eq. (76). It is important to note that while the integral in the second term of the square bracket of the above has a well defined limit for $\epsilon \to 0$, the coefficient multiplying it does not. In fact, due to the ϵ^{-1} nature of the coefficient, the modified mass-energy definition blows up at the $\epsilon \to 0$ limit, leading to the following conclusion

$$C_L = 0, \tag{101}$$

implying that the homogeneous Ricci scalar does not modify the tensor deviation density at all.

4.2. *Deviation from the general relativistic equations of motion*

The leading order Ricci scalar density deviation occurs at $\mathscr{O}\left(\epsilon^4\right)$, and is given by the following

$${}_{(4)}\mathfrak{R}_{part} \equiv {}_{(4)}\mathfrak{R} = -8\,\pi\,\gamma^2\,\epsilon^4\,\frac{M_L\,e^{-\frac{\gamma\left|Z^i_L\right|}{\sqrt{\epsilon}}}}{\left|Z^i_L\right|} \tag{102}$$

whose form is the same as the now extinct homogeneous solution, but instead of the unknown constant, there is just the mass coming as a parameter, as in the definition of Eq. (76). The only unknown in the above equation is the inverse length scale γ, which was related to the coefficient of the quadratic deviation of the GR action in Eq. (53). Derivation of the above has been given in Appendix A.

In GR, the non-vanishing and body zone independent terms in the surface integrals of Eq. (86) lead to the 3-force at different PN orders, and has been well

documented in PN literature, for example.[40,41] However, in this study only the leading order modification to the general relativistic equations of motion due to the quadratic deviation to the GR action will be quoted. Such a modification can be found by using the modified conservation law (51) along with the definitions of (83) and (86) with $\Lambda^{\mu\nu}$ replaced by $\tilde{\Lambda}^{\mu\nu}$. In Eq. (86), various components of the modified net energy-momentum-stress density tensor $\tilde{\Lambda}^{\mu\nu}$ will appear as integrand of the surface integrals. However, the components that have a non-vanishing and *body zone boundary radius* independent contribution to the surface integrals of Eq. (86), was found to be coming from the space-space component of $t^{\mu\nu}_{eff}$, or

$$t^{i\,j}_{eff} = \frac{\epsilon^4}{192\,\pi\,\gamma^4}\,\Big[2\,\left(\eta^{i\,k}\,\eta^{j\,l}\,+\,\eta^{i\,l}\,\eta^{j\,k}\,-\,\eta^{i\,j}\,\eta^{k\,l}\right)\,\mathfrak{R}\,\partial_k\mathfrak{R}\,\partial_l\tilde{\mathfrak{h}} + \left(\eta^{i\,k}\,\eta^{j\,l}\,-\,\frac{1}{2}\,\eta^{i\,j}\,\eta^{k\,l}\right)\,\left(4\,\partial_k\mathfrak{R}\,\partial_l\mathfrak{R}\,+\,\mathfrak{R}^2\,\partial_k\tilde{\mathfrak{h}}\,\partial_l\tilde{\mathfrak{h}}\right)\Big] \tag{103}$$

where Eq. (64) was used to replace the metric determinant, or $\frac{1}{(-g)^2}$ multiplying the quadratic f'' part of $t^{\mu\nu}_{eff}$ in Eq. (49), which appear as ϵ^4 above.

Eq. (103) has three different ϵ orders worth of terms inside the square bracket. When one utilizes the scalings of the leading order deviation of either $\tilde{\mathfrak{h}}$ or $\mathfrak{R}$, one notices that $\tilde{\mathfrak{h}} \sim \epsilon^2$ and $\mathfrak{R} \sim \epsilon^4$, following Eqs. (63), (73)-(75), and (102). From the number of times either $\tilde{\mathfrak{h}}$ and $\mathfrak{R}$ appear inside the square bracket in each product of fields, the first term scale as ϵ^{10}, the second as ϵ^8, and the last one as ϵ^{12}. Since the overall factor of ϵ^4 cancels away due to ϵ^{-4} multiplying each surface integral of Eq. (86), the three terms inside the square brackets of Eq. (103) are 5th, 4th, and 6th PN effects on the two body equations of motion. However, not all of them survive the surface integral. Only the last term in the square bracket survives and is a 6th PN effect with the modification to the 3-force on the L^{th} body given as follows

$$M_L\,\frac{dv^i_L}{ds} = F^i_{N+PN,GR} \,-\, \epsilon^{12}\,\frac{32\,\pi\,M_L\,M^3_{L'}}{9\,r^4_{1\,2}}\,e^{-\frac{2\,\gamma\,r_{1\,2}}{\sqrt{\epsilon}}}\,n^i \tag{104}$$

where M_L is the quasi-local mass enclosed by the L^{th} body zone boundary, and $M_{L'}$ is the quasi-local mass enclosed by the partner object's body zone boundary in the two-body system. $r_{1\,2}$ is the distance between the origin of the two body zones, whereas n^i is a unit vector pointing from the origin of body zone L to the origin of body zone L'. The negative sign on the modified/extra force imply that the direction of the force is antiparallel to n^i, which is then a repulsive modification to the usual attractive Newtonian and PN forces indicated by the first term in the RHS of the above, that is $F^i_{N+PN,GR}$. Compared to other works in literature on $f\,(R)$ theories that calculate the conserved potential (or the force) from modifications to the time-time component of the metric tensor, the above has an inverse quartic distance falloff (along with the Yukawa exponential falloff) at 6 PN order, which is

an extremely tiny modification to the general relativistic PN equations of motion at the very high energy scales of ultra-compact two-body interactions.

5. Discussions and conclusion

In the current study a particular redefinition of the standard gothic metric was utilized to drastically simplify the field equations of general $f(R)$ theories of gravity. Once the identification was made that $f'(R)$ acts like a massive scalar field, a fourth order system was reduced to two second order system of differential equations. The field equations were framed in the densitized approach following the methods of Landau and Lifshitz, and their redefinition of the field equations of GR. The use of conformal densities to define Landau-Lifshitz like field equations were used in scalar-tensor theories, for example in.[42,43] In this study, it is shown that the redefinition to a conformal density greatly suppresses the actual effect of an extra massive scalar degree of freedom on orbital dynamics.

Following the definition of a tensor density, a scalar density $\mathfrak{R}$, corresponding to the massive scalar degree of freedom was defined, which relates to the trace of the energy-momentum-stress density tensor in the same manner as the conformal metric tensor density relates to the energy-momentum-stress density tensor in the Landau-Lifshitz like tensor field equations of $f(R)$ theories. The differential equation that is satisfied by $\mathfrak{R}$ throws up a static solution that makes the mass-energy definition at the Newtonian/leading order diverge, leading to its removal from further usage. At the leading order, the particular/inhomogeneous $\mathfrak{R}$ is seen to be static as well, and has the Yukawa potential form with a cutoff/inverse length scale that is related to the coefficient of the quadratic modification (with $(\text{length})^2$ dimensions) to the Einstein-Hilbert action.

However, just finding out the particular form of the leading order deviation of $\mathfrak{R}$ is quite useless, unless one finds its corresponding effect on the conformal tensor density field equations, and then, on the PN equations of motion. The leading order $\mathfrak{R}$ appears as an effective source tensor that drives the behavior of $\tilde{\mathfrak{h}}^{\mu\nu}$, as well as the evolutionary dynamics of the quasi-local four-momentum (83), and correspondingly, the velocity evolution equations (86) through surface integrals of various components of the energy-momentum-stress density tensor. It was found in Eq. (104) that the modification to the GR PN equations of motion, due to quadratic modification at the action level, appear at the sixth PN order. On top of that, the force has an inverse fourth powered dependence on the distance along with a Yukawa falloff, which is a very tiny modification to GR equations of motion. The modified force is dependent on the masses enclosed by the body zone boundaries, and the modification to the equations of motion of one of the objects is independent of the mass enclosed in that object's body zone boundary - implying that the equivalence principle (as far as it holds in the GR PN scheme) is not affected by the quadratic

modification to GR. The modified force is also conservative, and hence will not be adding to any extra dissipation (other than GR) of the orbital energy of the binary system, although it might lead to very tiny modifications to the phase evolution of the binary, as encoded in the gravitational waves. More importantly, the force is repulsive, compared to the PN forces that appear on perturbing GR about Minkowski space-times. So in the course of the binary evolution, at length scales comparable to γ^{-1}, if the two objects have not merged into a single object already, they will feel a repulsive force that will compete with the attractive GR PN forces coming out of the *superpotentials* of the PN expansion of GR, which are themselves progressively short ranged. Putting a constraint on the new length scale γ^{-1} is beyond the scope of this study, mostly because the author is not aware of sixth PN equations of motion appearing in the literature.

Appendix A

The net solution to the Klein-Gordon problem (52) with the boundary condition choice of no incoming scalar radiation from past null Minkowskian infinity is given as follows

$$\mathfrak{R}\left(x^{\mu}\right) = -8\,\pi\,\gamma^{2}\,\epsilon^{-2}\int d^{4}y\,G_{\tilde{\gamma}}\left(x^{\mu},\,y^{\mu}\right)\,T_{m}\left(y^{\mu}\right) \tag{A.1}$$

$$\tilde{\gamma}^{2} \equiv \frac{\gamma^{2}}{\epsilon} \tag{A.2}$$

where the factor of ϵ^{-2} arises because the metric determinant $(-g)$ multiplying the RHS of Eq. (52), under the transformation (62), scale as ϵ^{-2}, as seen after Eq. (63). $G_{\tilde{\gamma}}\left(x^{\mu},\,y^{\mu}\right)$ is the retarded Green's function of the Klein-Gordon equation, as was given in[32] with the $(+\,,-\,,-\,,-)$ metric signature as

$$\begin{aligned} G_{\tilde{\gamma}}\left(t,\,q;\,x^{i},\,y^{i}\right) = &\int_{-\infty,\,\tilde{\gamma}}^{-\tilde{\gamma},\,\infty}\frac{d\,\omega}{2\,\pi}\,e^{-i\,\omega(t-q)}\,\frac{e^{i\,\sqrt{\omega^{2}-\tilde{\gamma}^{2}}\,\left|x^{i}-y^{i}\right|}}{4\,\pi\,\left|x^{i}-y^{i}\right|} \\ &+\int_{-\tilde{\gamma}}^{\tilde{\gamma}}\frac{d\omega}{2\,\pi}\,e^{-i\,\omega\,(t-q)}\,\frac{e^{-\,\sqrt{\tilde{\gamma}^{2}-\omega^{2}}\,\left|x^{i}-y^{i}\right|}}{4\,\pi\,\left|x^{i}-y^{i}\right|}. \end{aligned} \tag{A.3}$$

The notation $\int_{-\infty,\,\tilde{\gamma}}^{-\tilde{\gamma},\,\infty}$ involve two integrals, one from $-\infty$ to $-\tilde{\gamma}$, and the other from $\tilde{\gamma}$ to ∞. It is to be noted that the three integrals whose domains encompass all of ω space must be evaluated simultaneously in order for the solution to converge. $T_{m}\left(q,\,y^{i}\right)$, the trace of the classical matter energy-momentum tensor, which being a scalar, does not transform under any of the coordinate scalings.

Due to the the coordinate scalings in Eq. (65), the infinitesimal 4-volume element $d^{4}y$ transform from the asymptotic observer's frame to either of the body zones $\mathscr{B}_{L}$ in the following manner

$$d^{4}y \equiv dt\,\wedge\,d^{3}y \tag{A.4}$$

$$\rightarrow \epsilon^{-1}\,ds\,\wedge\,\epsilon^{6}\,d^{3}X_{L} \tag{A.5}$$

where $\wedge$ denotes the wedge product between 1-form dt and 3-form d^3y. In the current operational context $\wedge$ is effectively scalar multiplication.

Substituting Eq. (A.3) in Eq. (A.1), and transforming into the body zone coordinates by substituting Eq. (A.5) as the infinitesimal covariant volume element in Eq. (A.1), one obtains the following integral for the particular solution $\mathfrak{R}_{part}$ for the inhomogeneous Klein-Gordon equation

$$\begin{aligned}
\mathfrak{R}_{part}\left(t,\, x^i\right) = -\epsilon^3\; 8\,\pi\,\gamma^2 \int_{-\infty}^{\infty} ds' \int_{\mathscr{B}_L} d^3X_L \int_{-\infty,\,\tilde{\gamma}}^{-\tilde{\gamma},\,\infty} \frac{d\omega}{2\,\pi}\, e^{-i\,\omega\left(t-s'/\epsilon\right)} \\
\times \quad \frac{e^{i\,\sqrt{\omega^2-\tilde{\gamma}^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\,\left[\epsilon^4\,T_m\left(s'/\epsilon,\, X_L^i\right)\right] \\
-\epsilon^3 \quad 8\,\pi\,\gamma^2 \int_{-\infty}^{\infty} ds' \int_{\mathscr{B}_L} d^3X_L \int_{-\tilde{\gamma}}^{\tilde{\gamma}} \frac{d\omega}{2\,\pi}\, e^{-i\,\omega\left(t-s'/\epsilon\right)} \\
\times \quad \frac{e^{-\sqrt{\tilde{\gamma}^2-\omega^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\,\left[\epsilon^4\,T_m\left(s'/\epsilon,\, X_L^i\right)\right]. \qquad \text{(A.6)}
\end{aligned}$$

In the body zones that surround each of the objects, the energy-momentum tensor of classical matter $\left(T_{\mathscr{B}_L} = \lim_{\epsilon\,\to\,0}\epsilon^4\,T_m\right)$ will be assumed to have a quasi-stationary/adiabatic initial condition for solving the relaxed system of equations, in the absence of any other time dependent driving force inside the compact object, and is given by

$$T_{\mathscr{B}_L}\left(X^i\right) = \sum_{n=-\infty}^{\infty} \mathfrak{T}_n\left(\left|X_L^i\right|, X_L^\theta\right)\, e^{i\,n\,\mathscr{X}} \qquad \text{(A.7)}$$

$$\mathscr{X}\left(s/\epsilon,\, X_L^\phi\right) \equiv \mathscr{X} = X_L^\phi - \Omega\, s/\epsilon + \phi_0 \qquad \text{(A.8)}$$

where an axial+time symmetry was assumed for the initial condition, such that the time dependence of the source body in the body zone coordinate system repeats after every $T = \frac{2\,\pi}{\Omega}$, with a constant phase parameter ϕ_0. The weighing factors $\mathfrak{T}_n$ are coefficients in the series expansion of the trace T_m (or $T_{\mathscr{B}_{\mathscr{L}}}$) using stationary functions $e^{i\,n\,\mathscr{X}}$, and are functions of the radial and azimuthal coordinates. Eq. (A.7) physically implies that the variations in the energy-momentum tensor sourcing $\mathfrak{R}$ in the body zone of the first object is purely generated by the effect of the motion of the second object around it. The body zone coordinates $\left(X_L^i\right)$ were defined in Eq. (65), in which $\left|X_L^i\right|$ is the distance from the center of mass of the body L to any point in the body zone coordinates, as viewed in the respective body zones. The choice in Eq. (A.7) has a simplifying effect on the subsequent calculations and is justified by the adiabatic and stationarity in the co-moving frame approximations, as found in the literature on PN expansions.

Assuming quasi-periodicity for the source tensor, as in Eq. (A.7), one can substitute it into Eq. (A.6), to obtain

$$\begin{aligned}\mathfrak{R}_{part}\left(t,\,x^i\right) = &-\epsilon^3\,\frac{8\,\pi\,\gamma^2}{2\,\pi}\sum_{n=-\infty}^{\infty}\int_{\mathscr{B}_L} d^3X_L\left(\int_{-\infty,\,\tilde{\gamma}}^{-\tilde{\gamma},\,\infty} d\omega\, e^{-i\,\omega\,t}\right) e^{i\,n\left(X_L^{\phi}+\phi_0\right)}\\ &\times\quad \frac{e^{i\,\sqrt{\omega^2-\tilde{\gamma}^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\,\mathfrak{T}_n\left(\left|X_L^i\right|,X_L^{\theta}\right)\int_{-\infty}^{\infty} ds'\,e^{i\,(\omega-n\,\Omega)\,s'/\epsilon}\\ &-\epsilon^3\quad\frac{8\,\pi\,\gamma^2}{2\,\pi}\sum_{n=-\infty}^{\infty}\int_{\mathscr{B}_L} d^3X_L\left(\int_{-\tilde{\gamma}}^{\tilde{\gamma}} d\omega\, e^{-i\,\omega\,t}\right) e^{i\,n\left(X_L^{\phi}+\phi_0\right)}\\ &\times\quad \frac{e^{-\sqrt{\tilde{\gamma}^2-\omega^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\,\mathfrak{T}_n\left(\left|X_L^i\right|,X_L^{\theta}\right)\int_{-\infty}^{\infty} ds'\,e^{i\,(\omega-n\,\Omega)\,s'/\epsilon}.\end{aligned} \tag{A.9}$$

Transforming

$$s' \to \epsilon\, s \tag{A.10}$$

and using the integral representation of the delta function,

$$\int_{-\infty}^{\infty} e^{i\,k\,(x-x')}\,dk = 2\,\pi\,\delta\,(x-x')\,, \tag{A.11}$$

one obtains

$$\begin{aligned}\mathfrak{R}_{part}\left(t,\,x^i\right) = &-\epsilon^4\;8\,\pi\,\gamma^2\sum_{n=-\infty}^{\infty}\int_{\mathscr{B}_L} d^3X_L\,\frac{e^{i\,\sqrt{n^2\,\Omega^2-\tilde{\gamma}^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\\ &\times\quad \mathfrak{T}_n\left(\left|X_L^i\right|,X_L^{\theta}\right)\times e^{-i\,n\left(\Omega\,t-X_L^{\phi}-\phi_0\right)} \qquad\text{(A.12)}\\ &-\epsilon^4\;8\,\pi\,\gamma^2\sum_{n=-\infty}^{\infty}\int_{\mathscr{B}_L} d^3X_L\,\frac{e^{-\sqrt{\tilde{\gamma}^2-n^2\,\Omega^2}\left|Z_L^i-\epsilon^2\,X_L^i\right|}}{4\,\pi\,\left|Z_L^i-\epsilon^2\,X_L^i\right|}\\ &\times\quad \mathfrak{T}_n\left(\left|X_L^i\right|,X_L^{\theta}\right)\,e^{-i\,n\left(\Omega\,t-X_L^{\phi}-\phi_0\right)}. \qquad\text{(A.13)}\end{aligned}$$

The Green's function (A.3) can be written as an infinite sum of spherical harmonic functions, that are weighed by functions of the radial coordinates $\left|Z_L^i\right|$ and $\left|X_L^i\right|$, in the following way for $\left|Z_L^i\right| > \left|X_L^i\right|$

$$\frac{e^{i\,k\left|X^i-Z^i\right|}}{4\,\pi\,\left|X^i-Z^i\right|} = i\,k\sum_{\ell,\,m} j_{\ell}\left(k\,\left|X^i\right|\right)\,h_{\ell}^{(1)}\left(k\,\left|Z^i\right|\right)\,Y^*_{\ell\,m}\left(X^{\theta},\,X^{\phi}\right)\,Y_{\ell\,m}\left(Z^{\theta},\,Z^{\phi}\right) \tag{A.14}$$

$$\frac{e^{-\,k\left|X^i-Z^i\right|}}{4\,\pi\,\left|X^i-Z^i\right|} = \sum_{\ell,\,m}\frac{I_{\ell+\frac{1}{2}}\left(k\,\left|X^i\right|\right)\,K_{\ell+\frac{1}{2}}\,\left(k\,\left|Z^i\right|\right)}{\sqrt{\left|X^i\right|\,\left|Z^i\right|}}\,Y^*_{\ell\,m}\left(X^{\theta},\,X^{\phi}\right)\,Y_{\ell\,m}\left(Z^{\theta},\,Z^{\phi}\right)$$

$$Z_L^i \equiv \frac{x^i\,-\,z_L^i\,(s)}{\epsilon^2}\,;\quad L\,=\,1,\,2. \tag{A.15}$$

The various functions appearing above are as follows

- j_ℓ: Spherical Bessel function of first kind.
- $h^{(1)}_\ell$: Spherical Bessel function of third kind.
- $I_{\ell+\frac{1}{2}}$: Modified Bessel function of the first kind.
- $K_{\ell+\frac{1}{2}}$: Modified Bessel function of the second kind.
- $Y_{\ell\,m}$, $Y^*_{\ell\,m}$: Spherical harmonic functions and their complex conjugates.

After using Eqs. (A.14) and (A.15) in the above, one obtains for $\left|Z^i_L\right| > \left|X^i_L\right|$

$$\begin{aligned}
\mathfrak{R}_{part}\left(t,\, x^i\right) = -\epsilon^4\, 8\,\pi\, i\,\gamma^2 \sum_{n,\,\ell,\,m}^{|n|>\left\lfloor\frac{\tilde{\gamma}}{\Omega}\right\rfloor} \sqrt{n^2\,\Omega^2-\tilde{\gamma}^2} \int_{\mathscr{B}_L} d^3X_L\, e^{i\,n\left(X^\phi_L+\phi_0-\Omega\,t\right)} \\
\times \quad j_\ell\left(\epsilon^2\sqrt{n^2\,\Omega^2-\tilde{\gamma}^2}\,\left|X^i_L\right|\right)\, h^{(1)}_\ell\left(\sqrt{n^2\,\Omega^2-\tilde{\gamma}^2}\,\left|Z^i\right|\right) \\
\times \quad Y^*_{\ell\,m}\left(X^\theta_L,\, X^\phi_L\right)\, Y_{\ell\,m}\left(Z^\theta,\, Z^\phi\right)\, \mathfrak{T}_n\left(\left|X^i_L\right|,\, X^\theta_L\right) \\
-\epsilon^3 \quad 8\,\pi\,\gamma^2 \sum_{n,\,\ell,\,m}^{|n|<\left\lfloor\frac{\tilde{\gamma}}{\Omega}\right\rfloor} \int_{\mathscr{B}_L} d^3X_L\, e^{i\,n\left(X^\phi_L+\phi_0-\Omega\,t\right)} \\
\times \quad \frac{I_{\ell+\frac{1}{2}}\left(\epsilon^2\sqrt{\tilde{\gamma}^2-n^2\,\Omega^2}\,\left|X^i_L\right|\right)}{\sqrt{\left|X^i_L\right|}}\, \frac{K_{\ell+\frac{1}{2}}\left(\sqrt{\tilde{\gamma}^2-n^2\,\Omega^2}\,\left|Z^i\right|\right)}{\sqrt{|Z^i|}} \\
\times \quad Y^*_{\ell\,m}\left(X^\theta_L,\, X^\phi_L\right)\, Y_{\ell\,m}\left(Z^\theta,\, Z^\phi\right)\, \mathfrak{T}_n\left(\left|X^i_L\right|,\, X^\theta_L\right) \qquad \text{(A.16)}
\end{aligned}$$

one can substitute the following properties of spherical harmonic functions in Eq. (A.16)

$$Y^*_{\ell\,m} = (-1)^m\, Y_{\ell\,-m} \qquad \text{(A.17)}$$

$$Y_{\ell\,-m} = N_{\ell\,-m}\, P_{\ell\,-m}\left(\cos\, X^\theta_L\right)\, e^{-i\,m\,X^\phi} \qquad \text{(A.18)}$$

$$\int_0^{2\,\pi} dX^\phi_L\, Y^*_{\ell\,m}\, e^{i\,n\,X^\phi} = 2\,\pi\,(-1)^m\, N_{\ell\,-m}\, P_{\ell\,-m}\left(X^\theta_L\right)\, \delta_{m\,n} \qquad \text{(A.19)}$$

$$N_{\ell\,m} = (-1)^m \sqrt{\frac{(2\,\ell\,+\,1)}{4\,\pi}\,\frac{(\ell\,-\,m)!}{(\ell\,+\,m)!}} \qquad \text{(A.20)}$$

where $P_{\ell\,m}\left(X^\theta_L\right)$ are the associated Legendre polynomials and $\delta_{m\,n}$ is the Kronecker delta distribution. Since m takes values between $-\ell$ to ℓ, application of the Kronecker delta leads to the summation on ℓ in the first term of Eq. (A.16) going from the lower integral part of $\frac{\tilde{\gamma}}{\Omega}$ or $\left\lfloor\frac{\tilde{\gamma}}{\Omega}\right\rfloor$ to ∞, whereas the second terms' summation range takes ℓ from zero to $\left\lfloor\frac{\tilde{\gamma}}{\Omega}\right\rfloor$. The contributions to the particular solution from above the cutoff energy scale $\tilde{\gamma}$ is highly suppressed (especially at lower PN orders) because of the properties of j_ℓ and $h^{(1)}_\ell$ for asymptotically large values of ℓ, and

terms in $\mathfrak{R}$ series containing the same can be approximated to zero at the current orders of calculation. Hence one can safely set the first summation in RHS of Eq. (A.16) to zero and consider only the second summation. Hence it is to be noted that the factors that depend on ϵ in the above are only in the arguments of the Bessel function $I_{\ell+\frac{1}{2}}$.

Hence the only relevant factors that explicitly contain ϵ in both the sums in Eq. (A.16), can be expanded about $\epsilon \to 0$ as follows

$$I_{\ell+\frac{1}{2}}\left(\epsilon^2\,\Lambda\,|X_L|\right) \approx \epsilon^{2\,\ell+1}\,\frac{\Lambda^{\frac{2\,\ell+1}{2}}}{2^{\ell+\frac{1}{2}}\,\Gamma\left(\ell\,+\,\frac{3}{2}\right)}\,|X_L|^{\ell+\frac{1}{2}}\,. \tag{A.21}$$

The approximation of (A.21), when substituted in for the relevant factors of the RHS of Eq. (A.16), leads to the following polynomial series of ϵ about $\epsilon \to 0$

$$\begin{aligned}
\mathfrak{R}_{part}\left(t,\,x^i\right) \approx\, & -8\,\pi\,\gamma^2\sum_{L,\,\ell,\,m}^{\ell=\infty}\epsilon^{2\,\ell+4}\,\frac{\left(\tilde{\gamma}^2\,-\,m^2\,\Omega^2\right)^{\frac{2\,\ell+1}{4}}}{2^{\ell+\frac{1}{2}}\,\tilde{\gamma}\left(\ell\,+\,\frac{3}{2}\right)}\\
&\times e^{i\,m\,(\phi_0\,-\,\Omega\,t)}\,\frac{K_{\ell+\frac{1}{2}}\left(\sqrt{\tilde{\gamma}^2\,-\,m^2\,\Omega^2}\,\left|Z_L^i\right|\right)}{\sqrt{\left|Z_L^i\right|}}\,Y_{\ell\,m}\left(Z_L^\theta,\,Z_L^\phi\right)\\
&\times \mathfrak{M}_{L\,\ell\,m}
\end{aligned} \tag{A.22}$$

$$\begin{aligned}
\mathfrak{M}_{L\,\ell\,m} =\, & 2\,\pi\,(-1)^m\,N_{\ell\,-m}\int_{\mathscr{B}_L}\left|X_L^i\right|^{2+\ell}\sin\left(X_L^\theta\right)\\
&\times P_{\ell\,-m}\left(\cos X_L^\theta\right)\,\mathfrak{T}_m\left(\left|X_L^i\right|,\,X_L^\theta\right)\,d\left|X_L^i\right|\,dX_L^\theta.
\end{aligned} \tag{A.23}$$

From Eq. (A.22), one can immediately notice that the leading order deviation for the particular solution of the Ricci scalar for the L^{th} body zone comes at $\mathscr{O}\left(\epsilon^4\right)$ for $\ell = 0$ with $\mathfrak{M}_{L\,0\,0} \equiv M_L$ being the mass monopole, and was found to be

$${}_{(4)}\mathfrak{R}_{part} \equiv {}_{(4)}\mathfrak{R} \,=\, -8\,\pi\,\gamma^2\,\epsilon^4\,\frac{M_L\,e^{-\frac{\gamma\left|Z_L^i\right|}{\sqrt{\epsilon}}}}{\left|Z_L^i\right|}. \tag{A.24}$$

References

1. A. Einstein, Explanation of the Perihelion Motion of Mercury from the General Theory of Relativity, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 831 (1915).
2. C. M. Will, The 1919 measurement of the deflection of light, *Class. Quant. Grav.* **32**, p. 124001 (2015).
3. A. Einstein, Approximative Integration of the Field Equations of Gravitation, *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl* , 688 (1916).
4. B. P. Abbott *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**, p. 061102 (2016).
5. B. P. Abbott *et al.*, GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence, *Phys. Rev. Lett.* **116**, p. 241103 (2016).
6. B. P. Abbott *et al.*, GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence, *Phys. Rev. Lett.* **119**, p. 141101 (2017).

7. B. P. Abbott *et al.*, First search for nontensorial gravitational waves from known pulsars, *Phys. Rev. Lett.* **120**, p. 031104 (2018).
8. B. P. Abbott *et al.*, GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2, *Phys. Rev. Lett.* **118**, p. 221101 (2017).
9. K. Akiyama *et al.*, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, p. L1 (2019).
10. A. Einstein and L. Infeld, On the motion of particles in general relativity theory, *Canadian Journal of Mathematics* **1**, 209 (1949).
11. L. Infeld and A. Schild, On the Motion of Test Particles in General Relativity, *Reviews of Modern Physics* **21**, 408 (July 1949).
12. S. Chandrasekhar, The Post-Newtonian Equations of Hydrodynamics in General Relativity, **142**, p. 1488 (November 1965).
13. S. Chandrasekhar, The Post-Newtonian Effects of General Relativity on the Equilibrium of Uniformly Rotating Bodies. II. The Deformed Figures of the Maclaurin Spheroids, **147**, p. 334 (January 1967).
14. S. Chandrasekhar and Y. Nutku, The Second Post-Newtonian Equations of Hydrodynamics in General Relativity, **158**, p. 55 (October 1969).
15. S. Chandrasekhar, Conservation Laws in General Relativity and in the Post-Newtonian Approximations, **158**, p. 45 (October 1969).
16. S. Chandrasekhar and F. P. Esposito, The 2andonehalf-POST-NEWTONIAN Equations of Hydrodynamics and Radiation Reaction in General Relativity, **160**, p. 153 (April 1970).
17. A. A. Starobinsky, Disappearing cosmological constant in f(r) gravity, *JETP Letters* **86**, 157 (2007).
18. S. Nojiri and S. D. Odintsov, Introduction to modified gravity and gravitational alternative for dark energy, *eConf* **C0602061**, p. 06 (2006).
19. S. Capozziello, A. Stabile and A. Troisi, Newtonian limit off(r)gravity, *Physical Review D* **76** (November 2007).
20. S. Capozziello, C. Corda and M. F. De Laurentis, Massive gravitational waves from f (R) theories of gravity: Potential detection with LISA, *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **669**, 255 (2008).
21. S. Capozziello, A. Stabile and A. Troisi, The post-minkowskian limit of f(r)-gravity, *International Journal of Theoretical Physics* **49**, 1251 (2010).
22. T. Futamase and Y. Itoh, The post-newtonian approximation for relativistic compact binaries, *Living Reviews in Relativity* **10** (2007).
23. L. LANDAU and E. LIFSHITZ, Chapter 11 - the gravitational field equations, in *The Classical Theory of Fields (Fourth Edition)*, eds. L. LANDAU and E. LIFSHITZ, Course of Theoretical Physics, Vol. 2 (Pergamon, Amsterdam, 1975) pp. 259–294, fourth edition edn.
24. L. LANDAU and E. LIFSHITZ, Chapter 12 - the field of gravitating bodies, in *The Classical Theory of Fields (Fourth Edition)*, eds. L. LANDAU and E. LIFSHITZ, Course of Theoretical Physics, Vol. 2 (Pergamon, Amsterdam, 1975) pp. 295–344, fourth edition edn.
25. L. LANDAU and E. LIFSHITZ, Chapter 13 - gravitational waves, in *The Classical Theory of Fields (Fourth Edition)*, eds. L. LANDAU and E. LIFSHITZ, Course of Theoretical Physics, Vol. 2 (Pergamon, Amsterdam, 1975) pp. 345 – 357, fourth edition edn.
26. S. Nojiri and S. D. Odintsov, Unified cosmic history in modified gravity: from F(R) theory to Lorentz non-invariant models, *Phys. Rept.* **505**, 59 (2011).

27. G. Magnano and L. M. Sokolowski, On physical equivalence between nonlinear gravity theories and a general relativistic selfgravitating scalar field, *Phys. Rev. D* **50**, 5039 (1994).
28. S. Chakraborty, S. Pal and A. Saa, Dynamical equivalence of $f(R)$ gravity in Jordan and Einstein frames, *Phys. Rev. D* **99**, p. 024020 (2019).
29. S. Bahamonde, S. D. Odintsov, V. K. Oikonomou and P. V. Tretyakov, Deceleration versus acceleration universe in different frames of $F(R)$ gravity, *Phys. Lett. B* **766**, 225 (2017).
30. G. G. L. Nashed, W. El Hanafy, S. D. Odintsov and V. K. Oikonomou, Thermodynamical correspondence of $f(R)$ gravity in the Jordan and Einstein frames, *Int. J. Mod. Phys. D* **29**, p. 2050090 (2020).
31. T. P. Sotiriou and V. Faraoni, f(R) Theories Of Gravity, *Rev. Mod. Phys.* **82**, 451 (2010).
32. C. P. L. Berry and J. R. Gair, Linearized f(R) gravity: Gravitational radiation and Solar System tests, *Physical Review D - Particles, Fields, Gravitation and Cosmology* **83** (2011).
33. H.-J. Schmidt, The newtonian limit of fourth-order gravity, *Astronomische Nachrichten* **307**, 339.
34. P. Teyssandier, New solutions in linearized r+r2 gravity, *Astronomische Nachrichten* **311**, 209.
35. G. J. Olmo, The gravity lagrangian according to solar system experiments, *Phys. Rev. Lett.* **95**, p. 261102 (Dec 2005).
36. T. Koivisto, Covariant conservation of energy momentum in modified gravities, *Class. Quant. Grav.* **23**, 4289 (2006).
37. T. Futamase, Strong-field point-particle limit and the equations of motion in the binary pulsar, *Phys. Rev. D* **36**, 321 (Jul 1987).
38. Y. Itoh, T. Futamase and H. Asada, Equation of motion for relativistic compact binaries with the strong field point particle limit: Formulation, the first post-newtonian order, and multipole terms, *Phys. Rev. D* **62**, p. 064002 (Jul 2000).
39. D. Giulini, Laue's theorem revisited: Energy–momentum tensors, symmetries, and the habitat of globally conserved quantities, *Int. J. Geom. Meth. Mod. Phys.* **15**, p. 1850182 (2018).
40. Y. Itoh, T. Futamase and H. Asada, Equation of motion for relativistic compact binaries with the strong field point particle limit: The Second and half postNewtonian order, *Phys. Rev. D* **63**, p. 064038 (2001).
41. Y. Itoh, Equation of motion for relativistic compact binaries with the strong field point particle limit: Third postNewtonian order, *Phys. Rev. D* **69**, p. 064018 (2004).
42. S. Mirshekari and C. M. Will, Compact binary systems in scalar-tensor gravity: Equations of motion to 2.5 post-Newtonian order, *Phys. Rev. D* **87**, p. 084070 (2013).
43. S. M. Kopeikin, Post-Newtonian Lagrangian of an N-body System with Arbitrary Mass and Spin Multipoles, *Phys. Rev. D* **102**, p. 024053 (2020).

Alternatives to Λ: Torsion, generalized couplings, and scale invariance

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

C. M. J. Marques

Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa,
2829-516 Caparica, Portugal, and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

C. B. D. Fernandes, J. S. J. S. Oliveira, D. A. R. Pinheiro and B. A. R. Rocha

Faculdade de Ciências, Universidade do Porto,
Rua do Campo Alegre, 4150-007 Porto, Portugal, and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

We present a comparative analysis of current observational constraints on three recently discussed alternative models for explaining the low-redshift acceleration of the universe: the so-called steady-state torsion model, the generalized coupling model, and the scale invariant model by Maeder (an example of a broader class which we also briefly study). These are compared to the traditional parameterization of Chevallier, Polarski and Linder. Each of the candidate models is studied under two different assumptions: as genuine alternatives to ΛCDM (where a new degree of freedom would be expected to explain the recent acceleration of the universe without any cosmological constant) and as parametric extensions of ΛCDM (where both a cosmological constant and the new mechanism can coexist, and the relative contributions of both are determined by the data). Our comparative analysis suggests that, from a phenomenological point of view, all such models neatly divide into two classes, with different observational consequences.

Keywords: Cosmology; Dark energy; Torsion; Generalized couplings; Scale invariance.

1. Introduction

The observational evidence for the acceleration of the universe shows that our canonical theories of cosmology and particle physics are at least incomplete, and possibly incorrect. Is dark energy a cosmological constant (i.e. vacuum energy)? If the answer is yes, it is ten to some large power times smaller than our Quantum Field Theory based expectations. If the answer is no, then the Einstein Equivalence Principle must be violated. Either way, new physics is out there, waiting to be discovered; we must search for, identify and characterize this new physics. The CosmoESPRESSO team uses the universe as a laboratory to characterize, with precision spectroscopy and other observational, computational and theoretical tools, the behaviour of the

gravitational interaction, with the goal of determining what makes the universe accelerate. In what follows we highlight recent contributions of the CosmoESPRESSO team to this fundamental quest.

The search for the physical mechanism underlying the observed low-redshift acceleration of the universe is the most compelling goal of modern fundamental cosmology, and several theoretical possibilities beyond a cosmological constant can be envisaged in principle, each with its specific observational consequences.

Our goal here is to present a comparative study of the observational constraints on three classes of alternative models: the so-called steady-state torsion model[1] of Kranas *et al.*, the generalized coupling model[2] of Feng and Carloni, and the scale invariant model[3] of Maeder; the latter is an example of a broader class of models[4,5] first proposed by Canuto *et al.*, which we also briefly study. As a benchmark we use the traditional phenomenological parameterization of Chevallier, Polarski and Linder (henceforth CPL).[6,7] All models have common parameters (specifically, the matter density parameter, Ω_m) but also some specific ones, and a comparative analysis using a common data set is therefore interesting.

We take three models at face value and phenomenologically constrain them through a standard likelihood analysis using low-redshift background cosmology data. Specifically, we use the recent Pantheon dataset,[8] including its covariance matrix. We also use a compilation of 38 Hubble parameter measurements.[9] Occasionally we will also use a Planck prior[10] on the matter density, $\Omega_m = 0.0315 \pm 0.007$. The value of the he Hubble constant is always marginalized analytically, following the procedure detailed in Ref. 11. The analysis is done on a grid (since we are only dealing with background cosmology, there is no computational need for a full MCMC analysis), and we have explicitly verified that the grid sizes that have been used are sufficiently large for the results presented in the following sections not to be affected by these sizes. Moreover, the following section will also present an explicit validation test of our code for the supernova data.We will work in units where the speed of light is set to $c = 1$.

2. Preamble: The CPL parameterization

In the CPL parameterization the dark energy equation of state parameter is assumed to have the form[6,7]

$$w(z) = \frac{p(z)}{\rho(z)} = w_0 + w_a \frac{z}{1+z}\,, \tag{1}$$

where w_0 is its present value while w_a quantifies its possible evolution. This is manifestly phenomenological: it is not intended to mimic a particular dark energy model, but aims to describe generic departures from the ΛCDM behaviour (which corresponds to $w_0 = -1$ and $w_a = 0$). In principle it allows for both canonical and phantom fields, since there is no restriction on the two model parameters, at least on purely mathematical grounds.

We assume a flat Friedmann-Lemaître-Robertson-Walker model, in which case the Friedmann equation has the form

$$\frac{H^2(z)}{H_0^2} = \Omega_m(1+z)^3 + (1-\Omega_m)(1+z)^{3(1+w_0+w_a)} \exp\left(-\frac{3w_a z}{1+z}\right), \qquad (2)$$

where the matter parameter is $\Omega_m \equiv \kappa\rho_0/3H_0^2$ and $\kappa = 8\pi G$. This can now be constrained using the aforementioned data.

The case of a constant equation of state parameter (i.e. $w_a = 0$), for the case of the supernova dataset, have been used in Ref. 12 as a validation test of our analysis code, against the results of Ref. 8. In this case the one-sigma constraints on the two model parameters from the combined data sets are

$$\Omega_m = 0.27 \pm 0.02 \qquad (3)$$

$$w_0 = -0.92 \pm 0.06\,, \qquad (4)$$

which are compatible with ΛCDM.

For the full three-parameter CPL model, the one-sigma constraints on the three model parameters from the combined data sets are

$$\Omega_m = 0.26^{+0.03}_{-0.05} \qquad (5)$$

$$w_0 = -0.92^{+0.09}_{-0.08} \qquad (6)$$

$$w_a = 0.86^{+0.14}_{-0.24}; \qquad (7)$$

the reduced chi-square at the best fit is $\chi^2_\nu \sim 0.9$, so the model is slightly overfitting the data (a behaviour which is mainly driven by the Hubble parameter data). The first two of these constraints are compatible with the values for the w_0CDM analysis (with naturally larger uncertainties), but there is a clear preference for a positive slope $w_a > 0$. However there are strong degeneracies between the parameters, and the constraints do depend on the choice of priors. In the above we used the uniform prior on the matter density $\Omega_m = [0.05, 0.5]$, the choice being motivated by the aforementioned validation of our code. As an illustration of the sensitivity of our results to this choice, if instead one uses the narrower uniform prior $\Omega_m = [0.15, 0.45]$, one find

$$\Omega_m = 0.26^{+0.03}_{-0.05} \qquad (8)$$

$$w_0 = -0.92^{+0.07}_{-0.08} \qquad (9)$$

$$w_a = 0.74^{+0.21}_{-0.48}; \qquad (10)$$

in other words, there is no impact on the matter density and w_0, but there is a significant impact on w_a. Breaking these degeneracies requires additional data, for example from cosmic microwave background observations. In any case, our purpose here is to set up a benchmark for the constraining power of these data sets, against which to compare the constraints on the alternative models to be discussed in what follows.

3. Steady-state torsion

A possible extension of General Relativity consists in allowing for the presence of spacetime torsion. In such theories there is a further degree of freedom (in addition to the usual metric), which also gravitates. Mathematically, the torsion tensor is defined as the antisymmetric part of the affine connection; the symmetric part of the connection are the usual Christoffel symbols. Physically, this defines relation between the intrinsic angular momentum (i.e., the spin) of matter with the geometric properties of the underlying spacetime. The only non-trivial contraction of the torsion tensor is a torsion vector, and the general field equations including torsion are known as the Einstein-Cartan equations. Nominally the Einstein equations retain the usual form, but the presence of torsion implies that the Ricci tensor and the energy-momentum tensor are not symmetric. The Cartan equations relate the torsion tensor to the spin tensor, and similarly for the torsion and spin vectors.

The form of the underlying torsion tensor can be chosen such that the homogeneity and isotropy of FLRW universes is preserved,[13] and in this case the remaining degree of freedom is a scalar function ϕ which must depend only on time (a spatial dependence would violate the homogeneity assumption), but is otherwise arbitrary. Making the standard assumption of treating the metric and the torsion as independent objects and furhter assuming a flat universe, one finds the following Friedmann, Raychaudhuri and continuity equations[1]

$$H^2 = \frac{1}{3}\kappa\rho\frac{1}{3}\Lambda - 4\phi^2 - 4H\phi \tag{11}$$

$$\frac{\ddot{a}}{a} = -\frac{\kappa}{6}(\rho + 3p) + \frac{1}{3}\Lambda - 2\dot{\phi} - 2H\phi \tag{12}$$

$$\dot{\rho} = -3H\left(1 + 2\frac{\phi}{H}\right)(\rho + p) + 4\phi\left(\rho + \frac{\Lambda}{\kappa}\right) . \tag{13}$$

Here the dot denotes a derivative with respect to physical time, $H = \dot{a}/a$ is the Hubble parameter, and ρ and p are the density and pressure. In what follows we will assume barotropic fluids with a constant equation of state $p = w\rho$. It has been recently suggested that such universes may undergo accelerating phases.[1] We can conveniently define a torsion contribution

$$\Omega_\phi = -4\left(\frac{\phi_0}{H_0}\right)\left[1 + \frac{\phi_0}{H_0}\right] . \tag{14}$$

In Ref. 14 these models were constrained under the so-called steady-state torsion assumption of a constant fractional contribution of torsion to the volume expansion, that is $\phi/H = \lambda = const.$.

It is easy to find, in agreement with other recent works, that models without a cosmological constant (where torsion itself would be expected to yield the current acceleration of the universe) are strongly disfavoured by the data. Indeed, in this case, for which the matter density would be given by $\Omega_m = (1 + 2\lambda)^2$, the best fit parameters would have a reduced chi-square of at least 2.7 for the datasets under consideration.

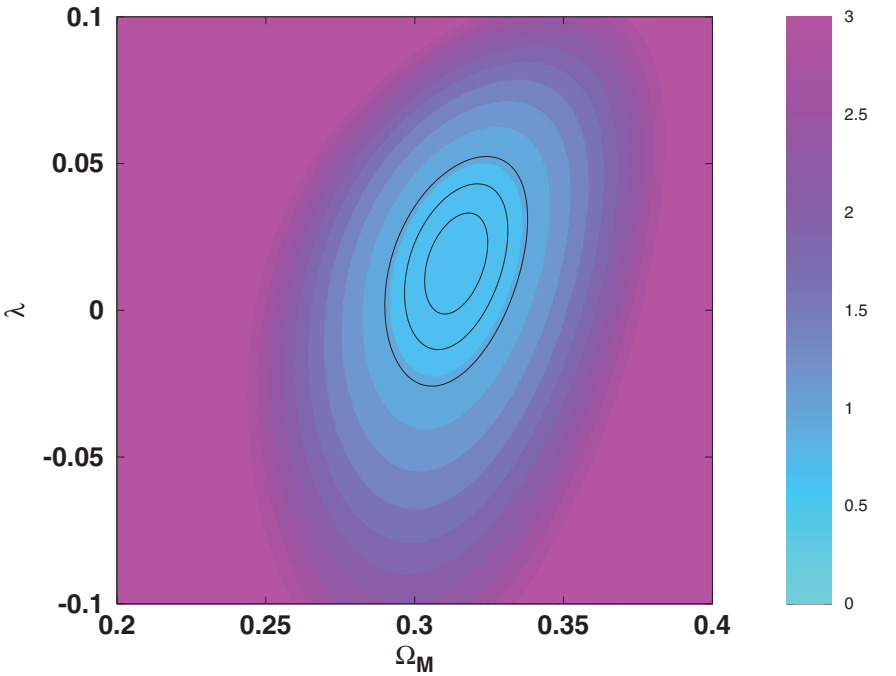


Fig. 1. Constraints on the λ–Ω_m parameter space for $w = 0$. The black lines represent the one, two and three sigma confidence levels, and the colormap depicts the reduced chi-square of the fit, with points with $\chi^2_\nu > 3$ shown in purple. Similar constraints can be found in Figure 2 of Ref. 14.

However, one can also treat these models as one-parameter extensions of ΛCDM, whereby one can constrain the relative contributions of the cosmological constant and of torsion. In this case, if one assumes that that matter has the standard equation of state, $w = 0$. As in the previous subsection will separately consider the cases without and with the aforementioned Planck prior on the matter density. Without the Planck prior, we find the following one-sigma posterior likelihoods for the two free parameters

$$\lambda_{w=0} = -0.07^{+0.05}_{-0.04} \tag{15}$$

$$\Omega_{m,w=0} = 0.18^{+0.06}_{-0.03}\,; \tag{16}$$

there is a clear degeneracy between the two parameters, and the preferred value of the matter density is lower. The inclusion of the Planck prior breaks the degeneracy and significantly improves the constraints, as shown in Fig. 1; the one-sigma posterior likelihood for the torsion parameter becomes

$$\lambda_{(w=0,Planck)} = 0.02^{+0.01}_{-0.02}\,, \tag{17}$$

which is consistent with the null result at just over one sigma.

Allowing for a non-zero (but still constant) equation of state, there is a weak degeneracy between w and the other model parameters, so although the constraints become weaker (as they must), both parameters are still well constrained by the data, provided the Planck prior is included. In this case we have

$$\lambda_{(w\neq 0,Planck)} = -0.01 \pm 0.02 \tag{18}$$

$$w_{Planck} = -0.05 \pm 0.03\,; \tag{19}$$

compared to the $w = 0$ case the best-fit value has changed sign, and the constraint is now consistent with the null result at one sigma.

Overall, we find no statistically significant preference for the presence of torsion. By itself it can't be responsible for the acceleration of the universe, and even if taken as an extension of the canonical ΛCDM paradigm the overall contribution to the Universe's energy budget is constrained to be no larger than a few percent. We also note that our constraints should be seen as conservative: an analysis including a full treatment of the cosmic microwave background should lead to stronger constraints.

4. Generalized couplings

The precise nature of the coupling between matter and the metric in the Einstein equations is at questionable assumption of the theory. One may therefore explore the possibility that this coupling is nontrivial. On such example is the Feng and Carloni's generalized coupling model,[2] which is equivalent to General Relativity in vacuum, but still allows for a different behaviour within a matter distribution.

In this case the Friedmann and Raychaudhuri equations, assuming a flat universe, can be written[2]

$$3qH^2 = \frac{256\kappa(1-pq)^3(q\rho+1)^2}{[4+q(\rho-3p)]^4} + q\Lambda - \kappa \tag{20}$$

$$6q(\dot{H}+H^2) = \frac{256\kappa(1-pq)^3(q\rho+1)[2-q(\rho+3p)]}{[4+q(\rho-3p)]^4} + 2(q\Lambda-\kappa)\,, \tag{21}$$

where q is a model-specific parameter defined as $q = \kappa/\lambda$ (where λ, not to be confused with the analogous torsion parameter, is interpreted as being akin to the vacuum energy density generated by matter fields) and p is the pressure of a fluid that is assumed to be barotropic, with an equation of state $p = w\rho$, where w is a constant equation of state parameter. The corresponding continuity equation takes the form

$$\dot{\rho} = -\frac{3H\rho(w+1)[q^2\rho^2 w(3w-1)+q\rho(1-7w)+4]}{q^2\rho^2 w(3w-1)-q\rho(3w^2+13w+2)+4}\,. \tag{22}$$

Note the model is effectively a bimetric theory.[2] In what follows we take the model as a phenomenological one and treat q (or a dimensionless version thereof) as a free parameter to be constrained by the data. It is convenient to define the dimensionless parameter $Q = q\rho_0$, where ρ_0 is the present-day critical density.

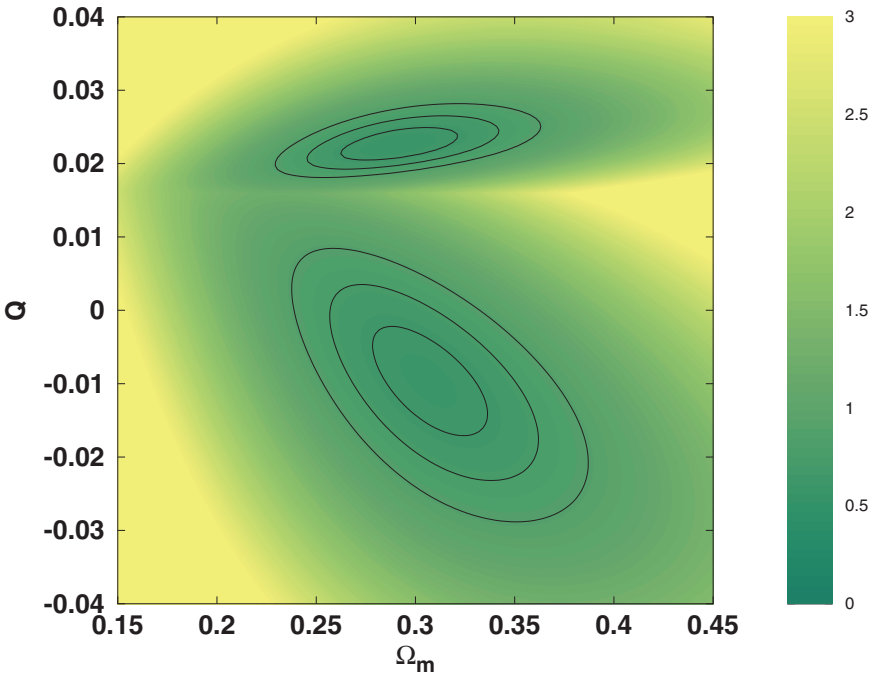


Fig. 2. Two-dimensional constraints on the $w = 0$ generalized coupling model. The $\Delta\chi^2 = 2.3$, $\Delta\chi^2 = 6.17$ and $\Delta\chi^2 = 11.8$ confidence levels are shown in black lines, and the color map depicts the reduced chi-square at each point in the parameter space, with points with $\chi^2_\nu > 3$ shown in yellow. Similar constraints can be found in Figure 3 of Ref. 12.

Since the model effectively has two types of vacuum energy, viz. the one generated by matter fields as well as the usual cosmological constant, one may wonder if the former is sufficient to yield an accelerating universe without invoking the latter. However, it is again simple to show that this can't be the case,[12] since in that case the minimum density would be $\Omega_m \sim 0.86$; clearly such high matter density universes would be incompatible with observations. Thus in what follows we treat this model as a phenomenological extension of ΛCDM, with the vacuum energy density of matter fields, Q, being an additional model parameter which we now constrain.

In the simpler case where the matter equation of state parameter has the standard value, $w = 0$, and agnostically allowing both positive and negative values of the model parameter Q, we find that while non-zero values of Q are preferred, the standard value is not significantly excluded, as shown in Fig. 2. We note the existence of two branches of the solution, one with $Q > 0$ and the other with $Q < 0$, with the former branch being slightly preferred. If we restrict the analysis to the

range $Q \leq 0$, the one-sigma constraints on the two model parameters are

$$\Omega_{m-} = 0.31 \pm 0.02 \tag{23}$$

$$Q_{-} = -0.010 \pm 0.006\,; \tag{24}$$

conversely, if we restrict the analysis to the range $Q \geq 0$ we find

$$\Omega_{m+} = 0.29 \pm 0.02 \tag{25}$$

$$Q_{+} = 0.023 \pm 0.003\,. \tag{26}$$

In all cases the reduced chi-square at the best fit is $\chi^2_\nu \sim 0.6$, so the model is clearly overfitting the data. All in all, there is no strong evidence for a non-zero Q.

In the general case, allowing the dark energy equation of state parameter w to become a further free parameter, one obtains the following one-sigma constraints on the parameters

$$\Omega_m = 0.29^{-0.09}_{+0.07} \tag{27}$$

$$Q = -0.018^{+0.005}_{-0.004} \tag{28}$$

$$w = -0.06^{+0.17}_{-0.08}\,. \tag{29}$$

The constraints on the matter density are now significantly weaker, but the two Q branches of the solution are still manifest, as are the degeneracies between the model parameters, as can be seen in Ref. 12. In this case the negative branch is also the preferred one. However, we should also point out that the matter equation of state parameter is already more tightly constrained than this (and this comment also applies to the above constraints on the torsion model). Recent analyses[15,16] constrain it, conservatively, to $|w| < 0.003$. Using this as a Gaussian prior and repeating the analysis, we recover the constraints on Q and Ω_m reported above for the $w = 0$ case, while the posterior for w itself simply recovers the prior.

5. Scale invariance: The specific Maeder model

Maeder's proposed scale invariant model[3] is a specific case of the scale-covariant theory of Canuto *et al.*[4,5] It is well known that the effects of scale invariance disappear upon the presence of matter; the assumption underlying scale invariant models is that at large (i.e., cosmological) scales empty space should still be scale invariant. This again leads to a bimetric theory, with a function λ (not to be confused with the parameters introduced in previous sections) playing the role of a scale transformation factor relating the ordinary matter frame to another frame which one assumes to still be scale invariant.

In this case, and with the further assumption of a flat homogeneous and isotropic universe, the Friedmann, Raychaudhuri, and continuity equations are[4,5]

$$\left(\frac{\dot{a}}{a}+\frac{\dot{\lambda}}{\lambda}\right)^2+\frac{k}{a^2}=\frac{1}{3}(\kappa\rho+\Lambda\lambda^2) \tag{30}$$

$$\frac{\ddot{a}}{a}+\frac{\ddot{\lambda}}{\lambda}+\frac{\dot{\lambda}}{\lambda}\frac{\dot{a}}{a}-\frac{\dot{\lambda}^2}{\lambda^2}=-\frac{\kappa}{6}(\rho+3p-2\Lambda\lambda^2) \tag{31}$$

$$\dot{\rho}+3(\rho+p)\frac{\dot{a}}{a}=-(\rho+3p)\frac{\dot{\lambda}}{\lambda}\,, \tag{32}$$

which match the standard equations if one chooses $\lambda = 1$. Note that for a homogeneous and isotropic model λ depends only on time, as does the scale factor.

The recent work of Maeder further postulates that the Minkowski metric is a solution of these Einstein equations, which leads to the following consistency conditions[3]

$$3\frac{\dot{\lambda}^2}{\lambda^2}=\Lambda\lambda^2 \tag{33}$$

$$2\frac{\ddot{\lambda}}{\lambda}-\frac{\dot{\lambda}^2}{\lambda^2}=\Lambda\lambda^2\,, \tag{34}$$

and further imply that

$$\lambda(t)=\sqrt{\frac{3}{\Lambda}}\,\frac{1}{t}\,. \tag{35}$$

We are again using $c = 1$, and constant equations of state, $p = w\rho$. Together with the solution for λ, the continuity equation yields

$$\rho\propto(1+z)^{3(1+w)}t^{1+3w}. \tag{36}$$

For a cosmological constant equation of state ($w = -1$) this becomes $\rho \propto t^{-2}$; in other words, this is effectively a model with a time-dependent cosmological constant, but no parametric ΛCDM limit. The author claims,[3] from a simple qualitative comparison, that with the choice $\Omega_m = 0.3$ the model is in good agreement with Hubble parameter data. In Ref. 12 this claim was assessed with a more thorough statistical analysis, and we summarize the results here.

With the aforementioned assumptions, the Friedmann equation for the Maeder model can be written

$$E^2(z,x)=\Omega_m(1+z)^{3(1+w)}x^{1+3w}+\frac{\Omega_\lambda}{x}E(z,x)\,, \tag{37}$$

where we have defined an effective parameter

$$\Omega_\lambda=\frac{2}{t_0H_0}\,, \tag{38}$$

which effectively quantifies the present age of the universe (in dimensionless units) and for convenience also introduced a dimensionless time $x = t/t_0$, with t_0 being

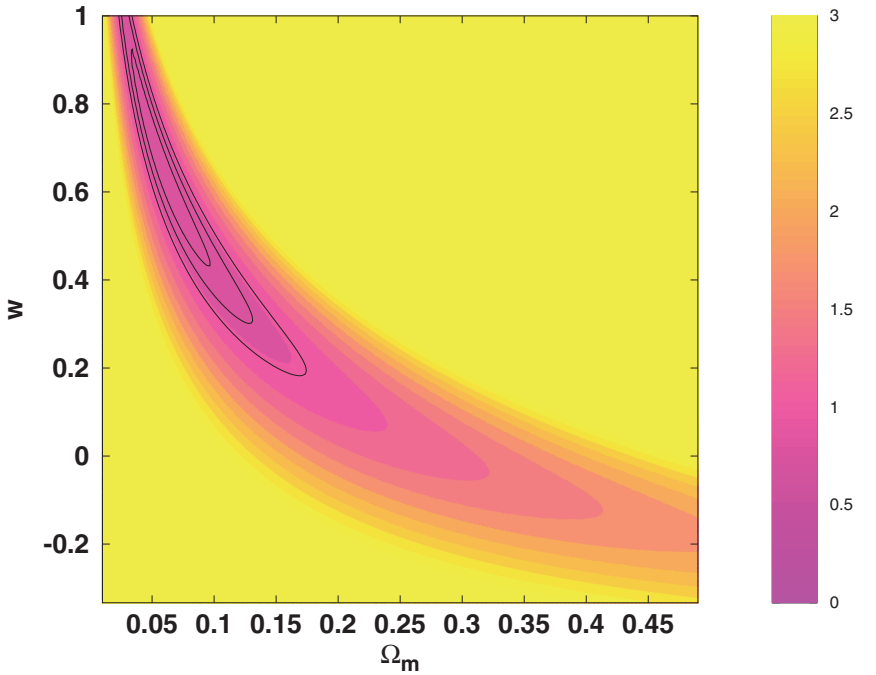


Fig. 3. Two-dimensional constraints on the Maeder model, with w as a free parameter. The black lines represent the one, two and three sigma confidence levels, and the color map depicts the reduced chi-square at each point in the parameter space, with points with $\chi^2_\nu > 3$ shown in yellow. Similar constraints can be found in Figure 6 of Ref. 12.

the current age of the universe. With these definitions the Friedmann equation can be re-written in the simpler form

$$E(z,x) = \frac{\Omega_\lambda}{2x}\left[1+\sqrt{1+M(z,x)}\right] \tag{39}$$

$$M(z,x) = \frac{4\Omega_m}{\Omega_\lambda^2}(1+z)^{3(1+w)}x^{3(1+w)}\,, \tag{40}$$

with the relation between redshift and (dimensionless) time being given by

$$\frac{dx}{dz} = -\frac{x}{1+z}\times\frac{1}{1+\sqrt{1+M(z,x)}} \tag{41}$$

and the initial condition $x = 1$ at $z = 0$.

In the $w = 0$ case we can write $\Omega_\lambda = 1 - \Omega_m$. and the one-sigma posterior constraint in the matter density is

$$\Omega_m = 0.26 \pm 0.02\,, \quad \chi^2_\nu = 1.3\,; \tag{42}$$

the inclusion of curvature as an additional parameter[12] slightly increases the preferred matter density but provides an equally poor fit. On the other hand, allowing

w as a further free parameter, one obtains the result shown in in Fig. 3. In this case the one sigma constraints are

$$\Omega_m = 0.06 \pm 0.02 \tag{43}$$

$$w = 0.60^{+0.16}_{-0.15}\,; \tag{44}$$

again the inclusion of curvature does not significantly change this.[12] In both cases the reduced chi-square is now $\chi^2_\nu = 0.8$, so the model is now slightly overfitting the data. Clearly there is a strong degeneracy between the matter density and the equation of state parameter (which are anticorrelated), and the best fit values of both parameters are very far from the standard ΛCDM ones.

6. Scale invariance: the general model

The previous section shows that the Maeder model is ruled out. One may therefore ask whether this conclusion extends to the more general model of Canute *et al.*, also introduced in the previous section. Here we present a very preliminary analysis of this issue. We will assume a generic power-law behaviour, $\lambda(t) \propto t^p$, choosing $\lambda_0 = 1$, and further assuming flat models[a]. This choice of λ also ensures that ΛCDM is recovered for $p = 0$.

Note that in the Maeder model there is no explicit cosmological constant Λ. In the general case it is still there, so we may again expect two classes of solutions. One has the usual Λ providing the acceleration, with the λ field providing a further contribution; in other words, this will be an extension of ΛCDM. The other has $\Lambda = 0$, meaning that the model will not have a ΛCDM limit, and the question is then whether the field λ can provide an alternative to acceleration in that case.

In this case the continuity equation gives

$$\rho \propto (1+z)^{3(1+w)} t^{-p(1+3w)}\,, \tag{45}$$

while the Friedmann equation gives

$$\left(E(z,x) + \frac{p}{2x}\Omega_\lambda\right)^2 = \Omega_m (1+z)^{3(1+w)} x^{-p(1+3w)} + \Omega_\Lambda x^{2p}\,. \tag{46}$$

For $p = -1$ the Maeder model is recovered with the further assumption that

$$\Omega_\Lambda = \frac{1}{4}\Omega_\lambda^2 = \frac{1}{(t_0 H_0)^2}\,. \tag{47}$$

In the general case the Friedmann equation can be re-written

$$E(z,x) = \frac{\Omega_\lambda}{2x}\left[-p + \sqrt{N(z,x)}\right] \tag{48}$$

$$N(z,x) = \frac{4}{\Omega_\lambda^2}\left[\Omega_m (1+z)^{3(1+w)} x^{2-p(1+3w)} + \Omega_\Lambda x^{2(1+p)}\right]\,, \tag{49}$$

[a] We will report on a more detailed analysis, relaxing some of these assumptions and exploring various scenarios, elsewhere.

and the relation between redshift and (dimensionless) time is now given by

$$\frac{dx}{dz} = -\frac{x}{1+z} \times \frac{1}{\sqrt{N(z,x) - p}}\,, \tag{50}$$

with the initial condition still being $x = 1$ at $z = 0$. One can easily check that the Maeder model equations are recovered in the appropriate limit. Note that Ω_λ is a dimensionless measure of the current age of the universe, and it must therefore be a positive quantity.

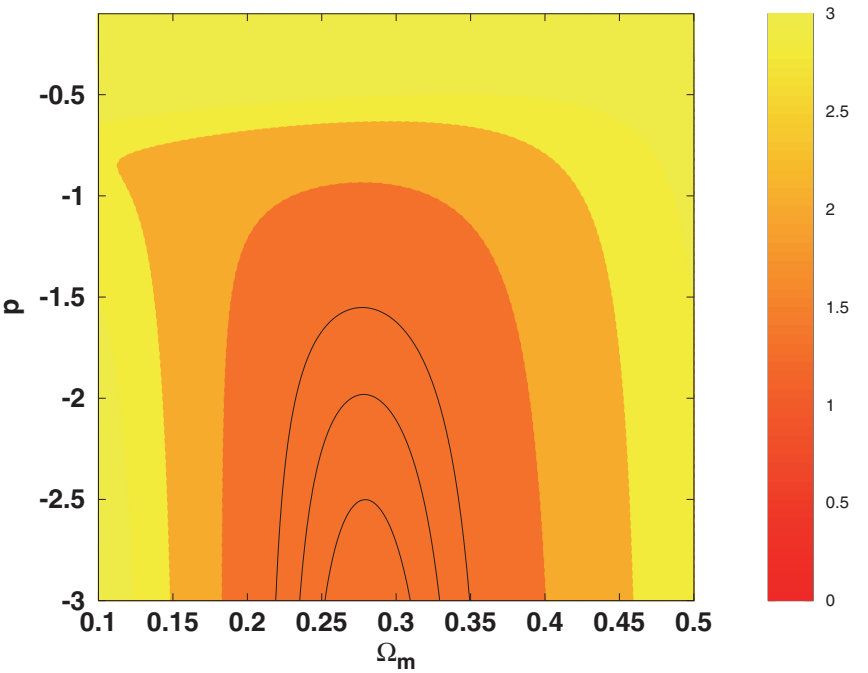


Fig. 4. Two-dimensional constraints on the Canuto *et al.* model, with $w = 0$ and $\Omega_\Lambda = 0$. The black lines represent the one, two and three sigma confidence levels, and the color map depicts the reduced chi-square at each point in the parameter space, with points with $\chi^2_\nu > 3$ shown in yellow.

Several different scenarios can now be considered. In what follows we only consider the simplest one. If we assume that $\Omega_\Lambda = 0$ and ordinary matter with an $w = 0$ equation of state we have two free parameters (Ω_m, p) and

$$-p\Omega_\lambda = 2(1 - \sqrt{\Omega_m})\,; \tag{51}$$

thus $p = 0$ corresponds to an $\Omega_m = 1$ universe, In this case we again find that this model does not provide a good fit to the data, cf. Fig. 4: while a matter density of around $\Omega_m \sim 0.28$ is preferred, the reduced chi-square of the best fit is quite poor,

being always larger than $\chi^2_\nu = 1.25$. An exploration of the wider parameter space, allowing for a cosmological constant, a non-zero equation of state, and curvature, will be reported elsewhere.

7. Outlook

We have briefly presented a comparison of three classes of models for the low-redshift acceleration of the universe against background low-redshift cosmological observations, further using the traditional CPL phenomenological parameterization as a benchmark. Each of these models contains, in principle, an additional mechanism, in addition to the cosmological constant, that could account for the recent acceleration of the universe.

We find that the steady-state torsion and generalized couplings have similar behaviours. The specific physical mechanism therein are ruled out as unique source of acceleration, but small (e.g. percent level) deviations from ΛCDM are still allowed by the data that we have considered. Constraints can of course be tightened by including further data. The scale invariant model provides an interesting contrast. Assuming the standard equation of state parameter, $w = 0$, the best-fit value is similar to the CPL one. However, these fits have a poor (specifically, high) reduced chi-square, indicating that the model does not provide a good fit to the data. An observationally viable model of this kind would require a highly non-standard 'matter' density (including a non-standard equation of state), which would conflict with other cosmological datasets. In passing, we also mention that another interesting phenomenological class is that of energy-momentum-powered models recently constrained in Ref. 17 and also discussed elsewhere in these proceedings.

Overall, we therefore conclude that ΛCDM is a remarkably robust paradigm. While it is clearly a phenomenological approximation to a still unknown more fundamental model, it is clearly a good one, and any plausible alternative model must be able to closely reproduce its behaviour in a broad range of cosmological settings.

Acknowledgments

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017.

References

1. D. Kranas, C. Tsagas, J. Barrow and D. Iosifidis, Friedmann-like universes with torsion, *Eur. Phys. J. C* **79**, p. 341 (2019).
2. J. C. Feng and S. Carloni, New class of generalized coupling theories, *Phys. Rev. D* **101**, p. 064002 (2020).

3. A. Maeder, An alternative to the ΛCDM model: The case of scale invariance, *Astrophys. J.* **834**, p. 194 (2017).
4. V. Canuto, S. Hsieh and P. Adams, Scale-Covariant Theory of Gravitation and Astrophysical Applications, *Phys. Rev. Lett.* **39**, 429 (1977).
5. V. Canuto, P. Adams, S. Hsieh and E. Tsiang, Scale Covariant Theory of Gravitation and Astrophysical Applications, *Phys. Rev. D* **16**, 1643 (1977).
6. M. Chevallier and D. Polarski, Accelerating universes with scaling dark matter, *Int. J. Mod. Phys. D* **10**, 213 (2001).
7. E. V. Linder, Exploring the expansion history of the universe, *Phys. Rev. Lett.* **90**, p. 091301 (2003).
8. A. G. Riess *et al.*, Type Ia Supernova Distances at Redshift >1.5 from the Hubble Space Telescope Multi-cycle Treasury Programs: The Early Expansion Rate, *Astrophys. J.* **853**, p. 126 (2018).
9. O. Farooq, F. R. Madiyar, S. Crandall and B. Ratra, Hubble Parameter Measurement Constraints on the Redshift of the Deceleration-acceleration Transition, Dynamical Dark Energy, and Space Curvature, *Astrophys. J.* **835**, p. 26 (2017).
10. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)].
11. F. K. Anagnostopoulos and S. Basilakos, Constraining the dark energy models with $H(z)$ data: An approach independent of H_0, *Phys. Rev.* **D97**, p. 063503 (2018).
12. C. B. D. Fernandes, C. J. A. P. Martins and B. A. R. Rocha, Constraining alternatives to a cosmological constant: Generalized couplings and scale invariance, *Phys. Dark Univ.* **31**, p. 100761 (2021).
13. M. Tsamparlis, Methods for Deriving Solutions in Generalized Theories of Gravitation: The Einstein-Cartan Theory, *Phys. Rev.* **D24**, 1451 (1981).
14. C. Marques and C. Martins, Low-redshift constraints on homogeneous and isotropic universes with torsion, *Phys. Dark Univ.* **27**, p. 100416 (2020).
15. D. B. Thomas, M. Kopp and C. Skordis, Constraining the Properties of Dark Matter with Observations of the Cosmic Microwave Background, *Astrophys. J.* **830**, p. 155 (2016).
16. I. Tutusaus, B. Lamine, A. Blanchard, A. Dupays, Y. Rousset and Y. Zolnierowski, Dark sectors of the Universe: A Euclid survey approach, *Phys. Rev. D* **94**, p. 123515 (2016).
17. M. C. F. Faria, C. J. A. P. Martins, F. Chiti and B. S. A. Silva, Low redshift constraints on energy-momentum-powered gravity models, *Astron. Astrophys.* **625**, p. A127 (2019).

Model-independent test of scalar-tensor gravity theory by reconstructing scalar mode of GW170817

Yuya Gushima* and Kazuhiro Hayama†

Department of Applied Physics, Fukuoka University, Fukuoka, 814-0180, Japan

** E-mail: sd211004@cis.fukuoka-u.ac.jp*

† E-mail: hayama@fukuoka-u.ac.jp

Einstein's general relativity predicts that a gravitational wave is allowed to have two polarizations called tensor-modes: plus and cross modes. On the other hand, the general metric theory of gravity predicts that a gravitational wave is allowed to have up to six polarizations: two scalar and two vector modes in addition to the tensor-modes. In case the number of laser-interferometric gravitational wave telescopes is larger than the one of the polarizations the gravitational waves have, all the polarizations can be reconstructed separately. Since it depends on theories of gravity which polarizations the gravitational waves have, the investigation of polarizations is important for the test of theories of gravity. In this paper, in order to test the scalar-tensor gravity theory, one of important alternative theories of gravity, we search for the scalar-mode of GW170817 observed by LIGO Livingstone, Hanford and Virgo without prior information about any tensor-scalar gravity theories. As a result, we found the maximum SNR of the scalar-mode of GW170817 was 2.77, the p-value was 0.01, and the band-limited root sum square of h was $1.55 \times 10^{-21}[1/\sqrt{Hz}]$ with the time window of 2[s] and frequency window of 60∼120[Hz].

Keywords: Gravitational wave; polarization; scalar-mode; test of gravity theory; GW170817.

1. Introduction

Since the gravitational wave (GW) was observed in 2015, many gravitational waves have been detected.[1] The observation of GWs will provide an opportunity to unveil the mystery of the early universe,[2] black holes[1] and dark matter. Einstein's general relativity predicts that a GW is allowed to have two polarization modes, plus mode and cross mode, which are called tensor modes. On the other hand, the general metric theory of gravity predicts that a gravitational wave is allowed to have up to six polarizations, two scalar and two vector modes in addition to the tensor modes.[3,4] This suggests, if either the scalar or the vector modes are found, it is necessary to modify the general relativity.

E. Hubble observed that the universe was expanding,[5] and George Gamow proposed the big bang cosmology.[6] However, the proposed theory could not explain the flatness problem and the horizon problem. Inflation universe theories were proposed to solve these problems.[7] Direct evidence of inflation has not been found yet, but if inflation theories are correct, a scalar field may be necessary for inflation to occur.

Instead of the general relativity, the general metric theory of gravity has been proposed, and the theory of gravity has been tested using the GWs detected so far.[8,9] Among the proposed theories of gravity, the scalar-tensor gravity theory predicts the gravitation is given through not only the tensor field but also by the scalar field. It is possible to reconstruct the tensor and scalar modes from GW observations when the GW is observed by three or more laser-interferometric gravitational wave telescopes and the direction to the GW is known.[10] In this paper, in order to test the scalar-tensor gravity theory, we search for a scalar mode of GW170817[11] observed by LIGO-Livingston, LIGO-Hanford[12] and Virgo.[13] The direction to GW170817 is known by the multi-messenger observation.[14] Normally, when exploring GWs, the correlation between the predicted waveform and the observed data is examined, which is called the matched filter method. For this purpose, we must assume a specific theory model. In this paper, we take a different way to test the gravity theories. We reconstruct the polarizations without any prior information about waveform. Since our method of reconstruction does not rely on any specific gravity theory, it is possible to test theory of gravity with high generality. In Sec. 2, we describe how reconstruct each polarization. In Sec. 3, we show the analysis result. Section 4 is devoted to the summary of this paper.

2. Reconstruction of polarization modes

2.1. *Polarization modes of a gravitational wave*

The GW propagating in the z direction, the bases of the six polarizations are defined by following tensors in which time, x, y, z component from the left to the right or the top to the bottom,

$$
\begin{aligned}
e^{+}_{ij} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad e^{\times}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
e^{b}_{ij} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad e^{l}_{ij} = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \\
e^{x}_{ij} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad e^{y}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},
\end{aligned}
\tag{1}
$$

where $+, \times, b, l, x$ and y denote plus, cross, breathing, longitudinal, vector-x, vector-y modes. Figure 1 shows that how each polarization affects test masses placed in

a circle. Using Eq. (1) and amplitude of each polarization $h_\alpha(\alpha = +, \times, b, l, x, y)$, the GW h_{ij} is given by

$$h_{ij} = h_+ e^+_{ij} + h_\times e^\times_{ij} + h_b e^b_{ij} + h_l e^l_{ij} + h_x e^x_{ij} + h_y e^y_{ij}. \tag{2}$$

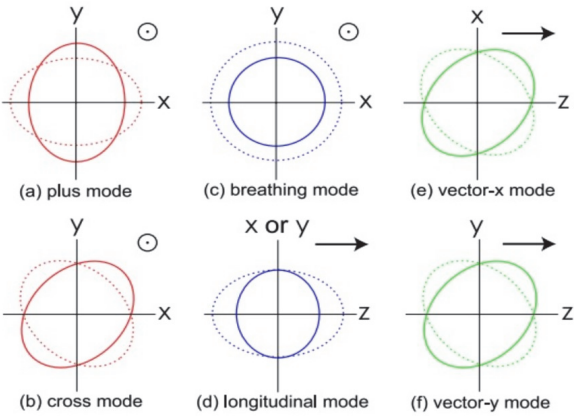


Fig. 1. Six gravitational wave polarizations in a general metric theory of gravity.[15]

2.2. *Antenna pattern function*

Suppose that a normal orthogonal system $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$ in which the unit vector $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ are directed to each detector arm. The coordination systems $(\hat{\mathbf{m}}, \hat{\mathbf{n}}, \hat{\mathbf{\Omega}})$ are given by the rotation angle(θ, ϕ, ψ) (Fig. 2). $\hat{\mathbf{\Omega}}$ is unit vector directed the direction to the GW, and the angle ψ around the $\hat{\mathbf{\Omega}}$ is an arbitrary angle. Using vector $\hat{\mathbf{m}}, \hat{\mathbf{n}}, \hat{\mathbf{\Omega}}$, the

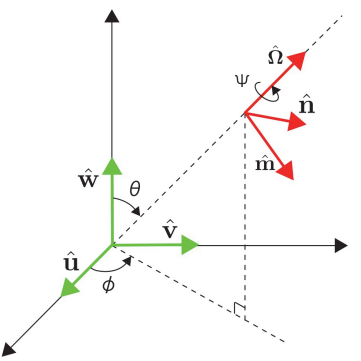


Fig. 2. Coordinate systems.[15]

polarization tensors are given by

$$\begin{aligned} \tilde{\mathbf{e}}_+ &= \hat{\mathbf{m}} \otimes \hat{\mathbf{m}} - \hat{\mathbf{n}} \otimes \hat{\mathbf{n}}, & \tilde{\mathbf{e}}_\times &= \hat{\mathbf{m}} \otimes \hat{\mathbf{n}} + \hat{\mathbf{n}} \otimes \hat{\mathbf{m}} \\ \tilde{\mathbf{e}}_b &= \hat{\mathbf{m}} \otimes \hat{\mathbf{m}} + \hat{\mathbf{n}} \otimes \hat{\mathbf{n}}, & \tilde{\mathbf{e}}_l &= \sqrt{2}\hat{\mathbf{\Omega}} \otimes \hat{\mathbf{\Omega}} \\ \tilde{\mathbf{e}}_x &= \hat{\mathbf{m}} \otimes \hat{\mathbf{\Omega}} + \hat{\mathbf{\Omega}} \otimes \hat{\mathbf{m}}, & \tilde{\mathbf{e}}_y &= \hat{\mathbf{n}} \otimes \hat{\mathbf{\Omega}} + \hat{\mathbf{\Omega}} \otimes \hat{\mathbf{n}} \end{aligned} \tag{3}$$

where the symbol $\otimes$ denotes a tensor product. The antenna pattern function $F_\alpha(\alpha = +, \times, b, l, x, y)$ of a detector is defined as

$$F_\alpha(\hat{\mathbf{\Omega}}) = \mathbf{D} : \tilde{\mathbf{e}}_\alpha(\hat{\mathbf{\Omega}}), \tag{4}$$

where the symbol : denotes contraction between tensors. $\mathbf{D}$ is a detector tensor and defined as

$$\mathbf{D} = \frac{1}{2}[\hat{\mathbf{u}} \otimes \hat{\mathbf{u}} - \hat{\mathbf{v}} \otimes \hat{\mathbf{v}}]. \tag{5}$$

The antenna pattern function depends on the position of detectors and the direction to the GW.

2.3. *Coherent network analysis*

Data from i-th detector are

$$x_i = h_i + \eta_i, \tag{6}$$

where h_i are the i-th detector response to the GW and η_i are the i-th detector noise. h_i is written as

$$h_i = F_{i+} h_+ + F_{i\times} h_\times + F_{is} h_s, \tag{7}$$

where $F_{i\alpha}(\alpha = +, \times, s)$ is the antenna pattern function of i-th detector. Here s denotes a scalar-mode. h_α are the amplitude of each polarization. As the antenna pattern function of the breathing-mode and the longitudinal-mode are degenerated, we can't decompose these scalar modes. In this paper, we suppose that the GW has two tensor modes and one scalar mode and the amplitude of vector modes is small enough to be ignored. If there are three detectors, from Eq. (6,7)

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} F_{1+} & F_{1\times} & F_{1s} \\ F_{2+} & F_{2\times} & F_{2s} \\ F_{3+} & F_{3\times} & F_{3s} \end{bmatrix} \begin{bmatrix} h_+ \\ h_\times \\ h_s \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}. \tag{8}$$

Equation (8) is written in a matrix as

$$\boldsymbol{x} = \boldsymbol{F}\boldsymbol{h} + \boldsymbol{\eta}. \tag{9}$$

$\boldsymbol{F}$ is not a square matrix and doesn't have its inverse matrix in general. But if we use Moore-Penrose pseudo-inverse matrix we can reconstruct each polarization. We consider the reconstruction without noise for simplicity. Multiplying Eq. (9) by transpose matrix $\boldsymbol{F^T}$, we get

$$\boldsymbol{F^T F h} = \boldsymbol{F^T x}. \tag{10}$$

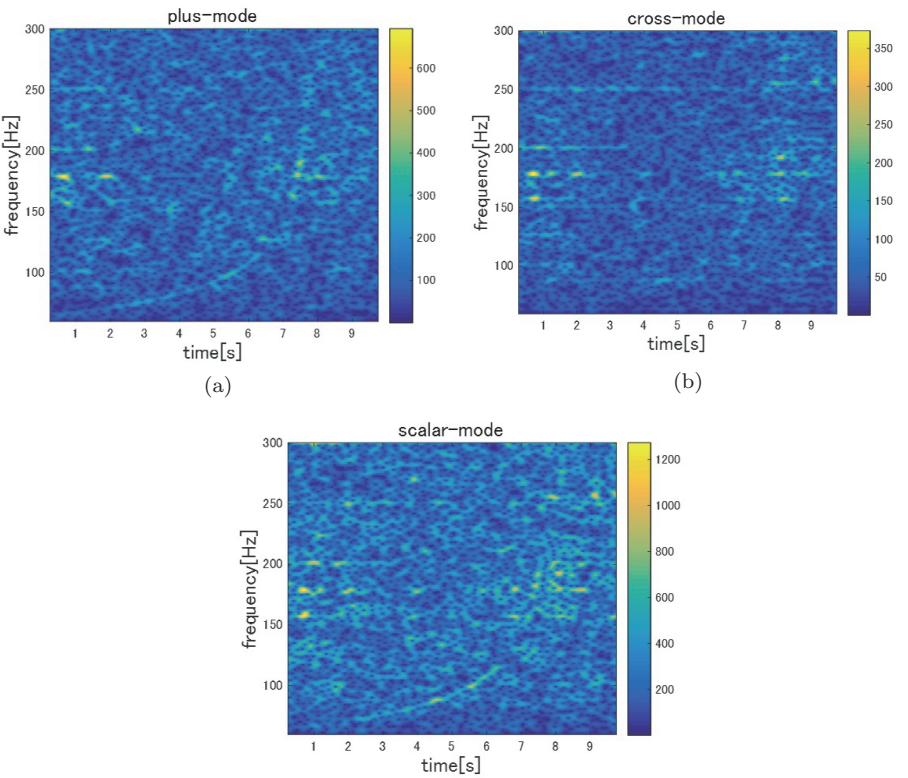
Since $\boldsymbol{F^T F}$ is a square matrix, we can find the inverse matrix. We finally get

$$\boldsymbol{h} = \boldsymbol{M^{-1} F^T x}, \qquad \boldsymbol{M} := \boldsymbol{F^T F}, \tag{11}$$

where $\boldsymbol{M^{-1} F^T}$ is the Moore-Penrose pseudo-inverse matrix.

3. Analysis method and results

Figure 3 shows spectrograms of three modes(plus, cross and scalar) reconstructed from GW170817.



(a) (b) (c)

Fig. 3. Spectrograms of the reconstructed plus-mode (a), cross-mode (b) and scalar-mode (c). The horizontal axis is time[s] since GPS time 1187008874[s]. The vertical axis is frequency [Hz].

Next, we describe the statistical evaluation method.

(1) Divide the data into 2, 3.5 or 5[s], and divide the frequency band of the data into 60[Hz] each from 60[Hz] to 300[Hz]. Figure 4 is the spectrogram on which red and black lines are drawn. The black lines are drawn at every 60[Hz] from 60[Hz]. The red lines are drawn at every 2[s] from 0[s].

(2) In each region, calculate the power of signal p, and create a histogram.

$$p = \Sigma {h_s}^2, \tag{12}$$

where h_s is the reconstructed scalar mode. Figure 5 is a histogram of the reconstructed scalar mode. The data used is divided into 2[s] and the frequency band is limited in 60∼120[Hz].

(3) Assuming that the probability distribution of p follows a normal distribution, find the mean μ and standard deviation σ from the distribution.

(4) Calculate the P-value and the signal-to-noise ratio(SNR).

$$\text{P-value} = \frac{n}{N}, \tag{13}$$

$$SNR = \frac{p_{gw} - \mu}{\sigma}, \tag{14}$$

where p_{GW} is the p in the region where GW170817 is in, n is the number of the p larger than p_{GW} and N is the total number of the p.

(5) Calculate the band limited root sum square of h_{rss} in the time-frequency domain with the highest SNR. The band limited root sum square of h, h_{rss}^{BL}, which we defined as follows

$$h_{rss}^{BL} = \sqrt{\sum_{t=t_1}^{t_2} \sum_{f=f_1}^{f_2} h_s(t,f)^2 dt}. \tag{15}$$

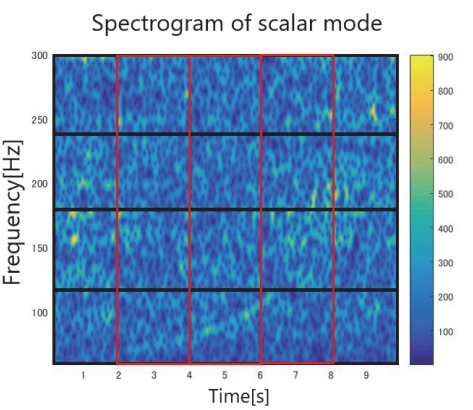


Fig. 4. Spectrogram of the reconstructed scalar-mode. The black lines are drawn at every 60[Hz] from 60[Hz]. The red lines are drawn at every 2[s] from 0[s].

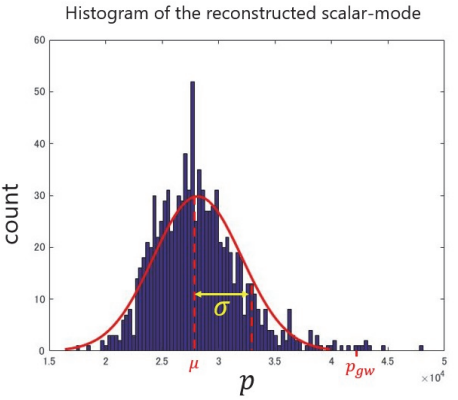


Fig. 5. Histogram of the reconstructed scalar-mode. The data used is divided into the 2[s] and the frequency band is limited in 60∼120[Hz]. μ is the mean value and σ is standard deviation of the p. p_{gw} is the p in the region where GW170817 is contained. n is the number of the p larger than p_{GW}.

Table 1. SNRs and P-values. The data used is divided into 2[s] and the frequency band is limited in 60∼120[Hz], 120∼180[Hz], 180∼240[Hz] and 240∼300[Hz].

frequency[Hz]	time[s]	2∼4	4∼6	6∼8
60∼120	P-value	0.13	0.01	0.60
	SNR	1.07	2.77	-0.308
120∼180	P-value	0.40	0.06	0.09
	SNR	0.0150	-0.453	1.39
180∼240	P-value	0.87	0.94	0.06
	SNR	-1.01	-1.33	1.59
240∼300	P-value	0.62	0.85	0.40
	SNR	-0.477	-0.897	0.0198

Figures 6, 7, 8, 9 shows spectrograms of the reconstructed scalar modes, and Tables 1, 2, 3 shows the SNRs and the P-values. In the Figs. 6, 7, 8, 9, the black lines are drawn at every 60[Hz] from 60[Hz], and the red lines are starting GPS time 1187008878∼1187008880[s], 1187008877∼1187008880.5[s], 1187008878∼1187008881.5[s], 1187008876∼1187008881[s], respectively.

The largest SNR was 2.77 with the time window of 2[s]. The P-value of the corresponding time-frequency region was 0.01. In the Fig. 6, the yellow frame is the region with the largest SNR. In case of the time window of 3.5[s], there are two regions where the SNR is larger than 2. The largest SNR was 2.54 with the time window of 3.5[s]. The P-value of the corresponding time-frequency region was 0.01. In the Fig. 7, the yellow frame is the region with the largest SNR. Also, The second

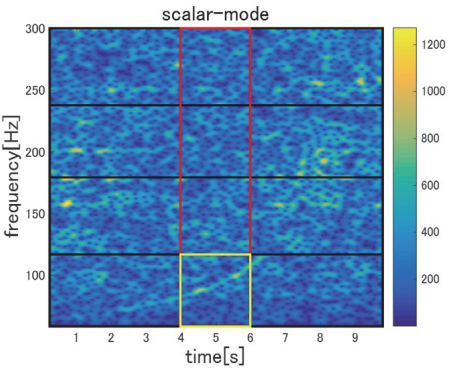


Fig. 6. Spectrogram of the reconstructed scalar-mode. The yellow frame is the region with highest SNR larger than 2.

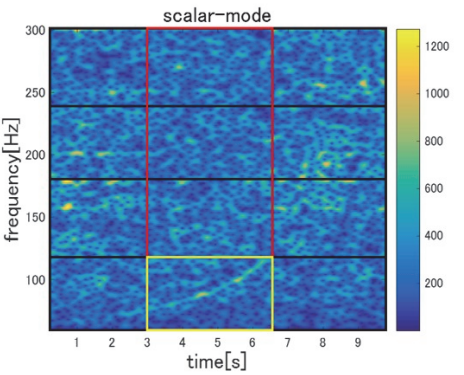


Fig. 7. Spectrogram of the reconstructed scalar-mode. The yellow frame is the region with highest SNR larger than 2.

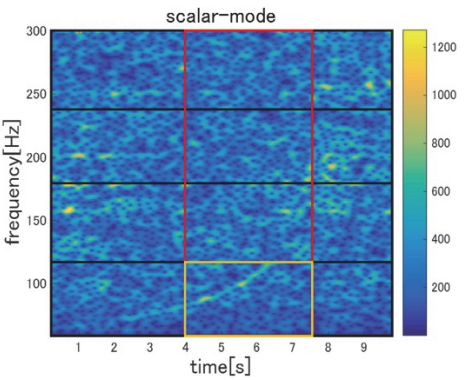


Fig. 8. Spectrogram of the reconstructed scalar-mode. The orange frame is the region with second highest SNR larger than 2.

Table 2. SNRs and P-values. The data used is divided into 3.5[s] and the frequency band is limited in 60∼120[Hz], 120∼180[Hz], 180∼240[Hz] and, 240∼300[Hz].

frequency[Hz]	time[s]	2∼5.5	3∼6.5	4∼7.5	5∼8.5
60∼120	P-value	0.03	0.01	0.03	0.23
	SNR	1.99	2.54	2.10	0.669
120∼180	P-value	0.61	0.67	0.25	0.09
	SNR	-0.377	-0.505	0.460	1.51
180∼240	P-value	0.96	0.97	0.734	0.04
	SNR	-1.59	-1.69	-0.790	1.94
240∼300	P-value	0.84	0.84	0.80	0.40
	SNR	-0.890	-0.887	-0.804	0.017

Table 3. SNRs and P-values. The data used is divided into 5[s] and the frequency band is limited in 60∼120[Hz], 120∼180[Hz], 180∼240[Hz] and 240∼300[Hz].

frequency[Hz]	time[s]	2∼7	3∼8
60∼120	P-value	0.02	0.06
	SNR	2.38	1.61
120∼180	P-value	0.44	0.27
	SNR	-0.0251	0.44
180∼240	P-value	0.96	0.58
	SNR	-1.50	-0.239
240∼300	P-value	0.81	0.71
	SNR	-0.860	-0.658

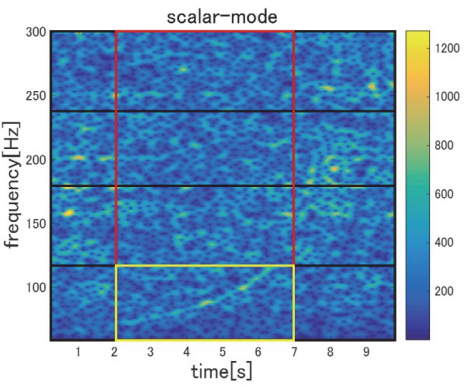


Fig. 9. Spectrogram of the reconstructed scalar-mode. The yellow frame is the region with highest SNR larger than 2.

largest SNR was 2.10. The P-value of the corresponding time-frequency region was 0.03. In the Fig. 8, the orange frame is the region with the second largest SNR. In case of the time window of 5[s], there is one region where the SNR is larger than 2. The largest SNR was 2.38 with the time window of 5[s]. The P-value of the corresponding time-frequency region was 0.02. In the Fig. 9, the yellow frame is the region with the largest SNR. Additionally, the h_{rss}^{BL} in the region with the maximum SNR was $1.55 \times 10^{-21}[1/\sqrt{Hz}]$.

4. Summary

We performed the model-independent test of the scalar-tensor gravity theory using GW170817 by reconstructing the tensor and scalar modes without any information about the waveform. We found the maximum SNR was 2.77 with the time window of 2[s] and the frequency band of 60∼120[Hz]. The P-value of the corresponding time-frequency region was 0.01, and the h_{rss}^{BL} of the same region was $1.55 \times 10^{-21}[1/\sqrt{Hz}]$. The SNR was small compared to threshold value of the GW detection by the matched filter analysis in recent years. In order to claim the detection of the scalar-mode, it is necessary to further increase the sensitivity of GW telescope and/or detect larger GWs. There are some future issues. First, we need evaluate the error that occurs in the process of the waveform reconstruction. Next, the method of reconstruction needs to be optimal for the scalar-mode. Our recent studies have shown that different locations have different amounts of leakage from the tensor mode. Therefore, by investigating it in detail, it will be possible to test the scalar-tensor gravity theory with high generality.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K03896. This research has made use of data, software and/or web tools obtained from Gravitational Wave Open Science Center (`https://www.gw-openscience.org`), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Center National de Recherche Scientifique (CNRS), the Italian Istituto Nazionaledella Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

References

1. B. P. Abbott, *et al.*, Phys. Rev. **X9**, 031040 (2019).
2. B. P. Abbott, *et al.*, Nature, **460**, 990 (2009).
3. D. M. Eardley, *et al.*, Phys. Rev. Lett. **30**, 884 (1973).
4. C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University Press, 1993).
5. E. P. Hubble, Proc. National Academy of Science of the USA, **15**, 168 (1929).
6. R. A. Alpher, H. Bethe, G. Gamow, Phys. Rev. **73**, 803 (1948).

7. K. Sato, Monthly Notices of Royal Astronomical Society, **195**, 467 (1981).
8. B. P. Abbott *et al.*, Phys. Rev. Lett. **123**, 011102 (2019).
9. B. P. Abbott *et al.*, Phys. Rev. Lett. **126**, 221101 (2016).
10. K. Hayama, A. Nishizawa, Phys. Rev. D **87**, 062003 (2013).
11. B. P. Abbott *et al.*, Phys. Rev. Lett. **119**, 161101 (2017).
12. LIGO Web page, `https://www.ligo.caltech.edu/MIT`
13. Virgo Web page, `https://www.virgo-gw.eu/`
14. LIGO Collab. *et al.*, Apj Lett. **848**, L13 (2017).
15. A. Nishizawa, *et al.*, Phys. Rev. D**79**, 082002 (2009).

Cosmology in the novel scalar-tensor representation of $f(R,T)$ gravity

Tiago B. Gonçalves

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisbon, Portugal
E-mail: tgoncalves@alunos.fc.ul.pt

João Luís Rosa

Institute of Physics, University of Tartu, W. Ostwaldi 1, 50411 Tartu, Estonia
E-mail: joaoluis92@gmail.com

Francisco S. N. Lobo

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisbon, Portugal
E-mail: fslobo@fc.ul.pt

We apply cosmological reconstruction methods to the $f(R,T)$ modified gravity, in its recently developed scalar-tensor representation. We do this analysis assuming a perfect fluid in a Friedmann-Lemaître-Robsertson-Walker (FLRW) universe. In this contribution we show the equations of motion obtained and we present the solutions found for one of the particular cases we analysed: an exponential evolution of the cosmological scale factor.

Keywords: Cosmology; modified gravity; scalar-tensor theories; accelerated expansion.

1. Introduction

The late-time universe is observed to be expanding in an accelerated way.[1,2] This behaviour can be modelled by introducing a dark energy component, often identified as a cosmological constant. However, we still lack the understanding for the dark energy mechanism at the fundamental level. Alternatively, it is possible to explain the accelerated expansion by modifying the theory of gravity. Many such modifications have been proposed to extend Einstein's General Relativity (GR) (e.g. see Refs. 3–9). For instance, one can modify the gravitational Lagrangian, which depends linearly on the Ricci curvature scalar R, to instead depend on a *general function* of R – thus, referred to as $f(R)$ gravity.[10] Resulting in modified Friedmann equations, $f(R)$ gravity can allow for the accelerated expansion without the need for dark energy.[11]

It is natural, then, to explore further extensions, such as $f(R,T)$ gravity[12] where the action depends on a general function of the curvature scalar R *and* of the trace of the stress-energy tensor T. Because of this explicit coupling between curvature and matter, the stress-energy tensor is not necessarily conserved, allowing the

appearance of a fifth force. Since its proposal, the $f(R,T)$ gravity has sparked a lot of interest leading, in some cases, to the study of its cosmological applications (e.g. see Refs. 13–19).

Recently, the original geometrical representation of the $f(R,T)$ gravity was reformulated into an equivalent scalar-tensor representation,[20] where the two extra degrees of freedom of the theory are expressed by means of two scalar fields. So far, this scalar-tensor theory was used to find thin shells' junction conditions[20] and thick brane solutions.[21] Since in scalar-tensor theories it is often easier to find analytical solutions, it calls for further study also of cosmological models in this representation of the $f(R,T)$ gravity. This study is undertaken in our work, modelling a Friedmann-Lemaître-Robsertson-Walker (FLRW) universe. We used reconstruction methods,[13–15,23,24] whereby the function $f(R,T)$ is not defined *a priori*, and instead we start from known cosmological solutions such as the time evolution of the spatial scale factor which seem to model well the observations. Here we give simply a preview of this work and some of its preliminary results. A more complete analysis is being incorporated into a paper under preparation.

2. The theory of the $f(R,T)$ gravity

We first introduce briefly the theory of $f(R,T)$ gravity, showing the main equations without detailing all of the steps. For further details we point to the main references. In its original geometrical formulation (see Ref. 12), the action of the $f(R,T)$ gravity is

$$S = \frac{1}{2\kappa^2}\int \sqrt{-g}\, f(R,T) d^4x + \int \sqrt{-g}\, \mathcal{L}_m d^4x, \tag{1}$$

where $\kappa^2 = 8\pi G/c^4$, G is the gravitational constant and c is the speed of light, g is the determinant of the metric $g_{\mu\nu}$, $f\,(R,T)$ is an arbitrary well-behaved function of the Ricci curvature scalar R and the trace of the stress-energy tensor T, and $\mathcal{L}_m$ is the matter Lagrangian. More explicitly, $R = g^{\mu\nu}R_{\mu\nu}$, where $R_{\mu\nu}$ is the Ricci tensor, and $T = g^{\mu\nu}T_{\mu\nu}$, where the stress-energy tensor $T_{\mu\nu}$ is defined in terms of the variation of the matter Lagrangian $\mathcal{L}_m$ as

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta\left(\sqrt{-g}\mathcal{L}_m\right)}{\delta g^{\mu\nu}}. \tag{2}$$

When reformulating into the dynamically equivalent scalar tensor representation (see Ref. 20 for further details), one defines two scalar fields

$$\varphi \equiv \frac{\partial f}{\partial R}, \qquad \psi \equiv \frac{\partial f}{\partial T}, \tag{3}$$

with the associated potential

$$V(\varphi,\psi) \equiv -f(R,T) + \varphi R + \psi T. \tag{4}$$

With these definitions, the action (1) can equivalently be written as

$$S = \frac{1}{2\kappa^2}\int_\Omega \sqrt{-g}\left[\varphi R + \psi T - V(\varphi,\psi)\right] d^4x + \int_\Omega \sqrt{-g}\mathcal{L}_m d^4x. \tag{5}$$

The variation of this action with respect to the metric $g_{\mu\nu}$ yields the modified field equations

$$\varphi R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \left(\varphi R + \psi T - V\right) - \left(\nabla_\mu \nabla_\nu - g_{\mu\nu} \Box\right) \varphi = \kappa^2 T_{\mu\nu} - \psi (T_{\mu\nu} + \Theta_{\mu\nu}), \quad (6)$$

where ∇_μ is the covariant derivative and $\Box \equiv \nabla^\sigma \nabla_\sigma$ is the D'Alembert operator, both defined in terms of the metric $g_{\mu\nu}$, and the tensor $\Theta_{\mu\nu}$ is defined as

$$\Theta_{\mu\nu} \equiv g^{\rho\sigma} \frac{\delta T_{\rho\sigma}}{\delta g^{\mu\nu}}. \quad (7)$$

On the other hand, the variation of the action (5) with respect to the scalar fields φ and ψ gives, respectively,

$$V_\varphi = R, \quad (8)$$

$$V_\psi = T, \quad (9)$$

where the subscripts in V_φ and V_ψ denote the partial derivatives of the potential $V(\varphi, \psi)$ with respect to the variables φ and ψ, respectively.

A general conservation equation can be found by taking the divergence of Eq. (6),

$$(\kappa^2 - \psi) \nabla^\mu T_{\mu\nu} = \left(T_{\mu\nu} + \Theta_{\mu\nu}\right) \nabla^\mu \psi + \psi \nabla^\mu \Theta_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \left[R \nabla^\mu \varphi + \nabla^\mu \left(\psi T - V\right)\right]. \quad (10)$$

Even though it is not a requirement of the theory, in this work we will impose the conservation of the stress-energy tensor, i.e. $\nabla^\mu T_{\mu\nu} = 0$, a condition that is automatically assured in GR. As detailed in the following section, for our cosmological model we use a FLRW metric and a perfect fluid approach in order to find the modified equations of motion.

In the following, we adopt a system of geometrized units in such a way that $G = c = 1$, and thus $\kappa^2 = 8\pi$.

3. Cosmology in the scalar-tensor representation

3.1. *Assumptions and equations of motion*

Let us first state the assumptions of our cosmological model with which we derived the equations of motion that are the focus of our study. We take the metric to be of the FLRW form, i.e., in spherical coordinates (t, r, θ, ϕ),

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2 \theta d\phi^2\right)\right], \quad (11)$$

where $a(t)$ is the scale factor and k is the curvature parameter which can take the values $k = \{-1, 0, 1\}$ corresponding to a hyperbolic, spatially flat, or hyperspherical universe, respectively.

We assume the stress-energy tensor $T_{\mu\nu}$ to be well described by that of an isotropic perfect fluid,

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \tag{12}$$

where ρ is the energy density, p is the isotropic pressure, and u^μ is the fluid 4-velocity satisfying the normalization condition $u_\mu u^\mu = -1$. We take the matter Lagrangian to be $\mathcal{L}_m = p$.[22] Under these considerations, the tensor $\Theta_{\mu\nu}$ takes the form

$$\Theta_{\mu\nu} = -2T_{\mu\nu} + p g_{\mu\nu}. \tag{13}$$

With these assumptions, and with all physical quantities depending only on the time coordinate t (which guarantees the homogeneity and isotropy of the solutions $\rho(t)$, $p(t)$, $\varphi(t)$, $\psi(t)$, etc.), the following equations of motion were obtained. From the two independent components of Eq. (6), we obtained the modified Friedmann equation and the modified Raychaudhuri equation,

$$\dot{\varphi}\left(\frac{\dot{a}}{a}\right) + \varphi\left(\frac{\dot{a}^2 + k}{a^2}\right) = \frac{8\pi}{3}\rho + \frac{\psi}{2}\left(\rho - \frac{1}{3}p\right) + \frac{1}{6}V, \tag{14}$$

$$\ddot{\varphi} + 2\dot{\varphi}\left(\frac{\dot{a}}{a}\right) + \varphi\left(\frac{2\ddot{a}}{a} + \frac{\dot{a}^2 + k}{a^2}\right) = -8\pi p + \frac{\psi}{2}(\rho - 3p) + \frac{1}{2}V, \tag{15}$$

respectively, where overdots denote derivatives with respect to time. From Eqs. (8) and (9) we obtained

$$V_\varphi = 6\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + k}{a^2}\right), \tag{16}$$

$$V_\psi = 3p - \rho, \tag{17}$$

respectively. By our choice, we imposed $\nabla_\mu T^{\mu\nu} = 0$ that leads to the usual continuity equation,

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p). \tag{18}$$

Consequently, from the general conservation equation (Eq. (10)) we obtained

$$2\dot{\psi}(\rho + p) = -\psi(\dot{\rho} - \dot{p}). \tag{19}$$

Finally, we impose an equation of state of the form

$$p = w\rho, \tag{20}$$

where w is a dimensionless parameter.

These equations of motion are the framework with which we searched for cosmological solutions in the $f(R,T)$ gravity. There is a redundancy in the equations obtained, since they are not all linearly independent, and one can be discarded without loss of generality. We chose to work with Eqs. (14), (16), (17), (18), (19), and (20) which form a system of six independent equations. This system is still

underdetermined since we have nine degrees of freedom: a, k, w, ρ, p, φ, ψ and V which contributes with two degrees of freedom (because it is a function of two variables, φ and ψ). In a first instance, we solved as much as it was possible without further constraining the system, obtaining some general solutions for $\rho(t)$, $p(t)$, $\psi(t)$, and the $V(\varphi, \psi)$ dependence on ψ. Only subsequently, did we provide further constraints to close the system and analyse particular cases.

3.2. *General solutions*

Taking the system of equations we have derived, Eqs. (14), (16), (17), (18), (19), and (20), we first found some general results. Beginning in a similar way that we would in GR, we used the equation of state (20) in the continuity equation (18) to find the general forms of the energy density and pressure as a function of a,

$$\rho = \rho_0 \left(\frac{a}{a_0}\right)^{-3(1+w)}, \tag{21}$$

$$p = w\rho_0 \left(\frac{a}{a_0}\right)^{-3(1+w)}, \tag{22}$$

respectively, where ρ_0 and a_0 are arbitrary constants from integration, and usually the subscript 0 denotes the value of the quantity today, at $t = t_0$. These solutions are identical to those in GR. This is only so because we required $\nabla_\mu T^{\mu\nu} = 0$ which, again, does not have to be generally true in modified gravity. Yet, it seems sensible to assume the validity of the continuity equation, at least on a first analysis.

Equipped with the results in Eqs. (21) and (22), we then solved Eq. (19) by integration and obtained

$$\psi = \psi_0 \left(\frac{a}{a_0}\right)^{\frac{3}{2}(1-w)}, \tag{23}$$

where ψ_0 is an integration constant. At this point, we were able to make the following considerations. The partial derivative V_ψ in Eq. (17) depends on ρ and p. We have the forms of $\rho(a)$ and $p(a)$ in Eqs. (21) and (22), and we can find a function $a(\psi)$ by inverting Eq. (23) (which we can do for $w \neq 1$, so, in what follows we will not be considering stiff matter). Thus, we combined these results to write the partial derivative V_ψ in Eq. (17) as depending only on ψ. We integrated the partial derivative to obtain a separable potential $V(\varphi, \psi) = V_0 + V_1(\varphi) + V_2(\psi)$, where V_0 is an arbitrary constant, $V_1(\varphi)$ is an arbitrary function of φ, and $V_2(\psi)$ is a function of ψ given by

$$V_2(\psi) = \frac{(1-3w)(1-w)}{(1+3w)} \rho_0 \psi_0^{\frac{2(1+w)}{(1-w)}} \psi^{-\frac{(1+3w)}{(1-w)}}, \tag{24}$$

which is undefined when $w = \{-1/3, 1\}$ and vanishes when $w = 1/3$ (corresponding to the case when the stress-energy tensor is traceless $T = 0$, such as for radiation).

This is as far as we were able to go without imposing further constraints. Up to this point, from the system of equations we are considering we have already used Eqs. (17), (18), (19), and (20), so we have only Eqs. (14) and (16) remaining . In balance, there are still five degrees of freedom at the moment: a, k, w, φ and $V_1(\varphi)$. Hence, we have to impose three further constraints in order to fully determine the system. We note that, as it stands, making a choice of the curvature and equation of state parameters, k and w, is not sufficient to determine the evolution of the universe modelled by $a(t)$. This is in contrast with GR where to each equation of state corresponds a unique evolution of the scale factor $a(t)$. Therefore, in this work we took the approach of first specifying the form of the scale factor, and then explored the parameter space of k and w in each particular case. The forms of the scale factor we chose are those known to model well different epochs of the universe: $a(t)$ given by an exponential or by a power law. By way of example, we present here only the particular case with an exponential scale factor, which can model the accelerated expansion observed in the present era.

3.3. *The particular case of an exponential expansion*

The universe is seen to be undergoing a phase of accelerated expansion. So, one of the particular cases we analysed was the de-Sitter solution that one obtains in GR, i.e., we impose in our system of equations an exponentially evolving scale factor of the form

$$a(t) = a_0 e^{\sqrt{\Lambda}(t-t_0)}, \tag{25}$$

where a_0, t_0 and Λ are constants. Once again, this particular choice does not require a unique value of w in $f(R,T)$ gravity and the solutions for ρ and ψ from Eqs. (21) and (23) become

$$\rho(t) = \rho_0 e^{-3\sqrt{\Lambda}(t-t_0)(1+w)}, \tag{26}$$

$$\psi(t) = \psi_0 e^{\frac{3}{2}\sqrt{\Lambda}(t-t_0)(1-w)}, \tag{27}$$

respectively. The pressure is simply given by $p\,(t) = w\rho\,(t)$.

To make further progress in order to find analytical solutions for $\varphi(t)$ and $V_1(\varphi)$, we set $k = 0$ (flat geometry) and left w arbitrary. With this choice of the curvature parameter and with the scale factor given by Eq. (25) the computation of Eq. (16) yields simply $V_\varphi = 12\Lambda$. Given that this is equivalent to having $dV_1/d\varphi = 12\Lambda$, we integrated it to obtain $V_1\,(\varphi) = 12\Lambda\varphi$. By doing so, we absorbed the arbitrary integration constant into V_0 in the full expression of the potential, which in this case reads as

$$V(\varphi,\psi) = V_0 + 12\Lambda\varphi + \frac{(1-3w)(1-w)}{(1+3w)}\rho_0\psi_0^{\frac{2(1+w)}{(1-w)}}\psi^{-\frac{(1+3w)}{(1-w)}}. \tag{28}$$

Having determined the form of the potential, we solved Eq. (14) to obtain $\varphi(t)$,

$$\varphi(t) = \varphi_0 e^{\sqrt{\Lambda}(t-t_0)} - \frac{V_0}{6\Lambda} - \frac{\rho_0}{3\Lambda}\left[\frac{8\pi e^{-3(1+w)\sqrt{\Lambda}(t-t_0)}}{(4+3w)} + \frac{4\,(1+w)\,\psi_0 e^{-\frac{3}{2}(1+3w)\sqrt{\Lambda}(t-t_0)}}{(1+3w)\,(5+9w)}\right], \tag{29}$$

where φ_0 is an integration constant. This solution $\varphi\,(t)$ is undefined at $w = \{-4/3, -5/9, -1/3\}$, and, from before, we were already not considering the $w = 1$ case. Apart from these discontinuities, the system does not pose further constraints to the equation of state. In the paper under preparation, we include a more complete analysis of this parameter space. However, here we show the solutions only for three particular choices of w.

A natural choice was to take $w = -1$, which in GR is the equation of state of constant dark energy density that drives an exponential expansion, corresponding to a cosmological constant. Given that we have the freedom to do so, we also set $w = 0$, i.e. the case in which pressureless (or dust-like) matter dominates. Alternatively, we also set $w = 1/3$ to model the case in which the energy density is dominated by radiation (remembering, for instance, that the stress-energy tensor of the electromagnetic field is traceless, i.e. $T = 3p - \rho = 0$ in the case of isotropic perfect fluid considered, corresponding to $w = 1/3$). That we tried different values for the equation of state, e.g. for matter and radiation, may be confusing at first given that we are considering an exponential growth of the universe, but we need only to remember that due to the extra degrees of freedom in $f(R, T)$ gravity, this cases can still be consistent within the system of equations we derived.

For each of these three values of the equation of state, $w = \{-1, 0, 1/3\}$, the energy density as given by Eq. (26) is plotted in Fig. 1. When space expands exponentially, the energy density is constant for $w = -1$ (cosmological constant), and undergoes exponential decay for $w = 0$ (dust) and $w = 1/3$ (radiation) — the decay is faster for radiation than for dust (due to the redshift of radiation). For each of the same three values of w, the scalar field $\psi(t)$ as given by Eq. (27) is plotted in Fig. 2. In the three cases, the field ψ grows exponentially with time — in descending order, it grows faster with the cosmological constant than with dust and than with radiation. Likewise, we plot the scalar field $\varphi\,(t)$ as given by Eq. (29) in Fig. 3. When $t \gg t_0$, all of the plotted $\varphi\,(t)$ solutions grow exponentially (in fact, they can all be approximated by $\varphi\,(t) \sim \varphi_0 \exp(\sqrt{\Lambda}t)$ in this limit). On the other hand, when $t < t_0$ the $\varphi\,(t)$ solutions show different behaviour between them. The two scalar fields do not have a straightforward physical interpretation; as a reminder, they were defined as the partial derivatives of the function $f\,(R, T)$, i.e. $\varphi = \partial f/\partial R$ and $\psi = \partial f/\partial T$. Finally, to each value of the equation of state corresponds a different form of the potential $V\,(\varphi, \psi)$ as given by Eq. (28),

$$w = -1: \qquad V(\varphi, \psi) = V_0 + 12\Lambda\varphi - 4\rho_0\psi, \tag{30}$$

$$w = 0: \qquad V(\varphi, \psi) = V_0 + 12\Lambda\varphi + \frac{\rho_0\psi_0^2}{\psi}, \tag{31}$$

$$w = 1/3: \qquad V(\varphi, \psi) = V_0 + 12\Lambda\varphi. \tag{32}$$

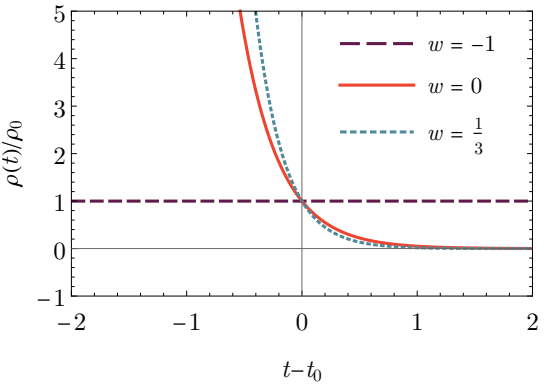


Fig. 1. Energy density $\rho(t)/\rho_0$ from Eq. (26) in the particular case where $a(t)$ is given by Eq. (25), setting $\Lambda = 1$, with the equation of state forced to be each of these values $w = \{-1, 0, 1/3\}$.

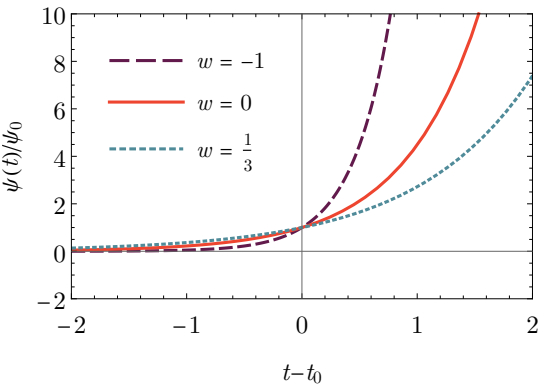


Fig. 2. Scalar field $\psi(t)/\psi_0$ from Eq. (27) in the particular case where $a(t)$ is given by Eq. (25), setting $\Lambda = 1$, with the equation of state forced to be each of these values $w = \{-1, 0, 1/3\}$.

As exemplified here, we were able to find solutions for all the variables in our system: $\rho(t)$, $p(t)$, $\varphi(t)$, $\psi(t)$ and $V(\varphi, \psi)$, for particular choices of $a(t)$, k and w. The next question we addressed in our work was: what form(s) of the function $f(R, T)$ generate these cosmologies?

3.4. *Reconstructing the function $f(R, T)$*

Given that we were able to obtain consistent solutions, one may hope that it is possible to reconstruct the explicit form of the function $f(R, T)$ that can lead to those solutions. For instance, we have found the form of the potential $V(\varphi, \psi)$ and, by definition, the potential is related to the function $f(R, T)$, as per Eq. (4) which

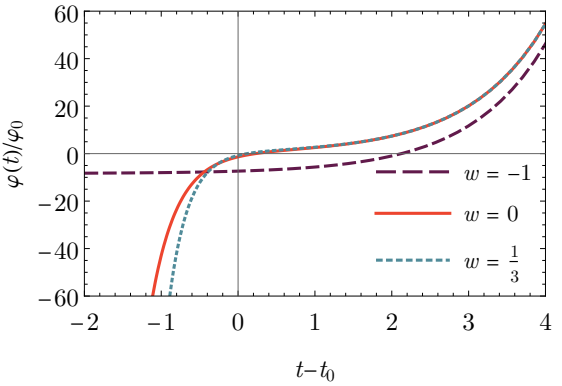


Fig. 3. Scalar field $\varphi(t)/\varphi_0$ from Eq. (29), setting $\varphi_0 = \rho_0 = \psi_0 = 1$, in the particular case where $a(t)$ is given by Eq. (25), setting $\Lambda = 1$, with the equation of state forced to be each of these values $w = \{-1, 0, 1/3\}$.

can be rewritten as

$$f(R,T) = -V(f_R, f_T) + f_R R + f_T T, \tag{33}$$

where we have changed the notation using $\varphi = f_R$ and $\psi = f_T$, with the subscripts R and T denoting the partial derivatives. Taking the partial derivatives of this equation with respect to R and T, we obtained a system of two PDEs which are satisfied for an arbitrary $f(R,T)$ if $V_\varphi = R$ and $V_\psi = T$. These two conditions are the equations of motion we obtained for the fields φ and ψ, in Eqs. (8) and (9), and so are always satisfied. It can be seen, e.g. from Eq. (28), that

$$V_\psi = T = (3w-1)\rho_0 \left(\frac{f_T}{\psi_0}\right)^{-\frac{2(1+w)}{(1-w)}}, \tag{34}$$

which is not defined for $w = 1$. If $w = \{-1, 1/3\}$, this equation does not give a relation between f_T and T. On the other hand, if $w = 0$, for instance, then $T = -\rho_0 \psi_0^2 / f_T^2$. This relation can be inverted and integrated to give

$$f(R,T) = g(R) \pm 2\psi_0 \sqrt{-\rho_0 T}, \tag{35}$$

where $g(R)$ is an arbitrary function of R. In the case of a flat universe ($k = 0$) with an exponentially evolving scale factor as given by Eq. (25), $V_\varphi = R = 12\Lambda$ which does not provide a direct relation between f_R and R, and so it does constrain the form of $g(R)$. Other particular cases are studied in our work and will be presented in the upcoming paper.

4. Summary and concluding remarks

Modified gravity has the potential to solve some of the remaining questions in cosmology, such as the current accelerated expansion. With this motivation, we

took the recently developed scalar-tensor representation of the $f(R,T)$ gravity to explore allowed FLRW cosmologies. To keep our contribution short, we presented here only one particular case: that of an exponential expansion. As opposed to GR, in $f(R,T)$ gravity there is not a one-to-one relation between the evolution of the scale factor and the equation of state. For instance, we were able to find a solution by choosing a flat, dust dominated, exponentially expanding universe, which would not have been possible in GR. The extra degrees of freedom in modified gravity may contribute to an effective dark energy, thus allowing an exponential expansion to occur even when the equation of state is that of matter or radiation. In the paper under preparation, we include other particular cases, such as power law evolution of the scale factor, non flat geometries, and other equations of state. We are also already working on testing whether finite time singularities can appear in $f(R,T)$ gravity.

Acknowledgments

TBG is funded by a PhD Research Fellowship in the context of the Fundação para a Ciência e Tecnologia (FCT) project "DarkRipple" with reference PTDC/FIS-OUT/29048/2017. JLR was supported by the European Regional Development Fund and the programme Mobilitas Pluss (MOBJD647). FSNL acknowledges support from the Fundação para a Ciência e a Tecnologia (FCT) Scientific Employment Stimulus contract with reference CEECINST/00032/2018, and funding from the research grants No. UIDB/FIS/04434/2020, UIDP/FIS/04434/2020, No. PTDC/FIS-OUT/29048/2017 and No. CERN/FIS-PAR/0037/2019.

References

1. S. Perlmutter *et al.* [Supernova Cosmology Project], "Measurements of Ω and Λ from 42 high redshift supernovae," Astrophys. J. **517**, 565-586 (1999) [arXiv:astro-ph/9812133 [astro-ph]].
2. A. G. Riess *et al.* [Supernova Search Team], "Observational evidence from supernovae for an accelerating universe and a cosmological constant," Astron. J. **116**, 1009-1038 (1998) [arXiv:astro-ph/9805201 [astro-ph]].
3. S. Nojiri and S. D. Odintsov, "Introduction to modified gravity and gravitational alternative for dark energy," eConf **C0602061**, 06 (2006) [arXiv:hep-th/0601213 [hep-th]].
4. F. S. N. Lobo, "The Dark side of gravity: Modified theories of gravity," [arXiv: 0807.1640 [gr-qc]].
5. S. Nojiri and S. D. Odintsov, "Unified cosmic history in modified gravity: From F(R) theory to Lorentz non-invariant models," Phys. Rept. **505**, 59-144 (2011) [arXiv:1011.0544 [gr-qc]].
6. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, "Modified Gravity and Cosmology," Phys. Rept. **513**, 1-189 (2012) [arXiv:1106.2476 [astro-ph.CO]].
7. S. Capozziello and M. De Laurentis, "Extended Theories of Gravity," Phys. Rept. **509**, 167-321 (2011) [arXiv:1108.6266 [gr-qc]].
8. E. N. Saridakis *et al.* [CANTATA], "Modified Gravity and Cosmology: An Update by the CANTATA Network," [arXiv:2105.12582 [gr-qc]].

9. P. Avelino, T. Barreiro, C. S. Carvalho, A. da Silva, F. S. N. Lobo, P. Martin-Moruno, J. P. Mimoso, N. J. Nunes, D. Rubiera-Garcia and D. Saez-Gomez, *et al.* "Unveiling the Dynamics of the Universe," Symmetry **8**, no. 8, 70 (2016) [arXiv:1607.02979 [astro-ph.CO]].
10. T. P. Sotiriou and V. Faraoni, "f(R) Theories Of Gravity," Rev. Mod. Phys. **82**, 451-497 (2010) [arXiv:0805.1726 [gr-qc]].
11. S. Capozziello, "Curvature quintessence," Int. J. Mod. Phys. D **11**, 483-492 (2002) [arXiv:gr-qc/0201033 [gr-qc]].
12. T. Harko, F. S. N. Lobo, S. Nojiri and S. D. Odintsov, "$f(R,T)$ gravity," Phys. Rev. D **84**, 024020 (2011) [arXiv:1104.2669 [gr-qc]].
13. M. Jamil, D. Momeni, M. Raza and R. Myrzakulov, "Reconstruction of some cosmological models in f(R,T) gravity," Eur. Phys. J. C **72**, 1999 (2012) [arXiv:1107.5807 [physics.gen-ph]].
14. M. J. S. Houndjo and O. F. Piattella, "Reconstructing $f(R,T)$ gravity from holographic dark energy," Int. J. Mod. Phys. D **21**, 1250024 (2012) [arXiv:1111.4275 [gr-qc]].
15. M. J. S. Houndjo, "Reconstruction of f(R, T) gravity describing matter dominated and accelerated phases," Int. J. Mod. Phys. D **21**, 1250003 (2012) [arXiv:1107.3887 [astro-ph.CO]].
16. F. G. Alvarenga, A. de la Cruz-Dombriz, M. J. S. Houndjo, M. E. Rodrigues and D. Sáez-Gómez, "Dynamics of scalar perturbations in $f(R,T)$ gravity," Phys. Rev. D **87**, no. 10, 103526 (2013) [erratum: Phys. Rev. D **87**, no. 12, 129905 (2013)] [arXiv:1302.1866 [gr-qc]].
17. H. Shabani and M. Farhoudi, "$f(R,T)$ Cosmological Models in Phase Space," Phys. Rev. D **88**, 044048 (2013) [arXiv:1306.3164 [gr-qc]].
18. H. Shabani and M. Farhoudi, "Cosmological and Solar System Consequences of f(R,T) Gravity Models," Phys. Rev. D **90**, no. 4, 044031 (2014) [arXiv:1407.6187 [gr-qc]].
19. P. H. R. S. Moraes, "Cosmological solutions from Induced Matter Model applied to 5D $f(R,T)$ gravity and the shrinking of the extra coordinate," Eur. Phys. J. C **75**, no. 4, 168 (2015) [arXiv:1502.02593 [gr-qc]].
20. J. L. Rosa, "Junction conditions and thin shells in perfect-fluid $f(R,T)$ gravity," Phys. Rev. D **103**, no. 10, 104069 (2021) [arXiv:2103.11698 [gr-qc]].
21. J. L. Rosa, M. A. Marques, D. Bazeia and F. S. N. Lobo, "Thick branes in the scalar-tensor representation of $f(R,T)$ gravity," [arXiv:2105.06101 [gr-qc]].
22. O. Bertolami, F. S. N. Lobo and J. Paramos, "Non-minimum coupling of perfect fluids to curvature," Phys. Rev. D **78**, 064036 (2008) [arXiv:0806.4434 [gr-qc]].
23. S. Capozziello, V. F. Cardone and A. Troisi, "Reconciling dark energy models with f(R) theories," Phys. Rev. D **71**, 043503 (2005) [arXiv:astro-ph/0501426 [astro-ph]].
24. T. Multamaki and I. Vilja, "Cosmological expansion and the uniqueness of gravitational action," Phys. Rev. D **73**, 024018 (2006) [arXiv:astro-ph/0506692 [astro-ph]].

On the interaction between electromagnetic, gravitational, and plasma related perturbations on LRS class II spacetimes

P. Semrén

Department of Physics, Umeå University, Umeå, SE-901 87, Sweden
E-mail: philip.semren@umu.se

We consider first order perturbations on locally rotationally symmetric (LRS) class II spacetimes. In particular, we investigate the interactions between electromagnetic, gravitational, and plasma related perturbations when allowing for a non-zero magnetic field on the background spacetime. In doing so we focus on mechanisms for generating and magnifying the magnetic fields, as these kinds of mechanisms may have some relevance when trying to explain the origin of the observed large scale cosmic magnetic fields. The equations governing the behavior of the perturbations are gathered from the Ricci identities for certain preferred vector fields, the Bianchi identities, Maxwell's equations, and from relations describing energy-momentum conservation and particle conservation for the cosmological plasma. After employing a simplified cold magnetohydrodynamic (MHD) description, and harmonically decomposing the spatial dependencies, we arrive at a closed set of ordinary differential equations in time. On analyzing this system, which decouples into an even and an odd subsector, we indeed observe possible mechanisms for generating magnetic field perturbations to first order.

Keywords: Cosmological perturbations; LRS Class II; plasma; MHD.

1. Introduction

Using the standard ΛCDM model of cosmology, assumed to be homogeneous and isotropic on average, most of our current observations can be explained. However, as the bounds on a possible anisotropy in our universe still remains quite unrestrictive,[1,2] it is still interesting to study more general anisotropic models. These models have a special relevance when considering electromagnetic perturbations in a cosmological context. When using the standard models as background spacetimes in a perturbative approach, the isotropy of these models make it is essentially impossible to have a non-zero magnetic field on the zeroth order level, since such a field would define a preferred spatial direction. Therefore, when considering electromagnetic perturbations on a FLRW background, previous works have often found it necessary to employ a non-linear perturbation theory to maintain gauge invariance and to be able to see any interesting interactions between the electromagnetic fields and the variables related to gravitational effects.[3–5] However, using an anisotropic model as background spacetime, there is no fundamental problem of having a non-zero magnetic field to zeroth order.

In this work, we take a first step towards a more general anisotropic model by using a $1 + 1 + 2$ covariant split of spacetime to investigate linear perturbations

on homogenous and hypersurface orthogonal locally rotationally symmetric (LRS) class II spacetimes. Since these spacetimes have a preferred spatial direction around which they are locally rotationally symmetric, we are able to include a magnetic field to zeroth order in a perturbative approach. The aim is then to investigate if this allows for interesting interactions between the electromagnetic fields and the gravitational variables already to first order in the perturbations.

As for linear perturbations on LRS class II cosmologies, these have been studied extensively before for perfect fluids, with and without vorticity.[6–8] A study of more general and dissipative perturbations can be found in Ref. 9 and in a forthcoming paper. Regarding electromagnetic perturbations, these have also been considered before, but, as far as we know, only when treating these perturbations as test fields or in rather restrictive cases.[10,11] This work therefore seeks to present a more general treatment of electromagnetic perturbations on LRS class II cosmologies when using a cold magnetohydrodynamic (MHD) approximation to describe the details of the cosmological plasma. However, in doing so, we will only outline the main steps and results. For more details, the reader is referred to Ref. 12 or a forthcoming paper to be published elsewhere.

2. Governing Equations

The studied LRS spacetimes are assumed to have two preferred directions, specified by a time-like vector field, u^a, and a space-like vector field n^a. These vector fields are assumed to be normalized as $u^a u_a = -1$ and $n^a n_a = 1$. The time-like direction could represent the 4-velocity of the cosmological fluid or some congruence of fundamental observers, whilst the spatial direction is the direction around which the spacetimes are locally rotationally symmetric.

2.1. *Spacetime Dynamics*

The characteristic directions u^a and n^a play a crucial role when describing the spacetime dynamics, as we will get many of our equations from the Ricci identities for these vector fields, namely

$$2\nabla_{[a}\nabla_{b]}u_c - R_{abcd}u^d = 0, \tag{1}$$

$$2\nabla_{[a}\nabla_{b]}n_c - R_{abcd}n^d = 0. \tag{2}$$

In addition to these equations, we will also make use of the contracted Bianchi identities

$$2\nabla_{[e}R_{d]b} + \nabla_a R^a{}_{bde} = 0, \tag{3}$$

$$\nabla^a G_{ab} = 0, \tag{4}$$

where $G_{ab} \equiv R_{ab} - g_{ab}R/2$ is the Einstein tensor. We will also implicitly enforce Einstein's field equations

$$G_{ab} = T_{ab} - \Lambda g_{ab}, \tag{5}$$

by using them to write the Ricci tensor and Ricci scalar in terms of the energy momentum tensor T_{ab} and the cosmological constant Λ.

Although the relations stated above will give us many of the equations that we need, they will in general not result in a closed set. To get the remaning relations, we also need to specify the detailed properties of the energy-momentum tensor.

2.2. *Plasma Dynamics*

In our case, where we want to describe a cosmological plasma together with its associated electromagnetic fields, we can divide the energy-momentum tensor as

$$T^{ab} = {T_{(F)}}^{ab} + {T_{(EM)}}^{ab}, \tag{6}$$

where ${T_{(F)}}^{ab}$ is due to the fluid whilst ${T_{(EM)}}^{ab}$ is due to the electromagnetic fields. The electromagnetic part can be written as

$${T_{(EM)}}^{ab} = {F^a}_c F^{bc} - \frac{1}{4} g^{ab} F^{cd} F_{cd}\,, \tag{7}$$

in terms of the Faraday tensor F_{ab}, which in turn is assumed to satisfy Maxwell's equations

$$\nabla_{[a} F_{bc]} = 0, \tag{8}$$

$$\nabla_b F^{ab} = j^a. \tag{9}$$

Here j^a is the 4-current density, which can be written as $j^a = \rho u^a + J^a$ in terms of the charge density ρ and the 3-current density J^a, where $u^a J_a = 0$.

Following the formalism in Ref. 13, the behavior of the fluid contribution can be deduced by first treating the cosmological plasma as a charged multi-component fluid, where the ith component is assumed to have the 4-velocity

$${u_{(i)}}^a = \gamma_{(i)} \left(u^a + {v_{(i)}}^a \right), \tag{10}$$

with the properties

$$u^a {v_{(i)}}_a = 0, \quad \gamma_{(i)} = \frac{1}{\sqrt{1 - {v_{(i)}}^2}}, \quad {v_{(i)}}^2 \equiv {v_{(i)}}_a {v_{(i)}}^a. \tag{11}$$

The dynamics of the ith component is then determined by imposing the following equations, describing conservation of energy-momentum and particles

$$\nabla_b {T_{(i)}}^{ab} = F^{ab} {j_{(i)}}_b + {I_{(i)}}^a, \tag{12}$$

$$\nabla_a \left(\mathcal{N}_{(i)} {u_{(i)}}^a \right) = 0. \tag{13}$$

Here, $\mathcal{N}_{(i)}$ is the number density for the ith component relative to its own rest frame, whilst ${T_{(i)}}^{ab}$ is the energy-momentum contribution from this component, so that

$${T_{(F)}}^{ab} = \sum_{(i)} {T_{(i)}}^{ab}. \tag{14}$$

As for the right hand side in Eq. (12), ${I_{(i)}}^a$ includes interactions, such as collisions, between the fluid components, whilst ${j_{(i)}}^a$ is the 4-current density of the ith component, which satisfies the relations

$$ {j_{(i)}}^a = q_{c(i)} \mathcal{N}_{(i)} {u_{(i)}}^a, \tag{15} $$

$$ j^a = \sum_{(i)} {j_{(i)}}^a. \tag{16} $$

Here $q_{c(i)}$ is the charge of the ith species. Furthermore, we will also assume that the ith component is a perfect fluid, so that

$$ {T_{(i)}}^{ab} = \left(\mu_{(i)} + p_{(i)}\right) {u_{(i)}}^a {u_{(i)}}^b + p_{(i)} g^{ab}, \tag{17} $$

where $\mu_{(i)}$ is the energy density and $p_{(i)}$ is the isotropic pressure relative to the rest frame of the fluid component. After specifying some reasonable equations of state $p_{(i)} = p_{(i)}(\mu_{(i)}, \mathcal{N}_{(i)})$, Eqs. (1)-(17) are enough to be able to get a closed system in the end. To simplify this system, we will in the following employ a cold MHD description. Instead of working directly with the individual fluid components, we then want to rewrite the system in terms of some collective variables. For this purpose, we will first assume that the plasma is sufficiently cold, so that the components can be seen as dust-like fluids with negligible pressures $p_{(i)}$. We will then, after having linearized the equations as described later on, rewrite the equations in terms of the following collective variables

$$ \mu_{(F)} \equiv \sum_i \mu_{(i)}, $$

$$ \mu_{(F)} {v_{(F)}}^a \equiv \sum_i \mu_{(i)} {v_{(i)}}^a, $$

where $\mu_{(F)}$ is the total energy density of the fluid and ${v_{(F)}}^a$ is an average plasma 3-velocity. In the following we will assume that the plasma consists of an electron component and a much more massive ion counterpart. Then, on following the usual approximations of the resistive MHD description, it is possible to derive an Ohm's law of the form[a]

$$ J^a = \frac{1}{\eta} \left(E^a + \epsilon^{abc} {v_{(F)}}_b B_c\right), $$

which will be used to describe the 3-current density. Here, η is the electrical resistivity whilst E^a and B^a are the electric and magnetic fields as defined by the relation

$$ F_{ab} = \epsilon_{abc} B^c + 2 u_{[a} E_{b]}, $$

where $\epsilon_{abc} \equiv \eta_{dabc} u^d \equiv 4! \sqrt{-g} \delta^0{}_{[d} \delta^1{}_a \delta^2{}_b \delta^3{}_{c]} u^d$ is the 3-dimensional Levi-Civita volume element. As for the charge density, total charge neutrality is assumed on

[a] A simplified derivation starting from the linearized and harmonically decomposed multi-component equations can be found in Ref. 12.

the background spacetime, where the plasma velocities ${u_{(i)}}^a$ are assumed to coincide with the fundamental observers u^a. To first order, however, the charge density will be allowed to be non-zero, and is determined by the electromagnetic fields through Maxwell's equations in the MHD description.

Before delving deeper into the perturbations, however, we first break the relevant equations down into smaller pieces by using covariant splits of spacetime.

3. Covariant Splits of Spacetime

Since the studied background spacetimes are equipped with two preferred directions, it is natural to decompose our variables with respect to the vector fields u^a and n^a. We start by performing a $1+3$ split with respect to u^a, after which the 3-dimensional hypersurfaces orthogonal to u^a are decomposed with respect to n^a, yielding a $1+1+2$ covariant split of spacetime.[14]

3.1. 1 + 3

Starting with the $1+3$ split, we first write the metric as

$$g_{ab} = U_{ab} + h_{ab}, \tag{18}$$

where $U_{ab} = -u_a u_b$ is a projection tensor along u^a, whilst h_{ab} projects onto the 3-dimensional hypersurfaces orthogonal to u^a. Using Eq. (18) we may then decompose every other quantity with respect to u^a. This has already been done for the electromagnetic and plasma-related variables in the previous section, but we note the following new variables which emerge as projections of the total energy-momentum tensor and the Weyl tensor

$$\begin{aligned}
\mu &\equiv T^{cd} u_c u_d, \\
q_a &\equiv -T^{cd} u_c h_{da}, \\
p &\equiv \frac{1}{3} T^{cd} h_{cd}, \\
\pi_{ab} &\equiv \left(h_{(a}{}^c h_{b)}{}^d - \frac{1}{3} h^{cd} h_{ab} \right) T_{cd} \equiv T_{\langle ab \rangle}, \\
E_{ab} &\equiv C_{acbd} u^c u^d, \\
H_{ab} &\equiv \frac{1}{2} \epsilon_a{}^{ef} C_{efbd} u^d.
\end{aligned}$$

Here μ is the total energy density, q_a is the total energy flow relative to u^a, p is the total isotropic pressure, and π_{ab} is the total anisotropic pressure, whilst E_{ab} and H_{ab} are the electric and magnetic parts of the Weyl tensor C_{abcd}. When defining the anisotropic pressure, we have introduced a notation with angular brackets to denote the Projected Symmetric and Trace-Free (PSTF) part of a tensor with respect to h_{ab}.

To describe the spacetime dynamics, we also have to decompose the covariant derivatives. For this purpose we define a time derivative

$$\dot{T}^{ab\ldots}{}_{cd\ldots} \equiv u^e \nabla_e T^{ab\ldots}{}_{cd\ldots},$$

and a spatial derivative that is fully projected onto the 3-dimensional hypersurfaces

$$D_e T^{ab\ldots}{}_{cd\ldots} \equiv h^a{}_f h^b{}_g \cdots h_c{}^h h_d{}^i \cdots h_e{}^j \nabla_j T^{fg\ldots}{}_{hi\ldots}.$$

Using these definitions, we may write the covariant derivative of u^a as

$$\nabla_a u_b = -u_a \dot{u}_b + \frac{1}{3}\Theta h_{ab} + \omega_{ab} + \sigma_{ab},$$

where $\Theta \equiv D^a u_a$ is the expansion rate, $\omega_{ab} \equiv D_{[a} u_{b]}$ is the vorticity, and $\sigma_{ab} \equiv D_{\langle a} u_{b\rangle}$ is the shear. When dealing with the vorticity we will in the following use the vorticity 3-vector

$$\omega^a \equiv \frac{1}{2}\epsilon^{abc}\omega_{bc},$$

rather than the vorticity tensor.

3.2. 1 + 1 + 2

Having decomposed the relevant quantities with respect to u^a, we perform an additional decomposition with respect to n^a. After writing the projection tensor h_{ab} as

$$h_{ab} = N_{ab} + n_a n_b,$$

we may then write the relevant 3-vectors as

$$\begin{aligned} q^a &= Q n^a + Q^a, & B^a &= \mathfrak{B} n^a + \mathfrak{B}^a, \\ \omega^a &= \Omega n^a + \Omega^a, & E^a &= \mathfrak{E} n^a + \mathfrak{E}^a, \\ \dot{u}^a &= \mathcal{A} n^a + \mathcal{A}^a, & J^a &= \mathfrak{J} n^a + \mathfrak{J}^a, \\ \dot{n}^a &= \mathcal{A} u^a + \alpha^a, & v_{(F)}{}^a &= \mathcal{V}_{(F)} n^a + \mathcal{V}_{(F)}{}^a, \end{aligned}$$

where the rightmost vectors in each relation are orthogonal to both u^a and n^b. As for PSTF 3-tensors ψ_{ab}, we decompose these as

$$\psi_{ab} = \Psi\left(n_a n_b - \frac{1}{2}N_{ab}\right) + 2n_{(a}\Psi_{b)} + \Psi_{ab},$$

where $n^a \Psi_a = 0$ and

$$\Psi_{ab} = \left(N_{(a}{}^c N_{b)}{}^d - \frac{1}{2}N_{ab}N^{cd}\right)\psi_{cd} \equiv \psi_{\{ab\}}.$$

Here we have introduced a notation with curly brackets to denote the PSTF part of a tensor with respect to N_{ab}. For the tensors relevant to us, we can therefore write

$$\pi_{ab} = \Pi\left(n_a n_b - \frac{1}{2}N_{ab}\right) + 2n_{(a}\Pi_{b)} + \Pi_{ab},$$

$$E_{ab} = \mathcal{E}\left(n_a n_b - \frac{1}{2}N_{ab}\right) + 2n_{(a}\mathcal{E}_{b)} + \mathcal{E}_{ab},$$
$$H_{ab} = \mathcal{H}\left(n_a n_b - \frac{1}{2}N_{ab}\right) + 2n_{(a}\mathcal{H}_{b)} + \mathcal{H}_{ab},$$
$$\sigma_{ab} = \Sigma\left(n_a n_b - \frac{1}{2}N_{ab}\right) + 2n_{(a}\Sigma_{b)} + \Sigma_{ab}.$$

Finally, we also need to decompose the spatial derivatives. For this purpose we define a spatial derivative along n^a

$$\hat{T}^{ab\ldots}{}_{cd\ldots} \equiv n^e D_e T^{ab\ldots}{}_{cd\ldots},$$

and a spatial derivative fully projected onto the 2-sheets orthogonal to both u^a and n^a

$$\delta_e T^{ab\ldots}{}_{cd\ldots} \equiv N^a{}_f\, N^b{}_g \cdots N_c{}^h N_d{}^i \cdots N_e{}^j D_j T^{fg\ldots}{}_{hi\ldots}.$$

With these definitions, we may write the spatial derivative of n^a as

$$D_a n_b = n_a a_b + \frac{1}{2}\phi N_{ab} + \xi\epsilon_{ab} + \zeta_{ab},$$

where $\epsilon_{ab} \equiv \epsilon_{abc}n^c$, $\phi \equiv \delta^c n_c$ describes an expansion of the 2-sheets, $\xi\epsilon_{ab} \equiv \delta_{[a}n_{b]}$ describes a twist of these sheets, $\zeta_{ab} \equiv \delta_{\{a}n_{b\}}$ is the distortion, and $a_a \equiv \hat{n}_a$.

4. Background Spacetimes

Having decomposed the equations using the $1+1+2$ covariant split of spacetime, we move on to consider these equations perturbatively, starting with the zeroth order background spacetime. We will assume that this spacetime is a homogeneous and hypersurface orthogonal member of LRS class II, which is characterized by that H_{ab}, ω_{ab} and ξ are all identically zero. To be able to use the same type of harmonics for all of the studied spacetimes, we will also assume that $\phi = 0$. These spacetimes can be described with the line element

$$ds^2 = -dt^2 + a_1^2(t)\,dz^2 + a_2^2(t)\left(d\vartheta^2 + f_\kappa(\vartheta)d\varphi^2\right),$$

where the scale factors a_1 and a_2 satisfy the relations

$$\Theta = \frac{\dot{a}_1}{a_1} + 2\frac{\dot{a}_2}{a_2},$$
$$\Sigma = \frac{2}{3}\left(\frac{\dot{a}_1}{a_1} - \frac{\dot{a}_2}{a_2}\right).$$

The function $f_\kappa(\vartheta)$ is determined by the 2D-curvature scalar

$$\mathcal{R} = \frac{2\kappa}{a_2^2} = 2(\mu + \Lambda) + \frac{3\Sigma^2}{2} - \frac{2\Theta^2}{3}.$$

When $\kappa = 1$, the sheets are 2-spheres with $f_1(\vartheta) = \sin^2\vartheta$. If $\kappa = 0$ the 2-sheets are flat and $f_0(\vartheta) = 1$ (or ϑ^2). Lastly, if $\kappa = -1$, the 2-sheets are pseudo-spheres with $f_{-1}(\vartheta) = \sinh^2\vartheta$.

As previously mentioned, we will to zeroth order assume charge neutrality and that the plasma velocities coincide with the fundamental observers. We will also assume that interactions between the plasma components, as described by the last term in Eq. (12), vanish to zeroth order. With these assumptions, our choice of background spacetimes implies that the only relevant non-zero dynamical quantities to zeroth order are

$$S^{(0)} = \left\{\Theta, \Sigma, \mu_{(F)}, \mathfrak{B}, \mathcal{E}, \mu, p, \Pi\right\},$$

which are found to satisfy the equations

$$\begin{aligned}
\dot{\Theta} &= -\frac{\Theta^2}{3} - \frac{3\Sigma^2}{2} - \frac{1}{2}(\mu + 3p) + \Lambda,\\
\dot{\Sigma} &= \frac{2}{3}(\mu + \Lambda) + \frac{\Sigma^2}{2} - \Sigma\Theta - \frac{2\Theta^2}{9} + \Pi,\\
\dot{\mu}_{(F)} &= -\mu_{(F)}\Theta,\\
\dot{\mathfrak{B}} &= \left(\Sigma - \frac{2\Theta}{3}\right)\mathfrak{B},
\end{aligned}$$

where

$$\begin{aligned}
3\mathcal{E} &= -2(\mu + \Lambda) - 3\Sigma^2 + \frac{2\Theta^2}{3} + \Sigma\Theta - \frac{3\Pi}{2},\\
\mu &= \mu_{(F)} + \frac{\mathfrak{B}^2}{2},\\
p &= \frac{\mathfrak{B}^2}{6},\\
\Pi &= -\frac{2\mathfrak{B}^2}{3}.
\end{aligned}$$

5. Perturbations

With the behavior of the background determined, we now move on to consider first order perturbations in the gravitational, electromagnetic and plasma-related variables. To ensure gauge invariance, we will use quantities which are zero on the background spacetime to describe the perturbations, as these are gauge invariant in accordance with the Stewart–Walker lemma.[15] However, since there are some non-zero variables on the background, we have to define some new variables to describe the perturbations in these quantities. For this purpose we take advantage of the background homogeneity and let the perturbations of the non-zero background variables be described by spatial gradients of these variables. Therefore, we define

the following first order quantities[b]

$$X_a \equiv \delta_a \mathcal{E}, \quad V_a \equiv \delta_a \Sigma, \quad W_a \equiv \delta_a \Theta, \quad \mu_{(F)_a} \equiv \delta_a \mu_{(F)},$$
$$Y_a \equiv \delta_a \Pi, \quad \mu_a \equiv \delta_a \mu, \quad p_a \equiv \delta_a p, \quad \mathcal{B}_a \equiv \delta_a \mathfrak{B}.$$

Collecting these variables with those that are naturally zero on the background, the complete set of gauge invariant first order quantities are

$$S^{(1)} \equiv \{X_a, V_a, W_a, \mu_a, p_a, Y_a, \mathcal{A}, \mathcal{A}_a, \Sigma_a, \Sigma_{ab}, \mathcal{E}_a, \mathcal{E}_{ab},$$
$$\mathcal{H}, \mathcal{H}_a . \mathcal{H}_{ab}, \alpha_a, a_a, \xi, \phi, \zeta_{ab}, \Omega, \Omega_a, Q, Q_a, \Pi_a,$$
$$\mu_{(F)_a}, \mathcal{V}_{(F)}, \mathcal{V}_{(F)_a}, \mathfrak{E}, \mathfrak{E}_a, \mathfrak{B}_a, \mathcal{B}_a, \rho, \mathfrak{J}, \mathfrak{J}_a\}.$$

To obtain the equations governing these variables to first order, we first rewrite the equations from Sec. 2 in terms of the variables in $S^{(1)}$ by using some commutation relations between the derivatives in the $1+1+2$ formalism. After removing terms which are of second order or higher in the variables in $S^{(1)}$, we arrive at a large set of coupled linear partial differential equations. To simplify the analysis of this system, we now perform a harmonic decomposition, whereby the system can be reduced to a closed set of ordinary differential equations in time.

6. Harmonic Decomposition

When performing the aforementioned harmonic decomposition, we will follow Ref. 7 and make use of the eigenfunctions $P^{k_\parallel}$ and $Q^{k_\perp}$, which are defined as functions satisfying

$$\hat{\Delta} P^{k_\parallel} = -\frac{k_\parallel^2}{a_1^2} P^{k_\parallel}, \quad \delta_a P^{k_\parallel} = 0, \quad \dot{P}^{k_\parallel} = 0,$$
$$\delta^2 Q^{k_\perp} = -\frac{k_\perp^2}{a_2^2} Q^{k_\perp}, \quad \hat{Q}^{k_\perp} = 0, \quad \dot{Q}^{k_\perp} = 0,$$

where $\hat{\Delta} \equiv n^a \nabla_a n^b \nabla_b$, and $\delta^2 = \delta_a \delta^a$, whilst $k_\parallel$ and $k_\perp$ are comoving dimensionless wave numbers along n^a and in the 2-sheets respectively. Using these harmonics, we may also define even and odd vector and tensor harmonics

$$Q^{k_\perp}{}_a = a_2 \delta_a Q^{k_\perp}, \qquad Q^{k_\perp}{}_{ab} = a_2^2 \delta_{\{a} \delta_{b\}} Q^{k_\perp},$$
$$\overline{Q}^{k_\perp}{}_a = a_2 \epsilon_{ab} \delta^b Q^{k_\perp}, \quad \overline{Q}^{k_\perp}{}_{ab} = a_2^2 \epsilon_{c\{a} \delta^c \delta_{b\}} Q^{k_\perp},$$

where the odd harmonics are denoted with bars. Equipped with these harmonics, we decompose scalars Ψ, 2-vectors Ψ_a, and PSTF 2-tensors Ψ_{ab} as

$$\Psi = \sum_{k_\parallel, k_\perp} \Psi^S_{k_\parallel k_\perp} P^{k_\parallel} Q^{k_\perp},$$

[b]Instead of using the $\delta_a G$ gradients to describe perturbations of a variable G that is non-zero on the background, we could have defined our perturbations using the spatial $\hat{G}$ derivatives. However, after performing a harmonic decomposition, it can be shown that $\hat{G}$ can be written in terms of $\delta_a G$ and the vorticity.[8]

$$\Psi_a = \sum_{k_\parallel, k_\perp} P^{k_\parallel} \left(\Psi^V_{k_\parallel k_\perp} Q^{k_\perp}{}_a + \overline{\Psi}^V_{k_\parallel k_\perp} \overline{Q}^{k_\perp}{}_a \right),$$

$$\Psi_{ab} = \sum_{k_\parallel, k_\perp} P^{k_\parallel} \left(\Psi^T_{k_\parallel k_\perp} Q^{k_\perp}{}_{ab} + \overline{\Psi}^T_{k_\parallel k_\perp} \overline{Q}^{k_\perp}{}_{ab} \right),$$

where the coefficients $\{\Psi^{S,V,T}_{k_\parallel k_\perp}, \overline{\Psi}^{V,T}_{k_\parallel k_\perp}\}$ now only depend on time. In the following we will omit the wave number indices on these coefficients when there is no risk of ambiguity.

7. Final System

After performing the harmonic decompositions described in the previous section, the first order linearized partial differential equations reduce to a set of constraints and ordinary differential equations in time for the harmonic coefficients. On using the constraints to solve for some of the coefficients in terms of the others, the system can be reduced significantly. In doing so, we also use the remaining freedom of specifying the dyad $\{u^a, n^a\}$ to first order to set a_a, $\mathcal{A}_a$, and $\mathcal{A}$ to zero. We are then left with two separate sectors, one so-called odd sector and one even sector, which decouple from each other. The odd sector consists of the coefficients

$$\{\Omega^S, \overline{\mathcal{E}}^T, \mathcal{H}^T, \overline{\mathcal{V}}_{(F)}{}^V, \mathfrak{E}^S, \mathfrak{E}^V, \overline{\mathfrak{B}}^V\},$$

whilst the even sector contains the coefficients

$$\{\overline{\Omega}^V, \overline{\mathcal{H}}^T, \mathcal{E}^T, \Sigma^T, \mu_{(F)}{}^V, \mathcal{V}_{(F)}{}^S, \mathcal{V}_{(F)}{}^V, \mathcal{B}^V, \overline{\mathfrak{E}}^V\}.$$

These sectors form closed sets in the sense that we have evolution equations for all of the coefficients above. As for the remaining coefficients, these can be given algebraically in terms of the ones above.

7.1. *Odd Sector*

The coefficients in the odd sector are found to satisfy the equations

$$\begin{aligned}
\dot{\Omega}^S &= \tilde{F}\Omega^S, \\
\dot{\overline{\mathcal{E}}}^T &= -\frac{3}{2}(F + \Sigma D)\,\overline{\mathcal{E}}^T + \frac{ik_\parallel}{a_1}(1 - D)\,\mathcal{H}^T \\
&\quad + P\Omega^S + \frac{1}{a_2}(1 + D)\left(\mu_{(F)}\overline{\mathcal{V}}_{(F)}{}^V + \mathfrak{B}\mathfrak{E}^V\right), \\
\dot{\mathcal{H}}^T &= \left(\frac{3ia_1\Pi}{k_\parallel k_\perp^2}\left(\Sigma + \frac{\Theta}{3}\right) + S\right)\Omega^S + \frac{1}{a_2}\mathfrak{B}\overline{\mathfrak{B}}^V \\
&\quad + \frac{ia_1}{a_2 k_\parallel B}\left(\left(\Sigma + \frac{\Theta}{3}\right)\left(3\Pi + \frac{k_\perp^2}{a_2^2}\right) - \frac{k_\parallel^2}{a_1^2}\tilde{F}\right)\left(\mu_{(F)}\overline{\mathcal{V}}_{(F)}{}^V + \mathfrak{B}\mathfrak{E}^V\right) \\
&\quad + \frac{ia_1}{2k_\parallel}\left(\frac{2k_\parallel^2}{a_1^2} - CB + 9\Sigma E + 3\Pi\right)\overline{\mathcal{E}}^T - \frac{3}{2}(2E + F)\,\mathcal{H}^T,
\end{aligned}$$

$$
\begin{aligned}
\dot{\overline{\mathcal{V}}}^{V}_{(F)} &= \frac{\mathfrak{B}}{\eta\mu_{(F)}}\left(\mathfrak{E}^{V} - \mathfrak{B}\overline{\mathcal{V}}_{(F)}{}^{V}\right) + \frac{1}{2}\left(\Sigma - \frac{2\Theta}{3}\right)\overline{\mathcal{V}}_{(F)}{}^{V},\\
\dot{\mathfrak{E}}^{S} &= \frac{k_{\perp}^{2}}{a_2}\overline{\mathfrak{B}}^{V} - \frac{2ia_1}{k_{\parallel}}\mathfrak{B}\left(\Sigma + \frac{\Theta}{3}\right)\Omega^{S} + \left(\Sigma - \frac{2\Theta}{3} - \frac{1}{\eta}\right)\mathfrak{E}^{S}, \qquad (19)\\
\dot{\mathfrak{E}}^{V} &= \frac{ik_{\parallel}}{a_1}\overline{\mathfrak{B}}^{V} - \frac{1}{2}\left(\Sigma + \frac{4\Theta}{3}\right)\mathfrak{E}^{V} - \frac{1}{\eta}\left(\mathfrak{E}^{V} - \mathfrak{B}\overline{\mathcal{V}}_{(F)}{}^{V}\right) - \frac{2a_2\dot{\mathfrak{B}}}{k_{\perp}^{2}}\Omega^{S}, \qquad (20)\\
\dot{\overline{\mathfrak{B}}}^{V} &= \left(\frac{ik_{\parallel}}{a_1}\mathfrak{E}^{V} - \frac{1}{a_2}\mathfrak{E}^{S}\right) - \frac{1}{2}\left(\Sigma + \frac{4\Theta}{3}\right)\overline{\mathfrak{B}}^{V}\\
&\quad - \frac{ia_1}{k_{\parallel}}\mathfrak{B}\left[\frac{2\left(\mathcal{R}a_2^{2} - k_{\perp}^{2}\right)}{a_2 B}\left(\frac{3\Sigma}{2}\overline{\mathcal{E}}^{T} + \frac{ik_{\parallel}}{a_1}\mathcal{H}^{T}\right)\right.\\
&\quad - \frac{2}{B}\left(3\Sigma\left(\Sigma + \frac{\Theta}{3}\right) + \frac{2k_{\parallel}^{2}}{a_1^{2}}\right)\left(\mu_{(F)}\overline{\mathcal{V}}_{(F)}{}^{V} + \mathfrak{B}\mathfrak{E}^{V}\right)\\
&\quad \left. - \frac{a_2}{k_{\perp}^{2}}\left(\mu + p - \frac{\Pi}{2} - 3\mathcal{E} + 2N + 3\Sigma\left(\Sigma + \frac{\Theta}{3}\right) + \frac{2k_{\parallel}^{2}}{a_1^{2}}\right)\Omega^{S}\right], \qquad (21)
\end{aligned}
$$

where we have defined the following auxiliary variables

$$
\begin{aligned}
\tilde{k}^{2} &= \frac{k_{\perp}^{2}}{a_2^{2}} + \frac{2k_{\parallel}^{2}}{a_1^{2}},\\
B &= \tilde{k}^{2} + \frac{9}{2}\Sigma^{2} + 3\left(\mathcal{E} + \frac{1}{2}\Pi\right),\\
CB &= \Sigma\left(\Theta - \frac{3\Sigma}{2}\right) - \frac{k_{\perp}^{2}}{a_2^{2}},\\
F &= \Sigma + \frac{2\Theta}{3},\\
\tilde{F} &= \Sigma - \frac{2\Theta}{3},\\
D &= C + \frac{\mu + p - 2\Pi}{B},\\
P &= \frac{2}{k_{\perp}^{2}B}\left(\mu + p - \frac{1}{2}\Pi\right)\left(B + 3\mathcal{E} + \mu + p - \frac{1}{2}\Pi + \mathcal{R} - \frac{k_{\perp}^{2}}{a_2^{2}}\right),\\
S &= \frac{2ia_1}{3k_{\parallel}k_{\perp}^{2}B}\left(\mu + p - \frac{1}{2}\Pi\right)\left(3\Sigma\left(\frac{k_{\perp}^{2}}{a_2^{2}} - \frac{k_{\parallel}^{2}}{a_1^{2}} + 3\Pi\right) + \Theta\left(\tilde{k}^{2} + 3\Pi\right)\right),\\
EB &= \frac{\Sigma}{2}\left(CB - \mathcal{E} - \frac{5}{2}\Pi\right) + \frac{\Theta}{3}\left(\mathcal{E} - \frac{1}{2}\Pi\right),\\
N &= \left(\mu + p - \frac{1}{2}\Pi\right)\left(1 + \frac{2}{a_2^{2}B}\left(\mathcal{R}a_2^{2} - k_{\perp}^{2}\right)\right).
\end{aligned}
$$

7.2. *Even Sector*

As for the even sector, its coefficients satisfy the equations

$$
\begin{aligned}
\dot{\mu}^V_{(F)} =& \left(\frac{\Sigma}{2} - \frac{4\Theta}{3} - \frac{3\Sigma\mu_{(F)}}{2B}\right){\mu_{(F)}}^V - \frac{3\Sigma\mu_{(F)}\mathfrak{B}}{2B}\mathcal{B}^V \\
&+ \frac{ia_1\mu_{(F)}}{2a_2k_\parallel}\left(\left(1 - \frac{3\Sigma}{B}\left(\Sigma + \frac{\Theta}{3}\right)\right)\mu_{(F)} - \frac{2k_\parallel^2}{a_1^2}\right){\mathcal{V}_{(F)}}^S \\
&- \mu_{(F)}\left(\left(1 + \frac{3\Sigma}{2B}\left(\Sigma - \frac{2\Theta}{3}\right)\right)\mu_{(F)} - \frac{k_\perp^2}{a_2^2}\right){\mathcal{V}_{(F)}}^V \\
&+ \frac{3a_2\Sigma\mu_{(F)}}{2}(C-1)\mathcal{E}^T - \frac{a_2B\mu_{(F)}}{2}(C-1)\Sigma^T \\
&- \frac{ia_2k_\parallel\mu_{(F)}}{2a_1}(J-2)\overline{\mathcal{H}}^T - \frac{ia_1\mu_{(F)}}{k_\parallel}\left(\frac{G}{B} + \frac{2k_\parallel^2}{a_1^2}\right)\overline{\Omega}^V \\
&+ \frac{\mathfrak{B}\mu_{(F)}}{2B}\left(2B + 3\Sigma\left(\Sigma - \frac{2\Theta}{3}\right)\right)\overline{\mathfrak{E}}^V, \\
\dot{\mathcal{B}}^V =& \frac{ia_1\mathfrak{B}\mu_{(F)}}{a_2k_\parallel B}\left(B - 3\Sigma\left(\Sigma + \frac{\Theta}{3}\right)\right){\mathcal{V}_{(F)}}^S \\
&- \frac{3\mathfrak{B}\Sigma}{B}{\mu_{(F)}}^V + 3a_2\mathfrak{B}\Sigma C\mathcal{E}^T - a_2\mathfrak{B}BC\Sigma^T - \frac{ia_2k_\parallel\mathfrak{B}J}{a_1}\overline{\mathcal{H}}^T \\
&- \frac{2ia_1\mathfrak{B}G}{k_\parallel B}\overline{\Omega}^V + \left(\frac{3\Sigma\mathfrak{B}^2}{B}\left(\Sigma - \frac{2\Theta}{3}\right) - \frac{k_\perp^2}{a_2^2}\right)\overline{\mathfrak{E}}^V \\
&+ \frac{3}{2}\left(\Sigma - \frac{2\Theta}{3} - \frac{2\Sigma\mathfrak{B}^2}{B}\right)\mathcal{B}^V - \frac{3\mathfrak{B}\Sigma\mu_{(F)}}{B}\left(\Sigma - \frac{2\Theta}{3}\right){\mathcal{V}_{(F)}}^V, \\
\dot{\overline{\mathfrak{E}}}^V =& \frac{2a_2^2k_\parallel^2\mathfrak{B}}{a_1^2k_\perp^2B}{\mu_{(F)}}^V + \frac{2ia_2k_\parallel\mathfrak{B}\mu_{(F)}}{a_1k_\perp^2B}\left(\Sigma + \frac{\Theta}{3}\right){\mathcal{V}_{(F)}}^S \\
&+ \left(\frac{2a_2^2k_\parallel^2\mu_{(F)}}{a_1^2k_\perp^2B}\left(\Sigma - \frac{2\Theta}{3}\right) - \frac{1}{\eta}\right)\mathfrak{B}{\mathcal{V}_{(F)}}^V \\
&+ \left(\frac{\mathcal{R}a_2^2 - k_\perp^2}{a_2B} - \frac{a_2^3\Sigma L}{k_\perp^2}\right)\mathfrak{B}\mathcal{E}^T + \frac{a_2^3\mathfrak{B}BL}{3k_\perp^2}\Sigma^T + \frac{2ia_2^3k_\parallel\mathfrak{B}L}{3a_1k_\perp^2}\overline{\mathcal{H}}^T \\
&+ \frac{4ia_2^2k_\parallel\mathfrak{B}}{a_1k_\perp^2B}(\mu + p + \Pi)\left(\Sigma + \frac{\Theta}{3}\right)\overline{\Omega}^V \\
&- \left(\frac{\Sigma}{2} + \frac{2\Theta}{3} + \frac{1}{\eta} + \frac{2a_2^2k_\parallel^2\mathfrak{B}^2}{a_1^2k_\perp^2B}\left(\Sigma - \frac{2\Theta}{3}\right)\right)\overline{\mathfrak{E}}^V \\
&+ \left(1 + \frac{a_2^2k_\parallel^2}{a_1^2k_\perp^2}\left(1 + \frac{2\mathfrak{B}^2}{B}\right)\right)\mathcal{B}^V, \\
\dot{\overline{\Omega}}^V =& -\left(\frac{2\Theta}{3} + \frac{\Sigma}{2}\right)\overline{\Omega}^V,
\end{aligned}
\tag{22}
\tag{23}
$$

$$
\begin{aligned}
\dot{\mathcal{V}}^S_{(F)} =& -\left(\Sigma+\frac{\Theta}{3}\right){\mathcal{V}_{(F)}}^S,\\
\dot{\mathcal{V}}^V_{(F)} =& \frac{1}{2}\left(\Sigma-\frac{2\Theta}{3}-\frac{2\mathfrak{B}^2}{\eta\mu_{(F)}}\right){\mathcal{V}_{(F)}}^V-\frac{\mathfrak{B}}{\eta\mu_{(F)}}\overline{\mathfrak{E}}^V,\\
\dot{\Sigma}^T =& \left(\Sigma-\frac{2\Theta}{3}\right)\Sigma^T-\mathcal{E}^T,\\
\dot{\overline{\mathcal{H}}}^T =& -\frac{ik_\parallel}{a_1a_2B}\left(1-\frac{2a_2^2\mathfrak{B}^2}{k_\perp^2}\right)\left({\mu_{(F)}}^V+\mathfrak{B}\mathcal{B}^V\right)+\frac{ia_2k_\parallel}{a_1k_\perp^2}\mathfrak{B}\mathcal{B}^V\\
&+\frac{2}{a_2B}\left(\mu+p+\Pi\right)\left(\Sigma+\frac{\Theta}{3}\right)\left(1-\frac{2a_2^2\mathfrak{B}^2}{k_\perp^2}\right)\overline{\Omega}^V\\
&+\frac{1}{a_2^2B}\left(1-\frac{2a_2^2\mathfrak{B}^2}{k_\perp^2}\right)\left(\Sigma+\frac{\Theta}{3}\right)\mu_{(F)}{\mathcal{V}_{(F)}}^S\\
&-\frac{ik_\parallel}{a_1a_2B}\left(1-\frac{2a_2^2\mathfrak{B}^2}{k_\perp^2}\right)\left(\Sigma-\frac{2\Theta}{3}\right)\left(\mu_{(F)}{\mathcal{V}_{(F)}}^V-\mathfrak{B}\overline{\mathfrak{E}}^V\right)\\
&+\frac{ia_1}{k_\parallel}\left(\frac{a_2^2\mathfrak{B}^2BL}{3k_\perp^2}+\frac{3\Pi}{2}\left(\Sigma+\frac{\Theta}{3}\right)\right)\Sigma^T\\
&-\frac{ia_1}{k_\parallel}\left(\frac{3\Pi}{2}+\left(1-C\right)\frac{k_\parallel^2}{a_1^2}-\frac{\mathfrak{B}^2}{a_2}\left(\frac{\mathcal{R}a_2^2-k_\perp^2}{a_2B}-\frac{a_2^3\Sigma L}{k_\perp^2}\right)\right)\mathcal{E}^T\\
&-\left(\frac{2a_2^2\mathfrak{B}^2L}{3k_\perp^2}+\frac{3}{2}\left(F+\frac{M}{B}\right)\right)\overline{\mathcal{H}}^T,\\
\dot{\mathcal{E}}^T =& -\frac{3}{2}\left(F+\Sigma C\right)\mathcal{E}^T+\frac{ik_\parallel}{2a_1}\left(J-2\right)\overline{\mathcal{H}}^T\\
&-\frac{1}{2}\left(\mu+p-2\Pi\right)\Sigma^T+\frac{3\Sigma}{2a_2B}\left({\mu_{(F)}}^V+\mathfrak{B}\mathcal{B}^V\right)+\frac{ia_1G}{a_2k_\parallel B}\overline{\Omega}^V\\
&-\frac{ia_1}{2a_2^2k_\parallel B}\left(B-3\Sigma\left(\Sigma+\frac{\Theta}{3}\right)\right)\mu_{(F)}{\mathcal{V}_{(F)}}^S\\
&-\frac{1}{a_2B}\left(B-\frac{3\Sigma}{2}\left(\Sigma-\frac{2\Theta}{3}\right)\right)\left(\mu_{(F)}{\mathcal{V}_{(F)}}^V-\mathfrak{B}\overline{\mathfrak{E}}^V\right),
\end{aligned}
$$

where we have introduced the additional auxiliary variables

$$
\begin{aligned}
M =& \ 2\left(\mathcal{E}+\frac{\Pi}{2}\right)\left(\Sigma+\frac{\Theta}{3}\right)+\frac{\Sigma}{a_2^2}\left(\mathcal{R}a_2^2-k_\perp^2\right),\\
JB =& \ \frac{a_1^2k_\perp^2}{a_2^2k_\parallel^2}\left(\mathcal{R}-\tilde{k}^2\right)-3\Sigma\left(\Sigma-\frac{2\Theta}{3}\right),\\
G =& \ \left(\mu+p+\Pi\right)\left(\mathcal{R}-\tilde{k}^2\right),\\
LB =& \ 3\Sigma\left(\frac{k_\perp^2}{a_2^2}-\frac{k_\parallel^2}{a_1^2}\right)+\Theta\tilde{k}^2.
\end{aligned}
$$

8. Discussion and Outlook

From Eqs. (19)-(23), which describe the evolution of the electromagnetic fields, we see that mechanisms for generating magnetic fields from perturbations in variables related to gravitational effects, such as the magnetic part of the Weyl tensor, indeed seem to exist. These mechanisms can be explicitly seen when solving the system numerically.[12] However, the mechanisms depend fundamentally on $\mathfrak{B}$ being non-zero, since the electromagnetic evolution equations decouple completely from the other first order equations when $\mathfrak{B} = 0$. Since a background anisotropy is needed to have a non-zero B-field, these mechanisms are therefore crucially dependent on the LRS nature of the zeroth order solution, at least to first order in the perturbations.

As we were, in the end, able to observe mechanisms for generating magnetic field perturbations, an interesting question is whether or not these results can be used to shed some light on the origin of the observed large scale cosmic magnetic fields. However, determining the viability of these mechanisms as candidates for amplifying primordial magnetic seeds would require a deeper investigation regarding the initial conditions and the primordial seed spectra, and is therefore left for future work.

References

1. D. L. Wiltshire, P. R. Smale, T. Mattsson and R. Watkins, Hubble flow variance and the cosmic rest frame, *Phys. Rev. D* **88**, p. 083529 (2013).
2. R.-G. Cai and Z.-L. Tuo, Direction Dependence of the Deceleration Parameter, *JCAP* **02**, p. 004 (2012).
3. G. Betschart, C. Zunckel, P. Dunsby and M. Marklund, Primordial magnetic seed field amplification by gravitational waves, *Phys. Rev. D* **72**, p. 123514 (2005).
4. C. Zunckel, G. Betschart, P. K. S. Dunsby and M. Marklund, On inhomogeneous magnetic seed fields and gravitational waves within the MHD limit, *Phys. Rev. D* **73**, p. 103509 (2006).
5. B. Mongwane, P. K. S. Dunsby and B. Osano, Cosmic Electromagnetic Fields due to Perturbations in the Gravitational Field, *Phys. Rev. D* **86**, p. 083533 (2012).
6. Z. Keresztes, M. Forsberg, M. Bradley, P. K. Dunsby and L. Gergely, Gravitational, shear and matter waves in Kantowski-Sachs cosmologies, *JCAP* **11**, p. 042 (2015).
7. M. Bradley, M. Forsberg and Z. Keresztes, Gravitational Waves in Locally Rotationally Symmetric (LRS) Class II Cosmologies, *Universe* **3**, p. 69 (2017).
8. R. Törnkvist and M. Bradley, General Perfect Fluid Perturbations of Homogeneous and Orthogonal Locally Rotationally Symmetric Class II Cosmologies, *Phys. Rev. D* **100**, p. 124043 (2019).
9. P. Semrén, Dissipative perturbations on LRS class II cosmologies using the 1+1+2 covariant split of spacetime, Master's thesis, Umeå University (2020).
10. G. Betschart and C. A. Clarkson, Scalar and electromagnetic perturbations on LRS class II space-times, *Class. Quant. Grav.* **21**, 5587 (2004).
11. M. Marklund and C. Clarkson, The General relativistic MHD dynamo, *Mon. Not. Roy. Astron. Soc.* **358**, p. 892 (2005).
12. P. Semrén, On the Interaction Between Electromagnetic, Gravitational, and Plasma Related Perturbations on LRS Class II Spacetimes, Master's thesis, Umeå University (2021).

13. M. Marklund, P. K. S. Dunsby, M. Servin, G. Betschart and C. Tsagas, Charged multifluids in general relativity, *Class. Quant. Grav.* **20**, 1823 (2003).
14. C. Clarkson, A Covariant approach for perturbations of rotationally symmetric space-times, *Phys. Rev. D* **76**, p. 104034 (2007).
15. J. M. Stewart and M. Walker, Perturbations of space-times in general relativity, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **341**, 49 (1974).

Condition for expansion-collapse duality between Einstein and Jordan frames

Dipayan Mukherjee*, H. K. Jassal† and Kinjalk Lochan‡
Indian Institute of Science Education and Research Mohali, SAS Nagar, Mohali-140306, Punjab, India
*E-mails: *ph17011@iisermohali.ac.in, †hkjassal@iisermohali.ac.in, ‡kinjalk@iisermohali.ac.in*

Quintessence fields, or scalar fields minimally coupled to gravity, are considered to be viable candidates for dark energy. It is well-known that some classes of modified theories of gravity can be recast as Einstein's general relativity with a minimally coupled scalar field, through conformal transformation. The 'universe' described by the initial and the final actions are referred to as the Jordan and Einstein frames, respectively. Although these conformally connected frames are mathematically equivalent, the equations of motion in these two frames may describe drastically different physical scenarios. Depending upon the choice of the Jordan frame action (or equivalently the scalar field potential in Einstein frame) it is possible that while the Einstein frame expands, the Jordan frame collapses. We classify quintessence models in the Einstein frame that are dual to $f(R)$ gravity theories in the Jordan frame, based on whether they possess such expansion-collapse duality. We derive a general condition for expansion-collapse duality, applicable to quintessence models with arbitrary time-dependent equations of state. The condition also takes into account the presence of other components in the Einstein frame universe. Such expansion-collapse duality between these conformally connected frames can lead to an effective description of a collapsing universe in terms of an expanding one, which is a topic of further exploration.

Keywords: $f(R)$ theories; quintessence; dark energy; Einstein and Jordan frames

1. Introduction

The quintessence model, or scalar fields minimally coupled to gravity, is a widely used implementation of dark energy in Einstein's gravity.[1–5] As an alternative to the cosmological constant model, quintessence fields (φ) can provide a simple dynamical description of the late-time acceleration of the universe. The equation of state of quintessence (w_φ) can violate the strong energy condition if the potential term ($V(\varphi)$) is set to be sufficiently flat. In general, w_φ can vary with time. One can introduce the time-dependence in w_φ by postulating a parameterization of the form $w_\varphi(a; w_0, w_1, \ldots)$,[6–10] where a is the scale factor and $w_0, w_1 \ldots$ are constant parameters of the model, usually constrained from different cosmological observations.

It is also well-known that some classes of modified theories of gravity can provide a dual description of minimally coupled scalar fields in Einstein gravity, in a conformally connected frame.[11–14] $f(R)$ theory is a simple modification of Einstein gravity, where the Ricci scalar (R) in the Einstein–Hilbert action is replaced with an arbitrary function of the Ricci scalar ($f(R)$) (for reviews, see [11, 12]). The

'universe' governed by the $f(R)$ theory is referred to as the *Jordan frame universe*. A conformal transformation allows for the Jordan frame action to be written as the Einstein–Hilbert action with a quintessence field (for reviews, see [11, 14, 15]). The 'universe' described by the latter action is referred to as the *Einstein frame universe*. Thus, one may consider that dark energy, as implemented by a quintessence field in Einstein gravity, has an equivalent description in terms of a pure $f(R)$ gravity theory in the conformally connected Jordan frame.

Although Einstein and Jordan frames are mathematically equivalent, the field equations in these two frames may describe drastically different universes. In this paper, we are interested in quintessence–$f(R)$ pairs for which, while the Einstein frame universe expands, the Jordan frame universe collapses. We derive a simple condition for such expansion-collapse duality for quintessence models with arbitrary, time-varying equation of state parameters. The condition is general enough to accommodate any other components in the Einstein frame, such as dust or radiation. As an example, we study the quintessence model with logarithmic equation of state[6, 10, 16] ($w_\varphi(a) = w_0 - w' \ln a$) and discuss the possibility of the expansion-collapse duality in the presence of non-relativistic matter and spatial curvature component.

A quintessence model–$f(R)$ theory pair with the expansion-collapse duality essentially provides a description of an expanding universe in terms of a collapsing one. This map between expanding and collapsing geometries may lead to potentially interesting results; for example, introducing perturbation in the background of FRW metric one can study how the evolution of perturbations in the expanding frame corresponds to the perturbations in the collapsing one. This will be pursued elsewhere.

This paper is organized as follows. In Sec. 2 we start with a brief review of the Jordan frame–Einstein frame correspondence in the context of $f(R)$ gravity. Our main result, the expansion-collapse duality condition is presented in Sec. 3. In Sec. 4 we conclude with a summary and discussion.

2. $f(R)$ gravity and quintessence: Einstein and Jordan frames

Let us consider that the Jordan frame universe is governed by the action of an $f(R)$ theory,

$$\mathcal{S}_J = \frac{1}{2\kappa^2} \int \mathrm{d}^4x \sqrt{-g} f(R), \tag{1}$$

where $\kappa = 8\pi G$, R is the Ricci scalar in Jordan frame, $f(R)$ is an arbitrary function of R. The Jordan frame metric $g_{\mu\nu}$ is the spatially flat FRW metric, i.e., $g_{\mu\nu} \equiv \mathrm{diag}(-1, a^2, a^2, a^2)$, and $a(t)$ is the scale factor in the Jordan frame (see [11, 12] for general reviews on $f(R)$ theories). One can then define a new metric using a conformal transformation,

$$\tilde{g}_{\mu\nu} = \Omega^2(x) g_{\mu\nu}. \tag{2}$$

If the conformal parameter Ω^2 is taken to be $\Omega^2 = \partial f/\partial R = F(R)$, then the same Jordan frame action can be recast in terms of the new metric $\tilde{g}_{\mu\nu}$ as,

$$\mathcal{S}_E = \int \mathrm{d}^4x\sqrt{-\tilde{g}}\frac{1}{2\kappa^2}\tilde{R} - \int \mathrm{d}^4x\sqrt{-\tilde{g}}\left[\frac{1}{2}\tilde{g}^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi + V(\varphi)\right], \tag{3}$$

where $\tilde{R}$ is the Ricci scalar in Einstein frame[17] and

$$\varphi = \frac{\sqrt{6}}{2\kappa}\ln F, \tag{4a}$$

$$V = \frac{1}{2\kappa^2}\frac{FR - f}{F^2}. \tag{4b}$$

The first part of the action in (3) is the Einstein–Hilbert action with respect to the metric $\tilde{g}_{\mu\nu}$, the second part is the action of a minimally coupled scalar field φ, with potential V. Equation (3) is the action in the Einstein frame universe.

On the other hand, if we consider the Einstein frame to be the physical frame, then the scalar field action in (3) can play the role of the quintessence model of dark energy. The equation of state parameter of φ can be computed as[5,9]

$$w_\varphi = \frac{P_\varphi}{\rho_\varphi} = \frac{\frac{1}{2}\dot{\varphi}^2 - V}{\frac{1}{2}\dot{\varphi}^2 + V}, \tag{5}$$

where P_φ and ρ_φ are the pressure and energy density associated with the quintessence field φ. As we can see from (5), for sufficiently flat potential ($V \gg \dot{\varphi}^2$) w_φ can become $w_\varphi < -1/3$, allowing for the scalar field to drive the acceleration in the Einstein frame.

As a simple example, let us consider a quintessence model with constant equation of state parameter w_φ. The quintessence potential in this case takes the form[16]

$$V(\varphi) = \frac{1}{2}(1-w)\rho_0 \exp\left(-\sqrt{3\kappa^2(1+w)}(\varphi)\right). \tag{6}$$

One can then solve the differential equation (4b), using (4a) and (6), and reconstruct the correspond $f(R)$ action in the Jordan frame as (see [18] for details)

$$f(R) = C_1R - A_1C_1^b \text{ or} \tag{7a}$$

$$f(R) = \frac{b-1}{b(A_1b)^{\frac{1}{b-1}}}R^{\frac{b}{b-1}}, \tag{7b}$$

where, $A_1 = \kappa^2\rho_0(1-w)$, $b = -\frac{3}{\sqrt{2}}\sqrt{1+w} + 2$.

In general, the equation of state of the quintessence field can be a function of time. One usually starts with a parameterized time-dependent ansatz for w_φ and then constrains the parameters of the model from cosmological observations (for different parameterizations of equation of state see [6–10]). In this work, we consider the logarithmic parameterization as an example of time-dependent equation of state ([6, 16]), given by

$$w_\varphi(\tilde{a}) = w_0 - w'\ln\tilde{a}, \tag{8}$$

where $\tilde{a}$ is the scale factor in Einstein frame, (w_0, w') are two constant parameters of the model. The values of (w_0, w') are constrained from different cosmological observations in [10, 16]. The quintessence field Φ for this model can be solved as a function of the scale factor as,[16]

$$\varphi(\tilde{a}) = \Phi(\tilde{a}) - \Phi_0 = \frac{2}{\sqrt{3}} \frac{1}{\kappa w'} \left[(1 + w_0)^{\frac{3}{2}} - (1 + w_0 - w' \ln \tilde{a})^{\frac{3}{2}} \right], \tag{9}$$

where Φ_0 is the value of the scalar field at current Einstein frame, i.e. at $\tilde{a} = 1$. We take $w' < 0$ to ensure that the field remains real valued in the late time limit of the Einstein frame $\tilde{a} \to \infty$. In the paper,[18] we used analytical approximations to reconstruct $F(R)$ corresponding to this quintessence model in two parts. First, we derived a perturbative solution in the small R limit, which is applicable in the near future of the Einstein frame. Second, we also obtained a non-perturbative, asymptotic form of $F(R)$ in the late time limit of the Einstein frame. However, as we will see later, the condition for the expansion-collapse duality does not directly depend on the form of the $f(R)$ function. We will hence not discuss the reconstruction of the Jordan frame action here. More details in this regard can be found in [18].

Having summarized the key features of the Einstein-Jordan frame correspondence, we now discuss the main results in the following section.

3. Expansion-collapse duality

A pure $f(R)$ action in the Jordan frame (1) gives rise to only a quintessence field in Einstein gravity in the Einstein frame (3). In order to incorporate the presence of any other matter components in the Einstein frame, the Jordan frame action (3) must be modified. Moreover, since we consider the Einstein frame to be the physical frame, we want the field equations in the Einstein frame to lead to the usual Friedmann equations. Essentially, we want the Einstein frame action to have the form

$$\begin{aligned} \mathcal{S}_E = & \int d^4x \sqrt{-\tilde{g}} \frac{1}{2\kappa^2} \tilde{R} - \int d^4x \sqrt{-\tilde{g}} \left[\frac{1}{2} \tilde{g}^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi + V(\varphi) \right] \\ & + \int d^4x \sqrt{-\tilde{g}} \mathcal{L}_{\mathrm{M}}(\psi_{\mathrm{M}}; \tilde{g}_{\mu\nu}), \end{aligned} \tag{10}$$

where $\mathcal{L}_{\mathrm{M}}(\psi_{\mathrm{M}}; \tilde{g}_{\mu\nu})$ is the Lagrangian density of some matter field ψ_{M}, which is independent of the quintessence field φ. In order to arrive at the Einstein frame action (10), we postulate the following Jordan frame action in the presence of matter,

$$\mathcal{S}_J = \frac{1}{2\kappa^2} \int \mathrm{d}^4x \sqrt{-g} f(R) + \int \mathrm{d}^4x \sqrt{-g} F^2(R) \mathcal{L}_{\mathrm{M}}(\psi_{\mathrm{M}}; F(R) g_{\mu\nu}). \tag{11}$$

It can be easily verified that the conformal transformation $\tilde{g}_{\mu\nu} = F g_{\mu\nu}$ gives back the desired Einstein frame action (10). It is to be noted that in the modified Jordan frame action (11), the matter field ψ_{M} is now non-minimally coupled to gravity

through the term $F(R)$. However, as we will see now, the exact form of the Jordan frame action is not required in order to predict the expansion-collapse duality.

The Einstein frame action of the form (10) leads to the Friedmann equation

$$\tilde{H}^2 = \frac{\kappa^2}{3}\left(\rho_\varphi + \sum \rho_I\right), \tag{12}$$

where $\tilde{H} = (\mathrm{d}\tilde{a}/\mathrm{d}\tilde{t})/\tilde{a}$ is the Einstein frame Hubble parameter, $\tilde{t}$ is the coordinate time in the Einstein frame, $\sum \rho_I$ denotes the summation of energy densities of all the component in the Einstein frame apart from φ. Noting that the Einstein frame metric is also a spatially flat FRW metric, $\tilde{g}_{\mu\nu} \equiv \mathrm{diag}(-1, \tilde{a}^2, \tilde{a}^2, \tilde{a}^2)$, one can relate the scale factors in the Jordan frame (a) and Einstein frame ($\tilde{a}$) as (see, for example, [11])

$$a = \frac{\tilde{a}}{\Omega} = \frac{\tilde{a}}{\sqrt{F}}. \tag{13}$$

Using (12) in (13), we find derivative of a with respect to $\tilde{a}$ to be (see [18] for details),

$$\frac{\mathrm{d}a}{\mathrm{d}\tilde{a}} = \frac{\mathrm{d}}{\mathrm{d}\tilde{a}}\left(\frac{\tilde{a}}{\sqrt{F}}\right) = F^{-1/2}\left(1 - \kappa\sqrt{1 + w_\varphi(\tilde{a})}\sqrt{\frac{\rho_\varphi}{6\tilde{H}^2}}\right). \tag{14}$$

As it can be seen from above, the sign of $\mathrm{d}a/\mathrm{d}\tilde{a}$ is determined by the term inside the parenthesis. The condition for $\mathrm{d}a/\mathrm{d}\tilde{a} < 0$, or expansion-collapse duality becomes

$$w_\varphi(\tilde{a}) > \mathcal{C}(\tilde{a}) \tag{15a}$$

$$\mathcal{C}(\tilde{a}) = \frac{6\tilde{H}^2}{\kappa^2\rho_\varphi(\tilde{a})} - 1. \tag{15b}$$

Thus, the condition of expansion-collapse duality between Einstein-Jordan frames can be written as a threshold imposed on the equation of state of the quintessence field φ, even when there are other components in the Einstein frame. The Hubble parameter ($\tilde{H}$) incorporates the contribution from all the other components present in the Einstein frame universe. Depending upon whether (15a) is satisfied, there might be a range of $\tilde{a}$, or possibly multiple ranges of $\tilde{a}$, during which the Jordan frame collapses while the Einstein frame expands. The possible 'turn around' of the Jordan frame ($\mathrm{d}a/\mathrm{d}\tilde{a} = 0$) can be determined by setting

$$w_\varphi(\tilde{a}) = \mathcal{C}(\tilde{a}), \tag{16}$$

provided such a solution does exist. Starting with the relation between Hubble parameters in the Jordan (H) and Einstein ($\tilde{H}$) frames (see [11]),

$$H = \sqrt{F}\tilde{H} - \frac{1}{2\sqrt{F}}\frac{\mathrm{d}F}{\mathrm{d}\tilde{t}}, \tag{17}$$

one can explicitly show that at the 'turn-around' of the Jordan frame, the Jordan frame Hubble parameter indeed vanishes, $H_{\text{turn-around}} = 0$.

It is also to be noted that the condition for expansion-collapse duality has a significantly simple form. Given an equation of state parameter w_φ, one only requires the knowledge of the energy densities of the components in the Einstein frame in order to check whether the condition is satisfied. The form of the Jordan frame action, even the solution of the field $\varphi(\tilde{a})$ or potential $V(\varphi)$ do not appear explicitly in the condition. Thus, as mentioned before, the reconstruction of Jordan frame action is not required in order to predict the expansion-collapse duality for a quintessence-$f(R)$ pair.

Let us now consider the example of quintessence model with logarithmic equation of state (8), where the Einstein frame universe also contains non-relativistic matter or dust. The condition (15) in this case becomes (see [18])

$$w_\varphi(\tilde{a}) > \mathcal{C}(\tilde{a}), \text{ where} \tag{18a}$$

$$w_\varphi(\tilde{a}) = w_0 - w' \ln \tilde{a}, \tag{18b}$$

$$\mathcal{C}(\tilde{a}) = 1 + 2\frac{\Omega_{m0}}{\Omega_{\varphi 0}} \exp\left(3w_0 \ln \tilde{a} - \frac{3}{2}w'(\ln \tilde{a})^2\right), \tag{18c}$$

Ω_{m0}, $\Omega_{\varphi 0}$ are the density parameters of dust and φ at $\tilde{a} = 1$. The evolution of the Jordan frame scale factor $a(\tilde{a})$ is plotted in Fig. 1(a). In the absence of dust ($\Omega_{m0} = 0$), the Jordan frame initially expands with the Einstein frame. The 'turn around' in the Jordan frame occurs at $\tilde{a} = \exp\left((w_0 - 1)/w'\right)$. In the late time limit of the Einstein frame ($\tilde{a} \to \infty$), the Jordan frame universe completely collapses ($a \to 0$). However, as the fraction of dust in the Einstein frame increases, the Jordan frame collapse staggers. For a realistic value of dust to dark energy ratio ($\Omega_{m0}/\Omega_{\varphi 0} \sim 0.3/0.7$), the Jordan frame does not collapse and the Jordan frame scale factor always increases monotonically with respect to that of the Einstein frame. In fact, it can be argued that in the presence of dust, quintessence model with observationally consistent logarithmic equation of state does not lead to the expansion-collapse duality (see [18]).

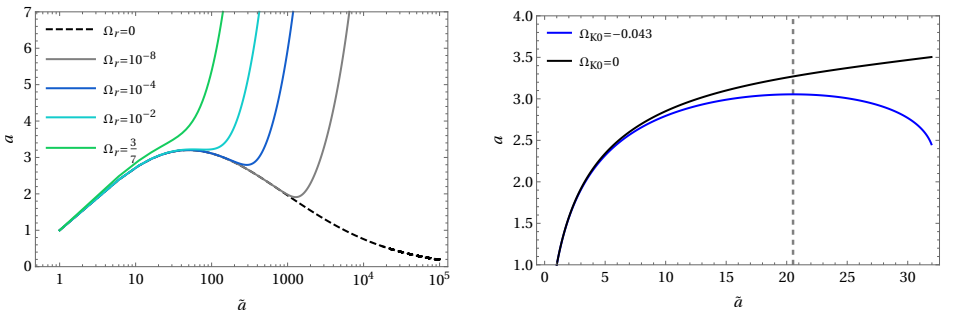
Finally, let us consider the effect of spatial curvature in the Jordan frame collapse for the same quintessence model. In the presence of positive spatial curvature and dust, the condition (15) becomes (see [18])

$$w_\varphi(\tilde{a}) > \mathcal{C}(\tilde{a}), \text{ where} \tag{19a}$$

$$w_\varphi(\tilde{a}) = w_0 - w' \ln \tilde{a}, \tag{19b}$$

$$\begin{aligned}\mathcal{C}(\tilde{a}) = {} & 1 + 2\frac{\Omega_{m0}}{\Omega_{\varphi 0}} \exp\left(3w_0 \ln \tilde{a} - \frac{3}{2}w'(\ln \tilde{a})^2\right) \\ & + 2\frac{\Omega_{K0}}{\Omega_{\varphi 0}} \exp\left((1 + 3w_0) \ln \tilde{a} - \frac{3}{2}w'(\ln \tilde{a})^2\right),\end{aligned} \tag{19c}$$

where $\Omega_K = -K/(\tilde{a}\tilde{H})^2 = \Omega_{K0}\tilde{a}^{-2}$, K is the spatial curvature parameter in the FRW metric. The expansion-collapse duality condition in this case is plotted in Fig. 2. We can see from (19c), the contribution from positive spatial curvature,[19] or negative Ω_{K0}, brings down the threshold $\mathcal{C}$, allowing for w_φ to eventually



(a) (b)

Fig. 1. The scale factor in the Jordan frame (a) is plotted as a function of the scale factor in the Einstein frame ($\tilde{a}$). The quintessence parameters are taken to be $(w_0, w') = (-0.87, -0.48)$, from [16]. (a) Jordan frame collapse in the presence of dust, without spatial curvature. $\Omega_r = \Omega_{m0}/\Omega_{\varphi 0}$ is the ratio of the density parameter of dust to that of the quintessence field, at current Einstein frame universe. In the absence of dust ($\Omega_r = 0$), the Jordan frame scale factor $a \to 0$ while the Einstein frame scale factor $\tilde{a} \to \infty$. However, in the presence of dust ($\Omega_r \neq 0$), the Jordan frame collapse eventually comes to a stop and the Jordan frame starts to expand again. For a realistic value of $\Omega_r = \Omega_{m0}/\Omega_{\varphi 0} \sim 0.3/0.7$, the Jordan frame does not go through a collapsing phase, it always expands monotonically with the Einstein frame. (b) Jordan frame collapse in the presence of dust and spatial curvature. In the presence of a small positive spatial curvature ($\Omega_{K0} = -0.043, \Omega_{m0} = 0.315,\ \Omega_{\varphi 0} = .728$), the Jordan frame collapse can be recovered (blue curve). The Jordan frame universe goes through a turn-around at $\tilde{a} \sim 20.52$ (represented by the vertical dashed line) after which the Jordan frame scale factor starts to decrease.

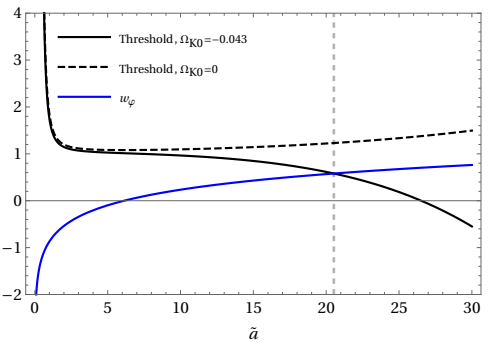


Fig. 2. The equation of state parameter w_φ and the threshold $\mathcal{C}$ (as in (19c)) are plotted as functions of the Einstein frame scale factor. The choice of parameters are: $(w_0, w') = (-0.87, -0.48)$ (from [16]), $\Omega_{m0} = 0.315$, $\Omega_{\varphi 0} = .728$, $\Omega_{K0} = -.043$ (from [19]). The 'turn-around' in the Jordan frame occurs at $w_\varphi(\tilde{a}) = \mathcal{C}(\tilde{a})$ (represented by the vertical dashed line), after which $(w_\varphi > \mathcal{C})$ the Jordan frame collapses. The plots demonstrate that in the absence of spatial curvature Ω_{K0}, w_φ (blue plot) never reaches $\mathcal{C}$ (black dashed plot), hence the Jordan frame collapse never begins. However, even a small positive spatial curvature ($\Omega_{K0} = -.043$) brings down the threshold $\mathcal{C}$ (black plot) enough for w_φ to overcome it eventually. For this case the 'turn-around' ($w_\varphi = \mathcal{C}$) occurs at $\tilde{a} \sim 20.52$, the Jordan frame starts to collapse afterwards.

overcome it. Therefore, the positive spatial curvature may recover the Jordan frame collapse in the presence of dust, at least for a finite duration. The evolution of the Jordan frame scale factor in this case is plotted in Fig. 1(b).

4. Summary and discussion

Dark energy, implemented by a quintessence field in the Einstein frame, can effectively be described by an $f(R)$ gravity theory in the Jordan frame. In this paper we classify quintessence models based on whether they lead to expansion-collapse duality between Einstein and Jordan frames, i.e., while the Einstein frame expands, the Jordan frame collapses or *vice versa*. We show that in order for the quintessence–$f(R)$ gravity pair to have such a feature, the equation of state of the quintessence must overcome a certain threshold. We derive a general condition for the expansion-collapse duality which is applicable to quintessence models with arbitrary, time-dependent equation of state parameters. The condition also holds true in the presence of any other components in the Einstein frame.

As an example, we consider the quintessence model with logarithmic equation of state parameter. We show that if the quintessence field is the sole component in the Einstein frame, this model does lead to the expansion-collapse duality, in fact the Jordan frame universe completely collapses in the late time limit of the Einstein frame. However, even the subdominant presence of dust in the Einstein frame prevents the Jordan frame collapse. If we also consider a small positive spatial curvature in the universe, the Jordan frame collapse can be recovered for a finite duration.

In general, the map between expanding and collapsing geometries at the field equation level may have potentially interesting implications on the cosmological perturbations in these frames. Further studies can reveal how perturbation in a collapsing universe corresponds to the perturbations in an expanding universe.

References

1. B. Ratra and P. J. E. Peebles, Cosmological consequences of a rolling homogeneous scalar field, *Phys. Rev. D* **37**, 3406 (Jun 1988).
2. I. Zlatev, L. Wang and P. J. Steinhardt, Quintessence, cosmic coincidence, and the cosmological constant, *Phys. Rev. Lett.* **82**, 896 (Feb 1999).
3. P. Brax and J. Martin, Robustness of quintessence, *Phys. Rev. D* **61**, p. 103502 (Apr 2000).
4. T. Barreiro, E. J. Copeland and N. J. Nunes, Quintessence arising from exponential potentials, *Phys. Rev. D* **61**, p. 127301 (May 2000).
5. S. Tsujikawa, Quintessence: A Review, *Classical and Quantum Gravity* **30**, p. 214003 (November 2013).
6. G. Efstathiou, Constraining the equation of state of the universe from distant type ia supernovae and cosmic microwave background anisotropies, *Monthly Notices of the Royal Astronomical Society* **310**, 842 (Dec 1999).
7. M. Chevallier and D. Polarski, Accelerating Universes with Scaling Dark Matter, *International Journal of Modern Physics D* **10**, 213 (January 2001).

8. H. K. Jassal, J. S. Bagla and T. Padmanabhan, Observational constraints on low redshift evolution of dark energy: How consistent are different observations?, *Physical Review D* **72** (Nov 2005).
9. E. J. Copeland, M. Sami and S. Tsujikawa, Dynamics of dark energy, *International Journal of Modern Physics D* **15**, 1753 (November 2006).
10. A. Tripathi, A. Sangwan and H. K. Jassal, Dark Energy Equation of State Parameter and Its Variation at Low Redshifts, *Journal of Cosmology and Astroparticle Physics* **2017**, 012 (June 2017).
11. A. De Felice and S. Tsujikawa, F(R) Theories, *Living Reviews in Relativity* **13** (December 2010).
12. T. P. Sotiriou and V. Faraoni, $f(R)$ Theories of gravity, *Reviews of Modern Physics* **82**, 451 (March 2010).
13. S. Capozziello and M. De Laurentis, Extended theories of gravity, *Physics Reports* **509**, 167 (Dec 2011).
14. S. Nojiri, S. Odintsov and V. Oikonomou, Modified gravity theories on a nutshell: Inflation, bounce and late-time evolution, *Physics Reports* **692**, p. 1–104 (Jun 2017).
15. V. Faraoni, Nine Years of f(R) Gravity and Cosmology, in *Accelerated Cosmic Expansion*, eds. C. Moreno González, J. E. Madriz Aguilar and L. M. Reyes Barrera (Springer International Publishing, Cham, 2014) pp. 19–32.
16. A. Sangwan, A. Mukherjee and H. K. Jassal, Reconstructing the Dark Energy Potential, *Journal of Cosmology and Astroparticle Physics* **2018**, 018 (January 2018).
17. R. M. Wald, *General Relativity*, repr. edn. (Univ. of Chicago Press, Chicago, 2009), OCLC: 554287751.
18. D. Mukherjee, H. K. Jassal and K. Lochan, $f(R)$ Dual Theories of Quintessence: Expansion-Collapse Duality (2021), `https://arxiv.org/abs/2105.10521`.
19. E. Di Valentino, A. Melchiorri and J. Silk, Planck evidence for a closed universe and a possible crisis for cosmology, *Nature Astronomy* **4**, 196 (Nov 2019).

The model of dark energy based on the quantum-mechanical uncertainty relation

Yu. V. Dumin

Sternberg Astronomical Institute (GAISh), Lomonosov Moscow State University, Universitetskii prosp. 13, Moscow, 119234, Russia
E-mail: dumin@yahoo.com

Space Research Institute (IKI), Russian Academy of Sciences Profsoyuznaya str. 84/32, Moscow, 117997, Russia

The Dark Energy became now a commonly-accepted paradigm of cosmology, but its physical essence remains absolutely unknown, and its numerical values are drastically different in the early and modern Universe. The Dark Energy is usually introduced in the contemporary literature either by postulating some additional terms in the Lagrangians or by employing the empirical equations of state. In the present work, we try to look at this problem from a more general point of view, namely, employing the quantum-mechanical uncertainty relation between the time and energy in the Mandelstam–Tamm form, which is appropriate for the long-term evolution of quantum systems. This leads us to the time-dependent effective Lambda-term, decaying as $1/t$. The corresponding cosmological model possesses a number of quite appealing features: (1) While in the standard cosmology there are a few very different expansion stages (governed by the Lambda-term, radiation, dust-like matter, and Lambda-term again), our model provides a universal description of the entire evolution of the Universe by the same "quasi-exponential" function. (2) As follows from the analysis of causal structure, the present-day cosmological horizon comprises a single domain developing from the Big Bang. Therefore, the problems of homogeneity and isotropy of the matter, the absence of topological defects, etc. should be naturally resolved. (3) At last, our model naturally explains the observed approximately flat 3D space, i.e., solution with zero curvature is formed "dynamically", starting from the arbitrary initial conditions.

Keywords: Dark Energy; Inflation; Uncertainty relation.

1. Introduction

The concept of Dark Energy (or the effective Λ-term in Einstein equations) represents now a commonly-accepted paradigm both in the cosmology of early and present-day Universe. A crucial role of the effective Λ-term in the early Universe was recognized in the early 1980s, when it was introduced as a driving force of the inflationary stage of the cosmological evolution;[1] while its significance at the present time was recognized in the very end of 1990s, when the accelerated expansion of the Universe was observed.[2] However, despite an active involvement into a huge number of theoretical works, a physical nature of the Dark Energy remains unknown. Besides, it is unclear why densities of the Dark Energy in the early and modern Universe are absolutely different.

In fact, the first inflationary models of the early Universe, discussed in the 1980s, were formulated on the quite solid physical basis steaming from the elementary-particle physics. Namely, it was assumed that the effective Λ-term was associated with a nonzero value of the potential energy of the Higgs field in its symmetry-broken state formed after a phase transition in the expanding and cooling Universe.[3] Unfortunately, a further research in this direction failed to derive a reasonable value of the Λ-term from the existing models of elementary particles, such as the electroweak or Grand Unification theories. As a result, a subsequent development of the inflationary concept of the early Universe was based mostly on the empirical Lagrangians, where some additional terms with a scalar field (the so-called inflaton) were added quite arbitrarily, without any attempts to associate them with the underlying theory of elementary particles. At last, when the accelerated expansion of the contemporary Universe was observed in the late 1990s, and it became necessary to introduce the Λ-term (or Dark Energy) into the "standard" cosmological model, this was done again only in terms of the empirical equations of state of the cosmological matter or, in a more elaborate theoretical approach, by means of the empirical Lagrangians similar to the inflationary ones.

In view of the above circumstances, there is now a clear need for a more solid conceptual basis for explanation of the Λ-term, i.e., stemming from other well-established physical laws and theories. One such approach is well known: this is the so-called Sakharov induced gravity,[4] going back to 1960s. The cornerstone of this concept is that the vacuum modes of the fundamental quantum fields (e.g., electroweak, chromodynamic, etc.) are perturbed if the space–time is curved. As a result, the effective Lagrangian acquires the additional terms, which are proportional to the various degrees of the curvature $\mathcal{R}$.[a] Thereby, the first-order term will represent the standard Einstein–Hilbert part of the Lagrangian (apart from the Λ-term); the term of zero order will be just the effective ("induced") Λ-term; and there should be also the higher-order terms of $\mathcal{R}$. About a decade later, this idea was implemented in the inflationary model by Starobinsky[5] and later evolved into a wide range of the so-called $f(\mathcal{R})$ theories. (However, it should be kept in mind that the ingredient responsible for the inflation in such theories is usually the quadratic rather than zero-order term of the curvature.) Unfortunately, since we do not have the ultimate theory of elementary particles, the exact form of the function f cannot be derived, and it is commonly specified "by hand", so that the required cosmological predictions can be obtained. From this point of view, the most of $f(\mathcal{R})$ theories are based on the same type of Lagrangians with arbitrary additional terms (but without the additional scalar fields); for a review, see Refs. 6 and 7.

The aim of the present paper is to discuss yet another conceivable universal approach to the emergence of the effective Λ-term, namely, based on the quantum-mechanical uncertainty relation between the time and energy. Let us mention that

[a]Strictly speaking, the Lagrangian can contain also a number of other invariants composed of the curvature tensor.

some qualitative ideas about a probable role of the uncertainty relation in cosmology were put forward some time ago by Coe,[8] but the corresponding quantitative treatment looked unreasonable. Here, we shall discuss another development of this idea, performed in our recent works.[9–11]

2. Basic Equations

2.1. *Uncertainty Relations*

As is known, the uncertainty relations are one of cornerstones of quantum mechanics. In the case of canonically conjugate variables (e.g., the coordinate x and momentum p_x), they can be derived in the rigorous and universal way by averaging the corresponding quantum commutators. This leads to the inequalities like $\Delta p_x \Delta x \geq \hbar/2$, and similarly for other pairs of conjugate quantities. The uncertainty relations are most widely employed in the problem of quantum measurements. However, they can be sometimes useful also in dealing with other properties of the quantum systems: a well-known example—which is considered in many textbooks—is estimating parameters of the ground state of a hydrogen atom.

Meanwhile, it should be kept in mind that the uncertainty relation between the time t and energy E is a quite exceptional case, because the time and energy are not the canonically conjugate operators in the non-relativistic quantum mechanics (instead, t is the independent variable in Schrödinger equation). So, the time–energy uncertainty relation is commonly derived by various "Gedankenexperiments"; and the resulting coefficients in the right-hand side, in general, can be different:

$$\Delta E \Delta t \geq C_{\mathrm{UR}} \hbar/2, \tag{1}$$

where C_{UR} is the numerical coefficient on the order of unity ('UR' implies here the uncertainty relation). In the standard Heisenberg's case, pertinent to the problem of measurements, $C_{\mathrm{UR}} = 1$. On the other hand, the case of the long-term evolution of a quantum system was considered by Mandelstam and Tamm[12] in 1945, and it was found that in this situation $C_{\mathrm{UR}} = \pi$. It is interesting that this work remained almost unknown for a few decades and attracted considerable attention of the scientific community only in the recent time; for more details, see Ref. 13. So, it is widely used now in such branches of physics as quantum optics, quantum information processing, etc.; and it was the aim of our recent studies[9–11] to employ it in cosmology.

2.2. *The Modified Friedmann Equation*

The starting point of our consideration is the standard Robertson–Walker cosmological metric:

$$ds^2 = c^2 dt^2 - R^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2\theta \, d\varphi^2) \right], \tag{2}$$

where t is time, r, θ, and φ are the dimensionless spherical coordinates, R is the scale factor of the Universe, c is the speed of light, and the coefficient k equals 1, 0, and –1 for the closed, flat, and open Universe.

Temporal evolution of this metric is governed by the well-known Friedmann equation (e.g., Ref. 14):

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3c^2}\,\rho - kc^2\frac{1}{R^2} + \frac{c^2}{3}\,\Lambda\,, \tag{3}$$

where H is the Hubble parameter, ρ is the energy density of matter in the Universe (both visible and dark), G and Λ are the gravitational and cosmological constants.

It is well known that Λ-term in the right-hand side of Eq. (3) can be formally attributed to the vacuum energy density $\rho_{\rm v}$:

$$\Lambda = \frac{8\pi G\rho_{\rm v}}{c^4}\,; \tag{4}$$

and then it becomes reasonable to estimate the vacuum energy from the uncertainty relation (1) with the equality sign, assuming that Δt is the total period of the cosmological evolution t. However, since the uncertainty relation should be applied to the energy rather than the energy density, it is necessary to specify the characteristic spatial volume; the most evident choice being the Planck volume $l_{\rm P}^3$ (where $l_{\rm P} = \sqrt{G\hbar/c^3}$ is the Planck length).[b]

Consequently, we arrive at the following effective Λ-term:

$$\Lambda(t) = \frac{4\pi C_{\rm UR}}{c\,l_{\rm P}}\,\frac{1}{t}\,; \tag{5}$$

and its substitution into Eq. (3) leads to the modified Friedmann equation:

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3c^2}\,\rho - kc^2\frac{1}{R^2} + \frac{4\pi C_{\rm UR}}{3\tau}\,\frac{1}{t}\,, \tag{6}$$

where $\tau = l_{\rm P}/c = \sqrt{G\hbar/c^5}$ is the Planck time.

A very unusual feature of this equation is the explicit dependence on time in the right-hand side. As far as we know, such situation was never considered in cosmology. However, as will be seen from the subsequent treatment, this time dependence does not result in any anomalous behavior.

The numerical coefficient $C_{\rm UR}$ remains essentially unknown. Since the uncertainty relation (1) was used with the equality sign, this coefficient can actually take the values from π up to infinity (but it cannot be less than π).

[b] As was pointed out by Starobinsky (private communication), yet another conceivable choice for the characteristic spatial scale might be the size of the cosmological horizon. However, as will be discussed below in Sec. 5, this option looks less reasonable.

3. The Simplest Solution and Its Main Properties

To get the simplest solution of Eq. (6), let us assume that the three-dimensional space is perfectly flat (i.e., $k \equiv 0$), and the energy density of ordinary matter is negligible (i.e., $\rho = 0$). Then, the resulting equation is easily integrated, and the solution can be written as

$$R(t) = R^* \exp\left[\sqrt{\frac{16\pi C_{\mathrm{UR}}}{3}}\sqrt{\frac{t}{\tau}}\,\right], \tag{7}$$

where $R^* = R(0)$ is the constant of integration; and we consider only the case of $R(t)$ increasing with time, i.e., corresponding to the expanding Universe. (In principle, there may be also a contracting solution, with minus sign in the argument of the exponential function.)

So, even the simplest solution (7) suggests us an interesting version of the cosmological model, outlined in Fig. 1. Instead of the standard model, composed of the four very different stages (namely, the inflationary one with exponential expansion, then the stages dominated by radiation and non-relativistic matter with power-like expansion, and finally the modern exponential stage), the uncertainty-mediated model describes everything by the universal "quasi-exponential" function (7). The corresponding Hubble parameter gradually decays with time as $1/\sqrt{t}$.

The age of the Universe T (i.e., the period from $t = 0$ to the present time) can be easily found from Eq. (6), remembering that $k \equiv 0$ and $\rho = 0$:

$$H_0^2 = \frac{4\pi C_{\mathrm{UR}}}{3\tau}\,\frac{1}{T}\,. \tag{8}$$

Next, taking into account that age of the Universe in the standard cosmological model T^* is commonly estimated as $1/H_0$, we get:

$$T \approx (T^*/\tau)\,T^*, \tag{9}$$

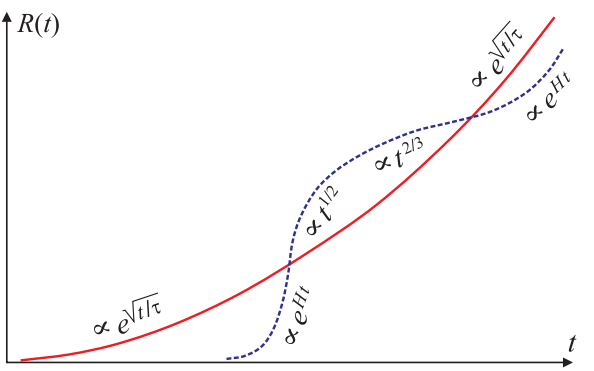


Fig. 1. Temporal variation of the scale factor of the Universe $R(t)$ in the standard cosmological model (dashed blue curve) and in the proposed uncertainty-mediated model (solid red curve).

if the numerical coefficient C_{UR} is assumed to be on the order of unity. Since $T^* \approx 4{\cdot}10^{17}\,\mathrm{s}$ and $\tau = 5{\cdot}10^{-44}\,\mathrm{s}$, this age in the proposed model will be greater than in the standard cosmology by the huge factor $(T^*/\tau) \approx 10^{61}$.

Nevertheless, this disagreement should not be a crucial failure of our model. Really, the "standard" age of the Universe T^* is measured from the onset of the hot Big Bang. If its evolution is supplemented by the preceding inflationary stage, this age will formally tend to infinity (because the expansion function e^{Ht} starts at $t \to -\infty$). So, the huge age of the Universe in the uncertainty-mediated model, described by the "universal" function (7), is not surprising.

4. The Model of Inflation

As is known, emergence of the inflationary paradigm in the early 1980s was caused by the inability of the "old" cosmological model—governed by radiation and ordinary matter—to resolve a few conceptual problems, such as a homogeneity and isotropy of the observed Universe, the absence of the topological defects, a surprising flatness of the three-dimensional space, etc. A presence of the early stage with very fast expansion (usually, the exponential one) was postulated as a universal remedy to treat all these issues.[1,3] So, it is interesting to check if the "quasi-exponential" expansion (7) can resolve the above-listed problems.

4.1. *Causal Structure of the Space–Time*

The problems of homogeneity and isotropy, as well as the absence of topological defects are closely related to the causal structure of the space–time under consideration. From our point of view, the most convenient way to analyze this issue is using the conformal diagrams.[15] Namely, in terms of the conformal time

$$\eta = \int dt/R(t) \tag{10}$$

the light rays in the curved Universe will be presented by the straight lines inclined at $\pm\pi/4$; and the causality (light) cones will look like the regular triangles, as shown in Fig. 2.

As distinct from the previous formulas, the null point of time was shifted here to the instant of observation; $\eta = -\eta_0$ corresponds to the "birth" of the Universe; and $\eta = -\eta^*$ is the instant at which the Universe is observed. Particularly, this may be the instant of plasma recombination—when the matter became transparent—if we speak about the observations in electromagnetic waves; or this may be the instant of symmetry-breaking phase transition of Higgs fields—when the ordinary matter was formed—if we speak about formation of the topological defects (such as the domain walls, cosmic strings, or monopoles, depending on the symmetry group involved). Accordingly, x_{p} is the size of the domain observable at the present time, and x_{f} is the size of the domain whose state can be "equilibrated" (causally connected) during the preceding cosmological evolution.

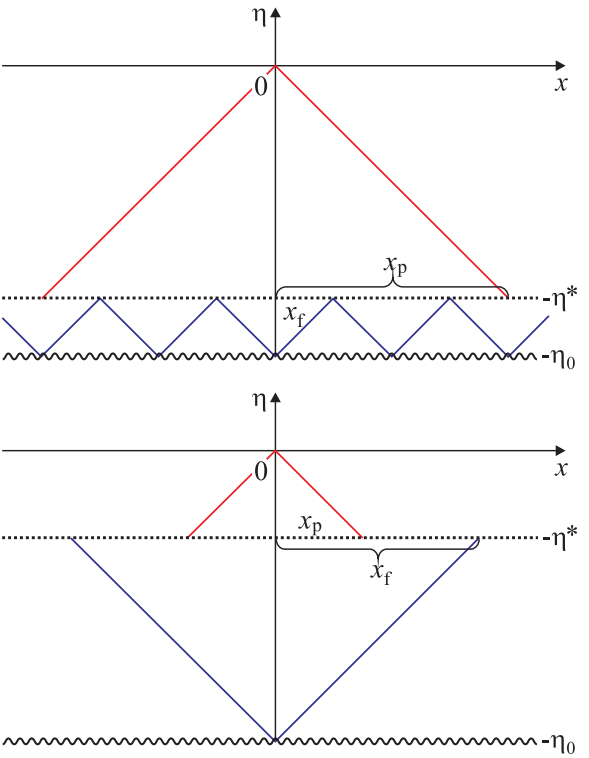


Fig. 2. Causal structure of space–time in the old "pre-inflationary" models (top panel) and in the inflationary models of the Universe (bottom panel). The lower (blue) triangles are the future causality cones developing from the birth of the Universe, and the upper (red) triangle is the past causality cone, representing the part of the Universe observed at the present time.

As is seen in the top panel of Fig. 2, a crucial drawback of the old "pre-inflationary" cosmology was that the observable region of space $x_{\rm p}$ (e.g., by the instant of recombination or by the instant of phase transition) comprised a large number of domains N developing independently from the birth of the Universe. This is because a difference between the conformal times η_0 and η^* in the case of power-like expansion of the Universe is rater small. As follows from the detailed calculations,[10]

$$N \approx \left\{\left[1-\left(\frac{T-t^*}{T}\right)^{\frac{1+3w}{3(1+w)}}\right]^{-1}-1\right\}^{-3} \gg 1\,, \tag{11}$$

where w parameterizes the effective equation of state of the cosmological matter, $p = w\rho$ ($w = 0$ for the non-relativistic matter, and $w = 1/3$ for the radiation); and $T(\eta_0)$ is the age of the Universe. As a result, one should expect that, for example, the observed CMB radiation will be strongly anisotropic (if t^* is associated with the instant of recombination), or a large number of topological defects will be formed

at the boundaries between the domains[16–19] (if t^* is associated with the instant of phase transition).

The situation changes radically if the inflationary stage with exponential expansion is present. Really, if $R(t) \propto e^{Ht}$, then

$$\eta(t) \propto -e^{-Ht} \to -\infty \quad \text{at} \quad t \to -\infty\,. \tag{12}$$

In other words, there will be a huge interval of conformal time before the instant η^*, corresponding to the observable structure. Consequently, the entire observable region $x_{\rm p}$ will be well within a single domain of causality $x_{\rm f}$ or, more exactly,[10]

$$N \approx \exp\bigl[-\sqrt{3\Lambda}\,c\,(T-t^*)\bigr] \ll 1\,, \tag{13}$$

as is seen in the bottom panel of Fig. 2. This is, in fact, the main reason why the inflationary model easily resolved the crucial problems of the "old" cosmology.

At last, what should we expect in the uncertainty-mediated cosmological model? Will the quasi-exponential expansion law (7) be a reasonable alternative to the pure exponential expansion? A quite straightforward but cumbersome calculation[10] results in

$$N \approx \left(\frac{16\pi C_{\rm UR}}{3}\right)^{3/2} \left(\frac{T-t^*}{\tau}\right)^{3/2} \exp\left[-4\sqrt{3\pi C_{\rm UR}}\left(\frac{T-t^*}{\tau}\right)^{1/2}\right] \ll 1\,, \tag{14}$$

i.e., the corresponding conformal diagram will be qualitatively the same as in the bottom panel of Fig. 2. In other words, the problems of homogeneity and isotropy of matter distribution, the absence of topological defects, etc. can be resolved by the uncertainty-mediated approach as efficiently as by the standard inflationary scenario.

4.2. *The Problem of Singularity*

A presence of the initial singularity in the power-like solution of the Friedmann equation with ordinary matter was also considered for a long time as a serious drawback of the cosmological model, but a considerably less attention is usually paid to this issue nowadays. Anyway, the quasi-exponential solution (7) of the uncertainty-mediated model evidently avoids the problem of singularity. However, this is achieved in a different way than in the standard inflationary model: in our case, the function $R(t)$ takes a finite value at $t = 0$ and becomes indefinite at $t < 0$; while in the standard exponential inflation the singularity is actually shifted to minus infinity.

Besides, let us mention that the uncertainty-mediated solution (7), in principle, if favorable for constructing a bounce-type model of the Universe, when a contraction at $t < 0$ is followed by the expansion at $t > 0$. Really, since the modified solution (7) with a minus sign in the argument of exponential function also satisfies the Friedmann equation, the decreasing and increasing branches can be fit to each other at $t = 0$.

4.3. *Formation of the Flat Three-Dimensional Space*

Yet another important conceptual problem of the old "pre-inflationary" cosmology was why the observed three-dimensional space is approximately flat with a high degree of accuracy. If we do not assume *a priori* that $k \equiv 0$ in the Friedmann equation (3), but the Λ-term is absent and the ordinary matter is described by the equation of state $p = w\rho$, then its energy density ρ will depend on the scale factor as $1/R^{3(1+w)}$, i.e., $1/R^3$ in the case of non-relativistic (dust-like) substance and $1/R^4$ in the case of radiation.[14] On the other hand, the curvature term decays as $1/R^2$ and, therefore, it will inevitably dominate after some time if the evolution begins from the arbitrary (generic) initial conditions. In other words, the approximately flat three-dimensional space cannot be formed "dynamically", and it is necessary to impose the artificial constraint $k \equiv 0$.

This problem is naturally resolved in the framework of classical inflationary scenario: if Λ is non-zero and approximately constant, then the second (curvature) term in the right-hand side of Eq. (3) becomes negligible as compared to the third (vacuum energy) term in the course of time and, therefore, the approximately flat three-dimensional space will be formed starting from the generic initial conditions.

Finally, what is the situation in the uncertainty-mediated cosmological model, described by the modified Friedmann equation (6)? In this case, it is not so easy to compare the magnitudes of the various terms in the right-hand side, because one of them depends immediately on t, while others depend on the function $R(t)$. However, assuming in the first approximation that $\rho = 0$ and utilizing the simplest solution (7), we get:

$$\lim_{t \to \infty} \frac{1/t}{1/R^2(t)} = \lim_{t \to \infty} \frac{R^2(t)}{t} = \infty\,, \tag{15}$$

as can be proved, for example, applying the L'Hôpital's rule two times. Consequently, we should expect that the role of the curvature term will become negligible in the course of time. Strictly speaking, this argumentation is not a rigorous mathematical proof, because the above-mentioned solution (7) was obtained under assumption $k \equiv 0$. However, a more accurate treatment in Ref. 11 leads to the same result. Therefore, the uncertainty-mediated model resolves the problem of flatness as efficiently as the standard inflationary scenario.

5. Discussion and Conclusions

1. As follows from the above consideration, the uncertainty-mediated cosmological model provides a unified description of the entire evolution of the Universe on the basis of the well-established physical principle—the quantum-mechanical uncertainty relation. Besides, it resolves all major problems of the early Universe and, therefore, can serve as a reasonable alternative to other inflationary scenarios.
2. Unfortunately, a few conceptual issues still remain questionable. Firstly, this is the basic idea of estimating the vacuum energy density from the uncertainty

relation. Secondly, it is not sufficiently clear what is the physical scale at which the uncertainty relation should be applied. We assumed above that this is the Planck volume. However, as was pointed out by Starobinsky (private communication), this might be also the cosmological horizon. In this case, the vacuum energy derived from the uncertainty relation would be extremely small and insignificant for the cosmological evolution. Nevertheless, from our point of view, the Planck volume is a more reasonable choice because the uncertainty relation should be applied to a fixed physical system, while the horizon is evidently an object changing with time.

3. It is also unclear how the ordinary matter can be created in the framework of the uncertainty-mediated scenario, since it was completely ignored in the simplest solution (7). So, it might be necessary to introduce the additional field(s) whose decay will produce the ordinary substance, as it is commonly postulated in other inflationary models. On the other hand, if the presence of ordinary matter is assumed *a priori*, it is unclear if it will destroy the inflationary stage (i.e., if the scale factor $R(t)$ will collapse into a singularity) under extrapolation backwards in time. To answer this question, it is necessary to obtain more general solutions of the modified Friedmann equation (6).
4. At last, a number of the additional physical effects, such as leptogenesis, bariosynthesis, formation of the large-scale structure, the spectrum of CMB fluctuations, etc. should be considered in more detail to draw the ultimate conclusion on viability of the uncertainty-mediated cosmological model.

References

1. A. H. Guth, The inflationary Universe: A possible solution to the horizon and flatness problems, *Phys. Rev. D* **23**, p. 347 (1981).
2. M. J. Mortonson, D. H. Weinberg and M. White, Dark Energy, *Chin. Phys. C* **40 (100001)**, p. 402 (2016), in *Review of Particle Physics* by C. Patrignani, et al. (Particle Data Group).
3. A. D. Linde, The inflationary Universe, *Rep. Prog. Phys.* **47**, p. 925 (1984).
4. A. D. Sakharov, Vacuum quantum fluctuations in curved space and the theory of gravitation, *Sov. Phys. Dokl.* **12**, p. 1040 (1968).
5. A. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Phys. Lett. B* **91**, p. 99 (1980).
6. J. Martin, C. Ringeval and V. Vennin, Encyclopædia inflationaris, *Phys. Dark Universe* **5–6**, p. 75 (2014).
7. J. Ellis and D. Wands, Inflation, *Chin. Phys. C* **40 (100001)**, p. 367 (2016), in *Review of Particle Physics* by C. Patrignani, et al. (Particle Data Group).
8. A. Coe, *Observable Universe Topology and Heisenberg Uncertainty Principle*, Preprint viXra: 1710.0143 (2017).
9. Y. V. Dumin, A unified model of Dark Energy based on the Mandelstam–Tamm uncertainty relation, *Grav. Cosmol.* **25**, p. 169 (2019).
10. Y. V. Dumin, Cosmological inflation based on the uncertainty-mediated Dark Energy, *Grav. Cosmol.* **26**, p. 259 (2020).

11. Y. V. Dumin, The problem of flatness of the Universe in the uncertainty-mediated inflationary model, *Grav. Cosmol.* **27**, p. 302 (2021).
12. L. I. Mandelstam and I. E. Tamm, The energy–time uncertainty relation in non-relativistic quantum mechanics, *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **9**, p. 122 (1945), in Russian.
13. S. Deffner and S. Campbell, Quantum speed limits: from Heisenberg's uncertainty principle to optimal quantum control, *J. Phys. A: Math. Theor.* **50**, p. 453001 (2017).
14. K. A. Olive and J. A. Peacock, Big-Bang cosmology, *Chin. Phys. C* **40 (100001)**, p. 355 (2016), in *Review of Particle Physics* by C. Patrignani, et al. (Particle Data Group).
15. C. W. Misner, Mixmaster Universe, *Phys. Rev. Lett.* **22**, p. 1071 (1969).
16. N. N. Bogoliubov, Field-theoretical methods in physics, *Suppl. Nuovo Cimento (Ser. prima)* **4**, p. 346 (1966).
17. T. W. B. Kibble, Topology of cosmic domains and strings, *J. Phys. A: Math. Gen.* **9**, p. 1387 (1976).
18. W. H. Zurek, Cosmological experiments in superfluid helium?, *Nature* **317**, p. 505 (1985).
19. Y. V. Dumin, Nonlocal quantum effects in cosmology, *Adv. High Energy Phys.* **2014**, p. 241831 (2014).

Conformal dilaton gravity, antipodal mapping and black hole physics on a warped spacetime

R. J. Slagter

ASFYON, Astronomisch Fysisch Onderzoek Nederland,
former: University of Amsterdam, Dept. Theor. Phys., The Netherlands
E-mail: info@asfyon.com

An exact time-dependent solution of a black hole is found in a conformally invariant gravity model on a warped Randall-Sundrum spacetime, by writing the metric $g_{\mu\nu} = \omega^{\frac{4}{n-2}}\tilde{g}_{\mu\nu}$. Here $\tilde{g}_{\mu\nu}$ represents the "un-physical" spacetime and ω the dilaton field, which will be treated on equal footing as any renormalizable scalar field. It is remarkable that the 5D and 4D effective field equations for the metric components and dilaton fields can be written in general dimension $n = 4, 5$. The location of the horizon(s) are determined by a quintic polynomial. This polynomial is related to the symmetry group of the icosahedron, isomorphic with the Galois group A_5. We applied the antipodal mapping on the axially symmetric black hole spacetime and make some connection with the information and firewall paradoxes. The dilaton field can be used to describe the different notion the in-going and outside observers have of the Hawking radiation by using different conformal gauge freedom. The disagreement about the interior of the black hole is explained by the antipodal map of points on the horizon. The free parameters of the solution can be chosen in such a way that $\tilde{g}_{\mu\nu}$ is singular-free and topologically regular, even for $\omega \to 0$.

Keywords: Conformal Invariance; Dilaton Field; Black Holes; Brane World Models; Antipodal Map; Quintic Polynomial

1. Introduction

One of the greatest challenges of modern physics will be the construction of a quantum-gravity model. In the vicinity of a horizon of a black hole and in the very early universe, these quantum gravity effects will come into play. The quantum features of a black hole were investigated, decades ago, by Hawking in his epic work on radiation effects of a black hole[1] by vacuum polarization. This radiation would be thermal and would contain no information. The black hole will eventually evaporate, and one could say that information is lost, because the anti-particles will fall into the black hole. This would violate quantum mechanics (QM). Related to this issue, is the holographic principle,[2] which states that the interior volume of spacetime of a black hole containing the information of the in-going particles is dual to the surface of the horizon. Could it be that the information is still at the horizon? The idea was extended to the well-known Anti-de Sitter/Conformal Field Theory (AdS/CFT) correspondence: is some way, the information must be present in the Hawking radiation. This model relies heavily on string theory, but would solve the information paradox, by introduction the notion of complementarity of the

in- and outside of the black hole. The in-going and out-going particles are entangled and the information of the in-going particle is also reflected back. However, this viewpoint conflict causality.[3] The previously emitted Hawking radiation and the corresponding in-going particles are independent systems and at the same time indirect entangled. Another solution for the information paradox, which don't rely on string theory, is the introduction of a firewall.[4] The entanglement between the in-going and out-going particles is broken by a high energetic shield. The freely in-falling observer encounters high-energy particles at the horizon. This viewpoint conflicts general relativity, i.e., violation of the equivalence principle. Free falling observers, when falling through the horizon, perceive spacetime as Minkowski, so will not notice the horizon at all. A fundamental issue which is omitted in all the treatments as described above, is the time-dependency of the spacetime structure near the horizon. The emitted Hawking particle will have a back-reaction effect on the spacetime.[5,6] Could it be possible, that the topology of the black hole must be revised? It is well know, that quantum field theory on a curved spacetime opens the possibility that a field theory can have different vacuum states. It can have intrinsic statistical features from a change in topology and not from from a priori statistical description of the matter fields. A spacetime with a given local geometry admits in principle, different possible global topologies. One can consider the modification of the spacetime topology of the form $\widehat{\mathcal{M}}/\Gamma$, where Γ is a discrete subgroup of isometries of $\mathcal{M}$,[7–10] without fixed points. $\widehat{\mathcal{M}}$ is non-singular and is obtained from its universal covering $\mathcal{M}$ by identifying points equivalent under Γ. A particular interesting case is obtained, when Γ is the antipodal transformation on $\mathcal{M}$

$$J: \quad P(X) \to \widehat{P}(\widehat{X}). \tag{1}$$

where the light-cone of the antipode of $P(X)$ intersects the light-cone of $P(X)$ only in two point (at the boundary of the spacetime). This is the so-called "elliptic interpretation"[11] of spacetime, where antipodal points represents in fact the same world-point or event. The future and past event horizon intersect each other as a projected cylinder $\mathbb{R}_1 \times \mathbb{S}_1/\mathbb{Z}_2$.[a] At the intersection one then identifies antipodal points. One must realize that the antipodal map is a boundary condition at the horizon, only observable by the outside observer. On a black hole spacetime, the inside is removed. So nothing can escape the interior, since there is no interior. The field theories formulated on $\mathcal{M}$ and $\widehat{\mathcal{M}}$ are globally different, while locally $\mathcal{M}$ and $\widehat{\mathcal{M}}$ are identically. The emitted radiation is only locally thermal. Antipodal identification, however, destroys the thermal features in the Fock space construction. In the construction, one needs unitary evolution operators for the in-going and out-going particles.[5,12] In order to avoid wormhole constellations or demanding "an other universe" in the construction of the Penrose diagram, it is essential that the asymptotic domain of $\mathcal{M}$ maps one-to-one onto the ordinary spacetime in order

[a]We work here in polar coordinates, because the spinning black hole we will consider, has a preferred spin axis. The antipodal identification is then $(U, V, z, \varphi) \to (-U, -V, -z, \pi + \varphi)$.

to preserve the metric. In fact, one deals with one black hole. A consequence is that time-inversion take place in region II of the Penrose diagram, so interchange of the creation and annihilation operators and entangling positive energy particles at the horizon with positive energy antiparticles at the antipodes. So the antipodal identification is not in conflict with the general CPT invariance of our world. Further, for the outside observer, the thermodynamically mixed state is replaced by a pure state. So the Hawking particles at opposite sides of the black hole are entangled. The former representation that observers has no access to the inside of the black hole is no longer valid. One arrives by this new geometrical description at pure quantum states for the black hole. It will solve, moreover, the information paradox and firewall problem as well.[b] The gravitational back-reaction as proposed by 't Hooft,[13,14] suggests a cut-off of high momenta, which avoids the firewall. The in-going particle has a back-reaction on the other particles, leading to an unitary S-matrix. The gravitational interaction between the in-going and out-going particles will be strong, because we are dealing here with a strongly curved spacetime near the horizon. Using a "cut-and-paste" procedure, one replaces the high-energy particles ("hard"), i.e., mass or momentum of the order of the Planck mass, by low-energy ("soft") particles far away. These hard particles just caused the firewall problem. Hard particles will also influence the local spacetime (to become non-Schwarzschild) and causes the Shapiro effect. The interaction with the soft particles is described by the Shapiro delay. Effectively, all hard particles are quantum clones of all soft particles. By this "firewall-transformation", we look only at the soft particle clones. They define the Hilbert space and leads to a unitary scattering matrix. The net result is that the black hole is actually in a pure state, invalidating the entanglement arguments in the firewall paradox. The entanglement issue can be reformulated by considering the two regions I and II in the maximally extended Penrose diagram of the black hole, as representing two "hemispheres" of the same black hole. It turns out that the antipodal identification keeps the wave functions pure[6] and the central $r = 0$ singularity has disappeared. This gravitational deformation will cause transitions from region I to II in the Penrose diagram. The fundamental construction then consist of the exchange of the position operator with the momentum operator of the in-going particles, which turn them into out-particles. Hereby, 't Hooft expands the moment distributions and position variables in partial waves in (θ, φ).[5] So the Hawking particles emerging from I are entangled with the particles emerging from II. An important new aspect is the way particles transmit the information they carry across the horizon. In the new model, the Hawking particles emerging from I are maximally entangled with the particles emerging from II. The particles form a pure state, which solves the information paradox. In order to describe the more realistic black holes, such as the axially symmetric Kerr black hole, it is not possible

[b]The technical aspects in constructing the unitary S-matrix can be found in the literature, as provided by the references.

to ignore the dynamics of the horizon. Moreover, one must incorporate gravitation waves. There is another reason to consider axially symmetry. A spherical symmetric system cannot emit gravitation waves.[15] Astronomers conjecture that most of the black holes in the center of galaxies are of the Kerr type. A linear approximation is, of course, inadequate in high-curvature situations. In the linear approximation, the waves don't carry enough energy and momentum to affect their own propagation. The notion of the "classical" Hartle-Hawking vacuum thermal state, with a temperature $T \sim \frac{1}{M} \sim \kappa$ and the luminosity $\frac{dM}{dt} \sim -\frac{1}{M^2}$ must also be revised when the mass reaches the order of the Planck mass. On the Kerr black hole spacetime no analog of the Hartle-Hawking vacuum state exists. The Killing field ξ^μ generates a bifurcate Killing horizon ($\xi^\mu \xi_\mu = -1$ at infinity) and possesses spacelike orbits near infinity.[16] Another aspect of the huge curvature in the vicinity of the horizon, will be the problem of constructing a renormalizable (and maintaining unitarity) quantum gravity model of the Standard Model fields, which must be incorporated in the Lagrangian. Up till now, no convincing theory of quantum gravity is available. Many attempts were made in order to make a renormalizable and unitary quantum gravity model. One also can try to construct a renormalizable model, by adding fourth order derivative terms of the curvature tensor (Euler-term). However, one looses unitarity. Also the "old" effective field theory (EFT) has its problems. One ignores what is going on at high energy. In order to solve the anomalies one encounters in calculating the effective action, one can apply the so-called conformal dilaton gravity (CDG) model.[5,6,17] CDG is a promising route to tackle the problems arising in quantum gravity model, such as the loss of unitarity close to the horizon. One assumes local conformal symmetry, which is spontaneously broken (for example by a quartic self-coupling of the Higgs field). Changing the symmetry of the action was also successful in the past, i.e., in the SM of particle physics. A numerical investigation of a black hole solution of a non-vacuum CDG model, was recently performed.[18] The key feature in CDG, is the splitting of the metric tensor $g_{\mu\nu} = \omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}$, with ω the dilaton field. Applying perturbation techniques (and renormalization/dimensional regularization), in order to find the effective action and its divergencies, one first integrate over ω (shifted to the complex contour), considered as a conventional renormalizable scalar field and afterwards over $\tilde{g}_{\mu\nu}$ and matter fields. The dilaton field is locally unobservable. It is fixed when we choose the global spacetime and coordinate system. If one applies this principle to a black hole spacetime, then the energy-momentum tensor of ω influences the Hawking radiation. When $\tilde{g}_{\mu\nu}$ is flat, then the handling of the anomalies simplifies considerably.[14] When $\tilde{g}_{\mu\nu}$ is non-flat, the problems are more deep-seated. It is well known, that the antipodal transformation, or inversion, is part of the conformal group.[19] So conformal invariant gravity models could fit very well the models of antipodal mapping as described above. In this context, the modification of GRT by an additional spacetime dimension could be an alternative compromise, because Einstein gravity on the brane will be modified by the very embedding itself and opens up

a possible new way to address the dark energy problem.[20,c] These models can be applied to the standard Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime and the modification on the Friedmann equations can be investigated.[21] Recently, Maldacena, et al.,[22] applies the RS model to two black hole spacetimes and could construct a traversable macroscopic wormhole solution by adding only a 5D U(1) gauge field (see also Maldacena[23]). However, an empty bulk would be preferable. In stead, one can investigate the contribution of the projected 5D Weyl tensor on the 4D brane. It carries information of the gravitational field outside the brane. If one writes the 5D Einstein equation in CDG setting, it could be possible that an effective theory can be constructed without an UV cutoff, because the fundamental scale M_5 can be much less than the effective scale M_{Pl} due to the warp factor. The physical scale is therefore not determined by M_{Pl}. In this manuscript we will apply the antipodal map on a spinning black hole spacetime in conformal dilaton gravity applied to a warped 5D spacetime.

2. Conformal Transformations and Antipodal Mapping

2.1. *The origin of the antipodal mapping*

The antipodal map originates from the so-called "elliptic" interpretation.[11] If one considers the hyperboloid H, $-t^2+x^2+y^2+z^2+w^2 = R^2$, then the space-like sections through the origin are ellipses and the time-like sections are hyperbola branches. Since the de Sitter spacetime can be isometrically embedded as a hyperboloid in $\mathbb{R}^5$, one can take $R^2 = -\frac{3}{\Lambda}$. If one suppresses the coordinates (z, w), we have the $\mathbb{R}^3$ Minkowski metric. Lorentz transformations (LT's) around the origin transforms H into itself. Circles on H represent space at different epochs. The bottle-neck parallel is a spatial geodesic, while the others are not. Further, the circumferences contract from $z = -\infty$ to $z = 0$ and then expand. A LT of $\mathbb{R}^3$ turns the bottle-neck into an ellipse, cut out of H with an angle $< 45^o$ with the (x, y)-plane. See figure 1. All the ellipses are equivalent space-like geodesics since each of them is transferred by a suitable automorphism into the bottle-neck, which is one of them. One defines the antipodal map

$$J : P(t, x, y) \to \widehat{P}(-t, -x, -y), \tag{2}$$

on H. The antipodicity is Lorentz invariant. When the angle approaches 45^o, then the ellipses degenerate into a couple of parallel generators (g_1, g_2) (null geodesics). The other plane of 45^o delivers the set (g_3, g_4). The sets (g_1, g_4) and (g_2, g_3) form, for example, the light-cones at the points M and $\widehat{M}$. If one moves upwards along t, the inner angles of the light-cones decrease. Note that the light-cones at P and $\widehat{P}$ has no point in common and the antipodes are joined by a space-like geodesic. Now Schrödinger proposed to identify P and $\widehat{P}$ with the same physical world-point or

[c]There is another argument in favor of a (warped) 5D spacetime. It turns out, as we shall see, that a surface in 4D can be immersed to 5D, like a Klein bottle.

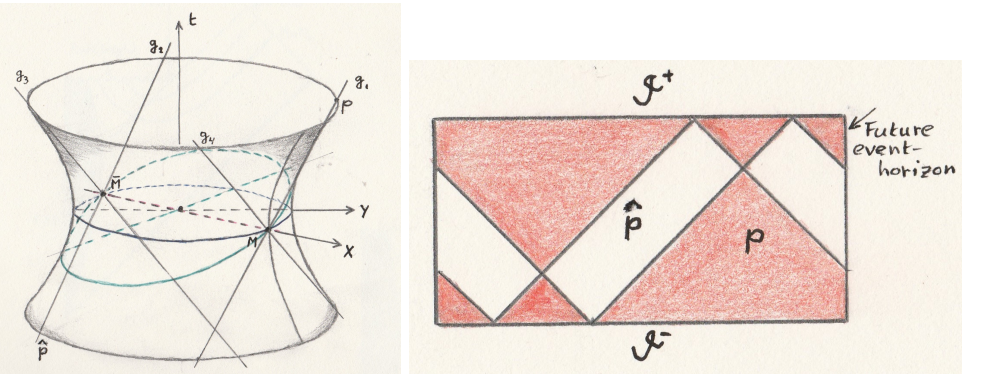


Fig. 1. Hyperboloid H representing the $\mathbb{R}^3$ spacetime of the compactified de Sitter universe $\mathbb{R}^5$ ((z, w) suppressed). Right: the Penrose diagram. The antipodal points P and $J(P)$ are spacelike separated. An observer moving in de Sitter spacetime cannot meet both P and $J(P)$. He cannot receive a message from $P, J(P)$. Moreover, he cannot receive a message from P and send a message to $J(P)$.

event. One half of H, containing no antipodal points, represents the "whole world". Thereafter, Schrödinger argues in a clever way that the total potential of experiences of any observer is complete and embraces the same events for any two observers, whatever their world lines be. But there is a price we have to pay for.[d] The direction of the arrow of time is lost (or the distinction between the "fore-cone" and "after-cone" is lost). The allotment of past and future is undecidable. The elliptic model is time-reversible. This can open perspective to the general CPT invariance of our world. The real problems arise, when one considers thermodynamical systems, as is the case for the Hawking effect in the vicinity of the horizon of a black hole. Then the entropy comes into play. Note, quoting Schrödinger, "the irreversible laws of thermodynamics can only be based on the statistical microscopically reversible systems on condition that statistical theory be autonomous in defining the arrow of time. If any other law of nature determines this arrow, the statistical theory collapses."

In a pseudo-polar frame $(\chi, T, \theta, \varphi)$ we can write the line element

$$ds^2 = -R^2 dT^2 + R^2 \cosh^2 T \Big[d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2) \Big], \tag{3}$$

where $0 < \chi < 2\pi$. The antipodal map becomes now

$$J : (T, \chi, \theta, \varphi) \to (-T, \pi + \chi, \pi - \theta, \pi + \varphi). \tag{4}$$

[d]This price is worth paying in the black hole situation, when the information paradox will be solved by the antipodal map. The antipodal half is not time orientable. There is a breakdown of the global distinction between past and future in the interior of the black hole.

We already mentioned that de Sitter can be embedded as a hyperboloid in 5D Minkowski. We then say that $J : X^\mu \to -X^\mu$ is an inversion.[e] There exist another coordinate system (introduced by de Sitter himself) in which the line element is written as

$$ds^2 = -\left(1 - \frac{\rho^2}{R^2}\right) dT'^2 + \frac{1}{(1 - \frac{\rho^2}{R^2})} d\rho^2 + \rho^2 (d\theta'^2 + \sin^2 \theta' d\varphi'^2), \tag{5}$$

where we have taken the velocity of the LT $\tanh T = \frac{t}{y}$. This is the static de Sitter and the spaces of constant time are all equivalent. There are singularities for $x = \pm R$ ($\chi = \pm 90^o$), i.e., the points $(M, \widehat{M})$. However, as also observed by Schrödinger, this static model is not adequate for applying the antipodal map. In order to apply the antipodal map on a black hole spacetime in a more general setting, one needs a time dependent spacetime.

2.2. *The "classical" Hawking effect and its problems*

The famous result of Hawking states, that a black hole will radiate at "sufficiently" late times like a black body at a temperature

$$kT \sim \frac{\kappa}{2\pi} = \frac{\hbar c^3}{8\pi GM}, \tag{6}$$

with κ the surface gravity and M the mass. The entropy should then be $S_{bh} = \frac{kc^3}{4\hbar} A$, with A the area of the horizon. However, one runs into problems by the back-reaction effect of the particle creation, which will alter the area. It is questionable if the ordinary laws of thermodynamics can be applied to a black hole. It is clear that these laws must be constrained to form quantum states with orthonormality and unitarity conditions. Suppose that an isolated black hole completely evaporate within a finite time. Loss of quantum coherence should then occur i.e., an initially pure quantum state should evolve to a mixed state. In general, in the classical picture, a black hole cannot causally influence its exterior, so it is hard to understand the mechanism by which thermal equilibrium could be achieved. Observe that the state of the field at late times in the region I of the Penrose diagram (and so the particles flux reaching infinity) is described by a density matrix by the S-matrix analysis. The particles present in region I are strongly correlated with the particles which entered the black hole at earlier times. Consider now in figure 2 the evolution of two Cauchy surfaces ("time" Σ_1 to "time" Σ_2). When the black hole disappears from the spacetime, then at late times, the entire state of the field is mixed. If one takes the "out" Hilbert space to be the Fock space of the particles propagating out to infinity at late times, one cannot describe particle creation and scattering by an ordinary S-matrix. The initial pure state will evolve to a final density matrix.

[e]The inversion $X_\mu \to -\frac{X_\mu}{X^2}$ (as well as the dilatations) is part of the conformal group.[19] We shall see in the next sections that in general the conformal group is a projective group from 5D. The fifth "degree of freedom" is a sort of gauge space.

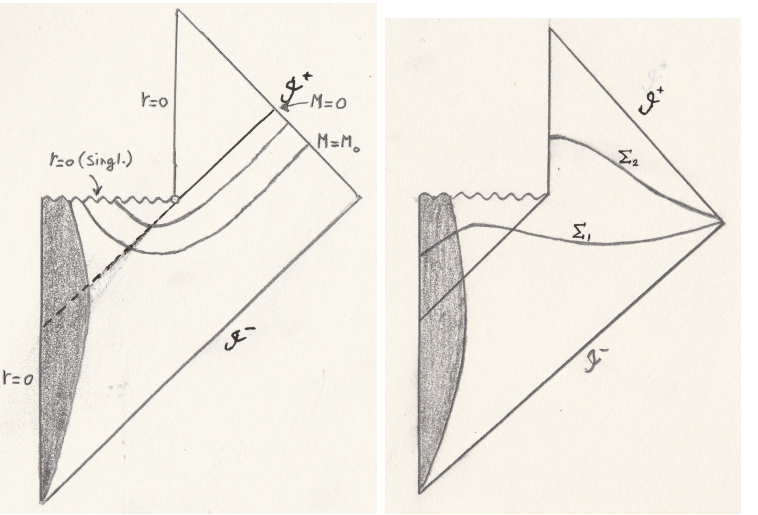


Fig. 2. Left: formation and evaporation of the Schwarschild black hole. The contour $M = 0$ lies at the retarded time corresponding to the final evaporation. The geometry is flat above this contour. Right: Loss of quantum coherence: evolution from a pure state to a mixed state.[16]

So we have a breakdown of quantum theory. The antipodal model, however, could "repair" this breakdown.

3. The Black Hole Solution on a 5D Warped Spacetime in Conformal Dilaton Gravity

3.1. *The 5D warped spacetime*

Let us consider the 5D spacetime warped spacetime[21,24]

$$ds^2 = \omega(t,r,y)^2\Big[-N(t,r)^2dt^2+\frac{1}{N(t,r)^2}dr^2+dz^2+r^2(d\varphi+N^\varphi(t,r)dt)^2+dy^2\Big], \quad (7)$$

where y is the extra dimension and ω a warp factor in the formulation of Randall-Sundrum's (RS) 5D warped spacetime with one large extra dimension and negative bulk tension Λ_5. The Standard Model (SM) fields are confined to the 4D brane, while gravity acts also in the fifth dimension. Originally, the RS model was applied to a 5-dimensional anti-de Sitter (AdS) spacetime with a positive brane tension. This is the so-called RS-1 model, with one brane. The RS-2 model treats two branes with $\mathbb{Z}_2$ symmetry. However, the effective cosmological constant on the brane can be zero by fine tuning with the negative Λ_5. In the RS model there is a bound state of the graviton confined to the wall as well as a continuum of Kaluza-Klein (KK) states. Four dimensional gravity is then recovered on the brane and the hierarchy problem seems to be solved. Since the pioneering publication of RS, many investigation were done in diverge domains. In particular, Shiromizu et.al.,[24,25] extended the RS model to a fully covariant curvature formalism. It this extended model, an effective

Einstein equation is found on the brane, with on the right-hand side a contribution from the 5D Weyl tensor which carries information of the gravitational field outside the brane. So the brane world observer may be subject to influences from the bulk. The field equations are (were we took an empty bulk)[18]

$$^{(5)}G_{\mu\nu} = -\Lambda_5\, ^{(5)}g_{\mu\nu}, \tag{8}$$

$$^{(4)}G_{\mu\nu} = -\Lambda_{eff}\, ^{(4)}g_{\mu\nu} + \kappa_4^{2\,(4)}T_{\mu\nu} + \kappa_5^4 \mathcal{S}_{\mu\nu} - \mathcal{E}_{\mu\nu}, \tag{9}$$

where we have written

$$^{(5)}g_{\mu\nu} = {}^{(4)}g_{\mu\nu} + n_\mu n_\nu, \tag{10}$$

with n^μ the unit normal to the brane. Here $^{(4)}T_{\mu\nu}$ is the energy-momentum tensor on the brane and $\mathcal{S}_{\mu\nu}$ the quadratic contribution of the energy-momentum tensor $^{(4)}T_{\mu\nu}$ arising from the extrinsic curvature terms in the projected Einstein tensor. Further,

$$\mathcal{E}_{\mu\nu} = {}^{(5)}C^\alpha_{\beta\rho\sigma} n_\alpha n^{\rho\,(4)} g_\mu^{\beta\,(4)} g_\nu^\sigma, \tag{11}$$

represents the projection of the bulk Weyl tensor orthogonal to n^μ. The effective gravitational filed equations on the brane are not closed. One must solve at the same time the 5D gravitation field in the bulk.

3.2. *The conformal dilaton gravity (CDG) model on a 5D warped spacetime*

One can distinguish several possible "routes" to the unification of GR and QFT. One can start, for example, with a given classical theory and applies heuristic quantization rules. One then can make a division in canonical and covariant approaches, i.e., uses a Hamiltonial formalism or employs covariance at some stage. The CDG model we consider here, is part of the covariant approach to quantum gravity. The key feature in CDG, is the splitting of the metric tensor[5,17]

$$g_{\mu\nu} = \omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}, \tag{12}$$

with ω the dilaton field and $\tilde{g}_{\mu\nu}$ the "un-physical" spacetime. At high energy, ω will be treated as a (renormalizable) quantum field. One can prove that the action (without matter terms for the time being)

$$S = \int d^n x \sqrt{-\tilde{g}} \Big[\frac{1}{2} \xi \omega^2 \tilde{R} + \frac{1}{2} \tilde{g}^{\mu\nu} \partial_\mu \omega \partial_\nu \omega + \Lambda \kappa^{\frac{4}{n-2}} \xi^{\frac{n}{n-2}} \omega^{\frac{2n}{n-2}} \Big], \tag{13}$$

is conformal invariant under

$$\tilde{g}_{\mu\nu} \to \Omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}, \quad \omega \to \Omega^{-\frac{n-2}{2}} \omega. \tag{14}$$

The covariant derivative is taken with respect to $\tilde{g}_{\mu\nu}$. For details, see Slagter.[18] Now we implement the 5D warped spacetime Eq. (7). So

$$ {}^{(5)}g_{\mu\nu} = \omega^{4/3}{}^{(5)}\tilde{g}_{\mu\nu}, \qquad {}^{(5)}\tilde{g}_{\mu\nu} = {}^{(4)}\tilde{g}_{\mu\nu} + n_\mu n_\nu, \tag{15} $$

and write again

$$ {}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^2 \bar{g}_{\mu\nu}. \tag{16} $$

Variation of the action leads to the field equations

$$ \xi\omega\tilde{R} - \tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\tilde{\nabla}_\nu\omega - \frac{2n}{n-2}\Lambda\kappa^{\frac{4}{n-2}}\xi^{\frac{n}{n-2}}\omega^{\frac{n+2}{n-2}} = 0 \tag{17} $$

and

$$ \omega^2\tilde{G}_{\mu\nu} = T^{\omega}_{\mu\nu} - \Lambda\tilde{g}_{\mu\nu}\kappa^{\frac{4}{n-2}}\xi^{\frac{2}{n-2}}\omega^{\frac{2n}{n-2}}, \tag{18} $$

with

$$ T^{\omega}_{\mu\nu} = \tilde{\nabla}_\mu\tilde{\nabla}_\nu\omega^2 - \tilde{g}_{\mu\nu}\tilde{\nabla}^2\omega^2 + \frac{1}{\xi}\Big(\frac{1}{2}\tilde{g}_{\alpha\beta}\tilde{g}_{\mu\nu} - \tilde{g}_{\mu\alpha}\tilde{g}_{\nu\beta}\Big)\partial^\alpha\omega\partial^\beta\omega. \tag{19} $$

From the 5D Einstein equations Eq. (8) one obtains $\omega(t,r,y) = \omega_1(t,r)\omega_2(y)$, with $\omega_2(y) = l$=constant (the length scale of the extra dimension). The dilaton equations Eq. (17) is superfluous. Note that the effective Einstein equations Eq. (9) contains the $\mathcal{E}_{\mu\nu}$, while $T_{\mu\nu}$ and $\mathcal{S}_{\mu\nu}$ are taken zero in our case. The dilaton equation is again superfluous.

It turns out that one can write the field equations for ω and N in the form $(n = 4, 5)$

$$ \ddot{\omega} = -N^4\omega'' + \frac{n}{\omega(n-2)}\Big(N^4\omega'^2 + \dot{\omega}^2\Big), \tag{20} $$

$$ \ddot{N} = \frac{3\dot{N}^2}{N} - N^4\Big(N'' + \frac{3N'}{r} + \frac{N'^2}{N}\Big) $$

$$ -\frac{n-1}{(n-3)\omega}\Big[N^5\Big(\omega'' + \frac{\omega'}{r} + \frac{n}{2-n}\frac{\omega'^2}{\omega}\Big) + N^4\omega'N' + \dot{\omega}\dot{N}\Big]. \tag{21} $$

One can solve these equations exact (we took $\Lambda_{eff} = 0$):

$$ \omega = \Big(\frac{a_1}{(r+a_2)t + a_3 r + a_4}\Big)^{\frac{1}{2}n-1}, $$

$$ N^2 = \frac{1}{5r^2}\frac{10a_2^3r^2 + 20a_2^2r^3 + 15a_2r^4 + 4r^5 + C_1}{C_2(a_3+t)^4 + C_3}, \tag{22} $$

with a_i some constants. There is a constraint equation

$$ \bar{\omega}'' = -\frac{2n}{n-2}\frac{\Lambda l\kappa^{\frac{4}{(n-2}}\xi^{\frac{n-2}{4(n-1)}}\bar{\omega}^{\frac{n+2}{n-2}}}{N^2} - \frac{\omega'N'}{N} - \frac{\omega'}{2r} + \frac{4}{n-2}\frac{\dot{\bar{\omega}}^2}{\bar{\omega}N^4} - \frac{\dot{\bar{\omega}}\dot{N}}{N^5}, \tag{23} $$

which l the dimension of y. The solution for the two dilaton fields ω and $\bar{\omega}$ differs only by the different exponent $\frac{3}{2}$ and 1 respectively. The solution For the metric

component is the same (apart from the constants). The solution for the angular momentum component is

$$N^{\varphi} = F_n(t) + \int \frac{1}{r^3 \bar{\omega}^{\frac{n-1}{n-3}}} dr. \tag{24}$$

The Ricci scalar for $\bar{g}_{\mu\nu}$ ($\Lambda = 0$) is given by

$$\bar{R} = \frac{12}{N^2}\Big[\dot{\bar{\omega}}^2 - N^4 \bar{\omega}'^2\Big], \tag{25}$$

with is consistent with the null condition for the two-dimensional (t, r) line element, when $\bar{R} = 0$. One can easily check that the trace of the Einstein equations is zero. Note that N^2 can be written as

$$N^2 = \frac{4\int r(r+a_2)^3 dr}{r^2[C_2(a_3+t)^4 + C_3]}. \tag{26}$$

So the spacetime seems to have two poles. However, the $r = 0$ is questionable. The conservation equations become

$$\bar{\nabla}^{\mu}\mathcal{E}_{\mu\nu} = \bar{\nabla}^{\mu}\Big[\frac{1}{\bar{\omega}^2}\Big(-\Lambda\kappa^{2\,(4)}\bar{g}_{\mu\nu}\bar{\omega}^4 + {}^{(4)}T_{\mu\nu}{}^{(\bar{\omega})}\Big)\Big], \tag{27}$$

which yields differential equations for $\ddot{N}'$ and $\dot{N}$ as boundary conditions at the brane. It can be described as the non-local conservation equation. In the high energy case close to the horizon, one must include the $\mathcal{S}_{\mu\nu}$ term. So the divergence of $\mathcal{E}_{\mu\nu}$ is constrained. In the non-conformal case, Eq. (27) contains on the right hand side also the quadratic correction $\mathcal{S}_{\mu\nu}$ of the matter fields on the brane. The effective field equations, Eq. (9), are then not a closed system. One needs the Bianchi equations. In fact, $\mathcal{E}_{\mu\nu}$ encodes corrections from the 5D graviton effects and are for the brane observer non-local. In our model under consideration, we have only the $T^{(\omega)}_{\mu\nu}$ term and no source terms (only the 5D Λ_5). But it still sources the KK modes. The dilaton ω plays the role of a "scalar field". But we don't need the 5D equations themselves, because the solution for N is the same! It is only the $\omega^{4/3}$ which represents the 5D contribution. There is no exchange of energy-momentum between the bulk and brane. If one applies the model to a FLRW model,[26] then the evolution equations are very complicated. Inhomogeneous and anisotropic effects from the 4D matter radiation distribution on the brane are sources for the 5D Weyl tensor $\mathcal{E}_{\mu\nu}$ and cause non-local back-reaction on the brane. One needs an approximation scheme in order to find the missing evolution equation for $\mathcal{E}_{\mu\nu}$.

The locations of the horizon's and ergo-spheres are found by solving $N^2 = 0$ and $\bar{g}_{tt} = 0$ respectively. N^2 becomes singular at coordinate time $t = t_H = -b_3 + \sqrt[4]{-\frac{C_3}{C_2}}$. However, $\bar{g}_{\mu\nu}$ can be made regular everywhere and singular free by suitable choices of the parameters b_i, c_i and C_i. For $C_1 = 0$, $\bar{g}_{\mu\nu}$ has one real zero $r_H = \sim |1.606 b_2|$ and two complex zero's $\sim (0.178 \pm 0.638I)b_2$. In figure 3 we plotted the possible graphs. If one ignores the contribution from the bulk, then N^2 has for $C_1 = 0$ no real roots, so only naked singularities. The contribution from the bulk then generates at least one horizon.

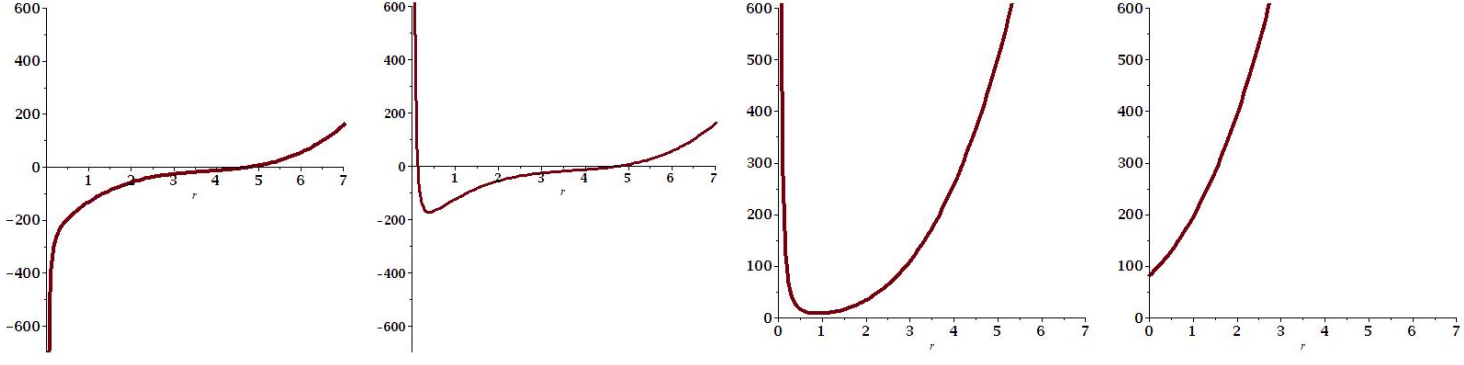


Fig. 3. Four possible plots of N^2 as function of r.

3.3. *Penrose diagram*

If we define the coordinates, $dr^* \equiv \frac{1}{N_1(r)^2} dr$ and $dt^* \equiv N_2(t)^2 dt$, then our induced spacetime can be written as

$$ds^2 = \omega^{4/3}\bar{\omega}^2\Big[\frac{N_1^2}{N_2^2}\Big(-dt^{*2} + dr^{*2}\Big) + dz^2 + r^2\Big(d\varphi + \frac{N^\varphi}{N_2^2}dt^*\Big)^2\Big], \tag{28}$$

with

$$N_1^2 = \frac{10b_2^3r^2 + 20b_2^2r^3 + 15b_2r^4 + 4r^5 + C_1}{5r^2}, N_2^2 = \frac{1}{C_2(t+b_3)^4 + C_3} \tag{29}$$

and

$$r^* = \frac{1}{4}\sum_{r_i^H} \frac{r_i^H \log(r - r_i^H)}{(r_i^H + b_2)^3}, \qquad t^* = \frac{1}{4C_2}\sum_{t_i^H} \frac{\log(t - t_i^H)}{(t_i^H + b_3)^3}. \tag{30}$$

The sum it taken over the roots of $(10b_2^3r^2 + 20b_2^2r^3 + 15b_2r^4 + 4r^5 + C_1)$ and $C_2(t+b_3)^4 + C_3$, i. e., r_i^H and t_i^H. This polynomial in r defining the roots of N_1^2, is a quintic equation, which has some interesting connection with Klein's icosahedral solution (see appendix). Further, one can define the azimuthal angular coordinate $d\varphi^* \equiv (d\varphi + \frac{N^\varphi}{N_2^2}dt^*)$, which can be used when an incoming null geodesic falls into the event horizon. φ^* is the azimuthal angle in a coordinate system rotating about the z-axis relative to the Boyer-Lindquist coordinates. Next, we define the coordinates[28] (in the case of $C_1 = C_3 = 0$ and 1 horizon, for the time being)

$$\begin{aligned} U_+ = e^{\kappa(r^* - t^*)}, \quad V_+ = e^{\kappa(r^* + t^*)} \qquad r > r_H \\ U_- = -e^{\kappa(r^* - t^*)}, \quad V_- = -e^{\kappa(r^* + t^*)} \qquad r < r_H, \end{aligned} \tag{31}$$

with κ a constant. The spacetime becomes

$$ds^2 = \omega^{4/3}\bar{\omega}^2\Big[\frac{N_1^2}{N_2^2}\log\Big(UV\Big)^{\frac{1}{2\kappa}} dUdV + dz^2 + r^2 d\varphi^{*2}\Big]. \tag{32}$$

In figure 4 we plotted the Penrose diagram (left). The antipodal points $P(X)$ and $\bar{P}(\bar{X})$ are physically identified. If we compactify the coordinates,

$$\tilde{U} = \tanh U, \qquad \tilde{V} = \tanh V, \tag{33}$$

then the spacetime can be written as

$$ds^2 = \omega^{4/3}\bar{\omega}^2\Big[H(\tilde{U}, \tilde{V})d\tilde{U}d\tilde{V} + dz^2 + r^2 d\varphi^{*2}\Big], \tag{34}$$

with

$$H = \frac{N_1^2}{N_2^2}\frac{1}{\kappa^2 \operatorname{arctanh}\tilde{U} \operatorname{arctanh}\tilde{V}(1 - \tilde{U}^2)(1 - \tilde{V}^2)}. \tag{35}$$

We can write r and t as

$$r = r_H + \Big(\operatorname{arctanh}\tilde{U} \operatorname{arctanh}\tilde{V}\Big)^{\frac{1}{2\kappa\alpha}}, \quad t = t_H + \Big(\frac{\operatorname{arctanh}\tilde{V}}{\operatorname{arctanh}\tilde{U}}\Big)^{\frac{1}{2\kappa\beta}}, \tag{36}$$

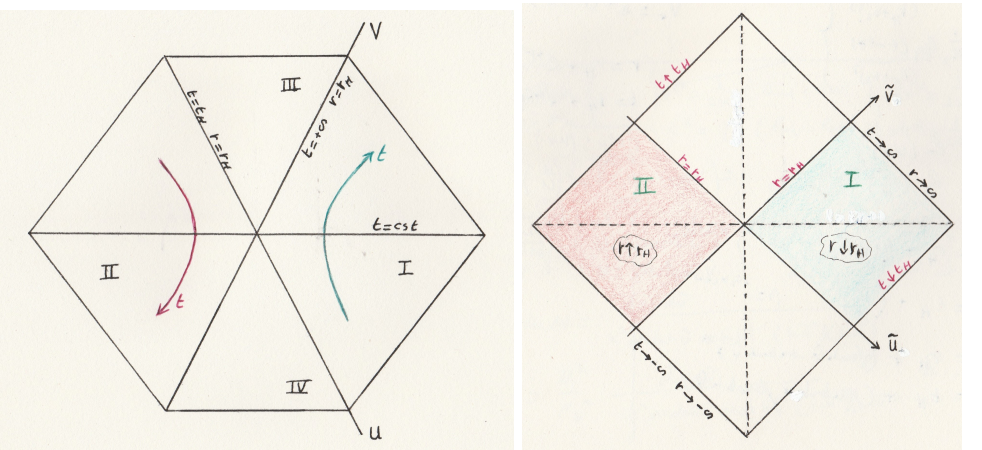


Fig. 4. Left: Kruskal diagram for $\bar{g}_{\mu\nu}$ in (U, V)- coordinates. If one approaches the horizon from the outside and passes the horizon, one approaches from the "otherside" the horizon. Right: Kruskal diagram for $\bar{g}_{\mu\nu}$ in $(\tilde{U}, \tilde{V})$- coordinates.

with

$$\alpha = \frac{r_H}{4(r_H + b_2)^3}, \qquad \beta = \frac{1}{4C_2(t_H + b_3)^3}. \tag{37}$$

Observe that N_1 and N_2 can be expressed in $(\tilde{U}, \tilde{V})$. The Penrose diagram is drawn in figure 4 (right). Note that ds^2 and H are invariant under $\tilde{U} \to -\tilde{U}$ and $\tilde{U} \to -\tilde{U}$. $\tilde{g}_{\mu\nu}$ is regular everywhere and conformally flat. The "scale-term" H is consistent with the features of the Penrose diagram. Now we have still the φ dependency. We assume no z-dependency. It is expected that the differential equation for ω can be separated in a (U, V) part and a φ part. The method of 't Hooft can then be applied. In the next sections, we will briefly come back to this issue.

4. Related Issues of the New Black Hole Solution

4.1. *Treatment of the quantum fields*

The physical identification in the de Sitter spacetime of $P(X)$ and $\widehat{P}(\widehat{X})$ are considered as different representations in Kruskal space of one and the same Schwarzschild event. There is only one world with one singularity and one exterior region. Fields which are symmetric under J are identified as

$$\Psi_{JS} = \frac{1}{2}\Big[\Psi(X) + \widehat{\Psi}(\widehat{X})\Big]. \tag{38}$$

One then builds these fields from fields with arguments specified in I.[8] Each of these fields, positive or negative frequency in I, can be extended to global spacetime surfaces. However, due to the time reversal, the inner product on the full Hilbert space have zero norm for the symmetric fields. One then defines negative frequency

functions $\Psi^{\uparrow}_{(-)}(X) = \Psi^{\downarrow}_{(+)}(JX)$ and $\Psi^{\downarrow}_{(-)}(X) = \Psi^{\uparrow}_{(+)}(JX)$, where the arrows stands for the solutions on the future/past singularity. The symmetric (anti-) solutions ($\epsilon = \pm 1$) are then

$$\Psi^{(1)}_{JS}(X) = \frac{1}{2}\Big[\Psi^{\uparrow}_{(+)}(X) + \epsilon\Psi^{\downarrow}_{(-)}(X)\Big], \quad \Psi^{(2)}_{JS}(X) = \frac{1}{2}\Big[\Psi^{\downarrow}_{(+)}(X) + \epsilon\Psi^{\uparrow}_{(-)}(X)\Big]. \quad (39)$$

Introducing then reflection and transmission coefficients, one can construct a wave function regular at the singularities, $\Psi^{(r)}_{JS} = \frac{\epsilon K}{(K+\epsilon)^2}\Big[\Psi^{(1)}_{JS} + \epsilon\Psi^{(2)}_{JS}\Big]$, with $K = e^{\pi\omega/\kappa}, \kappa = 1/4M$. Thereafter, one constructs hermitian field operators for the Fock space. Next, one needs the renormalized expectation value of the stress-energy tensor $< T_{\mu\nu} >$ in the "semiclassical" equations of Einstein $G_{\mu\nu} = 8\pi G < T_{\mu\nu} >$. If one assumes that there is a $r = 0$ singularity, then back-reaction will be small in the vicinity of the horizon (at least for massless fields). The spacetime can then be approximated by Schwarzschild geometry. The mass will decrease slowly with time and evaporates. In a flat spacetime, this is easily done, because the vacuum is well defined. One can calculate the zero-energy state and can construct finite quantum operators. In curved spacetime, the vacuum state is dependent of the boundary condition for the propagators (positive frequency modes). In principle, we can follow the method of Sanchez (for the de Sitter spacetime) for the dilaton field and our "un-physical" spacetime $\bar{g}_{\mu\nu}$ ($\Lambda = 0$),

$$< \bar{\omega}^2 > \bar{G}_{\mu\nu} =< T^{(\omega)}_{\mu\nu}(\bar{\omega}, \bar{g}_{\mu\nu} > - < \bar{\omega}^2 > \mathcal{E}_{\mu\nu}, \quad (40)$$

where $T^{(\omega)}_{\mu\nu}$ depends on the geometry and boundary conditions (see Eq. (19)). Further, $< T^{(\omega)} >= - < \bar{\omega}^2 > \bar{R}$, because $\mathcal{E}_{\mu\nu}$ is traceless. We have now contributions from the antipode:

$$< T^{(\omega)}_{\mu\nu} > \rightarrow < T^{(\omega)}_{\mu\nu} > \pm < \widehat{T}^{(\omega)}_{\mu\nu} >, \quad < \bar{\omega}^2 > \rightarrow < \bar{\omega}^2 > \pm < \widehat{\bar{\omega}^2} > . \quad (41)$$

In the simplified de Sitter space, one then easily construct Green functions[9]

$$G_{\alpha JS}(X, X') = e^{2\alpha}\Big[G(X.X') + G(X, JX')\Big],$$

$$G_{\alpha JA}(X, X') = e^{2\alpha}\Big[G(X.X') - G(X, JX')\Big], \quad (42)$$

with α labels the one parameter family of the de Sitter vacua. The expectation values for a scalar field and the energy momentum tensor can then be calculated. One obtain, for example,[8]

$$< \widehat{\Phi^2_{JS,JA}} >= \frac{1}{16\pi\cos\pi\nu}\left[m^2 + \left(\xi - \frac{1}{6}\right)R\right], \quad (43)$$

with $\nu = (9/4 - M^2/H^2)^{1/2}, M^2 = m^2 + \xi R$, m the mass of the field and $H \sim \Lambda$. In our case we have no scalar field, but instead ω. The expression for $T_{\tilde{U}\tilde{U}}$ becomes[30]

$$T^{(\omega)}_{UU} = \frac{c_1 e^{-2c_1 U}}{c_3^2 c_4^2 (c_2\rho + c_3)^2}(c_2 c_3 \rho^2 F(U) - c_1^2 c_2^2 \rho^2 + c_1 c_3^2), \quad (44)$$

which can be used to evaluate the expectation value. In order to apply the full antipodal map, one includes the φ-dependency in the dilaton equation. The relevant operator (d'Alembertian) can be separated in the used coordinate system. The relevant φ contribution comes from periodic Mathieu functions (in variable φ). They converge uniformly on all compact sets in the z-plane. Next, one applies the method of 't Hooft, by expanding the position variables $u^{\pm}(z,\varphi)$ and momentum distributions $p^{\pm}(z,\varphi)$ in the partial waves of Mathieu functions.[f] Further, one then calculates the gravitational shift $\delta\tilde{U}(z,\varphi)$, in order to carry a particle over from I to II, or back,[6] using the Shapiro delay.

4.2. *The surface gravity and the conformal gauge*

Since we have now the description of the antipodal map in our black hole spacetime, we will look more closely at the conformal invariance. First of all, one should rely in the dynamical situation on (conformal) Killing vectors in order to describe the spacetime symmetries. Our Lagrangian is conformal invariant under Eq. (14), so we can use the freedom of the conformal factor Ω. Remember, different ω means different notion of the vacuum state for the in-going and outside observer, so they will use different conformal gauge freedom. It is desirable that for the out-going observer, the surface gravity of the horizon is conformal invariant. Further, conformal transformations must preserve affinely parameterized null geodesics. This will deliver Ω for the in-going observer. We can define out-going and in-going null normals[18] for $\bar{g}_{\mu\nu}$

$$\begin{aligned} \bar{l}^{\mu} &= (1, N\sqrt{N^2 - r^2N^{\varphi 2}}, 0, 0), \\ \bar{m}^{\mu} &= \Big(-\frac{1}{2r^2N^{\varphi 2} - 2N^2}, -\frac{N}{2\sqrt{N^2 - r^2N^{\varphi 2}}}, 0, 0\Big). \end{aligned} \tag{45}$$

with $\bar{l}^{\mu}\bar{l}_{\mu} = \bar{m}^{\mu}\bar{m}_{\mu} = 0$, $\bar{l}^{\mu}\bar{m}_{\mu} = -1$. The surface gravity then becomes

$$\kappa = 2N\Big(\partial_r(\sqrt{N^2 - r^2N^{\varphi 2}}) + \partial_t\Big(\frac{1}{N}\Big)\Big) = 2N\Big(\partial_t\sqrt{\bar{g}_{rr}} - \partial_r\sqrt{\bar{g}_{tt}}\Big). \tag{46}$$

This is consistent with the metric definition of κ.

4.3. *The meaning of the warped spacetime*

Let us now return to $g_{\mu\nu} = \omega^{4/3}\bar{\omega}^2\bar{g}_{\mu\nu}$. In the CDG setting, the evaporation of the black hole is also determined by the complementarity transformation of ω between the in-going and outside observer. Our spacetime is now ($b_4 = b_3b_2$)

$$\begin{aligned} ds^2 = \omega^{4/3}\bar{\omega}^2\Big[(C_2(t+b_3)^4 + C_3)\frac{(10b_2^3r^2 + 20b_2^2r^3 + 15b_2r^4 + 4r^5 + C_1)}{5r^2}\cdot \\ \Big(-dt^{*2} + dr^{*2}\Big) + dz^2 + r^2\Big(d\varphi + \frac{N^{\varphi}}{N_2^2}dt^*\Big)^2\Big], \end{aligned} \tag{47}$$

[f] So the spherical harmonics are replaced by the Mathieu harmonics

with

$$\omega^{4/3}\bar{\omega}^2 = \frac{1}{(r+c_2)^2(t+c_3)^2(r+b_2)^2(t+b_3)^2}, \tag{48}$$

We observe that $\omega^{4/3}\bar{\omega}^2$ approaches zero for coordinate time $t \to \infty$, so $g_{\mu\nu}$ shrinks to zero, so the distant observer sees a gradually shrinking black hole when the metric time runs to infinity. Further, the only contribution from the 5D spacetime is the $\omega^{4/3}$. Remarkable, the projected Weyl component is necessary in order to obtain the same form of N^2 and to avoid naked singularities. So $\omega^{4/3} = [(r + c_2)^2(t + c_3)^2]^{-1}$ is the "scale" term from the 5D warped spacetime (the warpfactor in the RS model is the product of y-dependent part and ω part). Suppose one wants combine the conformal transformation with an internal symmetry transformation, i.e., a spacetime transformation. In particular, the scale transformations. One can proof in that case, $\Box \log \Omega = 0$,[19] which is consistent with our 2D null hypersurface of Eq. (32). Further, in dimension $n \neq 4$ only the scale-invariant theories based upon scalar fields (so ω from 5D) are conformally invariant. Conclusion: $\bar{\omega}$ of our $\bar{g}_{\mu\nu}$ can be used in non-vacuum models. An additional advantage of the warped spacetime in connection with cosmology and hierarchy problem, was already mentioned in the introduction. A new aspect will be the embedding of the 5D in the 4D spacetime and the relation with the 3D BTZ blackhole solution.

4.4. *The relation with the 3D Baňados-Teitelboim-Zanelli black hole*

In the spacetime under consideration, the dz^2 term can be omitted. One obtains then the 3D Baňados-Teitelboim-Zanelli (BTZ) black hole spacetime. It solves the Einstein equations with a negative cosmological constant.[29] The BTZ solution is related to the AdS/CFT correspondence and intensively studies in connection with black hole entropy issues. However, we should like to take the cosmological constant zero. In a former study,[30] an exact solution was found in a CDG setting in Eddington-Finkelstein retarded coordinates (U, ρ) (or advanced V) where the antipodal map $(U, V, \varphi) \to (-U, -V, \varphi + \pi)$ is applicable:

$$ds^2 = \frac{e^{-2c_1 U}}{(c_2\rho + c_3)^2}\Big[\pm\frac{c_1(c_3^2 - c_2^2\rho^2)}{c_2 c_3}dU^2 - 2dUd\rho + dz^2 + \rho^2\Big(d\xi + F(U)dU\Big)^2\Big], \tag{49}$$

which is Ricci flat, while $\tilde{R}^{(4)} = \frac{6c_1c_2}{c_3}$. The function $F(U)$ will be fixed when matter terms are incorporated (i.e. for example, a scalar gauge field). The metric Eq. (49) will then contain a term $b(U, \rho)^2 d\varphi^2$ and a relation like $(N^\xi)' = \frac{b}{\eta^2 X^2 + \omega^2}$ will be obtained. It has no curvature singularity. The location of the apparent horizon in U:

$$\rho_{AH} = \pm\frac{c_3}{\sqrt{c_2(c_2 + \frac{c_3}{c_1}F(U)^2)}}, \tag{50}$$

with

$$\frac{d\rho}{dU} = \frac{1}{2} e^{-2c_1 U} \cdot \begin{cases} -\frac{c_1}{c_2 c_3} & \ldots\ldots\ldots \rho \to 0 \\ \frac{c_3 F(U)^2 + c_1 c_2}{c_2^2 c_3} & \ldots \rho \to \infty \\ 0 & \ldots\ldots\ldots\ldots \rho = \rho_{AH} \end{cases} \tag{51}$$

which is independent of ω. Here C_i are constants and $F(U)$ a function determined by the non-diagonal contribution. Further, we have

$$\lim_{\rho \to 0} g_{UU} \to \pm \frac{c_1}{c_2 c_3 e^{2c_1 U}}, \tag{52}$$

So when the evaporation speeds up, it approaches zero. We are dealing here with null-radiation in the (ρ, z)-plane. One could compare this solution with that found by Chan[32] in standard GR of a spinning black hole. They also find a solution for $F(U)$ which is determined by an energy-momentum tensor of null spinning dust. It is again curious that the "uplifted" BTZ has the same solution, comparable with the "up-lifting" 5D[30]

5. Metric Fluctuation and Hawking Radiation

In the original deviation of the Hawking radiation, one uses the propagation of a linear quantized field in a classical background metric. However, near the horizon, high-frequencies metric fluctuations can contribute to the vacuum polarization and the impact of gravitational back reactions can be large. These zero-point fluctuations result in a modification of the Hawking radiation by gravitational waves.[31] One could question what the effect is of these waves in our CDG model, where we have instead the dilaton field. Of course, one should need a quantum gravitational approach, which is not available yet. So need some approximation. However, effect of the scattering of these quanta at the horizon can be investigated in the context of the antipodal mapping considered here.[g] Without the contribution of the metric fluctuations, the mean number of quanta reaching $\mathcal{J}^+$ takes the form

$$< \bar{n}_\lambda >_0 \sim \frac{1}{e^{2\pi E_+/\kappa} - 1}, \tag{53}$$

with E_+ the energy measured at $\mathcal{J}^+$ for the out modes. This is the Planck distribution with temperature $T = \kappa/2\pi = (8\pi M)^{-1}$. The correction terms can then be calculated by using the s-modes of a quantum massless scalar field and by using the fact that the in-going and out-going modes decouple.[31] One makes use of the mean energy flux, by calculating $dE/dU = 4\pi r^2 < T_{UU} >_{ren}$, where the renormalized surface gravity is used. However, in this approximation, the reflection conditions are at r=0, with in our antipodal map must be revised (we have no inside). We can

[g] A suitable approximation is the high-frequency approximation applied to a Vaidya spacetime, where the not-flat background spacetime is distorted by the gravitational waves.[33] A recent application was provided by Slagter.[34,35]

use the (U,U) energy-momentum component of our model and can apply Eq. (41) for the antipodal contribution.

Notice that the meaning of the local dilaton $\bar{\omega}$, is twofold. First, it determines the metric fluctuations (one also must incorporate in the dilaton equation the φ-dependency). Secondly, The in-going observer will use a different conformal gauge freedom Ω on $\bar{\omega}$ to describe the vacuum. Further, $\bar{\omega}$ is locally unobservable, unless we include metric fluctuations (gravitational waves. It will be necessary to compare this with the usual contribution using the Bunch-Davies method (and to taken into count the antipodal contribution). Note that the outside observer will use a different gauge and he/she experiences a mass $\sim \omega^2 N^2$ and Hawking radiation $\sim \partial_U(\omega^2 N^2)$, while for the in-going observer it is part of his vacuum. On the other hand, the outside observer is not aware of the antipodal identification. One could also say that they disagree about the observed scales. Or differently stated, they disagree about the back reaction from the Hawking radiation.

6. Conclusion and Outlook

We find an exact time dependent solution in the conformal dilaton gravity model on a warped 5D spacetime. The spacetime is written as ${}^{(5)}g_{\mu\nu} = \omega^{4/3\,(5)}\tilde{g}_{\mu\nu}$ and ${}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^{2\,(4)}\bar{g}_{\mu\nu}$. In our model, ω can be seen as the contribution from the bulk, while $\bar{\omega}$ is the brane component. It is conjectured that the different conformal gauge freedom, Ω, the in-going and outside observers possess, can be calculated by demanding a conformal invariant surface gravity and the preservation of affinely parameterized null geodesics. This means that the complementarity is expressed by the different notion of the vacuum state. The solution guarantees regularity of the action when $\omega \to 0$. We don't need a Weyl term in the action (generates negative metric states). In stead, we have a contribution from the bulk, i.e., the electric part of the 5D Weyl tensor. It is remarkable that the 5D field equations and the effective 4D equations can be written for general dimension n, with $n = 4, 5$. The energy-momentum tensor of the time-dependent dilaton, determining also the Hawking radiation, can be calculated exactly. By suitable choice of the parameters, the spacetime $\bar{g}_{\mu\nu}$ can be regular and singular free. This exact solution, nonetheless without mass terms, can be used to tackle the deep-seated problem of the black hole complementarity: the infalling and outside observer experience different ω by the choice of Ω. The next task is to incorporate mass into our model and investigate the dilaton-scalar field interaction. The conformal invariance will then spontaneously be broken.

Appendix: The Quintic Horizon Equation and Related Issues

Our quintic polynomial, determining the horizons,

$$f = r^5 + \frac{15}{4}b_2 r^4 + 5b_2^2 r^3 + \frac{5}{2}b_2^3 r^2 + \frac{C_1}{4} = 0, \tag{54}$$

can be written by a, so-called Tschirnhaus transformation, in the form

$$r^5 - \frac{15b_2}{16}(C_1 + b_2^5)r^2 - \frac{125b_2^3}{256}(C_1 + b_2^5)r - \frac{1}{16}(C_1 + b_2^5)^2 = 0 \qquad (55)$$

By scaling, this form can be reduced to the Bring-Jerrard form $r^5 + r - c$, with c a function of b_2 and C_1.[27] There is an interesting relation between the symmetry group of the icosahedron and our quintic equation. The symmetry group is isomorphic with the Galois group A_5 (of an irreducible quintic polynomial). The icosahedron is dual to the dodecahedron, i.e., their symmetries are isomorphic. The A_5 is interesting in physics, because it is a simple group having no invariant subgroups. It has three orbits, which are invariant under the antipodal map. So the connection with the Möbius group is clear. For details, we refer to Toth.[27] It is conjectured that our quintic polynomial (Eq. (54)) has a deep-seated relation with the 5D spacetime solution. Further, it is remarkable that the resulting quintic equation is independent of the dimension of our manifold ($n = 4, 5$). Moreover, the nice fitting of the antipodal map in our model cannot be a coincident. From Eq. (26) we observe that the derivative of f is $5r(r + b_2)^3$. So it is expected that our quintic equation results from a immersion[h] of a closed surface S in $\mathbb{R}^3$ into $\mathbb{R}^4$. This is currently under investigation by the author.

References

1. Hawking, S. (1975) *Particle Creation by Black Holes.* Comm. Math. Phys. **43**, 199.
2. 't Hooft, G. (1993) *Dimensional Reduction in Quantum Gravity.* Conference on Highlights of Particle and Condensed Matter Physics (SALAMFEST). C930308, 184.
3. Page, D. N. (1993) *Informatgion in black hole radiation.* Phys. Rev. Lett. **71**, 3743.
4. Almheiri, A., Marolf, D., Polchinski, J. and Sully, J. (2013) *Black Holes: Complementarity or Firewalls?.* JHEP **02**, 62.
5. 't Hooft, G. (2016) *Black Hole Unitarity and Antipodal Entanglement.* Found. Phys. **46**, 1185.
6. 't Hooft, G. (2018) *The Firewall Transformation for Black Holes and Some of its Implications.* arXiv: gr-qc/161208640v3.
7. 't Hooft, G. (1984) *Ambiguity of the Equivalence Princple and Hawking's Temperature.* J. Geom. and Physics **Volume 1, Issue 1**, 45.
8. Sanchez, N. and Whiting, B. F. (1987) *Quantum Field Theory and the Antipodal Identification of Black Holes.* Nucl. Phys. **B283**, 605.
9. Sanchez, N. (1986) *Two- and Four-Dimensional Semi-Classical Gravity and Conformal Mappings.* Cern-Th.4592/86.
10. Folacci, A. and Sanchez, N. (1987) *Quantum Field Theory and the Antipodal Identification of de Sitter Space. Elliptic inflation.*
11. Schrödinger, E. (1957) *Expanding Universe.* Cambridge Univ. Press, Cambridge.
12. 't Hooft, G. (2018) *Discreteness of Black Hole Microstates.* arXiv: gr-qc/180905367v2.
13. 't Hooft, G. (2018) *Virtual Black Holes and Spacetime Structure.* Found. Phys. **48**, 1149.

[h] An immersion is a differentiable function between differentiable manifold whose derivative is everywhere injective. It is also a topological embedding.

14. 't Hooft, G. (2015) *Diagonalizing the Black Hole Information Retrieval Process.* arXiv: gr-qc/150901695.
15. Zakharov, V. D. (1973) *Gravitational Waves in Einstein's Theory.* John Wiley & Sons, Inc., New York.
16. Wald, R. M. (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* The Univ. of Chicago Press, Chicago.
17. Alvarez, E., Herrero-Valea, M. and Martin, C. P. (2014) *Conformal and Non Conformal Dilaton Gravity* JHEP **10**, 214.
18. Slagter, R. J. (2021) *Conformal Dilaton Gravity and Warped Spacetimes in 5D* arXiv: 2012.00409.
19. Felsager, B. (1998) *Geometry, Particles and Fields.* Springer, New York.
20. Mannheim P. D. (2005) *Alternatives to dark matter and dark energy.* Prog. Part. Nucl. Phys. **56**, 340, arXiv: astro-ph/0505266v2.
21. Slagter, R. J. and Pan, S. (2016) *A new fate of a warped 5D FLRW model with a U(1) scalar gauge field.* Found. of Phys. **46**, 1075.
22. Maldacena, J., Milekhin, A. (2021) *Humanly traversable wormholes.* Phys. Rev. D **103**, 066007.
23. Maldacena, J. (2011) *Einstein Gravity from conformal gravity.* arXiv: hep-th/ 11055632.
24. Shiromizu, T., Maeda, K., Sasaki, M. (2000) *The Einstein equations on the 3-brane world.* Phys. Rev. D **62**, 024012.
25. Shiromizu, T., Maeda, K., Sasaki, M. (2003) *Low Energy Effective Theory for Two Branes System-Covariant Curvature Formulation.* Phys. Rev. D **7** 084022.
26. Maartens, R. (2010) it Brane-World Gravity. arXiv: hep-th/10043962V2.
27. Toth, G. (2002) *Finite Möbius Groups, Minimal Immersions of Spheres and Moduli.* Springer, Heidelberg.
28. Strauss, N. A., Whiting, B. F. and Franzen, A. T. (2020) *Classical Tools for Antipodal Identification in Reissner-Nordstrom Spacetime.* arXiv: gr-qc/200202501.
29. Compère, G. (2019) *Advanced lectures on General Relativity.* Lecture notes in Physics 952, Springer, Heidelberg.
30. Slagter, R. J. (2019) *On the Dynamical BTZ Black Hole in Conformal Gravity.* In Spacetime 1909-2019, Proc. Second H. Minkowski Meeting, Bulgaria, Albena.
31. Barrabès, C. and Hogan, P. A. (2013) *Advanced General Relativity.* Oxford Univ. Press, Oxford.
32. Chan, J. S. F., Chan, K. C. H. and Mann, R. B. (1996) *Interior Structure of a Spinning Black Hole in 2+1 Dimensions.* Phys. Rev. D **54** 1535 (arXiv: gr-qc/9406049).
33. Choquet-Bruhat, Y. and Geroch, R. P. (1969) *Global Aspects of the Cauchy Problem in General Relativity.* Commun. Math. Phys. **14**, 329.
34. Slagter, R. J. and Miedema, P. G. (2020) *On the Azimuthal Alignment of Quasars Spin Vector in Large Quasar Groups and Cosmic Strings.* Mon. Not. R. Astron. Soc., **459** 3054.
35. Slagter, R. J. (2021) *New Evidence of the Azimuthal Alignment of Quasars Spin Vector in the LQG U1.28, U1.27, U1.11, Cosmologically Explained.* arXiv: gr-qc/210212805.

From neutrino masses to the full size of the universe — Some intriguing aspects of the tetron model

B. Lampe

Theoretical Physics, Universität Hamburg, Hamburg, Germany
E-mail: Lampe.Bodo@web.de

The universe according to the tetron model consists of invisible tiny constituents, elastically bound with bond length about the Planck length and binding energy the Planck energy. A tetron transforms as the fundamental fermion(=octonion) representation **8** of SO(6,1). With respect to the decomposition $SO(6,1) \to SO(3,1) \times SO(3)$ it possesses spin 1/2 and isospin 1/2, i.e. a tetron represents an isospin doublet of Dirac spinors. The 24 known quarks and leptons arise as eigenmode excitations of a tetrahedral fiber structure, which is made up from 4 tetrons (plus 4 antitetrons) and extends into 3 additional 'internal' dimensions. While the laws of gravity are due to the elastic properties of the tetron bonds, particle physics interactions take place within the internal fibers. I will concentrate on two of the most intriguing features of the model:
- understanding small neutrino masses from the conservation of isospin, and, more in general, calculating the spectrum of quark and lepton masses. This is obtained from the tetron model's interpretation of the Higgs mechanism.
- the possibility to determine the full size of the universe from future dark energy measurements. This is obtained from the tetron model's interpretation of the dark energy phenomenon.

Finally, the dark energy equation of state, i.e. the equation of state of the tetron background will be derived.

Keywords: Tetron model; quark and lepton masses and mixings; neutrino masses; full size of the universe; dark energy equation of state.

1. Introduction

In the following sections some implications of the tetron model[1–4] of particle physics and cosmology will be presented.

In this model the universe itself is an elastic medium composed of invisible constituents which are bound at Planck energy. While the laws of gravity are due to the elastic properties of the medium, particle physics interactions take place within internal fibers, with the characteristic internal energy being the Fermi scale. All ordinary matter quarks and leptons are constructed as quasiparticle excitations of this internal fiber structure. Since the quasiparticles fulfill Lorentz covariant wave equations, they perceive the universe as a 3+1 dimensional spacetime continuum lacking a preferred rest system. Any type of mass/energy induces curvature on the spacetime continuum as determined by the Einstein equations.

Now in the following I want to explain some details of these statements.

2. Quark and Lepton Masses from Excitations of Internal Tetrahedrons

In the tetron model the ground state of our universe looks like illustrated in Figure 1. In this figure the large horizontal arrow stands for the 3 dimensions of physical space, while the tetrahedrons extend into 3 additional internal dimensions. The picture is a little misleading because in the tetron model physical space and internal dimensions are assumed to be completely orthogonal. This means the whole game is actually played within a large altogether 6 dimensional space, 3 physical dimensions and 3 internal ones.

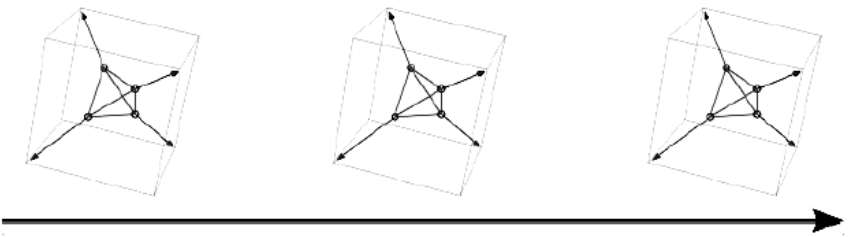


Fig. 1. The global ground state of the tetron model after the electroweak symmetry breaking.

If you ask: why this structure?, I can say at this point that the tetrahedral structure is introduced in order to explain the observed quark and lepton spectrum, which means to get exactly 24 excitation states with the correct multiplet structure.

Before discussing the excitations, let us first consider the ground state Figure 1 in some more detail. As you can see, each tetrahedron is made up from 4 constituents called 'tetrons', depicted as the white dots[a]. The white arrows denote the 'isospins', i.e. internal spin vectors of the tetrons. This means the tetrons have a spin in physical space and in addition an internal spin in internal space. As turns out, the interactions of the internal spins play an important role for particle physics and for electroweak symmetry breaking.

It is important to note, that the 4 isospin vectors in Figure 1 define tetrahedrons, too. Due to the pseudovector property of these vectors their tetrahedral symmetry group actually is a Shubnikov group.[5–7] This means, while the coordinate symmetry is S_4, the arrangement of isospin vectors respects the tetrahedral Shubnikov symmetry

$$G_4 := A_4 + CPT(S_4 - A_4) \tag{1}$$

where $A_4(S_4)$ is the (full) tetrahedral symmetry group and CPT the usual CPT operation except that P is the parity transformation in physical space only. Since

[a] Actually, antitetrons are needed as well. For details see the review.[3]

the elements of $S_4 - A_4$ contain an implicit factor of internal parity, the symmetry (1) certifies CPT invariance of the local ground state in the full of R^{6+1}.

2.1.

In the situation depicted in Figure 1 a symmetry breaking has already taken place, because the isospins are aligned between all the tetrahedrons.

Before the symmetry breaking, which means above a certain temperature, isospins are distributed randomly, corresponding to a local $SU(2) \times U_1$ symmetry[b], but when the universe cools down, there is a phase transition, and the isospins freeze into the aligned structure, breaking the symmetry from $SU(2) \times U_1$ to the discrete 'family group' G_4. And the important point to note is, this temperature can be identified with the Fermi scale.[2]

How does this work out in detail? Mathematically, a tetron is assumed to transform as the fundamental spinor representation of SO(6,1). This representation is 8-dimensional and sometimes called the octonion representation.

With respect to the decomposition of $SO(6,1) \to SO(3,1) \times SO(3)$ into the 3-dimensional base space and the 3-dimensional internal space, a tetron possesses spin 1/2 and isospin 1/2. This means it can rotate both in physical space and in internal space, and corresponds to the fact that a tetron Ψ decomposes into an isospin doublet $\Psi = (U, D)$ of two ordinary SO(3,1) Dirac fields U and D.

$$8 \to (1,2,2) + (2,1,2) = ((1,2) + (2,1), 2) \tag{2}$$

Using this, one can now rigorously define the isospin vectors used and drawn in Figure 1:

$$\vec{Q} = \Psi^\dagger \vec{\tau} \Psi \tag{3}$$

where τ are the internal spin Pauli matrices. Actually one has to distinguish chiral isospin vectors

$$\vec{Q}_L = \frac{1}{2}\Psi^\dagger(1-\gamma_5)\vec{\tau}\Psi \qquad\qquad \vec{Q}_R = \frac{1}{2}\Psi^\dagger(1+\gamma_5)\vec{\tau}\Psi \tag{4}$$

and for tetrons and antitetrons, but this complication will not be discussed in the present talk.[3]

A typical interaction Hamiltonian between such isospin vectors of 2 tetrons a and b looks like this

$$H_H = -J\,\vec{Q}_a \vec{Q}_b \tag{5}$$

So it has the form of a Heisenberg interaction - but for isospins, not for spins. The coupling J is called the 'isomagnetic exchange coupling'.

The mathematical treatment of the excitations arising from (5) is similar to that of magnons in ordinary magnetism, but the physics here is much different. Since

[b]Weak parity violation, i.e. the appearance of $SU(2)_L$, is discussed in.[3]

the isospin vibrations take place in internal space, they are pointlike in physical space. Furthermore, they inherit their fermion nature from the fermion property of the vibrating tetron.

In reality, the Hamiltonian is somewhat more complicated than (5) due to the appearance of antitetrons and of the fact that inner- and inter-tetrahedral interactions are present, the inner ones with 'exchange coupling' $j = -O(1)$ GeV (strong interaction scale) and the inter ones with $J = O(100)$ GeV (weak interaction scale). j and J have different sign, because j leads to the frustrated 'antiferromagnetic' ground state of a single tetrahedron, while J is responsible for the 'ferromagnetic'[c] alignment of neighboring tetrahedrons.[1]

This alignment can be shown to be the microscopic origin of the electroweak symmetry breaking, and furthermore it allows to calculate the quark and lepton masses. I refer here to references[1] and[2]. In these papers you can find all the details - how to construct the electroweak order parameter, the Higgs field and how to calculate the quark and lepton masses and mixings from the isospin couplings.

At this point it must be enough to show that among the 24 isospin excitations which are the quarks and leptons, there are 3 almost massless modes which correspond to the neutrinos. This has to do with the conservation of total isospin. Namely, the masses of the neutrinos are particularly suppressed because the 3 neutrino modes correspond to the vibrations of the 3 components of the total internal angular momentum vector in one tetrahedron

$$\vec{\Sigma} := \sum_{a=1}^{4} \vec{Q}_a = \sum_{a=1}^{4} \Psi_a^\dagger \vec{\tau} \Psi_a \tag{6}$$

Whenever this quantity is conserved

$$d\vec{\Sigma}/dt = 0 \tag{7}$$

the neutrino masses will strictly vanish. In fact, the Heisenberg type of interactions (5) conserve total internal angular momentum. Therefore, they fulfill (7) and give no contribution to the neutrino masses. Further details can be found in.[1]

So $\vec{\Sigma}$ is the total internal angular momentum (total isospin), and the conservation equation for the 3 components of $\vec{\Sigma}$ leads to 3 of the 24 eigenmodes being massless. Contributions to the neutrino masses come from tiny torsional interactions which violate the conservation of isospin.[1]

3. The Full Size of the Universe from a Simple Spring Model

From this point on, I do not want to give more details on the particle physics implications, but want to concentrate on gravity and cosmological aspects. In order to include gravity in the tetron model, it is assumed that there is not only an interaction among the isospin vectors in internal space, but also a binding among tetrons

[c] I am using the language of magnetism, although interactions of isospins and not of spins are considered here.

in physical space, and that this binding is elastic. In other words, our universe is a 3-dimensional elastic medium expanding within some larger 6-dimensional space, and it can acquire curvature both in space and time.

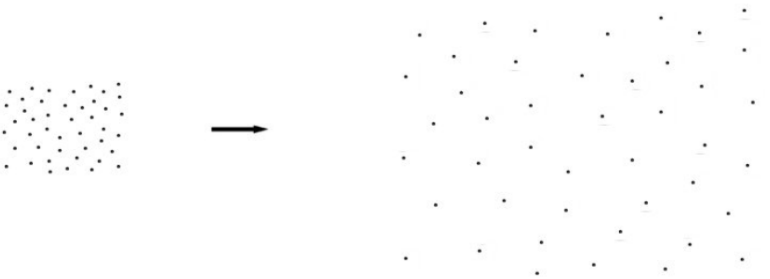


Fig. 2. Expansion of the empty universe consisting of tetrahedrons which look pointlike in physical space.

In 3-dimensional physical space, the expansion looks as shown in Figure 2. Note that in physical space the tetrahedrons are pointlike because they extend only into the internal dimensions. This means, you do not see the tetrahedral structure, you only see points which are bound with bond length about the Planck length and binding energy the Planck energy.

In the beginning, that means before the expansion started, the universe was created in a sudden so to say inflationary condensation process from an ultrahot tetron gas and afterwards was pushed to expansion by the condensation energy. Quark and matter particles were created as quasiparticle excitations of the tetrons gliding on the elastic medium. Since the quasiparticles fulfill Lorentz covariant wave equations, they perceive the universe as a 3+1 dimensional spacetime continuum lacking a preferred rest system. Any type of mass/energy induces curvature on the spacetime continuum as determined by the Einstein equations.

Now in the next step I want to draw a connection between the smallest and the largest scales of the universe, namely between the tetron binding structure at Planck length and the size of the universe as a whole. For that purpose, let us consider the binding energy of 2 tetrahedrons as a function of the bond length L, so this means the binding energy of 2 dots in Figure 3 as a function of their distance. It is assumed, that this function has a minimum at some bond length L_s and that at present we are at bond length L_0 roughly equal to the Planck length. Then the function looks like in Figure 3.

From this figure one concludes that the universe is expanding towards an equilibrium corresponding to an average bond length L_s. So the whole universe is carrying out an extremely low frequency breathing vibration $\omega \ll 10^{-10} yrs^{-1}$ around L_s, and all the tetron bonds on average are vibrating in accord with the universe. One

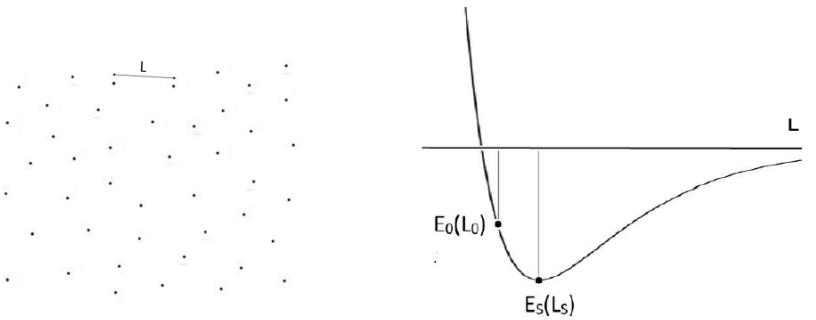


Fig. 3. Binding energy E of 2 neighboring tetrahedrons as a function of their distance L.

can work out the details of this picture and indeed show that this effect accounts for the present accelerated expansion as given by the dark energy observations.[4]

An important point to note is that in future more precise dark energy measurements may allow to extract the breathing frequency ω of the universe, and from ω one can determine the full size radius a of the universe according to the simple formula

$$a = \frac{L}{L_s}\frac{c}{\omega} \tag{8}$$

where c is the speed of light.

3.1.

The remainder of this section is devoted to the proof of (8) within a simple spring type model, and to carry the idea of the universe as an extremely low frequency oscillatory system to the end.

In addition, these considerations will allow to derive the dark energy equation of state,[11] i.e. the equation of state of the invisible tetron background substrate. The relation between tetron density and pressure is a characteristic property of the bound tetron system. Actually, the elastic tetron universe resembles a fluid with elastic bonds among its constituents rather than an ordered solid, and so the fluid equation seems an appropriate way of description[d].

By 'full size' is meant the diameter of the 3-dimensional elastic medium which according to the tetron model is our universe. As a measure of this size I shall take a(t) as appears in the standard FLRW line element

$$ds^2 = -c^2dt^2 + a(t)^2\left[\frac{dr^2}{1-kr^2} + \cdots\right] \tag{9}$$

a(t) is assumed to have dimension of length and dr^2 to be dimensionless.

[d]But note, the stiffness $\zeta = \frac{c^7}{\hbar G^2} \approx 10^{112}\frac{kg}{ms^2}$ of the tetron bonds is large enough, that this fluid pours out into only 3 of the 6 dimensions.

Following reference[4] a harmonic oscillatory term $\sim \omega^2(a - a_s)$ appears in the modified FLRW equations in addition to the cosmological constant Λ

$$\ddot{a} = -\frac{4\pi}{3}G\rho a - \omega^2(a - a_s) + \frac{\Lambda}{3}c^2 a \tag{10}$$

The assumption of a breathing vibration with $a(t) \sim L(t)$ then leads to an identical equation for L(t):

$$\ddot{L} = -\frac{4\pi}{3}G\rho L - \omega^2(L - L_s) + \frac{\Lambda}{3}c^2 L \tag{11}$$

The interpretation of the cosmological constant term within the tetron model will be postponed to later sections. For the moment, let us simply forget about Λ. Then the empty universe behaves harmonic with frequency ω, the reason being that its tetrahedral constituents follow a harmonic elastic interaction, i.e. the binding energy among the internal tetrahedrons in the neighborhood of the minimum at L_s can be approximated by a parabola, see Figure 3.

While characteristic frequencies ω and H_0(=Hubble constant) of the universe are tiny, the frequency $f_s = f(t_s)$ of a single tetron spring is extremely large and given by

$$f(t) = \frac{1}{T(t)} = \frac{c}{L(t)} \tag{12}$$

where $T(t) = L(t)/c$ is the time dependent Planck time.

In accordance with reference[4] I have introduced time dependent Planck quantities

$$L(t) = \sqrt{\frac{\hbar(t)G(t)}{c^3}} \qquad T(t) = \sqrt{\frac{\hbar(t)G(t)}{c^5}} \qquad M(t) = \sqrt{\frac{\hbar(t)c}{G(t)}} \tag{13}$$

with present day values at t=0:

$$L_0 = 1.6 \times 10^{-35} m \qquad M_0 = 2.2 \times 10^{-8} kg \qquad T_0 = 5.4 \times 10^{-44} s \tag{14}$$

Inverting (13), a time dependent Planck and Newton constant is induced

$$\hbar(t)c = E(t)L(t) \tag{15}$$

$$G(t) = c^4 L(t)/E(t) \tag{16}$$

No time dependence of c is indicated, because in the present model there is none - at least if one uses the so-called cosmic coordinates t and r appearing in the line element (9).

3.2.

Coming to the details of the spring model, it is assumed that at each point of the elastic universe each direction can be approximated as a serial connection of N harmonic springs which connect N+1 constituents (= internal tetrahedrons). These

can be thought to lie approximately on a straight line running from one end of the universe to the other. The spatial extension of the universe is then given by

$$a(t) = NL(t) \tag{17}$$

Clearly, this includes the additional very simplifying assumption that the expanding universe has the approximate form of a cube.

The mass μ of any one of the spring chains is given by the sum of all constituent rest masses M_{rest} in the chain, i.e.

$$\mu = NM_{rest} \tag{18}$$

where M_{rest} is the rest mass of one internal tetrahedron. The value of M_{rest} is unknown but should be typically of the order of the Planck mass.

Since in the neighborhood of L_s the binding energy $E(t) = E(L(t))$ is assumed to be quadratic in L, it must have the form

$$E(L) = E_s + \frac{d}{2}(L - L_s)^2 = M_{rest}c^2 - E_{bind} + \frac{1}{2}M_{rest}c^2\left(1 - \frac{L}{L_s}\right)^2 \tag{19}$$

i.e. E(L) is a sum of a constant energy E_s plus a variable component, which vanishes at $L(t_s) = L_s$. E_s comprises the binding energy of 2 tetrahedrons at L_s as well as their possible rest mass

$$E_s = E_{rest} - E_{bind} = M_{rest}c^2 - E_{bind} \tag{20}$$

It is to be noted that (19) does not meet the condition $E(L = \infty) = M_{rest}c^2$, so it holds only in the neighborhood of $L = L_s$. Note further that M_{rest} could in principle be any value, even much larger than E_{bind}. To have a bound state one needs a local minimum of $E(L)$ at L_s but not necessarily a negative $E(L_s)$, i.e. the condition

$$E_{bind} > E_{rest} \tag{21}$$

is not compelling - although present dark energy data seem to give a hint this inequality is true and furthermore in the simple spring model it will be needed for consistency of energy conservation eq. (27).

3.3.

The spring constant of a single spring is $d = M_{rest}f(t_s)^2$ with $f(t)$ from eq. (12). The basic reason why the breathing frequency ω of the universe is so small whereas the fundamental frequency $f = 1/T$ of its tetron constituents is so large, arises from the following fact: Consider one chain of strings stretching from one end of the universe to the other, each spring with a constant d. Then the serial connection of N springs is itself a harmonic oscillator with a much smaller combined spring constant

$$D = d/N \tag{22}$$

Note that the springs connected in parallel belong to different chains and do not contribute to the effective overall chain constant D.

Using $D = \mu\,\omega^2$ one can express the (extremely small) frequency ω of the universe in terms of the (extremely large) Planck frequency $f_s = f(t_s)$:

$$\omega^2 = \frac{D}{\mu} = \frac{d/N}{NM_{rest}} = \frac{1}{N^2} f_s^2 = \frac{c^2}{N^2 L_s^2} \tag{23}$$

Thus, the full extension (17) of the universe at equilibrium can be given as

$$a_s = NL_s = \frac{c}{\omega} \tag{24}$$

in a similar way as the observable(=Hubble) radius is given as c/H_0.

The present size of the universe is somewhat smaller than a_s and given by (8). These equations show that a precise enough measurement of the dark energy effect (i.e. of ω) can lead to an estimate for the full size of the universe.

3.4.

The spring model is also of use to better understand the FLRW theory within the general elasticity ansatz. The essence of the FLRW model is contained in the following equations
(i) the Friedmann equation for the Hubble parameter $\dot{a}/a$ and
(ii) the 'fluid equations' for the densities ρ_{tet} and ρ_{mat} of tetrons and (ordinary and dark) matter, respectively. Both, the tetron substrate and the matter content of the universe are assumed to be separate uniformly distributed perfect fluids with mass energy densities $\rho_{tet}(t)$ and $\rho_{mat}(t)$ and pressure $p_{tet}(t)$ and $p_{mat}(t)$.

3.4.1.

In the present model one can write down an equation for the conservation of energy in a matter-free spacially flat tetron universe

$$N^3 \frac{M_{rest}}{2} \dot{a}^2 + 3N^3 \frac{M_{rest} c^2}{2} \left(1 - \frac{a}{a_s}\right)^2 = -N^3 E_s \tag{25}$$

the conserved energy being the kinetic plus potential energy of the system of N^3 masses M_{rest} and $3N^3$ springs which make up the empty universe. The system is furthermore assumed to be in a breathing mode. Then, the difference in neighboring spring positions $x_{n+1}(t) - x_n(t)$ within one of the spring chains is n-independent and given by $L(t) - L_s$ for each spring n, and the velocity difference by $\dot{x}_{n+1} - \dot{x}_n = \dot{L}(t)$.

Comparing (25) to the corresponding matter-free FLRW equation

$$\frac{1}{2}\dot{a}^2 + \frac{3c^2}{2}\left(1 - \frac{a}{a_s}\right)^2 - \frac{1}{6}\Lambda c^2 a^2 = 0 \tag{26}$$

one finds 3 differences:
- There is no Λ-term in (25). Any cosmological constant contribution from tetrons has been removed in (25), because for the matter free elastic tetron substrate the binding forces do not drive the universe with Λa^2 to infinity, but with $\omega^2(a-a_s)^2$ to a_s. As will turn out later in section 3.4.3, in the presence of matter a cosmological constant reappears because the time dependent[4] Newton constant (16) implies a time dependent matter contribution to the cosmological constant.[8–10]

- There is a nonvanishing constant contribution $\sim E_s$ on the rhs, where there is 0 in FLRW. This has to do with the fact that the Friedmann equation only counts contributions to the spacetime curvature, where a constant binding energy of the tetron constituents does not matter. The FLRW model does not know about the cosmic constituents and uses the freedom to put their background binding energy to zero. In other words, a constant (like the binding energy $E_s = E(L_s)$ at equilibrium $a = a_s$) does not contribute to any cosmic expansion. E_s is the background tetron binding energy which leads to the flat elastic substrate being built from free tetrons, but does not at all contribute to its possible curvature or expansion.

- In order to have consistency with aforegoing formulas, in (26) I have pulled out a factor of 3 from the definition of the cosmic frequency ω.

3.4.2.

So far we have considered an empty universe built up from an elastic substrate of bound tetrons (or, more precisely, of bound internal tetrahedrons). If, in addition, matter is present in the universe, the modified Friedmann equation becomes

$$\frac{1}{2}\dot{a}^2 + \frac{3}{2}\omega^2(a-a_s)^2 = \frac{4\pi}{3c^2}G\rho_{mat}a^2 + \frac{1}{6}\Lambda_{mat}c^2a^2 - \frac{E_s}{M_{rest}} \tag{27}$$

where ρ_{mat}/c^2 denotes the mass density of ordinary plus dark matter. As before in (25), the constant term $\sim E_s$ is not relevant when forming the time derivative of (27) in order to obtain the equation for the curvature/acceleration $\ddot{a}$.

Equation (27) includes a cosmological constant contribution Λ_{mat} due to ordinary and dark matter. This contribution derives from the time dependence of the Newton constant $G = c^4L(t)/E(L(t))$ and will now be proven to be time dependent, too.

3.4.3.

In order to determine Λ_{mat} and its time dependence, we turn to the cosmological fluid equations. Both the expanding tetron background and the matter distribution will be approximated as separate fluids distributed homogeneously over the universe with energy densities $\rho_{tet}(t)$, $\rho_{mat}(t)$ and pressure $p_{tet}(t)$ and $p_{mat}(t)$.

For the tetronic fluid the appropriate form of the fluid equation is the ordinary one

$$\dot{\rho}_{tet} + 3(\rho_{tet} + p_{tet})\frac{\dot{a}}{a} = 0 \tag{28}$$

because the spring coupling d is constant and the time dependent Newton coupling not involved. In terms of a single tetrahedron with physical volume L^3 around it the density of the tetronic 'dark energy' fluid is

$$\rho_{tet}(t) = \frac{E}{V} = \frac{E_s}{L^3} + \frac{E_{rest}}{2L_s^3}\left(1 - \frac{L(t)}{L_s}\right)^2 \tag{29}$$

where $E_s = E(L_s)$ is the Planck energy at equilibrium and $M_{rest} = E_{rest}/c^2$ is the tetrahedral rest mass, cf. eq. (20).

Similarly, the dark energy pressure is obtained to be

$$p_{tet} = -\frac{\partial E}{\partial V} = \frac{E_{rest}}{3L_s^3}\left(1 - \frac{L}{L_s}\right) + O(L - L_s)^2 \tag{30}$$

and it can be checked that (29) and (30) indeed fulfill eq. (28).

These equations correspond to an equation of state parameter

$$w = \frac{p_{tet}}{\rho_{tet}} = \frac{E_{rest}}{3E_s}\left(1 - \frac{L}{L_s}\right) + O(L - L_s)^2 \tag{31}$$

To compare this with various other dark energy equations of state suggested in the literature one may consult.[11]

3.4.4.

For matter and dark matter the suitable form of the fluid equation can be derived from the Bianchi identity

$$T^{\mu}_{\nu;\mu} = 0 \tag{32}$$

for the energy-momentum tensor in general relativity. In case of a time dependent Newton and cosmological constant one has[8–10]

$$\frac{d}{dt}\left[G\rho_{mat} + \frac{c^4}{8\pi}\Lambda_{mat}\right] + 3G(p_{mat} + \rho_{mat})\frac{\dot{a}}{a} = 0 \tag{33}$$

On the other hand, for the late time cosmology under consideration, matter can be approximated in the standard way as uniformly distributed dust. Ordinary and dark matter should then fulfill the ordinary fluid equation

$$\dot{\rho}_{mat} + 3(p_{mat} + \rho_{mat})\frac{\dot{a}}{a} = 0 \tag{34}$$

by means of

$$p_{mat} = 0 \tag{35}$$

$$\rho_{mat}(a) = \rho_{mat}(a_s)\frac{a_s^3}{a^3} \tag{36}$$

Comparing (33) and (34) one concludes, that any imbalance coming from the time dependency of G must be cancelled by a time dependence of Λ_{mat} according to[8–10]

$$\dot{G}\rho_{mat} + \frac{c^4}{8\pi}\dot{\Lambda}_{mat} = 0 \tag{37}$$

In the present approach all time dependencies arise only through $a(t) = NL(t)$. Therefore using

$$\frac{d}{dt} = \dot{a}\frac{d}{da} = \dot{L}\frac{d}{dL} \tag{38}$$

one can calculate Λ_{mat} from the scale dependence of G eq. (16)

$$\Lambda_{mat}(a) = \Lambda_{mat}(a_s) + 8\pi \int_a^{a_s} da\ \rho_{mat}(a)\frac{d}{da}\frac{L}{|E(L)|} \tag{39}$$

with $L = a/N$. In the quadratic approximation used throughout this paper, where one considers the neighborhood of $a = a_s$, one can carry out the integral in (39) to obtain the a(t) dependence of Λ_{mat}:

$$\Lambda_{mat}(a) = \Lambda_{mat}(a_s) + \frac{8\pi\rho_{mat}(a_s)}{|E_s|}(L_s - L) + O(L - L_s)^2 \tag{40}$$

Here, $\rho_{mat}(a_s)L_s^3/|E_s|$ is the ratio of average energy of matter within a Planck volume L_s^3 over one tetrahedral binding energy E_s. Since there are much more bound tetrons than matter particles in the universe, the cosmological constant due to (40) is extremely small. This can be seen more explicitly by rewriting (40) as

$$\Lambda_{mat}(a) - \Lambda_{mat}(a_s) = 8\pi\frac{\rho_{mat}(a_s)}{\rho_{tet}(a_s)}\frac{L_s - L}{L_s^3} \approx 8\pi\frac{\rho_{mat}(a_0)}{\rho_{tet}(a_0)}\frac{L_0 - L}{L_0^3} \tag{41}$$

and using approximate values $L_0 - L \approx -L_0$ and

$$\rho_{mat}(a_0)/c^2 = 2.6\ 10^{-27}\frac{kg}{m^3} \tag{42}$$

$$\rho_{tet}(a_0)/c^2 = \frac{M_0}{L_0^3} = 0.54\ 10^{97}\frac{kg}{m^3} \tag{43}$$

3.5.

In the literature one often finds the argument that there should be enormous contributions to the cosmological constant from zero point fluctuations of the matter fields. This is 'proven' by summing up the zero point energies $\hbar\omega_k/2$ of all existing fields over all modes with wave vectors k

$$\rho_{quant} = \frac{\hbar}{2}\int \frac{d^3k}{(2\pi)^3}\omega_k \tag{44}$$

and using the Planck scale as UV cutoff. This leads to an enormous number of the order of $\rho_{quant} \sim E_s/L_s^3$ - actually the same order as ρ_{tet} eqs. (29) and (43) which is no wonder, because both the claimed quantum effects and the tetron properties are derived from the Planck scale.

If true, such a huge contribution ρ_{quant} to the cosmological constant would immediately tear the universe into pieces. This then is usually considered a fundamental unsolved problem and runs under the name 'cosmological constant problem'. The suggested solutions often lead to a cascade of fine tuning problems.

In the tetron model the situation is much simpler. As discussed before, the Planck energy $E_0 \sim E_s$ contained in the neighborhood volume L_0^3 of one bound constituent is the binding energy of tetrons in an empty flat spacetime and does not affect the curvature(=accelerated expansion), because it is a constant contribution in the conservation of energy law (27). Furthermore, quantum fluctuations do not exist by themselves at any point in the universe, but are an artefact of the discreteness of the universe[3] and arise only in the process of detecting and measuring a matter field gliding as quasiparticle excitation on the discrete tetron background.

For these reasons the cosmological constant does not get a contribution from quantum fluctuations.

4. Remark on the Strong Interaction

In the preceeding sections we have obtained an understanding of gravity as well as of the electroweak symmetry breaking. How does the strong interaction among quarks fit into the tetron picture?

The dominant features of the strong interaction are the linear attractive potential at low energies and asymptotic freedom at high energies. In the tetron model the strong interaction is related to disturbances by the triplet isospin excitations(=quarks) of the local frustrated ground state which is formed by a single tetrahedron of isospin vectors. As triplet states of G_4, quarks disturb the ground state's isomagnetism, whereas leptons are G_4-singlets, i.e. 'isomagnetically' neutral. They do not disturb the ground state and therefore do not take part in the strong interaction. As shown in,[1,3] the isomagnetic ground state energy of a single tetrahedron is roughly $E_{QCD} \approx 1$ GeV corresponding to a characteristic length scale $L_{QCD} \approx 10^{-15}$m.

The linear potential between two G_4-triplets, vulgo a quark Q and an antiquark $\bar{Q}$, arises as follows: Since the inner-tetrahedral exchange energy $j = E_{QCD}$ is relatively small, the 'antiferromagnetic' effects leading to the local frustrated tetrahedral state have a much longer range L_{QCD} than the weak interactions which are induced by the 'ferromagnetic' inter-tetrahedral exchange energies $J = O(100)$ GeV. The triplet excitations corresponding to the two quarks are characterized by small vibrations $\vec{\Delta}_Q$ and $\vec{\Delta}_{\bar{Q}}$ of the isospin vectors $\vec{S}$. When the distance between the two excitations becomes larger than L_{QCD}, an additional pair $\vec{\Delta}_q$ and $\vec{\Delta}_{\bar{q}}$ is excited on intermediate tetrahedrons in order to reduce the original 'antiferromagnetic' suspense between $\vec{\Delta}_Q$ and $\vec{\Delta}_{\bar{Q}}$. (I am using the language of magnetism, although interactions of isospins and not of spins are considered here.) The associated cost in energy is proportional to the number of $q\bar{q}$ pairs created, and the potential V between Q and $\bar{Q}$ therefore increases linearly with distance:

$$V = F|x| \qquad\qquad F \approx -j < \vec{\Delta}_Q \vec{\Delta}_{\bar{Q}} > /L_Q \tag{45}$$

where $< \vec{\Delta}_Q \vec{\Delta}_{\bar{Q}} >$ is the isospin correlation between the sites, on which the isospins vibrate, and L_{QCD} the length where all this becomes relevant. The confinement

energy is hence proportional to the original 'antiferromagnetic' exchange energy j induced on the disturbances Q and $\bar{Q}$. The ratio x/L_{QCD} is the number of times, an additional pair of excitations has to be created from the 'sea'.

In summary, a single quark Q increases the energy of the system in its neighborhood L_{QCD} not only by its flavor-dependent mass(=excitation energy) but by an additional energy necessary to 'pick up a $q\bar{q}$ pair from the sea'. This energy is flavor independent, because it does not depend on the flavor Q, which flavors q are excited. The flavors q correspond to an average of the light quarks u, d and s. So when a Q and a $\bar{Q}$ are torn apart, at some distance $x \approx L_{QCD}$ a light $q\bar{q}$ pair is formed, because otherwise the single quark Q could not endure the disturbance of the ground state. In the end a sort of string appears obtained by $Q\bar{q}q\bar{q}'...\bar{Q}$ pairs. Any time a new $q\bar{q}$ pair is created, energy is to be taken from the environment, so the associated cost in energy is proportional to the number of $q\bar{q}$ pairs and the potential between quark and antiquark therefore increases linearly with distance as indicated in (45).

Readers familiar with the quark model, will recognize that one is led directly to the classic ideas of the quark model. For example, using the linear potential (45) masses of mesons and baryons can be estimated just as in the quark model. Since mesons and baryons are G_4-singlets, the 'isomagnetic' disturbances induced by quarks get neutralized in these bound states, i.e. mesons and baryons do not disturb the frustrated 'antiferromagnetic' ground state of a single tetrahedron.

The role of the length L_{QCD} (and the energy E_{QCD}), below (above) which the creation of a light quark-antiquark is enforced, therefore is the same as in the Standard Model. At distances above L_{QCD} one has confinement, while below L_{QCD} the strong force diminishes. Virtual bound quark-antiquark pairs are formed which as gluons mediate an interaction of the original $Q\bar{Q}$ pair which effectively can be described by the QCD Lagrangian. As well known, this interaction dies out when the energies involved go to infinity, i.e. one has asymptotic freedom.

5. Summary

Introducing an additional level of matter, the tetron model offers a unified understanding of particle physics and gravitation. While particle physics processes are induced by 'isomagnetic' interactions in internal fibers with a tetrahedral substructure, gravity arises from elastic forces among the tetrons.

The internal tetrahedral structure is introduced in order to explain the elementary particle spectrum of quarks and leptons (6 singlets and 6 triplets after the spontaneous electroweak symmetry breaking). The 24 states of the 3 quark/lepton families are not truly fundamental particles but can be identified as quasi particle wave excitations of a tetrahedron formed by isospin vectors.

Since the quasiparticles fulfill Lorentz covariant wave equations, they perceive the universe as a 3+1 dimensional spacetime continuum lacking a preferred

rest system. Any type of mass/energy induces curvature on this continuum as determined by the Einstein equations.

The tetron model stands for a material understanding of the hitherto abstract particle symmetries $SU(3) \times SU(2)_L \times U(1)$ of the Standard Model of particle physics as well as of their breaking. In the tetron model the Higgs mechanism arises from the alignment of internal isospin vectors of neighboring tetrahedrons, and the Higgs field corresponds to a joint excitation of a neighboring tetron-antitetron pair. One may draw an analogy between particle physics and superconductivity. In this analogy the tetron model corresponds to the BCS theory of superconductivity, whereas the analog of the particle physics Standard Model in this comparison would be the Landau-Ginzberg approach.

References

1. B. Lampe, Int. J. Mod. Phys. A30 (2015)1550025.
2. B. Lampe, Mod. Phys. Lett. A 28 (2013) 1350018.
3. see the review B. Lampe, arXiv:1505.03477 [hep-ph] (2015).
4. B. Lampe, Progr. of Phys. 69 (2021) 2000072.
5. A.P. Cracknell, Progr. Theor. Phys. 35 (1966) 196.
6. A.S. Borovik-Romanov and H. Grimmer, International Tables for Crystallography D (2006) 105.
7. R.M. White, Quantum Theory of Magnetism, Springer Verlag, ISBN-10 3-540-65116-0.
8. H. Fritzsch and J. Sola, Class. Quant. Grav. 29 (2012) 215002.
9. H. Fritzsch and J. Sola, Mod. Phys. Lett. A, Vol. 30, No. 22 (2015) 1540034.
10. H. Fritzsch, J. Sola and Rafael C. Nunes, Eur. Phys. J. C77 (2017) 193.
11. R. Silva, R. S. Goncalves, J. S. Alcaniz and H. H. B. Silva, Astron. and Astrophys. A11 (2012) 537.

Summary parallel session AT5

Reinoud Jan Slagter
Astronomisch Fysisch Onderzoek Nederland (ASFYON),
University of Amsterdam (on leave)
E-mail: info@asfyon.com

Keywords: Conformal dilaton gravity.

It is believed that the understanding of the quantum mechanical properties of gravity at very high energy is one of the major challenges of modern physics. This circumstance occurs, for example, close to the horizon of a black hole or just after the big bang at the Planck scale. There are many ways to attack the problems. One can modify General Relativity rigorously or moderately. One can try to construct a renormalizable model, by adding fourth order derivatives of the curvature tensor to the Lagrangian, try to construct an effective field theory, apply superstring models and many more. However, all these modifications cause problems, such as the appearance of anomalies, loss of unitarity, loss of renormalizability, or one needs ghost fields and gauge-fixing terms. One could also use conformal dilaton gravity models. Conformal invariance is an exact symmetry and spontaneously broken, comparable with the Higgs mechanism.

In the parallel session diverse alternative models pass the review.

Beckwith uses information from a paper deriving a Lorentz-violating energy-momentum relation entailing an exact momentum cutoff as stated by G. Salesi. Salesi in his work allegedly defines Pre Planckian physics, whereas Beckwith restricts himself to an application to GW generation and DE formation in the first $10^{-39}s$ to $10^{-33}s$ or so in the early universe. This procedure is inacted due to an earlier work whereas referees exhibited puzzlement as to the physical mechanism for release of Gravitons in the very early universe. The calculation is meant to be complementary to work done in the Book "Dark Energy" by M. Li, X-D. Li, and Y. Wang, and also a calculation for Black hole destruction as outlined by Karen Freeze, et. al. The GW generation will be when there is sufficient early universe density so as to break apart Relic Black holes but he claims that this destruction is directly linked to a Lorentz violating energy-momentum G. Salesi derived, which he adopts, with a mass m added in the G. Salesi energy momentum results proportional to a tiny graviton mass, times the number of gravitons in the first 10^{-43} seconds.

Slagter presents an exact time-dependent solution of a black hole is found in a conformally invariant gravity model on a warped Randall-Sundrum spacetime, by

writing the metric $g_{\mu\nu} = \omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}$. Here $\tilde{g}_{\mu\nu}$ represents the "un-physical" spacetime and ω the dilaton field, which will be treated on equal footing as any renormalizable scalar field. In the case of a five-dimensional warped spacetime, he thereafter writes ${}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^{2(4)}\bar{g}_{\mu\nu}$. The dilaton field $\bar{\omega}$ can be used to describe the different notion the in-going and outside observers have of the Hawking radiation by using different conformal gauge freedom. The disagreement about the interior of the black hole is explained by the antipodal map of points on the horizon. The free parameters of the solution can be chosen in such a way that $\bar{g}_{\mu\nu}$ is singular-free and topologically regular, even for $\omega \to 0$. It is remarkable that the 5D and 4D effective field equations for the metric components and dilaton fields can be written in general dimension $n = 4, 5$. From the exact energy-momentum tensor in Eddington-Finkelstein coordinates, he is able to determine the gravitational wave contribution in the process of evaporation of the black hole. It is conjectured that, in context of quantization procedures in the vicinity of the horizon, unitarity problems only occur in the bulk at large extra-dimension scale.

Jain starts his presentation with the proposition that the zero modes can contribute towards divergence in the entanglement entropy and the nature of the divergent term can be either log or log(log). However, a clear understanding of what leads to these two different forms of zero mode divergence is still lacking. So, in order to throw some light along this direction, he investigates how these two different divergent behaviours can be seen as a signature of a crossover in the zero mode limit for the ground state entanglement.

Hartmann investigates an old hypothesis that there exists an intimate connection between weak interaction and gravity, symbolized by the relationship between the Fermi and Newton's constants. He analyzes the hypothesis that the effect of matter upon the metric that represents gravitational interaction in General Relativity is an effective one. He concludes that this leads to consider gravitation to be the result of the interaction of two neutral spinorial fields (g-neutrinos) with all kinds of matter and energy. He presents three examples with only one g-neutrino: two static and spherically symmetric configurations and a cosmological framework for an isotropic dynamical universe. Without self-interaction, the associated effective geometry is precisely the Schwarzschild metric. On the other hand, a self-interacting g-neutrino generates a new gravitational black-hole.

Lampe applies the tetron model of invisible tiny constituents in a cosmological model, elastically bound with bond length about the Planck length and binding energy the Planck energy. A tetron transforms as the fundamental fermion(=octonion) representation 8 of $SO(6,1)$. With respect to the decomposition $SO(6,1) \to SO(3,1) \otimes SO(3)$ a tetron possesses spin 1/2 and isospin 1/2, i.e. it represents an isospin doublet of Dirac spinors. The 24 known quarks and leptons arise as eigenmode excitations of a tetrahedral fiber structure, which is made up from 4 tetrons and extends into 3 additional 'internal' dimensions. While the laws of gravity are due to the elastic properties of the tetron bonds, particle physics interactions take

place within the internal fibers. He concentrates on two of the most intriguing features of the model: - understanding small neutrino masses from the conservation of isospin, and, more in general, calculating the spectrum of quark and lepton masses. This is obtained from the tetron model's interpretation of the Higgs mechanism. The possibility to determine the full size of the universe from future dark energy measurements. This is obtained from the tetron model's interpretation of the dark energy phenomenon.

Lapponi applies the model of a perfectly reflecting accelerating mirror to the null-shell of a collapsing black hole. He then investigates an extension of this model to mirror semi-transparency and derive a general expression for the corresponding Bogoliubov coefficients. He introduces the concept of "impulsive accelerated mirrors", corresponding to those mirrors that are accelerated via an impulsive force. He shows that this treatment guarantees analytic solutions of Bogolubov coefficients. In particular, he evaluates the corresponding particle production from the so-obtained Bogoliubov coefficients. Finally, he recognizes that the mirror, as a Gaussian quantum channel, acts between the spacetime regions of left-past and right-future. As a consequence he studies the loss/amplification properties of this quantum channel, alongside the noise it creates and evaluates its capacities in transmitting classical and quantum information.

Boundary conditions for the Klein-Gordon field on Lifshitz spacetimes

Lissa de Souza Campos

Physics Department, University of Pavia,
Pavia, PV 27100, Italy
E-mail: lissa.desouzacampos01@universitadipavia.it

Boundary conditions have physical consequences. On Lifshitz spacetimes, the Klein-Gordon equation gives rise to an initial-boundary value problem that admits a plethora of physically-sensible boundary conditions. Considering a free, scalar, massive quantum field theory on a four-dimensional Lifshitz spacetime with critical exponent $z = 2$, I layout how to construct two-point functions for ground and thermal states, of local Hadamard form, satisfying the canonical commutation relations, and compatible with Robin and mode-dependent boundary conditions. Each one relates to an inequivalent dynamics, but they are all equivalently physically-sensible—only an experiment could single one out. The results I present here are part of a joint work with C. Dappiaggi and D. Sina.[1]

Keywords: Klein-Gordon equation; Robin boundary conditions; Lifshitz spacetime.

Introduction

In this paper I outline part of a joint work with C. Dappiaggi and D. Sina presented in Ref. 1, elaborating on the talk I have given on the Sixteenth Marcel Grossmann Meeting.[2] The goal is to bring awareness to the fact that on Lifshitz spacetimes there are infinitely many physically-sensible boundary conditions naturally arising within a free, scalar quantum field theoretical framework. In particular, I show how to construct two-point functions for ground and thermal states on a four-dimensional Lifshitz spacetime Lif with dynamical critical exponent $z = 2$ compatible with Robin and mode-dependent boundary conditions.

Lif is a metric-solution of Hořava-Lifshitz gravity.[3] This gravity theory provides an alternative to General Relativity by breaking Lorentz invariance with a Lifshitz scaling between time t and space $\underline{\mathrm{x}}$ of the form:

$$t \mapsto \delta^z t \quad \text{and} \quad \underline{\mathrm{x}} \mapsto \delta \underline{\mathrm{x}} \quad \text{with} \quad z > 1. \tag{1}$$

Although the goal of introducing Lifshitz scaling in the theory was to allow its quantization, i.e. to be a theory of quantum gravity, such anisotropic scaling has manifold applications, from cosmology, through AdS/CFT correspondence, to condensed matter physics.[4] Since its initial proposition by Hořava in 2009,[5] it has gone through several developments,[6] and examples relating the dynamical critical exponent z with "real world" quantum critical systems have been elaborated.[7] Hořava-Lifshitz gravity itself lies beyong the scope of this work; here, the focus is on the quantization of a free scalar field on the background solution Lif.

Let us endow Lif with polar coordinates (t, r, θ, φ), where $t \in \mathbb{R}$, $r \in (0, \infty)$, $\theta \in [0, \pi)$ and $\varphi \in [0, 2\pi)$; its line-element reads

$$ds^2 = -\frac{r^4}{L^4}dt^2 + \frac{L^2}{r^2}dr^2 + r^2 d\theta^2 + r^2\theta^2 d\varphi^2, \tag{2}$$

where $L > 0$ is a length scale. The time coordinate t identifies a global, timelike, irrotational, Killing vector field ∂_t, making Lif a static spacetime. The Lifshitz horizon at $r = 0$ corresponds to a singularity of coordinate type, since neither the Ricci scalar, given by

$$\mathcal{R} = -\frac{22}{L^2}, \tag{3}$$

nor the Kretschmann scalars diverges there. The Lifshitz boundary at $r = \infty$ can be seen as an anisotropic conformal infinity,[8] but it also admits other definitions, see e.g. Ref. 9. I avoid going into details in this regard since I take $r \to \infty$ to be merely an asymptotic radial infinity for the bulk theory, as in Ref. 10. The crucial impact of this asymptotic behavior in the analysis performed here is that it results on Lif being non-globally hyperbolic.

Next, let us consider a free, scalar field $\Psi : \mathsf{Lif} \to \mathbb{R}$ with mass m_0 and coupled to the scalar curvature $\mathcal{R}$ through a parameter $\xi \in \mathbb{R}$. Its dynamics is governed by the Klein-Gordon equation:

$$(\Box - m_0^2 - \xi\mathcal{R})\Psi = 0, \tag{4}$$

where $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$ is the D'Alembertian operator, while g is the metric tensor associated to Equation (2). On Lif, the Klein-Gordon equation gives rise to an initial-boundary value problem. Given suitable initial data on a spacelike surface, corresponding solutions might not exist; if they exist, then each boundary condition selects a different solution, identifying an inequivalent dynamics. To guarantee such initial-boundary value problem is well-posed and physically meaningful, I follow the standard approach of looking for self-adjoint extensions of the radial part of the Klein-Gordon operator. This restriction guarantees the dynamics obtained are physically-sensible, as specified by the work of Ishibashi and Wald.[11,12]

For some boundary conditions, depending on the value of the effective mass, the two-point functions possess bound state contributions: *bound states* are poles of the two-point function. In our case, the set of bound states coincide with the set of poles of the Green function of the radial part of the Klein-Gordon equation. Since we impose self-adjointness on the radial part, they are, in fact, elements of the point spectrum of the radial Green function. Accordingly, they are a subset of the quasi-normal modes. Withall, since the physical meaning of bound state contributions is not well-understood, I rule them out by delineating the boundary conditions for which no bound states arise. Subsequently, by employing a prescription applicable on an static spacetime, as performed in Refs. 13–19, I obtain two-point functions that solve the Klein-Gordon equation in each entry, satisfy the canonical commutation relations and are of local Hadamard form. The analysis I present is consistent

with previous works regarding the scalar field on Lifshitz spacetimes[10, 20–23] and it generalizes them in different aspects: first, it takes into account more general boundary conditions; second, it specifies which of these boundary conditions yields physically sensible two-point functions; third, it provides explicit (though not in closed-form) expressions for the two-point functions.

The layout of this paper mirrors my talk.[2] First, in Section 1, taking into account the symmetries of the spacetime, I show that the Klein-Gordon equation reduces to a radial equation, I find the radial solutions and I study their square-integrability. Second, in Section 2, I construct the radial Green function. Then, in Section 3, I show how we can rule out the bound states and focus on physically-sensible boundary conditions. With all these ingredients in hands, in Section 4, I turn to spectral theory of differential operators and to algebraic quantum field theory to write down well-defined, physically-sensible, explicitly expressions for the two-point functions for the ground and thermal states.

1. Solving the Klein-Gordon equation

Let $Y_\ell^m(\theta,\varphi)$ be the eigenfunctions of the Laplacian operator on $\mathbb{R}^2$ in polar coordinates, with eigenvalues λ_ℓ^m. Since Lif is a static spacetime, consider the following ansatz

$$\Psi(t,r,\theta,\varphi) = e^{-i\omega t} R(r) Y_\ell^m(\theta,\varphi). \tag{5}$$

Substituting Equation (5) in the Klein-Gordon equation (4) entails that $R(r)$ is a solution of the radial equation:

$$AR(r) := R''(r) + \frac{5}{r}R'(r) + \left(\frac{L^6}{r^6}\omega^2 + \frac{L^2}{r^4}\lambda_\ell^m - \frac{L^2}{r^2}\right) R(r) = 0. \tag{6}$$

By applying transformations on both the independent and dependent variables,

$$r \mapsto u = \frac{iL^3\omega}{r^2} \in i(0,\infty) \quad \text{and} \quad R(u) = e^{-u/2}\left(\frac{u}{iL^3\omega}\right)^{\frac{1}{2}(2+\nu)} w(u), \tag{7}$$

the radial equation (6) yields a confluent hypergeometric equation

$$uw''(u) + (b_0 - u)w'(u) - a_0 w(u) = 0, \text{ with } \begin{cases} a_0 := \frac{1+\nu}{2} + \frac{i\lambda}{4L\omega}, \\ b_0 := 1+\nu, \\ \nu := \sqrt{4 + L^2\mu^2}. \end{cases} \tag{8}$$

Therefore, the radial solutions are given in terms of the confluent hypergeometric functions M and U.[24] Numerically satisfactory bases of solutions of Equations (6) and (8), close to the point u_0, are respectively given by $\{R_{1(u_0)}(u), R_{2(u_0)}(u)\}$ and $\{w_{1(u_0)}(u), w_{2(u_0)}(u)\}$, with $w_{j(u_0)}$ as in Table 1 and

$$R_{j(u_0)}(u) = e^{-u/2}\left(\frac{u}{iL^3\omega}\right)^{\frac{1}{2}(2+\nu)} w_{j(u_0)}(u), \text{ for } j \in \{1,2\}. \tag{9}$$

Table 1. Suitable bases of the confluent hypergeometric equation (8).

Endpoint	Suitable Basis
$u = 0 \equiv r = \infty$	$\begin{cases} w_{1(0)}(u) = M(a_0, b_0; u) \\ w_{2(0)}(u) = (iL^3\omega)^{\nu} u^{1-b_0} M(a_0 - b_0 + 1, 2 - b_0; u) \end{cases}$
$u = i\infty \equiv r = 0$	$\begin{cases} w_{1(\infty)}(u) = U(a_0, b_0; u) \\ w_{2(\infty)}(u) = e^u U(b_0 - a_0, b_0; -u) \end{cases}$

To obtain self-adjoint extensions of the radial part of the Klein-Gordon operator, we turn to Sturm-Liouville theory. Rewriting the radial equation as a Sturm-Liouville problem gives

$$AR(r) =: (L_{\omega^2} - \omega^2)R(r) = -\frac{1}{r^{-1}}\left(\frac{d}{dr}\left(\frac{r^5}{L^6}\frac{d}{dr}\right) + \frac{r}{L^4}\left(-\mu^2 r^2 + \lambda_\ell^m\right)\right) R(r) = 0. \tag{10}$$

Taking Equation (10) into account, we look for solutions $R(r)$ in the space of square-integrable functions with respect to the measure $r^{-1}dr$. In addition, the radial operator A is positive and symmetric on the Hilbert space $L^2((0,\infty), r^{-1}dr)$ when we take its domain to be $C_0^\infty((0,\infty))$.

Since the endpoints $r = 0$ and $r = \infty$ are singular, we perform an asymptotic analysis to check square-integrability of the radial solutions, as follows. As $r \to 0$ the radial solutions defined in Equation (9) together with Table 1 behave like

$$|R_{1(0)}(r)|^2 \sim e^{+\frac{L^3}{r^2}\operatorname{Im}(\omega)}, \tag{11a}$$

$$|R_{2(0)}(r)|^2 \sim e^{-\frac{L^3}{r^2}\operatorname{Im}(\omega)}. \tag{11b}$$

As $r \to \infty$, we have instead

$$R_{1(\infty)}(r) \sim r^{-2-\nu}, \tag{11c}$$

$$R_{2(\infty)}(r) \sim r^{-2+\nu}. \tag{11d}$$

According to Weyl's endpoint classification,[25] it follows that $r = 0$ is limit point, and that, if $\nu \in (0,1) \cup (1,2)$, then $r = \infty$ is limit circle. That is, the unique solution $R_0(r)$ that belongs to $L^2((0, r_0), r^{-1}dr)$ for all $r_0 < \infty$ is given by

$$R_0(r) = \Theta(-\operatorname{Im}(\omega))R_{1(0)}(r) + \Theta(\operatorname{Im}(\omega))R_{2(0)}(r). \tag{12}$$

Contrastingly, if $\nu \in (0,1) \cup (1,2)$, there is a one-parameter family of solutions $R_\gamma(r)$, with $\gamma \in [0, \pi)$, belonging to $L^2((r_0, \infty), r^{-1}dr)$ for all $r_0 < \infty$, given by

$$R_\gamma(r) = \cos(\gamma)R_{1(\infty)}(r) + \sin(\gamma)R_{2(\infty)}(r). \tag{13}$$

Define the Wronskian W for differentiable functions $f_1(x)$ and $f_2(x)$ as $W_x\left[f_1(x), f_2(x)\right] := f_1(x)\frac{df_2(x)}{dx} - \frac{df_1(x)}{dx}f_2(x)$. The solutions of Equation (13) satisfy the (generalized) Robin boundary conditions parametrized by γ:

$$\lim_{r\to\infty}\left(\cos(\gamma)W_r[R_\gamma(r), R_{1(\infty)}(r)] + \sin(\gamma)W_r[R_\gamma(r), R_{2(\infty)}(r)]\right) = 0. \tag{14}$$

When $\nu \notin (0,1)\cup(1,2)$, $r=\infty$ is also limit point and the only square-integrable solution to be considered is given by the principal solution, which coincides with (13) for $\gamma = 0$. This corresponds to imposing Dirichlet boundary condition and yields a unique dynamics. On this account, from now on let us focus on the more interesting scenario, where alternative boundary conditions give inequivalent dynamics, and consider only effective masses such that $\nu \in (0,1)\cup(1,2)$. In the latter case, given that $r=0$ is limit point and that $r=\infty$ is limit circle, from standard singular Sturm-Liouville theory[25] follows that A admits a one-parameter family of self-adjoint extensions in one-to-one correspondence with the (generalized) Robin boundary conditions.

2. Constructing the radial Green function

For R_0, R_γ and L_{ω^2} as obtained in the previous section, let $\mathcal{G}_\omega(r,r')$ be a Green function of the radial equation. It satisfies

$$((L_{\omega^2}-\omega^2)\otimes\mathbb{1})\mathcal{G}_\omega(r,r') = (\mathbb{1}\otimes(L_{\omega^2}-\omega^2))\mathcal{G}_\omega(r,r') = \frac{\delta(r-r')}{r^{-1}} \tag{15}$$

and it can be written as[26]

$$\mathcal{G}_\omega(r,r') = \frac{1}{\mathcal{N}_\omega}\left(\Theta(r'-r)R_0(r)R_\gamma(r') + \Theta(r-r')R_0(r')R_\gamma(r)\right). \tag{16}$$

For $\operatorname{Im}\omega < 0$, the normalization $\mathcal{N}_\omega$ reads:

$$\mathcal{N}_\omega = 4\nu\{B_0\cos(\gamma) - A_0\sin(\gamma)\}, \text{ with } \begin{cases} A_0 := \frac{\Gamma(1-b_0)}{\Gamma(a_0-b_0+1)}, \\ B_0 := \frac{\Gamma(b_0-1)}{\Gamma(a_0)}(iL^3\omega)^{-\nu}. \end{cases} \tag{17}$$

As shown in Ref. 1, the radial solutions defined above satisfy useful properties with respect to complex conjugation:

$$R_0(r,\overline{\omega}) = \overline{R_0(r,\omega)} \quad \text{and } R_\gamma(r,\overline{\omega}) = \overline{R_\gamma(r,\omega)}. \tag{18a}$$

As a consequence, the radial Green function and its normalization are such that

$$\mathcal{N}_{\overline{\omega}} = \overline{\mathcal{N}_\omega} \quad \text{and} \quad \mathcal{G}_{\overline{\omega}} = \overline{\mathcal{G}_\omega}. \tag{19}$$

In the next section, we study the poles of $\mathcal{G}_\omega(r,r')$ in the ω-complex plane. The properties above, in Equation (19), simplify our analysis since they imply that we can restrict our attention to either the upper or lower half of the ω-complex plane, and by conjugation, understand the behavior on the other half.

3. Rulying out the poles

The poles of the radial Green function are the frequencies ω for which

$$\mathcal{N}_\omega = 0. \tag{20}$$

As mentioned in the introduction, we refer to these frequencies as bound states; and in the scope of obtaining well-defined two-point functions, we exclude the boundary conditions for which they arise. Given the symmetry properties of the radial Green function with respect to complex conjugation, let us set $\mathrm{Im}(\omega) < 0$.

For Dirichlet and Neumann boundary conditions the analysis is straightforward. One just needs to take into account Equation (20) together with Equations (17) and the fact that the Euler Gamma function $\Gamma(z)$ satisfy $1/\Gamma(z) = 0 \iff z \in \mathbb{Z}_0^-$. We find that for Dirichlet boundary condition, $\gamma = 0$, the frequencies for which Equation (20) holds all lie in the upper half of the complex plane. That is, no bound states occur. For Neumann boundary condition, $\gamma = \pi/2$, bound states are also absent if $\nu \in (0, 1)$; however, if $\nu \in (1, 2)$, then for each λ_ℓ^m there is a unique pole with negative imaginary part at

$$\omega_p = -i\frac{\lambda_\ell^m}{2L(1-\nu)}. \tag{21}$$

For Robin boundary conditions, $\gamma \in (0, \pi/2) \cup (\pi/2, \pi)$, the analysis is less direct. First, we view Equation (20) as the following transcendental equation

$$\chi(\omega) := \tan(\gamma) - \Xi(\omega) = 0, \text{ with } \Xi(\omega) := \frac{B_0}{A_0}. \tag{22}$$

Then, by noting that Equation (22) admits a solution for all frequencies such that $\Xi(\omega) \in \mathbb{R}$, and by studying the function $\Xi(\omega)$, one can show that there is a one-to-one correspondence between the the boundary conditions for which bound states arise and the image $\mathbf{I} \subset \mathbb{R}$ of $\Xi(\omega)$ for $\omega = -i|\,\mathrm{Im}(\omega)|$. As a result, the physically-sensible boundary conditions are given by $\gamma \in \mathbb{R} \setminus \mathbf{I}$. Details of this analysis can be found in Ref. 1, but the distilled results are collected in the following table.

Table 2. Boundary conditions for which no bound states arise.

ν	mode-independent	mode-dependent
$(0, 1)$	$\gamma \in [0, \pi/2]$	$\gamma \in \left[0, \arctan\left(\left(\frac{-\lambda_\ell^m L^2}{4}\right)^{-\nu} \frac{\Gamma(\nu)}{\Gamma(-\nu)}\right)\right]$
$(1, 2)$	$\gamma = 0$	

4. Constructing the two-point functions

Analogously to the ansatz for the solutions to the Klein-Gordon equation, as in Equation (5), consider the following ansatz for the integral kernel of a positive

bisolution of the Klein-Gordon equation $G^+ \in \mathcal{D}'(\mathsf{Lif}^2)$, for $x = (t, r, \theta, \varphi) \in \mathsf{Lif}$,

$$G^+(x, x') = \lim_{\varepsilon \to 0^+} \int\limits_{\sigma(\triangle)} d\Sigma(\ell, m) \int_0^\infty d\omega e^{-i\omega(t-t'-i\varepsilon)} \widehat{G}_\omega(r, r') Y_\ell^m(\theta, \varphi) Y_\ell^m(\theta', \varphi'), \tag{23}$$

where $Y_\ell^m(\theta, \varphi)$ are the real-valued eigenfunctions of the Laplacian operator $\triangle$ with spectrum $\sigma(\triangle)$ and associated measure $d\Sigma(\ell, m)$, while $i\varepsilon$ implements the standard regularization.

To obtain an explicit expression for the radial component $\widehat{G}_\omega(r, r')$, in the following I employ the same steps as performed on other static spacetimes in Refs. 13–19. On one hand, imposing the canonical commutation relations on the antisymmetric part of Equation (23) entails that $\widehat{G}_\omega(r, r')$ is symmetric with respect to $r \leftrightarrow r'$ and gives the following identity

$$\int_{\mathbb{R}} \omega d\omega \widehat{G}_\omega(r, r') = \frac{\delta(r - r')}{r^{-1}}. \tag{24}$$

On the other hand, consider the radial Green function $\mathcal{G}_\omega(r, r')$ constructed in Section 2. Recalling that $\mathcal{G}_\omega(r, r')$ is not defined when $\mathrm{Im}(\omega) = 0$, let $\mathcal{C}^\infty$ be the infinite radius limit of a two semi-disks centered on the ω-complex plane and not touching the real axis, as illustrated in Ref. 1. For the boundary conditions for which the radial Green function has no poles, its spectral resolution[25,26] reads

$$\frac{1}{2\pi i} \oint_{\mathcal{C}^\infty} d(\omega^2) \mathcal{G}_\omega(r, r') = -\frac{\delta(r - r')}{r^{-1}}. \tag{25}$$

By performing the contour integral of Equation (25) and comparing it with Equation (24) we find the radial component of ansatz given in Equation (23):

$$\widehat{G}_\omega(r, r') = \frac{1}{4\pi\nu} \frac{\mathrm{Im}\left(B_0 \overline{A_0}\right)}{|B_0 \cos(\gamma) - A_0 \sin(\gamma)|^2} R_\gamma(r) R_\gamma(r'). \tag{26}$$

Alltogether, with respect to the Killing field ∂_t and for $\widehat{G}_\omega(r, r')$ given by Equation (26), a two-point function with integral kernel given by Equation (23) identifies the unique ground-state of a free, scalar, massive Klein-Gordon field on a Lifshitz spacetime. For a thermal-state at inverse-temperature β, its integral kernel reads instead

$$G_\beta^+(x, x') = \lim_{\varepsilon \to 0^+} \int\limits_{\sigma(\triangle)} d\Sigma(\ell, m) \int_0^\infty d\omega \, T_\omega(t, t') \widehat{G}_\omega(r, r') Y_\ell^m(\theta, \varphi) Y_\ell^m(\theta', \varphi'), \tag{27}$$

with

$$T_\omega(t, t') := \left[\frac{e^{-i\omega(t-t'-i\varepsilon)}}{1 - e^{-\beta\omega}} + \frac{e^{+i\omega(t-t'+i\varepsilon)}}{e^{\beta\omega} - 1} \right]. \tag{28}$$

Two-point functions with integral kernel given either by Equation (23) or Equation (27) identify Gaussian states of local Hadamard form[27] that satisfy the canonical commutation relations—they are physically-sensible two-point functions. Last, but most importantly, note that each boundary condition identifies a different two-point function.

5. Conclusions

The results presented here are part of the work in Ref. 1, where also generalizations to a hyperbolic and to a spherical Lifshitz topological black hole were considered. We demonstrated that on a Lifshitz spacetime, a Klein-Gordon field with effective mass $m^2 \in \left(-\frac{4}{L^2}, -\frac{3}{L^2}\right)$ admits a one-parameter family of Robin boundary conditions at $r \to \infty$ parametrized by $\gamma \in [0, \pi/2]$. In turn, each $\gamma \in [0, \pi/2]$ identifies a distinct two-point function. In addition, we found that for $m^2 \in \left(-\frac{4}{L^2}, -\frac{3}{L^2}\right) \cup \left(-\frac{3}{L^2}, -\frac{2}{L^2}\right)$, λ-mode dependent boundary conditions are also admissible. We call attention to the fact that, in the absence of a reason to justify selecting Dirichlet boundary condition over the others, all of these boundary conditions are of equal relevance—they all yield physically-sensible quantizations. Since these generalized quantizations arise naturally within the quantum field theoretical framework, we hope our results stimulates research in the direction of understanding their physical significance.

The analysis outlined here applies straightforwardly to higher dimensional Lifshitz spacetimes of critical exponent $z = 2$. That is not the case for $z > 2$, however, given that the choice of critical value affects drastically the classification, and hence the solutions, of the radial part of the Klein-Gordon equation. In Ref. 1, we constructed two-point functions not only on this spacetime, but also on Lifshitz topological black holes.[28] With this in mind, an interesting follow up work would be to perform a numerical analysis to understand how the topology and the choice of a boundary condition affects vacuum fluctuations, as performed for topological black holes of Einstein gravity in Ref. 29. Moreover, in this framework one can further directly obtain the transition rate of a static Unruh-DeWitt detector in the infinite interaction time limit, following the same prescription of Ref. 14, and study the effects of different boundary conditions and topologies in the perspective of a particle detector.

Acknowledgments

This work was part of a collaboration[1] with C.Dappiaggi and D.Sina, whom I thank and gratefully acknowledge; and it was partially based on the MSc thesis of D.S. submitted to the University of Pavia on March 2021. My work was supported by a PhD scholarship of the University of Pavia.

References

1. L. De Souza Campos, C. Dappiaggi and D. Sina, [arXiv:2103.15391 [hep-th]].
2. L. De Souza Campos, Talk: Boundary conditions for the Klein-Gordon field on Lifshitz Spacetimes, Sixteenth Marcel Grossmann Meeting, `https://indico.icranet.org/event/1/`, available on YouTube `https://youtu.be/5z7zhpiDOpw?list=PLr5RLbSWSonsaOnZukBDs0qsNIWM8AvRF&t=4007`.
3. S. A. Hartnoll, A. Lucas and S. Sachdev, [arXiv:1612.07324] (Pg.27).
4. P. Hořava, Class. Quant. Grav. **28**, 114012 (2011), [arXiv:1101.1081].
5. P. Hořava, Phys. Rev. D **79**, 084008 (2009), [arXiv:0901.3775].
6. A. Wang, Int. J. Mod. Phys. D **26**, no. 7, 1730014 (2017), [arXiv:1701.06087].
7. S. A. Hartnoll, Class. Quant. Grav. **26**, 224002 (2009), [arXiv:0903.3246].
8. P. Hořava and C. M. Melby-Thompson, Gen. Rel. Grav. **43**, 1391-1400 (2011) [arXiv:0909.3841].
9. S. F. Ross, Class. Quant. Grav. **28**, 215019 (2011) [arXiv:1107.4451].
10. S. Kachru, X. Liu and M. Mulligan, Phys. Rev. D **78**, 106005 (2008), [arXiv:0808.1725].
11. A. Ishibashi and R. M. Wald, Class. Quant. Grav. **20**, 3815-3826 (2003), [arXiv:gr-qc/0305012].
12. A. Ishibashi and R. M. Wald, Class. Quant. Grav. **21** (2004), 2981-3014, [arXiv:0402184].
13. C. Dappiaggi and H. Ferreira, Phys. Rev. D **94** (2016) no. 12, 125016, [arXiv:1610.01049].
14. L. Campos and C. Dappiaggi, Phys. Rev. D **103**, no. 2, 025021 (2021), [arXiv:2011.03812].
15. C. Dappiaggi, H. R. C. Ferreira and B. A. Juárez-Aubry, Phys. Rev. D **97** (2018) no. 8, 085022, [arXiv:1802.00283].
16. C. Dappiaggi and H. Ferreira, Rev. Math. Phys. **30** (2017) no. 02, 1850004, [arXiv:1701.07215].
17. C. Dappiaggi, H. Ferreira and A. Marta, Phys. Rev. D **98** (2018) no. 2, 025005, [arXiv:1805.03135].
18. F. Bussola, C. Dappiaggi, H. R. C. Ferreira and I. Khavkine, Phys. Rev. D **96** (2017) no. 10, 105016, [arXiv:1708.00271].
19. F. Bussola and C. Dappiaggi, Class. Quant. Grav. **36** (2019) no. 1, 015020, [arXiv:1806.00427].
20. A. Giacomini, G. Giribet, M. Leston, J. Oliva and S. Ray, Phys. Rev. D **85**, 124001 (2012), [arXiv:1203.0582].
21. G. M. Quinta, A. Flachi and J. P. S. Lemos, Phys. Rev. D **93**, no. 12, 124073 (2016) [arXiv:1604.00495].
22. C. Keeler, JHEP **01**, 067 (2014), [arXiv:1212.1728].
23. T. Andrade and S. F. Ross, Class. Quant. Grav. **30**, 065009 (2013), [arXiv:1212.2572].
24. NIST Digital Library of Mathematical Functions, Ch.13, `https://dlmf.nist.gov/13`, Accessed: 2021-03-28.
25. A. Zettl, *Sturm-Liouville Theory*, American Mathematical Society, (2005), 328p.
26. I. Stakgold and M.J. Holst, *Green's functions and boundary value problems*, John Wiley & Sons, 99 (2011).
27. H. Sahlmann and R. Verch, Commun. Math. Phys. **214** (2000), 705-731 [arXiv:0002021].
28. R. B. Mann, JHEP **06**, 075 (2009), [arXiv:0905.1136].
29. T. Morley, P. Taylor and E. Winstanley, [arXiv:2010.01562].

Dynamical system analysis of Bianchi-I spacetimes in $f(R)$ gravity

Saikat Chakraborty

Center for Gravitation and Cosmology, College of Physical Science and Technology, Yangzhou University, Yangzhou 225009, China,
and
International Center for Cosmology, Charusat University, Anand 388421, Gujrat, India
E-mail: saikatnilch@gmail.com; snilch@yzu.edu.cn

Kazuharu Bamba

Division of Human Support System, Faculty of Symbiotic Systems Science, Fukushima University, Fukushima 960-1296, Japan
E-mail: bamba@sss.fukushima-u.ac.jp

Alberto Saa

Department of Applied Mathematics, University of Campinas, 13083-859 Campinas, SP, Brazil
E-mail: asaa@ime.unicamp.br

Based on our original work published in Ref.[1], we investigate an autonomous system analysis in terms of new expansion-normalized variables for homogeneous and anisotropic Bianchi-I spacetimes in $f(R)$ gravity in the presence of anisotropic matter. It is demonstrated that with a suitable choice of the evolution parameter, the Einstein's equations are reduced to an autonomous 5-dimensional system of ordinary differential equations for the new variables. Furthermore, for a large class of functions $f(R)$, which includes several cases commonly considered in the literature, all the fixed points are polynomial roots, and thus they can be determined with good accuracy and classified for stability. In addition, typically for these cases, any fixed point corresponding to isotropic solutions in the presence of anisotropic matter will be unstable. The assumption of a perfect fluid as source and or the vacuum cases imply some dimensional reductions and even more simplifications. In particular, it is found that the vacuum solutions of $f(R) = R^{\delta+1}$ with δ a constant are governed by an effective bi-dimensional phase space which can be constructed analytically, leading to an exactly soluble dynamics. It is also shown that several results already reported in the literature can be re-obtained in a more direct and easy way by exploring our dynamical formulation.

Keywords: Modified gravity theories; autonomous system analysis

1. Introduction

According to recent various cosmological observations including Type Ia Supernovae, cosmic microwave background (CMB) radiation and large scale structure, in addition to inflation[2–4] in the early universe, the current expansion of the universe is also accelerating. This is the so-called dark energy problem. There are two representative approaches to study the issue of dark energy. One is to introduce some unknown matter called dark energy with the negative pressure in general relativity.

The other is to modify gravity at large scales (for reviews of dark energy and modified gravity theories, see, e.g., Refs.[5–17]). As one of the popular modified gravity theories, $f(R)$ gravity has been proposed.[18–20]

In this article, based on our original reference[1], we extend the so-called expansion-normalized variables[21] to write down the dynamical equations of $f(R)$ gravity for a homogeneous and anisotropic Bianchi-I metric in the presence of an anisotropic fluid, as a 5-dimensional system of ordinary differential equations. We show that some further assumptions may lead to considerable simplifications in the equations, and for several examples we end up with analytically soluble systems. For the sake of illustration, we consider explicitly the case of $f(R) = R^{1+\delta}$. We demonstrate that the formulation of[22,23] is recovered in the isotropic matter limit. Moreover, in a simpler and more direct way, we re-derive some uniqueness and stability properties of the Starobinsky's isotropic inflationary scenario in R^2 gravity,[24–26] which is consistent with the Planck 2018 results.[27,28]

The article is organized as follows. In section 2, we describe the dynamical equations for a Bianchi-I cosmology in $f(R)$ gravity in the presence of an anisotropic fluid. We discuss the isotropic fluid limit and introduce the new expansion-normalized variables for the system. In Section 3, we present cosmological applications. Finally, we summarize our results in Section 4.

2. Bianchi-I cosmology in $f(R)$ gravity with anisotropic fluid

The action describing $f(R)$ gravity is expressed as

$$S_{=}\frac{1}{2\kappa}\int d^4x\sqrt{-g}f(R) + S_M, \tag{1}$$

where $\kappa = 8\pi G$, $c = \hbar = 1$, and S_M stands for the usual matter contributions to the totalaction.

We consider the homogeneous and anisotropic Bianchi-I metric, which can be conveniently cast for our purposes in the following form[29–31]

$$ds^2 = -dt^2 + a^2(t)\sum_{i=1}^{3} e^{2\beta_i(t)}(dx^i)^2, \tag{2}$$

where $a(t)$ is the average scale factor and the three functions β_i, which characterize the anisotropies, are such that $\beta_1 + \beta_2 + \beta_3 = 0$. It is more convenient to employ the variables

$$\beta_{\pm} = \beta_1 \pm \beta_2. \tag{3}$$

The total amount of anisotropy in the metric (2) is given by the quantity

$$\sigma^2 = \dot{\beta}_1^2 + \dot{\beta}_2^2 + \dot{\beta}_3^2 = \frac{3}{2}\dot{\beta}_+^2 + \frac{1}{2}\dot{\beta}_-^2. \tag{4}$$

For $\sigma = 0$, one can show that the spatial coordinates x^i can be suitably rescaled to recast the Bianchi-I metric in the standard Friedmann-Lemaitre-Robertson-Walker

(FLRW) form. The Ricci scalar for the metric (2) reads

$$R = 6\dot{H} + 12H^2 + \sigma^2, \tag{5}$$

where the average Hubble parameter H is given by the standard expression

$$H = \frac{\dot{a}}{a}. \tag{6}$$

We assume the presence of an anisotropic barotropic fluid with energy momentum tensor parametrized as[32]

$$T^{\nu}_{\mu} = \mathrm{diag}\left(-\rho, p_1, p_2, p_3\right) = \mathrm{diag}\left(-\rho, \omega_1\rho, \omega_2\rho, \omega_3\rho\right), \tag{7}$$

and we define the anisotropic equation of state as

$$p_i = (\omega + \mu_i)\rho, \tag{8}$$

with $i = 1, 2, 3$, where ω is the average barotropic parameter and $\omega_i = \omega + \mu_i$, with $\mu_1 + \mu_2 + \mu_3 = 0$ by construction. We parameterize our fluid by the constants ω and $\mu_{\pm} = \mu_1 \pm \mu_2$.[31]

The dynamics of the Bianchi-I metric (2) under $f(R)$ gravity action (1), in the presence of and anisotropic barotropic fluid with energy-momentum tensor (7), can be described by the following set of equations,[31]

$$3H^2 = \frac{\kappa}{f'}\left(\rho + \frac{Rf' - f}{2\kappa} - \frac{3Hf''\dot{R}}{\kappa}\right) + \frac{\sigma^2}{2} \quad , \tag{9}$$

$$2\dot{H} + 3H^2 = -\frac{\kappa}{f'}\left(\omega\rho + \frac{\dot{R}^2 f''' + \left(2H\dot{R} + \ddot{R}\right) f''}{\kappa} - \frac{Rf' - f}{2\kappa}\right) - \frac{\sigma^2}{2}, \tag{10}$$

$$\ddot{\beta}_{\pm} + \left(3H + \frac{\dot{R}f''}{f'}\right)\dot{\beta}_{\pm} = \frac{\kappa\rho}{F}\mu_{\pm}, \tag{11}$$

$$\dot{\rho} + \left(3H\left(1 + \omega\right) + \boldsymbol{\delta}\cdot\dot{\boldsymbol{\beta}}\right)\rho = 0, \tag{12}$$

where $i = 1, 2, 3$, and

$$\boldsymbol{\delta}\cdot\dot{\boldsymbol{\beta}} = \mu_1\dot{\beta}_1 + \mu_2\dot{\beta}_2 + \mu_3\dot{\beta}_3 = \frac{3}{2}\mu_+\dot{\beta}_+ + \frac{1}{2}\mu_-\dot{\beta}_-. \tag{13}$$

Notice that in the presence of a perfect fluid, we have $\mu_+ = \mu_- = 0$ and the two equations (11) for β_+ and β_- can be substituted with

$$\dot{\sigma} + \left(3H + \frac{\dot{R}f''}{f'}\right)\sigma = 0. \tag{14}$$

In this case, there is no anisotropy in the matter sector and the single variable σ is sufficient to describe the total amount of metric anisotropy in the system.

In general, we have four functions of time $H(t)$, $\rho(t)$, $\beta_{\pm}(t)$ governing the dynamics. The existence of the constraint equation (9) implies that only three of them are indeed independent. Without loss of generality, we can choose them to be, for instance, $H(t)$ and $\beta_{\pm}(t)$. Given some specific form of the function $f(R)$, they can be determined by solving equations (10) and (11). The fluid energy density $\rho(t)$ can then be found using the energy constraint (9).

The traditional expansion-normalized variables were initially introduced for a better dynamical analysis of the standard FLRW model, see, e.g., Ref.[21]. Here, we expand the variables already introduced in[22,23] to include the case of the anisotropic barotropic fluid (7). In this regard, let us introduce the monotonically increasing variable

$$N = \epsilon \ln a, \tag{15}$$

known as the logarithmic time, where ϵ is defined to be $+1$ for expanding universe and -1 for a contracting one. Without loss of generality, we choose the scale factor at $t = 0$ to be $a_0 = 1$. Therefore, as time progresses in the forward (positive) direction, the logarithmic time N becomes positive and goes towards $+\infty$ in case of both expanding and contracting universes. One can notice that

$$\dot{N} = \epsilon H, \tag{16}$$

so that $\dot{N}$ is effectively always positive, justifying the use of N as the dimensionless evolution variable for both expanding and contracting universes. On the other hand, around a bounce or a turnaround point, this argument is not valid though and the expanding and contracting branches must be considered separately.

The expansion-normalized dynamical variables suitable for the equations (9) - (12) are the following dimensionless combinations

$$\begin{gathered} u_1 = \frac{\dot{R} f''}{f' H}, \; u_2 = \frac{R}{6H^2}, \; u_3 = \frac{f}{6 f' H^2}, \\ u_4^+ = \frac{\dot{\beta_+}^2}{4H^2}, \; u_4^- = \frac{\dot{\beta_-}^2}{12H^2}, \; u_5 = \frac{\kappa \rho}{3 f' H^2}. \end{gathered} \tag{17}$$

in terms of which the energy constraint (9) reads simply

$$g = 1 + u_1 - u_2 + u_3 - u_4^+ - u_4^- - u_5 = 0, \tag{18}$$

from where we have that one of the expansion-normalized variables can always be eliminated. Unless otherwise stated, we always choose the matter content variable u_5 to be expressed in terms of the others dynamical variables. The variable

$$u_4 = u_4^+ + u_4^- = \frac{\sigma^2}{6H^2} \tag{19}$$

is also relevant for our purposes. It is important to stress that the variables u_4^+ and u_4^- are both non-negative by construction. Now, let us introduce the quantity

$$\gamma(R) = \frac{f'}{R f''}, \tag{20}$$

which contains the information about the form of $f(R)$. Knowing the form of $f(R)$, γ can be determined in terms of the dynamical variables u_2, u_3 by inverting the relation

$$\frac{u_2}{u_3} = \frac{Rf'}{f}. \tag{21}$$

We return to the question of the invertibility of (21) in the last section. The 5-dimensional system of autonomous first order differential equations fully equivalent to (10) - (12) is given by

$$\begin{aligned}
\epsilon\frac{du_1}{dN} &= 1 + u_2 - 3u_3 - u_4 - 3\omega u_5 \\
&\quad -u_1\left(u_1 + u_2 - u_4\right), && (22)\\
\epsilon\frac{du_2}{dN} &= u_1u_2\gamma\left(\frac{u_2}{u_3}\right) - 2u_2\left(u_2 - u_4 - 2\right), && (23)\\
\epsilon\frac{du_3}{dN} &= u_1u_2\gamma\left(\frac{u_2}{u_3}\right) - u_3\left(u_1 + 2u_2 - 2u_4 - 4\right), && (24)\\
\epsilon\frac{du_4^+}{dN} &= -2u_4^+\left(1 + u_1 + u_2 - u_4\right) + 3\mu_+\sqrt{u_4^+}u_5, && (25)\\
\epsilon\frac{du_4^-}{dN} &= -2u_4^-\left(1 + u_1 + u_2 - u_4\right) + \mu_-\sqrt{3u_4^-}u_5, && (26)\\
\epsilon\frac{du_5}{dN} &= -u_5\Big(3\omega - 1 + u_1 + 2u_2 - 2u_4 && (27)\\
&\quad +3\mu_+\sqrt{u_4^+} + \mu_-\sqrt{3u_4^-}\Big).
\end{aligned}$$

Notice that differentiating (18) with respect to N and using the equations (22)-(27), we have

$$\epsilon\frac{dg}{dN} = -(u_1 + 2u_2 - 2u_4^+ - 2u_4^- - 1)g, \tag{28}$$

showing that the constraint $g = 0$ is indeed conserved along the solutions of our equations and the system (22) - (27) is effectively 5-dimensional.

The case of $f(R) = R^{1+\delta}$, with $\delta \neq 0$, is particularly important in our next examples. For this choice of $f(R)$, one has simply

$$\gamma = \delta^{-1}, \tag{29}$$

and the equations (23) and (24) can be considerably simplified. In this case, the right-handed side of the equations (22) - (27) involves only second degree polynomials in u_1, u_2, and u_3, and forth degree in $\sqrt{u_4^-}$ and $\sqrt{u_4^+}$. Hence, the task of finding the fixed points of our system reduce to finding polynomial roots, which may be performed in general with good accuracy. Notice that there are other relevant choices for $f(R)$ leading to polynomial fixed points. Besides of the trivial

extension $f(R) = \alpha R^{1+\delta} + \Lambda$, with α and Λ constants, for which (29) also holds. For $f(R) = R^a + \alpha R^b$, with $a \neq b$ constants, we have

$$\gamma = \frac{u_2}{(b+a-1)u_2 - abu_3}. \tag{30}$$

Notice that, as in the exponential case, the function γ does not depend on the parameter α. This, of course, does not mean that the dynamics in insensitive to the value of α, since the expansion-normalized variables (17) depend explicitly on α. The case $a = 1$ and $b = 2$ is the original Starobinsky inflationary scenario,[4] and for the vacuum case our approach reduces to that one considered recently in.[33]

3. Applications to cosmology

In the following, we consider the expanding universes ($\epsilon = 1$). We investigate the case $f(R) = R^{1+\delta}$, whose main motivations from a cosmological perspective can be found in,[22,23,34,35] for instance. The case with $\delta = 0$ is obviously pure GR, for which the corresponding system is lower-dimensional, and our approach simply does not apply. The case logarithmic case $f(R) = \ln R$ must be treated separately. Hence, we start considering $\delta \neq 0$ and $\delta \neq -1$. Since we will deal with vacuum solutions, we set $u_5 = 0$ in the equations (18) and (22) - (27). In this case, notice that (25) and (26) can be combined in only one equation for u_4. We can use (18) to write u_3 as

$$u_3 = u_2 - u_1 + u_4 - 1, \tag{31}$$

and we are left with only three dynamical variables u_1, u_2, and u_4. Now, there is an interesting point to notice[36] about the specific choice $f(R) = R^{1+\delta}$, with $\delta \neq -1$, namely that

$$\frac{u_2}{u_3} = \frac{Rf'}{f} = 1 + \delta, \tag{32}$$

which combined with the constraint (31) implies

$$\delta u_2 = (1+\delta)(u_1 - u_4 + 1), \tag{33}$$

and we are left in fact with a two-dimensional phase space spanned by the variables u_1 and u_4. The corresponding dynamical equations in this case are

$$\begin{aligned} \frac{du_1}{dN} &= \phi_1(u_1, u_4) \qquad (34)\\ &= -\delta^{-1}(1+2\delta)(u_1 - u_1^*)(u_1 - u_4 + 1), \\ \frac{du_4}{dN} &= \phi_4(u_1, u_4) \qquad (35)\\ &= -2\delta^{-1}(1+2\delta)u_4\,(u_1 - u_4 + 1), \end{aligned}$$

where

$$u_1^* = \frac{2(\delta - 1)}{1 + 2\delta}. \tag{36}$$

The phase space (u_1, u_4) associated with the system (34) - (35) has some interesting features. For instance, it has an one-dimensional invariant subspace (a continuous line of fixed points) corresponding to the straight line $u_1 - u_4 = -1$. However, from (31) we have that $u_3 = u_2$ on this line, which implies from (17) and (32) that $R = 0$ on $u_1 - u_4 = -1$. Besides of this invariant straight line, we find the isolated fixed $(u_1^*, 0)$, for $\delta \neq -\frac{1}{2}$. The case $\delta = -\frac{1}{2}$ is also discussed separately.

The stability of the isolated fixed point can be inferred from the linearization of (34) - (35). The Jacobian matrix of (34) - (35) at the point $(u_1^*, 0)$ reads

$$\left(\frac{\partial(\phi_1, \phi_4)}{\partial(u_1, u_4)}\right) = -\delta^{-1}(4\delta - 1)\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \tag{37}$$

from where we see that such fixed point is stable for $\delta > \frac{1}{4}$ or for $\delta < 0$. For the stability of the invariant straight line, we can consider the divergence of the vector field (ϕ_1, ϕ_2). One has

$$\nabla \cdot \boldsymbol{\phi} = \frac{\partial \phi_1}{\partial u_1} + \frac{\partial \phi_4}{\partial u_4} = \delta^{-1}((1 + 2\delta)u_4 + 4\delta - 1) \tag{38}$$

on the invariant line. Recalling that $u_4 \geq 0$, we find that the invariant line is entirely repulsive (positive divergence) for $\delta > \frac{1}{4}$ or for $\delta \leq -\frac{1}{2}$. For $-\frac{1}{2} < \delta \leq \frac{1}{4}$, we can obtain some attractive segments, depending on the value of u_4. We return to the physical interpretation of this $R = 0$ invariant line in a following sub-section. The case $\delta = -\frac{1}{2}$ is particularly curious, since the isolated fixed point is absent and we get a second one-dimensional invariant line, namely $u_1 = 0$, which is also entirely repulsive. On the other hand, the case $f(R) = \ln R$ cannot be incorporated in the present analysis since (32) is not valid for $\delta \to -1$, and in fact we have a three-dimensional phase space for such case.

The solutions of (34) and (35) are curves on the plane (u_1, u_4), and it turns out that such curves can be determined analytically. Notice that the solutions are such that

$$\frac{u_4'}{u_1'} = \frac{2u_4}{u_1 - u_1^*}, \tag{39}$$

which can be integrated as

$$u_4 = c\,(u_1 - u_1^*)^2\,, \tag{40}$$

with arbitrary c. Thus, the phase space trajectories of all solutions of (34) and (35) are simply parabolas centered in the isolated fixed point, irrespective of the value of δ, provided the fixed point exists. Since we know the trajectories graphs, one can infer the dynamics direction and, consequently, the dynamical properties of the fixed point and the invariant line, directly form the equations (34) and (35) as follows. Consider the phase space function $L = u_1 - u_4 + 1$. It is clear that $L = 0$ is the invariant line. On the other hand, $L = c$ constant is a parallel line located below the invariant line if $c > 0$, or above if $c < 0$. The invariant line is the boundary between two semiplanes with reverse dynamics direction, and the dynamical properties of the

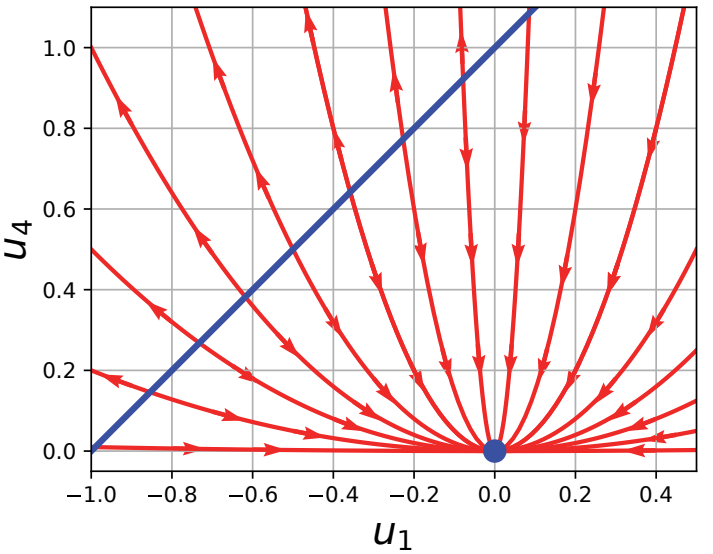


Fig. 1. Phase space for the system (34) - (35), for $\delta = 1$. The fixed point $(0, 0)$ is located in the semiplane below the critical line. The solutions are restricted to parabolas centered in the attractive fixed point. The region below the invariant line corresponds to the attraction basin of the fixed point. Any solution starting there tends to the fixed point asymptotically. All solutions starting in the region above the critical line will diverge to infinity. Notice that the critical line is entirely repulsive. Such phase space is rather generic, it is essentially the same for all theories of the type $f(R) = R^{1+\delta}$ such that the fixed point is attractive and is located below the invariant line.

fixed point and of the invariant line depend on the relative position between then, see Figs 1 and 2, which correspond, respectively, to the cases $\delta = 1$ and $\delta = \frac{1}{10}$. The former is the important case of the Starobinsky's inflationary scenario with $f(R) = R^2$.

We note that the solutions are constrained to the parabolas (40), the exact solutions of (34) and (35) boils down to a simple quadrature of a rational function

$$\frac{d\bar{u}_1}{c\bar{u}_1^3 - \bar{u}_1^2 - (u_1^* + 1)\bar{u}_1} = -\delta^{-1}(1 + 2\delta)dN, \tag{41}$$

with $u_1 = \bar{u}_1 + u_1^*$. For the case $\delta = -\frac{1}{2}$, u_4 is a constant and (34) also reduces to a simple rational quadrature. It has been shown that the vacuum solutions for the $f(R) = R^{1+\delta}$ case, for $\delta \neq -1$, are exactly soluble.

Since the stable fixed points of a cosmological model correspond to the cosmological histories which dominate the asymptotic evolution of the system, it worth to look more closely on them. By using (31) and (32), we see that the isolated fixed points are given by

$$u_3 = \frac{4\delta - 1}{\delta(1 + 2\delta)}, \tag{42}$$

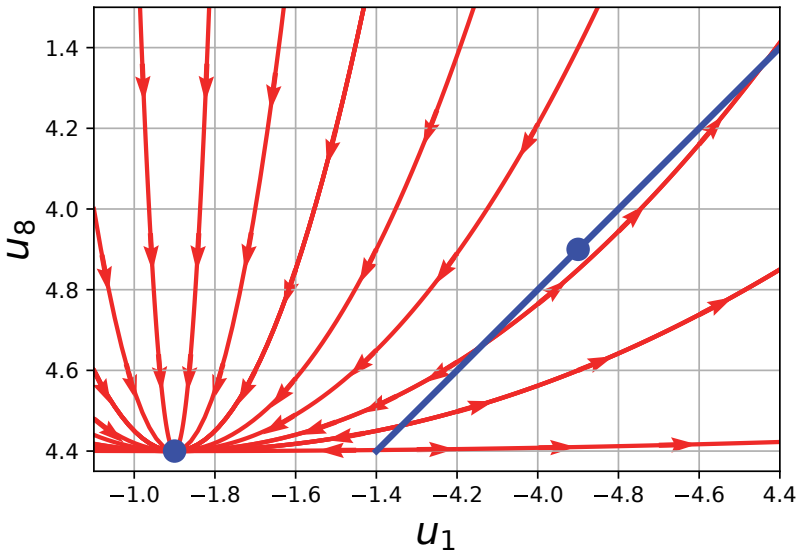


Fig. 2. Phase space for the system (34) - (35), for $\delta = \frac{1}{10}$. The fixed point $\left(-\frac{3}{2}, 0\right)$ is now located in the semiplane above the critical line. The solutions are also restricted to parabolas centered in the attractive fixed point. However, the attraction basin of the fixed point is now in the region above the critical line. Notice that the invariant line in this case has an attractive and a repulsive segment located, respectively, above and below the depicted point $\left(-\frac{1}{2}, \frac{1}{2}\right)$. The divergence (38) always vanishes in limit points between attractive and repulsive segments like this one.

with $\delta > \frac{1}{4}$ or $\delta < 0$. From the definition of u_3 and (5), we have that (42) implies that

$$\dot{H} = \Delta H^2, \tag{43}$$

where

$$\Delta = \frac{\delta - 1}{\delta(1 + 2\delta)}. \tag{44}$$

It is clear that for $\delta = 1$, the stable fixed point corresponds to de Sitter solution with $a(t) = e^{Ht}$, with constant H. (The case $H = 0$ corresponds to the flat Minkowski spacetime). This is namely the well known Starobinsky's inflationary solution. For $\delta \neq 1$, the solutions are

$$H(t) = \frac{H_0}{1 - \Delta H_0 (t - t_0)}, \tag{45}$$

where $H(t_0) = H_0$, which interpretation is straightforward. For $\Delta > 0$, which corresponds to $-\frac{1}{2} < \delta < 0$ or $\delta > 1$, we have a future finite time big rip singularity, while for $\Delta < 0$ ($\delta < -\frac{1}{2}$ or $0 < \delta < 1$), the Hubble parameter H decreases as t^{-1} for large t, *i.e.*, the solution asymptotically tends to a power law expansion.

The Starobinsky's R^2 inflationary scenario is unique among the $F(R) = R^{1+\delta}$ theories of gravity, since only for $\delta = 1$ the stable de Sitter fixed point $(0, 0)$ is available, a result indeed known for a long time, see,[24–26] for example. We can, however, easily prove a stronger result for generic $f(R)$ theories. The de Sitter solution $a(t) = e^{Ht}$, with constant and arbitrary H, implies $u_1 = u_4 = 0$, and also

$$R = 12H^2, \tag{46}$$

which, on the other hand, determine that $u_2 = 2$ and $u_3 = \frac{2f}{Rf'}$ and, hence, the constraint (18) reads

$$Rf'(R) = 2f(R). \tag{47}$$

Since we assume that de Sitter solution exists for arbitrary H, we have from (46) that it should exist for any $R > 0$, and hence equation (47) can be seen as a ordinary differential equation for $f(R)$, which unique solution is $f(R) = \alpha R^2$, establishing in this way a stronger result: the case R^2 is unique among all vacuum $f(R)$ theories with respect to the existence of a de Sitter solution with arbitrary H. The condition (47) was first obtained by Barrow and Ottewill in[24] by using a more intricate approach, but here we see that it appears from a very simple analysis of fixed points.

It is noted that in the case with anisotropic fluids, all isotropic fixed points are unstable (there is no asymptotically stable isotropic solutions in the presence of anisotropic matter). It is also mentioned that several known results as the existence of vacuum Kasner-like solutions for $-\frac{1}{2} \leq \delta \leq \frac{1}{4}$[34,35] are found. Moreover, for the case of exponential gravity,[37–40] $f(R) = e^{\alpha R}$, all isotropic fixed points in the presence of an anisotropic barotropic fluid are unstable.

4. Summary

In this article, we have introduced a new set of expansion-normalized variables for homogeneous and anisotropic Bianchi-I spacetimes in $f(R)$ gravity in the presence of anisotropic matter. In terms of these new dynamical variables, the full set of Einstein's equations boils down to a 5-dimensional phase space. As applications of the proposed dynamical approach, we have explicitly explored the $f(R) = R^{1+\delta}$ modified theory of gravity, and shown that its vacuum dynamics is exactly solvable. Furthermore, in a easier and more direct way, we have re-obtained several well known results for this particular choice of $f(R)$ such as Bleyer and Schmidt isotropic solutions.[41–43] We have also extended a uniqueness result for Starobisnki inflationary scenario, namely that the case R^2 is unique among all vacuum $f(R)$ theories with respect to the existence of a de Sitter solution with arbitrary H, a result obtained previously by Barrow and Ottewill by using a more intricate approach.[24]

Acknowledgments

The work of KB was supported in part by the JSPS KAKENHI Grant Number JP21K03547.

References

1. S. Chakraborty, K. Bamba and A. Saa, Phys. Rev. D **99**, 064048 (2019) [arXiv:1805.03237 [gr-qc]].
2. A. H. Guth, Phys. Rev. D **23**, 347 (1981).
3. A. D. Linde, Phys. Lett. **108B**, 389 (1982).
4. A. A. Starobinsky, Phys. Lett. **91B**, 99 (1980).
5. E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15**, (2006) 1753-1936. doi:10.1142/S021827180600942X [arXiv:hep-th/0603057 [hep-th]].
6. T. Padmanabhan, Gen. Rel. Grav. **40**, (2008) 529-564. doi:10.1007/s10714-007-0555-7 [arXiv:0705.2533 [gr-qc]].
7. R. Durrer and R. Maartens, Gen. Rel. Grav. **40**, (2008) 301-328. doi:10.1007/s10714-007-0549-5 [arXiv:0711.0077 [astro-ph]].
8. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, Phys. Rept. **513**, (2012) 1-189. doi:10.1016/j.physrep.2012.01.001 [arXiv:1106.2476 [astro-ph.CO]].
9. K. Bamba, S. Capozziello, S. Nojiri and S.D. Odintsov, Astrophys. Space Sci. **342**, 155 (2012) [arXiv:1205.3421].
10. A. de Felice and S. Tsujikawa, Living Rev. Rel. **13**, 3 (2010) [arXiv:1002.4928].
11. T.P. Sotiriou and V. Faraoni, Rev. Mod. Phys. **82**, 451 (2010) [arXiv:0805.1726].
12. V. Faraoni and S. Capozziello, Fundam. Theor. Phys. **170** (2010).
13. S. Nojiri and S. Odintsov, Phys. Rep. **505**, 59 (2011) [arXiv:1011.0544].
14. S. Capozziello and M. De Laurentis, Phys. Rept. **509**, 167 (2011) [arXiv:1108.6266].
15. Y. F. Cai, S. Capozziello, M. De Laurentis and E. N. Saridakis, Rept. Prog. Phys. **79**, 106901 (2016) [arXiv:1511.07586].
16. S. Nojiri, S. Odintsov, and V.K. Oikonomou, Phys. Rep. **692**, 1 (2017) [arXiv:1705.11098].
17. K. Bamba and S. D. Odintsov, Symmetry **7**, 220 (2015) [arXiv:1503.00442].
18. S. Capozziello, S. Carloni and A. Troisi, Recent Res. Dev. Astron. Astrophys. **1**, 625 (2003) [astro-ph/0303041].
19. S. Nojiri and S. D. Odintsov, Phys. Rev. D **68**, 123512 (2003) [hep-th/0307288].
20. S. M. Carroll, V. Duvvuri, M. Trodden and M. S. Turner, Phys. Rev. D **70**, 043528 (2004) [astro-ph/0306438].
21. J. Wainwright and G.F.R. Ellis, *Dynamical systems in cosmology*, Cambridge University Press (1997).
22. J. A. Leach, S. Carloni and P. K. S. Dunsby, Class. Quant. Grav. **23**, 4915 (2006) [gr-qc/0603012].
23. N. Goheer, J. A. Leach and P. K. S. Dunsby, Class. Quant. Grav. **24**, 5689 (2007) [arXiv:0710.0814].
24. J.D. Barrow. and A.C. Ottewill, J. Phys. A**16**, 2757 (1983).
25. K. Maeda, Phys. Rev. D **37**, 858 (1988).
26. J. D. Barrow and S. Hervik, Phys. Rev. D **74**, 124017 (2006) [gr-qc/0610013].
27. N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020) [erratum: Astron. Astrophys. **652**, C4 (2021)] doi:10.1051/0004-6361/201833910 [arXiv:1807.06209 [astro-ph.CO]].
28. Y. Akrami *et al.* [Planck], Astron. Astrophys. **641**, A10 (2020) doi:10.1051/0004-6361/201833887 [arXiv:1807.06211 [astro-ph.CO]].
29. T. S. Pereira, C. Pitrou and J. P. Uzan, JCAP **0709**, 006 (2007) [arXiv:0707.0736].
30. S. Chakraborty, Phys. Rev. D **98**, 024009 (2018) [arXiv:1803.01594].
31. S. Chakraborty, S. Pal, and A. Saa, Phys. Rev. D **99**, 024020 (2019) [arXiv:1812.01694].

32. O. Akarsu and C.B. Kilinc, Astr. Sp. Science, **326**, 315 (2010) [arXiv:1001.0550].
33. D. Muller, A. Ricciardone, A.A. Starobinsky, A. Toporensky, Eur. Phys. J. C **78**, 311 (2018) [arXiv:1710.08753].
34. J. D. Barrow and T. Clifton, Class. Quant. Grav. **23**, L1 (2006) [gr-qc/0509085].
35. T. Clifton and J. D. Barrow, Class. Quant. Grav. **23**, 2951 (2006) [gr-qc/0601118].
36. S. Carloni, P. K. S. Dunsby, S. Capozziello and A. Troisi, Class. Quant. Grav. **22**, 4839 (2005) [gr-qc/0410046].
37. M. Abdelwahab, S. Carloni, and P.K.S. Dunsby, Class. Quantum Grav. **25**, 135002 (2008) [arXiv:0706.1375].
38. E.V. Linder, Phys. Rev. D **80**, 123528 (2009) [arXiv:0905.2962].
39. S. I. Kruglov, Int. J. Mod. Phys. A **28**, 13501194 (2013) [arXiv:1204.6709].
40. S. D. Odintsov, D. Saez-Chillon Gomez, G.S. Sharov, Eur. Phys. J. C **77**, 862 (2017) [arXiv:1709.06800].
41. U. Bleyer and H.J. Schmidt, Int. J. Mod. Phys. A**5**, 4671 (1990).
42. H. J. Schmidt, Astron. Nachr. **311**, 165 (1990) [arXiv:gr-qc/0109004].
43. H. J. Schmidt, *Lectures on Mathematical Cosmology* [arXiv:gr-qc/0407095].

Cosmological implications in modified Hořava-Lifshitz gravity

Abdul Jawad

Department of Mathematics, COMSATS University Islamabad, Lahore-Campus, Lahore-54000, Pakistan
E-mail: jawadab181@yahoo.com; abduljawad@cuilahore.edu.pk

Kazuharu Bamba

Division of Human Support System, Faculty of Symbiotic Systems Science, Fukushima University, Fukushima 960-1296, Japan
E-mail: bamba@sss.fukushima-u.ac.jp

Farwa Khurshid

Department of Mathematics, COMSATS University Islamabad, Lahore-Campus, Lahore-54000, Pakistan
E-mail: farwakhursheed1@gmail.com

We imvestigate the cosmological implications for the Sharma-Mittal holographic dark energy such as some cosmological parameters and thermodynamic analysis. Taking into account the apparent horizon with interacting scenario of dark energy and dark matter, the framework of deformed Hořava-Lifshitz gravity is considered. The cosmological parameters include the Hubble parameter, equation of state parameter for the accelerating/deceleratng phases, deceleration parameter to explore the expansion rate and squared speed of sound for stability analysis.

Keywords: Modified gravity theories; autonomous system analysis

1. Introduction

Observational data reveals that the expansion of our cosmos is currently accelerating.[1,2] If Einstein's law of gravity is valid on cosmological scales, this acceleration indicates that our cosmos is governed by a mystery type of energy. This unknown energy component known as dark energy (DE) which has some peculiar characteristics, such as the fact that it is not concentrated on massive distances and that its pressure must be negative in order to cause the universe's current acceleration. DE has appeared among the most major challenges in modern cosmology and in theoretical physics. Dark energy makes up 68 percent of the entire content of the universe, whereas mass energies in the form of dark matter (DM) and ordinary matter make up 27 percent and 5%, respectively, according to the standard model of cosmology. Cosmological investigations of thermodynamics in modified gravity theories have been executed in Refs.[4–10] (for a recent review on thermodynamic properties of modified gravity theories, see, e.g.,[11]).

The accelerated expansion is currently under the limelight of modern research. One of the most challenging problems in modern physics is to find a physical

mechanism that may explain this accelerating universe. This acceleration can be easily accommodated within the scope of general relativity (GR) by adding DE. One of the most significant obstacles to achieving this goal is that GR is not a renormalizable theory in the high-energy limit, the ultraviolet (UV) limit and so there is no control over the theory along with its predictions.[12] Hořava-Lifshitz gravity is a new class of UV-complete gravity theories that attempts to solve this major issue. In the high-energy limit, the Hořava-Lifshitz theory of gravity is derived by rejecting the Lorentz symmetry using a Lifshitz-type re-scaling[13]

$$t \to b^z t^i, \quad x^i \to b x^i, \quad (i = 1, 2, 3, \ldots, d), \tag{1}$$

here z represents the exponent of dynamical scaling, d is a spacetime dimension that expresses a spatial dimension and b is any constant term. The Lorentz symmetry broken when $z \neq 1$ and recovered only when $z = 1$. Hořava-Lifshitz gravity[14] is nonrelativistic at high energies and it's expected that in the low-energy limit, the four-dimensional general covariance may be reconstituted. A lot of research has been done in the literature on modified Hořava-Lifshitz gravity. The deformed Hořava-Lifshitz[15] gravity model has been studied to investigate black hole quasinormal modes.

In this paper, we discuss some cosmological parameters in the framework of deformed Hořava-Lifshitz gravity with Sharma-Mittal holographic DE (SMHDE) model. We discuss the basic equations of SMHDE and deformed Hořava-Lifshitz gravity in the section **2**. In section **3**, we evaluate the cosmological parameters such as Hubble parameter, EoS parameter, deceleration parameter and squared speed of sound parameter. In the last, we discuss the final results.

2. The Deformed Hořava-Lifshitz Modified Gravity and SMHDE Model

The deformed Hořava-Lifshitz theory is a power-counting, renormalizable and ultraviolet total gravity theory. The total action of this gravity can be described as[12]

$$I = \int (\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_m) dt dx^3. \tag{2}$$

From the above expression $\mathcal{L}_m$ is a matter Lagrangian and $\mathcal{L}_1,\ \mathcal{L}_0$ are given by

$$\mathcal{L}_1 = \sqrt{g} N \left[\frac{k^2 \mu^2 (1 - 4\lambda)}{32(1 - 3\lambda)} R^2 - \frac{k^2}{2E^2} (C_{ij} - \frac{\mu E^2}{2} R_{ij})(C^{ij} - \frac{\mu E^2}{2} R^{ij}) \right], \tag{3}$$

$$\mathcal{L}_0 = \sqrt{g} N \left[\frac{k^2 \mu^2 (\Lambda_E R - 3\Lambda_E^2)}{8(1 - 3\lambda)} + \frac{2}{k^2} (K_{ij} K^{ij} - \lambda K^2) \right], \tag{4}$$

where g and N are the determinant of metric tensor and Lapse function respectively, $\mu,\ \Lambda_E$ are mass dimension parameters whereas dimensionless parameters are $E,\ \lambda,\ k^2$. In aditios, $R,\ R_{ij}$ are Ricci scalar and Ricci tensor respectively while the

extrinsic curvature K_{ij} and the Cotton tensor C^{ij} are defined as

$$K_{ij} = \frac{1}{2N}(\partial_t g_{ij} - \nabla_i N_j - \nabla_j N_i), C^{ij} = \frac{\nabla_k}{\sqrt{g}}\epsilon^{ikl}\bigg(- \frac{1}{4}R\delta^j_l + R^j_l\bigg). \tag{5}$$

here ϵ^{ikl} is the antisymmetric unit tensor. The total action of Hořava-Liftshitz gravity takes the form[12]

$$I = \int \bigg[\frac{k^2\mu^2(1-4\lambda)}{32(1-3\lambda)\mu^4 R + R^2} + \frac{k^2\mu}{2E^2}\epsilon^{ijk}R_{il}\nabla_j R^l_k - \frac{\mu^2 k^2}{8}R^{ij}R_{ij} - \frac{k^2}{2E^4}C_{ij}C^{ij} + \frac{2}{8\pi G}(K^{ij}K_{ij} - K^2\lambda)\bigg]dtdx^3. \tag{6}$$

If $c^2 = \frac{k^2\mu^2}{2}$, $G = \frac{k^2}{32\pi C}$ and $\lambda = 1$, the Einstein-Hilbert action in anti de Sitter form can be constructed from the aforementioned action. For the flat FRW universe model, the Friedmann equation of deformed Hořava-Lifshitz gravity can be written as

$$H^2 + \frac{H^4}{2E} = \frac{\rho_{_D} + \rho_{_m}}{3M_p^2}. \tag{7}$$

In the above equation, we can define $\frac{H^2}{2E} = \gamma$, $\frac{\rho_m}{3M_p^2} = \Omega_{_m}$ and $\frac{\rho_D}{3M_p^2} = \Omega_{_D}$. Therefore, the above equation can be re-written as

$$1 + \gamma = \Omega_{_D} + \Omega_{_m}. \tag{8}$$

Also, $\dot{\rho} + 3H(p + \rho) = 0$ is the equation of continuity where ρ and p represent the total energy density and pressure respectively. Interacting of DE and cold DM leads to the following conservation equation

$$\dot{\rho}_{_D} + 3H\rho_{_D}(1 + \omega_{_D}) = -\Gamma, \tag{9}$$

$$\dot{\rho}_{_m} + 3H\rho_{_m} = \Gamma. \tag{10}$$

The coupling parameter Γ shows the mutual exchange rate between DM and DE. The energy exchange from DM to DE when $\Gamma < 0$ and corresponds to reverse scenario for $\Gamma > 0$. In the literature the various models have been proposed for parameter interaction to understand the universe's dynamics. We consider following expression for Γ as

$$\Gamma = 3Hd^2\rho_m, \tag{11}$$

here the coupling constant d exchange the energy between cold DM and DE. Inserting the value of Γ in Eq. (10), we get

$$\rho_m = a^{3(d^2-1)}\rho_{m0}, \tag{12}$$

where ρ_{m0} is an integration constant.

Now we will discuss the SMHDE model. SMHDE model is the new form of HDE model which is proposed due to the inspiration of holographic principle.[16] The Renyi and Tsallis entropies are formulations in which probability distributions are

replaced with power-law distributions, yielding generalized entropies. Tsallis and Renyi entropies are combined in the Sharma-Mittal (SM) entropy. This entropy[17] leads to useful conclusions in cosmological scenarios, such as a description of the present accelerating universe that makes use of vacuum energy. The energy density of SMHDE is given by

$$\rho_{_D} = \frac{3c^2H^4}{8\pi R}\left[\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}}-1\right], \tag{13}$$

where c^2 is the unknown free parameter, $R,\ \delta$ are constants which provide Renyi and Tsallis HDE models under appropriate limits and IR cutoff is taken as Hubble horizon. We attain the following expression by taking the derivative of Eq. (13) w.r.t to t as

$$\dot{\rho}_{_D} = 4\frac{\dot{H}}{H}\rho_{_D} - \frac{3}{4}c^2H\dot{H}\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}-1}. \tag{14}$$

3. Cosmological Parameters

In this section, we discuss the cosmological parameters in the frame work of SMHDE with apparent horizon, such that the Hubble, EoS, deceleration and squared speed of sound parameters.

3.1. *Hubble Parameter*

Taking into account Eqs. (7), (10), (11), (12) and (14), we obtain

$$\dot{H} = \left(9(d^2-1)HH_0^2\Omega_{m0}M_{p0}^2a^{3(d^2-1)} * E\right)\left[6M_p^2H^2(E+H^2)-E\right.$$
$$\left.\times\left(\frac{3c^2H^3}{2\pi R}\left[\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}}-1\right]-\frac{3c^2H}{4}\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}-1}\right)\right]^{-1} \tag{15}$$

The Hubble parameter H versus redshift function $(1+z)$ trajectories are shown in **Figure 1** for different values of $d^2 = 0.4, 0.6, 0.8$ and $M_{p0}^2 = 1,\ E = 80,\ \delta = 0.5,\ c = 1,\ R = 0.1,\ \Omega_{m0} = 0.5,\ H_0 = 74.$ are the other constant parameters. The values of Hubble parameter lie within the range 74.010^{+10}_{-10} for all epochs.

3.2. *Equation of State Parameter*

For the identification of universe at different phases of acceleration and deceleration, we examine the EoS parameter. It is ratio of pressure to density i.e $\omega = \frac{p}{\rho}$. If we take $\omega = 0$, it is a non-relativistic matter while radiation period as decelerated phase of the universe corresponds to $0 < \omega < \frac{1}{3}$ and stiff fluid is for $(\omega = 1)$. During acceleration phase, it represents the phantom epoch for $(\omega < -1)$, quintessence epoch for $(-1 < \omega < \frac{-1}{3})$, and cosmological constant for $(\omega = -1)$. For SMHDE

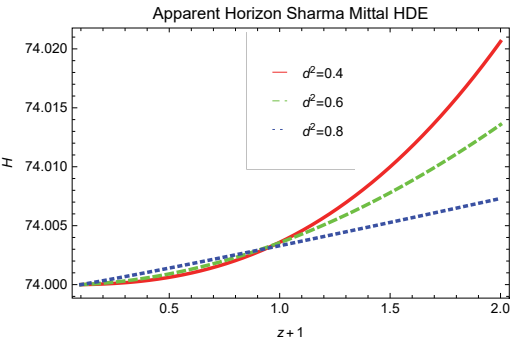


Fig. 1. (Color online) Plot of Hubble parameter H versus $1+z$ for Hubble horizon by choosing $d^2 = 0.4$ (solid red), 0.6 (dashed green), 0.8 (dotted blue).

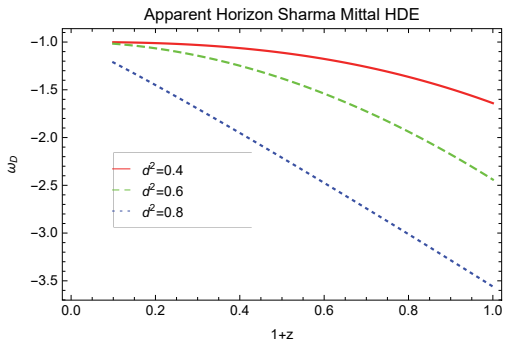


Fig. 2. (Color online) Plot of ω_{DE} versus $1+z$ for Hubble horizon by choosing $d^2 = 0.4$ (solid red), 0.6 (dashed green), 0.8 (dotted blue).

model we check the behavior of EoS parameter by using Eqs. (9), (11), (12) and (14), we get

$$\omega_{_D} = \frac{-3d^2H_0^2M_{p0}^2\Omega_{m0}a^{3(d^2-1)}}{\frac{3c^2H^4}{8\pi R}[(1+\frac{\delta\pi}{H^2})^{\frac{R}{\delta}}-1]} - 1 - \left[\frac{4}{3H^2} - \frac{c^2(1+\frac{\delta\pi}{H^2})^{\frac{R}{\delta}-1}}{\frac{3c^2H^2}{2\pi R}\left((1+\frac{\delta\pi}{H^2})^{\frac{R}{\delta}}-1\right)}\right]$$
$$\times\left(9(d^2-1)HH_0^2\Omega_{m0}M_{p0}^2a^{3(d^2-1)}E\right)\left[6M_p^2H^2(E+H^2)-E\right.$$
$$\left.\times\left(\frac{3c^2H^3}{2\pi R}\left[\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}}-1\right]-\frac{3c^2H}{4}\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}-1}\right)\right]^{-1}. \qquad (16)$$

In **Figure 2**, we plot $\omega_{_D}$ versus $1+z$ by taking same values of constants. The EoS parameter ($\omega_{_D}$) trajectories lie in the phantom regime which follow the cosmological constant behavior at $z=-1$.

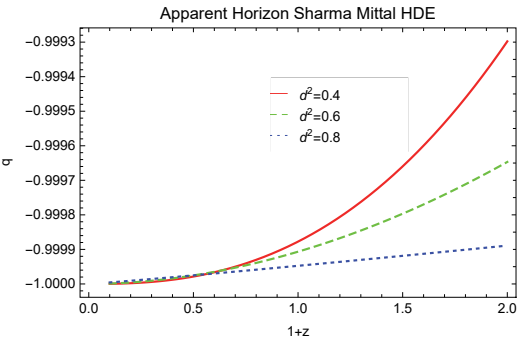


Fig. 3. Plot of q versus $1+z$ for Hubble horizon by choosing $d^2 = 0.4$ (solid red), 0.6 (dashed green), 0.8 (dotted blue).

3.3. *Deceleration Parameter*

We can analyze the accelerating and decelerating rate in different phases of the universe by using deceleration parameter. In terms of Hubble parameter, it is given by

$$q = -\frac{\ddot{a}}{aH^2} = -1 - \frac{\dot{H}}{H^2}. \tag{17}$$

The positive values of this parameter represent decelerated phase and expansion rate as constant when this parameter vanishes. The region described by $q < (-1, 0)$ is accelerating phase with power law expansion, $q = -1$ shows exponential expansion while and $q < -1$ describe super-exponential expansion of the accelerated phase of the expanding universe. In the underlying case, it takes the form

$$q = -1 - \frac{1}{H^2}\left[\left(9(d^2-1)HH_0^2\Omega_{m0}M_{p0}^2a^{3(d^2-1)}E\right)\left(6M_p^2H^2(E+H^2)\right.\right.$$
$$\left.\left. - E\left(\frac{3c^2H^3}{2\pi R}\left[\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}} - 1\right] - \frac{3c^2H}{4}\left(\frac{\delta\pi}{H^2}+1\right)^{\frac{R}{\delta}-1}\right)\right)^{-1}\right]. \tag{18}$$

Considering same values of constants, we plot q versus redshift $(1+z)$ as shown in **Figure 3**. The plot shows the expansion rate at power law level of the accelerating universe for all values of redshift parameter as well as d^2.

3.4. *Squared Speed of Sound Parameter*

In order to check the stability of model, we use squared speed of sound parameter for the underlying case. It is given by

$$v_s^2 == \frac{\dot{p}_D}{\dot{\rho}_D}, \tag{19}$$

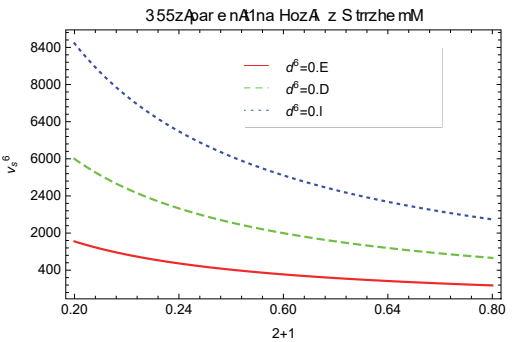


Fig. 4. Plot of υ_s^2 versus $1+z$ for Hubble horizon by choosing $d^2 = 0.4$ (solid red), 0.6 (dashed green), 0.8 (dotted blue).

If υ_s^2 is greater than zero then model is stable, otherwise unstable. For SMHDE model, υ_s^2 becomes

$$\begin{aligned}\upsilon_s^2 = &\Bigg(8RH_0^2\Omega_{m0}M_{p0}^2a^{3(d^2-1)\pi}(2(\delta\pi+H^2)(4d^2M_{p0}^2ER\pi+Ec^2H\\&+4d^2M_{p0}^2H^2R\pi)-c^2\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}}H(-ER\pi+d^2R\pi(1+E)\\&+2E\delta\pi+2EH^2))\Bigg)\Bigg(c^2\Bigg(-1+\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}}H^4(-2(\delta\pi+H^2)\\&\times(4M_{p0}^2ER\pi+Ec^2H+4M_{p0}^2H^2R\pi)+c^2\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}}H\\&\times(R\pi+2E\delta\pi+2EH^2))\Bigg)\Bigg)^{-1}+\dot{\omega}_D\left(\frac{4\dot{H}}{H}-\frac{3c^2H\dot{H}}{4\rho_D}\right.\\&\left.\times\left(1+\frac{\delta\pi}{H^2}\right)^{\frac{R}{\delta}-1}\right)^{-1}-1\end{aligned}\tag{20}$$

Figure 4 describes the plot of $\upsilon_s^2 > 0$ versus redshift parameter for same constant values as in previous cases. The plot expresses positive behavior for all the trajectories which lead to the stable behavior of the model.

4. Conclusions

In this paper, we have considered deformed Hořava-Lifshitz gravity with interaction of dark sector constituents of the universe. The SMHDE model with Hubble horizon is taken into account in order to discuss some cosmological parameters. These parameters are Hubble, EoS, deceleration and squared speed of sound. For different values of interaction parameter, we have investigated the behavior of these parameters graphically for some arbitrary values of model constants. The Hubble

parameter represents consistent results as 74.010^{+10}_{-10} for all trajectories while EoS parameter expresses phantom behavior of the universe. The deceleration parameter indicates the accelerating universe with power law expansion which leads to exponential expansion and squared speed of sound parameter gives the stable behavior of SMHDE models in the framework of deformed Hořava-Lifshitz gravity for all values of interaction parameter.

Acknowledgments

The work of KB was supported in part by the JSPS KAKENHI Grant Number JP21K03547.

References

1. Perlmutter, S., et al., Astrophy. J **517**, 565 (1999).
2. Spergel, D. N., et al., Astrophys. J **170**, 377 (2007).
3. Jawad, A. and Majeed, A.: Astrophy. Space Sci. **356**(2015)375; Jawad, A.: Eur. Phys. J. C **75**(2015)206; Jawad, A., Chattopadhyay, S. and Pasqua, A.: Astrophy. Space Sci. **346**, 273 (2013); Jawad, A., Chattopadhyay, S. and Pasqua, A.: Eur. Phys. J. Plus **128**, 88 (2013); Jawad, A., Chattopadhyay, S. and Pasqua, A.: Eur. Phys. J. Plus **129**, 54 (2014); Jawad, A., Pasqua, A. and Chattopadhyay, S.: Astrophy. Space Sci. **344**, 489 (2013); Jawad, A., Pasqua, A. and Chattopadhyay, S.: Eur. Phys. J. Plus **128**, 156 (2013); Jawad, A.: Astrophy. Space Sci. **353**, 691 (2014); Jawad, A.: Eur. Phys. J. Plus **129**, 207 (2014).
4. K. Bamba and C. Q. Geng, Phys. Lett. B **679**, 282 (2009).
5. K. Bamba, C. Q. Geng and S. Tsujikawa, Phys. Lett. B **688**, 101 (2010).
6. K. Bamba, C. Q. Geng, S. Nojiri and S. D. Odintsov, EPL **89**, 50003 (2010).
7. K. Bamba and C. Q. Geng, JCAP **1006**, 014 (2010).
8. K. Bamba and C. Q. Geng, JCAP **1111**, 008 (2011).
9. K. Bamba, R. Myrzakulov, S. Nojiri and S. D. Odintsov, Phys. Rev. D **85**, 104036 (2012).
10. K. Bamba, M. Jamil, D. Momeni and R. Myrzakulov, Astrophys. Space Sci. **344**, 259 (2013).
11. K. Bamba, Int. J. Geom. Meth. Mod. Phys. **13**, 1630007 (2016).
12. Horava, P., Phys. Rev. D **79**, 084008 (2009).
13. Minamitsuji, M., Phys. Lett. B **684**, 194-198 (2010).
14. Mukohyama, S., Phys. Rev. D **80**, 064005 (2009).
15. Chen, S., Jing, J., Phys. Lett. B **687**, 124-128 (2010).
16. Sharma, U.K., Dubey, V.C., Eur. Phys J. Plus **135**, 1-20 (2020).
17. Jawad, A., Bamba, K., Younas, M., Qummer, S., Rani, S, Symmetry **10**, 635 (2018).

Finite action principle and wormholes

Jan Chojnacki[1]
[1] *Faculty of Physics, University of Warsaw, ul. Pasteura 5, 02-093 Warsaw, Poland*

Jan Kwapisz[1,2]
[1] *Faculty of Physics, University of Warsaw, ul. Pasteura 5, 02-093 Warsaw, Poland*
[2] *CP3-Origins, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*

In this work, we elaborate on the finite action for wormholes in higher derivative theories as well as for wormholes. Both non-traversable and traversable wormholes in theories with higher curvature invariants posses finite action.

Keywords: Horava gravity; Wormholes; finite action

1. Introduction

The Finite Action Principle (FAP) proposes that the physical spacetimes are those for which the action is finite. This reasoning comes from the fact that in Euclidian path integral the weighting factor is $e^{-S[\Phi]}$. Hence lesser the action the more it contributes to the path integral. On the other hand, in the Lorentztian signature the more varying actions contribute less. For an infinitely varying action, one expects a distructive interference with the neighboring fields in such a configuration, see.[1]

This notion has recently received significant attention in the literature, see[2–4] for the cosmological considerations and[1,4] for the investigation of the black holes. Since it is expected that the quantum gravity should resolve the black-hole singularity problem, one may ask which of the microscopic actions remain finite for non-singular black holes and conversely interfere destructively for the singular ones. This we shall call the finite action selection principle. Only after the inclusion of higher-curvature operators, beyond the Einstein-Hilbert term, such selection principle can be satisfied.[1] Furthermore, in asymptotic safety, the quantum corrections to the Newtonian potential eliminate the classical-singularity.[5]

These findings suggest that by taking into account the higher curvatures one can resolve the singularities in black holes. Yet, an issue with the higher-curvature theory of quantum gravity is the existence of the particles with the negative mass-squared spectrum, known as *ghosts*, which makes the theory non-unitary. In this article, we explore possible resolution, namely, we investigate Horava-Lifshitz (H-L) gravity,[6] where the Lorentz Invariance (LI) is broken at the fundamental level (see[7] for a comprehensive progress report on this subject). Kinetic terms are first order in the time derivatives, while higher spatial curvature scalars regulate the UV behavior of the gravity. In our investigation we follow Horava[6] and assume that Wick

rotation is well defined. This can be motivated by the lack of the higher time-like curvature invariants. In the usual higher derivative constructions the existence of massive poles in the propagator makes the Wick rotation troublesome.[8] Conversely to the latter, the new poles are massless and non-tachyonic for suitable choice of parameters.[9] Furthermore, the existence of the foliation supports that claim. For the discussion of Minkowski path integral and FAP, see.[1]

In the previous article, we discussed the possibility that the Finite Action arguments applied to the projectable H-L gravity result in a flat, homogeneous, UV-complete, and ghost-free beginning of the universe.[4] However there is no known regular black hole solutions in HL gravity. Here we discuss the wormholes solutions in the context of the Finite Action Principle in HL gravity, as well as in the $f(R)$ theories. Interestingly the stable and traversable wormholes solutions are known only in the higher derivative gravities[10] (without exotic matter), so there seems to be a wormhole/non-singular BH trade-off after taking into account the Finite Action Principle.

2. Horava-Lifshitz gravity

In the Horava-Lifshitz gravity, space and time are scaled in a non-equivalent way. Diffeomorphism invariance is broken by the foliation of the 4-dimensional spacetime into 3-dimensional hypersurfaces of constant time, called leaves, making the theory power-counting renormalizable (see also the renormalization group studies of the subject[11–13]). The remaining symmetry respects transformations:

$$t \to \xi_0(t), \quad x^i \to \xi^i(t, x^k), \tag{1}$$

and is often referred to as the foliation-preserving diffeomorphism, denoted by $\mathrm{Diff}(M, \mathcal{F})$. The diffeomorphism invariance is still present on the leaves. The four-dimensional metric may be expressed in the Arnowitt-Deser-Misner (ADM)[14] variables:

$$(N, N^i, {}^{(3)}g_{ij}), \tag{2}$$

where N, N^i, ${}^{(3)}g_{ij}$ denote respectively the lapse function, shift vector, and 3-dimensional induced metric on the leaves. The theory is constructed from the following quantities:

$${}^{(3)}R_{ij}, \quad K_{ij}, \quad a_i, \quad {}^{(3)}\nabla_i, \tag{3}$$

where ${}^{(3)}R_{ij}$ is the 3-dimensional Ricci curvature tensor, ${}^{(3)}\nabla_i$ is the covariant derivative constructed from the 3-dimensional metric ${}^{(3)}g_{ij}$, and $a_i := \frac{N_{,i}}{N}$. Extrinsic curvature K_{ij} is the only object, invariant under general spatial diffeomorphisms containing exactly one time derivative of the metric tensor ${}^{(3)}g_{ij}$:

$$K_{ij} = \frac{1}{2N}\left(\frac{\partial\, {}^{(3)}g_{ij}}{\partial t} - {}^{(3)}\nabla_i N_j - {}^{(3)}\nabla_j N_i\right). \tag{4}$$

Quantities (2) are tensor/vectors with respect to Diff$(M, \mathcal{F})$ possessing the following mass dimensions:

$$[\,{}^{(3)}R_{ij}] = 2, \quad [K_{ij}] = 3, \quad [a_i] = 1, \quad [\,{}^{(3)}\nabla_i] = 1. \tag{5}$$

One may use (2) to construct, order by order, scalar terms appearing in the Lagrangian of the theory. Following[7,15] the action of the Horava gravity takes the form:

$$S_g = \zeta^2 \int dt dx^3 N \sqrt{{}^{(3)}g}\,(\mathcal{K} - V)\,, \tag{6}$$

where $\mathcal{K} = K_{ij}K^{ij} - \lambda K^2$ with $K = K_{ij}\,{}^{(3)}g^{ij}$, ${}^{(3)}g$ denotes the determinant of the 3-dimensional metric and $\zeta^2 = 1/16\pi G$. It may be expressed as the difference of the kinetic and potential part $\mathcal{L} = \mathcal{K} - V$ with $\mathcal{K} = \left(K_{ij}K^{ij} - \lambda K^2\right)$. At the 6th order, the potential part of the lagrangian contains over 100 terms.[7] The immense number of invariants is limited by imposing further symmetries. One possible restriction for the potential comes from the projectability condition $N = N(t)$, then terms proportional to $a_i \equiv 0$ vanish. Up to the sixth order (compatible with power counting renormalizability), the potential V restricted by the projectability condition is given by:

$$\begin{aligned} V = 2\Lambda\zeta^2 - {}^{(3)}R + \frac{1}{\zeta^2}\left(g_2\,{}^{(3)}R^2 + g_3\,{}^{(3)}R^{ij}\,{}^{(3)}R_{ij}\right) \\ + \frac{1}{\zeta^4}\left(g_4{}^{(3)}R^3 + g_5\,{}^{(3)}R\,{}^{(3)}R^{ij}\,{}^{(3)}R_{ij} + g_6\,{}^{(3)}R^i_j\,{}^{(3)}R^j_k\,{}^{(3)}R^k_i\right), \\ + \frac{1}{\zeta^4}\left(g_7{}^{(3)}R\nabla^2{}^{(3)}R + g_8(\nabla_i{}^{(3)}R_{jk})(\nabla^i{}^{(3)}R^{jk})\right), \end{aligned} \tag{7}$$

where Λ is the cosmological constant and α_{ij} are the coupling constants. For our purposes, we drop terms containing covariant derivatives ${}^{(3)}\nabla_i$. One should also mention that this *minimal theory*[16] suffers from the existence of spin 0 graviton, which is unstable in the IR. Various solutions to this problem have been proposed. One can add the additional local $U(1)$ symmetry.[7,17] Then by the introduction of new fields prevents the zero-mode from propagating. On the other hand, one can drop the projectability condition $a_i = 0$ and include the terms containing a_i in the potential term:

$$V = 2\Lambda\zeta^2 - {}^{(3)}R - \beta_0 a_i a^i + \sum_{n=3}^{6} \mathcal{L}_V^{(n)}, \tag{8}$$

then for the spin-0 mode to be stable one requires $0 < \beta_0 < 2$.[18,19]

3. Wormholes and finite action principle

Here, we take the first step in the direction of the investigations of the consequences of the Finite Action Principle in the context of wormholes (WH). The wormholes may be characterized in two classes: traversable and non-traversable.

The traversable WH, colloquially speaking, are such that one can go through it to the other side, see[20] for specific conditions. The pioneering Einstein-Rosen bridge has been found originally as a non-static, non-traversable solution to GR. The traversable solutions are unstable, however, they might be stabilized by an exotic matter or inclusion of the higher curvature scalar gravity.[10] This is important in the context of finite action since usually the divergences of black holes do appear in the curvature squared terms. Hence, due to the inclusion of the higher-order terms in the actions, the traversable wormholes are solutions to the equations of motions without the exotic matter. The exemplary wormhole spacetimes investigated here are the Einstein-Rosen bridge proposed in,[21] the Morris-Thorne (MT) wormhole,[20] the traversable exponential metric wormhole[22] and the wormhole solution discussed in the H-L gravity.[23] All of them have a finite action. Here, we shall discuss the exponential metric WH. For the exponential metric WH, the line element is given by:

$$ds^2 = -e^{-\frac{2M}{r}}dt^2 + e^{\frac{2M}{r}}\left(dr^2 + r^2 d\Omega^2\right). \tag{9}$$

This spacetime consists two regions: "our universe" with $r > M$ and the "other universe" with $r < M$. $r = M$ corresponds to the wormhole's throat. The spacial volume of the "other universe" is infinite when $r \to 0$. Such volume divergence is irrelevant to our discussion since it describes large distances in the "other universe". Hence, we further consider only $r \geq M$. The resulting Ricci and Kretschmann scalars have been calculated in[22] and the measure are non-singular everywhere:

$$\begin{aligned} R &= -\frac{2M^2}{r^4}e^{\frac{-2M}{r}}, \\ R_{\mu\nu\sigma\rho}R^{\mu\nu\sigma\rho} &= \frac{4M^2(12r^2 - 16Mr + 7M^2)}{r^8}e^{-4\frac{M}{r}}, \end{aligned} \tag{10}$$

resulting in the finite action for the Stelle gravity. Similarly for the H-L gravity:

$$\begin{aligned} {}^{(3)}R = R, \quad K^2 = K_{ij}K^{ij} = 0, \\ {}^{(3)}R_{ij}{}^{(3)}R^{ij} = \frac{2M^2(M^2 - 2Mr + 3r^2)}{r^8}e^{-\frac{4M}{r}}. \end{aligned} \tag{11}$$

Einstein-Rosen bridge The Einstein-Rosen (E-R) bridge smoothly glues together two copies of the Schwarzschild spacetime: the black hole and the white hole solutions corresponding to the positive and negative coordinate u. The metric tensor of the Einstein-Rosen wormhole proposed in[21] and discussed in e.g.[24] is given by:

$$ds^2 = \frac{-u^2}{u^2 + 4M}dt^2 + (u^2 + 4M)du^2 + \frac{1}{4}(u^2 + 4M)d\Omega^2. \tag{12}$$

The E-R bridge is non-traversable and geodesically incomplete in $u = 0$. This fact, however, does not impact the regularity of the curvature scalars. The 4-dimensional

Ricci scalar is:

$$R = \frac{2\left(64M^2 + 32Mu^2 + 4u^4 + u^2\right)}{\left(4M + u^2\right)^3}. \tag{13}$$

The second order curvature scalar $R_{\mu\nu}R^{\mu\nu}$ is:

$$\frac{4\left(48M^2 + 8\left(4M + u^2\right)^4 + (32M - 1)\left(4M + u^2\right)^2\right)}{\left(4M + u^2\right)^6}. \tag{14}$$

Both of which integrated with the measure are non-singular:

$$\sqrt{g} = \frac{1}{4}u(4M + u^2). \tag{15}$$

The wormhole solutions analyzed in this paper generally yield the finite action in both GR and H-L. The finite Action Principle suggests, that in the quantum UV regime, singular black-hole spacetimes may be replaced with the regular wormhole solutions.

Morris-Thorne wormhole The MT wormhole is defined in the spherically symmetric, Lorentzian spacetime by the line element:

$$ds^2 = -e^{2\Phi(r)}dt^2 + \frac{dr^2}{1 - \frac{b(r)}{r}} + r^2 d\Omega^2 \tag{16}$$

where $\Phi(r)$ is known as the redshift and there are no horizons if it is finite. Function $b(r)$ determines the wormhole's shape. We choose $\Phi(r)$, $b(r)$ to be:

$$\Phi(r) = 0, \quad b(r) = 2M\left(1 - e^{r_0 - r}\right) + r_0 e^{r_0 - r}, \tag{17}$$

where r_0 is the radius of the throat of the wormhole, such that $b(r_0) = r_0$. 4-dimensional curvature scalars for this spacetime have been calculated in.[25] The Ricci curvature scalar is singular at $r = 0$, however, the radial coordinate r varies between $r_0 > 0$ and infinity:

$$R = -2\left(2M - r_0\right)\frac{e^{r_0 - r}}{r^2}. \tag{18}$$

The resulting $S_s = \int_{r_{UV}=r_0}^{r_{IR}} \sqrt{g}R$ function is divergent as $r_{UV} \to r_0$ and cannot be expressed in terms of simple functions:

$$2(2M - r_0)\int_{r_{UV}=r_0}^{r_{IR}} \sqrt{\frac{r}{r - 2M(1 - e^{r_0 - r}) + r_0 e^{r_0 - r}}}\, e^{r_0 - r} dr. \tag{19}$$

However, this is only a coordinate singularity and one may get rid of it with a proper transformation.

Higher-order curvature scalars for Morris-Thorne wormhole are:

$$
\begin{aligned}
{}^{(3)}R &= \frac{2b'(r)}{r^2} \\
{}^{(3)}R_{ij}\,{}^{(3)}R^{ij} &= \frac{3r^2b'(r)^2 - 2rb(r)b'(r) + 3b(r)^2}{2r^6}, \\
{}^{(3)}R^i_j\,{}^{(3)}R^j_k\,{}^{(3)}R^k_i &= \frac{-9r^2b(r)b'(r)^2 + 5r^3b'(r)^3 + 15rb(r)^2b'(r) - 3b(r)^3}{4r^9},
\end{aligned} \tag{20}
$$

and integrated give action that is finite.

H-L wormhole Static spherically traversable symmetric wormholes have been constructed in[23] in the H-L theory through the modification of the Rosen-Einstein spacetime:

$$
ds^2 = -N^2(\rho)dt^2 + \frac{1}{f(\rho)}d\rho^2 + (r_0+\rho^2)^2d\Omega^2, \tag{21}
$$

with additional $\mathbf{Z}_2$ symmetry with respect to the wormhole's throat. There are solutions with $\lambda = 1$ asymptotically corresponding to the Minkowski vacuum. Explicitly we have:

$$
\begin{aligned}
f = N^2 &= 1 + \omega(r_0+\rho^2)^2 \\
&\quad - \sqrt{(r_0+\rho^2)\left(\omega^2(r_0+\rho^2)^3 + 4\omega M\right)}.
\end{aligned} \tag{22}
$$

The radius of the wormhole's throat is given by r_0. The parameters ω, and M are connected to the coupling constants in H-L action. See[23] for their explicit form. Ricci scalar of the H-L wormhole invariants are given by

$$
\begin{aligned}
{}^{(3)}R = &-\frac{1}{(\rho^2+r_0)^2}\Big[2(-10\rho^2\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} \\
&- 4r_0\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} + 16\rho^6\omega + 8\rho^2 \\
&+ 36\rho^4r_0\omega + 24\rho^2r_0^2\omega + 4r_0^3\omega + 4r_0 - 1)\Big], \\
{}^{(3)}R_{ij}\,{}^{(3)}R^{ij} = &\frac{1}{(\rho^2+r_0)^4}\Big\{2(-7\rho^2\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} \\
&-2r_0\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} \\
&+ 10\rho^6\omega + 6\rho^2 + 22\rho^4r_0\omega + 14\rho^2r_0^2\omega + 2r_0^3\omega + 2r_0 - 1)^2 + 4M \\
&+ \frac{4(\rho^2+r_0)}{\omega(4M+\omega(r_0+\omega^2)^3)}\Big[\rho^2\omega(-4(\rho^2+r_0) \\
&\times\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} + \omega(r_0+\omega^2)^3)\Big] \\
&-2\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} \\
&\times\left[-\sqrt{\omega(\rho^2+r_0)(4M+\omega(r_0+\omega^2)^3)} + \omega(\rho^2+r_0)^2+1)\right]^2\Big\}
\end{aligned} \tag{23}
$$

The kinetic terms with $K_{ij} = 0$ are vanishing, while the spacial Ricci scalar and higher curvature terms are finite.

From the point of view of the Finite Action Principle, all of the investigated wormhole spacetimes are included in the gravitational path integral.

4. Conclusions and discussion

The Finite Action Principle is a powerful tool to study quantum gravity theories and also QFTs in general. From the point of view of finite action selection principle[1,4] they are equally good theories, resolving the black holes singularities, assuming that ghost issue is resolved in the latter case. Yet none of the regular B-H solutions have been found in the context of H-L gravity[26] and the ones found for $f(R)$ suffer from mass inflation issue.[27–29] Hence, it is a strong suggestion that wormholes may appear in the UV regime of H-L gravity and can serve as a "cure" for singularities.[30–32]

Moreover, both the traversable and the non-traversable wormholes are on equal footing in the case of the Finite Action Principle, yet maybe the finite amplitude principle could distinguish between those,[3] since the amplitude to cross non-traversable wormhole should be different than the traversable one. However, this principle suggests that there is a trade-off between the resolution of black-hole singularities and the appearance of wormhole spacetimes due to higher curvature invariants. The wormhole solutions will remain in both the LI and H-L path integrals. The higher-order curvature scalars, generically present in the quantum gravity, stabilise the wormhole solutions without the need for an exotic matter.

Acknowledgements

We thank J. N. Borrisova, A. Eichhorn and A. Wang for inspiring discussions and careful reading of the manuscript. J.H.K. was supported by the Polish National Science Centre grant 2020/39/B/ST2/01279. J.H.K. would like to thank the CP3-Origins for the extended hospitality during this work. J.H.K acknowledges the NAWA Iwanowska scholarship PPN/IWA/2019/1/00048.

References

1. J. N. Borissova and A. Eichhorn, Universe **7**, 48 (2021), arXiv:2012.08570 [gr-qc].
2. J. D. Barrow and F. J. Tipler, Nature **331**, 31 (1988).
3. C. Jonas, J.-L. Lehners, and J. Quintin, Phys. Rev. D **103**, 103525 (2021), arXiv:2102.05550 [hep-th].
4. J. Chojnacki and J. Kwapisz, (2021), arXiv:2102.13556 [gr-qc].
5. L. Bosma, B. Knorr, and F. Saueressig, Phys. Rev. Lett. **123**, 101301 (2019), arXiv: 1904.04845 [hep-th].
6. P. Hořava, Physical Review D **79** (2009), 10.1103/physrevd.79.084008.
7. A. Wang, International Journal of Modern Physics D **26**, 1730014 (2017).
8. J. F. Donoghue and G. Menezes, Phys. Rev. D **100**, 105006 (2019), arXiv:1908.02416 [hep-th].
9. F. S. Bemfica and M. Gomes, Phys. Rev. D **84**, 084022 (2011), arXiv:1108.5979 [hep--th].

10. F. Duplessis and D. A. Easson, Physical Review D **92** (2015), 10.1103/physrevd.92.043516.
11. G. D'Odorico, F. Saueressig, and M. Schutten, Phys. Rev. Lett. **113**, 171101 (2014), arXiv:1406.4366 [gr-qc].
12. G. D'Odorico, J.-W. Goossens, and F. Saueressig, JHEP **10**, 126 (2015), arXiv:1508.00590 [hep-th].
13. A. O. Barvinsky, D. Blas, M. Herrero-Valea, S. M. Sibiryakov, and C. F. Steinwachs, Phys. Rev. Lett. **119**, 211301 (2017), arXiv:1706.06809 [hep-th].
14. R. Arnowitt, S. Deser, and C. W. Misner, General Relativity and Gravitation **40**, 1997–2027 (2008).
15. R. Maier and I. D. Soares, Physical Review D **96** (2017), 10.1103/physrevd.96.103532.
16. P. Horava, Class. Quant. Grav. **28**, 114012 (2011), arXiv:1101.1081 [hep-th].
17. P. Horava and C. M. Melby-Thompson, Phys. Rev. D **82**, 064027 (2010), arXiv:1007.2410 [hep-th].
18. D. Blas, O. Pujolas, and S. Sibiryakov, JHEP **04**, 018 (2011), arXiv:1007.3503 [hep--th].
19. S. Carloni, E. Elizalde, and P. J. Silva, Class. Quant. Grav. **28**, 195002 (2011), arXiv:1009.5319 [hep-th].
20. M. S. Morris and K. S. Thorne, Am. J. Phys. **56**, 395 (1988).
21. A. Einstein and N. Rosen, Phys. Rev. **48**, 73 (1935).
22. P. Boonserm, T. Ngampitipan, A. Simpson, and M. Visser, Physical Review D **98** (2018), 10.1103/physrevd.98.084048.
23. M. Botta Cantcheff, N. E. Grandi, and M. Sturla, Physical Review D **82** (2010), 10.1103/physrevd.82.124034.
24. M. O. Katanaev, Modern Physics Letters A **29**, 1450090 (2014).
25. B. Mattingly, A. Kar, W. Julius, M. Gorban, C. Watson, M. Ali, A. Baas, C. Elmore, B. Shakerin, E. Davis, and et al., Universe **6**, 11 (2020).
26. A. Wang, "private communication".
27. R. Carballo-Rubio, F. Di Filippo, S. Liberati, C. Pacilio, and M. Visser, JHEP **07**, 023 (2018), arXiv:1805.02675 [gr-qc].
28. A. Bonanno, A.-P. Khosravi, and F. Saueressig, Phys. Rev. D **103**, 124027 (2021), arXiv:2010.04226 [gr-qc].
29. M. Bertipagani, M. Rinaldi, L. Sebastiani, and S. Zerbini, Phys. Dark Univ. **33**, 100853 (2021), arXiv:2012.15645 [gr-qc].
30. F. S. N. Lobo, Int. J. Mod. Phys. D **25**, 1630017 (2016), arXiv:1604.02082 [gr-qc].
31. G. J. Olmo, D. Rubiera-Garcia, and A. Sanchez-Puente, in *14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories*, Vol. 2 (2017) pp. 1391–1396, arXiv:1601.00161 [gr-qc].
32. C. Bambi, A. Cardenas-Avendano, G. J. Olmo, and D. Rubiera-Garcia, Phys. Rev. D **93**, 064016 (2016), arXiv:1511.03755 [gr-qc].

Strange quark stars in Hořava gravity

Grigoris Panotopoulos

Departamento de Ciencias Físicas, Universidad de la Frontera, Casilla 54-D, 4811186 Temuco, Chile

grigorios.panotopoulos@ufrontera.cl

Non-rotating strange quark stars made of isotropic matter in Lorentz-violating theories of gravity are studied. In particular, Hořava gravity and Einstein-æther theory are considered. For quark matter we adopt both linear and non-linear equations-of-state, corresponding to the MIT bag model and color flavor locked phase, respectively. The new, modified structure equations generalize the usual Tolman-Oppenheimer-Volkoff equations valid in Einstein's General Relativity. A dimensionless parameter ν measures the deviation from the standard TOV equations, which are recovered in the appropriate limit. We compute some properties, such as masses, radii as well as the factor of compactness of the stars, and we show pictorially the impact of the parameter ν on the mass-to-radius relationships for several different equations-of-state. Other physical considerations, such as stability criteria, causality and energy conditions, are also considered, and they are all found to be fulfilled.

Keywords: Relativistic stars; Composition of astronomical objects; Theories of gravity other than GR.

1. Introduction

Einstein's General Relativity (GR)[1] is a relativistic, geometric theory of gravitation, which not only is beautiful but also very successful.[2,3] The classical tests and solar system tests,[4] and recently the direct detection of gravitational waves by the LIGO/VIRGO Collaborations[5] have confirmed a series of remarkable predictions of GR.

Despite its success, however, it it believed that GR must be somehow extended for several different reasons. One may mention, for example, the current cosmic acceleration,[6,7] one of the most dramatic discoveries in Cosmology in the end of the 90's. If a modified theory of gravity is assumed, correction terms to GR on cosmological scales are expected to appear, and those terms of geometrical origin could play the role of dark energy,[8] the fluid component that dominates the expansion of the Universe at late time. In this class of models one finds for instance $f(R)$ theories of gravity,[9–12] brane models[13–15] and scalar-tensor theories of gravity.[16–18]

Moreover, it has been known for a long time that GR is a classical, non-renormalizable theory of gravitation. A theory of gravity incorporating quantum mechanics in a consistent way is still one of the biggest challenges in Modern Theoretical Physics. In 2009 Hořava gravity[19,20] was proposed as a new candidate

theory for quantum gravity which explicitly breaks Lorentz invariance at any energy scale by introducing a preferred foliation of spacetime. Since then a lot of work has been done to prove, very successfully, its renormalizability by means of both power-counting arguments[21–26] and quantum field theory approaches.[27–30] Besides, a lot of effort has been made in order to investigate its phenomenological implications, e.g. concerning late-time Cosmology,[31,32] black hole physics,[33–39] binary systems,[40–42] and interior solutions of anisotropic stars.[43–45] In particular, after the multiple detections of gravitational waves by the LIGO-VIRGO Collaboration, and in particular the first merger observed from a binary of neutron stars,[46] a new era for gravitational astronomy get started. Very interestingly, Hořava gravity passes with flying colors all the theoretical and observational constraints which are available to date.[47] It is worth noticing that in the low-energy limit of Hořava gravity, its action written in a covariant form becomes equivalent to Einstein-æther theory[48] once the æther vector is restricted to be hypersurface-orthogonal at the level of the action.[49] In spherical symmetry it can be shown that the two theories share the same solutions.[50]

Therefore, it has become even more urgent to investigate the astrophysical implications and predictions of viable alternative theories of gravity, in order to explore non-standard scenarios and the possible signatures of deviations from GR to be observed in the forthcoming missions/space-based detectors. In this respect, compact objects,[51–53] such as neutron stars and white dwarfs, are relativistic stars of astrophysical and astronomical interest, which are characterized by ultra dense matter content and strong gravitational fields. Thanks to their properties, naturally they serve as ideal cosmic laboratories to study, test and constrain non-standard physics as well as non-conventional theories of gravity.

A new class of theoritized compact objects, that may be an alternative to neutron stars, are some as of today hypothetical objects which are supposed to be made of quark matter, and for that reason they are called strange quark stars.[54–59] Quark matter is by assumption absolutely stable,[60,61] and so it could be the true ground state of hadrons. That property provides us with a plausible explanation of some puzzling super-luminous supernovae,[62,63] which occur in about one out of every 1000 supernovae explosions, and which are more than 100 times more luminous than regular supernovae.

Our work is organized as follows. In Sec. 2 we briefly review the basics of Hořava gravity, and its connection to Einstein-æther theory, while in Sec. 3 we present the field equations as well as the structure equations describing hydrostatic equilibrium of static, spherically symmetric relativistic stars made of isotropic matter. In Sec. 4 we obtain and discuss our numerical results for quark stars. Finally, we finish our work with some concluding remarks in Sec. 5. We adopt the mostly negative metric signature $+,-,-,-$, and we work in natural units where $\hbar = 1 = c$. In those units all dimensionful quantities are measured in GeV $= 10^3$ MeV, and we make use of the conversion rules: 1 m $= 5.068 \times 10^{15}$ GeV^{-1} and 1 kg $= 5.610 \times 10^{26}$ GeV.[64]

2. Hořava gravity and Einstein-æther theory

The action of Hořava gravity[19,20] can be written in the preferred foliation as

$$\mathcal{S}_H = \frac{1}{16\pi G_H}\int dTd^3x\sqrt{-g}\left(K_{ij}K^{ij} - \lambda K^2 + \xi\mathcal{R} + \eta a_i a^i + \frac{L_4}{M_*^2} + \frac{L_6}{M_*^4}\right) + S_m[g_{\mu\nu},\psi]\,, \tag{1}$$

where G_H is the effective gravitational constant; g is the determinant of the metric $g_{\mu\nu}$; $\mathcal{R}$ is the Ricci scalar of the three-dimensional constant-T hypersurfaces; K_{ij} is the extrinsic curvature and K is its trace; and $a_i = \partial_i \ln N$, where N is the lapse function and S_m is the matter action where ψ collectively denotes the matter fields. The constant couplings $\{\lambda,\xi,\eta\}$ are dimensionless, and GR is identically recovered when they take the values $\{1,1,0\}$, respectively. Furthermore, L_4 and L_6 stand for the fourth-order and sixth-order operators respectively, which are suppressed by a characteristic mass scale M_*.

In the following, we consider the covariantized version of the low-energy limit of Hořava gravity, named the *khronometric* model, that is obtained by keeping only the operators up to second-order derivatives, which amounts to discarding L_4 and L_6 which instead contain the higher-order operators.

In order to write the action covariantly, let us first take the action of Einstein-æther theory:[48]

$$\mathcal{S}_{æ} = \frac{1}{16\pi G_{æ}}\int d^4x\sqrt{-g}\,(-R + L_{æ}) + S_m[g_{\mu\nu},\psi]\,, \tag{2}$$

where $G_{æ}$ is the "bare" gravitational constant; R is the four-dimensional Ricci scalar; u^a is a timelike vector field of unit norm, *i.e.*, $g_{\mu\nu}u^\mu u^\nu = 1$, from now on referred to as the "æther"; and

$$L_{æ} = -M^{\alpha\beta}{}_{\mu\nu}\nabla_\alpha u^\mu \nabla_\beta u^\nu\,, \tag{3}$$

with $M^{\alpha\beta}{}_{\mu\nu}$ defined as

$$M^{\alpha\beta}{}_{\mu\nu} = c_1 g^{\alpha\beta}g_{\mu\nu} + c_2\delta^\alpha_\mu\delta^\beta_\nu + c_3\delta^\alpha_\nu\delta^\beta_\mu + c_4 u^\alpha u^\beta g_{\mu\nu}\,, \tag{4}$$

where c_i's are dimensionless coupling constants.

Once the æther vector is taken to be hypersurface-orthogonal at the level of the action, that is

$$u_\alpha = \frac{\partial_\alpha T}{\sqrt{g^{\mu\nu}\partial_\mu T\partial_\nu T}}\,, \tag{5}$$

where the preferred time T is a scalar field (the *khronon*) which defines the preferred foliation, then the two actions in Eqs. (1) and (2) become equivalent if the parameters of the two theories are mapped into each other as[49]

$$\frac{G_H}{G_{æ}} = \xi = \frac{1}{1-c_{13}}\,, \qquad \frac{\lambda}{\xi} = 1 + c_2\,, \qquad \frac{\eta}{\xi} = c_{14}\,, \tag{6}$$

where $c_{ij} = c_i + c_j$. Moreover $G_{æ} = G_N\,(1 - \eta/2\xi)$, where G_N is the Newton's constant, which is needed to recover the Newtonian limit.[20,65] In what follows, we will thus consider the covariant formulation of the low-energy limit of Hořava gravity.

The variation of the action in Eq. (2) with respect to $g^{\alpha\beta}$ and T yields, respectively,[42]

$$G_{\alpha\beta} - T^{æ}_{\alpha\beta} = 8\pi G_{æ} T^{m}_{\alpha\beta}\,, \tag{7}$$

$$\partial_\mu \left(\frac{1}{\sqrt{\nabla^\alpha T \nabla_\alpha T}} \sqrt{-g} Æ^\mu \right) = 0\,, \tag{8}$$

where $G_{\alpha\beta} = R_{\alpha\beta} - R g_{\alpha\beta}/2$ is the Einstein tensor,

$$T^{æ}_{\alpha\beta} = \nabla_\mu \left(J_{(\alpha}{}^{\mu} u_{\beta)} - J^{\mu}{}_{(\alpha} u_{\beta)} - J_{(\alpha\beta)} u^\mu \right) + c_1 \left[(\nabla_\mu u_\alpha)(\nabla^\mu u_\beta) - (\nabla_\alpha u_\mu)(\nabla_\beta u^\mu) \right] + \left[u_\nu (\nabla_\mu J^{\mu\nu}) - c_4 \dot{u}^2 \right] u_\alpha u_\beta + c_4 \dot{u}_\alpha \dot{u}_\beta - \frac{1}{2} L_{æ} g_{\alpha\beta} + 2Æ_{(\alpha} u_{\beta)} \tag{9}$$

is the khronon stress-energy tensor,

$$J^{\alpha}{}_{\mu} = M^{\alpha\beta}{}_{\mu\nu} \nabla_\beta u^\nu\,, \quad \dot{u}_\nu = u^\mu \nabla_\mu u_\nu\,, \quad Æ_\mu = \left(\nabla_\alpha J^{\alpha\nu} - c_4 \dot{u}_\alpha \nabla^\nu u^\alpha \right) \left(g_{\mu\nu} - u_\mu u_\nu \right)\,, \tag{10}$$

and $T^{m}_{\alpha\beta}$ is the matter stress-energy tensor, defined as

$$T^{m}_{\alpha\beta} = \frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{\alpha\beta}}\,. \tag{11}$$

3. Field equations

The most general static, spherically symmetric geometry, in Schwarzschild-like coordinates, may be written down as follows

$$ds^2 = e^{A(r)} dt^2 - B(r) dr^2 - r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right)\,. \tag{12}$$

In addition, let us consider an interior spacetime filled by an isotropic fluid whose stress-energy tensor is

$$T^{m}_{\alpha\beta} = (\rho + p)\, v_\alpha v_\beta - p g_{\alpha\beta}\,, \tag{13}$$

where ρ is the density and p is the pressure of the fluid, and its 4-velocity v^α is given by

$$v^\alpha = \left(e^{-A(r)/2}, 0, 0, 0 \right)\,. \tag{14}$$

The æther vector field, which is by definition a unit timelike vector, in spherical symmetry is always hypersurface-orthogonal and takes the following general form:

$$u^\alpha = \left(F(r), \sqrt{\frac{e^{A(r)} F(r)^2 - 1}{B(r)}}, 0, 0 \right)\,. \tag{15}$$

However, in the following we will consider the case of a static aether which is aligned with the interior matter fluid 4-velocity v^α, *i.e.* when $F(r) = e^{-A(r)/2}$, which leads to

$$u^\alpha = \left(e^{-A(r)/2}, 0, 0, 0\right). \tag{16}$$

The field equations that we have to consider are the modified Einstein equations (0-0), (1-1) and (2-2) in Eq. (7), which can be written respectively as:

$$-\nu\left[4r^2A''(r) - \frac{2r^2A'(r)B'(r)}{B(r)} + r^2A'(r)^2 + 8rA'(r)\right] + \frac{rB'(r)}{B(r)} + B(r) - 1$$
$$= 8\pi G_{æ} r^2 B(r)\rho(r)\,, \tag{17}$$

$$\nu r^2 A'(r)^2 + rA'(r) - B(r) + 1 = 8\pi G_{æ} r^2 B(r) p(r)\,, \tag{18}$$

$$\frac{1}{2}r^2A''(r) - \frac{r^2A'(r)B'(r)}{4B(r)} - \nu r^2 A'(r)^2 + \frac{1}{4}r^2A'(r)^2 + \frac{1}{2}rA'(r) - \frac{rB'(r)}{2B(r)}$$
$$= 8\pi G_{æ} r^2 B(r) p(r)\,, \tag{19}$$

where $\nu = \frac{\eta}{8\xi}$. Furthermore, Eq. (8) is identically satisfied. In the following we shall set $8\pi G_{æ} = 1$. Finally, the conservation equation for the energy-momentum tensor takes the form

$$p'(r) + \frac{1}{2}A'(r)\left[\rho(r) + p(r)\right] = 0\,. \tag{20}$$

Notice that the standard Tolman-Oppenheimmer-Volkoff (TOV) equations,[66–69] which hold in GR,[1] are recovered setting $\nu = 0$ in the equations above.

Despite the fact that there are four equations in total, it turns out that only three of them are independent. Furthermore, since spherically symmetric solutions in Hořava gravity are identical to those of Einstein-æther theory, all of our conclusions will hold for both theories.[50]

In order to derive the modified TOV equations that will be used for the numerical integration, one can obtain $A'(r)$ from Eq. (18):

$$A'(r) = -\frac{r - r\sqrt{4\nu r^2 B(r)p(r) + 4\nu B(r) - 4\nu + 1}}{2\nu r^2}\,, \tag{21}$$

which corresponds to the only branch which admits the proper GR limit when $\nu \to 0$. Then, one has to substitute the latter in Eq. (17) and Eq. (20), that for brevity we do not show here. However, one can immediately notice the difference with respect to GR, since here the parameter ν (which in GR is identically zero) enters non-linearly in the resulting equations.

4. Properties of strange quark stars: Numerical treatment

In this section we investigate the properties of strange quark stars within the Lorentz-violating theories of gravity at hand. We integrate numerically the generalized structure equations, and after that we present and discuss the results.

4.1. *Vacuum solution*

In order to compute the radius and the mass of the objects, we need to match the solutions at the surface of the stars. For that we need to know the exterior (vacuum) solution first, and so we set the stress-energy tensor (i.e. both the pressure and the energy density of the matter content) to zero.

In order to calculate the total gravitational mass M of the fluid star appearing in the Newtonian potential, we will make use of the general vacuum solution found in Ref. 70 and summarized below:

$$e^A = \left(\frac{1 - Y/Y_-}{1 - Y/Y_+}\right)^{\frac{-Y_+}{2+Y_+}} , \tag{22}$$

$$B = \nu(Y - Y_-)(Y - Y_+) , \tag{23}$$

$$\frac{r_{\min}}{r} = \left(\frac{Y}{Y - Y_-}\right)\left(\frac{Y - Y_-}{Y - Y_+}\right)^{\frac{1}{2+Y_+}} , \tag{24}$$

where $Y = rA'$, $Y_\pm = (-1 \pm \sqrt{1 - 4\nu})/(2\nu)$, and $r_{\min}$ is an integration constant which is related to the gravitational radius $r_g = 2G_N M$ by

$$r_{\min}/r_g = (-Y_+)^{-1}(-1 - Y_+)^{(1+Y_+)/(2+Y_+)} . \tag{25}$$

The above solution agrees with the Schwarzschild solution[71] of GR to leading order in $1/r$.

4.2. *Equation-of-state*

Before we start integrating the structure equations, we must specify the matter content first, or in other words pick up a certain equation-of-state. Quark matter inside stars is commobly described by the MIT bag model.[72,73] In the simplest version there is a linear analytic function relating the energy density to the pressure of the fluid, which reads

$$p = k(\rho - \rho_s) , \tag{26}$$

where k is a dimensionless numerical factor, while ρ_s is the surface energy density. The MIT bag model is characterized by 3 parameters, namely i) the QCD coupling constant, α_c, ii) the mass of the strange (s) quark, m_s, and iii) the bag constant, B_0. In this work we shall consider the following 3 models:[74]

- The extreme model SQSB40 where $m_s = 100$ MeV, $\alpha_c = 0.6$ and $B_0 = 40$ MeV fm^{-3}. In this model $k = 0.324$ and $\rho_s = 3.0563 \times 10^{14}$ g cm^{-3}.

- The standard model SQSB56 where $m_s = 200$ MeV, $\alpha_c = 0.2$ and $B_0 = 56$ MeV fm^{-3}. In this model $k = 0.301$ and $\rho_s = 4.4997 \times 10^{14}$ g cm^{-3}.
- The simplified model SQSB60 where $m_s = 0 = \alpha_c$ and $B_0 = 60$ MeV fm^{-3}. In this model $k = 1/3$ and $\rho_s = 4.2785 \times 10^{14}$ g cm^{-3}.

Furthermore, at asymptotically large densities color superconductivity effects[75,76] become important. Quark matter is found to be in the so called color flavor locked (CFL) state,[77,78] in which quarks form Cooper pairs of different color and flavor, and where all quarks have the same Fermi momentum and electrons cannot be present. That quark state is described by a slightly more complicated equation of state (EoS), although still an analytic function, and it is given by the following non-linear relation[79–81]

$$\rho = 3p + 4B_0 - \frac{9\gamma\mu^2}{\pi^2}\,, \tag{27}$$

where γ and μ^2 are given by

$$\gamma = \frac{2\Delta^2}{3} - \frac{m_s^2}{6}\,, \tag{28}$$

and

$$\mu^2 = -3\gamma + \left(9\gamma^2 + \frac{4}{3}\pi^2(B_0 + p)\right)^{1/2}\,, \tag{29}$$

with Δ being the non-vanishing energy gap.

In the CFL state there are 19 viable models, but here we shall consider two, namely CFL4 and CFL10, characterized by the following parameters (look at table I in Ref. 81):

$$\Delta = 100 \text{ MeV}\,, \tag{30}$$
$$m_s = 150 \text{ MeV}\,, \tag{31}$$
$$B_0 = 60 \text{ MeV fm}^{-3}\,, \tag{32}$$

for CFL4, and

$$\Delta = 150 \text{ MeV}\,, \tag{33}$$
$$m_s = 150 \text{ MeV}\,, \tag{34}$$
$$B_0 = 80 \text{ MeV fm}^{-3}\,, \tag{35}$$

for CFL10.

4.3. *Numerical solution of structure equations*

The radius of the stars, R, is determined from the requirement that $p(r = R) = 0$, while the mass of the stars, M, is computed numerically using the vacuum solution presented before, and requiring that $B_{int}(r = R) = B_{ext}(r = R)$. Finally, the compactness of the object is computed by $C = G_N M/R$.

Our main numerical results are summarized in the figures below. At this point it should be mentioned that the 3 observed super-massive pulsars J1614-2230 ($(1.928 \pm 0.017)\ M_\odot$),[82,83] J0348+0432 ($(2.01 \pm 0.04)\ M_\odot$)[84] and J0740+6620 ($(2.14^{+0.20}_{-0.18})\ M_\odot$),[85] with masses $M \sim 2M_\odot$ have put stringent constraints on compact object modelling, since any EoS that does not cross the 2 solar mass strip must be ruled out. Furthermore, the recent observation of the highly massive pulsar in the binary J2215 + 5135 ($(2.27^{+0.17}_{-0.15})\ M_\odot$)[86] has put an even more tight constraint to be satisfied. However, our results will also depend on the assumed EoS, as the predicted mass-to-radius profiles are affected. For that reason, our goal is not to show that the models discussed here are able to pass all current tests, but just to study the general features of quark stars in the framework of Hořava gravity and Einstein-æther theory, and the impact of the modifications that they induce on the resulting mass-radius relationships. Moreover, the reader should also keep in mind that electrically charged stars or compact objects with anisotropic matter can have higher masses compared to their isotropic neutral counterparts, see e.g. Refs. 87, 88 and references therein.

In Fig. 1 we show the mass-to-radius (M_R) relations (left panel) and the compactness (right panel) of the quark stars with a linear EoS. Similarly, Fig. 2 shows the same properties of the stars with a non-linear EoS. We see that the M_R profiles are shifted downwards as the parameter ν increases. Therefore, if the highest star mass that a given EoS can support is lower than the 2 solar mass limit in GR, it will become even worse in Hořava gravity and Einstein-æther theory.

The speed of sound, defined by $c_s^2 \equiv dp/d\rho$, is shown in Fig. 3 for the models CFL4 and CFL10, while when the EoS is linear the sound speed is just a constant, $c_s^2 = k$. We have considered the cases where $\nu = 0.01, 0.02$, while for GR the speed of sound for the 19 CFL viable models can be seen in Fig. 2 (panel (b)) of Ref. 81. Clearly, throughout the object the speed of sound takes values in the range $0 < c_s^2 < 1$, as it should, and therefore causality is not violated. It should be stated that for each model the 2 curves corresponding to a different value of the parameter ν are indistinguishable.

The solutions obtained here should be able to describe realistic astrophysical configurations. Therefore, as a final check we investigate i) stability criteria, and ii) if the energy conditions are fulfilled or not. For the latter, we require that[89–93]

$$\rho \geq 0\,, \tag{36}$$

$$\rho - p \geq 0\,, \tag{37}$$

$$\rho - 3p \geq 0\,. \tag{38}$$

In Fig. 4 we show the normalized pressure, $\bar{p} \equiv p/B_0$, $3\bar{p}$, and normalized energy density, $\bar{\rho} \equiv \rho/B_0$ (from bottom to top) versus the normalized radial coordinate r/R for $\nu = 0.01, 0.02$ and $p_c(0) = 1.5\ B_0$ for the models CFL4 and CFL10. Notice that the lower curves corresponding to $\bar{p}$ and $3\bar{p}$ are indistinguishable for all four

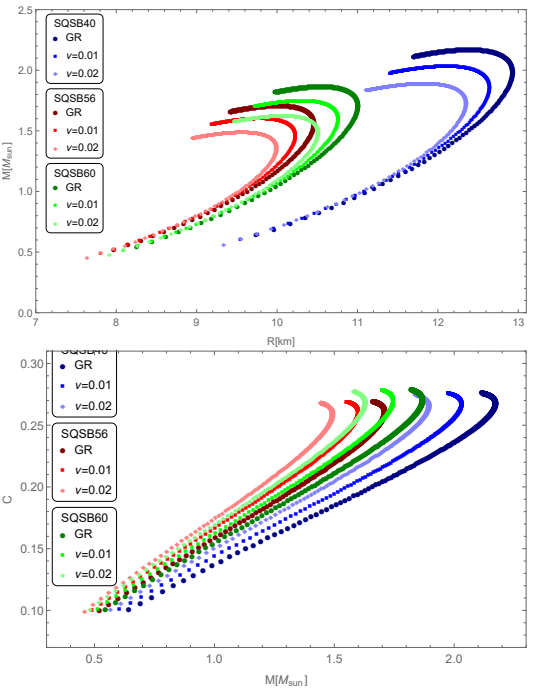


Fig. 1. Properties of strange quark stars with linear EoS. We have considered 3 MIT bag models, SQSB40 (in black), SQSB56 (in blue) and SQSB60 (in red) (see text) and $\nu = 0.01, 0.02$. The curves corresponding to GR ($\nu = 0$) are also shown for comparison reasons. **TOP:** M_R profiles (mass in solar masses and radius in km). **BOTTOM:** Compactness $C = G_N M/R$ vs mass (in solar masses).

cases, while the upper curves corresponding to the normalized energy density are clearly different for the two models. However, within each model the two curves corresponding to the two values of ν cannot be told apart. Regarding the models with a linear EoS we have obtained qualitatively very similar curves, which we do not show here. Clearly, all energy conditions are fulfilled throughout the star, and therefore we conclude that the solutions obtained here are realistic one, capable of describing realistic astrophysical configurations.

Regarding stability criteria, Bondi[94] suggested that for a stable Newtonian sphere the adiabatic index defined by

$$\Gamma \equiv c_s^2 \left[1 + \frac{\rho}{p}\right] \tag{39}$$

should be larger than 4/3. Moreover, the Harrison-Zeldovich-Novikov criterion requires that[95,96]

$$\frac{dM}{d\rho_c} > 0 \tag{40}$$

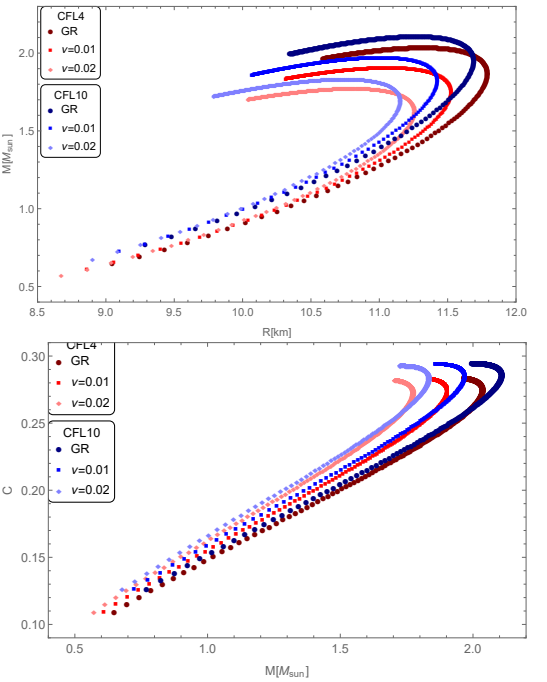


Fig. 2. Properties of strange quark stars with a non-linear EoS. We have considered 2 CFL models, CFL4 (in blue) and CFL10 (in black) (see text) and $\nu = 0.01, 0.02$. The curves corresponding to GR ($\nu = 0$) are also shown for comparison reasons. **TOP:** M_R profiles (mass in solar masses and radius in km). **BOTTOM:** Compactness $C = G_N M/R$ vs mass (in solar masses).

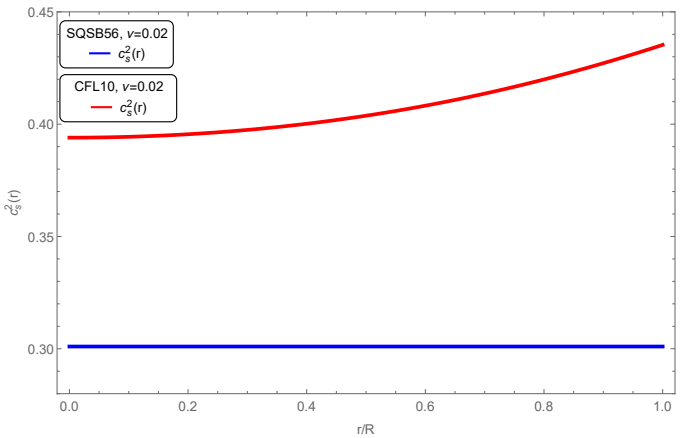


Fig. 3. Speed of sound, $c_s^2 = dp/d\rho$, vs normalized radial coordinate r/R for $\nu = 0.01, 0.02$ and for 2 color flavor locked states, CFL4 model (lower curves) and CFL10 model (upper curves) and for $\nu = 0.01, 0.02$.

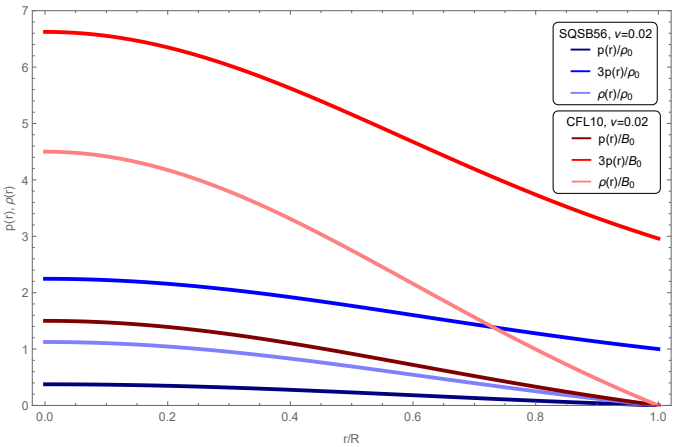


Fig. 4. Normalized pressure $\bar{p} \equiv p/B_0$, $3\bar{p}$ and energy density $\bar{\rho} \equiv \rho/B_0$ vs normalized radial coordinate r/R for $p_c = 1.5\ B_0$, $\nu = 0.01, 0.02$ and for 2 color flavor locked states, CFL4 model and CFL10 model. Shown are from bottom to top: i) Pressure (blue), ii) 3 times pressure (green), and iii) energy density (orange) for CFL 4 (upper) and for CFL10 (lower).

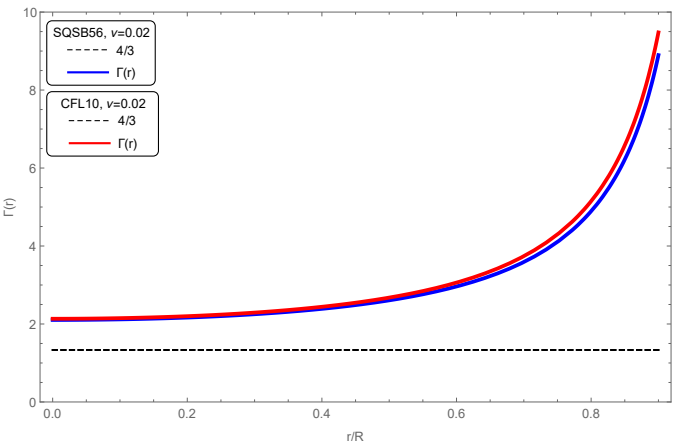


Fig. 5. Adiabatic index Γ vs normalized radial coordinate r/R for the model SQSB56 (linear EoS) and $\nu = 0.02$. The dashed horizontal line represents the Newtonian bound corresponding to $4/3$.

Fig. 5 shows that $\Gamma > 4/3$, as it should be. In Fig. 5 we have shown the adiabatic index only for one particular case, but we have checked that we have obtained almost identical figures for all the cases considered in the present work.

5. Conclusions

In summary, we have investigated in detail the properties of non-rotating, isotropic strange quark stars in Lorentz-violating theories of gravity. To be more precise, we have studied quark stars both in Hořava gravity and Einstein-æther theory, where a single dimensionless parameter ν measures the deviations from GR. As far as the EoS for the matter content is concerned, we adopted analytic functions, both linear and non-linear. The former corresponds to simple versions of the MIT bag model, while the latter corresponds to the CFL phase, where color superconductivity effects become important at very high densities. We numerically integrated the generalized structure equations, and we computed the factor of compactness, the radii and the masses of the objects upon matching the interior and the exterior solutions at the surface of the stars. We have presented graphically the mass-to-radius profiles for several different EoSs as well as for two distinct values of the parameter ν. The case $\nu = 0$ corresponding to GR is shown as well for comparison reasons. Our results show that the $M - R$ relationships are shifted downwards as ν increases, which implies a lower highest mass supported by a given EoS in comparison with the one obtained in GR for the same EoS. Finally, we have checked that causality is not violated, stability criteria are met, and energy conditions are fulfilled. Therefore we conclude that within the framework of the gravitational theories considered here, we have obtained well behaved solutions capable of describing realistic astrophysical configurations.

Acknowledgements

I wish to thank professor A. Wang for inviting me to MG-16, and the organizers for an excellent organization. I thank my colleagues I. Lopes and D. Vernieri for a fruitful collaboration. The work was financially supported by the Fundação para a Ciência e Tecnologia (FCT), Portugal, through the Project No. UIDB/FIS/00099/2020 and grant No. PTDC/FIS-AST/28920/2017.

References

1. A. Einstein, Annalen Phys. **49**, no. 7, 769 (1916).
2. S. G. Turyshev, Ann. Rev. Nucl. Part. Sci. **58** (2008) 207.
3. C. M. Will, Living Rev. Rel. **17** (2014) 4.
4. E. Asmodelle, arXiv:1705.04397 [gr-qc].
5. B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], Phys. Rev. Lett. **116** (2016) no. 6, 061102 [arXiv:1602.03837 [gr-qc]].
6. A. G. Riess et al. Astron. J. 116, 1009 (1998).
7. S. Perlmutter et al., Astrophys. J. 517, 565 (1999).
8. E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15** (2006) 1753 [hep-th/0603057].
9. T. P. Sotiriou and V. Faraoni, Rev. Mod. Phys. **82** (2010) 451 [arXiv:0805.1726 [gr-qc]].
10. A. De Felice and S. Tsujikawa, Living Rev. Rel. **13** (2010) 3 [arXiv:1002.4928 [gr-qc]].

11. W. Hu and I. Sawicki, Phys. Rev. D **76** (2007) 064004 [arXiv:0705.1158 [astro-ph]].
12. A. A. Starobinsky, JETP Lett. **86** (2007) 157.
13. D. Langlois, Prog. Theor. Phys. Suppl. **148** (2003) 181 [hep-th/0209261].
14. R. Maartens, Living Rev. Rel. **7** (2004) 7 [gr-qc/0312059].
15. G. R. Dvali, G. Gabadadze and M. Porrati, Phys. Lett. B **485** (2000) 208 [hep-th/0005016].
16. C. Brans and R. H. Dicke, Phys. Rev. **124** (1961) 925.
17. C. H. Brans, Phys. Rev. **125**, 2194 (1962).
18. J. C. B. Sanchez and L. Perivolaropoulos, Phys. Rev. D **81** (2010) 103505 [arXiv:1002.2042 [astro-ph.CO]].
19. P. Horava, Phys. Rev. D **79**, 084008 (2009) [arXiv:0901.3775 [hep-th]].
20. D. Blas, O. Pujolas and S. Sibiryakov, Phys. Rev. Lett. **104**, 181302 (2010) [arXiv:0909.3525 [hep-th]].
21. M. Visser, Phys. Rev. D **80**, 025011 (2009) [arXiv:0902.0590 [hep-th]].
22. T. P. Sotiriou, M. Visser and S. Weinfurtner, Phys. Rev. Lett. **102**, 251601 (2009) [arXiv:0904.4464 [hep-th]].
23. T. P. Sotiriou, M. Visser and S. Weinfurtner, JHEP **0910**, 033 (2009) [arXiv:0905.2798 [hep-th]].
24. D. Vernieri and T. P. Sotiriou, Phys. Rev. D **85**, 064003 (2012) [arXiv:1112.3385 [hep-th]].
25. D. Vernieri and T. P. Sotiriou, J. Phys. Conf. Ser. **453**, 012022 (2013) [arXiv:1212.4402 [hep-th]].
26. D. Vernieri, Phys. Rev. D **91**, no. 12, 124029 (2015) [arXiv:1502.06607 [hep-th]].
27. G. D'Odorico, F. Saueressig and M. Schutten, Phys. Rev. Lett. **113**, no. 17, 171101 (2014) [arXiv:1406.4366 [gr-qc]].
28. A. O. Barvinsky, D. Blas, M. Herrero-Valea, S. M. Sibiryakov and C. F. Steinwachs, Phys. Rev. D **93**, no. 6, 064022 (2016) [arXiv:1512.02250 [hep-th]].
29. A. O. Barvinsky, D. Blas, M. Herrero-Valea, S. M. Sibiryakov and C. F. Steinwachs, JHEP **1807**, 035 (2018) [arXiv:1705.03480 [hep-th]].
30. A. O. Barvinsky, M. Herrero-Valea and S. M. Sibiryakov, Phys. Rev. D **100**, no. 2, 026012 (2019) [arXiv:1905.03798 [hep-th]].
31. B. Audren, D. Blas, M. M. Ivanov, J. Lesgourgues and S. Sibiryakov, JCAP **1503**, no. 03, 016 (2015) [arXiv:1410.6514 [astro-ph.CO]].
32. N. Frusciante, M. Raveri, D. Vernieri, B. Hu and A. Silvestri, Phys. Dark Univ. **13**, 7 (2016) [arXiv:1508.01787 [astro-ph.CO]].
33. D. Blas and S. Sibiryakov, Phys. Rev. D **84**, 124043 (2011) [arXiv:1110.2195 [hep-th]].
34. E. Barausse, T. Jacobson and T. P. Sotiriou, Phys. Rev. D **83**, 124043 (2011) [arXiv:1104.2889 [gr-qc]].
35. P. Berglund, J. Bhattacharyya and D. Mattingly, Phys. Rev. D **85**, 124019 (2012) [arXiv:1202.4497 [hep-th]].
36. E. Barausse and T. P. Sotiriou, Phys. Rev. D **87**, 087504 (2013) [arXiv:1212.1334 [gr-qc]].
37. A. Wang, Phys. Rev. Lett. **110**, no. 9, 091101 (2013) [arXiv:1212.1876 [hep-th]].
38. E. Barausse and T. P. Sotiriou, Class. Quant. Grav. **30**, 244010 (2013) [arXiv:1307.3359 [gr-qc]].
39. T. P. Sotiriou, I. Vega and D. Vernieri, Phys. Rev. D **90**, no. 4, 044046 (2014) [arXiv:1405.3715 [gr-qc]].
40. K. Yagi, D. Blas, N. Yunes and E. Barausse, Phys. Rev. Lett. **112**, no. 16, 161101 (2014) [arXiv:1307.6219 [gr-qc]].

41. K. Yagi, D. Blas, E. Barausse and N. Yunes, Phys. Rev. D **89**, no. 8, 084067 (2014) Erratum: [Phys. Rev. D **90**, no. 6, 069902 (2014)] Erratum: [Phys. Rev. D **90**, no. 6, 069901 (2014)] [arXiv:1311.7144 [gr-qc]].
42. O. Ramos and E. Barausse, Phys. Rev. D **99**, no. 2, 024034 (2019) [arXiv:1811.07786 [gr-qc]].
43. D. Vernieri and S. Carloni, EPL **121**, no. 3, 30002 (2018) [arXiv:1706.06608 [gr-qc]].
44. D. Vernieri, Phys. Rev. D **98**, no. 2, 024051 (2018) [arXiv:1808.00974 [gr-qc]].
45. D. Vernieri, Phys. Rev. D **100**, no. 10, 104021 (2019) [arXiv:1906.07738 [gr-qc]].
46. B. P. Abbott *et al.* [LIGO Scientific and Virgo and Fermi-GBM and INTEGRAL Collaborations], Astrophys. J. **848**, no. 2, L13 (2017) [arXiv:1710.05834 [astro-ph.HE]].
47. A. Emir Gümrükçüoğlu, M. Saravani and T. P. Sotiriou, Phys. Rev. D **97**, no. 2, 024032 (2018) [arXiv:1711.08845 [gr-qc]].
48. T. Jacobson and D. Mattingly, Phys. Rev. D **64**, 024028 (2001) [gr-qc/0007031].
49. T. Jacobson, Phys. Rev. D **81**, 101502 (2010) Erratum: [Phys. Rev. D **82**, 129901 (2010)] [arXiv:1001.4823 [hep-th]].
50. D. Blas, O. Pujolas and S. Sibiryakov, JHEP **1104**, 018 (2011) [arXiv:1007.3503 [hep-th]].
51. S. L. Shapiro and S. A. Teukolsky, *Black holes, white dwarfs, and neutron stars: The physics of compact objects*, New York, USA: Wiley, 645 p. (1983).
52. D. Psaltis, Living Rev. Rel. **11**, 9 (2008) [arXiv:0806.1531 [astro-ph]].
53. D. R. Lorimer, Living Rev. Rel. **11**, 8 (2008) [arXiv:0811.0762 [astro-ph]].
54. C. Alcock, E. Farhi and A. Olinto, Astrophys. J. **310**, 261 (1986).
55. C. Alcock and A. Olinto, Ann. Rev. Nucl. Part. Sci. **38**, 161 (1988).
56. J. Madsen, Lect. Notes Phys. **516**, 162 (1999) [astro-ph/9809032].
57. F. Weber, Prog. Part. Nucl. Phys. **54**, 193 (2005) [astro-ph/0407155].
58. Y. L. Yue, X. H. Cui and R. X. Xu, Astrophys. J. **649**, L95 (2006) [astro-ph/0603468].
59. D. Leahy and R. Ouyed, Mon. Not. Roy. Astron. Soc. **387**, 1193 (2008) [arXiv:0708.1787 [astro-ph]].
60. E. Witten, Phys. Rev. D **30**, 272 (1984).
61. E. Farhi and R. L. Jaffe, Phys. Rev. D **30**, 2379 (1984).
62. E. O. Ofek *et al.*, Astrophys. J. **659**, L13 (2007) [astro-ph/0612408].
63. R. Ouyed, D. Leahy and P. Jaikumar, arXiv:0911.5424 [astro-ph.HE].
64. A. H. Guth, Phys. Rev. D **23**, 347 (1981) [Adv. Ser. Astrophys. Cosmol. **3**, 139 (1987)].
65. S. M. Carroll and E. A. Lim, Phys. Rev. D **70**, 123525 (2004) [hep-th/0407149].
66. R. C. Tolman, Phys. Rev. **55**, 364 (1939).
67. J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374 (1939).
68. S. Carloni and D. Vernieri, Phys. Rev. D **97**, no. 12, 124056 (2018) [arXiv:1709.02818 [gr-qc]].
69. S. Carloni and D. Vernieri, Phys. Rev. D **97**, no. 12, 124057 (2018) [arXiv:1709.03996 [gr-qc]].
70. C. Eling and T. Jacobson, Class. Quant. Grav. **23**, 5625 (2006) Erratum: [Class. Quant. Grav. **27**, 049801 (2010)] [gr-qc/0603058].
71. K. Schwarzschild, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1916**, 189 (1916) [physics/9905030].
72. A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn and V. F. Weisskopf, Phys. Rev. D **9**, 3471 (1974).
73. A. Chodos, R. L. Jaffe, K. Johnson and C. B. Thorn, Phys. Rev. D **10**, 2599 (1974).
74. D. Gondek-Rosinska and F. Limousin, arXiv:0801.4829 [gr-qc].
75. M. G. Alford, K. Rajagopal and F. Wilczek, Phys. Lett. B **422**, 247 (1998) [hep-ph/9711395].

76. M. G. Alford, K. Rajagopal and F. Wilczek, Nucl. Phys. A **638**, 515C (1998) [hep-ph/9802284].
77. M. G. Alford, K. Rajagopal and F. Wilczek, Nucl. Phys. B **537**, 443 (1999) [hep-ph/9804403].
78. K. Rajagopal and F. Wilczek, Phys. Rev. Lett. **86**, 3492 (2001) [hep-ph/0012039].
79. G. Lugones and J. E. Horvath, Phys. Rev. D **66**, 074017 (2002) [hep-ph/0211070].
80. C. Vasquez Flores and G. Lugones, Phys. Rev. D **82**, 063006 (2010) [arXiv:1008.4882 [astro-ph.HE]].
81. C. V. Flores and G. Lugones, Phys. Rev. C **95**, no. 2, 025808 (2017) [arXiv:1702.02081 [astro-ph.HE]].
82. P. Demorest, T. Pennucci, S. Ransom, M. Roberts and J. Hessels, Nature **467**, 1081 (2010) [arXiv:1010.5788 [astro-ph.HE]].
83. E. Fonseca *et al.*, Astrophys. J. **832**, no. 2, 167 (2016) [arXiv:1603.00545 [astro-ph.HE]].
84. J. Antoniadis *et al.*, Science **340**, 6131 (2013) [arXiv:1304.6875 [astro-ph.HE]].
85. H. T. Cromartie *et al.*, Nature Astronomy Letters (2019) [arXiv:1904.06759 [astro-ph.HE]].
86. M. Linares, T. Shahbaz and J. Casares, Astrophys. J. **859**, no. 1, 54 (2018) doi:10.3847/1538-4357/aabde6 [arXiv:1805.08799 [astro-ph.HE]].
87. G. Panotopoulos and Á. Rincón, Eur. Phys. J. C **79**, no. 6, 524 (2019).
88. I. Lopes, G. Panotopoulos and Á. Rincón, Eur. Phys. J. Plus **134**, no. 9, 454 (2019) [arXiv:1907.03549 [gr-qc]].
89. M. K. Mak and T. Harko, Chin. J. Astron. Astrophys. **2**, 248 (2002).
90. D. Deb, S. R. Chowdhury, S. Ray, F. Rahaman and B. K. Guha, Annals Phys. **387**, 239 (2017) [arXiv:1606.00713 [gr-qc]].
91. D. Deb, S. Roy Chowdhury, S. Ray and F. Rahaman, Gen. Rel. Grav. **50**, no. 9, 112 (2018) [arXiv:1509.00401 [gr-qc]].
92. P. Bhar, M. Govender and R. Sharma, Eur. Phys. J. C **77**, no. 2, 109 (2017) [arXiv:1607.06664 [gr-qc]].
93. G. Panotopoulos and Á. Rincón, Eur. Phys. J. Plus **134**, no. 9, 472 (2019) [arXiv:1907.03545 [gr-qc]].
94. H. Bondi, Proc. R. Soc. Lond. A **281** (1964) 39.
95. B. K. Harrison, *Gravitational Theory and Gravitational Collapse*, University of Chicago Press, Chicago, 1965.
96. Y. B. Zeldovich, and I. D. Novikov, *Relativistic Astrophysics, Vol. I: Stars and Relativity*, University of Chicago Press, Chicago, 1971.

Shadows of Kerr-like black holes in 4*D* Einstein–Gauss–Bonnet gravity and constraints from EHT observations

Sushant G. Ghosh

Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi 110025, India
Astrophysics Research Centre, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag 54001, Durban 4000, South Africa
E-mail: sghosh2@jmi.ac.in

Rahul Kumar Walia

Astrophysics Research Centre, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag 54001, Durban 4000, South Africa
E-mail: rahul.phy3@gmail.com

The M87* black hole shadow observation by the Event Horizon Telescope (EHT) has enabled us to test the modified gravity theories in the extreme-field regime and estimating the black hole parameters. Having this assertion, we investigate the Kerr-like rotating black holes in 4*D* Einstein-Gauss-Bonnet (EGB) gravity and deduce their shadows. Considering the inclination angle $\theta_0 = 17^o$, we show that the EGB black hole shadows are smaller and more distorted than for the Kerr black holes. Modelling the M87* black hole as the EGB black hole, we predict the shadow angular size $35.7888\mu as \leq \theta_d \leq 39.6192\mu as$. The M87* black hole shadow angular size $\theta_d = 42\pm3\mu as$, within the 1σ region, constrains the GB coupling parameter and the black hole spin parameter. Interestingly, the circularity deviation of the EGB black hole shadows is smaller than the bounded deduced for the M87* black hole.

Keywords: EGB gravity, Shadows, Astrophysical black holes, Parameter Estimation, EHT, M87* black hole shadow

1. Introduction

The uniqueness of the Einstein tensor to describe gravity in the four-dimensional ($4D$) spacetime is dictated by the Lovelock theorem.[1] However, if one or more conditions in the Lovelock theorem are relaxed, then modifications to the Einstein-Hilbert action exist that lead to covariant, conserved, and second-order field equations and propagate only gravitational degrees of freedom and thus are free from the ghost instabilities. One such Lagrangian-based theory of gravity is Einstein-Gauss-Bonnet (EGB) gravity that exists in the $D \geq 5$ and is motivated by the heterotic string theory.[2,3] EGB gravity supplements the Einstein-Hilbert action with quadratic corrections terms constructed from the curvature tensors invariants and reads as follows

$$\mathcal{I}_{\rm EGB} = \frac{1}{16\pi G_D}\int d^Dx\,\sqrt{-g}(\mathcal{L}_{\rm EH} + \alpha\,\mathcal{L}_{\rm GB}), \tag{1}$$

with

$$\mathcal{L}_{\rm EH} = R, \quad \mathcal{L}_{\rm GB} = R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} - 4R^{\mu\nu}R_{\mu\nu} + R^2. \tag{2}$$

Here, α is identified as the GB coupling constant and is related to the inverse string tension, making it positive-definite. The GB Lagrangian is a unique quadratic combination of the Riemann tensor that naturally emerges as a leading-order correction term in low-energy effective actions of heterotic string theory and $10D$ gauged supergravity. Boulware and Deser, in their seminal paper,[4] obtained the first spherically symmetric and static black hole solution for the EGB theory, afterward several intriguing black hole solutions are obtained.[5–7] It is worth mentioning that for $D < 5$ GB Lagrangian $\mathcal{L}_{\rm GB}$ turns into a total derivative, and thereby its contribution to the gravitational dynamics vanishes, rendering the theory indistinguishable from general relativity. However, in the presence of an additional non-minimally coupled scalar field dilaton with the canonical kinetic term, $\mathcal{L}_{\rm GB}$ leads to the non-trivial gravitational dynamics[8–11] and the resulting theory is Horndeski or Galilean.

Ever since the formulation of the EGB gravity theory, its $4D$ regularization has been a topic of great interest. In this line of research, Tomozawa[12] showed for the first time that the quantum corrections to gravity in a conformally flat metric in $4D$ appears as GB quadratic curvature forms, and the $4D$ black hole solution shows repulsive nature at $r \to 0$. In another attempt of regularization procedure, Cognolo *et al.*[13] used an "entropic" dimensional reduction of EGB gravity to $D \to 4$ within the classical Lagrangian formulation. Lately, the interest in the $4D$ EGB gravity theory is re-surged due to the regularization approach proposed by Glavan and Lin;[14] the GB coupling is re-scaled as $\alpha \to \alpha/(D-4)$ and the $4D$ EGB theory was obtained as the limit $D \to 4$ at the level of field equations. The aim for introducing this re-scaling is to generate a divergence that exactly cancels the vanishing contribution that the GB term makes to the field equations in $4D$. The extension to higher-order Lovelock gravity is presented in Refs.[15,16] Likewise, EGB theory is obtained in lower dimensions.[17,18] Interestingly, the Glavan and Lin's static and spherically symmetric black hole solution[14] matched with that obtained using the quantum correction by Tomozawa,[12] and Cognolo *et al.*[13]

However, Glavan and Lin's claim[14] that the resulting theory is of pure graviton was later proven to be spurious on several grounds. The covariant approach proposed in Ref.[14] is largely speculated to be valid only for specific higher-dimensional spacetimes with high degrees of symmetries, particularly maximally symmetric or spherically symmetric spacetimes. Recently, some studies have called into question the Glavan and Lin[14] regularization procedure for the less-symmetric spacetimes and also reported several other inconsistencies Refs.[17,19–22] Following that, the GB contribution arising in higher dimensions could be renormalized in such a way as to yield a non-trivial contribution also in $4D$, even without re-scaling the GB coupling.[16–18,23–26] Hennigar *et al.*[17] proposed another well defined $D \to 4$ limit of EGB gravity generalizing the previous work of Mann and Ross[27] in establishing the $D \to 2$ limit of general relativity and this regularization is applicable not only in $4D$ but also to $D < 4$. These alternate regularization procedures of EGB theory, leading to a divergence-free $4D$ action, describe the scalar-tensor theory of

gravity of the Horndeski type. These scalar-tensor models propagate the supplementary scalar mode in addition to the gravitational degree of freedom. Thus these alternate regularized theories are in line with Lovelock's theorem, as they introduce another dynamical field. Nevertheless, the spherically symmetric $4D$ black hole solution obtained in Ref.[14] remains valid for these regularised scalar-tensor theories.[16–18,23,28] This means that $4D$ EGB gravity can be viewed as both a dimensionally reduced theory and as a gravitational theory that displays known quantum corrections. As a result, both the Glavan and Lin theory[14] and scalar-tensor regularizations have received remarkable attention and more then 100 papers have been reported on $4D$ EGB gravity and its various solutions including their charged extension,[29,30] rotating counterparts,[31,32] Vaidya-like radiating black holes,[33,34] regular black holes.[35,36] The gravitational lensing of $4D$ EGB black holes have also been studied.[37–40]

The black hole shadow observations by the Event Horizon Telescope (EHT) Collaboration have unprecedentedly opened up an exciting arena to make a precision test of the gravitational theory in the strong and relativistic field regimes (in the vicinity of the unstable bound orbits around black holes).[41,42] The EHT analysis suggested that, based on *a priori* known estimates for the mass and distance from stellar dynamics, the M87* shadow size is consistent within 17% for a 68% confidence interval of the size predicted from the Kerr black hole general-relativistic-magneto-hydrodynamics (GRMHD) image.[43] However, several other studies altogether have not entirely precluded the possibility of non-Kerr black holes.[44–46] Using the M87* shadow angular size, constraints are placed on the second post-Newtonian metric coefficients, which were inaccessible in the earlier weak-field tests at the Solar-scale.[43] Therefore, it is both legitimate and timely to test the viability of the $4D$ EGB gravity theory using the M87* black hole shadow observations. This paper aims to present the detailed study of the rotating $4D$ EGB black hole shadow, parameter estimation of the black hole using the shadow observables, and constraining them using the M87* black hole shadow observed by the Event Horizon Telescope (EHT).

2. Rotating $4D$ EGB black hole shadows

Finding an exact analytic and rotating axially symmetric black hole solution of the EGB gravity is a notorious task due to the non-linearity involved in the field equations. However, there exists the rotating solution generating mechanisms such as the Newman-Janis algorithm[47] and the gravitational-decoupling method,[48] which have been widely used to construct rotating black hole solutions from their non-rotating counterparts. The Azreg-Aïnou's non-complexification procedure[49,50] for the modified Newman-Janis algorithm generates a unique imperfect fluid rotating solution from the seed spherically symmetric static solution. It has been applied to generate rotating solutions in several modified gravity theories.[32,51–56] The rotating

$4D$ EGB black hole metric, in Boyer-Lindquist coordinates, reads[31,32]

$$ds^2 = -\frac{\Delta}{\Sigma}\left(dt - a\sin^2\theta\, d\phi\right)^2 + \frac{\Sigma}{\Delta}dr^2 + \Sigma\, d\theta^2 + \frac{\sin^2\theta}{\Sigma}\left((r^2+a^2)\, d\phi - a\, dt\right)^2 \tag{3}$$

with

$$\Delta = r^2 + a^2 + \frac{r^4}{32\pi\alpha}\left[1 - \sqrt{1 + \frac{128\pi\alpha M}{r^3}}\right], \quad \Sigma = r^2 + a^2\cos^2\theta. \tag{4}$$

Thus rotating $4D$ EGB black holes are characterized by three parameters, mass (M), spin (a), and GB coupling parameter α, which also gives potential deviation from the Kerr solution. In the limit $\alpha \to 0$ or large r, the metric Eq. (3) smoothly recovers the Kerr black hole.[57] Because the static black hole solution of Ref.[14] is identical to that of regularized scalar-tensor theories and other quantum-corrected theories of gravity,[12–17,58,59] the rotating black hole metric (3) also corresponds to these theories. The rotating black hole admits up to two distinct horizons whose radii $r_- \leq r_+$ can be identified as real positive roots of the $\Delta = 0$. The variation of both horizon radii with GB coupling is shown in Fig. 1, it is evident that the event horizon radius decreases and Cauchy horizon radius increase with α. For a given value of spin a, there exists a extremal value of GB coupling $\alpha = \alpha_E$ for which degenerate horizons $r_- = r_+$ exists, such that for $\alpha > \alpha_E$ horizons disappear and the central singularity becomes globally naked. Similarly, for a given value of α, one can find the extremal value of spin $a = a_E$ which leads to degenerate horizons $r_- = r_+$. In this paper, we will only consider the black hole case viz., $\alpha \leq \alpha_E$.

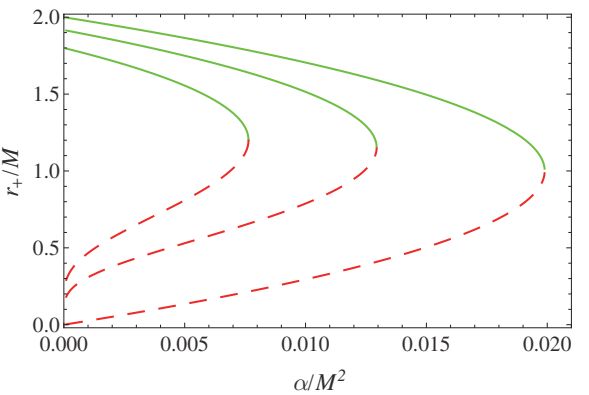


Fig. 1. Event horizon (solid green) and Cauchy horizon (dashed red) radii variation with α for different values of $a = 0, 0.40M, 0.60M$ (from outside to inside).

The optical appearance of the black hole in the presence of a bright background or the illuminated matter accretion flow is known as the shadow. The light from

the source gets strongly lensed by the black hole in the vicinity of the horizon and receive by a faraway observer on the opposite side of the source. Synge,[60] and Luminet[61] led the study of black hole shadow and calculated the capturing angle of the Schwarzschild black hole. For the first time, the shadow of the Kerr black hole was reported by Bardeen[62] in his pioneering work in 1973. The rotating EGB metric (3) belongs to Petrov type-D spacetimes, and thus the geodesics equations are completely integrable. The metric (3) carries two Killing vectors ∂_t and ∂_ϕ, associated with the time translational and rotational invariance of the spacetime geometry. The components of photon four-momentum p^μ projected along these Killing vectors are constant of motion, which in this case can be identified as the energy E and axial angular momentum magnitude L_z. We follow the Hamilton-Jacobi formalism to determine the null geodesics equations of motion around the rotating black hole, which read as follows[32,63]

$$\Sigma\frac{dt}{d\tau} = \frac{r^2+a^2}{\Delta}\left(E(r^2+a^2) - aL_z\right) - a(aE\sin^2\theta - L_z), \tag{5}$$

$$\Sigma\frac{dr}{d\tau} = \pm\sqrt{\mathcal{R}(r)}, \tag{6}$$

$$\Sigma\frac{d\theta}{d\tau} = \pm\sqrt{\Theta(\theta)}, \tag{7}$$

$$\Sigma\frac{d\phi}{d\tau} = \frac{a}{\Delta}\left(E(r^2+a^2) - aL_z\right) - \left(aE - \frac{L_z}{\sin^2\theta}\right), \tag{8}$$

where τ is the affine parameter along the null geodesics and

$$\mathcal{R}(r) = \left((r^2+a^2)E - aL_z\right)^2 - \Delta((aE-L_z)^2 + K), \tag{9}$$

$$\Theta(\theta) = K - \left(\frac{{L_z}^2}{\sin^2\theta} - a^2E^2\right)\cos^2\theta. \tag{10}$$

The constant K is the separability constant related to the Carter constant $\mathcal{Q}$ through $\mathcal{Q} = K + (aE - L_z)^2$. Carter constant appears as a conserved quantity associated with the hidden symmetry described by the second-rank Killing tensor. We introduce the impact parameters for the photons geodesics, which are constant along geodesics and defined in dimensionless form as follows[63]

$$\xi = \frac{L_z}{E} \ , \ \eta = \frac{K}{E^2}. \tag{11}$$

Photons may get scattered, captured, or follow bound orbits around the black hole depending on the values of (xi, eta). Because of the black hole rotation, photons can either co-rotate or counter-rotate along with the black hole, whose radii vary differently with black hole spin. At $\theta = \pi/2$, Carter's constant vanishes, and the photons follow the circular orbits with radii $r_p^\pm$, which can be determined by solving $Y = 0$. Whereas for $\theta \neq \pi/2$ the Carter constant is positive definite and the photons follow the non-planar orbits with radii $r_p^- < r_p < r_p^+$. The photons following the spherical orbits of constant coordinate radii r_p around the black hole are characterized by $\dot{r} = 0$ and $\ddot{r} = 0$. This results into the critical values of impact parameters

$(\xi_{crit}, \eta_{crit})$ for the unstable orbits[32]

$$\xi_{crit} = \frac{\left(a^2 + r^2\right)\Delta'(r) - 4r\Delta(r)}{a\Delta'(r)},$$
$$\eta_{crit} = \frac{r^2\left(8\Delta(r)\left(2a^2 + r\Delta'(r)\right) - r^2\Delta'(r)^2 - 16\Delta(r)^2\right)}{a^2\Delta'(r)^2}, \tag{12}$$

where $'$ stands for the derivative with the radial coordinate r. Furthermore, these spherical photons orbits are the non-planar orbits that periodically cross the equatorial plane and construct a photon region around the black hole. As a result, the photons following the spherical orbits, beside having a motion along the ϕ direction also move along the θ-direction. For visualizing the black hole shadow, we consider a distant observer at position (r_0, θ_0). The coordinates (X, Y) define the observer image plane, such that the stereographic projection of the shadow from the celestial sphere to the image plane is defined as follow

$$X = \lim_{r_0 \to \infty}\left(-r_0^2 \sin\theta_0 \frac{d\phi}{dr}\right),$$
$$Y = \lim_{r_0 \to \infty}\left(r_0^2 \frac{d\theta}{dr}\right). \tag{13}$$

For an asymptotically far observer, Eq. (13) leads to

$$X = -\xi_{crit}\csc\theta_0,$$
$$Y = \pm\sqrt{\eta_{crit} + a^2\cos^2\theta_0 - \xi_{crit}^2\cot^2\theta_0}\,. \tag{14}$$

and satisfy

$$X^2 + Y^2 = \eta_{crit} + \xi_{crit}^2 + a^2\cos^2\theta_0. \tag{15}$$

The parametric curve Y vs X delineates the shadow's boundary of the rotating EGB black hole. It is clear that the non-rotating black hole ($a = 0$) cast a perfectly circular shadow silhouette. For $\theta_0 \neq 0$ or π the rotating black hole shadow shifts in the direction perpendicular to the black hole rotation and appears distorted, whereas for $\theta_0 = 0, \pi$ shadows is centered at $(0, 0)$ and is perfectly circular for all values of a. The maximum off-center displacement of shadow appears for $\theta_0 = \pi/2$.

In April 2019, the EHT collaboration using the VLBI technology unveiled the first-ever horizon-scaled image of the supermassive black hole M87*.[41,42] The M87* image shows powerful relativistic jets, which could be emerged from magnetohydrodynamic interactions between the accretion disk and the rotating black hole. Considering the orientation of these jets in M87*, the inclination angle (angle between the rotational axis and the line of sight) is estimated to be 17^o.[64] Hereafter, for our analysis of EGB black hole shadows, we will consider the inclination angle $\theta_0 = 17^o$. The rotating EGB black holes shadows with varying a and α are depicted in Fig. 2. It is evident that the shadow size decreases with increasing α, such that

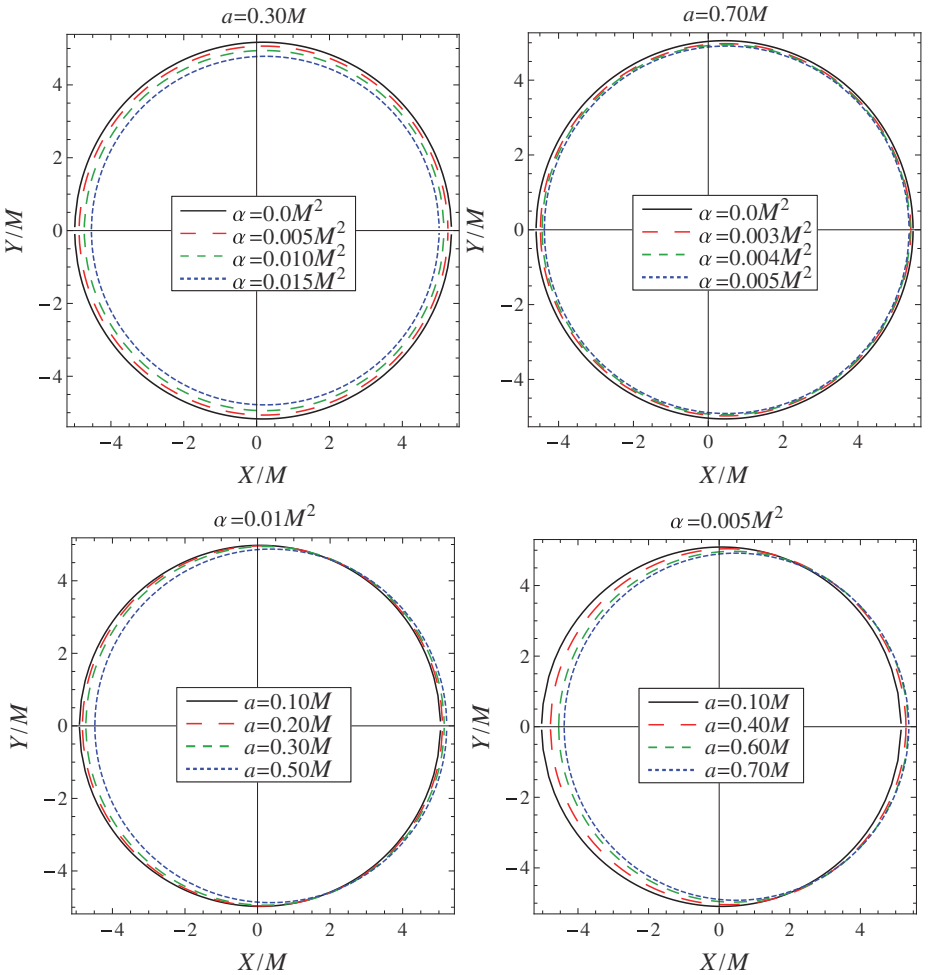


Fig. 2. Plot showing the rotating EGB black holes shadows with varying parameters a and α. Solid black curves in the upper panel are for the Kerr black holes.

the rotating EGB black hole shadows are smaller than the Kerr black hole shadows. Furthermore, the rotating black holes shadows are not perfectly circular. To characterize the shadow size and the deviation from the circularity, we introduce the shadow observables, namely, shadow area A and oblateness D as follows[65,66]

$$A = 2\int Y dX = 2\int_{r_p^-}^{r_p^+} \left(Y\frac{dX}{dr_p}\right) dr_p, \tag{16}$$

$$D = \frac{X_r - X_l}{Y_t - Y_b}, \tag{17}$$

where the points on the right, left, top, and bottom of the shadow boundary are designated by subscripts r, l, t and b. The shadow boundary, defined by (X, Y), is a function of spherical photon orbits radii r_p. Interestingly, photons with different orbit radii construct the different parts of the shadow boundary. Therefore, as one moves along the Y-axis in the shadow image, one sees deeper or far away from the black hole. The behavior of the shadow observables A and D with varying a and α is shown in Fig. 3. The shadow area monotonically decreases with α and a, and the oblateness increases with increasing a. To estimate the rotating EGB black hole parameters, we make a contour plot of observables A and D as functions of a and α in Fig. 4. Therein each solid red curve corresponds to constant values of A and dashed blue curve to D. The intersection point of observables A and D determines the unique and precise values of the black hole parameters a and α. Hence, from Fig. 4, it is clear that for a given set of $4D$ EGB black hole shadow observables, A and D, we can determine information about black hole spin and GB coupling parameter.

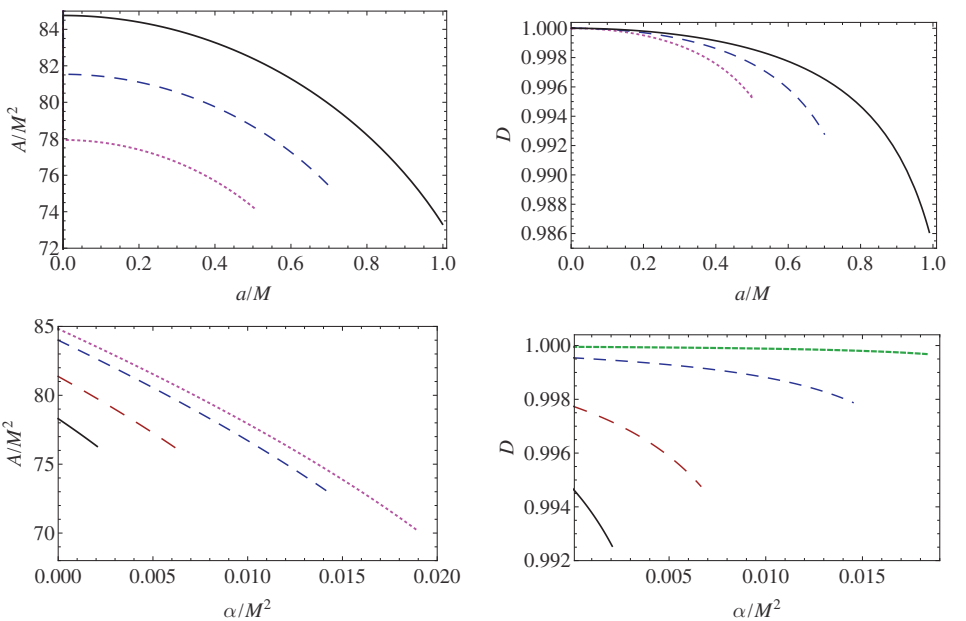


Fig. 3. *Upper*: The observables A and D vs a for $\alpha = 0.0M^2$ (solid black curve), for $\alpha = 0.005M^2$ (dashed blue curve), and for $\alpha = 0.01M^2$ (dotted magenta curve). *Bottom*: The observables A and D vs α for $a = 0.0M$ (dotted magenta curve), for $a = 0.1M$ (dotted green curve), for $a = 0.3M$ (dashed blue curve), for $a = 0.6M$ (long-dashed brown curve), and for $a = 0.8M$ (solid black curve).

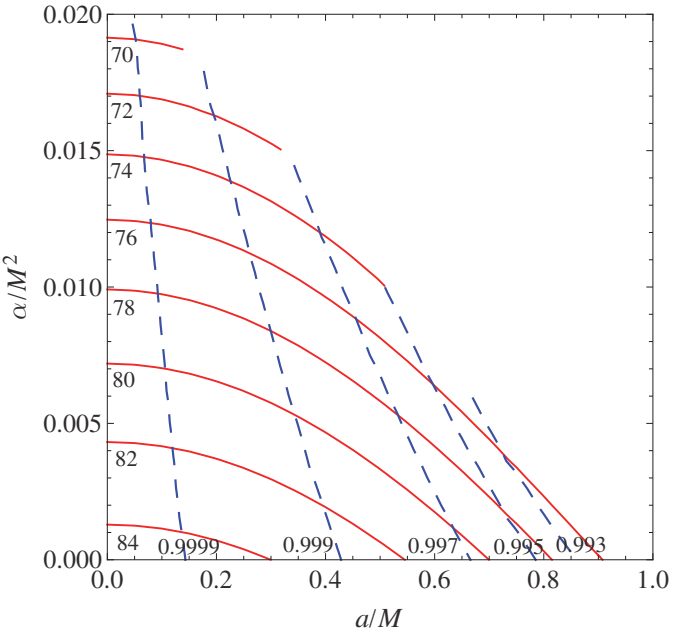


Fig. 4. Contour plots of the observables A and D in the plane (a, α) for the rotating EGB black holes. Each curve is labeled with the corresponding values of A and D. Solid red curves correspond to the A, and dashed blue curves are for the oblateness parameter D.

3. Constraints from the M87* shadow

Although the M87* black hole shadow is found to be consistent with the Kerr black hole image as predicted in the general relativity, the non-Kerr black holes are also not ruled out. Very recently, the EHT collaboration team has set stringent constraints on the physical charges of a large variety of modified gravity black holes by using the M87* shadow.[67] The M87* shadow is of crescent shape with the circularity deviation $\Delta C \leq 0.10$ (10%) in terms of root-mean-square deviation from average shadow radius, axis ratio $\lesssim 4/3$, whereas the angular diameter θ_d is $42 \pm 3\mu as$.[41,42] Here, we will model the M87* black hole as the rotating $4D$ EGB black hole and use the M87* shadow observables to place constraints on the black hole parameters.

For this, we define the shadow boundary with polar coordinates $(R(\varphi), \varphi)$ such that the origin is at the shadow center (X_O, Y_O). Figure (2) infers that the rotating black hole shadow is always $\mathcal{Z}_2$ symmetric around $Y = 0$. However, due to black hole rotation, the shadow center shifts from $X = 0$, and as a result, the shadow is asymmetric along the Y axis. It ascertains that the shadow center is $(X_O = |X_r + X_l|/2, Y_O = 0)$, where X_r and X_l are the maximum and minimum abscissas of

the shadow boundary in the image plane. The radial coordinate of shadow boundary from its center reads as

$$R(\varphi) = \sqrt{(X - X_O)^2 + (Y - Y_O)^2}, \quad \varphi \equiv \tan^{-1}\left(\frac{Y}{X - X_O}\right),$$

whereas shadow average radius $\bar{R}$ is defined as[68]

$$\bar{R}^2 = \frac{1}{2\pi}\int_0^{2\pi} R^2(\varphi) d\varphi. \tag{18}$$

We describe the circularity deviation ΔC to quantifies the shadow deviation from a perfect circle as a measure of the root-mean-square deviation of $R(\varphi)$ from the shadow average radius[68–70]

$$\Delta C = \frac{1}{\bar{R}}\sqrt{\frac{1}{2\pi}\int_0^{2\pi}\left(R(\varphi) - \bar{R}\right)^2 d\varphi}, \tag{19}$$

clearly, for a circular shadow of spherically symmetric black hole $\Delta C = 0$. Second observable is the shadow angular diameter θ_d, which for a far distant observer, at a distance r_0 from the black hole, is defined as

$$\theta_d = 2\frac{R_s}{r_0}, \quad R_s = \sqrt{A/\pi}, \tag{20}$$

and the third observable is the axis ratio

$$D_x = \frac{\Delta Y}{\Delta X}, \tag{21}$$

which is just the inverse of oblateness observable $D_x = 1/D$. We have calculated these three observable $\Delta C, \theta_d$, and D_x for the rotating $4D$ EGB black hole with $M = 6.5 \times 10^9 M_\odot$ and $r_0 = 16.8$ Mpc and plotted them in Fig. 5. The EHT bound for the M87* black hole shadow angular diameter $\theta_d = 39\,\mu$as within the 1σ region, shown as the black solid line, constrained the a and α. The shadow angular size for the non-rotating extremal EGB black hole with $a = 0, \alpha = 0.019894367M^2$ is $\theta_d = 35.7888\mu$as and for extremal Kerr black hole with $a = M, \alpha = 0$ is $\theta_d = 36.8632\mu$as. For comparison, at the inclination angle $\theta_o = 90^o$, the extremal Kerr black hole shadow angular size is 37.3534μas. The relative difference in shadow angular diameter $\delta\theta_d \equiv (\theta_d|_{Kerr} - \theta_d|_{EGB})/\theta_d|_{Kerr}$ is shown in Fig. 6. Clearly, $\delta\theta_d \leq 17\%$ and thus it is consistent with the Psaltis *et al.*[43] findings. Furthermore, the axis ratio and the circularity deviation for the M87* black hole shadow allow all parameter space of the EGB black hole.

4. Conclusions

The underlying theory of gravity in the extreme-field regime is currently unknown, and insights into it are likely to be gained through observations. In this paper, we have investigated the rotating $4D$ EGB black hole. The EGB gravity theory has been of great interest and importance due to second-order field equations and being

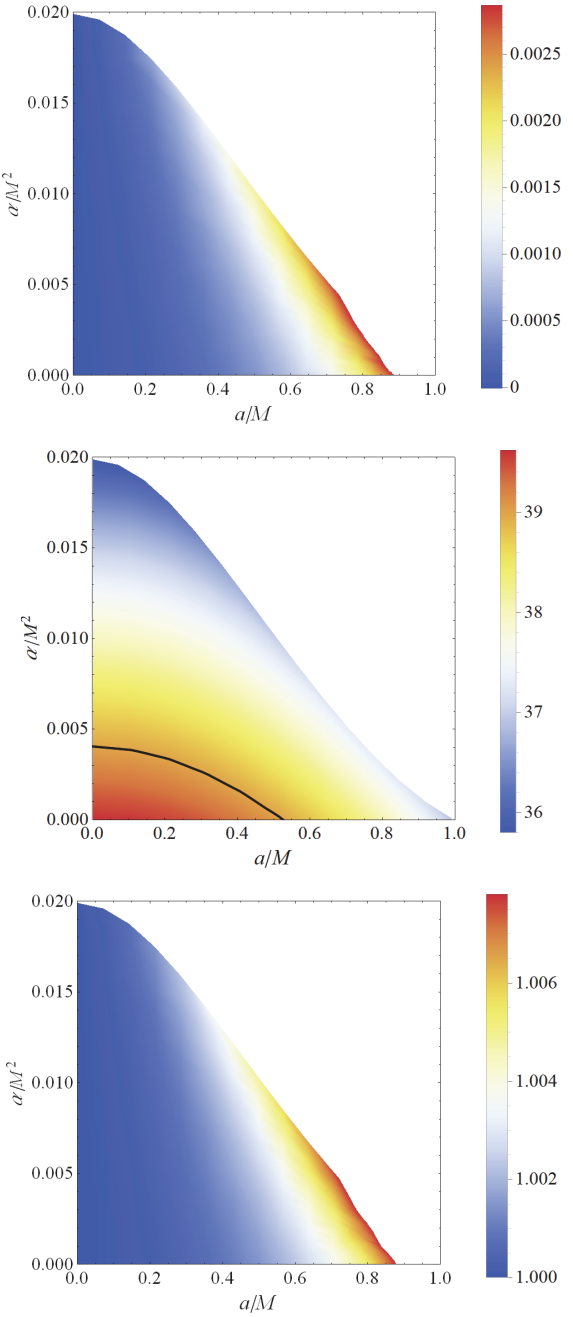


Fig. 5. Circularity deviation observable ΔC (*top*), the angular diameter θ_d (*middle*), and axis ratio D_x (*bottom*) as a function of (a, α) for the rotating EGB black holes. Black solid lines correspond to the M87* black hole shadow bounds $\theta_d = 39\mu$as within the 1σ region, such that the region above the black line is excluded by the EHT bounds.

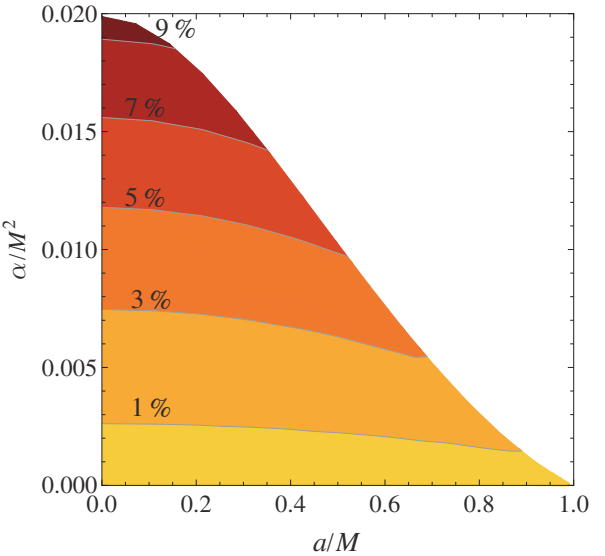


Fig. 6. Relative difference between the shadow angular diameter size of a rotating EGB black hole and a Kerr black hole with the same mass.

free from ghost instabilities. The rotating black holes possess two distinct horizons that eventually merge to form the degenerate horizon for the extremal values of the GB coupling parameter $\alpha = \alpha_E$. The null geodesics equations of motion were obtained in the first-order differential form, and the analytical expressions of the critical impact parameters for unstable spherical photon orbits are derived. The shadow contours are drawn for the rotating black hole for $\theta_0 = 17^o$ inclination angle and compared with those for the Kerr black hole. The rotating black hole shadows deviate from the circularity. The shadow observables A and D are calculated, and it is shown that the EGB black hole shadows are smaller and more distorted than those for the Kerr black hole. Furthermore, these observables are used to estimate the black hole parameters. We modeled the M87* black hole as the rotating $4D$ EGB black hole and used the deduced shadow observables ΔC, θ_d, and D_x for the M87* to constrain the EGB black hole parameters. We have found that only θ_d within 1σ region placed stringent bound the EGB parameters. Whereas ΔC and D_x allows all parameter space. The constraints deduced for $\theta_0 = 17^o$ are weaker than those deduced for $\theta_0 = 90^o$ in Ref.[32] However, it is important to account for systematic uncertainty when identifying observable shadow characteristics like the emission ring and center brightness depression, especially when using low-resolution data, to gravitational qualities like the size and shape of the critical curve. Future observations utilizing an improved ground or space-based array might significantly reduce these systematic errors, and we anticipate better constraints on the GB coupling parameter.

Acknowledgments

S.G.G. would like to thank DST INDO-SA bilateral project DST/INT/South Africa/P-06/2016, SERB-DST for the ASEAN project IMRC/AISTDF/CRD/2018/000042 and also IUCAA, Pune for the hospitality while this work was being done. R.K. would like to thanks UKZN and NRF for the post-doctoral fellowship.

References

1. D. Lovelock, J. Math. Phys. **13** 874 (1972).
2. C. Lanczos, Annals Math. **39** 842 (1938).
3. D. Lovelock, J. Math. Phys. **12** 498 (1971).
4. D. G. Boulware and S. Deser, Phys. Rev. Lett. **55**, 2656 (1985); J. T. Wheeler, Nucl. Phys. B **268**, 737 (1986).
5. S. Nojiri and S. D. Odintsov, Phys. Lett. B **521** 87 (2001); Erratum: [Phys. Lett. B **542** 301 (2002)]; Y. M. Cho and I. P. Neupane, Phys. Rev. D **66**, 024044 (2002); M. Cvetic, S. Nojiri and S. D. Odintsov, Nucl. Phys. **B 628**, 295 (2002); R. G. Cai, Phys. Rev. D **65**, 084014 (2002); I. P. Neupane, Phys. Rev. D **67**, 061501(R) (2003); **69**, 084011 (2004); A. Padilla, Class. Quant. Grav. **20**, 3129 (2003); N. Deruelle, J. Katz, and S. Ogushi, Class. Quant. Grav. **21**, 1971 (2004); M. H. Dehghani, Phys. Rev. D **69**, 064024 (2004); R. G. Cai and Q. Guo, Phys. Rev. D **69**, 104025 (2004); T. Torii and H. Maeda, Phys. Rev. D **71**, 124002 (2005); M. H. Dehghani and R. B. Mann, Phys. Rev. D **72**, 124006 (2005); M. H. Dehghani and S. H. Hendi, Phys. Rev. D **73**, 084021 (2006); M. H. Dehghani, G. H. Bordbar, and M. Shamirzaie, Phys. Rev. D **74**, 064023 (2006).
6. S. Jhingan and S. G. Ghosh, Phys. Rev. D **81**, 024010 (2010); S. G. Ghosh, M. Amir and S. D. Maharaj, Eur. Phys. J. C **77**, 530 (2017); S. G. Ghosh, Class. Quant. Grav. **35**, 085008 (2018).
7. S. Mignemi and N. R. Stewart, Phys. Rev. D **47**, 5259 (1993); P. Kanti, N. E. Mavromatos, J. Rizos, K. Tamvakis and E. Winstanley, Phys. Rev. D **54**, 5049 (1996); S. O. Alexeev and M. V. Pomazanov, Phys. Rev. D **55**, 2110 (1997); T. Torii, H. Yajima and K. i. Maeda, Phys. Rev. D **55**, 739 (1997); R. Konoplya, Phys. Rev. D **71**, 024038 (2005); B. Kleihaus, J. Kunz and E. Radu, Phys. Rev. Lett. **106**, 151104 (2011); A. Maselli, P. Pani, L. Gualtieri and V. Ferrari, Phys. Rev. D **92**, 083014 (2015).
8. T. P. Sotiriou and S. Y. Zhou, Phys. Rev. Lett. **112**, 251102 (2014).
9. T. P. Sotiriou and S. Y. Zhou, Phys. Rev. D **90**, 124063 (2014).
10. D. D. Doneva and S. S. Yazadjiev, Phys. Rev. Lett. **120**, 131103 (2018).
11. P. V. P. Cunha, C. A. R. Herdeiro and E. Radu, Phys. Rev. Lett. **123**, 011101 (2019).
12. Y. Tomozawa, arXiv:1107.1424 [gr-qc].
13. G. Cognola, R. Myrzakulov, L. Sebastiani and S. Zerbini, Phys. Rev. D **88**, 024006 (2013).
14. D. Glavan and C. Lin, Phys. Rev. Lett. **124**, 081301 (2020).
15. R. A. Konoplya and A. Zhidenko, Phys. Rev. D **101**, 084038 (2020).
16. A. Casalino, A. Colleaux, M. Rinaldi and S. Vicentini, Phys. Dark Univ. **31**, 100770 (2021).
17. R. A. Hennigar, D. Kubiznak, R. B. Mann and C. Pollack, JHEP **07**, 027 (2020).
18. L. Ma and H. Lu, Eur. Phys. J. C **80**, 1209 (2020).
19. W. Ai, Commun. Theor. Phys. **72**, 095402 (2020).
20. F. Shu, Phys. Lett. B **811**, 135907 (2020).

21. M. Gurses, T. C. Sisman and B. Tekin, Eur. Phys. J. C **80** 647 (2020).
22. S. Mahapatra, Eur. Phys. J. C **80**, 992 (2020).
23. H. Lu and Y. Pang, Phys. Lett. B **809**, 135717 (2020).
24. T. Kobayashi, JCAP **07**, 013 (2020).
25. J. Arrechea, A. Delhom and A. Jiménez-Cano, Chin. Phys. C **45**, 013107 (2021).
26. K. Aoki, M. A. Gorji and S. Mukohyama, Phys. Lett. B **810**, 135843 (2020).
27. R. B. Mann and S. F. Ross, Class. Quant. Grav. **10**, 1405 (1993).
28. P. G. S. Fernandes, P. Carrilho, T. Clifton and D. J. Mulryne, Phys. Rev. D **102**, 024025 (2020).
29. P. G. S. Fernandes, Phys. Lett. B **805**, 135468 (2020).
30. D. V. Singh, S. G. Ghosh and S. D. Maharaj, Phys. Dark Univ. **30**, 100730 (2020).
31. S. W. Wei and Y. X. Liu, Eur. Phys. J. Plus **136**, 436 (2021).
32. R. Kumar and S. G. Ghosh, JCAP **07**, 053 (2020).
33. S. G. Ghosh and S. D. Maharaj, Phys. Dark Univ. **30**, 100687 (2020).
34. S. G. Ghosh and R. Kumar, Class. Quant. Grav. **37**, 245008 (2020).
35. A. Kumar and S. G. Ghosh, arXiv:2004.01131 [gr-qc].
36. A. Kumar and R. Kumar, arXiv:2003.13104 [gr-qc].
37. S. U. Islam, R. Kumar and S. G. Ghosh, JCAP **09**, 030 (2020).
38. M. Heydari-Fard, M. Heydari-Fard and H. R. Sepangi, EPL **133**, 50006 (2021).
39. X. H. Jin, Y. X. Gao and D. J. Liu, Int. J. Mod. Phys. D **29**, 2050065 (2020).
40. R. Kumar, S. U. Islam and S. G. Ghosh, Eur. Phys. J. C **80**, 1128 (2020).
41. K. Akiyama *et al.*, Astrophys. J. **875**, L1 (2019).
42. K. Akiyama *et al.*, Astrophys. J. **875**, L6 (2019).
43. D. Psaltis *et al.*, Phys. Rev. Lett. **125**, 141104 (2020).
44. Y. Mizuno, *et al.*, Nature Astron. **2**, 585 (2018).
45. F. H. Vincent, M. Wielgus, M. A. Abramowicz, E. Gourgoulhon, J. P. Lasota, T. Paumard and G. Perrin, Astron. Astrophys. **646**, A37 (2021).
46. H. C. D. Lima, Junior, L. B. Crispino, P. P. Cunha and C. R. Herdeiro, Phys. Rev. D **103**, 084040 (2021).
47. E. Newman and A. Janis, J. Math. Phys. **6**, 915 (1965).
48. E. Contreras, J. Ovalle and R. Casadio, Phys. Rev. D **103**, 044020 (2021).
49. M. Azreg-Aïnou, Phys. Rev. D **90**, 064041 (2014).
50. M. Azreg-Aïnou, Eur. Phys. J. C **74**, 2865 (2014).
51. T. Johannsen and D. Psaltis, Phys. Rev. D **83**, 124015 (2011).
52. C. Bambi and L. Modesto, Phys. Lett. B **721**, 329 (2013).
53. S. G. Ghosh, Eur. Phys. J. C **75**, 532 (2015).
54. J. W. Moffat, Eur. Phys. J. C **75**, 175 (2015).
55. R. Kumar, S. G. Ghosh and A. Wang, Phys. Rev. D **101**, 104001 (2020).
56. R. Kumar and S. G. Ghosh, Eur. Phys. J. C **78**, 750 (2018).
57. R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
58. R. G. Cai, L. M. Cao and N. Ohta, JHEP **1004**, 082 (2010); R. G. Cai, Phys. Lett. B **733**, 183 (2014).
59. A. Kehagias and K. Sfetsos, Phys. Lett. B **678**, 123 (2009).
60. J. L. Synge, Mon. Not. R. Astron. Soc. **131**, 463 (1966).
61. J. P. Luminet, Astron. Astrophys. **75**, 228 (1979).
62. J. M. Bardeen, *Black Holes*, Edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973, p. 215).
63. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, New York, 1992).

64. R. Craig Walker, P. E. Hardee, F. B. Davies, C. Ly and W. Junor, Astrophys. J. **855**, 128 (2018).
65. R. Kumar and S. G. Ghosh, Astrophys. J. **892**, 78 (2020).
66. O. Y. Tsupko, Phys. Rev. D **95**, 104058 (2017).
67. P. Kocherlakota *et al.*, Phys. Rev. D **103**, 104047 (2021).
68. C. Bambi, K. Freese, S. Vagnozzi and L. Visinelli, Phys. Rev. D **100**, 044057 (2019).
69. T. Johannsen and D. Psaltis, Astrophys. J. **718**, 446 (2010).
70. T. Johannsen, Astrophys. J. **777**, 170 (2013).

Wormhole interaction in 2d Horăva-Lifshitz quantum gravity

Jan Ambjørn[a]
The Niels Bohr Institute, Copenhagen University,
Blegdamsvej 17, DK-2100 Copenhagen, Denmark
and
IMPP, Radboud University,
Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands
[a] *E-mail: ambjorn@nbi.dk*

Yuki Hiraga[b], Yoshiyasu Ito[c] and Yuki Sato[d]
Department of Physics, Nagoya University,
Chikusaku, Nagoya 464-8602, Japan
[b] *E-mail: hiraga@eken.phys.nagoya-u.ac.jp*
[c] *E-mail: ito@eken.phys.nagoya-u.ac.jp*
and
Institute for Advanced Research, Nagoya University,
Chikusaku, Nagoya 464-8602, Japan
[d] *E-mail: ysato@th.phys.nagoya-u.ac.jp*

A lattice regularization for the 2d projectable Horăva-Lifshitz (HL) quantum gravity is known to be the 2d causal dynamical triangulations (CDT), and the 2d CDT can be generalized so as to include all possible genus contributions non-perturbatively. We show that in the context of HL gravity, effects coming from such a non-perturbative sum over topologies can be successfully taken into account, if we quantize the 2d projectable HL gravity with a simple bi-local wormhole interaction. This conference paper is based on the article, Phys. Lett. B **816** (2021), 136205.[1]

Keywords: Wormholes; Low dimensional quantum gravity; Canonical quantization

1. Introduction

Two-dimensional models of quantum gravity are useful for understanding some aspects of non-perturbative physics, since they can be examined beyond the framework of perturbation theory by analytic computations in many cases. One of such models is the two-dimensional causal dynamical triangulations (2d CDT),[2] which is a lattice toy model of quantum gravity with a global time foliation. Global hyperbolicity imposed at the quantum level do not allow for any topology change in the CDT model, and the continuum limit can be described by quantum mechanics of a 1d universe.[2]

Yet another toy model for quantum gravity that has a global time foliation is 2d Horăva-Lifshitz (HL) quantum gravity that was introduced first in higher dimensions to resolve the issue of perturbative renormalizability by breaking the diffeomorphisms down to the foliation-preserving diffeomorphisms,[3,4] and so the model has a preferred foliation structure. The projectable version of the 2d HL quantum

gravity was discussed in the articles,[5,6] and in particular it was shown that the canonical quantization of the model yields exactly the same quantum Hamiltonian as the one obtained by the continuum limit of 2d CDT.[5] Therefore, one can interpret 2d CDT as a lattice regularization for the 2d projectable HL quantum gravity.

Although the creation of baby universes and wormholes (handles) is not allowed to occur in the original setup of CDT, one can generalize the 2d model to include such configurations in a manner consistent with the scaling limit,[7] and such a generalized model can be fully described by a string field theory for CDT[8a]. Here the word "string" refers to the 1d spatial universe. Based on the string field theory, one can take into account all genera (handles) as well as all baby universes of the two-dimensional spacetime. In addition, as long as one only asks for the amplitude between the states describing connected spatial universes separated a proper time t, this sum over topologies and baby universes can effectively be described by a one-body Hamiltonian.[11,12]

Having the equivalence between the continuum limit of 2d CDT and the 2d projectable HL quantum gravity, we discussed in the article[1] what kind of effective interaction should be added to the classical Lagrangian of the 2d projectable HL gravity when canonically quantized, in order to obtain the above mentioned one-body quantum Hamiltonian that includes all wormholes and baby universes obtained in 2d CDT. The answer is that it is enough to include a simple bi-local and spatial wormhole interaction that is compatible with the foliation-preserving diffeomorphisms.

This article is organized as follows. In Section 2, we give a brief introduction to 2d CDT and 2d HL quantum gravity, and explain the relation between the two. In Section 3 the string field theory for CDT is described and we introduce the one-body quantum Hamiltonian that includes all possible topologies. In Section 4, we introduce a simple wormhole interaction to the 2d projectable HL gravity and show that one can precisely recover the one-body Hamiltonian when canonically quantizing the model. Section 5 is devoted to discussions.

2. CDT and Horăva-Lifshitz gravity in two dimensions

In this Section, we review causal dynamical triangulations (CDT) and Horăva-Lifshitz gravity (HL) in two dimension, and explain the relation between the two.

2.1. 2*d CDT*

The starting point is a globally hyperbolic manifold equipped with a global time foliation:

$$\mathcal{M} = \bigcup_{t\in\mathbb{R}} \Sigma_t \ , \tag{1}$$

[a]The time-independent amplitudes can be also computed by the matrix model for CDT[9] and by the new scaling limit of the Hermitian one-matrix model.[10]

where each leaf Σ_t is a Cauchy "surface". 2d CDT is a model which quantizes 2d geometries with such a proper-time foliation. The geometries in the path integral are regularized by piecewise linear geometries constructed by gluing together special kinds of triangles.[2] Each triangle consists of one space-like edge and two time-like edges such that a square of the space-like edge is positive, $a_s^2 = \varepsilon^2$, while it is negative for the time-like edge, $a_t^2 = -\alpha\varepsilon^2$ with positive α.

If we prohibit a creation of baby universes, a reasonable choice of the action is the cosmological constant term that is regularized as follows:

$$\begin{aligned} S[g_{ij}] &= -\lambda_0 \int d^2x \, \sqrt{-\det(g_{ij})} \ , \\ \rightarrow \quad S_R[T;\alpha] &= -\frac{\mu}{\varepsilon^2} \times \left(\frac{\sqrt{4\alpha+1}}{4} \varepsilon^2 \times n(T) \right) \ , \end{aligned} \tag{2}$$

where the bare cosmological constant λ_0 is replaced by the dimensionless cosmological constant μ, and the quantity in the parentheses is the discretized area of the triangulation T, i.e. the area of each triangle times the number of triangles $n(T)$. The lattice action S_R in eq. (2) is called the Regge action.[13]

For computational convenience we implement a rotation to the Euclidean signature which can be done by replacing α with $-\alpha$:

$$iS_R[T;\alpha] \rightarrow -S_R^{(e)}[-\alpha;T] = -\mu \frac{\sqrt{4\alpha-1}}{4} n(T) =: -\mu\, n(T) \ , \tag{3}$$

where we have absorbed a numerical factor into the dimensionless cosmological constant. Note that this map is a bijection between individual Lorentzian and Euclidean geometries.

In CDT, the integration over diffeomorphism equivalent classes of metric g keeping both initial and final geometries fixed can be regularized by the sum over "all" triangulations. Therefore, the 2d Euclidean path-integral regularized by CDT is

$$G^{(0)}_{\text{lattice}}(l_1, l_2; \tau) = \sum_{T \in \mathcal{T}(l_1, l_2; \tau)} e^{-\mu n(T)} \ , \tag{4}$$

where $\mathcal{T}(l_1, l_2; \tau)$ is a set of triangulations such that the initial and final boundaries whose lengths are kept fixed to l_1 and l_2 are separated by τ Euclidean time steps.

One can compute the amplitude (4) analytically through the use of the generating function for the numbers $G^{(0)}_{\text{lattice}}(l_1, l_2; \tau)$.[2] As in the case of lattice QCD, tuning the UV relevant coupling constant μ/ε^2 to its critical value μ_c/ε^2 and taking $\varepsilon \rightarrow 0$ in a correlated manner, one can transmute the dimension of the lattice spacing to the renormalized coupling constant:

$$\lambda := \lim_{\substack{\varepsilon \rightarrow 0 \\ \mu \rightarrow \mu_c}} \frac{\mu - \mu_c}{\varepsilon^2} \ , \tag{5}$$

where λ is the renormalized cosmological constant. Introducing the renormalized quantities, boundary lengths and a proper time such that

$$\ell_1 := \varepsilon l_1 \ , \quad \ell_2 := \varepsilon l_2 \ , \quad t := \varepsilon \tau \ , \tag{6}$$

one can obtain the renormalized amplitude $G^{(0)}(\ell_1, \ell_2; t)$ that is known to satisfy the differential equation:[2,5]

$$-\frac{\partial}{\partial t} G^{(0)}(\ell_1, \ell_2; t) = H_a^{(0)}(\ell_1) G^{(0)}(\ell_1, \ell_2; t) \ , \tag{7}$$

where $H_a^{(0)}$ is the quantum Hamiltonian defined as

$$H_0^{(0)}(\ell) = -\frac{\partial}{\partial \ell} \ell \frac{\partial}{\partial \ell} + \lambda \ell \ , \quad H_{-1}^{(0)}(\ell) = -\ell \frac{\partial^2}{\partial \ell^2} + \lambda \ell \ , \quad H_{+1}^{(0)}(\ell) = -\frac{\partial^2}{\partial \ell^2} \ell + \lambda \ell \ . \tag{8}$$

Here the label a in eq. (7) specifies the ordering of the Hamiltonian. If we define a quantum state of the one-dimensional universe whose length is ℓ as $|\ell\rangle$, the amplitude $G^{(0)}(\ell_1, \ell_2; t)$ can be rewritten as

$$G^{(0)}(\ell_1, \ell_2; t) = \langle \ell_2 | e^{-t H_a^{(0)}(\ell_1)} | \ell_1 \rangle \ . \tag{9}$$

In fact, each ordering specifies the geometry of the quantum state $|\ell\rangle$, i.e. if $a = 0, -1, +1$, then the geometry of the one-dimensional universe is open, closed with a mark, and closed, respectively.[2,5] Marking a point on a closed universe is analogous to introducing a coordinate. The Hamiltonian is hermitian with respect to the following inner product:

$$\langle \phi | H_a^{(0)} | \psi \rangle = \int_0^\infty \phi^*(\ell) H_a^{(0)}(\ell) \psi(\ell) \ d\mu_a(\ell) \ , \quad \text{with} \quad d\mu_a(\ell) = \ell^a d\ell \ . \tag{10}$$

As a result, the physics of 2d CDT can be described by the quantum mechanics of a one-dimensional universe.

2.2. *2d projectable HL gravity*

The starting point is the same as that of CDT, i.e. a globally hyperbolic manifold equipped with a global time foliation (1). A natural parametrization for the metric on such a geometry is given by the Arnowitt-Deser-Misner metric:

$$g = -N^2 ds^2 + h_{11}(dx + N^1 ds)(dx + N^1 ds) \ , \tag{11}$$

where h_{11} is the spatial metric on the leaf; N and N^1 called lapse and shift functions, quantify the normal and tangential directions of the proper time to the leaf, respectively.

The 2d HL gravity is introduced as a theory that keeps the foliation structure,[3] or in other words, it is invariant under the foliation preserving diffeomorphisms (FPD):

$$s \to s + \xi^0(s) \ , \quad x \to x + \xi^1(s, x) \ . \tag{12}$$

Under the FPD, the fields transform as

$$\delta_\xi h_{11} = \xi^0 \partial_0 h_{11} + \xi^1 \partial_1 h_{11} + 2h_{11}\partial_1 \xi^1 \ , \tag{13}$$

$$\delta_\xi N_1 = \xi^\mu \partial_\mu N_1 + N_1 \partial_\mu \xi^\mu + h_{11}\partial_0 \xi^1 \ , \tag{14}$$

$$\delta_\xi N = \xi^\mu \partial_\mu N + N\partial_0 \xi^0 \ , \tag{15}$$

where $N_1 = h_{11}N^1$.

Note that if the lapse function N is a function of time, $N = N(t)$, it stays as a function of time under the FPD. The 2d projectable HL gravity satisfies this condition on N, and it is defined by the following action:

$$I = \frac{1}{\kappa}\int dsdx\ N\sqrt{h}\left((1-\eta)K^2 - 2\tilde{\lambda}\right) \ , \tag{16}$$

where η, $\tilde{\lambda}$ and κ are a dimensionless parameter, the cosmological constant and the dimensionless gravitational coupling constant, respectively; h is the determinant of the metric h_{11}, i.e. $h = h_{11}$; K is the trace of the extrinsic curvature K_{11} defined as

$$K_{11} = \frac{1}{2N}\left(\partial_0 h_{11} - 2\nabla_1 N_1\right) \ , \quad \text{with} \quad \nabla_1 N_1 := \partial_1 N_1 - \Gamma^1_{11} N_1 \ . \tag{17}$$

Here Γ^1_{11} is the spatial Christoffel symbol:

$$\Gamma^1_{11} = \frac{1}{2}h^{11}\partial_1 h_{11} \ . \tag{18}$$

In principle, one can add higher spatial derivative terms to the action (16), but they are not needed since 2d gravity is renormalizable without such terms and we will omit such terms.

The quantization of 2d projectable HL gravity was discussed in,[5,6] and in particular, it was shown that the quantum Hamiltonian coincides with the continuum Hamiltonian of 2d CDT when the following identification of the parameters is made:[5]

$$\lambda = \frac{\tilde{\lambda}}{2(1-\eta)} \ , \quad \eta < 1 \ , \quad \tilde{\lambda} > 0 \ , \quad \kappa = 4(1-\eta) \ . \tag{19}$$

where λ is the renormalized cosmological constant in 2d CDT (5)[b].

Let us briefly explain how to recover the quantum Hamiltonian (8) from the quantization of 2d HL gravity. Introducing the conjugate momentum of $\sqrt{h}$ as π, we have in the canonical formalism the Poisson bracket

$$\left\{\sqrt{h(s,x)}, \pi(s,x')\right\} = \delta(x-x') \, , \tag{20}$$

and corresponding to the Lagrangian (16) we have the Hamiltonian

$$H = \int dx\ \left[N_1\left(-\frac{\partial_1 \pi}{\sqrt{h}}\right) + N\left(\frac{\kappa}{4(1-\eta)}\pi^2\sqrt{h} + \frac{2}{\kappa}\tilde{\lambda}\sqrt{h}\right)\right] \ . \tag{21}$$

[b]We have set unimportant dimensionless gravitational constant as $\kappa = 4(1-\eta)$.

If we solve the momentum constraint at the classical level, i.e.

$$-\frac{\partial_1 \pi}{\sqrt{h}} = 0 \ , \quad \Rightarrow \quad \pi = \pi(s) \ , \tag{22}$$

the system reduces to a one-dimensional model with the Hamiltonian

$$H = N(s) \left(\frac{\kappa}{4(1-\eta)} \pi^2(s) \ell(s) + \frac{2}{\kappa} \tilde{\lambda} \ell(s) \right) \ , \quad \text{with} \quad \ell(s) := \int dx \ \sqrt{h(s,x)} \ . \tag{23}$$

Hereafter choosing the correct sign for the kinetic term, i.e. $\eta < 1$, we use the parametrization (19) with positive λ in order to discuss the relation to 2d CDT.

The classical 1d system with the Hamiltonian (23) can be alternatively described by the following action:

$$S = \int_0^1 ds \left(\frac{\dot{\ell}^2}{4N\ell} - \lambda N \ell \right) \ , \tag{24}$$

where $\dot{\ell} := d\ell/ds$. This system is invariant under the time reparametrization, $s \to s + \xi^0(t)$, which is ensured by the lapse function. In fact, the proper time,

$$t := \int_0^1 ds \ N(s) \ , \tag{25}$$

and the length, $\ell = \ell(t)$, are invariant under the time reparametrization, and so it makes sense to discuss the probability amplitude for a 1d universe to propagate in the proper time $t(> 0)$, starting from the state with the length ℓ_1 and ending up in the one with length ℓ_2.[5] Such an amplitude can be computed based on the path-integral, and we evaluate it by a rotation to the Euclidean signature for convenience. In our foliated spacetime, for $\eta < 1$, we can implement this procedure by a formal rotation, $s \to is$, which yields the amplitude:

$$G^{(0)}(\ell_2, \ell_1; t) = \int \frac{\mathcal{D}N(s)}{\mathrm{Diff}[0,1]} \int_{\ell(0)=\ell_1}^{\ell(1)=\ell_2} \mathcal{D}\ell(s) e^{-S_E[N(s),\ell(s)]} \ , \tag{26}$$

where Diff$[0,1]$ is the volume of the time reparametrization; S_E is the Euclidean action given by

$$S_E = \int_0^1 ds \left(\frac{\dot{\ell}^2}{4N\ell} + \lambda N \ell \right) \ , \tag{27}$$

where $\dot{\ell} := d\ell/ds$.

We set $N = 1$ as a gauge choice. One can show that the corresponding Faddeev-Popov determinant only gives an overall constant, which we will omit in the following. The amplitude (26) then becomes

$$G^{(0)}(\ell_2, \ell_1; t) = \int_{\ell(0)=\ell_1}^{\ell(t)=\ell_2} \mathcal{D}\ell(s) \exp\left[- \int_0^t ds \left(\frac{\dot{\ell}^2}{4\ell} + \lambda \ell \right) \right] \ , \tag{28}$$

which can be expressed in terms of the quantum Hamiltonian H that is unknown at the moment:

$$G^{(0)}(\ell_2, \ell_1; t) = \langle \ell_2 | e^{-tH} | \ell_1 \rangle \ , \tag{29}$$

where $|\ell\rangle$ is a quantum state of the 1d universe with the length ℓ. By standard methods (see e.g. the article[5]), one can determine the quantum Hamiltonian from eq. (28) and eq. (29). One obtains precisely the 2d CDT quantum Hamiltonians given by eq. (8) if the integral measures are chosen as

$$\mathcal{D}\ell(s) = \prod_{s=0}^{s=t} \ell^a(s) d\ell(s) \ , \tag{30}$$

where $a = 0, \pm 1$. The measures (30) are consistent with eq. (10) introduced in 2d CDT.

Therefore, we can conclude that 2d CDT is a lattice regularization for the 2d projectable HL quantum gravity.

3. Sum over all genera in 2d CDT

One can generalize the 2d CDT model so as to include spatial topology changes (splitting and joining interactions of 1d universe) in keeping with the foliation structure, and this is described by promoting the 1d quantum mechanics to a field theory of 1d universes, which is called the string field theory for CDT.[8] Here the word "string" means a 1d closed spatial universe.

We introduce an operator that creates a marked closed string with the length ℓ, $\Psi^\dagger(\ell)$, and an operator that annihilates a length ℓ closed string without a mark, $\Psi(\ell)$. They satisfy the commutation relation:

$$[\Psi(\ell), \Psi^\dagger(\ell')] = \delta(\ell - \ell'), \quad [\Psi(\ell), \Psi(\ell')] = [\Psi^\dagger(\ell), \Psi^\dagger(\ell')] = 0. \tag{31}$$

The vacuum $|\text{vac}\rangle$ is defined by $0 = \Psi(\ell)|\text{vac}\rangle = \langle \text{vac}|\Psi^\dagger(\ell)$.

The CDT amplitude (9) can be written in terms of the string field Hamiltonian obtained by sandwiching the one-body Hamiltonian:

$$G^{(0)}(\ell_1, \ell_2; t) = \langle \text{vac}|\Psi(\ell_2)\ e^{-t\hat{H}^{(0)}}\ \Psi^\dagger(\ell_1)|\text{vac}\rangle \ , \tag{32}$$

where

$$\hat{H}^{(0)} = \int_0^\infty \frac{d\ell}{\ell} \Psi^\dagger(\ell) H^{(0)}_{-1} \Psi(\ell) \ . \tag{33}$$

In order to incorporate topology change, one has to include suitable interactions, and the full string field Hamiltonian is given by:[8]

$$\begin{aligned} \hat{H} = \hat{H}^{(0)} &- \int_0^\infty d\ell\ \delta(\ell)\Psi(\ell) - g_s \int_0^\infty d\ell_1 \int_0^\infty d\ell_2\ \Psi^\dagger(\ell_1)\Psi^\dagger(\ell_2)(\ell_1 + \ell_2)\Psi(\ell_1 + \ell_2) \\ &- \alpha g_s \int_0^\infty d\ell_1 \int_0^\infty d\ell_2\ \Psi^\dagger(\ell_1 + \ell_2)\ell_1\Psi(\ell_1)\ell_2\Psi(\ell_2) \ , \end{aligned} \tag{34}$$

where the second, third and fourth terms mean a string vanishing into the vacuum, the splitting interaction with the string coupling g_s, and the joining interaction with the coupling αg_s, respectively; α is a constant introduced for counting handles. In general one can compute the following amplitude:

$$A(\ell_1, \cdots, \ell_m; \ell'_1, \cdots, \ell'_n; t) = \langle \text{vac}|\Psi(\ell'_1)\cdots\Psi(\ell'_n)e^{-t\hat{H}}\Psi^\dagger(\ell_1)\cdots\Psi^\dagger(\ell_m)|\text{vac}\rangle \ . \tag{35}$$

From now on, we focus on the full propagator $G(\ell_1, \ell_2; t)(:= A(\ell_1; \ell_2; t))$ that includes the sum over all genera (handles) and baby universes, and so we simply set $\alpha = 1$. In this case, somewhat miraculously, the full propagator in the multi-body system can be described by an effective one-body system:[11, 12]

$$G(\ell_1, \ell_2; t) = \langle \ell_2|e^{-tH_{-1}}|\ell_1\rangle \ , \quad \text{with} \quad H_{-1} = -\ell\frac{\partial^2}{\partial\ell^2} + \lambda\ell - g_s\ell^2 \ . \tag{36}$$

All the contributions coming from the sum over all genera and baby universes can be effectively described by the last term $-g_s\ell^2$. Although the Hamiltonian (36) is not bounded from below, it belongs to a class of Hamiltonians which are called "classical incomplete", where the Hamiltonians have discrete energy spectra and square integrable eigenfunctions.

The effective one-body system given by eq. (36) can be also described by the path-integral:[1, 14]

$$G(\ell_1, \ell_2; t) = \int_{\ell(0)=\ell_1}^{\ell(t)=\ell_2} \mathcal{D}\ell(s) \ \exp\left[-\int_0^t ds \ \left(\frac{\dot{\ell}^2(s)}{4\ell(s)} + \lambda\ell(s) - g_s\ell^2(s)\right)\right] \ . \tag{37}$$

If we choose the integral measure such that

$$\mathcal{D}\ell(s) = \prod_{s=0}^{s=t} \ell^a(s)d\ell(s) \ , \tag{38}$$

where $a = 0, \pm 1$, all possible orderings of the full one-body Hamiltonian (36) can be realized. As explained in the article,[14] in order for the functional integral to be well defined, the boundary conditions on $\ell(s)$ at infinity have to be chosen such that the kinetic term counteracts the unboundedness of the potential. As we will see, the information about the boundary conditions also appears in the classical Hamiltonian constraint (42) of 2d projectable HL gravity with a wormhole interaction.

In the next section, we will show that the path-integral (37) can be obtained quantizing the 2d projectable HL gravity with a wormhole interaction.

4. Wormhole interaction in 2d projectable HL gravity

We consider 2d projectable HL gravity with a space-like wormhole interaction given by the action:

$$\begin{aligned} I_w = \frac{1}{\kappa}\int dsdxN(s)\sqrt{h(s,x)}\left((1-\eta)K^2(s,x) - 2\tilde{\lambda}\right) \\ + \beta\int dsN(s)\int dx_1dx_2\sqrt{h(s,x_1)}\sqrt{h(s,x_2)} \ , \end{aligned} \tag{39}$$

where β is a dimensionfull coupling constant. One can show that the action (39) is invariant under the FPD (12) with the projectable lapse function, $N = N(t)$. The bi-local interaction in eq. (39) relates two distinct points at an equal time. The action is a simplified version of the general bi-local action suggested in the article,[15] made possible because HL gravity is invariant only under the FPD (12) and not the full set of diffeomorphisms.

Following the same procedure explained in the section 2.2, we quantize the system with the action (39). Introducing a conjugate momentum of $\sqrt{h}$ as π as before, we move on to the canonical formalism, and the Hamiltonian is

$$H = \int dx \ \left[N_1 \mathcal{C}^1(s,x) + N\mathcal{C}(s,x) \right] \ , \tag{40}$$

where

$$\mathcal{C}^1(s,x) = -\frac{\partial_1 \pi(s,x)}{\sqrt{h(s,x)}} \ , \tag{41}$$

$$\mathcal{C}(s,x) = \frac{\kappa}{4(1-\eta)} \pi^2(s,x)\sqrt{h(s,x)} + \frac{2}{\kappa}\tilde{\lambda}\sqrt{h(s,x)} - \beta\sqrt{h(s,x)} \int dx_2 \sqrt{h(s,x_2)} \ . \tag{42}$$

If we solve the momentum constraint (41) at the classical level, the system again reduces to the 1d system with the Hamiltonian:

$$H = N(s) \left(\frac{\kappa}{4(1-\eta)} \pi^2(s)\ell(s) + \frac{2}{\kappa}\tilde{\lambda}\ell(s) - \beta\ell^2(s) \right) \ , \tag{43}$$

where $\ell(s) := \int dx \ \sqrt{h(s,x)}$. Hereafter choosing the correct sign for the kinetic term, i.e. $\eta < 1$, we use the parametrization (19) with positive λ as before.

The classical Hamiltonian constraint (42) can be solved as

$$\pi^2 = -\lambda + \beta\ell \geq 0 \ , \tag{44}$$

for $\sqrt{\lambda}\ell \geq 1/\xi$ with $\xi := \beta/\lambda^{3/2}$, and otherwise, the classical Hamiltonian constraint (42) requires $\ell = 0$ on the constraint surface. As in the $\beta = 0$ case, when quantizing the system based on the path-integral, we don't have any problem with respect to the quantization around $\ell(s) = 0$.

Simply repeating the procedure in the section 2.2, one can show that the amplitude (37) can be precisely recovered by quantizing the 2d projectable HL gravity with the bi-local interaction based on the path-integral, if $\beta = g_s$.

5. Discussions

We have canonically quantized the 2d projectable HL gravity with a simple space-like wormhole interaction, and shown that the quantum Hamiltonian is equivalent to the one-body quantum Hamiltonian that includes contributions coming from all wormholes and baby universes obtained in the string field theory for CDT, if $g_s = \beta$ and $\lambda = \tilde{\lambda}/(2(1-\eta))$ where $\beta > 0$, $\lambda > 0$ and $\eta < 1$.

Let us consider the classical Hamiltonian constraint $\mathcal{C}$ (42) in the parameter region above. When $\sqrt{\lambda}\ell \geq 1/\xi$ where ξ is a dimensionless quantity defined by $\xi := g_s/\lambda^{3/2}$, the constraint surface is given by $\pi^2 = -\lambda + g_s\ell \geq 0$. However, when $\sqrt{\lambda}\ell < 1/\xi$, the only allowed solution is $\ell = 0$. In the case of the 2d projectable HL gravity ($\beta = g_s = 0$), with the parameter region corresponding to 2d CDT, i.e. $\lambda > 0$, the only solution to the classical Hamiltonian constraint is $\ell = 0$. Therefore, when $\sqrt{\lambda}\ell < 1/\xi$, the classical solution of the 2d projectable HL with the wormhole interaction will be close to that of the 2d projectable HL gravity, if one sits in the parameter region above; they can be quite different when $\sqrt{\lambda}\ell \geq 1/\xi$. Such a relation also holds at the quantum level. As shown in the article,[14] when $\sqrt{\lambda}\ell < 1/\xi$ the eigenfunctions of the one-body quantum Hamiltonian including all wormholes and baby universes, H_{-1} (36), can be well approximated by the eigenfunctions of the quantum Hamiltonian of 2d CDT without wormholes and baby universes, $H^{(0)}_{-1}$ (8). On the other hand, when $\sqrt{\lambda}\ell \geq 1/\xi$, their behaviors are quite different, and in this case, in order for the theory to be well-defined, the unbounded nature of the potential in H_{-1} will be counteracted by the kinetic term. This balance between the kinetic and potential terms is precisely what is reflected in the classical Hamiltonian constraint (42). This is also consistent with the boundary conditions on ℓ at infinity in the path-integral (38).

The picture of creation and annihilation of baby universes and wormholes is conceptually straightforward in the string field theory for CDT.[8] Nevertheless it is somewhat surprising that one from this can derive an effective one-body Hamiltonian which can describe propagation of a single spatial universe, i.e. the propagation where the spatial universe starts with the topology of a circle and at a later time t has the same topology, but where it in the intermediate times is allowed to split in two and either one part disappears in the vacuum (a baby universe), or the two parts join again at a later time (then changing the spacetime topology). This process of joining and splitting can be iterated at intermediate times and using string field theory we can perform the summation of all iterations and derive the effective one-body Hamiltonian (36). From the point of view of unitary evolution (or using Euclidean time, semigroup evolution, to be more precise), as given in eq. (29), it is difficult to understand how a complete set of intermediate states can both be given by the one spatial universe states $|\ell\rangle$ and by the complete multi-universe Fock states of the string field theory (for a recent discussion of this issue see i.e.[16]). However, this seems to be the case by explicit calculation, and we find it even more surprising that the simplest wormhole interaction term added to the classical action as in eq. (39) leads to precisely the same single universe quantum Hamiltonian as found in the string field theory for CDT.

References

1. J. Ambjørn, Y. Hiraga, Y. Ito and Y. Sato, "Wormholes in 2d Horăva-Lifshitz quantum gravity," Phys. Lett. B **816** (2021), 136205 doi:10.1016/j.physletb.2021.136205 [arXiv:2101.07401 [hep-th]].

2. J. Ambjørn and R. Loll, "Nonperturbative Lorentzian quantum gravity, causality and topology change," Nucl. Phys. B **536** (1998), 407-434 doi:10.1016/S0550-3213(98)00692-0 [arXiv:hep-th/9805108 [hep-th]].
3. P. Hořava, "Quantum gravity at a Lifshitz point," Phys. Rev. D **79** (2009) 084008, [arXiv:0901.3775, hep-th].
4. P. Hořava and C.M. Melby-Thompson, "General covariance in quantum gravity at a Lifshitz point," Phys. Rev. D **82** (2010) 064027, [arXiv:1007.2410, hep-th].
5. J. Ambjørn, L. Glaser, Y. Sato and Y. Watabiki, "2d CDT is 2d Hořava–Lifshitz quantum gravity," Phys. Lett. B **722** (2013), 172-175 doi:10.1016/j.physletb.2013.04.006 [arXiv:1302.6359 [hep-th]].
6. B. F. Li, A. Wang, Y. Wu and Z. C. Wu, "Quantization of (1+1)-dimensional Hořava-Lifshitz theory of gravity," Phys. Rev. D **90** (2014) no. 12, 124076 doi:10.1103/PhysRevD.90.124076 [arXiv:1408.2345 [hep-th]].
7. J. Ambjørn, R. Loll, W. Westra and S. Zohren, "Putting a cap on causality violations in CDT," JHEP **12** (2007), 017 doi:10.1088/1126-6708/2007/12/017 [arXiv:0709.2784 [gr-qc]].
8. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra and S. Zohren, "A String Field Theory based on Causal Dynamical Triangulations," JHEP **05** (2008), 032 doi:10.1088/1126-6708/2008/05/032 [arXiv:0802.0719 [hep-th]].
9. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra and S. Zohren, "A Matrix Model for 2D Quantum Gravity defined by Causal Dynamical Triangulations," Phys. Lett. B **665** (2008), 252-256 doi:10.1016/j.physletb.2008.06.026 [arXiv:0804.0252 [hep-th]].
10. J. Ambjørn, R. Loll, Y. Watabiki, W. Westra and S. Zohren, "A New continuum limit of matrix models," Phys. Lett. B **670** (2008), 224-230 doi:10.1016/j.physletb.2008.11.003 [arXiv:0810.2408 [hep-th]].
11. J. Ambjørn, R. Loll, W. Westra and S. Zohren, "Summing over all Topologies in CDT String Field Theory," Phys. Lett. B **678** (2009), 227-232 doi:10.1016/j.physletb.2009.06.031 [arXiv:0905.2108 [hep-th]].
12. J. Ambjørn, R. Loll, W. Westra and S. Zohren, "Stochastic quantization and the role of time in quantum gravity," Phys. Lett. B **680** (2009), 359-364 doi:10.1016/j.physletb.2009.09.016 [arXiv:0908.4224 [hep-th]].
13. T. Regge, "General Relativity without Coordinates," Nuovo Cim. **19** (1961), 558-571 doi:10.1007/BF02733251
14. J. Ambjørn, Y. Sato and Y. Watabiki, "Wormholes, a fluctuating cosmological constant and the Coleman mechanism," Phys. Lett. B **815** (2021), 136152 doi:10.1016/j.physletb.2021.136152 [arXiv:2101.00478 [hep-th]].
15. S. W. Hawking, "The Effective action for wormholes," Nucl. Phys. B **363** (1991), 117-131 doi:10.1016/0550-3213(91)90237-R
16. D. Marolf and H. Maxfield, "Transcending the ensemble: Baby universes, spacetime wormholes, and the order and disorder of black hole information," JHEP **08** (2020), 044, doi:10.1007/JHEP08(2020)044 [arXiv:2002.08950 [hep-th]].

Nature of singularities in vector-tensor theories of gravity

V. H. Satheeshkumar

Departamento de Física, Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, RJ 22290-240, Brazil
vhsatheeshkumar@gmail.com

The Vector-Tensor (VT) theories of gravity are a class of alternative theories to General Relativity (GR) that are characterized by the presence of a dynamical vector field besides the metric. They are studied in attempts to understand spontaneous Lorentz violation, to generate massive gravitons, and as models of dark matter and dark energy. In this article, I outline how the nature of singularities and horizons in VT theories differ greatly from GR even under the same ordinary conditions. This is illustrated with Einstein-aether theory where vacuum black hole solutions have naked singularities and vacuum cosmological solutions have new singularities that are otherwise absent in GR. It would be interesting to explore these deviations using gravitational waves.

Keywords: Vector-tensor theories of gravity, Modified theories of gravity, Lorentz violation, Spacetime singularities, Naked singularities, Event horizon, Universal horizon

1. Introduction

In the 1960s and early 1970s Roger Penrose, Stephen Hawking and Robert Geroch proved, independently and often in collaboration, a series of theorems on global features of spacetimes in GR. These theorems implied that singularities are inevitable in physically important situations of gravitational collapse and cosmology. The unpreventable existence of singularities for wide classes of rather general models in GR (and other classical theories of gravitation) marks the breakdown of the theory. That is why understanding singularities is crucial for us to find a replacement of GR at very high energies. Even after decades of work, very little is known about the structure and properties of spacetime singularities. For the latest developments on these issues, I refer the reader to the recent reviews by Witten,[1] and Penrose,[2] and also an accessible introduction to the subject by Joshi.[3]

In this paper, we are interested in studying how the presence of a dynamical vector field present in VT theories affects the nature of singularities in comparison to GR. There are many VT theories, but of them the Einstein-aether (EA) theory has the most general diffeomorphism-invariant action involving a spacetime metric and a vector field with the field equations being the second-order differential equations in terms of not only the metric but also the aether field. We study both timelike and spacelike singularities in EA theory. To this end, we consider both black hole and cosmological singularities. In both cases, by computing the Kretschmann scalar, we guarantee that we are dealing with curvature singularities, not conical or caustic singularities.

The paper is organized as follows. The Section 2 presents a quick overview of VT theories. Section 3, briefly outlines the EA theory. In Section 4 and 5, we present future and past singularities respectively. We end with a short summary in Section 6.

2. Vector-Tensor Theories of Gravity

One simple and straightforward extension of GR involves an introduction of a dynamical, timelike four-vector field in addition to the dynamical metric. This class of theories is referred to as Vector-Tensor theories of gravity. In some models, the four-vector is unconstrained, while in others it is constrained to have unit norm. The examples of unconstrained theories are Will-Nordtvedt theory[4,5] (1972), Hellings-Nordtvedt theory[6] (1973) and General vector-tensor theory[7] (1981). The constrained theories include Kostelecky-Samuel theory[8] (1989), Einstein-Aether theory[9] (2001) and Khronometric theory[10] (2010).

The most general action for such a theory which is Lagrangian-based and whose equations of motion involving the vector-field are linear and at most of second order is given by,

$$S = \frac{1}{16\pi G}\int \Big[(1+\omega\, g_{ab}u^a u^b)R - K^{ab}{}_{mn}\nabla_a u^m \nabla_b u^n + \lambda(g_{ab}u^a u^b + 1)\Big]\sqrt{-g}\, d^4x, \quad (1)$$

where

$$K^{ab}{}_{mn} = c_1 g^{ab}g_{mn} + c_2\delta^a_m\delta^b_n + c_3\delta^a_n\delta^b_m - c_4 u^a u^b g_{mn}, \quad (2)$$

the c_i being dimensionless coupling constants, and λ is a Lagrange multiplier enforcing the unit timelike constraint on the aether. The parameter ω is taken to be zero in constrained theories, while λ is set to zero in unconstrained theories.[11] Below, we give the conditions under which different VT theories can be obtained from the general action given above.

- Will-Nordtvedt theory: $\lambda = 0$, $c_1 = -1$ and $c_2 = c_3 = c_4 = 0$
- Hellings-Nordtvedt theory: $\lambda = 0$, $c_1 = 2$, $c_2 = 2\omega$, $c_1 + c_2 + c_3 = 0$ and $c_4 = 0$
- General vector-tensor theory: $\lambda = 0$, $c_1 = 2\epsilon - \tau$, $c_2 = -\eta$, $c_1 + c_2 + c_3 = -\tau$ and $c_4 = 0$
- Kostelecky-Samuel theory: $\omega = 0$ and u^a is not necessarily timelike.
- Einstein-Aether theory: $\omega = 0$ and u^a is always timelike unit vector.
- Khronometric theory: $\omega = 0$, $c_1 = -\epsilon$, $c_2 = \lambda_K$, $c_3 = \beta_K + \epsilon$ and $c_4 = \alpha_K + \epsilon$ where the limit $\epsilon \to \infty$ is to be taken.

Although a lot of work is done on VT theories, there are few reviews on the subject.[12] From the latest phenomenological studies,[13–15] it is evident that EA theory is the most representative of the VT theories and observationally relevant. For this reason, we shall restrict our discussion to EA theory.

3. Einstein-Aether Theory

The general action[9] of the EA theory is given by

$$S = \frac{1}{16\pi G}\int \Big[R - K^{ab}{}_{mn}\nabla_a u^m \nabla_b u^n + \lambda(g_{ab}u^a u^b + 1)\Big]\sqrt{-g}\, d^4x + S_{matter}, \quad (3)$$

where

$$K^{ab}{}_{mn} = c_1 g^{ab}g_{mn} + c_2\delta^a_m\delta^b_n + c_3\delta^a_n\delta^b_m - c_4 u^a u^b g_{mn}, \quad (4)$$

the c_i being dimensionless coupling constants, and λ is a Lagrange multiplier enforcing the unit timelike constraint on the aether, and $\delta^a_m\delta^b_n = g^{a\alpha}g_{\alpha m}g^{b\beta}g_{\beta n}$. In the weak-field, slow-motion limit EA theory reduces to Newtonian gravity with a value of Newton's constant $G_{\rm N}$ related to the parameter G in the action (3) by,

$$G = G_N\left(1 - \frac{c_{14}}{2}\right). \quad (5)$$

Here, the constant c_{14} is defined as

$$c_{14} = c_1 + c_4. \quad (6)$$

Note that if $c_{14} = 0$ the EA coupling constant G becomes the Newtonian coupling constant G_N, without necessarily imposing $c_1 = c_4 = 0$. For $c_{14} > 2$ the coupling constant G becomes negative, implying that the gravity is repulsive. The coupling constant vanishes when $c_{14} = 2$ which renders the action undefined. Thus, physically interesting region is $0 \le c_{14} < 2$. These free parameters have been severely constrained using many observational/experimental tests, including gravitational waves.[16]

The field equations are obtained by extremizing the action with respect to independent variables of the system. The variation with respect to the Lagrange multiplier λ imposes the condition that u^a is a unit timelike vector, thus

$$g_{ab}u^a u^b = -1, \quad (7)$$

while the variation of the action with respect u^a, leads to

$$\nabla_a K^{am}_{bn}\nabla_m u^n + c_4 u^m \nabla_m u_a \nabla_b u^a + \lambda u_b = 0, \quad (8)$$

the variation of the action with respect to the metric g_{mn} gives the dynamical equations,

$$G^{Einstein}_{ab} = T^{aether}_{ab} + 8\pi G T^{matter}_{ab}, \quad (9)$$

where

$$\begin{aligned} G^{Einstein}_{ab} &= R_{ab} - \frac{1}{2}g_{ab}R, \\ T^{aether}_{ab} &= \nabla_c[J^c{}_{(a}u_{b)} + u^c J_{(ab)} - J_{(a}{}^c u_{b)}] - \frac{1}{2}g_{ab}J^c_d\nabla_c u^d + \lambda u_a u_b \\ &\quad + c_1[\nabla_a u_c \nabla_b u^c - \nabla^c u_a \nabla_c u_b] + c_4 u^c \nabla_c u_a u^d \nabla_d u_b, \\ T^{matter}_{ab} &= \frac{-2}{\sqrt{-g}}\frac{\delta\left(\sqrt{-g}L_{matter}\right)}{\delta g_{ab}}, \end{aligned} \quad (10)$$

with

$$J^a{}_m = K^{ab}{}_{mn}\nabla_b u^n. \tag{11}$$

In a more general situation, the Lagrangian of GR is recovered, if and only if, the coupling constants are identically zero, i.e., $c_1 = c_2 = c_3 = c_4 = 0$.

4. Naked Singularities in Einstein-Aether Theory

In this section, I show that naked singularities appear in EA theory which are otherwise covered by event (Killing) horizons in GR. For this, we start with the most general spherically symmetric static metric,

$$ds^2 = -e^{2A(r)}dt^2 + e^{2B(r)}dr^2 + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2. \tag{12}$$

In accordance with equation (7), the aether vector is taken to be unitary and timelike,

$$u^a = (e^{-A(r)}, 0, 0, 0). \tag{13}$$

This choice is not the most general and is restricted to the scenario where aether is static. The aether must tip in a black hole solution as it cannot be timelike to be aligned with the null Killing vector on the horizon. As that is not the case with our choice, our solutions are valid only outside the Killing horizon. However, this is not going to be problematic for the solutions presented here.

The timelike Killing vector of the metric (12) is giving by

$$\chi^a = (-1, 0, 0, 0). \tag{14}$$

The Killing and the universal horizon[17] are obtained finding the largest root of

$$\chi^a\chi_a = 0, \tag{15}$$

and

$$\chi^a u_a = 0, \tag{16}$$

respectively, where χ^a is the timelike Killing vector. Thus

$$\chi^a\chi_a = -e^{2A(r)}, \tag{17}$$

$$\chi^a u_a = e^{A(r)}. \tag{18}$$

For the metric in Eq. (12), the tt, rr and $\theta\theta$ components of the vacuum field equations (10) are given by,

$$-\frac{e^{2(A-B)}}{2r^2}\left[c_{14}(-2r^2A'B' + r^2A'^2 + 2r^2A'' + 4rA') - 4rB' - 2e^{2B} + 2\right] = 0, \tag{19}$$

$$-\frac{1}{2r^2}\left(c_{14}r^2A'^2 + 4rA' - 2e^{2B} + 2\right) = 0, \tag{20}$$

$$\frac{r}{2e^{2B}}\left(2rA'' - 2rA'B' + 2A' - 2B' - (c_{14} - 2)rA'^2\right) = 0. \tag{21}$$

The Kretschmann scalar for the metric (12), is given by

$$K = \frac{4}{r^4 e^{4B}}\left(2B'^2r^2 + e^{4B} - 2e^{2B} + 1 + 2A'^2r^2 + r^4A''^2 + 2r^4A''A'^2 - 2r^4A''B'A' + r^4A'^4 - 2r^4A'^3B' + r^4B'^2A'^2\right). \quad (22)$$

It is not possible to solve these equations analytically for a general case. However, there exist closed form solutions for $c_{14} = 16/9$ and $c_{14} = 48/25$, both of which are within the permitted range. More details about the method employed to obtain these solutions are found in our papers.[18,19]

For $c_{14} = 16/9$, the metric functions and Kretschmann scalar are given by,

$$A = -\frac{3}{4}\ln(r) + \frac{3}{4}\ln\left[r + 4 + \sqrt{(r+8)r}\right], \quad (23)$$

$$B = -\frac{1}{2}\ln(2) + \frac{1}{2}\ln\left(\sqrt{\frac{r}{r+8}} + \frac{r}{r+8}\right), \quad (24)$$

$$K = \frac{768\sqrt{r+8}\left[2r(r+8)^2 + (384 + 104r + 7r^2)\sqrt{(r+8)r}\right]}{\left[r^{11/2}\left(r + 8 + \sqrt{(r+8)r}\right)^4\right]}. \quad (25)$$

From the above equation, we can see that there exists a curvature singularity at $r = 0$. The Killing and universal horizons for this case are obtained by finding the roots of the following,

$$\chi^a\chi_a = -r^{-\frac{3}{2}}\left(r + 4 + \sqrt{(r+8)r}\right)^{\frac{3}{2}} = 0, \quad (26)$$

$$\chi^a u_a = r^{-\frac{3}{4}}\left(r + 4 + \sqrt{(r+8)r}\right)^{\frac{3}{4}} = 0. \quad (27)$$

We can see easily that these equations do not have any real root. Thus, there exists neither Killing horizon nor universal horizon.

For $c_{14} = 48/25$, we see the same behavior as in the previous case, but the mathematical expressions are quite lengthy. So, instead of presenting them here, I refer the reader to our paper.[18]

5. Cosmological Singularities in Einstein-Aether Theory

In this section, I show that initial singularities appear in EA theory which are otherwise absent in GR. For this, we start with the most general isotropic and homogeneous universe is described by a Friedmann-Lemaître-Robertson-Walker (FLRW) metric,

$$ds^2 = -dt^2 + B(t)^2\left[\frac{dr^2}{1-kr^2} + r^2d\theta^2 + r^2\sin^2\theta d\phi^2\right], \quad (28)$$

where, $B(t)$ is the scale factor and k is a Gaussian curvature of the spacial slice at a given time. The appropriate aether vector for this metric is given by,

$$u^a = (1, 0, 0, 0). \tag{29}$$

The standard definitions of the Hubble parameter $H(t)$ and the deceleration parameter $q(t)$ are given by,

$$H(t) = \frac{\dot{B(t)}}{B(t)}, \tag{30}$$

$$q(t) = -\frac{\ddot{B(t)}B(t)}{\dot{B(t)}^2}, \tag{31}$$

where the symbol dot denotes the differentiation with respect to the time coordinate. The Friedmann-Lemaître equations are given by,

$$\left(1 + \frac{\beta}{2}\right)\left(\frac{\dot{B(t)}}{B(t)}\right)^2 = \frac{\Lambda}{3} - \frac{k}{B(t)^2}, \tag{32}$$

$$\left(1 + \frac{\beta}{2}\right)\frac{\ddot{B(t)}}{B(t)} = \frac{\Lambda}{3}, \tag{33}$$

where Λ is the cosmological constant. The Kretschmann scalar for FLRW metric is given by,

$$K = \frac{12}{B^4}\left[k^2 + 2k\dot{B}(t)^2 + \dot{B}(t)^4 + \ddot{B}(t)^2 B(t)^2\right]. \tag{34}$$

We can solve the above equations for nine different combinations of Λ and k. However, here I present only the three cases involving $k = -1$, because these solutions are non-singular in GR but are singular in EA theory. The theoretical consistency requires $\beta \equiv c_1 + 3c_2 + c_3 > -2$. A detailed discussion on other cases can be found in our article.[20]

For $\Lambda > 0, k = -1$, the Friedmann-Lemaître equations have two solutions that exist for $\beta + 2 > 0$,

$$B_1(t) = \frac{1}{2\sqrt{\Lambda}}\left[-\frac{3\, e^{\sqrt{\frac{2\Lambda}{3(\beta+2)}}(t_0 - t)}}{\epsilon\sqrt{\Lambda+3} + \sqrt{\Lambda}} + \frac{\epsilon\sqrt{\Lambda+3} + \sqrt{\Lambda}}{e^{\sqrt{\frac{2\Lambda}{3(\beta+2)}}(t_0 - t)}}\right], \tag{35}$$

$$B_2(t) = \frac{1}{2\sqrt{\Lambda}}\left[-\frac{3\, e^{\sqrt{\frac{2\Lambda}{3(\beta+2)}}(t - t_0)}}{\epsilon\sqrt{\Lambda+3} + \sqrt{\Lambda}} + \frac{\epsilon\sqrt{\Lambda+3} + \sqrt{\Lambda}}{e^{\sqrt{\frac{2\Lambda}{3(\beta+2)}}(t - t_0)}}\right]. \tag{36}$$

Both are singular at,

$$t_{sing}(B_1) \; = t_0 - \sqrt{\frac{3(\beta+2)}{8\Lambda}}\ln\left[\frac{2}{3}\Lambda + \frac{2}{3}\epsilon\sqrt{\Lambda(\Lambda+3)} + 1\right], \tag{37}$$

$$t_{sing}(B_2) \; = t_0 + \sqrt{\frac{3(\beta+2)}{8\Lambda}}\ln\left[\frac{2}{3}\Lambda + \frac{2}{3}\epsilon\sqrt{\Lambda(\Lambda+3)} + 1\right]. \tag{38}$$

The metric is not singular for $\beta = 0$ as the curvature invariant being $\frac{8\Lambda^2}{3}$. But, the metric is singular for $\beta + 2 > 0$ with $\beta \neq 0$. This means it exists only in EA theory but not in GR.

For $\Lambda = 0, k = -1$, we have two solutions satisfying $\beta + 2 > 0$,

$$B_1(t) = 1 + \sqrt{\frac{2}{2+\beta}}(t_0 - t) \tag{39}$$

$$B_2(t) = 1 + \sqrt{\frac{2}{2+\beta}}(t - t_0) \tag{40}$$

Both are singular at

$$t_{sing}(B_1) = t_0 + \sqrt{\frac{2+\beta}{2}}, \tag{41}$$

$$t_{sing}(B_2) = t_0 - \sqrt{\frac{2+\beta}{2}}. \tag{42}$$

For $\beta = 0$ which corresponds to GR, the solution exists and is never singular with the curvature invariant being zero. But, the metric is singular for $\beta + 2 > 0$ such that $\beta \neq 0$. This means it is a new singularity that exists only in EA theory but not GR.

For $\Lambda < 0, k = -1$, there exists two solutions satisfying $\beta + 2 > 0$,

$$B_1(t) = \sqrt{\frac{3}{|\Lambda|}} \sin\left(\sqrt{\frac{2|\Lambda|}{3(\beta+2)}}(t - t_0) + \sin^{-1}\sqrt{\frac{|\Lambda|}{3}}\right) \tag{43}$$

$$B_2(t) = \sqrt{\frac{3}{|\Lambda|}} \sin\left(\sqrt{\frac{2|\Lambda|}{3(\beta+2)}}(t_0 - t) + \sin^{-1}\sqrt{\frac{|\Lambda|}{3}}\right) \tag{44}$$

Both are singular at

$$t_{sing}(B_1) = t_0 - \sqrt{\frac{3(\beta+2)}{2|\Lambda|}} \sin^{-1}\sqrt{\frac{|\Lambda|}{3}} \tag{45}$$

$$t_{sing}(B_2) = t_0 + \sqrt{\frac{3(\beta+2)}{2|\Lambda|}} \sin^{-1}\sqrt{\frac{|\Lambda|}{3}} \tag{46}$$

For $\beta = 0$ the solution exists but is never singular with the curvature invariant being $\frac{8|\Lambda|^2}{3}$. But, the metric is singular for $\beta + 2 > 0$ with $\beta \neq 0$. This means it is a new singularity that exists only in EA theory.

The above results are independently confirmed by studying the focusing of congruence of timelike geodesics using the Raychaudhuri equation.[20]

6. Conclusions

Many of the deep conceptual issues related to gravity are closely connected with the very existence spacetime singularities. The singularity theorems predict only the existence of spacetime singularities under a set of physically reasonable conditions, but not say much else about them. In principle, spacetime singularities and event horizons are two totally different and independent concepts in gravitational physics. However, the Cosmic Censorship Conjecture (CCC) suggests that whenever a spacetime singularity occurs it always stays hidden within an event horizon. But, the mathematically rigorous formulation of the CCC, let alone its proof, is still an open problem. We have attempted to understand these issues in VT theories of gravity with the hope that we might look at GR from a fresh perspective.

In this paper, we investigated how the presence of timelike vector field in EA theory affects the nature of singularities in comparison to GR theory. Firstly, I have illustrated with two analytical solutions ($c_{14} = 16/9$ and $48/25$) that the Schwarzschild metric in EA theory is singular at $r = 0$ but it is not covered by neither Killing horizon nor universal horizon, whereas the corresponding case in GR has horizons. Thus we have naked singularities. Astrophysically, a naked singularity is distinguishable from a black hole of the same mass from the the luminosity of the accretion disk as it extends very close to the naked singularity.[21] Secondly, I have shown that for three different cases with $k = -1$, the FLRW vacuum solutions in EA theory are singular, but non-singular in GR. All these solutions are within the experimentally allowed parameter space. Initial singularity is in principle visible. However, from the data we know that our universe has $k = 0$, thus making this result observationally not viable. The important take away is that the new singular solutions show that GR and EA theories have different global structures even for the simple situations.

Acknowledgments

It is my pleasure to thank Roberto Chan and Maria de Fátima Alves da Silva for several useful discussions and collaboration on the works that produced some of the key results presented here. I am grateful to the organizers of MG16, especially Anzong Wang, for the invitation to give this talk. Finally, I would like to dedicate this article to T. Padmanabhan (1957–2021) who has significantly influenced me before and since I got to be in his orbit when I participated in the *IUCAA Vacation Students' Programme* during the summer of 2003.

References

1. E. Witten, Rev. Mod. Phys. **92**, no. 4, 045004 (2020) [arXiv:1901.03928 [hep-th]].
2. R. Penrose, "Singularity theorems," in *Topology and Physics*, Editors, Chen Ning Yang, Mo-Lin Ge, Yang-Hui He (World Scientific, Singapore, 2019) pp. 135 –171.
3. P. S. Joshi, "The Story of Collapsing Stars: Black Holes, Naked Singularities, and the Cosmic Play of Quantum Gravity," (Oxford University Press, USA, 2015).

4. C. M. Will and K. Nordtvedt, Jr., Astrophys. J. **177**, 757 (1972).
5. K. Nordtvedt, Jr. and C. M. Will, Astrophys. J. **177**, 775-792 (1972).
6. R. W. Hellings and K. Nordtvedt, Phys. Rev. D **7**, 3593-3602 (1973).
7. C. M. Will, *Theory and experiment in gravitational physics*, (Cambridge University Press, First Edition, UK, 1981). pp 126–130
8. V. A. Kostelecky and S. Samuel, Phys. Rev. D **40**, 1886-1903 (1989).
9. T. Jacobson and D. Mattingly, Phys. Rev. D **64**, 024028 (2001) [arXiv:gr-qc/0007031 [gr-qc]].
10. D. Blas, O. Pujolas and S. Sibiryakov, JHEP **04**, 018 (2011) [arXiv:1007.3503 [hep-th]].
11. C. M. Will, *Theory and experiment in gravitational physics*, (Cambridge University Press, Second Edition, UK, 2018). pp 118–121
12. A. Petrov, "Introduction to Modified Gravity" (Springer, Switzerland, 2020). pp 37–46
13. J. Oost, M. Bhattacharjee and A. Wang, Gen. Rel. Grav. **50**, no. 10, 124 (2018) [arXiv:1804.01124 [gr-qc]].
14. K. Lin, X. Zhao, C. Zhang, T. Liu, B. Wang, S. Zhang, X. Zhang, W. Zhao, T. Zhu and A. Wang, Phys. Rev. D **99**, no. 2, 023010 (2019) [arXiv:1810.07707 [astro-ph.GA]].
15. C. Zhang, X. Zhao, A. Wang, B. Wang, K. Yagi, N. Yunes, W. Zhao and T. Zhu, Phys. Rev. D **101**, no. 4, 044002 (2020) [erratum: Phys. Rev. D **104**, no. 6, 069905 (2021)] [arXiv:1911.10278 [gr-qc]].
16. J. Oost, S. Mukohyama and A. Wang, Phys. Rev. D **97**, no. 12, 124023 (2018) [arXiv:1802.04303 [gr-qc]].
17. K. Lin, O. Goldoni, M. F. da Silva and A. Wang, Phys. Rev. D **91**, no. 2, 024047 (2015) [arXiv:1410.6678 [gr-qc]].
18. R. Chan, M. F. A. Da Silva and V. H. Satheeshkumar, Eur. Phys. J. C **81**, no. 4, 317 (2021) [arXiv:2003.00227 [gr-qc]].
19. M. Campista, R. Chan, M. F. A. da Silva, O. Goldoni, V. H. Satheeshkumar and J. F. V. da Rocha, Can. J. Phys. **98**, no. 10, 917-928 (2020) [arXiv:1807.07553 [gr-qc]].
20. R. Chan, M. F. A. da Silva and V. H. Satheeshkumar, JCAP **05**, 025 (2020) [arXiv:1912.12845 [gr-qc]].
21. P. S. Joshi, D. Malafarina and R. Narayan, Class. Quant. Grav. **31**, 015002 (2014) [arXiv:1304.7331 [gr-qc]].

Hořava-Lifshitz and Einstein-Æther gravity in the light of Event Horizon Telescope observations of M87*

Emmanuel N. Saridakis

National Observatory of Athens, Lofos Nymfon, 11852 Athens, Greece
E-mail: msaridak@noa.gr

We investigate Hořava-Lifshitz and Einstein-Æther gravity in light of the Event Horizon Telescope (EHT) observations of the M87*. We calculate the corresponding photon effective potential, the unstable photon sphere radius, and finally the induced angular size, which combined with the mass and the distance can lead to a single prediction that quantifies the black hole shadow, namely the diameter per unit mass d. Since d_{M87*} is observationally known from the EHT Probe, we extract the corresponding parameter regions in order to obtain consistency. We find that Einstein-Æther black hole solutions agree with the shadow size of EHT M87*, if the involved Æther parameters are restricted within specific ranges, along with an upper bound on the dimensionless spin parameter a, which is verified by a full scan of the parameter space within 1σ-error.

Keywords: M87* observations; Event Horizon Telescope; Einstein-Æther gravity; Hořava-Lifshitz gravity.

1. Introduction

In the light of recording the first stunning new radio images of the supermassive black hole (BH) that exists at the center of nearby galaxy M87* by *"Event Horizon Telescope"* (EHT),[1,2] the black-hole shadow has become one of the most exciting events in observational astronomy. Historically the concept of the black hole shadow comes from the 70's with the seminal works.[3] In these studies it was found that the non-rotating BH has a perfect circular shadow, while by taking the rotation into account the shape of shadow is elongated due to the dragging effect.[4]

The dark area over a brighter background in the center of the image is termed as *"BH's shadow"*, and it is predicted by general relativity as the null geodesics in the strong gravity area.[5,6] In particular, photons with some critical angular momentum, will swirl around the BH one ring by one ring which makes the unstable boundary of the shadow, which is the same as the shining halo observed in EHT image. Nevertheless, given that the shape and size of the shadow contains traces of the geometry vicinity of the BH, and similarly to the quasinormal modes approach, one can consider the shadow as a potential probe to investigate the BH structure within different gravitational theories.

Horava-Lifshitz[7–10] and Einstein-Æther (EA) theory[11–14] are generally covariant theories of gravity, which violate the Lorentz invariance locally, and the latter possesses a dynamical, unit-norm and timelike vector field, called *"æther field"*, which defines a preferred timelike direction at each spacetime point. EA gravity includes a number of coupling constants called æther parameters. In the present work we

are interested in investigating Horava-Lifshitz and Einstein-Æther theory using the EHT observations of the M87*, based on.[15] Throughout the manuscript we adopt natural units where $\hbar = k_B = c = 1$.

2. Rotating black-hole solutions in Einstein-Æther gravity

The action of the EA theory is[11]

$$S = S_{EH} + S_{AE} = \frac{1}{16\pi G_{AE}} \int d^4x\sqrt{-g}\Big(R + \mathcal{L}_{AE}\Big), \tag{1}$$

which includes the standard Einstein-Hilbert action S_{EH} plus the æther action S_{AE}. In the above action, R, G_{EA} and $\mathcal{L}_{AE}$ refer respectively to the Ricci scalar, the Æther gravitational constant and the Lagrangian of the Æther field u^μ, which is defined as

$$\mathcal{L}_{AE} \equiv -\left(c_1 g^{\alpha\beta} g_{\mu\nu} + c_2 \delta^\alpha_\mu \delta^\beta_\nu + c_3 \delta^\alpha_\nu \delta^\beta_\mu - c_4 u^\alpha u^\beta g_{\mu\nu}\right)(\nabla_\alpha u^\mu)(\nabla_\beta u^\nu) + \lambda_0(u^2 + 1). \tag{2}$$

Here, λ_0 is a Lagrangian multiplier, ensuring that the Æther four-velocity u^α is always timelike (i.e. $u^2 = -1$). All of four coupling constants (c_1, c_2, c_3, c_4) in the above expression are dimensionless, and thus G_{AE} is linked to the Newtonian constant G_N via two of them, namely $G_{AE} = \frac{2G_N}{2-c_1-c_4}$.[16] As usual, variations of the total action with respect to $g_{\mu\nu}$, u^α, λ_0 yield, respectively, the field equations.[15]

There exists a corresponding slowly rotating black hole solution, with a spherically symmetric æther field configuration ($\lambda(r) = 0$), namely[15]

$$ds^2 = -e(r)dt^2 + \frac{dr^2}{e(r)} + r^2(d\theta^2 + \sin^2\theta d\phi^2) - \frac{4Ma}{r}\sin^2\theta dt d\phi, \tag{3}$$

with

$$e(r) = 1 - \frac{2M}{r} - \frac{27c_{13}}{256(1-c_{13})}\left(\frac{2M}{r}\right)^4. \tag{4}$$

Note that for convenience we have replaced the angular momentum J by introducing the rotation parameter[17] a through $a \equiv \frac{J}{M}$.

3. Black hole shadow

As usual, in order to study the geodesics structure of the photon trajectories, we begin with the Hamilton-Jacobi equation

$$\frac{\partial S}{\partial \lambda} = -\frac{1}{2} g^{\mu\nu} \frac{\partial S}{\partial x^\mu} \frac{\partial S}{\partial x^\nu}\ , \tag{5}$$

where S and λ denote respectively the Jacobi action of the particle (here photon) moving in the black hole spacetime, and the affine parameter of the null geodesic. Concerning the massless photon propagating on the null geodesics, the Jacobi action S can be separated as

$$S = -Et + J\phi + S_r(r) + S_\theta(\theta)\ , \tag{6}$$

where E and J respectively address the energy and angular momentum of the photon in the direction of the rotation axis. Furthermore, the functions $S_r(r)$ and $S_\theta(\theta)$ have only r and θ dependencies, respectively.

By inserting the Jacobi action (6) into the Hamilton-Jacobi equation (5), using also the metric components (3), we find that the photon trajectories are expressed in terms of the radial geodesics and the effective potential $U_{eff}(r)$ as:[15]

$$\left(\frac{dr}{d\lambda}\right)^2 + U_{eff}(r) = 0 \tag{7}$$

with

$$U_{eff}(r) = -1 + \frac{e(r)}{r^2}(\xi^2 + \eta) + \frac{4M\eta a}{r^3}, \tag{8}$$

where $\xi = \frac{J}{E}, \quad \eta = \frac{\mathcal{K}}{E^2}$. The above two impact parameters ξ and η are actually the principle quantities for determining the photon motion. We can extract the geometric shape of the shadow via the allowed values of ξ and η that satisfy the above conditions. Thus, we arrive at

$$2(\eta + \xi^2)re(r) - (\eta + \xi^2)r^2e'(r) + 12M\xi a = 0\ . \tag{9}$$

By solving this equation one acquires the radius r_{ps} of the photon sphere, which since we have taken the rotation effect into account is expected to be between the two values $r_{ps}^{\mp}$.

For slowly rotating BHs we immediately find that for the spherical-orbit photon motion the two parameters ξ and η have the form

$$\xi(r) = \frac{r^3[re'(r) - 2\,e(r)]}{4Ma[e(r) + re'(r)]}\ , \tag{10}$$

$$\eta(r) = \frac{-r^6[-2e(r) + re'(r)]^2 + 48M^2a^2r^2\left[e(r) + re'(r)\right]}{16M^2a^2\left[e(r) + e'(r)\right]^2}\ . \tag{11}$$

As usual, to describe the shadow as seen by a distant observer, one introduces the following two celestial coordinates X and Y:[6]

$$X = \lim_{r_*\to\infty}\left(-r_*^2\sin\theta_0\frac{d\phi}{dr}\right), \tag{12}$$

$$Y = \lim_{r_*\to\infty} r_*^2\frac{d\theta}{dr}, \tag{13}$$

where r_* and θ_0 are respectively the distance between the observer and the black hole, and the inclination angle between the line of sight of the observer and the rotational axis of the black hole. By applying the geodesics equations along with the expressions $\frac{d\phi}{dr} = \frac{d\phi/d\lambda}{dr/d\lambda}$ and $\frac{d\theta}{dr} = \frac{d\theta/d\lambda}{dr/d\lambda}$ we obtain

$$X = -\xi(r_{ps})\csc\theta_0\ , \tag{14}$$

$$Y = \sqrt{\eta(r_{ps}) - \xi^2(r_{ps})\cot^2\theta_0}\ , \tag{15}$$

and therefore these two celestial coordinates fulfill

$$X^2 + Y^2 = \xi^2(r_{ps}) + \eta_{\mathbf{ps}}(r_{ps}) \ , \tag{16}$$

where r_{ps} is the aforementioned radius of the unstable photon sphere. This is the expression of the EA BHs shadow in the slow rotation limit. As a self-consistency check, we can see that in the Schwarzschild limit, namely $a \to 0$, the above equation becomes $X^2 + Y^2 = 27M^2$, as expected.

3.1. *Einstein-Æther black hole solution*

In order to draw the shadow of the BH solution we need to calculate two essential quantities, namely the event horizon radius r_e and the radius of the unstable photon sphere r_{ps}. In order to find the even horizon of metric (3) we have to solve

$$r^4 - 2Mr^3 - \frac{27c_{13}}{16(1-c_{13})}M^4 = 0 \ , \tag{17}$$

where $c_{13} \equiv c_1 + c_3$. This leads to

$$r_{e_{1,2}} = \frac{3M^2}{2} - s \pm \frac{1}{2}\sqrt{-4s^2 + 3M^2 - \frac{M^2}{s}}, \tag{18}$$

$$r_{e_{3,4}} = \frac{3M^2}{2} + s \pm \frac{1}{2}\sqrt{-4s^2 + 3M^2 + \frac{M^2}{s}}, \tag{19}$$

where

$$s = \sqrt{\frac{3M^2Q + Q^2 + \Delta_0}{12Q}}, \quad Q = \left(\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}\right)^{1/3}, \tag{20}$$

$$\Delta_0 = -\frac{81c_{13}}{4(1-c_{13})}M^4, \quad \Delta_1 = \frac{729c_{13}}{2(1-c_{13})}M^7. \tag{21}$$

Solutions $r_{1,2}$ are imaginary and thus not physically interesting. Nevertheless, by setting c_{13} to zero the third solution becomes $r_3 = 2M$, as expected from Schwarzschild background. Thus, we deduce that r_3 addresses the event horizon radius r_e of the Einstein-Æther BH solution.

We proceed by setting $\eta = 0$, and inserting the ξ_{ps} from $U_{eff} = 0$

$$\xi_{ps} = \frac{2Ma}{e(r_{ps})r_{ps}}\left[-1 + \sqrt{1 + \frac{r_{ps}^4 e(r_{ps})}{4M^2a^2}}\right] \ , \tag{22}$$

into (9). Then, using $e(r)$ from Eq. (4) we finally result to the following equation

$$\begin{aligned} a^2\,(\alpha_1 - 8)\,(c_{13} - 1)\,r_{ps}^3\,\big[&81\alpha_1 c_{13} + 16\,(\alpha_1 + 16)\,(c_{13} - 1)\,r_{ps}^4 \\ &-48\,(\alpha_1 + 8)\,(c_{13} - 1)\,r_{ps}^3\big] = 0 \ , \end{aligned} \tag{23}$$

with

$$\alpha_1 = \sqrt{\frac{64a^2\,(c_{13} - 1)\,(r_{ps} - 2)r_{ps}^3 + 16\,(c_{13} - 1) + 27c_{13}}{(c_{13} - 1)a^2}} \ . \tag{24}$$

The solution of Eq. (23) will provide the radius of the photon sphere r_{ps}.

4. Confrontation with and M87* observations

We can now proceed to the investigation of the constraints on the æther parameter c_{13} that arise from the EHT Observations of the shadow of M87*. In particular, the size of the resulting shadow is sensitive on the æther parameter c_{13}, and hence the image of M87* is able to impose bounds on them.

In light of the report released by EHT collaboration for the shadow of M87* in,[1,2] for the angular size of the shadow, the mass and the distance to M87* one respectively has the values

$$\delta = (42 \pm 3)\,\mu\text{arcsec}, \tag{25}$$

$$M = (6.5 \pm 0.9) \times 10^9\, M_\odot, \tag{26}$$

$$D = 16.8^{+0.8}_{-0.7}\ \text{Mpc}, \tag{27}$$

where $M_\odot$ is the Sun mass. One can merge this information by introducing the single number d_{M87*}, which quantifies the size of M87*'s shadow in unit mass, defined as[17]

$$d_{M87*} \equiv \frac{D\delta}{M} \approx 11.0 \pm 1.5\,. \tag{28}$$

This combination can be used in order to confront with the theoretically predicted shadows (e.g. the above number is in agreement within 1σ-error with what we theoretically expect for the Schwarzschild BH, namely $d_{Schw} \simeq 10.4$[6]).

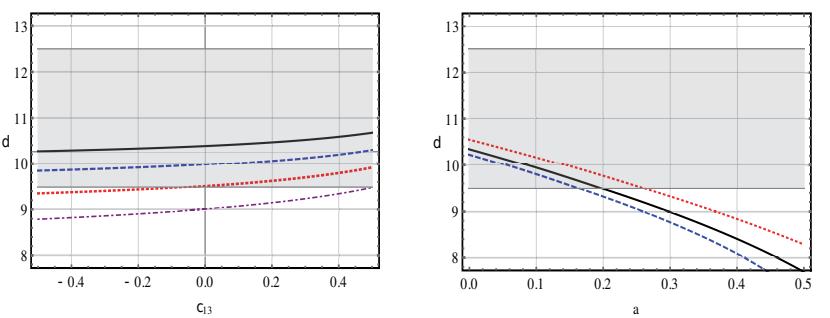


Fig. 1. **Left graph:** The predicted diameter per unit mass d for the Einstein-Æther black hole solution (3) with (4), as a function of the æther parameter c_{13}, for several values of the rotational parameter: $a = 0$ (black - solid), $a = 0.1$ (blue - dashed), $a = 0.2$ (red - dotted), $a = 0.3$ (purple - dashed-dotted). **Right graph:** d as a function of the rotational parameter a, for $c_{13} = -0.5$ (blue - dashed), $c_{13} = 0$ (black- solid), $c_{13} = 0.5$ (red- dotted). In both graphs the shaded area mark the observationally determined diameter per unit mass of M87*'s shadow, namely d_{M87*}, within 1σ-error.

In order to confront Einstein-Æther solutions with the above observational number, all we need is to calculate the angular size δ as a function of the Einstein-Æther parameters. The angular size can be immediately extracted from the profile (16), as long as we know the radius of the unstable photon sphere r_{ps} (as a function of

the Einstein-Æther parameter). Knowing δ can then easily lead to the predicted diameter per unit mass d. Note that in agreement with EHT, from now on we fix the value $\theta_0 = 17$ for the inclination angle.

In Fig. 1 we depict the diameter of the predicted shadow d as a function of c_{13} and a, on top of the observed one from M87* within 1σ. As we observe, the Æther BH solution is able to quantitatively describe the shadow size of M87*, provided that the dimensionless spin parameter a is constrained to specific values, dependent on c_{13}. Note that since our whole analysis has been restricted to the slow rotation case, the extracted upper bounds on a are perfectly consistent and justify our approximation. The curves in Fig. 1 imply that within these ranges of c_{13} it is possible to distinguish a naked singularity ($c_{13} < 0$) from BH ($c_{13} > 0$). Furthermore, in order to present the bounds on spin parameter a in a more transparent way, we perform a full scan of the two-dimensional parameter space and in Fig. 2 we depict the region which corresponds to a d in agreement with the observed value d_{M87*}. As it is clear from it as well as from the right panel in Fig. 1, by moving from negative to positive values of æther parameter c_{13}, the upper bound on the a grows. This may be useful in distinguishing between BH and naked singularity.

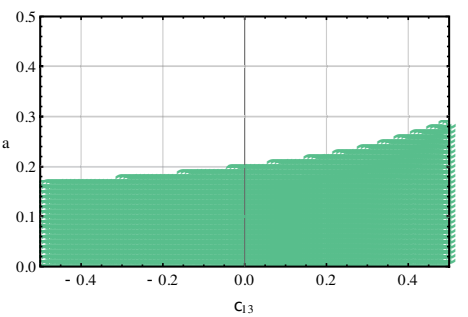


Fig. 2. The allowed parameter region (green) in the $c_{13} - a$ plane, for fixed BH mass ($M = 6 \times 10^9 M_\odot$), for the Einstein-Æther black hole solution (3) with (4), that leads to diameter per unit mass d in agreement with the observationally determined one d_{M87*} within 1σ-error.

We mention here that in order to create the allowed parameter space plots on the $c_{13} - a$ plane, we have fixed the BH mass to $M = 6 \times 10^9 M_\odot$. Nevertheless, one can also perform a similar parameter scan for the case of non-fixed BH mass, namely within the two-dimensional parameter space (c_{13}, M), for different fixed values of the rotation parameter a. In Fig. 3 we present such a graph. As we observe, for reasonable c_{13} values, the BH mass lies within the range we expect from EHT for M87*, i.e. $M = (6.5 \pm 0.9) \times 10^9\, M_\odot$. Furthermore, Fig. 3 shows the role of rotation on the allowed region of the $c_{13} - M$ plane, namely by increasing the rotation parameter we are led to larger region for the c_{13} parameter and still be consistent with EHT data within 1-σ error. In summary, Einstein-Æther BH solution is in agreement with M87* observation, and moreover we obtain a restriction to low rotational values.

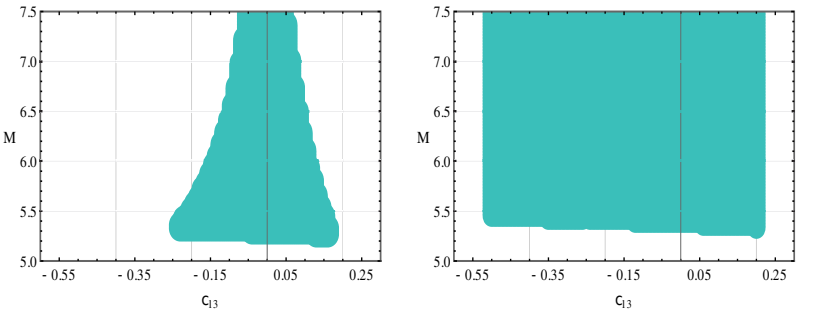


Fig. 3. The allowed parameter region (green) in the $c_{13} - M(10^9 M_{\odot})$ plane, for fixed rotation parameter values $a = 0$ (left graph) and $a = 0.2$ (right graph), for the Einstein-Æther black hole solution (3) with (4), that leads to diameter d in agreement with the observationally determined one d_{M87*} within 1σ-error.

5. Conclusions

In the present work we investigated Hořava-Lifshitz and Einstein-Æther gravity in light of the recent Event Horizon Telescope (EHT) observations of the M87*. In particular, the EHT provided the first visual evidence indicating directly the existence of a compact object such as a supermassive black hole candidate at the center of the M87* galaxy. Since the shape and size of the observed BH shadow contains information of the geometry in the vicinity of the BH, one can consider the shadow as a potential probe to investigate the BH structure within different gravitational theories, due to the fact that in modified gravity one obtain black holes which deviate from those of general relativity.[18] Hence, allowing for a modified gravity as the underlying theory enriches the calculation framework with different size-rotation features as well as with extra model parameters.

We first extracted the black hole solutions for EA gravity, and we calculated the corresponding effective potential $U_{eff}(r)$ for the photons, the resulting event horizon radius r_e and the radius of the unstable photon sphere r_{ps}. Then we straightforwardly calculated the induced angular size δ, which combined with the mass and the distance can lead to a single prediction that quantifies the black hole shadow size, namely the diameter per unit mass d. Since d_{M87*} is observationally known from the EHT Probe, we extracted the corresponding parameter regions of Einstein-Æther theory in order to obtain consistency.

Apart from the restriction of the æther parameter c_{13} to specific ranges, we found some constraints on the dimensionless spin parameter a in slow rotation limit, dependent on the values of æther parameters, which was verified by a full scan of the parameter space. Furthermore, our analysis has indicated that in the Einstein-Æther theory is possible to distinguish a naked singularity from BH. This is indeed a verification of the fact that in modified gravity one obtains different size-rotation features, depending on the extra model parameters, and hence the M87* observation can be in principle described in different ways comparing to general relativity,

especially in theories which possess various corrections on the Kerr-metric solutions, such is the case in Hořava-Lifshitz and Einstein-Æther gravity. In summary, Einstein-Æther black hole solutions are in agreement with EHT M87* observation.

References

1. K. Akiyama *et al.* [Event Horizon Telescope Collaboration], *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole*, Astrophys. J. Lett. **875**, L1 (2019).
2. K. Akiyama *et al.* [Event Horizon Telescope Collaboration], *First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole*, Astrophys. J. Lett. **875**, no. 1, L6 (2019).
3. J. L. Synge, *The Escape of Photons from Gravitationally Intense Stars*, Mon. Not. Roy. Astron. Soc. **131**, no. 3, 463 (1966).
4. R. Narayan, M. D. Johnson and C. F. Gammie, *The Shadow of a Spherically Accreting Black Hole*, Astrophys. J. Lett. **885**, no. 2, L33 (2019).
5. P. V. P. Cunha and C. A. R. Herdeiro, *Shadows and strong gravitational lensing: a brief review*, Gen. Rel. Grav. **50**, no. 4, 42 (2018).
6. S. Chandrasekhar, *The mathematical theory of black holes*, Oxford classic texts in the physical sciences, Oxford Univ. Press, Oxford (2002).
7. P. Horava, *Quantum Gravity at a Lifshitz Point*, Phys. Rev. D **79**, 084008 (2009).
8. E. N. Saridakis, *Horava-Lifshitz Dark Energy*, Eur. Phys. J. C **67**, 229-235 (2010).
9. S. Dutta and E. N. Saridakis, *Observational constraints on Horava-Lifshitz cosmology*, JCAP **01**, 013 (2010).
10. R. Garattini and E. N. Saridakis, *Gravity's Rainbow: A bridge towards Hořava–Lifshitz gravity*, Eur. Phys. J. C **75**, no. 7, 343 (2015).
11. T. Jacobson and D. Mattingly, *Gravity with a dynamical preferred frame*, Phys. Rev. D **64**, 024028 (2001).
12. T. Jacobson, *Extended Horava gravity and Einstein-aether theory*, Phys. Rev. D **81**, 101502 (2010) [erratum: Phys. Rev. D **82**, 129901 (2010)].
13. K. Yagi, D. Blas, E. Barausse and N. Yunes, *Constraints on Einstein-Æther theory and Hořava gravity from binary pulsar observations*, Phys. Rev. D **89**, no. 8, 084067 (2014) [erratum: Phys. Rev. D **90**, no. 6, 069902 (2014); erratum: Phys. Rev. D **90**, no. 6, 069901 (2014)].
14. C. Eling and T. Jacobson, *Spherical solutions in Einstein-aether theory: Static aether and stars*, Class. Quant. Grav. **23**, 5625-5642 (2006) [erratum: Class. Quant. Grav. **27**, 049801 (2010)].
15. M. Khodadi and E. N. Saridakis, *Einstein-Æther gravity in the light of event horizon telescope observations of M87**, Phys. Dark Univ. **32**, 100835 (2021).
16. S. M. Carroll and E. A. Lim, *Lorentz-violating vector fields slow the universe down*, Phys. Rev. D **70**, 123525 (2004).
17. C. Bambi, K. Freese, S. Vagnozzi and L. Visinelli, *Testing the rotational nature of the supermassive object M87* from the circularity and size of its first image*, Phys. Rev. D **100**, no. 4, 044057 (2019).
18. E. N. Saridakis *et al.* [CANTATA], *Modified Gravity and Cosmology: An Update by the CANTATA Network*, [arXiv:2105.12582 [gr-qc]].

Hořava-Lifshitz gravity in $(3+1)$ dimensions coupled with anisotropic matter and possible constraints from GRB 170817A

Tao Zhang[1,2,3] and Fu-Wen Shu[2,3,4,*]

[1] *School of Electrical Engineering, Jiujiang Vocational and Technical College, Jiujiang, Jiangxi 332007, China*

[2] *Department of Physics, Nanchang University, Nanchang, 330031, China*

[3] *Center for Relativistic Astrophysics and High Energy Physics, Nanchang University, Nanchang 330031, China*

[4] *GCAP-CASPER, Physics Department, Baylor University, Waco, TX 76798-7316, USA*

[*] *E-mail: shufuwen@ncu.edu.cn*

This paper is a summary of a talk given in the proceedings of Sixteenth Marcel Grossmann Meeting (MG16). We study a (3+1)-dimensional Hořava-Lifshitz (HL) gravity which coupling with an anisotropic electromagnetic (EM) field which is obtained from a (4+1)-dimensional HL gravity. The model has a noticeable feature that gravitational and electromagnetic waves have the same velocity in the Minkowski background and the Friedman-Robertson-Walker (FRW) background. Based on this characteristic we put forward a new way to restrict the parameter of the HL gravity by considering the possible Lorentz violation effect of the GRB 170817A. It turns out that in this way we can place a stringent constraint on the parameter of (4+1)-dimensional HL gravity.

Keywords: Hořava-Lifshitz gravity; Lorentz violation effect; GRB 170817A.

1. Introduction

The first gravitational wave (GW) event from a binary neutron star merger, GW170817, was detected by LIGO and Virgo Collaboration,[1] opened the era of multi-messenger astronomy, which create an opportunity to study gravity and cosmology by multiple observations. At the same time, its EM counterpart—GRB 170817A— was observed by Fermi-GBM with a time delay $\Delta t = (1.74 \pm 0.05)s$.[2] Analyzing the time delay between GW170817 and GRB 170817A, one can impose a stringent constraint on the speed deviation between GWs and the GRBs. This in turn may put restrictions on certain models of modified gravity. However, due to the lack of more similar GW events and the way of direct detection of the speed of GWs, it is necessary to develop new methods to restrict the GWs velocity. In view of this, we turn to consider the GW's counterpart: the gamma ray bursts (GRBs). It is of great interest to investigate whether we can limit the velocity of GWs from the GRB data. If this is possible, it may provide an independent approach in restricting the parameter of some modified models of gravity.

On the other hand, Lorentz invariance has been confirmed rigorously in low energy and the light speed is a constancy in General Relativity (GR). However, the viewpoint that the possible Lorentz violation (LV) will happen at energies comparable with the Planck scale ($E_{pl} \approx 1.22 \times 10^{19}$ GeV) has been proposed in many

quantum gravity theories. This possible LV effect may have been observed in some high-energy astrophysical phenomena such as gamma-ray bursts, active galactic nuclei (AGN) and Crab pulsars, etc., and the LV energy scale can be estimated by analyzing the observed time delay or spectra delay of the gamma rays.[3] Then we wonder if the possible LV is common in these high-energy astrophysical phenomena, the GRB 170817A would have the similar LV behavior. It can be then used to study GWs and restrict the parameter of some modified theories of gravity.

For this reason, in this work we consider a special modified model of gravity: a (3+1)-dimensional HL gravity couple with the electromagnetic field which is obtained from a $(4+1)-$dimensions HL gravity through a Kaluza-Klein (KK) reduction. The wave equations for gravitational and EM fields can be obtained by linear perturbations. One of the most significant characteristics of this model is that the speeds of gravitational and EM waves are the same, no matter it is in the Minkowski background or the Friedman-Robertson-Walker (FRW) background. Based on this feature, we can impose restrictions on the propagation velocity of GWs by considering the possible LV effect of the accompanied GRB, which in turn can be applied to limit the parameter of some gravity theories.

2. The (4+1)-dimensional Hořava-Lifshitz gravity

We begin with a (4+1)-dimensional HL gravity. The action of the non-projectable HL gravity in (4+1) dimensions is,[4]

$$S(g_{\mu\nu}, N_\rho, N) = \int dt d^4x N\sqrt{g}\Big[K_{\mu\nu}K^{\mu\nu} - \lambda K^2 + \beta\,{}^{(4)}R + \alpha a_\mu a^\mu + V(g_{\mu\nu}, N)\Big], \quad (1)$$

where μ, ν=1, 2, 3, 4. $V(g_{\mu\nu}, N)$ expresses the higher terms, we can ignore these terms because the most energetic particles of the GRB events are much less than the HL energy scale.

The Hamiltonian density $\mathcal{H}$ can be expressed as

$$\mathcal{H} = \sqrt{g}N\left[\frac{\pi_{\mu\nu}\pi^{\mu\nu}}{g} + \frac{\lambda}{1-4\lambda}\frac{\pi^2}{g} - \beta\,{}^{(4)}R - \alpha a_\mu a^\mu\right] + 2\pi^{\mu\nu}\nabla_\mu N_\nu + \sigma P_N, \quad (2)$$

where $\pi^{\mu\nu}$ is the conjugate momentum to $g_{\mu\nu}$ and we have added the the primary constraints by multiplying the Lagrange multipliers σ and N_ν.

In Eq.(2), we notice that $\lambda = 1/4$ is a kinetic conformal point which will eliminate a scalar mode, but it has no influence on gravity and electromagnetic modes which we are interest. In order to analyze the couplings between gravity modes and electromagnetic modes, we can perform a KK reduction and reduce the theory to (3+1) dimensions. The specific approach is to decompose the 4-dimensional Riemannian metric $g_{\mu\nu}$ in the following way[4]

$$g_{\mu\nu} = \begin{pmatrix} \gamma_{ij} + \phi A_i A_j & \phi A_j \\ \phi A_i & \phi \end{pmatrix}, \quad (3)$$

where γ_{ij} is a 3-dimensional Riemannian metric, A_i, A_j are the anti-symmetric electromagnetism vectors, and ϕ is a scalar field. After a canonical transformation,

the Hamiltonian density (2) can be rewritten by γ_{ij}, A_i, ϕ and their conjugate momenta p_{ij}, p_i and p. Variations with respect to variables γ_{ij}, A_i, ϕ, p^{ij}, p^i, p generate the equations of motion.[4]

For the scalar field ϕ and its conjugate momenta p, we can get their solutions from the equations of motion and find they are constants, regardless of whether it is in the Minkowski background or in the FRW background. As a consequence, one can assume $\bar{\phi} = 1$ and $\bar{p} = 0$ without loss of generality. Under this condition, for the Minkowski background we have

$$\bar{\gamma}_{ij} = \delta_{ij}, \bar{p}_{ij} = 0, \bar{N} = 1, \bar{A}_i = \bar{p}_i = 0, \bar{N}_i = \bar{N}_4 = 0. \tag{4}$$

and for the FRW background it becomes

$$\bar{\gamma}_{ij} = a(t)^2\delta_{ij},\ \bar{p}_{ij} = 0,\ \bar{p} = -a(t)^2\dot{a}(t),\ \bar{N} = 1,\ \bar{A}_i = \bar{p}_i = 0,\ \bar{N}_i = \bar{N}_4 = 0,\ \bar{\phi} = 1. \tag{5}$$

In order to get the wave equations of gravitational tensor and electromagnetic vectors, we can perform perturbations in both two backgrounds. The perturbations of the Minkowski background are

$$\begin{aligned} &\gamma_{ij} = \delta_{ij} + \epsilon h_{ij},\ p_{ij} = \epsilon\Omega_{ij},\ N_i = \epsilon n_i,\ N_4 = \epsilon n_4, \\ &N = 1 + \epsilon n,\ A_i = \epsilon\xi_i,\ p_i = \epsilon\zeta_i,\ \phi = 1 + \epsilon\tau,\ p = \epsilon\chi, \end{aligned} \tag{6}$$

and for the FRW background we have

$$\begin{aligned} &\gamma_{ij} = a(t)^2(\delta_{ij} + \epsilon h_{ij}),\ p_{ij} = \epsilon\Omega_{ij},\ p = -a(t)^2\dot{a} + \epsilon\chi, \\ &N_i = \epsilon n_i,\ N_4 = \epsilon n_4,\ N = 1 + \epsilon n,\ A_i = \epsilon\xi_i,\ p_i = \epsilon\zeta_i,\ \phi = 1 + \epsilon\tau. \end{aligned} \tag{7}$$

where h_{ij}, Ω_{ij}, n_i, n_4, n, ξ_i, ζ_i, τ, χ are linear perturbations for each field.

Inputting these perturbations into the equations of motion, the equations of motion at the linear order of ϵ can be obtained. Imposing the transverse traceless (TT) gauge conditions of tensors and vectors, and combining the constraint conditions, the wave equations of the EM vectors and the gravitational tensors in the Minkowski background are

$$\ddot{\xi}_i^T - \beta c^2\Delta\xi_i^T = 0, \tag{8}$$

$$\ddot{h}_{ij}^{TT} - \beta c^2\Delta h_{ij}^{TT} = 0. \tag{9}$$

Similarly, the wave equations in the FRW background can be obtained in the same way, they are

$$\ddot{\xi}_i^T + H\dot{\xi}_i^T - \beta c^2\triangle\xi_i^T = 0, \tag{10}$$

$$\ddot{h}_{ij}^{TT} + 3H\dot{h}_{ij}^{TT} - \beta c^2\triangle h_{ij}^{TT} = 0, \tag{11}$$

where $H = \dot{a}(t)/a(t)$ is the Hubble parameter.

From these wave equations, we can find that the propagation velocity of the EM waves and the GWs are the same, that is, $\sqrt{\beta}c$, no matter whether it is in the Minkowski background or in the FRW spacetime. In other words, in this model the

speed of GWs is exactly identified with the one of the EM waves. This remarkable feature allows us to examine possible LV of the model (and in turn place possible constraints on the parameter of the model) by detecting the deviation of the speed of the EM waves from the usual speed of light. In what follows we will present how can one get the deviation of the speed of the EM waves by analyzing the observed GRBs data.

3. Lorentz violation from GRB photons

Considering possible LV, the dispersion relation of a particle with energy E can be phenomenologically written as[3]

$$E^2 = p^2c^2 + m^2c^4 - \Sigma_n s_n E^2 \Big(\frac{E}{E_{LV,n}}\Big)^n, \tag{12}$$

where n is an integer, which corresponds to linear or higher-order dependence of the energy in question. $s_n = \pm 1$ is the sign factor of the LV correction, which indicates that the high-energy photons spread slower than the low-energy photons or vice versa. All GRB data we adopt in this paper have the sign $s_n = 1$. $E_{LV,n}$ is the nth-order LV energy scale which can be obtained from the time delay or the spectra lag of the high-energy particles generated from astrophysical sources such as the GRBs, the active galactic nuclei(AGN) and the Crab pulsars, etc.

In this paper, the GRB 170817A is a photon event which has the mass $m = 0$. Consider only the linear effect $n = 1$, Eq. (12) then becomes

$$E^2 = \Big(1 - s_1 \frac{E}{E_{LV,1}}\Big)p^2c^2. \tag{13}$$

The propagation velocity will be modified by the relation $\upsilon = \partial E/\partial P$. Comparing to the speed of EM waves in the previous section, the parameter β has the relation

$$\beta = \Big[1 - s_1\Big(\frac{E}{E_{LV,1}}\Big)\Big]^2. \tag{14}$$

From the relation (14), we can find that the propagation velocity of the GRBs $\sqrt{\beta}c$ depends on the energy of photons and LV energy scale $E_{LV,1}$, that is to say, the LV effect will lead to $\beta \neq 1$. The expression of $E_{LV,1}$ turns out to be

$$E_{LV,1} = s_1 \frac{1}{H_0} \frac{E_h - E_1}{\Delta t_{LV}} \int_0^z \frac{(1+z')dz'}{h(z')}, \tag{15}$$

where $h(z) = \sqrt{\Omega_\Lambda + \Omega_K(1+z)^2 + \Omega_m(1+z)^3 + \Omega_R(1+z)^4}$ and H_0, z, E_h are the Hubble parameter, the redshift and the energy of the high-energy photons, respectively. E_1 is the energy of the low energy photon that the energy is low and we can neglect it.

Removing the intrinsic emission time lag in the source, we can get the value of $E_{LV,1}$ from the time delay or spectral lag of GRB 170817A under ideal conditions. However, due to the lack of the data of high-energy photons (the GRB 170817A

was only observed by Fermi-GBM), the explicit time delay or spectral lag of the GRB 170817A was not detected.[5] Our strategy is to assume that the possible LV observed from other GRBs[6] is generic, that is to say, the GRB 170817A would have the similar LV behavior. We therefore take the value of $E_{LV,1}$ obtained from these GRB data. It turns out that $E_{LV,1}$ has the following characteristics: Firstly, using different GRB data will get different constraints on $E_{LV,1}$. Secondly, different analysis methods, energy sections or cosmological models will effect the result even if the same event such as GRB 090510. The full list of the restrictions on $E_{LV,1}$ can be found in our previous work,[7] which indicates that $E_{LV,1}$ ranges from 10^{17} GeV to 10^{20} GeV from the time lag of the GRBs (the newest GRB events,[8] where $E_{LV,1} \sim 10^{17}$ GeV, is also in this range).

As for the highest energy of the photon of the GRB 170817A, we estimate it by gamma-gamma absorption with the bulk Lorentz factors Γ in the early phase. The specific method is presented in Ref.[9]

$$E_{high} = \frac{\Gamma^2 (m_e c^2)^2}{E_{high,an}(1+z)^2}, \tag{16}$$

where z is the redshift of the GRBs, m_e is the mass of electron, $E_{high,an}$ is the spectral break/cutoff energy and E_{high} is the corresponding highest energy of photons. Generally speaking, one can assume $E_{high,an} = E_{peak}$ (the peak energy of the Band fit of the GBM data), which implies those photons have the maximum probabilities to undergo the absorption with the highest-energy photons (E_{high}). As to the value of the Lorentz factor, we apply the estimated values $(100 - 10^{2.5})$ as suggested in the literatures,[10,11] combining the observations data of the GRB 170817A ($E_{peak} = (185 \pm 62)$ keV and $z \simeq 0.01$ in[2]), we can have an estimation $E_{high} \sim (10 - 200)$ GeV.

Based on these two estimations, we can find that the speed of the GRB 170817A has a tiny deviation with the light speed c due to the possible LV effect, and the deviation depends on the highest energy photons E_{high} and LV energy scale $E_{LV,1}$. In the most stringent constraint $E_{LV,1} \sim 10^{20}$ GeV,[12] the deviation can be restricted in the range

$$|\triangle v/c| < (10^{-19} - 10^{-18}). \tag{17}$$

4. Result

With the constraint (17), now we can turn to discuss the constraint on the HL gravity. Recalling that the speed of electromagnetic and gravitational waves are the same as discussed in the previous sections. Therefore the speed of GW respects the same constraint as Eq. (17). This, in turn, puts a constraints on the parameter β of the model

$$|1 - \sqrt{\beta}| < (10^{-19} - 10^{-18}). \tag{18}$$

This is the main result of the paper. Some points should be addressed in what follows. Firstly, due to the lack of direct observation on the time delay of the GRB 170817A, we do not have direct value of $E_{LV,1}$ of this event. Instead, we estimate it by using values obtained from other GRB events. Therefore, the above result (18) depends on which GRB events we adopt. Secondly, up to date, the lack of more GW event companying EM counterpart makes it difficult to implement a statistic analysis. Hence, the system error of our result expect to be obvious. Moreover, this result has the limitation that it is only valid for the present HL gravity model which has the feature that the speed of EM waves and GWs are the same. In spite of these, our result is meaningful since it provide a new way to constraint the model parameter from the properties of matter (the LV effect of GRBs for example as what we did in this work). Meanwhile, with the development of detectors, the first two flaws will be wiped since we can expect with more events it is possible to directly observe the time lag of the EM counterpart.

5. Conclusion

In this talk we present a (3+1)-dimensional HL gravity couple to an anisotropic EM field, which can be obtained by a KK reduction from a (4+1)-dimensional HL gravity. The most remarkable feature of this model is that the electromagnetic and gravitational waves have the same speed, both in the Minkowski background and the FRW spacetime. Then we can use this feature to restrict the speed of gravitational wave and the parameter of model by analyzing the LV effect of the companying EM event, the GRB 170817A. We find, up to some estimations, the parameter β of the model can be received a more stringent constraint than the one obtained from GW analysis itself.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 11975116, and Jiangxi Science Foundation for Distinguished Young Scientists under grant number 20192BCB23007.

References

1. B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," Phys. Rev. Lett. **119**, no. 16, 161101 (2017) [arXiv:1710.05832 [gr-qc]].
2. B. P. Abbott *et al.* [LIGO Scientific and Virgo and Fermi-GBM and INTEGRAL Collaborations], "Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A," Astrophys. J. **848**, no. 2, L13 (2017) [arXiv:1710.05834 [astro-ph.HE]].
3. G. Amelino-Camelia, J. R. Ellis, N. E. Mavromatos, D. V. Nanopoulos and S. Sarkar, "Tests of quantum gravity from observations of gamma-ray bursts," Nature **393**, 763 (1998) [astro-ph/9712103].

4. J. Bellorin, A. Restuccia and F. Tello-Ortiz, "Anisotropic coupling of gravity and electromagnetism in Hořava-Lifshitz theory," Phys. Rev. D **98**, no. 10, 104018 (2018) [arXiv:1807.01629 [hep-th]].
5. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil and O. J. Roberts, *et al.* "An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A," Astrophys. J. Lett. **848**, no.2, L14 (2017) [arXiv:1710.05446 [astro-ph.HE]].
6. J. R. Ellis, N. E. Mavromatos, D. V. Nanopoulos, A. S. Sakharov and E. K. G. Sarkisyan, "Robust limits on Lorentz violation from gamma-ray bursts," Astropart. Phys. **25**, 402-411 (2006) [erratum: Astropart. Phys. **29**, 158-159 (2008)] [arXiv:0712.2781 [astro-ph]].
7. T. Zhang, F. W. Shu, Q. W. Tang and D. H. Du, "Constraints on Hořava–Lifshitz gravity from GRB 170817A," Eur. Phys. J. C **80**, no.11, 1062 (2020) [arXiv:2011.00816 [gr-qc]].
8. J. Zhu and B. Q. Ma, "Pre-burst events of gamma-ray bursts with light speed variation," Phys. Lett. B **820**, 136518 (2021) [arXiv:2108.05804 [astro-ph.HE]].
9. Q. W. Tang, F. K. Peng, X. Y. Wang and P. H. T. Tam, "Measuring the bulk Lorentz factors of gamma-ray bursts with Fermi," Astrophys. J. **806**, no. 2, 194 (2015) [arXiv:1412.3342 [astro-ph.HE]].
10. G. P. Lamb *et al.*, "The optical afterglow of GW170817 at one year post-merger," Astrophys. J. **870**, no. 2, L15 (2019) [arXiv:1811.11491 [astro-ph.HE]].
11. R. Gill, J. Granot, F. De Colle and G. Urrutia, "Numerical Simulations of an Initially Top-Hat Jet and the Afterglow of GW170817 / GRB170817A," Astrophys. J. **883**, no. 1, 15(2019) arXiv:1902.10303 [astro-ph.HE].
12. Z. Chang, Y. Jiang and H. Lin, "A unified constraint on the Lorentz invariance violation from both short and long GRBs," Astropart. Phys. **36**, 47 (2012) [arXiv:1201.3413 [astro-ph.HE]].

Summary of the parallel session AT6

Anzhong Wang
GCAP-CASPER, Physics Department, Baylor University, Waco, Texas, 76798-7316, USA
E-mail: Anzhong_Wang@baylor.edu

In this brief report, as the session convener of Hořava-Lifshitz (HL) Gravity (AT6), I summarize the main results of the 15 talks presented in the 16th Marcel Grossmann Meeting, July 5–10, 2021. This session mainly focused on classical and quantum aspects of HL gravity and some related theories, such as Einstein-aether theory and khronometric gravity, as well as their applications to cosmology and astrophysics.

Keywords: Anisotropic scaling; breaking Lorentz symmetry; quantizations; black holes; cosmology

1. A Brief Introduction to Hořava-Lifshitz gravity

Hořava-Lifshitz (HL) gravity is a theory proposed by Hořava in 2009.[1] To avoid Ostrogradsky's ghosts,[2,3] a problem that has been plaguing quantization of general relativity (GR) since the middle of 1970's,[4] Hořava chose to break the Lorentz invariance (LI) by a Lifshitz-type of anisotropic scaling between space and time at the ultra-high energy,[5,6] $t \to b^{-z}t,\ x^i \to b^{-1}x'^i,\ (i = 1, 2, ..., d)$, where z denotes the dynamical critical exponent, and d the spatial dimension of the spacetime. LI requires $z = 1$, while power-counting renomalizibility requires $z \geq d$. With the above scalings, the time and space have, respectively, the dimensions, $[t] = -z$, $\left[x^i\right] = -1$. Clearly, such a scaling breaks explicitly the LI and hence the diffeomorphism invariance. Hořava assumed that it is broken only down to the level

$$t \to \xi_0(t), \quad x^i \to \xi^i\left(t, x^k\right), \tag{1}$$

so the spatial diffeomorphism still remains. The above symmetry is often referred as to *the foliation-preserving diffeomorphism*, denoted by Diff(M, $\mathcal{F}$). Introducing the Arnowitt-Deser-Misner (ADM) variables $\left(N, N^i, g_{ij}\right)$,[7] where $N,\ N^i$ and g_{ij}, denote, respectively, the lapse function, shift vector, and d-dimensional metric of the leaves t = constant, it is found that they are scaling, respectively, as,[1] $N \to N,\ N^i \to b^{z-1}N^i,\ g_{ij} \to g_{ij}$, so that their dimensions are $N = 0,\ \left[N^i\right] = z - 1,\ [g_{ij}] = 0$. With the above symmetry, one can construct the basic operators of the fundamental ADM variables and their derivatives, which turn out to be,

$$R_{ij}, \quad K_{ij}, \quad a_i, \quad \nabla_i, \tag{2}$$

where $a_i \equiv N_{,i}/N$, and ∇_i denotes the covariant derivative with respect to g_{ij}, while R_{ij} and K_{ij} are the d-dimensional Ricci tensor, and the extrinsic curvature tensor of the leaves t= constant, respectively, constructed from the g_{ij} and g^{ij} where

$g_{ij}g^{ik} = \delta^k_j$, and $K_{ij} \equiv \left(-\dot{g}_{ij} + \nabla_i N_j + \nabla_j N_i\right)/(2N)$, with $\dot{g}_{ij} \equiv \partial g_{ij}/\partial t$. These basic quantities are vectors/tensors under the coordinate transformations (1), and have the dimensions, $[K_{ij}] = z,\ [R_{ij}] = 2,\ [a_i] = 1,\ [\nabla_i] = 1$. With the basic blocks of Eq.(2), we can build scalar operators order by order, so the total Lagrangian will take the form,[1]

$$\mathcal{L}_g = \sum_{n=0}^{2z} \mathcal{L}_g^{(n)}\left(N, N^i, g_{ij}\right), \tag{3}$$

where $\mathcal{L}_g^{(n)}$ denotes the part of the Lagrangian that contains operators of the nth-order only. Including all the terms that are allowed by the symmetry, it was found that there are about 100 terms appearing in the above action in the case $d = z = 3$.

To reduce the number of independent coupling terms, Hořava introduced *the projectability condition*,[1] which requires that the lapse function be a function of t only, $N = N(t)$, so that all the terms proportional to a_i and its derivatives will be dropped out. This will reduce the total number of the independent terms in Eq.(3) to 14 (including the cosmological constant and parity-violated terms).[8] However, despite of various desired features, this condition leads to several problems, including that the Newtonian limit does not exist, and the six-order derivative operators are eliminated, so the theory is still not power-counting renormalizable. In addition, it is not clear if this symmetry is still respected by radiative corrections. Even more fundamentally, the foliation-preserving diffeomorphism (1) allows one more degree of freedom in the gravitational sector, in comparing with that of general diffeomorphism. As a result, a spin-0 mode of gravitons appears generically. This mode is potentially dangerous and may cause ghosts and instability problems, which lead the constraint algebra dynamically inconsistent.

To solve these problems, various modifications have been proposed.[8] In some of these models the Einstein-aether theory[9] and khronometric gravity[10] are particular cases. For a recent review on these models, we refer readers to Ref. 8.

2. Summary of the 15 Talks

In this section, I summarize the main results of the 15 talks in a chronological order.

• Sato presented a joint work with Ambjorn, Hiraga and Ito. In particular, they quantized the 2D projectable HL gravity with a bi-local as well as spacelike wormhole interaction, and found that the resulting quantum Hamiltonian coincides with the one obtained through summing over all genus in the string field theory for 2D causal dynamical triangulations. This implies that the wormhole interaction can be interpreted as a splitting or joining interaction of 1D strings.

• Bellorín presented a joint work with Droguett, and showed that the Batalin-Fradkin-Vilkovisky (BFV) quantization scheme can be implemented in the nonprojectable 3D HL theory. The BFV quantization is based on the canonical formalism and suitable to incorporate the measure associated to the second-class constraints

that the theory has. They presented the Becchi-Rouet-Stora-Tyutin symmetry transformations in the canonical variables, and found that the theory is of rank one in the classification introduced by Fradkin and Fradkina. In particular, they found that the nonlocal gauge condition introduced in the projectable theory can be included among these gauges.

• Satheeshkumar presented a talk on how the nature of singularities and horizons in vector-tensor theories differ significantly from GR. In particular, by considering Einstien-aether theory, based on a joint work with Chan and da Silva, he showed that, due to the presence of the aether field, vacuum black holes in GR are replaced by naked singularities, while new singularities appear in some cosmological models of GR.

• Lappicy presented a joint work with Hell and Uggla on bifurcations and chaos in HL cosmology. In particular, they considered the initial singularity in spatially homogeneous Bianchi type VIII and IX vacuum models, where relativistic first principles are replaced with anisotropic scalings of Lifshitz type. Within this class of models, GR is shown to be a bifurcation where chaos becomes generic. To describe the chaotic features of generic singularities in HL cosmology, they introduced symbolic dynamics within Cantor sets and iterated function systems, and obtained the general properties of various cases.

• Shu presented his joint work with Zhang, Tang and Du on constraining HL gravity from GRB 170817A observations. In particular, they considered the HL gravity coupling with an anisotropic electromagnetic field, generated through a Kaluza-Klein reduction of a $(4+1)$-dimensional HL gravity, and found that the model has the same velocity for both gravitational and electromagnetic waves. By analyzing the possible Lorentz invariance violation effect of the GRB 170817A, they found a stringent constraint on a parameter β of the theory with $\left|1-\sqrt{\beta}\right| < (10^{-19}-10^{-18})$.

• Saridakis presented his joint work with Khodadi on Einstein-aether gravity in the light of Event Horizon Telescope (EHT) observations of M87, and calculated the photon effective potential, the unstable photon sphere radius, and the induced angular size of two analytical black hole solutions in the theory. Since d_{M87*} is observationally known from the EHT Probe, they extracted the corresponding parameter regions in order to obtain consistency, and found that the Einstien-aether black hole solutions agree with the shadow size of EHT $M87*$, if the involved aether parameters are restricted within specific ranges, along with an upper bound on the dimensionless spin parameter a, which is verified by a full scan of the parameter space within 1σ-error.

• Ajith presented a joint work with Yagi and Yunes on testing HL gravity with I-Love-Q. The authors studied how the I-Love-Q relations depend on the parameters of the theory, and found that the only un-constrained parameter disappears from the field equations in HL gravity. Therefore, the I-Love-Q relations are universal against not only the nuclear physics uncertainty but also the gravitational physics uncertainty within HL gravity.

• Ghosh presented a joint work with Kumar on the Einstein-Gauss-Bonnet (EGB) gravity under the limit $D \to 4$, and studied rotating black holes in regularized 4D EGB gravity and their horizon properties and shadows. In particular, they found that the $4D$ non-relativistic Horava-Lifshitz theory of gravity also admits the identical black hole solution. The effects of the GB coupling parameter on the shape and size of shadows were investigated and analyzed in recent M87 observations from EHT. They estimated the parameters associated with the 4D EGB gravity Kerr black holes using the shadow observables, and found that the inferred circularity deviation $\Delta C \leq 0.1$ for the M87 black hole, whereas shadow angular diameter $\theta_d = 42 \pm 3\mu as$ within the 1σ region on the GB parameter. Interestingly, the shadow axial ratio obeying undefined control sequence is in agreement with the EHT results and thus eventuates in the 4D EGB gravity black holes being suitable candidates for astrophysical black holes.

• Bamba presented a joint work with Chakraborty and Saa on dynamical system analysis of the homogeneous and anisotropic Bianchi-I spacetimes in $f(R)$ gravity coupled with an anisotropic matter. It was demonstrated that with a suitable choice of the evolution parameter, Einstein's equations are reduced to an autonomous 5D system of ordinary differential equations. Furthermore, for a large class of functions $f(R)$, which includes several cases commonly considered in the literature, all the fixed points are polynomial roots, and thus they can be determined with a good accuracy and classified for stability. In addition, any fixed point corresponding to isotropic solutions in the presence of anisotropic matter will be unstable. The assumption of a perfect fluid as source and or the vacuum cases imply some dimensional reductions and even more simplifications. In particular, it was found that the vacuum solutions of $f(R) = R^{\delta+1}$ (with δ = constant) are governed by an effective 2D phase space which can be constructed analytically, leading to an exactly solvable dynamics.

• Jawad presented various solutions of Einstein-aether gravity through a reconstruction approach. In order to discuss the current cosmic acceleration corresponding to his reconstructed models, he evaluated different cosmological parameters, and discussed the consistency of his results of cosmological parameters with current observational data.

• de Souza Campos presented her joint work with Dappiaggi and Sina on boundary conditions for the Klein-Gordon field in Lifshitz spacetimes. In particular, she considered a free, scalar, massive quantum field theory on a 4D Lifshitz spacetime with a critical exponent $z = 2$ and showed that there are two-point functions for ground and thermal states of the local Hadamard form and satisfying the canonical commutation relations, compatible with Robin's boundary conditions and with mode-dependent boundary conditions, depending on the value of the effective mass. Each of these generalized boundary conditions determine an inequivalent dynamics, but they are all equivalently physically sensible.

• Panotopoulos presented his recent work on non-rotating and isotropic strange quark stars in HL gravity and Einstein-æther theory. For quark matter he adopted

both linear and non-linear equations of state (EoS), corresponding to the MIT bag model and color flavor locked state, respectively. The new structure equations describe the generalization of the hydrostatic equilibrium of the usual Tolman-Oppenheimer-Volkoff (TOV) equations of GR, parameterized by a dimensionless deviation parameter v. He computed the mass, radius and compactness of the stars, and showed graphically the impact of v on the mass-to-radius profiles for different EoS describing quark matter. The energy conditions and stability criteria were also considered, and found that they are all fulfilled.

• Lin presented a joint work with Wang, Zhang, Zhao and Zhu on testing Einstein-aether theory by gravitational wave (GW) observations. In particular, he presented their recent studies of GWs produced by massive and compact objects, including the GW waveforms, polarizations, response function, Fourier transform, and energy loss rate through three different channels of radiation - the scalar, vector and tensor modes, and showed that combination of such theoretical predictions with future observations of GWs can bring severe constraints on the theory.

• Kwapisz presented a joint work with Chojnacki on finite action principle and black holes in HL gravity, and showed that the finite action selection principle works well for HL gravity in the context of black holes. In addition, wormholes possess a finite action and hence contribute to the path-integral of quantum gravity.

• Chojnacki presented a joint work with Kwapisz on their studies of the finite action principle and the beginning of the universe in HL gravity, and showed that field configurations with finite classical action describe a flat universe with a homogeneous and isotropic beginning without the ghost particles.

References

1. P. Horava, *Quantum gravity at a Lifshitz point*, Phys. Rev. D**79** (2009) 084008.
2. M. Ostrogradsky, Mem. Ac. St. Petersbourg, VI**4** (1850) 385.
3. R. P. Woodard, *The Theorem of Ostrogradsky*, Scholarpedia **10** (2015) 32243.
4. K. S. Stelle, *Renormalization of higher-derivative quantum gravity*, Phys. Rev. D**16** (1977) 953.
5. E. M. Lifshitz, *On the theory of second-order phase transitions I*, Zh. Eksp. Teor. Fiz. **11** (1941) 255.
6. E. M. Lifshitz, *On the theory of second-order phase transitions II*, Zh. Eksp. Teor. Fiz. **11** (1941) 269.
7. R. Arnowitt, S. Deser, and C. W. Misner, *The Dynamics of General Relativity*, Gen. Relativ. Gravit. **40**, 1997 (2008).
8. A. Wang, *Hořava gravity at a Lifshitz point: A progress report*, Inter. J. Mod. Phys. D**26** (2017) 1730014.
9. T. Jacobson, *Extended Horava gravity and Einstein-aether theory*, Phys. Rev. D**81** (2010) 101502; *ibid.*, D**82** (2010) 129901(E); *Undoing the twist: the Hořava limit of Einstein-aether*, *ibid.*, D**89** (2013) 081501.
10. D. Blas, O. Pujolas, and S. Sibiryakov, *Models of non-relativistic quantum gravity: The good, the bad and the healthy*, JHEP **04** (2011).

Non-local R^2-like inflation, gravitational waves and Non-Gaussianities

K. Sravan Kumar

Department of Physics, Tokyo Institute of Technology
1-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan
E-mail: sravan.k.aa@m.titech.ac.jp,
www.titech.ac.jp/english

The emergence of R^2 (Starobinsky) inflation from the semi-classical modification of gravity due to matter quantum fields (trace anomaly) clearly points out the importance of fundamental physics and the first principles in the construction of successful cosmological models. Along with the observational success, R^2 gravity is also an important step beyond general relativity (GR) towards quantum gravity. Furthermore, several approaches of quantum gravity to date are strongly indicating the presence of non-locality at small time and length scales. In this regard, ultraviolet (UV) completion of R^2 inflation has been recently studied in a string theory-inspired ghost-free analytic non-local gravity. We discuss the promising theoretical predictions of non-local R^2-like inflation with respect to the key observables such as tensor-to-scalar ratio, tensor tilt which tell us about the spectrum of primordial gravitational waves, and scalar Non-Gaussianities which tell us about the three-point correlations in the CMB fluctuations. Any signature of non-local physics in the early Universe will significantly improve our understanding of fundamental physics at UV energy scales and quantum gravity.

Keywords: Quantum gravity, Inflation, CMB, Gravitational Waves

1. Status of inflation and the quests for UV-completion

Inflationary cosmology has become the most important field of study in theoretical higher energy physics (THEP) in the recent years not only because of its success with respect to the data of Cosmic Microwave Background (CMB)[1–3] but also it gives a best chance to test our approaches to build physics at the fundamental scales such as Planck energy or length scales. Recent detection of Gravitational Waves from binary black hole merger events further corroborated to build new hopes and strategies to probe the physics of early Universe.[4,5]

Inflationary cosmology is so far developed in two equivalent ways. One is by geometrical modification i.e., by extension of General Relativity (GR) by higher curvature terms (for example "$R + R^2$ gravity which is also known as Starobinsky theory[6–9]). Second way is by the addition of hypothetical (scalar) matter fields which result in modification of right hand side of Einstein equations.[10,11] This second way motivated in the view of particle physics beyond the standard model and several ultraviolet (UV) complete approaches such as string theory and supergravity (SUGRA).[12] In this regard, inflationary cosmology is so far studied from the point of view of beyond the standard model of particle physics and as a low-energy limit of UV-complete approaches such as string theory/M-theory (see left panel of Fig. 1).

In the last decades plethora of inflationary models have been constructed with different scalar fields emerging from simplest set ups to all the way up to the most general scalar tensor theories[13,14] (see left panel of Fig. 1). However the latest data from Planck satellite[1] has found out Starobinsky and Higgs inflation be the most compatible models with the following predictions of spectral index and the tensor to scalar ratio

$$n_s = 1 - \frac{2}{N}\,, \qquad r = \frac{12}{N^2}\,, \tag{1}$$

where $N = 50 - 60$ number of e-folds before the end of inflation. The recent Planck data[1,2] has constrained $n_s = 0.9649 \pm 0.0042$ at 68%CL, $r < 0.036$ at 95% CL. The success of Starobinsky and Higgs inflation has lead to the emergence of equivalent frameworks of inflation in string theory and SUGRA.[12]

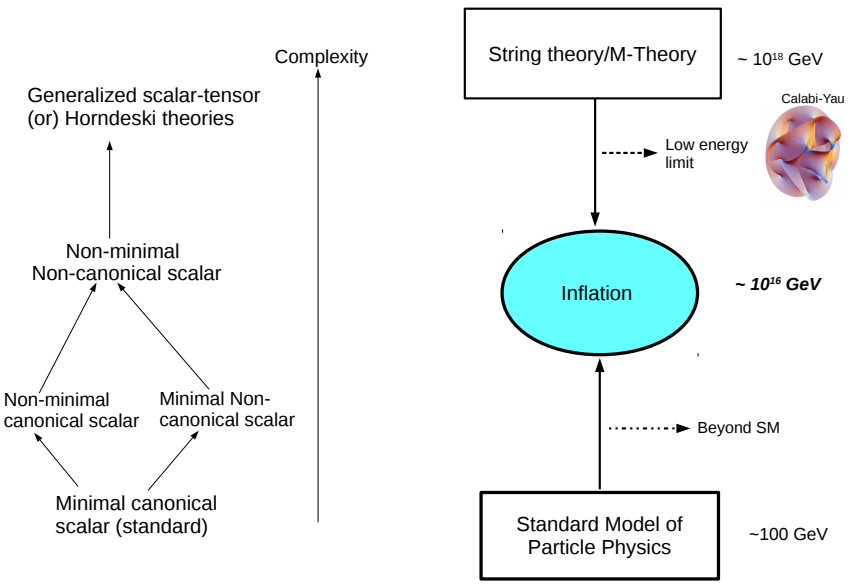


Fig. 1. In the right panel, arrow of complexity shows how one can start with a simplest scalar field Lagrangian and generalized all the way up to most general scalar-tensor theories whose equations of motion are second order in field derivative. In the right panel we can see that inflationary cosmology can be understood well either from beyond the standard model of particle physics or from low-energy limit of a UV-complete approaches such as string theory (This Figure taken from[13]).

The challenging problem of the current status of inflationary cosmology is to break several degeneracies in the frameworks of inflation and find definitive clues towards achieving quantum gravity and UV-completion. For this purpose, from observational point of view, one must go beyond the two observables (n_s, r). In this regard there are two more important observational parameters known as tensor tilt

(n_t) which is the tilt of the tensor power-spectrum and non-Gaussianity parameter (f_{NL}) which is related to the 3-point scalar correlations. If inflation is driven by single (canonical) scalar field with standard slow-roll then the tensor consistency relation[13] and the Maldacena consistency relation[15] given by

$$r = -8n_t \quad f_{NL}^{\rm sq} = \frac{5}{12}(1 - n_s) \tag{2}$$

serves as a crucial test. Here $f_{NL}^{\rm sq}$ is so-called squeezed shape of non-Gaussianity. If inflation is driven by multiple fields then the above consistency relations are understood to be violated.[16] If inflation is driven by non-canonical scalar fields and if non-standard quantum initial conditions are assumed we will find a strong signal in the equilateral ($f_{NL}^{\rm eq}$) and orthogonal ($f_{NL}^{\rm ortho}$) shapes of non-Gaussianities. These shapes are constrained by the latest Planck data as

$$f_{\rm NL}^{\rm loc} = 0.8 \pm 5.0\,, \quad f_{\rm NL}^{\rm equi} = -4 \pm 43\,, \quad f_{\rm NL}^{\rm ortho} = -26 \pm 21\,, \quad \text{at } 68\%\text{CL}. \tag{3}$$

Violation of consistency relations (2) and any detection of $f_{NL} \sim O(1)$ is definitely a new physics in the context of early Universe cosmology which will be crucial for further advancements in THEP. Moreover, in a quest to find signatures of new physics, recently a new program of research has gained significant attention known as "cosmological collider physics"[17] whose aim is to find signatures of higher energy particles interacting with primordial fields during inflation.

Despite numerous phenomenological approaches in understanding inflation and find new observables to probe physics of inflation, it is very vital to focus on consistency of our theoretical endeavors. The success of Starobinsky and Higgs inflationaris (which are almost identical during inflation) covey us a very crucial property that at the scale of inflation approximate "scale invariance" symmetry of the Lagrangian is important. The fact that various setups of SUGRA leading to Starobinsky-like or R^2-like inflation[12] indicate that $R + R^2$ is an important modification of GR. In the next section, we will discuss in detail about the foundations of Starobinsky inflation and how a very straightforward procedure from geoemertric modification of gravity lead us a UV-complete framework.[18–21]

2. From Starobinsky inflation to Stelle gravity and the emergence of ghost-free UV-non-local gravity

$R + R^2$ gravity should not be viewed as a toy model in the class of f(R) gravity. The theory naturally emerges from generic considerations and important questions about UV-completion which dates back the discovery of trace anomaly.[22] The trace anomaly comes from 1-loop quantum corrections to the graviton propagator due to matter conformal fields leading to an anomalous non-zero trace of energy momentum tensor which is proportional to curvature square term.

$$M_p^2 R = g^{\mu\nu}\langle T_{\mu\nu}\rangle = aF + bG + c\Box R\,, \tag{4}$$

where $F = W_{\mu\nu\rho\sigma}W^{\mu\nu\rho\sigma}$ is the the Weyl tensor square and $G = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2$ is the so-called Gauss-Bonnet term and $R_{\mu\nu\rho\sigma}$, $R_{\mu\nu}$, R Riemann tensor, Ricci tensor and Ricci scalar respectively. Here a, b, c are dimensionless coefficients quantified by the number of conformal fields. Trace anomaly has been widely studied and applied in various fundamental theories including anticipated UV complete theories such as string theory and supergravity (SUGRA).[22] Neglecting the coefficients a, b the trace anomaly equation reduces to the trace equation of the $R + R^2$ given by the action[6,23]

$$S_{R^2} = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + \frac{M_p^2}{12M^2}R^2\right], \tag{5}$$

where $M \ll M_p$ becomes the mass of a propagating scalar in this model, named scalaron. Independent of what is the right framework of UV-completion R^2 term is very natural term that appears as a classical addition to the GR and it naturally leads to a quasi-de Sitter evolution (which is "cosmic inflation") in the high curvature regime $R \gg M^2$. Therefore it is a matter of fact that $R + R^2$ gravity indeed an unavoidable and very important extension of GR. Further more if we add Weyl square term to the action of $R+R^2$ gravity (5) the resultant gravity theory becomes renormalizable (which is an important property to achieve quantum gravity) as it was proven by K. S. Stelle long ago[24]

$$S_{\text{Stelle}} = S_{R^2} = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + \frac{M_p^2}{12M^2}R^2 + \frac{M_p^2}{M_g^2}W_{\mu\nu\rho\sigma}W^{\mu\nu\rho\sigma}\right]. \tag{6}$$

However, Stelle gravity unfortunately falls short in becoming a consistent theory of quantum gravity because of the presence of a massive tensor ghost in the spectrum which spoils Unitarity. A natural question one can ask is that whether this ghost can be removed by any higher derivative extension of Stelle gravity. Due to Ostrogradsky theorem any finite derivative extension of Stelle gravity results in further introduction of ghosts in the spectrum and it becomes not a viable option. However, Krasnikov has shown that analytic infinite derivative extension of Stelle gravity (i.e., analytic non-local gravity) can give a ghost-free theory.[25]

$$S = \int d^4x\sqrt{-g}\left(\frac{M_p^2}{2}R + \frac{1}{2}\left[R\mathcal{F}_R\left(\Box_s\right)R + W_{\mu\nu\rho\sigma}\mathcal{F}_W\left(\Box_s\right)W^{\mu\nu\rho\sigma}\right]\right), \tag{7}$$

where $\Box_s = \frac{\Box}{\mathcal{M}_s^2}$ with $\Box$ is being the d'Alembertian and $\mathcal{F}_R, \mathcal{F}_W$ analytic infinite derivative operators or non-local form factors. This theory (7) has been studied over the years and understood to have good properties of UV-completion with respect to renormalizability and Unitarity (See[21] and the references therein). In a nutshell, the action (7) is a very logical extension of Stelle gravity and we very much straightforwardly end up with introducing non-locality in the gravity sector. Note that "Non-locality" is a very important common concept that appears in several approaches to quantum gravity such as string theory, loop quantum gravity,

causal sets and non-commutative gravity. Therefore, (7) is a culmination of various approaches which allow us to give a consistent UV-completion for Starobinsky inflation[a]. Further note that (7) should not be understood as an effective theory but rather a consistent part of a UV-complete theory. We do not have any cut off scales in the theory and the scale $\mathcal{M}_s$ represents the scale at which non-local nature of gravity is prominent. In the next section, we will discuss the cosmological application of (7) and identify the specific features of non-local gravity which can be probed in the future observations.

3. Non-local R^2-like inflation and future targets for CMB and gravitational wave observations

In this section, we apply UV non-local quadratic curvature gravity in the context of inflationary cosmology. The first and foremost difficulty is associated with solving equations of motion of an infinite derivative theory which perhaps appears to be an impossible task. However, this was consistency achieved[18] and it was also quite rigorously shown in[19] that the theory contains an inflationary attractor solution exactly same as the local $R + R^2$ or Starobinsky theory. To see this explicitly let us write down the background equations of motion for (7) Friendmann-Lemaître-Robertson-Walker background for which the Weyl tensor is zero

$$\bar{E}^{\mu}_{\nu} \equiv -\left[M_p^2 + 2\lambda\mathcal{F}\left(\frac{\Box}{\mathcal{M}_s^2}\right)\bar{R}\right]\bar{G}^{\mu}_{\nu}\delta^{\mu}_{\nu} - \frac{1}{2}\bar{R}\mathcal{F}\left(\frac{\Box}{\mathcal{M}_s^2}\right)\bar{R}\delta^{\mu}_{\nu} + 2\left(\nabla^{\mu}\partial_{\nu} - \delta^{\mu}_{\nu}\Box\right)\mathcal{F}\left(\frac{\Box}{\mathcal{M}_s^2}\right)R + \bar{\mathcal{K}}^{\mu}_{\nu} - \frac{\lambda}{2}\delta^{\mu}_{\nu}\left(\bar{\mathcal{K}}^{\sigma}_{\sigma} + \bar{\mathcal{K}}\right) = 0\,,$$

where

$$\bar{\mathcal{K}}^{\mu}_{\nu} = \frac{1}{\mathcal{M}_s^2}\sum_{n=1}^{\infty} f_n \sum_{l=0}^{n-1}\partial^{\mu}\frac{\Box^l}{\mathcal{M}_s^{2l}}\bar{R}\partial_{\nu}\left(\frac{\Box}{\mathcal{M}_s^2}\right)^{n-l-1}\bar{R}\,,$$

$$\bar{\mathcal{K}} = \sum_{n=1}^{\infty} f_n \sum_{l=0}^{n-1}\frac{\Box^l}{\mathcal{M}_s^{2l}}\bar{R}\left(\frac{\Box}{\mathcal{M}_s^2}\right)^{n-l}\bar{R}\,.$$

The trace equation is

$$\bar{E} = M_p^2\bar{R} - 6\bar{\Box}\mathcal{F}\left(\frac{\bar{\Box}}{\mathcal{M}_s^2}\right)\bar{R} - \bar{\mathcal{K}}^{\mu}_{\mu} - 2\bar{\mathcal{K}} = 0\,. \tag{8}$$

It was showed in[19] that the only solution of the above equations is $\bar{\Box}\bar{R} = M^2\bar{R}$ given the following Unique conditions on the form factor are satisfied.

$$\mathcal{F}'_R\left(\frac{M^2}{M_s^2}\right) = 0,\ \frac{M_P^2}{2\lambda} = 3M^2\mathcal{F}_1,\ \text{ where } \mathcal{F}_1 \equiv \mathcal{F}_R\left(\frac{M^2}{\mathcal{M}_s^2}\right) \tag{9}$$

[a]Worth to metion that Starobinsky inflation has also been embedded in SUGRA which gives an interesting route for UV-completion.[26] Here we focus on non-supersymmetric aspects of UV-completion.

Fig. 2. We depict here the hierarchy of mass scales in non-local R^2-like inflation.

where $'$ denotes derivative with respect to argument. Note that $\bar{\Box}\bar{R} = M^2\bar{R}$ is the trace equation of local $R + R^2$ which gives Starobinsky inflationary solutions. Here given the conditions (9) are met we can solve exactly the full non-local gravity equations of motion and also we have good inflationary solution. Now it is important to see how non-local effects of gravity appear through inflationary perturbations. Before we go this step we can draw a heuristic bound on the scale of non-locality shown in Fig. 2 based on the fact that we require near scale invariance at the scale of inflation. This can be seen heuristically by expanding the quadratic Ricci scalar part of the action as

$$S = \int d^4x\sqrt{-g}\left[\frac{M_p^2}{2}R + \frac{M_p^2}{12M^2}R^2 + \mathcal{O}\left(\frac{M_p^2 R\Box R}{M^2\mathcal{M}_s^2}\right)\right].$$

In the high curvature regime $R \gg M^2$ local quadratic curvature term is naturally dominant and it is known to be scale invariance.

3.1. *Scalar and tensor power spectrum*

To study perturbations around quasi-de Sitter we take the perturbed metric of the form

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)\,d\eta^2 + ((1-2\Psi)\,\delta_{ij} + 2h_{ij})\,dx^i dx^j\right], \tag{10}$$

where Φ, Ψ are Bardeen potentials and h_{ij} is tensor fluctuation. The study of non-local perturbed equations of motion[19] has revealed that $\Phi + \Psi \approx 0$ during inflation. Since $\delta W_{\mu\nu\rho\sigma} \propto (\Phi + \Psi)$ the Weyl tensor part in (7) does not contribute to the scalar perturbations during inflation. Following the computations in[19,20] the second order action of scalar degree of freedom in de Sitter approximation becomes

$$\delta^2 S_{(s)} = \frac{1}{2\mathcal{F}_1\bar{R}}\int d\tau d^3x\sqrt{-\bar{g}}\Upsilon\frac{\mathcal{W}\left(\frac{\bar{\Box}}{\mathcal{M}^2}\right)}{\mathcal{F}_R\left(\frac{\bar{\Box}}{\mathcal{M}^2}\right)}(\bar{\Box} - M^2)\Upsilon\,, \tag{11}$$

where $\Upsilon \approx 2\bar{R}\Psi$. In order not to have any more ghost degrees of freedom we choose the structure of form factor as

$$\mathcal{F}_R(\Box_s) = \mathcal{F}_1\frac{3e^{\gamma_S(\Box_s)}\left(\Box_s - \frac{M^2}{\mathcal{M}_s^2}\right) + \left(\frac{\bar{R}}{\mathcal{M}_s^2} + 3\frac{M^2}{\mathcal{M}_s^2}\right)}{3\Box_s + \frac{\bar{R}}{\mathcal{M}_s^2}}, \tag{12}$$

where $\bar{R} \approx$ constant and γ_S is an entire function. Therefore, the theory now has only one scalar(on) degree of freedom similar to local $R + R^2$ theory but however, the the scalar field here is non-local in nature.

Computation of the power spectrum and the scalar spectral index of curvature perturbation $\mathcal{R} = \Psi + H\frac{\delta R}{\dot{R}} = \Psi + \frac{24H^3}{24H\dot{H}}\Phi \approx -\frac{1}{\epsilon}\Psi$ we get

$$\mathcal{P}_{\mathcal{R}} = \frac{H^2}{16\pi^2\epsilon^2}\frac{1}{3\lambda\mathcal{F}_1\bar{R}}\bigg|_{k=aH}, \quad n_s \equiv \frac{d\ln\mathcal{P}_{\mathcal{R}}}{d\ln k}\bigg|_{k=aH} \approx 1 - \frac{2}{N}. \tag{13}$$

Similarly the computation of the second order action of tensor perturbations gives (with addition of the non-trivial contributions from the Weyl tensor part in (7))

$$\delta^2 S_{(T)} = \frac{1}{2}\int d^4x\sqrt{-\bar{g}}h^{\perp}_{ij}e^{2\gamma_T\left(\Box_s - \frac{2\bar{R}}{3\mathcal{M}_s^2}\right)}\left(\bar{\Box} - \frac{\bar{R}}{6}\right)h^{\perp ij}, \tag{14}$$

where

$$\mathcal{F}_W(\Box_s) = \frac{\mathcal{F}_1\bar{R}}{\mathcal{M}_s^2}\frac{e^{\gamma_T\left(\Box_s - \frac{2}{3}\frac{\bar{R}}{\mathcal{M}_s^2}\right)} - 1}{\Box_s - \frac{2\bar{R}}{3\mathcal{M}_s^2}}. \tag{15}$$

which fixes the tensor degree of freedom to be just massless graviton.

Computing the tensor power spectrum in the leading order in slow-roll we obtain

$$\mathcal{P}_{\mathcal{T}}|_{k=aH} = \frac{H^2}{\pi^2\mathcal{F}_1\bar{R}}e^{-2\gamma_T\left(\frac{\bar{R}}{6\mathcal{M}_s^2}\right)}. \tag{16}$$

As a result the key inflationary parameter called tensor-to-scalar ratio becomes

$$r = \frac{12}{N^2}e^{-2\gamma_T\left(\frac{-\bar{R}}{2\mathcal{M}_s^2}\right)}\bigg|_{k=aH}. \tag{17}$$

The tilt of the tensor power spectrum can be obtained as

$$\begin{aligned} n_t \equiv \frac{d\ln\mathcal{P}_{\mathcal{T}}}{d\ln k}\bigg|_{k=aH} &\approx -\frac{d\ln\mathcal{P}_{\mathcal{T}}}{dN}\left(1 + \frac{1}{2N}\right)\bigg|_{k=aH} \\ &\approx -\frac{3}{2N^2} - \left(\frac{2}{N} + \frac{3}{2N^2}\right)\frac{\bar{R}}{2\mathcal{M}_s^2}\gamma_T^{(\prime)}\left(-\frac{\bar{R}}{2\mathcal{M}_s^2}\right)\bigg|_{k=aH}, \end{aligned} \tag{18}$$

where $'$ denotes derivative with respect to the argument.

4. Primordial (scalar) non-Gaussianities

Beyond two point correlations it is interesting to see if there will be non-Gaussianities in non-local R^2-like inflation. As we discussed before the scalar(on) in the non-local R^2-like inflation is non-local in nature. Therefore, we generally expect the non-local interactions to play a role in predicting inflationary 3-point correlations different from the local theories. The primordial (scalar) non-Gaussianities in

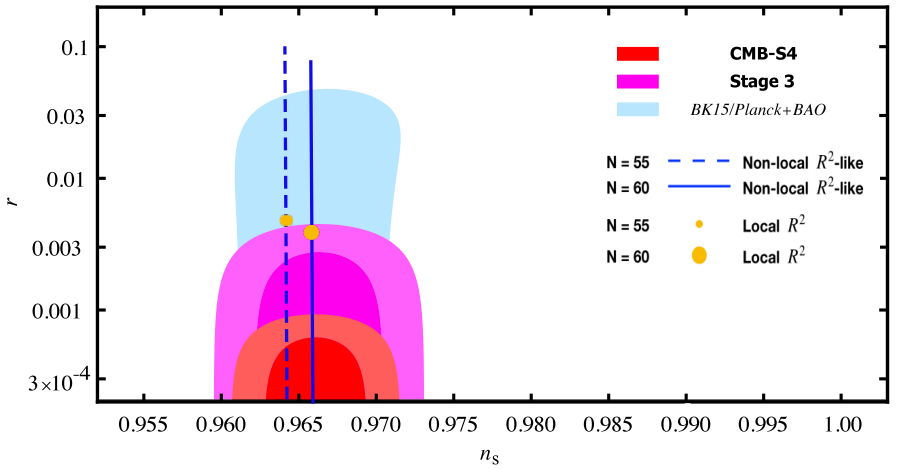


Fig. 3. We depict the predictions of non-local R^2-like inflation in the (n_s, r) plane of the latest CMB S-4 science paper about future forecast of detecting B-modes.[19, 21, 27]

the non-local R^2-like inflation are computed in[20] and here we briefly review the main results. The 3-point correlations and the parameter f_{NL} are defined by

$$\langle \mathcal{R}(\mathbf{k_1})\,\mathcal{R}(\mathbf{k_2})\,\mathcal{R}(\mathbf{k_3})\rangle = (2\pi)^3\,\delta^3(\mathbf{k_1}+\mathbf{k_2}+\mathbf{k_3})\,\mathcal{B}_{\mathcal{R}}(k_1,k_2,k_3)\ ,$$

and

$$B_{\mathcal{R}}(k_1,k_2,k_3) = (2\pi)^4\frac{1}{\prod_i k_i^3}\mathcal{P}_{\mathcal{R}}^{*2} f_{NL}(k_1,\,k_2,\,k_3)\ ,$$

where

$$f_{NL} = -\frac{5}{6}\frac{A_{\mathcal{R}}(k_1,\,k_2,\,k_3)}{\sum_i k_i^3}\,. \tag{19}$$

The interactions of the curvature perturbations in the non-local R^2-like inflation can be schematically explained as in Fig. 5.

To calculate bi-spectrum, first we expand our action to cubic order to the leading order in slow-roll parameter. Then we compute the correlations using the following definition

$$\langle \mathcal{R}(\mathbf{k_1})\,\mathcal{R}(\mathbf{k_2})\,\mathcal{R}(\mathbf{k_3})\rangle = -i\int_{-\infty}^{\tau_e} a d\tau \langle 0|[\mathcal{R}(\tau_e,\,\mathbf{k_1})\mathcal{R}(\tau_e,\,\mathbf{k_2})\mathcal{R}(\tau_e,\,\mathbf{k_3}),\,H_{int}]|0\rangle\ , \tag{20}$$

where $\boldsymbol{k}_i$ are wave vectors and H_{int} is the interaction Hamiltonian. It is related to a perturbation of the Lagrangian (7) expanded up to the 3rd order in curvature perturbations ($\mathcal{L}_3$) as $H_{int} \approx -\mathcal{L}_3$. We consider the mode functions as

$$\mathcal{R} \approx -\Psi/\epsilon, \quad \bar{\Box}\mathcal{R} = M^2\mathcal{R}\,. \tag{21}$$

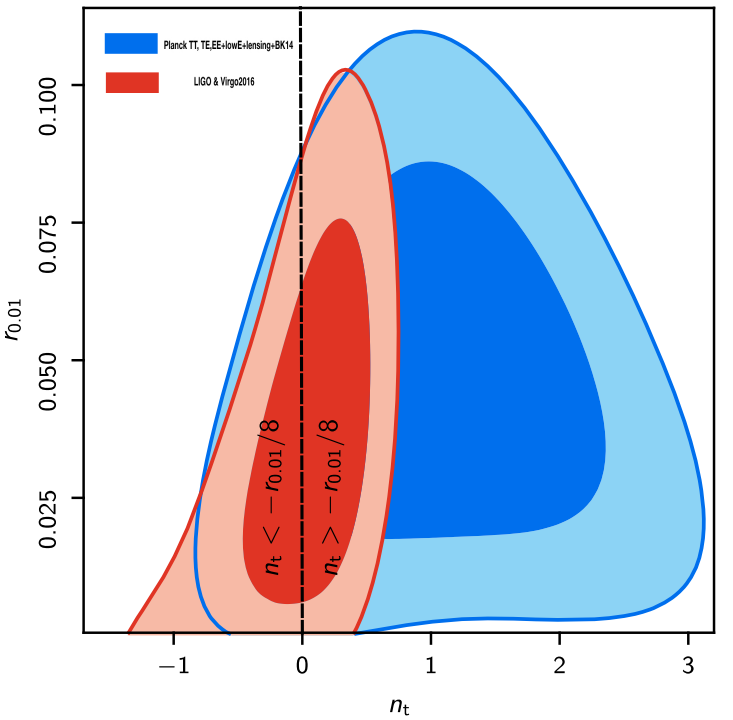


Fig. 4. We note that the predictions of non-local R^2-like inflation can be anywhere within the likelihood projected (n_t, r) plane of latest Planck 2018.[19, 21, 28]

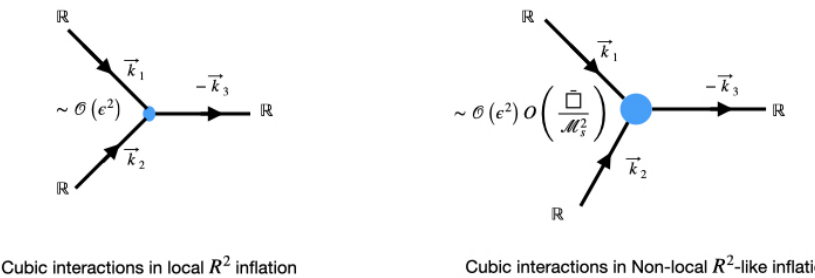


Fig. 5. In the above plot $\mathbb{R} = \{\mathcal{R}, \partial\mathcal{R}, \mathcal{D}^\mu\partial\mathcal{R}\}$ imply various tree level interactions of different modes of the curvature perturbation in the local R^2 and the non-local R^2-like inflation. $O\left(\frac{\bar{\Box}}{\mathcal{M}_s^2}\right)$ is some analytic non-local operator.

New Interactions arise via the commutation relation in de Sitter approximation

$$\Box\nabla_\mu\mathcal{R} = \nabla_\mu\Box\mathcal{R} + \frac{\bar{R}}{4}\nabla_\mu\mathcal{R}. \tag{22}$$

After numerous steps of computation the cubic order action in $\mathcal{R}$ of (7) in the leading order slow-roll approximation is obtained as[20]

$$\begin{aligned}\delta^{(3)}S_{(S)} =&4\epsilon M_p^2\int d\tau d^3x\Big\{T_1^*\mathcal{R}\nabla\mathcal{R}\cdot\nabla\mathcal{R}+T_2^*\mathcal{R}\mathcal{R}'^2+T_3^*\mathcal{H}^2\mathcal{R}^3\\&+T_4^*\mathcal{H}\mathcal{R}\mathcal{R}\mathcal{R}'+T_5^*\mathcal{H}^{-1}\nabla\mathcal{R}\cdot\nabla\mathcal{R}\mathcal{R}'+T_6^*\mathcal{H}^{-1}\mathcal{R}'^3\\&+T_7^*\mathcal{H}^{-2}\mathcal{R}'\nabla\mathcal{R}\cdot\nabla\mathcal{R}'\Big\},\end{aligned}$$

where T_i's are dimensionless constants whose can be found in[20]

Computing different shapes of f_{NL} parameter when $k=aH$ we get[20]

$$\begin{aligned}f_{NL}^{sq}\Big|_{k=aH}&\approx\frac{5}{12}(1-n_s)-23\epsilon^2\left[e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)}-1\right]-\frac{4\bar{R}}{\mathcal{M}_s^2}\epsilon^3e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)}\gamma_S'\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)\\f_{NL}^{eq}\Big|_{k=aH}&\approx\frac{5}{12}(1-n_s)-49\epsilon^2\left[e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_S^2}\right)}-1\right]-\frac{9\bar{R}}{\mathcal{M}_s^2}\epsilon^3e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)}\gamma_S'\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)\\f_{NL}^{ortho}\Big|_{k=aH}&\approx\frac{5}{12}(1-n_s)-43\epsilon^2\left[e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)}-1\right]-\frac{22\bar{R}_*}{3\mathcal{M}_s^2}\epsilon^3e^{\gamma_S\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right)}\gamma_S'\left(\frac{\bar{R}}{4\mathcal{M}_s^2}\right).\end{aligned}\tag{23}$$

These characterize the non-local interactions of curvature perturbation. Clearly from the above expressions single field Maldacena consistency relation is violated in non-local theories and one can obtain detectable level of non-Gaussianities due to non-local nature of interactions (See Fig. 6). This is totally non-trivial effect in comparison with all the known mechanisms of producing large non-Gaussianities in the local theories.

5. Concluding remarks

We conclude here with highlighting the notable results of early Universe cosmology in the context of UV-non-local ghost-free theories of gravity. First of all we note that the UV non-local gravity action (7) is a consistent and necessary extension of Stelle gravity in order to achieve UV-completion based on the first principles. Another point to note here is "non-locality" is an important concept that is needed for building a consistent theory of quantum gravity. We have discussed how one can have a successful realization of inflation with the action action (7) and how non-local nature of gravity can be seen through inflationary observables. These are the important conclusions we can draw here to probe non-local nature of gravity in the scope of future CMB and gravitational wave observations.[5,16,27–29]

- In the non-local R^2-like inflation the primordial scalar power spectrum remains same as in the case of local $R+R^2$ inflation. However the tensor power spectrum is modified and one can have any value of tensor-to-scalar

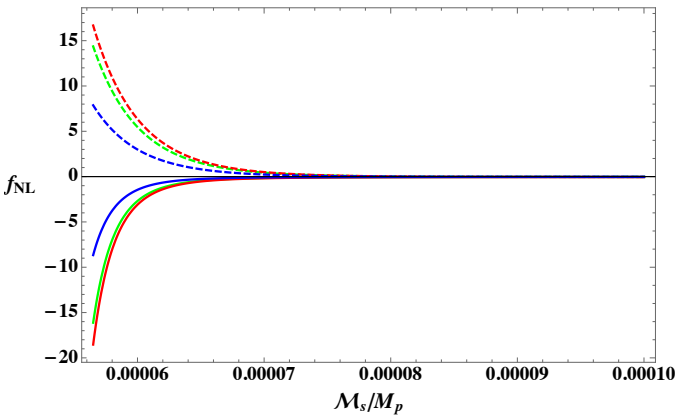


Fig. 6. In the above plots, f_{NL} versus the scale of non-locality $\mathcal{M}_s$ (in the units of M_p) is depicted for squeezed (blue), equilateral (red), and orthogonal (green) configurations for the polynomial entire functions γ_T given by eqs. (4.19) and (4.20) of arXiv:2003.00629 and represented by solid and dashed lines respectively. Here $N = 55$ of e-foldings is assumed. In the limit $\mathcal{M}_s \to M_p$ the predictions of the local R^2 model are recovered.

ratio[2, 20, 21] $r < 0.036$. As a result of modification of tensor power spectrum the tensor tilt in this theory is modified and ultimately tensor consistency relation $r = -8n_t$ is violated. Note that the violation of consistency relation is purely a non-local effect, the sound speed of primordial fluctuations are unity in this theory.

- The blue tensor tilt $n_t > 0$ is possible with non-local R^2-like inflation. Since not many frameworks of inflation can predict blue tilt any detection of it automatically favours non-local theory. Also $n_t < 0$ can also equally possible too.
- Non-local nature of gravity leads to detectable range of non-Gaussianities especial one can have squeezed limit of $f_{NL}^{\rm eq} \sim O(1)$ despite there is only one scalar degree of freedom and nature of inflation is standard-slow-roll. Further more, other shapes of f_{NL} can also be large despite the sound speed of primordial scalar field is unity.

In summary, non-local nature of gravity brings very unique predictions which are indeed very interesting in the view of future observations as well as further theoretical investigations.

Acknowledgments

KSK acknowledges financial support from JSPS and KAKENHI Grant-in-Aid for Scientific Research No. JP20F20320.

References

1. Planck, Y. Akrami *et al.*, (2018), 1807.06211.
2. BICEP, Keck, P. A. R. Ade *et al.*, Phys. Rev. Lett. **127**, 151301 (2021), 2110.00483.
3. Planck, Y. Akrami *et al.*, (2019), 1905.05697.
4. LIGO Scientific, Virgo, B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016), 1602.03837.
5. A. Ricciardone, J. Phys. Conf. Ser. **840**, 012030 (2017), 1612.06799.
6. A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).
7. A. A. Starobinsky, Nonsingular Model of the Universe with the Quantum Gravitational De Sitter Stage and its Observational Consequences, in **Moscow 1981, Proceedings of the second seminar "Quantum Theory of Gravity"*, 58-72*, 1981.
8. A. A. Starobinsky, JETP Lett. **34**, 438 (1981).
9. A. A. Starobinsky, Reheating in R^2 gravity, in *Fundamental Interactions, MGPI Press, Moscow, 1984, p. 55-79*, 1984.
10. A. H. Guth, Phys. Rev. **D23**, 347 (1981).
11. A. D. Linde, Phys. Lett. **B129**, 177 (1983).
12. A. Linde, Inflationary Cosmology after Planck 2013, in *Proceedings, 100th Les Houches Summer School: Post-Planck Cosmology: Les Houches, France, July 8 - August 2, 2013*, pp. 231–316, 2015, 1402.0526.
13. S. K. Korumilli, *Inflaton candidates: from string theory to particle physics*, PhD thesis, UBI, Covilha, 2017, 1808.03701.
14. J. Martin, C. Ringeval, and V. Vennin, Phys. Dark Univ. **5-6**, 75 (2014), 1303.3787.
15. J. M. Maldacena, JHEP **05**, 013 (2003), astro-ph/0210603.
16. P. D. Meerburg *et al.*, (2019), 1903.04409.
17. N. Arkani-Hamed and J. Maldacena, (2015), 1503.08043.
18. A. S. Koshelev, L. Modesto, L. Rachwal, and A. A. Starobinsky, JHEP **11**, 067 (2016), 1604.03127.
19. A. S. Koshelev, K. Sravan Kumar, and A. A. Starobinsky, JHEP **03**, 071 (2018), 1711.08864.
20. A. S. Koshelev, K. Sravan Kumar, A. Mazumdar, and A. A. Starobinsky, JHEP **06**, 152 (2020), 2003.00629.
21. A. S. Koshelev, K. S. Kumar, and A. A. Starobinsky, Int. J. Mod. Phys. D **29**, 2043018 (2020), 2005.09550.
22. M. J. Duff, Class. Quant. Grav. **11**, 1387 (1994), hep-th/9308075.
23. S. W. Hawking, T. Hertog, and H. S. Reall, Phys. Rev. **D63**, 083504 (2001), hep-th/0010232.
24. K. S. Stelle, Phys. Rev. **D16**, 953 (1977).
25. N. V. Krasnikov, Theor. Math. Phys. **73**, 1184 (1987), [Teor. Mat. Fiz.73,235(1987)].
26. S. V. Ketov and A. A. Starobinsky, Phys. Rev. **D83**, 063512 (2011), 1011.0240.
27. K. Abazajian *et al.*, (2019), 1907.04473.
28. CMB-S4, K. N. Abazajian *et al.*, (2016), 1610.02743.
29. P. D. Meerburg, M. Munchmeyer, J. B. Muñoz, and X. Chen, JCAP **1703**, 050 (2017), 1610.06559.

Palatini kinetic scalar-tensor theory: Analytical and numerical solutions

D. V. Gal'tsov* and D. S. Bushuev†

Faculty of Physics, Moscow State University, 119899, Moscow, Russia
**E-mail: galtsov@phys.msu.ru*
†E-mail:bushuev.ds16@physics.msu.ru

We investigate static spherically symmetric solutions in the Palatini kinetically coupled scalar-tensor theory, which reduces to gravity minimally coupled to a scalar field in Einstein frame. Using the fact that the Jordan and Einstein frame are related by a reversible disformal transformation, which can be solved in closed form, we derive the general solution in the Jordan frame and show that it does not contain black holes. There is a wormhole branch and a naked singularity branch between which lies a non-singular asymptotically flat solution with kernel $M_{1,1} \times S^2$. This theory strongly violates the null energy condition.

Keywords: Black holes, wormholes, Palatini scalar-tensor theories

1. Introduction

Recently[1] was discovered a particularly simple, albeit nontrivial, example of scalar-tensor theory with kinetic coupling, which strongly violates the null energy condition. This theory is a Palatini version of the well-known model with a scalar field coupled to Ricci tensor and Ricci scalar through derivatives.[2] With a certain ratio of the two coupling constants, the metric version of this theory belongs to Horndeski class,[3–5] and has an attractive feature to generate inflationary cosmological solutions without scalar potentials.[6–9]

The situation is different in the Palatini version of this theory, what is typical for scalar-tensor theories in general.[10–14] In this case, another ratio of the two coupling constants gives a theory admitting an Einstein frame in which the scalar coupling is minimally coupled. Moreover, the Einstein metric turns out to be the metric, the Levi-Civita connection of which is the Palatini connection of the original theory. The two metrics are related by an invertible disformal transformation.[15–17] The reversibilty of the disformal transformations ensures classical equivalence of the theories in two frames,[18–20] which probably extends to the quantum level.[21] It also opens a way to use disformal transformations as a generating tool to construct exact solutions of new theories.[1,22] Some examples of solutions of Palatini kinetic theory were presented in[1] demonstrating desingularization of solutions which are singular in the Einstein frame. For deeper discussion of singularities of general relativity in the context of modified gravity theories see.[23,24]

Further investigation of[25] showed a connection between this theory and the metric conformally coupled theory, both of which demonstrate a strong violation

of the null energy condition. The natural question is whether our theory allows for wormhole solutions.[26,27] Here we show that this is indeed the case.

2. The setup

We consider the action with two independent couplings of the derivatives of the scalar field $\phi_\mu \equiv \phi_{,\mu}$ to the Ricci tensor and the Ricci scalar[2,6–8]

$$S = \int d^4x\sqrt{-g}\left[R - \left(g_{\mu\nu} + \kappa_1 g_{\mu\nu}R + \kappa_2 R_{\mu\nu}\right)\phi^\mu\phi^\nu\right], \tag{1}$$

where $\phi^\mu = \phi_\nu g^{\mu\nu}$. In the metric formalism, an interesting case is $\kappa_1 = -\kappa_2/2$ when the Ricci tensor and Ricci scalar are combined into the Einstein tensor, which has zero covariant divergence. This case belongs to the Horndeski ghost-free class.

In the Palatini version, the action will read

$$S = \int d^4x\sqrt{-g}\big[(\hat{R}_{\mu\nu} - \phi_\mu\phi_\nu)g^{\mu\nu} - \hat{R}_{\alpha\beta}\phi_\mu\phi_\nu(\kappa_1 g^{\alpha\beta}g^{\mu\nu} + \kappa_2 g^{\alpha\mu}g^{\beta\nu})\big], \tag{2}$$

where $\hat{R}_{\alpha\beta}$ is the Ricci-tensor constructed with the Palatini connection $\hat{\Gamma}$. The corresponding Einstein equations are

$$\lambda\hat{R}_{\mu\nu} - \phi_\mu\phi_\nu(1 + \kappa_1\hat{R}) - 2\kappa_2\hat{R}_{\alpha(\mu}\phi_{\nu)}\phi^\alpha - g_{\mu\nu}L/2 = 0, \tag{3}$$

where the Lagrangian and the factor λ are

$$L = \hat{R}_{\mu\nu}Z^{\mu\nu} - \phi_\mu\phi_\nu g^{\mu\nu}, \;\; Z^{\mu\nu} = \lambda g^{\mu\nu} - \kappa_2\phi^\mu\phi^\nu, \;\; \lambda = (1 - \kappa_1 X), \tag{4}$$

and $X = \phi_\alpha\phi_\beta g^{\alpha\beta}$. Variation over ϕ gives rise to a scalar equation

$$\partial_\mu\left[\sqrt{-g}\left(\phi^\mu + \kappa_1\hat{R}\phi^\mu + \kappa_2\hat{R}_{\alpha\beta}g^{\beta\mu}\phi^\alpha\right)\right] = 0, \tag{5}$$

and variation over an independent connection gives:

$$\hat{\nabla}_\lambda\left(\sqrt{-g}Z^{\mu\nu}\right) = 0, \tag{6}$$

where the covariant derivative refers to the $\hat{\Gamma}$-connection.

Direct solution of the Einstein-scalar equations (3,5) does not seem possible. But as was shown in,[1] the Palatini equation (6) can be solved giving the Levi-Civita connection

$$\hat{\Gamma}^\lambda_{\mu\nu} = \hat{g}^{\lambda\tau}\left(\partial_\mu\hat{g}_{\lambda\nu} + \partial_\mu\hat{g}_{\mu\lambda} - \partial_\lambda\hat{g}_{\mu\nu}\right)/2 \tag{7}$$

of the second metric

$$\hat{g}_{\mu\nu} = \sqrt{\Lambda\lambda}\left(g_{\mu\nu} + \kappa_2\Lambda^{-1}\phi_\mu\phi_\nu\right), \quad \Lambda = 1 - (\kappa_1 + \kappa_2)X. \tag{8}$$

This is a disformal transformation, which can be inverted (for details see[1]). It is clear that setting $\kappa_1 + \kappa_2 = 0$, will greatly simplify the above relation, and we assume that from now on. Then, remarkably, when expressed in the second metric, the action (2) is transformed into an Einstein-Hilbert action plus a minimally coupled scalar:

$$S_E = \int\sqrt{-\hat{g}}\left[R_{\mu\nu}(\hat{g}) - \phi_\mu\phi_\nu\right]\hat{g}^{\mu\nu}d^4x. \tag{9}$$

This action gives (setting torsion to zero) the same equations in the Palatini version and in the metric version:

$$R_{\mu\nu} = \phi_\mu \phi_\nu, \qquad \hat{\Box}\phi = 0. \tag{10}$$

So our strategy will be first to integrate the system (10) and then to use an inverse of the disformal transformation (8) to obtain the desired metric $g_{\mu\nu}$. To be sure that we get the general solutions in the Jordan frame we need a general solution in the Einstein frame.

3. Einstein frame: Complete integration

Static solutions of the minimally coupled scalar-tensor gravity were investigated in the past in various contexts.[28–31] The most popular solution belongs to Fisher.[28] To make sure that it is unique in the class of static spherically symmetric solutions, we repeat the derivation here.

It is convenient to present the metric ansatz as

$$ds^2 = Adt^2 + \frac{dr^2}{A} + R(d\theta^2 + \sin^2\theta d\phi^2), \tag{11}$$

where A and R are functions of r only, and it is assumed that $R > 0$ (note that we are using R and not R^2 in the angular part). The non-zero components of the Ricci tensor are

$$R_{tt} = \frac{A}{2R}(A'R)', \tag{12}$$

$$R_{rr} = -\frac{1}{2AR^2}(2ARR'' + A'RR' - AR'^2 + A''R^2), \tag{13}$$

$$R_{\theta\theta} = \frac{1}{\sin^2\theta}R_{\phi\phi} = -\frac{1}{2}(AR')' + 1, \tag{14}$$

where the prime denotes the radial derivative. The scalar field satisfies the simple divergence equation

$$(AR\phi')' = 0. \tag{15}$$

With our choice of the action, the Einstein equations read $R_{\mu\nu} = \phi_\mu\phi_\nu$, so the equations for (tt) and $(\theta\theta)$ components of the Ricci tensor are sourceless and can be integrated once, giving

$$A'R = C = \text{const}; \quad AR' = 2(r + a), \quad a = \text{const}.$$

Similarly, integrating the scalar equation we get

$$\phi' = \frac{q}{AR},$$

where the integration constant q represents the scalar charge.

From the (rr) component of the Einstein equation we obtain a constraint

$$(2R - CR')A - 2(r + a)^2 + q^2 = 0,$$

Finally, we derive a separate equation for R:

$$\frac{2}{R'} - \frac{2(r+a)R''}{R'^2} = \frac{C}{R}.$$

To solve it, we put $R = e^u$, obtaining a non-linear equation for u:

$$u'' + u'^2 - \frac{u'}{x} + \frac{Cu'^2}{2x} = 0,$$

where $x = r + a$. This equation can be integrated once, giving

$$u' = \frac{2x}{x^2 + Cx + 2C_1}$$

with some new integration constant C_1. This constant, using the constraint equation, is reduced to $C_1 = -q^2/4$. Further integration gives

$$u = \ln\left(2x^2 + 2Cx - q^2\right) + \frac{2C}{\sqrt{C^2 + 2q^2}} \tanh^{-1} \frac{2x + C}{\sqrt{C^2 + 2q^2}} + u_0,$$

where u_0 is a trivial integration constant which can be changed by angular coordinates rescaling, so we put $u_0 = -\ln 2$. Rewriting the quadratic function under logarithm as

$$x^2 + Cx - \frac{q^2}{2} = (x - x_+)(x - x_-), \quad x_\pm = -\frac{1}{2}\left(C \pm \sqrt{C^2 + 2q^2}\right).$$

To get the metric function R, one has to take the power that leads to:

$$R = e^u = (x - x_-)(x - x_+)\left(\frac{x - x_-}{x - x_+}\right)^\gamma = r^2\left(1 - \frac{b}{r}\right)^{1-\gamma}, \tag{16}$$

where we have introduced two new parameters instead of C, q^2:

$$\gamma = \frac{C}{\sqrt{C^2 + 2q^2}} \quad b = \sqrt{C^2 + 2q^2}, \tag{17}$$

and a new radial coordinate $r = x - x_-$. The metric function A in the new variables reads

$$A = \frac{2(r + x_-)}{R'}.$$

Differentiating (16), we find

$$A = \left(1 - \frac{b}{r}\right)^\gamma. \tag{18}$$

Summarizing, we obtained the famous Fisher-Janis-Newman-Winicour (FJNW) solution, proving its uniqueness within the static spherically symmetric class:

$$ds^2 = -\left(1 - \frac{b}{r}\right)^\gamma dt^2 + \left(1 - \frac{b}{r}\right)^{-\gamma} dr^2 + r^2\left(1 - \frac{b}{r}\right)^{1-\gamma} d\Omega. \tag{19}$$

The derivative of the scalar field in terms of b, γ reads:

$$\phi' = \frac{b\sqrt{1-\gamma^2}}{\sqrt{2}\, r(r-b)}. \tag{20}$$

Note that the range of the radial variable is chosen $r \in (b, \infty)$ to ensure positivity of the function R. The region $r < b$ corresponds to cosmological solution.[32]

4. Disformal transformation

To pass to the Jordan frame, it is necessary to carry out the inverse disformal transformation

$$g_{\mu\nu} = \hat{g}_{\mu\nu}\lambda^{-1/2} - \kappa\phi_\mu\phi_\nu, \tag{21}$$

where

$$\lambda = (g_{rr}/w)^{-2/3}, \quad w = A = \left(1 - \frac{b}{r}\right)^{-\gamma}, \tag{22}$$

and g_{rr} obeys the following equation:

$$\left(g_{rr} - \frac{2x}{3\sqrt{3}}\right)^3 = w g_{rr}, \quad x = \frac{3\sqrt{3}\kappa q^2}{2r^2(r-b)^2}. \tag{23}$$

A real solution of this equation reads

$$g_{rr} = \frac{2x}{3\sqrt{3}} + \frac{1}{\sqrt{3}} \begin{cases} 2w\cos\left(\frac{1}{3}\arccos x/w\right), & x < w, \\ w^{2/3}B + w^{1/3}B^{-1}, & x > w, \end{cases} \tag{24}$$

where

$$B = \left(x + \sqrt{x^2 - w^2}\right)^{1/3}.$$

Two other Jordan frame metric components are

$$g_{tt} = \hat{g}_{tt}\lambda^{-1/2} \qquad g_{\theta\theta} \equiv Q = \hat{g}_{\theta\theta}\lambda^{-1/2} = R\lambda^{-1/2}. \tag{25}$$

It is easy to see that the above transformation to the Jordan frame preserves asymptotic flatness. Indeed, when $r \to \infty$, the variables $x \to 0$, $w \to 1$, so that

$$g_{rr} \sim 1 + x/\sqrt{3} \Rightarrow \lambda = 1 + O(1/r^4),$$

and we get as $r \to \infty$:

$$g_{tt} \simeq -1 + \frac{\gamma b}{r} \quad g_{rr} \simeq 1 - \frac{\gamma b}{r} \quad g_{\theta\theta} \sim r^2. \tag{26}$$

5. Wormholes

Consider first an asymptotic behavior of the Jordan metric as $r \to b$. Without loss of generality, let the dimensionless parameter $\kappa q^2/b^4$ be equal to one. Then a careful analysis of the behavior of various terms for $x \to \infty$, $w \to 0$ assuming arbitrary values $0 < \gamma < 1$ gives the following leading approximation for small $u = (r/b - 1)$:

$$ds^2 = -u^{2(2\gamma-1)/3}dt^2 + \frac{b^2du^2}{u^2} + b^2u^{(1-2\gamma)/3}d\Omega. \tag{27}$$

As $r \to b$, the metric component $g_{\theta\theta} = Q(r)$ explodes if $\gamma > 1/2$, while it goes to zero for $\gamma < 1/2$. One can suspect that in the first case the intrinsic volume of a region close to $r = b$, namely $r \in [b + \delta, r_0]$, such that $\delta \ll r_0 - b \ll b$, will be infinite as $\delta \to 0$. Indeed,

$$V_\delta = 4\pi \int_{b+\delta}^{r_0} \sqrt{h}dr = \frac{12\pi}{2\gamma - 1} b^3 \left(\frac{\delta}{b}\right)^{(1-2\gamma)/3}.$$

Thus, for $\gamma > 1/2$ one has the second infinite region of three-space. This region is not flat, and not Ricci flat. It is not asymptotically flat as $r \to b$ either, moreover the scalar field diverges there. But one can suspect the existence of an asymmetric wormhole, connecting our AF-region with the second sheet to the left of the local minimum of $Q(r)$ (a turning point) at some point $r_t > b$, if exists:

$$\left.\frac{dQ}{dr}\right|_{r=r_t} = 0, \quad \left.\frac{d^2Q}{dr^2}\right|_{r=r_t} > 0, \tag{28}$$

which could correspond to a wormhole throat.

Numerical analysis shows that for $\gamma < 1/2$, $g_{\theta\theta}$ does not have a local minimum for all positive values of the coupling constant κ. For $\gamma > 1/2$ a local minimum arises, with r_t depending on γ and κ.

5.1. *Numerical solutions*

Though we obtained solutions in the Jordan frame analytically, the algebraic functions involved are difficult for understanding the nature of the metric, so we undertake numerical plotting. On Fig. 1, 2 we plot the metric functions for various values of $1/2 < \gamma < 1$. In this case a local minimum of $Q(r)$ arises at $r = r_t$, whose position depends on γ and κ. This corresponds to the throat radius of an asymmetric wormhole. The values r_t can be found numerically with any desired accuracy. For $\kappa = 2$, the radii of the throat for some selected γ are

γ	3/5	3/4	0,95	0,99
r_t/b	1.041	1.078	1.072	1.041

With growing κ, the throat radius increases, as can be seen from the left panel of Fig. 1. Left to the minimum, the radial function Q increases again and tends to

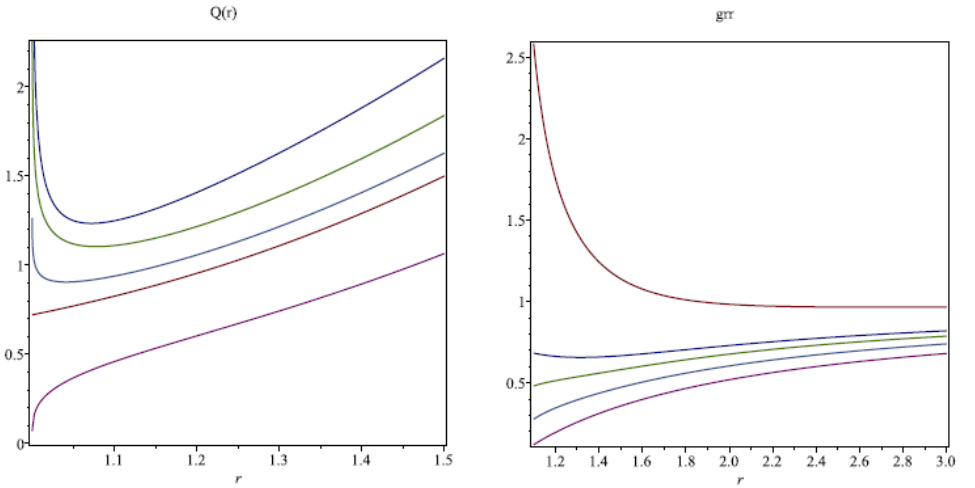


Fig. 1. Metric functions $g_{\theta\theta}$ and g_{tt} for various values of γ: $\gamma = 0.1$ (purple), $\gamma = 0.5$ (red), $\gamma = 0.6$ (light blue), $\gamma = 0.75$ (green), $\gamma = 0.95$ (blue).

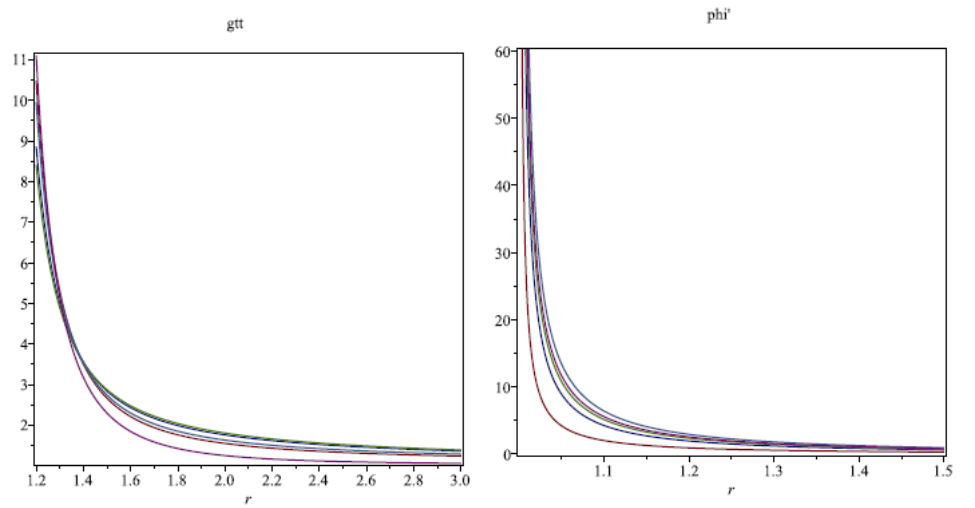


Fig. 2. Metric function g_{rr} and the derivative of the scalar field (ϕ') for various values of γ : $\gamma = 0.1$ (purple), $\gamma = 0.5$ (red), $\gamma = 0.6$ (light blue), $\gamma = 0.75$ (green), $\gamma = 0.95$ (blue).

infinity as $r \to b$ for $\gamma > 1/2$. From Fig. 2 (left panel) one can also see that $g_{tt} > 0$ for all $r > b$, so the solutions do not have horizons.

It can be shown that our wormhole is traversable: radial time-like and null geodesics freely pass through the throat. The tidal forces remain finite.

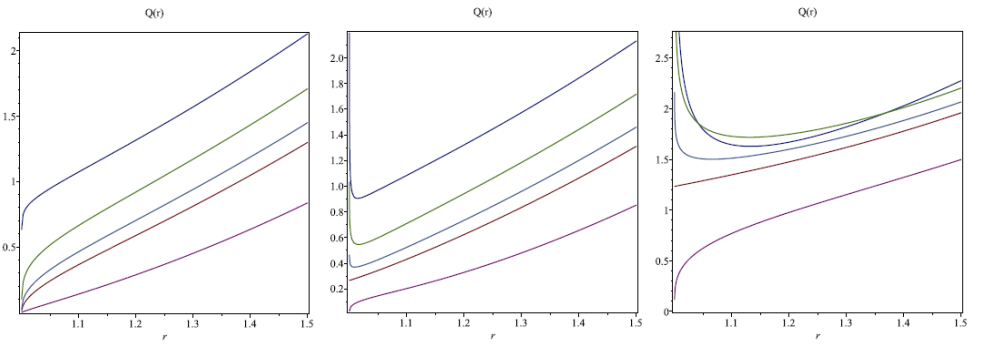


Fig. 3. Metric function ($g_{\theta\theta}$) for various values of κ: $\kappa = 0$ (left figure), $\kappa = 0.01$ (center), $\kappa = 10$ (right); and γ: $\gamma = 0.1$ (purple), $\gamma = 0.5$ (red), $\gamma = 0.6$ (light blue), $\gamma = 0.75$ (green), $\gamma = 0.95$ (blue).

5.2. *Geometry of the second sheet*

To investigate the geometry of the second leaf of the wormhole $r < r_t$, we first calculate the asymptotic behavior of the Ricci tensor and the Ricci scalar constructed from the Jordan metric (which is not the Palatini Ricci tensor):

$$R_{\mu\nu}(g)dx^\mu dx^\nu = -\frac{(1-2\gamma)^2}{6u^2}du^2 + d\Omega, \tag{29}$$

$$R(g) = \frac{2u^{(2\gamma-1)/3}}{b^2} - \frac{(1-2\gamma)^2}{6\mu^2}. \tag{30}$$

The rr-component of the Ricci tensor diverges as $u \to 0$, but the square of the Ricci tensor is finite for $\gamma > 1/2$:

$$R_{\mu\nu}R^{\mu\nu} = \frac{2u^{2(2\gamma-1)/3}}{b^4} + \frac{(2\gamma-1)^2}{36\,b^4}. \tag{31}$$

We then calculate the effective energy density and pressure by defining them in terms of the mixed components of the (diagonal) Einstein tensor, again constructed using the Jordan metric:

$$G^\nu_\mu(g) = (-\epsilon,\ p_r,\ p_\theta,\ p_\varphi). \tag{32}$$

This is not the physical energy density and pressure in Palatini's theory (see below), but only the effective ones used to illustrate geometry in terms of the notions of the Riemann geometry. The calculation gives for the energy density and radial pressure:

$$\epsilon = \frac{u^{(2\gamma-1)/3}}{b^2} - \frac{(2\gamma-1)^2}{12b^2}, \qquad p_r = -\frac{u^{(2\gamma-1)/3}}{b^2} - \frac{(2\gamma-1)^2}{12b^2}. \tag{33}$$

Their sum is negative, demonstrating violation of NEC in our theory. The tangential pressure is positive and finite

$$p_\theta = p_\varphi = (2\gamma-1)^2/12b^2. \tag{34}$$

In the sense of the Palatini theory, the Ricci scalar must be computed by contracting the Ricci tensor built from the Palatini connection, which is the Levi-Civita connection of Einstein's frame (i.e. $R_{\mu\nu}(\hat{g})$), with the Jordan contravariant metric tensor. Near the boundary $r = b$ the leading behavior of this quantity will be:

$$R_P = R_{\mu\nu}(\hat{g})\, g^{\mu\nu} = \frac{(1-2\gamma)^2}{6\,b^2}\frac{1}{u^{2-\gamma}}, \tag{35}$$

which diverges as $u \to 0$ for any $0 < \gamma < 1$. The question arises whether an observer will see the second sheet as a regular or a singular space. In fact, with independent connection, the curves of minimum length do not coincide with the curves along which the tangent vector is parallel transported. More generally, the matter action can depend only on the metric, or on the metric and the connection. The point particle action belongs to the first type, so the particle will move along the curves of minimal length. Such an observer will not see singularities on the internal sheet of the wormhole. If the notion of geodesic completeness is attributed to curves of minimal length, then this spacetime should be regarded as nonsingular indeed.

5.3. *Isotropic coordinates*

The asymptotic metric of the second sheet can be presented in isotropic coordinates t, ρ, θ, ϕ changing the radial variable as

$$\rho = b \exp\left(\nu^{-1} u^{\nu}\right), \quad \nu = (2\gamma - 1)/6.$$

The resulting metric reads

$$ds^2 = -\left(\nu \ln \frac{\rho}{b}\right)^4 dt^2 + f^2 dl_3^2, \quad dl_3^2 = d\rho^2 + \rho^2 d\Omega. \tag{36}$$

The conformal factor $f = \frac{b}{\rho\nu}\left(\ln \rho/b\right)^{-1}$, shown in Fig. 4 as a function of ρ, clearly demonstrates two asymptotic behavior of a wormhole:

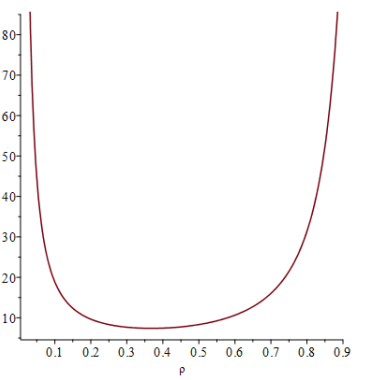


Fig. 4. The radial scale factor in isotropic coordinates.

5.4. *The soliton $\gamma = 1/2$*

In the Einstein frame the FJNW solution for $\gamma = 1/2$ satisfies the condition $g_{rr} = 1/|g_{tt}|$:

$$ds^2 = -\sqrt{1 - \frac{b}{r}}dt^2 + \frac{dr^2}{\sqrt{1 - \frac{b}{r}}} + r^2\sqrt{1 - \frac{b}{r}}d\Omega. \tag{37}$$

It has curvature invariants divergent as $r \to b$. But its Jordan frame counterpart turns out to be non-singular, namely

$$ds^2 = -dt^2 + \xi^{-2}dr^2 + b^2(d\theta^2 + \sin^2\theta d\varphi^2). \tag{38}$$

Passing to a new radial coordinate $z = b\ln\xi$, we transform the domain $r \in (b, \infty)$ into the complete real line $-\infty < z < \infty$

$$ds^2 = -dt^2 + dz^2 + b^2(d\theta^2 + \sin^2\theta d\varphi^2). \tag{39}$$

Therefore the manifold is isomorphic to the product of the two-dimensional Minkowski space and a sphere of radius b: $M_{1,1} \times S^2$. This manifold is geodesically complete in the Riemannian geometry sense. Thus, our solution can be regarded as a regular scalar-tensor soliton. Its striking feature is that, within the framework of our theory, it is supported by a singular scalar. The singularity now is transfered to the Palatini connection.

5.5. *Naked singularities*

The range of the power index $0 < \gamma < 1/2$ corresponds to Jordan metrics singular in the Riemannian sense. This can be seen already from inspection of the Ricci scalar constructed with the Jordan metric:

$$R(g) = \frac{2u^{(2\gamma-1)/3}}{b^2} - \frac{(1-2\gamma)^2}{6b^2}. \tag{40}$$

The constant term is due to the two-sphere sector of the metris, while the first term originates from radial component of the Ricci tensor and it tends to infinity for $\gamma < 1/2$.

Also, one can see that the three-volume of the region between $r = b - \delta$ and $r = b$ for $\gamma < 1/2$ tends to zero when $\delta \to 0$:

$$V_\delta = 12\pi b^3\,(\delta/b)^{(1-2\gamma)/3}. \tag{41}$$

The metric function $g_{\theta\theta}$ for various κ and γ is plotted in Fig. 3. From these plots one can see that in all three cases the metric has no horizons. More general numerical tests show that the static spherical sector of our theory does not contain black holes indeed.

5.6. *Classification of solutions*

Since our solution in the Einstein frame is generic in its class, we conclude that the above three types of static spherically symmetric solutions in the Jordan frame exhaust all possibilities, namely:

a) $\gamma > 1/2$ — asymmetric wormholes,
b) $\gamma = 1/2$ — regular soliton with the $M_{1,1} \times S_2$ core,
c) $\gamma < 1/2$ — asymptotically flat spacetime with a naked sigularity.

No solutions contain horizons.

Finally, we have two options for introducing the Kretschman scalar to investigate non-Ricci singularities. The Riemannian type scalar is simply $K_R = R_{\mu\nu\lambda\tau}(g)R^{\mu\nu\lambda\tau}(g)$. The Palatini-type Kretschmann scalar reads

$$K_P = R^{\mu}{}_{\nu\lambda\tau}(\hat{g})R^{\alpha}{}_{\beta\gamma\delta}(\hat{g})\, g_{\mu\alpha}\, g^{\nu\beta}\, g^{\lambda\gamma}\, g^{\tau\delta}.$$

In the asymptotic region of the second sheet (small u) the first one is:

$$K_R = \frac{1}{9\, b^4}\left(36u^{2(2\gamma-1)/3} - 2(2\gamma-1)^2 u^{(2\gamma-1)/3} + 12\gamma^2(\gamma-1)^2 + 6\gamma(\gamma-1) + 3/4\right). \tag{42}$$

It is finite in the cases a, b and singular in the case c as expected. The Palatini Kretschmann scalar always diverges as $u \to 0$:

$$K_P = \frac{(\gamma-1)^2(5\gamma^2 - 2\gamma + 2)}{4b^4\, u^{4(2-\gamma)/3}}, \tag{43}$$

reflecting singular nature of the Palatini connection.

6. Conclusion

To summarize: we have found that Palatini scalar-tensor theory with a scalar field kinetically coupled to the Ricci tensor and the Ricci scalar is an example of theory that does not admit solutions with an event horizon at least in the static case.

The theory has Palatini connection which is the Levi-Civita connection of the Einstein frame metic. The general static spherically symmetric solution in the Einstein frame is given by the Fisher family of metrics, specified by the power index $0 < \gamma < 1$. We have shown that for $1/2 < \gamma < 1$ the Jordan metric describes wormholes, interpolating between an asymptotically flat space and a Ricci non-flat second space of an infinite volume.The second sheet of the wormholes is not asymptotically flat. Its spacetime metric is regular, and the Riemannian Einstein tensor corresponds to an effective matter violating the null energy condition.

For $\gamma < 1/2$ the Jordan solutions are one-sheeted and represent naked singularities in asymptotically flat space. In the intermediate case $\gamma = 1/2$ the Jordan metric is non-singular, and have a core of cylindrical topology.

The work is supported by the Russian Foundation for Basic Research on the project 20-52-18012Bulg-a, and the Scientific and Educational School of Moscow State University “Fundamental and Applied Space Research”.

References

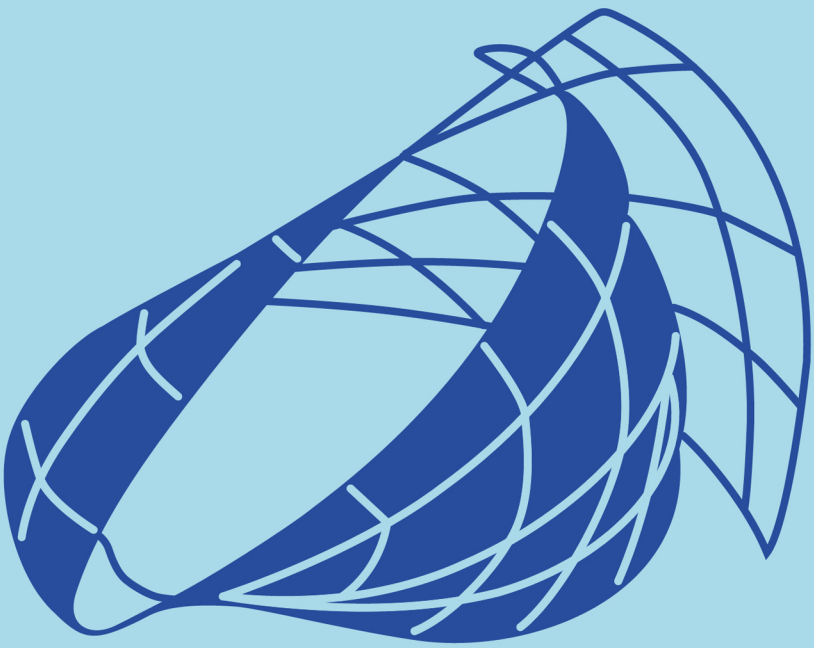
1. D. Gal'tsov and S. Zhidkova, "Ghost-free Palatini derivative scalar-tensor theory: Desingularization and the speed test," Phys. Lett. B **790** (2019) 453 [arXiv:1808.00492 [hep-th]].
2. L. Amendola, "Cosmology with nonminimal derivative couplings," Phys. Lett. B **301**, 175 (1993).
3. G. W. Horndeski, "Second-order scalar-tensor field equations in a four-dimensional space," Int. J. Theor. Phys. **10**, 363-384 (1974).
4. T. Kobayashi, "Horndeski theory and beyond: a review," Rept. Prog. Phys. **82**, no. 8, 086901 (2019) [arXiv:1901.07183 [gr-qc]].
5. C. Deffayet, X. Gao, D. A. Steer and G. Zahariade, "From k-essence to generalised Galileons," Phys. Rev. D **84** (2011), 064039 [arXiv:1103.3260 [hep-th]].
6. S. Capozziello and G. Lambiase, "Nonminimal derivative coupling and the recovering of cosmological constant," Gen. Rel. Grav. **31** (1999) 1005 [gr-qc/9901051].
7. S. V. Sushkov, "Exact cosmological solutions with nonminimal derivative coupling," Phys. Rev. D **80** (2009), 103505 [arXiv:0910.0980 [gr-qc]].
8. L. N. Granda and W. Cardona, "General Non-minimal Kinetic coupling to gravity," JCAP **1007**, 021 (2010). [arXiv:1005.2716 [hep-th]].
9. L. N. Granda and D. F. Jimenez, "Dynamical analysis for a scalar-tensor model with kinetic and nonminimal couplings," Int. J. Mod. Phys. D **27** (2017) no. 03, 1850030 [arXiv:1710.07273 [gr-qc]].
10. T. P. Sotiriou, "Unification of inflation and cosmic acceleration in the Palatini formalism," Phys. Rev. D **73**, 063515 (2006). [gr-qc/0509029].
11. T. Harko, T. S. Koivisto and F. S. N. Lobo, "Palatini formulation of modified gravity with a nonminimal curvature-matter coupling," Mod. Phys. Lett. A **26**, 1467 (2011). [arXiv:1007.4415 [gr-qc]].
12. X. Luo, P. Wu and H. Yu, "Non-minimal derivatively coupled quintessence in the Palatini formalism," Astrophys. Space Sci. **350**, no. 2, 831 (2014).
13. T. Helpin and M. S. Volkov, "Varying the Horndeski Lagrangian within the Palatini approach," arXiv:1906.07607 [hep-th].
14. N. Kaewkhao and B. Gumjudpai, "Cosmology of non-minimal derivative coupling to gravity in Palatini formalism and its chaotic inflation," Phys. Dark Univ. **20**, 20 (2018).
15. J. D. Bekenstein, "The Relation between physical and gravitational geometry," Phys. Rev. D **48**, 3641 (1993). [gr-qc/9211017].
16. D. Bettoni and S. Liberati, "Disformal invariance of second order scalar-tensor theories: Framing the Horndeski action," Phys. Rev. D **88**, 084020 (2013). [arXiv:1306.6724 [gr-qc]].
17. J. Sakstein and S. Verner, "Disformal Gravity Theories: A Jordan Frame Analysis," Phys. Rev. D **92**, no. 12, 123005 (2015). [arXiv:1509.05679 [gr-qc]].
18. G. Domenech, S. Mukohyama, R. Namba, A. Naruko, R. Saitou and Y. Watanabe, "Derivative-dependent metric transformation and physical degrees of freedom," Phys. Rev. D **92**, no. 8, 084027 (2015). [arXiv:1507.05390 [hep-th]].
19. K. Takahashi, H. Motohashi, T. Suyama and T. Kobayashi, "General invertible transformation and physical degrees of freedom," Phys. Rev. D **95**, no. 8, 084053 (2017). [arXiv:1702.01849 [gr-qc]].
20. E. Babichev, K. Izumi, N. Tanahashi and M. Yamaguchi, "Invertible field transformations with derivatives: necessary and sufficient conditions," [arXiv:1907.12333 [hep-th]].

21. A. Y. Kamenshchik and C. F. Steinwachs, "Question of quantum equivalence between Jordan frame and Einstein frame," Phys. Rev. D **91**, no. 8, 084033 (2015) [arXiv:1408.5769 [gr-qc]].
22. J. Ben Achour, H. Liu and S. Mukohyama, "Hairy black holes in DHOST theories: Exploring disformal transformation as a solution-generating method," JCAP **02** (2020), 023 [arXiv:1910.11017 [gr-qc]].
23. C. J. Fewster and G. J. Galloway, "Singularity theorems from weakened energy conditions," Class. Quant. Grav. **28** (2011) 125009 [arXiv:1012.6038 [gr-qc]].
24. C. J. Fewster and E. A. Kontou, "A new derivation of singularity theorems with weakened energy hypotheses," Class. Quant. Grav. **37** (2020) no. 6, 065010 [arXiv:1907.13604 [gr-qc]].
25. D. V. Gal'tsov, "Conformal and kinetic couplings as two Jordan frames of the same theory: Conformal and kinetic couplings," Eur. Phys. J. C **80** (2020) no. 5, 443 [arXiv:2001.03221 [gr-qc]].
26. V. A. Rubakov, "The Null Energy Condition and its violation," Phys. Usp. **57**, 128 (2014) [Usp. Fiz. Nauk **184**, no. 2, 137 (2014)] [arXiv:1401.4024 [hep-th]].
27. Mandal, S. "Revisiting Laws of Black Hole Mechanics and Violation of Null Energy Condition". Journal of High Energy Physics, Gravitation and Cosmology, 5, 82-111 (2019) arXiv:1907.12333 [hep-th].
28. I. Z. Fisher, "Scalar mesostatic field with regard for gravitational effects" Zh. Exp. Teor. Fiz., v. 18, 636-640 (1948), translated in arXiv: gr-qc/9911008.
29. O. Bergmann and R. Leipnik, *Space-Time Structure of a Static Spherically Symmetric Scalar Field*, Phys. Rev. **107** (1957) 1157.
30. A. I. Janis, E. T. Newman, J. Winicour, "Reality of the Schwarzschild singularity" Phys. Rev. Lett. **20**, 878 (1968).
31. K. A. Bronnikov, "Scalar-tensor theory and scalar charge," Acta Phys. Polon. B **4**, 251 (1973).
32. S. Abdolrahimi and A. A. Shoom, "Analysis of the Fisher solution," Phys. Rev. D **81** (2010), 024035 [arXiv:0911.5380 [gr-qc]].

PART B

PROCEEDINGS OF THE SIXTEENTH

MARCEL GROSSMANN MEETING ON GENERAL RELATIVITY



Editors

Remo Ruffini

Gregory Vereshchagin

World Scientific

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

This page intentionally left blank

PART B

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

Proceedings of the MG16 Meeting
on General Relativity
Online, 5–10 July 2021

Editors

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Gregory Vereshchagin
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Series Editor

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Names: Marcel Grossmann Meeting on General Relativity (16th : 2021 : Online) | Ruffini, Remo, editor. | Vereshchagin, Gregory, editor.

Title: The sixteenth Marcel Grossmann meeting on recent developments in theoretical and experimental general relativity, astrophysics, and relativistic field theories : proceedings of the MG16 meeting on general relativity online, 5–10 July 2021 / editors: Remo Ruffini, University of Rome "La Sapienza", Rome, Italy, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy, Gregory Vereshchagin, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy.

Other titles: Proceedings of the MG16 meeting on general relativity online, 5–10 July 2021

Description: New Jersey : World Scientific, [2023] | Includes bibliographical references.

Identifiers: LCCN 2022047088 | ISBN 9789811269769 (set ; hardcover) | ISBN 9789811266584 (v. 1) | ISBN 9789811266591 (v. 2) | ISBN 9789811266607 (v. 3) | ISBN 9789811266614 (v. 4) | ISBN 9789811269776 (set ; ebook)

Subjects: LCSH: General relativity (Physics)--Congresses. | Gravitation--Congresses. | Quantum gravity--Congresses. | Cosmology--Congresses. | Astrophysics--Congresses.

Classification: LCC QC173.6 .M37 2021 | DDC 523.01--dc23/eng/20220929

LC record available at https://lccn.loc.gov/2022047088

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2023 by Editors

All rights reserved.

This is an Open Access volume published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 (CC BY-NC) License. Further distribution of this work is permitted, provided the original work is properly cited.

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/13149#t=suppl

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

THE MARCEL GROSSMANN MEETINGS

Series Editor: REMO RUFFINI

Publications in the Series of Proceedings

Proceedings of the Sixteenth Marcel Grossmann Meeting on General Relativity
(Virtual Meeting, 2021)
Edited by G. Vereshchagin, R. Ruffini
World Scientific, 2022

Proceedings of the Fifteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2018)
Edited by E.S. Battistelli, R.T. Jantzen, R. Ruffini
World Scientific, 2022

Proceedings of the Fourteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2015)
Edited by M. Bianchi, R.T. Jantzen, R. Ruffini
World Scientific, 2017

Proceedings of the Thirteenth Marcel Grossmann Meeting on General Relativity
(Stockholm, Sweden, 2012)
Edited by K. Rosquist, R.T. Jantzen, R. Ruffini
World Scientific, 2015

Proceedings of the Twelfth Marcel Grossmann Meeting on General Relativity
(Paris, France, 2009)
Edited by T. Damour, R.T. Jantzen, R. Ruffini
World Scientific, 2012

Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity
(Berlin, Germany, 2006)
Edited by H. Kleinert, R.T. Jantzen, R. Ruffini
World Scientific, 2007

Proceedings of the Tenth Marcel Grossmann Meeting on General Relativity
(Rio de Janiero, Brazil, 2003)
Edited by M. Novello, S. Perez-Bergliaffa, R. Ruffini
World Scientific, 2005

Proceedings of the Ninth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2000)
Edited by V.G. Gurzadyan, R.T. Jantzen, R. Ruffini
World Scientific, 2002

Proceedings of the Eighth Marcel Grossmann Meeting on General Relativity
(Jerusalem, Israel, 1997)
Edited by T. Piran
World Scientific, 1998

Proceedings of the Seventh Marcel Grossmann Meeting on General Relativity
(Stanford, USA, 1994)
Edited by R.T. Jantzen and G.M. Keiser
World Scientific, 1996

Proceedings of the Sixth Marcel Grossmann Meeting on General Relativity
(Kyoto, Japan, 1991)
Edited by H. Sato and T. Nakamura
World Scientific, 1992

Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity
(Perth, Australia, 1988)
Edited by D.G. Blair and M.J. Buckingham
World Scientific, 1989

Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 1985)
Edited by R. Ruffini
World Scientific, 1986

Proceedings of the Third Marcel Grossmann Meeting on General Relativity
(Shanghai, People's Republic of China, 1982)
Edited by Hu Ning
Science Press – Beijing and North-Holland Publishing Company, 1983

Proceedings of the Second Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1979)
Edited by R. Ruffini
North-Holland Publishing Company, 1982

Proceedings of the First Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1975)
Edited by R. Ruffini
North-Holland Publishing Company, 1977

SPONSORS

International Center for Relativistic Astrophysics Network (ICRANet)
International Center for Relativistic Astrophysics (ICRA)

FREEDOM OF MOVEMENT FOR SCIENTISTS

The Marcel Grossmann Meetings were founded with the premise that scientists of all nations have a right to meet to exchange knowledge independent of national borders.

ACKNOWLEDGEMENTS

We acknowledge the outstanding job done before, during and after the meeting by the ICRANet/ICRA administrative and secretarial staff: Cristina Adamo, Silvia Latorre, Elisabetta Natale, and Cinzia di Niccolo. Finally this meeting and its proceedings could not have functioned without the dedicated IT support of the ICRANet system manager Gabriele Brandolini, with some temporary assistance from Domenico La Selva and Damiano Verzulli. We would like to thank Linda Kwan from World Scientific for valuable assistance during preparation of these proceedings.

ORGANIZING BODIES OF THE SIXTEENTH MARCEL GROSSMANN MEETING

INTERNATIONAL ORGANIZING COMMITTEE

Blair David, Choquet Bruhat Yvonne, Damour Thibault, De Bernardis Paolo, Everitt C. W. Francis, Fryer Chris, Haensch Theodor, Henneaux Marc, Jones Christine, Kerr Roy, Kleinert Hagen, Kunz Jutta, Laemmerzahl Claus, Longair Malcolm, Mirabel Felix, Mirzoyan Razmik, Piran Tsvi, Rueda Jorge, Ruffini Remo (chair), Sasaki Misao, Sato Humitaka, Sunyaev Rashid, 't Hooft Gerard, Weinberg Steven, Yau Shing-Tung, Zhang Bing

LOCAL ORGANIZING COMMITTEE

Adamo Cristina, Bianco Carlo Luciano, Brandolini Gabriele A., di Niccolo Cinzia, Latorre Silvia, La Selva Domenico, Li Liang, Loppini Alessandro, Natale Elisabetta, Verzulli Damiano, Vereshchagin Gregory (chair), Wang Yu

INTERNATIONAL COORDINATING COMMITTEE

ALBANIA: Hafizi M. - ARGENTINA: Arguelles C., Scoccola C., Reula O., Romero G.E. - ARMENIA: Sahakyan N. - AUSTRALIA: Blair D., Ju L., Lun A., Manchester D., Melatos A., Quinn P., Scott S.M., Steele J.D. - AUSTRIA: Aichelburg P.C., Schindler S. - BELARUS: Kilin S., Prakapenia M., Siutsou I. - BELGIUM: Henneaux M. - BOLIVIA: Aguirre C.B. - BOSNIA: Pasic V. - BRAZIL: Barres de Almeida U., Coelho Goulart J., Dalmolin F.T., de Lima Rafael C.R., Guzzo M., Maia C., Malheiro M., Romero Filho C.A., Shellard R.C., Zen Vasconcellos C. - BULGARIA: Yazadjiev S. - CANADA: Singh D., Smolin L., Turok N. - CHILE: Bauer F., Bunster W.C., Giacomini A. - CHINA (MAINLAND): Cai R., Cai Y., Cao Z., Chang J., Chen J., Chen X., Dai Z., Feng L.-L., Han W., Jing Y., Li T.-P., Lin W., Lou Y.-Q., Luo J., Mei J., Tam T., Wang A., Wang Y., Wu X.-P., Wu Y.-L., Yuan Y.-F., Zhang B.-B., Zhang S.-N., Zhao G. - CHINA (TAIWAN): Chen Chiang-Mei, Chen Pisin, Lee Da-Shin, Lee Wo-Lung, Ni Wei-Tou - COLOMBIA: Bargueño de Retes P., Gonzalez G., Higuera Garzon M.A., Núñez L., Romano A.E., Valenzuela Toledo C.A., Zuluaga J.I. - CROATIA: Dominis Prester D., Karlica M., Milekovic M., Smolcic V., Smolic I., Suric T. - CUBA: Perez Martinez A., Pérez Rojas H. - CZECH REPUBLIC: Bicak J., Stuchlik Z. - DENMARK: Naselsky P. - ECUADOR: Contreras E. - EGYPT: Tawfik A.N., Wanas M.I. - ESTONIA: Einasto J., Saar E. - FINLAND: Poutanen J., Volovik G. - FRANCE: Brillet A., Buchert T., Chardonnet P., Coullet P., de Freitas Pacheco J.A., Deruelle N., Iliopoulos J., Lamanna G., Mignard F. - GEORGIA: Lavrelashvili George, Machabeli Giorgi - GERMANY: Biermann P., Blumlein J., Di Piazza A., Fritzsch H., Genzel R., Gilmozzi R., Hasinger G., Hehl F., Keitel C., Kiefer C., Mirzoyan R.,

Neugebauer G., Nicolai H., Renn J., Ringwald A., Ruediger A. - GREECE: Batakis N.A., Cotsakis S., Vagenas E.C. - HUNGARY: Fodor G., Levai P. - ICELAND: Bjornsson G., Jakobsson P. - INDIA: Chakrabarti S.K., Iyer B., Padmanabhan T., Souradeep T. - IRAN: Baghram S., Bavarsad E., Eslam Panah B., Firouzjahi H., Haghighat M., Mansouri R., Mashhoon B., Shakeri S., Sobouti Y., Taghi Mirtorabi M. - IRELAND: O'Murchada N. - ISRAEL: Milgrom M., Nakar E., Pe'er A., Piran T. - ITALY: Belinski V., Bianchi M., Bianco C.L., Cherubini C., Della Valle M., Falciano S., Filippi S., Haardt F., Menotti P., Merafina M., Pani P., Ricci F., Treves A., Vereshchagin G.V., Vitale S., Xue S.- S. - JAPAN: Fujimoto M.-K., Makishima K., Nakamura T., Sato K., Shibata M. - KAZAKHSTAN: Abishev M., Aimuratov Y., Boshkayev K., Mychelkin E.G., Spitaleri C. - KOREA (PYEONGYANG): Kim J.S. - KOREA (SEOUL): Kim S.P., Kim S.-W., Lee H.K., Lee H.-W., van Putten M. - KYRGYZSTAN: Gurovich V.Ts. - LIBYA: Gadri M. - MEXICO: Breton N., Cervantes-Cota J.L., Fraija Cabrera N.I., García-Diaz A.A., Macías Alvarez A., Mielke Eckehard W., Quevedo H., Rodriguez L.F. - NETHERLANDS: Slagter R. - NEW ZEALAND: Visser M., Wiltshire D. - NORWAY: Elgaroy O., Fonseca Mota D., Knutsen H. - PAKISTAN: Qadir A., Qamar S. - PERU: Vargas T. - POLAND: Belczynski K., Demianski M., Lewandowski Jerzy, Nurowski P., Sokolowski L. - PORTUGAL: Costa M., Da Silva A., Lemos J.P.S., Lobo F., Moniz P., Silva L.O. - ROMANIA: Visinescu M. - RUSSIA: Aksenov A., Arkhangelskaja I., Bisnovatyi-Kogan G., Blinnikov S., Chechetikin V.M., Cherepaschuk A.M., Khriplovich I., Lipunov V.M., Lukash V.N., Novikov I., Rudenko V.N., Starobinsky A.A. - SERBIA: Djordjevic G., Jovanovic P., Knezevic Z., Pankov-Hzvojevic M., Popovic L., Prodanovic T., Sijacki D., Simic S. - SLOVAKIA: Balek V. - SLOVENIA: Cadez A., Gomboc A., Zavrtanik D. - SOUTH AFRICA: Larena J., Maharaj S. - SPAIN: Elizalde E., Ibanez J., Perez M.J., Verdaguer E. - SWEDEN: Abramowicz M.A., Marklund M., Ryde F. - SWITZERLAND: Durrer R., Jetzer P. - TURKEY: Aliev A., Gurses M. - UKRAINE: Novosyadlyj B., Zaslavskii O., Zhuk A. - UNITED ARAB EMIRATES: Fernini I. - UNITED KINGDOM: Cruise A.M., Frenk Carlos S., Green M., Mavromatos N., Perry M., Willingale R. - USA: Abel T., Ashtekar A., Bardeen J., Carlstrom J., Cornish N., Dermer C., Fan X., Flanagan E., Fraschetti F., Fryer C., Incera V., Jantzen R.T. (chairperson), Kolb R., Laguna P., Lousto C., Madau Piero, Mathews Grant, Matzner Richard, Melia Fulvio, Mester John, Michelson Peter, Nordtvedt Kenneth, Parker Leonard, Pretorius F., Pullin J., Shapiro I., Shapiro S., Shoemaker D., Smoot G., Stiavelli M., Teukolsky S., van Nieuwenhuizen P., Zhang B. - UZBEKISTAN: Ahmedov B., Zalaletdinov R.M. - VATICAN CITY: Gionti G. - VENEZUELA: Fuenmayor E. - VIETNAM: Long H.N.

MARCEL GROSSMANN AWARDS

Sixteenth Marcel Grossmann Meeting

Institutional Awards

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

Individual Awards

DEMETRIOS CHRISTODOULOU
"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

GERARD 't HOOFT
"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

TSVI PIRAN
"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".

STEVEN WEINBERG
"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Each recipient is presented with a silver casting of the TEST sculpture by the artist A. Pierelli. The original casting was presented to His Holiness Pope John Paul II on the first occasion of the Marcel Grossmann Awards.

15th Marcel Grossmann Meeting
July 2018, Rome, Italy

Institutional Awards

PLANCK SCIENTIFIC COLLABORATION (ESA)
"for obtaining important constraints on the models of inflationary stage of the Universe and level of primordial non-Gaussianity; measuring with unprecedented sensitivity gravitational lensing of Cosmic Microwave Background fluctuations by large-scale structure of the Universe and corresponding B-polarization of CMB, the imprint on the CMB of hot gas in galaxy clusters; getting unique information about the time of reionization of our Universe and distribution and properties of the dust and magnetic fields in our Galaxy"

- presented to Jean-Loup Puget, the Principal Investigator of the High Frequency Instrument (HFI)

HANSEN EXPERIMENTAL PHYSICS LABORATORY AT STANFORD UNIVERSITY
"to HEPL for having developed interdepartmental activities at Stanford University at the frontier of fundamental physics, astrophysics and technology"

- presented to Research Professor Leo Hollberg, HEPL Assistant Director

Individual Awards

LYMAN PAGE
"for his collaboration with David Wilkinson in realizing the NASA Explorer WMAP mission and as founding director of the Atacama Cosmology Telescope"

RASHID ALIEVICH SUNYAEV
"for the development of theoretical tools in the scrutinising, through the CMB, of the first observable electromagnetic appearance of our Universe"

SHING-TUNG YAU
"for the proof of the positivity of total mass in the theory of general relativity and perfecting as well the concept of quasi-local mass, for his proof of the Calabi conjecture, for his continuous inspiring role in the study of black holes physics"

14th Marcel Grossmann Meeting
July 2015, Rome, Italy

Institutional Award

EUROPEAN SPACE AGENCY (ESA)
"for the tremendous success of its scientific space missions in astronomy, astrophysics, cosmology and fundamental physics which have revolutionized our knowledge of the Universe and hugely benefited science and mankind"

- presented to its Director General Johann-Dietrich Woerner

Individual Awards

KEN'ICHI NOMOTO
"for heralding the role of binary systems in the evolution of massive stars"

MARTIN REES
"for fostering Research in black holes, gravitational waves and cosmology"

YAKOV G. SINAI
"for applying the mathematics of chaotic systems to physics and cosmology"

SACHIKO TSURUTA
"for pioneering the physics of hot neutron stars and their cooling"

FRANK C.N. YANG
"for deepening Einstein's geometrical approach to physics in the best tradition of Paul Dirac and Hermann Weyl"

T.D. LEE (award received by Yu-Qing Lou on behalf of Prof. T.D. Lee)
"for his work on white dwarfs motivating Enrico Fermi's return to astrophysics and guiding the basic understanding of neutron star matter and fields"

13th Marcel Grossmann Meeting
July 2012, Stockholm, Sweden

Institutional Award

ALBANOVA
for its innovative status as a joint institute established by Stockholm University and the Royal Institute of Technology and for fostering contributions to cosmology and astrophysics in the profound scientific tradition established by Oskar Klein.

- presented to the Rector of Stockholm University, Prof. Kåre Bremer.

Individual Awards

DAVID ARNETT
for exploring the nuclear physics and yet unsolved problems of the endpoint of thermonuclear evolution of stars, leading to new avenues of research in physics and astrophysics.

VLADIMIR BELINSKI and I.M. KHALATNIKOV
for the discovery of a general solution of the Einstein equations with a cosmological singularity of an oscillatory chaotic character known as the BKL singularity.

FILIPPO FRONTERA
for guiding the Gamma-ray Burst Monitor Project on board the BeppoSAX satellite, which led to the discovery of GRB X-ray afterglows, and to their optical identification.

12th Marcel Grossmann Meeting
July 2009, Paris, France

Institutional Award

INSTITUT DES HAUTES ÉSTUDES SCIENTIFIQUE (IHÉS)
for its outstanding contributions to mathematics and theoretical physics, and notably for having renewed basic geometrical concepts, and having developed new mathematical and physical aspects of spacetime.

- presented to Prof. Jean-Pierre Bourguignon

Individual Awards

JAAN EINASTO
for pioneering contributions in the discovery of dark matter and cosmic web and fostering research in the historical Tartu Observatory.

CHRISTINE JONES
for her fundamental contributions to the X-ray studies of galaxies and clusters tracing their formation and evolution and for her role in collaborations using clusters to study dark matter and in analyzing the effects of outbursts from supermassive black holes on the intracluster gas.

MICHAEL KRAMER
for his fundamental contributions to pulsar astrophysics, and notably for having first confirmed the existence of spin-orbit precession in binary pulsars.

11th Marcel Grossmann Meeting
July 2006, Berlin, Germany

Institutional Award

FREIE UNIVERSITÄT BERLIN
for the successful endeavor of re-establishing — in the spirit of the Humboldt tradition — freedom of thinking and teaching within a democratic society in a rapidly evolving cosmos

- presented to Dr. Dieter Lenzen, President of FUB

Individual Awards

ROY KERR
for his fundamental contribution to Einstein's theory of general relativity: "The gravitational field of a spinning mass as an example of algebraically special metrics."

GEORGE COYNE
for his committed support for the international development of relativistic astrophysics and for his dedication to fostering an enlightened relationship between science and religion.

JOACHIM TRUMPER
for his outstanding scientific contributions to the physics of compact astrophysical objects and for leading the highly successful ROSAT mission which discovered more than 200,000 galactic and extragalactic X-ray sources: a major step in the observational capabilities of X-ray astronomy and in the knowledge of our universe.

10th Marcel Grossmann Meeting
July 2003, Rio de Janeiro, Brazil

Institutional Award

CBPF (Brazilian Center for Research in Physics)
for its role as a teaching and research institution and as a place originating fundamental physics ideas in the exploration of the universe.

- presented to its founders Cesar Lattes, Josè Leite Lopez and Jayme Tiomno

Individual Awards

YVONNE CHOQUET-BRUHAT AND JAMES W. YORK, JR.
for separate as well as joint work in establishing the mathematical framework for proving the existence and uniqueness of solutions to Einstein's gravitational field equations.

YUVAL NE'EMAN
for his contributions to science, epistimology, mathematics and physics from subnuclear to space sciences.

9th Marcel Grossmann Meeting
July 2000, Rome, Italy

Institutional Award

SOLVAY INSTITUTES
for identifying and recording in discussions by the protagonists the crucial developments of physics and astrophysics in the twentieth century.

- presented to Jacques Solvay

Individual Awards

CECILLE AND BRYCE DEWITT
for promoting General Relativity and Mathematics research and inventing the "summer school" concept.

RICCARDO GIACCONI
for opening, five successive times, new highways for exploring the Universe.

ROGER PENROSE
for extending the mathematical and geometrical foundations of General Relativity.

8th Marcel Grossmann Meeting
June 1997, Jerusalem

Institutional Award

HEBREW UNIVERSITY
for its role as a cradle of Science and Humanities and for hosting the manuscripts of Albert Einstein.

- presented to M. Magidor, President of the Hebrew University of Jerusalem

Individual Awards

TULLIO REGGE
for his contributions to the interface between mathematics and physics leading to new fields of research of paramount importance in relativistic astrophysics and particle physics.

FRANCIS EVERITT
for leading the development of extremely precise space experiments utilizing superconducting technology to test General Relativity and the Equivalence Principle.

7th Marcel Grossmann Meeting
June 1994, Stanford, USA

Institutional Award

SPACE TELESCOPE SCIENCE INSTITUTE
for its critical role in the direction and operation of the Hubble Space Telescope, a truly unique international laboratory for the investigation and testing of general relativity in the context of modern astrophysics and cosmology.

- presented to Peter Stockman

Individual Awards

SUBRAHMANYAN CHANDRASEKHAR
for his contributions to the analysis of gravitational phenomena from Newton to Einstein and especially for leading the way to relativistic astrophysics with the concept of critical mass for gravitational collapse.

JIM WILSON
for having built on his experience in nuclear physics, thermonuclear reactions, and extensive numerical simulation to create a new testing ground for the novel concepts of relativistic astrophysics.

6th Marcel Grossmann Meeting
June 1991, Kyoto, Japan

Institutional Award

RITP
for keeping alive first in Hiroshima and them in Kyoto research in relativity, cosmology, and relativistic field theory and the development of a school of international acclaim.

- presented to Professor K. Tomita

Individual Awards

MINORU ODA
for participating in the pioneering work of the early sixties in X-ray astronomy and for his subsequent molding of an agile and diversified Japanese scientific space program investigating the deepest aspects of relativistic astrophysics.

STEPHEN HAWKING
for his contributions to the understanding of spacetime singularities and of the large scale structure of the Universe and of its quantum origins.

5th Marcel Grossmann Meeting
August 1988, Perth, Australia

Institutional Award

THE UNIVERSITY OF WESTERN AUSTRALIA
for its contributions to relativistic astrophysics.

- presented to the Vice Chancellor, Professor Robert Smith

Individual Awards

SATIO HAYAKAWA
for his contributions to research in gamma, X-ray and infrared radiation as well as cosmic rays.

JOHN ARCHIBALD WHEELER
for his contributions to geometrodynamics and Einstein's visions.

4th Marcel Grossmann Meeting
July 1985, Rome, Italy

Institutional Award

THE VATICAN OBSERVATORY
for its contributions to the origin and development of astrophysics.

- presented to His Holiness Pope John Paul II

Individual Awards

WILLIAM FAIRBANK
for his work in gravitation and low temperature physics.

ABDUS SALAM
for his work in unifying fundamental interactions.

Institutional Awards for the Spektrum-Roentgen-Gamma (SRG) mission

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

On Tuesday June 29, 2021, the following 31 astro-ph appeared:

1. https://arxiv.org/abs/2106.14517
2. https://arxiv.org/abs/2106.14518
3. https://arxiv.org/abs/2106.14519
4. https://arxiv.org/abs/2106.14520
5. https://arxiv.org/abs/2106.14521
6. https://arxiv.org/abs/2106.14522
7. https://arxiv.org/abs/2106.14523
8. https://arxiv.org/abs/2106.14524
9. https://arxiv.org/abs/2106.14525
10. https://arxiv.org/abs/2106.14526
11. https://arxiv.org/abs/2106.14527
12. https://arxiv.org/abs/2106.14528
13. https://arxiv.org/abs/2106.14529
14. https://arxiv.org/abs/2106.14530
15. https://arxiv.org/abs/2106.14531
16. https://arxiv.org/abs/2106.14532
17. https://arxiv.org/abs/2106.14533
18. https://arxiv.org/abs/2106.14534
19. https://arxiv.org/abs/2106.14535
20. https://arxiv.org/abs/2106.14536
21. https://arxiv.org/abs/2106.14537
22. https://arxiv.org/abs/2106.14541
23. https://arxiv.org/abs/2106.14542
24. https://arxiv.org/abs/2106.14543
25. https://arxiv.org/abs/2106.14544
26. https://arxiv.org/abs/2106.14545
27. https://arxiv.org/abs/2106.14546
28. https://arxiv.org/abs/2106.14547
29. https://arxiv.org/abs/2106.14548
30. https://arxiv.org/abs/2106.14549
31. https://arxiv.org/abs/2106.14550

S.A. LAVOCHKIN ASSOCIATION
presented to its Designer General **Alexander Shirshakov**

Dr Alexander Shirshakov

S.A. Lavochkin Association created the Navigator space platform carrying German eRosita and Russian ART-XC X-Ray Telescopes, organized the launch of SRG Orbital X-Ray Observatory to the second Lagrangian point of the Sun-Earth system at a distance of 1.5 million km from the Earth and managed the observatory flight and the daily reception of its scientific data on Earth for 23.5 months.

Dr Alexander Shirshakov, Designer General of the S.A. Lavochkin Association, is specialized in design, manufacture, testing, launch and control of S/C for scientific purposes. Among those S/C launched, there are the «Radiostron» Astrophysical Observatory (2011) and the «Spektr-RG» space observatory (2019), while the planned S/C launches are «Luna-25» and «Exomars».

Dr Shirshakov started his career in 1973, working as an engineer of the State Unitary Enterprise «NPO named by S.A. Lavochkin» in Khimki (Russian Federation). Starting from 1989 he has played multiple roles within the Lavochkin Association, been appointed head of the group, head of the sector, head of department, deputy head of the complex, head of the branch, director of the center, deputy head of the Design Bureau, deputy General Designer and deputy General Director.

Dr Shirshakov is an editorial board Member of the reviewed edition of «Vestnik of Lavochkin Association». Since 2017, he is also member of the General Designer council. He has been awarded Honored Mechanical engineer of the Russian Federation as well as Agency-level award of the Russian Federal Space Agency.

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

Professor Peter Predehl

eROSITA is the soft X-ray telescope on-board the Russian-German Spektr-RG mission which was successfully launched from Baikonur on July 13, 2019 and placed in a halo orbit around the L2 point. 30 years after ROSAT, eROSITA performs an all-sky survey with an unprecedented sensitivity, spectral and angular resolution. Clusters of galaxies are the largest collapsed objects in the Universe. Their formation and evolution is dominated by gravity, i.e. Dark Matter, while their large scale distribution and number density depends on the geometry of the Universe, i.e. Dark Energy. X-ray observations of clusters of galaxies provide information on the rate of expansion of the Universe, the fraction of mass in visible matter, and the amplitude of primordial fluctuations which are the origin of clusters of galaxies and the whole structure of the universe. eROSITA has been designed to detect at least 100.000 clusters of galaxies and to detect systematically more than 3 million obscured accreting Black Holes. eROSITA will also allow to study the physics of galactic X-ray source

populations, like pre-main sequence stars, supernova remnants and X-ray binaries. The eROSITA telescope consists of seven identical Wolter-1 mirror modules. A novel detector system has been developed by MPE on the basis of the successful XMM-Newton pn-CCD technology. MPE is the scientific lead institute of eROSITA, responsible for the development of the instrument, the operation, the analysis software and data archive. Peter Predehl led this development as Principal Investigator of eROSITA and German lead scientist of the SRG mission for more than 15 years until the completion of the first of eight surveys in 2020. At this time eROSITA has already discovered more than 1 million X-ray sources, more than all X-ray observatories of the last 50 years together. This demonstrates that the design goals of the mission will easily be fulfilled.

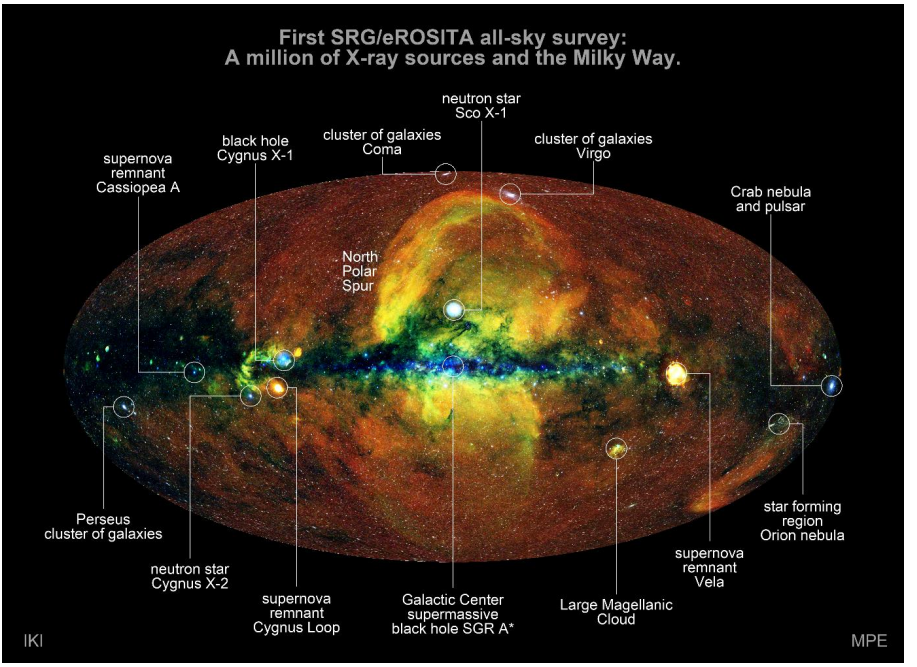
SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
presented to Professor **Rashid Sunyaev**

Professor Rashid Sunyaev

Space Research Institute (IKI) of the Russian Academy of Sciences was responsible for developing the overall concept and scientific program of the SRG Orbital observatory and played a leading role in developing the ART-XC telescope and the entire SRG observatory as part of the Russian space science program carried out by Roskosmos Corporation in the interests of the Russian Academy of Sciences.

During the flight to the L2 point of the Sun-Earth system, SRG with German (eRosita) and Russian (ART-XC named after Mikhail Pavlinsky) X-ray Telescopes aboard performed calibrations and long duration Performance Verification observations of a dozen of targets and deep fields. Starting in the middle of December 2019, the SRG scanned the whole sky three times. During these scans, SRG discovered two million point X-ray sources: mainly quasars, stars with hot and bright coronae, and more than 30 thousand clusters of galaxies. There is a competition and synergy in the search for clusters of galaxies between SRG and the ground-based Atacama Cosmology and South Pole Telescopes, which are searching for clusters of galaxies in microwave spectral band using Sunyaev-Zeldovich effect. SRG provided the X-Ray map of the whole sky in hard and soft bands, the last is now the best among existing. The huge samples of the X-ray selected quasars at the redshifts up to z = 6.2 and clusters of galaxies will be used for well-known cosmological tests and detailed study of the growth of the large scale structure of the Universe during and after reionization. SRG/eRosita is discovering every day several extragalactic objects which increased or decreased their brightness more than 10 times during half of the year after the previous scan of the same one-degree wide strip on the sky. A significant part of these objects has observational properties similar to the Events of Tidal Disruption of a star orbiting in the vicinity of the supermassive black hole. ART-XC discovered a lot of bright galactic and extragalactic transients.

Rashid Sunyaev is the Principal Investigator of SRG mission in Russia, director-emeritus of the Max-Planck Institute for Astrophysics and Maureen and John Hendricks distinguished visiting professor of the Institute for Advanced Study, Princeton.



Individual Awards

Professor **DEMETRIOS CHRISTODOULOU**

"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

Professor Demetrios Christodoulou

It was back in 1967 that Achille Papapetrou mentioned the case of the 16-year-old Demetrios Christodoulou to John Archibad Wheeler. Wheeler interviewed Demetrios in Paris and brought him immediately to Princeton where he was registered as an undergraduate at the university. After one year he entered the graduate school and started collaborating with me. At the time I was working with Wheeler on the effective potential approach to geodesics co-rotating and counter-rotating (see e.g. reference in The Classical Theory of Fields (Landau and Lifshitz, 1980) in the Kerr metric (later renamed as ISCO; see e.g. (Gravitation Misner, Thorne, Wheeler. 1973). In parallel, Frank Zerilli was working on the gravitational radiation emitted by the fall of a test particle in a Schwarzschild black hole (Zerilli 1970). From these limited conceptual arena Charles Misner and later Kip Thorne launched a program for the detection of gravitational waves on the Earth; see e.g. Misner 1974, Abbott et al. 2016, Abbott et al. 2017. See however Davis et al. 1972, Rodriguez et al. 2018 and J.A. Rueda et al. 2018.

A new approach started with the arrival of Demetrios: he was just creating mathematics following his needs. We identified the reversible and irreversible transformations of a Kerr black hole. Wheeler advanced a thermodynamic analogy. I addressed the need of identifying the concept of irreducible mass (from the Italian "irriducibile"), and was Demetrios's contribution to integrate, overnight, the differential equation for infinitesimal reversible transformations which led to the finite mass-energy formula of a Kerr black hole. That evening, while walking back home through IAS woods, I expressed to Wheeler the great relevance of the newly found formula by Demetrios and proposed to let Demetrios be the single author of this article, admiring his great mathematical talent. Wheeler agreed. The Editor of PRL objected since in that two pages article the Fig. 2 by Wheeler and myself was still unpublished. Actually that Fig. 2 followed a discussion I previously had with Penrose in Florence (Penrose 1961) which allowed us to present there, for the first time, a "Penrose Process". Some difficulties in achieving this process were obvious from the example in Fig. 2, which Roger later recognized himself (Penrose & Floyd 1971). The Editor finally agreed on our written request and the paper appeared on September 17, 1970 (Christodoulou, 1970). On January 1971 appeared my article with Johnny introducing the Black Hole (Ruffini & Wheeler, 1971), with the new physics we were developing in Princeton, including the concept of the "ergosphere". On march 1 1971 we submitted the mass formula of the Kerr Newmann metric, including the relation between the surface area of the horizon and the irreducible (Christodoulou & Ruffini, 1971). On March 11, 1971 the same results were independently confirmed by Steven Hawking, extending further the applicability of our equation (Hawking 1971).

The thesis was successfully discussed by a committee including Eugene Wigner (see Fig. 1), one of the closest collaborators of Albert Einstein and David Wilkinson (see Fig. 2), the head of the NASA WMAP mission, and Johnny and myself as supervisors. The new message was clear: Black Holes, far from being a sink of energy, were energy sources emitting "in principle" 50% of their mass energy, being extractable (Christodoulou & Ruffini, 1971).

Fig. 1 and Fig. 2: Demetrios during his thesis presentation with Eugene Wigner (Fig. 1) and David Wilkinson (Fig.2). Johnny and I were supervisors, ready to intervene in case of need, but no need of intervention was necessary! Wigner elaborated the aphorism of Niels Bohr "Interesting = wrong" in the most definite "very interesting if true = totally wrong".

Demetrios turned soon to the study of partial differential equations and mathematical physics, to which he dedicated for the rest of his career and results were published in four monographs: (Christodoulou and Klainerman 1994, Christodoulou 2007, Christodoulou 2009, Christodoulou 2019). In 1968, Johnny proposed to Demetrios the collapse of a "geon" composed of massless scalar field as a second topic for his thesis. It took almost forty years for him to solve this problem, extended by Demetrios to the focusing of gravitational waves leading to black hole formation (Christodoulou 2009).

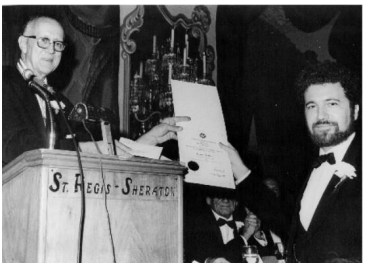


Fig. 3: Prof. Remo Ruffini receiving the Cressy Morrison Award of the New York Academy of Sciences, 1972 for the discovery of the first Black Hole in our galaxy Cygnus X1.

A "long march" started on 12 December 1970 with the launch of the Uhuru satellite by Riccardo Giacconi. Early in 1971 an almost daily conversation with him and Herb Gursky at the Smithsonian Astrophysical Observatory, leading to the discovery of binary X-ray sources. This was soon followedby the announcement of Cygnus X1 identified as the first black hole in our galaxy (Ruffini 1973); see e.g. Gursky & Ruffini 1975, which contained as well the first publicannouncement of the Discovery of Gamma Ray burst, as well as Giacconi & Ruffini 1980, 2009; see Figs. 3 and 4).

Fig. 4: In the second row, from left to right, there are, among others: E. T. Newman, S. Chandrasekhar (Nobel 1983), R. Giacconi (Nobel 2002), R. Ruffini, A. Treves, A. Hewish (Nobel 1974), D. Arnett, J.H. Taylor (Nobel 1993), J. Wilson, R. Penrose (Nobel 2020), as well as J. Bahcall, T. Damour, T. Piran et al.

Today, after fifty years, this "long march" has reached a definite result: through the grandest observational multi-wavelength effort in the history of mankind, from space, ground and underground observatories, we are finally finding evidence that black holes are "alive" and their "extractable energy" in our mass formula (Christodoulou & Ruffini, 1971), is the energy source of the most energetic cosmological sources: gamma ray bursts (GRBs),the active galactic nuclei (AGNs) as well as the ultra-high energy cosmic rays (UHECRs) (Ruffini et al. 2021 and references therein). Their "inner engine", has three independent components: 1) a Kerr black hole which is neither in a stationary state nor in vacuum, 2) a background magnetic field aligned with the black hole rotation axis, and 3) an extremely diluted fully ionized plasma (Moradi et al. 2021).There is no role in this inner engine for ISCO. Indeed a new electro dynamical field equations describe the synchrotron radiation emitted close to the black hole horizon, they point to a discrete and repetitive emission of "blackholic quanta" in the MeV and in the GeV. The magnitudes and the emission time scales of these quanta, for M87 and GRB 130427A, are expressed as a function of the above three parameters (Rueda & Ruffini, 2021). A long lasting GeV emission with a luminosity decreasing as a temporal power law, allows for the first time in GRBs, the determination of the black hole mass and spin as well as their time evolution perfectly fulfilling our mass energy formula (Christodoulou & Ruffini, 1971): a long lasting emission process profoundly different from the traditional process of continued gravitational contraction.

Remo Ruffini

Professor **GERARD 't HOOFT**

"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

Professor Gerard 't Hooft

Prof. Gerard 't Hooft has been a full Professor at the Utrecht University (the Netherlands), since 1977. Nowadays, he is an Emeritus Professor at that University. During his career, he has paid extended scientific visits to CERN (Geneva), Harvard, Stanford, Princeton and Duke University, NC. In 1999, together with M. Veltman, he received the Nobel Prize in Physics, awarded by The Royal Swedish Academy of Sciences, *"For elucidating the quantum structure of electroweak interactions in physics"*.

Prof. 't Hooft's main subjects of research includes:

– Gauge Theories for the sub-atomic particles and forces, various aspects and ingredients of what is now called "The Standard Model of the sub-atomic particles: renormalizability, topological features such as magnetic monopoles and instantons, 1/N expansions.

– Theories for the quantization of the gravitational force and black holes: producing models for the quantum properties of a black hole, as derived from Standard Model and General Relativity alone; its topological features such as antipodal identification.

– Fundamental theories underlying quantum mechanics, in particular returning determinism and reality to the dynamics of the tiniest material entities in his universe.

Prof. 't Hooft has been awarded the Wolf Prize of the State of Israel (1982), the Pius XI Medal (Vatican City, 1983), the Lorentz Medal (KNAW Amsterdam, 1986) as well as the Spinoza Premium (Netherlands Organization for Scientific Research NWO, 1995).

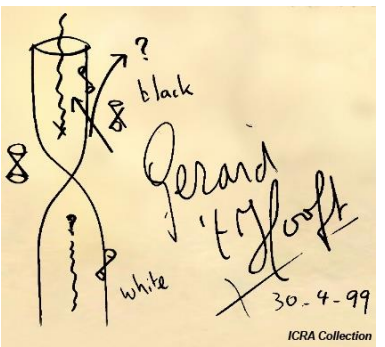


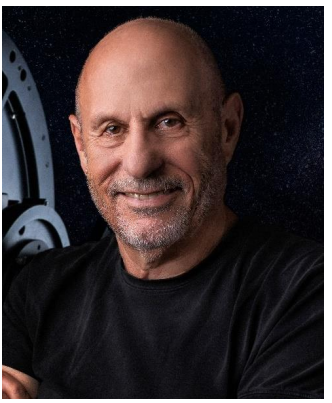
Fig. 2: The signature of Gerard 't Hooft on the wall of ICRA Room 301 (April 4, 1999).

A special event took place at ICRA on April 30, 1999. Prof. Ruffini invited Gerard 't Hooft to Rome to discuss a boundary condition for a quantum field on the black hole horizon, a topic Prof. Ruffini discussed in a previous article "Black-hole evaporation in the Klein-Sauter-Heisenberg-Euler formalism" with Thibault Damour (Phys. Rev. D 14, 332, 1976), but which needed to be examined in more detail. Prof. Ruffini planned to direct Gerard's attention to some specific aspects of this problem. Because we have traditionally been very attentive in spending ICRA travel funds, ICRA offered Gerard to come to Rome on a reduced fare weekend ticket arriving Friday and departing Monday. He had a great relaxing weekend together with Prof. Ruffini following his seminar, which among other things allowed Gerard to sign the wall in our ICRA Room (see Fig. 2), and during this splendid Rome spring weekend he also was able to find a missing factor of 2 in a formula in Prof. Ruffini's 1971 paper with Demetri Christodoulou on the black hole mass formula. The following October, Gerard received the Nobel prize, which meant that we could no longer get away with bringing him to Rome on a cheap ticket! Ever since Gerard has been in our MG IOC helping us with the preparation of the meetings. We are very happy to announce this MG16 Award to Gerard 't Hooft with the motivating phrase *"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon"*.

Remo Ruffini

Professor **TSVI PIRAN**

"for extending relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".



Professor Tsvi Piran

Tsvi Piran is the emeritus Schwartzmann professor at the Hebrew University of Jerusalem. He obtained his PhD in Physics, in 1976 from the Hebrew University working on the collisional Penrose process. Piran returned to the Hebrew University at 1981after being a post doc at Oxford and Texas and a long-term member at the IAS at Princeton. In 1982 he initiated and directed the first ever summer school on Gravitational Waves that took place at Les Houches. Piran was a visiting professor at Harvard, Columbia and New York and a Moore scholar at Caltech.

Piran's research deals with numerous aspects of relativistic astrophysics, ranging from the foundation of numerical relativity to modeling of observer relativistic phenomena and analytic work on the fate of gravitational collapse. Piran's research work focuses mostly on black holes and in particular on gamma-ray bursts. He was among the first to point out their cosmological origin and their association with merging neutron stars and heavy r-process nucleo synthesis. Piran's achievements were recognized in the 2019 EMET prize for Physics.

Professor **STEVEN WEINBERG**

"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Professor Steven Weinberg. Photo courtesy of Matt Valentine.

Steven Weinberg is a member of the Physics and Astronomy Departments at The University of Texas at Austin. His research has covered a broad range of topics in quantum field theory, elementary particle physics and cosmology. He has been honored with numerous awards, including the Nobel Prize in Physics, the National Medal of Science, the Heinemann Prize in Mathematical Physics and in 2020, the Breakthrough Prize. He is a member of the US National Academy of Sciences, Britain's Royal Society, and other academies in the USA and abroad. The American Philosophical Society awarded him the Benjamin Franklin Medal, with a citation that said he is "considered by many to be the preeminent theoretical physicist alive in the world today." His books for physicists include *Gravitation and Cosmology*, the three-volume work *The Quantum Theory of Fields*, *Cosmology* and published in April of 2021, *Foundations of Modern Physics*. Educated at Cornell, Copenhagen, and Princeton, he also holds honorary degrees from sixteen other universities. He taught at Columbia, Berkeley, M.I.T., and Harvard, where he was Higgins Professor of Physics, before coming to Texas in 1982.

Fig. 1: Chuo Pei Yuan and Cheng Ning Yang at MG2 in Trieste, Italy (1979).

The Sixteenth Marcel Grossmann Meeting (MG16) is a very special one in many respects: it will take place during a pandemic and in spite of the many difficulties, we have decided not to postpone it but to organize it as a virtual meeting. As described on the MG series webpage, these meetings started in 1975 with the first meeting at the International Centre for Theoretical Physics (ICTP) in Trieste (Italy) that I organized with Nobel Prize winner Abdus Salam. A second meeting followed in 1979, with a significantly larger participation including Nobel Laurate Cheng Ning Yang and a Chinese delegation led by Chuo Pei Yuan (see Fig. 1), including Fang Li-Zhi who had accompanied me during my entire first visit to China in 1979. The first truly international MG meeting followed in 1982 in Shanghai (China):this represented an especially important step forward both for the meeting and for China. A multi-millennia *"motto"* in China, which was then proclaimed on banners everywhere, read *"Friends from all over the world are welcomed"*.

We were soon at an impasse over the participation of scientists from Israel, since no diplomatic relations existed between China and Israel at that time and the Israeli scientists were not to be allowed to attend the meeting. A long negotiation began. The boundary conditions were clearly set by Steven Weinberg, a member of the present MG16 IOC: no MG meetings on Einstein's theory of general relativity could occur without the participation of Israeli scientists. The intervention of Yuval Ne'emann, also a member of the MG IOC then as well as the Minister of Science of Israel (see Fig.2), proposed a compromise that would admit at least one Israeli scientist. I went to Beijing alone, meeting every morning for a week with 12 Chinese representatives led by Chuo Pei Yuan going over all possible options. I stayed in an isolated villa not far from Tiananmen Square, accompanied by the 3 volumes of Matteo Ricci (RI MA TO) to keep me company. No solution was in sight the entire week. At the last moment, just before my departure, an agreement was finally reached allowing two Israeli scientists into China. The historic compromise would admit Gerard Tauber and Tsvi Piran into China using a special ICRA travel document I had proposed for them to be able to participate in the meeting, accepted by the Chinese Ambassador in Rome. This modified the thousand-year Chinese *"motto"* to read *"Scientists from all over the world are welcomed"*. The event was extremely beneficial for China and signaled the truly international nature of the MG meetings.

Fig. 2: From right to left: Chaim Weizmann, President of Israel; Yuval Ne'emann, Minister of Science of Israel; R. Ruffini.

I kept on meeting Tauber in the years which followed (see Fig. 3). Soon after, Yuval Ne'emann visited China. The development of bilateral relations, including military cooperation and economical tights, grow exponentially until the establishment of normal diplomatic relations between Israel and China in 1992.

Fig. 3: From right to left: Arrigo Finzi, Remo Ruffini, Gerard Tauber and Konrad Bleuler.

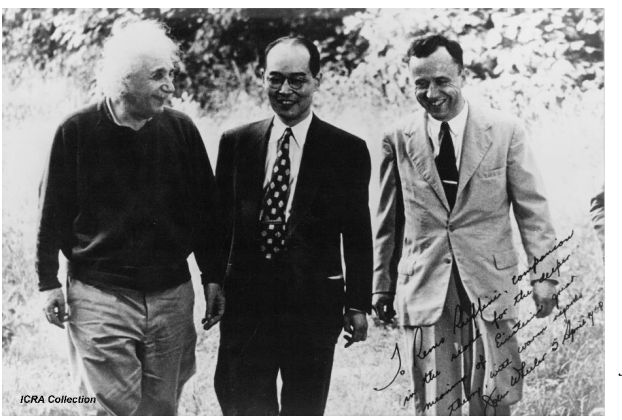


Fig. 4: Albert Einstein, Hideki Yukawa and John. A. Wheeler with a hand-written dedication to Remo Ruffini "To Remo Ruffini, companion in the search for the deeper meaning of Einstein great theory. With warm regards, John Wheeler 5 April 1968".

Given their key role played in the foundations of the MG meetings, I am very happy to propose on behalf of the MG16 IOC, two special Marcel Grossmann Individual Awards: one to Steven Weinberg for *"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory"* and another one to Tsvi Piran, *"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory"*, in the words of John A. Wheeler's photo dedication to me (see Fig. 4).

Remo Ruffini

PREFACE

Since 1975, the Marcel Grossmann Meetings on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories have been organized in order to provide opportunities for discussing recent advances in gravitation, general relativity and relativistic field theories, emphasizing mathematical foundations, physical predictions and experimental tests. The objective of these meetings is to elicit exchange among scientists that may deepen our understanding of spacetime structures as well as to review the status of ongoing experiments aimed at testing Einstein's theory of gravitation either from the ground or from space. Previous meetings have been held in Trieste (MG1: 1975) and (MG2: 1979), Shanghai (MG3: 1982), Rome (MG4: 1985, MG9: 2000), Perth (MG5: 1988), Kyoto (MG6: 1991), Stanford (MG7: 1994), Jerusalem (MG8: 1997), Rio (MG10: 2003), Berlin (MG11: 2006), Paris (MG12: 2009), Stockholm (MG13: 2012), MG14 in 2015 and MG15 in 2018 both in Rome.

Due to the COVID-19 pandemic spreading in the last two years the decision was taken to organize the Sixteenth Marcel Grossmann meeting for the first time in history entirely online. Despite numerous challenges, related to the organization of large worldwide event, MG16 showed the strongest ever interest from the scientific community with a record-breaking number of almost 1200 registered participants and of more than 1000 speakers.

The traditional six-day schedule has been modified to account for different time zones of the speakers and each day the program of the meeting was divided in three blocks with the reference to the Central European Summer Time. The first block was starting at 06:30 in the morning, allowing comfortable time for speakers from Asia and Oceania. The second block was held in the daytime in Europe and Africa. The third block was starting in the afternoon and ending at 19:30 allowing accommodation of the speakers from the Americas. Each day the blocks of plenary sessions were interchanging with the blocks of about 30 parallel sessions each, making this one of the most intense MG meetings ever. All this was possible thanks to recent developments in communication technologies. The Indico open-source software was selected as a web platform for this meeting, while Zoom platform was adopted for the video-conferencing. The meeting was streamed on ICRANet YouTube channel.

The meeting started on Monday July 5 with the Award ceremony. The individual awards went to Demetrios Christodoulou, Tsvi Piran, Gerard 't Hooft and Steven Weinberg, while the Institutional Awards went to the S.A. Lavochkin Association, to the Max Planck Institute for Extraterrestrial Physics – MPE and to the Space Research Institute IKI of the Russian Academy of Sciences. Overall there were 54 plenary talks, 4 public lectures and 5 roundtables and about 90 parallel sessions. The plenary session "Events in Relativistics Astrophysics" on Monday have seen the contributions from Rashid Sunyaev, Michael Kramer, James Miller-Jones, Felix Mirabel. The public lectures were delivered by Razmik Mirzoyan, Asghar Qadir

and Mohammad Bagheri. Plenary talks on Tuesday session "Black holes and the Quantum" by Juan Maldacena, Ahmed Almheiri, Gerard 't Hooft, Mihalis Dafermos, Sergiu Klainerman, Abhay Ashtekar and Frank Wilczek were bracketed by two roundtables on "New results from SRG/eRosita" with the participation of Andrea Merloni, Prof. Rashid Sunyaev, Alexander Lutovinov, Chandreyee Maitra, Esra Bulbul and "Solar neutrinos and Borexino" with the participation of Gianpaolo Bellini and Wick Haxton. Plenary talks on Wednesday in the session "Lambda CDM tensions" by George Efstathiou, Scolnic Daniel, Marc Kamionkowski, Wendy Freedman, Priya Natarajan and Licia Verde were followed by the roundtable "Precision cosmology" with the participation of Licia Verde, Marc Kamionkowski, Piero Rosati, and the public lecture by Francis Halzen. Two blocks of Thursday plenary sessions "Black holes in GRBs" and "Precision tests" included the talks by Roy Kerr, Yuan Ha, Lorenzo Amati, Elena Pian, Carlos Raúl Argüelles, Di Li, Jianglai Liu, Claus Lämmerzahl, Gerhard Heinzel and Ignazio Ciufolini and were followed by the roundtable "GRB 170817A and GRB 190829A" with the participation of Eleonora Troja, Liang Li, Rahim Moradi, Jorge Armando Rueda Hernandez. Two plenary blocks on Friday "Massive stars" and "Physics behind stellar collapse" included the talks by Selma de Mink, Norbert Langer, Jiri Bicak and Tomáš Ledvinka, Ivan De Mitri, Rahim Moradi and Giancarlo Cella. Finally, two plenary blocks on Saturday "Current and future missions" have seen the talks by Shuang-Nan Zhang, Weimin Yuan, Makoto Tashiro, Ruoyu Liu, Jean-Luc Atteia, Jim Hinton and Nicholas White and were followed by the roundtable "What is in our Galactic center" with the participation of Reinhard Genzel, Carlos Raúl Argüelles, Andreas Krut, Jorge Armando Rueda Hernandez, Eduar Becerra Vergara. The program of the meeting can be found at the official website http://www.icra.it/mg/mg16 and at ICRANet Indico website https://indico.icranet.org/event/1/.

These proceedings include about 400 papers containing the results presented at the Sixteenth Marcel Grossmann meeting. The plenary papers from the meeting have been published in International Journal of Modern Physics D as they were submitted. The table of contents includes also the links to YouTube videos with talks given at the meeting and cover plenary talks, public lectures, roundtables and all parallel sessions. The general link to the videos from MG16 is: https://www.youtube.com/watch?v=QFe1lsSid-o&list=PLr5RLbSWSonsaOnZukBDs0qsNIWM8AvRF.

As the editors we would like to express our gratitude to all the chairpersons of the parallel sessions at MG16, who peer-reviewed the papers submitted for these proceedings, as well as to the ICRANet secretariat office and in particular to Cinzia di Niccolo, Elisabetta Natale and Yasmina Di Domizio, as well as to ICRANet system manager Gabriele Brandolini for their help in preparation of this publication.

Remo Ruffini and Gregory Vereshchagin
November 2021

This page intentionally left blank

Contents

Publications in this Series v

Sponsors and Acknowledgements vii

Organizing Committees ix

Marcel Grossmann Awards xi

Preface xxxvi

PART A
PLENARY SESSIONS

New results from testing relativistic gravity with radio pulsars
Michael Kramer 3

Dragging of inertial frames by matter and waves
Jiří Bičák and Tomáš Ledvinka 22

Probes of the progenitors, engines and physics behind stellar collapse
Chris L. Fryer 39

The observation of high-energy neutrinos from the cosmos: Lessons learned for multimessenger astronomy
Francis Halzen 59

The first results of PandaX-4T
Jianglai Liu on behalf of the PandaX Collaboration 85

XRISM: X-ray imaging and spectroscopy mission
Makoto S. Tashiro and the XRISM team 95

The SVOM mission
J.-L. Atteia, B. Cordier and J. Wei on behalf of the SVOM Collaboration 104

Quantum field theory with boundary conditions at the horizons
Gerard 't Hooft 133

The development of general relativity and the cosmological constant
Asghar Qadir 143

The irreducible mass of Christodoulou-Ruffini-Hawking mass formula
Yuan K. Ha 154

Reshaping our understanding on structure formation with the quantum nature of the dark matter
C. R. Argüelles, E. A. Becerra-Vergara, A. Krut et al. 164

First results of LHAASO
Ruo-Yu Liu for the LHAASO Collaboration 180

On the MG16 awards 2021
Remo Ruffini 196

The white dwarf binary merger model of GRB 170817A
J. A. Rueda, R. Ruffini, Liang Li et al. 217

PARALLEL SESSIONS

Accretion

• MHD Processes Near Compact Objects
Chairperson: Sergey Moiseenko
YouTube link: https://youtu.be/2WMTg06ZmV8

A semi-implicit multidimensional unstructured gas dynamical solver for astrophysical applications
Ilya A. Kondratyev and Sergey G. Moiseenko 242

Magnetized neutron stars propagating through a non-uniform ISM
O. D. Toropina, M. M. Romanova and R. V. E. Lovelace 255

Calculation of the kinetic coefficients of arbitrary degenerate electrons in magnetized dense matter
M. V. Glushikhina and G. S. Bisnovatyi-Kogan 264

Modeling of magnetic fields of accretion discs, using no-z- and RZ-approximations
M. V. Pashentseva and E. A. Mikhailov 272

• Accretion Discs and Jets
Chairpersons: Audrey Trova and Shokoufe Faraji
YouTube link: https://youtu.be/29Wj9RCVEKw

Limiting effects in tori clusters
D. Pugliese and Z. Stuchlik 280

Hydrodynamical transport of angular momentum in accretion disks in the presence of nonlinear perturbations due to noise
Subham Ghosh and Banibrata Mukhopadhyay 295

Properties of accretion disc models in the background quadrupole
Shokoufe Faraji 307

Magnetized tori around a uniformly accelerating black hole
Shokoufe Faraji and Audrey Trova 317

Multifrequency behaviour of high mass X-ray binaries (Time lag between optical and X-ray outbursts)
Franco Giovannelli 321

Active Galactic Nuclei

• The Black Hole in M87
Chairpersons: Brian Punsly and Jorge Rueda
YouTube link: https://youtu.be/l1lTgksyJag

Rotation of the crescent image of M87* and polarization of its ESE hotspot
Krzysztof Nalewajko 339

Magnetic reconnection in jet-accretion disk systems
Chandra B. Singh, Elisabete M. de Gouveia Dal Pino, Luis H. S. Kadowaki et al. 344

• Machine Learning in Astronomy: AGN, Transient Events, Cosmology and Others
Chairpersons: Rahim Moradi and Yu Wang
YouTube link: https://youtu.be/ErqrmMZQsBk

Exact fractal model of the universe and possible machine learning methods for the verification of the fractality
A. A. Kirillov, E. P. Savelova and P. O. Vladykina 352

Estimating the photometric redshifts of galaxies and QSOs using regression techniques in machine learning
A. Momtaz, M. H. Salimi and S. Shakeri 368

Deep learning in quasar physics
F. Rastegar Nia, M. T. Mirtorabi, R. Moradi, Y. Wang and A. Vafaei Sadr 382

Cosmological density field emulation and gravitational wave inference based on dimensionality reduction and supervised machine learning
Miguel Conceição, António da Silva and Alberto Krone-Martins 391

Unsupervised photometric detection of galaxy cluster candidates in large surveys
Ana Carvalho, Alberto Krone-Martins and Antonio da Silva 409

• Multiwavelength and Multi-Messenger Observations of Active Galactic Nuclei
Chairpersons: Paolo Giommi and Narek Sahakyan

YouTube link: https://youtu.be/-Hyu2NQsExg

Time-dependent lepto-hadronic modeling of the emission processes in blazar jets
S. Gasparyan, D. Bégué and N. Sahakyan 429

Multiwavelength study of high-redshift blazars
G. Harutyunyan and D. Israyelyan 445

Alternative Theories

• Extended Theories of Gravity and Quantum Cosmology
Chairpersons: Yi-Fu Cai and Wentao Luo

YouTube links: https://youtu.be/ADRr9DfV5zM
https://youtu.be/eOzpiC1cFkU
https://youtu.be/kYzJds_JIp8

Quantum gravity phenomenology from thermodynamics of spacetime
A. Alonso-Serrano and M. Liška 462

Gauge theory of gravity based on the correspondence between the 1^{st} and the 2^{nd} order formalisms
David Benisty 479

$U(1)$ local strings in hybrid metric-Palatini gravity
Tiberiu Harko, Francisco S. N. Lobo and Hilberto M. R. da Silva 485

Inflationary supersymmetric FRLW quantum cosmology
N. E. Martínez-Pérez, C. Ramirez and V. M. Vázquez-Báez 499

Effective $f(R)$ actions for modified loop quantum cosmologies
Ana Rita Ribeiro, Daniele Vernieri and Francisco S. N. Lobo 517

Probing multiverse using gravitational wave observations
Moe Kukihara and Kazuhiro Hayama 531

Operator ordering ambiguity in observables of quantum cosmology
Harkirat Singh Sahota and Kinjalk Lochan 538

Decoupled quark stars in self-interacting Brans-Dicke gravity
M. Sharif and Amal Majid 548

Big-bounce in projectively invariant Nieh-Yan models: The Bianchi I case
Flavio Bombacigno, Simon Boudet, Gonzalo J. Olmo and Giovanni Montani 561

Late time cosmology with derivatives of matter Lagrangian
Shahab Shahidi 576

On the semiclassical and quantum picture of the Bianchi I polymer dynamics
E. Giovannetti, G. Montani and S. Schiattarella 588

Quantum corrections to the Bianchi II transition under local rotational invariance
Sara F. Uria, David Brizuela and Ana Alonso-Serrano 597

• **Mathematical Problems of Relativistic Physics: Classical and Quantum**
Chairpersons: A. Shadi Tahvildar-Zadeh and Michael Kiessling

YouTube links: https://youtu.be/9Dr3M9Kb2jo
https://youtu.be/gMOykpapJ5A

The hypercomplex medium as storage of physical equations
Alexander P. Yefremov 605

The Maxwell-Bopp-Landé-Thomas-Podolsky-Einstein system for a static point source
Érik Amorim 619

On recent developments in the theory of relativistic dissipative fluids
V. Hoang 627

Adiabatic solutions in general relativity as null geodesics on the space of boundary diffeomorphisms
Emine Şeyma Kutluk 635

The point spectrum of the Dirac Hamiltonian on the zero-gravity Kerr-Newman spacetime
M. Kiessling, E. Ling and A. S. Tahvildar-Zadeh 648

Causal fermion systems: Classical gravity and beyond
Felix Finster 661

Newman-Penrose-Debye formalism for fields of various spins in pp-wave backgrounds
Aleksandr Kulitskii and Elena Yu Melkumova 679

Gravitational geometric phase
Banibrata Mukhopadhyay, Tanuman Ghosh and Soumya Kanti Ganguly 689

Retarded potentials and radiation in odd dimensions
D. V. Gal'tsov and M. Khlopunov 699

Wave propagation in the anti-deSitter optical metric
D. García-Peláez and C. S. López-Monsalvo 713

New approaches to constrained dynamics and Hamilton-Jacobi procedures in general relativity
D. Salisbury, J. Renn and K. Sundermeyer 719

Orientability of space from electromagnetic quantum fluctuations
N. A. Lemos and M. J. Rebouças 725

Essential self-adjointness of Dirac operators under the influence of general-relativistic gravity
Michael K.-H. Kiessling, A. Shadi Tahvildar-Zadeh and Ebru Toprak 736

• **Wormholes, Energy Conditions and Time Machines**
Chairpersons: Francisco Lobo and Diego Rubiera-Garcia

YouTube links: https://youtu.be/tu_3Wqcd9Ys
https://youtu.be/NqN1c-2fv8Y

Relic magnetic wormholes as possible source of toroidal magnetic fields in galaxies
A. A. Kirillov and E. P. Savelova 743

Wormhole geometries induced by action-dependent Lagrangian theories
Ismael Ayuso, Francisco S. N. Lobo and José P. Mimoso 756

Gravitational lensing by wormholes in binary systems
S. Pietroni 774

Hyper-fast positive energy warp drives
E. W. Lentz 779

From black-bounce to traversable wormhole, and beyond
Alex Simpson 787

Tractor beams, pressor beams, and stressor beams within the context of general relativity
Matt Visser, Jessica Santiago and Sebastian Schuster 808

A singularity theorem for evaporating black holes
E.-A. Kontou, B. Freivogel and D. Krommydas 822

Circularly symmetric thin-shell wormholes in F(R) gravity with (2+1)-dimensions
Cecilia Bejarano, Ernesto F. Eiroa and Griselda Figueroa-Aguirre 831

Warp drive dynamic solutions considering different fluid sources
Osvaldo L. Santos-Pereira, Everton M. C. Abreu and Marcelo B. Ribeiro 840

Symmetries and geometry of spacetime: Towards a new paradigm
Francisco Cabral, Francisco S. N. Lobo and Diego Rubiera-Garcia 856

• Theories of gravity: Alternatives to the cosmological and particle standard models
Chairpersons: Stefano Bellucci and Orlando Luongo

YouTube links: https://youtu.be/-aQcrYoQfBM
https://youtu.be/WsCwPb5OQhY
https://youtu.be/JMnrUfBgqVU
https://youtu.be/pPhZY-bbsew

Thermodynamics of scalar-tensor gravity: A new approach
Valerio Faraoni 876

Two body dynamics in a quadratic modification of general relativity
Soham Bhattacharyya 883

Alternatives to Λ: Torsion, generalized couplings, and scale invariance
C. J. A. P. Martins, C. M. J. Marques, C. B. D. Fernandes et al. 907

Model-independent test of scalar-tensor gravity theory by reconstructing scalar mode of GW170817
Yuya Gushima and Kazuhiro Hayama 921

Cosmology in the novel scalar-tensor representation of $f(R,T)$ gravity
Tiago B. Gonçalves, João Luís Rosa and Francisco S. N. Lobo 932

On the interaction between electromagnetic, gravitational, and plasma related perturbations on LRS class II spacetimes
P. Semrén 943

Condition for expansion-collapse duality between Einstein and Jordan frames
Dipayan Mukherjee, H. K. Jassal and Kinjalk Lochan 958

The model of dark energy based on the quantum-mechanical uncertainty relation
Yu. V. Dumin 967

• Conformal Dilaton Gravity and Related Issues
Chairperson: Reinoud Jan Slagter

YouTube link: https://youtu.be/A3Ygi3YBs5A

Conformal dilaton gravity, antipodal mapping and black hole physics on a warped spacetime
R. J. Slagter 978

From neutrino masses to the full size of the universe — Some intriguing aspects of the tetron model
B. Lampe 999

Summary parallel session AT5
Reinoud Jan Slagter 1014

• Horava-Lifshitz Gravity
Chairperson: Anzhong Wang

YouTube links: https://youtu.be/vPT1dH1zITE
https://youtu.be/5z7zhpiDOpw

Boundary conditions for the Klein-Gordon field on Lifshitz spacetimes
Lissa de Souza Campos 1017

Dynamical system analysis of Bianchi-I spacetimes in $f(R)$ gravity
Saikat Chakraborty, Kazuharu Bamba and Alberto Saa 1026

Cosmological implications in modified Hořava-Lifshitz gravity
Abdul Jawad, Kazuharu Bamba and Farwa Khurshid 1038

Finite action principle and wormholes
Jan Chojnacki and Jan Kwapisz 1046

Strange quark stars in Hořava gravity
Grigoris Panotopoulos 1054

Shadows of Kerr-like black holes in $4D$ Einstein–Gauss–Bonnet gravity and constraints from EHT observations
Sushant G. Ghosh and Rahul Kumar Walia 1069

Wormhole interaction in 2d Horăva-Lifshitz quantum gravity
Jan Ambjørn, Yuki Hiraga, Yoshiyasu Ito and Yuki Sato 1084

Nature of singularities in vector-tensor theories of gravity
V. H. Satheeshkumar 1095

Hořava-Lifshitz and Einstein-Æther gravity in the light of Event Horizon Telescope observations of M87*
Emmanuel N. Saridakis 1104

Hořava-Lifshitz gravity in $(3 + 1)$ dimensions coupled with anisotropic matter and possible constraints from GRB 170817A
Tao Zhang and Fu-Wen Shu 1112

Summary of the parallel session AT6
Anzhong Wang 1119

• Ghost-Free Models of Modified Gravity: Massive Gravity, Horndeski and DHOST Theories, Other Related Models; Their Properties and Solutions
Chairpersons: Dmitry Gal'tsov and Michael Volkov

YouTube links: https://youtu.be/l8KHUPnT2D8
https://youtu.be/48OHIKpgNqs

Non-local R^2-like inflation, gravitational waves and Non-Gaussianities
K. Sravan Kumar 1124

Palatini kinetic scalar-tensor theory: Analytical and numerical solutions
D. V. Gal'tsov and D. S. Bushuev 1136

PART B

Black Holes: Theory and Observations/Experiments

• Theoretical and Observational Studies of Astrophysical Black Holes
Chairperson: Alexander Zakharov

YouTube links: https://youtu.be/fiv_MH-N2kw
https://youtu.be/GYoOb17GvE8

Reconstruction of a star motion in the vicinity of black hole from the redshift of the electromagnetic spectrum
S. O. Komarov and A. K. Gorbatsievich 1151

Shadows of hairy Kerr black holes and constraints from M87*
Sushant G. Ghosh and Misba Afrin 1167

Displacement memory and BMS symmetries
Shailesh Kumar 1179

Physical black holes in semiclassical gravity
Sebastian Murk and Daniel R. Terno 1196

• Black Hole Thermodynamics
Chairperson: Hernando Quevedo

YouTube links: https://youtu.be/XmZDf5mrXQk
https://youtu.be/amseL2qykfk

Information recovery from evaporating rotating charged black holes
Zhi-Wei Wang, Samuel L. Braunstein and Saurya Das 1212

Black hole thermodynamics from entanglement mechanics
S. Mahesh Chandran and S. Shankaranarayanan 1223

Thermodynamics of charged black hole
M. Sharif and Amjad Khan 1238

Rindler trajectories and Rindler horizons in the Schwarzschild spacetime
Kajol Paithankar and Sanved Kolekar 1250

Linear growth of the two-point function for the Unruh state in $1+1$ dimensional black holes
Paul R. Anderson, Zachary P. Scofield and Jennie Traschen 1255

Stress-energy tensor for a quantized scalar field in a four-dimensional black hole that forms from the collapse of a null shell
Shohreh Gholizadeh Siahmazg, Paul R. Anderson, Raymond D. Clark and Alessandro Fabbri 1265

Microscopic model building for black hole membranes from constraints of symmetry
Swastik Bhattacharya and S. Shankaranarayanan 1275

Einstein-Maxwell-Dilaton-Axion mass formulas for black holes with struts and strings
Dmitri Gal'tsov, Gérard Clément and Igor Bogush 1291

• Black Holes in Alternative Theories of Gravity
Chairpersons: Jutta Kunz and Kamal Hajian

YouTube links: https://youtu.be/FRNGJKhiw7c
https://youtu.be/Tjfmuut1Eo0

Holography for rotating black holes in $f(T)$ gravity
Masoud Ghezelbash 1308

Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity
Semin Xavier, Jose Mathew and S. Shankaranarayanan 1319

Universe in a black hole with spin and torsion
Nikodem Popławski 1327

Asymptotically flat black hole solution in modified gravity
Surajit Kalita and Banibrata Mukhopadhyay 1337

Shadow of a charged black hole surrounded by an anisotropic matter field
Javier Badía and Ernesto F. Eiroa 1343

Constraining modified gravity theories with physical black holes
Sebastian Murk 1351

Penrose suggestion as to pre-Planck-era black holes showing up in present universe data sets discussed, with a possible candidate as to GW radiation which may provide initial CMBR data
A. W. Beckwith 1359

Summary of the parallel session BH3
Kamal Hajian and Jutta Kunz 1372

Binaries

• **Explosive Events Associated with Compact-Object Binary Mergers**
Chairpersons: Chris Belczynski and Jorge Rueda

YouTube links: https://youtu.be/Dwq1ZU3gKrg
https://youtu.be/nw02ylI6R2M

Uncertainties in kilonova modeling
C. L. Fryer, C. J. Fontes, O. Korobkin et al. 1391

• **Post-Newtonian and Post-Minkowskian Corrections for Binary Gravitating Systems**
Chairperson: Johannes Bluemlein

YouTube link: https://youtu.be/wfiLG5r08yE

Tutti-Frutti method: Recent developments in the PN/PM/SF treatment of the gravitational two-body problem
Donato Bini and Andrea Geralico 1405

• **Multichannel Studies of Nonstationary Relativistic Stars**
Chairperson: Vladimir Lipunov

YouTube link: https://youtu.be/usn2PlU_qFA

GRB observations on cubesate satellites in the Universat–SOCRAT project
Sergey I. Svertilov, Michail I. Panasyuk, Vitaly V. Bogomolov et al. 1412

Multiwavelength observations of GRB160625B by MASTER, Lomonosov, Konus-Wind and three stage collapse
V. M. Lipunov, V. A. Sadovnichy, M. I. Panasyuk et al. 1429

The role of the magnetic fields in GRB outflows
N. Jordana-Mitjans, C. G. Mundell, S. Kobayashi et al. 1449

MASTER optical observations of the blazar TXS0506+056 during the IC170922A
V. M. Lipunov, K. Zhirkov, V. G. Kornilov et al. 1467

Boson Stars

• Scalar Fields in Cosmology
Chairpersons: Carlos Herdeiro and Alfredo Macias

YouTube links: https://youtu.be/SyLoguueGKk
https://youtu.be/f9nuo8Jvw-w

A short review on nonlinear perturbation theory of structure formation for modified gravity
Jorge L. Cervantes-Cota and Alejandro Aviles 1474

Testing modified gravity theories with marked statistics
Alejandro Aviles 1494

Dark matter as condensed phase of generic bosons
Elías Castellanos and Jorge Mastache 1513

Cosmic Microwave Background

• Cosmic Backgrounds from Radio to Far-IR
Chairperson: Carlo Burigana

YouTube link: https://youtu.be/4e3Cj5wahck

New Planck tSZ map and its cosmological analysis
H. Tanimura, M. Douspis and N. Aghanim 1527

The CMB dipole: Eppur si muove
R. M. Sullivan and D. Scott 1532

High angular resolution Sunyaev Zel'dovich observations: The case of MISTRAL
E. S. Battistelli, E. Barbavara, P. de Bernardis et al. 1542

Cosmological and astrophysical results exploiting magnification bias with high-z sub-millimetre galaxies
L. Bonavera, M. M. Cueli and J. Gonzalez-Nuevo 1557

The impact of the Lorentz symmetry violation on the CMB polarization
Seddigheh Tizchang, Rohoollah Mohammadi and She-Sheng Xue 1571

Cosmic backgrounds from the radio to the far-infrared: Recent results and perspectives from cosmological and astrophysical surveys
Carlo Burigana, Elia Sefano Battistelli, Laura Bonavera et al. 1579

• **New Horizons in Cosmology with CMB Spectral Distortions**
Chairpersons: Jens Chluba and Andrea Ravenni

YouTube links: https://youtu.be/uBYLO4Smw3o
https://youtu.be/5oAPfzAe35k

CMB μT cross-correlations as a probe of PBH scenarios
Ogan Özsoy and Gianmassimo Tasinato 1609

Theoretical and numerical aspects of CMB spectral distortions from non-thermal energy injections
Sandeep Kumar Acharya, Jens Chluba and Abir Sarkar 1628

BISOU: A balloon project to measure the CMB spectral distortions
B. Maffei, M. H. Abitbol, N. Aghanim et al. 1633

Cosmic microwave background spectral distortions constraints on decaying dark matter particles and axion-like particles using *COBE/FIRAS* and *EDGES*
Boris Bolliet 1645

The COSmic Monopole Observer (COSMO)
S. Masi, E. Battistelli, P. de Bernardis et al. 1654

• **Status of the $\mathbf{H_0}$ and $\boldsymbol{\sigma_8}$ Tensions: Theoretical Models and Model-Independent Constraints**
Chairpersons: Joan Solà Peracaula and Adrià Gómez-Valent

YouTube links: https://youtu.be/VNWZ1Bzjus4
https://youtu.be/zEBOCwvetKE
https://youtu.be/nzgC7qV9H_Y

Measuring the Hubble constant H_0 from gravitational lensing
Liliya L. R. Williams 1672

Extra components consistency in the Hubble tension and BBN
Osamu Seto and Yo Toda 1686

Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology
Nick E. Mavromatos 1693

Early and not so early dark energy. What do cosmological observations tell us about them?
Adrià Gómez-Valent, Ziyang Zheng, Luca Amendola, Valeria Pettorino and Christof Wetterich 1713

Renormalized $\rho_{\rm vac}$ without m^4 terms
Cristian Moreno-Pulido and Joan Solà Peracaula 1733

BD-ΛCDM and running vacuum models: Theoretical background and current observational status
Javier de Cruz Pérez, Joan Solà Peracaula, Adrià Gómez-Valent and Cristian Moreno-Pulido 1752

Cosmological tensions: Hints for a new concordance model?
E. Di Valentino 1770

Solving both H_0 and σ_8 tensions in $f(T)$ gravity
Emmanuel N. Saridakis 1783

Precision Cosmology and Hubble tension in the era of LSS surveys
G. Fanizza 1792

• **Effects of Primordial Perturbations Enhancement: From Black Holes Formation to CMB Anomalies**
Chairpersons: Antonio Enea Romano and Krzysztof Turzynski
YouTube link: https://youtu.be/frjjONXbd1M

Primordial black holes arise when the inflaton falls
Keisuke Inomata 1803

Effects of the modification of gravity on the production of primordial black holes
Sergio Andrés Vallejo Peña 1809

Cosmic Strings

• **Cosmic Strings**
Chairpersons: Reinoud Jan Slagter and Batool Imtiaz
YouTube link: https://youtu.be/ZpaU82ZUHzM

$U(1)$ local strings in generalized hybrid metric-Palatini gravity
Hilberto M. R. da Silva, Tiberiu Harko, Francisco S. N. Lobo and João Luís Rosa 1820

New evidence of the azimuthal alignment of quasars spin vector in Large Quasar Groups and cosmic strings
R. J. Slagter 1835

Summary parallel session cosmic strings I
Reinoud Jan Slagter and Batool Imtiaz 1848

• **From Cosmic Strings to Superstrings**
Chairpersons: Carlos Martins and Ivan Rybak
YouTube link: https://youtu.be/LJFtV_4aSAg

Scaling solutions of wiggly cosmic strings
A. R. R. Almeida and C. J. A. P. Martins 1851

High resolution calibration of string network evolution
J. R. C. C. C. Correia and C. J. A. P. Martins 1871

Radiation from Global Cosmic Strings using adaptive mesh refinement
Amelia Drew and E. P. S. Shellard 1891

Analysing the scaling density of axion strings
A. Lopez-Eiguren 1898

Electroweak axion string and superconductivity
Yu Hamada, Yoshihiko Abe and Koichi Yoshioka 1912

Dark Energy and Large Scale Structure

• **Dark Energy and the Accelerating Universe**
Chairpersons: Alexei Starobinky and David Polarski
YouTube links: https://youtu.be/JZEPRS_rqbE
https://youtu.be/3aG_AT4UzWE
https://youtu.be/ZlTWquG-pFk

Hints for the $H_0 - r_d$ tension in uncorrelated Baryon Acoustic Oscillations dataset
Denitsa Staicova 1923

Observational constraints on nonlinear matter extensions of general relativity
E.-A. Kolonia and C. J. A. P. Martins 1935

Constraining the dark energy-dark matter interaction model using low-redshift observations
Archana Sangwan, Joseph P. J. and S. Shankaranarayanan 1948

On the evolution of inhomogeneous perturbations in the ΛCDM model and $f(R)$ modified gravity theories
T. Schiavone and G. Montani 1961

Soft dark energy and soft dark matter
Emmanuel N. Saridakis 1970

A simple parametrisation for coupled dark energy
Vitor da Fonseca, Nelson J. Nunes and Tiago Barreiro 1979

• **Cosmography with Gravitational Lensing**
Chairpersons: Claudio Grillo and Mimoza Hafizi
YouTube link: https://youtu.be/FXplIAvUDBM

A tale of two double quasars: Hubble constant tension or biases?
L. J. Goicoechea and V. N. Shalyapin 1990

Dark Matter

• **Interacting Dark Matter**
Chairpersons: Nikolaos Mavromatos
YouTube links: https://youtu.be/zOshKJlwD-Y
https://youtu.be/JR44dh2GYik

Dark energy and dark matter unification from dynamical space time: BBN constraints
D. Benisty 2005

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2013

LHC experiments for long-lived particles of the dark sector
Vasiliki A. Mitsou 2029

Constraining the interactions in the dark sector with cosmological data
Adrià Gómez-Valent, Valeria Pettorino and Luca Amendola 2050

Running vacuum interacting with dark matter or with running gravitational coupling. Phenomenological implications
Joan Solà Peracaula 2069

Dark matter properties from the Fornax globular cluster timing: Dynamical friction and cored profiles
D. Blas 2089

Growth of linear perturbations in a universe with superfluid dark matter
S. Banerjee, S. Bera and D. F. Mota 2101

Interacting dark sector in the late Universe: Mapping fields and fluids, and observational signatures
Joseph P. J. and S. Shankaranarayanan 2119

The role of self interactions in the cosmological evolution of warm dark matter
R. Yunis, C. R. Argüelles, D. López Nacir et al. 2127

Interaction energy between a charged medium and its electromagnetic field as a dark matter candidate
Mayeul Arminjon 2139

• Dark Matter Searches with Liquid Xenon and Argon Detectors and Self Gravitating Systems and Dark Matter
Chairpersons: Marco Merafina and Soroush Shakeri and She-Sheng Xue

YouTube link: https://youtu.be/H9oGYnGq9pI

The maximum mass of dilute axion stars
Pierre-Henri Chavanis 2149

A dark matter solution for the XENON1T electron excess and the galactic center 511 keV line
Yasaman Farzan 2174

Preliminary results of rich galaxy clusters' spatial distribution analysis on CfA2 Redshift Survey data: Compact objects or dark matter presence at redshift less 0.032
I. V. Arkhangelskaja, A. M. Galper, L. N. Khanh and D. N. Dorosheva 2189

• Dark Matter: Beyond ΛCDM
Chairpersons: Carlos Argüelles and Andreas Krut

YouTube links: https://youtu.be/hdKeo5L7pYE
https://youtu.be/i0IPHXzmV-s

Probing the nature of dark matter with Milky Way subhaloes
M. R. Lovell 2202

Addressing classical cosmological back-reaction with multiple scales
Yonadav Barry Ginat 2217

Imaging formation process for DM profiles
Omar de J. Cabrera-Rosas and Tonatiuh Matos 2222

The self-gravitating Fermi gas in Newtonian gravity and general relativity
Pierre-Henri Chavanis 2230

• Dark Matter and Rare Processes
Chairpersons: Rita Bernabei and Zurab Berezhiani

YouTube links: https://youtu.be/wheVrbETP_0
https://youtu.be/dVdZ4TIDxt0

The dark matter: DAMA/LIBRA and its perspectives
R. Bernabei, P. Belli, V. Caracciolo et al. 2252

Dark matter directionality approach
R. Bernabei, P. Belli, V. Caracciolo et al. 2272

Collapse models under test by high sensitivity γ-ray and X-ray measurements
C. Curceanu, Kristian Piscicchia, Massimiliano Bazzi et al. 2288

Leptophilic dark matter at linear colliders
P. S. Bhupal Dev 2296

DM6 session: Dark matter and rare processes
R. Bernabei and Z. Berezhiani 2316

• **The Nature of Galactic Halos**
Chairpersons: Francesco De Paolis and Asghar Qadir
YouTube link: https://youtu.be/qIwVnNxi0n0

Primordial black holes as dark matter candidates in the Galactic halo
Lindita Hamolli, Mimoza Hafizi, Francesco De Paolis and Achille A. Nucita 2319

Giant cosmic ray halos around M31 and the Milky Way
S. Recchia, S. Gabici, F. A. Aharonian and V. Niro 2335

A nearly complete census of intergalactic gas using the kinematic Sunyaev-Zel'dovich effect
Chaves-Montero, Jonás 2345

Searching for Intermediate Mass Black Holes in the Milky Way's galactic halo
A. Franco, A. A. Nucita, F. De Paolis, F. Strafella and M. Maiorano . . . 2352

Virial clouds evolution from the last scattering up to the formation of first stars
Noraiz Tahir, Asghar Qadir, Muhammad Sakhi and Francesco De Paolis 2360

Testing Weyl-modified gravity on M31 and Milky Way
Muhammad Bilal and Asghar Qadir 2365

PART C

Education

• **Teaching Einsteinian Physics to School Students**
Chairpersons: David Blair and Matteo Luca Ruggiero
YouTube links: https://youtu.be/W-WV6J8kprg
https://youtu.be/UHYpwKQ09SU

Teaching relativity: A paradigm change
F. Herrmann and M. Pohlig 2371

Teaching relativity: Computer-aided modeling
F. Herrmann and M. Pohlig 2381

Solstices and Equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, and the stellar aberration of Sirius
Costantino Sigismondi and Silvia Pietroni 2388

Positional astrometry at arcsecond accuracy using historical instruments, with light equipment
Costantino Sigismondi and Lorenzo Ricciardi 2398

Daily, seasonal, and equinoctial solar paths on a school soccer field
Costantino Sigismondi 2411

Teaching relativity at the AstroCamp
C. J. A. P. Martins 2415

Sungrazing comets as General Relativistic gravitational probes
Silvia Pietroni and Costantino Sigismondi 2424

The three Summer solstice's markers of 1721 unveiled in the Basilica of Santa Maria degli Angeli in Rome
Costantino Sigismondi 2428

Einstein-First: Bringing children our best understanding of reality
A. Popkova, K. Adams, S. Boublil et al. 2438

Exact Solutions

• **Exact Solutions in Four and Higher Dimensions**
Chairpersons: David Blair and Matteo Luca Ruggiero

YouTube link: https://youtu.be/GHZUS5-4gVQ

Kundt spacetimes in the Einstein–Gauss–Bonnet theory
R. Švarc, J. Podolský and O. Hruška 2453

Exact decoupled solutions in curvature-matter coupled gravity
M. Sharif and Fizza Furqan 2464

Tolman-Oppenheimer-Volkov conditions beyond spherical symmetry
José P. Mimoso, Alan Maciel and Morgan Le Delliou 2479

A spherically symmetric stiff fluid spacetime in light of cosmic structure formation
Daniele Gregoris 2497

• **Exact Solutions (Including Higher Dimensions)**
Chairperson: Susan Scott

YouTube link: https://youtu.be/-hJjJvmmOew

Three-parameter solution for the null-surface formulation in 2+1 dimensions
Tina A. Harriott and J. G. Williams 2510

New exact stationary cylindrical anisotropic fluid solution of GR
M.-N. Célérier 2522

Early Universe

• **Quantum Fields**
Chairperson: Andrei Lebed

YouTube links: https://youtu.be/BLaTp0r0TkQ
https://youtu.be/ZifgaSDy5Vc

Hydrodynamic representation and energy balance for the Dirac and Weyl fermions in curved space-times
Tonatiuh Matos, Omar Gallegos and Pierre-Henri Chavanis 2533

Breakdown of the Equivalence Principle for a composite quantum body
A. G. Lebed 2551

Extended DeWitt-Schwinger subtraction scheme, heavy fields and decoupling
Antonio Ferreiro and Jose Navarro-Salas 2557

Renormalization and decoupling for the Yukawa model in curved spacetime
Sergi Nadal-Gisbert, Antonio Ferreiro and José Navarro-Salas 2562

Trace anomaly and evaporation of spherical black holes
P. Meda 2573

On decay of the false Unruh vacuum
A. Shkerin 2587

Behaviour of noise kernel in de Sitter and FRW space-times
Ankit Dhanuka and Kinjalk Lochan 2600

Breaking Buchdahl: Ultracompact stars in semiclassical gravity
Julio Arrechea, Carlos Barceló, Raúl Carballo-Rubio and Luis J. Garay 2608

Einstein anomaly with tensors of odd order in six dimensional curved space
Kohei Yamamoto and Satoshi Yajima 2619

Quantum memory and BMS symmetries
Sanved Kolekar and Jorma Louko 2623

• Topological Methods, Global Existence Problems, and Spacetime Singularities
Chairperson: Spiros Cotsakis

YouTube link: https://youtu.be/H8Itnc_C10c

Gravitational singularities, scattering maps for bouncing, and structure-preserving algorithms
Philippe G. LeFloch 2630

Brane-world asymptotics in a nonlinear fluid bulk
I. Antoniadis, S. Cotsakis and Ifigeneia Klaoudatou 2645

Primordial synchronization of Mixmaster spatial points
Spiros Cotsakis 2657

• The Early Universe
Chairperson: Stefano Ansoldi

YouTube links: https://youtu.be/m80hHWgOlFs
https://youtu.be/6tyaZ8MMVtw

Quintessential inflation from Lorentzian slow roll
David Benisty 2663

Condensed light, quantum black holes and L-CDM cosmology: Experimentally suggested and tested unified approach to dark matter, dark energy, cosmogenesis and two-stage inflation
Victor Borsevici 2672

Helical magnetic fields lead to baryogenesis
Ashu Kushwaha and S. Shankaranarayanan 2692

Polymer Quantization of the Isotropic Universe: Comparison with the Bounce of Loop Quantum Cosmology
G. Barca, E. Giovannetti, F. Mandini and G. Montani 2700

General relativistic evolution equations for density perturbations in open, flat and closed FLRW universes and the problem of structure formation
Pieter G. Miedema 2708

Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data
M. Berti 2726

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2737

Fundamental Interactions and Stellar Evolution

• Why and How the Sun and the Stars Shine: The Borexino Experiment
Chairpersons: Giampaolo Bellini, Dmitry Naumov, Gioacchino Ranucci, Gemma Testera

YouTube links: https://youtu.be/hh5wDnM8miU
https://youtu.be/TG-HgBf7W4s
https://youtu.be/HLiis2LFeEs

Experimental detection of the CNO cycle
B. Caccianiga, N. Rossi, G. Testera et al. 2753

Borexino detector performances
A. Caminata, M. Agostini, K. Altenmüller et al. 2765

Study of antineutrinos from the Earth and the Cosmos with the Borexino detector
Sandra Zavatarelli, M. Agostini, K. Altenmuller et al. 2774

Unveiling the engine of the Sun: Measurements of the *pp*-chain solar neutrinos with Borexino
D. Guffanti, A. C. Re and O. Smirnov 2785

Electron neutrino survival probability in the energy range 200 keV–15 MeV
Marco Pallavicini on behalf of the Borexino Collaboration 2804

The relevance of pp-chain and CNO-cycle neutrino measurements for solar physics
F. L. Villante and A. M. Serenelli 2815

Role of the CNO cycles in stars
A. Ianni 2835

Geoneutrino observation
Tadao Mitsui for the KamLAND Collaboration 2840

Synthesis of the session: Why and how the Sun and the stars shine
Gianpaolo Bellini 2845

• **Rotation in Stellar Evolution**
Chairperson: Georges Meynet

YouTube link: https://youtu.be/9LNoiwa0nv8

The internal rotation of low-mass stars from solar and stellar seismology
G. Buldgen and P. Eggenberger 2848

The rotation of supermassive stars
L. Haemmerlé 2865

Fast Transients

• **What Can We Learn from a Growing Sample of Fast Radio Bursts?**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/yo4n1SgfUrQ
https://youtu.be/NwGooPauhjU

Cosmology with high-redshift FRBs
A. Fialkov 2880

• **Non Standard Cosmological Probes**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/EEFUgiFeMck
https://youtu.be/Ryb15AINfMs

Closing the cosmological loop with the redshift drift
C. J. A. P. Martins, C. S. Alves, J. Esteves, A. Lapel and B. G. Pereira 2890

Gamma-Ray Bursts as potential cosmological probes
L. Izzo 2906

Surface brightness fluctuations: The method and future applications
Michele Cantiello 2915

Preliminary results of analysis of Ia supernovae redshift distributions on data of the Asiago Supernova and Open Supernova Catalogues
I. V. Arkhangelskaja 2930

• **Photospheric Emission in GRBs**
Chairpersons: Gregory Vereshchagin and Damien Bégué

YouTube links: https://youtu.be/ZifnzoUXIFc
https://youtu.be/n7vWAVhFXiU

Understanding prompt emission: Where do we stand?
Asaf Pe'er 2946

On explaining prompt emission from GRB central engines with photospheric emission model
M. Bhattacharya and P. Kumar 2957

Monte Carlo simulations of photospheric emission in Gamma Ray Bursts
T. M. Parsotan and D. Lazzati 2972

The photosphere emission spectrum of hybrid relativistic outflow for Gamma-Ray Bursts
Yan-Zhi Meng, Jin-Jun Geng and Xue-Feng Wu 2982

On diffusive photospheres in Gamma-Ray Bursts
G. V. Vereshchagin 2989

Summary of the parallel session GB3
G. V. Vereshchagin and D. Bégué 3002

• High and Very High Energy Emission from Gamma-Ray Bursts
Chairpersons: Francesco Longo and Fabian Schüssler

YouTube links: https://youtu.be/ZUkLyyYowaM
https://youtu.be/zwN1mNVeUzA

Synchrotron and synchrotron self-Compton emission components in GRBs detected at very high energies
Jagdish C. Joshi, Vikas Chand and Soebur Razzaque 3009

The VERITAS gamma-ray burst follow-up program
D. Ribeiro for the VERITAS Collaboration 3017

MAGIC view of Gamma-Ray Bursts at very high energies
A. Berti on behalf of the MAGIC Collaboration 3030

Prospects for VHE monitoring of gamma-ray bursts with SWGO
G. La Mura, U. Barres de Almeida, R. Conceição et al. 3041

Theoretical implications on the very high energy emission from GRB 190114C
D. Miceli, A. Berti, Z. Bosnjak et al. on behalf of the MAGIC Collaboration 3052

AGILE and GRBs: 13 years of observations
A. Ursi on behalf of the AGILE Team 3062

Searching for Gamma-Ray Bursts with the High-Altitude Water Cherenkov (HAWC) observatory
K. L. Engel for the HAWC Collaboration 3074

• **Electromagnetic Counterparts of Compact Binary Mergers**
Chairpersons: Jonathan Granot and Paz Beniamini

YouTube link: https://youtu.be/AS7bwaT48us

CALET search for gamma-ray counterparts of gravitational wave events
Masaki Mori for the CALET Collaboration 3084

• **Unusual and New Types of Gamma-Ray Bursts**
Chairperson: Binbin Zhang

YouTube link: https://youtu.be/VrF0iU8Q6us

Off-axis jet scenario for early afterglow emission of low-luminosity Gamma-Ray Burst GRB 190829A
Yuri Sato, Kaori Obayashi, Ryo Yamazaki, Kohta Murase and Yutaka Ohira 3095

• **Gamma-Ray Burst Correlations: Observational Challenges and Theoretical Interpretation**
Chairpersons: Maria Giovanna Dainotti and Liang Li

YouTube links: https://youtu.be/gcEy2h6y1jg
https://youtu.be/m2pCT6RAmI0
https://youtu.be/GTCvcVhBjN4

GRB prompt phase spectra under backscattering dominated model
Mukesh Kumar Vyas, Asaf Pe'er and David Eichler 3101

Applying models of pulsar wind nebulae to explain X-ray plateaux following short Gamma-Ray Bursts
L. C. Strang and A. Melatos 3107

Searching for strange quark planets
Xu Wang, Yong-Feng Huang and Bing Li 3118

Probe the universe by using Gamma-Ray Bursts with X-ray plateaus
Fan Xu and Yong-Feng Huang 3124

A new perspective on cosmology through Supernovae Ia and Gamma Ray Bursts
B. De Simone, V. Nielson, E. Rinaldi and M. G. Dainotti 3130

Theory of plateau phase in Gamma-Ray Bursts
Asaf Pe'er 3141

Exploring the canonical behaviour of long Gamma-Ray Bursts using an intrinsic multi-wavelength afterglow correlation
S. R. Oates, J. L. Racusin, M. De Pasquale et al. 3150

• **GRB 170817A and Binary Models**
Chairpersons: Marica Branchesi and Giulia Stratta
YouTube link: https://youtu.be/FRBvkaLX5WU

Kilonova emission observed so far: A comparison with AT2017gfo
A. Rossi 3162

• **Binary-Driven Hypernovae of Type 1, 2 and 3**
Chairpersons: Carlo Luciano Bianco, Christian Cherubini and Simonetta Filippi
YouTube link: https://youtu.be/8IFLfget3C0

Neutrinos and gamma-ray production from proton-proton interactions in binary-driven hypernovae framework
S. Campion, J. D. Melon Fuksman and J. A. Rueda Hernandez 3172

General relativistic turbulence in spherically symmetric core-collapse supernovae simulations
L. Boccioli, G. J. Mathews and E. P. O'Connor 3184

Gravitational Waves

• **Sources of Gravitational Waves**
Chairperson: Andrew Melatos
YouTube links: https://youtu.be/Lr8b9nKsFm4
https://youtu.be/xuqJOXPDyVY
https://youtu.be/qzOrgojHsMg
https://youtu.be/KCJoWMd71pE

Mountain formation by repeated, inhomogeneous crustal failure in a neutron star
A. D. Kerin and A. Melatos 3194

Gravitational waves from neutrino mass generating phase transitions
Nobuchika Okada and Osamu Seto 3206

Efficiency of registration of chirp bursts and signals of collapsing stars by the Euro-Asian network of GW interferometers
V. N. Rudenko, S. L. Andrusenko, D. P. Krichevskiy and G. D. Manucharyan 3219

Joint analysis method on gravitational waves and low-energy neutrinos to detect core-collapse supernovae
O. Halim, C. Casentini, M. Drago et al. 3228

• **Mid-frequency Gravitational Waves (0.1–10 Hz): Sources and Detection Methods**
Chairperson: Wei-Tou Ni
YouTube link: https://youtu.be/sJ6A7a73Vxw

A cryogenic and superconducting inertial sensor for the Lunar Gravitational–Wave Antenna, the Einstein Telescope and Selene-physics
F. Badaracco, J. V. van Heijningen, E. C. Ferreira and A. Perali 3245

Space gravitational wave antenna DECIGO and B-DECIGO
S. Kawamura and the DECIGO working group 3254

Summary of the parallel session GW2
Dongfeng Gao, Wei-Tou Ni, Jin Wang et al. 3261

• **Numerical Relativity and Gravitational Wave Observations**
Chairperson: Nigel Bishop

YouTube links: https://youtu.be/eZPytAU4Zmk
https://youtu.be/th-3KqkxDnU

Salient features of the optimised PyCBC IMBH search
Koustav Chandra, Archana Pai, V. Villa-Ortega et al. 3277

Matter shells modifying gravitational wave signals
Monos Naidoo, Nigel T. Bishop and Petrus J. van der Walt 3286

Odd-dimensional gravitational waves from a binary system on a three-brane
D. V. Gal'tsov and M. Khlopunov 3301

Developments in numerical relativity and gravitational wave observations
Nigel T. Bishop 3309

High Energy

• **Very High Energy Gamma Rays**
Chairpersons: Razmik Mirzoyan and Alessandro De Angelis

YouTube link: https://youtu.be/jHM1RH20ZyM

Insights into the Galactic Center environment from VHE gamma-ray observations with ground-based facilities
C. Fruck 3316

The TAIGA experiment
M. Tluczykont, I. I. Astapov, A. K. Awad et al. 3324

Science perspectives of the Southern Wide-field Gamma-ray Observatory (SWGO)
K. L. Engel for the SWGO Collaboration 3343

• **Future Missions for High-Energy Astrophysics**
Chairpersons: Filippo Frontera and Shaolin Xiong

YouTube link: https://youtu.be/JgDsZX6RUkU

Laue lenses: Focusing optics for hard X/soft Gamma-ray astronomy
L. Ferro, M. Moita, P. Rosati et al. 3355

ASTENA: A mission concept for a deep study of the transient gamma-ray sky and for nuclear astrophysics
E. Virgilli, F. Frontera, P. Rosati et al. on behalf of the ASTENA Collaboration 3368

Polarimetric prospects of a new hard X-soft gamma-ray space mission for next decades
M. Moita, L. Ferro, F. Frontera et al. 3385

• **The SRG Mission: First Results from eROSITA and ART-XC**
Chairperson: Andrea Merloni
YouTube link: https://youtu.be/l0t1B716UcM

Prospect for WHIM detection in the cosmic web by *SRG/eROSITA*
H. Tanimura and N. Aghanim 3400

• **eXTP – Enhanced X-Ray Timing and Polarimetry Mission**
Chairpersons: Marco Feroci and Fangjun Lu
YouTube link: https://youtu.be/3d08eKmsImI

The role of eXTP in the multi-messenger astronomy era
G. Stratta and Gor Oganesyan 3403

• **Observations of HE and UHE Cosmic Rays**
Chairpersons: Ivan De Mitri and Fabio Gargano
YouTube link: https://youtu.be/l9DZyBMWO5c

CALET on the ISS: The first 5 years
Pier Simone Marrocchesi for the CALET Collaboration 3427

The fluxes of charged cosmic rays as measured by the DAMPE satellite
Paolo Bernardini on behalf of the DAMPE Collaboration 3442

Recent results from the Pierre Auger Observatory
E. Roulet for the Pierre Auger Collaboration 3449

The HERD space mission
F.C.T. Barbato on behalf of the HERD Collaboration 3455

PART D

History of Relativity

• **The "Fall and Rise" of Betelgeuse**
Chairperson: Costantino Sigismondi
YouTube link: https://youtu.be/VmbrE2gYmOM

The observation of the stars in daytime and near the horizon
Costantino Sigismondi and Paolo Ochner 3471

Fall and Rise of Betelgeuse: The summary of HR1 session
Costantino Sigismondi and Paolo Ochner 3475

Photometry of Betelgeuse at daylight
Otmar Nickel 3479

Evidence for dynamical changes in Betelgeuse using multi-wavelength data
Sneha Kachhara, Sandip V. George, Ranjeev Misra and G. Ambika 3485

The curious case of Betelgeuse
Jacco Th. van Loon 3494

Variable stars observed from city sites: The 2500 AAVSO-SGQ database
Costantino Sigismondi and Paolo Ochner 3501

Betelgeuse: An introductory course to observational astronomy
Costantino Sigismondi and Tiziana Pompa 3507

The meridian line of the Vatican obelisk to study the stellar aberration
Costantino Sigismondi and Lorenzo Ricciardi 3513

Betelgeuse, Sirius and the stars in the roman *Settecento*
Costantino Sigismondi 3519

• **History of Relativity, Gravitation and Cosmology**
Chairperson: Luis Crispino

YouTube link: https://youtu.be/RNbPUSp95PQ

On Einstein's last bid to keep a stationary cosmology
Salvador Galindo-Uribarri and Jorge L. Cervantes-Cota 3536

Jayme Tiomno: Relativity, gravity, cosmology, and the Marcel Grossmann Meetings
William D. Brewer 3547

A look inside Feynman's route to gravitation
M. Di Mauro, S. Esposito and A. Naddeo 3563

Towards detecting gravitational waves: A contribution by Richard Feynman
M. Di Mauro, S. Esposito and A. Naddeo 3576

Stellar gravitational collapse, singularity formation and theory breakdown
Kiril Maltsev 3596

The Hamilton-Jacobi analysis by Peter Bergmann and Arthur Komar of classical general relativity
D. Salisbury 3626

• Time and Philosophy in Physics
Chairperson: Shokoufe Faraji

YouTube link: https://youtu.be/986v-V5JJEk

The passage of time and top-down causation
Barbara Drossel 3631

Explaining time's passage
Jonathan J. Dickau 3646

A glimpse to Feynman's contributions to the debate on the foundations of quantum mechanics
M. Di Mauro, S. Esposito and A. Naddeo 3657

Summary of the parallel session HR3
Shokoufe Faraji 3671

Neutron Stars

• Dense Matter in Compact Stars
Chairpersons: Alessandro Drago and Jorge Rueda

YouTube links: https://youtu.be/U5Yr0oDVhqY
https://youtu.be/MvMDFh-1_bc

Massive compact stars in the two-families scenario
P. Char, A. Drago and G. Pagliara 3677

Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond
J. Soldateschi, N. Bucciantini and L. Del Zanna 3684

Screening and elastic properties of the NS crust in the OCP approximation
D. Barba González, C. Albertus Torres and M. A. Pérez-García 3703

Tidal deformability as a probe of dark matter in neutron stars
D. Rafiei Karkevandi, S. Shakeri, V. Sagun and O. Ivanytskyi 3713

Binary neutron star mergers with quark matter equations of state
Atul Kedia, Hee Il Kim, Grant Mathews and In-Saeng Suh 3732

Probing dense matter physics with transiently-accreting neutron stars: The case of source MXB 1659-29
Melissa Mendes, Andrew Cumming, Charles Gale and Farrukh J. Fattoyev 3736

• **Compact Stars as Laboratories for Testing Strong Gravity**
Chairpersons: Aurora Perez Martinez and César Augusto Zen Vasconcellos

YouTube link: https://youtu.be/C1KGecdNEfs

Vacuum properties and astrophysical implications
A. Pérez Martínez, M. Pérez-Garcia, E. Rodríguez Querts and A. Romero Jorge 3756

Testing extended theories of gravity with GRBs
L. Mastrototaro 3762

• **Pulsar Power in Physics and Astrophysics and Pulsars and Pulsar Systems at High Energies**
Chairpersons: Andrea Possenti and Pak-Hin Tam

YouTube links: https://youtu.be/gAG29DZwUbM
https://youtu.be/ytdUBFrViHI

News and views regarding PSR J1757–1854, a highly-relativistic binary pulsar
A. D. Cameron, M. Bailes, V. Balakrishnan et al. 3774

On the origin of the unique isolated X-Ray pulsar 1E 161348-5055 with 6.7 hr. spin period
V. Yu. Kim 3785

Advantages of including globular cluster millisecond pulsars in Pulsar Timing Arrays
M. Maiorano, F. De Paolis, A. A. Nucita and A. Franco 3791

Searching for pulsars in globular clusters with the MeerKAT Radio Telescope
F. Abbate on behalf of the MeerTIME/TRAPUM Collaboration 3799

Precision Tests

• **Gravitational Lensing and Shadows**
Chairpersons: Perlick Volker and Oleg Tsupko

YouTube links: https://youtu.be/6DXXWpMJ3IQ
https://youtu.be/F-K1gXn71_Y
https://youtu.be/B-V8r2HMztw

Gravitational lensing by rotating Simpson–Visser black holes
Sushant G. Ghosh and Shafqat Ul Islam 3812

Killing tensors in foliated spacetimes and photon surfaces
Igor Bogush, Kirill Kobialko and Dmitri Gal'tsov 3827

Decoding black hole metrics from the interferometric pattern of relativistic images
V. Bozza 3839

Symplectic evolution of an observed light bundle
N. Uzun 3844

Shadow of black holes with a plasma environment in 4D Einstein-Gauss-Bonnet gravity
Javier Badía and Ernesto F. Eiroa 3856

Aspects of neutrino mass hierarchy in gravitational lensing
Himanshu Swami 3865

Photon regions in stationary axisymmetric spacetimes and umbilic conditions
K. V. Kobialko and D. V. Gal'tsov 3874

Gravitational lensing by charged accelerating black holes
Torben C. Frost 3885

• Experimental Gravitation
Chairpersons: Angela di Virgilio and Claus Lammerzahl

YouTube link: https://youtu.be/mXcxztQ0nyk
https://youtu.be/tE3gIUBviTc

A manmade experiment aimed to clarify the gravity law in the Solar system
Alexander P. Yefremov and Alexandra A. Vorobyeva 3905

Gravitomagnetic field generation using high permittivity materials in superconducting magnetic energy storage devices
G. V. Stephenson 3910

Large ring laser gyroscopes: Geometry stabilization and laser control
U. Giacomelli, N. Beverini, G. Carelli et al. 3920

Dark gravitomagnetism with LISA and gravitational waves space detectors
A. Tartaglia, M. Bassan, G. Pucacco, V. Ferroni and D. Vetrugno 3929

Light rays in the Solar system experiments: Phases and displacements
Pravin Kumar Dahal and Daniel R. Terno 3942

The Ginger project – preliminary results
C. Altucci, F. Bajardi, A. Basti et al. 3956

• **Variation of the Fundamental Constants, Tests of the Fundamental Symmetries and Probes of the Dark Sector**
Chairpersons: Angela Victor Flambaum and Yevgeny Stadnik

YouTube links: https://youtu.be/NAZA-0tWHak
https://youtu.be/juhDGBJ12Lg
https://youtu.be/zxGgDP2sn60

Varying fundamental constants and dark energy in the ESPRESSO era
C. J. A. P. Martins 3963

• **Dragging is Never Draggy: MAss and CHarge Flows in GR**
Chairperson: Oldrich Semerak

YouTube link: https://youtu.be/ho31IgLNxu8

Testing the general relativistic nature of the Milky Way rotation curve with Gaia DR2
Mariateresa Crosta 3970

Spinning cylinders in general relativity: A canonical form for the Lewis metrics of the Weyl class
L. Filipe O. Costa, José Natário and N. O. Santos 3982

Magnetized black holes: The role of rotation, boost, and accretion in twisting the field lines and accelerating particles
Ondřej Kopáček and Vladimír Karas 3999

Spinning particle: Is Newton-Wigner the only way?
V. Witzany 4010

Gravitomagnetic resonance and gravitational waves
Matteo Luca Ruggiero and Antonello Ortolan 4019

Quantum Gravity

• **Loop Quantum Gravity**
Chairpersons: Marcin Kisielowski and Jerzy Lewandowski

YouTube links: https://youtu.be/VaLPseYWh9E
https://youtu.be/WFIgMqrQ07U

Studying the EPRL spinfoam self-energy
Pietropaolo Frisoni 4026

A spin foam framework for the black-to-white hole transition
Farshid Soltani 4045

Holographic properties of the bulk-to-boundary transmission of information in regions of quantum space
Eugenia Colafranceschi 4062

• **Quantum Gravity Phenomenology**
Chairpersons: Giovanni Amelino-Camelia and Jerzy Kowalski-Glikman

YouTube links: https://youtu.be/vYnVb2zNl0o
https://youtu.be/icK8z80Hm-Y
https://youtu.be/svXLa0yhyYY

Minimal length discretization and properties of modified metric tensor and geodesics
Abdel Nasser Tawfik, Fady T. Farouk, F. Salah Tarabia and Muhammad Maher 4074

The structure of the multiverse from the entanglement entropy
Samuel Barroso Bellido 4082

Effective field theory from relativistic Generalized Uncertainty Principle
Vasil N. Todorinov, Saurya Das and Pasquale Bosso 4088

Stelle gravity as the limit of quantum gravity with maximal momentum
V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das 4107

Baryon asymmetry and minimum length
Saurya Das, Mitja Fridman, Gaetano Lambiase and Elias C. Vagenas . . . 4114

On quantum gravity and quantum gravity phenomenology
Douglas Edmonds, Djordje Minic and Tatsu Takeuchi 4126

WKB approach to the gravity-matter dynamics: A cosmological implementation
G. Maniccia and G. Montani 4146

Natural evidence for fuzzy sphere noncommutative geometry: Super-Chandrasekhar white dwarfs
Surajit Kalita, T. R. Govindarajan and Banibrata Mukhopadhyay 4159

A model of polymer gravitational waves: Theory and some possible observational consequences
Angel Garcia-Chung, James B. Mertens, Saeed Rastgoo, Yaser Tavakoli and Paulo Vargas Moniz 4166

• Loop Quantum Gravity: Cosmology and Black Holes
Chairpersons: Jorge Pullin and Parampreet Singh

YouTube links: https://youtu.be/CAuAK31Ukho
https://youtu.be/JlYnEzRiRR0

Primordial power spectrum from a matter-ekpyrotic scenario in loop quantum cosmology
Bao-Fei Li, Sahil Saini and Parampreet Singh 4178

The primordial power spectra in modified loop quantum cosmology
Bao-Fei Li, Javier Olmedo, Parampreet Singh and Anzhong Wang 4188

Primordial perturbations in kinetically dominated regimes of classical and quantum cosmology
B. Elizaga Navascués, R. Jiménez-Llamas and G. A. Mena Marugán 4193

Revisiting the Hamiltonian formalism of the Ashtekar–Olmedo–Singh black hole model
Alejandro García-Quismondo and Guillermo A. Mena Marugán 4211

A comparison of different choices of clocks in a reduced phase space quantization in loop quantum cosmology with an inflationary potential using effective techniques
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4228

Initial conditions in LQC/mLQCs
Bao-Fei Li, Parampreet Singh and Anzhong Wang 4234

Holonomy corrections in effective midisuperspace models
A. Alonso-Bardaji and D. Brizuela 4239

Infrared signatures of quantum bounce in collapsing geometry
Harkirat Singh Sahota and Kinjalk Lochan 4247

Effective black hole interior and the Raychadhuri equation
Keagan Blanchette, Saurya Das, Samantha Hergott and Saeed Rastgoo . . . 4256

Effect of loop quantization prescriptions on the physics of non-singular gravitational collapse
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4267

Summary of the parallel session QG3
Jorge Pullin and Parampreet Singh 4272

Strong Field

• Strong Electromagnetic and Gravitational Field Physics: From Laboratories to Early Universe

Chairpersons: Sang Pyo Kim and She-Sheng Xue

YouTube links: https://youtu.be/kexqTayqFiU
https://youtu.be/laymGp6x9Hg

Reliable equations of state of viscous strong and electroweak matter
A. Tawfik 4277

Neutral fermion pair production by Sauter-like magnetic step
T. C. Adorno, Zi-Wang He, S. P. Gavrilov and D. M. Gitman 4290

On the magnetic field screening in strong crossed electromagnetic field
S. Campion, J. A. Rueda Hernandez, S.-S. Xue and R. Ruffini 4306

Particle creation by strong fields and quantum anomalies
José Navarro-Salas 4317

Constraints on the non-minimal coupling of electromagnetic fields from astrophysical observations
Susmita Jana and S. Shankaranarayanan 4326

New partial resummation of the QED effective action
Silvia Pla and José Navarro-Salas 4337

Can a detector detect soft photons
Sanved Kolekar and Jorma Louko 4347

Breaking of the adiabatic invariance in the production of particles by strong fields
P. Beltrán-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla 4352

Dynamics of relativistic electrons in non-uniform magnetic fields and its applications in quantum computing and astrophysics
Srishty Aggarwal and Banibrata Mukhopadhyay 4362

Validity of the semiclassical approximation in 1+1 electrodynamics: Numerical solutions to the linear response equation
Ian M. Newsome 4374

On Kerr black hole perfect MHD processes in Doran coordinates
C. Cherubini, S. Filippi, A. Loppini et al. 4387

Tadpole contribution to magnetic photon-graviton conversion
N. Ahmadiniaz, F. Bastianelli, F. Karbstein and C. Schubert 4393

• **The Effects of (Non)Linear Electrodynamics on the Properties of Astrophysical/Gravitational Compact Objects**
Chairperson: Seyed Hossein Hendi

YouTube link: https://youtu.be/SAIAXtQhavE

Correspondence of gamma radiation coming from GRBs and magnetars based on the effects of nonlinear vacuum electrodynamics
Tursynbek Yernazarov, Medeu Abishev and Yerlan Aimuratov 4401

Absorption of massless scalar waves by electrically charged regular black holes
Marco A. A. de Paula, Luiz C. S. Leite and Luís C. B. Crispino 4410

White Dwarfs

• **White Dwarf Explosions**
Chairpersons: Robert Fisher and María Pilar Ruiz Lapuente

YouTube links: https://youtu.be/ndaW6u2xuOo
https://youtu.be/IwjYaQaqJeI

Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass
Friedrich K. Röpke, Florian Lach, Sabrina Gronow et al. 4420

Type Ia supernovae and their explosive nucleosynthesis: Constraints on progenitors
Shing-Chi Leung and Ken'ichi Nomoto 4427

Charged polarized white dwarfs with finite temperature as a possible source of type Ia supernovae
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4447

• **White Dwarfs, Magnetic Compact Stars and Nuclear Astrophysics**
Chairpersons: Manuel Malheiro and Jaziel Goulart Coelho

YouTube link: https://youtu.be/onicFElJQnA

CTCV J2056-3014 and other fast-spinning white dwarfs
C. V. Rodrigues, R. Lopes de Oliveira, A. Bruch et al. 4453

Gravitational waves from fast-spinning white dwarfs
M. F. Sousa, J. C. N. de Araujo and J. G. Coelho 4461

Highly magnetized white dwarfs: Implications and current status
B. Mukhopadhyay, M. Bhattacharya, A. J. Hackett et al. 4475

Electron captures and stability of white dwarfs
N. Chamel, L. Perot, A. F. Fantina et al. 4488

Massive hot white dwarfs: Consequences of finite temperature in the structure and on the onset of instabilities
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4508

A study of the infrared emission of SGR/AXPs in a disk scenario and its implications for their origin
Sarah Villanova Borges 4514

Particle acceleration and high energy emission in the white dwarf binaries AE Aquarii and AR Scorpii
P. J. Meintjes, S. T. Madzime, Q. Kaplan et al. 4522

Study the effects of anisotropy on the highly magnetized white dwarfs
Debabrata Deb, Banibrata Mukhopadhyay and Fridolin Weber 4532

List of Participants 4545

PART B

This page intentionally left blank

Reconstruction of a star motion in the vicinity of black hole from the redshift of the electromagnetic spectrum

S. O. Komarov* and A. K. Gorbatsievich
Theoretical Physics and Astrophysics Department, Belarusian State University, Minsk, Belarus
**E-mail: StasKomarov@tut.by*
www.KomarovSO.bsu.by

The problem of calculating the redshift of electromagnetic spectrum of the star, moving in the vicinity of Schwarzschild black hole is solved within the framework of the General Theory of Relativity. The inverse problem — determination the parameters of the motion of a star from observational data of redshift is considered. The approach that gives possibilities to solve the inverse problem is proposed. The approach is tested on the numerical model that gives possibilities to calculate redshift as function of time of observation for a star moving in the vicinity of Schwarzschild black hole. The parameters of the star in numerical model are close to parameters of the S-stars, moving in the vicinity of the Sgr A*.

Keywords: Black hole; gravitational redshift.

1. Introduction

It is well-known from astrophysical observations that supermassive black hole with mass $m_{SBH} \approx 4 \cdot 10^6 m_\odot$,[1,2] where $m_\odot$ is the mass of the Sun, exists in the Galactic Center.[1,3–5] Apart from this, large amount of stars exists in this region. For example, S-cluster includes stars closest to supermassive black hole.[2,6–8]

Astrophysical observations of such stars give possibility to study the structure of the Galactic Center and to test theories of gravity. The main source of information about the motion of these stars is their electromagnetic radiation. In the present work, we have performed theoretical investigation of the redshift of electromagnetic radiation of a star moving in the vicinity of a black hole. This problem includes two parts: the direct problem — calculation the redshift of the electromagnetic radiation of a star moving in external gravitational field and the inverse problem: — determination the motion of the star in external gravitational field if redshift as function of time of observation is known.

The direct problem within the framework of the General Relativity is considered in many papers see, e.g.,[6–12]. The mentioned studies consider different general relativistic effects such as Shapiro delay, gravitational redshift, and Doppler shift. To solve the direct problem, one needs to solve the boundary value problem for the isotropic geodesic that connects the source and the observer. In the cited studies, if the corresponding general relativistic effects are taken into account, this problem is solved using tables of impact parameters that make the accuracy of the solution limited by the step of the data in the tables. In our previous papers,[13–15] we developed a covariant approach that allows one to obtain compact expressions

for redshift as a function of observation time. We have solved the boundary value problem by numerically solving a non-linear equation, which allows for much more accurate solutions than table-based methods.

For the solution of inverse problem it is necessary to use statistical methods such us MCMC method.[6,8–10,12] But mentioned studies do not contain any approach for constructing initial guess for the solution of the problem. The method that is presented in this paper gives possibility to obtain such guesses from graphics.

As an example of solving the inverse problem, we consider a mathematical model of a star moving close to a supermassive black hole. For the demonstration purpose, we chose the parameters of motion corresponding to a star on a slightly closer orbit around the black hole than the orbits of known S-stars (see, e.g.,[2,5]). Such orbits allow us to test the approach in the strong-field regime apply to the cases when the sources are on very tight orbits near the Galactic Center. Sources with such orbits may be found with future observations.

In this paper, we use a system of units where the speed of light in the vacuum is equal to unity ($c = 1$), and the metric has signature $-\,+\,+\,+$.

2. Theoretical model

In the present paper consider only the case of spherically symmetric non-charged black hole. In General Relativity such a black hole can be described by the Schwarzschild metric (see, e.g.,[16]):

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1-2M/r} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\varphi^2 - \left(1-\frac{2M}{r}\right)\mathrm{d}t^2\,. \tag{1}$$

Here, $x^i = \{t,\, r,\, \theta,\, \varphi\}$ are Schwarzschild coordinates. Furthermore, $M = Gm_{SBH}$, where G is the gravitational constant. In our model, the mass of the black hole m_{SBH} is much larger than the mass of the star $m_s \sim m_\odot$. Because of this, we model source stars as test particles moving in the external gravitational field of the supermassive black hole. One can obtain the 4-velocity components of the star from the geodesic equation. They have the following form (see, e.g.,[16]):

$$\begin{aligned}
u^0 &= \frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{E}{(1-2M/r)}\,;\\
u^1 &= \frac{\mathrm{d}r}{\mathrm{d}\tau} = e_s\sqrt{E^2-(1-2M/r)(1+L^2/r^2)}\,;\\
u^2 &= \frac{\mathrm{d}\theta}{\mathrm{d}\tau} = 0\,;\\
u^3 &= \frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \frac{L}{r^2}\,,
\end{aligned} \tag{2}$$

where we chose the orientation of the spatial part of the coordinate system in such a way that the trajectory of the star lies in the plane $\theta = \pi/2$. Here L is the specific angular momentum of the star (in mass), E is its specific energy (in mass), and τ is its proper time. Factor e_s describes whether the considered part of the trajectory is receding or approaching.

From the system of equations (2), one can find the trajectory of the star in analytic form. In the case of finite motion, it has the following form (see, e.g.,[17]):

$$\frac{1}{r} = \frac{1}{r_s(\varphi,\, \delta,\, p_1,\, p_2)} = \frac{1}{p_1} - \frac{p_2 - p_1}{p_1 p_2}\,\mathrm{sn}^2\left[\frac{(\varphi-\delta)}{2}\sqrt{1 - \frac{4M}{p_2} - \frac{2M}{p_1}},\, k_s\right], \qquad (3)$$

where

$$k_s = \sqrt{\frac{p_2 - p_1}{p_1 p_2/(2M) - p_2 - 2p_1}},$$

where $\mathrm{sn}[\varphi,\, k]$ is the Jacobi sine of the first kind (see[18] for definition), δ is the longitude of pericenter, and p_1 and p_2 are pericenter and apocenter distances respectively. They are uniquely related to E and L as follows:

$$L = \frac{p_1 p_2}{\sqrt{\left(1 + \frac{p_1 + p_2}{2M}\right) p_1 p_2 - (p_1 + p_2)^2}}\,; E = \sqrt{\frac{(p_1 + p_2)\left(2M + \frac{p_1 p_2}{2M} - p_1 - p_2\right)}{\left(1 + \frac{p_1 + p_2}{2M}\right) p_1 p_2 - (p_1 + p_2)^2}}\,. \qquad (4)$$

The proper time of the star τ can be expressed as a function of its angular coordinate φ by using the well-known analytic formula (see, e.g.,[17]):

$$\tau = \tau_s(\varphi,\, p_1,\, p_2) + \tau_0\,. \qquad (5)$$

Here, $\tau_s(\delta,\, p_1,\, p_2) = 0$. We will not write down this expression explicitly due to cumbersomeness.

Astrophysical observations of stars in the vicinity of the Galactic Center use electromagnetic radiation in the wavelength range of $1\mu\mathrm{m} - 10\mathrm{m}$ (including observations of pulsars, see, e.g.,[3]). Such wavelengths are small compared to the typical orbit sizes of S-stars, which allows us to use the geometric optics approximation (see, e.g.,[19]). In this approximation, electromagnetic radiation propagates along a null geodesic with tangent vector k_i that satisfies the following relations: $k_j k^j = 0$ and $k_{i;j} k^j = 0$. Here we chose the following coordinate frame $\tilde{K} : \{t,\, r,\, \tilde{\theta},\, \tilde{\varphi}\}$. Therefore, the observer resides on the axis $\tilde{\theta} = 0$, $\tilde{\varphi} = 0$, we find that the trajectory of the light ray lies in the plane $\tilde{\varphi} = const$ and obtain (see, e.g.,[16]):

$$\begin{aligned}
k^0 &= \frac{\mathrm{d}t}{\mathrm{d}\nu} = \frac{1}{(1 - 2M/r)}\,;\\
k^1 &= \frac{\mathrm{d}r}{\mathrm{d}\nu} = e_r\sqrt{1 - (1 - 2M/r)D^2/r^2}\,;\\
k^2 &= \frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}\nu} = -\frac{D}{r^2}\,;\\
k^3 &= \frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\nu} = 0\,,
\end{aligned} \qquad (6)$$

where ν is an affine parameter along the ray and D is the impact parameter. Factor e_s describes whether the considered light trajectory is receding or approaching.

The sign in the expression for $\tilde{\theta}$ in equation (6) is chosen so that $D > 0$. We only consider zeroth-order trajectories, i.e. trajectories for which the increment of $\tilde{\theta}$ from the source to observer is less then π (see, e.g.,[20,21]).

From equations (6) and by using the boundary condition $r \to \infty$ for $\tilde{\theta} = 0$, we obtain the following analytic expression for the trajectory of the ray:

$$\frac{1}{r} = \frac{1}{r_r(\tilde{\theta},\, D)} = \frac{1}{P} - \frac{Qk^2}{2PM}\,\mathrm{cn}^2\left[\frac{\tilde{\theta}}{2}\sqrt{\frac{Q}{P}} + \mathrm{F}\left[\arccos\left(\sqrt{\frac{2M}{Qk^2}}\right), k\right], k\right],$$

where $r_r = r$ is satisfied for the points on the world line of the ray,

$$Q = \sqrt{P^2 + 4PM - 12M}\,; \qquad k = \sqrt{\frac{Q - P + 6M}{2Q}}\,, \tag{7}$$

$\mathrm{cn}\,[\varphi,\, k]$ and $\mathrm{F}\,[\varphi,\, k]$ are the Jacobi cosine and the elliptic integral of the first kind, respectively (see[18] for definition). If real, P has a physical meaning of the closest approach distance (see, e.g.,[17]). However, whether P is real or complex, it can be expressed through the impact parameter D as follows:

$$P = -\frac{2}{\sqrt{3}} D \sin\left[\frac{1}{3}\arcsin\left(\frac{3\sqrt{3}M}{D}\right) - \frac{\pi}{3}\right].$$

The angular coordinates in both coordinate systems are connected by the following relation (see Fig. 1):

$$\tilde{\theta} = \arccos[\cos(\varphi)\sin(i_0)]\,, \tag{8}$$

where angle i_0 is the inclination of the orbit of the star.

Redshift of the spectrum of electromagnetic radiation can be calculated from the formula (see, e.g.,[10]):

$$z = \frac{\delta\lambda}{\lambda} = \frac{(u_i)_s (k^i)_s}{(u_j)_o (k^j)_o} - 1\,. \tag{9}$$

Here, λ is the wavelength of emitted light, $\delta\lambda$ is the wavelength difference between received and emitted light. $(k_i)_o$ and $(k_i)_s$ denote the wave vector at the location of the observer and source, respectively. Likewise, $(u_i)_o$ and $(u_i)_s$ denote the 4-velocity vector of the observer and the star, respectively.

Consider the case of a stationary observer located at spatial infinity. Therefore, the spacetime around the observer is described to good accuracy by the Minkowsky metric. Therefore, for the observer we have $(u^i)_o = \{1,\, 0,\, 0,\, 0\}$. To calculate the components of the wave vector $(k_i)_s$, it is necessary to solve the boundary value problem for the system of differential equations (6) (see Fig. 2). The zeroth-order trajectory corresponds to the maximal intensity, and one may find it almost in all cases for S-stars (see, e.g.,[20,21]). Because of this, we will consider only light rays of zeroth order. Therefore, the solution is unique. For the chosen assumptions, solving the mentioned boundary value problem reduces to solving the following non-linear ordinary equation for the impact parameter D:

$$r_s(\varphi,\, p_1,\, p_2) = r_r(\tilde{\theta},\, D)\,. \tag{10}$$

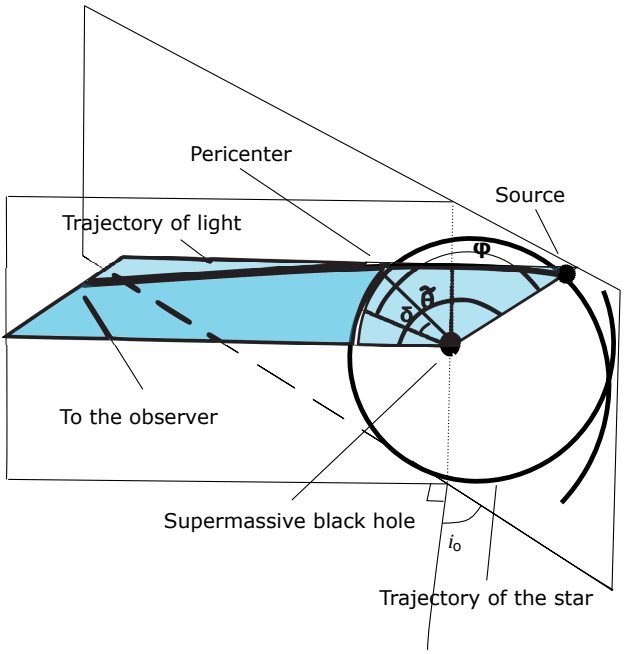


Fig. 1. For the derivation of formula (8)

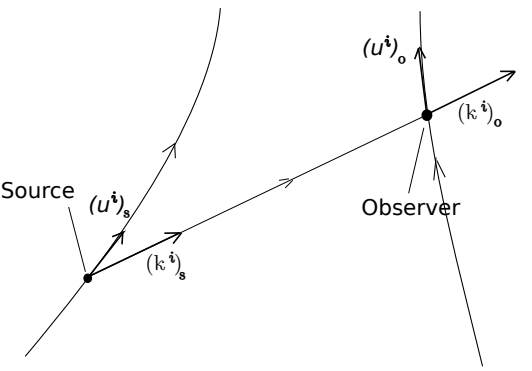


Fig. 2. $(k^i)_o$ and $(k^i)_s$ are tangent vectors to null geodesic that intersect both the worldline of the source and the worldline of the observer

Taking into account the stationarity of the observer, the relation between the angles (8), and substituting expressions (2), (6) into (9), we obtain the redshift in the following form:

$$z = -1 + \frac{E}{q} + \frac{DL}{r^2}\beta - e_s e_r \frac{1}{q}\sqrt{\left(E^2 - q\left(1 + \frac{L^2}{r^2}\right)\right)\left(1 - q\frac{D^2}{r^2}\right)}. \qquad (11)$$

Here, we denote $1 - 2M/r = q$ and $\beta = \sin(i_0)\sin(\varphi)/\sin(\tilde{\theta})$. The presented equations allow one to solve the direct problem: calculating the redshift of a star moving

in an external gravitational field of a supermassive black hole as a function of observation time. We illustrate the method by using a numerical model as shown in Fig. 3.

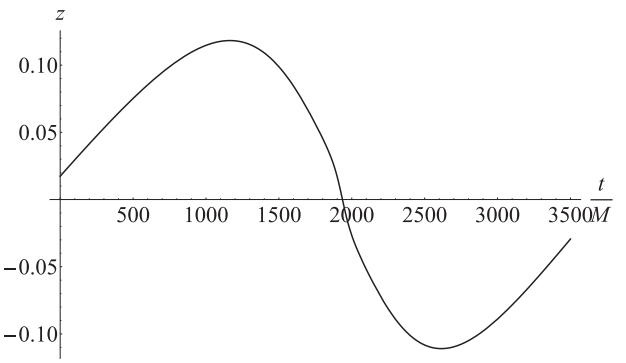


Fig. 3. Redshift of electromagnetic spectrum of a star in an external Schwarzschild gravitational field as a function observation time t. The pericenter distance of the stellar orbit is $p_1 = 60M$, its apocenter distance is $p_2 = 90M$, its inclination is $i_0 = 1.4\,\mathrm{rad}$, the longitude of pericenter is $\delta = 1\,\mathrm{rad}$ and the initial time of pericenter passage is $\tau_0 = 0M$

However, it is more interesting for astrophysical purposes to solve the inverse problem: determining the parameters of motion of a star in the external gravitational field of a supermassive black hole based on its redshift data. In the literature, the inverse problem is solved by minimizing the χ^2 function (see, e.g.,[9,10]):

$$\chi^2 = \sum_{j=1}^{N} \left[\frac{(z_j - z_{\mathrm{obs},j})^2}{\sigma_Z^2} \right] , \tag{12}$$

where z_j and $z_{\mathrm{obs},j}$ are the theoretical and observed values of the redshift, respectively, for the times of observation t_j ($j \in [1,\, N]$). σ_Z^2 is the dispersion of the redshift observation data. Since function $z_{\mathrm{obs},j}(t)$ has no explicit expression (at least, because D is the solution of non-linear equation 10), minimizing χ^2 can only be performed numerically (for example, using the Metropolis-Hastings algorithm[22]).

In this work, we present another approach based on deriving a system of equations expressed explicitly using elementary or special functions of the parameters of motion of the star. To obtain such equations for the inverse problem, one has to find expressions not only for z but also for $\mathrm{d}z/\mathrm{d}\tau$. We describe this calculation in the following section.

3. Derivative of the redshift function

3.1. *Newman-Penrose null tetrad and optical scalars*

In this section we will use the Newman-Penrose null tetrad (see, e.g.,[23,24]), determined along the world line of the ray emitted by the star:

$$k^i,\, n^i,\, m^i\, \bar{m}^i\,. \tag{13}$$

Here k^i is the wave vector of the ray. Symbol $^-$ denotes the complex conjugation. All vectors in (13) are null. All scalar products between vectors in (13) are equal to 0 apart from

$$k_i n^i = -1;\, m_i \bar{m}^i = 1\,. \tag{14}$$

Consider a congruence of isotropic geodesics that have tangent vectors k^i and intersect the world line of the observer at time t_o. Also, consider the Newman-Penrose tetrad (13) at all points of this congruence. Then the components of the vectors of the tetrad in the coordinate basis of $\tilde{K}$ have the form (we chose the affine parameter ν such that k_0=-1):

$$\begin{aligned}
&k_j = \left\{-1,\, e_r \frac{\sqrt{1-(1-2M/r)D^2/r^2}}{1-2M/r},\, -D,\, 0\right\};\\
&n_j = \left\{-\frac{1}{2}(1-2M/r),\, -\frac{e_r}{2}\sqrt{1-(1-2M/r)D^2/r^2},\right.\\
&\left.\frac{D}{2}(1-2M/r),\, 0\right\};\\
&m_j =\\
&\frac{1}{\sqrt{2}}\left\{0,\, i\frac{D}{r},\, ie_r r\sqrt{1-(1-2M/r)D^2/r^2},\, r\sin\tilde{\theta}\right\};\\
&\hat{m}_j =\\
&\frac{1}{\sqrt{2}}\left\{0,\, -i\frac{D}{r},\, -ie_r r\sqrt{1-(1-2M/r)D^2/r^2},\, r\sin\tilde{\theta}\right\}.
\end{aligned} \tag{15}$$

For the considered congruence, one can obtain the following equations (see, e.g.,[24–26], $\epsilon = k_{[i;j]} = 0$[14])

$$k_{i;j} m^i \bar{m}^j = -\rho\,, k_{i;j} m^i m^j = -\sigma\,, \qquad k_{i;j} k^j = k_{i;j} k^i = 0\,. \tag{16}$$

Here, ρ and σ are optical scalars. They are can be found numerically from well-known equations (see, e.g.,[23–26]): Only the sum $\rho+\sigma$ admits an analytical expression (see, e.g.,[25]):

$$\begin{aligned}
&\rho+\sigma = -\frac{\dfrac{\mathrm{d}}{\mathrm{d}\nu}\left(r\sin\tilde{\theta}\right)}{r\sin\tilde{\theta}} =\\
&-\frac{e_r}{r}\sqrt{1-\left(1-\frac{2M}{r}\right)\frac{D^2}{r^2}} + \frac{D}{r^2}\cot\tilde{\theta}\,.
\end{aligned} \tag{17}$$

We now write down components of the vector of 4-velocity of the star in the basis of the null tetrad (15):

$$u^j = \frac{1}{\sqrt{2}}\left(\bar{A} m^j + A\bar{m}^j\right) + Bk^j + Cn^j\,. \tag{18}$$

Here A, B, C — are coefficients of decomposition. We obtain $-k_j u^j = (1+z) = C$. Denoting the components of the Killing vector $\frac{\partial}{\partial t}$ as ξ^j, we have

$$\xi^j = \frac{1}{2}\left(1 - \frac{2M}{r}\right) k^j + n^j \,. \tag{19}$$

Furthermore

$$E = -u^i \xi_i = B + \frac{1}{2}\left(1 - \frac{2M}{r}\right)(1+z) \,. \tag{20}$$

From the relation for the norm of u_i, we obtain

$$u_j u^j = |A|^2 - 2(1+z)B = -1 \,. \tag{21}$$

From (20) and (21), it follows that

$$B = \frac{1+|A|^2}{2(1+z)} \,, \qquad |A|^2 = -1 + 2E(1+z) - (1-2M/r)(1+z)^2 \,.$$

Now, we express the time derivative of redshift, using the relation $k_{[j;l]} = 0$ satisfied for the considered congruence in Schwarzschild spacetime (see, e.g.,[14]):

$$\begin{aligned}
&\frac{\mathrm{d}z}{\mathrm{d}\tau} = -k_{j;l} u^j u^l = -|A|^2 k_{j;l} m^j \bar{m}^l - A^2 k_{j;l} m^j m^l - \\
&\bar{A}^2 k_{j;l} \bar{m}^j \bar{m}^l - 2\bar{A}(1+z) k_{j;l} n^j m^l - 2A(1+z) k_{j;l} n^j \bar{m}^l = \\
&|A|^2(\rho + \sigma\cos(2P_A)) - 2\sqrt{2}(1+z)\frac{D}{r^3}|A|\sin(P_A) + \\
&\frac{e_r}{r^2}(1+z)^2\sqrt{1 - \left(1 - \frac{2M}{r}\right)\frac{D^2}{r^2}} \,.
\end{aligned} \tag{22}$$

Here we use $A = |A| e^{iP_A}$, where $|A|$ and P_A are real. Numerical calculations show that optical scalar σ have very small value comparing to other terms in 22. Due to this will neglect the value of σ in calculations.

From equations (22) and (17) for the time derivative of redshift, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}z}{\mathrm{d}\tau} = \left[2E(1+z) - \left(1 - \frac{2M}{r}\right)(1+z)^2 - 1\right] \times \\
&\left[-\frac{e_r}{r}\sqrt{1 - \left(1 - \frac{2M}{r}\right)\frac{D^2}{r^2}} + \frac{D}{r^2}\cot\hat{\theta}\right] - \\
&2\frac{D}{r^3}\sin P_A \sqrt{2E(1+z) - \left(1 - \frac{2M}{r}\right)(1+z)^2 - 1} + \\
&\frac{e_r}{r^2}(1+z)^2\sqrt{1 - \left(1 - \frac{2M}{r}\right)\frac{D^2}{r^2}} \,.
\end{aligned} \tag{23}$$

To obtain an analytical formula, it is convenient to exclude the impact parameter D from equations (11) and (23). We obtain:

$$\frac{D}{r} = \tag{24}$$
$$\frac{\frac{L}{r}\beta\left((1+z)\left(1-\frac{2M}{r}\right)-E\right) \pm \sqrt{\left(E^2-\left(1-\frac{2M}{r}\right)\left(1+\frac{L^2}{r^2}\right)\right)\left(|A|^2+(\beta^2-1)\frac{L^2}{r^2}\right)}}{E^2-1+\frac{2}{r}+\frac{L^2}{r^2}(\beta^2-1)(1-\frac{2}{r})} =$$
$$\mathcal{F}_1(r,\, z,\, p_1,\, p_2,\, i_0,\, \tilde{\theta})\,;$$
$$\text{and } \frac{D}{r} = \mathcal{F}_2(r,\, z,\, \frac{\mathrm{d}z}{\mathrm{d}\tau},\, p_1,\, p_2,\, i_0,\, \tilde{\theta}) =$$
$$\frac{r\frac{\mathrm{d}z}{\mathrm{d}\tau}\left(|A|^2\cot\tilde{\theta}-\frac{2(1+z)}{r}|A|S\right)}{\left(1-\frac{2M}{r}\right)\left(\frac{(1+z)^2}{r}-|A|^2\right)^2+\left(|A|^2\cot\hat{\theta}-\frac{2(1+z)}{r}|A|S\right)^2} \pm$$
$$\frac{\sqrt{\left(1-\frac{2M}{r}\right)\left[\left(\frac{(1+z)^2}{r}-|A|^2\right)^2-r^2\left(\frac{\mathrm{d}z}{\mathrm{d}\tau}\right)^2\right]+\left(|A|^2\cot\tilde{\theta}-\frac{2(1+z)}{r}|A|S\right)^2}}{\left(1-\frac{2M}{r}\right)\left(\frac{(1+z)^2}{r}-|A|^2\right)^2+\left(|A|^2\cot\hat{\theta}-\frac{2(1+z)}{r}|A|S\right)^2} \times$$
$$\left(\frac{(1+z)^2}{r}-|A|^2\right). \tag{25}$$

Here, $S = \sin P_A$. To find an exact expression for $\sin P_A$, one may use the law of angular momentum conservation:

$$u_i\Psi^i = L = const\,, \tag{26}$$

where Ψ^i is the Killing vector field associated with the symmetry of the Schwarzschild metric relative to spatial rotation around an arbitrary axis (we chose it to be orthogonal to the orbit plane). Components of Ψ^i in the coordinate basis of $\hat{K}$ are given by (see, e.g.,[19])

$$\Psi^j = \left\{0,\, 0,\, (\cos i + \sin i\cot\tilde{\theta}\cos\tilde{\varphi}),\, \sin i\sin\tilde{\varphi}\right\}. \tag{27}$$

Equation of the orbital plane has the following form

$$-\sin i\sin\tilde{\theta}\cos\tilde{\varphi}+\cos i\cos\tilde{\theta}=0\,. \tag{28}$$

From equations (26), (27), (28), (18) and (15), it follows

$$e_r\frac{l}{r}\frac{1-\beta^2}{|A|\sqrt{1-\left(1-\frac{2M}{r}\right)\frac{D^2}{r^2}}}+e_P\beta\sqrt{1-\frac{l^2}{r^2}\frac{(1-\beta^2)}{|A|^2}} =$$
$$\beta\sin P_A+e_r\frac{\sqrt{1-\beta^2}}{\sqrt{1-\left(1-\frac{2M}{r}\right)\frac{D^2}{r^2}}}\cos P_A\,, \tag{29}$$

where e_P is defined as

$$e_P = \operatorname{sign}\left[e_r \beta l \sqrt{1 - \left(1 - \frac{2M}{r}\right)\frac{D^2}{r^2}} + e_s D \sqrt{E^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{L^2}{r^2}\right)}\right]. \tag{30}$$

The exact solution of (29) has the form

$$\sin P_A = e_P \sqrt{1 - \frac{L^2}{r^2}\frac{(1-\beta^2)}{|A|^2}}. \tag{31}$$

4. The inverse problem

4.1. *The surface of parameters of motion*

The main purpose of the present subsection is to obtain the relation between the parameters of motion of the star from one hand and the redshift z and the derivative $\mathrm{d}z/\mathrm{d}\tau$ for certain moments of proper time from another. From (25) and (25), we obtain:

$$\mathcal{F}_1(r,\, z(t_o),\, p_1,\, p_2,\, i_0,\, \tilde{\theta}) = \mathcal{F}_2(r,\, z(t_o),\, \frac{\mathrm{d}z}{\mathrm{d}\tau}(t_o),\, p_1,\, p_2,\, i_0,\, \tilde{\theta}). \tag{32}$$

For the known redshift data $z(t_o)$ and $\frac{\mathrm{d}z}{\mathrm{d}\tau}(t_o) = (z(t_o)+1)\frac{\mathrm{d}z}{\mathrm{d}\tau}(t_o)$ (t_o is a certain observation time), equation (32) allows one to implicitly express the constant parameters of motion p_1, p_2 and i_0 in the case, when the radial location of emission r and the angle $\tilde{\theta}$ are known. Therefore, more equations are needed to solve the problem. For this purpose, one can use equations (25) and (10). We express $\tilde{\theta}$ from (10). The impact parameter D in equation (10) can be expressed using (25). This way, we obtain:

$$\tilde{\theta} = f(r,\, \mathcal{F}_1(r,\, z(t_o),\, p_1,\, p_2,\, i_0,\, \tilde{\theta})). \tag{33}$$

Here f is some known explicit function. This equation can be solved for $\tilde{\theta}$ using the iteration method. Because the right-hand side of (33) depends on $\tilde{\theta}$ only through the small optical scalar ρ, it has little influence on the whole expression, and the solution of (33) converges quickly. For our numerical model, we have used only two iterations to obtain the solution in explicit form. Therefore, we obtain the following equation:

$$r = r_s(\varphi(\tilde{\theta}(r,\, z(t_o),\, p_1,\, p_2,\, i),\, i) - \delta,\, p_1,\, p_2). \tag{34}$$

The system (32), (34 obtained for certain observation time t_o contains 2 equations and 5 unknown variables: p_1, p_2, i_0, δ, r. Because of this, the solution is a 3-surface in a corresponding 5-dimensional space. Since we are only interested in the relations that connect the parameters of motion we must numerically calculate a projection of this space into a 4-dimensional space (with four coordinates: p_1, p_2, i_0, δ). The calculation results for our numerical model of the radiation of the star are presented (see Fig. 4–7) for different points of redshift data. To uniquely visualise the solution, we present it graphically on 2-dimensional sections of the mentioned 4-dimensional

space. As seen from Fig. 4–5, the 3-surfaces obtained for each data point do not coincide. Therefore, the intersection point of the surfaces gives an exact value of the parameters of motion of the star. This point can be determined from the obtained figures with high accuracy.

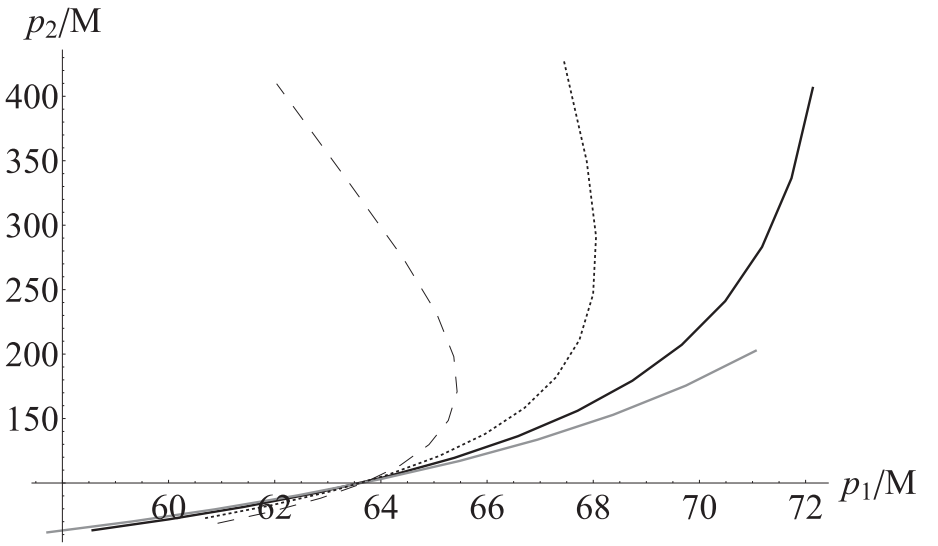


Fig. 4. 2-sections of the solution of (32), (34) for $i_0 = \pi/2$ and $\delta = 1$ by plane p_1, p_2 for different points of data (dashed: $t = 777M$, dotted: $t = 851M$, black: $t = 923M$, gray: $t = 992M$, see also Fig. 3)

Figures 4–7 also illustrate that in the case when the angular parameters are chosen to coincide with the exact solution, sections have a unique point of intersection (Fig. 6) that corresponds to the solution of the inverse problem. At the same time, if the angular parameters are not exact, a unique intersection point does not exist (Fig. 7). This is because the chosen 2-dimensional surface in the last case does not intersect with the solution in the whole 4-dimensional space of motion parameters. By using these figures, one may find if there exists a unique point of intersection in a certain region of the surface. Therefore, an initial approximation to the exact solution of the inverse problem can be found from these figures with a fairly good accuracy.

4.2. *Solution of the inverse problem*

A further improvement of the results can be obtain using the least squares method. For this purpose, we use functions $\tau_s(\varphi,\, p_1,\, p_2)$ and $r_s(\varphi,\, p_1,\, p_2)$. From (32), obtain the following equation:

$$\mathcal{F}_1(r_s(\varphi+\delta,\, E,\, L),\, z(\tau_s+\tau_0),\, p_1,\, p_2,\, i,\, \tilde{\theta}(\varphi,\, i)) = \\ \mathcal{F}_2(r_s(\varphi+\delta,\, E,\, L),\, z(\tau_s+\tau_0),\, \frac{\mathrm{d}}{\mathrm{d}\tau} z(\tau_s+\tau_0),\, p_1,\, p_2,\, i,\, \tilde{\theta}(\varphi,\, i))\,. \qquad (35)$$

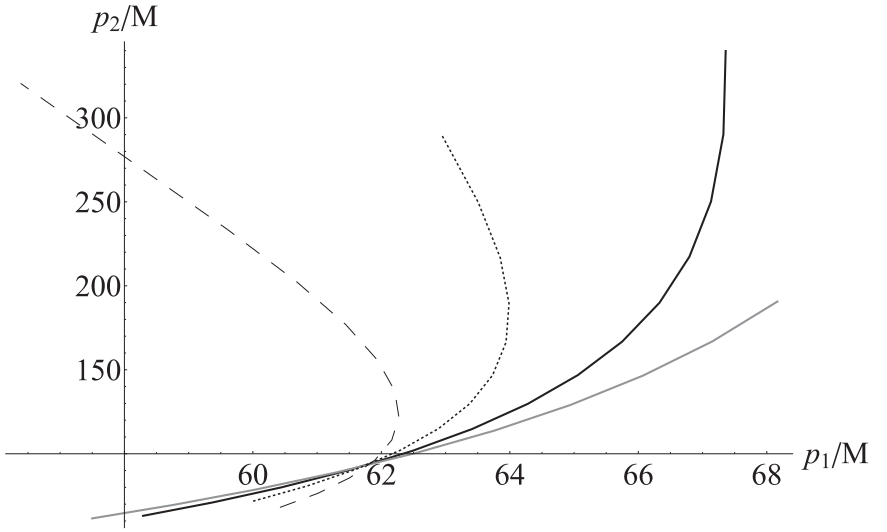


Fig. 5. 2-sections of solution of (32), (34) for $i_0 = \pi/2$ and $\delta = 1, 4$ by plane p_1, p_2 for different points of data (dashed: $t = 777M$, dotted: $t = 851M$, black: $t = 923M$, gray: $t = 992M$, see also Fig. 3)

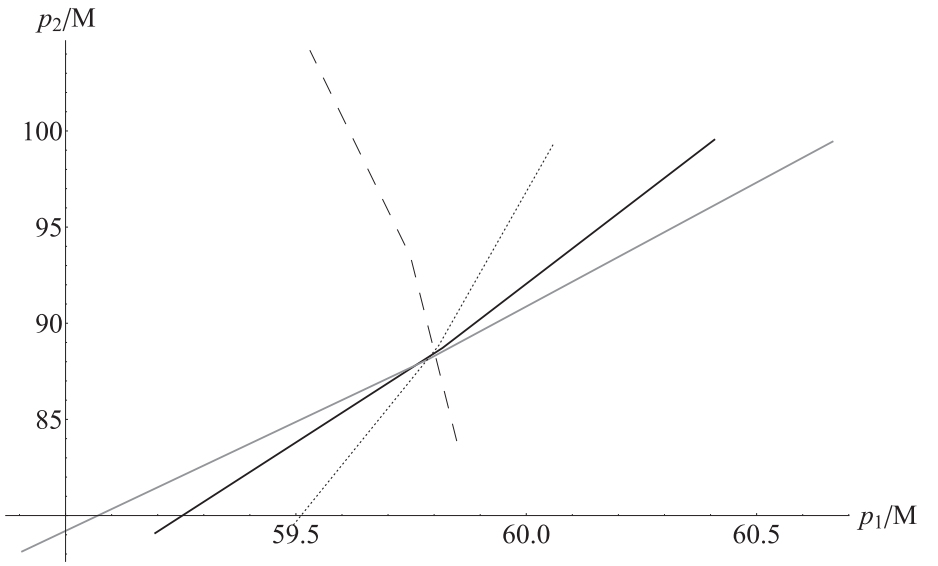


Fig. 6. Magnification of a part of figure (4)

Here, function $z(\tau)$ can be constructed based on the observational data for $z(t)$ and from function $\tau(t)$ given as an implicit function from:

$$\tau(t) = \int_0^t \frac{\mathrm{d}t'}{(1 + z(t'))}\,. \tag{36}$$

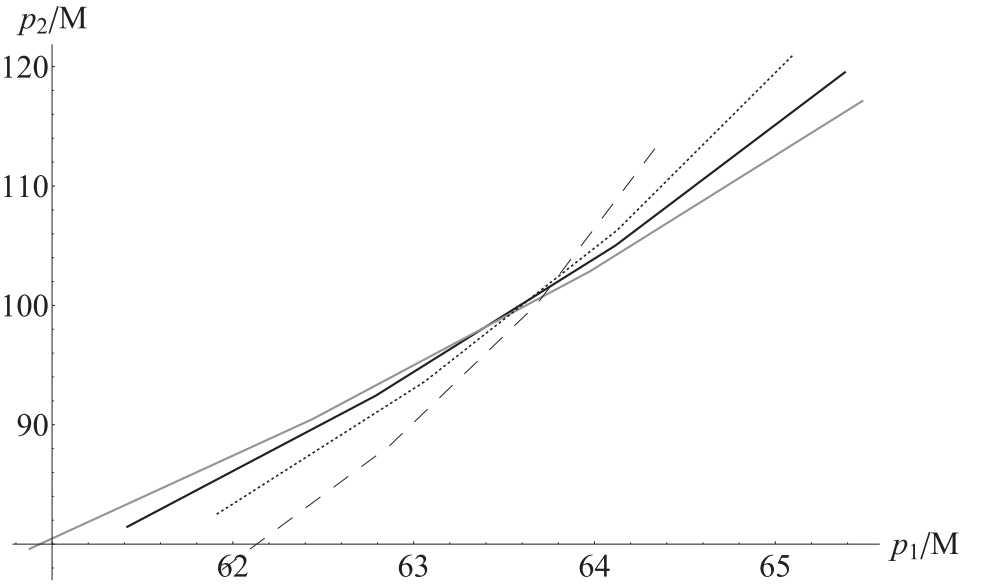


Fig. 7. Magnification of a part of figure (5)

Table 1. Results

Parameter	Initial approximation	Reconstructed value	Exact value
Pericenter distance, p_1/M	62.0	60.1	60.0
Apocenter distance, p_2/M	95.0	89.1	90.0
Orbital inclination, i_0, rad	1.5	1.48	1.4
Initial phase, δ, rad	0.9	1.0	1.0
Initial time of pericenter passage, τ_0/M	–	0.0	0.0

The left-hand side of (35) and the right-hand side of (35) are certain functions of φ. As follows from (35), these functions must be equal for a certain set of unknown parameters E, L, i_0, δ, τ_0. Therefore, one may find these parameters by using the least-squares method. As an example, we choose 10 points for different values of φ in the range [1rad, 1.4rad]. We obtain the initial approximation using the results from the previous subsection. We show the obtained results in Table 1.

5. Conclusion

The presented approach allows one to solve the inverse problem: reconstructing the motion of a star moving in the external gravitational field of a supermassive

black hole based on its redshift. The approach uses the properties of congruences of isotropic geodesics to account for the difference between trajectories of light that come to the observer from the different locations of the star during the observation time.

As the main result of the paper, we have provided a method for obtaining good starting values for the parameters of motion of a star. These starting values allow one to solve the inverse problem more accurately by using statistical methods. For this purpose, we used the graphs of the surfaces in the space of parameters of motion. However, the approach may also be formulated in terms of solving a system of equations. If one writes down equations (32) and (34) for four moments of time of observation (t_{o1}, t_{o2}, t_{o3}, t_{o4}), one may obtain a system of 8 equations for 8 unknown variables (p_1, p_2, δ, i_0 and four values of the radius of radiation: r_1, r_2, r_3, r_4). Therefore, one will obtain a complete system of equations. Even numerically, it is not easy to solve a system of 8 non-linear equations. However, in future work, developed mathematical methods will allow one to efficiently solve this system of equations and analyse the conditions for obtaining non-unique solutions. The problem of non-unique solutions can be solved, for instance, by adding more equations to the considered system. The last parameter of motion τ_0 can be determined from the least-squares method. Therefore, in principle, equations (32) and (34) can be used to obtain a unique solution of the inverse problem, rather than the graphic solution presented in our paper.

The obtained equations are exact equations in General Relativity (we only neglect the optical scalar σ). Therefore, one may use the presented approach for all possible sources moving at arbitrary distances from the black hole (if they can be approximated as test particles in an external gravitational field of the black hole).

Furthermore, the approach can be directly applied to pulsar timing data for a pulsar moving in an external gravitational field. A large number of pulsars will likely be detected in the Galactic Center in the near future (see, e.g.,[12]). Pulsars can move closer to the supermassive black hole than S-stars. This way, they may be even more interesting for testing theories of gravity. The arrival times of the radio pulses can be expressed through the redshift by using:

$$t_{TOA}^{(N)} = t_{TOA}^{(N-1)} + T_p(z+1) = t_{TOA}^{(N-1)} + T_p \frac{(k^i u_i)_s}{(k^i u_i)_o}\,.$$

Here T_p — is the pulsar period in the reference frame of the pulsar, z is the redshift, $t_{TOA}^{(j)}$ — is the time of arrival of the j-th the pulse. In the problem of reconstructing the pulsar motion in the neighbourhood of a supermassive black hole, there exists one more unknown parameter — T_p.

Another interesting application of the results of this paper is reconstructing the motion of a binary star in the vicinity of a black hole. Determining the motion of such objects is a very important problem in astrophysics and stellar mechanics (see, e.g.,[27–31]). An approach for solving the problem of determining only the relative motion of the binary components was presented in our previous paper.[32]

The approach can be applied directly to the redshift data for the stars, moving near the supermassive black hole in the Galactic Center (for example, the S62 star[2]) to test General Relativity. To do this, one may obtain the parameters of motion of a star by using the presented algorithm and calculate the redshift as a function of observation time for future observations. Then, comparing the obtained curve with the observational data will allow one to test the theory.

One may use statistical methods (for example, Metropolis-Hastings algorithm, see[22]) to increase the accuracy of reconstructing the motion of the star. Such methods allow one to calculate the likelihood probability distribution and optimise values of parameters of motion of the star according to the distribution. The approach can be generalised to reconstruct the motion of a star in the vicinity of a rotating black hole. We leave this problem for future work.

Acknowledgments

This research was supported by the Belarusian Republican Foundation for Fundamental Research. Grant №339/02 "Reconstruction of motion of a star in the gravitational field of a supermassive black hole by the redshift of the spectrum of its electromagnetic radiation."

References

1. S. Gillessen, P. Plewa, F. Eisenhauer et. al., An update on monitoring stellar orbits in the galactic center, *Astrophys. J.* **837**, 1 (2017).
2. F. Peibker, A. Eckart and M. Parsa, S62 on a 9.9 yr orbit around Sgr A*, *Astrophys. J.* **889**, 1 (2020).
3. R. Genzel, F. Eisenhauer and S. Gillessen, The Galactic Center massive black hole and nuclear star cluster, *Rev. Mod. Phys.* **82**, 3121 (2010).
4. M. Parsa, A. Eckart, B. Shahzamanian et. al., Investigating the relativistic motion of the stars near the supermassive black hole in the Galactic Center, *Astrophys. J.* **845**, 1 (2017).
5. S. Jia, J. R. Lu, S. Sakai, The Galactic Center: Improved relative astrometry for velocities, accelerations, and orbits near the supermassive black hole, *Astrophys. J.* **873**, 1 (2019).
6. GRAVITY Collaboration: R. Abuter, A. Amorim, N. Anugu et. al., Detection of the gravitational redshift in the orbit of the star S2 near the Galactic Centre massive black hole, *Astron. and Astrophys.* **615**, 1 (2018).
7. GRAVITY Collaboration: R. Abuter, A. Amorim, M. Bauböck et. al., Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic Centre massive black hole, *Astron. and Astrophys.* **636**, 1 (2020).
8. L. Iorio, Post-keplerian effects on radial velocity in binary systems and the possibility of measuring general relativity with the star S2 in 2018, *MNRAS* **472**, 2249 (2017).
9. M. Grould, F. H. Vincent, T. Paumard and G. Perrin, General relativistic effects on the orbit of the S2 star with GRAVITY, *Astron. and Astrophys.* **608**, A60, 1 (2017).
10. F. Zhang, Y. Lu and Q. Yu, On testing the Kerr metric of the massive black hole in the Galactic Center via stellar orbital motion: full general relativistic treatment, *Astrophys. J.* **809**, p. 27 (2015).

11. R. Angelil et al., Toward relativistic orbit fitting of Galactic Center stars and pulsars, *Astrophys. J.* **720**, 1303 (2010).
12. F. Zhang and P. Saha, Probing the spinning of the massive black hole in the Galactic Center via pulsar timing: A full relativistic treatment, *Astrophys. J.* **849**, p. 15 (2017).
13. A. Gorbatsievich, S. Komarov and A. Tarasenko, Optical appearance of a compact binary system in the neighbourhood of supermassive black hole, *arXiv:1702.08381* (2017).
14. S. Komarov, A. Gorbatsievich, A. Tarasenko, Redshift of a compact binary star in the neighborhood of a supermassive black hole, *Gen. Relativ. and Gravit.* **50**, p. 132 (2018).
15. A. N. Tarasenko, S. O. Komarov, A. K. Gorbatsievich, Redshift of radiation of a point-like source moving in the external Kerr field, *Doklady Natsional'noi akademii nauk Belarusi=Doklady of the National Academy of Sciences of Belarus* **62**, 668 (2018).
16. C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* (W. H. Freeman and company, San Francisco, 1973).
17. S. Chandrasechar, *The mathematical theory of black holes* (Oxford University press, New York, 1983).
18. G. A. Korn and T. M. Korn, *Mathematical handbook for scientists and engineers; definitions, theorems and formulas; for reference and review* (McGraw-Hill Book company, New York, San Francisco, Toronto, London, Sidney, 1968).
19. H. Stephani, *Relativity. An introduction to Special and General Relativity* (Cambridge University Press, 2004), Third English edition.
20. A. Tarasenko, Reconstruction of a compact object motion in the vicinity of a black hole by its electromagnetic radiation, *Phys. rev. D* **81**, p. 123005 (2010).
21. G. S. Bisnovatyi-Kogan and Y. Tsupko, Strong gravitational lensing by Schwarzschild black holes, *Astrophysics* **51**, 99 (2008).
22. W.K. Hastings, Monte Carlo sampling methods using markov chains and their applications, *Biometrica* **57**, p. 97 (1970).
23. R. Penrose and W. Rindler, *Spinors and space-time. V. 1 Two-spinor calculus and relativistic fields* (Cambridge University Press, 1987), in 2 parts.
24. E. Newman and R. Penrose, An approach to gravitational radiation by a method of spin coefficients, *J. of Math. Phys.* **3**, 566 (1961).
25. S. Pineault and R. C. Roeder, Applications of geometrical optics to the Kerr metric. i. analytical results, *Astrophys. J.* **212**, 541 (1977).
26. I. Novikov and V. Frolov, *Physics of Black Holes* (Springer Netherlands, 1989).
27. S. Naoz, The eccentric Kozai-Lidov effect and its applications, *Annu. Rev. Astron. Astrophys.* **54**, 441 (2016).
28. J. H. Van Landingham, M. C. Miller, D. P. Hamilton, and D. C. Richardson, The role of the Kozai-Lidov mechanism in black hole binary mergers in galactic centers, *Astrophys. J.* **828**, 1 (2016).
29. C. Petrovich and F. Antonini, Greatly enhanced merger rates of compact-object binaries in non-spherical nuclear star clusters, *Astrophys. J.* **846**, 1 (2017).
30. Y. Fang and Q.-G. Huangh, Secular evolution of compact binaries revolving around a spinning massive black hole, *Phys. rev. D* **99**, p. 103005 (2019).
31. F. Antonini and H. B. Perets, Secular evolution of compact binaries near massive black holes: gravitational wave sources and other exotica, *Astrophys. J.* **757**, 1 (2012).
32. S. Komarov, A. Gorbatsievich, Reconstruction of relative motion of a binary star in the vicinity of black hole by its redshift, *International Journal of Modern Physics A* **35**, p. 2040052 (2020).

Shadows of hairy Kerr black holes and constraints from M87*

Sushant G. Ghosh[1,2,*] and Misba Afrin[1,†]

[1] *Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi 110025, India*
[2] *Astrophysics and Cosmology Research Unit, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag 54001, Durban 4000, South Africa*
[*] *E-mail: sghosh2@jmi.ac.in* [†] *E-mail: me.misba@gmail.com*

In the wake of the *Event Horizon Telescope* (*EHT*) observations of the supermassive black hole M87*, efforts are underway to distinguish the black holes in general relativity (GR) and modified theories of gravity (MoG). We study the rotating hairy Kerr black holes with a deviation α and primary hair l_0, apart from rotation parameter a and mass M. Interestingly, the hairy Kerr black holes possess smaller sizes but more distorted shadows than the Kerr black holes. We find that, within 1σ uncertainty of the *EHT* observations, the inferred circularity deviation $\Delta C \leq 0.1$ for the M87* black hole is satisfied, whereas the shadow angular diameter $\theta_d = 42 \pm 3\mu as$, for a given choice of α, places bounds on the parameters a and l_0. Thusfore, the hairy Kerr black holes are inferred to be suitable candidates for astrophysical black holes.

Keywords: Black hole physics, gravitation, strong gravitational lensing, black hole shadow.

1. Introduction

The term "black hole shadow" has come to represent the interior of an "apparent boundary" as termed by Bardeen,[1] the inside of which is dark when the black hole is illuminated by distant, uniform, isotropic emission surrounding it. Indeed,the photons moving in unstable orbits around a black hole form a shadow silhouette that appears as a sharp boundary between bright and dark regions. The seminal work by Synge in 1966[2] gave the critical escape cone of photons from the surface of strong field stars. After that, Cunningham and Bardeen in 1973,[3] using geometrical optics, calculated the apparent position and energy flux of high-frequency radiation emitted by stars in a circular orbit around extremal Kerr black hole and also presented the first image of visible shapes for inner and outer circumferences of accretion disks (cf. Fig. 8[3]). Luminet in 1979[4] investigated the optical appearance of a thin accretion disk around spherically symmetric black holes and obtained the relative intrinsic intensity as well as the gravitational and Doppler spectral shifts of the relativistic photons. The shape and size of a shadow depend entirely upon the black hole parameters and spacetime geometry.[5–8, 8–24] However, it is expected that its shape and size will get influenced by the gravity governing a black hole and the matter around it.[5–10, 12–18, 20, 21, 25]

The first shadow image of the supermassive black hole M87* has been unveiled by the international *Event Horizon Telescope* (*EHT*) collaboration.[26–28] Using the

very long base interferometry (VLBI) technique, the shadow of the M87* – an asymmetric bright emission ring with a diameter of 42 ± 3 μas has been found to harbour an angular gravitational radius 3.8 ± 0.4 μas while also exhibiting a deviation from circularity $\Delta C \leq 0.1$. It is consistent with the shadow of a Kerr black hole as predicted by general relativity (GR). With the detection of the M87* black hole's horizon-scale image[26–28] the black hole shadows have become a physical reality, thus allowing probes into the background theory of gravity as well. The quantitative features are not sufficient to distinguish between black holes in different theories of gravity but using the M87* black hole shadow; one can investigate the viability of black holes in modified theories of gravity (MoG) in explaining the observational data while placing constraints on the black hole parameters using the *EHT* observations.[29]

The shadows cast by black holes in MoG have been found to be smaller and more distorted in comparison with the Kerr black hole shadow.[5,7,9–18] One of the motivations behind the multifarious works on black hole shadow is that by observing the size and deformation of shadow, the spin, mass parameter and possibly other global charges or *hair*-parameters of the black holes can be calculated.[19–22,29–31] The authors in,[23,24] considering the critical impact parameter as an analytical expression of charge, investigated the connection with shadows. In addition, in the context of the Reissner–Nordstrom metric, a tidal charge has been constrained[30] from estimates of shadow done by Doeleman et al.[32] The shadow size in Reissner–Nordstrom spacetime is found to decrease with increasing charge (cf. Figs. 1,2 in,[23] see also a discussion in[25]). Recently, Kocherlakota et al.[33] presented constraints on black hole charge from observations of M87* in 2017. Moreover, the shadows also has a vibrant use in testing theories of gravity.[29,34,35] While the the M87* has been presently inferred to be a black hole via the *EHT* observations,[26–28] there is a possibility that other exotic gravitationally intense objects cast a shadow. The hairy Kerr black holes under consideration are obtained using the gravitational decoupling (GD) approach.[36] It has a source that satisfies the strong energy condition (SEC) thus giving an extended Kerr metric which was termed as Kerr black holes with primary hair by Contreras *et. al.*[36] Hairy black holes are termed as stationary black hole solutions with new global charges, which are not associated with Gauss law.[8,37–39] (See, Ref.[37] for a recent review).

We investigate whether black hole shadow observations can constrain hairy Kerr black holes,[36] which has additional deviation parameter α and primary hair l_0 from the Kerr black hole. Furthermore, the horizons, ergoregions and shadow cast by the hairy Kerr black holes[36] are studied. An investigation of the deviation parameter's effects shows that the hairy Kerr black holes cast smaller and more distorted shadows than the Kerr black hole. Furthermore, within 1σ confidence of the EHT observational data, the parameter space of the hairy Kerr black holes is constrained with the bounds on the shadow observables of M87*.

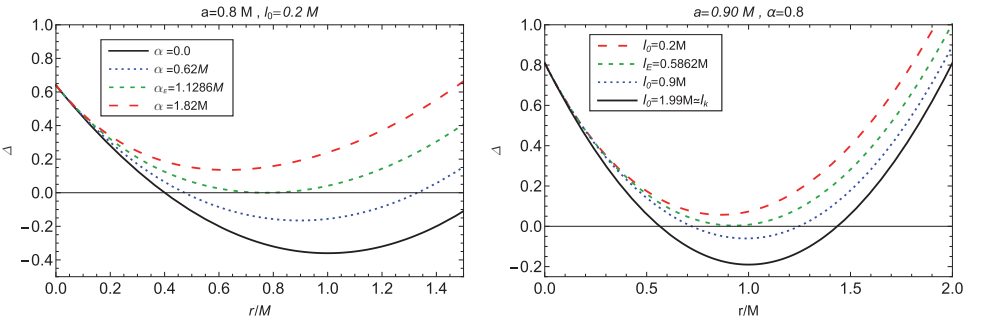


Fig. 1. Horizons of the hairy Kerr black holes are shown with varying α (*left*) and l_0 (*right*) and compared with the Kerr black holes ($\alpha = 0$ or $l \to l_k$).

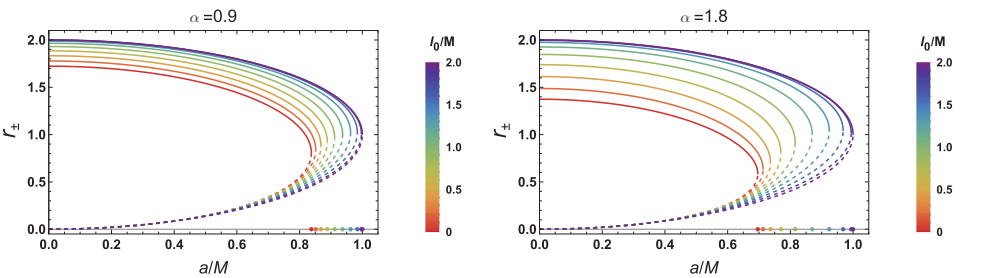


Fig. 2. Event horizon (solid lines) and Cauchy horizon (dashed lines) of hairy Kerr black holes with varying spin a for different values of l_0 and α parameters. The outermost lines are for the Kerr black hole in the limit $l_0 \to l_k$. The extremal spin a_E values are denoted by the colour coded points.

2. Hairy Kerr back holes

The GD approach, which is designed to obtain deformations of the known solutions of GR owing to the additional surrounding sources like dark matter or dark energy,[36,40,41] leads to deformed or hairy Schwarzchild black holes.[36] It has a dimensionless deviation parameter α and primary hair $l_0 \leq 2M$ which also determines the asymptotic flatness. The surrounding matter is described by the conserved stress-energy tensor $S_{\mu\nu}$ that satisfies the SEC. Hence, we consider the rotating counterpart of the spherically symmetric solution described by parameters M, a, l_0 and α, which in the Boyer-Lindquist coordinates reads[36]

$$\begin{aligned} ds^2 = &-\left[\frac{\Delta - a^2\sin^2\theta}{\Sigma}\right]dt^2 - 2a\sin^2\theta\left[1 - \frac{\Delta - a^2\sin^2\theta}{\Sigma}\right]dtd\phi \\ &+\sin^2\theta\left[\Sigma + a^2\sin^2\theta\left(2 - \frac{\Delta - a^2\sin^2\theta}{\Sigma}\right)\right]d\phi^2 + \frac{\Sigma}{\Delta}dr^2 \\ &+\Sigma d\theta^2, \end{aligned} \tag{1}$$

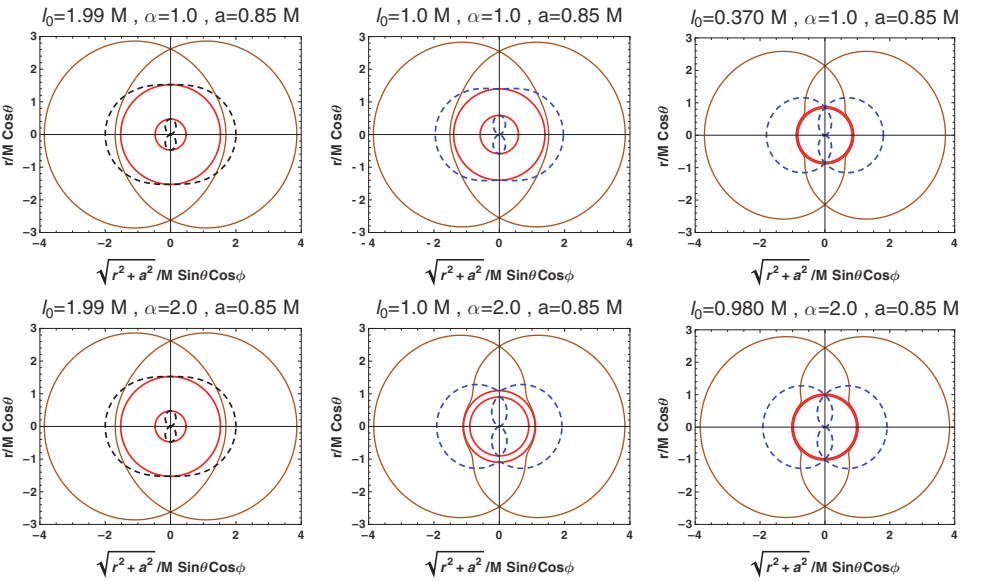


Fig. 3. The cross-section of event horizon (outer red line), SLS (outer blue dotted line) and photon region (bounded within the outer and inner orange lines) of hairy Kerr black holes. The black dotted curves correspond to the event horizon of Kerr black hole ($l_0 \to l_k$).

$$\text{with } \Delta = r^2 + a^2 - 2Mr + \alpha r^2 e^{-r/(M-\frac{l_0}{2})} \text{ and } \Sigma = r^2 + a^2\cos^2\theta.$$

Here α is related to l_0 via $l_0 = \alpha l$ and the metric (1) goes to Kerr in the absence of surrounding fluid ($\alpha = 0$). The metric (1) which is termed as hairy Kerr black holes,[36] is same as Kerr black hole with the Kerr mass M replaced by the mass function,

$$m(r) = M - \frac{\alpha}{2} r e^{-r/(M-\frac{l_0}{2})}.$$

Next, we shall investigate how the primary hair l_0 affects the horizons, photon region as well as shadow of the black hole. The horizons of the metric (1) correspond to spacetime points where $\Sigma \neq 0$ and $g^{\alpha\beta}\partial_\alpha r \partial_\beta r = g^{rr} = \Delta = 0$ and is given by the zeros of

$$r^2 + a^2 - 2Mr + \alpha r^2 e^{-r/(M-\frac{l_0}{2})} = 0. \tag{2}$$

The two roots of Eq. (2) are r_+ outer (event) horizon and r_- the inner (Cauchy) horizon. An analysis of Eq. (2) reveals that, depending on the values of M, a, α and l_0, there can exist a maximum of two distinct real positive roots, or equal roots corresponding respectively to the black holes with Cauchy and event horizons or extremal black holes.

From the horizon structure of the hairy Kerr black holes depicted in Fig. 1, we find that, $l_0 > l_E$ corresponds to hairy Kerr black holes with Cauchy and event horizons whereas $l_0 < l_E$ implies existence of a naked singularity.

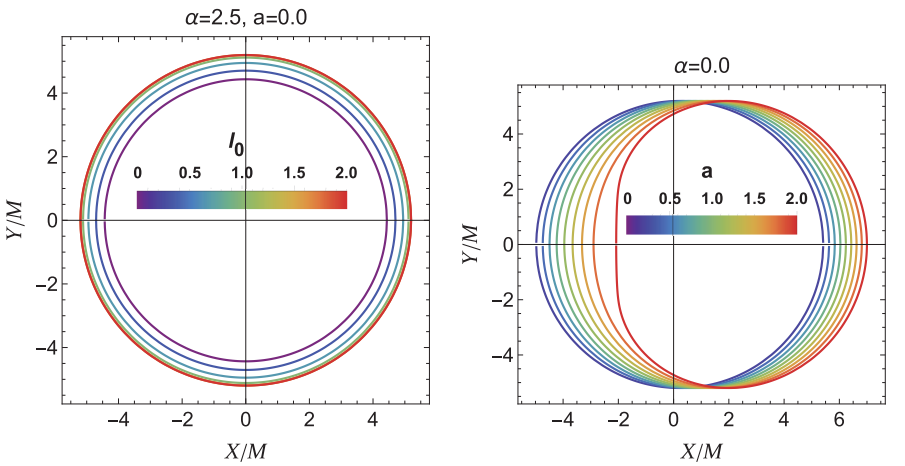


Fig. 4. Shadow geometry of non-rotating hairy black holes with varying l_0 parameter (left) and Kerr black hole ($\alpha = 0$) with varying spin a (right).

Furthermore, the asymptotic time-translational Killing vector, becomes null at the static limit surface (SLS) which yields[9]

$$r^2 + a^2\cos^2\theta - 2Mr + \alpha r^2 e^{-r/(M-\frac{l_0}{2})} = 0. \tag{3}$$

Thus, the real positive solutions ($r^{\pm}_{SLS}$) of Eq. (3) correspond to the radii of the SLS, the larger of the which corresponds to the outer SLS, denoted by r^{+}_{SLS}. Fig. 3 depicts the ergoregion of the hairy Kerr black hole (1) whereby it can be seen that, the increase in parameter l_0 has a diminishing effect on the ergoregion's size and may thus have consequence in the Penrose energy extraction process.[42]

3. Analytic black hole shadow

We use the Hamilton-Jacobi equation to obtain the geodesic equations in the first-order differential form,[43] which for the metric (1) read[44]

$$\Sigma\frac{dt}{d\lambda} = \frac{r^2+a^2}{\Delta}(E(r^2+a^2) - aL_z) - a(aE\sin^2\theta - L_z), \tag{4}$$

$$\Sigma\frac{dr}{d\lambda} = \pm\sqrt{\mathcal{R}(r)}\ , \tag{5}$$

$$\Sigma\frac{d\theta}{d\lambda} = \pm\sqrt{\Theta(\theta)}\ , \tag{6}$$

$$\Sigma\frac{d\phi}{d\lambda} = \frac{a}{\Delta}(E(r^2+a^2) - aL_z) - (aE - \frac{L_z}{\sin^2\theta}), \tag{7}$$

where λ is the affine parameter along the geodesics and the effective potentials $\mathcal{R}(r)$

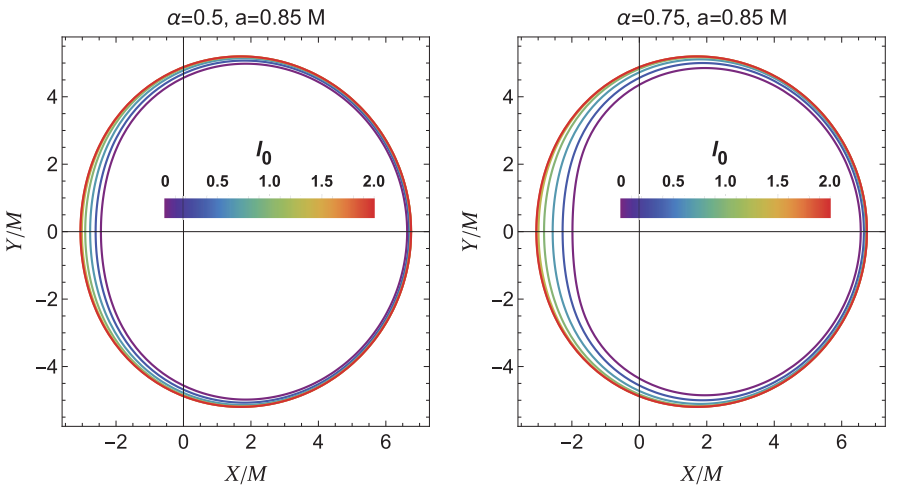


Fig. 5. Shadows of hairy Kerr black holes compared with Kerr black hole shadow (outermost solid curve corresponding to $l_0 \to l_k$).

and $\Theta(\theta)$ for radial and polar motions are given by

$$\mathcal{R}(r) = \Big((r^2+a^2)E - aL_z\Big)^2 - \Delta\Big(\mathcal{K} + (aE - L_z)^2\Big), \tag{8}$$

$$\Theta(\theta) = \mathcal{K} - \left(\frac{{L_z}^2}{\sin^2\theta} - a^2E^2\right)\cos^2\theta\ . \tag{9}$$

The constant $\mathcal{K}$ is the separability constant related to the Carter constant $\mathcal{Q}$ through $\mathcal{Q} = \mathcal{K} + (aE - L_z)^2$.[44] By introducing reduced quantities called impact parameters[44] $\xi = L_z/E$, $\eta = \mathcal{K}/E^2$, the degrees of freedom of the photon motion reduces from three to two. We are interested in the spherical photon orbits[44] whose radius $r = r_p$ are formed at the local extremum of the radial effective potential outside the horizon such that[44,45]

$$\mathcal{R} = \mathcal{R}' = 0 \text{ and } \mathcal{R}'' \leq 0. \tag{10}$$

Solving Eq. (10) we obtain the critical values of impact parameters $(\xi_{crit}, \eta_{crit})$ for the unstable orbits given by,

$$\xi_{crit} = \frac{\left(a^2+r^2\right)\Delta'(r) - 4r\Delta(r)}{a\Delta'(r)},$$

$$\eta_{crit} = \frac{r^2\left(8\Delta(r)\left(2a^2 + r\Delta'(r)\right) - r^2\Delta'(r)^2 - 16\Delta(r)^2\right)}{a^2\Delta'(r)^2}, \tag{11}$$

where $'$ stands for the derivative with respect to the radial coordinate. The Eq. (11) in the limit $\alpha \to 0$ reduces to the critical impact parameter for the Kerr black holes.[19]

The Carter constant is directly related to θ-velocity of the photons and the non-planar or 3D photon orbits only arise for $\eta_{crit} > 0$.[46] The roots of $\eta_{crit} = 0$ give

the radii of prograde (r_p^-) and retrograde (r_p^+) orbits. The exterior photon region, $r_+ < r_p^- < r_p < r_p^+$, of the hairy Kerr black holes is depicted in Fig. 3 as the region enclosed between the inner prograde and the outer retrograde orbits (cf. Fig. 3).

The gravitationally lensed image[47–50] of the photon region appears as a dark region outlined by a bright ring to a distant observer is termed the black hole shadow.[51,52] The shadow has been utilized for the measurement of various black hole parameters like its mass, spin angular momentum as well as other *hairs*,[22,29,35] whereby the black hole shadow is evidently a means to test Einstein's GR in the strong-field regime[13] besides the no hair theorem.[29,53] Assuming a distant observer ($r_o \to \infty$) at an inclination angle θ_0 with the rotation axis of the black hole a stereographic projection of the black hole shadow to the celestial image plane gives the shadow outline in the celestial coordinates (X,Y) as,[19]

$$\begin{aligned} X &= -\,\xi_{crit}\csc\theta_o, \\ Y &= \pm\sqrt{\eta_{crit} + a^2\cos^2\theta_o - \xi_{crit}^2\cot^2\theta_o}\ , \end{aligned} \tag{12}$$

which further gets simplified to $X = -\xi_{crit}$, $Y = \pm\sqrt{\eta_{crit}}$ for an equatorial observer ($\theta_o = \pi/2$). The parametric contour of ($X(r_p)$,$Y(r_p)$) traces the shadow of the hairy Kerr black holes with $m(r) = M - \frac{\alpha}{2} r e^{-r/(M-\frac{l_0}{2})}$.

Interestingly, Zakharov et al.[54] carried out an analysis of critical curve $\eta(\xi)$ which separates scatter and capture of a photon in Kerr metric (cf. see Fig. 2[54] and discussion therein) and considering critical values corresponding to multiple roots of the polynomial describing a radial photon motion as functions of ξ (see also the critical curve in Fig. 34[44]). Subsequently, it was shown[31] that for an equatorial observer for any spin parameter $0 \leq a \leq 1$, the maximum absolute value of Y corresponds to $X = -2a$ and hence, in principle, it is possible to estimate a by measuring the position of the maximum value for Y.[31] Moreover, Bardeen[3] and Chadrasekhar[44] considered the apparent shape of a black hole for spin $a = M$, but they do not suggest using the apparent shape of a black hole as GR test perhaps because the dark region (shadow) is too small to be detectable for all known estimates of black hole masses and distances.

We depict the shadows for the non-rotating hairy black hole ($a = 0$) and Kerr black holes ($\alpha = 0$) (1) in Fig. 4 (cf. Fig. 2[31]) . Further, a comparison of the hairy Kerr black hole shadows with that of the Kerr black hole is shown in Fig. 5. There is a decremental effect in the shadow size from that of the Kerr black hole (cf. Fig. 5) with the decrease in l_0 from the Kerr limit $l_0 = l_k$.

4. Constraints on hairy Kerr black hole parameters from *EHT* observation of M87*

The EHT image of M87* is nearly circular and of crescent shape encompassing a central depression in brightness with a flux ratio $\geq 10 : 1$;[26–28] in terms of root-mean-square deviation from the average shadow radius the circularity deviation

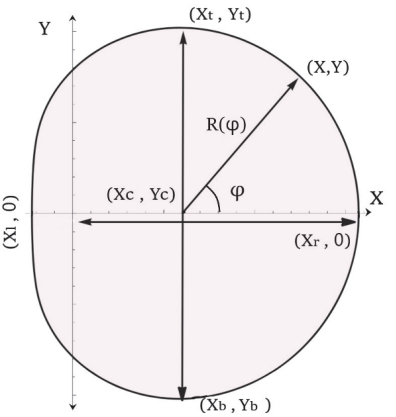


Fig. 6. The coordinates X_l and X_r indicate the most negative and most positive x-coordinates of the shadow while the Y_t, Y_b represent the top most and bottom most points. (X_c, Y_c) is the centre of the shadow and $R(\phi)$ is the polar shadow radius.[55]

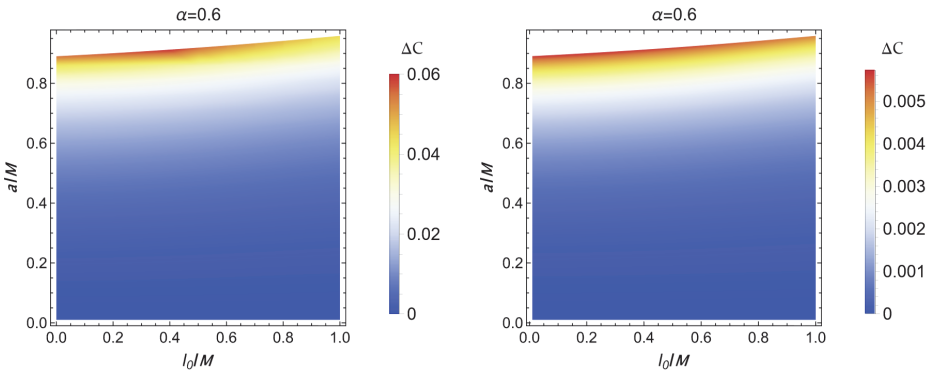


Fig. 7. Circularity deviation observable ΔC for hairy Kerr black hole shadows as a function of parameters (a/M and l_0/M) in agreement with the *EHT* observations of the M87* black hole, $\Delta C \leq 0.1$, is satisfied for the entire parameter space (a/M and l_0/M). The parameters of M87* used are $M = 6.5 \times 10^9 M_\odot$ and $d = 16.8$MPc. The inclination angle is $\theta_0 = 90^\circ$ (*left*) and 17° (*right*). The white region is forbidden for (a/M and l_0/M).

$\Delta C \leq 0.10$ (10%) and the angular size θ_d of the bright emission region in the observed image is $42 \pm 3\mu as$.[26–28]

The shadow outline of an axially symmetric black hole is depicted in Fig. 6 with the boundary of the black hole shadow by the polar coordinates $(R(\varphi), \varphi)$, with origin at the shadow centre (X_c, Y_c) such that $X_c = (X_r - X_l)/2$ and $Y_c = 0$ (cf. Fig. 6). The circularity deviation ΔC, which quantifies deviation from a perfect circle, is defined as[56]

$$\Delta C = \frac{1}{\bar{R}} \sqrt{\frac{1}{2\pi} \int_0^{2\pi} \left(R(\varphi) - \bar{R}\right)^2 d\varphi}, \tag{13}$$

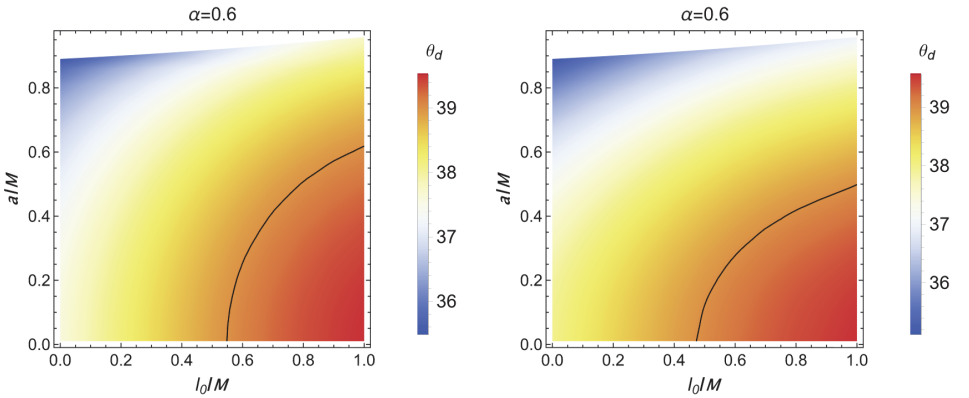


Fig. 8. Angular diameter observable θ_d for hairy Kerr black hole shadows as a function of parameters (a/M and l_0/M). The black solid curve correspond to $\theta_d = 39\mu$as within 1σ region of the measured angular diameter, $\theta_d = 42 \pm 3\mu$as, of the M87* black hole reported by the *EHT*. The parameters of M87* used are $M = 6.5 \times 10^9 M_\odot$ and $d = 16.8$Mpc. The inclination angle is $\theta_0 = 90°$ (*left*) and $17°$ (*right*). The white region is forbidden for (a/M and l_0/M).

with $\bar{R}$ – the average shadow radius.[56] The angular diameter of the shadow θ_d[13] is defined as

$$\theta_d = 2\frac{R_a}{d} \quad , \quad R_a = \sqrt{A/\pi}, \tag{14}$$

Here, φ is the polar angle between the x-axis and the vector connecting the centre (X_c, Y_c) with any point (X, Y) lying on the boundary of the shadow (cf. Fig. 6). where R_a is the areal radius of the shadow. The θ_d together with the ΔC will be useful to perform a comparison between the theoretical predictions for hairy Kerr BH shadows and the *EHT* observations.[56]

Assuming M87* a hairy Kerr black hole, we can calculate ΔC for metric (1) and use the *EHT* observational result $\Delta C \leq 0.1$ to put constraints on the hairy Kerr black hole parameters. The circularity deviation ΔC in the parameter space $(a/M - l_0/M)$ of the hairy Kerr black hole is depicted in Fig. 7. The Fig. 7 clearly indicates that the hairy Kerr black hole shadows of M87* for a given α, satisfy $\Delta C \leq 0.10$ for entire parameter space $(a/M - l_0/M)$ at both the inclination angles. The angular diameter of the shadow is calculated next, which apart from a, l_0, α, θ, depends on mass M and the distance d of the black hole and for the hairy Kerr black holes is depicted in Fig. 8 as function of a and l_0 α taking inclination angle $\theta_0 = 90°, 17°$. The black curves denote the *EHT* bound of 39μas, the 1σ bound of M87* on θ_d, and the region enclosed by it $39\mu as \leq \theta_d \leq 45\mu as$ serves as the parameter space $(a/M - l_0/M)$ consistent with the M87* observations.

5. Conclusions

A thorough analysis reveals that the l_0 and α quantitatively influences the event horizon structure, the ergosphere area, and the spherical photon region around

hairy Kerr black holes. Further, both the α and l_0 parameters are found to alter the shadow shape and size when compared to the Kerr black holes, namely, the shadows are smaller and more distorted, besides, the shadow size decreases, whereas the distortion increases with the decreasing value of parameter l_0.

The circularity deviation $\Delta C \leq 0.1$ is satisfied exhaustively for the entire $(l_0 - a)$ parameter space at both $\theta_0 = 90^\circ$ and $\theta_0 = 17^\circ$ inclination angles. The angular diameter satisfies $\theta_d = 42 \pm 3\mu$as within the 1σ region over a limited $(l_0 - a)$ space. However, the accordant parameter space for $\theta_0 = 17^\circ$ is smaller and more constricted than that at $\theta_0 = 90^\circ$. Thus θ_d puts a tighter bound on the conforming parameter space of the hairy Kerr black hole.

Therefore, our results constrain the hairy Kerr black holes parameters to ensure that the M87* shadow observations do not entirely rule out hairy Kerr black holes whereby the Kerr black hole has been studied in a more general setting, thus shedding light on the effect of surrounding dark matter on the Kerr black holes. The current resolution of the M87* black hole shadow angular diameter is of $\mathcal{O}(\mu$as) makes the distinguishability of the hairy black holes from the Kerr black hole bleak, at least from the present resolution of the *EHT* and the next-generation EHT with higher resolution may pin down the exact constraint. Thus to summarize, with the first direct image of the supermassive black hole M87* *EHT* observation, we have been able to place constraints on the hairy Kerr black hole parameters.

Acknowledgements

M.A. is supported by DST-INSPIRE Fellowship, Department of Science and Technology, Govt. of India

References

1. J. M. Bardeen, Timelike and null geodesics in the Kerr metric, *Les Houches summer school of theoretical physics: Black holes, Gordon and Breach science publishers Inc., United States* (1973).
2. J. L. Synge, The escape of photons from gravitationally intense stars, *Mon. Not. Roy. Astron. Soc.* **131**, 463 (1966).
3. C. T. Cunningham and J. M. Bardeen, The optical appearance of a star orbiting an extreme Kerr black hole, *ApJ* **183**, 237 (1973).
4. J. P. Luminet, Image of a spherical black hole with thin accretion disk, *Astron. Astrophys.* **75**, 228 (1979).
5. L. Amarilla, E. F. Eiroa and G. Giribet, Null geodesics and shadow of a rotating black hole in extended Chern-Simons modified gravity, *Phys. Rev. D* **81**, 124045 (2010).
6. L. Amarilla and E. F. Eiroa, Shadow of a rotating braneworld black hole, *Phys. Rev. D* **85**, 064019 (2012).
7. L. Amarilla and E. F. Eiroa, Shadow of a Kaluza-Klein rotating dilaton black hole, *Phys. Rev. D* **87**, 044057 (2013).
8. C. Herdeiro, E. Radu and H. Rúnarsson, Kerr black holes with Proca hair, *Class. Quantum Gravity* **33**, 154001 (2016).
9. M. Amir, B. P. Singh and S. G. Ghosh, Shadows of rotating five-dimensional charged EMCS black holes, *Eur. Phys. J. C* **78**, 399 (2018).

10. B. P. Singh and S. G. Ghosh, Shadow of Schwarzschild–Tangherlini black holes, *Ann. Phys.* **395**, 127 (2018).
11. Y. Mizuno, Z. Younsi, C. M. Fromm, O. Porth, M. De Laurentis, H. Olivares, H. Falcke, M. Kramer and L. Rezzolla, The current ability to test theories of gravity with black hole shadows, *Nature Astron.* **2**, 585 (2018).
12. S. G. Ghosh, M. Amir and S. D. Maharaj, Ergosphere and shadow of a rotating regular black hole, *Nucl. Phys. B* **957**, 115088 (2020).
13. R. Kumar and S. G. Ghosh, Rotating black holes in $4D$ Einstein-Gauss-Bonnet gravity and its shadow, *JCAP* **07**, 053 (2020).
14. M. Amir and S. G. Ghosh, Shapes of rotating nonsingular black hole shadows, *Phys. Rev. D* **94**, 024054 (2016).
15. R. Kumar, S. G. Ghosh and A. Wang, Gravitational deflection of light and shadow cast by rotating Kalb-Ramond black holes, *Phys. Rev. D* **101**, 104001 (2020).
16. M. Amir and S. G. Ghosh, Rotating Hayward's regular black hole as particle accelerator, *JHEP* **07**, 015 (2015).
17. R. Kumar, B. P. Singh and S. G. Ghosh, Shadow and deflection angle of rotating black hole in asymptotically safe gravity, *Annals Phys.* **420**, 168252 (2020).
18. R. Kumar, B. P. Singh, M. S. Ali and S. G. Ghosh, Shadows of black hole surrounded by anisotropic fluid in Rastall theory, *Phys. Dark Univ.* **34**, 100881 (2021).
19. K. Hioki and K.-I. Maeda, Measurement of the Kerr spin parameter by observation of a compact object's shadow, *Phys. Rev. D* **80**, 024042 (2009).
20. P. V. P. Cunha, C. A. R. Herdeiro and E. Radu, Spontaneously scalarized Kerr black holes in extended scalar-tensor–Gauss-Bonnet gravity, *Phys. Rev. Lett.* **123**, 011101 (2019).
21. P. V. P. Cunha, C. A. R. Herdeiro and E. Radu, EHT constraint on the ultralight scalar hair of the M87 supermassive black hole, *Universe* **5**, 220 (2019).
22. R. Kumar and S. G. Ghosh, Black hole parameter estimation from its shadow, *ApJ* **892**, 78 (2020).
23. A. F. Zakharov, F. de Paolis, G. Ingrosso and A. A. Nucita, Direct measurements of black hole charge with future astrometrical missions, *A&A* **442**, 795 (2005).
24. A. F. Zakharov, Constraints on a charge in the Reissner-Nordström metric for the black hole at the Galactic Center, *Phys. Rev. D* **90**, 062007 (2014).
25. A. F. Zakharov, Constraints on a tidal charge of the supermassive black hole in M87* with the EHT observations in April 2017, *arXiv:2108.01533* (2021).
26. K. Akiyama *et al.*, First M87 Event Horizon Telescope results. I. the shadow of the supermassive black hole, *ApJL* **875**, L1 (2019).
27. K. Akiyama *et al.*, First M87 Event Horizon Telescope results. IV. imaging the central supermassive black hole, *ApJL* **875**, L4 (2019).
28. K. Akiyama *et al.*, First M87 Event Horizon Telescope results. VI. the shadow and mass of the central black hole, *ApJL* **875**, L6 (2019).
29. M. Afrin and S. G. Ghosh, Constraining rotating black holes in Horndeski theory with EHT observations of M87*, *arXiv:2110.05258* (2021).
30. A. F. Zakharov, F. de Paolis, G. Ingrosso and A. A. Nucita, Shadows as a tool to evaluate black hole parameters and a dimension of spacetime, *New Astron. Rev.* **56**, 64 (2012).
31. A. F. Zakharov, A. A. Nucita, F. De Paolis and G. Ingrosso, Measuring the black hole parameters in the Galactic Center with RADIOASTRON, *New Astron.* **10**, 479 (2005).
32. S. S. Doeleman et al., Event-horizon-scale structure in the supermassive black hole candidate at the Galactic Centre, *Nature* **455**, 78 (2008).

33. P. Kocherlakota et al., Constraints on black-hole charges with the 2017 EHT observations of M87*, *Phys. Rev. D* **103**, 104047 (2021).
34. M. Kramer, D. C. Backer, J. M. Cordes, T. J. W. Lazio, B. W. Stappers and S. Johnston, Strong-field tests of gravity using pulsars and black holes, *New Astron. Rev.* **48**, 993 (2004).
35. R. Kumar, A. Kumar and S. G. Ghosh, Testing rotating regular metrics as candidates for astrophysical black holes, *Astrophys. J.* **896**, 89 (2020).
36. E. Contreras, J. Ovalle and R. Casadio, Gravitational decoupling for axially symmetric systems and rotating black holes, *Phys. Rev. D* **103**, 044020 (2021).
37. C. A. R. Herdeiro and E. Radu, Asymptotically flat black holes with scalar hair: a review, *Int. J. Mod. Phys. D* **24**, 1542014 (2015).
38. C. A. R. Herdeiro and E. Radu, Kerr black holes with scalar hair, *Phys. Rev. Lett.* **112**, p. 221101 (2014).
39. Y.-X. Gao and Y. Xie, Gravitational lensing by hairy black holes in Einstein-scalar-Gauss-Bonnet theories, *Phys. Rev. D* **103**, 043008 (2021).
40. J. Ovalle, Decoupling gravitational sources in general relativity: from perfect to anisotropic fluids, *Phys. Rev. D* **95**, 104019 (2017).
41. J. Ovalle, Decoupling gravitational sources in general relativity: the extended case, *Phys. Lett. B* **788**, 213 (2019).
42. R. Penrose and R. M. Floyd, Extraction of rotational energy from a black hole, *Nature* **229**, 177 (1971).
43. B. Carter, Global structure of the Kerr family of gravitational fields, *Phys. Rev.* **174**, 1559 (1968).
44. S. Chandrasekhar, *The mathematical theory of black holes* (Oxford Univ. Press, New York, 1985).
45. V. P. Frolov and A. Zelnikov, *Introduction to black hole physics* (Oxford Univ. Press, Oxford, 2011).
46. E. Teo, Spherical orbits around a Kerr black hole, *Gen. Relativ. Gravit.* **53**, 10 (2021).
47. V. Bozza, Gravitational lensing in the strong field limit, *Phys. Rev. D* **66**, 103001 (2002).
48. S. G. Ghosh, R. Kumar and S. U. Islam, Parameters estimation and strong gravitational lensing of nonsingular Kerr-Sen black holes, *JCAP* **03**, 056 (2021).
49. R. Kumar, S. U. Islam and S. G. Ghosh, Gravitational lensing by charged black hole in regularized 4*D* Einstein–Gauss–Bonnet gravity, *Eur. Phys. J. C* **80**, 1128 (2020).
50. S. U. Islam, R. Kumar and S. G. Ghosh, Gravitational lensing by black holes in the 4*D* Einstein-Gauss-Bonnet gravity, *JCAP* **09**, 030 (2020).
51. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon regions and shadows of Kerr-Newman-NUT black holes with a cosmological constant, *Phys. Rev. D* **89**, 124004 (2014).
52. P. V. P. Cunha, C. A. R. Herdeiro and M. J. Rodriguez, Does the black hole shadow probe the event horizon geometry?, *Phys. Rev. D* **97**, 084020 (2018).
53. P. V. P. Cunha, C. A. R. Herdeiro, E. Radu and H. F. Runarsson, Shadows of Kerr black holes with scalar hair, *Phys. Rev. Lett.* **115**, 211102 (2015).
54. A. F. Zakharov, Types of unbounded orbits in the Kerr metric, *Zh. Eksp. Teor. Fiz.* **91**, 3 (1986).
55. J. C. S. Neves, Upper bound on the GUP parameter using the black hole shadow, *Eur. Phys. J. C* **80**, 343 (2020).
56. C. Bambi, K. Freese, S. Vagnozzi and L. Visinelli, Testing the rotational nature of the supermassive object M87* from the circularity and size of its first image, *Phys. Rev. D* **100**, 044057 (2019).

Displacement memory and BMS symmetries

Shailesh Kumar

Indian Institute of Information Technology Allahabad,
Devghat, Jhalwa, Uttar Pradesh-211015, India.
E-mail: shaileshkumar.1770@gmail.com

This article reviews one of the most intriguing properties of black hole spacetimes known in the literature- *gravitational memory effect*, and its connection with asymptotic symmetries, also termed as Bondi-van der Burg-Metzner-Sachs (BMS) symmetries, emerging near the horizon of black holes. Gravitational memory is a non-oscillatory part of the gravitational wave amplitude which generates a permanent displacement for freely falling test particles or test detectors. We highlight a model scenario where asymptotic symmetries appear as a soldering freedom in the context of stitching of two black hole spacetimes, and examine the impact of the interaction between test detectors and horizon shells. Further, we provide a more realistic approach of computing displacement memory for near-horizon asymptotic symmetries which is analogous to the conventional memory originally obtained at asymptotic null infinity.

Keywords: Memory, Gravitational waves, BMS symmetries (asymptotic symmetries)

1. Introduction

The observational facets of gravitational waves (GWs)[1,2] have opened a new window to look for various aspects of black hole spacetimes; *gravitational memory*[3–8] is one of such intriguing features that has not been detected yet. GW induces a permanent relative change in the position of test detectors by imparting a memory to the configuration. This permanent relative change is referred to as gravitational memory. The term *memory* implies the information or properties of spacetimes from where it is being generated and carried by gravitational waves. The first practical computation of memory's evolution was done by Marc Favata using post-Newtonian formalism where he accounted for all stages of BBH coalescence.[9–12] In this direction, recently, there have been several implications of detecting GW memory using advanced detectors.[13–20]

On the other hand, it has been shown that the gravitational memory is closely related to the asymptotic symmetries of spacetimes originally discovered by Bondi-van der Burg-Metzner-Sachs (BMS) in the early sixties,[21] and such symmetries can also be recovered near the horizon of black holes which motivated us to probe the near horizon properties of black holes. The recent findings in this direction have provided some strong grounds for the information loss puzzle. In the context of asymptotic symmetries, the existence of soft hair on black holes is necessary for charge conservation of supertranslation and superrotation.[7,22–24] As the conservation principles are derived from the long-distance behaviour of fields close to spatial infinity, the presence of black holes should have no effect on them. We know that

the conserved charges can be expressed as bulk integrals over any Cauchy surface. A contribution from the future event horizon should be taken into account for conserved charges as future null infinity is no longer a Cauchy surface in the presence of a classical black hole. In this direction, Strominger and Hawking's latest discovery uses the asymptotic symmetries of the BMS group to prove that information is not lost rather stored in something called as a soft particle. Therefore, soft hair or low-energy quantum excitations may be carried by a black hole and leak information when it evaporates. This brings a direct motivation for the emergence of asymptotic symmetries near the horizon of black holes from conservation perspectives.

Let us understand how memory and BMS symmetries are inter-connected with each other. Classically, for a given spacetime geometry, BMS transformations produce an infinite class of spacetime metrics that are physically unique or distinct. Assume that BMS transformations act on a given metric $g_{\mu\nu}(x^\mu)$ with $x^\mu = (x^0, x^i)$, i.e., one time and three spatial coordinates. Such an action on the metric generates a completely different metric.

$$g_{\mu\nu}(x^\mu) \xrightarrow{\text{BMS transformation}} \tilde{g}_{\mu\nu}(x^\mu).$$

The metric $g_{\mu\nu}(x^\mu)$ and $\tilde{g}_{\mu\nu}(x^\mu)$ are distinct and this relative change implies the generation of *GW memory*, and also motivates us to seek for a connection between memory and BMS symmetries. This change can be understood in the following way- GWs generated from a black hole spacetime carrying information or properties in terms of BMS parameters would interact with the detector setup placed at the asymptotic null infinity, this would induce a permanent relative change in the initial configuration of the setup. A similar setup can also be considered at a place near the horizon of a black hole. A persistent effect similar to that of null-infinity may again be observed. It provides a physical meaning to the inter-connection between GW memory and asymptotic symmetries emerging near the horizon of black holes. Technically, $g_{\mu\nu}(x^\mu)$ can be thought of as a metric of a given asymptotically flat spacetime and $\tilde{g}_{\mu\nu}(x^\mu)$ is the resultant metric appears as a consequence of the interaction between GWs and detectors which implies a net relative change in the configuration and gives a definition to the memory. Briefly, if we have two nearby timelike geodesics or inertial detectors described by the tangent vector T^μ together with a deviation vector s^μ, and let us position them at the future null infinity. The evolution of the deviation vector before and after the interaction with gravitational waves will be captured in the geodesic deviation equation (GDE), written as

$$\frac{D^2 s^\mu}{d\tau^2} = -R^\mu{}_{\delta\sigma\lambda} T^\delta T^\lambda s^\sigma. \tag{1}$$

The solution of the GDE will give us a permanent relative change in the displacement vector s^μ which can further be related to supertranslation and will implicate the achievement of *BMS displacement memory effect*. Our study provides an analogous effect and its connection with asymptotic symmetries for near the horizon of black holes.

There are two methods to recover asymptotic symmetries near the horizon of a black hole. As a recent progress, Donnay et al. showed the first way of obtaining such symmetries near the horizon of a stationary black hole[25] with asymptotic form of the Killing vectors preserving the boundary conditions. It turns out that the near-horizon region of a stationary black hole spacetime induces supertranslations including semi-direct sum with extended asymptotic symmetry superrotations which is being represented by Virasoro algebra. Hence, one can recover asymptotic symmetries that would mimic the ones originally obtained at asymptotic null infinity[25,26] by preserving the near-horizon asymptotic structure of black holes. The second method for recovering asymptotic symmetries deals with the soldering of two spacetimes across a common null hypersurface.[27,28] It has been shown that we can stitch them in infinite ways by demanding that the induced metric remains invariant under the translations generated by the null generators of the shell.[27,28] The freedom for the choice of the intrinsic coordinates on null hypersurface in the null-direction is termed as *soldering freedom*, and also known as *BMS-like soldering freedom*. Since these appear as a metric preserving transformations, hence, known as BMS-like symmetries or BMS-like transformations. We shall discuss the related details in section 2.

The article is organized as follows. In section 2, we discuss the intrinsic formulation of null shells placed at the horizon and how near-horizon asymptotic symmetries are recovered in the context of stitching of two black hole spacetimes. Further, in section 3.1, we show how horizon shells carrying memory affect the displacement between two nearby test detectors or test particles for Schwarzschild and Extreme Reissner Nordström (ERN) black holes. We have also studied the impact of interaction between null geodesics and horizon shells; since we shall be completely focusing on timelike geodesics in this article, we do not include the discussion on null geodesics crossing the horizon shells. However, the study can be found in.[31,32] We further consider a more realistic approach in section 3.2 for determining the displacement memory effect and its connection with near-horizon asymptotic symmetries. In the end, we conclude our findings in section 4 by providing some remarks on possible future outlooks to our studies which might be relevant from theoretical as well as observational perspectives.

2. Horizon shell and asymptotic symmetries

In general relativity, a shell is a geometric configuration that can be used to investigate the propagation of thin distribution of null matter (e.g. neutrino) and impulsive gravitational waves (IGWs). The thin surface layer of null matter together with impulsive waves is precisely referred to as *thin-shell* or *thin null shell*.[29,30] The generated impulsive signals are usually produced during violent astrophysical phenomena like supernova explosions or coalescence of black holes. If we stitch two black hole spacetimes along a common null hypersurface which also happens to be the horizon of black holes, and stitching is consistent with *junction conditions*, we

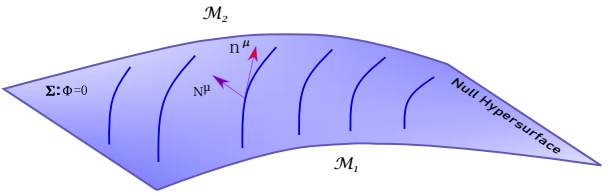


Fig. 1. Null hypersurface Σ separating manifolds $\mathcal{M}_{1,-}$ and $\mathcal{M}_{2,+}$ each with a different metric.

obtain *horizon shell*. The soldering formalism shows that the stess-energy tensor of the stitched spacetime satisfying Einstein field equation carries a singular term proportional to the Dirac delta distribution function, given by

$$T_{\mu\nu} = T^{+}_{\mu\nu}\mathcal{H}(\Phi) + T^{-}_{\mu\nu}\mathcal{H}(-\Phi) + S_{\mu\nu}\delta(\Phi), \tag{2}$$

where $\mathcal{H}(\Phi)$ is a Heaviside step function for a given null surface $\Sigma \equiv \Phi = 0$. The last term of Eq. (2) corresponds to the stress-energy tensor of the null surface, and further investigation on the same shows the generation of *impulsive gravitational wave* or thin surface layer of null matter or a mixture of both. The null hypersurface, representing the history of impulsive lightlike signals, separates spacetime manifold into two parts $(\mathcal{M}_1, \mathcal{M}_2)$ or $(\mathcal{M}_-, \mathcal{M}_+)$ as depicted in the Fig. 1 each with a different metric. n^μ is a null-normal or a generator of the null hypersurface, and N^μ is a transverse or auxiliary normal which is not tangent to the null surface, satisfying $n \cdot N = -1$. N^μ carries the transversal properties of Σ. One can study the intrinsic quantities of a horizon shell known as surface energy density (μ), surface current (J^A) and surface pressure (p) in terms of *transverse curvature* by analyzing the stress tensor on the null surface. The intrinsic quantities can be written in the following form

$$\mu = -\frac{1}{8\pi}\sigma^{AB}[\mathcal{K}_{AB}] \;\; ; \;\; j^A = \frac{1}{8\pi}\sigma^{AB}[\mathcal{K}_{VB}] \;\; ; \;\; p = -\frac{1}{8\pi}[\mathcal{K}_{VV}], \tag{3}$$

where '[ ]' denotes the difference between a quantity computed on the null surface Σ for both sides $\mathcal{M}_{1,2}$ separately. We notice that the null shell quantities depend on the induced metric σ^{AB} and jump in the extrinsic curvature $\mathcal{K}_{AB}$[a]. The other related details can be found in.[27,30–32]

An induced tensor field γ_{ab} on Σ is related to the jump of the induced metric g_{ab} and transverse normal N^μ which helps us to examine the intrinsic formulation of the horizon shell. The tensor field γ_{ab} can be written in terms of the jump in the derivative of the induced metric along the auxiliary normal N^μ which can further be written in terms of transverse curvature, i.e., $\gamma_{ab} = N^\mu[\partial_\mu g_{ab}] = 2[\mathcal{K}_{ab}]$. The analysis for intrinsic expression of stress tensor ensures that there is a part of γ_{ab}

[a]We have taken Kruskal-like coordinates. Also, Capital Latin letters denote spatial or spherical coordinates of the 2-sphere metric, and lower Latin letters denote hypersurface coordinates.

which does not contribute to stress-energy tensor. We denote this non-contributing part as $\hat{\gamma}_{ab}$. In general, one can write down γ_{ab} containing both null matter ($\bar{\gamma}_{ab}$) and GW degree of freedom ($\hat{\gamma}_{ab}$) in the following way

$$\gamma_{ab} = \hat{\gamma}_{ab} + \bar{\gamma}_{ab}, \tag{4}$$

with

$$\hat{\gamma}_{ab} = \gamma_{ab} - \frac{1}{2} g_*^{cd} \gamma_{cd} g_{ab} - 2n^d \gamma_{d(a} N_{b)} + \gamma^\dagger N_a N_b \tag{5}$$

$$\bar{\gamma}_{ab} = 16\pi \Big(g_{ac} S^{cd} N_d N_b + g_{bc} S^{cd} N_d N_a - \frac{1}{2} g_{cd} S^{cd} N_a N_b - \frac{2}{2} g_{ab} S^{cd} N_c N_d \Big), \tag{6}$$

where $\gamma^\dagger = \gamma_{cd} n^c n^d$, and g_*^{ab} is the pseudo-inverse of g_{ab}, i.e., $g_*^{ab} g_{bc} = \delta_c^a - n^a N_c, g_*^{cd} \gamma_{cd} = g^{AB} \gamma_{AB}$. The $\hat{\gamma}_{ab}$ carries the pure impulsive gravitational wave degree of freedom whereas $\bar{\gamma}_{ab}$ contains the null matter part of the horizon shell. In general, a null shell is being considered the combination of both IGWs and null matter. Further, using the expressions of intrinsic quantities of the null surface together with $\hat{\gamma}_{ab}$ and $\bar{\gamma}_{ab}$, in Kruskal coordinates, one obtains,

$$\bar{\gamma}_{VB} = 16\pi g_{BC} S^{VC} \;\; ; \;\; \bar{\gamma}_{AB} = -8\pi S^{VV} g_{AB}. \tag{7}$$

Now, we shall examine the interaction of such impulsive lightlike signals carrying BMS parameters on timelike geodesics. Let us first investigate the appearance of BMS symmetries in gluing formalism.

2.1. *Emergence of asymptotic symmetries*

Here, we investigate the emergence of near-horizon asymptotic symmetries in the context of soldering of two black hole spacetimes. This, in gluing formalism, can be achieved via obtaining the freedom in the choice of intrinsic coordinates along the null direction. The soldering freedom of stitching the two spacetimes along a common null surface provides BMS-like transformation on the horizon shell. It emerges as a coordinate transformation which preserves the induced metric on the null hypersurface Σ. This implies us to figuring out the Killing vectors of the hypersurface metric in a suitable coordinate system.[27,28] Therefore, the analysis is based on the Lie derivative of the induced metric along the Killing direction (say $Z^a \partial_a$ with components Z^a). We consider Kruskal coordinates (U, V, x^A) with coordinates (V, x^A) on Σ. Further, we also consider the metric with $g_{aV} = 0$. Therefore, the Killing equation for spatial metric g_{AB} is

$$\mathcal{L}_Z g_{AB} = 0 \implies Z^V \partial_V g_{AB} + Z^C \partial_C g_{AB} + (\partial_A Z^C) g_{CB} + (\partial_B Z^C) g_{AC} = 0. \tag{8}$$

Now, we may separately examine the emergence of near-horizon BMS symmetries-supertranslation and superrotation.

2.1.1. *Supertranslation*

The first special case is when metric does not depend on V parameter, i.e., $\partial_V g_{AB} = 0$. This induces a new type of translation which has angle dependent notion, termed as *Supertranslation*. It is similar to the one obtained at asymptotic null infinity for asymptotically flat spacetimes. Keeping in mind the impact of the Z-generated transformations on the null normal n^a of Σ, as a result, Eq. (8) gives

$$\partial_V Z^V = 0 \implies V \longrightarrow V + T(x^A) \ : \ BMS \ Supertranslation \tag{9}$$

This is instantly indentified as a supertranslation in the literature with $T(x^A)$ being a supertranslation parameter, where $x^A = (\theta, \phi)$. It is interpreted as an angle dependent translation, hence named *supertranslation*. The soldering group that keeps this structure preserved is still infinite dimensional. Let us now turn our discussion to investigate the extended form of asymptotic symmetry.

2.1.2. *Superrotation*

A new type of symmetry labeled as *superrotation* has just been discovered in an extended form of BMS symmetries near the horizon of black holes which mimics the one obtained at asymptotic null infinity.[25,27,28,33,34] It is a local conformal transformation of the spatial slice of the metric, or local conformal transformation of celestial sphere at null infinity.[22,33] Let us determine the extended BMS symmetry by considering the case when the spatial slice of the metric depends on the V parameter, i.e., $\partial_V g_{AB} \neq 0$. The analysis begins with the Eq. (8) in search of possible non-trivial soldering freedoms. If one performs the conformal transformation in spatial coordinates represented in complex coordinates via $z \longrightarrow f(z)$ and $\bar{z} \longrightarrow \bar{f}(\bar{z})$ such that the Eq. (8) can be written in the following way

$$Z^V \partial_V g_{AB} + \Omega(x^A) g_{AB} = 0, \tag{10}$$

where $x^A = (z, \bar{z})$ and $\Omega(x^A)$ denotes the conformal factor. We have the equation whose feasible solution can be written as, $g_{AB} = r^2(U, V)\tilde{g}_{AB}(x^A)$. It gives a suitable choice along V direction, i.e., $Z^V (= -\frac{V\Omega(x^A)}{2})$ which compensates the conformal transformation. This ensures that the metric remains preserved under such transformations. Thus, the analysis gives rise metric preserving extended BMS transformations known as *superrotation*-like symmetries.

3. Displacement memory and asymptotic symmetries

As a first approach, we wish to examine the relative change in the displacement vector between two nearby timelike geodesics which arises due to interaction with impulsive lightlike signals. We also discuss the appearance of near-horizon asymptotic symmetries in the context of soldering of two ERN and Schwarzschild spacetimes. Second, we would be considering a more realistic approach of computing displacement memory which is analogous to the one obtained at asymptotic null infinity. Let us first start with the case where horizon shell interacts with test detectors.

3.1. *Memory & BMS symmetries due to impulsive lightlike signals*

Cataclysmic processes such as black hole mergers and supernovae explosions produce shockwave type of gravitational radiations. We wish to estimate the finite difference in the displacement vector between two nearby timelike test particles or geodesics upon crossing the horizon shell[b]. It turns out that the asymptotic symmetries, associated with impulsive lightlike signals, leave footprints on test particles upon passing through them. As a result, we studied the effects of horizon shell for Schwarzschild and extreme RN spacetimes on the separation vector of two nearby timelike geodesics. The displacement between the geodesics is identified by supertranslation parameter which gives us *Supertranslation memory effect*. Let us consider a congruence having T^μ to be a tangent vector with $T\cdot T=-1$. A displacement or separation vector between two test particles is X^μ satisfying $T\cdot X=0$. Thus one can compute the relative change in the separation vector before and the passage of impulsive lightlike signals by analyzing the GDE Eq. (1). Riemann tensor is the memory generating factor for the given configuration. The solution of the GDE will generate a non-vanishing finite change in the deviation vector upon interacting with IGWs. Following the basic framework of the analyses from,[29,31,32] we use the expressions written in section 2 and $X^a=\tilde{g}^{ab}X_b$, we obtain components of deviation vectors as

$$X_V = 8\pi U g_{BC} S^{VC} X^B_{(0)} \tag{11}$$

$$X_A = X_{A(0)} + \frac{U}{2}\gamma_{AB}X^B_{(0)} + UV^-_{(0)A}, \tag{12}$$

where $V^-_{(0)a}=\frac{dX^-_a}{dU}\Big|_{U=0}$, and $\tilde{g}_{ab}=g_{ab}+(T_{(0)\mu}e^\mu_a)(T_{(0)\nu}e^\nu_b)$ with e_a defined as a triad on the null surface. The $X^B_{(0)}$ is some function evaluated on the null surface, and it is denoted by subscript (0). It is to note that when S^{VC} is nonzero, then we have $X_V\neq 0$. This implies that the particle will be displaced off from the initial two dimensional surface. On the other hand, if $S^{VC}=0$, the component X_V vanishes which means that the particle will reside on the initial two dimensional surface but with a relative displacement. In this particular consideration, the nonzero displacement vector X_A is written as

$$X_A = (1-4\pi U S^{VV})\Big(g_{AB}+\frac{U}{2}\hat{\gamma}_{AB}\Big)X^B_{(0)}. \tag{13}$$

The factor γ_{AB} is the one which carries the BMS memory part of the wave, and generates the distortion effect on the test particles. This sets our first goal to investigate the memory signal arises in the context of soldering of two black hole spacetimes. Further, we show our studies for extreme RN and Schwarzschild black holes. Let us understand these two cases separately.

[b]The study (B-memory) of interaction between null congruence and horizon shell can be found in Bhattacharjee et al.[31,32]

3.1.1. *Extreme RN case (ERN)*

As we know that 70% astrophysical black holes are near extremal and many supermassive black holes are also near extremal.[35–38] Further, Strominger and Vafa[39] determined the Bekenstein-Hawking area-entropy relation for extreme black hole. So extreme black holes are important from experimental as well as theoretical perspectives. Here, we investigate the asymptotic symmetries together with intrinsic properties of the shell and its interaction with test detectors in terms of BMS parameters. It is known that Carter investigated the maximal analytic extension of RN black hole for $e^2 = M^2$.[40] As the Carter's metric is not C^1 i.e. the first derivative of the metric component is discontinuous. Secondly, he certainly did a conceptual analysis without providing the exact Kruskal analogue for the extreme case. For our purpose, it is important to have an exact form of the Kruskal metric which enables us to write $U = 0$ on the horizon. This helps us to perform off-shell extension of the soldering transformation without any obscurities. Therefore, we adopt a Kruskal extension that unambiguously places the shell at $U = 0$, and also better suited for memory effect. The ERN metric in Kruskal coordinates can be written as[41]

$$ds^2 = -\frac{2M}{r^2}\psi(V)' dU dV + r^2(U)(d\theta^2 + \sin^2\theta d\phi^2), \tag{14}$$

where, $\psi(V)'$ is a regular, defined as, $\psi(V) = 4M\Big(lnV - \frac{M}{2V}\Big)$. Also $U = -(r - M)$ where $r = M$ is the horizon.

Immediately, by looking at the spherical part of the metric, we observe that supertranslation-like symmetries can be recovered, written as: $V \longrightarrow V + T(\theta, \phi)$. We also find that coordinate r is independent of V, thus interestingly, ERN consideration does not induce the superrotation-like symmetries whereas it is not the case with Schwarzschild discussed in section 3.1.2. For explicit details, we refer to Bhattacharjee et al.[32]

Further, we examine some measurable effects on timelike test particles due to interaction with horizon shell of ERN spacetime. In this process, we first extend the soldering transformation off the horizon shell to the linear order in U. The transformations are give by[31]

$$U_+ = UC(V, x^A) \;\; ; \;\; V_+ = F(V, x^A) + UA(V, x^A) \;\; ; \;\; x^A_+ = x^A + UB^A(V, x^A), \tag{15}$$

where $x^A \equiv (\theta, \phi)$, with

$$C = \frac{\partial_V \psi(V)}{\partial_V \psi(F)} \;\; ; \;\; A = \frac{M^2}{2}\frac{F_V}{\partial_V \psi(V)}\sigma_{AB}B^A B^B \;\; ; \;\; B^A = \partial_V \psi(V)\frac{1}{M^2 F_V}\sigma^{AB}F_B. \tag{16}$$

Here, σ_{AB} denotes the unit 2-sphere metric. One side of the spacetime $\mathcal{M}_-$ is completely ERN and the another side of the spacetime $\mathcal{M}_+$ is off-shell extended with the transformations (15). The process of obtaining the intrinsic quantities is

known as *off-shell extension of the soldering transformations*. It is to note here that one can also obtain the intrinsic properties of the horizon shell using *extrinsic curvature algorithm*. Both the results would exactly match. The benefit of the later approach is that it makes the computational algebra significantly simplified. As a result, we find that the surface current (j^A) together with surface energy density (μ) and pressure (p) is nonvanishing, and can be expressed in terms of supertranslation parameter $T(\theta, \phi)$. For example, the surface current is given by

$$j^A = \frac{1}{8M^2\pi}\sigma^{AB}\Big(T_B\frac{\psi(T)''}{\psi(T)'}\Big). \tag{17}$$

The expressions for μ and p can also be found in.[32] The presence of nonzero surface current induces a finite change in the X_V component of the deviation vector, i.e., $X_V \neq 0$; therefore, test particles get displaced off the initial 2-dimensional surface with a relative change in the displacement vector. The X_θ component is given by

$$X_\theta = X_{\theta(0)} + \frac{U}{2}\Big(\gamma_{\theta\theta}X^\theta_{(0)} + \gamma_{\theta\phi}X^\phi_{(0)}\Big) + UV^-_{(0)\theta}, \tag{18}$$

where,

$$\gamma_{\theta\theta} = 2\psi(V)'\Big(T_{\theta\theta} + \frac{T_\theta^2\psi(T)''}{\psi(T)'} - \frac{M}{\psi(T)'} + \frac{M}{\psi(V)'}\Big) \tag{19}$$

$$\gamma_{\theta\phi} = \gamma_{\phi\theta} = 2\psi(V)'\Big(\frac{T_\theta T_\phi}{\psi(T)'}\psi(T)'' + T_{\theta\phi} - T_\phi\cot\theta\Big). \tag{20}$$

This helps in determining GW degree of freedom γ_{ab}. One can determine X_ϕ component in a similar way. We notice that the deviation is written in terms of supertranslation parameter $T(\theta, \phi)$. The integration with respect to the geodesic parameters would give rise the displacement memory which would mimic the one obtained at null infinity. The Fig. 2 depicts the ultimate result of the test particles getting displaced off from the initial spatial slice with a comparison on Schwarzschild discussed in the section below. This completes our analyses of examining the role of near-horizon asymptotic symmetries on test particles upon interacting with impulsive lightlike signals together with the intrinsic properties of ERN horizon shell.

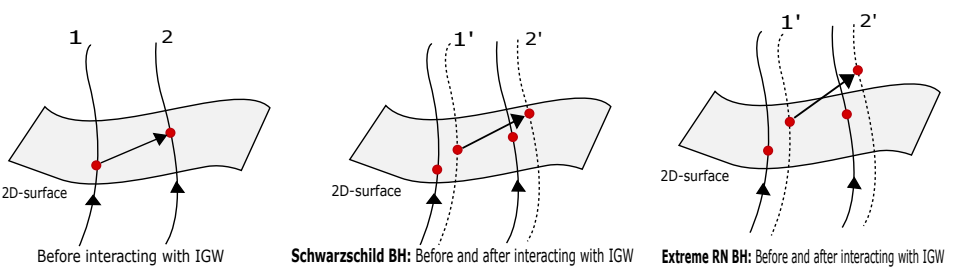


Fig. 2. Timelike geodesics 1 & 2 get displaced upon interacting with IGW, depicted as 1' & 2' with a new relative displacement vector.

3.1.2. *Schwarzschild case*

Now, we start with the Schwarzschild spacetime in order to examine the intrinsic formulation of the horizon shell and study the interaction with test particles.[31] Let us write down the metric in Kruskal coordinates,

$$ds^2 = -2G(r)dUdV + r^2(U,V)(d\theta^2 + \sin^2\theta d\phi^2), \tag{21}$$

where, $G(r) = \frac{16M^3}{r}$ and $UV - \Big(\frac{r}{2M} - 1\Big)e^{r/2M}$. The horizon is defined as $U = 0$. One can clearly see the supertranslation-like transformation, written as $V \longrightarrow V + T(\theta,\phi)$. Further, we notice that spherical part of the metric can be written as $\gamma_{\zeta\bar{\zeta}} = r^2\frac{d\zeta d\bar{\zeta}}{(1+\zeta\bar{\zeta})^2}$, and given the conformal transformations as discussed in (2.1.2), in contrast to ERN spacetime, we obtain superrotations for the null shell placed just outside the horizon $U = \epsilon$ with a compensation along null direction V which makes sure that the soldering transformations also preserve the form of the metric. These transformations mimic the ones originally obtained at asymptotic null infinity. We also compute the intrinsic quantities of the horizon shell which contain supertranslation parameter $T(\theta,\phi)$, can be seen in Bhattacharjee et al.[31]

Next, we show the non-vanishing finite change in the components of the deviation or displacement vector between two nearby timelike geodesics upon passing through the horizon shell. We follow Blau et al.[27] to extend the soldering transformations off the horizon shell in order to compute the induced tensor field γ_{ab}. The computation of the deviation vectors further require the GW degree of freedom which can be expressed as

$$\hat{\gamma}_{\theta\phi} = 2\nabla^{(2)}_{\theta}\partial_{\phi}T(\theta,\phi) \;\; ; \;\; \hat{\gamma}_{\theta\theta} = 2\Big(\nabla^{(2)}_{\theta}\partial_{\theta}T(\theta,\phi) - \frac{1}{\sin^2\theta}\nabla^{(2)}_{\phi}\partial_{\phi}T(\theta,\phi)\Big), \tag{22}$$

where $T(\theta,\phi)$ is a supertranslation parameter. Thus the framework of section 2 and section 3.1 generates the θ-component of the deviation vector,

$$\begin{aligned} X_{\theta} =& \Big(1 + \frac{U}{8m^2}(\nabla^2 T(\theta,\phi) - T(\theta,\phi)\Big)\Big(\Big(4m^2 + U(\nabla^{(2)}\partial_{\theta}T(\theta,\phi) \\ &- \frac{1}{\sin^2\theta}\nabla^{(2)}_{\phi}\partial_{\phi}T(\theta,\phi))\Big)X^{\theta}_{(0)} + U\nabla^{(2)}_{\theta}\partial_{\phi}T(\theta,\phi)X^{\phi}_{(0)}\Big). \end{aligned} \tag{23}$$

The X_{ϕ} component can also be computed in the similar way which again carries supertranslation parameter. It turns out that the surface current vanishes, i.e., $J^A = 0$, hence the test particles will remain on the spatial slice of the metric since $X_V = 0$, but with a relative change in the displacement as it can clearly be seen in Fig. 2, and opposes the result of ERN spacetime. Therefore, the nonvanishing displacement vector X_A depicts the BMS displacement memory in the context of soldering of two Schwarzschild spacetime geometries. One can further integrate Eq. (23) with respect to the parameter of the geodesics in order to have the explicit form of the displacement memory. Next, we shall discuss a more realistic approach of estimating the displacement memory which is analogous to the far region analysis.

3.2. *Memory and BMS symmetries: Analogous to far region*

In this section, we adopt a more realistic approach of computing displacement memory near the horizon of black holes and its possible connection with asymptotic symmetries. The analysis of this section is independent of section 2, i.e., it is not based on soldering of black hole spacetimes. We study displacement memory for non-extremal (fixed temperature) and extremal (zero temperature) black holes.[42] This shows an analogous effect of conventional GW-memory which was originally established at asymptotic null infinity (I^+). The emergence of asymptotic symmetries near the horizon of black holes (not in the context of null shell formalism) has been established by.[25,34] In this respect, we are interested in measuring the permanent relative change in the deviation or displacement of the test detectors induced due to the interaction with GW, and its connection with near-horizon BMS symmetries. In this realistic approach, as it can be seen in the schematic diagram below, the detectors are being placed near the horizon of a black hole ($\mathcal{H}^+$), and we estimate a relative change in the deviation vector of configuration before and after the passage of GWs. The displacement vector S^μ between the detector setup or geodesics evolves according to GDE Eq. (1). Let us consider the general form of the 4-dimensional near-horizon metric[26,43]

$$ds^2 = g_{vv}dv^2 + 2\kappa dvd\rho + 2g_{vA}dvdx^A + g_{AB}dx^Adx^B, \tag{24}$$

with following fall-off conditions for the horizon $\rho = 0$:

$$\begin{aligned} g_{vv} &= -2\kappa\rho + \mathcal{O}(\rho^2) \quad ; \quad k = 1 + \mathcal{O}(\rho^2) \\ g_{vA} &= \rho\theta_A + \mathcal{O}(\rho^2) \quad ; \quad g_{AB} = \Omega\gamma_{AB} + \rho\lambda_{AB} + \mathcal{O}(\rho^2) \end{aligned}$$

where θ_A, Ω and λ^{AB} are functions of (v, x^A). For computational purpose, we consider Ω to be unity. γ_{AB} represents the 2-sphere metric. In stereographic coordinates, $x^A = (\zeta, \bar{\zeta})$, the 2-sphere metric is $\gamma_{AB}dx^Adx^B = \frac{4}{(1+\zeta\bar{\zeta})^2}d\zeta d\bar{\zeta}$. The asymptotic Killing vectors preserving fall-off boundary conditions together with the charges

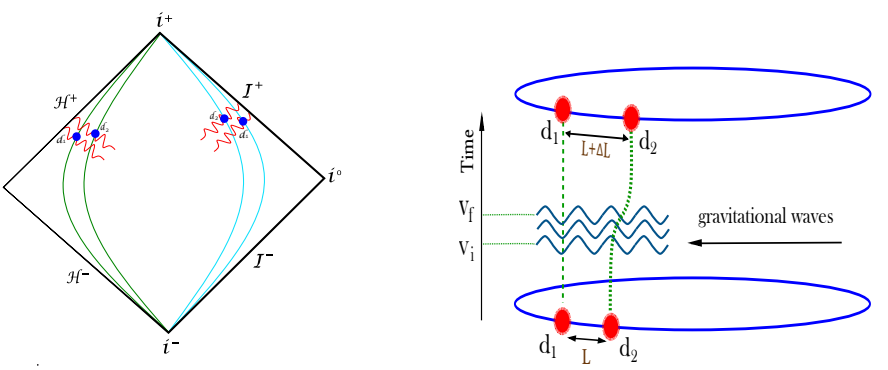


Fig. 3. Schematic diagrams depicting displacement memory effect for the detectors d_1 and d_2. Separation L gets modified permanently before and after the passage of GWs depicted as $L + \Delta L$.

can be found in Donnay et al.[26] It turns out that the variation of κ along Killing direction, when fixed temperature configuration considered, generates a copy of supertranslation together with a superrotation.[25,26] We use this fact in section 3.2.1. Let us consider the fixed temperature configuration first in order to compute displacement memory.

3.2.1. *Memory: Fixed temperature configuration*

The fixed temperature configuration enables us to take κ non-zero but constant. With this consideation, the solution of the GDE for the given metric is[42]

$$\Delta S^{\bar{\zeta}} = \frac{\rho(1+\zeta\bar{\zeta})^2}{4}\Big((\Delta\lambda_{\zeta\bar{\zeta}}S^{\bar{\zeta}} + \Delta\lambda_{\zeta\zeta}S^{\zeta}) - \kappa(\Delta v)^2(\mathcal{H}S^{\bar{\zeta}} + \mathcal{G}S^{\zeta})\Big) + \mathcal{O}(\rho^2), \quad (25)$$

where $S^{\bar{\zeta}}$ and S^{ζ} denote the $\bar{\zeta}$ and ζ components of the deviation vector. Here, we have also used the vA-component of the Einstein field equation, the $\mathcal{O}(\rho^0)$ term gives $\partial_v\theta_A = 0 \;\Rightarrow\; \theta_A = \mathcal{C}(x^A)$. Using vv-component of the Einstein field equation, we replace changes $\Delta\lambda_{AB}$ in Eq. (25) to obtain an explicit form of the memory which also ensures that λ_{AB} can be written in terms of θ_A. The resultant displacement memory is

$$\Delta S^{\bar{\zeta}} = \frac{\rho(1+\zeta\bar{\zeta})^2}{4}\Big(\Big((\kappa\tilde{G}\Delta v - \frac{2aQ}{(1+\zeta\bar{\zeta})^2})S^{\bar{\zeta}} + \kappa\tilde{B}S^{\zeta}\Big)\Delta v - \kappa(\Delta v)^2(\mathcal{H}S^{\bar{\zeta}} + \mathcal{G}S^{\zeta})\Big) + \mathcal{O}(\rho^2), \quad (26)$$

where $\tilde{G}, Q, \tilde{B}, \mathcal{H}$ and $\mathcal{G}$ are functions of $(\zeta, \bar{\zeta})$, also $\mathcal{H}$ is written in terms of metric parameter θ_A. Thus the analyses suggest that the displacement memory is restored in terms of metric parameters. This completes our study of achieving the displacement memory near the horizon of non-extremal black holes. Further, we relate it with the asymptotic symmetries.

Relation with BMS symmetry: We show the explicit relation between BMS symmetries and displacement memory. One can obtain the variation of the metric parameters λ_{AB}, θ_A and κ along the Killing direction. Since the the memory (25) or (26) is independent of v coordinate, we only mention the relevant expression of λ_{AB}-variation,[26] given by

$$\mathcal{L}_\chi\lambda_{AB} = f\partial_v\lambda_{AB} - \lambda_{AB}\partial_v f + \mathcal{L}_Y\lambda_{AB} + \theta_A\partial_B f + \theta_B\partial_A f - 2\nabla_A\nabla_B f. \quad (27)$$

On the other hand, the v component of the Killing vector for the fixed temperature configuration generates two sets of supertranslations $T(\phi)$ and $X(\phi)$, i.e.,

$$f(v, x^A) = T(x^A) + e^{-\kappa v}X(x^A). \quad (28)$$

Also, the Lie derivative of g_{vA} along Killing direction yields superrotation $Y^A(x^A)$.[26]

Now, in order to make the variation of λ_{AB} along the killing direction independent of v, we set v coefficients to be zero. Using the general solutions of λ_{AB} from relevant component of Einstein field equations, and for computational simplification switching off the supertranslation parameter T, we obtain $Y^{\zeta}(\zeta) = \tilde{a}e^{-\int \frac{\tilde{w}}{\tilde{p}} d\zeta}$ as a solution. where, $\tilde{w}$ and $\tilde{p}$ are functions of $(\zeta, \bar{\zeta})$ and $\tilde{a}$ appears as an integration constant; it would be a function of $\bar{\zeta}$ with respect to ζ differential equation. Similarly, one can also find the solution for $Y^{\bar{\zeta}}(\bar{\zeta})$. Therefore, we can find a solution for Y^A that will induce the desired shift in the displacement vector, and serves our purpose of establishing relation between displacement memory and asymptotic symmetries near the horizon of black holes.

3.2.2. *Memory: Zero temperature configuration & BMS symmetries*

We have also provided an explicit approach for zero-temperature configuration (extremal consideration). For this, the metric coefficients remain same as appear in the non-extreme case except the g_{vv} component which becomes $\mathcal{N}(\zeta, \bar{\zeta})\rho^2 + \mathcal{O}(\rho^3)$. The displacement memory with this consideration can be achieved by setting $\kappa = 0$ in Eq. (25), and written as

$$\Delta S_E^{\bar{\zeta}} = \frac{\rho}{4}(1 + \zeta\bar{\zeta})^2(\Delta\lambda_{\zeta\bar{\zeta}} S_E^{\bar{\zeta}} + \Delta\lambda_{\zeta\zeta} S_E^{\zeta}) + \mathcal{O}(\rho^2), \tag{29}$$

where subscript E stands for the change in the displacement vector for *extremal* or *zero temperature* configuration. One can again take the variation of λ_{AB} along Killing direction and obtain set of differential equations similar to the fixed-temperature analyses. The related details can be found in Bhattacharjee et al.[42]

3.2.3. *Memory & BMS symmetries: for a less generic form of the metric*

In this section, we consider a less generic form of the full metric (24) by setting $g_{vA} = 0$ which can be regarded as an asymptotic form of a metric near the horizon of a spherically symmetric black hole deformed in the spatial sector. The displacement memory for this setup can be computed in similar way as obtained for the full metric.

We find that the fixed temperature configuration does not produce very interesting result. However, the displacement memory is written in terms of $\Delta\lambda_{AB}$, and further set of conditions can be obtained in order to have the connection with BMS symmetries. The interesting finding appears if we consider zero temperature configuration. The change in the displacement vector is written in terms of $\Delta\lambda_{AB}$. Using the vv-component of Einstein field equation, the displacement memory is given by

$$\Delta S_E^{\bar{\zeta}} = \frac{\rho}{4}(1 + \zeta\bar{\zeta})^2 H(\zeta, \bar{\zeta})\Delta v S_E^{\bar{\zeta}} + \mathcal{O}(\rho^2). \tag{30}$$

We notice that the displacement memory is proportional to ρ whereas in the far region case[22] the memory is proportional to $\frac{1}{r}$. This implies that the displacement memory near the horizon of black holes is mimicking the one obtained null infinity.

One can further relate it with the BMS symmetries. It turns out that if we freeze off the superrotation, we have an exact solution: $X(\zeta, \bar{\zeta}) = X_1(\zeta) + X_2(\bar{\zeta})$. Hence, there is a supertranslation $X(\zeta, \bar{\zeta})$ that can induce the same shift in the displacement memory. We have also considered the three dimensional analyses for extreme and non-extreme cases, and its connection with near-horizon BMS symmetries in Bhattacharjee et al.[42]

4. Discussion and outlook

The primary motivation of this article is to provide a review study on the displacement memory effect near the horizon of black holes and its connection with asymptotic symmetries which is similar to the one established at null infinity for asymptotically flat spacetimes. In this respect, we started with a brief introduction of the intrinsic formulation of null shell. As a result, we show that the supertranslation-like transformations can be achieved in the context of soldering of two black hole spacetimes. We explicitly provided the results for Schwarzschild and ERN cases. However, interestingly, superrotation can not be recovered for ERN spacetime whereas it can be recovered for Schwarzschild. Further, as a result of interaction between horizon shell and test particles, we find that the particles remain on the initial 2-dimensional surface for Schwarzschild consideration as surface current is zero whereas for ERN spacetime, particles get displaced off from the initial 2-dimensional surface. We compute the components of deviation vectors which carry the supertranslation parameter ensuring that the memory can be obtained in terms of BMS parameters.

We have provided a detailed description of GW memory effect near the horizon of black holes as a more realistic approach which is analogous to the one obtained at null infinity. As a major distinguishing feature in order to establish a connection with asymptotic symmetries, we observed that there are two supertranslation parameters $T(x^A)$ and $X(x^A)$ and one superrotation $Y^A(x^A)$ in near horizon analyses whereas there is only one supertranslation in the far region case. We also notice that $\Delta\lambda_{AB}$ is the data available to be considered in the detection which mimics the data C_{zz} present near the null infinity.[22] The form of the GDE is also quite different with respect to the far region case. These are the brief and major distinguishing features between displacement memory obtained near the horizon of black holes and at asymptotic null infinity, together with its possible connection to BMS symmetries.

The observational features of GW will be extremely useful in investigating the signatures of the asymptotic symmetries in displacement memory effect. As a result, we might be able to look into such symmetries in greater detail as a firm evidence in near future. The theoretical aspects of our study might help as a model framework in the detection prospects of the near-horizon BMS memory effect. It is expected that present-day detectors like LIGO might not be able to play a crucial role in the detection prospects. We hope that advanced detectors like aLIGO or LISA might be able to capture this effect as LISA is looking for a much longer wavelength

opening up the detection realm to a wider range of gravitational wave sources. In this direction, building up the theoretical framework of post-Newtonian (PN) formalism having relevance for asymptotic symmetries, it is interesting to investigate the contribution of supertranslation-like symmetries in non-oscillatory signals of the gravitational wave polarizations. This would surely provide a more direct approach to gravitational wave data analysts in order to search for asymptotic symmetries in GW memory signal.

The alternative approach of detecting astrophysical signatures of asymptotic symmetries from gravitational lensing might not be an appropriate direction as supertranslated geometry of the spacetime will not lead to any deviations from standard results in general relativity for static configurations.[44] The underlying reason to the problem is that given a supertranslated Schwarzschild black hole, one can always choose a coordinate patch in a finite solid angular range where the metric will be given by the Schwarzschild. Hence, the signatures of asymptotic symmetries seem to be difficult from gravitational lensing. However, we still have a hope to detect such symmetries in black hole shadows and lensing if we have a dynamically evolving spacetime, and in this consideration, it is not sure whether one can again have some coordinate transformations which would give rise to indistinguishable features. So this might set a stronger grounds for detecting the asymptotic symmetries through deflection angle approach and black hole shadows. Interestingly, it is not difficult to understand that why gravitational memory is suitable for detecting such symmetries, because it appears as a physical effect where initial and final vacua differ by a BMS supertranslation. On another hand, the post-Newtonian tidal environment analysis especially in terms of BMS symmetries can also be explored from BMS-detection prospects.

Furthermore, as we have been investigating the issues from classical perspectives, the quantum memory effect has also got considerable attention in very recent, and it is yet to be explored extensively. This would certainly give a new meaning to the quantum treatment of the memory to catch on to the hawking information paradox. We know that the displacement memory is induced by the radiative energy flux, and it has been shown that there exists a new kind of gravitational memory- *spin memory effect* which is sourced by angular momentum flux.[45] It is interesting to examine the signatures of asymptotic symmetries in the context of spin memory effect near the horizon of black holes from BMS-detection point of view. Further, the algebra of asymptotic symmetries on null surface situated at a finite location of the manifold might determine some fascinating role of symmetries on GW memory.

Acknowledgement

I would like to express my sincere thanks to Srijit Bhattachrajee for proofreading this article and sharing his stimulating ideas during the preparation of the draft. I would also like to thank Arpan Bhattacharyya for illuminating discussions.

References

1. B. P. Abbott, R. Abbott and E. A. Abbott, Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, p. 061102 (Feb 2016).
2. R. Abbott, T. D. Abbott and E. A. Abraham, Gw190521: A binary black hole merger with a total mass of 150 $M_{\odot}$, *Phys. Rev. Lett.* **125**, p. 101102 (Sep 2020).
3. Y. B. Zel'dovich and A. G. Polnarev, Radiation of gravitational waves by a cluster of superdense stars, *Sov. Astron.* **18**, p. 17 (1974).
4. V. B. Braginsky and L. P. Grishchuk, Kinematic Resonance and Memory Effect in Free Mass Gravitational Antennas, *Sov. Phys. JETP* **62**, 427 (1985).
5. D. Christodoulou, Nonlinear nature of gravitation and gravitational wave experiments, *Phys. Rev. Lett.* **67**, 1486 (1991).
6. L. Bieri and D. Garfinkle, Perturbative and gauge invariant treatment of gravitational wave memory, *Phys. Rev. D* **89**, p. 084039 (Apr 2014).
7. A. Strominger and A. Zhiboedov, Gravitational Memory, BMS Supertranslations and Soft Theorems, *JHEP* **01**, p. 086 (2016).
8. A. A. Rahman and R. M. Wald, Black hole memory, *Phys. Rev. D* **101**, p. 124010 (Jun 2020).
9. M. Favata, The Gravitational-wave memory from eccentric binaries, *Phys. Rev. D* **84**, p. 124013 (2011).
10. M. Favata, Post-newtonian corrections to the gravitational-wave memory for quasicircular, inspiralling compact binaries, *Phys. Rev. D* **80**, p. 024002 (Jul 2009).
11. M. Favata, Nonlinear gravitational-wave memory from binary black hole mergers, *The Astrophysical Journal* **696**, L159 (Apr 2009).
12. M. Favata, Gravitational-wave memory revisited: Memory from the merger and recoil of binary black holes, *Journal of Physics: Conference Series* **154**, p. 012043 (Mar 2009).
13. K. Islo, J. Simon, S. Burke-Spolaor and X. Siemens, Prospects for Memory Detection with Low-Frequency Gravitational Wave Detectors (6 2019).
14. P. D. Lasky, E. Thrane, Y. Levin, J. Blackman and Y. Chen, Detecting gravitational-wave memory with LIGO: implications of GW150914, *Phys. Rev. Lett.* **117**, p. 061102 (2016).
15. O. M. Boersma, D. A. Nichols and P. Schmidt, Forecasts for detecting the gravitational-wave memory effect with Advanced LIGO and Virgo, *Phys. Rev. D* **101**, p. 083026 (2020).
16. D. Pollney and C. Reisswig, Gravitational memory in binary black hole mergers, *The Astrophysical Journal* **732**, p. L13 (Apr 2011).
17. A. M. Grant and D. A. Nichols, Persistent gravitational wave observables: Curve deviation in asymptotically flat spacetimes (9 2021).
18. M. Hübner, C. Talbot, P. D. Lasky and E. Thrane, Measuring gravitational-wave memory in the first ligo/virgo gravitational-wave transient catalog, *Phys. Rev. D* **101**, p. 023011 (Jan 2020).
19. T. Islam, S. E. Field, G. Khanna and N. Warburton, Survey of gravitational wave memory in intermediate mass ratio binaries (9 2021).
20. M. Hübner, P. Lasky and E. Thrane, Memory remains undetected: Updates from the second LIGO/Virgo gravitational-wave transient catalog, *Phys. Rev. D* **104**, p. 023004 (2021).
21. H. Bondi, M. G. J. Van der Burg and A. W. K. Metzner, Gravitational waves in general relativity, vii. waves from axi-symmetric isolated system, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **269**, 21 (1962).
22. A. Strominger, Lectures on the infrared structure of gravity and gauge theory (2018).

23. S. W. Hawking, M. J. Perry and A. Strominger, Soft hair on black holes, *Phys. Rev. Lett.* **116**, p. 231301 (Jun 2016).
24. P.-M. Zhang, C. Duval, G. W. Gibbons and P. A. Horvathy, Soft gravitons and the memory effect for plane gravitational waves, *Phys. Rev. D* **96**, p. 064013 (Sep 2017).
25. L. Donnay, G. Giribet, H. A. González and M. Pino, Supertranslations and superrotations at the black hole horizon, *Phys. Rev. Lett.* **116**, p. 091101 (Mar 2016).
26. L. Donnay, G. Giribet, H. A. González and M. Pino, Extended symmetries at the black hole horizon, *Journal of High Energy Physics* **2016**, p. 100 (Sep 2016).
27. M. Blau and M. O'Loughlin, Horizon Shells and BMS-like Soldering Transformations, *JHEP* **03**, p. 029 (2016).
28. S. Bhattacharjee and A. Bhattacharyya, Soldering freedom and bondi-metzner-sachs-like transformations, *Phys. Rev. D* **98**, p. 104009 (Nov 2018).
29. C. Barrabès and W. Israel, Thin shells in general relativity and cosmology: The lightlike limit, *Phys. Rev. D* **43**, 1129 (Feb 1991).
30. C. Barrabes and P. A. Hogan, *Singular Null Hypersurfaces in General Relativity: Light-Like Signals from Violent Astrophysical* (World Scientific Publishing Co. Pte. Ltd., 2003).
31. S. Bhattacharjee, S. Kumar and A. Bhattacharyya, Memory effect and bms-like symmetries for impulsive gravitational waves, *Phys. Rev. D* **100**, p. 084010 (Oct 2019).
32. S. Bhattacharjee and S. Kumar, Memory effect and bms symmetries for extreme black holes, *Phys. Rev. D* **102**, p. 044041 (Aug 2020).
33. G. Barnich and C. Troessaert, Symmetries of asymptotically flat four-dimensional spacetimes at null infinity revisited, *Phys. Rev. Lett.* **105**, p. 111103 (Sep 2010).
34. G. Barnich and C. Troessaert, Supertranslations call for superrotations (2012).
35. M. Volonteri, P. Madau, E. Quataert and M. J. Rees, The distribution and cosmic evolution of massive black hole spins, *The Astrophysical Journal* **620**, 69 (Feb 2005).
36. L. Gou, J. E. McClintock, R. A. Remillard, J. F. Steiner, M. J. Reid, J. A. Orosz, R. Narayan, M. Hanke and J. García, Confirmation via the continuum-fitting method that the spin of the black hole in Cygnus x-1 is extreme, *The Astrophysical Journal* **790**, p. 29 (Jun 2014).
37. J. E. McClintock, R. Shafee, R. Narayan, R. A. Remillard, S. W. Davis and L.-X. Li, The spin of the near-extreme kerr black hole GRS 1915+105, *The Astrophysical Journal* **652**, 518 (Nov 2006).
38. L. W. Brenneman and C. S. Reynolds, Constraining black hole spin via x-ray spectroscopy, *The Astrophysical Journal* **652**, 1028 (Dec 2006).
39. A. Strominger and C. Vafa, Microscopic origin of the bekenstein-hawking entropy, *Physics Letters B* **379**, 99 (Jun 1996).
40. B. Carter, The complete analytic extension of the reissner-nordström metric in the special case e2 = m2, *Physics Letters* **21**, 423 (1966).
41. S. Liberati, T. Rothman and S. Sonego, Nonthermal nature of incipient extremal black holes, *Phys. Rev. D* **62**, p. 024005 (Jun 2000).
42. S. Bhattacharjee, S. Kumar and A. Bhattacharyya, Displacement memory effect near the horizon of black holes, *Journal of High Energy Physics* **2021**, p. 134 (Mar 2021).
43. L. Donnay, G. Giribet, H. A. González and A. Puhm, Black hole memory effect, *Phys. Rev. D* **98**, p. 124016 (Dec 2018).
44. G. Compère and J. Long, Classical static final state of collapse with supertranslation memory, *Classical and Quantum Gravity* **33**, p. 195001 (Sep 2016).
45. S. Pasterski, A. Strominger and A. Zhiboedov, New gravitational memories, *Journal of High Energy Physics* **2016**, p. 53 (Dec 2016).

Physical black holes in semiclassical gravity

Sebastian Murk

Department of Physics and Astronomy, Macquarie University, Sydney, New South Wales 2109, Australia
and
Sydney Quantum Academy, Sydney, New South Wales 2006, Australia
E-mail: sebastian.murk@mq.edu.au

Daniel R. Terno

Department of Physics and Astronomy, Macquarie University, Sydney, New South Wales 2109, Australia
E-mail: daniel.terno@mq.edu.au

We derive and critically examine the consequences that follow from the formation of a regular black or white hole horizon in finite time of a distant observer. In spherical symmetry, only two distinct classes of solutions to the semiclassical Einstein equations are self-consistent. Both are required to describe the formation of physical black holes and violate the null energy condition in the vicinity of the outer apparent horizon. The near-horizon geometry differs considerably from that of classical solutions. If semiclassical physics is valid, accretion into a black hole is no longer possible after the horizon has formed. In addition, the two principal generalizations of surface gravity to dynamical spacetimes are irreconcilable, and neither can describe the emission of nearly-thermal radiation. Comparison of the required energy and timescales with established semiclassical results suggests that if the observed astrophysical black holes indeed have horizons, their formation is associated with new physics.

Keywords: Black holes; white holes; general relativity; semiclassical gravity; quantum aspects of black holes; energy conditions; surface gravity.

1. Introduction

Our current understanding of ultra-compact objects (UCOs) can be summarized as follows: the existence of astrophysical black holes (ABHs) — dark massive compact objects — is established beyond any reasonable doubt. However, it is unclear when, how, or if at all these UCOs develop the standard black hole attributes, such as horizons and singularities. A large number of models, often with deliberately designed features (or lack thereof), purport to describe ABHs. Translating the differences between these models into potentially observable differences between the signals received from the black hole candidates they describe is one of the most exciting topics in gravitational physics research.[1,2]

There is no unanimously agreed upon definition of a black hole.[3] The salient feature of a mathematical black hole (MBH)[4] is the event horizon that separates our outside world from an inaccessible interior. However, event horizons are global

teleological entities that are — even in principle — unobservable,[5,6] and observational, numerical, and theoretical studies focus on other characteristics of black holes. A trapped region is a spacetime domain where both ingoing and outgoing future-directed null geodesics emanating from a spacelike two-dimensional surface with spherical topology have negative expansion.[7–9] Its evolving outer boundary is the apparent horizon. This definition captures the most fundamental feature of black holes as spacetime regions that nothing, not even light, can escape. It is also local, and thus physically observable: the escape is not possible now, but the notion of "now" depends on the observer. Following the characterization scheme of Ref. 4, we refer to a trapped spacetime region that is bounded by an apparent horizon as a physical black hole (PBH).[10]

Although it cannot be observed, the event horizon is a useful asymptotic concept, and in many situations it is reasonable to assume that MBHs provide a good description of certain aspects of ABHs. Indeed, the predictions of various alternatives to black holes are most often compared with the standard Schwarzschild/Kerr paradigm.[1,2] However, using this asymptotic concept at finite times is logically unsatisfactory, and in the situations we consider below cannot be sustained.

Another question is whether the asymptotic picture is adequate to describe the relevant geometric and physical properties of black holes. We first formalize the minimal assumptions that underpin the widely used black hole notions presented in Fig. 1, namely regularity of the apparent horizon and its finite-time formation according to the clock of a distant observer. In spherical symmetry, these two assumptions uniquely determine the formation scenario for PBHs and suffice to provide an exhaustive description of their near-horizon geometry. Its properties are the subject of this contribution. We now outline the motivation and justification for these requirements.[10]

In classical general relativity (GR) non-spacelike singularities destroy predictability. According to the weak cosmic censorship conjecture,[16,17] spacetime singularities are concealed by event horizons. Quantum gravitational effects are expected to become important when the spacetime curvature is sufficiently strong,[8,18] i.e. when the Kretschmann scalar $\mathcal{K} := R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ reaches the Planck scale, that is $\mathcal{K} \gtrsim l_{\mathrm{P}}^{-4}$. The near-horizon geometry is believed to be well-described by semi-classical physics. Nontrivial quantum effects are obtained within the framework of quantum field theory in curved spacetime.[18] One of its most spectacular predictions is Hawking radiation: it not only completed black hole thermodynamics,[8,19] but has given rise to the infamous information loss paradox.[12] Regular (singularity-free) black holes were introduced to altogether eliminate singularities in classical gravitational collapse or as a way to resolve the problem of information loss.[4,20,21] Leaving aside the interior structure of black holes, we formulate the regularity criterion as the absence of curvature singularities at the apparent horizon. A precise mathematical formulation[10] is provided in Sec. 2.

To be considered a genuine physical object rather than merely a useful mathematical tool, the horizon must form in finite time of a distant observer, and there

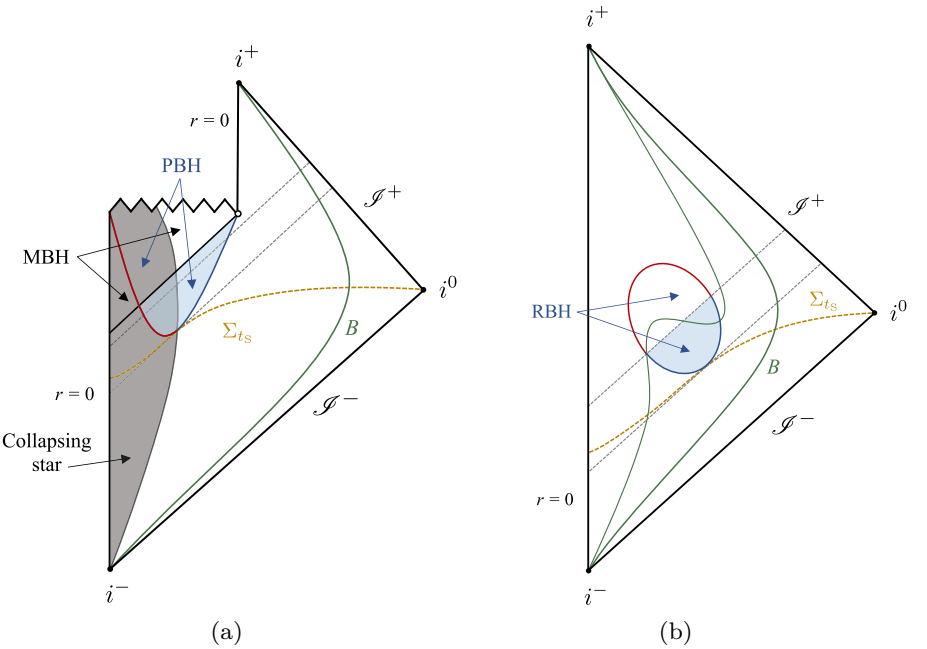


Fig. 1. Schematic Carter–Penrose diagram of the conventional formation and evaporation of (a) a black hole, and (b) a RBH. The outer apparent horizon $r_g(t)$ and the inner apparent horizon $r_{\rm in}(t)$ form the boundary of a PBH and are shown in blue and red, respectively. The equal time surface Σ_{t_S} is drawn as a dashed orange line. It is null at $\big(t_S, r_g(t_S)\big)$ and spacelike everywhere else.[11] The trajectory of a distant observer Bob is indicated in green and labeled "B". Dashed grey lines correspond to outgoing radial null geodesics. (a) Diagrams of this type are elaborations of the original sketch by Hawking, see Fig. 5 of Ref. 12. Spacetime regions corresponding to PBH (MBH) solutions are indicated by blue (black) arrows. Light signals emitted from the quantum ergosphere[13] (indicated by the light blue shading), i.e. from within the trapped region, but outside of the event horizon, reach $\mathscr{I}^+$ and are detected by Bob in his finite proper time. The collapsing matter and its surface are shown as in conventional depictions of the collapse. However, the matter in the vicinity of the outer apparent horizon $\big(t, r_g(t)\big)$ violates the NEC for $t \geqslant t_S$. Moreover, the energy density, pressure, and flux as seen by an infalling observer Alice vary continuously across it, and the equation of state dramatically differs from that of normal matter that may have been used to model the initial EMT of the collapse. (b) The asymptotic structure of a simple RBH spacetime[14] coincides with that of Minkowski spacetime. Conditions for the smooth joining of the inner and outer apparent horizon are described in Ref. 15. An unmarked green line represents a hypersurface $r = \text{const}$ that passes through the RBH.

should be some potentially (at least in principle) observable consequences of its formation. Moreover, if black holes do indeed emit Hawking radiation, and their evolution roughly resembles that of Fig. 1(a), then the apparent horizon forms in finite time t_S of a distant observer. We note that, as illustrated in Fig. 1(a), the outer apparent horizon is located outside of the event horizon.[8,12] Hence all signals that are emitted from the so-called quantum ergosphere[13,15] — part of the trapped region that lies outside of the event horizon — reach future null infinity $\mathscr{I}^+$. In addition, models of transient (even if long-lived) regular black holes (RBHs) imply

the finite-time formation and disappearance of the trapped region, as depicted in Fig. 1(b). This leads to our second requirement: the finite-time formation of an apparent horizon according to a distant observer.

Working in the framework of semiclassical gravity, we use classical notions (horizons, trajectories, etc.) and describe dynamics via the Einstein equations $G_{\mu\nu} = 8\pi T_{\mu\nu}$. We do not assume any specific matter content nor a specific quantum state ω that produces the expectation values of the energy-momentum tensor (EMT) $T_{\mu\nu} := \langle \hat{T}_{\mu\nu} \rangle_\omega$. Note that this EMT describes the total matter content — both the original collapsing matter and the produced excitations. We do not assume the presence of Hawking-like radiation, an event horizon, or a singularity. To simplify the exposition we work in asymptotically flat spacetimes, even if it is not essential for the resulting near-horizon geometries.

2. Spherically symmetric solutions

It is convenient to use Schwarzschild coordinates to impose the two assumptions described above. Their coordinate singularities allow us to identify the admissible solutions. A general spherically symmetric metric in four spacetime dimensions is given by

$$ds^2 = -e^{2h(t,r)} f(t,r) dt^2 + f(t,r)^{-1} dr^2 + r^2 d\Omega_2. \tag{1}$$

These coordinates provide geometrically preferred foliations with respect to Kodama time, a natural divergence-free preferred vector field.[9,22] In asymptotically flat spacetimes the time t represents the proper time of a distant static observer. Using the advanced null coordinate v, the metric is written as

$$ds^2 = -e^{2h_+} \left(1 - \frac{C_+}{r}\right) dv^2 + 2e^{h_+} dv dr + r^2 d\Omega_2. \tag{2}$$

The Misner–Sharp (MS) mass $C(t,r)/2$[9,23] is invariantly defined via

$$f(t,r) := 1 - C/r := \partial_\mu r \partial^\mu r, \tag{3}$$

and thus $C(t,r) \equiv C_+\big(v(t,r), r\big)$. The functions $h(t,r)$ and $h_+(v,r)$ play the role of integrating factors in coordinate transformations, such as

$$dt = e^{-h}(e^{h_+} dv - f^{-1} dr). \tag{4}$$

The apparent horizon is located at the Schwarzschild radius $r_g(t) \equiv r_+(v)$ that is the largest root of $f(t,r) = 0$.[9,24] In (v,r) coordinates the tangents to ingoing and outgoing radial geodesics are given by

$$l^\mu_{\text{in}} = (0, -e^{-h_+}, 0, 0), \qquad l^\mu_{\text{out}} = (1, \tfrac{1}{2} e^{h_+} f, 0, 0), \tag{5}$$

respectively. They are normalized to satisfy $l_{\text{in}} \cdot l_{\text{out}} = -1$, and their corresponding expansions are

$$\vartheta_{\text{in}} = -\frac{2e^{-h_+}}{r}, \qquad \vartheta_{\text{out}} = \frac{e^{h_+} f}{r}, \tag{6}$$

thus identifying the domain $r \leqslant r_+ \equiv r_g$ as a PBH.

Using the retarded null coordinate u, the metric is written as

$$ds^2 = -e^{2h_-}\left(1 - \frac{C_-}{r}\right)du^2 - 2e^{h_-}dudr + r^2 d\Omega_2. \tag{7}$$

In (u, r) coordinates the tangents to ingoing and outgoing radial geodesics are given by

$$l^\mu_{\text{in}} = (1, -\tfrac{1}{2}f(u,r)e^{h_-(u,r)}, 0, 0), \qquad l^\mu_{\text{out}} = (0, e^{-h_-(u,r)}, 0, 0), \tag{8}$$

respectively. Again, the tangents are normalized by $l_{\text{in}} \cdot l_{\text{out}} = -1$, and their respective expansions are given by

$$\vartheta_{\text{in}} = -\frac{e^{h_-}f}{r}, \qquad \vartheta_{\text{out}} = \frac{2e^{-h_-}}{r}. \tag{9}$$

As a result, the domain $r \leqslant r_- \equiv r_g$ is a white hole and the Schwarzschild radius is the location of the anti-trapping horizon. As in the case of the maximally extended Schwarzschild solution, these two scenarios describe different physical settings.

The Schwarzschild coordinates become singular as $r \to r_g$. We extract information about the EMT and therefore about the near-horizon geometry by studying how various divergences cancel to produce finite curvature scalars. Singular points are identified through the presence of incomplete geodesics in their vicinity and are excluded from manifolds representing the spacetime. These geodesics are inextendible in at least one direction, but the range of their generalized affine parameter (proper time for timelike geodesics) is bounded.[7,17,25,26] We focus on curvature singularities and formalize the regularity requirement as the demand that curvature scalars built from polynomials of Riemann tensor components are finite. This condition corresponds to the absence of essential scalar curvature singularities. More stringent conditions that involve higher covariant derivatives or regularity of individual components are not imposed.

In a general four-dimensional spacetime, there are 14 algebraically independent scalar invariants that can be constructed from the Riemann tensor.[27] We use two quantities that are expressed straightforwardly from components of the EMT, namely

$$\tilde{\mathrm{T}} := T^\mu{}_\mu\,, \qquad \tilde{\mathfrak{T}} := T_{\mu\nu}T^{\mu\nu}. \tag{10}$$

The Einstein equations relate them to the curvature scalars as $\tilde{\mathrm{T}} \equiv -\mathcal{R}/8\pi$ and $\tilde{\mathfrak{T}} \equiv R^{\mu\nu}R_{\mu\nu}/64\pi^2$, where $R_{\mu\nu}$ and $\mathcal{R}$ are the Ricci tensor and Ricci scalar, respectively. In spherical symmetry, if the two scalars of Eq. (10) are finite, then the rest are finite as well.[28]

It is convenient to introduce

$$\tau_t := e^{-2h}T_{tt}\,, \qquad \tau_t{}^r := e^{-h}T_t{}^r\,, \qquad \tau^r := T^{rr}. \tag{11}$$

The three Einstein equations for G_{tt}, $G_t{}^r$, and G^{rr} can then be written as

$$\partial_r C = 8\pi r^2 \tau_t / f, \tag{12}$$

$$\partial_t C = 8\pi r^2 e^h \tau_t{}^r, \tag{13}$$

$$\partial_r h = 4\pi r \left(\tau_t + \tau^r\right) / f^2. \tag{14}$$

Since $T^\theta{}_\theta \equiv T^\varphi{}_\varphi$ in spherical symmetry, the two scalars have the form

$$\tilde{\mathrm{T}} = \mathrm{T} + 2T^\theta{}_\theta, \qquad \tilde{\mathfrak{T}} = \mathfrak{T} + 2\big(T^\theta{}_\theta\big)^2, \tag{15}$$

where

$$\mathrm{T} := \left(\tau^r - \tau_t\right) / f, \qquad \mathfrak{T} := \left((\tau^r)^2 + (\tau_t)^2 - 2(\tau_t{}^r)^2\right) / f^2. \tag{16}$$

The curvature scalars are finite even if the individual terms are divergent in one of the two possible cases. The three effective EMT components τ_a, $a \in \{_t, _t{}^r, ^r\}$, may result in finite T and $\mathfrak{T}$, with $T^\theta{}_\theta$ [$G^\theta{}_\theta/8\pi$ with the metric functions that solve Eqs. (12)–(14)] being finite as well, or all four non-zero components $T^\mu{}_\nu$ may diverge, but these divergences cancel to produce the finite scalars of Eq. (15). We first describe the former possibility and then explain why the latter possibility is not realized.

A priori, to ensure that T and $\mathfrak{T}$ remain finite as $r \to r_g$, the effective EMT components τ_a must either diverge according to the same power law f^k, $k < 0$, converge to finite limits ($k = 0$), slowly converge to zero ($0 < k < 1$), or converge to zero with a variety of possible exponents k_a, $k_a \geqslant 1$. It was shown that only two classes of dynamic solutions (with the leading terms in the functions τ_a scaling as f^k, $k = 0, 1$) satisfy the regularity conditions.[29,30] The component $T^\theta{}_\theta$ is either finite due to the Einstein equations (or, possibly, due to some additional constraints on the functions C and h), resulting in admissible solutions, or divergent, indicating the inadmissibility of the solution as $T^\theta{}_\theta$ cannot have the same leading divergence as τ_a / f.[29] Moreover, a slow divergence (or a faster convergence of either τ_t or τ^r) is incompatible with Eqs. (13) and (14). Hence finite values of the two functions of Eq. (16) are a necessary condition to ensure that the curvature scalars are finite. We now briefly summarize the properties of the admissible solutions.

2.1. *Generic solution ($k = 0$)*

In principle, the solutions with $k = 0$ allow for $\tau_t \to \tau^r \to \mp\Upsilon^2$, $\tau_t{}^r \to \mp\Upsilon^2$ for some $\Upsilon(t) > 0$, but only

$$\tau_t \approx \tau^r = -\Upsilon^2 + \mathcal{O}(\sqrt{x}), \tag{17}$$

$$\tau_t{}^r = \mp\Upsilon^2 + \mathcal{O}(\sqrt{x}), \tag{18}$$

where $x := r - r_g$ denotes the coordinate distance from the horizon, describe valid solutions: taking $\tau_t \to \tau^r \to +\Upsilon^2$ results in complex-valued solutions of Eqs. (12)–(14) (see Ref. 31 for details).

The negative sign in the leading term of τ_t and τ^r leads to the violation of the null energy condition (NEC)[7,32,33] in the vicinity of the apparent horizon, i.e. a future-directed outward (inward) pointing radial null vector ℓ^μ does not satisfy $T_{\mu\nu}\ell^\mu\ell^\nu \geqslant 0$ for the contracting (expanding) Schwarzschild radius r_g.[31] An immediate consequence of this result is that accreting Vaidya black hole solutions in (v, r) coordinates cannot describe PBHs as they satisfy the NEC.[31]

The metric functions that solve Eqs. (12) and (14) are

$$C = r_g - 4\sqrt{\pi} r_g^{3/2} \Upsilon \sqrt{x} + \mathcal{O}(x), \tag{19}$$

$$h = -\frac{1}{2} \ln \frac{x}{\xi} + \mathcal{O}(\sqrt{x}), \tag{20}$$

where $\xi(t)$ is determined by the choice of time variable and the higher-order terms depend on the higher-order terms in the EMT expansion.[34] Eq. (13) must then hold identically. Both sides contain terms that diverge as $1/\sqrt{x}$, and their identification results in the consistency condition

$$r_g'/\sqrt{\xi} = 4\varepsilon_\pm \sqrt{\pi r_g}\, \Upsilon, \tag{21}$$

which describes the expansion or contraction of the Schwarzschild sphere depending on the sign of $\varepsilon_\pm = \pm 1$.

Useful information can be obtained by working with retarded and advanced null coordinates that result in regular metric functions.[31] If $r_g' > 0$, this is achieved by using the retarded null coordinate u. For $r_g' < 0$, the advanced null coordinate v leads to

$$C_+(v, r) = r_+(v) + \sum_{i \geqslant 1}^{\infty} w_i(v)(r - r_+)^i, \tag{22}$$

$$h_+(v, r) = \sum_{i \geqslant 1}^{\infty} \chi_i(v)(r - r_+)^i, \tag{23}$$

for some functions $w_i(v)$, $\chi_i(v)$, where $w_1 \leqslant 1$ due to the definition of r_+. This is the general form of the metric functions in (v, r) coordinates that ensures finite curvature scalars at the apparent horizon if $r_g' < 0$.[29] The components of the EMT in (v, r) and (t, r) coordinates are related by

$$\theta_v := e^{-2h_+} \Theta_{vv} = \tau_t\,, \tag{24}$$

$$\theta_{vr} := e^{-h_+} \Theta_{vr} = \left(\tau_t{}^r - \tau_t\right)/f, \tag{25}$$

$$\theta_r := \Theta_{rr} = \left(\tau^r + \tau_t - 2\tau_t{}^r\right)/f^2, \tag{26}$$

where $\Theta_{\mu\nu}$ labels the EMT components in (v, r) coordinates, which are finite. Hence $r_g' < 0$ corresponds to an evaporating PBH. Similarly, (u, r) coordinates are regular at the anti-trapping horizon r_g of an expanding white hole, $r_g' > 0$.[10,31]

The limiting form of the (tr) block of the EMT as $r \to r_g$ is

$$T^a{}_b = \begin{pmatrix} \Upsilon^2/f & -\varepsilon_\pm e^{-h} \Upsilon^2/f^2 \\ \varepsilon_\pm e^h \Upsilon^2 & -\Upsilon^2/f \end{pmatrix}, \qquad T_{\hat{a}\hat{b}} = \frac{\Upsilon^2}{f} \begin{pmatrix} -1 & \varepsilon_\pm \\ \varepsilon_\pm & -1 \end{pmatrix}, \tag{27}$$

where the second expression is written in the orthonormal frame. It makes the violation of the NEC particularly transparent.

In the test-field limit[18] quantum fields propagate on a given gravitational background, but the resulting EMT is not permitted to backreact on the geometry via the Einstein equations. It is instructive to compare the tensor of Eq. (27) with explicit results obtained in the test-field limit. Out of the three popular choices for the vacuum state,[8,18] only the Unruh vacuum results in an EMT with nonzero T_{tr} components. The state itself corresponds to the requirement that no particles impinge on the collapsing object from infinity.[35] In the context of a static maximally extended spacetime, its counterpart is a state with unpopulated modes at past null infinity and the white hole horizon.[8,18] Using various semi-analytical and numerical methods that are based on conformally coupled fields[36] and minimally coupled scalar field,[37] the expectation values of the renormalized components T_{tt}, $T_t{}^r$, and T^{rr} have been determined explicitly. They approach the same negative value as $r \to r_g$.

The experiences of observers in the vicinity of the apparent horizon depend on their trajectories. A static observer finds that the energy density $\rho := T_{\mu\nu}\, u^\mu u^\nu = -T^t_{\ t}$, pressure $p := T_{\mu\nu}\, n^\mu n^\nu = T^r_{\ r}$, and flux $\phi := T_{\mu\nu}\, u^\mu n^\nu$ diverge at the apparent horizon, where u^μ denotes the four-velocity and n^μ the outward-pointing radial spacelike vector. The experience of a radially-infalling observer Alice moving on the trajectory $x^\mu_{\rm A}(\tau) = (t_{\rm A}, r_{\rm A}, 0, 0)$ is different, and also differs from the infall into a classical eternal black hole.

First, crossing the apparent horizon of a PBH happens not only at some finite proper time τ_0, but also at a finite time $t_0(\tau_0)$, $r_g\big(t_0(\tau_0)\big) = r_{\rm A}(\tau_0)$ according to the clock of a distant Bob. This is particularly easy to see for radial null geodesics, where

$$\frac{dt}{dr} = -\frac{e^{-h(t,r)}}{f(t,r)} \to \pm\frac{1}{|r_g'|} \tag{28}$$

at $r = r_g$, the rhs is obtained using Eqs. (19)–(21),[29] and the upper (lower) signature corresponds to an outgoing (ingoing) null ray. We use this result in our estimate of the formation time in Sec. 4, and in showing that the simple generalizations of surface gravity to nonstationary spacetimes fail for PBHs (Sec. 3.2).

It is possible to find explicit relations between (t, r) and (v, r) coordinates in the vicinity of the apparent horizon,[30] namely

$$x(r_+ + y, v) = r_+ + y - r_g\big(t(v, r_+ + y)\big) = -r_g'' y^2/(2r_g'^2) + \mathcal{O}(y^3), \tag{29}$$

which relates the coordinates $x := r - r_g(t)$ and $y(r, v) := r - r_+(v)$. Another useful relation[38] is

$$w_1 = 1 - 2\sqrt{2\pi r_g^3 |r_g''|}\,\frac{\Upsilon}{|r_g'|}. \tag{30}$$

For an evaporating PBH ($r_g' < 0$) the energy density, pressure, and flux in Alice's frame are finite. However, upon approaching the anti-trapping horizon of an

expanding white hole ($r_g' > 0$), Alice encounters a firewall,

$$\rho_{\rm A} = T_{\mu\nu}\, u^\mu_{\rm A} u^\nu_{\rm A} = -\frac{\dot{r}^2_{\rm A}}{4\pi r_g X} + \mathcal{O}(1/\sqrt{X}), \tag{31}$$

where $X := r_A(\tau) - r_g\big(t_{\rm A}(\tau)\big)$,[28] which may be another indication of the instability[8] of such objects.

Violations of the NEC are bounded by quantum energy inequalities.[33,39] Outside of the singularities the lower bounds were shown to exist for the energy density $\langle \hat{T}_{\mu\nu} \rangle_\omega u^\mu u^\nu$ and its smeared averages. These are known to be state-independent for free fields. For spacetimes of small curvature, explicit bounds for a geodesic observer were derived in Ref. 40. A finite bound is violated by the $1/f^2$ divergence of the energy density that results in the logarithmic divergence of its smeared time average provided that the radial velocity does not go to zero.[11,29] However, this is not true for geodesics, and thus the inequality of Ref. 40 is inapplicable.

Observers that have crossed the apparent horizon but change their mind before traversing the quantum ergosphere can exit the black hole before it evaporates. Their experiences at the apparent horizon are more involved: for a geodesic observer that attempts to cross the apparent horizon from the inside, the energy density also diverges. However, it does so according to a weaker $1/f$ law, and thus the integrated energy density remains finite.

2.2. *Extreme solution ($k = 1$)*

The static solution with $k = 0$ is impossible, as in this case $\mathfrak{T}$ would diverge at the apparent horizon. Consequently, EMT components that allow for static solutions must behave differently. Many models of static nonsingular black holes assume finite values of energy density and pressure at the horizon.[4,21] With respect to the scaling behavior f^k of the effective EMT components of Eq. (11) this is the $k = 1$ solution, with

$$\tau_t \to E(t) f, \qquad \tau_t{}^r \to \Phi(t) f, \qquad \tau^r \to P(t) f, \tag{32}$$

where $\rho = E$ and $p = P$ at the apparent horizon. Any two functions can be expressed algebraically in terms of the third, and $8\pi r_g^2 E \leqslant 1$ to ensure that $C(t, r) - r_g > 0$ for $r > r_g$.

Only the extreme value of $E = (8\pi r_g^2)^{-1}$ corresponds to non-extreme dynamic black holes (i.e. those with $r_g \neq$ const whose trapped regions have nonzero volume).[30] As a result

$$C(t, r) = r - c_{32}(t) x^{3/2} + \mathcal{O}(x^2) \tag{33}$$

for some coefficient $c_{32}(t) > 0$, setting via Eq. (32) the scaling of other leading terms in the EMT. Consistency of Eqs. (13) and (14) implies $P = -E = -1/(8\pi r_g^2)$ and $\Phi = 0$. From the next-order expansion we obtain

$$h = -\frac{3}{2} \ln \frac{x}{\xi} + \mathcal{O}(\sqrt{x}), \tag{34}$$

as well as the consistency relation

$$r_g' = \varepsilon_\pm c_{32} \xi^{3/2} / r_g, \tag{35}$$

where the case $\varepsilon_\pm \equiv -1$ ($\varepsilon_\pm \equiv +1$) describes an evaporating PBH (an expanding white hole). The NEC violation is more subtle than in the $k = 0$ case. At $r = r_g(t)$ itself the NEC is marginally satisfied, as

$$T_{\hat{a}\hat{b}} = \frac{1}{8\pi r_g^2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{36}$$

However, the NEC is violated for some $x > 0$, as for both incoming and outgoing directions $l_{\rm in}$, $l_{\rm out}$, we have

$$T_{\hat{a}\hat{b}} l^{\hat{a}} l^{\hat{b}} = -\frac{3 c_{32} \sqrt{x}}{8\pi r_g^2} + \mathcal{O}(x). \tag{37}$$

Solutions with a time-independent apparent horizon or general static solutions do not require $w_1 = 1$ to satisfy Eqs. (24)–(26). Since $r_+(v) = r_g(t) = \text{const}$, it is possible to have non-extreme solutions. Then Eq. (13) implies $\Phi = 0$ and the identity $E = -P$ follows from Eq. (26), leading to a regular function $h(t, r)$. However, in this case Eq. (4) indicates that the apparent horizon cannot be reached in finite time t.

The curvature scalars of the solutions we have considered so far were rendered finite thanks to the constraint that Eq. (13) imposes on the dynamic behavior via Eqs. (21) and (35). For example, the Ricci scalar of the solution with $C(t, r)$ given by Eq. (33) includes the potentially divergent term

$$R_{\rm div} = \frac{3}{4\sqrt{x}} \left(\frac{r_g r_g'^2}{c_{32} \xi^3} - \frac{c_{32}}{r_g} \right) \tag{38}$$

that vanishes upon substitution of Eq. (35). However, a metric with $C(t, r) = r + c_2 x^2 + \ldots$ and the same $h(t, r)$ of Eq. (34), while being a solution of the Einstein equations, has a finite value of the Ricci scalar only in the static case $r_g' = 0$, as in this case

$$R_{\rm div} = -\frac{3 r_g r_g'^2}{c_2 \xi^3 x}. \tag{39}$$

3. Implications

The solutions described in Sec. 2 have many remarkable properties, and their near-horizon geometry differs considerably from that of Schwarzschild black hole spacetimes. First, we mention an immediate mathematical consequence of the firewall that we have described. The divergence of $\rho_{\rm A}$ [see Eq. (31)] points to the presence of a matter singularity. As a result, we observe that if an apparent/anti-trapping horizon forms, it is a surface with an intermediate singularity. The appearance of a negative energy firewall is the counterpart to arbitrarily large tidal forces that

could tear apart an observer falling into such a singularity. In these cases the fate of an observer depends on the integrated tidal stress.[8,41]

Consider observers attempting to enter an expanding white hole or exit an evaporating PBH.[10] In the former case, the radial velocity $\dot{R} < 0$ on a geodesic trajectory reverses the sign at some finite $R - r_g$ due to the divergent radial acceleration, and in fact the energy density remains finite. On the other hand, a massive test particle exits a PBH with $\dot{R} = 0$ and the energy density diverges, but its integrated smeared version remains finite.[11]

Since all curvature scalars remain finite, it is instructive to check the Ricci spinors

$$\Phi_{00} = \tfrac{1}{2}R_{11}, \quad \Phi_{22} = \tfrac{1}{2}R_{22}, \quad \Phi_{11} = \tfrac{1}{4}(R_{12} + R_{34}). \tag{40}$$

Using the natural Newman–Penrose tetrad that is built from the two null vectors of Eq. (5) and a pair of complex-conjugate vectors m^μ and $\bar{m}^\mu := m^{\mu *}$,

$$m = \frac{1}{\sqrt{2}r}\partial_\theta + \frac{i}{\sqrt{2}r\sin\theta}\partial_\phi, \qquad m \cdot \bar{m} = 1, \tag{41}$$

we find that the values of all nonzero spinors are finite at the apparent horizon. However, given the freedom of choice of these vectors $l^\mu_{\rm out} \to A l^\mu_{\rm out}$, $l^\mu_{\rm in} \to l^\mu_{\rm in}/A$, the values of the spinors Φ_{00} and Φ_{22} depend on this choice. By choosing $A = f(v, r)$ (this form of the tangent vectors may appear more natural in (t, r) coordinates), we find

$$\Phi_{00} \propto f, \qquad \Phi_{22} \propto f^{-1}, \tag{42}$$

again demonstrating that the apparent/anti-trapping horizon is a surface of intermediate singularity.[11]

3.1. *Black hole formation*

Consider now possibilities for horizon formation. The properties of the self-consistent solutions of the Einstein equations that we have described above lead to the identification of a unique scenario for black hole formation.[29,30] We consider only evaporating PBHs ($r_g' < 0$) due to the implications of the results presented in Sec. 2.1. Working in (v, r) coordinates, we assume that the first marginally trapped surface appears at some $v_{\rm S}$ at $r_+(v_{\rm S})$ that corresponds to a finite value of $t_{\rm S}$. For $v \leqslant v_{\rm S}$, the MS mass in its vicinity can be described by modifying Eq. (22) as

$$C(v, r) = \Delta(v) + r_*(v) + \sum_{i \geqslant 1}^{\infty} w_i(v)(r - r_*)^i, \tag{43}$$

where $r_*(v)$ corresponds to the maximum of $\Delta_v(r) := C(v, r) - r$, and the deficit function $\Delta(v) := C(v, r_*) - r_*(v) \leqslant 0$. At the advanced time $v_{\rm S}$, the location of the maximum corresponds to the first marginally trapped surface, $r_*(v_{\rm S}) = r_+(v_{\rm S})$, and $\sigma(v_{\rm S}) = 0$. For $v \geqslant v_{\rm S}$, the MS mass is described by Eq. (22). For $v \leqslant v_{\rm S}$, the (local) maximum of $\Delta_v(r)$ is determined by $\Delta_v(r)/dr = 0$, hence $w_1(v) - 1 \equiv 0$.

There are no *a priori* restrictions on the evolution of r_* before the PBH is formed. However, since an expanding Schwarzschild sphere describes white hole solutions $r'_+(v_S) < 0$. For $v > v_S$, the maximum of $C(v, r)$ does not coincide with $r_+(v)$ since the trapped region is of finite size. As a result, $w_1(v) < 1$ for $v > v_S$.

This means that at its formation a PBH is described by a $k = 1$ solution, which can be seen from Eq. (30), as $w_1 = 1$ implies $\Upsilon = 0$. It then immediately switches to the $k = 0$ solution, with matching decrease in $w_1(v)$ and increase in $\Upsilon\big(t(v, r_+)\big)$, and $w_1 < 1$ at all subsequent stages. The transition from f^1 to f^0 behavior is continuous as $\Upsilon(t_S) \equiv 0$ for the $k = 1$ solution and it increases thereafter.[30] A detailed discussion can be found in Ref. 10.

3.2. *Surface gravity*

The surface gravity κ plays an important role in GR and semiclassical gravity.[7–9] However, it is unambiguously defined only in stationary spacetimes (for a Schwarzschild black hole $\kappa = (2r_g)^{-1}$), where it is proportional to the Hawking temperature.

Generalizations of surface gravity to dynamical spacetimes are related to two (equivalent in stationary spacetimes) definitions of κ that are based on the peeling off properties and the inaffinity of null geodesics at the horizon, respectively.[9, 42] For sufficiently slowly evolving horizons with properties sufficiently close to their classical counterparts, these different generalizations of surface gravity are practically indistinguishable.[42] This agreement is important. Emission of Hawking-like radiation does not require the formation of an event or even an apparent horizon.[43–46] The role of the Hawking temperature is captured in various derivations either by the peeling[47] or the Kodama[48] surface gravity.

For PBHs, however, these quantities are very different. Here we consider only one version of the peeling surface gravity.[38] In Schwarzschild coordinates $\kappa_{\rm peel}$ is defined from the near-horizon behavior of null geodesics[9, 42] as the linear coefficient in the Taylor expansion of

$$\frac{dr}{dt} = \pm e^h f(t, r) \tag{44}$$

in powers of x as $x \to 0$ (i.e. $r \to r_g$). However, such an expansion is impossible for the metric functions of Eqs. (22)–(23). Alternatively, to be compatible with Eq. (28), $\kappa_{\rm peel}$ should be infinite, as we now demonstrate.

For differentiable C and h, the result is

$$\kappa_{\rm peel} = \frac{e^{h(t,r_g)}\,(1 - C'(t, r_g))}{2r_g}. \tag{45}$$

This quantity is undefined for $k = 0$ solutions as for a radial geodesic [Eq. (28)] it implies

$$\frac{dr}{dt} = \pm|r'_g| + \mathcal{O}(\sqrt{x}), \tag{46}$$

and $\kappa_{\rm peel}$ diverges (for both $k = 0$ and $k = 1$ solutions).[30]

The Kodama vector field has many useful properties of the Killing field to which, modulo possible rescaling, it reduces in the static case.[9,22,42] Similar to the Killing vector, it is most conveniently expressed in (v, r) coordinates,

$$K^\mu = (e^{-h_+}, 0, 0, 0). \tag{47}$$

It is covariantly conserved, and generates the conserved current

$$\nabla_\mu K^\mu = 0, \tag{48}$$

$$\nabla_\mu J^\mu = 0, \qquad J^\mu := G^{\mu\nu} K_\nu, \tag{49}$$

thereby giving a natural geometric meaning to the Schwarzschild coordinate time t. The MS mass is its Noether charge.

Since $K_{(\mu;\nu)} \neq 0$, the generalized Hayward–Kodama surface gravity is defined via[49]

$$\frac{1}{2} K^\mu (\nabla_\mu K_\nu - \nabla_\nu K_\mu) := \kappa_{\rm K} K_\nu, \tag{50}$$

evaluated on the apparent horizon. Hence

$$\kappa_{\rm K} = \frac{1}{2} \left(\frac{C_+(v,r)}{r^2} - \frac{\partial_r C_+(v,r)}{r} \right) \Bigg|_{r=r_+} = \frac{(1-w_1)}{2r_+}, \tag{51}$$

where Eq. (22) was used to obtain the final result. Thus at the formation of a black hole (i.e. of the first trapped surface) this version of surface gravity is zero. At the subsequent evolution stages that correspond to a $k = 0$ solution, $\kappa_{\rm K}$ is nonzero. However, it approaches the static value $\kappa = 1/(4M)$ only if the metric is close to the pure Vaidya metric with $w_1 \equiv 0$. This in turn contradicts the semiclassical results.[38]

4. Discussion

The solutions we have described follow from commonly obtained or assumed, but not always sufficiently articulated requirements: the formation of a minimally regular horizon in finite time of a distant observer. The two classes of spherically symmetric solutions that satisfy these requirements must violate the NEC in the vicinity of the outer apparent horizon, and the matter content is dominated by a null fluid. Unless we accept that semiclassical physics breaks down at the horizon scale, accretion is no longer possible after the horizon has formed: the solutions describe either evaporating black holes or expanding white holes.

It is still not clear how the collapsing matter actually behaves, and how the continuous transition from f^1 to f^0 behavior immediately after PBH formation (Sec. 3.1) can be realized in nature. A mechanism that converts the original matter into the exotic matter present in the vicinity of the forming apparent horizon is required to violate the NEC. However, the emission of collapse-induced radiation[43–45] is a nonviolent process that approaches at latter times the standard Hawking radiation and Page's evaporation law[8,50] $r_g' = -\alpha/r_g^2$, $\alpha \sim 10^{-3} - 10^{-4}$.

In addition, there is no obvious way to reconcile the two standard definitions of surface gravity that underpin different derivations of Hawking radiation to dynamical spacetimes (Sec. 3.2): the peeling surface gravity $\kappa_{\rm peel}$ diverges, whereas the Kodama surface gravity $\kappa_{\rm K}$ is zero at the instant of PBH formation and in contradiction with established semiclassical results thereafter. This has interesting repercussions for the formulation of the information loss paradox[38] and indicates that the Hawking radiation that is obtained on the Schwarzschild background is not an asymptotic form of the radiation that is emitted by a PBH.

Our results indicate that the observed ABHs may be horizonless UCOs rather than genuine PBHs with a horizon — not due to some exotic supporting matter or dramatic variation in the laws of gravity, but simply because the conditions for the formation of a horizon have not been met at the present moment of t. Even if no new physics is required to produce the mandatory NEC violation, the process may be too slow to transform the UCOs that we observe into PBHs. Eq. (28) of Sec. 2.1 sets the timescale of the last stages of infall according to Bob. Assuming that it is applicable through the radial interval of the order of r_g, we find $t_{\rm in} \sim r_g/r_g'$. For an evaporating macroscopic PBH, this is of the same order of magnitude as the Hawking process decay time $t_{\rm evp} \sim 10^3 r_g^3$. Such behavior was found in thin shell collapse models, where the exterior geometry is modeled by a pure outgoing Vaidya metric.[34] For a solar mass black hole, this time is about 10^{67}yr, indicating that it is simply too early for the horizon to form. It is also conceivable that the conditions are not met before evaporation is complete or before effects of quantum gravity become dominant.[8] Horizon avoidance may therefore occur due to the absence of new physics, and not because of it.

A more detailed understanding of the near-horizon geometry as well as the behavior of matter during gravitational collapse and PBH formation is needed to resolve all of the remaining questions. Coming of age of multimessenger astronomy and controversies surrounding the direct observation of ABHs make it a particularly timely project.

Acknowledgments

SM is supported by an International Macquarie University Research Excellence Scholarship and a Sydney Quantum Academy Scholarship. The work of DRT is supported by the ARC Discovery project grant DP210101279.

References

1. L. Barack, V. Cardoso, S. Nissanke, and T. P. Sotiriou (eds.), Class. Quantum Gravity **36**, 143001 (2019).
2. V. Cardoso and P. Pani, Living Rev. Relativ. **22**, 4 (2019).
3. E. Curiel, Nat. Astron. **3**, 27 (2019).
4. V. P. Frolov, J. High Energy Phys. **05**, 049 (2014).
5. M. Visser, Phys. Rev. D **90**, 127502 (2014).

6. V. Cardoso and P. Pani, Nat. Astron. **1**, 586 (2017).
7. S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, England, 1973).
8. V. P. Frolov and I. D. Novikov, *Black Hole Physics: Basic Concepts and New Developments* (Kluwer, Dordrecht, 1998).
9. V. Faraoni, *Cosmological and Black Hole Apparent Horizons* (Springer, Heidelberg, 2015).
10. R. B. Mann, S. Murk, and D. R. Terno, Int. J. Mod. Phys. D **31**, 2230015 (2022).
11. P. K. Dahal, S. Murk, and D. R. Terno, AVS Quantum Sci. **4**, 015606 (2022).
12. S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
13. J. W. York, Jr., Phys. Rev. D **28**, 2929 (1983).
14. R. Carballo-Rubio, F. Di Filippo, S. Liberati, and M. Visser, Phys. Rev. D **101**, 084047 (2020).
15. P. Binétruy, A. Helou and F. Lamy, Phys. Rev. D **98**, 064058 (2018).
16. J. M. M. Senovilla and D. Garfinkle, Class. Quantum Gravity **32**, 124008 (2015).
17. Y. Choquet-Bruhat, *General Relativity and the Einstein Equations* (Oxford University Press, Oxford, England, 2009).
18. N. D. Birrel and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1984).
19. R. M. Wald, Living Rev. Relativ. **4**, 6 (2001).
20. T. A. Roman and P. G. Bergman, Phys. Rev. D **28**, 1265 (1983).
21. S. A. Hayward, Phys. Rev. Lett. **96**, 031103 (2006).
22. G. Abreu and M. Visser, Phys. Rev. D **82**, 044027 (2010).
23. C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).
24. V. Faraoni, G. F. R. Ellis, J. T. Firouzjaee, A. Helou, and I. Musco, Phys. Rev. D **95**, 024008 (2017).
25. R. Penrose, *Structure of space-time*, in C. DeWitt, J. A. Wheeler (eds.), *Batelle Rencontres: 1967 Lectures in Mathematics and Physics*, pp. 121–235 (W. A. Benjamin, San Francisco, 1968).
26. F. J. Tipler, C. J. S. Clarke, and G. R. F. Ellis, *Singularities and horizons: A review article*, in A. Held (ed.), *General relativity and gravitation: One hundred years after the birth of Albert Einstein*, vol. 2, p. 97 (Plenum, New York, 1980).
27. H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt, *Exact Solutions of Einstein's Field Equations*, 2nd edn. (Cambridge University Press, Cambridge, England, 2003).
28. D. R. Terno, Phys. Rev. D **100**, 124025 (2019).
29. D. R. Terno, Phys. Rev. D **101**, 124053 (2020).
30. S. Murk and D. R. Terno, Phys. Rev. D **103**, 064082 (2021).
31. V. Baccetti, R. B. Mann, S. Murk, and D. R. Terno, Phys. Rev. D **99**, 124014 (2019).
32. P. Martín-Moruno and M. Visser, *Classical and Semi-classical Energy Conditions*, in *Wormholes, Warp Drives and Energy Conditions,* edited by F. N. S. Lobo (Springer, New York, 2017), p. 193.
33. E.-A. Kontou and K. Sanders, Class. Quantum Gravity **37**, 193001 (2020).
34. V. Baccetti, S. Murk, and D. R. Terno, Phys. Rev. D **100**, 064054 (2019).
35. W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
36. M. Visser, Phys. Rev. D **56**, 936 (1997).
37. A. Levi and A. Ori, Phys. Rev. Lett. **117**, 231101 (2016).
38. R. B. Mann, S. Murk, and D. R. Terno, Phys. Rev. D **105**, 124032 (2022).
39. C. J. Fewster, *Quantum Energy Inequalities*, in *Wormholes, Warp Drives and Energy Conditions*, edited by F. N. S. Lobo (Springer, New York, 2017), p. 215.

40. E.-A. Kontou and K. D. Olum, Phys. Rev. D **91**, 104005 (2015).
41. G. F. R. Ellis and A. R. King, Commun. Math. Phys. **38**, 119 (1974).
42. B. Cropp, S. Liberati, and M. Visser, Class. Quantum Gravity **30**, 125001 (2013).
43. P. Hájíček, Phys. Rev. D **36**, 1065 (1987).
44. C. Barceló, S. Liberati, S. Sonego, and M. Visser, Class. Quantum Gravity **23**, 5341 (2006).
45. T. Vachaspati, D. Stojkovic, and L. M. Krauss, Phys. Rev. D **76**, 024005 (2007).
46. A. Paranjape and T. Padmanabhan, Phys. Rev. D **80**, 044011 (2009).
47. C. Barceló, S. Liberati, S. Sonego, and M. Visser, Phys. Rev. D **83**, 041501(R) (2011).
48. F. Kurpicz, N. Pinamonti, and R. Verch, Lett. Math. Phys. **111**, 110 (2021).
49. S. A. Hayward, Class. Quantum Gravity **15**, 3147 (1998).
50. D. N. Page, Phys. Rev. D **13**, 198 (1976).

Information recovery from evaporating rotating charged black holes

Zhi-Wei Wang

Department of Physics and Astronomy, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
zhiweiwang.phy@gmail.com

Samuel L. Braunstein

Computer Science, University of York, York YO10 5GH, United Kingdom
sam.braunstein@york.ac.uk

Saurya Das*

Department of Physics and Astronomy, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
saurya.das@uleth.ca

In classical gravity, nothing can escape from a black hole, not even light. In particular, this happens for stationary black holes because their horizons are null. We show, on the other hand, that the apparent horizon and the region near $r = 0$ of an *evaporating* charged, rotating black hole are both timelike. This implies that there exists a channel, via which classical or quantum information can escape to the outside, as the black hole evaporates. Since astrophysical black holes have at least some rotation, our results apply to all black holes in nature. We discuss implications of our result.

Keywords: Black hole information loss; Hawking radiation; evaporating black holes; rotating black holes; Vaidya metric.

1. Introduction

In classical general relativity, a black hole is a region of spacetime from which nothing escapes. This result underwent a change with the discovery of Hawking radiation.[1,2] Studying the evolution of a vacuum quantum state in curved spacetime with a horizon, Hawking showed that the ingoing vacuum state has a probability to yield a non-vacuum outgoing state when the former enters the horizon of a black hole. This means that a black hole can radiate energy to infinity via this quantum mechanism.

Now, both classical and quantum information can enter the horizon of a black hole. However, as Hawking's work showed, only thermal Hawking radiation carrying very little information emerges from it.[3] Therefore, when the black hole evaporates completely, all information carried by the matter that was used to create the black hole is apparently lost forever. This violates unitarity in quantum theory which

*Author for correspondence (saurya.das@uleth.ca)

requires the information of a quantum system be conserved during its evolution. Further study of this apparent information loss paradox led to the idea of black hole complementarity,[4–6] the firewall paradox[7,8] etc.

When discussing the above problem, people often overlook an implicit hypothesis that the causal structure of a black hole is not disturbed by Hawking radiation itself. Parikh and Wilczek were the first to examine this assumption and show that the singularity of an evaporating charged black [white] hole may be accessible for an observer outside when the metric is time-dependent.[9] However, the metric used in their paper in fact seems to represent a white hole, contrary to their claim that it is a black hole (see Appendix A for a detailed analysis). Recently, numerical simulation also shows that the horizon of a charged black hole may be timelike.[10] We know on the other hand, that much like other astronomical objects, all black holes in nature are rotating and uncharged, and the probability of finding a black hole with zero rotation is practically nil.

Motivated by the above, in this work,[11] we show that for a rotating and charged black hole which is Hawking radiating, there is a classical channel through which information can escape, and following the above reasoning, it provides an extended window of information recovery from its interior. In the process, the black hole shrinks of course, but presumably at a faster rate than predicted by Hawking radiation, because of the additional outflow of information and associated matter.

We demonstrate the above rigorously by constructing the Penrose diagram for the above process. In particular, we prove that the region immediately surrounding $r = 0$ is timelike, and the apparent horizon is timelike as well. These two results and the Penrose diagram all together imply null geodesics originating from anywhere near the centre of the black hole to the apparent horizon may emerge to the outside Universe. This in turn provides a route for classical or quantum information to escape from the black hole. This escape of potentially a large amount of information must be taken into account in any attempt to resolve the information loss problem. What is most significant is perhaps the fact that the escaping information need not be thermal.

2. Coordinates for an evaporating rotating charged black hole

Since we are interested in evaporating black holes, we will start with the following time-dependent metric, which represents a rotating and charged Vaidya-type black hole,[12,13] as shown in Appendix B

$$\begin{aligned} ds^2 = &-\left(1-\frac{2Mr-Q^2}{\sigma^2}\right)du^2 + 2du\,dr + \sigma^2 d\theta^2 - 2a\sin^2\theta\,dr\,d\phi \\ &-\frac{2(2Mr-Q^2)a}{\sigma^2}\sin^2\theta\,du\,d\phi + \left(\frac{2Mr-Q^2}{\sigma^2}\,a^2\sin^2\theta + r^2 + a^2\right)\sin^2\theta\,d\phi^2, \end{aligned} \tag{1}$$

where $\sigma^2 \equiv r^2 + a^2 \cos^2\theta$, and $M = M(u)$, $Q = Q(u)$ and $a = a(u)$ denote smooth decreasing functions of the retarded time u.

Next, to construct a two-dimensional Penrose diagram for the above metric, we restrict ourselves to the symmetry axis along $\theta = 0$.[14] Then one gets, from Eq. (1):

$$ds^2 = -\left(1 - \frac{2Mr - Q^2}{r^2 + a^2}\right)du^2 + 2\,du\,dr = -\left[\frac{r^2 - 2Mr + a^2 + Q^2}{r^2 + a^2}\,du - 2\,dr\right]du. \tag{2}$$

We also assume without loss of generality that M, a and Q are proportional to each other and linear functions of u

$$\begin{aligned} M(u) &= M_0 \text{ (constant)}, && u < u_0 && (3)\\ &= \mu u + b \equiv \tilde{u}, && u_0 < u \leq -b/\mu, \mu < 0 && (4)\\ &= 0, && -b/\mu < u && (5)\\ a(u) &= \lambda_1 M(u), && 0 \leq \lambda_1 \leq 1 && (6)\\ Q(u) &= \lambda_2 M(u), && 0 \leq \lambda_2 \leq 1. && (7) \end{aligned}$$

This means that the black hole starts its process of evaporation at time $u = u_0$, and undergoes a mass decrease at a constant rate μ. Contrary to Ref. [13], where the angular momentum was held fixed, in this work we make it time-dependent. This is because on the one hand, the form of the metric continues to hold with that generalization and moreover, the condition $\lambda_1, \lambda_2 \leq 1$ ensures that there are no naked singularities.

The following points may be noted here. First, continuity of the functions $M(u), a(u), Q(u)$ in Eqs. (3-7) guarantees that the various patches of the Penrose diagram that will smoothly join to each other.[9] Second, considering arbitrary decay rates (i.e. not constant over time) is also straightforward. In this case, one can simply break down the non-linear decay function such as $M(u)$ into a series of linear functions of the form $M(u) = \sum_{i=1}^{N}(\mu_i u + b_i)\left[\Theta(u_i + \frac{1}{2}) - \Theta(u_i + \frac{1}{2})\right]$ $(u_0 \leq u_i)$, and similarly for $a(u)$ and $Q(u)$, for $N \gg 1$.

Now, following Ref. [9] and details presented in Ref. [11], we find a set of coordinates which are smooth across the horizon. To this end, we first write metric (2) in the so-called 'double null' form:

$$ds^2 = -\frac{g(\tilde{u}, r)}{\mu} d\tilde{u} dv\ , \tag{8}$$

where,

$$dv \equiv \frac{1}{g(\tilde{u}, r)}\left[\left(\frac{r^2 - 2Mr + (\lambda_1^2 + \lambda_2^2)M^2}{r^2 + {\lambda_1}^2 M^2}\right)\frac{d\tilde{u}}{\mu} - 2dr\right]. \tag{9}$$

with the integrating factor

$$g(\tilde{u}, r) = \left(\frac{r^2 - 2Mr + (\lambda_1^2 + \lambda_2^2)M^2}{r^2 + {\lambda_1}^2 M^2}\right)\frac{\tilde{u}}{\mu} - 2r\ . \tag{10}$$

We may then read off

$$\frac{\partial v}{\partial r} = \frac{r^2 + \lambda_1^2 \tilde{u}^2}{-r[r^2 + \lambda_1^2 \tilde{u}^2] + \frac{\tilde{u}}{2\mu}\left[r^2 - 2\tilde{u}r + (\lambda_1^2 + \lambda_2^2)\tilde{u}^2\right]} \tag{11}$$

$$\frac{\partial v}{\partial \tilde{u}} = \frac{\frac{1}{2a}(r^2 - 2Mr + (\lambda_1^2 + \lambda_2^2)\tilde{u}^2)}{-r[r^2 + \lambda_1^2 \tilde{u}^2] + \frac{\tilde{u}}{2\mu}\left[r^2 - 2\tilde{u}r + (\lambda_1^2 + \lambda_2^2)\tilde{u}^2\right]} \ . \tag{12}$$

As one can see, the common denominators $D(r)$ of the above expressions are cubic in r, which may be rewritten as

$$\begin{aligned} D(r) &= -r[r^2 + \lambda_1^2 \tilde{u}^2] + \frac{\tilde{u}}{2\mu}\left[r^2 - 2\tilde{u}r + (\lambda_1^2 + \lambda_2^2)\tilde{u}^2\right] \\ &= -r^3 - \lambda_1^2 \tilde{u}^2 r + \frac{\tilde{u}}{2\mu} r^2 - \frac{\tilde{u}^2}{\mu} r + \frac{\tilde{u}^3}{2\mu}(\lambda_1^2 + \lambda_2^2) \\ &= -r^3 + \frac{\tilde{u}}{2\mu} r^2 - (\lambda_1^2 \tilde{u}^2 + \frac{\tilde{u}^2}{\mu}) r + \frac{\tilde{u}^3}{2\mu}(\lambda_1^2 + \lambda_2^2). \end{aligned} \tag{13}$$

Since $D(-\infty) > 0$, $D(+\infty) < 0$, and $D(0) = \frac{\tilde{u}^3}{2\mu}(\lambda_1^2 + \lambda_2^2) < 0$ because of $\mu < 0$ and $\tilde{u} > 0$ ($\tilde{u} = 0$ represents a flat spacetime and hence is ignored here), the scenarios with positive roots for the cubic equation $D(r) = 0$ are well described by Fig. 1. No other scenario has a positive root.

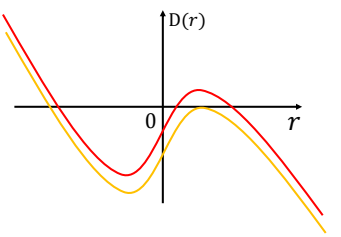


Fig. 1. Scenarios caused singularities for Eqs. (11) and (12).

For the case represented by the red line in Fig. 1, the denominators have three roots r_1, r_2, r_3, (with $r_1 > r_2 > 0 > r_3$) where v has coordinate singularities, and

$$v = \sum_{i=1}^{3} A_i \ln(r - r_i) \ , \tag{14}$$

with

$$\begin{aligned} A_1 &= \frac{r_1^2}{(r_1 - r^2)(r_1 - r_3)} + \lambda_1 \tilde{u}^2, \qquad A_2 = \frac{-r_2^2}{(r_1 - r^2)(r_2 - r_3)} + \lambda_1 \tilde{u}^2, \\ A_3 &= \frac{r_3^2}{(r_1 - r_3)(r_2 - r_3)} + \lambda_1 \tilde{u}^2 \ . \end{aligned} \tag{15}$$

Then the complete set of singularity free coordinates for the two patches may be defined as:

$$\begin{aligned} V_2(v) &\equiv e^{v/A_1} \\ &= (r-r_1)(r-r_2)^{A_2/A_1}(r-r_3)^{A_3/A_1} \ , \qquad r_2 \le r < \infty \qquad (16) \\ V_1(v) &\equiv k_1 - (-V_2)^{A_1/A_2} \\ &= k_2 + (r_1-r)^{A_1/A_2}(r_2-r)(r-r_3)^{A_3/A_2} \ , \qquad 0 \le r < r_2 \qquad (17) \end{aligned}$$

The constants k_i are determined by matching V_2 with V_1 in $r_2 < r < r_1$. If the denominator $D(r)$ of Eq. (11) is described by the gold line in Fig. 1, it has a negative root r_3 and a positive double root r_1 such that it can be factored as $(r-r_3)(r-r_1)^2$. This is nothing but a special case that $r_1 = r_2$ in the red-line scenario in Fig. 1.

3. Timelike regions

With the results in last section, it is easy to show that $r = 0$ and its immediate neighbourhood is *timelike*. For example, in the case of Eqs. (14-17) above,

$$\begin{aligned} ds^2 &= -\frac{g(\tilde{u},r)A}{V(\tilde{u},r)\mu}\, d\tilde{u}\, dV \qquad (18) \\ &\xrightarrow{r\to 0} -\left(\frac{\lambda_1^2+\lambda_2^2}{\lambda_1^2}\right) du^2 < 0 \ . \qquad (19) \end{aligned}$$

It may be noted that the 'ring singularity' of the metric is at $r = 0$ *and* $\theta = \pi/2$. It follows that except for that singular point, the region in the neighbourhood of $r = 0$ is regular and of finite curvature, except perhaps at the end of the evaporation process.

We now consider the apparent horizon and show that it is timelike as well. We follow the procedure of Ref. [15], and define the function:

$$f(u,r) = r - R_\pm \ , \qquad (20)$$

where $R_\pm$ represents the outer and inner apparent horizons of the evaporating rotating charged black hole. A rigorous calculations for $R_\pm$ yields

$$R_\pm = r_\pm + \frac{a\lambda_1(r_\pm^2+a^2)}{r_\pm^2 - (M+2a\mu\lambda_1)r_\pm}\mu + O(\mu^2), \qquad (21)$$

where

$$r_\pm = M \pm (M^2 - Q^2 - a^2)^{1/2} \qquad (22)$$

(see Appendix B for more details).

We assume that the evaporating process is sufficiently slow and that $|\mu|$ is sufficiently small. Therefore, we focus on the leading term in Eq. (21) for our future calculations. Thus, Eq. (20) reduces to

$$f(u,r) = r - M \mp (M^2 - Q^2 - a^2)^{1/2} \ , \qquad (23)$$

such that the apparent horizons are at

$$f(u,r) = 0, \ r_\pm = M \pm (M^2 - Q^2 - a^2)^{1/2} \ . \qquad (24)$$

Now since M, Q and a are functions of u the vector n_a, normal to the surface $f(u,r)=0$, will have components given by

$$n_u = f_{,u} = -\mu \mp \frac{\mu\ (M - \lambda_1 a - \lambda_2 Q)}{(M^2 - Q^2 - a^2)^{1/2}}\ , \qquad n_r = f_{,r} = 1\ , \tag{25}$$

and norm

$$n^2 = g^{ab} n_a n_b = 2\mu[-1 \mp (1 - \lambda_1^2 - \lambda_2^2)^{1/2}] > 0\ , \tag{26}$$

where we have used $\mu < 0$ from Eq. (4). Therefore, n_a is spacelike and the apparent horizon is timelike.

4. Penrose diagram

The information loss problem may be easily described using Penrose diagrams. The Penrose diagram in Fig. 2 depicts a static Schwarzschild (non-rotating, uncharged) black hole. In this case, the singularity is spacelike, and it is clear that information, from within the horizon, propagating along null or timelike geodesics cannot reach the outside Universe. The Penrose diagram in Fig. 3 describes an evaporating

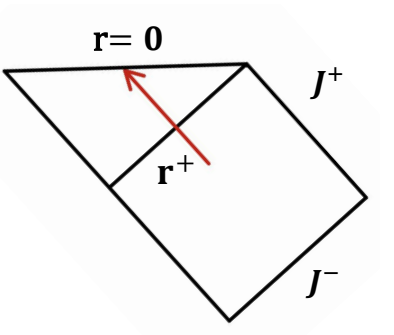


Fig. 2. Penrose diagram of a static Schwarzschild black hole.

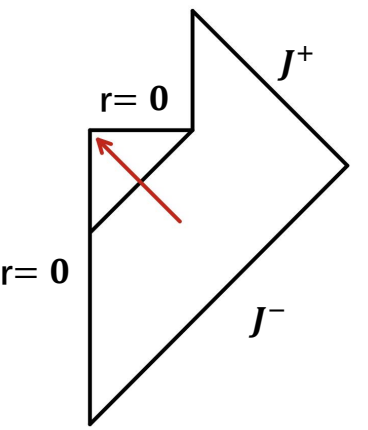


Fig. 3. Penrose diagram of an evaporating Schwarzschild black hole.

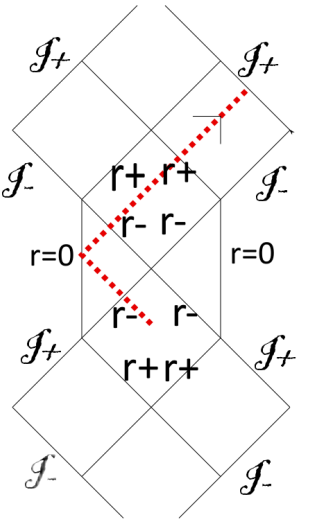


Fig. 4. Penrose diagram for stationary Kerr black hole.

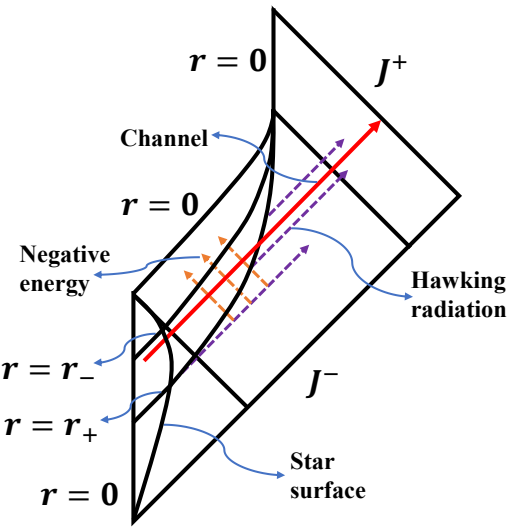


Fig. 5. Penrose diagram for an evaporating rotating (charged) black hole. This Penrose diagram describes a rotating (charged) star first collapsing into a rotating (charged) black hole. Then the black hole begins to evaporate, and both its inner and outer horizons at $\theta = 0$ follow timelike trajectories. After the black hole completely evaporates away, the spacetime is connected with a flat spacetime.

Schwarzschild black hole. The evaporating procedure of a Schwarzschild black hole does not change the fact that the horizon is along a null direction and hence does not improve the situation.

Similarly, Fig. 4 is the Penrose diagram of a maximally extended rotating black hole whose singularity is timelike. Although information propagating along null rays in such a Penrose diagram can exit a future horizon, it does so only by entering another Universe. Therefore, the information loss problem is not resolved in the present Universe.

Based on the results of Eqs. (19), (21), and (26), we can draw the Penrose diagram for the evaporating charged, rotating black hole (Fig. 5). As one can see, there is a channel (outgoing null rays in red) which can carry classical as well as quantum information all the way from the interior to infinity, crossing the timelike horizon along the way.

We assume $\theta = 0$ in Eq. (2) to simplify our calculations, it does not restrict our conclusion that there exists at least one such outgoing information channel. Even if future work shows that this channel is unique because its special azimuth, there is no limit in principle to the amount of information that can exit through this channel. Since our major calculations are carried out for a specific azimuth $\theta = 0$, its generalization to black holes in higher-dimensional spacetimes should be straightforward.

5. Discussion

We have shown in this article that the apparent horizon and the region close to $r = 0$ of an evaporating charged rotating black hole follow timelike trajectories. This implies that there exists a channel through which classical or quantum information can escape. Since proposals such as the fuzzball and firewall deal purely with quantum information, and in the context of Hawking radiation, there is no contradiction between those and our results here.[7,8,19] On the contrary, our approach complements the others.

It is known that matter flowing into the dynamic charged and rotating black hole may cause instability due to mass-inflation, resulting in the nature of the singularity changing.[20,21] However, we do not anticipate such a problem in our case since (a) our results pertain to the region of low curvature (not singularity), and (b) we have outgoing as opposed to infalling matter. That said a careful study of the effect for the spacetime under consideration is warranted and we hope to report on this in the future.

Since the horizon of a stationary black hole is null and the evaporating procedure is a quantum effect, emergence of such a classical channel is not a pure classical phenomenon. If the evaporation process stops for any reason before the black hole evaporates completely, our Fig. 5 would change and revert back to Fig. 4. However, since the Hawking temperature remains non-zero throughout, we do not consider this possibility.

Although our work does not completely resolve the problem of information loss, it complements other approaches in opening up a new information recovery channel. More work needs to be done to determine the extent of information that can be

extracted via this channel, and its interplay with the quantum resolution of the singularity.

Appendix A. The metric in Parikh and Wilczek's paper represents a white hole

To our knowledge, such an analysis was first done by in[9] for non-rotating black holes. However, contrary to their assuming the metric that they used to be that of an evaporating black hole, it is in fact that of a charged white hole. They study the metric[9]

$$ds^2 = -\left(1 - \frac{2Mr - Q^2}{r^2}\right)du^2 - 2du\,dr + r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{A.1}$$

where $M = M(u)$ and $Q = Q(u)$ denote smooth decreasing functions of the retarded time u. It is easy to see that Eq. (A.1) represents a white hole by taking $Q = 0$.[22] To show that Eq. (A.1) is the metric of a white hole, we now calculate the expansion of the null normal vector of this metric.

Since this metric is spherically symmetric, it is easy to figure out that the outward null normal vector l^μ and the inward null normal vector n^μ for this metric are

$$l^\mu = (0, 1, 0, 0) \quad \text{and} \quad n^\mu = \left(1, -\frac{r^2 - 2Mr + Q^2}{2r^2}, 0, 0\right). \tag{A.2}$$

It is easy to check that

$$l^\mu l_\mu = 0, \quad n^\mu n_\mu = 0 \quad \text{and} \quad l^\mu n_\mu = -1. \tag{A.3}$$

Finally, we calculate the expansion for both l^μ and n^μ by $\theta^{(A)} = A_{\mu;\nu}\sigma^{\mu\nu}$, where $\sigma^{\mu\nu} = g^{\mu\nu} + l^\mu n^\nu + n^\mu l^\nu$, to obtain

$$\theta^{(l)} = \frac{2}{r} \quad \text{and} \quad \theta^{(n)} = -\frac{r^2 - 2Mr + Q^2}{r^3}. \tag{A.4}$$

From Eq. (A.4) we know that $\theta^{(l)} > 0$ and $\theta^{(n)} < 0$ when r is larger than the apparent horizon. However, it is the vanishing of $\theta^{(n)}$ that defines an apparent horizon. Thus, the Vaidya-type metric in Parikh and Wilczek's paper represents a white hole.

Appendix B. The metric in our manuscript represents a black hole

To show that our metric represents an evaporating black hole, we calculate the expansion of the null normal vectors for the rotating charged Vaidya-type black hole, see Eq. (1) in the manuscript.

Suggested by the rotating charged stationary black hole,[23] we may conjecture that the outward null vector l^μ and the inward null vector n^μ for the rotating charged Vaidya-type black hole are

$$l^\mu = \left(\frac{r^2 + a^2}{\sigma^2}, \frac{\triangle}{2\sigma^2}, 0, \frac{a}{\sigma^2}\right) \quad \text{and} \quad n^\mu = (0, -1, 0, 0). \tag{B.1}$$

where $\sigma^2 = r^2 + a^2\cos^2\theta$ and $\triangle = r^2 + a^2 - 2Mr + Q^2$. It is easy to check that

$$l^\mu l_\mu = 0, \quad n^\mu n_\mu = 0 \quad \text{and} \quad l^\mu n_\mu = -1 \tag{B.2}$$

for the Vaidya-type metric in our manuscript. Then we calculate the expansion for both l^μ and n^μ by $\theta^{(A)} = A_{\mu;\nu}\sigma^{\mu\nu}$, where $\sigma^{\mu\nu} = g^{\mu\nu} + l^\mu n^\nu + n^\mu l^\nu$. Finally, we obtain

$$\theta^{(l)} = \frac{r\triangle + 2\,a\,a'(u)(r^2 + a^2\cos^2(\theta))}{(r^2 + a^2\cos^2(\theta))^2} \quad \text{and} \quad \theta^{(n)} = -\frac{2r}{r^2 + a^2\cos^2(\theta)}. \tag{B.3}$$

From Eq. (B.3) we see that $\theta^{(l)} > 0$ and $\theta^{(n)} < 0$ when r is larger than the apparent horizon and $\theta^{(l)} = 0$ defines the apparent horizon. Thus, the Vaidya-type metric in our manuscript represents a black hole.

The apparent horizon for a stationary rotating charged black hole is defined by $\triangle = 0$. Here, our calculations show that the time dependence of the angular momentum $a(u)$ makes the apparent horizon shift slightly from the stationary case with the size of the shift depending on the rate of change of $a(u)$ with respect to u.

If we insert our linear functions for M, Q and a, Eqs. (3-6) in the manuscript, into the expansion of the outward null normal vector and suppose the rate of change is very slow (μ is very small), we can obtain the corrected apparent horizon for $\theta = 0$ as

$$R_\pm = r_\pm + \frac{a\lambda_1(r_\pm^2 + a^2)}{r_\pm^2 - (M + 2a\mu\lambda_1)r_\pm}\mu + O(\mu^2). \tag{B.4}$$

Since we have proved $r_\pm$ is timelike, the apparent horizon of a slowly evaporating rotating charged black hole should also be timelike.

References

1. S. W. Hawking, Black hole explosions? *Nature* **248**, 30–31 (1974).
2. S. W. Hawking, Particle creation by black holes, *Communications in Mathematical Physics* **43**, 199–220 (1975).
3. S. W. Hawking, Breakdown of predictability in gravitational collapse, *Physical Review D* **14(10)**: 246 (1976).
4. G. 't Hooft, On the quantum structure of a black hole, *Nuclear Physics B* **256**: 727-745 (1985).
5. G. 't Hooft, The black hole interpretation of string theory, *Nuclear Physics B* **335**: 138–154 (1990).
6. L. Susskind, L. Thorlacius, J. Uglum, The stretched horizon and black hole complementarity, *Physical Review D* **48**: 3743 (1993).
7. S. L. Braunstein, Black hole entropy as entropy of entanglement, or it's curtains for the equivalence principle, *arXiv preprint* arXiv:0907.1190 (2009).
8. A. Almheiri, D. Marolf, J. Polchinski, J. Sully, Black holes: complementarity or firewalls?, *Journal of High Energy Physics* **2013(2)**: 1-20 (2013).
9. M. Parikh, F. Wilczek, Global structure of evaporating black holes. *Physics Letters B* **449** (1999) 24-29.

10. J. C. Schindler, A. Aguirre, A. Kuttner, Understanding black hole evaporation using explicitly computed Penrose diagrams, *Physical Review D* **101**: 024010 (2020).
11. S. L. Braunstein, S. Das, Z.-W. Wang, Information recovery from evaporating black holes, *International Journal of Modern Physics D* 2150069 (2021).
12. P. C. Vaidya, Curr. Sci. **12**, 183 (1943); Proc. Ind. Ac. Sci. A **33**, 264 (1951); Nature **171**, 260 (1953).
13. J. Jing and Y. Wang, Int. J. Theor. Phys. **35**, 1481 (1996).
14. B. Carter, Phys. Rev. **141**, 1242 (1966).
15. Y. Kaminaga, Class. Quant. Grav. **7**, 1135 (1990).
16. L. Susskind, *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe*, World Scientific (2004).
17. A. Ashtekar, J. Phys.: Conf. Ser. **189** 012003 (2009).
18. S. Das, Phys. Rev. **D89** (2014) 084068 [arXiv:1311.6539].
19. S. D. Mathur, Fortsch.Phys. 53 (2005) 793-827 [arXiv:hep-th/0502050].
20. C. Barrabès, W. Israel, E. Poisson, Class. Quant. Grav. **7**, L273–L278 (1990).
21. D. Rossi, *Reissner-Nordström Black Holes and Mass Inflation*, thesis, University of Bologna (2019).
22. E. Poisson, *A Relativist's Toolkit The Mathematics of Black-Hole Mechanics*, Cambridge University Press (2004) p.167.
23. E. T. Newman, R. Penrose, *Spin-coefficient formalism*, Scholarpedia, **4**(6):7445 (2009).

Black hole thermodynamics from entanglement mechanics

S. Mahesh Chandran* and S. Shankaranarayanan†

Dept. of Physics, IIT Bombay,
Mumbai, Maharashtra-400076, India
**E-mail: maheshchandran@iitb.ac.in*
†*E-mail: shanki@iitb.ac.in*

Since its inception, the Bekenstein-Hawking area relation for black-hole entropy has been the primary testing ground for various theories of quantum gravity. However, a key challenge to such theories is identifying the microscopic structures and explaining the exponential growth of microstates, providing a fundamental understanding of thermodynamic quantities. Since entropy is a single number, we explore other quantities to provide complete information about the black-hole microstates. We establish a one-to-one correspondence between entanglement energy, entropy, and temperature (quantum entanglement mechanics) and the Komar energy, Bekenstein-Hawking entropy, and Hawking temperature of the horizon (black-hole thermodynamics), respectively. We also show that this correspondence leads to the Komar relation and Smarr formula for generic 4-D spherically symmetric space-times. While offering an independent derivation of black-hole thermodynamics from field observables, the universality of results suggests that quantum entanglement is a *fundamental building block of space-time*.

Keywords: Entanglement Entropy; Black Hole Thermodynamics; Field Theory in Curved Space-times.

1. Introduction

Black holes are fascinating entities that typically arise from gravitationally collapsing bodies, such as stars at the end of their life cycle or star collisions. Despite their violent origin, black holes relax to a stationary state, which can then be fully described by a mere handful of variables such as their mass, charge and angular momentum. This makes it easier to probe various quantum and gravitational phenomena that come with it, further making it an ideal testing ground for ongoing research on quantum gravity.

Interestingly, like ordinary matter systems such as ideal gases, black holes also obey an equation of state. The physical parameters describing stationary black-holes satisfy what is known as the Komar relation[1–3]:

$$E_{\rm Komar} = 2\,T_{\rm H}\,S_{\rm BH}, \tag{1}$$

where $E_{\rm Komar}$ is related to the Hamiltonian of the Einstein-Hilbert action,[4] $T_{\rm H}$ is the Hawking temperature[5] and $S_{\rm BH}$ is the Bekenstein-Hawking entropy.[6] The above relation is a by-product of the Smarr formula[7,8]:

$$M = 2T_H\,S_{\rm BH} + 2\Omega_H J + \Phi_H Q \tag{2}$$

that relates the mass (M), angular momentum (J), entropy ($S_{\rm BH}$), electric charge (Q) of black-holes, angular velocity Ω_H, and Φ_H which is the potential difference between the horizon and infinity. It was shown that the Smarr formula, in its differential form and subject to certain assumptions, gives rise to the first law of black hole thermodynamics.[9] At the heart of this law is the analogy that connects thermodynamic entropy to surface area, and temperature to surface gravity.

Most of the effort in literature has been to understand the microscopic statistical origin of black-hole entropy,[10–13] and in extension, black hole thermodynamics. However, black-hole entropy has the problem of Universality[14]— wherein up to the leading order, several approaches using completely different microscopic degrees of freedom lead to Bekenstein-Hawking entropy.[10,13,14] The inability to conclusively distinguish such approaches, along with the fact that entropy is a single number, suggests that it may be impossible to identify the true degrees of freedom that give rise to black hole entropy. Therefore, it requires us to also identify other physical quantities corresponding to black hole thermodynamics, if we were to attempt resolving the universality problem.

Over the last decade or so, one such approach that has grown increasingly crucial to understanding black-hole physics, and quantum gravity in general, is quantum entanglement.[15–19] A fundamental feature of entanglement is the *Area law* — the entanglement entropy of blocks of low energy states of local Hamiltonian is often proportional to the measure of the boundary separating the block from its setting.[13,20–22] The area law has not only established a direct link between entanglement entropy and black hole entropy, but also led to the conjecture that space-time fabric might itself be built upon entanglement.[17] If entanglement is a necessity for the existence of space-time, then the *litmus-test* is to derive all the quantities of black-hole thermodynamics from entanglement.

In this talk, we show that the quantum scalar fields in a $(3+1)$-dimensional black-hole space-times with one or more horizons, provide a way to obtain the physical quantities from entanglement *aka entanglement mechanics*. We relate these quantities to black-hole thermodynamics, and further show that these quantities satisfy Smarr formula (2).[23]

2. Modeling scalar field in a background space-time

In this section, we recapitulate the procedure for probing entanglement mechanics of a minimally coupled scalar field against a static, spherically symmetric background space-time with one or more horizons. The line-element for such a space-time can in general be written as follows:

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2 d\Omega^2. \tag{3}$$

Depending on the form of $f(r)$, there is usually a handful of coordinate settings that help us understand the system better. Suppose the space-time in question has

a horizon (r_h), such as in the Schwarzschild case, it is useful to rewrite the metric in terms of proper-length coordinates[24]:

$$ds^2 = -f(r)dt^2 + d\rho^2 + r^2 d\Omega^2\,; \qquad \rho = \int_{r_h}^{r} \frac{dr}{\sqrt{f(r)}}. \tag{4}$$

Before modeling the scalar field in this setting, we have to address two fundamental aspects — i) Since entanglement entropy diverges in the continuum limit, the field has to be *regularized*. The regularisation provided by the more rigorous definition of quantum fields in terms of operator-valued distributions is difficult to calculate even for simple free theories.[25] We therefore resort to regularisation by discretization,[21,22] i.e., we treat quantum fields on a lattice, with the lattice spacing a fixed in units of proper length as depicted in Fig 1. (ii) To capture entanglement of the field across a space-time horizon, the field must be treated as a *bipartite* system made up of subsystems on either sides of the horizon. However, the proper length co-ordinates only capture the region on that side of the horizon where $f(r) > 0$. The physics in the region $f(r) < 0$ will hence remain inaccessible. The bipartition will then have to be performed as close to the horizon as possible, i.e., by tracing out a single degree of freedom along proper length co-ordinate. This is in fact a valid approximation for large horizon radius and/or small lattice spacing of discretized field.[23]

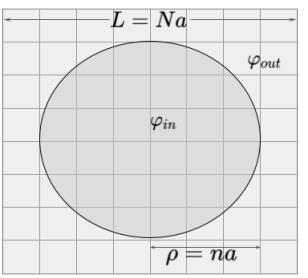


Fig. 1. Discretization scheme for a scalar field in proper length co-ordinates. The field essentially is replaced by a network of coupled harmonic oscillators placed at every grid point.

To set the model up, we begin with the action for a massive scalar field in an arbitrary space-time[26,27]:

$$S = \frac{1}{2}\int d^4x\sqrt{-g}\left[g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi - m_f^2\varphi^2\right]. \tag{5}$$

For the proper-length coordinate (4), we use the following spherical decomposition of the scalar field with appropriate scaling:

$$\dot{\varphi}(\rho,\theta,\phi) = \frac{f^{1/4}(r)}{r}\sum_{lm}\dot{\varphi}_{lm}(\rho)Z_{lm}(\theta,\phi) \tag{6}$$

$$\varphi(\rho,\theta,\phi)=\frac{f^{1/4}(r)}{r}\sum_{lm}\varphi_{lm}(\rho)Z_{lm}(\theta,\phi). \tag{7}$$

Substituting these in the action (5), leads to the following effective (1 + 1)-D Lagrangian:

$$L=\frac{1}{2}\sum_{lm}\int d\rho\left[\dot{\varphi}_{lm}^2-r^2\sqrt{f(r)}\left\{\partial_\rho\left(f^{1/4}(r)\frac{\varphi_{lm}}{r}\right)\right\}^2-f(r)\left\{m_f^2+\frac{l(l+1)}{r^2}\right\}\varphi_{lm}^2\right] \tag{8}$$

With the help of canonical conjugate momenta defined as $\pi_{lm}=\dot{\varphi}_{lm}$, we can now write down the Hamiltonian of the system. To regularize this Hamiltonian, we introduce lattice spacing a in the proper length co-ordinate as $\rho=ja$. The IR cut-off here is on the proper length, which is fixed to be $\rho_L=(N+1)a$. For each lattice point j, we obtain the corresponding lattice point in rescaled radial co-ordinate $r'=r/a$, by inverting the following expression for r_j:

$$j=\int_{\Delta_h}^{r_j}\frac{dr'}{\sqrt{f(r')}}. \tag{9}$$

where we have introduced the dimensionless parameters $\Delta_h=r_h/a$ and $r_j=r/a|_{\rho=ja}$. For convenience, we further define $f_j=f(r)|_{\rho=ja}$. It should be noted that the lattice points in radial co-ordinate $\{r_j\}$ are not equally spaced. On employing the midpoint discretization scheme,[13] we obtain a fully regularized Hamiltonian:

$$H=\frac{1}{2a}\sum_{lmj}\left[\pi_{lm,j}^2+r_{j+\frac{1}{2}}^2f_{j+\frac{1}{2}}^{1/2}\left\{f_j^{1/4}\frac{\varphi_{lm,j}}{r_j}-f_{j+1}^{1/4}\frac{\varphi_{lm,j+1}}{r_{j+1}}\right\}^2+f_j\left\{\Lambda^2+\frac{l(l+1)}{r_j^2}\right\}\varphi_{lm,j}^2\right], \tag{10}$$

where $\Lambda=a^2m_f^2$. Let us now factorize the Hamiltonian as $H=\tilde{H}/a$ and consider the following scaling transformations:

$$a\to\eta a;\quad m_f\to\eta^{-1}m_f;\quad r_h\to\eta r_h \tag{11}$$

Under these transformations, the parameters Λ and Δ_h remain invariant. The Hamiltonian H has therefore been factorized into a scale-dependent part $1/a$ and a scale-independent part $\tilde{H}$. Since the relations between entanglement measures of H and $\tilde{H}$ are well established,[23] it is sufficient to work with the scale-invariant part $\tilde{H}$. It can be seen that $\tilde{H}=\sum_{lm}\tilde{H}_{lm}$ exactly resembles a network of coupled harmonic oscillators:

$$\tilde{H}_{lm}=\frac{1}{2}\left[\sum_i\pi_{lm,i}^2+\sum_{ij}\varphi_{lm,i}K_{ij}\varphi_{lm,j}\right] \tag{12}$$

Here, K_{ij} is the coupling matrix, which contains all the relevant information about the interactions, and in which all information about entanglement entropy is encoded. To simplify further, we will focus on the massless case ($\Lambda = 0$), and also impose the Dirichlet boundary condition $\varphi_{lm,N+1} = 0$ to obtain a non-divergent scaling behavior. The coupling matrix will therefore have the following non-zero elements:

$$\begin{aligned} K_{11} =& f_1 \frac{l(l+1)}{r_1^2} + \frac{r_{3/2}^2}{r_1^2}\sqrt{f_1 f_{3/2}} \\ K_{jj\neq 1} =& f_j \frac{l(l+1)}{r_j^2} + \frac{\sqrt{f_j}}{r_j^2}\left\{ r_{j+\frac{1}{2}}^2 \sqrt{f_{j+\frac{1}{2}}} + r_{j-\frac{1}{2}}^2 \sqrt{f_{j-\frac{1}{2}}} \right\} \\ K_{j,j+1} =& K_{j+1,j} = -\frac{r_{j+\frac{1}{2}}^2}{r_j r_{j+1}} \left\{ f_{j+\frac{1}{2}}^2 f_j f_{j+1} \right\}^{1/4} \end{aligned} \tag{13}$$

Having fully described the model, we now introduce quantities associated with entanglement mechanics, namely, the entanglement entropy and entanglement energy, of a quantum field in a 4-D spherically symmetric space-time with at least one horizon[23,28]:

$$S_{\text{ent}} = -\operatorname{Tr} \rho_{red} \log \rho_{red} \quad ; \quad E_{\text{ent}} = \epsilon \operatorname{Tr} \left[\rho : H_{in} : \right], \tag{14}$$

where $: H_{in} :$ is the normal-ordered Hamiltonian corresponding to the reduced subsystem, ρ is the density matrix, and ρ_{red} is the reduced density matrix obtained by tracing over the field outside the horizon. The constant prefactor ϵ in the definition of entanglement energy accounts for the fact that it is not a unique measure;[23] it may be obtained *uniquely* by comparing with physical quantities of black-holes. While entanglement entropy (S_{ent}) essentially captures change in information content in the presence of the horizon, we may define entanglement energy (E_{ent}) as the disturbed vacuum energy or the *correlation energy* in the presence of the horizon. We further invoke the Komar relation(1) to define entanglement temperature as $T_{ent} = E_{ent}/2S_{ent}$.[23] Having formulated the quantities that describe entanglement mechanics for a given field, we proceed to simulate them in various space-times to infer its fundamental structure.

3. Entanglement mechanics of the field near space-time horizon

In this section, we discuss results from numerical simulations of entanglement mechanics in a variety of static, spherically symmetric space-times with one or more horizon(s).

3.1. *Schwarzschild Black Hole*

In the Schwarzschild space-time, $f(r) = 1 - r_h/r$, and the proper length (4) takes the form:

$$\rho = r\sqrt{1 - \frac{r_h}{r}} + \frac{r_h}{2} \ln \left[\frac{r}{r_h} \left\{ 1 + \sqrt{1 - \frac{r_h}{r}} \right\}^2 \right]. \tag{15}$$

where the horizon radius is $r_h = 2M$. On discretizing $\rho = ja$, we get a scale-invariant expression that connects lattice-points in the proper length and radial co-ordinates as follows:

$$j = r_j\sqrt{1-\frac{\Delta_h}{r_j}} + \frac{\Delta_h}{2}\ln\left[\frac{r_j}{\Delta_h}\left\{1+\sqrt{1-\frac{\Delta_h}{r_j}}\right\}^2\right], \tag{16}$$

where $\Delta_h = 2M/a$ and $r_j = r/a$ are dimensionless. We also see that $f_j = 1 - \Delta_h/r_j$. This confirms that the Hamiltonian in (10) is characterized by dimensionless parameters Λ and Δ_h, and is therefore invariant under the transformations:

$$a \to \eta a; \quad m_f \to \eta^{-1} m_f; \quad M \to \eta M \tag{17}$$

Now we focus on the scale-invariant Hamiltonian $\tilde{H}$, wherein we vary the rescaled horizon Δ_h, and assume that the entanglement mechanics of the field at the horizon can be approximated by tracing out the closest oscillator near the horizon. This approximation is reasonable for large values of Δ_h, wherein the radial distance of the closest oscillator from horizon is negligible ($r_1 \sim \Delta_h$).

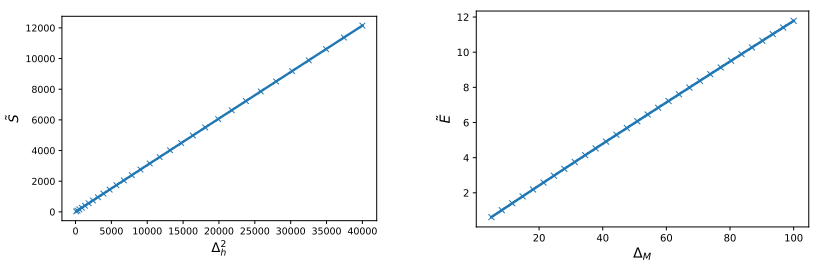
From Fig 2, we observe the following scaling relations:

$$\tilde{S} = c_s \Delta_h^2; \quad \tilde{E} = c_e \Delta_M, \tag{18}$$

where a linear fit fixes the values $c_s \sim 0.3$ and $c_e \sim 0.06$. For the original Hamiltonian H, we then have:

$$S = c_s \frac{r_h^2}{a^2}; \quad E = c_e \frac{M}{a^2}; \quad T = \frac{c_e}{4c_s r_h} = \frac{\pi c_e}{c_s} T_H \tag{19}$$

where T_H is the Hawking temperature. From the above relations, we see that (i) entanglement entropy exhibits an area law whereas entanglement energy scales linearly with horizon radius—the latter is fundamentally different from an area-law scaling observed in Minkowski space-time, (ii) entanglement temperature is independent of the UV cut-off a, and (iii) The entanglement mechanics follows the same laws of black-hole mechanics.[10] In the following sections, we will see if these observations extend for other space-times as well.



(a) (b)

Fig. 2. Entanglement Mechanics for Schwarzschild Black Hole.

3.2. *Reissner-Nordström*

We will now probe entanglement mechanics of the field in an asymptotically flat space-time with multiple horizons. The line-element for Reissner-Nordström black hole is given by (4), where

$$f(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}\,.$$

For $Q < M$, the roots are given by $r_\pm = M \pm \sqrt{M^2 - Q^2}$ where r_+ corresponds to the event-horizon and r_- refers to the internal Cauchy horizon. Thus, $f(r)$ is positive in two regions: 1. $0 < r < r_-$ and 2. $r_+ < r < \infty$.

3.2.1. *Cauchy horizon*

In terms of the dimensionless variable (χ), the Cauchy horizon is

$$r_- = Q\{\chi - \sqrt{\chi^2 - 1}\} \quad \text{where} \quad \chi = M/Q \in (1, \infty)\,.$$

To ensure that the proper length is positive definite quantity, we reverse the limits of integration in Eq. (4), i.e.,

$$\rho = \int_r^{r_h} \frac{dr}{\sqrt{f(r)}} = -\sqrt{Q^2 + r(r - 2\chi Q)} + \chi Q \ln\left[\frac{Q\sqrt{\chi^2 - 1}}{\chi Q - 2\left\{r + \sqrt{Q^2 + r(r - 2\chi Q)}\right\}}\right] \tag{20}$$

On discretizing $\rho = ja$, we convert the above expression into a dimensionless form:

$$j = -\sqrt{\Delta_Q^2 + r_j(r_j - 2\chi\Delta_Q)} + \chi\Delta_Q \ln\left[\frac{\Delta_Q\sqrt{\chi^2 - 1}}{\chi\Delta_Q - 2\left\{r_j + \sqrt{\Delta_Q^2 + r_j(r_j - 2\chi\Delta_Q)}\right\}}\right], \tag{21}$$

where $\Delta_Q \equiv Q/a$ and $r_j \equiv r/a$ are both dimensionless. We also see that

$$f_j = 1 - \frac{2\chi\Delta_Q}{r_j} + \frac{\Delta_Q^2}{r_j^2}. \tag{22}$$

The resulting Hamiltonian H is factorized into a scale-dependent part $1/a$ and a scale-independent part $\tilde{H}$. The latter is completely characterized by dimensionless parameters Λ, Δ_Q and χ, all of which are invariant under the scaling transformations:

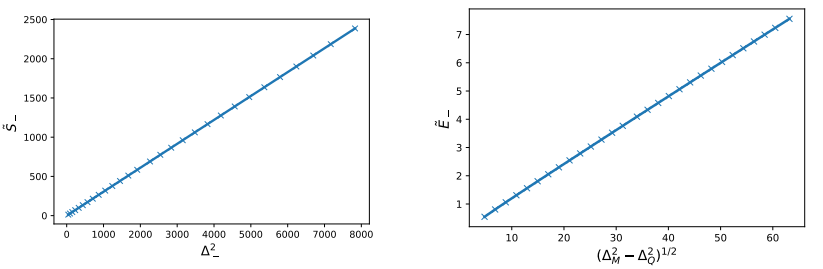
$$a \to \eta a; \quad m_f \to \eta^{-1} m_f; \quad M \to \eta M; \quad Q \to \eta Q \tag{23}$$

The IR cut-off on proper length is fixed at $r = 0$, leading to a certain discretization relation:

$$\Delta_Q \left[\chi \ln\sqrt{\frac{\chi + 1}{\chi - 1}} - 1\right] = N + 1 \tag{24}$$

The above relation tells us that if we fix Δ_q, then χ is discretized and vice versa. Here, we will consider the scenario where horizon changes on account of varying Δ_Q while keeping χ fixed. Physically, this corresponds to varying both mass and charge of the black hole proportionately to account for particles with a fixed mass-charge ratio (χ) that are entering the event horizon. As a result, both mass and charge have equally spaced discrete spectra:

$$Q_N = \frac{(N+1)a}{\left(\chi \ln \sqrt{\frac{\chi+1}{\chi-1}} - 1\right)}; \quad M_N = \chi Q_N; \tag{25}$$



(a) (b)

Fig. 3. Entanglement Mechanics at RN Cauchy horizon, when $\chi = 1.1$.

3.2.2. *Event horizon*

In terms of the dimensionless variable (χ), the event horizon is

$$r_+ = Q(\chi + \sqrt{\chi^2 - 1})\,.$$

From Eq. (4), we obtain:

$$\rho = \sqrt{Q^2 + r(r - 2\chi Q)} + \chi Q \ln \left[\frac{r - \chi Q + \sqrt{Q^2 + r(r - 2\chi Q)}}{Q\sqrt{\chi^2 - 1}} \right]. \tag{26}$$

On discretizing $\rho = ja$, we convert the above expression into a dimensionless form:

$$j = \sqrt{\Delta_Q^2 + r_j(r_j - 2\chi\Delta_Q)} + \chi\Delta_Q \ln \left[\frac{r_j - \chi\Delta_Q + \sqrt{\Delta_Q^2 + r_j(r_j - 2\chi\Delta_Q)}}{\Delta_Q\sqrt{\chi^2 - 1}} \right] \tag{27}$$

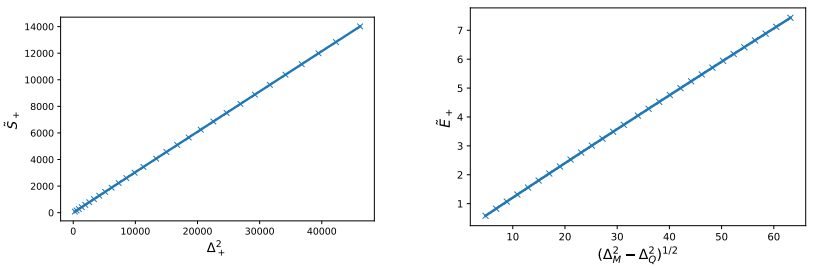
where $\Delta_Q \equiv Q/a$ and $r_j \equiv r/a$ are both dimensionless. From Figs 3 and 4, we obtain the following scaling relations for scale-invariant system $\tilde{H}$:

$$\tilde{S}_\pm = c_s \Delta_\pm^2; \quad \tilde{E}_\pm = c_e \sqrt{\Delta_M^2 - \Delta_Q^2}, \tag{28}$$

where a linear fit fixes the values $c_s \sim 0.3$ and $c_e \sim 0.12$ for both horizons. It can also be seen from here that in the limit $\Delta_Q \to 0$, we recover the values of c_e and c_s for Schwarzschild (18). As discussed above, $\tilde{S}_-$ and $\tilde{E}_-$ have discrete spectra. Since the entanglement energy is identical for both the horizons, we may therefore write for the total Hamiltonian (H):

$$S_\pm = c_s \frac{r_\pm^2}{a^2}; \quad E_+ = c_e \frac{\sqrt{M^2 - Q^2}}{a^2}; \quad T^{(\pm)} = \frac{\pi c_e}{c_s} T_H^{(\pm)} \tag{29}$$

where $T_H^{(-)}$ and $T_H^{(+)}$ are the Hawking temperatures of Cauchy and event horizon, respectively.



(a) (b)

Fig. 4. Entanglement Mechanics at RN event horizon, when $\chi = 1.1$.

3.3. *Schwarzschild de-Sitter*

We now move on to a space-time with multiple horizons that is asymptotically non-flat. A Schwarzschild black hole (of mass M) in a de-Sitter space-time (of radius l) is described by:

$$f(r) = 1 - \frac{2M}{r} - \frac{r^2}{l^2},$$

This space-time also has two horizons — r_b (event horizon) and r_c (cosmological horizon)[29]:

$$r_- = -\frac{2l}{\sqrt{3}} \cos\frac{\theta}{3}; \quad r_b = \frac{2l}{\sqrt{3}} \cos\frac{\pi + \theta}{3}; \quad r_c = \frac{2l}{\sqrt{3}} \cos\frac{\pi - \theta}{3} \tag{30}$$

where r_- is the third negative root, $\theta = \cos^{-1}(3\sqrt{3}\chi)$ and $\chi = M/l \in \left[0, 1/(3\sqrt{3})\right]$. $f(r)$ is positive in the region between the two horizons. Hence, in this region, we have two definitions for proper length — one w.r.t. the event horizon r_b which we refer to as ρ_b and the second w.r.t. the cosmological horizon r_c which we refer to as ρ_c.

3.3.1. *Event horizon*

The proper length with respect to the event horizon r_b is obtained as follows[23]:

$$\rho_b = \frac{2r_b l}{\sqrt{r_c(r_b - r_-)}} \Pi\left(\vartheta, \alpha^2, k\right), \tag{31}$$

where,

$$\vartheta = \sin^{-1}\sqrt{\frac{r_c(r-r_b)}{r(r_c-r_b)}}; \quad \alpha^2 = 1 - \frac{r_b}{r_c}; \quad k^2 = \frac{r_-(r_b-r_c)}{r_c(r_b-r_-)}. \tag{32}$$

On discretizing proper length $\rho_b = ja$, we convert the above expression into a dimensionless form:

$$j = \frac{2\Delta_b \Delta_l}{\sqrt{\Delta_c(\Delta_b - \Delta_-)}} \Pi\left(\vartheta, \alpha^2, k\right). \tag{33}$$

In terms of the dimensionless variables $\Delta_l = l/a$ and $r_j = r/a$, we have

$$f_j = 1 - \frac{2\Delta_M}{r_j} - \frac{r_j^2}{\Delta_l^2}, \tag{34}$$

and,

$$\begin{aligned} \Delta_- &= -\frac{2\Delta_l}{\sqrt{3}} \cos\frac{\theta}{3}; \quad \Delta_b = \frac{2\Delta_l}{\sqrt{3}} \cos\frac{\pi+\theta}{3}; \quad \Delta_c = \frac{2\Delta_l}{\sqrt{3}} \cos\frac{\pi-\theta}{3}; \\ \vartheta &= \sin^{-1}\sqrt{\frac{\Delta_c(r_j-\Delta_b)}{r_j(\Delta_c-\Delta_b)}}; \quad \alpha^2 = 1 - \frac{\Delta_b}{\Delta_c}; \quad k^2 = \frac{\Delta_-(\Delta_b-\Delta_c)}{\Delta_c(\Delta_b-\Delta_-)}. \end{aligned} \tag{35}$$

As a result, the Hamiltonian $\tilde{H}$ is fully characterized by dimensionless parameters Δ_l and Δ_M, both of which are invariant under the scaling transformations:

$$a \to \eta a; \quad m_f \to \eta^{-1} m_f; \quad M \to \eta M; \quad l \to \eta l \tag{36}$$

In the case of SdS, the IR cut-off on proper length is automatically fixed as we restrict ourselves to the region $\tilde{r}_b \le r_j \le \tilde{r}_c$:

$$N + 1 = \frac{2\tilde{r}_b \Delta_l}{\sqrt{\tilde{r}_c(\tilde{r}_b - \tilde{r}_-)}} \Pi\left(\frac{\pi}{2}, \alpha^2, k\right) \tag{37}$$

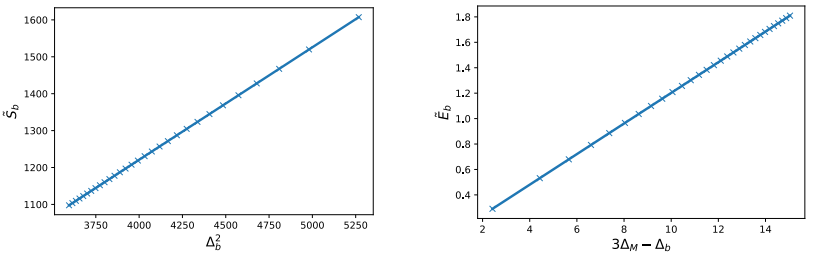
This is a discretization relation similar to what was obtained for RNBH. We will consider the case where we fix Δ_M and vary Δ_l by varying N. This is to ensure that χ is always between $[0, 1/(3\sqrt{3}]$. From Fig 5, we obtain the following scaling relations for the scale-invariant Hamiltonian ($\tilde{H}$):

$$\tilde{S}_b \sim c_s \Delta_b^2; \qquad \tilde{E}_b \sim c_e(3\Delta_M - \Delta_b) \tag{38}$$

where, $c_s \sim 0.3$ and $c_e \sim 0.12$ are the best-fit numerical values. In the limit $\Delta_l \to \infty$, we recover the prefactors of the Schwarzschild black hole (18). For the total Hamiltonian (H), the scaling relations become:

$$S_{b,N} \sim c_s \frac{r_{b,N}^2}{a^2}; \qquad E_{b,N} \sim c_e \frac{3M - r_{b,N}}{a^2}; \qquad T^{(b)} = \frac{\pi c_e}{c_s} T_H^{(b)} \tag{39}$$

where $T_H^{(b)}$ is the Hawking temperature of the event horizon in SdS.[29,30]



(a) (b)

Fig. 5. Entanglement Mechanics at SdS event horizon, when $\Delta_M = 25$.

3.3.2. *Cosmological horizon*

To explore the scaling properties of cosmological horizon, we define proper distance r_c as follows[31]:

$$\rho_c = \frac{2l}{\sqrt{r_c(r_b - r_-)}} \left[r_- F(\vartheta, k) - (r_c - r_-)\, \Pi\left(\vartheta, \alpha^2, k\right) \right], \tag{40}$$

where,

$$\sin\vartheta = \sqrt{\frac{(r_b - r_-)(r_c - r)}{(r_c - r_b)(r - r_-)}}; \quad \alpha^2 = \frac{r_b - r_c}{r_b - r_-}, \tag{41}$$

and the definition of k is the same as Eq. (32). On discretizing proper length $\rho_c = ja$, we convert the above expression into a dimensionless form:

$$j = \frac{2\Delta_l}{\sqrt{\Delta_c(\Delta_b - \Delta_-)}} \left[\Delta_- F(\vartheta, k) - (\Delta_c - \Delta_-)\, \Pi\left(\vartheta, \alpha^2, k\right) \right], \tag{42}$$

where in terms of dimensionless variables $\Delta_l = l/a$ and $r_j = r/a$, we can also rewrite:

$$\sin\vartheta = \sqrt{\frac{(\Delta_b - \Delta_-)(\Delta_c - r_j)}{(\Delta_c - \Delta_b)(r_j - \Delta_-)}}; \quad \alpha^2 = \frac{\Delta_b - \Delta_c}{\Delta_b - \Delta_-}, \tag{43}$$

Except for ϑ and α given above, the parameters used here follow the same definition as in (35). Now we impose an IR cut-off on proper length to restrict ourselves in the region $\tilde{r}_b \leq r_j \leq \tilde{r}_c$:

$$N + 1 = \frac{2\Delta_l}{\sqrt{\Delta_c(\Delta_b - \Delta_-)}} \left[\Delta_- K(k) - (\Delta_c - \Delta_-)\, \Pi\left(\frac{\pi}{2}, \alpha^2, k\right) \right] \tag{44}$$

This expression relates the number of oscillators N, Δ_l and χ, and we only need to fix two of these to fix the third. We will fix Δ_M here as we did for the event horizon, which leaves Δ_l with a discrete spectrum. From Fig 6, we see that the spectra for Δ_l obtained from ρ_b and ρ_c coincide exactly, and therefore the two discretization relations (37) and (44) are identical.

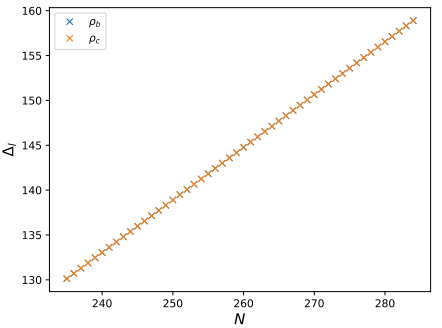


Fig. 6. Discretization of Δ_l from the cutoffs on proper lengths ρ_b and ρ_c, for $\Delta_M = 25$.

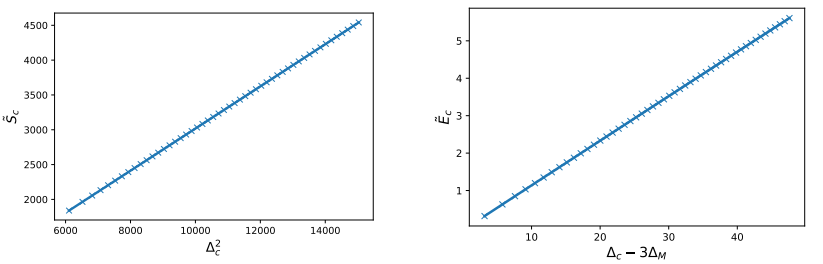
From Fig. 7, we obtain the following scaling relations for the scale-invariant system $\tilde{H}$:

$$\tilde{S}_c \sim c_s \Delta_c^2; \qquad \tilde{E}_c \sim c_e \left[\Delta_c - 3\Delta_M\right] \tag{45}$$

where $c_s \sim 0.3$ and $c_e \sim 0.12$ are the best-fit numerical values. For the total Hamiltonian H, the scaling relations become:

$$S_{c,N} \sim c_s \frac{r_{c,N}^2}{a^2}; \quad E_{c,N} \sim c_e \frac{r_{c,N} - 3M}{a^2}; \quad T^{(c)} \sim \frac{\pi c_e}{c_s} T_H^{(c)}, \tag{46}$$

where $T_H^{(c)}$ is the Hawking temperature of the cosmological horizon in SdS.



(a) (b)

Fig. 7. Entanglement Mechanics at SdS cosmological horizon, when $\Delta_M = 25$.

4. Black hole thermodynamics from entanglement mechanics

On numerically simulating the quantum entanglement mechanics near the horizon r_h for a variety of such space-times, we observe some universal properties associated with their scaling relations. Before we proceed, we first invoke the definition of entanglement energy wherein the pre-factor ϵ was introduced. Here, we may fix

this pre-factor as $\epsilon \sim 1.26$ upon imposing the condition that the entanglement temperature is identical to Hawking temperature ($T_{ent} = T_H$). While this rescales the proportionality constant c_e obtained numerically, the simulations nevertheless capture a universal one-to-one correspondence[23] between entanglement mechanics and black hole thermodynamics:

$$E_{\rm ent} = \frac{c_e}{a^2} E_{\rm Komar}; \quad S_{\rm ent} = \frac{c_s}{\pi a^2} S_{\rm BH}; \quad T_{\rm ent} = T_H \,. \tag{47}$$

Note that $E_{\rm ent}$ and $S_{\rm ent}$ depend on the UV cut-off (a), where as the temperature does not. Moreover, even with the pre-factor ϵ being assigned a new value, the constants of proportionality are found to be universal across all black-hole space-times, irrespective of whether they are asymptotically flat or non-flat:

$$c_e \sim 0.0955; \quad c_s \sim 0.3 \,. \tag{48}$$

Let us put these results in perspective: We have related quantum scalar fields near the horizon with the thermodynamic observables that describe any black-hole space-time (cf. Table 1). The fact that this is true for all 4-D spherically symmetric space-time suggests that entanglement of the quantum fields near the horizon carries crucial information about the black-hole thermodynamics. The above relations (47) go further and lead to the following universal and cut-off independent (a) result:

$$E_{\rm ent} = 2T_{\rm ent}\, S_{\rm ent} \qquad \Longleftrightarrow \qquad E_{\rm Komar} = 2T_H\, S_{\rm BH} \tag{49}$$

It is important to note that the above relation does not imply equality of their respective counterparts. On rearranging this relation further, we obtain the generalized Smarr formula of black hole thermodynamics, as summarized in Table 1.

Table 1. Generalized Smarr formula.

Space-time	**Entanglement Structure**	**Thermodynamic Structure**	**Smarr formula**	**Pressure**	**Potential**
Schwarzschild	$S_{\rm ent} = (c_s/a^2) r_h^2$ $E_{\rm ent} = (c_e/a^2) M$	$S_{BH} = \pi r_h^2$ $E_{\rm Komar} = M$	$M = 2T_{\rm H}\, S_{BH}$	—	—
Reissner-Nordström	$S_+ = (c_s/a^2) r_+^2$ $E_+ = (c_e/a^2)\sqrt{M^2 - Q^2}$	$S_{BH} = \pi r_+^2$ $E_{Komar} = \sqrt{M^2 - Q^2}$	$M = 2T_{\rm H}\, S_{BH} + Q^2/r_+$	—	Q/r_+
Schwarzschild-AdS	$S_{\rm ent} = (c_s/a^2) r_h^2$ $E_{\rm ent} = (c_e/a^2)[3M - r_h^2]$	$S_{BH} = \pi r_h^2$ $E_{Komar} = 3M - r_h^2$	$M = 2T_{\rm H}\, S_{BH} - r_h^3/l^2$	$3/8\pi l^2$	—
Schwarzschild-dS	$S_b = (c_s/a^2) r_b^2$ $E_b = (c_e/a^2)[3M - r_b^2]$	$S_{BH} = \pi r_b^2$ $E_{Komar} = 3M - r_b^2$	$M = 2T_{\rm H}\, S_{BH} + r_b^3/l^2$	$-3/8\pi l^2$	—

5. Conclusion

The Smarr formula for asymptotically flat and non-flat space-times is generally derived using Komar integral relations[32] or via the first law of thermodynamics with the help of scaling relations,[33] both of which makes use of Killing potential. In this talk, we developed an independent approach towards the derivation of the generalized Smarr formula, which essentially describes the equation of state for

black holes. The intrinsically quantum phenomenon of entanglement pertaining to the field near a horizon, gives rise to not just the thermodynamic quantities associated with the space-time, but also the exact relation connecting the same. This in turn gives entanglement mechanics an upper hand in addressing the universality problem of black hole entropy. While a wealth of other approaches may explain black-hole entropy, entanglement mechanics *alone* expands the analogy to include other thermodynamic quantities as well.

The results further complements two earlier results: First, Jacobson argued that entanglement provides a link between the presence of matter and the space-time geometry.[34] Second, in quantum gravity, Perez conjectured that degrees of freedom hidden from the classical space-time description but correlated to matter fields are necessary to maintain unitarity in the global evolution and prevent the information loss.[35] Our analysis here provides a crucial link to these results, placing entanglement at the center of a fundamental connection between space-time and matter, thereby holding the key to understanding black-hole horizons.

Acknowledgements

SMC is supported by Prime Minister's Research Fellowship offered by the Ministry of Education, Govt. of India. The work is supported by the MATRICS SERB grant.

References

1. D. Kastor, S. Ray and J. Traschen, Enthalpy and the mechanics of AdS black holes, *Classical Quantum Gravity* **26**, p. 195011 (2009).
2. T. Padmanabhan, Equipartition energy, noether energy and boundary term in gravitational action, *General Relativity and Gravitation* **44**, 2681 (2012).
3. T. Padmanabhan, General relativity from a thermodynamic perspective, *General Relativity and Gravitation* **46**, p. 1673 (2014).
4. P. T. Chruściel, On the relation between the einstein and the komar expressions for the energy of the gravitational field, *Annales de l'I.H.P. Physique théorique* **42**, 267 (1985).
5. S. W. Hawking, Particle creation by black holes, *Communications in Mathematical Physics* **43**, 199 (1975).
6. J. D. Bekenstein, Black holes and entropy, *Phys. Rev. D* **7**, 2333 (Apr 1973).
7. L. Smarr, Mass formula for kerr black holes, *Phys. Rev. Lett.* **30**, 71 (Jan 1973).
8. P. C. W. Davies, Thermodynamics of black holes, *Reports on Progress in Physics* **41**, 1313 (Aug 1978).
9. J. M. Bardeen, B. Carter and S. W. Hawking, The four laws of black hole mechanics, *Commun. Math. Phys.* **31**, 161 (1973).
10. R. M. Wald, The thermodynamics of black holes, *Living Reviews in Relativity* **4** (Jul 2001).
11. D. N. Page, Hawking radiation and black hole thermodynamics, *New Journal of Physics* **7**, 203 (Sep 2005).
12. S. Carlip, *Black Hole Thermodynamics and Statistical Mechanics*, in *Physics of Black Holes: A Guided Tour*, ed. E. Papantonopoulos (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), Berlin, Heidelberg, pp. 89–123.

13. S. Das, S. Shankaranarayanan and S. Sur, *Black hole entropy from entanglement: A review* (Nova Publishers, 2010), ch. 6.
14. S. Carlip, Black hole entropy and the problem of universality, *Journal of Physics: Conference Series* **67**, p. 012022 (May 2007).
15. G. Vidal, Entanglement renormalization, *Phys. Rev. Lett.* **99**, p. 220405 (Nov 2007).
16. B. Swingle, Entanglement renormalization and holography, *Phys. Rev. D* **86**, p. 065007 (Sep 2012).
17. M. Van Raamsdonk, Building up spacetime with quantum entanglement, *General Relativity and Gravitation* **42**, 2323 (2010).
18. C. Cao, S. M. Carroll and S. Michalakis, Space from hilbert space: Recovering geometry from bulk entanglement, *Phys. Rev. D* **95**, p. 024031 (Jan 2017).
19. K. Hyatt, J. R. Garrison and B. Bauer, Extracting entanglement geometry from quantum states, *Phys. Rev. Lett.* **119**, p. 140502 (Oct 2017).
20. L. Bombelli, R. K. Koul, J. Lee and R. D. Sorkin, Quantum source of entropy for black holes, *Phys. Rev. D* **34**, 373 (Jul 1986).
21. M. Srednicki, Entropy and area, *Phys. Rev. Lett.* **71**, 666 (Aug 1993).
22. J. Eisert, M. Cramer and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (Feb 2010).
23. S. M. Chandran and S. Shankaranarayanan, One-to-one correspondence between entanglement mechanics and black hole thermodynamics, *Phys. Rev. D* **102**, p. 125025 (Dec 2020).
24. S. Mukohyama, M. Seriu and H. Kodama, Thermodynamics of entanglement in schwarzschild spacetime, *Phys. Rev. D* **58**, p. 064001 (Jul 1998).
25. E. Witten, APS Medal for Exceptional Achievement in Research: Invited article on entanglement properties of quantum field theory, *Rev. Mod. Phys.* **90**, p. 045003 (2018).
26. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 1982).
27. T. Jacobson, *Introduction to Quantum Fields in Curved Spacetime and the Hawking Effect*, in *Lectures on Quantum Gravity*, eds. A. Gomberoff and D. Marolf (Springer US, Boston, MA, 2005), Boston, MA, pp. 39–89.
28. S. Mukohyama, M. Seriu and H. Kodama, Can the entanglement entropy be the origin of black-hole entropy?, *Phys. Rev. D* **55**, 7666 (Jun 1997).
29. S. Shankaranarayanan, Temperature and entropy of schwarzschild–de sitter spacetime, *Phys. Rev. D* **67**, p. 084026 (Apr 2003).
30. G. W. Gibbons and S. W. Hawking, Cosmological event horizons, thermodynamics, and particle creation, *Phys. Rev.* **D15**, 2738 (1977).
31. P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer Berlin Heidelberg, 1971).
32. A. Komar, Covariant conservation laws in general relativity, *Phys. Rev.* **113**, 934 (Feb 1959).
33. G. W. Gibbons, M. J. Perry and C. N. Pope, The First law of thermodynamics for Kerr-anti-de Sitter black holes, *Class. Quant. Grav.* **22**, 1503 (2005).
34. T. Jacobson, Thermodynamics of spacetime: The einstein equation of state, *Phys. Rev. Lett.* **75**, 1260 (Aug 1995).
35. A. Perez, No firewalls in quantum gravity: the role of discreteness of quantum geometry in resolving the information loss paradox, *Classical and Quantum Gravity* **32**, p. 084001 (Mar 2015).

Thermodynamics of charged black hole

M. Sharif* and Amjad Khan

Department of Mathematics, University of the Punjab, Quaid-e-Azam Campus, Lahore-54590, Pakistan
**E-mail: msharif.math@pu.edu.pk*
www.pu.edu.pk

This paper investigates thermodynamics as well as thermal stability of the Reissner-Nordström black hole with the effects of non-linear electrodynamics. We first calculate the expressions for Hawking temperature, Helmholtz free energy, internal energy, enthalpy and Gibbs free energy of this black hole and then study their graphical behavior in the presence of non-linear electrodynamic effects. We also investigate thermal stability of the considered system with two different methods, i.e., through heat capacity and through Hessian matrix. It is found from both the methods that the considered system is thermodynamically stable in the presence of non-linear electrodynamic parameter (α). Finally, we analyze the phase transitions of Hawking temperature as well as heat capacity in terms of entropy for different values of charge (q), horizon radius (r_+) and coupling parameter. We obtain that Hawking temperature changes its phase from positive to negative for increasing values of q and r_+ while it shows opposite trend for higher values of α. The heat capacity changes its phase from negative to positive for large values of charge, horizon radius and coupling parameter.

Keywords: Thermodynamics; Stability; Phase transitions.

1. Introduction

The study of final outcomes of self-gravitating objects is one of the most important subjects in general relativity. Black hole (BH) is a completely collapsed structure of massive stars which is defined as a thermodynamical object with very strong gravitational field such that anything even light cannot escape from it. Black hole thermodynamics has resemblance with classical thermodynamics in which mass is related to energy, surface area to entropy and surface gravity to temperature. This resemblance motivated Bekenstein[1] to find out the relation between entropy and area of the event horizon of BH. In this regard, he found that entropy is proportional to the area of event horizon of BH. Hawking's discovery[2] of black body radiations further confirmed the validity of this relation. Consequently, it is impossible to obtain thermal equilibrium between BH and thermal radiations leading to the concept of holographic principle.[3]

Hawking[4] evaluated the expressions of surface gravity, angular frequency and electric potential for any BH having mass, angular momentum and charge. He also discussed generalized second law of BH thermodynamics as well as thermal equilibrium condition for BHs. Motivated by the work of Hawking, Davies[5] discussed about the entropy, thermal stability, phase transition and third law of BH thermodynamics. He also discussed quantum effects of BHs as well as nature and origin of

BH thermal radiations. Wald[6] reviewed the laws of BH thermodynamics, physical correspondence among laws of BH as well as classical thermodynamics, derivation of Hawking radiations, formula for entropy of BH and validity of generalized second law of thermodynamics.

In literature, there is a wide discussion about thermodynamics of different BH geometries. Balart and Vagenas[7] proposed the exact solutions of various regular charged BHs in the presence of non-linear electrodynamic effects and observed that the behavior of some BHs asymptotically coincide to RN BH. Tharanath et al.[8] studied thermal quantities of four regular BHs and observed second order phase transition. Faizal and Khalil[9] found leading order corrections to the thermodynamic quantities of RN, Kerr as well as charged-AdS BHs and concluded that in all three cases, these BHs create remnants. Pradhan[10] investigated local thermal stability as well as logarithmic corrections for entropy of charged accelerating BHs and observed that these BHs undergo second order phase transitions. Sharif and Zunaira[11] investigated the effects of thermal fluctuations on thermodynamics of charged BH with Weyl corrections and concluded that in the presence of logarithmic corrections small BHs are unstable. Sinha[12] discussed thermodynamics as well as stability of asymptotically flat RN BH and observed that this BH exhibits phase transition which is different from the phase transition of Schwarzschild BH.

Davies[13] developed thermodynamical theory of BHs and found that phase transition occurs in the Kerr-Newman BH. Hawking and Page[14] determined the presence of phase transition in the Schwarzschild-AdS BH. Biswas and Chakraborty[15] investigated that whether phase transition is possible in Horava-Lifshitz gravity under the consideration of classical and topological choices of BH thermodynamics. They concluded that phase transition occurs in Horava-Lifshitz gravity from stable to unstable phase for increasing values of radius. Kubiznak and Mann[16] analyzed the first order small-large BHs phase transition in charged AdS BH which is analogous to liquid-gas phase transitions of fluids.

Chaturvedi et al.[17] investigated phase transition in the framework of thermodynamic geometry for charged AdS BH and found first order phase transition for fixed electric charge. Wei et al.[18] studied thermodynamics and found first order phase transition for charged AdS BHs in five-dimensions. Ovgun[19] studied thermodynamics as well as phase transition of a specific charged AdS type BH in $f(R)$ gravity coupled with Yang-Mills field by considering cosmological constant as thermal pressure. Wei and Liu[20] discussed a relation between phase transition and null geodesics of charged AdS BH. Saleh et al.[21] examined thermodynamics as well as phase transition of Bardeen BH surrounded by quintessence and found that the presence of quintessence induces phase transition in the considered BH. Kuang et al.[22] evaluated thermal quantities of AdS non-linear electrodynamic BH. Bhatacharya et al.[23] provided a general criteria to get information and other various results with an extremal limit of BH spacetime. They also evaluated critical values of second order phase transition without considering any specific BH and showed that these values are in agreement with those of any specific BH cases.

Gonzalez et al.[24] studied thermodynamics and stability of charged BHs with non-linear electrodynamics and found that small BHs are stable locally. Dayyani[25] explored thermodynamical properties and phase transition of dilaton BH with non-linear electrodynamics and observed zeroth order phase transition. Yu and Gao[26] derived the exact solutions of RN BH with non-linear electrodynamics. Javed et al.[27] observed the effect of non-linear electrodynamics on weak field deflection angle of RN BH. Recently, Fauzi and Ramadhan[28] discussed thermodynamics of charged BHs with non-linear electrodynamics and verified the first law of BH thermodynamics.

In this paper, we study thermodynamical quantities, thermal stability and phase transition for RN BH with non-linear electrodynamic effects. The paper is outlined as follows. In section **2**, we discuss thermodynamic quantities and their graphical behavior for this BH. Section **3** is devoted to investigate thermal stable configuration and section **4** analyzes phase transitions in terms of entropy. In the last section, we conclude all the results.

2. Thermodynamics

In this section, we discuss thermodynamic quantities such as Hawking temperature, entropy, Helmholtz-Gibbs free energy, internal energy and enthalpy of RN BH with non-linear electrodynamic effects. The line element of this BH is given as[26]

$$ds^2 = f(r)dt^2 - \frac{dr^2}{f(r)} - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{1}$$

with

$$f(r) = 1 + \frac{q^2}{r^2} - \left(\frac{2M}{r} + \frac{\alpha^2 r^2}{3}\right) + 2\alpha q, \tag{2}$$

where q, M and α represent charge, mass of BH and coupling constant, respectively. The line element (1) reduces to the RN BH when $\alpha = 0,\ q \neq 0$, Schwarzschild metric for $\alpha = 0,\ q = 0$ and it becomes Schwarzschild BH with the effects of non-linear electrodynamics for $\alpha \neq 0,\ q = 0$. Figure 1 shows the graphical analysis of the metric function (2) indicating the appearance of naked singularities for the increasing values of the coupling parameter. Setting $f(r_+) = 0$, the mass of BH in terms of r_+ can be obtained as

$$M = \frac{1}{6r_+}\left(3q^2 + r_+^2(6\alpha q - \alpha^2 r_+^2 + 3)\right), \tag{3}$$

where r_+ represents horizon radius of BH. To obtain the expression for Hawking temperature, we first evaluate surface gravity of the BH

$$\kappa = -\frac{\left(q + r_+ - \alpha r_+^2\right)\left(q - \alpha r_+^2 - r_+\right)}{2r_+^3}.$$

Consequently, the Hawking temperature ($T = \frac{\kappa}{2\pi}$) is

$$T = -\frac{\left(q + r_+ - \alpha r_+^2\right)\left(q - \alpha r_+^2 - r_+\right)}{4\pi r_+^3}. \tag{4}$$

The graphical sketch of Hawking temperature versus α is shown in Figure 2. It is found that Hawking temperature increases with the increasing values of coupling parameter throughout the considered domain.

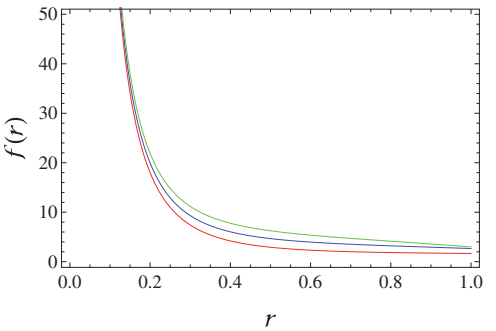


Fig. 1. Plot of the metric function versus r for $M = 1 = q$ with $\alpha = 1$(red), 2(blue) and 3(green).

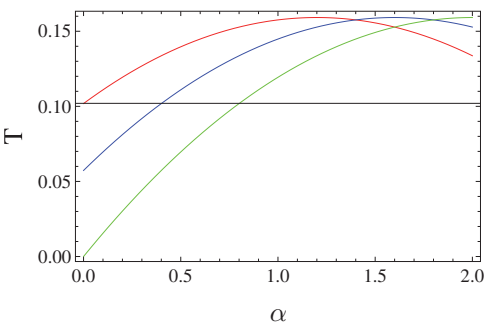


Fig. 2. Plot of Hawking temperature versus α for $r_+ = 0.5$ with $q = 0.3$(red), 0.4(blue) and 0.5(green).

According to Bekenstein area-entropy relation, the entropy of BH can be evaluated as[29]

$$S = \int_0^{2\pi}\int_0^{\pi} \sqrt{g_{\theta\theta}g_{\phi\phi}}d\theta d\phi = \pi r_+^2. \tag{5}$$

Now we calculate other thermal quantities by using the expressions of entropy and Hawking temperature. The expression of Helmholtz free energy can be

calculated as

$$F = \frac{1}{16}\left(4q^2\log(r_+) + r_+^2\left(-4\alpha q + \alpha^2 r_+^2 - 2\right)\right).$$

Figure 3 shows the graphical representation of Helmholtz free energy for different values of the coupling parameter. It is noted that Helmholtz free energy is an increasing function of r_+ and for increasing values of the coupling constant, it increases more rapidly. The increasing trend of Helmholtz free energy depicts that the reactions occurring inside the considered system are non-spontaneous and work done is required to make the system thermodynamically stable. The internal energy ($U = F + ST$) of the considered system can be obtained as

$$U = \frac{4q^2 r_+\log(r_+) - 4q^2 - 4\alpha q(r_+ - 2)r_+^2 + \alpha^2(r_+ - 4)r_+^4 - 2(r_+ - 2)r_+^2}{16r_+}.$$

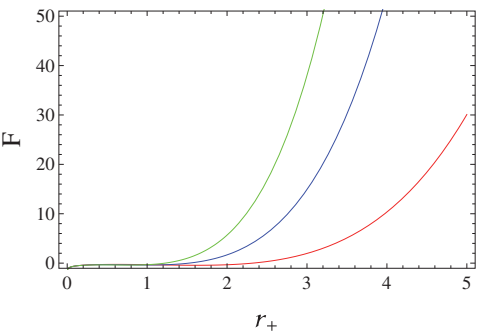


Fig. 3. Plot of Helmholtz free energy versus r_+ for $q = 1$ with $\alpha = 1$(red), 2(blue) and 3(green).

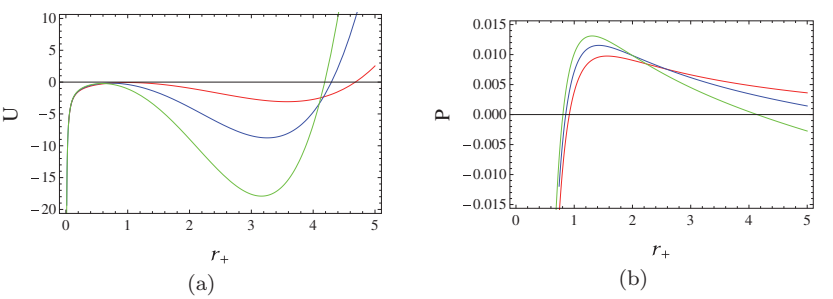


Fig. 4. (a) Plots of internal energy versus r_+ for $q = 1$ with $\alpha = 1$(red), 2(blue), 3(green). (b) Plot of pressure versus r_+ for $q = 1$ with $\alpha = 0.1$(red), 0.2(blue), 0.3(green).

The graphical representation of internal energy in terms of horizon radius is shown in Figure 4(a). It is observed that internal energy of the system fluctuates

for increasing values of r_+ and becomes positive towards the boundary of event horizon. The negative behavior of internal energy shows that the considered system is releasing heat for small values of r_+. The corresponding volume is given as[24]

$$V = \frac{4\pi r_+^3}{3}.$$

The pressure for the above calculated volume can be obtained by the relation

$$P = -\frac{dF}{dV} = -\frac{dF}{dr_+}\frac{dr_+}{dV},$$

which gives

$$P = -\frac{\left(q - \alpha r_+^2 + r_+\right)\left(q - r_+(\alpha r_+ + 1)\right)}{16\pi r_+^3}.$$

Figure 4(b) shows the graphical structure of pressure versus r_+ for different values of the coupling parameter. It is noted that pressure of the considered system firstly increases for certain values of horizon radius and then decreases monotonically for large radii. However, for increasing values of the coupling constant, the pressure of BHs with small radii increases slightly while shows opposite trend for large radii.

The expression of enthalpy ($H = VP + U$) can be calculated as follows

$$H = -\left(4q^2(r_+ + 3) - 12q^2 r_+ \log(r_+) + 4\alpha q(r_+ - 6)r_+^2 + \alpha^2(r_+ + 12)r_+^4 + 2(r_+ - 6)r_+^2\right)(48r_+)^{-1}.$$

The graphical analysis of enthalpy in terms of horizon radius is represented in Figure 5(a). It is observed that enthalpy of the considered system is an increasing function of horizon radius. The increasing values of the coupling parameter do not affect the enthalpy for BHs of both small as well as large radii. It is concluded that the considered system is absorbing heat from its surroundings. The Gibbs free energy ($G = H - ST$) of the considered system is obtained as

$$G = \frac{1}{48}\left(12q^2\log(r_+) - \left(2q + \alpha r_+^2\right)^2 - 2r_+^2\right).$$

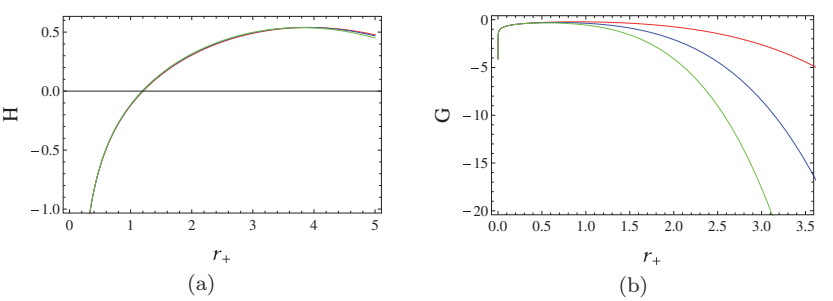


Fig. 5. (a) Plot of enthalpy (left) versus r_+ for $q = 1$ with $\alpha = 0.01$(red), 0.02(blue), 0.03(green). (b) Plot of Gibbs free energy (right) versus r_+ for $q = 1$ with $\alpha = 1$(red), 2(blue), 3(green).

Figure 5(b) represents the graphical structure of Gibbs free energy in terms of r_+ for different values of the coupling constant. It is noted that the Gibbs free energy of the system is inversely proportional to the horizon radius. The increasing values of the coupling parameter do not affect the Gibbs free energy of BHs with $r_+ < 1$ while for $r_+ \geq 1$, the Gibbs free energy is affected significantly and decreases more rapidly.

3. Thermal Stability

In this section, we discuss thermal stable configuration of RN BH with non-linear electrodynamic effects through the following two different methods:

- Heat Capacity;
- Hessian Matrix.

3.1. *Through Heat Capacity*

In order to analyze thermal stability of the considered system, the expression of heat capacity $(T\frac{\partial S}{\partial T})$ can be calculated as follows

$$C = \frac{\pi r_+^2 \left(q + r_+ - \alpha r_+^2\right)\left(q - \alpha r_+^2 - r_+\right)}{r_+^2 - 2\left(q^2 - \alpha q r_+^2\right)}. \tag{6}$$

The graphical analysis of heat capacity represents that it diverges at $r_+ = 0.56$ and this divergent point is known as Davies point (Figure 6). It is also noted that heat capacity changes its phase from negative to positive for increasing values of horizon radius which shows that the considered system is stable in the presence of non-linear electrodynamic effects.

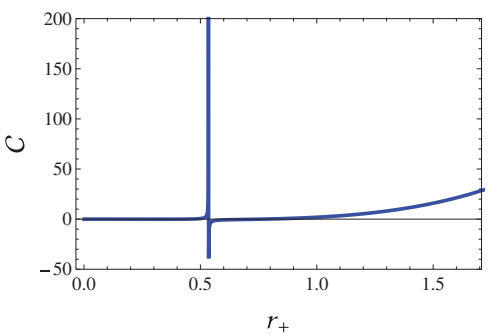


Fig. 6. Plot of heat capacity versus r_+ for $q = 1$ and $\alpha = 3$.

3.2. *Through Hessian Matrix*

We can also use Hessian matrix to investigate thermal stability of the considered system which contains second order partial derivatives of Helmholtz free energy and is defined as[30]

$$H = \begin{pmatrix} H_{mm} & H_{mn} \\ H_{nm} & H_{nn} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 F}{\partial T^2} & \frac{\partial^2 F}{\partial T \partial \upsilon} \\ \frac{\partial^2 F}{\partial \upsilon \partial T} & \frac{\partial^2 F}{\partial \upsilon^2} \end{pmatrix},$$

with $m \neq n$, and

$$H_{mm} = \frac{\partial^2 F}{\partial T^2}, \quad H_{mn} = \frac{\partial^2 F}{\partial T \partial \upsilon}, \quad H_{nm} = \frac{\partial^2 F}{\partial \upsilon \partial T}, \quad H_{nn} = \frac{\partial^2 F}{\partial \upsilon^2}.$$

It is noted that $H_{mm}H_{nn} = H_{mn}H_{nm}$, which implies that one of the eigenvalues of the Hessian matrix is zero. Therefore, to determine the stability of the considered geometry, we use trace of the Hessian matrix as

$$Tr(H) = H_{mm} + H_{nn},$$

where

$$\begin{aligned} H_{mm} = &- \left(4\pi^2 r_+^6 \left(9q^4 - 32\alpha q^3 r_+^2 + 2q^2 r_+^2 \left(17\alpha^2 r_+^2 - 8\right) + 4\alpha q r_+^4 (3\right.\right. \\ &\left.\left. - 2\alpha^2 r_+^2\right) - 3\alpha^4 r_+^8 - 4\alpha^2 r_+^6 + 3r_+^4)\right) \left(\left(-3q^2 + 2\alpha q r_+^2 + \alpha^2 r_+^4 + r_+^2\right)^3\right)^{-1}, \\ H_{nn} = &\left(r_+^2 \left(3\alpha + \left(q - \alpha r_+^2\right)^{-1}\right) + q\left(2r_+^2 \left(q - \alpha r_+^2\right)^{-2} - 1\right)\right) \\ &\times \left(4\left(\alpha - q r_+^{-2}\right)\right)^{-1}, \end{aligned}$$

hence

$$\begin{aligned} Tr(H) = &\left(r_+^2 \left(3\alpha + \left(q - \alpha r_+^2\right)^{-1}\right) + q\left(2r_+^2 \left(q - \alpha r_+^2\right)^{-2} - 1\right)\right) \\ &\times \left(4\left(\alpha - q r_+^{-2}\right)\right)^{-1} - \left(4\pi^2 r_+^6 \left(9q^4 - 32\alpha q^3 r_+^2 + 2q^2 r_+^2 \left(17\alpha^2 r_+^2 - 8\right)\right.\right. \\ &\left.\left. + 4\alpha q r_+^4 \left(3 - 2\alpha^2 r_+^2\right) - 3\alpha^4 r_+^8 - 4\alpha^2 r_+^6 + 3r_+^4)\right) \right. \\ &\left. \times \left(\left(-3q^2 + 2\alpha q r_+^2 + \alpha^2 r_+^4 + r_+^2\right)^3\right)^{-1}. \end{aligned}$$

A thermodynamical system is said to be stable if $Tr(H) \geq 0$.[31] Figure 7 shows the graphical analysis of the Hessian trace in terms of r_+ for different values of the coupling parameter. It is observed that the trace of Hessian matrix is not defined in the interval $0.5 < r_+ < 0.59$. It is also observed that the Hessian trace shows similar behavior to heat capacity and hence shows that the considered system is stable in the presence of non-linear electrodynamic effects.

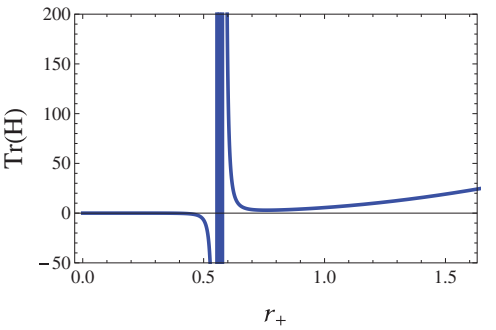


Fig. 7. Plot of Hessian trace versus r_+ for $q = 1$ and $\alpha = 3$.

4. Phase Transitions

Here we examine the phase transitions of Hawking temperature as well as heat capacity of RN BH with non-linear electrodynamic effects in terms of entropy. The Hawking temperature evaluated in Eq. (4) can be written in the form of entropy as

$$T = -\frac{\left(q - \alpha r_+^2 + r_+\right)\left(q - r_+(\alpha r_+ + 1)\right)}{4 r_+ S}.$$

Figure 8(a),(c),(e) represents the phase transitions of Hawking temperature in terms of S for different values of q, r_+ and α. It is noted that the Hawking temperature changes its phase from positive to negative for increasing values of q and r_+, respectively (Plot (a) and (c)). However, for increasing values of non-linear electrodynamic parameter, the Hawking temperature changes its phase from negative to positive (plot (c)). The expression of heat capacity for RN BH with non-linear electrodynamic effects in terms of entropy can be obtained from Eq. (6) as

$$C = \frac{S\left(q - \alpha r_+^2 + r_+\right)\left(q - r_+(\alpha r_+ + 1)\right)}{-2q^2 + 2\alpha q r_+^2 + r_+^2}.$$

The phase transitions of heat capacity versus entropy for different values of q, r_+ and α is shown in Figure 8(b),(d),(f). It is noted that heat capacity of the considered BH changes its phase from negative to positive for the increasing values of q, r_+ and α. It is also observed that the considered system is unstable for small values of the considered parameters while it shows stable behavior for large values of q, r_+ and α.

5. Concluding Remarks

In this paper, we have derived thermodynamical quantities for RN BH with non-linear electrodynamic effects and observed their behavior graphically. We have also discussed thermal stable configuration of this BH using both heat capacity as well as Hessian matrix and compared the results graphically. Firstly, we have calculated

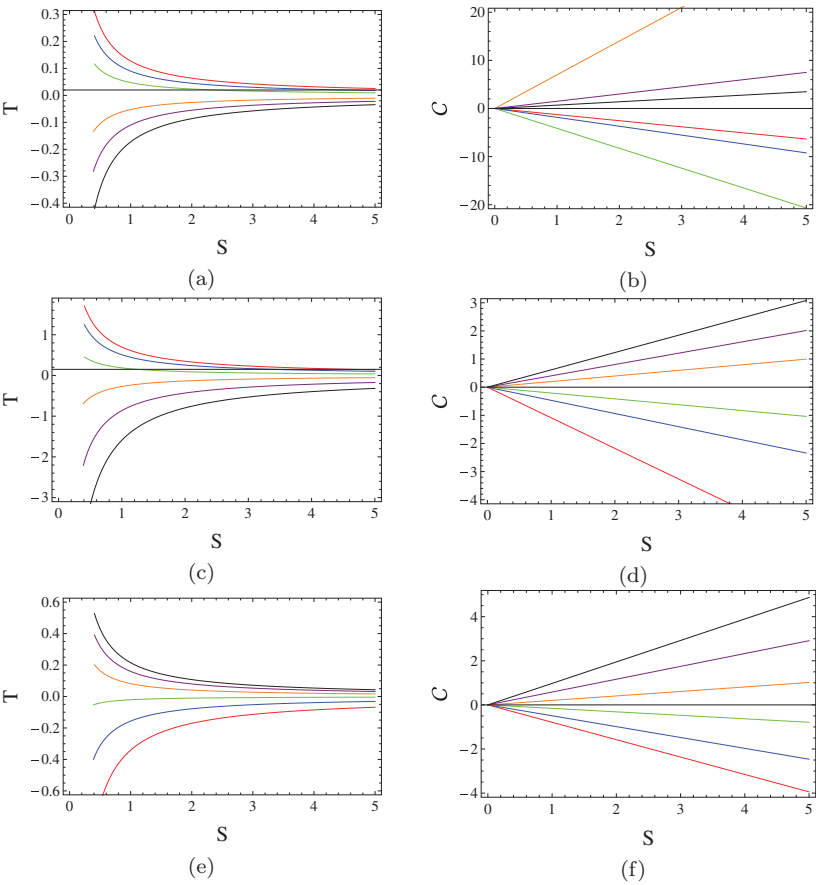


Fig. 8. Hawking temperature ((a), (c), (e)) versus S for $\alpha = 1$, $r_+ = 1$. We take $q = 1.7$(red), 1.8(blue), 1.9(green), 2.1(orange), 2.2(purple), 2.3(black) for plot (a), $\alpha = 1$, $q = 1$ with $r_+ = 0.4$(red), 0.5(blue), 0.6(green), 0.7(orange), 0.8(purple), 0.9(black) for plot (e) and $r_+ = 3$, $q = 1$ with $\alpha = 0.2$(red), 0.3(blue), 0.4(green), 0.5(orange), 0.6(purple), 0.7(black) plot (e). Heat capacity ((b), (d), (f)) versus S with $r_+ = 1$, $\alpha = 1$. We take $q = 1.1$(red), 1.2(blue), 1.3(green), 1.4(orange), 1.5(purple), 1.6(black) for plot (b), $\alpha = 2$, $q = 1$ with $r_+ = 0.7$(red), 0.8(blue), 0.9(green), 1.1(orange), 1.2(purple), 1.3(black) for plot (d) and $r_+ = 3$, $q = 1$, $\alpha = 0.2$(red), 0.3(blue), 0.4(green), 0.5(orange), 0.6(purple), 0.7(black) for plot (f).

the expressions of Hawking temperature with the help of surface gravity ($T = \frac{\kappa}{2\pi}$) and then Helmholtz free energy for the considered BH. The graphical representation of Helmholtz free energy shows that it is an increasing function of r_+ and the presence of non-linear electrodynamic effects only affect the Helmholtz free energy of BHs with large radii while BHs of small radii are unaffected. It is also noted that for increasing values of the coupling constant, it increases more rapidly. The increasing trend of Helmholtz free energy depicts that the reactions occurring inside the considered system are non-spontaneous. The internal energy of the system fluctuates and

has negative values for small values of r_+ which shows that the considered system is releasing heat. It is also observed that larger values of the coupling parameter induces large fluctuations in the internal energy of the system and vice-versa.

The enthalpy of the considered system is increasing continuously and showing that the considered system is absorbing heat while the increasing values of coupling parameter do not affect the enthalpy for BHs of both small as well as large radii (Figure 4(a)). Figure 5(a) shows that Gibbs free energy of the considered system is inversely proportional to the horizon radius. The increasing values of coupling parameter do not affect the Gibbs free energy of BHs with $r_+ < 1$ while for $r_+ \geq 1$, this free energy is affected significantly and decreases more rapidly.

Secondly, we have discussed thermal stability of the considered system with the help of heat capacity as well as Hessian matrix by observing their graphical behavior. It is noted that the heat capacity has Davies point at $r_+ = 0.56$ and changes its phase from negative to positive for increasing values of horizon radius which shows that the considered system is stable in the presence of coupling parameter (Figure 6). It is observed from Figure 7 that the trace of Hessian matrix is not defined in the interval $0.5 < r_+ < 0.59$ and shows similar behavior to heat capacity. It is concluded from both the methods that our system is thermodynamically stable in the presence of non-linear electrodynamic effects.

Finally, we have investigated the phase transitions of Hawking temperature as well as heat capacity in terms of entropy for the considered BH. It is found that Hawking temperature changes its phase from positive to negative for increasing values of both charge and horizon radius. This shows that BHs with small charge and radii are more hotter than the BHs with large charge and radii while it shows opposite behavior for increasing values of the coupling parameter. The heat capacity changes its phase from negative to positive for increasing values of the considered parameters which shows that BHs with small values of q, r_+ and α are unstable while BHs with large values are stable throughout the considered domain. We would like to mention here that all our results reduce to the RN BH for $\alpha = 0$.[9]

References

1. J. D. Bekenstein, Black holes and entropy, *Phys. Rev. D* **7**, 2333 (1973).
2. S. W. Hawking, Black hole explosions?, *Nature* **248**, 30 (1974).
3. R. Easther, D. Lowe, Holography, Cosmology and the Second Law of Thermodynamics, *Phys. Rev. Lett.* **82**, 4967 (1999).
4. S. W. Hawking, Black holes and thermodynamics, *Phys. Rev. D* **13**, 191 (1976).
5. P. C. Davies, Thermodynamics of black holes, *Rep. Prog. Phys.* **41**, 1313 (1978).
6. R. M. Wald, The Thermodynamics of black holes, *Living Rev. Relativ.* **4**, 1 (2001).
7. L. Balart, E. C. Vagenas, Regular black holes with a nonlinear electrodynamics source, *Phys. Rev. D* **90**, 124045 (2014).
8. R. Tharanath, J. Suresh, V. C. Kuriakose, Phase transitions and geometrothermodynamics of regular black holes, *Gen. Relativ. Gravit.* **46**, 47 (2015).
9. M. Faizal, M. M. Khalil, GUP-Corrected thermodynamics for all black objects and the existence of remnants, *Int. J. Mod. Phys. A* **30**, 1550144 (2015).

10. P. Pradhan, Horizon Areas and logarithmic correction to the charged accelerating black hole entropy, *Universe* **5**, 57 (2019).
11. M. Sharif, Z. Akhtar, Quasi-normal modes and thermal fluctuations of charged black hole with Weyl corrections, *Phys. Dark Universe* **29**, 100589 (2020).
12. A. K. Sinha, Thermodynamics of asymptotically flat Reissner–Nordstrom black hole, *Mod. Phys. Lett. A* **36**, 2150071 (2021).
13. P. C. W. Davies, The thermodynamic Theory of black holes, *Proc. Roy. Soc. Lond. A* **499**, 353 (1977).
14. S. W. Hawking, D. N. Page, Thermodynamics of black holes in anti-de Sitter space, *Comm. Math. Phys.* **87**, 577 (1983).
15. R. Biswas, S. Chakraborty, Black hole thermodynamics in Horava Lifshitz gravity and the related geometry, *Astrophys. Space Sci.* **193**, 332 (2011).
16. D. Kubiznak, R. B. Mann, P-V criticality of charged AdS black holes, *J. High Energy Phys.* **1207**, 033 (2012).
17. P. Chaturvedi, A. Das, G. Sengupta, Thermodynamic geometry and phase transitions of dyonic charged AdS black holes, *Eur. Phys. J. C* **110**, 77 (2017).
18. S. W. Wei, B. Liang, Y. X. Liu, Critical phenomena and chemical potential of charged AdS black hole, *Phys. Rev. D* **96**, 124018 (2017).
19. A. Ovgun, P-V criticality of a specific black hole in $f(r)$ gravity coupled with Yang-Mills field, *Ad. High Energy Phys.* **2018**, 8153721 (2018).
20. S. W. Wei, Y. X. Liu, Photon orbits and thermodynamic phase transition of d-dimensional charged AdS black holes, *Phys. Rev. D* **97**, 104027 (2018).
21. M. Saleh, B. B. Thomas, T. C. Kofane, Thermodynamics and phase transition from regular Bardeen black hole surrounded by quintessence, *Int. J. Theor. Phys.* **57**, 2640 (2018).
22. X. M. Kuang, B. Liu, A. Ovgun, Nonlinear electrodynamics AdS black hole and related phenomena in the extended thermodynamics, *Eur. Phys. J. C* **78**, 1 (2018).
23. K. Bhattacharya, et al., A general framework to study the extremal phase transition of black holes, *Phys. Rev. D* **99**, 124047 (2019).
24. H. A. Gonzalez, M. Hassaine, C. Martinez, Thermodynamics of charged black holes with a nonlinear electrodynamics source, *Phys. Rev. D* **80**, 104008 (2009).
25. Z. Dayyani, et al., Critical behavior and phase transition of dilaton black holes with nonlinear electrodynamics, *Eur. Phys. J. C* **78**, 152 (2018).
26. S. Yu, C. Gao, Exact black hole solutions with non-linear electrodynamic field, *Int. J. Mod. Phys. D* **29**, 2050032 (2020).
27. W. Javed, J. Abbas, A. Ovgun, Deflection angle of photon from magnetized black hole and effect of nonlinear electrodynamics, *Eur. Phys. J. C* **79**, 1 (2019); W. Javed, A. Hamza, A. Ovgun, Effect of non-linear electrodynamics on weak field deflection angle by black hole, *Phys. Rev. D* **101**, 103521 (2020).
28. M. I. Fauzi, H. S. Ramadhan, Nonlinear electrodynamics black holes: Study of thermodynamic properties, *AIP Conf. Proc.* **2234**, 040009 (2020).
29. S. Das, et al., General logarithmic corrections to black-hole entropy, *Class. Quantum Grav.* **19**, 2355 (2002).
30. B. Pourhassan, M. Faizal, Thermodynamics of a sufficient small singly spinning Kerr-AdS black hole, *Nucl. Phys. B* **913**, 834 (2016).
31. B. Cuadros-Melgar, et al., Late-time tails, entropy aspects, and stability of black holes with anisotropic fluids, *Eur. Phys. J. C* **80**, 848 (2020).

Rindler trajectories and Rindler horizons in the Schwarzschild spacetime

Kajol Paithankar
UM-DAE Centre for Excellence in Basic Sciences, Mumbai 400098, India
E-mail: kajol.paithankar@cbs.ac.in

Sanved Kolekar
Indian Institute of Astrophysics, 2nd Block, Koramangala, Bengaluru 560034, India
E-mail: sanved.kolekar@iiap.res.in

We investigate radial linear uniformly accelerated trajectories and their corresponding Rindler horizons in the black hole geometry. In a curved spacetime, a covariant definition for Rindler trajectories is provided in the context of the generalised Letaw-Frenet equations for trajectories with constant curvature scalar and vanishing torsion and hypertorsion. Interestingly, we arrive at a bound on magnitude of acceleration for Rindler trajectories such that, for acceleration greater than the bound value, the Rindler trajectory always falls into the black hole and the distance of closest approach for the trajectory to turn away is always greater than the Schwarzschild radius for all finite boundary data. We further investigate the past and future Rindler horizons using the analytical solution for the trajectories and discuss their features.

Keywords: Rindler trajectory, Schwarzschild spacetime, Letaw-Frenet equations.

1. Introduction

In a flat spacetime, Rindler trajectories are a special class of hyperbolic trajectories, moving along orbits of the boost Killing vector and lying in the right or left Rindler quadrant constrained by the future and past null Rindler horizons.[1] The investigation of quantum fields in this background leads to the Unruh effect, which states that the Minkowski vacuum appears as thermal to Rindler observer.[2] It has also been demonstrated that the Rindler observer in an inertial thermal bath with temperature T_B is unable to distinguish between the background thermal and quantum fluctuations as the corresponding reduced density matrix turns out to be symmetric in the two temperatures, the inertial bath temperature T_B and the acceleration temperature T_{unruh}.[3] An analogous case to be investigated would be in curved spacetime case of a black hole with Hawking radiation and an uniformly accelerated observer with arbitrary acceleration.

In this talk, we obtain a solution for radial Rindler trajectories in the Schwarzschild spacetime. In a curved spacetime, the flat spacetime hyperbolic trajectories are replaced by linearly uniformly accelerated (LUA) trajectories which are

locally hyperbolic at every point along the trajectory. We use the covariant definition in the context of the generalised Letaw-Frenet equations in curved spacetime for trajectories with constant curvature scalar and vanishing torsion and hyper-torsion. The accompanying Rindler horizons and the Rindler quadrant is expected to get modified in this case due to the presence of background curvature. We determine the causal past and future of the intercepts of the LUA trajectory at future and past null infinity respectively to analyse the corresponding Rindler horizons in the Schwarzschild background for such trajectories.

2. Rindler trajectories in Schwarzschild spacetime

In a curved spacetime, a generalisation of the Rindler trajectory involves, in addition to the constancy condition on the magnitude of acceleration, a further constraint on the trajectory that it must be *linear* with vanishing torsion and hyper-torsion. Such a construction is based on the generalised Letaw-Frenet equations in curved spacetime and the corresponding geometrical scalar invariants which lead to a covariant definition of the linear uniformly accelerated (LUA) trajectory.[4,5] For the LUA trajectory, the Letaw-Frenet equations take the following form of a constraint equation

$$w^i - |a|^2 u^i = 0 \tag{1}$$

where $w^i = u^j \nabla_j a^i$ with u^i and a^i being the velocity and acceleration vectors along the trajectory.

For the Schwarzschild spacetime,

$$ds^2 = \left(1 - \frac{r_s}{r}\right) dt^2 - \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 - r^2\, d\theta^2 - r^2 \sin^2\theta\, d\phi^2 \tag{2}$$

we consider radial trajectories which start from the boundary $r \to \infty$, travel towards the black hole and then turn back to infinity. The equation for radial LUA trajectory consistent with the constraint equation is

$$\frac{dt}{dr} = \pm\Big(1 - \frac{r_s}{r}\Big)^{-1} \frac{(|a|\, r + h)\, \sqrt{r}}{|a|\, \sqrt{(r - r_{min})(r - r_{max})(r - r_n)}} \tag{3}$$

where $|a|$ is the constant magnitude of acceleration, h specifies the initial data at spatial infinity which accounts for the nonlinear shift in the trajectory along the radial direction, and r_{min}, r_{max} and r_n are the roots of the cubic polynomial, $r\,(|a|r + h)^2 - r + r_s$. The roots r_{min} and r_{max} are the turning points of the LUA trajectories initially moving towards and away from the black hole respectively. The root r_n being negative does not have any physical significance.

These turning points are positive real only when the magnitude of acceleration satisfies the following bound $|a|_b$

$$|a| \leq |a|_b = \frac{2\left(-9\,h + h^3 + \sqrt{(3 + h^2)^3}\right)}{27\, r_s}. \tag{4}$$

A LUA trajectory violating the acceleration bound always falls into the black hole. The turning point r_{min} corresponding to $|a|_b$ gives the lower bound on the distance of closest approach r_b for the return trajectory,

$$r_b = \frac{2}{3|a|_b}\left(\frac{\sqrt{3+h^2}}{2} - h\right). \tag{5}$$

The bound $|a|_b$ has a positive real value only when $h < 1$. Hence, altogether, in the Schwarzschild spacetime for a radial LUA trajectory to turn back at r_{min} it should have initial data $h < 1$ and magnitude of acceleration $|a| < |a|_b$. For all finite boundary data $h < 1$, the distance of closest approach r_b is found to be greater than the Schwarzschild radius r_s.

3. Rindler horizons in Schwarzschild spacetime

For a radially inward moving LUA trajectory satisfying all the conditions to escape back to infinity, the explicit solution $t(r)$ is obtained by integrating Eq. (3) in terms of the elliptic integrals (For details, see[6]). The constant of integration is fixed by choosing $t = 0$ when the trajectory is at its turning point $r = r_{min}$. The nature of trajectories is well explained by this solution in whole spacetime except at radial infinity where the elliptic integral of third kind, encountered in solution, is ill-defined.

This problem is circumvented by re-deriving the solution $t(r)$ as a series expansion. For a LUA trajectory starting from $r_i > r_{min}$, the relevant factors appearing in the denominator of Eq. (3) are expanded around r_{max} and r_s to get

$$\frac{dt}{dr} = \pm\frac{1}{|a|}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{-1}{i}\binom{-1/2}{j}\frac{(-r_s)^i(-r_{max})^j}{(r)^{i+j}}\frac{(|a|\,r+h)}{\sqrt{(r-r_{min})(r-r_n)}} \tag{6}$$

where the notation $\binom{x}{y}$ refers to binomial coefficients. Integrating this equation we get the solution $t(r)$ in terms of the Appell Hypergeometric functions $\mathbf{F_1}$ as,

$$\begin{aligned} t(r) = \pm\frac{2}{|a|}\sqrt{\frac{r_{min}\,(r-r_{min})}{r\,(r_{min}-r_n)}}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}G(i,j) \\ \left(|a|\,(r_{min}-r_n)\,\mathbf{F_1}\left[\frac{1}{2}, 2-i-j, \frac{-1}{2}, \frac{3}{2}; \frac{r-r_{min}}{r}, \frac{-r_n\,(r-r_{min})}{r\,(r_{min}-r_n)}\right]\right. \\ \left. + (h+|a|\,r_n)\,\mathbf{F_1}\left[\frac{1}{2}, 1-i-j, \frac{1}{2}, \frac{3}{2}; \frac{r-r_{min}}{r}, \frac{-r_n\,(r-r_{min})}{r\,(r_{min}-r_n)}\right]\right) \end{aligned} \tag{7}$$

where the coefficients $G(i,j)$ are defined to be

$$G(i,j) = \binom{-1}{i}\binom{-1/2}{j}\left(\frac{-r_s}{r_{min}}\right)^i\left(\frac{-r_{max}}{r_{min}}\right)^j. \tag{8}$$

This solution satisfies the boundary condition $t(r_{min}) = 0$ and has a well defined asymptotic value denoted by $\mathcal{C}$. Using the properties of Hypergeometric functions

the solution $t(r)$ for $r \to \infty$ is written in the form, $t(r) = \pm\,(r + r_s\,\log\,(r/r_s) + \mathcal{C})$, with $\mathcal{C}$ determined to be

$$\mathcal{C} = \frac{h}{|a|} + \frac{r_{max}}{2} + \frac{h\,(r_{max} + 2\,r_s)}{|a|\,\sqrt{-r_n\,r_{min}}}\,\sin^{-1}\left(\sqrt{\frac{-r_n}{r_{min} - r_n}}\right) + r_s\,\log\left(\frac{4\,r_s}{r_{min} - r_n}\right) + \frac{\sqrt{\pi}}{|a|}\,\sqrt{\frac{r_{min}}{r_{min} - r_n}}\,(S_d + S_s) \tag{9}$$

where S_d and S_s are double and single convergent summation series respectively expressed as,

$$S_d = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} G(i,j)\;\left[\,|a|\,(r_{min} - r_n)\,H_1(i+j) + (h + |a|\,r_n)\,H_2(i+j)\,\right] \tag{10}$$

$$S_s = \sum_{k=2}^{\infty}\,(\,G(k,0) + G(0,k)\,)\;\left[\,|a|\,(r_{min} - r_n)\,H_1(k) + (h + |a|\,r_n)\,H_2(k)\,\right] \tag{11}$$

with the functions being defined as

$$H_1(k) = \frac{\Gamma(k-1)}{\Gamma\left(k - \frac{1}{2}\right)}\;\mathbf{F}\left[\frac{1}{2}, \frac{-1}{2}, k - \frac{1}{2}; \frac{-r_n}{r_{min} - r_n}\right]$$

$$H_2(k) = \frac{\Gamma(k)}{\Gamma\left(k + \frac{1}{2}\right)}\;\mathbf{F}\left[\frac{1}{2}, \frac{1}{2}, k + \frac{1}{2}; \frac{-r_n}{r_{min} - r_n}\right]$$

The LUA trajectory asymptotes to null trajectories near radial infinity as expected, since away from the black hole the scenario becomes asymptotically flat. The asymptotic expansion of LUA trajectory $t(r)$ matches to the form of the null trajectories representing the future and past Rindler horizons,

$$r + r_s \log\left|\frac{r}{r_s} - 1\right| \mp t = \mathcal{C}^{\pm}$$

where we then have the intercepts at radial infinity, $\mathcal{C}^+ = \mathcal{C}^- = \mathcal{C}$ due to the symmetry about the $t = 0$ axis.

The Rindler quadrant for the radial LUA trajectories in the Schwarzschild spacetime is then the union of the past and future null infinity with the past and future horizons in the Penrose diagram. The intersection points of these are the spatial infinity i^0, the past and future intercepts $\mathcal{C}^-$ and $\mathcal{C}^+$ and the bifurcation point of the past and future horizons, which is found to be,

$$r_{null} = r_s\left(1 + W\left[e^{-1-\frac{\mathcal{C}}{r_s}}\right]\right) \tag{12}$$

where $W[x]$ is the productlog function. Thus, unlike in the case of the Rindler quadrant in the flat spacetime, for the Schwarzschild case, the bifurcation point r_{null} is dependent on the asymptotic initial data h as well as the acceleration magnitude $|a|$ and the Schwarzschild radius r_s.

4. Discussion

For the case of Schwarzschild black hole the following observations are made, in comparison with flat spacetime, regarding the Rindler trajectories: In the flat spacetime, all the Rindler trajectories with the magnitude of acceleration in range 0 to ∞ lie in the same Rindler quadrant. This is due to the reason that the corresponding intercept $\mathcal{C}$, the Rindler horizons and the bifurcation point are independent of $|a|$ and just depend on h. In fact, for the limiting case $|a| \to \infty$, the turning point r_{min} is the bifurcation point as the Rindler trajectory coincides with the past and future null horizons. However, in the Schwarzschild case, the intercepts $\mathcal{C}$, the corresponding Rindler horizons and the bifurcation points are all functions of $|a|$ as well. Hence, increasing $|a|$ decreases not just the turning point r_{min} but also the bifurcation point r_{null}. However, it turns out that the turning point does not catch up with the bifurcation point. These observations provide an alternate perspective to look at the acceleration bounds $|a|_b$. The lowest possible value for r_{null} is limited by the black hole horizon to being the Schwarzschild radius r_s, which in turn limits the lowest possible value of r_{min} to $r_{min} = r_b$ and hence a maximum value for the acceleration $|a|$.

References

1. W. Rindler, Hyperbolic Motion in Curved Space Time, *Phys. Rev.* **119**, 2082 (1960).
2. W. G. Unruh, Notes on black-hole evaporation, *Phys. Rev. D* **14**, 870 (1976).
3. Sanved Kolekar and T. Padmanabhan, Indistinguishability of thermal and quantum fluctuations, *Class. Quant. Grav.* **32**, 202001 (2015).
4. J. Letaw, Stationary world lines and the vacuum excitation of noninertial detectors, *Phys. Rev. D* **23**, 1709 (1981).
5. Sanved Kolekar and Jorma Louko, Gravitational memory for uniformly accelerated observers, *Phys. Rev. D* **96**, 024054 (2017).
6. Kajol Paithankar and Sanved Kolekar, Bound on Rindler trajectories in a black hole spacetime, *Phys. Rev. D* **99**, 064012 (2019).
7. Kajol Paithankar and Sanved Kolekar, Rindler horizons in a Schwarzschild spacetime, *Phys. Rev. D* **100**, 084029 (2019).

Linear growth of the two-point function for the Unruh state in $1+1$ dimensional black holes

Paul R. Anderson* and Zachary P. Scofield
Department of Physics, Wake Forest University, Winston-Salem, NC 27109, USA
** E-mail: anderson@wfu.edu*

Jennie Traschen
Department of Physics, University of Massachusetts, Amherst, MA 01003, USA
E-mail: traschen@umass.edu

The symmetric two-point function for a massless, minimally coupled scalar field in the Unruh state is examined for Schwarzschild-de Sitter spacetime in two dimensions. This function grows linearly in terms of a time coordinate that is well-defined on the future black hole and cosmological horizons, when the points are split in the space direction. This type of behavior also occurs in two dimensions for other static black hole spacetimes when the field is in the Unruh state, and at late times it occurs in spacetimes where a black hole forms from the collapse of a null shell. The generalization to the case of the symmetric two-point function in two dimensions for a massive scalar field in Schwarzschild-de Sitter spacetime is discussed.

Keywords: Black hole, quantum field theory in curved space

1. Introduction

Hawking[1] predicted that black holes that form from collapse evaporate and at late times the particles are in a thermal distribution with a temperature that is proportional to the surface gravity of the black hole. In four dimensions (4D) it is very difficult to compute quantities such as the two-point function and the stress-energy tensor for a quantum field in the spacetime of a black hole that forms from collapse. Instead, it is significantly easier, although not easy, to do such computations in eternal black hole spacetimes which are either static or stationary. It is easier yet to work with a massless minimally coupled scalar field in spacetimes with two dimensions (2D) that have similar structures to the 4D black hole spacetimes.

Unruh[2] showed that, for an isolated eternal black hole in an asymptotically flat spacetime, there is a particular state for a quantum field that has the same properties as the *in* state for that field has at late times in a spacetime where the black hole forms from collapse. In particular, this state, called the Unruh state, has the same flux of particles at infinity for an eternal black hole as occurs at late times for a black hole that forms from collapse. Thus computations of various quantities for a quantum field in the Unruh state can give insight into the late time behaviors of these quantities in spacetimes where a black hole forms from collapse.

The simplest 4D black hole spacetimes to compute quantum effects in are those which are static and spherically symmetric outside of the event horizon where the metric has the general form

$$ds^2 = -f(r)dt^2 + f^{-1}(r)dr^2 + r^2 d\Omega^2 \ . \tag{1}$$

These are easily changed to 2D spacetimes by dropping the last term on the right. All 2D spacetimes are conformally flat and the massless, minimally coupled scalar field is conformally invariant in such spacetimes making it possible to analytically solve the mode equation in coordinates for which the metric is conformally flat.

In Ref. 3, the Hadamard Green's function, which is the symmetric two-point function, was computed and studied for a massless, minimally coupled scalar field in the Unruh state in 2D for Schwarzschild-de Sitter(SdS) spacetime, Schwarzschild spacetime, de Sitter space, a class of Bose-Einstein condensate analog black hole spacetimes, and a spacetime in which a null shell collapses to form a black hole. It was shown that in all of these cases there is linear growth in terms of a time coordinate T that is regular on the future horizon(s) when the points are split in the spatial direction.

In this proceeding we summarize some of the results of Ref. 3 for black holes. We also discuss the computation of the symmetric two-point function for a massive, minimally coupled scalar field in SdS. In this case, the mode functions and the two-point function must be computed numerically.

In Sec. 2, some properties of 2D SdS are reviewed along with the Unruh state for a massless, minimally coupled scalar field. The Hadamard Green's function for this field is displayed in Sec. 3 and its linear growth in time is discussed. Ongoing work related to the massive scalar field in 2D SdS is discussed in Sec. 4. In Sec. 5 the computation of the symmetric two-point function for a massless, minimally coupled scalar field in a 2D collapsing null shell spacetime is reviewed. Our results are summarized in Sec. 6. Throughout, our units are $\hbar = c = G = 1$.

2. Massless, Minimally Coupled Scalar Field in SdS

A Penrose diagram for SdS is given in Fig. 1. The 2D SdS metric is given by

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} \ , \tag{2a}$$

$$f(r) = 1 - \frac{2M}{r} - H^2 r^2 = \ -\frac{H^2}{r}(r - r_c)(r - r_b)(r + r_c + r_b) \ . \tag{2b}$$

Here $r_c > r_b$ are the locations of the cosmological and black hole horizons respectively, and $\frac{1}{3}\Lambda = H^2$. The two parameterizations are related by

$$M = \frac{r_b r_c (r_b + r_c)}{2(r_b^2 + r_c^2 + r_b r_c)}, \qquad H^2 = \frac{1}{r_b^2 + r_c^2 + r_b r_c} \tag{3}$$

It is useful to define the tortoise coordinate

$$r_*(r) = \int \frac{dr}{f} \ . \tag{4}$$

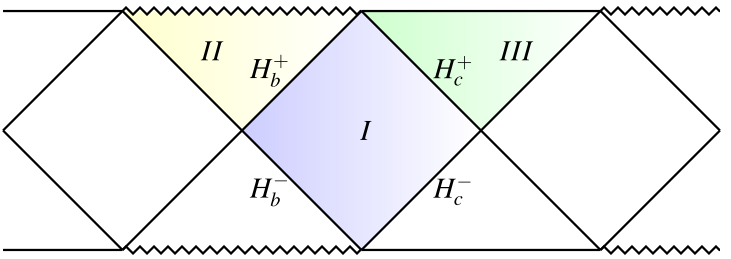


Fig. 1. Penrose diagram for SdS. Region I is the static patch between the black hole horizons and cosmological horizons, Region II is the interior of the black hole, and Region III is the cosmological region.

The ingoing and outgoing radial null coordinates are

$$u = t - r_* \ , \quad v = t + r_* \ . \tag{5}$$

One can define two sets of Kruskal coordinates, one for each horizon. The ones we are concerned with are

$$\begin{aligned} U_b &= \frac{1}{\kappa_b} e^{-\kappa_b u} \ , \quad r < r_b \ , \quad U_b = -\frac{1}{\kappa_b} e^{-\kappa_b u} \ , \quad r > r_b \\ V_c &= -\frac{1}{\kappa_c} e^{-\kappa_c v} \ , \quad r < r_c \ , \qquad V_c = \frac{1}{\kappa_c} e^{-\kappa_c v} \quad , \quad r > r_c \ . \end{aligned} \tag{6}$$

The surface gravities for the two horizons are

$$\begin{aligned} \kappa_b &= \frac{H^2}{2r_b}(r_c - r_b)(r_c + 2r_b) \ , \\ \kappa_c &= \frac{H^2}{2r_c}(r_c - r_b)(2r_c + r_b) \ , \end{aligned} \tag{7}$$

where for a horizon at $r = r_h$ we set $2\kappa_h = |f'(r_h)|$ so that each of the surface gravities is a positive quantity.

A set of coordinates that are good across $\mathcal{H}_b^+$ and $\mathcal{H}_c^+$ was found in Refs. 4, 5. Let

$$T = t + h(r) \ , \quad \text{where} \quad \frac{dh}{dr} = \frac{j}{f} \ , \qquad j(r) = -\gamma r + \frac{\beta}{r^2} \ , \tag{8}$$

and

$$\gamma = \frac{r_c^2 + r_b^2}{r_c^3 - r_b^3} \ , \quad \beta = \frac{r_c^2 r_b^2 (r_b + r_c)}{r_c^3 - r_b^3} \ . \tag{9}$$

Then the 2D SdS metric becomes

$$ds^2 = -f(r)dT^2 + 2j(r)drdT + \frac{1 - j^2}{f} dr^2 \ . \tag{10}$$

Note that for these choices of the constants γ and β, one finds that $j(r_b) = 1$ and $j(r_c) = -1$, which ensures that T interpolates between the ingoing null coordinate v at the future black hole horizon and the outgoing null coordinate u at the future cosmological horizon. These are ingoing and outgoing Eddington-Finklestein coordinates respectively. The metric (10) is well behaved in the static and cosmological regions and has the Eddington-Finklestein form on both the future black hole and future cosmological horizons. The coordinate T stays timelike and r stays spacelike beyond the cosmological horizon.

The definitions of r_* and T each contain an arbitrary constant which can be chosen such that

$$T = u \text{ on } \mathcal{H}_c^+ \ , \qquad \text{and} \quad T = v \text{ on } \mathcal{H}_b^+ \ . \tag{11}$$

The result is[3]

$$\begin{aligned} T &= t + h(r) \\ &= t + \frac{1}{2\kappa_b}\log\frac{|r-r_b|}{r_c-r_b} + \frac{1}{2\kappa_c}\log\frac{|r-r_c|}{r_c-r_b} + \frac{1}{2}\left(\frac{r_c}{r_b\kappa_b} - \frac{1}{\kappa_N}\right)\log\frac{r+r_c+r_b}{r_c+2r_b} \\ &\quad - \frac{r_b r_c}{2(r_c-r_b)}\log\frac{r^2}{r_c r_b} + \frac{r_c}{4r_b\kappa_b}\log\frac{r_c+2r_b}{2r_c+r_b} \ , \end{aligned} \tag{12}$$

and

$$\begin{aligned} r_*(r) &= \frac{1}{2\kappa_b}\log\frac{|r-r_b|}{r_c-r_b} - \frac{1}{2\kappa_c}\log\frac{|r-r_c|}{r_c-r_b} + \frac{1}{2\kappa_N}\log\frac{|r+r_c+r_b|}{r_c+2r_b} \\ &\quad - \frac{r_c}{4r_b\kappa_b}\log\frac{2r_c+r_b}{r_c+2r_b} - \frac{r_b r_c}{2(r_c-r_b)}\log\frac{r_b}{r_c} \ , \end{aligned} \tag{13}$$

where

$$\kappa_N = \frac{H^2}{2(r_c+r_b)}(2r_c+r_b)(r_c+2r_b) \ . \tag{14}$$

To relate the Kruskal coordinates to T and r, note that $u = T - (h(r) + r_*)$ and $v = T - h(r) + r_*$. Substitution gives

$$V_c = \frac{1}{\kappa_c}e^{-\kappa_c T}\tilde{V}_c \ , \quad \text{and} \quad U_b = \frac{1}{\kappa_b}e^{-\kappa_b T}\tilde{U}_b \ , \tag{15}$$

where $\tilde{V}_c$ and $\tilde{U}_b$ only depend on r and are given by

$$\begin{aligned} \tilde{V}_c &= e^{\kappa_c(h-r_*)} = \frac{r-r_c}{r_c-r_b}\left(\frac{r+r_c+r_b}{r_c+2r_b}\right)^{r_b/2r_c}\left(\frac{r_b}{r}\right)^{H^2 r_b(2r_c+r_b)/2} \ , \\ \tilde{U}_b &= -e^{\kappa_b(h+r_*)} = -\frac{(r-r_b)}{r_c-r_b}\left(\frac{r+r_b+r_c}{2r_c+r_b}\right)^{r_c/2r_b}\left(\frac{r_c}{r}\right)^{H^2 r_c(r_c+2r_b)/2} \ . \end{aligned} \tag{16}$$

Note that these expressions work both inside and outside the black hole and cosmological horizons.

The complete set of modes that comprise the Unruh state consist of modes that are positive frequency with respect to U_b on the past black hole horizon and modes that are positive frequency with respect to V_c on the past cosmological horizon.[6–8]

$$p^b_\omega = \frac{e^{-i\omega U_b(u)}}{\sqrt{4\pi\omega}} \text{ and } p^c_\omega = \frac{e^{-i\omega V_c(v)}}{\sqrt{4\pi\omega}} \tag{17}$$

Expanding the massless minimally coupled scalar field in terms of these gives

$$\phi = \int_0^\infty d\omega \left[a^b_\omega p^b_\omega + a^c_\omega p^c_\omega + a^{b\dagger}_\omega p^{b*}_\omega + a^{c\dagger}_\omega p^{c*}_\omega\right] , \tag{18}$$

where $a^{(b,c)}$ are the annihilation operators.

3. Symmetric Two-Point Function in SdS

For a massless, minimally coupled scalar field in 2D SdS in the Unruh state, the symmetric two-point function for an arbitrary splitting of the points is[3]

$$\begin{aligned} G^{(1)}(x,x') &= \langle U| \left(\phi(x)\phi(x') + \phi(x')\phi(x)\right) |U\rangle \\ &= \int_0^\infty d\omega [p^b_\omega(x) p^{b*}_\omega(x') + p^c_\omega(x) p^{c*}_\omega(x') + \text{c.c.}] \\ &= \frac{1}{2\pi}\int_{\omega_0}^\infty \frac{d\omega}{\omega}\left\{\cos[\omega(U_b - U_b^{'})] + \cos[\omega(V_c - V_c^{'})]\right\} \\ &= -\frac{1}{2\pi}\left\{ci[\omega_0|U_b - U_b^{'}|] + ci[\omega_0|V_c - V_c^{'}|]\right\} . \end{aligned} \tag{19}$$

Here ci is the cosine integral function which has the expansion $ci(z) = \gamma_E + \log z + O(z^2)$, and ω_0 is a small infrared cutoff.

Let x be on H^b_+ and x' on H^c_+ and take $T' = T$. Note that on H^+_b, $U_b = 0$, $v = T$ and on H^+_c, $V_c = 0$, $u = T$. Then

$$\begin{aligned} U_b - U_b^{'} &= -\kappa_b^{-1} e^{-\kappa_b T} , \qquad V_c - V_c^{'} = \kappa_c^{-1} e^{-\kappa_c T} , \\ 2\pi G^{(1)}(T,r;T,r') &= (\kappa_b + \kappa_c)T - \log\left[(\kappa_b\kappa_c)^{-1}\omega_0^2\right] - 2\gamma_E . \end{aligned} \tag{20}$$

Thus one sees an unexpected linear growth in the time coordinate T for this separation of the points. For a more general separation

$$2\pi G^{(1)}(T,r;T,r') = T(\kappa_b + \kappa_c) - \log\left(\frac{\omega_0^2}{\kappa_b\kappa_c}|\Delta\tilde{U}_b \Delta\tilde{V}_c|\right) - 2\gamma_E , \tag{21}$$

where $\Delta\tilde{U}_b$ and $\Delta\tilde{V}_c$ are functions of r and r', but not T.

In Ref. 3 a similar linear growth in a time coordinate for $G^{(1)}(x,x')$ in 2D for a massless minimally coupled scalar field was found in Schwarzschild spacetime for the Unruh state, de Sitter spacetime for the Bunch-Davies state, and a class of Bose-Einstein condensate analog black hole spacetimes for the Unruh state. A general argument was also given that linear growth should occur whenever there is a past horizon and no scattering for the Unruh state.

One may ask what these situations have in common. In each case the symmetric two-point function is computed for a massless minimally coupled scalar field in 2D in the Unruh state or its analog, the two-point function has an infrared divergence which requires an infrared cutoff, and each spacetime has a static patch and at least one past horizon. In the next two sections, two cases are discussed in which one or more of these conditions does not occur.

4. Massive Scalar Field in 2D SdS

One way to investigate the generality of the linear growth in time is to work with a massive scalar field. Work is in progress to compute the symmetric two-point function for this field. It is significantly more difficult than for the massless scalar field because the mode equation contains an effective potential term and does not have simple closed form solutions. Instead, when separation of variables is used, the radial part of the mode equation can be solved numerically.

The mode equation cannot be separated in Kruskal coordinates but it can be separated in terms of the coordinates t and r in (2b). The mode equation in this case is

$$-\frac{\partial^2 h}{\partial t^2}+\frac{\partial^2 h}{\partial r_*^2}-m^2 f=0\ . \tag{22}$$

The relevant solutions are of the form

$$h(t,r)=\frac{e^{-i\omega t}}{\sqrt{4\pi\omega}}\chi_\omega(r)\ , \tag{23}$$

with

$$\frac{d^2\chi}{dr_*^2}+(\omega^2-m^2 f)h=0\ . \tag{24}$$

A complete set of modes that specify the Boulware state can be obtained by combining modes which on the past black hole horizon have the form

$$h^b_\omega=\frac{e^{-i\omega u}}{\sqrt{4\pi\omega}}\ , \tag{25}$$

and those which on the past cosmological horizon have the form

$$h^c_\omega=\frac{e^{-i\omega v}}{\sqrt{4\pi\omega}}\ . \tag{26}$$

The modes that specify the Unruh state can be expanded in terms of these modes with the result

$$p^{(b,c)}_\omega=\int_0^\infty d\omega'\,[\alpha^{(b,c)}_{\omega\omega'}h^{(b,c)}_{\omega'}+\beta^{(b,c)}_{\omega\omega'}h^{(b,c)\,*}_{w'}]\ . \tag{27}$$

Since the mode equation reduces to its massless form on the past horizons, the initial conditions for the Unruh state are exactly the same as in the massless case (17).

Using the usual scalar product,[9] one finds[10]

$$\alpha^{(b,c)}_{\omega\omega'} = \frac{1}{2\pi}\sqrt{\frac{\omega'}{\omega}}\left(\kappa_{(b,c)}\right)^{-1-i\frac{\omega'}{\kappa_{(b,c)}}}\frac{\Gamma(-i\frac{\omega'}{\kappa_{(b,c)}})}{(-i\omega)^{-i\frac{\omega'}{\kappa_{(b,c)}}}} ,$$
$$\beta^{(b,c)}_{\omega\omega'} = \frac{1}{2\pi}\sqrt{\frac{\omega'}{\omega}}\left(\kappa_{(b,c)}\right)^{-1+i\frac{\omega'}{\kappa_{(b,c)}}}\frac{\Gamma(i\frac{\omega'}{\kappa_{(b,c)}})}{(-i\omega)^{i\frac{\omega'}{\kappa_{(b,c)}}}} . \tag{28}$$

Note that these are divergent in the limit $\omega' \to 0$ due to the poles in the Gamma functions. However, we have shown that because of scattering effects, these divergences are canceled in the integrand of Eq. (27). Also, it is important to note that the limit $\omega' \to 0$ needs to be taken before the limits $r \to r_b$ and $r \to r_c$. They cannot be interchanged.

At present, work is in progress to compute p^b_ω and p^c_ω numerically. It is not yet known whether the divergence that exists for the massless case for these mode functions continues to be present in the massive case.

5. Collapsing Null Shell Spacetime

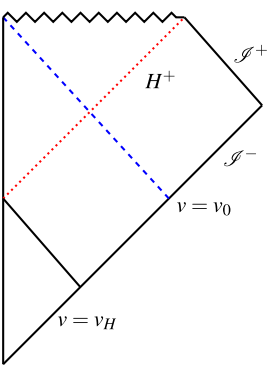


Fig. 2. Penrose diagram for a 2D spacetime in which a null shell collapses to form a black hole and there is a perfectly reflecting mirror at the origin. The trajectory of the shell is given by the dashed blue line and the event horizon is given by the dotted red line.

Another calculation that was done in Ref. 3 was to compute the two-point function for a massless scalar field in a 2D spacetime in which a null shell collapses to form a Schwarzschild black hole. Inside the shell the metric is the flat space metric

$$ds^2 = -dt^2 + dr^2 , \tag{29}$$

and outside it is Eq. (2b) with $H = 0$, which is the Schwarzschild metric. If one puts in a perfectly reflecting mirror and requires that the field vanish at the surface

of the mirror then one obtains the same conditions on the mode functions at $r = 0$ as in the 4D case. The Penrose diagram for this spacetime is shown in Fig. 2.

Inside the shell, the usual null coordinates are

$$u = t - r \ , \qquad v = t + r \ , \tag{30}$$

and outside the shell they are

$$u_s = t_s - r_* \ , \qquad v = t_s + t_* \ . \tag{31}$$

The coordinates r and v are continuous across the trajectory of the shell which is $v = v_0$, for some v_0. At the null shell surface one finds[11,12]

$$u(u_s) = v_H - 4MW\left[\exp\left(\frac{v_H - u_s}{4M}\right)\right] \ , \tag{32}$$

where $v_H = v_0 - 4M$ and W is the Lambert W function. On past null infinity both u_s and u are equal to $-\infty$. On the future horizon of the black hole $u_s = \infty$ and $u = v_H$.

Because of the mirror, the mode functions must vanish at $r = 0$. For the *in* state, at past null infinity they have the form

$$f_\omega^{\rm in} \to \frac{e^{-i\omega v}}{\sqrt{4\pi\omega}} \ . \tag{33}$$

The solution that satisfies these two conditions is

$$f_\omega^{\rm in} = \frac{1}{\sqrt{4\pi\omega}}(e^{-i\omega v} - e^{-i\omega u}) \ . \tag{34}$$

Since there is no scattering in 2D for the massless, minimally coupled scalar field, outside the shell the solution is

$$f_\omega^{\rm in} = \frac{1}{\sqrt{4\pi\omega}}(e^{-i\omega v} - e^{-i\omega u(u_s)}) \ . \tag{35}$$

Note that $f_\omega^{\rm in} = 0$ at $\omega = 0$. Thus there is no infrared divergence in the two-point function for this state and no infrared cutoff is necessary. One finds that[3]

$$2\pi G^{(1)}(x, x') = \log\left[\frac{|v - u(u_s')||v' - u(u_s)|}{|v - v'||u(u_s) - u(u_s')|}\right] \ . \tag{36}$$

Using $v = T - h(r) + r_*$ and $u = T - h(r) - r_*$, one finds for $T = T'$ and fixed values of r and r', that at late times

$$2\pi G(T, r; , T, r') \to \kappa T + \log(\kappa^2 T^2) + (r, r' \text{dependent terms}) \ . \tag{37}$$

Thus, there is not only a linear growth in T as there is for the Unruh state, but there is also a secondary logarithmic growth.

6. Conclusions

In Ref. 3, the behavior of the two point function was investigated for a massless, minimally coupled scalar field in the Unruh state in certain 2D spacetimes. It was found that linear growth in time of the form

$$2\pi G^{(1)}(T, r; T, r') = T \sum_J \kappa_J + (r, r' \text{dependent terms}) \tag{38}$$

occurs, where the sum is over the surface gravities of the past horizons and T is a time coordinate that is regular on the future horizons. Along with the existence of one or two past horizons, the Green's function for each spacetime is infrared divergent and thus requires an infrared cutoff to regularize this divergence.

The two-point function for this field in a 2D spacetime in which a null shell collapses to form a black hole was also computed. In this case there is no past horizon and thus no Unruh state. Instead, since the spacetime is asymptotically flat, the state for the massless, minimally coupled scalar field was chosen to be the initial vacuum state. By putting a perfectly reflecting mirror at the origin and requiring that the mode functions vanish there, the same type of condition that one finds in the 4D collapsing null shell spacetime if the shell is spherically symmetric was imposed. As a result of this condition, the two-point function has no infrared divergences.

It was found that at early times there is no linear growth in time of the form (38), however at late times there is growth of this form. There is also a subleading term that diverges logarithmically in time at late times. Thus, the linear growth in time found in the eternal black hole calculations also occurs, at least in 2D, for a black hole that forms from the collapse of a null shell. Perhaps this is not surprising since the Unruh state is supposed to give the leading order, late time behavior of quantities such as the two-point function and stress-energy tensor for the quantum field in the case of a black hole that forms from collapse. However, it does show that the linear growth in time does not depend on the existence of either a past horizon or an infrared divergence in the two-point function.

To see whether the linear growth in time only occurs for the massless scalar field, the case of a massive scalar field in the Unruh state in 2D SdS is being investigated. It has been shown that, when the mode functions for the Unruh state are expanded in terms of those for the Boulware state, an infrared diverge that occurs in the massless case is not present in the massive one. Work is in progress to determine whether the linear growth in time of the two-point function also occurs for a massive scalar field.

Acknowledgments

P.R.A. and J.T. would like to thank David Kastor for useful conversations. P.R.A. would also like to thank Alessandro Fabbri and Ian Moss for useful conversations. This work came about as a result of conversations at the NORDITA Gravitational

Physics with Lambda Workshop 2018. We would like to thank NORDITA for their generous support. This work was supported in part by the National Science Foundation under Grants No. PHY-1505875 and PHY-1912584 to Wake Forest University.

References

1. S. W. Hawking, *Commun. Math. Phys.* **43**, 199 (1975), Erratum: [*Commun. Math. Phys.* **46**, 206 (1976)].
2. W. G. Unruh, *Phys. Rev. D* **14**, 870 (1976).
3. P. R. Anderson and J. Traschen, arXiv:2012.08494 [hep-th].
4. R. Gregory, D. Kastor and J. Traschen, JHEP **1710**, 118 (2017).
5. R. Gregory, D. Kastor and J. Traschen, *Class. Quant. Grav.* **35**, 155008 (2018).
6. S. Tadaki and S. Takagi, *Prog. Theor. Phys.* **83**, 941 (1990).
7. S. Tadaki and S. Takagi, *Prog. Theor. Phys.* **83**, 1126 (1990).
8. D. Markovic and W. G. Unruh, *Phys. Rev.* D **43**, 332 (1991).
9. N. D. Birrell and P. C. W. Davies, *Quantum fields in Curved Space* (Cambridge University Press, Cambridge, UK, 1982).
10. P. R. Anderson, S. Gholizadeh Siahmazgi, R. D. Clark, and A, Fabbri, *Phys. Rev.* D **102**, 125035 (2020).
11. S. Massar and R. Parentani, *Phys. Rev.* D **54**, 7444 (1996).
12. A. Fabbri and J. Navarro-Salas, *Modeling black hole evaporation* (Imperial College Press, London, UK, 2005).

Stress-energy tensor for a quantized scalar field in a four-dimensional black hole that forms from the collapse of a null shell

Shohreh Gholizadeh Siahmazg*, Paul R. Anderson† and Raymond D. Clark
Department of Physics, Wake Forest University,
Winston Salem, North Carolina 27109, USA
** E-mail: ghols18@wfu.edu*
† E-mail: anderson@wfu.edu

Alessandro Fabbri
Departamento de Física Teórica and IFIC, Universidad de Valencia-CSIC
C. Dr. Moliner 50, 46100 Burjassot
E-mail: afabbri@ific.uv.es

A method is presented which allows for the numerical computation of the stress-energy tensor for a quantized massless minimally coupled scalar field in the region outside the event horizon of a 4D Schwarzschild black hole that forms from the collapse of a null shell. This method involves taking the difference between the stress-energy tensor for the *in* state in the collapsing null shell spacetime and that for the Unruh state in Schwarzschild spacetime. The construction of the modes for the *in* vacuum state and the Unruh state is discussed. Applying the method, the renormalized stress-energy tensor in the 2D case has been computed numerically and shown to be in agreement with the known analytic solution. In 4D, the presence of an effective potential in the mode equation causes scattering effects that make the the construction of the in modes more complicated. The numerical computation of the *in* modes in this case is given.

Keywords: Black holes; Quantum field theory in curved space; Stress-energy tensor.

1. Introduction

The expectation value of the renormalized stress-energy tensor operator for a quantized field is a useful way to study quantum effects curved space. It can also be used in the context of semiclassical gravity to compute the backreaction of the quantum field on the background geometry. In the case of four-dimensional, 4D, black holes, this quantity has to date only been computed for static black holes[1–4,6–15] and Kerr black holes.[16,17] However, to our knowledge, a full numerical computation of this quantity has not been done for a quantized field in a 4D spacetime in which a black hole forms from the collapse of a null shell, which is probably the simplest model for the formation of a black hole.

In Ref. 18, we developed a method to numerically compute the full renormalized stress-energy tensor for a massless minimally coupled scalar field in the case of a spherically symmetric black hole in 4D that forms from the collapse of a null shell. This method can be used in the region outside the null shell and future horizon, where by Birkhoff's theorem, the geometry is described by the Schwarzschild metric.

In this proceeding, we review this method with a focus on the computation of a complete set of *in* modes that can be used to construct the quantum field in the region outside the null shell. We also present new numerical results for a low frequency *in* mode on the future horizon and for a mode with relatively high frequency on the part of the future horizon close to the null shell trajectory.

In Sec. 2, we review the null shell spacetime and the metrics describing the geometry inside and outside of the null shell. In Sec. 3, we discuss the quantization of the massless minimally coupled scalar field in the null shell spacetime. In Sec. 4, we present our method to expand the *in* modes in terms of a complete set of modes in pure Schwarzschild spacetime and present our numerical results for the high and low frequency modes on the future horizon. In Sec. 5, a proper method for the renormalization of the stress-energy tensor is given. In this section, we summarize the application of our method in Ref. 18 to the case of a collapsing null shell spacetime which has a perfectly reflecting mirror at the spatial origin.

2. Collapsing null shell

The model we consider is a spherically symmetric null shell whose collapse results in the formation of a black hole. The Penrose diagram of the spacetime is depicted in Fig. 1 The spacetime inside the null shell is described by the flat metric

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2 \,,$$

and from Birkhoff's theorem, the metric outside the shell is the Schwarzschild metric

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt_s^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2$$

with $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. It is more convenient to use radial null coordinates to match the geometries inside and outside of the shell. In the interior,

$$u = t - r \,, \quad v = t + r \,.$$

and in the exterior region,

$$u_s = t_s - r_* \,, \quad v = t_s + r_* \,,$$

where $r_* = r + 2M \ln\left(\frac{r-2M}{2M}\right)$ is the tortoise coordinate in Schwarzschild spacetime. The null shell trajectory is $v = v_0$. We match the two spacetimes so that the v coordinate and the angular coordinates are continuous across the null shell trajectory. Applying this condition gives the following relation between the u an u_s coordinates[19,20]

$$u_s = u - 4M \log\left(\frac{v_H - u}{4M}\right),$$

where $v_H = v_0 - 4M$.

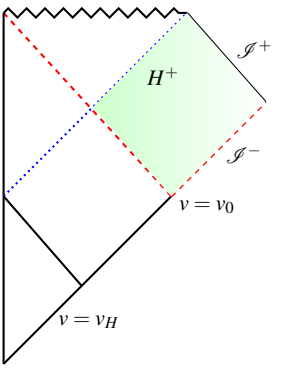


Fig. 1. Penrose diagram for a spacetime in which a null shell collapses to form a spherically symmetric black hole. The vertical line on the left corresponds to the surface $r = 0$ which is also the surface where $u = v$. The dashed red line on $v = v_0$ is the trajectory of the null shell. The horizon, H^+ is the dotted blue curve. Inside the shell trajectory H^+ corresponds to the surface $u = v_H$ and outside the shell trajectory it corresponds to $u_s = \infty$. A Cauchy surface is shown by the dashed line. It is the union of the surface $v = v_0$ with the part of $\mathscr{I}^-$ with $v > v_0$.

3. Massless minimally coupled scalar field

We consider a massless minimally coupled scalar field in the null shell spacetime. In a general static spherically symmetric spacetime, it can be expanded in the following way,

$$\phi = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \int_0^{\infty} [a_{\omega\ell m} f_{\omega\ell m} + a^{\dagger}_{\omega\ell m} f^*_{\omega\ell m}]$$

with $\Box\, f_{\omega\ell m} = 0$. In the region inside the null shell, $v < v_0$, separation of variables gives

$$f_{\omega\ell m} = \frac{Y_{\ell m}(\theta,\phi)}{r\sqrt{4\pi\omega}} \psi_{\omega\ell}(t,r) = \frac{Y_{\ell m}(\theta,\phi)}{r\sqrt{4\pi\omega}} e^{-i\omega t} \chi_{\omega\ell}(r), \tag{1}$$

while in the region outside the null shell, $v > v_0$, it gives

$$f_{\omega\ell m} = \frac{Y_{\ell m}(\theta,\phi)}{r\sqrt{4\pi\omega}} \psi_{\omega\ell}(t_s,r) = \frac{Y_{\ell m}(\theta,\phi)}{r\sqrt{4\pi\omega}} e^{-i\omega t_s} \chi_{\omega\ell}(r). \tag{2}$$

In the regions $v < v_0$ and $v > v_0$ respectively, the radial parts of the mode functions satisfy the differential equations

$$\frac{d^2\chi_{\omega\ell}}{dr^2} = -\left[\omega^2 - \frac{\ell(\ell+1)}{r^2}\right] \chi_{\omega\ell}. \tag{3}$$

$$\frac{d^2\chi_{\omega\ell}}{dr_*^2} = -\left[\omega^2 - \left(1 - \frac{2M}{r}\right)\left(\frac{2M}{r^3} + \frac{\ell(\ell+1)}{r^2}\right)\right] \chi_{\omega\ell}\,. \tag{4}$$

The *in* state is fixed by requiring that $\psi_{\omega\ell} = e^{-i\omega v}$ on past null infinity and it vanishes at $r = 0$. The solution with these properties has the form

$$\psi^{in}_{\omega\ell}(r,t) = C_\ell e^{i\omega t}\omega r j_\ell(\omega r) \tag{5}$$

inside the null shell, where C_ℓ is fixed by the aforementioned condition on past null infinity. Here j_ℓ is a spherical bessel function of the first kind. It is not possible for this solution to have the form $e^{-i\omega t_s}\chi_{\omega\ell}(r)$ outside the null shell. The solution in this region is more complicated.

4. Computation of $f^{in}_{\omega\ell m}$

We can compute $f^{in}_{\omega\ell m}$ outside the null shell and the event horizon by expanding it in terms of a complete set of modes since the geometry here is the Schwarzschild geometry. This problem can be mapped to the shaded part of pure Schwarzschild spacetime shown in Fig. 2. We choose a complete sets of modes that consists of the union of the modes $f^{H^+}_{\omega\ell m}$ that are positive frequency on future horizon and zero on future null infinity, and the modes $f^{\mathscr{I}^+}_{\omega\ell m}$ that are positive frequency on the future null infinity and zero on the future horizon, *i.e.*,

$$\begin{aligned} f^{in}_{\omega\ell m} = \sum_{\ell'=0}^{\infty}\sum_{m'=-\ell'}^{\ell'}\int_0^\infty d\omega' \Big[& A^{\mathscr{I}^+}_{\omega\ell m\omega'\ell'm'}\, f^{\mathscr{I}^+}_{\omega'\ell'm'} + B^{\mathscr{I}^+}_{\omega\ell m\omega'\ell'm'}\, (f^{\mathscr{I}^+}_{\omega'\ell'm'})^* \\ & + A^{H^+}_{\omega\ell m\omega'\ell'm'}\, f^{H^+}_{\omega'\ell'm'} + B^{H^+}_{\omega\ell m\omega'\ell'm'}\, (f^{H^+}_{\omega'\ell'm'})^* \Big]. \end{aligned} \tag{6}$$

The matching coefficients $A^{\mathscr{I}^+}_{\omega\ell m\omega'\ell'm'}$, $B^{\mathscr{I}^+}_{\omega\ell m\omega'\ell'm'}$, $A^{H^+}_{\omega\ell m\omega'\ell'm'}$, and $B^{H^+}_{\omega\ell m\omega'\ell'm'}$ can be found using the following scalar products on the Cauchy surface shown in Fig. 2.

$$A^{(\mathscr{I}^+,H^+)}_{\omega\ell m\omega'\ell'm'} = (f^{\text{in}}_{\omega\ell m}, f^{(\mathscr{I}^+,H^+)}_{\omega'\ell'm'}), \tag{7}$$

$$B^{(\mathscr{I}^+,H^+)}_{\omega\ell m\omega'\ell'm'} = -(f^{\text{in}}_{\omega\ell m}, (f^{(\mathscr{I}^+,H^+)}_{\omega'\ell'm'})^*). \tag{8}$$

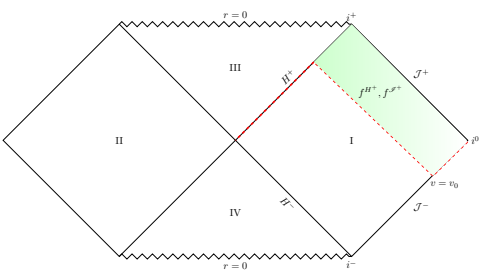


Fig. 2. Penrose diagram for Schwarzschild spacetime showing the Cauchy surface used for matching the in modes in the null shell spacetime to a complete set of modes in Schwarzschild spacetime in the region outside the past and future horizons. The Cauchy surface is denoted by the dashed red curve.

The reason one can expand the *in* modes in this way is that the same differential equations govern the evolution of the *in* modes in the shaded region in Fig. 1 and the shaded region in Fig. 2 because the metric is the same in both regions. However, one may notice that the union of the part of past null infinity with $v > v_0$ and the null shell surface does not form a Cauchy surface in pure Schwarzschild spacetime. We resolve this issue by adding the part of the future horizon with $-\infty < v < v_0$ to the union of $\mathscr{I}^-$ with $v > v_0$ and the null shell surface $v = v_0$, as shown in Fig. 2. We also need to specifity $\psi^{in}_{\omega\ell}$ on the Cauchy surface to evaluate the scalar products in Eqs. 7 and 8. On the part of the Cauchy surface with $v > v_0$ on past null infinity, $\psi^{in}_{\omega\ell} = e^{-i\omega v}$ and on the part where $v = v_0$, $\psi^{in}_{\omega\ell}$ is given by Eq. 5. For the part with $v < v_0$ on the future horizon, we can specify $\psi^{in}_{\omega\ell}$ any way we like so long as it is continuous at $v = v_0$.

Before computing the matching coefficients, we introduce a different complete set of modes that are defined by two linearly independent solutions to the radial mode equation in Schwarzschild spacetime with the following properties

$$\chi^{\infty}_{R} \to e^{i\omega r_*} , \qquad r_* \to \infty, \tag{9}$$

$$\chi^{\infty}_{L} \to e^{-i\omega r_*} , \qquad r_* \to \infty. \tag{10}$$

Near the horizon, they have the behaviors[21]

$$\chi^{\infty}_{R} \to E_R(\omega)e^{i\omega r_*} + F_R(\omega)e^{-i\omega r_*} , \qquad r_* \to -\infty, \tag{11}$$

$$\chi^{\infty}_{L} \to E_L(\omega)e^{i\omega r_*} + F_L(\omega)e^{-i\omega r_*} , \qquad r_* \to -\infty. \tag{12}$$

where E_R, E_L, F_R, and F_L are scattering parameters that can be determined numerically.

Evaluating the scalar products in Eqs. 7 and 8 gives the following results for the matching coefficients[18]

$$A^{H^+}_{\omega\omega' l} = -\frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{e^{i\omega' v_0}}{\omega' - i\epsilon}\psi^{\rm in}_{\omega\ell}(v_H, v_0) + \frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{1}{F^*_L(\omega', \ell)}\frac{e^{i(\omega'-\omega)v_0}}{\omega' - \omega + i\epsilon}$$
$$+ \frac{i}{2\pi\sqrt{\omega\omega'}}\int_{-\infty}^{v_H} du\ \left[\partial_u\psi^{\rm in}_{\omega\ell}(u, v_0)\right]\psi^{H^+}_{\omega'\ell}{}^*(u_s(u), v_0), \tag{13}$$

$$B^{H^+}_{\omega\omega' l} = \frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{e^{-i\omega' v_0}}{\omega' + i\epsilon}\psi^{\rm in}_{\omega\ell}(v_H, v_0) - \frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{1}{F_L(\omega', \ell)}\frac{e^{-i(\omega+\omega')v_0}}{\omega' + \omega - i\epsilon}$$
$$- \frac{i}{2\pi\sqrt{\omega\omega'}}\int_{-\infty}^{v_H} du\ \left[\partial_u\psi^{\rm in}_{\omega\ell}(u, v_0)\right]\psi^{H^+}_{\omega'\ell}(u_s(u), v_0). \tag{14}$$

$$A^{\mathscr{I}^+}_{\omega\omega' l} = -\frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{F^*_R(\omega',\ell)}{F^*_L(\omega',\ell)}\frac{e^{-i(\omega-\omega')v_0}}{\omega'-\omega+i\epsilon}$$
$$-\frac{i}{2\pi\sqrt{\omega\omega'}}\int_{-\infty}^{v_H} du\,\left[\psi^{\rm in}_{\omega\ell}(u,v_0)-e^{-i\omega v_0}\right]\partial_u\psi^{\mathscr{I}^+}_{\omega'\ell}{}^*(u_s(u),v_0), \qquad (15)$$

$$B^{\mathscr{I}^+}_{\omega\omega' l} = \frac{i}{2\pi}\sqrt{\frac{\omega'}{\omega}}\frac{F_R(\omega',\ell)}{F_L(\omega',\ell)}\frac{e^{-i(\omega+\omega')v_0}}{\omega'+\omega-i\epsilon}$$
$$+\frac{i}{2\pi\sqrt{\omega\omega'}}\int_{-\infty}^{v_H} du\,\left[\psi^{\rm in}_{\omega\ell}(u,v_0)-e^{-i\omega v_0}\right]\partial_u\psi^{\mathscr{I}^+}_{\omega'\ell}(u_s(u),v_0). \qquad (16)$$

In the case $\ell = 0$, the *in* mode functions have the form $f^{in}_{\omega 00} = \frac{\psi^{in}_{\omega 00}}{r\sqrt{4\pi\omega}}$ where $\psi^{in}_{\omega 0} = e^{-i\omega v} - e^{-i\omega u}$ for $v \le v_0$. Note that the terms with closed form in Eqs. 13–16 (v-dependent terms) are the only terms that contribute to the v-dependent part of $f^{in}_{\omega 00}$ and the integral terms contribute to the u-dependent part of $f^{in}_{\omega 00}$. We used the v-dependent terms in the matching coefficients to construct $f^{in}_{\omega 00}$ on H^+. The numerical results are shown in Fig. 3 and Fig. 4. In Fig. 3, the real and imaginary parts of the v-dependent part of the *in* mode function have been numerically computed on H^+. The results show that *in* mode function is continuous across the null shell as expected. For large values of ω, the effective potential in the mode equation is always small compared to ω^2 and one can ignore the scattering effects. Hence, one should expect to see the same behaviour as in the 2D case where there are no scattering effects. This is shown to be correct in Fig. 4. where the real and imaginary parts of $f^{in}_{\omega 00}$ are plotted for $M\omega = 9$.

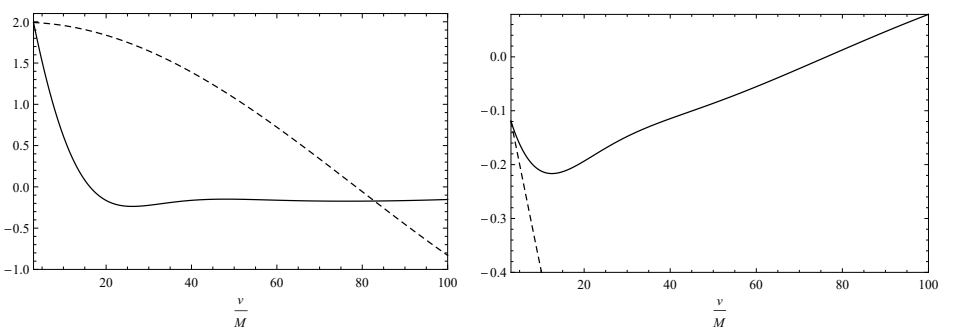


Fig. 3. Real part (left) and imaginary part (right) of $\psi^{in}_{\omega 00}(v)$ on H^+ for $v > v_0$. $\frac{v_0}{M} = 3$ and $M\omega = 0.02$. The dashed lines and solid lines correspond to the *in* modes in the 2D and the 4D cases respectively.

5. Stress-energy tensor and renormalization

One can renormalize the stress-energy tensor by subtracting from the unrenormalized expression for the stress-energy tensor for the in vacuum state, the unrenor-

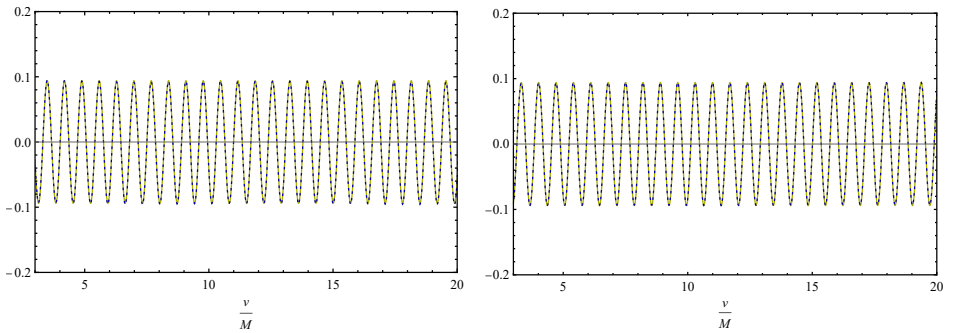


Fig. 4. Real part (left) and imaginary part (right) of $\psi^{in}_{\omega 00}(v)$ on H^+ for $v > v_0$. $\frac{v_0}{M} = 3$ and $M\omega = 9$. The dashed blue lines and solid yellow lines correspond to the *in* modes in the 2D and the 4D cases respectively.

malized stress-energy tensor for the Unruh state. Since the renormalization counterterms are local and thus do not depend on the state of the quantum field, this quantity will be finite. Then one can add back the unrenormalized stress-energy tensor for the Unruh state and subtract from it the renormalization counter terms. Schematically one can write

$$\langle \text{in}|T_{ab}|\text{in}\rangle_{\text{ren}} = \Delta\langle T_{ab}\rangle + \langle U|T_{ab}|U\rangle_{\text{ren}}, \tag{17}$$

where $\Delta\langle T_{ab}\rangle = \langle \text{in}|T_{ab}|\text{in}\rangle_{\text{unren}} - \langle U|T_{ab}|U\rangle_{\text{unren}}$. Note that the Unruh state is defined by a set of modes that are positive frequency on the past horizon with respect to the Kruskal time coordinate and the set of modes that have the form $\psi_{\omega\ell} = e^{-i\omega v}$ on past null infinity. The quantity $\langle U|T_{ab}|U\rangle_{\text{ren}}$ has been numerically computed for a massless minimally coupled scalar field in Schwarzschild spacetime. Thus what remains is to compute the difference between the unrenormalized expressions.

The unrenormalized stress-energy tensor can be computed by taking derivatives of the Hadamard Green's function as follows,[22]

$$\langle T_{ab}\rangle_{\text{unren}} = \frac{1}{4}\lim_{x'\to x}\left[\left(g_a^{c'}G^{(1)}_{;c';b}(x,x') + g_b^{c'}G^{(1)}_{;a;c'}(x,x')\right) - g_{ab}\,g^{cd'}G^{(1)}_{;c;d'}(x,x')\right], \tag{18}$$

where the quantity $g_a^{b'}$ parallel transports a vector from x' to x and is called the bivector of parallel transport.

5.1. *Stress-energy tensor in 2D*

In this section, we show how our method can be applied to the case of a 2D null shell spacetime which has a perfectly reflecting mirror at $r = 0$. There is no scattering

that means $E_R = F_L = 1$ and $E_L = F_R = 0$. The matching coefficients are,[18]

$$A^{\mathscr{I}^+}_{\omega\omega'} = -\frac{1}{2\pi}\sqrt{\frac{\omega}{\omega'}}(4M)^{i4M\omega'}e^{-i(\omega-\omega')v_H}\frac{\Gamma(1-i4M\omega')}{[-i(\omega-\omega')+\epsilon]^{1-i4M\omega'}}, \tag{19}$$

$$B^{\mathscr{I}^+}_{\omega\omega'} = \frac{1}{2\pi}\sqrt{\frac{\omega}{\omega'}}(4M)^{-i4M\omega'}e^{-i(\omega+\omega')v_H}\frac{\Gamma(1+i4M\omega')}{[-i(\omega+\omega')+\epsilon]^{1+i4M\omega'}}. \tag{20}$$

The expression for f^{in}_ω can be obtained by substituting Eqs. 19 and 20 into Eq. 6. Those for f^{Unruh}_ω can be obtained using similar expressions. See Ref. 18 for more details. Next, we construct the Hadamard form of the Green's function which in 2D is

$$G^{(1)}_{in}(x,x') = \int_0^\infty d\omega\ [f^{in}_\omega(x)f^{in\,*}_\omega(x') + f^{in\,*}_\omega(x)f^{in}_\omega(x')]\ . \tag{21}$$

We subtract off the corresponding expression for the Unruh state to obtain

$$\Delta G(x,x') = G^{(1)}_{in}(x,x') - G^{(1)}_{\text{Unruh}}(x,x'),$$

which gives

$$\Delta\langle T_{tt}\rangle = -(1-\frac{2M}{r})\lim_{x'\to x}\frac{1}{4}(\Delta G_{;t';r} + \Delta G_{;t;r'}). \tag{22}$$

Our method results in a complicated operation for $\Delta\langle T_{tt}\rangle$ which initially contains a triple integral. One of the integrals can be computed in closed form with the result[18]

$$\begin{aligned}\Delta\langle T_{tt}\rangle = \Re\Bigg\{ &\frac{i}{8\pi^3}\int_0^\infty d\omega_1\ \int_0^\infty d\omega_2\ e^{-2\pi M(\omega_1+\omega_2)} \\ &\times\Bigg\{ e^{i(\omega_2-\omega_1)u_s}\frac{(4M\omega_1 e^{\frac{v_H}{4M}})^{4iM\omega_1}}{(4M\omega_2 e^{\frac{v_H}{4M}})^{4iM\omega_2}}\frac{\Gamma(1-4iM\omega_1)\Gamma(1+4iM\omega_2)}{4M(\omega_1-\omega_2-i\epsilon)} \\ &+ e^{-i(\omega_2+\omega_1)u_s}(4M\omega_1 e^{\frac{v_H}{4M}})^{4iM\omega_1}(4M\omega_2 e^{\frac{v_H}{4M}})^{4iM\omega_2} \\ &\times\frac{\Gamma(1-4iM\omega_1)\Gamma(1-4iM\omega_2)}{4M(\omega_1+\omega_2)}\Bigg\}\Bigg\}\ .\end{aligned} \tag{23}$$

This quantity has been computed numerically and the results are shown in Fig. 5. The stress-energy tensor for a massless minimally coupled scalar field in the 2D collapsing null shell spacetime has been previously computed analytically using a different method[19,20,23–25] and the stress-energy tensor for the Unruh state has also been computed analytically.[19,20,23–30] Our results are shown with the dots in Fig. 5 and the result found by using previous methods is shown with a solid curve. They agree to more than ten digits. It is worth mentioning that in 2D, once $\Delta\langle T_{tt}\rangle$ is numerically computed, $\Delta\langle T_{rr}\rangle$ and $\Delta\langle T_{rt}\rangle$ can be easily determined.[18]

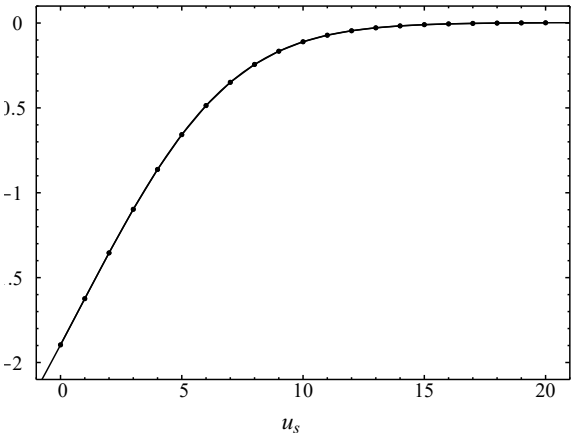


Fig. 5. The quantity $10^4 M^2 \langle T_{tt} \rangle$ is plotted for the massless minimally coupled scalar field in the region exterior to the null shell and to the event horizon. The dots correspond to the results of the numerical computations. The solid curve represents the analytic results in (reference).

6. Summary

We have reviewed a method of computing the stress-energy tensor for a massless minimally-coupled scalar field in a spacetime in which a spherically symmetric black hole is formed by the collapse of a null shell. This method primarily involves two parts. One part is the expansion of the *in* mode functions in terms of a complete set of modes in the part of a Schwarzschild black hole that is outside the event horizon. The matching coefficients of the expansion have been found in terms of the integrals of the mode functions and closed form terms. These matching coefficents have been used to numerically compute part of the *in* mode function on the future horizon of the black hoe.

The second part of the method is the renormalization of the stress-energy tensor which involves taking the difference between the stress-energy tensor for the "in" state in the collapsing null shell spacetime and that for the Unruh state in the Schwarzschild spacetime. Finally, we reviewed the computation of the stress-energy tensor in the corresponding 2D case using aformentioned method.

Acknowledgment

P. R. A. would like to thank Eric Carlson, Charles Evans, Adam Levi, and Amos Ori for helpful conversations and Adam Levi for sharing some of his numerical data. A.F. acknowledges partial financial support from the Spanish Ministerio de Ciencia e Innovación grant FIS2017-84440-C2-1-P and from the Generalitat Valenciana grant PROMETEO/2020/079. This work was supported in part by the National Science Foundation under Grants No. PHY-1308325, PHY-1505875, and PHY-1912584 to Wake Forest University. Some of the numerical work was done

using the WFU DEAC cluster; we thank the WFU Provost's Office and Information Systems Department for their generous support.

References

1. M. S. Fawcett, Commun. Math. Phys **89**, 103 (1983).
2. K. W. Howard and P. Candelas, Phys. Rev. Lett. **53**, 403 (1984).
3. K. W. Howard, Phys. Rev. D **30**, 2532 (1984).
4. B. P. Jensen and A. Ottewill, Phys. Rev. D **39**, 1130 (1989).
5. B. P. Jensen, J. G. McLaughlin, and A. C. Ottewill, Phys. Rev. D **43**, 4142 (1991).
6. B. P. Jensen, J. G. Mc Laughlin, and A. C. Ottewill, Phys. Rev. D **45**, 3002 (1992).
7. P. R. Anderson, W. A. Hiscock, and D. A. Samuel, Phys. Rev. Lett. **70**, 1739 (1993).
8. P. R. Anderson, W. A. Hiscock, and D. A. Samuel, Phys. Rev. D **51**, 4337 (1995).
9. P. R. Anderson, W. A. Hiscock, and D. J. Loranz, Phys. Rev. Lett. **74**, 4365 (1995).
10. E. D. Carlson, W. H. Hirsch, B. Obermayer, P. R. Anderson, and P. B. Groves, Phys. Rev. Lett. **91**, 051301 (2003).
11. P. R. Anderson, R. Balbinot, and A. Fabbri, Phys. Rev. Lett. **94**, 061301 (2005).
12. C. Breen and A. C. Ottewill, Phys. Rev. D **85**, 084029 (2012).
13. A. Levi and A. Ori, Phys. Rev. Lett. **117**, 231101 (2016).
14. A. Levi, Phys. Rev. D **95**, 025007 (2017).
15. N. Zilberman, A. Levi, A. Ori, Phys. Rev. Lett. **124**, 171302 (2020).
16. G. Duffy and A. C. Ottewill, Phys. Rev. D **77**, 024007 (2008).
17. A. Levi, E. Eilon, A. Ori, and M. van de Meent, Phys. Rev. Lett. **118**, 141102 (2017).
18. P. R. Anderson, S Gholizadeh Siahmazgi, R. D. Clark, and A. Fabbri, Phys. Rev. D **102**, 125035 (2020).
19. A. Fabbri and J. Navarro-Salas, *Modeling black hole evaporation* (Imperial College Press, London, UK, 2005).
20. S. Massar and R. Parentani, Phys. Rev. D **54**, 7444 (996).
21. P. R. Anderson, A. Fabbri, and R. Balbinot, Phys. Rev. D **91**, 064061 (2015).
22. S. M. Christensen, Phys. Rev. D **14**, 2490 (1976).
23. M. R. R. Good, P. R. Anderson, and C. R. Evans, Phys. Rev. D **94**, 065010 (2016).
24. P. R. Anderson, R. D. Clark, A. Fabbri, and M. R. R. Good, Phys. Rev. D **100**, 061703(R) (2019).
25. W. A. Hiscock, Phys. Rev. D **23**, 2813 (1981).
26. S. W. Hawking, Commun. Math. Phys. **43** 199 (1975).
27. W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
28. T. Elster, Phys. Lett. **94A**, 205 (1983).
29. E. T. Akhmedov, H. Godazgar, and F. K. Popov, Phys. Rev. D **93**, 024029 (2016).
30. P. C. W. Davies, S. A. Fulling, and W. G. Unruh, Phys. Rev. D **13**, 2720 (1976).
31. P. R. Anderson, R. Balbinot, A. Fabbri, and R. Parentani, Phys. Rev. D **87**, 124018 (2013).

Microscopic model building for black hole membranes from constraints of symmetry

Swastik Bhattacharya

Department of Physics, BITS Pilani Hyderabad; Hyderabad 500078, Telangana State, India
E-mail: swastik@hyderabad.bits-pilani.ac.in

S. Shankaranarayanan

Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India
E-mail: shanki@phy.iitb.ac.in

Einstein equations projected on black-hole horizons give rise to the equations of motion of a viscous fluid. This suggests a way to understand the microscopic degrees of freedom on the black-hole horizon by focusing on the physics of this fluid. In this talk, we shall approach this problem by building a crude microscopic model for the Horizon-fluid(HF) corresponding to asymptotically flat black-holes in 3+1 dimensions. The symmetry requirement for our model is that it should incorporate the S1 diffeo-symmetry on the black-hole horizon. The second constraint comes from the demand that the correct value of the Coefficient of the Bulk Viscosity of the HF can be deduced from the model. Both these requirements can be satisfied by an adoption of the eight vertex Baxter model on a S2 surface. We show that the adiabatic entropy quantisation proposed by Bekenstein also follows from this model. Finally, we argue the results obtained so far suggest that a perturbed black-hole can be described by a CFT perturbed by relevant operators and discuss the physical implications.

Keywords: Black Hole; Horizon-fluid; Bulk Viscosity; deformed CFT.

1. Introduction

Gravity, quantum theory, and thermodynamics[1–3] are related at some fundamental level. The laws of black-hole mechanics are formally similar to the Laws of Thermodynamics and suggest that black-holes have entropy and temperature.[4–8] A satisfactory explanation of black-hole thermodynamics requires a statistical mechanical origin of entropy. It has been argued that most of the black-hole degrees of freedom (DOF) reside on the horizon, as the black-hole entropy scales as area.[1,9] Also, black-hole thermodynamics deals only with equilibrium states.

Gravity causes the non-stationary black-hole horizon to interact with external fields (perturbations) continually. These interactions lead to the transfer of energy from the external fields to black-hole degrees of freedom. If we compute observables involving black-hole, we observe *dissipative* effects corresponding to interaction with fields. In macroscopic black-hole physics, this can be explicitly seen by projecting the equations of motion of external fields and gravity theory on the black-hole

event horizon leading to dissipative equations. In these scenarios, fluid dynamics description is useful as only average quantities resulting from the interactions at the microscopic level are observed on macroscopic scales.[10] Interestingly, it was shown black-hole horizon behaves like a viscous fluid and satisfies Damour-Navier-Stokes equation.[11–13]

Since we do not have a consistent model of quantum gravity, we explore here the viability of certain ideas in the context of a toy model for the non-stationary black-hole horizon that is slightly perturbed from stationarity. The toy model we construct takes into account *two different aspects of the black-hole horizon.* First, the model incorporates near-horizon symmetries of the stationary black-hole.[14–20] Second, the model accounts for the physics of transport phenomena of horizon-fluid.[21,22]

The first aspect is the constraint of symmetry on the microscopic theory. Thus, for example, we demand that the model must incorporate near-horizon symmetries. The second aspect can be viewed as the expectation that the transport phenomena exhibited by the horizon-fluid corresponding to a black-hole in the hydrodynamic limit should follow from this model.

A Conformal Field Theory (CFT) on the black-hole horizon can partly incorporate the near-horizon symmetries and is a natural candidate for the microscopic theory of stationary black-holes. We extend the analysis and show that a perturbed CFT can also incorporate this symmetry. Thus in our model, a non-stationary black-hole can be viewed as interacting with external fields. For the microscopic description, this means adding interaction terms to the stationary black-hole described by CFT. We can see this process as a perturbed black-hole relaxing to a stationary black-hole by emitting QNMs.[23] In practice, we shall deform the CFT by introducing only a mass term. However, it plays a vital role by determining the infra-red cutoff for the theory and, as a consequence, also determining the value of the coefficient of Bulk Viscosity, ζ.

We focus only on the Bulk Viscosity here, which means we only consider homogeneous processes, e.g., the increase of black-hole area in a spherically symmetric space-time due to infalling spherical mass shell. Modeling such processes is comparatively easier. Stationary, non-extremal black-holes in 4-dimensional general relativity exhibit an infinite-dimensional symmetry in the near-horizon region.[14–20,24,25] Thus, the near-horizon possess infinite-dimensional algebra such as $\mathcal{S}1$ diffeomorphism[19,20] or (near) BMS.[25–28] The CFT describing a stationary black-hole can incorporate the $\mathcal{S}1$ diffeosymmetry[19,20] as it possesses a representation of the Virasoro algebra.[20,29]

The perturbed CFT we choose possesses symmetries that lead to a representation of the Virasoro algebra.[30] These are *Integrable field theories* with an infinite number of conserved charges corresponding to an infinite number of symmetries.[31] The crucial point that allows us to model the black-hole horizon-fluid by such a perturbed CFT model is that one of the representations of the Virasoro algebra corresponding to the perturbed CFT is also a representation of the $\mathcal{S}1$

diffeomorphism symmetry. Throughout what follows, we use natural units setting $\hbar = c = G = k_B = 1$.

2. The microscopic model

We consider here the *Eight-vertex Baxter model*[32–37] for our purpose. We show that this model exhibits both the two aspects discussed above. The model has the following properties that form crucial ingredients for the microscopic model building of the horizon-fluid: First, it possesses *lattice* Virasoro algebra corresponding to $\mathcal{S}1$ diffeomorphism symmetry.[38,39] Second, it consists of two staggered 2D Ising lattices, and its free energy density is the same as that of the 2D Ising model. However, the two-sublattice symmetry is very different from that of the usual Ising model. Hence, the Baxter solution's critical indices are, in general, different from that of Ising.[38] Third, it exhibits a second-order phase transition. In the continuum limit, it is an Integrable Field Theory near the critical point and is a CFT at the critical point.[39–41]

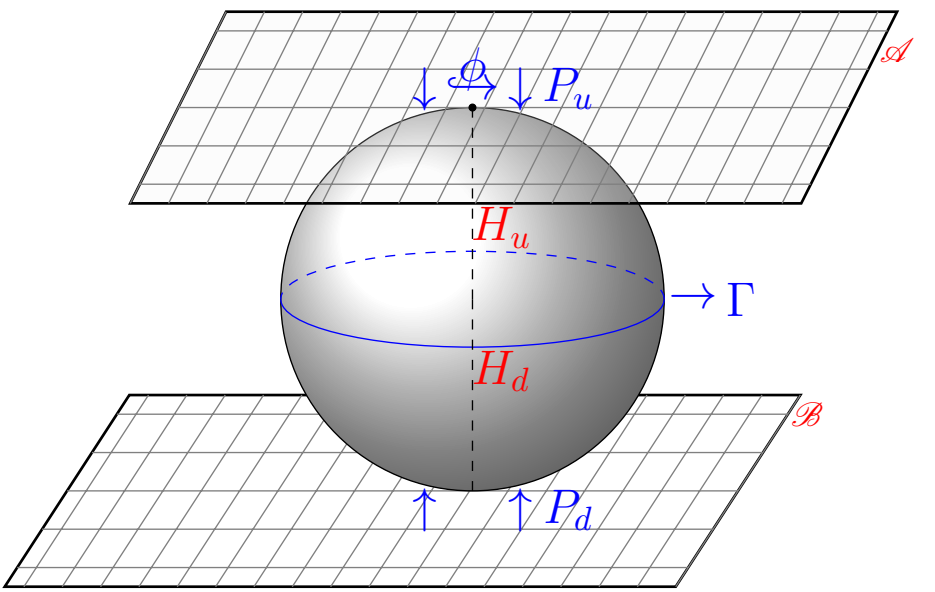


Fig. 1. The projection of the 8-vertex Baxter model from the two sub-lattices to S^2 surface of the horizon.

The Baxter model can be adapted to the cross-section of a black-hole event-horizon, a S^2 surface. As shown in 1, this model can be adopted on the two hemispheres of the S^2 surface through projection from two Baxter lattices. Let P_u (P_d) denote the map corresponding to the projection $\mathcal{A} \to H_u$ ($\mathcal{B} \to H_d$). For the

consistency of the model, we need to impose the condition $P_u^{-1} \circ \{\Gamma\} \equiv P_d^{-1} \circ \{\Gamma\}$, where, P^{-1} denotes the inverse map and Γ is equatorial plane of the S^2 surface.

The above condition retains the periodic boundary condition of the Baxter model. The projection allows relating the Euclidean boost parameter of the Baxter model[41] to the azimuthal angle in the spherical polar coordinate. We can then relate the Virasoro algebra (corresponding to the $\mathcal{S}1$ diffeomorphism) of the Euclidean boost parameter to the $\mathcal{S}1$ diffeomorphism of the azimuthal angle in the horizon-fluid model. Thus, the projection retains the model's main physical features[41] and *directly* incorporates a representation of the $S1$ diffeosymmetry in the model. *It is simply the diffeosymmetry of the azimuthal angle for the black-hole system.*

3. Eight-vertex model and deformed CFT

The Eight-vertex model has eight possible arrangements of arrows at a vertex with four distinct Boltzmann weights a, b, c, d. These satisfy two constraints:[37]

$$\frac{cd}{ab} = \frac{1-\Gamma}{1+\Gamma} \; ; \; \frac{a^2+b^2-c^2-d^2}{2(ab+cd)} = \Delta \tag{1}$$

For constant Γ and Δ, there exists a one-parameter family of Boltzmann weights (w) which satisfy the star–triangle relations and, hence, the eight-vertex model has a one-parameter family of commuting transfer matrices.[37] This allows one to parameterize the Boltzmann weights explicitly in terms of *spectral variable* (u):

$$\begin{aligned} a &= \mathrm{snh}(\lambda - u) \; b = \mathrm{snh}\, u \\ c &= k\,\mathrm{snh}\,\lambda, \qquad d = k\,\mathrm{snh}(\lambda - u) \end{aligned} \tag{2}$$

where, k is the elliptic modulus, and snh is the hyperbolic analogue of sn and is given by $\mathrm{snh}\;\; u = -i\;\; \mathrm{sn}(iu)$. It has been shown that the transfer matrix of the eight-vertex model commutes with the XYZ Hamiltonian:[43]

$$H_{\mathrm{XYZ}} = -\frac{1}{2}\sum_{j=1}^{N} J_\sigma \sigma_j^\sigma \sigma_{j+1}^\sigma \quad \text{where} \quad \sigma = x, y, z\,. \tag{3}$$

The coupling constants are related to the weights by the relation: $J_x : J_y : J_z = 1 : \Gamma : \Delta$. The spins σ_n's are related to the vertex weights by the vertex operator (V_n):

$$V_n = \frac{1}{2}\left[a + c + [a-c]\sigma_n^z \sigma_{n+1}^z + [b+d]\sigma_n^x \sigma_{n+1}^x + \sigma_n^y \sigma_{n+1}^y\right]$$

where,

$$\begin{aligned} a &= \exp\left[K_1 + K_2 + K''\right], \\ b &= \exp\left[-K_1 + K_2 + K''\right], \\ c &= \exp\left[-K_1 + K_2 - K''\right], \\ d &= \exp\left[K_1 - K_2 - K''\right]. \end{aligned} \tag{4}$$

The Hamiltonian of the Baxter model is:

$$H_{\text{Baxter}} = \sum_{\sigma} \left[K_1 \, \sigma_1 \sigma_3 + K_2 \, \sigma_2 \sigma_4 + K'' \, \sigma_1 \sigma_2 \sigma_3 \sigma_4 \right] \tag{5}$$

where σ_1, σ_3 (σ_2, σ_4) are the lattice spins in one (second) Ising sublattice with coupling constant K_1 (K_2) as shown in 2, and K'' refers to the four-spin coupling that connects the two sublattices.

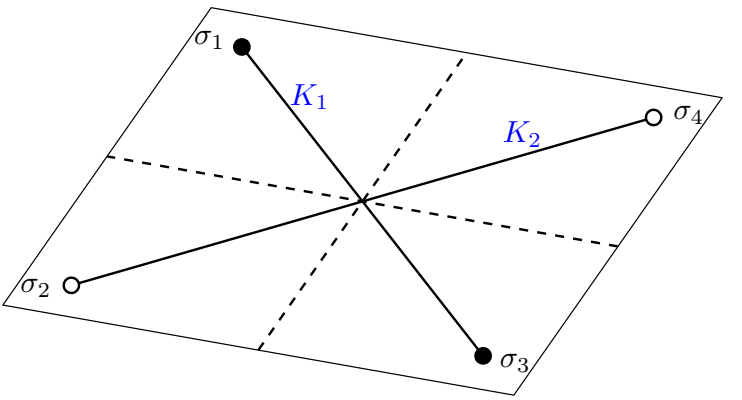


Fig. 2. A single face of the 8-Vertex Baxter model in the spin formulation. K_1 and K_2 are the horizontal and vertical spin-spin couplings of the two Ising sublattices.

It is possible to obtain the statistical weight (partition function) in the row-to-row transfer matrix and corner transfer matrix formulations.[37] In our case, it is convenient to use the corner transfer matrix (CTM) formulation as it directly connects the algebra with the conserved charges.

CTM can be expressed conveniently in terms of Fermions described by the following XYZ spin-chain Hamiltonian.[37,39–41] In this formulation, the partition function is evaluated by choosing a spin in the middle of the lattice and fixing the spins along the vertical and horizontal axes while summing over all spins in the interior of each quadrant.[37] The final sum over the spins along the axes may be interpreted as the trace of a product of four corner transfer matrices.

Thus we see that the eight vertex Baxter model is closely related to the XYZ spin chain. Now we can map it to a theory of Fermions by introducing the Corner Transfer Matrix(CTM).

The CTM operator can be viewed physically as connecting semi-infinite rows of arrows with a semi-infinite column of arrows of one quadrant of the lattice. In the thermodynamic limit, the following relation holds:[37]

$$\mathcal{A}(u) = \exp\left[-\frac{\pi u}{2K} \, \mathcal{L}_0 \right] , \tag{6}$$

where K is a complete elliptic integral associated with modulus k and

$$\mathcal{L}_0 = \frac{2K}{\pi} \sum_{j=-\infty}^{\infty} j\mathcal{H}_{XYZ}(j, j+1). \tag{7}$$

$\mathcal{H}_{XYZ}$ corresponds to the Hamiltonian of a spin chain with nearest neighbour (3) To keep the calculations transparent, we set $\Gamma = 0$, i.e., $cd = ab$ in (1). This corresponds to the condition $J_z = 0$ in the Hamiltonian (3) which is the well-known XY model.[37] $\mathcal{L}_0$ in Eq. (7) is diagonalized by the operators:

$$\Psi(l) = N_l \int d\alpha \, e^{-\imath\alpha l\pi/2K} \chi(\alpha) \tag{8}$$

where N_l is the normalization constant, and $\chi(\alpha) = \mathrm{sn}\,\alpha\, a^{\alpha}(\alpha) + \imath\sqrt{k}\,\mathrm{cn}\,\alpha\, a^{y}(\alpha)$. The integration over α is over one complete real period of the elliptic functions from $-2K + \imath K'/2$ to $2K + \imath K'/2$. Itoyama and Thacker showed that $\mathcal{L}_0$ could be expressed as

$$\mathcal{L}_0 = \sum_l l : \bar{\Psi}(l)\Psi(l) : + h\,, \tag{9}$$

where, h is a constant. $\mathcal{L}_0$ is embedded into a Virasoro algebra as a central element.[39–41] The normal ordering is defined by the relations, $\Psi(l)|h\rangle = 0(\forall\ l \geq 1), \overline{\Psi}(l)|h\rangle = 0, (\forall\ l \leq 1)$. Other Virasoro operators $\mathcal{L}_n$ can be constructed from these momentum space operators.[41] From (8), it follows that,

$$[\mathcal{L}_n, \mathcal{L}_m] = (n-m)\mathcal{L}_{n+m} + \frac{1}{12}c(n^3 - n)\delta_{n+m,0}. \tag{10}$$

As noted in Ref.,[41] the physical Hilbert space built from the state $|h\rangle$, forms the highest weight representation of the Virasoro algebra. Since the eigenvalues of $\mathcal{L}_0$ are doubled due to the zero modes of the operator $\Psi(0)$ and $\overline{\Psi}(0)$, the highest weight vector forms a two-dimensional representation under parity conjugation. At the critical point, the central charge $c = 1$ and $h = \frac{1}{8}$.

Using the following classical generators (l_n^{Diff}),

$$l_n^{Diff} = -\frac{1}{2}\zeta^{n+1}\frac{d}{d\zeta} - \frac{1}{2}\frac{d}{d\zeta}\zeta^{n+1}. \tag{11}$$

we can obtain other Virasoro algebra, different from the one described above. The difference is that l_n^{Diff} are generators of diffeo-transformation of the spectral rapidity parameter or the Euclidean boost parameter (α).[41,44] We can then construct the corresponding Virasoro algebra by defining the following $\mathcal{L}_n^{diff}$:

$$\begin{aligned}\mathcal{L}_n^{diff} &= : \int_{-K}^{3K} \frac{d\beta}{2\pi} B(\beta + 2K - \imath K')l_n B(\beta) : + h\delta_{n,0} \\ &= \sum_l \left(l + \frac{n-1}{2}\right) : \bar{\psi}(l)\psi(l+n) : + h\delta_{n,0}.\end{aligned} \tag{12}$$

This demonstrates that the eight-vertex model possesses the Virasoro algebra given by (12), which holds the key to incorporating near-horizon $S1$ diffeo-symmetry in the model of the horizon-fluid. The 2-D Euclideanized space-time (τ, q) can be identified with the 2-D Euclidean space (x_1, x_2) on which the horizon-fluid resides. The rapidity or the boost parameter in a Euclideanized space-time corresponds to an angle of rotation. A closer look reveals that in our case, the rapidity is the azimuthal angle (See (1)). Thus, l_n^{Diff} diffeomorphism algebra of the spectral rapidity corresponds to the l_n^{Diff} diffeomorphism symmetry on the black-hole horizon. Thus we see that the microscopic modelling of the horizon-fluid with a mass gap incorporates the l_n^{Diff} diffeomorphism symmetry on the black-hole horizon.

4. Continuum limit and modelling horizon-fluid

Long-range effects dominate the critical properties of this model; hence, a continuum approximation will suffice for our purpose. The eight-vertex model's continuum limit is a theory of massive Dirac Fermions (Ψ_1, Ψ_2)[41]

$$\mathcal{S}_{\text{Dirac}} = \int d\tau dq \left[\frac{1}{2} i\overline{\Psi}_1 \Big(\frac{\overleftrightarrow{\partial}}{\partial \tau} + \frac{\overleftrightarrow{\partial}}{\partial q} \Big) \Psi_1 + \frac{1}{2} i\overline{\Psi}_2 \Big(\frac{\overleftrightarrow{\partial}}{\partial \tau} - \frac{\overleftrightarrow{\partial}}{\partial q} \Big) \Psi_2 - m(\bar{\Psi}_1 \Psi_2 + \bar{\Psi}_2 \Psi_1) \right]. \tag{13}$$

The above action possesses an infinite sequence of conserved densities. Physically, this implies that besides the total angular momentum, the entire momentum distribution is conserved.[41] We can rewrite these operators as integrals of local densities in coordinate space:

$$\mathcal{L}_n = \int dq \, J_0^{(n)}(q) \,, \tag{14}$$

where J_0 is the zeroth component of a conserved current. $J_0^{(-1)}$ is the Hamiltonian plus the momentum operator, and $J_0^{(0)}$ (at $\tau = 0$) is the first moment of the Dirac Hamiltonian.[41] Integrability ensures the operators are related to the infinite sequence of conserved charges, with one for each new $\mathcal{L}_n$.[41]

The action (13) is useful to identify the infinite number of conserved charges. However, for our purpose here, starting from the action (13) it is cumbersome to look at the hydrodynamic or long-wavelength properties. Hence, to derive ζ, we will not make direct use of the action (13). Instead, to make contact with the macroscopic level, i.e., Damour's fluid description, we turn to the fact that the Free-energy density of the Baxter model is the same as that of a classical Ising model in 2-dimensions near the critical point.[37] On the other hand, the theory of the 2-dimensional Ising model can be described by a theory of a free massive scalar field φ in a two-dimensional Euclidean spacetime.[45] The key is to identify the microscopic field with the phenomenological order parameter.

The task of identifying the microscopic field with the phenomenological order parameter is non-trivial. For the transport theory, we need to describe the dynamics of the black-hole horizon, starting from a $(2+1)-$dimensional Hamiltonian.

However, the microscopic model we have constructed is characterized by the *2-D Euclidean space-time Hamiltonian*. This means that the microscopic model suffices to describe the quantum states for a system in thermal equilibrium with the desired accuracy but is inadequate if we consider how the system evolves with time. In other words, such Hamiltonians do not contain information about the dynamics of the black-hole horizon.

There are two ways to tackle this issue. Both approaches keep the thermal states the same but provide different inputs about the dynamics of the system. One way is to assume that a phenomenological Langevin equation gives the dynamics. From this, one can work out the transport theory and compute the correct coefficient of Bulk Viscosity for the Horizon-fluid. The other approach is to construct a $(2+1)-$dimensional Hamiltonian with the same symmetries as the mass gap possessed by the 2-D Euclidean space-time Hamiltonian. For a general process, this is a challenging problem. Fortunately, this difficulty is bypassed for the processes describing bulk viscosity of the fluid using the mean-field theory description. We should emphasize that in this case also, the thermal states remain the same. (The collection of the thermal states of the $(2+1)-$dimensional theory is equivalent to collecting the quantum states of the $1+1$ dimensional Euclideanized theory subject to appropriate boundary conditions.) We shall only discuss the second approach here as a Langevin-type equation can also be derived from the dynamical quantum theory under suitable approximation.

The mean-field description will not capture all the details of the symmetries of the quantum states of the model or can reproduce the correct scaling exponents; however, it works satisfactorily to describe horizon-fluid properties. Thus, the 2-D mean-field theory Hamiltonian of the microscopic model can be extended to the following $2+1$ dimensional space-time Hamiltonian:

$$H_{\text{eff}}(\varphi) = \int \left[\frac{1}{2}\big(\frac{\partial\varphi}{\partial t}\big)^2 + \frac{1}{2}\big(\nabla\varphi\big)^2 + \frac{m_{eff}^2}{2}\varphi^2\right] dA, \tag{15}$$

where the integral is over the area of the event horizon. The above Hamiltonian satisfies the essential requirement of possessing $Z2$ symmetry. Still, it suffers from the weakness that there is very little guidance otherwise in terms of symmetry requirements. One can view the above Hamiltonian as a Mean-Field description of the horizon-fluid near a critical point.

We are now in a position to identify the above Hamiltonian (15) with the horizon-fluid. As a first step, we write down the Landau-Ginzburg expression for the entropy functional corresponding to the horizon-fluid instead of using the Free Energy functional.[46] This is because the black-hole (of horizon area A) constantly interacts with the environment, leading to energy flow. Thus, we need a framework in which the energy density and the order parameter appear in the formalism on the same footing. As shown in Ref.,[46] the appropriate thermodynamic potential to consider in such a case, is the entropy functional and not free energy. The stationary state of the black-hole corresponds to the thermodynamic equilibrium state of the black-hole

with maximum entropy. The quasi-stationary state of a black-hole corresponds to the vacuum of the deformed CFT.

Following,[21] we assume that the entropy density of the horizon-fluid is homogeneous. We focus on a macroscopically small but finite element of the black-hole horizon area, denoted as $\mathcal{A}$. $\mathcal{A}$ satisfies the condition $\mathcal{A}/A \ll 1$. The order parameter (η) is taken to be:

$$\eta = C\sqrt{\mathcal{A}}\,.$$

We can fix the value of the constant C only by relating it to a macroscopic quantity. The entropy functional $(\mathcal{S})$ of the horizon-fluid of this finite element about $(T, \mathcal{A})$ is given by:[46]

$$\mathcal{S} = \mathcal{S}_0(T, \mathcal{A}) - a\,\eta^2 - b\,\eta^4, \tag{16}$$

where a, b are constants. Let the value of the entropy functional $\mathcal{S}$ in the quasi-stationary state by $\mathcal{S}_{QS}$ and the stationary state by $\mathcal{S}_S$, respectively. Note that $\mathcal{S}_S$ is the global maximum for entropy functional $\mathcal{S}$. We assume that the process of going from $\mathcal{S}_{QS}$ to $\mathcal{S}_S$ is a slow physical process so that we can use equilibrium thermodynamics to describe the quasi-stationary state. For the microscopic model of the horizon-fluid, this means that the ground state of the deformed CFT evolves slowly into a state of CFT. The slow evolution implies that the deformed CFT vacuum is likely to possess some of the symmetries of a CFT state. This is reminiscent of the adiabatic evolution of a quantum state.

Rewriting (16) in terms of the horizon-area at equilibrium, we get,

$$\begin{aligned} S_{\max} = \frac{\mathcal{A}_{\max}}{4} &= \mathcal{S}_0 - a\,C^2\mathcal{A}_{\max} - b\,C^4\mathcal{A}_{\max}^2\,, \\ \text{where}\quad a = -\frac{1}{4C^2};&\quad \mathcal{S}_0 - b\,C^4\,\mathcal{A}^2 = 0\,. \end{aligned} \tag{17}$$

Denote the deviation of any variable from its equilibrium value by a prefix δ. The change in the entropy functional is related to the change in the energy density of the horizon-fluid:

$$\delta\mathcal{H}(\delta\eta) = -T\,\delta\mathcal{S} = \frac{T}{2C^2}\delta\eta^2. \tag{18}$$

where we have $k_B = 1$. [Expansion around the maximum implies that the terms proportional to $\delta\eta$ cancel out as the expansion is being done around a maximum value of the entropy function. This leads to the condition $2b\eta_{max}^2 = -a$. Substituting this in the terms quadratic in $\delta\eta$, i.e. $(a{+}2b\eta_{max}^2)\delta\eta^2)$ leads to the above expression.]

Using the above expression, we can relate the order parameter (η) with the scalar field (φ) and the change in the energy density of the horizon fluid $(\delta\mathcal{H}(\delta\eta))$ with $H_{\text{eff}}(\varphi)$ (15), i.e.

$$\delta H = \int dA\,\delta\mathcal{H} = \delta H_{\text{eff}}\,. \tag{19}$$

For the homogeneous process, the spatial gradient terms in the Hamiltonian (15) can be ignored. Also, the field is taken to vary slowly so the kinetic term in the Hamiltonian (15) can be ignored. Using the universality near critical point, we have:[37,41]

$$\langle\varphi_*\rangle = \frac{\delta\eta}{\sqrt{\mathcal{A}}}, \tag{20}$$

where, $\langle\varphi_*\rangle$ is the average value of the field φ_* in the area $\mathcal{A}$. As seen from the above expression, φ_* is dimensionless, and is related to φ as follows: $\varphi = \sqrt{\Upsilon}\varphi_*$ where Υ has the dimensions of energy. The change in the energy of the horizon fluid is given by:

$$\delta H = \frac{\Upsilon}{2}\int dA\, m_*^2\,\varphi_*^2, \quad \text{where} \quad m_* = \frac{1}{C} \quad \text{and} \quad \Upsilon = T \tag{21}$$

Thus, we have made the connection between the microscopic model with black-hole thermodynamics. m_* has the dimensions of energy, i.e., $m_* = 1/(l_P^2\, C)$, l_P being the Planck length. C is dimensionless. Note that we have set $l_P = 1$. In the next section, we evaluate the correlations for this process (bulk viscosity). The thermal correlations considered here remain the same for Hamiltonian corresponding to action (13) and H_{eff}.

5. Bulk viscosity from the microscopic model

We use the above mean-field theory description to calculate the viscosity coefficient of the horizon-fluid from the correlations of the energy-momentum tensor of the scalar field given by the Hamiltonian (15). Jeon has developed such a description.[42] First, however, we need to make suitable changes to apply to the horizon-fluid.

According to Jeon, the coefficient of bulk-viscosity (ζ) of a viscous-fluid with stress-tensor is:[42]

$$\zeta = \frac{\beta}{2}\lim_{\omega\to 0}\lim_{\mathbf{q}\to 0}\sigma_{\bar{P}\bar{P}}, \tag{22}$$

where, $\sigma_{\bar{P}\bar{P}}$ is given by

$$\sigma_{\bar{P}\bar{P}}(\omega,\mathbf{q}) = \frac{1}{2\pi\mathcal{A}}\int d^2\mathbf{x}\int_{-\infty}^{\infty} dt e^{-i\mathbf{q}.\mathbf{x}+i\omega t}\langle\bar{P}(t,\mathbf{x})\bar{P}(0)\rangle, \tag{23}$$

where $\mathcal{A}$ is the area normalization of the spatial part,

$$\bar{P}(t,\mathbf{x}) = P(t,\mathbf{x}) - v_S^2\rho(t,\mathbf{x}) = \frac{1}{2}T_i^i(x^\mu) - v_S^2 T_{00}(x^\mu). \tag{24}$$

and v_S is the speed of sound of the field.

To apply this formalism to horizon-fluid, we need to make suitable changes in the formulation. As mentioned earlier, the stress-tensor of the horizon-fluid vanishes as the infalling matter-energy reaches the horizon and the horizon becomes quasi-stationary.[11–13] In other words, when the matter reaches an equilibrium at

a given temperature, the stress-tensor of the horizon-fluid is zero. Thus, the field-theoretic description of the horizon-fluid corresponds to the deviation of the energy-momentum tensor of the field ($T^H_{\mu\nu}$) from its thermal average at the thermal state ($\langle T^H_{\mu\nu}\rangle_{\rm equilibrium}$). In the case of normal-fluid, we use the stress-tensor (cf. 23). For the horizon-fluid on the other hand, the key quantity is the deviation of the energy-momentum tensor of the field, i.e.,

$$\delta T^H_{\mu\nu} = T^H_{\mu\nu} - \langle T^H_{\mu\nu}\rangle_{\rm equilibrium}\,.$$

Physically, this corresponds to the state when the expectation value of the stress-energy tensor of the perturbed CFT on the horizon is the thermal average. Thus, for the horizon-fluid, Eqs. (23) and (24) become:

$$\delta P(\bar{t},\mathbf{x}) = \frac{1}{2}\delta T^i_i(x^\mu) - \delta T_{00}(x^\mu) \tag{25}$$

$$\sigma_{\delta\bar{P},\delta\bar{P}}(\omega,\mathbf{q}) = \frac{1}{2\pi\mathcal{A}}\int d^2\mathbf{x}\int_{-\infty}^{\infty} dt e^{-i\mathbf{q}\cdot\mathbf{x}+i\omega t}\langle\delta\bar{P}(t,\mathbf{x})\delta\bar{P}(0)\rangle_0. \tag{26}$$

For the horizon-fluid, v_S is equal to the speed of light ($c=1$). For the Hamiltonian (15), we have:

$$\bar{P} = -\frac{1}{2}(\nabla\phi)^2 - m^2_{\rm eff}\phi^2. \tag{27}$$

We can ignore contributions from the $(\nabla\phi)$ term for the homogeneous perturbations responsible for the bulk viscosity. We can determine $\delta T_{\mu\nu}$ by systematically tracking the deviation of the field ϕ from its average value at equilibrium state, i.e.,

$$\phi = \langle\phi\rangle_0 + \delta\phi$$

where $\langle\rangle_0 \equiv \phi_0$ denotes the ensemble average of the density matrix.

For a field at a temperature T, the density matrix is the thermal density matrix at T. In the hydrodynamic limit, ϕ_0 is given by:

$$\phi_0^2 = \frac{\sum e^{-\beta H}\phi^2}{\sum e^{-\beta H}} = \frac{\int\mathcal{D}\phi\phi^2 e^{-\beta\int\frac{m^2_{\rm eff}}{2}\phi^2 d^2x}}{\int\mathcal{D}\phi e^{-\beta\int\frac{m^2_{\rm eff}}{2}\phi^2 d^2x}} = \frac{1}{2m^2_*}, \tag{28}$$

where star denotes that rescaled variables [see the discussion below Eq. (33)]. Writing $\phi = \phi_0 + \delta\phi$, (27) can be written as

$$\delta\bar{P} = -2m^2_*\phi_0\delta\phi = -2\frac{m_*}{\sqrt{2}}\delta\phi = \sqrt{2}m_*\delta\phi = \frac{\sqrt{2}}{C}\delta\phi. \tag{29}$$

The Coefficient of Bulk Viscosity is given by,

$$\zeta = 2\pi\,{\rm Im}\Big[\frac{1}{4\pi\mathcal{A}}\lim_{\mathbf{q}\to0}\lim_{\omega\to m_*}\int d^3x e^{-i\mathbf{q}.\mathbf{x}+i\omega t}\langle\big[\delta\bar{P}(t,\mathbf{x}),\delta\bar{P}(0)\big]\rangle\theta(-t)\Big]. \tag{30}$$

Substituting (29) in (30), and keeping in mind that the perturbations are homogeneous, one gets,

$$\zeta = {\rm Im}\Big[\lim_{\omega\to m_*}\frac{1}{C^2}\int_{-\infty}^{\infty} dt e^{i\omega t}\langle\big[\delta\hat{\phi}(t),\delta\hat{\phi}(0)\big]\rangle\theta(-t)\Big]. \tag{31}$$

$\langle\big[\delta\hat{\phi}(t,\mathbf{x}),\delta\hat{\phi}(0)\big]\rangle\theta(-t)$ corresponds to the advanced Green's function that appears because of the teleological nature of the event horizon [for detailed discussion, see[12,13,21]], and

$$\big[\delta\hat{\phi}(t),\delta\hat{\phi}(0)\big]=\big[\hat{\phi}(t),\hat{\phi}(0)\big] \tag{32}$$

To obtain the above commutation relation of the field operator, we write the Hamiltonian (15) as

$$H_{\rm HF}=\frac{T}{2}\int\left[\Big(\frac{\partial\varphi_*}{\partial t}\Big)^2+\big(\nabla\varphi_*\big)^2+m_*^2\varphi_*^2\right]dA\,. \tag{33}$$

where $\phi=\sqrt{T}\phi_*$, and ϕ_* is the rescaled (dimensionless) effective scalar field and T is taken to be a constant for convenience. Rewriting $\hat{\varphi}_*$ (also $\hat{\varphi}$) in terms of the creation and annihilation operators, i.e.,

$$\hat{\varphi_*}(t,\mathbf{x})=\sum_{\mathbf{k}_*}[\hat{a}_{\mathbf{k}_*}u_{\mathbf{k}_*}(t,\mathbf{x})+\hat{a}^\dagger_{\mathbf{k}_*}u^*_{\mathbf{k}_*}(t,\mathbf{x})], \tag{34}$$

where,

$$u_{\mathbf{k}_*}(t,\mathbf{x})=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2\omega_*}}e^{i(\mathbf{k}_*.\mathbf{x}-\omega_* t)} \tag{35}$$

and $\mathbf{k}_*$, ω_*, $\mathbf{x}$ and t are dimensionless variables. Expressing the dimensionless part of the above Hamiltonian (33) in the frequency domain, we have:

$$H_{\rm eff}(\varphi)=T\sum_{\mathbf{k}_*}(\hat{a}^\dagger_{\mathbf{k}_*}\hat{a}_{\mathbf{k}_*}+\frac{1}{2})\omega_*, \tag{36}$$

where the dispersion relation is given by:

$$\omega_*^2={\mathbf{k}_*}^2+\frac{1}{C^2}\quad\text{and}\quad C=\frac{1}{m_*}\,. \tag{37}$$

In the hydrodynamic limit ($\mathbf{k}_*\to 0$), the above expression reduces to $\omega_{*H}=1/C$. Substituting Eq. (32) in Eq. (31) and after a little algebraic manipulation in terms of dimensionless variables, we have:[47]

$$\zeta[\omega_{*H}]=\mathrm{Im}\left[-\frac{i}{4C^2}\frac{T\omega_{*H}}{E_\beta(\omega_{*H})}\int_{-\infty}^{\infty}\langle\hat{\varphi}(0)\hat{\varphi}(t)+\hat{\varphi}(t)\hat{\varphi}(0)\rangle e^{-i\omega_{*H}t}\theta(-t)dt\right], \tag{38}$$

where $E_\beta(\omega_{*H})$ is the average energy of excitation in the mode with frequency ω_{*H} at temperature T. The above integral exhibits a pole at $\omega'_*=\omega_{*H}$, which is the well-known pole at the hydrodynamical limit. Using the fact that the process corresponding to bulk viscosity is homogeneous, we have

$$\langle\hat{\varphi}^2_{\omega_H}(0)\rangle=\frac{1}{(2\pi)^2}\frac{1}{4m_*^2}=\frac{1}{(2\pi)^2}\frac{C^2}{4} \tag{39}$$

In the hydrodynamic limit, we get:

$$\zeta[\omega_{*H}]=-\frac{1}{128\pi^2}\frac{T\omega_{*H}}{E_\beta(\omega_{*H})}=-\frac{1}{128\pi^2}2\tanh\frac{1}{2C}\,. \tag{40}$$

Demanding that the above expression matches with the expression derived by Damour[11]:

$$\zeta = -\frac{1}{16\pi}\,,$$

we have

$$\tanh\frac{1}{2C} = \frac{1}{4\pi}. \tag{41}$$

Solving this equation numerically, we get, $C = 6.2696$ and hence, $m_* = 0.1595$. The negative sign of a transport coefficient arises due to $\theta(-t)$ in the response function of the black-hole horizon.[21]

6. Discussion

The minimal microscopic model for the horizon-fluid discussed here incorporates three simple ideas. First, the model contains an infinite-dimensional symmetry algebra, namely the Virasaro algebra, corresponding to black-holes' near-horizon $S1$ diffeo-symmetry. Second, the model describes a perturbed CFT that possesses a representation of the Virasoro algebra. The deformation is characterized by a mass term and is supposed to address the perturbed nature of the non-stationary black-hole horizon. Finally, the mass gap provides a simple derivation of ζ and area quantization, thus demonstrating that in the long-wavelength limit, it can connect with the known semiclassical Physics of black-holes. Historically, Bekenstein was the first to derive the quantization of area using adiabatic quantization. We obtained quantization of entropy by slowly evolving the black-hole horizon. Adiabatic quantization of a system is applicable when some of the parameters characterizing the system evolve slowly.

In this talk, only the homogeneous perturbations of the stationary black-hole have been considered. We need to construct a microscopic model, including general perturbations, which can describe the horizon-fluid. This complete theory would, of course, be richer than the toy model we have put forward. So the Baxter model serves only as an illustration in this sense. Nonetheless, an improved model should also share the first three critical features of the current model mentioned above. Thus, we expect a more comprehensive microscopic future model of the horizon-fluid to describe a system close to the critical point. It should represent the near-horizon symmetry algebra and a mass gap with a fixed, known value.

Our results also lead to an interesting suggestion. The twin requirements of the theory incorporating near-horizon symmetries($S1$ diffeomorphism) and possessing length scales due to external perturbations can be naturally satisfied if the theory on the non-stationary black-hole horizon is a deformed CFT. Similar types of arguments also lead to the expectation that the theory on the stationary black-hole horizon is a CFT.[19] Since the non-stationary black-hole eventually becomes stationary, it is reasonable to assume that the deformed CFT flows into a CFT as the black-hole horizon dynamically evolves. Note that the above argument does not

depend on any specific model. This leads us to conjecture that *the low energy theory on the black-hole horizon flows towards a critical point as the perturbed black-hole ultimately becomes stationary.* We think that this conjecture may be a good starting point to construct quantum theory models of the black-hole horizon.

References

1. Robert M. Wald. The thermodynamics of black holes. *Liv. Rev. Rela.*, 4:6, 2001.
2. Ted Jacobson and Renaud Parentani. Horizon entropy. *Found. Phys.*, 33(2):323–348, 2003.
3. Steven Carlip. Black Hole Thermodynamics and Statistical Mechanics. *Lect. Notes Phys.*, 769:89–123, 2009.
4. James M. Bardeen, B. Carter, and S. W. Hawking. The four laws of black hole mechanics. *Commun. Math. Phys.*, 31:161–170, 1973.
5. Ted Jacobson. Thermodynamics of space-time: The Einstein equation of state. *Phys. Rev. Lett.*, 75:1260–1263, 1995.
6. T. Padmanabhan. Thermodynamical aspects of gravity: new insights. *Rep. Prog. Phys.*, 73(4):046901, April 2010.
7. T. Padmanabhan. Distribution Function of the Atoms of Spacetime and the Nature of Gravity. *Entropy*, 17:7420–7452, October 2015.
8. S. W. Hawking. Particle creation by black holes. *Commun. Math. Phys.*, 43:199–220, 1975.
9. Saurya Das and S. Shankaranarayanan. Where are the black hole entropy degrees of freedom? *Class. Quant. Grav.*, 24:5299–5306, 2007.
10. Paul Romatschke and Ulrike Romatschke. *Relativistic Fluid Dynamics In and Out of Equilibrium.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2019.
11. T. Damour. Surface Effects in Black-Hole Physics. In R. Ruffini, editor, *Marcel Grossmann Meeting: General Relativity*, 1982.
12. R. H. Price and K. S. Thorne. Membrane viewpoint on black holes: Properties and evolution of the stretched horizon. *Phys. Rev. D.*, 33:915–941, February 1986.
13. Kip S Thorne, Richard H Price, and Douglas A Macdonald. *The Membrane Paradigm*, volume 19. Yale University Press, New Haven, 1986.
14. Romesh K. Kaul and Parthasarathi Majumdar. Quantum black hole entropy. *Phys. Lett.*, B439:267–270, 1998.
15. Romesh K. Kaul and Parthasarathi Majumdar. Schwarzschild horizon dynamics and SU(2) Chern-Simons theory. *Phys. Rev.*, D83:024038, 2011.
16. Jun-ichirou Koga. Asymptotic symmetries on Killing horizons. *Phys. Rev. D*, D64:124012, 2001.
17. M. Hotta, K. Sasaki, and T. Sasaki. Diffeomorphism on horizon as an asymptotic isometry of Schwarzschild black hole. *Class. Quant. Grav.*, 18:1823–1834, 2001.
18. M. Hotta. Holographic charge excitations on horizontal boundary. *Phys. Rev.*, D66:124021, 2002.
19. S. Carlip. Black hole entropy from conformal field theory in any dimension. *Phys. Rev. Lett.*, 82:2828, 1999.
20. Steven Carlip. Effective Conformal Descriptions of Black Hole Entropy. *Entropy*, 13:1355–1379, 2011.
21. Swastik Bhattacharya and S. Shankaranarayanan. Fluctuations in horizon-fluid lead to negative bulk viscosity. *Phys. Rev. D*, D93(6):064030, 2016.

22. Bethan Cropp, Swastik Bhattacharya, and S. Shankaranarayanan. Hints of quantum gravity from the horizon fluid. *Phys. Rev. D.*, D95(2):024006, 2017.
23. S. Chandrasekhar. *The mathematical theory of black holes*. Oxford Univ. Press, Oxford, 1992.
24. Glenn Barnich and Cedric Troessaert. Symmetries of asymptotically flat 4 dimensional spacetimes at null infinity revisited. *Phys. Rev. Lett.*, 105:111103, 2010.
25. Laura Donnay, Gaston Giribet, Hernan A. Gonzalez, and Miguel Pino. Supertranslations and Superrotations at the Black Hole Horizon. *Phys. Rev. Lett.*, 116(9):091101, 2016.
26. H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner. Gravitational waves in general relativity. 7. Waves from axisymmetric isolated systems. *Proc. Roy. Soc. Lond.*, A269:21–52, 1962.
27. R. K. Sachs. Gravitational waves in general relativity. 8. Waves in asymptotically flat space-times. *Proc. Roy. Soc. Lond.*, A270:103–126, 1962.
28. Andrew Strominger. Lectures on the Infrared Structure of Gravity and Gauge Theory. 2017.
29. Geoffrey Compére. The Kerr/CFT correspondence and its extensions. *Living Rev. Rel.*, 15:11, 2012.
30. AB Zamolodchikov. Higher-order integrals of motion in two-dimensional models of the field theory with a broken conformal. *JETP Lett*, 46(4), 1987.
31. Alexander B Zamolodchikov. Integrable field theory from conformal field theory. In *Integrable Systems and Quantum Field Theory*, pages 641–674. Elsevier, 1989.
32. Rodney J Baxter. Eight-vertex model in lattice statistics. *Physical Review Letters*, 26(14):832, 1971.
33. RJ Baxter and FY Wu. Exact solution of an ising model with three-spin interactions on a triangular lattice. *Physical Review Letters*, 31(21):1294, 1973.
34. R. J. Baxter. Corner transfer matrices of the eight-vertex model. i. low-temperature expansions and conjectured properties. *Journal of Statistical Physics*, 15(6):485–503, Dec 1976.
35. R. J. Baxter. Corner transfer matrices of the eight-vertex model. ii. the ising model case. *Journal of Statistical Physics*, 17(1):1–14, Jul 1977.
36. Rodney J Baxter. Variational approximations for square lattice models in statistical mechanics. *Journal of Statistical Physics*, 19(5):461–478, 1978.
37. Rodney J Baxter. *Exactly solved models in statistical mechanics*. Elsevier, 2016.
38. Leo P Kadanoff and Franz J Wegner. Some critical properties of the eight-vertex model. *Physical Review B*, 4(11):3989, 1971.
39. H. Itoyama and H. B. Thacker. Lattice Virasoro Algebra and Corner Transfer Matrices in the Baxter Eight Vertex Model. *Phys. Rev. Lett.*, 58:1395, 1987.
40. HB Thacker. Corner transfer matrices and lorentz invariance on a lattice. *Physica D: Nonlinear Phenomena*, 18(1-3):348–359, 1986.
41. H. Itoyama and H. B. Thacker. Integrability and Virasoro Symmetry of the Noncritical Baxter-Ising Model. *Nucl. Phys.*, B320:541–590, 1989.
42. S. Jeon, Phys. Rev. D **52**, 3591-3642 (1995) doi:10.1103/PhysRevD.52.3591 [arXiv:hep-ph/9409250 [hep-ph]].
43. Bill Sutherland. Two-dimensional hydrogen bonded crystals without the ice rule. *Journal of Mathematical Physics*, 11(11):3183–3186, 1970.
44. J.-L. Gervais. Infinite family of polynomial functions of the virasoro generators with vanishing poisson brackets. *Physics Letters B*, 160(4):277–278, 1985.
45. A M. Polyakov. *Gauge Fields and Strings*. Contemporary concepts in physics. Taylor & Francis, 1987.

46. Oliver Penrose and Paul C Fife. Thermodynamically consistent models of phase-field type for the kinetic of phase transitions. *Physica D*, 43(1):44–62, 1990.
47. R. Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29:255–284, January 1966.
48. R. Kubo, J. Phys. Soc. Jap. **12**, 570-586 (1957) doi:10.1143/JPSJ.12.570
49. J. D. Bekenstein. The quantum mass spectrum of the kerr black hole. *Lettere al Nuovo Cimento (1971-1985)*, 11(9):467–470, 1974.

Einstein-Maxwell-Dilaton-Axion mass formulas for black holes with struts and strings

Dmitri Gal'tsov[a,*], Gérard Clément[b,†] and Igor Bogush[a,‡]

[a] *Faculty of Physics, Moscow State University, 119899, Moscow, Russia*
[b] *LAPTh, Université Savoie Mont Blanc, CNRS, France*
Email: [*] *galtsov@phys.msu.ru,* [†] *gclement@lapth.cnrs.fr,* [‡] *igbogush@gmail.com*

Mass formulas are obtained for stationary axisymmetric solutions of the Einstein-Maxwell dilaton-axion theory, which have a regular rod structure on the axis of symmetry. Asymptotic mass, angular momentum and charge are expressed as the sums of masses, angular momenta and charges of rods dressed with field contributions. The calculation is based on a three-dimensional sigma model representation of the stationary EMDA system and the Tomimatsu approach proposed for the Einstein-Maxwell system. Our results provide an alternative interpretation of mass formulas and thermodynamics for black holes with Dirac and Misner strings. It is also applicable to aligned multiple black holes with struts.

Keywords: Black hole mass formulas, Dirac string, Misner string, therodynamics of black holes with NUT

1. Introduction

Mass formulas for black holes and the concept of irreversible mass were proposed by Christodoulou,[1] Hawking,[2] Christodoulou and Ruffini,[3] and Smarr[4] in the early 1970s, shortly before Hawking discovered the evaporation of a black hole.[5] They have played an important role in understanding the energy extraction from rotating black holes and the energy balance of merging black holes, which was brilliantly confirmed in the recent experiments of Ligo. The thermodynamics of black holes[6–9] then gave them a deep quantum interpretation. The mathematical foundations and detailed derivation of mass formulas in the Einstein-Maxwell theory were given by Carter.[10]

The original integral Smarr's mass formula[4] relates the total values of mass, angular momentum and electric charge of black holes in Einstein-Maxwell theory with the horizon area. The area term in this formula was originally interpreted mechanically as the work of *stresses* of the horizon. Further, this interpretation was forgotten in favor of the thermodynamic one. When interest turned to solutions with NUT, it was immediately found,[11–14] that the Misner string also contributes to the Smarr mass formula, and this contribution was included in the entropy term. This interpretation was revived recently in slightly different terms in a series of papers,[15–19] based on the Bonnor interpretation of the Misner strings as mild physical singularities.[20]

In our opinion, the most natural description of solutions with Misner and Dirac strings as well as aligned multicenter solutions with struts can be given in terms of

the rod structure, introduced by Harmark,[21] developing earlier ideas of Emparan and Reall.[22] The main novelty of this approach is the derivation of separate formulas for the partial masses for each component of the entire system, represented by a set of rods. In this interpretation, Misner's strings appear as independent components on the same base and as black holes.

Using the representation of vacuum stationary axisymmetric solutions in Weyl coordinates, we consider the solutions as generated by data on the polar axis, which look like distributional matter objects that are sources of the Poisson equation for the gravitational potential.[23] Accordingly, the total gravitational field can be viewed as a nonlinear superposition of the components such as black holes, struts, Misner and Dirac strings. Despite the nonlinearity of Einstein's equations, the total mass, angular momentum (and electric charge in the case of an electrovacuum) can be represented as the sum of the individual contributions of the constituents. This simple additivity is associated with conservation laws for the Komar and Gauss integrals. The magnetic charge and magnetic mass (NUT) in this description are due to the Misner and Dirac strings, which correspond to the individual elements in the rod system, so these charges do not enter the black hole horizon mass. Their contribution enters the asymptotic Komar mass as the proper masses of the Dirac and Misner strings along with the black hole contribution.

The difference between the horizon and the string rods is that the directions of the former are timelike, and the directions of the latter are spacelike. Both rods are Killing horizons and have an associated surface gravity. But the entropic interpretation of the surface contribution of spacelike rods (often encountered) does not seem convincing. An alternative interpretation may be similar to Smarr's original point of view.

This programme was previously performed for the Einstein-Maxwell system,[24,25] and here it is extended to Einstein-Maxwell-dilaton-axion gravity (EMDA). We show that Tomimatsu's proposal within the context of electrovacuum[26,27] on calculating Komar integrals over rods in terms of the boundary values of Ernst's potentials can be generalized to the sigma model representation of stationary dilaton-axion gravity. Surprisingly, the obtained mass formulas are very similar to the formulas obtained in the case of Einstein-Maxwell.

2. Stationary EMDA gravity

The EMDA gravity can be viewed as a consistent truncation of a toroidal reduction of the heterotic string,[28] or as a truncation of $N = 4$ four-dimensional supergravity.[29,30] Rotating electrically charged black hole solution of EMDA theory was obtained by Sen[28] within the first approach and independently, with the inclusion of NUT and magnetic charge, by Gal'tsov and Kechkin[31] in the second context. Recently, these solutions have attracted attention as an alternative to the Kerr metric in an effort to find astrophysical evidence for new physics.[32–34]

The mass formulas and thermodynamics of EMDA black holes have been discussed frequently in the past, though not in full generality, see, for example[35,36] (a more complete list of references prior to 2008 can be found in[36]). Our discussion here, based on generalization of,[25] covers solutions with Misner and Dirac strings, and is also applicable to hitherto unknown solutions for double black holes with struts that are expected to exist.

The EMDA action in our conventions reads:

$$S=\frac{1}{16\pi}\int\left\{R-2\partial_\mu\phi\partial^\mu\phi-\frac{1}{2}\mathrm{e}^{4\phi}\partial_\mu\kappa\partial^\mu\kappa-\mathrm{e}^{-2\phi}F_{\mu\nu}F^{\mu\nu}-\kappa F_{\mu\nu}\tilde{F}^{\mu\nu}\right\}\sqrt{-g}d^4x, \tag{1}$$

where $\tilde{F}^{\mu\nu}=\frac{1}{2}E^{\mu\nu\lambda\tau}F_{\lambda\tau}$. Dilaton ϕ and axion κ parameterize a coset $SL(2,R)/SO(2)$, and the $SL(2,R)$ group is the symmetry of the full action. To see this, one can introduce the complex scalar

$$\zeta=\kappa+i\mathrm{e}^{-2\phi} \tag{2}$$

and the self-dual Maxwell field $\mathcal{F}=\left(F+i\tilde{F}\right)/2$. Then the action takes the form

$$S=\frac{1}{16\pi}\int\left\{R-2\left|\partial\zeta(\zeta-\bar{\zeta})^{-1}\right|^2+(i\zeta\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu}+c.c.)\right\}\sqrt{-g}d^4x, \tag{3}$$

which is invariant under the $SL(2,R)$ transformations

$$\begin{aligned}&\zeta\to\frac{\alpha\zeta+\beta}{\gamma\zeta+\delta},\qquad \alpha\beta-\gamma\delta=1,\\ &F\to(\gamma\kappa+\delta)F+\gamma\mathrm{e}^{-2\phi}\tilde{F},\end{aligned} \tag{4}$$

which interchange the modified Maxwell equations and the Bianchi identity

$$\nabla_\nu G^{\mu\nu}=0,\quad \nabla_\nu\tilde{F}^{\mu\nu}=0, \tag{5}$$

$$G^{\mu\nu}=e^{-2\phi}F^{\mu\nu}+\kappa\tilde{F}^{\mu\nu}. \tag{6}$$

The Einstein equations are

$$R_{\mu\nu}=2\phi_{,\mu}\phi_{,\nu}+\frac{1}{2}\mathrm{e}^{4\phi}\kappa_{,\mu}\kappa_{,\nu}-\mathrm{e}^{-2\phi}\left(2F_{\mu\lambda}F^{\lambda}{}_{\nu}+\frac{1}{2}F^2g_{\mu\nu}\right). \tag{7}$$

For stationary and axisymmetric configurations the axion and dilaton kinetic term disappear from the Einstein equations in the $t-\varphi$ sector, and the dilaton enters only through the scale factor $\mathrm{e}^{-2\phi}$ in front of the Maxwell energy-momentum tensor. This is crucial for our derivation.

Assuming the existence of the Killing vector $k=\partial_t$, the metric can be represented as

$$ds^2=g_{\mu\nu}dx^\mu dx^\nu=-f(dt-\omega_i dx^i)^2+\frac{1}{f}h_{ij}dx^i dx^j, \tag{8}$$

where three-dimensional metric h_{ij}, three-dimensional rotation vector ω_i, $(i,j=1,2,3)$ and the scale factor f depend only on the space coordinates x^i. The spatial

parts of the Bianchi identities and the modified Maxwell equations can be solved by introducing electric v and magnetic u potentials

$$F_{i0} = \frac{1}{\sqrt{2}}\partial_i v, \tag{9}$$

$$G^{ij} = -\frac{f}{\sqrt{2h}}\epsilon^{ijk}\partial_k u. \tag{10}$$

The mixed R_0^i Einstein equations are solved by introducing the twist potential χ obtained by dualizing the rotation two-form:

$$f^2 h_{il}\frac{\epsilon^{ljk}}{\sqrt{h}}\partial_j\omega_k = \partial_i\chi + v\partial_i u - u\partial_i v. \tag{11}$$

The remaining equations reduce to those of three-dimensional gravitating sigma-model equations for six scalars $X^A = f, \chi, v, u, \phi, \kappa,\ A = 1, \ldots, 6$, and the three-dimensional metric h_{ij}:

$$S = \int \left[\mathcal{R}(h) - \mathcal{G}_{AB}(X)\,\partial_i X^A\,\partial_j X^B\,h^{ij}\right]\sqrt{h}\,d^3x, \tag{12}$$

where the target space metric $\mathcal{G}_{AB}$ can be presented as

$$dl^2 = \frac{1}{2}f^{-2}[df^2 + (d\chi + vdu - udv)^2] - f^{-1}[e^{2\phi}(du - \kappa dv)^2 + e^{-2\phi}dv^2] + 2d\phi^2 + \frac{1}{2}e^{4\phi}d\kappa^2. \tag{13}$$

The isometry group of this metric is $SO(3,2)$, as was identified in.[37] Based on isomorphism $SO(3,2) \sim Sp(4,R)$, a convenient 4×4 matrix representation of the coset $Sp(4,R)/SO(1,2)$ was suggested,[38] suitable for the generation technique. In terms of complex coordinates, one of which is (2) and two other are the following generalization of the Ernst potentials[39]

$$\Psi = u - \zeta v, \quad \mathcal{E} = if - \chi + v\Psi, \tag{14}$$

the target space is a three-dimensional Kähler space.

3. Rod structure

Consider the spacetime metric, admitting two commuting Killing vectors k, m corresponding to stationarity and axial symmetry. Let t, φ be chosen so that $k = \partial_t$, $m = \partial_\phi$ with $t \in R$ and $\varphi \in [0, \pi]$. The remaining coordinates will be primarily assumed of the Boyer-Lindquist type, r, θ with $0 < r < \infty$ and $\theta \in [0, \pi]$. We assume that the spacetime manifold has no strong naked curvature singularities, but can have visible line singularities along (a part of) the polar axis in the sense of Israel.[23] These includes cosmic strings (conical singularities), struts in aligned multiple black hole solutions (conical singularities of both positive or negative tension) and Misner strings in spacetimes with NUTs, without time periodicity being imposed. We also assume that spacetime is asymtotically flat, or

asymptotically locally flat. To such a spacetime one can ascribe a *rod structure* following Harmark.[21]

Let $\gamma_{ab},\ x^a = t,\ \varphi$ is the two-dimensional Lorentzian metric of the subspace spanned by the Killing vectors. Introduce the Weyl cordinates ρ, z such that ρ is

$$\rho = \sqrt{|\det \gamma|}\,, \tag{15}$$

and z to ensure the metric form:

$$ds^2 = \gamma_{ab}(\rho, z)dx^a dx^b + e^{2\nu}(d\rho^2 + dz^2)\,, \tag{16}$$

where ν is a function of (ρ, z). To find the rod structure, one has to solve the equation

$$\rho(r, \theta) = 0. \tag{17}$$

Generically, the solution splits the polar axis into a certain number of finite or semi-infinite intervals $(-\infty, z_1], [z_1, z_2], \ldots, [z_N, +\infty)$ called rods (we will label two semi-infinite rods by $n = \pm$, and the remaining finite ones by an index n corresponding to the left bound of the interval). Each rod can be equipped with a two-dimensional vector, called *rod direction*, via the following reasoning. At $\rho = 0$ the matrix $\gamma_{ab}(0, z)$ is degenerate by virtue of the definition (15), so it must have zero eigenvalues. For quasiregular spacetimes, such that on the symmetry axis there are no strong curvature singularities these eigenvalues must be non-degenerate except, perhaps, for a discrete set of "turning" points z_n which mark the ends of rods. The eigenvector v^a_n, satisfying the equation

$$\gamma_{ab}(0, z)v^b_n = 0 \tag{18}$$

on any segment $z \in [z_n, z_{n+1}]$, is called the n-s rod direction. By continuity, this vector can be extended to small $\rho \neq 0$, and a more accurate analysis shows that, in the leading order in ρ, its 2D norm behaves as

$$v^2_n = \gamma_{ab}v^a_n v^b_n \sim \pm a(z)\rho^2, \qquad e^{2\nu} \sim c^2 a(z), \tag{19}$$

where c is some constant and the sign $\pm$ corresponds to *spacelike* and *timelike* rod respectively.

Normalization of directional vectors can be chosen in different ways. One possibility, *Killing* normalization, relates to preferred normalisation an associated Killing vector in spacetime: $V = v^0 k + v^1 m$, namely, $v_0 = 1$ in the timelike case, and $v_1 = 1$ in spacelike. In view of behavior near the axis (19), in both cases the rod itself (i.e. the submanifold $\rho = 0$) will be the Killing horizon for so defined Killing vector. Another possibility is to choose the scale factor in such a way as to ensure the finiteness of its norm at $\rho \to 0$:

$$l^a_n = \lim_{\rho \to 0} \rho^{-1} e^{-\nu} v^a_n, \tag{20}$$

again with an option of further constant rescaling. An important property of so defined rod direction is its constancy (in view of (19) along the rod, i.e.,

$$\partial_z l_n^a = 0. \tag{21}$$

Timelike rods of finite length correspond to black hole horizons, infinite timelike rods describe acceleration horizons. The constant components of such a horizon rod are connected with its angular velocity Ω_H and the surface gravity $\varkappa_H$ of the Killing vector

$$\xi = k + \Omega_H\, m\,. \tag{22}$$

The surface gravity

$$\varkappa_H = (-\xi_{\mu;\nu}\xi^{\mu;\nu}/2)^{1/2} \tag{23}$$

can be expressed in the Weyl coordinates as

$$\varkappa_H = \lim_{\rho\to 0}\left(-\rho^{-2}e^{-2\nu}\gamma_{ab}v^a v^b\right)^{1/2}, \tag{24}$$

with $v^0 = 1$, $v^1 = \Omega_H$. It may be convenient to choose a canonical normalization of the directional vectors that corresponds to the unit surface gravity:

$$l_H = (1/\varkappa_H,\ \Omega_H/\varkappa_H)\,. \tag{25}$$

Spacelike rods correspond to line defects, such as cosmic strings, struts, Misner strings, which can be also Dirac strings associated with the corresponding vector potentials. These rods are Killing horizons for some spacelike Killing vectors. For them one also has constant angular velocities Ω_n and *spacelike* surface gravities

$$\varkappa_n = \lim_{\rho\to 0}\left(\rho^{-2}e^{-2\nu}\gamma_{ab}v_n^a v_n^b\right)^{1/2}, \tag{26}$$

so that the normalized directional vectors will be

$$l_n = (1/\varkappa_n,\ \Omega_n/\varkappa_n)\,. \tag{27}$$

Here $v_n^1 = 1$. With the normalized spacelike directional vector, the period (28) will be 2π. If the coordinate η associated with the spacelike rod Killing vector $l = \partial_\eta$, conical singularity is absent with perdiocity of η with

$$\Delta\eta = 2\pi \lim_{\rho\to 0}\left(\rho^2 e^{2\nu}\left(\gamma_{ab}l_n^a l_n^b\right)^{-1}\right)^{1/2}. \tag{28}$$

This is important in the situation when two spacelike rods meet at the turning point, where potentially a mismatch of periodicities may lead to *orbifold* singularities. This may happen for instantons.[41]

4. Komar charges

Following,[25] we start with the Komar definition[40] of the asymptotic mass and angular momentum:

$$M_\infty = \frac{1}{4\pi}\oint_{\Sigma_\infty} D^\nu k^\mu d\Sigma_{\mu\nu}, \quad J_\infty = -\frac{1}{8\pi}\oint_{\Sigma_\infty} D^\nu m^\mu d\Sigma_{\mu\nu}. \tag{29}$$

We also need a conserved electric charge as the surface integral. The definition of a conserved electric charge in EMDA follows from the modified Maxwell equations:

$$Q_\infty = \frac{1}{4\pi}\oint_{\Sigma_\infty} G^{\mu\nu} d\Sigma_{\mu\nu}. \tag{30}$$

Consider some stationary axisymmetric solution with an arbitrary rod structure z_n, l_n. Each rod must be surrounded by a thin cylinder Σ_n. Using Ostrogradski formula one can transform the asymptotic total mass to the sum of the local Komar masses M_n^d of rods ("direct" masses) and the bulk contribution coming from the fields: $M = \sum_n M_n^d + M_F$, where

$$M_n^d = \frac{1}{4\pi}\oint_{\Sigma_n} D^\nu k^\mu d\Sigma_{\mu\nu}, \qquad M_F = \frac{1}{4\pi}\int D_\nu D^\nu k^\mu dS_\mu, \tag{31}$$

where Σ_n are the spacelike surfaces bounding the various sources[a], and the second integral is over the bulk. Similarly, the asymptotic angular momentum will read $J = \sum_n J_n^d + J_F$, where

$$J_n^d = -\frac{1}{8\pi}\oint_{\Sigma_n} D^\nu m^\mu d\Sigma_{\mu\nu}, \qquad J_F = -\frac{1}{8\pi}\int D_\nu D^\nu m^\mu dS_\mu, \tag{32}$$

Using the well-known Killing lemma for k,

$$D_\nu D^\nu k^\mu = -[D_\nu, D^\mu]k^\nu = -R^\mu{}_\nu k^\nu, \tag{33}$$

and similarly for m, and an explicit form for the Ricci tensor in EMDA theory (7), one can express the bulk integrals as

$$M_F = -\frac{1}{4\pi}\int \mathrm{e}^{-2\phi}\left(F_{it}F^{it} - F_{i\varphi}F^{i\varphi}\right)\sqrt{|g|}d^3x, \tag{34}$$

$$J_F = \frac{1}{4\pi}\int \mathrm{e}^{-2\phi}F_{i\varphi}F^{it}\sqrt{|g|}\,d^3x. \tag{35}$$

Using the identities

$$F_{it}\tilde{F}^{it} - F_{i\varphi}\tilde{F}^{i\varphi} = 0, \qquad F_{i\varphi}\tilde{F}^{it} = 0 \tag{36}$$

[a]We use the metric signature (-+++) and the convention $d\Sigma_{\mu\nu} = 1/2\sqrt{|g|}\,\epsilon_{\mu\nu\lambda\tau}dx^\lambda dx^\tau$ with $\epsilon_{t\rho z\varphi} = 1$ in Weyl coordinates. We will label t, φ by an index a, and the remaining coordinates ρ, z by i, j. The two-dimensional Levi-Civita symbol ϵ_{ij} is defined with $\epsilon_{\rho z} = 1$ and $\epsilon_{xy} = 1$ respectively.

one can rewrite (34) and (35) as

$$M_F = -\frac{1}{4\pi}\int \left(F_{it}G^{it} - F_{i\varphi}G^{i\varphi}\right)\sqrt{|g|}d^3x, \tag{37}$$

$$J_F = \frac{1}{4\pi}\int F_{i\varphi}G^{it}\sqrt{|g|}\,d^3x. \tag{38}$$

In view of the modified Maxwell equations, these expressions can be presented in the full divergence form

$$M_F = -\frac{1}{4\pi}\int \partial_i\left[\sqrt{-g}\left(A_tG^{it} - A_\varphi G^{i\varphi}\right)\right]d^3x, \tag{39}$$

$$J_F = \frac{1}{4\pi}\int \partial_i\left(\sqrt{|g|}A_\varphi G^{it}\right)\,d^3x, \tag{40}$$

suggesting the representation of the asymptotic mass and angular momentum as the surface integrals over the rods[b]:

$$M_\infty = \sum_n M_n, \qquad M_n = \frac{1}{8\pi}\oint_{\Sigma_n}\left(g^{ij}g^{ta}\partial_j g_{ta} + 2\left(A_tG^{it} - A_\varphi G^{i\varphi}\right)\right)d\Sigma_i \tag{41}$$

$$J_\infty = \sum_n J_n, \qquad J_n = -\frac{1}{16\pi}\oint_{\Sigma_n}\left(g^{ij}g^{ta}\partial_j g_{\varphi a} + 4A_\varphi G^{it}\right)d\Sigma_i, \tag{42}$$

where we have also rewritten the direct Komar integrals in an explicit form. Thus we have succeeded in presenting the EMDA bulk contributions in the form of the integrals over the rods in the same was as it was done in[25] for the Einstein-Maxwell system. This procedure can be regarded as rod's dressing by the bulk field. So the resulting Smarr formulas for the asymptotic mass and angular momentum were written as sums of the dressed rod's contributions. Note, that in the general case, this is the only representation for the global M, J that can be found, since different forms of rod's contribution do not allow one to write asymptotic quantities directly in terms of physical charges as was possible for a single black hole.

From now on we assume the standard Weyl-Papaperou parametrization of the metric

$$ds^2 = -f(dt - \omega d\varphi)^2 + f^{-1}[e^{2k}(d\rho^2 + dz^2) + \rho^2 d\varphi^2], \tag{43}$$

and represent the four-potential as

$$\sqrt{2}\,A_\mu dx^\mu = vdt + Ad\varphi, \tag{44}$$

(note that v, u, A in this paper differs from these in[25] by $\sqrt{2}$ to be consistent with the EMDA literature). Computation of the direct rod Komar integrals integrals was given in.[4] It amounts to dualizing the rotation three-form

$$\partial_i\chi = -f^2\rho^{-1}\epsilon_{ij}\partial_j\omega + \left(u\partial_i v - v\partial_i u\right), \tag{45}$$

[b]We assume that the integrals over in infinite sphere vanish, otherwise they must be added too. This happens if the North and South Misner and Dirac strings are arranged non-symmetrically, by adding suitable constants to the asymptotic values of the rotation function ω and the azimuthal component of the four-potential A_φ.

and leads to

$$M_n^d = \frac{1}{8\pi}\int_n \omega\left[\partial_z\chi + (v\partial_z u - u\partial_z v)\right] dzd\varphi \tag{46}$$

Similarly for the angular momentum

$$J_n^d = -\frac{1}{16\pi}\int_n \omega\left\{2 - \omega\left[\partial_z\chi + (v\partial_z u - u\partial_z v)\right]\right\} dzd\varphi \tag{47}$$

Note the appearance of the electromagnetic potentials in the expressions for direct rod mass and angular momentum due to use of dualized twist potential, i.e., implicit use of Einstein equations. These terms will be exactly canceled by the field contributions to the same quantities.

To continue, we show that, apart from constancy (i.e. z-independence) of the rod angular velocity $\Omega = 1/\omega$ and the generalized surface gravity $\varkappa$, one can prove for any rod the following relations:

$$\lim_{\rho\to 0}\left(G^{\rho t} - \omega G^{\rho\varphi}\right)\sqrt{|g|} = 0, \tag{48}$$

$$\lim_{\rho\to 0}\left(v + A/\omega\right) \equiv -\sqrt{2}\Phi = \text{const}, \tag{49}$$

where Φ can be interpreted as an electric potential of the rod. Together with constancy of ω, the last relation gives also

$$\partial_z A = -\omega\partial_z v, \tag{50}$$

as $\rho \to 0$.

Using (48-50), and the EMDA dualization equation (10) one can rewrite the field contribution to the dressed rod mass (34) as

$$M_n^F = -\frac{1}{8\pi}\oint_{\Sigma_n}\left[\omega\left(v\partial_z u - u\partial_z v\right) - \partial_z(uA)\right] dzd\varphi. \tag{51}$$

Combining this with the direct Komar mass (46) we obtain the dressed rod's mass

$$M_n = \frac{1}{8\pi}\int_{\Sigma_n}\left[\omega\partial_z\chi + \partial_z(A\,u)\right] dzd\varphi. \tag{52}$$

It may seem surprising that this formula does not differ from that for Einstein-Maxwell system. No explicit scalar terms are seen.

For the bulk contribution to the angular momentum, one also applies the Eq. (48) and then magnetic dualization equation (10). The subsequent transformations are the same as in[25] and lead to the same result

$$J_n^F = \frac{1}{8\pi}\int_H \left[\omega^2 u\partial_z v + \omega\partial_z(Au)\right] dzd\varphi. \tag{53}$$

Combining with the direct Komar momentum, we obtain for the dressed rod:

$$J_n = \frac{1}{16\pi}\int_H \omega\left[-2 + \omega\partial_z\chi + \partial_z(A\,u) - \sqrt{2}\omega\Phi_n\partial_z u\right] dzd\varphi, \tag{54}$$

this expression also does not contain the dilaton and axion. Now consider Komar electric charge of a rod (30). Explicitly, it is given by the flux

$$Q_n = \frac{1}{4\pi}\int_{\Sigma_n}\sqrt{|g|}G^{t\rho}dzd\varphi. \tag{55}$$

Manipulating with indices and using dualization equation (10)

$$G^{t\rho} = \frac{g^{\rho\rho}}{g_{tt}}G_{t\rho} - \frac{g_{t\varphi}}{g_{tt}}G^{\varphi i} = e^{-2k}\left[G_{t\rho} + \frac{f\omega}{\rho}\partial_z u\right], \tag{56}$$

one finds that in the limit $\rho \to 0$ only the second term contributes, so

$$Q_n = \frac{1}{4\sqrt{2}\pi}\int_{\Sigma_n}\omega\partial_z u\, dzd\varphi. \tag{57}$$

After integration over φ

$$M_n = \frac{\omega_n}{4}\chi\Big|_{z_n}^{z_{n+1}} + \frac{1}{4}(Au)\Big|_{z_n}^{z_{n+1}} \tag{58}$$

Similarly, for the angular momentum (54) one finds

$$J_n = \frac{\omega_n}{8}\left\{-2(z_{n+1}-z_n) + \left[\omega_n\left(\chi - \sqrt{2}\Phi_n u\right) + Au\right]\Big|_{z_n}^{z_{n+1}}\right\}. \tag{59}$$

Finally, the electric charge will be

$$Q_n = \frac{\omega_n}{2\sqrt{2}}u\Big|_{z_n}^{z_{n+1}}. \tag{60}$$

The angular velocity of a rod is defined as a limit

$$\Omega_n = \lim_{\rho\to 0}\ \omega^{-1}(\rho, z), \qquad z_n < z < z_{n+1}. \tag{61}$$

Combining the above formulas we obtain rod Smarr mass

$$M_n = \frac{L_n}{2} + 2\Omega_n J_n + \Phi_n Q_n, \tag{62}$$

where $L_n = z_{n+1} - z_n$ is the n-s rod length (for infinite rods some length regularization is needed). Please note that this formula is identical; it is valid for any rod, regardless of the specific parameter values.

5. Length versus entropy

Although it looks like a one-dimensional object, any rod with finite surface gravity has a finite surrounding cylindrical area:

$$\mathcal{A}_n = \oint d\varphi\int_{z_n}^{z_{n+1}}\sqrt{|g_{zz}g_{\varphi\varphi}|}dz = 2\pi\int_{z_n}^{z_{n+1}}\sqrt{|e^{2k|}|}\omega|dz. \tag{63}$$

The surface gravity (by definition, positive) in Weyl coordinates reads

$$\kappa_n = |\Omega_n|\sqrt{|e^{-2k}|}. \tag{64}$$

Thus the integral (63) is simply proportional to the rod's length:

$$\frac{\kappa_n}{2\pi}\mathcal{A}_n = z_{n+1} - z_n = L_n. \tag{65}$$

This is true for both timelike rods (horizons) and spacelike ones (defects). In the first case, the surface gravity is proportional to the Hawking temperature, and the area of the horizon is proportional to the entropy.

$$T_H = \kappa_H/2\pi, \qquad S_H = \mathcal{A}_H/4, \tag{66}$$

therefore

$$T_H S_H = \frac{\kappa_H}{8\pi}\mathcal{A}_H = \frac{L_H}{4}, \tag{67}$$

and the Smarr formula can be rewritten as[4]

$$M_H = 2\Omega_H J_H + 2T_H S_H + \Phi_H Q_H. \tag{68}$$

For spacelike rods such an identification does not seem justified, so we prefer to leave Smarr formua in its "length" form (62).

For multicenter solutions this mass relation holds separately for each black hole constituent. As stated in,[24,25] this is also true when black holes have magnetic and/or NUT charges. In our interpretation, the contribution of these parameters to the total asymptotic mass comes from individual rods representing the Dirac and Misner strings.

6. EMDA rotating dyon with NUT

The metric of the EMDA dyon with the NUT parameter constructed in 1994[31] can be represented in the Kerr form:

$$ds^2 = -\frac{\Delta - a^2\sin^2\theta}{\Sigma}\left(dt - \omega d\varphi\right)^2 + \Sigma\left(\frac{dr^2}{\Delta} + d\theta^2 + \frac{\Delta\sin^2\theta}{\Delta - a^2\sin^2\theta}d\varphi^2\right), \tag{69}$$

with the modified coefficient functions:

$$\begin{aligned}
&\Delta = (r - r_-)(r - 2M) + a^2 - (N - N_-)^2,\\
&\Sigma = r(r - r_-) + (a\cos\theta + N)^2 - N_-^2,\\
&\omega = \frac{-2w}{\Delta - a^2\sin^2\theta}, \quad w = N\Delta\cos\theta + a\sin^2\theta\left(M(r - r_-) + N(N - N_-)\right).
\end{aligned}$$

Two new parameters r_-, N_- depend on mass M, NUT parameter N, electric and magnetic charges Q, P as follows:

$$r_- = \frac{M(Q^2 + P^2)}{M^2 + N^2}, \qquad N_- = \frac{N(Q^2 + P^2)}{2(M^2 + N^2)}, \tag{70}$$

and a is the rotation parameter. The electric and magnetic potentials and the complex axidilaton scalar (2) depend also on the asymptotic values of dilaton and

axion $\zeta_\infty = \kappa_\infty + i\mathrm{e}^{-2\phi_\infty}$:

$$v = -\frac{\sqrt{2}e^{\phi_\infty}}{\Sigma}\mathrm{Re}\left[\mathcal{Q}(r - r_- + i\delta)\right], \qquad u = -\frac{\sqrt{2}\mathrm{e}^{\phi_\infty}}{\Sigma}\mathrm{Re}\left[\mathcal{Q}\zeta_\infty(r - r_- + i\delta)\right]$$
$$\zeta = \frac{\zeta_\infty R + \mathcal{D}\zeta^*_\infty}{R + \mathcal{D}}, \quad R = r - \frac{\mathcal{M}^* r_-}{2M} + i\delta, \quad \delta = a\cos\theta + N - N_-,$$

where the complex mass, electromagnetic and axidilaton charges are introduced

$$\mathcal{M} = M + iN, \qquad \mathcal{Q} = Q + iP, \qquad \mathcal{D} \equiv D + iA = -\frac{\mathcal{Q}^{*2}}{2\mathcal{M}}. \tag{71}$$

The explicit expressions for the axion and dilaton fields are

$$\kappa = \kappa_\infty - e^{-2\phi_\infty}\frac{2\mathrm{Im}(R\mathcal{D}^*)}{|R + \mathcal{D}|^2}, \qquad e^{-2(\phi - \phi_\infty)} = \frac{\Sigma}{|R + \mathcal{D}|^2}, \tag{72}$$

where $|R|^2 - |\mathcal{D}|^2 = \Sigma$. The dilaton asymptotics renormalizes the electric and magnetic charges, and the axion determines their mixing. Indeed, for $r \to \infty$, the electric and magnetic potentials are read as follows:

$$v \sim -\frac{\sqrt{2}\,Q\,\mathrm{e}^{\phi_\infty}}{r}, \qquad u \sim \frac{\sqrt{2}\,\left(P\,\mathrm{e}^{-\phi_\infty} - \kappa_\infty Q\mathrm{e}^{\phi_\infty}\right)}{r}. \tag{73}$$

The axion and dilaton asymptotics $r \to \infty$ read as follows:

$$\zeta \sim \zeta_\infty - 2ie^{-2\phi_\infty}\frac{\mathcal{D}}{r}, \qquad \kappa \sim \kappa_\infty + 2e^{-2\phi_\infty}\frac{A}{r}, \qquad \phi \sim \phi_\infty + \frac{D}{r}. \tag{74}$$

Combining electric and magnetic potentials, the complex potential Ψ reads

$$\Psi = \frac{\sqrt{2}\mathrm{e}^{\phi_\infty}}{\Sigma}\left[(\zeta - \kappa_\infty)\mathrm{Re}\left[\mathcal{Q}(r - r_- + i\delta)\right] + \mathrm{e}^{-2\phi_\infty}\mathrm{Im}\left[\mathcal{Q}(r - r_- + i\delta)\right]\right] \tag{75}$$

Applying the inverse dualization equations (10), we find the azimuthal component of the Maxwell four-potential For the four-potential one-form $\mathrm{A} = A_\mu dx^\mu$ we obtain:

$$A_\mu dx^\mu = \frac{\mathrm{e}^{\phi_\infty}}{\Sigma}\left(-Q(r - r_-) + P\delta\right)\left(dt + \left(2Ny - a(1 - y^2)\right)d\varphi\right) - \mathrm{e}^{\phi_\infty}Pyd\varphi. \tag{76}$$

Its norm diverges on the Misner strings $\cos\theta = \pm 1$:

$$A_\mu A^\mu\big|_{\cos\theta \to \pm 1} \sim \frac{-\mathrm{e}^{2\phi_\infty}P^2}{\Sigma\sin^2\theta}, \tag{77}$$

thus, Misner rods are also Dirac string loci.

Gravitational dualization equation (11) gives the metric twist potential

$$\chi = 2\frac{M\delta - N(r - r_-)}{\Sigma}. \tag{78}$$

The ergosphere boundary is given by

$$r_E = M + r_-/2 + \sqrt{|\mathcal{M}|^2\left(1 - r_-/2M\right)^2 - a^2\cos^2\theta}, \tag{79}$$

and the event horizon radius (the outermost root of $\Delta = 0$) is equal to

$$r_H = M + r_-/2 + \sqrt{|\mathcal{M}|^2 \left(1 - r_-/2M\right)^2 - a^2}, \tag{80}$$

so that the horizon two-surface touches the ergosphere at the polar axis.

The transformation to the Weyl coordinates ρ, z is given by

$$\rho^2 = \Delta \sin^2\theta,$$
$$z = (\Delta'/2)\,\cos\theta = (r - M - r_-/2)\cos\theta.$$

Solving the equation $\rho = 0$, we find that the rod system consists of three rods:

- l_+ : northern Misner string $r \in (r_H, \infty)$, $\cos\theta = 1$;
- l_H the horizon rod $r = r_H$, $\theta \in [0, \pi]$;
- l_+ southern Misner string $r \in (r_H, \infty)$, $\cos\theta = -1$.

The normalized directional vectors are given by the formulas (25), (27) in terms of the corresponding surface gravities and angular velocities.

It is also useful to represent the solution in terms of prolate spheroidal coordinates x, y definied as

$$r = \sigma x + M + r_-/2, \qquad \cos\theta = y, \tag{81}$$

$$\sigma^2 = |\mathcal{M}|^2 - |\mathcal{D}|^2 - |\mathcal{Q}|^2 - a^2, \tag{82}$$

One obtains:

$$ds^2 = -f\left(dt - \omega d\varphi\right)^2 + f^{-1}\left[\sigma^2 e^{2k}(x^2 - y^2)\left(\frac{dx^2}{x^2-1} + \frac{dy^2}{1-y^2}\right) + \rho^2 d\varphi^2\right], \tag{83}$$

where the metric functions now read

$$\Sigma = (\sigma x + M)^2 - |\mathcal{D}|^2 + (ay + N)^2,$$
$$\Sigma f = \sigma^2(x^2-1) - a^2(1-y^2), \quad e^{2k} = \frac{f\Sigma}{\sigma^2(x^2-y^2)},$$
$$f\Sigma\omega = -2\left[N\sigma^2(x^2-1)y + a(1-y^2)\left(M\sigma x + |\mathcal{M}|^2 - |\mathcal{Q}|^2/2\right)\right].$$

The South segment of the Misner string is $x > 1$, $y = -1$, the horizon rod is $x = 1, -1 \le y \le 1$ and the North Misner string is $x > 1$, $y = 1$. As expected, the inverse rotation function on the rods is constant on them and represents their angular velocities. It is easy to find

$$\Omega_H = \frac{2a}{\nu^2}, \qquad \Omega_\pm = \mp\frac{1}{2N}. \tag{84}$$

where a new parameter is introduced

$$\nu^2 = 2M\sigma + 2|\mathcal{M}|^2 - |\mathcal{Q}|^2, \tag{85}$$

such that the boundary values of $\Sigma(x, y)$ are

$$\Sigma_\pm \equiv \Sigma(1, \pm 1) = \nu^2 \pm 2aN. \tag{86}$$

The other ingredients of the mass formulas depend on the values of electric and magnetic potentials $v_\pm$ and $u_\pm$ of the defect rods $l_\pm$. The magnetic potential in spheroidal coordinates is

$$u = \frac{\sqrt{2}}{\Sigma}\Big\{\mathrm{e}^{\phi_\infty}\kappa_\infty[-Q(\sigma x+M-r_-/2)+P\delta]+e^{-\phi_\infty}[P(\sigma x+M-r_-/2)+Q\delta]\Big\} \quad (87)$$

Using it, we obtain:

$$v_\pm = -\frac{\sqrt{2}e^{\phi_\infty}}{\nu^2 \pm 2aN}\mathrm{Re}\left[\mathcal{Q}(\sigma \pm ia + \mathcal{M}(1-|\mathcal{Q}|^2/2|\mathcal{M}|^2))\right], \quad (88)$$

$$u_\pm = -\frac{\sqrt{2}\mathrm{e}^{\phi_\infty}}{\nu^2 \pm 2aN}\mathrm{Re}\left[\mathcal{Q}\zeta(\sigma \pm ia + \mathcal{M}(1-|\mathcal{Q}|^2/2|\mathcal{M}|^2))\right] \quad (89)$$

For surface gravities it is convenient to use Eq. (64). One finds:

$$\varkappa_H = \frac{\sigma}{\nu^2}, \qquad \varkappa_n = \mp\frac{1}{2N}. \quad (90)$$

Note that the angular velocities and surface gravities of Misner strings coincide similarly to the case of Kerr-Newman-NUT.[25] But they enter into the mass formula in a different way, so on should not consider this coincidence as a matter of interpretation.

The electric charges of the rods are given by (57). One finds the Gauss charges of the horizon and strings as:

$$Q_H = \frac{\omega_H}{2\sqrt{2}}(u_+ - u_-) = \frac{e^{-\phi\infty}\nu^2\left(2\sigma N^2 + M\nu^2\right)\mathrm{Im}\left[\zeta_\infty e^{2\phi_\infty}\mathcal{M}\mathcal{Q}\right]}{|\mathcal{M}|^2\left(\nu^4 - 4a^2N^2\right)}, \quad (91)$$

$$Q_\pm = \frac{Nu_\pm}{\sqrt{2}}. \quad (92)$$

Combining them together, we obtain

$$Q_H + Q_+ + Q_- = Qe^{-\phi_\infty} + P\kappa_\infty e^{\phi_\infty}. \quad (93)$$

This coincides with the true asymptotic charge of the configuration, with account of axion mixing and dilaton renormalization.

The electric potential on the horizon and on the strings are

$$\Phi_H = e^{\phi\infty}\frac{\nu^2(QM-NP)+2\sigma N(QN+MP)}{2\nu^2|\mathcal{M}|^2}, \qquad \Phi_\pm = -e^{\phi_\infty}\frac{P}{2N}. \quad (94)$$

Rod masses can be expressed in the form (58), using the corresponding values of the twist potential (78). One obtains

$$M_H = \nu^2\cdot\frac{M\nu^2+2\sigma N^2}{\nu^4-4a^2N^2} + \frac{1}{4}(2N(v_+u_+ + v_-u_-) - \sqrt{2}\mathrm{e}^{\phi_\infty}P(u_+ + u_-)), \quad (95)$$

$$M_\pm = N\frac{-N\sigma \pm Ma}{\nu^2 \pm 2aN} + \frac{1}{4}(2Nv_\pm - \sqrt{2}P\mathrm{e}^{\phi_\infty})u_\pm, \quad (96)$$

where the boundary values of electric and magnetic potentials are given by (88), (89). One can distinguish Dirac string contributions to the masses of defects as the parts non-vanishing for $N = 0$. Other terms are due to Misner strings. Direct substitution of (88-89) leads to long expressions, but the one can show that the Maxwell contributions to the total mass cancel out, while the remaining part non-trivially leads to the identity

$$M_\infty = M_H + M_+ + M_- = M.$$

Now we turn to the angular momentum balance, using the rod's momenta the expression (59). The length of the horizon rod is $L_H = 2\sigma$. Other rod lengths diverge. As it was shown in Ref.,[25] to overcome this issue, one can regularize the angular momentum as follows:

$$\tilde{J}_\pm = J_\pm + \frac{\omega_\pm L_\pm}{4} \tag{97}$$

We then obtain the angular momenta of the three rods as

$$J_H = \frac{\omega_H}{2}\left(-\sigma + M_H - Q_H\Phi_H\right), \qquad \tilde{J}_\pm = N^2\left(\frac{\pm N\sigma - Ma}{\nu^2 \pm 2aN} \mp \frac{v_\pm u_\pm}{2}\right). \tag{98}$$

We have explicitly checked that

$$\tilde{J}_\infty = J_H + \tilde{J}_+ + \tilde{J}_- = aM.$$

Therefore, we have constructed horizon mass, angular momentum and charge in terms of the global parameters and checked their total balance.

7. Conclusion

We have shown that the derivation of mass formulas using the Tomimatsu approach for calculating Komar integrals around rods in electrovacuum can be generalized to the EMDA theory containing scalar fields coupled to the graviphoton. Using a three-dimensional reduction of the EMDA equations, we construct magnetic and twist potentials, which allow to express the integrals over the rods in the same way as in the Einstein-Maxwell theory, where Tomimatsu used for this purpose the Ernst potentials. It should be noted that the scalar dilaton and axion fields do not contribute to the resulting mass formulas, and their asymptotics only rotates electric and magnetic charges in the parameter space. For regular asymptotically flat configurations, the dilaton and axion charges are secondary and their variations do not enter the Smarr-type formulas.

In this approach, the magnetic and NUT charges do not affect the mass and angular momentum of the black hole calculated as Komar's integrals over the horizon. They contribute to the asymptotic mass and angular momentum via the respective individual rods representing the Dirac and Misner strings. As an illustration, we considered the case of a rotating EMDA dyon equipped with NUT. In this case, as in the case of dyonic Kerr-Neumann-NUT, the Dirac string contributes to the

asymptotic mass, but not to the asymptotic angular momentum, while the Misner string contributes to both.

The work is supported by the Russian Foundation for Basic Research on the project 20-52-18012Bulg-a, and the Scientific and Educational School of Moscow State University "Fundamental and Applied Space Research". I.B. is also grateful to the Foundation for the Advancement of Theoretical Physics and Mathematics "BASIS" for support.

References

1. D. Christodoulou, "Reversible and irreversible transforations in black hole physics," Phys. Rev. Lett. **25** (1970), 1596-1597.
2. S. W. Hawking, "Black holes in general relativity," Commun. Math. Phys. **25** (1972), 152-166.
3. D. Christodoulou and R. Ruffini, "Reversible transformations of a charged black hole," Phys. Rev. D **4** (1971), 3552-3555.
4. L. Smarr, "Mass formula for Kerr black holes," Phys. Rev. Lett. **30** (1973), 71-73 [erratum: Phys. Rev. Lett. **30** (1973), 521-521].
5. S. W. Hawking, "Black hole explosions," Nature **248** (1974), 30-31.
6. J. D. Bekenstein, "Black holes and the second law," Lett. Nuovo Cim. **4** (1972), 737-740.
7. J. M. Bardeen, B. Carter and S. W. Hawking, "The Four laws of black hole mechanics," Commun. Math. Phys. **31** (1973) 161.
8. S. W. Hawking, "Black Holes and Thermodynamics." Phys. Rev. D **13** (1976) 191.
9. G. W. Gibbons and S. W. Hawking. "Action Integrals and Partition Functions in Quantum Gravity," Phys. Rev. D **15** (1977) 2752.
10. B. Carter, "Black Hole Equilibrium States, Part 2: General Theory of Stationary Black Hole States", in Black Holes (Les Houches 1972) ed B and C. DeWitt, (Gordon Breach, New York 1973)) 125-214.
11. C. J. Hunter. "The Action of instantons with nut charge," Phys. Rev. D **59** (1999) 024009, [gr-qc/9807010].
12. S. W. Hawking and C. J. Hunter. "Gravitational entropy and global structure," Phys. Rev. D **59** (1999) 044025, [hep-th/9808085].
13. S. Carlip, "Entropy from conformal field theory at Killing horizons," Class. Quantum Grav. **16** (1999) 3327, [gr-qc/9906126].
14. R. B. Mann, "Misner string entropy," Phys. Rev. D **60** (1999) 104047, [hep-th/9903229].
15. R. A. Hennigar, D. Kubizňák and R. B. Mann, "Thermodynamics of Lorentzian Taub-NUT spacetimes," Phys. Rev. D **100** (2019) no. 6, 064055, [arXiv:1903.08668 [hep-th]].
16. R. Durka, "The first law of black hole thermodynamics for Taub-NUT spacetime," [arXiv:1908.04238 [gr-qc]].
17. S. Q. Wu and D. Wu, "Thermodynamical hairs of the four-dimensional Taub-Newman-Unti-Tamburino spacetimes," Phys. Rev. D **100** (2019) no. 10, 101501, [arXiv:1909.07776 [hep-th]].
18. Z. Chen and J. Jiang, "General Smarr relation and first law of a NUT dyonic black hole," Phys. Rev. D **100** (2019) no. 10, 104016, [arXiv:1910.10107 [hep-th]].
19. A. Ballon Bordo, F. Gray and D. Kubizňák, "Thermodynamics of Rotating NUTty Dyons," JHEP **05** (2020), 084, [arXiv:2003.02268 [hep-th]].

20. G. Clément, D. Gal'tsov and M. Guenouche, "Rehabilitating space-times with NUTs," Phys. Lett. B **750** (2015) 591, [arXiv:1508.07622 [hep-th]].
21. T. Harmark, "Stationary and axisymmetric solutions of higher-dimensional general relativity," Phys. Rev. D **70**(2004) 124002, [hep-th/0408141].
22. R. Emparan and H. S. Reall, "Generalized Weyl solutions," Phys. Rev. D **65** (2002), 084025, [arXiv:hep-th/0110258 [hep-th]].
23. W. Israel, "Line sources in general relativity," Phys. Rev. D **15** (1977) 935.
24. G. Clément and D. Gal'tsov, "On the Smarr formula for rotating dyonic black holes," Phys. Lett. B **773** (2017) 290, [arXiv:1707.01332 [gr-qc]].
25. G. Clément and D. Gal'tsov, "On the Smarr formulas for electrovac spacetimes with line singularities," Phys. Lett. B **802** (2020), 135270, [arXiv:1908.10617 [gr-qc]].
26. A. Tomimatsu, "On Gravitational Mass and Angular Momentum of Two Black Holes in Equilibrium," Prog. Theor. Phys. **70** (1983), 385.
27. A. Tomimatsu, "Equilibrium of Two Rotating Charged Black Holes and the Dirac String," Prog. Theor. Phys. **72** (1984), 73.
28. A. Sen, "Rotating charged black hole solution in heterotic string theory," Phys. Rev. Lett. **69** (1992), 1006-1009, [arXiv:hep-th/9204046 [hep-th]].
29. G. W. Gibbons, "Antigravitating Black Hole Solitons with Scalar Hair in N=4 Supergravity," Nucl. Phys. B **207** (1982), 337-349.
30. R. Kallosh, D. Kastor, T. Ortin and T. Torma, "Supersymmetry and stationary solutions in dilaton axion gravity," Phys. Rev. D **50** (1994), 6374-6384, [arXiv:hep-th/9406059 [hep-th]].
31. D. V. Galtsov and O. V. Kechkin, "Ehlers-Harrison type transformations in dilaton - axion gravity," Phys. Rev. D **50** (1994), 7394-7399, [arXiv:hep-th/9407155 [hep-th]].
32. A. Narang, S. Mohanty and A. Kumar, "Test of Kerr-Sen metric with black hole observations," [arXiv:2002.12786 [gr-qc]].
33. I. Banerjee, B. Mandal and S. SenGupta, "Implications of Einstein-Maxwell dilaton-axion gravity from the black hole continuum spectrum," Mon. Not. Roy. Astron. Soc. **500** (2020) no. 1, 481-492, [arXiv:2007.13980 [gr-qc]].
34. I. Banerjee, B. Mandal and S. SenGupta, "Signatures of Einstein-Maxwell dilaton-axion gravity from the observed jet power and the radiative efficiency," Phys. Rev. D **103** (2021) no. 4, 044046, [arXiv:2007.03947 [gr-qc]].
35. M. Rogatko, "Extrema of mass, first law of black hole mechanics and staticity theorem in Einstein-Maxwell axion dilaton gravity," Phys. Rev. D **58** (1998), 044011, [arXiv:hep-th/9807012 [hep-th]].
36. T. Ghosh and S. SenGupta, "Thermodynamics Of dilaton-axion black holes," Phys. Rev. D **78** (2008), 124005, [arXiv:0811.3679 [hep-th]].
37. D. V. Gal'tsov, "Integrable systems in stringy gravity," Phys. Rev. Lett. **74** (1995), 2863-2866, [arXiv:hep-th/9410217 [hep-th]].
38. D. V. Gal'tsov and O. V. Kechkin, "Matrix dilaton - axion for heterotic string in three-dimensions," Phys. Lett. B **361** (1995), 52-58, [arXiv:hep-th/9507164 [hep-th]].
39. D. V. Gal'tsov and S. A. Sharakin, "Matrix Ernst potentials for EMDA with multiple vector fields," Phys. Lett. B **399** (1997), 250-257, [arXiv:hep-th/9702039 [hep-th]].
40. A. Komar, "Covariant conservation laws in general relativity," Phys. Rev. **113** (1959), 934-936.
41. Y. Chen and E. Teo, "Rod-structure classification of gravitational instantons with U(1)xU(1) isometry," Nucl. Phys. B **838** (2010) 207, [arXiv:1004.2750 [gr-qc]].

Holography for rotating black holes in $f(T)$ gravity

Masoud Ghezelbash
Department of Physics and Engineering Physics,
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
E-mail: amg142@campus.usask.ca
www.usask.ca

We extend the black hole holography to the case of an asymptotically anti–de Sitter (AdS) rotating charged black holes in $f(T) = T + \alpha T^2$ gravity, where α is a constant. We find that the scalar wave radial equation at the near-horizon region implies the existence of the 2D conformal symmetries. We show that choosing proper central charges for the dual CFT, we produce exactly the macroscopic Bekenstein-Hawking entropy from the microscopic Cardy entropy for the dual CFT. These observations suggest that the rotating charged AdS black hole in $f(T)$ gravity is dual to a 2D CFT at finite temperatures.*

Keywords: $f(T)$ gravity, Black holes, Holography

1. Introduction

The idea of AdS/CFT correspondence was extended to the case of extremal rotating black holes, namely, the Kerr/CFT correspondence which was proposed by Guica et al.[6] The correspondence states that the physics of the extremal Kerr black holes, which are rotating with maximum angular velocity, can be described by a 2D CFT, living on the near-horizon region of the black holes. The correspondence was established by showing that one can microscopically reproduces the Bekenstein-Hawking entropy, using the CFT Cardy entropy formula. As one would expect, the Kerr/CFT correspondence is not only a peculiar property of extremal black holes but also non-extremal Kerr black holes. However, at the near-horizon region of the non-extremal Kerr black holes, one cannot indicate any conformal symmetries. In other words, the conformal symmetries are not the symmetries of the non-extremal Kerr black hole geometry (as they are for the case of the extremal Kerr black holes). However, it turns out that the "hidden" conformal symmetries can be revealed by looking at the solution space of the radial part of the Klein-Gordon equation, for a massless scalar probe in the near-horizon region of the Kerr black holes.[7] In this case, the radial equation, can be written as the $SL(2,R)_L \times SL(2,R)_R$ Casimir eigen-equation. Subsequently, the Kerr/CFT correspondence can be established by matching the microscopic CFT Cardy entropy to the macroscopic Bekenstein-Hawking entropy

*This presentation is entirely based on the published paper in *Phys. Rev.* **D 101**, 024020 (2020) by C. Bernard and M. Ghezelbash.

of the Kerr black holes with general angular momentum and mass parameters. The correspondence has been studied for several black holes solutions, for instance, in Refs.[8–12,12–18].

The usual theory of gravity, based on the Riemannian geometry, has been extended through several gravity theories. One of them is the teleparallel gravity (TG) theory, where the Ricci scalar R, is replaced by the teleparallel torsion scalar T. Moreover, the TG has been generalized to $f(T)$ gravity by replacing the torsion scalar T, with an arbitrary function of T, such as $f(T)$. In,[19] Awad et al. find an asymptotically rotating charged AdS black hole solution, in quadratic $f(T)$ gravity, where $f(T) = T + \alpha T^2$. A very natural question to be asked, is "Do we have any $f(T)$/Hidden CFT correspondence?", that we address in this article.

The outline of this presentation is organized as follows. In Sec. 2, we review the $f(T)$-Maxwell gravity theory as well as the rotating charged AdS black hole solutions, and its thermodynamics aspects. We also consider the massless Klein-Gordon wave equation, in the background of the rotating charged AdS black holes, in quadratic $f(T)$ gravity. In Sec. 3, we study the radial wave equation in the near-horizon region of the black holes, and rewrite it as the $SL(2,R)_L \times SL(2,R)_R$ Casimir equation. We also find the central charges of the dual CFT by matching the Cardy entropy for the dual CFT to the macroscopic Bekenstein-Hawking entropy. Therefore, we present evidence that the rotating charged AdS black holes in quadratic $f(T)$ gravity, can be considered holographically dual to the CFT. In the final section, we summarize our results and address some future works. We use the Planck units, in which $c = G = \hbar = k_B = 1$.

2. f(T) Rotating Black Holes

The basic variables in TG are tetrad fields ${e_a}^\mu$, where $a = (0,1,2,3)$ is the index of internal space and $\mu = (0,1,2,3)$ is the index of spacetime. The tetrad fields satisfy

$$ {e_a}^\mu {e^a}_\nu = \delta^\mu_\nu, \ {e_a}^\mu {e^b}_\mu = \delta^b_a. \tag{1} $$

The tetrad fields are related to the spacetime metric and its inverse

$$ g_{\mu\nu} = \eta_{ab} {e^a}_\mu {e^b}_\nu, \ g^{\mu\nu} = \eta^{ab} {e_a}^\mu {e_b}^\nu, \tag{2} $$

respectively, where $\eta_{ab} = \mathrm{diag}\,(-,+,+,+)$ is the metric of 4D Minkowski spacetime. Also, it can be shown that $e = \det(e_{a\mu}) = \sqrt{-g}$, where g is the determinant of the metric. In TG, we use the Weitzenbock connection

$$ {W^\alpha}_{\mu\nu} = {e_a}^\alpha \partial_\nu {e^a}_\mu = -{e^a}_\mu \partial_\nu {e_a}^\alpha, \tag{3} $$

to define the covariant derivative, by

$$ \nabla_\nu {e_a}^\mu = \partial_\nu {e_a}^\mu + {W^\mu}_{\rho\nu} {e_a}^\rho = 0. \tag{4} $$

The Weitzenbock connection is curvature-free, but it has a non vanishing torsion

$$ {T^\alpha}_{\mu\nu} = {W^\alpha}_{\nu\mu} - {W^\alpha}_{\mu\nu} = {e_i}^\alpha \left(\partial_\mu {e^i}_\nu - \partial_\nu {e^i}_\mu \right). \tag{5} $$

We define the torsion scalar by

$$T = T^{\alpha}{}_{\mu\nu} S_{\alpha}{}^{\mu\nu}, \tag{6}$$

where the superpotential tensor is

$$S_{\alpha}{}^{\mu\nu} = \frac{1}{2}\left(K^{\mu\nu}{}_{\alpha} + \delta^{\mu}_{\alpha} T^{\beta\nu}{}_{\beta} - \delta^{\nu}_{\alpha} T^{\beta\mu}{}_{\beta}\right). \tag{7}$$

We note that the Contortion tensor $K_{\alpha\mu\nu}$ is given by

$$K_{\alpha\mu\nu} = \frac{1}{2}\left(T_{\nu\alpha\mu} + T_{\alpha\mu\nu} - T_{\mu\alpha\nu}\right). \tag{8}$$

We consider a four-dimensional rotating charged AdS black hole solution in $f(T)$-Maxwell theory with a negative cosmological constant where

$$f(T) = T + \alpha T^2. \tag{9}$$

The dimensional negative parameter α is the coefficient of the quadratic term of the scalar torsion. The action of the $f(T)$-Maxwell theory in 4D, for an asymptotically AdS spacetimes, is given by

$$\mathcal{S} = \frac{1}{2\mathfrak{K}} \int d^4x\, |e| \left(f(T) - 2\Lambda - F \wedge {}^{*}F\right), \tag{10}$$

where $\Lambda = -3/l^2$ is the 4D cosmological constant, l is the length scale of AdS spacetime. The constant $\mathfrak{K}$ in (10) is related to the 4D Newton's gravitational constant G_4, by $\mathfrak{K} = 2\Omega_2 G_4$, where $\Omega_2 = 2\pi^{3/2}/\Gamma(3/2)$, is the volume of 2D unit sphere, and $\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}$.

In the action (10), $F = d\tilde{\Phi}$, where $\tilde{\Phi} = \tilde{\Phi}_{\mu} dx^{\mu}$ is the gauge potential one-form.

Varying action (10) with respect to the tetrad fields and the Maxwell potential Φ_{μ}, one finds the field equations for gravity

$$S_{\mu}{}^{\rho\nu} \partial_{\rho} T f''(T) + \left[e^{-1} e^{a}{}_{\mu} \partial_{\rho}\left(e e_{\alpha}{}^{\alpha} S_{\alpha}{}^{\rho\nu}\right) - T^{\alpha}{}_{\lambda\mu} S_{\alpha}{}^{\nu\lambda}\right] f'(T)$$

$$-\frac{\delta^{\nu}_{\mu}}{4}\left(f(T) + \frac{6}{l^2}\right) = -\frac{\mathfrak{K}}{2} \mathcal{T}_{\mathrm{em}}{}^{\nu}{}_{\mu}, \tag{11}$$

and the Maxwell's equations

$$\partial_{\nu}\left(\sqrt{-g} F^{\mu\nu}\right) = 0. \tag{12}$$

respectively.

In equation (11), $\mathcal{T}_{\mathrm{em}}{}^{\nu}{}_{\mu} = F_{\mu\alpha} F^{\nu\alpha} - 1/4 \delta_{\mu}{}^{\nu} F_{\alpha\beta} F^{\alpha\beta}$, is the energy-momentum tensor of the electromagnetic field. The rotating charged AdS black hole solution, is given by[19]

$$ds^2 = -A(r)(\Xi dt - \Omega d\phi)^2 + \frac{dr^2}{B(r)} + \frac{r^2}{l^4}\left(\Omega dt - \Xi l^2 d\phi\right)^2 + \frac{r^2}{l^2} dz^2, \tag{13}$$

where the range of coordinates are given by $-\infty < t, z < \infty$, $0 \leq r < \infty$ and $0 \leq \phi < 2\pi$.

In metric (13), we have

$$A(r) = r^2\Lambda_{eff} - \frac{M}{r} + \frac{3Q^2}{2r^2} + \frac{2Q^3\sqrt{6\,|\alpha|}}{6r^4}, \tag{14}$$

$$B(r) = A(r)\beta(r), \tag{15}$$

$$\beta(r) = \left(1 + Q\sqrt{6\,|\alpha|}/r^2\right)^{-2}, \tag{16}$$

$$\Xi = \sqrt{1 + \frac{\Omega^2}{l^2}}, \tag{17}$$

where $\Lambda_{eff} = \frac{1}{36|\alpha|}$, and M, Q, and Ω are the mass parameter, the charge parameter, and the rotation parameter, respectively. The parameter α cannot be zero, since the effective cosmological constant Λ_{eff}, and the metric functions $A(r)$ and $B(r)$ become singular. The gauge potential one-form $\tilde{\Phi}$ is given by

$$\tilde{\Phi}(r) = -\Phi(r)\left(\Omega d\phi - \Xi dt\right). \tag{18}$$

where $\Phi(r) = \frac{Q}{r} + \frac{Q^2\sqrt{6|\alpha|}}{3r^3}$. We note that the torsion scalar T, for the black hole solution (13), is given by

$$T(r) = \frac{4A'(r)B(r)}{rA(r)} + \frac{2B(r)}{r^2}, \tag{19}$$

where $A(r)$ and $B(r)$, in equations (14) and (15).

We notice that setting the rotational parameter $\Omega = 0$, we find the static charged black hole configuration, as in Ref.[20]. Moreover, turning off the mass parameter M and Q, the metric (13) reduces to, the 4D AdS metric in an uncommon coordinate system.

The horizons of the black holes are the positive roots of $A = 0$, among which, the outer one is denoted by r_+. The non-vanishing components of the contravariant metric tensor are given by

$$g^{tt} = \frac{l^4\left(A(r)\Omega^2 - r^2\Xi^2\right)}{A(r)r^2\left(\Xi^2l^2 - \Omega^2\right)^2},\ g^{rr} = B(r),\ g^{zz} = \frac{l^2}{r^2},$$

$$g^{\phi\phi} = \frac{A(r)\Xi^2l^4 - r^2\Omega^2}{A(r)r^2\left(\Xi^2l^2 - \Omega^2\right)^2},\ g^{t\phi} = g^{\phi t} = \frac{\Xi\Omega l^2\left(A(r)l^2 - r^2\right)}{A(r)r^2\left(\Xi^2l^2 - \Omega^2\right)^2}, \tag{20}$$

and the determinant of the metric is

$$\sqrt{-g} = \sqrt{\frac{A(r)}{B(r)}}\frac{r^2\left(\Xi^2l^2 - \Omega^2\right)}{l^3}. \tag{21}$$

We find the area of the outer horizon $\mathcal{A}$, by setting $dt = dr = 0$ in the metric (13), and find

$$\mathcal{A} = \int_0^{2\pi} d\phi \int_0^L dz\sqrt{-g|_{dt=dr=0}} = \frac{2\pi r_+^2\Xi L}{l}. \tag{22}$$

We discuss now the first law of black hole thermodynamics in $f(T)$ gravity. Generally, the first law of black hole thermodynamics

$$\delta\mathcal{Q} = \tau\delta S, \tag{23}$$

where $\delta\mathcal{Q}$ and δS are the heat flux and the entropy change, respectively, is violated in $f(T)$ gravity. The Hawking temperature $\tau = \kappa/2\pi$ in $f(T)$ gravity, where κ is the surface gravity, is the same as one in the Einstein gravity, since it is independent of dynamics of gravity. The black hole solutions in $f(T)$ gravity, violate the Clausius relation $dS = d\mathcal{Q}/\tau$, which suggest that black hols in $f(T)$ gravity, even in a static spacetime, are in non-equilibrium state and produce an intrinsic entropy production.[21]

The heat flux $\delta\mathcal{Q}$ along a Killing vector ξ^μ, is given by

$$\delta\mathcal{Q} = \frac{\kappa}{2\pi}\left(\frac{f'(T)\,d\mathcal{A}}{4}\right)\Big|_0^{d\lambda} + \frac{1}{8\pi}\int\limits_H k^\nu\partial^\mu f'(T)\left(\xi^\rho S_{\rho\nu\mu} - \partial_\nu\xi_\mu\right)d\mathcal{A}d\lambda, \tag{24}$$

where H is the black hole horizon, λ is the affine parameter, $k^\mu = dx^\mu/d\lambda$ is the tangent vector to H, and κ is the surface gravity of the surface H. The first term in the right-hand side of equation (24) provides the first law of black hole thermodynamics Ref.[21]. However, the second term, in general, is not equal to zero. This term maybe regarded as a contribution to the intrinsic entropy production δS_i, where

$$\frac{1}{8\pi}\int k^\nu\partial^\mu f'(T)\left(\xi^\rho S_{\rho\nu\mu} - \partial_\nu\xi_\mu\right)d\mathcal{A}d\lambda = -\tau\delta S_i. \tag{25}$$

The equation (25) suggests that the $f(T)$ black holes are in non-equilibrium thermodynamics, where

$$\delta\mathcal{Q} = \tau\delta S - \tau\delta S_i. \tag{26}$$

Miao et al.[21] showed that the first law of thermodynamics for the $f(T)$black holes can be recovered approximatively, if $f''(T) \ll 1$. In this approximation, the intrinsic entropy production term in equation (24) can be neglected, and the entropy of black holes in $f(T)$ gravity becomes

$$S = \frac{f'(T)\,\mathcal{A}}{4}. \tag{27}$$

We note that in the case of $f'(T) = 1$, the entropy production δS_i vanishes, and the entropy (27) reduces to the Bekenstein-Hawking entropy in Einstein gravity. Using equations (9) and (22) in (27), we find the entropy of black holes in equation (13), as

$$S = \frac{\pi\Xi L\left(7{r_+}^6 + 9\sqrt{6\,|\alpha|}Q{r_+}^4 + 18M\,|\alpha|\,{r_+}^3 - 54Q^2\,|\alpha|\,{r_+}^2 - 42\sqrt{6|\alpha|^3}Q^3\right)}{9l\left(Q\sqrt{6\,|\alpha|} + {r_+}^2\right)^2}. \tag{28}$$

We consider a massless scalar field ψ, in the background of the rotating charged AdS black holes (13), in quadratic $f(T)$ gravity. The scalar wave equation is given by[22]

$$\frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\psi\right)=0. \tag{29}$$

We consider the following ansatz for the scalar field

$$\psi\left(t,r,z,\phi\right)=e^{-i\omega t+ikz+im\phi}R\left(r\right), \tag{30}$$

where ω is the frequency of the scalar field, m is the azimuthal harmonic index, and k is the wave number. Substituting equations (20), (21), and (30) into equation (29), we find the radial equation

$$B(r)\frac{d^2R\left(r\right)}{dr^2}+\left(rB(r)\frac{dA(r)}{dr}+rA(r)\frac{dB(r)}{dr}+4A(r)B(r)\right)\frac{dR(r)}{dr}+V(r)R\left(r\right)=0, \tag{31}$$

where the potential $V(r)$, is given by

$$V(r)=\frac{r^2\left(\Xi l^2\omega-\Omega m\right)^2-A(r)l^2\left\{k^2l^4\Xi^4+k^2\Omega^4+l^2\left[m^2\Xi^2-2m\Xi\Omega\omega+\Omega^2\left(\omega^2-2\Xi^2k^2\right)\right]\right\}}{A(r)r^2\left(\Xi^2l^2-\Omega^2\right)^2}. \tag{32}$$

In the near-horizon region, we expand the metric function $A(r)$ as a quadratic polynomial in $(r-r_+)$, such as

$$A(r)\simeq K\left(r-r_+\right)\left(r-r_*\right), \tag{33}$$

where

$$K=15{r_+}^4\Lambda_{eff}-3Mr_++\frac{3Q^2}{2}, \tag{34}$$

$$r_*=r_+-\frac{2r_+\left(2{r_+}^4\Lambda_{eff}-Mr_++Q^2\right)}{10{r_+}^4\Lambda_{eff}-2Mr_++Q^2}. \tag{35}$$

We note that r_* is not necessarily any of the black hole horizons. In the near-horizon region, we consider the low-energy limit for the scalar fields, where $r_+\ll\frac{1}{\omega}$. Moreover, we consider a limit where the outer horizon r_+ is very close to r_*, in which, $|r_+-r_*|\ll r_+$. Using these two approximations, we find that the radial equation (31) simplifies to

$$\frac{d}{dr}\{\left(r-r_+\right)\left(r-r_*\right)\frac{d}{dr}R\left(r\right)\}+\left[\left(\frac{r_+-r_*}{r-r_+}\right)\mathcal{A}+\left(\frac{r_+-r_*}{r-r_*}\right)\mathcal{B}+\mathcal{C}\right]R\left(r\right)=0, \tag{36}$$

where the constants $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are given by

$$\mathcal{A}=\frac{\mathcal{D}m^2+\mathcal{E}m\omega}{K^2{r_+}^2{r_*}^3\left(\Xi^2l^2-\Omega^2\right)^2\left(r_+-r_*\right)^2\beta}+\frac{\mathcal{F}\omega^2}{K{r_+}^2\left(\Xi^2l^2-\Omega^2\right)^2\left(r_+-r_*\right)^2\beta}-C_1, \tag{37}$$

$$\mathcal{B}=\frac{\mathcal{G}m^2+\mathcal{I}m\omega}{K^2{r_+}^3r_*\left(\Xi^2l^2-\Omega^2\right)^2\left(r_+-r_*\right)^2\beta}+\frac{\mathcal{J}\omega^2}{K{r_+}^2\left(\Xi^2l^2-\Omega^2\right)^2\left(r_+-r_*\right)^2\beta}+C_2, \tag{38}$$

$$\mathcal{C}=-\frac{2m\Omega\left(\Xi l^2\omega-\Omega m/2\right)^2\left({r_+}^2+r_+r_*+{r_*}^2\right)}{\left(\Xi^2l^2-\Omega^2\right)^2K^2{r_+}^3{r_*}^3\beta}. \tag{39}$$

In equations (37)-(39), the constants C_1 and C_2 are given by $C_1 = C_2 = k^2l^2/K\beta(r_+-r_*)^2{r_+}^2$, and

$$\mathcal{D}=\Omega^2\left({r_+}^3+2{r_+}^2r_*+3{r_*}^2r_+\right)-l^4\Xi^2{r_*}^3K, \tag{40}$$

$$\mathcal{E}=2\Omega\Xi l^2\left(Kl^2{r_*}^3-{r_+}^3-2{r_+}^2r_*-3r_+{r_*}^2\right), \tag{41}$$

$$\mathcal{F}=-l^4\Omega^2, \tag{42}$$

$$\mathcal{G}=K\Xi^2l^4r_+r_*-\Omega^2\left(3{r_+}^2+2r_+r_*+{r_*}^2\right), \tag{43}$$

$$\mathcal{I}=2l^2\Xi\Omega\left[3{r_+}^2-r_+r_*\left(Kl^2-2\right)+{r_*}^2\right], \tag{44}$$

$$\mathcal{J}=l^4\Omega^2. \tag{45}$$

3. Holography for $f(T)$ Rotating Black Holes

To find the existence of the possible hidden symmetry, we introduce the following conformal coordinates ω^+, ω^- and y, in terms of the black hole coordinates t, r and ϕ

$$\omega^+=\sqrt{\frac{r-r_+}{r-r_*}}e^{2\pi T_R\phi+2n_Rt}, \tag{46}$$

$$\omega^-=\sqrt{\frac{r-r_+}{r-r_*}}e^{2\pi T_L\phi+2n_Lt}, \tag{47}$$

$$y=\sqrt{\frac{r_+-r_*}{r-r_*}}e^{\pi(T_L+T_R)\phi+(n_L+n_R)t}, \tag{48}$$

where T_L, T_R, n_L and n_R are constants. We also define the sets of *local* vector fields

$$H_1=i\partial_+, \tag{49}$$

$$H_0=i\left(\omega^+\partial_++\frac{1}{2}y\partial_y\right), \tag{50}$$

$$H_{-1}=i(\omega^{+2}\partial_++\omega^+y\partial_y-y^2\partial_-), \tag{51}$$

as well as

$$\bar{H}_1 = i\partial_-, \tag{52}$$

$$\bar{H}_0 = i\left(\omega^-\partial_- + \frac{1}{2}y\partial_y\right), \tag{53}$$

$$\bar{H}_{-1} = i(\omega^{-2}\partial_- + \omega^- y\partial_y - y^2\partial_+). \tag{54}$$

The vector fields (49)-(54) obey the $sl(2,R)_L \times sl(2,R)_R$ algebra, as

$$[H_0, H_{\pm 1}] = \mp iH_{\pm 1}, \quad [H_{-1}, H_1] = -2iH_0, \tag{55}$$

$$[\bar{H}_0, \bar{H}_{\pm 1}] = \mp i\bar{H}_{\pm 1}, \quad [\bar{H}_{-1}, \bar{H}_1] = -2i\bar{H}_0. \tag{56}$$

The quadratic Casimir operators of $sl(2,R)_L \times sl(2,R)_R$ algebra, are given by

$$\mathcal{H}^2 = \bar{\mathcal{H}}^2 = -H_0^2 + \frac{1}{2}(H_1H_{-1} + H_{-1}H_1) = \frac{1}{4}(y^2\partial_y^2 - y\partial_y) + y^2\partial_+\partial_-. \tag{57}$$

We notice that the Casimir operators (57), can be rewritten in terms of (t, r, ϕ) coordinates, as

$$\begin{aligned}\mathcal{H}^2 \;=\; & (r-r_+)(r-r_*)\frac{\partial^2}{\partial r^2} + (2r - r_+ - r_*)\frac{\partial}{\partial r} \\ & + \left(\frac{r_+ - r_*}{r - r_*}\right)\left[\left(\frac{n_L - n_R}{4\pi G}\partial_\phi - \frac{T_L - T_R}{4G}\partial_t\right)^2 + C_2\right] \\ & - \left(\frac{r_+ - r_*}{r - r_+}\right)\left[\left(\frac{n_L + n_R}{4\pi G}\partial_\phi - \frac{T_L + T_R}{4G}\partial_t\right)^2 - C_1\right],\end{aligned} \tag{58}$$

where $G = n_LT_R - n_RT_L$.

The Casimir operator (58) reproduces the radial equation (36), by choosing the right and left temperatures, as

$$T_R = \frac{r_+K\left(r_+ - r_*\right)\left(\Xi^2l^2 - \Omega^2\right)\sqrt{\beta r_+r_*\delta}}{4\pi\delta}, \tag{59}$$

$$T_L = \frac{r_+K(\Xi^2l^2 - \Omega^2)[{r_+}^4 + 2{r_+}^3r_* + 6{r_+}^2{r_*}^2 - 2{r_*}^3r_+(Kl^2 - 1) + {r_*}^4]\sqrt{\beta r_+r_*\delta}}{4\pi\left(r_+ + r_*\right)^3\delta}, \tag{60}$$

respectively, and

$$n_R = 0, \tag{61}$$

$$n_L = \frac{{r_*}^2r_+K\left(\Omega^2 - \Xi^2l^2\right)\sqrt{\beta r_+r_*\delta}}{2\Omega l^2\Xi\left(r_+ + r_*\right)^3}. \tag{62}$$

The constant δ which appears in equations (59), (60), (61) and (62), is given by $\delta = 3\Omega^2{r_+}^2 - r_+r_*\left(K\Xi^2l^4 - 2\Omega^2\right) + \Omega^2{r_*}^2$. Moreover, we find two constraints for

the parameters of the black hole solutions (13), such as

$$\frac{\Omega^2\Xi^2 l^4\left[-3{r_+}^2+r_* r_+\left(Kl^2-2\right)-{r_*}^2\right]^2}{K^2\beta {r_+}^3 r_*(r_+-r_*)^2(\Xi^2 l^2-\Omega^2)^2}=\frac{l^4\Omega^2}{{r_+}^2K(\Xi^2 l^2-\Omega^2)^2(r_+-r_*)^2\beta}, \quad (63)$$

and

$$\frac{\Omega^2\Xi^2 l^4\left(Kl^2{r_*}^3-{r_+}^3-2r_* {r_+}^2-3{r_*}^2 r_+\right)^2}{K^2\beta r_+{r_*}^5(r_+-r_*)^2(\Omega^2-\Xi^2 l^2)^2\left(Kl^4\Xi^2 r_* r_+-3\Omega^2{r_+}^2-2\Omega^2 r_* r_+-\Omega^2{r_*}^2\right)}$$
$$=\frac{l^4\Omega^2}{{r_+}^2K(\Xi^2 l^2-\Omega^2)^2(r_+-r_*)^2\beta}. \quad (64)$$

We note that equations (63) and (64) restrict the black hole parameters, in accord with existence of the real positive values for the outer horizon (and any other horizons), as the roots of the tri-quadratic algebraic equation

$$A(r)=0, \quad (65)$$

where $A(r)$ is given in (14).

We note that both right and left temperatures (59) and (60) are positive definite.[23] We emphasize that $SL(2,R)_L\times SL(2,R)_R$ is a *local* hidden symmetry, for the solution space of massless scalar field in the near region of the rotating charged AdS black holes (13), in quadratic $f(T)$ gravity. The *local* hidden symmetry is generated by the vector fields (49)–(54), which are not periodic under the angular identification $\phi\sim\phi+2\pi$. These symmetries can't be used to generate new global solutions from the old ones. This can be interpreted as a statement that the $SL(2,R)_L\times SL(2,R)_R$ symmetry is spontaneously broken down to $U(1)_L\times U(1)_R$ subgroup under the angular identification $\phi\sim\phi+2\pi$. As a result, we can identify the left and right temperatures of the dual CFT. We recall the Cardy entropy formula for the dual 2D CFT with temperatures T_L and T_R

$$S_{CFT}=\frac{\pi^2}{3}\left(c_L T_L+c_R T_R\right), \quad (66)$$

where c_L are c_R are the corresponding central charges for the left and right sectors. The central charges can be derived from the *asymptotic symmetry group* of the near-horizon (near-)extremal black hole geometry. There is no derivation for the central charges of the CFT dual to the non-extremal black holes, that we consider in this article. Of course, we expect that the conformal symmetry of the extremal black holes connects smoothly to those of the non-extremal black holes, for which the central charges are the same. The near-horizon extremal geometry for the spacetime (13) is still unknown, and it is not a straightforward task to find that, due to the tri-quadratic behaviour of the metric function $A(r)$. As a result, we turn the logic around and consider the favourite holographic situation, in which, the Cardy entropy (66) produces exactly the macroscopic entropy (27). Substituting equations (28),

(59), and (60) into equation (66), we find the central charges

$$c \equiv c_L = c_R = \frac{12\Xi\delta L\varpi(r_+ + r_*)^3}{lK{r_+}^2(\Xi^2 l^2 - \Omega^2)({r_+}^3 + 2{r_+}^2 r_* + 3r_+ {r_*}^2 - l^2 K {r_*}^3)(Q\sqrt{6|\alpha|} + {r_+}^2)^2\sqrt{\beta r_+ r_* \delta}}, \tag{67}$$

where $\varpi = {r_+}^2\left({r_+}^2 Q\sqrt{6\,|\alpha|}/2 + 7{r_+}^4/18 + M\,|\alpha|\,r_+ - 3Q^2\,|\alpha|\right) - 7\sqrt{6|\alpha|^3}Q^3/3$. We note that we only consider CFTs, in which the left and right central charges are equal, $c \equiv c_L = c_R$.[24,25]

4. Conclusions

In this presentation which is based entirely on the published paper,[23] we extend the concept of black hole holography to the non-extremal 4D rotating charged AdS black holes in $f(T)$-Maxwell theory with a negative cosmological constant. We explicitly construct the hidden conformal symmetry for the rotating black holes in$f(T)$-Maxwell theory with a negative cosmological constant. We mainly consider the near-horizon region, as the metric function which determines the event horizon, is a triple-quadratic equation. In this region, we show that the radial equation of the scalar wave function could be written as the $SL(2,R)_L \times SL(2,R)_R$ squared Casimir equation, indicating a *local* hidden conformal symmetry acting on the solution space. The conformal symmetry is spontaneously broken under the angular identification $\phi \sim \phi + 2\pi$, which suggests the rotating charged AdS black holes in quadratic $f(T)$ gravity, should be dual to the finite temperatures T_L and T_R mixed state, in the 2D CFT. Instead of calculating the central charges using the *asymptotic symmetry group*, we calculated the central charges by assuming the Cardy entropy for the dual CFT, matches the macroscopic Bekenstein-Hawking entropy. These results suggest that rotating charged AdS black holes in quadratic $f(T)$ gravity with particular values of M, Ω, Q, and $|\alpha|$, are dual to a 2D CFT.

It is an open question to find the near-horizon (near-)extremal geometry of the rotating charged AdS spacetime in quadratic $f(T)$ gravity. We may calculate the central charges using the *asymptotic symmetry group* to confirm our results in this article. We can also study on various kinds of superradiant scattering off the near-extremal black hole as a further evidence to support the holographic picture for the non-extremal 4D rotating charged AdS black holes in $f(T)$-Maxwell theory with a negative cosmological constant. We leave addressing these open questions for future articles.

Acknowledgments

The author would like to acknowledge the financial support by the Natural Sciences and Engineering Research Council of Canada.

References

1. G. 't Hooft, *Conf. Proc.* **C930308**, 284 (1993).
2. L. Susskind, *J. Math. Phys.* **36**, 6377 (1995).
3. J. D. Bekenstein, *Phys. Rev. D* **7**, 2333 (1973).
4. S. W. Hawking, *Commun. Math. Phys.* **43**, 199 (1975).
5. J. M. Maldacena, *Int. J. Theor. Phys.* **38**, 1113 (1999), *Adv. Theor. Math. Phys.* **2**, 231(1998).
6. M. Guica, T. Hartman, W. Song and A. Strominger, *Phys. Rev. D* **80**, 124008 (2009).
7. A. Castro, A. Maloney and A. Strominger, *Phys. Rev. D* **82**, 024008 (2010).
8. A. M. Ghezelbash, *JHEP* **08**, 045 (2009).
9. A. M. Ghezelbash, *Mod. Phys. Lett. A* **27**, 1250046 (2012).
10. B. Chen and J. Long *JHEP* **08**, 065 (2010).
11. B. Chen, A. M. Ghezelbash, V. Kamali and M. R. Setare, *Nucl. Phys. B* **848**, 108 (2011).
12. A. M. Ghezelbash, V. Kamali and M. R. Setare, *Phys. Rev. D* **82**, 124051(2010).
13. A. M. Ghezelbash and H. M. Siahaan, *Class. Quant. Grav.* **30**, 135005 (2013).
14. Y.-Q. Wang and Y.-X, Liu, *JHEP* **08** 087 (2010).
15. B. Chen, J. Long, and J.-j. Zhang, *Phys. Rev. D* **82**, 104017 (2010).
16. A. M. Ghezelbash and H. M. Siahaan, *Gen. Rel. Grav.* **46**, 1783 (2014).
17. H. M. Siahaan, *Class. Quant. Grav.* **33**, 155013 (2016).
18. H. M. Siahaan, *Class. Quant. Grav.* **35**, 155002 (2018).
19. A. M. Awad, G. G. L. Nashed and W. El Hanafy, arXiv:1903.12471[gr-qc].
20. A. M. Awad, S. Capozziello, and G. G. L. Nashed, *JHEP* **07**, 136 (2017).
21. R.-X. Miao, M. Li and Y.-G. Miao, *JCAP* **1111**, 033 (2011).
22. C. Bernard, *Phys. Rev. D* **96**, 105025 (2017).
23. C. Bernard and M. Ghezelbash, *Phys. Rev. D* **101**, 024020 (2020).
24. S. El-Showk and K. Papadodimas, *JHEP* **10**, 106 (2012).
25. G. Compère, *Living Rev. Rel.* **15**, 11 (2012).

Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity

Semin Xavier
Department of Physics, Indian Institute of Technology, Mumbai, Maharashtra 400076
E-mail: seminxavier@iitb.ac.in

Jose Mathew
Department of Physics, The Cochin College, Kochi, Kerala, India
E-mail: josecherukara@gmail.com

S. Shankaranarayanan
Department of Physics, Indian Institute of Technology, Mumbai, Maharashtra 400076
E-mail: shanki@phy.iitb.ac.in

The evidence is mounting that the universe is currently undergoing a phase of accelerated expansion. One possible alternative is the modification in gravity in the largest possible scales. This leads to the many questions related to black-holes: violation of Birkhoff theorem and no-hair theorem. To confirm/infirm, we need to obtain exact black-hole solutions in these modified gravity theories. In this talk, we focus on the exact spherically symmetric solutions in $f(R)$ theories of gravity. We explicitly show that some $f(R)$ models contain an infinite number of exact static, Ricci-flat spherically symmetric vacuum solutions and, hence, violate Birkhoff's theorem in $f(R)$ theories. We analytically derive two exact vacuum black-hole solutions for the same class of $f(R)$ theories. The two black-hole solutions have the event-horizon at the same point; however, their asymptotic features are different. Our results suggest that the no-hair theorem may not hold for generic modified gravity theories. We discuss the implications to distinguish modified gravity theories from general relativity.

Keywords: $f(R)$ theory, no-hair theorem, based on the work[1]

1. Introduction

General Relativity (GR) is a successful description of gravity and is in excellent agreement with experiments and observations. However, paradoxically, the very success of GR suggests that GR might have classical and quantum corrections in the strong gravity regime.[2–6] Recently, strong gravity diagnostic parameters have been proposed to distinguish GR and modified gravity theories.[7]

In GR, static black-hole solutions are the final stage of gravitationally collapsed objects. As a consequence of the no-hair theorem,[8] GR predicts that the Kerr metric represents the most general astrophysical black holes.[9] In the case of spherical symmetry, Birkhoff's theorem ensures that the most general spherically symmetric, electrovac solution of the Einstein-Maxwell equations is the static Reissner-Nordstrom solution.[10,11] The absence of spin-0 modes in the linearized field equations leads to this result. However, there is no Birkhoff theorem for the Kerr metric. Outside

a rotating star, the metric is not described by Kerr. In comparison, the Schwarzschild solution describes the space-time outside a non-rotating star.

In this talk, we show that the Schwarzschild solution is not an unique solution for $f(R)$ theories of gravity. More specifically, we show that there exists an infinite number of spherically symmetric vacuum solutions in 4-D $f(R)$ theories of gravity *without* transforming to a conformal frame.

$f(R)$ theories of gravity are the most straightforward modifications to GR.[6] The higher-order Ricci scalar terms encapsulate modifications to GR. Although the equations of motion are higher-order, they do not suffer from Oströgradsky instability.[12] A deviation from general relativity may introduce changes in both the generation and the propagation of gravitational waves. Unlike GR, $f(R)$ theories have an extra field equation and longitudinal mode.[7, 13]

Although Schwarzschild black-hole is a solution to vacuum $f(R)$ theories of gravity. The additional dynamical equation by the trace equation in $f(R)$ theories may give rise to new solutions other than GR. The reason for such assumption is that, in $f(R)$ gravity has 11 dynamical variables — 10 metric variables ($g_{\mu\nu}$) and the Ricci scalar (R). In other words, in $f(R)$ theories, the scalar curvature R, plays a non-trivial role in determining the metric itself. Recently, the effects on the extra mode are used to obtain diagnostic parameters to distinguish GR and modified gravity theories using the Quasi-normal mode spectrum of the identical black-hole solutions in the two theories.[7] Here, the focus is to test the validity of this assumption. In other words, to verify whether the black-hole solutions in the two theories are indeed identical. We show that $f(R)$ admits multiple space-time geometries with the horizon for the same stress-tensor configuration (vacuum in this case).

We use $(-,+,+,+)$ signature for the 4-D space-time metric,[14] Greek alphabets for the 4-D space-time, and $\kappa^2 = 8\pi G/c^4$ where G is the Newton's constant. We shall denote the derivative of any function with r by an overprime, and overdot denotes the partial derivative w.r.t Ricci scalar (R).

2. $f(R)$ theory and the model

The modified vacuum action for $f(R)$ gravity is:[6]

$$S[g_{\mu\nu}] = \frac{1}{2\kappa^2}\int d^4x\sqrt{-g}\, f(R) \tag{1}$$

where $f(R)$ is arbitrary function of the Ricci scalar R. Using the metric formalism[6] the modified Einstein tensor ($\mathcal{G}_{\mu\nu}$) is obtained, i.e.,

$$\mathcal{G}_{\mu\nu} \equiv \dot{f}(R)\, R_{\mu\nu} - \nabla_\mu\nabla_\nu \dot{f}(R) + g_{\mu\nu}\Box\dot{f}(R) - \frac{f(R)}{2}g_{\mu\nu} = 0\,, \tag{2}$$

where $\dot{f}(R) = \partial f/\partial R$ and $\Box = \nabla^\mu\nabla_\mu$. For the above $f(R)$ action (1), the trace of the field equation (2) is dynamical:

$$R\,\dot{f}(R) + 3\,\Box\dot{f}(R) - 2\,f(R) = 0 \tag{3}$$

In $f(R)$, the scalar curvature R, plays a non-trivial role in determining the metric itself. In general, the field equations for $f(R)$ gravity (2) are fourth-order partial differential equations, instead of second-order partial differential equations of GR. Hence, we expect more class of solutions in $f(R)$ theories of gravity.

One may still find the trivial solution where for which $R = R_0 =$ constant. In such a case, the field equations reduce to the Einstein field equations with an effective cosmological constant and an effective gravitational constant.[15] This includes the case where $R = 0$. However, our interest in this talk is to look for non-trivial solutions.

In our model, we consider $f(R)$ to be a polynomial in R, i.e.,

$$f(R) = \beta_0 + \beta_1 R + \beta_2 R^2 + \cdots , \tag{4}$$

where β_i's $(i = 0, 1, 2 \cdots)$ are constants having appropriate dimensions. The above form of $f(R)$ can be written in a binomial form, i.e.,

$$f(R) = (\alpha_0 + \alpha_1 R)^p , \tag{5}$$

where, p is the power index, we are interested in the strong-gravity corrections to GR, we take $p > 1$. In principle, p need not be an integer. The α_0 and α_1 being positive constants related to the β_i's in (4). For simplicity consider $p = 1$, the above from (3) reduces to:

$$f(R) = \alpha_0 + \alpha_1 R . \tag{6}$$

Thus, α_0 acts like the cosmological constant and α_1 is a dimensionless constant that modifies Newton's constant. [Note that α_1 is dimensionless and α_0 has dimensions of $[L]^{-2}$.]

3. A class of exact solutions for $f(R)$ theory

Consider the static, spherically symmetric metric of the form,

$$ds^2 = -A(r)e^{\delta(r)}dt^2 + \frac{dr^2}{A(r)} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \tag{7}$$

where $A(r)$ and $\delta(r)$ are arbitrary functions of the Schwarzschild radial coordinate r. Substituting the above metric in the modified Einstein's equations (2) for the model $f(R)$ (5), leads to the following three field equations:

$$\mathcal{G}_t^t \equiv T_3[A(r), \delta(r)] = 0 \tag{8a}$$

$$\mathcal{G}_r^r \equiv T_4[A(r), \delta(r)] = 0 \tag{8b}$$

$$\mathcal{G}_\theta^\theta = \mathcal{G}_\phi^\phi \equiv T_5[A(r), \delta(r)] = 0 \tag{8c}$$

where T_3, T_4 and T_5 are functions of $A(r)$ and $\delta(r)$, and their derivatives. More specifically, (i) T_3 and T_5 are non-linear, and *contain up to 4th order derivatives* of $A(r)$ and $\delta(r)$, and (ii) T_4 is non-linear and *contain up to 3rd order derivatives* of

$A(r)$ and $\delta(r)$. (iii) Even in the special case of $\delta(r) = 0$, $\mathcal{G}_t^t \neq \mathcal{G}_r^r$. Hence, we do not expect to get identical solutions as in GR.

The field equations (8) contain up to fourth-order derivatives in $A(r)$ and $\delta(r)$. Thus, an exact solution to these equations will contain up to four independent constants.

Ref. 1 contains the details of the procedure used to reduce these highly non-trivial equations into a product of two second-order non-linear differentials in $A(r)$ and $\delta(r)$. Interestingly, both the procedures lead to the following equation:

$$\frac{2}{r}\,\frac{(p-\frac{1}{2})}{p\,(p-1)}\,\frac{T_1[A(r),\delta(r)]}{(\Phi(r)+4)}\,\frac{T_2[A(r),\delta(r)]}{(\Phi(r)-2)} = 0 \tag{9}$$

where $\Phi(r) = r\left(\delta^{'}(r) + [\ln A(r)]^{'}\right)$ and prime denotes derivative w.r.t r.

At this point, we would like to stress the following points: First, we have obtained the Eq. (9) using two different approaches. This implies that Eq. (9) is a unique differential equation for this $f(R)$ model with the metric (7). Second, Eq. (9) is a product of two second-order non-linear differentials of $T_1[A(r),\delta(r)]$ and $T_2[A(r),\delta(r)]$.[1] It simplifies the procedure to obtain the exact black-hole solutions for any value of p using $T_1[A(r),\delta(r)]$ or $T_2[A(r),\delta(r)]$. Third, the above equation imposes constrains on p, $A(r)$ and $\delta(r)$. More specifically, if for a non-trival solution,

$$p \neq 0, \frac{1}{2}, 1\,;\, \Phi(r) \neq -4 \text{ or } 2\,. \tag{10}$$

Since $p = 1$ is not allowed. It means we are looking for non-trivial solutions valid for modified theories of gravity, where $p = 1$ will give solutions of GR. While $p = 0$ and 1 will lead to divergence, $p = 1/2$ will lead to trivial solutions. Fourth, the above condition of $\Phi(r)$ implies that

$$\delta(r) + \ln A(r) \neq \ln(r^2) \text{ or } \ln(r^{-4}) \tag{11}$$

If we assume $\delta(r) =$ constant, then $A(r) \neq c_0 r^2 + c_1 r^{-4}$, where c_0, c_1 are constants. Lastly, non-trivial solutions for the above equation (9) are possible if T_1 or T_2 vanish, i.e.,

$$T_1[A(r),\delta(r)] = 0 \qquad \text{or} \qquad T_2[A(r),\delta(r)] = 0 \tag{12}$$

In principle, from Eq. (12) we can identity that for a given $A(r)$, we can have two functional forms of $\delta(r)$. Either $T_1 = 0$ or $T_2 = 0$ will give two choices of $\delta(r)$. While, the functional form of $A(r)$ can be obtained from the condition (10) on $\Phi(r)$, we get

$$\delta'(r) + (\ln[A(r)])^{'} = \mu(r) \quad \mu(r) \neq \frac{-4}{r} \text{ or } \frac{2}{r} \tag{13}$$

Thus, for a given $\mu(r)$, we have a functional relation between $A(r)$ and $\delta(r)$. Substituting this relation in the constraint relation (12), we obtain two set of differential equation in terms of $A(r)$ or $\delta(r)$.

Thus, Eq. (12) give two branches of solutions for $A(r)$ and $\delta(r)$ by setting $T_1 = 0$ or $T_2 = 0$. Using the arbitrary function $\mu(r)$, we can obtain an infinite solutions for the spherically symmetric metric (7).

At this point, it is important to validate whether any arbitrary function $\mu(r)$ satisfying $T_1 = 0$ or $T_2 = 0$ is indeed a solution to the vacuum field equations (8)?

To address this, for any p, we write a formal solution to the above equation Eq. (13) as

$$A(r) = e^{-\delta(r)}\gamma(r) \quad \text{where} \quad \gamma(r) = \exp\left(\int \mu(r)dr\right). \tag{14}$$

Substituting $A(r)$ in-terms of $\delta(r)$ in $T_2[A(r), \delta(r)] = 0$, we obtain a differential equation in $\delta''(r)$. Substituting these in Eq. (8), we get,

$$\begin{aligned} T_3[\delta(r), \gamma(r)] &= 4\alpha_1 p(p-1)(p-2)\gamma^7(r)e^{-\delta(r)}\left[\delta'(r)r - 8\right]^2 T_2[\delta(r), \gamma(r)]^2 \\ &= T_5[\delta(r), \gamma(r)] \end{aligned} \tag{15a}$$

$$T_4[\delta(r), \gamma(r)] = -2\alpha_1 p(p-1)\gamma^5(r)e^{-\delta(r)}\left[\delta'(r)r - 8\right]\left(r\ln[\gamma(r)]' + 4\right) T_2[\delta(r), \gamma(r)] \tag{15b}$$

Since, we have obtained the above expressions using the condition $T_2[A(r), \delta(r)] = 0$, we have $\mathcal{G}^t_t = \mathcal{G}^r_r = \mathcal{G}^\theta_\theta = \mathcal{G}^\phi_\phi = 0$. This implies that $A(r)$ given by Eq. (14) satisfying $T_2[A(r), \delta(r)] = 0$ is an exact solution for the $f(R)$ model (5). Since $A(r)$ depends on the arbitrary function $\mu(r)$, for the same observer with Schwarzschild time t, there are *infinite number of exact static, spherically symmetric solutions* for this model.

This is the key result of this talk. The Birkhoff theorem in GR guarantees that static Schwarzschild solution[11] is the most general spherically symmetric vacuum solution. However, the trace equation (3) provides a non-trivial dynamical structure in $f(R)$ theories. This provides an infinite set of static solutions for $f(R)$ theories of gravity. To our knowledge, such an explicit calculation is new for any modified theories of gravity. All the earlier analyses, either restrictive or use a conformal frame to confirm/infirm Birkhoff theorem.[16,17] Here, we have not made any approximation or performed a conformal transformation. Our results point that the Birkhoff theorem is not valid for all modified gravity theories.

As noted in Eq. (12), we can obtain non-trivial solutions if T_1 or T_2 vanish. Until now, we have shown that $T_2 = 0$ yields an infinitely many vacuum spherically symmetric solutions. Unlike T_2, T_1 is not a common factor of the field equations (15); hence, $T_1 = 0$ alone can not provide valid solutions. Thus, beside $T_1 = 0$, we must use either of the three equations (8) to obtain a unique solution.

4. Two vacuum black-hole solutions

In the earlier section, we showed that there exists an infinite number of spherically symmetric vacuum solutions for the $f(R)$ model (5). In this section, we obtain two black-hole solutions which satisfy the condition $T_2[A(r), \delta(r)] = 0$.

As we have mentioned earlier, $\delta(r) = 0$ is a trivial solution. Hence, setting $\delta(r) = 0$, we obtain the following solution for $A(r)$

$$A(r) = 1 + C_2 r^2 - \frac{C_3}{r^2} \quad \text{where} \quad C_2 = \frac{\alpha_0}{12\alpha_1} \tag{16}$$

which satisfies the null-energy condition.[14] C_3 is a constant of integration and can take any real value. We like to list the following important points regarding the above solution: First, it is easy to verify that the above solution satisfies the modified Einstein's equations (8). Second, C_2 is a positive constant, since, α_0 and α_1 are positive constants. Physically, C_2 acts like an effective cosmological constant. For $C_3 > 0$, the metric (7) has a horizon at

$$r_h = \sqrt{\frac{\sqrt{1 + 4C_2C_3} - 1}{2C_2}} \tag{17}$$

In the limit of $\alpha_0 \to 0$, $C_2 \to 0$, the metric (7) has a horizon at $r = \sqrt{C_3}$. Thus, $\alpha_0 \to 0$ is a smooth limit. Third, the term c_3/r^2 is a reminiscence of the charge in the Reissner-Nordström solution in GR.[14] In our case, the $1/r^2$ term is present in the absence of the $1/r$ term. While in GR, C_3/r^2 term can not exist without the mass term. This result is similar to the one obtained sometime back in the context of black-holes on the brane.[18] Physically, C_3 corresponds to the mass of the black-hole. Lastly, it is easy to verify that the above solution satisfies the modified Einstein's equations (8). The Kretschmann scalar for the above form of $A(r)$ is

$$R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} = \frac{\alpha_0^2}{6\alpha_1^2} + \frac{56C_3^2}{r^8}. \tag{18}$$

Thus, the metric has a singularity at $r = 0$ and is finite everywhere else. For finite α_0, the Kretschmann scalar is a positive constant at asymptotic infinity, which corresponds to asymptotic de Sitter space-times in GR.[14]

To obtain second solution, assume $A(r)$ in Eq. (16) and $\delta(r) \neq 0$. Now $A(r)$ in Eq. (16) in $T_2[A(r), \delta(r)] = 0$, this leads to the following differential equation for $\delta(r)$:

$$\delta''(r) + \frac{1}{2}\delta'(r)^2 + \frac{(5C_2r^4 + 2r^2 + C_3)}{(C_2r^4 + r^2 - C_3)}\frac{\delta'(r)}{r} = 0 \tag{19}$$

This differential equation is highly non-linear; however, it has the following exact solution:

$$e^{\delta(r)/2} = \frac{C_4}{2} - \frac{C_5}{2(12\alpha_1)^{3/2}}\frac{(2C_2r^2 + 1)}{(4C_2C_3 + 1)}\frac{1}{(C_2r^4 + r^2 - C_3)^{1/2}} \tag{20}$$

where C_4 and C_5 are arbitrary constants. [C_4 is dimensionless while C_5 has dimension of $[L]$.] The horizon for this new solution is given by (17). The Kretschmann scalar again can be evaluated, and near origin is singular, i.e.

$$R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} \sim \frac{\Gamma(C_2, C_3, C_4, C_5)}{r^8} \tag{21}$$

and is finite everywhere else. Thus, the constants C_4 and C_5 determine the behavior of the second black-hole solution. Thus, we have reconfirmed the results in the earlier section and shown that at least two black-hole solutions correspond to the same matter configuration (in this case, vacuum). To our knowledge, this is a novel result for any modified theories of gravity and confirms that Birkhoff's theorem is not valid for all modified gravity theories.

5. Conclusions and Discussion

In this talk, we obtained many exact static spherically symmetric vacuum solutions for $f(R)$ gravity. We obtained two exact vacuum black-hole solutions to the $f(R)$ model to emphasize this unique feature. We showed that two solutions have the event-horizon at the same point; however, their asymptotic features are different. Our results point that the Birkhoff theorem is not valid for all modified gravity theories. The two black-hole solutions have a space-time singularity at the origin.

Unlike in the literature, we have obtained the exact solutions *without* transforming to a conformal frame. Our analysis shows the deficiency of finding solutions in the conformally transformed frame.[1] The conformal transformations are not well-defined near $T_2 \to 0$ and, hence, the conformal frame will not be able to pick off the solutions we have obtained. However, the solution corresponding to $T_1[A(r), \delta(r)] = 0$ will be well-defined in the conformal frame.

One of the prospects of gravitational wave observations is to find signatures for the modified gravity theories. We have shown that if the modified theories belong to one of the degenerate classes with $T_2 = 0$, then our analysis shows that the prospect of detection needs different methodologies than the one that is currently used.[19]

Acknowledgments

This work was made possible by the use of GRTensor 3 package for MAPLE. SX is financially supported by the MHRD fellowship at IIT Bombay. The work is supported by the MATRICS SERB grant.

References

1. S. Xavier, J. Mathew and S. Shankaranarayanan, Infinitely degenerate exact ricci-flat solutions in f(r) gravity, *Class. Quant. Grav.* **37**, p. 225006 (2020).
2. K. S. Stelle, Renormalization of higher-derivative quantum gravity, **16**, 953.
3. S. Capozziello and M. D. Laurentis, Extended theories of gravity (August 2011).
4. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, Modified gravity and cosmology, *Physics Reports 513, 1 (2012), 1-189* (June 2011).
5. S. Nojiri, S. D. Odintsov and V. K. Oikonomou, Modified gravity theories on a nutshell: Inflation, bounce and late-time evolution (May 2017).
6. T. P. Sotiriou and V. Faraoni, f(r) theories of gravity, *Rev. Mod. Phys. 82:451-497,2010* (May 2008).
7. S.Shankaranarayanan, Strong gravity signatures in the polarization of gravitational waves, *Int. J. Mod. Phys. D* **28**, p. 1944020 (2019).

8. J. D. Bekenstein, Novel "no-scalar-hair" theorem for black holes, *Phys. Rev. D* **51**, p. R6608 (1995).
9. P. T. Chruściel, J. L. Costa and M. Heusler, Stationary black holes: Uniqueness and beyond, *Living Rev. Relativity 15, (2012), 7* (May 2012).
10. N. Voje Johansen and F. Ravndal, On the discovery of birkhoff's theorem, *Gen. Rel. Grav.* **38**, 537 (2006).
11. H.-J. Schmidt, The tetralogy of birkhoff theorems, *Gen. Rel. Grav.* **45**, 395 (2013).
12. R. P. Woodard, Avoiding dark energy with 1/r modifications of gravity, *Lect. Notes Phys.* **720**, 403 (2007).
13. S. Bhattacharyya and S. Shankaranarayanan, Quasinormal modes as a distinguisher between general relativity and f(r) gravity, *Phys. Rev. D 96, 064044 (2017)* (April 2017).
14. C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).
15. P. Cañate, L. G. Jaime and M. Salgado, Spherically symmetric black holes in $f(r)$ gravity: Is geometric scalar hair supported ? (September 2015).
16. J. Oliva and S. Ray, Birkhoff's theorem in higher derivative theories of gravity, *Class. Quant. Grav. 28:175007,2011* (April 2011).
17. S. Capozziello and D. Sáez-Gómez, Scalar-tensor representation of $f(r)$ gravity and birkhoff's theorem (July 2011).
18. N. Dadhich, R. Maartens, P. Papadopoulos and V. Rezania, Black holes on the brane, *Phys. Lett. B* **487**, 1 (2000).
19. L. Barack *et al.*, Black holes, gravitational waves and fundamental physics: a roadmap, *Class. Quant. Grav.* **36**, p. 143001 (2019).

Universe in a black hole with spin and torsion

Nikodem Popławski

Department of Mathematics and Physics, University of New Haven,
West Haven, CT 06516, USA
E-mail: NPoplawski@newhaven.edu
www.newhaven.edu

We consider gravitational collapse of a spherically symmetric sphere of a fluid with spin and torsion into a black hole. We use the Tolman metric and the Einstein–Cartan field equations with a relativistic spin fluid as a source. We show that gravitational repulsion of torsion prevents a singularity and replaces it with a nonsingular bounce. Quantum particle production during contraction strengthens torsion in opposing shear. Particle production during expansion can produce enormous amounts of matter and generate a finite period of inflation. The resulting closed universe on the other side of the event horizon may have several bounces. Such a universe is oscillatory, with each cycle larger in size then the previous cycle, until it reaches the cosmological size and expands indefinitely. Our universe might have therefore originated from a black hole existing in another universe.

Keywords: Einstein-Cartan theory; torsion; gravitational collapse.

1. Einstein–Cartan gravity and spin fluid

In general relativity (GR), the affine connection $\Gamma_{ij}^{\ \ k}$ is constrained to be symmetric. Einstein–Cartan–Sciama–Kibble (EC) gravity removes this constraint by regarding the antisymmetric part of the connection, the torsion tensor $S^k_{\ ij} = \Gamma_{[ij]}^{\ \ \ k}$, as a variable.[1] The total Lagrangian density is $-\frac{1}{2\kappa}R\sqrt{-g} + L_{\rm m}$, where R is the Ricci scalar constructed from the connection and $L_{\rm m}$ is the Lagrangian density of matter, as in GR.[2]

Varying the Lagrangian with respect to the contortion tensor $C_{ijk} = S_{ijk} + S_{jki} + S_{kji}$ gives the Cartan equations:

$$S^j_{\ ik} - S_i\delta^j_k + S_k\delta^j_i = -\frac{1}{2}\kappa s_{ik}^{\ \ \ j},$$

where $S_i = S^k_{\ ik}$ and $s^{ijk} = 2(\delta L_{\rm m}/\delta C_{ijk})/\sqrt{-g}$ is the spin tensor. Varying the Lagrangian with respect to the metric tensor g_{ik} gives the Einstein equations with the Ricci tensor. They can be put into a GR form with the modified energy–momentum tensor:

$$G^{ik} = \kappa T^{ik} + \kappa^2\Big(-s^{ij}_{\ \ \ [l}s^{kl}_{\ \ \ j]} - \frac{1}{2}s^{ijl}s^k_{\ jl} + \frac{1}{4}s^{jli}s_{jl}^{\ \ \ k} + \frac{1}{8}g^{ik}(-4s^l_{\ j[m}s^{jm}_{\ \ \ l]} + s^{jlm}s_{jlm})\Big).$$

Dirac spinors, representing fermions, couple to torsion through the covariant derivative in the Lagrangian and therefore are the source of torsion. At macroscopic

scales, Dirac fields can be averaged and described as a spin fluid:[3]

$$s_{ij}{}^{k} = s_{ij}u^{k}, \quad s_{ij}u^{j} = 0.$$

The terms in the effective energy–momentum tensor that are quadratic in the spin tensor do not vanish after averaging:

$$G^{ij} = \kappa\Big(\epsilon - \frac{1}{4}\kappa s^2\Big)u^i u^j - \kappa\Big(p - \frac{1}{4}\kappa s^2\Big)(g^{ij} - u^i u^j),$$

where

$$s^2 = \frac{1}{2}s_{ij}s^{ij} > 0 \propto n_{\mathrm{f}}^2$$

is the averaged square of the spin density.[4] The Einstein–Cartan equations for a spin fluid are therefore equivalent to the GR Einstein equations for a perfect fluid with:

$$\tilde{\epsilon} = \epsilon - \alpha n_{\mathrm{f}}^2, \quad \tilde{p} = p - \alpha n_{\mathrm{f}}^2,$$

where ϵ and p are the thermodynamic energy density and pressure, n_{f} is the number density of fermions, and $\alpha = \kappa(\hbar c)^2/32$ with $\kappa = 8\pi G/c^4$.

2. Gravitational collapse of a fluid sphere

A spherically symmetric gravitational field is given by the Tolman metric:[5]

$$ds^2 = e^{\nu(\tau,R)}c^2 d\tau^2 - e^{\lambda(\tau,R)}dR^2 - e^{\mu(\tau,R)}(d\theta^2 + \sin^2\theta\, d\phi^2),$$

where ν, λ, and μ are functions of a time coordinate τ and a radial coordinate R. Coordinate transformations $\tau \to \tilde{\tau}(\tau)$ and $R \to \tilde{R}(R)$ do not change the form of the metric. The components of the Einstein tensor corresponding to this metric that do not vanish identically are:

$$G_0^0 = -e^{-\lambda}\Big(\mu'' + \frac{3\mu'^2}{4} - \frac{\mu'\lambda'}{2}\Big) + \frac{e^{-\nu}}{2}\Big(\dot{\lambda}\dot{\mu} + \frac{\dot{\mu}^2}{2}\Big) + e^{-\mu},$$

$$G_1^1 = -\frac{e^{-\lambda}}{2}\Big(\frac{\mu'^2}{2} + \mu'\nu'\Big) + e^{-\nu}\Big(\ddot{\mu} - \frac{\dot{\mu}\dot{\nu}}{2} + \frac{3\dot{\mu}^2}{4}\Big) + e^{-\mu},$$

$$G_2^2 = G_3^3 = -\frac{e^{-\nu}}{4}(\dot{\lambda}\dot{\nu} + \dot{\mu}\dot{\nu} - \dot{\lambda}\dot{\mu} - 2\ddot{\lambda} - \dot{\lambda}^2 - 2\ddot{\mu} - \dot{\mu}^2)$$

$$-\frac{e^{-\lambda}}{4}(2\nu'' + \nu'^2 + 2\mu'' + \mu'^2 - \mu'\lambda' - \nu'\lambda' + \mu'\nu'),$$

$$G_0^1 = \frac{e^{-\lambda}}{2}(2\dot{\mu}' + \dot{\mu}\mu' - \dot{\lambda}\mu' - \dot{\mu}\nu'),$$

where a dot denotes differentiation with respect to $c\tau$ and a prime denotes differentiation with respect to R.

In the comoving frame of reference, the spatial components of the four-velocity u^μ vanish. The nonzero components of the energy–momentum tensor for a spin fluid, $T_{\mu\nu} = (\tilde{\epsilon} + \tilde{p})u_\mu u_\nu - \tilde{p}g_{\mu\nu}$, are: $T_0^0 = \tilde{\epsilon}$, $T_1^1 = T_2^2 = T_3^3 = -\tilde{p}$. The Einstein field equations $G_\nu^\mu = \kappa T_\nu^\mu$ in this frame of reference are:

$$G_0^0 = \kappa\tilde{\epsilon}, \quad G_1^1 = G_2^2 = G_3^3 = -\kappa\tilde{p}, \quad G_0^1 = 0.$$

The covariant conservation of the energy–momentum tensor gives

$$\dot{\lambda} + 2\dot{\mu} = -\frac{2\dot{\tilde{\epsilon}}}{\tilde{\epsilon} + \tilde{p}}, \quad \nu' = -\frac{2\tilde{p}'}{\tilde{\epsilon} + \tilde{p}},$$

where the constants of integration depend on the allowed transformations $\tau \to \tilde{\tau}$ and $R \to \tilde{R}$.

If the pressure is homogeneous (no pressure gradients), then $\tilde{p}' = 0$, which gives $\nu' = 0$. Therefore, $\nu = \nu(\tau)$ and a transformation $\tau \to \tilde{\tau}$ can bring ν to zero and $g_{00} = e^{\nu}$ to 1. The system of coordinates becomes synchronous. Defining $r(\tau, R) = e^{\mu/2}$ turns the metric into[5]

$$ds^2 = c^2 d\tau^2 - e^{\lambda(\tau,R)} dR^2 - r^2(\tau, R)(d\theta^2 + \sin^2\theta \, d\phi^2).$$

Every particle in a collapsing fluid sphere is represented by a value of R that ranges from 0 (at the center) to R_0 (at the surface). The Einstein field equations reduce to

$$\begin{aligned}
&\kappa\tilde{\epsilon} = -\frac{e^{-\lambda}}{r^2}(2rr'' + r'^2 - rr'\lambda') + \frac{1}{r^2}(r\dot{r}\dot{\lambda} + \dot{r}^2 + 1),\\
&-\kappa\tilde{p} = \frac{1}{r^2}(-e^{-\lambda}r'^2 + 2r\ddot{r} + \dot{r}^2 + 1),\\
&-2\kappa\tilde{p} = -\frac{e^{-\lambda}}{r}(2r'' - r'\lambda') + \frac{\dot{r}\dot{\lambda}}{r} + \ddot{\lambda} + \frac{1}{2}\dot{\lambda}^2 + \frac{2\ddot{r}}{r},\\
&2\dot{r}' - \dot{\lambda}r' = 0.
\end{aligned}$$

Integrating the last equation gives

$$e^{\lambda} = \frac{r'^2}{1 + f(R)}, \tag{1}$$

where f is a function of R satisfying a condition $1 + f > 0$.[6] Substituting this relation into the second field equation gives $2r\ddot{r} + \dot{r}^2 - f = -\kappa\tilde{p}r^2$, which is integrated to

$$\dot{r}^2 = f(R) + \frac{F(R)}{r} - \frac{\kappa}{r}\int \tilde{p}r^2 dr,$$

where F is a positive function of R. Substituting the last two equations into the first field equation gives a relation $\kappa(\tilde{\epsilon} + \tilde{p}) = F'(R)/(r^2 r')$, leading to

$$\dot{r}^2 = f(R) + \frac{\kappa}{r}\int_0^R \tilde{\epsilon} r^2 r' dR. \tag{2}$$

If the mass of the sphere is M, then the Schwarzschild radius $r_g = 2GM/c^2$ of the black hole that forms from the sphere is equal to

$$r_g = \kappa \int_0^{R_0} \tilde{\epsilon} r^2 r' dR.$$

These two equations give $\dot{r}^2(\tau, R_0) = f(R_0) + r_g/r(\tau, R_0)$. If $r_0 = r(0, R_0)$ is the initial radius of the sphere and the sphere is initially at rest, then $\dot{r}(0, R_0) = 0$. Consequently, the value of R_0 is determined by

$$f(R_0) = -\frac{r_g}{r_0}.$$

Substituting $r = e^{\mu/2}$ and (1) into the first conservation law gives the first law of thermodynamics for the effective energy density and pressure:

$$\frac{d}{d\tau}(\tilde{\epsilon} r^2 r') + \tilde{p}\frac{d}{d\tau}(r^2 r') = 0. \tag{3}$$

3. Collapse of a spin fluid sphere

If we assume that the spin fluid is composed of an ultrarelativistic matter in kinetic equilibrium, then $\epsilon = h_\star T^4$, $p = \epsilon/3$, and $n_{\mathrm{f}} = h_{n\mathrm{f}} T^3$, where T is the temperature of the fluid, $h_\star = (\pi^2/30)(g_{\mathrm{b}} + (7/8)g_{\mathrm{f}})k^4/(\hbar c)^3$, and $h_{n\mathrm{f}} = (\zeta(3)/\pi^2)(3/4)g_{\mathrm{f}}k^3/(\hbar c)^3$. For standard-model particles, $g_{\mathrm{b}} = 29$ and $g_{\mathrm{f}} = 90$. The effective energy density and pressure are thus:

$$\tilde{\epsilon} = h_\star T^4 - \alpha h_{n\mathrm{f}}^2 T^6, \quad \tilde{p} = \frac{1}{3} h_\star T^4 - \alpha h_{n\mathrm{f}}^2 T^6.$$

Since the pressure has no gradient, the temperature only depends on τ, and so does the energy density. This scenario describes a homogeneous sphere. The first law of thermodynamics (3) gives

$$r^2 r' T^3 = g(R), \tag{4}$$

where g is a function of R. Putting this relation into (2) gives

$$\dot{r}^2 = f(R) + \frac{\kappa}{r}(h_\star T^4 - \alpha h_{n\mathrm{f}}^2 T^6) \int_0^R r^2 r' dR. \tag{5}$$

Equations (4) and (5) give the function $r(\tau, R)$, which with (1) gives $\lambda(\tau, R)$.[7] The integration of (5) also contains the initial value $\tau_0(R)$. The metric therefore depends on three arbitrary functions: $f(R)$, $g(R)$, and $\tau_0(R)$.

We seek a solution of (4) and (5) as

$$f(R) = -\sin^2 R, \quad r(\tau, R) = a(\tau)\sin R, \tag{6}$$

where $a(\tau)$ is a nonnegative function of τ. For this choice of functions, (4) gives $a^3 T^3 \sin^2 R \cos R = g(R)$, in which separation of the variables τ and R leads to

$$g(R) = \mathrm{const} \cdot \sin^2 R \cos R, \quad a^3 T^3 = \mathrm{const}.$$

Consequently,

$$aT = a_0 T_0, \quad \frac{\dot{T}}{T} + \frac{H}{c} = 0, \tag{7}$$

where $a_0 = a(\tau = 0)$ and $T_0 = T(\tau = 0)$ are the values at the initial time, and $H = c\dot{a}/a$. Substituting (6) into (5) gives

$$\dot{a}^2 + 1 = \frac{\kappa}{3}(h_\star T^4 - \alpha h_{n\mathrm{f}}^2 T^6)a^2, \tag{8}$$

which has a form of the Friedmann equation for the scale factor a as a function of the cosmic time τ in a closed, homogeneous universe. The quantity H is the Hubble parameter of this universe. Using (7) in (8) yields[8]

$$\dot{a}^2 = -1 + \frac{\kappa}{3}\Big(\frac{h_\star T_0^4 a_0^4}{a^2} - \frac{\alpha h_{nf}^2 T_0^6 a_0^6}{a^4}\Big). \tag{9}$$

The relations (6) determine the constants:

$$\sin R_0 = \Big(\frac{r_g}{r_0}\Big)^{1/2}, \quad a(0) = \Big(\frac{r_0^3}{r_g}\Big)^{1/2}.$$

Substituting the initial values $a(0)$ and $\dot{a}(0) = 0$ into (8), in which the second term on the right-hand side is negligible, gives $Mc^2 = (4\pi/3)r_0^3 h_\star T_0^4$. This relation indicates the equivalence of mass and energy of a fluid sphere with radius r_0 and determines T_0. An event horizon for the entire sphere forms when $r(\tau, R_0) = r_g$, which is equivalent to $a = (r_g r_0)^{1/2}$. Equation (9) has two turning points, $\dot{a} = 0$, if

$$\frac{r_0^3}{r_g} > \frac{3\pi G\hbar^4 h_{nf}^4}{8h_\star^3} \sim l_{\text{Planck}}^2,$$

which is satisfied for astrophysical systems that form black holes.[9]

Substituting (6) into (1) gives $e^{\lambda(\tau,R)} = a^2$. Consequently, the square of an infinitesimal interval in the interior of a collapsing spin fluid is given by

$$ds^2 = c^2 d\tau^2 - a^2(\tau) dR^2 - a^2(\tau)\sin^2 R(d\theta^2 + \sin^2\theta\, d\phi^2).$$

This metric has a form of the closed Friedmann–Lemaître–Robertson–Walker metric and describes a part of a closed universe with $0 \le R \le R_0$ (like dust).

4. Nonsingular bounce

Equation (9) can be solved analytically in terms of an elliptic integral of the second kind, giving the function $a(\tau)$ and then $r(\tau, R) = a(\tau)\sin R$. The value of a never reaches zero because as a decreases, the right-hand side of (9) becomes negative, contradicting the left-hand side. The change of the sign occurs when $a < (r_g r_0)^{1/2}$, that is, after the event horizon forms. Consequently, all particles with $R > 0$ fall within the event horizon but never reach $r = 0$ (the only particle at the center is the particle that is initially at the center, with $R = 0$). A singularity is replaced with a nonsingular bounce.[7] Nonzero values of a give finite values of T and thus finite values of ϵ, p, and n_{f}.

After the bounce, the matter expands on the other side of the event horizon as a new universe. This universe has a closed geometry (constant positive curvature). The quantity $a(\tau)$ is the scale factor. The universe is oscillatory: the value of a oscillates between the two turning points. The value of R_0 does not change. A turning point at which $\ddot{a} > 0$ is a bounce, and a turning point at which $\ddot{a} < 0$ is a crunch. This universe has an infinite number of identical cycles.

The Raychaudhuri equation for a congruence of geodesics without four-acceleration and rotation is $d\theta/ds = -\theta^2/3 - 2\sigma^2 - R_{\mu\nu}u^\mu u^\nu$, where θ is the expansion scalar, σ^2 is the shear scalar, and $R_{\mu\nu}$ is the Ricci tensor. For a spin fluid, the last term in this equation is equal to $-\kappa(\tilde{\epsilon} + 3\tilde{p})/2$. Consequently, the necessary and sufficient condition for avoiding a singularity in a black hole is $-\kappa(\tilde{\epsilon} + 3\tilde{p})/2 > 2\sigma^2$. For a relativistic spin fluid, $p = \epsilon/3$, this condition is equivalent to

$$2\kappa\alpha n_{\mathrm{f}}^2 > 2\sigma^2 + \kappa\epsilon. \tag{10}$$

Without torsion, the left-hand side of (10) would be absent and this inequality could not be satisfied, resulting in a singularity. Torsion may provide a necessary condition for preventing a singularity.[10] In the absence of shear, this condition would be also sufficient.[11]

The presence of shear opposes the effects of torsion. The shear scalar σ^2 grows with decreasing a like $\sim a^{-6}$, which is the same power law as that for n_{f}^2. Therefore, if the initial shear term dominates over the initial torsion term in (10), then it will dominate at later times during contraction and a singularity will form. To avoid a singularity if the shear is present, n_{f}^2 must grow faster than $\sim a^{-6}$. Consequently, fermions must be produced in a black hole during contraction.

The production rate of particles in a contracting or expanding universe[12] can be phenomenologically given by

$$\frac{1}{c\sqrt{-g}}\frac{d(\sqrt{-g}n_{\mathrm{f}})}{dt} = \frac{\beta H^4}{c^4}, \tag{11}$$

where $g = -a^6 \sin^4 R \sin^2\theta$ is the determinant of the metric tensor and β is a nondimensional production rate. With particle production, the second equation in (7) turns into[8]

$$\frac{\dot{T}}{T} = \frac{H}{c}\Big(\frac{\beta H^3}{3c^3 h_{n\mathrm{f}} T^3} - 1\Big). \tag{12}$$

Particle production changes the power law $n_{\mathrm{f}}(a)$:

$$n_{\mathrm{f}} \sim a^{-(3+\delta)},$$

where δ varies with τ. Putting this relation into (11) gives

$$\delta \sim -a^\delta \dot{a}^3.$$

During contraction, $\dot{a} < 0$ and thus $\delta > 0$. The term $n_{\mathrm{f}}^2 \sim a^{-6-2\delta}$ grows faster than $\sigma^2 \sim a^{-6}$ and a singularity is avoided. Particle production and torsion together reverse the effects of shear, generating a bounce.[7]

5. Closed universe in a black hole

The dynamics of the nonsingular, relativistic universe in a black hole is described by equations (8) and (12), where $H = c\dot{a}/a$. These equations, with the initial conditions $a(0) = (r_0^3/r_g)^{1/2}$ and $\dot{a}(0) = 0$, give the functions $a(\tau)$ and $T(\tau)$.

The shear would enter the right-hand side of the first equation as an additional positive term that is proportional to a^{-4}. When the universe becomes nonrelativistic, the term $h_\star T^4$ changes into a positive term proportional to a^{-1}. The cosmological constant enters as a positive term proportional to a^2.

Particle production increases the maximum size of the scale factor that is reached at a crunch. Consequently, a new cycle is larger and lasts longer then the previous cycle. R_0 is given by $\sin^3 R_0 = r_g/a(0)$, where $a(0)$ is the maximum scale factor in the first cycle. Since the maximum scale factor in the next cycle is larger, the value of $\sin R_0$ decreases. As cycles proceed, R_0 approaches π (the value for a completely closed universe).

A parent black hole creating a new, baby universe becomes an Einstein–Rosen bridge (unidirectional wormhole) to that universe.[13] If the universe is closed, it is analogous to the 2-dimensional surface of a 3-dimensional baloon (sphere). The universe would be mathematically the 3-dimensional hypersurface of a 4-dimensional hypersphere. The 3-dimensional space in which the balloon expands is not analogous to any higher dimensional space. Points off the surface of the balloon are not in the universe in this analogy. The balloon radius represents the scale factor a. The universe expands, which can be described locally by the Hubble law.

This scenario occurs also if the fermionic matter is described by the Dirac Lagrangian density instead of the spin fluid.[14]

6. Inflation and oscillations

During expansion ($H > 0$), if β is too big, then the right-hand side of (12) could become positive. In this case, the temperature would grow with increasing a, which would lead to eternal inflation. Consequently, there is an upper limit to the production rate: the maximum of the function $(\beta H^3)/(3c^3 h_{nf} T^3)$ must be lesser than 1.

If $(\beta H^3)/(3c^3 h_{nf} T^3)$ increases after a bounce to a value that is slightly lesser than 1, then T would become approximately constant. Accordingly, H would be also nearly constant and the scale factor a would grow exponentially, generating inflation.[8] Since the energy density would be also nearly constant, the universe would produce enormous amounts of matter and entropy. Such an expansion would last until the right-hand side of drops below 1. Consequently, inflation would last a finite period of time. After this period, the effects of torsion weaken and the universe smoothly enters the radiation-dominated expansion, followed by the matter-dominated expansion.

Torsion and particle production can therefore generate finite inflation without scalar fields and reheating, which is consistent with the astronomical data.[15] If quantum effects in the gravitational field near a bounce do not produce enough matter, then the closed Universe reaches the maximum size and then contracts to another bounce, beginning the new cycle. Because of matter production, a new cycle reaches larger size and last longer than the previous cycle.[16] When the Universe reaches a

size at which the cosmological constant becomes dominant, it avoids another contraction and expands to infinity.[17]

7. Our universe in a black hole

These results suggest that every black hole may create a new, closed, baby universe. Accordingly, our Universe may be closed and may have born in the interior of a black hole existing in a parent universe.[13,18] Its energy would be equal to zero minus the energy of the wormhole connecting the Universe to the parent universe.[19] This hypothesis could solve the black hole information paradox: the information goes through the Einstein–Rosen bridge to the baby universe on the other side of the black hole's event horizon. The motion of matter through an event horizon is unidirectional: it defines the past and future. Time asymmetry at the event horizon may induce time asymmetry everywhere in the baby universe and could explain why time flows in one direction. Recent astronomical data seem to favor the closed geometry of the Universe,[20] thus providing some support for the black-hole cosmology.

8. Summary

The conservation law for total angular momentum in curved spacetime, consistent with Dirac equation, may require torsion.[2] The simplest theory with torsion, the Einstein–Cartan gravity, has the same Lagrangian as GR, but the affine connection contains the torsion tensor, generated by spin. Gravitational collapse of a sphere of a spin fluid creates an event horizon. The matter within the horizon collapses to extremely high densities, at which torsion acts like gravitational repulsion. Semiclassically and without shear, torsion prevents a singularity and replaces it with a nonsingular bounce. With shear, torsion prevents a singularity if the number of fermions increases during contraction through quantum particle production.[7] Particle production during expansion produces enormous amounts of matter and can generate a finite period of inflation. The resulting closed universe on the other side of the event horizon may have several bounces. Such a universe is oscillatory, with each cycle larger in size then the previous cycle, until it reaches a size at which the cosmological constant dominates, and then expands indefinitely.

In addition to eliminating gravitational singularities, torsion may also remove loop divergences in Feynman diagrams in quantum electrodynamics, resulting in finite values of bare (before renormalization) quantities such as the mass and electric charge of the electron.[21] Consequently, torsion may be a physical phenomenon that ensures physics is finite.

Acknowledgments

This work was funded by the University Research Scholar program at the University of New Haven. The author is thankful to Michael Del Grosso, Jordan Cubero, Gabriel Unger, and Shantanu Desai for valuable discussions.

References

1. L. P. Eisenhart, *Non-Riemannian Geometry* (American Mathematical Society, 1927); E. Schrödinger, *Space-time Structure* (Cambridge University Press, 1954); J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, 1954).
2. D. W. Sciama, Proc. Camb. Phil. Soc. **54**, 72 (1958); T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961); D. W. Sciama, in *Recent Developments in General Relativity*, p. 415 (Pergamon, 1962); Rev. Mod. Phys. **36**, 463 (1964); Rev. Mod. Phys. **36**, 1103 (1964); F. W. Hehl and B. K. Datta, J. Math. Phys. **12**, 1334 (1971); F. W. Hehl, Gen. Relativ. Gravit. **4**, 333 (1973); Gen. Relativ. Gravit. **5**, 491 (1974); F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976); V. de Sabbata and M. Gasperini, *Introduction to Gravitation* (World Scientific, 1985); F. W. Hehl and J. D. McCrea, Found. Phys. **16**, 267 (1986); V. de Sabbata and C. Sivaram, *Spin and Torsion in Gravitation* (World Scientific, 1994); N. Popławski, *Classical Physics: Spacetime and Fields*, arXiv:0911.0334; N. J. Popławski, Phys. Lett. B **690**, 73 (2010); Phys. Lett. B **727**, 575 (2013).
3. K. Nomura, T. Shirafuji, and K. Hayashi, Prog. Theor. Phys. **86**, 1239 (1991).
4. F. W. Hehl, P. von der Heyde, and G. D. Kerlick, Phys. Rev. D **10**, 1066 (1974).
5. R. A. Tolman, Proc. Natl. Acad. Sci. USA **20**, 169 (1934); J. R. Oppenheimer and H. Snyder, Phys. Rev. **56**, 455 (1939).
6. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, 1975).
7. N. Popławski, Zh. Eksp. Teor. Fiz. **159**, 448 (2021); J. Exp. Theor. Phys. **132**, 374 (2021).
8. N. Popławski, Astrophys. J. **832**, 96 (2016); Int. J. Mod. Phys. D **27**, 1847020 (2018).
9. G. Unger and N. Popławski, Astrophys. J. **870**, 78 (2019).
10. F. W. Hehl, Abh. Braunschw. Wiss. Ges. **18**, 98 (1966); A. Trautman, Bull. Acad. Polon. Sci., Serie Sci. Math. Astr. Phys. **20**, 185 (1972); W. Kopczyński, Phys. Lett. A **39**, 219 (1972); W. Kopczyński, Phys. Lett. A **43**, 63 (1973); A. Trautman, Symp. Math. **12**, 139 (1973); Nature Phys. Sci. **242**, 7 (1973).
11. B. Kuchowicz, Gen. Relativ. Gravit. **9**, 511 (1978); M. Gasperini, Phys. Rev. Lett. **56**, 2873 (1986); Y. N. Obukhov and V. A. Korotky, Class. Quantum Grav. **4**, 1633 (1987); N. J. Popławski, Phys. Lett. B **694**, 181 (2010); Phys. Lett. B **701**, 672 (2011); Gen. Relativ. Gravit. **44**, 1007 (2012).
12. L. Parker, Phys. Rev. Lett. **21**, 562 (1968); Phys. Rev. **183**, 1057 (1969); Y. B. Zel'dovich, J. Exp. Theor. Phys. Lett. **12**, 307 (1970); L. Parker, Phys. Rev. D **3**, 346 (1971); Phys. Rev. D **3**, 2546 (1971); Y. B. Zel'dovich and A. A. Starobinskii, J. Exp. Theor. Phys. Lett. **26**, 252 (1977); V. A. Beilin, G. M. Vereshkov, Y. S. Grishkan, N. M. Ivanov, V. A. Nesterenko, and A. N. Poltavtsev, J. Exp. Theor. Phys. **51**, 1045 (1980).
13. N. J. Popławski, Phys. Lett. B **687**, 110 (2010).
14. N. Popławski, Phys. Rev. D **85**, 107502 (2012); J. Magueijo, T. G. Zlosnik, and T. W. B. Kibble, Phys. Rev. D **87**, 063504 (2013); J. L. Cubero and N. J. Popławski, Class. Quantum Grav. **37**, 025011 (2020).
15. S. Desai and N. J. Popławski, Phys. Lett. B **755**, 183 (2016).
16. J. D. Barrow and M. P. Dąbrowski, Mon. Not. Roy. Astron. Soc. **275**, 850 (1995); J. D. Barrow and C. Ganguly, Int. J. Mod. Phys. D **26**, 1743016 (2017).
17. H. Bondi, *Cosmology* (Cambridge University Press, 1960); E. A. Lord, *Tensors, Relativity and Cosmology* (McGraw-Hill, 1976).
18. I. D. Novikov, J. Exp. Theor. Phys. Lett. **3**, 142 (1966); R. K. Pathria, Nature **240**, 298 (1972); V. P. Frolov, M. A. Markov, and V. F. Mukhanov, Phys. Lett. B **216**, 272 (1989); Phys. Rev. D **41**, 383 (1990); L. Smolin, Class. Quantum Grav. **9**, 173 (1992);

S. Hawking, *Black Holes and Baby Universes and other Essays* (Bantam Dell, 1993); W. M. Stuckey, Am. J. Phys. **62**, 788 (1994); D. A. Easson and R. H. Brandenberger, J. High Energ. Phys. **06**, 024 (2001); J. Smoller and B. Temple, Proc. Natl. Acad. Sci. USA **100**, 11216 (2003).
19. N. J. Popławski, Class. Quantum Grav. **31**, 065005 (2014); N. Popławski, Mod. Phys. Lett. A **33**, 1850236 (2018).
20. E. Di Valentino, A. Melchiorri, and J. Silk, Nature Astron. **4**, 196 (2020).
21. N. Popławski, Found. Phys. **50**, 900 (2020); M. Del Grosso and N. Popławski, arXiv:2107.01612.

Asymptotically flat black hole solution in modified gravity

Surajit Kalita[1] and Banibrata Mukhopadhyay[2]
Department of Physics, Indian Institute of Science, Bangalore-560012, India
[1]*E-mail: surajitk@iisc.ac.in*
[2]*E-mail: bm@iisc.ac.in*

From time to time, different observations suggest that Einstein's theory of general relativity (GR) may not be the ultimate theory of gravity. Various researchers have suggested that the $f(R)$ theory of gravity is the best alternative to replace GR. Using $f(R)$ gravity, one can elucidate the various unexplained physics of compact objects, such as black holes, neutron stars, and white dwarfs. Researchers have already put effort into finding the vacuum solution around a black hole in $f(R)$ gravity. However, for a long time, they could not find an asymptotically flat vacuum solution. In this article, we show that the asymptotically flat vacuum solution of $f(R)$ gravity is possible and thereby use it to explain the spherical accretion flow around the black hole.

Keywords: Black hole, modified gravity, spherical accretion.

1. Introduction

Einstein's theory of general relativity (GR) is undoubtedly the most beautiful theory to explain gravity. It can easily explain a large number of phenomena where Newtonian gravity falls short, such as the deflection of light in strong gravity, generation of the gravitational waves, perihelion precession of Mercury's orbit, gravitational redshift of light, to mention a few. However, following several recent observations in cosmology and the high-density regions in the universe, such as the vicinity of compact objects, it seems that GR may not be the ultimate theory of gravity. Starobinsky first used one of the modified theories of gravity, namely R^2-gravity,[1] to explain the cosmology of the very early universe. Eventually, researchers have proposed different modifications of GR, e.g., various models of $f(R)$ gravity, to elaborate on various unexplained physics related to compact objects. Capozziello and his collaborators used different forms of $f(R)$ gravity to explain the massive neutron stars.[2] Similarly, Kalita & Mukhopadhyay used $f(R)$ gravity to unify the physics of all white dwarfs, including those possessing sub- and super-Chandrasekhar limiting masses,[3] and show that they can be detected by various proposed gravitational wave detectors.[4]

According to Birkhoff's theorem, the spherically symmetric vacuum solution in GR is static, which is the famous Schwarzschild solution. In the same context, the vacuum solution of $f(R)$ gravity is also an interesting problem. Multamäki & Vilja first obtained the solutions for a static, spherically symmetric vacuum spacetime in $f(R)$ gravity,[5] and argued that for a large class model, Schwarzschild-de Sitter metric is an exact solution of the field equations. Eventually, many researchers have

obtained several solutions for different modified theories of gravity in various spacetime geometries. Capozziello and his collaborators obtained spherically symmetric as well as axially symmetric vacuum solutions in $f(R)$ gravity using Noether symmetry.[6,7] These new solutions alter the black hole's event horizon, thereby various critical orbits, such as marginally stable, marginally bound, and photon orbits. In this way, modified theories of gravity change the dynamics of the particles moving around the black hole.

The above-mentioned solutions, introduced in these literatures, always contain some diverging terms in the metric components. Hence these solutions never reduce to the Minkowski metric at the asymptotically flat limit. This asymptotic flatness is essential in the context of physical problems such as accretion disc, as a disc is expected to extend in a vast region around the compact object. Therefore, at the outermost portion of the disc, which is very far from the compact object, no physics should be violated as given by the Schwarzschild/Minkowski metric. Many of the models mentioned above assume scalar curvature, $R \neq 0$ throughout the spacetime, which is again questionable as for the Schwarzschild metric, $R = 0$, which needs to be satisfied at the asymptotic flat limit. Kalita & Mukhopadhyay, for the first time, obtained the spherically symmetric asymptotically flat vacuum solution in the premise of $f(R)$ gravity.[8] In this article, we primarily discuss this solution and its applications in the spherical accretion flow around a black hole.

2. Basic formalism

The 4-dimensional action for $f(R)$ gravity in vacuum is given by[9]

$$S = \frac{c^4}{16\pi G} \int \sqrt{-g}\; f(R)\, \mathrm{d}^4x\,, \tag{1}$$

where c is the speed of light, G is Newton's gravitational constant, R is the scalar curvature, and $g = \det(g_{\mu\nu})$ is the determinant of the metric $g_{\mu\nu}$. Varying this action with respect to $g_{\mu\nu}$, with appropriate boundary conditions, we obtain the modified Einstein equation for the $f(R)$ gravity, given by

$$f'(R)G_{\mu\nu} + \frac{1}{2}g_{\mu\nu}\left[Rf'(R) - f(R)\right] - \left(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box\right) f'(R) = 0, \tag{2}$$

where $\Box$ is the d'Alembertian operator given by $\Box = \nabla^\mu\nabla_\mu$ and ∇_μ is the covariant derivative. For $f(R) = R$, Equation (2) reduces to the well-known Einstein field equation in GR.

3. Asymptotically flat vacuum solution in $f(R)$ gravity

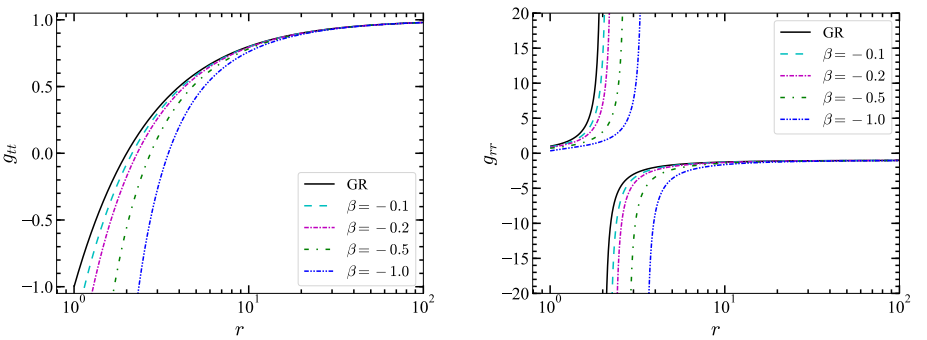
To obtain the spherically symmetric solution of the modified Einstein equation, we first assume $g_{\mu\nu} = \mathrm{diag}(s(r), -p(r), -r^2, -r^2\sin^2\theta)$, where s, p are the functions of the radial coordinate r only. We know that far from the compact objects, the metric should reduce first to the Schwarzschild metric, and at very far from the compact object, it should reduce to the Minkowski metric as the effect of gravity is

negligible there. Using these conditions at the asymptotic limit, one can show that for $f'(R) = 1 + \beta/r$ with β being a constant, in $c = G = M = 1$ unit, the temporal and radial components of the metric are given by[8]

$$g_{tt} = s(r) = 1 - \frac{2}{r} - \frac{\beta(\beta-6)}{2r^2} + \frac{\beta^2(13\beta-66)}{20r^3} - \frac{\beta^3(31\beta-156)}{48r^4} + \ldots, \tag{3}$$

$$g_{rr} = -p(r) = -\frac{16r^4}{(\beta+2r)^4}\frac{1}{s(r)}. \tag{4}$$

It is, of course, obvious from the above expressions that as $\beta \to 0$, it reduces to the Schwarzschild metric, and as $r \to \infty$, it reduces to the Minkowski metric, i.e., the flat spacetime.



(a) temporal component (b) radial component

Fig. 1. Variation of temporal and radial metric elements as functions of distance r. All the quantities are in dimensionless units of $c = G = M = 1$.

From Equations (3) and (4), it is observed that in $f(R)$ gravity, the vacuum solution is different from the Schwarzschild metric. This deviation indicates the violation of Birkhoff's theorem, which states that any spherically symmetric vacuum solution is essentially the Schwarzschild solution. Figure 1 describes the variations of the metric elements g_{tt} and g_{rr} with respect to r for different β. It is evident from the figure that as $r \to \infty$, the solutions reduce to the flat Minkowski metric, which, however, deviates near the black hole. $\beta = 0$ represents the Schwarzschild solution, with the radius of the event horizon $r_{\rm H} = 2$. In $f(R)$ gravity, as β deviates from 0, the solution is no longer the Schwarzschild solution, and the physics related to these solutions alter. For example, one can easily obtain the event horizon from Equations (3) and (4) by solving $g_{rr} \to \infty$, which shows that $r_{\rm H}$ shifts from 2 as β deviates from 0. Not only the radius of the event horizon, but one can also obtain the various marginal orbits in $f(R)$ gravity from Equations (3) and (4). Table 1 shows the radii of various orbits and their critical angular momenta for different β. We choose β to be negative because it guarantees that gravity is always attractive.

Table 1. Various characteristic parameters of space-time for different values of β: $r_{\rm H}$ is the event horizon, $r_{\rm MB}$ the marginally bound orbit, $r_{\rm MS}$ the marginally stable orbit, $L_{\rm MB}$ and $L_{\rm MS}$ are their corresponding specific angular momenta, and $r_{\rm ph}$ is the photon orbit. All the values are in dimensionless unit considering $c = G = M = 1$.

β	$r_{\rm H}$	$r_{\rm MB}$	$r_{\rm MS}$	$r_{\rm ph}$	$L_{\rm MB}$	$L_{\rm MS}$
0 (GR)	2.00	4.00	6.00	3.00	4.00	3.46
-0.1	2.15	4.30	6.45	3.20	4.15	3.61
-0.2	2.30	4.60	6.90	3.40	4.29	3.75
-0.5	2.74	5.52	8.28	3.98	4.71	4.15
-1.0	3.47	7.07	10.64	4.94	5.37	4.78
-1.5	4.18	8.66	13.08	5.89	5.99	5.37

4. Spherical accretion flow in $f(R)$ gravity

The vacuum solution for the $f(R)$ gravity can be used to understand the physics of black hole accretion. Here we explore the simplest class of it, i.e., spherical accretion flow around a black hole. In the Newtonian framework, such a flow is known as the Bondi accretion flow.[10] Since this flow is spherically symmetric, it does not possess any angular momentum, i.e., the flow is not rotating. The equations for conservation of energy and mass flux are respectively given by

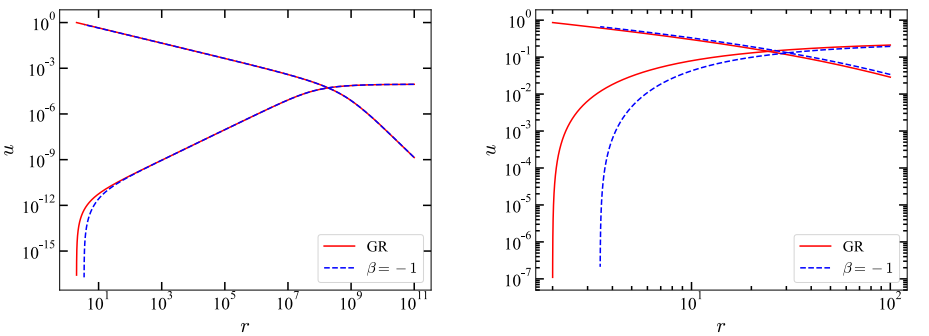
$$T^{\mu}_{0;\mu} = 0, \quad \text{and} \quad (\rho U^{\mu})_{;\mu} = 0, \tag{5}$$

where T^{μ}_{ν} is the stress-energy tensor, ρ is the density, and U^{μ} is the four-velocity of the fluid. Using these equations, for a spherically symmetric metric, $g_{\mu\nu} = \mathrm{diag}(-e^{2\phi(r)}, e^{2\lambda(r)}, r^2, r^2\sin^2\theta)$, the radial velocity ($u$) gradient equation for such flow is given by[8,11]

$$\frac{\mathrm{d}u}{u}\left[V^2 - \frac{u^2}{u^2+e^{-2\lambda}}\right] + \frac{\mathrm{d}r}{r}\left[2V^2 + r(V^2-1)(\phi'+\lambda') + \frac{r\lambda' e^{-2\lambda}}{u^2+e^{-2\lambda}}\right] = 0, \tag{6}$$

where $u = \mathrm{d}r/\mathrm{d}t$, $V^2 = 4T/3(1+4T)$ with T being the temperature of the fluid. The general idea of an accretion disc is that there is a Keplerian disc outside, where the matter has a minimal radial velocity. As the Keplerian disc terminates at a smaller distance from the black hole, the matter has a significant radial velocity, and it starts falling faster to the black hole. We assume that in this region, matter starts exhibiting spherical accretion flow. In other words, we assume that as the matter comes close enough to the black hole, it loses all its angular momentum, resulting in radial fall to the black hole (in reality, matter exhibits some angular momentum, even if sub-Keplerian). Figure 2 shows the accretion and wind flows for the spherical accretion flow in $f(R)$ gravity, which are obtained by solving the Equation (6). It is evident from the figure that the velocity is highly dependent on the outside temperature ($T_{\rm out}$), where matter starts behaving as a spherical accretion flow. The flow reaches the speed of light at the event horizon. As the event horizon shifts in

the presence of $f(R)$ gravity, the flow reaches this maximum speed at different radii for different values of β. However, since the solution is asymptotically flat, both the accretion and wind curves match the well-known case for the Schwarzschild solution far from the black hole.



(a) $T_{\text{out}} = 10^4\,\text{K}$ (b) $T_{\text{out}} = 10^{11}\,\text{K}$

Fig. 2. Spherical accretion flow in modified gravity: red solid line corresponds to the Schwarzschild spacetime and blue dashed line corresponds to the $f(R)$ gravity with $\beta = -1$.

5. Conclusions

$f(R)$ gravity is one of the best bets to replace GR, and researchers have already proposed different models to explain various exciting phenomena related to cosmology and compact objects. For more than a decade, various researchers have put effort into obtaining the vacuum solution of $f(R)$ gravity. However, none of those solutions are asymptotically flat, which means they do not reduce to the Minkowski metric as one moves very far from any gravity source, and hence they may not be applicable to astrophysical phenomena. In this article, we have argued that asymptotically flat vacuum spacetime is also possible in $f(R)$ gravity. We have shown here that the vacuum solution in $f(R)$ gravity is not the Schwarzschild solution, which indicates the violation of Birkhoff's theorem in the presence of the modified gravity. Since the solution is no longer the Schwarzschild solution, the quantities related to the event horizon and marginal orbits also alter, which further affects the spherical accretion flow around a black hole.

References

1. A. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Physics Letters B* **91**, 99 (March 1980).
2. A. V. Astashenok, S. Capozziello and S. D. Odintsov, Maximal neutron star mass and the resolution of the hyperon puzzle in modified gravity, *Phys. Rev. D* **89**, p. 103509 (May 2014).

3. S. Kalita and B. Mukhopadhyay, Modified Einstein's gravity to probe the sub- and super-Chandrasekhar limiting mass white dwarfs: a new perspective to unify under- and over-luminous type Ia supernovae, *J. Cosmology Astropart. Phys.* **9**, p. 007 (September 2018).
4. S. Kalita and B. Mukhopadhyay, Gravitational Wave in f(R) Gravity: Possible Signature of Sub- and Super-Chandrasekhar Limiting-mass White Dwarfs, *ApJ* **909**, p. 65 (March 2021).
5. T. Multamäki and I. Vilja, Spherically symmetric solutions of modified field equations in $f(R)$ theories of gravity, *Phys. Rev. D* **74**, p. 064022 (September 2006).
6. S. Capozziello, A. Stabile and A. Troisi, Spherically symmetric solutions in $f(R)$ gravity via the Noether symmetry approach, *Classical and Quantum Gravity* **24**, 2153 (April 2007).
7. S. Capozziello, M. De Laurentis and A. Stabile, Axially symmetric solutions in $f(R)$-gravity, *Classical and Quantum Gravity* **27**, p. 165008 (August 2010).
8. S. Kalita and B. Mukhopadhyay, Asymptotically flat vacuum solution in modified theory of Einstein's gravity, *European Physical Journal C* **79**, p. 877 (October 2019).
9. A. De Felice and S. Tsujikawa, $f(R)$ Theories, *Living Reviews in Relativity* **13**, p. 3 (June 2010).
10. H. Bondi, On spherically symmetrical accretion, *MNRAS* **112**, p. 195 (1952).
11. F. C. Michel, Accretion of Matter by Condensed Objects, *Ap&SS* **15**, 153 (January 1972).

Shadow of a charged black hole surrounded by an anisotropic matter field

Javier Badía[1,2,*] and Ernesto F. Eiroa[1,†]

[1] *Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), Casilla de Correo 67, Sucursal 28, 1428, Buenos Aires, Argentina*
[2] *Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria Pab. I, 1428, Buenos Aires, Argentina*
[*] *E-mail: jbadia@iafe.uba.ar*
[†] *E-mail: eiroa@iafe.uba.ar*

A certain type of matter with anisotropic pressures can add to the Reissner-Nordström metric a term proportional to a power of the radial coordinate. Using the standard method of separating variables for the Hamilton-Jacobi equation, we study the shadow of the corresponding rotating solution, obtained through the Newman-Janis algorithm. We define and calculate three observables in order to characterize the position, size and shape of the shadow.

Keywords: Black hole shadow; anisotropic fluid; electromagnetic field.

1. Introduction

There has been a renewed interest in the study of black hole shadows since the release of the first reconstructed image of the surroundings of the supermassive black hole M87* by the Event Horizon Telescope (EHT) collaboration.[1,2] While the shadow of a non-rotating black hole is a circle, a rotating black hole has an asymmetric shadow[3] whose shape depends on the black hole mass and spin, as well as the inclination angle of the observer. In alternative theories of gravity the shadow also presents characteristics that depend on the parameters of the specific model.[4]

Anisotropic fluids in general relativity have been studied in the context of compact objects such as stars or black holes, and in particular spherically symmetric black holes surrounded by an anistropic fluid have recently been introduced.[5] The anisotropy allows the matter around the black hole to remain static by means of a negative radial pressure, and the resulting metric has a very general form, also being found in some alternative theories of gravity.[6] In this work we study the shadows of the charged and rotating version[7] of these black holes. We adopt units such that $G = c = 1$.

2. Black hole solution

The energy-momentum tensor corresponding to the anisotropic fluid adopted by Cho & Kim[5] has the form

$$T_{\mu}{}^{\nu} = \mathrm{diag}(-\rho, p_1, p_2, p_2) \tag{1}$$

in spherical coordinates, with a radial pressure p_1 and an angular pressure p_2. A barotropic equation of state

$$p_i = w_i \rho \tag{2}$$

is assumed. It can be shown that in order to have a static black hole solution the value $w_1 = -1$ must be chosen, leaving w_2 as the only free parameter, which we will simply name w. It is also possible to add to the black hole an electric charge,[8] leading to the solution

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2), \tag{3}$$

with

$$f(r) = 1 - \frac{2m(r)}{r}, \tag{4}$$

$$m(r) = M - \frac{Q^2}{2r} + \frac{K}{2r^{2w-1}}, \tag{5}$$

where M, Q, and K are integration constants. M and Q are the mass and the electric charge of the black hole, respectively, while K is related to the energy density of the anisotropic fluid. It can be seen that by changing the equation of state of the fluid—that is, by changing w—one arrives at a different power of r in the last term of $f(r)$. The squared charge Q^2 is clearly positive; however, there is no impediment to continuing the solution to negative values, replacing Q^2 by a parameter q which may take either sign, and we will do so in the following. When q is negative the second term in $m(r)$ can no longer be interpreted as arising from an electric charge, but it can be found for example in some braneworld models.[9]

Applying the Newman-Janis algorithm to the spherically symmetric solution (3) leads to the corresponding axisymmetric metric

$$ds^2 = -\frac{\rho^2\Delta}{\Sigma}dt^2 + \frac{\Sigma\sin^2\theta}{\rho^2}(d\varphi - \Omega\, dt)^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2, \tag{6}$$

where

$$\rho^2 = r^2 + a^2\cos^2\theta, \tag{7}$$

$$\Delta = r^2 - 2m(r)r + a^2, \tag{8}$$

$$\Sigma = (r^2 + a^2)^2 - a^2\Delta\sin^2\theta, \tag{9}$$

$$\Omega = \frac{2am(r)r}{\Sigma}, \tag{10}$$

and

$$m(r) = M - \frac{q}{2r} + \frac{K}{2r^{2w-1}}, \tag{11}$$

as in the spherically symmetric case. When $K = 0$ the Kerr-Newman metric is recovered, with q replaced by Q^2. It is important to keep in mind that outside of general relativity, the metric obtained through the Newman-Janis algorithm may

correspond to a different energy-momentum tensor than that of the original spherically symmetric metric.[10] We will assume throughout this work that $w > 0$ so that the spacetime is asymptotically flat. The weak, strong and dominant energy conditions impose various restrictions[5,7] on the allowed values of w and K, but since they do not affect the calculation of the shadow we will not take them into account.

We require the presence of an event horizon, so that the spacetime does not contain a naked singularity. The event horizon is located at the largest solution of $\Delta(r) = 0$, so that its disappearance corresponds to parameter values for which $\Delta(r)$ has a double root. The values of w and K for which this happens can be found parametrically as functions of the radius r of the double root,[11] and are given by

$$w = \frac{a^2 + q - Mr}{\Delta_{\mathrm{KN}}} \tag{12}$$

and

$$K = \frac{\Delta_{\mathrm{KN}}}{r^{2r(r-M)/\Delta_{\mathrm{KN}}}}, \tag{13}$$

where $\Delta_{\mathrm{KN}} = r^2 - 2Mr + a^2 + q$ is the functional form of Δ in a Kerr-Newman-like spacetime. Plotting these curves for various values of $a^2 + q$ shows the regions in parameter space for which an event horizon exists, as seen in Figure 1, in which we have set $M = 1$ for simplicity. It can be seen from the figure that unlike for the Kerr-Newman spacetime, where $a^2 + q \leq M^2$ is a necessary condition for the existence of an event horizon, the presence of the fluid allows for black holes with $a^2 + q > M^2$.

3. The black hole shadow

We will briefly review the standard method for finding the null geodesics for photons and the shadow, adapted to our selected spacetime.[3,11] The metric (6) is independent of the t and φ coordinates, leading to the conserved quantities $E = -p_t$ and $L = p_\varphi$; however, there is an additional hidden symmetry with an associated conserved quantity, the Carter constant $\mathcal{Q}$. This constant can be found by assuming a separable solution for the Hamilton-Jacobi equation, leading to the first-order equations of motion

$$\rho^2 \dot{t} = \frac{r^2 + a^2}{\Delta} P(r) + aL - a^2 \sin^2\theta\, E, \tag{14}$$

$$\rho^2 \dot{\varphi} = \frac{aP(r)}{\Delta} + \frac{L}{\sin^2\theta} - aE, \tag{15}$$

$$(\rho^2 \dot{r})^2 = \mathcal{R}(r), \tag{16}$$

$$(\rho^2 \dot{\theta})^2 = \Theta(\theta), \tag{17}$$

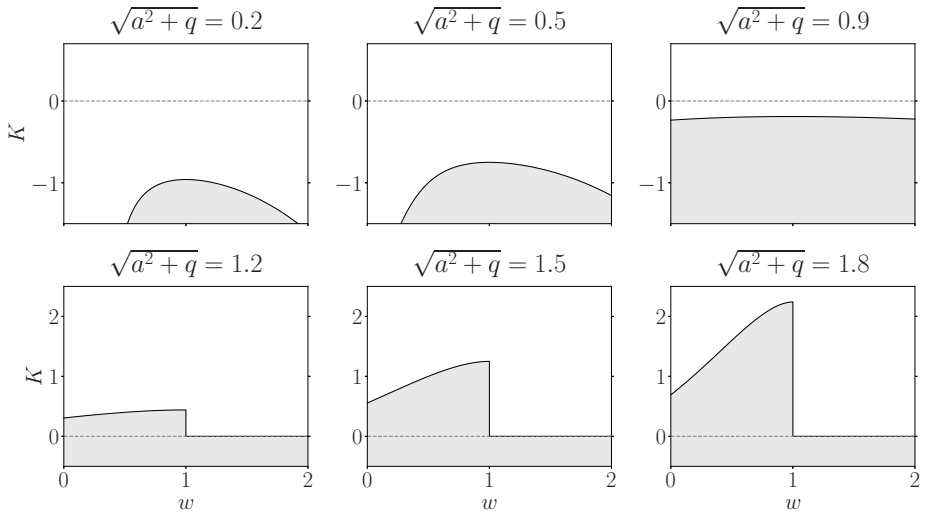


Fig. 1. Presence of an event horizon for our adopted metric. The shaded regions indicate the parameter values for which the spacetime contains a naked singularity, while in the white regions there is an event horizon. The solid curves separating them are described paremetrically by Eqs. (12) and (13). We have set $M = 1$, so that all parameters are dimensionless.

where

$$P(r) = E(r^2 + a^2) - aL, \tag{18}$$

$$\mathcal{R}(r) = P(r)^2 - \Delta[(L - aE)^2 + \mathcal{Q}], \tag{19}$$

$$\Theta(\theta) = \mathcal{Q} + \cos^2\theta \left(a^2E^2 - \frac{L^2}{\sin^2\theta}\right). \tag{20}$$

For convenience, we define the impact parameters $\xi = L/E$ and $\eta = \mathcal{Q}/E^2$. The trajectories that make up the shadow contour have the same impact parameters as the spherical photon orbits—that is, the solutions of the equations of motion that stay at a constant value of r. These can be found by solving $\mathcal{R}(r) = 0 = \mathcal{R}'(r)$, and the corresponding impact parameters are given parametrically by

$$\xi = \frac{4m(r)r^2 - (r + m(r) + m'(r)r)(r^2 + a^2)}{a(r - m(r) - m'(r)r)} \tag{21}$$

$$\eta = r^3 \frac{4a^2(m(r) - m'(r)r) - r(r - 3m(r) + m'(r)r)^2}{a^2(r - m(r) - m'(r)r)^2}. \tag{22}$$

A distant observer at an inclination angle $\theta = \theta_{\rm o}$ can use the celestial coordinates (α, β), which are given by

$$\alpha = -\frac{\xi}{\sin\theta_{\rm o}}, \tag{23}$$

$$\beta = \pm\sqrt{\eta + \cos^2\theta_{\rm o}\left(a^2 - \frac{\xi^2}{\sin^2\theta_{\rm o}}\right)}; \tag{24}$$

they correspond to horizontal and vertical displacement in the image plane, respectively.

Figure 2 shows the shadows produced by a black hole with spin $a/M = 0.9$ as seen by a distant equatorial observer, for various values of the parameters w, K and q. We find the expected behaviors from the Kerr and Kerr-Newman spacetimes: the shadow is asymmetrical and displaced as a consequence of the spin of the black hole, and its size decreases as the charge increases.

To better characterize the size and shape of the shadow, we define three observables[11,12] that can be calculated from a given shadow contour: the area of the shadow, its oblateness, and its horizontal displacement. The area is simply

$$A = 2\int \beta\, d\alpha, \tag{25}$$

where the plus sign in Eq. (24) has been chosen; the oblateness measures the deviation of the shadow from circularity and is defined as

$$D = \frac{\Delta\alpha}{\Delta\beta}, \tag{26}$$

where $\Delta\alpha$ and $\Delta\beta$ are the extent of the shadow in the horizontal and vertical directions respectively; finally, the horizontal displacement is more properly described as the α-coordinate of the centroid, given by

$$\alpha_c = \frac{1}{A}\int 2\alpha\beta\, d\alpha. \tag{27}$$

Working with observables has the benefit of making it easier to explore larger areas of parameter space at once; this can be seen in Figs. 3, 4 and 5, showing the values of A, D and α_c for $a/M = 0.9$ and various values of the other parameters. All plots are shown as functions of q, whose maximum value is determined by the requirement that there exist an event horizon.

It is clear from Fig. 3 that the shadow becomes smaller as q increases or as K decreases. This is to be expected, since both parameters appear with opposite signs in the metric. The oblateness and horizontal displacement show more interesting features. For $w < 1$ we see that the shadow becomes more compressed and more displaced as the charge increases. For $w = 2$ and $K > 0$, though, the charge can become large enough to reach a region where the behavior reverses, with D becoming greater than one and α_c becoming greater than zero, reflecting a shadow

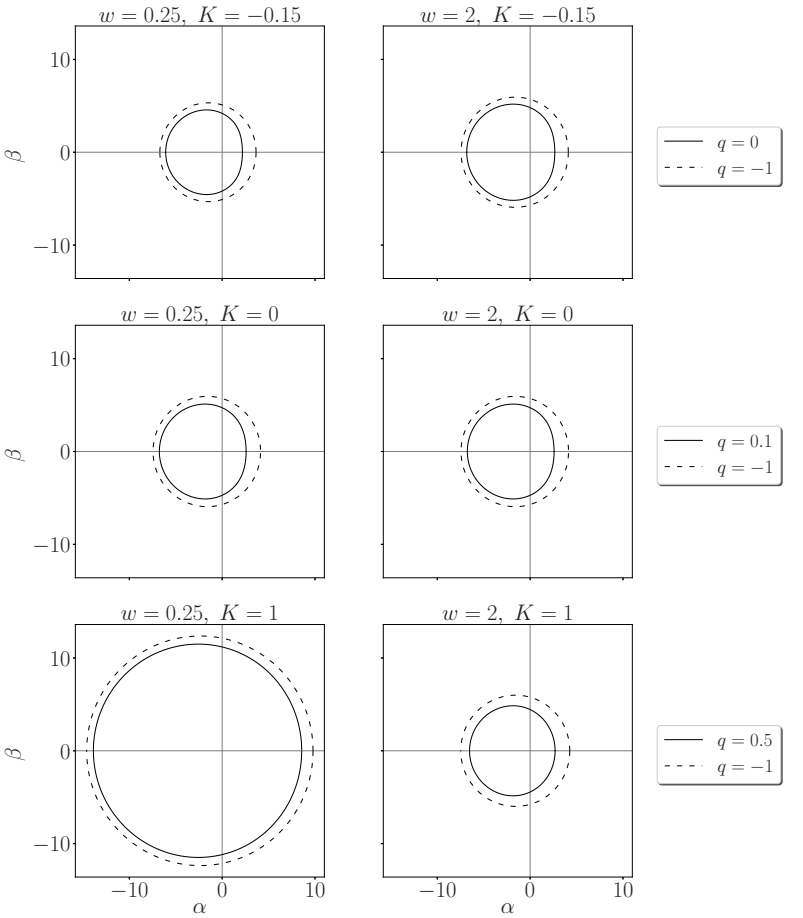


Fig. 2. The shadows of a black hole with $a/M = 0.9$ for an observer at $\theta_{\rm o} = \pi/2$, for various values of the parameters. All quantities have been adimensionalized by setting $M = 1$.

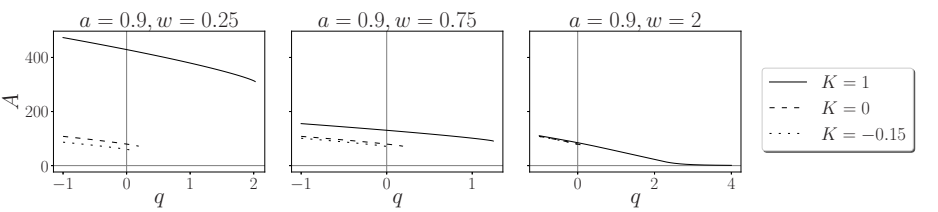


Fig. 3. The area of the black hole shadow for some values of the parameters. We set $M = 1$, so that all quantities are dimensionless.

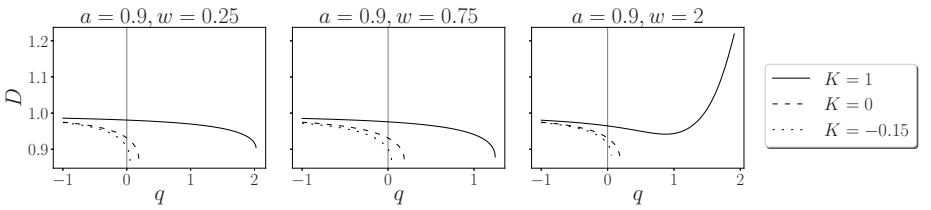


Fig. 4. The oblateness of the black hole shadow for some values of the parameters. We set $M = 1$.

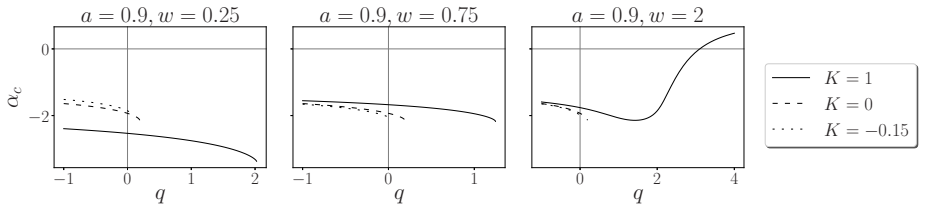


Fig. 5. The horizontal position of the shadow centroid for some values of the parameters. We set $M = 1$.

that is wider than it is tall—unlike the Kerr shadow—and displaced in the opposite direction. However, the energy-momentum tensor of the fluid can have the physically undesirable property of having negative energy[7, 11] in this region of parameter space, so the corresponding shadows should be interpreted with care.

4. Discussion

In this work we have considered the effect of a fluid with anisotropic pressure on the shadow of a charged and rotating black hole. The fluid makes a contribution to the gravitational field of the black hole, and the resulting Hamilton-Jacobi equation for the geodesics is separable, so that a standard method for finding the shadow contour can be used. We have produced plots of the shadow for various values of the charge as well as the fluid anisotropy and density, and we have found that for many values of the parameters the shadow is qualitatively similar to the well known Kerr-Newman case, exhibiting the typical deviation from circularity due to the spin and shrinking as the charge increases or the energy density of the fluid decreases. However, in some regions of parameter space where the fluid has negative energy, the opposite behavior is seen, with the shadow becoming narrow and displaced in the opposite direction.

Acknowledgments

This work has been supported by CONICET and Universidad de Buenos Aires.

References

1. K. Akiyama *et al.* (Event Horizon Telescope), First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, L1 (2019).
2. K. Akiyama *et al.* (Event Horizon Telescope), First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring, *Astrophys. J. Lett.* **875**, L5 (2019).
3. J. Bardeen, *Black Holes*, Proceedings of École d'été de Physique Théorique, Les Houches 1972, 215-240, edited by C. De Witt and B.S. De Witt (Gordon and Breach Science Publishers, New York, 1973).
4. V. Perlick and O. Y. Tsupko, Calculating black hole shadows: review of analytical studies, arXiv:2105.07101 [gr-qc]
5. I. Cho and H. Kim, Simple black holes with anisotropic fluid, *Chinese Phys. C* **43**, 025101 (2018).
6. R. Kumar, S. G. Ghosh, and A. Wang, Gravitational deflection of light and shadow cast by rotating Kalb-Ramond black holes, *Phys. Rev. D* **101**, 104001 (2020).
7. H. Kim, B. Lee, W. Lee, and Y. Lee, Rotating black holes with an anisotropic matter field, *Phys. Rev. D* **101**, 064067 (2020).
8. V. V. Kiselev, Quintessence and black holes, *Class. Quant. Grav.* **20**, 1187 (2003).
9. A. N. Aliev and A. E. Gümrükçüoglu, Charged rotating black holes on a 3-brane, *Phys. Rev. D* **71**, 104027 (2005).
10. D. Hansen and N. Yunes, Applicability of the Newman-Janis algorithm to black hole solutions of modified gravity theories, *Phys. Rev. D* **88**, 104020 (2013).
11. J. Badía and E. F. Eiroa, Influence of an anisotropic matter field on the shadow of a rotating black hole, *Phys. Rev. D* **102**, 024066 (2020).
12. R. Kumar and S. G. Ghosh, Black Hole Parameter Estimation from Its Shadow, *Astrophys. J.* **892**, 78 (2020).

Constraining modified gravity theories with physical black holes

Sebastian Murk

Department of Physics and Astronomy, Macquarie University,
Sydney, New South Wales 2109, Australia
and
Sydney Quantum Academy,
Sydney, New South Wales 2006, Australia
E-mail: sebastian.murk@mq.edu.au

We review the constraints modified theories of gravity must satisfy to be compatible with the spherically symmetric black hole solutions of semiclassical gravity that describe the formation of an apparent horizon in finite time of a distant observer. The constraints are satisfied in generic modified gravity theories with up to fourth-order derivatives in the metric, indicating that the semiclassical solutions correspond to zeroth-order terms in perturbative solutions of these theories. As a result, it may not be possible to distinguish between the semiclassical theory and modifications including up to fourth-order derivatives based on the observation of an apparent horizon alone.

Keywords: Black holes; general relativity; modified gravity; alternative theories of gravity; quantum aspects of black holes.

1. Introduction

The predictions of general relativity (GR) have been confirmed in numerous experiments and are so far compatible with all currently available astrophysical and cosmological data. In particular, strong evidence for the existence of astrophysical black holes (ABHs) — massive dark compact objects — has accumulated over the last few decades. The precise nature of ABHs is still under debate, but contemporary models describe them as ultra-compact objects (UCOs) with or without a horizon.[1] Due to their compactness, they provide excellent opportunities to probe strong gravity.[2]

The development of alternative theories of gravity has received much attention in recent years in an attempt to alleviate some of the perceived shortcomings of GR (such as the presence of singularities) and incorporate quantum gravitational effects.[3,4] As a preliminary test, any modified theory of gravity (MTG) should be compatible with the observed ABHs candidates and provide a model to describe them. Present-day astrophysical observations such as the detection of gravitational waves from coalescing compact systems[5,6] and measurements of the M87 black hole shadow[7] probe the (dynamic) strong-field regime of gravity and limit the extent to which MTG can deviate from GR while still being compatible with observational data.

There is no single, unanimously agreed upon definition of what exactly constitutes a black hole as different lines of research focus on different aspects. Consequently, the most useful and/or practical definition of a black hole depends on the context and highlights the relevant features. Nevertheless, a widely accepted feature of black holes is the presence of a trapped spacetime region from which nothing can escape.[8] We adapt the terminology of Ref. 9 and refer to a trapped region that is bounded by an apparent horizon as a physical black hole (PBH). The definition of an apparent horizon (see Sec. 2) is an observer-dependent notion that is, at least in principle, physically observable (in contrast to the global, observer-independent notion of an event horizon).[10] It is therefore possible that GR and various MTG can be distinguished observationally based on distinct properties of PBHs and their horizons in these models.

2. Mathematical preliminaries

We work in units where $\hbar = c = G = 1$, use the $(-+++)$ signature of the metric, and describe dynamics using the semiclassical Einstein equations $G_{\mu\nu} = 8\pi T_{\mu\nu}$ or modifications thereof, where $T_{\mu\nu} \equiv \langle \hat{T}_{\mu\nu} \rangle_\omega$ on the rhs denotes the expectation value of the renormalized energy-momentum tensor (EMT) that describes both the collapsing matter and the produced excitations. No assumptions are made about the matter content or the quantum state ω.

The field equations of a generally covariant metric theory of gravity can be derived from its action.[11] Here, we omit discussions of the matter Lagrangian and boundary conditions and focus on the gravitational action

$$\mathcal{S} = \int \sqrt{-g}\, \mathcal{L}_{\mathrm{g}}\, d^4x \tag{1}$$

determined by the gravitational Lagrangian density $\mathcal{L}_{\mathrm{g}}(g^{\mu\nu}, R_{\mu\nu\rho\sigma})$. In classical GR and semiclassical gravity, it is strictly linear in the Ricci scalar, i.e. $\mathcal{L}_{\mathrm{g}}(g^{\mu\nu}, R_{\mu\nu\rho\sigma}) = R$. Modifications typically include higher-order curvature corrections, which we organize according to powers of derivatives in the metric, i.e.

$$\begin{aligned} \mathcal{L}_{\mathrm{g}}\sqrt{-g} &= \frac{{M_{\mathrm{Pl}}}^2}{16\pi}\big(R + \lambda\mathcal{F}(g^{\mu\nu}, R_{\mu\nu\rho\sigma})\big) \\ &= \frac{{M_{\mathrm{Pl}}}^2}{16\pi}R + a_1 R^2 + a_2 R_{\mu\nu}R^{\mu\nu} + a_3 R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} + \ldots, \end{aligned} \tag{2}$$

where λ sets the scale of the perturbative analysis, the cosmological constant term was omitted, M_{P} denotes the Planck mass, and the coefficients a_1, a_2, a_3 are dimensionless.

We limit our exposition to spherical symmetry and work with a general spherically symmetric metric in Schwarzschild coordinates given by

$$ds^2 = -e^{2h(t,r)}f(t,r)dt^2 + f(t,r)^{-1}dr^2 + r^2 d\Omega, \tag{3}$$

where

$$f(t,r) := \partial_\mu r \partial^\mu r \equiv 1 - C(t,r)/r \tag{4}$$

provides a coordinate-independent definition of the Misner–Sharp mass[12] $C(t,r)/2$, r denotes the areal radius, and the function $h(t,r)$ plays the role of an integrating factor in coordinate transformations, such as

$$dt = e^{-h}\left(e^{h_+}dv - f^{-1}dr\right). \tag{5}$$

Explicit calculations can be simplified by introducing the effective EMT components

$$\tau_t := e^{-2h}T_{tt}\,, \qquad \tau_t{}^r := e^{-h}T_t{}^r, \qquad \tau^r := T^{rr}. \tag{6}$$

The apparent horizon plays a central role in our analysis. Its outer component is located at the Schwarzschild radius $r_g(t)$ that is defined as the largest root of $f(t,r) = 0$.[13] Albeit an observer-dependent notion in general, it is unambiguously defined in spherical symmetry. To be of physical relevance, it must form in finite time according to the clock of a distant observer.[14,15] Moreover, to maintain predictability of the theory, it must be a regular surface,[16,17] i.e. the curvature scalars

$$\mathrm{T} := T^\mu{}_\mu = -R/8\pi + \mathcal{O}(\lambda), \qquad \mathfrak{T} := T^{\mu\nu}T_{\mu\nu} = R^{\mu\nu}R_{\mu\nu}/64\pi^2 + \mathcal{O}(\lambda^2), \tag{7}$$

must be finite at $r = r_g$. Mathematically, this requirement is expressed as

$$\mathrm{T} = (\tau^r - \tau_t)/f \;\to\; g_1(t)f^{k_1}, \tag{8}$$

$$\mathfrak{T} = \left[(\tau_t)^2 - 2(\tau_t{}^r)^2 + (\tau^r)^2\right]/f^2 \;\to\; g_2(t)f^{k_2}, \tag{9}$$

for some functions $g_{1,2}(t)$ and $k_{1,2} \geqslant 0$. While there a infinitely many solutions in principle, only two distinct classes of dynamic solutions that satisfy the regularity requirement were found to be self-consistent.[18,19] With respect to the scaling behavior $\tau \sim f^k$ of the effective EMT components of Eq. (6), they correspond to the values $k \in \{0,1\}$, where $k \equiv k_1 \equiv k_2$. Both of them violate the null energy condition (NEC) in the vicinity of the outer apparent horizon.[15,18,20] We briefly summarize their properties in what follows.

3. Semiclassical PBH solutions in spherical symmetry

In spherical symmetry, the relevant components G_{tt}, $G_t{}^r$, G^{rr} of the semiclassical Einstein equations can then be written as

$$\partial_r C = 8\pi r^2 \tau_t/f, \tag{10}$$

$$\partial_t C = 8\pi r^2 e^h \tau_t{}^r, \tag{11}$$

$$\partial_r h = 4\pi r(\tau_t + \tau^r)/f^2, \tag{12}$$

respectively.

3.1. $k = 0$ *solutions*

Close to the horizon (i.e. as $r \to r_g$), the limiting form of the effective EMT components is given by

$$\tau_t = -\Upsilon^2 + \sum_{j\geqslant\frac{1}{2}}^{\infty} e_j x^j, \qquad {\tau_t}^r = \pm\Upsilon^2 + \sum_{j\geqslant\frac{1}{2}}^{\infty} \phi_j x^j, \qquad \tau^r = -\Upsilon^2 + \sum_{j\geqslant\frac{1}{2}}^{\infty} p_j x^j \quad (13)$$

for some time-dependent function $\Upsilon(t) > 0$, where $j \in \mathbb{Z}\frac{1}{2}$ labels half-integer and integer coefficients and powers of the coordinate distance $x \coloneqq r - r_g$ from the apparent horizon, and the lower signature of ${\tau_t}^r$ corresponds to evaporation (accretion). Taking $\tau_t \to \tau^r \to +\Upsilon^2 + \mathcal{O}(\sqrt{x})$ results in complex-valued solutions of the Einstein equations Eqs. (10)–(12).[20] The leading terms of the metric functions that solve Eqs. (10) and (12) are given by

$$C = r_g - c_{12}\sqrt{x} + \sum_{j\geqslant 1}^{\infty} c_j x^j, \tag{14}$$

$$h = -\frac{1}{2}\ln\frac{x}{\xi} + \sum_{j\geqslant\frac{1}{2}}^{\infty} h_j x^j, \tag{15}$$

where the leading coefficient $c_{12} = 4\sqrt{\pi} r_g^{3/2}\Upsilon$, $\xi(t)$ is determined by the choice of time variable, and higher-order terms must be matched with those of the EMT expansion. Eq. (11) then necessitates that the dynamic behavior of the horizon is governed by

$$r_g' = \pm c_{12}\sqrt{\xi}/r_g \tag{16}$$

where the lower (upper) sign corresponds to evaporation (accretion).

3.2. $k = 1$ *solution*

For $k = 1$, the limiting form of the effective EMT components close to the horizon (i.e. as $r \to r_g$) is given by

$$\tau_t = Ef + \sum_{j\geqslant 2}^{\infty} e_j x^j, \qquad {\tau_t}^r = \Phi f + \sum_{j\geqslant 2}^{\infty} \phi_j x^j, \qquad \tau^r = Pf + \sum_{j\geqslant 2}^{\infty} p_j x^j, \tag{17}$$

where E, Φ, and P denote energy density, flux, and pressure at the horizon. The metric functions that solve Eqs. (10) and (12) are given by

$$C = r - c_{32}x^{3/2} + \mathcal{O}(x^2), \tag{18}$$

$$h = -\frac{3}{2}\ln\frac{x}{\xi} + \mathcal{O}(\sqrt{x}), \tag{19}$$

where the leading coefficient $c_{32} = 4r_g^{3/2}\sqrt{-\pi e_2/3}$, and e_2 denotes the $\mathcal{O}(x^2)$ coefficient of τ_t, see Eq. (17). According to Eq. (11), the dynamic behavior of the

horizon is governed by

$$r_g' = \pm c_{32}\xi^{3/2}/r_g. \tag{20}$$

where the lower (upper) sign corresponds to evaporation (accretion). For $k = 1$, only one solution is self-consistent. It describes PBHs at the instant of their formation[19] and is uniquely defined by the energy density $E := -T^t_{\ t}$ and pressure $P := T^r_{\ r}$ at the horizon, which take on their maximal possible values $E = -P = 1/(8\pi r_g^2)$. Hence, we refer to it as extreme-valued $k = 1$ solution.

4. Modified Einstein equations in spherical symmetry

The modified Einstein equations are obtained through variation of the gravitational action[11] [cf. Eqs. (1)–(2)], and can be represented schematically as

$$G_{\mu\nu} + \lambda\mathcal{E}_{\mu\nu} = 8\pi T_{\mu\nu}\,, \tag{21}$$

where the terms $\mathcal{E}_{\mu\nu}$ denote deviations from the Einstein equations that result from the variation of $\mathcal{F}(g^{\mu\nu}, R_{\mu\nu\rho\sigma})$ [cf. Eq. (2)]. The explicit form of the Einstein equations in MTG then immediately follows from Eqs. (10)–(12):

$$fr^{-2}e^{2h}\partial_r C + \lambda\mathcal{E}_{tt} = 8\pi T_{tt}\,, \tag{22}$$

$$r^{-2}\partial_t C + \lambda\mathcal{E}_t^{\ r} = 8\pi T_t^{\ r}, \tag{23}$$

$$2f^2 r^{-1}\partial_r h - fr^{-2}\partial_r C + \lambda\mathcal{E}^{rr} = 8\pi T^{rr}. \tag{24}$$

The metric functions that solve the modified Einstein equations Eq. (21) are

$$C_\lambda =: \bar{C}(t,r) + \lambda\Sigma(t,r), \tag{25}$$

$$h_\lambda =: \bar{h}(t,r) + \lambda\Omega(t,r), \tag{26}$$

where Σ and Ω denote the perturbative corrections, and the bar labels the semiclassical metric functions introduced in Sec. 3. Schematically, substitution of Eqs. (25)–(26) into the modified Einstein equations Eq. (21) yields an equation of the form

$$\bar{G}_{\mu\nu} + \lambda\tilde{G}_{\mu\nu} + \lambda\bar{\mathcal{E}}_{\mu\nu} = 8\pi\left(\bar{T}_{\mu\nu} + \lambda\tilde{T}_{\mu\nu}\right), \tag{27}$$

where objects labeled by the tilde correspond to first-order terms in the expansion of λ that involve the perturbative corrections Σ and Ω, and terms of order $\mathcal{O}(\lambda^2)$ and higher have been omitted. The modified gravity contributions can be expressed solely in terms of unperturbed quantities, i.e. $\bar{\mathcal{E}}_{\mu\nu} \equiv \mathcal{E}_{\mu\nu}(\bar{C}, \bar{h})$. For a detailed account of the perturbative analysis, the reader is referred to Ref. 21.

5. Constraints for modified gravity theories

Any arbitrary MTG must satisfy several constraints[21] to be compatible with the spherically symmetric PBH solutions of semiclassical gravity described in Sec. 3. The constraints are twofold: first, the terms $\bar{\mathcal{E}}_{\mu\nu}$ that encode deviations from the semiclassical theory must follow a particular structure when expanded in terms of

the coordinate distance $x := r - r_g$ from the apparent horizon. Second, several identities between their coefficients must hold.

To be compatible with PBH solutions of the $k = 0$ type (Sec. 3.1), the MTG terms $\bar{\mathcal{E}}_{\mu\nu}$ must conform to the expansion structures prescribed by

$$\bar{\mathcal{E}}_{tt} = \frac{æ_{\bar{1}}}{x} + \frac{æ_{\overline{12}}}{\sqrt{x}} + æ_0 + \sum_{j\geqslant\frac{1}{2}}^{\infty} æ_j x^j, \tag{28}$$

$$\bar{\mathcal{E}}_t{}^r = \frac{œ_{\overline{12}}}{\sqrt{x}} + œ_0 + \sum_{j\geqslant\frac{1}{2}}^{\infty} œ_j x^j, \tag{29}$$

$$\bar{\mathcal{E}}^{rr} = ø_0 + \sum_{j\geqslant\frac{1}{2}}^{\infty} ø_j x^j, \tag{30}$$

and the three identities

$$æ_{\bar{1}} = \sqrt{\bar{\xi}} œ_{\overline{12}} = \bar{\xi} ø_0, \qquad æ_{\overline{12}} = 2\sqrt{\xi} œ_0 - \xi ø_{12}, \tag{31}$$

between their lowest-order coefficients must be satisfied.

Similarly, to be compatible with the extreme-valued $k = 1$ solution (Sec. 3.2), the MTG terms must follow the expansion structures prescribed by

$$\bar{\mathcal{E}}_{tt} = \frac{æ_{\overline{32}}}{x^{3/2}} + \frac{æ_{\bar{1}}}{x} + \frac{æ_{\overline{12}}}{\sqrt{x}} + æ_0 + \sum_{j\geqslant\frac{1}{2}}^{\infty} æ_j x^j, \tag{32}$$

$$\bar{\mathcal{E}}_t{}^r = œ_0 + \sum_{j\geqslant\frac{1}{2}}^{\infty} œ_j x^j, \tag{33}$$

$$\bar{\mathcal{E}}^{rr} = \sum_{j\geqslant\frac{3}{2}}^{\infty} ø_j x^j, \tag{34}$$

and the two identities

$$æ_{\overline{32}} = 2\bar{\xi}^{3/2} œ_0 - \bar{\xi}^3 ø_{32}, \qquad æ_{\bar{1}} = 2\bar{\xi}^{3/2}\left(h_{12} œ_0 + œ_{12}\right) - \bar{\xi}^3\left(2h_{12} ø_{32} + ø_2\right). \tag{35}$$

between their lowest-order coefficients must hold.

For both types of solutions $k \in \{0, 1\}$, no additional constraints can be obtained through the consideration of higher-order terms ($\geqslant \mathcal{O}(x^{3/2})$) in the metric functions as in this case the modified Einstein equations contain too many additional independent variables. A comprehensive analysis of this result and a detailed derivation of Eqs. (28)–(35) is provided in Ref. 21.

If a particular MTG does not satisfy the constraints for one or both classes of semiclassical PBHs, the theory in question may still possess solutions corresponding to PBHs, but their underlying mathematical description must then differ from that of Sec. 3.1 and/or Sec. 3.2, which may or may not lead to observable differences based on properties of the associated near-horizon geometry.

So far, the constraints imposed by Eqs. (28)–(35) have been investigated in MTG with up to fourth-order derivatives in the metric.[22] In particular, it has been demonstrated that

(1) generic $\mathfrak{f}(R)$ theories, where $\mathcal{L}_{\rm g} = \mathfrak{f}(R)$, $\mathfrak{f}(R) =: R + \lambda\mathcal{F}(R)$ [cf. Eq. (2)], $\mathcal{F}(R) = \beta R^q$, and $\beta, q \in \mathbb{R}$
(2) generic MTG with up to fourth-order derivatives in the metric, where the particular choice of $\mathcal{L}_{\rm g} = (-\alpha R_{\mu\nu}R^{\mu\nu} + \beta R^2 + \gamma\kappa^{-2}R)$ in the action is motivated by the desire to avoid ghosts (massive states of negative norm that result in an apparent lack of unitarity),[23,24] and $\kappa^2 = 32\pi$

satisfy all of the constrains identically (i.e. without requiring any additional conditions) and are thus compatible with the existence of spherically symmetric semiclassical PBHs. The procedure for obtaining explicit expressions for the MTG terms $\bar{\mathcal{E}}_{\mu\nu}$ in terms of unperturbed quantities is outlined in Ref. 22, and explicit expressions for the relevant MTG coefficients of Eqs. (31) and (35) in these two theories are provided in the following linked Github repository.

6. Discussion

We have outlined a procedure to test the compatibility of arbitrary metric MTG with the PBH solutions of semiclassical gravity, which are characterized by the formation of a regular apparent horizon in finite time of a distant observer. We find that generic MTG with up to fourth-order derivatives in the metric identically satisfy all of the constraints for both classes of solutions, indicating that the semiclassical PBHs correspond to zeroth-order terms in perturbative solutions of these models. As a result, the observation of an apparent horizon by itself is insufficient to distinguish between the semiclassical theory and modifications including up to fourth-order derivatives in the metric. A detailed analysis of the response of the near-horizon geometry to perturbations is required, and may allow to identify potentially observable differences between PBHs of various theoretical frameworks.

Acknowledgments

I would like to thank Daniel Terno for useful discussions and helpful comments. This work was supported by an International Macquarie University Research Excellence Scholarship and a Sydney Quantum Academy Scholarship.

References

1. L. Barack, V. Cardoso, S. Nissanke, and T. P. Sotiriou (eds.), Class. Quantum Gravity **36**, 143001 (2019).
2. V. Cardoso and P. Pani, Living Rev. Relativ. **22**, 4 (2019).
3. C. P. Burgess, Living Rev. Relativ. **7**, 5 (2004).
4. J. F. Donoghue and B. R. Holstein, J. Phys. G: Nucl. Part. Phys. **42**, 103102 (2015).

5. LIGO Scientific Collaboration and Virgo Collaboration, Phys. Rev. Lett. **116**, 061102 (2016).
6. LIGO Scientific Collaboration and Virgo Collaboration, Astrophys. J. Lett. **882**, L24 (2019).
7. Event Horizon Telescope Collaboration, Phys. Rev. Lett. **125**, 141104 (2020).
8. E. Curiel, Nat. Astr. **3**, 27 (2019).
9. V. P. Frolov, J. High Energy Phys. **05**, 049 (2014).
10. M. Visser, Phys. Rev. D **90**, 127502 (2014).
11. T. Padmanabhan, Phys. Rev. D **84**, 124041 (2011).
12. C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).
13. V. Faraoni, G. F. R. Ellis, J. T. Firouzjaee, A. Helou, and I. Musco, Phys. Rev. D **95**, 024008 (2017).
14. S. Murk and D. R. Terno, arXiv:2110.12761 (2021).
15. P. K. Dahal, S. Murk, and D. R. Terno, arXiv:2110.00722 (2021).
16. B.-L. B. Hu and E. Verdaguer, *Semiclassical and Stochastic Gravity: Quantum Field Effects on Curved Spacetime* (Cambridge University Press, 2020).
17. L. Parker and D. Toms, *Quantum Field Theory in Curved Spacetime* (Cambridge University Press, 2009).
18. D. R. Terno, Phys. Rev. D **101**, 124053 (2020).
19. S. Murk and D. R. Terno, Phys. Rev. D **103**, 064082 (2021).
20. V. Baccetti, R. B. Mann, S. Murk, and D. R. Terno, Phys. Rev. D **99**, 124014 (2019).
21. S. Murk and D. R. Terno, Phys. Rev. D **104**, 064048 (2021).
22. S. Murk, arXiv:2110.14973 (2021).
23. K. S. Stelle, Gen. Relativ. Gravit. **9**, 353 (1978).
24. T. P. Sotiriou and V. Faraoni, Rev. Mod. Phys. **82**, 451 (2010).

Penrose suggestion as to pre-Planck-era black holes showing up in present universe data sets discussed, with a possible candidate as to GW radiation which may provide initial CMBR data

A. W. Beckwith

Physics Department, College of Physics, Chongqing University, City, State ZIP/Zone, People's Republic of China
E-mail: rwill9955b@gmail.com, abeckwith@uh.edu

What we are doing is three-fold. First, we examine the gist of the Penrose suggestion as to signals from a prior universe showing up in the CMBR. That is, this shows up as data in the CMBR. Second, we give a suggestion as to how super massive black holes could be broken up s of a prior universe cycle by pre-big-bang conditions, with say millions of pre-Planck black holes coming up out of a breakup of prior universe black holes. Three, we utilize a discussion as to Bose–Einstein condensates set as gravitons as to composing the early universe black holes. The BEC formulation gives a number N of gravitons, linked to entropy, per black hole, which could lead to contributions to the alleged CMBR perturbations, which were identified by Penrose et al.

Keywords: Minimum scale factor, Cosmological constant, Space-time bubble, Penrose singularity, Prior universe black holes.

1. First, What Does Penrose Suggest About the CMBR Data Set and Preuniverse Massive Black Holes?

The abstract has a clue, as part of Ref. 1 states as to what we want to explain in the CMBR, i.e., circular rings in the CMBR "data."

In Ref. 1 there is a well crafted suggestion by Gurzadyan and Penrose as to an initial quote:

> The significance of individual low-variance circles in the true data has been disputed; yet a recent independent analysis has confirmed CCC's expectation that CMB circles have a non-Gaussian temperature distribution. Here we examine concentric sets of low-variance circular rings in the WMAP data, finding a highly nonisotropic distribution.

Here is the nuts and bolts as to what Penrose cosmology is about Ref. 2.

1.1. *There is initial inflationary expansion of the universe, but the caveat is that matter–energy is sucked up in super-massive black holes*

That is, rather than have a purported infinite expansion, and we see the following dynamic. We connect a countable sequence of open Friedmann–Lemaître–Robertson–Walker metric (FLRW) spacetimes, each representing a big bang followed by an

infinite future expansion. Penrose noticed that the past conformal boundary of one copy of FLRW spacetime can be "attached" to the future conformal boundary of another, after an appropriate conformal rescaling. result g_{ab} reset in a conformal reset with matter from black holes collected and reset to a new value of g_{ab} at the start of cosmological expansion with matter–energy from black holes being recycled conformally to a new expansion cycle.[2]

1.2. *Next, let us view the Penrose suggestion as to black holes from a prior universe*

In order to see this, consider a suggestion as to black holes, being the template for a start to the present universe,[2] and also Ref. 3 which has the Penrose suggestion of an imprint of a prior universe black holes having an effect upon the CMBR spectrum. The CMBR spectrum is a real datum, but the worth of getting this information would be in terms of having what was said in Ref. 3 as to the "ghost" of prior universe black hole radiation. To get a glimpse of where this is going the author invites readers to look at Ref. 4 as to the cosmic maelstrom such "signals" would have to pass through.

> Figure 1 shows a conformal diagram representing the effect of a highly energetic event occurring at the space-time point H. In CCC, H is taken to be a Hawking point, where virtually the entire Hawking radiation of a previous-aeon supermassive black hole is concentrated at H by the conformal compression of the hole's radiating future. The horizontal line at the bottom stands for the crossover surface dividing the previous cosmic aeon from our own and describes our conformally stretched big bang. In conventional inflationary cosmology, X would represent the graceful exit turn-off of inflation. In each case, the future light cone of H represents the outer causal boundary of physical effects initiated at H, and such effects can reach D only within the roughly 0.08 radian spread indicated at the top of the diagram.[3]

2. What Can We Expect from the Transition from a Prior Universe to the Planckian Regime of Micro Black Holes? A Transition from Initially Gigantic Black Holes to Micro Black Holes

2.1. *In a word, we would likely have in the prior universe a massive black hole, which would be broken up into millions (billions?) of Planck-sized black holes*

In a word the GW radiation and thermal/photonic input would have to fight through a thicket of pairs of micro black holes which would be in binary configuration generating their OWN GW background.

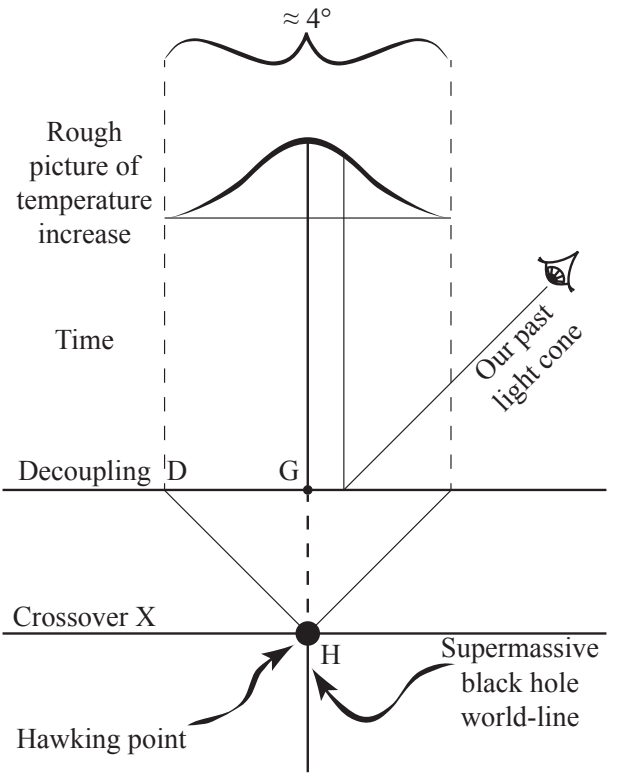


Fig. 1. Competing black hole radiation, and can we see this today in the CMBR?

We first will discuss this "binary black holes" signal background which the Planckian early universe stars would have to impinge upon, in order to come to our attention.

Now for the discussion of the millions (more than that) of micro-sized black hole pairs which would create a generalized GW signature.

To evaluate the above in terms of our model, we need to refer to a formula given in Ref. 4, 18, on page 16 of that document which reads as a change in power from rotating Planck-sized black holes separated say by a Planck length.

$$\dot{E} = GW - (\Delta \text{ in energy}) = \frac{32(M_1 M_2)^2 (M_1 + M_2)}{5R^2 M_{\text{Planck}}^5} \xrightarrow[M_1 = M_2 = M_{\text{Planck}}]{} \frac{64}{5R_{\text{Planck}}^2}$$
$$\equiv (\Delta \text{ in power from rotating binary black holes}). \tag{1}$$

For M about the size of a Planck-sized black hole, it likely would fade out almost too quickly to be very measurable.

2.2. *We also can consider the following gravitons as a Bose–Einstein condensate in low mass black holes, and its relevance to signal propagation*

This is a way to get measurable GW signals from a black hole, which have a chance of being detected.

We will be looking at Ref. 5, specifically page 181, where we have the following scaling arguments to work through, if gravitons are Bose–Einstein condensates (BECs) for small black holes. The following are scaling value to consider, if we want BEC.

Why we consider BECs and Eq. (2)–Eq. (6), i.e., if there is a break up of massive black holes into say Planck mass-sized black holes, as or about the Planck era, very likely will not have a surviving signal which has a chance of being measurable in the CMBR data. That is, the discussion of Eq. (2)–Eq. (6) below uses the device of having BEC condensation in gravitons for masses up to about 10 grams or so, and in doing so a dodge as to getting entropy counts per black hole.

That is after the black hole masses, as given in Eq. (2)–Eq. (6) are likely built up by the consolidation of two mini black holes going through an inspiral collapse, as has been modeled in GW.

$$m \approx \frac{M_{\mathrm{P}}}{\sqrt{N_{\mathrm{graviton}}}} \tag{2}$$

$$M_{\mathrm{BH}} \approx \sqrt{N_{\mathrm{graviton}}} \cdot M_{\mathrm{P}} \tag{3}$$

$$R_{\mathrm{BH}} \approx \sqrt{N_{\mathrm{graviton}}} \cdot l_{\mathrm{P}} \tag{4}$$

$$S_{\mathrm{BH}} \approx k_B N_{\mathrm{graviton}} \tag{5}$$

$$T_{\mathrm{BH}} \approx \frac{T_{\mathrm{P}}}{\sqrt{N_{\mathrm{graviton}}}}. \tag{6}$$

Here, the first term, m, is in the effective mass of a graviton. This is my take as to how to make all this commensurate as to special relativity.

$$m \approx \frac{m_g}{\sqrt{1-\left(\frac{\nu_g}{c}\right)^2}} \approx \frac{M_{\mathrm{P}}}{\sqrt{N_{\mathrm{graviton}}}} \approx 10^{-10}\,\mathrm{g} \tag{7}$$

$$\therefore N_{\mathrm{graviton}} \approx 10^{10} \tag{8}$$

With this, if say one has a 1 gram black hole, about 10^5 times larger than a Planck mass, one would be having say an entropy generated this way of about 10^{10}, assuming Planck normalization.

3. So, Then What are the Number of Gravitons Emitted via a Spinning Planck-Sized Black Holes Component Binary in Terms of Gravitons?[6] What Does This Say About an Optimal Black-Hole Size as to Perhaps See Measurable GW/Graviton Generation Effects

Likely from the situation in Ref. 6 for items as of about a Planck length, and involving Planck-sized masses, we would see the following equation for a rotating rod, of mass M, and of velocity V, of its end, for graviton production.

$$N_g \approx \frac{32G}{45\hbar c} M_{\text{rod}}^2 \cdot \left(\frac{V_{\text{tip}}}{c}\right)^4 \propto 7.5 \times 10^{-8} M_{\text{rod}}^2 \Rightarrow N_g \geq 1 \text{ iff } M_{\text{rod}} \geq 4 \times 10^6 \text{ g}. \quad (9)$$

If we have an equivalent situation with respect to two black holes in a binary state, we would likely need to have approximately black holes of masses 10^5 g to 10^6 g—i.e., 10^{10} to 10^{11} times larger than Planck mass—to have a measurable GW/Graviton signal which would be commensurate with experimental data sets. If we had say 10^5 to 10^6 g black holes, then the value of gravitons released per second, from a BEC condensate of gravitons for a mini black hole would me many times larger than Eq. (8) above.

We don't know the exact values, but this leads to our next point, which is the stages of black holes, before the Planckian era, to at the point of time (and space) where 1 to 10^5 g black holes would be composed of gravitons by BEC condensation of gravitons, for a release in.

Considering this, what can we say about the regimes of black-hole masses, just before the Planckian era, during the Planckian era, and right after the Planckian era?

We are assuming the following. A moderately large number (10^6 or more) of super massive black holes which would be in the center of galaxies, and which would be broken up and recycled in the CCC cosmology regime, with masses dropping from about 10^{41} g, down to about 10^{-4} to 10^{-5} g, before recombination by Planck era recombination into a tier of black holes which would be at least 1 gram in mass, scaling up to 10^5 g in mass so as to allow for BEC generation of gravitons through entropy production as in Eq. (2)–Eq. (7) above.

In doing so, we purport to use the datum given in Ref. 6 that masses of say much lower than 10^5 to 10^6 g for black holes likely do not have much chance of producing gravitons which would be detectable in the present era. Indeed, a minimum mass of about 1 to 10 grams for a black hole would be needed for a Bose–Einstein condensation via gravitons for a "light, low-mass" black hole which would be able to by Eq. (2)–Eq. (6), Eq. (8), and Eq. (9) to have at least 10^{10} gravitons per second generated (entropy for a BEC black hole).

We then would to a round off approximation state this hierarchy of black-hole behavior and size to consider.

Table 1. Scaling of mass of black holes, and their purported number, If CCC cosmology (Penrose) assumed for GW radiation release (may affect the CMBR)

End of prior universe time frame	Super massive end-of-time black hole 10^{41} to 10^{44} g.	10^6 to 10^9 of black holes, usually from centers of galaxies
Planck era black hole formation assuming merging of micro black-hole pairs	Micro black holes 10^{-5} to 10^{-4} g (approximately the Planck mass value).	10^{40} to 10^{45} black holes, assuming not too much destruction of matter–energy from pre-Planck to Planck conditions
Post-Planck era black holes: Can use Eq. (2)–Eq. (6) to have 10^{10} gravitons/second released per black hole	Normal-sized black holes 10 g to 10^6 g	10^{20} to, at most, 10^{25} black holes with repeated black-hole pairs forming a single black hole multiple times.

4. Why We Would Have the Figures from Table 1 to Consider for Contributions to the CMBR and the Penrose Suggestion

The formula which is for luminosity from a black hole and in page 16 of Ref. 4 the text states that the two black holes emit GW with a wave frequency 2 times the rotation frequency of the orbit of the two black holes to each other.

If we assume that we are still using this approximation above,[4] we can see support for our choice of Planck length as the minimum separation distance between the two black holes via using Planck units normalized to 1 as yielding

$$R\ (\text{separation}) \simeq r_g^{\text{eff}} = \frac{M_1 + M_2}{M_{\text{Planck}}^2} \xrightarrow[M_1=M_2=M_{\text{Planck}}=1]{} 1 \equiv R_{\text{Planck}}. \tag{10}$$

Going to Clifford Will,[7] we see on page 252 a loss or shrinkage of the period for the rotating black hole pair defined by P

$$\frac{\dot{P}}{P} = \frac{\mathrm{d}P}{\mathrm{d}t} \cdot \frac{1}{P} = -\frac{3}{2}\frac{\dot{E}}{E}. \tag{11}$$

Whereas, with the mechanics version of P for a sphere to be defined by, where M is a mass of a star, and we assume a binary system with two masses of equal mass M, so that, if R is the separation between the two masses[8] on page 188 would be

$$P = R\sqrt{\frac{2\pi^2 R}{GM}} \xrightarrow[M_1=M_2=M_{\text{Planck}}=1]{} R_{\text{Planck}}\sqrt{\frac{2\pi^2 R_{\text{Planck}}}{GM_{\text{Planck}}}}. \tag{12}$$

For Planck-sized masses, this means that the period of the binary Planck mass black hole pair would be vanishingly small.

The frequency of rotation would be half that of the GW emitted by these two Planck mass black holes which would collapse into each other. Note that the frequency we have stated for this last step, is given in Eq. (13). That is, could we have the following quantization contribution to initial frequency?

Our final concluding point to this chapter is to review the physics of Fig. 1, and then to ask, can we ascertain the GW radiation of Planck era black hole stars in a

binary configuration contributing to a buildup of generating frequency getting. If

$$\Delta E \Delta t \approx \hbar \Rightarrow \hbar\omega\Delta t \approx \hbar\omega \left(\frac{2}{3a_{\min}}\right)^{\frac{1}{\gamma}} \Rightarrow \omega \approx \hbar^{-1}\left(\frac{2}{3a_{\min}}\right)^{-\frac{1}{\gamma}}. \tag{13}$$

We claim that if we take the energy as consistent with a change in value as given by Eq. (1) that this will lead to a frequency which may, if $a_{\min} \approx 10^{-25}$–$10^{-20}$ (range from 10^{-25} to 10^{-20}) lead to

$$\omega \approx \hbar^{-1}\left(\frac{3}{2}\right)^{\frac{1}{\gamma}} \cdot 10^{\frac{25}{\gamma}} \propto \left(\frac{3}{2}\right)^{\frac{1}{\gamma}} \cdot 10^{\frac{25}{\gamma}}\,\text{Hz}. \tag{14}$$

Whereas note that the frequency is, say dependent upon the choice of γ and that this could be very different from the Planck frequency

$$\omega_{\text{P}} \approx 1.885 \times 10^{43}\,\text{Hz}. \tag{15}$$

We have then that if one had a redshift, of $z \approx 10^{25}$, that this would mean a present value of frequency as of about 1 Hz, whereas we can consider what would be gained by looking at the contribution near the CMBR, $z \approx$ 1,100 or so for the CMBR, whereas this would mean roughly that we would be looking in the regime of the CMBR:

$$\omega_{\text{signal from Planck to CMBR}} \propto \left(\frac{3}{2}\right)^{\frac{1}{\gamma}} 10^{\frac{25}{\gamma}} \times 10^{-3}\,\text{Hz}. \tag{16}$$

However, we have in doing this, that the duration of this frequency signal would be very minimal, due to the decay of the period, this would be going on for less than a nanosecond.

If so then we would need to refer to Eq. (2)–Eq. (6) and the value of

$$E_{\text{BEC–Graviton}} \approx \frac{k_B T_{\text{BH}}}{2} \approx \frac{k_B \times 10^{-5} \times T_{\text{P}}}{2}$$
$$\Rightarrow \omega_{\text{BEC–Graviton}} \approx 10^{38}\,\text{Hz} \Rightarrow \omega_{\text{BEC–Graviton-to-CMBR}} \approx 10^{38}\,\text{Hz}. \tag{17}$$

Needless to state, that unlike the case of (12), one would likely have the duration of the signal last long enough as to imprint directly on the CMBR. That is, look at Ref. 8. Also, for this I refer to the Zeldovich 4 conference Abhay Ashtekar presentation.[9] Ashtekar referred to a removal of bogus data points in the CMBR (Figure 1[9]).

Now looking at what was discussed by Abhay Ashtekar in Zeldovich 4, on September 7, 2020.[9]

In our Fig. 2, we copy what was done by Ashtekar, in Zeldovich 4 as to what was part of anisotropic fits to the E and B polarization, as given is made easier, if there is a nonsingular start to the universe which I discussed in detail in Ref. 10, and that further polarization states which may be analyzed in detail could be ascertained in Ref. 11.

If one has a nonsingular start to the universe, modeled on a multiverse generalization of Penrose CCC cosmology[10] then the details of a break up of black holes

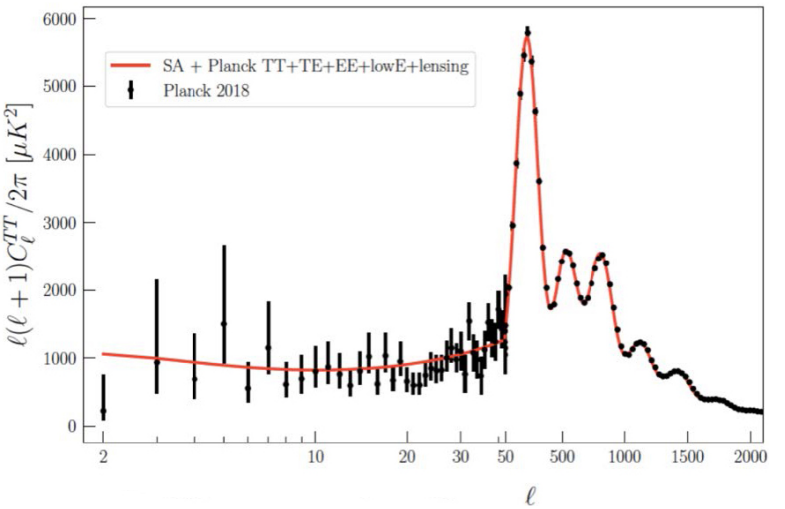


Fig. 2. Filling in the data points left out in terms of CMBR cosmic microwave background[9]

would not be so startling, i.e., these are the details from Ref. 10 as given by the following generalization of CCC cosmology.[10]

4.1. *Looking now at the modification of the Penrose CCC (cosmology)*

We now outline the generalization for Penrose CCC (Cosmology) just before inflation which we state we are extending Penrose's suggestion of cyclic universes, black hole evaporation, and the embedding structure our universe is contained within. This multiverse has black holes and may resolve what appears to be an impossible dichotomy. The text following is largely from Ref. 10 and has serious relevance to the final part of the conclusion. That there are N universes undergoing Penrose "infinite expansion"[2] contained in a mega universe structure. Furthermore, each of the N universes has black hole evaporation, with Hawking radiation from decaying black holes. If each of the N universes is defined by a partition function, called $\{\Theta_i\}_{i=N}^{i=1}$, then there exist an information ensemble of mixed minimum information correlated about 10^7–10^8 bits of information per partition function in the set $\{\Theta_i\}_{i=N}^{i=1}\big|_{\text{before}}$, so minimum information is conserved between a set of partition functions per universe.[12]

$$\{\Theta_i\}_{i=N}^{i=1}\big|_{\text{before}} \equiv \{\Theta_i\}_{i=N}^{i=1}\big|_{\text{after}} \tag{18}$$

However, there is nonuniqueness of information put into partition function $\{\Theta_i\}_{i=N}^{i=1}$. Also Hawking radiation from black holes is collated via a strange attractor collection in the mega universe structure to form a new inflationary regime for each of the N universes represented.

Our idea is to use what is known as CCC cosmology,[12] which can be thought of as the following. First. Have a big bang (initial expansion) for the universe which is represented by $\{\Theta_i\}_{i=N}^{i=1}$. Verification of this mega structure compression and expansion of information with stated nonuniqueness of information placed in each of the N universes favors ergodic mixing of initial values for each of N universes expanding from a singularity beginning. The n_f stated value, will be $S_{\text{entropy}} \approx n_f$.[12,13] How to tie in this energy expression, as in Eq. (16) will be to look at the formation of a nontrivial gravitational measure as a new big bang for each of the N universes as by $n(E_i)$. The density of states at energy E_i for partition function.[12,14]

$$\{\Theta_i\}_{i=1}^{i=N} \propto \left\{ \int_0^\infty \mathrm{d}E_i \cdot n(E_i) \cdot \mathrm{e}^{-E_i} \right\}_{i=1}^{i=N} . \tag{19}$$

Each E_i identified with Eq. (13) above, are with the iteration for N universes.[2,12] Then the following holds, by asserting the following claim to the universe, as a mixed state, with black holes playing a major part.

4.1.1. *Claim 1*

See the below representation[12] of mixing for assorted N partition function per CCC cycle.

$$\frac{1}{N}\sum_{j=1}^{N} \Theta_j\big|_{j \text{ before nucleation regime}} \xrightarrow[\text{vacuum–nucleation transfer}]{} \Theta_i\big|_{i \text{ fixed after nucleation regime}} \tag{20}$$

For N number of universes, with each $\Theta_j\big|_{j \text{ before nucleation regime}}$ for $j = 1$ to N being the partition function of each universe just before the blend into the right-hand side of Eq. (20) above for our present universe. Also, each independent universes as given by $\Theta_j\big|_{j \text{ before nucleation regime}}$ is constructed by the absorption of one to ten million black holes taking in energy.[2] Furthermore, the main point is done in Ref. 10 in terms of general ergodic mixing.[12]

4.1.2. *Claim 2*

$$\Theta_j\big|_{j \text{ before nucleation regime}} \approx \sum_{k=1}^{\max} \widetilde{\Theta}_k\Big|_{\text{black holes } j\text{th universe}} \tag{21}$$

What is done in Claims 1 and 2[10] is to come up as to how a multi dimensional representation of black hole physics enables continual mixing of spacetime largely as a way to avoid the anthropic principle,[10] as to a preferred set of initial conditions.

5. Conclusion

If one has a nonsingular start of expansion of the universe and ergodic mixing of initial conditions of space-time from other universes, how does this relate to the

breaking up of black holes from Table 1? In Ref. 10 in order to do away with the anthropic principle, the following references in terms of ergodic mixing of the partition function of the universe was utilized, as far as a multiverse. But there is one final piece. Assume that we have

$$\omega_{\text{Earth}} \leq 10^{-25}\omega_{\text{initial}}. \tag{22}$$

We will be of course assuming an equivalence between a graviton count and information,[16] and we can in future work compare this with the Rosen[3] value of energy for a mini universe of (from a Schrödinger equation) with ground state mass of $m = \sqrt{\pi}M_{\text{Planck}}$ and an energy of

$$E_{\hat{n}} = \frac{-Gm^5}{2\pi^2\hbar^2\hat{n}^2}. \tag{23}$$

Our preliminary supposition is that Eq. (23) could represent the initial energy of a pre-Planckian universe and that Eq. (24) would be thermally based energy dumped into the space-time bubble assumed in Ref. 10. That is,

$$E_{\text{universe}} = 10^{41} \times E_{\text{BEC–Graviton}} \approx 10^{41} \times \left(\frac{k_B T_{\text{BH}}}{2} \approx \frac{k_B \times 10^{-5} \times T_{\text{P}}}{2}\right) \tag{24}$$

is the thermal energy dumped in due to the use of cyclic conformal cosmology. Here we specify that initially it would have that the value of Eq. (24) would exactly counter balance the energy given in a negative form by Rosen as of Eq. (23).

Now use the following approximation of the universe, initially having the entropy of a black hole. That is, we are using Ng infinite quantum statistics,[18] while area denotes the surface area of the regime of space-time:

$$S_{\text{universe}} \propto S_{\text{BH}} \simeq \frac{A}{4l_{\text{Planck}}^2} \approx \frac{9n_Q}{4} \approx n_{\text{graviton}}. \tag{25}$$

This way of noting entropy and the signals of the prior universe black holes being generated secondarily is a surface area which is commensurate with the utilization of Eq. (2)–Eq. (6) for BEC condensation by gravitons for early-universe black holes. This is in tandem with the quantum fluctuations as seen in Figure 2 below. Also see Appendix A as well as the physics.[3,17,18]

The bubble nucleation, plus the details of cosmology leading to black holes from a prior universe showing up:

> For thirty years Oxford mathematician Roger Penrose has challenged one of the key planks of cosmology, namely the concept of inflation, now over 40 years old, according to which our universe expanded at an enormous rate immediately after the big bang. Instead, fifteen years ago, Penrose proposed a counterconcept of conformal cyclic cosmology by which inflation is moved to before the big bang and which introduces the idea of preceding eons. The concept has been disputed by most physicists, but Penrose and colleagues believe that new evidence has come to light which requires closer inspection

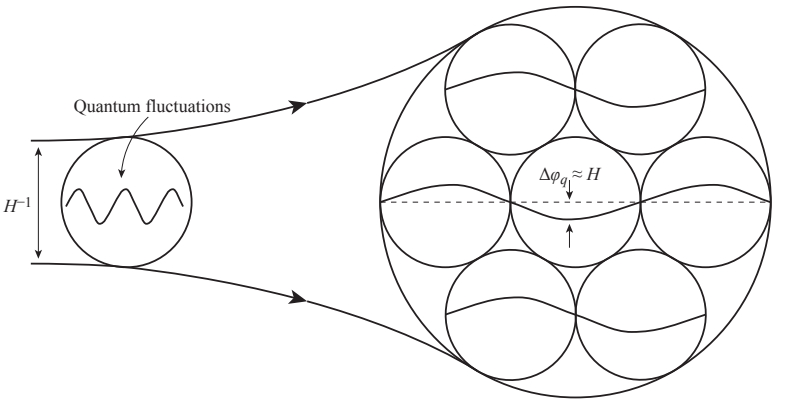


Fig. 3. We here are examining how the universe has self replicating regimes of spacetime[10, 14]

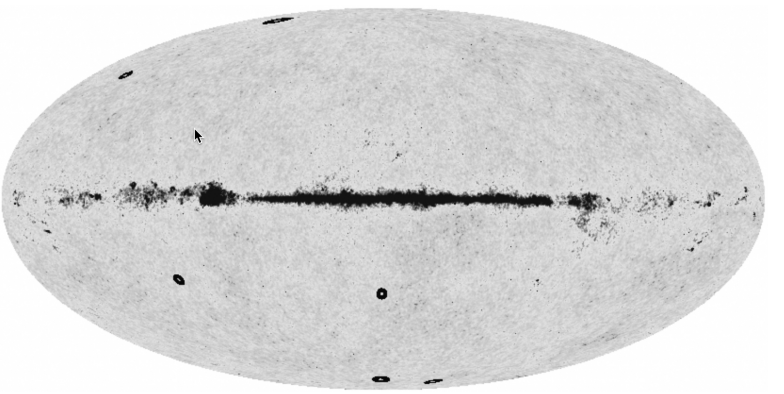


Fig. 4. Mollweide view showing how the CMBR spectrum has "rings" in it from black holes from a prior universe

Source: This from Ref. 19 with Ref. 3 having the data points used to construct this image in.

and argument—the research is published today in the Monthly Notices of the Royal Astronomical Society.[19]

Recent analysis of the cosmic microwave background (CMB) by Penrose, An, Meissner, and Nurowski has revealed, both in the Planck and WMAP satellite data (at 99.98% confidence), a powerful signal that had never been noticed previously, namely numerous circular spots $\approx$ 8 times the diameter of the full moon. The brightest six (Figure 1) are $\approx$ 30 times the average CMB temperature variations seen at precisely the same locations in the Planck and WMAP data. These spots were overlooked previously owing to a belief that the very early exponentially expanding *inflationary* phase of standard cosmology should have obliterated any such features.

Judicious application of Eq. (2)–(6) plus Table 1 above leads to this phenomenon.

Appendix A. Examining How Many Gravitons Might Be Produced by Initially Planck-Sized Black Holes

Alexander D. Dolgov and Damian Ejlli[4] inform us that a mass of a primordial black hole is

$$M_{\text{early black hole}} \approx 4 \times 10^{38} t \frac{\text{g}}{\text{s}}. \tag{26}$$

A Planck mass is of the value 10^{-5} g, i.e., almost, is then obtainable when

$$t_{\text{formation}} \approx 10^{-43}\,\text{s} \Rightarrow M_{\text{early black hole}} \approx 10^{-5}\text{g}. \tag{27}$$

Note that $t_{\text{formation}} \approx 10^{-43}\,\text{sec} \geq 5.39 \times 10^{-44}\,\text{s}$ leads to almost a Planck mass, $2.176434(24) \times 10^{-5}\,\text{g} = M_{\text{Planck}}$.

The mechanism of how Planck-sized black holes could generate GW comes from,[4] initial friction in the early universe environment, leading to coupling of early primordial binary black hole systems which in turn would collapse and form larger black holes—i.e., in fact the argument in Ref. 18 is stated on page 15 as follows.

> For PBH masses below a few grams dynamical friction would be an efficient mechanism of PBH cooling leading to frequent binary formation. Moreover, dynamical friction could result in the collapse of small PBHs into much larger black holes with the mass of the order of Mb (18). This process would be accompanied by a burst of GW emission

What is called Mb in this situation is given in Ref. 4 on page 4.

> As we see in what follows, generation of gravitational waves would be especially efficient from such high density clusters of primordial black holes. Let us assume that the spectrum of perturbations is the flat Harrison–Zeldovich one and that a perturbation with some wavelength λ crossed horizon at moment t_{in}. The mass inside horizon at this moment was
>
> $$Mb(t_{\text{in}}) = m_2 Plt_{\text{in}}.\ (4) \tag{28}$$
>
> It is the mass of the would-be high density cluster of PBHs.

We then from here have the mechanism of black hole formation comes from binary pair formation of small black holes which collapse into a larger set of black holes. This chain of black-hole pair production and collapse would then lead to an accretion procedure along the lines of Eq. (25). Eventually these black hole clusters would form the mega black holes as seen in the center of spiral galaxies.

References

1. V. G. Gurzadyan and R. Penrose, On CCC-predicted concentric low-variance circles in the CMB sky, *Eur. Phys. J. Plus* **128**, Article 22 (2013). https://doi.org/10.1140/epjp/i2013-13022-4.

2. R. Penrose, Before the big bang: An outrageous new perspective and its implications for particle physics *Proc. EPAC 2006* (Edinburgh, Scotland, 2006) pp. 2759–2762, http://accelconf.web.cern.ch/accelconf/e06/PAPERS/THESPA01.PDF
3. D. An, K. A. Meissner, P. Nurowski, and R. Penrose, *Apparent Evidence for Hawking Points in the CMB Sky* (2020), https://arxiv.org/abs/1808.01740.
4. A. D. Dolgov and D. Ejlli, Relic gravitational waves from light primordial black holes, *Phys. Rev. D* **84**, Article 024028 (2011). https://doi.org/10.1103/PhysRevD.84.024028.
5. P.-H. Chavanis, Self gravitating Bose–Einstein condensates, in *Quantum Aspects of Black Holes*, ed. X. Calmet, Fundamental Theories of Physics, Vol. 178 (Springer Nature, Cham, Switzerland, 2012), pp. 151–194, https://doi.org/10.1007/978-3-319-10852-0_6.
6. G. L. Murphy, Gravitons from a spinning rod, *Aust. J. Phys.* **31**, 205–207 (1978), https://www.publish.csiro.au/ph/pdf/PH780205.
7. C. M. Will, *Theory and Experiment in Gravitational Wave Physics*, 2nd edn. (Cambridge University Press, New York, 2018).
8. S. Passaglia, W. Hu, and H. Motohashi, Primordial black holes and local non-Gaussianity in canonical inflation, *Phys. Rev. D* **99**, Article 043536 (2019), https://arxiv.org/abs/1812.08243.
9. A. Ashtekar, Quantum gravity in the sky? Alleviating tensions in the CMB using Planck scale physics, Zeldovich 4, (2020), http://www.icranet.org/images/stories/Meetings/ZM4/presentations/Ashtekar.pdf.
10. A. W. Beckwith, A Solution of the cosmological constant, using multiverse version of Penrose CCC cosmology, and enhanced quantization compared, *J. High Energy Phys. Gravit. Cosmol.* **7**, 559–571 (2021), https://doi.org/10.4236/jhepgc.2021.72032.
11. N. Poplawski, Cosmological constant from QCD vacuum and torsion, *Ann. Phys.* **523**, 291–295 (2011) https://doi.org/10.1002/andp.201000162.
12. H. Dye, On the ergodic mixing theorem, *Trans. Amer. Math. Soc.* **118**, 123–130 (1965), http://www.ams.org/journals/tran/1965-118-00/S0002-9947-1965-0174705-8/S0002-9947-1965-0174705-8.pdf.
13. P. D. Naselsky, D. Novikov, and I. Novikov, *The Physics of the Cosmic Microwave Background* (Cambridge University Press, Cambridge, UK, 2006).
14. V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, New York, 2005), https://doi.org/10.1017/CBO9780511790553.
15. N. Rosen, Quantum mechanics of a miniuniverse, *Int J. Theor. Phys.* **32** (8), 1435–1440 (1993), https://doi.org/10.1007/BF00675204.
16. I. Haranas and I. Gkigkitzis, The mass of graviton and its relation to the number of information according to the holographic principle, *Int. Sch. Res. Notices* **2014**, Article 718251, https://doi.org/10.1155/2014/718251.
17. Y. J. Ng, Holographic foam, dark energy and infinite statistics, *Phys. Lett. B*, **657**, 10–14 (2007), https://doi.org/10.1016/j.physletb.2007.09.052.
18. R. Letzter, *Physicists Think They've Spotted the Ghosts of Black Holes from Another Universe*, Live Science (2018), https://www.livescience.com/63392-black-holes-from-past-universes.html.
19. R. Penrose, *Hawking Points in the Cosmic Microwave Background—A Challenge to the Concept of Inflation*, University of Oxford Mathematical Institute (2018). https://www.maths.ox.ac.uk/node/36137.

Summary of the parallel session BH3

Kamal Hajian
Hanse-Wissenschaftskolleg Institute for Advanced Study, Lehmkuhlenbusch 4, 27733 Delmenhorst, Germany
E-mail: kamalhajian@ipm.ir

Jutta Kunz
Institute of Physics, University of Oldenburg, 26111 Oldenburg, Germany
E-mail: jutta.kunz@uni-oldenburg.de

We summarize the talks presented at the BH3 session (Black Holes in Alternative Theories of Gravity) of the 16th Marcel Grossmann Meeting held online on July 5-10 2021.

Keywords: Gravity; Alternative theories; Black holes.

1. Introduction

Black holes have received tremendous interest in recent years, originating on the one hand from the discovery of gravitational waves, generated in the merger of stellar black holes, by the LIGO collaboration,[1] and on the other hand from the observation of the shadow of the supermassive black hole at the center of M87 by the EHT collaboration.[2] The analysis of observational data of black holes is typically based on the *Kerr hypothesis*, namely the assumption that rotating black holes are well described by the Kerr solution of General Relativity. Just like General Relativity itself, also the *Kerr hypothesis* is still consistent with all available data.

However, there are a number of reasons suggesting that General Relativity will be superseded by a new theory of gravity, which would reduce to General Relativity in a limit. From a theoretical side these reasons include the problem of quantization of gravity and the presence of singularities in solutions of General Relativity, while from an observational side the need for *Dark Matter* and *Dark Energy* in a cosmological context seem most provoking.

All these reasons have led to a large number of suggestions for alternative theories of gravity based on deep theoretical reasoning or simply phenomenological modelling, see e.g.,[3–6] While alternative theories of gravity should be consistent with observations in the weak gravity regime, thus in particular, with observations in the solar system, the strong gravity regime is so far much less constrained. In the strong gravity regime there exist certainly high precision data from pulsar observations (see, e.g.[7]), but the unknown equation of state of matter at extreme densities and pressures may lead to degeneracies with the unknown theory of gravity. Therefore black holes may present cleaner probes of the strong gravity regime.

In General Relativity black holes are remarkably simple objects in terms of the characterization of the space-times, as expressed by the *no-hair theorem* (see, e.g.[8]). Consequently, Kerr black holes are uniquely determined by their mass and their angular momentum, and all their higher multipole moments can be fully expressed by these two lowest moments. Moreover, when considering the fields of the Standard Model of Particle Physics, no astrophysical rotating black holes other than the *bald* Kerr black holes arise. Only microscopic black holes might carry fields of the Standard Model as hair (see, e.g.[9,10]). However, if one allows for a hypothetical complex massive scalar field or Proca field, then General Relativity allows for rotating Kerr black holes with scalar or Proca hair that might be of astrophysical relevance.[11,12]

Alternative theories of gravity typically introduce further degrees of freedom, often in the form of gravitational scalar or vector fields.[3–6] Prominent theories with additional scalar fields are Horndeski and beyond Horndeski theories,[13–16] while their counterparts with vector fields are generalized Proca and beyond generalized Proca theories.[17–20] But further degrees of freedom may also be present when the tensorial part of gravity is modified, yielding, for instance, de Rham-Gabadadze-Tolley massive gravity[21] and bigravity theories.[22,23] While General Relativity can be formulated in terms of curvature, torsion or non-metricity, as expressed by the *geometrical trinity of gravity*,[24] the associated generalizations of these formulations are no longer equivalent.

Up to now most work on black holes in alternative theories of gravity has been done in metric theories. Here the recent years have witnessed the emergence of new types of black holes due to newly discovered phenomena. These are, for instance, curvature-induced spontaneously scalarized black holes that arise in certain Horndeski models with higher curvature terms, when curvature is sufficiently strong to induce a tachyonic instability in the background of a Schwarzschild or Kerr black hole solution.[25–29] Similarly, spin-induced spontaneously scalarized black holes emerge in the presence of sufficient curvature and spin.[30–33] While these solutions satisfy circularity, this is not necessarily true for rotating black holes in beyond Horndeski theories, where also so-called disformed black holes with a non-circular geometry can arise (see, e.g.[34,35]).

A very important aspect of black holes in alternative theories of gravity is of course their stability and their response to perturbations. In particular, the study of quasi-normal modes (QNMs) is very valuable here, since these are also of relevance for the analysis of the ringdown phase of black hole mergers. Whereas the QNMs of the Schwarzschild and Kerr black holes are known since quite some time (see, e.g.[36–40]), the study of the QNMs of black holes in alternative theories of gravity is still in its infancy. While some work has been done for static black holes (see, e.g.[41]), very little is known so far even for the case of slow rotation only. The presence of additional degrees of freedom in alternative theories of gravity, yields a much more intricate spectrum, though, since various modes forbidden in General Relativity will be present in alternative theories of gravity.

In the following we will summarize the progress on black holes in alternative theories of gravity reported in our session of the 16th Marcel Grossmann Meeting.

2. Summary of the presentations

The 16th Marcel Grossmann Meeting was held online on July 5-10 2021. This was the first time that the meeting was held completely online, mainly because of the COVID-19 pandemic around the world. In the parallel session "Black Holes in Alternative Theories of Gravity" abbreviated as BH3, 16 physicists presented their recent research activities in two days, Tuesday 06 July 2021 09:30-12:30 and Thursday 08 July 2021 16:30-19:30 CEST. Here we recall the main points of the talks in BH3, and provide a summary of the presentations.

2.1. *Asymptotically flat hairy black holes in massive bigravity*

Presented by: Mikhail Volkov
In collaboration with: Romain Gervalle
Based on the Ref.[42].

The ghost-free theories of bigravity were first introduced by Hassan and Rosen in Ref.[43]. In these theories, there are two metrics denoted by $g_{\mu\nu}$ and $f_{\mu\nu}$. Both of the metrics have the usual Einstein-Hilbert action, and there is an interaction term between the two metrics. The metric $g_{\mu\nu}$ can be coupled to some matter, while $f_{\mu\nu}$ is not coupled to any matter. Analyzing the propagating degrees of freedom, there are two gravitons in such theories, one massive and one massless. These family of theories have been shown to be able to describe the accelerating expansion of the universe without a cosmological constant.

Any black hole as a vacuum solution in general relativity is a solution to bigravity theories by the choice of $g_{\mu\nu} = f_{\mu\nu}$. For example, the Schwarzschild metric is a solution to these theories if one chooses $g_{\mu\nu} = f_{\mu\nu}$ be equal to the Schwarzschild metric. However, this solution is not a stable solution in these theories. An interesting question arises here: if one relaxes the constraint $g_{\mu\nu} = f_{\mu\nu}$, is it possible to have spherically symmetric asymptotically flat solutions other than the Schwarzschild. The terminology "hairy black holes" has been used for such solutions in bigravity, and the "bald Schwarzschild" for the special case of $g_{\mu\nu} = f_{\mu\nu}$ to be Schwarzschild. It has been a debate to answer this question[44–46] whether there are hairy black holes in ghost-free bigravity or not. Romain Gervalle and Mikhail Volkov in Ref.[42] have tried to answer this interesting question by constructing spherically asymptotically flat black hole solutions numerically. They have found that there exist such black hole solutions (as pairs of solutions, one for $g_{\mu\nu}$ and one for $f_{\mu\nu}$) with the same horizon and surface gravity. Free parameters of the Schwarzschild hairy black holes are constrained by studying the stability of the solution. To this end, the $g_{\mu\nu}$ metric should be close to the Schwarzschild metric, the $f_{\mu\nu}$ metric should not be coupled to the matter, and the mass of the black hole should be between $0.2M_{\odot}$ and $0.3 \times 10^6 M_{\odot}$.

2.2. *An overview of quasinormal modes in modified and extended gravity*

Presented by: Aurélien Barrau
In collaboration with: Flora Moulin and Killian Martineau
Based on the Ref.[47].

Quasi-normal modes (QNMs) are dissipative perturbations around a black hole background solution. They satisfy purely outgoing and ingoing boundary conditions at infinity and at the event horizon respectively. The oscillatory time dependence of these modes is described by

$$\psi \propto e^{i\omega t} = e^{i(\omega_R + i\omega_I)t}. \tag{1}$$

The frequency ω_R characterizes the oscillatory behavior of the mode, while for the $\omega_I > 0$ and $\omega_I < 0$ the mode grows or decays exponentially, respectively. The main question in the QNM analysis is to find the behavior of the complex function ω.

In this talk, Aurélien Barrau provided an overview on QNMs, and presented their results for perturbations around a spherically symmetric black hole in some alternative theories of gravity. Their analysis is based on Wentzel-Kramers-Brillouin (WKB) approximation method.[48–50] The theories under consideration have been chosen to be the massive gravity,[21,22,51,52] Modified Scalar-Tensor-Vector (STV) Gravity,[53] Horava-Lifshitz,[54–57] $\hbar$-correction (quantum correction),[58] and loop quantum gravity (based on the model presented in Ref.[59]). For each one of these choices, the diagrams of ω_R and ω_I for some suitable multipole number ℓ and overtone number n were illustrated and compared qualitatively.

2.3. *Constraining modified gravity theories with physical black holes*

Presented by: Sebastian Murk
In collaboration with: Daniel R. Terno
Based on the Ref.[60].

In this work, the authors emphasize that a physical black hole is a celestial object which has a smooth apparent horizon and trapped surface.[61] Requesting the smoothness of these surfaces constrains the gravitational models which govern the dynamics of such solutions. To be more accurate, the existence of semiclassical physical black holes in modified theories of gravity induces some necessary conditions.

In order to find such necessary conditions, Sebastian Murk as a PhD student and his colleague have focused on the spherically symmetric black hole solutions which are presented as expansions in the coordinate distance from the apparent horizon and do not require a General Relativity solution as the zeroth-order perturbative solution of the modified theory. The only condition which they impose is the regularity of apparent horizon for black hole solutions. They pick finiteness of T^{μ}_{μ} and $T^{\mu\nu}T_{\mu\nu}$ as the regularity condition on the trapped horizon. By perturbing the Einstein equation with a new term $G_{\mu\nu} + \lambda\epsilon_{\mu\nu} = 8\pi T_{\mu\nu}$ and a generic ansatz for a

spherically symmetric black hole metric, the necessary conditions on the component of the $\epsilon_{\mu\nu}$ are investigated. In the end, the results are studied more for a special modified gravity, the Starobinsky model.[62]

2.4. *Black holes, stationary clouds and magnetic fields*

Presented by: Nuno Santos
In collaboration with: Carlos A.R. Herdeiro
Based on the Ref.[63].

Stationary bosonic clouds are stable configuration of scalar fields around a black hole resulting in a bound state. Suggested by many examples, in order to have such a bound state, two conditions are necessary: (1) the possibility for superradiance, (2) a confinement mechanism. The former is necessary to synchronize the bosonic cloud with the black hole rotation, and the latter is needed to make the configuration stable. Interestingly, stationary clouds are characterized by a discrete number of nodes in the radial direction, n, the orbital angular momentum, ℓ , and the azimuthal harmonic index m. In this regard, they resemble orbital configurations in a Hydrogen atom.[11,12,64,65]

In this presentation, the bosonic clouds are studied around a Reissner-Nordström black hole immersed in a magnetic field. These family of black hole solutions are called Reissner-Nordström-Melvin black holes.[66] They are solutions to Einstein-Maxwell theory, and their asymptotics resemble a magnetic Melvin universe. These black holes are stationary and axially symmetric. The presence of an external magnetic field provides both of the conditions for the stability of a bosonic cloud, i.e., the ergoregion and the confinement mechanism. The scalar field in this analysis has been considered to be complex, massless and minimally coupled to gravity. So, the theory is described by the action:

$$\mathcal{I} = \frac{1}{4\pi}\int d^4x\sqrt{-g}\big[\frac{R}{4} - \frac{F^2}{4} - (\nabla^\mu\Psi^*)(\nabla_\mu\Psi)\big]. \tag{2}$$

The bosonic clouds in this model, and on the Reissner-Nordström-Melvin solution are studied, and it is shown that for specific mass to charge ratios, there exist stable bosonic clouds.

2.5. *Bardeen black hole from a self-dual radius in spacetime*

Presented by: Michael Florian Wondrak
In collaboration with: Marcus Bleicher, Piero Nicolini, and Euro Spallucci
Based on the Ref.[67,68].

String T-duality is an equivalence between two string theories on spacetimes with at least one compactified extra dimension – provided that the compactification radii are inversely related to each other, i.e. $R_1 = R$ and $R_2 = \frac{(R^\star)^2}{R}$. The special case of $R = R^\star$ is called self-dual radius. Beginning from T-duality in bosonic string theory, Michael Florian Wondrak first gave an overview on how quantum

field propagators are deformed.[69] The scalar propagator obtained via Schwinger's proper time formalism is read to be

$$G(k) = -\frac{\ell_0 K_1(\ell_0\sqrt{k^2+m^2})}{\sqrt{k^2+m^2}}. \tag{3}$$

In this relation, $K_\nu(x)$ is the Modified Bessel function of second kind, and ℓ_0 is the zero point length.[70] Studying the potential induced by this propagator, a spherically symmetric black hole solution was derived,

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2, \qquad f(r) = 1 - \frac{2Mr^2}{(r^2+\ell_0^2)^{3/2}}. \tag{4}$$

This metric resembles the Bardeen black hole with a new interpretation of the UV cut-off in terms of the zero point length ℓ_0 instead of a magnetic monopole charge. Then, the thermodynamics of this black hole was studied and it was shown that at the end of evaporation process, the evaporation stops with a cold remnant instead of a final explosion. The remnants of these black holes were investigated regarding a possible fraction of dark matter. With a mass below $5\times10^{-8}\,M_\odot$, they comply with recent constraints on primordial black holes.[71] In the rest of the talk, observability of ℓ_0 was discussed: From the hydrogen energy spectrum, an upper bound of 4×10^{-19} meter was deduced.

2.6. *Asymptotically flat black hole solution in modified gravity*

Presented by: Surajit Kalita
In collaboration with: Banibrata Mukhopadhyay
Based on the Ref.[72].

$f(R)$ gravities are among alternative theories of gravity which have observational motivations to be studied. In these theories, one replaces the Ricci scalar R in the Lagrangian of General Relativity by a function of R, which is called $f(R)$.

$$\mathcal{I} = \frac{1}{16\pi}\int d^4x\sqrt{-g}f(R). \tag{5}$$

Although the equations of motion are higher-order in derivatives, they do not suffer from Ostrogradsky instability. In this presentation, Surajit Kalita presented the construction of spherically symmetric asymptotic flat black hole solutions to a subset of $f(R)$ theories. This subset is parametrized by a constant B in the relation $\frac{df}{dR} = 1 + \frac{B}{r}$. The metric ansatz is chosen to be $g_{\mu\nu} = \mathrm{diag}(g_{tt}, g_{rr}, r^2, r^2\sin^2\theta)$. By inserting this ansatz in the equations of motion of $f(R)$ gravities, the unknown components (g_{tt}, g_{rr}) are found as an expansion if powers of B and $\frac{1}{r}$. Then, some properties of these solutions are analyzed, including marginally stable and bound orbits, and spherical accretion flows. Moreover, from these solutions it is deduced that the Birkhoff theorem can be violated in $f(R)$ gravities.

2.7. *Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity*

Presented by: Semin Xavier
In collaboration with: Jose Mathew and S. Shankaranarayanan
Based on the Ref.[73].

In this talk, Semin Xavier presented their results for an infinite number of solutions to a subset of $f(R)$ gravities. The theory is described by the action

$$\mathcal{I} = \frac{1}{16\pi}\int d^4x\sqrt{-g}f(R), \qquad f(R) = (\alpha_0 + \alpha_1 R)^p, \tag{6}$$

for a real number $p > 1$, and constants α_0 and α_1. The metric ansatz is chosen to be spherically symmetric:

$$ds^2 = -A(r)e^{\delta(r)}dt^2 + \frac{dr^2}{A(r)} + r^2(d\theta^2 + \sin^2\theta d\varphi^2). \tag{7}$$

Solving the equations of motion for the unknown functions $A(r)$ and $\delta(r)$ it is shown that there are an infinite number of possibility for these functions. As a result, it is discussed that in $f(r)$ gravities the Birkhoff theorem may be violated. In order to show this explicitly, two black hole solutions are constructed and studied.

2.8. *Does the Penrose suggestion as to black holes from a prior universe showing up in today's universe have credibility? Examined from a singular, and nonsingular beginning of cosmological expansion*

Presented by: Andrew Walcott Beckwith
Based on the Ref.[74].

Conformal Cyclic Cosmology (CCC) is a cosmological model in the framework of General Relativity, which is proposed by Roger Penrose.[75] In this model of cosmology, the universe iterates through infinite cycles, with the future timelike infinity of each previous iteration being identified with the past timelike infinity of the next universe. From the observational point of view, it has been suggested that the black holes in the previous universe can have implications in our universe, imprinted in the cosmic microwave background (CMB).[76,77]

In this talk, Andrew Walcott Beckwith discussed on the feasibility and credibility of the proposed methods for checking CCC using CMB data. He discussed the Penrose singularity theorem, and investigated the two cases of the CCC: cycling through singular or non-singular starting/ending points.

2.9. *Analytical computation of quasi-normal modes of slowly-rotating black-holes in dCS gravity*

Presented by: Manu Srivastava
In collaboration with: Yanbei Chen and S. Shankaranarayanan
Based on the Ref.[78].

Gravitational waves have provided interesting data about the inspiral and ring-down phases of black hole mergers. The waves in the ring-down phase are quasi-normal modes of the black hole merger. Therefore, by studying the data in the ring-down phase we can investigate the quasi-normal modes of the system.[36,37,40] However, quasi-normal mode frequencies depend on the gravitational theory. So, they provide a new tool to distinguish and examine alternative theories of gravity.

In this line of research, Manu Srivastava and his collaborators have focused on dynamical Chern-Simons gravity (dCS), which is described by the following action:[79]

$$\begin{aligned}\mathcal{I} =&\frac{1}{16\pi}\int d^4x\sqrt{-g}R - \frac{\beta}{2}\int d^4x\sqrt{-g}(\nabla_\mu\vartheta\nabla^\mu\vartheta + V(\vartheta))\\ &+\frac{\alpha}{4}\int d^4x\sqrt{-g}\vartheta * RR + \mathcal{I}_{\rm matter},\end{aligned} \tag{8}$$

in which ϑ is a pseudo-scalar field, and $*RR$ is the Pontryagin density

$$*RR \equiv \frac{1}{2}\epsilon^{cdef}R^a{}_{bef}R^b{}_{acd}. \tag{9}$$

where ϵ^{cdef} is the Levi-Civita tensor. The background is chosen to be a slowly rotating black hole introduced in Ref.[80]. The parameter of slow rotation is denoted by a. In the analysis, quasi-normal modes in the axial and polar sectors are studied up to linear order in a and quadratic order in α. The results of this study, along with the data from gravitational wave observations, can be used as a test for dCS gravity and to constrain coupling parameters.

2.10. *Scalar perturbations of Kerr black-holes in hybrid metric-Palatini gravity*

Presented by: João Luís Rosa
In collaboration with: José P. S. Lemos and Francisco S. N. Lobo
Based on the Ref.[81].

In General Relativity, the connection $\Gamma^\lambda_{\mu\nu}$ is assumed to be the metric connection, which is related to the metric by the relation

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\lambda\sigma}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right). \tag{10}$$

However, in the Palatini formulation of gravity (see a review in Ref.[82]), the connection $\hat{\Gamma}^\lambda_{\mu\nu}$ is considered independent of the metric. If we denote the Ricci scalar which is built upon $\hat{\Gamma}$ by $\mathcal{R}$, then the generalized hybrid metric-Palatini (GHMP) gravity[83–85] (see a review in Ref.[86]) is described by the following action:

$$\mathcal{I} = \frac{1}{16\pi}\int d^4x\sqrt{-g}f(R,\mathcal{R}) + \mathcal{I}_{\rm matter}. \tag{11}$$

By the equations of motion, it turns out that the $\mathcal{R}$ is the Ricci scalar calculated for a metric which is conformal to $g_{\mu\nu}$. Motivated by this, João Luís Rosa and his

collaborators have shown that for any vacuum solution to GR, there is a non-trivial function $f(R, \mathcal{R})$ such that it is a solution to the GHMP too.

In this context, the Kerr black hole can also be considered as a solution to the GHMP. The stability of the Kerr constrains the $f(R, \mathcal{R})$, and this is analyzed in this presentation by studying massive scalar perturbations around the Kerr black hole. In particular, it is shown that the Kerr black hole is stable for some specific $f(R, \mathcal{R})$ theories and masses for the scalar perturbation.

2.11. *Emergent magnetic monopoles in degenerate theory*

Presented by: Suvikranth Gera
In collaboration with: Sandipan Sengupta
Based on the Ref.[87].

Magnetic monopoles are theoretical counterparts for electric charges in electric-magnetic duality.[88,89] Although they have been studied extensively, there is not yet observational evidence for their existence in nature.

In this talk, Suvikranth Gera presented a non-invertible metric which resembles a magnetic monopole in the first-order formulation of gravity.[90–92] Unlike the usual metric formulations, the non-invertible metric is well-defined in the tetrad formulations. In the presentation, he first presented the metric explicitly, and then followed by the calculation the spin-connection and its field strength in the first-order formulation. The magnetic charge is calculated, and the topological origin and its observability is discussed at the end, and it is mentioned that this solution has not any curvature singularity. Moreover, for the observers moving on timelike geodesics, this emergent magnetic charge is not accessible observationally, although it affects the curvature of the space-time.

2.12. *Black holes in metric-affine gravity: properties and observational discriminators*

Presented by: Diego Rubiera-Garcia
Based on the Ref.[93].

A generic connection $\Gamma^{\lambda}_{\mu\nu}$ can be independent of the metric (see the review in Ref.[82]). Such a generic connection is called affine connection. Accordingly, it can be regarded as an independent field in the Lagrangian. These models of gravity are called metric-affine gravities. In a subset of such models, which is called "Ricci based gravities," the connection $\Gamma^{\lambda}_{\mu\nu}$ appears in the Lagrangian only through a symmetric Ricci tensor $R_{\mu\nu}$,

$$\mathcal{I} = \frac{1}{16\pi}\int d^4x\sqrt{-g}\mathcal{L}_G[g_{\mu\nu}, R_{\mu\nu}(\Gamma)] + \mathcal{I}_{\text{matter}}[g_{\mu\nu}, \psi_m]. \tag{12}$$

In this talk, Diego Rubiera-Garcia presented an overview on recent results for the black holes in these models of gravity. It is shown that for spherically symmetric black holes, the curvature singularity can be smoothed out, and the geodesic

completeness is restored. Moreover, for such solutions there exist "double critical curves." This feature makes the shadow of the black hole to appear as two or more bright rings.

2.13. *Holography for rotating black holes in $f(T)$ gravity*

Presented by: Masoud Ghezelbash
In collaboration with: Canisius Bernard
Based on the Ref.[94].

General relativity is based on the Riemann curvature, while the torsion and non-metricity are considered to vanish. However, it is equivalent to make gravitational theories solely based on torsion, or non-metricity. These equivalent formulations in the literature are referred as Teleparallel and Symmetric Teleparallel (or coincident) gravities respectively. Similar to the Ricci scalar R which is made from the Riemann tensor, one can appropriately define the torsion scalar T, or the non-metricity scalar Q. According to the equivalence alluded to above, one can have analogous theories for $f(R)$, namely $f(T)$ or $f(Q)$.

It is useful to study gravitational features in GR, in the Teleparallel gravity or Symmetric Teleparallel gravity. In this talk, Masoud Ghezelbash has focused on the Kerr/CFT correspondence in $f(T)$ gravities. Kerr/CFT is a correspondence between the near horizon region of extremal black holes (originally Kerr black hole[95]) with a (chiral) two dimensional CFT. In this correspondence, the entropy of the black hole is calculated via the Cardy-formula for the entropy in a CFT:[96,97]

$$S = \frac{\pi^2}{3}(c_L T_L + c_R T_R), \tag{13}$$

in which the c's are the central charge of the left and right sectors, and the T's are Frolov-Thorne temperatures.[98] The result of such analyses shows that the entropy is proportional to the angular momentum associated with the axial symmetry which eventually enhances to the Virasoro sectors in the CFT (a review on Kerr/CFT and its extensions can be found in Ref.[99]). Besides, there are many works attempting to realize black hole microstates from the CFT using the Kerr/CFT correspondence (see e.g.[100–108]).

In this study, the model of gravity has been chosen to be

$$\mathcal{I} = \frac{1}{16\pi}\int d^4x\sqrt{-g}\left(f(T) - 2\Lambda - F^2\right), \tag{14}$$

in which Λ is the cosmological constant, and F^2-term is the Maxwell theory. A rotating charged black hole as a solution to this theory (introduced in Ref.[109]) is chosen to be studied, and the wave equation for a massless scalar field is approximated in the near horizon geometry of this black hole. Instead of a chiral Virasoro, a full Virasoro algebra is reported in this near horizon, and the Cardy formula (13) is used to reproduce the black hole entropy.

2.14. *Universe in a black hole with spin and torsion*

Presented by: Nikodem Poplawski
Based on the Ref.[110].

The Einstein-Cartan theory (EC) of gravity[111–115] is the simplest theory with torsion and curvature. In this theory, the Lagrangian is the same as the Einstein gravity

$$\mathcal{I} = \frac{1}{2\kappa}\int d^4x\sqrt{-g}(R + \mathcal{L}_{\rm matter}), \tag{15}$$

but in addition to the metric, the torsion is considered as a dynamical field. In this talk, Nikodem Poplawski studies the collapse of spherically symmetric fermionic matter to form a black hole in EC. He explains that if a fermionic field is considered as a perfect fluid with energy density ϵ and pressure p as the matter in this theory, the effect of the dynamics of torsion can be absorbed in the energy-momentum tensor[116] by

$$T_{\mu\nu} = \epsilon u_\mu u_\nu - p(g_{\mu\nu} - u_\mu u_\nu) \quad \rightarrow \quad \tilde{T}_{\mu\nu} = \tilde{\epsilon} u_\mu u_\nu - \tilde{p}(g_{\mu\nu} - u_\mu u_\nu), \tag{16}$$

in which

$$\tilde{\epsilon} = \epsilon - \alpha n_f^2, \qquad \tilde{p} = p - \alpha n_f^2. \tag{17}$$

In this relation, u_μ is the four-velocity, n_f is the number density of fermions, and $\alpha = \frac{\kappa}{32}$. Using $\tilde{T}_{\mu\nu}$, the equation of motion can be written as the usual Einstein equation:

$$G_{\mu\nu} = \kappa\tilde{T}_{\mu\nu}. \tag{18}$$

In this setup, the collapse is studied by the Tolman ansatz[117] for the metric:

$$ds^2 = e^{\nu(\tau,R)}d\tau^2 - e^{\lambda(\tau,R)}dR^2 - e^{\mu(\tau,R)}(d\theta^2 + \sin^2\theta d\varphi^2), \tag{19}$$

and solved for the unknown functions (ν, μ, λ). The results show that in this collapse, the singularity is prevented by the effects of the torsion. The geometry of the universe on the other side of the horizon is calculated, and it is shown that the geometry is an oscillating FLRW metric and it describes a closed universe. It is also calculated that due to the pair particle production, the frequency of the oscillation is reduced in time.

2.15. *Absorption by deformed black holes*

Presented by: Renan B. Magalhães
In collaboration with: Luiz C.S. Leite and Luís C.B. Crispino
Based on the Ref.[118].

The usual way to study alternative theories of gravity is to change the Lagrangian or the dynamical fields. Then, the equations of motion follow, and one can study the solutions to the new equations. However, in Ref.[119], Johannsen and

Psaltis introduced a parametric deviation approach, which is the reverse of the procedure mentioned above. They deform black holes by some parameters such that the black holes remain smooth, free of pathology, and have some suitable properties. Then, the equations of motion which could have such solutions are investigated. Following this method, Konoplya and Zhidenko in Ref.[120] have deformed the Kerr black hole by some parameters, while keeping its suitable features intact.

In this talk, Renan B. Magalhães presented the analysis of the absorption cross section of a massless scalar field for the static Konoplya-Zhidenko black hole. The metric of this black hole is similar to the Schwarzschild black hole with a deformed mass:

$$ds^2 = -f dt^2 + \frac{dr^2}{f} + r^2 d\Omega^2, \qquad f = 1 - \frac{2M}{r}, \qquad M \to M + \frac{1}{2}\sum_i \frac{\eta_i}{r^i}, \tag{20}$$

with the special choice of $\eta_i = \delta_{i2}$. So, the f is chosen to be $f = 1 - \frac{2M}{r} + \frac{\eta}{r^3}$ for some parameter η. In order to have a horizon, $\eta_{\min} < \eta < 0$ in which $\eta_{\min} = \frac{-32}{27M^3}$. On this specific background, Renan B. Magalhães and his collaborators have studied the wave equation for a scalar field $\Box\psi = 0$. He presented the result of the radial potential for different spherical harmonic modes, illustrating them from numerical calculations. Moreover, the absorption cross sections are derived numerically.

2.16. *Shadow of a charged black hole surrounded by an anisotropic matter field*

Presented by: Javier Badía
In collaboration with: Ernesto F. Eiroa
Based on the Ref.[121].

Observation of the shadow of a supermassive black hole at the center of galaxy M87 has been one of the main progresses in black hole physics in recent years.[122] In parallel with this observation, theoretical studies of black hole shadows have been one of the active lines of research (see Ref.[123] for a review). Especially, the presence of matter fields surrounding black holes, and its effect on their shadows has been investigated (see e.g.[124–127]).

In this talk, Javier Badía presents the results of the calculation of shadows for rotating charged black holes surrounded by an anisotropic matter field.[128,129] In the spherical coordinates (t, r, θ, φ) the anisotropic matter is considered to be a perfect fluid which is described by the energy-momentum tensor

$$T_\mu^{\ \nu} = \mathrm{diag}(-\rho, p_1, p_2, p_2), \qquad p_1 = -\rho, \quad p_2 = \mathrm{w}\rho. \tag{21}$$

The rotating charged black hole surrounded by this matter is derived by the Newman-Janis algorithm to the following spherically symmetric spacetime:

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2, \qquad f(r) = 1 - \frac{2m(r)}{r}, \qquad m(r) = M - \frac{Q^2}{2r} + \frac{K}{2r^{2\mathrm{w}-1}}. \tag{22}$$

The M, Q, and K are integration constants, and w is restricted to be $\frac{1}{2} < \mathrm{w} < 1$ by physical energy conditions. The resulting metric by the Newman-Janis algorithm is the Kerr metric in which $M \to m(r)$, i.e.

$$\begin{aligned} ds^2 =& -(1-f)dt^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2 - 2fa\sin^2\theta\, dt d\varphi \\ &+ \left(r^2 + a^2 + fa^2\sin^2\theta\right)\sin^2\theta\, d\varphi^2, \\ \rho^2 =& r^2 + a^2\cos^2\theta\,, \qquad \Delta = r^2 - 2m(r)r + a^2\,, \qquad f = \frac{2m(r)r}{\rho^2}, \end{aligned} \tag{23}$$

and $m(r) = M - \frac{Q^2}{2r} + \frac{K}{2r^{2\mathrm{w}-1}}$.

On this background, the Hamilton-Jacobi equation for the null geodesics is separable, and reduces to ordinary differential equations with radial derivatives. The result of the shadow calculations for an asymptotic observer on the equator is presented by illustrations. Moreover, three observables which characterize the shadows, named as the area, oblateness, and centroid of the shadows are discussed.[121,130]

Acknowledgments

We would like to thank the organizers of MG16 for making this interesting meeting with numerous inspiring talks and discussions possible in spite of the pandemic, and we would like to thank the speakers of our session for their notable contributions and for sharing with us their latest results. K.H. thanks also the administration of Hanse-Wissenschaftskolleg (Institute for Advanced Study) for the extension of his fellowship. J.K. gratefully acknowledges support by the DFG Research Training Group 1620 *Models of Gravity* and the COST Actions CA15117 *CANTATA* and CA16104 *GWverse*.

References

1. B. P. Abbott *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**, p. 061102 (2016).
2. K. Akiyama *et al.*, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J.* **875**, p. L1 (2019).
3. C. M. Will, The Confrontation between general relativity and experiment, *Living Rev. Rel.* **9**, p. 3 (2006).
4. V. Faraoni and S. Capozziello, *Beyond Einstein Gravity: A Survey of Gravitational Theories for Cosmology and Astrophysics* (Springer, Dordrecht, 2011).
5. E. Berti *et al.*, Testing General Relativity with Present and Future Astrophysical Observations, *Class. Quant. Grav.* **32**, p. 243001 (2015).
6. E. N. Saridakis *et al.*, Modified Gravity and Cosmology: An Update by the CANTATA Network (5 2021).
7. L. Shao *et al.*, Testing Gravity with Pulsars in the SKA Era, *PoS* **AASKA14**, p. 042 (2015).
8. P. T. Chrusciel, J. Lopes Costa and M. Heusler, Stationary Black Holes: Uniqueness and Beyond, *Living Rev. Rel.* **15**, p. 7 (2012).

9. M. S. Volkov and D. V. Gal'tsov, Gravitating nonAbelian solitons and black holes with Yang-Mills fields, *Phys. Rept.* **319**, 1 (1999).
10. B. Kleihaus, J. Kunz and F. Navarro-Lerida, Rotating black holes with non-Abelian hair, *Class. Quant. Grav.* **33**, p. 234002 (2016).
11. C. A. R. Herdeiro and E. Radu, Kerr black holes with scalar hair, *Phys. Rev. Lett.* **112**, p. 221101 (2014).
12. C. Herdeiro, E. Radu and H. Rúnarsson, Kerr black holes with Proca hair, *Class. Quant. Grav.* **33**, p. 154001 (2016).
13. G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, *Int. J. Theor. Phys.* **10**, 363 (1974).
14. C. Charmousis, E. J. Copeland, A. Padilla and P. M. Saffin, General second order scalar-tensor theory, self tuning, and the Fab Four, *Phys. Rev. Lett.* **108**, p. 051101 (2012).
15. T. Kobayashi, M. Yamaguchi and J. Yokoyama, Generalized G-inflation: Inflation with the most general second-order field equations, *Prog. Theor. Phys.* **126**, 511 (2011).
16. J. Gleyzes, D. Langlois, F. Piazza and F. Vernizzi, Healthy theories beyond Horndeski, *Phys. Rev. Lett.* **114**, p. 211101 (2015).
17. G. Tasinato, Cosmic Acceleration from Abelian Symmetry Breaking, *JHEP* **04**, p. 067 (2014).
18. L. Heisenberg, Generalization of the Proca Action, *JCAP* **05**, p. 015 (2014).
19. G.-P. Nicosia, J. Levi Said and V. Gakis, Generalised Proca theories in teleparallel gravity, *Eur. Phys. J. Plus* **136**, p. 191 (2021).
20. L. Heisenberg, R. Kase and S. Tsujikawa, Beyond generalized Proca theories, *Phys. Lett. B* **760**, 617 (2016).
21. C. de Rham, G. Gabadadze and A. J. Tolley, Resummation of Massive Gravity, *Phys. Rev. Lett.* **106**, p. 231101 (2011).
22. S. F. Hassan and R. A. Rosen, On Non-Linear Actions for Massive Gravity, *JHEP* **07**, p. 009 (2011).
23. S. F. Hassan and R. A. Rosen, Exact Solution to the 'Auxiliary Extra Dimension' Model of Massive Gravity, *Phys. Lett. B* **702**, 90 (2011).
24. J. B. Jiménez, L. Heisenberg and T. S. Koivisto, The Geometrical Trinity of Gravity, *Universe* **5**, p. 173 (2019).
25. D. D. Doneva and S. S. Yazadjiev, New Gauss-Bonnet Black Holes with Curvature-Induced Scalarization in Extended Scalar-Tensor Theories, *Phys. Rev. Lett.* **120**, p. 131103 (2018).
26. H. O. Silva, J. Sakstein, L. Gualtieri, T. P. Sotiriou and E. Berti, Spontaneous scalarization of black holes and compact stars from a Gauss-Bonnet coupling, *Phys. Rev. Lett.* **120**, p. 131104 (2018).
27. G. Antoniou, A. Bakopoulos and P. Kanti, Evasion of No-Hair Theorems and Novel Black-Hole Solutions in Gauss-Bonnet Theories, *Phys. Rev. Lett.* **120**, p. 131102 (2018).
28. P. V. P. Cunha, C. A. R. Herdeiro and E. Radu, Spontaneously Scalarized Kerr Black Holes in Extended Scalar-Tensor–Gauss-Bonnet Gravity, *Phys. Rev. Lett.* **123**, p. 011101 (2019).
29. L. G. Collodel, B. Kleihaus, J. Kunz and E. Berti, Spinning and excited black holes in Einstein-scalar-Gauss–Bonnet theory, *Class. Quant. Grav.* **37**, p. 075018 (2020).
30. A. Dima, E. Barausse, N. Franchini and T. P. Sotiriou, Spin-induced black hole spontaneous scalarization, *Phys. Rev. Lett.* **125**, p. 231101 (2020).

31. S. Hod, Onset of spontaneous scalarization in spinning Gauss-Bonnet black holes, *Phys. Rev. D* **102**, p. 084060 (2020).
32. C. A. Herdeiro, E. Radu, H. O. Silva, T. P. Sotiriou and N. Yunes, Spin-induced scalarized black holes (9 2020).
33. E. Berti, L. G. Collodel, B. Kleihaus and J. Kunz, Spin-induced black-hole scalarization in Einstein-scalar-Gauss-Bonnet theory (9 2020).
34. T. Anson, E. Babichev, C. Charmousis and M. Hassaine, Disforming the Kerr metric, *JHEP* **01**, p. 018 (2021).
35. J. Ben Achour, H. Liu, H. Motohashi, S. Mukohyama and K. Noui, On rotating black holes in DHOST theories, *JCAP* **11**, p. 001 (2020).
36. K. D. Kokkotas and B. G. Schmidt, Quasinormal modes of stars and black holes, *Living Rev. Rel.* **2**, p. 2 (1999).
37. H.-P. Nollert, TOPICAL REVIEW: Quasinormal modes: the characteristic 'sound' of black holes and neutron stars, *Class. Quant. Grav.* **16**, R159 (1999).
38. L. Rezzolla, Gravitational waves from perturbed black holes and relativistic stars, *ICTP Lect. Notes Ser.* **14**, 255 (2003).
39. E. Berti, V. Cardoso and A. O. Starinets, Quasinormal modes of black holes and black branes, *Class. Quant. Grav.* **26**, p. 163001 (2009).
40. R. A. Konoplya and A. Zhidenko, Quasinormal modes of black holes: From astrophysics to string theory, *Rev. Mod. Phys.* **83**, 793 (2011).
41. J. L. Blázquez-Salcedo, Z. Altaha Motahar, D. D. Doneva, F. S. Khoo, J. Kunz, S. Mojica, K. V. Staykov and S. S. Yazadjiev, Quasinormal modes of compact objects in alternative theories of gravity, *Eur. Phys. J. Plus* **134**, p. 46 (2019).
42. R. Gervalle and M. S. Volkov, Asymptotically flat hairy black holes in massive bigravity, *Phys. Rev. D* **102**, p. 124040 (2020).
43. S. F. Hassan and R. A. Rosen, Bimetric Gravity from Ghost-free Massive Gravity, *JHEP* **02**, p. 126 (2012).
44. M. S. Volkov, Hairy black holes in the ghost-free bigravity theory, *Phys. Rev. D* **85**, p. 124043 (2012).
45. R. Brito, V. Cardoso and P. Pani, Massive spin-2 fields on black hole spacetimes: Instability of the Schwarzschild and Kerr solutions and bounds on the graviton mass, *Phys. Rev. D* **88**, p. 023514 (2013).
46. F. Torsello, M. Kocic and E. Mortsell, Classification and asymptotic structure of black holes in bimetric theory, *Phys. Rev. D* **96**, p. 064003 (2017).
47. F. Moulin, A. Barrau and K. Martineau, An overview of quasinormal modes in modified and extended gravity, *Universe* **5**, p. 202 (2019).
48. B. F. Schutz and C. M. Will, BLACK HOLE NORMAL MODES: A SEMIANALYTIC APPROACH, *Astrophys. J. Lett.* **291**, L33 (1985).
49. S. Iyer and C. M. Will, Black Hole Normal Modes: A WKB Approach. 1. Foundations and Application of a Higher Order WKB Analysis of Potential Barrier Scattering, *Phys. Rev. D* **35**, p. 3621 (1987).
50. S. Iyer, BLACK HOLE NORMAL MODES: A WKB APPROACH. 2. SCHWARZSCHILD BLACK HOLES, *Phys. Rev. D* **35**, p. 3632 (1987).
51. C. de Rham and G. Gabadadze, Generalization of the Fierz-Pauli Action, *Phys. Rev. D* **82**, p. 044020 (2010).
52. C. de Rham, Massive Gravity, *Living Rev. Rel.* **17**, p. 7 (2014).
53. J. W. Moffat, Black Holes in Modified Gravity (MOG), *Eur. Phys. J. C* **75**, p. 175 (2015).
54. P. Horava, Quantum Gravity at a Lifshitz Point, *Phys. Rev. D* **79**, p. 084008 (2009).

55. E. O Colgain and H. Yavartanoo, Dyonic solution of Horava-Lifshitz Gravity, *JHEP* **08**, p. 021 (2009).
56. A. Kehagias and K. Sfetsos, The Black hole and FRW geometries of non-relativistic gravity, *Phys. Lett. B* **678**, 123 (2009).
57. H. Lu, J. Mei and C. N. Pope, New Black Holes in Five Dimensions, *Nucl. Phys. B* **806**, 436 (2009).
58. P. Bargueño, S. Bravo Medina, M. Nowakowski and D. Batic, Quantum Mechanical Corrections to the Schwarzschild Black Hole Metric, *EPL* **117**, p. 60006 (2017).
59. E. Alesci and L. Modesto, Hawking radiation from loop black holes, *J. Phys. Conf. Ser.* **360**, p. 012036 (2012).
60. S. Murk and D. R. Terno, Spherically symmetric black holes in metric gravity, *Phys. Rev. D* **104**, p. 064048 (2021).
61. V. P. Frolov, Information loss problem and a 'black hole' model with a closed apparent horizon, *JHEP* **05**, p. 049 (2014).
62. A. A. Starobinsky, A New Type of Isotropic Cosmological Models Without Singularity, *Phys. Lett. B* **91**, 99 (1980).
63. N. M. Santos and C. A. R. Herdeiro, Black holes, stationary clouds and magnetic fields, *Phys. Lett. B* **815**, p. 136142 (2021).
64. W. E. East, F. M. Ramazanoğlu and F. Pretorius, Black Hole Superradiance in Dynamical Spacetime, *Phys. Rev. D* **89**, p. 061503 (2014).
65. C. A. R. Herdeiro and E. Radu, Dynamical Formation of Kerr Black Holes with Synchronized Hair: An Analytic Model, *Phys. Rev. Lett.* **119**, p. 261101 (2017).
66. G. W. Gibbons, A. H. Mujtaba and C. N. Pope, Ergoregions in Magnetised Black Hole Spacetimes, *Class. Quant. Grav.* **30**, p. 125008 (2013).
67. P. Nicolini, E. Spallucci and M. F. Wondrak, Quantum Corrected Black Holes from String T-Duality, *Phys. Lett. B* **797**, p. 134888 (2019).
68. M. F. Wondrak and M. Bleicher, Constraints on the String T-Duality Propagator from the Hydrogen Atom, *Symmetry* **11**, p. 1478 (2019).
69. M. Fontanini, E. Spallucci and T. Padmanabhan, Zero-point length from string fluctuations, *Phys. Lett. B* **633**, 627 (2006).
70. T. Padmanabhan, Duality and zero point length of space-time, *Phys. Rev. Lett.* **78**, 1854 (1997).
71. B. Carr and F. Kühnel, Primordial Black Holes as Dark Matter: Recent Developments, *Ann. Rev. Nucl. Part. Sci.* **70**, 355 (2020).
72. S. Kalita and B. Mukhopadhyay, Asymptotically flat vacuum solution in modified theory of Einstein's gravity, *Eur. Phys. J. C* **79**, p. 877 (2019).
73. S. Xavier, J. Mathew and S. Shankaranarayanan, Infinitely degenerate exact Ricci-flat solutions in f(R) gravity, *Class. Quant. Grav.* **37**, p. 225006 (2020).
74. A. W. Beckwith, Does the penrose suggestion as to black holes from a prior universe showing up in today's universe have credibility? examined from a singular, and nonsingular beginning of cosmological expansion, *Journal of High Energy Physics, Gravitation and Cosmology* **7**, 149 (2021).
75. R. Penrose, Before the big bang: An outrageous new perspective and its implications for particle physics, *Conf. Proc. C* **060626**, 2759 (2006).
76. D. An, K. A. Meissner, P. Nurowski and R. Penrose, Apparent evidence for Hawking points in the CMB Sky, *Mon. Not. Roy. Astron. Soc.* **495**, 3403 (2020).
77. N. J. Popławski, Black Hole Genesis and origin of cosmic acceleration (12 2019).
78. M. Srivastava, Y. Chen and S. Shankaranarayanan, Analytical computation of quasinormal modes of slowly rotating black holes in dynamical Chern-Simons gravity, *Phys. Rev. D* **104**, p. 064034 (2021).

79. R. Jackiw and S. Y. Pi, Chern-Simons modification of general relativity, *Phys. Rev. D* **68**, p. 104012 (2003).
80. N. Yunes and F. Pretorius, Dynamical Chern-Simons Modified Gravity. I. Spinning Black Holes in the Slow-Rotation Approximation, *Phys. Rev. D* **79**, p. 084043 (2009).
81. J. a. L. Rosa, J. P. S. Lemos and F. S. N. Lobo, Stability of Kerr black holes in generalized hybrid metric-Palatini gravity, *Phys. Rev. D* **101**, p. 044055 (2020).
82. G. J. Olmo, Palatini Approach to Modified Gravity: f(R) Theories and Beyond, *Int. J. Mod. Phys. D* **20**, 413 (2011).
83. T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Metric-palatini gravity unifying local constraints and late-time cosmic acceleration, *Phys. Rev. D* **85**, p. 084016 (Apr 2012).
84. S. Capozziello, T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Hybrid metric-palatini gravity, *Universe* **1**, 199 (2015).
85. N. Tamanini and C. G. Böhmer, Generalized hybrid metric-palatini gravity, *Phys. Rev. D* **87**, p. 084031 (Apr 2013).
86. T. Harko and F. S. N. Lobo, *Extensions of f(R) Gravity: Curvature-Matter Couplings and Hybrid Metric-Palatini Theory* (Cambridge University Press, 11 2018).
87. S. Gera and S. Sengupta, Magnetic monopole as a spacetime defect, *Phys. Rev. D* **104**, p. 044038 (2021).
88. P. A. M. Dirac, Quantised singularities in the electromagnetic field, *Proc. Roy. Soc. Lond. A* **133**, 60 (1931).
89. P. A. M. Dirac, The Theory of magnetic poles, *Phys. Rev.* **74**, 817 (1948).
90. A. A. Tseytlin, On the First Order Formalism in Quantum Gravity, *J. Phys. A* **15**, p. L105 (1982).
91. R. K. Kaul and S. Sengupta, Degenerate spacetimes in first order gravity, *Phys. Rev. D* **93**, p. 084026 (2016).
92. R. K. Kaul and S. Sengupta, New solutions in pure gravity with degenerate tetrads, *Phys. Rev. D* **94**, p. 104047 (2016).
93. D. Rubiera-Garcia, From fundamental physics to tests with compact objects in metric-affine theories of gravity, *Int. J. Mod. Phys. D* **29**, p. 2041007 (2020).
94. C. Bernard and M. Ghezelbash, Hidden conformal symmetry of the rotating charged AdS black holes in quadratic $f(T)$ gravity, *Phys. Rev. D* **101**, p. 024020 (2020).
95. M. Guica, T. Hartman, W. Song and A. Strominger, The Kerr/CFT Correspondence, *Phys. Rev. D* **80**, p. 124008 (2009).
96. J. L. Cardy, Operator Content of Two-Dimensional Conformally Invariant Theories, *Nucl. Phys. B* **270**, 186 (1986).
97. H. W. J. Bloete, J. L. Cardy and M. P. Nightingale, Conformal Invariance, the Central Charge, and Universal Finite Size Amplitudes at Criticality, *Phys. Rev. Lett.* **56**, 742 (1986).
98. V. P. Frolov and K. S. Thorne, Renormalized Stress - Energy Tensor Near the Horizon of a Slowly Evolving, Rotating Black Hole, *Phys. Rev. D* **39**, 2125 (1989).
99. G. Compère, The Kerr/CFT correspondence and its extensions, *Living Rev. Rel.* **15**, p. 11 (2012).
100. H. Afshar, D. Grumiller, W. Merbis, A. Perez, D. Tempo and R. Troncoso, Soft hairy horizons in three spacetime dimensions, *Phys. Rev. D* **95**, p. 106005 (2017).
101. H. Afshar, D. Grumiller and M. M. Sheikh-Jabbari, Near horizon soft hair as microstates of three dimensional black holes, *Phys. Rev. D* **96**, p. 084032 (2017).
102. H. Afshar, S. Detournay, D. Grumiller, W. Merbis, A. Perez, D. Tempo and R. Troncoso, Soft Heisenberg hair on black holes in three dimensions, *Phys. Rev. D* **93**, p. 101503 (2016).

103. G. Compère, K. Hajian, A. Seraj and M. M. Sheikh-Jabbari, Wiggling Throat of Extremal Black Holes, *JHEP* **10**, p. 093 (2015).
104. G. Compère, K. Hajian, A. Seraj and M. M. Sheikh-Jabbari, Extremal Rotating Black Holes in the Near-Horizon Limit: Phase Space and Symmetry Algebra, *Phys. Lett. B* **749**, 443 (2015).
105. K. Hajian, M. M. Sheikh-Jabbari and H. Yavartanoo, Extreme Kerr black hole microstates with horizon fluff, *Phys. Rev. D* **98**, p. 026025 (2018).
106. K. Hajian and A. Seraj, Symplectic Structure of Extremal Black Holes, *Springer Proc. Phys.* **208**, 61 (2018).
107. B. Chen, A. M. Ghezelbash, V. Kamali and M. R. Setare, Holographic description of Kerr-Bolt-AdS-dS Spacetimes, *Nucl. Phys. B* **848**, 108 (2011).
108. A. M. Ghezelbash, V. Kamali and M. R. Setare, Hidden Conformal Symmetry of Kerr-Bolt Spacetimes, *Phys. Rev. D* **82**, p. 124051 (2010).
109. A. M. Awad, G. G. L. Nashed and W. El Hanafy, Rotating charged AdS solutions in quadratic $f(T)$ gravity, *Eur. Phys. J. C* **79**, p. 668 (2019).
110. N. Popławski, Gravitational Collapse of a Fluid with Torsion into a Universe in a Black Hole, *J. Exp. Theor. Phys.* **132**, 374 (2021).
111. D. Sciama, On a Nonsymmetric theory of the pure gravitational field, *Proc. Cambridge Phil. Soc.* **54**, p. 72 (1958).
112. D. W. Sciama, The Physical structure of general relativity, *Rev. Mod. Phys.* **36**, 463 (1964), [Erratum: Rev. Mod. Phys. 36, 1103–1103 (1964)].
113. T. W. B. Kibble, Lorentz invariance and the gravitational field, *J. Math. Phys.* **2**, 212 (1961).
114. F. W. Hehl and B. K. Datta, Nonlinear spinor equation and asymmetric connection in general relativity, *J. Math. Phys.* **12**, 1334 (1971).
115. F. W. Hehl, P. Von Der Heyde, G. D. Kerlick and J. M. Nester, General Relativity with Spin and Torsion: Foundations and Prospects, *Rev. Mod. Phys.* **48**, 393 (1976).
116. F. W. Hehl, G. D. Kerlick and P. Von Der Heyde, General relativity with spin and torsion and its deviations from einstein's theory, *Phys. Rev. D* **10**, 1066 (1974).
117. R. C. Tolman, Effect of imhomogeneity on cosmological models, *Proc. Nat. Acad. Sci.* **20**, 169 (1934).
118. R. B. Magalhães, L. C. S. Leite and L. C. B. Crispino, Absorption by deformed black holes, *Phys. Lett. B* **805**, p. 135418 (2020).
119. T. Johannsen and D. Psaltis, A Metric for Rapidly Spinning Black Holes Suitable for Strong-Field Tests of the No-Hair Theorem, *Phys. Rev. D* **83**, p. 124015 (2011).
120. R. Konoplya and A. Zhidenko, Detection of gravitational waves from black holes: Is there a window for alternative theories?, *Phys. Lett. B* **756**, 350 (2016).
121. J. Badía and E. F. Eiroa, Influence of an anisotropic matter field on the shadow of a rotating black hole, *Phys. Rev. D* **102**, p. 024066 (2020).
122. K. Akiyama *et al.*, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, p. L1 (2019).
123. V. Perlick and O. Y. Tsupko, Calculating black hole shadows: review of analytical studies (5 2021).
124. V. Perlick, O. Y. Tsupko and G. S. Bisnovatyi-Kogan, Influence of a plasma on the shadow of a spherically symmetric black hole, *Phys. Rev. D* **92**, p. 104031 (2015).
125. O. Y. Tsupko, Deflection of light rays by a spherically symmetric black hole in a dispersive medium, *Phys. Rev. D* **103**, p. 104019 (2021).
126. V. Perlick and O. Y. Tsupko, Light propagation in a plasma on Kerr spacetime: Separation of the Hamilton-Jacobi equation and calculation of the shadow, *Phys. Rev. D* **95**, p. 104003 (2017).

127. O. Y. Tsupko and G. S. Bisnovatyi-Kogan, Gravitational lensing in the presence of plasmas and strong gravitational fields, *Grav. Cosmol.* **20**, 220 (2014).
128. I. Cho and H.-C. Kim, Simple black holes with anisotropic fluid, *Chin. Phys. C* **43**, p. 025101 (2019).
129. H.-C. Kim, B.-H. Lee, W. Lee and Y. Lee, Rotating black holes with an anisotropic matter field, *Phys. Rev. D* **101**, p. 064067 (2020).
130. J. Badia and E. F. Eiroa, Shadow of axisymmetric, stationary and asymptotically flat black holes in the presence of plasma (6 2021).

Uncertainties in kilonova modeling

C. L. Fryer*, C. J. Fontes, O. Korobkin, M. Mumpower and R. Wollaeger
Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
**E-mail: fryer@lanl.gov*
https://ccsweb.lanl.gov/astro/index.html

E.M. Holmbeck
The Observatories of the Carnegie Institute for Science, Pasadena, CA, 91101, USA

R. O'Shaugnessy
Rochester Institute of Technology, Rochester, NY 14623

The detection of the merger of a neutron star binary in both gravitational waves and a broad spectrum of electromagnetic waves (GW170817) provided the most compelling evidence to date that such mergers produce heavy r-process elements. The inferred rate of these mergers coupled to the estimated r-process production suggests that these mergers could produce nearly all of the r-process elements in the universe. However, uncertainties in the merger rate and the amount of r-process production per merger means that scientists can not constrain the fraction of the merger r-process contribution to better than 1–100% of the total amount in the universe. The total r-process mass synthesized is best constrained by the observations themselves and uncertainties in the inferred production quantity follows from the uncertainties in modeling the emission from the NSM ejecta. In this paper, we review these modeling uncertainties.

Keywords: Neutron Stars; Gravitational Waves, r-Process

1. Kilonovae, a Key to Understanding r-Process

Many of the heaviest elements in the universe are produced through a process of rapid neutron capture (r-process), building isotopes far from stability that then decay to the long-lived heavy isotopes observed in the universe. This rapid neutron capture requires extreme conditions with high densities and large fractions of free neutrons. Scientists identified neutron star mergers (NSMs) as a potential source of these r-process elements over 40 years ago[1,2] and evidence has been growing steadily to argue that these mergers could dominate the r-process production.

Although the evolution of the ejecta from these mergers reaches ideal conditions (high densities, temperatures and neutron fractions) for the production of r-process elements, observational validation of r-process production in NSMs has been much more difficult to obtain. Until recently, theoretical and observational evidence relied on galactic chemical evolution arguments and their associated uncertainties.[3,4]

This situation changed with the joint electromagnetic- and gravitational-wave observations of NSM GW170817.[5] Between the gamma-, x-, optical, infra-red, and radio emission, it became clear that the NSM in GW170817 produced both a relativistic jet and a lower-velocity mass ejecta that argued for both neutron-rich dynamical ejecta and higher electron-fraction wind ejecta components.

Although initial analyses argued strongly that GW170817 produced a large fraction of r-process ($0.04M_\odot$), as more studies were completed, uncertainties in the inferred r-process production from GW170817 grew.[6] Many of these differences arose from different assumptions about the properties of the outflow material and a difference in which observations were used to constrain the data. But these are just some of the uncertainties in the analysis of the optical and infra-red emission from NSMs. In this paper, we review these uncertainties. In section 2, we review the uncertainties in the properties (distributions of the velocity, density, entropy and neutron fraction) of the outflow from these mergers. Section 3 reviews some of the microphysics uncertainties in these calculations. Section 4 details alternative power sources that can alter the mass inferred from observations.

2. Uncertainties in the Outflows

The emission from astrophysical transients depends upon the ejecta properties (e.g. composition, velocity and angular distributions). Unlike many transients that are produced by a single outburst, the ejecta from NSMs arises from multiple components including ejecta from the initial tidal tails produced in the dynamical merger and a outflow arising from a disk. The disk outflow is driven by viscous forces and neutrino emission from a post-disruption torus of material (Figure 1). We will refer to the former as *dynamical* ejecta and the latter as *wind* ejecta. The bulk of the dynamical material is ejected along the orbital plane. This material has a high-neutron fraction, ideally suited for the production of the most massive elements (heavy r-process). In contrast, the wind ejecta is more isotropic and consists of material spanning a broad range of neutron fractions that depend on both the properties of the disk and the fate of the merged core.

Many of the initial studies of GW170817 used spherically-symmetric models, varying the fraction of heavy r-process to lighter elements.[6] Some of these initial models used two components to differentiate between the dynamical ejecta primarily flowing along the orbital plane and a more symmetric or axis-aligned wind outflow[7] shown in Figure 1. However, even these two-component models made simplifying assumptions using a single composition for each component and a fixed velocity distribution of the ejecta. These assumptions oversimplify the nature of the ejecta and can have order-of-magnitude effects on the inferred mass of r-process ejecta. Before we study how these two components fit together, we first review some of the outflow uncertainties in the individual dynamical and wind components.

Although models of the dynamical ejecta have produced a range of ejecta masses, e.g. compare the results in,[8,9] these numerical differences are slowly diminishing as

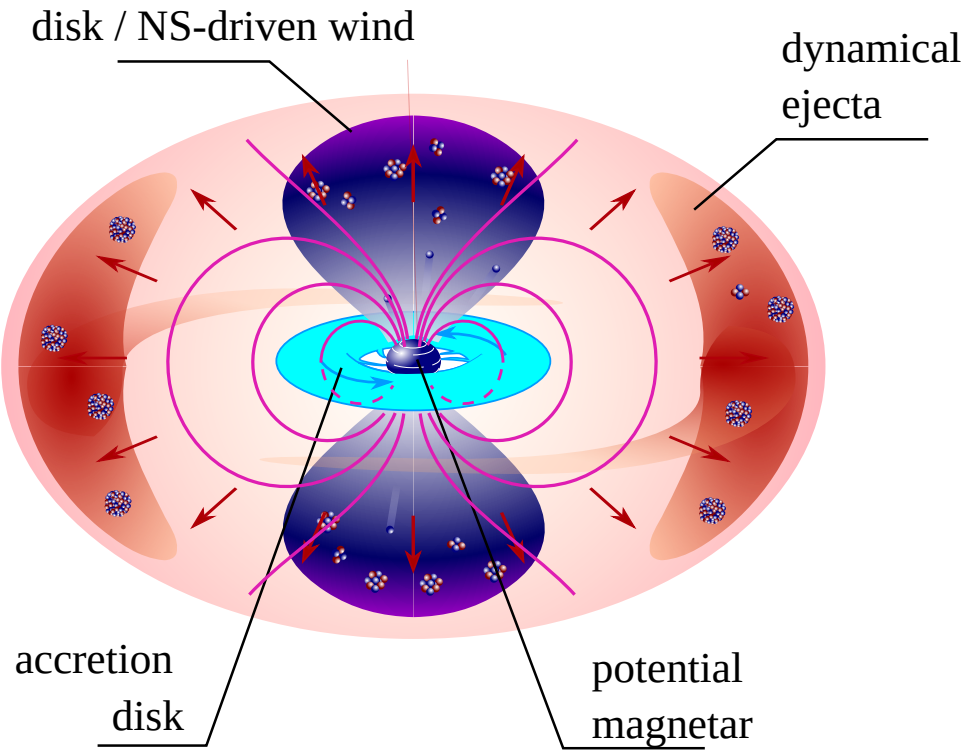


Fig. 1. Anatomy of the NSM outflows. The material ejected in the merger phase flows predominantly along the orbital plane and is very neutron rich. The merged object produces a neutron star or black hole core surrounded by a torus of high angular-momentum material. Outflows driven by viscous forces in this disk, coupled with neutrino emission, produce the "wind" ejecta with its lower neutron fraction. These outflows produce the observed kilonova light-curve emission from the ultra-violet to the infra-red powered by radioactive decay. When material falls back on this merged core and/or if a magnetar forms, an additional power source can contribute additional energy to the observed transient.

computational methods improve. Even with this convergence, the ejecta mass will vary with the individual masses of the merging compact objects.[8] The composition of this ejecta is much better constrained than the ejecta mass. At high neutron fractions (above 0.8), the final composition does not depend sensitively on the neutron fraction. Because of this, the small variation in the neutron fraction does not alter the Lanthanide composition considerably and the composition of this ejecta is fairly stable. In addition, the final composition is also less sensitive to the details of the outflow (density and temperature evolution). For this reason, the composition of this ejecta is well-understood with the dominant uncertainties depending on our lack of understanding of the nuclear physics (Sec. 3). Finally, due to the dense forest of absorption lines in the opacity produced by the heavy r-process lanthanides and actinides, the light curves do not depend sensitively on the exact relative abundance

of these elements.[10] The primary uncertainties in models of the dynamical ejecta emission lie in our understanding of the total ejecta mass and outflow properties (velocity distribution).

The properties of the wind ejecta are much more uncertain. The mass, velocity distribution and composition of this ejecta varies considerably from model to model. Whereas the neutron fraction is high and exhibits only small variations for dynamical ejecta, the neutron fraction ranges across a broad range in the wind ejecta. Further, at the lower neutron fractions (between 0.5-0.8), the yields become much more sensitive to the exact value of the neutron fraction, so the composition of the wind ejecta can vary wildly. To further complicate these light-curve models, disk models of this wind predict a range of compositions that vary with angle and time.[11] This composition variation is not taken into account in most of the two-component models currently in the literature. In addition, current wind models typically only follow the wind out to, at most, 1 second and much of the ejecta is below the escape velocity (but still accelerating) at this time. Although models are being constructed to follow this outflow to later times, we still do not have accurate models of this ejecta's velocity distribution.

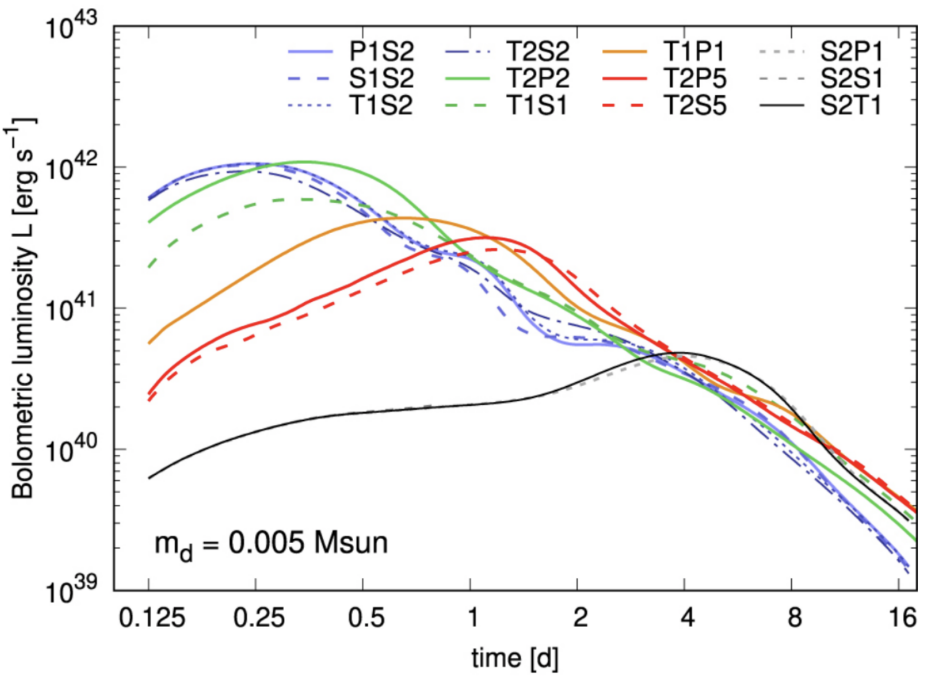


Fig. 2. Bolometric Light curves versus time for a broad range of models varying the morphology of the two components in a dynamical plus wind model.[12] Varying the morphology alone can change the bolometric luminosity by over an order of magnitude.

Although we have mostly discussed the role of the composition, the morphology of the ejecta (velocity and angular distribution) may be even more important. Figure 2 shows a broad set of models with the same explosion energy and mass, but varying the angular morphology of the ejecta.[12] In this study, two-component models were developed, each with a different morphology. In these models, although the energy of the ejecta is the same, the velocity distribution of the ejecta can vary. Changing the morphology alters both the peak luminosity and the time of that peak, varying both by an order of magnitude.

At this time, a complete study of all the outflow uncertainties has not been done. We end this section with the discussion of a final study varying ejecta mass and velocity of a two-component model.[13] This grid of models used a single dynamical ejecta composition and one of two wind compositions. It focused on a primary pair of morphologies for these two components, using a default r-process composition for the dynamical ejecta and two different compositions for the wind. But the compositions were fixed for each component. Although the peak velocity was changed, the velocity distribution was unaltered. Depending on the relative velocities of the two components, the dynamical ejecta, with its high-opacity lanthanides, can obscure the wind ejecta. In these cases, the luminosity can depend sensitively on the viewing angle. In other cases, the wind ejecta is sufficiently fast that the light-curve is similar, regardless of viewing angle. Although more massive ejecta models are, on average, brighter, lower mass models can be brighter than higher mass models based on this viewing angle.

3. Uncertainties in the Microphysics and its Implementation

The outflow models depend on detailed general-relativistic, magneto-hydrodynamic models including neutrino transport (and the related neutrino microphysics). We do not have enough space to discuss all of this microphysics in detail. Instead we focus on the uncertainties in the nuclear physics and atomic physics with a brief discussion on each.

Unstable neutron-rich nuclei participating in the r-process have many unmeasured properties that consequently influence kilonova signals.[14,15] Uncertainties in nuclear binding energies and reaction rates impact the flow of material to heavier mass regions as well as the subsequent flow back towards lighter mass regions in ejecta with low electron fraction.[16] Properties that depend on excited nuclear states, including half-lives and branching ratios, additionally control the timescale over which energy is released as well as potentially stored.[17,18] The creation of the heaviest elements in merger events[19] may produce distinct observable signals, e.g. via the production of long-lived species like ^{254}Cf or via unique electromagnetic signatures.[20,21,21,22] Central to the reduction of large nuclear uncertainties that impact kilonovae are experimental studies undertaken at radioactive beam facilities around the world, see e.g.[23,24]

A major uncertainty in the application of atomic physics to the modeling of kilonovae concerns the bound-bound (line absorption) contribution to the radiative opacity. While photons traveling through a plasma can be absorbed in a number of atomic processes, line absorption is the dominant mechanism in the dynamical ejecta due to the dense forest of lines associated with near-neutral lanthanide and actinide elements. Complete sets of lanthanide opacities have been recently produced for the purpose of kilonova modeling[25–27] using different atomic physics methods, which is another potential source of uncertainty. In an effort to mitigate this issue, we have made available to the public our lanthanide opacites[28] so that kilonova modelers can use a consistent set of opacities, thereby ruling out emission differences that could arise from atomic physics implementations. As a specific illustration of these concepts, we present in Figure 3 the opacity of Nd, calculated under the assumption of local thermodynamic equilibrium (LTE), at a temperature of $T = 0.5$ eV and mass density $\rho = 10^{-13}$ g/cm^3, using four different models (see[26] for details). Note that the line features enhance the opacity by six to eight orders of magnitude compared to what would occur if only free-electron processes were included. The Planck mean opacity, $\kappa^{\rm P}$, is also displayed for each model, indicating a maximum discrepancy of about 40% between the various models for this integrated quantity. Detailed light-curve and spectral comparisons were carried out with these four models.[26] Only

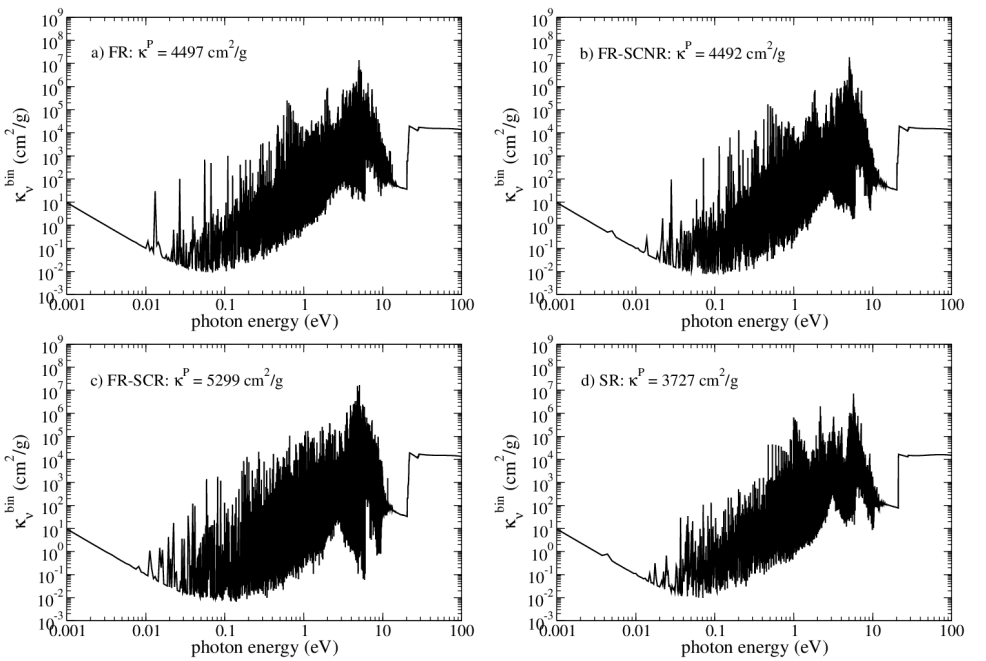


Fig. 3. The LTE line-binned opacity for neodymium at $T = 0.5$ eV and $\rho = 10^{-13}$ g/cm^3 using four different models described in:[26] a) FR, b) FR-SCNR, c) FR-SCR, and d) SR. The Planck mean opacity, obtained via integration of the line-binned opacity, is also listed in each panel.

modest differences were noted in the light-curves, i.e. differences of 10–20% in the peak of the light curve and a maximum shift of half a day in the time of peak luminosity. Similar, modest differences were observed in the spectra produced with these models, i.e. certain spectral features are shifted redward are blueward, but the overall spectral characteristics are similar.

From a more general perspective, we note that opacities are typically calculated from first principles, rather than measured. The line contribution depends on quantities such as transition energies, line strengths (quantum mechanical matrix elements) and atomic level populations. While it is beyond the scope of this article to provide a detailed discussion of these concepts, we mention that recent work[29,30] has been performed to compare calculated energies with benchmark values in the NIST Atomic Spectra Database[31] for a number of important, low-lying levels of the lanthanides. Also, most simulations employ the LTE approximation when calculating the level populations. This assumption is expected to be valid up to about one week post merger. But, as with many astrophysical transients, the LTE conditions become less and less valid as the expansion continues and the density becomes so low that the thermalization timescale from collisions becomes long compared to evolutionary timescales of the ejecta. A recent study[32] explores the possibility that non-thermal, β-decay electrons, produced from the radioactive decay of r-process nuclei, are primarily responsible for the heating and collisional ionization in the ejecta.

With the large number of atomic levels, the number of line features is enormous (in the 10s of millions). Implementing these opacities into a numerical calculation is intractable even from a computational memory standpoint. In addition, because of the high velocities in the ejecta, modeling radiation transport becomes even more difficult. For a single strong line, the Sobolev approximation can be used to determine the likelihood of a photon to both be absorbed as it expands out of a medium and its likelihood of escaping this line when re-emitted. Transient explosions exhibit homologous outflow conditions where the velocity of the ejecta is proportional to the radius. As a photon moves out through this ejecta, its energy spans a broad range with respect to the rest frame of the outflow. Even if the photon does not interact with a line feature at one radius, it may interact with this feature at a different radius.

This treatment includes the fact that the velocity of the ejecta spans a range such that, even if a photon does not interact with an absorption feature in the rest frame of the outflowing material at one position, it may, as it transports outward, interact with other material. Hence, a single line feature can affect a larger energy band of photons. On the other hand, if most of the re-emitted photons arise from a direct de-excitation of the absorption feature, the photons remain trapped in a line feature in material moving at a constant velocity. For the homologous outflow of astrophysical transients, the photon can escape more quickly, effectively reducing the opacity. Implementing the Sobolev approximation is already difficult, but in

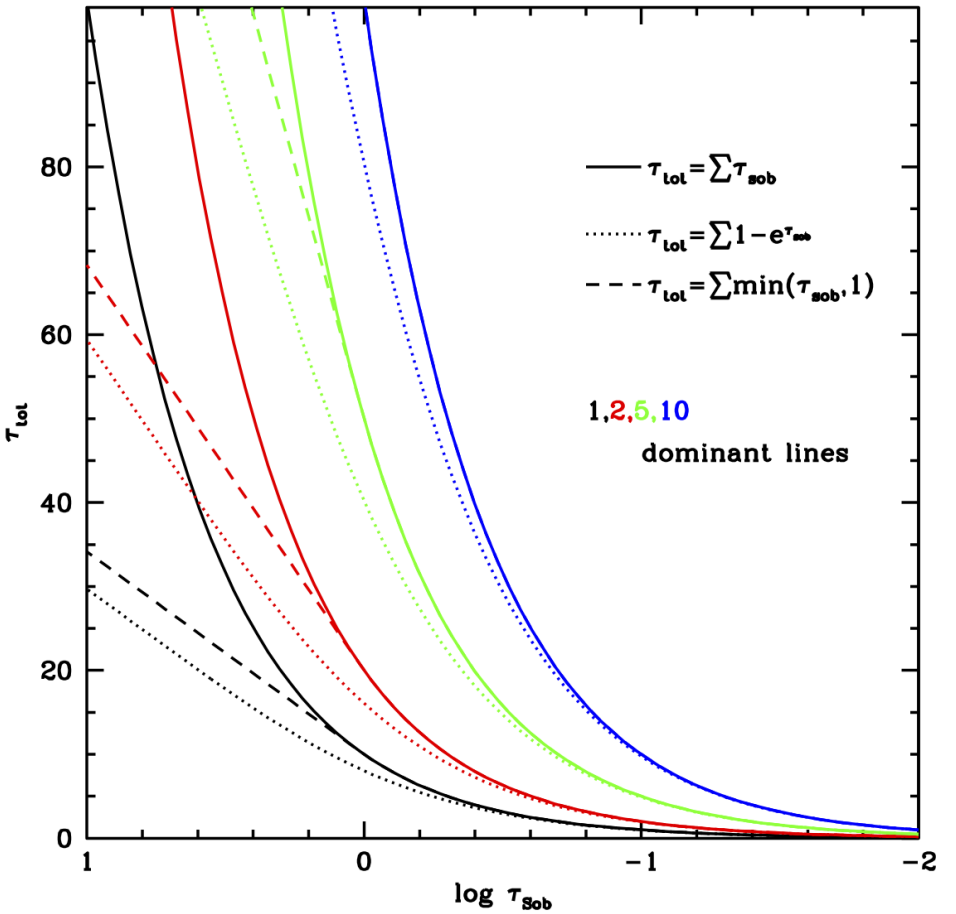


Fig. 4. A number of prescriptions have been proposed to model the expansion opacities. This figure shows the total optical depth using three different prescriptions for different numbers of dominant line features as a function of the optical depth of the line feature. There is some convergence with large number of lines, showing why a simple averaged opacity can work nearly as well as a more complicated expansion procedure.

a medium where a forest of lines exists, producing an accurate model for spectra becomes increasingly difficult. The deficiencies of the Pinto & Eastman expansion opacity[33] were outlined in its inception and a number of alternative approaches have been developed (for a review, see[34]). A number of methods have been developed to calculate the optical depth and these different models produce a range of results based on these prescriptions (Figure 4). For a simplified model with pure neodymium, Figure 5 compares the bolometric luminosity for different approaches: Sobolev, expansion opacity and a simple binned treatment. Compared to many of our other uncertainties, the opacity implementation errors seem low.

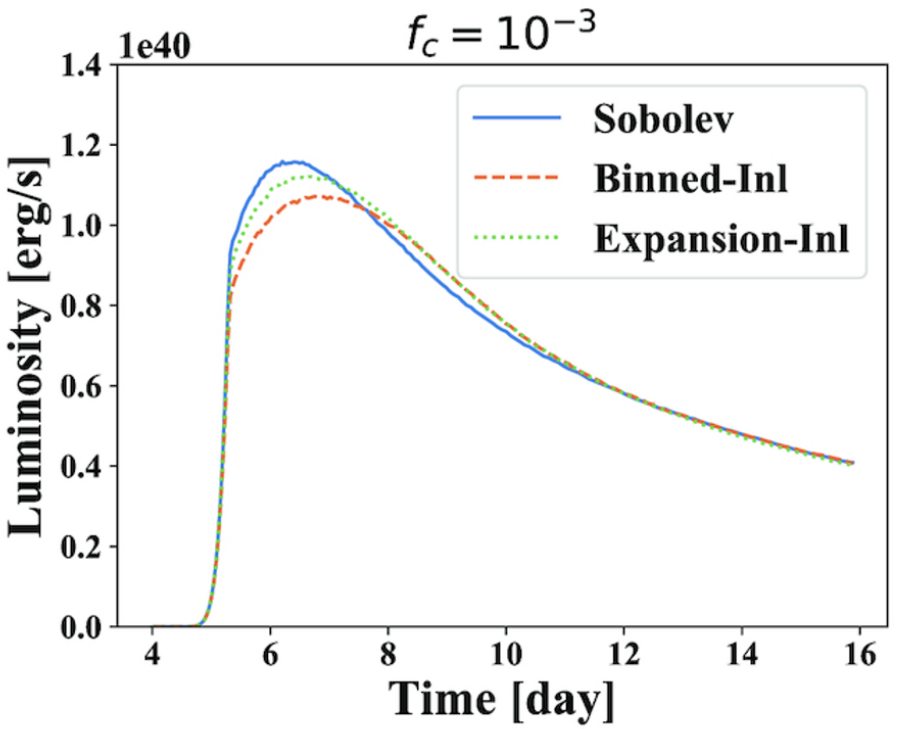


Fig. 5. Bolometric light curves using three different implementations of the opacity. The quantity $f_c = 10^{-3}$ indicates the cut off value for retaining oscillator strengths in the atomic physics model.

4. Alternative Energy Sources

Most calculations assume that the power source behind the kilonova emission arises solely from radioactive decay. But, just as with core-collapse supernovae, the light-curves and spectra can be powered by a variety of sources. For kilonovae, a number of additional power sources have been proposed to drive the observed emission: magnetar emission, shock heating in the ejecta, and fallback accretion. In this section, we will discuss each of these energy sources in turn, reviewing both their basic physics and expected features.

The black hole accretion disk engine paradigm argued that long-duration bursts are produced in stellar collapse, while short-duration GRBs are produced by the merger of compact remnants. Under this paradigm, scientists predicted that short-duration GRBs should be offset from their formation region due to the momentum imparted on compact binaries at birth.[35,36] Although other power sources do not differentiate between short- and long-duration bursts in this manner, magnetar engines for short-duration bursts have remained a competitive model for these bursts. The basic idea behind this engine is that the merged core of a NSM is expected to

be rapidly rotating. If strong, dipole magnetic fields are generated in this merged core, then the rotational energy in the core can power the GRB. Whether or not the magnetar drives the initial burst of gamma-rays, magnetars do provide a natural explanation for the roughly hall of all short-duration bursts that exhibit a long-lived ("plateau") phase in the X-ray that can last up to 10^5s.[37] The magnetar emission, if reprocessed by the expanding ejecta, can power the optical and even infra-red emission,[38,39] and can dominate the observed emission, altering the inferred r-process yield in the merger.

In supernovae, shock heating is believed to dominate the heating in type II supernovae, shock breakout events, and many superluminous supernovae. Shock heating may also play a role in kilonova emission. Shocks between the jet and the wind ejecta have been invoked to explain additional emission at early times (first day) in the kilonova emission.[40] At later times (beyond a day), material falls back onto the compact remnant, likely forming a disk and further outflows. The potential energy released in this fallback can be tapped to provide an additional power source. Depending upon the timing of the fallback, this accretion emission

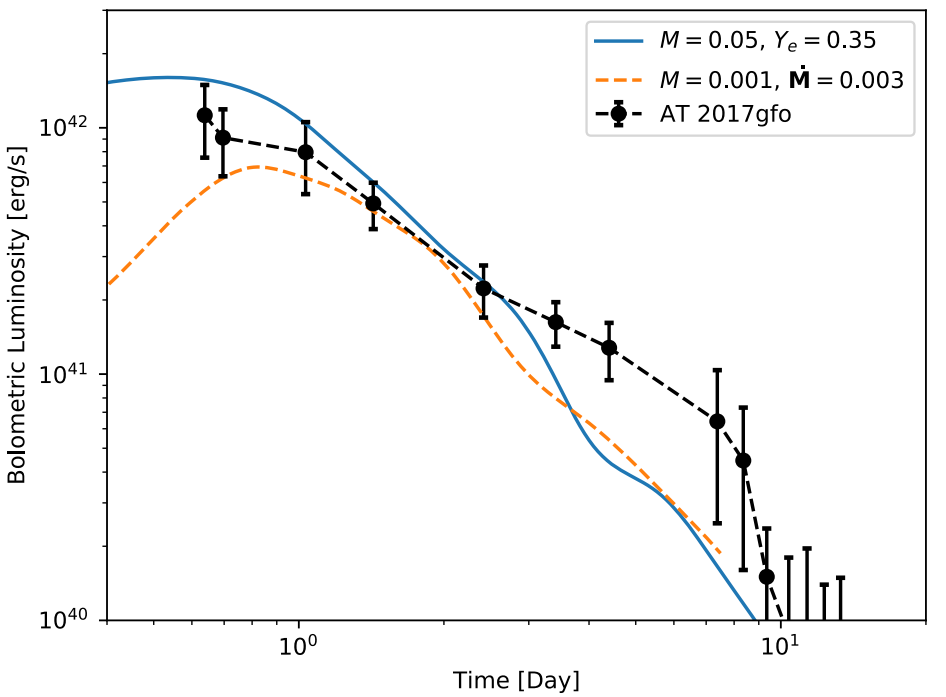


Fig. 6. Bolometric luminosity versus time for a 0.05 $M_\odot$ 1D spherical wind model (blue), a 0.001 $M_\odot$ 1D spherical model with a $\dot{M} = 0.003$ $M_\odot$/s central fallback luminosity source (dashed orange), and the EM counterpart of GW170817, AT 2017gfo (dashed black, with error bars).

can play a major role in the emission after the first day (Figure 6). The accretion luminosity is proportional to the potential energy released:

$$L_{\rm acc} \propto G M_{\rm core} \dot{M}_{\rm fallback} / r_{\rm ISCO} \quad (1)$$

where G is the gravitational constant, $M_{\rm core}$ is the mass of the merged core, $r_{\rm ISCO}$ is its innermost stable circular orbit if the core is a black hole or the neutron star radius if it is a neutron star, and $\dot{M}_{\rm fallback}$ is the fallback accretion rate. Whether the merged core is a black hole or a neutron star, the fallback material is likely to have sufficient angular momentum to initially be centrifugally supported in an accretion disk, providing a mechanism by which a fraction of the potential energy released in the infall can be converted to both energy and mass ejection.

5. Conclusions

In addition to being the primary mechanism behind short-duration gamma-ray bursts, NSMs have the potential to be the dominant source of r-process elements in the universe. To validate this claim, scientists must be able to accurately infer the mass and abundances from observations of the ejecta-driven transient from these mergers (a.k.a. kilonovae). In this paper, we reviewed a broad set of uncertainties affecting the modeling of kilonova light-curves and the current status of our efforts to characterize and, hopefully constrain, them. Much more work must be done to infer accurate r-process masses from these observations.

In this short paper, we focused mostly on broad-band light-curves. Spectra are key to constraining the different models and a number of studies have focused on trying to find line features in the kilonova observations e.g.[41]. Gamma-ray decay lines, particularly in the remnants of the NSM ejecta, also have the potential to probe detailed yields.[21,42] So despite the difficulties that face astronomers in understanding the yields from these mergers, the future is looking bright.

References

1. J. M. Lattimer and D. N. Schramm, Black-Hole-Neutron-Star Collisions, *ApJ Letters* **192**, p. L145 (September 1974).
2. J. M. Lattimer and D. N. Schramm, The tidal disruption of neutron stars by black holes in close binaries., *ApJ* **210**, 549 (December 1976).
3. P. Beniamini, K. Hotokezaka and T. Piran, r-process Production Sites as Inferred from Eu Abundances in Dwarf Galaxies, *ApJ* **832**, p. 149 (December 2016).
4. M. Safarzadeh and E. Scannapieco, Simulating neutron star mergers as r-process sources in ultrafaint dwarf galaxies, *MNRAS* **471**, 2088 (October 2017).
5. B. P. Abbott and a. et, Multi-messenger Observations of a Binary Neutron Star Merger, *ApJ Letters* **848**, p. L12 (October 2017).
6. B. Côté, C. L. Fryer, K. Belczynski, O. Korobkin, M. Chruślińska, N. Vassh, M. R. Mumpower, J. Lippuner, T. M. Sprouse, R. Surman and R. Wollaeger, The Origin of r-process Elements in the Milky Way, *ApJ* **855**, p. 99 (March 2018).

7. R. T. Wollaeger, O. Korobkin, C. J. Fontes, S. K. Rosswog, W. P. Even, C. L. Fryer, J. Sollerman, A. L. Hungerford, D. R. van Rossum and A. B. Wollaber, Impact of ejecta morphology and composition on the electromagnetic signatures of neutron star mergers, *MNRAS* **478**, 3298 (August 2018).
8. O. Korobkin, S. Rosswog, A. Arcones and C. Winteler, On the astrophysical robustness of the neutron star merger r-process, *MNRAS* **426**, 1940 (November 2012).
9. M. Shibata and K. Hotokezaka, Merger and Mass Ejection of Neutron Star Binaries, *Annual Review of Nuclear and Particle Science* **69**, 41 (October 2019).
10. W. Even, O. Korobkin, C. L. Fryer, C. J. Fontes, R. T. Wollaeger, A. Hungerford, J. Lippuner, J. Miller, M. R. Mumpower and G. W. Misch, Composition Effects on Kilonova Spectra and Light Curves. I, *ApJ* **899**, p. 24 (August 2020).
11. J. M. Miller, T. M. Sprouse, C. L. Fryer, B. R. Ryan, J. C. Dolence, M. R. Mumpower and R. Surman, Full Transport General Relativistic Radiation Magnetohydrodynamics for Nucleosynthesis in Collapsars, *ApJ* **902**, p. 66 (October 2020).
12. O. Korobkin, R. T. Wollaeger, C. L. Fryer, A. L. Hungerford, S. Rosswog, C. J. Fontes, M. R. Mumpower, E. A. Chase, W. P. Even, J. Miller, G. W. Misch and J. Lippuner, Axisymmetric Radiative Transfer Models of Kilonovae, *ApJ* **910**, p. 116 (April 2021).
13. R. T. Wollaeger, C. L. Fryer, E. A. Chase, C. J. Fontes, M. Ristic, A. L. Hungerford, O. Korobkin, R. O'Shaughnessy and A. M. Herring, A Broad Grid of 2D Kilonova Emission Models, *ApJ* **918**, p. 10 (September 2021).
14. Y. L. Zhu, K. A. Lund, J. Barnes, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower and R. Surman, Modeling Kilonova Light Curves: Dependence on Nuclear Inputs, *ApJ* **906**, p. 94 (January 2021).
15. J. Barnes, Y. L. Zhu, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower and R. Surman, Kilonovae Across the Nuclear Physics Landscape: The Impact of Nuclear Physics Uncertainties on r-process-powered Emission, *ApJ* **918**, p. 44 (September 2021).
16. B. Côté, M. Eichler, A. Yagüe López, N. Vassh, M. R. Mumpower, B. Világos, B. Soós, A. Arcones, T. M. Sprouse, R. Surman, M. Pignatari, M. K. Pető, B. Wehmeyer, T. Rauscher and M. Lugaro, ^{129}I and ^{247}Cm in meteorites constrain the last astrophysical source of solar r-process elements, *Science* **371**, 945 (February 2021).
17. M. R. Mumpower, T. Kawano and P. Möller, Neutron-γ competition for β -delayed neutron emission, *PRC* **94**, p. 064317 (December 2016).
18. G. W. Misch, T. M. Sprouse and M. R. Mumpower, Astromers in the Radioactive Decay of r-process Nuclei, *ApJ Letters* **913**, p. L2 (May 2021).
19. E. M. Holmbeck, A. Frebel, G. C. McLaughlin, M. R. Mumpower, T. M. Sprouse and R. Surman, Actinide-rich and Actinide-poor r-process-enhanced Metal-poor Stars Do Not Require Separate r-process Progenitors, *ApJ* **881**, p. 5 (August 2019).
20. Y. Zhu, R. T. Wollaeger, N. Vassh, R. Surman, T. M. Sprouse, M. R. Mumpower, P. Möller, G. C. McLaughlin, O. Korobkin, T. Kawano, P. J. Jaffke, E. M. Holmbeck, C. L. Fryer, W. P. Even, A. J. Couture and J. Barnes, Californium-254 and Kilonova Light Curves, *ApJ Letters* **863**, p. L23 (August 2018).
21. O. Korobkin, A. M. Hungerford, C. L. Fryer, M. R. Mumpower, G. W. Misch, T. M. Sprouse, J. Lippuner, R. Surman, A. J. Couture, P. F. Bloser, F. Shirazi, W. P. Even, W. T. Vestrand and R. S. Miller, Gamma Rays from Kilonova: A Potential Probe of r-process Nucleosynthesis, *ApJ* **889**, p. 168 (February 2020).
22. X. Wang, N3AS Collaboration, N. Vassh, FIRE Collaboration, T. Sprouse, M. Mumpower, R. Vogt, J. Randrup and R. Surman, MeV Gamma Rays from Fission: A Distinct Signature of Actinide Production in Neutron Star Mergers, *ApJ Letters* **903**, p. L3 (November 2020).

23. A. Spyrou, S. N. Liddick, F. Naqvi, B. P. Crider, A. C. Dombos, D. L. Bleuel, B. A. Brown, A. Couture, L. Crespo Campo, M. Guttormsen, A. C. Larsen, R. Lewis, P. Möller, S. Mosby, M. R. Mumpower, G. Perdikakis, C. J. Prokop, T. Renstrøm, S. Siem, S. J. Quinn and S. Valenta, Strong Neutron-γ Competition above the Neutron Threshold in the Decay of ^{70}Co, *Physical Review Letters* **117**, p. 142701 (September 2016).
24. O. Hall, T. Davinson, A. Estrade, J. Liu, G. Lorusso, F. Montes, S. Nishimura, V. H. Phong, P. J. Woods, J. Agramunt, D. S. Ahn, A. Algora, J. M. Allmond, H. Baba, S. Bae, N. T. Brewer, C. G. Bruno, R. Caballero-Folch, F. Calviño, P. J. Coleman-Smith, G. Cortes, I. Dillmann, C. Domingo-Pardo, A. Fijalkowska, N. Fukuda, S. Go, C. J. Griffin, R. Grzywacz, J. Ha, L. J. Harkness-Brennan, T. Isobe, D. Kahl, L. H. Khiem, G. G. Kiss, A. Korgul, S. Kubono, M. Labiche, I. Lazarus, J. Liang, Z. Liu, K. Matsui, K. Miernik, B. Moon, A. I. Morales, P. Morrall, M. R. Mumpower, N. Nepal, R. D. Page, M. Piersa, V. F. E. Pucknell, B. C. Rasco, B. Rubio, K. P. Rykaczewski, H. Sakurai, Y. Shimizu, D. W. Stracener, T. Sumikama, H. Suzuki, J. L. Tain, H. Takeda, A. Tarifeño-Saldivia, A. Tolosa-Delgado, M. Wolińska-Cichocka and R. Yokoyama, β-delayed neutron emission of r-process nuclei at the N = 82 shell closure, *Physics Letters B* **816**, p. 136266 (May 2021).
25. D. Kasen, B. Metzger, J. Barnes, E. Quataert and E. Ramirez-Ruiz, Origin of the heavy elements in binary neutron-star mergers from a gravitational wave event, *Nature* **551**, 80 (November 2017).
26. C. J. Fontes, C. L. Fryer, A. L. Hungerford, R. T. Wollaeger and O. Korobkin, A line-binned treatment of opacities for the spectra and light curves from neutron star mergers, *MNRAS* **493**, 4143 (April 2020).
27. M. Tanaka, D. Kato, G. Gaigalas and K. Kawaguchi, Systematic opacity calculations for kilonovae, *MNRAS* **496**, 1369 (August 2020).
28. K. Olsen, C. J. Fontes, C. L. Fryer, A. L. Hungerford, R. T. Wollaeger and O. Korobkin. NIST-LANL Opacity Database (ver. 1.0), [Online]. Available: [Online]. Available: `https://nlte.nist.gov/OPAC`. National Institute of Standards and Technology, Gaithersburg, MD., (2020).
29. G. Gaigalas, D. Kato, P. Rynkun, L. Radžiūtė and M. Tanaka, Extended Calculations of Energy Levels and Transition Rates of Nd II-IV Ions for Application to Neutron Star Mergers, *ApJ Supp* **240**, p. 29 (February 2019).
30. L. Radžiūtė, G. Gaigalas, D. Kato, P. Rynkun and M. Tanaka, Extended Calculations of Energy Levels and Transition Rates for Singly Ionized Lanthanide Elements. I. Pr-Gd, *ApJ Supp* **248**, p. 17 (May 2020).
31. A. Kramida, Yu. Ralchenko, J. Reader and and NIST ASD Team. NIST Atomic Spectra Database (ver. 5.8), [Online]. Available: `https://physics.nist.gov/asd` [2021, October 1]. National Institute of Standards and Technology, Gaithersburg, MD., (2020).
32. K. Hotokezaka, M. Tanaka, D. Kato and G. Gaigalas, Nebular emission from lanthanide-rich ejecta of neutron star merger, *MNRAS* **506**, 5863 (October 2021).
33. P. A. Pinto and R. G. Eastman, The Physics of Type IA Supernova Light Curves. II. Opacity and Diffusion, *ApJ* **530**, 757 (February 2000).
34. J. I. Castor, *Radiation Hydrodynamics* 2004.
35. C. L. Fryer, S. E. Woosley and D. H. Hartmann, Formation Rates of Black Hole Accretion Disk Gamma-Ray Bursts, *ApJ* **526**, 152 (November 1999).
36. J. S. Bloom, S. Sigurdsson and O. R. Pols, The spatial distribution of coalescing neutron star binaries: implications for gamma-ray bursts, *MNRAS* **305**, 763 (May 1999).

37. L. C. Strang, A. Melatos, N. Sarin and P. D. Lasky, Inferring properties of neutron stars born in short gamma-ray bursts with a plerion-like X-ray plateau, *MNRAS* **507**, 2843 (October 2021).
38. R. T. Wollaeger, C. L. Fryer, C. J. Fontes, J. Lippuner, W. T. Vestrand, M. R. Mumpower, O. Korobkin, A. L. Hungerford and W. P. Even, Impact of Pulsar and Fallback Sources on Multifrequency Kilonova Models, *ApJ* **880**, p. 22 (July 2019).
39. L. Piro, E. Troja, B. Zhang, G. Ryan, H. van Eerten, R. Ricci, M. H. Wieringa, A. Tiengo, N. R. Butler, S. B. Cenko, O. D. Fox, H. G. Khandrika, G. Novara, A. Rossi and T. Sakamoto, A long-lived neutron star merger remnant in GW170817: constraints and clues from X-ray observations, *MNRAS* **483**, 1912 (February 2019).
40. H. Klion, P. C. Duffell, D. Kasen and E. Quataert, The effect of jet-ejecta interaction on the viewing angle dependence of kilonova light curves, *MNRAS* **502**, 865 (March 2021).
41. D. Watson, C. J. Hansen, J. Selsing, A. Koch, D. B. Malesani, A. C. Andersen, J. P. U. Fynbo, A. Arcones, A. Bauswein, S. Covino, A. Grado, K. E. Heintz, L. Hunt, C. Kouveliotou, G. Leloudas, A. J. Levan, P. Mazzali and E. Pian, Identification of strontium in the merger of two neutron stars, *Nature* **574**, 497 (October 2019).
42. M.-R. Wu, P. Banerjee, B. D. Metzger, G. Martínez-Pinedo, T. Aramaki, E. Burns, C. J. Hailey, J. Barnes and G. Karagiorgi, Finding the Remnants of the Milky Way's Last Neutron Star Mergers, *ApJ* **880**, p. 23 (July 2019).

Tutti-Frutti method: Recent developments in the PN/PM/SF treatment of the gravitational two-body problem

Donato Bini* and Andrea Geralico†

Istituto per le Applicazioni del Calcolo "M. Picone," CNR I-00185 Rome, Italy
**E-mail: donato.bini@gmail.com*
†E-mail: andrea.geralico@gmail.com

We review some recent results obtained at the sixth Post-Newtonian level of approximation for the Hamiltonian description of a two-body system, by using several methods whose combination has led to the so-called "Tutti-Frutti" approach.

Keywords: Two-body system; Post-Newtonian and Post-Minkowskian approximation; Scattering angle

1. Introduction

The two-body gravitational interaction is a fundamental physics problem whose descriptions necessitates a fully general relativistic treatment. The latter, in turn, is much difficult and over the years several approximation methods have been implemented for the mathematical treatment of this problem, ranging from weak-field and slow-motion (Post-Newtonian (PN) approximation[1]), weak-field but eventually relativistic motions (Post-Minkowskian (PM) approximation[2,3]), extreme-mass-ratio limit (i.e., the condition in which the mass of one of the two bodies is much larger of the other, discussed in the framework of perturbation theory and gravitational self-force (GSF)[4]), Effective Field Theory (EFT),[5] numerical relativity (NR).[6]

All these analytical (and semi-analytical) treatments have been used to build up a Hamiltonian description of the system. Since as soon as one raises the level of approximation considered the number of terms entering this Hamiltonian raises as well, in 1999 A. Buonanno and T. Damour[7,8] introduced the so-called Effective One-Body (EOB) approach aiming at a partial resummation of the Hamiltonian itself. Indeed, the EOB is especially useful since it condensates in a few gravitational potentials the essential characteristics of the interaction, and can be also continuously (and easily) updated as soon as new results become available in the literature by using whatever approximation scheme.

Let us recall also that 1) the gravitational interaction of two bodies is actually compatible with two basic scenarios: capture (the more massive of the two attracts and then swallows the other; the system in this case spirals undergoing ellipticlike motions) and scattering (the two bodies can be close enough but they have enough energy to resist the attraction; the system in this case undergoes hyperboliclike or paraboliclike motions). 2) Both cases are of particular importance in view of the

possibility to detect gravitational wave signals from Earth-based interferometers operating now (e.g., Ligo[9] and Virgo[10]) and also (more likely) by future, forthcoming satellite missions involving space-based interferometers (Lisa[11]). 3) The mathematical description of the dynamics by any available approximation method has strong limitations when the gravitational field becomes too strong, as for example in the case of the capture. In this condition one is only left with NR.

To add fuel to fire one should also take into account that new difficulties arise when taking into account the emission of gravitational radiation (energy, angular momentum and linear momentum) by the system as soon as a purely conservative scenario is no more possible. Indeed, starting from the 2.5PN level of approximation the problem is no more conservative, so that one has to deal with radiation-reaction effects. Furthermore, starting from the 4PN order the Hamiltonian of the system also includes a nonlocal part, which accounts for the past history of the system.

This picture clearly explains the difficulties which one encounters when trying to reach the 5PN (and beyond) level of accuracy in the model. Luckily, the various concomitant effects can be still computed separately so one can decide to limit to the study of the conservative and local part of the Hamiltonian, and later including nonlocal effects and radiation-reaction induced effects. This is the spirit of the recently developed "Tutti-Frutti" (TF) method, which combines several theoretical formalisms: PN, PM, multipolar-post-Minkowskian (MPM), EFT, GSF, EOB, and Delaunay averaging.[12–16]

2. EOB Hamiltonian

The EOB approach rewrites the (real, conservative) two-body Hamiltonian $H(r, p_r, p_\phi)$ in terms of an effective Hamiltonian $H_{\rm eff}(r, p_r, p_\phi)$ (ϕ is an ignorable coordinate in the conservative case)

$$H = M\sqrt{1 + 2\nu(\hat{H}_{\rm eff} - 1)}\,, \tag{1}$$

with

$$M = m_1 + m_2\,, \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}\,, \qquad \nu = \frac{m_1 m_2}{(m_1 + m_2)^2} \tag{2}$$

the total mass of the system (M), the reduced mass (μ) and the symmetric mass ratio (ν). The effective Hamiltonian (per unit of reduced mass, μ) $\hat{H}_{\rm eff}$ written in the standard p_r-gauge or DJS gauge[17] involves several potentials: A, $\bar{D}$, $\hat{Q}$, etc.,

$$\hat{H}^2_{\rm eff} = A(u,\nu)[1 + p^2_\phi u^2 + A(u,\nu)\bar{D}(u,\nu)p^2_r + \hat{Q}(u,\nu;p_r)]\,, \tag{3}$$

where $u = GM/r$ and

$$\hat{Q}(u,\nu;p_r) = p^4_r q_4(u,\nu) + p^6_r q_6(u,\nu) + p^8_r q_8(u,\nu) + \dots\,. \tag{4}$$

The various potentials 1) have a polynomial structure in terms of the symmetric mass-ratio ν, e.g.,

$$\begin{aligned} A(u,\nu) &= 1 - 2u + \nu a^{\nu^1}(u) + \nu^2 a^{\nu^2}(u) + \nu^3 a^{\nu^3}(u) + \dots \,, \\ \bar{D}(u,\nu) &= 1 + \nu \bar{d}^{\nu^1}(u) + \nu^2 \bar{d}^{\nu^2}(u) + \nu^3 \bar{d}^{\nu^3}(u) + \dots \,, \\ q_4(u,\nu) &= 1 + \nu q_4^{\nu^1}(u) + \nu^2 q_4^{\nu^2}(u) + \nu^3 q_4^{\nu^3}(u) + \dots \,, \end{aligned} \tag{5}$$

etc., with the degree of the polynom increasing with PN approximation order; 2) have both a local part and a nonlocal part

$$\begin{aligned} A(u,\nu) &= A_{\rm loc}(u,\nu) + A_{\rm nl}(u,\nu) \,, \\ \bar{D}(u,\nu) &= \bar{D}_{\rm loc}(u,\nu) + \bar{D}_{\rm nl}(u,\nu) \,, \\ \hat{Q}(u,\nu;p_r) &= \hat{Q}_{\rm loc}(u,\nu;p_r) + \hat{Q}_{\rm nl}(u,\nu;p_r) \,, \end{aligned} \tag{6}$$

the nonlocal part starting at 4PN.

3. The Tutti-Frutti approach and the determination of the conservative dynamics at 6PN

The starting point of the TF approach is to consider the total two-body conservative action, which at the 6PN accuracy has both a local-in-time part and a nonlocal-in-time part (starting at 4PN)

$$S_{\rm tot}^{\leq 6{\rm PN}} = S_{\rm loc}^{\leq 6{\rm PN}} + S_{\rm nonloc}^{\leq 6{\rm PN}} \,. \tag{7}$$

The main steps are summarized below.

(1) Fix completely the nonlocal part of the Hamiltonian by using Delaunay averaging along ellipticlike orbits.
The nonlocal action is given by

$$S_{\rm nonloc}^{4+5+6{\rm PN}} = -\int dt\, H_{\rm nonloc}^{4+5+6{\rm PN}}(t) \,, \tag{8}$$

with

$$H_{\rm nonloc}^{4+5+6{\rm PN}}(t) = \frac{G\mathcal{M}}{c^3} {\rm Pf}_{2r_{12}/c} \int \frac{dt'}{|t-t'|} \mathcal{F}_{\rm 2PN}^{\rm split}(t,t') \,. \tag{9}$$

Here, $\mathcal{M}$ denotes the total ADM conserved mass-energy of the binary system, r_{12} entering the timescale of the the *partie finie* (Pf) operation is the harmonic-coordinate radial distance, and $\mathcal{F}_{\rm 2PN}^{\rm split}(t,t')$ is the time-split version of the fractionally 2PN-accurate gravitational-wave energy flux, i.e.,

$$\mathcal{F}_{\rm 2PN}^{\rm GW}(t) \propto I_{ab}^{(3)}(t) I_{ab}^{(3)}(t) + O(\eta^2) \quad \to \quad \mathcal{F}_{\rm 2PN}^{\rm split}(t,t') \propto I_{ab}^{(3)}(t) I_{ab}^{(3)}(t') + O(\eta^2) \,, \tag{10}$$

with $\eta \equiv \frac{1}{c}$, the superscript in parenthesis denoting repeated time-derivatives of the quadrupole moment I_{ab}. Taking then the (Delaunay) time average of the

harmonic-coordinates Hamiltonian (9) gives a gauge-invariant function of two orbital parameters

$$\langle \delta H^{4+5+6\mathrm{PN}}_{\mathrm{nonloc,h}} \rangle = \frac{1}{\oint dt_h} \oint \delta H^{4+5+6\mathrm{PN}}_{\mathrm{nonloc,h}}(t_h) dt_h \,. \tag{11}$$

Next, do the same computation in EOB coordinates. Comparison between the two results allows fixing the nonlocal parts of the EOB potentials.

(2) Use SF information about small-eccentricity ellipticlike motion to determine (part of) the local EOB Hamiltonian.
Compute the averaged redshift factor, z_1, i.e., a (first) gauge-invariant variable associated with the conservative dynamics of the two-body system, along ellipticlike orbits by using first-order SF techniques in a small-eccentricity expansion limit, including but high powers of the eccentricity (up to the eighth order). $z_1 \sim \langle \partial_{m_1} H \rangle$ incorporates both local and nonlocal contributions in the Hamiltonian, but it is limited from the fact that analytic computations are possible only at first-order in the symmetric mass-ratio ν. Combining this result with the information about the nonlocal part of the EOB Hamiltonian specified at the point 1 will allow one to determine all the linear-in-ν local EOB potentials. The remaining (i.e., higher order in ν) coefficients are still undetermined.

(3) Compute another gauge-invariant quantity, the scattering angle, χ, along hyperboliclike orbits to determine most of the remaining coefficients.
The total scattering angle $\chi^{\mathrm{tot}} = \chi^{\mathrm{loc}} + \chi^{\mathrm{nonloc}}$ can be expressed as a large-angular momentum (or equivalently large-eccentricity) expansion, with coefficients having a precise mass-structure (or ν-structure) as recently shown by Damour.[18] Therefore, one needs to separately compute both the local and the nonlocal contributions to the scattering angle. The latter is defined by

$$\chi^{\mathrm{nonloc}}(E, J, \nu) = \frac{\partial W^{\mathrm{nonloc}}(E, J, \nu)}{\partial J} \,, \tag{12}$$

where E and J are the total energy and angular momentum in the c.m. frame, respectively, and

$$W^{\mathrm{nonloc}}(E, J; \nu) \equiv \int_{-\infty}^{+\infty} dt \, H^{\mathrm{nonloc}}(t) \,, \tag{13}$$

is the integrated nonlocal action. In order to compute the local contribution χ^{loc}, instead, it is convenient to convert the local EOB Hamiltonian into the so-called energy-gauge.[19] Imposing then that χ^{tot} satisfies the prescribed ν-structure will fix most of the parametrizing coefficients of the EOB potentials.

All the above steps used jointly have lead to the determination of most of the coefficients entering the conservative Hamiltonian of the two-body system. More precisely at 5PN there remain only 2 quantities to be determined ($\bar{d}_5^{\nu^2}$ and $a_6^{\nu^2}$) and at 6PN there remain only 4 more quantities to be determined ($q_{45}^{\nu^2}$, $\bar{d}_6^{\nu^2}$, $a_7^{\nu^2}$, and $a_7^{\nu^3}$).

Within the EFT approach, the conservative scattering angle can be decomposed into a potential-graviton contribution, and a radiation-graviton one.[5] The $O(G^4)$ potential-graviton contribution to the radial action has been recently derived in Ref.,[20] whereas Ref.[21] has computed the potential-graviton contribution to the 5PN two-body Hamiltonian. Although the TF decomposition of the two-body dynamics into local-in-time and non-local-in-time parts is closely linked to the EFT decomposition, one cannot simply identify the TF time-symmetric local-in-time dynamics to the EFT time-symmetric potential-graviton dynamics. However, one can compare the total conservative scattering angle derived within the two approaches. This allows one to complete the result of Ref.[20] by providing the explicit expression (at the 6PN accuracy) of the complementary radiation-graviton contributions to the scattering angle, or equivalently, to the radial action. Furthermore, adding the radiation-graviton contributions to the 5PN Hamiltonian obtained in Ref.[22] yields explicit expressions for the two 5PN undetermined $O(G^5)$ and $O(G^6)$ TF parameters $\bar{d}_5^{\nu^2}$ and $a_6^{\nu^2}$, involving either π^2 terms and rational coefficients entering various (local-in-time) radiation-graviton contributions to the conservative effective 5PN action.[22] Our results are in agreement (for the π^2 contributions) with those of Ref.[21] A comparison with the results of Ref.[22] is currently under consideration.

4. Radiation-reaction effects

The presence of a radiation-reaction force implies a modification of Hamilton equations as

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}\,, \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} + \boldsymbol{\mathcal{F}}_{\rm rr}\,. \tag{14}$$

The work done by this force on the system implies energy, angular momentum and linear momentum losses by the system itself, as well as to the fact that the four-velocity of the center-of-mass is no more conserved, leading to recoil effects. For instance, the scattering angle will be modified as

$$\chi = \chi^{\rm cons} + \delta^{\rm rr}\chi\,, \tag{15}$$

where the radiation-reaction corrections to the conservative value can be written as

$$\delta^{\rm rr}\chi = -\frac{1}{2}\left[\frac{\partial \chi^{\rm cons}}{\partial E}E^{\rm rad} + \frac{\partial \chi^{\rm cons}}{\partial J}J^{\rm rad}\right]\,, \tag{16}$$

in terms of the energy and angular momentum losses $E^{\rm rad}$ and $J^{\rm rad}$. The latter include instantaneous and tail (i.e. hereditary) contributions, and admit a double PM and PN expansion. At the fractional 2PN order they read

$$\begin{aligned} E^{\rm rad}_{\rm 2PN} &= \nu^2\left(E_{\rm N} + \eta^2 E_{\rm 1PN} + \eta^3 E^{\rm tail}_{\rm 1.5PN} + \eta^4 E_{\rm 2PN}\right)\,,\\ J^{\rm rad}_{\rm 2PN} &= \nu^2\left(J_{\rm N} + \eta^2 J_{\rm 1PN} + \eta^3 J^{\rm tail}_{\rm 1.5PN} + \eta^4 J_{\rm 2PN}\right)\,, \end{aligned} \tag{17}$$

where tails start at the 1.5PN order. For example, the Newtonian values are given by

$$E_{\rm N} = \frac{37}{15}\pi\frac{p_\infty^4}{j^3} + \frac{1568}{45}\frac{p_\infty^3}{j^4} + \frac{122}{5}\pi\frac{p_\infty^2}{j^5} + O\left(\frac{1}{j^6}\right),$$
$$J_{\rm N} = \frac{16}{5}\frac{p_\infty^3}{j} + \frac{28}{5}\pi\frac{p_\infty^2}{j^2} + \frac{176}{5}\frac{p_\infty}{j^3} + O\left(\frac{1}{j^4}\right), \tag{18}$$

where $j \equiv cJ/(GM\mu)$ is a dimensionless angular momentum parameter, and the linear momentum at infinity, p_∞, is related to the binding energy by $\bar{E} \equiv (E - Mc^2)/(\mu c^2) = \frac{1}{2}p_\infty^2$, at the Newtonian level. At higher PN orders the coefficients of the above expansion become functions of ν: linear functions of ν at 1PN, quadratic at 2PN, etc., but the structure of the PM expansion is exactly the same.

We have computed in Ref.[23] the changes of 4-momentum during the scattering process (between the two asymptotic states of the two bodies labelled by $a = 1, 2$) which are linear order in radiation-reaction, i.e.,

$$\Delta p_{a\mu} \equiv p_{a\mu}^+ - p_{a\mu}^- = \Delta p_{a\mu}^{\rm cons} + \Delta p_{a\mu}^{\rm rr}, \tag{19}$$

where the radiation-reacted contribution $\Delta p_{a\mu}^{\rm rr}$ is the sum of a relative-motion term and a recoil one, which are linear in the radiative losses of energy, linear-momentum, and angular momentum. We have also shown how the polynomial dependence of $\Delta p_{a\mu}^{\rm rr}$ can be exploited to yield some identity relating the various radiative losses. Adding radiation-reaction effects at higher PN orders is an open issue for challenging future works.

References

1. L. Blanchet, "Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries," Living Rev. Rel. **17**, 2 (2014) [arXiv:1310.1528 [gr-qc]].
2. L. Bel, T. Damour, N. Deruelle, J. Ibanez and J. Martin, "Poincaré-invariant gravitational field and equations of motion of two pointlike objects: The postlinear approximation of general relativity," Gen. Rel. Grav. **13**, 963 (1981).
3. K. Westpfahl, "High-Speed Scattering of Charged and Uncharged Particles in General Relativity," Fortsch. Phys. **33**, 417 (1985).
4. S. L. Detweiler, "Perspective on gravitational self-force analyses," Class. Quant. Grav. **22**, S681 (2005) [gr-qc/0501004].
5. W. D. Goldberger and I. Z. Rothstein, "An Effective field theory of gravity for extended objects," Phys. Rev. D **73**, 104029 (2006) [arXiv:hep-th/0409156 [hep-th]].
6. F. Pretorius, "Numerical relativity using a generalized harmonic decomposition," Class. Quant. Grav. **22**, 425 (2005) [gr-qc/0407110].
7. A. Buonanno and T. Damour, "Effective one-body approach to general relativistic two-body dynamics," Phys. Rev. D **59**, 084006 (1999) [gr-qc/9811091].
8. A. Buonanno and T. Damour, "Transition from inspiral to plunge in binary black hole coalescences," Phys. Rev. D **62**, 064015 (2000) [gr-qc/0001013].
9. See the LIGO website at http://www.ligo.org
10. See the VIRGO website at http://www.virgo-gw.eu
11. See the LISA website at https://lisa.nasa.gov/

12. D. Bini, T. Damour and A. Geralico, "Novel approach to binary dynamics: application to the fifth post-Newtonian level," Phys. Rev. Lett. **123**, no.23, 231104 (2019) [arXiv:1909.02375 [gr-qc]].
13. D. Bini, T. Damour and A. Geralico, "Binary dynamics at the fifth and fifth-and-a-half post-Newtonian orders," Phys. Rev. D **102**, no.2, 024062 (2020) [arXiv:2003.11891 [gr-qc]].
14. D. Bini, T. Damour and A. Geralico, "Sixth post-Newtonian local-in-time dynamics of binary systems," Phys. Rev. D **102**, no.2, 024061 (2020) [arXiv:2004.05407 [gr-qc]].
15. D. Bini, T. Damour and A. Geralico, "Sixth post-Newtonian nonlocal-in-time dynamics of binary systems," Phys. Rev. D **102**, no.8, 084047 (2020) [arXiv:2007.11239 [gr-qc]].
16. D. Bini, T. Damour, A. Geralico, S. Laporta and P. Mastrolia, "Gravitational scattering at the seventh order in G: nonlocal contribution at the sixth post-Newtonian accuracy," Phys. Rev. D **103**, no.4, 044038 (2021) [arXiv:2012.12918 [gr-qc]].
17. T. Damour, P. Jaranowski, and G. Schäfer, "On the determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation," Phys. Rev. D **62**, 084011 (2000) [arXiv:gr-qc/0005034].
18. T. Damour, "Classical and quantum scattering in post-Minkowskian gravity," Phys. Rev. D **102**, no.2, 024060 (2020) [arXiv:1912.02139 [gr-qc]].
19. T. Damour, "High-energy gravitational scattering and the general relativistic two-body problem," Phys. Rev. D **97**, no. 4, 044038 (2018) [arXiv:1710.10599 [gr-qc]].
20. Z. Bern, J. Parra-Martinez, R. Roiban, M. S. Ruf, C. H. Shen, M. P. Solon and M. Zeng, "Scattering Amplitudes and Conservative Binary Dynamics at $\mathcal{O}(G^4)$," Phys. Rev. Lett. **126**, no.17, 171601 (2021) [arXiv:2101.07254 [hep-th]].
21. J. Blümlein, A. Maier, P. Marquard and G. Schäfer, "The fifth-order post-Newtonian Hamiltonian dynamics of two-body systems from an effective field theory approach: potential contributions," Nucl. Phys. B **965**, 115352 (2021) [arXiv:2010.13672 [gr-qc]].
22. S. Foffa and R. Sturani, "Hereditary terms at next-to-leading order in two-body gravitational dynamics," Phys. Rev. D **101**, no.6, 064033 (2020) [erratum: Phys. Rev. D **103**, no.8, 089901 (2021)] [arXiv:1907.02869 [gr-qc]].
23. D. Bini, T. Damour and A. Geralico, "Radiative contributions to gravitational scattering," Phys. Rev. D **104**, no.8, 084031 (2021) [arXiv:2107.08896 [gr-qc]].

GRB observations on cubesate satellites in the Universat–SOCRAT project

Sergey I. Svertilov[1], Michail I. Panasyuk, Vitaly V. Bogomolov[2] and Mikhail I. Prokhorov[3]

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, 119234, Russia
Department of Physics, Lomonosov Moscow State University, Moscow, 119234, Russia
Email: [1]*sis@coronas.ru,* [2]*bogovit@rambler.ru,* [3]*prokhorov.mi17@physics.msu.ru*

Anatoly F. Iyudin[4], Pavel A. Klimov[5], Vladislav I. Osedlo[6], Vasily L. Petrov[7], Oleg Yu. Peretyat'ko[8], Yury K. Zaiko[9] and Ivan V. Yashin[10]

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, 119234, Russia
Email: [4]*aiyudin@srd.sinp.msu.ru,* [5]*pavel.klimov@gmail.com,* [6]*osedlo@mail.ru,* [7]*vas@srd.sinp.msu.ru,* [8]*pouoleg@mail.ru,* [9]*zaikoiurii@yandex.ru,* [10]*iv_n@bk.ru*

Vladimir M. Lipunov

Department of Physics, Lomonosov Moscow State University, Moscow, 119234, Russia
Sternberg Astronomical Institute, Lomonosov Moscow State University, Moscow, 119234, Russia
Email: lipunov2007@gmail.com

Vitaliy Yu. Prokopyev and Alexey S. Styuf

Division for Aerospace Research, Novosibirsk State University, Novosibirsk, 630090, Russia
Email: vprok@cosmos.nsu.ru

Multi-channel astronomy is one of the most important and rapidly developing field of modern physics. The well known result of multi-channel observations is simultaneous detection of gravitational wave and gamma ray bursts associated with neutron star merger event. This observation open a new window to study the Universe, and actually triggered the broad scale study of astrophysical phenomena in hard X-Rays and gamma rays by orbital experiments together with ground based observations of gravitational waves, neutrino and ultra-high energy cosmic rays. From this perspective it appears that small satellites of CubeSat type are quite appropriate for multi-channel observations of astrophysical transients because it is the cheapest way to realize all-sky monitoring observations by orbital instruments. Presently at D.V. Skobeltsyn Institute of Nuclear Physics of the M.V. Lomonosov Moscow State University (SINP MSU) a new project named Universat–SOCRAT is under development which is intended for operational monitoring of near-Earth's radiation environment and monitoring of electromagnetic transients in the optical, UV, X-ray and gamma ranges. Here we discuss the first results of charged particles, gamma quanta fluxes and UV-emission measurements from the upper atmosphere in several CubeSat missions, which were successfully launched in 2019, 2020.

Keywords: Multimessenger astronomy; Universat-SOCRAT project.

1. Introduction

One of the most important and interesting fields of modern physics and astrophysics is the study of extreme processes and phenomena leading to the most powerful explosions, acceleration of particles to the maximum possible energies, generation of superstrong gravitational and electromagnetic fields. Such phenomena include cosmic gamma-ray bursts (GRBs) associated with the collapse of massive stars (hypernovae), merging of relativistic compact objects (neutron stars and black holes), leading to the generation of gravitational waves and short-term increases in gamma-ray fluxes, nonstationary processes in neutron stars with a superstrong magnetic field (magnetars), dynamic processes in the nuclei of active galaxies, pulsars and close binary systems, which can be associated with the acceleration of particles to high and extremely high energies and the generation of high and ultrahigh energy neutrinos.

Despite the fact that these extreme phenomena have been studied very intensively in recent years, there are still many unsolved problems. As for GRBs, this is, first of all, the definition of the mechanism of functioning of the "central engine", the formation of relativistic jets and the propagation of relativistic shock waves in them, the so-called shells. The related problems concern with the study of the possibility of particle acceleration on relativistic shock waves, and of the conditions for generation of polarized electromagnetic radiation. In this regard, especially important information can be obtained as a result of multi-wavelength observations of the so-called intrinsic emission of a GRB, especially in the optical and gamma ranges. Among the observational tasks, it should be noted that it is necessary to observe GRBs with the maximum possible temporal and spectral resolution, as well as at a high level of sensitivity, which will make it possible to expand the range of observed burst sources up to the most distant objects in the Universe ($z \sim 10-12$). The study of the fine temporal structure of GRB intrinsic emission, as well as of the so-called precursors, including their observations in the optical range, is very actual.

An important task is studying of the so-called short hard GRBs, includes the need to discover their intrinsic optical emission. It is known, that such bursts are associated with the processes of neutron stars merging in close binaries; therefore, their study is particularly relevant after recent discovery of gravitational-waves bursts generated during the merger of relativistic compact objects.

Considering the relevance of studying extreme astrophysical phenomena, it should be emphasized that for further progress in understanding their nature, multi-wave (at least simultaneous in the optical, X-ray and gamma ranges), multi-channel and multi-satellite observations are required. Multi-channel observations mean the simultaneous detection of events by space-based and ground-based facilities, including ground-based optical, radio and gamma telescopes, gravitational-wave antennas, and facilities designed to detect neutrinos and ultra-high energy cosmic rays. Multi-satellite observations are necessary to ensure the detection of each GRB on several spacecraft in order to localize the GRB source using the triangulation technique.

2. Advantages of CubeSat constellation for GRB observations

The main success of multi-channel astrophysical observations is connected with simultaneous detection of gravitation wave and gamma-ray burst in August 17, 2017 from neutron star merging.[1] It seems that one of the cheapest and very effective ways of multi-channel and multi-satellite observations of such events and GRBs in general is use of small satellites grouping, in particular the CubeSat constellations. Due to the relatively low cost and ease of manufacture, it is possible to carry out more, or less regular launches of CubeSat satellites as a by-pass mission. Thus, a constellation of nano-satellites can be created, which will significantly increase the efficiency of space radiation and electromagnetic transient monitoring.

The solution for the localization problem can be achieved by applying the triangulation method for observations of the selected area with different satellites. For this purpose the number of CubeSats should be launched for the joint observations of a given burst. To realize the triangulation technique in minimal variant 3-4 CubeSats are necessary (see Fig. 1). Optimally they should be launched as by-pass mission together with the main spacecraft of multi-satellite group, i.e. on the near circular solar-synchronous orbits with altitude about 400 – 600 km, inclination about 98° and longitude deviation of the ascending node $< 5°/year$. The error box of GRB source localization is defined by time delay of signals detected on each satellite. The longer delay the more accurate will be localization. Delay itself depends on distance between satellites, i.e. the larger distance corresponds to the longer delay. In the case of near-Earth orbits the distance between satellites is limited by about 10 000 – 15 000 km, because all satellites should be from one side of the Earth to avoid the screening of GRB source. By this the necessity of maintaining of given distance between satellites is a separate technical problem. It can be solved by the use of thrusters. Off cause CubeSat constellation can be used also for observations of transient events from the Earth atmosphere, including Terrestrial Gamma Ray Flashes (TGF) and monitoring of space radiation. However, there will be other requirements to the distance between satellites in the case of triangulation technique application for TGFs. In the last case the given area of Atmosphere should be observed by all satellites of grouping, thus the maximal distance should be no more than about 1000 km. For successive realization of multi-satellite mission for GRB observations certain conditions should be met:

- presence of satellite Internet (global star/iridium) plus high frequency transmitter;
- stabilization better than $1°/s$, satellite orientation on the "Zenith–Nadir" axis and velocity vector with accuracy $+/-10°$, information on satellite axes orientation with accuracy 0.1°;
- possibility of quasi-autonomous navigation, in particular with the use of millisecond pulsars as beacons;

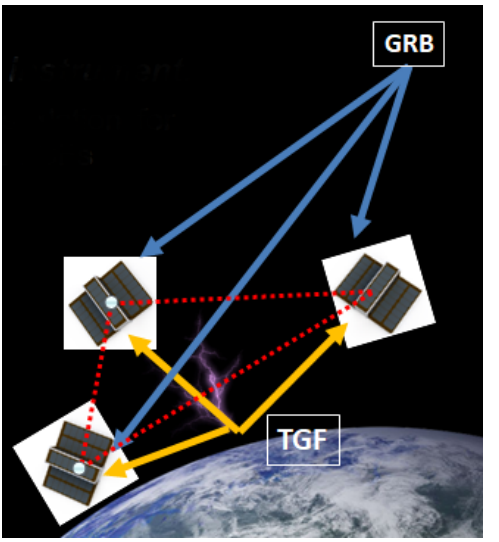


Fig. 1. The CubeSat constellation in minimal configuration.

- different data types, such as monitoring, event-by-event, telegrams about burst triggers (GCN);
- data daily volume no less 100 Mb;
- exact timing is needed;
- orientation accuracy $\sim 1^\circ$ is useful;
- fast telegrams are needed;
- necessity of charge particle variation control.

The last demand is particularly actually if the satellites are launched to the polar orbit because of the problem of false triggering by electron precipitation rises. It can be solved by monitoring of electron flux with the use of special electron detectors or instruments which are able to select true gamma-quantum flux rises on the background of electron precipitation. The last is illustrated by Figs 2, 3, on which the examples of GRB imitations by electron precipitation and the map with such imitation distributions from the BDRG/Lomonosov data[2] are presented.

For instruments, that can be used for GRB detection on-board CubeSats one of the main criteria is the detector area, that is sufficient for GRB registration, should be larger $\sim 50 cm^2$ in the energy range of photons from $\sim$ 20 keV up to several MeV. Such detector units can also be used to detect TGFs and X-ray and gamma ray emission of solar flares.

We note, that presently the space project of Moscow University Universat-SOCRAT is under realization. In the frame of this project the launches of a number of satellites including CubeSats, that can be used as the base of constellation for GRB study, is foreseen.

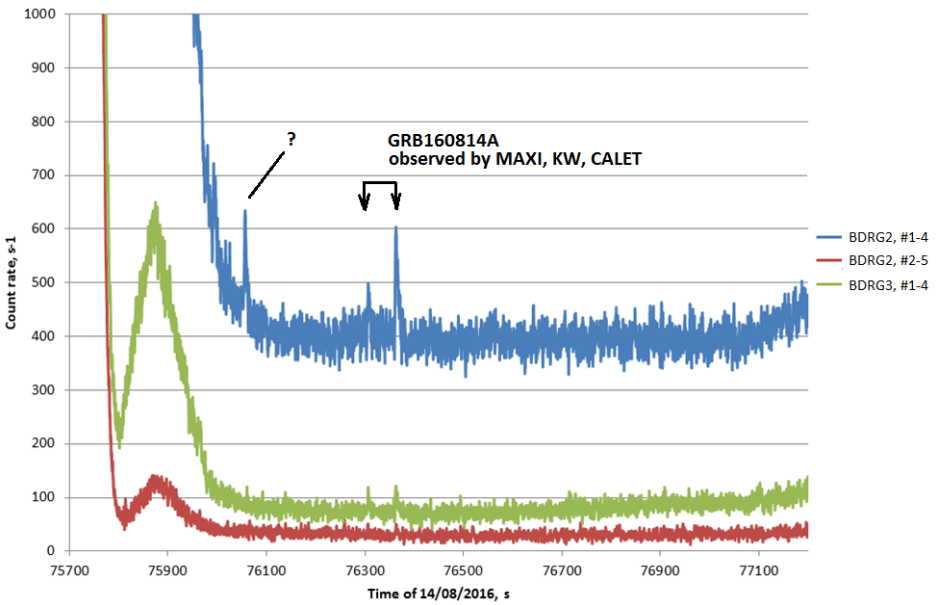


Fig. 2. Example of GRB and burst imitations by magnetosphere electron precipitation detected BDRG/Lomonosov near radiation belts.

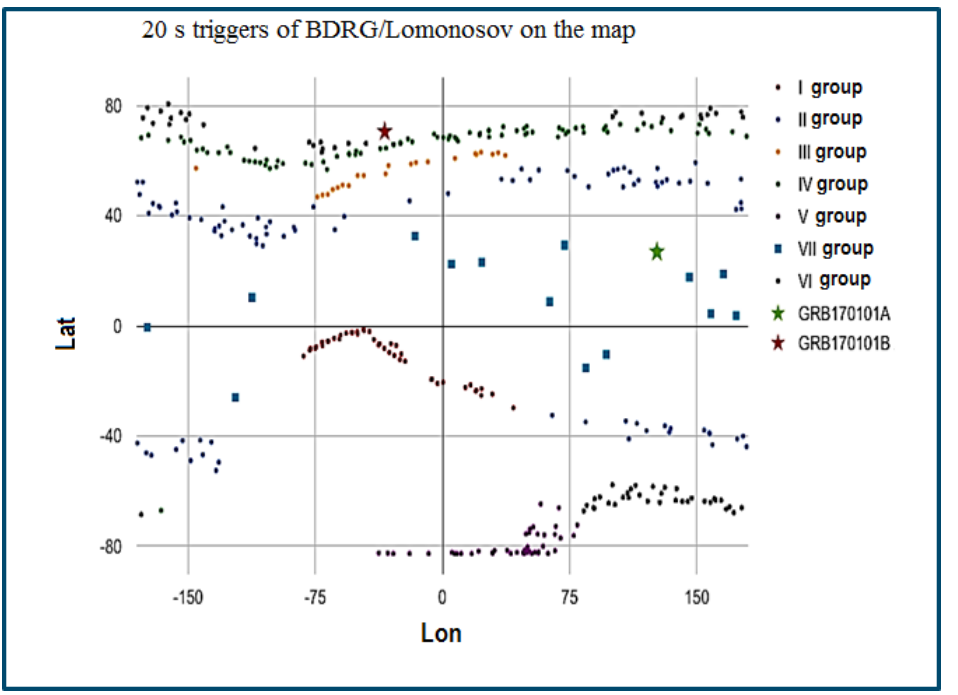


Fig. 3. Geographic distribution of different types of BDRG/Lomonosov triggers. By stars the real GRBs are marked.

2.1. *Universat–SOCRAT space project*

Under the Universat–SOCRAT project it is planned to create a grouping of small satellites for real time monitoring in the near-Earth space of potentially dangerous hazards, i.e. the radiation environment; dangerous objects of the natural (asteroids, meteors) and technogenic origin (space debris), as well as electromagnetic transients (cosmic gamma-ray bursts, optical, ultraviolet and gamma ray flashes from the Earth's atmosphere.[3, 4])

Within the framework of the Universat–SOCRAT project several small spacecraft should be launched on specially selected orbits. In the minimal version, the group of satellites should consist of three spacecraft. One spacecraft of medium mass (small satellite) should be launched on a low solar-synchronous orbit with a height of about 500–650 km and an inclination of 97–98°. Two other satellites of lower mass (micro satellites) should be launched on an orbit close to circular with a height of about 1500 km and an inclination of $\sim 80°$ and on an elliptical orbit with an apogee of about 8000 km, a perigee of 600–700 km, inclination 63.4° and argument of perigee $\sim 310°$.

The small satellite payload should include instruments for monitoring of space radiation, a set of instruments for optical monitoring of hazardous objects, a set of instruments for studying of atmospheric phenomena in the optical range, a set of instruments for monitoring in gamma- range, and special unit for data collection. The payload should also include three-component magnetometer. The payload of each micro satellite should include instruments for space radiation monitoring, a compact gamma spectrometer, a wide field of view optical camera, an ultraviolet detector and an electronics unit for data collection. Payload of all three satellites also should include the special electronic unit for data collection from detector units, its transmission to the board systems and feeding too the instruments power supply and commands.

The basic multi-satellite group can be supplemented with nano-satellites of the CubeSat type. Due to the relative cheapness and ease of manufacture, it is possible to carry out more or less regular launches of satellites as a by-pass mission. Thus, a constellation of nanosatellites can be created, which will significantly increase the efficiency of space radiation and electromagnetic transient monitoring.

The use of several nano-satellites of CubeSat type equipped with several multi-directional spectrometers of energetic protons and electrons will provide a solution of the problem of elaboration of a three-dimensional dynamic model of radiation in near-Earth space. For this, the satellites should be launched onto specially selected orbits and provide measurements that will allow us to calculate the current distribution of particle fluxes in a large volume of near-Earth space, i.e. from low orbits to geostationary, and, as a result, determine the current levels of radiation doses for a wide range spacecraft orbits, and also give a forecast of the radiation situation at low altitudes. Simultaneous measurements of particle fluxes on several spacecraft located at different points in near-Earth space will also allow sufficiently reliable

separation of temporal and spatial effects in detected variations of the instrument output readings.

The nano-satellite constellation also gives advantages for GRB and TGF study. As it was mentioned above, the crucial point from this view is necessity of good localization of the transient source. This can be achieved with the use of triangulation technique. In the case of GRB observations about 3–4 satellites will be enough. As for TGF study, optimally the spacecraft grouping should consist of 16 satellites pairwise launched in 8 double orbits, i.e. for each pair of orbits the angle between the planes of the orbits should be about For the all constellation spacecraft the main operational mode of the most instruments is on duty mode when the all instruments are switched-on and operate continuously. Instrument switching between the operational modes is carried out by commands from the Earth or by the internal programs. To optimize the payload energy consumption, the data exchange between instrument electronic and satellite board should be foreseen, including information on changing the parameters of the spacecraft power system and payload switching into energy-saving modes (changes in the instrument operating modes or their partial switching-off).

2.2. *Instruments for gamma transient monitoring*

Instruments for study of gamma transients in the Universat-SOCRAT project will include gamma-ray flash monitor (GFM), and a tracking gamma-ray spectrometer (TGS) of the high resolution and sensitivity, that will be sensitive also to neutrons with energies from about 1 MeV up to dozens of MeV.[5] The separate unit for data analysis and control of GFM and TGS should be also foreseen in the payload. This unit should contain digital electronics unit that will provide the record of data stream with a time resolution of $\sim 10\mu s$, and generate the triggers enabling the detection of different type events, including GRB/TGF flux increases. The data record must be referenced to the UTC time with an accuracy of $\sim 10\mu s$ to allow the positioning of the burst, if observed simultaneously by several space missions, using triangulation, and for the comparison with the data of ground based lightning nets.

To realize the triangulation technique sufficiently light X-ray and gamma-ray detectors aimed for the GRB/TGF detection only, without spectrometer capabilities, will be used on the CubeSats. The composition of each spacecraft of small satellite constellation of the cubesat type should include gamma-ray detectors in the amount of:

— Cubesat 1U – 1 instrument;
— Cubesat 3U – at least 1 instrument;
— Cubesat 6U – at least 3 instruments.

If one instrument is installed on the satellite, its axis should be directed along the nadir - zenith axis to nadir, and the field of view should not be obscured within $\pm 60^\circ$ from the axis of the instrument in the nadir direction. If two instruments are

installed on the spacecraft, their axes should be normal to each other, while the axis of one detector should be directed along the nadir—zenith axis to nadir, and the axis of the other detector should be directed along the spacecraft velocity vector in the case of the orbital orientation of the spacecraft (along the line "zenith - nadir" and the velocity vector), while the field of view should not be obscured within $\pm 60^\circ$ from the axis of the detector. If three instruments are installed on the spacecraft, their detector axes should be normal to each other, directed along mutually perpendicular edges of the cube, as if forming a Cartesian coordinate system, while the axis of one detector should be directed to the nadir, the other along the velocity vector, the axis of the third should complement the three-axial coordinate system to complete. Each gamma detector should be placed in a cubesat-type spacecraft in such a way that the surface of its input window coincides with the surface of one of the side faces of the satellite. The instrument should consist of detector assemblies and electronics boards. Various options for the use of detector units are considered. In the simplest version there are considered 4 detector units each from scintillation crystals with a size of $45 \times 45 \times 5$ mm, viewed by a photo-sensor. There are considered as possible scintillators CsI(Tl) crystals (a cheap option) or Ce:GAGG (an optimal option). For signal pickup, photodetectors such as tube photomultiplier tubes (PMTs) or silicon PMT arrays (Si-PM) can be used.

To provide identification of different types of space radiation (electrons, gamma-quanta), a variant with a fosphich detector can be used when a combination of a plastic scintillator/inorganic scintillator (CsI(Tl) or Ce:GAGG), viewed by a single photodetector, is used while the separation of signals from different scintillators is carried out by analyzing the time profiles of the PMT output signals.

A multi-pixel version of the detector can be also considered when it consists of individual crystals (pixels) of a relatively small size, each of which is viewed by its own photodetector. In this case, a variant can be considered both using standard arrays of silicon PMTs, and using original photodetectors and corresponding non-standard electronics.

2.3. *Instruments for electromagnetic transient monitoring in ultraviolet and optics*

Instruments for study transient luminous events (TLEs) in UV and optical bands should include position sensitive spectrometer, i.e. small lens telescope with high time resolution for spectral measurements of TLE and lightning and UV and infrared (IR) detector-photometer, which is similar to that used in Tatiana-2 and Vernov missions.[6] This instrument will be added by channels in far UV range. It allows comparison with data of previous experiments on study UV flashes in the Atmosphere. Spectral measurements are necessary to determine the type and altitude of TLE generation as well as to reveal lightning discharges by typical 777 nm line and by the absence of signal in the range of oxygen absorption lines about 762 nm. Axes of both instruments should be directed toward Nadir with open angles about 90°.

For the TLE observations on the CubeSats will be used compact UV detector with a wide field of view (AURA-2) or its improved version, i.e. telescope (AURA-2T).[7] These instruments are the development of the AURA (Atmospheric Ultraviolet Radiation) detector already launched on board the VDNH-80 satellite. In this instrument SiPM MicroFC-60035-SMT silicon photomultipliers with a sensitive area of 6×6 mm, spectral sensitivity from 300 to 800 nm, a maximum spectral sensitivity at 420 nm, and a quantum efficiency of about 41% at 420 nm are used as photodetectors. Silicon photomultiplier tubes, unlike traditional PMTs used for fast photometry, have a number of advantages that are significant for experiments on small spacecraft. It is compact (thickness about 1 mm), low voltage ($25 - 70$ V) and low weight. The main advantages of the new detector options are high temporal resolution ($1\mu s - 10ms$) and spatial resolution, higher sensitivity due to an increase in the area of the optical system.

The operation of the detectors in the monitoring mode under the condition of simultaneous operation on several spacecraft will make it possible to control transient activity in the Earth's atmosphere, both of thunderstorm and extra-thunderstorm nature with a large exposure. Cartography and monitoring of UV transients is an important task for understanding the interconnections of various energetic processes in the atmosphere. Also, the quasi-stationary UV glow of the atmosphere can be an additional indicator of the state of the geomagnetic situation and the effect of the solar wind on the atmosphere and magnetosphere of the Earth.

3. First results of Moscow University CubeSat missions

The first stage of the Universat–SOCRAT program began to be implemented on July 5, 2019 after a successful launch from the Vostochny cosmodrome of three nanosatellites of the cubesat type (SOCRAT, AmurSat, VDNH-80). These satellites are equipped with instruments DéCoR (Detector of Cosmic Radiation) for monitoring space radiation, as well as prototypes of instruments for observing transient phenomena in the Earth's atmosphere. In particular, scintillation fosphich detectors, which are prototype of the instrument intended for gamma transient observation on cubsats, are installed on two satellites (AmurSat, VDNH-80). These instruments detect charged particles and gamma rays in the energy release range of 0.1 - 2.0 MeV. The geometric factor of these instruments is $\sim 50 cm^2 sr$. One of the cubesats (VDNH-80) also contains an optical photometer AURA, consisting of four silicon photomultipliers, whose input windows are covered by different light filters. Thus, the instrument provides observations of the Earth's atmosphere in the range from ultraviolet to red. The satellites were launched into a sun-synchronous orbit with an altitude of ~ 800 km. This creates favorable conditions for monitoring space radiation in various areas of near-Earth space, including zones of captured radiation, areas of precipitation, etc. Such an orbit also allows observations of flash phenomena both in the near-equatorial atmosphere and at high latitudes.

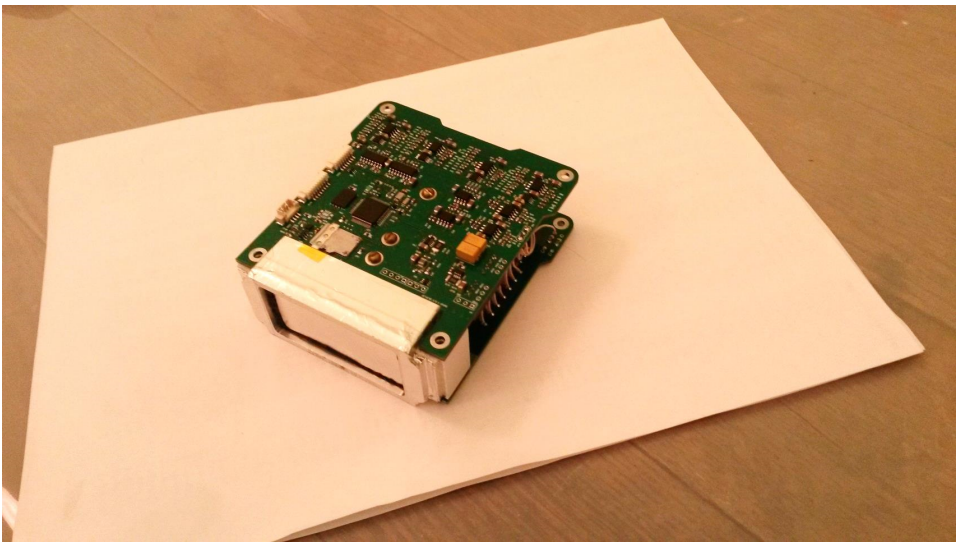
The implementation of the Universat-SOCRAT program was continued by the successful launch on 28 September 2020 of three more spacecraft of the cubesat type, on which DeCoR and AURA-2 instruments are installed. One of them, the DECART 6U satellite, contains three identical DeCoR instruments installed in such a way that their axes are mutually normal to each other, which makes it possible to estimate the angular distributions of the detected fluxes. DeCoR devices are also installed on the Norbi 6U satellite, developed jointly with the Novosibirsk State University, as well as on the Yarilo-2 spacecraft of the 1.5U format, developed jointly with the N.E. Bauman Moscow State Technical University. The AURA-2 instrument is operated on-board of DECART satellite.

3.1. *Parameters, structure and functioning principles of DeCoR payload*

The DéCoR payload is an instrument for measuring the fluxes and spectra of charged particles and gamma rays in the energy release range of 0.1 - 2 MeV. The main scientific goals are the study of fast variations of electron fluxes in the areas of precipitation and the slot between radiation belts, as well as the study of the dynamics of particle fluxes and gamma rays at low orbits in dependence on geomagnetic conditions. An important factor that allows more efficient scientific research is the installation of the instrument on two satellites, sequentially flying through the same region of near-Earth space. A photo of the instrument and sketch of detector unit are shown in Fig. 4.

The DeCoR instrument is adapted in order to be mounted in the cubesat satellite body. Its detector is a two-layer assembly of a plastic scintillator and a CsI (Tl) crystal (see Fig. 4b). The detector has a sensitive area of $\sim 18cm^2$. The assembly is viewed by two PMTs. Separate digitization of the initial and final parts of the PMT output pulse ("Fast component" and "Slow component") from each PMT allows determining the scintillator in which the interaction took place from the pulse shape analysis pulse and, thus allows separate measuring of electrons and gamma ray fluxes. Data and commands are exchanged with the satellite's on-board systems by the instrumentmicrocontroller of the device using the UART protocol. For the accumulation and storage of scientific data non-volatile memory with a capacity of 16 MB is foreseen in the instrument. For transmission to the Earth, the DeCoR instrument generates both monitoring data (particle counting rates 1 time per second) and detailed data about all interactions in the detector with a time resolution of $20\mu s$ (so call event by event mode), which are accumulated in the instrument's memory by a command that determines the moment of interesting time from the view of the researcher . The main way to conduct a scientific experiment is to regularly collect and transmit monitoring data to the Earth, as well as activate from time to time the detailed recording mode on pre-calculated sections of the satellite trajectory passing through the zones of possible variations of the trapped and quasi-trapped particle fluxes.

a



b

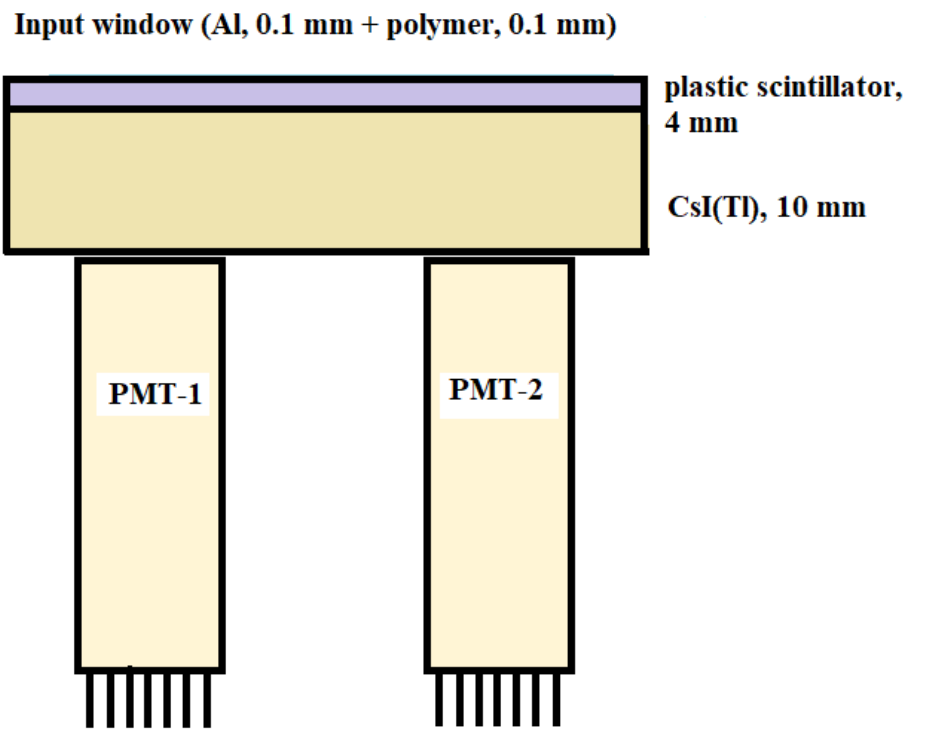


Fig. 4. The photo (a) and general view (b) of DeCoR instrument.

After viewing data of the "Monitoring" type, the researcher identifies the time points at which the flow variations of interest actually occurred. For these moments, detailed data is requested for transmission to Earth. The used data transfer rate allows you to transmit detailed data about an interesting phenomenon, for example, about the precipitation of magnetosphere electrons lasting several minutes, for A magnetometer is also installed in the DeCoR instrument, which makes it possible to measure the magnitude of the magnetic field along three mutually perpendicular axes. Using the outputs of this magnetometer allows taking into account spacecraft rotation in the study of trapped and precipitating electrons and even use it to estimate the pitch-angle distribution of particles. The magnetometer data are included directly in the monitoring information frames containing the parameters of the measured radiation.

3.2. *The results of observations with DeCoR instrument on-board AmurSat and VDNH-80 satellites*

Space missions with AmurSat and VDNH-80 satellites were the first Moscow University experience of monitor observations on cubesats. Due to technical problems, first of all connected with necessity of solar battery charging and limited transmitter capabilities, during both missions data were obtained from DeCoR instrument for about hundred and a half orbits. Taking into account that background conditions appropriate for GRB observations were on about a half of orbit, real instrument field of view about 1 sr and GRB detection threshold about $5 \cdot 10^{-7} erg/cm^2$, the expected number of detecting GRBs was no more than one. Really there were no detected any significant gamma-quantum flux increasing, which can be discussed as GRB candidates. Nevertheless sufficiently rich experience was obtained in view of radiation monitoring, especially of electron flux variation study by two spacecraft.

As noted above, an important advantage of multi-satellite experiments is the possibility of simultaneous measurements at different points in the near-Earth space on the one hand and measurements in the same regions with sequential passage of satellites through them on the other. Such measurements are illustrated in Fig. 5, 6, which show the orbital projections of the AmurSat and VDNKh-80 satellites and the outputs of two DeCoR instruments installed on these satellites, which were recorded on March 18, 2020. The difference in the positions of the satellites was about half of the orbit, so at the same time one spacecraft was taking measurements in the North Polar Region, and the other in the South.

Instruments on both satellites were switched on over Kaluga-city, when the satellites were moving in the direction of the northern polar region. At the moment of switching on the DeCoR instrument on the VDNH-80 satellite, the AmurSat cubesat was in the southern hemisphere, so the DeCoR instrument on it was switched on after ~ 45 minutes. At the same time, the AmurSat satellite continued to move along a similar trajectory with the VDNH-80 spacecraft.

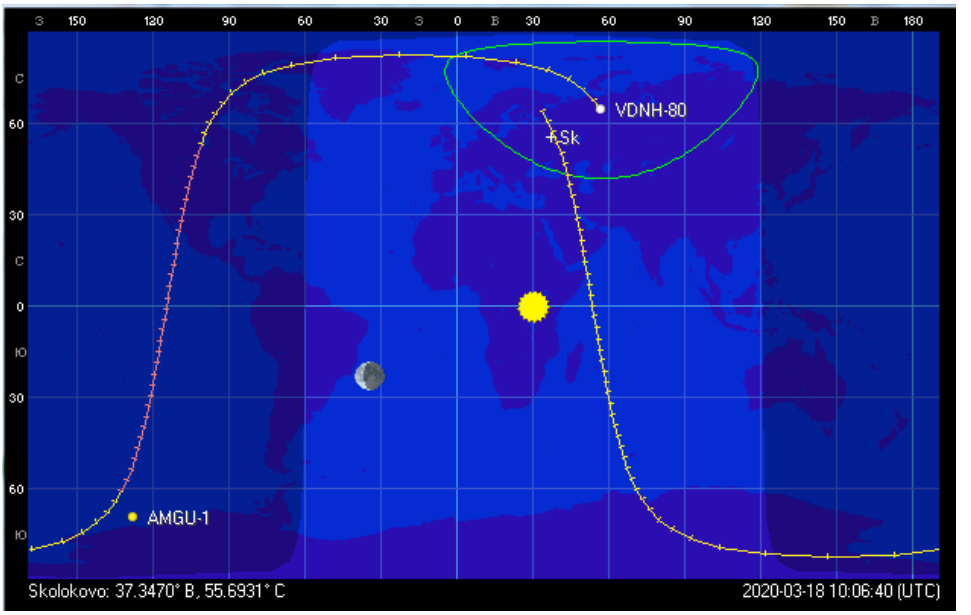


Fig. 5. Projection of VDNH-80 satellite orbit onto the Earth map when the DeCoR instrument was switched on March 18, 2020.

The orbital projection of the VDNH-80 satellite during the switching on of the DeCoR instrument on March 18, 2020 is shown in Fig. 5. It also shows the position of the AmurSat satellite at the moment when the DeKoR instrument was switched on the VDNH-80 satellite. The DeCoR instrument on the AmurSat satellite was switched on in $\sim$ 45 minutes above the communication point in Kaluga, thus, the data array transmitted from the AmurSat satellite corresponds to a trajectory close to that shown in the figure.

The time variation of the counting rate of electrons with energy $>$ 300 keV along the orbit of the VDNH-80 satellite, recorded in the monitor mode with a time resolution of 5 s are shown in Fig. 6a. The maxima are clearly visible at the moments when the satellite is in the outer radiation belt of the Earth, as well as some small increases in the polar region. Significant increases in the counting rate near the geomagnetic equator are also clearly visible.

The counting rates in a similar channel of the DeCoR instrument installed on the AmurSat satellite are shown in Fig. 6b. Noteworthy is the repeatability of the features visible when two satellites pass the same polar region with an interval of $\sim$ 45 minutes. During the flight of the geomagnetic equator, a smooth increase in the electron counting rate is also seen. The moving average line shows modulation with a period of $\sim$ 40 s associated with the rotation of the satellite. As noted above, the presence of such modulation indicates that the electron flux is non-isotropic in the near-equatorial regions.

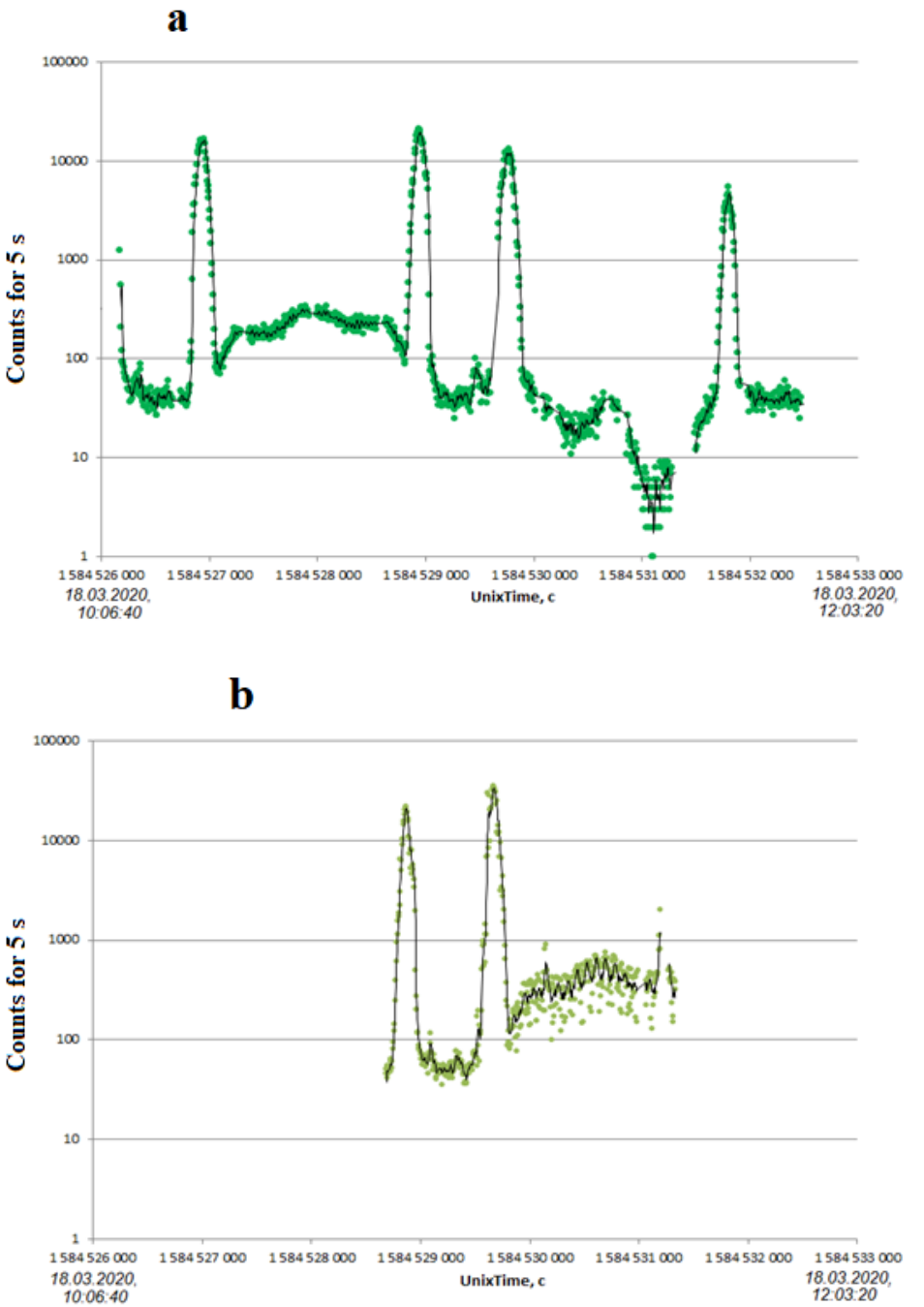


Fig. 6. Time variation of the count rate for electrons with energy > 300 keV, recorded in the monitor mode of the DeCoR instrument operated on-board VDNH-80 satellite on March 18, 2020 (a), similar readings of the DeCoR instrument on -board AmurSat satellite (b).

3.3. *The first experience of observations with DéCoR instruments on-board Norbi and DEKART satellites*

Just now the stage of flight tests of Norbi and DEKART missions was completed and regular observations were begun. The example of DeCoR instrument data obtained from Norbi CubeSat on July 14, 2021 is presented in Fig. 7. There it can be seen the monitor time dependent (with 5 s time resolution) count rates in plastic detector (mainly electrons) and CsI(Tl) (mainly gamma quanta) recorded for several consequent orbits. All typical variations caused by the near-Earth radiation flux distribution can be seen. There are seen peaks connected with the crossing of the outer belt regions and South–Atlantic Anomaly, as well, as in precipitation areas.

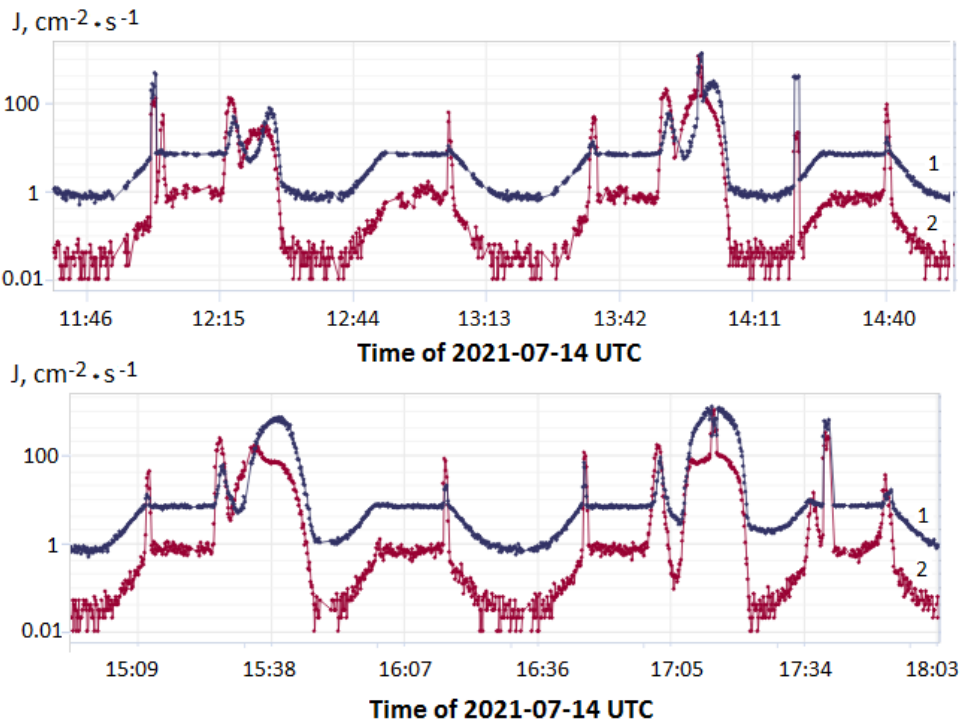


Fig. 7. Time variation of the count rate for electrons with energy > 300 keV (red curve) and for gamma-quanta (dark blue curve), recorded in the monitor mode of the DeCoR instrument operated on-board Norbi satellite on July 14, 2021.

It is necessary to note that gamma quantum count rate dependence along the orbit contains the intervals of quasi-constant background that correspond near-equatorial and polar cup regions. These intervals are quite appropriate for GRB detection. Nevertheless on these intervals flux increasing caused by electron precip-

itation can be observed. The example of such event detected on 04:49 September 28, 2021 is presented in Fig. 8.

The example of output data from one detector unit of DéCoR instrument on-board DEKART satellite is presented in Fig. 9. As it could be seen, time dependent gamma quantum and electron count rates, recorded with 0.5 s time resolution along the orbit, are quite similar to those obtained by Norbi satellite. Thus, it gives opportunity for multi-satellite measurements of time effects in near-Earth radiation fluxes and increase probability of GRB detection, including possibility of simultaneous GRB observation on two satellites.

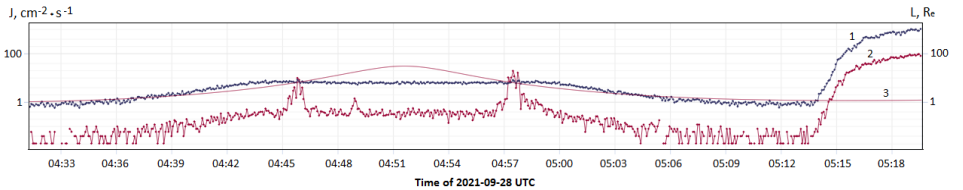


Fig. 8. Time variation of the count rate for electrons with energy > 300 keV (red curve), and gamma-quanta (dark blue curve), and L values (violet curve) recorded in the monitor mode of the DeCoR instrument operated on-board Norbi satellite on September 28, 2021.

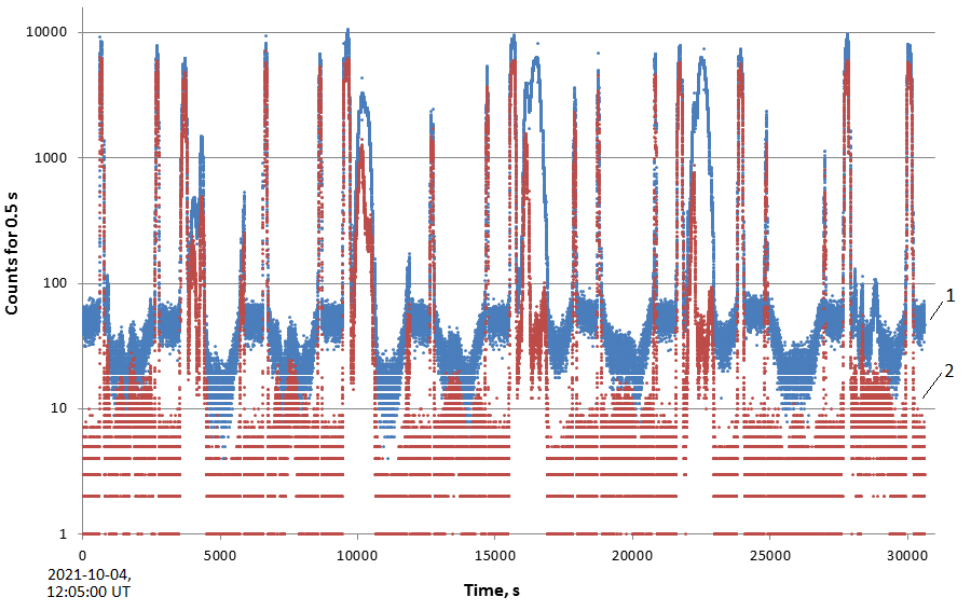


Fig. 9. Time variation of the count rate for electrons with energy > 300 keV (red curve) and for gamma-quanta (dark blue curve) recorded in the monitor mode of the DeCoR detector on-board DEKART satellite on October 4, 2021.

4. Conclusion

The successful realization of the CubeSat missions in frame of Universat–SOCRAT project make it possible for the first time to create a prototype of nano-satellite constellation a space system for monitoring of space radiation and electromagnetic transients including GRBs.

The first stage of this project was beginning from the launching of three CubeSats in July, 2019 and then three CubeSats in September, 2020. The prototypes of the instruments, which should be used in future Universat–SOCRAT missions were tested in these orbital experiments with CubeSats. Results of these tests confirmed that instruments operate well as in terms of command control as in terms of detector parameters. The particle flux measurements carried out during flight tests demonstrate the presence of variations corresponding to the expected changes in outputs at low polar orbit. During flight tests, both monitoring data with a time resolution of ~ 0.5, 1 and 5 s and event by event data with a time resolution of $\sim 20\mu s$ were successfully obtained from both DeCoR instruments installed on-board launched cubesats. Obtaining and processing of these data allows us to analyze both slow and fast particle flux variations. Thus, flight tests have confirmed the suitability of the instruments for the scientific research for which they were intended including principal possibility of GRB observation.

References

1. Abbott B. P., Abbott R., Abbott T. D., et al., "Multi-messenger Observations of a Binary Neutron Star Merger", *The Astrophysical Journal Letters*, V.848, Iss. 2, article id. L12, 59 pp., 2017.
2. Svertilov S.I., Panasyuk M.I., Bogomolv V.V., et al., "Wide-Field Gamma-Spectrometer BDRG: GRB Monitor on-board the Lomonosov Mission", *Space Science Review*, article id. V.214, 8, 22 pp., 2018.
3. Osedlo V.I., Sadovnichii V.A., Panasyuk M.I. et al., "Multi-satellite project "Universat-Socrat" for natural and artificial hazards monitoring", *Advances in the Astronautical Sciences*, V.163, pp. 45-52, 2018.
4. Sadovnichiy V.A., Panasyuk M.I., Lipunov V.M., et al., "Project "Universat-SOCRAT" of Multiple Small Satellites for Monitoring of Natural and Technogenic Space Hazards", *Open Astronomy*, V.27. Iss. 1, pp. 126-131, 2018.
5. Panasyuk M.I., Klimov P.A., Svertilov S.I., et al., "Universat-SOCRAT multi-satellite project to study TLEs and TGFs", *Progress in Earth and Planetary Science*, V.6, pp. 35-54, 2019.
6. Panasyuk M.I., Svertilov S.I., Bogomolov V.V., et al., "RELEC mission: Relativistic electron precipitation and TLE study on-board small spacecraft", *Advances in Space Research*, V.57, pp. 835–849, 2016.
7. Svertilov S.I., Panasyuk M.I., Petrov V.L., et al., "Multi-Satellite Project UNIVERSAT-SOCRAT of Cubesat Grouping for Spacecraft and Aviation Radiation Hazard Warning System and First Experience of Moscow University Cubesat Missions", *Advances in the Astronautical Sciences*, V.173, pp. 171-188, 2018.

Multiwavelength observations of GRB160625B by MASTER, Lomonosov, Konus-Wind and three stage collapse

V. M. Lipunov[1,2,*], V. A. Sadovnichy[3], M. I. Panasyuk[1,4], I. V. Yashin[4], S. I. Svertilov[1,4], D. Svinkin[5], E. Gorbovskoy[2], S. G. Simakov[2], G. V. Lipunova[2], V. G. Kornilov[1,2], D. Frederiks[5], V. Topolev[1,2], R. Rebolo[6], M. Serra[6] , N. Tiurina[2], E. Minkina[1,2], V. V. Bogomolov[1,4], A. V. Bogomolov[4], A. F. Iyudin[4], A. Chasovnikov[1,2], A. Gabovich[8], N.M. Budnev[7], O. A. Gress[2,7], G. Antipov[2], D. Vlasenko[1,2], P. Balanutsa[2], R. Podesta[5], K. Zhirkov[1,2], A. Kuznetsov[2], V. Vladimirov[2], F. Podesta[5], C. Francile[5], Yu. Sergienko[8], A. Tlatov[9], O. Ershova[7], D. Cheryasov[2] and V. Yurkov[8]

[1] *Lomonosov Moscow State University, Physics Department, Vorobievy Hills, 1, Moscow 119991, Russia*

[2] *Lomonosov MSU, SAI, Universitetsky, 13, Moscow 119234, Russia*

[3] *Lomonosov Moscow State Universit, GSP-1, Vorobievy Hills, 1, Moscow, 119991, Russia*

[4] *Lomonosov Moscow State University, Skobeltsyn Institute of Nuclear Physics (SINP MSU), Vorobievy Hills, 1, Moscow, 119991, Russia*

[5] *Ioffe Institute of theoretical physics, 26 Politekhnicheskaya, St Petersburg 194021, Russia*

[6] *Instituto de Astrofísica de Canarias, Lactea, E38205, LaLaguna, Tenerife, Spain*

[7] *Irkutsk State University, 20, Gagarin, Irkutsk 664003, Russia*

[8] *Blagoveschensk State Pedagogical University, Lenin, 104, Amur, Blagoveschensk 675000, Russia*

[9] *Kislovodsk Solar Station Pulkovo Observatory, Gagarina 100, Kislovodsk 357700, Russia*

**E-mail: lipunov@sai.msu.ru*

The detailed continuous fast optical photometry analysis obtained by MASTER Global Network for the GRB160625B optical counterpart MASTER OT J203423.51+065508.0 is presented. There are also hard X-ray and gamma-ray emission obtained by the Lomonosov and Konus-Wind spacecrafts detectors. We detected quasiperiodic emission components in the intrinsic optical emission of GRB160625B and propose a three-stage collapse scenario for this long and bright GRB. We associate quasiperiodic fluctuations with forced precession of a Spinar, i.e. self-gravitating rapidly rotating super dense body, whose evolution is determined by a powerful magnetic field. The spinar's mass lead it to collapse into a black hole at the end of an evolution.

Keywords: Gamma-ray burst

1. Introduction

Gamma-ray bursts (GRB) still remain a mysterious phenomenon. We proceed from the assumption believe that long gamma-ray burst occurs as a result of the collapse of the rapidly rotating core of a massive star. The fast rotation of the core slows down the collapse and extends time available to produce electromagnetic radiation. Two scenarios are possible here. In the first scenario (MacFadyen & Woosley, 2000),[1] a black hole first forms, and then the fallback of the supernova envelope, which has a supercritical torque, forms a heavy superdense disk. Due to the generation of magnetic fields by this disk, axial jets with a large gamma factor are generated, which we observe. In another scenario, a rapidly rotating magnetized object, a Spinar, is first formed, which is slowly compressed due to the dissipation of the rotational moment (Lipunov & Gorbovskoy, 2007).[2] In this case, a jet with a Poynting-Umov energy flow is formed along the rotation axis. The operating time of the central gamma-ray burst engine changes depending on the dissipation rate. In general, both scenarios require a fairly large torque in the collapsing stellar core. And here the authors are impressed by a binary scenario in which fast rotation occurs due to the tidal influence of the second component in a very close binary system (Tutukov & Cherepashchuk, 2016).[3] In this scenario, the centrifugal barrier is a consequence of the large torque acquired as a result of the natural evolution of the binary system. In the case of GRB160625B our attention was attracted by the quasiperiodic brightness fluctuations during the time of the central engine operation, and below we try to interpret them as a consequence of a Spinar Paradigm at work (Lipunova & Lipunov, 1997; Lipunov & Gorbovskoy, 2007;2008; Lipunova et al., 2009)[2,4–6]

In this paper, we present multiwavelength observations of GRB160625B, one of the brightest gamma-ray bursts in the history of their study. We have concentrated on the temporal behavior of itsintrinsic electromagnetic radiation. We present optical and gamma-ray data, recorded during the time of the operation of the central engine of GRB160625B. We made an attempt to find traces of the duality of the GRB system. We have suspected traces of the dual nature of long GRBs. Of course, we do not have one hundred percent proof of this scenario, but this study can serve as an example of searching for and finding the dual nature of long GRBs.

2. Observations

GRB160625B was observed by the large number of space and ground telescopes in a wide range of electromagnetic waves from gamma-ray to radio. For the first time in the GRB study history, the polarization of its own prompt optical emission synchronous with the gamma one was discovered (Troja et al., 2017a).[7] Now we present the details of synchronous observations in the optical (MASTER OT J203423.51+065508.0) and gamma-ray ranges.

GRB 160625B triggered Fermi observatory at 2016-06-25 22:40:16UT , firstly as a short pulse (Fermi-GBM trigger 488587220, Burns et al.2016),[8] then the

Fermi-LAT triggered (Dirirsa et al. 2016)[9] at 22:43:24.82UT, at 22:51:16.03 Fermi-GBM triggered again (trigger 488587880).The GBM light curve consists of multiple peaks over approximately 700 seconds, the first one being 1 second long soft peak. The main peak, corresponding to the LAT trigger, was very hard and about 25 seconds long. The peak that triggered GBM for the second time was soft and about 30 seconds long (Burns et al. 2016).[8] Swift has initiated a series of observations and discovered uncatalogued X-ray sources at R.A.,Dec.(2000)=20:34:23.25 +06:55:10.5 with an uncertainty of 3.8 arcsec (Melandri et al. 2016), publishied it at 2016-06-26 06:54:13UT.

This long, extremely bright GRB 160625B also triggered BDRG detectors on-board Lomonosov Space Observatory of Moscow State University (Sadovnichii et al. 2017)[10](Fig. 1) and Konus-Wind at 22:40:19.875UT (Svinkin et al. 2016)[11](Fig. 1) as well as the CALET Gamma-ray Burst Monitor (CGBM) at 22:40:15.49 (Nakahira et al. 2016).[12]

The optical counterpart was discovered by RATIR (Troja et al. 2017b)[14] starting 8.53h after LAT trigger. MASTER started observation in polarization filters (Lipunov 2010, 2019, Kornilov 2012)[17–19] 31 sec after GBM notice time (57 sec after GBM, i.e. 131 sec before LAT Trigger, Dirirsa et al.2016)[9] at 2016-06-25 22:41:13UT (Gorbovskoy et al. 2016),[20] but published the circular in GCN (Barthelmy et al. 1998ab)[21,22] at 16/06/28 14:05:38. The GCN publication was delayed by the installation of the new MASTER telescope in Argentina at this time (MASTER-OAFA). MASTER measurements probed the structure of the magnetic field at an early development stage of the jet, closer to a central black hole, and show that the prompt emission phase is produced via fast-cooling synchrotron radiation in a large-scale magnetic field that was advected from the black hole and distorted by dissipation processes within the jet (Troja et al. 2017a).[7] The optical data obtained by MASTER telescopes-robots (Lipunov et al. 2010,2019)[17,18] have the best temporal resolution with the minimum exposure time of 5 seconds. This resolution made it possible to suspect quasiperiodic variability in the optical range, which we try to associate with the dual nature of the long GRB.

The detection of a substantial (8.3 ± 0.8 per cent from our most conservative estimation) variable linear polarization of a prompt optical flash that accompanied the extremely energetic and long prompt γ-ray emission from GRB 160625B was discovered by MASTER (Gorbovskoy et al. 2016a, Lipunov et al. 2016a, Troja et al. 2017a).[7,15,16]

2.1. *MASTER*

MASTER Global Robotic Net of Lomonosov Moscow State University in 2016 consisted of orbital MASTER-Shok detectors on-board Lomonosov spacecraft and of 8 ground observatories with identical scientific equipment distributed all over the Earth: MASTER-Amur, -Tunka, -Ural, -Kislovodsk, -Tavrida in Russia, -SAAO (South African Astronomical Observatory), -IAC (Tenerife, Spain, Teide

observatory of Institute of Astrophysics of the Canary Islands), -OAFA (Argentina, San Juan National University Astronomical Observatory named by Felix Aguilar), see Lipunov 2010, 2019[17,18](Fig. 1). Identical equipment includes twin wide-field (MASTER-II, 4-8 square degrees up to 20m unfiltered limit at 180s exposition) and very wide field (MASTER-VWFC, 800 square degrees, up to 15m at summary image) optical channels (Kornilov et al. 2012)[19] which allows us to follow a target 24h per day in one photometric channel. MASTER-VWFC are the ground-based analogue of the MASTER-SHOCK cameras, that were installed on board the Lomonosov space observatory (Sadovnichy et al. 2017,2018, Lipunov et al. 2018, Svertilov et al.2018, Park et al. 2018).[10,23,25,37,38] Observations by MASTER-VWFC are unfiltered.

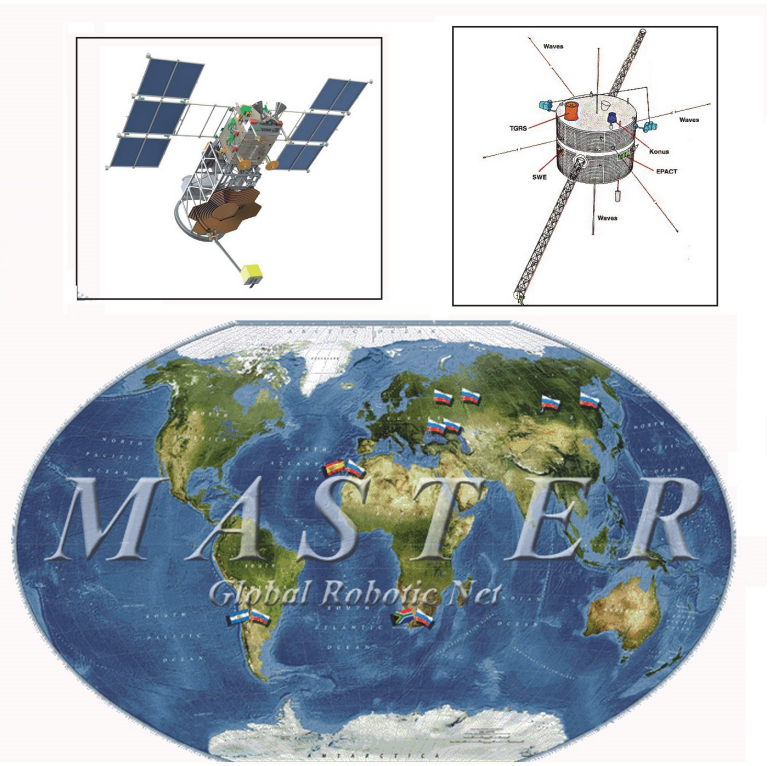


Fig. 1. MASTER Global Robotic Net. The upper left part shows Lomonosov satellite and the upper right part shows Konus-Wind. The flags indicate equipped sites of network stations.

MASTER own developed photometer can observe simultaneously in 2 orthogonally oriented polarization filters in both tubes of twin MASTER telescope

(Kornilov et al. 2012)[19] or in BVRI filters and unfiltered. Mount has fast positioning with the speed up to 30 deg/s.

MASTER key factors is own software for real-time reduction. It includes full robotization of all processes: hardware control, weather control, efemerides, central planner, automatic evening/morning calibration,and primary image reduction, astrometry and photometry, extraction of new optical sources and notification of the main station of them). Such features let us to discover significant and variable linear polarization during the prompt optical flash of GRB 160625B (Troja et al. 2017a),[7] to discover GRB optical counterparts (Lipunov et al. 2016, Gorbovskoy et al. 2016, Sadovnichy et al. 2018, Laskar et al. 2019),[20,24–27] to discover Smooth Optical Self-similar Emission of GRB subclass (Lipunov et al. 2017),[28] to make the most optical support to GW150914 event (Abbott et al. 2016ab, Lipunov et al. 2017,2018),[29–31] to independently discover Kilonova GW170817(Abbott et al. 2017, Lipunov et al. 2017),[32,33] to make the most optical support to the follow-up of a rare IceCube neutrino multiplet (Aartsen et al. 2017),[34] to discover V404Cygni polarization variability (Lipunov et al. 2017),[35] to make optical observationsthat revealed a strong evidence for a high-energy neutrino progenitor - blazar TXS 0506+056 for IC170922A (Lipunov et al. 2020)[36] , to discover more than 2000 optical transients of 10 different types and other.

2.2. *MASTER GRB160625B observation*

MASTER-IAC robotic telescope pointed to the GRB 160625B at 2016-06-25 22:41:13UT, and observed Swift X-ray error box (Melandri et al., 2016)[13] by MASTER-IAC wide field camera with a 5 second exposition time without filter and imaged several thousand frames.

MASTER-Tavrida started first frame exposition 12 seconds after LAT notice time (66s after trigger time) at 22:44:30UT.

MASTER-Tavrida and MASTER-IAC very wide field of view cameras (VWFC) observed GRB error-box when the Fermi-LAT detected the main impulse of GRB with a high coordinate accuracy of 0.5 degrees, 131 seconds after the first message.

MASTER OT J203423.51+065508.0 optical counterpart was detected at 22:43:30UT (+2.5 sec of exposition) UT with m_{OT}=8.60 and was getting brighter up to a maximum of 7.86m(see Fig. 2). The prompt optical emission strongly correlated with Konus-Wind gamma ray light curve (Svinkin et al., 2016).[11]

MASTER-IAC robotic telescope started observation of the error box 43 sec after LAT notice time or 95 sec after trigger time at 2016-06-25 22:44:59UTby main MASTER-II telescope in two polarizations. The OT was 8.6m at the moment.

As a result, MASTER not only registered the entire event of the GRB explosion in the optical range with the best time resolution (2.5 sec), but for the first time in the history of gamma-ray burst research, it detected the polarization of prompt optical emission from the gamma-ray burst while the flash was still going on (Troja et al. 2017a).[7]

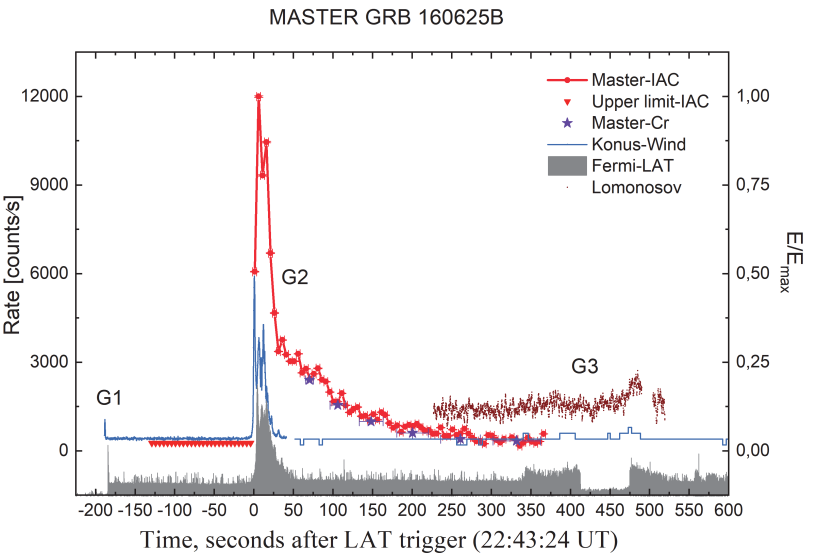


Fig. 2. Red curve - observations of the very wide-field MASTER-IAC cameras, reduced to the proper time of the system (compression of the burst time scale relative to the observed one). The purple dots corresponds to the MASTER-Tavrida measurements taken with a lower temporal resolution. The gray curve corresponds to the emission detected by Fermi-LAT (see Toja et al. 2017a.[7]

The gamma-ray burst GRB160625B turned out to be one of the most powerful cosmic explosions of this type, which appeared as a narrow jet of relativistic particles accelerated by the electromagnetic field of a rapidly rotating black hole at the other end of the Universe forming before our eyes.

An analysis of the MASTER's polarization observations made it possible for the first time to detect the polarization of GRB's intrinsic optical radiation and directly showed that the muzzle of the most powerful space gun was formed by an ordered powerful magnetic field of a forming black hole (Troja et al., 2017).[7]

This magnificent astrophysical experiment succeeded thanks to the interaction of scientists from several countries, who created unique robotic equipment to detect gamma rays, infrared radiation, and photons in the optical range born by the GRB event.

2.3. *LOMONOSOV Observations*

The GRB monitor (BDRG) aboard the Lomonosov (Sadovnichii et al. 2017)[10] observatory (hereafter BDRG/Lomonosov (Svertilov et al., 2018)[37] was built for the early detection of GRBs in the gammas of 0.01-3.0 MeV energy range and for generation of triggers for those events. The BDRG consisted of 3 identical detector units

connected to the electronic unit. The BDRG instrument detector units (blocks) were mounted on the spacecraft payload platform in such a way that their axes are oriented 90° to each other. Each detector has a cosine angular dependence for a sensitive area not shaded by satellite construction elements within $\sim 60°$ of its axis. The monitor's central axis, relative to which the detector axes are inclined, is directed toward to the local zenith. Thus, the total field of view for all three detectors is about 2π sr; and one quarter of this field, i.e. $\pi/2$ sr, is the value of a solid angle, within which limits the GRB position error can be estimated with sufficiently good accuracy through the comparison of all three detector outputs.

BDRG operates in two main observational modes: the monitor or continuous mode, and the burst mode. In the monitor mode all instrument outputs were recorded and stored continuously with time resolutions adjustable by commands from Earth. On the other hand, the burst mode was activated by on-board instrument triggers to record detailed information of each photon during the before-burst, burst, and after-burst time intervals. The BDRG trigger initiated the estimation of GRB positions and relayed the trigger not only to other GRB observation instruments onboard the Lomonosov spacecraft, i.e. SHOK optical cameras (Lipunov et al., 2018)[23] and Ultrafast Flash Observatory (Park et al., 2018),[38] but to the ground telescopes as well through the Gamma-ray Coordinates Network (GCN, Barthelmy et al. 1998ab)[21] viathe Global Star transmitter.

Each BDRG detector unit consisted of a thin layer (0.3 cm) of NaI(Tl) crystals optically coupled to a considerably thicker layer (1.7 cm) of CsI(Tl) crystals situated underneath. The diameter of the detectors is 13 cm, and both layers are read by a single photomultiplier tube (PMT). Thus, the overall detector area is about 130 cm2. The thickness of the NaI(Tl) layer is optimized for the soft part of energy range, and the working ranges of the units are 0.01-0.5 MeV for the NaI(Tl) layer and 0.05-3 MeV for CsI(Tl). As such, the NaI(Tl) layer serves as the main detector for hard X-ray timing, while the CsI(Tl) operates as an active shield against background gamma-rays. Additionally, the CsI(Tl) crystals can also detect gamma-rays with energies up to a few MeV. The difference in decay times for the NaI(Tl) ($\sim$0.25 ms) and CsI(Tl) ($\sim$2.0 ms) crystals permits the separation of light flashes in the scintillators through special electronic circuits that differentiate pulse shapes.

The information provided by the BDRG units consisted of a number of different categories for the data frames generated continuously (continuous mode) as well as irregularly by various triggers (burst mode). The continuous data stream included three types of frame corresponding to the instruments' monitoring, spectrum, and event. Monitoring frames provided count rates in 8 energy channels for the NaI(Tl) and CsI(Tl) scintillator crystals for each of the BDRG detector units, while spectrum frames contain 724 channel spectra for NaI(Tl) and CsI(Tl), separately. Event frames gave the primary values for energy release within the NaI(Tl) and CsI(Tl) crystals, combined with time data for a fixed number of detected gamma-quanta. Likewise, information about the main parameters for all GRB triggers was stored

and transferred in the form of "trigger logs". There are three trigger types categorized as "fast," "slow" and "super-slow," with characteristic times of 10 ms, 1 s, and 20 s, respectively. Corresponding to each trigger type, three data frame sequences for the monitoring, spectrum, and event were generated continuously in a manner similar to the continuous mode discussed above. A portion of data collected before the trigger was always included for all trigger types.

On 25 June 2016 near the GRB160625B trigger, the BDRG/Lomonosov operated in monitoring mode. The background environment at the time of event was very complicated. The Lomonosov satellite was flying through the radiation belts.

The count rate variations in the BDRG/Lomonosov gamma-quanta channels are illustrated by the curves in the upper panel of Fig. 3, in which count-rates of three BDRG-1 NaI detector channels 10 – 35, 35 – 170, 170 – 650 keV shown with the time resolution of 0.1 s are plotted. Full data is available on the link(http://master.sai.msu.ru/static/MG16MASTER/BDRG160625B.txt)

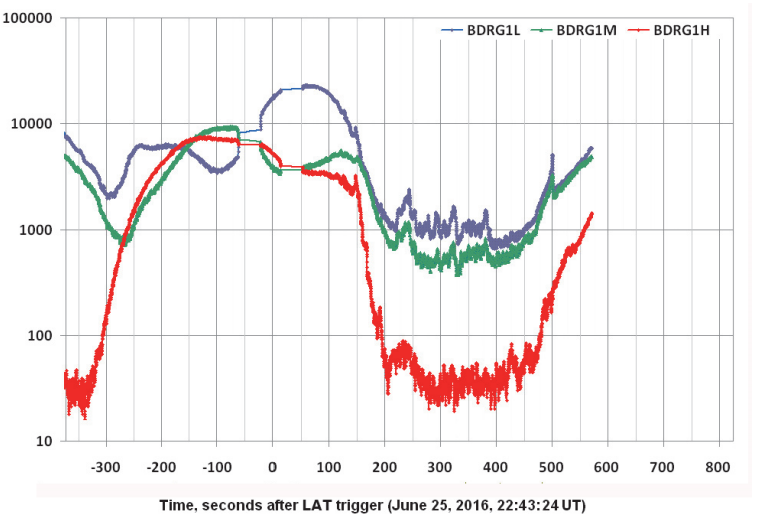


Fig. 3. GRB160626B γ-ray light curves.The BDRG-1(NaI detector) counts vs time in the 10-35keV (BDRG1L), 35-170 keV (BDRG1M), 170-650 keV (BDRG1H) ranges.

The clear count-rate increases in the time interval from about -150 to 350 seconds relative to the Swift trigger time (-300 to 200 seconds relative to the LAT trigger time) corresponds to the satellite crossing outskirts of the external radiation belt. After that the satellite flew into the region of the South polar cap, where the background is smaller, and after that it began to cross the outer belt again and the background began to increase. Thus, due to such background variations the main pulse (G2) as well as the precursor (G1) could not be observed on by the

Lomonosov satellite detectors. Only the "last tail" (G3) could be observed. However, during the corresponding time interval (from about 350 to 550 seconds from LAT trigger time) count variations in the BDRG/Lomonosov were in reality combination of GRB counts plus variations caused by unstable electron fluxes in the polar cap. The last component can give the significant input in gamma ray channels of the instrument due to electron bremsstrahlung. Thus, to obtain the real GRB light curve it is necessary to clean the detector outputs from electron background variations.

To realize this procedure we used the outputs from different BDRG detectors. Because of different BDRG detector orientation the given GRB source was observed by separate detectors at different angles. In the case of GRB160625B only BDRG1 detector unit was illuminated, the angle between the detector axis and direction toward the GRB source was about 56°, while for BDRG2 and BDGR3 these angles were about 136° and 116° respectively. It means that GRB source was out of field of view (FOV) of both latter detector units. On the other hand, since counting rate of the irregular variations observed in the polar cap regions are mainly due to the quasi-trapped electron fluxes having an anisotropic but rather wide pitch-angle distribution, they will exhibit similar temporal behavior in separate, although differently oriented detectors. It allows us to use regression analysis of detector unit count rates obtained for two detectors during time interval of GRB observation to estimate regression coefficients, which then can be used for rejection of the part of counting rate variations caused by electron fluxes. Because during the time of GRB160625B observation BDRG2 unit was switched-off, for regression analysis only the BDRG1 and BDRG3 unit outputs were used. To be exact, we selected time interval from 200 to 450 seconds after LAT trigger (22:43:24UTC) to estimate regression coefficients, which corresponds to the time between G2 and G3 events, when expected input to detectors counting rates from GRB160625B was negligible.

2.4. *Konus-Wind observations*

GRB 160625B triggered Konus-Wind GRB spectrometer (KW, Aptekar et al. 1995)[41] at T0= 81619.875 s UT (22:40:19.875; Svinkin et al., 2016).[11] The burst was detected by the S2 detector, which observes the Northern ecliptic hemisphere; the incident angle was 65.2 deg. The propagation delay from Earth to WIND is 3.356 s for this GRB; correcting for this factor, the KW trigger time corresponds to the Earth-crossing time 81616.519 s UT (22:40:16.519).

2.4.1. *Time history*

Count rates for Konus-Wind are recorded in three energy bands in the triggered mode: 17–70 keV (B1), 70–300 keV (B2), and 300–1170 keV (B3). The record starts at T0 - 0.512 s and continues to T0 + 229.376 s with an accumulation time varying from 2 to 256 ms. Waiting-mode count rate data are available up to T0 + 250 s

in the same energy bands with a coarse temporal resolution of 2.944 s. A source activity after time interval T0 + 250 s may be traced in the housekeeping mode with temporal resolution of 3.68s for the 70–300 keV energy range.

The prompt-emission light curve (Fig. 4) can be divided into three episodes. It starts, at ∼ T0 - 0.3 s, with a short, spectrally-soft initial pulse (precursor) which has duration of about 1 s. The precursor is followed, starting at ∼ T0 + 180s, by the main, extremely bright and spectrally-hard emission episode lasting for about 40 s. The final episode observed by KW in the housekeeping mode starts at ∼ T0 + 530 s and has duration of about 150 s. The total burst duration is about 680 s.

The first episode was localized only by Fermi (GBM) with a position uncertainty of about 2 deg (statistical only). The difference in the arrival time of the gamma-ray signals at Fermi-GBM and Konus-Wind provides additional significant constraints on the gamma-ray localization of the episode (Zhang et al., 2018).[42] The triangulation of the first episode is consistent with the source position determined by Swift-XRT in the main episode, supporting the association of the precursor and the burst.

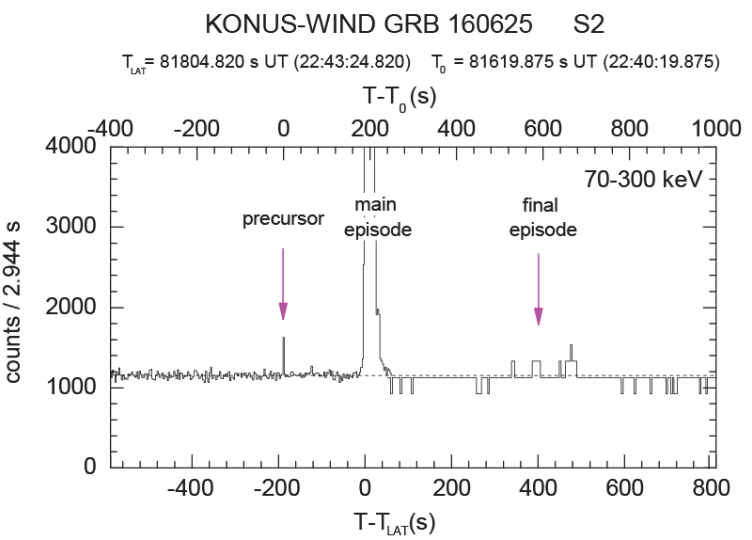


Fig. 4. GRB 160625B light curve recorded in the KW waiting and housekeeping modes in the 70-300 keV band (B2).

Let us discuss the time-resolved spectral analysis. In the triggered mode, Konus-WIND measures 64 energy spectra in 128 channels of two overlapping energy bands: 20–1170 keV (PHA1) and 244 keV–15 MeV (PHA2). The first four spectra have a fixed accumulation time of 64 ms; after that, the accumulation time varies over 0.256-8.192 s, depending on the current intensity of the burst. Five initial energy

spectra covered the precursor (T0 – T0+8.448 s) and 35 covered the main episode (T0+180.480 - T0 + 237.824 s).

The spectral analysis was performed with XSPEC version 12.9.0i (Arnaud 1996)[43] with the three model: Band GRB function (Band et al. 1993):[44] $f(E) \sim E^{\alpha} exp(-(2+\alpha)E/Epeak)$ for $E < Epeak(\alpha-\beta)/(2+\alpha)$, and $f(E) \sim E^{\beta}$ for $E > Epeak(\alpha-\beta)/(2+\alpha)$, where α is the power law photon index at low energies, Epeak is the peak energy in the EF(E) spectrum, and β is the high-energy photon index; a cutoff power law model (CPL): $f(E) \sim E^{\alpha} exp(-(2+\alpha)E/Epeak)$; and a simple power law. The spectral models were normalized to the energy flux in the 20 keV – 10 MeV range, a standard band for the KW GRB spectral analysis.

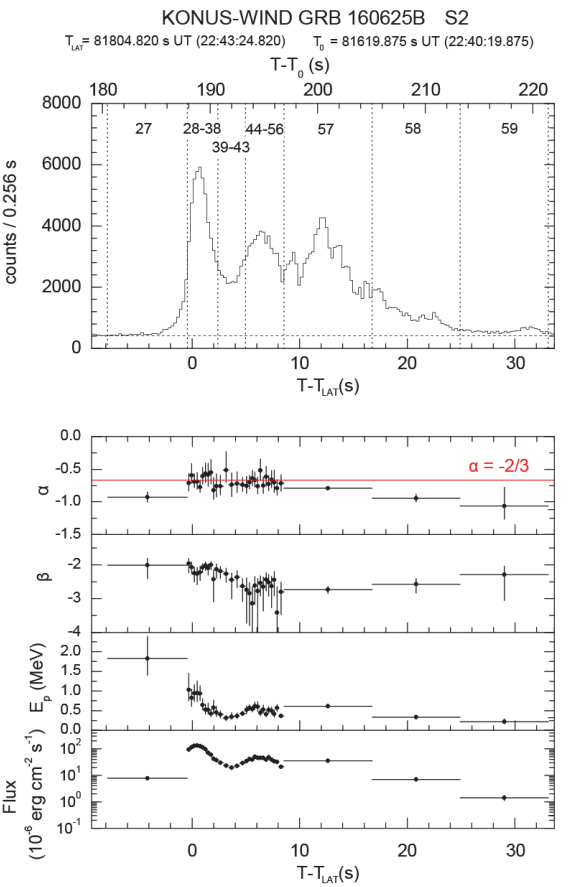


Fig. 5. Spectral evolution of the gamma-ray emission during the prompt phase of the burst main episode. The Konus-WIND light curve in the combined B1+B2+B3 energy band (17-1170 keV) is shown with 256 ms resolution, along with the temporal behavior of the Band spectral model parameters Epeak, α, and β obtained from the time-resolved fits.

Typically, the spectral channels are rebinned to have at least 10 counts per energy bin to ensure Gaussian-distributed errors and the correctness of the χ^2 statistic.

For the precursor, constrained spectral parameters of the CPL model are available only for the sum of the first four spectra, while for the main episode, a good count statistic is achieved for 35 individual spectra between T0+180.480 s and T0+237.824 s, the spectrum between T0+ 229.632 s and T0+237.824 s is well described by a simple power law. The good spectral coverage enables us to construct the temporal behavior of the model parameters (α, β, Epeak) and to trace in detail the evolution of the spectral composition of radiation over the course of the main episode (Fig. 5).

Spectrum 27 was measured at the onset of the very intense initial pulse of the main episode (Fig. 5). The emission at this moment is hard; Epeak reaches the highest value for the burst ($\sim$ 1.8 MeV). After the summit of the brightest pulse, Epeak starts to decrease gradually (spectra 28-38), down to $\sim$ 300 keV in spectra (39-43), and then grows up to $\sim$ 600 keV at the time of the second peak (spectra 44-56), staying at this level during the last peak (spectrum 57). Spectra 58-60 describe the decay of the main episode showing the typical decrease in Epeak. During the brightest part of the main episode the low-energy spectral index is approximately constant and is consistent with the synchrotron emission in the slow-cooling regime $\alpha = -2/3$ (Preece et al., 1998).[45] The high-energy index β shows no significant correlation with energy flux and is typical for long GRBs ($\beta \sim$ -2).

Spectrum 1-4 (time –averaged), corresponding to the peak of the initial pulse is well fitted with a CPL model with $\alpha \sim$ -0.4, and Epeak $\sim$ 70 keV (χ^2/dof = 16.5/29). The fit with a single blackbody (BB) component yields kT= 17.3 (-1.6, +1.6) keV, energy flux 1.92 (-0.19,+0.19) $10^{-6} erg/cm^2/s$, with $\chi^2/dof = 32.1/30$; the fit underestimates count rate at energies below $\sim$ 30 keV. Thus, despite the BB kT is consistent with the value found by Zhang et al., 2018,[42] we argue that BB model cannot be favoured for the spectrum.

3. Modeling

Looking at the optical light curve starting from 20 seconds relative to the main pulse and up to 200 - 250 seconds, one can suspect quasiperiodic brightness oscillations with an amplitude significantly exceeding the random brightness measurement error. Recall that a five-second exposure on ultra-wide field cameras is practically without delay. Such variability at times of the order of ($\tau \sim$ 10-20 seconds) is caused by internal physical processes in the operation of the central engine. However, it would be tempting to associate this phenomenon with the interaction of a relativistic jet with quasiperiodic inhomogeneous layers in the radial distribution of the progenitor's stellar wind. This type of inhomogeneity can be caused by the presence of a close second component in the collapsing star's binary system. Calculations carried out earlier (Lipunova et al. 2009)[6] show that in systems with an orbital period of less than 1 - 2 hours, tidal forces lead to a critical increase in the spin

moment of the collapsar. As a result, it is under these conditions that the formation of a massive accretion disk (Woosley 1993)[46] or a spinar (Lipunov & Gorbovskoy, 2007)[2] prolongs the magneto-rotational collapse, which leads to the phenomenon of a gamma-ray burst.

First of all, we will discuss the possible reasons for the quasiperiodic optical emission, which is an intrinsic emission at the stage of the central engine operation and must reflect some quasiperiodic oscillations of the central engine. This circumstance is supported by the fact that we detect oscillations after 30 seconds (in our own frame of reference) from the trigger, that is, at the same time when the optical and gamma spectra become one. Recently, Suvorov and Kokkotas (2021)[47] discussed quasiperiodic pulsations of X-rays from some short GRBs in the magnetar model. Actually, the only property of the magnetars used here is the one inherent to all radio transmitters they loose energy approximately according to the magnetic dipole law (Pacini, 1967)[48] and their total power changes over time $L \sim t^{-2}$. In connection with the gamma-ray bursts, this circumstance was expressed by Lipunova & Lipunov (1998).[4] The merit of the works of Suvorov and Kokkotas (2021)[47] is the attraction of the idea that the observed quasiperiodic oscillations are associated with the free precession of a neutron star with an anomalously strong magnetic field. In fact, the magnetar model is only a special case of a more general model of the magnetorotational collapse - Spinar Paradigm (Lipunov & Lipunov, 1998; Lipunov & Gorbovskoy, 2007, 2008; Lipunova et al., 2009).[2,4–6] In the framework of Spinar Paradigm, it is assumed that the initial rotational moment of the body is so great that centrifugal forces have a significant effect on the collapse process. In particular, this model successfully explains not only the plateau phenomenon, but also a sharp (by several orders of magnitude) cliff at the end of the plateau (Lipunov & Gorbovskoy, 2007).[2] In fact, an approximate non-stationary model of gravitational-rotational collapse (Lipunov & Gorbovskoy, 2008)[5] includes all relativistic effects plus the contribution of the nuclear forces of the neutron liquid. The collapse character depends on three main parameters: the core mass M of the collapsing star, the generalized Kerr parameter $a_0 = I\omega_0 c/GM^2$ and the ratio of magnetic and gravitational energy $\alpha_m = U_m/U_{gr} << 1$ which remains constant in the approximation of the conservation of the magnetic flux. A black hole or a neutron star can also be the end product of the collapse (Lipunova et al., 2009).[6] In the process of formation of both types of objects, an intermediate object is formed - a spinar - which can experience not only free, but also forced precession. In addition, if a neutron star can only slow down, decreasing the overall luminosity, the spinar can accelerate as it evolves, and this will even be accompanied at certain stages by an increase in luminosity. It is precisely by the spinar precession that we propose to explain the suspected oscillations of the intrinsic optical emission of the GRB160625B gamma burst. Remarkably, the magneto-rotational collapse model naturally explains the existence of the precursor, with which we begin.

In the spinar model, the precursor of the gamma-ray bursts occurs during the first abrupt stop of the collapse due to the increase in centrifugal force (Fig. 6).

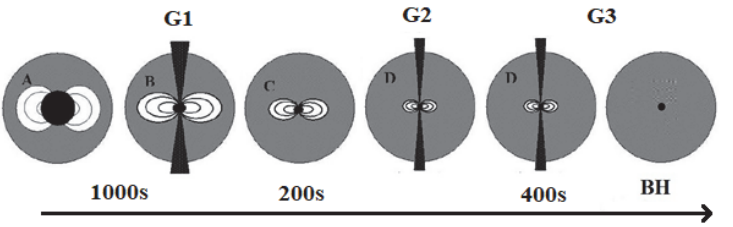


Fig. 6. Scheme of the three-stage magneto-rotational collapse in the Spinar Paradigm.

At this moment, half of the gravitational energy accumulated by the free-falling progenitor core is converted into heat. The second half of the work of gravitational forces turns into the energy of the spinar rotation. In other words, the centrifugal forces stop the collapse and a spinar is formed - which continues to slowly contract, losing its extreme moment and constantly increasing its radiation power (see Fig. 6). This is the formation of the precursor! Its maximum power $E_{precursor} \approx GM^2/2R_{sp}$ is determined by the radius of the formed spinar (Lipunov & Gorbovskoy, 2007)[2] $R_{sp} \approx a_0^2 R_g/2$, where R_g – Shwarcshild radius:

$$E_{precursor} \approx GM^2/2R_{sp} \approx (1/2a_0^2)Mc^2 \quad (1)$$

If the core torque is not too high $a_0 < 5-10m^{-1}(m = M/M_0)$, then this energy directed mainly along the spinar rotation axis will pierce the progenitor shell and we will see the precursor (Lipunov & Gorbovskoy, 2007).[2] However, immediately after the formation of the Spinar, the energy released by the spinar will already be determined by the magnitude of the magnetic energy $L_{sp} = U_m\omega \approx \alpha_m U_{gr}\omega$. Accordingly, the energy released by the spinar in one spin period and the corresponding pressure impulse on the shell will be reduced by a factor of $\alpha_m << 1$, hence the jet breakdown will happen. But if the torque is lost, the luminosity of the spinar will grow $L_{sp} = U_m\omega \approx \varepsilon_m U_{gr}\omega$ and finally break through the shell. This will be the time close to the beginning of the gamma-ray burst t_{GRB}. However, the time elapsed from the precursor to the start of the gamma-ray burst will be determined by the rate of loss of the spinar torque at the moment of its formation:$\Delta t \approx I\omega/U_m$. We can express the magnetic energy through the magnetic flux $\Phi = \pi BR^2 = \Phi_{28}10^{28}Gs\,cm^2$, thus normalizing the flux to the characteristic of magnetars the does not change during the collapse, according to our assumption. Then we get the precursor time:

$$\Delta t_1 \approx -800s\,\Phi_{28}^{-2}m_{10}^3a_0^3 \quad (2)$$

Kerr Black Hole released energy can reach 44% (Kip Thorne, 1974),[49] we get another important ratio of precursor fluence to gamma-ray burst fluence $E_{pre}/E_{GRB} \approx a_0^{-2}$. Usually there is a ratio $E_{pre}/E_{GRB} \approx 1-10\%$ (Troja, 2007),[50] corresponding

to the generic Kerr parameter $a_0 \sim$ 3-10 . For the masses of nuclei close to M $\sim$ M_{OV}, a larger Kerr parameter is required. However, in this case, it is necessary to include the contribution of nuclear forces (Lipunova et al., 2009).[6]

After the formation of the spinar, the direction of the magnetic flux of course does not have to coincide with its axis of rotation. We recall that we are considering the case of the conservation of the magnetic flux without an accompanying generation of the type of the dynamo mechanism. Therefore, a spinar can participate in a free precession, especially when it turns into a magnetar. However, a spinar is not a neutron star, whose surface is a hard boundary, beyond which there is a vacuum, or a highly discharged magnetosphere with a Julian-Goldreich density (Lipunov, 1992).[51] Spinar is an idealization of a superdense rotating body surrounded by a gas-dynamic plasma. Its evolution can be described by the equation:

$$dI\omega/dt = K_{\parallel} + K_{\perp} \tag{3}$$

On the right is the moment of forces, the parallel component of which leads to a change in the spinar's rotational moment in magnitude, and the perpendicular component leads to forced precession. It is clear that the spinar is not a rigid body, but the experience of studying the precession of accretion disks under the action of the magnetic moments of the seat shows that even thin disks successfully precess, though in a differential way (Lipunov & Shakura, 1980).[52] The maximum value of both moments of forces in equation 3 is determined as $|K| \approx U_m = \varepsilon_m U_{gr}$. The forced precession frequency turns out to be of the order of $\Omega = |K_{\perp}/I\omega| \approx \varepsilon_m \omega$. As the spinar radius approaches the event horizon $R \to R_g/2$ spinar frequency tends to the $\omega \to c/R_g$ as $R \to Rg$. So far as $U_m \to \varepsilon_m Mc^2$ we get an estimate of the precession period:

$$T = 2\pi/\Omega = (2\pi/\omega)\varepsilon_m^{-1} \approx 5000s\, \Phi_{28}^{-2} m_{10}^3 r^{3/2} \tag{4}$$

Here $r = R/Rg$.

For the precursor GRB160725B in its own frame of reference $\Delta t_{pre} \approx -70s$ (we took z = 1.406 (GCN circular №19600), and the characteristic time of variations is T $\approx$ 10 sec. From 2 and 4 we obtain the estimate $a \approx 10$ and $\Phi_{27}^{-2} m^3$. Accordingly, the magnetic flux turns out to be quite reasonable $\Phi_{27}^{-2} \approx 140 m^{3/2}$.

After the main pulse (G2), the luminosity of the spinar begins to decrease. This is a sure sign that the role of nuclear forces is becoming important. The role of nuclear forces can become important only if the mass of the collapsing nucleus does not greatly exceed the Oppenheimer-Volkov limit. Let us recall that the Oppenheimer-Volkov limit essentially depends on the contribution of centrifugal forces to the equilibrium of the neutron star. So if the fraction of the rotational energy in the virial theorem is 10-20%, the Openhemer Volkov limit can increase by 2-3 times, depending on the equation of state of the neutron star (Lipunov, 1992).[51] With a mass of 4-5 solar masses, the spinar will be supported by nuclear forces until it freezes. This will happen during the after

$$\Delta t_3 \approx I\omega/U_m \approx 6I\omega R/\Phi^2 \approx 10s\, m^2 \Phi_{27}^{-2} r^2$$

So far as $r \to 1/2$ relativistic effects should be taken into account immediately.

Next, we applied a dynamic model of gravimagnetic collapse, which allows us to describe the evolution of a collapsing spinar from the moment of the loss of stability to the collapse into a black hole (Lipunov & Gorbovskoy, 2008; Lipunova et al., 2009).[5,6] In this formulation, a simple non-stationary three-parameter model of collapse is obtained with the decisive role of the rotation and the magnetic field. The input parameters of the theory are the mass, angular momentum and magnetic field of the collapsar. The approximate model includes: centrifugal force, relativistic effects of the Kerr metric, pressure of nuclear matter, dissipation of angular momentum under the influence of a spinar magnetic field, decrease in the dipole magnetic moment as a result of compression and the effects of general relativity (a black hole has no hair), neutrino cooling, time dilation due to gravitational redshift.

In Fig. 7 the results of calculating our approximate model are shown, the parameters of which are given in Fig. 1. Obviously, the main time intervals between the events of the gamma-ray burst G1, G2, G3 are in good agreement with observations. Of course, we do not pretend to describe a detailed light curve within the framework of our approximate model. We can only compare with observations the total energy release of the central engine at times G1, G2 and G3.

Table 1. Core collapse with the mass $M = 3.52M_\odot, M_{OV} = 2.5M_\odot$, effective Kerr parameter $a_0 = 4.9$ and magnetic parameter $a_m = 10^{-4}$.

Stage	Time(obs),s	E(obs)/E90	Time,s	E/E90	R/Rg	B,Gs	a_0
G1	-180	0.02	-75	0.03	37	10^{13}	4.9
G2	0	0.88	0.93	2	3	$2 * 10^{14}$	2.4
G3	$\sim$450	190	0.04	0.04	3	$1.2 * 10^{15}$	1

We associate a slight increase in the gamma flux in G3 with an increase in the power of the central engine at the moment when the spinar loses its stability and collapses into a black hole. Note that such a three-stage collapse occurs only when the mass of the collapsing nucleus exceeds but is comparable with the Oppenheimer – Volkov limit.

4. Discussion

We presented multi-wavelength observations of GRB160625B, one of the brightest gamma-ray bursts in the history of their study. The authors of the article have already published a paper, concerning the first in the history discovery of variable polarization of the intrinsic emission of gamma-ray bursts (Troja et al., 2017).[7] Here we have concentrated on the temporal behavior of itsintrinsic electromagnetic radiation and presented optical and gamma-ray data, recorded during the time of the operation of the central engine of GRB160625B. We made an attempt to find

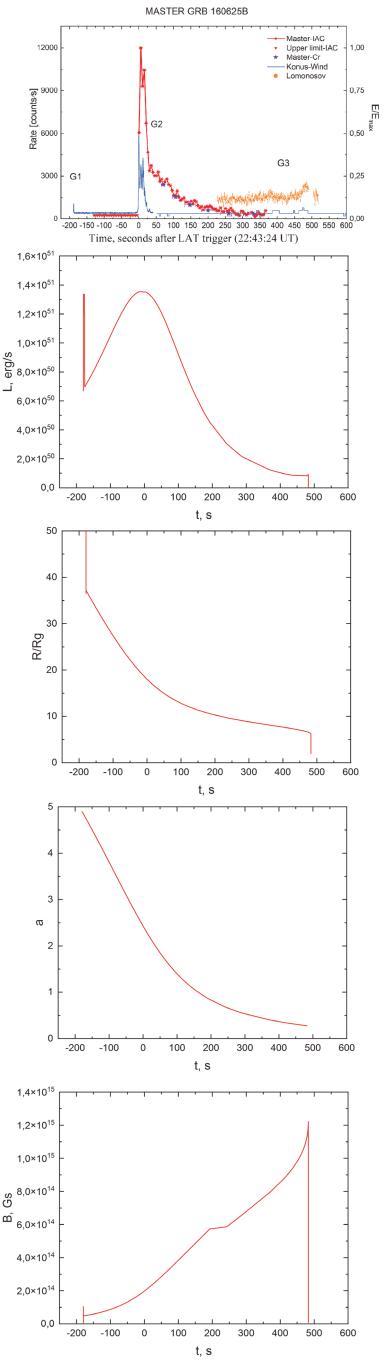


Fig. 7. In the diagram of the GRB160625B event in the observer's frame of reference is shown. In the top diagram, we have shown three principal G1 events - the position of the precursor (Δt = −200 s), the main peak of the G2 pulse (t = 0) and finally the last episode of increased G3 activity (Δt = 400–500 s).

traces of the duality of the GRB system, suspected traces of the dual nature of long GRBs. This study can serve as an example of searching for and finding the dual nature of long GRBs.

Long bright GRB160625B is rare, but not unique. GRB 080319B (Racusin et al., 2008)[53] also had a significant gap between the optical flux and the flux obtained by extrapolating the standard spectrum to the low frequency range. For GRB 080319B Kumar and Narayan (Kumar, Narayan 2009)[54] (see also Lazar et al 2009)[55] proposed a relativistic turbulence model to explain the fine structure of the light curve observed in the gamma range. It is assumed that the matter of the dropped shell is divided into two phases: the turbulent cells and everything in between. In this case, gamma emission is generated in both phases (reverse Compton), and synchrotron radiation of electrons in the inter-cell space is responsible for the radiation flux in the optics.

This explains a number of observed properties of gamma-ray bursts and, what is important in the context of this work, the huge excess of the flux in the optical range over what is dictated by the Band function. However, the natural consequence of this theory is a diminishing of the polarization over time, since the assumed turbulence entangles the magnetic field, leading to depolarization. In fact, the MASTER polarization observations (see Troja et al., 2017)[7] show an increase in polarization. Thus, in the case of GBR 160625B, the relativistic turbulence model does not work.

We interpret the features of the GRB160625B radiation within the framework of the model of a three-stage magnetorotational collapse of the core of a massive star whose mass exceeds but is comparable to the Oppenheimer-Volkov limit. In this model, the first G1 event in Spinar Paradigm will be explained by an abrupt stop of the nucleus collapse at the centrifugal barrier and the formation of a spinar. In this case, half of the accumulated gravitational energy is converted into rotational energy, and the second can be converted into jet energy along the spinar's axis of rotation. Further, as the rotational moment is lost, the spinar is compressing, gradually increasing its rotational energy, mainly emitted along the axis of rotation with a power of the proportional to the Umov-Poynting vector electromagnetic energy flow $E_{G1} = U_m\omega$. Of course, the observer will not see a smooth curve, since the jet must accumulate enough energy to pierce the progenitor shell (Lipunov & Gorbovskoy, 2007).[2] Naturally, in a model based on the conservation laws with an approximate description of the torque dissipation, one should not rely on the exact repetition of the light curve at the moment of the mainpulse G2. Therefore, we only achieved the coincidence of the moment of the maximum energy release with observations and the interval preceding the collapse of a heavy spinar with a powerful magnetic field and a mass exceeding the Oppenheimer-Volkov limit for a non-rotating neutron star.

MASTER and Lomonosov Space Observatory are supported by Lomonosov Moscow State University Development program (equipment). VL, PB are supported by RFBR grant 19-29-11011. NB is supported by FZZE-2020-0017).

References

1. MacFadyen A. I., Woosley S. E., 2001, ApJ 550, 410
2. Lipunov V., Gorbovskoy E., 2007, ApJ, 665L, 97L
3. Tutukov, A. V., Cherepashchuk, A. M., 2016, ARep, 60, 461
4. Lipunova G.V. and Lipunov V.M., 1998, A&A, 329, L29
5. Lipunov V., Gorbovskoy E., 2008, MNRAS, 383, 1397L
6. Lipunova G. et al. 2009, MNRAS,397,1695
7. Troja et al., 2017a, Natur, 547, 425T
8. Burns et al. 2016, GCN Circular, 19581,1
9. Dirirsa et al., 2016, GCN 19580, 1
10. Sadovnichii V., Panasyuk M. et al. 2017, SSRv, 212, 1705S
11. Svinkin et al. 2016, GCN, 19604, 1
12. Nakahira et al. 2016, GCN Circular, 19617,1
13. Melandri et al. 2016, GCN Circular, 19585,1
14. Troja et al. 2017b, GCN Circuar, 19588,1
15. Gorbovskoy et al. 2016a, GCN Circular, 19612,1
16. Lipunov et al. 2016a, GCN Circular, 19683,1
17. Lipunov V. et al. 2010, Advances in Astronomy, 2010, 30L
18. Lipunov et al. 2019, ARep, 63, 293L
19. Kornilov V. et al. 2012, ExpAst, 33, 173
20. Gorbovskoy et al. 2016b, MNRAS, 455, 3312
21. Barthelmy S. D. et al. 1998a, AIP Conference Proceedings, 428, 129
22. Barthelmy S. D. et al. 1998b, AIP Conference Proceedings,428, 99
23. Lipunov V. et al. 2018, Space Sci Rev, 214, 6
24. Lipunov et al. 2016b, ApJ, 833, 198L
25. Sadovnichy, V., Panasyuk M. et al. 2018, ApJ, 861, 48S
26. Ershova et al., 2020, ARep, 64, 126
27. Laskar et al. 2019, ApJ, 884, 121
28. Lipunov et al. 2017d, ApJ, 845, 52L
29. Lipunov et al. 2017a, MNRAS, 465, 3656L
30. Abbott B. P. et al. 2016, ApJL, 826, 13
31. Lipunov et al. 2018, NewA, 63, 48L GW
32. Abbott B. P. et al. 2017a, ApJL, 848, 12
33. Lipunov et al2017c, ApJL, 850, 1
34. Aartsen M. et al. 2017, A&A, 607A, 115I
35. Lipunov et al. 2019, NewA, 72, 42
36. Lipunov V. et al. 2020, ApJ, 896L, 19L
37. Svertilov S.I., Panasyuk M.I.et al. 2018, SSRv, 214, 1
38. Park I.H., M.I., etal.. 2018 Space Sci Rev, 214, 14
39. Mazaeva et al. 2016, GCN Circular, 19605, 1
40. von Kienlin et al.2003, A&A 411, L299-L305
41. Aptekar et al.. 1995,SSRv, 71, 265A
42. Zhang, B. et al. 2018, NatAstron 2, 69
43. Arnaud 1996 ASPC, 101, 17A
44. Band D.,1993, ApJ, 413, 281
45. Preece et al. 1998 ApJ, 506L, 23
46. Woosley, S. E. 1993, ApJ, 405, 273
47. Suvorov A.G., Kokkotas K.D., 2021, MNRAS, 502, 2482-2494
48. Pacini, F. 1967, Natur 216, 567-568

49. Kip S. Thorne 1974, ApJ, 191, 507-519
50. Troja, E., et al. 2007, ApJ, 665, 599
51. Lipunov V. M. 1992, Astrophysics of Neutron Stars, Springer
52. Lipunov V. M., Shakura N. I., SvAL, 6, 14L
53. Racusin, J.et al. 2008, Nature 455, 183
54. Pawan Kumar, Ramesh Narayan, 2009, MNRAS, 395 (1), 472
55. Lazar M., Poedts S. 2009, A&A, 494, 311-315

The role of the magnetic fields in GRB outflows

Jordana-Mitjans, N.[1,*]; Mundell, C. G.[1]; Kobayashi, S.[2]; Smith, R. J.[2]; Guidorzi, C.[3,4,5]; Steele, I. A.[2]; Shrestha, M.[2]; Gomboc, A.[6]; Marongiu, M.[7]; Martone, R.[3]; Lipunov, V.[8]; Gorbovskoy, E. S.[8]; Buckley, D. A. H.[9]; Rebolo, R.[10]; Budnev, N. M.[11]

[1] *Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK*
[2] *Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool, L3 5RF, UK*
[3] *Department of Physics and Earth Science, University of Ferrara, via Saragat 1, I-44122, Ferrara, Italy*
[4] *INFN – Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy*
[5] *INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 101, 40129 Bologna, Italy*
[6] *Center for Astrophysics and Cosmology, University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia*
[7] *INAF – Osservatorio Astronomico di Cagliari - via della Scienza 5 - I-09047 Selargius, Italy*
[8] *Lomonosov Moscow State University, SAI, Physics Department, 13 Univeristetskij pr-t, Moscow 119991, Russia*
[9] *South African Astronomical Observatory PO Box 9, Observatory 7935, Cape Town, South Africa*
[10] *Instituto de Astrofísica de Canarias (IAC), Calle Vía Láctea s/n, E-38200 La Laguna, Tenerife, Spain*
[11] *Applied Physics Institute, Irkutsk State University, 20, Gagarin blvd, 664003, Irkutsk, Russia*
[*] *E-mail: N. Jordana@bath.ac.uk*

Gamma-ray bursts (GRBs) are bright extragalactic flashes of gamma-ray radiation and briefly the most energetic explosions in the Universe. Their catastrophic origin —the merger of compact objects or the collapse of massive stars— drives the formation of a newborn compact remnant (black hole or magnetar) that powers two highly relativistic jets. As these jets continue to travel outwards, they collide with the external material surrounding the dying star, producing a long-lasting afterglow that can be seen across the entire electromagnetic spectrum, from the most energetic gamma-ray emission to radio wavelengths. But how can such material be accelerated and focused into narrow beams? The internal shock model proposes that repeated collisions between material blasted out during the explosion can produce the gamma-ray flash. The competing magnetic model credits primordial large-scale ordered magnetic fields that collimate and accelerate the relativistic outflows. To distinguish between these models and ultimately determine the power source for these energetic explosions, our team studies the polarization of the light during the first minutes after the explosion (using novel instruments on fully autonomous telescopes around the globe) to directly probe the magnetic field properties in these extragalactic jets. This technology allowed the detection of highly polarized optical light in GRB 120308A[1] and confirmed the presence of mildly magnetized jets with large-scale primordial magnetic fields in a reduced sample of GRBs (e.g. GRB 090102,[2] GRB 110205A,[3] GRB 101112A,[3] GRB 160625B[4]). Here we discuss the observations of the most energetic and first GRB detected at very high TeV energies, GRB 190114C,[5] which opens a new frontier in GRB magnetic field studies suggesting that some jets can be launched highly magnetized and that the collapse and destruction of these magnetic fields at very early times may have powered the explosion itself.

Additionally, our most recent polarimetric observations of the jet of GRB 141220A[6] indicate that, when the jetted ejected material is decelerated by the surrounding environment, the magnetic field amplification mechanisms at the front shock —needed to generate the observed synchrotron emission— produce small magnetic domains. These measurements validate theoretical expectations and contrast with previous observations that suggest large magnetic domains in collisionless shocks (i.e. GRB 091208B[7]).

Keywords: High energy astrophysics; Gamma-ray bursts; Magnetic fields; Polarimetry; Shocks; Jets

Gamma-ray bursts (GRBs) are briefly the most intense sources of gamma-ray photons in the Universe to the extent that they outshine any other kind of gamma-ray emitter present in the sky map. The amount of energy released in a matter of milliseconds to hundreds of seconds provides a unique opportunity to study physics in extreme environments (e.g., GRB central engines, jet composition, energy dissipation, acceleration, collimation, shock physics) and test the physical models regarding the nature of their progenitors.[8,9] Additionally, their brightness and cosmological distances make them useful for probing the composition of the early Universe.[10,11]

GRBs with long duration (typically $> 2\,$s; see Ref. 12) are related to the death of massive stars as single collapsars[13,14] or in binary systems,[15] and short GRBs ($< 2\,$s) to mergers from an old population of degenerate compact stars. Possible merger candidates include neutron star binaries[16,17] or binaries containing a stellar black hole and a neutron star companion.[18–20] Due to their inspiral and coalescence, they are emitters of gravitational waves[21] and they are also promising candidate sources of high-energy neutrinos.[22]

GRBs are produced when the central engine (a black hole or magnetar) drives out two bipolar relativistic jets that burrow through the stellar ejecta and emit the characteristic prompt gamma-ray emission via internal dissipation mechanisms[23] —i.e. internal shocks[24] or reconnection.[25] Later on, the relativistic ejecta are decelerated by the circumburst medium by a pair of external shocks:[26,27] a short-lived reverse shock[28,29] and a forward shock (see Fig. 1). This lagging emission called the afterglow radiates via synchrotron emission and can be detected seconds to years after the burst at wavelengths across the electromagnetic spectrum.[30,31] In this context, a variety of afterglow light curves are expected depending on the relative contributions of the reverse and forward shock.

1. Magnetic fields in GRBs

Magnetic fields play an important role in the physics of GRBs by affecting how the jet is launched (formation, acceleration, collimation and composition of the jet) and in shaping the afterglow emission (particle acceleration by the shock). Depending on the driving mechanism of the relativistic outflow —if dominated by kinetic or magnetic energy— GRBs jets can be launched as baryonic (also named hydrodynamical jets[32]) or magnetized jets (i.e., Poynting-flux jets[33,34]). The degree

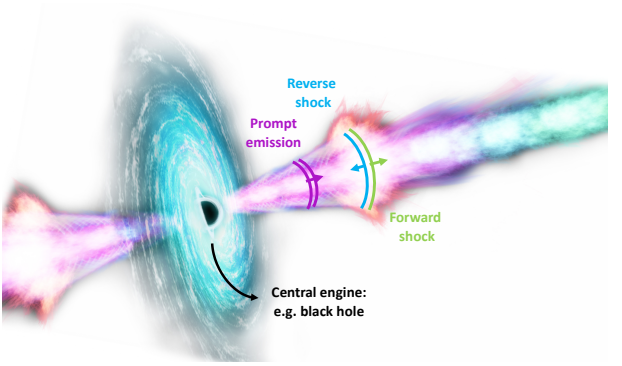


Fig. 1. Schematic representation of a GRB outflow. The steps for the formation of a GRB are: (1) an extragalactic massive star collapses or two compact objects merge, (2) there is accretion onto a newborn black hole, and (3) the ejecta are collimated and accelerated to relativistic speeds along the narrow beam of a jet. Note that the emission we detect (prompt gamma-ray emission and afterglow) comes from sites at different distances from the central engine. Additionally, the afterglow (reverse and forward shock) is formed at the contact discontinuity between the outflow and the circumburst medium.

of magnetization of an outflow is usually characterized by the parameter σ, which is defined as the ratio of magnetic to kinetic energy flux.[35]

In a baryonic jet ($\sigma \ll 1$), neutrino annihilation at the polar region of a hyper-accreting system would launch a thermally-driven fireball that accelerates the material.[13,36] In this model, tangled magnetic fields are locally generated in shocks.[37]

In a magnetized jet ($\sigma \gg 1$), large-scale magnetic fields extract rotational energy from the black hole, accretion disk or magnetar (in a magnetohydrodynamic extraction[33,38,39]), which accelerates the material. The magnetization parameter σ influences the dynamics of the jet and it is thought to decrease with increasing distance from the central engine.[40] In this scenario, the magnetic field configuration is likely toroidal far from the black hole.[41] The detection or non-detection of this large-scale magnetic field remnant is crucial to support or discard the magnetic jet model. Furthermore, a plausible scenario is that GRB outflows are launched with a combination of both hot-baryonic and cold-magnetic components, leading to a broad range of magnetization degrees σ with consequences in the prompt and afterglow emission phase.[34,35,42]

2. The polarization of GRBs

GRBs are extragalactic collimated sources that are not possible to spatially resolve. Polarization offers a tool to study the magnetic field properties and the jet geometry on length scales orders of magnitude smaller than what imaging can probe. Additionally, it can give an extra dimension of information on the jet physics and underlying emission processes and provide complementary information that can

break degeneracies in the temporal and spectral evolution of the emission. Overall, the expected polarization signals in GRBs depend on the magnetic field topology (i.e. ordered or random depending on the jet model), the jet angular geometry,[43,44] and the emission mechanisms. While the emission mechanisms for the prompt emission are still under debate,[45] the afterglow emission is widely assumed to be synchrotron.[46] In the lab frame, synchrotron-emitted photons are linearly polarized,[47] with $P \approx 70\%$. However, the intrinsic polarization from synchrotron emission can be averaged out if the magnetic field topology is random.

During the afterglow stage, the forward shock emission is produced by shocked ambient medium, with tangled magnetic fields locally produced and amplified by e.g., a two-stream magnetic instability (Weibel instability[37,48]). Consequently, the forward shock emission is insensitive to the magnetic field structure of the original ejecta and it is expected to be unpolarized (see Section 2.2). In contrast, the prompt and reverse shock emission is still sensitive to the properties of the central engine ejecta and measuring the polarization allows discriminating between competing jet models (see Section 2.1).

2.1. *The polarization of baryonic and magnetized jets*

Distinct polarization signatures are predicted for magnetized and baryonic jet models. We expect unpolarized emission if the jet is baryonic[37] and polarized emission in a magnetized jet[49] from large-scale ordered magnetic fields advected from the central engine (see Fig. 2). Even if the magnetic fields become slightly distorted during the internal shocks, a mildly magnetized jet is still expected to produce large polarization levels, as only a small area of the jet is observed due to the relativistic beaming. However, note that the maximum synchrotron polarization is reduced to $P \approx 50\%$ due to relativistic aberration effects —i.e., the rotation of the polarization vectors.[50] In the case of a highly magnetized jet (see Fig. 2), reconnection mechanisms are thought to power the prompt emission,[25,51] which would in principle distort the order in the magnetic fields.

2.1.1. *Prompt gamma-ray emission*

Polarization studies of the prompt gamma-ray emission offer great prospects for the study of jet models.[52] However, current studies have yielded a wide range of polarization results,[52–59] i.e. $P = 0\% - 100\%$, with the discrepancy usually related to data with low signal-to-noise, instrumental bias or selection effects in the data analysis.[52,54] Furthermore, time-resolved analyses suggest that the polarization degree in time-integrated studies has been averaged out given that the polarization angle evolves through single pulses (see Refs. 60–62). More importantly, the interpretation of these results is complex given that the underlying physics of the prompt emission remain unknown.[45]

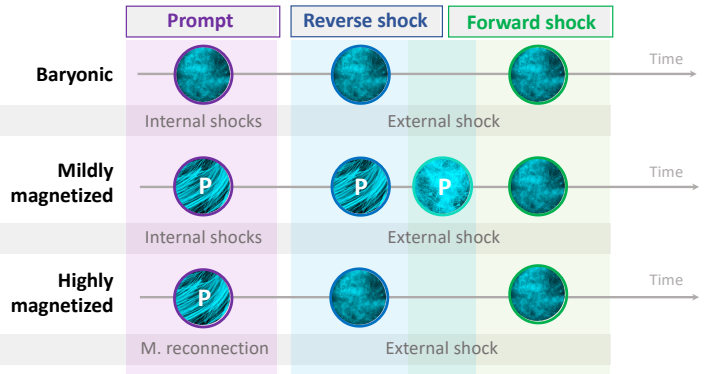


Fig. 2. Summary of the magnetic field structures expected in each jet model (baryonic and magnetized) for the different shocks (internal and external) and assuming synchrotron as the emitting mechanism. The white **P** letter denotes those magnetic field structures that would give rise to significant polarization (large-scale ordered or locally tangled magnetic fields). Note that emission from different origins can be blended into the overall received; e.g. in a mildly magnetized jet, the polarization can be reduced from the maximum expected given a mix of reverse and unpolarized forward shock emission.

2.1.2. *The early optical afterglow*

At optical bands, the polarization measurements from the early afterglow are the domain of robotic telescopes with rapid follow-up capabilities[69–72] (see Fig. 3). The first robust evidence of large-scale ordered magnetic fields in the fireball was the detection of $P = 10\% \pm 1\%$ polarization during the steep decay of GRB 090102 bright reverse shock.[2] This result suggested that primordial magnetic fields do play a role in GRBs outflows and that they can persist at great distances from the central source —up to the external shock radius. The GRB 120308A measurements further proved this idea with a time-resolved polarimetric light curve during a reverse-forward shock interplay[1] (see Fig 4); the polarization decayed steadily from $P = 28\% \pm 4\%$ to $P = 16^{+5}_{-4}\%$ with constant polarization angle due to the increase of unpolarized forward shock photons. Additionally, polarization degrees of $P = 13^{+13}_{-9}\%$ and $P \approx 6\%$ were also detected in GRB 110205A and GRB 101112A.[3]

Overall, current observations of polarized reverse shocks and the modelling of the early GRB afterglow suggest mildly magnetized jets at the deceleration radius,[35,73] with magnetization degrees $\sigma = 0.1 - 1$. These intermediate values of magnetization allow the existence of bright reverse shocks and are also consistent with the measurements of polarized reverse shocks (see Fig. 2). However, we note that some observations suggest higher magnetization at different stages of the jet. The $P < 8\%$ constraint during the fireball deceleration of GRB 060418[74] implies that the reverse shock could have been suppressed due to high magnetization in the fireball at the external shock radius, with $\sigma \geq 1$. Additionally, the optical emission

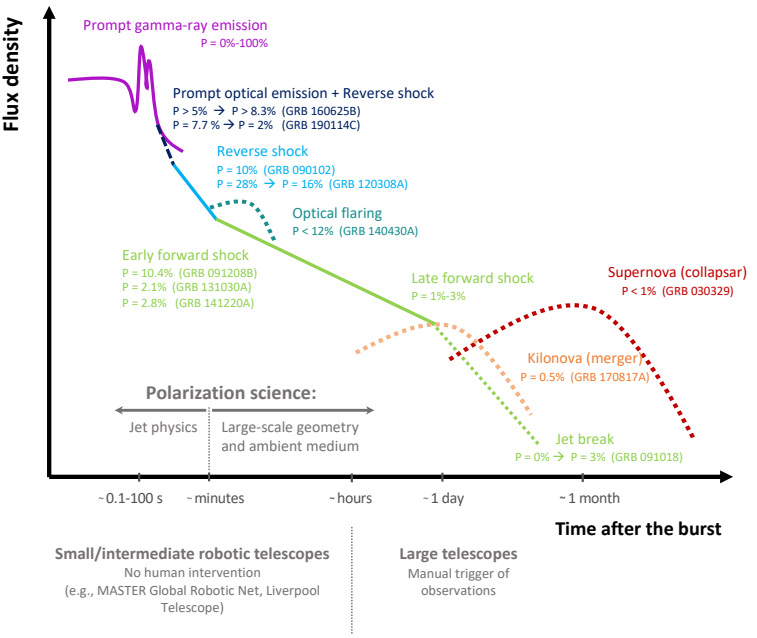


Fig. 3. Schematic representation of the different types of GRB emission and relevant polarization measurements. Note that, with the exception of the prompt gamma-ray emission, the represented "canonical" light curve refers to optical wavelengths. *References:* GRB 030329[63] GRB 090102,[2] GRB 091018,[64] GRB 091208B,[7] GRB 120308A,[1] GRB 131030A,[65] GRB 140430A,[66] GRB 141220A[6] GRB 160625B,[4] GRB 170817A,[67] GRB 190114C,[5] late-time forward shocks[68] and the AstroSat CZT/POLAR gamma-ray polarization catalogues.[52,59]

in between gamma-ray pulses of GRB 160625B[4] was interpreted as reverse shock and presented a significant increase from $P > 5.2\% \pm 0.6\%$ to $P > 8.3\% \pm 0.5\%$ simultaneous to a gamma-ray pulse —implying that polarized prompt emission photons contributed at optical wavelengths. The polarization lower limits were quite low through observations, which led to speculations regarding distorting mechanisms during the prompt gamma-ray emission, such as reconnection.[25,51]

2.1.3. *The highly magnetized jet of GRB 190114C*

Both MASTER II from the Master Global robotic net and RINGO3 instrument from the Liverpool Telescope observed GRB 190114C, the first GRB detected at very high energies (in the TeV domain[76]). Interestingly, the afterglow emission decay rate presented a steep to shallow transition, which suggested reverse shock contribution.[5] Our observations report remarkably low polarization just after the end of the gamma-ray flash ($P = 7.7\%$), a sharp drop of polarization one minute later ($P = 2\%$) and constant levels during the following half an hour. In Ref. 5, we suggest that the $P \approx 2\%$ constant polarization is due to differential dust absorption

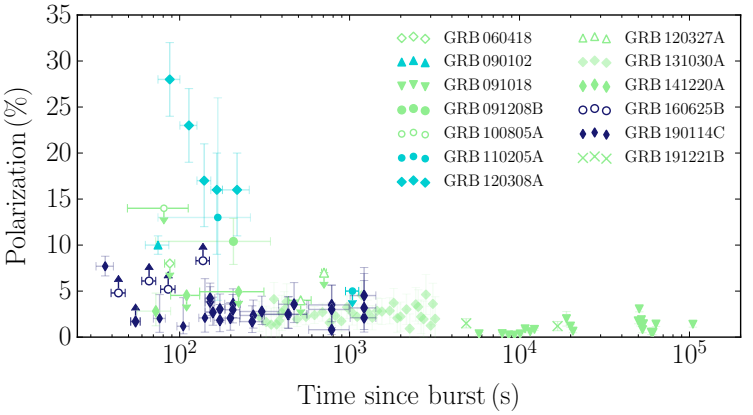


Fig. 4. Rest-frame polarization measurements of GRBs at optical bands —adapted from Ref. 1. Overall, the polarization dependency with time indicates that the regime in which we can infer the jet physics corresponds to observations before ~ 5 min postburst. In contrast, the polarization at later times gives information about the jet geometry and dust. *References:* GRB 060418,[74] GRB 090102,[2] GRB 091018,[64] GRB 091208B,[7] GRB 110205A,[3] GRB 100805A,[3] GRB 120308A,[1] GRB 120327A,[3] GRB 131030A,[65] GRB 141220A,[6] GRB 160625B,[4] GRB 190114C,[5] and GRB 191221B.[75]

in the line of sight —mainly due to the highly obscured environment in which the GRB was formed— which means that the intrinsic polarization of the jet is very small for reverse shock emission. Our low optical polarization measurements are also consistent with $P = 0.87\% \pm 0.13\%$ measured 2.2 h later at millimetre frequencies, during the reverse shock emission.[77]

The temporal and spectral modelling of GRB 190114C emission indicates a clear interplay between the reverse and forward shock and more magnetization in the reverse shock, which suggests the existence of a primordial large-scale magnetic field ejected from the rotating black hole. However, GRB 190114C polarimetric observations reveal that the magnetic field in the ejecta was mostly random oriented in space, which does not agree with what we expect from previous measurements of reverse shocks[1,2] (see Fig. 4).

In Ref. 5, we propose that the polarization was low because the large-scale magnetic field catastrophically collapsed during the first tens of seconds of the gamma-ray flash via magnetic reconnection mechanisms and that the $P = 7.7\%$ measurement is a relic from this emission.[25,51] As opposed to previous polarization measurements,[1,2] these findings suggest that at least some energetic GRBs can be launched highly magnetized and that magnetic dissipation mechanisms can power a bright gamma-ray prompt. Additionally, it opens a debate about the nature of the bright early-time steep emission, which could be a reconnection tail instead of a reverse shock. That is because the reverse shock emission could have been suppressed if, at the deceleration radius, the outflow was still in the highly magnetized regime.[42]

These results pin down timescales and distances for which large-scale magnetic fields survive in astrophysical jets and challenge the current models for the production of GRBs.

2.2. *The polarization from collisionless shocks*

Polarimetric studies of the late afterglow have measured low degrees of polarization at optical wavelengths during the decay of the long-lived forward shock[68] and have provided information on the angular structure of jets[43] and the dust properties of GRB host galaxies[78] (see Fig. 4). Additionally, these polarization levels agree with theoretical predictions and suggest that forward shocks are intrinsically unpolarized due to tangled magnetic fields locally generated in shocks.[37]

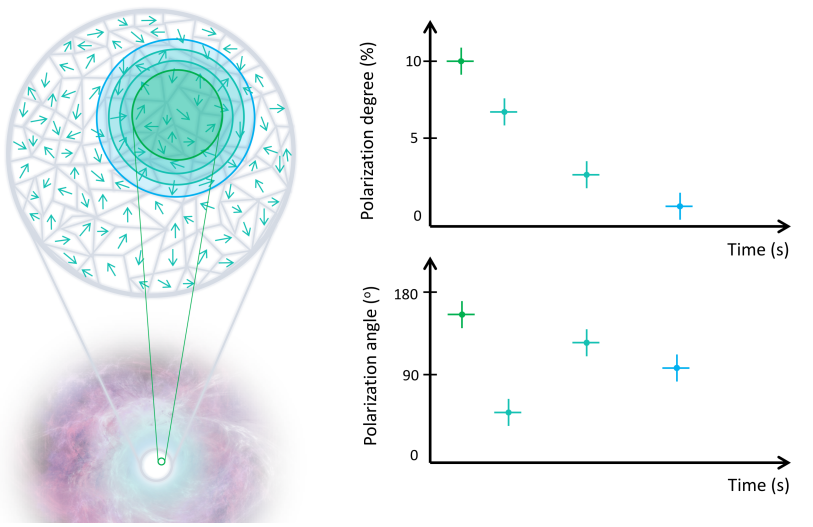


Fig. 5. Schematic representation of the evolution of the polarization degree and angle in the patches model.[79]

However, Ref. 7 observations set a debate on the intrinsic polarization of forward shocks at early times given the detection of significant polarization in the GRB 091208B forward shock ($P = 10.4\% \pm 2.5\%$), which favoured the patches model.[79] Phenomenologically, the patches model implies that as the outflow decelerates and the visible emitting region increases, more patches with coherent magnetic fields (i.e., highly polarized; $P \approx 50\%$) are visible. As the orientation of the magnetic field is random from one domain to another, the overall observed polarization decreases with time (see Fig. 5). Consequently, GRB 091208B measurement still agrees with the $P \approx 1\% - 3\%$ polarization found in late-time forward shocks. Physically, the patches model proposes that the amplification mechanism of the magnetic field at the shock front —needed to increase the magnetic energy density from interstellar

to the observed levels— leads to the formation of large magnetic domains (i.e. large coherent length scales for the magnetic field).

The GRB 091208B measurement gives rise to new possible mechanisms for magnetic field amplification, as well as adding a problematic degeneracy in early-time polarimetric studies. Note that the GRB 091208B polarization level is comparable to that measured in the reverse shock of GRB 090102.[2] If forward shocks are polarized at early times, it becomes difficult to disentangle the possible polarized components that probe the jet physics from the contribution of the shocked ambient material. Consequently, a good understanding of the typical polarization levels of early-time forward shocks is crucial.

2.2.1. *The polarization of the early-time classical forward shock of GRB 141220A*

In Ref. 6, we report the early-time observations of GRB 141220A forward shock. Interestingly, this afterglow presented temporal and spectral properties typical of late-time forward shocks but starting as soon as 1.4 min postburst. Furthermore, our polarimetric observations at 2.2-3.4 min postburst are earlier than GRB 091208B measurement (2.5-11.8 min), which allow us to probe smaller emitting regions. If the length scale of the magnetic field was the same as in GRB 091208B, we would expect $P = 20\%$ polarization for GRB 141220A at the time of observations (as predicted by the patches model; see Fig 5). Instead, we measure $P = 2.8^{+2.0}_{-1.6}\,\%$ in GRB 141220A, which we find is compatible with that induced by dust in the GRB environment and agrees with theoretical predictions of unpolarized forward shocks,[37,80] i.e. $P_{\rm th} = 0\% - 2\%$ (but see Refs. 81, 82).

The GRB 141220A and current early-time forward shock polarization measurements suggest that GRB 091208B is an outlier (see Fig. 4). We speculate that GRB 091208B scenario could be similar to the reverse-forward shock interplay in GRB 120308A.[1] As the polarization measurement was integrated over a large time window, we suggest the presence of a more stable underlying polarized component and ordered large-scale magnetic fields in the fireball.

3. Outlook

After 15 years of progress in early-time polarization studies of GRBs afterglows, we can confidently confirm the presence of large-scale ordered magnetic fields in a subset of bright afterglows, suggesting that the magnetic field ejected from the central source does play a role in the jet launching. In general, we have not measured the maximum polarization allowed by theory in reverse shocks ($P = 50\%$) and polarization observations indicate that highly polarized emission at $P = 28\%$ level is not that common (GRB 120308A;[1] see Fig. 4). Therefore, mechanisms that can change the magnetic field topology might play an important role at the early stages of GRB jets.

Observations of reverse shocks remain rare,[73] and a way to overcome the lack of reverse shocks in the optical is to look at millimetre and radio wavelengths. The predictions are that, if the forward shock peaks at optical bands, the reverse shock will peak at $1 - 100\,\mathrm{GHz}$ bands and $\sim 0.01 - 0.1\,\mathrm{days}$ postburst with fluxes of $0.01 - 0.1\,\mathrm{mJy}$.[83] However, if the magnetization is high at the external shock radius, the reverse shock would be suppressed.[42] The lack of reverse shocks at optical and radio wavelengths would indirectly imply that jets are still highly magnetized at large distances from the central source. Consequently, the rate and the polarization of reverse shocks at optical[1–5] and lower frequencies[77] can greatly constrain the magnetization of GRB jets and further test the mildly and highly magnetized jet regimes.

Great advancement will be made on constraining GRB jet models with a faster target acquisition and the increase of the polarimeters' sensitivity.[84] This will allow gathering a large polarimetric dataset of measurements at early times, sampling most of the GRB population at lower luminosities. Furthermore, other TeV GRBs have been discovered[85–88] —something we did not expect before the Cherenkov Telescope Array (CTA). With the start of the CTA era, we will be able to track the earliest stages of the magnetic field and its evolution in the most energetic systems. Consequently, we will need new polarization technology to optimize CTA follow-up as well as other multimessenger triggers from gravitational waves and possibly neutrinos.

References

1. C. G. Mundell, D. Kopač, D. M. Arnold, I. A. Steele, A. Gomboc, S. Kobayashi, R. M. Harrison, R. J. Smith, C. Guidorzi, F. J. Virgili, A. Melandri and J. Japelj, Highly polarized light from stable ordered magnetic fields in GRB 120308A, *Nature* **504**, 119 (December 2013).
2. I. A. Steele, C. G. Mundell, R. J. Smith, S. Kobayashi and C. Guidorzi, Ten per cent polarized optical emission from GRB090102, *Nature* **462**, 767 (December 2009).
3. I. A. Steele, D. Kopač, D. M. Arnold, R. J. Smith, S. Kobayashi, H. E. Jermak, C. G. Mundell, A. Gomboc, C. Guidorzi, A. Meland ri and J. Japelj, Polarimetry and Photometry of Gamma-Ray Bursts with RINGO2, *ApJ* **843**, p. 143 (Jul 2017).
4. E. Troja, V. M. Lipunov, C. G. Mundell, N. R. Butler, A. M. Watson, S. Kobayashi, S. B. Cenko, F. E. Marshall, R. Ricci and A. Fruchter, Significant and variable linear polarization during the prompt optical flash of GRB 160625B., *Nature* **547**, 425 (Jul 2017).
5. N. Jordana-Mitjans, C. G. Mundell, S. Kobayashi, R. J. Smith, C. Guidorzi, I. A. Steele, M. Shrestha, A. Gomboc, M. Marongiu, R. Martone, V. Lipunov, E. S. Gorbovskoy, D. A. H. Buckley, R. Rebolo and N. M. Budnev, Lowly Polarized Light from a Highly Magnetized Jet of GRB 190114C, *ApJ* **892**, p. 97 (April 2020).
6. N. Jordana-Mitjans, C. G. Mundell, R. J. Smith, C. Guidorzi, M. Marongiu, S. Kobayashi, A. Gomboc, M. Shrestha and I. A. Steele, Coherence scale of magnetic fields generated in early-time forward shocks of GRBs, *MNRAS* **505**, 2662 (August 2021).
7. T. Uehara, K. Toma, K. S. Kawabata, S. Chiyonobu, Y. Fukazawa, Y. Ikejiri, T. Inoue, R. Itoh, T. Komatsu, H. Miyamoto, T. Mizuno, O. Nagae, H. Nakaya, T. Ohsugi,

K. Sakimoto, M. Sasada, H. Tanaka, M. Uemura, M. Yamanaka, T. Yamashita, R. Yamazaki and M. Yoshida, GRB 091208B: First Detection of the Optical Polarization in Early Forward Shock Emission of a Gamma-Ray Burst Afterglow, *ApJ* **752**, p. L6 (June 2012).
8. B. Zhang, Open questions in GRB physics, *Comptes Rendus Physique* **12**, 206 (April 2011).
9. P. Kumar and B. Zhang, The physics of gamma-ray bursts & relativistic jets, *Phys. Rep.* **561**, 1 (February 2015).
10. R. Salvaterra, M. Della Valle, S. Campana, G. Chincarini, S. Covino, P. D'Avanzo, A. Fernández-Soto, C. Guidorzi, F. Mannucci, R. Margutti, C. C. Thöne, L. A. Antonelli, S. D. Barthelmy, M. de Pasquale, V. D'Elia, F. Fiore, D. Fugazza, L. K. Hunt, E. Maiorano, S. Marinoni, F. E. Marshall, E. Molinari, J. Nousek, E. Pian, J. L. Racusin, L. Stella, L. Amati, G. Andreuzzi, G. Cusumano, E. E. Fenimore, P. Ferrero, P. Giommi, D. Guetta, S. T. Holland, K. Hurley, G. L. Israel, J. Mao, C. B. Markwardt, N. Masetti, C. Pagani, E. Palazzi, D. M. Palmer, S. Piranomonte, G. Tagliaferri and V. Testa, GRB090423 at a redshift of z˜8.1, *Nature* **461**, 1258 (October 2009).
11. A. Cucchiara, A. J. Levan, D. B. Fox, N. R. Tanvir, T. N. Ukwatta, E. Berger, T. Krühler, A. Küpcü Yoldaș, X. F. Wu, K. Toma, J. Greiner, F. E. Olivares, A. Rowlinson, L. Amati, T. Sakamoto, K. Roth, A. Stephens, A. Fritz, J. P. U. Fynbo, J. Hjorth, D. Malesani, P. Jakobsson, K. Wiersema, P. T. O'Brien, A. M. Soderberg, R. J. Foley, A. S. Fruchter, J. Rhoads, R. E. Rutledge, B. P. Schmidt, M. A. Dopita, P. Podsiadlowski, R. Willingale, C. Wolf, S. R. Kulkarni and P. D'Avanzo, A Photometric Redshift of z ˜9.4 for GRB 090429B, *ApJ* **736**, p. 7 (July 2011).
12. C. Kouveliotou, C. A. Meegan, G. J. Fishman, N. P. Bhat, M. S. Briggs, T. M. Koshut, W. S. Paciesas and G. N. Pendleton, Identification of two classes of gamma-ray bursts, *ApJ* **413**, L101 (August 1993).
13. S. E. Woosley, Gamma-ray bursts from stellar mass accretion disks around black holes, *ApJ* **405**, 273 (March 1993).
14. A. I. MacFadyen and S. E. Woosley, Collapsars: Gamma-Ray Bursts and Explosions in "Failed Supernovae", *ApJ* **524**, 262 (October 1999).
15. C. L. Fryer, P. A. Mazzali, J. Prochaska, E. Cappellaro, A. Panaitescu, E. Berger, M. van Putten, E. P. J. van den Heuvel, P. Young, A. Hungerford, G. Rockefeller, S.-C. Yoon, P. Podsiadlowski, K. Nomoto, R. Chevalier, B. Schmidt and S. Kulkarni, Constraints on Type Ib/c Supernovae and Gamma-Ray Burst Progenitors, *PASP* **119**, 1211 (November 2007).
16. B. Paczynski, Gamma-ray bursters at cosmological distances, *ApJ* **308**, L43 (September 1986).
17. D. Eichler, M. Livio, T. Piran and D. N. Schramm, Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars, *Nature* **340**, 126 (July 1989).
18. B. Paczynski, Cosmological gamma-ray bursts., *Acta Astronomica* **41**, 257 (January 1991).
19. R. Narayan, B. Paczynski and T. Piran, Gamma-Ray Bursts as the Death Throes of Massive Binary Stars, *ApJ* **395**, p. L83 (August 1992).
20. S. E. Woosley and J. S. Bloom, The Supernova Gamma-Ray Burst Connection, *ARA&A* **44**, 507 (September 2006).
21. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Physical Review Letters* **119**, p. 161101 (October 2017).

22. S. S. Kimura, K. Murase, P. Mészáros and K. Kiuchi, High-energy Neutrino Emission from Short Gamma-Ray Bursts: Prospects for Coincident Detection with Gravitational Waves, *ApJ* **848**, p. L4 (October 2017).
23. R. Sari and T. Piran, Variability in Gamma-Ray Bursts: A Clue, *ApJ* **485**, 270 (August 1997).
24. M. J. Rees and P. Meszaros, Unsteady Outflow Models for Cosmological Gamma-Ray Bursts, *ApJ* **430**, p. L93 (August 1994).
25. B. Zhang and H. Yan, The Internal-collision-induced Magnetic Reconnection and Turbulence (ICMART) Model of Gamma-ray Bursts, *ApJ* **726**, p. 90 (Jan 2011).
26. R. Sari, Hydrodynamics of Gamma-Ray Burst Afterglow, *ApJ* **489**, L37 (November 1997).
27. T. Piran, Gamma-ray bursts and the fireball model, *Phys. Rep.* **314**, 575 (June 1999).
28. R. Sari and T. Piran, Predictions for the Very Early Afterglow and the Optical Flash, *ApJ* **520**, 641 (Aug 1999).
29. S. Kobayashi, Light Curves of Gamma-Ray Burst Optical Flashes, *ApJ* **545**, 807 (Dec 2000).
30. E. Costa, F. Frontera, J. Heise, M. Feroci, J. in't Zand, F. Fiore, M. N. Cinti, D. Dal Fiume, L. Nicastro, M. Orlandini, E. Palazzi, M. Rapisarda#, G. Zavattini, R. Jager, A. Parmar, A. Owens, S. Molendi, G. Cusumano, M. C. Maccarone, S. Giarrusso, A. Coletta, L. A. Antonelli, P. Giommi, J. M. Muller, L. Piro and R. C. Butler, Discovery of an X-ray afterglow associated with the γ-ray burst of 28 February 1997, *Nature* **387**, 783 (June 1997).
31. J. van Paradijs, P. J. Groot, T. Galama, C. Kouveliotou, R. G. Strom, J. Telting, R. G. M. Rutten, G. J. Fishman, C. A. Meegan, M. Pettini, N. Tanvir, J. Bloom, H. Pedersen, H. U. Nørdgaard-Nielsen, M. Linden-Vørnle, J. Melnick, G. Van der Steene, M. Bremer, R. Naber, J. Heise, J. in't Zand, E. Costa, M. Feroci, L. Piro, F. Frontera, G. Zavattini, L. Nicastro, E. Palazzi, K. Bennett, L. Hanlon and A. Parmar, Transient optical emission from the error box of the γ-ray burst of 28 February 1997, *Nature* **386**, 686 (April 1997).
32. A. Shemi and T. Piran, The Appearance of Cosmic Fireballs, *ApJ* **365**, p. L55 (December 1990).
33. V. V. Usov, Millisecond pulsars with extremely strong magnetic fields as a cosmological source of gamma-ray bursts, *Nature* **357**, 472 (June 1992).
34. S. S. Komissarov, N. Vlahakis, A. Königl and M. V. Barkov, Magnetic acceleration of ultrarelativistic jets in gamma-ray burst sources, *MNRAS* **394**, 1182 (April 2009).
35. B. Zhang and S. Kobayashi, Gamma-Ray Burst Early Afterglows: Reverse Shock Emission from an Arbitrarily Magnetized Ejecta, *ApJ* **628**, 315 (July 2005).
36. M. A. Aloy, H. T. Janka and E. Müller, Relativistic outflows from remnants of compact object mergers and their viability for short gamma-ray bursts, *A&A* **436**, 273 (June 2005).
37. M. V. Medvedev and A. Loeb, Generation of Magnetic Fields in the Relativistic Shock of Gamma-Ray Burst Sources, *ApJ* **526**, 697 (Dec 1999).
38. R. D. Blandford and R. L. Znajek, Electromagnetic extraction of energy from Kerr black holes., *MNRAS* **179**, 433 (May 1977).
39. R. D. Blandford and D. G. Payne, Hydromagnetic flows from accretion disks and the production of radio jets., *MNRAS* **199**, 883 (June 1982).
40. Zhang, Gamma-Ray Burst Prompt Emission, *International Journal of Modern Physics D* **23**, p. 1430002 (December 2014).
41. H. C. Spruit, F. Daigne and G. Drenkhahn, Large scale magnetic fields and their dissipation in GRB fireballs, *A&A* **369**, 694 (Apr 2001).

42. D. Giannios, P. Mimica and M. A. Aloy, On the existence of a reverse shock in magnetized gamma-ray burst ejecta, *A&A* **478**, 747 (Feb 2008).
43. E. M. Rossi, D. Lazzati, J. D. Salmonson and G. Ghisellini, The polarization of afterglow emission reveals γ-ray bursts jet structure, *MNRAS* **354**, 86 (Oct 2004).
44. R. Gill, J. Granot and P. Kumar, Linear polarization in gamma-ray burst prompt emission, *MNRAS* **491**, 3343 (January 2020).
45. K. Toma, T. Sakamoto, B. Zhang, J. E. Hill, M. L. McConnell, P. F. Bloser, R. Yamazaki, K. Ioka and T. Nakamura, Statistical Properties of Gamma-Ray Burst Polarization, *ApJ* **698**, 1042 (June 2009).
46. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *ApJ* **497**, L17 (April 1998).
47. G. B. Rybicki and A. P. Lightman, *Radiative processes in astrophysics* 1979.
48. E. S. Weibel, Spontaneously Growing Transverse Waves in a Plasma Due to an Anisotropic Velocity Distribution, *Phys. Rev. Lett.* **2**, 83 (Feb 1959).
49. J. Granot and A. Königl, Linear Polarization in Gamma-Ray Bursts: The Case for an Ordered Magnetic Field, *ApJ* **594**, L83 (September 2003).
50. M. Lyutikov, V. I. Pariev and R. D. Blandford, Polarization of Prompt Gamma-Ray Burst Emission: Evidence for Electromagnetically Dominated Outflow, *ApJ* **597**, 998 (Nov 2003).
51. W. Deng, H. Zhang, B. Zhang and H. Li, Collision-induced Magnetic Reconnection and a Unified Interpretation of Polarization Properties of GRBs and Blazars, *ApJ* **821**, p. L12 (Apr 2016).
52. M. Kole, N. De Angelis, F. Berlato, J. M. Burgess, N. Gauvin, J. Greiner, W. Hajdas, H. C. Li, Z. H. Li, A. Pollo, N. Produit, D. Rybka, L. M. Song, J. C. Sun, J. Szabelski, T. Tymieniecka, Y. H. Wang, B. B. Wu, X. Wu, S. L. Xiong, S. N. Zhang and Y. J. Zhang, The POLAR gamma-ray burst polarization catalog, *A&A* **644**, p. A124 (December 2020).
53. W. Coburn and S. E. Boggs, Polarization of the prompt γ-ray emission from the γ-ray burst of 6 December 2002, *Nature* **423**, 415 (May 2003).
54. R. E. Rutledge and D. B. Fox, Re-analysis of polarization in the γ-ray flux of GRB 021206, *MNRAS* **350**, 1288 (June 2004).
55. C. Wigger, W. Hajdas, K. Arzner, M. Güdel and A. Zehnder, Gamma-Ray Burst Polarization: Limits from RHESSI Measurements, *ApJ* **613**, 1088 (October 2004).
56. D. Götz, P. Laurent, F. Lebrun, F. Daigne and Ž. Bošnjak, Variable Polarization Measured in the Prompt Emission of GRB 041219A Using IBIS on Board INTEGRAL, *ApJ* **695**, L208 (April 2009).
57. D. Götz, S. Covino, A. Fernández-Soto, P. Laurent and Ž. Bošnjak, The polarized gamma-ray burst GRB 061122, *MNRAS* **431**, 3550 (June 2013).
58. D. Yonetoku, T. Murakami, S. Gunji, T. Mihara, K. Toma, Y. Morihara, T. Takahashi, Y. Wakashima, H. Yonemochi, T. Sakashita, N. Toukairin, H. Fujimoto and Y. Kodama, Magnetic Structures in Gamma-Ray Burst Jets Probed by Gamma-Ray Polarization, *ApJ* **758**, p. L1 (October 2012).
59. T. Chattopadhyay, S. V. Vadawale, E. Aarthy, N. P. S. Mithun, V. Chand, A. Ratheesh, R. Basak, A. R. Rao, V. Bhalerao, S. Mate, B. Arvind, V. Sharma and D. Bhattacharya, Prompt Emission Polarimetry of Gamma-Ray Bursts with the AstroSat CZT Imager, *ApJ* **884**, p. 123 (October 2019).
60. J. M. Burgess, M. Kole, F. Berlato, J. Greiner, G. Vianello, N. Produit, Z. H. Li and J. C. Sun, Time-resolved GRB polarization with POLAR and GBM. Simultaneous spectral and polarization analysis with synchrotron emission, *A&A* **627**, p. A105 (July 2019).

61. V. Sharma, S. Iyyani, D. Bhattacharya, T. Chattopadhyay, A. R. Rao, E. Aarthy, S. V. Vadawale, N. P. S. Mithun, V. B. Bhalerao, F. Ryde and A. Pe'er, Time-varying Polarized Gamma-Rays from GRB 160821A: Evidence for Ordered Magnetic Fields, *ApJ* **882**, p. L10 (September 2019).
62. S.-N. Zhang, M. Kole, T.-W. Bao, T. Batsch, T. Bernasconi, F. Cadoux, J.-Y. Chai, Z.-G. Dai, Y.-W. Dong, N. Gauvin, W. Hajdas, M.-X. Lan, H.-C. Li, L. Li, Z.-H. Li, J.-T. Liu, X. Liu, R. Marcinkowski, N. Produit, S. Orsi, M. Pohl, D. Rybka, H.-L. Shi, L.-M. Song, J.-C. Sun, J. Szabelski, T. Tymieniecka, R.-J. Wang, Y.-H. Wang, X. Wen, B.-B. Wu, X. Wu, X.-F. Wu, H.-L. Xiao, S.-L. Xiong, L.-Y. Zhang, L. Zhang, X.-F. Zhang, Y.-J. Zhang and A. Zwolinska, Detailed polarization measurements of the prompt emission of five gamma-ray bursts, *Nature Astronomy* **3**, 258 (January 2019).
63. K. S. Kawabata, J. Deng, L. Wang, P. Mazzali, K. Nomoto, K. Maeda, N. Tominaga, H. Umeda, M. Iye, G. Kosugi, Y. Ohyama, T. Sasaki, P. Höflich, J. C. Wheeler, D. J. Jeffery, K. Aoki, N. Kashikawa, T. Takata, N. Kawai, T. Sakamoto, Y. Urata, A. Yoshida, T. Tamagawa, K. Torii, W. Aoki, N. Kobayashi, Y. Komiyama, Y. Mizumoto, J. Noumaru, R. Ogasawara, K. Sekiguchi, Y. Shirasaki, T. Totani, J. Watanabe and T. Yamada, On the Spectrum and Spectropolarimetry of Type Ic Hypernova SN 2003dh/GRB 030329, *ApJ* **593**, L19 (August 2003).
64. K. Wiersema, P. A. Curran, T. Krühler, A. Meland ri, E. Rol, R. L. C. Starling, N. R. Tanvir, A. J. van der Horst, S. Covino, J. P. U. Fynbo, P. Goldoni, J. Gorosabel, J. Hjorth, S. Klose, C. G. Mundell, P. T. O'Brien, E. Palazzi, R. A. M. J. Wijers, V. D'Elia, P. A. Evans, R. Filgas, A. Gomboc, J. Greiner, C. Guidorzi, L. Kaper, S. Kobayashi, C. Kouveliotou, A. J. Levan, A. Rossi, A. Rowlinson, I. A. Steele, A. de Ugarte Postigo and S. D. Vergani, Detailed optical and near-infrared polarimetry, spectroscopy and broad-band photometry of the afterglow of GRB 091018: polarization evolution, *MNRAS* **426**, 2 (Oct 2012).
65. O. G. King, D. Blinov, D. Giannios, I. Papadakis, E. Angelakis, M. Balokovic, L. Fuhrmann, T. Hovatta, P. Khodade, S. Kiehlmann, N. Kylafis, A. Kus, I. Myserlis, D. Modi, G. Panopoulou, I. Papamastorakis, V. Pavlidou, B. Pazderska, E. Pazderski, T. J. Pearson, C. Rajarshi, A. N. Ramaprakash, A. C. S. Readhead, P. Reig, K. Tassis and J. A. Zensus, Early-time polarized optical light curve of GRB 131030A., *MNRAS* **445**, L114 (November 2014).
66. D. Kopač, C. G. Mundell, J. Japelj, D. M. Arnold, I. A. Steele, C. Guidorzi, S. Dichiara, S. Kobayashi, A. Gomboc, R. M. Harrison, G. P. Lamb, A. Melandri, R. J. Smith, F. J. Virgili, A. J. Castro-Tirado, J. Gorosabel, A. Järvinen, R. Sánchez-Ramírez, S. R. Oates and M. Jelínek, Limits on Optical Polarization during the Prompt Phase of GRB 140430A, *ApJ* **813**, p. 1 (November 2015).
67. S. Covino, K. Wiersema, Y. Z. Fan, K. Toma, A. B. Higgins, A. Melandri, P. D'Avanzo, C. G. Mundell, E. Palazzi, N. R. Tanvir, M. G. Bernardini, M. Branchesi, E. Brocato, S. Campana, S. di Serego Alighieri, D. Götz, J. P. U. Fynbo, W. Gao, A. Gomboc, B. Gompertz, J. Greiner, J. Hjorth, Z. P. Jin, L. Kaper, S. Klose, S. Kobayashi, D. Kopac, C. Kouveliotou, A. J. Levan, J. Mao, D. Malesani, E. Pian, A. Rossi, R. Salvaterra, R. L. C. Starling, I. Steele, G. Tagliaferri, E. Troja, A. J. van der Horst and R. A. M. J. Wijers, The unpolarized macronova associated with the gravitational wave event GW 170817, *Nature Astronomy* **1**, 791 (October 2017).
68. S. Covino, G. Ghisellini, D. Lazzati and D. Malesani, *Polarization of Gamma-Ray Burst Optical and Near-Infrared Afterglows*, in *Gamma-Ray Bursts in the Afterglow Era*, eds. M. Feroci, F. Frontera, N. Masetti and L. Piro, Astronomical Society of the Pacific Conference Series, Vol. 312 2004, p. 169.

69. V. M. Lipunov, A. V. Krylov, V. G. Kornilov, G. V. Borisov, D. A. Kuvshinov, A. A. Belinsky, M. V. Kuznetsov, S. A. Potanin, G. A. Antipov, N. V. Tyurina, E. S. Gorbovskoy and I. Chilingaryan, MASTER: The Mobile Astronomical System of Telescope-Robots, *Astronomische Nachrichten* **325**, 580 (October 2004).
70. I. A. Steele, The Liverpool telescope, *Astronomische Nachrichten* **325**, 519 (October 2004).
71. K. S. Kawabata, O. Nagae, S. Chiyonobu, H. Tanaka, H. Nakaya, M. Suzuki, Y. Kamata, S. Miyazaki, K. Hiragi, H. Miyamoto, M. Yamanaka, A. Arai, T. Yamashita, M. Uemura, T. Ohsugi, M. Isogai, Y. Ishitobi and S. Sato, Wide-field one-shot optical polarimeter: HOWPol, **7014**, p. 70144L (July 2008).
72. A. N. Ramaprakash, C. V. Rajarshi, H. K. Das, P. Khodade, D. Modi, G. Panopoulou, S. Maharana, D. Blinov, E. Angelakis, C. Casadio, L. Fuhrmann, T. Hovatta, S. Kiehlmann, O. G. King, N. Kylafis, A. Kougentakis, A. Kus, A. Mahabal, A. Marecki, I. Myserlis, G. Paterakis, E. Paleologou, I. Liodakis, I. Papadakis, I. Papamastorakis, V. Pavlidou, E. Pazderski, T. J. Pearson, A. C. S. Readhead, P. Reig, A. Słowikowska, K. Tassis and J. A. Zensus, RoboPol: a four-channel optical imaging polarimeter, *MNRAS* **485**, 2355 (May 2019).
73. J. Japelj, D. Kopač, S. Kobayashi, R. Harrison, C. Guidorzi, F. J. Virgili, C. G. Mundell, A. Melandri and A. Gomboc, Phenomenology of Reverse-shock Emission in the Optical Afterglows of Gamma-Ray Bursts, *ApJ* **785**, p. 84 (Apr 2014).
74. C. G. Mundell, I. A. Steele, R. J. Smith, S. Kobayashi, A. Melandri, C. Guidorzi, A. Gomboc, C. J. Mottram, D. Clarke, A. Monfardini, D. Carter and D. Bersier, Early Optical Polarization of a Gamma-Ray Burst Afterglow, *Science* **315**, p. 1822 (March 2007).
75. D. A. H. Buckley, S. Bagnulo, R. J. Britto, J. Mao, D. A. Kann, J. Cooper, V. Lipunov, D. M. Hewitt, S. Razzaque, N. P. M. Kuin, I. M. Monageng, S. Covino, P. Jakobsson, A. J. van der Horst, K. Wiersema, M. Böttcher, S. Campana, V. D'Elia, E. S. Gorbovskoy, I. Gorbunov, D. N. Groenewald, D. H. Hartmann, V. G. Kornilov, C. G. Mundell, R. Podesta, J. K. Thomas, N. Tyurina, D. Vlasenko, B. van Soelen and D. Xu, Spectropolarimetry and photometry of the early afterglow of the gamma-ray burst GRB 191221B, *MNRAS* **506**, 4621 (September 2021).
76. MAGIC Collaboration, V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González, W. Bednarek, L. Bellizzi, E. Bernardini, A. Berti, J. Besenrieder, W. Bhattacharyya, C. Bigongiari, A. Biland, O. Blanch, G. Bonnoli, Ž. Bošnjak, G. Busetto, A. Carosi, R. Carosi, G. Ceribella, Y. Chai, A. Chilingaryan, S. Cikota, S. M. Colak, U. Colin, E. Colombo, J. L. Contreras, J. Cortina, S. Covino, G. D'Amico, V. D'Elia, P. da Vela, F. Dazzi, A. de Angelis, B. de Lotto, M. Delfino, J. Delgado, D. Depaoli, F. di Pierro, L. di Venere, E. Do Souto Espiñeira, D. Dominis Prester, A. Donini, D. Dorner, M. Doro, D. Elsaesser, V. Fallah Ramazani, A. Fattorini, A. Fernández-Barral, G. Ferrara, D. Fidalgo, L. Foffano, M. V. Fonseca, L. Font, C. Fruck, S. Fukami, S. Gallozzi, R. J. García López, M. Garczarczyk, S. Gasparyan, M. Gaug, N. Giglietto, F. Giordano, N. Godinović, D. Green, D. Guberman, D. Hadasch, A. Hahn, J. Herrera, J. Hoang, D. Hrupec, M. Hütten, T. Inada, S. Inoue, K. Ishio, Y. Iwamura, L. Jouvin, D. Kerszberg, H. Kubo, J. Kushida, A. Lamastra, D. Lelas, F. Leone, E. Lindfors, S. Lombardi, F. Longo, M. López, R. López-Coto, A. López-Oramas, S. Loporchio, B. Machado de Oliveira Fraga, C. Maggio, P. Majumdar, M. Makariev, M. Mallamaci, G. Maneva, M. Manganaro, K. Mannheim, L. Maraschi, M. Mariotti, M. Martínez, S. Masuda, D. Mazin, S. Mićanović, D. Miceli, M. Minev, J. M. Miranda, R. Mirzoyan, E. Molina, A. Moralejo, D. Morcuende, V. Moreno, E. Moretti, P. Munar-Adrover,

V. Neustroev, C. Nigro, K. Nilsson, D. Ninci, K. Nishijima, K. Noda, L. Nogués, M. Nöthe, S. Nozaki, S. Paiano, J. Palacio, M. Palatiello, D. Paneque, R. Paoletti, J. M. Paredes, P. Peñil, M. Peresano, M. Persic, P. G. Prada Moroni, E. Prandini, I. Puljak, W. Rhode, M. Ribó, J. Rico, C. Righi, A. Rugliancich, L. Saha, N. Sahakyan, T. Saito, S. Sakurai, K. Satalecka, K. Schmidt, T. Schweizer, J. Sitarek, I. Šnidarić, D. Sobczynska, A. Somero, A. Stamerra, D. Strom, M. Strzys, Y. Suda, T. Surić, M. Takahashi, F. Tavecchio, P. Temnikov, T. Terzić, M. Teshima, N. Torres-Albà, L. Tosti, S. Tsujimoto, V. Vagelli, J. van Scherpenberg, G. Vanzo, M. Vazquez Acosta, C. F. Vigorito, V. Vitale, I. Vovk, M. Will, D. Zarić and L. Nava, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).

77. T. Laskar, K. D. Alexander, R. Gill, J. Granot, E. Berger, C. G. Mundell, R. Barniol Duran, J. Bolmer, P. Duffell, H. van Eerten, W.-f. Fong, S. Kobayashi, R. Margutti and P. Schady, ALMA Detection of a Linearly Polarized Reverse Shock in GRB 190114C, *ApJ* **878**, p. L26 (Jun 2019).
78. S. Klose, E. Palazzi, N. Masetti, B. Stecklum, J. Greiner, D. H. Hartmann and H. M. Schmid, Prospects for multiwavelength polarization observations of GRB afterglows and the case GRB 030329, *A&A* **420**, 899 (Jun 2004).
79. A. Gruzinov and E. Waxman, Gamma-Ray Burst Afterglow: Polarization and Analytic Light Curves, *ApJ* **511**, 852 (February 1999).
80. T. Inoue, K. Asano and K. Ioka, Three-dimensional Simulations of Magnetohydrodynamic Turbulence Behind Relativistic Shock Waves and Their Implications for Gamma-Ray Bursts, *ApJ* **734**, p. 77 (June 2011).
81. J. Mao and J. Wang, Application of Jitter Radiation: Gamma-Ray Burst Prompt Polarization, *ApJ* **776**, p. 17 (October 2013).
82. J. Mao and J. Wang, Linear Polarization, Circular Polarization, and Depolarization of Gamma-ray Bursts: A Simple Case of Jitter Radiation, *ApJ* **838**, p. 78 (April 2017).
83. D. Kopač, C. G. Mundell, S. Kobayashi, F. J. Virgili, R. Harrison, J. Japelj, C. Guidorzi, A. Melandri and A. Gomboc, Radio Flares from Gamma-ray Bursts, *ApJ* **806**, p. 179 (June 2015).
84. C. M. Gutiérrez, D. Arnold, D. Copley, C. M. Copperwheat, E. Harvey, H. Jermak, J. Knapen, A. McGrath, A. Oria, R. Rebolo, I. A. Steele and M. Torres, The new 4-m robotic telescope, *Astronomische Nachrichten* **340**, 40 (January 2019).
85. H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E. O. Angüner, M. Arakawa, C. Arcaro, C. Armand, H. Ashkar, M. Backes, V. Barbosa Martins, M. Barnard, Y. Becherini, D. Berge, K. Bernlöhr, E. Bissaldi, R. Blackwell, M. Böttcher, C. Boisson, J. Bolmont, S. Bonnefoy, J. Bregeon, M. Breuhaus, F. Brun, P. Brun, M. Bryan, M. Büchele, T. Bulik, T. Bylund, M. Capasso, S. Caroff, A. Carosi, S. Casanova, M. Cerruti, T. Chand, S. Chandra, A. Chen, S. Colafrancesco, M. Curyło, I. D. Davids, C. Deil, J. Devin, P. deWilt, L. Dirson, A. Djannati-Ataï, A. Dmytriiev, A. Donath, V. Doroshenko, J. Dyks, K. Egberts, G. Emery, J. P. Ernenwein, S. Eschbach, K. Feijen, S. Fegan, A. Fiasson, G. Fontaine, S. Funk, M. Füßling, S. Gabici, Y. A. Gallant, F. Gaté, G. Giavitto, L. Giunti, D. Glawion, J. F. Glicenstein, D. Gottschall, M. H. Grondin, J. Hahn, M. Haupt, G. Heinzelmann, G. Henri, G. Hermann, J. A. Hinton, W. Hofmann, C. Hoischen, T. L. Holch, M. Holler, D. Horns, D. Huber, H. Iwasaki, M. Jamrozy, D. Jankowsky, F. Jankowsky, A. Jardin-Blicq, I. Jung-Richardt, M. A. Kastendieck, K. Katarzyński, M. Katsuragawa, U. Katz, D. Khangulyan, B. Khélifi, J. King, S. Klepser, W. Kluźniak, N. Komin, K. Kosack, D. Kostunin, M. Kreter, G. Lamanna, A. Lemière, M. Lemoine-Goumard, J. P. Lenain, E. Leser, C. Levy, T. Lohse, I. Lypova, J. Mackey, J. Majumdar,

D. Malyshev, V. Marandon, A. Marcowith, A. Mares, C. Mariaud, G. Martí-Devesa, R. Marx, G. Maurin, P. J. Meintjes, A. M. W. Mitchell, R. Moderski, M. Mohamed, L. Mohrmann, C. Moore, E. Moulin, J. Muller, T. Murach, S. Nakashima, M. de Naurois, H. Ndiyavala, F. Niederwanger, J. Niemiec, L. Oakes, P. O'Brien, H. Odaka, S. Ohm, E. de Ona Wilhelmi, M. Ostrowski, I. Oya, M. Panter, R. D. Parsons, C. Perennes, P. O. Petrucci, B. Peyaud, Q. Piel, S. Pita, V. Poireau, A. Priyana Noel, D. A. Prokhorov, H. Prokoph, G. Pühlhofer, M. Punch, A. Quirrenbach, S. Raab, R. Rauth, A. Reimer, O. Reimer, Q. Remy, M. Renaud, F. Rieger, L. Rinchiuso, C. Romoli, G. Rowell, B. Rudak, E. Ruiz-Velasco, V. Sahakian, S. Sailer, S. Saito, D. A. Sanchez, A. Santangelo, M. Sasaki, R. Schlickeiser, F. Schüssler, A. Schulz, H. M. Schutte, U. Schwanke, S. Schwemmer, M. Seglar-Arroyo, M. Senniappan, A. S. Seyffert, N. Shafi, K. Shiningayamwe, R. Simoni, A. Sinha, H. Sol, A. Specovius, M. Spir-Jacob, Ł. Stawarz, R. Steenkamp, C. Stegmann, C. Steppa, T. Takahashi, T. Tavernier, A. M. Taylor, R. Terrier, D. Tiziani, M. Tluczykont, C. Trichard, M. Tsirou, N. Tsuji, R. Tuffs, Y. Uchiyama, D. J. van der Walt, C. van Eldik, C. van Rensburg, B. van Soelen, G. Vasileiadis, J. Veh, C. Venter, P. Vincent, J. Vink, H. J. Völk, T. Vuillaume, Z. Wadiasingh, S. J. Wagner, R. White, A. Wierzcholska, R. Yang, H. Yoneda, M. Zacharias, R. Zanin, A. A. Zdziarski, A. Zech, A. Ziegler, J. Zorn, N. Żywucka, F. de Palma, M. Axelsson and O. J. Roberts, A very-high-energy component deep in the γ-ray burst afterglow, *Nature* **575**, 464 (November 2019).

86. O. Blanch, F. Longo, A. Berti, S. Fukami, Y. Suda, S. Loporchio, S. Micanovic, J. G. Green, V. Pinter, M. Takahashi and MAGIC Collaboration, GRB 201216C: MAGIC detection in very high energy gamma rays, *GRB Coordinates Network* **29075**, p. 1 (December 2020).

87. V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, K. Asano, D. Baack, A. Babić, A. Baquero, U. Barres de Almeida, J. A. Barrio, J. Becerra González, W. Bednarek, L. Bellizzi, E. Bernardini, M. Bernardos, A. Berti, J. Besenrieder, W. Bhattacharyya, C. Bigongiari, A. Biland, O. Blanch, G. Bonnoli, Ž. Bošnjak, G. Busetto, R. Carosi, G. Ceribella, M. Cerruti, Y. Chai, A. Chilingarian, S. Cikota, S. M. Colak, E. Colombo, J. L. Contreras, J. Cortina, S. Covino, G. D'Amico, V. D'Elia, P. da Vela, F. Dazzi, A. de Angelis, B. de Lotto, M. Delfino, J. Delgado, C. Delgado Mendez, D. Depaoli, F. di Pierro, L. di Venere, E. D. Souto Espiñeira, D. Dominis Prester, A. Donini, D. Dorner, M. Doro, D. Elsaesser, V. Fallah Ramazani, A. Fattorini, G. Ferrara, L. Foffano, M. V. Fonseca, L. Font, C. Fruck, S. Fukami, R. J. García López, M. Garczarczyk, S. Gasparyan, M. Gaug, N. Giglietto, F. Giordano, P. Gliwny, N. Godinović, J. G. Green, D. Green, D. Hadasch, A. Hahn, L. Heckmann, J. Herrera, J. Hoang, D. Hrupec, M. Hütten, T. Inada, S. Inoue, K. Ishio, Y. Iwamura, J. Jormanainen, L. Jouvin, Y. Kajiwara, M. Karjalainen, D. Kerszberg, Y. Kobayashi, H. Kubo, J. Kushida, A. Lamastra, D. Lelas, F. Leone, E. Lindfors, S. Lombardi, F. Longo, R. López-Coto, M. López-Moya, A. López-Oramas, S. Loporchio, B. Machado de Oliveira Fraga, C. Maggio, P. Majumdar, M. Makariev, M. Mallamaci, G. Maneva, M. Manganaro, K. Mannheim, L. Maraschi, M. Mariotti, M. Martínez, D. Mazin, S. Mender, S. Mićanović, D. Miceli, T. Miener, M. Minev, J. M. Miranda, R. Mirzoyan, E. Molina, A. Moralejo, D. Morcuende, V. Moreno, E. Moretti, V. Neustroev, C. Nigro, K. Nilsson, D. Ninci, K. Nishijima, K. Noda, S. Nozaki, Y. Ohtani, T. Oka, J. Otero-Santos, S. Paiano, M. Palatiello, D. Paneque, R. Paoletti, J. M. Paredes, L. Pavletić, P. Peñil, C. Perennes, M. Persic, P. G. Prada Moroni, E. Prandini, C. Priyadarshi, I. Puljak, W. Rhode, M. Ribó, J. Rico, C. Righi, A. Rugliancich, L. Saha, N. Sahakyan, T. Saito, S. Sakurai, K. Satalecka, F. G. Saturni, B. Schleicher, K. Schmidt, T. Schweizer, J. Sitarek, I. Šnidarić, D. Sobczynska,

A. Spolon, A. Stamerra, D. Strom, M. Strzys, Y. Suda, T. Surić, M. Takahashi, F. Tavecchio, P. Temnikov, T. Terzić, M. Teshima, N. Torres-Albà, L. Tosti, S. Truzzi, A. Tutone, J. van Scherpenberg, G. Vanzo, M. Vazquez Acosta, S. Ventura, V. Verguilov, C. F. Vigorito, V. Vitale, I. Vovk, M. Will, D. Zarić, MAGIC Collaboration and L. Nava, MAGIC Observations of the Nearby Short Gamma-Ray Burst GRB 160821B, *ApJ* **908**, p. 90 (February 2021).

88. H. E. S. S. Collaboration, H. Abdalla, F. Aharonian, F. Ait Benkhali, E. O. Angüner, C. Arcaro, C. Armand, T. Armstrong, H. Ashkar, M. Backes, V. Baghmanyan, V. Barbosa Martins, A. Barnacka, M. Barnard, Y. Becherini, D. Berge, K. Bernlöhr, B. Bi, E. Bissaldi, M. Böttcher, C. Boisson, J. Bolmont, M. de Bony de Lavergne, M. Breuhaus, F. Brun, P. Brun, M. Bryan, M. Büchele, T. Bulik, T. Bylund, S. Caroff, A. Carosi, S. Casanova, T. Chand, S. Chandra, A. Chen, G. Cotter, M. Curyło, J. Damascene Mbarubucyeye, I. D. Davids, J. Davies, C. Deil, J. Devin, L. Dirson, A. Djannati-Ataï, A. Dmytriiev, A. Donath, V. Doroshenko, L. Dreyer, C. Duffy, J. Dyks, K. Egberts, F. Eichhorn, S. Einecke, G. Emery, J. P. Ernenwein, K. Feijen, S. Fegan, A. Fiasson, G. Fichet de Clairfontaine, G. Fontaine, S. Funk, M. Füßling, S. Gabici, Y. A. Gallant, G. Giavitto, L. Giunti, D. Glawion, J. F. Glicenstein, M. H. Grondin, J. Hahn, M. Haupt, G. Hermann, J. A. Hinton, W. Hofmann, C. Hoischen, T. L. Holch, M. Holler, M. Hörbe, D. Horns, D. Huber, M. Jamrozy, D. Jankowsky, F. Jankowsky, A. Jardin-Blicq, V. Joshi, I. Jung-Richardt, E. Kasai, M. A. Kastendieck, K. Katarzyński, U. Katz, D. Khangulyan, B. Khélifi, S. Klepser, W. Kluźniak, N. Komin, R. Konno, K. Kosack, D. Kostunin, M. Kreter, G. Lamanna, A. Lemière, M. Lemoine-Goumard, J. P. Lenain, F. Leuschner, C. Levy, T. Lohse, I. Lypova, J. Mackey, J. Majumdar, D. Malyshev, D. Malyshev, V. Marandon, P. Marchegiani, A. Marcowith, A. Mares, G. Martí-Devesa, R. Marx, G. Maurin, P. J. Meintjes, M. Meyer, A. Mitchell, R. Moderski, L. Mohrmann, A. Montanari, C. Moore, P. Morris, E. Moulin, J. Muller, T. Murach, K. Nakashima, A. Nayerhoda, M. de Naurois, H. Ndiyavala, J. Niemiec, L. Oakes, P. O'Brien, H. Odaka, S. Ohm, L. Olivera-Nieto, E. de Ona Wilhelmi, M. Ostrowski, S. Panny, M. Panter, R. D. Parsons, G. Peron, B. Peyaud, Q. Piel, S. Pita, V. Poireau, A. Priyana Noel, D. A. Prokhorov, H. Prokoph, G. Pühlhofer, M. Punch, A. Quirrenbach, S. Raab, R. Rauth, P. Reichherzer, A. Reimer, O. Reimer, Q. Remy, M. Renaud, F. Rieger, L. Rinchiuso, C. Romoli, G. Rowell, B. Rudak, E. Ruiz-Velasco, V. Sahakian, S. Sailer, H. Salzmann, D. A. Sanchez, A. Santangelo, M. Sasaki, M. Scalici, J. Schäfer, F. Schüssler, H. M. Schutte, U. Schwanke, M. Seglar-Arroyo, M. Senniappan, A. S. Seyffert, N. Shafi, J. N. S. Shapopi, K. Shiningayamwe, R. Simoni, A. Sinha, H. Sol, A. Specovius, S. Spencer, M. Spir-Jacob, Ł. Stawarz, L. Sun, R. Steenkamp, C. Stegmann, S. Steinmassl, C. Steppa, T. Takahashi, T. Tam, T. Tavernier, A. M. Taylor, R. Terrier, J. H. E. Thiersen, D. Tiziani, M. Tluczykont, L. Tomankova, M. Tsirou, R. Tuffs, Y. Uchiyama, D. J. van der Walt, C. van Eldik, C. van Rensburg, B. van Soelen, G. Vasileiadis, J. Veh, C. Venter, P. Vincent, J. Vink, H. J. Völk, Z. Wadiasingh, S. J. Wagner, J. Watson, F. Werner, R. White, A. Wierzcholska, Y. W. Wong, A. Yusafzai, M. Zacharias, R. Zanin, D. Zargaryan, A. A. Zdziarski, A. Zech, S. J. Zhu, J. Zorn, S. Zouari, N. Żywucka, P. Evans and K. Page, Revealing x-ray and gamma ray temporal and spectral similarities in the GRB 190829A afterglow, *Science* **372**, 1081 (June 2021).

MASTER optical observations of the blazar TXS0506+056 during the IC170922A

V. M. Lipunov[1,2,*], K. Zhirkov[1,2], V. G. Kornilov[1,2], E. Gorbovskoy[2], N. M. Budnev[3], D. A. H. Buckley[4], R. Rebolo[5], M. Serra[5] , R. Podesta[6], C. Francile[6], V. Topolev[1,2], N. Tiurina[2], O. A. Gress[2,3], P. Balanutsa[2], A. Chasovnikov[1,2], D. Vlasenko[1,2], E. Minkina[1,2], A. Gabovich[7], G. Antipov[2], A. Kuznetsov[2], V. Vladimirov[2], S. Svertilov[8], F. Podesta[6], V. Grinshpun[1,2], A. Tlatov[9], D. Cheryasov[2] and V. Senik[2]

[1] *Lomonosov Moscow State University, Physics Department, Vorobievy Hills, 1, Moscow 119991, Russia*

[2] *Lomonosov MSU, SAI, Universitetsky, 13, Moscow 119234, Russia*

[3] *Irkutsk State University, 20, Gagarin, Irkutsk 664003, Russia*

[4] *South African Astrophysical Observatory, PO Box 9, 7935 Observatory, Cape Town, South Africa*

[5] *Instituto de Astrofísica de Canarias, Lactea, E38205, LaLaguna, Tenerife, Spain*

[6] *Observatorio Astronomico Felix Aguilar(OAFA), San Juan National University, Avda Benavides s/n, Rivadavia, El Leonsito, Argentina;*

[7] *Blagoveschensk State Pedagogical University, Lenin, 104, Amur, Blagoveschensk 675000, Russia*

[8] *Lomonosov Moscow State University, Skobeltsyn Institute of Nuclear Physics (SINP MSU), Vorobievy Hills, 1, Moscow, 119991, Russia*

[9] *Kislovodsk Solar Station Pulkovo Observatory, Gagarina 100, Kislovodsk 357700, Russia*

**E-mail: lipunov@sai.msu.ru*

We present MASTER Global Robotic Net (Lipunov et al. 2010) earliest optical alert observations of IceCube-170922A error box. We discovered fast variability of blazar TXS 0506+056 27 sec after notice time (73s after the trigger time) at 2017-09-22 20:55:43 UT by MASTER-Tavrida robotic telescope. MASTER found the blazar TXS 0506+056 to be in the off-state after one minute and then switched to the on-state no later than two hours after the event. The effect is observed at a 50-sigma significance level. We also analysed own unique 16-years light curve of blazar TXS 0506+056 (518 data set).

Keywords: Gamma-ray burst

1. Introduction

High energy cosmic neutrinos sources are still mysterious. MASTER Global Robotic Net (Lipunov et al. 2010)[19] starts Ice Cube alert and follow up observations a few

years ago. On 22 September 2017 the robotic telescope of the MASTER global network automatically imaged the error box of the high-energy neutrino event IceCube-170922A (Kopper et al. 2017). Observations were carried out 27 seconds after receiving the alert, i.e., 73 seconds after the IceCube-170922A neutrino event was detected by the IceCube observatory at the South Pole (Lipunov et al. 2018a).[23] Observations started at 84 degree zenith distance (MASTER has fully opened roof as alert telescope).

We calibrated these images using the Gaia (Brown et al. 2018) catalog as the source of reference stars, and found the BL Lacertae type blazar TXS 0506+056 (IceCube et al. 2017) to be in the off-state after one minute and then switched to the on-state no later than two hours after the event. The effect is observed at *delta* m =0.790+-0.016 (a 50-*sigma* significance level).[27]

IceCube registered events with error regions, whose sizes are currently comparable to one square degree. Therefore finding a blazar within the error box of a VHE neutrino event cannot be considered sufficient to prove that blazars are actually progenitors of these particles. Detecting some non-standard event from the supposed source at a time close to the neutrino event is required. For example, a blazar emitting gamma and cosmic rays and showing a sharp flux variation near the neutrino detection time would provide compelling evidence of the association of the neutrino event with a known astrophysical object. The first candidate object for an astrophysical neutrino event was the blazar TXS 0506+056 (IceCube et al. 2017)[14,15] found inside the error box of the IceCube-170922A neutrino event. This blazar turned out to be located at a distance of $\tilde{3}$.7 billion light years (its redshift is z = 0.3365+/- 0.0010) (Paiano et al. 2018).[29]

TXS 0506+056 blazar was registered in the gamma-ray active state several months before the neutrino event. Detection of high-energy particles (175Gev) began one week after, and the optical, x-ray, and gamma-ray emission was observed with low temporal resolution and showed no appreciable variations near the detection time . Therefore although when combining the available data suggested that TXS 0506+056 was a very promising high energy neutrino source optical candidate, the temporal resolution of multi-messenger data did not provide conclusive evidence at the time and the object remained just a likely, but still debatable, candidate. In this letter we report conclusive detection of light variation of the blazar TXS 0506+056 just several minutes after the neutrino event, which ended no later than after two hours. For comparison, nearest ASAS-SN, Kiso/KWFC and Kanata/HONIR optical observations do not show the same decrease in optical brightness (because they started 18 hours after MASTER observations when effect disappeared. Although the blazar was in the gamma-ray active state, this state started several months before the neutrino event. Detection of high-energy particles (175Gev) began one week after, and the optical, x-ray, and gamma-ray emission was observed with low temporal resolution and showed no appreciable variations near the detection time. Therefore although when combining the available data suggested that TXS 0506+056 was a

very promising high energy neutrino source optical candidate, the temporal resolution of multi-messenger data did not provide conclusive evidence at the time and the object remained just a likely, but still debatable, candidate . In this letter we report conclusive detection of light variation of the blazar TXS 0506+056 just several minutes after the neutrino event, which ended no later than after two hours. For comparison, nearest ASAS-SN, Kiso/KWFC and Kanata/HONIR optical observations do not show the same decrease in optical brightness (because they started 18 hours after MASTER observations when effect disappeared.

2. MASTER alert optical observations of IceCube-170922A error box

MASTER Global Robotic Net, as the leader of early gamma-ray burst observations (Troja et al. 2017; Ershova et al. 2020, Lipunov et al. 2016, Lipunov et al. 2017, Gorbvoskoy et al. 2016)[10, 11, 18, 20, 21, 32] has an almost 20-year long experience with real-time rapid pointings to GRB alerts within the first minute of the alert . In 2015 MASTER started actively participation in the program of fast optical support of major physical and astrophysical experiments, such as detection of very high energy neutrinos (ANTARES (Dornic et al. 2015;[9] Gress et al. 2019[12]), IceCube (Aartsen et al. 2017)[1] , Baksan (Lipunov et al. 2019a),[25] gravitational waves (LIGO/VIRGO collaboration (Abbott et al. 2016),[2] and Fast Radio Bursts (FRB (Lipunov et al. 2018b)[24]). The favorable arrangement of MASTER sites makes it possible to inspect all gravitational-wave error boxes. MASTER made the crucial contribution to the optical support of the first gravitational-wave event GW 150914 by inspecting the most part of the error box (Abbott et al. 2016, Lipunov et al. 2017).[3, 21] On 17 August 2017 MASTER, together with 5 other telescopes, performed the first ever optical localization of a gravitational-wave source independently discovered Kilonova from GW 170817 (Abbot et al. 2017, Lipunov et al. 2017).[3, 22]

At the 22nd of September, 2017, MASTER received an alert from IceCube event and MASTER-Tavrida telescope acquired the first three images, starting from 2017-09-22 20:55:43UT 27 sec after notice time (i.e.73s after the trigger time). The field of view of the MASTER telescope has a size of four square degrees (Lipunov et al. 2010)[19] and fully covers the final field of view of IceCube (Lipunov et al. 2020[27]).

MASTER limiting magnitude was 19.0m at 180-second frames, despite the large zenith distance (84 degrees). Hence the TXS 0506+056 blazar at the time of the alert was a 15.12 +- 0.01 magnitude object in all three frames acquired over $\tilde{1}5$ minutes (the light curve is shown in Fig. 2). There was the faintest blazar brightness over the full period of this alert. After 2 hours, at 2017-09-22 23:11:36 UT, the flux from the blazar increased in brightness by a factor of two and reached 14.33m+-0.01m. We emphasize that here in the text we give the averaged values for the triples of frames in the first minutes and after two hours (compare[27]). Hence our observations show at an extremely high confidence level of 50 sigma that within several minutes of the neutrino event, the blazar was in an anomalously extinguished state. This conclusion

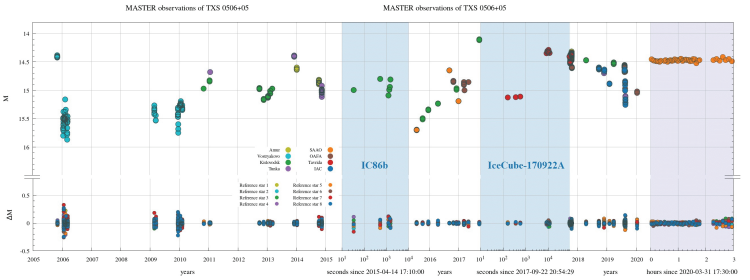


Fig. 1. MASTER light curve of the TXS 0506+056 blazar for the 16 years. Archive light curve of the TXS 0506+056 blazar based on observations made by MASTER Global Net robotic telescope of from 2005 until now (red point). Below we see the photometry of 8 reference stars, to the pink and blue panels represent three very narrow episodes in time. The first of these is April 2015 when IceCube IC86b saw a 3.5 sigma excess of the neutrino flux over the background (IceCube 2017) . The second is the 22 September 2017 event (IceCube 2017). Logarithmic time is shown in seconds from the neutrino trigger. It is easy to easily see the rapid change in the luminosity of the blazar in $\widetilde{2}$ times. Finally, the third episode is a uniform blazar monitoring timeline in the first quarter of 2020. With these new observations, the total number of observations submitted reached 518 (See the photometry table)[27]

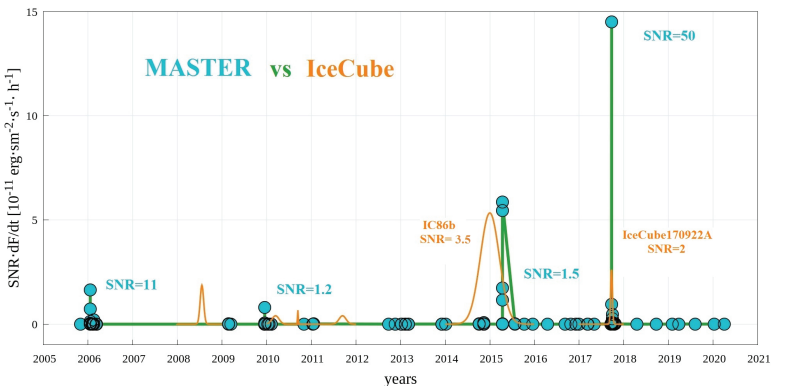


Fig. 2. MASTER optical variability rate history for blazar TXS0506+056. Flux deriviation the multiplied by the signal-to-noise ratio (blue). The orange curves schematically show the Gaussian analysis of the archive data of IceCube.[27]

was fully confirmed during the two days after the alert, when the MASTER-SAAO robotic telescope joined the blazar observation campaign (Lipunov et al. 2018)[23, 24]

To undestand, was it a unique event for the TXS 0506+056 blazar or no, we analysed 518 2x2 square-degrees images in MASTER Global Robotic Net database, starting from 2005, when we had only one MASTER I telescope located near Moscow (Lipunov et al. 2010,[19] MASTERVostryakovo).[27] All MASTER telescopes has identical equipment, that gives us possibility to make photometry in one system (Lipunov et al. 2010, 2019b).[19, 26]

Fig. 1 shows our photometry of the blazar over the last 16 years. We chose 8 Gaia catalog stars, having the brightness and color similar to those of the blazar, as photometric reference stars, and estimated the errors of individual photometric measurements from the scatter of the magnitudes of these reference stars. We also checked these stars for rapid and long-term variability and found them to be quite stable. We found three times when the brightness of the blazar varied $\tilde{0}$.5m more than 1-3 significance level. The first such time was in 2006, when IceCube neutrino observatory was not yet operating. The second time was in April 2015 (substantial increase of neutrino signal IC86b). The MASTER observation date is in April but statistically close to Gaussian IceCube half year window 9/2014 to 3/2015. And the third time was in September 2017, when the IceCube-170922 (IC86C) event occurred.[27]

The event of September 22, 2017 has outstanding characteristics in terms of flux derivation and signalto- noise ratio. Recently, we conducted detailed blazar monitoring over several nights at the right end of Fig. 2. As we see in the usual state, the blazar is stable at times of several hours and even days with an accuracy of $\tilde{0}$.02 mag. This means that the instability we discovered on 22 September 2017, a few minutes after the neutrino alert, has a reliability of 50 sigma by this criterion(Fig. 2). This result shows the power of fast alert observations of sources of ultrahigh energy particles.

The event of September 22, 2017 has outstanding characteristics in terms of flux derivation and signal-to- noise ratio. Recently, we conducted detailed blazar monitoring over several nights. In the usual state, the blazar is stable at times of several hours and even days with an accuracy of $\tilde{0}$.02 mag. This means that the instability we discovered on 22 September 2017, a few minutes after the neutrino alert, has a reliability of about 40 sigma, and by this criterion. This result shows the power of fast alert observations of sources of ultrahigh energy particles.

3. Discussion

We find for the adopted set of Hubble cosmological parameters (Planck Collaboration et al. 2016)[4] H0 = 67.8 km/s/Mpc (the Hubble constant), $Omega_m = 0.308$, $\Omega_\Lambda = 0.692$ (the matter and vacuum density) that several minutes after the neutrino event the optical isotropic luminosity of the blazar was L_{opt} $4.3{*}10^{45}$ erg/s and after two hours it returned to the typical level within several weeks of the neutrino event, $9.7 * 10^{45}$ erg/s (we include galactic absorbtion AB=0.4 (Schlegel et al. 1998)[31]). The generally accepted picture (Schlegel et al. 1998)[31] is that blazar radiation arises from relativistic jet directed toward us. The boosted jet gamma factor is moderate Γ 10. In the shock wave at the front of the jet there is an acceleration of protons to ultrahigh energies, which in turn collide with target photons and generate pion production. The decay of pions, in turn, gives rise to a muon neutrino which registers an IceCube detection and high gamma photons detected by the Fermi gamma-ray observatory.

During the period within $\tilde{2}$ weeks around the neutrino event detection time the 0.1 – 100 Gev gamma-ray luminosity was $1.3 * 10^{47}$ erg/s (IceCube et al 2017). Note that the neutrino luminosity of the quasar was equal to about $4 * 10^{47}$ erg/s, which is appreciably higher and, evidently, closer to the gamma-ray luminosity. However, this is not surprising because neutrinos and gamma-ray emission have the same source of energy — that of high-energy protons accelerated by the central supermassive black hole.

The event that we discovered, namely the decrease of the brightness of the TXS 0506+056 blazar near the neutrino detection time, provides complementary and very compelling evidence for the link between the blazar and the IceCube-170922 neutrino event. We analyzed archival data (MASTER unique 518 photometry data for 16 years), which we found to be consistent with this fact. We also propose a hypothesis explaining the anti correlation of the optical and neutrino flux. An increase in neutrino flux means that up to half of the protons disappear. If we assume that these protons produce synchrotron optical radiation, then any increase in neutrino luminosity will lead to a decrease in the optical brightness of the blazar.

MASTER and Lomonosov Space Observatory are supported by Lomonosov Moscow State University Development program (equipment). VL, PB are supported by RFBR grant 19-29-11011. NB is supported by FZZE-2020-0017).

References

1. Aartsen M. et al., 2017, A&A 607, A115
2. Abbott B.P. et al. 2016, ApJL, 826, 13A
3. Abbot B. P. et al. 2017, ApJL, 848 (2), L12
4. Planck Collaboration et al. 2016, A&A, 594, A13
5. Ageron M.et al. 2011, NIMPh, 656, 11
6. Balkanov V. et al. 2002, NuPhS, 110, 504B
7. Boliev M.M. et al. 2018, PPN , 49 (4), 585
8. Brown A.G.A. et al. 2018, A&A, 616, A1
9. Dornic D. et al. 2015, ATel, 7987, 1
10. Ershova O. et al. 2020, ARep, 64 (2), 126
11. Gorbovskoy et al. 2016, MNRAS, 455, 3312
12. Gress O. et al. 2019, RMxAA, 51, 89
13. Hayakava S., Yamamoto Y. 1963, PThPh, 30, 71
14. IceCube Collaboration et al. 2017, Sci, 361, eaat1378
15. IceCube Collaboration. 2017, Sci, 361, 147
16. Kopper C.et al. 2017, GCN Circ, 21916, 1
17. Kornilov V. et al. 2012, ExpAst, 33, 173
18. Laskar et al. 2019, ApJ, 884, 121
19. Lipunov V. et al. 2010, AdAst, 2010, 30L
20. Lipunov V. et al. 2016, MNRAS, 455, 3312
21. Lipunov V. et al. 2017, MNRAS , 465 (3), 3656
22. Lipunov V. et al. 2017, ApJL, 850, 1L
23. Lipunov V.M. et al. 2018a, GCN Circ, 22942, 1
24. Lipunov V. et al. 2018b,ATel, 11902, 1
25. Lipunov V. et al. 2019a, GCN Circ, 26539, 1

26. Lipunov V. et al. 2019b, ARep, 63,293
27. Lipunov V. et al. 2020, ApJ, 896L, 19L
28. ORiordan M., Pe'er A., McKinney J. C. 2017, ApJ, 843, 2, id. 81
29. Paiano S. et al. 2018, ApJ, 854, L32
30. Paliya V.S.et al. 2020, a-ph: 20034.060112v2 , https://arxiv.org/abs/2003.06012
31. Schlegel D., Finkbeiner D., Davis M. 1998, ApJ, 843, 525
32. Troja et al., 2017a, Natur, 547, 425T

A short review on nonlinear perturbation theory of structure formation for modified gravity

Jorge L. Cervantes-Cota

Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Carretera México Toluca Km. 36.5, Ocoyoacac, C.P. 52750, Edo. Mex., México. E-mail: jorge.cervantes@inin.gob.mx

Alejandro Aviles

Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Apartado Postal 18-1027, Col. Escandón, Ciudad de México,11801, México. and Consejo Nacional de Ciencia y Tecnología, Av. Insurgentes Sur 1582, Colonia Crédito Constructor, Del. Benito Juárez, 03940, Ciudad de México, México E-mail: avilescervantes@gmail.com

In this proceedings contribution we summarize a series of results on nonlinear perturbation theory for modified gravity models, in which a scalar field changes the gravitational dynamics at cosmic scales. We discuss its effects on large scale structure, focusing on screenings and the power spectrum of matter fields and tracers. We also discuss a redshift space distortions model for modified gravity and the departures from the LCDM model.

Keywords: Perturbation theory; modified gravity.

1. Introduction

In recent years we have witnessed important developments on the understanding of large scale structure (LSS) formation. The experimental confirmation of baryon acoustic oscillation (BAO) feature and measurements of redshift space distortions (RSD) have boosted the search for an increasing accuracy in the determination of the cosmological parameters from the LSS side, apart from those of the cosmic microwave background (CMB) measurements, and other probes.

On the other hand, there is a growing interest to test gravity at cosmological scales. One the one hand, because we do not really know the origin of dark energy (being this a constant or a field), so it can be due to some modified gravity (MG). On the other hand, because the level of understanding of the LSS phenomena mentioned in the previous paragraph opens the possibility to test gravity at cosmic scales at high precision. Due to these facts, a natural interest arose to develop perturbation theory (PT) for MG models. At the same time, MG simulations have been carried out, see for example Cautun et al.[1] Both PT theory and simulations allow us to understand the role that gravity plays to form structure.

Nonlinear PT for LSS in MG models have been developed in the last decade or so, beginning with approach proposed by Koyama et al[2] for standard perturbation theory (SPT) to one-loop. Later, we and other collaborators developed the

formalism for Lagrangian perturbation theory (LPT).[3–6] Nonlinear PT is limited to weakly nonlinear scales, but it is nevertheless important because of the BAO and RSD effects are manifested in these scales. Non-linearities in MG show special features such as screening mechanisms due to the fifth force of the scalar degree of freedom present in these models.[7]

Since our goal is to test gravity at cosmic scales, one employs toy MG models, such as f(R) or Dvali-Gabadadze-Porrati (DGP) models to achieve a percent-level of precision in the power spectrum (PS) or two point correlation function, that will allow to describe the physics that new generation of stage IV experiments will be able to reach soon, such as DESI,[8] EUCLID[9] and LSST,[10] among others. Therefore, a whole machinery of tools are being developed, including precision theory, mocks, simulations, emulators, among other technical improvements, see for instance Alam et al.[11]

There are a few pieces to be understood to study LSS in PT, and that we tackle in the present work: description of the dynamics of density fields composed of dark matter particles, and here one has LPT and SPT; then, one has to deal with biased tracers; to understand how the RSD description is modelled in MG; to add the small-scale effective field theory (EFT) corrections coming from UV-cutoffs of PT; and finally, to apply a IR-resummation of modes to cancel out unwanted features in the power spectrum.

Our primer to PT for MG is organized as follows: We first explain, in section 2, how LPT is obtained and the role of growth functions is emphasized, and in section 3 we discuss their use to screen the fifth force. In section 4, we explain how LPT results can be used to obtain SPT ones, establishing a bridge between these two formalisms. In section 5, we present the power spectra for matter fields and in section 6 we introduce the bias parameter that contributes to power spectra for tracers. In section 7 we review recent results in which we modeled RSD for MG models. Finally, in section 8, EFT and IR-resummation explanations are shortly addressed, before we conclude in section 9.

2. LPT for MG

At the large scales we are working, baryons behave essentially as cold dark matter. So, from now on we consider the whole matter that follows trajectories with Eulerian comoving coordinates $\mathbf{x}$. The Lagrangian displacement vector field $\mathbf{\Psi}$ relates the initial (Lagrangian) $\mathbf{q}$ and Eulerian $\mathbf{x}$ positions of particles as $\mathbf{\Psi}(\mathbf{q},t) = \mathbf{x}(\mathbf{q},t) - \mathbf{q}$, chosen such that $\mathbf{x}(\mathbf{q}, t_{ini}) = \mathbf{q}$, with t_{ini} an early time where the evolution of all scales of interest remains linear and the overdensities are quite small, $\delta(\mathbf{x}, t_{ini}) = \delta(\mathbf{q}) \ll 1$. We consider $\mathbf{\Psi}$ a longitudinal field, since it is the only contribution to 1-loop. Using matter conservation,

$$(1 + \delta(\mathbf{x}, t))d^3x = (1 + \delta(\mathbf{q}))d^3q, \tag{1}$$

one gets the known relation between density and Lagrangian fields

$$\delta(\mathbf{x},t) = \frac{1+\delta(\mathbf{q})-J}{J} \simeq \frac{1-J}{J}, \tag{2}$$

where $J_{ij} = \partial x_i/\partial q^j = \delta_{ij} + \Psi_{i,j}$ is the Jacobian matrix of the coordinate transformation and J its determinant. Since $\mathbf{\Psi}$ is a potential field, the Jacobian matrix is symmetric.

The fundamental field in LPT is the displacement field, and its equation of motion is given by the geodesic equation,

$$\hat{\mathcal{T}}\mathbf{\Psi}(\mathbf{q},t) = -\frac{1}{a}\nabla_{\mathbf{x}}\psi(\mathbf{x},t)|_{\mathbf{x}=\mathbf{q}+\mathbf{\Psi}}. \tag{3}$$

where $\hat{\mathcal{T}} \equiv \frac{d^2}{dt^2} + 2H\frac{d}{dt}$. We use $\nabla_{\mathbf{x}} = \partial/\partial\mathbf{x}$ to denote partial differentiation with respect to Eulerian coordinates, and $\nabla = \partial/\partial\mathbf{q}$ for differentiation with respect to Lagrangian coordinates. The two scalar gravitational potentials of the metric are related by $\phi - \psi = \varphi/2$, assuming Newtonian gauge, where φ is the scalar field mediating the fifth force.

The gravity model can be set out in the Einstein frame, where the Ricci scalar in the Lagrangian is not coupled to the scalar field, but the couplings are given by the scalar field and the matter Lagrangian, such as in Symmetron gravity.[12] Another possibility is to lay out the theory in the Jordan frame, where explicit couplings exist between the Ricci and the scalar field, such as Brans-Dicke type of couplings, and in addition, couplings to matter may exist. For the sake of concreteness, we opt for the latter approach here, but we will show below some results for the former, as well. Then, the Poisson equation is

$$\frac{1}{a^2}\nabla^2_{\mathbf{x}}\psi = A_0\delta(\mathbf{x},t) - \frac{1}{2a^2}\nabla^2_{\mathbf{x}}\varphi. \tag{4}$$

clearly, the scalar field modifies the Newtonian potential (ψ), and then the force upon to test particles. The Klein-Gordon equation in the quasi-static limit takes the form[2]

$$(3+2\omega_{\rm BD})\frac{1}{a^2}\nabla^2_{\mathbf{x}}\varphi = -8\pi G\bar{\rho}\delta + \mathrm{NL}, \tag{5}$$

where NL contains all nonlinear terms to be defined next that will encode the mode couplings.

It is useful to consider PT in the Fourier space, in which modes are decoupled for LCDM and become simple for MG models at first order approximation. The full expansion for Klein-Gordon equation becomes

$$\begin{aligned} -\frac{k^2}{2a^2}\varphi(\mathbf{k}) = &-(A(k)-A_0)\tilde{\delta}(\mathbf{k}) + \frac{k^2/a^2}{6\Pi(k)}\delta I(\mathbf{k}) \\ &- \frac{(3+2\omega_{\rm BD})k^2/a^2}{3\Pi(k)}\frac{1}{2a^2}[(\nabla^2_{\mathbf{x}}\varphi - \nabla^2\varphi)](\mathbf{k}), \end{aligned} \tag{6}$$

where $[(\cdots)](\mathbf{k})$ means Fourier transform of $(\cdots)(\mathbf{q})$ and we defined

$$A(k) = A_0\left(1 + \frac{k^2/a^2}{3\Pi(k)}\right), \tag{7}$$

$$\Pi(k) = \frac{1}{3a^2}\big((3 + 2\omega_{\rm BD})k^2 + M_1 a^2\big), \tag{8}$$

$$A_0 = A(k = 0, t) = 4\pi G\bar{\rho}. \tag{9}$$

$A(k)$ is the gravitational strength in the MG cases, while A_0 is for GR. In Brans-Dicke type theories we identify $3 + 2\omega_{\rm BD} = 1/2\beta^2$, allowing us to recognize the β parameter as the strength of the matter to scalar field coupling.

There are two nonlinear terms in eq. (6), the second and third terms of the right hand side (r.h.s.). The second term encodes the nonlinear gravity model, given by

$$\begin{aligned}\delta I(\mathbf{k}) = {} & \frac{1}{2}\int \frac{d^3k_1 d^3k_2}{(2\pi)^3}\delta_{\rm D}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2)M_2(\mathbf{k}_1, \mathbf{k}_2)\varphi(\mathbf{k}_1)\varphi(\mathbf{k}_2) \\ & + \frac{1}{3}\int \frac{d^3k_1 d^3k_2 d^3k_3}{(2\pi)^6}\delta_{\rm D}(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3)M_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)\varphi(\mathbf{k}_1)\varphi(\mathbf{k}_2)\varphi(\mathbf{k}_3).\end{aligned} \tag{10}$$

The M_i functions carry the gravitational properties of particular MG models. It is convenient to insert eq. (6) into eq. (10) to obtain an equation in terms of δ expansions that is useful to find the final expression for the displacements.

The third term accounts for the difference of the Laplacian in x- and q- coordinates, the so called frame-lagging term:

$$\begin{aligned}[(\nabla^2_{\mathbf{x}}\varphi - \nabla^2\varphi)](\mathbf{k}) = {} & [-2\Psi_{i,j}\varphi_{,ij} - \Psi_{i,ij}\varphi_{,j} + 3\Psi_{i,j}\Psi_{j,k}\varphi_{,ki} + 2\Psi_{i,j}\Psi_{j,ik}\varphi_{,k} \\ & + \Psi_{l,li}\Psi_{i,j}\varphi_{,j}](\mathbf{k}),\end{aligned} \tag{11}$$

computed up to third-order PT. This term contributes at different scales, but it is more important at large scales, especially if the theory is expected to reduce to GR in that limit.

It is convenient to recast eq. (3) in terms of only Lagrangian coordinates, then by taking the $\mathbf{x}$-divergence and transforming to Lagrangian coordinates with the aid of $\nabla_{\mathbf{x}i} = (J^{-1})_{ji}\nabla_i$, we get in Fourier space

$$\big[(J^{-1})_{ij}\hat{\mathcal{T}}\Psi_{i,j}\big](\mathbf{k}) = -A(k)\tilde{\delta}(\mathbf{k}) + \frac{2\beta^2k^2}{k^2 + m^2a^2}\delta I(\mathbf{k}) + \frac{1}{2}\frac{m^2}{k^2 + m^2a^2}[(\nabla^2_{\mathbf{x}}\varphi - \nabla^2\varphi)](\mathbf{k}), \tag{12}$$

and

$$\tilde{\delta}(\mathbf{k}) = \int d^3q e^{-i\mathbf{q}\cdot\mathbf{k}}\delta(\mathbf{x}) = \left[\frac{1 - J(\mathbf{q})}{J(\mathbf{q})}\right](\mathbf{k}). \tag{13}$$

Eq. (12) is solved perturbatively using $(J^{-1})_{ij} = \delta_{ij} - \Psi_{i,j} + \Psi_{i,k}\Psi_{k,j} + \cdots$. One obtains the following equation for the displacement at third order PT:

$$\begin{aligned}(\hat{\mathcal{T}} - A(k))[\Psi_{i,i}](\mathbf{k}) = {} & [\Psi_{i,j}\hat{\mathcal{T}}\Psi_{j,i}](\mathbf{k}) - \frac{A(k)}{2}[\Psi_{i,j}\Psi_{j,i}](\mathbf{k}) - \frac{A(k)}{2}[(\Psi_{l,l})^2](\mathbf{k}) \\ & - [\Psi_{i,k}\Psi_{k,j}\hat{\mathcal{T}}\Psi_{j,i}](\mathbf{k}) + \frac{A(k)}{6}[(\Psi_{l,l})^3](\mathbf{k}) \\ & + \frac{A(k)}{2}[\Psi_{l,l}\Psi_{i,j}\Psi_{j,i}](\mathbf{k}) + \frac{A(k)}{3}[\Psi_{i,k}\Psi_{k,j}\Psi_{j,i}](\mathbf{k}) \\ & + \frac{k^2/a^2}{6\Pi(\mathbf{k})}\delta I(\mathbf{k}) + \frac{M_1}{6\Pi(k)}\frac{1}{a^2}[(\nabla^2_{\mathbf{x}}\varphi - \nabla^2\varphi)](\mathbf{k}). \end{aligned} \tag{14}$$

One then can analyze solutions at different PT orders. To linear order one has

$$(\hat{\mathcal{T}} - A(k,t))[\Psi^{(1)}_{i,i}] = 0, \tag{15}$$

which is the same equation for the linear matter overdensity. Hence, the first order solution is

$$\Psi^{(1)}_i(\mathbf{k},t) = i\frac{k_i}{k^2}\delta_L(\mathbf{k},t) = i\frac{k_i}{k^2}D_+(k,t)\delta_L(\mathbf{k},t_0), \tag{16}$$

where we notice the normalization is fixed by eq. (1), and we choose the growth function such that $D_+(k \to 0, t_0) = 1$, where t_0 denotes present time. More generally, Lagrangian displacements are expanded as

$$\Psi_i(\mathbf{k}) = \sum_{n=0}^{\infty}\Psi^{(n)} = \sum_{n=1}^{\infty}\frac{i}{n!}\int_{\mathbf{k}_{1\cdots n}=\mathbf{k}} L^{(n)}_i(\mathbf{k}_1,...,\mathbf{k}_n)\delta_L(\mathbf{k}_1)\cdots\delta_L(\mathbf{k}_n). \tag{17}$$

Hereafter we make use of the shorthand notation

$$\int_{\mathbf{k}_{1\cdots n}=\mathbf{k}} = \int\frac{d^3\mathbf{k}_1}{(2\pi)^3}\cdots\frac{d^3\mathbf{k}_n}{(2\pi)^3}(2\pi)^3\delta_{\rm D}(\mathbf{k}-\mathbf{k}_{1\cdots n}), \tag{18}$$

and $\mathbf{k}_{1\cdots n} = \mathbf{k}_1 + \cdots + \mathbf{k}_n$ denotes the sum of arbitrary number of momenta.

The LPT kernels $L^{(n)}_i$ are obtained order by order using eq. (12). At first order, reading eq. (16) above, we have

$$L^{(1)}_i(\mathbf{k}) = \frac{k_i}{k^2}, \tag{19}$$

while to second order we obtain[3]

$$L^{(2)}_i(\mathbf{k}_1,\mathbf{k}_2) = \frac{3}{7}\frac{k_i}{k^2}\left(\mathcal{A}(\mathbf{k}_1,\mathbf{k}_2) - \mathcal{B}(\mathbf{k}_1,\mathbf{k}_2)\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2}\right), \tag{20}$$

with $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$,

$$\mathcal{A}(\mathbf{k}_1,\mathbf{k}_2) = \frac{7D^{(2)}_{\mathcal{A}}(\mathbf{k}_1,\mathbf{k}_2)}{3D_+(k_1)D_+(k_2)}, \qquad \mathcal{B}(\mathbf{k}_1,\mathbf{k}_2) = \frac{7D^{(2)}_{\mathcal{B}}(\mathbf{k}_1,\mathbf{k}_2)}{3D_+(k_1)D_+(k_2)}, \tag{21}$$

and the second order growth functions are the solutions to equations

$$(\hat{\mathcal{T}} - A(k))D_{\mathcal{A}}^{(2)} = \Bigg[A(k) + (A(k) - A(k_1))\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_2^2} + (A(k) - A(k_2))\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1^2} - \left(\frac{2A_0}{3}\right)\frac{k^2}{a^2}\frac{M_2(\mathbf{k}_1\mathbf{k}_2)}{6\Pi(k)\Pi(k_1)\Pi(k_2)}\Bigg]D_+(k_1)D_+(k_2), \tag{22}$$

$$(\hat{\mathcal{T}} - A(k))D_{\mathcal{B}}^{(2)} = \Big[A(k_1) + A(k_2) - A(k)\Big]D_+(k_1)D_+(k_2), \tag{23}$$

with appropriate initial conditions. The second and third terms in the r.h.s. of eq. (22) stem from the frame-lagging contributions. The fourth term is the second order contribution of δI, which in MG is responsible of the screening mechanism. Since $\mathcal{A}$ and $\mathcal{B}$ depend on $\mathbf{k}_1$ and $\mathbf{k}_2$, the decomposition in eq. (20) is arbitrary and we adopt it because they take values of order unity and the connection to ΛCDM is direct. For this case we obtain

$$\begin{aligned} D_{\mathcal{A}}^{(2)}(t) = D_{\mathcal{B}}^{(2)}(t) &= (\hat{\mathcal{T}} - A_0)^{-1}\left[\frac{3}{2}\Omega_m H^2 D_+^2\right] \\ &= \frac{3}{7}D_+^2(t) + \frac{4}{7}(\hat{\mathcal{T}} - A_0)^{-1}\left[\frac{3}{2}\Omega_m H^2 D_+^2\left(1 - \frac{f^2}{\Omega_m}\right)\right], \end{aligned} \tag{24}$$

thus $\mathcal{A} = \mathcal{B}$ are only time dependent. For $f = \Omega_m^{1/2}$ we get $\mathcal{A}^{\rm EdS} = \mathcal{B}^{\rm EdS} = 1$ and the standard kernels in Einstein-de Sitter (EdS) are recovered.

The third order kernel $L_i^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ and many more details are provided in Aviles & Cervantes,[3] therefore we do not reproduce them here.

3. Screenings by the growth functions

As explained above, MG models introduce a fifth force that changes the Newtonian collapse at cosmic scales. A known effect in nonlinear MG is that its spherical solutions show the screening of fifth force, to drive the solution to the Newtonian one. This happens in large density regions that produce an effective weak coupling that effectively cancels the fifth force; there are other screening mechanisms stemming from different theoretical reasons; see, e.g., K. Koyama.[13] Although screenings are more usually studied in the highly non-linear regime, they also leave imprints in quasi-linear scales that can be captured by cosmological PT.

3.1. *Einstein frame: Symmetron case*

The previous LPT formalism was done in the Jordan frame, but it can be equally well translated to the Einstein frame, for example to study PT within Symmetron fields.

The Symmetron achieves its screening through the coupling of the scalar field to the matter Lagrangian, generating an effective potential with two parameters

(using $\hat{n} = \hat{m} = 1/2$): the scalar field mass (m) and the strength of the coupling (β), see details in reference.[12]

To convert theories posed in the Einstein frame to the Jordan frame, it is convenient to define a rescaled field:

$$\chi(\mathbf{q}, t) \equiv -\frac{2\beta}{C}\frac{\varphi(\mathbf{q}, t)}{M_{\rm Pl}}, \tag{25}$$

hereafter we denote $C \equiv C(\bar{\varphi})$ unless otherwise is explicitly stated.And again, this new scalar field follows a similar Klein Gordon equation

$$\begin{aligned} -\frac{k^2}{2a^2}\chi(\mathbf{k}) = &-(A(k) - A_0)\tilde{\delta}(\mathbf{k}) + \frac{k^2/a^2}{6\Pi(k)}\delta\mathcal{I}(\chi) \\ &- \frac{C}{2\beta^2}\frac{k^2/a^2}{3\Pi(k)}\frac{1}{2a^2}[(\nabla^2_{\mathbf{x}}\chi - \nabla^2\chi)](\mathbf{k}), \end{aligned} \tag{26}$$

and the contribution from screenings is given by

$$\begin{aligned} \delta\mathcal{I}(\mathbf{k}) = &\sum_{n=2}^{\infty}\frac{(-1)^{n+1}}{2^n n!}\frac{C^n\kappa_{n+1}}{\beta^{n+1}}[\chi^n](\mathbf{k}) \\ &+ \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2^n n!}\frac{2A_0C^n\beta_{n+1}}{\beta^{n+1}}[\chi^n\delta](\mathbf{k}), \end{aligned} \tag{27}$$

where $3 + 2\omega_{\rm BD}(a) = \frac{C}{2\beta^2}$.

Solutions to first order are formally the same as eq. (16). To second order one can write the solutions as:

$$\Psi^{i(2)}(\mathbf{k}) = \frac{ik^i}{2k^2}\int_{\mathbf{k}_{12}=\mathbf{k}}\frac{3}{7}\Big(\bar{D}^{(2)}_{\rm NS}(\mathbf{k}_1, \mathbf{k}_2) - \bar{D}^{(2)}_{\rm S}(\mathbf{k}_1, \mathbf{k}_2)\Big)\delta_1\delta_2, \tag{28}$$

where we denote $\delta_{1,2} \equiv \delta_L(\mathbf{k}_{1,2})$ and where we split the second order growth in non-screened (NS) and screening (S) pieces. These growth functions $D^{(2)}$ are solutions to the equations:

$$\begin{aligned} \big(\hat{\mathcal{T}} - A(k)\big)D^{(2)}_{\rm NS}(\mathbf{k}_1, \mathbf{k}_2) = &\left(A(k) - (A(k_1) + A(k_2) - A(k))\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{k_1^2k_2^2}\right. \\ \left.+(A(k) - A(k_1))\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_2^2} + (A(k) - A(k_2))\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1^2}\right)&D_+(k_1)D_+(k_2) \end{aligned} \tag{29}$$

$$\big(\hat{\mathcal{T}} - A(k)\big)D^{(2)}_{\rm S}(\mathbf{k}_1, \mathbf{k}_2) = \left(\frac{2A_0}{3}\right)^2\frac{k^2}{a^2}\frac{M_2(\mathbf{k}_1, \mathbf{k}_2)D_+(k_1)D_+(k_2)}{6\Pi(k)\Pi(k_1)\Pi(k_2)}, \tag{30}$$

and the normalized growth functions are defined as

$$\bar{D}^{(2)}_{\rm S,NS}(\mathbf{k}_1, \mathbf{k}_2, t) = \frac{7}{3}\frac{D^{(2)}_{\rm S,NS}(\mathbf{k}_1, \mathbf{k}_2, t)}{D_+(k_1)D_+(k_2)}. \tag{31}$$

In an EdS universe one obtains the well known result $\bar{D}^{(2)}_{\rm NS} = 1 - (\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2)^2$, while in ΛCDM one gets the same result multiplied by a function that varies slowly with time, such that nowadays $\bar{D}^{(2)\Lambda{\rm CDM}}_{\rm NS} \simeq 1.01 \bar{D}^{(2){\rm EdS}}_{\rm NS}$. As expected, the screening second order growth, $\bar{D}^{(2)}_{\rm S}$, is zero in both EdS and ΛCDM models.

The function $\bar{D}^{(2)}_{\rm S}$ encodes the non-linearities of the effective potential of the scalar field and it yields the second order screening effects that drive the theory to GR at small scales. The total second-order growth function, as can be read from eq. (28), is given by $D^{(2)} = D^{(2)}_{\rm NS} - D^{(2)}_{\rm S}$, such that negative values of $D^{(2)}_{\rm S}$ enhance the growth of perturbations (*anti-screening effects*), while positive values of it yield the standard suppression of the fifth force. In the Symmetron case there are configurations of interacting wave modes that instead of driving the theory towards GR, they drive the theory away from it. In fact, from eq. (30), all triangular configurations with

$$k_1^2 + k_2^2 < k_{AS}^2(a), \tag{32}$$

where the anti-screening wavenumber is defined as

$$k_{AS}^2(a) \equiv \frac{\kappa_3 \beta a^2}{\beta_2} - 2m^2 a^2, \tag{33}$$

will contribute with a negative source to the second order screening as long as the r.h.s. of the above equation is positive.

Analogously, each higher perturbative order carries its own screening and it is efficient over a certain k interval. We emphasize that $\bar{D}^{(2)}_{\rm S}$ and $\bar{D}^{(3)}_{\rm S}$ with positive values will screen the fifth force, while negative values will anti-screen it instead. Given this, the M_i MG functions of each model will determine their screening properties. We plot in Fig. 1 the normalized second and third order screening growth functions $\bar{D}^{(2)}_{\rm S}(\mathbf{k}_1, \mathbf{k}_2)$ and $\bar{D}^{(3)}_{\rm S}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ for the Symmetron, again for different triangle configurations.

The vertical lines correspond to the *screening wavenumber*

$$k_{M_1} = a\sqrt{M_1(a)}, \tag{34}$$

that characterizes the scale at which the screening is present; in fact it is close to the maximum screening growth of the largest triangular contribution (squeezed modes). This scale might serve to parametrize the nonlinear growth to eventually make (e.g. N-point statistics) computations more efficient.

3.2. *Jordan frame: the f(R) case*

As it is known, a f(R) Lagrangian can be transformed into the Jordan frame, and so we can use the formalism of section 2 and, in fact, also the screening/non-screened division of the previous subsection. Here we consider the HS n=1 model for different $f_{R0} = -10^{-4}, -10^{-8}, -10^{-12}$ corresponding to F4, F8, and F12 models. In the

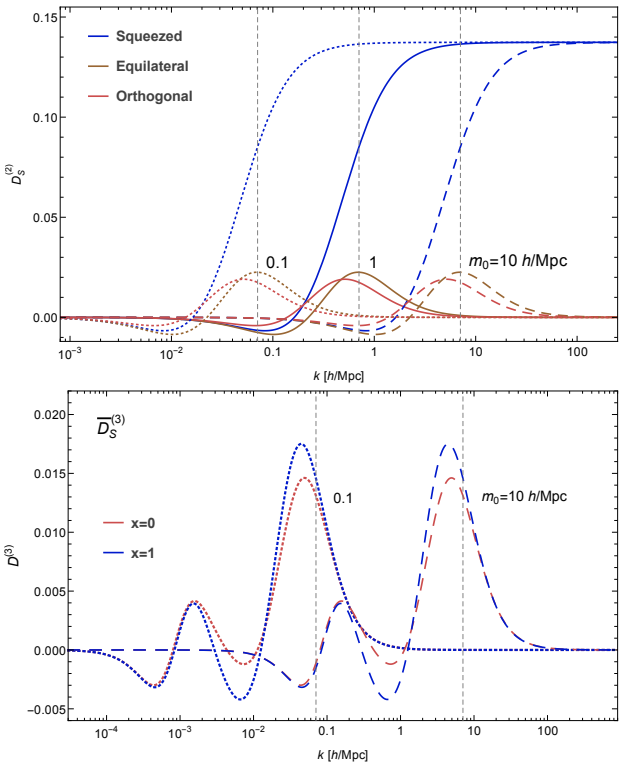


Fig. 1. We show the displacements at second and third order, for the Symmetron model with $a_{ssb} = 0.33$, $\beta_0 = 1$ and different values of m_0, and considering different triangle configurations; results are shown for $z = 0$. The vertical lines are located at $k = a\sqrt{M_1}$, showing the characteristic scale at which the screenings are present. In the upper plot we show the $\bar{D}_{\rm S}^{(2)}$ functions. The plot below shows $\bar{D}_{\rm S}^{(3)}(\mathbf{k}, -\mathbf{p}, \mathbf{p})$ for two values of cosine angle $\mathrm{x} = \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$. In both plots one observes regions of antiscreening behavior ($\bar{D}_{\rm S} < 0$).

upper panel of Fig. 2 we show plots for the second order screening growth functions, $\bar{D}_{\rm S}^{(2)}$, evaluated at redshift $z = 0$ and for different triangular configurations.

Note that in f(R) the screening scale, eq. (34), goes as $k_{M_1} \propto \sqrt{1/|f_{R0}|}$. The lower panel of Fig. 2 shows the growth $D_S^{(3)}(\mathbf{k}, -\mathbf{p}, \mathbf{p})$ for the double squeezed configuration with additionally $|\mathbf{k}| = |\mathbf{p}|$ and for different values of the cosine angle $x = \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$.

In order to study the effects of the background evolution on the screening growth, in Fig. 3 we show it for the F4 model at different redshifts $z = 0, 3$, and 10. The upper panel uses a ΛCDM background cosmology and the lower panel an EdS background evolution. In ΛCDM the screening curves are narrower and reach smaller maxima, this is expected because the cosmic acceleration attenuates the clustering of dark matter, and hence the nonlinear effects. Instead, in an EdS background the pattern of the growth is preserved. The vertical lines again denote the scale k_{M_1}, which are in ΛCDM slightly shifted in comparison with the EdS results.

It is worthy to observe that in f(R) the nonlinearities of the Klein-Gordon equation lead to screening for any configuration. This is manifested in Figs. 2 and 3 because the screening growth functions always take positive values.

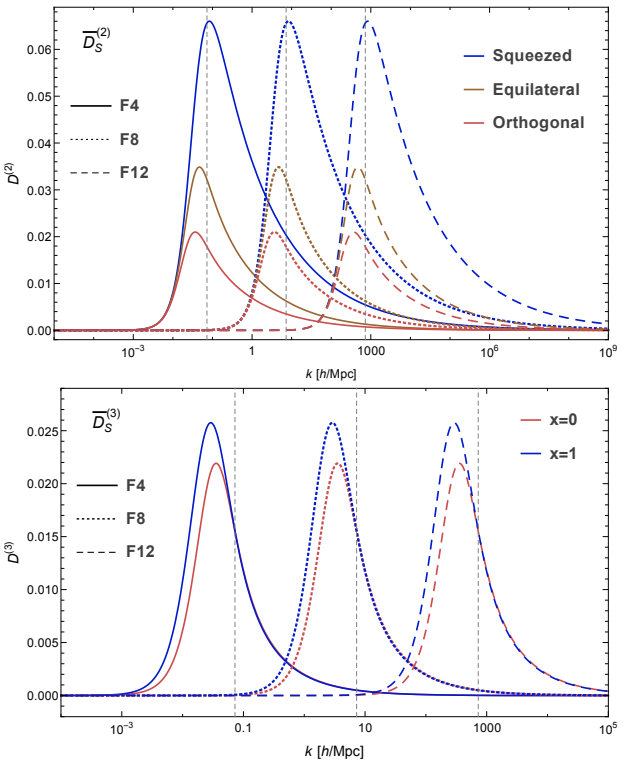


Fig. 2. The upper plot shows the $\bar{D}_{\rm S}^{(2)}$ functions for F4, F8, and F12 models, considering different triangle configurations. The vertical lines are located at $k = a\sqrt{M_1}$, showing the characteristic scale at which the screenings are present. The lower panel shows the third order growth $\bar{D}_{\rm S}^{(3)}$ for different angles $x = \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$ and with $k = p$. The f(R) n=1 shows no antiscreening.

We have considered other models that show different screening mechanisms, such as DGP based on the Vainshtein mechanism. In this case, the second and third order screening growths always act to attenuate the fifth force that modifies Newtonian gravity: no antiscreening is found. We refer to further details on the screening treatment to reference.[7]

4. From Lagrangian to standard PT

The formalism presented above was developed for LPT. However, one can demonstrate that from the kernel solutions of LPT one can obtain the kernels in SPT,

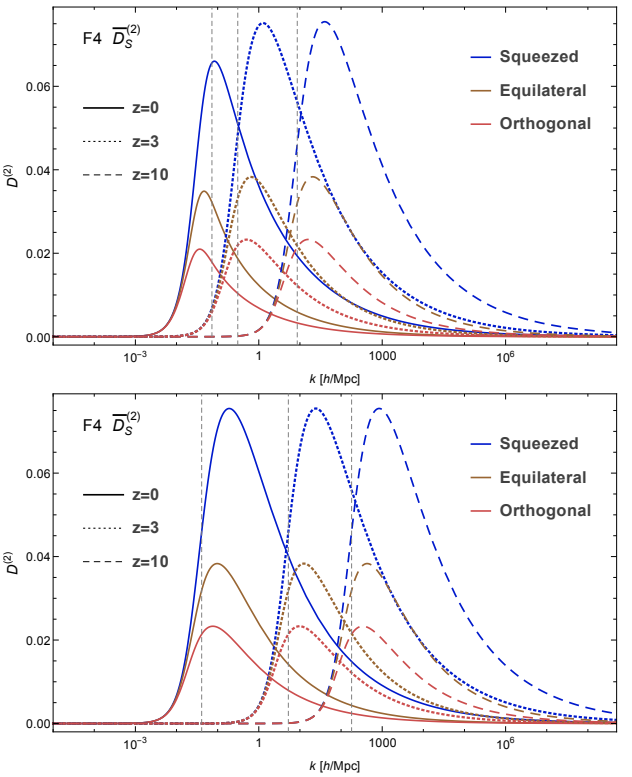


Fig. 3. $\bar{D}_{\rm S}^{(2)}$ functions for the F4 for redshifts $z = 0, 3, 10$, and considering different triangle configurations. The vertical lines are located at $k = a\sqrt{M_1}$, showing the characteristic scale at which the screenings are present. The upper panel shows a ΛCDM background evolution while the bottom panel an EdS background evolution.

so establishing a formal connection between these two formalisms. This was done for general cosmologies in reference,[4] where it is shown that the SPT power spectrum computed from the LPT formalism coincides with the Eulerean, SPT power spectrum. Here omit details that can be found in the above reference, where it is shown that one can express the overdensities as

$$\delta^{(n)}(\mathbf{k}) = \sum_{\ell=1}^{n} \sum_{m_1+\cdots+m_\ell=n} \frac{k_{i_1}\cdots k_{i_\ell}}{\ell! m_1! \cdots m_\ell!} \times \int_{\mathbf{p}_{1\cdots n}=\mathbf{k}} L_{i_1}^{(m_1)}(\mathbf{p}_1, ..., \mathbf{p}_{m_1}) \cdots L_{i_\ell}^{(m_\ell)}(\mathbf{p}_{m_{\ell-1}+1}, ..., \mathbf{p}_{m_\ell}) \times \delta_L(\mathbf{p}_1) \cdots \delta_L(\mathbf{p}_n). \tag{35}$$

This expression provides the SPT F_n kernels out of the LPT kernels (L_n):

$$F_n(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \sum_{\ell=1}^{n} \sum_{m_1+\cdots+m_\ell=n} \frac{k_{i_1}\cdots k_{i_\ell}}{\ell! m_1!\cdots m_\ell!} \times L_{i_1}^{(m_1)}(\mathbf{k}_1,...,\mathbf{k}_{m_1})\cdots L_{i_\ell}^{(m_\ell)}(\mathbf{k}_{m_{\ell-1}+1},...,\mathbf{k}_{m_\ell}) \tag{36}$$

with $\mathbf{k} = \mathbf{k}_1 + \cdots + \mathbf{k}_n$.

As an example of these results here we show the $n = 2$ kernel:

$$F_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{1}{2} + \frac{3}{14}\mathcal{A} + \left(\frac{1}{2} - \frac{3}{14}\mathcal{B}\right)\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{\mathbf{k}_1^2\mathbf{k}_2^2} + \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{2}\left(\frac{1}{k_1^2}+\frac{1}{k_2^2}\right). \tag{37}$$

and, analogously, for the velocity field kernels one finds a similar relation general relationship. For $n = 2$ we obtain:

$$G_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{3}{14}\mathcal{A}(f_1+f_2) + \frac{3\dot{\mathcal{A}}}{14H} + \left(\frac{f_1+f_2}{2} - \frac{3}{14}\mathcal{B}(f_1+f_2) - \frac{3\dot{\mathcal{B}}}{14H}\right)\frac{(\mathbf{k}_1\cdot\mathbf{k}_2)^2}{\mathbf{k}_1^2\mathbf{k}_2^2} + \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{2}\left(\frac{f_2}{k_1^2}+\frac{f_1}{k_2^2}\right), \tag{38}$$

where $f_{1,2} = f(k_{1,2})$ are growth rates for two arbitrary modes. $\mathcal{A}$ and $\mathcal{B}$ are given by eq. (21).

In Fig. 4 we plot the ratios of the SPT 1-loop to linear matter power spectra, for the ΛCDM model, as well as for HS n=1 F4, and the normal branch of DGP with crossover scale fixed by the Hubble constant, $r_c = H_0^{-1}$. We fixed the cosmological parameters to $\Omega_m = 0.281$, $\Omega_b = 0.046$, $h = 0.697$, $n_s = 0.971$, and $\sigma_8 = 0.82$, corresponding to WMAP 9 years best fit to ΛCDM.[14] For F4 model, the background cosmology is indistinguishable to that in ΛCDM. This is not the case for DGP. However, as it is usual in the literature, for DGP we fix to a ΛCDM background. In such a way we can compare the differences in the growth of perturbations due to the fifth force and not to a different Hubble flow. Despite this, the power spectrum in DGP suffers a scale independent shift because $A(k,t) = A(t) \neq A_0$. We plot the the power spectra with and without screenings; the later are achieved by setting $M_2 = M_3 = 0$ in eq. (10).

5. Matter power spectra

We now construct the power spectra for STP for general cosmologies. We follow the approach developed in Aviles et al.[15] As described in section 2, the Poisson equation is given by eq. (4), whereas the fluid equations are:

$$\partial_t\delta(\mathbf{x},t) + \frac{1}{a}\partial_i\left[(1+\delta)v^i\right] = 0, \tag{39}$$

$$\partial_t v^i(\mathbf{x},t) + \frac{1}{a}v^j\partial_j v^i + Hv^i + \frac{1}{a}\partial_i\psi = 0, \tag{40}$$

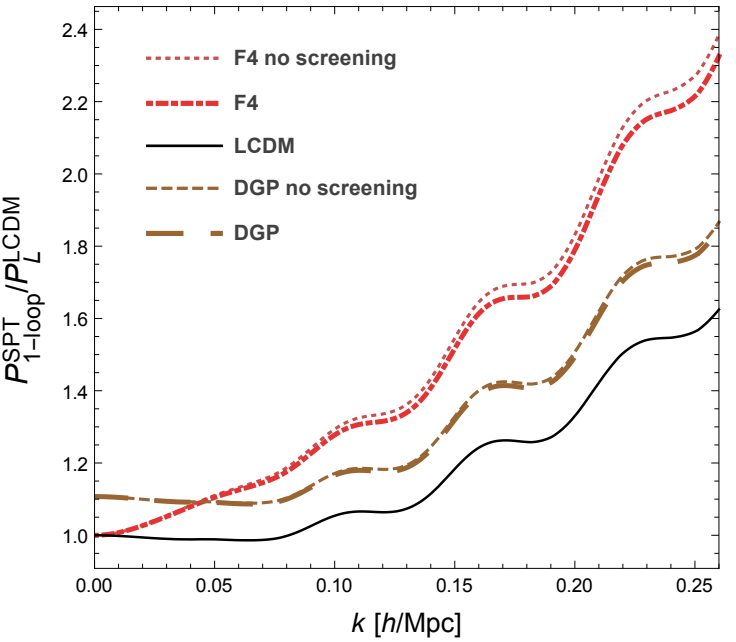


Fig. 4. Ratio of 1-loop SPT to ΛCDM linear matter power spectra for ΛCDM, F4, and DGP models at redshift $z = 0$. The role of screening is self evident at nonlinear scales. We fix cosmological parameters to the best fit of the WMAP Nine-year results.[14]

where $v^i(t) = \frac{dx^i(t)}{d\tau} = a\dot{x}^i(t)$ is the CDM fluid's element peculiar velocity; $\mathbf{x}$ is its comoving coordinate, t the cosmic time, and $d\tau = \frac{1}{a}dt$ the conformal time, such that the total velocity is $v^i_T = aHx^i + v^i$.

We use the (dimensionless) velocity divergence

$$\theta(\mathbf{x}, t) = -\frac{\partial_i v^i}{aHf_0}, \tag{41}$$

where $f_0(t)$ is an arbitrary function of time that will be fixed to be the logarithmic growth rate at a convenient scale. We assume the transverse piece of the velocity is negligible at large scales, hence it is a longitudinal field fully specified by θ.

In Fourier space the continuity and Euler equations can be written as

$$\frac{1}{H}\frac{\partial\delta(\mathbf{k})}{\partial t} - f_0\theta(\mathbf{k}) = f_0 \int_{\mathbf{k}_{12}=\mathbf{k}} \alpha(\mathbf{k}_1, \mathbf{k}_2)\theta(\mathbf{k}_1)\delta(\mathbf{k}_2), \tag{42}$$

$$\frac{1}{H}\frac{\partial f_0\theta(\mathbf{k})}{\partial t} + \left(2 + \frac{\dot{H}}{H^2}\right) f_0\theta(\mathbf{k}) - \frac{A(k)}{H^2}\delta(\mathbf{k}) - \frac{S(\mathbf{k})}{H^2} = f_0^2 \int_{\mathbf{k}_{12}=\mathbf{k}} \beta(\mathbf{k}_1, \mathbf{k}_2)\theta(\mathbf{k}_1)\theta(\mathbf{k}_2), \tag{43}$$

with

$$\alpha(\mathbf{k}_1,\mathbf{k}_2)=1+\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1^2},\qquad \beta(\mathbf{k}_1,\mathbf{k}_2)=\frac{k_{12}^2(\mathbf{k}_1\cdot\mathbf{k}_2)}{2k_1^2k_2^2}. \tag{44}$$

where $S(k)$ is related to eq. (10), see Aviles et al.[15] One finds perturbative solutions at the different orders according to eqs. (42, 43) and the nonlinear expansion (10). The solutions are same those given in eqs. (37) and (38), and so for higher orders.

Having found the kernels up to third order, we can now construct spectra and cross-spectra of the velocity and density fields. At 1-loop in PT, one can use the general expressions

$$P_{ab}(k)=P_{ab}^{L}(k)+P_{ab}^{22}(k)+P_{ab}^{13}(k), \tag{45}$$

where a and b refer to δ or θ fields, and linear power spectra $P_{ab}^{L}(k)$

$$P_{\delta\delta}^{L}(k)\equiv P_L(k),\quad P_{\delta\theta}^{L}(k)=\frac{f(k)}{f_0}P_L(k),\quad P_{\theta\theta}^{L}(k)=\left(\frac{f(k)}{f_0}\right)^2P_L(k), \tag{46}$$

and leading nonlinear contributions

$$P_{\delta\delta}^{22}(k)=2\int_{\mathbf{p}}\big[F_2(\mathbf{p},\mathbf{k}-\mathbf{p})\big]^2P_L(p)P_L(|\mathbf{k}-\mathbf{p}|), \tag{47}$$

$$P_{\delta\theta}^{22}(k)=2\int_{\mathbf{p}}F_2(\mathbf{p},\mathbf{k}-\mathbf{p})G_2(\mathbf{p},\mathbf{k}-\mathbf{p})P_L(p)P_L(|\mathbf{k}-\mathbf{p}|), \tag{48}$$

$$P_{\theta\theta}^{22}(k)=2\int_{\mathbf{p}}\big[G_2(\mathbf{p},\mathbf{k}-\mathbf{p})\big]^2P_L(p)P_L(|\mathbf{k}-\mathbf{p}|), \tag{49}$$

$$P_{\delta\delta}^{13}(k)=6P_L(k)\int_{\mathbf{p}}F_3(\mathbf{k},-\mathbf{p},\mathbf{p})P_L(p), \tag{50}$$

$$P_{\delta\theta}^{13}(k)=3P_L(k)\int_{\mathbf{p}}\big[F_3(\mathbf{k},-\mathbf{p},\mathbf{p})G_1(\mathbf{k})+G_3(\mathbf{k},-\mathbf{p},\mathbf{p})\big]P_L(p), \tag{51}$$

$$P_{\theta\theta}^{13}(k)=6P_L(k)\int_{\mathbf{p}}G_3(\mathbf{k},-\mathbf{p},\mathbf{p})G_1(\mathbf{k})P_L(p). \tag{52}$$

The above expressions are valid for any cosmological model such as MG models, dark energy, or massive neutrinos—in general, for models that introduce a new degree of freedom to the dynamics. For the different models, these expressions for power spectra and kernels are the same, the difference being the corresponding growth solutions through $\mathcal{A}$ and $\mathcal{B}$ (see e.g. eq. 37) that are inside the kernels.

6. Bias Expansion and tracer's power spectra

We now turn to the description of dark matter halos and galaxies, instead of only matter fields. That is, we go into tracers, that are more closely related to observations.

We expand the tracer densities in terms of a set of operators, including the leading curvature operators, labeled with "m" to make reference to matter fields,

in a manner similar to McDonald and Roy[16]:

$$\delta(\mathbf{x}) = c_\delta \delta_m + c_{\nabla^2\delta}\nabla^2\delta_m + \frac{1}{2}c_{\delta^2}\delta_m^2 + \frac{1}{2}c_{s^2}s^2 + \frac{1}{6}c_{\delta^3}\delta_m^3 + \frac{1}{2}c_{\delta s^2}\delta s^2 + c_\psi\psi + c_{st}st + \frac{1}{2}c_{s^3}s^3, \tag{53}$$

$$\theta(\mathbf{x}) = \theta_m + c_{\nabla^2\theta}\nabla^2\theta_m. \tag{54}$$

Since for MG models an additional scale shows up, the bias coefficients are re-scaled. The first order terms are:

$$\delta^{(1)}(\mathbf{k}) = (c_\delta - c_{\nabla^2\delta}k^2)\delta_m^{(1)}(\mathbf{k}), \text{ and} \tag{55}$$

$$\theta^{(1)}(\mathbf{k}) = (1 - c_{\nabla^2\theta}k^2)\theta_m^{(1)}(\mathbf{k}). \tag{56}$$

The expansion goes to third order that can be found in Aviles et al.[15]

One can now find the tracers' power spectra, that will be given in terms of the above biases and matter power spectra of the previous section. We obtain:

$$P_{\delta\delta}(k) = (b_1 - b_{\nabla^2\delta}k^2)^2 P_L(k) + b_1^2 P_{m,\delta\delta}^{\rm loop}(k) + 2b_1 b_2 P_{b_1 b_2}(k) + 2b_1 b_{s^2} P_{b_1 b_{s^2}}(k) + b_2^2 P_{b_2^2}(k) + 2b_2 b_{s^2} P_{b_2 b_{s^2}}(k) + b_{s^2}^2 P_{b_{s^2}^2}(k) + 2b_1 b_{3nl}\sigma_3^2(k)P_L(k), \tag{57}$$

$$P_{\delta\theta}(k) = (b_1 - b_{\nabla^2\delta}k^2)(1 - b_{\nabla^2\theta}k^2)\frac{f(k)}{f_0}P_L(k) + b_1 P_{m,\delta\theta}^{\rm loop}(k) + b_2 P_{b_2,\theta}(k) + b_{s^2}P_{b_{s^2},\theta}(k) + b_{3nl}\sigma_3^2(k)\frac{f(k)}{f_0}P_L(k), \tag{58}$$

$$P_{\theta\theta}(k) = (1 - b_{\nabla^2\theta}k^2)^2\left(\frac{f(k)}{f_0}\right)^2 P_L(k) + P_{m,\theta\theta}^{\rm loop}(k), \tag{59}$$

with

$$P_{b_1 b_2}(k) = \int_{\mathbf{k}_{12}=\mathbf{k}} F_2(\mathbf{k}_1,\mathbf{k}_2)P_L(\mathbf{k}_1)P_L(\mathbf{k}_2), \tag{60}$$

$$P_{b_1 b_{s^2}}(k) = \int_{\mathbf{k}_{12}=\mathbf{k}} F_2(\mathbf{k}_1,\mathbf{k}_2)S_2(\mathbf{k}_1,\mathbf{k}_2)P_L(\mathbf{k}_1)P_L(\mathbf{k}_2), \tag{61}$$

$$P_{b_2^2}(k) = \frac{1}{2}\int_{\mathbf{k}_{12}=\mathbf{k}} P_L(\mathbf{k}_1)\big[P_L(\mathbf{k}_2) - P_L(\mathbf{k}_1)\big], \tag{62}$$

$$P_{b_2 b_{s^2}}(k) = \frac{1}{2}\int_{\mathbf{k}_{12}=\mathbf{k}} P_L(\mathbf{k}_1)\left[P_L(\mathbf{k}_2)S_2(\mathbf{k}_1,\mathbf{k}_2) - \frac{2}{3}P_L(\mathbf{k}_1)\right], \tag{63}$$

$$P_{b_{s^2}^2}(k) = \frac{1}{2}\int_{\mathbf{k}_{12}=\mathbf{k}} P_L(\mathbf{k}_1)\left[P_L(\mathbf{k}_2)[S_2(\mathbf{k}_1,\mathbf{k}_2)]^2 - \frac{4}{9}P_L(\mathbf{k}_1)\right], \tag{64}$$

and

$$P_{b_2,\theta}(k) = \int_{\mathbf{k}_{12}=\mathbf{k}} G_2(\mathbf{k}_1,\mathbf{k}_2) P_L(\mathbf{k}_1) P_L(\mathbf{k}_2), \tag{65}$$

$$P_{b_{s^2},\theta}(k) = \int_{\mathbf{k}_{12}=\mathbf{k}} G_2(\mathbf{k}_1,\mathbf{k}_2) S_2(\mathbf{k}_1,\mathbf{k}_2) P_L(\mathbf{k}_1) P_L(\mathbf{k}_2), \tag{66}$$

with renormalized bias parameters

$$b_1 = c_\delta + \left[\frac{31}{24} c_{\delta^2}\mathcal{C}_{\delta^2} + \frac{1}{2} c_{\delta^3}\mathcal{C}_{\delta^3} + \frac{1}{3} c_{\delta s^2}\mathcal{C}_{\delta s^2} + \frac{68}{63} c_{s^2}\mathcal{C}_{s^2} - \frac{16}{63} c_{st}\mathcal{C}_{st}\right]\sigma^2, \tag{67}$$

with $\mathcal{C}_\mathcal{O}$ constants of order unity and they are required in cosmologies beyond EdS, as we have seen for $\mathcal{C}_{s^2}$ and $\mathcal{C}_{st}$. At 1-loop in the PS the rest of biasing parameters remain equal after renormalization: $b_2 = c_{\delta^2}$ and $b_{s^2} = c_{s^2}$.

7. Redshift space distortions in MG

We now introduce a model for RSD in alternative cosmologies. An object located at a comoving distance $\mathbf{x}$ is observed to be at an apparent position $\mathbf{s}$, such that the map between real and redshift space positions is given by the non-relativistic, longitudinal Doppler effect,

$$\mathbf{s} = \mathbf{x} + \mathbf{u}, \tag{68}$$

where $\mathbf{u}$ is given in terms of the divergence field $\mathbf{u}(\mathbf{k}) = i f_0 \hat{\mathbf{n}} \frac{\mathbf{k}\cdot\hat{\mathbf{n}}}{k^2}\theta(\mathbf{k})$.

Clearly, the map to redshift coordinates conserves the number of tracers, $\big[1 + \delta_s(\mathbf{s})\big]d^3s = \big[1 + \delta(\mathbf{x})\big]d^3x$, yielding

$$(2\pi)^3\delta_{\rm D}(\mathbf{k}) + \delta_s(\mathbf{k}) = \int d^3x \big(1 + \delta(\mathbf{x})\big) e^{-i\mathbf{k}\cdot(\mathbf{x}+\mathbf{u}(\mathbf{x}))}, \tag{69}$$

and the redshift-space PS becomes[17,18]

$$(2\pi)^3\delta_{\rm D}(\mathbf{k}) + P_s(\mathbf{k}) = \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \Big[1 + \mathcal{M}(\mathbf{J} = \mathbf{k}, \mathbf{x})\Big], \tag{70}$$

with the velocity moments generating function

$$1 + \mathcal{M}(\mathbf{J},\mathbf{x}) = \left\langle \big(1 + \delta(\mathbf{x}_1)\big)\big(1 + \delta(\mathbf{x}_2)\big) e^{-i\mathbf{J}\cdot\Delta\mathbf{u}} \right\rangle, \tag{71}$$

where $\Delta\mathbf{u} = \mathbf{u}(\mathbf{x}_2) - \mathbf{u}(\mathbf{x}_1)$ and $\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$. Function $\mathcal{M}$ (or its Fourier transform) plays a central role in RSD. Different expansion procedures of eq. (71) yield different approaches to RSD modeling, grouped in Vlah et al[18] as: direct Lagrangian, moment expansion, streaming model, and smoothing kernel. We will follow here the moment expansion (ME) approach, in which the exponential in the generating function is expanded and the moments are evaluated.

The m-th density weighted velocity field moment of the generating function is an m-rank tensor defined as[17,18]

$$\Xi^{\mathrm{m}}_{i_1\cdots i_{\mathrm{m}}}(\mathbf{x}) \equiv i^{\mathrm{m}} \frac{\partial^{\mathrm{m}}}{\partial J_{i_1}\cdots\partial J_{i_{\mathrm{m}}}}\big[1+\mathcal{M}(\mathbf{J},\mathbf{x})\big]\Big|_{\mathbf{J}=0} = \langle\big(1+\delta_1\big)\big(1+\delta_2\big)\Delta u_{i_1}\cdots\Delta u_{i_{\mathrm{m}}}\rangle, \tag{72}$$

with $\delta_1=\delta(\mathbf{x}_1)$ and $\delta_2=\delta(\mathbf{x}_2)$. The PS in the moment expansion approach becomes

$$(2\pi)^3\delta_{\mathrm{D}}(\mathbf{k}) + P_s(\mathbf{k}) = \sum_{\mathrm{m}=0}^{\infty}\frac{(-i)^{\mathrm{m}}}{\mathrm{m}!}k_{i_1}\ldots k_{i_{\mathrm{m}}}\tilde{\Xi}^{\mathrm{m}}_{i_1\cdots i_{\mathrm{m}}}(\mathbf{k}), \tag{73}$$

where the $\tilde{\Xi}^{\mathrm{m}}_{i_1\cdots i_{\mathrm{m}}}(\mathbf{k})$ are the Fourier moments of the generating function —the Fourier transforms of their configuration space counterparts, $\Xi^{\mathrm{m}}_{i_1\cdots i_n}(\mathbf{x})$.

Now, to compute the moments we make the following ansatz:

$$P^{\mathrm{m}}(k,\mu) \equiv \frac{(-i)^{\mathrm{m}}}{\mathrm{m}!}k_{i_1}\cdots k_{i_{\mathrm{m}}}\tilde{\Xi}^{\mathrm{m}}_{i_1\cdots i_{\mathrm{m}}}(\mathbf{k}) = \sum_{n=0}^{\mathrm{m}}\mu^{2n}f_0^{\mathrm{m}}I_n^{\mathrm{m}}(k), \tag{74}$$

such that the total PS in redshift coordinates (s) is

$$P_s^{\mathrm{ME}}(k,\mu) = \sum_{\mathrm{m}=0}^{\infty}P^{\mathrm{m}}(k,\mu), \tag{75}$$

up to a Dirac delta function localized at $\mathbf{k}=0$. Whereas the first equality (definition) in eq. (74) depends on $\mathbf{k}$, implicitly on k and $\mu\equiv\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}$, the second equality is an ansatz with explicit angular dependence as even powers of μ. This ansatz has been shown to work, at least for moments $\mathrm{m}=0,\ldots,4$, in Aviles et al,[15] where the functions $I_n^{\mathrm{m}}(k)$ are given for the momenta necessary to compute the 1-loop expansion.

8. Considering UV and IR modes

There are still two important ingredients before we have a final model: EFT counterterms and infrared (IR) resummation of modes.

• As it is known, UV modes need to be cut off in the integrals, since small scale modes are not properly captured in PT, so it will not otherwise render the correct dynamics. The EFT for large scale structure formalism, see for instance Baumann et al[19] and Hertzberg et al,[20] cuts-off the loop integrals, by directly smoothing the overdensity fields by an arbitrary scale, and introduces counterterms necessary to remove the cut-off dependence on the final expressions. In redshift space, the situation is more complex, because the counterterms not only model small scales, non-perturbative physics, but also the non-linear mathematical map between real space and redshift space densities, see various authors.[21–23] In our case, the final expression turns to:

$$P_s^{\mathrm{EFT}}(k,\mu) = P_s^{\mathrm{ME}}(k,\mu) + (\alpha_0+\alpha_1\mu^2+\cdots)k^2P_L(k) + \tilde{c}\big(\mu k f_0\big)^4P_s^K(k,\mu) + P_{\mathrm{shot}}, \tag{76}$$

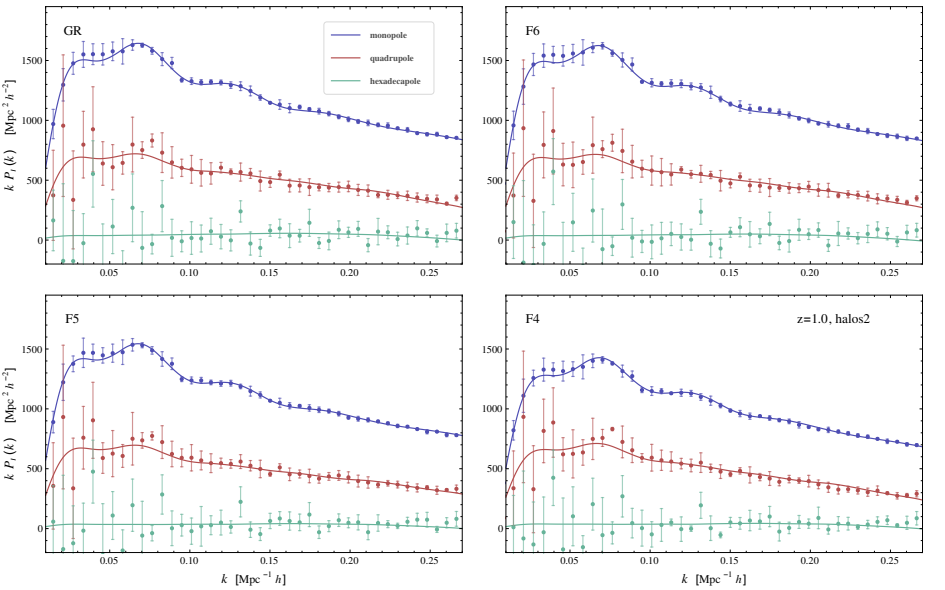


Fig. 5. Redshift-space power spectrum multipoles for halo catalogue 2 ($4.5{\times}10^{12} < M_h < 1{\times}10^{13}$ $M_\odot h^{-1}$) at redshift $z = 1$.

where the α_i stem from the nonlinear mathematical map between real space and redshift space densities and the $\tilde{c}$ from nonlinearities from the Fingers of God (FoG) that couple velocities and density fields. $P_{\rm shot}$ models the stochastic terms, uncorrelated with long wave-length fluctuations.

• Despite the success of SPT-EFT in modeling the broadband PS, the theory yet gives poor results in modeling the BAO since long-wavelength displacement fields, though being essentially linear, stream largely contributing to damp features in the PS in a manner that is non-perturbative under an SPT scheme.Then, in order to model the spread and degradation of the BAO oscillations due to large scale bulk flows, we employ IR-resummations[24] as implemented by Ivanov et al.[22,25]

The final result is a complete PS model for RSD that can be compared to dark matter halos or galaxies. In Fig. 5 we plot our analytical results against RSD multipoles for the HS n=1 model, extracted from `ELEPHANT` simulations,[1] with halo masses in the interval $4.5 \times 10^{12} < M_h < 1 \times 10^{13}$ $M_\odot h^{-1}$ evaluated at $z = 1$. These results show that theory renders precise results for wave modes as big as $k = 0.25$ h/Mpc.

9. Conclusions

In the last few years the PT for LSS has been developed to a point that most problems have been tackled, and it seems now ready to be feasibly compared against

simulations and observations, to render consistent results and to be able to extract cosmological parameter information at the level that stage-IV galaxy surveys should provide, among other probes.

The present contribution presents some PT developments for models that possess an extra degree of freedom, that is, models of MG, dynamical dark energy, or considering massive neutrinos within LCDM.[26,27] We explained mainly our results, but, because of the limited space here, we did not provide a full list of references, however, this can be found in the our cited papers in the bibliography.

Acknowledgments

The authors acknowledge support by CONACyT project 283151.

References

1. M. Cautun, E. Paillas, Y.-C. Cai, S. Bose, J. Armijo, B. Li and N. Padilla, The Santiago–Harvard–Edinburgh–Durham void comparison – I. SHEDding light on chameleon gravity tests, *Mon. Not. Roy. Astron. Soc.* **476**, 3195 (2018).
2. K. Koyama, A. Taruya and T. Hiramatsu, Non-linear Evolution of Matter Power Spectrum in Modified Theory of Gravity, *Phys. Rev. D* **79**, p. 123512 (2009).
3. A. Aviles and J. L. Cervantes-Cota, Lagrangian perturbation theory for modified gravity, *Phys. Rev. D* **96**, p. 123526 (2017).
4. A. Aviles, M. A. Rodriguez-Meza, J. De-Santiago and J. L. Cervantes-Cota, Nonlinear evolution of initially biased tracers in modified gravity, *JCAP* **11**, p. 013 (2018).
5. G. Valogiannis and R. Bean, Convolution Lagrangian perturbation theory for biased tracers beyond general relativity, *Phys. Rev. D* **99**, p. 063526 (2019).
6. G. Valogiannis, R. Bean and A. Aviles, An accurate perturbative approach to redshift space clustering of biased tracers in modified gravity, *JCAP* **01**, p. 055 (2020).
7. A. Aviles, J. L. Cervantes-Cota and D. F. Mota, Screenings in Modified Gravity: A perturbative approach, *Astron. Astrophys.* **622**, p. A62 (2019).
8. A. Aghamousa *et al.*, The DESI Experiment Part I: Science,Targeting, and Survey Design (10 2016).
9. L. Amendola *et al.*, Cosmology and fundamental physics with the Euclid satellite, *Living Rev. Rel.* **21**, p. 2 (2018).
10. A. Abate *et al.*, Large Synoptic Survey Telescope: Dark Energy Science Collaboration (11 2012).
11. S. Alam *et al.*, Testing the theory of gravity with DESI: Estimators, predictions and simulation requirements (11 2020).
12. P. Brax and P. Valageas, Impact on the power spectrum of Screening in Modified Gravity Scenarios, *Phys. Rev.* **D88**, p. 023527 (2013).
13. K. Koyama, Cosmological Tests of Modified Gravity, *Rept. Prog. Phys.* **79**, p. 046902 (2016).
14. G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack and E. L. Wright, Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results, *APJS* **208**, p. 19 (October 2013).

15. A. Aviles, G. Valogiannis, M. A. Rodriguez-Meza, J. L. Cervantes-Cota, B. Li and R. Bean, Redshift space power spectrum beyond Einstein-de Sitter kernels, *JCAP* **04**, p. 039 (2021).
16. P. McDonald and A. Roy, Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS, *JCAP* **08**, p. 020 (2009).
17. R. Scoccimarro, Redshift-space distortions, pairwise velocities and nonlinearities, *Phys. Rev. D* **70**, p. 083007 (2004).
18. Z. Vlah and M. White, Exploring redshift-space distortions in large-scale structure, *JCAP* **1903**, p. 007 (2019).
19. D. Baumann, A. Nicolis, L. Senatore and M. Zaldarriaga, Cosmological Non-Linearities as an Effective Fluid, *JCAP* **1207**, p. 051 (2012).
20. M. P. Hertzberg, Effective field theory of dark matter and structure formation: Semi-analytical results, *Phys. Rev. D* **89**, p. 043521 (2014).
21. A. Perko, L. Senatore, E. Jennings and R. H. Wechsler, Biased Tracers in Redshift Space in the EFT of Large-Scale Structure (10 2016).
22. M. M. Ivanov, M. Simonović and M. Zaldarriaga, Cosmological Parameters from the BOSS Galaxy Power Spectrum, *JCAP* **05**, p. 042 (2020).
23. S.-F. Chen, Z. Vlah and M. White, Consistent Modeling of Velocity Statistics and Redshift-Space Distortions in One-Loop Perturbation Theory, *JCAP* **07**, p. 062 (2020).
24. L. Senatore and M. Zaldarriaga, The IR-resummed Effective Field Theory of Large Scale Structures, *JCAP* **02**, p. 013 (2015).
25. M. M. Ivanov and S. Sibiryakov, Infrared Resummation for Biased Tracers in Redshift Space, *JCAP* **07**, p. 053 (2018).
26. A. Aviles and A. Banerjee, A Lagrangian Perturbation Theory in the presence of massive neutrinos, *JCAP* **10**, p. 034 (2020).
27. A. Aviles, A. Banerjee, G. Niz and Z. Slepian, Clustering in Massive Neutrino Cosmologies via Eulerian Perturbation Theory (6 2021).

Testing modified gravity theories with marked statistics

Alejandro Aviles

Consejo Nacional de Ciencia y Tecnología,
Av. Insurgentes Sur 1582, Colonia Crédito Constructor, Del. Benito Juárez, 03940,
Ciudad de México, México,
and
Departamento de Física, Instituto Nacional de Investigaciones Nucleares
Apartado Postal 18-1027, Col. Escandón, Ciudad de México,11801, México
E-mail: avilescervantes@gmail.com

In the last two decades, Modified Gravity (MG) models have been proposed to explain the accelerated expansion of the Universe. However, one of the main difficulties these theories face is that they must reduce to General Relativity (GR) at sufficiently high energy densities, such as those found in the solar system. To achieve this, MG theories typically employ so-called screening mechanisms: nonlinear effects that bring them to GR at the appropriate limits. For this reason, low-energy regions where the screenings do not operate efficiently, such as cosmic voids, are identified as ideal laboratories for testing GR. Hence, the use of *marked* statistics that up-weight low energy densities have been proposed for being implemented with data from future galaxy surveys. In this proceeding note, we show how to construct theoretical templates for such statistics and test their accuracy with the use of N-body simulations.

Keywords: Modified Gravity; Cosmic Large Scale Structure; Testing General Relativity.

1. Modified gravity in Cosmology: a very brief introduction

The standard model of Cosmology, the so-called ΛCDM, is the most successful theory to date for describing the large-scale behaviour of our Universe. The outstanding way this model fits the cosmic background radiation (CMB) anisotropies measurements from the Planck satellite[1]—and prior to that from the WMAP and COBE experiments—leaves small room for modifications of it. However, among other drawbacks in Cosmology, nowadays cosmic acceleration is still not well understood, and in models that modify the background history of the Universe at late times, the CMB differs mainly because the journey of photons from the last-scattering surface to us is affected. This is a secondary effect called the integrated Sachs-Wolfe effect and its influence over the CMB is mainly localized at small multipoles, where the error on measurements is nonetheless large because of the cosmic variance—the fact that we have only one universe and hence a small number of long wave-length modes to average over. Such a liberty has led to a lot of proposals (literally hundreds of them) as alternatives to the cosmological constant to explain the speeding up of the expansion of the Universe. Unfortunately, there is much more literature proposing new gravity models or that makes simple tests to them (the vast majority focusing only on the homogeneous and isotropic cosmology; I will come back to this in a moment), than literature that explore methods to probe the clustering due to gravity

at cosmological scales. Of course, this is a highly unbalanced situation and observational cosmologists cannot test each of the proposed models. These alternatives can be categorized into two big branches: dark energy (DE) and Modified Gravity (MG); though, one should be careful since there is not a well defined boundary between them. In this proceedings note, we will focus on the latter assuming well defined, and to the particular case of chameleon fields[2,3] although the techniques described here have been applied to other kinds of theories.

Perhaps the most difficult aspect of modifying gravity is the following. Observations indicate that the *true* model of the universe should be close to ΛCDM, at least from the epoch of the primordial ("Big-Bang") Nucleosynthesis until now. In particular the expansion history should be close to that of ΛCDM, perhaps with small differences with the aim of, for example, resolving the Hubble tension.[4] However, as a rule of thumb, MG models that slightly depart from ΛCDM at the background level, become quite different at the perturbative level.[5,6] That is, two models that have almost the same expansion history, typically predict a completely different clustering. This is because in MG the two scalar gravitational potentials of the metric, thinking of in Newtonian gauge, are different even in the absence of anisotropic stresses, which makes GR a very special theory and not easy to mimic. This would mean that the gravitational potential appearing in the geodesic potential is not the same to the gravitational potential sourced by the energy density fields in the Poisson equation; and what is particularly important to this discussion is that the homogeneous and isotropic expansion is completely insensitive to this. For example, take a look at figure 1, where we show the power spectrum of matter density fluctuations in GR and in the F5 gravitational model (we will define what is F5 in appendix A). The F5 gravitational model is indistinguishable for any practical purpose from GR at the background level, but their predicted matter power spectra differ by a lot. This is the reason why it does not make too much sense to test MG models at the background level only, for example using supernovae data only. However, for the sake of fairness we must say there are exceptions to this rule, for example the recently proposed No-Slip gravity,[7] tailored with the specific purpose that the two scalar gravitational potentials become equal.

When gravity is modified, in particular in the infrared, one should be careful to not spoil the well-tested regimes of validity of GR, as for example in the solar system or regions with much larger densities,[a] but at the same time provide the accelerated expansion of the Universe. In order to deal with this unquestionable observational constriction, MG models often invoke mechanisms that yield the theory to GR in the appropriate limits. These are generically called *screening* mechanisms: nonlinear effects that drive MG theories to GR when either the environmental energy

[a]It is perhaps surprising to consider the solar system a high energy density region, but it is very large compared to the mean background energy density $3H_0^2/8\pi G \sim 10^{-29}\,\mathrm{g/cm^3}$. The latter is the one that is important at the largest cosmological scales, and hence drives the background history of the Universe.

density is large, or there are large gravitational tidal fields, among other scenarios. The most popular is the *chameleon* mechanism,[2,3] in which a scalar field mediates a universal fifth force, but its mass depends on the environmental density where it resides: as larger is the ambient density, the lighter is the mass of the scalar field and smaller the range of the force. For example, in successful theories, and for human-like densities ($\sim$ 1 gram per cubic centimeter) the range of the fifth-force becomes so ridiculously small that it becomes effectively invisible to any experiment. On the other hand, at extremely low ambient densities the range of the fifth-force becomes quite large, with cosmological scope.

The most studied MG theory in Cosmology, one can affirm with high confidence, is the $f(R)$ gravity,[8–10] and in particular the Hu-Sawicki model.[11] Although, these theories were first understood as higher-order gravity theories, it was soon recognized that under a few assumptions they can be described as second order differential equations for the metric plus a scalar field[12,13] which propagates only an additional massive mode. In this proceeding note we will lead with the Hu-Sawicki model to exemplify our analytical results and to take advantage of the fact that we have good, *state-of-the-part* N-body simulations for them—the Extended LEnsing PHysics using ANalaytic ray Tracing `Elephant` cosmological suite of simulations.[14,15] Even though $f(R)$ is not a representative model of the vast universe of MG models (or more narrowly of Hordenski theories), the techniques reviewed here can be applied to different models and can be beneficial to study MG that rely on screening mechanisms. The material presented in this note is extracted mainly from Refs. 16–19; we refer the reader to these works for further details.

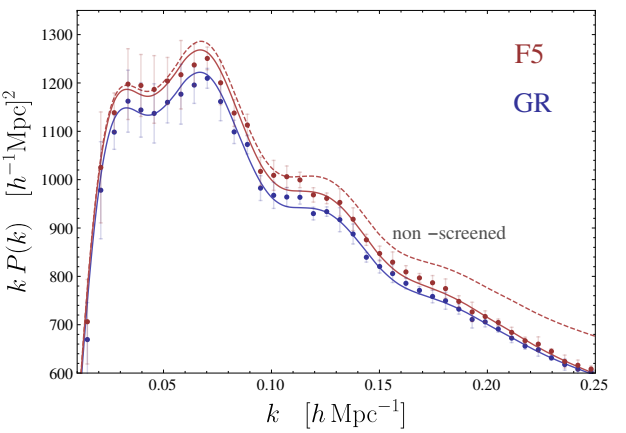


Fig. 1. Matter power spectrum in the models GR (blue line) and Hu-Sawicki F5 (red lines). The latter is further shown with and without screenings with solid and dashed lines, respectively. To generate this analytical results we use the Standard Perturbation Theory with effective field theory counterterms and infrared-resummations of Refs. 20, 21. The mock data is extracted from the `Elephant` simulations.

2. The importance of up-weight low energy density regions

In figure 1 we show the chameleon screening in action. The black line is the analytical power spectrum of the ΛCDM. The red coloured lines show the matter power spectrum in the Hu-Sawicki F5 model with screenings (solid line) and without screenings (dashed line), together with data from the `Elephant` simulations. We note the screening mechanism drives the theory to GR, more evident at small scales (high wave-numbers k), while without the screenings the differences between the two models are larger. This behaviour raises the question if the power spectrum, and its counterpart in configuration space, the correlation function, are the most clever way to test the gravitational theories from its cosmological clustering. After all, by construction these statistics give more weight to high density regions, since they sum over all tracers evenly (either particles, halos, galaxies...), and tracers tend to clump together by gravitational attraction. Meanwhile, cosmological voids are underrepresented by the power spectrum—as well as by the correlation function and their N-points generalizations—even though they are the largest regions in our universe, and as such, they contain a significant share of the total energy budget. But clumpy regions of space, being denser, behave very close to GR due to the screenings mechanism. Hence, the standard statistics used in Cosmology may not be the best option to test infrared modifications of GR.

With the above ideas in mind, M. White proposed in Ref. 16 to use *marked correlation functions*[b] to test gravity at cosmological scales. The proposal consists in weighting the tracer field density fluctuations with a function (the mark) whose value depends on the environmental density, such that objects residing in low density regions become up-weighted in 2-point statistics, while objects residing in high density become down-weighted. In this way, statistics of such marked density fields will probe low density regions where gravitational fifth-force screenings are inefficient and the effects of hypothetical MG theories would be more pronounced.

The construction of a marked density field $\delta_M(\mathbf{x}, t)$ is a three step process, depicted in figure 2. It goes as follows,

(1) One starts with the matter fluctuations field $\delta(\mathbf{x}, t)$ and convolve it with a window function $W_R(\mathbf{x})$ of width R (a Gaussian is used in this work) to obtain the smoothed density

$$\delta_R(\mathbf{x}, t) = \int d^3y \; W_R(|\mathbf{x} - \mathbf{y}|)\delta(\mathbf{y}, t). \tag{1}$$

The scale R is chosen from the beginning and it determines the size of the region which is used to define the environmental density of the tracers. In this sense,

[b]Marked statistics have a long history;[22] they have been used to assign properties to objects, such as the luminosity, color, and morphology of galaxies,[23,24] and to break degeneracies between Halo Occupation Distribution (HOD; a semi-analytical method to populate halos from N-body simulations with galaxies) and cosmologies.[25]

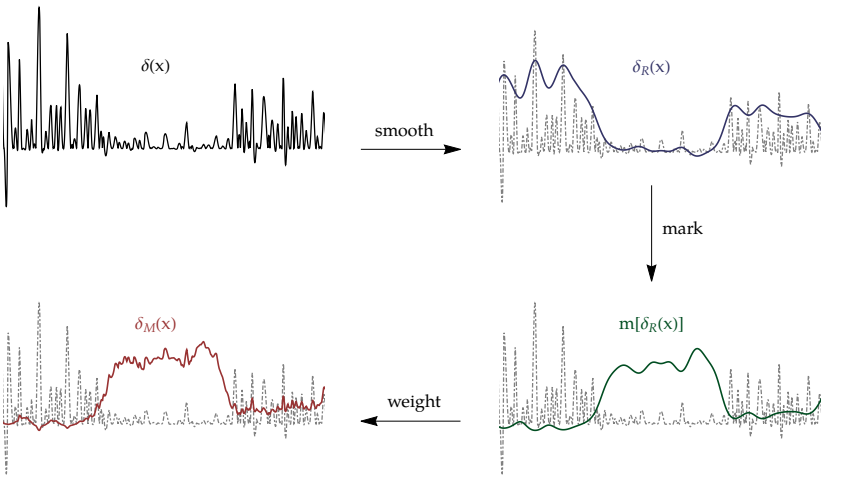


Fig. 2. 1-dimensional cartoon for the construction of the marked density field.

it is good to have sufficiently large values for R, typically comparable to, or a little bit smaller than, the size of cosmic voids. Here we use $R = 10\,h^{-1}\,\mathrm{Mpc}$.

(2) Construct the mark function $m(\delta_R(\mathbf{x}))$. We choose a function that decays with the environmental (or smoothed) density, such that low density regions are up-weighted, as required. For example, a simple exponential $m = \exp[-\alpha(1 + \delta_R(\mathbf{x}))]$ can do the job. However it is better to have more free parameters such that we have more liberty to find an optimal mark for a particular MG model. The most used weight is the White-mark[16]

$$m(\delta(\mathbf{x}), R) = \left(\frac{1+\delta_*}{1+\delta_*+\delta_R(\mathbf{x})}\right)^p, \tag{2}$$

with $p > 0$ and $\delta_* > 0$ dimensional parameters. It was find with numerical simulations that the choices $p = 10$ and $\delta_* = 4$ maximize both the Fisher information and the signal-to-noise.[26] However, these parameters are not good for perturbative analysis, so we will use the parameters $p = 7$ and $\delta_* = 10$, unless otherwise is stated. Other authors working with simulations have used different parameters and mark functions.[18,26–28]

(3) Weight the density field of tracers $1+\delta_X(\mathbf{x})$ with the mark function $m(\delta(\mathbf{x}), R)$ to obtain the marked density field

$$1 + \delta_M(\mathbf{x}, t) = \big(1 + \delta_X(\mathbf{x}, t)\big) m[\delta_R(\mathbf{x})]. \tag{3}$$

The density field can be that of any tracer of the underlying matter density, it can even be the matter density field itself. But we observe tracers, as galaxies for example, in the sky. These are related through a biasing function F to the dark matter as

$$1 + \delta_X(\mathbf{x}) = F_{\mathbf{x}}[\delta(\mathbf{x}), \nabla^2\delta(\mathbf{x}), ...; \mathbf{x}] \tag{4}$$

where the stochasticity of this relation is emphasized by writing $\mathbf{x}$ as an additional argument. The dependence on the non-local curvature operator $\nabla^2\delta(\mathbf{x})$ is expected to be small in GR, but a well defined bias expansion within chameleon theories must contain higher-order derivative operators.[29,30]

3. Marked correlation function

A marked correlation function (mCF)[16,22–25,31,32] is defined as the sum of pairs of objects (i, j) separated by a distance r, and weighted by the ratio of the mark function value to the mean mark $m_i/\bar{m}$ at each point and divided by the number of pairs $n(r)$,

$$\mathcal{M}(r) = \sum_{i,j|r_{ij}=r} \frac{m_i m_j}{n(r)\bar{m}^2}, \tag{5}$$

so it quantifies the deviation of a pair of objects to have the mean mark $\bar{m}$ subjected to be located at a separation r. By construction it can be written as

$$\mathcal{M}(r) = \frac{1 + W(r)}{1 + \xi(r)} \tag{6}$$

where $\xi(r)$ is the standard 2-point correlation function of tracers

$$1 + \xi(r) = \left\langle \big(1 + \delta_X(\mathbf{x} + \mathbf{r})\big)\big(1 + \delta_X(\mathbf{x})\big) \right\rangle, \tag{7}$$

and $W(r)$ is the 2-point correlation function of marked fields:

$$1 + W(r) = \frac{1}{\bar{m}^2} \left\langle \big(1 + \delta_M(\mathbf{x} + \mathbf{r})\big)\big(1 + \delta_M(\mathbf{x})\big) \right\rangle. \tag{8}$$

At large scales both ξ and W are small and one gets

$$\mathcal{M}(r) = 1 + W(r) - \xi(r). \tag{9}$$

Now, one may show that $\xi(r) \in W(r)$,[17] hence the mark function becomes a 2-point statistic of the clustering of marks with the clustering of objects effectively factorized out. The presence of the mean mark is not necessary to estimate the mCF, but it plays an important role in the analytical renormalization of the theory.

4. Perturbative treatment of the marked correlation function

It is well known that a good modeling of N-point statistics of density fields in configuration space requires the use of Lagrangian Perturbation Theory (LPT),[33] since the spread and degradation of the BAO oscillations is due to large scale bulk flows, long wave-length Lagrangian displacement fields that are well described in the Lagrangian framework, even at the linear order.[34] However the derivation of the marked correlation function within the Lagrangian framework is quite large and technical, so we prefer to omit it in this note; the interested reader can find it in Ref. 17, and later on we will display the main results. However, in this section we develop the Standard Perturbation Theory (SPT) approach to the marked correlation function, which is easier, but yet shows the main ingredients and insights of the theory.

We start by expanding the mark in a power series of δ_R

$$m(\delta_R; C_i) = C_0 + C_1\delta_R + \frac{1}{2}C_2\delta_R^2 + \cdots \tag{10}$$

with $C_i = m^{(i)}[0]$, the i-th derivative of $m[\delta_R]$ evaluated at $\delta_R = 0$. The fewer parameters are needed to model the mark, the better the convergence will be. We notice that $C_1 < 0$ enhances low density regions.

For the sake of simplicity we will assume a local relation between matter and tracer overdensities

$$1 + \delta_X(\mathbf{x}) = c_0 + c_1\delta(\mathbf{x}) + \frac{c_2}{2}\delta^2(\mathbf{x}) + \cdots, \tag{11}$$

with $c_i = \partial F_{\mathbf{x}}^{(i)}[\delta_R = 0]/\partial\,\delta_R$ the bare bias parameters and $F_{\mathbf{x}}$ given by eq. (4). A Taylor expansion is an unnatural and restrictive assumption for biased tracers, but being pragmatic we note that it works at the level of precision required by our simulations.

The mean mark is given by the mark weighted by the tracer density field

$$\begin{aligned}\bar{m} &= \langle m[\delta_R(\mathbf{x})](1 + \delta_X(\mathbf{x}))\rangle \\ &= \left\langle \left(C_0 + C_1\delta_R(\mathbf{x}) + \frac{C_2}{2}\delta_R^2(\mathbf{x}) + \cdots\right)\left(c_0 + c_1\delta(\mathbf{x}) + \frac{c_2}{2}\delta^2(\mathbf{x}) + \cdots\right)\right\rangle \\ &= c_0C_0 + c_1C_1\sigma_R^2 + \frac{1}{2}c_0C_2\sigma_{RR}^2 + \frac{1}{2}C_0c_2\sigma^2 + \cdots,\end{aligned} \tag{12}$$

with zero-lag correlators defined as

$$\sigma^2 \equiv \langle(\delta(0))^2\rangle, \quad \sigma_R^2 \equiv \langle\delta_R(0)\delta(0)\rangle, \quad \sigma_{RR}^2 \equiv \langle(\delta_R(0))^2\rangle. \tag{13}$$

We will use also the correlation and cross-correlation functions

$$\xi(r) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle, \tag{14}$$

$$\xi_R(r) \equiv \langle\delta_R(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle, \tag{15}$$

$$\xi_{RR}(r) \equiv \langle\delta_R(\mathbf{x})\delta_R(\mathbf{x}+\mathbf{r})\rangle. \tag{16}$$

We have to write eq. (12) in terms of renormalized bias parameters,[35]

$$b_n^E = \int \frac{d\lambda}{2\pi} e^{-\lambda^2\sigma^2/2} \tilde{F}_{\mathbf{x}}(\lambda)(i\lambda)^n, \tag{17}$$

where $\tilde{F}_{\mathbf{x}}(\lambda)$ is the Fourier transform of $F_{\mathbf{x}}$ with spectral parameter λ (dual to δ). We are using the label "E" to distinguish Eulerian from Lagrangian biases. Equation (17) leads to the relation between bare bias c_n and renormalized bias b_n^E parameters:[36,37]

$$b_n^E = \sum_{k=0}^{\infty} \frac{\sigma^{2k}}{2^k k!} c_{n+2k}. \tag{18}$$

Analogously, we introduce the "resummed" expansion parameters B_n as

$$B_n = \frac{B_n^*}{B_0^*} \qquad \text{with} \qquad B_n^* = \int \frac{d\Lambda}{2\pi} e^{-\Lambda^2\sigma_{RR}^2/2} \tilde{m}(\Lambda)(i\Lambda)^n, \tag{19}$$

where $\tilde{m}(\Lambda)$ is the Fourier transform of $m(\delta_R)$, and Λ is a spectral parameter, dual to δ_R. One easily finds[17] the following relation between the expansion parameters B_n and the Taylor coefficients of the mark function C_n

$$B_n(C_n, \sigma_{RR}^2) = \frac{\sum_{k=0}^{\infty} C_{n+2k}\sigma_{RR}^{2k}/(2^k k!)}{\sum_{k=0}^{\infty} C_{2k}\sigma_{RR}^{2k}/(2^k k!)}. \tag{20}$$

Inserting the b_n^E and B_n parameters in eq. (12) we get

$$\bar{m} = B_0^* \big[1 + b_1^E B_1 \sigma_R^2 + \cdots\big]. \tag{21}$$

We compute now

$$\begin{aligned}
\bar{m}^2(1+W(r)) &= \langle m[\delta_R(\mathbf{x}_1)]\big(1+\delta_X(\mathbf{x}_1)\big) m[\delta_R(\mathbf{x}_2)]\big(1+\delta_X(\mathbf{x}_2)\big)\rangle \\
&= \int \frac{d\lambda_1 d\lambda_2 d\Lambda_1 d\Lambda_2}{(2\pi)^4} \langle e^{i(\lambda_1\delta_1 + \Lambda_1\delta_{R,1} + \lambda_2\delta_2 + \Lambda_2\delta_{R,2})}\rangle \tilde{F}_{\mathbf{x}}(\lambda_1)\tilde{F}_{\mathbf{x}}(\lambda_2)\tilde{m}(\Lambda_1)\tilde{m}(\Lambda_2) \\
&= \int \frac{d\lambda_1 d\lambda_2 d\Lambda_1 d\Lambda_2}{(2\pi)^4} \tilde{F}_{\mathbf{x}}(\lambda_1)\tilde{F}_{\mathbf{x}}(\lambda_2)\tilde{m}(\Lambda_1)\tilde{m}(\Lambda_2) e^{-\frac{1}{2}(\lambda_1^2+\lambda_2^2)\sigma^2 - \frac{1}{2}(\Lambda_1^2+\Lambda_2^2)\sigma_{RR}^2} \\
&\quad \times \Big[1 - (\lambda_1\Lambda_1 + \lambda_2\Lambda_2)\sigma_R^2 - \lambda_1\lambda_2\xi(r) - \Lambda_1\Lambda_2\xi_{RR}(r) - (\lambda_1\Lambda_2 + \lambda_2\Lambda_1)\xi_R(r) + \cdots\Big] \\
&= \bar{m}^2\Big[1 + (b_1^E)^2\xi(r) + B_1^2\xi_{RR}(r) + 2b_1^E B_1 \xi_R(r) + \cdots\Big],
\end{aligned} \tag{22}$$

where the ellipsis denotes second order terms in $\xi_{,R,RR}(r)$. In the second equality, we have shifted to Fourier space ($\delta \to \lambda$, $\delta_R \to \Lambda$) and in the third equality we used the cumulant expansion theorem,

$$\langle e^{iX}\rangle = \exp\left[\sum_{N=1}^{\infty} \frac{i^N}{N!}\langle X^N\rangle_c\right], \tag{23}$$

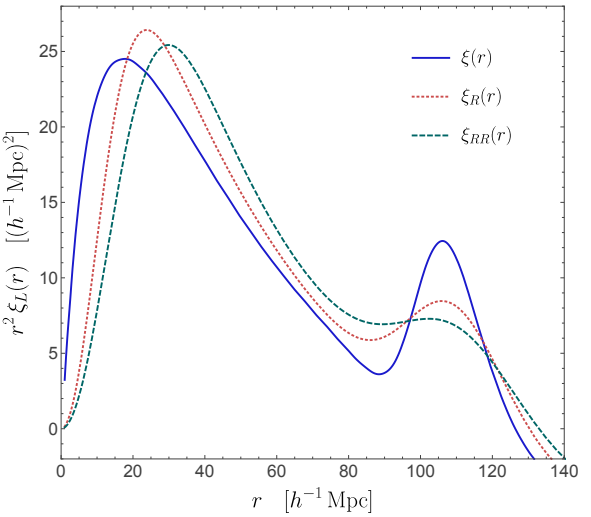


Fig. 3. Smoothed correlation functions defined in defined in eqs. (14)-(16) and appearing in the Eulerian marked correlation function [Eq. (24)]. We use a Gaussian damping kernel W_R of width $R = 10\,h^{-1}$Mpc to smooth the density fields.

and expanded out of the exponential all terms but those containing σ^2 and σ^2_{RR}, such that we can use Eqs. (17) and (19) to get biases from spectral parameters, as we did in the last equality. The mCF becomes

$$\mathcal{M}^E(r) = \frac{1+W(r)}{1+\xi_X(r)} = \frac{1+(b_1^E)^2\xi(r)+B_1^2\xi_{RR}(r)+2b_1^E B_1\xi_R(r)+\cdots}{1+(b_1^E)^2\xi(r)+\cdots}. \qquad (24)$$

In figure 3 we show the correlation functions defined in eqs. (14)-(16) and used to construct the above mCF. We use a Gaussian smoothing kernel W_R with width $R = 10\,h^{-1}$Mpc.

We notice that the zero-lag correlators σ^2 and σ^2_{RR} do not appear in the mCF as is guaranteed because we are using renormalized b and B parameters. Meanwhile, the cross-covariance σ^2_R is canceled out by the mean mark squared appearing in the definition of the mCF in eq. (5). We remark that a process of renormalization of the Taylor expansion coefficients C is not strictly necessary because the scale R is physical, and chosen from the beginning by the observer to mark the tracers, so there is no impediment for σ^2_{RR} to appear in the final expressions. However, the use of parameters B instead of the Taylor coefficients leads to simpler final equations. Furthermore, in real applications, we have to use a mark computed by the number density of galaxies. Another advantage of using the renormalized B parameters is that the effect of this reassignment can be included simply by re-scaling the B parameters. For the application to dark matter halos and galaxies, one should treat them as free parameters and fit them with simulations.

5. Lagrangian space

In Lagrangian space one considers regions of space, at an initial early time with spatial coordinates $\mathbf{q}$. We relate matter and tracers overdensities, in an analogous way to eq. (4), by

$$1 + \delta_X(\mathbf{q}) = F[\delta(\mathbf{q})], \tag{25}$$

with the difference that the overdensity fields are taken at an initial early time.[c] Assuming number conservation of tracers, $(1 + \delta_X(\mathbf{x}))d^3x = (1 + \delta_X(\mathbf{q}))d^3q$, one can relate the function $F_{\mathbf{x}}$, introduced in eq. (4), and F through

$$F_{\mathbf{x}}[\delta(\mathbf{x})] = \int \frac{d^3k}{(2\pi)^3} \int d^3q\, e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{q})} \int \frac{d\lambda}{2\pi} \tilde{F}(\lambda) e^{i\lambda\delta(\mathbf{q}) - i\mathbf{k}\cdot\mathbf{\Psi}(\mathbf{q},\mathbf{t})}, \tag{26}$$

with $\mathbf{\Psi}(\mathbf{q})$ the Lagrangian displacement vector, that maps Lagrangian coordinates $\mathbf{q}$ to Eulerian coordinates $\mathbf{x}$ as $\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, t)$. Equivalently to eq. (17), we introduce the renormalized Lagrangian local biases[35] with

$$b_n = \int \frac{d\lambda}{2\pi} e^{-\lambda^2\sigma^2/2} \tilde{F}(\lambda)(i\lambda)^n. \tag{27}$$

The next step is to evolve initially biased tracers with overdensity $\delta_X(\mathbf{q})$ using Convolution Lagrangian perturbation theory[38,39] and thereafter assign them a mark $m[\delta_R(\mathbf{x})]$ at the moment of observation. The use of the Lagrangian approach has some advantages with respect to the Eulerian. In the first place it is well known that the two-point correlation function is poorly modeled within the Eulerian approach, particularly at the BAO peak position; second, the (renormalized) Lagrangian bias parameters are obtained through the peak background split prescription,[40–43] and hence are physically appealing. But the price to pay is that the mCF is cumbersome to compute, since it becomes a double Gaussian convolution:[17]

$$1 + W(r) = \int \frac{d^3q\, e^{-\frac{1}{2}(\mathbf{r}-\mathbf{q})^T\mathbf{A}^{-1}(\mathbf{r}-\mathbf{q})}}{(2\pi)^{3/2}|\mathbf{A}|^{1/2}} \int \frac{d^3Q\, e^{-\frac{1}{2}(\mathbf{R}-\mathbf{Q})^T\mathbf{C}^{-1}(\mathbf{R}-\mathbf{Q})}}{(2\pi)^{3/2}|\mathbf{C}|^{1/2}} \tag{28}$$

$$\times \Big(1 + \mathcal{I}^0(\mathbf{r}, \mathbf{q}, \mathbf{R}, \mathbf{Q}; b_i, B_i)\Big), \tag{29}$$

with $\mathcal{I}^0$ a function, whose expression is quite large and not very illuminating, and contains the information of the mark (through the expansion parameters B_i) and of the tracers (through the bias parameters b_i). It can be found in eq. (B.8) of Ref. 17. The matrices $\mathbf{A}$ and $\mathbf{C}$ are given by

$$A_{ij}(\mathbf{q}) = 2 \int \frac{d^3p}{(2\pi)^3} \big(1 - e^{i\mathbf{p}\cdot\mathbf{q}}\big) \frac{p_i p_j}{p^4} P_L(p), \tag{30}$$

[c]This initial time is taken early enough that all scales of interest are still in the linear regime and one can safely take $\delta(\mathbf{q}) \ll 1$, but late enough that the background universe is already in the Einstein-de Sitter phase.

and

$$C_{ij}(\mathbf{q}) = -\frac{1}{4}A_{ij}(\mathbf{q}) + \frac{1}{6\pi^2}\int_0^\infty P_L(p)dp\ \delta_{ij}, \tag{31}$$

with $P_L(p)$ the linear matter power spectrum. We notice that if $1+\mathcal{I}$ is not a function of $\mathbf{Q}$, it can be pulled out of the $\mathbf{Q}$ integral and due to that C_{ij} depends only on $\mathbf{q}$, the integration over $\mathbf{Q}$ gives 1. This is the case of the "standard" correlation function in Convolution Lagrangian Perturbation Theory (CLPT),[38] reducing the double Gaussian convolution in eq. (28), to a single three-dimensional convolution.

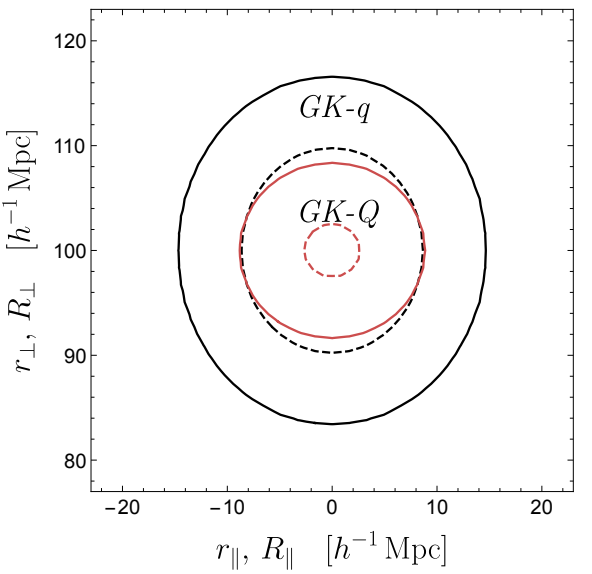


Fig. 4. Gaussian kernels in eq. (28). *GK-q* is the kernel of the q-integral (black curves) and *GK-Q* the kernel of the Q-integral (red curves). Dashed and solid curves show the regions that enclose the 68% and 95% of the volume respectively. A similar plot can be found in Ref. 44. Figure adapted from Ref. 17.

In figure 4 we show level plots for the Gaussian kernels appearing in eq. (28). *GK-q* is the kernel of the q-integral and *GK-Q* the kernel of the Q-integral. These are shown as a function of $r_\parallel$ ($R_\parallel$) and $r_\perp$ ($R_\perp$), the components of $\mathbf{r}$ ($\mathbf{R}$) parallel and perpendicular to the Lagrangian coordinate $\mathbf{q}$ ($\mathbf{Q}$) with $Q = q = 100\,h^{-1}\,\mathrm{Mpc}$ fixed. To have a sense of the meaning of this plot, consider that the kernel *GK-q* is the probability distribution of finding two dark matter particles separated by a distance $\mathbf{r}$ at the redshift of evaluation ($z = 0.5$ in the plot), given that they were separated by a distance $\mathbf{q}$ at the early initial time.[44] We notice that both kernels have their maximum value at $Q = q = 100\,h^{-1}\,\mathrm{Mpc}$, but the *GK-Q* is more sharply

peaked because $|\mathbf{C}| < |\mathbf{A}_L|$. This observation suggests to approximate

$$\frac{e^{-\frac{1}{2}(\mathbf{R}-\mathbf{Q})^T\mathbf{C}^{-1}(\mathbf{R}-\mathbf{Q})}}{(2\pi)^{3/2}|\mathbf{C}|^{1/2}} \approx \delta_{\rm D}(\mathbf{R}-\mathbf{Q}). \tag{32}$$

By doing so, one arrives at

$$1 + W^{\rm W16}(r) = \int \frac{d^3q}{(2\pi)^{3/2}|A_L|^{1/2}} e^{-\frac{1}{2}(\mathbf{r}-\mathbf{q})^T\mathbf{A}_L^{-1}(\mathbf{r}-\mathbf{q})}\Big(1 + J^0(\mathbf{q},\mathbf{r};b_i,B_i)\Big) \tag{33}$$

which is the generalization of the Zeldovich approximation (ZA) result obtained in Ref. 16. The expression for $J^0(\mathbf{q},\mathbf{r};b_i,B_i)$ is given by eq. (4.18) of Ref. 17.

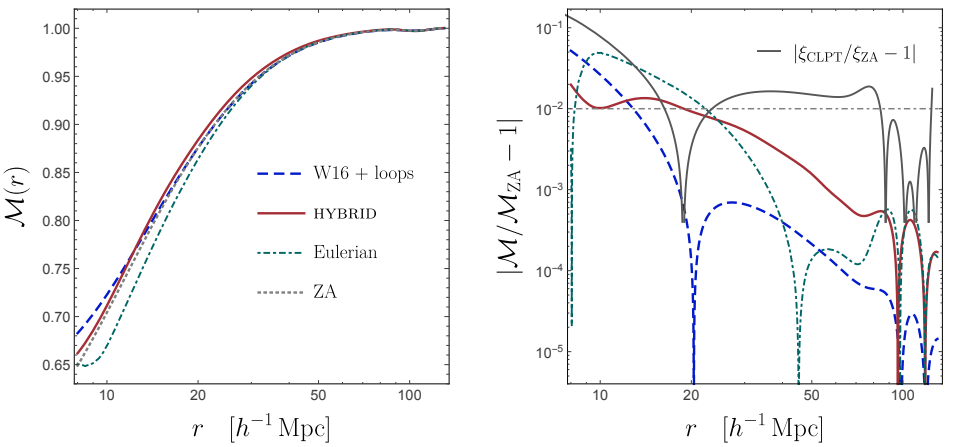


Fig. 5. Marked correlation function modeling comparison. The left panel shows the mCF with parameters $b_1 = 1.2$, $b_2 = 0.2$, $B_1 = -0.7$, $B_2 = B_1^2$, for the different methods: solid red curve is the HYBRID model presented in Ref. 17; dashed blue the method of Eq. (33), which is the method introduced in Ref. 16, plus 1-loop nonlinear contributions; in dotted gray, the ZA; and in dot-dashed green, the Eulerian linear model of eq. (24). The right panel shows the relative differences with respect to the ZA. The gray dot-dashed horizontal line denotes the 1% differences. The solid black curve in the right panel shows the relative differences between the CLPT 1-loop and ZA standard correlation functions. Figure adapted from Ref. 17.

Figure 5 shows the mCF using different analytical methods for the MG Hu-Sawicki F5 model: solid red curve is the one presented above and developed in Ref. 17—called HYBRID from now on, because it evolves the tracers using LPT but marks the tracers using the Eulerian smoothed matter overdensity fields at the moment of observation. The dashed blue is the method of eq. (33), which is that introduced in Ref. 16 with the addition of leading non-linear (loop) corrections; in dotted gray, the Zeldovich Approximation, which in this work means the linear model of Ref. 16; and, in dotdashed green, the Eulerian linear model of eq. (24).

We are using Lagrangian local biases $b_1 = 1.2$ and $b_2 = 0.2$, and mark parameters $B_1 = -0.7$, $B_2 = B_1^2$. In the right panel we show the relative difference between the models and the ZA. The agreement between them is very good, even for the linear Eulerian theory, below 1% for scales above $\sim 20\,h^{-1}\,\mathrm{Mpc}$. The relative differences between the CLPT standard correlation function with and without loop contributions (ξ_{CLPT} and ξ_{ZA}, respectively) is shown in the right panel (black curve); by comparing it with the dashed blue curve (which shows the relative differences of the W16+1-loop and ZA mCFs) we confirm that mCFs which efficiently enhance low density regions in the sky are indeed more linear than the standard correlation function.

6. Comparison to simulations

We use the `Elephant` suite of simulations run with the `ECOSMOG` code.[14] This suite contains five realisations of the initial conditions and for each realisation we have one simulation of ΛCDM (GR) together with three simulations of the Hu-Sawicki $f(R)$ model with parameters $f_{R0} = -10^{-4}$ (F4), $f_{R0} = -10^{-5}$ (F5) and $f_{R0} = -10^{-6}$ (F6). It also contains galaxy mock catalogs that were made with a Halo Occupation Distribution (HOD) method. The HOD parameters for the ΛCDM model are the best-fit parameter values from the CMASS data.[45] For the $f(R)$ models, we tune the HOD parameters so that we reproduce the correlation function in the ΛCDM model. The simulations were ran in a box of size $L = 1024\,h^{-1}\,\mathrm{Mpc}$ with $N = 1024^3$ particles and the cosmological parameters used to make the initial conditions were $\Omega_b = 0.046$, $\Omega_\Lambda = 0.719$, $\Omega_m = 0.281$, $h = 0.697$, $\sigma_8 = 0.82$ and $n_s = 0.971$, and for our analysis we choose snapshots at redshift $z = 0.5$. To compute the mark for each of our tracers (dark matter particles, halos or mock galaxies) we binned the particles/halos/mock galaxies to a grid with gridsize of $20\,h^{-1}\,\mathrm{Mpc}$ (corresponding to a $N = 52^3$ grid) using a Nearest Grid Point assignment scheme to get an estimate for the density for which the mark depends on.

Both, the standard and weighted 2-point correlation functions were computed using the Correlation Utilities and Two-point Estimates (`CUTE`)[46] code.[d] From this the marked correlation function follows simply as $\mathcal{M}(r) = (1 + W(r))/(1 + \xi(r))$.

We consider the White-mark with $\rho_* = 10$, $p = 7$, corresponding to coefficients $C_0 = 1$, $C_1 = -0.64$, $C_2 = 0.46$ in the Taylor expansion of the mark function in eq. (10). In our analytical models we smooth the matter fields that assign the mark with a top-hat filter W_R of radius $R = 10\,h^{-1}\,\mathrm{Mpc}$.

We first confront different analytical methods against matter data, since in this case the mCF does not suffer from biasing-marking degeneracies and we can observe more neatly the effects of the weights. The expansion parameters B are obtained from eq. (20): $B_1^{\mathrm{GR}} = -0.6495$, $B_1^{\mathrm{F6}} = -0.6497$, $B_1^{\mathrm{F5}} = -0.6505$, $B_1^{\mathrm{F4}} = -0.6522$,

[d] https://github.com/damonge/CUTE

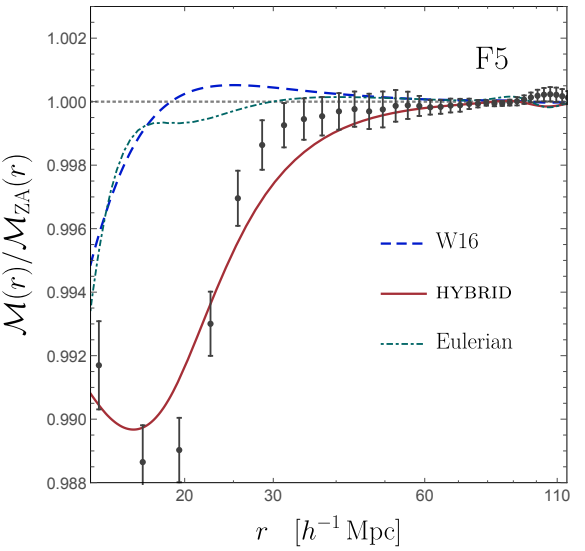


Fig. 6. Ratios of different mCFs analytical models to the ZA mCF for the F5 gravity (we obtain similar results for F4, F6 and GR). We show the W16 model plus 1-loop corrections (dashed blue); the HYBRID model (solid red); and the linear Eulerian model of eq. (24) (dot-dashed green). Figure adapted from Ref. 17.

$B_2^{\rm GR} = 0.4628$, $B_2^{\rm F6} = 0.4502$, $B_2^{\rm F5} = 0.4495$, $B_2^{\rm F4} = 0.4480$; which are numerically very close to C_1 and C_2 values because $C_0 = 1$ and the variances σ^2_{RR} are small.

In figure 6 we show the ratios of the different analytical methods to the ZA model. We do it for the Hu-Sawicki F5 gravity model, and we have checked also the gravity theories GR, F4 and F6, and obtained similar results. The dashed blue curve shows the analytical result with the W16 model plus 1-loop corrections; solid red, the HYBRID method; and dot-dashed green, the linear Eulerian model of eq. (24). The differences between the analytical methods are apparent but small, being lesser than the 1%. At scales $r > 40\,h^{-1}\,{\rm Mpc}$, all analytic models are indistinguishable and they are within the errors of the simulation data. At smaller scales, the method HYBRID outperforms the other perturbative approaches and captures reasonably the trend of the data all the way up to the smoothing scale $R = 10\,h^{-1}\,{\rm Mpc}$.

Finally, in figure 7 we use the HYBRID model to fit the data for halos and galaxies (HOD), finding very good agreement between simulations and theory for $r > 30\,h^{-1}\,{\rm Mpc}$. In the left panel we show that the MG models are differentiated by the data and the analytical curves in the interval $30 < r < 50\,h^{-1}\,{\rm Mpc}$, with the exception of the F5 and F6 models, which become indistinguishable within the error bars. However, we notice that the HOD mCFs for the different MG models do not look differently enough to be distinguished by our data. We think that the use of different parameters δ_* and p or different mark functions can help to discern between different gravitational patterns; see, for example, Ref. 18.

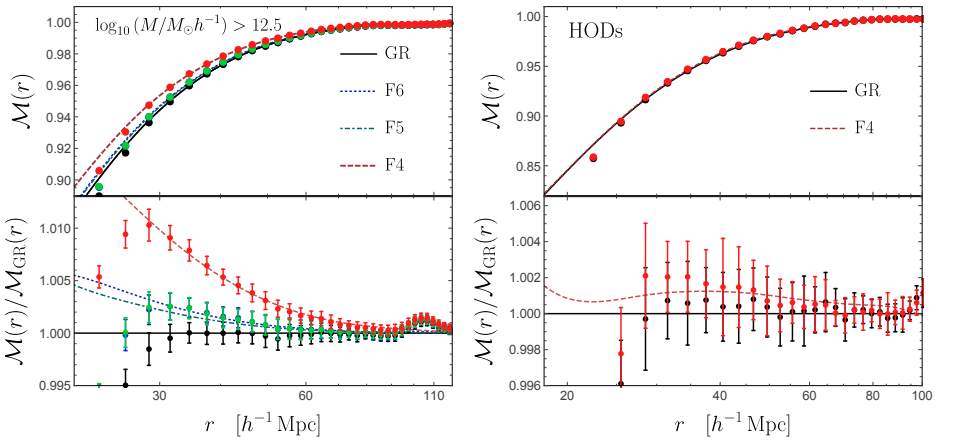


Fig. 7. Halos and galaxies mCFs. Left panel is for halo masses $12.5 < \log_{10}[M/(M_\odot h^{-1})] < 15$ and gravitational models F4, F5, F6 and GR. Right panel is for HOD and shows only the F4 and GR models. The linear local biases are fitted by using the ZA correlation function. Figure adapted from Ref. 17.

7. Marked power spectrum

Recently, it was discovered that the marked statistics can be very useful for the estimation of cosmological parameters, particularly the mass scale of the primordial neutrinos.[47] The reason is that given the large velocity dispersion of neutrinos, they do not contribute to the small scale clustering (see, e.g. Ref. 48). But instead, they free-stream over large distances, allowing them to populate the cosmic voids considerably. On the other hand, the influence of neutrinos on high density regions, for example halos, is quite marginal because their contribution to the total matter density is very small. Thus, statistics that up-weight low density regions can be very beneficial for parameter estimation. In fact, the Fisher formalism results are quite impressive: the marked power spectrum of the matter field outperforms the constraining power of the standard power spectrum by at least a factor of 2 on all cosmological parameters, and in particular of up to 80 times for the sum of the neutrino masses, when a combination of two different mark functions is used.

The analytical description of the marked power spectrum has been studied in detail in a couple of recent works. First, for matter in real space[49] and later for tracers in redshift space.[19] The main complication of such studies is that the mark parameters optimal for improving the cosmological parameter estimation are not able for a perturbative treatment, and the linear theory fails at all scales. The large-scale theory contains non-negligible contributions from all perturbative orders; in Ref. 19 a reorganization of the theory that contains all terms relevant on large-scales is proposed, with an explicit form at one-loop and the following structure at

infinite-loop:

$$\mathcal{M}(k,\mu) = \big[C_0 - C_1 W_R(k)\big]\big[(\tilde{a}_0 + \tilde{a}_2\mu^2)P_L(k) + \tilde{b}_0\big] + \text{EFT} + \text{shot noise}$$

with $\tilde{a}_0$, $\tilde{a}_2$ and $\tilde{b}_0$ free parameters of the theory. EFT refers to the standard 1-loop corrections to the linear power spectrum and the effective field theory terms, while the shot noise includes the stochastic contributions.

8. Summary

We have reviewed the analytical study of marked statistics that up-weight low density regions in the Universe.[16–19] The idea behind marked statistics is to assign a value (the mark) to each entity in a catalogue and perform statistics over the resulting weighted objects. An efficient way to suppress non-linearities is by choosing a mark that gives more weight to objects that reside in low density regions, which become less important as one consider higher density regions. This marking process can be particularly useful for testing general relativity with cosmological probes, since modified gravity models that provide cosmic acceleration, often rely on screening mechanisms to hide its range and strength on high density environments, and the screenings switch off in regions that are more depleted of matter, like cosmic voids. Therefore, the effects of an hypothetical MG are expected to be more pronounced in low density regions. It is then natural to search for such marked statistics, as proposed in Ref. 16. Further, these statistics can potentially improve the constraining power of all cosmological parameters, in particular the mass of the primordial neutrinos, as it was shown in Ref. 47.

Acknowledgments

I want to thank Alfredo Macias and Carlos Herdeiro for inviting me to give a talk in the BS1 parallel session "Scalar fields in Cosmology" of the Sixteenth Marcel Grossmann Meeting – MG16, July 5-10 2021. I also want to thank Jorge L. Cervantes-Cota, Baojiu Li, Kazuya Koyama, Hans Winther, Elisa Massara, Oliver Philcox and Martin White for their collaboration in developing the material that led to this work. I acknowledge support by CONACyT project 283151 and CONACyT Ciencia de Frontera grant No. 102958. I would like to thank the DiRAC Data Centric system at Durham University (`www.dirac.ac.uk`) for computational facilities.

Appendix A. Hu-Sawicki MG model

The Hu-Sawicki model[11] is a particular realisation of $f(R)$ gravity that is able to evade the strong constraints coming from local test of gravity and still give rise to interesting observable signatures on cosmological scales. $f(R)$ theories consist on replace the Einstein-Hilbert Lagrangian density $\sqrt{-g}R$ by a general function of the Ricci scalar $\sqrt{-g}\big(R + f(R)\big)$; see Ref. 50 for a review. The Hu-Sawicki model is

defined by the function

$$f(R) = -M^2 \frac{c_1 (R/M^2)^n}{c_2 (R/M^2)^n + 1} \tag{A.1}$$

where the energy scale is chosen to be $M^2 = \Omega_m H_0^2$. In this parametrized model, at high curvature ($R \ll M^2$) the function $f(R)$ approaches a constant, recovering GR with cosmological constant, while at low curvature it goes to zero, recovering GR; the manner these two behaviors are interpolated is dictated by the free parameters. Given that $d^2 f(R)/dR^2 > 0$ for $R > M^2$, the solutions are stable and the scalar tensor gravity description is possible.[12] In order to be have a similar background evolution than the ΛCDM model it is also necessary that $c_1/c_2 = 6\Omega_\Lambda/\Omega_m$, thus leaving two parameters to fix the model. One choose these parameters to be n and $f_{R0} \equiv df(R)/dR|_{R=R_0}$, with $R_0 = 3H_0^2(\Omega_m + 4\Omega_\Lambda)$ the Ricci scalar of the background cosmological metric evaluated nowadays.

In this work we have considered a fixed value $n = 1$, and hence the model has only one free parameter f_{R0}. Taking this parameter to zero we recover GR. The three choices for the parameters we are considering in this paper (F4, F5 and F6) are such that they lie in the region around where the best constraints lie today. The F4 model corresponds to $|f_{R0}| = 10^{-4}$ (is in tension with local experiments), the F5 model to $|f_{R0}| = 10^{-5}$ (agrees with most experiments and observations, but are in tension with others) and the F6 model to $|f_{R0}| = 10^{-6}$ (which is still allowed). All these models are completely compatible with observations of the expansion history of the Universe, but present large deviations from the ΛCDM in the clustering and formation of structures.

References

1. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters (2018).
2. J. Khoury and A. Weltman, Chameleon fields: Awaiting surprises for tests of gravity in space, *Phys. Rev. Lett.* **93**, p. 171104 (2004).
3. J. Khoury and A. Weltman, Chameleon cosmology, *Phys. Rev.* **D69**, p. 044026 (2004).
4. L. Verde, T. Treu and A. G. Riess, Tensions between the Early and the Late Universe, *Nature Astron.* **3**, p. 891 (7 2019).
5. E. Bertschinger, On the Growth of Perturbations as a Test of Dark Energy, *Astrophys. J.* **648**, 797 (2006).
6. E. Bertschinger and P. Zukin, Distinguishing Modified Gravity from Dark Energy, *Phys. Rev.* **D78**, p. 024015 (2008).
7. E. V. Linder, No Slip Gravity, *JCAP* **03**, p. 005 (2018).
8. A. A. Starobinsky, A New Type of Isotropic Cosmological Models Without Singularity, *Phys. Lett. B* **91**, 99 (1980).
9. S. M. Carroll, V. Duvvuri, M. Trodden and M. S. Turner, Is cosmic speed - up due to new gravitational physics?, *Phys. Rev.* **D70**, p. 043528 (2004).
10. S. Capozziello, Curvature quintessence, *Int. J. Mod. Phys.* **D11**, 483 (2002).
11. W. Hu and I. Sawicki, Models of f(R) Cosmic Acceleration that Evade Solar-System Tests, *Phys. Rev.* **D76**, p. 064004 (2007).

12. G. Magnano and L. M. Sokolowski, On physical equivalence between nonlinear gravity theories and a general relativistic selfgravitating scalar field, *Phys. Rev. D* **50**, 5039 (1994).
13. L. G. Jaime, L. Patino and M. Salgado, Robust approach to f(R) gravity, *Phys. Rev. D* **83**, p. 024039 (2011).
14. B. Li, G.-B. Zhao, R. Teyssier and K. Koyama, ECOSMOG: an Efficient COde for Simulating MOdified Gravity, *JCAP* **2012**, p. 051 (Jan 2012).
15. M. Cautun, E. Paillas, Y.-C. Cai, S. Bose, J. Armijo, B. Li and N. Padilla, The Santiago–Harvard–Edinburgh–Durham void comparison – I. SHEDding light on chameleon gravity tests, *Mon. Not. Roy. Astron. Soc.* **476**, 3195 (2018).
16. M. White, A marked correlation function for constraining modified gravity models, *JCAP* **1611**, p. 057 (2016).
17. A. Aviles, K. Koyama, J. L. Cervantes-Cota, H. A. Winther and B. Li, Marked correlation functions in perturbation theory, *JCAP* **01**, p. 006 (2020).
18. S. Alam, C. Arnold, A. Aviles *et al.*, Testing the theory of gravity with DESI: estimators, predictions and simulation requirements (11 2020).
19. O. H. E. Philcox, A. Aviles and E. Massara, Modeling the Marked Spectrum of Matter and Biased Tracers in Real- and Redshift-Space, *JCAP* **03**, p. 038 (2021).
20. A. Aviles and J. L. Cervantes-Cota, Lagrangian perturbation theory for modified gravity, *Phys. Rev.* **D96**, p. 123526 (2017).
21. A. Aviles, G. Valogiannis, M. A. Rodriguez-Meza, J. L. Cervantes-Cota, B. Li and R. Bean, Redshift space power spectrum beyond Einstein-de Sitter kernels, *JCAP* **04**, p. 039 (2021).
22. C. Beisbart and M. Kerscher, Luminosity- and morphology-dependent clustering of galaxies, *Astrophys. J.* **545**, p. 6 (2000).
23. R. K. Sheth, The halo-model description of marked statistics, *Mon. Not. Roy. Astron. Soc.* **364**, p. 796 (2005).
24. R. K. Sheth, A. J. Connolly and R. Skibba, Marked correlations in galaxy formation models, *Submitted to: Mon. Not. Roy. Astron. Soc.* (2005).
25. M. White and N. Padmanabhan, Breaking Halo Occupation Degeneracies with Marked Statistics, *Mon. Not. Roy. Astron. Soc.* **395**, p. 2381 (2009).
26. G. Valogiannis and R. Bean, Beyond δ: Tailoring marked statistics to reveal modified gravity, *Phys. Rev.* **D97**, p. 023535 (2018).
27. C. Hernández-Aguayo, C. M. Baugh and B. Li, Marked clustering statistics in $f(R)$ gravity cosmologies, *Mon. Not. Roy. Astron. Soc.* **479**, 4824 (2018).
28. J. Armijo, Y.-C. Cai, N. Padilla, B. Li and J. A. Peacock, Testing modified gravity using a marked correlation function, *Mon. Not. Roy. Astron. Soc.* **478**, 3627 (2018).
29. V. Desjacques, D. Jeong and F. Schmidt, Large-Scale Galaxy Bias, *Phys. Rept.* **733**, 1 (2018).
30. A. Aviles, M. A. Rodriguez-Meza, J. De-Santiago and J. L. Cervantes-Cota, Nonlinear evolution of initially biased tracers in modified gravity, *JCAP* **1811**, p. 013 (2018).
31. S. Gottloeber, M. Kerscher, A. V. Kravtsov, A. Faltenbacher, A. Klypin and V. Mueller, Spatial distribution of galactic halos and their merger histories, *Astron. Astrophys.* **387**, p. 778 (2002).
32. R. Skibba, R. K. Sheth, A. J. Connolly and R. Scranton, The luminosity-weighted or 'marked' correlation function, *Mon. Not. Roy. Astron. Soc.* **369**, 68 (2006).
33. M. White, The Zel'dovich approximation, *Mon. Not. Roy. Astron. Soc.* **439**, 3630 (2014).
34. T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, Equivalence Principle and the Baryon Acoustic Peak, *Phys. Rev.* **D92**, p. 043514 (2015).

35. T. Matsubara, Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture, *Phys. Rev.* **D78**, p. 083519 (2008), [Erratum: Phys. Rev.D78,109901(2008)].
36. A. Aviles, Renormalization of Lagrangian bias via spectral parameters, *Phys. Rev.* **D98**, p. 083541 (2018).
37. A. Eggemeier, R. Scoccimarro and R. E. Smith, Bias Loop Corrections to the Galaxy Bispectrum, *Phys. Rev.* **D99**, p. 123514 (2019).
38. J. Carlson, B. Reid and M. White, Convolution Lagrangian perturbation theory for biased tracers, *Mon. Not. Roy. Astron. Soc.* **429**, p. 1674 (2013).
39. Z. Vlah, M. White and A. Aviles, A Lagrangian effective field theory, *JCAP* **1509**, p. 014 (2015).
40. N. Kaiser, On the Spatial correlations of Abell clusters, *Astrophys. J.* **284**, L9 (1984).
41. H. J. Mo, Y. P. Jing and S. D. M. White, High-order correlations of peaks and halos: A Step toward understanding galaxy biasing, *Mon. Not. Roy. Astron. Soc.* **284**, p. 189 (1997).
42. R. K. Sheth and G. Tormen, Large scale bias and the peak background split, *Mon. Not. Roy. Astron. Soc.* **308**, p. 119 (1999).
43. F. Schmidt, D. Jeong and V. Desjacques, Peak-Background Split, Renormalization, and Galaxy Clustering, *Phys. Rev.* **D88**, p. 023515 (2013).
44. S. Tassev, Lagrangian or Eulerian; Real or Fourier? Not All Approaches to Large-Scale Structure Are Created Equal, *JCAP* **1406**, p. 008 (2014).
45. M. Manera, R. Scoccimarro, W. J. Percival, L. Samushia, C. K. McBride, A. J. Ross, R. K. Sheth, M. White, B. A. Reid, A. G. Sánchez, R. de Putter, X. Xu, A. A. Berlind, J. Brinkmann, C. Maraston, B. Nichol, F. Montesano, N. Padmanabhan, R. A. Skibba, R. Tojeiro and B. A. Weaver, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues, *Mon. Not. Roy. Astron. Soc.* **428**, 1036 (Jan 2013).
46. D. Alonso, CUTE solutions for two-point correlation functions from large cosmological datasets, *arXiv e-prints* , p. arXiv:1210.1833 (Oct 2012).
47. E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal and D. N. Spergel, Using the Marked Power Spectrum to Detect the Signature of Neutrinos in Large-Scale Structure, *Phys. Rev. Lett.* **126**, p. 011301 (2021).
48. J. Lesgourgues and S. Pastor, Massive neutrinos and cosmology, *Phys. Rept.* **429**, 307 (2006).
49. O. H. E. Philcox, E. Massara and D. N. Spergel, What does the marked power spectrum measure? Insights from perturbation theory, *Phys. Rev. D* **102**, p. 043516 (2020).
50. K. Koyama, Cosmological Tests of Modified Gravity, *Rept. Prog. Phys.* **79**, p. 046902 (2016).

Dark matter as condensed phase of generic bosons

Elías Castellanos

Mesoamerican Centre for Theoretical Physics
Universidad Autónoma de Chiapas. Ciudad Universitaria,
Carretera Zapata Km. 4, Real del Bosque (Terán), 29040,
Tuxtla Gutiérrez, Chiapas, México
E-mail: ecastellanos@mctp.mx

Jorge Mastache

Consejo Nacional de Ciencia y Tecnología,
Av. Insurgentes Sur 1582, Col Crédito Constructor,
Del. Benito Juárez, CP 03940. Mexico, and

Mesoamerican Centre for Theoretical Physics
Universidad Autónoma de Chiapas. Ciudad Universitaria,
Carretera Zapata Km. 4, Real del Bosque (Terán), 29040,
Tuxtla Gutiérrez, Chiapas, México
E-mail: jhmastache@mctp.mx

We analyze the properties of a generic bosonic cloud that we interpreted as a condensed phase of generic bosons or a Bose–Einstein condensate (BEC) type halo surrounding a Schwarzschild–type black hole. We model the halo as a condensed phase of generic bosons in terms of a massive scalar field that satisfies a self–interacting Klein–Gordon equation. To model the density of particles of the corresponding cloud, we apply the so–called Thomas–Fermi approximation that allows us to extract relevant properties of the system. By using galaxy data from a subsample of SPARC data base, we find the best fits of the BEC model by using the Thomas–Fermi approximation. We show that in the centre of galaxies we must have a supermassive compact central object, i.e., supermassive black hole, additionally the cloud or the halo behaves as a weakly interacting BEC composed of ultralight bosons.

Keywords: Dark matter; Bose–Einstein condensate.

1. Introduction

Bose–Einstein condensates (BECs) play a fascinating role in modern physics, relating many models spreading from microscopic well proved behavior of ultracold quantum gases to galactic and cosmological scales. Nevertheless, there is an issue not well understood in this scenario that deserves more in-depth study, i.e., the non–trivial conditions in which scalar fields can form BECs. However, it seems to be that scalar fields in the form of a BEC formed by generic bosons can describe the basic properties of dark matter (DM) in the universe.[1–5] According to this line of thought, DM consists of a particular type of spin–zero bosons, such as ultra–light scalar field dark matter or fuzzy dark matter, weakly interacting massive particles,

axions, etc., (depending on the specific model under consideration) which have not yet been observed. The bosonic character of these particles, by using the theory of relativistic Bose gases,[6,7] also opens the door for the existence of scalar field dark matter in the form of BECs.[8,9]

The relation between scalar fields and (relativistic) condensates (BEC's) has already a long history, which is non trivial field of study even in flat space time, see for instance[6,7,10–14] and references therein. The aforementioned relation remains a very important topic that must be still fully understood.

Let us remark that in the core of the current astrophysics and cosmology, scalar fields seem to be a relevant candidate to describe dark matter (DM).[1–4,15] Nowadays, the observations suggest that the amount of dark matter in the universe is around 26% of the total energy density. In other words, the existence of dark matter can be found in the following observations, for instance, from the kinematic of galaxies and clusters,[16,17] the physics of the cosmic microwave background radiation (CMBR) and baryon acoustic oscillations (BAO),[18,19] as well as observations from Supernovae Type Ia (SNIa) and Gravitational Lensing (GL).[20,21]

Moreover, dark matter could consist of some type of generic scalar field(–particle) of spin–zero, for instance, Weakly Interacting Massive Particles (WIMPs), axions and others. Although these particles have been not yet observed, scalar fields interpreted as dark matter open up a very interesting model to confront observations. On the other hand, the theory of relativistic bosonic gases and its transitions, under certain circumstances as BEC's open the door to the interpretation of dark matter in the form of a condensate of generic bosonic particles.[8,9,22]

At this point is important to mention that there are some models related to the above ideas in the literature, for instance, the so–called hairy wigs models.[23,24] In the above mentioned approach it is shown that these systems can form stable structures for enough time, make them plausible candidates to describe dark matter galactic halos. Even more, since it is generally accepted that almost all the galaxies host a supermassive black hole at the center, and together with the assumption that dark matter is some kind of scalar field, this leads to the analysis of the existence of bound or quasi–bound scalar field configurations surrounding these compact objects. In fact, in the case of a Schwarzschild black hole and a massive scalar field without self–interactions, it was found that such quasi–bound states exist.[25]

Furthermore, when self–interactions are present for these scalar field configurations surrounding black holes and when quasi–bound states exist, the system can be also analyzed from the Bose–Einstein condensation point of view.[5,26,27] It is quite interesting that in the aforementioned scenario a Gross–Pitaevskii like–equation can be deduced from the corresponding Klein–Gordon equation. In other words, it is possible to study the system by using the formalism behind the Bose–Einstein condensation to extract relevant information. More precisely, in the formalism developed in refs.[5,26,27] it was shown that for a spherically symmetric test–scalar field configurations surrounding different types of black holes, the system can be interpreted as a trapped BEC. In other words, the corresponding Klein–Gordon

equation can be rewritten in a form of a Gross–Pitaevskii equation, which describes the dynamics of usual laboratory BEC's, under certain circumstances. Additionally, in the three quoted references, it was analyzed the limit of validity of the so–called Thomas–Fermi approximation which is capable to give us an approximated solution (in principle, by using a simple algebraic procedure) for the system under consideration.

Here it is important to mention that in the standard theory of condensates (for instance, ref. [28]) the Thomas–Fermi approximation is an algebraic procedure that is useful for exploring relevant properties associated with the system when interactions are present. Moreover, the Thomas–Fermi approximation is valid when the kinetic energy is negligible with respect to the potential and interaction ones. Then, the kinetic energy can be neglected from the very beginning in the corresponding Gross–Pitaevskii equation. As a consequence, the non–linear differential equation becomes an algebraic equation for the order parameter, which in principle, is easy to solve. Additionally, the Thomas–Fermi approximation is valid when the corresponding scattering length, which describes the interaction among the particles within the system, is much smaller than the mean inter–particle spacing for sufficient large clouds. In other words, when the system is diluted enough and contains a large number of particles. Finally, we must add that the Thomas–Fermi approximation fails for trapped condensates near the edge of the cloud, due to the divergent behavior of the kinetic energy (i.e. the total kinetic energy per unit area diverges on the boundary of the system[28]). The last assertion can be used to define the validity of the Thomas–Fermi approximation and sets the limit in which we can extract information from the system. Therefore, in the case of the scalar field configurations viewed as BEC's surrounding black holes, the divergent behavior of the kinetic energy on the predicted boundary of the system can be used to test the region in which the Thomas–Fermi approximation is valid. In other words, we are not able to extract information beyond the region in which the approximation is valid. Deeper research is needed to support the description of these systems as dark matter. In addition, according to the results obtained in refs. ,[5,26] observable features related to the *scalar cloud* within the Thomas–Fermi approximation, for discriminating between different types of black holes are tiny.

The main goal of the present manuscript is to identify the particle density distribution obtained from the Thomas–Fermi approximation (or the BEC density distribution) as a DM density profile surrounding galaxies (i.e., we restrict our analysis to the region in which the Thomas–Fermi approximation is valid).

Galaxy rotation curves cannot be explained by luminous matter alone and we have to appeal to an extra matter component, i.e. DM or BEC–DM in our scenario, to explain the observations.[29,30] Several DM density profiles have been proposed, we can categorize the profiles by its central behavior. Profiles whose densities grow with a power law of $\rho \sim r^{-1}$ are known as cuspy profiles, i.e. Navarro–Frenk–White.[31,32] Profiles whose density tends to a constant value at the center of the galaxy are know as core profiles, i.e. isothermal, Burket profiles.[33,34] Most cuspy profiles come

from numerical simulations while the later are phenomenological proven, but must of them do not offer a clear explanation for the DM fundamental nature. Moreover, there is a tension between cores and cusp profiles because, on one hand, cuspy profiles are the one predicted from numerical simulations, on the other hand, observations seem to prefer cored profiles, the so called core–cusp problem.[35]

We are able to constrain the parameters of the BEC–DM model within the Thomas–Fermi approximation by using a total of 20 high resolution, circular galaxies, see reference.[27] The mass density of the BEC is given in terms of the mass of the bosonic particles M_Φ, a frequency ω, the coupling constant λ (which describes the interparticle interaction within the system) and, the only astronomical parameter, the mass of the black hole M. Except for the mass of the black hole, the profile is only given by the underlying particle model, the BEC indeed, which we expect to be equal for all galaxies, leaving the mass of the black hole as the only astronomical parameter that may vary from one galaxy to other.

2. Rotation Curves and Bose–Einstein Condensates

In this section, we summarize some important results obtained in refs. [5,26,27] In the aforementioned references a Gross–Pitaevskii like–equation is deduced from the corresponding Klein–Gordon equation in spherically symmetric and static black hole spacetimes. Also, a self–interacting scalar potential is assumed, allowing to link the system with the BEC's point of view. It is important to mention that both equations contain the same information related to the system. After the deduction of the Gross–Pitaevskii like–equation we apply the Thomas–Fermi approximation, which allows us to deduce with a very simple procedure, the corresponding density profile that we assume as a galactic dark matter halo.

In order to deduce the density profile we consider a test scalar field(–particle) in a spherically symmetric and static spacetime where the metric which in standard spherical coordinates is given by

$$ds^2 = -f(r)c^2dt^2 + \frac{1}{f(r)}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\,. \tag{1}$$

Thus, the Klein–Gordon equation for a complex test–scalar field Φ with a scalar potential $V(\Phi)$ in a spacetime with metric $g_{\mu\nu}$ can be written as follows:

$$\frac{1}{\sqrt{-g}}\partial_\mu\Big(\sqrt{-g}g^{\mu\nu}\partial_\nu\Phi\Big) - \frac{d\,V(\Phi\Phi^*)}{d\,\Phi^*} = 0, \tag{2}$$

where the star is complex conjugation and g is the determinant of the metric. If we want to link the system with a weakly interacting Bose–Einstein condensate of some generic bosonic particles, we assume a scalar potential of the form

$$V(\Phi\Phi^*) = \mu^2\Phi\Phi^* + \frac{\lambda}{2}(\Phi\Phi^*)^2, \tag{3}$$

where μ is the scalar mass parameter which is related to the mass of the bosonic particles M_Φ through the inverse of the Compton wavelength of the particles $\mu = \dfrac{M_\Phi c}{\hbar}$.

Additionally, λ is the self–interaction coupling constant which is interpreted as the scattering length of the particles.

We can find solutions of the Klein–Gordon equation by using the following ansatz

$$\Phi(t,r) = e^{i\,\omega\,t}\,\frac{u(r)}{r}, \tag{4}$$

where we assume that the frequency ω is *real*. Notice that in general the function $u(r)$ is a complex–valued classical function that can be interpreted as the macroscopical wave function of the system or the order parameter as in standard theory of condensates. By using the ansatz (4) the corresponding Klein–Gordon equation reduces to a Gross–Pitaevskii–like equation of the form

$$\left(-\frac{d^2}{dr_*^2} + V_{\rm eff}(r) + \lambda_{\rm eff}(r)\,\frac{|u(r)|^2}{r^2}\right) u(r) = \frac{\omega^2}{c^2}\,u(r)\,, \tag{5}$$

where we have also introducing the r_* coordinate defined as $d\,r_* = \frac{d\,r}{f(r)}$, i.e., the so–called tortoise coordinate.

Notice that in equation (5) we define the effective trapping potential as

$$V_{\rm eff}(r) = f(r)\,\left(\mu^2 + \frac{f'(r)}{r}\right) \tag{6}$$

where the prime indicates derivatives respect to the r coordinate, together with an effective self–interaction parameter with the following functional form

$$\lambda_{\rm eff}(r) = \lambda\,f(r)\,. \tag{7}$$

Here is important to mention that the effective potential in equation (6) is caused by the curvature of the spacetime itself, together with the contribution of the mass parameter μ. In other words, the effective potential V_{eff} allows the bosonic cloud to admit the existence of quasi–bound states. Additionally, λ is *modulated* by the influence of the spacetime geometry, i.e., the interactions show a position-dependent behavior as was shown in refs.[5,26] Finally, we must mention that the term ω^2/c^2 can be also identified with effective chemical potential.

As was mentioned in the introduction, the Thomas–Fermi approximation assumes that the kinetic energy is negligibly small in comparison to the potential energy and the self-interaction energy. Then, we can neglect the kinetic energy in equation (5) from the very beginning, with the result of an algebraic equation, from which we can obtain the density of particles $\rho(r) \equiv \dfrac{|u(r)|^2}{r^2}$.

The solution for the Gross–Pitaevskii–like equation (5) within the Thomas–Fermi approximation is then given by

$$\frac{|u(r)|^2}{r^2} \equiv \rho(r) = \left(\frac{\omega^2}{c^2} - V_{\rm eff}(r)\right)\,\frac{1}{\lambda_{\rm eff}(r)}\,. \tag{8}$$

Notice that the above equation is well defined as long as the right-hand side is positive. The value of $\rho(r)$ is zero outside the region delimited by the equation

$V_{\text{eff}}(r) = \omega^2/c^2$. Moreover, the equation $V_{\text{eff}}(r) = \omega^2/c^2$ sets the size of the cloud within de Thomas–Fermi approximation. The region in which the condensate lies is a spherical shell of inner radius r_{min} and outer radius r_{max}, where r_{min} and r_{max} are precisely the solutions of the equation $V_{\text{eff}}(r) = \omega^2/c^2$. Additionally, as was point it out in,[26] the Thomas–Fermi approximation becomes arbitrarily good if λN (with N the total number of particles within the condensate) becomes sufficiently big. In the present case, λN increases if the ω/c is chosen very close to the mass parameter μ. Then we expect that the Thomas–Fermi approximation becomes quite good if the corresponding parameters are chosen properly.

For simplicity, let us assume that the function $f(r)$ in the metric (1) in ordinary units is given by:

$$f(r) = \left(1 - \frac{\alpha}{r}\right), \tag{9}$$

where $\alpha \equiv 2GM/c^2$ for the Schwarzschild metric, where G is the gravitational constant and M the mass of the black hole.

Let us remark that the model under consideration comes from *first principles* in the sense that we only assume a scalar distribution in a form of some kind of Bose–Einstein condensate composed of generic bosonic particles surrounding a black hole. Then, we can obtain the density of particles or the density profile that we interpreted as a galactic dark matter halo by using the so–called Thomas–Fermi approximation.

The SPARC catalog have an observing data sample of 175 galaxies from $HI/H\alpha$ studies with large range of luminosities and Hubble types.[36] Our sample consist on galaxies with large number of observational data points, preferably with observations close to the galactic center, their rotation curve is smooth, with no relevant wiggles and extended to large radii, have none or small bulge. The total rotational velocity is computed taken the values for the gas and the stars from the mass model of the SPARC catalog. The bulge and the stellar disk model is given by the catalog.

The mass model include the four main components of a galaxy: the budge (when is present), V_b, the gas disk, V_g, the stellar disk, $V_\star$, and the BEC halo, V_{bec}. The total gravitational potential of the galaxy is the sum of each component of the galaxy, thus the observed rotation velocity is,

$$V_{\text{tot}}^2 = V_g^2 + \Upsilon_b V_b^2 + \Upsilon_\star V_\star^2 + V_{\text{bec}}^2\,, \tag{10}$$

where $\Upsilon_\star$ and Υ_b are the mass-to-light ratio of the star and bulge disk, respectively. We take the mass-to-light ratio of the star, $\Upsilon_\star$, as free parameter. We assume an heuristic relation between the bulge and the stellar disk so $\Upsilon_b = 1.4\Upsilon_\star$ is assumed.[36] Is well know that the value of $\Upsilon_\star$ is model dependent which precise value rely on extinction, star formation history, IMF, among others. We ignore a priori any knowledge of the IMF and treat $\Upsilon_\star$ as an extra free parameter.

The halo model that we are proposing have four free parameters, the mass of the black hole, M, the mass parameter of the scalar field μ, the frequency of the

field, ω and the self–interaction parameter, λ. The mass of the black hole is the only galactic parameter that can change from one galaxy to another.

We assume that galaxies with a central black hole could possibly create enough gravitational potential well in order to create a condensate cloud of almost non–interactive particles that could affect the dynamics of the galaxy. In the present work, we estimate the parameters of the scalar field(–particle) from rotation curves. It is worth pointing out that this in principle is an *universal profile* extracted from a particle model in which only the astronomical parameter is the mass of the black hole, other parameters are inherent of the particle which means should be the same for all galaxies.

The density profile is simple enough in order to compute the mass as function of the radius, the integral over the volume is given by

$$M(r) = \frac{\mu}{\lambda}\left[\alpha^2\omega^2 r + \frac{1}{2}\alpha\omega^2 r^2 - \frac{1}{3}\left(\mu^2 - \omega^2\right) r^3 \right. \\ \left. -\alpha\log(r) + \alpha^3\omega^2\log(r-\alpha)\right] \tag{11}$$

This expression should be evaluated between $r_{\min}$ and $r_{\max}$. Dimensional analysis of the expression for the mass let us conclude that the r^3 term is negligible because $\mu \sim \omega$, therefore the two most relevant terms are the ones proportional to $\log r$ and r^2, that dominates at small and large radii, respectively. And because $\alpha \gg \omega$ we can approximate the mass as,

$$M(r) \approx \frac{\mu\alpha}{\lambda}\left[\frac{1}{2}\omega^2 r^2 - \log(r)\right], \tag{12}$$

from the last expression we clear see that $r_{\min}$ can be approximately computed when $M(r) = 0$, which gives $r_{\min}^2 \approx -\mathcal{W}(-\omega^2)/\omega^2$, where $\mathcal{W}$ is the Lambert function. The value of $r_{\max}$ can also be approximated when the term proportional to r^3 gets bigger to the r^2 term in equation (11), which would lead to a negative values for the mass. Thus, $r_{\max} \sim 3\alpha/2(1-\mu^2/\omega^2)$, given the bound we fix for ω, this would lead to $r_{\max} \sim \mathcal{O}(\alpha 10^{14})$. Then, the region where the Thomas–Fermi approximation is valid is given by

$$\frac{3}{2}\frac{\alpha}{1-\mu^2/\omega^2} \gtrsim r \gtrsim \frac{[-\mathcal{W}(-\omega^2)]^{1/2}}{\omega} \tag{13}$$

Notice that if we do not have observation for $r \sim \mathcal{O}(r_{\min})$ then the parameters μ, ω and λ will be degenerate for the fitting, this is, we can find any combination of $\mu\omega^2/\lambda = \text{cte}$ to valid for one galaxy. We can break the degeneracy with α because $r_{\max}$ highly depends on its value.

We assume circular rotation velocities of test particles in the plane of the galaxy and also spherical DM halos and BEC distributions. For this distribution of matter the circular velocity at radius r is given by $V_H^2(r) = GM(r)/r$. From this last expression and equation (12) we see that in the case of the BEC at large radii the velocity grows as $v \propto r^{1/2}$. Our objective is to test the realization of the BEC model,

equation (11), through rotation curves of a sample of high-resolution galaxies. Form Fig. 1 notice that we may have no presence of BEC at the center of the galaxy (i.e. for $r < r_{\min}$), which make sense if we have a central super-massive object as a black hole. Thus, this profile is neither core or cuspy, we may call it null central profile. The extension of the condensate reaches up to $r_{\max}$, which is a finite value close to the observational limit of the galaxy.

Let us now do an analysis of the equations in order to found bounds to the parameters. The frequency ω is particularly important because the region where the Tomas–Fermi approximation is valid is given by the condition $V_{\text{eff}}(r) = \omega^2/c^2$. The region $r_{\max} > r > r_{\min}$ where the condensate could form is given by the roots of the polynomial $(\mu^2 - \omega^2)r^4 - \alpha\mu^2 r^3 - \alpha r + \alpha^2 = 0$. The extreme case where $\omega \to \mu$ gives a region where $r_{\max} \to \infty$. The closer the value of ω is to μ the larger the potential well of the condensate. For instance, assuming a black hole of $M = 10^{13}\mathrm{M}_\odot$ and $\mu = \mu_{\max}$, if ω is just 10% less than μ then $r_{\max} \sim 4$ pc, which will make the model not testable for currents observations. However, if the percentile difference is of the order $\mathcal{O}(10^{-4})$ then $r_{\max} \sim 50$ kpc, enough extension for a Milky Way galaxy size. If we assume that all galaxies are surrounded for a dark matter halo then the frequency parameter is the most constrained of all four parameters.

The maximum radius at which the approximation is valid, at least the maximum observed radius, also depends on the value of ω, the smaller the value of ω the smaller is the region. We find out that if $1 - \mu/\omega < 10^{-5}$ the maximum radius is of the order of $r_{\max} \sim \mathcal{O}(10)$ kpc. Therefore we fix the value of ω such that $1 - \mu/\omega = 10^{-7}$ in order to make sure that the maximum observed galaxy rotation curve is well inside the region where the Thomas–Fermi approximation is valid.

A special case is to compute the maximum value of the mass parameter $\mu_{\max}$, which is defined when $V_{\text{eff}}(r_\wedge) = \mu^2_{\max}$ at the radius $r_\wedge$, from this condition we obtain that $\mu^2_{\max} = f(r_\wedge)/r^2_\wedge$. Using the equation of the corresponding V_{eff} is quite to easy to compute that $f(r_\wedge) = 1/3$. Therefor $r_\wedge = 3GM/c^2$. For instance, for a black hole of $M_{\text{bh}} = 10^{13}\mathrm{M}_\odot$, the radii where the potential is maximum is $r_\wedge \sim 1.4$ pc. Is important to notice that the maximum value of $\mu_{\max}$ is given only in terms of the mass of the black hole.

We take the ansatz $V_{\text{eff}}(r_\vee) = 0.95\,\mu^2_{\min}$ in order to compute $\mu_{\min}$, which range (0.7 – 0.85) $\mu_{\max}$. Taking $V_{\text{eff}}(r_\vee) = \mu^2_{\min}$ makes no physical sense because there will be no potential well where the condensate could form.

The scattering length (or the self–interacting parameter λ) acts as a weight parameter, this should be or the order of $\mathcal{O}(10^{-70})$ 1/m to fit current observational velocities. The smaller the value of λ the higher velocities we can compute, this will also mean that the scalar field is behaving almost as an ideal condensate. Heuristically speaking, we found that the bounds should be the order of $10^{-50} > \lambda[\mathrm{m}^{-1}] > 10^{-90}$.

We use the observed rotation curve, stellar, and gas component as an input for the numerical code, in order to obtain the properties of the BEC. When fitting

Eq. (10) to the observed rotation curves, we apply a non-linear least-squares method to perform the fit, minimizing the residual sum of the χ^2-test. The χ^2-goodness-of-fit test, that tell us how close are the theoretical to the observed values. In general the χ^2-test statistics is of the form:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{V_{\mathrm{obs}_i} - V_{\mathrm{model}_i}(r, \rho_0, r_s, r_c)}{\sigma_i} \right)^2 , \tag{14}$$

where σ is the standard deviation, and n is the number of observations. Comparison of the fits derived can tell us which of the DM models is preferred. More important are the differences between the reduced $\chi^2_{\mathrm{red}} = \chi^2/(n-p-1)$ values, where n is the number of observations and p is the number of fitted parameters. The uncertainties in the rotation velocity are reflected in the uncertainties in the model parameters.

Table 1. Numerical fits for the BEC model by using the SPARC data base. For each Galaxy we reported the values of: Υ relation total mass-to-light ratio on luminosity, the reduced χ^2 in terms of the degree of freedom for the BEC model χ^2/dof, which correspond to 3 and 4 dof, respectively. The small χ^2 values are mainly due to the large data error bars. $\log_{10} M_{\mathrm{BH}}/M_\odot$ is the mass of the black hole given in solar units $M_{BH} = 10^x M_\odot$. The free parameters for the BEC model are: λ and μ. The last column is the $\ln B$ factor in comparison with the NFW model.

Galaxy ID	BEC model in the Thomas–Fermi approximation					$\ln B$
	$\Upsilon_\star^{\mathrm{bec}}$	χ^2/dof	$\log_{10} M_{\mathrm{BH}}/M_\odot$	$\log_{10} \lambda\ \mathrm{pc}^{-1}$	$\mu[\mathrm{pc}^{-1}]$	
IC2574	$0.21^{+1.07}_{-0.21}$	5.37	12.00 ± 0.73	-89.83 ± 0.78	4.06 ± 0.97	4.44
NGC2403	0.68 ± 0.46	14.14	10.64 ± 0.51	-91.88 ± 0.96	4.12 ± 0.30	4.42
NGC2841	0.95 ± 0.16	2.50	11.20 ± 0.10	-91.43 ± 0.17	4.01 ± 1.00	2.23
NGC2903	$0.44^{+1.39}_{-0.44}$	5.75	10.72 ± 0.94	-92.77 ± 0.42	2.34 ± 0.97	1.47
NGC3198	0.73 ± 0.04	2.58	10.96 ± 0.16	-91.30 ± 0.47	4.24 ± 1.00	1.40
NGC3521	0.54 ± 0.10	0.21	11.71 ± 1.35	-89.99 ± 1.49	7.87 ± 0.95	1.61
NGC4559	$0.52^{+2.16}_{-0.52}$	0.24	10.78 ± 0.97	-91.03 ± 0.91	5.89 ± 1.64	1.24
NGC6015	0.94 ± 0.05	7.16	11.08 ± 0.40	-91.35 ± 0.54	3.60 ± 1.03	1.35
NGC6503	0.63 ± 0.10	5.09	10.72 ± 0.34	-91.58 ± 0.51	4.33 ± 0.46	1.10
NGC6946	0.50 ± 0.06	1.91	10.77 ± 0.61	-90.59 ± 0.98	9.79 ± 1.54	2.06
NGC7331	0.38 ± 0.03	0.90	11.05 ± 0.25	-90.36 ± 0.46	10.20 ± 0.94	1.37
NGC7793	$0.71^{+1.18}_{-0.71}$	0.78	10.70 ± 7.54	-90.79 ± 42.99	7.54 ± 4.65	1.39
UGC02953	$0.67^{+1.08}_{-0.67}$	14.86	11.12 ± 1.00	-91.58 ± 1.59	4.01 ± 0.45	1.54
UGC03205	0.25 ± 0.16	5.10	10.78 ± 0.47	-91.32 ± 0.19	3.74 ± 1.19	2.54
UGC03580	0.81 ± 0.07	3.11	10.78 ± 0.47	-91.57 ± 0.84	4.23 ± 0.98	1.92
UGC05253	$0.52^{+0.98}_{-0.52}$	6.56	11.05 ± 0.95	-91.54 ± 1.62	4.30 ± 1.01	1.47
UGC07524	$1.11^{+2.64}_{-1.11}$	0.59	10.54 ± 1.05	-90.66 ± 1.86	8.42 ± 1.02	1.96
UGC08699	0.56 ± 0.01	1.44	10.73 ± 0.25	-92.16 ± 0.34	3.51 ± 0.05	1.29
UGC09133	$0.58^{+0.97}_{0.58}$	23.93	11.35 ± 2.36	-90.65 ± 2.61	5.21 ± 1.60	1.57
UGC11455	0.54 ± 0.08	3.60	11.67 ± 1.12	-89.78 ± 1.24	8.56 ± 1.00	2.95
UGC11914	$0.64^{+0.84}_{-0.64}$	2.16	12.01 ± 0.99	-90.83 ± 2.16	3.96 ± 0.95	0.92

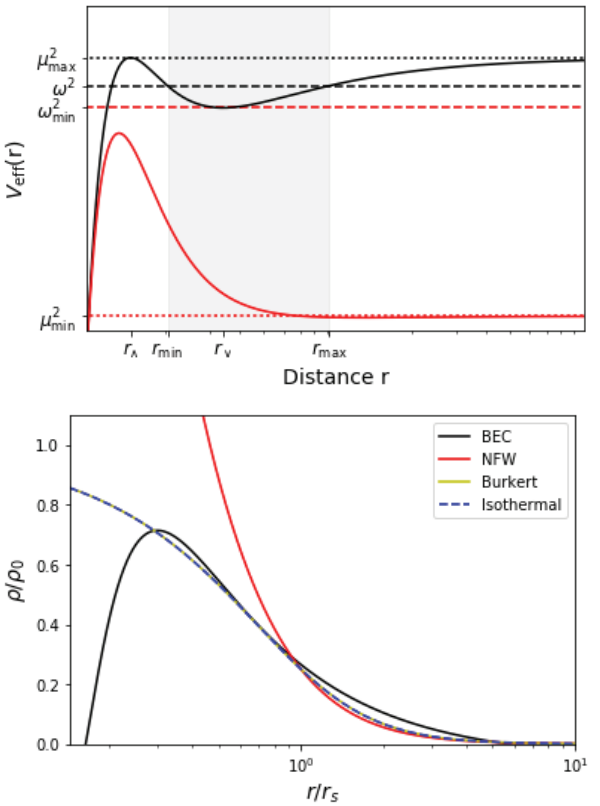


Fig. 1. *Up:* Plot of the potential $V_{\rm eff}$, Eq. (6). Solid lines are two examples for different values of mass parameter μ, the black lines is given $\mu_{\rm max} = V_{\rm eff}(r_\wedge))$ and the red line is given $\mu_{\rm min}$. The ω parameter defines the the region where the Thomas-Fermi approximation is valid, $r_{\rm max} > r > r_{\rm min}$ and could take the value $\omega_{\rm max}(=\mu_{\rm max}) > \omega > \omega_{\rm min}(= V_{\rm eff}(r_\vee))$. *Down:* Plot of the profile density of BEC (solid black), NFW (solid red), Burkert (solid yellow) and isothermal (dashed blue).

3. Discussion

According to the results summarized in the present work, it seems to be that the interpretation of dark matter as some kind of BEC is, in fact, a good model to describe the kinematics of the galaxy rotation curves. Even more, we have shown that the corresponding Thomas–Fermi approximation can give insights in a simple way that match with the observations of rotation curves in galaxies. From the simple analysis, we notice that density is proportional $\rho \propto r^{-1}$ for all radii, in contrast with most DM profiles that behave as $\rho \propto r^{-3}$ at large radii. The mass of the BEC behaves as $m_{\rm bec} \propto r^2$, and therefore $v_{\rm bec} \propto r^{1/2}$, this coincides with NFW in the limit $r \ll r_{\rm max}$ where $v_{\rm nfw} \propto r^{1/2}$, with $r_{\rm max} \sim 2.163 r_s$.

We have obtained an average $\mu = 5.43 \pm 2.24$ pc^{-1} which corresponds to an average boson mass of $M_\Phi = (3.47 \pm 1.43) \times 10^{-23}$ eV. Previous analysis has found

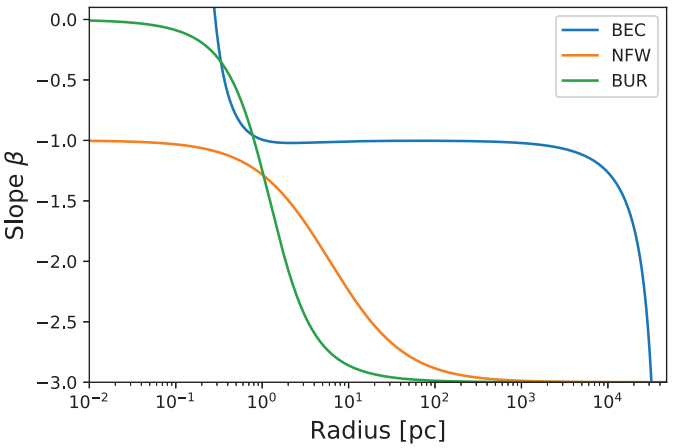


Fig. 2. The slope for different profiles. For BEC (blue line) notice that the density is bounded between some $r_{\min}$ and $r_{\max}$, in this region the slope is almost constant with value $\beta = -1$. For NFW (orange line) we notice its cuspy behavior close to the galactic centre and then dilutes as $\rho \propto r^{-3}$, while core profiles as Burkert (green line) reach constant value in the centre $\beta = 0$ and $\rho \propto r^{-3}$ at large radii.

masses for ultralight dark matter (see[42] for a compilation of results), which constrain the mass of the boson in the order of 10^{-22} eV. Notice that the latter result agrees with the one obtained in the present work. For the self–interacting parameter we get the values $\log_{10}(\lambda\ [\mathrm{pc}^{-1}]) = -91.09 \pm 0.74$. In other words, the halo viewed as a Bose–Einstein condensate can be interpreted almost as an ideal condensate of some generic bosons, which also agree with the results reported in.[26,27]

Thus, the corresponding mass of the black holes at the galaxy centre is given by $\log_{10} M/M_{\odot} = 11.08 \pm 0.43$. This kind of scenarios, e.g Abell 85 and Holm 15A, has been reported in[43,44] with a core that host supermassive black holes of mass 10^9–10^{11} $M_{\odot}$. Although recent observation of supermassive black holes in galaxies supports our fitting for the mass of the black hole, we could also think of the possibility that the black hole may not be the only option to trap the boson cloud. In future works, we could use a weak field approximation to also include the mass of stars, bulge, and gas in the center of the galaxy, even more, the self-gravitating effect of the bosonic halo or the effects caused by the rotation of the halo as a contribution to the gravitational potential and therefore reduce the contribution of the black hole as the only source of the potential. Indeed, the BEC model on average fits quite well and is not a bad fitting at all given the simplicity of the model is something that must be taken into account. The above results are sumarized in Table (1).

The BEC model in the Thomas–Fermi approximation is strong enough as the NFW model to describe galaxy rotation curves. More important, we cannot ignore that the BEC model comes from first principles and that makes use the Thomas–Fermi approximation and assume that is only a compact central object that conden-

sates the relativistic bosonic cloud. This model can be extended and generalized to include bulge and central stars contribution that may diminish the mass of the central black hole, or the–self gravitating nature of the halo and also with rotation (and maybe charged). The good fitting of the BEC model using galaxy rotation curves may be indicating that the nature of dark matter may be close to be unveiled.

The fit in this model helps us to set physical bounds to the parameters of the Bose–Einstein dark matter approach. These bounds can be used to extend the analysis in bigger samples of galaxies. Finally, more general scenarios can be taken into account in order to study the corresponding rotation curves, for instance, in charged and rotating black holes spacetimes and perhaps as discrimination criteria of the nature of supermassive black holes. Even more, we are able to extend also the analysis reported in the present manuscript in order to study the eventual relation between dark matter and the so–called boson stars from the Bose–Einstein condensation point of view, in the relativistic and non–relativistic regime.[39–41]

Acknowledgments

This work was partially supported by CONACyT México under Grant No. 304001. J.M. acknowledges CONACYT Cátedras program and the support of the MCTP/UNACH.

References

1. S. J. Sin, *Late-time phase transition and the galactic halo as a Bose liquid*, *Phys. Rev.* D 50 (1994) 3650.
2. F. S. Guzmán, T. Matos and H. Villegas, *Scalar fields as dark matter in spiral galaxies: comparison with experiments*, *Astron. Nachr.* 320 (1999) 97
3. F. S. Guzmán and T. Matos, *Quintessence at Galactic Level? Class. Quantum Grav.* 17 (2000) L9
4. J. Magaña, T. Matos, *A brief review of the scalar field dark matter model*, *J. Phys. Conf. Ser.* 378 (2012) 012012
5. E. Castellanos, C. Escamilla–Rivera, C. Lämmerzahl and A. Macías, *Scalar field as a Bose-Einstein condensate in a Schwarzschild-de Sitter spacetime*, *Int. J. Mod. Phys.* D 26 (2017) 1750032
6. J. Bernstein and S. Dodelson, *Relativistic Bose gas*, *Phys. Rev. Lett.* 66 (1991) 683
7. L. Parker and Y. Zhang, *Ultrarelativistic Bose-Einstein condensation in the Einstein universe and energy conditions*, *Phys. Rev.* D 44 (1991) 2421
8. C. G. Boehmer and T. Harko, *Can dark matter be a Bose-Einstein condensate? JCAP* (2007) 025
9. L.A. Ureña–López, *Bose-Einstein condensation of relativistic scalar field dark matter*, *JCAP* (2009) 014
10. L. Dolan and R. Jackiw, Phys. Rev. D 9 (1974) 3320.
11. Weinberg, Phys. Rev. D 9 (1974) 3357.
12. E. Castellanos and T. Matos, Int. J. Mod. Phys. B 27 (2013) 11.
13. M. Grether, M. de Llano and G. A. Baker, Jr., Phys. Rev. Lett. 99 (2007) 200406.
14. Tonatiuh Matos, Elías Castellanos and Abril Suárez, *Bose–Einstein Condensation and Symmetry Breaking of a Complex Charged Scalar Field*, Eur. Phys. J. C 77:500 (2017).

15. S. U. Ji and S. J. Sin, *Late-time phase transition and the galactic halo as a Bose liquid. II. The effect of visible matter*, *Phys. Rev.* D50 (1994) 3656
16. F. Zwicky, Helv. Phys. Acta **6**, 110 (1933) [Gen. Rel. Grav. 41 207 (2009)]. doi:10.1007/s10714-008-0707-4
17. V. C. Rubin, N. Thonnard and W. K. Ford, Jr., Astrophys. J. 238 471 (1980). doi:10.1086/158003
18. N. Aghanim *et al.* [Planck Collaboration], arXiv:1807.06209 [astro-ph.CO].
19. D. J. Eisenstein *et al.* [SDSS Collaboration], Astrophys. J. 633 560 (2005) doi:10.1086/466512 [astro-ph/0501171].
20. P. Astier *et al.* [SNLS Collaboration], Astron. Astrophys. 447 31 (2006) doi:10.1051/0004-6361:20054185 [astro-ph/0510447].
21. N. Kaiser and G. Squires, Astrophys. J. 404 441 (1993). doi:10.1086/172297
22. A. Suárez, V. Robles and T. Matos, *A Review on the Scalar Field/Bose-Einstein Condensate Dark Matter Model*, *Astrophys. Space Sci. Proc.* 38 (2014) 107
23. J. Barranco et al., Phys. Rev. Lett. 109 (2012) 081102.
24. N. Sanchis-Gual, J. C. Degollado, P. J. Montero and J. A. Font, *Quasistationary solutions of self-gravitating scalar fields around black holes*, *Phys. Rev.* D 91 (2015) 043005.
25. J. Barranco, A. Bernal, J. C. Degollado, A. Diez-Tejedor, M. Megevand, D. Núñez and O. Sarbach, *Self-gravitating black hole scalar wigs* [arXiv:1704.03450]
26. Elías Castellanos, Juan Carlos Degollado, Claus Lämmerzahl, Alfredo Macías and Volker Perlick, *Bose–Einstein condensates in charged black–hole spacetimes* JCAP 01(2018) 043.
27. Elías Castellanos, Celia Escamilla–Rivera and Jorge Mastache, *Is a Bose–Einstein condensate a good candidate for dark matter? A test with galaxy rotation curves*, Int. J. Mod. Phys. D. 29 9 (2020) 2050063.
28. C. J. Pethick and H. Smith, *Bose–Einstein Condensation in Dilute Gases*, Cambridge University Press, Cambridge (2002)
29. P. Salucci, Astron. Astrophys. Rev. 27 no. 1, 2 (2019) doi:10.1007/s00159-018-0113-1 [arXiv:1811.08843 [astro-ph.GA]].
30. P. Salucci, Found. Phys. 48 no. 10, 1517 (2018) doi:10.1007/s10701-018-0209-5 [arXiv:1807.08541 [astro-ph.GA]].
31. J. F. Navarro, C. S. Frenk and S. D. M. White, Astrophys. J. 462 563 (1996) doi:10.1086/177173 [astro-ph/9508025].
32. B. Moore, T. R. Quinn, F. Governato, J. Stadel and G. Lake, Mon. Not. Roy. Astron. Soc. 310 1147 (1999) doi:10.1046/j.1365-8711.1999.03039.x [astro-ph/9903164].
33. J. E. Gunn and J. R. Gott, III, Astrophys. J. 176 1 (1972). doi:10.1086/151605
34. A. Burkert, IAU Symp. **171**, 175 (1996) [Astrophys. J. 447 L25 (1995)] doi:10.1086/309560 [astro-ph/9504041].
35. W. J. G. de Blok, Adv. Astron. 789293 (2010) doi:10.1155/2010/789293 [arXiv: 0910.3538 [astro-ph.CO]].
36. Federico Lelli et al. 2016 AJ 152 157.
37. F. C. van den Bosch and R. A. Swaters, Mon. Not. Roy. Astron. Soc. 325 1017 (2001) doi:10.1046/j.1365-8711.2001.04456.x [astro-ph/0006048].
38. Elías Castellanos, Guillermo Chacón-Acosta and Jorge Mastache, Non-relativistic Boson Stars as N-Body Quantum Systems, arXiv:2012.03445 [gr-qc] (2020).
39. F. Schunck and E. Mielke, Class. Quantum Grav. 20 (2003) R301–R356.
40. F. Schunck and E. Mielke, General relativistic boson stars, Class. Quantum Grav. 20 (2003) R301-R356.

41. J. Eby, et al., Approximation methods in the study of boson stars, Phys. Rev. D 98
42. P. H. Chavanis, arXiv:1810.08948 [gr-qc].
43. O. Lopez-Cruz, et al. The Brightest Cluster Galaxy in Abell 85: The Largest Core Known so far DOI:10.1088/2041-8205/795/2/L31
44. K. Mehrgan, et al. A 40-billion solar mass black hole in the extreme core of Holm 15A, the central galaxy of Abell 85 arXiv:1907.10608

New Planck tSZ map and its cosmological analysis

H. Tanimura*, M. Douspis and N. Aghanim

Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, Bâtiment 121, 91405 Orsay, France

**E-mail: hideki.tanimura@ias.u-psud.fr*

We constructed a new all-sky Compton parameter map (y-map) of the thermal Sunyaev-Zel'dovich (tSZ) effect by applying the MILCA component separation algorithm to the 100 to 857 GHz frequency channel maps from the Planck data release 4. The Planck team performed several improvements for the new channel maps in terms of noises and systematics, and it allowed us to produce a new y-map with reduced noises by $\sim$7% and minimal survey strips compared to the previous version released in 2015. We computed the tSZ angular power spectrum of the new y-map and performed a cosmological analysis. The results showed $S_8 = 0.764\,^{+0.015}_{-0.018}\,(stat)\,^{+0.031}_{-0.016}\,(sys)$, including systematic uncertainties from a hydrostatic mass bias and pressure profile model. The value is fully consistent with recent KiDS and DES weak-lensing observations. It is also consistent with the Planck CMB's result within 2σ, while our result is slightly lower.

Keywords: Clusters; intracluster medium; large-scale structure of Universe; cosmic background radiation

1. S_8 tension

The current standard cosmological model called the Λ cold dark matter (ΛCDM) model, mainly consisting of dark energy and dark matter, provides a wonderful fit to many cosmological data. However, a slight discrepancy was found in the latest data analyses between the CMB anisotropies ($z \sim 1100$) and other low-redshift ($z \sim 0-1$) cosmological probes for the $S_8(\equiv \sigma_8(\Omega_m/0.3)^{0.5})$ cosmological parameter representing the amplitude of the structure growth in the Universe. For example, the S_8 value was precisely measured to be $S_8 = 0.830 \pm 0.013$ with the Planck CMB observation.[1] However, a lower value of $S_8 \sim 0.77 \pm 0.03$ was measured using the population of galaxy clusters detected by Planck at low redshift ($z < 0.6$),[2,3] thus called "S_8 tension". Furthermore, other low-redshift observations with gravitational lensing by the Kilo-Degree Survey[4] (KiDS) and the Dark Energy Survey[5,6] (DES) experiments also found lower S_8 values. These results indicate that the cosmic structure growth is slower than the prediction based on the CMB measurement and may demand modifications to the standard cosmological model to explain all these measurements.

We use the thermal Sunyaev Zel'dovish (tSZ) effect to investigate the value of the S_8 parameter. The tSZ signals provide independent constraints on cosmological parameters. In particular, it is sensitive to the normalization of the matter

power spectrum, commonly parameterized by σ_8, and to the total matter density, Ω_m. It is also sensitive to other cosmological parameters, e.g., the baryon density parameter Ω_b, the Hubble constant H_0, and the primordial spectral index n_s. However, these variations are small relative to those of Ω_m and σ_8.[7] Thus, in our analysis, we varied only Ω_m and σ_8, and other cosmological parameters were fixed to the fiducial values obtained from the Planck CMB measurements.[1]

2. New all-sky tSZ map

The tSZ signal is subdominant relative to the CMB and other foreground emissions in the Planck bands. Thus a tailored component separation algorithm is required to reconstruct the tSZ map. We adopted the MILCA (Modified Internal Linear Combination Algorithm)[8] used for the Planck y-map reconstruction in 2015 and applied it to the 100 to 857 GHz frequency channel maps from the Planck data release 4 (PR4).

The Planck PR4 data implemented several improvements from the previous version: the usage of foreground polarization priors during the calibration stage to break scanning-induced degeneracies, the correction of bandpass mismatch at all frequencies, and the inclusion of 8% of data collected during repointing maneuvers, etc. These improvements allowed us to produce a new y-map with reduced noises by ∼7% and minimal survey strips than the previous version released in 2015 (Fig. 1).

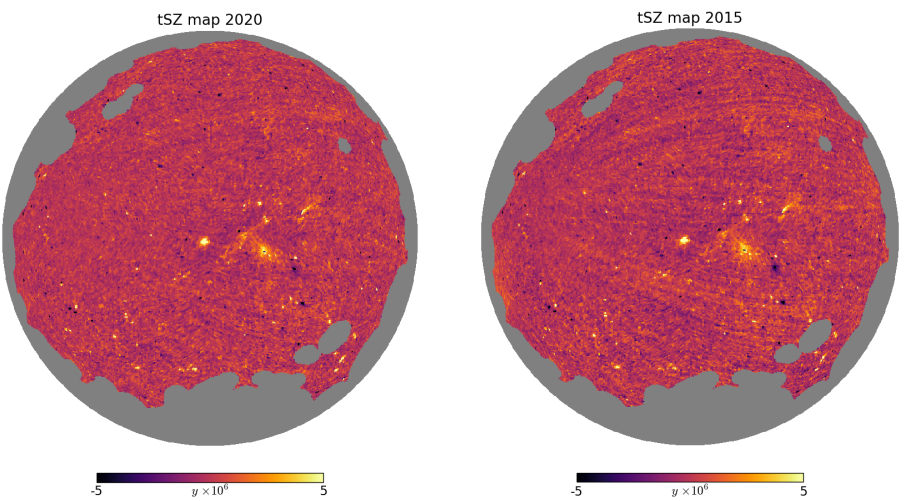


Fig. 1. MILCA all-sky Compton y parameter maps reconstructed in this work (*left*) and 2015[9] (*right*) in orthographic projections. The pixel resolution is changed from Nside = 2048 to Nside = 128 for visualization purposes.

3. Cosmological analysis and result

For our cosmological analyses, we also reconstructed two y-maps from the first and last halves of the data for a cosmological analysis. We considered a cross-angular power spectrum between these y-maps to avoid the bias induced by the noise in the auto-angular power spectrum. This tSZ cross-power spectrum is affected by residual foreground emissions from radio and IR point sources and CIB emission. Thus we fitted our measurement with the cosmology-dependent tSZ model, including the radio and IR point source models[10] and the CIB model[11] (*left* panel in Fig. 2).

The cosmological analysis of the tSZ cross-power spectrum allowed us to set constraints on cosmological parameters, mainly of the S_8 parameter. We obtained $S_8 = 0.764\,^{+0.015}_{-0.018}$. Our obtained S_8 value is fully consistent with recent weak lensing results from KiDS and DES. It is also consistent with the Planck CMB's result[1] within 2σ, while it is slightly lower by $\sim 1.7\sigma$ (*right* panel in Fig. 2).

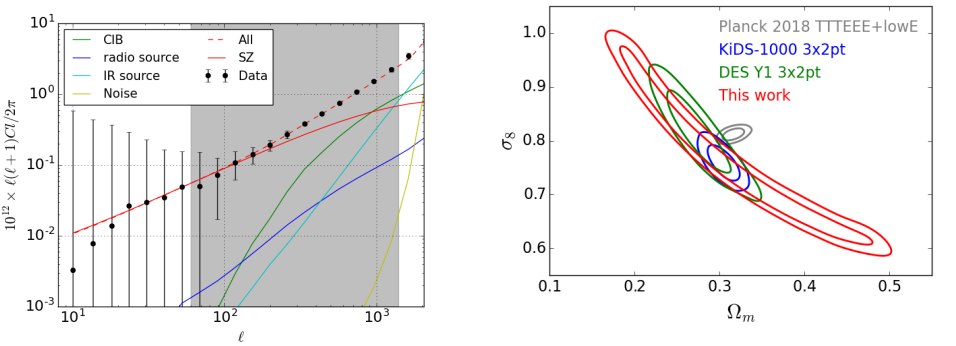


Fig. 2. *Left*: Cross-power spectrum (black) fitted with the tSZ (red), foreground models, and fiducial model of instrumental noise. The considered foregrounds are CIB (green), radio sources (blue), infrared sources (cyan) and noise (yellow). The sum of the tSZ and foreground models is shown in red dashed line. *Right*: Posterior distribution of the cosmological parameters, σ_8 and Ω_m, with 68% and 95% confidence interval contours obtained from our cosmological analysis. It is compared with the Planck CMB's (gray), KiDS-1000 3x2pt (blue), and DES Y1 3x2pt results (green).

4. Systematic effects in the cosmological analysis

Finally, we considered systematic uncertainties in our cosmological analysis introduced by the mass bias and the pressure profile model. First, we consider a systematic uncertainty from the mass bias. To investigate its impact on our cosmological analysis, we replaced our fiducial mass bias (CCCP:[12] $1-b = 0.780 \pm 0.092$) prior with two other mass bias models from WtG[13] ($1-b = 0.688 \pm 0.072$) and BIFFI[14] ($1-b = 0.877 \pm 0.015$). The left panel in Fig. 3 shows their comparison in σ_8 and Ω_m. The result shows that the S_8 value increases as the mass bias increases

(or $1-b$ decrease). This systematic uncertainty may cause a $\sim 1\sigma$ shift in the S_8 parameter from our fiducial value.

In addition, the pressure profile model may affect the cosmological analysis with the tSZ power spectrum. To investigate its impact on our cosmological analysis, we replaced the pressure profile model (P13[15]) with two other models from A10[16] and PACT21[17] . The right panel in Fig. 3 shows their comparison in σ_8 and Ω_m. The PACT21 model reduces the S_8 discrepancy between the Planck CMB and our tSZ results by $\sim 1\sigma$ relative to our fiducial model. It may imply that the discrepancy can be further reconciled or amplified by a better understanding of baryonic physics in galaxy clusters.

In conclusion, we included these systematic uncertainties and obtained $S_8 = 0.764\,^{+0.015}_{-0.018}\,(stat)\,^{+0.031}_{-0.016}\,(sys)$.

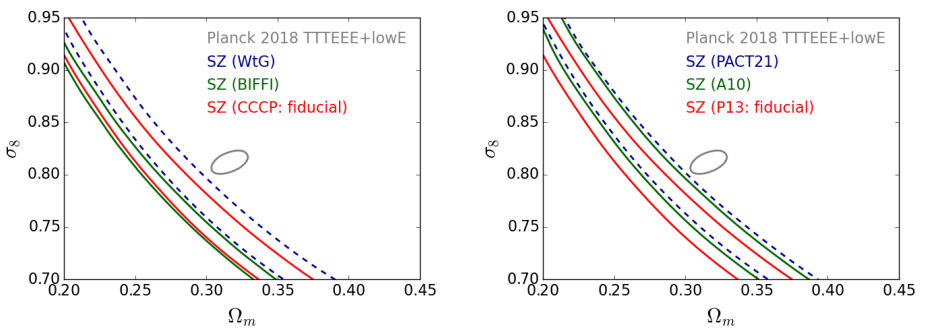


Fig. 3. *Left*: Posterior distribution of the cosmological parameters, σ_8 and Ω_m, with 68% confidence interval contours obtained from the Planck CMB analysis in gray and our tSZ analysis with three different mass bias priors. Our fiducial model of CCCP is shown in red. It is compared with the WtG model in blue and BIFFI model in green. *Right*: Posterior distribution of the cosmological parameters, σ_8 and Ω_m, with 68% confidence interval contours obtained from the Planck CMB analysis in gray and our tSZ analysis with three different pressure profile models. Our fiducial model of PA13 is shown in red. It is compared with the PACT21 model in blue and A10 model in green.

Acknowledgement

This research has been supported by the funding for the ByoPiC project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement ERC-2015-AdG 695561.

References

1. Planck Collaboration, *A&A* **641**, p. A6 (September 2020).
2. Planck Collaboration, *A&A* **571**, p. A20 (November 2014).
3. Planck Collaboration, *A&A* **594**, p. A24 (September 2016).

4. C. Heymans *et al.*, *A&A* **646**, p. A140 (February 2021).
5. Dark Energy Survey Collaboration, *Phys. Rev. D* **98**, p. 043526 (August 2018).
6. DES Collaboration, *arXiv e-prints* , p. arXiv:2105.13549 (May 2021).
7. E. Komatsu and U. Seljak, *MNRAS* **336**, 1256 (November 2002).
8. G. Hurier *et al.*, *A&A* **558**, p. A118 (October 2013).
9. Planck Collaboration, *A&A* **594**, p. A22 (September 2016).
10. J. Delabrouille *et al.*, *A&A* **553**, p. A96 (May 2013).
11. A. Maniyar *et al.*, *A&A* **645**, p. A40 (January 2021).
12. H. Hoekstra *et al.*, *MNRAS* **449**, 685 (May 2015).
13. A. von der Linden *et al.*, *MNRAS* **443**, 1973 (September 2014).
14. V. Biffi *et al.*, *ApJ* **827**, p. 112 (August 2016).
15. Planck Collaboration, *A&A* **550**, p. A131 (February 2013).
16. M. Arnaud *et al.*, *A&A* **517**, p. A92 (July 2010).
17. E. Pointecouteau *et al.*, *arXiv e-prints* , p. arXiv:2105.05607 (May 2021).

The CMB dipole: Eppur si muove

R. M. Sullivan* and D. Scott

Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada

** E-mail: rsullivan@phas.ubc.ca*

www.ubc.ca

The largest temperature anisotropy in the cosmic microwave background (CMB) is the dipole. The simplest interpretation of the dipole is that it is due to our motion with respect to the rest frame of the CMB. As well as creating the ℓ=1 mode of the CMB sky, this motion affects all astrophysical observations by modulating and aberrating sources across the sky. It can be seen in galaxy clustering, and in principle its time derivative through a dipole-shaped acceleration pattern in quasar positions. Additionally, the dipole modulates the CMB temperature anisotropies with the same frequency dependence as the thermal Sunyaev-Zeldovich (tSZ) effect and so these modulated CMB anisotropies can be extracted from the tSZ maps produced by *Planck*. Unfortunately this measurement cannot determine if the dipole is due to our motion, but it does provide an independent measure of the dipole and a validation of the y maps. This measurement, and a description of the first-order terms of the CMB dipole, are outlined here.

Keywords: Cosmic Microwave Background; Cosmic Microwave Background dipole; Special relativity; Thermal Sunyaev-Zeldovich effect.

1. The CMB Sky from *Planck*

Planck[a] was a space-based telescope that measured the microwave sky in nine wavebands, allowing it to capture not only the cosmic microwave background (CMB) but also several Galactic and extragalactic foreground components. This is most clearly seen in figure 51 from Ref. 1, which shows the various wavebands of the *Planck* satellite and the frequency spectra of the foreground signals across those bands. One signal of interest to this study is the thermal Sunyaev-Zeldovich (tSZ) effect, which produces so-called y-type distortion signals. This comes from CMB photons being inverse-Compton scattered, mostly through hot galaxy clusters, which makes holes (or lowers the flux) at low frequencies and up-scatters (or makes an excess flux) at high frequencies. This signal allows us to construct a novel and independent measure of the CMB dipole because temperature anisotropies stemming from the CMB dipole contaminate the y maps. It can also be used as a valuable test of the quality of the y maps. We will start in Sec. 2 by setting out relevant notation for

[a] *Planck* (`http://www.esa.int/Planck`) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states and led by Principal Investigators from France and Italy, telescope reflectors provided through a collaboration between ESA and a scientific consortium led and funded by Denmark, and additional contributions from NASA (USA).

the unboosted CMB sky. Next, in Sec. 3 we will boost the CMB sky, and explore the relevant terms that arise from that boost in the subsections. Of particular relevance, in Sec. 3.3 we will discuss our measurement of the dipole modulation terms that mix with the tSZ effect. We will finish in Sec. 4 with a short discussion and conclusion regarding our work.

2. The Unboosted CMB Sky

To derive the connection between the y map and the dipole we will begin by defining some useful terms regarding the unboosted CMB sky:

$$x \equiv \frac{h\nu}{k_{\rm B}T}; \tag{1}$$

$$I \equiv \frac{2k_{\rm B}^3T^3}{h^2c^2}\frac{x^3}{e^x-1}; \tag{2}$$

$$f(x) \equiv \frac{xe^x}{e^x-1}; \tag{3}$$

$$Y(x) \equiv x\frac{e^x+1}{e^x-1} - 4. \tag{4}$$

These are the dimensionless frequency, the standard Planck blackbody intensity function, the frequency dependence of the CMB anisotropies and the relative frequency dependence of the tSZ effect or y type distortions, respectively. Thus, to first order the anisotropies of intensity measured by *Planck* can be written as

$$\frac{\delta I(\hat{\mathbf{n}})}{If(x)} = \frac{\delta T(\hat{\mathbf{n}})}{T_{\rm CMB}} + y(\hat{\mathbf{n}})Y(x), \tag{5}$$

where $\hat{\mathbf{n}}$ is the line of sight direction on the sky and we have only considered the temperature anisotropies and the y signals here.

3. The Boosted CMB Sky

If we apply a boost to Eq. 5, with a dimensionless velocity $\boldsymbol{\beta}$, we find

$$\begin{aligned}\frac{\delta I'(\hat{\mathbf{n}}')}{If(x)} = {} & \beta\mu + \frac{\delta T(\hat{\mathbf{n}}')}{T_{\rm CMB}}(1+3\beta\mu) \\ & + Y(x)\left(y(\hat{\mathbf{n}}')(1+3\beta\mu) + \beta\mu\frac{\delta T(\hat{\mathbf{n}}')}{T_{\rm CMB}}\right) \\ & + \beta\mu y(\hat{\mathbf{n}}')\left(Y^2(x) - x\frac{dY(x)}{dx}\right) + \mathcal{O}(\beta^2),\end{aligned} \tag{6}$$

where $\mu = \cos(\theta)$, and θ is the angle between the boost $\boldsymbol{\beta}$ and the line of sight $\hat{\mathbf{n}}'$. The first line has the same frequency dependence as thermal fluctuations and so appear in typical CMB temperature anisotropy maps. Crucially for our analysis, the middle line has the same frequency dependence as y-type distortions and thus describes the signals in the y map. The final line has more obscure frequency dependence and is

not discussed here. Additionally, the direction of the incoming photons will change from $\hat{\mathbf{n}}$ to $\hat{\mathbf{n}}'$, where $\hat{\mathbf{n}}' = \hat{\mathbf{n}} - \nabla(\hat{\mathbf{n}} \cdot \beta)$; this deflection of the photons by $\nabla(\hat{\mathbf{n}} \cdot \beta)$ is due to aberration, an effect that is not unique to the microwave sky and occurs for all astronomical observations. We will now discuss each of these terms in turn.

3.1. *The CMB dipole: $\beta\mu$*

In the first line of Eq. 6, the term $\beta\mu$ describes the pure CMB dipole, as discussed previously. This mainly (or perhaps entirely) comes from our local motion with respect to the CMB rest frame and it has been previously measured in Refs. 2, 3, and 4, and most recently in Refs. 5, 6, and 7. Taking the large dipole as being solely caused by our motion, the velocity is $v = (369.82 \pm 0.11)\,\mathrm{km\,s^{-1}}$ in the direction $(l, b) = (264.^{\circ}021 \pm 0.^{\circ}011, 48.^{\circ}253 \pm 0.^{\circ}005)$[5] and can be easily seen in the CMB frequency maps, such as in Fig. 1.

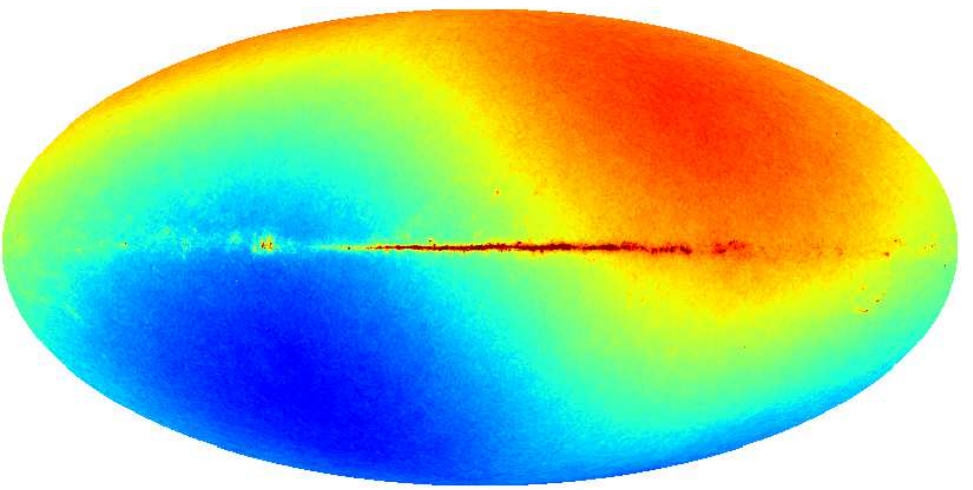
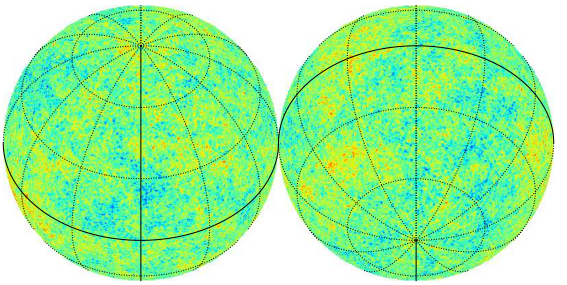


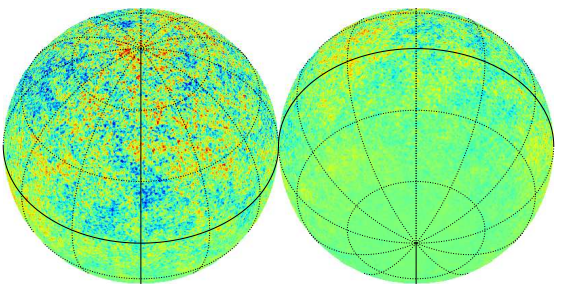
Fig. 1. *Planck* 100-GHz channel map from the NPIPE (PR4) data release,[8] showing the dominant $\ell = 1$ mode or dipole across the sky. The temperature difference across the sky here is 3.36 mK.

3.2. *Aberration and modulation of the CMB anisotropies: $(1 + 3\beta\mu)\delta T(\hat{\mathbf{n}}')/T_{\mathrm{CMB}}$*

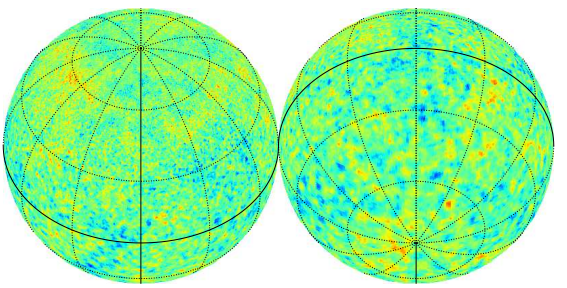
The second term in the first line of Eq. 6 is the dipole aberration and modulation of the temperature anisotropies of the CMB. The modulation causes the temperature anisotropies to be brighter in the forwards direction, and dimmer in the reverse direction. The aberration causes the anisotropies to be more condensed in the forwards direction, and more stretched out in the reverse direction (effectively the same as $\ell = 1$ lensing). These two effects can be seen in Fig. 2. This effect was first measured in Ref. 9 to about $4\,\sigma$.



(a) The unboosted CMB sky.



(b) The modulated CMB sky.



(c) The aberrated CMB sky.

Fig. 2. Here the CMB sky is shown unboosted in (a), with a modulation from a boost of 90 % the speed of light in (b), and with aberration from a boost of 90 % of the speed of light in (c). In the case of modulation, the anisotropies are more intense in the forward direction and less so in the reverse direction, whereas the aberration condenses the anisotropies in the forwards direction and causes them to be more spread out in the reverse direction.

3.3. *Temperature modulation and the tSZ effect:* $Y(x)\beta\mu\delta T(\hat{\mathbf{n}}')/T_{\rm CMB}$

The second line of Eq. 6 shows the dipole-generated signals in the y maps produced by *Planck*. The first half is the same modulation and aberration terms as were seen

in the temperature anisotropies; however, the final term is due to the second-order expansion of the intensity about $T_{\rm CMB}$ and adds a contribution to the y maps from the temperature anisotropies.

We can look for this signal by cross-correlating a template map, derived from the CMB temperature data, with a y map. To this end, we use the so-called `2D-ILC` CMB temperature map which was produced by the "Constrained ILC" component-separation method designed by Ref. 10 to explicitly null out the contribution from the y-type spectral distortions in the CMB map. We also use the `SMICA-NOSZ` temperature map, similarly produced with the express intent of removing the y-type spectral distortions, and which was generated for the *Planck* 2018 data release.[11] Likewise, we use the corresponding `2D-ILC` y map, and the *Planck* `MILCA` y map, which explicitly null out the contributions from a (differential) blackbody spectral distribution in the y map.[12,13] If we multiply our CMB map with $\beta\mu$ and cross-correlate that with our tSZ map, then we can directly probe the dipole modulation.

In Ref. 9 a quadratic estimator was used to determine the dipole aberration and modulation, in essence using the auto-correlation of the CMB fluctuation temperature maps. In this work we use the fact that we know the true CMB fluctuations with excellent precision and therefore know the signal that should be present in the y map. Thus, we fully exploit the angular dependence of the modulation signal and remove much of the cosmic variance that would be present in the auto-correlation. In order to implement this idea we define three templates, B_i (with $i = 1, 2, 3$) as

$$B_i(\hat{\mathbf{n}}) = \beta\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}_i \, \frac{\delta T}{T_0}(\hat{\mathbf{n}}), \tag{7}$$

where $\beta = v/c^5$ is 1.23357×10^{-3} and $\hat{\mathbf{m}}_1, \hat{\mathbf{m}}_2, \hat{\mathbf{m}}_3$ are the CMB dipole direction, an orthogonal direction in the Galactic plane, and the third remaining orthogonal direction (see Fig. 3). Due to the presence of the CMB dipole, the signal B_1 should be present in the y map and so we can directly cross-correlate B_1 with our y map to pull out the signal. Likewise, the cross-correlation of B_2 and B_3 with our y map should give results consistent with noise.

Our y simulations are generated by first computing the power spectra of our data y maps; specifically we apply the `MASTER` method using the `NaMASTER` routine[14] to account for the applied mask.[15] Then we generate y maps using this power-spectrum with the `HEALPix`[16] routine `synfast`. We finally apply a Gaussian smoothing of $5'$ to model the telescope beam. For each analysis method we estimate the amplitude of the dipole ($\hat{\beta}_i$) in each of the three orthogonal directions. We apply the same analysis procedure on a suite of 1000 y simulations, generated with and without the dipolar modulation term.

We use two methods of cross-correlation: the first is performed directly in map-space; and the second is performed in harmonic space. For both methods we first apply our mask to the templates B_i and the y map.

In the map-space method we then locate all peaks (i.e., local maxima or minima) of the template map B_i and select a patch of radius $2\overset{\circ}{.}0$ around each peak. For every

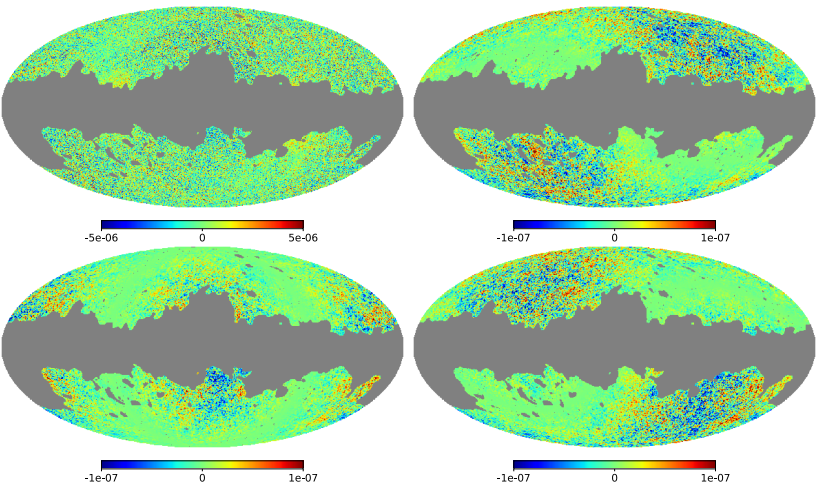


Fig. 3. Map of the tSZ effect from the `MILCA` component-separation method in y-map units (top left) and the expected modulated CMB signal (top right) generated using the `SMICA-NOSZ` CMB map in units of T_0. The bottom left and right panels are the CMB anisotropies modulated in orthogonal directions to the CMB dipole. Note that the map of the tSZ effect (top left) has a different scale bar when compared to the other three (i.e., the modulation signal is about 50 times weaker).

peak we obtain an estimate of $\hat{\beta}_i$ through the simple operation

$$\hat{\beta}_{i,p} = \beta \frac{\sum_{k\in D(p)} B_{i,k} y_k}{\sum_{k\in D(p)} B_{i,k}^2}, \tag{8}$$

where $D(p)$ is the collection of all *unmasked* pixels in a $2\overset{\circ}{.}0$ radius centred on pixel p, and p is the position of a peak. Equation 8 is simply a cross-correlation in map space and by itself offers a highly noisy (and largely unbiased) estimate. We then combine all individual peak estimates with a set of weights (w_p) to give our full estimate:

$$\hat{\beta}_i = \frac{\sum_p w_{i,p} \hat{\beta}_{i,p}}{\sum_p w_{i,p}}. \tag{9}$$

We choose w_p to be proportional to the square of the dipole, and use weights that are proportional to the square of the Laplacian at the peak;[17] this favours sharply defined peaks over shallow ones. Finally we account for the scan strategy of the *Planck* mission by weighting by the 217-GHz hits map,[18] denoted H_p^{217}. The weights are then explicitly

$$w_{i,p} = |\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}_i|_p^2 \left(\nabla^2 (B_i)\big|_p \right)^2 H_p^{217}. \tag{10}$$

We apply the method for each of our simulated y maps, in exactly the same way as for the data.

Under the assumption that the y map contains the template (B_i), the y multipoles are Gaussian random numbers with mean and variance given by

$$s^i_{\ell m} = \int d\Omega\, \beta\, \hat{\mathbf{m}}_i \cdot \hat{\mathbf{n}}\, \frac{\delta T}{T_0} M(\Omega) Y^*_{\ell m}, \tag{11}$$

$$\sigma^2_\ell = C^y_\ell + N^y_\ell, \tag{12}$$

respectively, where $M(\Omega)$ is the mask over the sphere, $Y_{\ell m}$ are the spherical harmonics, and the $\hat{\mathbf{m}}_i$ are as defined in Eq. 7. We can obtain an estimate of β_i by taking the cross-correlation with inverse-variance weighting. Our estimator is therefore

$$\hat{\beta}_i = \beta \sum_{i'} \left[\sum_{\ell m}^{\ell_{\max}} s^i_{\ell m} (s^{i'}_{\ell m})^* / \sigma^2_\ell \right]^{-1} \sum_{\ell m}^{\ell_{\max}} s^{i'}_{\ell m} (y_{\ell m})^* / \sigma^2_\ell. \tag{13}$$

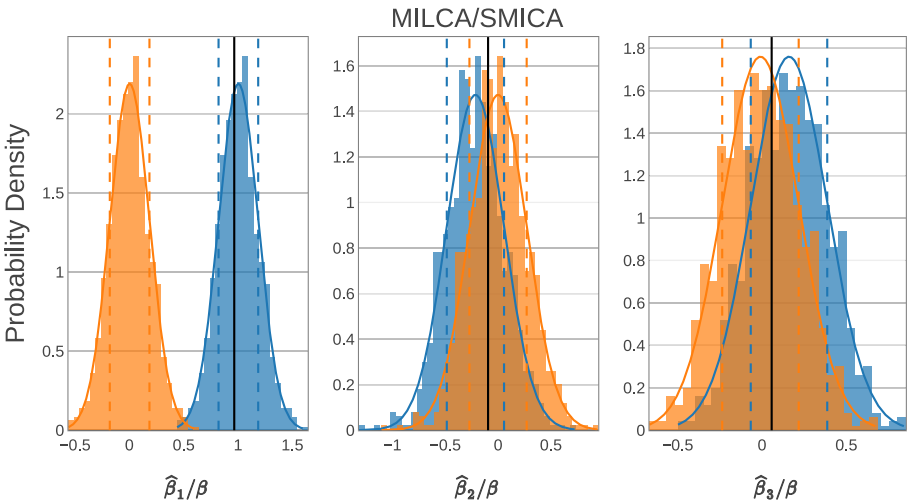


Fig. 4. Histograms of $\hat{\beta}_i/\beta$ values (with 1, 2, and 3 corresponding to the CMB dipole direction, the Galactic plane, and a third orthogonal direction) using the map-space analysis for `MILCA` y maps, and for CMB template maps `SMICA-NOSZ`. Blue histograms are simulations with the dipolar modulation term, while orange histograms are simulations without modulation. Black vertical lines denote the values of the real data, demonstrating that they are much more consistent with the existence of the dipolar modulation term than without it. Dashed lines show the 68 % regions for a Gaussian fit to the histograms. To see the full results with all data analysis combinations see Ref. 19.

First we compare the consistency of the data with our two sets of simulations (with and without the dipole term). This comparisons shown in Figs. 4 and 5 have blue histograms being the simulations *with* the dipole term and orange histograms being *without*. The data (black line) for `2D-ILC` and `MILCA` can clearly be seen to be consistent with the simulations with the dipole term. Further details and analysis may be found in Ref. 19.

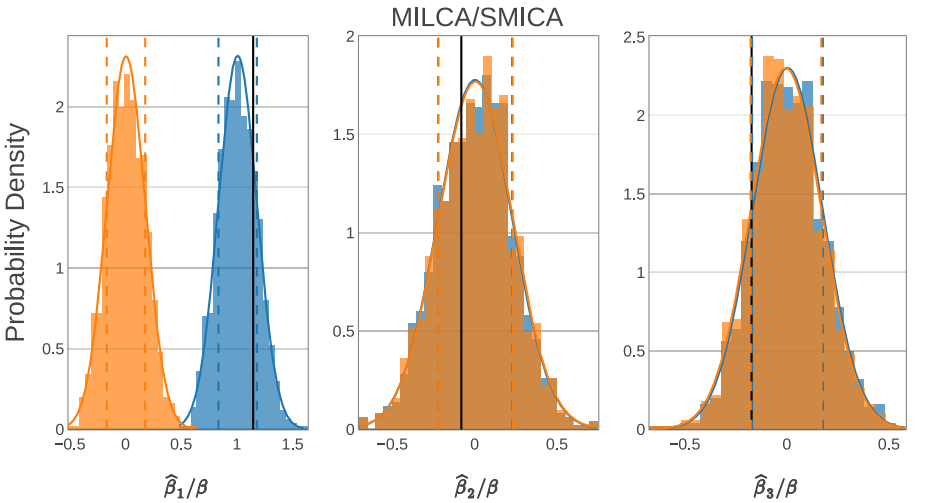


Fig. 5. As in Fig. 4, except now for the harmonic-space analysis.

4. Conclusion: Distinguishing Intrinsic and Extrinsic CMB Dipoles

The frequency-dependent part of the dipolar-modulation signal is agnostic to the source of the large CMB dipole. Therefore, its measurement is an independent determination of the CMB dipole. While it may be tempting to use this signal to distinguish an intrinsic dipole, it has been shown that an intrinsic dipole and a dipole induced by a velocity boost would in fact have the same dipolar-modulation signature on the sky.[20,21]

Due to the existence of the CMB dipole, a tSZ map necessarily contains a contaminating signal that is simply the dipole modulation of the CMB anisotropies. This occurs because CMB experiments do not directly measure temperature anisotropies, but instead measure intensity variations that are conventionally converted to temperature variations. This contamination adds power to the tSZ map in a Y_{20} pattern, with its axis parallel to the dipole direction. We have measured this effect and determined a statistically independent value of the CMB dipole, which is consistent with direct measurements of the dipole. Using a conservative multipole cut on the y map, the significance of the detection of the dipole modulation signal is around 5 or $6\,\sigma$, depending on the precise choice of data set and analysis method.

The question as to whether an intrinsic dipole could ever be observationally distinguished from an extrinsic dipole (i.e. a Doppler boost) remains an open question. The terms discussed in Eq. 6 are based on the assumption of a CMB blackbody spectrum and cannot be used to distinguish the two, as they would naturally arise whether the CMB dipole is caused by a boost, or if there is for some other reason a dipole on the sky with the same magnitude and direction.

Acknowledgements

We would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada. Some of the results in this paper have been derived using the `HEALPix` package. Results are based on observations obtained with Planck (http://www.esa.int/Planck), an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada. We would also like to thank Dagoberto Contreras for collaboration on topics within this paper.

References

1. Planck Collaboration X, *Planck* 2015 results. X. Diffuse component separation: Foreground maps, *A&A* **594**, p. A10 (2016).
2. A. Kogut, C. Lineweaver, G. F. Smoot, C. L. Bennett, A. Banday, N. W. Boggess, E. S. Cheng, G. de Amici, D. J. Fixsen, G. Hinshaw, P. D. Jackson, M. Janssen, P. Keegstra, K. Loewenstein, P. Lubin, J. C. Mather, L. Tenorio, R. Weiss, D. T. Wilkinson and E. L. Wright, Dipole Anisotropy in the COBE Differential Microwave Radiometers First-Year Sky Maps, *ApJ* **419**, p. 1 (December 1993).
3. D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer and E. L. Wright, The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set, *ApJ* **473**, p. 576 (December 1996).
4. G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker and E. L. Wright, Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results, *ApJS* **180**, 225 (February 2009).
5. Planck Collaboration I, *Planck* 2018 results. I. Overview, and the cosmological legacy of *Planck*, *A&A* **641**, p. A1 (2020).
6. Planck Collaboration II, *Planck* 2018 results. II. Low Frequency Instrument data processing, *A&A, in press* (2019).
7. Planck Collaboration III, *Planck* 2018 results. III. High Frequency Instrument data processing, *A&A, in press* (2019).
8. Planck Collaboration Int. LVII, *Planck* intermediate results. LVII. NPIPE: Joint *Planck* LFI and HFI data processing, *A&A* **643**, p. 42 (2020).
9. Planck Collaboration XXVII, *Planck* 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove, *A&A* **571**, p. A27 (2014).
10. M. Remazeilles, J. Delabrouille and J.-F. Cardoso, CMB and SZ effect separation with constrained Internal Linear Combinations, *MNRAS* **410**, 2481 (February 2011).
11. Planck Collaboration IV, *Planck* 2018 results. IV. Diffuse component separation, *A&A, in press* (2019).
12. G. Hurier, J. F. Macías-Pérez and S. Hildebrandt, MILCA, a modified internal linear combination algorithm to extract astrophysical emissions from multifrequency sky maps, *A&A* **558**, p. A118 (October 2013).
13. Planck Collaboration XXII, *Planck* 2015 results. XXII. A map of the thermal Sunyaev-Zeldovich effect, *A&A* **594**, p. A22 (2016).
14. D. Alonso, J. Sanchez, A. Slosar and L. D. E. S. Collaboration, A unified pseudo-Cℓ framework, *Monthly Notices of the Royal Astronomical Society* **484**, 4127 (01 2019).

15. E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet and F. Hansen, MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets, *ApJ* **567**, 2 (March 2002).
16. K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke and M. Bartelmann, HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, *ApJ* **622**, 759 (April 2005).
17. V. Desjacques, Baryon acoustic signature in the clustering of density maxima, *Phys. Rev. D* **78**, p. 103503 (November 2008).
18. Planck Collaboration VIII, *Planck* 2015 results. VIII. High Frequency Instrument data processing: Calibration and maps, *A&A* **594**, p. A8 (2016).
19. Planck Collaboration Int. LVI, *Planck* intermediate results. LVI. Detection of the CMB dipole through modulation of the thermal Sunyaev-Zeldovich effect: Eppur si muove II, *A&A* **644**, p. 100 (2020).
20. A. Challinor and F. van Leeuwen, Peculiar velocity effects in high-resolution microwave background experiments, *Phys. Rev. D* **65**, p. 103001 (May 2002).
21. A. Notari and M. Quartin, CMB all-scale blackbody distortions induced by linearizing temperature, *ArXiv e-prints* (October 2015).

High angular resolution Sunyaev Zel'dovich observations: The case of MISTRAL

E. S. Battistelli[1,*], E. Barbavara[1], P. de Bernardis[1], F. Cacciotti[1], V. Capalbo[1], E. Carretti[2], F. Columbro[1], A. Coppolecchia[1], A. Cruciani[3], G. D'Alessandro[1], M. De Petris[1], F. Govoni[4], G. Isopi[1], L. Lamagna[1], P. Marongiu[4], S. Masi[1], L. Mele[1], E. Molinari[4], M. Murgia[4], A. Navarrini[4], A. Orlati[2], A. Paiella[1], G. Pettinari[5], F. Piacentini[1], T. Pisanu[4], S. Poppi[4], G. Presta[1] and F. Radiconi[1]

[1] *Dipartimento di Fisica, Sapienza Università di Roma - Piazzale Aldo Moro, 5, I-00185 Rome, Italy*

[2] *INAF - Istituto di Radioastronomia - Via P. Gobetti, 101 - I-40129 Bologna, Italy*

[3] *INFN - Sezione di Roma, Piazzale Aldo Moro 5 - I-00185, Rome, Italy*

[4] *INAF - Osservatorio Astronomico di Cagliari, Via della Scienza 5 - I-09047 Selargius (CA), Italy*

[5] *Istituto di Fotonica e Nanotecnologie - CNR, Via Cineto Romano 42, I-00156 Rome, Italy*

[*] *E-mail: elia.battistelli@roma1.infn.it*

The MIllimeter Sardinia radio Telescope Receiver based on Array of Lumped elements kids, MISTRAL, is a millimetric ($\simeq 90GHz$) multipixel camera being built for the Sardinia Radio Telescope. It is going to be a facility instrument and will sample the sky with 12 arcsec angular resolution, 4 arcmin field of view, through 408 Kinetic Inductance Detectors (KIDs). The construction and the beginning of commissioning is planned to be in 2022. MISTRAL will allow the scientific community to propose a wide variety of scientific cases including protoplanetary discs study, star forming regions, galaxies radial profiles, and high angular resolution measurements of the Sunyaev Zel'dovich (SZ) effect with the investigation of the morphology of galaxy cluster and the search for the Cosmic Web.

Keywords: Millimetric astronomy, large radio telescopes, Sunyaev Zel'dovich effect

1. Introduction

High angular resolution millimetric observation are key to understand a wide variety of scientific cases. Among others, the interaction of Cosmic Microwave Background (CMB) photons with the hot electron gas in galaxy clusters and surrounding medium (the Sunyaev Zel'dovich, SZ effect) promises to study galaxy clusters and their deviation from relaxed behaviour. This include high angular resolution measurements of the SZ effect with the investigation of the morphology of galaxy cluster and the search for the Cosmic Web. Also, protoplanetary discs study, star forming regions, and galaxies radial profiles are among the important scientific cases one could achieve with high angular resolution millimetric observation.

Only a handful of instruments are capable of measuring, for instance, galaxy clusters with high enough angular resolution to resolve them. Among them, we mention

MUSTANG2 at the Green Bank Telescope (GBT[a]) and NIKA2 at IRAM telescope[b]. MUSTANG2[1] is a millimetric cryogenic camera installed at the 100m GBT. It is operating in the frequency range between 75GHz and 105GHz and samples 4 arcmin f.o.v. in the sky with 223 microstrip-coupled Transition Edge Sensors (TESs). The resulted diffraction limited angular resolution is of the order of 9 arcsec. A frequency domain multiplexing technique is used to read MUSTANG2 detectors through flux-ramp modulation technique.

Another receiver which addresses similar science goals, although it takes data at higher frequencies, is NIKA2.[2–4] NIKA2 is a 2 channel millimetric camera measuring at 150GHz (one array) and 260GHz (two arrays). It observes from the 30m IRAM telescope with a field of view (f.o.v.) of 6.5 arcmin while the angular resolution is around 18 arcsec at 150GHz and around 10 arcsec at 260GHz. NIKA2 hosts 3 arrays of Kinetic Inductance Detectors (KIDs), each of around 2000 pixels, cooled down at 100mK with a dilution refrigerator. Both MUSTANG2 and NIKA2 are very active in the field of millimetric astronomy especially for the high angular resolution detection of the SZ effect.

A third instrument being prepared for such scientific cases is the MIllimeter Sardinia radio Telescope Receiver based on Array of Lumped elements kids, MISTRAL, to be fielded at the 64m Sardinia Radio Telescope (SRT[c]). MISTRAL will use an array of 408 KIDs with a f.o.v. of 4 arcmin and an angular resolution of around 12 arcsec. It is a cryogenic instrument with detectors cooled down at 270mK and with frequency domain multiplexing, ROACH2 based read-out.

2. Sardinia Radio Telescope and its upgrade

2.1. *Sardinia Radio Telescope*

The Sardinia Radio Telescope (SRT - Lat. 39.4930 N; -Long. 9.2451 E), is a Gregorian configured fully steerable 64m primary mirror radio telescope which can work from 300MHz and 116GHz[5] (see Fig. 1). It is a multipurpose instrument operated in either single dish or Very Long Baseline Interferometer mode and managed by the Italian Astrophysics National Institute (INAF). The telescope manufacturing started in 2003 and was completed in August 2012 when the technical commissioning started. The scientific exploitation of the SRT started on 2016 with an Early Science Program, while regular proposals started in 2018.

SRT has a f/0.33 primary focus and f/2.34 secondary focus allowing to host the low frequency receivers ($\nu < 2GHz$) in the primary focus, and high frequency receivers in the Gregorian room. The 64-m primary mirror (M1) is composed of 1008 electromechanically controlled aluminum elements by actuators. The secondary mirror (M2) is a 7.9m mirror composed of 49 aluminum elements adjustable for focus

[a] https://greenbankobservatory.org/science/telescopes/gbt/
[b] https://www.iram-institute.org/EN/30-meter-telescope.php
[c] http://www.srt.inaf.it/

Fig. 1. The 64-m Sardinia Radio Telescope (image credits F. Guidi).

operations. Both M1 and M2 are shaped to minimize spillover and the standing waves between the receiver and the subreflector.[5]

2.2. *Observational site*

The location of the SRT is in the island of Sardinia, ≃50km north of Cagliari, at 600m a.s.l.. It is a region suited for radio-observations and the telescope was designed to be operating up to 116GHz. Estimation of sky opacity, based on recorded dedicated radiometer data, reports[d] an opacity $\tau < 0.15Np$ (50th percentile) at 93GHz during the winter nights. The average precipitable water vapour (PWV) in the same conditions is mainly 8mm. For comparison, the GBT reports $\tau < 0.125$ (50th percentile) at 86GHz, and an average $PWV < 9mm$ (50th percentile)[e]. The above estimation is confirmed by radiosondes profiles taken at Cagliari airport and scaled for the SRT site. This shown $PWV < 11mm$ (50th precentile) and opacities $< 0.2Np$ (50th percentile) at 100GHz.[6]

2.3. *PON upgrade*

In 2018, a National Operational Program (PON) grant was assigned to INAF with the aim to exploit to the full the SRT capability to reach mm wavelenghts up to 116GHz.[7] One of the Working Packages includes a metrological system to measure

[d] `http://hdl.handle.net/20.500.12386/28787`
[e] `https://www.gb.nrao.edu/mustang/wx.shtml`

with high precision deformations of the SRT due to gravitational and thermal effects; by means of the active surface it will be possible to correct the deformations of the main mirror allowing the surface roughness needed for observations up to 116 GHz. Also, the metrological system will allow high precision pointing measuring and correcting the subreflector displacement from the optical axis together with the deformation of the whole structure.

In addition an upgrade of the SRT receiver fleet is foreseen including an extension to higher frequencies up to 116GHz, thus exploiting to the full its optical surface. This grant was divided in 9 working packages which include new receivers like a 19 feeds, double polarization Q-band heterodyne receiver, a triple band (K, Q, W) coherent receiver, a 16 double polarization W-band heterodyne receiver, and a W-band, total power, 408 pixels bolometric[f] receiver: MISTRAL.

3. MISTRAL

3.1. *Introduction*

MISTRAL is a facility instrument to be installed at the SRT in 2022. It operates in an atmospheric window in the frequency range 78-103GHz, namely W-band, that is interesting for a number of scientific reasons, and most importantly it provides low optical depth and allows high efficiency observations. It will be installed in the Gregorian room of the SRT and as such, it needs to meet several requirements and limitations. These include:

- 250kg maximum weight;
- 700mm x 700mm x 2400mm (h) maximum occupation. In addition, the top part should not interfere with the Gregorian room structure;
- adeguate Radio Frequency shielding;
- ability to work from 30° elevation to 80° and the need not to be damaged in the range 0-90° elevation;
- long ($\simeq$100m) long helium lines for the Pulse Tube head-compressor connection;
- remote operations;
- MISTRAL will be installed on the Gregorian room turret (a rotating structure which hosts MISTRAL an all other receivers) and needs to be agile enough to be inserted in the SRT focus or removed through rotation of the turret itself (see Fig. 2).

[f] Actually MISTRAL will implement Kinetic Inductance Detectors which are not properly bolometers.

3.2. *Cryostat*

MISTRAL consists of a cryostat, being constructed by QMC instruments[g], with a Pulse Tube cryocooler, an He-10 sorption refrigerator, custom optics and detectors. We use a Sumitomo[h] RP-182B2S-F100H Pulse Tube (PT) cryocooler. From Sumitomo user manuals, PT heads work properly in vertical position with inclination no larger than +/-20°. In order to have the PT heads as close as possible to vertical position during observations, we have positioned the head with an inclination of 57.5° with respect to the focal plane. This allows a nominal observation elevation of 37.5-77.5°. Nevertheless, we have tested higher inclinations and there is no impact on the cryogenics with inclination of +/-25° (resulting in elevation range of 32.5-82.5°) with no degradation of the thermal performance. MISTRAL has a $40K$ shield, a $4.2K$ shield and a sub-Kelvin stage. The PT is used to cool down the $40K$ and $4.2K$ shields and provides a cooling power of $36W$ for the $40K$ shield and $1.5W$ for the $4.2K$ shield.

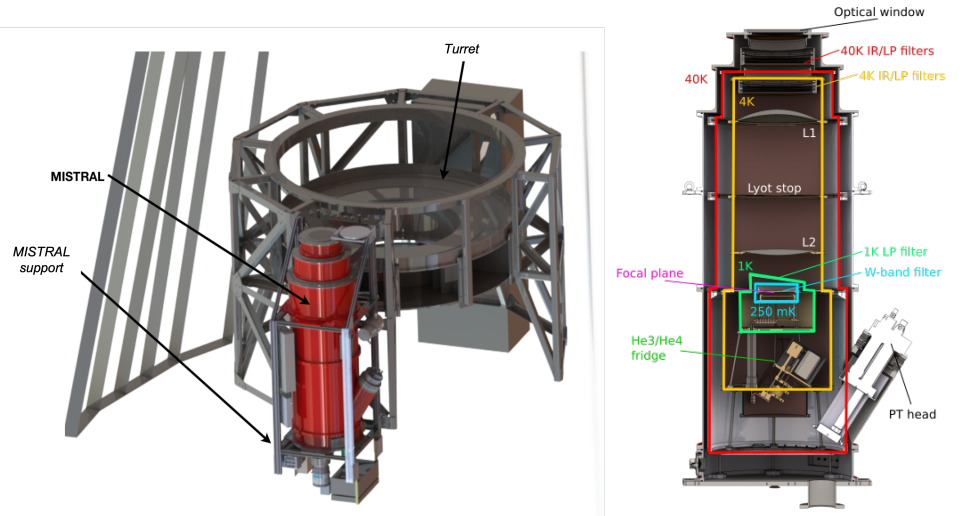


Fig. 2. Left: SRT Turret. with MISTRAL installed. Right: a cut of MISTRAL cryostat highlighting the main parts.

The PT head is controlled by a PT compressor which cannot be inclined with the telescope elevation angle. For this reason, MISTRAL compressor will not be located at the Gregorian room, but rather in a compressor room at the base of the radio telescope, that can rotate with the azimuth angle but does not change inclination with the elevation angle. This requires about 100m of flexible lines for the

[g] http://www.terahertz.co.uk/qmc-instruments-ltd
[h] https://www.shicryogenics.com/

compressor-to-PT head connection. Sumitomo (as well as other companies) report 20m maximum length for this operations so on-site tests were performed in order to verify if a PT can keep the $4.2K$ base temperature and with which dissipated power.[8] We made these tests using a Sumitomo RP-082B2S cryocooler, which is similar to the one chosen for MISTRAL but with lower cooling power (i.e. $0.9W$ at $4.2K$). It was shown that this cryocooler was able to hold a base temperature of $4.2K$ with a power load of 0.6W with $119m$ of flexible lines. Since the power dissipation requirement for MISTRAL is of the order of $0.6W$, the $1.5W$ PT will be able to satisfy the requirement with some margin. In addition, a vibration dumper, and a remote PT valve, will allow to reduce vibration into the detectors.

Attached to the $40K$ and the $4.2K$ stages, it will be applied radiation shields as well as radiation filters allowing to block thermal input of the sub-K stages of MISTRAL. The sub-K fridge is a Chase[i] He^3/He^4 sorption refrigerator (Twin GL10 sorption fridge) allowing a nominal cooling power as in the following table:

Stage	temperature no load	temperature loaded	Load applied
He^4	$0.840K$	$0.913K$	$150\mu W$
He^3	$0.299K$	$0.332K$	$30\mu W$
Ultra cold He^3	$0.195K$	$0.251K$	$20\mu W$

The total room to be cooled down is of the order of $1.5m^3$ and the total weight of MISTRAL is around $250kg$. The top part of the cryostat has been shaped in such a way it does not interfere with the Gregorian room walls while rotating in and out of the focus position (see Fig. 2).

3.3. *Optics and quasi-optics*

The optical design of MISTRAL includes a set of radiation quasi-optical filters, anchored at the different thermal stages of the cryostat, an Anti Reflection Coated (ARC) Ultra High Molecular Weight (UHMW) polyetilene window, and two ARC Silicon lenses able to image the Gregorian Focus on the array of detectors. Detectors are coupled to radiation through open space (filled array) so cryogenic cold stop, placed in between the two lenses, is needed.

Quasi-optical filters are a combination of metal mesh filters, thin IR filters, sub-mm low pass filters (LPE), and a final 78-103GHz Band pass filter produced at QMC instruments[j]. The final configuration still needs to be settled down as it depends on cryogenic tests and performance that still need to be carried on. Nevertheless, the filter chain will include:

- $300K$: ARC UHMW window of 190mm outer diameter, plus two blocking thermal filters;

[i] `https://www.chasecryogenics.com/`

[j] `http://www.terahertz.co.uk/qmc-instruments-ltd`

- $40K$: two additional thermal filters plus a Low Pass Edge filter;
- $4K$: two more thermal filters (IR5 and IR6) and another Low Pass filter;
- $1K$: Low Pass filters;
- 0.350 K: Low Pass filters;
- 0.250 K: Band Pass Filter centered at 90GHz with a bandwidth either of 30GHz or of 20GHz (to be defined).

Some characteristics of the SRT optical configuration (which were adopted to make the SRT suitable for a wide range of applications) limit the Diffraction Limited Field Of View (DLFOV) usable by the mm camera. As mentioned, we plan to implement a naked Kinetic Inductance Detectors (KIDs) array (a filled array) optically coupled to the antenna by a relay optics. This is the preferred solution for satisfying the following requirements:

- to possibly rescale the Gregorian telescope focal plane;
- to insert a cold aperture stop;
- to provide a gaussian wavelength independent beam telescope configuration;
- to provide a telecentric optics.

The first requirement allow us to tune the size of the KID array preserving the telescope DLFOV. The second condition allows us to introduce an aperture conjugated to the secondary mirror, reduces the spillover contribution due to subreflector and/or primary mirror, ensuring the highest edge taper in a cold aperture. The third condition enables a magnification of the relay optics and the output beam waist location wavelength-independent, at least inside the expected bandwidth, which is a considerable innovation point. The fourth requirement ensures a homogeneous illumination of the array, apart from a possible aberrations impact, suitable for the solution of a naked array maintaining an adequate Strehl Ratio.

Silicon is a suitable material for lens fabrication at millimetri wavelenghts due to its high dielectric constant, low loss tangent, and high thermal conductivities. The two Silicon lenses allow to report 4 arcmin of the SRT focus onto the array of 408 KIDs. They are anti-reflection coated with Rogers RO3003. Their diameter is 290mm and 240mm respectively while the aperture cold stop diameter of 125mm. All the lenses+cold stop system is kept at 4K. The in band average simulations report excellent values with a Strehl Ratio from 0.97 to 0.91 for on-axis and F. O. V. edge positions. Analogously, the FWHM is 12.2 arcsec on axis, and 12.7 arcsec at 45mm off axis.

3.4. *Detectors*

MISTRAL will take advantage of the high sensitivity and the capability of Frequency Domain Multiplexing of KIDs cryogenic detectors. KIDs are superconductive detectors where millimetric radiation with higher energy with respect to the Cooper pair binding energy, can break Cooper pairs producing a change in the pop-

ulation densities and thus in the kinetic inductance. In fact, the inductance L of a thin superconductor is dominated by the kinetic inductance L_k, which depends on the Cooper pair density.

If we integrate this inductance into a resonating RLC circuit, with a well defined resonant frequency, we will see a change in the impedance of the circuit when the inductance is illuminated. This change in impedance can be measured by exciting the circuit with an RF tone at the resonant frequency and monitoring both amplitude and phase of the outcoming tone. The amplitude, and similarly the phase, will change because of the variation in L_k. The general effect is that a decrease in the Cooper pair density increases L_k, thus it lowers the resonant frequency. Furthermore, the increase of the quasi-particle density increases the internal dissipations, thus reducing the quality factor of the resonator. KIDs are intrinsically easy to multiplex in the frequency domain because an array of resonators with different resonant frequencies can be simultaneously readout by sending a comb of Radio Frequency (RF) tones through a single feed-line.

MISTRAL KIDs are Ti-Al bilayer 10 + 30 nm with critical temperature Tc=945mK and are fabricated at CNR-IFN[k].[9,10] They are front-illuminated 3rd order Hilbert crude absorber with backshort on the opposite side of the wafer. The detector array is composed of 408 KIDs detectors. They are 3mm x 3mm each and are arranged on an equilateral triangle every 4.2mm on a 4 inches silicon wafer (see Fig. 3). They sample the focal plane with a FWHM angular spacing of 10.6 arcsec lower the each pixel angular resolution.

As mentioned, KIDs behave as high quality factor, Q, LC resonators. High values of Q allow to multiplex thousands of KIDs, with different frequencies, all coupled to the same feedline. KIDs are in fact intrinsically easy to multiplex in the frequency domain, because an array of resonators with different resonant frequencies can be simultaneously readout by sending a comb of RF tones through a single feed-line. We will use ROACH2 based Frequency Domain Multiplexing which was originally developed for the OLIMPO experiment.[11]

3.5. *Magnetic shield*

Another important part of MISTRAL is a 1-mm thick Cryoperm 10 μ-metal shield, characterized by an high magnetic permeability ($\mu > 7000$), that acts as a shielding from spurious magnetic fields such as the geomagnetic field. When a KID, usually composed of a thin wire, is exposed to an external magnetic field, a shielding current will be induced. Consequentially, its kinetic inductance changes. The external magnetic field will increase the kinetic inductance of the circuit, thus increasing the total inductance, eventually reducing the quality factor of the resonator. This effect was observed on the MISTRAL detector array[13] and required a magnetic shield to be lower than the earth magnetic shield.

[k] `https://www.roma.ifn.cnr.it/`

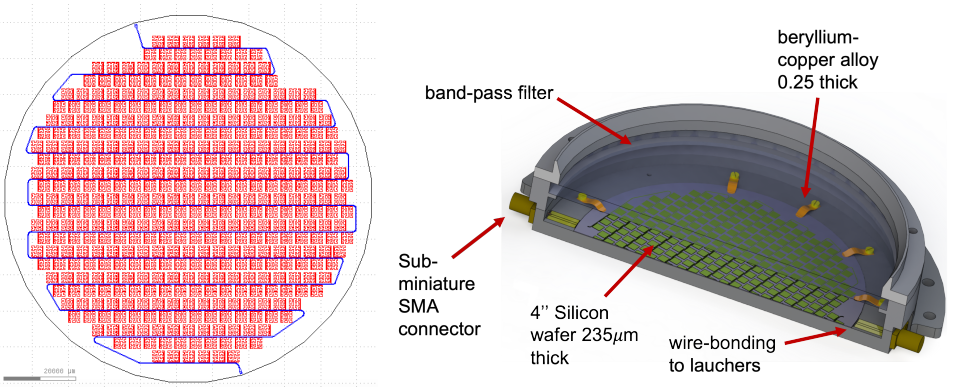


Fig. 3. Left: a design of the focal plane of MISTRAL with its feedline.[12] Right: KIDs holder with the surrounding components.

4. Sensitivity forecast and scanning strategy

4.1. *Noise estimate*

A crucial step to predict the performance on the sky is estimating the expected noise which should affect our detectors. Electrical Noise Equivalent Power (NEP) should be compared with Photon noise NEP due to warm optics and atmosphere emission. The latter, should be considered in its emission part and turbulence part which will act as correlated noise.

Electrical NEP of MISTRAL detectors was measured to be temperature dependent. If we assume a final detectors temperature of 270mK we have average values of our test-bed array of detectors of $NEP_{avg} = 3.5 \times 10^{-16} W/\sqrt{Hz}$ with a maximum and a minimum range of $NEP_{max} = 7.4 \times 10^{-16} W/\sqrt{Hz}$ and $NEP_{min} = 1.6 \times 10^{-16} W/\sqrt{Hz}$ respectively. These numbers need to be compared with estimated photon noise and turbulence noise from atmosphere. The photon noise part was estimated to be, within the 30GHz bandwidth, of $NEP_{ph} = 0.9 \times 10^{-16} W/\sqrt{Hz}$ (which converts into a Noise Equivalent Flux Density of 0.30mJy/beam). To this, in order to estimate the real sensitivity of MISTRAL, we have to add the atmospheric effects due to turbulence. An a-priori calculation of this effect is quite complicated because it also depends on the kind of filtering and on the scanning strategy: the best option would be directly measuring the atmosphere with MISTRAL. This will happen during the commissioning phase, but it is possible to give an early estimate by using experimental data taken at different frequencies. We have SRT data in the K-band (at 22GHz) and we can assume that the ratio between atmospheric fluctuations is equal to the squared ratio of the opacities. In this way, we can infer the atmospheric fluctuation noise at 90GHz using atmospheric modelling softwares such as *am*[1]. This increases the atmospheric noise estimation a factor 10 with respect to

[1] `https://www.cfa.harvard.edu/~spaine/am/`

the photon noise level even in the best 5% percentile resulting in a noise Noise Equivalent Flux Density of $3mJy/beam$. The final result is that MISTRAL detectors will be dominated by atmospheric noise rather than by electric noise, as desired.[14]

4.2. *Scanning strategy*

The final noise level of our maps will depend on the detectors sensitivity, on the atmospheric contamination but also on the scanning strategy and the filtering of the data. We can clearly clean observations very aggressively at the cost of loosing large angular scales on sky. Among the many scientific goals of MISTRAL, the possibility of retrieving large angular scales at the level of its field of view (i.e. 4 arcmin) and even more would be crucial for the study of galaxy clusters and, for instance, for the search for the Cosmic Web in close-by clusters. Scanning strategy drives also the sensitivity as we can more or less efficiently remove atmosphere emission and its fluctuations

At the SRT it is already in use an on-the-fly map scanning strategy which allows orthogonal scans in R.A./Dec or Azimuth/Elevation each composed by parallel subscans. In addition, we are working on the scanning strategy based on Lissajou daisy scanning already in use on different Telescopes such as, for instance, the GBT[m] which has the advantage to reduce to the minimum the overhead due to Telescope inversion and optimize the integration time on the center of an observed field (see Fig. 4). This kind of scanning strategy can easily be analysed with Fourier space filtering of (typically) 0.1Hz (cut-on) and 7.5Hz (cut-off frequency).

5. Science case

MISTRAL will be a facility instrument. Thus, it will be open to the scientific community to decide what kind of scientific output it can achieve and propose the observations to the Time Allocation Committee of the Sardinia Radio Telescope. Nevertheless, we will list in the following a few science cases including those that we think are the most interesting.

Protoplanetary discs millimetric measurements in star forming regions allow to break the degeneracy present in InfraRed (IR) measurements due to the optically thick nature of the hot inner disc. Measuring the SED at mm wavelengths with high angular resolution (i.e. $\simeq$ 10 arcsec) can add information to planetary formation theories.[15] Star formation in molecular clouds with high enough angular resolution allow to distinguish starless cores with respect to those hosting protostars. This is for instance what was attempted with the Large Millimeter Telescope.[n,16] Continuum resolved galaxies observations can give information about morphology and radial

[m] `https://www.gb.nrao.edu/scienceDocs/GBTog.pdf`
[n] `http://lmtgtm.org/`

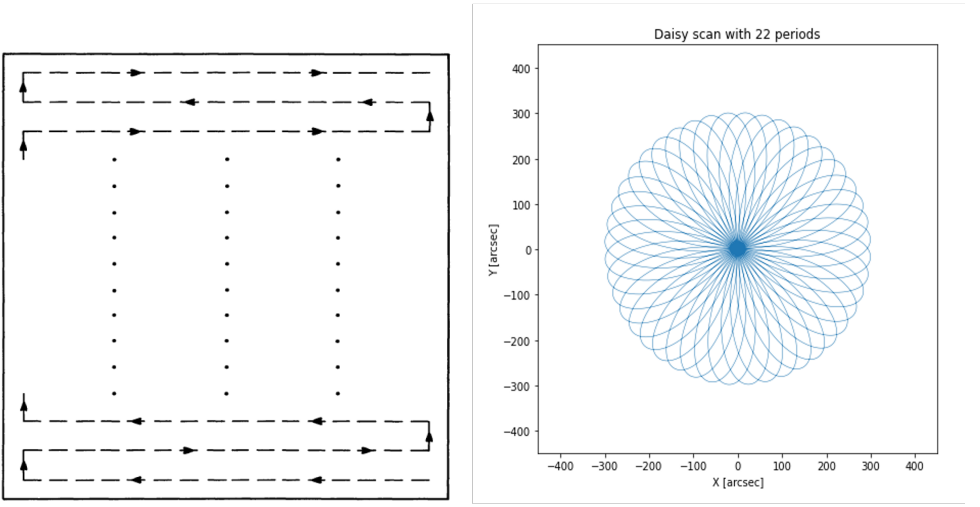


Fig. 4. Left: on-the-fly map scanning strategy already implemented at the SRT. Right: simulation of 22 periods dasy scanning over a 10 arcmin sky area.

profiles (gas column profiles, dust temperature profiles).[17] Spatially resolved high-z radio galaxies can provide info about cold dust re-emission.[18]

When the CMB photons scatter off a hot electron gas, for example in galaxy clusters, they undergo inverse Compton scattering which is visible in the frequency spectrum of the CMB. The resulting distorted spectrum is thus:

$$\frac{\Delta I(x)}{I_0} = y\frac{x^4 e^x}{(e^x-1)^2}\left(x\coth\frac{x}{2}-4\right) = yg(x) \tag{1}$$

where: $I_0 = \frac{2h}{c^2}\left(\frac{k_b T_{CMB}}{h}\right)^3$, T_{CMB} is the CMB temperature, $x = \frac{h\nu}{k_b T_{CMB}}$ is the adimensional frequency, and $y = \int n_e \sigma_T \frac{k_B T_e}{m_e c^2}$ is the Comptonization parameter, the integral along the line of sight dl of the electron density n_e multiplied by the electron temperature T_e (see Fig. 5). This is the Sunyaev Zel'dovich (SZ) effect[19,20] and can be used to study the physics of Galaxy clusters. In fact, galaxy clusters can experience a wide variety of situation including collisions, merging, and non-relaxed situation. In addition, not only galaxy clusters can host relativistic electrons. Hydrodynamical simulations[21,22] suggested that galaxy clusters occupy the knots of the so called Cosmic Web and Warm to Hot Intergalactic Medium (WHIM) is disposed connecting galaxy cluster forming filaments that could be seen through the SZ effect itself.

What is probably one of the most impacting expected result of MISTRAL is in fact the possibility to observe and resolve the SZ effect through galaxy clusters and surrounding medium. With respect to moderate to low angular resolution observations, high angular resolution ($\simeq$ 10 arcsec) detections can investigate non relaxed

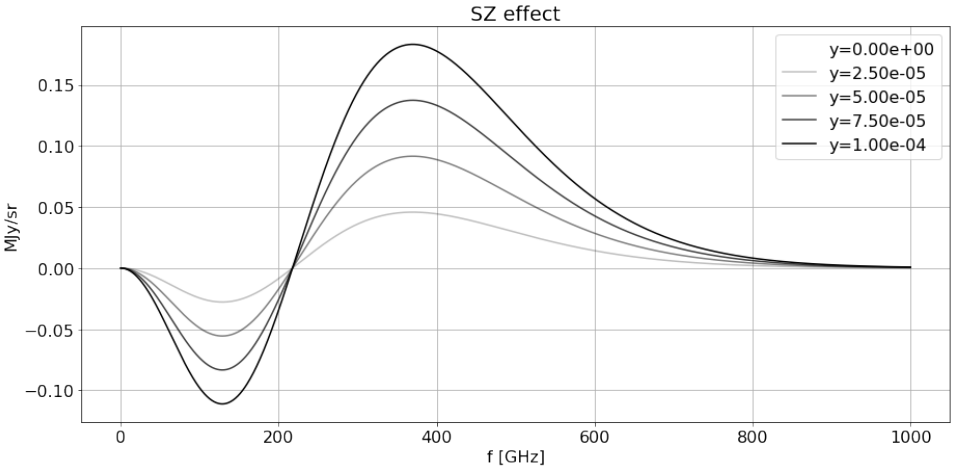


Fig. 5. The SZ effect with different Comptonization parameters.

clusters, study merging clusters, understanding the self similarities of galaxy cluster, study the pressure profiles and understand the AGN feedback expected in some environments.

Understanding the self similarities of galaxy clusters, their identical appearance or scaled properties regardless of their mass or distance is crucial. Self-similarity seems demonstrated but this is not always the case. Galaxy clusters in fact interact, collide, merge. This brings to important degeneracies and deviation from specific models. High angular resolution SZ effect measurements allow to solve the density vs temperature profile degeneracy together with X-ray information especially for high z clusters. In order to understand the physics in different environments, it is very important to study the Intra Cluster Medium (ICM) thermodynamics, including the impact of feedback, bulk and turbulent motions, substructures and cluster asphericity.

Sometimes, clusters are assumed to follow the β profile,[23] the generalized Navarro Frank and Withe (gNFW)[24] profile, or the Universal model.[25] High angular resolution SZ measurements allow to disentangle between models and to identify the most appropriate ones. Even relaxed clusters experience pressure fluctuations and compressions of the ICM. The study of pressure power spectrum allows to separate relaxed vs non-relaxed clusters where larger perturbations increase in the outer shells of the clusters (where SZ is more sensitive with respect to X-ray). Resolved SZ measurements allow to infer cluster masses from integrated Y mass (Y-M) relations. The Y-M relation is sensitive to the cluster astrophysics, radiative cooling, star formation, energy injection from stars and AGN feedback. In general, it allows to study of the pressure profiles and fluctuations of the ICM. This, again, can allow to understanding the self-similarity of galaxy clusters.

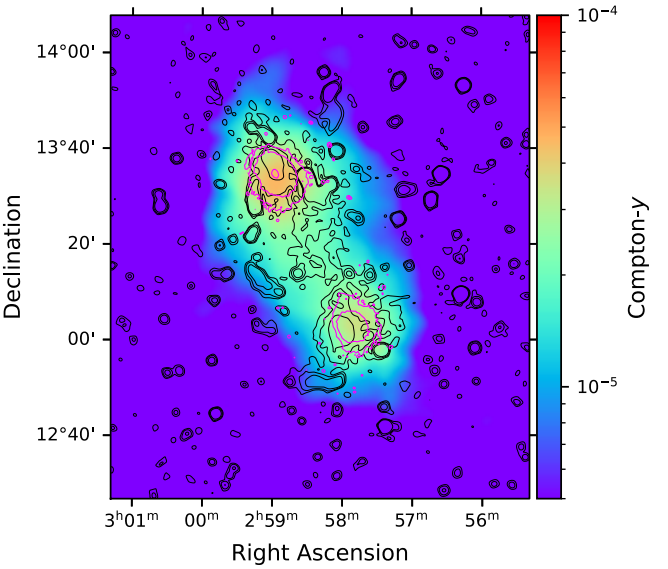


Fig. 6. Planck PR2 MILCA Compton-y map of Abell 401 (north-east) and Abell 399 (south-west). Black contour level are from the LOFAR map[26] at 0.0035, 0.008 and 0.02 Jy. Magenta contour level at 0.00008, 0.0015 and 0.001 Counts/s refer to the XMM-Newton map of the system.[27] The figure shows a good agreement between the radio and SZ signal both in the clusters and in the intracluster region, while the X-ray emission is appreciable only in correspondence with the two clusters.

In addition, point sources can be a contamination and need to be correctly subtracted not to bias the estimate of the y-parameter in a galaxy cluster. This is for instance what was achieved with MUSTANG2 at the GBT.[28]

Galaxy clusters experience hierarchical growth through mergers. Pre-mergers clusters should be connected through the Cosmic Web and simulations endorse this[21,22] (see, e.g. Figure 6). On the other hand, post mergers clusters dissipate an enormous quantity of energy with turbulence and shocks (pressure discontinuity up to mach number 3). The *Bullet* cluster or *El Gordo* are example of this unrelaxed occurrence.

The wide majority of the SZ is in fact localized towards galaxy clusters but the baryons distribution is still an open issue for modern cosmology: half of the baryons are still missing. Magneto-hydrodynamical and hydrodynamical simulations predict that they are structured in the cosmic web, filaments, and voids in the form of Warm Hot Intergalactic Medium (WHIM).[21,22] WHIM are expected to be distributed as over-desities in filamentary structures between galaxy clusters: high angular resolution SZ measurements can detect WHIM better than X-ray under low density circumstances. This is probably the most challenging and most rewarding of the achievements that an experiment like MISTRAL can achieve.

6. Conclusions

High angular resolution millimetric observation are key to understand a wide variety of scientific cases including structure formation and missing baryons issue. MISTRAL is a facility millimetric camera that will study several scientific case and that will be installed and commissioned at the Sardinia Radio Telescope during 2022. MISTRAL is going to be a unique camera: agile cryostat, 408 KIDs, optics allowing 12 arcsec resolution and a 4 arcmin f.o.v. at 90GHz. Sapienza University will do the technical commissioning, will participate to the astronomical commissioning, and will release MISTRAL to the scientific community.

Acknowledgments

The Enhancement of the Sardinia Radio Telescope (SRT) for the study of the Universe at high radio frequencies is financially supported by the National Operative Program (Programma Operativo Nazionale - PON) of the Italian Ministry of University and Research R̈esearch and Innovation 2014-2020; Notice D.D. 424 of 28/02/2018 for the granting of funding aimed at strengthening research infrastructures, in implementation of the Action II.1 - Project Proposal PIR01_00010.

References

1. S. R. Dicker, P. A. R. Ade, J. Aguirre *et al.*, MUSTANG2: a large focal plan array for the 100 meter Green Bank Telescope, in *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, eds. W. S. Holland and J. Zmuidzinas, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9153 July 2014.
2. A. Catalano, R. Adam, P. A. R. Ade, P. André, H. Aussel, A. Beelen, A. Benoît, A. Bideaud, N. Billot, O. Bourrion, *et al.*, The nika2 instrument at 30-m iram telescope: Performance and results, *Journal of Low Temperature Physics* **193**, 916 (Mar 2018).
3. R. Adam *et al.*, The nika2 large-field-of-view millimetre continuum camera for the 30 m iram telescope, *Astronomy and Astrophysics* **609**, p. A115 (January 2018).
4. L. Perotto *et al.*, Calibration and performance of the nika2 camera at the iram 30-m telescope, *Astronomy and Astrophysics* **637**, p. A71 (May 2020).
5. P. Bolli, A. Orlati, L. Stringhetti *et al.*, Sardinia radio telescope: General description, technical commissioning and first light, *Journal of Astronomical Instrumentation* **04**, p. 1550008 (Dec 2015).
6. F. T. Nasir, F. Buffa and G. L. Deiana, Characterization of the atmosphere above a site for millimeter wave astronomy, *Experimental Astronomy* **29**, 207 (Apr 2011).
7. F. Govoni *et al.*, The high-frequency upgrade of the sardinia radio telescope, *2021 XXXIVth General Assembly and Scientific Symposium of the International Union of Radio Science* **-**, 1 (- 2021).
8. A. Coppolecchia, E. Battistelli, Masi *et al.*, *Pulse tube cooler with* $> 100m$ *flexible lines optimized for operation of cryogenic detector arrays at large radiotelescopes* 19^{th} *International Workshop on Low Temperature Detectors* 2021.
9. A. Paiella *et al.*, Development of lumped element kinetic inductance detectors for the w-band, *Journal of Low Temperature Physics* **184**, 97 (July 2016).

10. A. Coppolecchia *et al.*, W-band lumped element kinetic inductance detector array for large ground-based telescopes, *Journal of Low Temperature Physics* **199**, 130 (April 2020).
11. A. Paiella, E. S. Battistelli, M. G. Castellano *et al.*, Kinetic inductance detectors and readout electronics for the olimpo experiment, *Journal of Physics: Conference Series* **1182**, p. 012005 (Feb 2019).
12. A. Paiella *et al.*, MISTRAL and its KIDs (19th International Workshop on Low Temperature Detectors (LTD19), 2021).
13. A. Paiella, E. S. Battistelli, P. de Bernardis *et al.*, The cryogenic detectors of MISTRAL *14^{th} Workshop On Low Temperature Electronics* April 2021.
14. G. Isopi, The mistral receiver: read-out electronics and observational strategies, *MASTER degree thesis* **-** (October 2021).
15. M. S. Petersen, R. A. Gutermuth, E. Nagel, G. W. Wilson and J. Lane, Early science with the large millimetre telescope: new mm-wave detections of circumstellar discs in ic 348 from lmt/aztec, *MNRAS* **488**, 1462 (Sep 2019).
16. A. D. Sokol, R. A. Gutermuth, R. Pokhrel *et al.*, Early science with the large millimetre telescope: An lmt/aztec 1.1 mm survey of dense cores in the monoceros r2 giant molecular cloud, *MNRAS* **483**, 407 (Feb 2019).
17. W. F. Wall, I. Puerari, R. Tilanus *et al.*, Continuum observations of m 51 and m 83 at 1.1 mm with aztec, *MNRAS* **459**, 1440 (Jun 2016).
18. A. Humphrey, M. Zeballos, I. Aretxaga *et al.*, Aztec 1.1-mm images of 16 radio galaxies at $0.5 < z < 5.2$ and a quasar at z= 6.3, *MNRAS* **418**, 74 (Nov 2011).
19. R. A. Sunyaev and Y. B. Zeldovich, The observations of relic radiation as a test of the nature of x-ray radiation from the clusters of galaxies, *Comments on Astrophysics and Space Physics* (Nov 1972).
20. M. Birkinshaw, The sunyaev-zel'dovich effect, *Physics Reports* **310**, p. 97–195 (Mar 1999).
21. R. Cen and J. P. Ostriker, Where are the baryons?, *The Astrophysical Journal* **514**, 1 (Mar 1999).
22. T. Tuominen, J. Nevalainen, E. Tempel *et al.*, An eagle view of the missing baryons, *Astronomy & Astrophysics* **646**, p. A156 (Feb 2021).
23. A. Cavaliere and R. Fusco-Femiano, Reprint of 1976A&A....49..137C. X-rays from hot plasma in clusters of galaxies., *aap* **500**, 95 (May 1976).
24. J. F. Navarro, C. S. Frenk and S. D. M. White, The structure of cold dark matter halos, *The Astrophysical Journal* **462**, p. 563 (May 1996).
25. M. Arnaud, The β-model of the intracluster medium. Commentary on: Cavaliere A. and Fusco-Femiano R., 1976, A&A, 49, 137, *aap* **500**, 103 (June 2009).
26. F. Govoni *et al.*, A radio ridge connecting two galaxy clusters in a filament of the cosmic web, *Science* **364**, 981 (June 2019).
27. M. Murgia *et al.*, A double radio halo in the close pair of galaxy clusters abell 399 and abell 401, *Astronomy and Astrophysics* **509**, p. A86 (January 2010).
28. S. R. Dicker, E. S. Battistelli, T. Bhandarkar *et al.*, Observations of compact sources in galaxy clusters using mustang2, *MNRAS on line* **-** (Sep 2021).

Cosmological and astrophysical results exploiting magnification bias with high-z sub-millimetre galaxies

L. Bonavera[1,2,*], M.M. Cueli[1,2,†] and J. Gonzalez-Nuevo[1,2,‡]

[1] *Physics Department, University of Oviedo, Oviedo, 33007, Spain*

[2] *Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), C. Independencia 13, 33004 Oviedo, Spain*

*E-mail: * bonaveralaura@uniovi.es, † mcueli@uniovi.es, ‡ gnuevo@uniovi.es*

The high-z submillimeter galaxies (SMGs) can be used as background sample for gravitational lensing studies thanks to their magnification bias, which can manifest itself through a non-negligible measurement of the cross-correlation function between a background and a foreground source sample with non-overlapping redshift distributions. In particular, the choice of SMGs as background sample enhances the cross-correlation signal so as to provide an alternative and independent observable for cosmological studies regarding the probing of mass distribution.

In particular the magnification bias can be exploited in order to constrain the free astrophysical parameters of a Halo Occupation Distribution model and some of the main cosmological parameters. Urged by the improvements obtained when adopting a pseudo-tomographic analysis, It has been adopted a tomographic set-up to explore not only a ΛCDM scenario, but also the possible time evolution of the dark energy density in the ω_0CDM and $\omega_0\omega_a$CDM frameworks.

Keywords: Cosmology; Gravitational lensing; Sub-millimeter galaxies.

1. Magnification bias

Magnification bias is due to the gravitational lensing effect and consists in the apparent excess in the number of high redshift sources (the lensed sources) close to the position of low redshift galaxies (the lenses). In fact, the gravitational lensing deflects the light rays of high redshift sources causing the stretching of the apparent sky area in the region affected by the lensing. This also causes a boost in the flux density of the high redshift sources, making them more likely to be detected above a given instrument flux density limit (e.g. Ref. 1). This is sketched in Fig. 1(a). The red objects in the backward plane (source plane) are the sources with flux density limit above the instrument detection limit, and thus detectable while the orange ones are those with a flux density below the instrument detection limit, and thus not detectable unless boosted by the magnification bias. The blue objects in the lens plane are the possible lenses that might cause the magnification bias effect. Finally in the observer plane, the blue objects of the lens plane will be observable together with the red objects of the source plane and some of the orange objects whose flux have been boosted. The peculiarity of such magnified orange objects is that they will be found close to some of the blue objects (their lenses).

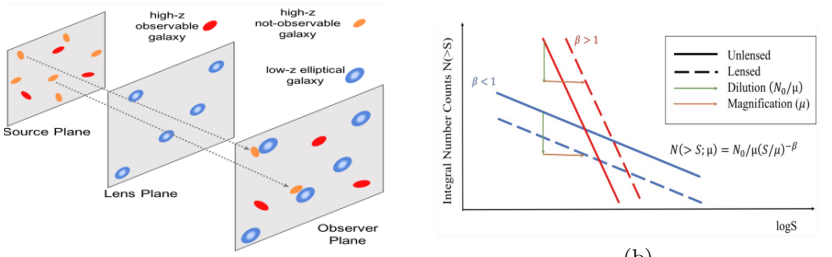


Fig. 1. Scheme of magnification bias (a) and how it affect the integral number counts of the lensed sources (b). See text for details.

On this respect, magnification bias main advantage relies on the steepness of background sources counts and on the fact that it does not require either ellipticity measurements or knowledge on the galaxies' orientation (as shear does). In particular, given that n_0 are the unlensed integrated background source number counts (number of background sources per solid angle and redshift with observed flux density larger than S in the absence of gravitational lensing):

$$n_0(>S,z) \equiv \int_S^\infty \frac{dN}{dS dz\, d\Omega} dS \tag{1}$$

the background source number counts at an angular position $\vec{\theta}$ within an image are modified by dilution and magnification following Ref. 2:

$$n(>S,z;\vec{\theta}) = \frac{1}{\mu(\vec{\theta})} n_0\left(> \frac{S}{\mu(\vec{\theta})}, z \right) \tag{2}$$

where $\mu(\vec{\theta})$ is the magnification field at angular position $\vec{\theta}$. Assuming a power-law behaviour of the unlensed integrated number counts with β the source number count steepness, i.e. $n_0(>S,z) = AS^{-\beta}$,

$$\frac{n(>S,z;\vec{\theta})}{n_0(>S,z)} = \mu^{\beta-1}(\vec{\theta}) \tag{3}$$

Depending on the value of β the source counts might increase because of lensing: Fig. 1(b) sketched the increase or decrease of the source number counts for $\beta > 1$ (in red) and $\beta < 1$ (in blue), respectively.

2. Cross-correlation

Magnification bias is usually observed by means of angular cross-correlation function (CCF) measurements between two samples of sources: how many background sources fall 'near' the position of a lens, repeated for each object in the lens sample. Since the samples are at different redshifts, the detection of a non-zero signal would be exclusively due to the gravitational lensing induced by the lenses.

2.1. *The estimator*

Thus, magnification bias can be measured through the cross-correlation between the low and high redshift objects.

The CCF of two source populations (D_f and D_b) is the fractional excess probability, relative to a random distribution (R_f and R_b), of finding a D_f source separated by an angle θ from a D_b source within an infinitesimal solid angle $d\theta$. The CCF can be computed using a modified version of the Landy & Szalay (Ref. 3) estimator (see Ref. 4):

$$\tilde{w}_{fb}(\theta) = \frac{D_f D_b(\theta) - D_f R_b(\theta) - D_b R_f(\theta) + R_f R_b(\theta)}{R_f R_b(\theta)} \tag{4}$$

In particular, $D_f D_b$, $D_f R_b$, $D_b R_f$ and $R_f R_b$ are the normalized foreground-background, foreground-Poisson, background-Poisson and Poisson-Poisson pair counts for a given angular separation θ. Given two source samples with non-overlapping redshift distributions, the excess signal when computing their cross-correlation with respect to the random case is due to the magnification bias. Being such signal related to lensing and thus to cosmological distances and the galaxy halo characteristics, it can be used to constrain cosmological and astrophysical parameters.

2.2. *The halo model*

To achieve this, the adopted theoretical description of the cross-correlation is the one by Cooray & Sheth (Ref. 5):

$$w_{fb}(\theta) = 2(\beta - 1)\int_0^\infty \frac{dz}{\chi^2(z)} n_f(z) W^{\text{lens}}(z) \int_0^\infty l \frac{dl}{2\pi} P_{\text{gal-dm}}(l/\chi^2(z), z) J_0(l\theta) \tag{5}$$

$$W^{\text{lens}}(z) = \frac{3}{2}\frac{H_0^2}{c^2}\left[\frac{E(z)}{1+z}\right]^2 \int_z^\infty dz' \frac{\chi(z)\chi(z'-z)}{\chi(z')} n_b(z') \tag{6}$$

where l is the multipole, H_0 is the Hubble constant, $E(z)$ the quantity where the contribution to the energy density are included (see section 2.3 for details), $n_b(z)$ ($n_f(z)$) is the normalized background (foreground) redshift distribution, $\chi(z)$ is the comoving distance, J_0 is the zeroth-order Bessel function of the first kind, $W^{\text{lens}}(z)$ is the lensing kernel and β is the source number count steepness, commonly fixed to 3 for submillimeter galaxies (Ref. 6 and reference therein). Therefore, the cross-correlation signal can be interpreted under the halo model parametrization (Ref. 5) considering that both galaxy samples trace the same dark matter distribution around redshift $z \sim 0.4$. This dark matter distribution is traced directly by the foreground galaxies while, in the case of the background sample, it is traced thanks to the weak lensing effect. According to this model, the galaxy distribution power spectrum is parametrized as the sum of a 2-halo term, related to the correlations between one halo traced by the foreground galaxies and another one traced by the

background sources and that therefore dominates at large scales, and a 1-halo term, that describes the correlation between sub-halos (traced by both samples) inside the same halo and therefore more important at small scales. Moreover, the halo model also suggests a simple parametrization of the cross-correlation between the galaxy and dark matter distributions (see Refs. 5,7,8). According to the halo model (Cooray & Sheth, Refs. 5):

$$P_{\text{g-dm}}(k,z) = P^{1h}_{\text{g-dm}}(k,z) + P^{2h}_{\text{g-dm}}(k,z) \tag{7}$$

$$P^{\text{1h}}_{\text{g-dm}}(k,z) = \int_0^\infty dM\, M \frac{n(M,z)}{\bar{\rho}(z)} \frac{\langle N_g \rangle_M}{\bar{n}_g(z)} |u_{\text{dm}}(k,z|M)||u_{\text{g}}(k,z|M)|^{p-1}$$

$$P^{\text{2h}}_{\text{g-dm}}(k,z) = P^{\text{lin}}(k,z)\Big[\int_0^\infty dM\, M \frac{n(M,z)}{\bar{\rho}(z)} b_1(M,z) u_{\text{dm}}(k,z|M)\Big] \cdot$$
$$\cdot \Big[\int_0^\infty dM\, n(M,z) b_1(M,z) \frac{\langle N_g \rangle_M}{\bar{n}_g(z)} u_g(k,z|M)\Big]$$

In these equations $\bar{\rho}$ is the background density, $\langle N_g \rangle_M$ the mean number of galaxies within a halo of mass M, $\bar{n}_g(z)$ the mean number density of galaxies at redshift z, p is set to 1 for central galaxies and to 2 for satellites (Ref. 5), k is the wave number, $n(M,z)$ is the Sheth & Tormen halo mass function by Ref. 9, $P^{lin}(k,z)$ is the linear matter power spectrum, $b_1(M,z)$ is the linear large-scale bias (via the peak background split), and $u_g(k,z|M)$ is the normalized Fourier transform of the galaxy density distribution within a halo, which is assumed to equal the dark matter density profile, i.e. $u_g(k,z|M) = u_{dm}(k,z|M)$. Halos are defined as overdense regions whose mean density is 200 times the mean background density of the universe according to the spherical collapse model, and the halo density profiles $\rho(r)$ adopted is the one by Ref. 10, hereafter NFW, with the concentration parameter in Ref. 11.

The halo occupation distribution (HOD) assumed is the one by Zheng et al. in Ref. 12:

$$N_{\text{cen}}(M_h) = \begin{cases} 0 & \text{if } M_h < M_{\text{min}} \\ 1 & \text{otherwise} \end{cases} \tag{8}$$

$$N_{\text{sat}}(M_h) = N_{\text{cen}}(M_h) \cdot \left(\frac{M_h}{M_1}\right)^{\alpha} \tag{9}$$

In this model, the mean number of galaxies is represented by N_{gal} where, the distinction between central and satellite galaxies is made: $N_{gal} = N_{cen} + N_{sat} = 1 + N_{sat}$. All halos above a minimum mass M_{min} host a galaxy at their centre, while any remaining galaxies are classified as satellites and are distributed in proportion to the halo mass profile (see Ref. 12). Halos host satellites when their mass exceeds the M_1 mass, and the number of satellites is a power-law function of halo mass $N_{\text{sat}}(M_h)$. These parameters define the adopted HOD.

2.3. *The barotropic index of the dark energy*

In the theoretical model, a possible redshift evolution of the dark energy (DE) can be introduced by allowing different values for the dark energy equation of state parameter, ω. For a flat cosmology, a redshift-dependent function $f(z)$ in the quantity

$$E(z) \equiv \sqrt{\Omega_r(1+z)^4 + \Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\rm DE} f(z)} \tag{10}$$

takes into account the possible time evolution of the dark energy density, where Ω_x is the present-day density parameters for radiation r, matter (M), curvature (k) and dark energy DE. In particular, by assuming that the barotropic index of dark energy

$$\omega(z) = \omega_0 + \omega_a \frac{z}{1+z} \tag{11}$$

the dark energy density has a redshift dependence given by $\rho(z) = \rho_0 f(z)$, where

$$f(z) = (1+z)^{3(1+\omega_0+\omega_a)} e^{-3\omega_a \frac{z}{1+z}} \tag{12}$$

being ρ_0 the dark energy density at $z = 0$. As for the dark energy density parameter, it can be written as

$$\Omega_{\rm DE}(z) = \frac{\Omega_{\rm DE} f(z)}{E(z)^2} \tag{13}$$

In this framework, the ΛCDM is recovered when $\omega_0 = -1$ and $\omega_a = 0$.

3. Background and foreground samples

In order to achieve the cross-correlation measurement, a clear separation of the background and foreground sources is needed. For example, in the recent works

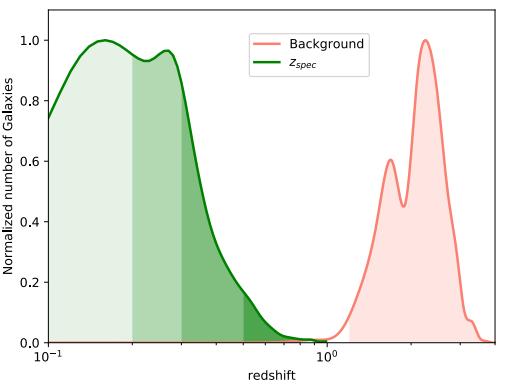


Fig. 2. Redshift distribution of background (orange) and foreground objects (green). The different shades of green indicate the bins of redshift used in Ref. 13. Credit: Bonavera et al., A&A, in press, DOI 10.1051/0004-6361/202141521, 2021, reproduced with permission © ESO.

by Bonavera et al. (Refs. 6, 13), the redshift distribution of the foreground and background galaxies sample is the one in Fig. 2 in green and orange respectively. The different shades of green indicate the bins of redshift used in Ref. 13.

As for the foreground sample, different kind of sources at relatively low redshift can be used. For example, in their recent work (Ref. 14) Bonavera et al. adopt the same sample of quasi-stellar object (QSOs) with $0.2 < z < 1.0$ as in Ref. 15. They are selected from the publicly available Sloan Digital Sky Survey (SDSS-II and SDSS-III) Baryon Oscillation Spectroscopic Survey (BOSS) catalogues. In Refs. 6, 13, 16–18 the galaxies in the Galaxy and Mass Assembly (GAMA, Ref. 19) survey with $0.2 < z < 0.8$ have been used and in Ref. 13 those with $0.1 < z < 0.8$ (see Fig. 2).

Currently the best way to measure the CCF is to choose as background samples the high-z sub-millimetre galaxies (SMGs). In fact they have steep number counts $\beta \sim 3$ (Refs. 20, 21), shown by the Herschel Space Observatory (Herschel; Ref. 22) and the South Pole Telescope (SPT; Ref. 23) observations. Such property enhances their magnification, as explained in Sect. 1. Moreover, the SMGs are faint in the optical avoiding the possibility of being confused with the foreground lens sample, which is on its turn invisible at sub-millimetre (sub-mm) wavelength (Refs. 24, 25). Finally, the redshifts of the SMGs are usually greater than $z > 1 - 1.5$, which guarantees no overlap with the foreground sample. This makes the SMGs the perfect background sample for magnification bias studies by means of CCF measurements and thus allowing the tracing of the mass density (baryonic and dark matter) and its evolution with time.

Moreover, it has been confirmed when, due to their magnification bias, Dunne et al. in Ref. 26 made a serendipitous direct observation of high-redshift SMGs. They were performing a study of gas tracers with Atacama Large Millimeter Array (ALMA) observation of galaxies by targeting a statistically complete sample of twelve galaxies selected at 250μm with z=0.35 and magnified SMGs appears around the position of half of them.

4. Recent results with SMGs magnification bias

The first attempt to measure the CCF was carried out in Ref. 27 with Herschel/SPIRE galaxies and low-z galaxies as background and foreground samples, respectively. Their results were a strong confirmation of this lensing-induced effect.

To study this possible bias, Ref. 28 relied on much better statistics and carried out a more detailed analysis by measuring the CCF with Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS) high-z sources at $z > 1.5$ and two optical samples selected from SDSS (Sloan Digital Sky Survey, Ref. 27) and GAMA (Ref. 19) surveys, with redshifts $0.2 < z < 0.6$. The resulting CCF was measured with high significance, $> 10\sigma$ and using realistic simulations, they concluded that the signal was entirely explained with the magnification bias produced by the weak lensing effect caused by galaxy groups/clusters whose halo masses are in the range

of $10^{13.2} - 10^{14.5} M_{\odot}$ and that are signposted by the brightest galaxies in the optical samples. Later on, Gonazelz-Nuevo et al. in Ref. 16 and Bonavera et al. in Ref. 13 showed furthermore that the SMGs properties make them the perfect background sample for constraining the free parameters of a halo occupation distribution (HOD) model. In particular, the results by Ref. 16 suggest that the lenses are massive galaxies or clusters, with a minimum mass of $M_{min} = 10^{13} M_{\odot}$. They also carried out tomographic studies on the HOD parameters: they divided the foreground sample in four bins of redshift, $0.1 < z < 0.2$, $0.2 < z < 0.3$, $0.3 < z < 0.5$ and $0.5 < z < 0.8$. They main findings were that while M_1 is almost redshift independent, M_{min} evolves: it increases with redshift, as predicted by theoretical estimations. Bonavera et al. in Ref. 14 study with the magnification bias the mass properties of a sample of QSOs whose position signpost the lenses at $0.2 < z < 1.0$ obtaining $M_{min} = 10^{13.6^{+0.9}_{-0.4}} M_{\odot}$. This again suggests that the lensing is actually produced by halos of the size of clusters placed close to the QSOs positions.

Furthermore, magnification bias trough CCF measurements can be used to constrain the halo mass function. In particular, Cueli et al. in Ref. 18 successfully test such possibility according to two common parametrisations: the Sheth & Tormen and Tinker fits. They find general agreement with traditional values for the involved parameters, with a slight difference in the Sheth & Tormen fit for intermediate and high masses, where the results suggest a hint at a somewhat higher number of halos.

5. Cosmological studies with magnification bias

Moreover, some of the main cosmological parameters can be also estimated using the magnification bias, as in Ref. 6. They use the CCF measured between a foreground sample of GAMA galaxies with spectroscopic redshifts in the range of $0.2 < z < 0.8$ and a background sample of H-ATLAS galaxies with photometric redshifts $z > 1.2$ (as described in section 3) to constrain the astrophysical parameters (M_{min}, M_1, and α) and the Ω_M, σ_8, and H_0 cosmological ones (the matter density parameter, the present root-mean-square matter fluctuation averaged over a sphere of radius $8h^{-1}$Mpc where h is the dimensionless Hubble constant, and the Hubble density parameters). These parameters are estimated through a Markov chain Monte Carlo analysis, using the Python package `emcee` (Ref. 29). They study various cases and in particular they perform a run by setting flat priors to those in Table 1.

Table 1. Flat priors for the ΛCDM run in Ref. 6.

$\log M_{min}$	$\mathcal{U}[\ 11.6,\ 13.6]$	Ω_M	$\mathcal{U}[\ 0.1,\ 0.8]$
$\log M_1$	$\mathcal{U}[\ 13.0,\ 14.5]$	σ_8	$\mathcal{U}[\ 0.6,\ 1.2]$
α	$\mathcal{U}[\ 0.5,\ 1.37]$	h	$\mathcal{U}[\ 0.5,\ 1.0]$

Note: The HOD parameters names and flat priors are listed in the first two columns. The cosmological ones in the third and fourth columns.

They obtain a lower limit at 95% confidence level (CL) on $\Omega_M > 0.24$, a slight trend towards $H_0 > 70$ km s^{-1} Mpc^{-1} values and an upper limit at 95% CL of $\sigma_8 < 1$. Such results are summarised in the $\Omega_M - \sigma_8$ plane in Fig. 3 (a) with the dot-dashed blue line.

Given the large number of studies that can be performed with magnification bias through CCF data, a systematic analysis of possible bias that might affect the estimation have been carried out. In particular, in Ref. 17 Gonzalez-Nuevo et al. take into account different biases at cosmological scales in the source samples to carefully measure unbiased CCF. More precisely, their background sample consists of H-ATLAS galaxies with $z > 1.2$ whereas they use two independent foreground samples with $0.2 < z < 0.8$: GAMA galaxies with spectroscopic redshifts and SDSS galaxies with photometric redshifts. These independent samples allowed them to perform a pseudo-tomographic study that yields to constrain $\Omega_M = 0.50^{+0.14}_{-0.20}$ and $\sigma_8 = 0.75^{+0.07}_{-0.10}$. Such analysis also suggests that a tomographic approach might improve the results.

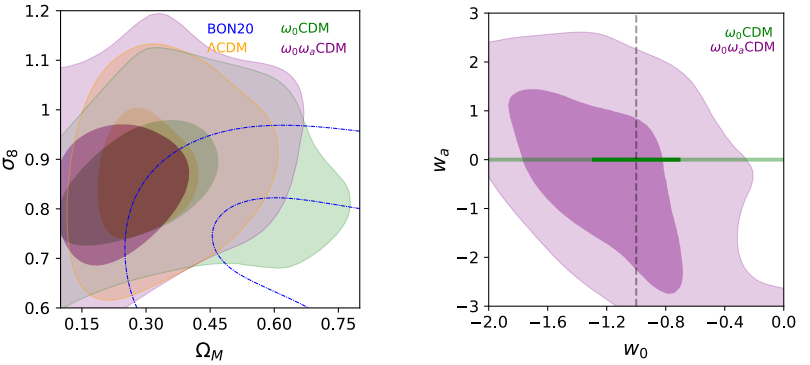
Driven by the conclusions mentioned above, Bonavera et al. in Ref. 13 adopt the unbiased sample by Gonzalez-Nuevo et al. in Ref. 17 dividing the foreground sample into four redshift bins ($0.1-0.2$, $0.2-0.3$, $0.3-0.5$ and $0.5-0.8$). The background objects are a sample of H-ATLAS galaxies with photometric redshifts $z > 1.2$. The redshift distribution are shown in Fig. 2 and the redshift bins of the foreground sample are highlighted with different shades of green. They use magnification bias in tomography to jointly constrain the astrophysical HOD parameters (M_{min}, M_1 and α) in each one of the selected redshift bins together with the Ω_M, σ_8, and H_0 cosmological ones. In particular, the analysis is carried out both in the ΛCDM scenario and with the introduction of the dark energy density related ω_0 and ω_a parameters in the ω_0CDM and $\omega_0\omega_a$CDM frameworks. The parameters priors for the three described runs are listed in Table 2.

Table 2. Flat priors for the ΛCDM, ω_0CDM and $\omega_0\omega_a$CDM models in Ref. 13.

$\log M_{min}\mathcal{U}[a_i, b_i]$	$\log M_1\mathcal{U}[c_i, d_i]$	Ω_M	$\mathcal{U}$[0.1, 0.8]
$[a_1, b_1] = \mathcal{U}[10.0, 13.0]$	$[c_1, d_1] = \mathcal{U}[12.0, 15.5]$	σ_8	$\mathcal{U}$[0.6, 1.2]
$[a_2, b_2] = \mathcal{U}[11.0, 13.0]$	$[c_2, d_2] = \mathcal{U}[12.0, 15.5]$	h	$\mathcal{U}$[0.5, 1.0]
$[a_3, b_3] = \mathcal{U}[11.5, 13.5]$	$[c_3, d_3] = \mathcal{U}[12.5, 15.5]$	ω_a	$\mathcal{U}$[-3.0, 3.0]
$[a_4, b_4] = \mathcal{U}[13.0, 15.5]$	$[c_4, d_4] = \mathcal{U}[13.0, 15.5]$	ω_0	$\mathcal{U}$[-2.0, 0.0]

Note: For the four different bins of redshift, the M_{min} flat priors are listed in the first column while the M_1 ones in the second column. The α prior is $\alpha = \mathcal{U}[\ 0.5,\ 1.37]$ in all bins. The cosmological parameters names and flat priors are in the third and fourth columns.

As for the HOD parameters, M_{min} shows a trend towards higher values at higher redshift confirming the findings in Ref. 16. For the ΛCDM model, they obtain a mean(maximum) posterior value [68% CL] of $\Omega_M = 0.33(0.26)$ [0.17,0.41]



(a) (b)

Fig. 3. Contour plot of the 2D posterior distributions with contours set to 0.393 and 0.865. (a) Ω_M and σ_8 for the non-tomographic case in Ref. 6 (dot-dashed blue line) and the tomographic runs in Ref. 13 within ΛCDM (in yellow), ω_0CDM (in green) and $\omega_0\omega_a$CDM (in purple). (b) ω_0 and ω_a for the $\omega_0\omega_a$CDM model. Credit: Bonavera et al., A&A, in press, DOI 10.1051/0004-6361/202141521, 2021, reproduced with permission © ESO.

and of σ_8= 0.87 [0.75,1], being H_0 not yet constrained. For the ω_0CDM model, they find similar results on Ω_M and σ_8 and a mean(maximum) posterior value [68% CL] of $\omega_0 = -1(-0.97)$ $[-1.56, -0.47]$. For the $\omega_0\omega_a$CDM model, $\omega_0 = -1.09(-0.92)$ $[-1.72, -0.66]$ and $\omega_a = -0.19(-0.20)$ $[-1.88, 1.48]$. These results are summarised in Tables 3 and 4, where the the mean, peak and 68%CL of the posterior distributions for the estimated cosmological parameters are given. In particular, the first column gives the parameter name, the second and third columns in Table 3 are the results for the ΛCDM model in the non-tomographic and tomographic cases, and in Table 4 are those for the ω_0CDM and $\omega_0\omega_a$CDM models in the tomographic case.

The $\Omega_M - \sigma_8$ plane is shown in Fig. 3(a) for the ΛCDM (in yellow), ω_0CDM (in green) and $\omega_0\omega_a$CDM (in purple) models. The results in the $\omega_0 - \omega_a$ plane are shown in Fig. 3(b). The ω_0 results are shown in Fig. 4 for the ω_0CDM (in green) and $\omega_0\omega_a$CDM (in purple) cases, where they are compared with other results from literature. Moreover, the tomographic analysis presented in Ref. 13 confirms that magnification bias results do not show the degeneracy found with cosmic shear measurements and that, related to dark energy, do show a trend of higher ω_0 values for lower H_0 values.

6. Conclusion

In conclusion, all these works confirm that the SMGs are a perfect background sample for magnification bias studies and serendipitous direct observation of magnified SMGs have been performed with ALMA (i.e. Ref. 26). It has been demonstrated in Ref. 28 and Ref. 16 for astrophysical studies and in Ref. 6 for cosmological ones.

Table 3. Cosmological results for the ΛCDM from Bonavera et al. in Refs. 6, 13.

	ΛCDM nontomo		ΛCDM	
	$\mu(Peak)$	$68\%CL$	$\mu(Peak)$	$68\%CL$
Ω_M	0.54(0.67)	[0.46,0.80]	0.33(0.26)	[0.17,0.41]
σ_8	0.78(0.74)	[0.63,0.85]	0.87(0.87)	[0.75,1.00]
h	0.76(-)	[0.68,1.00]	0.72(0.72)	<0.79

Note: Mean(Peak) and $68\%CL$ of the posterior distributions for the cosmological parameters (listed in the first column) estimated according to the ΛCDM model in the non-tomographic case (second and third columns) and in the tomographic one (fourth and fifth columns) models in the tomographic case. With the exception of the estimated cosmological parameters, the cosmology is fixed to the *Planck* one.

Table 4. Cosmological results for the ω_0CDM and $\omega_0\omega_a$CDM models from Bonavera et al. in Ref. 13.

	ω_0CDM		$\omega_0\omega_a$CDM	
	$\mu(Peak)$	$68\%CL$	$\mu(Peak)$	$68\%CL$
Ω_M	0.38(0.26)	[0.13,0.47]	0.34(0.21)	[0.11,0.41]
σ_8	0.87(0.85)	[0.73,0.98]	0.88(0.84)	[0.72,1.01]
h	0.70(-)	<0.75	0.70(-)	<0.76
ω_0	-1.00(-0.97)	[-1.56,-0.47]	-1.09(-0.92)	[-1.72,-0.66]
ω_a	-	-	-0.19(-0.20)	[-1.88,1.48]

Note: Mean(Peak) and $68\%CL$ of the posterior distributions for the cosmological parameters (listed in the first column) estimated according to the ω_0CDM (second and third columns) and $\omega_0\omega_a$CDM (fourth and fifth columns) models in the tomographic case. With the exception of the estimated cosmological parameters, the cosmology is fixed to the *Planck* one.

In particular magnification bias proves to be useful as an independent and additional cosmological probe for ΛCDM and beyond ΛCDM models, as done in Ref. 13.

In these works, the performance of magnification bias through CCF measurements has been compared in non-tomographic and tomographic analysis. The conclusion is that tomography improves non-tomographic studies despite the worsening in the statistics of the measured CCFs.

Moreover, the CCF can be computed adopting different foreground samples, e.g. galaxies in Ref. 6 or QSOs in Ref. 14 and the measurements are improved by taking into account observational biases correction (i.e. Ref. 17). In addition, variation of the ingredients of the models, as the HMF, might be tested with magnification bias (e.g. Ref. 18).

Up to now, the main conclusions are the fact that astrophysical results are robust for all models (ΛCDM and beyond ΛCDM) and that ΛCDM compatible

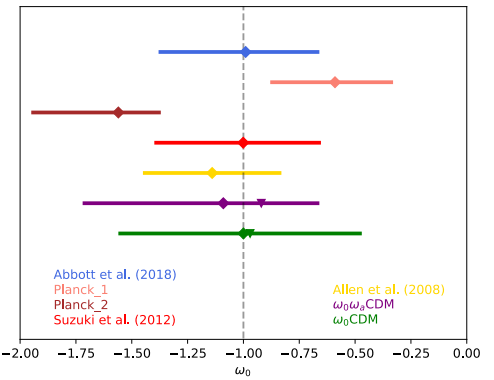


Fig. 4. Results on ω_0 within the ω_0CDM (green) and $\omega_0\omega_a$CDM (purple) models compared with those by DES (blue), Planck1 (basewwaplikHMTTlowllowEBAO, salmon), Planck2 (basewplikHMTTlowllowE, brown), Supernovae (red), x-ray measurements (yellow). Credit: Bonavera et al., A&A, in press, DOI 10.1051/0004-6361/202141521, 2021, reproduced with permission © ESO.

results have been found with tomographic and non-tomographic approaches, i.e. $\Omega_M \sim 0.33$, $\sigma_8 \sim 0.87$, $\omega_0 \sim -1$ and $\omega_a \sim -0.19$.

Acknowledgments

LB, JGN and MMC acknowledge the PGC 2018 project PGC2018-101948-B-I00 (MICINN/FEDER). MMC acknowledges support from PAPI-20-PF-23 (Universidad de Oviedo). LB acknowledges Carlo Burigana forhis help in the clarity of the manuscript.

References

1. P. Schneider, J. Ehlers and E. E. Falco, *Gravitational Lenses* 1992.
2. M. Bartelmann and P. Schneider, Weak gravitational lensing, *Physics Reports* **340**, 291 (2001).
3. S. D. Landy and A. S. Szalay, Bias and Variance of Angular Correlation Functions, *The Astrophysical Journal* **412**, p. 64 (July 1993).
4. D. Herranz, Foreground-Background Galaxy Correlations in the Hubble Deep Fields, in *Cosmological Physics with Gravitational Lensing*, eds. J. Tran Thanh Van, Y. Mellier and M. Moniez January 2001.
5. A. Cooray and R. Sheth, Halo models of large scale structure, *Physics Reports* **372**, 1 (Dec 2002).
6. L. Bonavera, J. González-Nuevo, M. M. Cueli, T. Ronconi, M. Migliaccio, L. Dunne, A. Lapi, S. J. Maddox and M. Negrello, Cosmology with the submillimetre galaxies magnification bias: Proof of concept, *Astronomy & Astrophysics* **639**, p. A128 (July 2020).
7. U. Seljak, Analytic model for galaxy and dark matter clustering, *Monthly Notices of the Royal Astronomical Society* **318**, 203 (October 2000).

8. J. Guzik and U. Seljak, Galaxy-dark matter correlations applied to galaxy-galaxy lensing: predictions from the semi-analytic galaxy formation models, *Monthly Notices of the Royal Astronomical Society* **321**, 439 (March 2001).
9. R. K. Sheth and G. Tormen, Large-scale bias and the peak background split, *Monthly Notices of the Royal Astronomical Society* **308**, 119 (September 1999).
10. J. F. Navarro, C. S. Frenk and S. D. M. White, The Structure of Cold Dark Matter Halos, *The Astrophysical Journal* **462**, p. 563 (May 1996).
11. J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack and A. Dekel, Profiles of dark haloes: evolution, scatter and environment, *Monthly Notices of the Royal Astronomical Society* **321**, 559 (March 2001).
12. Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, N. Katz and C. G. Lacey, Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies, *The Astrophysical Journal* **633**, 791 (November 2005).
13. L. Bonavera, M. M. Cueli, J. González-Nuevo, T. Ronconi, M. Migliaccio, A. Lapi, J. M. Casas and D. Crespo, Cosmology with the sub-millimetre galaxies magnification bias: Tomographic Analysis, *arXiv e-prints* , p. arXiv:2109.12413 (September 2021).
14. L. Bonavera, J. Gonz'alez-Nuevo, S. L. Suárez Gómez, A. Lapi, F. Bianchini, M. Negrello, E. Díez Alonso, J. D. Santos and F. J. de Cos Juez, QSOs sigposting cluster size halos as gravitational lenses: halo mass, projected mass density profile and concentration at z˜0.7, *Journal of Cosmology and Astroparticle Physics* **2019**, p. 021 (September 2019).
15. F. Bianchini, G. Fabbian, A. Lapi, J. Gonzalez-Nuevo, R. Gilli and C. Baccigalupi, Broadband Spectral Energy Distributions of SDSS-selected Quasars and of Their Host Galaxies: Intense Activity at the Onset of AGN Feedback, *The Astrophysical Journal* **871**, p. 136 (February 2019).
16. J. González-Nuevo, A. Lapi, L. Bonavera, L. Danese, G. de Zotti, M. Negrello, N. Bourne, A. Cooray, L. Dunne, S. Dye, S. Eales, C. Furlanetto, R. J. Ivison, J. Loveday, S. Maddox, M. W. L. Smith and E. Valiante, H-ATLAS/GAMA: magnification bias tomography. Astrophysical constraints above ˜1 arcmin, *Journal of Cosmology and Astroparticle Physics* **2017**, p. 024 (October 2017).
17. J. González-Nuevo, M. M. Cueli, L. Bonavera, A. Lapi, M. Migliaccio, F. Argüeso and L. Toffolatti, Cosmological constraints on the magnification bias on sub-millimetre galaxies after large-scale bias corrections, *Astronomy & Astrophysics* **646**, p. A152 (February 2021).
18. M. M. Cueli, L. Bonavera, J. González-Nuevo and A. Lapi, A direct and robust method to observationally constrain the halo mass function via the submillimeter magnification bias: Proof of concept, *Astronomy & Astrophysics* **645**, p. A126 (January 2021).
19. S. P. Driver, D. T. Hill, L. S. Kelvin, A. S. G. Robotham, J. Liske, P. Norberg, I. K. Baldry, S. P. Bamford, A. M. Hopkins, J. Loveday, J. A. Peacock, E. Andrae, J. Bland-Hawthorn, S. Brough, M. J. I. Brown, E. Cameron, J. H. Y. Ching, M. Colless, C. J. Conselice, S. M. Croom, N. J. G. Cross, R. de Propris, S. Dye, M. J. Drinkwater, S. Ellis, A. W. Graham, M. W. Grootes, M. Gunawardhana, D. H. Jones, E. van Kampen, C. Maraston, R. C. Nichol, H. R. Parkinson, S. Phillipps, K. Pimbblet, C. C. Popescu, M. Prescott, I. G. Roseboom, E. M. Sadler, A. E. Sansom, R. G. Sharp, D. J. B. Smith, E. Taylor, D. Thomas, R. J. Tuffs, D. Wijesinghe, L. Dunne, C. S. Frenk, M. J. Jarvis, B. F. Madore, M. J. Meyer, M. Seibert, L. Staveley-Smith, W. J. Sutherland and S. J. Warren, Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release, *Monthly Notices of the Royal Astronomical Society* **413**, 971 (May 2011).

20. A. W. Blain, Galaxy-galaxy gravitational lensing in the millimetre/submillimetre waveband, *Monthly Notices of the Royal Astronomical Society* **283**, 1340 (December 1996).
21. M. Negrello, F. Perrotta, J. González-Nuevo, L. Silva, G. de Zotti, G. L. Granato, C. Baccigalupi and L. Danese, Astrophysical and cosmological information from large-scale submillimetre surveys of extragalactic sources, *Monthly Notices of the Royal Astronomical Society* **377**, 1557 (June 2007).
22. G. L. Pilbratt, J. R. Riedinger, T. Passvogel, G. Crone, D. Doyle, U. Gageur, A. M. Heras, C. Jewell, L. Metcalfe, S. Ott and M. Schmidt, Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy, *Astronomy & Astrophysics* **518**, p. L1 (July 2010).
23. J. E. Carlstrom, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, S. Busetti, C. L. Chang, E. Chauvin, H. M. Cho, T. M. Crawford, A. T. Crites, M. A. Dobbs, N. W. Halverson, S. Heimsath, W. L. Holzapfel, J. D. Hrubes, M. Joy, R. Keisler, T. M. Lanting, A. T. Lee, E. M. Leitch, J. Leong, W. Lu, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, K. K. Schaffer, D. Schwan, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, C. Tucker, K. Vand erlinde, J. D. Vieira and R. Williamson, The 10 Meter South Pole Telescope, *Publications of the ASP* **123**, p. 568 (May 2011).
24. M. W. Auger, T. Treu, A. S. Bolton, R. Gavazzi, L. V. E. Koopmans, P. J. Marshall, K. Bundy and L. A. Moustakas, The Sloan Lens ACS Survey. IX. Colors, Lensing, and Stellar Masses of Early-Type Galaxies, *The Astrophysical Journal* **705**, 1099 (November 2009).
25. M. Negrello, R. Hopwood, G. De Zotti, A. Cooray, A. Verma, J. Bock, D. T. Frayer, M. A. Gurwell, A. Omont, R. Neri, H. Dannerbauer, L. L. Leeuw, E. Barton, J. Cooke, S. Kim, E. da Cunha, G. Rodighiero, P. Cox, D. G. Bonfield, M. J. Jarvis, S. Serjeant, R. J. Ivison, S. Dye, I. Aretxaga, D. H. Hughes, E. Ibar, F. Bertoldi, I. Valtchanov, S. Eales, L. Dunne, S. P. Driver, R. Auld, S. Buttiglione, A. Cava, C. A. Grady, D. L. Clements, A. Dariush, J. Fritz, D. Hill, J. B. Hornbeck, L. Kelvin, G. Lagache, M. Lopez-Caniego, J. Gonzalez-Nuevo, S. Maddox, E. Pascale, M. Pohlen, E. E. Rigby, A. Robotham, C. Simpson, D. J. B. Smith, P. Temi, M. A. Thompson, B. E. Woodgate, D. G. York, J. E. Aguirre, A. Beelen, A. Blain, A. J. Baker, M. Birkinshaw, R. Blundell, C. M. Bradford, D. Burgarella, L. Danese, J. S. Dunlop, S. Fleuren, J. Glenn, A. I. Harris, J. Kamenetzky, R. E. Lupu, R. J. Maddalena, B. F. Madore, P. R. Maloney, H. Matsuhara, M. J. Michałowski, E. J. Murphy, B. J. Naylor, H. Nguyen, C. Popescu, S. Rawlings, D. Rigopoulou, D. Scott, K. S. Scott, M. Seibert, I. Smail, R. J. Tuffs, J. D. Vieira, P. P. van der Werf and J. Zmuidzinas, The Detection of a Population of Submillimeter-Bright, Strongly Lensed Galaxies, *Science* **330**, p. 800 (November 2010).
26. L. Dunne, L. Bonavera, J. Gonzalez-Nuevo, S. J. Maddox and C. Vlahakis, Overdensity of SMGs in fields containing z $\sim$ 0.3 galaxies: magnification bias and the implications for studies of galaxy evolution, *Monthly Notices of the Royal Astronomical Society* **498**, 4635 (November 2020).
27. L. Wang, A. Cooray, D. Farrah, A. Amblard, R. Auld, J. Bock, D. Brisbin, D. Burgarella, P. Chanial, D. L. Clements, S. Eales, A. Franceschini, J. Glenn, Y. Gong, M. Griffin, S. Heinis, E. Ibar, R. J. Ivison, A. M. J. Mortier, S. J. Oliver, M. J. Page, A. Papageorgiou, C. P. Pearson, I. Pérez-Fournon, M. Pohlen, J. I. Rawlings, G. Raymond, G. Rodighiero, I. G. Roseboom, M. Rowan-Robinson, D. Scott, P. Serra, N. Seymour, A. J. Smith, M. Symeonidis, K. E. Tugwell, M. Vaccari, J. D. Vieira,

L. Vigroux and G. Wright, HerMES: detection of cosmic magnification of submillimetre galaxies using angular cross-correlation, *Monthly Notices of the Royal Astronomical Society* **414**, 596 (June 2011).

28. J. González-Nuevo, A. Lapi, M. Negrello, L. Danese, G. De Zotti, S. Amber, M. Baes, J. Bland-Hawthorn, N. Bourne, S. Brough, R. S. Bussmann, Z. Y. Cai, A. Cooray, S. P. Driver, L. Dunne, S. Dye, S. Eales, E. Ibar, R. Ivison, J. Liske, J. Loveday, S. Maddox, M. J. Michałowski, A. S. G. Robotham, D. Scott, M. W. L. Smith, E. Valiante and J. Q. Xia, Herschel-ATLAS/GAMA: SDSS cross-correlation induced by weak lensing, *Monthly Notices of the Royal Astronomical Society* **442**, 2680 (August 2014).
29. D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, emcee: The MCMC Hammer, *Publications of the ASP* **125**, p. 306 (Mar 2013).

The impact of the Lorentz symmetry violation on the CMB polarization

Seddigheh Tizchang

School of Particles and Accelerators, Institute for Research in Fundamental Sciences (IPM) PO BOX19395-5531, Tehran, Iran
E-mail: s.tizchang@ipm.ir

Rohoollah Mohammadi

Iranian National Science and Technology Museum (INMOST), PO BOX: 11369-14611, Tehran, Iran
School of Astronomy, Institute for Research in FundamentalSciences (IPM), PO BOX 19395-5531,Tehran, Iran
E-mail: rmohammadi@ipm.ir

She-Sheng Xue

ICRANet and Department of Physics, Sapienza University of Rome P.le A. Moro 5, 00185 Rome, Italy
E-mail: xue@icra.it

In the standard cosmological scenario, no circular polarization is predicted for Cosmic Microwave Background (CMB) radiation. However, in the frame of moving particle, Lorentz symmetry violation leads to circular polarization for CMB radiation. We estimate the circular polarization power spectrum $C_{Vl}^{(S)}$ in CMB radiation due to Compton scattering in the presence of the Lorentz symmetry violation. We show that the V-mode power spectrum can be obtained in terms of linear polarization power spectrum at the last scattering surface.

Keywords: Circular polarization, CMB, power spectrum

1. Introduction

Numerous valuable information could be obtained from temperature anisotropies and polarization of the CMB. In particular, tempreture anisotropy in CMB radiation could generate polarization for the CMB. The theoretical calculation predicts at the most 10% of linear polarization for the CMB known as E-mode and B-mode which is originated from Compton scattering of photons on the electron plasma at the recombination epoch. The E-mode and B-mode polarization are different polarization pattern arisen from scalar and tensor temperature anisotropy, respectively.[1–3] Besides, the standard model of cosmology peredict no circular polarization for the CMB due to Compton scattering, and circular polarization has not been measured yet. However, most recently the CLASS experiment has established tightest upper bound on angular power spectrum of the CMB at 95% confidence from $0.4\mu K^2$ to $13.5\mu K^2$ on $\ell(\ell+1)C_\ell^{VV}/(2\pi))$ at multipole moments $1 \leq \ell \leq 120$.[4] The circular polarization for the CMB is a rare process so that observation of any level of

circular polarization could be a hint for the existence of new physics. In particular, Faraday conversion could be a source of the CMB circular polarization in which linear polarized light propagates through a medium and lead to various indexes of refraction along the two transverse axes. Besides, Compton scattering of CMB from the electron in the presence of a trivial background field such as a magnetic field[5,6] or non-trivial background like Non-Commutative space-time[7,8] and Lorentz symmetry violation[6,9] will rise circular polarization for the CMB. Moreover, scattering of CMB photon from non-standard particles like vector or fermionic dark matter can lead to circular polarization of the CMB[10,11] and cosmic photons.[12]

In this study, we would like to explore the effects of Lorentz symmetry violation (LIV) as a non-trivial background on the angular power spectrum of the CMB. The energy scale for LIV is the Planck scale which is out of reach based on direct experiments. An alternative way to study such effects is the Standard Model Extension (SME) based on effective field theory approach.[13] The Lorentz violation at the SME is induced by spontaneous symmetry breaking and result in new terms including LIV parameters and can behave such as constant backgrounds. The phenomenological aspects of LIV have been widely studied in the literature which total obtained constraints are gathered at Ref.[14].

In this paper, we give a brief overview of the SME in section 2. Then the effect of LIV on the CMB polarization is given in section 3. Finally, in section 4 we give a conclusion and summary.

2. The SME

We assume the CMB photon scatter from cosmic electron while the Lorentz invariance is violated. The LIV effect is appliable by the SME. The minimal version of the SME contains all possible gauge invariant operators, made by all the SM fields and LIV is specified with new couplings.[13] The QED part of SME including charged fermion interaction with photon is defined as follows:

$$\mathcal{L}^{\rm LIV}_{\rm QED} = \frac{1}{2} i\bar{\psi}(\gamma^{\nu} + c^{\mu\nu}\gamma_{\mu}) \overleftrightarrow{D}_{\mu} \psi - \bar{\psi} m_e \psi, \tag{1}$$

where ψ is the fermion field, $\mathcal{D}_{\mu}$ shows the covariant derivative with $u \overleftrightarrow{D}_{\mu} v \equiv u D_{\mu} v - v D_{\mu} u$,[15] the mass of electron is m_e, and γ_{μ} indicates the Dirac gamma matrices and the coefficient for Lorentz symmetry violation is $c^{\mu\nu}$ in fermion sector. We assume $c^{\mu\nu}$ to be symmetric and traceless tensor. For convenience,the constraint on the SME parameters are expressed in a standard reference frame, *i.e* Sun-centered, celestial equatorial frame (SCCEF) with coordinates $\{T, X, Y, Z\}$, the coordinates of lab frame are denoted by $\{t, x, y, z\}$. Conversion to a lab frame at convenient time t is defined by[16]

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} \cos\chi\cos\Omega t & \cos\chi\sin\Omega t & -\sin\chi \\ -\sin\Omega t & \cos\Omega t & 0 \\ \sin\chi\cos\Omega t & \sin\chi\sin\Omega t & \cos\chi \end{pmatrix} \begin{pmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{pmatrix}. \tag{2}$$

If we consider the lab frame on the earth Ω would be the earth sidereal rotation frequency equivalent to $2\pi/(23\text{h }56\text{ min})$ and $0 \leq \chi \leq \pi$. Some experimental bounds on the components of c parameter in the SCCEF frame is given in Tab 1. Therefore,

$c_{\mu\nu}$	Experimental bounds	system
c_{TT}	2×10^{-15}	Collider physics[17]
c_{YY}	3×10^{-15}	Astrophysics[18]
c_{ZZ}	5×10^{-15}	Astrophysics[18]
$c_{(XY)}$	3×10^{-15}	Astrophysics[18]
$c_{(YZ)}$	1.8×10^{-15}	Astrophysics[18]
$c_{(XZ)}$	3×10^{-15}	Astrophysics[18]
$c_{(TX)}$	-30×10^{-14}	Collider physics[17]
$c_{(TY)}$	-80×10^{-15}	Collider physics[17]
$c_{(TZ)}$	-11×10^{-13}	Collider physics[17]

we also use correlation between components of laboratory frame and SCCEF given in Ref[16] to transfer the result to SCCEF frame.

3. Effect of LIV on the CMB polarization

The theoretical aspect of polarization of the CMB photon is characterized by four Stokes parameters: $[I, Q, U, V]$[1] which can be related to component of density matrix as:

$$\rho = \frac{1}{2}\begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix}, \tag{3}$$

where I indicates the total radiation intensity (temperature), Q and U indicates linear polarization and V is circular polarization of the CMB radiation which satisfy the inequality $I^2 \geq Q^2 + U^2 + V^2$. Compton scattering cannot generate circular polarization V, so in the standard model of cosmology circular polarization is not anticipated for the CMB. For an ensemble of photon which are assumed free before as well as after scattering, the time evolution for Stokes parameters will be obtained by solving the following quantum Boltzmann equation, given as[1]

$$\begin{aligned}(2\pi)^3\delta^3(0)2\,p^0\frac{d}{dt}\rho_{ij}(0,\mathbf{p}) &= i\langle[H_I^0(0),\hat{\mathcal{D}}_{ij}(\mathbf{p})]\rangle \\ &\quad -\frac{1}{2}\int_{-\infty}^{\infty}dt\langle[H_I^0(t),[H_I(0),\hat{\mathcal{D}}_{ij}(\mathbf{p})]]\rangle,\end{aligned} \tag{4}$$

where p^0 represents the energy of photon and the photon number operator is given with $\hat{\mathcal{D}}_{ij}(\mathbf{p}) = \hat{a}_i^\dagger(\mathbf{p})\hat{a}_j(\mathbf{p})$ where $\hat{a}_i^\dagger(\mathbf{p})$ and $\hat{a}_j(\mathbf{p})$ are rising and lowering operators, respectively.

In the Eq. (4), H_I^0 is an interacting Hamiltonian that includes all possible interactions at leading order. First term on the right-hand side of the above equation is known as 'forward scattering' where the incoming particles interact without exchange the momenta and this term is linearly proportional to H_I^0. The second term is called 'Damping term' and quadratically proportional to H_I^0. In the case of Compton scattering, the contribution of first term of Eq. (4) is equal to zero. Moreover, time evolution of the circular polarization is zero ($\dot{V} = 0$) for both terms of right-hand side of Eq. (4). Therefore, contribution of any new physics in polarization of photon in forward scattering would be considerable.

We obtain the Boltzmann equation for circular polarization of CMB radiation in the presence of LIV by using the Eq. (4), and reconstructing the Stokes parameters[6] as follows:

$$\begin{aligned}
&\frac{d}{d\eta}\Delta_I^{(S)} + iK\mu\Delta_I^{(S)} + 4[\dot{\psi} - iK\mu\varphi] = C_{e\gamma}^I,\\
&\frac{d}{d\eta}\Delta_P^{\pm(S)} + iK\mu\Delta_P^{\pm(S)} = C_{e\gamma}^{\pm} - i\kappa_{\rm LIV}^{\pm}\,\Delta_V^{(S)},\\
&\frac{d}{d\eta}\Delta_V^{(S)} + iK\mu\Delta_V^{(S)} = C_{e\gamma}^V + \frac{i}{2}\Big[\kappa_{\rm LIV}^{+}\,\Delta_P^{-(S)} + \kappa_{\rm LIV}^{-}\,\Delta_P^{+(S)}\Big],
\end{aligned} \tag{5}$$

where η represents conformal time. Scalar perturbation corresponding to the potential of Newton and spacial curvature perturbation are shown with ψ and ϕ. $\Delta_P^{\pm(S)} = Q^{(S)} \pm iU^{(S)}$ is a linear combination of $Q^{(S)}$ and $U^{(S)}$ polarization. The scalar product of wave vector $\mathbf{K}$ and the CMB propagating direction is given by μ. $C_{e\gamma}^I$, $C_{e\gamma}^{\pm}$ and $C_{e\gamma}^V$ represents the contribution of the SM Compton scattering in $\Delta_P^{\pm(S)}$ polarization and $\Delta_V^{(S)}$ polarization,[1] respectively. $\kappa_{\rm LIV}^{\pm}$ indicates the involvement of Compton scattering by considering fermionic part of the SME Lagrangian at Boltzmann equations (Eq. (5)) which are obtained as follows:

$$\kappa_{\rm LIV} = -a(\eta)\,\frac{3}{4}\,\frac{\sigma_T}{\alpha}\,\frac{m_e^2}{p_0}\,\frac{\bar{n}_e}{m_e},\;\; \kappa_{\rm LIV}^{\pm} = \kappa_{\rm LIV}\,c_{\{\alpha\beta\}}(\rho_Q^{\{\alpha\beta\}} \pm i\rho_U^{\{\alpha\beta\}}), \tag{6}$$

where $\rho_Q^{\{\alpha\beta\}}$ and $\rho_U^{\{\alpha\beta\}}$ are defined as:

$$\begin{aligned}
\rho_Q^{\{\alpha\beta\}} &= -\Big[\epsilon^1\cdot\hat{c}\cdot\epsilon^2 + \sum_{f=e,p}\frac{m_f v_b^2}{2k_0}(\hat{v}\cdot\epsilon^2\hat{v}\cdot\hat{c}\cdot\epsilon^1 + \hat{v}\cdot\epsilon^1\hat{v}\cdot\hat{c}\cdot\epsilon^2) - \frac{v_b}{2}(\hat{v}\cdot\epsilon^2\hat{k}\cdot\hat{c}\cdot\epsilon^1\\
&\quad + \hat{v}\cdot\epsilon^1\hat{k}\cdot\hat{c}\cdot\epsilon^2) + \sum_{f=e,p}\frac{2v_b^4 m_f^2}{k_0^2}(\hat{v}\cdot\hat{c}\cdot\hat{v}\,\hat{v}\cdot\epsilon^2\hat{v}\cdot\epsilon^1) + 2v_b^2\,\hat{k}\cdot\hat{c}\cdot\hat{k}\,\hat{v}\cdot\epsilon^1\,\hat{v}\cdot\epsilon^2\Big],\\
\rho_U^{\{\alpha\beta\}} &= \Big[\frac{1}{2}(\epsilon^1\cdot\hat{c}\cdot\epsilon^1 + \epsilon^2\cdot\hat{c}\cdot\epsilon^2) + \sum_{f=e,p}\frac{m_f v_b^2}{2k_0}(\hat{v}\cdot\epsilon^1\hat{v}\cdot\hat{c}\cdot\epsilon^1 - \hat{v}\cdot\epsilon^2\hat{v}\cdot\hat{c}\cdot\epsilon^2)\\
&\quad + \frac{v_b}{2}(\hat{v}\cdot\epsilon^2\hat{k}\cdot\hat{c}\cdot\epsilon^2 - \hat{v}\cdot\epsilon^1\hat{k}\cdot\hat{c}\cdot\epsilon^1) + \sum_{f=e,p}\frac{v_b^4 m_f^2}{2k_0^2}\hat{v}\cdot\hat{c}\cdot\hat{v}(\hat{v}\cdot\epsilon^1\hat{v}\cdot\epsilon^1\\
&\quad + \hat{v}\cdot\epsilon^2\hat{v}\cdot\epsilon^2) + \frac{v_b^2}{2}\,\hat{k}\cdot\hat{c}\cdot\hat{k}\,(\hat{v}\cdot\epsilon^1\,\hat{v}\cdot\epsilon^1 + \hat{v}\cdot\epsilon^2\,\hat{v}\cdot\epsilon^2)\Big],
\end{aligned} \tag{7}$$

here $\alpha, \beta = 0, i, j$ represent time-like and space-like LIV components of coefficient c. The direction of bulk velocity of electrons is shown by $\hat{v}$. It is convenient, to describe the statistical properties of CMB radiation via spherical harmonic. Then the Stokes parameter could be expanded over suitable multipole moment in spin weighted basis[2,3]

$$\Delta^{(S)}_{I,P,V}(\eta, \mathbf{K}, \mu) = \sum_{l=0}^{\infty}(2l+1)(-i)^l \Delta^l_{I,P,V}(\eta, \mathbf{K}) P_l(\mu), \tag{8}$$

with $P_l(\mu)$ being the Legendre polynomial of rank l. In particular, the value of $\Delta^{(S)}_{I,P,V}$ at the present time η_0 in direction $\hat{\mathbf{n}}$ are achieved by integrating the Boltzmann equations (Eq. (5)) along the line of sight and summation over all Fourier modes $\mathbf{K}$ as follows:

$$\begin{aligned}\Delta^{\pm(S)}_P(\hat{\mathbf{n}}) &= \int d^3\mathbf{K}\xi(\mathbf{K}) e^{\pm 2i\phi_{K,n}} \Delta^{\pm(S)}_P(\mathbf{K}, \mathbf{p}, \eta),\\ \Delta^{(S)}_V(\hat{\mathbf{n}}) &= \int d^3\mathbf{K}\xi(\mathbf{K}) \Delta^{(S)}_V(\mathbf{K}, \mathbf{p}, \eta),\end{aligned} \tag{9}$$

where $\phi_{K,n}$ are characterized to rotate the $\hat{\mathbf{n}}$ and $\hat{\mathbf{K}}$ dependent basis to a fixed frame in the sky. The random variable $\xi(\mathbf{K})$ is used to define the initial amplitude of the mode. Then $\Delta^{(S)}_V$ and $\Delta^{\pm(S)}_P$ are obtained as:

$$\begin{aligned}\Delta^{(S)}_V(\mathbf{K}, \mu, \eta_0) &= \frac{1}{2}\int_0^{\eta_0} d\eta\, \dot{\eta}_{e\gamma}\, e^{ix\mu - \eta_{e\gamma}} \left[3\mu\Delta^{(S)}_{V1} + i(\frac{\kappa^{+}_{LV}}{\dot{\eta}_{e\gamma}}\Delta^{+(S)}_P + \frac{\kappa^{-}_{LV}}{\dot{\eta}_{e\gamma}}\Delta^{-(S)}_P)\right],\\ &\approx \frac{1}{2}\int_0^{\eta_0} d\eta\, \dot{\eta}_{e\gamma}\, e^{ix\mu - \eta_{e\gamma}} \left[3\mu\Delta^{(S)}_{V1} + 2i\,\rho^{\{\alpha\beta\}}_Q c_{\{\alpha\beta\}}\frac{\kappa_{LV}}{\dot{\eta}_{e\gamma}}\Delta^{(S)}_P + ...\right],\end{aligned} \tag{10}$$

$$\begin{aligned}\Delta^{\pm(S)}_P(\mathbf{K}, \mu, \eta_0) = \int_0^{\eta_0} d\eta\, \dot{\tau}_{e\gamma}\, e^{ix\mu - \tau_{e\gamma}} \Big[&\frac{3}{4}(1-\mu^2)\Pi(K,\eta)\\ &- i\frac{\kappa^{\mp}_{LV}}{\dot{\eta}_{e\gamma}}\Delta^{(S)}_V\Big],\end{aligned} \tag{11}$$

where $\Pi = \Delta^{(S)}_{T2} + \Delta^{(S)}_{P2} + \Delta^{(S)}_{P0}$, the differential and total optical depth caused by Thomson scattering (σ_T) at the time of η are given by $\dot{\eta}_{e\gamma} = a(\eta)\, n_e\, \sigma_T$ and $\eta_{e\gamma} = \int_{\eta}^{\eta_0} \dot{\eta}_{e\gamma}\, d\eta$ with n_e being electron number density, and $x = \mathbf{K}\,(\eta_0 - \eta)$. The $\Delta^{(S)}_P$ is linear polarization defined as

$$\Delta^{(S)}_P(\mathbf{K}, \mu, \eta) = \int_0^{\eta} d\eta\, \dot{\eta}_{e\gamma}\, e^{ix\mu - \eta_{e\gamma}} \left[\frac{3}{4}(1-\mu^2)\Pi(K,\eta)\right]. \tag{12}$$

The $\Delta^{(S)}_V$ arised from Compton scattering in presence of LIV (Eq. (10)) can be estimated in terms of linear polarization. The term $\frac{\kappa_{LV}}{\dot{\eta}_{e\gamma}}\rho^{\{\alpha\beta\}}_Q c_{\{\alpha\beta\}}$ represents the strength of LIV contribution at circular polarization. Therefore, we consider the dominant term in $\rho^{\{\alpha\beta\}}_Q$ (Eq. (7)).

The correlation function of circular polarization in the presence of LIV in terms of primordial power spectrum is obtained as:

$$\begin{aligned} C_l^V &= \frac{1}{2l+1}\sum_m \left\langle a^*_{V,lm}\, a_{V,lm}\right\rangle \\ &\approx \frac{1}{2l+1}\int d^3K P_v(\mathbf{K}) \times \\ &\sum_m \Big| \int d\Omega Y^*_{lm} \int_0^{\eta_0} d\eta\, \dot\eta_{e\gamma}\, e^{ix\mu-\eta_{e\gamma}}\, 2i\, \rho_Q^{\{\alpha\beta\}} c_{\{\alpha\beta\}}\, \frac{\kappa_{\rm LIV}}{\dot\eta_{e\gamma}}\, \Delta_P^{(S)}|^2, \end{aligned} \tag{13}$$

and $P_\nu(K)$ is the velocity power spectrum which can be expressed as a function of primordial scalar spectrum:

$$P_\nu(\mathbf{K},\tau) \sim P_\phi^{(S)}(\mathbf{K},\tau). \tag{14}$$

Finally, we estimate the circular power spectrum C_l^V regarding to linearly polarized two point function C_l^P as follows:

$$C_l^V \approx (\tilde\kappa_{\rm LIV}^{\rm avg})^2 C_l^P, \tag{15}$$

where

$$\tilde\kappa_{\rm LIV} = c_{\{\alpha\beta\}}\, \rho_Q^{\{\alpha\beta\}} \frac{\kappa_{\rm LIV}}{\dot\tau_{e\gamma}}, \qquad \tilde\kappa_{\rm LIV}^{\rm avg} = \frac{1}{z_{lss}}\int_0^{z_{lss}} dz\ \tilde\kappa_{\rm LIV}(z) \simeq 10^{-3}. \tag{16}$$

here $\tilde\kappa_{\rm LIV}^{\rm avg}$ is the mean of $\tilde\kappa_{\rm LIV}(z)$ over redshift from last scattering surface $z_{lss} \simeq 1100$ up to today $z = 1$. Similar to Ref,[19] we assume $c^{\alpha\beta}$ tensor with symmetric component $\alpha = \beta = T, X, Y, Z$ at the order of 10^{-15} and $\chi = \pi/4$.[14] Current experimental value for linearly polarized power spectrum for the CMB is at the order of $C_l^P \equiv 0.1\mu K^2$ for $l < 250$.[20] Applying the value of linearly polarized power spectrum, we estimate the V-mode power spectrum generated from Compton scattering in the presence of LIV as follows

$$C_l^V \approx 0.8\ (nK)^2 \big(\frac{c^{\alpha\beta}}{10^{-15}}\big)^2. \tag{17}$$

As we mentioned before, the tightest experimental bound on the circular polarization is obtained by CLASS experiment which is $0.4\mu K^2$ to $13.5\mu K^2$ for $1 \le l \le 120$. By improving the sensitivity of experiments in the future, we expect the obtained constraint in Eq. (17) would be comparable to experimental bounds.

4. Summary

In this work we studied the circular polarization of CMB due to forward scattering with cosmic electron in the presence of LIV and scalar perturbations. We derived the Boltzmann equations describing the time evolution of CMB polarization. We also showed that the circular power spectrum of CMB in the presence of LIV can be expressed as linear power spectrum. Circular polarization of CMB might be originated from different sources so that the possibility to separate the contribution of

different sources would be valuable. The dominant term of circular polarization for the CMB due to LIV is linearly proportional to wavelength $\lambda = (1/k_0)$ (Eq. (7)) while the wavelength dependence for other effects could be different. This various dependency to wavelength leads to a different distribution of V-mode power spectrum arising from different models and can help us to compare our result with other bounds on circular polarization.

References

1. A. Kosowsky, Cosmic microwave background polarization, *Annals Phys.* **246**, 49 (1996).
2. U. Seljak and M. Zaldarriaga, A Line of sight integration approach to cosmic microwave background anisotropies, *Astrophys. J.* **469**, 437 (1996).
3. M. Zaldarriaga and U. Seljak, An all sky analysis of polarization in the microwave background, *Phys. Rev. D* **55**, 1830 (1997).
4. I. L. Padilla, J. R. Eimer, Y. Li, G. E. Addison, A. Ali, J. W. Appel, C. L. Bennett, R. Bustos, M. K. Brewer, M. Chan, D. T. Chuss, J. Cleary, J. Couto, S. Dahal, K. Denis, R. Dünner, T. Essinger-Hileman, P. Fluxá, D. Gothe, S. K. Haridas, K. Harrington, J. Iuliano, J. Karakla, T. A. Marriage, N. J. Miller, C. Núñez, L. Parker, M. A. Petroff, R. Reeves, K. Rostem, R. W. Stevens, D. A. N. Valle, D. J. Watts, J. L. Weiland, E. J. Wollack and Z. Xu, Two-year cosmology large angular scale surveyor (CLASS) observations: A measurement of circular polarization at 40 GHz, *The Astrophysical Journal* **889**, p. 105 (jan 2020).
5. A. Cooray, A. Melchiorri and J. Silk, Is the cosmic microwave background circularly polarized?, *Phys. Lett. B* **554**, 1 (2003).
6. E. Bavarsad, M. Haghighat, Z. Rezaei, R. Mohammadi, I. Motie and M. Zarei, Generation of circular polarization of the CMB, *Phys. Rev. D* **81**, p. 084035 (2010).
7. S. Tizchang, S. Batebi, M. Haghighat and R. Mohammadi, The generation of B-mode and circular polarization of cosmic photons due to Non-Commutative space-Time background, *PoS* **ICHEP2018**, p. 142 (2019).
8. S. Tizchang, S. Batebi, M. Haghighat and R. Mohammadi, Cosmic microwave background polarization in non-commutative space-time, *Eur. Phys. J. C* **76**, p. 478 (2016).
9. S. Alexander, E. McDonough, A. Pullen and B. Shapiro, Physics Beyond The Standard Model with Circular Polarization in the CMB and CMB-21cm Cross-Correlation, *JCAP* **01**, p. 032 (2020).
10. S. Modares Vamegh, M. Haghighat, S. Mahmoudi and R. Mohammadi, Impact of the vector dark matter on polarization of the CMB photon, *Phys. Rev. D* **100**, p. 103024 (2019).
11. M. Haghighat, S. Mahmoudi, R. Mohammadi, S. Tizchang and S. S. Xue, Circular polarization of cosmic photons due to their interactions with Sterile neutrino dark matter, *Phys. Rev. D* **101**, p. 123016 (2020).
12. M. Cermeño, C. Degrande and L. Mantani, Circular polarisation of gamma rays as a probe of dark matter interactions with cosmic ray electrons (3 2021).
13. D. Colladay and V. A. Kostelecky, CPT violation and the standard model, *Phys. Rev. D* **55**, 6760 (1997).
14. V. A. Kostelecky and N. Russell, Data Tables for Lorentz and CPT Violation (1 2008).
15. D. Colladay and V. A. Kostelecky, Lorentz violating extension of the standard model, *Phys. Rev. D* **58**, p. 116002 (1998).

16. V. A. Kostelecky and C. D. Lane, Constraints on Lorentz violation from clock comparison experiments, *Phys. Rev. D* **60**, p. 116010 (1999).
17. B. Altschul, Laboratory Bounds on Electron Lorentz Violation, *Phys. Rev. D* **82**, p. 016002 (2010).
18. B. Altschul, Synchrotron and inverse compton constraints on Lorentz violations for electrons, *Phys. Rev. D* **74**, p. 083003 (2006).
19. S. Tizchang, R. Mohammadi and S.-S. Xue, Probing Lorentz violation effects via a laser beam interacting with a high-energy charged lepton beam, *Eur. Phys. J. C* **79**, p. 224 (2019).
20. P. A. R. Ade *et al.*, Planck 2015 results. XX. Constraints on inflation, *Astron. Astrophys.* **594**, p. A20 (2016).

Cosmic backgrounds from the radio to the far-infrared: Recent results and perspectives from cosmological and astrophysical surveys*

Carlo Burigana;[1,2,3,a] Elia Sefano Battistelli;[4,5,b] Laura Bonavera;[6,7,c]
Tirthankar Roy Choudhury;[8,d] Marcos Lopez-Caniego;[9,e] Constantinos Skordis;[10,f]
Raelyn Marguerite Sullivan;[11,g] Hideki Tanimura;[12,h] Seddigheh Tizchang;[13,i]
Matthieu Tristram;[14,j] Amanda Weltman[15,k]

[1] *INAF–IRA, Via Piero Gobetti 101, 40129 Bologna, Italy*†
[2] *Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy*
[3] *INFN, Sezione di Bologna, Via Irnerio 46, 40126, Bologna, Italy*
[4] *Dipartimento di Fisica, Università di Roma "La Sapienza", P.le Aldo Moro 2, 00185, Rome, Italy*
[5] *INAF-IAPS Roma, Via del Fosso del Cavaliere 100, 00133 Roma, Italy*
[6] *Departamento de Física, Universidad de Oviedo, C. Federico García Lorca 18, 33007 Oviedo, Spain*
[7] *Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), C. Independencia 13, 33004 Oviedo, Spain*
[8] *National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Ganeshkhind, Pune 411007, India*
[9] *Aurora Technology B.V. for the European Space Agency, Villanueva de la Canada, Madrid, 28692, Spain*
[10] *CEICO, FZU - Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 21, Prague*
[11] *Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia, Canada*
[12] *Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, Bâtiment 121, 91405 Orsay, France*
[13] *School of Particles and Accelerators, Institute for Research in Fundamental Sciences (IPM), PO BOX19395-5531, Tehran, Iran*
[14] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France*
[15] *High Energy Physics, Cosmology & Astrophysics Theory Group, University of Cape Town, Private Bag, Cape Town, South Africa, 7700*

[a] *burigana@ira.inaf.it* — [b] *elia.battistelli@roma1.infn.it* — [c] *bonaveralaura@uniovi.es*
[d] *tirth@ncra.tifr.res.in* — [e] *marcos.lopez.caniego@sciops.esa.int* — [f] *skordis@fzu.cz*
[g] *rsullivan@phas.ubc.ca* — [h] *hideki.tanimura@ias.u-psud.fr* — [i] *s.tizchang@ipm.ir*
[j] *tristram@ijclab.in2p3.fr* — [k] *amanda.weltman@uct.ac.za*

Cosmological and astrophysical surveys in various wavebands, in particular from the radio to the far-infrared, offer a unique view of the universe's properties and the formation and evolution of its structures. After a preamble on the so-called tension problem, which occurs when different types of data are used to determine cosmological parameters, we discuss the role of fast radio bursts in cosmology, in particular for the missing baryon

*Based on talks presented at the Sixteenth Marcel Grossmann Meeting on General Relativity, July 2021.

†Istituto Nazionale di Astrofisica – Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy

problem, and the perspectives from the analysis of the 21 cm redshifted line from neutral hydrogen. We then describe the *Planck* Legacy Archive, its wealth of scientific information and next developments, and the promising perspectives expected from higher resolution observations, in particular for the analysis of the thermal Sunyaev-Zel'dovich effect. Three cosmological results of the *Planck* mission are presented next: the implications of the map of Comptonization fluctuations, the dipole analysis from cross-correlating cosmic microwave background anisotropy and Comptonization fluctuation maps, and the constraints on the primordial tensor-to-scalar perturbation ratio. Finally, we discuss some future perspectives and alternative scenarios in cosmology, such as the study of the Lorentz invariance violation with the cosmic microwave background polarization, the introduction of new gravitational degrees of freedom to solve the dark matter problem, and the exploitation of the magnification bias with high-redshift sub-millimeter galaxies to constrain cosmological parameters.

Keywords: Cosmology; background radiations; observational cosmology; radio, microwave; submillimeter; infrared emission; large-scale structure of the universe; galaxy clusters; radio sources; IR sources; gravitational lenses and luminous arcs; relativity and gravitation; modified theories of gravity; dark matter; special relativity; radiation mechanisms, polarization.

1. Introduction

Over the last century or so, cosmology has progressed from a field that was essentially philosophy to a precision science. The current standard cosmological model is the Λ cold dark matter (ΛCDM) model, which mainly consists of a cosmological constant (Λ) or dark energy (DE) component and cold dark matter (CDM) component, with general relativity (GR) as the assumed theory of gravity. It makes predictions on the smallest and largest scales, that, despite some open questions, are largely verified by a host of observational cosmology experiments.

Cosmological and astrophysical surveys in various wavebands, in particular from the radio to the far-infrared, are crucial to solve many problems in cosmology and to understand the origin and evolution of the cosmic structures at various scales through cosmic time. The accurate observations of the properties of the cosmic microwave background (CMB) began the era of precision cosmology, furthering our understanding of the early universe and its evolution up to the present time and beyond. Similarly, the information contained in the radio sky is of increasing relevance to unravel the complexity of cosmic evolution and to answer a wide range of open questions, thanks to the spectacular experimental improvement achieved in the past years and expected in the future.

The structure of this paper is as follows. In Sect. 2, we begin with a preamble to the so-called tension problem, which occurs when different data are used to determine cosmological parameters. We then discuss the role of fast radio bursts (FRBs) in cosmology to solve the missing baryons problem, followed by the perspectives represented by the analysis of the 21 cm redshifted line from neutral hydrogen (HI). Sect. 3 describes the wealth of scientific information and tools publicly available at the *Planck* Legacy Archive (PLA), with its current and future developments, and the perspectives expected from higher resolution observations at millimeter (mm)

and sub-mm wavelengths, in particular for the analysis of the thermal Sunyaev-Zel'dovich (SZ) effect towards galaxy clusters (GCs), focusing on a forthcoming well-defined facility instrument. In Sect. 4 we discuss three important cosmological results based on the data from the *Planck* mission: the implications of the analysis of the thermal SZ (tSZ) map, the dipole analysis based on the cross-correlation of CMB anisotropy and tSZ maps, and the constraints on the ratio between the primordial tensor and scalar perturbations. Finally, Sect. 5 is dedicated to future perspectives and alternative scenarios in cosmology. We consider three specific examples: the study of Lorentz invariance violation with the analysis of CMB polarization, the solution to the dark matter (DM) problem introducing new gravitational degrees of freedom, and the exploitation of the magnification bias with high-z sub-mm galaxies to constrain a set of cosmological parameters.

2. Cosmology in the radio

2.1. *Cosmology with radio astrophysics and tensions in cosmology*

Amongst the successes of modern cosmology, is the remarkable precision in the estimation of the cosmological parameters that describe the contents, geometry and history of our universe. Herein also lies the current greatest tension in the data. Indeed, in spite of the increasing precision in observational cosmology, it is possible we have lost some accuracy: a number of tensions now appear within our data, tensions that more cosmology experiments do not appear to solve. These tensions may require a change in the underlying theory or they may be resolved with novel observational probes of the large scale picture of the universe. Here we will describe an example of the latter, namely, the FRBs, the novel discovery of our century, as potential instruments for cosmology. Indeed we will highlight more broadly the use of radio astrophysics - encompassing both the use of radio transients and intensity mapping tools, and the potential they hold for groundbreaking cosmology discovery.

The Hubble constant, H_0, is anything but constant across observations at different scales (see Ref. 1 for a recent review of most of the data). A brief summary however, is that historically there were two independent measures for H_0, very far apart in value with measurements of roughly $50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ or $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ from the early and late universe respectively. However, the error bars on each measurement was large enough, that each was within errors of the other and it was assumed that with improved observations, the error bars would get smaller and the values would get closer and converge. Indeed, this appeared to be the case early this century, but every new data set over the last several years has the values diverging with ever smaller error bars, that critically no longer allow for a simple convergence in the value of H_0. Measurements from the early universe, such as *Planck*, Dark Energy Survey[2,3] (DES), and baryonic acoustic oscillations (BAO) and Big Bang nucleosynthesis (BBN) observations, all point to a smaller value of around $H_0 \simeq 67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and measurements from the late universe, using Cepheids as distance ladders, or lensing for example, find $H_0 \simeq 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with each

measurement having error bars in the range $\simeq 1 \div 4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Thus this tension is growing with data rather than resolving.

There are longer standing issues with the standard cosmological model, of course, the lack of direct observational evidence for DM which makes up roughly 25% of the total energy budget of the universe is of particular concern. No less resolved is the issue of DE, here making up $\sim 70\%$ of the energy budget of the universe and yet only observed in indirect cosmological experiments with no direct detection and also no compelling theoretical explanation for why it even exists or why it appears to be dominating the total energy budget of the universe at a cosmological time when we are both possible and here to observe it. Early universe inflation is a key component in the model, solving several open problems such as the flatness and horizon problems and yet we still have no theory of inflation that fits within an ultraviolet complete theory with a natural candidate for the inflaton lacking. One hopes that these three problems are resolved together with a leap in the theory space - yet the obligation remains to continue to search for observational data that may yet play a role.

Other problems have lingered for very long, such as the missing baryon problem, where even the $\sim 5\%$ of the universe that we know exists is not entirely accounted for. Indeed adding up all of the observed contributions to the baryons accounts for only $\sim 70\%$ of the total expected. Roughly 30% are considered missing, and expected to be somewhere in the warm hot intergalactic medium (WHIM). Indeed, we should say were considered missing, as we discuss below how to find these missing baryons using the FRBs.

2.1.1. *Fast radio bursts and their role in cosmology*

As the name suggests, FRBs are very bright ($\sim$ Jy), brief ($\sim\mu$s to $\sim$ ms scale) transients, observed in the broad spectrum of the radio. The first was discovered[4] in 2007 and ever since the search has been on to discover more and their properties, first through searches in archival data and more recently through purpose built radio telescopes and arrays to find thousands if not tens of thousands of bursts every year. Our understanding of the properties of FRBs is still evolving; it appears that many but not all repeat, it is not yet known if any of the repeaters repeat in a periodic fashion, and indeed it is likely that ultimately FRBs will fall in several classes as has happened with transients historically. Their other properties such as polarisation, rotation measures etc are all still unknown as the observations do not all point to a single pattern and it is not always possible to disentangle host galaxy effects from propagation effects from intrinsic properties of the bursts. It does appear certain that they are found in host galaxies, and that they lie at cosmological distances,[5] thus making them excellent candidates for cosmology as their propagation will probe the intergalactic medium (IGM) and thus give us a number of novel ways to constrain cosmological parameters (see, e.g., Refs. 6–8 and references therein). From a theoretical standpoint, the progenitor mechanism

driving FRBs is not yet known though there are strong hints that at least some are driven by magnetars through one of many physical possible mechanisms. A full database of theories,[9] once outnumbered the observations, but as the amount of data grows, the possibilities for theoretical explanations shrinks and so it is likely that in a few years we will have only a handful of possible contenders remaining, perhaps matching future classes of FRBs.

There are a number of ways FRBs can be used as cosmological probes. Here we focus on a single application. After emission from the source, the photons from the burst travel to the observer through the IGM, and are slowed down differentially as a function of their wavelength, λ, or frequency, $\nu = c/\lambda$, c being the speed of light. The dispersion measure, $M_D \simeq \int n_e dl$, where n_e is the electron number density and the integral along the line of sight runs from the source to the observer, thus contains cosmological information about the distribution of electrons. Indeed, this simple equation combined with precision cosmology constraints from the early universe allows us to make a prediction for the fraction of baryons in the IGM,[7] which can then be experimentally verified. Indeed, the so-called missing baryon problem is not longer an outstanding problem as the results of Ref. 10 used a handful of well located FRBs with the observed dispersion measures to show that indeed the missing baryons are in the IGM as expected, and are playing the predicted role of dispersing the FRB signal. This longstanding open problem is thus resolved not due to any great technological or theoretical breakthrough, but simply through the use of a few transient observations and an understanding of the contributions to M_D from our own galaxy. This suggests that there is immense and untapped potential for great discovery with FRBs and that the future is very bright for this young field.

2.2. *High-redshift universe with redshifted 21 cm line*

The redshifted 21 cm line of HI is one of the most useful probes of the early universe. Several experiments are ongoing and are being planned to detect the signal at high redshift, z. Detection of the signal will help in understanding the first stars in the universe, the formation and evolution of galaxies and also constraining cosmological parameters. This section summarizes the status of observational efforts and theoretical understanding.

The HI evolution can be used to study the thermal and ionization history of the universe and also the evolution of the galaxies.[11,12] The very first HI atoms formed during the so-called 'recombination' epoch. The hydrogen remains in the neutral form until the first stars form which can then start ionizing the HI in the IGM, a process known as *reionization*. This epoch is believed to end around redshifts $z \sim 5.5$ by when almost all the HI in the IGM gets ionized. In the *post-reionization* era, the only regions where HI can survive are the high-density regions, e.g., the interstellar medium (ISM) of the galaxies.

The evolution of HI can be efficiently tracked by the redshifted 21 cm signal arising from hyperfine transition of the ground state of the atom. The signal is

observed, either in emission or absorption, in contrast to the CMB signal using low-frequency radio telescopes. Since the hyperfine transition is forbidden, the signal amplitude turns out to be extremely weak and hence is appreciable only when the HI density is significant. There are broadly three regimes where the detection of the 21 cm radiation is being planned: (i) the *cosmic dawn* when the ultraviolet (UV) and X-ray radiation from the very early galaxies lead to a strong absorption signal around $z \sim 20$; (ii) the *epoch of reionization* when the heated and ionized IGM leads to an emission signal around $z \sim 15 - 6$; and (iii) the *post-reionization* signal when HI in galaxies can be detected either directly or through intensity mapping.

Among the different ways to detect the 21 cm signal, perhaps the one that requires simplest of the instruments is the signal averaged over large regions of the sky, i.e., the *global signal*. This signal is believed to be most prominent during the cosmic dawn corresponding to $\nu \simeq 50 - 100$ MHz. The amplitude of the cosmological signal is weak and is buried under other astrophysical signals few orders of magnitude stronger, which makes it extremely challenging to separate it out.

Recently, there has been a claim of a detection of the 21 cm signal from cosmic dawn at $z \sim 18$ by the Experiment to Detect the Global EoR Signature[26] (EDGES), although the recovered signal seems to be 3–4 times larger than that predicted by standard galaxy formation models. This lead to several theoretical interpretation beyond the standard calculations, e.g., exotic physics in the DM sector[27] or strong

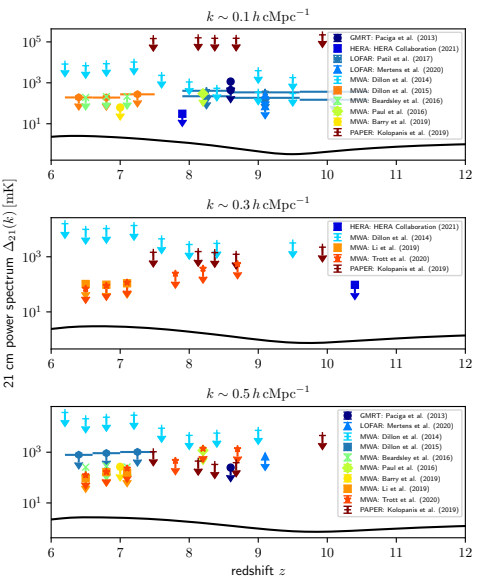


Fig. 1. Observational upper limits to the 21 cm PS, $\sqrt{k^3 P_{21}(k)/2\pi^2}$, for three different comoving wavenumbers k; here $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ and cMpc denotes comoving Mpc. Data from different telescopes, namely, the Giant Metrewave Radio Telescope[13] (GMRT), the Low Frequency Array[14,15] (LOFAR), the Murchison Widefield Array[16–22] (MWA), the Precision Array for Probing the Epoch of Reionization[23] (PAPER), the Hydrogen Epoch of Reionization[24] Array (HERA), are shown in comparison with the signal predictions from the Semi-numerical Code for ReIonization with PhoTon-conservation[25] (`SCRIPT`).

radio background from the first galaxies.[28–30] There have also been concerns regarding the foreground subtraction and other systematics in the data,[31] hence it is important to validate the signal through other low-frequency telescopes.

An alternate way of detecting the 21 cm signal is through the spatial fluctuations. These can, in principle, be useful throughout the cosmic history starting from the cosmic dawn to the post-reionization universe. During the cosmic dawn, the fluctuations arise from those in the galactic radiation, while in the epoch reionization it is the patchiness in the ionization field that drives the 21 cm fluctuation signal. In the post-reionization era, the fluctuations in the 21 cm signal essentially trace the distribution of galaxies and hence the underlying matter fluctuations.

At present, a number of low-frequency interferometers are attempting to detect the 21 cm power spectrum (PS) from the epoch of reionization at scales $\sim 10h^{-1}$ cMpc. Fig. 1 shows the current upper limits (see Ref. 32 for a complete compilation of available data). These observational attempts are ably complemented by numerical and semi-numerical theoretical models to help interpret the data. The result from the one such semi-numerical simulation `SCRIPT`[a] is shown in Fig. 1 for a model of reionization consistent with all other available observations.[25] Upcoming facilities like the Square Kilometre Array[b] (SKA) will be able to reach a noise level of ~ 1 mK (~ 0.3 mK) within ~ 100 (1000) hours of observations and hence should constrain reionization models with high statistical significance.

3. Observational results and perspectives in the microwaves

3.1. *The Planck Legacy Archive, present and future*

Planck was an ESA space satellite that measured the microwave sky in nine wavebands, allowing it to not only capture the CMB to incredible precision, but also to accurately study a number of astrophysical diffuse emissions and discrete sources (see fig. 51 in Ref. 33, which shows the various wavebands of the *Planck* satellite, as well as the foregrounds signals and their frequency spectra across those bands).

The PLA hosts and serves over 150 TB of products from *Planck* that are publicly available via the PLA web interface[c] (Fig. 2). This interface provides direct access to the users to a wide variety of products of the *Planck* Collaboration[d] through the Low Frequency Instrument (LFI) Data Processing Centre (DPC) in Trieste, Italy, by the High Frequency Instrument (HFI) DPC in Paris, France, and, more recently, by the US *Planck* Data Center in Pasadena and Berkeley, California, USA. The

[a] `https://bitbucket.org/rctirthankar/script`

[b] `https://www.skatelescope.org`

[c] `https://pla.esac.esa.int`

[d] The *Planck* Collaboration acknowledges the support of: ESA; CNES, and CNRS/INSU-IN2P3-INP (France); ASI, CNR, and INAF (Italy); NASA and DoE (USA); STFC and UKSA (UK); CSIC, MINECO, JA, and RES (Spain); Tekes, AoF, and CSC (Finland); DLR and MPG (Germany); CSA (Canada); DTU Space (Denmark); SER/SSO (Switzerland); RCN (Norway); SFI (Ireland); FCT/MCTES (Portugal); ERC and PRACE (EU). A description of the *Planck* Collaboration and a list of its members, indicating which technical or scientific activities they have been involved in, can be found at `http://www.cosmos.esa.int/web/planck/planck-collaboration`.

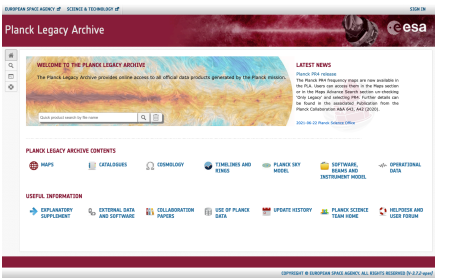


Fig. 2. *Left*: the *Planck* Legacy Archive home page.

PLA is located at the European Space Astronomy Centre (ESAC) near Madrid, Spain, and is maintained by the Data Science and Archives Division personnel.

In the last years, four major releases of *Planck* products have taken place, providing users with the most complete cosmology data set to date.

The *Planck* products can be divided into various categories. *First*, the time-ordered data, that comes in multiple flavors: raw, semi-calibrated and calibrated science ready timelines and compressed timelines (called HPRs). *Second*, the frequency maps produced by the DPCs, in most cases for each detector, detector pair and frequency, and for different combinations of the mission coverage (e.g., yearly maps, nominal mission, half-mission, half-ring, full mission, etc); *third*, the maps of the CMB; and, *fourth*, the so-called astrophysical component maps that contain the diffuse emission from our Galaxy as produced by the dust grains, synchrotron emission from spiraling electrons moving along the magnetic field lines, cosmic infrared background (CIB), etc. See Fig. 3, *left* panel. These CMB and astrophysical maps are derived in the scheme offered by the Hierarchical Equal Area isoLatitude Pixelization of a sphere[34] (`HEALPix`) using different component separation methods, e.g., `Commander`, Spectral Matching Independent Component Analysis (`SMICA`), Spectral Estimation Via Expectation Maximisation (`SEVEM`) and Generalized Needlet Internal Linear Combination (`GNILC`), the four official component separation methods used to extract the CMB maps in the third *Planck* Release (PR3). *Fifth*, source catalogues in various forms, from Galactic and extragalactic compact source catalogues, available in the *Planck* Catalogue of Compact Sources (PCCS), to galactic cold cores, SZ GCs, high-z lensing candidates, etc (Fig. 3, *right* panel). *Sixth*, cosmology products, e.g., the *Planck* likelihood code and associated files, the different flavors of the *Planck* CMB angular PS as a function of the multipole ℓ (e.g., low-ℓ, high-ℓ, temperature-only, temperature and polarization, etc), lensing products and more (Fig. 4, *left* panel).

Advance Search panels are available to extensively query the PLA database, in addition to embedded links to the *Planck* Explanatory Supplement documentation,[e] multiple data download options, and Helpdesk support.

[e] `https://wiki.cosmos.esa.int/planck-legacy-archive`

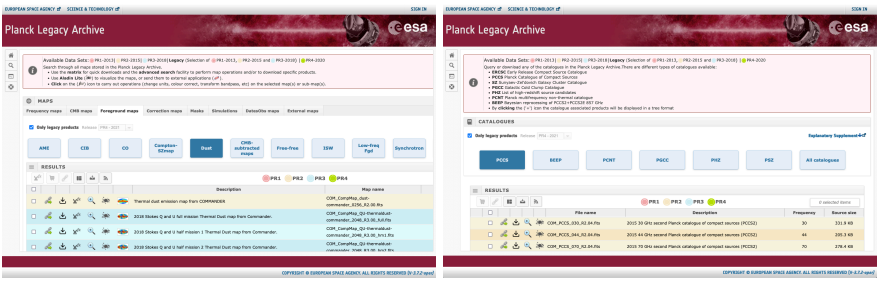


Fig. 3. *Left*: the *Planck* maps area in the PLA, with emphasis to the astrophysical components, and in particular to the Galaxy dust emission. *Right*: the *Planck* source catalogues area in the PLA, with emphasis to the PCCS.

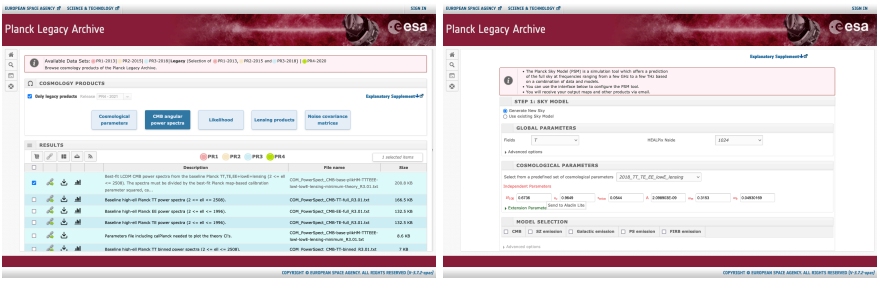


Fig. 4. *Left*: the *Planck* cosmology products area in the PLA, with emphasis to the CMB angular PS. *Right*: the `Planck Sky Model` simulation tool area in the PLA.

Three major releases of *Planck* products took place in 2013, 2015, and 2018 and a selection of products have been tagged as "Legacy" to identify the version of each product most appropriate for general scientific use. In 2021 a new release of products will take place with a joint reprocessing of LFI+HFI time-ordered data that includes additional information not used in previous releases. In addition, EU funded projects reprocessing *Planck* data, or combining it with other experiments, are expected to the deliver to the PLA higher level data products of interest. The PLA also offers specialized tools that facilitate the processing of *Planck* products. These tools are mainly designed to help users who are not familiar with some of the particularities of the *Planck* products, and can be categorized into distinct groups: map operations including component subtraction, unit conversion, colour correction, bandpass transformation, and masking of map-cutouts/full-sky maps; component separation codes, map-making codes and effective beam-averaging. In addition, the PLA includes an interface to the latest version of the `Planck Sky Model` simulation tool (Fig. 4, *right* panel), with a simple user interface that allows users to simulate the microwave/sub-mm sky with *Planck*, as well as future CMB experiments and custom-defined instruments.

In the coming years it is planned to continue releasing data products from improved reprocessing of *Planck* data and combining *Planck* with other experiments.

3.2. *High angular resolution SZ observations with MISTRAL*

High angular resolution mm observations are key to understand a wide variety of scientific cases. Among the others, the interaction of CMB photons with the hot electron gas in GCs and surrounding medium promises to study GCs and their deviation from relaxed behaviour. When the CMB photons scatter off a hot electron gas, for example in GCs, in fact, they undergo inverse Compton scattering which is visible in the frequency spectrum of the CMB. The resulting distorted spectrum is

$$\frac{\Delta I(x)}{I_0} = y \frac{x^4 e^x}{(e^x - 1)^2} \left(x \coth \frac{x}{2} - 4 \right) = y g(x) \,, \tag{1}$$

where $I_0 = (2h_P/c^2)(k_B T_{CMB}/h_P)^3$, T_{CMB} is the CMB temperature, $x = h_P \nu/(k_B T_{CMB})$ is the dimensionless frequency, and $y = \int n_e \sigma_T k_B T_e/(m_e c^2) dl$ is the Comptonization parameter, the integral along the line of sight of the electron density n_e multiplied by the electron temperature T_e; σ_T is cross-section of Thomson scattering, k_B and h_P are the Boltzmann and Planck constants. This is the tSZ effect[35,36] and it can be used to study both relaxed and non-relaxed GCs. In fact, GCs can experience a wide variety of situation including collisions, merging, and non-relaxed situation. In addition, not only GCs can host relativistic electrons. Hydrodynamical simulations[37,38] suggested that GCs occupy the knots of the so called cosmic web (CW) and WHIM is disposed connecting GCs forming filaments that could be seen through the SZ effect itself.

An instrument being prepared for such scientific cases is the MIllimeter Sardinia radio Telescope Receiver based on Array of Lumped elements kids[39] (MISTRAL), to be fielded at the 64m Sardinia Radio Telescope[f] (SRT). MISTRAL will use an array of 408 kinetic inductance detectors (KIDs) with a field of view of $4'$ and an angular resolution of $\sim 12''$. It is a cryogenic instrument with detectors cooled down at 270 mK and with frequency domain multiplexing (FDM), Reconfigurable Open Architecture Computing Hardware 2 (ROACH2) based read-out.

MISTRAL is a facility instrument to be installed at the SRT in 2022. It operates in an atmospheric window in the frequency range $78 \div 103$ GHz, namely W-band, that is interesting for a number of scientific reasons, and most importantly it provides low optical depth and allows high efficiency observations. It will be installed in the Gregorian room of the SRT (Fig. 5).

MISTRAL consists of a cryostat, being constructed by Queen Mary College (QMC) Instruments,[g] with a Pulse Tube (PT) cryocooler, an He-10 sorption refrigerator, custom optics and detectors. In order to have the PT heads as close as possible to vertical position during observations, we have positioned the head with an inclination of $57.5°$ with respect to the focal plane. This allows observation elevations in the range of $32.5° \div 82.5°$ with no degradation of the thermal performance.

[f] `http://www.srt.inaf.it`
[g] `http://www.terahertz.co.uk/qmc-instruments-ltd`

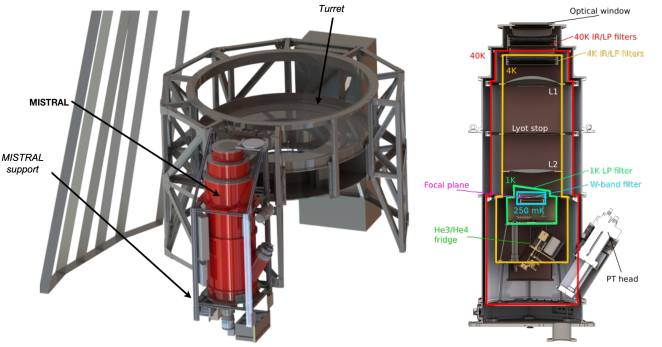


Fig. 5. *Left*: SRT Turret with MISTRAL installed. *Right*: a cut of MISTRAL cryostat highlighting the main parts. Acknowledgement: MISTRAL Collaboration.

The optical design of MISTRAL includes a set of radiation quasi-optical filters, anchored at the different thermal stages of the cryostat, an Anti Reflection Coated (ARC) Ultra High Molecular Weight (UHMW) polyetilene window, and two Silicon lenses able to image the Gregorian Focus on the array of detectors. Detectors are coupled to radiation through open space (filled array) so cryogenic cold stop, placed in between the two lenses, is needed.

MISTRAL will take advantage of the high sensitivity and capability of FDM of KIDs cryogenic detectors. KIDs are superconductive detectors where mm radiation with higher energy with respect to the Cooper pair binding energy, can break Cooper pairs producing a change in the population densities and thus in the kinetic inductance. In fact, the inductance L of a thin superconductor is dominated by the kinetic inductance L_k, which depends on the Cooper pair density. The detector array is composed of 408 KIDs detectors. They are $3\,\text{mm} \times 3\,\text{mm}$ each and are arranged on an equilateral triangle every 4.2 mm on a 4 inches silicon wafer. They sample the focal plane with a full width half maximum (FWHM) angular spacing of $10.6''$ lower the each pixel angular resolution. The ROACH2 based FDM, originally developed for the Osservatorio nel Lontano Infrarosso Montato su Pallone Orientabile[h] (OLIMPO) experiment,[40,41] will be used.

MISTRAL will be a facility instrument: it will be open to the scientific community to decide what kind of scientific output it can achieve and propose the observations to the Time Allocation Committee of the SRT. Nevertheless, we will list some scientific cases below, including those we find most interesting.

Protoplanetary discs mm measurements in star forming regions allow to break the degeneracy present in infrared (IR) measurements due to the optically thick nature of the hot inner disc.[42] Observations of star formation in molecular clouds with high enough angular resolution allow to distinguish starless cores with respect to those hosting protostars.[43] Continuum resolved galaxies observations can give

[h]Far Infrared Observatory Mounted on Orientable Balloon. See `http://olimpo.roma1.infn.it.`

information about morphology and radial profiles, e.g., gas column profiles, dust temperature profiles.[44] Spatially resolved high-z radio galaxies can provide information about cold dust re-emission.[45]

What is probably the most impacting expected result of MISTRAL is the possibility to observe and resolve the SZ effect through GCs and surrounding medium. With respect to moderate to low angular resolution observations, high ($\simeq 10''$) etections can investigate non relaxed GCs, study merging GCs, understanding the self similarities of GCs, their identical appearance regardless of their mass or distance. Study the pressure profiles and understand the active galactic nuclei (AGN) feedback expected in some environments. Often GCs are assumed to be spherical and isothermal. Nevertheless, they interact, collide and merge, bringing to important degeneracies and deviation from specific models. High angular resolution SZ measurements allow to disentangle between models and to identify the most appropriate ones. Even relaxed GCs experience pressure fluctuations and compressions of the intra cluster medium which would be interesting to study.

GCs experience hierarchical growth through mergers. Pre-mergers GCs should be connected through the CW and simulations endorse this.[37,38] WHIM are expected to be distributed as over-densities in filamentary structures between GCs: high angular resolution SZ measurements can detect WHIM better than X-ray under low density circumstances. This is probably the most challenging and most rewarding of the achievements that an experiment like MISTRAL can achieve.

4. *Planck* cosmological implications

4.1. *New Planck thermal SZ map and its cosmological analysis*

The ΛCDM model provides a wonderful fit to many cosmological data. However, a slight discrepancy was found in the latest data analyses between the CMB anisotropies ($z \sim 1100$) and other cosmological probes at low ($z \sim 0 \div 1$) redshift, z, for the $S_8(\equiv \sigma_8(\Omega_M/0.3)^{0.5})$ cosmological parameter representing the amplitude of the structure growth in the universe; here Ω_M and σ_8 are the usual non-relativistic matter density parameter and the density contrast on a scale of 8 $[H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})]^{-1}$ Mpc.

For example, the S_8 value was precisely measured to be $S_8 = 0.830 \pm 0.013$ with the *Planck* CMB observation.[46] However, a lower value of $S_8 \sim 0.77 \pm 0.03$ was measured using the population of GCs detected by *Planck* at low-z[47,48] ($z < 0.6$), thus called "S_8 tension". Furthermore, other low-z observations with gravitational lensing by the Kilo-Degree Survey[49] (KiDS) and DES experiments also found lower S_8 values. These results indicate that the cosmic structure growth is slower than the prediction based on the CMB measurement and may demand modifications to the standard cosmological model to explain all these measurements.

One of the most interesting signals extensively studied with *Planck* is the tSZ effect, which produces what are known as y distortions, or y type signals. These primarily come from CMB photons being inverse-Compton scattered through hot

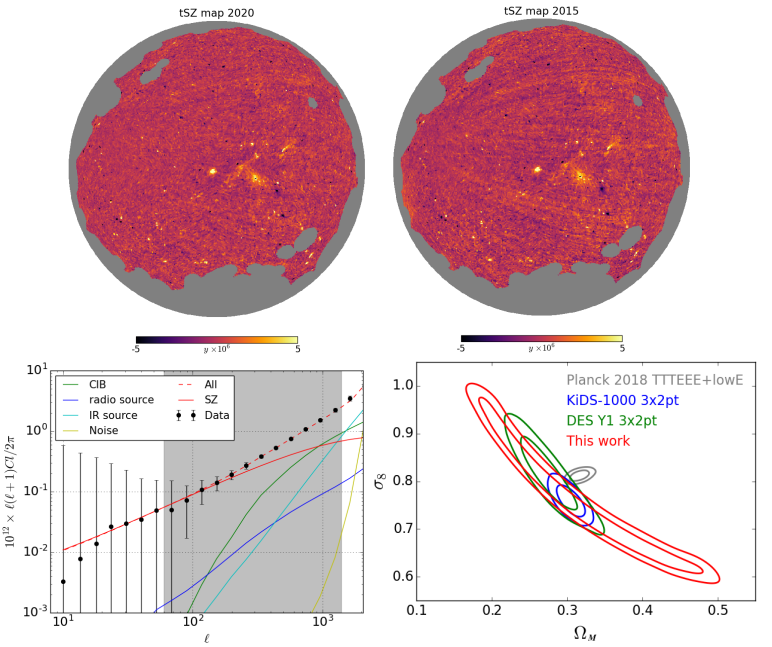


Fig. 6. *Top*: Orthographic projections of the all-sky Compton y parameter maps reconstructed with `MILCA` (*left*) and in 2015[50] (*right*), with appreciable differences in the striping features. Pixel resolution changed for visualization purposes from $N_{\rm side} = 2048$ to $N_{\rm side} = 128$ (i.e. from $\simeq 1.72'$ to $\simeq 27.48'$ pixel size), according to the sky pixelization in Ref. 34. *Bottom*: Cross-PS (black) fitted with the tSZ (red) and foreground models, namely the CIB (green), radio (blue) and IR (cyan) sources, plus noise (yellow), the sum of the tSZ, foreground models and noise being shown in red dashed line (*left*); posterior distribution of the cosmological parameters σ_8 and Ω_M, with 68% and 95% confidence level (CL) interval contours obtained from our cosmological analysis (*right*), compared with the *Planck* CMB's (gray), KiDS-1000 3x2pt (blue), and DES Y1 3x2pt results (green). Acknowledgements: H. Tanimura, M. Douspis, N. Aghanim (CNRS/Univ. Paris-Saclay).

GCs (see Sect. 3.2), which makes holes, or lowers the flux of low frequency photons and up scatters, or makes an excess flux of high frequency photons.

The tSZ signal is subdominant relative to the CMB and other foreground emissions in the *Planck* bands. Thus a tailored component separation algorithm is required to reconstruct the tSZ map. We adopted the Modified Internal Linear Combination Algorithm[51] (`MILCA`) used for the *Planck* y-map reconstruction in 2015 and applied it to the 100 to 857 GHz frequency channel maps (see Sect. 3.1) from the fourth *Planck* Release (PR4).

The *Planck* PR4 data implemented several improvements from the previous version: the usage of foreground polarization priors during the calibration stage to break scanning-induced degeneracies, the correction of bandpass mismatch at all frequencies, the inclusion of 8% of data collected during repointing maneuvers, etc. With these improvements we produced a new y-map with smaller noises (by $\sim$7%) and reduced strips than the previous version released in 2015 (Fig. 6, *top* panels).

We also reconstructed two y-maps from the first and last halves of the data for a cosmological analysis and considered a cross-angular PS between these y-maps to avoid the bias induced by the noise in the auto-angular PS (see Ref. 52 for more details). This tSZ cross-PS is affected by residual foreground emissions from radio and IR point sources and CIB emission. Thus we fitted our measurement with the cosmology-dependent tSZ model, including the radio and IR point source models,[53] the CIB model[54] and the noise (Fig. 6, *left bottom* panel).

The cosmological analysis of the tSZ cross-PS allowed us to set constraints on cosmological parameters, mainly of the S_8 parameter. We obtained $S_8 = 0.764\,^{+0.015}_{-0.018}\,(\mathit{statistical})\,^{+0.031}_{-0.016}\,(\mathit{systematic})$, in which the systematic uncertainty includes the contributions from the mass bias and pressure profile model. Our obtained S_8 value is fully consistent with recent weak lensing results from KiDS and DES. It is also consistent with the *Planck* CMB's result[46] within 2σ, while it is slightly lower by $\sim 1.7\sigma$ (Fig. 6, *right bottom* panel).

4.2. *The CMB dipole and the thermal SZ effect: Eppur Si Muove*

The largest temperature anisotropy in the CMB is the dipole pattern at $\ell = 1$. The simplest interpretation of the dipole is that it is due to our motion with respect to the rest frame of the CMB (with debate over the possibility of alternative explanations), and is measured to be $\beta = v/c = (1.23357 \pm 0.00036) \times 10^{-3}$ in the direction of the constellation Crater, $(l, b) = (264.021^\circ \pm 0.011^\circ, 48.253^\circ \pm 0.005^\circ)$, see Ref. 55. As well as creating the ℓ=1 mode of the CMB sky, this motion affects all astrophysical observations by modulating and aberrating sources across the sky, such as the CMB temperature anisotropies. These and other effects will be discussed.

As discussed above, an accurate mapping of the tSZ signal on the sky has been achieved with the *Planck* satellite. This allows us to construct a novel and independent measure of the CMB dipole because the y maps are contaminated by temperature anisotropies stemming from the CMB dipole. This is also valuable as a test of the quality of the y maps.

To derive the connection between the y map and the dipole we will begin by defining some useful terms regarding the unboosted CMB sky,

$$I(x) \equiv \frac{2k_{\rm B}^3 T_{CMB}^3}{h_P^2 c^2} \frac{x^3}{e^x - 1} \tag{2}$$

$$f(x) \equiv \frac{x e^x}{e^x - 1} \tag{3}$$

$$Y(x) \equiv x \frac{e^x + 1}{e^x - 1} - 4\,, \tag{4}$$

which are the standard Planck blackbody intensity function, the frequency dependence of CMB anisotropies, and the relative frequency dependence of the tSZ effect

or y type distortions, x being the dimensionless frequency (see Sect. 3.2). Thus, to first-order the anisotropies of intensity measured by *Planck* can be written as

$$\frac{\delta I(\hat{\mathbf{n}})}{If(x)} = \frac{\delta T(\hat{\mathbf{n}})}{T_{CMB}} + y(\hat{\mathbf{n}})Y(x)\,, \tag{5}$$

where we have included just the temperature anisotropies and the y signals here.

If we apply a boost to Eq. (5) with a dimensionless velocity $\boldsymbol{\beta}$, it becomes

$$\begin{aligned}\frac{\delta I'(\hat{\mathbf{n}}')}{If(x)} = {} & \beta\mu + \frac{\delta T(\hat{\mathbf{n}}')}{T_{CMB}}(1+3\beta\mu) \\ & + Y(x)\left(y(\hat{\mathbf{n}}')(1+3\beta\mu) + \beta\mu\frac{\delta T(\hat{\mathbf{n}}')}{T_{CMB}}\right) \\ & + \beta\mu y(\hat{\mathbf{n}}')\left(Y^2(x) - x\frac{dY(x)}{dx}\right) + \mathcal{O}(\beta^2)\,;\end{aligned} \tag{6}$$

here $\mu = \cos(\theta)$, and θ is the angle from the direction of the boost $\boldsymbol{\beta}$ and the line of sight. Additionally, the aberration effect deflects the direction of the incoming photons from $\hat{\mathbf{n}} \to \hat{\mathbf{n}}'$ where $\hat{\mathbf{n}}' = \hat{\mathbf{n}} - \nabla(\hat{\mathbf{n}}\cdot\boldsymbol{\beta})$. The first line has the same frequency dependence as thermal fluctuations. The term $\beta\mu$ simply describes the pure CMB dipole, as discussed previously. The second term represents the dipole aberration and modulation of the temperature anisotropies of the CMB. This was first measured in Ref. 56 to 5σ. Crucially for our analysis, the middle line has the same frequency dependence as y type distortions, and so encompasses the signals in the y-map. The final line has more obscure frequency dependencies and so will not be discussed here. See Refs. 57, 58 for further details.

The second line of Eq. (6) shows the signals in the y maps produced by *Planck*. The first half is the same boost as was seen in the temperature anisotropies, however, the final term is due to the second-order expansion of the intensity about T_{CMB} and adds a contribution to the y maps from the temperature anisotropies. We measured this signal in Ref. 57 to 6σ by cross-correlating the expected modulated temperature map with the y maps. We confirmed the dipole direction by cross-correlating the temperature maps modulated in two additional perpendicular directions. To test consistency and rule out different systematics, we used the Internal Linear Combination (`ILC`) in two dimensions, `2D-ILC`, and `MILCA` y-maps, the temperature anisotropies from both `2D-ILC` and `SMICA`, and finally both a harmonic space analysis and a map space analysis.

The question as to whether an intrinsic dipole could be observationally distinguished from an extrinsic dipole remains an open question. The terms discussed in Eq. (6) are based on the assumption of a CMB blackbody spectrum[i] and cannot be used to distinguish the two, as they would naturally arise whether the CMB dipole

[i]See instead Ref. 59 for a study exploiting the relaxation of the blackbody assumption and references to other methods.

were caused by a boost, or if there were simply a dipole in the sky with the same magnitude and direction.

4.3. *Planck constraints on the primordial tensor-to-scalar ratio*

Discovering the stochastic field of gravitational waves generated during the early phases of the universe is one of the most ambitious goals of modern cosmology. Various models of inflation predict cosmological perturbations from quantum fluctuations[60–66] in terms of tensor component (gravitational waves) and scalar component (density variations). Inflationary gravitational waves entering the horizon since recombination epoch generate a tensor component contributing to temperature (T), polarization (E and B modes) and cross-correlation TE angular PS, C_ℓ^X where X stands for a given mode[j], of CMB anisotropy,[67–69] in particular at large angular scale (or at low multipoles, $\ell \lesssim 150$, where $\ell \sim 180°/\theta$ approximately relates ℓ to the angular scale θ). C_ℓ^X is typically represented in terms of $D_\ell^X = \ell(\ell+1)C_\ell^X/(2\pi)$, which is nearly flat at low multipoles for $X = T$, although a power suppression at very large scales, first observed by the Cosmic Background Explorer (COBE) an then confirmed by the Wilkinson Microwave Anisotropy Probe (WMAP) and by the *Planck* satellite, is predicted in several inflationary models possibly in connection with universe geometry and topology.

While the E and TE modes contain both scalar and tensor signals mainly coming from the epochs of recombination and reionization, the model does not predict primordial scalar fluctuations in the B-mode. The detection of the primordial B-mode then constitutes a direct way to unravel tensor perturbations from the early universe, and the efforts to measure of the ratio, r, between the amplitudes of primordial tensor and scalar perturbations is the main scientific aim of many ongoing and future CMB projects. When CMB photons pass through gravitational potentials produced by cosmic structures, part of the E-mode power is transformed to B-mode power, the so-called lensing B-mode. This signal, already detected by a number of ground-based observatories and by *Planck*, is particularly important at small angular scale[70] ($\ell \gtrsim 150$), but it may mask the primordial B-mode also at intermediate angular scale and, for low r, even at large angular scale.

Assuming the spectra of scalar and tensor perturbations are described by pure power-laws, $\mathcal{P}_\mathrm{s}(k) = A_\mathrm{s}(k/k_0)^{n_\mathrm{s}-1}$, $\mathcal{P}_\mathrm{t}(k) = A_\mathrm{t}(k/k_0)^{n_\mathrm{t}}$, they are fully defined by the two amplitude parameters $A_\mathrm{s}, A_\mathrm{t}$ and by the so-called (comoving wavenumber) pivot scale k_0, the definition of $r = A_\mathrm{t}/A_\mathrm{s}$ being then related to the choice of k_0. Among the various choices adopted in the literature, $k_0 = 0.05\,\mathrm{Mpc}^{-1}$ approximately corresponds to the middle of the logarithmic range of k probed by *Planck*. For historical reasons, a scale-invariant spectrum corresponds to $n_\mathrm{s} = 1$ or to $n_\mathrm{t} = 0$ for scalar or tensor perturbations, respectively.

[j]For simplicity, we use here two letters only for cross modes.

The multipole range $2 \lesssim \ell \lesssim 150$ is the most advantageous to directly extract r from B-mode analyses. According to theoretical predictions, the primordial D_ℓ^B is close to the maximum at $50 \lesssim \ell \lesssim 150$, the region of the so-called recombination bump (i); at large scale, $\ell \lesssim 10$, D_ℓ^B exhibits a pronounced bump generated at late epochs by the reionization process (ii); at $10 \lesssim \ell \lesssim 50$, $\log(D_\ell^B)$ increases almost linearly with $\log(\ell)$ (iii). Exploiting higher ℓ's is of particular relevance for a very accurate treatment (delensing) of lensing B-mode. C_ℓ^B increases almost linearly with r, the reionization bump features depending on r and on the details of the reionization process, which is parametrized to first approximation by the corresponding Thomson optical depth, τ. In general, the detection of primordial B-modes requires exquisite experimental sensitivity and accuracy as well as very precise treatment of the polarized foreground emissions (see, e.g., Ref. 71), particularly for low values of r, which is why a primordial B-mode tensor signal has not yet been detected. On the other hand, recent CMB projects set significant upper limits to r. The great sensitivity achievable by ground-based experiments on selected sky areas has been used to strongly constrain (i), the current tightest B-mode limits on r from these scales coming from the Background Imaging of Cosmic Extragalactic Polarization (BICEP)/Keck measurements[72] (BK15), while the *Planck* all-sky survey allows to constrain (i), (ii) and (iii). In particular, Ref. 73 presents B-mode limits on r from *Planck* data at $\ell \lesssim 30$.

A recent reanalysis of *Planck* data alone and in combination with BK15 (see Ref. 74 for a previous joint analysis) has been performed in Ref. 75 exploiting the whole multipole range $2 \leq \ell \leq 150$. In this study, the *Planck* PR4 data (see Sect. 3.1) were adopted, where the `NPIPE`[76] processing pipeline was used to create calibrated frequency maps in temperature and polarization from LFI and HFI data, providing several improvements in noise and systematics levels in both frequency and component-separated maps. Various sky masks were considered in order to retain different sky fractions (from 30% to 70%) and to test the impact of data treatment. In polarization, the CMB sky was separated from foregrounds applying the `Commander`[77] code to a model with three components, namely the CMB, synchrotron, and thermal dust emission, starting from the PLA PR4 maps but downgraded from $N_{\rm side} = 2048$ to $N_{\rm side} = 16$ and to $N_{\rm side} = 1024$ respectively for the analysis at $2 \leq \ell \leq 35$ and at $35 \leq \ell \leq 150$. An updated version of the LOw-ℓ LIkelihood on POlarised Power-spectra (`LoLLiPoP`) code previously employed by the *Planck* Collaboration in the reionization analysis[78] was used to derive the likelihood from *cross*-power spectra for the CMB maps reprocessed as outlined above. Here, *cross* denotes that the angular PS is extracted analyzing different sets of detectors: indeed, for these *cross*-power spectra the bias is zero when the noise is uncorrelated between maps. Unbiased estimates of the angular PS were derived using an extension of the quadratic maximum likelihood estimator and a classical pseudo-C_ℓ estimator at $2 \leq \ell \leq 35$ and at $35 \leq \ell \leq 150$, respectively. The C_ℓ covariance is deduced from end-to-end simulations and thus includes CMB sample variance,

statistical noise, residuals from systematics, uncertainties from foreground subtraction, and correlations generated by masking. These uncertainties are propagated through the likelihood up to the level of cosmological parameters.

Applying this analysis to the B-mode alone, fixing the other cosmological parameters to the fiducial ΛCDM spectrum of *Planck* 2018 results, and considering the cleanest 50% fraction of the sky, the values of r (for $k_0 = 0.05\,\mathrm{Mpc}^{-1}$) retrieved from the posterior distribution analysis at $2 \leq \ell \leq 35$ (that includes the reionization bump) or at $35 \leq \ell \leq 150$ (that includes the recombination bump) are $r_{0.05} = -0.014^{+0.108}_{-0.111}$ or $r_{0.05} = +0.069^{+0.114}_{-0.113}$, respectively. The two multipole windows contribute almost equally to the overall *Planck* sensitivity to r. Combining them gives $r_{0.05} = 0.033 \pm 0.069$, while setting also a positive value of r as a physical prior implies $r_{0.05} < 0.158$ at 95% CL.

A tighter result, $r_{0.05} = -0.031 \pm 0.046$, and a stronger constraint, $r_{0.05} < 0.069$ at 95% CL, slightly improving the BK15 95% CL limit $r_{0.05} < 0.072$, is derived applying the same method but jointly considering the three polarization modes B, E and EB.[k]

These results on r do not depend significantly on the choice of other ΛCDM cosmological parameters. In particular, the constraints on r are unchanged when r is retrieved together with τ: considering all polarization modes $r_{0.05} = -0.015 \pm 0.045$ and $\tau = 0.0577 \pm 0.0056$, while considering only the B-mode $r_{0.05} = 0.025 \pm 0.064$ and τ is undetermined because of the noise.

The *Planck* temperature data are crucial to determine the other ΛCDM parameters and then, when combined with polarization data, to exhaustively verify if the limits on r depend on the other parameters. Furthermore, although *Planck* temperature data, when considered alone, are found to be about two times less sensitive than *Planck* polarization data in constraining r, the tensor contribution to temperature fluctuations allows to further tight the constraints on r derived from polarization data alone. In fact, including also *Planck* temperature data in the analysis using the *Planck* public low-ℓ temperature-only likelihood based on the PR3 CMB map recovered from the `Commander` and the High-ℓ Likelihood on Polarised Power-spectra (`HiLLiPoP`) at $\ell > 30$, and marginalizing over the nuisance and the other ΛCDM cosmological parameters, from the posterior distribution for r the authors[75] derive the currently most stringent upper limit based on *Planck* data alone: $r_{0.05} < 0.056$ at 95% CL.

Finally, in Ref. 75 the *Planck* temperature and polarization data analyzed as summarized above have been combined with the BK15 data, assuming their mutual independence, i.e. simply multiplying their likelihood distributions, as justified given their so different sky coverage, obtaining $r_{0.05} < 0.044$ at 95% CL which currently represents the most stringent upper limit.

[k]The primordial EB is expected to be null in the ΛCDM model. On the other hand, special rotation effects of the polarization field (such as birefringence) can produce some EB signal.

5. Future and alternative perspectives

5.1. *Lorentz invariance violation and CMB polarization*

The standard model of cosmology can explain the CMB radiation from the recombination epoch till today. As discussed above, it predicts the existence of some degrees of linear polarization known as E-modes and B-modes (see also Refs. 79–81). The linear polarization can be induced to the CMB by means of Compton scattering of CMB photons and cosmic electrons in presence of cosmological perturbations. However, the circular polarization for the CMB is a rare process in the standard model of cosmology, so that observation of any level of circular polarization can be a hint for the existence of new physics. The circular polarization for the CMB can be generated through Faraday conversion in which linear polarized light propagates through a medium resulting in different indexes of refraction along the two transverse axes.[82,83] Moreover, the CMB photons scattering from vector or fermionic DM can lead to circular polarization of the CMB.[84,85] Furthermore, circular polarization for the CMB can be induced by electron-photon Compton scattering in presence of a background field such as a magnetic field[82] or non-commutativity in space and time.[83,86]

In this study, we would like to explore the effects of Lorentz invariance violation (LIV) as a new background in the generation of circular polarization (see Ref. 87 for details). To do that we assume the CMB photon and cosmic electron scatter while the particle Lorentz symmetry is violated. To evaluate the LIV effect in the Compton scattering, we focus on an effective field theory approach known as standard model extension (SME). The minimal version of the SME contains every gauge invariant and observer-covariant operator, made by all the standard model (SM) fields which violate Lorentz invariance specified with new couplings.[88] In the SME, LIV is considered in a coordinate that the speed of light is constant c in all frames. The kinetic term in electron sector of electromagnetic quantum electrodynamics (QED) Lagrangian is modified with $c^{\mu\nu}$ tensor as[89]

$$\mathcal{L}_{\rm QED}^{\rm LIV} = \frac{1}{2} i\bar{\psi}(\gamma^{\nu} + c^{\mu\nu}\gamma_{\mu})\overleftrightarrow{D}_{\mu}\psi - \bar{\psi}m_e\psi\,, \tag{7}$$

where $\mathcal{D}_{\mu}$ represents the covariant derivative with $u\overleftrightarrow{D}^{\mu}v \equiv uD^{\mu}v - vD^{\mu}u$, ψ is the fermion field, γ_{μ} indicates the Dirac gamma matrices, and the mass of electron is m_e. We assume $c^{\mu\nu}$ to be symmetric and traceless tensor. The $c^{\mu\nu}$ tensor depends on the frame and it is estimated in the standard reference frame, i.e. Sun-centered, celestial equatorial frame (SCCEF).[90] All the available bounds on LIV parameters are given in SCCEF frame to be comparable with each other. Therefore, we also use correlation between components of laboratory frame and SCCEF given in Ref. 90 to transfer the result to SCCEF frame.

The polarization of the scattered CMB photon from cosmic electron in presence of LIV is characterized by four Stokes parameters:[79] $[I, Q, U, V]$. I indicates the overall radiation intensity, Q and U represent linear polarization and V is circular

polarization of the CMB radiation satisfying the inequality $I^2 \geq Q^2 + U^2 + V^2$. The time evolution of Stokes parameters, defining the CMB polarization, can be expressed through the quantum Boltzmann equation, schematically as[79]

$$\frac{df}{dt} = C[f]\,, \tag{8}$$

where the left-hand side involves time derivative of Stokes parameters which includes the gravitational effects and space time structure; besides, all possible interactions appear at the right-hand side of Boltzmann equation.

After expanding the intensity and polarization of the CMB radiation in terms of multipole moments in spin-weighted basis, the Boltzmann equations regarding to electron-photon Compton scattering in presence of LIV by considering only scalar perturbation in the metric, are obtained as

$$\begin{aligned}
&\frac{d}{d\eta}\Delta_I^{(S)} + iK\mu\Delta_I^{(S)} + 4[\dot{\psi} - iK\mu\varphi] = C_{e\gamma}^I \\
&\frac{d}{d\eta}\Delta_P^{\pm(S)} + iK\mu\Delta_P^{\pm(S)} = C_{e\gamma}^{\pm} - i\kappa_{\rm LIV}^{\pm}\,\Delta_V^{(S)} \\
&\frac{d}{d\eta}\Delta_V^{(S)} + iK\mu\Delta_V^{(S)} = C_{e\gamma}^V + \frac{i}{2}\Big[\kappa_{\rm LIV}^{+}\,\Delta_P^{-(S)} + \kappa_{\rm LIV}^{-}\,\Delta_P^{+(S)}\Big]\,;
\end{aligned} \tag{9}$$

here η is the conformal time, ψ and ϕ are the scalar perturbation corresponding to the potential of Newton and the spatial curvature perturbation, μ is the scalar product of the wave vector $\mathbf{K}$ and the CMB photon propagating direction, $\Delta_P^{\pm(S)} = Q^{(S)} \pm iU^{(S)}$ is a linear combination of linear polarizations, $C_{e\gamma}^I$, $C_{e\gamma}^{\pm}$ and $C_{e\gamma}^V$ indicate the involvement of the SM Compton scattering in $\Delta_P^{\pm(S)}$ and $\Delta_V^{(S)}$ parameters, respectively. In presence of LIV, the terms $\kappa_{\rm LIV}$ and $\kappa_{\rm LIV}^{\pm}$ in Eq. (9) are given by

$$\kappa_{\rm LIV} = -a(\eta)\,\frac{3}{4}\,\frac{\sigma_T}{\alpha}\,\frac{m_e^2}{k_0}\,\frac{\bar{n}_e}{m_e}\,,\;\; \kappa_{\rm LIV}^{\pm} = \kappa_{\rm LIV}\, c_{\{\alpha\beta\}}(\rho_Q^{\{\alpha\beta\}} \pm i\rho_U^{\{\alpha\beta\}})\,; \tag{10}$$

they define the Compton scattering contribution, k_0 being the momentum of the CMB photon. In general, the CMB polarization at the direction $\hat{\mathbf{n}}$ and at the present time η_0 is derived by integrating Eq. (9) along the line of sight and summing over all Fourier modes $\mathbf{K}$. The circular polarization due to LIV Compton scattering can finally be estimated in terms of linear polarization as follows

$$\begin{aligned}
\Delta_V^{(S)}(\mathbf{K},\mu,\eta_0) \approx \frac{1}{2}\int_0^{\eta_0} d\eta\,\dot{\eta}_{e\gamma}\,e^{ix\mu-\eta_{e\gamma}}\;\Big[&3\mu\Delta_{V1}^{(S)} + 2i\,\rho_Q^{\{\alpha\beta\}}c_{\{\alpha\beta\}}\,\frac{\kappa_{\rm LIV}}{\dot{\eta}_{e\gamma}}\,\Delta_P^{(S)} \\
&+ 2\,\frac{\kappa_{\rm LIV}^{-}\kappa_{\rm LIV}^{+}}{\dot{\eta}_{e\gamma}^2}\Delta_V^{(S)}\Big]\,,
\end{aligned} \tag{11}$$

where

$$\Delta_P^{(S)}(\mathbf{K},\mu,\eta) = \int_0^{\eta} d\eta\,\dot{\eta}_{e\gamma}\,e^{ix\mu-\eta_{e\gamma}}\;\Big[\frac{3}{4}(1-\mu^2)\Pi(K,\eta)\Big]\,, \tag{12}$$

$$\rho_Q^{\{\alpha\beta\}} = -\Big[\epsilon^1\cdot\hat{c}\cdot\epsilon^2 + \sum_{f=e,p}\frac{m_f v_b^2}{2k_0}(\hat{v}\cdot\epsilon^2\hat{v}\cdot\hat{c}\cdot\epsilon^1 + \hat{v}\cdot\epsilon^1\hat{v}\cdot\hat{c}\cdot\epsilon^2) - \frac{v_b}{2}(\hat{v}\cdot\epsilon^2\hat{k}\cdot\hat{c}\cdot\epsilon^1 + \hat{v}\cdot\epsilon^1\hat{k}\cdot\hat{c}\cdot\epsilon^2)\Big], \quad (13)$$

$\Pi = \Delta_{T2}^{(S)} + \Delta_{P2}^{(S)} + \Delta_{P0}^{(S)}$ and $\dot{\eta}_{e\gamma} = a(\eta)n_e\sigma_T$ and $\eta_{e\gamma} = \int_\eta^{\eta_0}\dot{\eta}_{e\gamma}(\eta)d\eta$ are the differential optical depth and total optical depth. For a detail calculation of Eqs. (9)-(13) see Ref. 87. The angular PS, C_ℓ^V, arisen from CMB-cosmic electron forward scattering in presence of LIV is obtained as

$$C_\ell^V \approx \frac{1}{2\ell+1}\int d^3K P_v(\mathbf{K})\times \sum_m \Big|\int d\Omega Y_{l\ell m}^*\int_0^{\eta_0} d\eta\,\dot{\eta}_{e\gamma}\,e^{ix\mu-\eta_{e\gamma}}\,2i\,\rho_Q^{\{\alpha\beta\}}c_{\{\alpha\beta\}}\,\frac{\kappa_{\rm LIV}}{\dot{\eta}_{e\gamma}}\,\Delta_P^{(S)}\Big|^2; \quad (14)$$

here $P_v(\mathbf{K})$ indicates the velocity PS which can be estimated as a function of primordial scalar spectrum $P_v(\mathbf{K},\tau)\sim P_\phi^{(S)}(\mathbf{K},\tau)$. Therefore, we can estimate the C_ℓ^V in terms of linearly polarized angular PS as follows:

$$C_\ell^V \approx (\tilde{\kappa}_{\rm LIV}^{\rm avg})^2 C_\ell^P, \quad (15)$$

with

$$\tilde{\kappa}_{\rm LIV} = c_{\{\alpha\beta\}}\,\rho_Q^{\{\alpha\beta\}}\frac{\kappa_{\rm LIV}}{\dot{\tau}_{e\gamma}}, \qquad \tilde{\kappa}_{\rm LIV}^{\rm avg} = \frac{1}{z_{lss}}\int_0^{z_{lss}} dz\;\tilde{\kappa}_{\rm LIV}(z)\simeq 10^{-3}, \quad (16)$$

where $\tilde{\kappa}_{\rm LIV}^{\rm avg}$ is average of $\tilde{\kappa}_{\rm LIV}(z)$ over redshift from last scattering surface, $z_{lss}\simeq 1100$, up to today, $z=0$. Based on recent experimental astrophysics bounds on different components of $c^{\alpha\beta}$ tensor,[91] we assume all component of $c^{\alpha\beta}$ tensor at the order of 10^{-15}. Considering the experimental value for the CMB linearly polarized angular PS of the order of[92] $C_\ell^P\equiv 0.1\mu{\rm K}^2$ for $\ell<250$, we can estimate the V-mode angular PS arisen from Compton scattering in presence of LIV as follows:

$$C_\ell^V \approx 0.8\ {\rm nK}^2\Big(\frac{c^{\alpha\beta}}{10^{-15}}\Big)^2. \quad (17)$$

The current experimental bound on the circular polarization coming from the Cosmology Large Angular Scale Surveyor (CLASS) experiment with the 40 GHz polarimeter ranging from $0.4\mu{\rm K}^2$ to $13.5\mu{\rm K}^2$ for $1\leq\ell\leq 120$. We expect that by improving the sensitivity of experiments in future, the constraint obtained in Eq. (17) would be comparable to experimental bounds.

5.2. *New gravitational degrees of freedom and the DM problem*

The nature of DM is the deepest problem of modern cosmology. Assuming GR, DM appears to be a necessary ingredient in theoretical models constructed for explaining astrophysical and cosmological observations on $\sim$ kpc scales or larger. This essentially boils down to a mismatch between the observed dynamics of visible matter and its gravitational influence. Many theories of what is DM have been proposed.[93]

The DM microphysical nature, however, is not crucial to explain many cosmological observations. This leads to a simple description so that the DM dynamics are effectively governed by the collisionless Boltzmann equation, coupled to gravity.

The less explored alternative is that DM is not the cause behind all these observed phenomena, but rather that GR (and also Newtonian gravity) breaks down at ultra-low curvatures and gravitational accelerations and a new description of gravity is necessary. Changing the law of gravity, however, is not as arbitrary as it may seem.[94] Lovelock's theorem[95] tells us that GR is the unique 4-dimensional theory based on a local and diffeomorphism-invariant action where the metric is the only dynamical degree of freedom. If a non-GR gravitational theory is then assumed to explain the DM phenomenon, it is inevitable that one or more of these assumptions must be broken, and this generically leads to new dynamical degrees of freedom, that is, new dynamical fields.

Milgrom proposed[96–98] that DM is only an apparent phenomenon and that one can instead fit galactic rotation curves via a modification of the inertia/dynamics of baryons or of the gravitational law, when accelerations become smaller than $a_0 \sim 1.2 \times 10^{-10}\,\mathrm{m/s^2}$. Modifying the gravitational law was further explored by Bekenstein & Milgrom[99] who proposed that when the gradients of the potential Φ are smaller than a_0, non-relativistic gravity is effectively governed by

$$\nabla \cdot \left(\frac{|\nabla \Phi|}{a_0} \nabla \Phi \right) = 4\pi G_{\mathrm{N}} \rho \,, \tag{18}$$

where G_{N} is the Newtonian gravitational constant, and ρ the matter density. These type of models are referred to as Modified Newtonian Dynamics (MOND).

MOND has enjoyed success in fitting galactic rotation curves and reproducing the baryonic Tully-Fisher relation, while on GC scales it is found that either a_0 must be larger by a factor of $4 \div 5$ or that some DM is still needed at GC cores; see Ref. 100 for a review. As MOND is inherently non-relativistic, it is difficult to test in cosmological settings, since systems such as the CMB require a relativistic treatment. Relativistic theories that yield MOND behavior have been proposed, the most well-known being the Tensor-Vector-Scalar (TeVeS) theory,[101,102] making clear predictions regarding gravitational lensing and cosmology. In cases where the CMB and matter power spectra (MPS) have been computed, no theory has been shown to fit *all* of the cosmological data while preserving MOND phenomenology in galaxies. TeVeS in particular can be compatible with the MPS of large scale structure[103,104] but fails to fit the CMB. Moreover, it leads to a tensor mode gravitational wave (GW) speed different than the speed of light,[105] in contradiction with the Laser Interferometer Gravitational-Wave Observatory (LIGO)/Virgo observations of GW along with an electromagnetic counterpart.

The non-relativistic equation (18) can only be an effective description, due to Lovelock's theorem. To make it into a fully fledged relativistic theory, one can introduce additional fields. Building on TeVeS and later developments,[105,106] Skordis & Zlosnik[107] proposed a simple relativistic theory which in addition to the metric

$g_{\mu\nu}$ has a scalar field ϕ and unit time-like vector field A^μ. Perturbing these fields on a Minkowski background, results in perturbations of ϕ and A_μ which mix with the metric perturbation via gauge transformations, i.e. the metric potentials mix with the perturbations of ϕ and A_μ. Thus, ϕ and A_μ are gravitational fields and not DM fields, because due to this mixing it is impossible to distinguish observationally which of the three is the cause behind the "DM phenomenon", in analogy to the electric/magnetic field mixing resulting in the electromagnetic field.

The theory contains a free function $\mathcal{F}(\mathcal{Y},\mathcal{Q})$ of two arguments: $\mathcal{Q} = A^\mu\nabla_\mu\phi$ and $\mathcal{Y} = q^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi$ where $q_{\mu\nu} = g_{\mu\nu} + A_\mu A_\nu$ is the 3-metric orthogonal to A^μ. It must obey certain conditions in order for the theory to serve its purpose and to be observationally viable. To recover MOND on galactic scales, it is necessary that the reduced function $\mathcal{J}(\mathcal{Y}) \equiv \mathcal{F}(\mathcal{Y},\mathcal{Q}_0)/(2-K_\mathrm{B})$ where $\mathcal{Q}_0$ is a non-zero constant, has the limit $\mathcal{J} \to \mathcal{Y}^{3/2}/a_0$ when $\mathcal{Y} \sim |\nabla\varphi|^2$ is small. Imposing that and considering virialized systems in the weak-field approximation, leads to linear equations involving the metric potential Φ and a scalar perturbation φ, sourced by the (baryonic) matter density. These contain an additional term which is not present in MOND, but looks like a mass term for the potential, $\mu^2\Phi$, where here $\mu^2 = 2\mathcal{K}_2\mathcal{Q}_0^2/(2-K_\mathrm{B})$. It is determined by the other parameters of the theory: $\mathcal{Q}_0$, appearing in as part of the free function, K_B, related to the coupling strength of the vector field and taking values in the range $0 < K_\mathrm{B} < 2$, and $\mathcal{K}_2$, related to the coupling strength of the scalar field. In addition, the resulting equations depend on Milgrom's constant a_0. Thus in the quasistatic weak-field situation, at least two parameter combinations are important: a_0 and μ, although, depending on how the GR limit is attained, there may be further parameters as part of the free function. Due to the presence of the mass term μ, the theory departs from MOND behaviour at ultra-low accelerations determined by μ.

Consider now setups involving the relativistic nature of the theory. The form of the action of the theory ensures that the right amount of gravitational lensing is ensured whenever Φ correctly determines the dynamics and the tensor mode GW speed is equal to the speed of light in all situations. Meanwhile, on Friedmann-Lemaître-Robertson-Walker (FLRW) $\phi = \bar{\phi}(t)$ while A^μ aligns trivially with the time direction so that $\mathcal{Y} \to 0$ and $\mathcal{Q} \to \dot{\bar{\phi}}$ and the action reduces to that of shift-symmetric K-essence[108] and the low-energy limit of the ghost-condensate theory.[109] Defining then the reduced function $\mathcal{K}(\mathcal{Q}) \equiv -\mathcal{F}(0,\mathcal{Q})/2$, we require that $\mathcal{K} \approx \mathcal{K}_2\left(\mathcal{Q}-\mathcal{Q}_0\right)^2 + \ldots$, with non-zero $\mathcal{Q}_0$ (same constants as what appear above), the energy density of ϕ scales precisely as a^{-3}, $a = 1/(1+z)$ being here the cosmic scale factor, i.e. like pressureless matter, plus small corrections. The equation of state and speed of sound of ϕ also scale as a^{-3} and depend on the constants $\mathcal{K}_2$ and $\mathcal{Q}_0$ of the action. In the late universe limit these tend to zero so that the linear perturbations of ϕ and A_μ can be re-casted by linear transformations into a density contrast δ and velocity divergence θ of pressureless matter. In this limit, the field equations of ϕ and A_μ and their contribution to the Einstein equations, become identical to

pressureless-dust equations, i.e. as for CDM. This ensures that the CMB and MPS spectra calculated within this theory, become parametrically close to those from ΛCDM, leading to excellent fits to the CMB angular PS from *Planck* and the MPS obtained from galaxy clustering.

To summarize, the theory is not a MOND extrapolation to cosmology. Rather, it recovers MOND in quasistatic situations on galactic scales, and it tends to ΛCDM behaviour on cosmological scales. It depends on a free function, which is, however, constrained so that these two limits are attained. Doing that, the theory depends on a minimum of 4 parameters, $K_{\rm B}$, $\mathcal{Q}_0$, $\mathcal{K}_2$ and a_0, while additional parameters could be part of the free function and be related to the GR limit or the early universe.

5.3. *Exploiting the magnification bias with high-z sub-mm galaxies*

Magnification bias is due to gravitational lensing and consists in the apparent excess in the number of high-z sources (the lenses sources) close to the position of low-z galaxies (the lenses). In fact, the gravitational lensing deflects the light rays of high-z sources causing the stretching of the apparent sky area in the region affected by the lensing. This also causes a boost in the flux density of the high-z sources, making them more likely to be detected above a given instrument flux density limit (see, e.g., Ref. 110). The magnification bias can be measured through the cross-correlation between the low and high-z objects. Given two source samples with non-overlapping redshift distributions, the excess signal when computing their cross-correlation with respect to the random case is due to the magnification bias. Being such signal related to lensing and thus to cosmological distances and the galaxy halo characteristics, it can be used to constrain cosmological parameters.

In order to get the best cross-correlation measurements, at the moment the optimal choice is to use as high-z sources (background samples) the high-z sub-mm galaxies (SMGs) first of all because of their steep number counts, shown with recent observations by the *Herschel* Space Observatory[111] and the South Pole Telescope[112] (SPT) strengthening their magnification. Moreover, the SMGs are faint in the optical which means that they do not get confused with the foreground lens sample, being such sample invisible at sub-mm wavelength. Lastly, the SMGs redshift are usually above $z > 1$ so that their selection as background sample ensures no overlap with the foreground objects. In fact, due to their extreme magnification bias, Dunne et al. (Ref. 113) make a serendipitous direct observation of high-z SMGs. While performing a study of gas tracers with Atacama Large Millimeter/submillimeter Array (ALMA) observation of galaxies targeting a statistically complete sample of twelve galaxies selected at 250 μm with $z = 0.35$, magnified SMGs appears around the position of half of them.

As shown in Refs. 114, 115, these advantages make the SMGs the perfect background sample for tracing the (baryonic and DM) mass density profiles of galaxies and GCs and their time-evolution and for constraining the free parameters of a

halo occupation distribution (HOD) model. In the adopted HOD model[116] all halos above a mass M_{min} host a central galaxy and those above a mass M_1 host satellites (galaxies in the halo) whose number is a power-law, $(M/M_1)^{\alpha}$, of the halo mass M. In particular, the results in Ref. 114 suggests that the lenses are massive galaxies or GCs, with a minimum mass of $10^{13} M_{\odot}$. Later on, Bonavera et al. (Ref. 115) study with the magnification bias the mass properties of a sample of quasars whose position signpost the lens at $0.2 < z < 1.0$ obtaining a $M_{min} = 10^{13.6^{+0.9}_{-0.4}} M_{\odot}$. This again suggests that the lensing is actually produced by halos of the size of GCs placed close to the quasar positions.

Moreover, some of the main cosmological parameters can be also estimated using the magnification bias, as in Ref. 117. They use the cross-correlation data measured between a foreground sample of Galaxy And Mass Assembly (GAMA) galaxies with spectroscopic redshifts in the range $0.2 < z < 0.8$ and a background sample of *Herschel* Astrophysical Terahertz Large Area Survey (H-ATLAS) galaxies with photometric redshifts $z > 1.2$ to constrain the HOD astrophysical parameters (M_{min}, M_1, and α) and some of the cosmological ones (Ω_M, σ_8, and H_0). These parameters are estimated through a Markov chain Monte Carlo analysis. They obtain a lower limit at 95% CL on $\Omega_M > 0.24$, a slight trend towards $H_0 > 70$ values and a tentative peak around 0.75 with an upper limit at 95% CL of $\sigma_8 < 1$. Such results are summarized in the $\Omega_M - \sigma_8$ plane in Fig. 7 (*left* panel).

Furthermore, magnification bias trough cross-correlation measurements can be used to constrain the halo mass function. In particular, Ref. 118 successfully tests such possibility according to two common parametrizations, the Sheth & Tormen[119] and the Tinker et al.[120] fits.

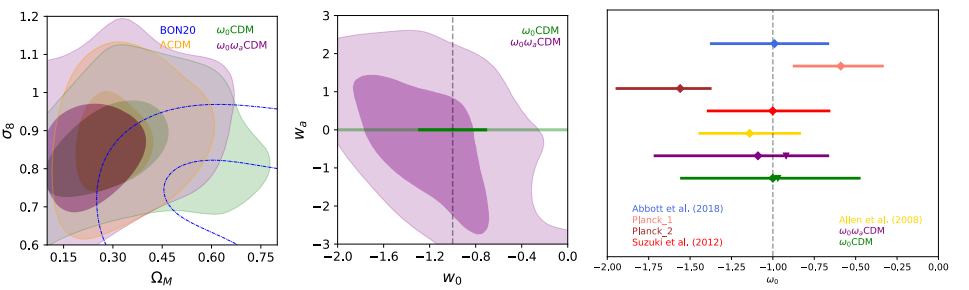


Fig. 7. *Left*: Ω_M and σ_8 contour plot for the non-tomographic case in Ref. 117 (dot-dashed blue line) and the tomographic runs in Ref. 122 for ΛCDM (in yellow), ω_0CDM (in green) and $\omega_0\omega_a$CDM (in purple) models. *Central*: ω_0 and ω_a contour plot for the $\omega_0\omega_a$CDM model. In both panels, the contours of the 2-dimensional posterior distributions are set to 0.393 and 0.865. *Right*: ω_0 results for the ω_0CDM (green) and $\omega_0\omega_a$CDM (purple) models compared with those by DES (blue), *Planck*_1 (base_w_wa_plikHM_TT_lowl_lowE_BAO, salmon), *Planck*_2 (base_w_plikHM_TT_lowl_lowE, brown), Supernovae (red), X-ray measurements (yellow). Adapted from Bonavera et al., A&A, in press, DOI 10.1051/0004-6361/202141521, 2021 (Ref. 122), reproduced with permission from Astronomy & Astrophysics, © ESO.

They find general agreement with traditional values for the involved parameters, with a slight difference in the Sheth & Tormen fit for intermediate and high masses, where the results suggest a hint at a somewhat higher number of halos. Given the large number of studies that can be performed with magnification bias through cross-correlation data, a systematic analysis of possible bias that might affect the estimation is being carried out. In particular, González-Nuevo et al. (Ref. 121) take into account different biases at cosmological scales in the source samples to carefully measure unbiased cross-correlation functions. More precisely, their background sample consists of H-ATLAS galaxies with $z > 1.2$ whereas they use two independent foreground samples with $0.2 < z < 0.8$: GAMA galaxies with spectroscopic redshifts and Sloan Digital Sky Survey (SDSS) galaxies with photometric redshifts. These independent samples allowed them to perform a pseudo-tomographic study that yields to constrain $\Omega_M = 0.50^{+0.14}_{-0.20}$ and $\sigma_8 = 0.75^{+0.07}_{-0.10}$. Such analysis also suggests that a tomographic approach might improve the results.

Driven by the conclusions mentioned above, adopting the unbiased sample by Ref. 121, Bonavera et al.[122] use magnification bias in tomography to jointly constrain the HOD astrophysical parameters in each selected redshift bin together with the Ω_M, σ_8, and H_0 cosmological ones in the ΛCDM scenario and for DE models in the ω_0CDM and $\omega_0\omega_a$CDM frameworks characterized by the ω_0 and ω_a parameters defining the DE barotropic index as $\omega(a) = \omega_0 + \omega_a(1 - a)$. The foreground sample has been divided into four redshift bins (0.1–0.2, 0.2–0.3, 0.3–0.5 and 0.5–0.8) and the sample of H-ATLAS galaxies has photometric redshifts $z > 1.2$. As for the HOD parameters, M_{min} shows a trend towards higher values at higher z confirming the findings in Ref. 114. For the ΛCDM model, they obtain a maximum posterior value [68% CL] of $\Omega_M = 0.26$ [0.17,0.41] and of $\sigma_8 = 0.87$ [0.75,1] (implying $S_8 \simeq 0.81$, in good agreement with the result in Sect. 4.1), being H_0 not yet constrained. For the ω_0CDM model, they find similar results on Ω_M and σ_8 and a maximum posterior value [68% CL] of $\omega_0 = -1$ $[-1.56, -0.47]$. For the $\omega_0\omega_a$CDM model, $\omega_0 = -1.09$ $[-1.72, -0.66]$ and $\omega_a = -0.19$ $[-1.88, 1.48]$. Fig. 7 shows the results in the $\Omega_M - \sigma_8$ plane for the ΛCDM, ω_0CDM and $\omega_0\omega_a$CDM models (*left* panel) and in the $\omega_0 - \omega_a$ plane (*central* panel). The results on ω_0 for the ω_0CDM and $\omega_0\omega_a$CDM cases are shown in Fig. 7 (*right* panel), together with a comparison with other results from literature. Moreover, the tomographic analysis presented in Ref. 122 confirms that magnification bias results do not show the degeneracy found with cosmic shear measurements and that, related to DE, do show a trend of higher ω_0 values for lower H_0 values (see Ref. 123 for details).

In conclusion, SMGs turn out to be a perfect background sample for magnification bias analyses, useful in astrophysical studies and as a further independent cosmological probe for ΛCDM and beyond ΛCDM models and this can be implemented adopting different foreground samples. Up to now ΛCDM compatible results have been found with tomographic and non-tomographic approaches and serendipitous direct observation of magnified SMGs have been performed with ALMA.

Acknowledgments

CB acknowledges partial support from the INAF PRIN SKA/CTA project FORECaST. LB acknowledges the PGC 2018 project PGC2018-101948-B-I00 (MICINN/FEDER). TRC acknowledges support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.02-0700. CS acknowledges support from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement n. 617656 "Theories and Models of the Dark Sector: Dark Matter, Dark Energy and Gravity" and from the European Structural and Investment Funds and the Czech Ministry of Education, Youth and Sports (MSMT) (Project CoGraDS - CZ.02.1.01/0.0/0.0/15003/0000437). RS acknowledges the support of the Natural Sciences and Engineering Research Council of Canada. HT acknowledges support from the ERC-2015-AdG 695561 (ByoPiC project). AW acknowledges funding from the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation of South Africa. Some of the results in this paper have been derived using the `HEALPix`[34] package. The use of the PLA is acknowledged. It is a pleasure to thank N. Aghanim, D. Contreras, M. M. Cueli, M. Douspis, B. Gaensler, J. Gonzalez-Nuevo, R. Mohammadi, D. Scott, S. S. Xue, the MISTRAL Collaboration and the *Planck* Collaboration for long-standing collaborations on the various topics discussed in this paper.

References

1. E. Di Valentino, O. Mena, S. Pan *et al.*, *Classical and Quantum Gravity* **38**, 153001 (2021).
2. Dark Energy Survey Collaboration, *Phys. Rev. D* **98**, 043526 (2018).
3. DES Collaboration, arXiv:2105.13549 (2021).
4. D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic and F. Crawford, *Science* **318**, 777 (2007).
5. S. P. Tendulkar, C. G. Bassa, J. M. Cordes *et al.*, *ApJL* **834**, L7 (2017).
6. A. Walters, A. Weltman, B. M. Gaensler, Y.-Z. Ma and A. Witzemann, *ApJ* **856**, 65 (2018).
7. A. Walters, Y.-Z. Ma, J. Sievers and A. Weltman, *Phys. Rev. D* **100**, 103519 (2019).
8. A. Weltman and A. Walters, arXiv:1905.07132 (2019).
9. E. Platts, A. Weltman, A. Walters *et al.*, *Physics Reports* **821**, 1 (2019).
10. J. P. Macquart, J. X. Prochaska, M. McQuinn *et al.*, *Nature* **581**, 391 (2020).
11. R. Barkana and A. Loeb, *Physics Reports* **349**, 125 (2001).
12. T. R. Choudhury, *Current Science* **97**, 841 (2009).
13. G. Paciga, J. G. Albert, K. Bandura *et al.*, *MNRAS* **433**, 639 (2013).
14. A. H. Patil, S. Yatawatta, L. V. E. Koopmans *et al.*, *ApJ* **838**, 65 (2017).
15. F. G. Mertens, M. Mevius, L. V. E. Koopmans *et al.*, *MNRAS* **493**, 1662 (2020).
16. J. S. Dillon, A. Liu, C. L. Williams *et al.*, *Phys. Rev. D* **89**, 023002 (2014).
17. J. S. Dillon, A. R. Neben, J. N. Hewitt *et al.*, *Phys. Rev. D* **91**, 123011 (2015).
18. A. P. Beardsley, B. J. Hazelton, I. S. Sullivan *et al.*, *ApJ* **833**, 102 (2016).
19. S. Paul, S. K. Sethi, M. F. Morales *et al.*, *ApJ* **833**, 213 (2016).

20. N. Barry, M. Wilensky, C. M. Trott *et al.*, *ApJ* **884**, 1 (2019).
21. W. Li, J. C. Pober, N. Barry *et al.*, *ApJ* **887**, 141 (2019).
22. C. M. Trott, C. H. Jordan, S. Midgley *et al.*, *MNRAS* **493**, 4711 (2020).
23. M. Kolopanis, D. C. Jacobs, C. Cheng *et al.*, *ApJ* **883**, 133 (2019).
24. The HERA Collaboration, arXiv:2108.02263 (2021).
25. T. R. Choudhury and A. Paranjape, *MNRAS* **481**, 3821 (2018).
26. J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen and N. Mahesh, *Nature* **555**, 67 (2018).
27. R. Barkana, *Nature* **555**, 71 (2018).
28. C. Feng and G. Holder, *ApJL* **858**, L17 (2018).
29. A. Fialkov and R. Barkana, *MNRAS* **486**, 1763 (2019).
30. A. Chatterjee, P. Dayal, T. R. Choudhury and R. Schneider, *MNRAS* **496**, 1445 (2020).
31. R. Hills, G. Kulkarni, P. D. Meerburg and E. Puchwein, *Nature* **564**, E32 (2018).
32. J. Raste, G. Kulkarni, L. C. Keating *et al.*, *MNRAS* **507**, 4684 (2021).
33. Planck Collaboration, *A&A* **594**, A10 (2016).
34. K. M. Górski, E. Hivon, A. J. Banday *et al.*, *ApJ* **622**, 759 (2005).
35. R. A. Sunyaev and Y. B. Zel'dovich, *Comments on Astrophysics and Space Physics* **4**, 173 (1972).
36. M. Birkinshaw, *Physics Reports* **310**, 97 (1999).
37. R. Cen and J. P. Ostriker, *ApJ* **514**, 1 (1999).
38. T. Tuominen, J. Nevalainen, E. Tempel *et al.*, *A&A* **646**, A156 (2021).
39. E. S. Battistelli, E. Barbavara and P. de Bernardis *et al.*, in *these proceedings*.
40. A. Paiella, E. S. Battistelli, M. G. Castellano *et al.*, *Journal of Physics: Conference Series* **1182**, 012005 (2019).
41. S. Masi, P. de Bernardis, A. Paiella *et al.*, *JCAP* **07**, 003 (2019).
42. M. S. Petersen, R. A. Gutermuth, E. Nagel, G. W. Wilson and J. Lane, *MNRAS* **488**, 1462 (2019).
43. A. D. Sokol, R. A. Gutermuth, R. Pokhrel *et al.*, *MNRAS* **483**, 407 (2019).
44. W. F. Wall, I. Puerari, R. Tilanus *et al.*, *MNRAS* **459**, 1440 (2016).
45. A. Humphrey, M. Zeballos, I. Aretxaga *et al.*, *MNRAS* **418**, 74 (2011).
46. Planck Collaboration, *A&A* **641**, A6 (2020).
47. Planck Collaboration, *A&A* **571**, A20 (2014).
48. Planck Collaboration, *A&A* **594**, A24 (2016).
49. C. Heymans *et al.*, *A&A* **646**, A140 (2021).
50. Planck Collaboration, *A&A* **594**, A22 (2016).
51. G. Hurier *et al.*, *A&A* **558**, A118 (2013).
52. H. Tanimura, M. Douspis and N. Aghanim, in *these proceedings*.
53. J. Delabrouille *et al.*, *A&A* **553**, A96 (2013).
54. A. Maniyar *et al.*, *A&A* **645**, A40 (2021).
55. Planck Collaboration, *A&A* **641**, A1 (2020).
56. Planck Collaboration, *A&A* **571**, A27 (2014).
57. Planck Collaboration, *A&A* **644**, 100 (2020).
58. R. M. Sullivan and D. Scott, in *these proceedings*.
59. T. Trombetti, C. Burigana and F. Chierici, *A&A* **646**, A75 (2021).
60. V. F. Mukhanov and G. V. Chibisov, *Sov. Journal of Exp. and Theor. Phys. Lett.* **33**, 532 (1981).
61. V. F. Mukhanov and G. V. Chibisov, *Sov. Journal of Exp. and Theor. Phys.* **56**, 258 (1982).
62. S. W. Hawking, *Physics Letters B* **115**, 295 (1982).

63. A. H. Guth and S. Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982).
64. A. A. Starobinsky, *Physics Letters B* **117**, 175 (1982).
65. J. M. Bardeen, P. J. Steinhardt and M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
66. V. F. Mukhanov, *Sov. Journal of Exp. and Theor. Phys. Lett.* **41**, 493 (1985).
67. U. Seljak, *ApJ* **482**, 6 (1997).
68. M. Kamionkowski, A. Kosowsky and A. Stebbins, *Phys. Rev. Lett.* **78**, 2058 (1997).
69. U. Seljak and M. Zaldarriaga, *Phys. Rev. Lett.* **78**, 2054 (1997).
70. M. Zaldarriaga and U. Seljak, *Phys. Rev. D* **58**, 023003 (1998).
71. Planck Collaboration, *A&A* **586**, A133 (2016).
72. BICEP2 Collaboration, Keck Array Collaboration, *Phys. Rev. Lett.* **121**, 221301 (2018).
73. Planck Collaboration, *A&A* **641**, A10 (2020).
74. BICEP2/Keck Collaboration, Planck Collaboration, *Phys. Rev. Lett.* **114**, 101301 (2015).
75. M. Tristram, A. J. Banday, K. M. Górski *et al.*, *A&A* **647**, A128 (2021).
76. Planck Collaboration, *A&A* **643**, A42 (2020).
77. H. K. Eriksen, J. B. Jewell, C. Dickinson *et al.*, *ApJ* **676**, 10 (2008).
78. Planck Collaboration, *A&A* **596**, A108 (2016).
79. A. Kosowsky, *Annals Phys.* **246**, 49 (1996).
80. U. Seljak and M. Zaldarriaga, *ApJ* **469**, 437 (1996).
81. M. Zaldarriaga and U. Seljak, *Phys. Rev. D* **55**, 1830 (1997).
82. A. Cooray, A. Melchiorri and J. Silk, *Phys. Lett. B* **554**, 1 (2003).
83. E. Bavarsad, M. Haghighat, Z. Rezaei *et al.*, *Phys. Rev. D* **81**, 084035 (2010).
84. S. Modares Vamegh, M. Haghighat, S. Mahmoudi and R. Mohammadi, *Phys. Rev. D* **100**, 103024 (2019).
85. M. Haghighat, S. Mahmoudi, R. Mohammadi, S. Tizchang and S. S. Xue, *Phys. Rev. D* **101**, 123016 (2020).
86. S. Tizchang, S. Batebi, M. Haghighat and R. Mohammadi, in *proceedings of The 39th Intern. Conf. on High Energy Physics (ICHEP2018)*, p.142, Proceedings of Science (PoS) (2019).
87. S. Tizchang, R. Mohammadi and S. S. Xue, in *these proceedings*.
88. D. Colladay and V. A. Kostelecky, *Phys. Rev. D* **55**, 6760 (1997).
89. D. Colladay and V. A. Kostelecky, *Phys. Rev. D* **58**, 116002 (1998).
90. V. A. Kostelecky and C. D. Lane, *Phys. Rev. D* **60**, 116010 (1999).
91. V. A. Kostelecky and N. Russell, in *proceedings of the Fourth Meeting on CPT and Lorentz Symmetry*, p. 308, World Scientific (2008).
92. Planck Collaboration, *A&A* **594**, A20 (2016).
93. G. Bertone, D. Hooper and J. Silk, *Physics Reports* **405**, 279 (2005).
94. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, *Physics Reports* **513**, 1 (2012).
95. D. Lovelock, *J. Math. Phys.* **12**, 498 (1971).
96. M. Milgrom, *ApJ* **270**, 365 (1983).
97. M. Milgrom, *ApJ* **270**, 371 (1983).
98. M. Milgrom, *ApJ* **270**, 384 (1983).
99. J. Bekenstein and M. Milgrom, *ApJ* **286**, 7 (1984).
100. B. Famaey and S. McGaugh, *Living Reviews in Relativity* **15**, 10 (2012).
101. R. H. Sanders, *ApJ* **480**, 492 (1997).
102. J. D. Bekenstein, *Phys. Rev. D* **70**, 083509 (2004), [Erratum: *Phys. Rev. D* **71**, 069901 (2005)].
103. C. Skordis, D. F. Mota, P. G. Ferreira and C. Boehm, *Phys. Rev. Lett.* **96**, 011301 (2006).

104. S. Dodelson and M. Liguori, *Phys. Rev. Lett.* **97**, 231301 (2006).
105. C. Skordis and T. Zlosnik, *Phys. Rev. D* **100**, 104013 (2019).
106. T. G. Zlosnik, P. G. Ferreira and G. D. Starkman, *Phys. Rev. D* **74**, 044037 (2006).
107. C. Skordis and T. Zlosnik, *Phys. Rev. Lett.* **127**, 161302 (2021).
108. R. J. Scherrer, *Phys. Rev. Lett.* **93**, 011301 (2004).
109. N. Arkani-Hamed, H.-C. Cheng, M. A. Luty and S. Mukohyama, *JHEP* **05**, 074 (2004).
110. P. Schneider, J. Ehlers and E. E. Falco, *Gravitational Lenses*, Springer-Verlag Berlin Heidelberg (1992).
111. G. L. Pilbratt, J. R. Riedinger, T. Passvogel *et al.*, *A&A* **518**, L1 (2010).
112. J. E. Carlstrom, P. A. R. Ade, K. A. Aird *et al.*, *PASP* **123**, 568 (2011).
113. L. Dunne, L. Bonavera, J. Gonzalez-Nuevo, S. J. Maddox and C. Vlahakis, *MNRAS* **498**, 4635 (2020).
114. J. González-Nuevo, A. Lapi, L. Bonavera *et al.*, *JCAP* **10**, 024 (2017).
115. L. Bonavera, J. González-Nuevo, S. L. Suárez Gómez *et al.*, *JCAP* **09**, 021 (2019).
116. Z. Zheng, A. A. Berlind, D. H. Weinberg *et al.*, *ApJ* **633**, 791 (2005).
117. L. Bonavera, J. González-Nuevo, M. M. Cueli *et al.*, *A&A* **639**, A128 (2020).
118. M. M. Cueli, L. Bonavera, J. González-Nuevo and A. Lapi, *A&A* **645**, A126 (2021).
119. R. K. Sheth and G. Tormen, *MNRAS* **308**, 119 (1999).
120. J. Tinker, A. V. Kravtsov, A. Klypin *et al.*, *ApJ* **688**, 709 (2008).
121. J. González-Nuevo, M. M. Cueli, L. Bonavera *et al.*, *A&A* **646**, A152 (2021).
122. L. Bonavera, M. M. M. Cueli M., J. González-Nuevo J. *et al.*, *A&A*, in press, DOI 10.1051/0004-6361/202141521, arXiv:2109.12413 (2021).
123. L. Bonavera, M. M. Cueli and J. González-Nuevo, in *these proceedings*.

CMB μT cross-correlations as a probe of PBH scenarios

Ogan Özsoy

CEICO, Institute of Physics of the Czech Academy of Sciences, Prague, Na Slovance 1999/2, 182 21, Czechia
E-mail: ozsoy@fzu.cz

Gianmassimo Tasinato

Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom
E-mail: g.tasinato@swansea.ac.uk

We initiate a new method for probing inflationary models that can produce primordial black hole populations, using only CMB physics at relatively large scales. In these scenarios, profile of the primordial scalar power spectrum exhibit a universal dip feature that is followed by a rapid growth towards small scales, leading to a peak responsible for PBH formation. Focusing on scales around the dip that are well separated from the peak, we first analytically compute expressions for the curvature bispectrum. We then show that the amplitude of the bispectrum is enhanced for the squeezed configuration around the position of the dip, and it acquires a characteristic scale dependence that can be probed by cross correlations between CMB μ-distortions and temperature anisotropies. We quantitatively study the properties of such cross-correlations and how they depend on the underlying model, discussing how they can be tested by the next generation of CMB μ-distortion experiments. This method allows one to experimentally probe inflationary PBH scenarios using well-understood CMB physics, without considering non-linearities associated with PBH formation and evolution.

Keywords: Inflation, Primordial Black Holes, CMB distortions

1. Introduction

Gravitational waves from merging black holes, in conjunction with other probes, can be used for testing our understanding of cosmology. There is the intriguing possibility that part of black holes from merging events have primordial origin, arising from the collapse of overdense regions in the early stages of our universe evolution.[1–3] The existence of such overdense regions might be attributed to the dynamics of inflationary cosmology. In this framework, it is well known that to produce primordial black holes (PBHs) (see *e.g.* reviews[4–7]) with astrophysically relevant masses, the primordial scalar power spectrum should increase by several orders of magnitude with respect to its values at large CMB scales. Within the context of single-field inflation, such an amplification can be realized during a short slow-roll violating non-attractor period during which the would-be decaying mode grows to influence evolution of super-horizon fluctuations. In this set-up, a peak in the curvature perturbation spectrum can be produced, and the non-linear process

of PBH production and subsequent evolution depends on the details of the peak, as first explicitly discussed in.[8]

The possible existence of primordial black hole (PBH) populations in our universe demands a detailed understanding of inflationary mechanisms that can give rise to them. Intriguingly, in inflationary scenarios capable of producing PBHs, the profile of the spectrum shows universal features such as the existence of pronounced dip, followed by a rapid growth with a well defined slope[9] towards its peak. The position of the dip is not random, but depends on other global properties of the spectrum, as well as on details of the non-attractor inflationary evolution.[10] It typically occurs at scales much larger than the peak, $k_{\rm dip}/k_{\rm peak} \simeq 10^{-2}$. Since the occurrence of dip seems to be a rather universal feature of power spectrum in single-field models of PBH formation, probing the physics associated with it offers an indirect way of testing these models, independently of the details of non-linear PBH formation and evolution mechanisms.

In this work, we ask whether we can probe the PBH formation mechanism by identifying large scale observables away from the scales associated with the peak of the spectrum.[11] For this purpose, we focus on scales around the dip feature to compute bispectrum of curvature perturbation analytically using the gradient expansion formalism[12,13] and show that it is generically amplified around the dip scale for squeezed configurations, exhibiting a characteristic scale dependence. We then show the influence of such a scale dependent squeezed bispectrum on the properties of cross-correlations between μ distortions and temperature T anisotropies and discuss the prospects of detectability of the $\langle \mu T \rangle$ signal with future spectral distortion experiments.

2. Curvature perturbation at super-horizon scales

Models of primordial black hole (PBH) formation based on single field inflation often include a short phase of non-attractor evolution, characterized by a transient growth of would be the decaying mode which influences the super-horizon evolution of cosmological scalar perturbations[a]. In this section, we review and utilize the *gradient expansion formalism*[b] – first introduced in[12] – to study the statistics of curvature fluctuations in scenarios including a transient non-attractor phase. In this framework, iterative solutions of the comoving curvature perturbation are generated at any desired order in a small-k expansion, in terms of analytic functions controlled by the background evolution. For example, the amplitude of curvature fluctuations at a fiducial late time $\tau = \tau_*$ (*e.g.* at the reheating surface) is mapped to its value at an initial time around horizon exit $\tau = \tau_k$ in terms of a complex-valued k-dependent

[a]Direct enhancement of tensor perturbations on super-horizon scales can be also achieved by devising an analogue non-attractor phase within the generalized scalar-tensor theories of single field inflation.[14,15]

[b]For a detailed account on the spectral profile of the scalar power spectrum that leads to PBH formation and its applications in the context of induced gravitational waves, we refer the reader to[13] where further developments of the gradient expansion formalism is discussed.

coefficient:

$$\mathcal{R}_k(\tau_*) = \alpha_k \mathcal{R}_k(\tau_k) = \left(\alpha_k^R + i\alpha_k^I\right) \mathcal{R}_k(\tau_k)\,, \tag{1}$$

where in the last equality we split the enhancement factor into its real and imaginary parts:

$$\alpha_k \equiv \alpha_k^R + i\alpha_k^I\,. \tag{2}$$

Once expanded up to second, k^2-order in the gradient expansion, the real and imaginary parts of α_k are given by

$$\alpha_k^R = 1 + D(\tau_k)\, v_{\mathcal{R}}^R - F(\tau_k)\, k^2, \tag{3}$$

$$\alpha_k^I = D(\tau_k)\, v_{\mathcal{R}}^I\,. \tag{4}$$

The quantity $v_{\mathcal{R}}^R$ and $v_{\mathcal{R}}^I$ denote the real and imaginary part of the k-dependent fractional velocity of the curvature perturbation, which is defined as

$$v_{\mathcal{R}}(\tau_k) = \left.\frac{\mathcal{R}_k'}{3\mathcal{H}_k\mathcal{R}_k}\right|_{\tau=\tau_k}\,. \tag{5}$$

The full k dependence of the expressions (3) and (4) on super-horizon scales is encoded in the quantities $v_{\mathcal{R}}^R$, $v_{\mathcal{R}}^I$ and in the functions $D(\tau_k), F(\tau_k)$ (see Appendix A and B), given by the following nested integrals of the pump field appearing in Eq. (15)[12, 13]:

$$D(\tau) = 3\mathcal{H}_k \int_{\tau}^{\tau_*} \mathrm{d}\tau' \, \frac{z^2(\tau_k)}{z^2(\tau')}\,, \tag{6}$$

$$F(\tau) = \int_{\tau}^{\tau_*} \frac{d\tau'}{z^2(\tau')} \int_{\tau_k}^{\tau'} \mathrm{d}\tau'' z^2(\tau'')\,. \tag{7}$$

Expressions (6) and (7) indicate that if the pump field increases with time – as in standard slow-roll inflation, where $z \propto a(\tau)$ – the functions D, F rapidly decrease to zero after horizon crossing (*i.e.* $\alpha_k \to 1$). Hence the curvature perturbation in (1) settles to a constant shortly after horizon exit ($\mathcal{R}_k(\tau_*) \simeq \mathcal{R}_k(\tau_k)$). On the contrary, in inflationary models containing phases of non-attractor evolution, $z(\tau)$ transiently decreases and the functions D, F can grow, amplifying the spectrum of curvature perturbation with $|\alpha_k| \gg 1$ at super-horizon scales: see Appendix B.

The curvature perturbation power spectrum. We define the late-time power spectrum evaluated at $\tau = \tau_*$ as

$$\langle \mathcal{R}_k(\tau_*) \mathcal{R}_{k'}(\tau_*) \rangle = (2\pi)^3\, P_{\mathcal{R}}(\tau_*, k)\, \delta\left(\vec{k} + \vec{k}'\right)\,. \tag{8}$$

Using Eq. (1), we can then relate the power spectrum at late times to the power spectrum evaluated at horizon crossing via

$$P_{\mathcal{R}}(\tau_*, k) \equiv \frac{2\pi^2}{k^3}\mathcal{P}_{\mathcal{R}}(\tau_*, k) = \frac{2\pi^2}{k^3}\left[\, |\alpha_k|^2\, \mathcal{P}_{\mathcal{R}}(\tau_k)\,\right] \equiv |\alpha_k|^2 P_{\mathcal{R}}(\tau_k), \tag{9}$$

where $\mathcal{P}_{\mathcal{R}}(\tau_k) \equiv k^3 P_{\mathcal{R}}(\tau_k, k)/2\pi^2$, and $|\alpha_k|^2 = (\alpha_k^R)^2 + (\alpha_k^I)^2$.

A non-linear expression for $\mathcal{R}$. We assume that $\mathcal{R}_k(\tau_k)$ is a Gaussian random variable: nevertheless, the super-horizon evolution typically introduces non-linearities. In fact, we can go beyond the linear theory used for Eq. (1) to compute the bispectrum of the late time curvature perturbation $\mathcal{R}_k(\tau_*)$. For the purpose of deriving an analytic expression for the bispectrum, we will use the following non-linear expression for the curvature perturbation, first derived in[16]

$$\mathcal{R}_k(\tau_*) = \alpha_k \mathcal{R}_k(\tau_k) + \frac{F(\tau_k)}{2}\left\{\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\left[4k'^2 - \vec{k}'.(\vec{k}-\vec{k}')\right]\mathcal{R}_{k'}(\tau_{k'})\mathcal{R}_{|\vec{k}-\vec{k}'|}(\tau_{|\vec{k}-\vec{k}'|})\right\}. \tag{10}$$

where the last term represents the non-linear contribution parametrized by the convolution of the Gaussian variable $\mathcal{R}_k(\tau_k)$.

Bispectrum. We define the late-time bispectrum of curvature perturbations as

$$\langle \mathcal{R}_{k_1}(\tau_*)\mathcal{R}_{k_2}(\tau_*)\mathcal{R}_{k_3}(\tau_*)\rangle = (2\pi)^3 B_{\mathcal{R}}\left(k_1,k_2,k_3\right)\delta\left(\vec{k}_1+\vec{k}_2+\vec{k}_3\right), \tag{11}$$

and using (10), at leading order in the non-linear term, the bispectrum is given by[16,17]

$$B_{\mathcal{R}} = \frac{(2\pi^2)^2}{2(k_1k_2k_3)^3}\Big[\mathrm{Re}[\alpha^*_{k_1}\alpha_{k_2}]F(\tau_{k_3})\left\{5\left(k_1^2+k_2^2\right)-k_3^2\right\}k_3^3\,\mathcal{P}_{\mathcal{R}}(\tau_{k_1})\mathcal{P}_{\mathcal{R}}(\tau_{k_2}) + \text{ perms }\Big], \tag{12}$$

where the permutations are cyclic among the three external momenta $\{\vec{k}_1,\vec{k}_2,\vec{k}_3\}$. Correspondingly, we define the scale-dependent non-linearity parameter $f_{\rm NL}$ as

$$f_{\rm NL}\left(k_1,k_2,k_3\right) = \frac{5}{6}\frac{B_{\mathcal{R}}\left(k_1,k_2,k_3\right)}{\left[P_{\mathcal{R}}\left(\tau_*,k_1\right)P_{\mathcal{R}}\left(\tau_*,k_2\right)+\text{ perms }\right]}, \tag{13}$$

Using (9) and (12) in (13), we then obtain the scale-dependent $f_{\rm NL}$ as

$$f_{\rm NL} = \frac{5}{12}\frac{\left(\mathrm{Re}[\alpha^*_{k_1}\alpha_{k_2}]F(\tau_{k_3})\left\{5\left(k_1^2+k_2^2\right)-k_3^2\right\}k_3^3\,\mathcal{P}_{\mathcal{R}}(\tau_{k_1})\mathcal{P}_{\mathcal{R}}(\tau_{k_2})+\text{ perms }\right)}{\left[\left|\alpha_{k_1}\alpha_{k_2}\right|^2\mathcal{P}_{\mathcal{R}}\left(\tau_{k_1}\right)\mathcal{P}_{\mathcal{R}}\left(\tau_{k_2}\right)k_3^3+\text{ perms }\right]}. \tag{14}$$

Eq. (14) inform us that the size and the scale-dependence of the $f_{\rm NL}$ parameter depend on the function $F(\tau_k)$ and the enhancement factor α_k, whose behavior is tightly related with the background dynamics during inflation. We now turn to study the scale dependence of the power spectrum and bispectrum on a representative set-up capable of producing a large PBH population during inflation.

2.1. *Characteristics of the scalar power and bispectrum*

In order to study the spectral shape and enhancement in the power spectrum and bispectrum, we consider a representative scenario that instantly connects at $\tau = \tau_0$

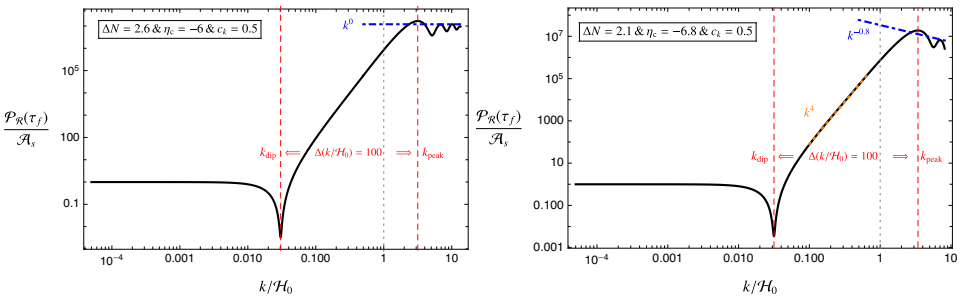


Fig. 1. Scale dependence of the power spectrum for an example ultra slow-roll model ($\eta_{\rm c} = -6$) (Left) and constant-roll model ($\eta_{\rm c} = -6.8$)(Right) as function of $k/\mathcal{H}_0$. The vertical gray-dotted lines separate the interval of wave-numbers leaving the horizon during the initial slow-roll era (left side) and the non-attractor era (right side). We define $\mathcal{A}_s = H^2/(8\pi^2\epsilon_{\rm sr}M_{\rm pl}^2)$.

an initial slow-roll era, with $\eta_{\rm sr} = 0$, to a slow-roll violating, non-attractor phase with constant $\eta_{\rm c} \leq -6$ lasting until $\tau = \tau_f$. Here η denotes the second slow-roll parameter, $H\,\eta \equiv d\ln\epsilon/dt$ and the first slow-roll parameter is given by $H\,\epsilon = -d\ln H/dt$. The pump field $z(\tau)$ appearing in the eqs. (6) and (7) is assumed to have a profile (we take $\tau < 0$):

$$z(\tau) = \begin{cases} z_0\,(\tau/\tau_0)^{-1} & \tau/\tau_0 \geq 1\,, \\ z_0\,(\tau/\tau_0)^{-(\eta_{\rm c}+2)/2} & \tau_f/\tau_0 \leq \tau/\tau_0 \leq 1\,, \end{cases} \tag{15}$$

describing collectively the initial slow-roll and the slow-roll violating phases, the latter being controlled by the negative $\mathcal{O}(1)$ "slow-roll" parameter η_c. The duration of the non-attractor phase can be parametrized as $\Delta N = \ln(\tau_0/\tau_f)$ in e-fold numbers. We relate the quantity z_0 with a constant slow-roll parameter $\epsilon_{\rm sr}$ via $z_0 = -a(\tau_0)\sqrt{2\epsilon_{\rm sr}}M_{\rm pl}$. For simplicity we parameterize the scale factor as in de Sitter space: $a = -1/(H\tau)$ with a constant Hubble rate H during inflation. We also indicate with $\mathcal{H}_0$ the size of the comoving horizon at the time of the transition to the non-attractor era.

Spectral profile of the power spectrum: We proceed with determining the corresponding growth rate of the power spectrum. For this purpose, we first re-write the dimensionless power spectrum as (see *e.g.* Eq. (9)),

$$\mathcal{P}_{\mathcal{R}}(\tau_f, k) \equiv |\alpha_k|^2\mathcal{P}_{\mathcal{R}}(\tau_k) = \left[(\alpha_k^R)^2 + (\alpha_k^I)^2\right]\mathcal{P}_{\mathcal{R}}(\tau_k)\,, \tag{16}$$

and we evaluate this quantity at the end of the non-attractor era $\tau_* = \tau_f$, to characterize the shape of the power spectrum using α_k^R (3), α_k^I (4) and $\mathcal{P}_{\mathcal{R}}(\tau_k)$ with the help of formulas we derived in Appendix A and B. In this way, we represent our results in Figure 1 for two different sets of parameters associated with the two-phase model of PBH formation characterized by the pump field profile of (15). For understanding the physical implications of our findings within the gradient expansion formalism, it is convenient to introduce a fixed quantity

$$c_k \equiv -k\tau_k \leq 1\,, \tag{17}$$

which determines the size of a mode k with respect to the horizon $(aH)^{-1}$ at time $\tau = \tau_k$, corresponding to the horizon crossing epoch. We stress that by virtue of the relation (17) and the super-horizon gradient formalism, all modes we consider in Figure 1 (and in general in this work) are outside the horizon at the initial time τ_k. We then distinguish modes whose momenta lie in the following ranges:

i) modes that become super-horizon during the initial slow-roll era, *i.e.* modes satisfying $\tau_k/\tau_0 > 1$ or equivalently $k/\mathcal{H}_0 < c_k \leq 1$, and
ii) modes that leave the horizon during the non-attractor $\eta_c \leq -6$ phase, $c_k < k/\mathcal{H}_0$.

Focusing on these regimes separately, we discuss below the spectral behavior of $\mathcal{P}_\mathcal{R}$ and some of its global features.

The behaviour of late time $\mathcal{P}_\mathcal{R}$ shown in Figure 1 reflects the accuracy of the gradient expansion formalism in capturing characteristic features of the power spectrum in inflationary backgrounds that transiently violates slow-roll conditions which is required to generate PBHs.[18] In particular, we notice a pronounced dip feature occurring at relatively large scales, associated with modes that still leave the horizon in the initial slow-roll era (*i.e.* $k_{\rm dip} \ll \mathcal{H}_0$) far away from the peak, which occurs at $k_{\rm peak} \simeq 3\mathcal{H}_0$.[19] The presence of such a pronounced dip in the spectrum is a universal feature, being virtually present in all single field models based on non-attractor evolution that are aiming to generate a sizeable peak in the power spectrum for producing PBH, say of order 10^7 (see e.g.[20–29]).

The dip and its properties. It is worth mentioning that such dip feature arise due to competing contributions that appear in the power spectrum that are weighted by opposite signs.[35] The properties of the dip feature turn out to be related with global features of the spectrum profile. As shown in a recent work,[10] the location of the dip position $k_{\rm dip}$ in momentum space can be related with a characteristic pivot scale $k_\star$ associated with the duration of non-attractor evolution by the expression $k_{\rm dip}/k_\star \simeq \mathcal{O}(1) \times (\Delta\mathcal{P}_\mathcal{R}/\mathcal{P}_\mathcal{R})^{-1/4}$. For the scenarios we consider with a pronounced peak in the power spectrum, we expect $k_\star$ to be related with $k_{\rm peak}$, and $(\Delta\mathcal{P}_\mathcal{R}/\mathcal{P}_\mathcal{R})^{-1/4} \simeq \left(10^7\right)^{-1/4} \simeq 10^{-2}$. A close examination of Figure 1 confirms these arguments and shows that

$$k_{\rm dip} \simeq 10^{-2}\, k_{\rm peak} \tag{18}$$

is a robust relation, valid in all single field inflationary scenarios that can produce a pronounced peak in the power spectrum. In light of these considerations, we assume (18) for the rest of this work.

Spectral profile of the bispectrum: Let us concentrate on modes that exit during the initial slow-roll era, $k/\mathcal{H}_0 < c_k$, to investigate the scale-dependence of the bispectrum. We expect the bispectrum to have features and be amplified around the scale $k_{\rm dip}$ corresponding to the position of the dip in the power spectrum: in fact, non-linearities are usually enhanced at the location of rapid changes in the

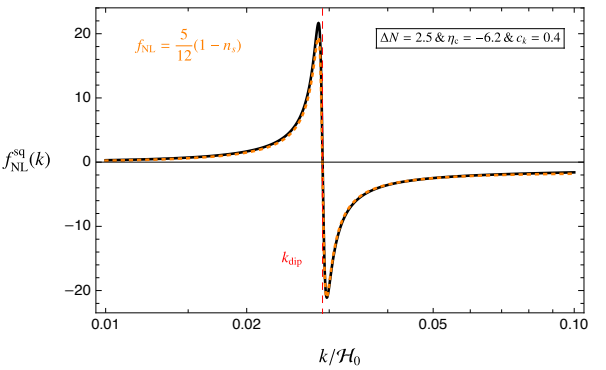


Fig. 2. Scale dependence of the non-linearity parameter $f_{\rm NL}$ around $k_{\rm dip}$ for the squeezed configuration for a transient constant-roll model that can generate a 10^7 enhancement in the power spectrum: $\{\Delta N = 2.5, \eta_c = -6.2, c_k = 0.4\}$. In the right panel, we represent the accuracy of the consistency relation $f_{\rm NL} = 5(1 - n_s)/12$ in capturing the behavior of the $f_{\rm NL}$ in the squeezed limit, using (16) and $n_s - 1 \equiv {\rm d}\ln\mathcal{P}_{\mathcal{R}}(\tau_f, k)/{\rm d}\ln k$.

power spectrum.[30] In what follows, we confirm this expectation for the system under consideration.

We now analyze the squeezed limit of the scale dependent non-linearity parameter[c]. For scales that exit the horizon during the initial slow-roll era, noting the scale invariance of $\mathcal{P}_{\mathcal{R}}(\tau_k)$ factors, we take the squeezed limit $\vec{k}_3 \to 0$ ($k_1 \simeq k_2 \equiv q$) of (14) which yields as[11,16]

$$f_{\rm NL}(q, q, k_3 \to 0) \equiv f_{\rm NL}^{\rm sq} = \frac{5}{12}\left(\frac{4\ {\rm Re}[\alpha_q^* \alpha_{k_3}]\, F(\tau_{k_q}) q^2}{|\alpha_q \alpha_{k_3}|^2}\right). \tag{19}$$

Focusing on a representative choice of parameters that can generate an enhancement of order 10^7 in the power spectrum, in Figure 2 we plot the profile of $f_{\rm NL}$ in the squeezed configuration $k_3 \ll k_1 = k_2$ around the dip scale $k_{\rm dip}$ using (19) and with the help of formulas derived in Appendix B. We observe that the squeezed limit $f_{\rm NL}$ initially grows positive, to then rapidly decrease to negative values around $k_{\rm dip}$, and to finally increase again from negative values, almost in a symmetrical fashion. Intriguingly, as can be seen from the overlap of orange dotted curve with the black solid one in the right panel of Figure 2, the scale dependence of $f_{\rm NL}$ in the squeezed limit agrees very well with the one inferred from Maldacena's consistency condition: $f_{\rm NL} = 5(1-n_s)/12$. The typical magnitude for the squeezed limit of $|f_{\rm NL}|$ around $k_{\rm dip}$ results of order $\mathcal{O}(10)$, compatible with the large value for the spectral tilt around the dip region.[13]

It is worth stressing that the findings we present here agree well with previous literature[34–36] that numerically studied non-Gaussianity in inflationary setups that

[c]A scale-dependent $f_{\rm NL}$ arises in a variety of other inflationary contexts, see e.g. the early works.[31–33]

include a transient non-attractor era. However, the main advantage of our formulas (16) and (19) is the analytic control they provide us to study features in the power and bispectrum in a model independent way without restoring to numerical techniques. In particular, these formulas are flexible enough to ensure an accurate description of the power and bispectrum in various inflationary backgrounds that exhibit a transient, slow-roll violating phase(s) characterized by two essential parameters: the duration ΔN of the non-attractor phase and the value of the slow-roll parameter $\eta_c < 0$ in this phase.

3. μT correlations as a probe of the PBH generation mechanism

In the previous section we studied scale dependence of the squeezed bispectrum around the scales corresponding to the dip scale of the power spectrum. If the dip feature occurs at relatively large scales, say $10\,\mathrm{Mpc}^{-1} \leq k_{\rm dip} \leq 10^4\,\mathrm{Mpc}^{-1}$, the properties of the resulting curvature spectrum can be probed through CMB $\mu-$distortions, using well controlled CMB physics in the linear regime.[37–39] Building on the ideas first introduced by Pajer&Zaldarriaga,[40] this possibility was suggested in our recent work.[11] We stress that the range of scales we identified for $k_{\rm dip}$ above is well motivated for PBH populations with astrophysical masses. For example, given that $k_{\rm dip}/k_{\rm peak} \simeq 10^{-2}$ with a dip scale located around $k_{\rm dip} \simeq 10^3 - 10^4\,\mathrm{Mpc}^{-1}$ corresponds to a peak scale $k_{\rm peak} \simeq 10^5 - 10^6\,\mathrm{Mpc}^{-1}$ that correspond to the formation of PBH in the mass range[6,41] $M_{\rm pbh} \simeq 1 - 100\,M_\odot$.

We now discuss how to use the cross-correlation among CMB spectral μ distortions and temperature anisotropies for probing the statistics of curvature perturbation in inflationary scenarios capable of generating PBHs. The method we propose allows one to indirectly probe the inflationary PBH formation mechanism at relatively large scales, away from the scales where the curvature spectrum grow and has a peak. For a detailed discussion on the formulas we will introduce below, we refer the reader to the works.[11,40,42,43]

We begin by relating the initial curvature perturbation $\mathcal{R}_k$ to the spherical harmonic coefficients of the CMB temperature anisotropies $\Theta(\hat{n}) = \sum_{lm} a^T_{lm}\, Y_{lm}(\hat{n})$ and μ distortion anisotropies $\mu(\hat{n}) = \sum_{lm} a^\mu_{lm}\, Y_{lm}(\hat{n})$ which are are given by[40,42]

$$a^T_{lm} = \frac{12\pi}{5}(-i)^l \int \frac{\mathrm{d}^3k}{(2\pi)^3}\, \mathcal{R}_k\, \Delta_l(k)\, Y^*_{lm}(\hat{k}), \tag{20}$$

$$a^\mu_{lm} \simeq 18.4\pi(-i)^l \int \frac{\mathrm{d}^3k_1\, \mathrm{d}^3k_2}{(2\pi)^6} Y^*_{lm}\left(\hat{k}_+\right)\, \mathcal{R}_{k_1}\, \mathcal{R}_{k_2}\, W\left(\frac{k_+}{k_s}\right) j_l\left(k_+\chi_*\right) \times \left\langle \cos\left(c_s k_1 \tau\right) \cos\left(c_s k_2 \tau\right)\right\rangle_p \left[e^{-\left(k_1^2+k_2^2\right)/k_D^2}\right]^{z_i}_{z_f}, \tag{21}$$

where p denotes time averaging over the period acoustic oscillations; $\vec{k}_+ \equiv \vec{k}_1 + \vec{k}_2$; c_s is the sound speed of the radiation perturbations; $\chi_* = \tau_0 - \tau_* \simeq 14\,\mathrm{Gpc}$ is the comoving distance between the last scattering surface and today; and

$\Delta_l(k)$ is the transfer function during radiation domination. Furthermore, in (21), $W(k) = 3k^{-3}[\sin(k) - k\cos(k)]$ is a top-hat filter function in Fourier space that smears the dissipated energy over a volume of radius $k_s^{-1} \gtrsim k_D(z_f)^{-1}$ where k_D denotes the diffusion damping scale during a radiation dominated universe. It depends on redshift as $k_D(z) \simeq [(1+z)/10^5]^{3/2}\ 130\,\mathrm{Mpc}^{-1}$ where the range of redshifts associated with the μ distortions are given by

$$z_f \equiv 5 \times 10^4 < z < 2 \times 10^6 \equiv z_i. \tag{22}$$

We then define angular correlators involving anisotropies labeled by $\{i, j\}$ as

$$\left\langle \left(a^i_{lm}\right)^* a^j_{l'm'} \right\rangle = \delta_{ll'}\delta_{mm'} C^{ij}_l\,, \qquad i\,=\,\mu, T. \tag{23}$$

In the following discussion, we make use of the definition (23) to study angular correlations $\langle \mu T\rangle$ and to relate them to scale dependent bispectrum present around the dip scale of the PBH forming inflationary scenarios.

Using the definition (11) of the bispectrum together with (20) and (21), the cross correlations $C_l^{\mu T}$ for a squeezed bispectrum ($\vec{k}_+ = \vec{k}_1 + \vec{k}_2 \to 0$) can be expressed as,[40,42]

$$C_l^{\mu T} \simeq \frac{27.6}{20\pi^3} \int \mathrm{d}\ln k_+\, \Delta_l(k_+)\, j_l\left(k_+\chi_*\right) \times \int \mathrm{d}\ln q \left[k_+^3 q^3 B_{\mathcal{R}}\left(q, q, k_+ \to 0\right)\right] \left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}, \tag{24}$$

where we take the filter function $W \to 1$ in the squeezed limit $k_+ \ll k_s \simeq k_D(z_f)$, and relabel $|\vec{k}_1| \simeq |-\vec{k}_2| \equiv q$. Using the general definition (13) of the non-linearity parameter, the squeezed limit bispectrum inside the $\mathrm{d}\ln q$ integral in (24) can be expressed as

$$B_{\mathcal{R}}\left(q, q, k_+ \to 0\right) \simeq \frac{12}{5} \underbrace{f_{\mathrm{NL}}^{\mathrm{sq}}(q)|\alpha_q|^2|\alpha_{k_+}|^2}_{\equiv f_{\mathrm{NL}}^{\mathrm{eff}}(q)}\ P_{\mathcal{R}}(\tau_q) P_{\mathcal{R}}(\tau_{k_+}). \tag{25}$$

In (25), we defined the bispectrum in terms of the two copies of the power spectrum evaluated at around horizon crossing τ_k in the initial slow-roll era. Notice that with this parametrization, (25) represent the scale dependent generalization of the purely local bispectrum through the scale dependent $f_{\mathrm{NL}}^{\mathrm{eff}}(q)$.

Noticing that $P_{\mathcal{R}}(\tau_p) = 2\pi^2\mathcal{P}_{\mathcal{R}}(\tau_p)/p^3$ and $\mathcal{P}_{\mathcal{R}}(\tau_p) \simeq 2.1 \times 10^{-9}$ for mode exit during the initial slow-roll era (see e.g. (A.3)), we insert (25) in (24) to express the angular cross correlation as

$$C_l^{\mu T} \simeq 2.7 \times 10^{-17}\, \frac{2\pi}{l(l+1)}\ b_{(\mathrm{pbh})}(l), \tag{26}$$

where we define

$$b_{(\mathrm{pbh})}(l) \equiv \frac{6l(l+1)}{\ln\left(\frac{k_D(z_i)}{k_D(z_f)}\right)} \int \mathrm{d}\ln k_+\ \Delta_l(k_+)\ j_l\left(k_+\chi_*\right) \int \mathrm{d}\ln q\ f_{\mathrm{NL}}^{\mathrm{eff}}(q) \left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}, \tag{27}$$

as the key quantity that determines the multipole dependence and the size of the $\langle \mu T \rangle$ correlator of Eq. (26). Notice that this quantity depends on the transfer function $\Delta_l(k)$ which we decompose as[43]

$$\Delta_l(k) = \rho(l)\Delta_l^{(\mathrm{SW})}(k)\,. \tag{28}$$

$\Delta_l^{(\mathrm{SW})}(k)$ is the transfer function in the large-scale Sachs-Wolfe (SW) limit $l \to 0$,

$$\Delta_l^{(\mathrm{SW})}(k) \;=\; \frac{1}{3} j_l(k\chi_*), \tag{29}$$

and

$$\rho(l) \simeq 1.08\left[1 - 0.022l - 1.72\times 10^{-4} l^2 + 2\times 10^{-6} l^3 - 4.56\times 10^{-9} l^4\right], \tag{30}$$

is an analytic fit that includes high-l corrections to the SW approximation.[43] Using (28) and (29), we can then simply take the integral over the long momenta k_+ in (27) and factorize the outcome as

$$b_{(\mathrm{pbh})}(l) = \rho(l)\, b_{(\mathrm{pbh})}^{(\mathrm{SW})}, \tag{31}$$

with

$$b_{(\mathrm{pbh})}^{(\mathrm{SW})} = \left[\ln\left(\frac{k_D\,(z_i)}{k_D\,(z_f)}\right)\right]^{-1} \int \mathrm{d}\ln q\; f_{\mathrm{NL}}^{\mathrm{eff}}(q)\left[e^{-2q^2/k_D^2(z)}\right]_{z_f}^{z_i}. \tag{32}$$

Notice that for a purely local spectrum with constant non-linearity parameter, *i.e.* $f_{\mathrm{NL}}^{\mathrm{eff}}(q) \to f_{\mathrm{NL}}^{(\mathrm{p})}$, we get $b_{(\mathrm{pbh})}^{\mathrm{sw}} = f_{\mathrm{NL}}^{(\mathrm{p})}$ and hence reproduce the standard results derived for $C_l^{\mu T}$ in the SW limit.[40,42] However, $b_{(\mathrm{pbh})}^{\mathrm{sw}}$ can not influence the l dependence of the angular correlator $C_l^{\mu T}$(26) but it can significantly influence its overall amplitude through a non-trivial scale dependence of $f_{\mathrm{NL}}^{\mathrm{eff}}$. We will concretely see an example of this fact in the PBH forming inflationary scenarios below.

Using the analytic formulas we provide in Appendix A, B, we derive an analytic expression for the scale dependence of the effective non-linearity parameter $f_{\mathrm{NL}}^{\mathrm{eff}}$ (25) in Appendix C. In this way, we compute the expression (32) analytically to obtain

$$\frac{b_{(\mathrm{pbh})}^{(\mathrm{SW})}}{f_{\mathrm{NL}}^{(\mathrm{p})}} = 1 + \frac{1}{\ln\left(\frac{k_D(z_i)}{k_D(z_f)}\right)} \sum_{n=2}^{6} c_{\mathrm{NL}}^{(n)} \frac{\Gamma(n/2)}{2^{(n+2)/2}} \left(\frac{k_D(z)}{\mathcal{H}_0}\right)^n \Bigg|_{z_f}^{z_i}, \tag{33}$$

where the dependence of the coefficients $c_{\mathrm{NL}}^{(n)}$ on the background parameters in the non-attractor era, *i.e.* $\{\eta_{\mathrm{c}}, \Delta N\}$ can be obtained from the formulas we provide in Appendix C, B and A. Armed with (33), we can quantify the extent of which the scale dependence in the bispectrum can influence the μT angular correlator $C_l^{\mu T}$. For this purpose, we use (33) to compute $b_{(\mathrm{pbh})}^{(\mathrm{SW})}$ in terms of the location of the dip in momentum space k_{dip} by noting the relationship $\mathcal{H}_0 \simeq 100 k_{\mathrm{dip}}/3 \simeq 33\, k_{\mathrm{dip}}$ we derived in Section 2.1. The resulting dependence of $b_{(\mathrm{pbh})}^{(\mathrm{SW})}$ on the location of the dip feature k_{dip} is shown in Figure 3 for two different choices of parameter set that describes an inflationary model including a transient non-attractor era. We observe

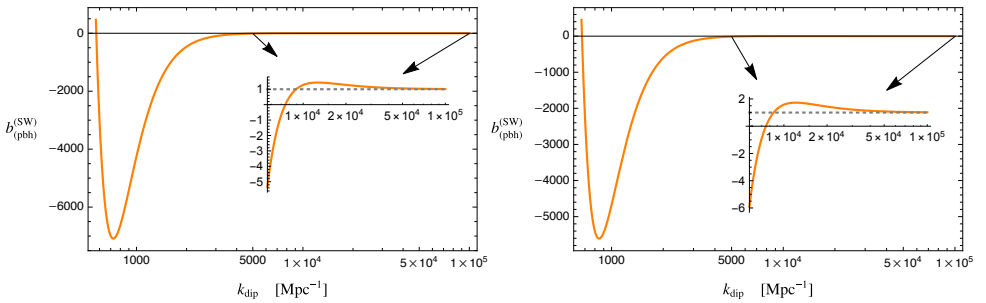


Fig. 3. Parameter $b^{(\mathrm{SW})}_{(\mathrm{pbh})}$ (33) for $f^{(\mathrm{p})}_{\mathrm{NL}} = 1$ as a function of the location of the dip feature $k_{\mathrm{dip}} \simeq 3\mathcal{H}_0/100$ in an inflationary model that contains a transient non-attractor phase parametrized by the parameter sets $\{\Delta N = 2.6, \eta_{\mathrm{c}} = -6, c_k = 0.5\}$ (Left) and $\{\Delta N = 2.5, \eta_{\mathrm{c}} = -6.2, c_k = 0.4\}$ (Right).

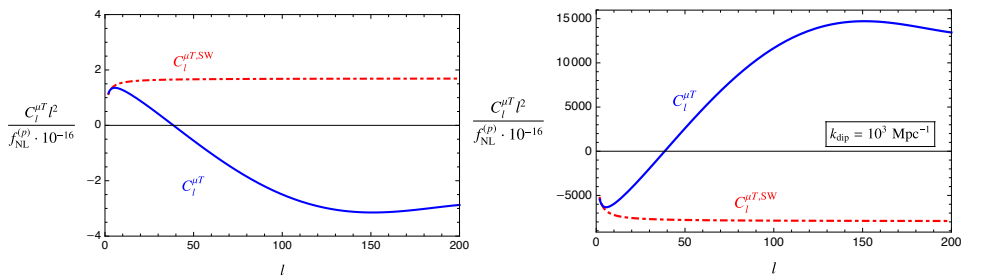


Fig. 4. Multipole dependence of $C_l^{\mu T}$ for a purely local type bispectrum (Left) and for a PBH forming inflationary scenario that contains a transient constant-roll era ($\eta_{\mathrm{c}} = -6.2$) that has a duration of $\Delta N = 2.5$ e-folds and with a dip feature that is located at $k_{\mathrm{dip}} = 10^3\,\mathrm{Mpc}^{-1}$(Right).

that for phenomenologically interesting values of k_{dip}, $b^{(\mathrm{SW})}_{(\mathrm{pbh})}$ obtains *large negative* values and hence can alter the *overall amplitude* and *sign* of the $\langle \mu T \rangle$ correlator significantly, in particular compared to a inflationary scenario endowed with a purely local type bispectrum where $b^{(\mathrm{SW})} \to f^{(\mathrm{p})}_{\mathrm{NL}}$.[40]

To further illustrate these points, in Figure 4 we present the multipole dependence of $C_l^{\mu T}$ (26) for a purely local bispectrum and for an inflationary model that contains an intermediate non-attractor phase where we show the resulting angular correlator for both within the Sachs-Wolfe approximation (dot-dashed line), and by taking into account full transfer effects (solid line). We observe that compared to the scale independent local bispectrum case (left panel in Figure 4), the overall amplitude of the angular correlator is enhanced significantly and its behavior from large to small scales is inverted due to the large negative $b^{(\mathrm{SW})}_{(\mathrm{pbh})}$ acquires for the $k_{\mathrm{dip}} = 10^3\,\mathrm{Mpc}^{-1}$. In other words, for a PBH forming inflationary scenario, μ distortions become anti-correlated with temperature anisotropies at large scales contrary to the case arise for a scale independent local bispectrum. It is worth stressing that this behavior arise because in PBH forming inflationary scenarios,

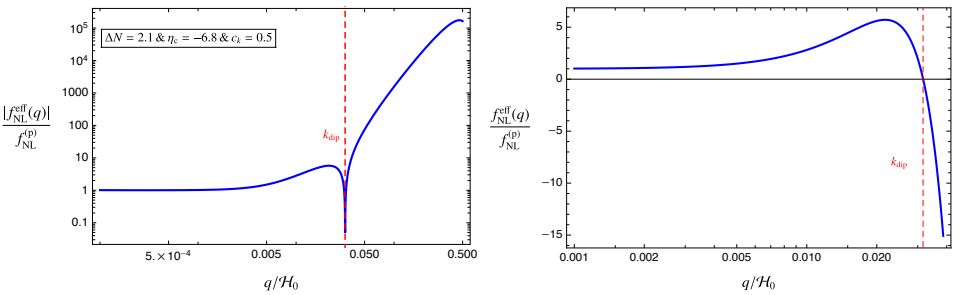


Fig. 5. Scale dependence of the effective non-linearity parameter $f_{\rm NL}^{\rm eff}$ (25) in the squeezed limit $k_3/q \to 0$ for an inflationary model that exhibit an intermediate constant-roll phase $\eta_{\rm c} = -6.8$ that lasts $\Delta N = 2.1$ e-folds.

the effective scale dependent bispectrum changes sign around the dip feature and grows large in the negative direction as we show in the right panel of Figure 5. This result is interesting on its own as one can in principle distinguish between these scenarios by just looking at the sign of the $C_l^{\mu T}$ at large scales at small l. In summary, our results imply that phases of non-attractor inflation can qualitatively and quantitatively change the properties of the cross-correlations between CMB temperature anisotropies and μ-type distortions.

3.1. *Prospects of detectability for* $\langle \mu T \rangle$

To assess the prospects of detectability for the μT correlator, we now estimate the cumulative signal-to-noise ratio, S/N using[42]

$$\left(\frac{S}{N}\right)^2 = \sum_{l=2}^{l_{\rm max}} (2l+1) \frac{\left(C_l^{\mu T}\right)^2}{C_l^{TT} C_l^{\mu\mu,N}} \,, \tag{34}$$

where C_l^{TT} is the CMB temperature anisotropy power spectrum and $C_l^{\mu\mu,N}$ is the noise level for μ distortions. For an experiment as PIXIE[44] the noise is given by $C_l^{\mu\mu,N} \simeq 4\pi\, \mu_{\rm min}^2\, e^{l^2/84^2}$ [40] where $\mu_{\rm min}$ denotes the minimum detectable μ distortion signal. Using (20), we denote the TT correlator in (34) as

$$C_l^{TT} = \frac{36\pi}{25} \int {\rm d}\ln k\, \mathcal{P}_{\mathcal{R}}(k) \Delta_l^2(k). \tag{35}$$

Therefore to estimate the S/N, we require the full knowledge the of the transfer function Δ_l in (35). Taking these effects into account, it was found that the signal to noise ratio can be estimated as 40% of the result that one would obtain by adopting the SW limit,[42,43] adopting $\rho_l \to 1$, $C_l^{TT,{\rm SW}} = 2\pi \mathcal{P}_{\mathcal{R}}^{(0)}/(25\, l(l+1))$ [d]. Following the

[d]Note that the smallest scales we are interested corresponds to $l_{\rm max} = 200$, so we can safely assume the scale indepedent part of the power spectrum in (35) although it has non-trivial scale dependence (enhancement) at much smaller scales for inflationary scenarios we are interested in this work.

discussion above, we carry the sum in (34) up to $l_{\rm max} = 200$ using (26) and (31) to obtain[11]

$$\left(\frac{S}{N}\right) \simeq 0.35 \times 10^{-3} |f_{\rm NL}^{\rm (p)} b_{\rm (pbh)}^{\rm (SW)}| \left(\frac{10^{-8}}{\mu_{\rm min}}\right). \tag{36}$$

This result implies that $|f_{\rm NL}^{\rm (p)} b_{\rm (pbh)}^{\rm (SW)}| \gtrsim 2892$ should be observable for a PIXIE like experiment. On the other hand, for a spectrometer comparable to PRISM[45] with $\mu_{\rm min} = 10^{-9}$, $|f_{\rm NL}^{\rm (p)} b_{\rm (pbh)}^{\rm (SW)}| \gtrsim 290$ is required for the detectability. At this point, it is worth mentioning that for a local type $f_{\rm NL}^{\rm (p)} \sim 10^{-2}$ at CMB scales as predicted by standard slow-roll backgrounds, such a signal would be challenging to detect.

On the other hand, considering the 2σ limits on the local type bispectrum by Planck[46] at CMB scales: $-11.1 < f_{\rm NL}^{\rm (p)} < 9.3$ ($f_{\rm NL}^{\rm (p)} = -0.9 \pm 5.1$, 68%CL) and the typical values of $|b_{\rm (pbh)}^{\rm (SW)}| \simeq$ a few $(10^2 - 10^3)$ that can be obtained in inflationary scenarios that can generate PBH populations (See figure 3), we conclude that the impact of the scale dependent non-Gaussianity on the $\langle \mu T \rangle$ correlator should be observable by PIXIE or PRISM for an interesting range of $k_{\rm dip}$ values relevant for the formation of PBHs within the mass range $M_{\rm pbh} \simeq 1 - 100\, M_\odot$.

4. Discussion

We presented a method for probing inflationary scenarios of PBH formation, using only CMB physics at relatively large scales. We based our arguments on the characteristic properties of the spectrum of curvature perturbation in single-field inflationary models that can generate a large population of PBHs. In these models, the curvature perturbation spectrum is characterized by a pronounced dip followed by a rapid growth towards a peak responsible for PBH formation. By making use of the gradient expansion formalism,[12] we analytically computed the properties of the power spectrum and for the first time of the bispectrum around the dip position, which occurs at scales well larger than the peak. Focussing our attention to the squeezed configuration, we found that the bispectrum can be enhanced around the position of the dip; it also acquires a characteristic momentum dependence that is controlled by the underlying inflationary mechanism. Here, we propose to probe such an enhanced squeezed bispectrum through the correlations it induces between CMB μ-distortions and CMB temperature fluctuations. For this purpose, we extended the methods first explored in[40] to include the case of scale-dependent non-Gaussianity from PBH formation mechanisms, finding analytical expression for quantities controlling μT correlations, and discussing how future CMB μ-distortion experiments can test this observable. Interestingly, the method we propose would allow one to experimentally probe inflationary PBH scenarios using well-understood CMB physics at scales much larger than the peak of the spectrum, without considering non-linearities associated with PBH formation and evolution. In particular, owing the relation between relevant scales associated with μ distortions and the

features present in the PBH forming inflationary scenarios, our findings are relevant for PBH masses within the $M_{\rm pbh} \sim 1 - 100 M_\odot$. These finding suggest that μT correlations can be used as a useful tool to distinguish between astrophysical vs primordial origin of BHs.

Acknowledgments

The work of OÖ is supported by the European Structural and Investment Funds and the Czech Ministry of Education, Youth and Sports (Project CoGraDS-CZ.02.1.01/0.0/0.0/15003/0000437). GT is partially funded by the STFC grant ST/T000813/1.

Appendix A. The power spectrum $\mathcal{P}_{\mathcal{R}}(\tau_k)$ and fractional velocity $v_{\mathcal{R}}$

For the initial slow-roll phase, the solution for the curvature perturbation that reduces to the standard Bunch Davies vacuum can be written as[13]

$$\mathcal{R}_k^{\rm sr} = \frac{iH}{M_{\rm pl}} \frac{e^{-ik\tau}}{\sqrt{4\epsilon_{\rm sr} k^3}} \, (1 + ik\tau), \qquad \tau_k/\tau_0 > 1, \tag{A.1}$$

where $\epsilon_{\rm sr} \ll 1$ is the slow-roll parameter in the initial slow-roll stage which we assume to be constant. Using the solution (A.1) in the definition, the real and imaginary part of the fractional velocity (evaluated at $\tau = \tau_k$) can be obtained as

$$v_{\mathcal{R}}^R(c_k) = -\frac{c_k^2}{3(1+c_k^2)}, \quad v_{\mathcal{R}}^I(c_k) = -\frac{c_k^3}{3(1+c_k^2)}, \qquad \tau_k/\tau_0 > 1, \tag{A.2}$$

where c_k is defined in (17). It is clear from (A.2) that the imaginary part of $v_{\mathcal{R}}$ includes an extra factor of c_k compared to the real part. We note that unless $c_k = 1$, this translates into an extra suppression for the imaginary part of the fractional velocity.

On the other hand, using (A.1), the power spectrum evaluated at around horizon crossing is given by

$$\mathcal{P}_{\mathcal{R}}(\tau_k) = \frac{k^3}{2\pi^2} |\mathcal{R}_k(\tau_k)|^2 = \frac{H^2}{8\pi^2 \epsilon_{\rm sr} M_{\rm pl}^2} \left(1 + c_k^2\right), \qquad \tau_k/\tau_0 > 1. \tag{A.3}$$

Next, we need to develop an expression for $\mathcal{R}(\tau_k)$ and the fractional velocity continuous through the transition at $\tau = \tau_0$. This was done using a matching procedure for $\mathcal{R}_k$ and its derivative in[11] which gives

$$\mathcal{R}_k = \frac{iH}{M_{\rm pl}} \frac{(\tau/\tau_0)^{\eta_{\rm c}/2}}{\sqrt{4\epsilon_{\rm sr} k^3}} (-k\tau)^{3/2} \left[c_1 H_\nu^{(1)}(-k\tau) + c_2 H_\nu^{(2)}(-k\tau) \right], \qquad \tau_k/\tau_0 < 1. \tag{A.4}$$

where $\nu = (3+\eta_c)/2$. The scale dependent coefficients are given by

$$c_1 = y_0^{3/2} e^{iy_0} \frac{y_0\left(H_\nu^{(2)}(y_0) + iH_{\nu-1}^{(2)}(y_0)\right) - H_{\nu-1}^{(2)}(y_0)}{H_{\nu-1}^{(1)}(y_0)H_\nu^{(2)}(y_0) - H_\nu^{(1)}(y_0)H_{\nu-1}^{(2)}(y_0)}, \tag{A.5}$$

$$c_2 = y_0^{3/2} e^{iy_0} \frac{(y_0)\left(H_\nu^{(1)}(y_0) + iH_{\nu-1}^{(1)}(y_0)\right) - H_{\nu-1}^{(1)}(y_0)}{H_\nu^{(1)}(y_0)H_{\nu-1}^{(2)}(y_0) - H_\nu^{(2)}(y_0)H_{\nu-1}^{(1)}(y_0)}. \tag{A.6}$$

where we defined $y \equiv -k\tau$. Then for $\tau_k/\tau_0 < 1$, the real and the imaginary part of the fractional velocity can be obtained as

$$v_{\mathcal{R}}^R(\tau) = -\frac{y}{3}\left[\frac{f_1 f_3 - y_0\left(f_1 f_4 + f_2 f_3\right) + y_0^2\left(f_1 f_3 + f_2 f_4\right)}{f_3^2 - 2y_0 f_3 f_4 + y_0^2\left(f_3^2 + f_4^2\right)}\right], \tag{A.7}$$

$$v_{\mathcal{R}}^I(\tau) = -\frac{y}{3}\left[\frac{y_0^2\left(f_1 f_4 - f_2 f_3\right)}{f_3^2 - 2y_0 f_3 f_4 + y_0^2\left(f_3^2 + f_4^2\right)}\right], \tag{A.8}$$

where we define the functions $f_n = f_n(y, y_0, \nu)$, for $n = 1, 2, 3, 4$ in terms of the Bessel function of the first and second kind as

$$\begin{aligned} f_1(y, y_0, \nu) &= J_{\nu-1}(y_0)Y_{\nu-1}(y) - Y_{\nu-1}(y_0)J_{\nu-1}(y), \\ f_2(y, y_0, \nu) &= J_\nu(y_0)Y_{\nu-1}(y) - Y_\nu(y_0)J_{\nu-1}(y). \end{aligned} \tag{A.9}$$

Furthermore, we note the following relations $f_4 = f_1(y, y_0, \nu+1)$, $f_3 = -f_2(y_0, y, \nu)$. The continuity of the real and the imaginary part of the fractional velocity can be confirmed explicitly from the eqs in (A.7) and (A.8) as they reduce to their slow-roll counterparts provided in (A.2) at the transition point $\tau_k = \tau_0$. Using the relations we noted earlier, power spectrum evaluated at $\tau = \tau_k$ for modes that leave the horizon during the non-attractor phase ($\tau_k/\tau_0 < 1$) can be obtained as

$$\mathcal{P}_{\mathcal{R}}(\tau_k) = \frac{H^2}{8\pi^2\epsilon_{\rm sr}M_{\rm pl}^2}\frac{c_k^{2\nu}\pi^2}{4}\left(\frac{k}{\mathcal{H}_0}\right)^{2-2\nu}\left[f_3^2 - 2y_0 f_3 f_4 + y_0^2(f_3^2 + f_4^2)\right]_{\tau=\tau_k} \tag{A.10}$$

Similarly, the continuity of the power spectrum can be confirmed explicitly by evaluating (A.10) at the transition point $\tau_k = \tau_0$ at which it reduced to (A.3).

Appendix B. The functions $D(\tau_k)$, $F_k(\tau_k)$

For the two phase background model parametrized by the pump field profile in Eq. (15), scale dependent functions $D(\tau_k)$, $F(\tau_k)$ (see Eqs. (6) and (7)) are calculated in[13] for modes that exit during the initial slow-roll era ($k/\mathcal{H}_0 < c_k$) and in[11]

for mode exit during the non-attractor era ($e^{\Delta N} > k/\mathcal{H}_0 > c_k$). Below, we present the explicit formulas for these functions for both stages of inflationary evolution.

$$D(\tau_k) \equiv \mathcal{C}_0^D + \mathcal{C}_3^D \left(\frac{k}{\mathcal{H}_0}\right)^3, \tag{B.1}$$
$$\frac{F(\tau_k)}{\tau_0^2} \equiv \mathcal{C}_{-2}^F \left(\frac{k}{\mathcal{H}_0}\right)^{-2} + \mathcal{C}_0^F + \mathcal{C}_1^F \left(\frac{k}{\mathcal{H}_0}\right), \qquad \frac{k}{\mathcal{H}_0} < c_k,$$

where the coefficients are given by

$$\mathcal{C}_0^D = 1, \qquad \mathcal{C}_3^D \simeq -\frac{3\, e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)c_k^3}, \qquad \mathcal{C}_{-2}^F = \frac{c_k^2}{6},$$
$$\mathcal{C}_0^F \simeq -\frac{\eta_c\, e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)(\eta_c+1)}, \qquad \mathcal{C}_1^F \simeq \frac{e^{-(\eta_c+3)\Delta N}}{c_k(\eta_c+3)} \tag{B.2}$$

where $\tau_f/\tau_0 \equiv x_f = e^{-\Delta N}$ with ΔN denoting the duration of the slow-roll violating ($\eta_c \leq -6$) era. On the other hand, for modes that leave the horizon during the non-attractor era, we have

$$D(\tau_k) \equiv \tilde{\mathcal{C}}_{\eta_c+3}^D \left(\frac{k}{\mathcal{H}_0}\right)^{\eta_c+3} + \tilde{\mathcal{C}}_0^D \tag{B.3}$$
$$\frac{F(\tau_k)}{\tau_0^2} \equiv \tilde{\mathcal{C}}_{\eta_c+1}^F \left(\frac{k}{\mathcal{H}_0}\right)^{\eta_c+1} + \tilde{\mathcal{C}}_{-2}^F \left(\frac{k}{\mathcal{H}_0}\right)^{-2} + \tilde{\mathcal{C}}_0^F, \qquad c_k < \frac{k}{\mathcal{H}_0} < e^{\Delta N}$$

where

$$\tilde{\mathcal{C}}_{\eta_c+3}^D = -\frac{3\, e^{-(\eta_c+3)\Delta N}}{(\eta_c+3)c_k^{\eta_c+3}}, \qquad \tilde{\mathcal{C}}_0^D = -\frac{3}{\eta_c+3}, \qquad \tilde{\mathcal{C}}_0^F = \frac{e^{-2\Delta N}}{2(\eta_c+1)},$$
$$\tilde{\mathcal{C}}_{-2}^F = \frac{c_k^2}{2(\eta_c+3)}, \qquad \tilde{\mathcal{C}}_{\eta_c+1}^F = \frac{e^{-(\eta_c+3)\Delta N}}{(\eta_c+1)(\eta_c+3)c_k^{\eta_c+1}}. \tag{B.4}$$

Appendix C. $f_{\rm NL}^{\rm eff}$ in the squeezed limit

Using the definition (25) with (19), for mode exit during the initial slow-roll era, the squeezed limit effective non-linearity parameter is given by

$$f_{\rm NL}^{\rm eff} = \frac{5}{3}\left(\mathrm{Re}[\alpha_q^* \alpha_{k_3}]\, F(\tau_{k_q}) q^2\right). \tag{C.1}$$

From (C.1), one can show that in the large scale tail of the squeezed limit, *i.e.* $q > k_+$ as $q \to 0$, $f_{\rm NL}^{\rm eff}$ obtains an "initial" value $f_{\rm NL}^{(0)} = 5c_{\rm NL}^{(0)}(c_q)/3$ where $c_{\rm NL}^{(0)} \simeq (1 + v_{\mathcal{R}}^R - \mathcal{C}_{-2}^F)^2 \mathcal{C}_{-2}^F$. Since, we send $q \to 0$ to reach this initial value, we do not expect that $f_{\rm NL}^{(0)}$ can reflect an accurate initial value for the squeezed configuration of (C.1). To keep our discussion on the observability of $\langle \mu T \rangle$ (see Section 3.1) as general as possible, we therefore replace this initial value with a fiducial primordial $f_{\rm NL}^{(\rm p)}$ to characterize the effective non-linearity parameter at large scales, *i.e.* at CMB scales. Using the formulas we derived for the functions $D(\tau_k), F(\tau_k)$ above and the

definitions of the enhancement factor α_k in eqs.(3) and (4), at leading order in the $k_+/q \ll 1$ expansion, we can then express $f_{\rm NL}^{\rm eff}$ in terms of simple power law expansion as[11]

$$f_{\rm NL}^{\rm eff} \simeq f_{\rm NL}^{\rm (p)} \left[1 + \sum_{n=2}^{6} c_{\rm NL}^{(n)}(\eta_{\rm c}, \Delta N, c_q) \left(\frac{q}{\mathcal{H}_0} \right)^n \right] + \mathcal{O}\left(\frac{k_+}{q} \right). \tag{C.2}$$

The coefficients $c_{\rm NL}^{(n)}$ of this expansion can be expressed in terms of the coefficients $\mathcal{C}$ of the F, D functions of Appendix B and the fractional velocities $v_{\mathcal{R}}^R, v_{\mathcal{R}}^I$ we defined in (A.2) as

$$c_{\rm NL}^{(0)} \simeq (1 + v_{\mathcal{R}}^R - \mathcal{C}_{-2}^F)^2 \mathcal{C}_{-2}^F, \qquad c_{\rm NL}^{(2)} \simeq \frac{\mathcal{C}_0^F}{\mathcal{C}_{-2}^F}, \qquad c_{\rm NL}^{(3)} \simeq \frac{\mathcal{C}_1^F}{\mathcal{C}_{-2}^F} \tag{C.3}$$

$$c_{\rm NL}^{(4)} = -\frac{(\mathcal{C}_0^F)^2}{\mathcal{C}_{-2}^F(1 + v_{\mathcal{R}}^R - \mathcal{C}_{-2}^F)}, \qquad c_{\rm NL}^{(5)} \simeq (1 + v_{\mathcal{R}}^R - \mathcal{C}_{-2}^F) \left[v_{\mathcal{R}}^R \mathcal{C}_3^D - 2\mathcal{C}_1^F \right] \mathcal{C}_0^F, \tag{C.4}$$

$$c_{\rm NL}^{(6)} \simeq (1 + v_{\mathcal{R}}^R - \mathcal{C}_{-2}^F) \left[v_{\mathcal{R}}^R \mathcal{C}_3^D - \mathcal{C}_1^F \right] \mathcal{C}_1^F, \tag{C.5}$$

where we neglected terms proportional to $v_{\mathcal{R}}^I$ for $c_k \ll 1$ because in this regime $v_{\mathcal{R}}^R \gg v_{\mathcal{R}}^I$ as can be realized from (A.2).

References

1. S. Hawking, Gravitationally collapsed objects of very low mass, *Mon. Not. Roy. Astron. Soc.* **152**, p. 75 (1971).
2. B. J. Carr and S. W. Hawking, Black holes in the early Universe, *Mon. Not. Roy. Astron. Soc.* **168**, 399 (1974).
3. B. J. Carr, The Primordial black hole mass spectrum, *Astrophys. J.* **201**, 1 (1975).
4. B. Carr, F. Kuhnel and M. Sandstad, Primordial Black Holes as Dark Matter, *Phys. Rev.* **D94**, p. 083504 (2016).
5. M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, Primordial black holes—perspectives in gravitational wave astronomy, *Class. Quant. Grav.* **35**, p. 063001 (2018).
6. B. Carr and F. Kuhnel, Primordial Black Holes as Dark Matter: Recent Developments, *Ann. Rev. Nucl. Part. Sci.* **70**, 355 (2020).
7. A. M. Green and B. J. Kavanagh, Primordial Black Holes as a dark matter candidate, *J. Phys. G* **48**, p. 4 (2021).
8. C. Germani and I. Musco, Abundance of Primordial Black Holes Depends on the Shape of the Inflationary Power Spectrum, *Phys. Rev. Lett.* **122**, p. 141302 (2019).
9. C. T. Byrnes, P. S. Cole and S. P. Patil, Steepest growth of the power spectrum and primordial black holes, *JCAP* **1906**, p. 028 (2019).
10. G. Tasinato, An analytic approach to non-slow-roll inflation, *Phys. Rev. D* **103**, p. 023535 (2021).
11. O. Özsoy and G. Tasinato, CMB μT cross correlations as a probe of primordial black hole scenarios, *Phys. Rev. D* **104**, p. 043526 (2021).
12. S. M. Leach, M. Sasaki, D. Wands and A. R. Liddle, Enhancement of superhorizon scale inflationary curvature perturbations, *Phys. Rev.* **D64**, p. 023512 (2001).
13. O. Özsoy and G. Tasinato, On the slope of the curvature power spectrum in non-attractor inflation, *JCAP* **04**, p. 048 (2020).

14. M. Mylova, O. Özsoy, S. Parameswaran, G. Tasinato and I. Zavala, A new mechanism to enhance primordial tensor fluctuations in single field inflation, *JCAP* **1812**, p. 024 (2018).
15. O. Ozsoy, M. Mylova, S. Parameswaran, C. Powell, G. Tasinato and I. Zavala, Squeezed tensor non-Gaussianity in non-attractor inflation, *JCAP* **1909**, p. 036 (2019).
16. Y.-i. Takamizu, S. Mukohyama, M. Sasaki and Y. Tanaka, Non-Gaussianity of super-horizon curvature perturbations beyond δ N formalism, *JCAP* **1006**, p. 019 (2010).
17. Y.-i. Takamizu, Application of beyond δN formalism: Varying sound speed, *Phys. Rev. D* **89**, p. 043528 (2014).
18. H. Motohashi and W. Hu, Primordial Black Holes and Slow-Roll Violation, *Phys. Rev.* **D96**, p. 063503 (2017).
19. M. Biagetti, G. Franciolini, A. Kehagias and A. Riotto, Primordial Black Holes from Inflation and Quantum Diffusion, *JCAP* **1807**, p. 032 (2018).
20. J. Garcia-Bellido and E. Ruiz Morales, Primordial black holes from single field models of inflation, *Phys. Dark Univ.* **18**, 47 (2017).
21. J. M. Ezquiaga, J. Garcia-Bellido and E. Ruiz Morales, Primordial Black Hole production in Critical Higgs Inflation, *Phys. Lett.* **B776**, 345 (2018).
22. G. Ballesteros and M. Taoso, Primordial black hole dark matter from single field inflation, *Phys. Rev.* **D97**, p. 023501 (2018).
23. M. P. Hertzberg and M. Yamada, Primordial Black Holes from Polynomial Potentials in Single Field Inflation, *Phys. Rev.* **D97**, p. 083509 (2018).
24. M. Cicoli, V. A. Diaz and F. G. Pedro, Primordial Black Holes from String Inflation, *JCAP* **1806**, p. 034 (2018).
25. O. Ozsoy, S. Parameswaran, G. Tasinato and I. Zavala, Mechanisms for Primordial Black Hole Production in String Theory, *JCAP* **1807**, p. 005 (2018).
26. R. Mahbub, Primordial black hole formation in inflationary α-attractor models, *Phys. Rev. D* **101**, p. 023533 (2020).
27. G. Ballesteros, J. Rey, M. Taoso and A. Urbano, Primordial black holes as dark matter and gravitational waves from single-field polynomial inflation, *JCAP* **07**, p. 025 (2020).
28. J. Liu, Z.-K. Guo and R.-G. Cai, Analytical approximation of the scalar spectrum in the ultraslow-roll inflationary models, *Phys. Rev. D* **101**, p. 083535 (2020).
29. K. Kefala, G. P. Kodaxis, I. D. Stamou and N. Tetradis, Features of the inflaton potential and the power spectrum of cosmological perturbations (10 2020).
30. J. Chluba, J. Hamann and S. P. Patil, Features and New Physical Scales in Primordial Observables: Theory and Observation, *Int. J. Mod. Phys.* **D24**, p. 1530023 (2015).
31. X. Chen, Running non-Gaussianities in DBI inflation, *Phys. Rev.* **D72**, p. 123518 (2005).
32. C. T. Byrnes, S. Nurmi, G. Tasinato and D. Wands, Scale dependence of local fNL, *JCAP* **1002**, p. 034 (2010).
33. C. T. Byrnes, M. Gerstenlauer, S. Nurmi, G. Tasinato and D. Wands, Scale-dependent non-Gaussianity probes inflationary physics, *JCAP* **1010**, p. 004 (2010).
34. V. Atal and C. Germani, The role of non-gaussianities in Primordial Black Hole formation, *Phys. Dark Univ.* **24**, p. 100275 (2019).
35. S. Passaglia, W. Hu and H. Motohashi, Primordial black holes and local non-Gaussianity in canonical inflation, *Phys. Rev. D* **99**, p. 043536 (2019).
36. Q. Gao, Primordial black holes and secondary gravitational waves from chaotic inflation (2 2021).
37. W. Hu, D. Scott and J. Silk, Power spectrum constraints from spectral distortions in the cosmic microwave background, *Astrophys. J. Lett.* **430**, L5 (1994).

38. R. Khatri, R. A. Sunyaev and J. Chluba, Does Bose-Einstein condensation of CMB photons cancel \mu distortions created by dissipation of sound waves in the early Universe?, *Astron. Astrophys.* **540**, p. A124 (2012).
39. J. Chluba and R. A. Sunyaev, The evolution of CMB spectral distortions in the early Universe, *Mon. Not. Roy. Astron. Soc.* **419**, 1294 (2012).
40. E. Pajer and M. Zaldarriaga, A New Window on Primordial non-Gaussianity, *Phys. Rev. Lett.* **109**, p. 021302 (2012).
41. J. García-Bellido, Massive Primordial Black Holes as Dark Matter and their detection with Gravitational Waves, *J. Phys. Conf. Ser.* **840**, p. 012032 (2017).
42. J. Ganc and E. Komatsu, Scale-dependent bias of galaxies and mu-type distortion of the cosmic microwave background spectrum from single-field inflation with a modified initial state, *Phys. Rev. D* **86**, p. 023518 (2012).
43. J. Chluba, E. Dimastrogiovanni, M. A. Amin and M. Kamionkowski, Evolution of CMB spectral distortion anisotropies and tests of primordial non-Gaussianity, *Mon. Not. Roy. Astron. Soc.* **466**, 2390 (2017).
44. A. Kogut *et al.*, The Primordial Inflation Explorer (PIXIE): A Nulling Polarimeter for Cosmic Microwave Background Observations, *JCAP* **07**, p. 025 (2011).
45. P. André *et al.*, PRISM (Polarized Radiation Imaging and Spectroscopy Mission): An Extended White Paper, *JCAP* **02**, p. 006 (2014).
46. Y. Akrami *et al.*, Planck 2018 results. IX. Constraints on primordial non-Gaussianity, *Astron. Astrophys.* **641**, p. A9 (2020).

Theoretical and numerical aspects of CMB spectral distortions from non-thermal energy injections

Sandeep Kumar Acharya
Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester, M13 9PL, United Kingdom
E-mail: sandeep.acharya@manchester.ac.uk

Jens Chluba
Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester, M13 9PL, United Kingdom
E-mail: jens.chluba@manchester.ac.uk

Abir Sarkar
Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester, M13 9PL, United Kingdom
E-mail: abirsarkar.150490@gmail.com

Compton scattering of photons with thermal electrons is one of the most ubiquitous phenomenon in nature. Numerical approaches to solve the evolution of photons in such situations typically assume the energy exchange to be much smaller compared to the temperature of electrons. In this work, we solve the photon evolution in its full generality, without the diffusion approximation, and show the differences between the exact solution and approximate solutions. We point out the importance of solving exact kinematics of Compton scattering for the computation of Cosmic Microwave Background (CMB) spectral distortions at redshifts $(z) \lesssim 3 \times 10^4$.

Keywords: Cosmology; Cosmic Microwave background; CMB spectral distortions.

1. Photon evolution equation

The Boltzmann equation for the evolution of photon spectrum in contact with a thermal electron bath can be written as,[1]

$$\frac{\mathrm{d}n_0}{\mathrm{d}\tau} = \int P^{\mathrm{th}}(\omega_0 \rightarrow \omega) \left[\mathrm{e}^{\frac{\omega-\omega_0}{\theta_e}} n(1+n_0) - n_0(1+n)\right] d\omega, \tag{1}$$

where n_0 and n are the photon occupation number at frequency ω_0 and ω respectively, $\theta_e (= \frac{T_e}{m_e})$ is the non-dimensional electron temperature, $d\tau (= \mathrm{c} N_e \sigma_{\mathrm{T}} dt)$ is the optical depth of Compton scattering of photons with the electrons. The exact

kinematics of Compton scattering is stored in the kernel $P^{\rm th}(\omega_0 \rightarrow \omega)$ which can be computed using `CSpack`[a] .[2] The kernel satisfies the detailed balance relation,

$$P^{\rm th}(\omega \rightarrow \omega_0) = \frac{\omega_0^2}{\omega^2} {\rm e}^{\frac{\omega-\omega_0}{\theta_e}}\, P^{\rm th}(\omega_0 \rightarrow \omega), \tag{2}$$

which ensures that the Planck spectrum is the equilibrium solution of photon evolution equation.

In literature, one often uses the diffusion approximation[1] which assumes that the energy exchange between photons and the electrons is much smaller than the electron temperature. Then one can do a Taylor series expansion in $\frac{\omega-\omega_0}{\theta_e}$ in Eq. 1. Further assuming Compton scattering to be non-relativistic, we obtain Kompaneets equation,[3,4]

$$\frac{{\rm d}n_0}{{\rm d}\tau} = \frac{\theta_e}{x_0^2}\frac{\partial}{\partial x_0} x_0^4 \left[\frac{\partial}{\partial x_0} n_0 + n_0(1+n_0)\right], \tag{3}$$

where $x_0 (= \frac{\omega_0}{\theta_e})$ is the dimensionless frequency. The photon equilibrium solution is given by,

$$\frac{\partial}{\partial x_0} n_0 + n_0(1+n_0) = 0 \tag{4}$$

We can obtain a relativistically exact equation within diffusion approximation by using the Ansatz (without stimulated scattering terms),[5,6]

$$\frac{{\rm d}n_0}{{\rm d}\tau} = \frac{1}{\omega_0^2}\partial_{\omega_0}\left[\mathcal{D}(\omega_0)\,\partial_{\omega_0}\Delta n_0 + \mathcal{A}(\omega_0)\Delta n_0\right], \tag{5}$$

where $A(\omega_0)$ and $D(\omega_0)$ are tuned to match the relativistically exact first and second moment of photon evolution equation.[5] We call this the improved Fokker-Planck (FP) approximation. Note that the zeroth moment which is the photon number conservation equation is automatically satisfied by Eq. 5. Since the coefficients of Eq. 5 are a function of photon energy ω_0, the equilibrium solution will not be unique and will be a function of photon energy which is unphysical. Therefore, we expect that Eq. 5 can not be a reliable substitute of Eq. 1. We will confirm it with numerical test problems in the next section.

2. Numerical solutions for test problems

As a test problem, we consider monochromatic photon injection at frequency x_{inj}. In Fig. 1, we plot the first and second moment of the evolving photon spectrum as a function of $y = \int \theta_e d\tau$, which can be thought of as a time variable. With time, the photon spectrum spreads out. Therefore, the absolute value of the moments increases until equilibrium is reached at which point there is no more evolution. Since the injected photon energy is higher than the electron temperature ($x = \omega/\theta_e$),

[a] `CSpack` is available at `www.chluba.de/CSpack`

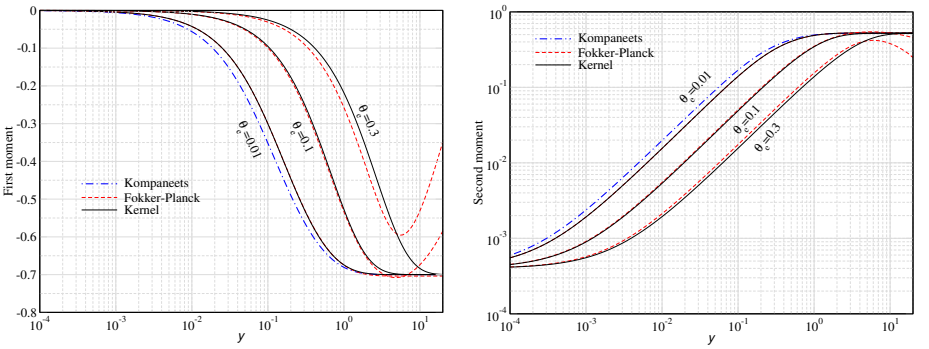


Fig. 1. Evolution of first (left panel) and second (right panel) moments of photon field for injection at $x_{\rm inj} = 10$. Stimulated terms were neglected. The electrons are kept at a constant temperature $\theta_e = 0.1$. We compare the solutions of the Kompaneets equation and the improved FP approach (Eq. 5) with the solution for the exact scattering kernel (Eq. 1). The Kompaneets solutions are independent of the temperature. The temperature dependence is captured well by the improved FP approach at low value of the y parameter. However, this approach fails as expected once equilibrium is reached.

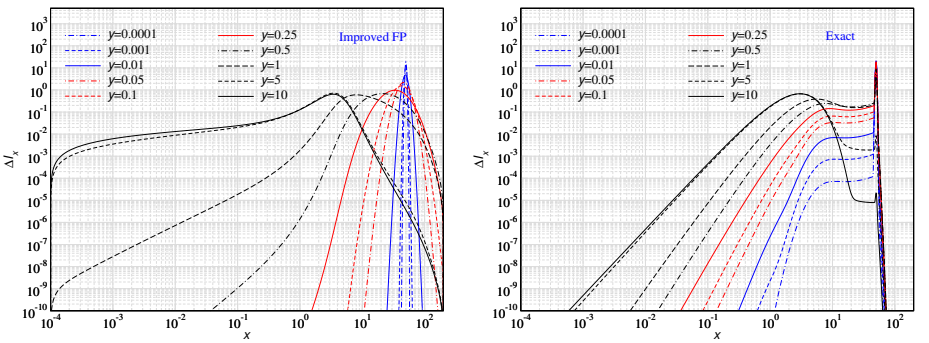


Fig. 2. Snapshots of solution for photon field as a function of frequency for different y parameters and numerical approaches as annotated. On the y-axis, we plot the dimensionless intensity, $I_x = x^3\, n$. Photon were injected at $x_{\rm inj} = 50$ at $y = 0$.

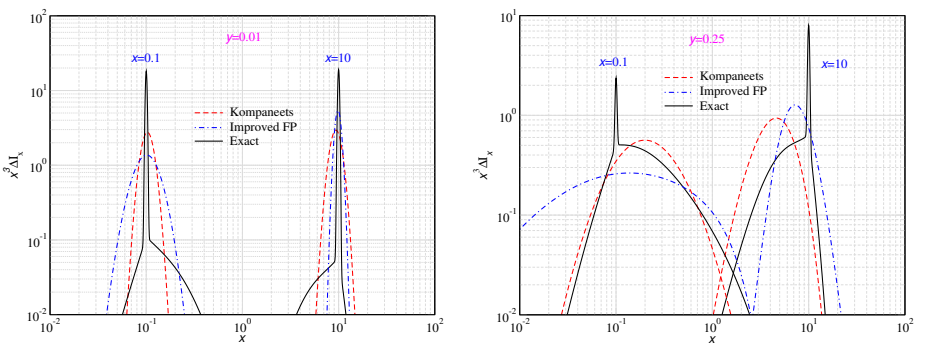


Fig. 3. Comparison of solutions with different approximations and the exact solution with $\theta_e = 0.1$ and various values of injection frequency (x) and y as shown in the figure.

photons lose energy to the electrons on average. This makes the first moment negative for this particular choice of initial photon energy. The second moment which captures the dispersion of the photon spectrum is always positive. The improved FP solution captures the evolution of moments much better than the Kompaneets equation but they do not approach the equilibrium solution for reasons as already explained before.

In Fig. 2, we plot the snapshots of photon evolution and compare the improved FP and the exact solution for a few values of y. The different solutions look significantly different from each other even if the moments of the spectrum are similar. The improved FP solution is much smoother due to the diffusion approximation while in case of the exact solution, we still see a signature of photon injection at $x_{inj} = 50$ at $y \sim 10$. In Fig. 3, we compare Kompaneets, improved FP and the exact solution for few different cases to further illustrate the differences of different approximations.

3. Implications for CMB spectral distortions

The CMB spectrum is accurately described by the Planck spectrum up to a level of 10^{-5}.[7] Injection of non-thermal photons at $z \lesssim 2 \times 10^6$ gives rise to CMB spectral distortions as the CMB can not relax to the Planck spectrum as photon number changing processes become inefficient.[8] As opposed to the test problems considered in the previous section in which electrons were at a constant temperature, in the expanding universe, both the CMB photons and the background electrons evolve together. Therefore, in this scenario, we solve two equations which are the photon evolution equation (Eq. 1) and the temperature of the electrons.[5] We evolve the photon spectrum from the redshift of photon injection up to today when the CMB photons reach us. Here, we consider photon injection at a single redshift.

In Fig. 4, we plot the CMB spectral distortion solutions for photons with energy and redshifts as shown and compare the exact solution with the Kompaneets

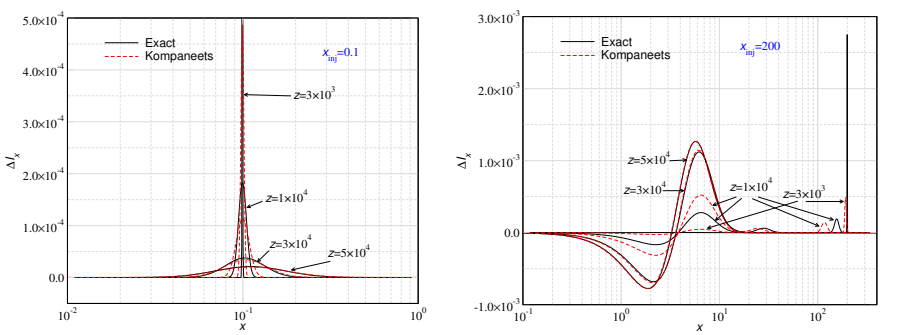


Fig. 4. Spectral distortion solution for energy injection at different redshifts with frequency of injection as denoted. The Kompaneets solution generally overestimates the broadening and mean shifts in the photon distribution. For injection at $x = 200$, the exact solution falls back with respect to the Kompaneets treatment, which overestimates the energy exchange with electrons.

approximation. The electron temperature at these redshifts are of the order of few eV, therefore, relativistic corrections are not important, and the improved FP approximation is identical to Kompaneets approximation. The Kompaneets approximation overestimates the broadening of photon spectrum. At higher frequencies, the Klein-Nishina corrections reduce the energy exchange from the energetic photons to the background electrons which is not captured by the Kompaneets solution. The energy gain by the electrons shows up as y-distortion[4] which, consequently, is not predicted correctly in Kompaneets approximation.

4. Conclusions

In this work, we solve the evolution of photons in contact with a thermal bath of electrons via Compton scattering and compare the exact solutions with various approximations found in literature. We point out and explain the differences between these solutions and show that the various approximations do not capture the physics correctly for the whole photon evolution history. We study CMB spectral distortions from photon injections and show that the exact solution differs significantly from Kompaneets solution and will be important for determining CMB spectral distortion constraints for photon injection scenarios.

References

1. S. Y. Sazonov and R. A. Sunyaev, The Profile of a Narrow Line after Single Scattering by Maxwellian Electrons: Relativistic Corrections to the Kernel of the Integral Kinetic Equation, *ApJ* **543**, 28 (November 2000).
2. A. Sarkar, J. Chluba and E. Lee, Dissecting the Compton scattering kernel I: Isotropic media, *MNRAS* **490**, 3705 (December 2019).
3. A. Kompaneets, The establishment of thermal equilibrium between photons and electrons, *Sov.Phys. JETP* **31**, p. 876 (1956).
4. Y. B. Zeldovich and R. A. Sunyaev, The Interaction of Matter and Radiation in a Hot-Model Universe, *ApSS* **4**, 301 (July 1969).
5. S. K. Acharya, J. Chluba and A. Sarkar, Comparison of numerical methods for computing the repeated Compton scattering of photons in isotropic media, *MNRAS* **507**, 2052 (October 2021).
6. R. Belmont, Numerical computation of isotropic Compton scattering, *A&A* **506**, 589 (November 2009).
7. D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer and E. L. Wright, The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set, *ApJ* **473**, 576 (December 1996).
8. J. Chluba and R. A. Sunyaev, The evolution of CMB spectral distortions in the early Universe, *MNRAS* **419**, 1294 (January 2012).

BISOU: A balloon project to measure the CMB spectral distortions

B. Maffei[1,*], M. H. Abitbol[2], N. Aghanim[1], J. Aumont[3], E. Battistelli[4], J. Chluba[5], X. Coulon[1], P. De Bernardis[4], M. Douspis[1], J. Grain[1], S. Gervasoni[1], J. C. Hill[6,7], A. Kogut[8], S. Masi[4], T. Matsumura[9], C. O'Sullivan[10], L. Pagano[11], G. Pisano[4], M. Remazeilles[5], A. Ritacco[1], A. Rotti[5], V. Sauvage[1], G. Savini[12], S. L. Stever[13,9], A. Tartari[14], L. Thiele[15], N. Trappe[10]

[1] *Institut d'Astrophysique Spatiale, CNRS-Université Paris-Saclay, Orsay, 91405, France*
[*] *E-mail: Bruno.Maffei@universite-paris-saclay.fr*

[2] *University of Oxford, Department of Physics, Denys Wilkinson Building, Oxford OX1 4LS, UK*

[3] *IRAP - CNRS, Toulouse, France*

[4] *Dipartimento di Fisica, Università di Roma "La Sapienza", Italy*

[5] *JBCA, School of Physics and Astronomy, The University of Manchester, UK*

[6] *Department of Physics, Columbia University, New York, NY, USA 10027*

[7] *Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA 10010*

[8] *NASA - Goddard Space Flight Center, Greenbelt MD 20771 USA*

[9] *Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba, 277-8583 Japan*

[10] *Department of Experimental Physics, National University of Ireland, Maynooth, Ireland*

[11] *Dipartimento di Fisica e Scienze della Terra - Università degli Studi di Ferrara, Italy*

[12] *Physics and Astronomy Department, University College London, UK*

[13] *Okayama University, Kita-ku, Okayama, 700-8530 Japan*

[14] *Dipartimento di Fisica "E. Fermi" - Università di Pisa - INFN, Pisa, Italy*

[15] *Department of Physics, Princeton University, Princeton, NJ, 08544 USA*

The BISOU (Balloon Interferometer for Spectral Observations of the Universe) project aims to study the viability and prospects of a balloon-borne spectrometer, pathfinder of a future space mission dedicated to the measurements of the CMB spectral distortions. We present here a preliminary concept based on previous space mission proposals, together with some sensitivity calculation results for the observation goals, showing that a 5-σ measurement of the y-distortions is achievable.

Keywords: CMB, Spectral distortions, Balloon project.

1. Introduction

With the success of the ESA Planck mission, the concordance cosmological model is established as the reference framework. However, outstanding questions about this model are still unanswered. In particular the simplest inflationary model proposed as the origin of the initial matter perturbations is favoured by Planck measurement of the spectral index and low non-Gaussianity. Nevertheless, it still needs to be confirmed through the measurement of its smoking gun signature: the relic background of primordial gravitational waves. The latter can only be observed through the Cosmic Microwave Background (CMB) polarisation: namely B-modes. The CMB frequency spectrum is another key observable to probe the cosmological model. Its intensity was precisely measured by COBE/FIRAS almost three decades ago,[1,2] with deviations limited to $\Delta I/I \approx 10^{-5}$. Since then, not much progress has been achieved in measuring its deviation from a true blackbody. However, while the space mission proposals PIXIE[3] and PRISTINE (to ESA F-mission call) have not been successful, following two white papers,[4,5] the ESA Voyage 2050 programme has selected this topic amongst its three themes.

The BISOU (Balloon Interferometer for Spectral Observations of the Universe) project aims to study the viability and prospects of a balloon-borne spectrometer, pathfinder of a future space mission dedicated to the absolute measurement of the CMB spectrum. While PIXIE and PRISTINE were targeting both the measurement of the CMB polarisation, namely a first detection of the CMB B-mode polarisation, and the absolute measurement of the CMB spectral distortions, BISOU's main goal is to perform a first measurement of the later. However, secondary science will also include measurement of the Cosmic Infrared Background (CIB) emission, and potentially the polarisation of the dust up to 2 THz together with Galactic emission lines such as CI, CII, NII and OI.

Taking into account the specificity of a balloon flight in term of requirements and conditions (i.e. residual atmosphere, observation strategy for instance), this CNES Phase 0 study will evaluate if such a spectrometer is sensitive enough to measure at least the Compton y-distortion while consolidating the instrument concept and improving the readiness of some of its key sub-systems.

2. Science goals

While the CMB has a nearly perfect blackbody emission spectrum, deviations from it, referred to as spectral distortions, are expected. These distortions encode information about the full thermal history of the Universe from the early stages (primordial distortions from inflation and cosmological recombination lines) until today (star formation and galaxy clusters). Many of these processes are part of our standard cosmological model and are detailed in several publications.[4,6]

Spectral distortions result from processes that affect the thermal equilibrium between matter and radiation. One of the standard distortions, known as the Compton y-distortion, is created in the regime of inefficient energy transfer (optically

thin scattering) between electrons and photons, relevant at redshifts $z < 5 \times 10^4$. Processes creating this type of distortion are dominated by the inverse-Compton scattering of CMB photons off hot electrons during the epoch of reionization and structure formation, also known as the thermal Sunyaev-Zeldovich (tSZ) effect.

Chemical potential or μ-type distortions, on the other hand, are generated by energy release at earlier stages ($z > 5 \times 10^4$), when interactions are still extremely efficient (optically thick scattering) and able to establish kinetic equilibrium between electrons and photons under repeated Compton scattering and photon emission processes.

The COBE-FIRAS limit, $|y| < 1.5 \times 10^{-5}$ (95% C.L.), is roughly one order of magnitude larger than the expected signal,[7] $y \approx 2 \times 10^{-6}$. Signal from μ-type distortions will be even fainter. It is therefore crucial to have reliable models of all the emissions that will be much stronger than these signals in order to properly subtract them.

2.1. *Modeling the signals*

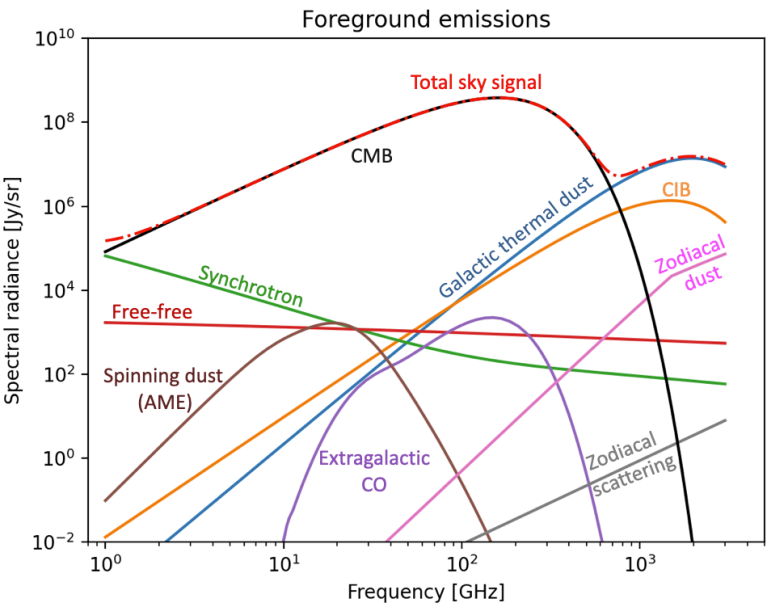


Fig. 1. Foreground emission contributions with respect to the CMB blackbody emission, together with the sum of all astrophysical signals (CMB + foregrounds).

Several astrophysical foregrounds contribute meaningfully to the sky signal at frequencies relevant to CMB spectral distortions. In addition to the ones already modeled[6] from Planck data,[8] additional components from the zodiacal dust have

been included. Figure 1 shows the individual contributions together with the sum of all sky signals, including the CMB. At low frequencies (below 70 GHz), the brightest foregrounds are from the synchrotron, the free-free and the so-called anomalous microwave emissions. High frequencies foregrounds (above 100 GHz) are mainly due to the emission of the Galactic thermal dust, the cumulative redshifted emission from thermal dust in distant galaxies, called the cosmic infrared background (CIB), and the zodiacal thermal dust emission. Additional foregrounds contributing to the total sky signal are the cumulative CO emission from distant galaxies at intermediate frequencies and the zodiacal scattering at high frequencies.

2.2. *CMB spectral distortions modeling*

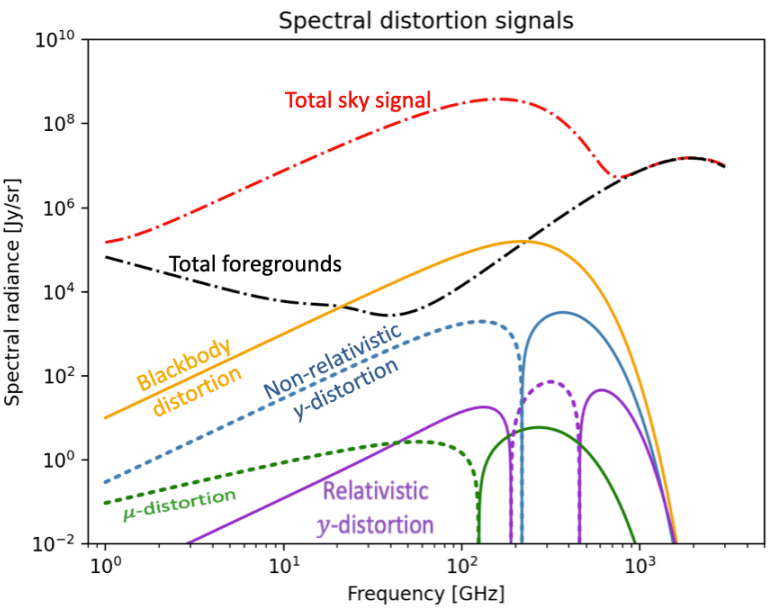


Fig. 2. Models of the spectral distortions: Black body distortion ΔT_{CMB} (orange), non-relativistic y-distortion (blue), relativistic y-distortion (purple) and μ-distortion (green). This is compared to the total foreground signal (dashed black) and the total sky signal - Foregrounds + CMB (dashed red).

Following the same assumptions and processes as the ones presented in Abitbol et al.,[6] figure 2 shows the signals associated with the various distortions of the spectrum, with respect to the total emission of the foregrounds and the total sky signal (including the CMB emission).

Four contributions are considered and modeled. First a blackbody distortion that represents a first order temperature deviation ΔT_{CMB} to the true CMB blackbody

spectrum. Then a cumulative thermal SZ y-distortion, including both standard non-relativistic and relativistic contributions (from Hill et al.[7]). The intracluster medium (ICM), the intergalactic medium and reionization contributions are included in the Compton-y signal. A relativistic correction to thermal SZ distortion[7,6] is modeled using the moment-based approach. Finally the chemical potential μ-distortion is generated assuming only signals from acoustic damping and adiabatic cooling. The r-type distortion, sometimes called "residual" distortion, is expected to have a contribution smaller than the μ-type distortion and is not represented in figure 2.

3. Instrument concept

The initial concept is based on a similar one that had been proposed by the PIXIE team[3] (shown in figure 3 left) and also used for the PRISTINE F-class ESA mission proposal. The instrument is a Fourier Transform Spectrometer (FTS) with two inputs and two outputs. Both inputs are going through a separate telescope, both sets of optics being identical in order to minimise the systematics. Because PIXIE has several observation modes, an external calibrator (cooled blackbody) could be located in front of one of the apertures at any one time. In the "spectral distortion" mode, for which the absolute spectrum needs to be measured, one input is directed towards the sky, the second towards a blackbody whose temperature is set to the CMB one, 2.726 K, so that a differential measurement can be achieved. This process can be alternated in order to cancel any asymmetry in both optical systems. After going through a set of polarisers, each of the two outputs is focused on a dual-polarisation multimoded bolometric detector, so that each of the four detectors measures an interference fringe pattern between orthogonal linear polarizations from the two input beams. In order to limit the photon noise, the whole instrument is cooled to about 3 K, the detectors being at sub-K temperature.

3.1. *Balloon specificity*

The previous studies for the instrument concept were performed on the basis of a space mission. For the case of a balloon project, the conditions, the requirements and therefore the instrument concept will be different. Stratospheric balloon projects might be considered to be in near-space conditions, but there are still many differences, some of these being the potential flight time, access to the sky, thermal environment, etc...

A few will have strong impacts on the payload concept. First the residual atmosphere at an altitude of about 40 km will be of the order of 3 mbar with a temperature of the order of 270 K. This will not only lead to an additional photon noise contribution, but also to the necessity of hosting the overall instrument inside a dewar in order to keep it at a temperature of the order of 3 K under vacuum. Even with a small pressure difference, the dewar will then need to have a window through which the telescope will point towards the sky (see figure 3 right). The

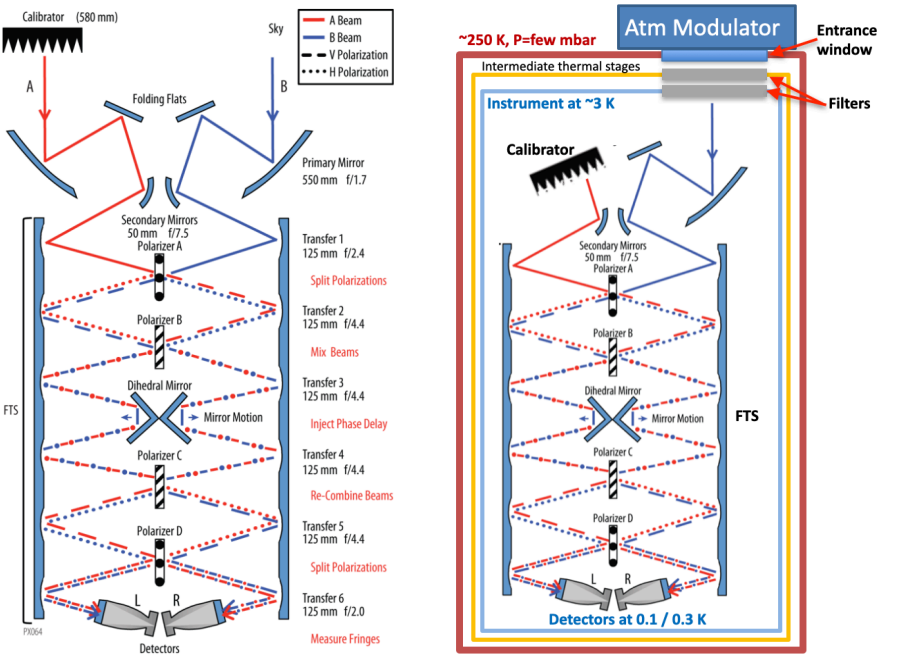


Fig. 3. Left: Original instrument concept from the PIXIE space mission proposal.[3] Right: BISOU instrument scheme.

outside dewar shell being at ambient temperature, intermediate thermal stages will be necessary with thermal filters in order to limit the thermal background load and potential straylight.

This is also preventing the calibrator to be outside the dewar and will then have to be located inside in order to keep it at a very steady temperature of about 3 K.

Due to the mass, power and dimensions constraints of a typical balloon payload, it is unlikely that two sets of optics can be used with a telescope primary diameter of 35 to 40 cm allowing for a multi-moded beam FWHM of about 1.5 deg.

3.2. *Gondola, dewar and cooling chain*

Assuming that the balloon flight will be provided by CNES, our study is based on the use of the CARMEN gondola[9] design that has been used lately for the PILOT balloon project[10] for instance, allowing for a maximum all-included payload mass of 750 kg (including ballast). While CNES has so far flown these types of payload for a typical 35-hour flight duration, it is planned that a first 5-day test-flight will happen in 2022. We will therefore base our sensitivity calculations on the assumption that this type of flight will be available in the future.

The whole instrument needs to be cooled to about 3 K, where the detectors only will be cooled to a sub-K temperature. A trade-off is being made in order to

define the optimum one, but it is highly likely that 300 mK will be low enough (section 4.1) for the focal plane, allowing for the use of a ^{3}He sorption cooler. Due to mass and power limitations, mechanical coolers for balloon platforms are not yet mature enough, even if some developments are being investigated, most notably by NASA. We chose a proven cooling solution using liquid helium to cool the overall instrument, between 4 and 2 K depending on the bath pressure, using the natural low pressure at high altitude. Several intermediate thermal stages and shields cooled with the helium vapour retrieved from the bath will allow for heat load reduction on the 3 K stage. Typical intermediate temperature stages are 150 K, 80 K and 20 K.

3.3. *Atmosphere*

As already mentioned, the residual atmosphere at 40 km altitude will still create a large photon noise contribution with respect to the extremely faint signal that we are trying to observe. For instance, atmospheric effects at ground level and even at balloon altitudes are shown in Masi et al.[11] with respect to spectral distortion signals. The signal level due to the atmosphere is still 2 to 3 orders of magnitude higher that the non-relativistic y-distortion signal depending on the frequency. In theory, assuming a reliable model, this contribution could be removed.

More important, will be the variations of the atmosphere with altitude, observation elevation and time. This cannot be modelled and will therefore need an atmosphere modulation strategy at a high enough frequency to be compatible with the variation timescale. Such a modulator is already used for COSMO,[11] but for BISOU, another modulator, such as a cold internal beam stirrer is also being considered.

For the time being, at that stage of the study, the atmosphere is not yet considered in our sensitivity and systematic effects calculations.

3.4. *The instrument*

Following the conditions of a balloon project (section 3.1), several points had to be adapted and trade-offs have already been made. With emphasis on the spectral distortions, these measurements will always be performed against the calibrator. Due to the fact that we are using only one telescope to save mass (with therefore an unbalanced optical system), the calibrator will be kept fixed in front of one of the FTS inputs for simplicity and risk mitigation.

With the help of an instrument model being developed, and assuming some initial parameters such as telescope diameter, flight duration and observation efficiency, or again spectral resolution, calculations presented in section 4 are showing that the sensitivity is strongly dependant on three key parameters: the temperature and emissivity of the window, and the maximum frequency of observation as the photon noise is increasing with frequency. The implications of these issues will be

discussed in the following sections. However, we see straight away that the dewar window, the warmest optical element, needs to have a very low emissivity, a temperature as low as possible and be as thin as possible to minimise the systematics.

In order to use a very thin window, a valve at the entrance of the dewar, opening only when the pressure difference is small (at a certain altitude), could be used. Such a technique was used for the Archeops[12] balloon project for instance.

The emissivity of the window could be reduced by the application of a copper/gold grid as was done on the ArTéMiS[13] ground based camera.

Finally while the dewar will be at a temperature ranging from 300 to 270 K, the window could have a mount thermally isolated from the structure and actively cooled with the vapour coming from the helium bath. Such a development will be studied with a dedicated R&D programme. This should bring the window temperature below 200 K.

4. Preliminary sensitivity estimates

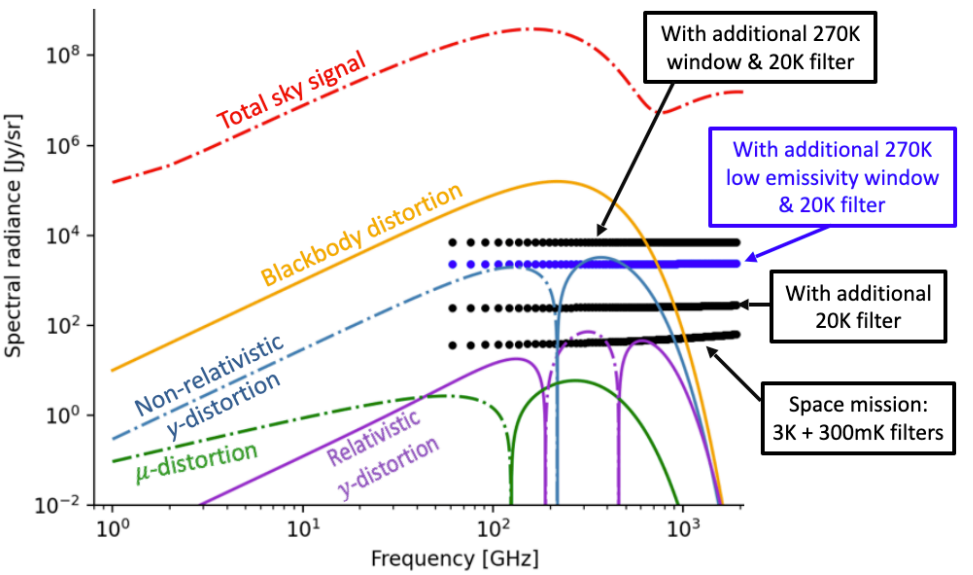


Fig. 4. Estimated instrument sensitivity for different concepts, compared to various spectral distortion signals. All calculations are assuming a 5-day flight. The highest sensitivity assumes a space mission, while the lowest sensitivity is for a balloon flight for which the noise is limited by the emission of the dewar window with pessimistic characteristics.

4.1. *Photometric model*

A preliminary photometric model has been developed, taking into account only the dewar window and a minimum of two spectral filters on the lowest temperature

stages (supposed to be at 20 and 3 K and with a top-hat transmission profile) for the optical input. The emission of all the components in the optical path is modeled as blackbody emission (at the temperature of the component) multiplied by the emissivity of that component at the specified temperature. The load on the detector is estimated by adding the power contributions of the optical components as well as the contribution from the sky, using for each:

$$P(\nu, T) = \int_{\nu_{min}}^{\nu_{max}} eff(\nu) A\Omega(\nu) \epsilon(\nu, T) B(\nu, T) d\nu$$

where $A\Omega$ is the throughput of the multi-moded optics, eff the transmission/efficiency of the optical system, B the blackbody function, and for this first version of our model, the emissivity ϵ is taken as a constant for each component.

We integrate the power received by the detector over the full range of frequencies. In order to assume the worst case scenario for the detector load, when the optical path difference of the Fourier Transform Spectrometer is null, the power is integrated over the full frequency range that we will observe.

The total Noise Equivalent Power (NEP) is then calculated by adding in quadrature the photon NEP from the signal arriving on the detector and the detector NEP which is assumed to be about four times lower than the photon NEP. For the time being, we assume that the other contributions to the total NEP are small in comparison to the photon one. Depending on the assumptions made, for the temperature and emissivity of the optical components for instance, the total NEP is of the order of a few 10^{-16} W Hz$^{-1/2}$ for a space mission to about 10^{-14} W Hz$^{-1/2}$ for a balloon configuration, high enough to allow for the use of detectors at a temperature of about 300 mK.

4.2. *Results*

Following classic calculations,[3] from the total NEP, the detected noise for a fixed integration time τ is given by:

$$\delta P = \frac{NEP_{total}}{\sqrt{\tau/2}}$$

where τ is taken to be 90 hours, corresponding to a flight duration of 5 days with 75% observation efficiency. From this, the noise at the detector may in turn be referred to the specific intensity, leading to an equivalent instrumental sensitivity that can be compared to spectral emissions. These results are shown on figure 4 for different instrument configurations. For all these configurations, a FTS spectral resolution of 15 GHz, a lowest frequency ν_{min}=60 GHz, and a linear tapered filter to limit the high frequency contribution above 600 GHz have been used.

Figure 4 shows the impact of the window on the sensitivity as it will be the warmest element on the optical chain. Starting from a "space configuration" where

the warmest element might be around 3 K (filter at 3 K and 300/100 mK with 0.1% emissivity each), which will be the most sensitive configuration, followed by a configuration with an added filter at T=20 K, then the same configuration with the window at T=270 K added, this with 2 emissivity assumptions for the window, one with ϵ=0.1% like all other filters, and a lower one with $\epsilon = 3 \times 10^{-4}$ (in blue), something that we hope to achieve.

We then conclude that in order to have a first detection of the y-parameter with a decent signal-to-noise ratio, the window will need to be cooled further as previously discussed.

Another parameter which is also crucial, is the maximum observation frequency ν_{max}. All the photon contributions being higher with increasing frequency, the overall sensitivity will drop with increasing ν_{max}. On the other hand, in order to retrieve the "easy" science goals, such as the y-parameter spectral distortions, ΔT_{CMB} and T_{CIB} for the measurement of the Cosmic Infrared Background, measurements have to go to high frequencies. Therefore a trade-off needs to be reached in order to optimise the S/N for each parameter.

This work has started and will be presented in future publications. However, we can already say that, not taking into account the atmosphere, S/N ratios of 5 for the y-distortions, about 30 for ΔT_{CMB} and about 10 for T_{CIB}, can be reached if ν_{max} were to be 1200 GHz.

Obviously a more detailed study will need to take place for a better optimisation, including more sophisticated models and better parameter knowledge.

5. On-going and future work

Aside from more detailed model of the instrument and sensitivity calculations, several points need to be addressed by the end of the Phase 0 study.

The next important hurdle to cross is understanding how to deal with the atmosphere. More accurate models and data for high altitude are being gathered within the consortium but as already stated, a modulator seems inevitable. Concept work and the study of the associated effects of this modulator will be paramount.

Other points to be developed (not exhaustive) can be listed as:

- Observation plan and scanning strategy;
- Calibration strategy, together with the accurate characterisation and control of the calibrator;
- Study on how to increase the sensitivity, by increasing the number of detectors for instance, or having two channels;
- A first cut to the study of the systematic effects, in order to answer at least if the unbalanced optical system between the two inputs will not be a limiting factor;
- Updated science case.

6. Conclusion

So far, assuming that the atmospheric problem can be solved, it seems that according to our preliminary estimates, such a balloon borne instrument will be able to have valid scientific outputs, and not be limited to a technological demonstrator for future space missions. Namely, a measurement of the CIB, a better constraint on ΔT_{CMB}, and more importantly a first 5-σ measurement of the y-distortions if not better.

If these results are confirmed by the end of Phase 0, a proposal to CNES and other funding bodies will be submitted to start a more detailed Phase A study, for a first test-flight on the horizon 2026.

Acknowledgments

The authors acknowledge the support of the French space agency, Centre National d'Etudes Spatiales (CNES).

Kavli IPMU is supported by World Premier International Research Center Initiative (WPI), MEXT, Japan.

References

1. J. C. Mather *et al.*, A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite, *ApJL* **354**, p. L37 (May 1990).
2. D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer and E. L. Wright, The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set, *ApJ* **473**, p. 576 (December 1996).
3. A. Kogut, D. J. Fixsen, D. T. Chuss, J. Dotson, E. Dwek, M. Halpern, G. F. Hinshaw, S. M. Meyer, S. H. Moseley, M. D. Seiffert, D. N. Spergel and E. J. Wollack, The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations, *JCAP* **2011**, p. 025 (July 2011).
4. J. Chluba, M. H. Abitbol, N. Aghanim *et al.*, New horizons in cosmology with spectral distortions of the cosmic microwave background, *Experimental Astronomy* (May 2021).
5. J. Delabrouille, M. H. Abitbol, N. Aghanim *et al.*, Microwave spectro-polarimetry of matter and radiation across space and time, *Experimental Astronomy* (July 2021).
6. M. H. Abitbol, J. Chluba, J. C. Hill and B. R. Johnson, Prospects for measuring cosmic microwave background spectral distortions in the presence of foregrounds, *MNRAS* **471**, 1126 (October 2017).
7. J. C. Hill, N. Battaglia, J. Chluba, S. Ferraro, E. Schaan and D. N. Spergel, Taking the Universe's Temperature with Spectral Distortions of the Cosmic Microwave Background, *Physical Review Letters* **115**, p. 261301 (December 2015).
8. Planck Collaboration, Planck 2015 results. X. Diffuse component separation: Foreground maps, *A&A* **594**, p. A10 (September 2016).
9. CNES, *CNES Balloons capabilities* `https://www.hemera-h2020.eu/facilities-2/cnes-balloons/`.
10. A. Mangilli *et al.*, Inflight performance of the PILOT balloon-borne experiment, *Experimental Astronomy* **48**, 265 (December 2019).
11. S. Masi *et al.*, The COSmic Monopole Observer (COSMO), *These proceedings* (2021).

12. A. Benoît *et al.*, Archeops: a high resolution, large sky coverage balloon experiment for mapping cosmic microwave background anisotropies, *Astroparticle Physics* **17**, 101 (May 2002).
13. M. Talvard *et al.*, ArTeMiS: filled bolometer arrays for next generation sub-mm telescopes, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, eds. J. Zmuidzinas, W. S. Holland, S. Withington and W. D. Duncan, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6275 (June 2006).

Cosmic microwave background spectral distortions constraints on decaying dark matter particles and axion-like particles using *COBE/FIRAS* and *EDGES*

Boris Bolliet

Department of Physics, Columbia University, New York, NY 10027, USA
E-mail: boris.bolliet@gmail.com

We present new cosmic microwave background (CMB) spectral distortions constraints on low mass (10^{-10}-10^{4} eV) dark matter particles that decay into photons. The constraints are first presented in a model-independent manner and then applied to axions and decaying excited states. For the first time, we place constraints using the full distortion spectra compared to the *COBE/FIRAS* and *EDGES* measurements. This proceeding summarizes results obtained with Jens Chluba and Richard Battye, and published in MNRAS 507 (2021)[a].

Keywords: Cosmology: cosmic microwave background – theory – observations; Dark Matter.

1. Introduction

Spectral distortions are a powerful probe of the thermal history of the universe.[1] They can constrain dark matter candidates with coupling to the electro-magnetic sector, since these typically inject energy into the photon-baryon plasma, generating departure of the CMB spectrum from the black body law. In the past decade, a number of studies have derived constraints on the dark matter cross-section, the mass lifetime and abundance of the decaying or annihilating dark matter particles[e.g.].[2–5] Spectral distortions are often characterized by the chemical potential μ and the Compton-y parameter, which are directly proportional to the injected energy and compared with the COBE/FIRAS bounds, $\mu < 9 \times 10^{-5}$ and $y < 1.5 \times 10^{-5}$ (95%CL). However, the μ and y distortions are generally not sufficient to characterize the full distortions. This is the case when the injected photons do not have time to undergo many scatterings, i.e., *comptonize*. For instance, for distortions from decaying particles with a lifetime comparable to the age of the universe, there may be a y distortion but also a peak in intensity located at the rest mass energy of the decaying particle. Here, we present the first constraints that use such features in the distortion spectra.

This is a challenging task: it requires to solve the Boltzmann equation for photons, simultaneously with the time evolution of the electron temperature and the ionisation history. To do so, we use the code `Cosmotherm` as well as packages recently

[a] https://arxiv.org/abs/2012.07292

developed at the University of Manchester, namely `CSpack` for computing energy exchange from Compton scattering[6] and `BRpack` to compute Bremsstrahlung emission coefficients.[7] We use a standard reionization modeling[8] and a refined treatment of helium and hydrogen recombination initially based on Recfast.[9,10] To run these codes we developed a python package, named `specdist`[b], so that we could compute in parallel many distortion spectra and the corresponding ionisation histories.

2. Photon injection from decaying particles

We focused on a dark matter particle that decays into two photons, each carrying half of the rest mass energy of the decaying particle. We also considered dark matter excited states where one photon carries the energy difference between the excited and ground state of the decaying particle. The time evolution of the dark matter particle population is described by the exponential decay law with a decay rate Γ that is the inverse of the lifetime. The decay rate determines the time of energy injection and consequently the type of distortion. For short lifetimes we expect a μ-distortion until redshift $z \approx 5 \times 10^4$, and a y-distortion at lower redshift. For photon injection happening after recombination, we expect a more complicated shape due to photons that do not have time to scatter and instead simply redshift as the universe expands.

For the spectrum of the injected photons we used a narrow gaussian that approximates a delta function centered at the rest-mass energy of the decaying particle. The normalization of the spectrum is directly proportional to the abundance of decaying particles which we can write in term of the dark matter fraction $f_{\rm dm}$, that we constrain. To do this, once the distortion is computed we proceed with the exercise of comparing it with CMB frequency spectra measurements such as the measurements from *COBE/FIRAS*[11] and the *EDGES* brightness temperature at 78 MHz.[12]

3. Complementary constraints from ionization histories

We also use complementary constraints from the perturbed ionisation histories that affect the CMB temperature and polarization anisotropy power spectra. The left panel of figure 1 shows several ionisation histories for a particle that decays during the end of recombination, with different masses labelled in dimensionless unit. For some of these we found an interesting time evolution with periodic spikes when we switched on collisional ionisation.

In principle these ionisation histories can be fed to Boltzmann codes such as `class`[16] or `camb`[17] to compute the CMB anisotropy power spectra and extract constraints using, e.g., the *Planck* likelihood.[18] A faster method is to use the projection methods developed by Luke Hart and Jens Chluba:[13] the ionisation histories are

[b] https://github.com/CMBSPEC/specdist

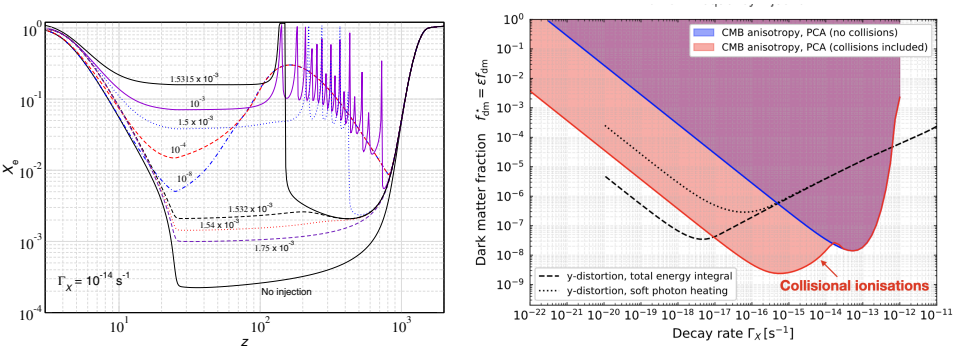


Fig. 1. **Left:** Non-linear responses in the ionization history around critical injection frequencies for $\Gamma = 10^{-14}\,\mathrm{s}^{-1}$ and $\Delta\rho_\gamma/\rho_\gamma = 3\times10^{-5}$. For $10^{-3} \lesssim x_{\rm inj,0} = h\nu_{\rm inj,0}/k_{\rm B}T_{\rm CMB} \lesssim 1.53\times10^{-3}$ the free electron fraction shows intermittent bursts due to photon injection, free-free absorption and collisional processes. **Right:** Derived 95% CL dark matter fraction limit, $f^*_{\rm dm} = \epsilon f_{\rm dm}$, for a range of lifetimes and $x_{\rm inj,0} \lesssim 10^{-8}$. The CMB anisotropy limits were obtained with the projection method,[13] while the other two are derived assuming $4y \lesssim 6\times10^{-5}$ (95% CL) from *COBE/FIRAS*. The dotted curve corresponds to soft photon heating, and the solid dashed line to the total energy integral with the y-era visibility function. For the CMB anisotropy constraints we considered a case where collisional ionisations are not included (blue region) and a case where these are included (red region). As can be seen, atomic collisions are driving post-recombination decay constraints ($\Gamma \lesssim 10^{-14}\,\mathrm{s}^{-1}$). We emphasize that these constraints are for very low mass decaying particles ($m \ll 1$eV) where all the injected energy in the form of photons is converted into heat, in contrast with scenario where the energy is deposited and leads to direct ionisation.[14, 15]

projected onto eigenmodes obtained from a principal component analysis, allowing to derive constraints with a simple root-finding routine.

As shown in the right panel of figure 1, benefiting from the exquisite measurements of *Planck* and *WMAP*, constraints from CMB anisotropy are the strongest for dark matter candidates that decay near recombination, when the CMB is emitted. Nonetheless, *COBE/FIRAS* spectrometer measurements actually beat imager constraints for particles that decay far from recombination, since spectral distortions probe nearly the entire thermal history. Note that the figure illustrates this point with the y-distortion constraints only, while we can now make use of the full distortion spectra.

4. Constraints from the full distortion spectra

In a forthcoming work we will be using machine learning methods in order to build an optimal spectral distortion emulator. For this work, we computed a vast library of distortions so that at any point in the parameter space spanned by the lifetime, mass and abundance of the decaying particle we can interpolate the precomputed spectra and compute the likelihood values given the *COBE/FIRAS* and *EDGES* measurements. A subset of the distortion spectra is shown in figure 2. Our model independent constraints are shown on figure 3. On the left panel, we see the maximally allowed dark matter fraction as a function of the injection energy in eV.

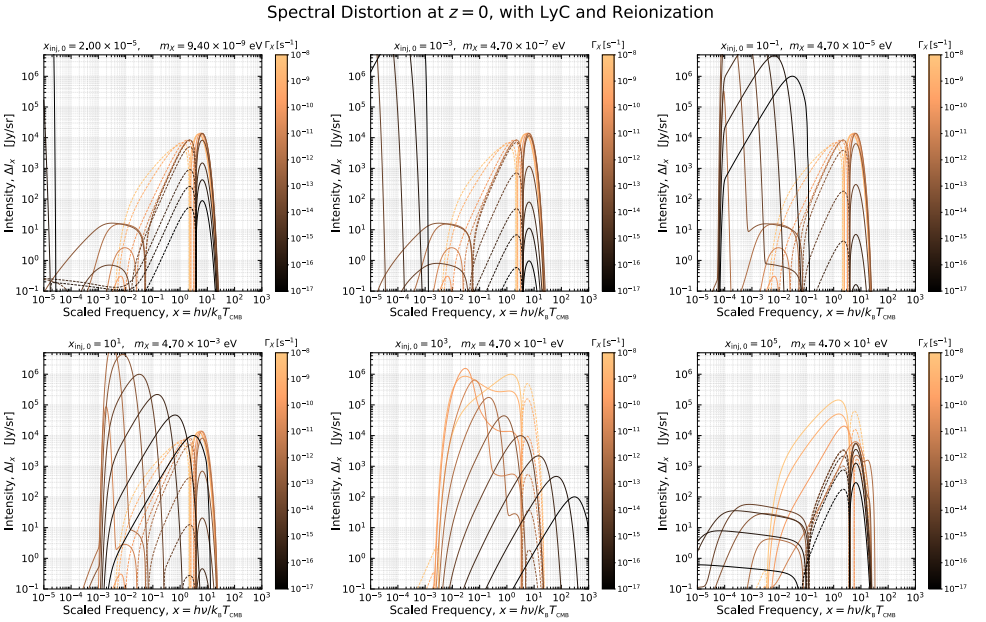


Fig. 2. Solutions for the spectral distortion created by photon injection from decaying particles for several masses, m, and lifetimes, Γ, including simple modelling for reionization in addition to H I, He I and He II photo-ionization (LyC stands for Lyman continuum absorption). All spectra are normalized such that $\Delta\rho/\rho|_{\rm inj} = 3 \times 10^{-5}$. Negative parts of the signal are shown as dashed lines.

The thick line correspond to a a short lifetime, for a particle that decays at early time, towards the end of the μ-era. It is essentially the constraints that one would obtain with a simple μ-estimate calculation by integrating the injected energy over time. There are two differences though: a region near 0.1-10 eV where the chemical potential is actually negative, and a small bump near 1meV associated with photons that have redshifted down to 78MHz where we impose that the CMB brightness temperature should not exceed the *EDGES* measurements. For longer lifetimes, the straight horizontal line constraints corresponding to μ or y estimates is nearly absent: the constraints are driven by the extra redshifted photons that do not scatter efficiently and end up in the *COBE/FIRAS* or *EDGES* frequency range. For decays happening near or after recombination (dotted and dashed-dotted lines respectively), we see strong constraints on the dark matter fraction for energy higher than the ionisation threshold of Hydrogen and Helium (the vertical dashed line on the figure). These constraints are driven by the y-distortion caused by heating from Lyman continuum absorption, and analogous processes for Helium. On the right panel of figure 3, we show the full constraints in the lifetime, injection energy plane with darker colors corresponding to more excluded regions. One piece of modeling that is absent in the constraints of figure 3 are quantum effects associated with the

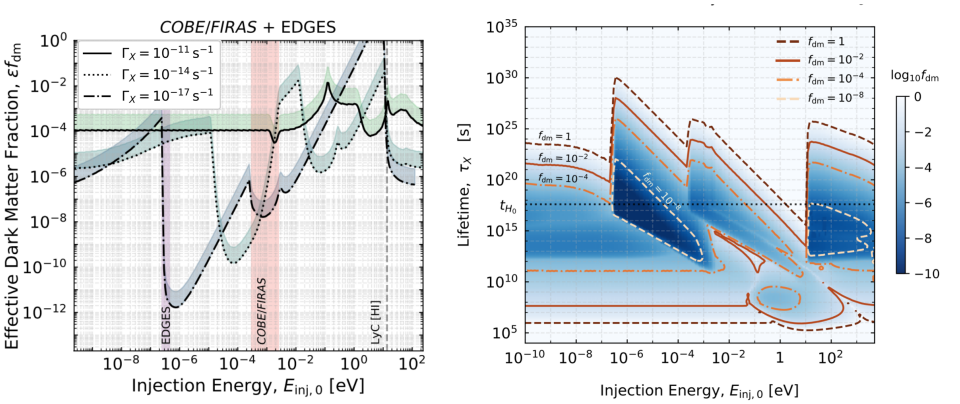


Fig. 3. **Left:** Constraints on the effective DM fraction for decaying particle scenarios (95% CL) for various combinations of data from *COBE/FIRAS*, *EDGES* and *Planck*. The upper panels show the constraints for lifetimes $\Gamma \leq 10^{-11}\,{\rm s}^{-1}$ and various combinations of datasets when assuming decay to occur in vacuum. The lower panels show several examples when including stimulated decay. See main text for discussion. **Right:** Constraints (95% CL) on decaying particle models when subsequently including data from *COBE/FIRAS*, *EDGES* and *Planck*. The addition of *EDGES* affects the constraints at $E_{\rm inj} \simeq 3 \times 10^{-7} - 3 \times 10^{-4}\,{\rm eV}$, while the addition of *Planck* has the largest effect at high energies above $E_{\rm inj} \simeq 13.6{\rm eV}$.

bosonic or fermionic nature of the decaying particle. We treated the case of bosons, for which the emission is stimulated by the ambient CMB photon bath.[19] The effect of stimulated decay is relatively simple: the decay rate becomes frequency dependent with a term that includes the photon occupation number, evaluated at the injection frequency. One has

$$\Gamma_{\rm stim} \approx [1 + 2n(x_{\rm inj})]\,\Gamma \tag{1}$$

where $\Gamma_{\rm stim}$ is the decay rate in the stimulated case and Γ the would be decay rate without stimulated emission, and where we use the Planck black body law to compute $n(x_{\rm inj})$ with $x_{\rm inj} \equiv h\nu_{\rm inj}/k_{\rm B}T_{\rm CMB}$. For large injection frequencies $\nu_{\rm inj}$, both the stimulated decay rate and standard decay rate are equal because $n(x_{\rm inj}) \approx 0$, but for low injection frequencies the difference between $\Gamma_{\rm stim}$ and Γ is proportional to $1/\nu_{\rm inj}$ and is larger at higher redshifts. For instance, for a decay with lifetime roughly the age of the universe today, we expect a small y-distortion and an emission peak centered around the injection frequency. With stimulated decay accounted for, such decaying particle would already have completely decayed a long time ago, in the μ-era, generating a standard μ-distortion. This also implies that with stimulated decay, in the case of low $\nu_{\rm inj}$, one needs much more particles to yield the same amount of energy injection for a fixed lifetime. With this extra piece of modelling we can revisit the constraints of figure 3.

Constraints with stimulated decay accounted for are presented in figure 4 and are the main model independent results of our work. By *model independent*, we mean that at this stage we have not chosen a specific type of particles (other than

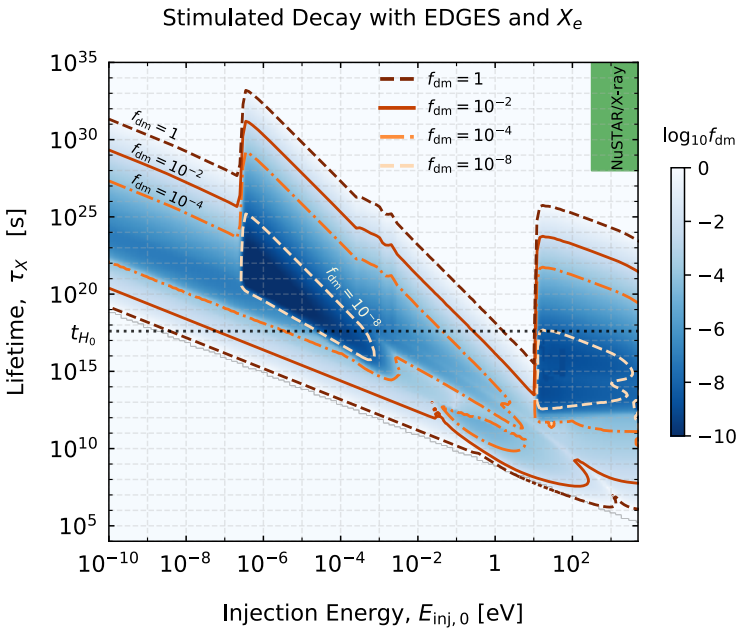


Fig. 4. Same as in figure 3 but with stimulated decay accounted for. The dominant effect of simulated decay is to shift and compress the limits along the τ_X axis for low injection energies/frequencies. We also mark the lifetime corresponding to the age of the Universe ($\tau_X \equiv t_{H_0} \simeq 4.4 \times 10^{17}$ s). For comparison, we have added the current level of constraints of sterile neutrino decay in the X-ray frequency range, namely the lifetime has to be longer than 10^{28} s for masses between $5 - 50$keV (green area labeled 'NuSTAR/X-ray'), see Fig 5 of Roach et al.[20]

being bosons). At high injection frequency, namely above 10 eV, the contours are unaffected by the stimulated decay, as expected. But at lower frequency we see an upward tilt of the constraint regions compared to the vacuum decay case of figure 3, towards longer lifetimes, consistent with the discussion above. For comparison, on the top right corner of figure 4 we have added a green exclusion region which comes from the non-detection of X-ray lines expected in radiative decay of sterile neutrinos.[20]

5. Constraints on axion-like particles

We translated our constraints to axion-like particles (ALPs) models. ALPs are pseudo Nambu-Goldston bosons proposed as dark matter candidates and motivated by solutions to the strong CP problem.[21] For these types of particles there is a fundamental relationship between the mass and the decay rate, parameterized the coupling constant $g_{a\gamma\gamma}$.[22] Then, we can replace the x-axis of figure 4 by the ALPs mass, and map the y-axis, which was the lifetime, to the coupling constant. Constraints on ALPs are shown in figure 5. The top right corner of the plot is not constrained

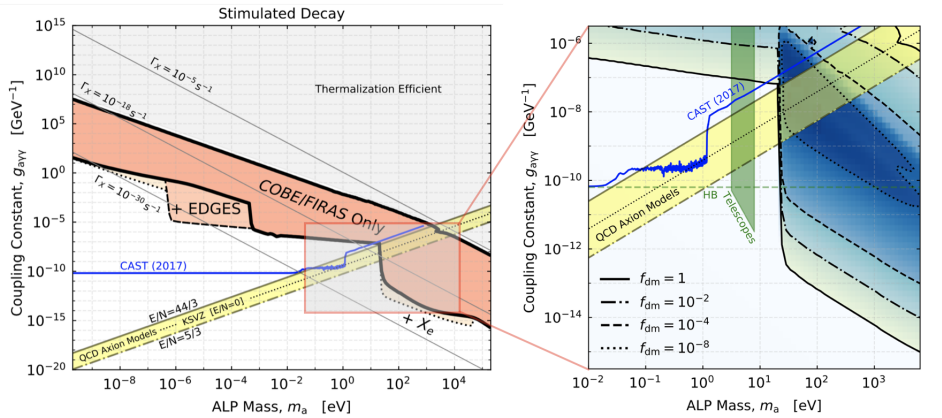


Fig. 5. **Left:** Cumulative *COBE/FIRAS*, *EDGES* and *Planck* constraints (95% CL) on ALPs models and $f_{\rm dm} = 1$. These constraints include blackbody-induced stimulated decay. For reference, we also show that constraint from *CAST*[23] and the physically-motivated region valid for the KSVZ axion model over a range of anomaly coefficients E/N (yellow region). Spectral distortions provide competitive constraints at axion masses of $m_a c^2 \gtrsim 27\,$eV. **Right:** Same as left panel but zoomed in on the high-mass end and for various values of $f_{\rm dm}$. Stimulated decay is unimportant in this domain. Axions with masses $m_a c^2 \gtrsim 27$eV are heavily constrained even if they only constitute a small fraction of the DM. For comparison we have added constraints from Horizontal Branch Stars (green dashed line labeled 'HB') and optical telescopes searches focused on Abell clusters (green shaded area).[24,25] We refer to Fig 91.1 of PDG18[26] for further details on these optical constraints and to Regis et al[27] for the most recent updates.

because it corresponds to very short lifetimes, and all injected energy and photons are thermalized thanks to efficient double Compton and Bremsstrahlung scatterings. The upper part of the pink band is constrained by the μ and y-distortion limits, and the lower part with the peaks correspond to constraints driven by the full spectra. The thick line shows the constraints using *COBE/FIRAS* residuals and the dashed line is the extended constraints when we add the 78 MHz *EDGES* data. The dotted lines are the constraints when we add CMB anisotropy from perturbed recombination. On the right we show a zoomed in region of the plot and we also show the constraints assuming ALPs make up only a fraction of the dark matter. Most of this part of the constraint plot is determined by the y-distortion due to heating from Lyman continuum absorption and it shows that spectral distortion constraints are competitive against other astrophysical constraints on axions, e.g., constraints from *Horizontal Branch* (HB) stars (the green dashed line) and *optical telescopes* (the green area).

6. Conclusion

Here we have presented the first constraints on photon injection processes using the full distortion spectra and ionisation histories. Our work shows that we are still harvesting insights into dark matter from the 30 years old *COBE/FIRAS* spectrometer

data. There is little doubt that the next spectrometer mission will be game changing for dark matter models. Our next steps will be to build a spectral distortion emulator to alleviate the computationally expensive calculations of distortion spectra and ionisation histories, and provide a forecast analysis for a next spectrometer mission such as Voyage 2050.[28]

References

1. J. Chluba and R. A. Sunyaev, The evolution of cmb spectral distortions in the early universe, *Monthly Notices of the Royal Astronomical Society* **419**, p. 1294–1314 (Nov 2011).
2. J. Chluba, Green's function of the cosmological thermalization problem – ii. effect of photon injection and constraints, *Monthly Notices of the Royal Astronomical Society* **454**, p. 4182–4196 (Oct 2015).
3. Y. Ali-Haïmoud, J. Chluba and M. Kamionkowski, Constraints on dark matter interactions with standard model particles from cosmic microwave background spectral distortions, *Physical Review Letters* **115** (Aug 2015).
4. V. Poulin, J. Lesgourgues and P. D. Serpico, Cosmological constraints on exotic injection of electromagnetic energy, *Journal of Cosmology and Astroparticle Physics* **2017**, p. 043–043 (Mar 2017).
5. Y. Ali-Haïmoud, Testing dark matter interactions with cmb spectral distortions, *Physical Review D* **103** (Feb 2021).
6. A. Sarkar, J. Chluba and E. Lee, Dissecting the compton scattering kernel i: Isotropic media, *Monthly Notices of the Royal Astronomical Society* **490**, p. 3705–3726 (Oct 2019).
7. J. Chluba, A. Ravenni and B. Bolliet, Improved calculations of electron–ion bremsstrahlung gaunt factors for astrophysical applications, *Monthly Notices of the Royal Astronomical Society* **492**, p. 177–194 (Dec 2019).
8. V. Poulin, P. D. Serpico and J. Lesgourgues, Dark matter annihilations in halos and high-redshift sources of reionization of the universe, *Journal of Cosmology and Astroparticle Physics* **2015**, p. 041–041 (Dec 2015).
9. J. Chluba, Could the cosmological recombination spectrum help us understand annihilating dark matter?, *Monthly Notices of the Royal Astronomical Society* **402**, p. 1195–1207 (Feb 2010).
10. J. Chluba and Y. Ali-Haïmoud, Cosmospec: fast and detailed computation of the cosmological recombination radiation from hydrogen and helium, *Monthly Notices of the Royal Astronomical Society* **456**, p. 3494–3508 (Jan 2016).
11. D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer and E. L. Wright, The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set, *apj* **473**, p. 576 (December 1996).
12. J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen and N. Mahesh, An absorption profile centred at 78 megahertz in the sky-averaged spectrum, *Nature* **555**, p. 67–70 (Mar 2018).
13. L. Hart and J. Chluba, Improved model-independent constraints on the recombination era and development of a direct projection method (12 2019).
14. V. Poulin, J. Lesgourgues and P. D. Serpico, Cosmological constraints on exotic injection of electromagnetic energy, *jcap* **2017**, p. 043 (March 2017).
15. M. Lucca, N. Schöneberg, D. C. Hooper, J. Lesgourgues and J. Chluba, The synergy between CMB spectral distortions and anisotropies, *jcap* **2020**, p. 026 (February 2020).

16. D. Blas, J. Lesgourgues and T. Tram, The cosmic linear anisotropy solving system (class). part ii: Approximation schemes, *Journal of Cosmology and Astroparticle Physics* **2011**, p. 034–034 (Jul 2011).
17. A. Lewis, A. Challinor and A. Lasenby, Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models, *The Astrophysical Journal* **538**, p. 473–476 (Aug 2000).
18. N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo and et al., Planck 2018 results, *Astronomy & Astrophysics* **641**, p. A5 (Sep 2020).
19. A. Caputo, M. Regis, M. Taoso and S. J. Witte, Detecting the stimulated decay of axions at radio frequencies, *Journal of Cosmology and Astroparticle Physics* **2019**, p. 027–027 (Mar 2019).
20. B. M. Roach, K. C. Ng, K. Perez, J. F. Beacom, S. Horiuchi, R. Krivonos and D. R. Wik, Nustar tests of sterile-neutrino dark matter: New galactic bulge observations and combined impact, *Physical Review D* **101** (May 2020).
21. D. J. Marsh, Axion cosmology, *Physics Reports* **643**, p. 1–79 (Jul 2016).
22. G. G. Raffelt, Astrophysical axion bounds, *Axions* , p. 51–71.
23. CAST Collaboration, New CAST Limit on the Axion-Photon Interaction, *Nature Phys.* **13**, 584 (2017).
24. M. A. Bershady, M. T. Ressell and M. S. Turner, Telescope search for a 3-ev to 8-ev axion, *Phys. Rev. Lett.* **66**, 1398 (Mar 1991).
25. D. Grin, G. Covone, J.-P. Kneib, M. Kamionkowski, A. Blain and E. Jullo, Telescope search for decaying relic axions, *Physical Review D* **75** (May 2007).
26. M. Tanabashi and P. D. Group, Review of particle physics, *Phys. Rev. D* **98**, p. 030001 (Aug 2018).
27. M. Regis, M. Taoso, D. Vaz, J. Brinchmann, S. L. Zoutendijk, N. F. Bouché and M. Steinmetz, Searching for light in the darkness: Bounds on alp dark matter with the optical muse-faint survey, *Physics Letters B* **814**, p. 136075 (Mar 2021).
28. J. Chluba, M. H. Abitbol, N. Aghanim, Y. Ali-Haïmoud, M. Alvarez, K. Basu, B. Bolliet, C. Burigana, P. de Bernardis, J. Delabrouille and et al., New horizons in cosmology with spectral distortions of the cosmic microwave background, *Experimental Astronomy* **51**, p. 1515–1554 (May 2021).

The COSmic Monopole Observer (COSMO)

S. Masi*, E. Battistelli, P. de Bernardis, F. Columbro, A. Coppolecchia, G. D'Alessandro, M. De Petris, L. Lamagna, E. Marchitelli, L. Mele, A. Paiella, F. Piacentini, G. Pisano

Physics Department, Sapienza University of Rome
and INFN Sezione di Roma, P.le A. Moro 2, Rome, 00185, Italy
**E-mail: silvia.masi@roma1.infn.it*
https://cosmo.roma1.infn.it

M. Bersanelli, C. Franceschet, E. Manzan, D. Mennella, S. Realini

Physics Department, University of Milan
and INFN Sezione di Milano, Via Celoria 16, 20133 Milan, Italy

S. Cibella, F. Martini, G. Pettinari

IFN-CNR
Via Cineto Romano 42, 00156 Roma, Italy

G. Coppi, M. Gervasi, A. Limonta, M. Zannoni

Physics Department, University of Milan Bicocca
and INFN Sezione di Milano Bicocca, P.zza delle Scienze 1, 20126 Milan, Italy

L. Piccirillo

Department of Physics and Astronomy, University of Manchester
Oxford road, M13 9PL Manchester, UK

C. Tucker

School of Physics and Astronomy, University of Cardiff
The Parade, CF24 3AA Cardiff, UK

The COSmic Monopole Observer (COSMO) is an experiment to measure low-level spectral distortions in the isotropic component of the Cosmic Microwave Background (CMB). Deviations from a pure blackbody spectrum are expected at low level (< 1 ppm) due to several astrophysical and cosmological phenomena, and promise to provide important independent information on the early and late phases of the universe. They have not been detected yet, due to the extreme accuracy required, the best upper limits being still those from the COBE-FIRAS mission. COSMO is based on a cryogenic differential Fourier Transform Spectrometer, measuring the spectral brightness difference between the sky and an accurate cryogenic blackbody. The first implementation of COSMO, funded by the Italian PRIN and PNRA programs, will operate from the Concordia station at Dome-C, in Antarctica, and will take advantage of a fast sky-dip technique to get rid of atmospheric emission and its fluctuations, separating them from the monopole component of the sky brightness. Here we describe the instrument design, its capabilities, the current status. We also discuss its subsequent implementation in a balloon-flight, which has been studied within the COSMOS program of the Italian Space Agency.

Keywords: Cosmic Microwave Background; Spectral Distortions; Fourier Transform Spectrometer.

1. Introduction

Spectral distortions of the CMB represent a research path orthogonal and synergic to CMB polarization studies. Their detection can shed light on cosmic reionization, on the physics of recombination, on dark matter and in general on any energy release in the primeval fireball, and even on the very early universe and cosmic inflation. See[1] and references therein for a comprehensive review. The expected deviations are very small, < 1 ppm, but have characteristic spectral signatures, allowing, in principle, their separation from overwhelming foreground emission and the CMB monopole itself. In figure 1 we compare the largest spectral distortions of the CMB in the mm-wave region to the most important foregrounds.

Current upper limits for spectral distortions are at a level of 0.01% of the peak brightness of the CMB, and were obtained more than 20 years ago by COBE – FIRAS.[2,3] Current technologies promise to achieve much higher sensitivity, the limiting factor remaining the control of instrumental systematic effects and the emission of astrophysical foregrounds. In order to avoid the very large background produced by the earth atmosphere at CMB frequencies, the final measurement must be carried out from space. To this purpose, several proposals for space-based experiments have been submitted, like the PIXIE,[4] CORE[5] and PRISTINE missions. Meanwhile,

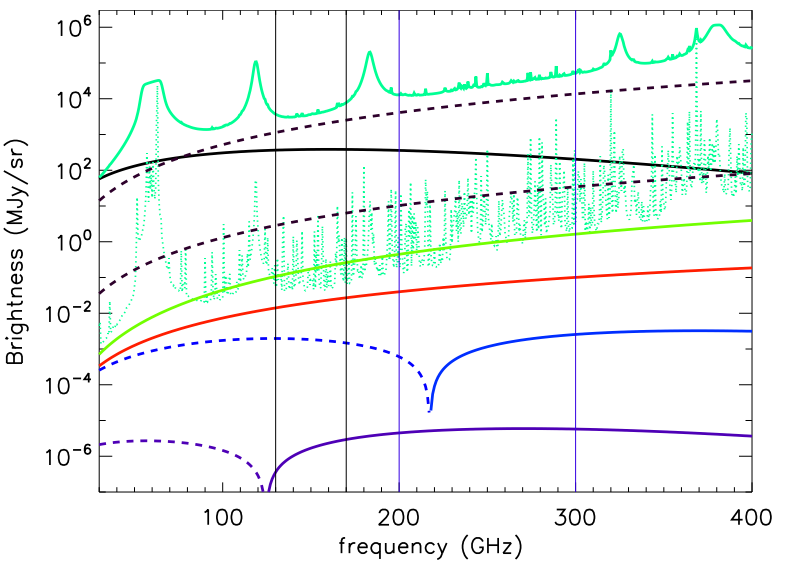


Fig. 1. Plot of isotropic brightness from (top to bottom, continuous lines): Antarctic atmosphere in Dome-C (zenith), CMB, interstellar dust (Galactic average), far infrared background (FIRB) from unresolved galaxies, absolute value of the y-distortion ($y = 1.77 \times 10^{-6}$, negative branch dashed), absolute value of the μ-distortion ($\mu = 2 \times 10^{-8}$, $\beta = 2.19$, negative branch dashed). The two dashed lines represent the emission of the window of the receiver, for the ground based (upper) and the balloon-borne (lower) implementations of COSMO. The dotted line represents atmospheric emission at stratospheric balloon altitude (zenith). The two couples of vertical lines define the observation bands of COSMO in its ground-based implementation.

ground-based and near-space efforts are necessary to test and refine methods, and possibly to detect the largest spectral distortions. Here we focus on the COSMO project, a pathfinder, staged effort to be carried out in Dome-C (Antarctica) first, and then on a stratospheric balloon.

The problem of spectral distortions of the CMB monopole, as evident from figure 1, is that they are very small with respect to other isotropic backgrounds, including instrument emission, atmospheric emission (if present), astrophysical foregrounds from the interstellar medium and the other galaxies along the line of sight, and the CMB monopole itself. Moreover the instrument has to be cryogenic to reduce its own emission, and operate where atmospheric emission is very small and/or can be measured and subtracted. For sure, advanced strategies for components separation are in order (see e.g.[6]).

In the following section we focus on the COSMO experiment, attempting sub-orbital measurements of spectral distortions.

2. The COSMO experiment

2.1. *Measurement method and general design*

COSMO is a cryogenic differential Fourier Transform Spectrometer (FTS) in Martin Puplett Interferometer (MPI) configuration:[7] a two input ports differential instrument comparing the brightness of the sky (port A) to the brightness of an accurate internal blackbody (port B), exactly like in COBE-FIRAS.[8] A simplified instrument model is described by equation 1:

$$I_{sky}(x) = \mathcal{R} \int_0^\infty A\Omega(\sigma)[B_{sky}(\sigma) - B_{ref}(T_{ref}, \sigma)]e(\sigma)[1 + cos(4\pi\sigma x)]d\sigma \quad (1)$$

where $I_{sky}(x)$ is the signal measured by the detector as a function of the optical path difference x between the two arms of the interferometer, which is scanned by means of a moving roof-mirror; σ is the wavenumber, $A\Omega(\sigma)$ is the optical throughput of the detector, $B_{sky}(\sigma)$ is the sky brightness present at input port A, $B_{ref}(T_{ref}, \sigma)$ is the brightness of the reference blackbody at temperature T_{ref} present at input port B, $e(\sigma)$ is the spectral efficiency of the instrument, the constant $\mathcal{R}$ is the responsivity of the instrument. During the calibration phase, an external blackbody at a known temperature T_{ext} fills input port A: in these conditions the measured signal is

$$I_{cal}(x) = \mathcal{R} \int_0^\infty A\Omega(\sigma)[B_{ext}(T_{ext}, \sigma) - B_{ref}(T_{ref}, \sigma)]e(\sigma)[1 + cos(4\pi\sigma x)]d\sigma. \quad (2)$$

Antitransforming equation 2 one gets the quantity $\mathcal{R}A\Omega(\sigma)e(\sigma)$, which can then be used in the antitransform of equation 1 to estimate $B_{sky}(\sigma)$.

Both B_{ref} and B_{ext} should be at cryogenic temperatures, close to T_{CMB}, in order to avoid saturation or non-linearity effects in the detectors, and extreme requirements on the knowledge of T_{ref}, T_{ext} and the blackbody emissivities.

Since the COSMO instrument operates in environments (Dome-C or the stratosphere) where atmospheric emission is present, special care must be taken to measure the emission from the atmosphere, and the emission of the vacuum window separating the cryogenic instrument from the external environment. Both these emissions are not present in a space-based instrument, while they contribute very significantly to the measured B_{sky} in the case of a suborbital experiment like COSMO. Moreover, the cryogenic external blackbody will necessarily have a vacuum window, adding its emission to B_{ext}.

2.2. *Coping with atmospheric emission*

In the ground-based implementation, COSMO will operate from the French-Italian Concordia base, in Dome-C (Antarctica). This is likely to be one of the best sites on Earth for mm-wave astronomical observations, being extremely cold and dry. But still atmospheric emission is present. The first defense against atmospheric emission is the selection of the frequency interval. In its ground-based implementation, COSMO uses two frequency channels matching mm-wave atmospheric windows, covering the frequency bands 110-170 GHz and 200-300 GHz. Each frequency channel is detected by an independent focal plane array. In this way radiation from the high frequency band does not overload the low-frequency detectors, so that higher sensitivity can be achieved in the low frequency channel.

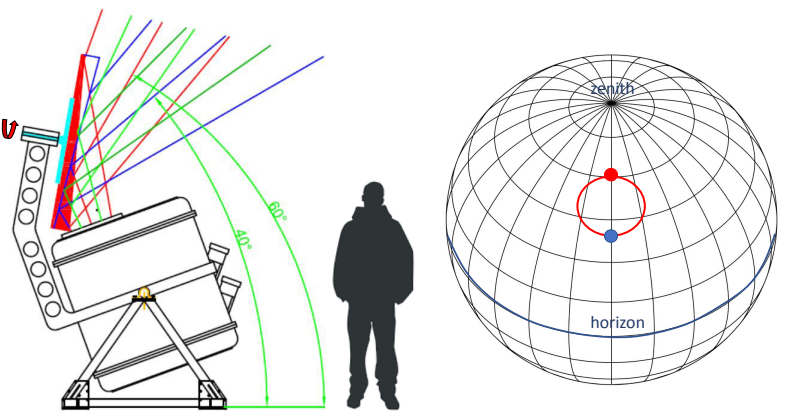


Fig. 2. Geometry of the sky scan for the COSMO experiment. **Left:** The COSMO instrument with the external spinning wedged flat mirror, steering the instrument beam in the sky while spinning. The arrow identifies the spin axis, while the mirror is sketched in two positions (red and blue) corresponding to maximum and minimum elevation of the beams, respectively. The central and marginal beams are also shown in the same colors. **Right:** Resulting scan of the beam (in red) over the celestial sphere. The two spots mark the maximum and minimum elevations explored by the beam. The circular scans, combined with Earth rotation, result in the coverage of several sky patches at different Galactic latitudes.

COSMO uses fast Kinetic Inductance Detectors (KID) and fast elevation scans to separate atmospheric emission and its long-term fluctuations from the monopole of the sky brightness. A fast spinning, wedged flat mirror (>1000 rpm) steers the boresight direction on a circle, 20^o in diameter, scanning a range of elevations (and the corresponding optical depths of the atmosphere) while the cryogenic interferometer scans the optical path difference (see figure 2).

In this way the recorded interferogram (equation 1) will be the one due to the sky monopole (constant during the sky scan) plus the one due to atmospheric emission, which is modulated by the beam elevation variations during the sky scan. A sample expected interferogram is reported in figure 3 for the low-frequency band of COSMO.

By properly sampling the measured interferograms, the interferograms corresponding to the different elevations e scanned by the instrument can be retrieved. These can be antitrasformed to estimate the measured brightness spectra for the different elevations. For each frequency, a cosec(e) law can be used to estrapolate to zero air-mass, thus estimating the isotropic component of the sky brightness (see figure 4).

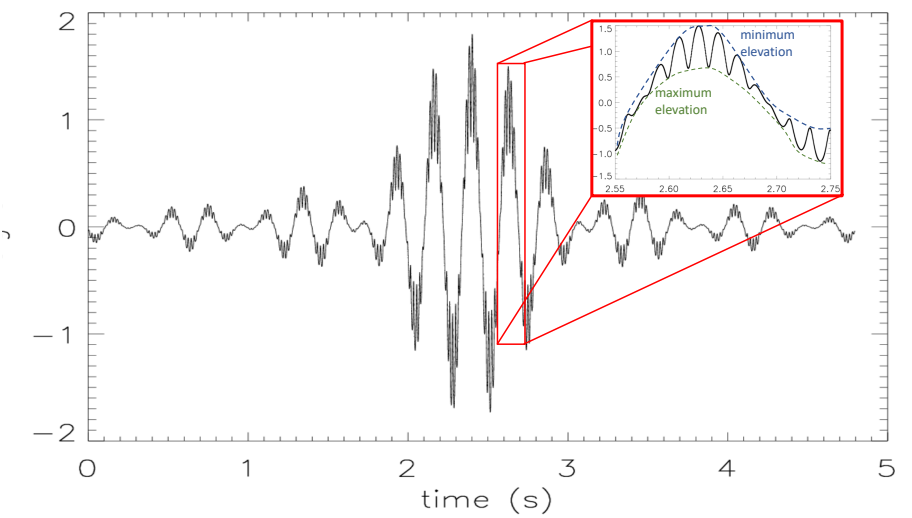


Fig. 3. Example of expected interferogram obtained while the flat wedged mirror is spinning. For each position x of the roof mirror in the MPI (corresponding to a different time on the abscissa), the signal changes due to the fast variation of the observed elevation. Fast (KID) detectors are used to follow these variations. In this example one full scan of the roof mirror is shown (at a mechanical speed of $\sim$ 0.25 cm/s, from -1.27 cm to +1.27 cm of optical path difference). The zoomed inset shows how the interferograms corresponding to the maximum and minimum elevations can be obtained by properly resampling the measured interferogram. Millions of interferograms like this one will be measured during an observation campaign.

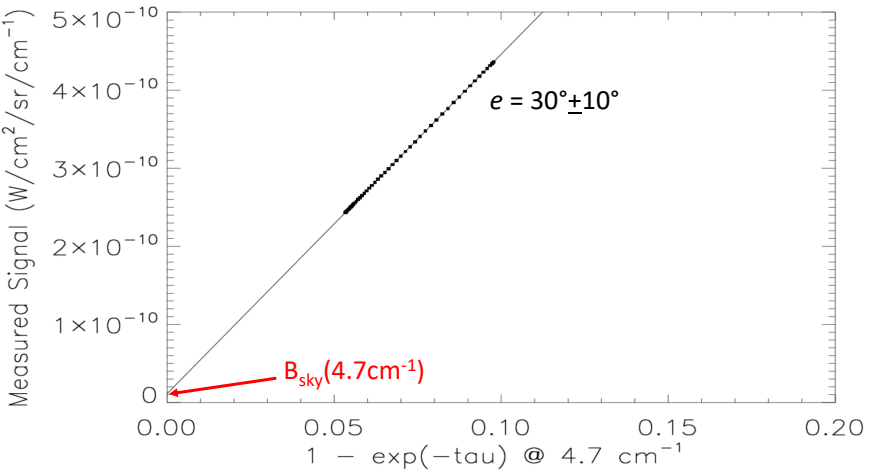


Fig. 4. Extrapolation to zero air-mass to estimate the isotropic component of the sky brightness for each frequency.

Since the instrument measures one interferogram every few seconds, slower atmospheric fluctuations are removed in the process, mitigating the effects of $1/f$ noise, which is characteristic of atmospheric fluctuations.

This measurement method is allowed by the use of fast ($\tau \sim 60\mu s$) KIDs, similar to the ones developed for the OLIMPO experiment.[9] In table 1 we report a possible scanning configuration of the COSMO instrument in its first ground-based implementation.

2.3. *Coping with window emission*

Both the cryogenic reference blackbody and the FTS of COSMO are enclosed in a vacuum shell, necessary for the cryogenic operations. External radiation enters the vacuum jacket through a vacuum window, made from a transparent material for mm waves. The best solution is to use a slab of polypropylene, which has to be ~ 1 cm thick in the ground based implementation, to withstand atmospheric pressure. This can be significantly thinner at balloon altitude: we assume a 25 μm thick window for that implementation. As visible in figure 1, the emission of the receiver window is very important in both cases. The way we cope with it is to measure spectra keeping the internal reference and the external calibrator at the two input ports and steady, for different, well controlled temperatures of the window. This procedure will produce a calibration of the contribution from window emission in the measured spectra as a function of the temperature of the window. This will be monitored during the sky measurements, so that will be possible to subtract

Table 1. Scanning characteristics of the COSMO instrument[a]

circle radius	5	deg
beam FWHM	0.5	deg
wedged mirror spin	600	rpm
time per beam	200	μs
time for one forward plus one reverse sky dip	0.1	s
maximum wavenumber	20	cm^{-1}
sampling step	125	μm
resolution	5-15	GHz
time to complete one interferogram	25.6	s
sky dips per interferogram element	2	

Note: [a]Example of certainly feasible combination of wedge and roof mirror scans; faster scans are currently under consideration.

window emission using the calibration. The expected accuracy of this process has been investigated: preliminary results show that the residuals after subtraction are subdominant when targeting at the measurement of the y distortion.

3. Implementation of the ground-based COSMO experiment

3.1. *Selection of the site and installation*

Due to the high altitude, extremely low temperatures in the winter, and absence of solar illumination for several months in a row, the Dome-C (Antarctica) site offers unmatched excellent conditions for astronomical observations in the mm-wave atmospheric windows (see e.g.[10,11]). Its latitude (75^oS) allows for crosslinked sky scans at constant elevation (which are not possible from the South Pole), similar to those obtained in circumpolar stratospheric balloon flights (see e.g.[12]). The presence of the French-Italian Concordia station provides high quality logistics support (power, communications, resident personnel) for the operations of a modern CMB experiment. The instrument will be hosted in a thermally insulated ISO20 shipping container, with retractable roof, and mounted on a palafitte on top of a hardened snow berm, following a consolidated tradition for medium-size astronomical instruments in Antarctica. The container and the instrument will be oriented so that the elevation movement steers the boresight along the local meridian. As described below (see §4), this simple sky scan strategy allows to observe both high and intermediate Galactic latitudes, totalizing a coverage of $\sim 5\%$ of the sky.

3.2. *The optical system*

3.2.1. *Generalities*

Angular resolution is not the main driver of this experiment, which is aimed at measurements of the *isotropic* monopole component of the CMB brightness. We designed

the instrument with and optical aperture of 22 cm in diameter, producing a diffraction limited beam at the longest wavelength (2.4 mm) of $\sim 0.7^o$ FWHM. This is the result of a tradeoff between the size of the optical aperture and the need of monitoring the effect of *anisotropic* foregrounds. The telescope and the Fourier Transform Spectrometer are located inside the cryostat to minimize their radiative background. A sketch of the optical setup, based on polyethylene lenses, is shown in the left panel of figure 5, while the main data are reported in table 2.

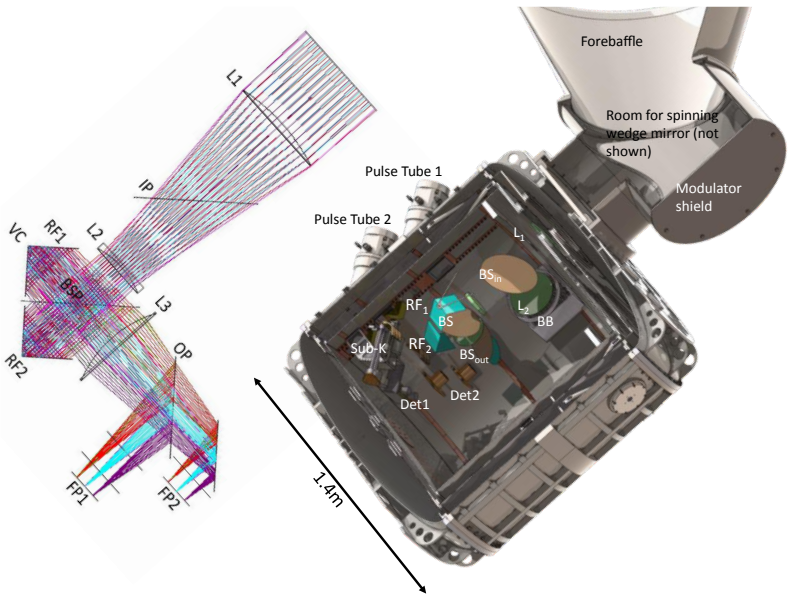


Fig. 5. **Left:** Ray tracing and optical design of COSMO. L1 is the light collector lens, L2 produces a nearly parallel beam for the FTS, L3 focuses radiation of the focal planes FP1 and FP2. The Differential Fourier Transform Spectrometer is costituted by the input polarizer IP, the beam-splitter polarizer BSP, the roof mirrors RF1 and RF2, the output polarizer OP. The input polarizer transmits radiation from the sky and reflects radiation from the cryogenic reference blackbody. **Right:** rendering of the accommodation of the COSMO instrument inside its cryogenic system.

Table 2. Optical parameters of the COSMO instrument.

optical aperture diameter	220	mm
effective focal length	726	mm
multimode pixel antenna aperture diameter	20	mm
focal planes	2	
number of detectors per focal plane	9	
projected pixel to pixel distance (x and y)	0.75	deg
beam FWHM	0.75	deg

3.2.2. *The cryogenic Fourier Transform Spectrometer*

The cryogenic FTS is a MPI, with two delay lines terminated with roof-mirrors. The interferogram is obtained introducing a variable difference in the lengths of the two delay lines. This is obtained with the linear movement of one of the two roof-mirrors (RF1 in figure 5). The maximum optical path difference introduced by the motion of the moving roof mirror is +/- 2 cm, corresponding to a coarse spectral resolution of ∼10 GHz. Since the FTS operates at cryogenic temperature, we have to minimize the friction produced by the motion of the roof-mirror. We have designed a cryomechanism based on harmonic steel flexure blades, operating in the elastic regime. The actuator is a coil moving in the radial magnetic field of a strong permanent magnet (figure 6). The position is sensed by a LVDT. Preliminary room-temperature tests confirm the frictionless operation of the device.

3.2.3. *The detectors*

The output beam from the FTS is naturally split in two exit ports by the output polarizer. In each output port we place a focal plane array, consisting of multimode horn antennas feeding multimode lumped elements kinetic inductance detectors (LEKIDs), on a 4 inch Si wafer. The reflected beam serves the high frequency band array, while the transmitted one serves the low frequency band array.

The design of the COSMO LEKIDs was optimized starting from the experience gained with the LEKID arrays realized for the OLIMPO experiment.[25,26] COSMO detectors are 30 nm thick Aluminum LEKIDs, each of which has a front–illuminated V order Hilbert curve absorber/inductor. The large absorber area, $\sim 8\,\mathrm{mm} \times 8\,\mathrm{mm}$ for both channels, serves to guarantee the efficient absorption of all the modes

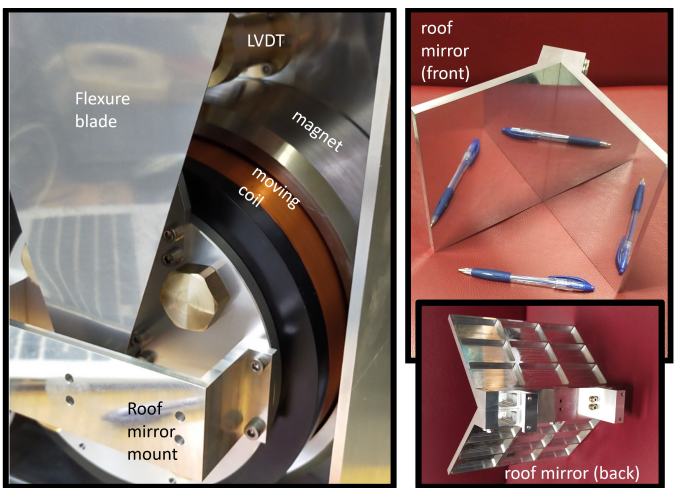


Fig. 6. **Left:** photo of the cryogenic roof-mirror scanner for the MPI FTS. **Right:** the roof mirror. Its projected size is 200 × 200 mm.

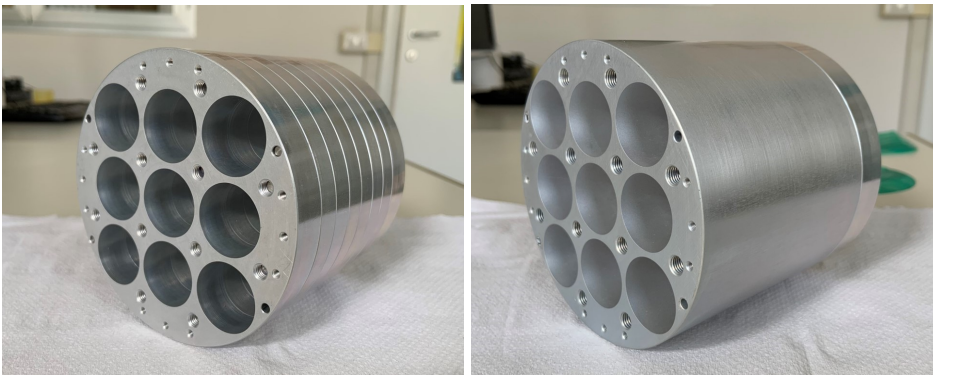


Fig. 7. **Left:** photo of the low-frequency array, made of seven antenna plates and a bottom plate with circular waveguides. **Right:** photo of the high-frequency array, which is made as one piece, with a separate bottom plate of circular waveguides.

which propagate through the multimode waveguides. In particular, the size of the waveguide diameter, 4.5 mm for the 150 GHz channel and 4 mm for the 255 GHz channel, allows the propagation of 10 to 19 modes in the 120-180 GHz range and of 23 to 42 modes in the 210-300 GHz range. Because of the size of the focal planes and the horn apertures, each detector array hosts 9 pixels.

The forecast performance for such detectors, in terms of detector noise equivalent power (NEP) is $\mathrm{NEP_{det}} \sim 3.8 \times 10^{-17}\,\mathrm{W}/\sqrt{\mathrm{Hz}}$, with a response time $\tau \sim 60\,\mu\mathrm{s}$, when operated at about 300 mK. These estimations take into account the absorber volume, the superconductor critical temperature, and the simulated absorber efficiency. This performance is more than adequate for the ground-based implementation of COSMO, which is likely to be dominated by atmospheric noise and systematic effects rather than by detector sensitivity.

3.2.4. *The detector feedhorns*

The radiation coming from the FTS is coupled to the detectors by two arrays of nine multimoded feed-horns working in the $120 - 180$ GHz and $210 - 300$ GHz range, respectively. Since COSMO will perform only total intensity measurements, we chose a smooth-walled profile which is easier to manufacture than a corrugated one.

The low-frequency array consists of Winston cone antennas while the high-frequency array hosts linear profiled antennas. The antenna design is the result of a trade-off between the multimode requirement on the antenna waveguide, the mechanical constraint on the antenna aperture and the optimization of the antenna directivity within the cryostat aperture window. The forecasted FWHM of the main beam is around 20°, the first side-lobe level is below -15 dB and the far side-lobe level is below -30 dB.

The arrays are machined with CNC milling by Pasquali SRL in Milan, and are obtained by superimposing metal plates through dowel-pins and tightening them with screws, as shown in figure 7. The arrays are cylindrical with a diameter of 101 mm to minimize the mass to be cooled as much as possible; the feed-horns are arranged on a square footprint with a 26 mm center-to-center distance.

Both the arrays are made in ergal (Al7075) to conform to the focal plane and avoid differential thermal contractions during the cryogenic cooling of the experiment. They are mechanically coupled to a band-pass filter on the front (aperture) side and to an interface to the focal plane wafer holder on the back. To this purpose, we used the experience gained with the LSPE-SWIPE experiment[15] to design a flared waveguide interface to directly illuminate the KIDs detectors.

3.2.5. *The reference blackbody*

One of the two inputs of FTS looks at a cryogenic reference blackbody. This is a cavity with the inner surface covered by a microwave absorber. The cavity structure is in copper, to achieve a very uniform temperature. We have optimized the shape of the cavity both through geometrical ray tracing (maximizing the number of internal reflections) and via finite elements electromagnetic simulations (minimizing the S_{11} parameter). The performance in the ray-tracing approximation is assessed assuming a reflectance of the coating as a function of the wavenumber $\bar{\nu}$ as in ref.,[16] $r(\bar{\nu}) = 0.08 + \frac{0.06[cm^{-1}]}{\bar{\nu}}$. The number of internal reflections is always $N > 6$, corresponding to the cavity reflectance $R = r^N < 10^{-6}$ over the band 100-300 GHz. This indirectly provides the cavity emissivity as $(1 - R)$.

Due to the size of the calibrator, much larger than the wavelengths of interest, we used *HFSS* electromagnetic analysis only to simulate the performance of a thin cut section of the cavity. This is a few wavelengths thick, and is representative in terms of the absorbing properties of the cavity. However it is not representative of the direction of the damped outgoing radiation. The simulations have been run for the lowest frequencies of the 150 GHz band, and the corresponding dielectric properties of the absorbing coating (relative permittivity $\epsilon_r = 3.50$ and dielectric loss tangent $tg(\delta) = 0.032$ at 100 GHz at 4 K) have been taken from.[17] The input radiation is set as an *Incident Gaussian Beam*, determined by the lens L1 feeding the blackbody cavity. The reflectance, computed as the ratio of the Poynting vector fluxes of the scattered and incident radiation, is $R(120\,\mathrm{GHz}) = 3.2 \cdot 10^{-6}$. Diffractive effects and reflection coefficients at non-normal incidence, not included in the ray tracing approach, produce discrepancies between the two models.

Finite elements thermal simulations have been performed, considering a perfect thermal contact between the copper external structure and the internal absorbing coating, and including an irradiating internal surface of the cavity to a 4 K background. The external structure temperature was set to 2.725 K, from a starting

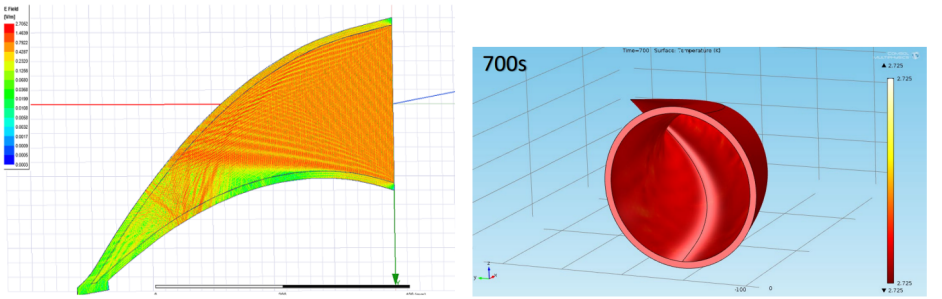


Fig. 8. **Left:** Results of an HFSS simulation of the total electric field propagating within the blackbody cavity, demonstrating its gradual attenuation (colors correspond to a logarithmic scale for field amplitudes. **Right:** Results of a finite elements thermal simulation displaying the temperature at the surface of the internal absorbing coating 700s after cool-down. The temperature gradients are <0.1 mK.

point of 4 K. The thermal properties of the coating are taken from.[18] Preliminary results show a maximum thermal gradient <0.1 mK (see fig.8).

3.2.6. *The atmospheric modulator*

The other input of the FTS looks at the sky through an optical window and a beam-steering mirror, which is used to modulate atmospheric emission and measure it as described in §2.2. The atmospheric scan has to be very fast (up to 2800 rpm) to measure atmospheric emission and subtract it in real time, so that 1/f noise from atmospheric emission is efficiently removed. A light-weighted, all 6061 Aluminum design for the mirror has been optimized to minimize the inertia. The off-axis inertia moments are null using optimized counterweights, allowing us to use simple and reliable electrical motors and shaft bearings.

One of the most dangerous problems in this kind of measurements is the pickup of ground radiation. A custom forebaffle has been designed as the first defense against ground spillover. This forebaffle moves together with the receiver, and will be complemented by much larger steady ground shields. The internal surface will be fully covered with eccosorb absorber. The forebaffle is largely oversized, and has been shaped so that the spillover from its inner surface is not modulated by the spin of the wedge mirror. Heaters and temperature sensors will allow to test any residual spillover contribution to the measured signal by means of custom calibration measurements.

3.3. *The cryogenic system*

A large cryogenic system, based on pulse-tubes and a ^{3}He evaporation cooler is used to cool the optical system and the detector arrays. The system is similar to the one

developed for the QUBIC experiment (see[13]), adapted for operation in the harsh Antarctic environment. To this purpose, we used the experience gained with the BRAIN refrigerator, successfully operated in Dome-C more than 10 years ago.[14] In the right panel of figure 5 we show the implementation of the COSMO instrument inside its cryogenic system. Using pulse tube refrigerators, the cryogenic system can operate continuously and does not rely on the delivery of cryogenic liquids, which would be very difficult in dome-C. Of course, continuous power is required, which amounts to about 10 kW, most of which are used for the pulse tube compressors and to warm-up the environment inside the experiment shelter to above 0 Celsius.

3.4. *Readout electronics*

The arrays of KIDs are read applying a comb of frequencies tuned to the resonances of the different pixels, and reading the S21 parameter for each pixel. A FPGA-based electronic system generates a set of tones corresponding to the KIDs' resonances. The combination of tones is sent to the detectors; the return signal is acquired and analyzed by means of a transceiver which detects the change in amplitude and phase of the tones. The system is based on a commercial architecture (Xilinx Virtex-5 NI7966R). The Virtex5 architecture was selected because it is the most powerful of the family being available as a rad-hard component. This specification not necessary for Antarctica, but is strategic in the perspective of the implementation of COSMO on a stratospheric balloon . We have prepared a state machine which, in the first state, performs a sweep to find the resonance tones of the KIDs. In the second state, a lookup table is generated with the values of the resonances found. With a CORDIC algorithm these tones are generated to compose a comb, which the third state of the machine transmits to the KIDs and acquires in amplitude and phase to measure the variation of the working point of the detectors. We have already tested the functionality of the state machine in the lab by means of Nb kids, fit for a 4 K cryostat. We are currently working to move the entire process of modulation and demodulation on board of the FPGAs, to reach the target rate of acquisition, fit for the operation with the FTS.

4. Performance forecast for the the ground-based COSMO experiment

The performance of the instrument is assessed via ILC[19] (Internal Linear Combination) based simulations. Absolutely-calibrated maps are used as input, containing the superposition of the PySM[20] maps of CMB anisotropies and thermal dust, as the main galactic foreground. The isotropic distortion maps are added as a y-distortion map with $y = 1.77 \cdot 10^{-6}$ and a μ-distortion map with $\mu = 2.0 \cdot 10^{-8}$. The input multi-frequency maps, with a spectral resolution $\Delta\nu = 15\,\mathrm{GHz}$, are smoothed with a FWHM=1°. The noise realization, added as Gaussian noise, is dominated by the photon noise from the cryostat window, whose emission is modeled as a 240 K

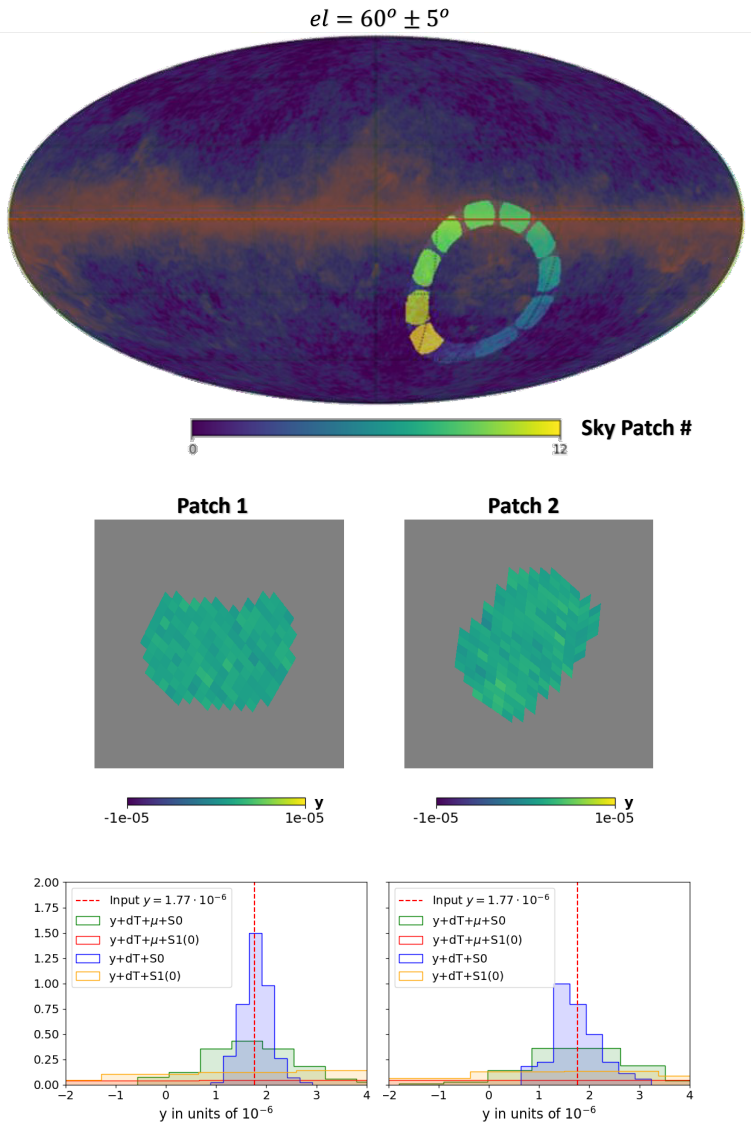


Fig. 9. **Top:** Sky coverage map divided in 11 patches, overlapped to the 270GHz PySM map which includes thermal dust emission and CMB anisotropy. **Middle:** Simulated measured maps for the y distortion sky component, extracted using the ILC method, in two sample sky patches at high Galctic latitude. **Bottom:** Histograms of the simulated estimates for the y values, for the two sky patches above.

grey-body with 1% emissivity, and by the atmospheric emission (modeled based on the *a.m.* software with PWV=0.15mm).

The ILC machinery allows to extract the y-distortion map from the dominant components. Different orders of the thermal dust emission are subtracted from the

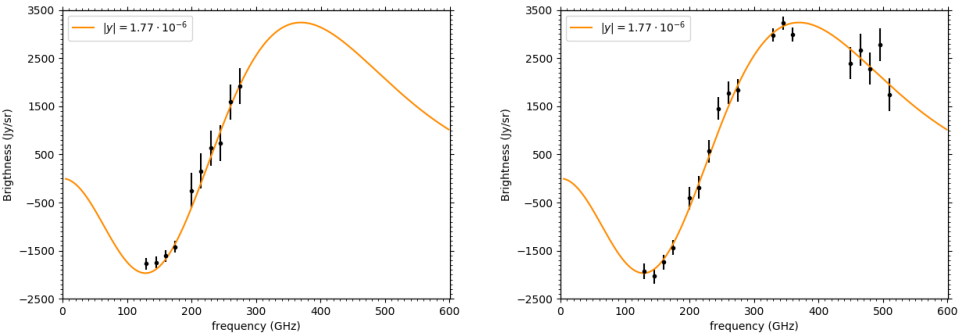


Fig. 10. **Left:** Simulated measurement of the spectral distortion of the CMB after one year of integration in Dome-C, assuming photon noise limited performance dominated by the emissions of the cryostat window and the atmosphere and assuming 15 GHz resolution. **Right:** Simulated measurement of the spectral distortion of the CMB after 15 days of integration on a stratospheric balloon, assuming photon noise limited performance dominated by the CMB, residual atmosphere and warm mirror, and assuming 15 GHz resolution.

solution as in.[21] Histograms of the output Compton-y maps, reported in Fig.9, show that the best ILC solution is provided by subtracting the 0^{th} order of the thermal dust, in combination with the CMB anisotropy.

Assuming a scanning strategy at fixed central elevation, combined with the spinning wedge mirror modulation, 5% of the sky is observed every day, which can be divided for the analysis in 11 sky-patches at different Galactic latitudes (see fig.9). The ILC is independently applied to the patches and the weighted average of the output Compton-y maps provides the estimate of the isotropic y distortion as $y = (1.70 \pm 0.28) \cdot 10^{-6}$.

In figure 10 (left panel) we show the expected spectrum of the spectral distortion we expect to measure after 1 year of integration in Dome-C.

5. Implementation of the balloon-borne COSMO experiment

The spectrum of the CMB at mm wavelengths has been measured from stratospheric balloon platforms in the past (see e.g.[22]). Here we plan to exploit the recent advaces in detector sensitivity and speed, and the atmospheric modulation technique validated with the ground based implementation of COSMO, to take advantage of the very low residual atmospheric emission present at stratospheric altitude (see fig.1). With an atmospheric emission ∼300 times dimmer than in the ground based case, we expect that possible atmospheric residuals, which in the ground based implementation might escape the removal method described in §2.2, will be negligible. In these conditions we expect that the limiting factor will become the emission of the warm components of the instrument, i.e. the room temperature sky scan mirror and the cryostat window. Due to the reduced atmospheric pressure at balloon altitude,

the latter can be made very thin (order of 25 μm), implying a factor $\sim$100 reduction of its thermal emission. We have a long-term plan to reuse the LSPE-SWIPE gondola and cryostat ([23]) to fly a modified COSMO instrument in the stratosphere in a time-scale of 4 years. This activity is synergic to the development of the BISOU experiment.[24] Due to the reduced atmospheric load, we can extend our spectral coverage to significantly higher frequencies, in order to easy the separation of spectral distortions from local foreground emission. We rerun our simple simulation (§4) for the case of an observation in a 15-days flight, with photon-noise limited detectors (here the radiative background is dominated by the CMB, the thin window and the sky modulator mirror) obtaining the performance forecast shown in figure 10 (right panel). The S/N ratio for the extraction of the largest spectral distortion is quite high (>10).

6. Conclusions

The measurement of spectral distortions in the CMB poses hard experimental challenges, but represents a fundamental cosmology tool, able to test a variety of cosmological phenomena. In particular, this kind of measurement seems to be the best, among very few, to shed light on the pre-recombination evolution of the universe. COSMO represents a pioneering sub-orbital approach to investigate spectral distortions at mm wavelengths. It copes with atmospheric emission and its fluctuations using fast detectors and a fast atmospheric emission modulation technique, able to separate its spectral brightness from the isotropic cosmological component. The instrument is in an advanced stage of development, and promises to detect the largest spectral distortion (the y distortion due to ionized matter in the universe) in 2 years of integration from the French-Italian station of Concordia (Antarctica). A further implementation on a stratospheric balloon platform is under study, and promises significant advantages and a necessary validation step in view of the final space mission devoted to spectral distortions.

Acknowledgments

The COSMO activity is funded in Italy by PNRA (National Program for Antarctic Research) and PRIN (Programs with Relevant National Interest). The balloon-borne implementation is being studied within the COSMOS program of the Italian Space Agency.

References

1. J. Chluba, M.H. Abitbol, N. Aghanim, et al., New horizons in cosmology with spectral distortions of the cosmic microwave background, *Experimental Astronomy*, (2021), https://doi.org/10.1007/s10686-021-09729-5

2. J. Mather, et al., A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite, *Ap.J.Letters* **354**, 37 (1990).
3. D. Fixsen, et al., The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set, *Ap.J.* **473**, 576 (1996).
4. A. Kogut, et al., The Primordial Inflation Explorer (PIXIE): a nulling polarimeter for cosmic microwave background observations, *Journal of Cosmology and Astroparticle Physics*, **07**, 025 (2011).
5. P. de Bernardis, et al., Exploring cosmic origins with CORE: The instrument, *Journal of Cosmology and Astroparticle Physics* , **04**, 015 (2018).
6. J. Chluba, et al., Rethinking CMB foregrounds: systematic extension of foreground parametrizations, *MNRAS*, **472**, 1195, (2017).
7. D.H. Martin and E. Puplett, Polarised interferometric spectrometry for the millimetre and submillimetre spectrum, *Infrared Physics*, **10**, 105 (1970)
8. D.J. Fixsen, et al., Calibration of the COBE FIRAS Instrument, *The Astrophysical Journal*, **420**, 457 (1994).
9. S. Masi, et al., Kinetic Inductance Detectors for the OLIMPO experiment: in-flight operation and performance, *Journal of Cosmology and Astroparticle Physics*, **07**, 003 (2019).
10. P. Tremblin, et al., Site testing for submillimetre astronomy at Dome C, Antarctica. *Astronomy and Astrophysics*, **535**, A112, (2011).
11. E. Battistelli, et al., Intensity and polarization of the atmospheric emission at millimetric wavelengths at Dome Concordia, *MNRAS*, **423**, 1293 (2012).
12. S. Masi, et al., Instrument, method, brightness, and polarization maps from the 2003 flight of BOOMERanG, *Astronomy and Astrophysics*, **458**, 687, (2006).
13. S. Masi, et al., QUBIC V: Cryogenic system design and performance, *Journal of Cosmology and Astroparticle Physics*, accepted for publication (arXiv:2008.10659) (2021).
14. G. Polenta, et al., The BRAIN CMB polarization experiment, *New Astronomy Reviews*, **51**, 256 (2007).
15. F. Columbro, et al., SWIPE Multi-mode Pixel Assembly Design and Beam Pattern Measurements at Cryogenic Temperature, *Journal of Low Temperature Physics*, **199**, 312, (2020)
16. J. C. Mather, et al., Calibrator Design for the COBE Far Infrared Absolute Spectrophotometer (FIRAS), *ApJ*, **512**, 511 (1999)
17. J. W. Lamb, Miscellaneous data on materials for millimetre and submillimetre optics, *Int J Infrared Milli Waves*, **17**, 1997–2034 (1996)
18. L. Valenziano et al, Planck-LFI: design and performance of the Kelvin Reference Load Unit, *JINST*, **4**, (2009)
19. M. Remazeilles, et al., Foreground component separation with generalized Internal Linear Combination, *Monthly Notices of the Royal Astronomical Society*, **418**, 467–476 (2011)
20. B. Thorne et al., The Python Sky Model: software for simulating the Galactic microwave sky, *Monthly Notices of the Royal Astronomical Society*, **469**, 2821–2833 (2017)
21. A. Rotti and J. Chluba, Combining ILC and moment expansion techniques for extracting average-sky signals and CMB anisotropies, *Monthly Notices of the Royal Astronomical Society*, **500**, 976–985 (2021)
22. D.P. Woody, P. L. Richards, Near-millimeter spectrum of the microwave background *Astrophysical Journal*, **248**, 18–37 (1981)

23. The LSPE Collaboration, The large scale polarization explorer (LSPE) for CMB measurements: performance forecast, *Journal of Cosmology and Astroparticle Physics*, accepted for publication (arXiv:2008.11049) (2021)
24. B. Maffei, et al., these proceedings.
25. A. Paiella, et al., Kinetic Inductance Detectors for the OLIMPO experiment: design and pre-flight characterization, *Journal of Cosmology and Astroparticle Physics*, **01**, 039 (2019).
26. A. Paiella, et al., Kinetic Inductance Detectors and readout electronics for the OLIMPO experiment, *Journal of Physics: Conference Series*, **1182**, 012005 (2019).

Measuring the Hubble constant H_0 from gravitational lensing

Liliya L.R. Williams

Minnesota Institute for Astrophysics,
University of Minnesota–Twin Cities,
116 Church Street SE, Minneapolis, MN 55454, USA
E-mail: llrw@umn.edu

First proposed in 1964 by Sjur Refsdal, gravitational lensing provides a straightforward and elegant geometrical way of estimating the Hubble constant from cosmologically distant variable sources. The method relies on observationally determined time delays between light arriving through different multiple images, and the mass models of the lens, which are constrained by observed image properties and other information. While the time delays are obtained with increasing precision, the mass models, which are subject to lensing degeneracies, remain the main source of systematic uncertainty. Various modeling groups have adopted different strategies for dealing with degeneracies. In this talk I will describe the basics of extracting H_0 from lensing, the observational successes, modeling challenges, current results, and future prospects.

Keywords: Cosmological parameters – gravitational lensing

1. Introduction

The idea of measuring H_0 from multiple image (i.e., strong) gravitational lensing is due to the Norwegian astronomer, Sjur Refsdal.[1] Because H_0 is the property of global cosmology, with units of inverse time, and because light travel from lensed sources spans cosmological distances, light travel time will depend on H_0. Therefore there should be a way to extract H_0 from gravitational lensing.

Light travel time along null geodesics, from the source, passing the lens and arriving at the observer consists of two contributions: geometrical and gravitational, which are added linearly, since lensing almost always is in the weak field regime of General Relativity:

- **geometrical** – path length, related to the angular location of the image on the sky
- **gravitational** – integrated Newtonian potential along the line of sight, a.k.a. the Shapiro delay

Figure 1 illustrates the basic geometry. The observer on the left sees images at angular locations on the sky, $\vec{\theta}_i$, $\vec{\theta}_j$. The unlensed source position, $\vec{\beta}$, is also indicated. The light travel time, or the arrival time of the image seen at $\vec{\theta}$ is given by $t(\vec{\theta})$, and consists of the two aforementioned terms. The gravitational time delay is calculated as an integral over the observer sky of the sky-projected normalized (dimensionless) surface mass density in the lens, $\kappa(\vec{\theta})$, with a logarithmic kernel. The prefactor in

front of the square brackets is important because it involves cosmological angular diameter distances, each of which scales inversely with H_0. This is where the depence on the Hubble constant comes from in multple image gravitational lensing.

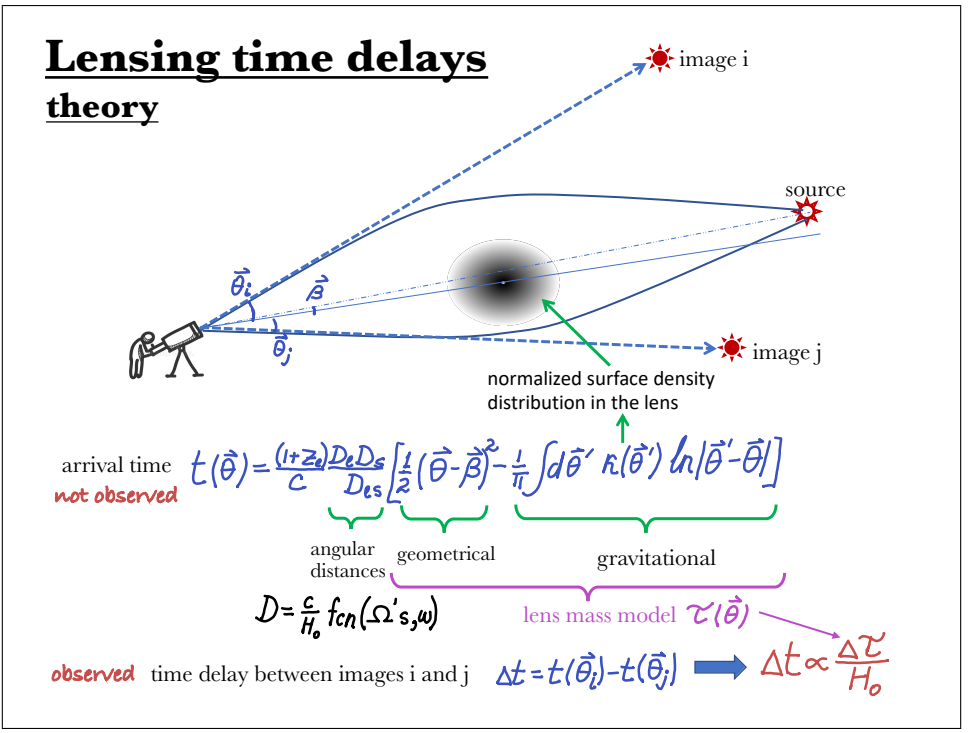


Fig. 1. A sketch of a multiple image (i.e., strong lensing) geometry. The light rays paths for two images are shown, $\vec{\theta}_{i,j}$, as well as the direction towards the unobservable source, $\vec{\beta}$. Hubble constant, H_0 enters through the angular diameter distance in the equation for the arrival time, $t(\vec{\theta})$. See § 1 for details.

The expression in the square brackets is the model arrival time, $\tau(\vec{\theta})$, because it is calculated from a lens model, which recovers the mass distribution in the lens, $\kappa(\vec{\theta})$, as well as the unobservable $\vec{\beta}$. The total light travel time, sometimes called the arrival time, $t(\vec{\theta})$ is not an observable. What we can observe is the difference between the arrival times of two images, Δt, related to $\Delta\tau$ through H_0:

$$\Delta t \propto \Delta\tau / H_0. \tag{1}$$

This is the essence of the method. Basically, to extract H_0 from lensing, one needs to measure time delays, and to construct a good model of the mass distribution in the lens. Thus, the basic principle of the method is simple, and very appealing because it is a one-step determination of H_0, and works on cosmological distances, bypassing all the rungs of the traditional astronomical distance ladder method,

which uses overlaping distance determination methods to reach the Hubble flow. The complications in the lensing measurement of H_0 come from uncertainties in the determination of observational values, as well as lens mass model building.

2. Overview of uncertainties

There are 4 main sources of uncertainties.

- *Line-of-sight (LoS) structures.* The Universe contains not just the source, the lens, and the observer, but also a wide range of other mass structures; in fact, a whole spectrum of mass density fluctuations. The main lens is usually by far the most massive object in the direction of the source, but other, smaller mass clumps also exist. Their influence on the lensing has been recently studied by a number of authors, including Ref. 2, who concluded that the corresponding uncertainty in the determination of H_0 for one lens contributes $\sim 0.6 - 2.4\%$ of scatter, with little bias. If a handful of lenses are used, which is usually the case, the scatter goes down to $\sim 0.5\%$, and at this level it is subdominant to other sources of uncertainty.
- Uncertainties in *cosmological parameters*, like energy densities Ω's, contribute very little to the uncertainties in H_0, typically $< 0.1\%$, if one uses cosmological parameter values that are broadly consistent with the concordance cosmological model ΛCDM.
- *Time delays between images;* see § 3.
- *Mass modeling of the lens;* see § 4.

3. Time delays between lensed images

To measure H_0 one generally uses lens systems that have quasars are sources. Fluxes of most quasars vary in time on scales from seconds to years. The variability is stochastic, its origin being the variations in the conditions of the inner accretion disk around supermassive black holes, residing at the centers of source galaxies. The upper left panel in Figure 2 shows an example of lightcurves (flux vs. time) of images of one particular lens system over about 14 years,[3] and the zoomed in portion below it concentrates on a much narrower time scale, and also resolves time variations of two close images, called A_1 and A_2 (see the sky-plot of the lens on the very left). Observing lightcurves require a lot of effort and dedication. The main group envolved in this is COSMOGRAIL, based at the EPFL in Switzerland.

The lightcurves show features that are common to all images, but displaced in the time direction. (There are also offsets in flux, since images are magnified by different amount by the lens.) These displacements can be measured and are the lensing-induced time delays. There are many algorithms that recover time delays from lightcurves: six examples are shown on the right, for one particular pair of images in this lens system. The two bottom values are two different ways of "averaging" over these individual determinations. In this case, the best estimate is about 36 days,

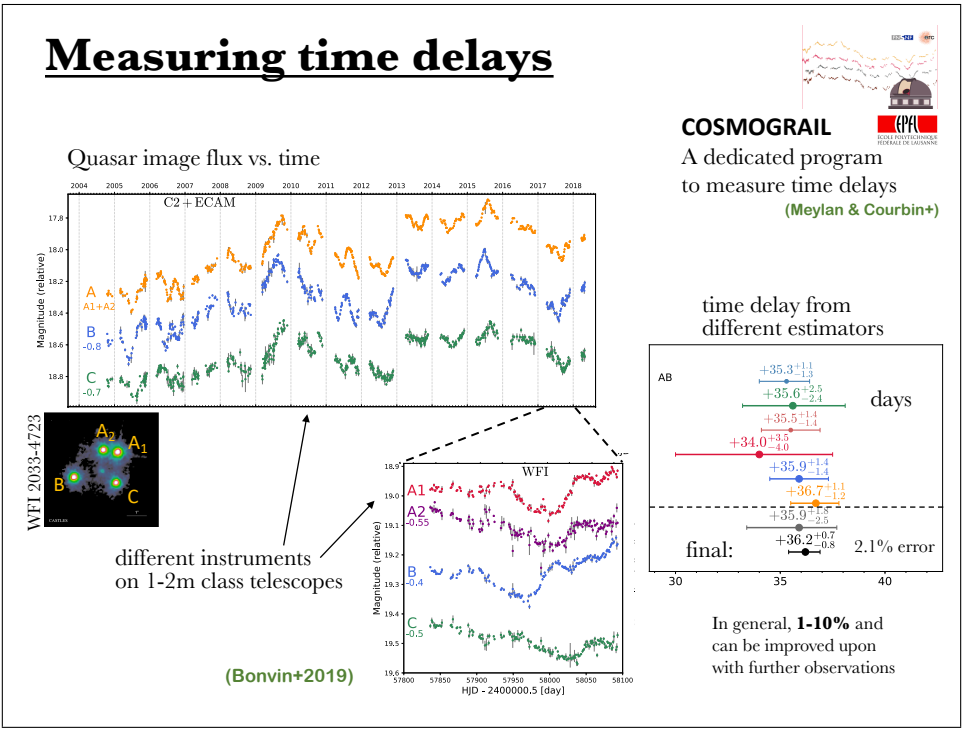


Fig. 2. An example of image lightcurves in one particular quad lens system (left). The results of time delay estimation using different methods are shown in the panel on the right. See § 3 for details.

with an error of 2.1%. In galaxy-scale lens systems, time delays of the order of few days to few weeks are common. Precision ranges from $1 - 10\%$, and can be improved upon with longer base-line, and possible improvements in the cadence of observations.

4. Mass modeling of lens systems

4.1. *Data constraints*

We start with assembling observational data that will help in constraining the mass distribution in the lensing galaxy. These are

- quasar image positions
- quasar time delays
- quasar image flux ratios
- Einstein ring, formed by the extended and usually merging multiple images of the host galaxy of the source quasar

While the first two are always used in deriving H_0, the last two have caveats.

Ratios of quasar image fluxes do constrain mass distribution in the lens, but not necessarily the aspects of the mass distribution that affect H_0. Specifically, image fluxes, and hence flux ratios can be affected by stars, or ΛCDM subhalos in the lensing galaxy that have relatively little mass, but happen to lie very close to the light path of the images. Interpreting these flux ratios as being due to the mass distribution on galaxy scales will lead to wrong conclusions. Because of this image flux ratios are usually not used.

The extended Einstein rings are spread over many pixels, so in principle can provide a lot of additional information. However, because the *Hubble Space Telescope*'s camera pixels are about a factor of 10 larger in linear dimension than the positional accuracy for quasar point images, the positional information from extended rings is not as constraining. Furthermore, the rings are often faint, with low signal-to-noise ratio. Finally, the shape of the (unlensed) source is unknown.

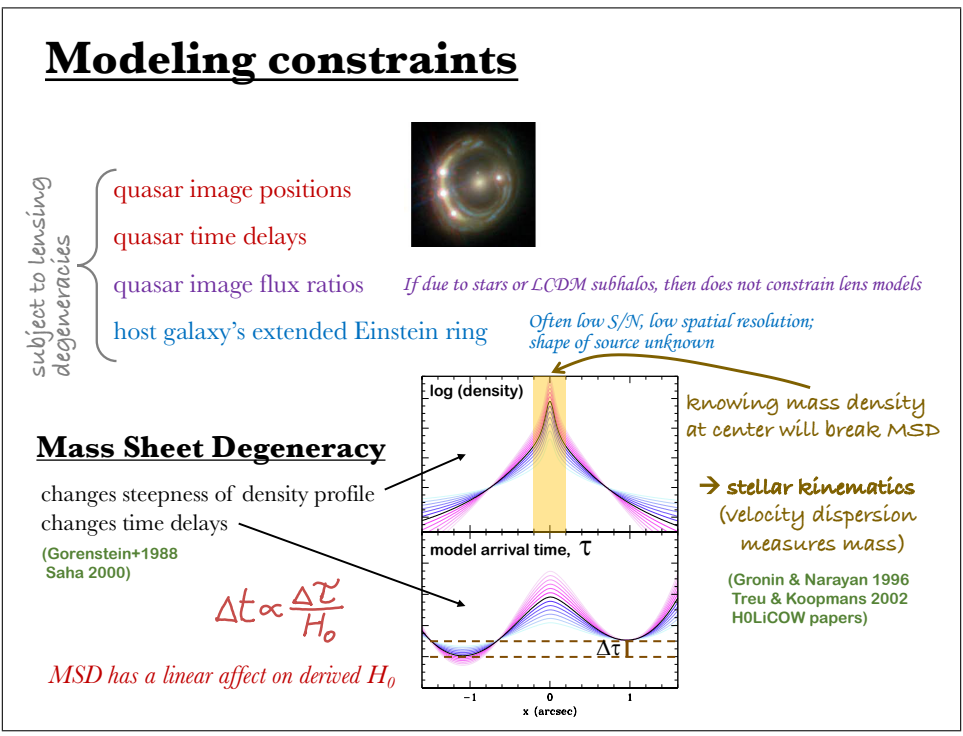


Fig. 3. Lens mass modeling constraints (top), and an illustration of models connected by Mass Sheet Degeneracy (MSD, bottom). See § 4.2 for details.

4.2. *Lensing degeneracies*

All the lensing observables are subject to lensing degeneracies. This means that a given set of observables can be reproduced by more than one lens mass distribution. The most famous degeneracy, the Mass Sheet Degeneracy (MSD) is an exact mathematical degeneracy, which was first recognized and described by Ref. 4. The bottom half of Figure 3 illustrates MSD. If the model arrival time surface τ is stretched or shrunk uniformly, i.e., $\tau \rightarrow \lambda\tau$, where λ is a constant, then the derived H_0 will be affected similarly, since observed Δt values have to stay the same (see eq. 1). In terms of the lens mass distribution this amounts to (approximately) changing the slope of the density profile, so MSD is sometimes referred to as the steepness degeneracy.[5] Steeper density profiles result in larger derived values of H_0.

The mathematically exact MSD is not the main problem; because of the limited precision of observations, it is actually the approximate version of MSD which we must try to limit in order to obtain good accuracy in derived H_0. This breaking of (approximate) MSD can be achieved if we use an independent method (not lensing) to measure the steepness of the lens density profile. The method usually used is the stellar kinematics: the motion of stars near the center of the lensing galaxy, as measured by the velocity dispersion, depends on the mass distribution.[6,7] The sense of the relation is as follows: If stellar kinematics indicate a lot of mass at the very center, that means the density profile is steep.

However, there are caveats which have been pointed out in the recent papers. Ref. 8 shows that the kinematic priors used in the analysis can have a significant effect on the results. Ref. 9 demonstrate that using kinematics can result in a bias in derived H_0 of a few percent or more. And Ref. 10 conclude that unresolved stellar kinematics—which is what being used today—have limited power to constrain the mass profile. Fortunately, spatially resolved kinematics—which will be possible with the *James Webb Space Telescope*—where stellar velocity dispersion is measured as a function of 2D projected position within the lensing galaxy, will be very helpful in breaking MSD.[11]

Other degeneracies exist as well, most of which affect the shape, or the morphology of the mass distribution in the lens. Known examples are the Source-Position Transformation (SPT), which is a generalization of the MSD,[15] monopole degeneracy,[5,16] and a whole range of other shape degeneracies,[17] which we are yet to classify and fully understand. An example is presented in Figure 4; it shows three different mass models of the same lens system, from Ref. 12 (left panel; parametric with one mass component), 13 (middle panel; parametric with two offset mass components), and 14 (right panel; free-form method). There are many common features in the three mass maps, but the differences in the shape of the mass distributions

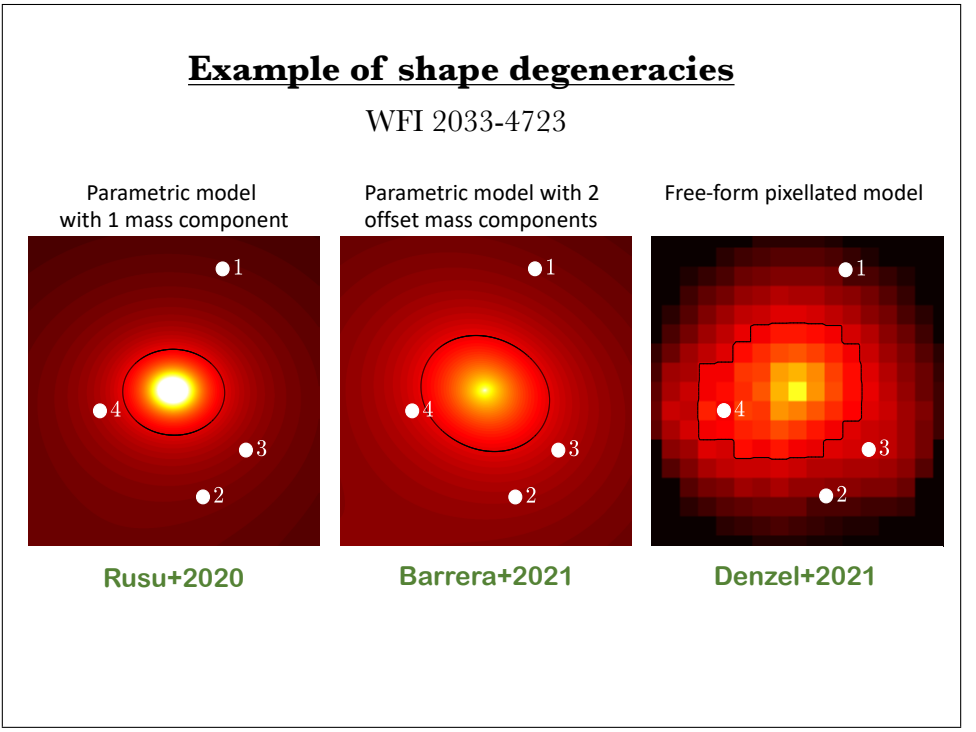


Fig. 4. Three different mass models of the same quad lens system: Ref. 12 (left panel; one mass component), 13 (middle panel; two offset mass components), and 14 (right panel; free-form). These illustrate shape degeneracies in lens mass modeling. See § 4.2 for details.

are also apparent. (Note that the H0LiCOW reconstruction [left panel; Ref. 12)] does not reproduce quasar image positions to even $\sim 5\sigma$, while the the other two [middle & right panels] reproduce these to $\sim 1\sigma$.)

Lensing degeneracies are the largest source of uncertainty in any lensing reconstruction, and hence in the determination of H_0.

4.3. *Constructing lens mass models*

Because lensing data is limited, and because of the existance of lensing degeneracies, lens mass models are not unique. Therefore, we need to appeal to data to constrain our models, based on our existing knowledge about galaxies. Different modeling groups have varying philosophies as to how much of the prior knowledge to incorporate into their models, and how exactly to parametrize this information.

One way to classify different modeling methods is by the number of models parameters they use. This is the horizontal axis in Figure 5. Please keep in mind that these numbers are approximate. The simplest models represent the lens mass distribution by one mass component, with some density profile slope, ellipticity, position

angle, and a tidal term (lensing external shear), which captures the contribution from the external mass. These add up to $\sim 5-6$ parameters. Somewhat more detailed lens mass models use two superimposed mass components, where the two are co-centered, and usually represent baryons (stars), and dark matter, respectively. These result in $\sim 9-15$ parameters. These two types of models are the ones used by the H0LiCOW Collaboration,[18] whose results for H_0 are reported in many papers.

However, there is evidence that these models are not sufficient to describe lenses accurately. For example, as already noted above, one and two co-centered mass component models struggle to reproduce quasar image positions to observational precision.[12] It appears that mass models need more freedom in order to reproduce lens systems more accurately. For example, Ref. 19–22 have shown that offsetting the 2 mass components, which results in deviations from purely elliptical mass distributions, allows better fits for many quad lens systems. The models with the most parameters are the free-form models, which allow each mass pixel in the lens to vary almost independently from others. These models—PixeLens and Glass—have also been used to estimate H_0.[14,17,23] I will present the results in § 5.

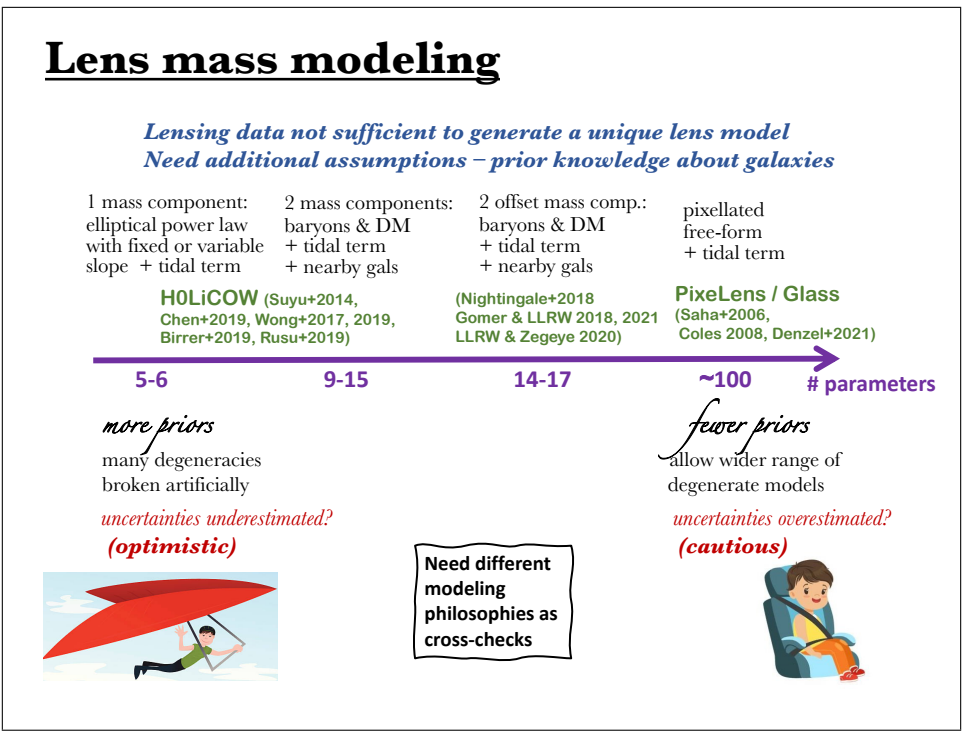


Fig. 5. A few lens mass modeling approaches exist in the literature. One way to classify them is by the number of models parameters they use, represented (approximately) on the horizontal axis. See § 4.3 for details.

Models with fewer parameters and restrictive priors represent a more optimistic view of our knowledge about the galaxies acting as lenses. They suppress lensing degeneracies through the strong assumptions they use, and as a result may underestimate uncertainties in H_0. On the other hand, models with many parameters have fewer priors, and are cautious about how much knowledge we can assume about galaxies. These allow a wider range of degenerate models to be present, and could overestimate the uncertainties. In the face of limited observational data, the best compromise is to have results from different modeling methods act as cross-checks.

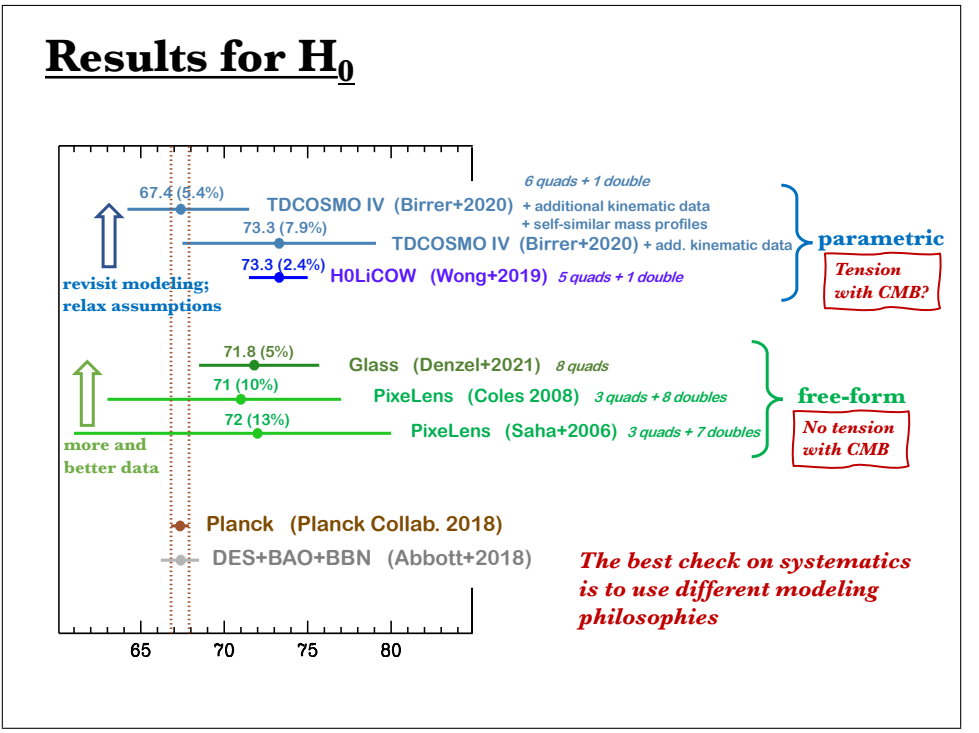


Fig. 6. Summary of the current results of H_0 estimation for strong lensing. See § 5 for details.

5. Results

Figure 6 summaries the results for H_0 determination from multiple image lensing from two groups with very different modeling approaches: parameteric models from the H0LiCOW Collaboration,[18,24] and free-form models.[14,25] (I am a part of the latter group.)

The top portion of Figure 6 presents the most recent H0LiCOW results, which used lens models with one and two co-centered mass components (TDCOSMO is

a continuation of that group's effort). The H0LiCOW's derived value of $H_0 = 73.3$ km/s/Mpc with 2.4% error[24] is probably the most widely quoted Hubble constant result from gravitational lensing. It is in tension with the Planck result,[26] and the other "early Universe" measurement.[27] About a year ago, the TDCOSMO Collaboration revisited the measurement, relaxing their earlier assumptions.[10] They dropped the assumption on the lens mass density profile, and instead added kinematic data from a larger sample galaxies that do not act as lenses. This recovers the same H_0 value as before, but increases the erorrs to about 7.9%. If one also assumes self-similarity in the galaxy mass distribution, then the central derived value becomes $H_0 = 67.4$, with uncertainty of 5.4%, which is now completely consistent with the early Universe Planck value. (Note that these 3 papers used nearly the same set of lens systems, namely, 5-6 quads and 1 double lens. Quads provide 3 times as many lensing data constrains compared to doubles.)

The middle portion of Figure 6 summarizes the results fro the free-form method. The two earlier papers,[17,23] used 3 quads and 7-8 doubles, and derived H_0 with rather large uncertainties, $10 - 13\%$, which easily encompassed the Planck result. In the past year, the number of quads in the sample was increased to 8, and the uncertainties shunk to 5%. (Most of these quads are in common with those in H0LiCOW and TDCOSMO samples.) The cenral value of H_0 remained nearly the same in all 3 papers. The latest value is $\sim 1.1\sigma$ away from that of the Planck Collaboration. There is no tension between the free-form model results and those from the early Universe probes.

In the estimation of H_0, systematics are the main source of concern. The best way to check on systematics is to use different modeling philosophies and compare their results. The parametric and the free-form methods discussed above have no common modeling assumptions, and yet their most recent results differ by at most $\sim 1\sigma$, and show no significant tension with the Planck early Universe value.

6. Future directions

Many exciting developments in the field of lensing determination of H_0 are just around the corner! In addition to multiply imaged quasars, we will use multiply imaged Supernova Type Ia, Gamma-Ray Bursts, and Fast Radio Bursts.[28–34] Measurements of image time delays from these will be much more precise. Furthermore, since the light from these sources will fade with time (in contrast to that of quasars), it will make it easier to characterize the light from the lensing galaxy, which will help with its mass modeling. Type Ia Superova, whose instrinsic flux is known, will also aide in breaking the mass sheet degeneracy.

In this talk I discussed lensing systems with galaxies as the main lens, but galaxy clusters can also be used,[35,36] especially if the number of lensed images is in the few hundreds. While the mass distribution in galaxy clusters is more complicated than in individual galaxies, clusters lense tens or even hundred(s) of background

sources, resulting in 1-few hundred of lensed images, which can constrain the mass distribution suffiently well to yield 1-few percent level accuracy on H_0.[37]

In the next several years, these improvements should be able to deliver $\sim 1\%$ accurary and precision determination of H_0 from gravitational lensing.

References

1. S. Refsdal, On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect, *MNRAS* **128**, p. 307 (January 1964).
2. D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du and A. Benson, Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses, *MNRAS* **491**, 6077 (February 2020).
3. V. Bonvin, M. Millon, J.-H. Chan, F. Courbin, C. Rusu, D. Sluse, S. Suyu, K. Wong, C. Fassnacht, P. Marshall *et al.*, Cosmograil-xviii. time delays of the quadruply lensed quasar wfi2033- 4723, *Astronomy & Astrophysics* **629**, p. A97 (2019).
4. M. V. Gorenstein, E. E. Falco and I. I. Shapiro, Degeneracies in Parameter Estimates for Models of Gravitational Lens Systems, *Astroph. J* **327**, p. 693 (April 1988).
5. P. Saha, Lensing Degeneracies Revisited, *Astron. J.* **120**, 1654 (October 2000).
6. N. A. Grogin and R. Narayan, A New Model of the Gravitational Lens 0957+561 and a Limit on the Hubble Constant, *Astroph. J* **464**, p. 92 (June 1996).
7. T. Treu and L. V. E. Koopmans, The internal structure of the lens PG1115+080: breaking degeneracies in the value of the Hubble constant, *MNRAS* **337**, L6 (December 2002).
8. S. Birrer, A. Amara and A. Refregier, The mass-sheet degeneracy and time-delay cosmography: analysis of the strong lens RXJ1131-1231, *J. Cosmology & Astroparticle Phys.* **2016**, p. 020 (August 2016).
9. M. Gomer and L. L. R. Williams, Galaxy-lens determination of H_0: constraining density slope in the context of the mass sheet degeneracy, *J. Cosmology & Astroparticle Phys.* **2020**, p. 045 (November 2020).
10. S. Birrer, A. J. Shajib, A. Galan, M. Millon, T. Treu, A. Agnello, M. Auger, G. C. F. Chen, L. Christensen, T. Collett, F. Courbin, C. D. Fassnacht, L. V. E. Koopmans, P. J. Marshall, J. W. Park, C. E. Rusu, D. Sluse, C. Spiniello, S. H. Suyu, S. Wagner-Carena, K. C. Wong, M. Barnabè, A. S. Bolton, O. Czoske, X. Ding, J. A. Frieman and L. Van de Vyvere, TDCOSMO. IV. Hierarchical time-delay cosmography – joint inference of the Hubble constant and galaxy density profiles, *Astron. & Astroph.* **643**, p. A165 (November 2020).
11. A. Yıldırım, S. H. Suyu and A. Halkola, Time-delay cosmographic forecasts with strong lensing and JWST stellar kinematics, *MNRAS* **493**, 4783 (April 2020).
12. C. E. Rusu, K. C. Wong, V. Bonvin, D. Sluse, S. H. Suyu, C. D. Fassnacht, J. H. H. Chan, S. Hilbert, M. W. Auger, A. Sonnenfeld, S. Birrer, F. Courbin, T. Treu, G. C. F. Chen, A. Halkola, L. V. E. Koopmans, P. J. Marshall and A. J. Shajib, H0LiCOW XII. Lens mass model of WFI2033-4723 and blind measurement of its time-delay distance and H_0, *MNRAS* **498**, 1440 (October 2020).
13. B. Barrera, L. L. R. Williams, J. P. Coles and P. Denzel, Bridging the Gap Between Simply Parametrized and Free-Form Pixelated Models of Galaxy Lenses: The Case of WFI 2033-4723 Quad, *arXiv e-prints* , p. arXiv:2108.04348 (August 2021).
14. P. Denzel, J. P. Coles, P. Saha and L. L. R. Williams, The Hubble constant from eight time-delay galaxy lenses, *MNRAS* **501**, 784 (February 2021).

15. P. Schneider and D. Sluse, Source-position transformation: an approximate invariance in strong gravitational lensing, *Astron. & Astroph.* **564**, p. A103 (April 2014).
16. J. Liesenborgs and S. De Rijcke, Lensing degeneracies and mass substructure, *MNRAS* **425**, 1772 (September 2012).
17. P. Saha and L. L. R. Williams, Gravitational Lensing Model Degeneracies: Is Steepness All-Important?, *Astroph. J* **653**, 936 (December 2006).
18. S. H. Suyu, V. Bonvin, F. Courbin, C. D. Fassnacht, C. E. Rusu, D. Sluse, T. Treu, K. Wong, M. W. Auger, X. Ding *et al.*, H0licow–i. h 0 lenses in cosmograil's wellspring: program overview, *Monthly Notices of the Royal Astronomical Society* **468**, 2590 (2017).
19. M. R. Gomer and L. L. R. Williams, The impact of ΛCDM substructure and baryon-dark matter transition on the image positions of quad galaxy lenses, *MNRAS* **475**, 1987 (April 2018).
20. J. W. Nightingale, R. J. Massey, D. R. Harvey, A. P. Cooper, A. Etherington, S.-I. Tam and R. G. Hayes, Galaxy structure with strong gravitational lensing: decomposing the internal mass distribution of massive elliptical galaxies, *MNRAS* **489**, 2049 (October 2019).
21. M. R. Gomer and L. L. R. Williams, Galaxy-lens determination of H_0: the effect of the ellipse + shear modelling assumption, *MNRAS* **504**, 1340 (June 2021).
22. L. L. R. Williams and D. Zegeye, Two-component mass models of the lensing galaxy in the quadruply imaged supernova iPTF16geu, *The Open Journal of Astrophysics* **3**, p. 10 (September 2020).
23. J. Coles, A New Estimate of the Hubble Time with Improved Modeling of Gravitational Lenses, *Astroph. J* **679**, 17 (May 2008).
24. K. C. Wong, S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger, M. W. Auger, S. Birrer, J. H. H. Chan, F. Courbin, S. Hilbert, O. Tihhonova, T. Treu, A. Agnello, X. Ding, I. Jee, E. Komatsu, A. J. Shajib, A. Sonnenfeld, R. D. Blandford, L. V. E. Koopmans, P. J. Marshall and G. Meylan, H0LiCOW – XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, *MNRAS* **498**, 1420 (October 2020).
25. P. Saha and L. L. R. Williams, A Portable Modeler of Lensed Quasars, *Astron. J.* **127**, 2604 (May 2004).
26. Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego,

P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei and A. Zonca, Planck 2018 results. VI. Cosmological parameters, *Astron. & Astroph.* **641**, p. A6 (September 2020).

27. T. M. C. Abbott, F. B. Abdalla, J. Annis, K. Bechtol, J. Blazek, B. A. Benson, R. A. Bernstein, G. M. Bernstein, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, C. L. Chang, T. M. Crawford, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, A. E. Evrard, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, J. W. Henning, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, E. Krause, K. Kuehn, S. Kuhlmann, N. Kuropatkin, O. Lahav, A. R. Liddle, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, A. Manzotti, M. March, J. L. Marshall, R. Miquel, J. J. Mohr, T. Natoli, P. Nugent, R. L. C. Ogando, Y. Park, A. A. Plazas, C. L. Reichardt, K. Reil, A. Roodman, A. J. Ross, E. Rozo, E. S. Rykoff, E. Sanchez, V. Scarpine, M. Schubnell, D. Scolnic, I. Sevilla-Noarbe, E. Sheldon, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, A. R. Walker, R. H. Wechsler, J. Weller, W. Wester, W. L. K. Wu, J. Zuntz, Dark Energy Survey Collaboration and South Pole Telescope Collaboration, Dark Energy Survey Year 1 Results: A Precise H_0 Estimate from DES Y1, BAO, and D/H Data, *MNRAS* **480**, 3879 (November 2018).
28. A. Goobar, R. Amanullah, S. R. Kulkarni, P. E. Nugent, J. Johansson, C. Steidel, D. Law, E. Mörtsell, R. Quimby, N. Blagorodnova, A. Brandeker, Y. Cao, A. Cooray, R. Ferretti, C. Fremling, L. Hangard, M. Kasliwal, T. Kupfer, R. Lunnan, F. Masci, A. A. Miller, H. Nayyeri, J. D. Neill, E. O. Ofek, S. Papadogiannakis, T. Petrushevska, V. Ravi, J. Sollerman, M. Sullivan, F. Taddia, R. Walters, D. Wilson, L. Yan and O. Yaron, iPTF16geu: A multiply imaged, gravitationally lensed type Ia supernova, *Science* **356**, 291 (April 2017).
29. S. H. Suyu, T.-C. Chang, F. Courbin and T. Okumura, Cosmological Distance Indicators, *SSRv* **214**, p. 91 (August 2018).
30. S. H. Suyu, S. Huber, R. Cañameras, M. Kromer, S. Schuldt, S. Taubenberger, A. Yıldırım, V. Bonvin, J. H. H. Chan, F. Courbin, U. Nöbauer, S. A. Sim and D. Sluse, HOLISMOKES. I. Highly Optimised Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals, *Astron. & Astroph.* **644**, p. A162 (December 2020).
31. A. Zitrin and D. Eichler, Observing Cosmological Processes in Real Time with Repeating Fast Radio Bursts, *Astroph. J* **866**, p. 101 (October 2018).

32. C. Grillo, P. Rosati, S. H. Suyu, I. Balestra, G. B. Caminha, A. Halkola, P. L. Kelly, M. Lombardi, A. Mercurio, S. A. Rodney and T. Treu, Measuring the Value of the Hubble Constant "à la Refsdal", *Astroph. J* **860**, p. 94 (June 2018).
33. J. Wagner, J. Liesenborgs and D. Eichler, Multiply imaged time-varying sources behind galaxy clusters. Comparing fast radio bursts to QSOs, SNe, and GRBs, *Astron. & Astroph.* **621**, p. A91 (January 2019).
34. X. Ding, T. Treu, S. Birrer, G. C. F. Chen, J. Coles, P. Denzel, M. Frigo, A. Galan, P. J. Marshall, M. Millon, A. More, A. J. Shajib, D. Sluse, H. Tak, D. Xu, M. W. Auger, V. Bonvin, H. Chand, F. Courbin, G. Despali, C. D. Fassnacht, D. Gilman, S. Hilbert, S. R. Kumar, J. Y. Y. Lin, J. W. Park, P. Saha, S. Vegetti, L. Van de Vyvere and L. L. R. Williams, Time delay lens modelling challenge, *MNRAS* **503**, 1096 (May 2021).
35. P. L. Kelly, S. A. Rodney, T. Treu, R. J. Foley, G. Brammer, K. B. Schmidt, A. Zitrin, A. Sonnenfeld, L.-G. Strolger, O. Graur, A. V. Filippenko, S. W. Jha, A. G. Riess, M. Bradac, B. J. Weiner, D. Scolnic, M. A. Malkan, A. von der Linden, M. Trenti, J. Hjorth, R. Gavazzi, A. Fontana, J. C. Merten, C. McCully, T. Jones, M. Postman, A. Dressler, B. Patel, S. B. Cenko, M. L. Graham and B. E. Tucker, Multiple images of a highly magnified supernova formed by an early-type cluster galaxy lens, *Science* **347**, 1123 (March 2015).
36. P. L. Kelly, S. A. Rodney, T. Treu, L. G. Strolger, R. J. Foley, S. W. Jha, J. Selsing, G. Brammer, M. Bradač, S. B. Cenko, O. Graur, A. V. Filippenko, J. Hjorth, C. McCully, A. Molino, M. Nonino, A. G. Riess, K. B. Schmidt, B. Tucker, A. von der Linden, B. J. Weiner and A. Zitrin, Deja Vu All Over Again: The Reappearance of Supernova Refsdal, *ApJLett* **819**, p. L8 (March 2016).
37. A. Ghosh, L. L. R. Williams and J. Liesenborgs, Free-form grale lens inversion of galaxy clusters with up to 1000 multiple images, *MNRAS* **494**, 3998 (May 2020).

Extra components consistency in the Hubble tension and BBN

Osamu Seto

Institute for the Advancement of Higher Education, Hokkaido University,
Sapporo 060-0817, Japan,
Department of Physics, Hokkaido University,
Sapporo 060-0810, Japan
E-mail: seto@particle.sci.hokudai.ac.jp

Yo Toda

Department of Physics, Hokkaido University,
Sapporo 060-0810, Japan
E-mail: y-toda@particle.sci.hokudai.ac.jp

The standard ΛCDM cosmological model now seems to face some puzzles. One of the most serious problems is the so-called Hubble tension; the values of the Hubble constant H_0 obtained by local measurements look inconsistent with that inferred from Cosmic Microwave Background (CMB). Although introducing extra energy components such as the extra radiation or Early Dark Energy appears to be promising, such extra components could alter the abundance of light elements synthesized by Big Bang Nucleosynthesis (BBN). We perform a Monte Carlo simulation to evaluate the effect of those extra component scenarios to solve the Hubble tension on the BBN prediction.

Keywords: Cosmology; Big Bang Nucleosynthesis; Dark energy; Radiation; Hubble constant

1. Introduction

The ΛCDM model has been successful in explaining the evolution of our Universe. However, the tension of the Hubble constant is now apparent between the measurement of the local universe and the distant universe. Assuming ΛCDM model, Planck measurements of Cosmic Microwave Background (CMB) anisotropy infers the Hubble constant $H_0 = 67.4 \pm 0.5$ km/s/Mpc.[1] Other distant observations such as the Atacama Cosmology Telescope,[2] Baryon Acoustic Oscillation(BAO),[3] and the combined analysis of BAOand Big Bang Nucleosynthesis (BBN) (independent of CMB)[4] all infer $H_0 \sim 67$ km/s/Mpc. On the other hand, local measurements of H_0 by the SH0ES collaboration with Cepheids and type Ia supernovae (SNe Ia) in Ref.[5] and Ref.[6] and by the H0LiCOW collaboration with lensed quasars[7] have reported as $H_0 \sim 73$ km/s/Mpc. Another local measurement using the Tip of the Red Giant Branch (TRGB) as distance ladders has obtained a value between Planck and the SH0ES, $H_0 \sim 70$ km/s/Mpc.[8] It appears that the discrepancy is more than 3σ significance.

Several ideas have been proposed to extend ΛCDM model to resolve this tension. Among such ideas, we would like to consider the extra radiation and Early Dark Energy[9–18] solutions to the Hubble tension and the consistency with BBN. These extra components are promising solutions. However, the theoretical abundance of light elements synthesized by BBN should be different from the values of the standard scenario. Through the consideration, we confirm that extra radiation and Early Dark Energy are promising solutions. However, two scenarios are constrained by BBN measurements. Extra radiation is constrained by the helium abundance and Early Dark Energy is constrained by the deuterium abundance.[19]

2. H_0 from CMB

First, we explain how the Hubble constant is derived from the distant observation, especially the CMB anisotropy. The measured anisotropy includes the information of the angular size $\theta_* = r_*/D_{M*}$, where r_* is the comoving sound horizon at the recombination and D_{M*} is the comoving angular diameter distance. Planck team directly and strictly measured $\theta_* = 1.041 \times 10^{-2}$. A simple calculation according to the definition shows that $\theta_* \propto H_0/\sqrt{\rho_{\rm early}}$, where $\rho_{\rm early}$ is the energy density in the early universe. Therefore, in order for the Hubble constant from CMB to approach the local value, we need to increase ρ in the early universe. Then, we will consider two promising ways to increase ρ; Extra radiation and Early Dark Energy.

3. Modeling

3.1. *Extra radiation*

One simple "solution" to the Hubble tension is increasing the effective number of neutrinos $N_{\rm eff}$, which is expressed as

$$\Omega_r = \left(1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}\right)\Omega_\gamma. \tag{1}$$

Here,

$$\Omega_i = \left.\frac{\rho_i}{3M_P^2 H^2}\right|_{t=t_0}, \tag{2}$$

with M_P being the reduced Planck mass, are the present values of density parameters for i spices. γ and r stand for CMB photon and total radiation, respectively.

3.2. *Early Dark Energy*

Another solution to the Hubble tension is introducing Early Dark Energy. In the Early Dark Energy scenario, the dark energy density in the early universe was much larger than today and after the critical point, the energy density decreases faster than the background energy densities do. In our analysis, we consider a model where

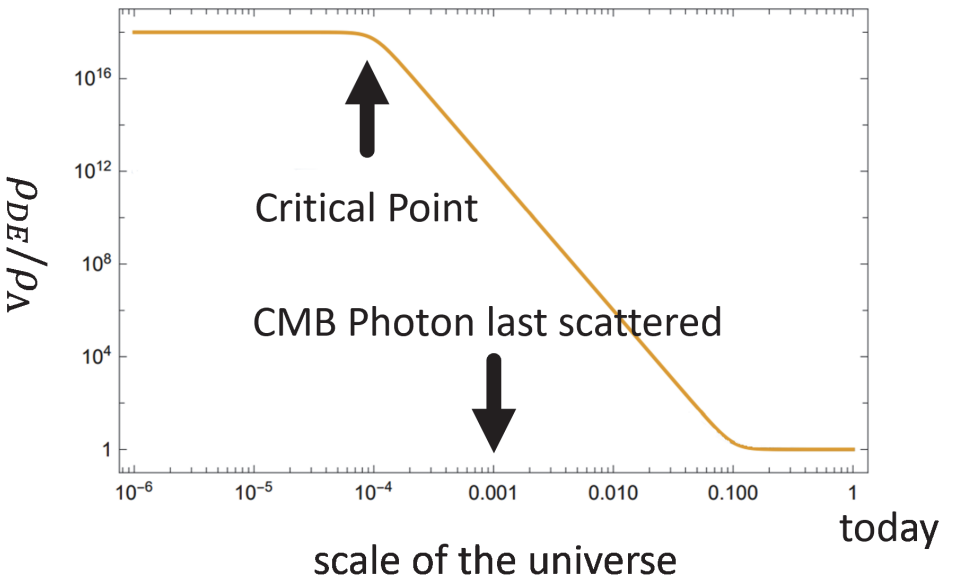


Fig. 1. Energy density evolution of Early Dark Energy

the energy density of Early Dark Energy ρ_{DE} decreases as the kination of scalar field after the critical point at $z = 3000$ around the matter-radiation equality. This setup is according to a preferred parameter set in Ref.[9]. The typical evolution of Dark Energy density $\rho_{DE}(a)$ normalized by ρ_{Λ} is shown in Fig. 1.

4. Data and Analysis

We perform a Markov-Chain Monte Carlo (MCMC) analysis on a $N_{\rm eff}$ model and the Early Dark Energy model described in the previous section. We use the public MCMC code `CosmoMC-planck2018`[20] with implementing the above Early Dark Energy scenarios by modifying its equation file in `camb`. For estimation of light elements, we have used `PArthENoPE_marcucci`.[21,22]

4.1. *Data sets*

We analyze models with referring to the following cosmological observation data sets. We include both temperature and polarization likelihoods for high l ($l = 30$ to 2508 in TT and $l = 30$ to 1997 in EE and TE) and low$l$ `Commander` and lowE `SimAll` ($l = 2$ to 29) of Planck (2018) measurement of the CMB temperature anisotropy.[1] We also include Planck lensing.[23] For constraints on low redshift cosmology, we include data of BAO from 6dF,[24] DR7[25] and DR12.[26] We also include Pantheon[27] of the local measurement of light curves and luminosity distance of supernovae as well as SH0ES (R19)[6] of the local measurement of the Hubble constant from the

Hubble Space Telescope observation of Supernovae and Cepheid variables. Finally, we include the data sets of helium mass fraction Y_P measurement[28] and deuterium abundance D/H measurement[29] to impose the constraints from BBN.

5. Result and Discussion

As is well known, an increase of $N_{\rm eff}$ affects the fit with the observation of light elements, because it contributes to the cosmic expansion at BBN epoch and alters the proton to neutron p/n ratio. This leads to increasing both the helium mass fraction Y_P and the deuterium abundance D/H. Thus, larger relativistic degrees are disfavored by the helium mass fraction measurement, while a little favored by deuterium measurement. This can be seen in Fig. 2.

On the other hand, in the Early Dark Energy scenarios, the Early Dark Energy increases the helium mass fraction little because the cosmic expansion rate in the BBN epoch increase little. Therefore, the Early Dark Energy scenario is consistent with the helium fraction measurement. However, increasing $\Omega_b h^2$ to adjust the CMB

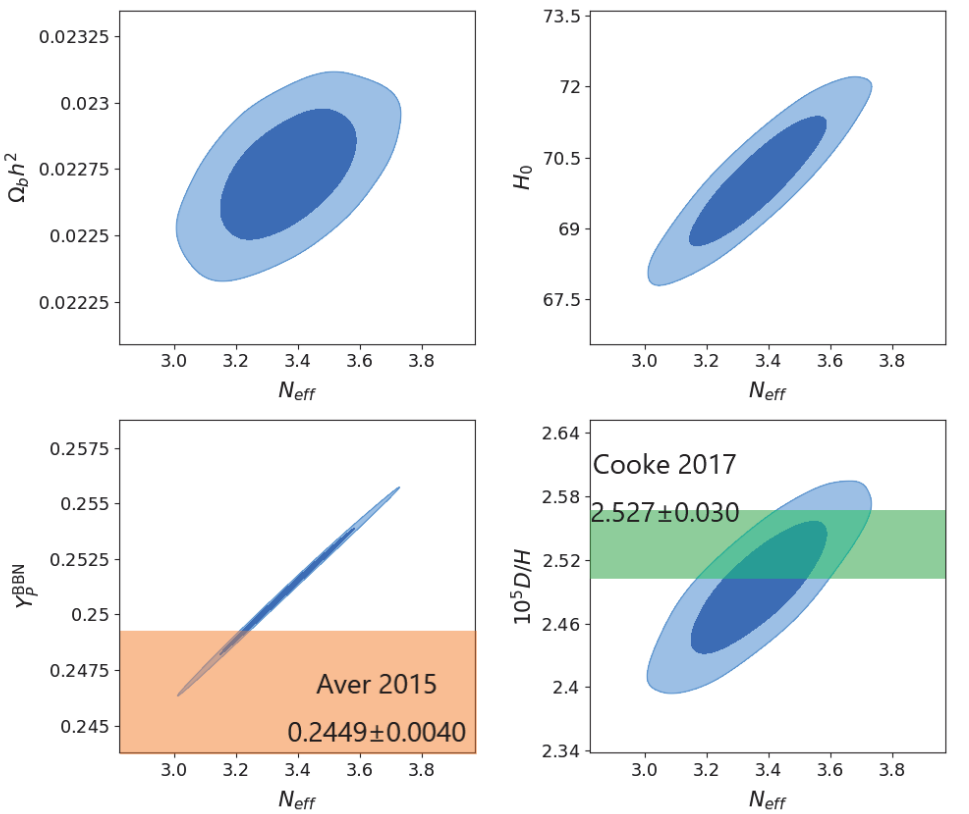


Fig. 2. The posterior and the constraints on the $N_{\rm eff}$ model

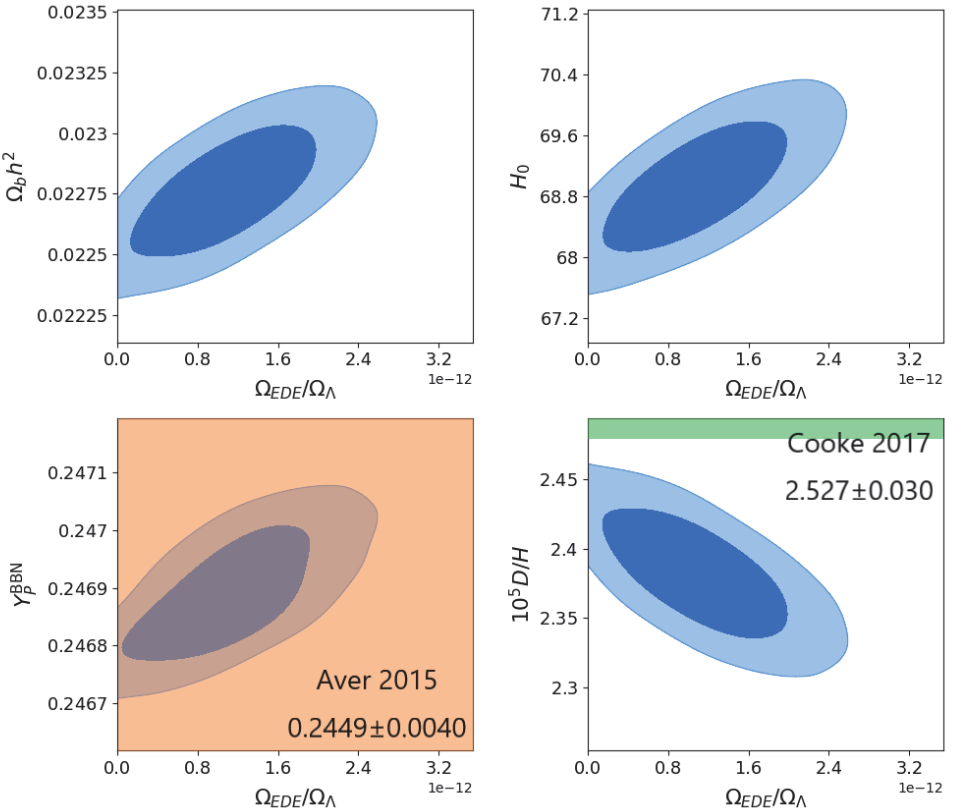


Fig. 3. The posterior and the constraints on the Early Dark Energy model

fit reduces the D/H abundance significantly. Thus, Early Dark Energy is disfavored from the deuterium measurement. This can be seen in Fig. 3.

6. Conclusion

The shorten sound horizon scale at the recombination epoch by introducing extra energy components such as the extra radiation or the Early Dark Energy is a promising solution to the Hubble tension. However, the compatibility with successful BBN would be another concern, because the extra radiation contribute the cosmic expansion or the inferred baryon asymmetry would differ from that in the ΛCDM.

We have performed analyses on the Early Dark Energy models and $N_{\rm eff}$ model with paying attention to the fit to BBN. Not only $N_{\rm eff}$ model but also the Early Dark Energy model is subject to the BBN constraints (as shown in Fig. 4.). Extra radiation is constrained by the helium abundance, while Early Dark Energy is constrained

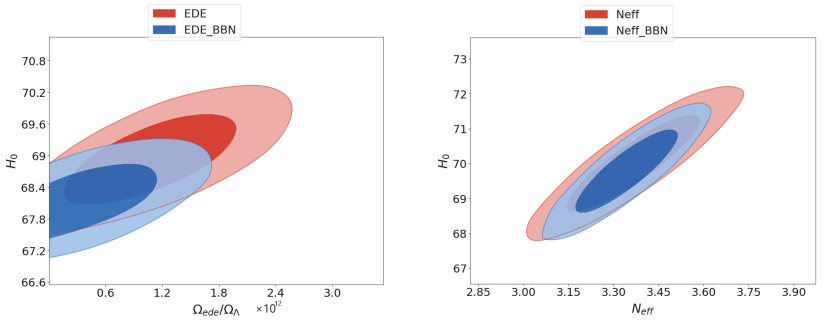


Fig. 4. Comparison of constraints based on data sets with and without BBN for the Early Dark Energy (left) model and the $N_{\rm eff}$ model (right)

by the deuterium abundance. Therefore, when we introduce extra components for the Hubble tension resolution, we should pay attention to the BBN as well.

Acknowledgments

This work of O.S. is supported in part the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants No. 19K03860, No. 19H05091, No. 19K03865, and No. 21H00060.

References

1. N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020).
2. S. Aiola *et al.* [ACT], JCAP **12**, 047 (2020).
3. G. E. Addison, D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw and J. L. Weiland, Astrophys. J. **853**, no. 2, 119 (2018).
4. N. Schöneberg, J. Lesgourgues and D. C. Hooper, JCAP **10**, 029 (2019).
5. A. G. Riess *et al.*, Astrophys. J. **855**, no. 2, 136 (2018).
6. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Astrophys. J. **876**, no. 1, 85 (2019).
7. K. C. Wong, S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger and M. W. Auger, *et al.* Mon. Not. Roy. Astron. Soc. **498**, no. 1, 1420-1439 (2020).
8. W. L. Freedman, B. F. Madore, D. Hatt, T. J. Hoyt, I. S. Jang, R. L. Beaton, C. R. Burns, M. G. Lee, A. J. Monson and J. R. Neeley, *et al.* Astrophys. J. **882**, 34 (2019).
9. V. Poulin, T. L. Smith, T. Karwal and M. Kamionkowski, Phys. Rev. Lett. **122**, no. 22, 221301 (2019).
10. P. Agrawal, F. Y. Cyr-Racine, D. Pinner and L. Randall, [arXiv:1904.01016 [astro-ph.CO]].
11. F. Niedermann and M. S. Sloth, Phys. Rev. D **103**, no. 4, L041303 (2021).
12. V. Poulin, T. L. Smith, D. Grin, T. Karwal and M. Kamionkowski, Phys. Rev. D **98**, no. 8, 083525 (2018) doi:10.1103/PhysRevD.98.083525.
13. S. Alexander and E. McDonough, Phys. Lett. B **797**, 134830 (2019) doi:10.1016/j.physletb.2019.134830.

14. M. X. Lin, G. Benevento, W. Hu and M. Raveri, Phys. Rev. D **100**, no. 6, 063542 (2019).
15. T. L. Smith, V. Poulin and M. A. Amin, Phys. Rev. D **101**, no. 6, 063523 (2020).
16. J. Sakstein and M. Trodden, Phys. Rev. Lett. **124**, no. 16, 161301 (2020).
17. A. Chudaykin, D. Gorbunov and N. Nedelko, JCAP **08**, 013 (2020).
18. M. Braglia, W. T. Emond, F. Finelli, A. E. Gumrukcuoglu and K. Koyama, Phys. Rev. D **102**, no. 8, 083513 (2020).
19. O. Seto and Y. Toda, Phys. Rev. D **103**, no. 12, 123501 (2021).
20. A. Lewis and S. Bridle, Phys. Rev. D **66**, 103511 (2002).
21. O. Pisanti, A. Cirillo, S. Esposito, F. Iocco, G. Mangano, G. Miele and P. D. Serpico, Comput. Phys. Commun. **178**, 956-971 (2008).
22. L. E. Marcucci, G. Mangano, A. Kievsky and M. Viviani, Phys. Rev. Lett. **116**, no. 10, 102501 (2016).
23. N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A8 (2020).
24. F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders and F. Watson, Mon. Not. Roy. Astron. Soc. **416**, 3017-3032 (2011).
25. A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden and M. Manera, Mon. Not. Roy. Astron. Soc. **449**, no. 1, 835-847 (2015).
26. S. Alam *et al.* [BOSS], Mon. Not. Roy. Astron. Soc. **470**, no. 3, 2617-2652 (2017).
27. D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan and A. G. Riess, *et al.* Astrophys. J. **859**, no. 2, 101 (2018).
28. E. Aver, K. A. Olive and E. D. Skillman, JCAP **07**, 011 (2015).
29. R. J. Cooke, M. Pettini and C. C. Steidel, Astrophys. J. **855**, no. 2, 102 (2018).

Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology

Nick E. Mavromatos*,†

* *Theoretical Particle Physics and Cosmology Group, Physics Department, King's College, London, Strand, London WC2R 2LS, UK,*

† *Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, 9, Heroon Polytechneiou Str., Zografou Campus, Athens 157 80, Greece*
E-mail: nikolaos.mavromatos@cern.ch

I review a string-inspired cosmological model with gravitational anomalies in its early epochs, which is based on fields from the (bosonic) massless gravitational multiplet of strings, in particular gravitons and Kalb Ramond (KR), string-model independent, axions (the dilaton is assumed constant). I show how condensation of primordial gravitational waves, which are generared at the very early eras immediately after the big bang, can lead to inflation of the so called running vacuum model (RVM) type, without external inflatons. The role of the slow-roll field is played by the KR axion, but it does not drive inflation. The non-linearities in the anomaly terms do. Chiral fermionic matter excitations appear at the end of this RVM inflation, as a result of the decay of the RVM vacuum, and are held responsible for the cancellation of the primordial gravitational anomalies. Chiral anomalies, however, survive in the post-inflationary epochs, and can lead to the generation of a non perturbative mass for the KR axion, which could thus play the role of dark matter in this Universe. As a result of the condensed gravitational anomaly, there is a Lorentz-invariance violating KR axion background, which remains undiluted during the RVM inflation, and can lead to baryogenesis through leptogenesis in the radiation era, in models with sterile right-handed neutrinos. I also discuss the phenomenology of the model in the modern era, paying particular attention to linking it with a version of RVM, called type II RVM, which arguably can alleviate observed tensions in the current-epoch cosmological data.

Keywords: String Cosmology; Running Vacuum; Inflation, Axions; Dark Matter.

1. Introduction

Although the plethora of the available cosmological data today[1] agree very well with the standard concordance model of Cosmology,[2] also known as ΛCDM, nonetheless there are compelling reasons for seeking alternative models, that go beyond it. Theoretically, given that the ΛCDM model is a global (de Sitter) solution of Einstein's general relativity (GR) theory, with a positive cosmological constant $\Lambda > 0$ added without explanation, one needs to seek microscopic frameworks of quantum gravity, which the GR can be embedded to, that could provide a detailed origin of the

observed dark energy component in the current-Universe energy budget.[a] Moreover, one of the most important drawbacks of GR *per se* is that it is a non-renormalizable theory, and as such, it cannot provide by itself a framework for quantising gravity along with the rest of the fundamental interactions in nature (assuming of course that gravity can admit conventional quantization; for alternative views, treating gravity as an entropic force, associated with changes of information pertaining to the positions of material bodies, see Ref. 4). On the observational side, there appear to exist discrepancies ("tensions"[5,6]) between the results of data analyses from Planck Collaboration, based on ΛCDM,[1] and local direct measurements of the Hubble parameter today, the so called H_0 tension,[7–9] but also tensions in the growth-of-galaxies data, the so-called σ_8 tension.[10] Thus, although such tensions could admit more mundane astrophysical explanations,[11–14] or even disappear in future data, being due to statistical uncertainties, nonetheless, several theoretical attempts are currently on the way in order to provide an explanation for them or alleviate them.

One phenomenological framework, which deals successfully with such an alleviation of both the H_0 and the σ_8 tensions, is a mildly modified version of the running vacuum model of cosmology (RVM),[15–20] the so-called type II RVM.[21] In this talk, I will review a microscopic framework for obtaining the RVM, actually a version of RVM somewhat analogous to the aforementioned type II RVM, which pertains to a string-inspired[22] cosmological model with *gravitational anomalies* in the early Universe.[23–28]

The structure of the talk is as follows: in the next section 2, I describe briefly the RVM framework, and then the string-inspired gravitational theory, with gravitational anomalies coupled to *massless* string-model independent (Kalb-Ramond (KR)) axions fields, which can give rise to an RVM inflation, without external inflatons. Only bosonic fields from the massless gravitational string multiplet enter the early-Universe era string-effective action as external fields. Primordial gravitational waves can lead to condensation of the gravitational anomalies, which in turn induces the RVM inflation. In section 3, I describe briefly the post inflationary epochs, placing the emphasis on how a gravitational-anomaly-induced Lorentz-violating (LV) KR axion background, which remans undiluted at the end of the RVM inflationary era, can lead to leptogenesis in models with right-handed sterile neutrinos (RHN) in their spectra. I also explain how the stringy KR axion field can play the rôle of Dark matter, by acquiring a non-perturbative mass during the Quantum Chromo-Dynamics (QCD) epoch of the Universe. In section 4, I speculate on potential links of this stringy RVM effective theory to a modified version of RVM, somewhat analogous (but with crucial differences) to the aforementioned type-II RVM. The crucial

[a]See, however, the arguments of Ref. 3 claiming that the "observed" dark energy might be an artefact of supernova data interpretation based on the assumption of homogeneity and isotropy of the universe at large (cosmological) scales, in the context of the Friedman-Lemaitre-Robertson-Walker (FLRW) framework. In our approach here, the large-scale isotropy and homogeneity of the Universe is assumed.

modifications are due to quantum-graviton-mode integration in the respective path integral of the string-inspired theory, and could lead to observable departures from ΛCDM and potential alleviation of tensions in the current-era data. Finally, section 5 contains our conclusions and outlook.

2. Running Vacuum Model (RVM) Framework, its String-Theory Embedding and Inflation from Gravitational Anomalies

I commence the discussion by first summarising briefly the basic features of the RVM framework. This is a phenomenologically successful effective cosmological framework, which leads to observable, in principle, deviations from ΛCDM,[19,20] and alleviation of the current cosmological tensions.[21] It describes a smooth evolution of the Universe from a dynamical inflationary phase till the present era.[18]

2.1. *Features of RVM cosmology*

The basic feature of the RVM cosmology is that the vacuum energy density is a function of even powers of the Hubble parameter at a given era of the Universe, $H(t)$ (with t the cosmic time), due to general covariance. To describe the entire evolution of the Universe from inflation till the present epoch, it is sufficient to truncate this series to quartic powers of $H(t)$, thus writing::[15–18]

$$\rho^{\Lambda}_{\rm RVM}(H) = \frac{\Lambda(H)}{\kappa^2} = \frac{3}{\kappa^2}\left(c_0 + \nu H^2 + \alpha\frac{H^4}{H_I^2} + \dots\right) > 0\,, \tag{1}$$

with κ the four-dimensional gravitational constant $\kappa = \sqrt{8\pi\,\mathrm{G}} = M_{\rm Pl}^{-1}$, with $M_{\rm Pl} = 2.43 \times 10^{18}$ GeV (we work in natural units $\hbar = c = 1$ throughout this work), $H_I \sim 10^{-5} M_{\rm Pl}$, the inflationary scale, as inferred from the latest Planck-Collaboration data,[1] and $c_0 \geq 0$, ν and α constants, throughout the Universe evolution. In the conventional RVM, $\nu > 0$ and $\alpha > 0$, while $c_0 > 0$ is the current-epoch cosmological constant.[1,2] The $\dots$ in (1) denote terms of higher orders in $H^2(t)$. In general there is also dependence on $\dot{H}$ (with the overdot denoting cosmic-time-t derivative), but for a given era of the Universe, one can express $\dot{H}$ in terms of the cosmic deceleration of that era (assumed approximately constant) and of H^2, hence the expression (1) suffices for our purposes here. The defining feature of the RVM is its de-Sitter type equation of state,[15]

$$p^{\rm vac}_{\rm RVM}(H(t)) = -\rho^{\rm vac}_{\rm RVM}(H(t)) < 0, \tag{2}$$

despite the fact that the pressure $p^{\rm vac}_{\rm RVM}$ and energy $\rho^{\rm vac}_{\rm RVM}$ densities are time-dependent functions of $H(t)$. We note for completeness that a similar feature characterises the super-critical (Liouville) string cosmologies, where the cosmic time is identified with (the zero mode of) a world-sheet renormalization-group scale.[29]

From the conservation of the total stress tensor of the vacuum plus radiation and matter excitations, the latter being characterised by an equation of state w_m, one

can easily arrive at the following evolution equation for the Hubble parameter:[18]

$$\dot{H} + \frac{3}{2}\,(1+\omega_m)\,H^2\left(1-\nu-\frac{c_0}{H^2}-\alpha\,\frac{H^2}{H_I^2}\right)=0\ . \tag{3}$$

At early stages of the Universe, one may ignore the c_0 term, thus arriving at a solution for $H(a)$ as a function of the scale factor a (in units of the present-era scale factor):[18,19]

$$H(a)=\left(\frac{1-\nu}{\alpha}\right)^{1/2}\ \frac{H_I}{\sqrt{D\,a^{3(1-\nu)(1+\omega_m)}+1}}\,, \tag{4}$$

where $D>0$ is an integration constant. For the early Universe, one may assume without loss of generality that $D\,a^{3(1-\nu)(1+\omega_m)}\ll 1$, which from (4) leads to an (unstable) dynamical early de Sitter phase, characterised by an approximately constant $H_{\rm de\ Sitter}\simeq\left(\frac{1-\nu}{\alpha}\right)^{1/2}\ H_I$. This inflationary era is due to the non-linear terms H^4 in (1), and is not characterised by external inflaton fields.

After inflation, the RVM evolution (3) is characterised by radiation-dominance ($w_m=1/3$) and, subsequently, by a matter-dominated ($w_m=0$) era, both occurring for considerably larger values of the scale factor than the RVM inflation. In such cases, the corresponding quantities $D\,a^{3(1-\nu)(1+\omega_m)}\gg 1$. One can therefore determine the RVM vacuum energy density (1) during radiation as[18,30] $\rho^{\rm vac\ radiation\ era}_{\rm RVM}\simeq\frac{3H_I^2}{\kappa^2\,\alpha}\frac{1}{(1+Da^4)^2}$, and the radiation energy density $\rho_{\rm rad}(a)\simeq\frac{3H_I^2}{\kappa^2\,\alpha}\frac{Da^4}{(1+Da^4)^2}$, thus recovering the familiar a^{-4} scaling for $Da^4\gg 1$. As the evolution continues, matter starts to dominate ($\omega_m=0$), and at that point the cosmological constant term c_0 can no longer be ingored, leading eventually to a modern-era scaling of the Hubble parameter $H\ll H_I$ (obtained as the solution of the evolution equation (3) at this era):[18,30] $H^2_{\rm modern}(a)\simeq\frac{H_0^2}{1-\nu}\Big[(1-\Omega^0_\Lambda)\,a^{-3(1-\nu)}+\Omega^0_\Lambda-\nu\Big]$, with H_0 the present-day Hubble parameter, and $\Omega^0_\Lambda=\frac{c_0+H_0^2\nu}{H_0^2}$. Notice that the presence of the parameter $\nu>0$ implies observable deviations from the ΛCDM concordance model of cosmology[19,20] (which is recovered in the limit $\nu=0$, in which case Ω^0_Λ coincides with the present-era cosmological constant (in units of the current-era critical density)).

We remark at this stage that the non-linearities-induced inflation (4) is different[19,26,30] from Starobinsky's model of inflation,[31] which is also due to conformal-("trace")-anomaly-induced non-linear curvature corrections to Einstein's GR theory. Starobinsky's model does not have in its energy density the crucial H^4 term of the RVM (*cf.* (1)). As discussed in Refs. 19, 26, it is an inflation characterised mainly by $\dot{H}\simeq$ constant rather than $H\simeq$ constant, which is the case of RVM. The crucial for inflation H^4 term of the RVM energy density (1) is also missing in the case of a quantum field theory of a scalar field non-minimally coupled to a gravity background.[32] As we shall discuss below, a direct (geometric in origin) induction of an RVM H^4 term in the corresponding energy density can be achieved through *condensates* of *gravitational anomalies*, induced by primordial gravity waves,[33] in the string-inspired model for cosmology discussed in Refs. 23, 24, 25, 26, 27, 28, which we now come to describe briefly.

2.2. *String-inspired RVM, gravitational anomalies and inflation*

The basic assumption[24] towards the construction of the string-inspired cosmology model that will lead to RVM, is that at early stages of the Universe, only fields from the bosonic massless gravitational multiplet of strings[22] appear as external fields. Assuming constant dilatons, this assumption implies that the effective action in the four dimensional space time, after string compactification, will consist only of gravitons and antisymmetric tensor fields $B_{\mu\nu} = -B_{\nu\mu}$, $\mu, \nu = 0, \dots 3$. As a result of an appropriate U(1) gauge symmetry in the closed string sector:[22] $B_{\mu\nu} \to B_{\mu\nu} + \partial_\mu \theta_\nu(X) - \partial_\nu \theta_\mu(X)$, where $\theta_\mu(X)$ are gauge parameters, the field $B_{\mu\nu}$ appears in the effective four space-time dimensional action only through its field strength. Due to the Green-Schwarz anomaly cancellation mechanism,[22] the latter is modified by appropriate gauge and gravitational Chern-Simons terms. In our model, only the latter are present in the early Universe,[24] and as such the modified field strength $\mathcal{H}_{\mu\nu\rho}$ of the field $B_{\mu\nu}$ reads (in differential form language):[22] $\mathcal{H} = \mathbf{dB} + \frac{\alpha'}{8\,\kappa}\,\Omega_{3\mathrm{L}}$, with $\alpha' = M_s^{-2}$ the Regge slope, M_s the string mass scale, and the gravitational Chern Simons term is defined as $\Omega_{3\mathrm{L}} = \omega^a{}_c \wedge \mathbf{d}\,\omega^c{}_a + \frac{2}{3}\omega^a{}_c \wedge \omega^c{}_d \wedge \omega^d{}_a$, where $\wedge$ denotes the exterior product among forms, $\mathbf{d}$ is the exterior derivative one form. and $\omega^a_{\mu\ b}$ is the spin connection, with Latin indices $a, b = 0, \dots 3$ denoting tangent space indices. The field strength satisfies a Bianchi identity:

$$\varepsilon_{abc}{}^{\mu}\,\mathcal{H}^{abc}{}_{;\mu} = \frac{\alpha'}{32\,\kappa}\,\sqrt{-g}\,R_{\mu\nu\rho\sigma}\,\widetilde{R}^{\mu\nu\rho\sigma} \tag{5}$$

where the semicolon denotes covariant derivative with respect to the standard Christoffel connection, $\varepsilon_{\mu\nu\rho\sigma}$ is the gravitationally covariant Levi-Civita tensor density, totally antisymmetric in its indices, and $\widetilde{R}_{\mu\nu\rho\sigma} = \frac{1}{2}\varepsilon_{\mu\nu\lambda\pi}R^{\lambda\pi}{}_{\rho\sigma}$ is the dual Riemann curvature tensor (we follow the conventions of Ref. 24). To lowest (quadratic) order in derivatives, the target-space effective action (for constant dilatons) reads:

$$S_B = \int d^4x\sqrt{-g}\Big(\frac{1}{2\kappa^2}[-R - \frac{1}{6}\,\mathcal{H}_{\lambda\mu\nu}\mathcal{H}^{\lambda\mu\nu} + \dots\Big), \tag{6}$$

with the $\dots$ representing higher-derivative terms. We mention at this stage the rôle of $\mathcal{H}_{\mu\nu\rho}$ as a totally antisymmetric component of torsion in string-inspired effective gravitational theories,[22,34] in the sense that the action (6) can be expressed in terms of a generalised curvature scalar with respect to such a torsion. As we shall describe below, the torsion is associated with an axion field.[34]

Indeed, on implementing (5) as a constraint in the respective path-integral of the action (7) by means of a pseudoscalar Lagrange mutl[plier field $b(x)$, and integrating out $\mathcal{H}$, one obtains the effective action[34,35]

$$\begin{aligned} S_B^{\mathrm{eff}} &= \int d^4x\sqrt{-g}\Big[-\frac{1}{2\kappa^2}\,R + \frac{1}{2}\,\partial_\mu b\,\partial^\mu b + \sqrt{\frac{2}{3}}\,\frac{\alpha'}{96\,\kappa}\,b(x)\,R_{\mu\nu\rho\sigma}\,\widetilde{R}^{\mu\nu\rho\sigma} + \dots\Big] \\ &= \int d^4x\,\sqrt{-g}\Big[-\frac{1}{2\kappa^2}\,R + \frac{1}{2}\,\partial_\mu b\,\partial^\mu b - \sqrt{\frac{2}{3}}\,\frac{\alpha'}{96\,\kappa}\,\mathcal{K}^\mu(\omega)\,\partial_\mu b(x) + \dots\Big], \end{aligned} \tag{7}$$

where we took into account that the gravitational anomaly is a total derivative $\sqrt{-g}\,R_{\mu\nu\rho\sigma}\,\widetilde{R}^{\mu\nu\rho\sigma} = \sqrt{-g}\,\mathcal{K}^{\mu}(\omega)_{;\mu} = \partial_\mu\Big(\sqrt{-g}\,\mathcal{K}^{\mu}(\omega)\Big) = -2\,\partial_\mu\Big[\sqrt{-g}\,\varepsilon^{\mu\nu\alpha\beta}\,\omega^{ab}_{\nu}\,\Big(\partial_\alpha\,\omega_{\beta ab} + \frac{2}{3}\,\omega_{\alpha a}{}^{c}\,\omega_{\beta cb}\Big)\Big]$. One observes from (7) that the Lagrange multiplier field $b(x)$ became a fully dynamical axion-like field, the so called Kalb-Ramond (KR) or string-model-independent axion.[34,35] Classically, as follows from the saddle-point of the path integral over $\mathcal{H}$, one has the relation:[34] $3\sqrt{2}\,\partial_\sigma\bar{b} = -\sqrt{-g}\,\epsilon_{\mu\nu\rho\sigma}\,\overline{\mathcal{H}}^{\mu\nu\rho}$, where the bar denotes classical fields. Hence the association of the torsion with the axion field $b(x)$.

The effective action (7) forms the basis of our cosmological model.[24] In a FLRW space-time background the gravitational anomaly terms vanish.[36] This is not true, however, in the presence of primordial gravity waves (GW) perturbations, which violate CP symmetry:

$$ds^2 = dt^2 - a^2(t)\Big[(1 - h_+(t,z))\,dx^2 + (1 + h_+(t,z))\,dy^2 + 2h_\times(t,z)\,dx\,dy + dz^2\Big], \tag{8}$$

using standard notation for the graviton polarizations of the GW. In the metric (8), the gravitational anomaly term is *non zero*.[33] In fact, on assuming an inflationary space time with constant Hubble parameter $H \simeq$ constant, one can compute the anomaly condensate in the presence of GW CP-violating perturbations:[33]

$$\begin{aligned}\langle R_{\mu\nu\rho\sigma}\,\widetilde{R}^{\mu\nu\rho\sigma}\rangle &\simeq \frac{16}{a^4}\,\kappa^2\int^{\mu}\frac{d^3k}{(2\pi)^3}\,\frac{H^2}{2\,k^3}\,k^4\,\Theta = \frac{1}{\pi^2}\Big(\frac{H}{M_{\rm Pl}}\Big)^2\,\mu^4\,\Theta \\ &= \frac{2}{3\pi^2}\,\frac{1}{96\times 12}\,\Big(\frac{H}{M_{\rm Pl}}\Big)^3\,\Big(\frac{\mu}{M_{\rm Pl}}\Big)^4\,M_{\rm Pl}\,\times\,\mathcal{K}^0(t)\,,\end{aligned} \tag{9}$$

to leading order in the slow-roll parameter

$$\Theta = \sqrt{\frac{2}{3}}\,\frac{\alpha'\,\kappa}{12}\,H\,\dot{\bar{b}} \ll 1\,, \tag{10}$$

with the overdot denoting derivative with respect to the cosmic time t. The quantity μ is an UltraViolet (UV) cutoff in the momenta of the graviton modes.

Assuming isotropy and homogeneity in the Universe, the condensate (9) implies the following solution of the Euler-Lagrange equation of the KR axion $b(t)$:[24]

$$\dot{\bar{b}} = \sqrt{\frac{2}{3}}\,\frac{\alpha'}{96\,\kappa}\,\mathcal{K}^0 \simeq \text{constant}, \tag{11}$$

provided $\mu/M_s \simeq 15\,\Big(M_{\rm Pl}/H\Big)^{1/2}$. Combining the slow-roll condition (10), with the transplanckian conjecture, *i.e.* the absence of transplanckian graviton modes ($\mu \leq M_{\rm Pl}$), one can show that for the inflationary Hubble parameter $H = H_I = 10^{-5}\,M_{\rm Pl}$, as dictated by data,[1] the string mass scale is restricted to the range:[27,28]

$$2.6\times 10^{-5}\,M_{\rm Pl} \lesssim M_s \lesssim 10^{-4}\,M_{\rm Pl}. \tag{12}$$

Eq. (11) admits a spontaneous-Lorentz-violating (LV) KR axion background solution, parametrised as:[24,26]

$$b(t) = \overline{b}(0) + \sqrt{2\epsilon} H M_{\rm Pl}\, t, \quad H \simeq \text{constant}, \tag{13}$$

with $\overline{b}(0)$ a boundary condition for the field $b(x)$ at the onset of inflation at $t = 0$. For phenomenological reasons[1] we may take[24] $\epsilon = \mathcal{O}(10^{-2})$. Thus, the KR axion is a slow-roll field, but, as we shall discuss next, it does not drive inflation.

To this end, we remark that the condensate (9) implies a condensate of the quantity:[24] $\langle b(t) R_{\mu\nu\rho\sigma} \widetilde{R}^{\mu\nu\rho\sigma} \rangle$, which remains approximately constant during the entire duration of inflation, thus corresponding to a de-Sitter-type (cosmological-constant) contribution to the effective target-space action (7), provided that

$$\frac{|\overline{b}(0)|}{M_{\rm Pl}} \gg \sqrt{2\,\epsilon}\,\mathcal{N} = \mathcal{O}(10), \quad \overline{b}(0) < 0, \tag{14}$$

where the end of inflation[37] has been set to $t_{\rm inf} H \simeq \mathcal{N}$, with $\mathcal{N}$ the number of e-foldings of the inflationary era, which phenomenologically[1] lies in the range $\mathcal{N} = 60 - 70$.

Expanding the effective action (7) about such de-Sitter background configurations, one can show that the equation of state of the total energy $\rho_{\rm total} = \rho^b + \rho^{\rm gCS} + \rho^{\rm condensate}$ and pressure $p_{\rm total} = p^b + p^{\rm gCS} + p^{\rm condensate}$ densities of this fluid, consisting of contributions from the KR axion (superscript b), fluctuations of the gravitational anomaly (Chern-Simons) term (superscript gCS) and the condensate term (superscript "condensate"), assumes an RVM form:[27]

$$p_{\rm total} = -\rho_{\rm total} < 0, \tag{15}$$

with the total energy density being dominated by the H^4 term during the inflationary era

$$0 < \rho_{\rm total} \simeq 3\kappa^{-4} \Big[-1.65 \times 10^{-3} \Big(\kappa\, H\Big)^2 + \frac{\sqrt{2}}{3} |\overline{b}(0)|\, \kappa \times 5.86 \times 10^6 \, (\kappa\, H)^4 \Big] \tag{16}$$

under the condition (14). The reader should notice that the coefficient of the H^2 term is negative, due to the contribution of the anomalous gravitational Chern-Simons terms, and this is a difference from the conventional RVM form (1). We stress that it is the dominance of the condensate term that makes the total energy density (16) positive, and thus leads to a proper RVM fluid. Without the formation of this condensate, the equation of state of the remaining terms curiously acquire the form of an RVM-like (i.e. de Sitter-type) "phantom-matter":[27,38,39]

$$p^b + p^{\rm gCS} = -(\rho^b + \rho^{\rm gCS}) > 0\ , \tag{17}$$

violating the weak energy conditions.[b] Such exotic matter can be used for the stabilisation of traversable wormholes,[40] so in this respect our anomalous gravitational theory, in conditions in which the condensate $\langle b(x)\, R_{\mu\nu\rho\sigma}\, \widetilde{R}^{\mu\nu\rho\sigma}\rangle$ is not formed, might find applications to this problem as well.

The dominance of the non-linear H^4 term in (16) at early eras of this string-inspired Universe, will lead to an RVM type inflation (4), as we discussed in the previous subsection 2.1. It is in this sense that an inflationary phase arises dynamically in a self consistent way in our approach, so our assumption of a constant H in the computation of (9) is justified *a posteriori*.

2.3. *Potential origin of GW: the dynamically-broken supergravity example*

So far, we have assumed the existence of primordial GW perturbations (8), without examining in detail their microscopic origin. In Refs. 26, 27, we have discussed several scenarios which lead to the creation of GW in a pre-RVM-inflationary era of our string-inspired cosmology. One of the most relevant ones to our model, consistent with the basic assumption that only fields from the massless gravitational string multiplet appear as external fields at the very early stages of the Universe, is the scenario considering an underlying superstring theory, whose low energy limit will result in a supergravity model. Although the precise supergravity model depends crucially on the microscopic higher-dimensional string theory considered, for our qualitative purposes here, we may consider a simplified but highly non-trivial example for our (3+1)-dimensional gravitational theory at the very early stages of our cosmology, that of $N = 1$ (3+1)-dimensional supergravity.[41]

The model contains massless spin-2 gravitons, $g_{\mu\nu}(x)$, $\mu, \nu = 0, \ldots 3$, and their spin-3/2 supersymmetric partners, the gravitinos, $\psi_\mu(x)$. The supergravity can be broken dynamically in this model as a consequence of gravitino condensation.[42,43] The corresponding effective potential $\widetilde{V}(\sigma)$ of the gravitino-condensate scalar field

[b]The computation leading to (17) is based on the fact that the graviton variations of the anomalous gravitational Chern-Simons terms in (7) yield the Cotton tensor $C_{\mu\nu}$:[36]

$$\delta\Big[\int d^4x\sqrt{-g}\, b\, R_{\mu\nu\rho\sigma}\, \widetilde{R}^{\mu\nu\rho\sigma}\Big] = 4\int d^4x\sqrt{-g}\, \mathcal{C}^{\mu\nu}\, \delta g_{\mu\nu} = -4\int d^4x\sqrt{-g}\, \mathcal{C}_{\mu\nu}\, \delta g^{\mu\nu},$$

where

$$\mathcal{C}^{\mu\nu} \equiv -\frac{1}{2}\Big[v_\sigma\left(\varepsilon^{\sigma\mu\alpha\beta}R^{\nu}{}_{\beta;\alpha} + \varepsilon^{\sigma\nu\alpha\beta}R^{\mu}{}_{\beta;\alpha}\right) + v_{\sigma\tau}\left(\widetilde{R}^{\tau\mu\sigma\nu} + \widetilde{R}^{\tau\nu\sigma\mu}\right)\Big],$$

with $v_\sigma \equiv \partial_\sigma b = b_{;\sigma}$, $v_{\sigma\tau} \equiv v_{\tau;\sigma} = b_{;\tau;\sigma}$. This implies that the Einstein's equations stemming from (7) read: $R^{\mu\nu} - \frac{1}{2}\, g^{\mu\nu}\, R - \mathcal{C}^{\mu\nu} = \kappa^2\, T_b^{\mu\nu}$, where $T_b^{\mu\nu}$ is the KR axion stress tensor. Taking into account conservation properties of the Cotton tensor,[36] $\mathcal{C}^{\mu\nu}{}_{;\mu} = \frac{1}{8}\, v^\nu\, R^{\alpha\beta\gamma\delta}\, \widetilde{R}_{\alpha\beta\gamma\delta}$, we observe that there is an exchange of energy between the KR axion and the gravitational anomaly, when $\mathcal{C}_{\mu\nu}$ is non trivial, as is the case of GW perturbations of the metric (*cf.* (8)). We note that classically the Cotton tensor is traceless $g^{\mu\nu}\mathcal{C}_{\mu\nu} = 0$. However, the formation of the condensate $\langle b(t) R_{\mu\nu\rho\sigma}\, \widetilde{R}^{\mu\nu\rho\sigma}\rangle$, as a result of (9), introduces in a sense a quantum-gravity-induced trace of $\langle \mathcal{C}^\mu_\mu\rangle$, which leads to the addition of a de Sitter type cosmological-constant term in the effective action (7).[24]

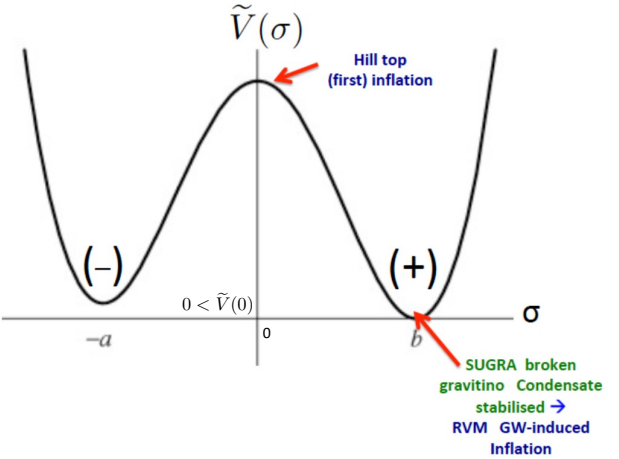


Fig. 1. The biased double-well gravitino potential $\widetilde{V}(\sigma)$ for dynamical supergravity breaking that could characgterise the early epochs of the string-inspired RTVM Cosmology.[26] The bias between the two ($\pm$) vacua is assumed to be due to percolation effects in the early Universe,[45,46] and leads to the formation of unstable domain walls, whose collapse or collisions lead to GW. There is also a first hill-top inflation near the origin $\sigma = 0$.

$\sigma = \langle \overline{\psi}_\mu \psi^\mu \rangle$, evaluated at one loop in a local de Sitter space-time background,[44] after graviton and gravitino Euclidean path-integration, assumes a double-well shape (see fig. 1), defining two local minima, at each of which the potential takes on non-negative values, compatible with the breaking of global and local supersymmetry. There could be a statistical bias, in the sense of unequal occupation numbers of these two local minima/"vacua", as a result of percolation effects.[45,46] Such a situation leads to the formation of "biased" domain walls in the theory, whose non-spherical collapse or collision, leads to GW. The $N = 1$ supergravity model is also characterised by a very early hill-top inflationary phase,[47] which is not necessarily slow roll, and has no observable consequences.

This first inflationary phase is consistent with the study of the theory in a local de Sitter (LDS) background, in the absence of a mass parameter,[44] given the existence of an appropriate general coordinate transformation[48] that maps the LDS space-time without a mass parameter to a cosmological (global) de Sitter one (GDS), and leaves the effective action invariant.[c]

[c]The local de Sitter-Schwarzschild metric with a mass parameter M is described by the following invariant element

$$ds^2 = \left(1 - 2\frac{M}{\bar{r}} - \frac{\Lambda}{3}\bar{r}^2\right) c^2 dT^2 - \left(1 - 2\frac{M}{\bar{r}} - \frac{\Lambda}{3}\bar{r}^2\right)^{-1} d\bar{r}^2 - \bar{r}^2(d\bar{\theta}^2 + \sin^2\bar{\theta}\mathrm{d}\bar{\phi}^2) \qquad (18)$$

in de Sitter-Schwarzschild coordinates. In the case $M = 0$, as required by the isotropy and homogeneity of space, the following transformation[48]

$$x^\mu \equiv \{cT, \bar{r}, \bar{\theta}, \bar{\phi}\} \to x'^\mu \equiv \{ct, r, \bar{\theta}, \bar{\phi}\} \text{ (comoving frame) :}$$
$$t = T + \frac{1}{2H}\ln\left(1 - H^2\bar{r}^2\right), \quad r = \frac{\bar{r}}{\sqrt{1 - H^2\bar{r}^2}} e^{-HT} = \bar{r}\, e^{-Ht}, \quad H^2 \equiv \frac{\Lambda}{3} > 0\,. \qquad (19)$$

This hill-top inflation is a different phase from our RVM inflation. In fact, it ensures that any spatial inhomogeneities or isotropies are washed out well before entrance into the RVM inflationary phase, so the assumptions and calculations of Refs. 24, 26 are formally justified. The RVM inflation occurs[26] at the phase where the gravitino has settled in its lowest minimum (see fig. 1), is massive and the supergravity is dynamically broken. This phase coincides with the exit phase from the hill-top inflation.[47] This exit phase is characterised by the creation of the biased domain walls and GW, as a consequence of wall collisions. The massive gravitino can be integrated out of the path integral.

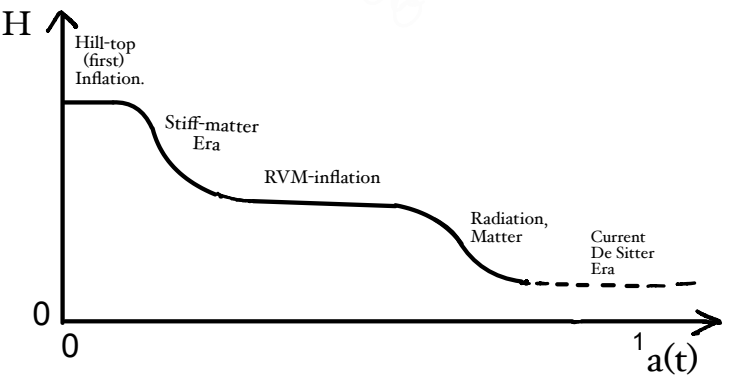


Fig. 2. Schematic evolution of the Hubblle parameter (H) as a function of the scale factor a(t) in the string-inspired RVM universe,[27] from the Big-Bang (a=0) till the current epoch (a=1, in units of today's scale factor).

Embedding this simple supergravity model into our string framework implies[26,28] that the exit from the hill-top inflation may be succeeded by dominance of the KR axion matter,[26] which, due to the lack of a potential, constitutes a stiff matter fluid (with equation of state $w = +1$).[49,50] This era interpolates between the first-hill-top inflation and the GW-induced RVM inflation (see fig. 2). Indeed, the presence of GW as a consequence of domain-wall collapse or collisions, implies the appearance of CP-violating anomalous gravitational Chern Simons terms coupled

where t and r denote co-moving frame time and radial space coordinates, respectively, maps the metric (18) to a standard cosmological de Sitter space-time:

$$ds^2 = c^2\,dt^2 - a(t)^2\left[dr^2 + r^2 d\Omega^2\right], \quad d\Omega^2 = d\bar{\theta}^2 + \sin^2\bar{\theta}\mathrm{d}\bar{\phi}^2\ , \tag{20}$$

where $a(t) = e^{\sqrt{\frac{\Lambda}{3}}\,t} \equiv e^{H\,t}$ (H = constant), is the exponentially expanding scale factor of the de-Sitter/FLRW (inflationary) Universe. The above result is exact, valid for every non-negative value of the cosmological constant $\Lambda \geq 0$. Notice that in this form of the metric there is no preferred origin of space, as appropriate for a space-time generated by a uniform homogeneous and isotropic fluid. The apparent existence of an origin at zero for the radial coordinate in the local form of the pure de-Sitter metric (18) with $M = 0$ is thus an artefact of an inappropriate choice of coordinates, much like the internal observer de Sitter horizon at $\sqrt{\Lambda/3}$.

to the KR axions in the effective action (7). As the cosmic time elapses, conditions for the condensation of GW can develop, leading to the anomaly condensate (9), and the eventual RVM inflationary era, driven by the H^4 term in (16).

Before closing this subsection, we would like to discuss some formal properties of the one-loop corrected effective action of supergravity in the de-Sitter background, which we shall make use of in section 4. As discussed in Ref. 43 (see also Ref. 30), the Euclidean (E) one-loop effective action of the dynamically-broken $N = 1$ supergravity model, in a given gauge, evaluated in a de Sitter background, with (one-loop corrected) cosmological constant $1 \gg \Lambda > 0$ reads:

$$\Gamma^{(\mathrm{E})} \simeq S_{\mathrm{cl}} - \frac{24\pi^2}{\Lambda^2}\big(\alpha_0^F + \alpha_0^B + \big(\alpha_1^F + \alpha_1^B\big)\,\Lambda + \big(\alpha_2^F + \alpha_2^B\big)\,\Lambda^2 + \dots\big)\ , \tag{21}$$

where the superscripts $B(F)$ denote terms arising from integration of massless (quantum) gravitons (B) (and gravitinos (F)), f is the scale of supergravity (and global suprsymmetry) breaking, $\sigma_c < f$,[43] is the value of the gravitino condensate field at the minimum of its effective potential (see fig. 1) and S_{cl} denotes the classical supergravity action with tree-level (bare) cosmological constant $\Lambda_0 < 0$

$$\frac{\Lambda_0}{\kappa^2} \equiv \sigma_c^2 - f^2\ , \tag{22}$$

which is necessarily *negative* for reasons of incompatibility of (unbroken) supergravity with de Sitter vacua.[41–43] The one-loop renormalised cosmological constant Λ, though, is *positive*, due to quantum corrections (*cf.* (25)), compatible with dynamically broken supergravity. Taking into account that the Euclidean de Sitter volume is[44] $24\pi^2/\Lambda^2 \to \int d^4x\sqrt{\widehat{g}_{\mathrm{E}}}$, we may express the effective action (21) in a generally covariant form (using the notation $\widehat{\dots}$ for quantities evaluated in the background de Sitter space-time):

$$\Gamma^{(\mathrm{E})} \simeq -\frac{1}{2\kappa^2}\int d^4x\sqrt{\widehat{g}_{\mathrm{E}}}\left[\left(\widehat{R} - 2\Lambda_1\right) + \alpha_1\,\widehat{R} + \alpha_2\,\widehat{R}^2\right]\ , \tag{23}$$

where we have replaced Λ by the curvature scalar,

$$\widehat{R} = 4\Lambda. \tag{24}$$

The various quantities entering (23) are given by[30,43]

$$\Lambda_1 = -\,\kappa^2\,\left(-\frac{\Lambda_0}{\kappa^2} + \alpha_0^F + \alpha_0^B\right)\ , \tag{25}$$

where

$$\begin{aligned}
\alpha_0^F &= \kappa^4\,\sigma_c^4\left(0.100\,\ln\left(\frac{\kappa^2\,\sigma_c^2}{3\mu^2}\right) + 0.126\right)\ ,\\
\alpha_0^B &= \kappa^4\,\left(f^2 - \sigma_c^2\right)^2\left(0.027 - 0.018\ln\left(\frac{3\kappa^2\left(f^2 - \sigma_c^2\right)}{2\mu^2}\right)\right)\ ,
\end{aligned} \tag{26}$$

and

$$\alpha_1 = \frac{\kappa^2}{2}\left(\alpha_1^F + \alpha_1^B\right)\ ,\quad \alpha_2 = \frac{\kappa^2}{8}\left(\alpha_2^F + \alpha_2^B\right)\ , \tag{27}$$

with

$$\alpha_1^F = 0.067\,\kappa^2\sigma_c^2 - 0.021\,\tilde{\kappa}^2\sigma_c^2 \ln\left(\frac{\Lambda}{\mu^2}\right) + 0.073\,\kappa^2\sigma_c^2 \ln\left(\frac{\kappa^2\sigma_c^2}{\mu^2}\right) \ ,$$
$$\alpha_2^F = 0.029 + 0.014 \ln\left(\frac{\kappa^2\sigma_c^2}{\mu^2}\right) - 0.029 \ln\left(\frac{\Lambda}{\mu^2}\right) \ ,$$
$$\alpha_1^B = -0.083\Lambda_0 + 0.018\,\Lambda_0 \ln\left(\frac{\Lambda}{3\mu^2}\right) + 0.049\,\Lambda_0 \ln\left(-\frac{3\Lambda_0}{\mu^2}\right) \ ,$$
$$\alpha_2^B = 0.020 + 0.021 \ln\left(\frac{\Lambda}{3\mu^2}\right) - 0.014 \ln\left(-\frac{6\Lambda_0}{\mu^2}\right) \ , \tag{28}$$

with the replacement (24) understood in all the above expressions.

The scale μ^2 is a proper-time ultraviolet cutoff, used to regularise UV divergences.[42–44] Supergravity breaks dynamically for μ^2 close to Planck scale.[42] We then set from now on

$$\mu^2 \sim M_{\rm Pl}^2 = \kappa^{-2}. \tag{29}$$

It can be arranged that the masses of the gravitino condensate and the gravitino become very large compared to the low-energy mass scales. They can be taken to be above the grand-unification scale, even close to Planck scale, for our purposes,[26–28] provided the scale f is chosen appropriately. Since the effective action (23) is evaluated after one-loop integration of (weak) graviton perturbations about the de Sitter background, it is actually a quantum-gravity effective action.

In the broken supergravity RVM phase, where the hypermassive gravitino and gravitino-condensate fields, with masses of order of the Planck mass, are integrated out from the low-energy effective action, leading to Planck-mass suppressed terms, the only leading quantum corrections are those due to massless graviton integration. As can be seen from the expressions (25),(26),(27),(28), such corrections involve $\widehat{R}^n \ln(\kappa^2\widehat{R})$, $n = 1, 2$ terms (*cf.* coefficients with superscript B). Passing onto a slowly-varying global de Sitter background (via the coordinate transformations[48] (19)), we may approximately replace $\widehat{R} \to 12H^2$, where H is the approximately constant inflationary Hubble parameter.

In the framework of our cosmology, this RVM phase will also involve the KR axion and gravitational anomaly terms, in addition to the aforementioned quantum-gravity corrections of the effective action (23) contributing terms of the form $H^{2n'}\ln(\kappa^2 H^2)$, $n' = 1, 2$. From the pertinent expressions (25),(26),(27),(28), it becomes clear that, on assuming $\kappa^2|\Lambda_0| < 1$, for the scale of the bare cosmological constant (22), and taking into account (14), these quantum-gravity-induced corrections are subdominant compared to the gravitational-anomaly-condensate-induced H^4 term in (16). Thus, our discussion and conclusions on the rôle of anomalies in inducing an RVM inflation are not affected.[27,28]

As we speculated though in Refs. 27, 28, quantum-gravity corrections of the form $H^2\ln(\kappa^2 H^2)$ could lead to observable deviations from the ΛCDM in the modern era (which again is characterised by an approximately de Sitter cosmological space-time

background, with $H \simeq H_0$). In particular, they could lead to a modified version of RVM, somewhat analogous to the type-II RVM considered in Ref. 21 for the simultaneous alleviation of the H_0 and σ_8 tensions. We shall briefly discuss these issues in section 4.

3. After RVM Inflation: KR-Axion Dark Matter and Leptogenesis

In this section we review briefly the situation charactrerising the exit from the RVM inflation and post inflationary eras of the model. For details we refer the reader to Refs. 25, 24, 27, 26.

At the exit from RVM inflation, gauge fields and chiral fermionic matter, including massive sterile right-handed neutrinos (RHN) of interest to us here, are assumed to be generated from the decay of the RVM vacuum.[25] Chiral fermions have their own gravitational and chiral anomalies. The former are assumed[24,26] to cancel the primordial gravitational anomalies due to the Green-Schwarz counterterms. The chiral anomalies, however, remain in the radiation and matter eras:[25]

$$\begin{aligned}&\partial_\mu\Big[\sqrt{-g}\,\Big(\sqrt{\frac{3}{8}}\frac{\alpha'}{\kappa}\,J^{5\mu} - \frac{\alpha'}{\kappa}\,\sqrt{\frac{2}{3}}\,\frac{1}{96}\,\mathcal{K}^\mu\Big)\Big]\\ &= \sqrt{\frac{3}{8}}\,\frac{\alpha'}{\kappa}\,\Big(\frac{\alpha_{\rm EM}}{2\pi}\,\sqrt{-g}\,F^{\mu\nu}\,\widetilde{F}_{\mu\nu} + \frac{\alpha_s}{8\pi}\,\sqrt{-g}\,G^a_{\mu\nu}\,\widetilde{G}^{a\mu\nu}\Big)\,,\end{aligned} \tag{30}$$

where $J^{5\mu} = \sum_{i=\text{chiral fermions}} \overline{\psi}_i\gamma^5\gamma^\mu\psi_i$, $F_{\mu\nu}$ denotes the electromagnetic (EM) Maxwell tensor, $G^a_{\mu\nu}$ is the gluon field strength, with $a = 1,\dots 8$ an adjoint SU(3) colour index, $\alpha_{\rm EM}$ the electromagnetic fine structure constant, and α_s the strong-interactions fine structure constant. The gluon terms in (30) may lead, through non-perturbative instanton effects in the post-inflationary QCD era, to the generation of a potential and a mass m_b term, for the KR axion, which, due to (12), is within phenomenologically acceptable ranges,[24,28]

$$1.17\times 10^{-8} \lesssim m_b/(\text{eV}) \lesssim 1.17\times 10^{-5}. \tag{31}$$

In this way, the KR axion field can play the rôle of a dominant component of dark matter.[25] In view of the association of the KR axion with torsion, then, one obtains a geometric origin of the dark matter sector of this Universe.

Moreover, the spontaneous-LV KR axion background (13), which remains undiluted at the exit phase from RVM inflation, can trigger the generation of Leptogenesis in models involving RHN, according to the mechanism discussed in Refs. 51, 52. The pertinent lagrangian of the RHN and their interactions with the standard model (SM) sector is:

$$\mathcal{L} = \mathcal{L}_{\rm SM} + i\overline{N}\,\gamma^\mu\,\partial_\mu\,N - \frac{m_N}{2}(\overline{N^c}N + \overline{N}N^c) - \overline{N}\gamma^\mu\,B_\mu\,\gamma^5 N - \sum_f\, y_f\overline{L}_f\tilde{\phi}^d N + \text{h.c.} \tag{32}$$

where h.c. denotes hermitian conjugate, $\mathcal{L}_{\rm SM}$ denotes the SM Lagrangian, N is the RHN field, of (Majorana) mass m_N, $\tilde{\phi}$ is the SU(2) adjoint of the Higgs field

ϕ ($\tilde{\phi}_i^d \equiv \varepsilon_{ij}\phi_j$, $i,j = 1,2$, SU(2) indices), L_f is a SM lepton doublet, with f a generation index, $f = e,\mu,\tau$, y_f is a Yukawa coupling of the "Higgs portal" interactions between the SM sector and RHN, which can lead to the generation of SM light neutrino masses, via, *e.g.*, the seesaw mechanism. The quantity B_μ is associated with the KR axion background (13) induced by the anomaly condensates during the RVM inflation:

$$B_\mu = M_{\rm Pl}^{-1}\,\dot{\bar{b}}\,\delta_{\mu 0}\,. \tag{33}$$

As discussed in Ref. 24, during the radiation era, B_μ is slowly varying with the cosmic temperature T ($B_0 \propto T^3$), so the Lagrangian (32) resembles that of a Standard Model Extension (SME) Lagrangian in a LV and CPT Violating (CPTV) (approximately constant) background.[53,54] A lepton asymmetry is then generated due to asymmetric tree-level decays of the RHN N into SM leptons and anti-leptons in the background (33) between the channels I and II:[51,52] Channel $I: N \to l^- h^+$, $\nu\, h^0$, and Channel II : $N \to l^+ h^-$, $\overline{\nu}\, h^0$, where $\ell^\pm$ are charged leptons, $h^\pm$ are the charged Higgs fields, which, at the high temperatures, above the electroweak symmetry breaking, at which this leptogenesis takes place,[51,52] do not decouple from the physical spectrum, ν ($\overline{\nu}$) are the light SM neutrinos (antineutrinos), and h^0 is the neutral Higgs field. The lepton asymmetry reads (assuming that the dominant lepton asymmetry is generated by one sterile neitrino, although extension to more RHN species, as required by the seesaw mechanism, is straightforward):[52]

$$\frac{\Delta L^{TOT}(T=T_D)}{s} \sim q\,\frac{B_0(T_D)\,m_N^2}{T_D^3} \sim q\,3.5\times 10^{11}\left(\frac{m_N}{M_{\rm Pl}}\right)^2, \quad q>0, \tag{34}$$

where $s \propto T^3$ is the entropy density of the Universe, $T_D \simeq m_N$ denotes the freeze-out temperature, and $0 < q = \mathcal{O}(10)$ is a numerical factor expressing theoretical uncertainties in the approximate analytic (Padé) methods used.[52] The lepton asymmetry (34) can then be communicated to the baryon sector via Baryon-minus-Lepton-number ($B-L$) conserving sphaleron processes in the SM sector.

4. Modern Era Phenomenology: Links of the Stringy RVM to a Modified Type-II RVM and Potential Data Tension Alleviation

In the modern era, the energy density of this stringy Universe also assumes an RVM form,

$$\rho_0 = \frac{3}{\kappa^2}\left(c_0 + \nu_0\, H_0^2\right)\,, \tag{35}$$

where now $\nu_0 > 0$, due to contributions from cosmic electromagnetic background fields[24] (we ignore terms of order H_0^4, as they are negligible in the current era). Comparison with the data[19,20] indicates $\nu_0 = \mathcal{O}(10^{-3})$. The present-day cosmological constant $c_0 > 0$ cannot be uniquely determined in our string-inspired cosmology, given that there might be various contributions to it,[28] some associated with details

of the extra dimensional geometry of the underlying microscopic string/brane theory. For comparison with data, we therefore should treat c_0 as a phenomenological parameter at this stage.

What is important to realise is that, as discussed in subsection 2.3, integrating out massless quantum graviton fluctuations in the path integral of the current-era gravitational theory, which is also characterised by a de Sitter background, generates terms in the one-loop effective action of the form (taking into account (29)):

$$\delta\mathcal{L}^{1-\text{loop}}_{\text{quant. grav.}} = \sqrt{-\widehat{g}}\left[\widetilde{\alpha}_0 + \widehat{R}\Big(c_1 + c_2\ln(\frac{1}{12}\kappa^2\widehat{R})\Big)\right] + \dots \tag{36}$$

From the supergravity example[30,42,43] we have that the constant coefficients $c_i \propto \kappa^2\mathcal{E}_0$, or $c_i \propto \kappa^2\mathcal{E}_0\ln(\kappa^4|\mathcal{E}_0|)$, $i = 1, 2$, (*cf.* (28)) with $\mathcal{E}_0$ a bare (constant) vacuum energy density scale (the ellipses $\dots$ denote terms of quadratic and higher order in $\widehat{R} = 12H^2$, which are subdominant in the current epoch ($H = H_0$)). The structures (36) appear generic for weak quantum gravity corrections about de Sitter backgrounds,[44] and we may therefore conjecture that they can play a rôle in the current era phenomenology.

From the graviton equations of the one-loop corrected effective Lagrangian, it can be readily seen that the terms (36) will imply corrections to the effective stress-energy tensor in the current era of the form,[28]

$$\delta\rho_0^{\text{vac}} = \frac{1}{2}\widetilde{\alpha}_0 + 3(c_1 - c_2)H_0^2 + 3c_2H_0^2\ln(\kappa^2H_0^2)\ , \tag{37}$$

which should be added to (35). Moreover, the supergravity prototype[30,42,43] indicates that the one-loop correction (dark-energy-type) term $\frac{1}{2}\widetilde{\alpha}_0$ is constant, independent of $\ln H^2$ terms. This will lead to some crucial differences from the standard type II RVM used in Ref. 21 to alleviate the H_0 and σ_8 tensions, which is characterised only by a mild cosmic-time t dependence of an effective gravitational constant, $\kappa^2_{\text{eff}}(t) = \kappa^2/\varphi(t)$, with $\varphi(t)$ a phenomenological non-dynamical function.

Another issue to be clarified is the sign of the bare cosmological constant $\Lambda_0 \propto \mathcal{E}_0\,\kappa^2$ which, as in the supergravity example of subsection 2.3, enters the computation of the effective action about de Sitter backgrounds in generic quantum gravity (QG) models that could describe the current-era. Such a sign depends on the details of the underlying QG theory. In the supergravity example this is negative (22) (Anti de Sitter type (AdS)), due to fundamental reasons of compatibility with supersymmetry.[42,43] One may also use such negative bare cosmological constant scales as formal regulators in the quantization procedure, as, *e.g.*, has been considered in the context of black hole physics[55] motivated by the AdS/conformal-field-theory correspondence.[56] We stress though that the one-loop corrected cosmological constant is always positive in such constructions, compatible with the current phenomenology (and also with the broken supergravity scenario[30,42,43]). We plan to study such issues, and their modern-era phenomenology, in future works.

Before closing the section, we mention that in string theory there might be a mixing of the KR axion with the other, string-model dependent, axions that exist

as a result of the compactification procedure.[35] In Ref. 24, we have argued that, depending on details of this mixing, one might obtain KR axion backgrounds that may be characterised by an approximately constant $\dot{b}$ even in the current era of the string Universe. Such backgrounds can therefore be parametrised by

$$\dot{b}_0 \sim \sqrt{2\,\epsilon'}\, H_0\, M_{\rm Pl}. \tag{38}$$

Phenomenological considerations, associated with the rôle of the KR axion background as dark matter, imply then that $\epsilon' \sim 10^{-2}$, in order to reproduce a dark matter energy density due to the KR axion in the phenomenologically right ballpark.[1,2] This, in turn, leads to a curious coincidence in the order of magnitude of the slow-roll parameters entering the expressions for $\dot{b}$ in the RVM-inflationary (13) (ϵ) and modern (38) eras (ϵ'),

$$\epsilon' \sim \epsilon \simeq 10^{-2}\ , \tag{39}$$

which needs to be understood further, in the context of the underlying microscopic string theory models. We do note though[25,28] that the background (38), under the assumption (39), is compatible with the stringent current experimental bounds of LV and CPTV discussed in ref. 54.

5. Conclusions and Outlook

In this talk, I reviewed a string-inspired model of a running vacuum (RVM) cosmology, which seems consistent with the current phenomenology, but also provides a geometrical origin of both, RVM inflation and dark matter. Crucial rôles are played by the gravitational anomalies that characterise the model at early eras, and the KR axion fields from the fundamental massless multiplet of the underlying string theory. The KR axion field is associated with a totally antisymmetric torsion in the underlying string theory. The RVM inflation arises from the non-linear H^4 term in the vacuum energy density (16), which is exclusively due to primordial GW that induce non trivial condensates of the gravitational Chern-Simons terms that are present in the very early stages of this string-inspired cosmology. Such terms do not arise in the context of anomaly-free quantum field theories, or Starobinsky inflation, and, as such, are rather exclusive to our string-inspired RVM. The anomaly condensates also imply LV KR axion backgrounds that induce baryogenesis through leptogenesis. In this sense, we may dare state that our very existence, that is the dominance of matter over antimatter in the Cosmos, is due to cosmic anomalies. *We are anomalously made of star stuff*, to paraphrase the famous quote by Carl Sagan.[23,24]

There are several open issues that we need to understand, and explore further, such as: (i) potential hints in the cosmic-microwave-background data[1] about the LV KR axion backgrounds, (ii) prospects of getting some phenomenological indications in early-Universe data about the negative, anomaly-induced, coefficient of the H^2 term in the RVM-like cosmic energy density (16) during the RVM inflation.

This, however, is a difficult task, in view of the phase transition occurring in our model at the exit from the RVM inflation, (iii) the rich phenomenology of both, the string-model-independent KR axion and the other, string-model dependent, axions, that exist in string theory,[35,57] (iv) the precise rôle of quantum-gravity-induced $H^2\ln(\kappa^2 H^2)$ corrections in the vacuum energy density of the current era, comparing the model with the type-II RVM,[21] in an attempt to discuss potential alleviation of the observed tensions in the cosmological data, and (v) last but not least: in view of the association of the KR axion with (totally antisymmetric) torsion, the phenomenological comparison of the stringy RVM with other cosmologies with torsion[58,59] (in which the torsion has more components than the totally antisymmetric one characterising our model) or other teleparallel-gravity models[60] (where torsion mimics gravity, in contrast to our case, which contains also graviton fields). We hope to be able to report on these important issues in the near future.

Acknowledgements

NEM thanks A. Gómez-Valent and J. Solà peracaula for the invitation to speak in the *CM3* parallel session: "*Status of the H_0 and σ_8 tensions: theoretical models and model-independent constraints*" of the MG16 Marcel Grossmann virtual Conference, July 5-10 2021. He also acknowledges participation in the COST Association Action CA18108 "*Quantum Gravity Phenomenology in the Multimessenger Approach (QG-MM)*". This work is funded in part by the UK Science and Technology Facilities research Council (STFC) under the research grant ST/T000759/1.

References

1. N. Aghanim *et al.* [Planck], Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020) doi:10.1051/0004-6361/201833910 [arXiv:1807.06209 [astro-ph.CO]].
2. P. A. Zyla *et al.* [Particle Data Group], Review of Particle Physics, PTEP **2020**, no. 8, 083C01 (2020) doi:10.1093/ptep/ptaa104.
3. For a recent review see: Mohayaee R., Rameez M. and Sarkar S., Do supernovae indicate an accelerating universe?, Eur. Phys. J. Spec. Top. (2021), doi:10.1140/epjs/s11734-021-00199-6. [arXiv:2106.03119 [astro-ph.CO]].
4. E. P. Verlinde, On the Origin of Gravity and the Laws of Newton, JHEP **04**, 029 (2011) doi:10.1007/JHEP04(2011)029 [arXiv:1001.0785 [hep-th]].
5. L. T. Verde, T. Treu T. and A. G. Riess, Tensions between the Early and the Late Universe, Nature Astron. **3**, 891 (2019) doi:10.1038/s41550-019-0902-0. [arXiv:1907.10625 [astro-ph.CO]].
6. L. Perivolaropoulos L. and F. Skara F. (2021), Challenges for ΛCDM: An update, [arXiv:2105.05208 [astro-ph.CO]].
7. A. G. Riess *et al.*, A 2.4% Determination of the Local Value of the Hubble Constant, Astrophys. J. **826** (2016) no. 1, 56 (2016). doi:10.3847/0004-637X/826/1/56. [arXiv:1604.01424 [astro-ph.CO]].
8. A. G. Riess, *et al.*, New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope: Implications for the Hubble Constant, Astrophys. J. **855** no. 2, 136 (2018). doi:10.3847/1538-4357/aaadb7. [arXiv:1801.01120 [astro-ph.SR]].

9. A. G. Riess, S. Casertano S., W. Yuan W. J. B. Bowers , L. Macri L., J. C. Zinn and D. Scolnic, Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM, Astrophys. J. Lett. **908** no. 1, L6 (2021). doi:10.3847/2041-8213/abdbaf. [arXiv:2012.08534 [astro-ph.CO]].
10. E. Di Valentino *et al.*, Cosmology intertwined III: $f\sigma_8$ and S_8, Astropart. Phys. **131**, 102604 (2021). doi:10.1016/j.astropartphys.2021.102604, [arXiv:2008.11285 [astro-ph.CO]] and references therein.
11. G. Efstathiou (2021), To H0 or not to H0?, doi:10.1093/mnras/stab1588. [arXiv:2103.08723 [astro-ph.CO]].
12. E. Mortsell, A. Goobar, J. Johansson and S. Dhawan (2021), The Hubble Tension Bites the Dust: Sensitivity of the Hubble Constant Determination to Cepheid Color Calibration, [arXiv:2105.11461 [astro-ph.CO]].
13. W. L Freedman (2021), Measurements of the Hubble Constant: Tensions in Perspective, [arXiv:2106.15656 [astro-ph.CO]].
14. W. L. Freedman, Cosmology at a Crossroads, Nature Astron. **1**, 0121 (2017). doi:10.1038/s41550-017-0121. [arXiv:1706.02739 [astro-ph.CO]].
15. I. L. Shapiro and J. Sola, On the scaling behavior of the cosmological constant and the possible existence of new forces and new light degrees of freedom, Phys. Lett. B **475**, 236-246 (2000) doi:10.1016/S0370-2693(00)00090-3 [arXiv:hep-ph/9910462 [hep-ph]].
16. I. L. Shapiro and J. Solà, Scaling behavior of the cosmological constant: Interface between quantum field theory and cosmology, JHEP **02**, 006. doi:10.1088/1126-6708/2002/02/006 (2002). [arXiv:hep-th/0012227 [hep-th]].
17. J. Solà, Dark energy: A Quantum fossil from the inflationary Universe?, J. Phys. A **41**, 164066 (2008) doi:10.1088/1751-8113/41/16/164066 [arXiv:0710.4151 [hep-th]], and references therein.
18. S. Basilakos S., J. A. S. Lima and J. Solà, From inflation to dark energy through a dynamical Lambda: An attempt at alleviating fundamental cosmic puzzles, Int. J. Mod. Phys. D **22**, 1342008 (2013). doi:10.1142/S021827181342008X. [arXiv:1307.6251 [astro-ph.CO]].
19. J. Solà and A. Gómez-Valent, The $\bar{\Lambda}$CDM cosmology: From inflation to dark energy through running Λ, Int. J. Mod. Phys. D **24**, 1541003 (2015). doi:10.1142/S0218271815410035 [arXiv:1501.03832 [gr-qc] and references therein.
20. J. Solà, Cosmological constant vis-a-vis dynamical vacuum: bold challenging the ΛCDM, Int. J. Mod. Phys. A **31**, no. 23, 1630035 (2016). doi:10.1142/S0217751X16300350. [arXiv:1612.02449 [astro-ph.CO]].
21. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Perez and C. Moreno-Pulido, Running vacuum against the H_0 and σ_8 tensions, EPL **134** no. 1, 19001 (2021). doi:10.1209/0295-5075/134/19001. [arXiv:2102.12758 [astro-ph.CO]].
22. J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*, and *String theory. Vol. 2: Superstring theory and beyond*, (Cambridge University Press (Cambridge, UK, 2010, 2011), doi:10.1017/CBO9780511816079, doi:10.1017/CBO9780511618123).
23. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula, Do we Come from a Quantum Anomaly?, Int. J. Mod. Phys. D **28**, no. 14, 1944002 (2019) doi:10.1142/S0218271819440024 [arXiv:1905.04685 [hep-th]].
24. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula Gravitational and Chiral Anomalies in the Running Vacuum Universe and Matter-Antimatter Asymmetry, Phys. Rev. D **101** no. 4, 045001 (2020). doi:10.1103/PhysRevD.101.045001. [arXiv:1907.04890 [hep-ph]].

25. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula , Quantum Anomalies in String-Inspired Running Vacuum Universe: Inflation and Axion Dark Matter, Phys. Lett. B **803**, 135342 (2020). doi:10.1016/j.physletb.2020.135342 [arXiv:2001.03465 [gr-qc]].
26. N. E. Mavromatos and J. Solà Peracaula, Stringy-Running-Vacuum-Model Inflation: from primordial Gravitational Waves and stiff Axion Matter to Dynamical Dark Energy, Eur. Phys. J. Spec. Top. (2021), doi:10.1140/epjs/s11734-021-00197-8. [arXiv:2012.07971 [hep-ph]].
27. N. E. Mavromatos and J. Solà Peracaula , Inflationary physics and transplanckian conjecture in the Stringy Running-Vacuum-Model: from the phantom vacuum to the true vacuum [arXiv:2105.02659 [hep-th]] (2021).
28. N. E. Mavromatos (2021), Geometrical origins of the Universe dark sector: string-inspired torsion and anomalies as seeds for inflation and dark matter, [arXiv:2108.02152 [gr-qc]].
29. J. R. Ellis, N. E. Mavromatos and D. V. Nanopoulos, Time dependent vacuum energy induced by D particle recoil, Gen. Rel. Grav. **32**, 943-958 (2000). doi:10.1023/A:1001993226227. [arXiv:gr-qc/9810086 [gr-qc]].
30. S. Basilakos, N. E. Mavromatos and J. Solà, Starobinsky-like inflation and running vacuum in the context of Supergravity, Universe **2**, no. 3, 14 (2016) doi:10.3390/universe2030014 [arXiv:1505.04434 [gr-qc]].
31. A. A. Starobinsky, A New Type of Isotropic Cosmological Models Without Singularity, Phys. Lett. B **91**, 99-102 (1980).
32. C. Moreno-Pulido C. and J. Solà Peracaula J., Running vacuum in quantum field theory in curved spacetime: renormalizing ρ_{vac} without $\sim m^4$ terms, Eur. Phys. J. C **80**, no. 8, 692 (2020). doi:10.1140/epjc/s10052-020-8238-6. [arXiv:2005.03164 [gr-qc]].
33. S. H. S. Alexander, M. E. Peskin and M. M. Sheikh-Jabbari, Leptogenesis from gravity waves in models of inflation, Phys. Rev. Lett. **96**, 081301 (2006). doi:10.1103/PhysRevLett.96.081301 [hep-th/0403069].
34. M. J. Duncan, N. Kaloper N.and K. A. Olive, Axion hair and dynamical torsion from anomalies, Nucl. Phys. B **387**, 215 (1992,).
35. P. Svrcek and E. Witten E., Axions In String Theory, JHEP **06**, 051 doi:10.1088/1126-6708/2006/06/051 (2006). [arXiv:hep-th/0605206 [hep-th]].
36. R. Jackiw R. and S. Y.Pi, Chern-Simons modification of general relativity, Phys. Rev. D **68**, 104012 (2003), doi:10.1103/PhysRevD.68.104012. [gr-qc/0308071].
37. J. Martin, C. Ringeval and V. Vennin, Encyclopædia Inflationaris, Phys. Dark Univ. **5-6**, 75-235 (2014). doi:10.1016/j.dark.2014.01.003 [arXiv:1303.3787 [astro-ph.CO]], and references therein.
38. J. Grande, J. Solà and H. Stefancic, LXCDM: A Cosmon model solution to the cosmological coincidence problem?, JCAP **08**, 011 (2006). doi:10.1088/1475-7516/2006/08/011. [arXiv:gr-qc/0604057 [gr-qc]].
39. J. Grande, A. Pelinson and J. Solà, Dark energy perturbations and cosmic coincidence, Phys. Rev. D **79**, 043006 (2009). doi:10.1103/PhysRevD.79.043006. [arXiv:0809.3462 [astro-ph]]
40. M. S. Morris and K. S. Thorne and U. Yurtsever, Wormholes, Time Machines, and the Weak Energy Condition, Phys. Rev. Lett. **61**, 1446-1449 (1988). doi:10.1103/PhysRevLett.61.1446.
41. P. Van Nieuwenhuizen, Supergravity, Phys. Rept. **68**, 189-398 (1981) doi:10.1016/0370-1573(81)90157-5.
42. J. Alexandre, N. Houston and N. E. Mavromatos, Dynamical Supergravity Breaking via the Super-Higgs Effect Revisited, Phys. Rev. D **88**, 125017 (2013) doi:10.1103/PhysRevD.88.125017 [arXiv:1310.4122 [hep-th]].

43. J. Alexandre, N. Houston and N. E. Mavromatos, Starobinsky-type Inflation in Dynamical Supergravity Breaking Scenarios, Phys. Rev. D **89**, no. 2, 027703 (2014) doi:10.1103/PhysRevD.89.027703 [arXiv:1312.5197 [gr-qc]].
44. E. S. Fradkin and A. A. Tseytlin, One Loop Effective Potential in Gauged O(4) Supergravity, Nucl. Phys. B **234**, 472 (1984) doi:10.1016/0550-3213(84)90074-9.
45. Z. Lalak, S. Lola, B. A. Ovrut and G. G. Ross, Large scale structure from biased nonequilibrium phase transitions: Percolation theory picture, Nucl. Phys. B **434**, 675-696 (1995) doi:10.1016/0550-3213(94)00557-U [arXiv:hep-ph/9404218 [hep-ph]].
46. D. Coulson, Z. Lalak and B. A. Ovrut, Biased domain walls, Phys. Rev. D **53**, 4237-4246 (1996) doi:10.1103/PhysRevD.53.4237.
47. J. Ellis and N. E. Mavromatos, Inflation induced by gravitino condensation in supergravity, Phys. Rev. D **88**, no. 8, 085029 (2013) doi:10.1103/PhysRevD.88.085029 [arXiv:1308.1906 [hep-th]].
48. L. Lanczos, *Bemerkung zur de Sitterschen Welt*, Physikalische Zeitschrift, **23**, 539-543 (1922).
49. Y. B. Zel'dovich 1961 (1962), The equation of state at ultrahigh densities and its relativistic limitations, Zh. Eksp. Teor. Fiz. **41**, 1609-1615 (1961) (Sov.Phys.JETP **14** (1962) 1143-1147). The stiff matter in that model is assumed to be baryons, in contrast to our model, in which the stiff matter consists of potential-free KR axions.
50. P. H. Chavanis, Cosmology with a stiff matter era, Phys. Rev. D **92**, no. 10, 103004 (2015). doi:10.1103/PhysRevD.92.103004. [arXiv:1412.0743 [gr-qc]].
51. M. de Cesare, N. E. Mavromatos and S. Sarkar, On the possibility of tree-level leptogenesis from Kalb–Ramond torsion background, Eur. Phys. J. C **75** no. 10, 514 (2015), doi:10.1140/epjc/s10052-015-3731-z. [arXiv:1412.7077 [hep-ph]].
52. T. Bossingham T., N. E. Mavromatos and S. Sarkar, The role of temperature dependent string-inspired CPT violating backgrounds in leptogenesis and the chiral magnetic effect, Eur. Phys. J. C **79** no. 1, 50 (2019). doi:10.1140/epjc/s10052-019-6564-3. [arXiv:1810.13384 [hep-ph]] and references therein.
53. D. Colladay and V. A. Kostelecky, Lorentz violating extension of the standard model, Phys. Rev. D **58**, 116002 (1998). [hep-ph/9809521].
54. V. A. Kostelecky and N. Russell, , Data Tables for Lorentz and CPT Violation, Rev. Mod. Phys. **83**, 11 (2011). [arXiv:0801.0287 [hep-ph]], and references therein.
55. N. E. Mavromatos and E. Winstanley, Infinitely colored black holes, Class. Quant. Grav. **17**, 1595-1611 (2000) doi:10.1088/0264-9381/17/7/302 [arXiv:hep-th/9909018 [hep-th]].
56. J. M. Maldacena, The Large N limit of superconformal field theories and supergravity, Adv. Theor. Math. Phys. **2**, 231-252 (1998) doi:10.1023/A:1026654312961 [arXiv:hep-th/9711200 [hep-th]].
57. A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper and J. March-Russell, String Axiverse, Phys. Rev. D **81**, 123530 (2010) doi:10.1103/PhysRevD.81.123530 [arXiv:0905.4720 [hep-th]].
58. S. Capozziello, R. Cianci, C. Stornaiolo and S. Vignolo, f(R) cosmology with torsion, Phys. Scripta **78**, 065010 (2008), doi:10.1088/0031-8949/78/06/065010, [arXiv:0810.2549 [gr-qc]] and references therein.
59. J. Magueijo, T. G. Zlosnik and T. W. B. Kibble, Cosmology with a spin, Phys. Rev. D **87**, no. 6, 063504 (2013), doi:10.1103/PhysRevD.87.063504. [arXiv:1212.0585 [astro-ph.CO]].
60. Y. F. Cai, S. Capozziello, M. De Laurentis and E. N. Saridakis, f(T) teleparallel gravity and cosmology, Rept. Prog. Phys. **79** (2016) no. 10, 106901 (2016), doi:10.1088/0034-4885/79/10/106901 [arXiv:1511.07586 [gr-qc]] and references therein.

Early and not so early dark energy. What do cosmological observations tell us about them?

Adrià Gómez-Valent[1,*], Ziyang Zheng[1], Luca Amendola[1], Valeria Pettorino[2] and Christof Wetterich[1]

[1] *Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*

[2] *AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité F-91191 Gif-sur-Yvette, France*

Cosmological data still allow for the presence of a non-negligible amount of dark energy at very high redshifts, namely during the matter- and radiation-dominated epochs. This is the so-called early dark energy (EDE), which could help to mitigate the tensions that affect the standard model of cosmology since (i) it reduces the sound horizon at the baryon-drag epoch, hence giving room to higher values of H_0 than those found in the ΛCDM; and (ii) it could potentially decrease the number of large-scale structures in the Universe due its negative pressure and its inability to cluster efficiently for large enough values of its sound speed. Here we put constraints on the fraction of EDE using two methods: first, we use a perfect fluid parameterization that produces plateaux in $\Omega_{ede}(z)$ during the relativistic and non-relativistic matter-dominated eras. Second, we apply a tomographic approach to constrain the EDE density in redshift bins, which allows us to reconstruct the evolution of the EDE fraction before and after the decoupling of the Cosmic Microwave Background (CMB) photons. We have employed *Planck* data 2018, the Pantheon compilation of supernovae of Type Ia (SNIa), data on galaxy clustering, the prior on the absolute magnitude of SNIa by SH0ES, and weak lensing data from KiDS+VIKING-450 and DES-Y1. Using our minimal parameterization we find that EDE is not able to loosen the cosmological tensions, and show that the constraints on the EDE fraction weaken considerably when its sound speed takes lower values. Thanks to our binned analysis we are able to put tight constraints on the EDE fraction around the CMB decoupling time, $\lesssim 0.4\%$ at 2σ c.l. We confirm previous results that a significant EDE fraction in the radiation-dominated epoch loosens the H_0 tension, but tends to worsen the tension for σ_8. A subsequent presence of EDE in the matter-dominated era helps to alleviate this problem. When both the SH0ES prior and weak lensing data are considered in the fitting analysis in combination with data from CMB, SNIa and baryon acoustic oscillations, the EDE fractions are constrained to be $\lesssim 2.6\%$ in the radiation-dominated epoch and $\lesssim 1.5\%$ in the redshift range $z \in (100, 1000)$ at 2σ c.l. The two tensions remain with a statistical significance of $\sim 2-3\sigma$ c.l. This contribution to the proceedings of the CM3 parallel session of the MG16 Marcel Grossmann virtual Conference: "Status of the H_0 and σ_8 tensions: theoretical models and model-independent constraints" is based on the paper arXiv:2107.11065,[1] which appeared in the arXiv shortly after my talk of July 6th 2021.

Keywords: Cosmology: observations – Cosmology: theory – cosmological parameters – dark energy – dark matter.

*Speaker. E-mail: gomez-valent@thphys.uni-heidelberg.de

1. Introduction

The cosmological constant, Λ, is (together with other assumptions as the Cosmological Principle, the existence of cold dark matter (CDM) and an inflationary period prior to the radiation-dominated era) a key building block of the standard model of Cosmology. By adding this very simple term in Einstein's field equations it is possible to produce the late-time acceleration of the Universe[2,3] and help to explain a large variety of cosmological observations[4–6] without increasing excessively the mathematical complexity of the equations. However, the model is not free from extremely intricate theoretical conundrums, as the famous old cosmological constant[7–9] and coincidence problems, see e.g.[10] The former might be strongly intertwined with the latter, which should not be considered as a "why now" problem,[11] but as a matter of why the energy density associated to Λ, i.e. $\rho_\Lambda = \Lambda/8\pi G$, takes the value $\rho_\Lambda \sim \mathcal{O}(10^{-47})$ GeV^4 in natural units, hence being of the same order of magnitude of the non-relativistic matter energy density $\rho_m(z)$ when $z \lesssim 1$. This is pivotal for the model to exhibit an excellent phenomenological performance. We can soften the coincidence problem if we depart from the ΛCDM by allowing some sort of dynamical dark energy (DE) density instead of considering a rigid ρ_Λ.[12–14] This is a more appealing framework, since it is actually hard to believe in an immutable entity like the cosmological constant. Why should the component in charge of the current acceleration of the Cosmos be insensitive e.g. to the Universe's evolution and energy content? Such dynamics could lead to the presence of a non-negligible fraction of early dark energy (EDE) in the Universe during the radiation-dominated epoch (RDE) and/or the matter-dominated era (MDE). Of particular relevance concerning the coincidence problem are the quintessence models with scaling solutions, as the one with an exponential potential originally proposed by C. Wetterich in the late 80's,[14] in which the EDE fraction follows the dominant component in the Universe in both, the RDE and MDE,[15] making more natural the scenario with $\rho_{de} \sim \mathcal{O}(\rho_m)$ at low redshifts.

Apart from the aforementioned theoretical problems there are also some observational tensions affecting the ΛCDM, making the concordance model less concordant than it was thought to be ten years ago. At least, if these tensions are not induced by systematic errors in the data. The cosmic microwave background (CMB) temperature, polarization and lensing data from *Planck* 2018 leads to a value of the Hubble parameter $H_0 = (67.36 \pm 0.54)$ km/s/Mpc when analyzed under the assumption of the ΛCDM,[4] and a similar result is obtained from measurements of the baryon acoustic oscillations (BAO) and the deuterium abundance, again in the standard model.[16] These values are at odds with the one measured by the SH0ES team, $H_0 = (73.2 \pm 1.3)$ km/s/Mpc,[17] which is obtained with the cosmic distance ladder method and does not rely on the assumption of any cosmological model. There exists a $\sim 4.1\sigma$ tension between them, which has been persistently and consistently increasing in the last years,[18,19] see also the reviews.[20,21] On the other hand, galaxy clustering (through redshift-space distortions, RSD), direct peculiar velocity and

weak lensing (WL) measurements suggest that the Universe is less clumpy than preferred by the CMB data under the ΛCDM, see e.g.[22–25] The tension is usually formulated in terms of the root-mean-square (*rms*) of mass fluctuations at scales of $8h^{-1}$ Mpc, σ_8, and related composite quantities like $S_8 = \sigma_8(\Omega_m^{(0)}/0.3)^{0.5}$, with $\Omega_m^{(0)}$ the current matter energy fraction in the Universe. The first indication of its existence appeared already almost a decade ago[26] and it is still there.[27–29] The tension is in this case less significant from a statistical point of view than the aforesaid one of H_0, of about $2-3\sigma$ depending on the data source and the large-scale structure (LSS) estimator employed to quantify the tension. For instance, by combining the KiDS-1000 WL results with BOSS and 2dFLenS data the authors of[25] have shown that it is possible to reduce by a factor ~ 2 the uncertainty on S_8, and also to write the tension only in terms of σ_8, since BOSS puts tight constraints on $\Omega_m^{(0)}$ and the degeneracy in the $\sigma_8 - \Omega_m^{(0)}$ plane can be broken. They find $S_8 = 0.766^{+0.020}_{-0.014}$ and $\sigma_8 = 0.760^{+0.021}_{-0.023}$, which are in 3.1σ and 2.2σ tension with *Planck*, respectively.

Whether the H_0 and S_8/σ_8 tensions are to some degree physical or not is still unclear and under intense debate.[30–32] Hopefully this question will be resolved in the near future. In the meanwhile, theoreticians have worked very hard to find ways of loosening the tensions, taking for granted that they are real, see the reviews[21,33,34] and the complete lists of references therein. It is important to remark, though, that it is very difficult to find models in the literature capable of relieving both tensions at a time in a significant way. There are only some few exceptions that offer better perspectives, as e.g. the running vacuum model of type-II studied in[29] or the Brans-Dicke ΛCDM model explored in.[35,36] The latter, though, might encounter some problems when trying to match the cosmological and local values of the gravitational coupling through an appropriate screening mechanism,[37] and the effect of the *Planck* 2018 CMB polarization data might also hinder its overall fitting performance. Other models are only able to loosen one of the tensions, while worsening the other. This is the case e.g. of the class of new EDE models based on early phase transitions,[38–40] ultra-light axions[41,42] or alike.[43] In practice, all these models fight against the H_0 tension in a very similar way. There is a new component in the energy budget of the Universe that acts as a cosmological constant deep in the RDE and has an associated (constant) density which is of about $5-10\%$ the radiation energy density around the matter-radiation equality time, when the EDE fraction reaches its maximum. This excess of energy with respect to the standard model decreases the sound horizon at the baryon-drag epoch, and this forces the Hubble parameter to be larger in order to decrease the angular diameter distance to the last-scattering surface and keep in this way intact the location of the first peak of the temperature CMB anisotropies. The latter is very well constrained by *Planck*. After that moment, the energy density dilutes typically faster than radiation, leaving no imprint in the late-time universe, where a cosmological constant is still assumed to produce the current positive acceleration, as in the standard ΛCDM. Although it is possible to obtain posterior values of the Hubble parameter

much closer to the SH0ES measurement[17] in the context of these new EDE models, they require larger values of the current matter energy density in order to compensate the early Sachs-Wolfe effect introduced by EDE in the pre-recombination epoch, and this enhances the LSS formation processes in the late-time Universe, which in turn exacerbates the S_8/σ_8 tension.[44–46]

Here we first explore EDE models with scaling solutions in the MDE and RDE, or just in the MDE. As already discussed above, they can alleviate in some sense the coincidence problem. They could in principle have an impact on the cosmological tensions as well, depending on how strong are the constraints on the EDE fraction imposed by the the cosmological data. In these models, EDE might not be negligible after the recombination era, and this could lead to a softening of the S_8/σ_8 tension. Moreover, the shape of $\Omega_{ede}(z) = \rho_{ede}(z)/\rho_c(z)$, with $\rho_c(z) = 3H^2(z)/8\pi G$ the critical energy density in the Universe, is very different from the one encountered in the new EDE models mentioned in the previous paragraph, since in this case the EDE fraction is constant during the epochs at which the scaling behavior comes into play. Hence, it is clearly worth to study these EDE models and to determine to what extent they can alleviate the tensions, if they do at all. In view of the existing tensions, there is a clear interest of revisiting these models, whose seeds where already present more than thirty years ago.[14] In addition, we also perform a more model-independent analysis, applying a tomographic method to reconstruct the shape of $\Omega_{ede}(z)$ that is preferred by different combinations of cosmological data sets. We will see that this is very useful to extract more general information about how is EDE constrained in the various epochs of the cosmic expansion.

2. Early dark energy

EDE affects observables in several ways. The presence of EDE at decoupling can change the position and height of the peaks in the CMB[47–49] and can also impact it through the early integrated Sachs-Wolfe effect. Furthermore, EDE suppresses the growth of structure:[50–52] a smaller number of clusters can form with respect to the ΛCDM[53] because of the negative pressure of EDE and also because of its large sound speed, which does not allow it to cluster; the lensing potential is also weaker, with an impact on weak lensing and the CMB peaks at large multipoles. Thus, EDE can potentially have a direct impact on the cosmological tensions. In the following we first design a general parameterization able to mimic the background dominant component, and then proceed with a tomographic analysis in different redshift bins.

2.1. *Parametric EDE*

We build now a simple parameterization of the DE density that allows us to reproduce the behavior of uncoupled quintessence models with scaling solutions. For this reason, we want our parameterization to be able to generate two plateaux in $\Omega_{de}(z)$. The first plateau occurs in the RDE, and the second one in the MDE,

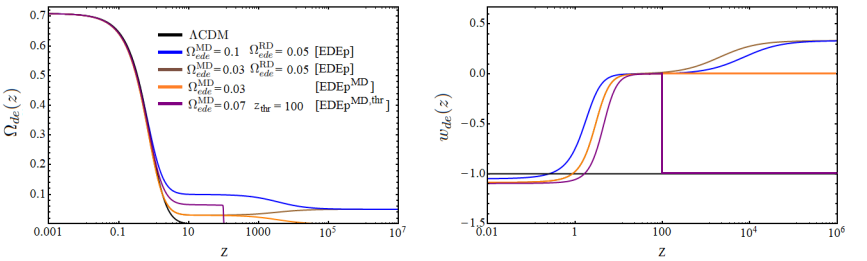


Fig. 1. Functions $\Omega_{de}(z)$ (left plot) and $w_{de}(z) = p_{de}(z)/\rho_{de}(z)$ (right plot) obtained using the parametrizations described in Sec. 2.1, for some illustrative values of the parameters.

and can have different heights in principle. This is what happens for instance in quintessence models with a single exponential potential $V(\phi) = V_0 e^{-\sqrt{8\pi G}\lambda\phi}$, where the fractions of EDE in the RDE and MDE depend only on one parameter, λ.[14,15] By introducing a second exponential potential it is possible to control the height of the two plateaux independently.[54]

With this aim in mind, we generalize the parameterizations proposed earlier in[48,49] and we consider here a DE density with the following form,

$$\rho_{de}(z) = \rho_1(1+z)^4 + \rho_2(1+z)^3 + \rho_3(1+z)^{3(1+w)}\,, \tag{1}$$

parameterized by the constant energy densities ρ_1, ρ_2 and ρ_3 and by a constant equation of state parameter (EoS) w. We call this parameterization EDEp, where the "p" reminds us of the 'plateaux' that characterise it. The last term of (1) is able to mimic the behavior of a late-time dynamical DE with the wCDM form,[55] whereas the first two terms produce the plateaux in the RDE and MDE. It is useful to write the constants ρ_1 and ρ_2 in terms of dimensionless parameters, as follows

$$\rho_1 = \chi_1 \Omega_{r,*}^{(0)} \rho_c^{(0)}\,; \qquad \rho_2 = \chi_2 \Omega_{m,*}^{(0)} \rho_c^{(0)}\,, \tag{2}$$

where $\Omega_{r,*}^{(0)}$ and $\Omega_{m,*}^{(0)}$ are the current density parameters of radiation and matter, respectively, computed considering three massless neutrinos, and $\rho_c^{(0)} = \rho_c(z=0)$. We consider standard General Relativity and a flat Friedmann-Lemaître-Robertson-Walker universe. $\chi_1, \chi_2, \rho_3 > 0$ in order the DE density to be positive during the expansion. Deep in the RDE and MDE the EDE fractions are constant,

$$\Omega_{ede}^{\rm RD} = \frac{\chi_1}{1+\chi_1}\,; \qquad \Omega_{ede}^{\rm MD} = \frac{\chi_2}{1+\chi_2}\,. \tag{3}$$

The present dark energy density, $\rho_{de}^{(0)}$, can be directly computed from the Hubble parameter, $H_0 = 100h$ km/s/Mpc, and the reduced CDM and baryon density parameters, $\omega_{cdm} = \Omega_{cdm}^{(0)} h^2$ and $\omega_b = \Omega_b^{(0)} h^2$. Thus, one of the three ρ_i's appearing in (1) can be expressed in terms of the other two, e.g. $\rho_3 = \rho_{de}^{(0)} - \rho_1 - \rho_2$, so in this EDE parametrization we deal with three additional parameters with respect to the ΛCDM. We have nine cosmological parameters in total, namely the spectral

index n_s and amplitude A_s of the primordial power spectrum, H_0, ω_b, ω_{cdm}, the optical depth to reionization, τ, together with w, χ_1 and χ_2 (or, equivalently, $\Omega_{ede}^{\rm RD}$ and $\Omega_{ede}^{\rm MD}$). We consider a massive neutrino of 0.06 eV and two massless neutrinos.

The DE fluid is covariantly conserved so its associated pressure reads,

$$\dot{\rho}_{de}+3H(\rho_{de}+p_{de})=0 \longrightarrow p_{de}(z)=\frac{\rho_1}{3}(1+z)^4+w\rho_3(1+z)^{3(1+w)}\,, \qquad (4)$$

where the dot refers to a derivative with respect to the cosmic time. The corresponding equation of state parameter can be obtained from the ratio of (4) and (1), i.e. $w_{de}(z)=p_{de}(z)/\rho_{de}(z)$. It is clear that $w_{de}=1/3$ and $w_{de}=0$ in the RDE and MDE, respectively, and $w_{de}\approx w$ at present for low values of χ_2. For the perturbations, in our main analyses we take the sound speed of the DE fluid to be equal to the speed of light in the DE rest frame,[56] i.e. $\hat{c}_s=1$, so in our model the DE does not cluster efficiently. We will study also what happens when $\hat{c}_s<1$.

If we set $\rho_1=0$ ($\Omega_{ede}^{\rm RD}=0$) in (1) EDE is completely negligible during the RDE,

$$\rho_{de}(z)=\rho_2(1+z)^3+\rho_3(1+z)^{3(1+w)}\,. \qquad (5)$$

We denote this particular case of EDEp as EDEp$^{\rm MD}$, to remind us that the plateau is in this case following the matter component. If we turn the first term on in (5) at a particular 'threshold' redshift $z_{\rm thr}$, in the MDE we obtain,

$$\rho_{de}(z)=\rho_2(1+z)^3\theta(z_{\rm thr}-z)+\rho_3(1+z)^{3(1+w)}\,, \qquad (6)$$

with θ the Heaviside step function. This is what we call EDEp$^{{\rm MD},thr}$. For $z_{\rm thr}\to\infty$ we recover the pure EDEp$^{\rm MD}$ parametrization (5). We show typical shapes of the functions $\Omega_{de}(z)$ and $w_{de}(z)$ obtained with EDEp, EDEp$^{\rm MD}$ and EDEp$^{{\rm MD},thr}$ in Fig. 1.

2.2. *Tomographic EDE*

We further consider the possibility of binning the amount of $\Omega_{de}(z)$ in 11 bins to perform a tomographic analysis using the data sets described in Sec. 3. Here, again, $\rho_{de}^{(0)}$ can be directly determined from H_0, ω_b and ω_{cdm}; the other constant densities ρ_i, with $i=A,...,J$ (see Table 1 in[1] for details), are left free in our Monte Carlo (MC) runs, together with w and the six usual ΛCDM parameters. We keep $\hat{c}_s=1$. Our main aim is to see how much EDE we can have in each bin, and therefore which shape of $\Omega_{de}(z)$ is preferred by the data, regardless of its complexity. The corresponding fitting results and reconstructed shapes of $\Omega_{de}(z)$ are shown and discussed in Sec. 4.2.

3. Methodology and data

We have implemented the various parametrizations of Sec. 2.1 and also the binned $\rho_{de}(z)$ described in Sec. 2.2 in our own modified version of the Einstein-Boltzmann code `CLASS`.[57] We have constrained the parameters of our models through a Bayesian

exploration of the parameter space, employing the MC sampler `MontePython`.[58] We have used flat priors for the cosmological parameters in common with the ΛCDM model, with widths that fully respect the *Planck* 2018 uncertainties.[4] Regarding the priors on the EDE fractions, previous studies in the literature showed that they are always lower than $\sim 10\%$,[48,49] so we have used $0 < \chi_i < 0.12$ in the MC analyses of the parametrizations of Sec. 2.1, and similar priors for each bin of the tomographic study of Sec. 2.2.

Now we list the various data sets used in this study:

- CMB: We consider the full *Planck* 2018 TTTEEE+lowE and TTTEEE+lowE+lensing likelihoods.[4] We denote these data sets as CMBpol and CMBpolens, respectively.
- Supernovae of Type Ia (SNIa): We use the observed SNIa apparent magnitude and redshifts of the Pantheon compilation.[5] The absolute magnitude of these SNIa, M, is left free in the fitting analysis, and we impose a prior on it. See the next item of the list for details.
- Prior on M: We use in some of our fitting analyses the SH0ES effective calibration prior on the absolute magnitude of the SNIa as provided in,[59] $M_{\rm SH0ES} = -19.2191 \pm 0.0405$. It is obtained from the calibration of nearby SNIa (at $z \lesssim 0.01$) with Cepheids.[60] It is better to use this prior rather than the one on H_0,[17,61] especially when it is combined with data from SNIa compilations that include the same SNIa in the Hubble flow considered by the SH0ES team (as in the Pantheon compilation[5]) because in this way we avoid double-counting issues. In some of our Tables we also provide the best-fit values of M. This allows us to quantity the 'M tension', i.e. the tension between the latter and $M_{\rm SH0ES}$. We show in Sec. 4.2 that the statistical level of the SH0ES-*Planck* tension can be in some cases quite different when formulated in terms of H_0 and M.
- BAO: We have employed the data reported in.[6,22,62–65]
- WL: In some of our fitting analyses we employ the KiDS +VIKING-450+DESY1 prior $S_8 = 0.762^{+0.025}_{-0.024}$.[66] The author of[67] has raised some concerns about the use of σ_8 and derived quantities as S_8. He suggests the use of σ_{12}, defined as the *rms* linear theory variance at the fixed scale of 12 Mpc, and $S_{12} = \sigma_{12}(\omega_m/0.14)^{0.4}$. We provide the values of these parameters in some of our Tables, together with the usual σ_8 and S_8.
- RSD: Data on anisotropic clustering of galaxies in redshift space, and from the direct measurement of peculiar velocities.[6,22,68–75] We call this data set RSD in short because most of these points are obtained from the analysis of redshift-space distortions.

The corresponding fitting results are presented and discussed in Sec. 4. In order to quantify the impact of the SH0ES prior on the fitting results and on the ability of the models to loosen the H_0/M tension we study and compare the constraints on our

Table 1. The mean fit values and 68.3% confidence limits for the ΛCDM, the wCDM, EDEp (1) and EDEp$^{\rm MD}$ (5), using the CMBpol+SNIa and CMBpolens+SNIa+M data sets (cf. Sec. 3). For $\Omega_{ede}^{\rm RD}$ and $\Omega_{ede}^{\rm MD}$ we also show the 2σ limits inside the parentheses. Constraints on τ are provided in.[1] See the comments in Sec. 4.1.

CMBpol+SNIa				
Parameter	ΛCDM	wCDM	EDEp	EDEp$^{\rm MD}$
ω_b	$0.02239^{+0.00014}_{-0.00015}$	$0.02237^{+0.00016}_{-0.00015}$	$0.02238^{+0.00017}_{-0.00016}$	$0.02234^{+0.00015}_{-0.00016}$
ω_{cdm}	$0.1199^{+0.0014}_{-0.0013}$	0.1204 ± 0.0014	$0.1218^{+0.0016}_{-0.0015}$	$0.1208^{+0.0014}_{-0.0015}$
n_s	0.9659 ± 0.0044	$0.9646^{+0.0044}_{-0.0045}$	$0.9642^{+0.0044}_{-0.0045}$	$0.9642^{+0.0043}_{-0.0046}$
H_0 [km/s/Mpc]	$67.60^{+0.59}_{-0.61}$	$68.55^{+1.12}_{-1.10}$	68.71 ± 1.16	$68.61^{+1.09}_{-1.16}$
σ_8	$0.811^{+0.007}_{-0.008}$	0.823 ± 0.014	$0.817^{+0.014}_{-0.015}$	$0.818^{+0.015}_{-0.014}$
r_d [Mpc]	$147.02^{+0.28}_{-0.32}$	146.92 ± 0.30	$146.18^{+0.66}_{-0.43}$	$146.75^{+0.36}_{-0.30}$
w	-1	$-1.039^{+0.035}_{-0.039}$	$-1.050^{+0.041}_{-0.040}$	$-1.053^{+0.038}_{-0.042}$
$\Omega_{ede}^{\rm RD}$ (%)	0	0	$< 0.91\,(< 2.08)$	0
$\Omega_{ede}^{\rm MD}$ (%)	0	0	$< 0.27\,(< 0.69)$	$< 0.29\,(< 0.69)$
CMBpolens+SNIa+M				
Parameter	ΛCDM	wCDM	EDEp	EDEp$^{\rm MD}$
ω_b	$0.02257^{+0.00015}_{-0.00014}$	$0.02241^{+0.00014}_{-0.00016}$	$0.02245^{+0.00015}_{-0.00019}$	0.02239 ± 0.00015
ω_{cdm}	0.1179 ± 0.0012	0.1200 ± 0.0015	0.1212 ± 0.0016	0.1206 ± 0.0012
n_s	0.9709 ± 0.0044	$0.9658^{+0.0044}_{-0.0047}$	$0.9659^{+0.0045}_{-0.0044}$	$0.9656^{+0.0041}_{-0.0042}$
H_0 [km/s/Mpc]	$68.56^{+0.56}_{-0.54}$	$70.55^{+0.86}_{-0.88}$	$70.63^{+0.86}_{-0.82}$	$70.20^{+0.52}_{-0.69}$
σ_8	$0.811^{+0.007}_{-0.008}$	0.838 ± 0.014	$0.830^{+0.014}_{-0.013}$	0.835 ± 0.010
r_d [Mpc]	$147.36^{+0.28}_{-0.29}$	$146.98^{+0.33}_{-0.30}$	$146.21^{+0.73}_{-0.44}$	146.79 ± 0.29
w	-1	$-1.098^{+0.035}_{-0.032}$	$-1.099^{+0.034}_{-0.032}$	$-1.099^{+0.030}_{-0.025}$
$\Omega_{ede}^{\rm RD}$ (%)	0	0	$< 1.14\,(2.44)$	0
$\Omega_{ede}^{\rm MD}$ (%)	0	0	$< 0.22\,(0.52)$	$< 0.22\,(0.54)$

EDE models obtained: (i) with a minimal data set composed by CMBpol+SNIa; and (ii) adding on top of the latter the SH0ES prior on M, i.e. using CMBpol+SNIa+M. The SNIa data help to break the strong degeneracies found in the w-H_0 plane when only CMB data are used in the analysis.[76]

The properties and limitations of the EDEp and EDEp$^{\rm MD}$ parametrizations are already grasped with the aforementioned minimal data sets (cf. Table 1). For EDEp$^{{\rm MD},thr}$ we study also the effect of the CMB lensing and BAO+RSD data when combined with CMBpol+SNIa. We provide the corresponding constraints in figures 2 and 3. For the analyses of the binned $\rho_{de}(z)$ described in Sec. 2.2 we report our results in Table 2, where we explicitly test the impact of BAO and the weak lensing data, by considering not only the minimal data sets described in the previous paragraph, but also adding the information on BAO and BAO+WL. In addition, we redo the fitting analyses considering also the CMB lensing in order to quantify

its impact. For further explanations on the data we refer the reader to our work[1] and the original data sources.

4. Results and discussion

We present and discuss now the results obtained from the fitting analyses of the EDE parametrizations and the tomographic EDE described in Secs. 2.1 and 2.2, respectively, using the methodology and data sets of the previous section.

4.1. *Results for the parametric analysis*

The mean fit values and corresponding uncertainties for the various cosmological parameters in the EDEp and EDEp$^{\rm MD}$ parametrizations obtained with the baseline CMBpol+SNIa dataset are reported in Table 1. The constraints on the fraction of early dark energy in the EDEp parametrization in the radiation- and matter-dominated epochs, $\Omega_{ede}^{\rm RD}$ and $\Omega_{ede}^{\rm MD}$, are very strong. They lie below $\sim 2\%$ and $\sim 0.7\%$, respectively, at the 2σ c.l. It is interesting to observe that the upper limit of $\Omega_{ede}^{\rm MD}$ in the EDEp$^{\rm MD}$ parametrization coincides with the one obtained in the more general EDEp. The constraints on $\Omega_{ede}^{\rm MD}$ and $\Omega_{ede}^{\rm RD}$ in EDEp are quite independent. Actually, we have checked that the correlation coefficient between these two parameters is pretty small, $\sim 5.6\%$. As already noticed in,[48] the low upper limits on $\Omega_{ede}^{\rm RD}$ and $\Omega_{ede}^{\rm MD}$ are due to the very tight constraint on the fraction of EDE around the CMB decoupling time. The latter acts as an anchor for $\Omega_{ede}^{\rm RD}$ and even more for $\Omega_{ede}^{\rm MD}$, since in the last scattering surface the matter energy density is already ~ 3 times larger than the radiation one.

Another result from Table 1 to remark is that EDEp cannot alleviate significantly the H_0 and σ_8 tensions. The shape of the early dark energy density seems to be too restricted in these parametrization. There is a slight increase of H_0 in EDEp and EDEp$^{\rm MD}$ with respect to the ΛCDM, but it is mainly due to the dynamics of the late-time DE, and this is why the major part of the effect is already found with the wCDM parametrization.[55]

When we include the SH0ES prior in our fitting analysis we increase, of course, the value of the Hubble parameter, see Table 1. The tension with the distance ladder determination is now only of $\sim 1.7\sigma$, but this is again mainly thanks to the lowering of w, which now lies more in the phantom region (3σ away below $w = -1$). The values of w and H_0 are almost identical to those found in the wCDM parametrization, and EDE does not have any important impact on the H_0 tension in the context of the EDEp parametrization. The loosening of the H_0 tension is accompanied by a slight worsening of the σ_8 one due to the positive correlation between H_0 and σ_8. Phantom dark energy leads to lower values of the DE density in the past and this produces, in turn, an increase of the structure formation processes in the Universe. It seems that the EDE density has a too restricted form in the parameterizations under study here. They allow for plateaux and generalize previous

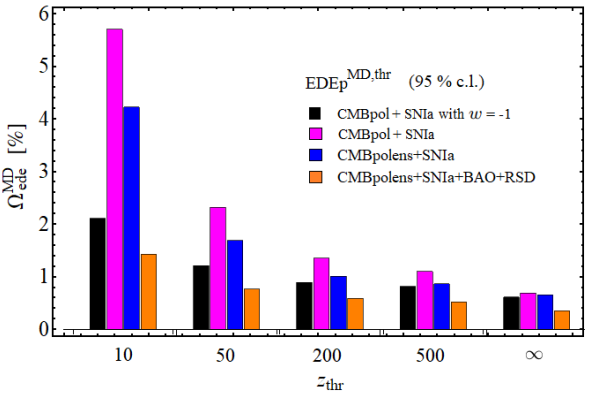


Fig. 2. Constraints on $\Omega_{ede}^{\rm MD}$ at 95% c.l. obtained from the fitting analyses of the EDEp$^{\rm MD,thr}$ parametrization (6), and under the data sets shown in the legend.

studies, but still seem to be quite constrained and unable to resolve the tensions. In the next section we then investigate the binned tomographic approach described in Sec. 2.2. Before moving on, however, we discuss briefly two options to weaken the constraints on the fraction of EDE.

Much weaker constraints on the fraction of EDE can be obtained by allowing for values of $\hat{c}_s^2 \ll 1$. The only difference between our EDE and cold dark matter during the MDE is found at the perturbations level. Both are pressureless fluids at the background level, but have different $\hat{c}_s$. The latter is equal to 1 for EDE and 0 for the dark matter. Hence, we expect the constraints on $\Omega_{ede}^{\rm MD}$ to loosen if we decrease the value of $\hat{c}_s$ of EDE, and even obtain a full degeneracy between the fraction of EDE and CDM during the MDE in the limit $\hat{c}_s \to 0$. This is actually what happens, as we explicitly show in Fig. 2 of.[1] We have to say, though, that the change in the sound speed does not help to alleviate the cosmological tensions, neither. For instance, under the CMBPol+SNIa data set the Hubble parameter remains close to ~ 67.5 km/s/Mpc, and the 1σ uncertainty is lower than ~ 0.8 km/s/Mpc regardless of the value of $\hat{c}_s$ under consideration when the late-time DE dynamics is switched off.

One can also get weaker constraints on EDE in the MDE by activating EDE at lower redshifts. In order to study this effect we explore the EDEp$^{{\rm MD},thr}$ parametrization (6). In Fig. 2 we provide the 2σ c.l. bounds on $\Omega_{ede}^{\rm MD}$ obtained with the baseline data set CMBpol+SNIa with and without late-time dark energy dynamics, and also adding the CMB lensing and the BAO+RSD data sets. When $z_{\rm thr} \to \infty$ we recover the constraints obtained in the EDEp$^{\rm MD}$ model, of course, but when we allow for lower values of the threshold redshift (below the CMB decoupling one) we get larger upper bounds on $\Omega_{ede}^{\rm MD}$, which depend on the concrete data set under consideration and also on $z_{\rm thr}$. We report the results obtained with $z_{\rm thr} = 10, 50, 200, 500, \infty$. The addition of the *Planck* 2018 CMB lensing to the CMBpol+SNIa baseline data set leads to stronger constraints on $\Omega_{ede}^{\rm MD}$. Its value decreases by a $\sim 25\%$ $\forall z_{\rm thr}$. When

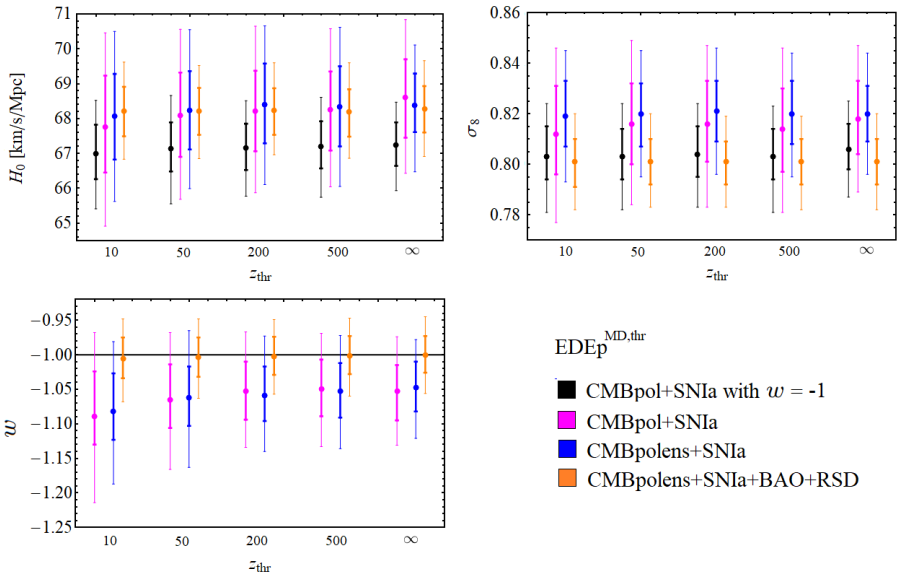


Fig. 3. Constraints for H_0, σ_8 and w at 1σ and 2σ c.l. obtained with the EDEp$^{\mathrm{MD},thr}$ parametrization (6) for different values of the threshold redshift z_{thr} and using the same combined data sets as in Fig. 2.

we also include the BAO+RSD data the decrease is even bigger, $\sim 75\%$. If we turn off the late-time dynamics of DE we also obtain tighter bounds on $\Omega_{ede}^{\mathrm{MD}}$, just because in this case we remove the degeneracy between this parameter and w. Our Fig. 2 can be compared with Figs. 6-7 of[48] and Fig. 11 of,[49] which were obviously obtained with older data sets. In[48] the authors employed CMB data from WMAP9 combined with small scale measurements from the South Pole Telescope (SPT), whereas in[49] the authors employed the *Planck* 2015 CMB likelihood and studied the impact of some other background and weak lensing data sets, as described in Sec. 4 in that reference. The results presented in this section constitute a significant update obtained with the *Planck* 2018 likelihood and also other more recent background and LSS data (cf. Sec. 3 for details).

In Fig. 3 we provide the corresponding constraints on H_0, σ_8 and w at 1σ and 2σ c.l. in the EDEp$^{\mathrm{MD,thr}}$ parameterization for the same scenarios explored in Fig. 2. They support some of the comments made in the previous paragraphs of this section, e.g. (i) the values of σ_8 and H_0 remain close to those found in the ΛCDM model. In other words, the tensions are not significantly alleviated in this class of scaling early dark energy models; (ii) phantom values of $w < -1$ allow us to decrease the H_0 tension very slightly; and (iii) larger values of w lead to lower values of σ_8 due to the presence of a larger fraction of dark energy at low redshifts. This is why we get $w \sim -1$ when we include the BAO+RSD data set.

The dedicated analysis presented in this section updated and generalized previous constraints on this class of early dark energy models, and motivated the study of

Table 2. The mean fit values and 68.3% confidence limits for the most important cosmological (main+derived) parameters, obtained under different data sets (with and without CMB lensing) for the binned $\rho_{de}(z)$ described in Sec. 2.2. See Fig. 4 for the constraints on $\Omega_{de}(z)$, and Sec. 4.2 for a thorough discussion on these results. Here we have employed the following notation: Base≡CMBpol+SNIa and BaseL≡CMBpolens+SNIa.

Parameter	Base	Base+M	Base+M+BAO	Base+M+BAO+WL
ω_b	$0.02257^{+0.00022}_{-0.00023}$	$0.02277^{+0.00023}_{-0.00025}$	$0.02282^{+0.00024}_{-0.00025}$	0.02259 ± 0.00021
ω_{cdm}	$0.1222^{+0.0020}_{-0.0021}$	0.1221 ± 0.0022	$0.1223^{+0.0020}_{-0.0021}$	$0.1200^{+0.0013}_{-0.0014}$
n_s	$0.9727^{+0.0072}_{-0.0073}$	$0.9752^{+0.0067}_{-0.0069}$	$0.9760^{+0.0070}_{-0.0071}$	$0.9740^{+0.0061}_{-0.0067}$
H_0 [km/s/Mpc]	$68.29^{+1.26}_{-1.36}$	$70.86^{+1.00}_{-1.10}$	$70.38^{+0.84}_{-0.89}$	$69.85^{+0.76}_{-0.77}$
M	-19.405 ± 0.032	$-19.342^{+0.023}_{-0.024}$	$-19.350^{+0.019}_{-0.020}$	$-19.365^{+0.018}_{-0.017}$
σ_8	$0.854^{+0.023}_{-0.025}$	$0.880^{+0.025}_{-0.029}$	$0.877^{+0.026}_{-0.029}$	$0.833^{+0.016}_{-0.017}$
S_8	$0.869^{+0.025}_{-0.029}$	0.863 ± 0.029	$0.866^{+0.026}_{-0.028}$	$0.819^{+0.014}_{-0.016}$
σ_{12}	$0.840^{+0.019}_{-0.022}$	0.843 ± 0.022	$0.843^{+0.020}_{-0.022}$	$0.806^{+0.012}_{-0.013}$
S_{12}	$0.851^{+0.022}_{-0.025}$	$0.854^{+0.024}_{-0.027}$	$0.855^{+0.023}_{-0.026}$	$0.810^{+0.014}_{-0.016}$
r_d [Mpc]	$145.66^{+0.90}_{-0.70}$	$145.22^{+0.95}_{-0.90}$	$145.06^{+0.97}_{-0.89}$	$146.51^{+0.64}_{-0.51}$
w	$-1.037^{+0.043}_{-0.041}$	-1.070 ± 0.038	$-1.048^{+0.037}_{-0.034}$	$-1.037^{+0.032}_{-0.031}$
Parameter	BaseL	BaseL+M	BaseL+M+BAO	BaseL+M+BAO+WL
ω_b	$0.02257^{+0.00021}_{-0.00023}$	$0.02274^{+0.00023}_{-0.00025}$	$0.02277^{+0.00022}_{-0.00024}$	$0.02266^{+0.00021}_{-0.00020}$
ω_{cdm}	0.1215 ± 0.0016	0.1211 ± 0.0017	0.1214 ± 0.0016	$0.1193^{+0.0013}_{-0.0014}$
n_s	$0.9721^{+0.0063}_{-0.0071}$	$0.9748^{+0.0067}_{-0.0071}$	$0.9746^{+0.0068}_{-0.0069}$	$0.9727^{+0.0061}_{-0.0068}$
H_0 [km/s/Mpc]	$68.25^{+1.27}_{-1.26}$	$70.84^{+1.04}_{-1.07}$	$70.21^{+0.80}_{-0.84}$	$70.00^{+0.76}_{-0.73}$
M	$-19.407^{+0.031}_{-0.030}$	$-19.343^{+0.024}_{-0.025}$	-19.355 ± 0.018	-19.362 ± 0.016
σ_8	$0.845^{+0.017}_{-0.020}$	$0.868^{+0.019}_{-0.021}$	$0.864^{+0.019}_{-0.021}$	$0.839^{+0.013}_{-0.015}$
S_8	$0.858^{+0.019}_{-0.021}$	$0.848^{+0.020}_{-0.022}$	$0.853^{+0.019}_{-0.020}$	$0.824^{+0.012}_{-0.013}$
σ_{12}	$0.831^{+0.014}_{-0.015}$	$0.831^{+0.015}_{-0.016}$	$0.833^{+0.015}_{-0.016}$	$0.810^{+0.010}_{-0.011}$
S_{12}	$0.841^{+0.016}_{-0.017}$	$0.840^{+0.017}_{-0.020}$	0.843 ± 0.018	$0.815^{+0.011}_{-0.013}$
r_d [Mpc]	$145.94^{+0.69}_{-0.57}$	$145.65^{+0.82}_{-0.71}$	$145.51^{+0.83}_{-0.71}$	$146.46^{+0.56}_{-0.44}$
w	$-1.035^{+0.040}_{-0.039}$	$-1.067^{+0.037}_{-0.036}$	$-1.050^{+0.034}_{-0.035}$	-1.045 ± 0.032

the next section, in which we will reconstruct the shape of $\Omega_{de}(z)$ without sticking to a restricted family of parametrizations.

4.2. *Results for tomographic dark energy*

Now we provide the results obtained in the tomographic model described in Sec. 2.2 in order to see whether more general shapes of $\Omega_{de}(z)$ can loosen the cosmological tensions. This is in fact suggested by previous analyses in the literature, see e.g.[38–43, 77]

Our results are presented in Table 2 and Fig. 4. They confirm that a significant (and non-constant) fraction of EDE in the RDE can alleviate the H_0 tension if it can be kept below $\sim 0.6\%$ at 2σ c.l. around the CMB decoupling time, i.e. at $z \sim 1000 - 2000$. For instance, from the analysis of the CMBpolens+SNIa and CMBpolens+SNIa+M+BAO data sets we obtain $H_0 = (68.25^{+1.27}_{-1.26})$ km/s/Mpc

and $H_0 = (70.21^{+0.80}_{-0.84})$ km/s/Mpc, respectively. They are 2.77σ and 1.95σ below the SH0ES measurement, and the central value is significantly lower when the SH0ES prior is not considered. The authors of[44] reported similar results in the context of the ultra-light axion model. Nevertheless, if we compare the values of M obtained from these data sets with $M_{\rm SH0ES}$ we still obtain a tension of 3.80σ and 3.07σ, respectively. This means that in terms of M, the tension is bigger than when it is formulated in terms of H_0, and the capability of EDE of alleviating the tension is much lower, at least in the EDE framework we are considering here.

We also see that the large fraction of $\Omega_{de}(z)$ required in the RDE to loosen the H_0 tension, which can be of about $\sim 4-5\%$ at 2σ c.l. according to some data sets that include the SH0ES prior on M but no LSS information, leads to higher values of ω_{cdm} and, to a lesser extent, also of n_s, which in turn exacerbates the σ_8/S_8 tension. The former is needed to reduce the early integrated Sachs-Wolfe effect introduced by the EDE. This is aligned with previous works that also consider EDE in the pre-recombination epoch, see e.g.[44–46,78,79] Table 2 shows that the tension decreases $\sim 1\sigma$ when it is analyzed through the LSS estimators σ_{12} and S_{12}.[67] Nevertheless, it does not disappear. Under the CMBpolens+SNIa+M data set the σ_8 tension is of 3.63σ, whereas for the σ_{12} parameter it is of 2.64σ, and a similar decrease is observed when other data combinations are employed in the fitting analysis.

When the SH0ES prior on M is taken into account in absence of LSS data, we get a $\sim 2\sigma$ evidence for the presence of a non-null EDE fraction during the RDE, similar to.[41] Nevertheless, the inclusion of the weak lensing data from KiDS+VIKING-450 and DES-Y1 (cf. Sec. 3) forces the EDE fraction in the RDE to be again compatible at 1σ with 0 in order to allow ω_{cdm} to take values closer to the ΛCDM ones and not to worsen the σ_8/S_8 tension. Notice that, as expected, the upper bound on the fraction of EDE in the MDE that we obtain in our binned analysis is larger than the one found with EDEp and its variants, even when the prior on S_8 is not included (see Table 1 and Fig. 4). Indeed, higher fractions of EDE in the MDE (see again Fig. 4) also allow to keep the amount of LSS more under control. With the CMBpolens+SNIa+M+BAO+WL we obtain $H_0 = (70.00^{+0.76}_{-0.73})$ km/s/Mpc and $M = -19.362 \pm 0.016$. They are in 2.13σ and 3.28σ tension with the SH0ES values, respectively. Again, the tension in M is larger than in H_0 by $\sim 1\sigma$. Regarding the LSS estimators, we obtain $S_8 = 0.824^{+0.012}_{-0.013}$ and $\sigma_{12} = 0.810^{+0.010}_{-0.011}$. The former is in 2.25σ tension with the KiDS+VIKING-450+DESY1 value, whereas the latter is compatible at 1σ with the value obtained in the ΛCDM. The tensions in H_0 and S_8 can be kept at $\sim 2\sigma$ c.l. under this concrete data set, as advocated in.[78]

The effect of the CMB lensing from *Planck* 2018 on the EDE fraction can be appreciated by direct comparison of the plots in the left and right columns of Fig. 4. When the CMB lensing is included, the upper bound on $\Omega_{de}(z)$ is reduced by $\sim 1\%$ for $z \gtrsim 5000$, by a $\sim 0.5\%$ for $3000 \lesssim z \lesssim 5000$ and by a smaller fraction at lower redshifts. Some differences are found, though, depending on the other data sets employed in the fitting analyses. The preferred matter densities decrease, although

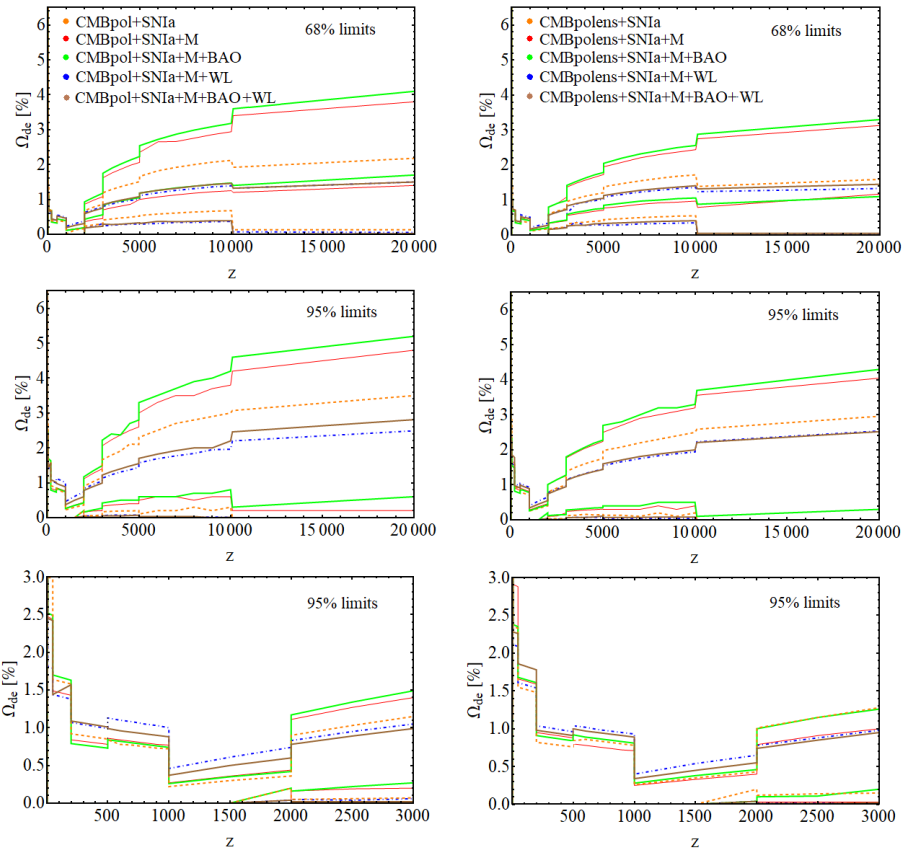


Fig. 4. *Left plots:* Reconstructed shapes of $\Omega_{de}(z)$ obtained from the fitting analyses without the Planck 2018 CMB lensing data (cf. Sec. 3). In the first and second rows we show the constraints in the region $z \in [0, 2 \cdot 10^4]$ at 1σ and 2σ c.l., respectively. In the last row we zoom in the redshift range $z \in [0, 3000]$ to better appreciate the details in the MDE. The tightest upper bound on $\Omega_{de}(z)$ is obtained around the CMB decoupling time, i.e. at $z \sim 1000 - 2000$, where the data force $\Omega_{de}(z) \lesssim 0.6\%$ at 2σ c.l.; *Right plots:* The same, but including the CMB lensing data. See the comments in Sec. 4.2.

they are still compatible at 1σ with the ones inferred without including the CMB lensing. This leads also to a slight decrease on the LSS estimators when the weak lensing prior is not considered. When the latter is included, the values of σ_{12} and S_{12} (and also σ_8 and S_8) remain stable under the addition of the CMB lensing likelihood.

We would like to remark the very low upper bounds that we obtain for the EDE fraction in the redshift range $z \in (1000, 2000)$, cf. Fig. 4. The exact value for these upper bounds depend on the specific data set, e.g. the constraints are a little bit weaker when the WL prior is used in the fitting analysis, but the EDE fraction when $\hat{c}_s = 1$ is, in any case, very strongly constrained in that epoch and lies always below $\sim 0.6\%$ regardless of the data set under consideration. Under the full data set, the

EDE fractions are constrained to be below 2.6% in the RDE and $\lesssim 1-1.5\%$ in the redshift range $z \in (100, 1000)$ at 2σ c.l. This limits strongly the possible impact of EDE on the value of the Hubble parameter or the present amount of structure.

5. Conclusions

In this work we have studied the phenomenological performance of a family of flexible parametrizations for the dark energy density that are able to mimic the scaling behavior that is encountered in a wide variety of quintessence models. DE has been treated as a perfect fluid with two plateaus in $\Omega_{de}(z)$, one in the matter-dominated epoch and another one in the radiation-dominated era. We have put very tight constraints on the fraction of EDE in the radiation- and matter-dominated epochs in the context of these models. The CMBpol+SNIa data set already forces these two quantities to lie below 2.44% and 0.52% at 2σ c.l., respectively. These strong constraints are necessary to respect the upper bound on the amount of EDE at the last scattering surface, as we have explicitly checked in our tomographic analysis. We have found that this class of scaling EDE models does not lead to a significant alleviation of the H_0 and σ_8 tensions. Larger EDE fractions are allowed by the data if: (i) the sound speed for DE is fixed at values $\hat{c}_s^2 \ll 1$, since in this limit dark energy behaves as dark matter during the matter-dominated epoch at the background and perturbations levels and, hence, there is a huge degeneracy between these two components. If we would assume a vanishing sound speed for dark energy there would be no way to distinguish it from dark matter. We find indeed no bounds on the dark energy fraction in this case. This changes drastically already for a rather small sound speed of EDE of the order 10^{-4}; and (ii) if EDE is switched on at later times, already in the matter-dominated era. Nevertheless, the cosmological tensions remain also in these cases.

EDE can only have a larger impact on the cosmological tensions if $\Omega_{de}(z)$ takes more flexible shapes that allow to respect the very strict constraints found around the CMB decoupling time ($\Omega_{de}(z_{dec}) \lesssim 0.4\%$ at 2σ c.l.), while still leading to a significant EDE fraction in other epochs of the cosmic expansion. The strong bound on EDE for redshift $200 < z < 1000$ ($\Omega_{de}(z) \lesssim 1\%$ at 2σ c.l.) is rather impressive since it is entirely based on the different clustering properties of dark energy and dark matter.

In general, when the SH0ES prior is not included in the fitting analysis there is no significant shift in the value of the Hubble parameter when compared to the ΛCDM, although the uncertainties clearly grow by a factor $2-3$. This allows to decrease the H_0 tension to the 2.66σ c.l. under the minimal CMBpol+SNIa data set. The addition of the CMB lensing has a very mild effect, increasing the tension up to the 2.73σ level.

When the BAO data and the SH0ES prior are also considered, the H_0 tension is reduced to the $\sim 2\sigma$ level by increasing the EDE fraction at the radiation-dominated epoch. A 2σ preference for a non-null EDE density in that epoch is obtained,

$\Omega_{de}(z) \lesssim 4\%$ at 95% c.l. Our reconstructed $\Omega_{de}(z)$ allows for non-peaked shapes, in contrast to what one finds e.g. in models based on ultra-light axions.[41,80] The model needs, though, values of the current dark matter density much larger than the ones typically encountered in the concordance model. This is to lower the early integrated Sachs-Wolfe effect down, which is enhanced by the presence of EDE in the pre-recombination epoch. This automatically leads to an increase of σ_8 and S_8 that worsens the LSS tension, which lies now at the $3 - 3.5\sigma$ c.l. In terms of σ_{12}, the tension is somewhat lower, but still remains at the $\sim 2.7\sigma$ level, as when the M prior is not used in the analysis. It is also to be noted that the H_0 tension still stays at the 3σ level when formulated in terms of the absolute magnitude of SNIa M.

Finally, when we use the most complete data set, taking also the weak lensing data into account, we find that the model can keep the H_0 and S_8 tensions at the $\sim 2\sigma$ c.l., thanks also to a slight increase of the EDE fraction in the matter-dominated epoch, although, again, the M tension remains at $\gtrsim 3\sigma$.

In view of our results, it seems unlikely that EDE alone can provide a satisfactory resolution of the cosmological tensions. Whether the latter have or not a physical origin, or whether their statistical significance is as high as claimed by some sectors of the cosmological community, is still a matter of discussion and is certainly not a closed subject. Here we conclude that, in any case, if the data sets employed in this study do not suffer from any important systematic errors, uncoupled EDE is not able to relieve completely the tensions. Under our full data set CMBpolens+SNIa+M+BAO+WL they remain at $2 - 3\sigma$ c.l.

Nothing prevents, though, the solution to the cosmological tensions to be multi-sided, rather than due to a single new physics component. EDE could still play a significant role in this story. Some interesting directions to explore in the future are: (i) a possible coupling of EDE to dark matter, (ii) the impact of more complicated behavior of the EDE sound speed, parameterizing its time dependence, or performing a tomographic analysis similar to the one we have carried out in this work for the EDE density; and (iii) the potential degeneracy between EDE and other cosmological parameters, as the neutrino masses, which could in principle help to soften the S_8 tension while keeping the needed amount of non-relativistic matter at the CMB decoupling time. The latter would be needed, together with a higher EDE fraction in the radiation-dominated epoch, in order to increase the value of the Hubble parameter. We leave these investigations for a future work.

Acknowledgments

The author is very grateful to Prof. Joan Solà Peracaula for his kind invitation to give this talk and co-chair the CM3 session with him. He was funded by the Deutsche Forschungsgemeinschaft (DFG) - Project number 415335479 - during the time at which the MG16 meeting was held and the research of this project carried out.

References

1. A. Gómez-Valent, Z. Zheng, L. Amendola, V. Pettorino and C. Wetterich, Early dark energy in the pre- and post-recombination epochs (7 2021).
2. S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
3. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
4. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
5. D. M. Scolnic *et al.*, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *Astrophys. J.* **859**, p. 101 (2018).
6. R. Neveux *et al.*, The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the quasar sample between redshift 0.8 and 2.2, *Mon. Not. Roy. Astron. Soc.* **499**, 210 (2020).
7. S. Weinberg, The Cosmological Constant Problem, *Rev. Mod. Phys.* **61**, 1 (1989), [,569(1988)].
8. J. Martin, Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask), *Comptes Rendus Physique* **13**, 566 (2012).
9. J. Solà Peracaula, Cosmological constant and vacuum energy: Old and new ideas, *J. Phys. Conf. Ser.* **453**, p. 012015 (2013).
10. P. J. E. Peebles and B. Ratra, The Cosmological Constant and Dark Energy, *Rev. Mod. Phys.* **75**, 559 (2003).
11. H. E. S. Velten, R. F. vom Marttens and W. Zimdahl, Aspects of the cosmological "coincidence problem", *Eur. Phys. J. C* **74**, p. 3160 (2014).
12. P. J. E. Peebles and B. Ratra, Cosmology with a Time Variable Cosmological Constant, *Astrophys. J. Lett.* **325**, p. L17 (1988).
13. B. Ratra and P. J. E. Peebles, Cosmological Consequences of a Rolling Homogeneous Scalar Field, *Phys. Rev. D* **37**, p. 3406 (1988).
14. C. Wetterich, Cosmology and the Fate of Dilatation Symmetry, *Nucl. Phys. B* **302**, 668 (1988).
15. E. J. Copeland, A. R. Liddle and D. Wands, Exponential potentials and cosmological scaling solutions, *Phys. Rev. D* **57**, 4686 (1998).
16. G. E. Addison, D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw and J. L. Weiland, Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy, *Astrophys. J.* **853**, p. 119 (2018).
17. A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn and D. Scolnic, Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM, *Astrophys. J. Lett.* **908**, p. L6 (2021).
18. A. G. Riess *et al.*, A 2.4% Determination of the Local Value of the Hubble Constant, *Astrophys. J.* **826**, p. 56 (2016).
19. A. G. Riess *et al.*, New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope: Implications for the Hubble Constant, *Astrophys. J.* **855**, p. 136 (2018).
20. L. Verde, T. Treu and A. G. Riess, Tensions between the Early and the Late Universe, *Nature Astron.* **3**, p. 891 (7 2019).
21. E. Di Valentino *et al.*, Snowmass2021 - Letter of interest cosmology intertwined II: The hubble constant tension, *Astropart. Phys.* **131**, p. 102605 (2021).

22. H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, *Mon. Not. Roy. Astron. Soc.* **465**, 1757 (2017).
23. S. Joudaki *et al.*, KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering, *Mon. Not. Roy. Astron. Soc.* **474**, 4894 (2018).
24. A. H. Wright, H. Hildebrandt, J. L. van den Busch, C. Heymans, B. Joachimi, A. Kannawadi and K. Kuijken, KiDS+VIKING-450: Improved cosmological parameter constraints from redshift calibration with self-organising maps, *Astron. Astrophys.* **640**, p. L14 (2020).
25. C. Heymans *et al.*, KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, *Astron. Astrophys.* **646**, p. A140 (2021).
26. E. Macaulay, I. K. Wehus and H. K. Eriksen, Lower Growth Rate from Recent Redshift Space Distortion Measurements than Expected from Planck, *Phys. Rev. Lett.* **111**, p. 161301 (2013).
27. S. Nesseris, G. Pantazis and L. Perivolaropoulos, Tension and constraints on modified gravity parametrizations of $G_{\rm eff}(z)$ from growth rate and Planck data, *Phys. Rev. D* **96**, p. 023542 (2017).
28. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Possible signals of vacuum dynamics in the Universe, *Mon. Not. Roy. Astron. Soc.* **478**, 4357 (2018).
29. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Perez and C. Moreno-Pulido, Running vacuum against the H_0 and σ_8 tensions, *EPL* **134**, p. 19001 (2021).
30. G. Efstathiou, A Lockdown Perspective on the Hubble Tension (with comments from the SH0ES team) (7 2020).
31. E. Mortsell, A. Goobar, J. Johansson and S. Dhawan, The Hubble Tension Bites the Dust: Sensitivity of the Hubble Constant Determination to Cepheid Color Calibration (5 2021).
32. W. L. Freedman, Measurements of the Hubble Constant: Tensions in Perspective (6 2021).
33. E. Di Valentino *et al.*, Cosmology intertwined III: $f\sigma_8$ and S_8, *Astropart. Phys.* **131**, p. 102604 (2021).
34. E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess and J. Silk, In the realm of the Hubble tension—a review of solutions, *Class. Quant. Grav.* **38**, p. 153001 (2021).
35. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Brans–Dicke Gravity with a Cosmological Constant Smoothes Out ΛCDM Tensions, *Astrophys. J. Lett.* **886**, p. L6 (2019).
36. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Brans–Dicke cosmology with a Λ-term: A possible solution to ΛCDM tensions, *Class. Quant. Grav.* **37**, p. 245003 (2020).
37. A. Gómez-Valent and P. Hassan Puttasiddappa, Difficulties in reconciling non-negligible differences between the local and cosmological values of the gravitational coupling in extended Brans-Dicke theories (5 2021).
38. F. Niedermann and M. S. Sloth, New early dark energy, *Phys. Rev. D* **103**, p. L041303 (2021).
39. A. Gogoi, R. K. Sharma, P. Chanda and S. Das, Early Mass-varying Neutrino Dark Energy: Nugget Formation and Hubble Anomaly, *Astrophys. J.* **915**, p. 132 (2021).

40. F. Niedermann and M. S. Sloth, Resolving the Hubble tension with new early dark energy, *Phys. Rev. D* **102**, p. 063527 (2020).
41. V. Poulin, T. L. Smith, T. Karwal and M. Kamionkowski, Early Dark Energy Can Resolve The Hubble Tension, *Phys. Rev. Lett.* **122**, p. 221301 (2019).
42. T. L. Smith, V. Poulin and M. A. Amin, Oscillating scalar fields and the Hubble tension: a resolution with novel signatures, *Phys. Rev. D* **101**, p. 063523 (2020).
43. P. Agrawal, F.-Y. Cyr-Racine, D. Pinner and L. Randall, Rock 'n' Roll Solutions to the Hubble Tension (4 2019).
44. J. C. Hill, E. McDonough, M. W. Toomey and S. Alexander, Early dark energy does not restore cosmological concordance, *Phys. Rev. D* **102**, p. 043507 (2020).
45. M. M. Ivanov, E. McDonough, J. C. Hill, M. Simonović, M. W. Toomey, S. Alexander and M. Zaldarriaga, Constraining Early Dark Energy with Large-Scale Structure, *Phys. Rev. D* **102**, p. 103502 (2020).
46. G. D'Amico, L. Senatore, P. Zhang and H. Zheng, The Hubble Tension in Light of the Full-Shape Analysis of Large-Scale Structure Data, *JCAP* **05**, p. 072 (2021).
47. M. Doran, M. J. Lilley, J. Schwindt and C. Wetterich, Quintessence and the separation of CMB peaks, *Astrophys. J.* **559**, 501 (2001).
48. V. Pettorino, L. Amendola and C. Wetterich, How early is early dark energy?, *Phys. Rev. D* **87**, p. 083009 (2013).
49. P. A. R. Ade *et al.*, Planck 2015 results. XIV. Dark energy and modified gravity, *Astron. Astrophys.* **594**, p. A14 (2016).
50. P. G. Ferreira and M. Joyce, Structure formation with a selftuning scalar field, *Phys. Rev. Lett.* **79**, 4740 (1997).
51. M. Doran, J.-M. Schwindt and C. Wetterich, Structure formation and the time dependence of quintessence, *Phys. Rev. D* **64**, p. 123520 (2001).
52. R. R. Caldwell, M. Doran, C. M. Mueller, G. Schafer and C. Wetterich, Early quintessence in light of WMAP, *Astrophys. J. Lett.* **591**, L75 (2003).
53. M. Grossi and V. Springel, The impact of early dark energy on non-linear structure formation, *Mon. Not. Roy. Astron. Soc.* **394**, p. 1559–1574 (Apr 2009).
54. T. Barreiro, E. J. Copeland and N. J. Nunes, Quintessence arising from exponential potentials, *Phys. Rev. D* **61**, p. 127301 (2000).
55. M. S. Turner and M. J. White, CDM models with a smooth component, *Phys. Rev. D* **56**, p. R4439 (1997).
56. G. Ballesteros and J. Lesgourgues, Dark energy with non-adiabatic sound speed: initial conditions and detectability, *JCAP* **10**, p. 014 (2010).
57. D. Blas, J. Lesgourgues and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, *JCAP* **07**, p. 034 (2011).
58. B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, *JCAP* **02**, p. 001 (2013).
59. D. Camarena and V. Marra, Local determination of the Hubble constant and the deceleration parameter, *Phys. Rev. Res.* **2**, p. 013028 (2020).
60. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *Astrophys. J.* **876**, p. 85 (2019).
61. M. J. Reid, D. W. Pesce and A. G. Riess, An Improved Distance to NGC 4258 and its Implications for the Hubble Constant, *Astrophys. J. Lett.* **886**, p. L27 (2019).

62. P. Carter, F. Beutler, W. J. Percival, C. Blake, J. Koda and A. J. Ross, Low Redshift Baryon Acoustic Oscillation Measurement from the Reconstructed 6-degree Field Galaxy Survey, *Mon. Not. Roy. Astron. Soc.* **481**, 2371 (2018).
63. E. A. Kazin *et al.*, The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature, *Mon. Not. Roy. Astron. Soc.* **441**, 3524 (2014).
64. T. M. C. Abbott *et al.*, Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1, *Mon. Not. Roy. Astron. Soc.* **483**, 4866 (2019).
65. H. du Mas des Bourboux *et al.*, The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests, *Astrophys. J.* **901**, p. 153 (2020).
66. S. Joudaki *et al.*, KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear, *Astron. Astrophys.* **638**, p. L1 (2020).
67. A. G. Sanchez, Arguments against using h^{-1}Mpc units in observational cosmology, *Phys. Rev. D* **102**, p. 123511 (2020).
68. K. Said, M. Colless, C. Magoulas, J. R. Lucey and M. J. Hudson, Joint analysis of 6dFGS and SDSS peculiar velocities for the growth rate of cosmic structure and tests of gravity, *Mon. Not. Roy. Astron. Soc.* **497**, 1275 (2020).
69. F. Simpson, C. Blake, J. A. Peacock, I. Baldry, J. Bland-Hawthorn, A. Heavens, C. Heymans, J. Loveday and P. Norberg, Galaxy and mass assembly: Redshift space distortions from the clipped galaxy field, *Phys. Rev. D* **93**, p. 023525 (2016).
70. C. Blake *et al.*, Galaxy And Mass Assembly (GAMA): improved cosmic growth measurements using multiple tracers of large-scale structure, *Mon. Not. Roy. Astron. Soc.* **436**, p. 3089 (2013).
71. C. Blake *et al.*, The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9, *Mon. Not. Roy. Astron. Soc.* **415**, p. 2876 (2011).
72. F. G. Mohammad *et al.*, The VIMOS Public Extragalactic Redshift Survey (VIPERS): Unbiased clustering estimate with VIPERS slit assignment, *Astron. Astrophys.* **619**, p. A17 (2018).
73. L. Guzzo *et al.*, A test of the nature of cosmic acceleration using galaxy redshift distortions, *Nature* **451**, 541 (2008).
74. Y.-S. Song and W. J. Percival, Reconstructing the history of structure formation using Redshift Distortions, *JCAP* **10**, p. 004 (2009).
75. T. Okumura *et al.*, The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at $z \sim 1.4$, *Publ. Astron. Soc. Jap.* **68**, p. 38 (2016).
76. G. Alestas, L. Kazantzidis and L. Perivolaropoulos, H_0 tension, phantom dark energy, and cosmological parameter degeneracies, *Phys. Rev. D* **101**, p. 123516 (2020).
77. A. Chudaykin, D. Gorbunov and N. Nedelko, Combined analysis of Planck and SPTPol data favors the early dark energy models, *JCAP* **08**, p. 013 (2020).
78. R. Murgia, G. F. Abellán and V. Poulin, Early dark energy resolution to the Hubble tension in light of weak lensing surveys and lensing anomalies, *Phys. Rev. D* **103**, p. 063502 (2021).
79. T. L. Smith, V. Poulin, J. L. Bernal, K. K. Boddy, M. Kamionkowski and R. Murgia, Early dark energy is not excluded by current large-scale structure data, *Phys. Rev. D* **103**, p. 123542 (2021).
80. V. Poulin, T. L. Smith, D. Grin, T. Karwal and M. Kamionkowski, Cosmological implications of ultralight axionlike fields, *Phys. Rev. D* **98**, p. 083525 (2018).

Renormalized $\rho_{\rm vac}$ without m^4 terms

Cristian Moreno-Pulido* and Joan Solà Peracaula

Departament de Física Quàntica i Astrofísica, and Institute of Cosmos Sciences, Universitat de Barcelona, Av. Diagonal 647, E-08028 Barcelona, Catalonia, Spain

**E-mails: cristian.moreno@fqa.ub.edu, sola@fqa.ub.edu*

The cosmological constant (CC) term, Λ, in Einstein's equations has been for about three decades a fundamental building block of the concordance or standard ΛCDM model of cosmology. Although the latter is not free of fundamental problems, it provides a good phenomenological description of the overall cosmological observations. However, an interesting improvement in such a phenomenological description and also a change in the theoretical status of the Λ-term occurs upon realizing that the vacuum energy is actually a "running quantity" in quantum field theory in curved spacetime. Several works have shown that this option can compete with the ΛCDM with a rigid Λ term. The so-called, "running vacuum models" (RVM) are characterized indeed by a vacuum energy density, $\rho_{\rm vac}$, which is evolving with time as a series of even powers of the Hubble parameter and its time derivatives. This form has been motivated by semi-qualitative renormalization group arguments in previous works. Here we review a recent detailed computation by the authors of the renormalized energy-momentum tensor of a non-minimally coupled scalar field with the help of adiabatic regularization procedure. The final result is noteworthy: $\rho_{\rm vac}(H)$ takes the precise structure of the RVM, namely a constant term plus a dynamical component $\sim H^2$ (which should be detectable in the present universe) including also higher order effects $\mathcal{O}(H^4)$ which can be of interest during the early stages of the cosmological evolution. Besides, it is most remarkable that such renormalized form of the vacuum energy density does not carry dangerous terms proportional to m^4, the quartic powers of the masses of the fields, which are a well-known source of exceedingly large contributions to the vacuum energy density and are directly responsible for extreme fine tuning in the context of the cosmological constant problem.

Keywords: Cosmology, Dark Energy, Quantum Field Theory.

1. Introduction

Since the mid nineties, the cosmological constant (CC) term, Λ, in Einstein's equations has been a crucial ingredient of the 'concordance' or standard ΛCDM model of cosmology.[1] From the phenomenological point of view it was favored only as of the time Λ became a physically measured quantity some twenty years ago.[2] Nowadays, precision cosmology[3] is able to do better measurements of Λ, or more precisely of the associated parameter $\Omega^0_{\rm vac} = \rho^0_{\rm vac}/\rho^0_c$, established by observations to be $\Omega^0_{\rm vac} \sim 0.7$. Here, the vacuum energy density (VED) is associated to Λ and defined as $\rho^0_{\rm vac} = \Lambda/(8\pi G_N)$, where G_N is Newton's constant and $\rho^0_c = 3H_0^2/(8\pi G_N)$ is the critical density in our times. However, there is a lot of uncertainty with the CC with respect its nature and origin at a fundamental level. For instance it is commonly

*Speaker

assumed that Λ may be related with the Zero-Point Energy (ZPE) of the quantum matter fields and with the Higgs potential of the Standard Model (SM). However, if ones compares the observational value $\rho_{\rm vac}$ with the ZPE of a typical SM particle such as the electron we are left with a tiny value of $\rho_{\rm vac}^{\rm obs}/\rho_{\rm ZPE} \sim 10^{-34}$. If compared with the ground state energy of the effective Higgs potential, $V_{\rm eff}$, we find $\rho_{\rm vac}^{\rm obs}/\langle V_{\rm eff}\rangle \sim 10^{-56}$. This issue is generically called the cosmological constant problem[4,5] and affects all forms of dark energy (DE).[6–10] Another aspect of this intriguing topic is the fact that, at the present time, $\rho_{\rm vac}^{\rm obs}$ and $\rho_{\rm CDM}$, the energy density associated to Cold Dark Matter (CDM), are observed to be of the same order of magnitude, despite the fact that $\rho_{\rm CDM}$ is assumed to decrease with the expansion as a^{-3}, being a the scale factor, whereas $\rho_{\rm vac}$ maintains constant.

The astonishing theoretical problems, though, are not the only ones concerning the concordance model. Currently, some important measurements seems to be in tension with the ΛCDM, emphasizing the discrepant values of the Hubble parameter at the present time, H_0, obtained from independent measurements of the early and the local universe.[11] It is still unknown if these tensions are the result of systematic errors, but the chance that a deviation from the ΛCDM model could provide an explanation for such discrepancies[12] is totally open.

As it has been shown in the literature, models mimicking a time-evolving $\rho_{\rm vac}$, which are allowed by the cosmological principle, could help in alleviating the mentioned problems, see e.g.[13–18] and.[19–25] In this work, we review recent work[26] on the possibility of having dynamical vacuum energy density (VED), $\rho_{\rm vac}$, in the context of quantum field theory (QFT) in curved spacetime with a spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric. We want to focus on the dynamics associated to the running vacuum model (RVM);[27–29] for a review, see[30–32] and references therein. Related studies are e.g.[33,34] and,[35,36] but also others extending the subject to the context of supersymmetric theories[37,38] and to supergravity.[39] Recently, some works in the framework of the effective action of string theories[40–45] have been dealed with the topic.

In the calculations reviewed here, the renormalization of the energy-momentum tensor is done using the adiabatic regularization procedure (ARP).[46–49] The renormalization process is a WKB approximation of the field modes in an expanding FLRW background. Once the VED is renormalized, obtained upon inclusion of the renormalized value of ρ_Λ (associated to Λ) at a given scale, it does not include the dangerous contributions depending on the fourth power of the particle masses ($\sim m^4$) and therefore it is free from huge induced corrections to the VED.[26] Additionally, we find a RVM-like form for the VED in the current times, $c_0 + \nu H^2$, with c_0 related to the usual constant term and $|\nu| \ll 1$. This last parameter represents only a small (dynamical) correction to the constant term and, depending on the sign of ν, it can mimic quintessence or phantom DE.

First, in Sec. 2 our framework is defined, consisting of a real scalar field non-minimally coupled to gravity, and we present the classical energy-momentum tensor (EMT). Later on, in Sections 3 and 4, we present the quantum fluctuations in the

adiabatic vacuum by means of a WKB expansion of the field modes in the FLRW background and we discuss the adiabatic regularization of the EMT. In Sec. 5 we start with the renormalization of the EMT in the FLRW context through the ARP, which is then needed in Sec. 6 to derive the precise result of $\rho_{\rm vac}$ from the renormalized ZPE up to terms of adiabatic order 4. In our case that means up to $\mathcal{O}(H^4)$. We also demonstrate that the values of the VED at different scales are related by an expression which does not contain quartic powers of the masses. To continue, in Sec. 7, we connect the computed VED in this work with the well-known running vacuum model (RVM), which originally was justified from renormalization group arguments in curved spacetime in previous works. A summary of the conclusions and discussion appear in Sec. 8. At the end, we define our conventions and collect some useful formulas in Appendix.

2. Energy-Momentum Tensor for non-minimally coupled scalar field

Einstein's equations read[a]

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}^{\rm matter}\,, \tag{1}$$

where $T_{\mu\nu}^{\rm matter}$ is the EMT of matter. They can also be written as

$$\frac{1}{8\pi G_N}G_{\mu\nu} + \rho_\Lambda g_{\mu\nu} = T_{\mu\nu}^{\rm matter}\,, \tag{2}$$

where $G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ and $\rho_\Lambda \equiv \Lambda/(8\pi G_N)$ is the VED associated to Λ. The associated contribution of Λ to $T_{\mu\nu}^{\Lambda} \equiv -\rho_\Lambda g_{\mu\nu}$. But, in general, there are more contributions to the total VED, associated to the quantum fluctuations of the fields, and also to their classical ground state energy. In this work, for practical reasons, we will assume that there is only one (matter) field in the form of a real scalar field, ϕ, and we will denote by $T_{\mu\nu}^{\phi}$ the piece of the EMT associated to it. Thus $T_{\mu\nu}^{\rm tot} = T_{\mu\nu}^{\Lambda} + T_{\mu\nu}^{\phi}$. We neglect other incoherent matter contributions such as dust and radiation because they can be added without altering the QFT considerations developed here.

A non-minimal coupling between the scalar field and gravity is assumed, without any classical potential for ϕ . The part of the action associated to ϕ is

$$S[\phi] = -\int d^4x\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\nu\phi\partial_\mu\phi + \frac{1}{2}(m^2+\xi R)\phi^2\right)\,, \tag{3}$$

where ξ is the non-minimal coupling between ϕ and gravity. In the special case $\xi = 1/6$, the massless ($m = 0$) action has conformal symmetry, i.e. symmetric under simultaneous rescalings of the $g_{\mu\nu}$ and ϕ with a local function $\alpha(x)$: $g_{\mu\nu} \to e^{2\alpha(x)}g_{\mu\nu}$ and $\phi \to e^{-\alpha(x)}\phi$. However, we will keep ξ general as it enables a richer phenomenology.

[a]Some useful geometrical quantities can be seen in the Appendix.

The Klein-Gordon (KG) equation is satisfied by the field ϕ,

$$(\Box - m^2 - \xi R)\phi = 0. \tag{4}$$

Here $\Box\phi = g^{\mu\nu}\nabla_\mu\nabla_\nu\phi = (-g)^{-1/2}\partial_\mu\left(\sqrt{-g}\, g^{\mu\nu}\partial_\nu\phi\right)$. On the other hand, the EMT takes this form :

$$T_{\mu\nu}(\phi) = -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = (1-2\xi)\partial_\mu\phi\partial_\nu\phi + \left(2\xi - \frac{1}{2}\right) g_{\mu\nu}\partial^\sigma\phi\partial_\sigma\phi$$
$$- 2\xi\phi\nabla_\mu\nabla_\nu\phi + 2\xi g_{\mu\nu}\phi\Box\phi + \xi G_{\mu\nu}\phi^2 - \frac{1}{2}m^2 g_{\mu\nu}\phi^2. \tag{5}$$

In this work we consider a spatially flat FLRW metric in the conformal frame. Being η the conformal time we have $ds^2 = a^2(\eta)\eta_{\mu\nu}dx^\mu dx^\nu$, with $\eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$ the Minkowski metric. Differentiation with respect the conformal time is denoted by $()' \equiv d\,()\,/d\eta$. It is convinient to define $\mathcal{H}(\eta) \equiv a'/a$ and, since $dt = ad\eta$, it is related with the usual Hubble rate $H(t) = \dot{a}/a$ as $\mathcal{H}(\eta) = aH(t)$. Despite the fact we will do intermediate calculations in conformal time, at the end it will be useful to express the VED in terms of the usual Hubble rate $H(t)$, because it is the easiest way to compare with other RVM results in the literature.

3. Quantum fluctuations and WKB ansatz

Now, we can take into account the quantum fluctuations of the field ϕ as an expansion around the background field (or classical mean field), ϕ_b:

$$\phi(\eta, x) = \phi_b(\eta) + \delta\phi(\eta, x). \tag{6}$$

The vacuum expectation value (VEV) of the field is $\langle 0|\phi(\eta,x)|0\rangle = \phi_b(\eta)$, whereas we assume that the VEV of the fluctuations vanishes: $\langle 0|\delta\phi|0\rangle = 0$. The vacuum state to which we are referring to is the so-called adiabatic vacuum, we will make some comments about this below.

The corresponding EMT decomposes itself as $\langle T^\phi_{\mu\nu}\rangle = \langle T^{\phi_b}_{\mu\nu}\rangle + \langle T^{\delta_\phi}_{\mu\nu}\rangle$, where $\langle T^{\phi_b}_{\mu\nu}\rangle = T^{\phi_b}_{\mu\nu}$ is the contribution from the background part and $\langle T^{\delta\phi}_{\mu\nu}\rangle$ is related with the quantum fluctuations. In particular, the $\langle T^{\delta\phi}_{00}\rangle$ is associated to the ZPE density of the scalar field in the FLRW background. Thus, the total vacuum contribution to the EMT reads

$$\langle T^{\rm vac}_{\mu\nu}\rangle = T^\Lambda_{\mu\nu} + \langle T^{\delta\phi}_{\mu\nu}\rangle = -\rho_\Lambda g_{\mu\nu} + \langle T^{\delta\phi}_{\mu\nu}\rangle\,. \tag{7}$$

This means that the total vacuum EMT receives contributions from the cosmological constant term and from the quantum fluctuations. A renormalized version of this equation will lead us to a renormalized VED, as we will see later.

The KG equation (4) is satisfied independently by the classical field and the quantum part (6). Let us concentrate on the fluctuation $\delta\phi$. Its Fourier decomposition in frequency modes $h_k(\eta)$ is

$$\delta\phi(\eta,\mathbf{x}) = \frac{1}{(2\pi)^{3/2}a}\int d^3k\left[A_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}h_k(\eta) + A^\dagger_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{x}}h^*_k(\eta)\right]. \tag{8}$$

$A_{\mathbf{k}}$ and $A^{\dagger}_{\mathbf{k}}$ are the (time-independent) annihilation and creation operators, with commutation relations

$$[A_{\mathbf{k}}, {A_{\mathbf{k}'}}^{\dagger}] = \delta(\mathbf{k}-\mathbf{k}'), \qquad [A_{\mathbf{k}}, A_{\mathbf{k}'}] = 0. \tag{9}$$

Introducing the Fourier expansion (8) in $\left(\Box - m^2 - \xi R\right)\delta\phi = 0$ we find that the frequency modes of the fluctuations satisfy

$$h_k'' + \Omega_k^2 h_k = 0\,, \qquad \Omega_k^2(\eta) \equiv \omega_k^2(m) + a^2\,(\xi - 1/6)R\,, \tag{10}$$

with $\omega_k^2(m) \equiv k^2 + a^2 m^2$. As we can see, h_k depends uniquely on the modulus $k \equiv |\mathbf{k}|$ of the momentum. Since $\Omega_k(\eta)$ is a nontrivial function of the conformal time, there is no closed form of the solution of (10). Instead of looking to an analytical solution, we will generate an approximate solution from a recursive method based on the phase integral ansatz

$$h_k(\eta) \sim \frac{1}{\sqrt{W_k(\eta)}} \exp\left(i\int^{\eta} W_k(\tilde{\eta})d\tilde{\eta}\right). \tag{11}$$

Therefore, the function $W_k(\eta)$ satisfies the following differential equation:

$$W_k^2 = \Omega_k^2 - \frac{1}{2}\frac{W_k''}{W_k} + \frac{3}{4}\left(\frac{W_k'}{W_k}\right)^2. \tag{12}$$

This non-linear differential equation can be solved using the WKB approximation. The solution is valid for large k (i.e. for short wave lengths) and the function Ω_k is slowly varying for weak fields. The notion of vacuum we work with it is called the adiabatic vacuum,[50–52] and can be defined as the quantum state annihilated by all the operators A_k of the Fourier expansion of the scalar field, see[46–49] for details. The physical interpretation of the modes (10) with frequencies dependending on time, must be understood in terms of field observables rather than in particle language. Therefore, for a more physical interpretation of the vacuum effects of the expanding background, we must compute the renormalized EMT in the FLRW spacetime, but first we need to regularize it.

4. Adiabatic regularization of the EMT

The adiabatic (slowly varying) Ω_k, is susceptible to be computed using Eq. (12) to generate an (asymptotic) series solution. In this context, the series is obtained through the adiabatic regularization procedure (ARP)[b] and it is organized in what we call adiabatic orders. First, the quantites considered of adiabatic order 0 are: k^2 and a. Of adiabatic order 1 are: a' and $\mathcal{H}$. Then $a'', a'^2, \mathcal{H}'$ and $\mathcal{H}^2$ are quantities

[b]This method was introduced for minimally coupled massive scalar fields in.[51,52] In,[50] it is generalized for other couplings. A review can be found in the classic books.[46,47] The ARP has been used in the context of QFT in curved backgrounds[35,36] and also it has been extended to other fields, such as the spin one-half fields in.[53,54]

of adiabatic order 2. We can sum up by saying that each extra derivative in conformal time increases the adiabatic order one unit. As a consequence, the "effective frequency" W_k can be written as an asymptotic expansion:

$$W_k = \omega_k^{(0)} + \omega_k^{(2)} + \omega_k^{(4)} + \dots, \tag{13}$$

where each $\omega_k^{(j)}$ is an adiabatic correction of order j. This leads to an expansion of the mode function h_k in even order adiabatic terms. This is justified by arguments of general covariance, since only terms of even adiabatic order (an even number of time derivatives) are allowed in the field equations.

4.1. *Introducing the renormalization parameter*

We start by defining the $0th$ order terms

$$\omega_k^{(0)} \equiv \omega_k = \sqrt{k^2 + a^2 M^2}. \tag{14}$$

In this approach the WKB expansion is performed off-shell, at an arbitrary mass scale M replacing the scalar field mass m in (14). In consequence, ARP can be formulated in such a way that we can relate the adiabatically renormalized theory at two different scales.[36] If $M = m$ we obtain the renormalized theory on-shell. In the computation of the EMT, the parameter $\Delta^2 \equiv m^2 - M^2$ will appear in the correction terms and it has to be considered of adiabatic order 2 since it appears in the WKB expansion together with other terms of the same adiabatic order.[36] If $\Delta = 0$, then $M = m$ and corresponds to the usual on-shell ARP, see.[46,47] With the help of this formalism one can explore the evolution of the VED throughout the cosmological history as it has been done for the first time in.[26] As we shall see, it will be convenient to consider m to be of the order of magnitude as the masses of Grand Unified Theory (GUT) fields in order to explore the dynamics of the VED in the low energy domain $M^2 \ll m^2$ (corresponding to the late universe). For simplicity, we model here all particles in terms of real scalar fields as in.[26] For a generalization to fermion fields, see.[55]

4.2. *Regularized ZPE*

Our starting point is the initial solution $W_k \approx \omega_k^{(0)}$ indicated in Eq. (14). For $a = 1$, it corresponds to the standard Minkowski space modes. Since $a = a(\eta)$ we have to find a better approximation. We use the initial solution $\omega_k^{(0)}$ in (12) and expand the RHS in powers of ω_k^{-1}, then collect the new terms up to adiabatic order 2 to find $\omega_k^{(2)}$. Now, we repeat the process with $W_k \approx \omega_k^{(0)} + \omega_k^{(2)}$ on the RHS of the same equation, expand again in ω_k^{-1}, collect contributions of adiabatic order 4... After some steps, the expansion seems to be organized in powers of $\omega_k^{-1} \sim 1/k$ (i.e. a short wavelength expansion). The UV divergent terms of the ARP are precisely the first lowest powers of ω_k^{-1}, which are present in the first adiabatic orders of the expansion. Higher adiabatic orders come later in the iteration, and represent finite

contributions[50–52] since they decay quickly at large k and the associated integrals are manifestly convergent. In our case, the divergent terms of the EMT are present up to $4th$ adiabatic order so that, at least, we have to compute all the terms up to this order (more detail below in Eq. (16)). After renormalization, we will obtain a finite expression for the EMT and then compute the vacuum energy density. But first, we need to consider the EMT associated to the fluctuations, using Eq. (5) and (6). For the 00-component,

$$\langle T_{00}^{\delta\phi}\rangle = \Big\langle \frac{1}{2}\left(\delta\phi'\right)^2 + \left(\frac{1}{2}-2\xi\right)\sum_i \partial_i\delta\phi\partial_i\delta\phi + 6\xi\mathcal{H}\delta\phi\delta\phi' - 2\xi\delta\phi\nabla^2\delta\phi + 3\xi\mathcal{H}^2\delta\phi^2 + \frac{a^2m^2}{2}\left(\delta\phi\right)^2\Big\rangle, \tag{15}$$

where $\nabla^2 \equiv \sum_{i=1}^3 \partial_i^2$ and $\left(\delta\phi'\right)^2 \equiv \left(\delta\partial_0\phi\right)^2 = \left(\partial_0\delta\phi\right)^2$. We may now substitute the Fourier expansion of $\delta\phi$, as given in (8), into Eq. (15) and apply the commutation relations (9). After symmetrizing the operator field product $\delta\phi\delta\phi'$ with respect to the annihilation and creation operators, we end up with

$$\begin{aligned}\langle T_{00}^{\delta\phi}\rangle = \frac{1}{8\pi^2a^2}\int dkk^2\Bigg[&2\omega_k + \frac{a^4M^4\mathcal{H}^2}{4\omega_k^5} - \frac{a^4M^4}{16\omega_k^7}(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 + 8\mathcal{H}'\mathcal{H}^2 + 4\mathcal{H}^4)\\ &+ \frac{7a^6M^6}{8\omega_k^9}(\mathcal{H}'\mathcal{H}^2 + 2\mathcal{H}^4) - \frac{105a^8M^8\mathcal{H}^4}{64\omega_k^{11}}\\ &+ \bar{\xi}\left(-\frac{6\mathcal{H}^2}{\omega_k} - \frac{6a^2M^2\mathcal{H}^2}{\omega_k^3} + \frac{a^2M^2}{2\omega_k^5}(6\mathcal{H}''\mathcal{H} - 3\mathcal{H}'^2 + 12\mathcal{H}'\mathcal{H}^2)\right.\\ &\left.- \frac{a^4M^4}{8\omega_k^7}(120\mathcal{H}'\mathcal{H}^2 + 210\mathcal{H}^4) + \frac{105a^6M^6\mathcal{H}^4}{4\omega_k^9}\right)\\ &+ \bar{\xi}^2\left(-\frac{1}{4\omega_k^3}(72\mathcal{H}''\mathcal{H} - 36\mathcal{H}'^2 - 108\mathcal{H}^4) + \frac{54a^2M^2}{\omega_k^5}(\mathcal{H}'\mathcal{H}^2 + \mathcal{H}^4)\right)\Bigg]\\ &+ \frac{1}{8\pi^2a^2}\int dkk^2\Bigg[\frac{a^2\Delta^2}{\omega_k} - \frac{a^4\Delta^4}{4\omega_k^3} + \frac{a^4\mathcal{H}^2M^2\Delta^2}{2\omega_k^5} - \frac{5}{8}\frac{a^6\mathcal{H}^2M^4\Delta^2}{\omega_k^7}\\ &+ \bar{\xi}\left(-\frac{3a^2\Delta^2\mathcal{H}^2}{\omega_k^3} + \frac{9a^4M^2\Delta^2\mathcal{H}^2}{\omega_k^5}\right)\Bigg] + \ldots,\end{aligned} \tag{16}$$

where we have defined $\bar{\xi} \equiv \xi - 1/6$ and we have integrated $\int \frac{d^3k}{(2\pi)^3}(...)$ over solid angles and expressed the final integration in terms of $k = |\mathbf{k}|$. Let us note the presence of the Δ-dependent terms in the last two rows, which contribute at $2th$ (Δ^2) and $4th$ (Δ^4) adiabatic order.

5. Renormalization of the ZPE in curved spacetime

The ZPE part of the EMT, as given by Eq. (16) can be split into two parts as follows:

$$\langle T_{00}^{\delta\phi}\rangle(M) = \langle T_{00}^{\delta\phi}\rangle_{Div}(M) + \langle T_{00}^{\delta\phi}\rangle_{Non-Div}(M), \tag{17}$$

where

$$\langle T_{00}^{\delta\phi}\rangle_{Div}(M) = \frac{1}{8\pi^2 a^2}\int dk k^2 \Bigg[2\omega_k + \frac{a^2\Delta^2}{\omega_k} - \frac{a^4\Delta^4}{4\omega_k^3} - 6\bar{\xi}\mathcal{H}^2\left(\frac{1}{\omega_k} + \frac{a^2M^2}{\omega_k^3} + \frac{a^2\Delta^2}{2\omega_k^3}\right) - 9\frac{\bar{\xi}^2}{\omega_k^3}(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4)\Bigg], \tag{18}$$

which containts the powers $1/\omega_k^n$ up to $n = 3$, manifestly UV-divergent.

On the other hand, the non-divergent part of (17) involves the integrals with powers of $1/\omega_k$ higher than 3 which are perfectly finite. Computing these convergent integrals (see Eq. (45) in Appendix), the result reads

$$\langle T_{00}^{\delta\phi}\rangle_{Non-Div}(M) = \frac{m^2\mathcal{H}^2}{96\pi^2} - \frac{1}{960\pi^2 a^2}\left(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 2\mathcal{H}^4\right) + \bar{\xi}\frac{3\Delta^2\mathcal{H}^2}{8\pi^2} + \frac{1}{16\pi^2 a^2}\bar{\xi}\left(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4\right) + \frac{9}{4\pi^2 a^2}\bar{\xi}^2(\mathcal{H}'\mathcal{H}^2 + \mathcal{H}^4) + \ldots \tag{19}$$

where the dots in the last expression correspond to higher adiabatic orders. Let us now focus on the divergent part of the ZPE, Eq. (18). The adiabatic series is an asymptotic series representation of Eq. (15). Thefore, such series is not convergent but it provides an approximation, which is obtained once the series is cut at a particular order.[36] As a consequence there is some arbitrariness in the way of choosing the leading adiabatic order, which we can use in our favour. There is, nonetheless, some previous steps to do before obtaining a meaningful result. First, we are going to set the arbitrary scale at the mass of the scalar field. That is, $M = m$ and hence $\Delta = 0$ (cf. Sec. 4.1). The divergent part (18) is reduced in this case to

$$\langle T_{00}^{\delta\phi}\rangle_{Div}(m) = \frac{1}{8\pi^2 a^2}\int dk k^2 \Bigg[2\omega_k(m) - \bar{\xi}6\mathcal{H}^2\left(\frac{1}{\omega_k(m)} + \frac{a^2m^2}{\omega_k^3(m)}\right) - \bar{\xi}^2\frac{9}{\omega_k^3(m)}(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4)\Bigg]. \tag{20}$$

What we are going to do next, in order to renormalize the ZPE and the EMT, is to subtract the terms that appear up to $4th$ adiabatic order at the arbitrary mass scale M. This procedure is enough to cancel the divergent terms through the ARP, as it is discussed in the literature.[46–48]

5.1. *Renormalized ZPE for generic M*

As said before, our proposal for the renormalized ZPE in curved spacetime at the scale M is[26]:

$$\begin{aligned}\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M) &= \langle T_{00}^{\delta\phi}\rangle(m) - \langle T_{00}^{\delta\phi}\rangle^{(0-4)}(M)\\ &= \langle T_{00}^{\delta\phi}\rangle_{Div}(m) - \langle T_{00}^{\delta\phi}\rangle_{Div}(M) - \bar{\xi}\frac{3\Delta^2\mathcal{H}^2}{8\pi^2} + \ldots,\end{aligned} \tag{21}$$

here $(0-4)$ means that the expansion is up to fourth adiabatic order and the dots in (21) denote perfectly finite terms of higher adiabatic order. Using now Eq. (20), we arrive to

$$\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M) = \frac{1}{8\pi^2 a^2}\int dk k^2\left[2\omega_k(m) - \frac{a^2\Delta^2}{\omega_k(M)} + \frac{a^4\Delta^4}{4\omega_k^3(M)} - 2\omega_k(M)\right]$$
$$+\bar{\xi}\frac{6\mathcal{H}^2}{8\pi^2 a^2}\left\{\int dk k^2\left[\frac{1}{\omega_k(M)} + \frac{a^2M^2}{\omega_k^3(M)} + \frac{a^2\Delta^2}{2\omega_k^3(M)} - \frac{1}{\omega_k(m)} - \frac{a^2m^2}{\omega_k^3(m)}\right] - \frac{a^2\Delta^2}{2}\right\}$$
$$-\bar{\xi}^2\frac{9\left(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4\right)}{8\pi^2 a^2}\int dk k^2\left[\frac{1}{\omega_k^3(m)} - \frac{1}{\omega_k^3(M)}\right] + \ldots \tag{22}$$

In this equation we have introduced new notation in order to distinguish between the off-shell energy mode $\omega_k(M) \equiv \sqrt{k^2 + a^2M^2}$ (formerly denoted just as ω_k) and the on-shell one $\omega_k(m) \equiv \sqrt{k^2 + a^2m^2}$. The involved computations of these convergent integrals are not so straightforward but, after some algebra and the help of Mathematica,[56] we find

$$\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M) = \frac{a^2}{128\pi^2}\left(-M^4 + 4m^2M^2 - 3m^4 + 2m^4\ln\frac{m^2}{M^2}\right)$$
$$-\bar{\xi}\frac{3\mathcal{H}^2}{16\pi^2}\left(m^2 - M^2 - m^2\ln\frac{m^2}{M^2}\right) + \bar{\xi}^2\frac{9\left(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4\right)}{16\pi^2 a^2}\ln\frac{m^2}{M^2} + \ldots \tag{23}$$

The obtained expression vanishes for $M = m$ as expected from Eq. (21). However, let us remind the reader that this only happens because we have computed the on-shell value $\langle T_{\mu\nu}^{\delta\phi}\rangle_{\rm Ren}(m)$ up to adiabatic order 4 in Eq. (23). Nevertheless, $\langle T_{\mu\nu}^{\delta\phi}\rangle_{\rm Ren}(m)$ can be computed up to an arbitrary, but finite, adiabatic order. Beyond $4th$ order one always obtains subleading and perfectly finite corrections. We expect that these effects become suppressed when the physical mass m of the quantum field is large, so that they satisfy the Appelquist-Carazzone decoupling theorem.[57]

So far, we have obtained a renormalized ZPE in curved spacetime (23) which, despite the fact it is finite, it includes the quartic powers of the masses corresponding to undesired contributions to the vacuum energy.

6. Renormalized vacuum energy density

The renormalization in the context of QFT in curved spacetime implies the consideration of the higher derivative (HD) terms in the classical effective action,[46] beyond the usual Einstein-Hilbert (EH) term with a cosmological constant, Λ. In particular, in the FLRW background in four dimensions, the only surviving HD term is $H_{\mu\nu}^{(1)}$, as suggested by the geometric structure of (23). $H_{\mu\nu}^{(1)}$ is obtained by functionally differentiating R^2 with respect the metric (see Appendix). So, the full action contains the EH+HD terms and also the matter part which, in our simplified case, only consists of a scalar field ϕ non-minimally coupled to gravity, Eq. (3).

As usual, the modified Einstein's equations are obtained through the variation of the new action with respect to $g_{\mu\nu}$, extending (2) as follows:

$$\frac{1}{8\pi G_N(M)}G_{\mu\nu}+\rho_\Lambda(M)g_{\mu\nu}+a_1(M)H^{(1)}_{\mu\nu}=T^{\phi_b}_{\mu\nu}+\langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M)\,, \tag{24}$$

where the dependence in M indicates we are considering renormalized quantities. In particular, the classical part of the matter EMT does not depend on it. Our goal is to relate the theory at different renormalization points[c], and with this purpose in mind let us subtract Einstein's equations as written in (24) at two different scales, M and M_0. We find

$$\langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M)-\langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M_0)=f_{G_N^{-1}}G_{\mu\nu}+f_{\rho_\Lambda}g_{\mu\nu}+f_{a_1}H^{(1)}_{\mu\nu}, \tag{25}$$

here $f_X(m,M,M_0)\equiv X(M)-X(M_0)$ for the various couplings involved $X=G_N^{-1},\rho_\Lambda,a_1$. Since we know the expression of the renormalized EMT within the ARP, namely Eq (23), we can obtain the renormalization shift of the couplings G_N^{-1}, ρ_Λ and a_1 in (25) between the two scales M and M_0. This is possible by taking the value of the components G_{00} and $H^{(1)}_{00}$ (from Appendix) and comparing with (23). Notice that the first term of (23), in the *r.h.s*, is of $0th$ adiabatic order and is associated to $f_{\rho_\Lambda}(m,M,M_0)$ and, therefore, to the running of ρ_Λ. By doing this comparision we find that the functions are

$$\begin{aligned}
f_{G_N^{-1}}(m,M,M_0)&=\frac{\overline{\xi}}{16\pi^2}\left[M^2-M_0^2-m^2\ln\frac{M^2}{M_0^2}\right],\\
f_{\rho_\Lambda}(m,M,M_0)&=\frac{1}{128\pi^2}\left(M^4-M_0^4-4m^2(M^2-M_0^2)+2m^4\ln\frac{M^2}{M_0^2}\right),\\
f_{a_1}(M,M_0)&=-\frac{\overline{\xi}^2}{32\pi^2}\ln\frac{M^2}{M_0^2}.
\end{aligned} \tag{26}$$

6.1. *Running Vacuum Energy Density. Absence of $\sim m^4$ terms*

At this point we can come back to the definition of vacuum energy (7) which was used in this work. Considering $\langle T^{\rm vac}_{\mu\nu}\rangle=p_{\rm vac}g_{\mu\nu}+(p_{\rm vac}+\rho_{\rm vac})\,u_\mu u_\nu$, with $u_\mu=(a,0,0,0)$, and equating with Eq. (7), and taking the 00-component of the equality (keeping also in mind that $g_{00}=-a^2(\eta)$ in the conformal frame), we obtain

$$\rho_{\rm vac}(M)=\rho_\Lambda(M)+\frac{\langle T^{\delta\phi}_{00}\rangle_{\rm Ren}(M)}{a^2}. \tag{27}$$

We write the explicit dependence in the renormalization point M since we are talking about the renormalized quantity at that scale. Equation (27) means that

[c]We do not pretend to compute the renormalized couplings from first principles, in particular the VED. Renormalization program allows us to compare the theory at different renormalization points but, at the end, an input from the experiment is needed to predict its value at another scale.

the total VED at an arbitrary scale M is not only receiving contributions from the renormalized cosmological term but also from the quantum fluctuations of fields (i.e. from the renormalized ZPE of the scalar field, in our simplified model). Subtracting the renormalized result at two scales, M and M_0, and using (25), we find:

$$\begin{aligned}
&\rho_{\rm vac}(M)-\rho_{\rm vac}(M_0)=\rho_\Lambda(M)-\rho_\Lambda(M_0)+\frac{\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M)-\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M_0)}{a^2}\\
&=f_{\rho_\Lambda}(m,M,M_0)+\frac{f_{G_N^{-1}}(m,M,M_0)G_{00}+f_{\rho_\Lambda}(m,M,M_0)g_{00}+f_{a_1}(M,M_0)H_{00}^{(1)}}{a^2}\\
&=\frac{f_{G_N^{-1}}(m,M,M_0)}{a^2}G_{00}+\frac{f_{a_1}(M,M_0)}{a^2}H_{00}^{(1)}\\
&=\frac{3\mathcal{H}^2}{a^2}\,f_{G_N^{-1}}(m,M,M_0)-\frac{18}{a^4}\left(\mathcal{H}'^2-2\mathcal{H}''\mathcal{H}+3\mathcal{H}^4\right)\,f_{a_1}(M,M_0)\,. \qquad (28)
\end{aligned}$$

As we see the term $f_{\rho_\Lambda}(m,M,M_0)$ has disappeared. Finally, from (26) we obtain[26]

$$\begin{aligned}
\rho_{\rm vac}(M)=\rho_{\rm vac}(M_0)&+\frac{3}{16\pi^2}\overline{\xi}H^2\left[M^2-M_0^2-m^2\ln\frac{M^2}{M_0^2}\right]\\
&+\frac{9}{16\pi^2}\overline{\xi}^2\left(\dot{H}^2-2H\ddot{H}-6H^2\dot{H}\right)\ln\frac{M^2}{M_0^2}\,. \qquad (29)
\end{aligned}$$

In addition, we have used Eq. (44) from Appendix to write the final result in terms of the Hubble function in cosmic time ($H=\mathcal{H}/a$). The result (29) is the value of the VED at a particular energy scale M, related with the value of the VED at another renormalization scale M_0, i.e. it expresses the 'running' of the VED. Notice that if $\xi=1/6$ ($\overline{\xi}=0$), then the VED is independent of the renormalization scale. However, this is not likely to happen since one also has to consider the contribution from fermions and vector boson fields. It is also important to notice that the running is slow for small H, as it depends as $\mathcal{O}(H^2)$ times a mass scale squared and on $\mathcal{O}(H^4)$ contributions. Finally, we remark that there is no contributions of quartic mass scales.

7. Running vacuum connection

As explained in a footnote in previous sections, the result we were seeking (represented now by Eq. (29)) was not the calculated value of the VED at a given scale, e.g. it says nothing on the actual value of $\rho_{\rm vac}(M_0)$ and hence it has nothing to do with the cosmological constant problem mentioned in the Introduction. In other words, we have not provided the calculation of the value itself of the vacuum energy at particular point of the cosmic history. This result can nevertheless be useful to study the 'running' of the VED from one scale to another. We can rephrase this by saying that if $\rho_{\rm vac}$ is known at some scale M_0, Eq. (29) can be used to obtain the value of $\rho_{\rm vac}$ at another scale M. Such connection was suggested long ago from renormalization group arguments in curved spacetime[27–29] – see[30,31] and references therein for a review of the running vacuum model (RVM). What is more, it can even

set a framework for the possible variation of the so-called fundamental constants of nature[58] with the evolution of the universe. In the next subsection we are going to suggest a possible interpretation of Eq. (29) in the context of the RVM. For this purpose, let us assume that we set one arbitrary scale to the renormalized VED at some Grand Unified Theory (GUT) scale (i.e., set $M_0 = M_X$, where $M_X \sim 10^{16}$ GeV). This value is associated also with the inflationary scale. It is natural to assume that fundamental parameters, such as e.g. $\rho_{\rm vac}$, are defined at that primeval scale, related with the very beginning in the cosmological history. Additionally, a large scale such as the GUT scale insures that all particle masses can be active degrees of freedom to some extent.

7.1. *RVM in the currect universe*

Equation (29) can, in principle, be used to explore the value of the VED throughout the cosmological history. However, for the study of the very early universe (in particular, inflation) other contributions can appear, which will not be addressed here since in the present work we are just going to study the current universe dynamics in order to compare with the late time RVM phenomenology. For related studies in different contexts, cf.[59–62]. See also[42–45] for a stringy version of the RVM with implications on the mechanism of inflation.

Consider $\rho_{\rm vac}(M_X)$, the value of the VED at $M_0 = M_X$. The value of this energy density is unknown, but our aim is to relate it with the current value of the VED, $\rho^0_{\rm vac} \equiv \rho_{\rm vac}(H_0)$, by fixing the numerical value of the second scale at $M = H_0$, where H_0 today's value of the Hubble function– it can be considered an estimation for the energy scale associated to the the present FLRW universe. This association is commonly made in the aforementioned references on the RVM.[30] So, (29) applied to the present universe is

$$\rho^0_{\rm vac} \approx \rho_{\rm vac}(M_X) - \frac{3}{16\pi^2}\overline{\xi} H_0^2 \left[M_X^2 + m^2 \ln \frac{H_0^2}{M_X^2}\right]. \tag{30}$$

Here we ignore all terms of order $\mathcal{O}(H^4)$ (including $\dot{H}^2, H\ddot{H}$ and $H^2\dot{H}$) for the present universe ($H = H_0$). We can use this relation to find the value of $\rho_{\rm vac}(M_X)$, and the result is

$$\rho_{\rm vac}(M_X) = \rho^0_{\rm vac} - \frac{3\nu_{\rm eff}}{8\pi} H_0^2 M_P^2 , \tag{31}$$

where we have defined $\nu_{\rm eff}$, the 'running parameter' for the VED:

$$\nu_{\rm eff} = -\frac{\overline{\xi}}{2\pi}\frac{M_X^2}{M_P^2}\left(1 + \frac{m^2}{M_X^2}\ln\frac{H_0^2}{M_X^2}\right). \tag{32}$$

Notice that for $\xi = 1/6$ (or $\overline{\xi} = 0$), it vanishes and there is no running, as mentioned previously. In the case $\xi = 0$ (or $\overline{\xi} = -1/6$) and $m^2/M_X^2 \ll 1$ we obtain $\nu_{\rm eff} \simeq \frac{1}{12\pi}\frac{M_X^2}{M_P^2} \ll 1$. More in general, the structure obtained for $\nu_{\rm eff}$ is very similar to that one obtained previously in other works with the RVM approach, see.[30]

In these contexts, the parameter is related with the one-loop β-function for the renormalization group equation of $\rho_{\rm vac}$. Although, in our case, there is and additional logarithmic contribution $\ln H_0^2/M_X^2$, coming from the direct QFT calculation developed here. At the end of the day it does not make significant differences in practice since $\nu_{\rm eff}$ is a constant parameter which is fitted directly to cosmological data as an effective coefficient. It is natural to expect a small $\nu_{\rm eff}$, i.e. $|\nu_{\rm eff}| \ll 1$, because of the ratio $M_X^2/M_P^2 \sim 10^{-6}$, but notice that $\nu_{\rm eff}$ depends on ξ and also on contributions from other fields (fermions and bosons) and their multiplicities. Heavy fields of the order of the GUT scale can contribute in a significant way ($m \sim M_X$). As mentioned, $\nu_{\rm eff}$ needs to be fitted the RVM to tbe cosmological data, and this has been done in detail e.g. in,[15,16] the results indicate that $\nu_{\rm eff} \sim 10^{-3}$.

Now, we can us Eq. (29) and (31) to study the late universe. By ignoring $\mathcal{O}(H^4)$, we may estimate the current VED by taking $M = H$ around the present times:

$$\rho_{\rm vac}(H) = \rho_{\rm vac}^0 - \frac{3\nu_{\rm eff}}{8\pi}\, H_0^2\, M_P^2 + \frac{3\nu_{\rm eff}(H)}{8\pi}\, H^2\, M_P^2\,, \tag{33}$$

where

$$\nu_{\rm eff}(H) = -\frac{\overline{\xi}}{2\pi}\frac{M_X^2}{M_P^2}\left(1 + \frac{m^2}{M_X^2}\ln\frac{H^2}{M_X^2}\right). \tag{34}$$

Notice that the last expression is not constant as a difference from (32). However, since the dynamical behaviour of $\nu_{\rm eff}(H)$ is logarithmic, and considering a value of H close to H_0, we can approximate $\nu_{\rm eff}(H)$ by (32). Then, equation (33) may be written as[55]

$$\rho_{\rm vac}(H) \simeq \rho_{\rm vac}^0 + \frac{3\nu_{\rm eff}}{8\pi}\,(H^2 - H_0^2)\, M_P^2 = \rho_{\rm vac}^0 + \frac{3\nu_{\rm eff}}{8\pi G_N}\,(H^2 - H_0^2)\,. \tag{35}$$

This is exactly the canonical form of the RVM formula.[30] Such approximation holds fine even if we study the CMB epoch, since the difference between $\nu_{\rm eff}(H)$ from $\nu_{\rm eff}$ can be estimated to be less than 8% (for $m \simeq M_X$) or much less when $m \ll M_X$. This is the typical parametrization that has been used in the literature about the RVM.[15,16] As a matter of fact, this parametrization can be promoted to a more general form, by considering also a term proportional to $\dot{H}$ in the running equation for the VED. In the following subsection we are going to give grounds for the appearence of these kind of factors, not yet present in (35).

7.2. *More geometric structures for vacuum in curved spacetime*

Using the definition of the EMT associated to vacuum given in (7) we have derived the expression of the vacuum energy density in equation (27). However, we expect that the definition (7) can be generalized as follows:

$$\left\langle T_{\mu\nu}^{\rm vac}\right\rangle = T_{\mu\nu}^{\Lambda} + \left\langle T_{\mu\nu}^{\delta\phi}\right\rangle + \alpha_1 R g_{\mu\nu} + \alpha_2 R_{\mu\nu} + \mathcal{O}(R^2). \tag{36}$$

$\mathcal{O}(R^2)$ represents tensors of adiabatic order 4, that is R^2, $R_{\mu\nu}R^{\mu\nu}$, ... and α_i are parameters of dimension +2 in natural units. This new form for the vacuum EMT

can be justified from the phenomenological point of view, but we also remind the reader that in a more realistic picture, contributions from other fields (bosons and fermions) are also expected, and general covariance leads to a generic form as represented in (36). Generation of new terms is also possible by considering string-inspired mechanisms as it has been shown recently in.[40–43]

Let us work with the more general form for the vacuum EMT, (36). It is useful to redefine the coefficients of that expression as $\alpha_i(M) \equiv \frac{\lambda_i}{16\pi^2} M^2$. We can check in Appendix the expression of R and R_{00} in flat FLRW spacetime, we obtain after a straightforward calculation:

$$\rho_{\rm vac}(H) \simeq \rho^0_{\rm vac} + \frac{3}{8\pi G_N}\left(\tilde{\nu}_{\rm eff}\, H^2 + \bar{\nu}\,\dot{H}\right), \tag{37}$$

where

$$\tilde{\nu}_{\rm eff} = \frac{1}{2\pi}\,\frac{M_X^2}{M_P^2}\left[\left(\frac{1}{6}-\xi\right)\left(1+\frac{m^2}{M_X^2}\ln\frac{H_0^2}{M_X^2}\right)+4\lambda_1+\lambda_2\right] \tag{38}$$

and

$$\bar{\nu} = \frac{2\lambda_1+\lambda_2}{2\pi}\,\frac{M_X^2}{M_P^2}\,. \tag{39}$$

As done before, the logarithmic part of $\tilde{\nu}_{\rm eff}(H)$ can be neglected and we may approximate $\tilde{\nu}_{\rm eff}(H_0) \simeq \tilde{\nu}_{\rm eff}$. The parameter $\bar{\nu}$ is treated as constant here, but we can not discard also a possible running of λ_i, acquiring a dependence in M. Although, if this were the case, it is reasonable to expect a logarithmic evolution of the renormalization effects. The above formula is a generalization of Eq. (35) incorporating the additional coefficient $\bar{\nu}$, accompanying $\sim \dot{H}$. Finally, let us comment that several generalized forms of the RVM beyond the usual one containing the H^2 term has been studied before in the literature, see for instance.[13,14]

8. Discussion and conclusions

In this presentation we have investigated the possible dynamics of vacuum in the context of QFT in FLRW spacetime. Specifically, we have revisited the calculation of the renormalized energy-momentum tensor (EMT) of a real quantum scalar field non-minimally coupled to the FLRW background.[26] The approach is based on adiabatic regularization and renormalization of the EMT, starting from the WKB approximation of the field modes in curved spacetime. The renormalized EMT is defined as the difference of its on-shell value and its value at an arbitrary renormalization point M up to $4th$ adiabatic order. It is sufficient with this subtraction since the divergent terms are present up to $4th$ adiabatic order. The renormalized EMT is perfectly finite and acquires a dependence on M. We can use this fact to compare the renormalized result at different times of the history of the universe characterized by different energy scales. At this point we can identify the VED from the renormalized EMT. Such VED depends on $\rho_\Lambda(M)$ but also in the ZPE part involving

the quantum fluctuations of the scalar field. An interesting implication is that the combination of both quantities is free from terms $\sim m^4$, well-known sources of large contributions to the VED.

This QFT calculation leads us to a renormalized VED which has the usual form of the running vacuum models (RVM's),[30] in which $\rho_{\rm vac} = \rho_{\rm vac}(H)$ consists of an additive constant plus a series of powers of H (the Hubble rate) and its time derivatives. In previous works, the RVM was motivated from general renormalization group arguments in QFT in curved spacetime (cf.[30] and references therein). But, in this work, we have reviewed proof that the RVM form of the VED for the current universe emerges also from this calculation of QFT in the FLRW spacetime involving adiabatic regularization. It is found that the powers of H (and its time derivatives) in the EMT carry an even number of time derivatives of the scale factor, which is mandatory if we impose the general covariance of the action. In particular, the lowest order dynamical component of the VED is $\sim \nu\, H^2$. The dimensionless coefficient ν is expected to be small ($|\nu| \ll 1$). However, at the end, the model has to be fitted with the help of a complet set of cosmological data. This lower order term $\sim H^2$ has the main role in the study of the dynamics of the vacuum in the late universe, whereas the higher order terms may play a major role in the very early universe, for example to describe inflation (despite the fact that we have not exploited this idea in this work). In previous works, the RVM has been succesfully confronted to a large number of cosmological data and the effective parameter ν has been fitted. The encountered results say that ν is positive and around $\sim 10^{-3}$, see[13–16] . To conclude, let us remark that our QFT calculation has some limitations, for instance it has been simplified by the use of just one single field in the form of a real quantum scalar field. For this reason, further investigations have to be done in order to generalize these results for multiple fields involving scalar, fermionic and vectorial degrees of freedom. However, we do not expect major changes beyond computational details and the main results presented here are likely to be maintained.[55]

Acknowledgements

This work is based on the invited talk presented in the CM3 parallel session: "Status of the H_0 and σ_8 Tensions: Theoretical Models and Model-Independent Constraints" of the MG16 Marcel Grossmann virtual Conference, July 5-10 2021. We would like to thank the organizers of MG16 as well as the chairpersons of the session for the excellence of the event and for the invitation.

Work partially funded by projects PID2019-105614GB-C21 and FPA2016-76005-C2-1-P (MINECO, Spain), 2017-SGR-929 (Generalitat de Catalunya) and CEX2019-000918-M (ICCUB). CMP is also partially supported by the fellowship 2019 FI_B 00351 (Generalitat de Catalunya). JSP acknowledges participation in the COST Association Action CA18108 "*Quantum Gravity Phenomenology in the Multimessenger Approach (QG-MM)*".

Appendix: Conventions, geometrical quantities and useful formulas

Through this work we have used natural units. This means that the gravitational constant is written as $G_N = 1/M_P$, being M_P is the Planck mass and $\hbar = c = 1$.

Conventions regarding the geometrical quantities can be summarized as follows: the metric has the signature $(-,+,+,+)$; the Riemann tensor is $R^{\lambda}{}_{\mu\nu\sigma} = \partial_\nu \Gamma^{\lambda}_{\mu\sigma} + \Gamma^{\rho}_{\mu\sigma}\Gamma^{\lambda}_{\rho\nu} - (\nu \leftrightarrow \sigma)$; the Ricci tensor is $R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}$; and finally, Ricci scalar reads $R = g^{\mu\nu}R_{\mu\nu}$. These elections correspond to $(+,+,+)$ in the Misner-Thorne-Wheeler[63] conventions.

On the other hand, the Einstein tensor is defined as $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ and the field equations are $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}$. The (non-vanishing) Christoffel symbols, with a line element in terms of the conformal time $ds^2 = a^2(\eta)\eta_{\mu\nu}dx^\mu dx^\nu$ and $\eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$, are:

$$\Gamma^0_{00} = \mathcal{H}, \qquad \Gamma^0_{ij} = \mathcal{H}\delta_{ij}, \qquad \Gamma^i_{j0} = \mathcal{H}\delta^i_j\,. \tag{40}$$

The Ricci scalar and the 00-components of the curvature tensors are:

$$R = 6\frac{a''}{a^3} = \frac{6}{a^2}\left(\mathcal{H}' + \mathcal{H}^2\right) = 6\left(\frac{\dot{a}^2}{a^2} + \frac{\ddot{a}}{a}\right) = 6(2H^2 + \dot{H}) \tag{41}$$

and

$$R_{00} = -3\mathcal{H}' = -3a^2(H^2 + \dot{H})\,, \qquad G_{00} = 3\mathcal{H}^2 = 3a^2H^2\,. \tag{42}$$

Primes are indicating differentiation with respect to conformal time, meanwhile we use dots to represent derivatives with respect cosmic time. The process of renormalization introduces the need for the higher order curvature tensor $H^{(1)}_{\mu\nu}$, obtained by the variation of R^2 with respect the metric in the higher derivative vacuum action:

$$H^{(1)}_{\mu\nu} = \frac{1}{\sqrt{-g}}\frac{\delta}{\delta g^{\mu\nu}}\int d^4x\sqrt{-g}R^2 = -2\nabla_\mu\nabla_\nu R + 2g_{\mu\nu}\Box R - \frac{1}{2}g_{\mu\nu}R^2 + 2RR_{\mu\nu}\,. \tag{43}$$

In particular, the 00-component is related with the vacuum energy density and in the conformally flat metric is

$$H^{(1)}_{00} = \frac{-18}{a^2}\left(\mathcal{H}'^2 - 2\mathcal{H}''\mathcal{H} + 3\mathcal{H}^4\right) = -18a^2\left(\dot{H}^2 - 2H\ddot{H} - 6H^2\dot{H}\right)\,. \tag{44}$$

Computing integrals with powers of $1/\omega_k$

The following expression is useful to compute all the finite ($n > N$) and UV-divergent ($n \leq N$) integrals involved in the EMT computation:

$$I(n,Q) \equiv \int \frac{d^Nk}{(2\pi)^N}\frac{1}{(k^2+Q^2)^{n/2}} = \frac{1}{(4\pi)^{N/2}}\frac{\Gamma\left(\frac{n-N}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(Q^2\right)^{\frac{N-n}{2}}\,, \tag{45}$$

where $k \equiv |\mathbf{k}|$ and Q is an arbitrary scale. We have corrected typos in this formula as compared to that of Appendix B of,[26] with no consequences in the calculation.

References

1. P. J. E. Peebles, Astrophys. J. **284** 439 (1984); *Principles of Physical Cosmology* (Princeton Univ. Press, Princeton, 1993).
2. A. G. Riess *et al.,* Astron. J. **116** 1009 (1998); S. Perlmutter *et al.,* Astrophys. J. **517** 565 (1999).
3. N. Aghanim *et al.* [Planck Collab.], Astron. Astrophys. **641** A6 (2020); P.A.R. Ade *et al.* [Planck Collab.], Astron. Astrophys. **594** A13 (2016).
4. S. Weinberg, Rev. Mod. Phys. **61** 1 (1989).
5. E. Witten, *The Cosmological Constant From The Viewpoint Of String Theory*, Contribution to the 4*th* International Symposium on Sources and Detection of Dark Matter in the Universe (DM 2000), 27-36, hep-ph/0002297.
6. V. Sahni and A. Starobinsky, Int. J. Mod. Phys. D **9** 373 (2000).
7. P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75** 559 (2003).
8. T. Padmanabhan, Phys. Rept. **380** 235 (2003).
9. E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15** 1753 (2006).
10. L. Amendola and S. Tsujikawa, *Dark Energy*, Cambridge U. Press (2010) & (2015), and references therein.
11. L. Verde, T. Treu and A. G. Riess, Nat. Astron. **3** 891 (2019), and references therein.
12. A. G. Riess, *et al.*, Astrophys. J. **876** 85 (2019).
13. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, Astrophys. J. **811** L14 (2015); Astrophys. J. **836** 43 (2017).
14. S. Basilakos, D. Polarski and J. Solà, Phys. Rev. D **86** 043010 (2012); A. Gómez-Valent, J. Solà and S. Basilakos, JCAP **01** 004 (2015); A. Gómez-Valent and J. Solà, MNRAS **448** 2810 (2015); A. Gómez-Valent, E. Karimkhani and J. Solà, JCAP **1512** 048 (2015).
15. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, Phys. Lett. B **774** 317 (2017).
16. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, EPL **121** 39001 (2018); MNRAS **478** 4357 (2018).
17. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Astrophys. J. **886** L6 (2019).
18. M. Rezaei, M. Malekjani and J. Solà Peracaula, Phys. Rev. D **100** 023539 (2019).
19. E. D. Valentino, A. Melchiorri, Olga Mena and S. Vagnozzi, Phys. Rev. D **101** 063502 (2020).
20. E. D. Valentino, A. Melchiorri and O. Mena, Phys. Rev. D **96** 043503 (2017); E.D. Valentino, A. Melchiorri and J. Silk, Phys. Lett. B **761** 242 (2016).
21. J. Ooba, B. Ratra, and N. Sugiyama, Astrophys. Space Sci. **364** 176 (2019).
22. C.-G. Park and B. Ratra, Astrophys. J. **882** 158 (2019); Astrophys. Space Sci. **364** 82 (2019); Astrophys. J. **868** 83 (2018).
23. M. Martinelli, N. B. Hogg, S. Peirone, M. Bruni and D. Wands, MNRAS **488** 3423 (2019) ; V. Salvatelli, N. Said, M. Bruni, A. Melchiorri and D. Wands, Phys. Rev. Lett. **113** 181301 (2014).
24. A. A. Costa, X. D. Xu, B. Wang and E. Abdalla, JCAP **1701** 028 (2017).
25. Y. H. Li, J. F. Zhang and X. Zhang, Phys. Rev. D **93** 023002 (2016); Phys. Rev. D **90** 123007 (2014); Phys. Rev. D **90** 063005 (2014).
26. C. Moreno-Pulido and J. Solà Peracaula, Eur. Phys. J. C **80** 692 (2020).
27. I. L. Shapiro and J. Solà, JHEP **02** 006 (2002); Nucl. Phys. B Proc. Suppl. **127** 71 (2004).
28. J. Solà, J. Phys. A **41** 164066 (2008).
29. I. L. Shapiro and J. Solà, Phys. Lett. B **682** 105 (2009).

30. J. Solà, J. Phys. Conf. Ser. **453** 012015 (2013) [arXiv:1306.1527]; AIP Conf. Proc. **1606** 19 (2015)[arXiv:1402.7049]; J. Phys. Conf. Ser. **283** 012033 (2011) [arXiv:1102.1815].
31. J. Solà and A. Gómez-Valent, Int. J. Mod. Phys. D **24** 1541003 (2015).
32. A. Gómez Valent, *Vacuum energy in Quantum Field Theory and Cosmology* (PhD Thesis), [arXiv:1710.01978]; J. de Cruz Pérez, *Implications of Dynamical Dark Energy in the expansion of the Universe and the Structure Formation* (PhD. Thesis), [arXiv:2105.14800].
33. A. Babic, B. Guberina, R. Horvat and H. Stefancic, Phys. Rev. D **71** 124041 (2005).
34. M. Maggiore, Phys. Rev. D **83** 063514 (2011); M. Maggiore, L. Hollenstein, M. Jaccard and E. Mitsou, Phys. Lett. B **704** 102 (2011); L. Hollenstein, M. Jaccard, M. Maggiore and E. Mitsou, Phys. Rev. D **85** 124031 (2012)
35. K. Kohri and H. Matsui, JCAP **06** 006 (2017).
36. A. Ferreiro and J. Navarro-Salas, Phys. Lett. B **792** 81 (2019).
37. N. Bilic, Phys. Rev. D **83** 105003 (2011); Rom. J. Phys. **57** 793 (2012).
38. N. Bilic, S. Domazet and B. Guberina, Phys. Lett. B **707** 221 (2012).
39. S. Basilakos, N. E. Mavromatos and J. Solà, Universe **2** 14 (2016).
40. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula, Phys. Rev. D **101** 045001 (2020) ; Int. J. Mod. Phys. D **28** 1944002 (2019).
41. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula, Phys. Lett. B **803** 135342 (2020).
42. N. E. Mavromatos and J. Solà Peracaula, Eur. Phys. J. Spec. Top. **230** 2077 (2021).
43. N. E. Mavromatos, *Geometrical origins of the Universe dark sector: string-inspired torsion and anomalies as seeds for inflation and dark matter*, arXiv:2108.02152 (to appear in Phil.Trans.Roy.Soc.Lond.A).
44. N. E. Navromatos and J. Solà Peracaula, *Inflationary physics and transplanckian conjecture in the Stringy Running-Vacuum-Model: from the phantom vacuum to the true vacuum*, arXiv:2105.02659.
45. N. E. Navromatos, *Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology*, arXiv:2108.03998.
46. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge U. Press (1982).
47. L. E. Parker and D. J. Toms *Quantum Field Theory in Curved Spacetime: quantized fields and gravity*, Cambridge U. Press (2009).
48. W. Fulling, *Aspects of Quantum Field Theory in Curved Space-Time*, Cambridge U. Press, (1989).
49. V. F. Mukhanov and S. Winitzki, *Quantum Effects in Gravity*, Cambridge U. Press (2007).
50. T. S. Bunch, J. Phys. A **13** 1297 (1980).
51. L. Parker and S. A. Fulling, Phys. Rev. D **9** 341 (1974); S. A. Fulling and L. Parker, Ann. Phys. **87** 176 (1974); S. A. Fulling, L. Parker and B. L. Hu, Phys. Rev. D **10** 3905 (1974).
52. T. S. Bunch and L. Parker, Phys. Rev. D **20** 2499 (1979).
53. A. Landete, J. Navarro-Salas and F. Torrenti, Phys Rev. D **88** 061501 (2013); Phys. Rev. D **89** 044030 (2014); A. del Rio, J. Navarro-Salas and F. Torrenti, Phys. Rev. D **90** 084017 (2014).
54. J. F. Barbero G., A. Ferreiro, J. Navarro-Salas and E. J. S. Villaseñor, Phys.Rev. D **98** 025016 (2018).
55. J. Solà Peracaula, C. Moreno-Pulido and S. Cheraghchi, in preparation.
56. *Mathematica*, Wolfram Research, Inc.. URL: *https://www.wolfram.com/mathematica*.

57. T. Appelquist and J. Carazzone, Phys. Rev. D **11** 2856 (1975).
58. H. Fritzsch and J. Solà, Class. Quant. Grav. **29** 215002 (2012); Mod. Phys. Lett. A **30** 1540034 (2015); Eur. Phys. J. C **77** 193 (2017).
59. S. Basilakos, J. A. S. Lima and J. Solà, MNRAS **431** 923 (2013); E. L. D. Perico, J. A. S. Limta, S. Basilakos and J. Solà, Phys. Rev. D **88** 063531 (2013).
60. J. Solà and A. Gómez-Valent, Int. J. Mod. Phys. D **24** 151003 (2015).
61. J. Solà, Int. J. Mod. Phys. D **24** 1544027 (2015).
62. J. Solà Peracaula and H. Yu, Gen. Rel. Grav. **52** 17 (2020).
63. C. W. Misner, K. S. Thorn and J. A. Wheeler, *Gravitation*, Freeman, San Francisco (1973).

BD-ΛCDM and running vacuum models: Theoretical background and current observational status

Javier de Cruz Pérez[1,*], Joan Solà Peracaula[2], Adrià Gómez-Valent[3] and Cristian Moreno-Pulido[2]

[1] *Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*

[2] *Departament de Física Quàntica i Astrofísica, and Institute of Cosmos Sciences, Universitat de Barcelona, Av. Diagonal 647, E-08028 Barcelona, Catalonia, Spain*

[3] *Dipartamento di Fisica, and INFN Sezione di Roma Tor Vergata, via della Ricerca Scientifica 1, 00133 Roma Italy*

[*] *E-mail: decruz@fqa.ub.edu, sola@fqa.ub.edu, cristian.moreno@fqa.ub.edu, agvalent@roma2.infn.it*

We present an analysis of the Brans-Dicke cosmological model with a cosmological constant and cold dark matter (BD-ΛCDM). We find that the BD-ΛCDM is favored by the overall cosmological data (SNIa+BAO+$H(z)$+LSS+CMB) when it is compared with the standard model of cosmology. The BD-ΛCDM model can be viewed from the GR perspective as a Running Vacuum Model (RVM) with a time evolving vacuum energy density. Due to this fact and also to its time evolving effective gravitational coupling, the model can alleviate the σ_8 and the H_0 tensions at a time. We also present the results for different types of RVM's when they are tested in the light of the cosmological data and we show that a mild dynamics for the vacuum energy density can help to smooth out the aforementioned tensions, thus improving the performance of ΛCDM model.

Keywords: Cosmology, Dark Energy, General Relativity.

1. Introduction

The high quality observations performed during the last two decades, have allowed to demonstrate, with a high confidence range, that the Universe is in expansion and to be more precise in accelerated expansion.[1,2] The standard model of cosmology, the so-called ΛCDM, is based on the assumption that the observed accelerated expansion is due to the existence of a repulsive force, exerted by the Λ term, which counteracts the attractive gravitational force and pushes the clusters of galaxies at a speed that increases with the cosmic expansion. This term, which is called the cosmological constant (CC), is associated with a mysterious form of energy, usually called dark energy (DE) presumably permeating all corners of the Universe as a whole. The cosmological constant is not the only element beyond the conventional matter, namely: baryons, photons and neutrinos, demanded by the observations since it is also essential the presence of large amounts of what is commonly call, cold dark matter (CDM). It is important to remark that the concordance model is

[*]Speaker

defined in the framework of General Relativity (GR), whose field equations were found by Einstein in 1915.[3] The fact that the ΛCDM model, with so few ingredients (encoded in 6 degrees of freedom), has remained robust and unbeaten for a long time turns out to be pretty impressive. It is consistent with a large body of very accurate cosmological data, like the cosmic microwave background (CMB), the baryonic acoustic oscillations (BAO) or the large scale structure (LSS) data.

However, this does not mean at all that we really understand the primary (dynamical) cause for such an acceleration or that we are in position to propose a cosmological model explaining the speeding up of the cosmos at the level of fundamental physics, say quantum field theory (QFT) in curved spacetime, quantum gravity or string theory. For instance, if we try to explain the origin of the cosmological constant in the context of QFT we will immediately realize that we have to face the well-known old cosmological constant problem. This problem was formulated by Zel'dovich[4,5] (see also[6–9]) and basically consists in the tremendous mismatch between the computed value of the vacuum energy density, considering different contributions from QFT, and the experimental measure, being the discrepancy between both values in natural units of at least 55 orders of magnitude.

Thus, in spite of the many virtues of the ΛCDM model, the lack of an explanation for the presence of the CC, leads inevitably to an unsatisfactory theoretical picture which has motivated the search of a wide range of alternatives beyond the standard model. What is more, aside from the theoretical problems some persisting tensions with the cosmological data (being the most important ones, the tension on the local value of the H_0 parameter[10,11] and the one affecting the LSS data[12]) point out that the ΛCDM model might be performing insufficiently at the observational level.

Here we study two alternatives where the aforementioned tension could be alleviated. It has been suggested that it would help to alleviate the tensions if DE would be a dynamical quantity (see for instance the possibility of an early dark energy[13,14]), i.e. slowly evolving with the cosmic expansion. This could be achieved, for example, through scalar field models[15–20] (see[21] for an updated analysis), where the energy density associated to the Λ term is replaced by the energy density of the scalar field, which varies throughout the cosmic history. In this paper, instead of considering a scalar field model we focus on another promising possibility, namely that vacuum energy density $\rho_{\rm vac} = \Lambda/8\pi G_N$ might be a time evolving quantity whose dynamics is triggered by the quantum effects coming from the matter fields. We call these types of models Running Vacuum Models (RVM's). Phenomenologically, these models, have recently been carefully confronted against the wealth of the cosmological data, obtaining a significant success.[22–27]

On the other hand there exists the possibility to consider a completely different approach, where a theoretical framework, different from GR, is assumed. In this regard, Brans & Dicke[28] proposed the first historical attempt to extend Einstein's GR by promoting the Newtonian coupling G_N into a variable one in the cosmic time $G(t)$. Gravity is then not only mediated by the metric but also by an scalar field

denoted by ψ. The theory contains an extra *d.o.f.* encoded in ψ and a dimensionless parameter $\omega_{\rm BD}$. In order to recover GR a necessary (though not sufficient) condition is to demand large values of $\omega_{\rm BD}$. We consider the presence of the CC and all the species considered in the standard model, so we call the model studied in the context of Brans-Dicke (BD) gravity, BD-ΛCDM.

In this work we present a summary of the main results recently obtained by confronting the overall cosmological observations with the BD-ΛCDM model and related RVM's and we show how they deal with the aforementioned tensions. See the original references[?, 27, 29, 30] for more details.

2. Background equations for the BD-ΛCDM model

We will consider the original BD-action extended with a cosmological constant density term as it is essential to mimic the conventional ΛCDM model based on GR and reproduce its main successes. We call this the 'BD-ΛCDM model', i.e. the version of the ΛCDM within the BD paradigm. The BD action reads, written in the Jordan frame, as follows [a]:

$$S_{\rm BD} = \int d^4x\sqrt{-g}\left[\frac{1}{16\pi}\left(R\psi - \frac{\omega_{\rm BD}}{\psi}g^{\mu\nu}\partial_\nu\psi\partial_\mu\psi\right) - \rho^0_{\rm vac}\right] + S_m\,. \tag{1}$$

For the sake of convenience we denote the constant vacuum energy density as $\rho^0_{\rm vac}$. The (dimensionless) factor in front of the kinetic term of ψ, i.e. $\omega_{\rm BD}$, will be referred to as the BD-parameter and we consider the canonical option, $\omega_{\rm BD}$ =const. The last term of (1) stands for the matter action S_m, which is constructed from the Lagrangian density of the matter fields. There is no potential for the BD-field ψ in the original BD-theory, but we admit the presence of a CC term associated to $\rho^0_{\rm vac}$.

The modified field equations (with respect to GR) can be obtained after performing variation of the action (1) with respect to both the metric and the scalar field ψ. While the first variation yields

$$\psi\, G_{\mu\nu} + \left(\Box\psi + \frac{\omega_{\rm BD}}{2\psi}\left(\nabla\psi\right)^2\right) g_{\mu\nu} - \nabla_\mu\nabla_\nu\psi - \frac{\omega_{\rm BD}}{\psi}\nabla_\mu\psi\nabla_\nu\psi = 8\pi\left(T_{\mu\nu} - g_{\mu\nu}\rho^0_{\rm vac}\right)\,, \tag{2}$$

the second variation gives the wave equation for ψ

$$\Box\psi = \frac{8\pi}{2\omega_{\rm BD}+3}\left(T - 4\rho^0_{\rm vac}\right)\,. \tag{3}$$

We have used the definition $(\nabla\psi)^2 \equiv g^{\mu\nu}\nabla_\mu\psi\nabla_\nu\psi$. In the first equation, $G_{\mu\nu} = R_{\mu\nu} - (1/2)Rg_{\mu\nu}$ is the Einstein tensor, and on its *r.h.s.* $T_{\mu\nu} = -(2/\sqrt{-g})\delta S_m/\delta g^{\mu\nu}$ is the energy-momentum tensor of matter, being $T \equiv T^\mu_\mu$ its trace. The total energy-momentum tensor is the sum of the matter and vacuum contributions and it takes

[a] We use natural units, $\hbar = c = 1$ and $G_N = 1/m^2_{\rm Pl}$, where $m_{\rm Pl} \simeq 1.22\times10^{19}$ GeV is the Planck mass. As for the geometrical quantities, we adopt the $(+,+,+)$ convention of the popular classification by Misner, Thorn and Wheeler.[31]

the perfect fluid form:

$$\tilde{T}_{\mu\nu} = T_{\mu\nu} - \rho^0_{\rm vac} g_{\mu\nu} = p\, g_{\mu\nu} + (\rho + p) u_\mu u_\nu \,, \tag{4}$$

with $\rho \equiv \rho_m + \rho_\gamma + \rho_{\rm ncdm} + \rho^0_{\rm vac}$ and $p \equiv p_m + p_\gamma + p_{\rm ncdm} + p^0_{\rm vac}$. The matter part $\rho_m \equiv \rho_b + \rho_{\rm cdm}$, contains the pressureless contribution from baryons and cold dark matter and photons are of course relativistic, so $p_\gamma = \rho_\gamma/3$. The functions $\rho_{\rm ncdm}$ and $p_{\rm ncdm}$ ('ncdm' means non-CDM) include the effect of massive and massless neutrinos, and therefore must be computed numerically. As in GR, we have considered a constant vacuum energy density, $\rho^0_{\rm vac}$, in the BD-action (1), being its equation of state $p^0_{\rm vac} = -\rho^0_{\rm vac}$ the usual one.

Let us write down the field equations in the flat (to know more about the nonflat models see the references[32–44]) FLRW metric, $ds^2 = -dt^2 + a^2 \delta_{ij} dx^i dx^j$. Using the previous definitions for the total energy density ρ and pressure p we get the two independent equations:

$$3H^2 + 3H\frac{\dot{\psi}}{\psi} - \frac{\omega_{\rm BD}}{2}\left(\frac{\dot{\psi}}{\psi}\right)^2 = \frac{8\pi}{\psi}\rho \tag{5}$$

and

$$2\dot{H} + 3H^2 + \frac{\ddot{\psi}}{\psi} + 2H\frac{\dot{\psi}}{\psi} + \frac{\omega_{\rm BD}}{2}\left(\frac{\dot{\psi}}{\psi}\right)^2 = -\frac{8\pi}{\psi}p \,, \tag{6}$$

whereas from (3) we obtain

$$\ddot{\psi} + 3H\dot{\psi} = \frac{8\pi}{2\omega_{\rm BD} + 3}(\rho - 3p) \,. \tag{7}$$

Here dots stand for derivatives with respect to the cosmic time and $H = \dot{a}/a$ is the Hubble rate. In the limit $\psi \to 1/G_N$ and $\omega_{\rm BD} \to \infty$ we recover the standard equations. The connection between GR and the $\omega_{\rm BD} \to \infty$ limit is sometimes not as straightforward as one might naively think .[45,46]

Matter and the BD-field are not in interaction, as a result, the local conservation law adopts the usual form:

$$\dot{\rho} + 3H(\rho + p) = \sum_N \left[\dot{\rho}_N + 3H(\rho_N + p_N)\right] = 0 \,, \tag{8}$$

where the sum is over all components, i.e. baryons, dark matter, neutrinos, photons and vacuum. We assume that all of the components are separately conserved in the main periods of the cosmic evolution. For convenience, we will use a dimensionless BD-field, φ, and the inverse of the BD-parameter, according to :

$$\varphi(t) \equiv G_N \psi(t) \,, \qquad\qquad \epsilon_{\rm BD} \equiv \frac{1}{\omega_{\rm BD}} \,. \tag{9}$$

As stated before G_N gives the local value of the gravitational coupling. Note that a nonvanishing value of $\epsilon_{\rm BD}$ entails a deviation from GR. Being $\varphi(t)$ a dimensionless quantity, we can recover GR by enforcing the simultaneous limits $\epsilon_{\rm BD} \to 0$ *and* $\varphi \to 1$.

Baseline

Parameter	ΛCDM	type I RRVM	type I RRVM$_{\rm thr.}$	type II RRVM	BD-ΛCDM
H_0(km/s/Mpc)	$68.37^{+0.38}_{-0.41}$	$68.17^{+0.50}_{-0.48}$	$67.63^{+0.42}_{-0.43}$	$69.02^{+1.16}_{-1.21}$	$69.30^{+1.38}_{-1.33}$
ω_b	$0.02230^{+0.00019}_{-0.00018}$	$0.02239^{+0.00023}_{-0.00024}$	$0.02231^{+0.00020}_{-0.00019}$	$0.02245^{+0.00025}_{-0.00027}$	0.02248 ± 0.00025
$\omega_{\rm cdm}$	$0.11725^{+0.00094}_{-0.00084}$	$0.11731^{+0.00092}_{-0.00087}$	$0.12461^{+0.00201}_{-0.00210}$	$0.11653^{+0.00158}_{-0.00160}$	$0.11629^{+0.00148}_{-0.00151}$
$\nu_{\rm eff}$	-	$0.00024^{+0.00039}_{-0.00040}$	$0.02369^{+0.00625}_{-0.00563}$	0.00029 ± 0.00047	-
$\epsilon_{\rm BD}$	-	-	-	-	$-0.00109 \pm^{+0.00135}_{-0.00141}$
$\varphi_{\rm ini}$	-	-	-	$0.980^{+0.031}_{-0.027}$	$0.972^{+0.030}_{-0.037}$
φ_0	-	-	-	$0.973^{+0.036}_{-0.033}$	$0.963^{+0.036}_{-0.041}$
$\tau_{\rm reio}$	$0.049^{+0.008}_{-0.007}$	$0.051^{+0.008}_{-0.009}$	$0.058^{+0.007}_{-0.009}$	0.051 ± 0.008	0.051 ± 0.008
n_s	$0.9698^{+0.0039}_{-0.0036}$	$0.9716^{+0.0044}_{-0.0047}$	0.9703 ± 0.038	$0.9762^{+0.0081}_{-0.0091}$	
σ_8	0.796 ± 0.007	$0.789^{+0.013}_{-0.014}$	$0.768^{+0.010}_{-0.009}$	$0.791^{+0.013}_{-0.012}$	$0.790^{+0.013}_{-0.012}$
S_8	0.796 ± 0.011	$0.791^{+0.014}_{-0.013}$	$0.797^{+0.012}_{-0.011}$	$0.781^{+0.021}_{-0.020}$	$0.777^{+0.021}_{-0.022}$
r_s (Mpc)	$147.90^{+0.30}_{-0.31}$	$147.99^{+0.35}_{-0.36}$	147.81 ± 0.30	$146.30^{+2.39}_{-2.30}$	$145.72^{+2.44}_{-2.90}$
$\chi^2_{\rm min}$	2290.20	2289.72	2272.44	2288.74	2289.40
ΔDIC	-	-2.70	+13.82	-4.59	-3.53

Note: The mean values and 68.3% confidence limits for the models under study using our Baseline dataset, which is comporsed by: the full Pantheon likelihood,[47] 13 BAO data points,[48–53] 31 data points on $H(z_i)$, at different redshifts,[54–61] 14 points on the observable $f(z_i)\sigma_8(z_i)$[50,52,62–66] and finally the full Planck likelihood for the CMB data.[10] We display the fitting values for the usual parameters, to wit: H_0, the reduced density parameter for baryons ($w_b = \Omega_b^0 h^2$) and also for CDM ($w_{\rm cdm} = \Omega_{\rm cdm}^0 h^2$), being $\Omega_i^0 = 8\pi G_N \rho_i^0/3H_0^2$ and h the reduced Hubble constant, the reionization optical depth $\tau_{\rm reio}$, the spectral index n_s and the current matter density rms fluctuations within spheres of radius $8h^{-1}$ Mpc, i.e. σ_8. We include also a couple of derived parameters, namely: the sound horizon at the baryon drag epoch r_s and $S_8 \equiv \sigma_8\sqrt{\Omega_m^0/0.3}$. For the RRVM's we provide the value of $\nu_{\rm eff}$, and for the type II and BD-ΛCDM[27] we also report the initial and current values of φ, $\varphi_{\rm ini}$ and φ_0, respectively. The parameter $\epsilon_{\rm BD} \equiv 1/\omega_{\rm BD}$ (inverse of the Brans-Dicke parameter[28]) controls the dynamics of the scalar field.[29,30] We finally provide the corresponding values of $\chi^2_{\rm min}$ and ΔDIC.

3. Connection of the BD-ΛCDM model with the Running Vacuum Model

So far, no analytical solutions to the system (5)-(7) have been found, for this reason our actual analysis proceeds numerically. However, it is possible to search for approximate solutions valid in the different epochs of cosmic history, which can provide a qualitative understanding of the numerical results obtained. Actually, a first attempt in this direction trying to show that BD-ΛCDM can mimic the Running Vacuum Model (RVM) was done in,[67,68] we refer the reader to these references for details. See also.[69,70] We are interested in looking for solutions in the Matter Dominated Epoch (MDE) in the form of a power-law ansatz in which the BD-field φ evolves very slowly:

$$\varphi(a) = \varphi_0\, a^{-\epsilon} \qquad (|\epsilon| \ll 1)\,. \tag{10}$$

The ϵ parameter must be a very small parameter in absolute value since $G(a) \equiv G(\varphi(a))$ cannot depart too much from G_N and $\varphi_0 \equiv \varphi(z=0)$. For $\epsilon > 0$, the effective coupling increases with the expansion and hence is asymptotically free

since $G(a)$ is smaller in the past, which is the epoch when the Hubble rate (with natural dimension of energy) is bigger. For $\epsilon < 0$, is the other way around and $G(a)$ decreases with the expansion.

Plugging the power-law ansatz into the cosmological equations we end up with the following pair of Friedmann-like equations to $\mathcal{O}(\epsilon)$:[67,68]

$$H^2 = \frac{8\pi G}{3}\left(\rho_m^0 a^{-3+\epsilon} + \rho_{\rm DE}(H)\right) \tag{11}$$

and for the acceleration equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho_m^0 a^{-3+\epsilon} + \rho_{\rm DE}(H) + 3p_{\rm vac}^0\right), \tag{12}$$

with $G = G_N/\varphi_0$. The first of the above equations can be understood as an effective Friedmann's equation with time-evolving cosmological term, in which the DE appears as if it were a dynamical quantity:

$$\rho_{\rm DE}(H) = \rho_{\rm vac}^0 + \frac{3\,\bar{\nu}_{\rm eff}}{8\pi G} H^2\,. \tag{13}$$

Here

$$\bar{\nu}_{\rm eff} \equiv \epsilon\left(1 + \frac{1}{6}\,\omega_{\rm BD}\epsilon\right) \tag{14}$$

is the coefficient controlling the dynamical character of the effective dark energy (13). The structure of this dynamical dark energy (DDE) is reminiscent of the Running Vacuum Model (RVM), see[71–73] and references therein. Notice from (11) that, to $\mathcal{O}(\epsilon)$:

$$\Omega_m + \Omega_\Lambda = 1 - \bar{\nu}_{\rm eff}\,, \tag{15}$$

so, due to the presence of $\bar{\nu}_{\rm eff}$, we find a slight deviation from the usual sum rule of GR. Only in the case $\epsilon = 0$, we have $\bar{\nu}_{\rm eff} = 0$ and then we recover the usual cosmic sum rule. As stated, the parameter $\bar{\nu}_{\rm eff}$ becomes associated to the dynamics of the dark energy. We shall see more about the most important features of the RVM in the next section. All in all, it turns out that the BD-RVM (we may call it in this way for convenience) can cure the H_0-tension due to the evolution of the effective gravitational $G_{\rm eff}$ and, additionally, by mimicking the RVM it can also alleviate the well-known σ_8-tension thanks to the role played by $\bar{\nu}_{\rm eff}$. Going a little further in this effective picture from (12) we can define the EoS parameter for the effective DDE:

$$w_{\rm eff}(z) = \frac{p_{\rm vac}^0}{\rho_{\rm DE}(H)} \simeq -1 + \frac{3\bar{\nu}_{\rm eff}}{8\pi G\rho_{\rm vac}^0}\,H^2(z) = -1 + \frac{\bar{\nu}_{\rm eff}}{\Omega_\Lambda}\,\frac{H^2(z)}{H_0^2}\,, \tag{16}$$

where use has been made of (13). As it is clear from the above equation the BD-RVM, unlike the original RVM, does not describe a dark energy of pure vacuum form but a DE whose EoS departs, in a mild way, from the pure vacuum. Actually, for $\epsilon > 0$ ($\epsilon < 0$) we have $\bar{\nu}_{\rm eff} > 0$ ($\bar{\nu}_{\rm eff} < 0$) and the effective DDE behaves quintessence (phantom)-like.

Baseline + H_0

Parameter	ΛCDM	type I RRVM	type I RRVM$_{\rm thr.}$	type II RRVM	BD-ΛCDM
H_0 (km/s/Mpc)	$68.75^{+0.41}_{-0.36}$	$68.77^{+0.49}_{-0.48}$	$68.14^{+0.43}_{-0.41}$	$70.93^{+0.93}_{-0.87}$	$71.23^{+1.01}_{-1.02}$
ω_b	$0.02240^{+0.00019}_{-0.00021}$	$0.02238^{+0.00021}_{-0.00023}$	$0.02243^{+0.00019}_{-0.00018}$	$0.02269^{+0.00025}_{-0.00024}$	$0.02267^{+0.00026}_{-0.00023}$
$\omega_{\rm cdm}$	$0.11658^{+0.00080}_{-0.00083}$	$0.11661^{+0.00084}_{-0.00085}$	$0.12299^{+0.00197}_{-0.00203}$	$0.11602^{+0.00162}_{-0.00163}$	$0.11601^{+0.00161}_{-0.00157}$
$\nu_{\rm eff}$	-	$-0.00005^{+0.00040}_{-0.00038}$	$0.02089^{+0.00553}_{-0.00593}$	$0.00038^{+0.00041}_{-0.00044}$	-
$\epsilon_{\rm BD}$	-	-	-	-	$-0.00130\pm^{+0.00136}_{-0.00140}$
$\varphi_{\rm ini}$	-	-	-	$0.938^{+0.018}_{-0.024}$	$0.928^{+0.024}_{-0.026}$
φ_0	-	-	-	$0.930^{+0.022}_{-0.029}$	$0.919^{+0.028}_{-0.033}$
$\tau_{\rm reio}$	$0.050^{+0.008}_{-0.007}$	$0.049^{+0.009}_{-0.008}$	$0.058^{+0.008}_{-0.009}$	0.052 ± 0.008	0.052 ± 0.008
n_s	$0.9718^{+0.0035}_{-0.0038}$	0.9714 ± 0.0046	$0.9723^{+0.0040}_{-0.0039}$	$0.9868^{+0.0072}_{-0.0074}$	$0.9859^{+0.0073}_{-0.0072}$
σ_8	0.794 ± 0.007	0.795 ± 0.013	0.770 ± 0.010	$0.794^{+0.013}_{-0.012}$	$0.792^{+0.013}_{-0.012}$
S_8	$0.788^{+0.010}_{-0.011}$	0.789 ± 0.013	0.789 ± 0.011	$0.761^{+0.018}_{-0.017}$	$0.758^{+0.019}_{-0.018}$
r_s (Mpc)	$147.97^{+0.29}_{-0.31}$	$147.94^{+0.35}_{-0.36}$	$147.88^{+0.33}_{-0.29}$	$143.00^{+1.54}_{-1.96}$	$142.24^{+1.99}_{-2.12}$
$\chi^2_{\rm min}$	2302.14	2301.90	2288.82	2296.38	2295.36
ΔDIC	-	-2.36	+10.88	+5.52	+6.25

Note: Same as in Table 1, but also considering the prior on $H_0 = (73.5 \pm 1.4)$ km/s/Mpc from SH0ES.[11]

4. Background equations for the RVM's

In the following we are going to present the theoretical framework where the previously mentioned RVM's are placed[71–73] and references therein. We consider two types of dynamical vacuum energy (DVE) scenarios. In type I scenario the vacuum is in interaction with matter, in contrast, in type II matter is conserved at the expense of an exchange between the vacuum and a slowly evolving gravitational coupling $G(H)$. In both cases, the combined cosmological 'running' of these quantities insures the accomplishment of the Bianchi identity (and the local conservation law). Let us therefore consider a generic cosmological framework described by the spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric. The vacuum energy density in the RVM can be expressed as :[71,72]

$$\rho_{\rm vac}(H) = \frac{3}{8\pi G_N}\left(c_0 + \nu H^2 + \tilde{\nu}\dot{H}\right) + \mathcal{O}(H^4)\,. \tag{17}$$

The $\mathcal{O}(H^4)$ terms play no role in the post-inflationary epoch so they can be neglected. The expression (17) can be motivated from the explicit QFT calculations on a FLRW background.[74] The value of the additive constant c_0 is fixed by the boundary condition $\rho_{\rm vac}(H_0) = \rho^0_{\rm vac}$. The two dynamical components H^2 and $\dot{H}$ are dimensionally homogeneous and in principle independent. The dimensionless coefficients ν and $\tilde{\nu}$ encode the dynamics of the vacuum at low energy and we naturally expect for both of them $|\nu, \tilde{\nu}| \ll 1$. An estimate of ν in QFT indicates that it is of order 10^{-3} at most.[75] In the calculations performed in[74] these coefficients are expected to be of order $\sim M_X^2/m_{\rm Pl}^2 \ll 1$, being M_X of order of a typical Grand Unified Theory (GUT) scale. Taking into account the multiplicity of particles in a GUT, the theoretical estimate on ν could be much larger but still subject to a value below 1.

We are interested in a particular form for the vacuum energy density obtained by imposing $\tilde{\nu} = \nu/2$. Consequently

$$\rho_{\rm vac}(H) = 3/(8\pi G_N)\left[c_0 + \nu\left(H^2 + \frac{1}{2}\dot{H}\right)\right]. \tag{18}$$

We will call this form of the vacuum energy density the 'RRVM' since it realizes the generic RVM density (17) through the Ricci scalar $\mathcal{R} = 12H^2 + 6\dot{H}$, namely

$$\rho_{\rm vac}(H) = \frac{3}{8\pi G_N}\left(c_0 + \frac{\nu}{12}\mathcal{R}\right) \equiv \rho_{\rm vac}(\mathcal{R})\,. \tag{19}$$

Due to the fact that the condition $\mathcal{R}/H^2 \ll 1$ is fulfilled in the early epochs of the cosmological evolution we do not generate any conflict with the BBN nor with any other feature of the modern universe. Of course, early on the RVM has its own mechanism for inflation see[71, 72, 76, 77] for more details.

4.1. *Type I RRVM*

Taking into account, the dynamical character presented for the vacuum energy density, the Friedmann and the acceleration equations for the different species involved, read

$$3H^2 = 8\pi G_N\left(\rho_m + \rho_{\rm ncdm} + \rho_\gamma + \rho_{\rm vac}(H)\right), \tag{20}$$

$$3H^2 + 2\dot{H} = -8\pi G_N\left(p_{\rm ncdm} + p_\gamma + p_{\rm vac}(H)\right). \tag{21}$$

As for the BD-model, we define the total nonrelativistic matter density as the sum of the CDM component and the baryonic one: $\rho_m = \rho_{\rm cdm} + \rho_b$. Therefore the total (relativistic and nonrelativistic) matter density is $\rho_t = \rho_m + \rho_\gamma + \rho_{\rm ncdm}$. In a similar way, the total matter pressure reads $p_t = p_{\rm ncdm} + p_\gamma$ (with $p_\gamma = (1/3)\rho_\gamma$). Making use of the functions employed by the system solver `CLASS` [78] we distinguish between the contributions of the nonrelativistic neutrinos $\rho_h = \rho_{\rm ncdm} - 3p_{\rm ncdm}$ and the one from the relativistic neutrinos $\rho_\nu = 3p_{\rm ncdm}$. This separation allows to compute $\mathcal{R}/12 = H^2 + (1/2)\dot{H}$ appearing in (19) in terms of the energy densities and pressures using (20) and (21):

$$\mathcal{R} = 8\pi G_N\left(\rho_m + 4\rho_{\rm vac} + \rho_h\right). \tag{22}$$

We can safely neglect the contribution of neutrinos from (22) since it remains well below the contribution of CDM and baryons throughout the whole cosmic history. This fact allows us to solve for the vacuum density as a function of the scale factor a as follows:

$$\rho_{\rm vac}(a) = \rho^0_{\rm vac} + \frac{\nu}{4(1-\nu)}(\rho_m(a) - \rho^0_m)\,, \tag{23}$$

where ' 0' (used as subscript or superscript) always refers to current quantities. For $a = 1$ we confirm the correct normalization: $\rho_{\rm vac}(a = 1) = \rho^0_{\rm vac}$. Matter does not follow the standard dilution law, remember that it is in interaction with vacuum,

therefore $\rho_m(a)$ is not just $\sim a^{-3}$. The local conservation law for CDM and vacuum can be expressed as follows:

$$\dot{\rho}_{\rm cdm} + 3H\rho_{\rm cdm} = -\dot{\rho}_{\rm vac}\,. \tag{24}$$

We assume that baryons are self conserved, which implies $\dot{\rho}_b + 3H\rho_b = 0$, and as a consequence the total matter contribution (ρ_m) satisfies the same local conservation law (24) as CDM: $\dot{\rho}_m + 3H\rho_m = -\dot{\rho}_{\rm vac}$. Using it together with (23) we find

$$\dot{\rho}_m + 3H\xi\rho_m = 0 \tag{25}$$

where, for convenience, we have defined

$$\xi \equiv \frac{1-\nu}{1-\frac{3}{4}\nu}. \tag{26}$$

We encode the deviations with respect to the standard model in terms of the effective parameter $\nu_{\rm eff} \equiv \nu/4$:

$$\xi = 1 - \nu_{\rm eff} + \mathcal{O}\left(\nu_{\rm eff}^2\right)\,. \tag{27}$$

Having reached this point it is straightforward to find the expression for the matter energy densities:

$$\rho_m(a) = \rho_m^0 a^{-3\xi}\,, \quad \rho_{\rm cdm}(a) = \rho_m^0 a^{-3\xi} - \rho_b^0 a^{-3}\,. \tag{28}$$

As expected, by setting $\xi = 1$ ($\nu_{\rm eff} = 0$) we recover the ΛCDM expressions. The small departure is precisely what gives allowance for a mild dynamical vacuum evolution:

$$\rho_{\rm vac}(a) = \rho_{\rm vac}^0 + \left(\frac{1}{\xi} - 1\right)\rho_m^0\left(a^{-3\xi} - 1\right)\,. \tag{29}$$

The vacuum becomes rigid if $\xi = 1$ ($\nu_{\rm eff} = 0$).

4.2. *Type II RRVM*

Unlike the first kind of models presented, for type II models matter is conserved. However, vacuum can still evolve provided the gravitational coupling also evolves with the expansion: $G = G(H)$. Let us define an auxiliary variable $\varphi = G_N/G$ – in the manner of a Brans-Dicke, without being really so. Notice that $\varphi \neq 1$ in the cosmological domain, but remains very close to it, see Tables 1 and 2. The modified Friedmann's equation for type-II model can be written as

$$3H^2 = \frac{8\pi G_N}{\varphi}\left[\rho_t + C_0 + \frac{3\nu}{16\pi G_N}(2H^2 + \dot{H})\right]\,, \tag{30}$$

with $C_0 = 3c_0/(8\pi G_N)$. The link between the dynamics of φ and that of $\rho_{\rm vac}$ is given by the Bianchi identity:

$$\frac{\dot{\varphi}}{\varphi} = \frac{\dot{\rho}_{\rm vac}}{\rho_t + \rho_{\rm vac}}\,, \tag{31}$$

where ρ_t is as before the total matter energy density and $\rho_{\rm vac}$ adopts exactly the same form as in (19). In the absence of an analytical expression for the vacuum energy density, we can show its approximate behaviour close to the present time by keeping only the terms linear in $\nu_{\rm eff}$ (recall that $|\nu_{\rm eff}| \ll 1$):

$$\rho_{\rm vac}(a) = C_0(1 + 4\nu_{\rm eff}) + \nu_{\rm eff}\rho_m^0 a^{-3} + \mathcal{O}(\nu_{\rm eff}^2)\,. \tag{32}$$

Again, for $\nu_{\rm eff} = 0$ the vacuum energy density is constant, but otherwise it shows a moderate dynamics of $\mathcal{O}(\nu_{\rm eff})$ as in the type I case (29). One can also show that $\rho_{\rm vac}(a) \ll \rho_r(a) = \rho_r^0 a^{-4}$ for $a \ll 1$ and therefore the $\rho_{\rm vac}$ for the type II model does not perturb the normal thermal history (as in the type I model). Regarding the auxiliary variable φ, in the current epoch it exhibits a very mild evolution, almost logarithmic.

4.3. *Threshold redshift scenario for type I models*

One possibility that has been explored in the literature is to consider that vacuum is a dynamical quantity only close to the present time and that remains constant for the rest of the cosmic history (see e.g.[79,80]). So, in this scenario, we keep deactivated the interaction between the vacuum energy density and the CDM until the late universe when dark energy becomes apparent. We denote the threshold value of the scale factor when the activation takes places by a_*. According to this scenario the vacuum energy density was constant prior to $a = a_*$ and it just started to evolve for $a > a_*$. It is important to remark that while $\rho_{\rm vac}$ is a continuous function, its derivative is not. We mimic such situation through a Heaviside step function $\Theta(a - a_*)$. Therefore, we assume that in the range $a < a_*$ (hence for $z > z_*$) we have

$$\begin{aligned} &\rho_{\rm cdm}(a) = \rho_{\rm cdm}(a_*)\left(\frac{a}{a_*}\right)^{-3}, \\ &\rho_{\rm vac}(a) = \rho_{\rm vac}(a_*) = \text{const.} \qquad\qquad (a < a_*)\,, \end{aligned} \tag{33}$$

where $\rho_{\rm cdm}(a_*)$ and $\rho_{\rm vac}(a_*)$ are computed from (28) and (29), respectively. In the complementary range, instead, i.e. for $a > a_*$ ($0 < z < z_*$) near our time, the original equations (28) and (29) are the ones considered.

The threshold procedure is only applied within type I models with the main purpose of preserving the standard evolution law for the matter energy density when the redshift is sufficiently high. In fact, the threshold redshift value does not need to be very high and when it is fixed by optimization it turns out to be of order $z_* \simeq 1$. An important consequence of such threshold is that the cosmological physics during the CMB epoch is exactly as in the ΛCDM. On the contrary for type II models there is still some evolution (very mild though) of $\rho_{\rm vac}$ at the CMB epoch, but the matter density follows the same law as in the standard model.

5. Cosmological perturbations

In order to perform a complete analysis of the different models under study we need to consider the evolution of the perturbed cosmological quantities throughout the cosmic history. We refer the reader to the references[23, 25, 30, 81] to know for the details of the perturbation equations since here we just display some basic equations. We consider the perturbed, spatially flat, FLRW metric $ds^2 = a^2(\eta)[-d\eta^2 + (\delta_{ij} + h_{ij})dx^i dx^j]$, in which h_{ij} represents the metric fluctuations which are coupled to the matter density perturbations $\delta_m = \delta\rho_m/\rho_m$ and $d\eta = dt/a$ is the conformal time. In the case of the BD model, at deep subhorizon scales, the differential equation for the matter density contrast is:

$$\delta_m'' + \mathcal{H}\delta_m' - 4\pi G_{\rm eff}(\bar{\varphi})a^2\bar{\rho}_m\delta_m = 0, \tag{34}$$

where here $()' \equiv d()/d\eta$ and $\mathcal{H} = aH$. In this section when we add a bar over a quantity we indicate that it is a background quantity. We have employed the definition:

$$G_{\rm eff} = \frac{G_N}{\bar{\varphi}}\left(\frac{4+2\omega_{\rm BD}}{3+2\omega_{\rm BD}}\right), \tag{35}$$

which represents the effective coupling that modifies the Poisson term of the perturbed equation with respect to the standard model.

In the case of the type I RRVM since baryons do not interact with the time evolving vacuum energy density their perturbed conservation equations are not directly affected. On the other hand the equation for the CDM takes the following form:

$$\delta_{\rm cdm}' + \frac{h'}{2} - \frac{\rho_{\rm vac}'}{\rho_{\rm cdm}}\delta_{\rm cdm} = 0, \tag{36}$$

which is obviously affected by the dynamics of the vacuum in a nontrivial way. We will present the details of the perturbation equations for the type II RRVM elsewhere.

6. Data and methodology

We fit the BD-ΛCDM, the different RVM's together with the standard model to the wealth of the cosmological data compiled from distant type Ia supernovae (SNIa), baryonic acoustic oscillations (BAO), different values of the Hubble function $H(z_i)$, the large scale structure (LSS) formation data embodied in the $f(z_i)\sigma_8(z_i)$ observable and the CMB data from the Planck satellite, see the references in the caption of Table 1. In order to compare the theoretical predictions with the available cosmological data we define the total χ^2 function as:

$$\chi^2_{\rm tot} = \chi^2_{\rm SNIa} + \chi^2_{\rm BAO} + \chi^2_H + \chi^2_{f\sigma_8} + \chi^2_{\rm CMB}. \tag{37}$$

The above terms are defined in the standard way from the data including the corresponding covariance matrices. In particular, the χ^2_H part may contain or not the

H_0 value measured by[11] depending on the setup indicated in the tables. To obtain the posterior distributions and the corresponding constraints for the dataset described above we have run the Monte Carlo sampler `MontePython`[82] together with the Einstein-Boltzmann system solver `CLASS`.[78] The latter has been properly modified in order to implement the background and the linear perturbation equations for the different models.

7. Discussions and conclusions

The main fitting results of our analysis are displayed in Tables 1 and 2 and in Figure 1. In the tables we compare, for two different datasets denoted as "Baseline" and "Baseline+H_0" (see the corresponding captions for more details), the concordance ΛCDM model with different RRVM's and with the BD-ΛCDM model.

In order to make a fair comparison between models with a different numbers of free parameters we employ the Deviance Information Criterion (DIC):[83] $\Delta\mathrm{DIC} = \mathrm{DIC}_{\Lambda\mathrm{CDM}} - \mathrm{DIC}_{\mathrm{X}}$, representing X one of the nonstandard models under study. The DIC is defined as

$$\mathrm{DIC} = \chi^2(\overline{\theta}) + 2p_D\,. \tag{38}$$

Here $p_D = \overline{\chi^2} - \chi^2(\overline{\theta})$ is the effective number of parameters of the model, and $\overline{\chi^2}$ and $\overline{\theta}$ the mean of the overall χ^2 distribution and the parameters, respectively. As it can be seen from the definition of ΔDIC, if $\Delta\mathrm{DIC} < 0$, means that the ΛCDM fits better the cosmological data, whereas for $\Delta\mathrm{DIC} > 0$ is the other way around.

Taking a look at the first table, it can be appreciated that a mild dynamics of the vacuum is very much welcome, specially in the case where the time evolution of ρ_{vac} is activated close to the moment when the contribution of this quantity becomes relevant, namely at $z \simeq 1$. It is in this particular scenario where the impact on the description of the cosmological data string SNIa+BAO+$H(z)$+LSS+CMB becomes extraordinary significant on statistical terms, since $\Delta\mathrm{AIC} > +10$ which means that a very strong evidence, in favor of this model, is found, when it is compared with the ΛCDM model. At very high redshift the physics of the type I RRVM with a threshold remain basically unaltered with respect to the ΛCDM, however, the activation of the dynamics of the vacuum at $z_* \simeq 1$ allows to suppress an exceedingly amount of structure in the universe, thus leading to a better description of the $f(z)\sigma_8(z)$ data set. On the contrary, non of the other models beyond the standard one is preferred by the cosmological data, and the DIC indicates moderate evidence against them ($-5 < \mathrm{DIC} < -2$). We would like to remark that even if the RRVM's under study are not favoured by the data, in all cases the values of $S_8 \equiv \sigma_8(0)\sqrt{\Omega_m^0/0.3}$ (or the analogous observable $\tilde{S}_8 \equiv S_8/\sqrt{\varphi_0}$ for the type II RRVM and the BD-ΛCDM models) remain compatible with recent weak lensing and galaxy clustering measurements,[84] hence smoothing out the σ_8-tension. Just by including the H_0 prior in

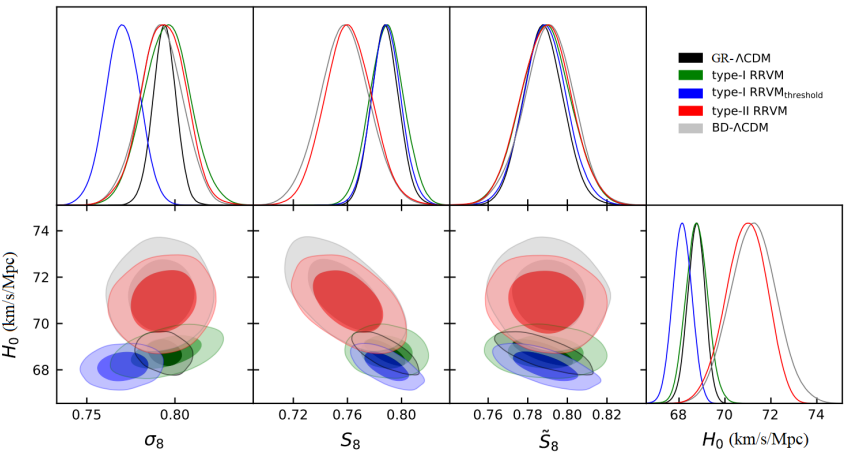


Fig. 1. The contour plots at 1σ and 2σ confidence level in the H_0-$\sigma_8, S_8, \tilde{S}_8$ planes and the corresponding one-dimensional posteriors for the GR- and BD-ΛCDM and the RRVM's obtained from the fitting analyses with our Baseline+H_0 data set. The type II model manifestly alleviates the H_0 tension without spoiling the σ_8 one, whereas the type I model with threshold redshift $z_* \simeq 1$ can fully solve the latter but cannot address the former.

the data set the results change in a very significant way as it can be observed from Table 2. The ΛCDM model, as well as, the two versions of the type I RRVM model are not able to accommodate high values of the H_0 parameter, hence, all of them show their inability to alleviate the H_0-tension. However, the DIC still decides very strongly in the case of the type I RRVM with threshold. On the other hand, the type II RRVM and the BD-ΛCDM model seem to have no problem to fit the H_0 prior since in both cases the fitting value for this parameter is $\simeq 71$km/s/Mpc. To fully appreciate how the cosmological models deal with the H_0 and the σ_8 tensions simultaneously, we have included Figure 1, where we have depicted the contour plots for the different models in the H_0-$\sigma_8, S_8, \tilde{S}_8$ planes of the parameter space. As mentioned before, the RRVM's are able to keep the values of the different parameters related with the structure formation $(\sigma_8, S_8, \tilde{S}_8)$ at an intermediate value between Planck measurements[10] and cosmic shear data.[84] Remarkably enough, in the case of the type II RRVM and the BD-ΛCDM models, the contour lines in Figure 1 show a preference for relatively high values of H_0 (the tension is lowered at $\sim 1.6\sigma$) while keeping $\sigma_8 \equiv \sigma_8(0)$ in the low value range (the tension remains at $\sim 0.4\sigma$), thus smoothing out both tensions at a time. We conclude that the models studied in this paper provide very interesting alternatives, either considering a time-evolving vacuum energy density or by extending the paradigm of General Relativity, to alleviate the H_0 and the σ_8 tensions. The better performance of the aforementioned models is reconfirmed by the Deviance Information Criterion, which points out that in some cases a very strong evidence in favor of a nonstandard model is found.

Acknowledgments

I would like to thank J. Solà Peracaula and A. Gómez-Valent for inviting me to speak in the CM3 parallel sesion "Status of the H_0 and σ_8 Tensions: Theoretical Models and Model-Independent Constraints of the MG16 Marcel Grossman Virtual Conference, July 5-10 2021. We also would like to thank the organizers of the MG16 for their great effort organizing such a great event. JdCP is supported by a FPI fellowship associated to the project FPA2016-76005-C2-1-P. JSP and CMP are partially supported by projects FPA2016-76005-C2-1-P (MINECO), 2017-SGR-929 (Generalitat de Catalunya) and MDM-2014-0369 (ICCUB). CMP is partially supported by the fellowship 2019 FI-B 00351. AGV is funded by the INFN – Project number 22425/2020.

References

1. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998), arXiv:astro-ph/9805201.
2. S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999), arXiv:astro-ph/9812133.
3. A. Einstein, The Field Equations of Gravitation, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 844 (1915).
4. Y. Zel'dovich, Cosmological constant and elementary particles, *Sov. Phys. JETP. Lett.* **6**, p. 3167 (1967).
5. Y. Zel'dovich, The cosmological constant and the theory of elementary particles, *Sov. Phys. Ups.* **11**, p. 381 (1968).
6. S. Weinberg, The Cosmological Constant Problem, *Rev. Mod. Phys.* **61**, 1 (1989).
7. P. J. E. Peebles and B. Ratra, The Cosmological Constant and Dark Energy, *Rev. Mod. Phys.* **75**, 559 (2003), arXiv:astro-ph/0207347.
8. T. Padmanabhan, Cosmological constant: The Weight of the vacuum, *Phys. Rept.* **380**, 235 (2003), arXiv:hep-th/0212290.
9. E. J. Copeland, M. Sami and S. Tsujikawa, Dynamics of dark energy, *Int. J. Mod. Phys. D* **15**, 1753 (2006), arXiv:hep-th/0603057.
10. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)].
11. M. J. Reid, D. W. Pesce and A. G. Riess, An Improved Distance to NGC 4258 and its Implications for the Hubble Constant, *Astrophys. J.* **886**, p. L27 (2019), arXiv:1908.05625.
12. E. Macaulay, I. K. Wehus and H. K. Eriksen, Lower Growth Rate from Recent Redshift Space Distortion Measurements than Expected from Planck, *Phys. Rev. Lett.* **111**, p. 161301 (2013), arXiv:1303.6583.
13. A. Gómez-Valent, Z. Zheng, L. Amendola, V. Pettorino and C. Wetterich, Early dark energy in the pre- and post-recombination epochs (7 2021).
14. G. Benevento, W. Hu and M. Raveri, Can Late Dark Energy Transitions Raise the Hubble constant?, *Phys. Rev. D* **101**, p. 103517 (2020), arXiv:2002.11707.
15. P. J. E. Peebles and B. Ratra, Cosmology with a Time Variable Cosmological Constant, *Astrophys. J.* **325**, p. L17 (1988).
16. B. Ratra and P. J. E. Peebles, Cosmological Consequences of a Rolling Homogeneous Scalar Field, *Phys. Rev.* **D37**, p. 3406 (1988).

17. R. D. Peccei, J. Solà and C. Wetterich, Adjusting the Cosmological Constant Dynamically: Cosmons and a New Force Weaker Than Gravity, *Phys. Lett.* **B195**, 183 (1987).
18. C. Wetterich, Cosmology and the Fate of Dilatation Symmetry, *Nucl. Phys.* **B302**, 668 (1988), arXiv:1711.03844.
19. C. Wetterich, The Cosmon model for an asymptotically vanishing time dependent cosmological 'constant', *Astron. Astrophys.* **301**, 321 (1995), arXiv:hep-th/9408025.
20. L. Amendola, Coupled quintessence, *Phys. Rev.* **D62**, p. 043511 (2000), arXiv:astro-ph/9908023.
21. A. Gómez-Valent, V. Pettorino and L. Amendola, Update on coupled dark energy and the H_0 tension, *Phys. Rev. D* **101**, p. 123513 (2020), arXiv:2004.00610.
22. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, First evidence of running cosmic vacuum: challenging the concordance model, *Astrophys. J.* **836**, p. 43 (2017).
23. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Possible signals of vacuum dynamics in the Universe, *Mon. Not. Roy. Astron. Soc.* **478**, 4357 (2018).
24. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, The H_0 tension in light of vacuum dynamics in the Universe, *Phys. Lett.* **B774**, 317 (2017), arXiv:1705.06723.
25. A. Gómez-Valent and J. Solà Peracaula, Density perturbations for running vacuum: a successful approach to structure formation and to the σ_8-tension, *Mon. Not. Roy. Astron. Soc.* **478**, 126 (2018), arXiv:1801.08501.
26. J. Solà Peracaula, A. Gómez-Valent and J. de Cruz Pérez, Signs of Dynamical Dark Energy in Current Observations, *Phys. Dark Univ.* **25**, p. 100311 (2019), arXiv:1811.03505.
27. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Running vacuum against the H_0 and σ_8 tensions, *EPL* **134**, p. 19001 (2021).
28. C. Brans and R. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev* **124**, p. 925 (1961).
29. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Brans–Dicke Gravity with a Cosmological Constant Smoothes Out ΛCDM Tensions, *Astrophys. J. Lett.* **886**, p. L6 (2019).
30. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Brans–Dicke cosmology with a Λ-term: a possible solution to ΛCDM tensions, *Class. Quant. Grav.* **37**, p. 245003 (2020).
31. C. W. Misner, K. S. Thorn and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1974).
32. J. Ooba, B. Ratra and N. Sugiyama, Planck 2015 Constraints on the Non-flat ΛCDM Inflation Model, *Astrophys. J.* **864**, p. 80 (2018).
33. J. Ooba, B. Ratra and N. Sugiyama, Planck 2015 Constraints on the Nonflat ϕCDM Inflation Model, *Astrophys. J.* **866**, p. 68 (2018).
34. C.-G. Park and B. Ratra, Observational constraints on the tilted flat-XCDM and the untilted nonflat XCDM dynamical dark energy inflation parameterizations, *Astrophys. Space Sci.* **364**, p. 82 (2019).
35. C.-G. Park and B. Ratra, Observational constraints on the tilted spatially-flat and the untilted nonflat ϕCDM dynamical dark energy inflation models, *Astrophys. J.* **868**, p. 83 (2018).
36. C.-G. Park and B. Ratra, Measuring the Hubble constant and spatial curvature from supernova apparent magnitude, baryon acoustic oscillation, and Hubble parameter data, *Astrophys. Space Sci.* **364**, p. 134 (2019).

37. C.-G. Park and B. Ratra, Using SPT polarization, *Planck* 2015, and non-CMB data to constrain tilted spatially-flat and untilted nonflat ΛCDM, XCDM, and ϕCDM dark energy inflation cosmologies, *Phys. Rev. D* **101**, p. 083508 (2020).
38. J. Ryan, Y. Chen and B. Ratra, Baryon acoustic oscillation, Hubble parameter, and angular size measurement constraints on the Hubble constant, dark energy dynamics, and spatial curvature, *Mon. Not. Roy. Astron. Soc.* **488**, 3844 (2019).
39. N. Khadka and B. Ratra, Using quasar X-ray and UV flux measurements to constrain cosmological model parameters, *Mon. Not. Roy. Astron. Soc.* **497**, 263 (2020), arXiv:2004.09979.
40. N. Khadka and B. Ratra, Quasar X-ray and UV flux, baryon acoustic oscillation, and Hubble parameter measurement constraints on cosmological model parameters, *Mon. Not. Roy. Astron. Soc.* **492**, 4456 (2020).
41. S. Cao, J. Ryan, N. Khadka and B. Ratra, Cosmological constraints from higher redshift gamma-ray burst, HII starburst galaxy, and quasar (and other) data, *Mon. Not. Roy. Astron. Soc.* **501**, 1520 (2021).
42. S. Cao, J. Ryan and B. Ratra, Using Pantheon and DES supernova, baryon acoustic oscillation, and Hubble parameter data to constrain the Hubble constant, dark energy dynamics, and spatial curvature, *Mon. Not. Roy. Astron. Soc.* **504**, 300 (2021).
43. N. Khadka and B. Ratra, Do quasar X-ray and UV flux measurements provide a useful test of cosmological models? (7 2021).
44. S. Cao, J. Ryan and B. Ratra, Cosmological constraints from HII starburst galaxy, quasar angular size, and other measurements (9 2021).
45. V. Faraoni, The $\omega \to \infty$ limit of Brans Dicke theory, *Phys. Lett.* **A245**, 26 (1998), arXiv:gr-qc/9805057.
46. V. Faraoni, Illusions of general relativity in Brans-Dicke gravity, *Phys. Rev.* **D59**, p. 084021 (1999), arXiv:gr-qc/9902083.
47. D. M. Scolnic *et al.*, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *Astrophys. J.* **859**, p. 101 (2018), arXiv:1710.00845.
48. P. Carter, F. Beutler, W. J. Percival, C. Blake, J. Koda and A. J. Ross, Low Redshift Baryon Acoustic Oscillation Measurement from the Reconstructed 6-degree Field Galaxy Survey, *Mon. Not. Roy. Astron. Soc.* **481**, 2371 (2018), arXiv:1803.01746.
49. E. A. Kazin *et al.*, The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature, *Mon. Not. Roy. Astron. Soc.* **441**, 3524 (2014), arXiv:1401.0358.
50. H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, *Mon. Not. Roy. Astron. Soc.* **465**, 1757 (2017), arXiv:1606.00439.
51. T. M. C. Abbott *et al.*, Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1, *Mon. Not. Roy. Astron. Soc.* **483**, 4866 (2019), arXiv:1712.06209.
52. R. Neveux *et al.*, The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the quasar sample between redshift 0.8 and 2.2, *Mon. Not. Roy. Astron. Soc.* **499**, 210 (2020).
53. H. du Mas des Bourboux *et al.*, The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests, *Astrophys. J.* **901**, p. 153 (2020).

54. R. Jiménez, L. Verde, T. Treu and D. Stern, Constraints on the equation of state of dark energy and the Hubble constant from stellar ages and the CMB, *Astrophys. J.* **593**, 622 (2003), arXiv:astro-ph/0302560.
55. J. Simon, L. Verde and R. Jimenez, Constraints on the redshift dependence of the dark energy potential, *Phys. Rev.* **D71**, p. 123001 (2005), arXiv:astro-ph/0412269.
56. D. Stern, R. Jiménez, L. Verde, M. Kamionkowski and S. A. Stanford, Cosmic Chronometers: Constraining the Equation of State of Dark Energy. I: H(z) Measurements, *JCAP* **1002**, p. 008 (2010), arXiv:0907.3149.
57. M. Moresco *et al.*, Improved constraints on the expansion rate of the Universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers, *JCAP* **1208**, p. 006 (2012), arXiv:1201.3609.
58. C. Zhang, H. Zhang, S. Yuan, T.-J. Zhang and Y.-C. Sun, Four new observational $H(z)$ data from luminous red galaxies in the Sloan Digital Sky Survey data release seven, *Res. Astron. Astrophys.* **14**, 1221 (2014), arXiv:1207.4541.
59. M. Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$, *Mon. Not. Roy. Astron. Soc.* **450**, L16 (2015), arXiv:1503.01116.
60. M. Moresco, L. Pozzetti, A. Cimatti, R. Jiménez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro and D. Wilkinson, A 6% measurement of the Hubble parameter at $z \sim 0.45$: direct evidence of the epoch of cosmic re-acceleration, *JCAP* **1605**, p. 014 (2016), arXiv:1601.01701.
61. A. L. Ratsimbazafy, S. I. Loubser, S. M. Crawford, C. M. Cress, B. A. Bassett, R. C. Nichol and P. Väisänen, Age-dating Luminous Red Galaxies observed with the Southern African Large Telescope, *Mon. Not. Roy. Astron. Soc.* **467**, 3239 (2017), arXiv:1702.00418.
62. K. Said, M. Colless, C. Magoulas, J. R. Lucey and M. J. Hudson, Joint analysis of 6dFGS and SDSS peculiar velocities for the growth rate of cosmic structure and tests of gravity, *Mon. Not. Roy. Astron. Soc.* **497**, 1275 (2020).
63. F. Simpson, C. Blake, J. A. Peacock, I. Baldry, J. Bland-Hawthorn, A. Heavens, C. Heymans, J. Loveday and P. Norberg, Galaxy and mass assembly: Redshift space distortions from the clipped galaxy field, *Phys. Rev.* **D93**, p. 023525 (2016), arXiv:1505.03865.
64. C. Blake *et al.*, Galaxy And Mass Assembly (GAMA): improved cosmic growth measurements using multiple tracers of large-scale structure, *Mon. Not. Roy. Astron. Soc.* **436**, p. 3089 (2013), arXiv:1309.5556.
65. F. G. Mohammad *et al.*, The VIMOS Public Extragalactic Redshift Survey (VIPERS): Unbiased clustering estimate with VIPERS slit assignment, *Astron. Astrophys.* **619**, p. A17 (2018), arXiv:1807.05999.
66. T. Okumura *et al.*, The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at $z \sim 1.4$, *Publ. Astron. Soc. Jap.* **68**, p. 38 (2016), arXiv:1511.08083.
67. J. Solà Peracaula, Brans–Dicke gravity: From Higgs physics to (dynamical) dark energy, *Int. J. Mod. Phys.* **D27**, p. 1847029 (2018), arXiv:1805.09810.
68. J. de Cruz Pérez and J. Solà Peracaula, Brans–Dicke cosmology mimicking running vacuum, *Mod. Phys. Lett.* **A33**, p. 1850228 (2018), arXiv:1809.03329.
69. N. Banerjee and D. Pavon, Cosmic acceleration without quintessence, *Phys. Rev. D* **63**, p. 043504 (2001).
70. N. Banerjee and D. Pavon, A Quintessence scalar field in Brans-Dicke theory, *Class. Quant. Grav.* **18**, p. 593 (2001), arXiv:gr-qc/0012098.

71. J. Solà, Cosmological constant and vacuum energy: old and new ideas, *J. Phys. Conf. Ser.* **453**, p. 012015 (2013), arXiv:1306.1527.
72. J. Solà and A. Gómez-Valent, The ΛCDM cosmology: From inflation to dark energy through running Λ, *Int. J. Mod. Phys. D* **24**, p. 1541003 (2015).
73. A. Gómez-Valent, Vacuum energy in Quantum Field Theory and Cosmology, PhD thesis, ICC, Barcelona U.2017. arXiv:1710.01978.
74. C. Moreno-Pulido and J. Solà, Running vacuum in quantum field theory in curved spacetime: renormalizing ρ_{vac} without $\sim m^4$ terms, *Eur. Phys. J. C* **80**, p. 692 (2020), arXiv:2005.03164.
75. J. Solà, Dark energy: A Quantum fossil from the inflationary Universe?, *J. Phys.* **A41**, p. 164066 (2008), arXiv:0710.4151.
76. J. A. S. Lima, S. Basilakos and J. Sola, Expansion History with Decaying Vacuum: A Complete Cosmological Scenario, *Mon. Not. Roy. Astron. Soc.* **431**, 923 (2013).
77. J. Solà, The cosmological constant and entropy problems: mysteries of the present with profound roots in the past, *Int. J. Mod. Phys. D* **24**, p. 1544027 (2015).
78. D. Blas, J. Lesgourgues and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, *JCAP* **1107**, p. 034 (2011), arXiv:1104.2933.
79. V. Salvatelli, N. Said, M. Bruni, A. Melchiorri and D. Wands, Indications of a late-time interaction in the dark sector, *Phys. Rev. Lett.* **113**, p. 181301 (2014), arXiv:1406.7297.
80. M. Martinelli, N. B. Hogg, S. Peirone, M. Bruni and D. Wands, Constraints on the interacting vacuum–geodesic CDM scenario, *Mon. Not. Roy. Astron. Soc.* **488**, 3423 (2019), arXiv:1902.10694.
81. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Dynamical dark energy vs. Λ = const in light of observations, *EPL* **121**, p. 39001 (2018).
82. B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, *JCAP* **1302**, p. 001 (2013), arXiv:1210.7183.
83. D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde, Bayesian measures of model complexity and fit, *J. Roy. Stat. Soc.* **64**, p. 583 (2002).
84. C. Heymans *et al.*, KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, *Astron. Astrophys.* **646**, p. A140 (2021).

Cosmological tensions: Hints for a new concordance model?

E. Di Valentino

School of Mathematics and Statistics, University of Sheffield, Hounsfield Road, Sheffield S3 7RH, United Kingdom
E-mail: e.divalentino@sheffield.ac.uk

The Cosmic Microwave Background (CMB) temperature and polarization anisotropy measurements have provided strong confirmation of the ΛCDM model of structure formation. Even if this model can explain incredibly well the observations in a vast range of scales and epochs, with the improvement of the experimental sensitivity, a few interesting tensions between the cosmological probes, and anomalies in the CMB data, have emerged with different statistical significance. While some portion of these discrepancies may be due to systematic errors, their persistence across probes strongly hints at cracks in the standard ΛCDM cosmological scenario. The most statistically significant is the Hubble constant puzzle and I will show a couple of interesting extended cosmological scenarios that can alleviate it.

Keywords: CMB; cosmology; H_0 tension.

1. Introduction

The aphorism "All the models are wrong, but some are useful" it is particularly true for the standard model adopted in cosmology. Actually, the model that has now practically been selected as the "standard" cosmological model is the Lambda Cold Dark Matter (ΛCDM) scenario, where Λ is the cosmological constant, that provides a remarkable fit to the bulk of available cosmological data.

However, despite its incredible success, ΛCDM harbours large areas of phenomenology and ignorance. For example, it still cannot explain key concepts in our understanding of the structure and evolution of the Universe, at the moment based on unknown quantities, that are also its largest components. In addition, their physical evidence comes from cosmological and astrophysical observations only, without strong theoretical motivations. In particular, the ΛCDM models is based on three unknown pillars:

- *Inflation*: an early stage of accelerated expansion which produces the initial, tiny, density perturbations, needed for structure formation,
- *Dark Matter*: a clustering matter component to facilitate structure formation,
- *Dark Energy*: an energy component to explain the current stage of accelerated expansion.

Furthermore, the ΛCDM model is based on the simplest form for these unknown quantities, mostly motivated by the computational simplicity, i.e. the theoretical

predictions under CDM for several observables are, in general, easier to compute and include fewer free parameters than most other solutions. These specific solutions for ΛCDM are:

- Inflation is given by a single, minimally coupled, slow-rolling scalar field;
- Dark Matter (DM) is a pressureless fluid made of cold, i.e., with low momentum, and collisionless particles;
- Dark Energy (DE) is a cosmological constant term.

Therefore, the 6 parameter ΛCDM model lacks the deep underpinnings a model requires to approach fundamental physics laws. It can be rightly considered, at best, as an effective theory of an underlying physical theory, yet to be discovered. In this situation, we must be careful not to cling to the model too tightly or to risk missing the appearance of departures from the paradigm. With the improvement of the number and the accuracy of the observations, deviations from ΛCDM may be expected. And, actually, discrepancies among key cosmological parameters of the models have emerged with different statistical significance. While some proportion of these discrepancies may have a systematic origin, their persistence across probes should require multiple and unrelated errors, strongly hinting at cracks in the standard cosmological scenario and the necessity of new physics. These tensions can indicate a failure of the canonical ΛCDM model.

2. The H_0 tension

The most statistically significant, long-lasting and widely persisting tension is the H_0 disagreement between the Planck estimate obtained assuming a "vanilla" ΛCDM cosmological model, i.e. $H_0 = 67.27 \pm 0.60$ km/s/Mpc at 68% CL,[1] and the local measurements obtained by the SH0ES collaboration, the so called R19[2] ($H_0 = 74.03 \pm 1.42$ km/s/Mpc at 68% CL) or R20[3] using the parallax measurements of Gaia EDR3 ($H_0 = 73.2 \pm 1.3$ km/s/Mpc at 68% CL).

However there are other many important measurements of the H_0 parameter. On the same side of Planck, i.e. preferring smaller values of H_0 we have the ground based CMB telescope, South Pole Telescope (SPT-3G)[4] that gives $H_0 = 68.8 \pm 1.5$ km/s/Mpc at 68% CL in ΛCDM, or Atacama Cosmology Telescope (ACT-DR4) + Wilkinson Microwave Anisotropy Probe (WMAP)[5] finding $H_0 = 67.6 \pm 1.1$ km/s/Mpc at 68% CL in ΛCDM. Moreover, there are the Baryon Acoustic Oscillations (BAO) + Big Bang Nucleosynthesis (BBN) from BOSS and eBOSS estimates,[6] that provide $H_0 = 67.35 \pm 0.97$ km/s/Mpc at 68% CL. These estimates are CMB independent but still ΛCDM dependent.

Moreover, we have many other direct late time measurements on the same side of SH0ES, i.e. preferring higher values of H_0 (see for example the compilation in Ref.[7] and references therein). For example, we have the measurements obtained from the Supernovae calibrated with the Cepheids,[3,8–10] or with the Tip of the Red Giant

Branch (TRGB),[8,11–14] where we find the well discussed Freedman et al. 2020 measurement[12] or the more updated Freedman 2021[15] that is completely in the middle between early and late measurements, so it can't really say anything about the tensions. Moreover, this TRGB-SNIa measurement has been reanalysed independently by Ref.[16] and we can see a shift towards higher values, more in agreement with SH0ES. Then we have the measurements obtained from the Supernovae calibrated with MIRAS,[17] variable red giant stars from older stellar populations, or the Megamaser Cosmology Project,[18] that measures H_0 avoiding any distance ladder and by providing the geometric distance directly into the Hubble flow. We have also measurements of the Hubble constant based on the Tully-Fisher Relation,[19,20] that are using as calibrators Cepheids and TRGB, or on the Surface Brightness Fluctuation,[21,22] that are substitutive distance ladder for long range indicator, and can be calibrated by both Cepheids and TRGB giving the same result,[21] or on Type II Supernovae[23] used as standardisable candles and calibrated again with both Cepheids and TRGB. Finally we have the strong lensing results, and, while the previous H0LiCOW[24] and STRIDES[25] results were more in agreement with SH0ES, the latest TDCOSMO+SLAC[26] constraints are shifted towards lower values, but with error bars much more weak, so they are not able to discriminate between SH0ES and Planck. This happens because the previous analyses were assuming a specific form for the mass density profile of the lens, and while this assumption is relaxed in a more agnostic approach, the constraints change significantly. This means that these kinds of measurements, even if cosmological model independent, are instead astrophysical model dependent.

The robustness of the H_0 disagreement can be proven considering several combinations of the late time measurements, as it has been done in Ref.,[7] where it is shown that also removing some of the most precise and strong measurements the discordance ranges between $\sim 3-6\sigma$, making it very difficult to be explained by systematic errors. Actually, we see a high precision and consistency of the data at both ends, the indirect ones preferring a lower value and the direct ones a higher value. Therefore, a possible solution should have enough rigor to explain multiple observations. A recent Review[15] seems to suggest that the situation is improving during the last 4 decades, with the H_0 measurement converging towards the same value. However, while in the past the tension was within the same types of measurements and at the same redshifts and thus pointing directly to systematics, now there are no late universe measurements below the early ones and vice versa, as expected in case of a scatter around the true value. Therefore, since it is hard to believe that a single type of systematic error could effectively resolve the Hubble constant tension, and because the discordance is robust under the removal of the measurements of any single type of object, mode or calibration, it is difficult to believe that unrelated systematic errors could coherently bias the measurements in the same direction and solve the H_0 puzzle.

3. Solutions

An alternative possibility for solving the Hubble tension is instead to change the assumptions about the cosmological model considered to derive the indirect constraints, modifying directly the three unknown pillars on which the standard vanilla ΛCDM is based.

3.1. *First pillar: Dark matter*

For example, we can consider modifications in the dark matter sector.

A classical extension is the effective number of relativistic degrees of freedom $N_{\rm eff}$,[27] i.e. additional relativistic matter at recombination, corresponding to a modification of the expansion history of the universe at early times, before the recombination epoch. The radiation density ρ_r can be recast as a function of the photon density ρ_γ, where we it is assumed the ratio $T_\nu/T_\gamma = (4/11)^{1/3}$ between the background temperatures of neutrinos and photons under the approximation of instantaneous neutrino decoupling:

$$\rho_r = \rho_\gamma \left[1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}\right]. \tag{1}$$

The expected value is $N_{\rm eff} = 3.046$,[28–32] if we assume standard electroweak interactions and three active massless neutrinos, but can be larger in presence of extra relativistic particles at recombination, such as, for example, sterile neutrinos[33–37] or thermal axions,[38–45] or decaying dark matter.[46–48] Additional relativistic degrees of freedom other than the three standard model neutrinos will produce more radiation, smearing and shifting of the acoustic peaks in the damping tail of the CMB temperature power spectrum, and delaying the matter to radiation equivalence,[49,50] with a corresponding increase of the early Integrated Sachs-Wolfe (eISW) effect, and therefore of the amplitude of the peak around multipoles $\ell \sim 200$. This parameter was often considered to increase the H_0 estimate because of the strong positive correlation present between these two quantities. However, $N_{\rm eff}$ should be around $N_{\rm eff} \approx 3.95$ in order to obtain a value of H_0 from Planck 2018 + BAO + Pantheon in perfect agreement with SH0ES,[51] and this value is significantly excluded by the latest Planck 2018 data. In particular, even when $N_{\rm eff}$ can vary freely, the H_0 tension is still above 3.6σ, and therefore this is no longer a suitable solution.

3.2. *Second pillar: Dark energy*

For example, we can consider modifications in the dark energy sector.

A classical extension is a varying dark energy equation of state, that is a modification of the expansion history of the universe at late times, after the recombination

epoch. In this case we replace the cosmological constant Λ in the expansion rate of the Universe with a Dark Energy with an equation of state w free to vary. A dark energy component with a time-varying equation of state $w \equiv p_{\rm DE}/\rho_{\rm DE}$ modifies the Hubble rate through the first Friedmann equation:

$$H^2(z) = H_0^2 \left[\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_{\rm DE}(1+z)^{3(1+w)} + \Omega_k(1+z)^2\right] \quad (2)$$

where Ω_r, Ω_m, $\Omega_{\rm DE}$ and Ω_k are the density parameters, evaluated at present time, for radiation, matter, dark energy and curvature for, and for the case $w = -1$ the dark energy component acts as a cosmological constant of density parameter $\Omega_{\rm DE}$. Changing the dark energy equation of state w, we are introducing a geometrical degeneracy with the Hubble constant that will be unconstrained using the CMB data only, resulting in agreement with SH0ES within 2σ at the price of a phantom dark energy with $w < -1$. Actually, a likelihood analysis of Planck 2018 gives $w = -1.58^{+0.16}_{-0.35}$ at 68% CL and $H_0 > 69.9\,{\rm km\,s^{-1}\,Mpc^{-1}}$ at 95% CL.[1] A phantom Dark energy implies that the energy density of this component increases with time in an expanding universe that will end in a Big Rip.[52] Moreover, it has many theoretical problems, but there exist models that expect an effective energy density with a phantom equation of state that can avoid them.

3.3. *Third pillar: Inflation*

For example, we can consider modifications in the inflationary sector.

Actually, a strong correlation is present between the scalar spectral index n_s of the primordial density fluctuations and the Hubble constant H_0. Therefore it is natural to ask if the simple Harrison-Zel'dovich model is able to solve the Hubble tension, or if there is support for a more complicated perturbation spectrum.

The inflationary theory predicts that the primordial spectrum should be a power law with a value of the spectral index to be nearly one, $n_s \sim 1$, reflecting the constancy of the Hubble horizon during inflation, but at the same time not exactly one, due to the dynamics of the inflaton field. An exact value of $n_s = 1$ is indeed not expected in inflation and would coincide with the phenomenological model proposed by Harrison,[53] Zel'dovich,[54] and Peebles and Yu,[55] known as Harrison - Zel'dovich (HZ) spectrum, proposed well before the formulation of inflation, and corresponding to perfect scale-invariance of the fluctuations. This model has one parameter fewer than standard vanilla ΛCDM model, and it is therefore less complicated (from the point of view of the number of parameters). While it is still possible to have inflationary models with spectral index nearly identical to HZ, a measurement of n_s close to, but different from, one should be considered as a further corroboration of inflation. The presence of the Hubble tension in current cosmological data does not rule out a HZ spectrum at high statistical significance because it predicts a Hubble constant value in agreement with SH0ES.[56] See also Ref.[57] for an updated analysis of the available CMB data in light of the different inflationary scenarios.

3.4. *Examples of formally successful models in solving H_0*

It is possible to classify the formally successful models in solving H_0 looking at the Hubble constant that a model predicts once the Planck 2018 data are analysed (see Table B1 of Ref.[58]). In this case, without any consideration about the goodness of the fit, excellent solutions in solving the H_0 tensions are late time solutions, i.e. mostly Dark Energy and Modified gravity models.

In these class of models we can find for example the Parker Vacuum Metamorphosis (VM),[59–61] which has a phase transition in the nature of the vacuum, is physically motivated by quantum gravitational effects and can mimic a phantom DE behaviour at low redshifts. This model can solve the Hubble tension in an excellent way.[62–64] For this model the expansion rate above and below the phase transition for a flat universe is for $z > z_t$:

$$\frac{H^2}{H_0^2} = \Omega_m(1+z)^3 + \Omega_r(1+z)^4 - M\left\{1 - \left[3\left(\frac{4}{3\Omega_m}\right)^4 M(1-M-\Omega_r)^3\right]^{-1}\right\}, \quad (3)$$

and for $z \leq z_t$:

$$\frac{H^2}{H_0^2} = (1-M)(1+z)^4 + M\,, \quad (4)$$

where $M = m^2/(12H_0^2)$ and the redshift of the phase transition is

$$z_t = -1 + \frac{3\Omega_m}{4(1-M-\Omega_r)}\,. \quad (5)$$

Therefore, above the phase transition, the universe behaves as one with matter and radiation plus a constant, and after the phase transition it effectively has a dark radiation component that rapidly redshifts away leaving a de Sitter phase. The original model did not include an explicit high redshift cosmological constant and therefore we have:

$$\Omega_m = \frac{4}{3}\left[3M(1-M-\Omega_r)^3\right]^{1/4}\,, \quad (6)$$

i.e. the parameter M is fixed and depends on the matter density, and this model has the same number of degrees of freedom as ΛCDM. For this model, the DE equation of state behaviour is phantom today, and more deeply phantom in the past. In the case without the cosmological constant, instead, there is no DE above the phase transition. If we now fit the current cosmological data with this model, we obtain H_0 exactly in agreement with SH0ES even if BAO and Pantheon SnIa data are included.[63] However, this considerably worsened the fit of the full Planck2018+BAO+Pantheon data combination because the model fails to recover the shape of H(z) at low redshifts.

Additionally, we can also consider an extended VM where M is an independent parameter. In this case, the massive scalar field has a vacuum expectation value

that manifests as a cosmological constant, and the conditions that the transition happens in the past plus the Ω_{DE} is always positive are assumed:

$$\frac{4}{3}(1-M-\Omega_r) \leq \Omega_m \leq \frac{4}{3}\left[3M(1-M-\Omega_r)^3\right]^{1/4} , \tag{7}$$

This more ad hoc VM model that includes a cosmological constant, i.e. allowing the vacuum criticality parameter M to float, is even better in fitting the CMB cosmological data alone,[63] but still can't recover the shape of H(z) at low redshifts preferred by BAO+Pantheon.

The problem is that BAO+Pantheon measurements constrain the product of H_0 and the sound horizon r_s. Therefore, in order to have a higher H_0 value in agreement with SH0ES, we need r_s near 137 Mpc. However, Planck, by assuming ΛCDM, prefers r_s near 147 Mpc. Therefore, a cosmological solution that can increase H_0 and at the same time can lower the sound horizon inferred from CMB data is promising to put in agreement all the measurements.[65] Therefore, the late time solutions, as wCDM, are disfavored by the full dataset combination because they increase H_0 but leave r_s unaltered. On the contrary, the early time solutions, as $N_{\rm eff}$ or Early Dark Energy,[66–68] move in the right direction both the parameters, but can't solve completely the H_0 tension with SH0ES, as we can see in Fig. 3 of Ref.[69]

At this point, it is possible to classify the formally successful models in solving H_0 looking at the Hubble constant that a model predicts once the Planck 2018 data combined with additional probes are analysed (see Table B2 of Ref.[58]). In this case, again without any consideration about the goodness of the fit, excellent solutions in solving the H_0 tensions are mostly early time solutions.

However, we should be careful in ruling out all the late time solutions because of the disagreement with the BAO data. Let's consider for example an Interacting model IDE, where DM and DE share interactions other than gravitational. Actually, dark matter and dark energy are described as separate fluids not sharing interactions beyond gravitational ones, even if from a microphysical perspective it is hard to imagine how non-gravitational DM-DE interactions can be avoided, if not forbidden by a fundamental symmetry. At the background level, the conservation equations for the pressureless DM and DE components can be decoupled into two separate equations with an inclusion of an arbitrary function, Q, known as the coupling or interacting function:

$$\dot{\rho}_c + 3\mathcal{H}\rho_c = Q , \tag{8}$$

$$\dot{\rho}_{DE} + 3\mathcal{H}(1+w)\rho_{DE} = -Q . \tag{9}$$

We assume the phenomenological form for the interaction rate:

$$Q = \mathcal{H}\xi\rho_{DE} \tag{10}$$

proportional to the dark energy density ρ_{DE} and the conformal Hubble rate $\mathcal{H}$, via a negative dimensionless parameter ξ quantifying the strength of the coupling, to

avoid early-time instabilities. In this scenario of IDE the tension on H_0 between the Planck satellite and SH0ES is completely solved. The coupling could affect the value of the present matter energy density Ω_m. Therefore, if within an interacting model Ω_m is smaller (because for negative ξ the dark matter density will decay into the dark energy one), a higher value of H_0 would be required in order to satisfy the peaks structure of CMB observations, which accurately determine the value of $\Omega_m h^2$ (see Table 1 of Ref.[70]). Therefore, we can safely combine the two datasets together, and we obtain a non-zero dark matter-dark energy coupling ξ at more than 5 standard deviations, i.e. $\xi = -0.66^{+0.09}_{-0.13}$ at 68% CL. For this model we have only an upper limit for the matter density Ω_m, however the IDE scenario is still in agreement with the local structures. Actually, within interacting cosmologies the growth of dark matter perturbations is larger than in uncoupled models.[71] This feature is general for models with negative coupling and in which the energy exchange among the dark sectors is proportional to ρ_{DE}, due to a suppression of the friction term and an enhancement of the source term in the differential growth equation.

For this model, the addition of low-redshift measurements, as BAO data, is extremely important, because it breaks the correlation between the parameters that can produce a fake detection for the coupling (see Ref.[72]). And Planck 2018 + BAO still hints at the presence of a coupling, albeit at a lower statistical significance. Also for these datasets the Hubble constant value is larger than that obtained in the case of a pure ΛCDM scenario, enough to bring the H_0 tension at 2.4σ.[73] This residual tension is however of difficult interpretation. In fact, BAO are formed in the early universe, when baryons are strongly coupled to photons, and the gravitational collapse due to the CDM is counterbalanced by the radiation pressure. Sound waves that propagate in the early universe imprint a characteristic scale on the CMB. Since the scale of these oscillations can be measured at recombination, BAO is considered a "standard ruler". These fluctuations have evolved and we can observe BAO at low redshifts in the distribution of galaxies. Since the data reduction process leading to these measurements requires assumptions about the fiducial cosmology, BAO is model dependent, and the model assumed to extrapolate the data is the standard ΛCDM scenario. In other words, the tension between Planck + BAO and SH0ES could be due to a statistical fluctuation in this case because the modified scenario of interacting dark energy could affect the BAO measurements. In fact, the full procedure which leads to the BAO constraints carried out by the different collaborations might be not necessarily valid in extended DE models. For instance, the BOSS collaboration advises caution when using their BAO measurements (both the pre- and post- reconstruction measurements) in more exotic dark energy cosmologies, because their reliability has been assessed only for modifications of the expansion history before recombination, and late modifications that can be parametrized by $w_0 w_a$. Therefore, BAO constraints themselves might need to be revised in a non-trivial manner when applied to constrain extended dark energy cosmologies, such as the IDE model that is showing a coupling different from ΛCDM at more than 5σ.

The IDE model can also solve the M_B tension. It has been pointed out in Ref.[74] that the SH0ES Collaboration, combining together anchors, cepheids and calibrators, produces a constraint on the SN absolute magnitude M_B, that depends on the astrophysical properties of the sources and is independent of cosmology. After that, in order to obtain the H_0 measurement, it uses the Hubble-ow supernovae in the redshift range $0.023 \leq z \leq 0.15$ to probe the luminosity distance-redshift relation, and it adopts a cosmography with $q_0 = 0.55$ and $j_0 = 1$. Therefore, the use of a prior on the Hubble constant could bias the conclusions about hockey-stick dark energy scenarios, i.e. models that have a transition at very-low redshift ($z \leq 0.1$), that can solve the Hubble tension producing a disagreement with M_B at more than 5σ.[75] Therefore, rather than explaining the H_0 tension, one should instead focus on the supernova absolute magnitude tension M_B, because the SH0ES team does not directly measure H_0. The M_B approach allows to skip the cosmography part, and to avoid the double counting of SN in the redshift range $0.023 \leq z \leq 0.15$ when combined with the Pantheon sample. Actually, imposing a H_0 prior, the whole information on the shape of the SN magnitude-redshift relation is lost, as also noticed by Ref.[76] If we consider the M_B parameter, instead, we have that the Planck constraint on the the SN absolute magnitude using the parametric-free inverse distance ladder predicts $M_B = -19.401 \pm 0.027$ mag, while the the SN measurements from SH0ES corresponds to $M_B = -19.244 \pm 0.037$ mag, corresponding to a 3.4σ tension. A classification of the models that can solve the M_B tension has been performed in Ref.,[77] finding that the best model is a closed universe with the electron mass free to vary.[78]

When an IDE model is considered, instead, the M_B tension is always at 2.4σ considering a combination of Pantheon + BAO + BBN, and is completely solved below 1σ when a gaussian prior on M_B is included, with an indication for a coupling different from zero at more than 95% CL.[79] Therefore, an IDE model can alleviate the M_B tension, as well as the H_0 disagreement, and as it is shown in Ref.[79] it is not disfavoured by the Pantheon magnitude-redshift relation.

4. Conclusions

To summarize the state-of-the-arte we have a long-standing Hubble tension that is presenting a serious limitation to precision cosmology. A solution to this problem is mandatory at this point.

On one side early time modifications of the standard models seem most promising for putting in agreement CMB and SH0ES, and agree also with BAO+Pantheon data, but unfortunately they don't completely solve the H_0 tension. On the other hand, late time modifications are instead more powerful in solving the H_0 tension, but are producing a disagreement with the additional BAO+Pantheon data. However, we can have working models, such as the simple IDE scenario, that can relieve the H_0 tension hinting to an interaction different from zero at more than 5σ. However, when BAO data are added in the analysis the Hubble constant tension is

restored at about 2.4σ, and this residual tension is difficult to interpret because of the reliability of the BAO data for this model. Alternative solutions can involve the inflationary paradigm, and in this context we saw that a HZ spectrum can solve the H_0 tension without new physics and with fewer parameters free to vary, but is disfavoured by a model comparison with the standard ΛCDM model.

Concluding, more than 300 models have been proposed (see Ref.[58]) to solve the Hubble tension, but at the moment no specific proposal makes a strong case for being highly likely or far better than all others. We therefore need new observations and the investigation of alternative theoretical models and solutions.

Acknowledgment

EDV is supported by a Royal Society Dorothy Hodgkin Research Fellowship.

References

1. N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020) [erratum: Astron. Astrophys. **652**, C4 (2021)] doi:10.1051/0004-6361/201833910 [arXiv:1807.06209 [astro-ph.CO]].
2. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Astrophys. J. **876**, no. 1, 85 (2019) doi:10.3847/1538-4357/ab1422 [arXiv:1903.07603 [astro-ph.CO]].
3. A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn and D. Scolnic, Astrophys. J. Lett. **908**, no. 1, L6 (2021) doi:10.3847/2041-8213/abdbaf [arXiv:2012.08534 [astro-ph.CO]].
4. D. Dutcher *et al.* [SPT-3G], Phys. Rev. D **104**, no. 2, 022003 (2021) doi:10.1103/PhysRevD.104.022003 [arXiv:2101.01684 [astro-ph.CO]].
5. S. Aiola *et al.* [ACT], JCAP **12**, 047 (2020) doi:10.1088/1475-7516/2020/12/047 [arXiv:2007.07288 [astro-ph.CO]].
6. S. Alam *et al.* [eBOSS], Phys. Rev. D **103**, no. 8, 083533 (2021) doi:10.1103/PhysRevD.103.083533 [arXiv:2007.08991 [astro-ph.CO]].
7. E. Di Valentino, Mon. Not. Roy. Astron. Soc. **502**, no. 2, 2065-2073 (2021) doi:10.1093/mnras/stab187 [arXiv:2011.00246 [astro-ph.CO]].
8. M. J. Reid, D. W. Pesce and A. G. Riess, Astrophys. J. Lett. **886**, no. 2, L27 (2019) doi:10.3847/2041-8213/ab552d [arXiv:1908.05625 [astro-ph.GA]].
9. L. Breuval, P. Kervella, R. I. Anderson, A. G. Riess, F. Arenou, B. Trahin, A. Mérand, A. Gallenne, W. Gieren and J. Storm, *et al.* Astron. Astrophys. **643**, A115 (2020) doi:10.1051/0004-6361/202038633 [arXiv:2006.08763 [astro-ph.SR]].
10. C. R. Burns *et al.* [CSP], Astrophys. J. **869**, no. 1, 56 (2018) doi:10.3847/1538-4357/aae51c [arXiv:1809.06381 [astro-ph.CO]].
11. J. Soltis, S. Casertano and A. G. Riess, Astrophys. J. Lett. **908**, no. 1, L5 (2021) doi:10.3847/2041-8213/abdbad [arXiv:2012.09196 [astro-ph.GA]].
12. W. L. Freedman, B. F. Madore, T. Hoyt, I. S. Jang, R. Beaton, M. G. Lee, A. Monson, J. Neeley and J. Rich, doi:10.3847/1538-4357/ab7339 [arXiv:2002.01550 [astro-ph.GA]].
13. W. Yuan, A. G. Riess, L. M. Macri, S. Casertano and D. Scolnic, Astrophys. J. **886**, 61 (2019) doi:10.3847/1538-4357/ab4bc9 [arXiv:1908.00993 [astro-ph.GA]].
14. I. S. Jang and M. G. Lee, Astrophys. J. **836**, no. 1, 74 (2017) doi:10.3847/1538-4357/836/1/74 [arXiv:1702.01118 [astro-ph.CO]].

15. W. L. Freedman, Astrophys. J. **919**, no. 1, 16 (2021) doi:10.3847/1538-4357/ac0e95 [arXiv:2106.15656 [astro-ph.CO]].
16. G. S. Anand *et al.* Astrophys. J., **932**, 1, 15 (2022) e-Print 2108.00007 [astro-ph. CO].
17. C. D. Huang, A. G. Riess, W. Yuan, L. M. Macri, N. L. Zakamska, S. Casertano, P. A. Whitelock, S. L. Hoffmann, A. V. Filippenko and D. Scolnic, doi:10.3847/1538-4357/ab5dbd [arXiv:1908.10883 [astro-ph.CO]].
18. D. W. Pesce, J. A. Braatz, M. J. Reid, A. G. Riess, D. Scolnic, J. J. Condon, F. Gao, C. Henkel, C. M. V. Impellizzeri and C. Y. Kuo, *et al.* Astrophys. J. Lett. **891**, no. 1, L1 (2020) doi:10.3847/2041-8213/ab75f0 [arXiv:2001.09213 [astro-ph.CO]].
19. E. Kourkchi, R. B. Tully, G. S. Anand, H. M. Courtois, A. Dupuy, J. D. Neill, L. Rizzi and M. Seibert, Astrophys. J. **896**, no. 1, 3 (2020) doi:10.3847/1538-4357/ab901c [arXiv:2004.14499 [astro-ph.GA]].
20. J. Schombert, S. McGaugh and F. Lelli, Astron. J. **160**, no. 2, 71 (2020) doi:10.3847/1538-3881/ab9d88 [arXiv:2006.08615 [astro-ph.CO]].
21. J. P. Blakeslee, J. B. Jensen, C. P. Ma, P. A. Milne and J. E. Greene, Astrophys. J. **911**, no. 1, 65 (2021) doi:10.3847/1538-4357/abe86a [arXiv:2101.02221 [astro-ph.CO]].
22. N. Khetan, L. Izzo, M. Branchesi, R. Wojtak, M. Cantiello, C. Murugeshan, A. Agnello, M. Della Valle, C. Gall and J. Hjorth, *et al.* Astron. Astrophys. **647**, A72 (2021) doi:10.1051/0004-6361/202039196 [arXiv:2008.07754 [astro-ph.CO]].
23. T. de Jaeger, B. E. Stahl, W. Zheng, A. V. Filippenko, A. G. Riess and L. Galbany, Mon. Not. Roy. Astron. Soc. **496**, no. 3, 3402-3411 (2020) doi:10.1093/mnras/staa1801 [arXiv:2006.03412 [astro-ph.CO]].
24. K. C. Wong, S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger and M. W. Auger, *et al.* Mon. Not. Roy. Astron. Soc. **498**, no. 1, 1420-1439 (2020) doi:10.1093/mnras/stz3094 [arXiv:1907.04869 [astro-ph.CO]].
25. A. J. Shajib *et al.* [DES], Mon. Not. Roy. Astron. Soc. **494**, no. 4, 6072-6102 (2020) doi:10.1093/mnras/staa828 [arXiv:1910.06306 [astro-ph.CO]].
26. S. Birrer, A. J. Shajib, A. Galan, M. Millon, T. Treu, A. Agnello, M. Auger, G. C. F. Chen, L. Christensen and T. Collett, *et al.* Astron. Astrophys. **643**, A165 (2020) doi:10.1051/0004-6361/202038861 [arXiv:2007.02941 [astro-ph.CO]].
27. G. Steigman, D. N. Schramm and J. E. Gunn, Phys. Lett. B **66**, 202-204 (1977) doi:10.1016/0370-2693(77)90176-9
28. G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P. D. Serpico, Nucl. Phys. B **729**, 221-234 (2005) doi:10.1016/j.nuclphysb.2005.09.041 [arXiv:hep-ph/0506164 [hep-ph]].
29. P. F. de Salas and S. Pastor, JCAP **07**, 051 (2016) doi:10.1088/1475-7516/2016/07/051 [arXiv:1606.06986 [hep-ph]].
30. K. Akita and M. Yamaguchi, JCAP **08**, 012 (2020) doi:10.1088/1475-7516/2020/08/012 [arXiv:2005.07047 [hep-ph]].
31. J. Froustey, C. Pitrou and M. C. Volpe, JCAP **12**, 015 (2020) doi:10.1088/1475-7516/2020/12/015 [arXiv:2008.01074 [hep-ph]].
32. J. J. Bennett, G. Buldgen, P. F. De Salas, M. Drewes, S. Gariazzo, S. Pastor and Y. Y. Y. Wong, JCAP **04**, 073 (2021) doi:10.1088/1475-7516/2021/04/073 [arXiv:2012.02726 [hep-ph]].
33. A. Melchiorri, O. Mena, S. Palomares-Ruiz, S. Pascoli, A. Slosar and M. Sorel, JCAP **01**, 036 (2009) doi:10.1088/1475-7516/2009/01/036 [arXiv:0810.5133 [hep-ph]].
34. J. Kopp, M. Maltoni and T. Schwetz, Phys. Rev. Lett. **107**, 091801 (2011) doi:10.1103/PhysRevLett.107.091801 [arXiv:1103.4570 [hep-ph]].
35. J. Kopp, P. A. N. Machado, M. Maltoni and T. Schwetz, JHEP **05**, 050 (2013) doi:10.1007/JHEP05(2013)050 [arXiv:1303.3011 [hep-ph]].

36. S. Gariazzo, P. F. de Salas and S. Pastor, JCAP **07**, 014 (2019) doi:10.1088/1475-7516/2019/07/014 [arXiv:1905.11290 [astro-ph.CO]].
37. S. Hagstotz, P. F. de Salas, S. Gariazzo, M. Gerbino, M. Lattanzi, S. Vagnozzi, K. Freese and S. Pastor, [arXiv:2003.02289 [astro-ph.CO]].
38. A. Melchiorri, O. Mena and A. Slosar, Phys. Rev. D **76**, 041303 (2007) doi:10.1103/PhysRevD.76.041303 [arXiv:0705.2695 [astro-ph]].
39. S. Hannestad, A. Mirizzi, G. G. Raffelt and Y. Y. Y. Wong, JCAP **08**, 015 (2007) doi:10.1088/1475-7516/2007/08/015 [arXiv:0706.4198 [astro-ph]].
40. S. Hannestad, A. Mirizzi, G. G. Raffelt and Y. Y. Y. Wong, JCAP **04**, 019 (2008) doi:10.1088/1475-7516/2008/04/019 [arXiv:0803.1585 [astro-ph]].
41. M. Archidiacono, S. Hannestad, A. Mirizzi, G. Raffelt and Y. Y. Y. Wong, JCAP **10**, 020 (2013) doi:10.1088/1475-7516/2013/10/020 [arXiv:1307.0615 [astro-ph.CO]].
42. E. Giusarma, E. Di Valentino, M. Lattanzi, A. Melchiorri and O. Mena, Phys. Rev. D **90**, no. 4, 043507 (2014) doi:10.1103/PhysRevD.90.043507 [arXiv:1403.4852 [astro-ph.CO]].
43. E. Di Valentino, E. Giusarma, M. Lattanzi, O. Mena, A. Melchiorri and J. Silk, Phys. Lett. B **752**, 182-185 (2016) doi:10.1016/j.physletb.2015.11.025 [arXiv:1507.08665 [astro-ph.CO]].
44. W. Giarè, E. Di Valentino, A. Melchiorri and O. Mena, Mon. Not. Roy. Astron. Soc. **505**, no. 2, 2703-2711 (2021) doi:10.1093/mnras/stab1442 [arXiv:2011.14704 [astro-ph.CO]].
45. W. Giarè, F. Renzi, A. Melchiorri, O. Mena and E. Di Valentino, [arXiv:2110.00340 [astro-ph.CO]].
46. Z. Berezhiani, A. D. Dolgov and I. I. Tkachev, Phys. Rev. D **92**, no. 6, 061303 (2015) doi:10.1103/PhysRevD.92.061303 [arXiv:1505.03644 [astro-ph.CO]].
47. K. Vattis, S. M. Koushiappas and A. Loeb, Phys. Rev. D **99**, no. 12, 121302 (2019) doi:10.1103/PhysRevD.99.121302 [arXiv:1903.06220 [astro-ph.CO]].
48. N. Blinov, C. Keith and D. Hooper, JCAP **06**, 005 (2020) doi:10.1088/1475-7516/2020/06/005 [arXiv:2004.06114 [astro-ph.CO]].
49. Z. Hou, R. Keisler, L. Knox, M. Millea and C. Reichardt, Phys. Rev. D **87**, 083008 (2013) doi:10.1103/PhysRevD.87.083008 [arXiv:1104.2333 [astro-ph.CO]].
50. M. Archidiacono, E. Giusarma, S. Hannestad and O. Mena, Adv. High Energy Phys. **2013**, 191047 (2013) doi:10.1155/2013/191047 [arXiv:1307.0637 [astro-ph.CO]].
51. S. Vagnozzi, Phys. Rev. D **102**, no. 2, 023518 (2020) doi:10.1103/PhysRevD.102.023518 [arXiv:1907.07569 [astro-ph.CO]].
52. R. R. Caldwell, M. Kamionkowski and N. N. Weinberg, Phys. Rev. Lett. **91**, 071301 (2003) doi:10.1103/PhysRevLett.91.071301 [arXiv:astro-ph/0302506 [astro-ph]].
53. E. R. Harrison, Phys. Rev. D 1, 2726 (1970).
54. Y. B. Zel'dovich, Mon. Not. Roy. Astron. Soc. 160, (1972).
55. P. J. E. Peebles and J. T. Yu Astrophys. J. 162, 815 (1970).
56. E. Di Valentino, A. Melchiorri, Y. Fantaye and A. Heavens, Phys. Rev. D **98**, no. 6, 063508 (2018) doi:10.1103/PhysRevD.98.063508 [arXiv:1808.09201 [astro-ph.CO]].
57. M. Forconi, W. Giarè, E. Di Valentino and A. Melchiorri, [arXiv:2110.01695 [astro-ph.CO]].
58. E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess and J. Silk, Class. Quant. Grav. **38**, no. 15, 153001 (2021) doi:10.1088/1361-6382/ac086d [arXiv:2103.01183 [astro-ph.CO]].
59. L. Parker and D. A. T. Vanzella, Phys. Rev. D **69**, 104009 (2004) doi:10.1103/PhysRevD.69.104009 [arXiv:gr-qc/0312108 [gr-qc]].

60. L. Parker and A. Raval, Phys. Rev. D **62**, 083503 (2000) [erratum: Phys. Rev. D **67**, 029903 (2003)] doi:10.1103/PhysRevD.62.083503 [arXiv:gr-qc/0003103 [gr-qc]].
61. R. R. Caldwell, W. Komp, L. Parker and D. A. T. Vanzella, Phys. Rev. D **73**, 023513 (2006) doi:10.1103/PhysRevD.73.023513 [arXiv:astro-ph/0507622 [astro-ph]].
62. E. Di Valentino, E. V. Linder and A. Melchiorri, Phys. Rev. D **97**, no. 4, 043528 (2018) doi:10.1103/PhysRevD.97.043528 [arXiv:1710.02153 [astro-ph.CO]].
63. E. Di Valentino, E. V. Linder and A. Melchiorri, Phys. Dark Univ. **30**, 100733 (2020) doi:10.1016/j.dark.2020.100733 [arXiv:2006.16291 [astro-ph.CO]].
64. E. Di Valentino, S. Pan, W. Yang and L. A. Anchordoqui, Phys. Rev. D **103**, no. 12, 123527 (2021) doi:10.1103/PhysRevD.103.123527 [arXiv:2102.05641 [astro-ph.CO]].
65. L. Knox and M. Millea, Phys. Rev. D **101**, no. 4, 043533 (2020) doi:10.1103/PhysRevD.101.043533 [arXiv:1908.03663 [astro-ph.CO]].
66. V. Poulin, T. L. Smith, T. Karwal and M. Kamionkowski, Phys. Rev. Lett. **122**, no. 22, 221301 (2019) doi:10.1103/PhysRevLett.122.221301 [arXiv:1811.04083 [astro-ph.CO]].
67. J. C. Hill, E. Calabrese, S. Aiola, N. Battaglia, B. Bolliet, S. K. Choi, M. J. Devlin, A. J. Duivenvoorden, J. Dunkley and S. Ferraro, *et al.* [arXiv:2109.04451 [astro-ph.CO]].
68. V. Poulin, T. L. Smith and A. Bartlett, [arXiv:2109.06229 [astro-ph.CO]].
69. N. Arendse, R. J. Wojtak, A. Agnello, G. C. F. Chen, C. D. Fassnacht, D. Sluse, S. Hilbert, M. Millon, V. Bonvin and K. C. Wong, *et al.* Astron. Astrophys. **639**, A57 (2020) doi:10.1051/0004-6361/201936720 [arXiv:1909.07986 [astro-ph.CO]].
70. E. Di Valentino, A. Melchiorri, O. Mena and S. Vagnozzi, Phys. Dark Univ. **30**, 100666 (2020) doi:10.1016/j.dark.2020.100666 [arXiv:1908.04281 [astro-ph.CO]].
71. G. Caldera-Cabral, R. Maartens and B. M. Schaefer, JCAP **07**, 027 (2009) doi:10.1088/1475-7516/2009/07/027 [arXiv:0905.0492 [astro-ph.CO]].
72. E. Di Valentino and O. Mena, Mon. Not. Roy. Astron. Soc. **500**, no. 1, L22-L26 (2020) doi:10.1093/mnrasl/slaa175 [arXiv:2009.12620 [astro-ph.CO]].
73. E. Di Valentino, A. Melchiorri, O. Mena and S. Vagnozzi, Phys. Rev. D **101**, no. 6, 063502 (2020) doi:10.1103/PhysRevD.101.063502 [arXiv:1910.09853 [astro-ph.CO]].
74. D. Camarena and V. Marra, Phys. Rev. Res. **2**, no. 1, 013028 (2020) doi:10.1103/PhysRevResearch.2.013028 [arXiv:1906.11814 [astro-ph.CO]].
75. D. Camarena and V. Marra, Mon. Not. Roy. Astron. Soc. **504**, 5164-5171 (2021) doi:10.1093/mnras/stab1200 [arXiv:2101.08641 [astro-ph.CO]].
76. G. Efstathiou, Mon. Not. Roy. Astron. Soc. **505**, no. 3, 3866-3872 (2021) doi:10.1093/mnras/stab1588 [arXiv:2103.08723 [astro-ph.CO]].
77. N. Schöneberg, G. Franco Abellán, A. Pérez Sánchez, S. J. Witte, V. Poulin and J. Lesgourgues, [arXiv:2107.10291 [astro-ph.CO]].
78. T. Sekiguchi and T. Takahashi, Phys. Rev. D **103**, no. 8, 083507 (2021) doi:10.1103/PhysRevD.103.083507 [arXiv:2007.03381 [astro-ph.CO]].
79. R. C. Nunes and E. Di Valentino, Phys. Rev. D **104**, no. 6, 063529 (2021) doi:10.1103/PhysRevD.104.063529 [arXiv:2107.09151 [astro-ph.CO]].

Solving both H_0 and σ_8 tensions in $f(T)$ gravity

Emmanuel N. Saridakis
National Observatory of Athens, Lofos Nymfon, 11852 Athens, Greece
E-mail: msaridak@noa.gr

We report how to alleviate both the H_0 and σ_8 tensions simultaneously within $f(T)$ gravity. In particular, we consider the parametrization $f(T) = -T - 2\Lambda/M_P^2 + \alpha T^\beta$, where two out of the three parameters are independent. This model can efficiently fit observations solving the two tensions. To our knowledge, this is the first time where a modified gravity theory can alleviate both H_0 and σ_8 tensions simultaneously, hence, offering an additional argument in favor of gravitational modification.

Keywords: H_0 tension; σ_8 tension; $f(T)$ gravity; MG16 Proceedings.

1. Introduction

It is now well established that the Universe at late times experienced the transition from the matter era to the accelerated expansion phase. Although the simplest explanation would be the consideration of the cosmological constant, the corresponding problem related to the quantum-field-theoretical calculation of its value, as well as the possibility of a dynamical nature, led to two main paths of constructing extended scenarios. The first is to maintain general relativity as the underlying theory of gravity, and consider new, exotic forms of matter that constitute the concept of dark energy.[1,2] The second is to construct extended or modified theories of gravity, that posses general relativity as a low-energy limit, but which in general provide the extra degrees of freedom that can drive the dynamical universe acceleration.[3,4]

With the accumulation of cosmological data, experimental tensions may arise within ΛCDM cosmology. If they were to remain under the increasing precision of experimental observations, they would constitute, in a statistical sense, clear indications of new physics beyond ΛCDM. One recently well-debated tension relates the value of the Hubble parameter at present time H_0 measured from the cosmic microwave background (CMB) temperature and polarization data by the Planck Collaboration[5] to be $H_0 = 67.37 \pm 0.54\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, to the one from local measurements of the Hubble Space Telescope[6] yielding $H_0 = 74.03 \pm 1.42\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. Recent analyses with the combination of gravitational lensing and time-delay effects data reported a significant deviation[7] at 5.3σ. Another potential tension concerns the measurements of the parameter σ_8, which quantifies the gravitational clustering of matter from the amplitude of the linearly evolved power spectrum at the scale of $8h^{-1}$Mpc. Specifically, a possible deviation was noticed between measurements of CMB and LSS surveys, namely, between Planck[5] and SDSS/BOSS.[8,9] Although these two tensions could in principle arise from unknown systematics, the possibility

of physical origin puts the standard lore of cosmology into additional investigations, by pointing to various extensions beyond ΛCDM.

In this work we consider systematically the H_0[10] and σ_8[11] tensions, and report how to alleviate both simultaneously within $f(T)$ gravity, based on the analysis of.[12] We exploit the effective field theory (EFT) of torsional gravity, a formalism that allows for a systematic investigation of the background and perturbations separately. This approach was first developed early for curvature gravity,[13] and recently it was extended to torsional gravity.[14,15] In order to address cosmological tensions, we identify the effects of gravitational modifications within the EFT on the dynamics of the background and of linear perturbation levels. This will allow us to construct specific models of $f(T)$ gravity providing adequate deviation from ΛCDM that can alleviate H_0 and σ_8 tensions.

2. Effective field theory approach

We start our analysis by the question of whether both tensions can be alleviated simultaneously via gravitational modifications in a much general framework. Indeed, the H_0 tension reveals a universe that is expanding faster at late times than that from a cosmological constant preferred by CMB data, while a lower value of σ_8 than the one of CMB most likely ΛCDM would imply that matter clusters either later on or less efficiently. Hence, these two observations seem to indicate that there might be "less gravitational power" at intermediate scales, which phenomenologically advocates a possible modification of gravitation. Accordingly, in this work we address the aforementioned question within the EFT framework for torsional gravity.

For a general curvature-based gravity, the action following the EFT approach in the unitary gauge, invariant by space diffeomorphisms, which expanded around a flat FRW metric $ds^2 = -dt^2 + a^2(t)\,\delta_{ij}dx^i dx^j$, is given by

$$\begin{aligned} S = \int d^4x \Big\{ \sqrt{-g}\Big[\frac{M_P^2}{2}\Psi(t)R - \Lambda(t) - b(t)g^{00} \\ + M_2^4(\delta g^{00})^2 - \bar{m}_1^3\delta g^{00}\delta K - \bar{M}_2^2\delta K^2 - \bar{M}_3^2\delta K^\nu_\mu \delta K^\mu_\nu \\ + m_2^2 h^{\mu\nu}\partial_\mu g^{00}\partial_\nu g^{00} + \lambda_1\delta R^2 + \lambda_2\delta R_{\mu\nu}\delta R^{\mu\nu} + \mu_1^2\delta g^{00}\delta R\Big] \\ + \gamma_1 C^{\mu\nu\rho\sigma}C_{\mu\nu\rho\sigma} + \gamma_2\epsilon^{\mu\nu\rho\sigma}C_{\mu\nu}{}^{\kappa\lambda}C_{\rho\sigma\kappa\lambda} \\ + \sqrt{-g}\Big[\frac{M_3^4}{3}(\delta g^{00})^3 - \bar{m}_2^3(\delta g^{00})^2\delta K + \cdots\Big]\Big\}\,, \end{aligned} \tag{1}$$

where $M_P = (8\pi G_N)^{-1/2}$ is the reduced Planck mass with G_N the Newtonian constant. R is the Ricci scalar corresponding to the Levi-Cività connection, $C^{\mu\nu\rho\sigma}$ is the Weyl tensor, δK^ν_μ is the perturbation of the extrinsic curvature, and the functions $\Psi(t)$, $\Lambda(t)$, $b(t)$ are determined by the background evolution.

When the underlying theory also includes torsion,[14] one can generalize the EFT action as[14,15]

$$S = \int d^4x\sqrt{-g}\Big[\frac{M_P^2}{2}\Psi(t)R - \Lambda(t) - b(t)g^{00} + \frac{M_P^2}{2}d(t)T^0\Big] + S^{(2)}\,. \tag{2}$$

To compare with the effective action for curvature-based gravity (1), one reads that at background level there is additionally the zeroth part T^0 of the contracted torsion tensor $T^{0\mu}_{\ \ \mu}$, with its time-dependent coefficient $d(t)$. Furthermore, the perturbation part $S^{(2)}$ contains all operators of the perturbation part of (1), plus pure torsion terms including δT^2, $\delta T^0 \delta T^0$, and $\delta T^{\rho\mu\nu}\delta T_{\rho\mu\nu}$, and extra terms that mix curvature and torsion, namely, $\delta T \delta R$, $\delta g^{00}\delta T$, $\delta g^{00}\delta T^0$, and $\delta K \delta T^0$, where $T \equiv \frac{1}{4}T^{\rho\mu\nu}T_{\rho\mu\nu} + \frac{1}{2}T^{\rho\mu\nu}T_{\nu\mu\rho} - T_{\rho\mu}{}^{\rho}T^{\nu\mu}{}_{\nu}$ is the torsion scalar. Adding the matter action S_m to the effective action of torsional-based gravity (2) and then performing variation, one obtains the Friedmann equations to be:[14]

$$H^2 = \frac{1}{3M_P^2}\big(\rho_m + \rho_{DE}^{\rm eff}\big) \ , \tag{3}$$
$$\dot{H} = -\frac{1}{2M_P^2}\big(\rho_m + \rho_{DE}^{\rm eff} + p_m + p_{DE}^{\rm eff}\big) \ ,$$

and where

$$\rho_{DE}^{\rm eff} = b + \Lambda - 3M_P^2\Big[H\dot{\Psi} + \frac{dH}{2} + H^2(\Psi - 1)\Big] \ , \tag{4}$$
$$p_{DE}^{\rm eff} = b - \Lambda + M_P^2\Big[\ddot{\Psi} + 2H\dot{\Psi} + \frac{\dot{d}}{2} + (H^2 + 2\dot{H})(\Psi - 1)\Big] \ ,$$

are, respectively, the effective DE density and pressure in the general torsional gravity. Moreover, we treat the matter sector as dust that satisfies the conservation equation $\dot{\rho}_m + 3H(\rho_m + p_m) = 0$, which in terms of redshift leads to $\rho_m = 3M_P^2 H_0^2 \Omega_{m0}(1+z)^3$, with Ω_{m0} the value of $\Omega_m \equiv 8\pi G_N \rho_m/(3H^2)$ at present.

3. Model independent analysis

In general, to avoid the H_0 tension one needs a positive correction to the first Friedmann equation at late times that could yield an increase in H_0 compared to the ΛCDM scenario. As for the σ_8 tension, we recall that in any cosmological model, at sub-Hubble scales and through the matter epoch, the equation that governs the evolution of matter perturbations in the linear regime is[16]

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G_{\rm eff}\rho_m \delta \ , \tag{5}$$

where $G_{\rm eff}$ is the effective gravitational coupling given by a generalized Poisson equation. In general, $G_{\rm eff}$ differs from the Newtonian constant G_N, and thus contains information from gravitational modifications (note that $G_{\rm eff} = G_N$ in ΛCDM cosmology). Solving for $\delta(a)$ provides the observable quantity $f\sigma_8(a)$, following the definitions $f(a) \equiv d\ln\delta(a)/d\ln a$ and $\sigma(a) = \sigma_8\delta(1)/\delta(a=1)$. Hence, alleviation of the σ_8 tension may be obtained if $G_{\rm eff}$ becomes smaller than G_N during the growth of matter perturbations and/or if the "friction" term in (5) increases.

To grasp the physical picture, we start with a simple case: $b(t) = 0$ and $\Lambda(t) = \Lambda = const$ [b and Λ are highly degenerate as shown in (4)], while $\Psi(t) = 1$. Hence, from (2), with the above coefficient choices, the only deviation from ΛCDM

at the background level comes from the term $d(t)T^0$, and we remind that in FRW geometry $T^0 = H$ when evaluated on the background. In this case, the first Friedmann equation in (3), using for convenience the redshift $z = -1 + a_0/a$ as the dimensionless variable and setting $a_0 = 1$, yields

$$H(z) = -\frac{d(z)}{4} + \sqrt{\frac{d^2(z)}{16} + H^2_{\Lambda\mathrm{CDM}}(z)}\ , \tag{6}$$

where $H_{\Lambda\mathrm{CDM}}(z) \equiv H_0\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$ is the Hubble rate in ΛCDM, with $\Omega_m = \rho_m/(3M_p^2H^2)$ the matter density parameter and primes denote derivatives with respect to z. Accordingly, if $d < 0$ and is suitably chosen, one can have $H(z \to z_{\mathrm{CMB}}) \approx H_{\Lambda\mathrm{CDM}}(z \to z_{\mathrm{CMB}})$ but $H(z \to 0) > H_{\Lambda\mathrm{CDM}}(z \to 0)$; i.e., the H_0 tension is solved [one should choose $|d(z)| < H(z)$, and thus, since $H(z)$ decreases for smaller z, the deviation from ΛCDM will be significant only at low redshift]. Additionally, since the friction term in (5) increases, the growth of structure gets damped, and therefore, the σ_8 tension is also solved [note that since we have imposed $\Psi = 1$, then $G_{\mathrm{eff}} = G_N$ as one can verify from (2) and (3); namely, the contributions from T^0 vanish at first order in perturbations].

Furthermore, for typical values that lie well within the 1σ intervals of the $H(z)$ redshift surveys, it is expected that CMB measurements will be sensitive to such a deviation from the ΛCDM scenario for nonvanishing T^0 at early times. Actually, the T^0 operator acts in a similar way as a conventional cosmological constant. Thus, it adds yet another new functional form to parametrize the background and leads to more flexibility in fitting redshift and clustering measurements.

4. $f(T)$ gravity and cosmology

In this section we propose concrete models of torsional modified gravity that can be applied to alleviate the two cosmological tensions based on the torsional EFT dictionary. In particular, we focus on the well-known class of torsional gravity, namely, the $f(T)$ gravity,[17] which is characterized by the action $S = \frac{M_P^2}{2}\int d^4x e f(T)$, with $e = \det(e^A_\mu) = \sqrt{-g}$ and e^A_μ the vierbein, and thus by the Friedmann equations

$$\begin{aligned} H^2 &= \frac{\rho_m}{3M_P^2} + \frac{T}{6} - \frac{f}{6} + \frac{Tf_T}{3} \\ \dot{H} &= -\frac{1}{2M_P^2}(\rho_m + p_m) + \dot{H}(1 + f_T + 2Tf_{TT}), \end{aligned} \tag{7}$$

with $f_T \equiv \partial f/\partial T$, $f_{TT} \equiv \partial^2 f/\partial T^2$, where we have applied $T = 6H^2$ in flat FRW geometry (we follow the convention of[14]). Therefore, $f(T)$ gravity can arise from the general EFT approach to torsional gravity by choosing $\Psi = -f_T$, $\Lambda = \frac{M_P^2}{2}(Tf_T - f)$, $b = 0$, $d = 2\dot{f}_T$,[14] and can restore GR by choosing $f(T) = -T - 2\Lambda/M_P^2$.

The above EFT approach holds for every $f(T)$ gravity by making a suitable identification of the involved time-dependent functions. For instance, we consider the following ansatz: $f(T) = -[T + 6H_0^2(1 - \Omega_{m0}) + F(T)]$, where $F(T)$ describes

the deviation from GR [note, however, that in FRW geometry, apart from the regular choice $F = 0$, the ΛCDM scenario can also be obtained for the special case $F(T) = c\,T^{1/2}$ too, with c a constant]. Under these assumptions, the first Friedmann equation becomes

$$T(z) + 2\frac{F'(z)}{T'(z)}T(z) - F(z) = 6H^2_{\Lambda CDM}(z)\ . \tag{8}$$

In order to solve the H_0 tension, we need $T(0) = 6H_0^2 \simeq 6(H_0^{CC})^2$, with $H_0^{CC} = 74.03\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ following the local measurements,[6] while in the early era of $z \gtrsim 1100$ we require the Universe expansion to evolve as in ΛCDM, namely $H(z \gtrsim 1100) \simeq H_{\Lambda CDM}(z \gtrsim 1100)$. This requirement follows from the fact that we are considering modifications kicking in only at late times, and therefore, the results in tension from CMB analysis performed within ΛCDM remain unaffected. This implies $F(z)|_{z\gtrsim 1100} \simeq cT^{1/2}(z)$ (the value $c = 0$ corresponds to standard GR, while for $c \neq 0$ we obtain ΛCDM too). Note that, in this case the effective gravitational coupling is given by[18]

$$G_{\rm eff} = \frac{G_N}{1 + F_T}\ . \tag{9}$$

Therefore, the perturbation equation at linear order (5) becomes

$$\delta'' + \left[\frac{T'(z)}{2T(z)} - \frac{1}{1+z}\right]\delta' = \frac{9H_0^2\Omega_{m0}(1+z)}{[1 + F'(z)/T'(z)]T(z)}\delta\ , \tag{10}$$

where $\delta \equiv \delta\rho_m/\rho_m$ is the matter overdensity. Since around the last scattering moment $z \gtrsim 1100$ the Universe should be matter-dominated, we impose $\delta'(z)|_{z\gtrsim 1100} \simeq -\frac{1}{1+z}\delta(z)$, while at late times we look for $\delta(z)$ that leads to an $f\sigma_8$ in agreement with redshift survey observations.

By solving (8) and (10) with initial and boundary conditions at $z \sim 0$ and $z \sim 1100$, we can find the functional forms for the free functions of the $f(T)$ gravity that we consider, namely, $T(z)$ and $F(z)$, that can alleviate both H_0 and σ_8 tensions. We find two such forms for $F(T)$. Both models approach the ΛCDM scenario at $z \gtrsim 1100$, with Model-1 approaching $F = 0$ and hence restoring GR, while Model-2 approaches $F \propto T^{1/2}$, and thus it reproduces ΛCDM. In particular, we find that we can well fit the numerical solutions of Model-1 by

$$F(T) \approx 375.47\Big(\frac{T}{6H_0^2}\Big)^{-1.65}\ , \tag{11}$$

and of Model-2 by

$$F(T) \approx 375.47\Big(\frac{T}{6H_0^2}\Big)^{-1.65} + 25T^{1/2}\ . \tag{12}$$

Note that, the first term of Model-2, which coincides with Model-1, provides a small deviation to ΛCDM at late times, while it decreases rapidly to become negligible in the early Universe. In addition, we examine $G_{\rm eff}$ given by (9) for the two models (11) and (12), which are displayed in Fig. 1. As expected, at high redshifts in both

models, $G_{\rm eff}$ becomes G_N, recovering the ΛCDM paradigm. At very low redshifts, $G_{\rm eff}$ becomes slightly higher than G_N, increasing slightly the gravitational strength. This gravitational modification is in competition at late times with the accelerating expansion. It turns out that the effect of an increased cosmic acceleration with respect to ΛCDM in our $f(T)$ gravity models dominates over the stronger gravitational strength in the clustering of matter. We check that both models can easily pass the BBN constraints (which demand[19] $|G_{\rm eff}/G_N - 1| \leq 0.2$), as well as the ones from the Solar System [which demand[20] $|G'_{\rm eff}(z=0)/G_N| \leq 10^{-3}h^{-1}$ and $|G''_{\rm eff}(z=0)/G_N| \leq 10^{5}h^{-2}$].

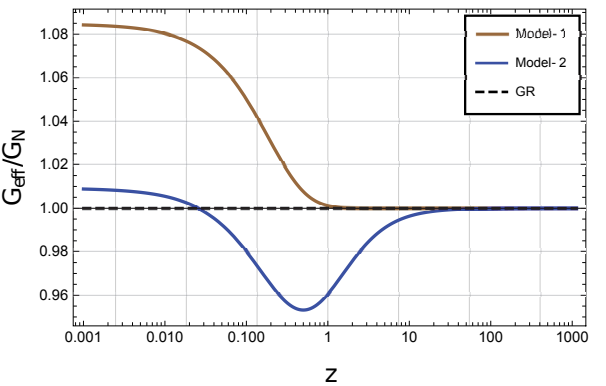


Fig. 1. Redshift evolution of $G_{\rm eff}/G_N$ in Model-1 (brown solid line) and Model-2 (blue solid line) and their comparison to the GR case (black dashed line).

Now we show how Model-1 and Model-2 can alleviate the H_0 and σ_8 tension by solving the background and perturbation equations. In Fig. 2 we present the evolution of $H(z)$, while in Fig. 3 the evolution of $f\sigma_8$, for two $f(T)$ models, and we compare them with ΛCDM. We stress that the H_0 tension can be alleviated as $H(z)$ remains statistically consistent for all CMB and CC measurements at all redshifts. We remind the reader that the two $f(T)$ models differing merely by a term $\propto T^{1/2}$, which does not affect the background as explained before, are degenerate at the background level. However, at the perturbation level, the two models behave differently as their gravitational coupling $G_{\rm eff}$ differs.

We further stress that both models can alleviate the σ_8 tension, and fit efficiently to BAO and LSS measurements. Note that at high redshifts ($z \geq 2$), Model-2 approaches ΛCDM slower than Model-1, but in a way that is statistically indistinguishable for present-day data. Nevertheless, future high-redshift surveys such as eBOSS for quasars and Euclid[24] for galaxies have the potential to discriminate among the predictions of $f(T)$ gravity and the ΛCDM scenario. Moreover, the clusters and CMB measurements on σ_8 are in good agreement in our models, as the CMB preferred values in ΛCDM get further lowered than local ones from rescaling

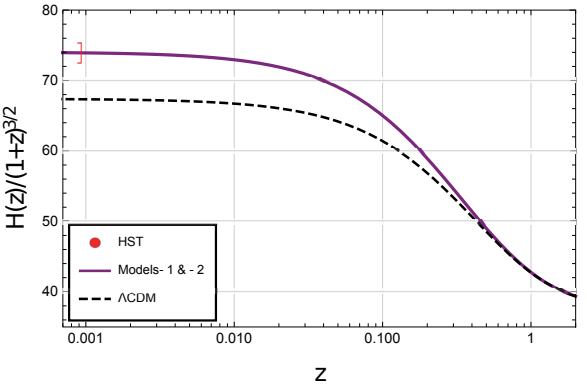


Fig. 2. Evolution of the Hubble parameter $H(z)$ in the two $f(T)$ models (purple solid line) and in ΛCDM cosmology (black dashed line). The red point represents the latest data from extragalactic Cepheid-based local measurement of H_0 provided in.[6]

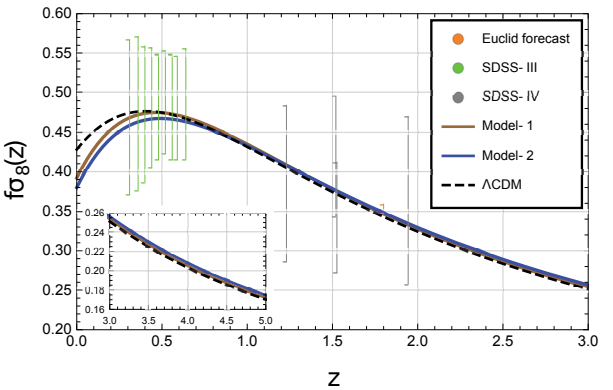


Fig. 3. Evolution of $f\sigma_8$ in Model-1 (brown solid line) and Model-2 (blue solid line) of $f(T)$ gravity and in ΛCDM cosmology (black dashed line). The green data points are from BAO observations in SDSS-III DR12,[21] the gray data points at higher redshift are from SDSS-IV DR14,[22] while the red point around ~ 1.8 is the forecast from Euclid.[23] The subgraph in the left bottom displays $f\sigma_8$ at high redshift $z = 3 \sim 5$, which shows that the curve of Model-2 is above the one of Model-1 and ΛCDM scenario and hence approaches ΛCDM slower than Model-1.

σ_8 by the ratio of the growth factors in $f(T)$ gravity and ΛCDM. More explicitly,

$$\sigma_8^{f(T)}(z=0) = \frac{D^{f(T)}(z=0)}{D^{\Lambda}(z=0)} \frac{D^{\Lambda}(z_{\rm eff})}{D^{f(T)}(z_{\rm eff})} \sigma_8^{\Lambda}(z=0), \tag{13}$$

where $D(z)$ is the growth factor, $f(T)$ and Λ denote our models and ΛCDM respectively, and $z_{\rm eff}$ is the effective redshift of the measurements ($z_{\rm eff} \sim 0.1$ for clusters experiments and $z_{\rm eff} \sim 1100$ for CMB temperature fluctuations observations). It turns out that, as at high redshift, $z_{\rm eff} \sim 1100$, the growth factor is the same in either our $f(T)$-models or in ΛCDM, but at low redshift, $z_{\rm eff} \sim 0.1$, the growth factor is approximately 1.03 bigger in the later compared to the formers, cluster

σ_8-measurements get bigger by about such an amount reducing the gap with CMB preferred value in this modified gravity scenario.

In short summary, we conclude that the class of $f(T)$ gravity: $f(T) = -T - 2\Lambda/M_P^2 + \alpha T^\beta$, where only two out of the three parameters Λ, α, and β are independent (the third one is eliminated using Ω_{m0}), can alleviate both H_0 and σ_8 tensions with suitable parameter choices. Moreover, such kinds of models in $f(T)$ gravity could also be examined through galaxy-galaxy lensing effects,[25] strong lensing effects around black holes[26] and gravitational wave experiments.[15]

5. Conclusions

In this work we reported how $f(T)$ gravity can alleviate both H_0 and σ_8 tensions simultaneously. Working within the EFT framework, torsional gravity theories can be identified as the EFT operators that allow us to extract the evolution equations of the background and of the perturbations in a model-independent manner. This allows us to address in a systematic way how tensions amongst the observational measurements, such as the ones on H_0 and σ_8, can be relaxed. Following these considerations, we constructed concrete models from specific Lagrangians in $f(T)$ gravity, describing cosmological scenarios where these tensions fade away. Imposing initial conditions at the last scattering that reproduce the ΛCDM scenario, and imposing the late-time values preferred by local measurements, we reconstructed two particular forms of $f(T)$. These models are well described by the parametrization: $f(T) = -T - 2\Lambda/M_P^2 + \alpha T^\beta$. To our knowledge, this is the first time where both H_0 and σ_8 tensions are simultaneously alleviated by a modified gravity theory.

References

1. E. J. Copeland, M. Sami and S. Tsujikawa, *Dynamics of dark energy*, Int. J. Mod. Phys. D **15**, 1753-1936 (2006).
2. Y. F. Cai, E. N. Saridakis, M. R. Setare and J. Q. Xia, *Quintom Cosmology: Theoretical implications and observations*, Phys. Rept. **493**, 1-60 (2010)
3. E. N. Saridakis *et al.* [CANTATA], *Modified Gravity and Cosmology: An Update by the CANTATA Network*, [arXiv:2105.12582 [gr-qc]].
4. S. Capozziello and M. De Laurentis, *Extended Theories of Gravity*, Phys. Rept. **509**, 167-321 (2011).
5. N. Aghanim *et al.* [Planck Collaboration], *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641**, A6 (2020) [erratum: Astron. Astrophys. **652**, C4 (2021)].
6. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, *Large Magellanic Cloud Cepheid Standards Provide a 1Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM*, Astrophys. J. **876**, no. 1, 85 (2019).
7. K. C. Wong *et al.*, *H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes*, Mon. Not. Roy. Astron. Soc. **498**, no.1, 1420-1439 (2020).
8. S. Alam *et al.* [BOSS Collaboration], *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, Mon. Not. Roy. Astron. Soc. **470**, no. 3, 2617 (2017).

9. M. Ata *et al.*, *The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2*, Mon. Not. Roy. Astron. Soc. **473**, no. 4, 4773 (2018).
10. E. Di Valentino, L. A. Anchordoqui, O. Akarsu, Y. Ali-Haimoud, L. Amendola, N. Arendse, M. Asgari, M. Ballardini, S. Basilakos and E. Battistelli, *et al. Snowmass2021 - Letter of interest cosmology intertwined II: The hubble constant tension*, Astropart. Phys. **131**, 102605 (2021).
11. E. Di Valentino, L. A. Anchordoqui, Ö. Akarsu, Y. Ali-Haimoud, L. Amendola, N. Arendse, M. Asgari, M. Ballardini, S. Basilakos and E. Battistelli, *et al. Cosmology intertwined III: $f\sigma_8$ and S_8*, Astropart. Phys. **131**, 102604 (2021).
12. S. F. Yan, P. Zhang, J. W. Chen, X. Z. Zhang, Y. F. Cai and E. N. Saridakis, *Interpreting cosmological tensions from the effective field theory of torsional gravity*, Phys. Rev. D **101**, no.12, 121301 (2020).
13. N. Arkani-Hamed, H. C. Cheng, M. A. Luty and S. Mukohyama, *Ghost condensation and a consistent infrared modification of gravity*, JHEP **0405**, 074 (2004).
14. C. Li, Y. Cai, Y. F. Cai and E. N. Saridakis, *The effective field theory approach of teleparallel gravity, $f(T)$ gravity and beyond*, JCAP **1810**, 001 (2018).
15. Y. F. Cai, C. Li, E. N. Saridakis and L. Xue, *$f(T)$ gravity after GW170817 and GRB170817A*, Phys. Rev. D **97**, no. 10, 103513 (2018).
16. F. K. Anagnostopoulos, S. Basilakos and E. N. Saridakis, *Bayesian analysis of $f(T)$ gravity using $f\sigma_8$ data*, Phys. Rev. D **100**, no. 8, 083517 (2019).
17. Y. F. Cai, S. Capozziello, M. De Laurentis and E. N. Saridakis, *f(T) teleparallel gravity and cosmology*, Rept. Prog. Phys. **79**, no.10, 106901 (2016).
18. S. Nesseris, S. Basilakos, E. N. Saridakis and L. Perivolaropoulos, *Viable $f(T)$ models are practically indistinguishable from ΛCDM*, Phys. Rev. D **88**, 103010 (2013).
19. C. J. Copi, A. N. Davis and L. M. Krauss, *A New nucleosynthesis constraint on the variation of G*, Phys. Rev. Lett. **92**, 171301 (2004).
20. S. Nesseris and L. Perivolaropoulos, *The Limits of Extended Quintessence*, Phys. Rev. D **75**, 023517 (2007).
21. Y. Wang, G. B. Zhao, C. H. Chuang, M. Pellejero-Ibanez, C. Zhao, F. S. Kitaura and S. Rodriguez-Torres, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: a tomographic analysis of structure growth and expansion rate from anisotropic galaxy clustering*, Mon. Not. Roy. Astron. Soc. **481**, no. 3, 3160 (2018).
22. H. Gil-Marín *et al.*, *The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: structure growth rate measurement from the anisotropic quasar power spectrum in the redshift range* $0.8 < z < 2.2$, Mon. Not. Roy. Astron. Soc. **477**, no. 2, 1604 (2018).
23. L. Taddei, M. Martinelli and L. Amendola, *Model-independent constraints on modified gravity from current data and from the Euclid and SKA future surveys*, JCAP **1612**, 032 (2016).
24. R. Laureijs *et al.* [EUCLID Collaboration], *Euclid Definition Study Report*, arXiv:1110.3193 [astro-ph.CO].
25. Z. Chen, W. Luo, Y. F. Cai and E. N. Saridakis, *New test on General Relativity using galaxy-galaxy lensing with astronomical surveys*, Phys. Rev. D **102**, no. 10, 104044 (2020).
26. S. Yan, C. Li, L. Xue, X. Ren, Y. F. Cai, D. A. Easson, Y. Yuan and H. Zhao, *Testing the equivalence principle via the shadow of black holes*, Phys. Rev. Res. **2**, no. 2, 023164 (2020).

Precision Cosmology and Hubble tension in the era of LSS surveys

G. Fanizza

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edificio C8, Campo Grande, P-1740-016, Lisbon, Portugal
and
CERN, Theory Department, CH-1211 Geneva 23, Switzerland
E-mail: gfanizza@fc.ul.pt

We present a fully relativistic framework to evaluate the impact of stochastic inhomogeneities on the prediction of the Hubble-Lemaître diagram. In this regard, we relate the fluctuations of the luminosity distance-redshift relation in the Cosmic Concordance model to the intrinsic uncertainty associated to the estimation of cosmological parameters from high-redshift surveys (up to z = 4). Within this framework and according to the specific of forthcoming surveys as Euclid Deep Survey and LSST, we show that the cosmic variance associated with the measurement of the Hubble constant will not exceed 0.1 %. Thanks to our results, we infer that deep surveys will provide an estimation of the the Hubble constant H_0 which will be more precise than the one obtained from local sources, at least in regard of the intrinsic uncertainty related to a stochastic distribution of inhomogeneities.

Keywords: Cosmology; Large Scale Structure; Inhomogeneous Universe; Hubble tension

1. Introduction

Along this work, we will summarize the results obtained in Fanizza et al. (2021)[1] and presented at the parallel session *"Status of the H_0 and σ_8 Tensions: Theoretical Models and Model-Independent Constraints"* of the 16th Marcel Grossmann meeting. In particular, we will quantify the cosmic variance concerning the estimation of H_0 from forthcoming high redshift survey with limited sky coverage.

Indeed, the best-fit of the CMB data[2] allows us to infer the value of few cosmological parameters with the highest precision achievable so far in cosmology. Among these parameters, the estimation of H_0, namely the current expansion rate of the Universe, has gained great interest for the cosmologists because of an increasing tension emerging against the value of H_0 itself as estimated from local measurements.[3]

This tension is almost of 5σ, since CMB measurements provide[2] a value for $H_0 = 67.36 \pm 0.54$ km s^{-1} Mpc^{-1}, whereas late time estimations based on local probes, such as Supernovae Ia (SnIa), return[3] $H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1}.

Given this tension and working within the conservative framework of ΛCDM model, our main interest is to understand whether, in view of the forthcoming Large Scale Structure (LSS) surveys, there could be any theoretical bias increasing the standard deviation in order to alleviate the above-mentioned tension. A first attempt to address this point has been provided in Ben-Dayan et al. (2014),[4] where

the effect of velocity dispersion of local SnIa (at redshift lower than 0.1) has been studied and it has been shown that it can introduce a further intrinsic error in the local estimation of H_0 of $\sim 1\%$. This result is in agreement with subsequent analysis,[5] leading to an effect which is large but unfortunately not enough to resolve the tension. Our goal here is to extend the analysis to higher redshifts ($z \leq 3.85$), including then also lensing effects to the analysis, in order to forecast the estimated precision for forthcoming surveys, just like Euclid Deep Survey (EDS) and LSST. To this aim, we will assume that the new generation of standard candles known as Superluminous Supernovae (SLSNe) will provide a suitable dataset, following what has been recently claimed in Inserra et al. (2021).[6]

In Sect. 2 we present the estimator for the cosmic variance in realistic high redshift surveys. In Sect. 3, we recall and discuss the analytic tool, such as the 2-point correlation function of the luminosity distance-redshift relation and its monopole, needed for our estimation. In Sect. 4, we generate a survey according to the specific of forthcoming surveys and forecast the expected cosmic variance. In Sect. 5 our results are summarized and discussed.

2. Local measurements

In order to extend the analysis of Ben-Dayan et al. (2014)[4] to higher redshift, we first need to go beyond the linear relation between luminosity distance and redshift $d_L = z/H_0$. To this aim, we then refer to the luminosity distance-redshift relation $d_L(z)$ in the homogenous and isotropic ΛCDM model

$$d_L(z) = \frac{1+z}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_{m0}(1+z')^3 + 1 - \Omega_{m0}}}, \tag{1}$$

where H_0 is indeed the Hubble constant and Ω_{m0} is the energy density of the matter today. Eq. (1) then provides an estimator for H_0 at higher redshifts as

$$H_0 = \frac{1+z}{d_L(z)} \int_0^z \frac{dz'}{\sqrt{\Omega_{0m}(1+z')^3 + 1 - \Omega_{0m}}}\,. \tag{2}$$

By assuming then *a priori* the value of Ω_{m0} and given that d_L and z can be independently observed, a measurement of H_0 can be directly inferred from the high-redshift surveys thanks to Eq. (2).

Eq. (2) provides also the starting point to estimate the intrinsic uncertainty associated to H_0 given by the presence of inhomogeneities all around our observed Universe. This is the lowest theoretical uncertainty we can reach, according to the sample of sources we have access to and is usually named *cosmic variance*. We hence have that, if we take into account also inhomogeneities in the Universe, the luminosity distance-redshift relation is modified accordingly as

$$\widetilde{d_L}(z, \mathbf{n}) = d_L(z) \left[1 + \delta^{(1)}(z, \mathbf{n}) + \delta^{(2)}(z, \mathbf{n})\right], \tag{3}$$

where $\mathbf{n}$ is the observed direction of the given source and $\delta^{(1)}$ and $\delta^{(2)}$ are linear and second order corrections to the luminosity distance-redshift relation.

At this point, the inhomogeneous value of the Hubble constant $\widetilde{H_0}$ can be evaluated as

$$\widetilde{H_0} \equiv \frac{1+z}{\widetilde{d_L}(z)} \int_0^z \frac{dz'}{\sqrt{\Omega_{0m}(1+z')^3 + 1 - \Omega_{0m}}} = H_0 \left[1 - \delta^{(1)} - \delta^{(2)} + \left(\delta^{(1)}\right)^2 \right](z, \mathbf{n}) \,. \tag{4}$$

We notice that $\widetilde{H_0}$ is now a function of z and $\mathbf{n}$. This is in line with the fact that inhomogeneities may render the estimation of H_0 biased by local structures. In principle, these inhomogeneities may affect not only the precision but also the accuracy of H_0 and then shift its averaged value within the sample. Here we neglect this effect since the correction to the Hubble diagram due to the non-linearities has been shown to be small in the regime of redshift of our interest.[7–9]

Because of these fluctuations, it is possibile to define an estimator for the variance associated to the background value of H_0 inferred from a finite survey of N sources observed at positions $(z_i, \mathbf{n}_i)$, where $i = 1, \ldots, N$ runs over the number of the sources. Indeed, the variance associated to the average value of $\widetilde{H_0}$ inferred from a finite survey of N sources will be then

$$\sigma^2_{H_0} = \overline{\left(\sum_{i=1}^{N} \frac{\widetilde{H_0}(z_i, \mathbf{n}_i)}{N} - H_0 \right) \left(\sum_{j=1}^{N} \frac{\widetilde{H_0}(z_j, \mathbf{n}_j)}{N} - H_0 \right)} = \frac{1}{N^2} \sum_{i,j=1}^{N} \overline{\left(\widetilde{H_0}(z_i, \mathbf{n}_i) \widetilde{H_0}(z_j, \mathbf{n}_j) - H_0^2 \right)} = \frac{H_0^2}{N^2} \sum_{i,j=1}^{N} \overline{\delta^{(1)}(z_i, \mathbf{n}_i)\, \delta^{(1)}(z_j, \mathbf{n}_j)} \,. \tag{5}$$

where $\overline{\cdots}$ represents the *ensemble average* over all the possibile configurations of perturbations. Last equality of Eq. (5) states that our estimator for the cosmic variance of H_0 precisely corresponds to the sum of the 2-point correlation function of the luminosity distance-redshift relation over all the possible pairs of sources in the survey. It then links the details of the surveys, such as angular and redshift distributions of the sources, to the theoretical expressions of the inhomogeneities. We also notice that nonlinear terms in Eq. (5) exactly cancel.

In the next section, these general preliminaries will be applied to the case of linear perturbations of the luminosity distance. This will provide the explicit expression for $\sigma^2_{H_0}$ due to all the linear relativistic corrections.

3. Analytical expressions for leading order effects

Eq. (5) explicitly shows that the cosmic variance $\sigma^2_{H_0}$ is sourced at the leading order only by linear perturbations. It is then enough to take into account the linear relativistic corrections involved in the $\delta^{(1)}$. By restricting our analysis to scalar

perturbations as given in the Longitudinal Gauge without anisotropic stress[a], the relativistic effects occurring at linear level are well-known:[12–15] we have lensing of the photon geodesics along the line-of-sight due to cosmic structures, peculiar velocity (also known as Doppler) due to the free falling motion of the sources within local inhomogeneities and then fluctuations of the gravitational potential around the source position and along the photon geodesics, leading to local and integrated Sachs-Wolfe effects and time delay. However, effects due to gravitational potentials are negligible with respect to lensing and peculiar velocity. We then focus the rest of our discussion only to the lensing and peculiar velocity.

The impact of the these effects on $\sigma^2_{H_0}$ is then given by

$$\frac{\sigma^2_{H_0}}{H_0^2} = \frac{1}{N^2}\sum_{i,j}\sum_{E,E'}\int \frac{dk}{k}\mathcal{P}_\psi(k)\mathcal{W}_{Ei,E'j}\,, \tag{6}$$

where $i,j = 1,\ldots,N$ and E and E' label the leading relativistic effects such as Lensing ($E,E' = L$) or Peculiar Velocity ($E,E' = PV$). $\mathcal{P}_\psi(k)$ in Eq. (6) is the so-called dimensionless power spectrum of gravitational perturbations in Fourier space

$$\mathcal{P}_\psi(k) \equiv \frac{k^3}{2\pi^2}|\psi_k|^2\,. \tag{7}$$

We remark that kernels $\mathcal{W}_{Ei,E'j}$ are functions of the (times) redshifts of the i-th and j-th sources (η_i) z_i and (η_j) z_j and their angular separation $\nu \equiv \mathbf{n}_i\cdot\mathbf{n}_j$. In this way, the only terms involved in our analysis are the auto-correlations of lensing and peculiar velocity

$$\begin{aligned}\mathcal{W}_{PVi,PVj} &= \Xi_i\Xi_j G_i G_j k^2\left\{\frac{\Delta\eta_i\,\Delta\eta_j(1-\nu^2)}{R^2}j_2(kR)\right.\\ &\left.+\frac{\nu}{3}\left[j_0\left(kR\right)-2j_2\left(kR\right)\right]\right\}(\eta_i,\eta_j,\nu)\\ \mathcal{W}_{Li,Lj} &= \frac{1}{\Delta\eta_i}\frac{1}{\Delta\eta_j}\int_{\eta_i}^{\eta_o}d\eta\frac{\eta-\eta_i}{\eta_o-\eta}\int_{\eta_j}^{\eta_o}d\eta'\frac{\eta'-\eta_j}{\eta_o-\eta'}\frac{g(\eta)g(\eta')}{g^2(\eta_o)}\left[k^4H^4j_4(kR)\right.\\ &\left.-8k^3H^2L\,j_3(kR)+k^2\left(8L^2-6H^2\right)j_2(kR)+4\,k\,L\,j_1(kR)\right](\eta,\eta',\nu)\,,\end{aligned} \tag{8}$$

and their cross-correlation

$$\begin{aligned}\mathcal{W}_{PVi,Lj} = \frac{\Xi_i}{\Delta\eta_j}G_i\int_{\eta_j}^{\eta_o}d\eta\frac{\eta-\eta_j}{\eta_o-\eta}\frac{g(\eta)}{g(\eta_o)}&\left\{-k^3\Delta\eta^2\frac{(\Delta\eta_i-\nu\Delta\eta)(\Delta\eta-\nu\Delta\eta_i)^2}{R^3}j_3(kR)\right.\\ &+\,k^2\Delta\eta\left(3\Delta\eta\frac{\Delta\eta_i-\nu\,\Delta\eta}{R^2}-2\nu\right)j_2(kR)\\ &\left.-\frac{k\Delta\eta}{R}\left[k^2\Delta\eta(\Delta\eta_i-\nu\Delta\eta)-2\nu\right]j_1(kR)\right\}(\eta_i,\eta,\nu)\,,\end{aligned} \tag{9}$$

[a]This assumption is in line with the fact that we take into account sources located after the decoupling. General expressions in presence of anisotropic stress can be found in Marozzi et al. (2014)[10] for the non-linear luminosity distance-redshift relation and in Fanizza et al. (2018)[11] for the non-linear redshift.

where $j_n(x)$ are the spherical Bessel functions of n-th order and we define

$$\begin{aligned} R(\eta_x,\eta_y,\nu) =& \sqrt{\Delta\eta_x^2+\Delta\eta_x^2-2\Delta\eta_x\Delta\eta_y\nu} \\ L(\eta_x,\eta_y,\nu) = \frac{\Delta\eta_x\,\Delta\eta_y\nu}{R\,(\eta_x,\eta_y,\nu)}\,, &\qquad H(\eta_x,\eta_y,\nu) = \frac{\Delta\eta_x\,\Delta\eta_y\sqrt{1-\nu^2}}{R\,(\eta_x,\eta_y,\nu)} \\ G_i = \int_{\eta_{in}}^{\eta_i} d\eta\,\frac{a(\eta)}{a(\eta_i)}\frac{g(\eta)}{g(\eta_o)}\,, &\qquad \Xi_i = \left(1-\frac{1}{\mathcal{H}_i\Delta\eta_i}\right)\,, \end{aligned} \tag{10}$$

with $\Delta\eta_x \equiv \eta_o - \eta_x$. In Eqs. (10), R is the Euclidean distance between two sources and L and H are respectively the normalized scalar and (modulo of the) vector products between the two directions of the sources. Moreover, G_i is the growth factor for the velocity potential of the i-th source, $g(\eta)$ is the growth factor of the gravitational potential, $a(\eta)$ is the scale factor, $\mathcal{H}(\eta) = a'(\eta)/a(\eta)$ and η_o is the value of the conformal time today. The detailed analytic derivation for all the different contributions $\mathcal{W}_{Ei,E'j}$ is reported in Fanizza et al. (2021).[1] We just remark that the kernel $\mathcal{W}_{PVi,PVj}$ is in agreement with the one found in Ben-Dayan et al. (2014).[4]

The values of the 2-point correlation function

$$\xi_{E,E'}(z_i,z_j,\nu) \equiv \int \frac{dk}{k}\mathcal{P}_\psi(k)\mathcal{W}_{Ei,E'j} \tag{11}$$

for the kernels in Eqs. (8) and (9) are then shown in Fig. 1 for the particular cases of aligned ($\nu = 1$) and anti-aligned ($\nu = -1$) sources. In the regime of interest for us ($0.15 \leq z \leq 3.85$), it is clear from these plots that lensing is always leading with respect to the other terms when the sources are aligned, whereas peculiar velocity turns out to be the leading effect for the anti-aligned sources at redshift smaller than 1.

Another interesting feature emerging from Eqs. (8) and (9) regards the angular average of $\xi_{E,E'}$. It turns out that

$$\int_{-1}^{1} d\nu\,\xi_{E,L}(z_i,z_j,\nu) = 0 \tag{12}$$

where E may label all the relativistic effects such as lensing itself, peculiar velocity, time delay and (integrated) Sachs-Wolfe effect. This means that, in the limit of full sky coverage surveys and large number of sources[b], lensing does not affects at all the estimation $\sigma^2_{H_0}$. Hence, it follows that in this ideal case the leading contribution to the cosmic variance is entirely addressed to the monopole of $\xi_{PV,PV}$. This can be easily evaluated from the angular integration of the $\mathcal{W}_{PVi,PVj}$ in Eq. (8) and results are shown in Fig. 2. In this ideal case, σ_{H_0} is entirely given by the doppler 2-point correlation function.

[b]See Yoo (2020)[16] for a detailed discussion of this ideal case.

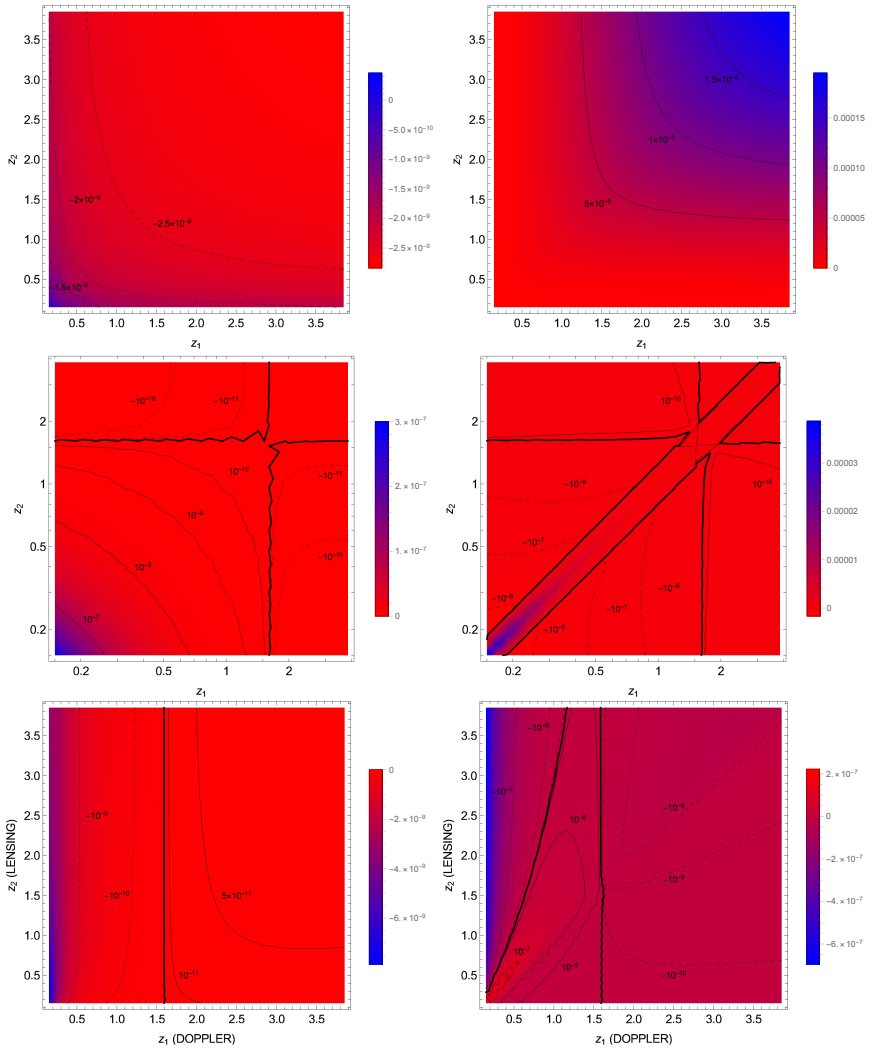


Fig. 1. 2-point correlation functions of Lensing (top panels), Peculiar Velocities (middle panels) and cross-correlation between Lensing and Peculiar Velocity (bottom panels). Left panels refer to anti-aligned sources ($\nu = -1$) wheres right ones consider aligned sources ($\nu = 1$). These plots are obtained by using linear power spectrum for the gravitational potential, following the cosmological parameters provided in Fanizza et al. (2021).[1]

However, realistic surveys deal with finite sky-coverages which can be also very narrow. In this case, monopole of the 2-point correlation functions of $\widetilde{d_L}(z, \mathbf{n})$ is no longer enough and higher multipoles must be taken into account for the evaluation of $\sigma^2_{H_0}$. We redirect the reader interested in the detailed multipoles analysis of $\xi_{E,E'}$ to Fanizza et al. (2021).[1] In the next section, we will apply the numerical results shown in Fig. 1 to the estimator for $\sigma^2_{H_0}$ in Eq. (6).

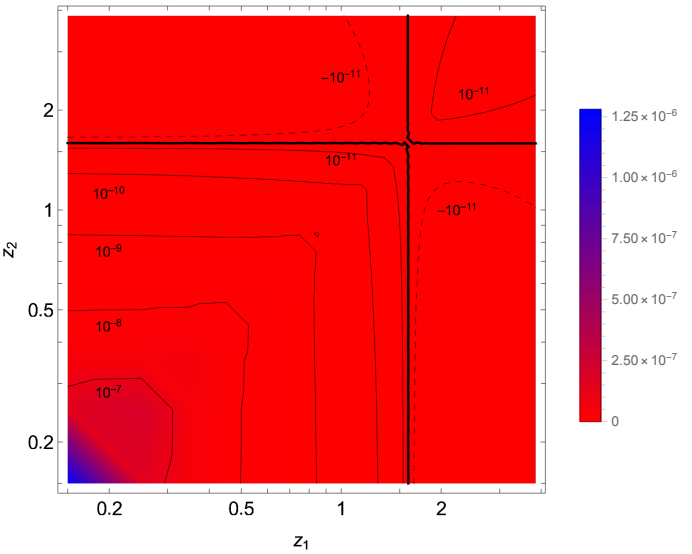


Fig. 2. Monopole of the doppler 2-point correlation function. Thick black lines indicate where function is 0. Dashed lines stand for negative values whereas continues lines refer to positive values.

4. Cosmic variance for next generation surveys

In this section, we apply the analytical and numerical results previously obtained to the case of Superluminous Supernovae (SLSNe). Following the specifics of forthcoming surveys like EDS[17] and LSST,[18] we consider the expected detection rate of SLSNe claimed in Inserra et al. (2021)[6] and reported in our Fig. 3. With these histograms in mind, we have generated two random surveys for the distribution of SLSNe with the following properties

- EDS: 135 sources in 7 redshift bins with $0.5 \leq z \leq 3.5$ and redshift bin width $\Delta z = 0.5$. For this survey, the angular distribution covers two

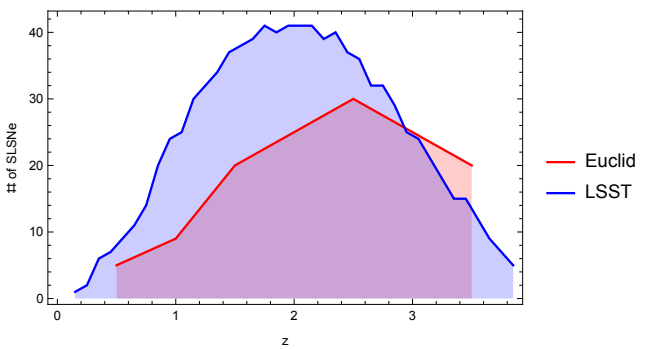


Fig. 3. Simulated SLSNe distributions for Euclid Deep Survey and LSST. From Inserra et al. (2021).[6]

line-of-sights at North and South Poles, with angular opening of 20 deg^2 per line-of-sight,

- LSST: 929 sources in 38 redshift bins with $0.15 \leq z \leq 3.85$ and redshift bin width $\Delta z = 0.1$. For this survey, the angular distribution spans a broad solid angle of 9000 deg^2.

Results are summarized in Table 1. Here we see that the dispersion associated to the measure of H_0, namely $\sigma_{H_0} \equiv \sqrt{\sigma^2_{H_0}}$ is of $\sim 0.2\%$ for EDS but its value drops of almost 1 order of magnitude for LSST, where it contributes with a dispersion of $\sim 0.03\%$. This difference is mainly addressed to the fact that, for larger sky coverage and large number of sources, the total effect due to lensing must tend to 0. The specific of LSST are indeed in line with regime: first of all, the number of sources adopted in our forecast of LSST is almost 1 order of magnitude higher than the one of EDS. Secondly, also the sky coverage is larger.

Table 1. Forecasts for the variance of H_0 in EDS and LSST. In the first line, lensing 2-point correlation function is considered. In the second line, the contributions from 2-point correlation function of peculiar velocities are shown. Values are reported for linear power spectrum truncated at $k_{UV} = 0.1h$ Mpc^{-1} and Non-Linear HaloFit model truncated at $k_{UV} = 10h$ Mpc^{-1}.

$\sigma^2_{H_0}/H_0^2$	EDS (Linear)	LSST (Linear)	EDS (Non-Linear)	LSST (Non-Linear)
Lensing	5.1×10^{-6}	7.6×10^{-8}	1.1×10^{-5}	7.8×10^{-7}
Doppler	2.1×10^{-9}	2.9×10^{-10}	-	-

Another fact that we underline is that doppler effect is always subdominant. Hence we have that the total cosmic variance due to lensing and doppler is

$$\sigma_{H_0} \approx \sigma_{H_0 L} \left[1 + \frac{1}{2} \frac{\sigma^2_{H_0 PV}}{\sigma^2_{H_0 L}} + \mathcal{O}\left(\left(\frac{\sigma^2_{H_0 PV}}{\sigma^2_{H_0 L}} \right)^2 \right) \right] . \tag{13}$$

Again from Table 1, we then get that the doppler effect corrects the total cosmic variance associated to H_0 by 0.2% for LSST and by 0.02% for EDS.

This analytical estimation can be naturally extended to $k_{UV} = 10\,h$ Mpc^{-1} where the HaloFit model[19,20] is taken into account to model the non-linear scales. Results for the lensing are reported again in Table 1 and show that $\sigma^2_{H_0\,NL} = 1.1 \times 10^{-5}\, H_0^2$ for EDS and $\sigma^2_{H_0\,NL} = 7.8 \times 10^{-7}\, H_0^2$ for LSST.

It follows then, for EDS, even if non-linear scales are expected to enhance the lensing correction by almost one order of magnitude, the intrinsic error associated to the measurement of H_0 is almost insensitive to the non-linear scales, since it becomes $\sigma_{H_0\,NL}/H_0 = 0.003$. On the contrary, non-linear scales increase by roughly a factor 3 the dispersion of H_0 within the specific of LSST, raising σ_{H_0} to the value $\sigma_{H_0\,NL}/H_0 = 0.0009$.

Our analysis, provides also a test for the claim done in Ben-Dayan et al. (2014)[4] about small redshift surveys: it has been stated there that the analysis is insensitive

to smaller scales fluctuations due to the incoherence of such contributions. According to our investigation, we conclude that this is a reasonable expectation only for EDS. We address this feature to the fact that EDS covers smaller regions in the sky with higher angular density of sources.

5. Summary and conclusions

In this work, we have studied the impact of cosmological inhomogeneities on the estimation of H_0 from the high redshift Hubble diagram. Our analysis considers the possibility, discussed in Inserra et al. (2021),[6] that EDS[17] and LSST[18] will have access to a statistically relevant number of Superluminous Supernovae in the next decade. In this regards, less conservative studies about the Hubble diagram at high redshifts ($z \leq 1.5$) have been also investigated by exploiting exact inhomogeneous models in general relativity[21–24] or by considering strongly inhomogeneous dynamical dark energy models.[25] These attempts look interesting especially in light of recent analysis done in Krishnan et al. (2020)[26] and Krishnan et al. (2021)[27] about H0LiCOW[28] and TDCOSMO[29] data and in Dainotti et al. (2021)[30] for the SNe Ia Pantheon sample, suggesting that H_0 could be a decreasing function of redshift already at late time.

However, our approach is more conservative, since it is based on linear perturbations within the Cosmic Concordance model. In this framework, the 2-point correlation function of luminosity distance-redshift relation has been analytically derived and investigated numerically. It turns out that lensing is the leading effect at the considered redshift, as one may expect.

Our first analytical estimations of the cosmic variance for limited-sky-coverage surveys indicate then that forthcoming high-redshift surveys are well-suited to provide a precise determination of cosmological parameters, such as H_0. Moreover, these forecasted errors for LSST and EDS are stable enough to be quite insensitive to the role of non-linearities in the matter power spectrum. Despite our analysis has been performed entirely within the ΛCDM model, from the geometrical structure of the light-cone 2 general features emerge:

- Lensing 2-point correlation function has vanishing monopole, hence in the limit of large sky coverage and huge number of sources, the cosmic variance must be dominated by Doppler effect also on high redshift surveys.
- As a consequence of line-of-sight integration and statistic isotropy, non-linear scales play an important role only when two sources are almost aligned.

However, realistic surveys deal with limited sky coverage and this makes lensing contribution no longer vanishing. In fact, according to the specific of EDS and LSST and to what has been claimed in Inserra et al. (2021),[6] we forecast that the intrinsic error from cosmic variance associated to H_0 is of $\sim 0.03\,\%$ for LSST and $0.3\,\%$ for EDS for the linear power spectrum. Non-linear scales contribute marginally to this

estimation within the specific of EDS. For what regard the specific of LSST, the situation is a way worse. Indeed, in this case, we get that non-linear scales enhance our forecast by almost a factor 3. This is a direct consequence of the fact that lensing 2-point correlation function strongly depends on small scales fluctuations for the diagonal entries of the covariance matrix.

Our analysis extends the one performed in Ben-Dayan et al. (2014),[4] where only close Supernovae (up to $z = 0.1$) have been considered. Indeed, we took into account also lensing corrections on top of the peculiar motion of the sources. In fact, the former is the leading relativistic effect expected at those redshifts.[7] Interestingly, our analysis points out that surveys have an intrinsic error for H_0 which tends to decrease when higher redshift sources are considered, whereas low redshift surveys discussed in Ben-Dayan et al. (2014)[4] admits a quite high dispersion for H_0 of $\sim 1\,\%$.

On one hand, our results are not able to alleviate the tension between local and distant measurements of the Hubble constant. However, they indicate that the analysis of fainter sources does not increase the theoretical uncertainty on H_0. The price to pay stands in the fact that the Hubble diagram at higher redshift is no longer model independent.

References

1. G. Fanizza, B. Fiorini and G. Marozzi, Cosmic variance of H0 in light of forthcoming high-redshift surveys, *Phys. Rev. D* **104**, p. 083506 (2021).
2. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
3. A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn and D. Scolnic, Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM, *Astrophys. J. Lett.* **908**, p. L6 (2021).
4. I. Ben-Dayan, R. Durrer, G. Marozzi and D. J. Schwarz, The value of H_0 in the inhomogeneous Universe, *Phys. Rev. Lett.* **112**, p. 221301 (2014).
5. D. Camarena and V. Marra, Impact of the cosmic variance on H_0 on cosmological analyses, *Phys. Rev. D* **98**, p. 023537 (2018).
6. C. Inserra *et al.*, The first Hubble diagram and cosmological constraints using superluminous supernovae, *Mon. Not. Roy. Astron. Soc.* **504**, 2535 (2021).
7. I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier and G. Veneziano, Average and dispersion of the luminosity-redshift relation in the concordance model, *JCAP* **1306**, p. 002 (2013).
8. P. Fleury, C. Clarkson and R. Maartens, How does the cosmic large-scale structure bias the Hubble diagram?, *JCAP* **03**, p. 062 (2017).
9. G. Fanizza, M. Gasperini, G. Marozzi and G. Veneziano, Generalized covariant prescriptions for averaging cosmological observables, *JCAP* **02**, p. 017 (2020).
10. G. Marozzi, The luminosity distance–redshift relation up to second order in the Poisson gauge with anisotropic stress, *Class. Quant. Grav.* **32**, p. 045004 (2015), [Erratum: Class.Quant.Grav. 32, 179501 (2015)].
11. G. Fanizza, J. Yoo and S. G. Biern, Non-linear general relativistic effects in the observed redshift, *JCAP* **09**, p. 037 (2018).

12. C. Bonvin, R. Durrer and M. Gasparini, Fluctuations of the luminosity distance, *Phys. Rev. D* **73**, p. 023523 (2006), [Erratum: Phys.Rev.D 85, 029901 (2012)].
13. I. Ben-Dayan, G. Marozzi, F. Nugier and G. Veneziano, The second-order luminosity-redshift relation in a generic inhomogeneous cosmology, *JCAP* **11**, p. 045 (2012).
14. G. Fanizza, M. Gasperini, G. Marozzi and G. Veneziano, A new approach to the propagation of light-like signals in perturbed cosmological backgrounds, *JCAP* **08**, p. 020 (2015).
15. O. Umeh, C. Clarkson and R. Maartens, Nonlinear relativistic corrections to cosmological distances, redshift and gravitational lensing magnification. II - Derivation, *Class. Quant. Grav.* **31**, p. 205001 (2014).
16. J. Yoo, Maximum Cosmological Information from Type-Ia Supernova Observations, *Phys. Rev. D* **101**, p. 043507 (2020).
17. R. Laureijs *et al.*, Euclid Definition Study Report (2011).
18. P. A. Abell *et al.*, LSST Science Book, Version 2.0 (12 2009).
19. R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou and H. M. P. Couchman, Stable clustering, the halo model and non-linear cosmological power spectra, *Monthly Notices of the Royal Astronomical Society* **341**, 1311 (Jun 2003).
20. R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, Revising the halofit model for the nonlinear matter power spectrum, *The Astrophysical Journal* **761**, p. 152 (Dec 2012).
21. L. Cosmai, G. Fanizza, M. Gasperini and L. Tedesco, Discriminating different models of luminosity-redshift distribution, *Class. Quant. Grav.* **30**, p. 095011 (2013).
22. A. E. Romano, Hubble trouble or Hubble bubble?, *Int. J. Mod. Phys. D* **27**, p. 1850102 (2018).
23. L. Cosmai, G. Fanizza, F. Sylos Labini, L. Pietronero and L. Tedesco, Fractal universe and cosmic acceleration in a Lemaître–Tolman–Bondi scenario, *Class. Quant. Grav.* **36**, p. 045007 (2019).
24. S. A. Vallejo-Peña and A. E. Romano, Coordinate independent approach to the calculation of the effects of local structure on the luminosity distance, *JCAP* **03**, p. 023 (2020).
25. R.-G. Cai, Z.-K. Guo, L. Li, S.-J. Wang and W.-W. Yu, Chameleon dark energy can resolve the Hubble tension, *Phys. Rev. D* **103**, p. 121302 (2021).
26. C. Krishnan, E. O. Colgáin, Ruchika, A. A. Sen, M. M. Sheikh-Jabbari and T. Yang, Is there an early Universe solution to Hubble tension?, *Phys. Rev. D* **102**, p. 103525 (2020).
27. C. Krishnan, E. O. Colgáin, M. M. Sheikh-Jabbari and T. Yang, Running Hubble Tension and a H0 Diagnostic, *Phys. Rev. D* **103**, p. 103509 (2021).
28. K. C. Wong *et al.*, H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, *Mon. Not. Roy. Astron. Soc.* **498**, 1420 (2020).
29. M. Millon *et al.*, TDCOSMO. I. An exploration of systematic uncertainties in the inference of H_0 from time-delay cosmography, *Astron. Astrophys.* **639**, p. A101 (2020).
30. M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi and G. Lambiase, On the Hubble constant tension in the SNe Ia Pantheon sample, *Astrophys. J.* **912**, p. 150 (2021).

Primordial black holes arise when the inflaton falls

Keisuke Inomata
Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA
E-mail: inomata@uchicago.edu

Primordial black holes (PBHs) can be produced when large density perturbations enter the horizon in the early universe. They can be dark matter (DM) and the black holes detected by the LIGO-Virgo collaborations. In this proceeding, we show that the large enhancement of the perturbations, required for the DM PBHs and LIGO/Virgo PBHs, can be realized in inflation models with a downward step. This enhancement mechanism is related to the particle production associated with the non-adiabatic evolution of the inflaton. This proceeding is based on our original paper.[1]

Keywords: Primordial black holes; inflaton; dark matter.

1. Introduction

Primordial black holes (PBHs) are one of the hottest topics in Cosmology because they are one of the candidates of dark matter (DM) and the black holes (BHs) detected by LIGO-Virgo collaborations.[2–7] PBHs are produced when the large density perturbations enter the horizon during a radiation- or matter-dominated era. The power spectrum of the density perturbations sensitively depends on the inflaton potential.

Throughout this proceeding, we focus on single-field inflation models. In the previous works, the large perturbations for the PBH scenarios are realized by the flat region in the inflaton potential.[8–25] When the inflaton rolls on the flat region, it gets decelerated due to the Hubble friction, which is called the ultra-slow-roll (USR) period. Once the inflaton gets decelerated, the perturbations are enhanced. This can be seen from the form of the power spectrum in the slow-roll approximation, $\mathcal{P}_\mathcal{R} \simeq H^2/(8\pi^2\epsilon M_{\rm Pl}^2)$, where H is the Hubble parameter and ϵ is the slow-roll parameter, related to the inflaton velocity through $\epsilon \equiv -\dot{H}/H^2 = \dot{\phi}^2/(2H^2M_{\rm Pl}^2)$.

On the other hand, in this proceeding, we show that a downward step in the inflaton potential can also realize the large enhancement of the perturbations. Although the perturbation enhancement with the step-like feature is discussed in previous works,[26,27] the $\mathcal{O}(10^7)$ enhancement of $\mathcal{P}_\mathcal{R}$, required for the PBH scenarios, has never been realized. One example of the inflaton potential with the downward step is shown in Fig. 1, whose concrete parameterization will be given in Sec. 3. In the following, we explain why the downward step can enhance the perturbations.

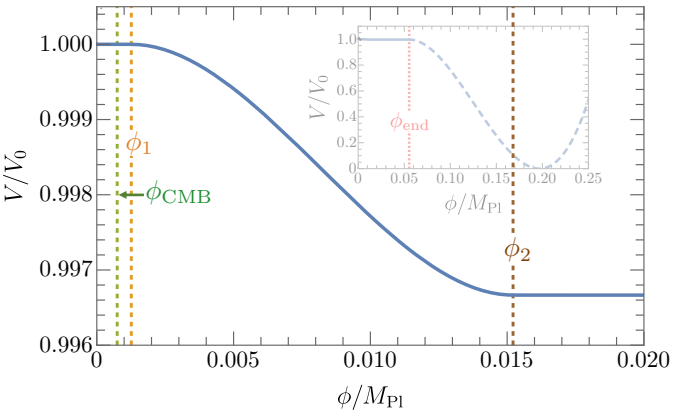


Fig. 1. The inflaton potential of Eq. (10) that realizes the large enhancement of perturbations, with the downward step at $\phi_1 \leq \phi \leq \phi_2$ highlighted and an inset for the full range. The parameters are $n_s = 0.97$, $\epsilon_1 = 7.43 \times 10^{-10}$, $\epsilon_2 = 0.01$, $\epsilon_3 = 10^{-9}$, and $\Delta N_{\rm step} = 0.5$. $\phi_{\rm end}$ denotes the end of inflation (red vertical dotted line) and corresponds to 50 e-folds from $\phi_{\rm CMB}$. This figure is taken from our original paper.[1]

2. A toy model

In this section, we explain the reason why the downward step causes the large enhancement by using a toy model. In canonical single-field inflation, the curvature perturbation $\mathcal{R}_k$ obeys the following equation:[26,28,29]

$$\mathcal{R}_k'' + (2+\eta)\mathcal{H}\mathcal{R}_k' + k^2\mathcal{R}_k = 0, \tag{1}$$

where $\eta \equiv \epsilon'/(\mathcal{H}\epsilon)$, $\mathcal{H}$ is the comoving Hubble parameter, and the prime denotes the derivative with respect to the conformal time, $d\tau \equiv dt/a$ with $\tau = 0$ at the end of inflation. We can rewrite this equation as

$$\frac{{\rm d}^2\mathcal{R}_k}{{\rm d}x^2} - \frac{2+\eta}{x}\frac{{\rm d}\mathcal{R}_k}{{\rm d}x} + \mathcal{R}_k \simeq 0, \tag{2}$$

where $x \equiv -k\tau$ and we have used the relation $\mathcal{H} \simeq -1/\tau$. Then, we obtain the general solution for the curvature perturbation:

$$\mathcal{R}_k \approx C_1 G_\nu^{(1)}(-k\tau) + C_2 G_\nu^{(2)}(-k\tau), \tag{3}$$

where $\nu = 3/2 + \eta/2$, C_1 and C_2 are constant in time, and $G_\nu^{(1)}$ and $G_\nu^{(2)}$ are defined with the Hankel functions of the first ($H_\nu^{(1)}$) and the second kind ($H_\nu^{(2)}$) as

$$G_\nu^{(j)}(-k\tau) \equiv (-k\tau)^\nu H_\nu^{(j)}(-k\tau), \tag{4}$$

with $j \in (1,2)$.

As a toy model, we consider the case where ϵ changes from $\epsilon_1 (\ll 1)$ to $\epsilon_2 (> \epsilon_1)$ with constant positive η and, after that, ϵ decreases with $\eta = -6$, which corresponds to a USR phase. Specifically, we parameterize η as

$$\eta = \eta_c \Theta(\tau - \tau_1)\Theta(\tau_2 - \tau) - 6\Theta(\tau - \tau_2). \tag{5}$$

Once ϵ_1/ϵ_2 is fixed, we have $\tau_2/\tau_1 = (\epsilon_1/\epsilon_2)^{1/\eta_c}$ and $\tau/\tau_2 = (\epsilon(\tau)/\epsilon_2)^{1/6}$ for $\tau > \tau_2$ with $\epsilon(\tau)$ being the value at τ. Substituting this into Eq. (2), we obtain the solution of the curvature perturbation,

$$\mathcal{R}_k = \begin{cases} D_1 G^{(1)}_{3/2}(-k\tau) & (\tau < \tau_1) \\ E_1 G^{(1)}_{\nu_c}(-k\tau) + E_2 G^{(2)}_{\nu_c}(-k\tau) & (\tau_1 \le \tau \le \tau_2) \\ F_1 G^{(1)}_{-3/2}(-k\tau) + F_2 G^{(2)}_{-3/2}(-k\tau) & (\tau_2 \le \tau) \end{cases},$$

where $\nu_c = 3/2 + \eta_c/2$. Imposing the Bunch-Davies vacuum condition as $D_1 = -\sqrt{\pi/2}H/(\sqrt{4k^3\epsilon_1}M_{\rm Pl})$ and the continuity of $\mathcal{R}$ and $\mathcal{R}'$ at τ_1 and τ_2, we determine the other coefficients. Hereafter, as the simplest case, let us consider the large limit of η_c, which corresponds to the step-function like downward step. In this limit, Eq. (1) in $\tau_1 < \tau < \tau_2$ can be approximated as

$$\mathcal{R}_k'' + \frac{\epsilon'}{\epsilon}\mathcal{R}_k' \approx 0. \tag{6}$$

From this relation, we can obtain $\mathcal{R}_k'(\tau_2) = (\epsilon_1/\epsilon_2)\mathcal{R}_k'(\tau_1)$. We can see that $\mathcal{R}_k'$ undergoes a jump down at the transition, by a relative factor of (ϵ_1/ϵ_2), while $\mathcal{R}$ itself is continuous, $\mathcal{R}_k(\tau_1) = \mathcal{R}_k(\tau_2)$, for any finite ϵ. From the relation between the curvature perturbation and the inflaton fluctuation $\delta\phi$ during the inflation, $\mathcal{R} \simeq \delta\phi/(\sqrt{2\epsilon}M_{\rm Pl})$, we can see that, while the curvature perturbations remain constant, the inflaton fluctuations get amplified during the increase of ϵ (the rolling down of the step). This amplification of the inflaton fluctuations can be regarded as the particle production. Note that the amplification of the curvature perturbations occurs after the rolling down phase ($\tau > \tau_2$) because the curvature perturbation evolves as $\mathcal{R}_k \propto 1/\sqrt{\epsilon}$ in $\tau > \tau_2$.

In the limit of $\eta_c \to \infty$, the coefficients F_1 and F_2 approach the following values:

$$\begin{aligned} F_{1,\rm lim} =& D_1 \frac{3\epsilon_2 + (2\epsilon_2+\epsilon_1)(-k\tau_1)^2 + i(\epsilon_2+\epsilon_1)(-k\tau_1)^3}{2\epsilon_2}, \\ F_{2,\rm lim} =& -D_1 {\rm e}^{-2ik\tau_1}(1 - i(-k\tau_1))\frac{(3\epsilon_2 - 3i\epsilon_2(-k\tau_1) + (\epsilon_1-\epsilon_2)(-k\tau_1)^2)}{2\epsilon_2}, \end{aligned} \tag{7}$$

where note $\tau_1 = \tau_2$ in this limit. Using these expressions, we derive the analytic form of the final power spectrum after the perturbations exit the horizon. Here, we define τ_3 as the time at the end of the USR phase so that $\epsilon_3 = \epsilon(\tau_3)$. After the USR phase, the inflation continues with $\epsilon = \epsilon_3$. In this case, the power spectrum in the large-scale limit, $k \ll 1/|\tau_1|$ becomes

$$\mathcal{P}_\mathcal{R}(k) \simeq \frac{H^2}{8\pi^2 M_{\rm Pl}^2\epsilon_1}\left[1 - \frac{2}{15}\sqrt{\frac{\epsilon_2}{\epsilon_3}}(-k\tau_1)^2 + \frac{1}{225}\frac{\epsilon_2}{\epsilon_3}(-k\tau_1)^4\right], \tag{8}$$

up to $\mathcal{O}((-k\tau_1)^4)$, and where we have also assumed $\epsilon_3/\epsilon_2 \ll 1$. On the other hand, for modes that were inside the horizon at the transition ($k \gtrsim 1/|\tau_1|$), the power spectrum finally becomes

$$\mathcal{P}_\mathcal{R}(k) \simeq A\frac{1-\cos(-2k\tau_1)}{2}, \quad A = \frac{H^2}{8\pi^2 M_{\rm Pl}^2\epsilon_1}\frac{\epsilon_2}{\epsilon_3}, \tag{9}$$

where we have assumed $\epsilon_1/\epsilon_2 \ll 1$. From this expression, we can see that the power spectrum is enhanced by a factor of $\mathcal{O}(\epsilon_2/\epsilon_3)$.

3. A concrete model

Next, we discuss a concrete inflaton potential with a downward step, which realizes the large perturbation enhancement. Specifically, we consider the following potential:

$$V(\phi) = V_b(\phi) F\left(\phi; \phi_1, \phi_2, h\right), \tag{10}$$

where V_b is the base potential and F induces the downward step. We take the following form of V_b:

$$V_b(\phi) \simeq V_0 \left(1 - \frac{\beta \phi^2/M_{\rm Pl}^2}{1 + \phi/\phi_{\rm CMB}}\right), \tag{11}$$

where we determine the parameters to make the power spectrum on large scales consistent with CMB observation, $\mathcal{P}_\mathcal{R} \simeq 2 \times 10^{-9}$ and $n_s = 0.97$. Note that the inflaton potential changes from this form around the end of the inflation, which is irrelevant to the enhancement mechanism. The concrete expression of the function F is given by

$$F(\phi; \phi_1, \phi_2, h) \equiv \begin{cases} 1 & (\phi < \phi_1) \\ 1 - hS\left(\frac{\phi - \phi_1}{\phi_2 - \phi_1}\right) & (\phi_1 \le \phi \le \phi_2) \\ 1 - h & (\phi > \phi_2) \end{cases}, \tag{12}$$

where $S(x) \equiv x^2(3 - 2x)$ and it changes from 0 to 1 with the change of x from 0 to 1. The parameter h determines the step height, normalized by V_0, and $\phi_2 - \phi_1$ determines the e-folds for the inflaton to pass through the step, which is denoted by $\Delta N_{\rm step}$. With this concrete inflaton potential, we numerically calculate the power spectrum and get the results in Fig. 2. The peak height of the power spectrum in Fig. 2 is $\mathcal{P}_\mathcal{R} \sim \mathcal{O}(10^{-2})$, which is large enough for the scenarios of the DM PBHs and the LIGO-Virgo PBHs.[30] From this figure, we can see that the power spectrum is enhanced by a factor of $\mathcal{O}(\epsilon_3/\epsilon_2)$, which is consistent with the analytical estimate in the previous section. We can also see that the perturbation enhancement does not occur on the very small scales, where Eq. (6) is invalid. This is because Eq. (6) is valid only when the transition timescale, $\tau_2 - \tau_1$, is much smaller than the timescale of the perturbation oscillation, $1/k$.

4. Conclusion

In this proceeding, we have proposed a new type of single-field inflation model for the PBH scenarios. We have shown that a downward step feature in the inflaton potential can enhance the power spectrum by a factor of $\mathcal{O}(10^7)$, required for the PBH scenarios. This perturbation enhancement can be interpreted as the particle production that occurs during the inflaton rolling down of the step.

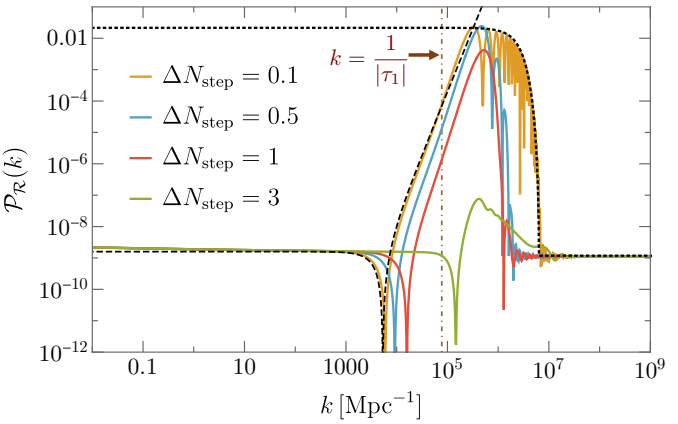


Fig. 2. The power spectrum in the inflaton potential of Eq. (10) with different $\Delta N_{\rm step}$. Except for $\Delta N_{\rm step}$, the parameters are the same as in Fig. 1. The black dotted and dashed lines are the approximate forms of the power spectrum. See our original paper[1] for details. This figure is taken from our original paper.[1]

Acknowledgments

The author was supported by the Kavli Institute for Cosmological Physics at the University of Chicago through an endowment from the Kavli Foundation and its founder Fred Kavli.

References

1. K. Inomata, E. Mcdonough and W. Hu, Primordial Black Holes Arise When The Inflaton Falls (4 2021).
2. S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli and A. G. Riess, Did LIGO detect dark matter?, *Phys. Rev. Lett.* **116**, p. 201301 (2016).
3. S. Clesse and J. García-Bellido, The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO, *Phys. Dark Univ.* **15**, 142 (2017).
4. M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914, *Phys. Rev. Lett.* **117**, p. 061101 (2016).
5. A. Kashlinsky, LIGO gravitational wave detection, primordial black holes and the near-IR cosmic infrared background anisotropies, *Astrophys. J. Lett.* **823**, p. L25 (2016).
6. A. Kashlinsky, R. G. Arendt, F. Atrio-Barandela, N. Cappelluti, A. Ferrara and G. Hasinger, Looking at cosmic near-infrared background radiation anisotropies, *Rev. Mod. Phys.* **90**, p. 025006 (2018).
7. J. García-Bellido, J. F. Nuño Siles and E. Ruiz Morales, Bayesian analysis of the spin distribution of LIGO/Virgo black holes, *Phys. Dark Univ.* **31**, p. 100791 (2021).
8. P. Ivanov, P. Naselsky and I. Novikov, Inflation and primordial black holes as dark matter, *Phys. Rev. D* **50**, 7173 (1994).

9. S. M. Leach and A. R. Liddle, Inflationary perturbations near horizon crossing, *Phys. Rev. D* **63**, p. 043508 (2001).
10. S. M. Leach, M. Sasaki, D. Wands and A. R. Liddle, Enhancement of superhorizon scale inflationary curvature perturbations, *Phys. Rev. D* **64**, p. 023512 (2001).
11. S. Inoue and J. Yokoyama, Curvature perturbation at the local extremum of the inflaton's potential, *Phys. Lett. B* **524**, 15 (2002).
12. N. C. Tsamis and R. P. Woodard, Improved estimates of cosmological perturbations, *Phys. Rev. D* **69**, p. 084005 (2004).
13. W. H. Kinney, Horizon crossing and inflation with large eta, *Phys. Rev. D* **72**, p. 023515 (2005).
14. J. Garcia-Bellido and E. Ruiz Morales, Primordial black holes from single field models of inflation, *Phys. Dark Univ.* **18**, 47 (2017).
15. J. M. Ezquiaga, J. Garcia-Bellido and E. Ruiz Morales, Primordial Black Hole production in Critical Higgs Inflation, *Phys. Lett. B* **776**, 345 (2018).
16. K. Kannike, L. Marzola, M. Raidal and H. Veermäe, Single Field Double Inflation and Primordial Black Holes, *JCAP* **1709**, p. 020 (2017).
17. C. Germani and T. Prokopec, On primordial black holes from an inflection point, *Phys. Dark Univ.* **18**, 6 (2017).
18. G. Ballesteros and M. Taoso, Primordial black hole dark matter from single field inflation, *Phys. Rev.* **D97**, p. 023501 (2018).
19. M. P. Hertzberg and M. Yamada, Primordial Black Holes from Polynomial Potentials in Single Field Inflation, *Phys. Rev. D* **97**, p. 083509 (2018).
20. S.-L. Cheng, W. Lee and K.-W. Ng, Superhorizon curvature perturbation in ultraslow-roll inflation, *Phys. Rev. D* **99**, p. 063524 (2019).
21. C. T. Byrnes, P. S. Cole and S. P. Patil, Steepest growth of the power spectrum and primordial black holes, *JCAP* **1906**, p. 028 (2019).
22. S. Passaglia, W. Hu and H. Motohashi, Primordial black holes and local non-Gaussianity in canonical inflation, *Phys. Rev. D* **99**, p. 043536 (2019).
23. M. Drees and Y. Xu, Overshooting, Critical Higgs Inflation and Second Order Gravitational Wave Signatures, *Eur. Phys. J. C* **81**, p. 182 (2021).
24. P. Carrilho, K. A. Malik and D. J. Mulryne, Dissecting the growth of the power spectrum for primordial black holes, *Phys. Rev. D* **100**, p. 103529 (2019).
25. K.-W. Ng and Y.-P. Wu, Constant-rate inflation: primordial black holes from conformal weight transitions (2 2021).
26. V. Miranda, W. Hu, C. He and H. Motohashi, Nonlinear Excitations in Inflationary Power Spectra, *Phys. Rev. D* **93**, p. 023504 (2016).
27. K. Kefala, G. P. Kodaxis, I. D. Stamou and N. Tetradis, Features of the inflaton potential and the power spectrum of cosmological perturbations, *Phys. Rev. D* **104**, p. 023506 (2021).
28. M. Sasaki, Large Scale Quantum Fluctuations in the Inflationary Universe, *Prog. Theor. Phys.* **76**, p. 1036 (1986).
29. V. F. Mukhanov, Quantum Theory of Gauge Invariant Cosmological Perturbations, *Sov. Phys. JETP* **67**, 1297 (1988).
30. M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, Primordial black holes—perspectives in gravitational wave astronomy, *Class. Quant. Grav.* **35**, p. 063001 (2018).

Effects of the modification of gravity on the production of primordial black holes

Sergio Andrés Vallejo Peña
ICRANet, Piazza della Repubblica 10, I–65122 Pescara, Italy
Instituto de Fisica, Universidad de Antioquia, A.A.1226, Medellin, Colombia
E-mail: sergio.vallejo@udea.edu.co

The enhancement of the spectrum of primordial comoving curvature perturbation $\mathcal{R}$ can induce the production of primordial black holes (PBH) which could account for part of present day dark matter. As an example of the effects of the modification of gravity on the production of PBHs, we investigate the effects on the spectrum of $\mathcal{R}$ produced by the modification of gravity in the case of G-inflation, deriving the relation between the unitary gauge curvature perturbation ζ and the comoving curvature perturbation $\mathcal{R}$, and identifying a background dependent enhancement function $\mathcal{E}$ which can induce large differences between the two gauge invariant variables.

When ζ is not constant in time it is different from $\mathcal{R}$, for example on sub-horizon scales, or in models exhibiting an anomalous super-horizon growth of ζ, but since this growth cannot last indefinitely, eventually they will coincide. We derive the general condition for super-horizon growth of ζ, showing that slow-roll violation is not necessary. Since the abundance of PBHs depends on the statistics of the peaks of the comoving density contrast, which is related to the spectrum of $\mathcal{R}$, it is important to take into account these effects on the PBHs abundance in modified gravity theories.

Keywords: Modified gravity, primordial curvature perturbations, primordial black holes.

1. Introduction

The study of primordial perturbations is fundamental in any cosmological model, since it allows to make predictions of the conditions which provided the seeds for the anisotropies of the cosmic microwave background (CMB) radiation or for the process of structure formation. Among the different theoretical scenarios proposed to explain the accelerated expansion of the Universe, Horndeski's theory[1] has received a lot of attention, both in the context of inflation and dark energy.

The calculation of the equation for cosmological perturbations for these theories have been so far performed in the so called unitary gauge, also known as uniform field gauge. While the unitary gauge has some computational convenience in general relativity when only a scalar field is present, in general it is not directly related to observations, which depend on the comoving curvature perturbations $\mathcal{R}$. The production of PBHs[2–6] is an example of phenomenon depending on $\mathcal{R}$[7] and not on the unitary gauge curvature perturbations ζ. Another example are the numerical codes developed for the solution of the Boltzman's equations in a perturbed Friedman-Lemaître-Robertson-Walker (FLRW) Universe, which are using equations in the synchronous gauge,[8] which for adiabatic perturbations coincides

approximately with the comoving gauge,[9] justifying the use of the comoving slices gauge for early Universe calculations.

The comoving gauge can differ from the unitary gauge in modified gravity theories because the effective energy momentum tensor arising from the modification of gravity can produce some effective entropy terms, which are absent in $K(X)$ theories, but are present in any more complicated Hordenski's theory. In this letter we compute the general relation between $\mathcal{R}$ and ζ in G-Inflation.[10,11] As an application we use this relation to investigate the effects of the modification of gravity on the power spectrum of $\mathcal{R}$, and its implications on the production of PBHs.

2. G-inflation

In G-inflation the scalar field Φ is minimally coupled to gravity according to the action[10,12]

$$S = \int d^4x\sqrt{-g}\left(\frac{M_{Pl}^2}{2}R + L(\Phi, X)\right),$$

where $X = -g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi/2$, R is the Ricci scalar and we use a system of units in which $c = \hbar = 1$. The Lagrangian density of the scalar field corresponds to

$$L(\Phi, X) = K(\Phi, X) + G(\Phi, X)\Box\Phi\,, \tag{1}$$

where K and G are arbitrary functions. The corresponding effective stress-energy-momentum tensor (EST) is given by

$$T_{\mu\nu} = L_{,X}\nabla_\mu\Phi\nabla_\nu\Phi + P_\Phi g_{\mu\nu} + \nabla_\mu\Phi\nabla_\nu G + \nabla_\nu\Phi\nabla_\mu G\,, \tag{2}$$

where

$$L_{,X} = \partial_X L = K_X(\Phi, X) + G_X(\Phi, X)\Box\Phi\,, \tag{3}$$

$$P_\Phi = L - \nabla_\mu\left(G\nabla^\mu\Phi\right) = K - g^{\mu\nu}\nabla_\mu\Phi\nabla_\nu G\,. \tag{4}$$

3. The perturbed effective energy-stress-momentum tensor

The most general scalar perturbations with respect to a flat FLRW background can be written as

$$ds^2 = a^2\Big\{-(1+2A)d\tau^2 + 2\partial_i B dx^i d\tau + \left[\delta_{ij}(1-2C) + 2\partial_i\partial_j E\right]dx^i dx^j\Big\}. \tag{5}$$

For the decomposition of the scalar field and the EST into their background and perturbation parts we use the notation

$$\Phi(x^\mu) = \phi(\tau) + \delta\phi(x^\mu)\,, \tag{6}$$

$$T^\mu{}_\nu = \overline{T}^\mu{}_\nu + \delta T^\mu{}_\nu\,. \tag{7}$$

The background components of the EST are

$$\overline{T}^0{}_0 = -\,\overline{\rho} = K(\phi,\chi) + \frac{3\mathcal{H}\phi'^3}{a^4}G_\chi(\phi,\chi) - \frac{\phi'^2}{a^2}\left[K_\chi(\phi,\chi) + G_\phi(\phi,\chi)\right]\,, \tag{8}$$

$$\overline{T}^0{}_i = \overline{T}^i{}_0 = 0\,, \tag{9}$$

$$\overline{T}^i{}_j = \delta^i{}_j\overline{P}\,,$$

$$\overline{P} = K(\phi,\chi) - \frac{\mathcal{H}\phi'^3}{a^4}G_\chi(\phi,\chi) + \frac{\phi'^2}{a^2}\left[G_\phi(\phi,\chi) + \frac{\phi''}{a^2}G_\chi(\phi,\chi)\right]\,, \tag{10}$$

where the primes stand for derivatives with respect to τ, χ is given by $\chi = \frac{\phi'^2}{2a^2}$, and the subscripts ϕ and χ denote partial derivatives with respect to these quantities, i.e. $G_\phi(\phi,\chi) = \partial_\phi G(\phi,\chi)$ and $G_\chi(\phi,\chi) = \partial_\chi G(\phi,\chi)$. In order to define the comoving slices gauge we need this component of the perturbed EST

$$\delta T^0{}_i = -\left(K_\chi + 2G_\phi - \frac{3\mathcal{H}\phi'}{a^2}G_\chi\right)\frac{\phi'^2}{a^2}\partial_i\delta\phi - \frac{\phi'^2}{a^4}G_\chi\partial_i\left(\delta\phi' - \phi' A\right)\,, \tag{11}$$

where $\mathcal{H} = a'/a$. The remaining components of the perturbed EST are not relevant to the computations done in this letter, and we will give them in a future work. Under a gauge transformation of the form $(\tau, x^i) \to (\tau + \delta\tau, x^i + \delta x,^i)$ the perturbations $\delta\phi$, A, B, C, and E transform according to[13]

$$\delta\phi \to \delta\phi - \phi'\delta\tau\,, \tag{12}$$

$$A \to A - \mathcal{H}\delta\tau - \delta\tau'\,, \tag{13}$$

$$B \to B + \delta\tau - \delta x'\,, \tag{14}$$

$$C \to C + \mathcal{H}\delta\tau\,, \tag{15}$$

$$E \to E - \delta x\,. \tag{16}$$

4. Evolution of curvature perturbations in the unitary gauge

In single scalar field models the unitary gauge is defined by the condition $\delta\phi_u = 0$. From the gauge transformation in eq.(12) we can see that the time translation $\delta\tau_u$ necessary to go to the unitary gauge is given by

$$\delta\tau_u = \frac{\delta\phi}{\phi'}\,. \tag{17}$$

Using eq.(15) we can compute the curvature perturbation in the unitary gauge ζ

$$\zeta \equiv -C_u = -C - \mathcal{H}\delta\tau_u = -C - \mathcal{H}\frac{\delta\phi}{\phi'}\,. \tag{18}$$

which is by construction gauge invariant. We can also define other gauge invariant quantities such as the unitary gauge lapse function

$$A_u \equiv A - \mathcal{H}\delta\tau_u - \delta\tau_u' = A - \mathcal{H}\frac{\delta\phi}{\phi'} - \left(\frac{\delta\phi}{\phi'}\right)'\,. \tag{19}$$

The second order action for ζ in Horndeski's theories was computed in[14]

$$S_\zeta^{(2)} = \int dt d^3x a^3 \left[\mathcal{G}_S \dot{\zeta}^2 - \frac{\mathcal{F}_S}{a^2} (\partial_i \zeta)^2 \right] , \tag{20}$$

where $\mathcal{G}_S$ and $\mathcal{F}_S$ are functions of $K(\phi, \chi)$ and $G(\phi, \chi)$ and their derivatives. The Lagrange equations for this action give the equation of motion of ζ

$$\zeta'' + \left(2\mathcal{H} + \frac{\mathcal{G}_S'}{\mathcal{G}_S} \right) \zeta' - c_s^2 \overset{(3)}{\Delta} \zeta = 0 , \tag{21}$$

where $c_s^2(\tau) = \mathcal{F}_S / \mathcal{G}_S$.

For the Fourier transform of the above equation we use the notation

$$\zeta_k'' + \left(2\mathcal{H} + \frac{\mathcal{G}_S'}{\mathcal{G}_S} \right) \zeta_k' + c_s^2 k^2 \zeta_k = 0 . \tag{22}$$

5. Comoving slices gauge in G-inflation

The comoving slices gauge is defined by the condition $\delta T^0{}_i = 0$. In G-inflation, combing eqs.(12-13) with eq.(11) we have that under an infinitesimal time translation

$$\delta T^0{}_i \to \delta T^0{}_i + \partial_i \left(\frac{\phi'^2}{a^4} D\delta\tau \right) , \tag{23}$$

where

$$D = a^2(2G_\phi + K_\chi) + G_\chi(-4\mathcal{H}\phi' + \phi'') , \tag{24}$$

from which we get the time translation $\delta\tau_c$ required to go to the comoving slices gauge

$$\delta\tau_c = \frac{1}{\phi' D} \Big[-\phi' G_\chi (3\mathcal{H}\delta\phi + \phi' A - \delta\phi') + a^2(2G_\phi + K_\chi)\delta\phi \Big] . \tag{25}$$

Note that in the particular case in which G does not depend explicitly on χ, i.e. $G(\phi, \chi) = G(\phi)$ the above transformation reduces to

$$\delta\tau_c = \frac{\delta\phi}{\phi'} , \tag{26}$$

and the comoving gauge coincides with the unitary gauge, since in this case the system is equivalent to a $K(X)$ theory.[15,16]

The comoving curvature perturbation $\mathcal{R}$ is then defined as

$$\mathcal{R} \equiv -C_c = -C - \mathcal{H}\delta\tau_c . \tag{27}$$

Our goal is to derive the relation between ζ and $\mathcal{R}$, and we can achieve this by performing the gauge transformation between the unitary and comoving slices gauge.

Using the general gauge transformation defined in eq.(25), when $\delta\phi = 0$ and $A = A_u$, we get

$$\delta\tau_{uc} = -\frac{\phi' G_\chi}{D} A_u , \tag{28}$$

from which we obtain

$$\mathcal{R} = \zeta + \mathcal{H}\frac{\phi' G_\chi}{D} A_u \,. \tag{29}$$

The gauge invariant variable A_u can be expressed in terms of ζ using the perturbed Einstein's equation $\delta G^0{}_i = \delta T^0{}_i / M_{Pl}^2$ in the unitary gauge, which using eq.(11) gives

$$-\zeta' + \mathcal{H} A_u = -\frac{\phi'^3 G_\chi}{2 M_{Pl}^2 a^2} A_u \,. \tag{30}$$

We can then combine eq.(29) and eq.(30) to obtain the relation between $\mathcal{R}$ and ζ only

$$\mathcal{R} = \zeta + \mathcal{H}\frac{\phi' G_\chi}{D}\left(\frac{\phi'^3 G_\chi}{2 M_{Pl}^2 a^2} + \mathcal{H}\right)^{-1} \zeta' = \zeta + \mathcal{E}(\tau)\zeta' \,, \tag{31}$$

where we have defined the enhancement factor $\mathcal{E}(\tau)$, a quantity depending only on the background, which can induce a significant difference between the curvature perturbations on comoving and uniform field slices. The relation between the power spectrum of ζ and $\mathcal{R}$ is then given by

$$P_\mathcal{R} = \frac{k^3}{2\pi^2}|\mathcal{R}_k|^2 = P_\zeta + \frac{k^3}{2\pi^2}\Delta \,, \tag{32}$$

where

$$\Delta = \left[\mathcal{E}\zeta^*\zeta' + \mathcal{E}^*\zeta'^*(\zeta + \mathcal{E}\zeta')\right] . \tag{33}$$

Note that the above relations are valid on any scale, since they are just based on gauge transformations, without assuming any sub or super horizon limit. This implies that the spectra of $\mathcal{R}$ and ζ could be different due to a change in the evolution of both sub-horizon and super-horizon modes during the time interval when $\mathcal{E}(\eta)$ is large. On sub-horizon scales the effect is always present, since ζ is oscillating and $\zeta' \neq 0$, while for super-horizon scales the effect could be suppressed if $\zeta \approx 0$, but even for models conserving ζ there could be an effect, since the freezing does not happen immediately after horizon crossing. We will discuss later the implication on the production of PBHs.

6. Conservation of $\mathcal{R}$ and ζ

From eq.(31) we can reach the important conclusion that

$$\zeta = const \Rightarrow \zeta = \mathcal{R} = const \,; \tag{34}$$

however the opposite is not true, i.e.

$$\mathcal{R} = const \nRightarrow \zeta = const \,, \tag{35}$$

which can have important implications for conservation laws of $\mathcal{R}$ and non-Gaussianity consistency conditions.[17] As explained previously, $\mathcal{R}$ is the quantity

related to observations, so it would be inconsistent to infer constraints on ζ from CMB observations for example, since the latter depend on $\mathcal{R}$. From a theoretical point of view the models approximately conserving ζ on super-horizon scales may be incompatible with observations for large enhancement functions $\mathcal{E}(\tau)$, because $\mathcal{R}$ could be not conserved, implying for example a violation of the non-Gaussianity consistency condition or a miss-estimation of PBHs abundance.

Nevertheless it should be noted that the super-horizon growth of perturbations cannot last indefinitely, or the entire perturbative treatment of the problem would breakdown, leading to inhomogeneities much larger than those imprinted in the CMB for example. For this reason it is expected that for any model compatible with observations the super-horizon growth of ζ should be only temporary, and according to eq.(34), at some time after horizon crossing $\zeta \approx \mathcal{R}$. This simplifies the calculation of $\mathcal{R}$, whose evolution can be then traced during and after reheating, too, contrary to ζ.

The only exception to this argument could be very small scales ζ modes which leave the horizon very late, and whose super-horizon growth could continue until horizon re-enter, without affecting the validity of the perturbative treatment of the problem. For these small scale modes the difference between $\mathcal{R}$ and ζ could be important, but it would still be computationally convenient to solve the equation for ζ and then obtain $\mathcal{R}$ using the gauge transformation given in eq.(31).

7. Enhancement of curvature perturbations

As already observed for comoving curvature perturbation $\mathcal{R}$ in general relativity for standard kinetic term single field models,[18] a temporary violation of slow-roll conditions can lead to the anomalous growth of what would normally be a decaying mode. A similar mechanism can induce the growth of ζ, as we will show in this section. We can re-write eq.(22) in the form

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(a^3\mathcal{H}\mathcal{G}_S\frac{\mathrm{d}\zeta_k}{\mathrm{d}a}\right)+a\mathcal{F}_S\frac{k^2}{\mathcal{H}}\zeta_k=0\,, \tag{36}$$

from which it is possible to find a super-horizon scale solution of the form

$$\zeta_k=A+B\int\frac{\mathrm{d}a}{a}f\,,\quad f=\frac{1}{a^2\mathcal{H}\mathcal{G}_S}\,, \tag{37}$$

where A and B are constants. For standard slow-roll models the function f decreases as the scale factor increases, implying that ζ tend to a constant value, i.e. the second term in eq.(37) is a decaying mode. If the function f is a growing function of a then the second term in eq.(37) becomes a growing mode, and there can be a super-horizon growth. It follows that the general condition for super-horizon growth of ζ_k is then

$$\frac{\mathrm{d}f}{\mathrm{d}a}\geq 0\,, \tag{38}$$

or equivalently

$$\frac{\mathrm{d}f}{\mathrm{d}a} = \frac{1}{a'}\frac{\mathrm{d}f}{\mathrm{d}\tau} = \frac{1}{a\mathcal{H}}f' \geq 0\,. \tag{39}$$

During inflation $a\mathcal{H} > 0$ and this condition reduces to

$$f' \geq 0\,. \tag{40}$$

In the case of a minimally coupled single scalar field the unitary gauge and the comoving gauge coincide, and the general condition given above takes the form[18]

$$3 - \epsilon + \eta \leq 0\,, \tag{41}$$

where the slow-roll parameters are defined according to

$$\epsilon \equiv -\frac{a}{\mathcal{H}^2}\left(\frac{\mathcal{H}}{a}\right)' = \frac{a^2(\overline{\rho}+\overline{P})}{2M_{Pl}^2\mathcal{H}^2} \quad , \quad \eta \equiv \frac{\epsilon'}{\epsilon\mathcal{H}}\,. \tag{42}$$

In G-Inflation the condition given in eq.(40) implies that

$$f' = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{1}{a^2\mathcal{H}\mathcal{G}_S}\right) = \frac{3 - \epsilon + \mathcal{G}_S'/\mathcal{H}\mathcal{G}_S}{a^2\mathcal{G}_S} = \frac{\gamma}{\delta} \leq 0\,, \tag{43}$$

which gives the general condition for super-horizon growth in an expanding Universe. For a contracting Universe the inequality would be inverted.

As can be seen from the above equation the super-horizon growth can be achieved in different cases, corresponding to γ and δ having opposite signs, contrary to what happened for the standard kinetic term single field scenario, in which δ sign is fixed. Note also that contrary to standard kinetic term single field models, the super-horizon growth does not depend only on the slow-roll parameters, implying that it can occur also during slow-roll.

The anomalous super-horizon growth of ζ, and consequently of $\mathcal{R}$, can increase the abundance of PBHs, since it affects the statistics of the density perturbations peaks which can seed the PBHs.

8. Production of primordial black holes

The super-horizon growth of $\mathcal{R}_k$ could produce primordial black holes which could possibly account for part of dark matter[2,4,7,19–26] and produce gravitational waves (GW) detectable with future GW detectors such as LISA.[5,7] In this session we will show how to obtain some approximate estimation of the effetcs of the modification of gravity on the PBH production, without considering any specific model, leaving this to a future work.

The mass M of PBHs produced by the mode $\mathcal{R}_k$ re-entering the horizon during the radiation domination can be approoximated as[7]

$$M = \gamma M_H\Big|_F\,, \tag{44}$$

where $\gamma \approx 0.2$ is a correction factor, and $M_H\Big|_F$ is the horizon mass $M_H \equiv (4\pi/3)\overline{\rho}(a\mathcal{H})^{-3}$ at the time of PBH formation, corresponding to the horizon crossing time

$$k = (a^2\mathcal{H})\Big|_F . \tag{45}$$

Note the above is just a rough estimation, and a more accurate treatment would involve the use of a scaling relation.[27,28]

The present time fraction f_{PBH} of PBHs of mass M against the total dark matter component can then be approximated as[7]

$$f = 2.7 \times 10^8 \left(\frac{\gamma}{0.2}\right)^{1/2} \left(\frac{g_{*F}}{106.75}\right)^{-1/4} \left(\frac{M}{M_\odot}\right)^{-1/2} \beta ,$$

where g_{*F} is the number of relativistic degrees of freedom at formation, the quantity β is the energy density fraction of PBHs at formation time

$$\beta \equiv \frac{\overline{\rho}_{PBH}}{\overline{\rho}}\Big|_F , \tag{46}$$

which can be written in terms of the probability of the density contrast $P(\delta)$ as[3,29]

$$\beta(M) = \gamma \int_{\delta_t}^{1} P(\delta) \mathrm{d}\delta , \tag{47}$$

where δ_t is the threshold for PBH formation. Assuming the density perturbations follow a Gaussian distribution β is given by

$$\beta(M) \approx \frac{\gamma}{\sqrt{2\pi}\nu(M)} \exp\left[-\frac{\nu(M)^2}{2}\right] , \tag{48}$$

where $\nu(M) \equiv \delta_t/\sigma(M)$, and $\sigma(M)$ is an estimation of the standard deviation of the density contrast on scale R from the variance

$$\sigma^2(M) = \int \mathrm{d}\ln k W^2(kR)\mathcal{P}_\delta(k) = \int \mathrm{d}\ln k W^2(kR) \left(\frac{16}{81}\right) (kR)^4 \mathcal{P}_\mathcal{R}(k) , \tag{49}$$

where $W(kR)$ is a window function smoothing over the comoving scale $R(M) = (a^2\mathcal{H})^{-1}\Big|_F = 2GM/a_F\gamma^{-1}$, and the relation between δ and $\mathcal{R}$ has been used in the second equality. It should be mentioned that eq.(48) can be used as a guideline, but more accurate calculations would involve the use of the results of numerical simulations.[30,31] The choice of the window function could also affect[32,33] the results of the calculation.

Our aim here is not make an accurate estimation of the PBHs abundance for a specific model, but to show why in general it can be impacted by the modification of gravity, and the approximations adopted so far are enough to serve this general purpose. According to the equations above, the PBH fraction β is affected by the

power spectrum of $\mathcal{R}$ since this can increase the standard deviation of the density field $\sigma(M)$. Note that the above approximations to estimate the PBHs abundance can receive important corrections depending on the shape of power spectrum, on non- gaussianity, and non-linear statistics.[34–36] Due to the importance of all these different effects it is difficult to find a general model independent analytical formula to estimate the PBHs abundance for a generic G-inflation theory, but any enhancement of the power spectrum is expected, according to eq.(32), to affect the probability of production of PBHs. Beside this, numerical relativity simulations of the PBHs formation are based on general relativity, so the effects of the modification of gravity on the process of gravitational collapse are at the moment not fully understood and would require investigations beyond the scope of this paper.[37]

At the end of its anomalous super-horizon growth, ζ will coincide with $\mathcal{R}$, and the consequent enhancement of the spectrum will lead to an increased PBH abundance. Contrary to what happens for standard kinetic term single field models in general relativity,[18] in the case of G-inflation this power spectrum enhancement can be achieved also during slow-roll, as long as the condition in eq.(43) is satisfied, which can be attained by an appropriate choice of the function $\mathcal{G}_S$. We expect a similar behavior for more complex modified gravity theories as well.

9. Conclusions

We have computed the effective energy-stress-tensor for G-inflation theories in the comoving slices gauge and have used it to derive a general relation between the unitary gauge curvature ζ and the comoving curvature perturbation $\mathcal{R}$, involving an enhancement function which depends on the evolution of the background, and which can cause a large difference between the two gauge invariant quantities.

When ζ is not constant in time it differs from $\mathcal{R}$, for example on sub-horizon scales, or in models exhibiting an anomalous super-horizon growth of ζ, but since this growth cannot last indefinitely, eventually they will coincide. We have derived the general condition for super-horizon growth of ζ, showing that slow-roll violation is not necessary, and discussed how the the enhancement of the spectrum of $\mathcal{R}$ can affect the PBH abundance.

We expect similar results to hold for other modified gravity theories such as other Horndeski's theories,[1] since also for these theories there can be effective entropy or anisotropy terms which can modify the evolution of curvature perturbations. In the future it will be interesting to extend this study to other modified gravity theories or to multi-fields systems, and to use observations to constraints the different types of theories. It would also be important to perform numerical simulations of the PBHs formation taking into account the non perturbative effects of the modification of gravity on the process of black hole formation.

References

1. G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, *Int. J. Theor. Phys.* **10**, 363 (1974).
2. K. M. Belotsky, A. D. Dmitriev, E. A. Esipova, V. A. Gani, A. V. Grobov, M. Yu. Khlopov, A. A. Kirillov, S. G. Rubin and I. V. Svadkovsky, Signatures of primordial black hole dark matter, *Mod. Phys. Lett.* **A29**, p. 1440005 (2014).
3. B. J. Carr, The primordial black hole mass spectrum., *Astrophysical Journal* **201**, 1 (Oct 1975).
4. M. Y. Khlopov, Primordial Black Holes, *Res. Astron. Astrophys.* **10**, 495 (2010).
5. M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914, *Phys. Rev. Lett.* **117**, p. 061101 (2016), [erratum: Phys. Rev. Lett.121,no.5,059901(2018)].
6. A. S. Josan, A. M. Green and K. A. Malik, Generalised constraints on the curvature perturbation from primordial black holes, *Phys. Rev.* **D79**, p. 103520 (2009).
7. M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, Primordial black holes—perspectives in gravitational wave astronomy, *Class. Quant. Grav.* **35**, p. 063001 (2018).
8. M. Bucher, K. Moodley and N. Turok, The General primordial cosmic perturbation, *Phys. Rev.* **D62**, p. 083508 (2000).
9. A. E. Romano, S. Mooij and M. Sasaki, Adiabaticity and gravity theory independent conservation laws for cosmological perturbations, *Phys. Lett.* **B755**, 464 (2016).
10. T. Kobayashi, M. Yamaguchi and J. Yokoyama, G-inflation: Inflation driven by the Galileon field, *Phys. Rev. Lett.* **105**, p. 231302 (2010).
11. S. Hirano, T. Kobayashi and S. Yokoyama, Ultra slow-roll G-inflation, *Phys. Rev. D* **94**, p. 103515 (2016).
12. C. Deffayet, O. Pujolas, I. Sawicki and A. Vikman, Imperfect Dark Energy from Kinetic Gravity Braiding, *JCAP* **1010**, p. 026 (2010).
13. H. Kodama and M. Sasaki, Cosmological Perturbation Theory, *Prog. Theor. Phys. Suppl.* **78**, 1 (1984).
14. T. Kobayashi, M. Yamaguchi and J. Yokoyama, Generalized G-inflation: Inflation with the most general second-order field equations, *Prog. Theor. Phys.* **126**, 511 (2011).
15. A. E. Romano, General background conditions for K-bounce and adiabaticity, *Eur. Phys. J.* **C77**, p. 147 (2017).
16. J. Garriga and V. F. Mukhanov, Perturbations in k-inflation, *Phys. Lett.* **B458**, 219 (1999).
17. A. E. Romano, S. Mooij and M. Sasaki, Global adiabaticity and non-Gaussianity consistency condition, *Phys. Lett.* **B761**, 119 (2016).
18. R. Saito, J. Yokoyama and R. Nagata, Single-field inflation, anomalous enhancement of superhorizon fluctuations, and non-Gaussianity in primordial black hole formation, *JCAP* **06**, p. 024 (2008).
19. B. J. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, New cosmological constraints on primordial black holes, *Phys. Rev. D* **81**, p. 104019 (2010).
20. B. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, Constraints on Primordial Black Holes (2 2020).
21. J. Yokoyama, Formation of MACHO primordial black holes in inflationary cosmology, *Astron. Astrophys.* **318**, p. 673 (1997).
22. J. Garcia-Bellido and E. Ruiz Morales, Primordial black holes from single field models of inflation, *Phys. Dark Univ.* **18**, 47 (2017).
23. J. Garcia-Bellido, A. D. Linde and D. Wands, Density perturbations and black hole formation in hybrid inflation, *Phys. Rev.* **D54**, 6040 (1996).

24. P. Ivanov, P. Naselsky and I. Novikov, Inflation and primordial black holes as dark matter, *Phys. Rev.* **D50**, 7173 (1994).
25. G. F. Chapline, Cosmological effects of primordial black holes, *Nature* **253**, 251 (Jan 1975).
26. K. M. Belotsky, V. I. Dokuchaev, Y. N. Eroshenko, E. A. Esipova, M. Yu. Khlopov, L. A. Khromykh, A. A. Kirillov, V. V. Nikulin, S. G. Rubin and I. V. Svadkovsky, Clusters of primordial black holes, *Eur. Phys. J.* **C79**, p. 246 (2019).
27. J. Yokoyama, Cosmological constraints on primordial black holes produced in the near critical gravitational collapse, *Phys. Rev. D* **58**, p. 107502 (1998).
28. J. C. Niemeyer and K. Jedamzik, Near-critical gravitational collapse and the initial mass function of primordial black holes, *Phys. Rev. Lett.* **80**, 5481 (1998).
29. A. M. Green, A. R. Liddle, K. A. Malik and M. Sasaki, A New calculation of the mass fraction of primordial black holes, *Phys. Rev.* **D70**, p. 041502 (2004).
30. T. Nakama, T. Harada, A. G. Polnarev and J. Yokoyama, Identifying the most crucial parameters of the initial curvature profile for primordial black hole formation, *JCAP* **01**, p. 037 (2014).
31. M. Shibata and M. Sasaki, Black hole formation in the Friedmann universe: Formulation and computation in numerical relativity, *Phys. Rev. D* **60**, p. 084002 (1999).
32. K. Ando, K. Inomata and M. Kawasaki, Primordial black holes and uncertainties in the choice of the window function, *Phys. Rev. D* **97**, p. 103528 (2018).
33. K. Tokeshi, K. Inomata and J. Yokoyama, Window function dependence of the novel mass function of primordial black holes, *JCAP* **12**, p. 038 (2020).
34. C. Germani and I. Musco, Abundance of Primordial Black Holes Depends on the Shape of the Inflationary Power Spectrum, *Phys. Rev. Lett.* **122**, p. 141302 (2019).
35. V. Atal and C. Germani, The role of non-gaussianities in Primordial Black Hole formation, *Phys. Dark Univ.* **24**, p. 100275 (2019).
36. C. Germani and R. K. Sheth, Nonlinear statistics of primordial black holes from Gaussian curvature perturbations, *Phys. Rev. D* **101**, p. 063520 (2020).
37. C.-Y. Chen, Threshold of primordial black hole formation in modified theories of gravity (2019).

$U(1)$ local strings in generalized hybrid metric-Palatini gravity

Hilberto M. R. da Silva

Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal and Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, PT4150-762 Porto, Portugal

Tiberiu Harko

Astronomical Observatory, 19 Ciresilor Street, 400487 Cluj-Napoca, Romania
Faculty of Physics, Babes-Bolyai University, 1 Kogalniceanu Street, 400084 Cluj-Napoca, Romania
School of Physics, Sun Yat-Sen University, Xingang Road, 510275 Guangzhou, People's Republic of China

Francisco S. N. Lobo

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisbon, Portugal

João Luís Rosa

Institute of Physics, University of Tartu, W. Ostwaldi 1, 50411 Tartu, Estonia

In this work we will explore $U(1)$ local cosmic string solutions in the context of the generalized hybrid metric-Palatini theory of gravity in its scalar-tensor representation. Using a general static cillindrically symmetric metric to find the dynamical equations for this particular case, we will simplify the equations by imposing boost invariance along t and z directions. The physical and geometrical properties of the cosmic strings are determined by the two scalar fields, as well by an effective field potential, functionally dependent on both scalar fields. While for some forms of the potential, the dynamical equations can be solved exactly, for more general formas of the potential the solutions are found numerically. In this way, we obtain a large class of stable stringlike astrophysical configurations, whose basic parameters (string tension and radius) depend essentially on the effective field potential, and on the boundary conditions.

Keywords: Cosmic Strings, Modified Gravity, hybrid metric-Palatini theory

1. Introduction

The main motivation to investigate hybrid metric-Palatini theories resides on the fact that these theories are able to overcome flaws of both the metric and the Palatini approaches to $f(R)$ gravity. For example, in both the metric and the Palatini formalisms of $f(R)$ gravity, one is able to model the late-time cosmic acceleration period without invoking dark energy sources,[1] but both approaches present unavoidable disadvantages: the metric $f(R)$ was shown to be inconsistent with solar-system constraints unless chameleon mechanisms are considered,[2,3] whereas

the Palatini $f(R)$ gravity induces microscopic instabilities, surface singularities in polytropic star models, and is unable to describe the evolution of cosmological perturbations.[4,5] The HMPG successfully unifies the late-time cosmic acceleration period with the weak-field solar system dynamics without the need for chameleon mechanisms,[6] thus being a viable and relevant modification to GR. We refer the reader to Refs.[7–9] for recent reviews on the topic.

Inspired by the success of electroweak theory,[10–12] which unifies the weak and electromagnetic interaction under the gauge group $SU(2)\times U(1)$ at a scale of around $10^2 GeV$, Grand Unified Theories (GUT) propose the unification of electroweak and strong interactions under a more general symmetry that takes place at higher energy scales, around $10^{16} GeV$, the Grand Unification Scale. GUT theories are supported by the observation that the coupling "constants" of the Standard Model for Particle Physics seem to slowly vary with the energy scale, converging to a common value at the Grand Unification Scale.[13]

These symmetries presented at higher energies are spontaneously broken as the system lowers its energy state. In several GUT scenarios proposed, a universal covering group G, which would be effective above the GUT scale, would spontaneously break into the Standard Model $SU(3) \times SU(2) \times U(1)$, where $SU(3)$ is the symmetry group of quantum chromodynamics, describing the strong interaction, and $SU(2) \times U(1)$ is the aforementioned electroweak group.

These phase transitions may have left behind some relics that can help shed some light into earlier times of our Universe.[14] These relics are known as topological defects and are a well known, and studied, phenomena in physics, particularly in the context of condensed matter (namely metal crystallization,[15] liquid crystals,[16] superfluid helium-3 and helium-4,[17] and superconductivity[18]).

The underlying idea behind topological defect formation is the one of Spontaneous symmetry breaking, which is the principle behind the Higgs-Englert mechanism.[19]

Cosmic strings are one of the possible topological defects formed after spontaneous symmetry breaking (SSB) during phase transitions in the history of the Universe.

The type of strings to be considered in this work are local $U(1)$ cosmic strings, which are an extension of the global $U(1)$ strings to include gauge fields. Local strings differ from the global cosmic strings in what concerns the symmetry that is effective above the spontaneous breaking scale; in the case of local strings, the lagrangian remains invariant under local transformations of the type $\phi(x) \longrightarrow e^{i\alpha(x)}\phi(x)$.

The study of the properties and dynamics of cosmic strings in the context of modified theories of gravity is crucial in the advent of powerful observatories, such as LISA, as it may allow us to constrain both Modified Gravity theories and Grand Unified theories.

2. Generalized Hybrid Metric-Palatini Gravity

The generalized HMPG theory is described by an action S of the form

$$S = \frac{1}{2\kappa^2}\int_\Omega \sqrt{-g} f\left(R,\mathcal{R}\right) d^4x + \int_\Omega \sqrt{-g}\,\mathcal{L}_m d^4x, \tag{1}$$

where $\kappa^2 \equiv 8\pi G/c^4$, G is the gravitational constant and c the speed of light, Ω is the spacetime manifold described by a system of coordinates x^a, g is the determinant of the spacetime metric g_{ab}, where Latin indices run from 0 to 3, $R = g^{ab}R_{ab}$ is the Ricci scalar of the metric g_{ab} and where R_{ab} is the Ricci tensor, $\mathcal{R} \equiv \mathcal{R}^{ab}g_{ab}$ is the Palatini Ricci scalar, where the Palatini Ricci tensor $\mathcal{R}_{ab}$ is defined in terms of an independent connection $\hat{\Gamma}^c_{ab}$ in the usual form as $\mathcal{R}_{ab} = \partial_c\hat{\Gamma}^c_{ab} - \partial_b\hat{\Gamma}^c_{ac} + \hat{\Gamma}^c_{cd}\hat{\Gamma}^d_{ab} - \hat{\Gamma}^c_{ad}\hat{\Gamma}^d_{cb}$, where ∂_a denotes partial derivatives with respect to the coordinates x^a, $f\left(R,\mathcal{R}\right)$ is a well-behaved function of R and $\mathcal{R}$, and $\mathcal{L}_m$ is the matter Lagrangian density considered to be minimally coupled to the metric g_{ab}. Equation (1) depends on two independent variables, namely the metric g_{ab} and the independent connection $\hat{\Gamma}^c_{ab}$, and thus two equations of motion can be obtained.

Taking a variation of Eq. (1) with respect to the metric g_{ab} leads to the modified field equations

$$\frac{\partial f}{\partial R}R_{ab} + \frac{\partial f}{\partial \mathcal{R}}\mathcal{R}_{ab} - \frac{1}{2}g_{ab}f\left(R,\mathcal{R}\right) - \left(\nabla_a\nabla_b - g_{ab}\Box\right)\frac{\partial f}{\partial R} = \kappa^2 T_{ab}, \tag{2}$$

where ∇_a denotes covariant derivatives and $\Box = \nabla^a\nabla_a$ the d'Alembert operator, both with respect to g_{ab}, and T_{ab} is the energy-momentum tensor defined as usual:

$$T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\,\mathcal{L}_m)}{\delta(g^{ab})}\ . \tag{3}$$

On the other hand, taking a variation of Eq. (1) with respect to the independent connection $\hat{\Gamma}^c_{ab}$ yields

$$\hat{\nabla}_c\left(\sqrt{-g}\frac{\partial f}{\partial \mathcal{R}}g^{ab}\right) = 0\,, \tag{4}$$

where $\hat{\nabla}_a$ is the covariant derivative written in terms of the independent connection $\hat{\Gamma}^c_{ab}$. Since $\sqrt{-g}$ is a scalar density of weight 1, then $\hat{\nabla}_c\sqrt{-g} = 0$ and Eq. (4) can be rewritten in the form $\hat{\nabla}_c\left(\frac{\partial f}{\partial \mathcal{R}}g^{ab}\right) = 0$. This result implies the existence of a new metric, h_{ab}, conformally related to the metric g_{ab} via

$$h_{ab} = g_{ab}\frac{\partial f}{\partial \mathcal{R}}\,, \tag{5}$$

in such a way that the independent connection is the Levi-Civita connection of the metric h_{ab}, i.e., $\hat{\Gamma}^c_{ab}$ can be written as

$$\hat{\Gamma}^a_{bc} = \frac{1}{2}h^{ad}\left(\partial_b h_{dc} + \partial_c h_{bd} - \partial_d h_{bc}\right)\ . \tag{6}$$

2.1. *Scalar-tensor representation of generalized HMPG with matter*

It is sometimes useful to represent the generalized HMPG theory in a dinamically equivalent scalar-tensor representation in which the two extra scalar degrees of freedom of the theory are explicitly carried by two scalar fields. To obtain this representation, we introduce two auxiliary fields α and β into Eq. (1) and rewrite it in the form

$$S = \frac{1}{2\kappa^2}\int_\Omega \sqrt{-g}\left[f(\alpha,\beta) + \frac{\partial f}{\partial \alpha}(R-\alpha) + \frac{\partial f}{\partial \beta}(\mathcal{R}-\beta)\right] d^4x + \int_\Omega \sqrt{-g}\,\mathcal{L}_m d^4x. \tag{7}$$

At this point one verifies that if $\alpha = R$ and $\beta = \mathcal{R}$ one recovers Eq. (1). This equivalence between the two representations is only guaranteed if the determinant of the Hessian matrix of the function $f(\alpha,\beta)$ is non-zero, i.e., if $f_{\alpha\alpha}f_{\beta\beta} - f_{\alpha\beta}^2 \neq 0$, where the subscripts α and β denote partial derivatives with respect to these variables. Defining two scalar fields as $\varphi = \partial f(\alpha,\beta)/\partial\alpha$ and $\psi = -\partial f(\alpha,\beta)/\partial\beta$, where the negative sign in ψ is imposed to avoid the presence of ghosts, Eq. (7) takes the form

$$S = \frac{1}{2\kappa^2}\int_\Omega \sqrt{-g}\left[\varphi R - \psi\mathcal{R} - V(\varphi,\psi)\right] d^4x + \int_\Omega \sqrt{-g}\,\mathcal{L}_m d^4x, \tag{8}$$

where the function $V(\varphi,\psi)$ plays the role of an interaction potential between the two scalar fields and it is defined as

$$V(\varphi,\psi) = -f(\alpha,\beta) + \varphi\alpha - \psi\beta\,. \tag{9}$$

Recalling the conformal relation between h_{ab} and g_{ab} in Eq. (5), which can now be written in the form $h_{ab} = -\psi\, g_{ab}$ by taking into consideration the definition of ψ, one can derive a relationship between R and $\mathcal{R}$ as

$$\mathcal{R} = R + \frac{3}{\psi^2}\partial^a\psi\partial_a\psi - \frac{3}{\psi}\Box\psi\,, \tag{10}$$

which can be used to eliminate $\mathcal{R}$ from Eq. (8) and gives the final form of the action

$$S = \frac{1}{2\kappa^2}\int_\Omega \sqrt{-g}\Big[(\varphi-\psi)R - \frac{3}{2\psi}\partial^a\psi\partial_a\psi - V(\varphi,\psi)\Big] d^4x + \int_\Omega \sqrt{-g}\,\mathcal{L}_m d^4x. \tag{11}$$

Equation (11) is now a function of three independent variables, namely, the metric g_{ab} and the two scalar fields φ and ψ. Taking a variation of Eq. (11) with respect to the metric g_{ab} yields the modified field equations in the scalar-tensor representation. Varying the action (11) with respect to the metric g_{ab} provides the following gravitational equation

$$\begin{aligned}(\varphi-\psi)G_{ab} = {}& \kappa^2 T_{ab} + \nabla_a\nabla_b\varphi - \nabla_a\nabla_b\psi + \frac{3}{2\psi}\partial_a\psi\partial_b\psi \\ & - \left(\Box\varphi - \Box\psi + \frac{1}{2}V + \frac{3}{4\psi}\partial^c\psi\partial_c\psi\right) g_{ab}\,.\end{aligned} \tag{12}$$

Note that Eq. (12) could be obtained directly from Eq. (2) via the introduction of the definitions of φ, ψ and $V(\varphi,\psi)$, which further emphasizes the equivalence between the two representations.

Finally, the equations of motion for the scalar fields φ and ψ can be obtained via a variation of Eq. (11) with respect to these fields, respectively, which after algebraic manipulations can be written in the forms

$$\Box\varphi+\frac{1}{3}\left(2V-\psi V_{\psi}-\varphi V_{\varphi}\right)=\frac{\kappa^2 T}{3}\,, \tag{13}$$

$$\Box\psi-\frac{1}{2\psi}\partial^a\psi\partial_a\psi-\frac{\psi}{3}\left(V_{\varphi}+V_{\psi}\right)=0\,, \tag{14}$$

respectively.

Notice from Eq.(11) that the coupling between the scalar fields and the Ricci scalar is the combination $\varphi-\psi$. Since φ and ψ are arbitrary functions, it is not guaranteed that this combination preserves the positivity of the coupling. We thus introduce a redefinition of the scalar field φ as $\xi^2=\varphi-\psi$. With this redefinition, any solution obtained for which ξ is a real function preserves the positivity of the coupling $(\varphi-\psi)R$. Equations (12) to (14) thus become

$$\xi^2 G_{ab}=\kappa^2 T_{ab}+\nabla_a\nabla_b\xi^2+\frac{3}{2\psi}\partial_a\psi\partial_b\psi-\left(\Box\xi^2+\frac{1}{2}\bar{V}+\frac{3}{4\psi}\partial^c\psi\partial_c\psi\right)g_{ab}, \tag{15}$$

$$\Box\xi^2+\frac{1}{2\psi}\partial^a\psi\partial_a\psi+\frac{1}{6}\left(4\bar{V}-\xi\bar{V}_{\xi}\right)=\frac{\kappa^2 T}{3}, \tag{16}$$

$$\Box\psi-\frac{1}{2\psi}\partial^a\psi\partial_a\psi-\frac{\psi}{3}\left(\frac{1}{2\xi}\bar{V}_{\xi}+\bar{V}_{\psi}\right)=0, \tag{17}$$

where $\bar{V}(\xi,\psi)$ is the potential written in terms of the scalar fields ξ and ψ and the subscript ξ denotes a partial derivative with respect to this scalar field. In the next section, we will use the equations of motion (15)–(17) to find cosmic string solutions. Finally, one can also obtain a relationship between the potential $\bar{V}$ and the function $f(R,\mathcal{R})$ from Eq.(9) as

$$\bar{V}(\xi,\psi)=-f(R,\mathcal{R})+\xi^2 R+\psi\left(R-\mathcal{R}\right), \tag{18}$$

where we have used the fact that the scalar-tensor representation is only defined if $\alpha=R$ and $\beta=\mathcal{R}$. This equation becomes a PDE for $f(R,\mathcal{R})$ by replacing $\psi=f_{\mathcal{R}}$ and $\xi^2=f_R+f_{\mathcal{R}}$.

Hence Eq. (18) becomes

$$V\left(\sqrt{\frac{\partial f(R,\mathcal{R})}{\partial R}+\frac{\partial f(R,\mathcal{R})}{\partial\mathcal{R}}},\frac{\partial f(R,\mathcal{R})}{\partial\mathcal{R}}\right)=-f(R,\mathcal{R})+R\frac{\partial f(R,\mathcal{R})}{\partial R}+\mathcal{R}\frac{\partial f(R,\mathcal{R})}{\partial\mathcal{R}}. \tag{19}$$

3. Dynamical equations of local strings in hybrid metric-palatini using Vilenkin's approximation

In this work, we consider a straight infinite Abelian-Higgs cosmic string, using Vilenkin's approximation[20]

$$T_t^t = T_z^z = -\sigma(r)\,, \tag{20}$$

where σ is the string tension. We further assume cylindrical symmetry with a general metric of the form:

$$ds^2 = -e^{2(K-U)}dt^2 + e^{2(K-U)}dr^2 + e^{-2U}W^2 d\theta^2 + e^{2U}dz^2, \tag{21}$$

where t, r, θ and z denote the time, radial, angular and axial cylindrical coordinates, respectively, and K, U and W are functions of r alone. This metric is invariant under a set of transformations of the type: $x^0 \to x^0 + c_1$, $x^3 \to x^3 + c_2$, $x^0 \to -x^0$ and $x^3 \to -x^3$, where c_1 and c_2 are constants, which renders this metric static and cylindrically symmetric.

Since in this model the matter field couples minimally with curvature, it is possible to show that the energy conservation equation still holds, i.e.,

$$\nabla_a T^a{}_b = 0, \tag{22}$$

which provides $K'\sigma = 0$ and, apart from the trivial vacuum solution $\sigma = 0$, this implies that $K' = 0$, where the prime represents a differentiation w.r.t. r. Thus, we consider from now on that $e^K = 1$.

Note that local gauge strings preserve boost invariance along t and z,[20] so that this requires $U = 0$. Hence the only surviving non-trivial metric tensor component is $g_{\theta\theta} = W^2(r)$. From a geometric point of view $W(r)$ is the radius of the coordinate circles r = constant, z = constant, parameterized by the angle θ. Since in this geometry the perimeter of a circle equals $2\pi W$, in the following we will call the only remaining metric tensor component $W^2(r)$ a *circular radius*. On the other hand $W^2(r)$ also has the geometric meaning of a length that may be counted from any zero point, with its value at $r = 0$ not distinguished geometrically. Hence the metric of the cosmic string reduces to the form

$$ds^2 = -dt^2 + dr^2 + W^2(r)d\theta^2 + dz^2. \tag{23}$$

Applying this symmetry, the gravitational field equations simplify considerably. Equation (15) provides three independent field equations, which are

$$\xi^2\frac{W''}{W} + 2\xi\xi'\frac{W'}{W} + \frac{3\psi'^2}{4\psi} + 2\left(\xi'^2 + \xi\xi''\right) + \frac{\bar{V}}{2} = -\kappa^2\sigma, \tag{24}$$

$$2\xi\xi'\frac{W'}{W} - \frac{3\psi'^2}{4\psi} + \frac{\bar{V}}{2} = 0, \tag{25}$$

$$2\left(\xi'^2 + \xi\xi''\right) + \frac{3\psi'^2}{4\psi} + \frac{\bar{V}}{2} = \frac{d^2}{dr^2}\xi^2 + \frac{3\psi'^2}{4\psi} + \frac{\bar{V}}{2} = 0\,, \tag{26}$$

whereas the scalar field equations for ξ and ψ, provided by Eqs. (16) and (17), give

$$2\left(\xi'^2+\xi\xi''\right)+2\xi\xi'\frac{W'}{W}+\frac{\psi'^2}{2\psi}+\frac{1}{6}\left(\bar{V}-\xi\bar{V}_\xi\right)=-\frac{2\kappa^2}{3}\sigma, \tag{27}$$

$$\psi''+\frac{W'}{W}\psi'-\frac{\psi'^2}{2\psi}-\frac{\psi}{3}\left(\bar{V}_\psi+\frac{1}{2\xi}\bar{V}_\xi\right)=0\,. \tag{28}$$

The system of Eqs. (24)-(28) is a system of five equations from which only four are linearly independent. This statement can be proven by taking a radial derivative of Eq. (25), using Eq. (24) to cancel the factors W'', using Eq. (27) to cancel σ, and finally using Eqs. (26) and (28) to cancel the second-order derivatives of the scalar fields ξ'' and ψ'', respectively. As a result, one recovers Eq. (25), thus proving that the system of equations is linearly dependent. Thus, one only needs to consider four of these equations to completely determine the solution in the sections that follow. Given its complexity, we chose to discard Eq. (24) from the analysis.

Furthermore, an equation for the potential $\bar{V}$ can be obtained by summing the field equations in Eqs. (25) and (26), yielding

$$\bar{V}=-2\left(\xi'^2+\xi\xi''\right)-2\xi\xi'\frac{W'}{W}. \tag{29}$$

This equation is particularly useful to obtain an equation for W' in terms of the scalar fields ξ and ψ and their derivatives after setting an explicit form of the potential $\bar{V}$.

The system of basic equations describing the structure of a cosmic string can thus be reformulated in the form of a first-order dynamical system. By defining $\alpha=\xi^2$, and by introducing two extra dynamical variables as $u=\alpha'$ and $v=\psi'$, the dynamical system takes the form

$$\frac{d\alpha}{dr}=u,\qquad \frac{d\psi}{dr}=v, \tag{30}$$

$$\frac{dW}{dr}=\frac{1}{u}\left(\frac{3v^2}{4\psi}-\frac{\bar{V}}{2}\right)W, \tag{31}$$

$$\frac{du}{dr}=-\frac{3v^2}{4\psi}-\frac{\bar{V}}{2}, \tag{32}$$

$$\frac{dv}{dr}=-\frac{v}{u}\left(\frac{3v^2}{4\psi}-\frac{\bar{V}}{2}\right)+\frac{v^2}{2\psi}+\frac{\psi}{3}\left(\bar{V}_\psi+\frac{1}{2\sqrt{\alpha}}\bar{V}_{\sqrt{\alpha}}\right), \tag{33}$$

where Eq. (30) is the explicit definition of u and v, and Eqs. (31)–(33) are reformulations of Eqs. (25), (27), and (28), respectively. Once the functional form of the potential $\bar{V}(\xi,\psi)$ is specified, the system of Eqs. (30)-(33) represents a system of ordinary, strongly nonlinear, differential equations for the variables $\left(\alpha=\xi^2,\psi,W,u,v\right)$. To solve this system, one has to impose a set of boundary conditions at some radius $r=r_0$, i.e., $\alpha\left(r_0\right)=\alpha_0$, $\psi\left(r_0\right)=\psi_0$, $W\left(r_0\right)=W_0$, $u\left(r_0\right)=u_0$, and $v\left(r_0\right)=v_0$,

respectively, which specify the boundary values of the variables on, or nearby the string axis. Moreover, we will also impose the condition $u(r_0) \neq v(r_0)$. Once the system is solved, the string tension can be obtained from Eq. (27), and it is given by

$$\frac{2}{3}\kappa^2\sigma = \bar{V} - \frac{v^2}{2\psi} - \frac{1}{6}\left(\bar{V} - \sqrt{\alpha}\bar{V}_{\sqrt{\alpha}}\right). \tag{34}$$

An important physical characteristic of the string-like objects is their mass per unit length m_s, defined as

$$m_s(R_s) = \int_0^{2\pi} d\theta \int_0^{R_s} \sigma(r)W(r)dr = 2\pi \int_0^{R_s} \sigma(r)W(r)dr, \tag{35}$$

where R_s is the radius of the string, defined as the distance from the center where the string tension vanishes, $\sigma(R_s) = 0$, and $\sigma(r) \equiv 0, \forall r \geq R_s$. Note that, in general, the solutions obtained for σ do not satisfy the property $\sigma(r) \equiv 0, \forall r \geq R_s$, and this condition must be imposed manually by performing a matching between the string spacetime and an exterior cosmological spacetime. This matching must be performed via the use of the junction conditions of the theory, previously used in.[21] However, we do not pursue this analysis here as it is out of the scope of this paper. Using Eqs. (26) and Eq. (24) the mass per unit length of the string can be expressed as

$$\kappa^2 m_s(R_s) = 3\pi \int_0^{R_s} \left[\bar{V} - \frac{v^2}{2\psi} - \frac{1}{6}\left(\bar{V} - \sqrt{\alpha}\bar{V}_{\sqrt{\alpha}}\right)\right] W dr. \tag{36}$$

4. Solutions to the dynamical equations with specific potentials

In this section we will investigate the application of the set of equations deduced on the previous section to different potential configurations, for a more complete set of possible potential configurations, we refer the reader to the original article Ref. 22.

4.1. *Constant potential*

We consider the case when the potential V is a constant, so that $V = \Lambda = \text{constant}$. In this case Eq. (19) takes the form

$$-f(R,\mathcal{R}) + R\frac{\partial f(R,\mathcal{R})}{\partial R} + \mathcal{R}\frac{\partial f(R,\mathcal{R})}{\partial \mathcal{R}} = \Lambda, \tag{37}$$

and it has the general solution

$$f(R,\mathcal{R}) = Rg\left(\frac{\mathcal{R}}{R}\right) + \mathcal{R}h\left(\frac{R}{\mathcal{R}}\right) - \Lambda, \tag{38}$$

where g and h are arbitrary functions.

For a constant potential Eq. (28) simplifies to

$$\frac{W'}{W} = -\frac{\psi''}{\psi'} + \frac{1}{2}\frac{\psi'}{\psi}, \tag{39}$$

which allows us to write Eq. (25) in the form

$$\frac{d}{dr}\xi^2 = \frac{3\psi'^2/4\psi - \Lambda/2}{W'/W} = \frac{3\psi'^2/4\psi - \Lambda/2}{-\psi''/\psi' + (1/2)\psi'/\psi}. \tag{40}$$

To facilitate the analysis, we introduce now a new function $h = \psi'^2/\psi$. The radial derivative of this function can be written in terms of ψ and its derivatives as

$$h' = 2h\left(\frac{\psi''}{\psi'} - \frac{1}{2}\frac{\psi'}{\psi}\right). \tag{41}$$

This definition allows us to rewrite Eq. (40) in terms of h as

$$\frac{d}{dr}\xi^2 = -\frac{(3/4)h - \Lambda/2}{h'/2h}, \tag{42}$$

which can then be differentiated with respect to r and inserted into Eq. (26) to cancel the dependency in $d^2\xi^2/dr^2$ and ψ''. As a result, we obtain an equation depending solely in h of the form

$$\frac{(2\Lambda - 3h)\left(3h'^2 - 2hh''\right)}{4h'^2} = 0. \tag{43}$$

This equation is undefined for $h' = 0$, as the denominator vanishes in this case. Thus, in the following we ignore the solution corresponding to $h = 2\Lambda/3 =$ constant, giving $h' = 0$, and $W =$ constant. The general solution of Eq. (43) is given by

$$h(r) = \frac{c_2}{(r + 2c_1)^2}, \tag{44}$$

where c_1 and c_2 are arbitrary integration constants. Recalling that $h' = \psi'^2/\psi$, Eq. (44) becomes a separable ODE for ψ which can be directly integrated and provides the general solution

$$\psi(r) = \left[c_3 \pm \frac{\sqrt{c_2}}{2}\ln(r + 2c_1)\right]^2, \tag{45}$$

where c_3 is an arbitrary integration constant. Inserting Eq. (45) into Eq. (42) and integrating gives for ξ^2 the expression

$$\xi^2(r) = \xi_0^2 + \frac{1}{4}\left[3c_2\ln(r + 2c_1) - \Lambda r\,(r + 4c_1)\right], \tag{46}$$

where ξ_0^2 is an integration constant. Inserting Eq. (45) into Eq. (39) we obtain the solution for W

$$W(r) = W_0\,(r + 2c_1), \tag{47}$$

where W_0 is a constant of integration. Hence the cosmic string metric tensor component $W(r)$ is the same in both $V = 0$ and $V = \Lambda$ cases. Finally, the string tension can be computed via Eq. (27), leading to

$$\kappa^2\sigma = \frac{\Lambda}{2} - \frac{3c_2}{4\,(2c_1 + r)^2}. \tag{48}$$

On the string axis $r = 0$ we obtain for the string tension the value

$$\kappa^2\sigma_0 = \kappa^2\sigma(0) = \frac{\Lambda}{2} - \frac{3c_2}{16c_1^2}. \tag{49}$$

The condition of the positivity of the string tension imposes the condition $c_2/c_1^2 < 8\Lambda/3$ on the integration constants.

For $\Lambda = 0$ we reobtain the expression corresponding to the case $V = 0$. However, by an appropriate choice of the integration constants, and by assuming $\Lambda > 0$, the string tension can be made positive in this model for all $r > 0$. Moreover, $\lim_{r\to\infty}\sigma(r) = \Lambda/2\kappa^2$, and hence at infinity the string tension becomes equal to the cosmological constant. However, in this case one can obtain a finite radius string configuration, with the radius R_s determined by the condition $\sigma(R_s) = 0$, and given by

$$R_s = \sqrt{\frac{3}{2}\frac{c_2}{\Lambda}} - 2c_1. \tag{50}$$

For a positive string tension at the origin $r = 0$, the string radius is also positive.

As for the mass of the string we obtain

$$m_s(R_s) = \frac{2\pi W_0}{\kappa^2}\left\{\frac{1}{2}\Lambda\left[R_s + c_1(\Lambda W_0 + 2) - \frac{3c_2W_0}{8c_1}\right] - \frac{6c_1c_2}{8c_1\left[R_s + c_1(\Lambda W_0 + 2)\right] - 3c_2W_0}\right\}. \tag{51}$$

By an appropriate choice of the integration constants, giving the boundary conditions of the fields φ and ψ for $r = 0$, one can always satisfy the condition $m_s(R_s) > 0, \forall R_s$. In the limit $R_s \to \infty$, we obtain $m_s(R_s) \approx \left(\pi W_0\Lambda/\kappa^2\right)R_s$, that is, for large distances the mass of the string linearly increases with its radius.

4.2. $\bar{V}(\xi,\psi) = \bar{V}_0\xi^2\psi^2$

Next we will consider string type solutions in the Generalized Hybrid Metric Palatini Gravity under the assumption that the potential $\bar{V}$ is given by $\bar{V} = \bar{V}_0\xi^2\psi^2 = \bar{V}_0\alpha\psi^2$, with $\bar{V}_0$ constant. Equation (19) then becomes

$$-f(R,\mathcal{R}) + R\frac{\partial f(R,\mathcal{R})}{\partial R} + \mathcal{R}\frac{\partial f(R,\mathcal{R})}{\partial \mathcal{R}} = \bar{V}_0\frac{\partial f(R,\mathcal{R})}{\partial \mathcal{R}}\left(\frac{\partial f(R,\mathcal{R})}{\partial \mathcal{R}} + \frac{\partial f(R,\mathcal{R})}{\partial R}\right) \tag{52}$$

and a particular solution for the function $f(R,\mathcal{R})$ is

$$f(R,\mathcal{R}) = \sqrt{\frac{R}{\bar{V}_0}}(R - \mathcal{R}). \tag{53}$$

For this potential the field equations describing the string-like structure take the form

$$\frac{d\alpha}{dr} = u, \qquad \frac{d\psi}{dr} = v, \tag{54}$$

$$\frac{dW}{dr} = \frac{1}{2u}\left(\frac{3v^2}{2\psi} - \bar{V}_0\alpha\psi^2\right)W, \tag{55}$$

$$\frac{du}{dr} = -\frac{3v^2}{4\psi} - \frac{\bar{V}_0\alpha\psi^2}{2}, \tag{56}$$

$$\frac{dv}{dr} = -\frac{v}{u}\left(\frac{3v^2}{4\psi} - \frac{\bar{V}_0\alpha}{2}\right) + \frac{v^2}{2\psi} + \frac{2V_0}{3}\psi^2\left(\alpha + \frac{\psi}{2}\right). \tag{57}$$

For this model the string tension is given by

$$\frac{2\kappa^2}{3}\sigma = \frac{7}{6}\bar{V}_0\alpha\psi^2 - \frac{v^2}{2\psi}. \tag{58}$$

The metric function W^2 and the string tension σ are depicted in Fig. 1, for a varying initial condition $\psi'(0) = \psi_0$, while all the other initial conditions are fixed. In this case the radial metric function is an increasing function of the radial coordinate r, and its rate of increase is strongly dependent on the variations in the numerical values of ψ_0. Similarly to the previous cases, the string tension is a monotonically decreasing function of r, and it vanishes at a finite value of r, $r = R_s$, which uniquely defines the string radius. The string radius is weakly dependent on the variation of ψ_0, however, significant variations in σ do appear for small values of r.

The variations of the potential and of the function ψ are represented in Fig. 2. $\bar{V}$ is a slowly decreasing positive function of r, strongly dependent on the initial condition for ψ'. The function ψ takes negative values, and show a strong dependence on ψ_0.

The behavior of the function $\xi^2(r)$ is depicted in Fig. 3. ξ^2 is positive for $r \in [0, R_s]$, and thus the physical nature of the gravitational coupling in the present model is guaranteed. ξ^2 is a monotonically decreasing function of r, and its variation depends significantly on the numerical values of the initial conditions for $\psi'(0)$.

5. Conclusions

In this work we studied the existence and physical properties of local $U(1)$ cosmic strings in the context of the generalized hybrid metric-Palatini gravity. The theory is an extension to General Relativity, combining both metric and Palatini formalism. A main success of the theory is the possibility to generate long-range forces that pass the classical local tests of gravity at the Solar System level, thus avoiding some problematic features of the standard $f(R)$ theories. Another interesting advantage of the theory is that it admits an equivalent scalar-tensor representation, simplifying greatly the dynamical equations. The type of strings studied in this work are local gauge strings, using an approximation to the Vilenkin-prescribed energy-momentum tensor and different potential configurations.

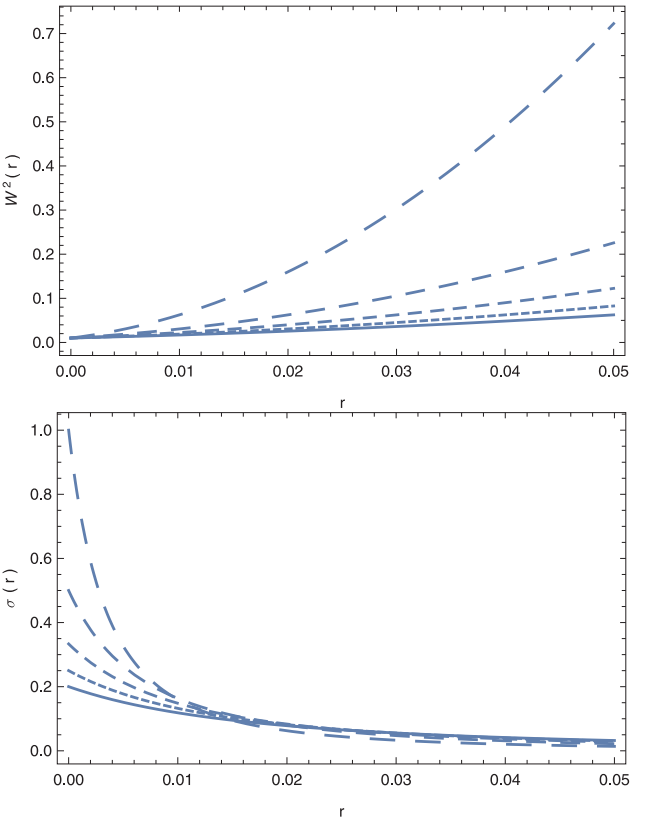


Fig. 1. Variations of the metric function $W^2(r)$ (top panel), and of the string tension $\sigma(r)$ (bottom panel) as a function of r (with all quantities in arbitrary units) for the $\bar{V}(\xi,\psi)=\bar{V}_0\xi^2\psi^2$ potential, for $\psi_0=-0.025$ (solid curve), $\psi_0=-0.020$ (dotted curve), $\psi_0=-0.015$ (short dashed curve), $\psi_0=-0.01$ (dashed curve), and $\psi_0=-0.005$ (long dashed curve), respectively. For $\bar{V}_0$ we have adopted the value $\bar{V}_0=10$, while the boundary conditions used to numerically integrate the field equations are $u_0=-0.01$, $\alpha_0=0.025$, $W(0)=0.10$, and $v_0=0.10$, respectively.

The field equations determine the string tension σ. These equations must be solved by choosing a functional form for the potential, and by imposing some appropriate boundary conditions at $r=0$ on the two scalar fields $\left(\xi^2,\psi\right)$, on their derivatives, and for $W^2(0)$. Since in the present approach these boundary conditions are arbitrary, and since the second order, strongly nonlinear system of the gravitational field equations is extremely sensitive to the variation of the boundary conditions, many types of cosmic string structures can be obtained by adopting some specific forms of V, and different sets of initial conditions.

The gravitational field equations describing cosmic string structures in generalized HMPG can be solved exactly in the simple cases of the vanishing and constant potentials.

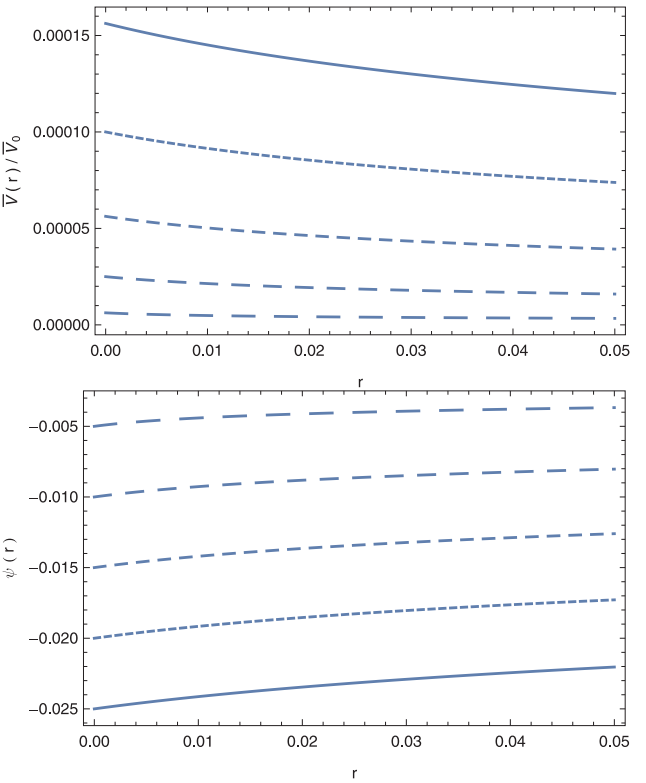


Fig. 2. Variations of the potential $\bar{V}(\xi,\psi)=\bar{V}_0\xi^2\psi^2$ (top panel), and of the function ψ (bottom panel) as a function of r (with all quantities in arbitrary units) for the $\bar{V}(\xi)=\bar{V}_0\xi^2\psi^2$ potential, for $\psi_0=-0.025$ (solid curve), $\psi_0=-0.020$ (dotted curve), $\psi_0=-0.015$ (short dashed curve), $\psi_0=-0.01$ (dashed curve), and $\psi_0=-0.005$ (long dashed curve), respectively. For $\bar{V}_0$ we have adopted the value $\bar{V}_0=10$, while the boundary conditions used to numerically integrate the field equations are $u_0=-0.01$, $\alpha_0=0.025$, $W(0)=0.10$, and $v_0=0.10$, respectively.

In the case of the constant, nonzero potential, the field equations can be solved exactly, and some simple expressions for the geometrical and physical parameters can be obtained. In this case, the string tension can be made positive by an appropriate choice of the potential. A solution with a constant string tension $\kappa^2\sigma=\Lambda/2$ can also be constructed, as well as a solution having $W(r)=W_0r$, which can describe the standard general relativistic string if $W_0^2=1-8\pi G\mu$. The string radius R_s can be uniquely defined, and it is given in terms of the constant potential, as well as two integration constants. Under the assumption that the string tension is positive in $r=0$, the string radius is also positive.

For other, more complicated forms of the potential it seems very difficult, if not impossible, to obtain exact analytical solutions of the field equations, and hence they must be solved numerically.

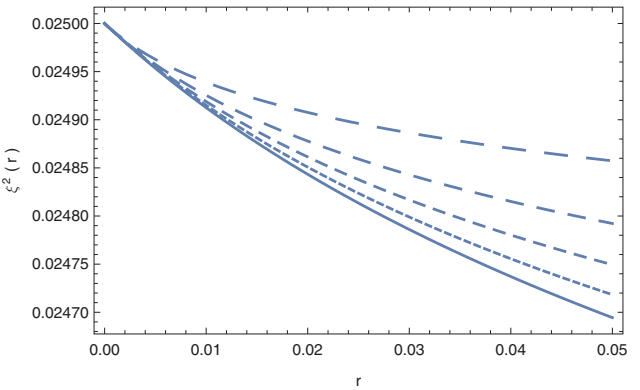


Fig. 3. Variation of ξ^2 as a function of r (with all quantities in arbitrary units) for the $\bar{V}(\xi,\psi) = \bar{V}_0\xi^2\psi^2$ potential, for $\psi_0 = -0.025$ (solid curve), $\psi_0 = -0.020$ (dotted curve), $\psi_0 = -0.015$ (short dashed curve), $\psi_0 = -0.01$ (dashed curve), and $\psi_0 = -0.005$ (long dashed curve), respectively. For $\bar{V}_0$ we have adopted the value $\bar{V}_0 = 10$, while the boundary conditions used to numerically integrate the field equations are $u_0 = -0.01$, $\alpha_0 = 0.025$, $W(0) = 0.10$, and $v_0 = 0.10$, respectively.

Since the potentials depend on at least one extra constant parameter, together with the five boundary conditions we obtain a very large boundary parameter space, containing from six to nine arbitrary parameters. The large number of parameters allows the construction of a large number of different numerical cosmic string models. However, we have restricted the set of parameters, as well as the physical nature of the solutions, by imposing three physical constraints, namely, that the string tension is positive inside the string, and it vanishes at the vacuum boundary, that the string must have a well defined and unique radius R_s, obtained from the condition $\sigma(R_s) = 0$, and that $\xi^2 > 0, \forall r \in [0, R_s]$. Even after imposing this set of restrictions, a large variety of string models in generalized HMPG theory can be obtained.

In conclusion, in the present work we have considered specific cosmic string models that are solutions of the field equations of the generalized HMPG theory. Modified gravity theories may have profound implications on the formation, properties and structure of cosmic strings, interesting and important topological objects that may have been generated in the early Universe. Hence, the theoretical investigations of strings in modified gravity models may therefore be a worthwhile pathway for future research.

Acknowledgments

FSNL acknowledges support from the Fundação para a Ciência e a Tecnologia (FCT) Scientific Employment Stimulus contract with reference CEECINST/00032/2018, and the research grants No. UID/FIS/04434/2020, No. PTDC/FIS-OUT/29048/2017 and No. CERN/FIS-PAR/0037/2019. JLR was supported by the European Regional Development Fund and the programme Mobilitas Pluss (MOBJD647).

References

1. T. P. Sotiriou and V. Faraoni, Rev. Mod. Phys. **82** (2010), 451-497 doi:10.1103/RevModPhys.82.451 [arXiv:0805.1726 [gr-qc]].
2. J. Khoury and A. Weltman, Phys. Rev. Lett. **93** (2004), 171104 doi:10.1103/PhysRevLett.93.171104 [arXiv:astro-ph/0309300 [astro-ph]].
3. J. Khoury and A. Weltman, Phys. Rev. D **69** (2004), 044026 doi:10.1103/PhysRevD.69.044026 [arXiv:astro-ph/0309411 [astro-ph]].
4. G. J. Olmo, Int. J. Mod. Phys. D **20** (2011), 413-462 doi:10.1142/S0218271811018925 [arXiv:1101.3864 [gr-qc]].
5. D. Sáez-Chillón Gómez, Phys. Lett. B **814** (2021), 136103 doi:10.1016/j.physletb.2021.136103 [arXiv:2011.11568 [gr-qc]].
6. T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Phys. Rev. D **85** (2012), 084016 doi:10.1103/PhysRevD.85.084016 [arXiv:1110.1049 [gr-qc]].
7. T. Harko and F. S. N. Lobo,
8. S. Capozziello, T. Harko, T. S. Koivisto, F. S. N. Lobo and G. J. Olmo, Universe **1** (2015) no. 2, 199-238 doi:10.3390/universe1020199 [arXiv:1508.04641 [gr-qc]].
9. T. Harko and F. S. N. Lobo, Int. J. Mod. Phys. D **29** (2020) no. 13, 2030008 doi:10.1142/S0218271820300086 [arXiv:2007.15345 [gr-qc]].
10. S. Weinberg, Phys. Rev. Lett. **19** (1967), 1264-1266 doi:10.1103/PhysRevLett.19.1264
11. A. Salam and J. C. Ward, Nuovo Cim. **11** (1959), 568-577 doi:10.1007/BF02726525
12. A. Salam and J. C. Ward, Phys. Lett. **13** (1964), 168-171 doi:10.1016/0031-9163(64)90711-5
13. U. Amaldi, W. de Boer, P. H. Frampton, H. Furstenau and J. T. Liu, Phys. Lett. B **281** (1992), 374-382 doi:10.1016/0370-2693(92)91158-6
14. R. Jeannerot, J. Rocher and M. Sakellariadou, Phys. Rev. D **68** (2003), 103514 doi:10.1103/PhysRevD.68.103514 [arXiv:hep-ph/0308134 [hep-ph]].
15. N. D. Mermin, Rev. Mod. Phys. **51** (1979), 591-648 doi:10.1103/RevModPhys.51.591
16. I. Chuang, B. Yurke, R. Durrer and N. Turok, Science **251** (1991), 1336-1342 doi:10.1126/science.251.4999.1336
17. M. M. Salomaa and G. E. Volovik, Rev. Mod. Phys. **59** (1987), 533-613 [erratum: Rev. Mod. Phys. **60** (1988), 573-573] doi:10.1103/RevModPhys.59.533
18. A. A. Abrikosov, Sov. Phys. JETP **5** (1957), 1174-1182
19. P. W. Higgs, Phys. Rev. Lett. **13** (1964), 508-509 doi:10.1103/PhysRevLett.13.508
20. A. Vilenkin, Phys. Rev. D **23** (1981), 852-857 doi:10.1103/PhysRevD.23.852
21. J. L. Rosa, J. P. S. Lemos and F. S. N. Lobo, Phys. Rev. D **98** (2018) no. 6, 064054 doi:10.1103/PhysRevD.98.064054 [arXiv:1808.08975 [gr-qc]].
22. João Luís Rosa, José P. S. Lemos, Junction conditions for generalized hybrid metric-Palatini gravity with applications, Physical Review D. Volume 104, Issue 12, article id. 124076

New evidence of the azimuthal alignment of quasars spin vector in Large Quasar Groups and cosmic strings

R. J. Slagter

ASFYON, Astronomisch Fysisch Onderzoek Nederland,
former: University of Amsterdam, Dept. Theor. Phys., The Netherlands
E-mail: info@asfyon.com

There has been observational evidence about spin axes of quasars in large quasar groups correlated over hundreds of Mpc. This is seen in the radio spectrum as well as in the optical range. There is not yet a satisfactory explanation of this "spooky" alignment. This alignment cannot be explained by mutual interaction at the time that quasars manifest themselves optically. A cosmological explanation could be possible in the formation of superconducting vortices (cosmic strings) in the early universe, just after the symmetry-breaking phase of the universe. We gathered from the NASA/IPAC and SIMBAD extragalactic databases the right ascension, declination, inclination, position angle and eccentricity of the host galaxies of 3 large quasar groups to obtain the azimuthal and polar angle of the spin vectors. The alignment of the azimuthal angle of the spin vectors of quasars in their host galaxy is confirmed in the large quasar group U1.27 and compared with two other groups in the vicinity, i.e., U1.11 and U1.28, investigated by Clowes (2013). It is well possible that the azimuthal angle alignment fits the predicted azimuthal angle dependency in the theoretical model of the formation of general relativistic superconducting vortices, where the initial axially symmetry is broken just after the symmetry breaking of the scalar-gauge field.

Keywords: Quasar groups; alignment spin vectors; host galaxy; cosmic strings; scalar-gauge field.

1. Introduction

A large quasar group (LQG) is a cluster of quasars that makes the largest astronomical structures in the current universe. Their sizes can be of the order of hundreds of Mpc. Astronomers believe that a quasar is an active galactic nuclei (AGN) with a vibrant eruption of radiation both optical and in radio range originated by a spinning (Kerr-) black hole, surrounded by an accretion disk. A LQG has an internal non-uniform distribution of spin vectors seen in the radio spectrum[1] and the optical spectrum as observed by Hutsemekers et al.[2,3] This coherence is mysterious and cannot be explained by mutual interaction at the time scale of primordial galaxies formation but rather by use of a more advanced method[4–6] In a recent study, Slagter[7] found that the azimuthal angle of the spin vector of quasars in their host galaxies in six quasar groups, show preferred directions. This is demonstrated through an emergent azimuthal angle dependency of the general relativistic Nielsen-Olesen (NO) vortices at the point after the symmetry breaking at grand unified theory (GUT)-scale. This review focuses on three more LQG, studied by

Clowes.[8,9] In section 2 we outline the theoretical model and in section 3 we present a theoretical proof of the breaking of the axial symmetry and in section 4 we present the data analysis.

2. The model

2.1. *The U(1) scalar-gauge field*

Superconductivity can be described by the famous Ginsburg-Landau model.[10] It is a beautiful example of a quantum field model, where a scalar field acts as an order parameter. This construction was made before the introduction of the scalar Higgs field of the Standard Model of particle physics! For the type II superconductivity of the Ginzburg-Landau (GL) theory,[11,12] the electro-magnetic (EM) gauge invariance is broken and the well-known Meissner effect occurs.[14] One says that the phase symmetry is spontaneously broken and the EM field acquires a length scale, which introduces a penetration depth of the gauge field A_μ in the superconductor and a coherence length of Φ. In the relativistic case one says that the photon acquires mass. Because we have three space dimensions, these solutions of the GL theory behave like magnetic flux vortices (or Nielsen-Olesen (NO) strings[13]) extended to tubes and carry a quantized magnetic flux $2\pi n$, with n an integer, the topological charge or winding number of the field. It was discovered by Abrikosov[14] that these vortices can form a lattice. These localized vortices (or solitons) in the GL-theory are observed in experiments. The phenomenon of magnetic flux quantization in the theory of superconductivity is characteristic for so-called ordered media. After the discovery of the electroweak unification, one could formulate the relativistic vortex solution, nicely expressed by Nielsen and Olesen. We write the scalar field and gauge field as

$$\Phi = \eta X(t,r)e^{in\varphi}, \qquad A_\mu = \frac{n}{e}\big[P(t,r)-1\big]\nabla_\mu\varphi,. \tag{1}$$

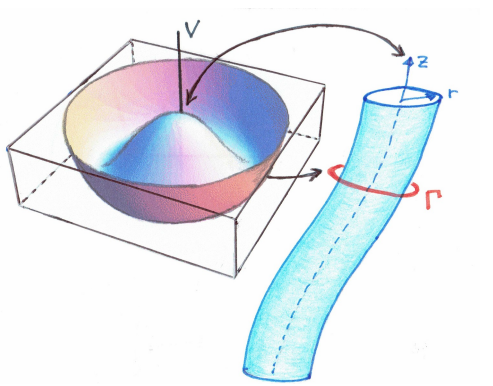


Fig. 1. Mapping the degenerated minima of the potential to position space.

The model is invariant under the group U(1) of local gauge transformations (of the second kind)

$$\Phi(\mathbf{x}) \to e^{i\chi(x)}\Phi(x), \qquad A_a(\mathbf{x}) \to A_a(x) + \frac{1}{e}\partial_a\chi(\mathbf{x}). \tag{2}$$

The scalar potential is the well-known "mexican hat" potential (see figure 1) $V(\Phi) = \frac{1}{8}\lambda\left(\Phi\Phi^* - \eta^2\right)^2$ with η the vacuum expectation value. Further, $\frac{m_\Phi}{m_A} = \frac{e^2}{\lambda}$ is the ratio of the scalar to gauge masses. This potential leads to a nonzero η and spontaneously breaking of the U(1) symmetry (note that the parameters are in general temperature dependent[a]. We say that the "false" vacuum for $\Phi = 0$ is broken and the true vacuum is then at the lowest points of the potential. The points of the false vacuum form one or more continuous curves or filaments in space. Such a filament of non-zero Higgs field energy is called a (cosmic-) string (CS) and first proposed by Kibble (1976). The mass and dimensions of the CS are largely determined by the energy scale at which the relevant phase transition occurs. The mass in the GUT scale can be of the order $10^{21} gcm^{-1}$ and there is no restriction for the length. The forces existing between the vortices are the electromagnetic- and scalar force. When the vortices get close together, the problem becomes non-linear and the resulting forces depends on the ratio $\frac{e^2}{\lambda}$. For details, see, for example, the text book of Felsager.[15]

The vortex solution possesses mass, so it will couple to gravity. The resulting self-gravitating cosmic strings (CS) still show all the features of superconductivity, but the stability conditions complicate considerably. The stability of the formed lattices depends critically on the parameters of the model, certainly when gravity comes into play. The force between the gauged vortices depends on the the strength of the self-interaction potential of the Higgs field, the gauge-coupling constant, the energy scale at which the phase transition takes place and the spacetime structure. When the mass of the Higgs field is greater than the mass of the gauge field, vortices will repel each other. So gravity could balance the vortices. The energy of the vortex

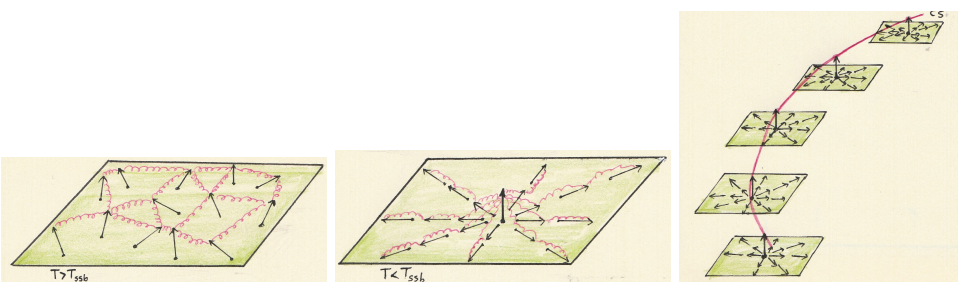


Fig. 2. Formation of the cosmic string. When the temperature drops, not all the "points" in the configuration can follow the dictated new symmetry.

[a]The string structure is cylindrical symmetric, so one uses the polar coordinates (t, r, z, φ). This choice is very convenient, when one considers rotating compact objects, such as a Kerr spacetime.

grows by increasing multiplicity n, so configurations with $n > 1$ can be seen as multi-soliton states and it is energetically favourable for these to decay into n well separated $n = 1$ solitons. Vortices with high multiplicity can be formed during the symmetry breaking. The total vortex number n is the sum of multiplicities $n_1, n_2, ..$ of isolated points (zero's of Φ).[12]

2.2. *The high-frequency approximation*

A linear approximation of wavelike solutions of the Einstein equations is not adequate when one is dealing with high curvature (or high energy scale), i.e., close to the horizons of black holes or in the early stage of the universe at the time of mass formation by the Higgs mechanism. There will be a "back-reaction" on the background spacetime. There is a powerful approximation method which can deal with these non linearities: the multiple-scale method. Pioneer work was done by ChoquetBruhat.[16] One expands the relevant fields[17]

$$V_i = \sum_{n=0}^{\infty} \frac{1}{\omega^n} F_i^{(n)}(\mathbf{x}, \xi), \tag{3}$$

where ω represents a dimensionless parameter ("frequency"), which will be large. Further, $\xi = \omega\Theta(\mathbf{x})$, with Θ a scalar (phase) function on the manifold. The small parameter $\frac{1}{\omega}$ can also be the ratio of the characteristic wavelength of the perturbation to the characteristic dimension of the background. On warped spacetimes it could also be the ratio of the extra dimension to the background dimension. In the vacuum case, we expand the metric

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \frac{1}{\omega} h_{\mu\nu}(\mathbf{x}, \xi) + \frac{1}{\omega^2} k_{\mu\nu}(\mathbf{x}, \xi) + ..., \tag{4}$$

where we defined

$$\frac{dg_{\mu\nu}}{dx^\sigma} = g_{\mu\nu,\sigma} + \omega l_\sigma \dot{g}_{\mu\nu}, \quad g_{\mu\nu,\sigma} = \frac{\partial g_{\mu\nu}}{\partial x^\sigma}, \quad \dot{g}_{\mu\nu} = \frac{\partial g_{\mu\nu}}{\partial \xi}, \tag{5}$$

with $l_\mu = \frac{\partial \Theta}{x^\mu}$. One then says that

$$V_i = \sum_{n=-m}^{\infty} \frac{1}{\omega^n} F_i^{(n)}(\mathbf{x}, \xi) \tag{6}$$

is an approximate wavelike solution of order n of the field equation, if $F_i^{(n)} = 0, \forall n$. One can substitute the expansion into the field equations. The Ricci tensor then expands as

$$R_{\mu\nu} \to \omega R_{\mu\nu}^{(-1)} + \left(\bar{R}_{\mu\nu} + R_{\mu\nu}^{(0)}\right) + \frac{1}{\omega} R_{\mu\nu}^{(1)} + ... \tag{7}$$

By equating the subsequent orders to zero, we obtain

$$R_{\mu\nu}^{(-1)} = 0 = \frac{1}{2} \bar{g}^{\beta\lambda} (l_\lambda l_\mu \ddot{h}_{\beta\nu} + l_\nu l_\beta \ddot{h}_{\mu\lambda} - l_\lambda l_\beta \ddot{h}_{\mu\nu} - l_\nu l_\mu \ddot{h}_{\beta\lambda}), \tag{8}$$

$$R^{(0)}_{\mu\nu} + \bar{R}_{\mu\nu} = 0, \qquad R^{(1)}_{\mu\nu} = 0, \qquad .... \tag{9}$$

Here we used $l_\mu l^\mu = 0$. The rapid variation is observed in the direction of l_μ. In the radiative outgoing Eddington-Finkelstein coordinates, we have $x^1 = u = \Theta(\mathbf{x}) = t - r$ and $l_\mu = (1, 0, 0, 0)$, while the bar stands for the background.

We now expand the scalar and gauge field

$$A_\mu = \bar{A}_\mu(\mathbf{x}) + \frac{1}{\omega} B_\mu(\mathbf{x}, \xi) + \frac{1}{\omega^2} C_\mu(\mathbf{x}, \xi) + ..., \tag{10}$$

$$\Phi = \bar{\Phi}(\mathbf{x}) + \frac{1}{\omega} \Psi(\mathbf{x}, \xi) + \frac{1}{\omega^2} \Xi(\mathbf{x}, \xi) + ..., \tag{11}$$

where we write the subsequent orders of the scalar field as

$$\bar{\Phi} = \eta \bar{X}(t, r) e^{in_1\varphi}, \quad \Psi = Y(t, r, \xi) e^{in_2\varphi}, \quad \Xi = Z(t, r, \xi) e^{in_3\varphi}, \tag{12}$$

with n_i the winding numbers. So we suppose that the winding numbers of the subsequent orders of the scalar field expansion coefficients are different. This can happen by the quantum fluctuations just after the symmetry breaking scale.

2.3. *The exitation and decay of the vortices*

The energy of the string in flat spacetime is given by

$$E = \frac{1}{2} \frac{n^2}{e^2 r^2} (\partial_r P)^2 + \frac{1}{2} \eta^2 (\partial_r X)^2 + \frac{1}{2} n^2 \eta^2 \frac{P^2 X^2}{r^2} + \frac{1}{8} \beta \eta^4 (X^2 - 1)^2. \tag{13}$$

The energy is proportional with n^2, so there can be no exact ground state for the string carrying multiple flux quanta (the expression changes when gravity comes into play and new features will emerge). So the configuration will dissociate into n well separated $n = 1$ vortices.

It turns out, when one evaluates the field equations, that an imprint will be left over of the azimuthal dependency of the orientation of the clustering of Abrikosov vortices. See figure 3. The azimuthal angle φ does not reach of course, the partial differential equations (PDE) in the unperturbed case. So the axial symmetry is dynamically broken. The azimuthal dependency emerges already to first order in the approximation. For example,

$${}^4T^{(0)}_{\varphi\varphi} = e^{-2\bar{\gamma}} r^2 \dot{Y} (\partial_t \bar{X} - \partial_r \bar{X}) \mathbf{cos}[(n_2 - n_1)\varphi] + e^{2\bar{\psi} - 2\bar{\gamma}} \frac{\dot{B}}{\bar{W}_1^2 \epsilon} (\partial_r \bar{P} - \partial_t \bar{P}). \tag{14}$$

However, in $T^{(1)}_{\varphi\varphi}$ there appears terms like $\cos(n_2 - n_1)\varphi$ and $\cos(n_3 - n_1)\varphi$. The perturbative appearance of a nonzero energy-momentum component $T_{t\varphi}$ can be compared with the phenomenon of bifurcation along the Maclaurin-Jacobi sequence of equilibrium ellipsoids of self-gravitating compact objects, signalling the onset of secular instabilities.[18] This shows a similarity with the Goldstone-boson modes of spontaneously broken symmetries of continuous groups. The recovery of the SO(2) symmetry from the equatorial eccentricity takes place at a time comparable to the

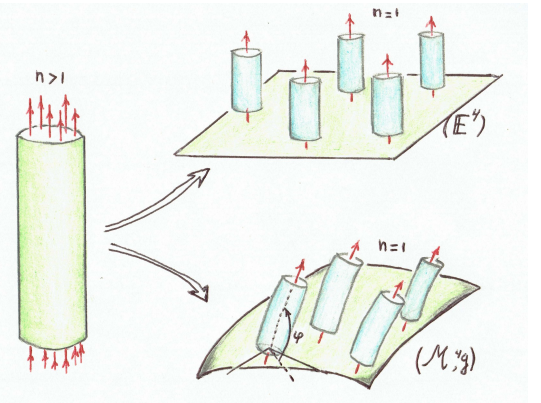


Fig. 3. Excitation and decay of a high multiplicity vortex into correlated vortices of unit flux $n = 1$. Top: the Abrikosov lattice in Euclidean space. Bottom: correlated vortices with preferred azimuthal angle φ in curved spacetime after the symmetry breaking.

emission of gravitational waves. The particular ellipsoid orientation in the frame (r, φ, z) expressed as $\varphi_0 \equiv \varphi(t_0)$, is at $t > t_0$ and determined by the transformation $\varphi \to \varphi_0 - Jt$, where J is the rotation frequency (circulation or "angular momentum") of the coordinate system. The angle φ_0 is fixed arbitrarily at the onset of symmetry breaking.

2.4. *Contribution from the bulk*

It is conjectured that the imprint of the azimuthal dependency will be to faint to contribute to observational evidence, after the expansion of the universe. However, let us consider the warped 5D spacetime

$$ds^2 = \mathcal{W}(t,r,y)^2\Big[e^{2(\gamma(t,r)-\psi(t,r))}(-dt^2+dr^2)+e^{2\psi(t,r)}dz^2 + r^2e^{-2\psi(t,r)}d\varphi^2\Big]+dy^2, \tag{15}$$

with $\mathcal{W} = W_1(t,r)W_2(y)$ is the warp factor. All standard model fields reside on the brane, while gravity can propagate into the bulk. An extended treatment of the warped five-dimensional spacetime and the effective gravitational equations can be found in Shiromizu et al.[19] and Slagter.[20] The warpfactor can be solved exactly in the case of the spacetime Eq.(15) from the 5D Einstein equations

$$\mathcal{W}^2 = W_2(y)^2W_1(t,r)^2 = \frac{e^{2\sqrt{-\frac{1}{6}\Lambda_5}(y-y_0)}}{\tau r} \cdot\Big(d_1e^{(\sqrt{2\tau})t}-d_2e^{-(\sqrt{2\tau})t}\Big)\Big(d_3e^{(\sqrt{2\tau})r}-d_4e^{-(\sqrt{2\tau})r}\Big), \tag{16}$$

If we inspect, for example, the several orders of the (t, φ) components of the energy momentum tensor

$${}^4\bar{T}_{t\varphi} = 0, \qquad {}^4T^{(0)}_{t\varphi} = \bar{X}\bar{P}\dot{Y}\mathbf{sin}[(n_2-n_1)\varphi] \tag{17}$$

$$
\begin{aligned}
{}^4T^{(1)}_{t\varphi} = &\Big[\partial_t\bar{X}Y(n_1-n_2-\bar{P})+\bar{X}(\bar{P}\partial_t Y+\epsilon B\dot{Y})\Big]\mathbf{sin}[(n_2-n_1)\varphi]\\
&+\bar{X}\bar{P}\dot{Z}\mathbf{sin}[(n_3-n_1)\varphi]+\frac{e^{2\bar{\psi}-2\bar{\gamma}}}{\bar{W}_1^2}\dot{Y}h_{14}(\partial_t\bar{X}-\partial_r\bar{X})\mathbf{cos}[(n_2-n_1)\varphi]\\
&+\bar{X}^2\bar{P}\epsilon B_0-\frac{1}{8}\beta(\bar{X}^2-\eta^2)^2h_{14}\\
&-\frac{e^{2\bar{\psi}-2\bar{\gamma}}}{2\bar{W}_1^2}h_{14}\Big[\frac{e^{2\bar{\psi}}}{r^2\bar{W}_1^2\epsilon^2}(\partial_t\bar{P}-\partial_r\bar{P})^2+\partial_r\bar{X}^2-\partial_t\bar{X}^2+e^{2\bar{\gamma}}\frac{\bar{X}^2\bar{P}^2}{r^2}\Big], \qquad (18)
\end{aligned}
$$

then it is observed that the warpfactor enters in the denominator for some terms. So the φ terms can dominate at late times (see also Eq.(14)). It will evident that an emergent imprint of a preferred azimuthal angle φ_0 is left over on the lattice of vortices when the ground state is reached ($n=1$). The configuration will become again axially symmetric.

3. A fundamental proof of the breaking of axially symmetry: The pure gravitational wave case

So far, we found that temporarily off-diagonal terms occurred in the perturbative approach of the Einstein scalar gauge field. What remains unclear is if the breaking of the axially symmetry already appears in the vacuum case like in the vicinity of the black hole spacetime. It is conjectured that the formation of primordial (Kerr-) black holes (and so quasars) happened in the early stages of the evolution of the universe before the stars were formed. Therefore, consider the radiative Vaidya spacetime in Eddington-Finkelstein coordinates [b]

$$
ds^2=-\Big(1-\frac{2M(u)}{r}\Big)du^2-2dudr+r^2(d\theta^2+\sin^2\theta d\varphi^2), \qquad (19)
$$

which is the Schwarzschild black hole spacetime with $u=t-r-2M\log(\frac{r}{2M}-1)$. Here we used $l_\mu l^\mu=0$. In the radiative coordinates, we have $x^1=u=\Theta(\mathbf{x})$ and $l_\mu=(1,0,0,0)$. From Eq.(8) we obtain

$$
h_{rr}=h_{r\theta}=h_{r\varphi}=0, \qquad h_{\varphi\varphi}=-\sin^2\theta h_{\theta\theta}. \qquad (20)
$$

From the zero-order equations Eq.(9) we obtain

$$
\ddot{k}_{rr}=0, \qquad \dot{h}_{\theta\theta}=r\partial_r\dot{h}_{\theta\theta}, \qquad \dot{h}_{\theta\varphi}=r\partial_r\dot{h}_{\theta\varphi}. \qquad (21)
$$

So one writes

$$
h_{\theta\theta}=r\alpha(u,\theta,\varphi,\xi), \quad h_{\theta\varphi}=r\beta(u,\theta,\varphi,\xi), \quad h_{\varphi\varphi}=-r\alpha\sin^2\theta. \qquad (22)
$$

Further, we have

$$
\ddot{k}_{r\theta}=\frac{1}{r}\Big(2\dot{\alpha}\cot\theta+\partial_\theta\dot{\alpha}+\frac{1}{\sin^2\theta}\partial_\varphi\dot{\beta}\Big), \qquad (23)
$$

[b] This spacetime is also applied to describe the evaporation of a black hole by hawking radiation in a quantum mechanical way.

$$\ddot{k}_{r\varphi} = \frac{1}{r}\Big(\dot{\beta}\cot\theta - \partial_\varphi\dot{\alpha} + \partial_\theta\dot{\beta}\Big), \tag{24}$$

$$\frac{dM}{du} = -\frac{\ddot{k}_{\phi\phi} + \sin^2\theta\ddot{k}_{\theta\theta}}{4\sin^2\theta} - \frac{1}{2}r\dot{h}_{uu} - \frac{1}{4}\Big(\dot{\alpha}^2 + \frac{\dot{\beta}^2}{\sin^2\theta}\Big) + \frac{1}{4}\Big(\ddot{\alpha}^2 + \frac{\ddot{\beta}^2}{\sin^2\theta}\Big). \tag{25}$$

Not all the components of $h_{\mu\nu}$ and $k_{\mu\nu}$ are physical, so one needs some extra gauge conditions. Suitable choice of α and β (Choquet-Bruhat uses, for example, $\alpha = 0, \beta = g(u)h(\xi)\sin\theta$), leads to a solution to second order which is in general not axially symmetric. We can integrate these zero order equations with respect to ξ. One obtains then some conditions on the background fields, because terms like $\int \dot{\alpha}d\xi$ disappear. From Eq. (25), we obtain

$$\frac{dM}{du} = -\frac{1}{4\tau}\int_0^\tau \Big(\dot{\alpha}^2 + \frac{\dot{\beta}^2}{\sin^2\theta}\Big)d\xi, \tag{26}$$

which is the back-reaction of the high-frequency disturbances on the mass M. τ is the period of $\dot{h}_{\mu\nu}$. This expression can be substituted back into Eq.(25). However, in the non-vacuum case, the right-hand side will also contain contributions from the matter fields. In order to obtain propagation equations for $h_{\mu\nu}$ and $k_{\mu\nu}$, one proceeds with the next order equation $R^{(1)}_{\mu\nu} = 0$. First of all, Eq.(23) and (24) are consistent with $R^{(1)}_{r\varphi} = 0$ and $R^{(1)}_{r\theta} = 0$. Further, one obtains propagation equations for α and β and for some second order perturbations, such as $k_{\varphi\varphi}$. Moreover, the (φ, θ)-dependent part of the PDE's for α and β (say $A(\theta, \varphi), B(\theta, \varphi)$) can be separated (for the case $k_{\theta\varphi} \neq 0$):

$$\partial_\varphi B + 2\sin\theta\cos\theta A + \sin^2\theta\partial_\theta A = 0, \tag{27}$$

$$\sin^2\theta\partial_{\theta\theta}A + 7\sin\theta\cos\theta\partial_\theta A + 4\cot\theta\partial_\varphi B + 2(5\cos^2\theta - 1)A + 2\partial_{\theta\varphi}B = 0. \tag{28}$$

A non-trivial simple solution is

$$A = \frac{\cos\theta(sin\varphi + \cos\varphi)}{sin^3\theta}, \quad B = \frac{\sin\varphi - \cos\varphi}{\sin^2\theta} + G(\theta), \tag{29}$$

with $G(\theta)$ arbitrary. So the breaking of the spherically and axially symmetry is evident.

4. Data analysis on quasar groups

In our model it is conjectured that the azimuthal dependency expresses itself in the φ alignment of quasar spin vectors in LQG.[7] From the NASA/IPAC extragalactic database and SIMBAD we extract for the three LQG U1.11, U1.27 and U.28 the right ascension, declination, inclination, position angle and eccentricity of the host galaxies. The 3D orientation of the spin vectors can then be calculated by[21–23] by

$$\cos^2 i = \frac{\epsilon^2 - \epsilon_0^2}{1 - \epsilon_0^2}, \tag{30}$$

$$\sin\theta = -\cos i\sin\delta \pm \sin i\sin p\cos\delta, \tag{31}$$

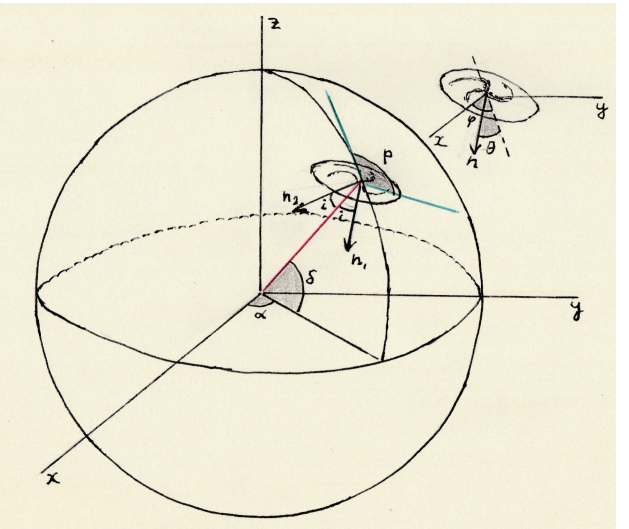


Fig. 4. The two polar coordinates (φ, θ)

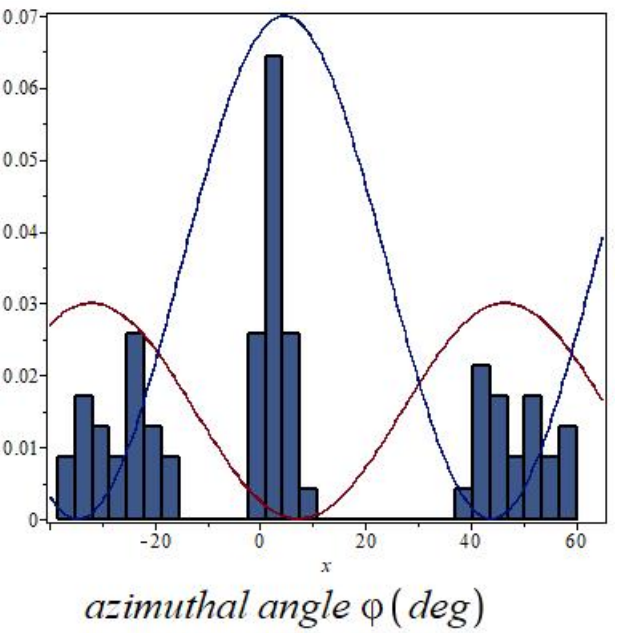


Fig. 5. Plot of the azimuthal angle φ in degrees. This shows the distribution of the azimuthal angle of the spin vectors in the LQG U1.27 (N=71) with a best-fit of two trigonometric functions with a phase shift range of 45°.

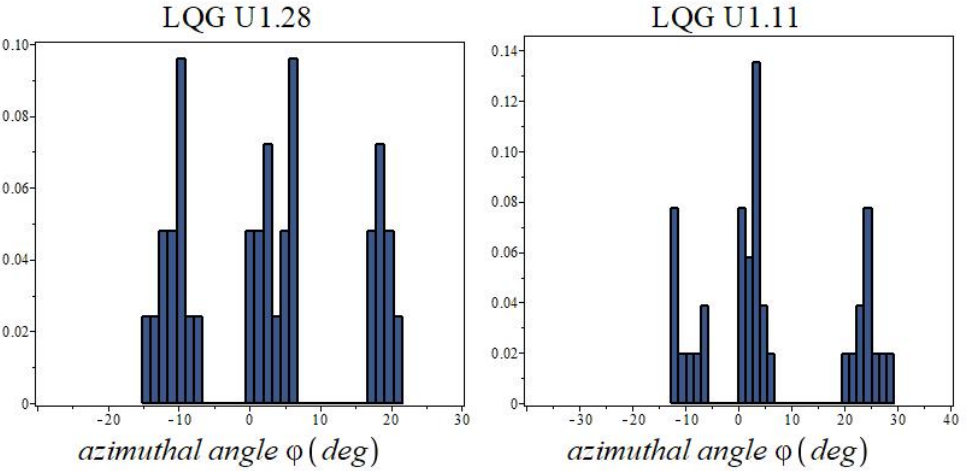


Fig. 6. Plot showing the distribution of the azimuthal angle φ for the LQG U1.28 (N = 34) and U1.11 (N = 38).

$$\sin\varphi = \frac{-\cos i\cos\delta\sin\alpha + \sin i(\mp\sin p\sin\delta\sin\alpha \mp \cos p\cos\alpha}{\cos\theta}. \tag{32}$$

We used for the intrinsic flatness (ϵ_0) the value 0.2, which is the standard value when we have no information about the morphological types of the analyzed galaxies. In figure 5 and 6 we plotted the azimuthal angle. Without statistical analysis one can conclude that the preferred orientations are evident. In the case of LQG U1.27 (see Table 1 for the data), we fitted two trigonometric functions on the distribution, which can theoretically be explained.

5. Conclusion

We find in a nonlinear approximation, an emergent azimuthal-angle dependency of Nielsen-Olesen vortices just after the symmetry breaking at GUT scale. Using a approximation scheme, the azimuthal-angle dependency appears in the first and second order field equations as trigonometrical functions $\sin(n_i - n_j)\varphi$ and $\sin(n_i - n_j)\varphi(i > j)$, with n_i the multiplicities of subsequent perturbation terms of the scalar field. Vortices with high multiplicity decay into a lattice with entangled Abrikosov vortices. The stability of this lattice of correlated flux $n = 1$ vortices with preferred azimuthal-angle is guaranteed by the contribution from the bulk spacetime by means of the warp factor: the cosmic string becomes super-massive for some time during the evolution. These so-called super-massive cosmic strings arise in a natural way in string theory or M-theory., i.e., brane-world models and are produced when the universe underwent phase transitions at energies much higher than the GUT scale, $G\mu > 1$. The imprint on the effective 4D brane spacetime can be

Table 1. Data for the LQG U1.27 (N=71) from NASA/IPAC and SIMBAD. The successive columns represent: right ascension, declination, redshift, inclination, eccentricity, position angle, azimuthal angle and polar angle. We omitted the other two values $\pi - \theta$ in the last column, belonging to the minus sign of φ.

RA	Dec	z	inc	ecc	PA(deg)	φ(rad)	θ(rad)
160.413150	14.591740	1.221	.830	.69	87	$\pm$.298 / $\pm$.427	.575 / 1.083
160.840103	14.600057	1.271	.606	.83	113	.571 / .013	.305 / .796
161.128848	16.045854	1.233	.526	.87	45	.034 / .782	.619 / .102
161.187666	15.317133	1.237	.741	.75	68	.049 / .763	.421 / .925
161.335962	14.290068	1.270	.586	.84	63	.063 / .678	.751 / .275
161.516893	14.044789	1.290	.547	.86	6	.227 / .889	-.155 / .263
161.567993	16.753510	1.282	.772	.73	48	.186 / 1.04	.779 / .294
161.601093	14.502540	1.372	.660	.80	60	.007 / .775	.322 / .792
162.056815	16.480304	1.290	.506	.88	170	.82 / .215	.336 / -.167
162.248981	12.889527	1.368	.677	.79	99	.419 / .156	.434 / .890
162.344193	15.726700	1.263	.772	.73	42	.257 / 1.05	.258 / .699
162.351272	15.698897	1.301	.435	.91	8	.129 / .762	-.190 / .306
162.409269	21.808144	1.235	.526	.87	33	.127 / .849	-.067 / .612
162.423677	15.306867	1.341	.676	.79	3	.379 / 1.00	-.175 / .240
162.449082	16.371282	1.300	.566	.85	142	.745 / .225	.080 / .589
162.505093	15.565017	1.255	.357	.94	17	.038 / .671	-.154 / .357
162.676146	16.015594	1.269	.771	.73	125	.744 / .340	.843 / .359
162.767371	16.316926	1.253	.287	.96	170	.592 / .004	.322 / -.224
162.820881	13.193341	1.337	.483	.89	102	.400 / .175	.700 / .243
162.831696	14.436524	1.315	.659	.80	54	.084 / .812	.743 / .287
162.845783	11.981222	1.309	.437	.91	159	.707 / .132	.344 / -.039
162.857213	12.796204	1.283	.641	.81	139	.779 / .275	.595 / .207
162.884258	14.937553	1.367	.756	.74	94	.354 / .211	1.014 / .494
162.918357	20.655882	1.174	.567	.85	49	.062 / .799	.743 / .082
162.937035	12.974699	1.316	.461	.90	110	.455 / .107	.654 / .207
163.041780	16.928818	1.339	.436	.91	59	.077 / .576	.656 / .083
163.092251	12.515036	1.316	.355	.94	21	.035 / .647	-.082 / .331
163.098712	14.090468	1.256	.622	.82	101	.415 / .127	.365 / .852
163.100377	20.776172	1.203	.504	.88	25	.160 / .827	-.120 / .525
163.190868	13.682636	1.356	.482	.89	60	.057 / .590	.181 / .643
163.238226	10.992647	1.266	.483	.89	89	.285 / .305	.291 / .674
163.242363	20.284854	1.253	.356	.94	41	.026 / .611	-.111 / .569
163.552829	14.959790	1.231	.678	.79	120	.625 / .184	.330 / .812
163.591268	21.358676	1.257	.606	.83	120	.579 / .164	.161 / .862
163.648532	10.304548	1.260	.385	.93	16	.084 / .671	-.064 / .271
163.677979	10.722398	1.335	.725	.76	144	.872 / .393	.247 / .550
163.694730	19.952953	1.220	.566	.85	20	.246 / .889	-.116 / .478
163.845984	13.102997	1.358	.588	.84	11	.295 / .889	.296 / -.085
163.854942	19.298998	1.201	.773	.73	146	.906 / .531	.651 / .133
163.857047	11.617507	1.293	.527	.87	119	.538 / .027	.650 / .260
163.924334	11.298387	1.331	.845	.68	88	.249 / .335	.648 / 1.042
163.984288	18.788475	1.277	.355	.94	87	.262 / .305	.027 / .683
164.046989	17.140997	1.344	.845	.68	72	.012 / .780	.505 / 1.066
164.091276	14.566966	1.243	.207	.98	129	.409 / .138	.412 / -.091
164.156218	15.013202	1.371	.787	.72	109	.541 / .147	.979 / .483
164.158278	10.052025	1.273	.435	.91	128	.544 / .023	.507 / .170
164.230688	14.829509	1.229	.625	.82	133	.696 / .257	.208 / .671
164.308435	18.798158	1.285	.382	.91	152	.614 / .106	-.133 / .484

Table 1. (*Continued*)

RA	Dec	z	inc	ecc	PA(deg)	φ(rad)	θ(rad)
164.521231	20.061417	1.273	.587	.94	72	.095 / .549	.211 / .895
164.633396	17.082247	1.286	.462	.90	17	.175 / .750	-.139 / .398
164.668752	17.904320	1.269	.527	.87	50	.062 / .700	.101 / .684
164.730550	8.230752	1.246	.677	.79	8	.401 / .953	-.025 / .199
164.869078	16.782794	1.300	.623	.82	116	.533 / .122	.272 / .829
165.025092	9.444098	1.252	.567	.85	65	.019 / .556	.348 / .666
165.070384	19.606880	1.240	.832	.69	114	.597 / .372	.423 / 1.040
165.166670	16.952878	1.300	.802	.71	127	.739 / .454	.354 / .851
165.452792	8.368669	1.196	.413	.92	9	.153 / .671	-.071 / .197
165.676652	8.655867	1.240	.411	.92	60	.046 / .481	.205 / .500
166.268588	8.759810	1.241	.566	.85	57	.072 / .606	.322 / .610
166.589189	8.686458	1.244	.548	.86	53	.098 / .618	.286 / .571
166.902560	9.020778	1.228	1.02	.55	87	.161 / .347	.860 / 1.175
166.935900	9.924171	1.225	.383	.93	14	.143 / .613	-.071 / .251
167.532914	10.802888	1.211	.248	.97	129	.374 / .052	.006 / .378
167.539961	7.868564	1.208	.740	.75	10	.508 / .967	.015 / .219
168.567405	10.390996	1.210	.693	.78	151	.803 / .472	.168 / .460
168.938775	8.249943	1.194	.547	.86	106	.349 / .012	.381 / .665
169.508801	10.550690	1.215	.526	.87	134	.548 / .234	.196 / .540
169.596744	9.084715	1.197	.787	.72	98	.303 / .018	.620 / .934
170.081761	8.984773	1.229	.641	.81	56	.193 / .613	.373 / .662
170.246993	10.185907	1.208	.384	.93	77	.085 / .271	.196 / .551
170.290711	7.999635	1.141	.461	.90	93	.195 / .143	.320 / .599
169.508801	10.550690	1.215	.526	.87	134	.548 / .234	.196 / .540
169.596744	9.084715	1.197	.787	.72	98	.303 / .018	.620 / .934
170.081761	8.984773	1.229	.641	.81	56	.193 / .613	.373 / .662
170.246993	10.185907	1.208	.384	.93	77	.085 / .271	.196 / .551
170.290711	7.999635	1.141	.461	.90	93	.195 / .143	.320 / .599

caused by wavelike disturbances triggered by the huge mass of the cosmic strings in the bulk. One conjectures that these disturbances could act as an effective dark energy field. Another possible effect can be the formation of massive KK-modes as the imprint of the 5D gravitational field on the 4D brane. We used the azimuthal-angle correlation for the explanation of the recently observed alignment of polarization axes of quasars in large quasar groups. The detailed behavior of this alignment can be explained with our model. The two different orientations perpendicular to each other in quasars groups of less richness could be a second order effect in our model.

More data of high-redshift quasars will be needed in order to test the second order effect predicted in our model.

6. The Data

The data of LQG U1.27 underlying this article are gathered in Table 1.

References

1. A. R. Taylor, and P. Jagannathan, *Alignments of radio galaxies in deep radio imaging of ELAIS N1*, Mon. Not. Roy. Astr. Soc. **459**, 459 (2016).
2. D. Hutsemekers, R. Cabanac, H. Lamy and D. Sluse, (2005) *New optical polarization measurements of quasi-stellar objects. The data*, Astron. Astrophys. **441**, 915, (2005).
3. D. Hutsemekers, L. Braibant, V. Pelgrims, and D. Sluse, *Alignment of quasar polarizations with large-scale structures*, Astron. Astrophys. **572**, A18 (2014).
4. R. J. Slagter, *Alignment of quasar polarizations on large scales explained by warped cosmic strings*, J. of Mod. Phys. **7, no 6**, 501,k (2016).
5. R. J. Slagter, *Alignment of quasar polarizations on large scales explained by warped cosmic strings PART II: The second order contribution*, J. of Mod. Phys. **8, no 2**, 163 (2017).
6. R. J. Slagter, *Evidence of cosmic strings by the observation of the alignment of quasar polarization axes on Mpc scale*, Int. J. Mod. Phys. D **27**, 185009 (2018).
7. R. J. Slagter and P. G. Miedema, *On the azimuthal alignment of quasars spin vector in large quasar groups and cosmic strings*, MNRAS **501**, 2, 3054 (2021).
8. R. G. Clowes, et al. *A structure in the early universe at z = 1.3 that exceeds the homogeneity scale of the R-W concordance cosmology*, Mon. Not. Roy. Astron. Soc., **429**, 2910 (2013).
9. R. G. Clowes, et al., *Two close large quasar groups of size = 350 Mpc at z = 1.2*, Mon. Not. Roy. Astron. Soc., **419**, 556 (2012)
10. V. Ginzburg, V. L. and L. D. Landau, *On the theory of superconductivity*, Zh. Eksp. Teor. Fiz. **20**, 1064 (1950).
11. A. Jaffa and C. Taubes, *Vortices and monopoles*, 2nd edn. (Birkhauser, Boston, 1981).
12. N. Manton and P. Sutcliffe, *Topological solitons* (Cambridge University press, Cambrigde, 2007).
13. H. B. Nielsen and P. Olesen, *Vortex line models for dual strings* Nucl. Phys. **B61**, 45, (1973).
14. A. A. Abrikosov, *On the Magnetic properties of superconductors of the second group*, Soviet Physics JETP **5**, 1174 (1957).
15. B. Felsager, *Geometry, particles and fields*, (Odense University press, Odense, 1987).
16. Y. Choquet-Bruhat, (1969) Commun. Math. Phys. **12**, 16 (1969).
17. R. J. Slagter, *Gravitational waves from spinning non-Abelian cosmic strings*, Class. and Quantum Grav. **18**, 463 (2001).
18. D. Gondek-Rosinska and E. Gourgoulhon, Jacobi-like bar mode instability of relativistic rotating bodies. Phys. Rev. **D66**, 044021 (2002).
19. T. Shiromizu, K. Maeda and M. Sasaki, *The Einstein equations on the 3-brane world.* Phys. Rev. **D62** 024012 (2000).
20. R. J Slagter and S. Pan, *A New Fate of a Warped 5D FLRW Model with a U(1) Scalar Gauge Field*, Found. of Phys. **46**, 1075 (2016).
21. P. Pajowska, W. Godlowski, Z. Zhu, J. Poliela and E. Panko, *Investigation of the orientation of galaxies in clusters: The importance, methods and results of research*, J. Cosm. Astroparticles **02**, 005 (2019).
22. W. Godlowski, *Galactic orientation within the Local Supercluster*, Mon. Not. R. Astron. Soc. **265**, 874 (1993).
23. W. Godlowski, Acta Phys. Pol. **B42**, 2323 (2011).

Summary parallel session cosmic strings I

Reinoud Jan Slagter
Astronomisch Fysisch Onderzoek Nederland (ASFYON),
University of Amsterdam (on leave)
E-mail: info@asfyon.com

Batool Imtiaz*
Department of Astronomy, School of Physical Sciences,
University of Science and Technology of China, Hefei, Anhui 230026, China
CAS Key Laboratory for Researches in Galaxies and Cosmology,
University of Science and Technology of China, Hefei, Anhui 230026, China
**E-mail: batool24@mail.ustc.edu.cn*

Keywords: Cosmic strings.

Cosmic strings (CS's) are topological defects formed at the GUT symmetry breaking scale in the Einstein-U(1) scalar- gauge field model. This model shows a surprising resemblance with superconductivity and the relativistic Nielsen-Olesen quantized magnetic flux vortex solution. In the standard model of particle physics this scalar-gauge field is responsible for the spontaneously broken symmetry (Higgs mechanism). So one could say that this quantum field with the Mexican hat potential has lived up to its reputation. In cosmological context, the confined regions of the false vacuum of the scalar field form a locus of trapped energy, i.e., a CS. The mass and dimension of a CS is largely determined by the energy scale at which the phase transition takes place. It is believed that in the FLRW model a scale-invariant cosmic string-network is formed. Observational bounds, however, predict a negligible contribution of CS's to large-scale inhomogeneities such as the angular distribution in the CMB radiation. A renewed interest occurred when it was realized that CS's could be produced within the framework of superstring theory inspired cosmological models, i.e., brane-world models. Supersymmetric GUT's can even demand the existence of CS. These super-massive CS's could be produced when the universe underwent phase transitions at energies much higher than the GUT scale, so their gravitational impact increases. Although evidence of CS are not yet found, new observational windows are opened by these super-massive CS's.

In this CS1 parallel session one could present cosmic string-related subjects, such as observational results on CS, gravitational waves and CS, vortex solutions and GRT, cosmic strings and higher dimensional models.

Silva evaluates the possibility of the existence of gravitationally bound string-like objects in the framework of the general hybrid metric-Palatini gravity theory,

in which the gravitational action is represented by functions in terms of the Ricci and Palatini scalars, respectively. This theory recognizes equivalent scalar-tensor representations in two independent scalar fields. Assuming cylindrical symmetry and metric invariance, he obtains a gravitational field equation that theoretically describes the structure of the cosmic string. The physical and geometric properties of the cosmic string are determined by two scalar fields, as well as the effective field potential, which functionally depends on two scalar fields. The field equation can be solved exactly for constant and zero potential, respectively, with the corresponding string tensions taking negative and positive values. In addition, for a more general class of potential, which has an additive algebraic structure and multiplication in two scalar fields, the equations of the gravitational field are solved numerically. For each potential, he investigated the effect of changing the parameters of the potential and boundary conditions on the structure of the cosmic string. Thus, he obtains a large class of stable string configurations, the main parameters of which (string tension and radius) substantially depend on the effective field potential and boundary conditions.

Rosa studied 5-dimensional braneworld scenarios in the scalar-tensor representation of the generalized hybrid metric-Palatini gravitational theory. He starts by considering a model for a brane supported purely by the gravitational scalar fields of the theory and then considers other distinct cases where the models are also supported by an additional matter scalar field. He investigates the stability of the gravity sector and shows that the models are all robust against small fluctuations of the metric. In particular, in the presence of the additional scalar field, he finds that the profile of the gravitational zero mode may be controlled by the parameters of the model, being also capable of developing internal structure.

Slagter tried to explain the observational evidence of the correlation of the spin axes of quasars in large quasar groups over hundreds of Mpc. This is seen in the radio spectrum as well as in the optical range. There is not yet a satisfactory explanation of this "spooky" alignment. This alignment cannot be explained by mutual interaction at the time that quasars manifest themselves optically. A cosmological explanation could be possible in the formation of superconducting vortices (cosmic strings) in the early universe, just after the symmetry-breaking phase of the universe. He gathers from the NASA/IPAC and SIMBAD extragalactic databases the right ascension, declination, inclination, position angle and eccentricity of the host galaxies of 3 large quasar groups to obtain the azimuthal and polar angle of the spin vectors. The alignment of the azimuthal angle of the spin vectors of quasars in their host galaxy is confirmed in the large quasar group U1.27 and compared with two other groups in the vicinity, i.e., U1.11 and U1.28, investigated by Clowes (2013). It is well possible that the azimuthal angle alignment fits the predicted azimuthal angle dependency in the theoretical model of the formation of general relativistic superconducting vortices, where the initial axially symmetry is broken just after the symmetry breaking of the scalar-gauge field.

Blasi presented research in gravitational wave astronomy which could open a new window of exploration in fundamental physics. The recent analysis by the NANOGrav collaboration based on the 12.5-year pulsar-timing data set has shown evidence for a common red process which is compatible with a stochastic background of gravitational waves. He discusses the interpretation of this signal in terms of physics beyond the Standard Model with focus on cosmic strings, highly energetic topological defects arising as a consequence of cosmological phase transitions in the early Universe.

Scaling solutions of wiggly cosmic strings

A. R. R. Almeida

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal, and
Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre 687, 4169-007 Porto, Portugal
E-mail: Ana.Almeida@astro.up.pt

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

Cosmic string networks form during cosmological phase transitions as a consequence of the Kibble mechanism. The evolution of the simplest networks is accurately described by the canonical Velocity Dependent One-Scale (VOS) model. However, numerical simulations have demonstrated the existence of significant quantities of short-wavelength propagation modes on the strings, known as wiggles, which motivated the recent development of a wiggly string extension of the VOS. Here we summarize recent progress in the physical interpretation of this model through a systematic study of the allowed asymptotic scaling solutions of the model. The modeling mainly relies on three mechanisms: the universe's expansion rate, energy transfer mechanisms (e.g., the production of loops and wiggles), and the choice of the scale in which wiggles are coarse-grained. We consider the various limits in which each mechanism dominates and compare the scaling solutions for each case, in order to gain insight into the role of each mechanism in the overall behavior of the network. Our results show that there are three scaling regimes for the wiggliness, consisting of the well-known Nambu-Goto solution, and non-trivial regimes where the amount of wiggliness can grow as the network evolves or, for specific expansion rates, become a constant. We also demonstrate that full scaling of the network is more likely in the matter era than in the radiation epoch, in agreement with numerical simulations.

Keywords: Cosmology; Topological Defects; Cosmic Strings; Wiggly Strings.

1. Introduction

Cosmic string networks arise in many theories of unification beyond the standard model. In the context of cosmology, they are predicted to have been formed in the early universe as a consequence of the Kibble mechanism.[1] Due to the non-linearity and non-trivial interactions intrinsic to these networks, a complete quantitative description of their evolution, as well as of their observational consequences, proves to be a difficult task.[2]

An accurate description of cosmic strings is accomplished through two complementary approaches: numerical simulations (Nambu-Goto or Abelian-Higgs) and

analytic models. The analytic modeling of string networks relies on the statistical physics of the network so as to derive a description of its thermodynamics. The analytic framework for the simplest Nambu-Goto string networks is the Velocity Dependent One-Scale (VOS) model.[3,4] As the direct comparisons with Nambu-Goto and Abelian-Higgs numerical simulations[5–8] demonstrate, the VOS is able to capture the key large-scales properties of the network extremely well. The VOS is built upon the one-scale assumption embodied in the original model developed by Kibble,[9] which expresses the evolution of the network in terms of a single length scale. This length scale can be the inter-string distance, the correlation length ξ, or the average curvature radius R (note these length scales are expected to be identical in this formalism). Various frameworks attempting to describe the dynamics of string defects have been developed, like the three-scale approach,[10] but it was only with the advent of the VOS that a model capable of making robust predictions in various cosmological epochs emerged. The key difference in the VOS framework lies in an additional macroscopic quantity, a root-mean-square (RMS) velocity, whose evolution is described by an extension of the Newton's Second Law. The equations of the standard Nambu-Goto VOS model are rigorously deduced from the Nambu-Goto effective action.

In spite of the successes of the VOS, the model fails to give an accurate description of realistic networks of strings, whose worldsheets are expected to have additional degrees of freedom. In particular, the aforementioned numerical simulations of cosmic strings in expanding universes have demonstrated the existence of great amounts of short-wavelength propagation modes on the strings, known as wiggles, on scales orders of magnitude below the correlation length. The presence of small-scale structure is simply a byproduct of the energy loss phenomena of the network, and thus, if one aims to have an accurate analytic framework of string networks, it is imperative that these small-scale wiggles are accounted for. This motivated the development of a wiggly extension of the VOS that describes the evolution of both large-scale and small-scale properties of string networks.[11,12]

Here we continue the work recently reported in Ref. 13 studying the properties of wiggly cosmic strings, through an improvement of its physical interpretation and modeling. We carry out a systematic study of the asymptotic scaling solutions of the wiggly generalization of the VOS, so as to explore the allowed scaling regimes, other than the trivial Nambu-Goto solution, where wiggliness grows, disappears, or evolves towards a constant value. One of the open questions in the wiggly model is whether the wiggliness reaches scaling, and if so, under what physical conditions. Numerical studies[6] hint at scaling being reached at least in the matter-dominated era.

Another aim of this mathematical exploration is to determine the role that each mechanism has in the overall behavior of the network, as we know the nature of the scaling solutions is highly dependent on the dominant physical mechanism.[12] We consider the various limits in which each mechanism dominates and compare

the scaling solutions for each case. Finally, this study also aims to pave the way for a future calibration of the wiggly model. Having determined the various possible scaling regimes, and particularly the behavior of the small-scale structure, we can plan future field theory and Nambu-Goto simulations that are optimized for this calibration and able to test the consistency relations of the solutions, as well as to make numerical predictions of the wiggliness on the strings.

This paper is organized as follows. Section 2 gives a brief overview of the wiggly extension of the VOS. In Section 3, we discuss the asymptotic scaling solutions without and with cosmological expansion, followed by a discussion of their physical implications. Finally we present some conclusions and an outlook in Section 4.

2. Wiggly Strings Dynamics

A wiggly extension of the VOS is accomplished by noting that, in direct contrast with Nambu-Goto strings, wiggly strings have their local string tension T differ from the energy density in the string's local rest frame U. In particular, we can define a parameter w that ranges from 0 to 1 (unity being the value of a Nambu-Goto string), such that

$$\frac{T}{\mu_o} = w, \quad \frac{U}{\mu_o} = \frac{1}{w}, \tag{1}$$

where μ_o is the string mass per unit length. The total energy of a piece of string is given by

$$E = a \int \epsilon U d\sigma = \mu_o a \int \frac{\epsilon}{w} d\sigma, \tag{2}$$

which can be decomposed into two contributions: one for the bare string segments E_o,

$$E_o = \mu_o a \int \epsilon d\sigma, \tag{3}$$

and another due to the small-scale wiggles E_w:

$$E_w = \mu_o a \int \frac{1-w}{w} \epsilon d\sigma. \tag{4}$$

Wiggly strings can be conceived as containing a mass current that renormalizes the bare energy per unit length.[14] We define the renormalization factor

$$\mu \equiv \frac{E}{E_o} \geq 1, \tag{5}$$

which is a measure of the energy due to the wiggles of the network, known as the renormalized string mass per unit length. μ is a scale-dependent quantity, that is, it depends on the renormalization scale ℓ at which the wiggliness is being measured. Each energy contribution yields a characteristic length scale

$$\rho \equiv \frac{\mu_o}{L^2} \tag{6}$$

$$\rho_o \equiv \frac{\mu_o}{\xi^2}, \tag{7}$$

which are related by

$$\mu = \left(\frac{\xi}{L}\right)^2. \tag{8}$$

This procedure makes it clear that a quantitative description of small-scale structure implies that the one-scale assumption implicit to the VOS no longer holds. Instead, we have two distinct length scales L and ξ, which will have different dynamics as it will become apparent in their evolution equations. Thus, an averaged model for wiggly cosmic string evolution implies three evolution equations, instead of two, with this additional equation accounting for the evolution of the network's small-scale structure. We note the assumption of $\xi \sim R$ still holds in the wiggly formalism, but it can be tested numerically.[6] It should also be noted that while ξ measures a physical length, L is the length that a Nambu-Goto string with the same total energy would have, and thus should not be interpreted as a physical length.

The equations for the evolution of the network's dynamical quantities are obtained from averaging the microscopic string equations of motion, and so one needs to define the averaging procedure. There are two natural approaches to this. The first consists of averaging over the total energy, so as to give more weight to string segments with larger mass currents. For a generic quantity Q, this is defined as:

$$\langle Q \rangle = \frac{\int Q \frac{\epsilon}{w} d\sigma}{\int \frac{\epsilon}{w} d\sigma}. \tag{9}$$

An alternative approach consists of averaging over the bare string energy

$$\langle Q \rangle_0 = \frac{\int Q \epsilon d\sigma}{\int \epsilon d\sigma} \tag{10}$$

The two procedures are related by:

$$\langle Q \rangle = \frac{\langle QU \rangle_0}{\langle U \rangle_0} = \frac{\langle Q/w \rangle_0}{\mu}. \tag{11}$$

In the wiggly formalism[11,12] the first averaging method is the one adopted (and is the one we adopt in what follows).

The various interactions by which the network loses energy still need to be discussed and included in the evolution equations. In particular, we need to define the phenomenological terms that account for the energy transfers between the bare string and the wiggles. While in the standard VOS model, long string intercommutings had no effect on the evolution of the network, in wiggly strings, they lead to the formation of kinks that increase the energy due to wiggles, and thus, need to be taken into consideration. The energy transfer from the bare string to the wiggles is modeled by:

$$\left(\frac{1}{\rho_o}\frac{d\rho_o}{dt}\right)_{\text{wiggles}} = -cs(\mu)\frac{v}{\xi},, \tag{12}$$

in which s approaches 0 in the Nambu-Goto limit. Besides accounting for the gain in small-scale structure, s should also include the effects of kink decay by gravitational radiation.

Another mechanism by which the network loses energy is the production of loops. In the VOS model, this is encoded in a phenomenological parameter c, the loop chopping efficiency. The loop chopping parameters of the wiggly model are defined in analogy with the VOS definition, but now the energy transferred to the loops may come from the bare string or from the wiggles

$$\left(\frac{1}{\rho_o}\frac{d\rho_o}{dt}\right)_{\text{loops}} = -cf_o(\mu)\frac{v}{\xi}, \tag{13}$$

$$\left(\frac{1}{\rho_w}\frac{d\rho_w}{dt}\right)_{\text{loops}} = -cf_1(\mu)\frac{v}{\xi}, \tag{14}$$

such that

$$\left(\frac{1}{\rho}\frac{d\rho}{dt}\right)_{\text{loops}} = \left(\frac{1}{\rho}\frac{d\rho_0}{dt}\right)_{\text{loops}} + \left(\frac{1}{\rho}\frac{d\rho_w}{dt}\right)_{\text{loops}}. \tag{15}$$

For simplicity, one can define an overall energy loss parameter

$$\left(\frac{1}{\rho}\frac{d\rho}{dt}\right)_{\text{loops}} \equiv -cf(\mu)\frac{v}{\xi}, \tag{16}$$

which using equations (13–15) can be shown to yield:

$$f(\mu) = \frac{f_o(\mu)}{\mu} + \left(1 - \frac{1}{\mu}\right) f_1(\mu). \tag{17}$$

Note that f has an explicit dependence on μ so as to account for the fact that loop production is favored in regions of the long string network with more small-scale structure than average.[6, 15, 16] In what follows we adopt the assumptions introduced in Ref. 12

$$\begin{aligned} f_o(\mu) &= 1 \\ f(\mu) &= 1 + \eta\left(1 - \frac{1}{\sqrt{\mu}}\right) \\ s(\mu) &= D\left(1 - \frac{1}{\mu^2}\right), \end{aligned} \tag{18}$$

where two additional phenomenological parameters η and D were introduced, which should be interpreted as probabilities for small-scale structure loss and gain, respectively.

Finally, we emphasize that the renormalized string mass per unit length is defined at a renormalization scale ℓ that need not be fixed but can be time dependent (e.g., the correlation length). In fact, the choice of scale is equivalent to altering

what defines a wiggle, and consequently, must have a direct influence on the energy distribution between the bare string and the wiggles. One accounts for this by introducing the following scale-drift terms

$$\frac{1}{\mu}\frac{\partial\mu}{\partial\ell}\frac{d\ell}{dt} \sim \frac{d_m - 1}{l}\frac{d\ell}{dt}, \tag{19}$$

$$\frac{\partial v^2}{\partial\ell}\frac{d\ell}{dt} = \frac{1-v^2}{1+\langle w^2\rangle}\frac{\partial\langle w^2\rangle}{\partial\ell}\frac{d\ell}{dt}, \tag{20}$$

where $d_m(\ell)$ is the multifractal dimension of a string segment at scale ℓ. It should be emphasized that while (19) is a mere geometric identity, (20) is a consequence of ensuring energy conservation at all scales. We will also assume the following phenomenological relation

$$d_m(\mu) = 2 - \frac{1}{\mu^b}, \tag{21}$$

where b is a free parameter that imposes the transition limit between two regimes ($d_m = 2$ for Brownian networks, $d_m = 1$ for small scales).

We are now ready to derive the system of equations that compose the wiggly model[11,12]

$$2\frac{dL}{dt} = HL\left[3 + v^2 - \frac{(1-v^2)}{\mu^2}\right] + \frac{cfv}{\sqrt{\mu}} \tag{22}$$

$$2\frac{d\xi}{dt} = H\xi\left[2 + \left(1 + \frac{1}{\mu^2}\right)v^2\right] + v\left[k\left(1 - \frac{1}{\mu^2}\right) + c\left(f_o + s\right)\right] + \left[d_m\left(\ell\right) - 1\right]\frac{\xi}{\ell}\frac{d\ell}{dt} \tag{23}$$

$$\frac{dv}{dt} = \left(1-v^2\right)\left[\frac{k}{\xi\mu^2} - Hv\left(1 + \frac{1}{\mu^2}\right) - \frac{1}{1+\mu^2}\frac{\left[d_m\left(\ell\right) - 1\right]}{v\ell}\frac{d\ell}{dt}\right] \tag{24}$$

$$\frac{1}{\mu}\frac{d\mu}{dt} = \frac{v}{\xi}\left[k\left(1 - \frac{1}{\mu^2}\right) - c\left(f - f_o - s\right)\right] - H\left(1 - \frac{1}{\mu^2}\right) + \frac{\left[d_m\left(\ell\right) - 1\right]}{\ell}\frac{d\ell}{dt}, \tag{25}$$

where $H \equiv \dot{a}/a$ is the Hubble parameter (which we will assume to be of the form $a \propto t^{\lambda}$) and k the momentum parameter already introduced in the VOS formalism. k essentially quantifies the local curvature of strings. It should be noted that equations (22), (23) and (25) are not independent, being related by Eq. (8).

3. Scaling Solutions

We carry out a mathematical exploration of the landscape of scaling solutions of the wiggly model, so as to characterize the behavior of the wiggliness of the network. One important yet unanswered question in the wiggly model is whether the wiggliness is able to scale, i.e, evolve towards a constant value, and if so, what are the underlying physical conditions of the network that lead to this behavior. In spite of current

numerical simulations not giving a definitive answer to this question,[6] they hint at scaling being reached in the matter-dominated era ($\lambda = 2/3$), with the behavior in the radiation era being more complex. This may be indicative of an absence of a scaling solution or of a slower scaling regime (since there is less Hubble damping).

As shown by the system of equations (22)–(25), the evolution of the network is driven by three main mechanisms: the universe's expansion, energy loss by intercommutation, and the choice of the scale in which wiggles are coarse-grained. In what follows, we study the effects of each of them on the possible asymptotic scaling solutions by considering the case where each mechanism dominates and by doing a subsequent comparison of the obtained scaling solutions in each case. This allows us to gain insight into the role that each mechanism has on the evolution of the network.

From a physical point of view, we expect three types of scaling solution regimes. Firstly, we evidently expect the Nambu-Goto solution $\mu = 1$, as the wiggly model reduces to the VOS in the appropriate limit. Secondly, we also expect a regime where the wiggliness scales, $\mu = m_o$. Finally, a regime where small-scale structure grows $\mu \propto t^{\gamma}$ as the network evolves is also expected. In general we will assume

$$\begin{aligned} L &= L_o t^{\alpha} \\ v &= v_o t^{\beta} \\ \mu &= m_o t^{\gamma}, \end{aligned} \tag{26}$$

or alternatively, using Eq. (8):

$$\xi = L_o m_o^{1/2} t^{\alpha+\gamma/2}. \tag{27}$$

It should also be noted that we will discard ultra-relativistic $v = 1$ solutions in our analysis, as well as other branches of solutions that are mathematically allowed but have no physical meaning. This includes solutions in Minkowski space $H = 0$ (expanding universes $H \neq 0$) that hold for non-null values of the momentum parameter $k \neq 0$ (null values of momentum parameter $k = 0$). The reason behind this is we expect the VOS to hold for $k = 0$ in Minkowski space,[6,7,14,17] and in an analogous manner, we expect $k \neq 0$ when $H \neq 0$.[6]

3.1. *Scaling Solutions Without Expansion*

First, we consider the allowed scaling solutions in Minkowski space. We note that linear scaling has been observed in numerical simulations in Minkowski space.[6,18] We also note that a more detailed analysis of some of these solutions has already been published in Ref. 13.

3.1.1. *No dynamical mechanism*

We start by considering the simplest case where no dynamical mechanisms act on the network. In this case we find a trivial solution that depicts a network where all

four dynamical quantities remain constant but undetermined

$$\begin{aligned} L &= L_o \\ v &= v_o \\ \mu &= m_o \\ \xi &= \sqrt{m_o} L_o. \end{aligned} \tag{28}$$

The values of these quantities could be determined by measurement in numerical simulations. The physical interpretation of this solution is clear: the network is in equilibrium due to the absence of energy loss mechanisms, and thus its dynamical quantities have no time evolution.

3.1.2. *With running averaging scale*

In this case we assume the fractal dimension introduced in Eq. (21) with the specific choice $b = 2$, as numerical simulations suggest this value is reasonable.[6] We also assume a generic power-law shaped averaging scale

$$l = l_o t^{\delta}, \quad 0 < \delta \leq 1. \tag{29}$$

In the absence of energy loss mechanisms, we expect that the characteristic length scale of the network remains constant. We find two distinct scaling solutions, the first being the Nambu-Goto limit

$$\begin{aligned} L &= L_o \\ v &= v_o \\ \mu &= 1 \\ \xi &= L_o, \end{aligned} \tag{30}$$

which is analogous to that of Eq.(28), with the difference that μ is no longer arbitrary but restricted to the Nambu-Goto value. This solution gives a clear depiction of how, when there is no small-scale structure, a variation in the coarse-graining scale makes no difference. In other words, this solution exists for any value of δ and has no explicit dependence on its choice.

The second solution yields growing wiggliness

$$\begin{aligned} L &= L_o \\ v &= v_o t^{-\delta} \\ \mu &= m_o t^{\delta} \\ \xi &= \sqrt{m_o} L_o t^{\delta/2}, \end{aligned} \tag{31}$$

subject to the following constraint:

$$v_o^2 m_o^2 = 1. \tag{32}$$

We note that in the fixed scale limit $\delta \longrightarrow 0$ we recover the scaling exponents of Eq. (29). In addition, note that choosing our renormalization scale to be the correlation

length $l \propto \xi$, yields $\delta = \delta/2 = 0$, implying $\gamma = 0$, or in other words, the only possible solution in this case is the Nambu-Goto one. The physical interpretation of this solution is clear. The presence of small-scale structure made this solution explicitly dependent on δ. In particular, we are always able to find small-scale structure for all choices of scale, and a variation in the scale simply modifies the way energy migrates from the bare string to the wiggles, with the total energy of the network staying constant. Moreover, the growth of the wiggliness is compensated by a decrease in the velocity, which suggests that not only is μ scale-dependent, but v also depends on δ. This agrees with the mesoscopic velocity interpretation of the wiggly model, instead of the traditional interpretation of a microscopic RMS one embodied in the VOS.[11,12]

3.1.3. *With energy losses*

We now consider the case where the network is losing energy due the various energy loss mechanisms inherent to the network, as previously described. In particular, we assume the energy loss terms are modeled after Eq. (18). It is clear that the presence of energy losses must lead to a different behavior for the network's characteristic length scales L and ξ, as they are now forced to scale in time. We find two solution regimes. The first solution implies non-trivial wiggliness:

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= m_o \\ \xi &= \sqrt{m_o} L_o t, \end{aligned} \tag{33}$$

such that

$$L_o = \frac{c v_o}{2} \frac{(1+\eta) m_o^{1/2} - \eta}{m_o} \tag{34}$$

$$D\left(1 - \frac{1}{m_o^2}\right) = \eta\left(1 - \frac{1}{m_o^{1/2}}\right), \quad \eta \geq D. \tag{35}$$

This allows for a wide range of values of the wiggliness m_o, depending on the values of the small-scale structure parameters D and η. In particular, one retrieves the Nambu-Goto case $m_o = 1$ under the condition $D = \eta$. For realistic values of D and η, Eq. (35) has two real solutions and a pair of complex conjugate solutions. Having in mind that realistic values of the wiggliness imply $m_o \geq 1$, it follows that for values of the parameters such that $D/\eta < 1/4$ the only physically meaningful solution consists of $m_o = 1$, while in the range $D/\eta \in]1/4, 1[$, we find solutions with $m_o > 1$. From a physical point of view, this should be interpreted as: if the amount of wiggliness generated by the network is a small fraction of the amount that is lost, as in the case of $D/\eta < 1/4$, then no wiggliness can asymptotically survive on the string network, resulting in the Nambu-Goto solution. On the other hand, if this

ratio takes up large enough values, then the small-scale structure of the network is able to survive.

In the second regime, we find a solution where small-scale structure builds-up in the network:

$$\begin{aligned} L &= L_o t^{\alpha} \\ v &= v_o \\ \mu &= m_o t^{2-2\alpha} \\ \xi &= \sqrt{m_o} L_o t \end{aligned} \tag{36}$$

with the following constraints and conditions

$$L_o = \frac{c v_o}{2\alpha} \frac{1+\eta}{m_o^{1/2}} \tag{37}$$

$$\alpha = \frac{1+\eta}{1+D} < 1 \tag{38}$$

$$\gamma = 2\frac{D-\eta}{1+D}, \quad D > \eta, \tag{39}$$

which yields Eq. (33) in the limit where the two small-scale structure parameters balance each other $D \longrightarrow \eta$.

The set of solutions (33) (36) allows us to infer the exact physical conditions that dictate the nature of the scaling regime. As seen in Eq. (33), when the amount of small-scale structure removed (e.g., loop production) is greater than its gain (e.g., kink formation upon intercommutation), $\eta > D$, the wiggliness of the network remains constant. On the other hand, as illustrated by Eq. (36), when there is more generation of small-scale structure on the network than loss, $D > \eta$, we find that the long string wiggliness must grow. Moreover, another interesting difference between these two solutions lies in the behavior of μ and L (or equivalently, the total energy density of the network), with ξ scaling linearly ($\xi \propto t$) in both solutions. The fact that ξ maintains its linear scaling independently of the scaling of L just further emphasizes the two scale feature of the wiggly VOS extension.

3.1.4. *With energy losses and a running averaging scale*

If one allows for both energy losses and a varying renormalization scale $\delta \neq 0$, one finds two possible solutions. The first yields constant wiggliness such that:

$$\begin{aligned} L &= \frac{c(1+\eta)v_o}{2} t \\ v &= v_o \\ \mu &= 1 \\ \xi &= L, \end{aligned} \tag{40}$$

which is analogous to that of Eq.(33) with the caveat that m_o is no longer arbitrary but restricted to the Nambu-Goto case. This solution depicts, yet again, that if the network is devoid of small-scale structure a change in the coarse-graining scale makes no difference.

In the second solution, we find that the small-scale structure of the network grows as the network evolves

$$\begin{aligned} L &= L_o t^{\alpha} \\ v &= v_o t^{-\gamma} \\ \mu &= m_o t^{\gamma} \\ \xi &= \sqrt{m_o} L_o t^{\frac{1}{3}+\frac{2}{3}\alpha}, \end{aligned} \tag{41}$$

subject to

$$L_o = \frac{c(1+\eta)v_o}{2\alpha m_o^{1/2}} \tag{42}$$

$$\alpha = \frac{(\frac{2}{3}-\delta)(1+\eta)}{\frac{2}{3}(1+\eta)+2(D-\eta)} < 1, \quad D > \eta \tag{43}$$

$$\gamma = \frac{2}{3} - \frac{2}{3}\alpha = \frac{2(D-\eta)+\delta(1+\eta)}{3(D-\eta)+(1+\eta)} \tag{44}$$

$$\delta = \gamma v_o^2 m_o^2 \in \left]0, \frac{2}{3}\right[\tag{45}$$

Note that if we consider the fixed scale limit $\delta \longrightarrow 0$, we recover Eq. (36) $\gamma \longrightarrow 0$, $\alpha \longrightarrow 1$. We can also take the limit of no energy losses $D, \eta \longrightarrow 0$ on (41) where we partially recover the previous solution of Eq. (31), namely the dynamics for the velocity and wiggliness $\gamma = -\beta = \delta$, but with different behavior of both length scales. Unlike the other solutions so far, δ is restricted because Eq. (45) must always hold. Specifically, the averaging scale is now bounded from above, which can be attributed to the energy losses of the network. In other words, if the scale does not evolve slowly enough, we are unable to find small-scale structure. Furthermore, we note that the total energy of the network, more specifically the exponent α, explicitly depends on δ, hinting at an apparent scale dependence of the total energy of the network. However, one would expect this quantity to be scale-invariant, that is, the way the network loses energy should not depend on the choice of averaging scale choice. Having said this, it is clear that the physics behind this behavior needs to be explored in future studies.

3.2. *Scaling Solutions With Expansion*

We are now concerned with solutions in expanding universes, where we expect $k \neq 0$.[6]

3.2.1. *No dynamical mechanisms*

We start with the case of no energy losses and a fixed scale. In this case, we find the three expansion regimes. Firstly, we find the canonical Nambu-Goto VOS solution

$$\begin{aligned} L &= \left(\frac{k^2}{4\lambda(1-\lambda)}\right)^{1/2} t \\ v &= \left(\frac{1-\lambda}{\lambda}\right)^{1/2} \\ \mu &= 1 \\ \xi &= \left(\frac{k^2}{4\lambda(1-\lambda)}\right)^{1/2} t\,. \end{aligned} \tag{46}$$

The range of expansion rates of this solution depends on the physical interpretation of the velocity. Interpreted as a microscopic velocity, we require $v_o^2 < 1$ and we find that the radiation epoch is the limiting case $\lambda > 1/2$. This implies that scaling can be achieved in the matter-dominated era, but not in the radiation epoch, since in the latter the network does not lose enough energy by Hubble damping. Alternatively, one can also interpret this velocity as an average one, with now the physical constraint being $v_o^2 \leq 1/2$ (corresponding to the average velocity of loops in Minkowski space), leading to the matter era becoming the limiting case $\lambda \geq 2/3$.

Secondly, we find a solution of non-trivial constant wiggliness, only viable in the matter era

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= m_o \\ \xi &= \sqrt{m_o} L \\ \lambda &= \frac{2}{3} \end{aligned} \tag{47}$$

$$L_o = \frac{3k}{2(m_o^3 + m_o)^{1/2}}, \quad v_o = \frac{1}{(1+m_o^2)^{1/2}}, \tag{48}$$

which in principle holds for any value of the wiggliness m_o. We note that by choosing $m_o = 1$, we retrieve the previous solution (46). This solution also demonstrates that scaling is more easily reached in the matter era than in the radiation era, in agreement with Nambu-Goto numerical simulations.[6] In addition, the behavior depicted by this solution is analogous to the one previously identified for chiral superconducting strings.[19]

Thirdly, a regime of growing wiggliness is also allowed, which implies decreasing velocities $\beta < 0$, and $\alpha = 3/2\lambda$, as well as the consistency relations $3/2\lambda + \gamma/2 - \beta = 1$, and $\beta + \gamma \geq 0$. This narrows the range of allowed expansion rates to slower rates

than the matter era one $\lambda < 2/3$. There are two solution regimes, depending on the value of the expansion rate. For very slow expansion rates we have

$$\begin{aligned} L &= L_o t^{3/2\lambda} \\ v &= v_o t^{-\lambda} \\ \mu &= m_o t^{2-5\lambda} \\ \xi &= \frac{v_o k}{2-4\lambda} t^{1-\lambda}, \end{aligned} \tag{49}$$

subject to the following constraints:

$$m_o = \left(\frac{v_o k}{L_o (2-4\lambda)} \right)^2 \tag{50}$$

$$\lambda \leq \frac{1}{3}. \tag{51}$$

Conversely, in the intermediate expansion rate regime, we find

$$\begin{aligned} L &= L_o t^{3/2\lambda} \\ v &= v_o t^{-\gamma} \\ \mu &= m_o t^{\gamma} \\ \xi &= \sqrt{m_o} L_o t^{\lambda+\frac{1}{3}} \\ \gamma &= \frac{2}{3} - \lambda, \end{aligned} \tag{52}$$

with the following constraints:

$$v_o = \frac{2}{3} \left(\frac{L_o}{(\frac{4}{3}\lambda - \frac{4}{9})^{1/4} k} \right)^{2/3}, \quad m_o = \left(\frac{k}{(\frac{4}{3}\lambda - \frac{4}{9})^{1/2} L_o} \right)^{2/3} \tag{53}$$

$$\lambda \in \left] \frac{1}{3}, \frac{2}{3} \right[. \tag{54}$$

We note that in the limit $\lambda \longrightarrow 1/3$ this solution matches the scaling solution obtained in the slow regime. In addition, we also recover the full scaling solution (47) in the limit $\lambda \longrightarrow 2/3$, apart from differences in the coefficients for the velocity.

It is instructive to consider the particular case of the radiation epoch ($\lambda = 1/2$), where we find

$$L \propto t^{3/4}, \quad v \propto t^{-1/6} \quad \mu \propto t^{1/6} \quad \xi \propto \sqrt{m_o} L_o t^{5/6}, \tag{55}$$

These relations provide an opportunity for testing with Nambu-Goto simulations where energy loss terms can be switched off on demand.

3.2.2. *With a running averaging scale*

We consider the averaging scale previously introduced in Eq. (29). In this case we still find the Nambu-Goto solution

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= 1 \\ \xi &= \sqrt{m_o} L_o t \\ \lambda &> \frac{1}{2}, \end{aligned} \tag{56}$$

$$L_o = \left(\frac{k^2}{4\lambda(1-\lambda)} \right)^{1/2}, \quad v = \left(\frac{1-\lambda}{\lambda} \right)^{1/2}. \tag{57}$$

We observe that the introduction of a running averaging scale neither affected the scaling behavior of the previous obtained solution (46) nor did it impose any restriction on δ.

We also find a full scaling solution

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= m_o \\ \xi &= \sqrt{m_o} L_o t \\ \lambda &< \frac{2}{3} \end{aligned} \tag{58}$$

$$v_o = \left(\frac{1 + m_o^2 (2\lambda^{-1} - 3)}{1 + m_o^2} \right)^{1/2} \tag{59}$$

$$\delta = \left(1 - \frac{3}{2}\lambda \right) m_o^2 (1 + m_o^2). \tag{60}$$

This solution demonstrates how the inclusion of a renormalization scale decreased the value of the expansion rate for which full scaling can occur: this is now achieved for expansion rates lower than the matter era one.

Lastly, we find two possible solutions with growing wiggliness, which entail the same dynamics for the characteristic length scale and velocity as found for the expansion case, namely $\alpha = 3\lambda/2$, $\beta < 0$, $3/2\lambda + \gamma/2 - \beta = 1$, and $\beta + \gamma \geq 0$. For very slow expansion rates we have

$$\begin{aligned} L &= L_o t^{3/2\lambda} \\ v &= v_o t^{-\lambda} \\ \mu &= m_o t^{2-5\lambda} \\ \xi &= \sqrt{m_o} L_o t^{1-\lambda} \\ \lambda &\leq \frac{1}{3}, \end{aligned} \tag{61}$$

where the scaling coefficients are related by:

$$m_o = \left(\frac{kv_o}{L_o(2-4\lambda+\delta)}\right)^2, \tag{62}$$

while in the intermediate expansion regime we determine:

$$\begin{aligned} L &= L_o t^{3/2\lambda} \\ v &= v_o t^{-\gamma} \\ \mu &= m_o t^{\gamma} \\ \xi &= \sqrt{m_o} L_o t^{1/3+\lambda} \\ \frac{1}{3} &< \lambda < \frac{2}{3}, \end{aligned} \tag{63}$$

such that

$$v_o = \left(\frac{1}{(3\lambda-1)m_o^2}\right)^{1/2}, \quad m_o = \left(\frac{kv_o}{L_o(\frac{2}{3}+\delta)}\right)^2 \tag{64}$$

$$\gamma = \frac{2}{3} - \lambda. \tag{65}$$

Note that the two regimes match for an expansion rate $\lambda = \frac{1}{3}$.

3.2.3. *With energy losses*

We finally consider the most realistic case, where the network is losing energy in the context of an expanding universe. Firstly, we recover the canonical VOS Nambu-Goto solution

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= 1 \\ \xi &= L_o t, \end{aligned} \tag{66}$$

$$L_o = \left(\frac{k(k+c)}{4\lambda(1-\lambda)}\right)^{1/2}, \quad v = \left(\frac{k(1-\lambda)}{\lambda(k+c)}\right)^{1/2}. \tag{67}$$

We note that Eq. (46) is a particular case of Eq. (66) for which $c = 0$, which can be physically interpreted as the network's intrinsic energy losses being unnecessary for the network to reach linear scaling because Hubble damping is enough to ensure this.

Secondly, there is also a non-trivial wiggliness solution

$$\begin{aligned} L &= L_o t \\ v &= v_o \\ \mu &= m_o \\ \xi &= \frac{k}{\lambda v_o(1+m_o^2)} t, \end{aligned} \tag{68}$$

subject to the following constraints:

$$L_o^2 = \frac{k}{\lambda}\frac{k + c(1+\eta)m_o^2 - c\eta m_o^{3/2}}{m_o(1+m_o^2)(\lambda + (2-3\lambda)m_o^2)} \tag{69}$$

$$v_o^2 = \frac{k}{\lambda}\frac{\lambda + (2-3\lambda)m_o^2}{(1+m_o^2)(k + c(1+\eta)m_o^2 - c\eta m_o^{3/2})} \tag{70}$$

$$\begin{aligned}\left(\lambda + m_o^2(2-3\lambda)\right)\left((k+cD)\left(1-\frac{1}{m_o^2}\right) - c\eta\left(1-\frac{1}{m_o^{1/2}}\right)\right)\\ = \lambda\left(1-\frac{1}{m_o^2}\right)\left(k + (1+\eta-\eta m_o^{-1/2})m_o^2 c\right).\end{aligned} \tag{71}$$

We note that the previous solution (47) is recovered by direct substitution of $\lambda = 2/3$, and $c = 0$ on this solution. So as to understand more about its physical implications, we consider the case study of large amounts of wiggliness. Before proceeding to this analysis, we define new parameters

$$\begin{aligned} k_{eff} &\equiv k + c(D-\eta)\\ c_{eff} &\equiv c(1+\eta).\end{aligned} \tag{72}$$

This is not only convenient, but also makes sense from a physical point of view: it is known that the presence of small-scale structure on the long strings modifies its typical curvature, as well as enhances energy losses.[12] The large wiggliness limit $m_o \gg 1$ yields:

$$\begin{aligned} L_o &= \left(\frac{k c_{eff}}{\lambda m_o^3(2-3\lambda)}\right)^{1/2}\\ v_o &= \left(\frac{k(2-3\lambda)}{c\lambda m_o^2(1+\eta)}\right)^{1/2}\\ \lambda &= \frac{2k_{eff}}{c_{eff}+3k_{eff}}.\end{aligned} \tag{73}$$

This solution clearly illustrates that when there are great amounts of wiggliness on the network, the overall energy density increases while the velocity decreases, as expected. In addition, it is interesting to note that in the radiation era, the effective parameters balance one another

$$k_{eff} = c_{eff}. \tag{74}$$

This relation constitutes an opportunity for testing with Nambu-Goto numerical simulations.

Thirdly, there also exists a growing wiggliness regime. This has the same implications as found for the expanding case, namely a decreasing velocity $\beta < 0$, although $\alpha = 3\lambda/2$ no longer holds, and instead we have $\alpha > 3\lambda/2$. Physically, this

makes sense as the energy loss terms enable faster energy loss. We again distinguish two regimes according to their expansion rates. For slow expansion rates,

$$\begin{aligned} L &= L_o t^{\alpha} \\ v &= v_o t^{-\lambda} \\ \mu &= m_o t^{\gamma} \\ \xi &= \sqrt{m_o} L_o t^{1-\lambda}, \end{aligned} \tag{75}$$

$$\alpha = \frac{1}{2} \frac{2c_{eff} + \lambda(3k_{eff} - c_{eff})}{k_{eff} + c_{eff}} \tag{76}$$

$$\gamma = 2(1 - \lambda - \alpha) = \frac{2k_{eff} - \lambda(c_{eff} + 5k_{eff})}{k_{eff} + c_{eff}} \tag{77}$$

$$\gamma \in]0, 2 - 5\lambda[\tag{78}$$

$$\lambda \leq \frac{k_{eff}}{3k_{eff} + c_{eff}}. \tag{79}$$

It is worthy of note that one recovers the previous slow expansion solution (49) in the limit $c \longrightarrow 0$.

As far as the intermediate expansion rate regime is concerned, we find

$$\begin{aligned} L &= L_o t^{\alpha} \\ v &= v_o t^{-\gamma} \\ \mu &= m_o t^{\gamma} \\ \xi &= \sqrt{m_o} L_o t^{\frac{1}{3} + \frac{2}{3}\alpha} \end{aligned} \tag{80}$$

$$L_o^2 = \frac{k c_{eff}}{(2\alpha - 3\lambda)(\lambda + \frac{2}{3}\alpha - \frac{2}{3}) m_o^3} \tag{81}$$

$$v_o^2 = \frac{k(2\alpha - 3\lambda)}{c_{eff}(\lambda + \frac{2}{3}\alpha - \frac{2}{3}) m_o^2} \tag{82}$$

$$\alpha = \frac{\frac{2}{3}c_{eff} + \lambda(3k_{eff} + c_{eff})}{\frac{2}{3}c_{eff} + 2k_{eff}} \tag{83}$$

$$\gamma = \frac{2}{3} - \frac{2}{3}\alpha = \frac{2k_{eff} - \lambda(3k_{eff} + c_{eff})}{c_{eff} + 3k_{eff}} \tag{84}$$

$$\lambda \in \left] \frac{k_{eff}}{3k_{eff} + c_{eff}}, \frac{2k_{eff}}{3k_{eff} + c_{eff}} \right[, \tag{85}$$

where one is also able to retrieve the previous solution Eq. (52) in the limit $c \longrightarrow 0$. It should be emphasized that the domain of the expansion rate is now broader, when in comparison to Eq. (52), which is due to the two energy loss mechanisms acting on the network (not just Hubble damping).

Note that the two regimes match for the expansion rate

$$\lambda = \frac{k_{eff}}{3k_{eff} + c_{eff}}, \tag{86}$$

which is half of the expansion rate of Eq. (73). We also note that in the no energy losses limit $c \longrightarrow 0$, the expansion rate approaches $\lambda = 1/3$.

Finally, it is also instructive to consider a particular case study where the two effective parameter balance each other $k_{eff} = c_{eff}$ which yields in the slow regime

$$\alpha_{slow} = \frac{1+\lambda}{2}, \quad \gamma_{slow} = 1 - 3\lambda, \quad \lambda \leq \frac{1}{4}, \tag{87}$$

while in the intermediate regime we find:

$$\alpha_{int} = \frac{1}{4} + \frac{3}{2}\lambda, \quad \gamma_{int} = \frac{1}{2} - \lambda, \quad \lambda \in \left]\frac{1}{4}, \frac{1}{2}\right[. \tag{88}$$

This solution leads to full scaling in the radiation epoch and growing wiggliness regimes for slower expansion rates than the radiation one. The transition between slow and intermediate regimes takes place at $\lambda = 1/4$.

4. Conclusions and Further Work

We have improved the physical interpretation of the wiggly generalization of the VOS model,[11,12] building upon a mathematical exploration of the landscape of possible scaling solutions. Not only did we successfully characterize the behavior of the wiggliness in various regimes, but we also have managed to determine the effect that each dynamical mechanism has on the scaling solutions.

In Minkowski space, we find that the network is in a trivial equilibrium solution in the absence of energy loss mechanisms. Allowing the network to lose energy, we find that the network's correlation length always maintains its linear scaling, while the behavior of the characteristic length scale depends on the dynamics of the wiggliness. If the wiggliness is constant, L scales linearly, while if the wiggliness grows in time, L exhibits slower growth specified by the energy loss parameters. These results agree with linear scaling observed in Minkowski space simulations.[6,18] Moreover, the inclusion of a renormalization scale led to a possible regime where small-scale structure grows in time, but is compensated by a decreasing velocity, with the network's total energy density remaining constant and the domain of the scale remaining unrestricted due to energy conservation. Lastly, when the network is subject to both mechanisms, we determined that the averaging scale becomes bounded from above, as a consequence of the network losing energy.

In power-law expanding universes, we determined that the three scaling regimes were possible, primarily depending on the expansion rate. For slow and intermediate expansion rates one has growing wiggliness solutions, while in fast expansion regimes one has the Nambu-Goto regime. For a specific value of expansion rate, one finds a full scaling solution. In the absence of energy losses, this expansion rate coincides with the matter era one, whereas the inclusion of energy loss mechanisms decreases this value, possibly reaching the radiation era for an appropriate value of the energy losses. Our results also demonstrate that scaling is more easily achieved in the matter era rather than the radiation epoch, in agreement with numerical simulations[6 16].[15] It is also interesting to note that the three scaling behaviors here determined have also been found for the case of chiral superconducting strings[19] and other current-carrying strings.[20] The case where all three mechanisms act on the network was not included in this work for reasons of space. Lastly, one aspect that remains unclear and should be subject to future studies, is the role of the coarse-graining scale, and in particular, the apparent scale dependence of the total energy of the network.

These analytic results pave the way for a future calibration of the wiggly model. While our results show qualitative agreement with previous Nambu-Goto numerical simulations, the current available data does not allow for a detailed comparison. Our work therefore motivates additional, higher resolution simulations that cover a broad range of expansion rates and that are able to extract accurate numerical diagnostics for the wiggliness.

Acknowledgements

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COM-PETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017.

References

1. T. W. Kibble, Topology of cosmic domains and strings, *Journal of Physics A: Mathematical and General* **9**, p. 1387 (1976).
2. A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and other Topological Defects* (Cambridge University Press, Cambridge, U.K., 1994).
3. C. J. A. P. Martins and E. P. S. Shellard, Quantitative string evolution, *Phys. Rev. D* **54**, 2535 (1996).
4. C. J. A. P. Martins and E. P. S. Shellard, Scale-invariant string evolution with friction, *Phys. Rev. D* **53**, R575 (Jan 1996).
5. J. N. Moore, E. P. S. Shellard and C. J. A. P. Martins, Evolution of abelian-higgs string networks, *Physical Review D* **65**, p. 023503 (2001).
6. C. J. A. P. Martins and E. P. S. Shellard, Fractal properties and small-scale structure of cosmic string networks, *Physical Review D* **73**, p. 043515 (2006).

7. J. R. C. C. C. Correia and C. J. A. P. Martins, Extending and Calibrating the Velocity dependent One-Scale model for Cosmic Strings with One Thousand Field Theory Simulations, *Phys. Rev. D* **100**, p. 103517 (2019).
8. J. R. C. C. C. Correia and C. J. A. P. Martins, Quantifying the effect of cooled initial conditions on cosmic string network evolution, *Phys. Rev. D* **102**, p. 043503 (2020).
9. T. W. B. Kibble, Evolution of a system of cosmic strings, *Nuclear Physics B* **252**, 227 (1985).
10. D. Austin, E. J. Copeland and T. W. B. Kibble, Evolution of cosmic string configurations, *Physical Review D* **48**, p. 5594 (1993).
11. C. J. A. P. Martins, E. P. S. Shellard and J. P. P. Vieira, Models for small-scale structure of cosmic strings: Mathematical formalism, *Phys. Rev. D* **90**, p. 043518 (Aug 2014).
12. J. P. P. Vieira, C. J. A. P. Martins and E. P. S. Shellard, Models for small-scale structure on cosmic strings. II. Scaling and its stability, *Phys. Rev. D* **94**, p. 096005 (2016), [Erratum: Phys.Rev.D 94, 099907 (2016)].
13. A. R. R. Almeida and C. J. A. P. Martins, Scaling solutions of wiggly cosmic strings, *Phys. Rev. D* **104**, p. 043524 (2021).
14. C. J. A. P. Martins, *Defect evolution in cosmology and condensed matter: Quantitative analysis with the velocity-dependent one-scale model* (Springer, 2016).
15. B. Allen and E. P. S. Shellard, Cosmic-string evolution: A numerical simulation, *Physical review letters* **64**, p. 119 (1990).
16. D. P. Bennett and F. R. Bouchet, High-resolution simulations of cosmic-string evolution. i. network evolution, *Physical Review D* **41**, p. 2408 (1990).
17. C. J. A. P. Martins, J. N. Moore and E. P. S. Shellard, Unified model for vortex-string network evolution, *Physical review letters* **92**, p. 251601 (2004).
18. M. Sakellariadou and A. Vilenkin, Cosmic-string evolution in flat spacetime, *Physical Review D* **42**, p. 349 (1990).
19. M. F. Oliveira, A. Avgoustidis and C. J. A. P. Martins, Cosmic string evolution with a conserved charge, *Physical Review D* **85**, p. 083515 (2012).
20. C. J. A. P. Martins, P. Peter, I. Y. Rybak and E. P. S. Shellard, Charge-velocity-dependent one-scale linear model, arXiv:2108.03147 (2021).

High resolution calibration of string network evolution

J. R. C. C. C. Correia

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal, and
Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre 687, 4169-007 Porto, Portugal
E-mail: jose.correia@astro.up.pt

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

The canonical velocity-dependent one-scale (VOS) model for cosmic string evolution contains a number of free parameters which cannot be obtained ab initio. Therefore it must be calibrated using high resolution numerical simulations. We exploit our state of the art graphically accelerated implementation of the evolution of local Abelian-Higgs string networks to provide a statistically robust calibration of this model. In order to do so, we will make use of the largest set of high resolution simulations carried out to date, for a variety of cosmological expansion rates, and explore the impact of key numerical choices on model calibration, including the dynamic range, lattice spacing, and the choice of numerical estimators for the mean string velocity. This sensitivity exploration shows that certain numerical choices will indeed have consequences for observationally crucial parameters, such as the loop chopping parameter. To conclude, we will also briefly illustrate how our results impact observational constraints on cosmic strings.

Keywords: Cosmic strings; Velocity-dependent One-Scale model; Abelian-Higgs string simulations; Observational consequences

1. Introduction

Topological defects are expected consequences of phase transitions in the early Universe, formed via means of the Kibble mechanism.[1] Depending on the symmetry broken different defects, with different dimension, can appear. The most studied defects are the filament-like cosmic strings. These are generically predicted in many candidates of Grand Unified Theories[2] and superstring theory,[3] and are cosmologically safe, in the sense that they are not expected to overclose the Universe.

Due to their ubiquitous nature they have become a primary target for observational facilities, both current[4,5] and forthcoming.[6,7] Analytical and observational studies of defect networks (strings included) often rely on a combination of semi-analytical modelling and simulations.

Here we present an up-to-date model calibration using high-resolution simulations, that ran in one of Europe's largest supercomputers, Piz Daint, using up to

4096 GPUs. In order to ensure robustness, we will also explore how sensitive the model calibration is to certain numerical choices. We will also discuss our recently developed visualization strategy and avenues it might open in the study of defects. Some of this work has been published in Ref. 8.

2. Simulation Setup

We begin with a brief description of our simulations. Take a Lagrangian density which, while originally invariant under $U(1)_L$ transformations, is already in the broken symmetry phase,

$$\mathcal{L} = |D_\mu \phi|^2 - \frac{\lambda}{4}(|\phi|^2 - 1)^2 - \frac{1}{4e^2} F^{\mu\nu} F_{\mu\nu} \,, \tag{1}$$

where ϕ is a complex scalar field, $D_\mu \phi = \partial\phi - ieA_\mu$ are the covariant derivatives, A_μ is a gauge field, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is a gauge field strength and the scalar and gauge couplings are λ and e, respectively. The defect network that will form corresponds to homotopically stable solutions of the equations of motion of this Lagrangian. In fact a cosmic a string is the result of a specific solution of the equations of motion, which maps ϕ to the symmetry breaking scale (here set to u; minima of potential) infinitely away from the string, and to zero at its core (maxima of potential). The equations of motion in Friedmann-Lemaitre-Robertson-Walker Universes and under the assumption of the temporal gauge ($A_0 = 0$) are

$$\ddot{\phi} + 2\frac{\dot{a}}{a}\dot{\phi} = D^j D_j \phi - \frac{a^2 \lambda}{2}(|\phi|^2 - 1)\phi \tag{2}$$

$$\dot{F}_{0j} = \partial_j F_{ij} - 2a^2 e^2 Im[\phi^* D_j \phi] \,. \tag{3}$$

Abelian-Higgs field theory simulations of cosmic strings[9,10] merely use the discretization procedure of lattice gauge theory[8]—which places scalar fields on lattice points distanced by $\Delta = 0.5$, and gauge fields on lattice links—to provide a discretized set of equations to update fields every conformal timestep. Due to the existence of Hubble damping and due to the topological nature of the potential, the initial conditions eventually relax into a configuration with multiple string networks. For simplicity, and to mimic what the fields are post-phase-transition, the initial conditions are set as follows: null on every field, except ϕ which is set to have a random phase and unit magnitude.

We will additionally set scalar and gauge couplings to $\lambda = 2$ and $e = 1$, respectively, which sets the Bogmolnyi ratio to $\frac{\lambda}{2e^2} = 1$ (a usual choice in the literature) and force the gauge and scalar core to have the same width. Given that the string radius varies as $r \propto \frac{1}{\sqrt{\lambda} a}$ and this could cause strings to "slip" through the lattice, we will make each coupling vary as

$$\lambda = \lambda_0 a^{2(1-\beta)} \qquad\qquad e = e_0 e a^{(1-\beta)} \tag{4}$$

where β is set to zero. This unphysical constant comoving width trick was first proposed in Ref. 11 and has been used throughout the literature to avoid resolution

problems. We note that while previous work has suggested that $\beta = 0$ and $\beta = 1$ dynamics seem to agree, it is unknown presently if changing the value of β would somehow pose a systematic error source in model calibration. This particular effect will be studied in a future publication and is beyond the scope of the present work.

In order to characterize the defect network that forms and its evolution throughout cosmic time, we also need to characterize the mean comoving distance between strings, ξ, and the mean squared velocity, $\langle v^2 \rangle$. For velocity estimators we will use two different forms, one first proposed in Ref. 12 and the other in Ref. 10. The first one is based on a boosted static string (detailed derivation in Ref. 10) and will be referred to as the scalar field velocity estimator. It has the following form

$$< v^2 >_{\phi} = \frac{2R}{1+R}, \qquad R = \frac{\int |\Pi|^2 \mathcal{W} dx^3}{\int (\sum_i |D_i^+ \phi|^2) \mathcal{W} dx^3}\,. \tag{5}$$

where $\mathcal{W}$ is weight function, chosen to localize the estimator on strings. We will use the Lagrangian density $\mathcal{L}$ to this end. Additionally, we will use the equation of state based velocity estimator,

$$< v^2 >_{\omega} = \frac{1}{2}\left(1 + 3\frac{\int p \mathcal{W} dx^3}{\int \rho \mathcal{W} dx^3}\right). \tag{6}$$

We remark that the work of Reef. 10 showed the second estimator to perform better in Minkowski space than the scalar based estimator, i.e. it is in better agreement with the analytical expectation for the velocity of an oscillating string in flat space. Such disagreement is lattice space dependent, and indeed an exploration of this feature in the opposite low velocity limit is published in Ref. 13.

One way to characterize the typical distance between strings is by counting the total amount of string in the box, L,

$$\xi_W = \sqrt{\frac{\mathcal{V}}{L}}\,, \tag{7}$$

where $\mathcal{V}$ is box volume, and L as mentioned the total amount of string. L can be counted in multiple ways, but we will use an integer called winding. To explain it, remember one can think of a string as a magnetic flux tube, where the flux is quantized around the string,

$$\Phi = \oint dx^{\mu} A_{\mu} = \frac{2\pi}{e} \tag{8}$$

and the integer n is the winding. If a cell face is pierced by a string, then the lattice version of the winding[14] will be non-zero (1 for the choices of λ and e made above), and therefore we can say that a segment of length Δx pierced the plaquette. Summing all the segments (taking care not to double-count) we then obtain the total length of string in the box, L.

Note that output of cells pierced by windings forms the basis of our visualization strategy. Given that this novel aspect of our simulation has thus far not been described elsewhere, we will now describe the strategy itself and what avenues it opens for the study of defect networks.

2.1. *Visualization strategy I — In-situ windings*

One of the most stringent and common bottlenecks for many High-Performance Computing simulations is related to Input/Output. Not only does computational throughput grow much faster than the typical read/write speeds of most storage solutions, but also, very often, the amount of data required to do science exceeds the amount of storage available. It is thus no surprise that in recent years there has been a push towards techniques of in-situ visualization, where the output data is heavily reduced beforehand and a smaller subset is used afterwards. There are many literature examples of successful application of such techniques.[15–19] In this section we will detail how this technique is applied in our multiGPU cosmic string simulations[20] and the resulting performance gains.

In order to use an in-situ approach to outputting string positions in the lattice, we use ParaView Catalyst version 5.8.0. There are then two relevant components: the Adaptor code (written in C++) and a Python script which applies three filters to the data and then outputs an Unstructured grid (in *.pvtu format).

In the case of the Adaptor, it begins (if necessary) by creating a vtkMultiBlock-Dataset, where each rank contains then a vtkImageData with the correct information about block extent (in number of points), lattice spacing and origin of each block or sub-domain of the grid. Afterwards a series of vtkFloatArray are either created or updated by copying the contents of 7 different arrays pinned host memory buffers updated by GPUs before-hand. Six of these arrays (one for each cell face) contain only ± 1 in case a string pierces a cell face and 0 otherwise. The last array is merely an OR of the previous six, and will be used to determine which cells are pierced by strings and should therefore be outputted.

This determination is done in the Python script, which applies a Threshold filter to the data, selecting only cells wherein the last array is non-zero. In other words, from this point on we already have a reduced dataset. The Merge Block filter is then applied to collect pieces of strings spread throughout several sub-domain. To finalize the Parallel Unstructured Grid writer is used to output data in the *.pvtu format (alongside multiple block files *.vtu).

Table 1. A summary of typical output sizes for a timestep either with Raw output of all cells (HDF5) or using only the unstructured grid outputs from our in-situ approach.

Lattice size	Output data size (MB)		Time taken (s)	
	HDF5 estimated	In-situ measured	HDF5 estimated	In-situ measured
512^3	5296	$[8.7, 25.0]$	0.9	$[2.6, 3.5]$
1024^3	42368	$[11.0, 126.0]$	0.9	$[2.7, 3.2]$
2048^3	338944	$[33.0, 199.0]$	3.0	$[2.9, 5.6]$
4096^3	2711552	$[99.0, 144.0]$	23.9	$[4.1, 8.2]$

Note: Given that the in-situ approach size depends on the number of cells that are part of a string, we indicate the range of sizes obtained over the conformal time ranges where we output.

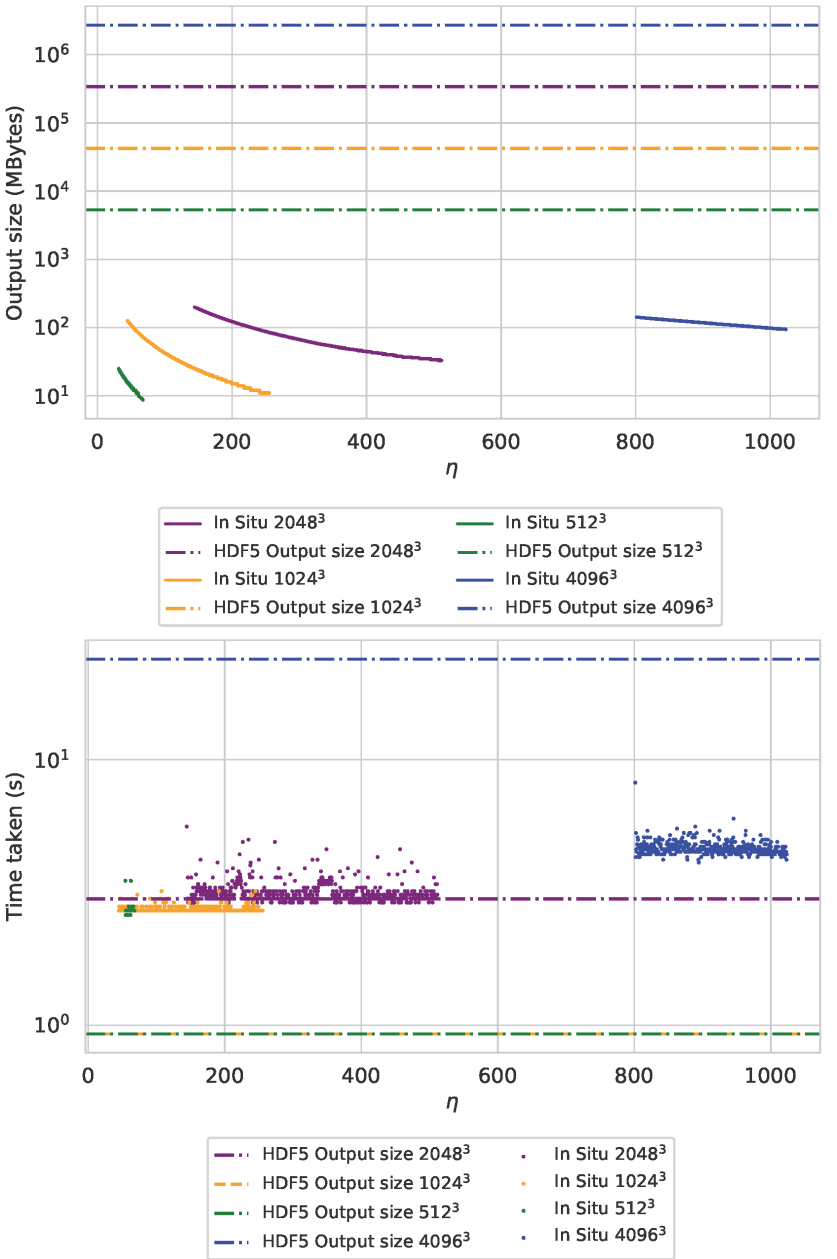


Fig. 1. Top: Size of raw HDF5 data output (windings for six cell faces, dashed lines) and of that for the In-Situ approach, where Unstructured grids are output (full lines), for lattice sizes: 4096^3 (blue), 2048^3 (purple), 1024^3 (orange) and 512^3 (green). Bottom: corresponding output times.

Note that no rendering is done so far: the in-situ aspect outputs heavily reduced data, but it does not produce a string centerline. The next section will describe how to create such a centerline by connecting each cell of each individual string and applying an averaging to string positions, such that the effect of taxicab geometry is greatly reduced.

Before we move on, we must detail the gains obtained from using this particular approach to data output, by comparing to the previous method of output of the simulation, which was to output all cells throughout the domain in HDF5 files (file-per-process approach). All benchmarks were conducted at Piz Daint, the Swiss National Supercomputer, the $4th$ most powerful supercomputer in Europe at the time of writing.[21] This machine features 5704 nodes, each equipped with an Nvidia Tesla P100, together with a *scratch* partition which is a Cray Sonexion 3000 Lustre filesystem with 8.8PB capacity. The system contains 40 object storage targets, and can handle a file per process approach well as long as there are not thousands of files in a single folder (this can be circumvented by grouping files according to rank for instance).

The first, more obvious consequence is how this results in less data being output. The smallest size of single HDF5 file with 7 single precision floating point datasets of dimension $(256, 256, 256)$ is around 662 MB in our test set. In order to determine the total typical output size for the benchmarks we are about to describe, we need to multiply the number by the number of ranks used. Keeping the problem size per process fixed. We will then run the benchmarks for in-situ with the following lattice sizes 512^3, 1024^3, 2048^3 and 4096^3 with domain decomposition of $(2, 2, 2)$, $(4, 4, 4)$, $(8, 8, 8)$ and $(16, 16, 16)$. In all benchmarks we output data every five timesteps. The obtained total timestep output sizes are shown in Table 1 and in Figure 1. We note that as the networks evolve, and the strings interact, the energy density of strings decreases which has an immediate and obvious effect on the number of cells output: they decrease with conformal time. This is translated in the total timestep size, which decreases with conformal time for every lattice size. The other important conclusion is that there is a vast reduction in total storage space required, at worst two orders of magnitude, at best four. This not only allows us to output data at a larger temporal rate (more timesteps), but also allows the possibility of adding more data to the outputs, if necessary.

We must also ask if there a performance (wall-clock time) speed-up or a performance penalty, or if, even including a costly filter like Merge Blocks, we can output data at similar rates as in the previous case. In order to do so, we will use the maximum measured writing bandwidth of our file per process approach with HDF5 files with 8, 64 and 512 writers for 512^3, 1024^3 and for both 2048^3 and 4096^3. The three maximum bandwidth are, respectively 5712, 45640 and 113300 MB/s. Note that we will use the maximum measured bandwidth of 512 writers for 4096 as it is already very close to the estimated peak bandwidth of the Lustre filesystem on Piz Daint.

Using these maximum bandwidths and the estimated storage needed for each file allows us to calculate a typical time taken. For in-situ we measure the time taken to output each timestep where data is written out. Our results, shown in table 1 and in figure 1, indicate that in-situ is definitely worth it if the file size becomes extremely large and the bandwidth is already saturated (close to the theoretical maximum bandwidth of the filesystem). This is evident when comparing the 2048^3 and 4096^3 cases where the amount of time taken is roughly comparable for 2048^3 but in the 4096^3 the amount of time taken for in-situ is roughly an order of magnitude below.

In conclusion, the technique described herein makes studying small-scale structure of strings possible with large lattices, not only reducing the amount of storage required by the simulation towards feasible amounts, but also permitting a significant speed-up, circumventing the limitation imposed by the maximum bandwidth of the filesystem.

2.2. *Visualization of string networks II — post-processing*

After the simulation runs, we can use the Unstructured grid output to construct string centerlines. The first step is to group neighboring cells via the Connectivity filter, and then use the PassArrays filter to pass all the data arrays to the input of the custom filter. The centerline script loops over each region identified by the connectivity filter and by each cell. The problem of trying to connect cells based on the order of the cellId in the Unstructured grid is that they are ordered by index (i,j,k) and not by the order they should be in a string.

The way to solve this issue is to begin in a cell in the string, ask for the neighbors according to the connectivity filter and then use the information contained in each of the winding arrays to understand if the connectivity established via cell faces is physically valid, i.e. if magnetic flux passes through that cell face. Note that we choose the positive, +1 direction for the magnetic flux but the script would work if working in the opposite direction too. Once a valid neighbor is identified, the vortexBounds for the specific cell and the next one are used to find cell center coordinates. These coordinates used to trace a vtkLine connecting one cell to the next, passing in the middle of a cell face. Cell by cell and segment by segment, the string centerline begins taking shape, as a collection of vtkLines (hence the output is of vtkPolyDataCollection type).

There are however some special cases to be mentioned as they can cause issues in the correct visualization and analysis.These are merely a consequence of the phenomenology of a string network. When two strings meet at one point (forming an X-shaped region), they can exchange ends, i.e. intercommute, and two new strings emerge immediately afterwards. On the other hand, if strings are to either meet in two points or to self-intersect, it is very likely that a closed loop forms. Loops are

Fig. 2. String cells colored by region in two close-up screenshots of 2048^3 radiation era simulation. We show in addition the output of the centerlines custom filter with smoothing via a Hanning window. A Loop (in blue) is shown at the center of the top panel screenshot. An intercomutation event (in red) is shown on the left-hand side of the bottom panel.

extremely important for observational consequences, as they shrink and dissipate energy[a] therefore generating some observational fingerprints.

Intercommutation events (or X-regions) can cause the string centerline to stop at the boundary of the box but not go through all the cells of the entire region. In effect the initial version of the custom filter would draw a single string (corresponding to two legs of the X-region, so to say). This can be avoided by comparing the number of cells used for the creation of the centerline so far and the total number of cells in the region. If one is much smaller, then the centerline drawing must restart at a cell not used previously and create the other two legs. Loops require us to close the path of the centerline, in other words, to connect the last cell and the first. Identifying

[a]Here lies a disagreement between two types of cosmic string simulation in the literature: in what type of energy do loops decay? Gravitational waves, massive radiation? This can alter the overall energy loss of the network, subsequent evolution of the network, and observational imprints. The scientific goal of the numerical techniques developed in the present work is to address this issue.

whether a loop is present can be done using the number of points and the number of cells. Should they differ by one, it is necessary to close this path.

After dealing with these two problematic cases we now have the string positions along every stair-case-like segment.This staircase behavior is no more, no less a consequence of the discrete nature of the lattice and of not having complete information about where exactly in each cell face a string lies (one assumes perfectly in the middle but this is not entirely realistic either). There is therefore some small-scale structure on the resulting strings in a scale roughly of the lattice size. Given that this is not a physical effect, we wish to remove it. In Ref. 12 the smoothed string was obtained by averaging string positions along the string path. Here we convolve each string with a window function in order to remove such artificial structure. The specific window function and the window size can be chosen in the filter. Given that numpy and scipy are used for this procedure, possible window functions by default include Hanning, Haming, Bartlett, Blackman. Figure 2 shows a collection of cells in two long strings and one loop either in cells as the resulting output of the Connectivity filter or as string centerlines (in white) smoothed over with the Hanning window function. Qualitatively this is more representative of a natural string. An example of the completed centerlines for a 4096^3 lattice simulation, with $\Delta x = 0.5$ aand $\Delta\eta = 0.1$ in radiation epoch is shown in Fig. 3.

3. Extended Velocity-dependent One-Scale model

The canonical semi-analytic model for string network evolution is the so-called Velocity dependent One-Scale model of Refs. 22, 23. It can be thought of as thermodynamic model in the sense that it describes the time evolution of two network averaged quantities, the mean string separation L, and the root mean squared velocity v. The cosmological history of these quantities is described in terms of relevant physical effects. For instance, the expansion of the Universe tends to act as a friction force on the strings (thereby affecting velocities) and also acts to dilute the network (affecting the mean separation). The model takes the following form, in comoving coordinates,

$$\frac{d\xi}{d\eta} = \frac{m\xi}{(1-m)\eta}v^2 + F(v) \tag{9}$$

$$\frac{dv}{d\eta} = (1-v^2)\left[\frac{k(v)}{\xi} - \frac{2mv}{(1-m)\eta}\right], \tag{10}$$

where $\xi = La$ is the comoving mean string separation, η is the conformal time, related to t by means of $d\eta = adt$, $\mathcal{H}$ is the conformal Hubble parameter, m is an expansion rate such that the scale factor $a \propto t^m$, and $k(v)$ and $F(v)$ are two phenomenological functions, related to the curvature of strings and to energy loss processes, respectively. In the original VOS, $k(v)$ can be analytically determined by comparison with the helicoidal string ansatz. However, for the present work, we will

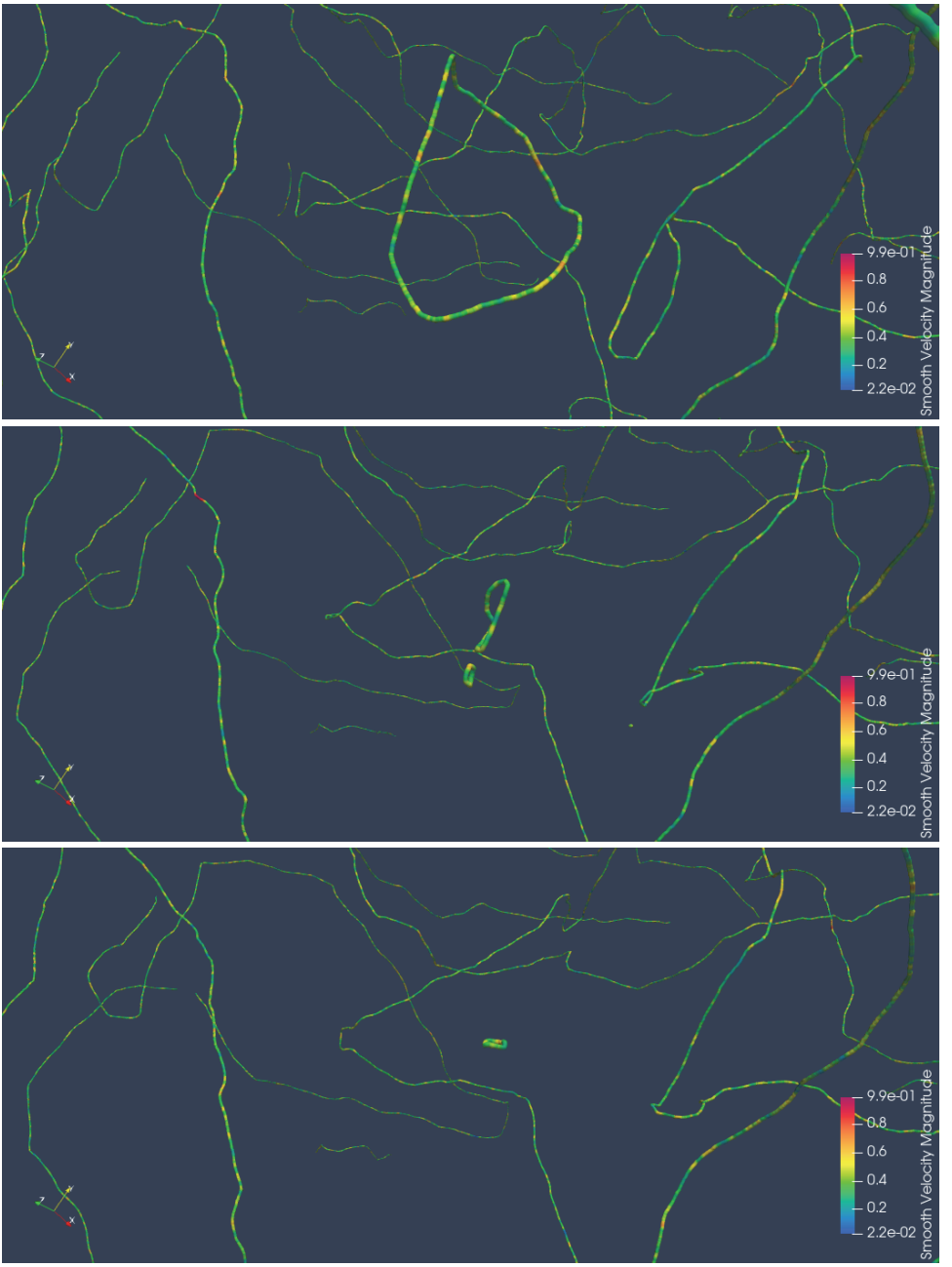


Fig. 3. Comparison of the mean rate of change of correlation length ξ/η (top left) and the mean velocity $\langle v \rangle$ (top right) with the solid lines corresponding to the calibration and the shaded regions to the uncertainty of the measurements of each estimator for three different box sizes. The bottom plots show how these differences impact the momentum parameter $k(v)$ (bottom left) and in the energy loss parameter $F(v)$ (bottom right).

allow a more general form,

$$k(v) = k_0 \frac{1-(qv^2)^\beta}{1+(qv^2)^\beta}. \tag{11}$$

where q, k_0 andβ are free parameters, which in the case of q and k_0 have a clear physical meaning. The former corresponds to the inverse of the maximum defect velocity squared $q \approx 1/v^2$ and the latter to the maximal value of $k(v)$, obtained in the low velocity limit. Note that the helicoidal ansatz corresponds to setting $q = 2$, $k_0 = \frac{\sqrt{2}\pi}{2}$ and $\beta = 3$. Letting these parameters free allows us to verify if small-scale structure on strings deviates from the Nambu-Goto expectation significantly, as this would be reflected on parameters q, k_0 and β taking different values from the analytical expectation.

The energy loss function $F(v)$ was originally devised to only include a term linearly proportional to the velocity to model energy loss via loop chopping. However, recent work on radiation on Abelian-Higgs strings,[10] then prompted the authors of Ref. 24 to introduce a radiative loss term when studying domain wall evolution. Such a term is described by a power law of curvature,

$$F(v) = cv + d[k_0 - k]^r \,, \tag{12}$$

where c, d and r are free parameters which express the loop chopping parameter and the energy loss normalization and exponent, respectively. Although the original VOS, with the helicoidal ansatz and no explicit energy loss, had only one parameter, the extended version has 6. This is not a problem per se, as long as one can extract the correct velocity dependencies of each function $k(v)$ and $F(v)$. It is for this reason we will simulate string networks in Universes where with a power law scale factor, $a \propto m$ and with expansion rates m in the range $m \in [0.5, 0.95]$ (i.e. not only matter and radiation epoch are simulated).

We note that since string networks, under any Universe with a power law scale factor ($a \propto t^m \propto \eta^{m/(1-m)}$), are expected to evolve according to the so-called linear scaling regime where,

$$\xi \propto \eta \propto d_H \qquad v = const \,, \tag{13}$$

we can easily use the measured $\dot{\xi} \sim \frac{\xi}{\eta}$ and v to extract the proper velocity dependencies of $k(v)$ and $F(v)$, and also calibrate model parameters. To this end, we will use the same Markov Chain Monte Carlo pipeline of Ref. 25 to assess posterior distributions on all model parameters, and thus unveil possible correlations, and predict likelihood maxima and uncertainties. All model parameters are sampled from uniform distributions, with the range of these distributions adapted to each dataset as necessary. The logarithm of the likelihood is to be given by the χ^2 statistic. We will use 32 walkers and a minimum of 10000 steps.

4. Dynamic range and lattice size

String simulations have a problem of separation of scales, as it is often necessary to resolve scales all the way from the size of the horizon, down to scales comparable to

the radius of the strings. When it comes to better resolving small-scale structure, there are two possible ways to do so, either by decreasing lattice spacing or by increasing lattice size, which in turn increases dynamic range. The latter follows because the lattice size directly determines the final simulation time: half-a-light crossing time or when the horizon size is half the box size. If a larger horizon can be reached by the end of the simulation, then for the same lattice spacing we can better resolve scales of a fraction of the horizon.

We now explore the effects of increasing lattice size, while keeping spacing constant throughout ($\Delta x = 0.5$). To this end, we calibrate the VOS with 1024^3, 2048^3 and 4096^3 lattices, with expansion rates $m \in [0.45, 0.95]$, 10 simulation runs each. For now we will use the equation of state velocity estimator. The resulting calibrations are summarized in table 2 and the corresponding posterior plots are found in figure 3 of Ref. 13, remade in Fig. 4.

Table 2. Calibrated VOS model parameters for our three different lattice sizes, 1024^3, 2048^3 and 4096^3, all with the same lattice spacing $\Delta x = 0.5$, and two different choices of velocity estimators, $\langle v_\omega^2 \rangle$ and $\langle v_\phi^2 \rangle$ (in the top and bottom parts of the table, respectively),

Lattice size	Δx	Velocity estimator	d	r	β	k_0	q	c
1024^3	0.5	$\langle v_\omega^2 \rangle$	$0.32^{+0.04}_{-0.04}$	$1.51^{+0.48}_{-0.37}$	$1.82^{+0.34}_{-0.30}$	$1.27^{+0.08}_{-0.06}$	$2.41^{+0.13}_{-0.13}$	$0.15^{+0.05}_{-0.07}$
2048^3			$0.37^{+0.02}_{-0.02}$	$1.27^{+0.17}_{-0.15}$	$2.33^{+0.21}_{-0.20}$	$1.21^{+0.03}_{-0.03}$	$2.57^{+0.06}_{-0.06}$	$0.03^{+0.02}_{-0.03}$
4096^3			$0.39^{+0.02}_{-0.02}$	$1.36^{+0.15}_{-0.13}$	$2.32^{+0.20}_{-0.18}$	$1.18^{+0.03}_{-0.03}$	$2.59^{+0.05}_{-0.05}$	$0.00^{+0.01}_{-0.01}$
1024^3	0.5	$\langle v_\phi^2 \rangle$	$0.35^{+0.23}_{-0.10}$	$2.39^{+1.58}_{-0.94}$	$2.79^{+0.73}_{-0.56}$	$1.06^{+0.05}_{-0.05}$	$2.95^{+0.18}_{-0.19}$	$0.44^{+0.04}_{-0.05}$
2048^3			$0.33^{+0.05}_{-0.04}$	$1.86^{+0.39}_{-0.32}$	$2.65^{+0.28}_{-0.26}$	$1.05^{+0.03}_{-0.03}$	$2.84^{+0.08}_{-0.08}$	$0.31^{+0.02}_{-0.02}$
4096^3			$0.36^{+0.03}_{-0.03}$	$1.72^{+0.26}_{-0.23}$	$2.50^{+0.21}_{-0.20}$	$1.06^{+0.02}_{-0.02}$	$2.83^{+0.06}_{-0.06}$	$0.23^{+0.01}_{-0.01}$

Note: Displayed values correspond to 16th, 50th, 84th percentiles of the posterior distributions.

We see that most parameters remain relatively unchanged by increasing the lattice size, being subject only to a reduction of uncertainties (smaller contours in Fig. 4). The only exceptions are parameters c and (to a lesser extent) d. We remark as well that these differences are qualitatively larger when comparing 1024^3 with 2048^3, than if comparing 2048^3 to 4096^3, which might hint that there is a minimum lattice size/dynamic range for model calibration. Gradually, c is even reduced to zero at 4096^3 which seems to suggest that as lattice resolution increases, loop-chopping is eventually replaced by radiative losses, bu this is at odds with information from visually seeing the network evolve—as loops are formed at various instants and at various sizes—as can be inferred from Fig. 3.

In Ref.[13] we found, by comparing asymptotic quantities at all expansion rates, that the main culprit for the changes we discussed in the previous paragraph was the behavior of ξ/η. In fact, while the velocities remain mostly unchanged, ξ/η is

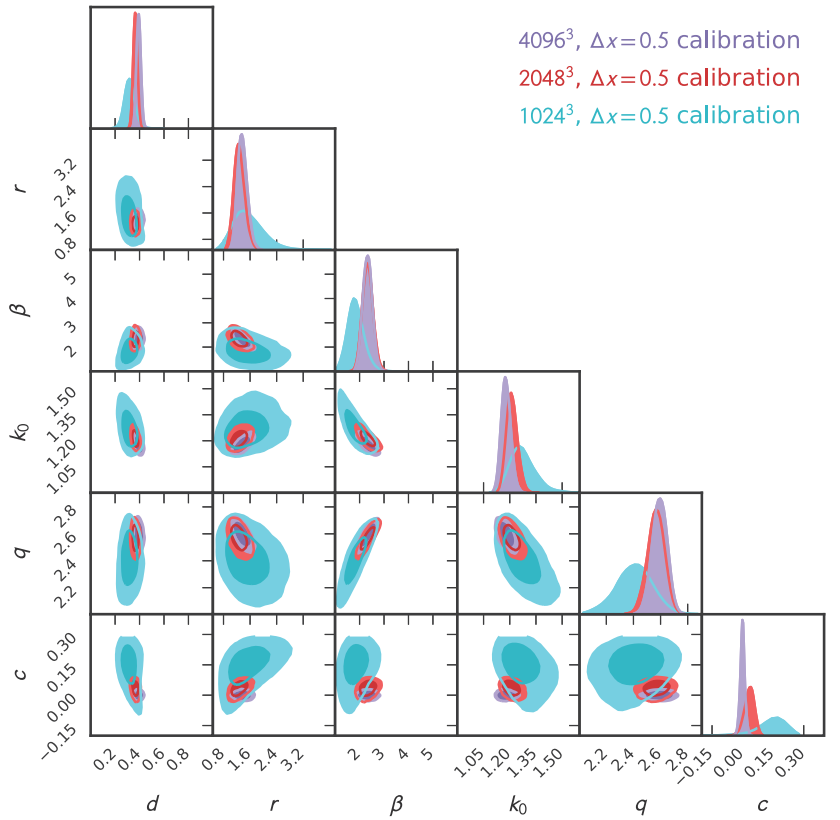


Fig. 4. Corner plots for the MCMC calibration of the VOS model, obtained with the velocity estimator $\langle v_\omega \rangle$, lattice spacing $\Delta x = 0.5$, box sizes 4096^3, 2048^3 and 1024^3. Remake of figure 3 of Ref. 13.

decreases (this is present in the literature too[9, 12]) with lattice size, as does it uncertainty. The reduction of the uncertainties is responsible for the narrower posterior distributions, and the lower mean value translates itself in change in a downwards shift of $F(v)$. This is reflected in the parameters that control the normalization of $F(v)$, i.e. in c and (to a lesser extent), via anti-correlation d. Again the differences are larger when going from 1024^3 to 2048^3 than from 2048^3 to 4096^3.

Going back to our previous discussion of c, it appears the calibration suggested the loop-chopping parameter would eventually go to zero, however the visualization strategy showed loops being produced at different sizes and different size. Perhaps, as a cross-check, we should also repeat this analysis using the scalar field velocity estimator. The resulting parameters are summarized in the bottom half of Table 2 and in figure 4 of Ref. 13. This now leads to a surprisingly different calibration, with very different model parameters, namely c is statistically non-zero. This prompts an investigation on the biases of each velocity estimator and on possible solutions.

5. Lattice spacing and velocity estimators

Here we investigate why both velocity estimators give vastly different calibrations and suggest one possible solution. From the point-of-view of the extended VOS, the high expansion rates can be some of the most important (statistically) for two reasons. First, in the low-velocity limit, both velocity functions are reduced to one parameter each,

$$\lim_{v\to 0} F(v) = cv \qquad \lim_{v\to 0} k(v) = k_0\,, \tag{14}$$

and this means that overall the model becomes reduced to only two parameters. Uncertainties on observed asymptotic quantities also decrease with decreasing velocity (increasing expansion rate). Statistically this will mean that the high expansion rate limit will be important in determining c and k_0, which will in turn affect all variables correlated with them, and that the lower uncertainties give a larger statistical weight to higher expansion rates. This leads us to suspect that the high expansion rate behavior of both velocity estimators might be the cause, as the difference between them increases with expansion rate.[10]

If we compare velocity estimators in the range $m \in [0.5, 0.95]$ (see figure 3 of Ref. 26) the difference is maximal at $m = 0.95$, and of about $\sim 10\%$. For even larger expansion rates, in the range $m \in [0.95, 0.997]$ (left panel of figure 5 of Ref. 13), the velocity estimators can disagree up to 60% in the worst case scenario for lattice spacing $\Delta x = 0.5$, In the work of Ref. 10, which explored the differences between the two estimators in Minkowski space (the opposite limit), it was shown that lattice spacing could reduce this disagreement. But even at high expansion rate this remains true: the disagreement is heavily reduced when we halve lattice spacing $\Delta x = 0.25$ as can be seen on the right panel of the aforementioned Figure 5. There is however one interesting difference between two opposite limits of high expansion rate and Minkowski: in the first the equation of state estimator approximates the scalar field, in the latter the opposite occurs. There is no a priori reason as to why both should behave the same way under lattice spacing reductions, therefore this is not in contradiction with previous results.

As expected, the differences and their reduction under lattice spacing, are also reflected in the velocity functions $k(v)$ and $F(v)$. This as can be seen in Figure 6 of Ref. 13, where two conclusions can be drawn. First, the equation of state velocity estimator yields unphysical $F(v)$ at lattice spacing $\Delta x = 0.5$, leading one to question its validity at the low velocity limit. Second, as the lattice is made finer, this brings both estimators in better agreement, namely by approximating $F(v)$ and $k(v)$ obtained via the equation of state estimator to the scalar field estimator. Note that this doesn't solve the problem of unphysical $F(v)$ completely, although for our intents and purposes it brings agreement for expansion rates $m \in [0.93, 0.95]$.

Now we compare the resulting model parameters for the two choices of lattice spacing. The results are summarized in Table 3 and in the corresponding posterior contours of figures Figures 7 and 8 of Ref. 13. We also showcase the contours of the

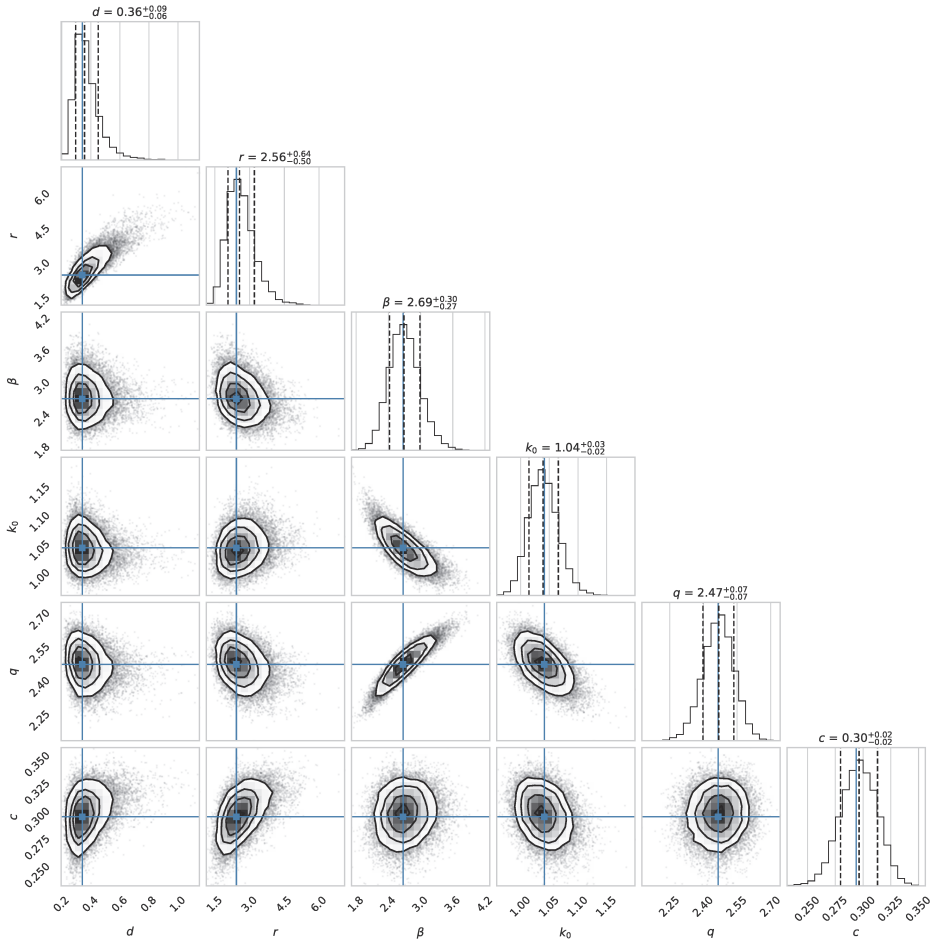


Fig. 5. Corner plots for the MCMC calibration of the VOS model, obtained with the velocity estimator $\langle v_\phi \rangle$, lattice spacing $\Delta x = 0.25$, box size 4096^3.

4096^3, $\Delta x = 0.25$, $\langle v_\phi^2 \rangle$ calibration in Fig. 5. Since the equation of state velocity estimator is highly sensitive to lattice spacing at large expansion rate, it is no surprise that the calibration drastically changes for all model parameters under a reduction of spacing, as can be seen both in the table and in the first aforementioned figure of Ref.[13] With the finer calibration, it is also obvious the value of c is no longer consistent with zero, that indeed loop-chopping is still a viable energy loss mechanism.

The calibrations performed with the scalar field velocity estimator mostly reduce uncertainties (with one exception) as the lattice spacing is reduced, as can be inferred from the lower half of Table 3 and from the contours of Figure 8 from

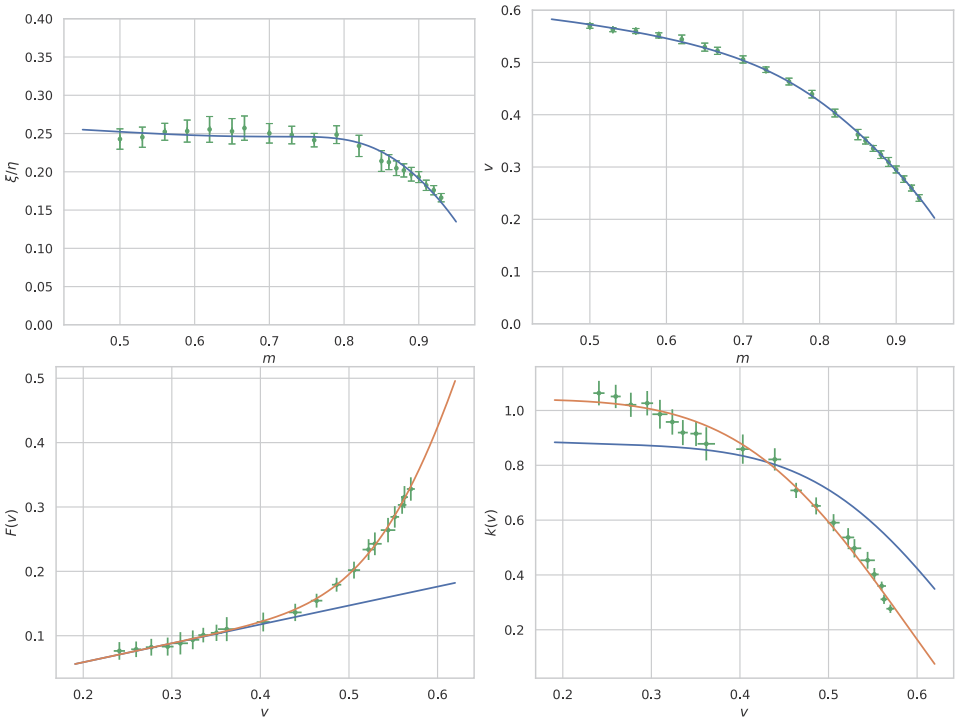


Fig. 6. A showcase of the VOS predictions for the mean rate of change of correlation length ξ/η (top left) and the mean velocity $\langle v \rangle$ (top right) with the solid blue lines corresponding to the calibration and the green points to the measured quantities with statistical $1-\sigma$ uncertainties. The bottom plots show the functions $F(v)$ and $k(v)$ with green points showing measured values and the blue lines showing the standard VOS prediction with the extended VOS prediction in orange.

Ref. 13. The only exception to this is the parameter q which is again expected, as it is related to the maximal defect velocity and the estimator changes mostly in the high velocity limit.

Thus lowering lattice spacing brings both calibrations, with different velocity estimators, into better agreement with the only different parameter being c. Overall this means one of the most important parameters from an observational point of view, c, is the most sensitive in the model, being somewhat affected by resolution effects and choice of velocity estimator.

To conclude we present the VOS predictions of the best-case scenario calibration, with lattice 4096^3, spacing, $\Delta x = 0.25$ and velocity estimator $\langle v_\phi^2 \rangle$ in Fig. 6. The VOS model predicts reasonably well, for all expansion rates, the values of ξ/η and v. We can see the measured value of $k(v)$ and $F(v)$ seem to be well described by the generalized forms, while the standard forms fail to predict the proper velocity dependency.

Table 3. Calibrated VOS model parameters for our two choices of lattice spacing Δx and corresponding lattice sizes, for the two different choices of velocity estimators, $\langle v_\omega^2 \rangle$ and $\langle v_\phi^2 \rangle$, further described in the main text.

Lattice size	Δx	Velocity estimator	d	r	β	k_0	q	c
2048^3	0.5	$\langle v_\omega^2 \rangle$	$0.37^{+0.02}_{-0.02}$	$1.27^{+0.17}_{-0.15}$	$2.33^{+0.21}_{-0.20}$	$1.21^{+0.03}_{-0.03}$	$2.57^{+0.06}_{-0.06}$	$0.03^{+0.02}_{-0.03}$
4096^3	0.25		$0.34^{+0.07}_{-0.05}$	$2.32^{+0.52}_{-0.40}$	$2.62^{+0.29}_{-0.26}$	$1.06^{+0.03}_{-0.02}$	$2.37^{+0.06}_{-0.07}$	$0.25^{+0.02}_{-0.02}$
2048^3	0.5	$\langle v_\phi^2 \rangle$	$0.33^{+0.05}_{-0.04}$	$1.86^{+0.39}_{-0.32}$	$2.65^{+0.28}_{-0.26}$	$1.05^{+0.03}_{-0.03}$	$2.84^{+0.08}_{-0.08}$	$0.31^{+0.02}_{-0.02}$
4096^3	0.25		$0.36^{+0.09}_{-0.06}$	$2.56^{+0.64}_{-0.50}$	$2.69^{+0.30}_{-0.27}$	$1.04^{+0.03}_{-0.02}$	$2.47^{+0.07}_{-0.07}$	$0.30^{+0.02}_{-0.02}$

Note: Displayed values correspond to 16th, 50th, 84th percentiles of the posterior distributions.

6. Observational Impact

We will now a highlight the impact of different calibrations on observational footprints of cosmic string networks. Note that a detailed study is beyond the scope of this manuscript, and this is merely illustrative of the need for an accurate calibration.[13]

We can use the CMBACT4 code[27,28] to compute the Cosmic Microwave Background anisotropies generated by a string network, assuming different VOS calibrations. This code assumes the Unconnected Segment Model,[27] where ensembles of straight, randomly oriented segments with separations and velocities given by the VOS model are used to compute power spectra.

We cancompare four different calibrations: standard VOS compatible with Nambu-Goto, standard VOS calibrated for Abelian-Higgs and two extended VOS calibrations present in the previous section, a worst (1024^3, $\Delta x = 0.5$, $\langle v_\omega^2 \rangle$) and best ($4096^3$, $\Delta x = 0.25$, $\langle v_\phi^2 \rangle$) case scenario. The resulting TT, TE, EE and BB can be found in figure 9 of Ref. 13, all normalized to a string tension $G\mu = 1.0 \times 10^{-7}$. It is clear that the Abelian-Higgs spectra are in better agreement between each other, than the Nambu-Goto one, which is expected.[29–34]

However, there are still some noticeable differences between Abelian-Higgs calibrations. The most discrepant overall is the 1024^3 calibration—the worst-case scenario. The 4096^3 calibration and the standard VOS one are in better agreement, although it should be noted there still exist some scale-dependent differences. To exemplify, at $l = 10$, the relative difference between the two TT spectra is around 16%, 30% and 11% for scalar, vector, tensor C_l, respectively.

The exact quantifying of how much of these differences are uniquely due to the calibrations or due to the USM requires a more in-depth, detailed study. Nevertheless, this illustrative comparison highlights the need for accurate VOS calibrations.

7. Conclusion

In this work we used our recently developed multiGPU Abelian-Higgs code and high-end computational resources of Piz Daint to obtain a robust calibration of the canonical semi-analytical model of string evolution. Using high-resolution simulations from lattice sizes ranging from 1024^3 to 4096^3 and various choices of expansion rate. We explored the impact of two different numerical choices on the obtained calibrations. This allowed us to not only reduce statistical uncertainties but also to correct for systematic error sources. Since we tested the sensitivity of model parameters to each of these choices we uncovered that the loop-chopping parameter is highly sensitive to both resolution effects and also to choice of velocity estimator. We showed that a minimum dynamic range of about 512.0 conformal time units and a minimum lattice spacing of $\Delta x = 0.25$ are necessary to minimize the impacts of these systematic error sources.

We then illustrated one of the main motivations to pursue an accurate, high-resolution calibration of the VOS: observational consequences. We specifically showed that different calibrations can give rise to scale-dependent differences on the computed CMB anisotropy power spectra. Although one cannot (at least without a more in-depth study) exclude that these differences are caused by the various approximations of the Unconnected Segment Model, it is still clear that precise and accurate calibrations are necessary to for both current and future constraints.

Acknowledgements

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COM-PETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017. J. R. C. is supported by an FCT fellowship (SFRH/BD/130445/2017). We gratefully acnknowledge the support of NVIDIA corporation with the donation of the Quadro P5000 GPU used in this research. We acknowledge PRACE for awarding us access to Piz Daint at CSCS, Switzerland, through Preparatory Access proposal 2010PA4610, Project Access proposal 2019204986 and Project Access proposal 2020225448. Technical support from Jean Favre at CSCS is gratefully acknowledged.

References

1. T. W. B. Kibble, Topology of Cosmic Domains and Strings, *J. Phys.* **A9**, 1387 (1976).
2. R. Jeannerot, J. Rocher and M. Sakellariadou, How generic is cosmic string formation in SUSY GUTs, *Phys. Rev.* **D68**, p. 103514 (2003).
3. S. Sarangi and S. H. H. Tye, Cosmic string production towards the end of brane inflation, *Phys. Lett.* **B536**, 185 (2002).
4. P. A. R. Ade *et al.*, Planck 2013 results. XXV. Searches for cosmic strings and other topological defects, *Astron. Astrophys.* **571**, p. A25 (2014).

5. B. P. Abbott *et al.*, Constraints on cosmic strings using data from the first Advanced LIGO observing run, *Phys. Rev.* **D97**, p. 102002 (2018).
6. F. Finelli *et al.*, Exploring cosmic origins with CORE: Inflation, *JCAP* **1804**, p. 016 (2018).
7. P. Binetruy, A. Bohe, C. Caprini and J.-F. Dufaux, Cosmological Backgrounds of Gravitational Waves and eLISA/NGO: Phase Transitions, Cosmic Strings and Other Sources, *JCAP* **1206**, p. 027 (2012).
8. K. G. Wilson, Confinement of quarks, *Phys. Rev. D* **10**, 2445 (Oct 1974).
9. N. Bevis, M. Hindmarsh, M. Kunz and J. Urrestilla, CMB power spectrum contribution from cosmic strings using field-evolution simulations of the Abelian Higgs model, *Phys. Rev.* **D75**, p. 065015 (2007).
10. M. Hindmarsh, J. Lizarraga, J. Urrestilla, D. Daverio and M. Kunz, Scaling from gauge and scalar radiation in Abelian Higgs string networks, *Phys. Rev.* **D96**, p. 023525 (2017).
11. W. H. Press, B. S. Ryden and D. N. Spergel, Dynamical Evolution of Domain Walls in an Expanding Universe, *Astrophys. J.* **347**, 590 (1989).
12. M. Hindmarsh, S. Stuckey and N. Bevis, Abelian Higgs Cosmic Strings: Small Scale Structure and Loops, *Phys. Rev. D* **79**, p. 123504 (2009).
13. J. R. C. C. C. Correia and C. J. A. P. Martins, High resolution calibration of the cosmic strings velocity dependent one-scale model, *Phys. Rev. D* **104**, p. 063511 (2021).
14. K. Kajantie, M. Karjalainen, M. Laine, J. Peisa and A. Rajantie, Thermodynamics of gauge invariant U(1) vortices from lattice Monte Carlo simulations, *Phys. Lett.* **B428**, 334 (1998).
15. J. J. Camata, V. Silva, P. Valduriez, M. Mattoso and A. L. Coutinho, In situ visualization and data analysis for turbidity currents simulation, *Computers & Geosciences* **110**, 23 (2018).
16. M. Rautenhaus, M. Böttinger, S. Siemen, R. Hoffman, R. M. Kirby, M. Mirzargar, N. Röber and R. Westermann, Visualization in meteorology—a survey of techniques and tools for data analysis tasks, *IEEE Transactions on Visualization and Computer Graphics* **24**, 3268 (2018).
17. D. Mu, J. Moran, H. Zhou, Y. Cui, R. Hawkins, M. Tatineni and S. Campbell, In-situ analysis and visualization of earthquake simulation (2019).
18. R. Sohrabi, S. Omlin and S. A. Miller, Geyser: 3d thermo-hydrodynamic reactive transport numerical simulator including porosity and permeability evolution using gpu clusters, *Computational Geosciences* **23**, 1317 (Dec 2019).
19. A. KAGEYAMA, N. SAKAMOTO, H. MIURA and N. OHNO, Interactive exploration of the in-situ visualization of a magnetohydrodynamic simulation, *Plasma and Fusion Research* **15**, 1401065 (2020).
20. J. R. C. C. C. Correia and C. J. A. P. Martins, Abelian–Higgs cosmic string evolution with multiple GPUs, *Astron. Comput.* **34**, p. 100438 (2021).
21. Top500, Top500 november 2020 list `https://top500.org/lists/top500/2020/11/`, (2018).
22. C. J. A. P. Martins and E. P. S. Shellard, Quantitative string evolution, *Phys. Rev.* **D54**, 2535 (1996).
23. C. Martins and E. Shellard, Extending the velocity dependent one scale string evolution model, *Phys. Rev. D* **65**, p. 043514 (2002).
24. C. J. A. P. Martins, I. Y. Rybak, A. Avgoustidis and E. P. S. Shellard, Extending the velocity-dependent one-scale model for domain walls, *Phys. Rev.* **D93**, p. 043534 (2016).

25. J. R. C. C. C. Correia and C. J. A. P. Martins, Quantifying the effect of cooled initial conditions on cosmic string network evolution, *Phys. Rev. D* **102**, p. 043503 (2020).
26. J. R. C. C. C. Correia and J. A. P. Martins, Extending and Calibrating the Velocity dependent One-Scale model for Cosmic Strings with One Thousand Field Theory Simulations, *Phys. Rev.* **D100**, p. 103517 (2019).
27. L. Pogosian and T. Vachaspati, Cosmic microwave background anisotropy from wiggly strings, *Phys. Rev. D* **60**, p. 083504 (1999).
28. T. Charnock, A. Avgoustidis, E. J. Copeland and A. Moss, CMB constraints on cosmic strings and superstrings, *Phys. Rev. D* **93**, p. 123503 (2016).
29. B. Allen and E. P. S. Shellard, Cosmic string evolution: A numerical simulation, *Phys. Rev. Lett.* **64**, 119 (1990).
30. K. D. Olum and V. Vanchurin, Cosmic string loops in the expanding universe, *Phys. Rev.* **D75**, p. 063521 (2007).
31. D. P. Bennett and F. R. Bouchet, High resolution simulations of cosmic string evolution. 1. network evolution, *Phys. Rev.* **D41**, p. 2408 (1990).
32. J. J. Blanco-Pillado, K. D. Olum and B. Shlaer, Large parallel cosmic string simulations: New results o n loop production, *Phys. Rev.* **D83**, p. 083514 (2011).
33. A. Lazanu, E. P. S. Shellard and M. Landriau, CMB power spectrum of Nambu-Goto cosmic strings, *Phys. Rev. D* **91**, p. 083519 (2015).
34. A. Lazanu and P. Shellard, Constraints on the Nambu-Goto cosmic string contribution to the CMB power spectrum in light of new temperature and polarisation data, *JCAP* **02**, p. 024 (2015).

Radiation from Global Cosmic Strings using adaptive mesh refinement

Amelia Drew* and E. P. S. Shellard

Centre for Theoretical Cosmology, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
**E-mail: a.drew@damtp.cam.ac.uk*

We present results from adaptive mesh refinement (AMR) simulations of global cosmic strings. Using the public code, GRChombo, we perform a quantitative investigation of the dynamics of single sinusoidally displaced string configurations. We study a wide range of string energy densities $\mu \propto \ln \lambda$, defined by the string width parameter λ over two orders of magnitude. We investigate the resulting massless (Goldstone boson or axion) and massive (Higgs) radiation signals, using quantitative diagnostic tools to determine the eigenmode decomposition. Given analytic radiation predictions for global Nambu-Goto strings, we compare the oscillating string decay with a backreaction model accounting for radiation energy losses, finding excellent agreement. We establish that backreaction decay is accurately characterised by the inverse square of the amplitude being proportional to the inverse tension μ for $3 \lesssim \lambda \lesssim 100$. The investigation of massive radiation at small to intermediate amplitudes finds evidence that it is suppressed exponentially relative to the preferred massless channel with a $\sqrt{\lambda}$ dependence in the exponent. We conclude that analytic radiation modelling in the thin-string (Nambu-Goto) limit provides the appropriate cosmological limit for global strings.

Keywords: Cosmic strings; adaptive mesh refinement.

1. Introduction

Cosmic strings are a fundamental prediction of many physically motivated field theories with symmetry-breaking phase transitions. They are a potential source of dark matter axions,[1] as well as gravitational waves.[2] However, their complex non-linear evolution means that significant uncertainties in their observational signatures and other properties persist in the literature.

The simplest example of cosmic string defects are 'global strings' formed from the breaking of a $U(1)$ symmetry with a single complex scalar field φ, given by

$$\mathcal{L} = (\partial_\mu \bar{\varphi})(\partial^\mu \varphi) - \frac{\lambda}{4}(\bar{\varphi}\varphi - \eta^2)^2 \,. \tag{1}$$

The symmetry breaking scale with this potential is set by the constant λ, which also determines the width of the string core $\delta \approx (\sqrt{\lambda}\,\eta)^{-1}$, where η is also a constant. Massive ϕ and massless (Goldstone) ϑ radiative modes are generated when these global strings oscillate, which can be analysed respectively using the following

diagnostics:

$$\Pi_\phi \equiv \frac{\phi_1\dot{\phi}_1 + \phi_2\dot{\phi}_2}{\phi}, \qquad \mathcal{D}_i\phi \equiv \frac{\phi_1\nabla_i\phi_1 + \phi_2\nabla_i\phi_2}{\phi},$$
$$\Pi_\vartheta \equiv \frac{\phi_1\dot{\phi}_2 - \phi_2\dot{\phi}_1}{\phi}, \qquad \mathcal{D}_i\vartheta \equiv \frac{\phi_1\nabla_i\phi_2 - \phi_2\nabla_i\phi_1}{\phi}. \tag{2}$$

Despite their apparent simplicity, the study of cosmic strings using large-scale numerical simulations poses a very significant computational challenge. This is due to the vast difference in scale between the typical string width δ and the string curvature scale (set by the Hubble radius $R \lesssim H^{-1}$, where H is the Hubble scale). This is characterised by the ratios $\ln R/\delta \sim 70$ and $\ln R/\delta \sim 100$ for QCD axion and GUT scale strings respectively. This challenge is typically addressed by using the Nambu-Goto action (e.g. Ref. 3), or field theory simulations with a fixed (comoving) core width (e.g. Ref. 4). This work addresses some of the uncertainties introduced by these methods by implementing field theory simulations of individual string configurations using adaptive mesh refinement.

2. Simulations

In this work, we perform adaptive mesh simulations of single, sinusoidally displaced strings with $0.3 \leq \lambda \leq 100$ and spatial periodicity (wavelength) $L = 32$. After an initial period of dissipative damping, strings are subsequently evolved from a desired initial amplitude A_0, with both linear and nonlinear evolution regimes investigated. The oscillation amplitude is measured by calculating the position of the string core i.e. where there is non-zero winding, and the massive and massless diagnostics Π_ϕ and $\mathcal{D}_i\vartheta$ are extracted on a cylinder far from the string core. Our simulations are performed using the adaptive mesh refinement code, GRChombo.[5]

In this study, we measure the string radiation damping rate for a range of $1 \leq \lambda \leq 100$, and investigate the massless radiation spectrum for $\lambda = 1$ and $\lambda = 10$ at linear and nonlinear oscillation amplitudes.[6] We also investigate the massive radiation in detail for strings with $\lambda \leq 2.5$.[7]

3. Results

It has been previously shown[8] that for a sinusoidally displaced string, the massless quadrupole radiation rate can be described using a simple analytic backreaction model, derived from the Kalb-Ramond string action. The effect of radiation energy losses on the string oscillation amplitude A is given by

$$\frac{1}{\varepsilon^2} - \frac{1}{\varepsilon_0^2} = \frac{\beta\, t}{\bar{\mu}L} \quad \Rightarrow \quad \varepsilon = \varepsilon_0 \left(1 + \frac{\beta\varepsilon_0^2 t}{\bar{\mu}L}\right)^{-1/2}. \tag{3}$$

Here, $\varepsilon = 2\pi A/T$ is the amplitude relative to the oscillation time period T of the string, where ε_0 is measured at time $t = 0$. The normalised energy per unit length

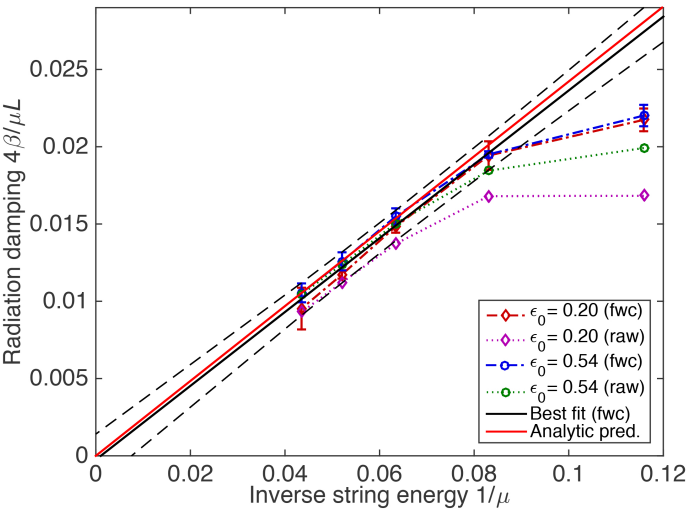


Fig. 1. Radiative damping rate measured using the inverse square model, plotted as a function of inverse string density μ^{-1}. We apply a finite size correction (fwc) to the data and exclude the result at $\lambda = 1$ from the best fit, due to string width effects which limit radiative damping at small λ. The analytic prediction in red uses an effective string radius cutoff $R = 3.5$. The dashed black lines show the error in the extrapolated slope. We observe that the damping rate tends to zero as $\mu \to \infty$ ($\lambda \to \infty$).

$\bar{\mu} = \mu/\eta^2 \approx 2\pi \ln(\sqrt{\lambda}\,\eta\, R)$, where the cutoff R is related to the radius of curvature of the string, $\beta = \pi^3/4 \approx 7.75$ is a constant and L is the spatial period of the string.

Figure 1 shows the main result of the work: the damping rate of a string measured using the inverse square model (3) depends linearly on the inverse of the string tension. We find that the model provides a good fit to the decay rate using the parameters $\beta = 7.6 \pm 1.6$ and $\ln R = 1.3 \pm 0.3$ ($R \approx 3.6$) (after applying a modest finite width correction of $\sim 8\%$ to account for internal oscillations within the string core). These values are consistent with the analytic damping coefficient β, and indicate an effective radial cutoff $R \approx L/8$, about half the string radius of curvature $L/4$. Hence, the inverse square radiation model provides an accurate picture of an oscillating and radiating global string, predicting both the correct power law dependence $\varepsilon^{-2} \propto \mu^{-1}$ and magnitude of the radiation damping.

We also investigate the spectrum of the massless eigenmodes emitted by oscillating strings for $\lambda = 1$ and $\lambda = 10$. Figure 2 gives an example of the massless radiation emitted from a $\lambda = 1$ string with intermediate initial amplitude $A_0 = 4$, showing an apparent dominant quadrupole signal. Figure 3 shows the dominant massless radiative modes for a $\lambda = 1$ string with initial amplitude $A_0 = 1$. We observe the $\{mn\} = \{2\,0\}$ quadrupole mode is dominant (where m and n denote the θ and z eigenvalues respectively), with a $\{1\,1\}$ self-field mode present from the start of the simulation and higher modes present at $\lesssim 100\times$ magnitude. We observe similar results for higher initial amplitudes in the nonlinear regime, $A_0 = 4$ and $A_0 = 8$,

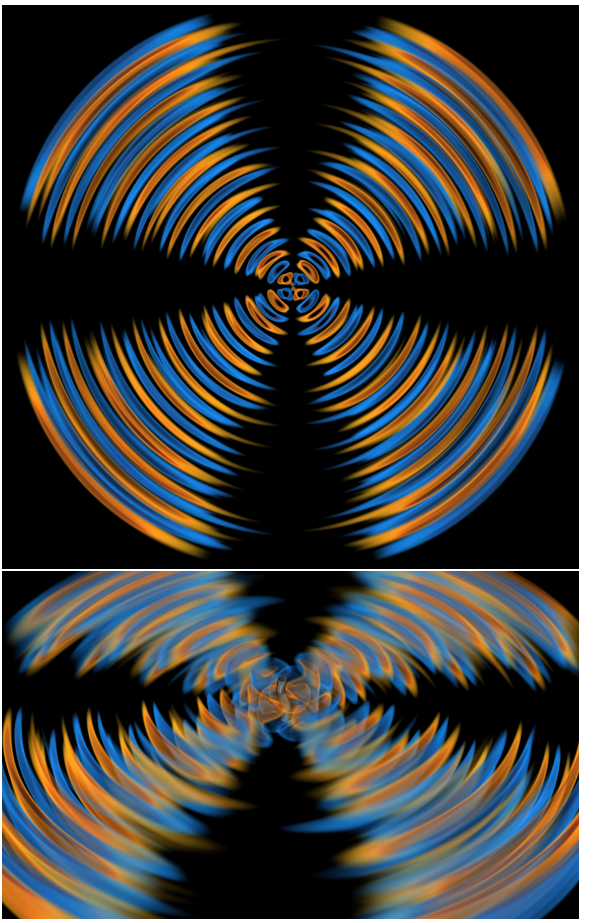


Fig. 2. Volume rendering in 3D space (x, y, z) of the massless radiation $\mathcal{D}\vartheta \cdot \hat{\mathbf{r}}$ from a $\lambda = 1$ string with initial amplitude $A_0 = 4$. The radiation is emitted from a string at the centre of the grid, with the quadrupole mode $\{mn\} = \{2\,0\}$ clearly dominant.

but with higher total magnitude and higher relative amplitude of the subdominant modes as A_0 increases. The amplitudes and spectra are similar for the more massive $\lambda = 10$ with some subtle differences discernible, including a higher initial radiation amplitude and a slightly slower amplitude decay rate. The former arises because, for $\lambda = 1$, finite size effects become important as the string core $\phi < 1$ extends into the radiation zone, causing some suppression of the quadrupole amplitude. The latter is a consequence of the $\lambda = 1$ string being lighter, so there is a larger relative effect from radiation backreaction.

Finally, we briefly present results from an investigation into the massive radiation from light strings with $\lambda \lesssim 3$. Overall, radiation via massive modes is subdominant; even for quasinonlinear intermediate amplitude regimes with $A_0 = 4$ ($\varepsilon \approx 0.7$),

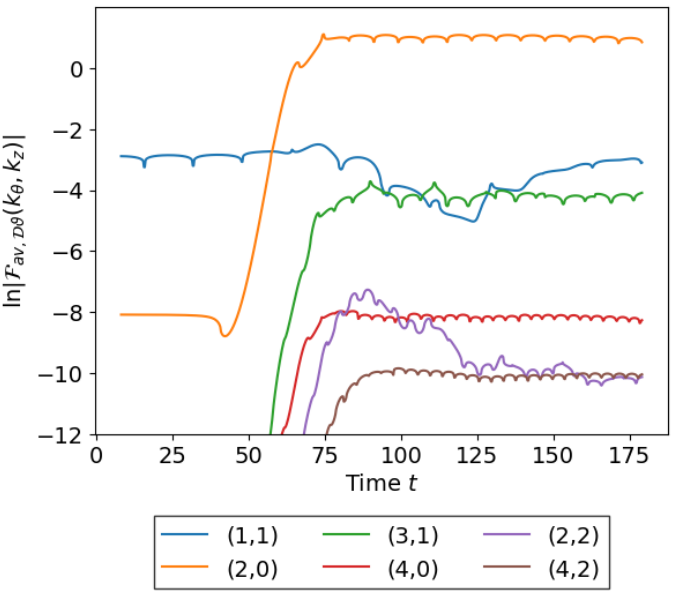


Fig. 3. Dominant 2D Fourier modes of the massless radiation $\mathcal{D}\vartheta \cdot \hat{\mathbf{r}}$ from a $\lambda = 1$ string with an initial amplitude $A_0 = 1$, measured on a cylinder at $R = 64$ and time averaged over approximate half-period $\Delta t = 33/2$.

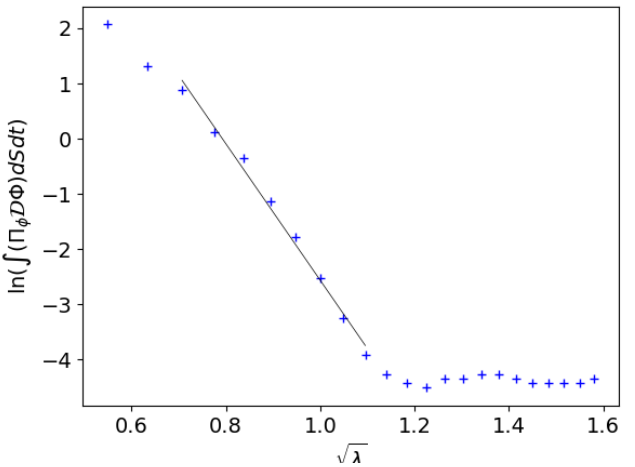


Fig. 4. Log-linear plot of the massive radiation diagnostic $\Pi_\phi \mathcal{D}\phi$ integrated over a diagnostic cylinder S at $R = 64$ from $90 \leq t \leq 228$ for $\varepsilon = 0.5$ and a range of $0.3 \leq \lambda \leq 2.5$. The black line indicates an exponential fit to the data for $0.5 \leq \lambda \leq 1.2$.

the magnitude of the massive signal is $\sim 10^3\times$ smaller than the massless emission. We propose a simple phenomenological model adapted from Ref. 9 to obtain an expression for the energy loss per period via massive radiation,

$$E(L, A_{\rm rel}, \lambda) = \sqrt{L A_{\rm rel}} \exp\left[-\gamma\sqrt{\lambda}\frac{L}{A_{\rm rel}}\right], \tag{4}$$

where $A_{\rm rel} = 4\,A/L$ and γ is a constant to be determined. Figure 4 shows the extracted massive signal $\Pi_\phi \mathcal{D}\phi$ for $A_0 = 4$ ($A_{\rm rel} = 0.5$), integrated over time from $90 \leq t \leq 228$ to give a quantitative estimate for the energy loss through this channel. We observe a clear exponential decay from $0.7 \lesssim \sqrt{\lambda} \lesssim 1.1$ ($0.5 \lesssim \lambda \lesssim 1.2$), consistent with the prediction in (4). We measure

$$\gamma = 0.193\ (\pm\ 0.007) \tag{5}$$

and the intercept $\ln\sqrt{L A_{\rm rel}} = 9.80\ (\pm\ 0.41)$. The first two points are excluded due to finite width effects that appear to suppress radiation, as are $\sqrt{\lambda} \gtrsim 1.1$ ($\lambda \gtrsim 1.3$) where numerical resonance effects become significant. The latter appears to be due to limited grid resolution (despite the use of AMR simulations), but we note that in the small and intermediate amplitude regimes studied the massive radiation always remains subdominant relative to massless modes for all λ.

4. Conclusions

We have presented results from adaptive mesh refinement simulations of global cosmic strings with a range of widths given by $0.3 \leq \lambda \leq 100$. By accurately fitting the oscillation amplitude to an inverse square decay model, we have demonstrated that dual radiation predictions from the Kalb-Ramond action are accurate as we asymptotically approach the small width regime. This lends weight to the Nambu-Goto approach as an accurate method to approximate strings in a cosmological context. We further conclude that the primary radiation channel for sinusoidal string configurations is the massless quadrupole eigenmode $\{mn\} = \{2\,0\}$, strongly dominating other radiation modes even at high amplitude. Finally, we conclude that massive radiation is subdominant in quasilinear regimes and decays exponentially with $\sqrt{\lambda}$ as expected.

Acknowledgements

We thank Josu Aurrekoetxea, Katy Clough, Thomas Helfer, Eugene Lim, Miren Radia and Ulrich Sperhake for useful conversations, as well as Kacper Kornet for invaluable computing support. This work was undertaken primarily on the COSMOS supercomputer at DAMTP, University of Cambridge, funded by BEIS National E-infrastructure capital grants ST/J005673/1 and STFC grants ST/H008586/1, ST/K00333X/1 and ST/P000673/1, as well as the Cambridge CSD3 part of the STFC DiRAC HPC Facility (with STFC grants ST/P002307/1, ST/R002452/1 and ST/R00689X/1). AD is supported by a Junior Research Fellowship from Homerton

College, Cambridge, and previously by an EPSRC iCASE Studentship in partnership with Intel (EP/N509620/1, Voucher 16000206).

References

1. R. L. Davis, Cosmic Axions from Cosmic Strings, *Phys. Lett. B* **180**, 225 (1986).
2. T. Vachaspati and A. Vilenkin, Gravitational Radiation from Cosmic Strings, *Phys. Rev. D* **31**, 3052 (1985).
3. B. Allen and E. P. S. Shellard, Cosmic-String Evolution: A Numerical Simulation, *Phys. Rev. Lett.* **64**, 119 (1990).
4. R. L. Davis and E. P. S. Shellard, Do Axions Need Inflation?, *Nucl. Phys. B* **324**, 167 (1989).
5. K. Clough et al., `GRChombo`: Numerical Relativity with Adaptive Mesh Refinement, *Class. Quant. Grav.* **32** (2016).
6. A. Drew and E. P. S. Shellard, Radiation from Global Topological Strings using Adaptive Mesh Refinement: Methodology and Massless Modes, *arXiv:1910.01718* (2019).
7. A. Drew and E. P. S. Shellard, Radiation from Global Topological Strings using Adaptive Mesh Refinement: Massive Modes, *in prep.* (2021).
8. R. A. Battye and E. P. S. Shellard, Global String Radiation, *Nucl. Phys. B* **423**, 260 (1994).
9. K. D. Olum and J. J. Blanco-Pillado, Radiation from Cosmic String Standing Waves, *Phys. Rev. Lett.* **84**, 4288 (2000).

Analysing the scaling density of axion strings

A. Lopez-Eiguren

Institute of Cosmology, Department of Physics and Astronomy, Tufts University, Medford, MA 02155, USA
E-mail: asier.lopez_eiguren@tufts.edu

In the QCD axion dark matter scenario with post-inflationary Peccei-Quinn symmetry breaking, the number density of axions, and hence the dark matter density, depends on the length of string per unit volume at cosmic time t, by convention written ζ/t^2. The expectation has been that the dimensionless parameter ζ tends to a constant ζ_0, a feature of a string network known as scaling. It has recently been claimed that in larger numerical simulations ζ shows a logarithmic increase with time. This case would result in a large enhancement of the string density at the QCD transition, and a substantial revision to the axion mass required for the axion to constitute all of the dark matter. With a set of new simulations of global strings we compare the standard scaling (constant-ζ) model to the logarithmic growth. We also study the approach to scaling, through measuring the root-mean-square velocity v as well as the scaled mean string separation x. We find good evidence for a fixed point in the phase-space analysis in the variables (x, v), providing a strong indication that standard scaling is taking place. We show that the approach to scaling can be well described by a two parameter velocity-one-scale (VOS) model, and show that the values of the parameters are insensitive to the initial state of the network. We conclude that the apparent corrections to ζ are artifacts of the initial conditions, rather than a property of the scaling network.

Keywords: Axionic Dark Matter, Topological Defects, Strings, Field Theory Simulations

1. Introduction

Axions[1, 2] were proposed to solve the strong-CP problem, a puzzling question of why QCD does not seem to break Charge Parity (CP) symmetry in strong interactions, but these models also inevitably produce dark matter by releasing their initial potential energy density. However, the analysis of the axions would not only provide us with the information about the dark matter nature or the solution to the strong-CP problem; axion dark matter searches and cosmic microwave background experiments would provide complementary evidences of inflation, a measurement of the axion mass would also immediately determine or constrain the energy scale of cosmic inflation and recent phenomenological analysis has also indicated a possible rich interplay between QCD axions and the electroweak hierarchy problem. Therefore, the analysis of the QCD axions will provide important information about unsolved problems in physics.

The axion particles can be produced in two different scenarios in the non-thermal case. One of these scenarios is described by the so-called misalignment mechanism.[3] In this scenario, the nonperturbative effects at the QCD phase transition generate a mass and the axion field relaxes to its minimum, oscillating around it. The axions

created in this process have zero momentum. The other scenario is based on the formation of topological strings after the spontaneous symmetry breaking of the Peccei-Quinn symmetry.[4] These strings show scaling behaviour and therefore emit axion particles in order to reduce their energy. After the QCD phase transition, the network of strings vanishes due to the presence of domain walls. The misalignment mechanism produces a dominant axion contribution when the axions are created before the end of inflation, otherwise axionic string radiation would produce the bulk of the axion population.

As we have mentioned, the axion[1,2] is a particle, the angular excitation of the "Peccei-Quinn" (PQ) field,[5,6] proposed to solve the strong-CP problem. The axion model has one undetermined parameter, the vacuum value of the field, f_a. The mass of the axion, m_a, can be written by using this parameter: $m_a = (\chi(T)/f_a)^{1/2}$. In this expression $\chi(T)$ is the temperature-dependent topological susceptibility of QCD. The value of f_a also determines the amount of axion dark matter that would have been produced in the early Universe, which means that it should be possible to predict the axion mass using the dark matter densities measured by the recent experiments such as Planck.[7] But to do so, we need to understand the efficiency of axion production in cosmology.

Thus, the vacuum value of the axion field, f_a, is an important energy scale to determine the axion mass. However, there is also another important energy scale to analyse the cosmological behaviour of the axion field. This energy scale is the QCD symmetry breaking scale. In the cosmological evolution of the axion field the first important event is the PQ phase transition, which is broken at a high temperature $T \sim f_a > 109$ GeV. In this phase transition a network of axionic strings are formed. These strings decay gradually emitting axion particles. At lower temperatures, temperatures similar to the QCD symmetry breaking, domain walls are formed between strings and the complex hybrid network annihilates into axions in about one Hubble time.

As we have shown the axion production analysis is complicated because topological defect, axionic strings, appear in the axionic field[8] and because the only reliable way to analyse their dynamics is through field simulations. However, numerical simulations cannot cover the whole string evolution and they must be extrapolated. The prediction of the axion density depends on having an accurate description of axion string evolution.

A physical basis is very important for the extrapolation of the simulation results. In the case of strings, the one-scale model[9] and its velocity-dependent improvement[10–12] can give the appropriate basis. These models predict that the system will acquire a scaling regime, where the mean string separation grows in proportion to cosmic time t, and the RMS velocity of the strings is constant. By scaling we mean that at distances much larger than the string width, network length-scales such as the mean string separation are proportional to the cosmic time t.

In the following, we will analyse if the axionic strings follow the standard scaling picture,[13–15] where the dynamical evolution of the network is independent of the

string width and tension. Or on the contrary, the system follows the model recently put forward,[16–21] which suggest a logarithmic correction to scaling based on the logarithmic growth in the effective string tension of a global string with their mean separation.

2. Model and Network Parameters

2.1. *Field dynamics*

The simplest axion model can be constructed using a singlet scalar field with a U(1) symmetry, Φ, with action

$$S = \int d^4x\sqrt{-g}\Big(\frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{4}\lambda(\Phi^2-\eta^2)^2\Big), \tag{1}$$

where we have written the field as a two-component vector, and the U(1) symmetry is realised as a rotation on the vector.

In a FLRW metric, and when the field is coupled to a thermal bath of weakly-coupled particles, the equations of motion take the form

$$\Phi'' + 2\frac{a'}{a}\Phi' - \nabla^2\Phi = -a^2\lambda(\Phi^2-\eta^2(T))\Phi, \tag{2}$$

where a is the scale factor, a prime denotes differentiation with respect to conformal time τ, and in the radiation era $a \propto \tau$. The free energy of the system is minimised at the field magnitude $\eta(T)$, where $\eta^2(T) = d(T_{\rm c}^2 - T^2)$, $T_{\rm c} \simeq \eta$ is the critical temperature of the PQ phase transition, and d is a constant computable in perturbation theory. For $T \gg T_{\rm c}$, it is energetically favourable for the field to fluctuate around $\Phi = 0$. Well below the critical temperature it is energetically favourable for the magnitude of the field to take the value η, with a massless pseudoscalar fluctuation mode (the axion) and a scalar mode of mass $m_{\rm sca} = \sqrt{2\lambda}\eta$. During the phase transition, the direction in field space is chosen at random in uncorrelated regions of the universe, with the result that the field is forced to stay zero along lines.[22] These lines form the cores of the axion strings.[8] The size of the core is approximately $w_0 = m_{\rm sca}^{-1}$.

2.2. *Network parameters from field averages*

The evolution of the network of strings produced after the symmetry breaking can be analysed using the string length ℓ and the RMS velocity v of the strings.

In order to compute the string length we can use a couple of estimators. On the one hand we can use the number of plaquettes pierced by strings. This estimator is known as the winding length $\ell_{\rm w}$. The length of the strings is obtained by multiplying the number of plaquettes pierced by strings by the physical lattice spacing $a\delta x$ and then correcting the value by factor of 2/3 to compensate for the Manhattan effect.[23] The plaquettes pierced by strings are identified by working out the "winding" phase of the field around each plaquette of the lattice.[24] This is an estimate of the length

of string measured in the "universe frame", that is, observers comoving with the expansion of the universe.

One can obtain other measures of length taking into account that the energy of a string configuration is proportional to its length. In this case the estimators are constructed using local functions of the fields. Neglecting the expansion of the universe and considering the energy components to be weighted, we can write the rest-frame length $\ell_{\rm r}$ and the mean square velocity as

$$\ell_{\rm r} = \frac{E + f_V L}{\mu(1 - f_V)}, \tag{3}$$

$$v_L^2 = \frac{E + L}{E + f_V L}, \tag{4}$$

where the subscript L denotes the use of the Lagrangian to obtain the estimate. The energy, E, and the lagrangian, L, are defined as:

$$E = E_\pi + E_D + E_V, \tag{5}$$

$$L = E_\pi - E_D - E_V, \tag{6}$$

where E_π is the weighted kinetic energy, E_D is the weighted gradient energy and E_V is the weighted potential energy. Likewise, μ is the weighted mass per unit length of a static string and f_V is the fraction contributed by the potential energy density. In this case their values are $\mu = 0.892\eta^2$ and $f_V = 0.368$. For more information about the weighting procedure and computation of the values of μ and f_V we refer the reader to.[15]

An alternative way of estimating the string velocity comes from the pressure,

$$p\mathcal{V} = E_\pi - \frac{1}{3}E_D - E_V, \tag{7}$$

giving another mean square velocity estimator

$$v_\omega^2 = \frac{1 + 3\omega + 2f_V}{2 + f_V(1 + 3\omega)}, \tag{8}$$

where $\omega = p\mathcal{V}/E$ is the equation of state parameter of the strings. A third estimate for the string velocity can be constructed from the ratio of the kinetic to gradient energies,[25]

$$R_{\rm s} = \frac{E_\pi}{E_D}, \tag{9}$$

which can be rearranged to give

$$v_{\rm s}^2 = \frac{2R_{\rm s}}{1 + R_{\rm s}}. \tag{10}$$

Following the analysis performed in,[15] the estimators that we are going to use in this paper will be the rest frame length, $\ell_{\rm r}$, and the velocity estimator derived from the kinetic and gradient energies, v_s.

In a cosmological simulations one can express the string length in terms of Hubble lengths per Hubble volume, or

$$\zeta_{\rm r} = \frac{\ell_{\rm r} t^2}{\mathcal{V}}. \tag{11}$$

When investigating scaling in string networks, it is more transparent to parametrise the string density by the mean string separation, which is obtained from measures of the string length via

$$\xi_{\rm r} = \sqrt{\frac{\mathcal{V}}{\ell_{\rm r}}}. \tag{12}$$

The above estimators were derived for a Minkowski space-time. In an expanding background, one can view the space-time coordinates as representing comoving position and conformal time, from which physical lengths follow by multiplication by the scale factor a.

3. Simulations

In order to perform simulations of axionic string we discretise the equations of motion (2) and evolve them on a cubic lattice with periodic boundary conditions. The system is evolved in conformal time τ. The simulations analysed in this paper are the same set of simulations used in,[14,15] where we analyzed lattices with 1024, 2048 and 4096 sites per dimension with spatial resolution of $\delta x \eta = 0.5$ and conformal time steps of $\delta\tau\eta = 0.1$. We will give a short summary of the procedure and we refer the reader interested in more details to.[14,15]

The field configuration is initiated at conformal time $\tau_{\rm start}$ by setting the canonical momentum to zero and the scalar field to be a Gaussian random field with power spectrum $P_\Phi(k) = A\left[1 + (k l_\phi)^2\right]^{-1}$, were A is chosen so that $\langle\Phi^2\rangle = \eta^2$ and l_ϕ is the field correlation length in comoving coordinates. We use different values of l_ϕ in order to cover a range of string separations in the initial conditions. The values used are $l_\phi = 5, 10, 20, 40$, where $l_\phi = 40$ case have been used only for $s = 1$ case. In order to allow the strings to form, and to remove the energy excess in the field fluctuations around the string configurations, we evolve this configuration with a diffusion equation with unit diffusion constant until conformal time $\tau_{\rm diff}$. We then apply the second-order time evolution equation (2).

Similarly to our previous paper, we extract data from simulations with both fixed comoving string width and fixed physical string width. We promote the scalar self-coupling constant to be a time dependent parameter $\lambda = \lambda_0/a^{2(1-s)}$ following the PRS method.[26] This makes the comoving string width decrease with conformal time as:

$$w(\tau) = \frac{w_0}{a^s(\tau)}\,. \tag{13}$$

The physical equation of motion, where the physical string width remains constant at $w_0 = 1/\sqrt{2\lambda_0}\eta$, and the comoving width decreases with time, corresponds to $s = 1$. With $s = 0$ the comoving width is constant at w_0 and the physical string width increases in time.

For the $s = 1$ case, it is difficult to avoid the string width being larger than the Hubble length at early times, which also means that the relaxation of the field to its equilibrium value is longer than a Hubble time. In order to speed up the string formation, we arrange the time-dependence of the coupling so that strings are formed and diffused with a constant comoving width, equal to their final comoving width. At the end of the diffusion period, the string width is much smaller than its physical value w_0. The string width is then allowed to grow by setting $s = -1$ until $\tau_{\rm cg}$, which is when the string core has expanded to its correct physical width w_0. After conformal time $\tau_{\rm cg}$, the physical evolution with $s = 1$ starts. We call this procedure core growth. Simulations end at conformal time $\tau_{\rm end}$.

Table 1 contains all simulation parameter choices that have been considered in the procedures described above. Four simulations with different random number seeds were carried out at each parameter choice. The data are analysed in cosmic time $t = (\tau/\tau_{\rm end})^2\tau_{\rm end}/2$.

Table 1. Run parameters used in simulations. See text for explanation.

Model	$s = 1$	$s = 0$
$l_\phi\eta$	(5,10,20,40)	(5,10,20)
$\tau_{\rm start}\eta$	50	50
$\tau_{\rm diff}\eta$	70	70
$s_{\rm cg}$	-1	–
$\tau_{\rm cg}\eta$	271.11	–
$\tau_{\rm end}\eta$	1050	1050

4. Results

In Fig. 1 we show the evolution of the mean string separation (12) and the RMS velocities for all the simulations performed with size 4096. In these figures the solid line represents the mean obtained by averaging over realisations and the shaded regions the 1σ standard deviations. In both estimators the uncertainties are worked out by propagating the fluctuations in the weighted energies E_π, E_D and E_V.The color code is as follows: black for $l_\phi\eta = 5$, red for $l_\phi\eta = 10$, blue for $l_\phi\eta = 20$, and green for $l_\phi\eta = 40$. Note that we only have $l_\phi\eta = 40$ simulations in the $s = 1$ case. The end of the core growth period is shown as a vertical green dashed line.

Figure 1 show that, independently of the initial field correlation length, all simulations are compatible, *i.e.* all of them give separation and velocity data which are within 1σ of each other. Moreover, the behaviour of both estimators (ξ and v) qualitatively agrees with the standard scaling, showing a tendency towards linear growth in ξ and a constant RMS velocity.

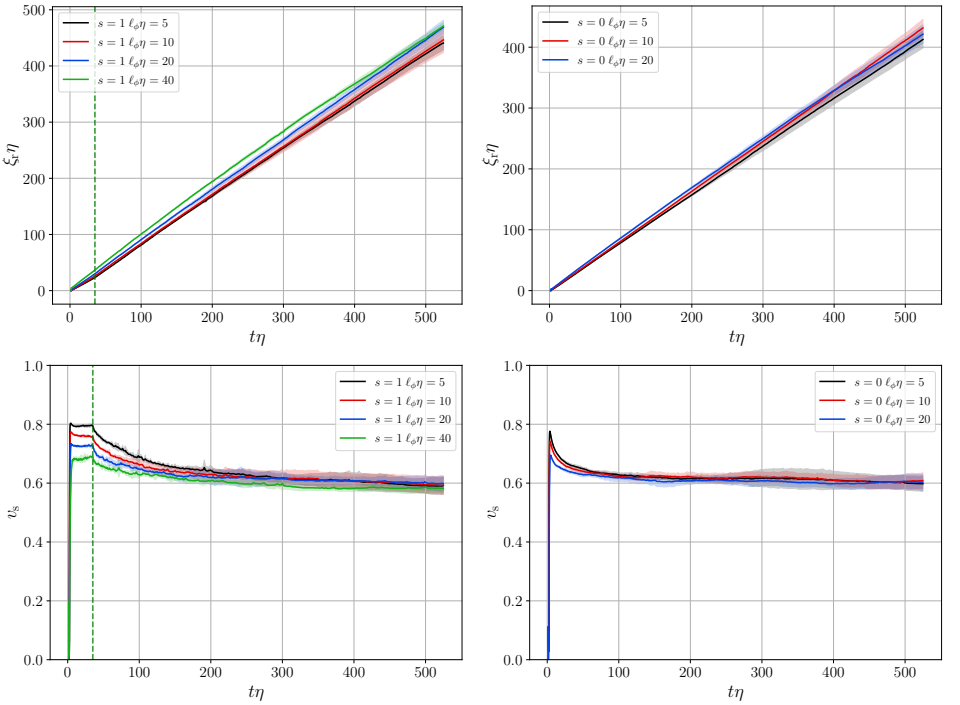


Fig. 1. Mean string separation ξ_r and velocity v_s evolution for $s = 1$ (right) and $s = 0$ (left).

Now we are going to analyse more in detail the scaling behavior of the system. First we will analyse the Length estimator and compare the different behaviors proposed in the literature for the scaling regime. Afterwards using the both length and velocity estimators we will analyse the approach to the scaling phase.

4.1. *Length Estimator*

As we have previously mentioned after a relatively short period of relaxation, the mean string separation ξ tends to a line that can be well fitted with

$$\xi = 2\beta(t - t_0). \tag{14}$$

where β and t_0 are parameters to be fitted. β is predicted to be a universal parameter, while t_0 is a phenomenological fit parameter to reduce the effect on estimates of β of the initial conditions and evolution of the RMS velocity. This model is known as the standard scaling and in this case the scaling value of the string length in terms of Hubble lengths per Hubble volume, ζ, is $\zeta_0 = 1/4\beta^2$.

We measure the parameters β and t_0 for each one of the individual runs we performed with a linear fit over the ranges in conformal time defined by the boundary

times $\tau_{\rm b} = [6, 10]\tau_s$, and $\tau_s\eta = [25, 50, 100]$ for $N = [1k, 2k, 4k]$. That is, in the $1k$ case the fitting range is $\tau_{\rm b} = [150, 250]$. The times are chosen in the last half of the conformal time range to minimise the biases coming from the initial conditions.

The values obtained are shown in Fig 2. The color code in this figure is the same as in Fig. 1 and the style of each point represent the size of the simulation box, circles represent $1k$ simulations, triangles pointing downwards represent $2k$ simulations and triangles pointing upwards represent $4k$ simulations. The parameter t_f is taken to be as the end of the fitting period.

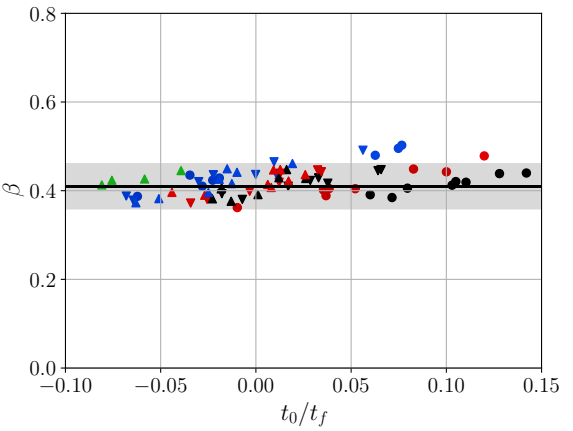


Fig. 2. Parameters β and t_0 for each one of the individual runs.

In Fig. 2 we have also added the mean value for β and its 1σ standard deviation obtained in.[15] As we can see all the points agree with the average value within 2σ and there is no correlation between the value of t_0 and β, showing that the initial conditions are not affecting the value obtained for β.

We now turn to the alternative model recently put forward: logarithmic[16–19,21] correction to scaling

$$\zeta(t) = \zeta_0^* + \alpha^* \log(\eta t) \tag{15}$$

where ζ_0^* and α^* are the fitting parameters. We performed fits over the ranges used previously and we show the values in Fig. 3, where we use the value of t_0 computed previously.

In Fig. 3 we have also added the mean value for β and its 1σ standard deviation obtained in,[15] but in this case translated to ζ_0. One can see that there is a clear correlation between the values of ζ_0^* and α^* with respect to t_0, showing that the initial conditions are affecting the values. Moreover, in the case that $t_0 = 0$ the value of α^* is zero and ζ_0^* agrees with the value obtained in.[15]

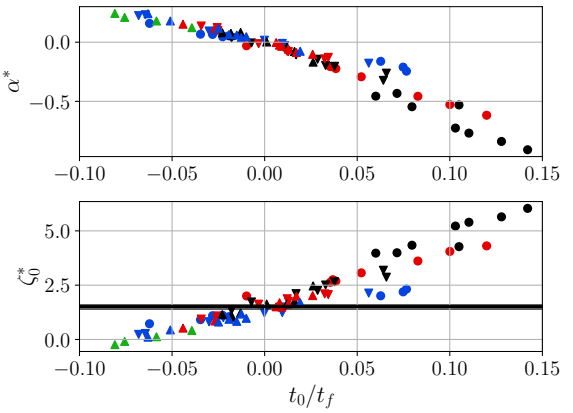


Fig. 3. Parameters ζ_0^* and α^* for each one of the individual runs.

4.2. *Length and Velocity Estimators*

As we have seen in the previous section, there is a small departure from standard scaling in the earlier phases of the simulations. It is necessary to understand this departure to improve the estimates of the asymptotic behavior of v and ξ, or more precisely the asymptotic values of the *scaled mean string separation*,

$$x = \xi/t. \tag{16}$$

In the previous Section, we have studied scaling using ξ only. Now we will also use the RMS velocity data, which gives extra information about the approach to scaling, and we will avoid the use of t_0. We will see that in doing so we will improve the accuracy of the estimate of β, while remaining consistent with our previous estimate. The analysis of this Section will be performed in simulations with size 4096.

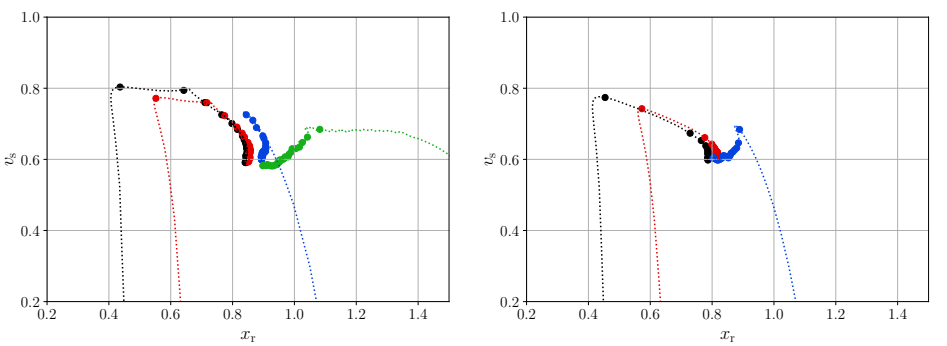


Fig. 4. The evolution of the network in the phase space (x, v).

In Fig. 4 we show the evolution of the network in the phase space (x, v). The phase space representation shows clearly the different regimes in which the network evolves. A rapid increase in the RMS velocity, while the inter-string distance remains nearly constant, happens after the end of the diffusion period, when strings are accelerated under their curvature. In the case of $s = 1$, after the diffusion period we impose a core growth period as part of the preparation of the initial conditions, where the velocities remain approximately constant but the scaled mean string separation changes. This change is different for different initial conditions. For correlation lengths $l_\phi \eta = 5,\ 10$, the scaled mean string separation grows, whereas for correlation lengths $l_\phi \eta = 20,\ 40$ it decreases. Finally, when the physical equations of motion (2) are being solved, the system starts to spiral towards an apparent common fixed point for all simulations.

The evolution of the scaled mean string separation and the RMS velocity can be modeled by the velocity-dependent one-scale (VOS) model.[9–12] Assuming that the statistical distribution of string configurations and velocities has a universal form the model parametrised it by the string separation ξ and the RMS velocity v.

The model uses the following equations to describe the dynamical system,

$$\frac{d\xi}{dt} = H\xi(1 + v^2) + \frac{1}{2}cv, \tag{17}$$

$$\frac{dv}{dt} = (1 - v^2)\left(\frac{k}{\xi} - 2Hv\right), \tag{18}$$

where H is the Hubble parameter. There are two phenomenological parameters, k and c. The parameter c describes the efficiency of the energy loss mechanism, while the parameter k describes the correlation between the string curvature and the velocity.

Using the dimensionless mean string separation variable x (16), and $H = 1/2t$ since we are considerning a radiation-dominated universe, we have,

$$t\dot{x} = \frac{1}{2}x\left(v^2 - 1\right) + \frac{c}{2}v \tag{19}$$

$$t\dot{v} = \left(1 - v^2\right)\left(\frac{k}{x} - v\right) \tag{20}$$

This dynamical system has a fixed point in the relevant region $0 \leq x$, $0 \leq v < 1$,

$$x_* = \sqrt{k(c + k)} \qquad v_* = \sqrt{\frac{k}{(c + k)}}. \tag{21}$$

The fixed point is a stable spiral. We plot flows in the phase diagram predicted by the VOS model in Fig. 5 for the global best fit (x_*, v_*) (computed from a χ^2 analysis; for more details see[15]), along with the mean values of selected (x, v) from the simulations. The stable spiral form is clearly visible in the streamlines.

To show in a more quantitative way how well the VOS model describe the evolution of the system, we show in Fig. 6 the relative difference between the simulation

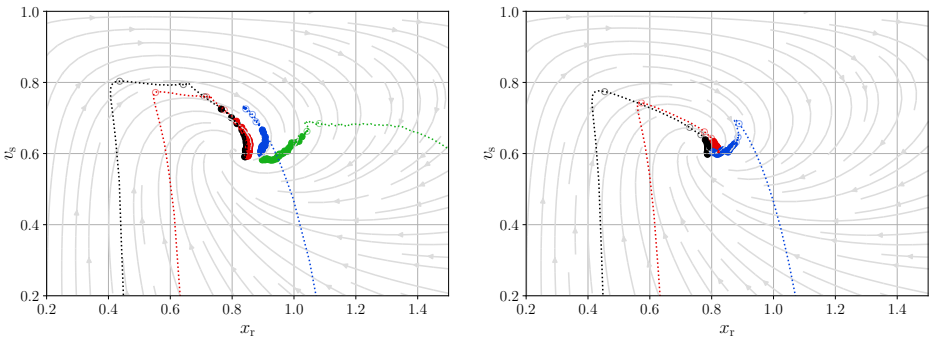


Fig. 5. The evolution of the network in the phase space (x, v) with stream lines.

time series data and the VOS best-fit model for each (s, l_ϕ), where the VOS model is initialized at $t_{\rm fit}$, which is the same as in the previous Section. Shaded regions correspond to the uncertainties propagated from simulation estimators. It can be observed that the mean relative difference always lies below 5% level of deviation, with zero deviation always within the errors.

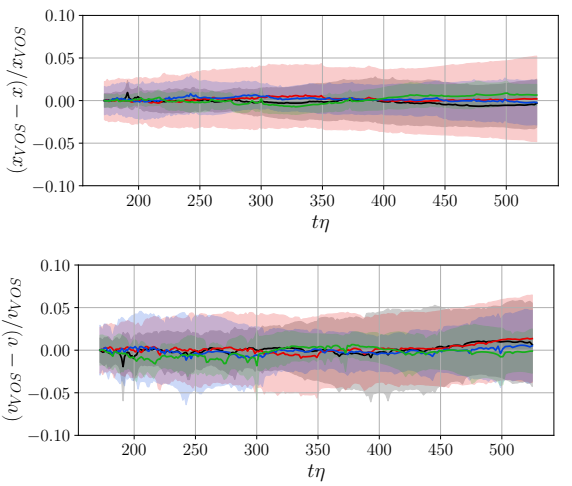


Fig. 6. Relative errors.

To understand better how the network evolves it is fundamental to analyse it in terms of the length density parameter ζ (11). In Fig. 7 we show ζ plottted against the logarithm of time, in order to emphasise the earlier times when the network is further away from scaling. The case that we are showing is this plot is $s = 1$, where we have also added the best-fit VOS models for each correlation length, and their extrapolation to larger values of time. The asymptotic values obtained in[15]

($\zeta_{r,*} = 1.50 \pm 0.11$) is also shown in the figure. The solid line represents the central value and the shaded region represents the 1σ error. In the figure we can see that the simulations approach the asymptotic value of ζ from below and that at the end of the simulations they are still slowly increasing, but within 20% of its asymptotic value. This increase is more visible in the simulations that start very underdense and therefore have much more to evolve until they reach scaling. The extrapolated values of the VOS model show that the system will reach the asymptotic value and become constant.

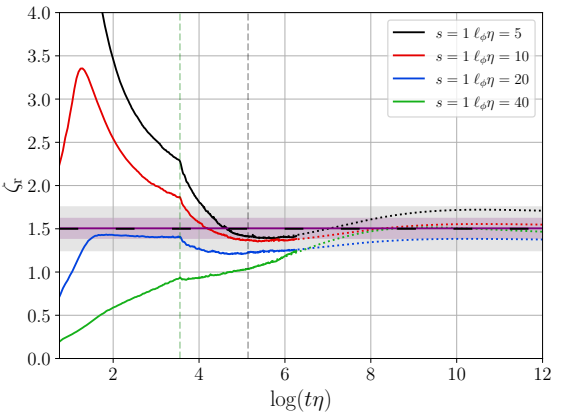


Fig. 7. ζ evolution.

5. Conclusions

In this paper we have analysed the evolution of the mean string separation ξ and the RMS velocity for an axion string network in the radiation era. The axion model has been constructed using a singlet scalar field with a U(1) symmetry, where the field has been considered as a two-component vector and the U(1) symmetry has been realised as a rotation on the vector.

Firstly, we have fitted the mean string separation (12) from the simulations we have performed with two two-parameter models: the standard scaling model with the usual time offset t_0 to account for the initial string evolution and a constant parameter β, and a model with a logarithmically increasing ζ (11). Using linear fits to obtain the values for β we have seen that the value does not depend on the initial conditions, or on the value of t_0. The coefficients of the logarithmically increasing model can be understood in terms of the dependence of ζ on the initial string evolution parameter t_0, and describe a disguised approach to scaling for non-zero values of t_0. We find that the value of the parameter describing the logarithmic increase is consistent with zero when $t_0/t_f \to 0$, where t_f is the final fitting time. In this case the constant term of the model is consistent with the standard scaling values.

Secondly, we have fitted the results of our set of numerical simulations to a two-parameter VOS model (18). We find that the evolution of the network is well described, qualitatively and quantitatively, by the VOS model. Qualitatively, we see that after the effects of the initial conditions are gone the systems turns into a curved approach to the fixed point in the (x, v) plane. A deeper analysis of the spiral behavior of the system showed that the apparent logarithmic growth in ζ is a transient epoch of the system, and that the length of this transient epoch depends on how far the initial conditions are from the fix point. After the transient epoch the system tends to a constant ζ evolution.

Acknowledgments

This work has been performed in collaboration with Mark Hindmarsh, Joanes Lizarraga and Jon Urrestilla. ALE (ORCID ID 0000-0002-1696-3579) is supported by the National Science Foundation grant PHY-1820872. This work has been possible thanks to the computational resources on the STFC DiRAC HPC facility obtained under the dp116 project. Our simulations also made use of facilities at the i2Basque academic network and CSC Finland.

References

1. S. Weinberg, A New Light Boson?, *Phys. Rev. Lett.* **40**, 223 (1978).
2. F. Wilczek, Problem of Strong P and T Invariance in the Presence of Instantons, *Phys. Rev. Lett.* **40**, 279 (1978).
3. J. M. Cline, TASI Lectures on Early Universe Cosmology: Inflation, Baryogenesis and Dark Matter, *PoS* **TASI2018**, p. 001 (2019).
4. R. L. Davis, Cosmic Axions from Cosmic Strings, *Phys. Lett. B* **180**, 225 (1986).
5. R. D. Peccei and H. R. Quinn, CP Conservation in the Presence of Instantons, *Phys. Rev. Lett.* **38**, 1440 (1977).
6. R. D. Peccei and H. R. Quinn, Constraints Imposed by CP Conservation in the Presence of Instantons, *Phys. Rev. D* **16**, 1791 (1977).
7. P. A. R. Ade *et al.*, Planck 2015 results. XIII. Cosmological parameters, *Astron. Astrophys.* **594**, p. A13 (2016).
8. R. L. Davis, Cosmic Axions from Cosmic Strings, *Phys. Lett. B* **180**, 225 (1986).
9. T. W. B. Kibble, Evolution of a system of cosmic strings, *Nucl. Phys. B* **252**, p. 227 (1985), [Erratum: Nucl.Phys.B 261, 750 (1985)].
10. C. J. A. P. Martins and E. P. S. Shellard, Quantitative string evolution, *Phys. Rev. D* **54**, 2535 (1996).
11. C. J. A. P. Martins and E. P. S. Shellard, Extending the velocity dependent one scale string evolution model, *Phys. Rev. D* **65**, p. 043514 (2002).
12. C. J. A. P. Martins, Scaling properties of cosmological axion strings, *Phys. Lett. B* **788**, 147 (2019).
13. A. Lopez-Eiguren, J. Lizarraga, M. Hindmarsh and J. Urrestilla, Cosmic Microwave Background constraints for global strings and global monopoles, *JCAP* **07**, p. 026 (2017).
14. M. Hindmarsh, J. Lizarraga, A. Lopez-Eiguren and J. Urrestilla, Scaling Density of Axion Strings, *Phys. Rev. Lett.* **124**, p. 021301 (2020).

15. M. Hindmarsh, J. Lizarraga, A. Lopez-Eiguren and J. Urrestilla, Approach to scaling in axion string networks, *Phys. Rev. D* **103**, p. 103534 (2021).
16. M. Gorghetto, E. Hardy and G. Villadoro, Axions from Strings: The Attractive Solution, *JHEP* **07**, p. 151 (2018).
17. A. Vaquero, J. Redondo and J. Stadler, Early seeds of axion miniclusters, *JCAP* **04**, p. 012 (2019).
18. M. Kawasaki, T. Sekiguchi, M. Yamaguchi and J. Yokoyama, Long-term dynamics of cosmological axion strings, *PTEP* **2018**, p. 091E01 (2018).
19. M. Buschmann, J. W. Foster and B. R. Safdi, Early-Universe Simulations of the Cosmological Axion, *Phys. Rev. Lett.* **124**, p. 161103 (2020).
20. M. Gorghetto, E. Hardy and G. Villadoro, More axions from strings, *SciPost Phys.* **10**, p. 050 (2021).
21. M. Buschmann, J. W. Foster, A. Hook, A. Peterson, D. E. Willcox, W. Zhang and B. R. Safdi, Dark Matter from Axion Strings with Adaptive Mesh Refinement (8 2021).
22. T. W. B. Kibble, Topology of Cosmic Domains and Strings, *J. Phys. A* **9**, 1387 (1976).
23. L. Fleury and G. D. Moore, Axion dark matter: strings and their cores, *JCAP* **01**, p. 004 (2016).
24. T. Vachaspati and A. Vilenkin, Formation and Evolution of Cosmic Strings, *Phys. Rev. D* **30**, p. 2036 (1984).
25. M. Hindmarsh, J. Lizarraga, J. Urrestilla, D. Daverio and M. Kunz, Scaling from gauge and scalar radiation in Abelian Higgs string networks, *Phys. Rev. D* **96**, p. 023525 (2017).
26. W. H. Press, B. S. Ryden and D. N. Spergel, Dynamical Evolution of Domain Walls in an Expanding Universe, *Astrophys. J.* **347**, 590 (1989).

Electroweak axion string and superconductivity

Yu Hamada

Theory Center, High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan
E-mail: yuhamada@post.kek.jp

Yoshihiko Abe and Koichi Yoshioka

Department of Physics, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan
E-mail: y.abe@gauge.scphys.kyoto-u.ac.jp
yoshioka@gauge.scphys.kyoto-u.ac.jp

We study axion strings with the electroweak gauge flux in the DFSZ axion model and show that these strings, called electroweak axions, exhibit superconductivity without fermionic zero modes. We also show that the primordial magnetic field in the early universe can induce a large electric current along the string. A pair of the strings carrying such a large current feels a net attractive force between them and can form a Y-shaped junction in the early universe, whose formation probability is roughly estimated to be 1/2.

Keywords: Axion; Cosmic string; Soliton; Higgs

1. Introduction

The strong CP problem is one of the mysteries in the Standard Model (SM) of particle physics. The problem can be naturally solved by the Peccei-Quinn mechanism, in which a global symmetry denoted by $U(1)_{\rm PQ}$ is assumed to be spontaneously broken and provides a (pseudo) Nambu-Goldstone particle, the axion.[1–4] The axion can also play a role of cold dark matter.

The DFSZ model[5,6] is a popular model providing the axion as well as the KSVZ model.[7,8] In the DFSZ model, the scalar sector of the SM is extended to have two Higgs doublets and one SM-singlet complex scalar. The scalar fields and the SM fermions are assumed to be charged under the $U(1)_{\rm PQ}$ symmetry, which is spontaneously broken by a vacuum expectation value (VEV) of the complex scalar. The axion is a linear combination of imaginary components of the doublets and the complex scalar.

It is known that if the $U(1)_{\rm PQ}$ symmetry is broken after the cosmological inflation, the axion strings, which are global cosmic strings, are created by the Kibble-Zurek mechanism.[9,10] To avoid dominating the energy of the universe, the strings should lose their energy by reconnection. When the probability of the reconnection is sufficiently close to unity, the number density of the strings approach $\mathcal{O}(1)$ values per Hubble length, called the scaling regime.

Table 1. Quantum charges of scalar fields.

	H_1	H_2	S
$SU(2)_W$	**2**	**2**	**1**
$U(1)_Y$	1	1	0
$U(1)_{\rm PQ}$	1	-1	1

It has been believed that axion strings almost always reconnect when they collide. This is because the interaction between the strings is thought to be dominated by exchange of the (massless) axion as a long-range force. This is analogous to the case of the magnetic flux tube in the type-II superconductor. However, when there is a strong attractive force between them, they can form a bound state called Y-junction.[11,12] When the Y-junction is formed, the reconnection does not work successfully, making the evolution of the string network non-trivial. Is such an attractive force realized for axion strings?

Sometimes, cosmic strings can be superconducting when $U(1)_{\rm EM}$ is broken inside them. This is called the superconducting string.[13] Electric current flows along the string without dissipation and can induce magnetic attractive interaction between the strings. This enables them to form Y-junctions!

In this work,[14] we show that the axion string in the DFSZ model can become a superconducting string after the breaking of the electroweak (EW) symmetry.[a] This is because the charged fields, the charged Higgs and W bosons, acquire nonzero values inside the string and the $U(1)_{\rm EM}$ symmetry is spontaneously broken there. This is a similar situation to superconductivity of non-Abelian vortices[21,22] and of the $U(1) \times \tilde{U}(1)$ model considered by Witten.[13] Remarkably, the amount of the supercurrent can be of order of the $U(1)_{\rm PQ}$ breaking scale resulting in large magnetic energy. As a consequence, the strings feel a large magnetic interaction, which can overcome the one from the axion exchange. The formation probability of the Y-junctions is roughly estimated to be 1/2.

2. DFSZ axion model

Throughout this article, we concentrate on the bosonic sector of the DFSZ model. The particle contents and the charge assignments under the SM gauge group and the $U(1)_{\rm PQ}$ are shown in Tab. 1. We introduce a SM-singlet complex scalar S and two $SU(2)_W$ doublets, H_1 and H_2, both with the hypercharge $Y = 1$. The Lagrangian which describes the EW and scalar sectors is written as

$$\mathcal{L} = -\frac{1}{4}\left(Y_{\mu\nu}\right)^2 - \frac{1}{4}\left(W^a_{\mu\nu}\right)^2 + \sum_{i=1,2} |D_\mu H_i|^2 + |\partial_\mu S|^2 - V(H_1, H_2, S). \qquad (1)$$

[a] Our mechanism of superconductivity is different from one using fermionic zeromodes.[15–20] In the DFSZ model, the fermionic zeromodes should be irrelevant for the dynamics of the strings because the mass scale of the SM fermions are much lighter than the $U(1)_{\rm PQ}$ breaking scale.

Here, $Y_{\mu\nu}$ and $W^a_{\mu\nu}$ describe field strength tensors of the hypercharge and weak gauge interactions, respectively, with μ (ν) and a being Lorentz and weak iso-spin indices, respectively. D_μ represents the covariant derivative acting on the Higgs fields, and the index i runs $i = 1, 2$. The scalar potential $V(H_1, H_2, S)$ being invariant under the charge assignments of Tab. 1 is

$$V(H_1, H_2, S) = V_H + V_S + V_{\rm mix}, \tag{2}$$

where each part is given by

$$\begin{aligned} V_H =& m_{11}^2 H_1^\dagger H_1 + m_{22}^2 H_2^\dagger H_2 + \frac{\beta_1}{2}\left(H_1^\dagger H_1\right)^2 + \frac{\beta_2}{2}\left(H_2^\dagger H_2\right)^2 \\ &+ \beta_3\left(H_1^\dagger H_1\right)\left(H_2^\dagger H_2\right) + \beta_4\left(H_1^\dagger H_2\right)\left(H_2^\dagger H_1\right), \end{aligned} \tag{3}$$

$$V_S = - m_S^2|S|^2 + \lambda_S|S|^4, \tag{4}$$

$$V_{\rm mix} = \left(\kappa S^2 H_1^\dagger H_2 + {\rm h.c.}\right) + \kappa_{1S}|S|^2|H_1|^2 + \kappa_{2S}|S|^2|H_2|^2, \tag{5}$$

with $m_S^2 > 0$ which admits S to acquire a non-zero VEV: $\langle S\rangle = v_s$. From astrophysical constraints, v_s is supposed to be $\mathcal{O}(10^{8-12})$ GeV. Without loss of generality, we can suppose that the Higgs fields develop VEVs as $\langle H_1\rangle = (0, v_1)^{\rm T}$, $\langle H_2\rangle = (0, v_2)^{\rm T}$ with $v_1, v_2 \in \mathbb{R}$. For simplicity, we take $v_1 = v_2 \equiv v$ in this article. Then the electroweak scale, $v_{\rm EW}$ ($\simeq 246$ GeV), can be expressed by these VEVs as $v_{\rm EW}^2 = 4v^2$.

3. Electroweak axion strings

Similarly to other axion models, the DFSZ axion model provides a vortex string solution known as the axion string corresponding to the breaking of $U(1)_{\rm PQ}$. On the other hand, after the electroweak phase transition, the axion string can contain flux tubes of the $SU(2)_W \times U(1)_Y$ gauge fields like the Abrikosov-Nielsen-Olesen vortex[23,24] since the two Higgs doublets also acquire the VEVs. We call such vortex strings the EW axion strings.

3.1. *Axion string in DFSZ model*

We first review the conventional axion string in this subsection. Let us consider a case that the $U(1)_{\rm PQ}$ symmetry is spontaneously broken by $\langle S\rangle \neq 0$ but the EW symmetry remains unbroken, $\langle H_1\rangle = \langle H_2\rangle = (0, 0)^{\rm T}$. This situation is realized in the early universe when the temperature of the thermal bath T_{th} satisfies $v_{EW} \ll T_{th} \ll v_s$. In this case, as is well-known, a vortex-string configuration associated with the global $U(1)_{\rm PQ}$ symmetry exists as a solution to the equation of motion (EOM), which is called the axion string in the literature. The configuration located on the z-axis is described by the following ansatz

$$S = v_s e^{i\theta}\phi(r), \qquad H_1 = H_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{6}$$

where r and θ are the distance from the z-axis and the rotational angle, respectively, *i.e.*, $x + iy = re^{i\theta}$. The profile function $\phi(r)$ satisfies the boundary conditions

$$\phi(0) = 0, \qquad \phi(\infty) = 1. \tag{7}$$

The detailed form of $\phi(r)$ is determined by solving the EOM. This string has a winding number associated with the $U(1)_{\rm PQ}$ symmetry, and hence is topologically stable. The $U(1)_{\rm PQ}$ symmetry is restored at the string core because of $\phi(0) = 0$.

After these axion strings are produced in the early universe, their string network evolves to the scaling regime. This is because the interaction between the strings is dominated by exchange of massless axion particles, resulting in the long-range repulsive force for two parallel axion strings with the same topological charge. The potential of the interaction $V_{\rm st.}$ is approximately given as $V_{\rm st.} \sim -v_s^2 \log R$ with R being the distance between the pair.

3.2. *Electroweak axion string with W-flux*

Next, we discuss the EW axion strings. Let us consider a string configuration after the EW phase transition $T_{th} \lesssim v_{\rm EW}$. The two doublets also acquire the VEVs and can have winding phases.[14] There are three patterns of the winding phases, only one of which (called type-C EW axion string in Ref. 14) is studied in the following.

We consider a configuration that has winding phases in $U(1)_{\rm PQ}$ and in the $U(1)_{W^1}$ subgroup of the $SU(2)_W$ gauge group. Its asymptotic form is given as

$$S \sim v_s e^{i\theta}\,, \tag{8}$$

$$H_1 \sim \frac{1}{2} v e^{i\theta} e^{i\theta\sigma_1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad H_2 \sim \frac{1}{2} v e^{-i\theta} e^{i\theta\sigma_1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{9}$$

at $r \to \infty$. The exponential factor with σ_1 means a rotation in the direction of $U(1)_{W^1}$. Since these configuration are singular at $r = 0$, they must be smeared by introducing profile functions as follows:

$$\begin{cases} S = v_s e^{i\theta} \phi(r) \\ H_1 = \frac{1}{2} v_1 e^{i\theta} \begin{pmatrix} f(r) e^{i\theta} - h(r) e^{-i\theta} \\ f(r) e^{i\theta} + h(r) e^{-i\theta} \end{pmatrix} \\ H_2 = \frac{1}{2} v_2 e^{-i\theta} \begin{pmatrix} h(r) e^{i\theta} - f(r) e^{-i\theta} \\ h(r) e^{i\theta} + f(r) e^{-i\theta} \end{pmatrix} \end{cases} \tag{10}$$

The profile functions $f(r), h(r)$ and $\phi(r)$ satisfy the following boundary conditions,

$$f(0) = \phi(0) = 0, \quad \partial_r h|_{r=0} = 0, \quad f(\infty) = h(\infty) = \phi(\infty) = 1. \tag{11}$$

Note that $h(r)$ does not necessarily vanish at the center of the string. Due to the non-zero value of $h(0)$, the EW symmetry is not restored inside the string.

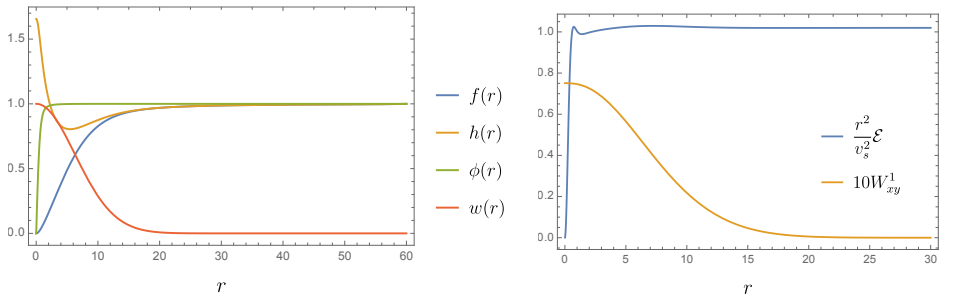


Fig. 1. Numerical solution for the EW axion string. Also we adopt a length unit as $v_s^{-1} = 0.5$. (left): Plots of profile functions. The profile function of the W^1 field, $w(r)$, approaches to zero as $r \to \infty$ starting from unity at $r = 0$. (right): Plots of energy density divided by v_s^2/r^2 and the W^1-flux density multiplied by 10. The tension integrated over $0 \le r \le 120\, v_s^{-1}$ is 140.604. The total value of the W^1-flux is calculated to 19.3208, which is consistent with Eq. (14).

To cancel the gradient energy from the windings for $U(1)_{W^1}$ at infinity, the W^1 gauge field is induced as

$$W_i^1 = \frac{-2}{g}\frac{\epsilon_{ij}x_j}{r^2}(1 - w(r)), \tag{12}$$

and $A_i = Z_i = 0$. The profile function $w(r)$ satisfies

$$w(0) = 1, \quad w(\infty) = 0. \tag{13}$$

This string has the confined W^1-flux

$$\Phi_{W^1} = \oint_{r=\infty} dx_i\, W_i^1 = \frac{4\pi}{g}. \tag{14}$$

This feature is similar to the electroweak strings[25,26] in the SM and its extended models. The detailed behavior of the profile functions, ϕ, f, h, and w are obtained by solving the EOMs numerically. The obtained numerical solutions are shown in Fig. 1.

The most important feature of this string is that the $U(1)_{\rm EM}$ symmetry is spontaneously broken inside the string core because the W^1 boson acquires a non-zero value Eqs. (12) and (14). (The charged component of the Higgs doublets also acquires a non-zero value, see Ref. 14). Actually, that is generally inevitable for a configuration with a non-vanishing winding number for charged components,[21,22] i.e., a configuration whose asymptotic form is $\exp[i\hat{T}(\theta)]H_i(\theta = 0)$ with a non-Abelian generator $\hat{T}(\theta)$ satisfying $[\hat{T}(\theta), \hat{Q}] \neq 0$, where $\hat{Q}$ is the generator of $U(1)_{\rm EM}$. ($\hat{T} = \sigma_1\theta$ in our case.) Due to this breaking of $U(1)_{\rm EM}$, the string is a superconducting string. We discuss its consequence in the next section.

In Ref. 14, other two types of the EW axion strings are studied. Since they do not break $U(1)_{\rm EM}$ inside it and are not superconducting ones, we do not consider them in this article.

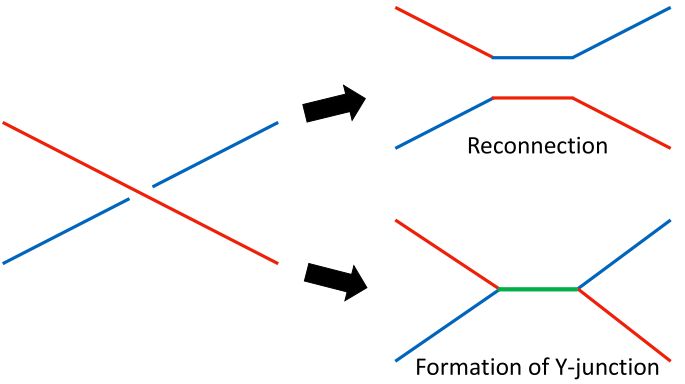


Fig. 2. Reconnection process and formation of a Y-junction.

4. Superconductivity and Y-junction formation

In the EW axion string presented above, the W^1-flux is confined, and the $U(1)_{\rm EM}$ symmetry is broken by the charged Higgs components and the W^1 gauge field, which means that there exists a zero mode (massless excitation) propagating on the string with the speed of light. It can carry an electric current without resistance, resulting in a superconducting current. We will show that such a current changes the interaction between the strings and can lead to formations of the Y-junctions in the early universe.

Firstly, let us assume the presence of a primordial magnetic field (PMF) in the universe after the electroweak phase transition. There are the observational lower and upper bounds on the present strength of intergalactic magnetic fields for a coherent length $\lambda \gtrsim 0.1\,\mathrm{Mpc}$ as $3 \times 10^{-16}\,\mathrm{G} \lesssim B_0 \lesssim 10^{-11}\,\mathrm{G}$, given by Refs. 27, 28. For simplicity, we further assume that the power spectrum of PMF is scale invariant and almost coherent over the entire Hubble horizon.

We then show that PMF can induce large superconducting currents on the strings. Such large currents are sufficient to form the Y-junctions. The scale invariant PMF evolves as

$$B(t) \propto a(t)^{-2}\,, \tag{15}$$

where $B(t)$ and $a(t)$ are the magnetic field strength and the scale factor at the cosmological time t, respectively. It is convenient to define a ratio of the energy density of PMF (ρ_B) to that of photon (ρ_γ),

$$\epsilon \equiv \frac{\rho_B}{\rho_\gamma} \sim B(t)^2 G_N t^2, \tag{16}$$

with G_N being the Newton constant. The ratio ϵ is a constant as the universe expands. If we take the upper limit on B_0, we have $\epsilon \sim 10^{-11}$.

In the early universe, the superconducting strings move with velocity $v_{\rm str}$ ($\sim \mathcal{O}(1)$) in the presence of PMF, and hence feel the electric fields $E \sim B(t) v_{\rm str}$, which

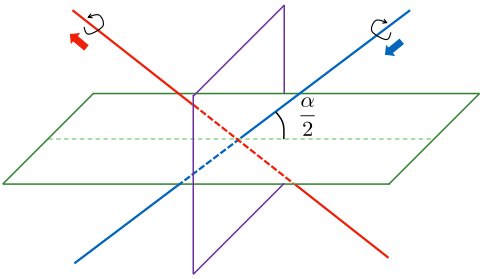


Fig. 3. Two strings (red and blue lines) collide with the crossing angle α. The black arrows indicate the directions of the topological winding number of $U(1)_{\rm PQ}$. The red and blue arrows indicate the directions of the flowing electric currents, which are parallel for $\alpha = 0$. The green and purple planes are orthogonal and include the collision point.

induce the superconducting current,

$$J_{\rm PMF} \sim e^2 B(t) \xi v_{\rm str}\,, \tag{17}$$

with ξ being a string typical length. Since there are no superconducting strings before the electroweak phase transition, we can assume that the scaling property $\xi \sim t$ holds at least just after the phase transition, and obtain

$$J_{\rm PMF} \sim 10^{12} \left(\frac{\epsilon}{10^{-11}}\right)^{1/2} {\rm GeV} \sim 10^{12} \left(\frac{B_0}{10^{-11}\,{\rm G}}\right) {\rm GeV}\,. \tag{18}$$

Note that this is independent of t. For the upper limit of B_0, yielding $\epsilon \sim 10^{-11}$, the induced current Eq. (18) is $\sim 10^{12}\,{\rm GeV}$. On the other hand, it is shown[14] that the EW axion string can carry an electric current at most $J_{\rm max} \simeq v_s/e$ with e the $U(1)_{\rm EM}$ coupling constant. Thus the current saturates at this value for $v_s \simeq 10^{9-12}\,{\rm GeV}$. In the following argument, we use this maximum current as those the strings carry in the early universe.

Let us discuss the formation of Y-junctions. They can be formed by collisions of two superconducting strings when they feel an attractive force and are trapped in the potential. We suppose that the two strings collide with the crossing angle α and that they are identical and parallel for $\alpha = 0$. The superconducting currents are assumed to flow in the same direction for $\alpha = 0$ (the opposite case will be considered later). The three-dimensional dynamics of the collision event is rather complicated and difficult to analyze. However a qualitative picture can be understood by reducing it on two orthogonal planes including the colliding point[29–31] (see Fig. 3). On one of them (purple plane in Fig. 3), the situation is regarded as a collision (scattering) event of two point-like vortices in two dimensions while, on the other plane (green plane in Fig. 3), as a vortex-antivortex collision resulting in the annihilation in two dimensions. Note that this picture reproduces the reconnection process shown in the upper-right picture in Fig. 2 when the strings have no superconducting current; two scattered vortices on the purple plane and nothing on the green one.

We concentrate on the two-dimensional analysis of the superconducting vortex-vortex collision since the current does not change the annihilation. Unlike the non-superconducting case leading to the 90° scattering, the long-range magnetic interaction plays a crucial role in our case. On the reduced plane, the vortex-vortex pair feel a net interaction potential,

$$V \sim \frac{v_s^2}{e^2} \cos\frac{\alpha}{2} \log r \,, \tag{19}$$

where r is the distance between the two vortices and we have ignored the axion-mediated repulsive interaction, which should be subleading by a factor e^2 compared to Eq. (19). If the vortices do not have sufficient kinetic energy to escape to infinity, they become trapped by the potential. Thus we obtain a condition to form a bound state,

$$\frac{v_s^2}{e^2} \cos\frac{\alpha}{2} \log\frac{L}{\delta} \geq (\gamma - 1)\mu \,, \tag{20}$$

where γ and μ are the Lorentz factor and the tension of the colliding vortices. L and δ are the IR and UV cutoff for the potential, which we take as the Hubble radius and the width of the strings, respectively. As considering just after the electroweak phase transition, the logarithmic factor gives $\log(10^{17} \times v_s/v_{\rm EW}) \sim 50$. In addition, we take a mildly relativistic velocity ~ 0.6 for the vortices, yielding $\gamma \sim 1.25$, and thus Eq. (20) gives

$$\cos\frac{\alpha}{2} \geq 10^{-3} \,, \tag{21}$$

which means that they *almost always* form the bound state except for $\alpha \simeq \pi$.

On the other hand, they cannot form such a bound state in the case that the currents flow in the opposite directions (anti-parallel) on the vortex-vortex plane because they always feel a repulsive force. Therefore, we obtain a rough estimation for the formation probability of the bound state,

$$\begin{cases} \text{Y-junction formed for current in the same directions} \\ \text{not formed for current in the opposite directions} \end{cases}$$
$$\Rightarrow \text{probability} \simeq \frac{1}{2} \,. \tag{22}$$

Note that this is based on the two-dimensional analysis focusing on the collision point. This picture may break down for the case that the dynamics after colliding is dominated by other parts of the strings than the collision point. Furthermore, the formed Y-junction could be peeled off, depending on the velocities and the crossing angle of the strings. For a more detailed study including these effects, it is necessary to perform a three-dimensional simulation, which is beyond the scope of the present paper.

Finally, we discuss a consequence of the formation of Y-junctions. When a superconducting string intersects with itself, it forms the Y-junction with probability 1/2,

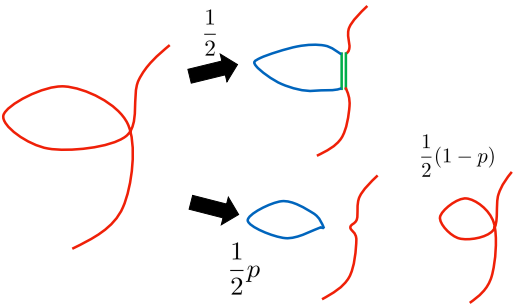


Fig. 4. Self-intersection of a superconducting string. The Y-junction (green doubled line) is formed with probability $\frac{1}{2}$. Otherwise, it reconnects producing a small loop (blue line) or passes through. The reconnection probability is $p \sim 1$.

produces a small loop with probability $p/2$, or passes through with probability $(1-p)/2$ (see Fig. 4), where p is the reconnection probability without the current ($p \sim 1$). If such a Y-junction-connected loop (upper one in Fig. 4) is produced, it is not obvious whether the strings sufficiently loose their energy, and the scaling behavior of the energy density of strings is not ensured. Once the energy density of the universe is dominated by such superconducting strings, it causes large impacts on cosmology, and hence some parameter region in the DFSZ axion model may be constrained.

5. Conclusion

We have studied the axion strings with the electroweak gauge flux, and their superconductivity in the DFSZ model. The superconducting current is induced by PMF in the early universe, and a large electric current can flow along the string. This large current may realize a net attractive force between the axion strings, which could form the so-called Y-junctions in the early universe. By considering the string collision reduced onto two-dimensional planes, the probability of the formation of the Y-junctions is estimated to be $\sim 1/2$.

Once such Y-junctions are formed, they can affect the evolution of the string network and make non-trivial whether the network evolves to the scaling solution. If no obeying the scaling behavior, the string network could dominate the energy density of the the universe and the model is severely constrained. To conclude whether this is true or not, we need detailed numerical simulations on the time evolution of the network taking into account both the axionic and magnetic interactions.

References

1. R. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev. D* **16** (1977) 1791.
2. R. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440.

3. S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223.
4. F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279.
5. A. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*, *Sov. J. Nucl. Phys.* **31** (1980) 260.
6. M. Dine, W. Fischler and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett. B* **104** (1981) 199.
7. J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43** (1979) 103.
8. M. A. Shifman, A. Vainshtein and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys. B* **166** (1980) 493.
9. T. W. B. Kibble, *Some Implications of a Cosmological Phase Transition*, *Phys. Rept.* **67** (1980) 183.
10. W. H. Zurek, *Cosmological Experiments in Superfluid Helium?*, *Nature* **317** (1985) 505.
11. L. M. Bettencourt, P. Laguna and R. A. Matzner, *Nonintercommuting cosmic strings*, *Phys. Rev. Lett.* **78** (1997) 2066 [hep-ph/9612350].
12. L. Bettencourt and T. Kibble, *Nonintercommuting configurations in the collisions of type I U(1) cosmic strings*, *Phys. Lett. B* **332** (1994) 297 [hep-ph/9405221].
13. E. Witten, *Superconducting Strings*, *Nucl. Phys. B* **249** (1985) 557.
14. Y. Abe, Y. Hamada and K. Yoshioka, *Electroweak axion string and superconductivity*, *JHEP* **06** (2021) 172 [2010.02834].
15. G. Lazarides and Q. Shafi, *Superconducting Strings in Axion Models*, *Phys. Lett. B* **151** (1985) 123.
16. A. Iwazaki, *Spontaneous magnetization of axion domain wall and primordial magnetic field*, *Phys. Rev. Lett.* **79** (1997) 2927 [hep-ph/9705456].
17. N. Ganoulis and G. Lazarides, *Fermionic Zero Modes for Cosmic Strings*, *Nucl. Phys. B* **316** (1989) 443.
18. G. Lazarides, C. Panagiotakopoulos and Q. Shafi, *Cosmic Superconducting Strings and Colliders*, *Nucl. Phys. B* **296** (1988) 657.
19. H. Fukuda, A. V. Manohar, H. Murayama and O. Telem, *Axion strings are superconducting*, 2010.02763.
20. P. Agrawal, A. Hook, J. Huang and G. Marques-Tavares, *Axion string signatures II: A cosmological plasma collider*, 2010.15848.
21. M. G. Alford, K. Benson, S. R. Coleman, J. March-Russell and F. Wilczek, *The Interactions and Excitations of Nonabelian Vortices*, *Phys. Rev. Lett.* **64** (1990) 1632.
22. M. G. Alford, K. Benson, S. R. Coleman, J. March-Russell and F. Wilczek, *Zero modes of nonabelian vortices*, *Nucl. Phys. B* **349** (1991) 414.
23. A. A. Abrikosov, *On the Magnetic properties of superconductors of the second group*, *Sov. Phys. JETP* **5** (1957) 1174.
24. H. B. Nielsen and P. Olesen, *Vortex Line Models for Dual Strings*, *Nucl. Phys.* **B61** (1973) 45.
25. Y. Nambu, *String-Like Configurations in the Weinberg-Salam Theory*, *Nucl. Phys.* **B130** (1977) 505.
26. T. Vachaspati, *Vortex solutions in the Weinberg-Salam model*, *Phys. Rev. Lett.* **68** (1992) 1977.
27. A. Neronov and I. Vovk, *Evidence for strong extragalactic magnetic fields from fermi observations of tev blazars*, *Science* **328** (2010) 73–75.

28. K. Jedamzik and A. Saveliev, *Stringent limit on primordial magnetic fields from the cosmic microwave background radiation*, *Physical Review Letters* **123** (2019).
29. E. J. Copeland and N. Turok, *Cosmic String Interactions*.
30. E. P. S. Shellard, *Cosmic String Interactions*, *Nucl. Phys. B* **283** (1987) 624.
31. E. P. S. Shellard, *Understanding intercommuting*, *Proceedings of Yale Workshop: Cosmic Strings: The Current Status* (1988).

Hints for the $H_0 - r_d$ tension in uncorrelated Baryon Acoustic Oscillations dataset

Denitsa Staicova

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria
E-mail: dstaicova@inrne.bas.bg

Baryon Acoustic Oscillations (BAO) datasets use very precise measurements of the spatial distribution of large-scale structures as a distance ladder to help constrain cosmological parameters. In a recent article,[1] we combined 17 uncorrelated BAO measurements in the effective redshift range $0.106 \leq z \leq 2.36$ with the Cosmic Chronometers data, the Pantheon Type Ia supernova and the Hubble Diagram of Gamma Ray Bursts and Quasars to obtain that the ΛCDM model fit infers for the Hubble constant: $69.85 \pm 1.27 km/sec/Mpc$ and for the sound horizon distance: $146.1 \pm 2.15 Mpc$. Beyond the ΛCDM model we test Ω_kCDM and wCDM and we get $\Omega_k = -0.076 \pm 0.012$, $w = -0.989 \pm 0.049$ accordingly. In this proceeding we present elaborate on our findings and we compare them to other recent results in the literature.

Keywords: Baryon Acoustic Oscillations, Dark Energy, Dark Matter, Large Scale Structure, Hubble Tension

1. Introduction

The $\sim 4\sigma$ tension between the Hubble parameter measured by late universe observations by the SH0ES collaboration[2] and the one measured from the cosmic microwave background (CMB) by the Planck Collaboration[3] is one of the major stumbling block in front of modern cosmology and the theories aiming to explain the evolution of the Universe.[4–22] The default ΛCDM model which uses a combination between cold dark matter and dark energy components has been shown to fit remarkably well current astronomical observations yet it fails to explain not only the beginning of the universe (the inflationary epoch) but also the Hubble tension and the related σ_8 tension.

In a recent article[1] we selected 17 uncorrelated BAO points from the largest collection of BAO data points (333 points). We then combined them with the Cosmic Chronometers data, the Pantheon Type Ia supernova, and the Hubble Diagram of Gamma-Ray Bursts and Quasars. From this combination of datasets, referred in the article sometimes as the *full* dataset, we found: the Hubble constant yields $69.85 \pm 1.27 km/sec/Mpc$, the sound horizon distance is $146.1 \pm 2.15 Mpc$ and the matter energy density – $\Omega_m = 0.271 \pm 0.016$. If one uses the so called *Riess prior* (denoted here as R19) to constrain H_0 by the model-independent local universe measurement,[2] one gets: $H_0 = 71.40 \pm 0.89$, $r_d = 143.5 \pm 2.0$ and $\Omega_m = 0.267 \pm 0.017$. Beyond the ΛCDM model we test Ω_KCDM and wCDM. The spatial curvature is

$\Omega_k = -0.076 \pm 0.012$ and the dark energy equation of states is $w = -0.989 \pm 0.049$. In this proceeding we discuss how our results are situated regarding other published results by emphasizing on the need to consider the H_0-tension in the context of the $H_0 - r_d$ plane or even of the $H_0 - r_d - \Omega_m$ plane.

2. Overview of the used datasets

The Baryon acoustic oscillations (BAO) are fluctuations in the photon-baryonic plasma that froze at recombination at the so called drag sound horizon. Because the sound horizon can be calculated rather simply from basic assumptions of the pre-recombination plasma, they provide a standard ruler which can be seen in the clustering of large scale structures. This provides an independent way to probe cosmological parameters, complimentary to this of the Supernova and the CMB surveys, see.[23–25] The BAO peak can be measured from objects with different nature and using different methods. For example: the BOSS experiment measures the clustering of different galaxies: emission-line galaxies (ELGs), luminous red galaxies (LRGs), and quasars, and also from the correlation function of the Lyman-alpha (Lyα) absorption lines in the spectra of distant quasars etc. The peak can be seen on different redshifts, providing us with a standard ruler evolving with the Universe since the recombination epoch.[26,27]

The final dataset we use is a set of uncorrelated data points from different **BAO** measurements: the Sloan Digital Sky Survey (SDSS), the WiggleZ Dark Energy Survey, Dark Energy Camera Legacy Survey (DECaLS), the Dark Energy Survey (DES), the 6dF Galaxy Survey (6dFGS).[28–43] To this dataset, we add **cosmic chronometers (CCs)** (30 uncorrelated CC measurements of $H(z)$[44–47]), and **standard candles (SCs)** (the Pantheon Type Ia supernova dataset[48–51]), and quasars[52] and gamma-ray bursts[53] (186 points).

3. Theoretical background

We use the following theoretical setup. If one assumes a Friedmann-Lemaître-Robertson-Walker metric with the scale parameter $a = 1/(1+z)$, where z is the redshift, one gets for the Friedmann equation for the ΛCDM model:

$$E(z)^2 = \Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda, \tag{1}$$

where Ω_r, Ω_m, Ω_Λ, and Ω_k are respectively the fractional densities of radiation, matter, dark energy, and the spatial curvature at redshift $z = 0$. Here $E(z) = H(z)/H_0$, and $H(z) = \dot{a}/a$ is the Hubble parameter at z, while H_0 is the Hubble parameter today. The radiation density can be computed as $\Omega_r = 1 - \Omega_m - \Omega_\Lambda - \Omega_k$. For wCDM the Friedmann equation is generalized to $\Omega_\Lambda \to \Omega^0_{DE}(1+z)^{-3(1+w)}$, while $\Omega_k = 0$ represents a flat universe.

Since in cosmology one deals with the measurements of angles and redshifts, it is needed to connect the different cosmological distances with the observational

quantities. The comoving angular diameter distance:[54,55]

$$D_M = \frac{c}{H_0} S_k \left(\int_0^z \frac{dz'}{E(z')} \right), \tag{2}$$

where one accounts for non-zero spatial curvature with:

$$S_k(x) = \begin{cases} \frac{1}{\sqrt{\Omega_k}} \sinh\left(\sqrt{\Omega_k} x\right) & \text{if} \quad \Omega_k > 0 \\ x & \text{if} \quad \Omega_k = 0 \\ \frac{1}{\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k} x\right) & \text{if} \quad \Omega_k < 0 \end{cases}. \tag{3}$$

The other distances we use are the Hubble distance $D_H(z) = c/H(z)$, the angular diameter distance $D_A = D_M/(1+z)$ and the volume averaged distance:

$$D_V(z) \equiv [z D_H(z) D_M^2(z)]^{1/3}. \tag{4}$$

As said before the BAO use as a standard ruler depends only on the sound horizon r_d at the drag epoch ($z_d \approx 1060$) when photons and baryons decouple:

$$r_d = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)} dz, \tag{5}$$

where $c_s \approx c\,(3 + 9\rho_b/(4\rho_\gamma))^{-0.5}$ is the speed of sound in the baryon-photon fluid with the baryon $\rho_b(z)$ and the photon $\rho_\gamma(z)$ densities, respectively.[56] One needs to acknowledge that since the actual measured quantities are the projections $\Delta z = r_d H/c$ and $\Delta\theta = r_d/(1+z)D_A(z)$, where Δz and $\Delta\theta$ are the redshift and the angular separation, from BAO one can get information only about the quantity $r_d \times H$. Thus in order to decouple these quantities, one needs some kind of independent measurement of H_0 or an assumption for r_d. Here we take r_d as an independent parameter but use the additional datasets (CC+SC) to decouple the two variables. In a recent paper,[57] we instead use the combination $r_d \times H_0$ as a parameter to break the degeneracy.

4. Numerical methods

To deal with the possible correlations in the BAO dataset, we perform a covariance analysis based on the one proposed in Ref.[58]. We transform the standard covariance matrix for uncorrelated points C_{ii} into C_{ij} as follows:

$$C_{ii} = \sigma_i^2 \quad \longrightarrow \quad C_{ij} = \sigma_i^2 + 0.5\sigma_i\sigma_j \tag{6}$$

by adding randomly certain number of nondiagonal elements while keeping it symmetric. Here $\sigma_i\sigma_j$ are the published 1σ errors of the data points i, j. With this approach we show that the effect of up to 25% random correlations with this magnitude results in less than 10% deviation in the final results, thus it is minimal and the points can be considered uncorrelated.

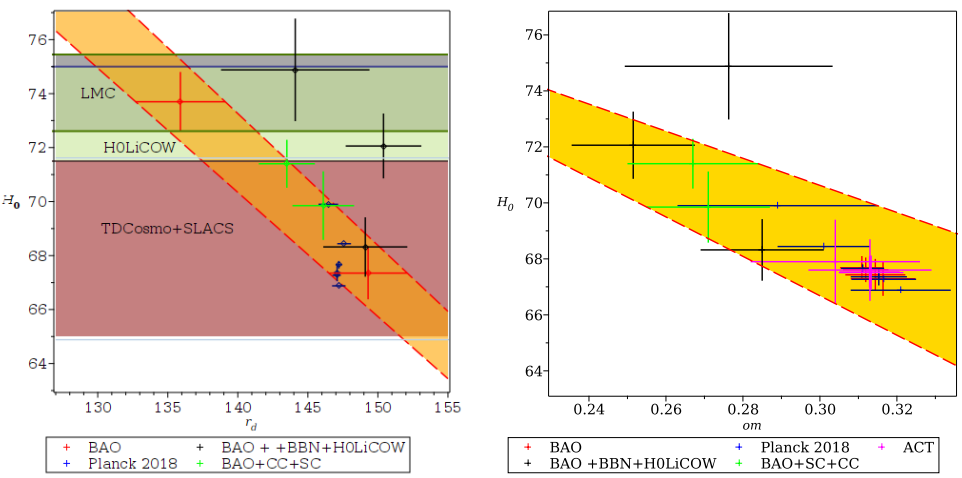
We use a nested sampler as implemented within the open-source package *Polychord*[59] with the *GetDist* package[60] to present the results. The priors we use

to obtain the results and the averaged mean values across the 3 models (ΛCDM, wCDM, Ω_kCDM) can be found in Table A1.

5. Comparison with other results

The complete numerical results for the three models and the three datasets (BAO+R19, BAO+SC+CC, BAO+SC+CC+R19) can be found in Table 3 in.[1] In this article we will focus on comparing our results with other works, by using only the *full* datasets with and without the R19 prior, referred to as BAO+SC+CC+R19 and BAO+SC+CC.

Specifically, on Fig. 1, we use some known measurements in the $H_0 - r_d$ and $H_0 - \Omega_m$ plane. It has been discussed in a number of papers, more notably,[61] that the H_0-tension is actually extended to a tension with respect to the main parameters H_0, r_d and Ω_m, since increasing Ω_m to reduce r_d leads to the opposite to the desired effect on H_0. To demonstrate this, we have plotted the following points: TDCosmo IV,[62] H0LiCOW XIII,[63] LMC Cepheids,[2] Planck 2018,[3] eBOSS SDSS-IV,[64] BAO +BBN+H0LiCOW and BAO +BBN+CC,[65] all listed for convenience in Table A2. Note, some references do not measure r_d, others omit mentioning Ω_m.



(a) (b)

Fig. 1. (a) $H_0 - r_d$ plane comparison of different results. (b) $H_0 - \Omega_m$ plane comparision. Our points are with green, for legend see the text.

On Fig. 1(a) we consider only $H_0 - r_d$ and we see that our results (in green) fit nicely between the results of the pure BAO points by eBOSS (in red) and the Planck 2018 points (in blue).

On this plot, the points we obtain with the R19 prior (top green point) are very close to the local universe measurements (LMC), while the points we obtain with

a uniform prior (the lower green point) are close to the early universe (i.e. Planck 2018 with bue). The numbers we obtain are surprisingly close to a measurement of the Tip of the Red Giants Branch $H_0 = 69.8 \pm 1.9(km/s)/Mpc$[66] — an independent measurement based on the distances to nearby galaxies (in this case based on TRGB stars in the Large Magellanic Cloud, grounded on detached eclipsing binaries).

As for r_d, we see that our points without the R19 prior are again close to the Planck measurement (Planck 2018 $r_d = 147.09 \pm 0.26Mpc$, ours: $r_d = 146.1 \pm 2.2Mpc$). The final measurements from the completed SDSS lineage of experiments in large-scale structure provide $r_d = 149.3 \pm 2.8Mpc$.[67] Using BAO, SNea, the ages of early-type galaxies, and local determinations of the Hubble constant, Ref.[68] reports $r_d = 143.9 \pm 3.1$ Mpc. Thus, one can see clearly the tension between different results from the early and late universe confirming the "tensions in the $r_d - H_0$ plane."[61] The model with the R19 prior gives $r_d = 143.5 \pm 2.0Mpc$. Importantly, the choice of a prior for r_d has a critical effect on H_0 decreasing or increasing the inferred value to a large extend independently of the prior on H_0.

Another part of the $H_0 - r_d$ puzzle is the effect of the matter energy density Ω_m. A comparison of different results for Ω_m can be seen on Fig. 1(b), where we have plotted the $H_0 - \Omega_m$ plane. Note that the eBOSS points used here are different from the ones used in Table 2, as we use different set of published points (Table 4 in Ref.[64]). Also, we add the points by ACT which can be found in Table 4 in Ref.[69]. Our points are once again in green and as we can see they fit very close to Plancks's results in blue which come with very small error on H_0 but rather large on Ω_m ($\Omega_m^{Planck} = .289, 0 \pm 0.03$). In our case, the matter density is lower than the expected. One can also note that removing the top right point (belonging to BAO +BBN+H0LiCOW, see Table 2) will make the points along with the error bars to lie on approximately one plane, similar to the one in the $H_0 - r_d$ plane and thus hinting at possible degeneracy (in orange and yellow respectively). This has been commented already in[61] and is related to the fact we do not know how exactly the phase space of all parameters looks like. The tension for $H_0 - r_d - \Omega_m$ appears in different measurements (Planck, BAO, ACT etc), some from the early Universe, others from the late one and the independent local measurements. This once again poses the question whether we need to consider the full parameter space when discussing the tension.

Finally, we would like to discuss the results we obtained about the spacial curvature energy density. As already mentioned we get $\Omega_k = -0.076 \pm 0.012$ which excludes flat universe at 68% CL and points to a closed Universe ($k = 1$). Due to the rather small prior used to obtain this number, we now repeat the inference with a larger prior $\Omega_k \in [-0.3, 0.3]$. The results can be seen on Fig. 2. Once again, when calculating for the full dataset with or without the R19 prior, we obtain values excluding flat universe. This has been reported already, for example by the Planck 2018 collaboration[3] for CMB alone which found a preference for a closed universe at 3.4σ. Also,[70] using the CC, Pantheon, and BAO measurements concluded that negative Ω_k also relieves the H_0 tension.

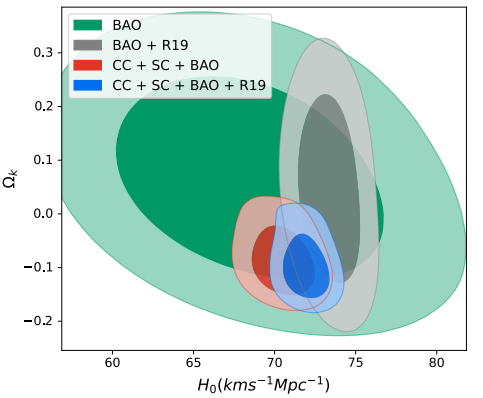


Fig. 2. 2d contour plot of Ω_k vs H_0 for the datasets BAO, BAO + R19, CC+SC+BAO and CC+SC+BAO+R19.

The issue of a possibility for a deviation from a flat Universe is extremely important and has been discussed in a number of works. On Fig. 3, we plot some of the published in the literature results with respect to Ω_k. The reference for the results are accordingly ($N = 1..12$) data from: Planck18Plk, Planck18CamSpec, ACT+WMAP, ACT+Planck, CC+Quasars,CC+SN, BAO+BBN+H0LiCOW, BAO+BBN+CC, taken from[71] to which we add the values we measured with the following dataset: BAO+R19, BAO+CC+SC (same as BAO+CC+SC+R19) and the extended prior points: BAO+CC+SCl, BAO+CC+SC+R19l. One can see that our results seem to add to a mounting evidence that maybe we observe an effect related to a non-flat universe.

The results for the full dataset for the extended prior can be seen in Table A3. We see now that the reported earlier in Ref.[1] lower matter energy density is no longer an issue as we get $\Omega_m \sim 0.326$, also we see that now H_0 is a bit larger, thus pointing to further alleviation of the H_0-tension due to a non-zero spatial curvature. Critically, we see that again $\Omega_k = 0$ is excluded from the 68% CL for both models. Thus we can claim this effect persists with the increase of the prior and it is a feature of the dataset.

On a side note, we mention the result we obtained in Ref.[1] with respect to the wCDM model. For BAO + R19 we get $w < -1$ for ($w = -1.067 \pm 0.065$), while the full datasets seem to tend to $w \geq -1$ ($w = -0.989 \pm 0.049$). Since $w = -1$ is included in the error, the results are essentially consistent with a cosmological constant.

To compare the extended models to ΛCDM we used well known statistical measures, i.e. the Akaike information criteria (AIC)[72–74,a]. We find that the ΛCDM

[a]The definition of AIC we use is $AIC = \chi^2_{min} + 2p$, where p is the number of parameters of the model.

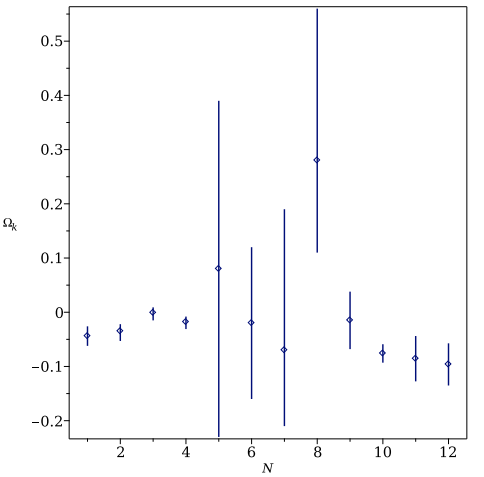


Fig. 3. Comparison of different measurements of Ω_k, see legend in text. Our points are $N = 9-12$, where points with $N = 11, 12$ are with larger prior for Ω_k.

model remains the best fit to the data with difference of 2.3 and 7.6 AIC units for the wCDM and the Ω_kCDM respectively. One can see that wCDM has a little bit more support than Ω_kCDM but also as we mentioned, $w = -1$ enters into the the 68% CL of w for the full dataset which may explain its closeness to ΛCDM in terms of AIC. Our results are consistent with the eBoss collaboration official results.[67] Interestingly, we see that in our case $w \geq -1$ for the full dataset, which differs from the estimations in some of the different cases considered in.[67] Our dataset differs from theirs by the inclusion of the quasars and the GRB data and the exclusion of Planck points, so this may point to a local universe effect.

6. Conclusion

We use a set of 17 uncorrelated BAO points to infer the cosmological parameters for 3 different models: ΛCDM, Ω_kCDM and wCDM. We find that by choosing the sound horizon at drag epoch as an independent parameter and adding additional datasets such as SNe, GRB and quasars, we are able to break the degeneracy between H_0 and r_d and to constrain the cosmological parameters for the different models. The Hubble parameter obtained from the full dataset is very close to the TGRB measurement, while the one with a R19 prior is close to the local universe measurement. We show that we are able to alleviate the H_0-tension to certain degree but not entirely. Another interesting result is the prediction of a non-flat but closed Universe at 68 % CL, which has been confirmed with an increased prior and seems to add to the mounting evidences there may be a deviation from flatness.

Acknowledgments

D.S. is thankful to Bulgarian National Science Fund for support via research grants KP-06-N38/11. We have received partial support from European COST actions CA18108.

References

1. D. Benisty and D. Staicova, Testing late-time cosmic acceleration with uncorrelated baryon acoustic oscillation dataset, *Astron. Astrophys.* **647**, p. A38 (2021).
2. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *Astrophys. J.* **876**, p. 85 (2019).
3. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
4. E. Di Valentino *et al.*, Cosmology Intertwined I: Perspectives for the Next Decade (2020).
5. E. Di Valentino *et al.*, Cosmology Intertwined II: The Hubble Constant Tension (2020).
6. E. Di Valentino *et al.*, Cosmology Intertwined III: $f\sigma_8$ and S_8 (2020).
7. G. Efstathiou, A Lockdown Perspective on the Hubble Tension (with comments from the SH0ES team) (2020).
8. S. Borhanian, A. Dhani, A. Gupta, K. G. Arun and B. S. Sathyaprakash, Dark Sirens to Resolve the Hubble-Lemaître Tension (2020).
9. A. Hryczuk and K. Jodłowski, Self-interacting dark matter from late decays and the H_0 tension, *Phys. Rev.* **D102**, p. 043024 (2020).
10. A. Klypin, V. Poulin, F. Prada, J. Primack, M. Kamionkowski, V. Avila-Reese, A. Rodriguez-Puebla, P. Behroozi, D. Hellinger and T. L. Smith, Clustering and Halo Abundances in Early Dark Energy Cosmological Models (2020).
11. M. M. Ivanov, Y. Ali-Haïmoud and J. Lesgourgues, H0 tension or T0 tension?, *Phys. Rev.* **D102**, p. 063515 (2020).
12. A. Chudaykin, D. Gorbunov and N. Nedelko, Combined analysis of Planck and SPTPol data favors the early dark energy models, *JCAP* **2008**, p. 013 (2020).
13. K.-F. Lyu, E. Stamou and L.-T. Wang, Self-interacting neutrinos: solution to Hubble tension versus experimental constraints (2020).
14. G. Alestas, L. Kazantzidis and L. Perivolaropoulos, H_0 Tension, Phantom Dark Energy and Cosmological Parameter Degeneracies (2020).
15. P. Motloch and W. Hu, Lensinglike tensions in the *Planck* legacy release, *Phys. Rev.* **D101**, p. 083515 (2020).
16. N. Frusciante, S. Peirone, L. Atayde and A. De Felice, Phenomenology of the generalized cubic covariant Galileon model and cosmological bounds, *Phys. Rev.* **D101**, p. 064001 (2020).
17. D. Staicova, Special cases of the Multi-Measure Model – understanding the prolonged inflation (11 2020).
18. D. Staicova and M. Stoilov, Cosmology from multimeasure multifield model, *Int. J. Mod. Phys. A* **34**, p. 1950099 (2019).
19. D. Benisty, E. I. Guendelman, A. van de Venn, D. Vasak, J. Struckmeier and H. Stoecker, The Dark side of the torsion: Dark Energy from kinetic torsion (9 2021).
20. L. Aresté Saló, D. Benisty, E. I. Guendelman and J. d. Haro, Quintessential inflation and cosmological seesaw mechanism: reheating and observational constraints, *JCAP* **07**, p. 007 (2021).

21. D. Benisty, Reducing the σ_8 tension with the Redshift Space Distortion data set (2020).
22. S. Bahamonde, K. F. Dialektopoulos, C. Escamilla-Rivera, G. Farrugia, V. Gakis, M. Hendry, M. Hohmann, J. L. Said, J. Mifsud and E. Di Valentino, Teleparallel Gravity: From Theory to Cosmology (6 2021).
23. W. Handley, Curvature tension: evidence for a closed universe (8 2019).
24. E. Di Valentino, A. Melchiorri and J. Silk, Cosmic Discordance: Planck and luminosity distance data exclude LCDM (2020).
25. X. Luo, Z. Huang, Q. Qian and L. Huang, Reaffirming the Cosmic Acceleration without Supernova and CMB (2020).
26. A. Cuceu, J. Farr, P. Lemos and A. Font-Ribera, Baryon Acoustic Oscillations and the Hubble Constant: Past, Present and Future, *JCAP* **10**, p. 044 (2019).
27. W. K. Wu, P. Motloch, W. Hu and M. Raveri, Hubble constant difference between CMB lensing and BAO measurements, *Phys. Rev. D* **102**, p. 023510 (2020).
28. F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders and F. Watson, The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant, *Mon. Not. Roy. Astron. Soc.* **416**, 3017 (2011).
29. A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden and M. Manera, The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$, *Mon. Not. Roy. Astron. Soc.* **449**, 835 (2015).
30. W. J. Percival *et al.*, Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Data Release 7 Galaxy Sample, *Mon. Not. Roy. Astron. Soc.* **401**, 2148 (2010).
31. R. Tojeiro *et al.*, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: galaxy clustering measurements in the low redshift sample of Data Release 11, *Mon. Not. Roy. Astron. Soc.* **440**, 2222 (2014).
32. C. Blake *et al.*, The WiggleZ Dark Energy Survey: Joint measurements of the expansion and growth history at z ¡ 1, *Mon. Not. Roy. Astron. Soc.* **425**, 405 (2012).
33. H.-J. Seo *et al.*, Acoustic scale from the angular power spectra of SDSS-III DR8 photometric luminous galaxies, *Astrophys. J.* **761**, p. 13 (2012).
34. L. Anderson *et al.*, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample, *Mon. Not. Roy. Astron. Soc.* **427**, 3435 (2013).
35. S. Sridhar, Y.-S. Song, A. J. Ross, R. Zhou, J. A. Newman, C.-H. Chuang, F. Prada, R. Blum, E. Gaztañaga and M. Landriau, The clustering of LRGs in the DECaLS DR8 footprint: distance constraints from baryon acoustic oscillations using photometric redshifts (2020).
36. J. E. Bautista *et al.*, The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations at redshift of 0.72 with the DR14 Luminous Red Galaxy Sample, *Astrophys. J.* **863**, p. 110 (2018).
37. T. M. C. Abbott *et al.*, Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1, *Mon. Not. Roy. Astron. Soc.* **483**, 4866 (2019).
38. J. Hou *et al.*, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the Quasar Sample in configuration space between redshift 0.8 and 2.2 (2020).
39. M. Ata *et al.*, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, *Mon. Not. Roy. Astron. Soc.* **473**, 4773 (2018).
40. N. G. Busca *et al.*, Baryon Acoustic Oscillations in the Ly-α forest of BOSS quasars, *Astron. Astrophys.* **552**, p. A96 (2013).

41. V. de Sainte Agathe *et al.*, Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14, *Astron. Astrophys.* **629**, p. A85 (2019).
42. P. Carter, F. Beutler, W. J. Percival, C. Blake, J. Koda and A. J. Ross, Low Redshift Baryon Acoustic Oscillation Measurement from the Reconstructed 6-degree Field Galaxy Survey, *Mon. Not. Roy. Astron. Soc.* **481**, 2371 (2018).
43. E. A. Kazin *et al.*, The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature, *Mon. Not. Roy. Astron. Soc.* **441**, 3524 (2014).
44. M. Moresco, L. Verde, L. Pozzetti, R. Jimenez and A. Cimatti, New constraints on cosmological parameters and neutrino properties using the expansion rate of the Universe to z˜1.75, *JCAP* **07**, p. 053 (2012).
45. M. Moresco *et al.*, Improved constraints on the expansion rate of the Universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers, *JCAP* **1208**, p. 006 (2012).
46. M. Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$, *Mon. Not. Roy. Astron. Soc.* **450**, L16 (2015).
47. M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro and D. Wilkinson, A 6% measurement of the Hubble parameter at $z \sim 0.45$: Direct evidence of the epoch of cosmic re-acceleration, *JCAP* **1605**, p. 014 (2016).
48. S. Perlmutter *et al.*, Measurements of Omega and Lambda from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
49. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
50. D. Scolnic *et al.*, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *Astrophys. J.* **859**, p. 101 (2018).
51. F. K. Anagnostopoulos, S. Basilakos and E. N. Saridakis, Observational constraints on Barrow holographic dark energy, *Eur. Phys. J.* **C80**, p. 826 (2020).
52. C. Roberts, K. Horne, A. O. Hodson and A. D. Leggat, Tests of ΛCDM and Conformal Gravity using GRB and Quasars as Standard Candles out to $z \sim 8$ (2017).
53. M. Demianski, E. Piedipalumbo, D. Sawant and L. Amati, Cosmology with gamma-ray bursts: I. The Hubble diagram through the calibrated $E_{\rm p,i}$ - $E_{\rm iso}$ correlation, *Astron. Astrophys.* **598**, p. A112 (2017).
54. N. B. Hogg, M. Martinelli and S. Nesseris, Constraints on the distance duality relation with standard sirens (2020).
55. M. Martinelli *et al.*, Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes (2020).
56. E. Aubourg *et al.*, Cosmological implications of baryon acoustic oscillation measurements, *Phys. Rev. D* **92**, p. 123516 (2015).
57. D. Benisty and D. Staicova, A preference for Dynamical Dark Energy? (7 2021).
58. L. Kazantzidis and L. Perivolaropoulos, Evolution of the $f\sigma_8$ tension with the Planck15/ΛCDM determination and implications for modified gravity theories, *Phys. Rev.* **D97**, p. 103503 (2018).
59. W. J. Handley, M. P. Hobson and A. N. Lasenby, PolyChord: Nested sampling for cosmology, *Mon. Not. Roy. Astron. Soc.* **450**, L61 (2015).
60. A. Lewis, GetDist: a Python package for analysing Monte Carlo samples (2019).
61. L. Knox and M. Millea, Hubble constant hunter's guide, *Physical Review D* **101** (Feb 2020).

62. S. Birrer *et al.*, TDCOSMO - IV. Hierarchical time-delay cosmography – joint inference of the Hubble constant and galaxy density profiles, *Astron. Astrophys.* **643**, p. A165 (2020).
63. K. C. Wong *et al.*, H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, *Mon. Not. Roy. Astron. Soc.* **498**, 1420 (2020).
64. S. Alam *et al.*, Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory, *Phys. Rev. D* **103**, p. 083533 (2021).
65. R. C. Nunes and A. Bernui, $\theta_{\rm BAO}$ estimates and the H_0 tension (2020).
66. W. L. Freedman, B. F. Madore, T. Hoyt, I. S. Jang, R. Beaton, M. G. Lee, A. Monson, J. Neeley and J. Rich, Calibration of the Tip of the Red Giant Branch (TRGB) (2 2020).
67. S. Alam *et al.*, The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory (7 2020).
68. L. Verde, J. L. Bernal, A. F. Heavens and R. Jimenez, The length of the low-redshift standard ruler, *Mon. Not. Roy. Astron. Soc.* **467**, 731 (2017).
69. S. Aiola *et al.*, The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters, *JCAP* **12**, p. 047 (2020).
70. E.-K. Li, M. Du and L. Xu, General Cosmography Model with Spatial Curvature, *Mon. Not. Roy. Astron. Soc.* **491**, 4960 (2020).
71. E. Di Valentino *et al.*, Snowmass2021 - Letter of interest cosmology intertwined IV: The age of the universe and its curvature, *Astropart. Phys.* **131**, p. 102607 (2021).
72. K. P. Burnham and D. R. Anderson, Multimodel inference: Understanding aic and bic in model selection, *Sociological Methods & Research* **33**, 261 (2004).
73. A. R. Liddle, Information criteria for astrophysical model selection, *Mon. Not. Roy. Astron. Soc.* **377**, L74 (2007).
74. F. K. Anagnostopoulos, S. Basilakos and E. N. Saridakis, Bayesian analysis of $f(T)$ gravity using $f\sigma_8$ data, *Phys. Rev.* **D100**, p. 083517 (2019).

Appendix A.

Table A1. The priors used to obtain each cosmological parameter.

Parameter	Prior	Average Value [1]	Average Value R19 [2]
Ω_m	$[0.; 1.]$	0.26 ± 0.017	0.26 ± 0.016
Ω_Λ	$[0.; 1 - \Omega_m]$	0.749 ± 0.013	0.751 ± 0.013
H_0	$[50; 100]$	70.19 ± 1.11	-
H_0^{R19}	74.03 ± 1.42	-	71.68 ± 0.9
$r_d/r_{d,fid}$	$[0.9, 1.1]$	0.996 ± 0.019	0.97 ± 0.013
r_d	$[100; 200]$	145.8 ± 2.37	143.3 ± 1.9
w	$[-1.25; -0.75]$	-0.989 ± 0.049	-0.989 ± 0.049
Ω_k	$[-0.1; 0.1]$	-0.076 ± 0.017	-0.076 ± 0.012

Note: [1]average value for the parameter under the flat prior for H_0, [2]average value for the parameter under the Gaussian R19 prior for H_0. H_0 is in $(km/s)/Mpc$, r_d in Mpc.

Table A2. Numbers used to draw Fig 1.

Mission	Reference	H_0 (in km/s/Mpc)	r_d (in Mpc)	Ω_m
TDCosmo IV	62	$74.5^{+5.6}_{-6.1}$	-	-
TDCOSMO+SLACS	62	$67.4^{+4.1}_{-3.2}$	-	-
H0LiCOW XIII	63	$73.3^{+1.7}_{-1.8}$	-	-
LMC DEBs	2	74.22 ± 1.82	-	-
LMC DEBs and NGC 4258 and Milky Way:	2	74.03 ± 1.42	-	-
Planck 2018				
TT+lowE	3	66.88 ± 0.92	147.21 ± 0.48	0.321 ± 0.013
TE+lowE	3	68.44 ± 0.91	147.59 ± 0.49	0.301 ± 0.012
EE+lowE	3	69.9 ± 2.7,	146.46 ± 0.70	$0.289^{+0.026}_{-0.033}$
TT,TE,EE+lowE	3	67.27 ± 0.60	147.05 ± 0.30	0.3166 ± 0.0084
TT,TE,EE+lowE+lensing	3	67.36 ± 0.54	147.09 ± 0.26	0.3153 ± 0.0073
TT,TE,EE+lowE+lensing+BAO	3	67.66 ± 0.42	147.21 ± 0.23	0.3111 ± 0.0056
eBOSS				
BAO+BBN	64	67.35 ± 0.97	149.3 ± 2.8	0.314 ± 0.008
BAO and distance ladder	64	73.7 ± 1.1	135.9 ± 3.2	-
BAO + SDSS (BAO) + BBN	65	$68.32^{+0.98}_{-1.1}$	$151.9^{+3}_{-2.8}$	$0.27^{+0.015}_{-0.016}$
BAO +BBN+H0LiCOW	65	$74.88 \pm 1.95.1$	144.1 ± 5.3	0.2763 ± 0.027
BAO +BBN+CC	65	$72.06^{+1.2}_{-1.3}$	$150.4^{+2.7}_{-3.3}$	0.2515 ± 0.016

Table A3. The results for the Ω_kCDM model with the increased prior on $\Omega_k \in [-0.3, 0.3]$.

Dataset	Parameter	Value
CC+SC+BAO	$H_0(kms^{-1}Mpc^{-1})$	70.48 ± 1.23
	Ω_k	-0.086 ± 0.042
	Ω_m	0.326 ± 0.027
	Ω_Λ	0.766 ± 0.029
	$r_d(Mpc)$	145.961 ± 2.676
	rat	0.984 ± 0.016
CC+SC+BAO+R19	$H_0(kms^{-1}Mpc^{-1})$	71.91 ± 0.87
	Ω_k	-0.096 ± 0.039
	Ω_m	0.327 ± 0.026
	Ω_Λ	0.776 ± 0.024
	r_d (Mpc)	143.452 ± 1.948
	rat	0.967 ± 0.013

Observational constraints on nonlinear matter extensions of general relativity

E.-A. Kolonia

Department of Physics, University of Patras,
26504 Patras, Greece

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

We present a phenomenological analysis of current observational constraints on classes of FLRW cosmological models in which the matter side of Einstein's equations includes, in addition to the canonical term, a term proportional to some function of the energy-momentum tensor ($T^2 = T_{\alpha\beta}T^{\alpha\beta} = \rho^2 + 3p^2$), or of its trace ($T = \rho - 3p$). Qualitatively, one may think of these models as extensions of general relativity with a nonlinear matter Lagrangian. As such they are somewhat different from the usual dynamical dark energy or modified gravity models: in the former class of models one adds further dynamical degrees of freedom to the Lagrangian (often in the form of scalar fields), while in the latter the gravitational part of the Lagrangian is changed. We study both of these models under two different scenarios: (1) as phenomenological two-parameter or three-parameter extensions of the standard ΛCDM, in which case the model still has a cosmological constant but the nonlinear matter Lagrangian leads to additional terms in Einstein's equations, which cosmological observations tightly constrain, and (2) as alternatives to ΛCDM, where there is no cosmological constant, and the nonlinear matter term would have to provide the acceleration (which would be somewhat closer in spirit to the usual modified gravity models). A comparative analysis of the observational constraints obtained in the various cases provides some insight on the level of robustness of the Λ model and on the parameter space still available for phenomenological alternatives.

Keywords: Cosmology; Dark energy; Modified gravity; Observational constraints.

1. Introduction

The observational evidence for the acceleration of the universe shows that our canonical theories of cosmology and particle physics are at least incomplete, and possibly incorrect. Mapping the dark side of the universe, in order to ascertain the physical mechanism behind this acceleration, in a compelling observational task for current and future facilities. The CosmoESPRESSO team uses the universe as a laboratory to address, with precision spectroscopy and other observational, computational and theoretical tools, this and other grand-challenge questions. In what follows we highlight recent contributions of the CosmoESPRESSO team to this fundamental quest, pertaining to dark energy phenomenology.

There has been some recent interest in the so-called energy-momentum-squared gravity models,[1] where the matter part of Einstein's equations is modified by the addition of a term proportional to $T^2 \equiv T_{\mu\nu}T^{\mu\nu}$, where $T_{\mu\nu}$ is the energy-momentum tensor. This we later extended[2,3] to the more generic form $(T^2)^n$, dubbed energy-momentum-powered gravity. Reference 4 provided low redshift constraints on these models, using in particular the Pantheon Type Ia supernova compilation[5] and a compilation of 38 Hubble parameter measurements.[6]

In practical terms, we may think of these models as extensions to the canonical ΛCDM, in which case the model still has a cosmological constant but the nonlinear matter Lagrangian leads to additional terms in Einstein's equations, and cosmological observations can constrain these additional model parameters. Typically there are two such additional parameters: the power n of the nonlinear part of the Lagrangian and a further parameter (to be defined below) quantifying the contribution of this term to the energy budget of the universe.

Alternatively, we may ask whether a suitably chosen nonlinear Lagrangian can reproduce the low redshift acceleration of the universe in a model which at low redshift only contains matter (plus a subdominant amount of radiation) but no true cosmological constant. In principle such a scenario is conceivable.[1] It is also somewhat closer in spirit to the usual modified gravity models—with the caveat that, as previously mentioned, in the latter models the modification occurs in the gravitational part of the Lagrangian and not in the matter part.

The analysis of Ref. 4 found that these models do not solve the cosmological constant problem *per se*, but they can phenomenologically lead to a recent accelerating universe without a cosmological constant at the cost of having preferred values of the cosmological parameters that are somewhat different from the standard ΛCDM ones. Here we revisit and update these constraints, and also provide new constraints on models where the new terms depend on the trace of the energy-momentum tensor, $T = \rho - 3p$. As before, we can consider both the scenario with a cosmological constant (in which case the model is an extension of ΛCDM) and the scenario without a cosmological constant (in which case we can check whether such models can accelerate at all). These have been qualitatively studied in the literature,[7,8] and in what follows we provided quantitative constraints on them. We note that in all the analysis that follows the Hubble constant is analytically marginalized as discussed in Ref. 9.

2. Energy-momentum-powered models

The general action for these models is[2]

$$S = \frac{1}{2\kappa}\int \left[R + \eta(T^2)^n - 2\Lambda\right] d^4x + S_{matter}\,, \tag{1}$$

where $\kappa = 8\pi G$, and η is a constant quantifying the contribution of the T^2-dependent term. In a flat Friedmann-Lemaitre-Robertson-Walker universe and assuming a perfect fluid we have $T^2 = \rho^2 + 3p^2$ and the generalized Friedmann and

Raychaudhuri equations and the corresponding continuity equation can be written

$$3\left(\frac{\dot{a}}{a}\right)^2 = \Lambda + \kappa\rho + \eta(\rho^2+3p^2)^{n-1}\left[\left(n-\frac{1}{2}\right)(\rho^2+3p^2)+4np\rho\right] \tag{2}$$

$$6\frac{\ddot{a}}{a} = 2\Lambda - \kappa(\rho+3p) - \eta(\rho^2+3p^2)^{n-1}\left[(n+1)(\rho^2+3p^2)+4np\rho\right] \tag{3}$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho+p)\frac{\kappa\rho + n\eta\rho(\rho+3p)(\rho^2+3p^2)^{n-1}}{\kappa\rho + 2n\eta(\rho^2+3p^2)^{n-1}\left[\left(n-\frac{1}{2}\right)(\rho^2+3p^2)+4np\rho\right]}. \tag{4}$$

As usual, the Bianchi identity implies that only two of these equations are independent.

If we consider the low redshift limit of these models, further assuming that the universe is composed of matter and possibly also a cosmological constant, we can simplify the Einstein equations to

$$3\left(\frac{\dot{a}}{a}\right)^2 = \Lambda + \kappa\rho + \left(n-\frac{1}{2}\right)\eta\rho^{2n} \tag{5}$$

$$6\frac{\ddot{a}}{a} = 2\Lambda - \kappa\rho - (n+1)\eta\rho^{2n} \tag{6}$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\rho\frac{\kappa + n\eta\rho^{2n-1}}{\kappa + (2n-1)n\eta\rho^{2n-1}}\,. \tag{7}$$

Broadly speaking, we inspection of the equations leads to the expectation that $n < 1/2$ may be interesting at late times. In general these equations need to be solved numerically. However, there are three particular cases for which analytic solutions can be found (at least approximate, low redshift solutions), corresponding to the values $n = 1$, $n = 1/2$ and $n = 0$ (the latter actually corresponds to the ΛCDM case). These have been studied in the literature, in a general mathematical context,[1,2,10] and also observationally constrained in Ref. 4.

Generically we can treat n as a further model parameter, to be constrained by observations. In order to do this we define a dimensionless cosmological density r, via $\rho = r\rho_0$, where ρ_0 is the present day density, as well as a generic parameter

$$Q = \frac{\eta}{\kappa}\rho_0^{2n-1}\,. \tag{8}$$

With these assumptions, and keeping for the time being the matter assumption, the continuity equation expressed in terms of redshift has the form

$$\frac{dr}{dz} = \frac{3r}{1+z}\times\frac{1+nQr^{2n-1}}{1+(2n-1)nQr^{2n-1}}\,; \tag{9}$$

$$E^2(z) = \frac{H^2(z)}{H_0^2} = \Omega_\Lambda + \Omega_M r + \left(n-\frac{1}{2}\right)Q\Omega_M r^{2n}\,, \tag{10}$$

where, since we must have $E(0) = 1$, the model parameters are related by the requirement that $\Omega_\Lambda = 1-\Omega_M[1+(n-1/2)Q]$, and therefore the Friedmann can be

recast in the two alternative forms

$$E^2(z) = \Omega_\Lambda + \Omega_M r + (1 - \Omega_M - \Omega_\Lambda) r^{2n} \tag{11}$$

$$E^2(z) = 1 + \Omega_M (r - 1) + \left(n - \frac{1}{2}\right) Q \Omega_M (r^{2n} - 1)\,, \tag{12}$$

the first one is generic, while the second applies only if $\Omega_\Lambda \neq 0$. On the other hand, if $\Omega_\Lambda = 0$ we can also use the flatness assumption to eliminate Q in the continuity equation, writing it as

$$\frac{dr}{dz} = \frac{3r}{1+z} \times \frac{(2n-1)\Omega_M + 2n(1-\Omega_M) r^{2n-1}}{(2n-1)[\Omega_M + 2n(1-\Omega_M) r^{2n-1}]}\,. \tag{13}$$

This shows that in a phenomenological sense these models could explain the recent acceleration of the universe without invoking a cosmological constant but relying instead on the nonlinearities of the matter Lagrangian in a matter-only universe with $n = 0$. However, we note that even if these models can lead to accelerating universes without a cosmological constant, this is not per se sufficient to solve the 'old' cosmological constant problem of why it should be zero. For n close to but not equal two zero, two things happen: the (formerly) constant term in the Friedmann equation becomes slowly varying, and the continuity equation implies that the matter density does not behave exactly as $r \propto (1+z)^3$. In what follows we discuss the extent to which deviations from the $n = 0$ case are observationally allowed, and also the most general parameter space where a standard cosmological constant is also allowed.

We can also consider a generalization: instead of considering a universe with a matter fluid, we can assume that this fluid has a constant equation of state $w = p/\rho = const.$ (with the matter case corresponding to $w = 0$). In this case the continuity equation becomes

$$\frac{dr}{dz} = \frac{3r}{1+z}(1+w) \times \frac{1 + nQ f_1(n,w) r^{2n-1}}{1 + 2nQ f_2(n,w) r^{2n-1}}\,, \tag{14}$$

where for convenience we defined

$$f_1(n,w) = (1+3w)(1+3w^2)^{n-1} \tag{15}$$

$$f_2(n,w) = (1+3w^2)^{n-1} \left[\left(n - \frac{1}{2}\right)(1+3w^2) + 4nw\right]\,. \tag{16}$$

In this case the Friedmann equation can be written

$$E^2(z) = \Omega_\Lambda + \Omega_M r + f_2(n,w) Q \Omega_M r^{2n}\,, \tag{17}$$

together with the consistency relation $\Omega_\Lambda = 1 - \Omega_M[1 + f_2 Q]$. It follows that we can also re-write it as

$$E^2(z) = \Omega_\Lambda + \Omega_M r + (1 - \Omega_M - \Omega_\Lambda) r^{2n} \tag{18}$$

$$E^2(z) = 1 + \Omega_M (r - 1) + f_2(n,w) Q \Omega_M (r^{2n} - 1)\,, \tag{19}$$

where again the first is generic—and indeed identical to Eq. (11), although the redshift dependence of r will now be different—while the second holds for $\Omega_\Lambda \neq 0$. Here, if $\Omega_\Lambda = 0$ the continuity equation can also be written in a way that eliminates the parameter Q,

$$\frac{dr}{dz} = \frac{3r}{1+z}(1+w) \times \frac{\Omega_M f_2 + n(1-\Omega_M) f_1 r^{2n-1}}{f_2[\Omega_M + 2n(1-\Omega_M) r^{2n-1}]} \,. \tag{20}$$

Table 1. One sigma posterior likelihoods on the power n, the matter density Ω_M and the constant equation of state w (when applicable) for various flat energy-momentum-powered models containing matter, with or without a cosmological constant. The last column lists the reduced chi-square for each best-fit model.

Model assumptions	Ω_M	n	w	χ^2_ν
$\Omega_\Lambda = 0$, $w = 0$	0.39 ± 0.08	0.04 ± 0.04	N/A	0.64
$\Omega_\Lambda \neq 0$, $w = 0$	$0.29^{+0.05}_{-0.03}$	Unconstrained	N/A	0.64
$\Omega_\Lambda = 0$, $w = const.$	$0.28^{+0.12}_{-0.10}$	$-0.08^{+0.06}_{-0.02}$	$-0.11^{+0.07}_{-0.04}$	0.62

Note: The specific assumptions for each case are described in the main text. The constraints come from the combination of the Pantheon supernova data and Hubble parameter measurements.

3. Constraints on energy-momentum-powered models

In what follows we briefly summarize constraints on the generic energy-momentum-powered models, revising and updating the analysis in Ref. 4. We carry out a standard likelihood analysis, using the datasets already mentioned in the introduction, and separately considering three different theoretical scenarios within this class of models. An overview of the results can be found in Table 1.

The constraints for the $\Omega_\Lambda = 0$ matter case are summarized in the top panel pf Fig. 1 and also in the first row of Table 1. As expected given the form of the Friedmann and continuity equations, there is a clear degeneracy between the two parameters. The best-fit values are about one standard deviation away from the canonical values $n = 0$ and $\Omega_M \sim 0.3$, and a non-zero $n \sim 0.04$ and a slightly higher matter density are preferred. However, at the two sigma level the results are consistent with ΛCDM; one should also bear in mind that the $n = 0$ does correspond to the ΛCDM case.

Constraints on the $\Omega_\Lambda \neq 0$ matter case are shown in the bottom panel of Fig. 1 and also in the second row of Table 1. Here there is a strong degeneracy between Q and n, both of which are unconstrained. On the other hand, the matter density is still well constrained (indeed, the constraint is tighter than in the case without a cosmological constant) and fully consistent with the canonical ΛCDM value.

Finally, constraints on the $\Omega_\Lambda = 0$ case while allowing for a constant equation of state, $w = const$ are in Fig. 2 and also in the final row of Table 1. In this case

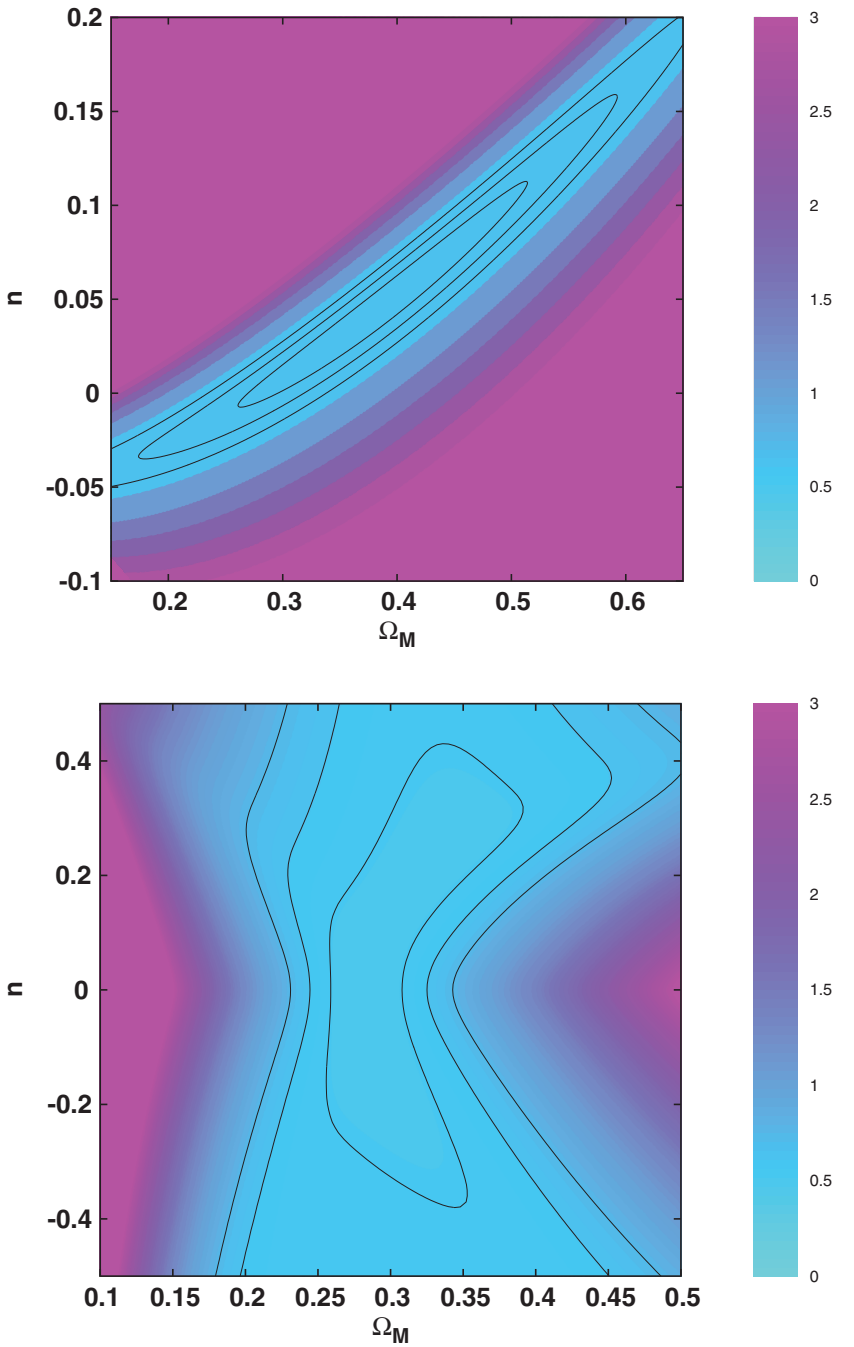


Fig. 1. Constraints on the n–Ω_M parameter space for flat universes with $w = 0$. Top and bottom panels correspond to the $\Omega_\Lambda = 0$ and $\Omega_\Lambda \neq 0$ cases, respectively. The black solid curves show the one, two, and three sigma confidence levels, while the color map depicts the reduced chi-square.

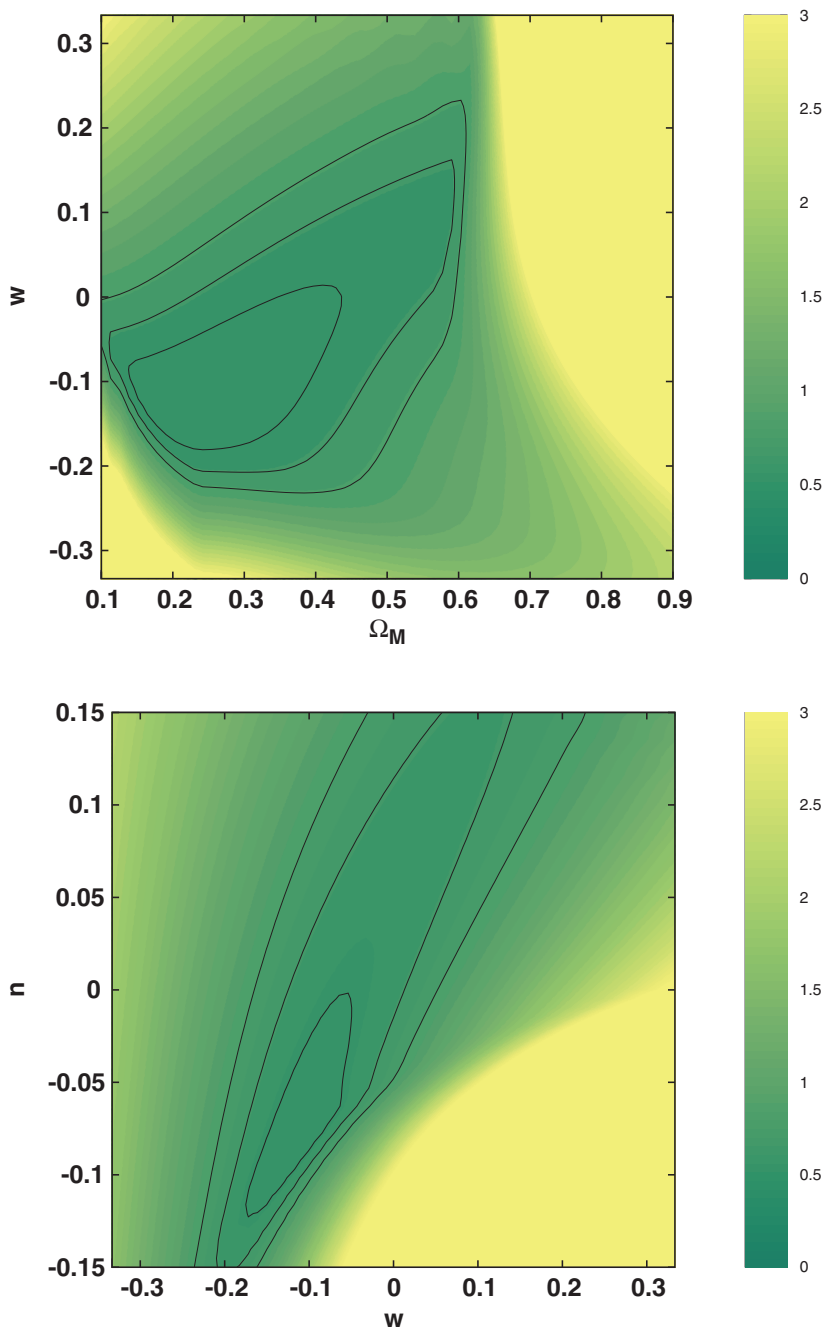


Fig. 2. Constraints on the n–Ω_M–w parameter space for flat universes with $\Omega_\Lambda = 0$. The black solid curves show the one, two, and three sigma confidence levels, while the color maps depict the reduced chi-square.

one can constrain the three model parameters. The preferred value of the matter density is again consistent with the canonical ΛCDM value, while there is a mild preference (less than two standard deviations) for negative values on the exponent n and the equation of state parameter w.

Overall, we note that the best-fit value for the matter density is compatible, within the uncertainties, with the standard one, and there is no significant evidence for deviations from ΛCDM. On the other hand, it is worthy of note that the values of the reduced chi-square for all the best-fit models is significantly below unity, so the models clearly overfit the low redshift data that we are considering.

4. A simple $f(R,T)$ model

We will now explore the modified gravity model recently discussed in Ref. 7. This is actually one case of a larger set of $f(R,T)$ models, to be discussed elsewhere. A class of modified gravity models now dubbed $f(R)$ gravity, R being the scalar curvature, was first considered in Ref. 11, but these models are subject to tight cosmological constraints.[12] A phenomenologically broader (if physically less well motivated) class is that of the so-called $f(R,T)$ models, where T is the trace of the stress energy tensor. A particular subclass of these models has separable function, $f(R,T) = R + f_2(T)$,. These models have been the subject of several mathematical studies but so far they have not been put through a detailed comparison with cosmological observations, with the exception of the recent qualitative analysis of Ref. 8.

Qualitatively, the main difference is that here the new terms depend on the trace of the energy-momentum tensor $T = \rho - 3p$, while in the model considered in the previous sections they depended on $T^2 \equiv T_{\mu\nu}T^{\mu\nu} = \rho^2 + 3p^2$. The procedure for studying the two models should otherwise be similar.

This model,[7] also previously considered in Ref. 8, has the action

$$S = \frac{1}{2\kappa}\int \left[R + \xi\sqrt{T} - 2\Lambda\right] d^4x + S_{matter}\,. \tag{21}$$

In a flat FLRW universe the Friedmann and Raychaudhuri equations now have the following form[a]

$$3\left(\frac{\dot{a}}{a}\right)^2 = \Lambda + \kappa\rho + \xi\frac{(\rho - p)}{\sqrt{\rho - 3p}} \tag{22}$$

$$6\frac{\ddot{a}}{a} = 2\Lambda - \kappa(\rho + 3p) + \frac{\xi}{2}\frac{(\rho - 7p)}{\sqrt{\rho - 3p}}\,. \tag{23}$$

As a simple comparison, in the $p = 0$ case this model leads to a Friedmann equation of the form

$$3H^2 = \Lambda + \kappa\rho + \xi\sqrt{\rho}\,, \tag{24}$$

[a]We note that there is a minus sign typo in the last term of Eq. (10) in Ref. 7.

while in the energy-momentum-powered model, choosing $n = 1/4$, one has

$$3H^2 = \Lambda + \kappa\rho - \frac{\eta}{4}\sqrt{\rho}\,. \tag{25}$$

We note that the two Friedmann equations coincide (if one identifies $\xi = -\eta/4$), but the corresponding continuity equations differ in the two cases.

We will again assume constant equations of state ($p = w\rho$), use the standard definitions of Ω_M and Ω_Λ together with $\rho = r\rho_0$ and additionally define

$$\zeta = \frac{\xi}{2\kappa\sqrt{\rho_0}}\,. \tag{26}$$

We can then rewrite the Friedmann equation as follows

$$E^2(z) = \Omega_\Lambda + \Omega_M r + 2\zeta\frac{(1-w)}{\sqrt{1-3w}}\Omega_M\sqrt{r}\,. \tag{27}$$

In principle there are therefore 3 free parameters, since the $E(0) = 1$ condition requires that the model parameters are related by $\Omega_\Lambda = 1 - \Omega_M[1 + 2\zeta(1 - w)/\sqrt{1-3w}]$. We can also rewrite it as

$$E^2(z) = \Omega_\Lambda + \Omega_M r + (1 - \Omega_M - \Omega_\Lambda)\sqrt{r} \tag{28}$$

$$E^2(z) = 1 + \Omega_M(r-1) + 2\zeta\frac{(1-w)}{\sqrt{1-3w}}\Omega_M(\sqrt{r}-1)\,; \tag{29}$$

the first of these is generic, while the second is only valid if $\Omega_\Lambda \neq 0$. However, note that in general the parameters (ζ, w) still affect the continuity equation, which can be written as

$$\frac{dr}{dz} = \frac{3r}{1+z}(1+w) \times \frac{\sqrt{1-3w} + \zeta/\sqrt{r}}{\sqrt{1-3w} + (1-w)\zeta/\sqrt{r}}\,. \tag{30}$$

We note that the usual behaviour, $r \propto (1+z)^3$, is recovered for $\zeta = 0$ and that (less trivially) this also occurs for the matter case ($w = 0$) for any value of the parameter ζ. As an illustration of the role of this parameter we can solve the continuity in the $\zeta \to 0$ limit. One finds

$$r(z) = \left[\left(1 + \frac{w\zeta}{\sqrt{1-3w}}\right)(1+z)^{3(1+w)/2} - \frac{w\zeta}{\sqrt{1-3w}}\right]^2\,, \tag{31}$$

which again has the appropriate limits.

5. Observational constraints on the $\sqrt{T}$ model

The model can now be constrained, and in particular we can again consider both the scenario with a cosmological constant (in which case the model is an extension of ΛCDM) and the scenario without a cosmological constant (in which case we can check whether such models can account for the recent acceleration of the universe at all). An overview of the results can be found in Table 2.

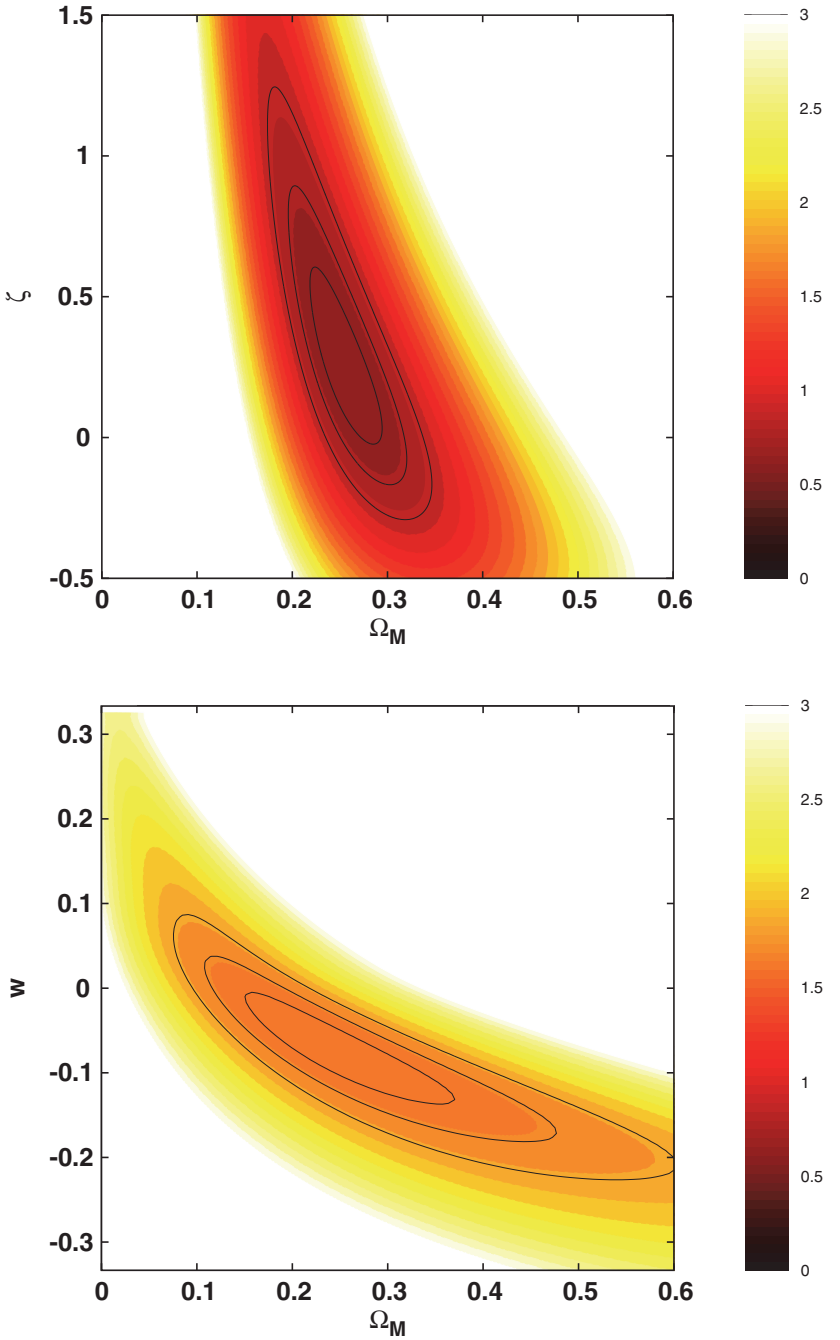


Fig. 3. Constraints on flat $\sqrt{T}$ models. The top panel shows constraints for $\Omega_\Lambda \neq 0$ and $w = 0$, and the bottom panel shows constraints for $\Omega_\Lambda = 0$ and $w = const.$ The black solid curves show the one, two, and three sigma confidence levels, and the color maps depict the reduced chi-square.

Table 2. One sigma posterior likelihoods on the matter density Ω_M, the coupling ζ and the constant equation of state w (when applicable) for various flat energy-momentum-powered models containing matter, with or without a cosmological constant. The last column lists the reduced chi-square for each best-fit model.

Model assumptions	Ω_M	ζ	w	χ^2_ν
$\Omega_\Lambda = 0$, $w = 0$	0.15 ± 0.02	2.78 ± 1.72	N/A	1.80
$\Omega_\Lambda \neq 0$, $w = 0$	$0.25^{+0.03}_{-0.02}$	$0.23^{+0.22}_{-0.18}$	N/A	0.63
$\Omega_\Lambda = 0$, $w = const.$	$0.24^{+0.08}_{-0.07}$	Unconstrained	$-0.08^{+0.04}_{-0.05}$	1.80

Note: The specific assumptions for each case are described in the main text. The constraints come from the combination of the Pantheon supernova data and Hubble parameter measurements.

Starting with the case of $\Omega_\Lambda = 0$ and $w = 0$, we effectively have only one independent parameter, since the matter density and coupling parameter are related through $(1 + 2\zeta)\Omega_M = 1$. The obtained constraints are listed in the first row of Table 2. It is clear from the very large value of the reduced chi-square that this model, containing only matter but no cosmological constant, does not fit the data. In other words, this model can not be a true alternative to ΛCDM.

The top panel of Fig. 3 and the second row of Table 2 summarize the constraints for the case of $\Omega_\Lambda \neq 0$ and $w = 0$. In this case we have two independent parameters, and the model is effectively a one parameter extension of ΛCDM. Here, as in the previously discussed case of energy-momentum-powered models, the model slightly overfits the data, and there is no statistically significant preference for a non-zero coupling parameter ζ, The best fit value of the matter density is not significantly changed with respect to its value, for the same datasets, in the ΛCDM model.[13]

Finally, the bottom panel of Fig. 3 and the third row of Table 2 show the constraints for $\Omega_\Lambda = 0$ and $w \neq 0$, where we have three independent parameters. There are now additional degeneracies between the parameters, but the main conclusion remains the same as for the $w = 0$ case: without a cosmological constant this model severely underfits the data, and therefore it is not viable as an alternative to ΛCDM.

6. Outlook

We have discussed observational current low redshift background constraints on classes of FLRW cosmological models in which the matter side of Einstein's equations includes, in addition to the canonical term, either a term proportional to a function of the energy-momentum tensor ($T^2 = \rho^2 + 3p^2$), or of its trace ($T = \rho - 3p$). Both of these can be phenomenologically thought of as extensions of general relativity with a nonlinear matter Lagrangian.

We considered both models under two different scenarios: (1) as phenomenological two-parameter or three-parameter extensions of the standard ΛCDM, in which case the model still has a cosmological constant but the nonlinear matter Lagrangian

leads to additional terms in Einstein's equations, which cosmological observations tightly constrain, and (2) as alternatives to ΛCDM, where there is no cosmological constant, and the nonlinear matter term would have to provide the acceleration (which would be somewhat closer in spirit to the usual modified gravity models).

Overall, our analysis suggests that the ΛCDM paradigm is fairly robust or, pragmatically, it is a good phenomenological approximation to a still unknown more fundamental model. In other words, if there is no true cosmological constant, the alternative mechanism must effectively be like one, at least at low redshifts. On the other hand, for parametric extensions of ΛCDM, subdominant (ca. 10% level) contributions are allowed by the low redshift background cosmology data that we have considered. These constraints can of course be tightened by including additional datasets, such as that from cosmic microwave background observations. Our work can be extended to broader classes of $f(R,T)$ models, a discussion of which is left for a subsequent publication.

Acknowledgments

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017. The project that led to this work was started during AstroCamp 2020.

References

1. M. Roshan and F. Shojai, Energy-Momentum Squared Gravity, *Phys. Rev.* **D94**, p. 044002 (2016).
2. C. V. R. Board and J. D. Barrow, Cosmological Models in Energy-Momentum-Squared Gravity, *Phys. Rev.* **D96**, p. 123517 (2017), [Erratum: Phys. Rev. D 98, no. 12, 129902 (2018)].
3. O. Akarsu, N. Katirci and S. Kumar, Cosmic acceleration in a dust only universe via energy-momentum powered gravity, *Phys. Rev.* **D97**, p. 024011 (2018).
4. M. C. F. Faria, C. J. A. P. Martins, F. Chiti and B. S. A. Silva, Low redshift constraints on energy-momentum-powered gravity models, *Astron. Astrophys.* **625**, p. A127 (2019).
5. A. G. Riess *et al.*, Type Ia Supernova Distances at Redshift > 1.5 from the Hubble Space Telescope Multi-cycle Treasury Programs: The Early Expansion Rate, *Astrophys. J.* **853**, p. 126 (2018).
6. O. Farooq, F. R. Madiyar, S. Crandall and B. Ratra, Hubble Parameter Measurement Constraints on the Redshift of the Deceleration–acceleration Transition, Dynamical Dark Energy, and Space Curvature, *Astrophys. J.* **835**, p. 26 (2017).
7. N. Godani and G. C. Samanta, Estimation of cosmological parameters, stability analysis and energy conditions in viable modified gravity, *Chin. J. Phys.* **66**, 787 (2020).
8. H. Velten and T. R. P. Caramês, Cosmological inviability of $f(R,T)$ gravity, *Phys. Rev. D* **95**, p. 123536 (2017).

9. F. K. Anagnostopoulos and S. Basilakos, Constraining the dark energy models with $H(z)$ data: An approach independent of H_0, *Phys. Rev.* **D97**, p. 063503 (2018).
10. N. Katırcı and M. Kavuk, $f(R, T_{\mu\nu}T^{\mu\nu})$ gravity and Cardassian-like expansion as one of its consequences, *Eur. Phys. J. Plus* **129**, p. 163 (2014).
11. H. A. Buchdahl, Non-linear Lagrangians and cosmological theory, *Mon. Not. Roy. Astron. Soc.* **150**, p. 1 (1970).
12. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, Modified Gravity and Cosmology, *Phys. Rept.* **513**, 1 (2012).
13. C. B. D. Fernandes, C. J. A. P. Martins and B. A. R. Rocha, Constraining alternatives to a cosmological constant: Generalized couplings and scale invariance, *Phys. Dark Univ.* **31**, p. 100761 (2021).

Constraining the dark energy-dark matter interaction model using low-redshift observations

Archana Sangwan*, Joseph P J†, and S. Shankaranarayanan‡
Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India
* *E-mail: arch06san@gmail.com,* † *E-mail: josephpj@iitb.ac.in*
‡ *E-mail: shanki@phy.iitb.ac.in*

Various observations have shown that dark energy accounts for nearly two-thirds of the energy density of the Universe. The simplest model to explain the nature of dark energy is the cosmological constant (ΛCDM) model. Although Planck observations supports using ΛCDM model as the base cosmological model, there exist some inconsistencies in parameter estimates when compared with independent observations. The most important is the inconsistency in the H_0 estimates from the Planck collaboration which reports $H_0 = 67.5^{+0.5}_{-0.5}\ km\ s^{-1}\ Mpc^{-1}$, a considerably lower value when compared with the direct local distance ladder measurements. This value shows a discrepancy at the level greater than 4σ with the constraints reported by SH0ES collaboration in 2019, $H_0 = 74.3^{+1.42}_{-1.42}$ $km\ s^{-1}\ Mpc^{-1}$. These disagreements, called the Hubble tension, point towards a new physics that deviates from the standard ΛCDM model and to resolve this various methods have been proposed. In this work, a quintessence scalar field with an inverse power potential ($V(\phi) \sim \phi^{-n}$) is assumed as a description of dark energy and we focus on an interacting dark energy dark matter model where the interacting term is taken to be linear in the field (Φ). We study in detail the evolution of the model and provide constraints on the model parameters using low redshift cosmological observations of Type Ia Supernovae (SN), baryon acoustic oscillations (BAO), direct measurements of Hubble parameter (Hz) and high redshift HII galaxy measurements (HIIG). We find that the model agrees with the existing values of the nonrelativistic matter density parameter, Ω_m and dark energy equation of state parameter, w_0. The analysis shows that the observations prefer a negative value of coupling constant and gives the best fit value of $H_0 = 69.9^{+0.46}_{-1.02}\ km\ s^{-1}\ Mpc^{-1}$ and thereby can be used to alleviates the H_0 tension between Planck measurements and the observations considered.

Keywords: Cosmological parameter, observations, dark energy - dark matter interaction, based on the work [arXiv:2102.12367].[1]

1. Introduction

Many cosmological observations suggest that, at the present epoch, the energy content of the universe is dominated by dark energy, and to explain the observed accelerated expansion of the universe, this dark energy component must have a negative pressure.[2–5] However, the nature of dark energy is still a mystery, and many models are proposed in the literature to explain what dark energy might be. The simplest model is the ΛCDM model, which is also the most consistent model for cosmological observations. The latest results from the Planck mission provide good support to the ΛCDM model as the base cosmological model and also improve upon the uncertainties in parameter estimation.[6,7] With more accurate and precise measurements, the

uncertainties from various experiments have been studied and are in disagreement with each other. The most glaring tension is found in the estimation of the Hubble parameter reported by the Planck collaboration and the low redshift cosmological probes.[8–10] In some cases, the tension stands at level greater than 4σ, e.g., Planck 2018 results, which assume standard ΛCDM model as the base cosmological model, report Hubble parameter to be $H_0 = 67.4 \pm 0.5\ km\ s^{-1}\ Mpc^{-1}$,[7] whereas SH0ES collaboration estimates $H_0 = 73.5 \pm 1.4\ km\ s^{-1}\ Mpc^{-1}$.[11] Another discrepancy, although not as severe, is found in the estimation of σ_8 reported by Planck 2018 and low redshift observation with stands at a level of 2σ-3σ level.[12–17] The high precision of the data from Planck collaboration and low redshift probes insinuate that the H_0 tension between these experiments cannot result from the systematics and hence, points towards a new physics that deviates from the standard ΛCDM model.

Many models have been proposed and studied to resolve these disagreements in literature. One such proposal is to consider a non-zero interaction between dark energy and dark matter which could alleviate these disagreements.[18–21] In this work, we focus on an interacting dark sector model, which is introduced in.[22] In most cases, this interaction term is introduced in fluid equations by hand, which might not be consistent with the classical field theory action. In the Ref.,[22] instead of putting it by hand, the authors derive the interaction term, which is consistent with both fluid descriptions and classical field theory description. We use this interaction term in our work and study whether the interaction model is consistent with low redshift observations such as direct measurement of Hubble parameter (Hz),[23–28] measurements of HII galaxy (HIIG),[29,30] baryon acoustic oscillations (BAO),[31–36] and type Ia supernovae (SN).[37] We constrain the model parameters, and after analyzing, find that using only the background observations, we cannot distinguish between interacting and noninteracting scenarios. Two see a clear distinction between the two cases; we have to study the evolution of perturbed quantities.

2. Dark Energy Dark Matter Interaction Model

The interacting model is described by the action

$$S = \int d^4x\sqrt{-g}\left(\frac{1}{2\kappa^2}R - \frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - U(\phi) - \frac{1}{2}e^{2\alpha(\phi)}g^{\mu\nu}\nabla_\mu\chi\nabla_\nu\chi - e^{4\alpha(\phi)}V(\chi)\right). \tag{1}$$

where dark energy is denoted by the scalar field ϕ, $U(\phi)$ is the scalar field potential, dark matter is represented by the scalar fieldχ, and $\alpha(\phi)$ denotes the dark energy dark matter interaction (For more details, see[22]).

The Einstein equation in terms of the scalar field (ϕ) and dark matter fluid, specified by four-velocity energy density and pressure, is given by[22]

$$G_{\mu\nu} = 16\pi G\left[\nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}\nabla^\sigma\phi\nabla_\sigma\phi - g_{\mu\nu}V(\phi) + p_m g_{\mu\nu} + (\rho_m + p_m)u_\mu u_\nu\right], \tag{2}$$

and the energy-momentum tensor for the dark matter fluid is given by

$$T_{\nu}^{(m)\mu} = p_m g_{\mu\nu} + (\rho_m + p_m) u_\mu u_\nu \,. \tag{3}$$

The interaction term is described by $\nabla_\mu T_{\nu}^{(m)\mu} = Q_\nu^{(\mathrm{F})}$, and by demanding that the interaction term Q_ν must have a field theory description, they obtained a unique form of interaction term, given by[22]

$$Q_\nu^{(\mathrm{F})} = T^{(m)} \nabla_\nu \alpha(\phi) \,. \tag{4}$$

We study the evolution of cosmological equations in the presence of the interaction and obtain constraints on the model parameters by assuming a specific form of scalar field potential and the interaction term. In this work, we assume a quintessence scalar field described by the inverse power law potential and an interaction term that is linear in the field (ϕ),

$$U(\phi) \sim \frac{1}{\phi^n}, \qquad n \;=\; 1, 2, \tag{5}$$

$$\alpha(\phi) \sim \phi, \qquad \alpha(\phi) = C\phi \tag{6}$$

where C is the interaction strength. The evolution of the scalar field is described by the Klein-Gordon equation given by

$$(\ddot{\phi} + 3H\dot{\phi} + U_{,\phi})\dot{\phi} = Q \,,$$

where $Q = Q_0^{(\mathrm{F})}$. The scalar field potential in terms of dimensionless parameter $\tilde{\phi}$ is

$$U(\tilde{\phi}) = \frac{\kappa}{2} m_{\mathrm{Pl}}^2 \tilde{\phi}^{-n} \tag{7}$$

where κ is a constant and m_{Pl}, is the Planck mass. The evolution of non-relativistic matter density in the presence of interaction is given by

$$\dot{\rho}_m + 3H\rho_m = -Q \,.$$

For interaction term, $Q = -\alpha_{,\phi}\dot{\phi}\rho_m$, on solving the above equation we get, $\rho_m = \rho_{m_0} e^{\alpha(\phi)-\alpha(\phi_0)} a^{-3}$.

We use these equations to study the universe's evolution in the presence of nonzero interaction and establish the consistency of the model with the latest low redshift observations. The observations used in the analysis are: a) the direct measurements of Hubble parameter (Hz) with observations in the redshift range of $0.070 \leq z \leq 1.965$, b) the baryon acoustic oscillations (BAO) data ($0.106 \leq z \leq 2.36$), c) High-redshift HII Galaxy measurements (HIIG) which spans in a range $0.0088 \leq z \leq 2.42935$, and d) type Ia Supernovae (SN) observations with data points between range $0.05 \leq z \leq 1.4$. To obtain the constraints on model parameters, we use the chi-squared minimization technique, which tells about the goodness of the fit by comparing the observed values of an observable with the theoretically expected values for a specific model (dark energy dark matter interaction model in our case).

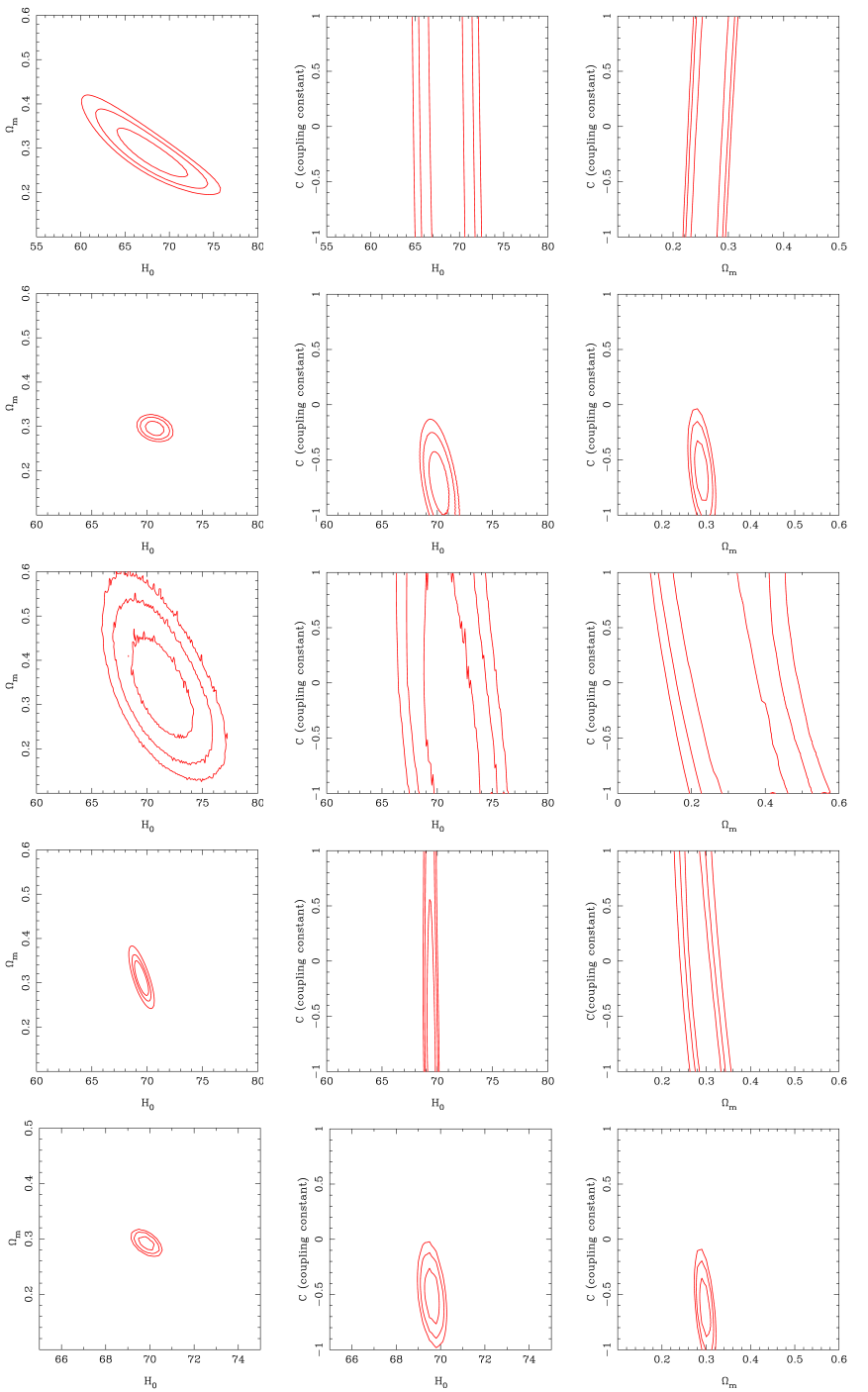


Fig. 1. 1,2,3-σ confidence regions for Hz data (I row), BAO+Hz data (II row), HIIG data (III row), SN+Hz data (IV row) and all four data sets (V row).

3. Observational Constraints

In the analysis, the parameter describing the model are: the Hubble parameter at present (H_0), nonrelativistic matter density parameter (Ω_m), dark energy equation of state parameter at present (w_0), and the interaction strength (C). The priors used in the analysis are given in table below.

Parameter	Lower Limit	Upper Limit
H_0	60.0	80.0
Ω_m	0.1	0.6
w_0	-1.0	1.0
C	-1.0	1.0

We perform the chi-squared minimization technique to obtain the $1, 2, 3 - \sigma$ confidence regions for the four observations we considered in the analysis and provide the results in two-dimensional planes in Fig. 1. We assume $n = 1$ in the analysis and have marginalized over the other parameters to obtain the results in the two-dimensional plane. The results from Hz observation are shown in the first row. The allowed range from Hz data for Ω_m and H_0 parameters are fairly large as compared to the allowed uncertainties obtained from BAO measurements (second row) and SN data (fourth row). On the other hand, the interaction strength parameter C is not constrained by Hz observations, and the entire range is allowed by the observations. BAO+Hz data (second row) provides the tightest constraints on the cosmological parameters, and these observations also obtain the upper limit on interaction strength.

Constraints from HIIG data are presented in the third row (where we have not included the systematic error in the analysis for HIIG data); the HIIG data is unable to constrain the H_0 parameter and provides the largest allowed range for Ω_m among the data sets considered in the analysis. In the fourth row, results from SN+Hz are presented. SN+Hz data also provides narrow constraints on H_0 and Ω_m parameters and gives an upper limit on the interaction strength. Finally, the results from combining the data sets (Hz+BAO+HIIG+SN) are given in row fifth. We find that by performing combined analysis, we get a very narrow range of uncertainties in parameter estimation, and the best fit values of parameters are given by $H_0 = 69.9\ km\ s^{-1}\ Mpc^{-1}$, $\Omega_m = 0.29$, and $C = -0.47$. From the analysis, we find that all the observations are consistent with negative values of C, and the upper on C is given by SN+Hz and BAO+Hz observations which drive the combined results.

We show the results in $\Omega_m - w_0$ plane obtained from Hz (first row, left), BAO+Hz (first row, right), HIIG (the second row, left), and SN+Hz (second row right) observations in Fig. 2. For all the observations, we find that the model is consistent with the cosmological constant model, and BAO+Hz data provides the narrowest

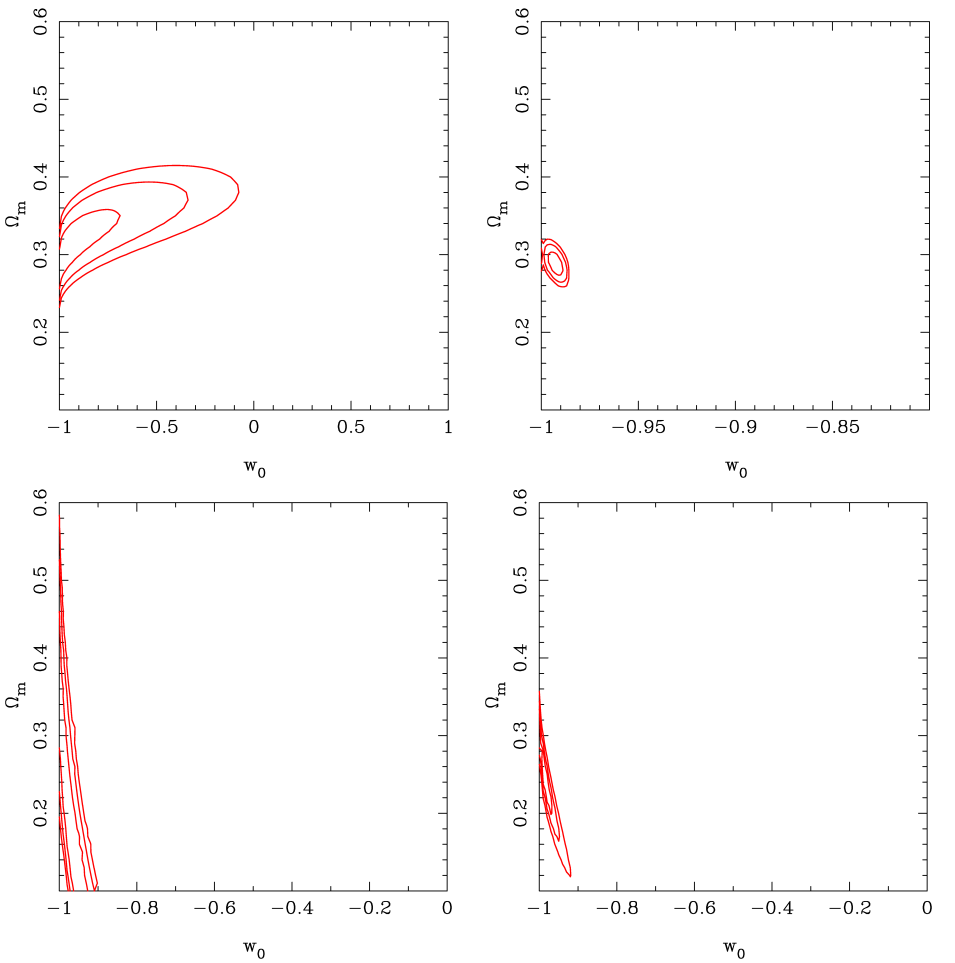


Fig. 2. 1,2,3-σ confidence regions in 'w_0-Ω_m' plane. The first row shows constraints from Hz data (left) and BAO+Hz observations (right). The second row shows constraints from HIIG measurements (left) and SN+Hz observations (right).

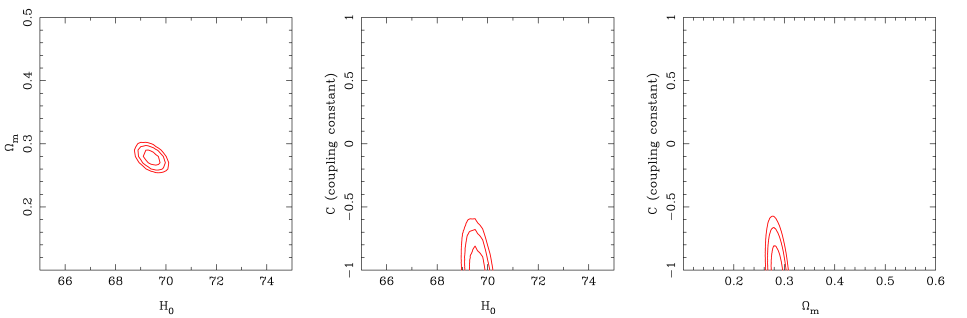


Fig. 3. 1,2,3-σ confidence regions for combined analysis of all four data sets.

constraints on Ω_m and w_0 parameters followed by Sn+Hz measurements. However, Hz data also allows for non-accelerating universe scenarios.

We have also performed the chi-squared analysis for the $n = 2$ case. And we find that the key conclusions from $n = 1$ remain the same that all the observations constrain H_0 to be close to 70 $km\ s^{-1}\ Mpc^{-1}$ and the narrowest constraints are obtained from BAO+Hz and SN+Hz, followed by Hz and HIIG observation. The change in results from individual observations is not significant as we move from $n = 1$ to $n = 2$ case. The only difference worth mentioning is that as n changes from 1 to 2, the contours show a slight shift value of C towards -1, and the model seems to prefer negative values of interaction strength. This preference is more notable in the results from the combined analysis of all the previously mentioned observations. In Fig. 3, we show the 1,2,3-σ confidence regions for combined analysis from all the four data sets considered in the analysis. The effect on Ω_m and H_0 constraints is not significant as we go from $n = 1$ to $n = 2$ case, and the constraints obtained from each dataset are consistent with each other.

Examining the constraints obtained through the evolution of background cosmological equations and using low redshift observations, it is evident that we cannot distinguish between interacting and noninteracting scenarios. Therefore, in the next section, we study the evolution of first-order perturbations in the presence of a nonzero interacting term.

4. Evolution of Perturbations

In this section, we study the evolution of quantities that depends upon the evolution of first-order cosmological perturbations and are related to the observations in current cosmological experiments. The cosmological quantities we discuss are related to structure formation, gravitational lensing and integrated Sachs-Wolfe effect, denoted by δ_m, $\Phi + \Psi$, and $\Phi' + \Psi'$ respectively.

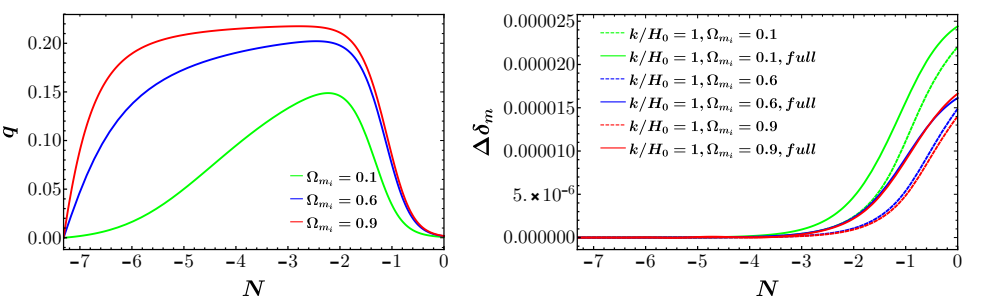


Fig. 4. Evolution of q (left panel) and $\Delta\delta_m$ (right panel) as a function of N with $C = -0.6$.

Structure formation: To study the effect of interaction between dark energy and dark matter on structure formation, we study the evolution of matter density per-

turbation denoted by δ_m, for various length scales. The parameter δ_m is defined as $\delta_m(t,x,y,z) \equiv \frac{\delta\rho_m(t,x,y,z)}{\overline{\rho_m}(t)}$, where $\overline{\rho_m}(t)$ is the mean matter density at time t.

In Fig. 4, the left plot shows the evolution of the scaled interaction term given by $q = \frac{Q}{H^3 M_{Pl}^2}$ as a function of N for different initial values of matter density parameter Ω_m, where $N = ln(a)$ is the number of e-foldings. We see that for a larger value of initial matter density, the contribution of the interaction term is greater and becomes significant at an earlier time than for smaller initial matter density values. The right plot shows the evolution of $\Delta\delta_m = \delta_{m_i} - \delta_{m_{ni}}$ as a function of N. The plot shows the difference in the evolution of matter density perturbations in the interacting and noninteracting models denoted by parameters δ_{m_i} and $\delta_{m_{ni}}$ respectively. The dotted lines are for the case when the perturbed interaction term is neglected, and the solid line is when we consider the perturbed interaction term. The different colors correspond to the different initial values of the matter density parameter. We see that at around $N \sim 2$, the curves deviate from one another, showing that considering the interaction term changes the evolution of matter perturbations, and the maximum change is observed in smaller Ω_{m_i} case. This demonstrates that its possible to detect the signature of dark energy and dark matter interactions in future large scale structure observations.[38–40]

Weak gravitational lensing: The quantity $\Phi + \Phi$ contributes to the weak gravitational lensing effect that is the result of deviation in photon trajectory because of the mass distribution in space. This section discusses the evolution of Φ as in standard cosmology, $\Phi = \Psi$. The evolution of Φ as a function of N is shown in Fig. 5. The plot on the left is for dark energy dark matter interaction scenario with interaction strength $C = -0.6$, and the right plot is for noninteracting scenarios. The different curves correspond to different length scales represented by different values of k. The curves are obtained by evolving the system of equations with the same initial conditions starting at a redshift of $z_i = 1500$. We find that the effect of interaction is clear on the smallest scale considered in the analysis, i.e., at $k/H_0 = 1$. Therefore, we can use the evolution of scalar perturbations to distinguish between the interacting and noninteracting dark energy dark matter scenarios by the future observations of weak gravitational lensing.[41]

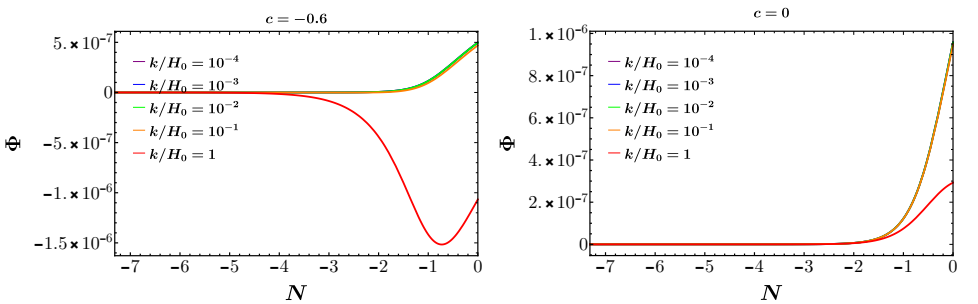


Fig. 5. Evolution of Φ as a function of N. Left: $C = -0.6$, Right: $C = 0$.

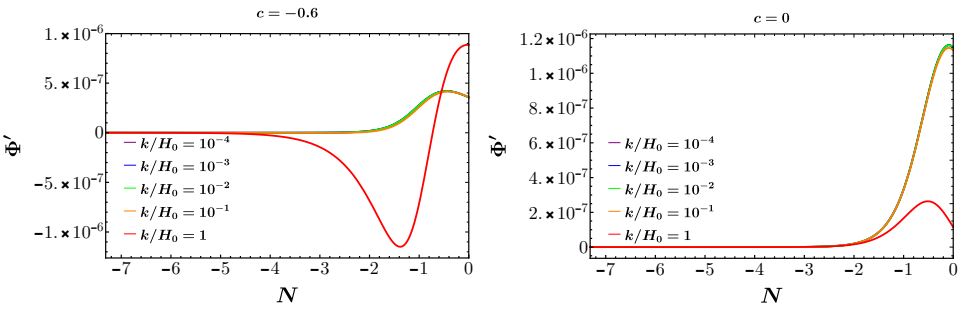


Fig. 6. Evolution of Φ' as a function of N. Left: $C = -0.6$, Right: $C = 0$.

Integrated Sachs-Wolfe effect: This effect manifests in cosmic microwave background radiation anisotropies because of the change in gravitational potential in the path of photons as the photons travel after decoupling from the last scattering surface towards the observer. This effect is given by $\Phi' + \Psi'$, where $'$ denotes the rate of change with respect to conformal time. The change in the evolution of Φ' as a function of N for interacting case with interaction strength $C = -0.6$ (left plot) and noninteracting (right plot) scenarios is given Fig. 6. The different curves correspond to different length scales. The change in the evolution of Φ' because of interaction is observed at $N \sim 3$ where the Φ' is suppressed as compared to the noninteracting scenario. And the maximum change is observed on the smaller scales, and hence, this effect could be observed in the CMB anisotropies. Therefore, using the ISW observations, we can distinguish between interacting and noninteracting scenarios.

5. Conclusion

In this work, we consider a dark energy dark matter interaction model and check the consistency of the model with low redshift observations. In this model, the interaction term is obtained by assuming a one-to-one mapping between the classical field description and the fluid description of the interacting model. We perform the chi-squared minimization technique and obtain the constraints on model parameters. The observations we use in the analysis are direct measurements of Hubble parameter (Hz), baryon acoustic oscillation (BAO) data, measurements of HII galaxies (HIIG), and type Ia supernovae data (SN). We obtain narrow constraints on Hubble parameter H_0, non-relativistic matter density parameter Ω_m, and dark energy equation of state parameter at present w_0. The BAO+Hz and SN+Hz observations provide an upper limit on interaction parameter C. The constraints from these observations are consistent with one another, and the best fit value from all the observations are close to $H_0 \sim 70\ km\ s^{-1}\ Mpc^{-1}$. So there is no tension in the estimated value of H_0 from these observations for the interaction model considered.

The Interaction model is consistent with the abovementioned low redshift observations used in the analysis.

From background analysis, we also observed that it is difficult to distinguish between the interacting and noninteracting models, and therefore we must look at the evolution of perturbations. Therefore, we studied the evolution of first order perturbation quantities such as δ_m, Φ, and Φ' in the interacting and noninteracting case. We find that the evolution of these quantities is modified in the presence of dark energy dark matter interaction terms. Therefore, using the observations of structure formation, weak gravitational lensing, and integrated Sach-Wolfe effect, which depends upon the evolution perturbed quantities mentioned above and carries the signature of the interaction, we can distinguish between the interacting and noninteracting scenarios.

Acknowledgments

The work is partially supported by the ISRO-Respond grant. J.P.J. is supported by CSIR Senior Research Fellowship, India.

References

1. J. P. Johnson, A. Sangwan and S. Shankaranarayanan, Cosmological perturbations in the interacting dark sector: Observational constraints and predictions (2 2021).
2. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
3. S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
4. D. N. Spergel *et al.*, Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology, *Astrophys. J. Suppl.* **170**, p. 377 (2007).
5. D. M. Scolnic *et al.*, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *Astrophys. J.* **859**, p. 101 (2018).
6. N. Aghanim *et al.*, Planck 2018 results. I. Overview and the cosmological legacy of Planck, *Astron. Astrophys.* **641**, p. A1 (2020).
7. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)].
8. V. Marra, L. Amendola, I. Sawicki and W. Valkenburg, Cosmic variance and the measurement of the local Hubble parameter, *Phys. Rev. Lett.* **110**, p. 241305 (2013).
9. L. Verde, P. Protopapas and R. Jimenez, Planck and the local Universe: Quantifying the tension, *Phys. Dark Univ.* **2**, 166 (2013).
10. K. C. Wong *et al.*, H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, *Mon. Not. Roy. Astron. Soc.* **498**, 1420 (2020).
11. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *Astrophys. J.* **876**, p. 85 (2019).

12. P. A. R. Ade *et al.*, Planck 2015 results. XV. Gravitational lensing, *Astron. Astrophys.* **594**, p. A15 (2016).
13. P. A. R. Ade *et al.*, Planck 2015 results. XXVII. The Second Planck Catalogue of Sunyaev-Zeldovich Sources, *Astron. Astrophys.* **594**, p. A27 (2016).
14. K. Kuijken *et al.*, Gravitational Lensing Analysis of the Kilo Degree Survey, *Mon. Not. Roy. Astron. Soc.* **454**, 3500 (2015).
15. S. Joudaki *et al.*, KiDS-450: Testing extensions to the standard cosmological model, *Mon. Not. Roy. Astron. Soc.* **471**, 1259 (2017).
16. S. Joudaki *et al.*, KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering, *Mon. Not. Roy. Astron. Soc.* **474**, 4894 (2018).
17. T. M. C. Abbott *et al.*, Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing, *Phys. Rev. D* **98**, p. 043526 (2018).
18. E. Di Valentino, A. Melchiorri and O. Mena, Can interacting dark energy solve the H_0 tension?, *Phys. Rev. D* **96**, p. 043503 (2017).
19. W. Yang, A. Mukherjee, E. Di Valentino and S. Pan, Interacting dark energy with time varying equation of state and the H_0 tension, *Phys. Rev. D* **98**, p. 123527 (2018).
20. E. Di Valentino, A. Melchiorri, O. Mena and S. Vagnozzi, Nonminimal dark sector physics and cosmological tensions, *Phys. Rev. D* **101**, p. 063502 (2020).
21. A. Gómez-Valent, V. Pettorino and L. Amendola, Update on coupled dark energy and the H_0 tension, *Phys. Rev. D* **101**, p. 123513 (2020).
22. J. P. Johnson and S. Shankaranarayanan, Cosmological perturbations in the interacting dark sector: Mapping fields and fluids, *Phys. Rev. D* **103**, p. 023510 (2021).
23. J. Simon, L. Verde and R. Jimenez, Constraints on the redshift dependence of the dark energy potential, *Phys. Rev. D* **71**, p. 123001 (2005).
24. D. Stern, R. Jimenez, L. Verde, S. A. Stanford and M. Kamionkowski, Cosmic Chronometers: Constraining the Equation of State of Dark Energy. II. A Spectroscopic Catalog of Red Galaxies in Galaxy Clusters, *he Astrophysical Journal Supplement* **188**, 280 (May 2010).
25. A. L. Ratsimbazafy, S. I. Loubser, S. M. Crawford, C. M. Cress, B. A. Bassett, R. C. Nichol and P. Väisänen, *Mon. Not. Roy. Astron. Soc.* **467**, 3239 (2017).
26. C. Zhang, H. Zhang, S. Yuan, S. Liu, T.-J. Zhang and Y.-C. Sun, Four new observational H(z) data from luminous red galaxies in the Sloan Digital Sky Survey data release seven, *Research in Astronomy and Astrophysics* **14**, 1221 (October 2014).
27. M. Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at z $\sim$ 2, *Mon. Not. Roy. Astron. Soc.* **450**, L16 (2015).
28. M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro and D. Wilkinson, A 6% measurement of the Hubble parameter at $z \sim 0.45$: direct evidence of the epoch of cosmic re-acceleration, *JCAP* **05**, p. 014 (2016).
29. A. L. González-Morán, R. Chávez, R. Terlevich, E. Terlevich, F. Bresolin, D. Fernández-Arenas, M. Plionis, S. Basilakos, J. Melnick and E. Telles, Independent cosmological constraints from high-z H ii galaxies, *Mon. Not. Roy. Astron. Soc.* **487**, 4669 (2019).
30. R. Terlevich, E. Terlevich, J. Melnick, R. Chávez, M. Plionis, F. Bresolin and S. Basilakos, On the road to precision cosmology with high-redshift H II galaxies, *Mon. Not. Roy. Astron. Soc.* **451**, 3001 (2015).
31. A. Font-Ribera *et al.*, Quasar-Lyman α Forest Cross-Correlation from BOSS DR11 : Baryon Acoustic Oscillations, *JCAP* **05**, p. 027 (2014).

32. J. E. Bautista *et al.*, Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with SDSS DR12 Lyα-Forests, *Astron. Astrophys.* **603**, p. A12 (2017).
33. M. Ata *et al.*, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, *Mon. Not. Roy. Astron. Soc.* **473**, 4773 (2018).
34. A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden and M. Manera, The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$, *Mon. Not. Roy. Astron. Soc.* **449**, 835 (2015).
35. F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders and F. Watson, The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant, *Mon. Not. Roy. Astron. Soc.* **416**, 3017 (October 2011).
36. S. Alam *et al.*, The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample, *Mon. Not. Roy. Astron. Soc.* **470**, 2617 (2017).
37. M. Betoule *et al.*, Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples, *Astron. Astrophys.* **568**, p. A22 (2014).
38. L. Amendola *et al.*, Cosmology and fundamental physics with the Euclid satellite, *Living Rev. Rel.* **21**, p. 2 (2018).
39. R. Maartens, F. B. Abdalla, M. Jarvis and M. G. Santos, Overview of Cosmology with the SKA, *PoS* **AASKA14**, p. 016 (2015).
40. M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A. J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R. J. Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J. M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J. P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H. J. Stiepel, K. Stuurwold, M. Tagger,

Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A. J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus and J. van Zwieten, LOFAR: The LOw-Frequency ARray, *Astronomy & Astrophysics* **556**, p. A2 (August 2013).

41. H. Hildebrandt *et al.*, KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing, *Mon. Not. Roy. Astron. Soc.* **465**, p. 1454 (2017).

On the evolution of inhomogeneous perturbations in the ΛCDM model and $f(R)$ modified gravity theories

T. Schiavone

Department of Physics "E. Fermi", University of Pisa, Polo Fibonacci, Largo B. Pontecorvo 3, I-56127, Pisa, Italy
and
INFN, Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Polo Fibonacci, Largo B. Pontecorvo 3, I-56127, Pisa, Italy
E-mail: tiziano.schiavone@pi.infn.it

G. Montani

ENEA, Fusion and Nuclear Safety Department, C.R. Frascati, Via E. Fermi 45, I-00044 Frascati (Rome), Italy
and
Physics Department, "Sapienza" University of Rome, P.le Aldo Moro 5, I-00185 Rome, Italy

We focus on weak inhomogeneous models of the Universe at low redshifts, described by the Lemaître-Tolman-Bondi (LTB) metric. The principal aim of this work is to compare the evolution of inhomogeneous perturbations in the ΛCDM cosmological model and $f(R)$ modified gravity theories, considering a flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric for the background. More specifically, we adopt the equivalent scalar-tensor formalism in the Jordan frame, in which the extra degree of freedom of the $f(R)$ function is converted into a non-minimally coupled scalar field. We investigate the evolution of local inhomogeneities in time and space separately, following a linear perturbation approach. Then, we obtain spherically symmetric solutions in both cosmological models. Our results allow us to distinguish between the presence of a cosmological constant and modified gravity scenarios, since a peculiar Yukawa-like solution for radial perturbations occurs in the Jordan frame. Furthermore, the radial profile of perturbations does not depend on a particular choice of the $f(R)$ function, hence our results are valid for any $f(R)$ model.

Keywords: Cosmology, Modified gravity, Inhomogeneous Universe, Late Universe, Large Scale Structure

Introduction

The well-known ΛCDM model,[1,2] which includes a cosmological constant Λ and a cold dark matter (CDM) component, is based on General Relativity (GR), and it provides a robust and accurate description of our Universe, supported by cosmological data. Moreover, the Universe appears homogeneous and isotropic at very large scales[3] (cosmological principle), and this kind of geometry is described by the FLRW line element. In this work, based on a paper in preparation,[4] we investigate possible deviations from two pillars of the standard ΛCDM model, focusing on a local inhomogeneous description of the Universe and modified gravity models.

Here, we adopt the LTB metric,[2] a spherically symmetric solution of the Einstein field equations, to account for local deviations of the Universe from the homogeneity. Hence, we regard the weakly inhomogeneous LTB solution as the standard FLRW geometry plus small spherically symmetric perturbations.

Furthermore, in recent years increasing interest raised to provide features able to distinguish between GR and modified gravity approaches. Some open problems in cosmology, such as the Hubble constant tension,[5] may require new physics.[6] The aim of the present work is to compare the evolution of linear inhomogeneous perturbations in the ΛCDM and $f(R)$ modified gravity models.[7,8] More in detail, we used the equivalent formalism in the Jordan frame,[7–9] and we adopt the $f(R)$ Hu-Sawicki model,[10] which is a promising theoretical framework to reproduce the cosmic acceleration in the late Universe via the dynamics of a non-minimally coupled scalar field without a cosmological constant.

Then, we investigate the dynamics of inhomogeneous perturbations in the two cosmological models abovementioned. Moreover, the separable variables method is implemented in the analysis at the first-order of perturbation, and our approach is the same used in Ref. 11. We obtain an analytic expression only for the radial dependence of the linear perturbations, while their time evolution must be numerically evaluated. The most relevant issue, emerging from our analysis, is the different morphology of the radial dependence of the LTB solution in the ΛCDM formulation and Hu-Sawicki dynamics. Indeed, a distinctive Yukawa-like behavior occurs in the Jordan frame gravity regardless of the choice of a specific $f(R)$ model. Hence, we have found a peculiar feature to distinguish a modified gravity model from a cosmological constant scenario.

This work is organized as follows: in Sec. 1 we introduce the framework of the $f(R)$ extended gravity models; in Sec. 2 we adopt the LTB metric; then, in Sec. 3 we implement a perturbation approach to deal with local inhomogeneities. Finally, we summarize our work in Sec. 4. We adopt the metric signature $(-,+,+,+)$, and we set $c = 1$.

1. $f(R)$ models in the Jordan frame

A generalization of the Einstein-Hilbert theory is provided by the so-called $f(R)$ modified gravity models.[7,8] In this context, the gravitational Lagrangian density is expressed as a free function f of the Ricci scalar R, namely an extra degree of freedom compared to GR. The extended field equations in $f(R)$ gravity are fourth-order partial differential equations in the metric tensor components. In particular, if $f(R) = R$, the gravitational field equations end up in the Einstein equations in GR, trivially.

It is often convenient to reformulate $f(R)$ gravity in the scalar-tensor representation in the Jordan frame.[7–9] In doing so, the extra degree of freedom of the $f(R)$ function is converted into a non-minimally coupled scalar field. It can be checked

that the following action

$$S_J = \frac{1}{2\,\chi}\int_\Omega d^4x\,\sqrt{-g}\,\left[\phi\,R - V\,(\phi)\right] + S_M\,(g_{\mu\nu},\psi) \tag{1}$$

is dynamically equivalent to the action in the $f\,(R)$ metric formalism, if $f''\,(R) \neq 0$, where we have defined a scalar field $\phi = f'\,(R) = df/dR$, and a scalar field potential $V\,(\phi) = \phi\,R\,(\phi) - f\,(R\,(\phi))$. In Eq. (1), $\chi \equiv 8\,\pi\,G$ is the Einstein constant, G is the Newton constant, g is the determinant of the metric tensor $g_{\mu\nu}$, while S_M is the matter action, and ψ denotes matter fields. The advantage of the Jordan frame is that the corresponding field equations[8] are now second-order differential equations, but one has to deal with the non-minimally coupling.

The increasing attention to these extended gravity models is motivated by the consideration that the actual cosmic acceleration of the Universe might be obtained by geometry rather than a cosmological constant. A geometrical modification, indeed, could be regarded as an effective matter source.

Among several suggested $f\,(R)$ models,[7,8,12,13] we focus on the Hu-Sawicki proposal,[10] which is largely studied in the late Universe. The deviation $F\,(R) \equiv f\,(R) - R$ with respect to the GR scenario for the Hu-Sawicki (HS) model with $n = 1$ assumes the following form

$$F\,(R) = -m^2\frac{c_1\,R/m^2}{c_2\,R/m^2 + 1}\,, \tag{2}$$

where $m^2 \equiv \frac{\chi\,\rho_{m0}}{3}$ is related to the present matter density ρ_{m0}, while c_1 and c_2 are dimensionless parameters. It can be checked[10] that an effective cosmological constant is obtained for $R \gg m^2$. Furthermore, c_1 and c_2 can be constrained[10] specifying $F_{R0} \equiv dF\,/\,dR\;(z = 0)$ at the present cosmic time (redshift $z = 0$).

Finally, we write the scalar field potential $V\,(\phi)$ in the Jordan frame for the HS model with $n = 1$:

$$V\,(\phi) = \frac{m^2}{c_2}\,\left[c_1 + 1 - \phi - 2\sqrt{c_1\,(1-\phi)}\right]\,. \tag{3}$$

It can be shown[14] that the potential (3) ends in an asymptotically stable de Sitter Universe.

2. Inhomogeneous solutions in GR and in the Jordan frame

An inhomogeneous Universe can be described by using the LTB line element,[2] which is a spherically symmetric solution of the Einstein equations. The Universe in the LTB geometry appears inhomogeneous, but isotropic, from a single preferred point located at the center of the spherical symmetry. For instance, a cosmological dust or a spherical overdensity mass are well described by such a metric.

In the synchronous gauge, the LTB line element may be regarded as a generalization of the FLRW one, and it is given by

$$ds^2 = -dt^2 + e^{2\alpha}dr^2 + e^{2\beta}\,\left(d\theta^2 + sin^2\theta\,d\phi^2\right)\,, \tag{4}$$

where r is the radial distance from the preferred point, while $\alpha = \alpha(t, r)$ and $\beta = \beta(t, r)$ are two metric functions.

Considering a pressure-less dust ($p = 0$) and a cosmological constant Λ in GR, we can write the Einstein equations in the LTB metric. For instance, the 01 component of field equations is written as

$$\frac{\dot{\beta}'}{\beta'} - \dot{\alpha} + \dot{\beta} = 0 \tag{5}$$

where $\dot{()}$ and $()'$ denote derivatives with respect to time t and radial coordinate r, respectively. The other non-zero equations are the 00 and 11 components,[4] while different components vanish due to spherical symmetry.

It is possible to simplify the LTB metric within the framework of GR, using Eq. (5), which allows us to find a relation between the metric functions α and β. Following the approach in Ref. 2, the LTB line element (4) becomes

$$ds^2 = -dt^2 + \frac{[(a\,r)']^2}{1 - r^2\,K^2} dr^2 + (a\,r)^2 \left(d\theta^2 + sin^2\theta\, d\phi^2\right) , \tag{6}$$

where $a = a(t, r) = e^{\beta}/r$ is a generalized scale factor in a inhomogeneous geometry, and $K = K(r)$. Note that if $a(t, r)$ and $K(r)$ are independent of r, the metric (6) coincides with the FLRW line element, describing an isotropic and homogeneous geometry. Moreover, the 00 and 11 components of the Einstein equations in the LTB metric provides a generalization of the Friedmann equations in the ΛCDM model.

Concerning the $f(R)$ modified gravity scenario in the equivalent Jordan frame, the 01 component of the gravitational field equations in the LTB metric is written as

$$\frac{\dot{\beta}'}{\beta'} - \dot{\alpha} + \dot{\beta} = -\frac{1}{2\,\phi\,\beta'} \left(\dot{\phi}' - \dot{\alpha}\,\phi'\right) . \tag{7}$$

Furthermore, we can compute[4] also the 00 and 11 of field equations, as well as the scalar field equation in the Jordan frame. Observe that ϕ, ρ, α, and β are all functions of t and r in a inhomogeneous scenario.

Comparing Eq. (7) with the respective equation (5) obtained in GR, an extra term occurs in the Jordan frame due to non-minimal coupling between ϕ and the metric functions α, β. As a consequence, we can not relate α and β to rewrite the LTB metric as in the GR scenario. Moreover, the scalar field potential $V(\phi)$ does not affect Eq. (7), but it is included in the remaining field equations. $V(\phi)$ represents an extra degree of freedom with respect to GR, and the dynamics in the Jordan frame deviates from that of the Einstein theory.

3. Perturbation approach in GR and in the Jordan frame

In this Section, we investigate the different evolution of local inhomogeneities within the framework of GR and modified gravity.[4] Here, we stress one important

difference: we can not adopt the LTB metric in the form (6) in the Jordan frame, but we have to refer to the line element in Eq. (4), as motivated in Sec. 2.

Then, we consider the LTB metric to describe local inhomogeneities as small linear perturbations with respect to a flat FLRW metric as background. Hence, we can write all the physical quantities as a background term, denoted with (), plus a linear perturbation, which depends on t and r, and it is denoted with δ. For instance, the energy density is written as $\rho(t,r) = \bar{\rho}(t) + \delta\rho(t,r)$. Similarly, we rewrite the scale factor $a(t,r)$ in GR, the scalar field $\phi(t,r)$, the metric functions $\alpha(t,r)$, and $\beta(t,r)$ in the Jordan frame. We also assume that perturbation terms are small corrections if compared with respective background quantities, i.e. $\delta\rho \ll \bar{\rho}$. Furthermore, we consider $K^2(r)$ in the LTB metric (6) as a linear term, since it is related to inhomogeneities. It should be noted that, comparing the LTB line element (4) and a flat FLRW metric, we have to require two constraints: $\bar{\alpha}(t) = \ln(\bar{a}(t))$ and $\bar{\beta}(t,r) = \ln(\bar{a}(t)\, r)$.

Concerning the scalar-tensor formalism in the Jordan frame, we expand the scalar field potential $V(\phi)$ to the first order in $\delta\phi$. It can be easily checked that the zeroth order term $V\left(\bar{\phi}(t)\right)$ depends only on the background scalar field.[4]

Once we have split background and linear contributions, we focus on the dynamics in the ΛCDM cosmological model and in the Jordan frame gravity. Hence, we rewrite the field equations in GR and in the Jordan frame, separating background and perturbation terms.

3.1. *Comparing background solutions*

To study the time evolution, we define a dimensionless variable $\tau = t/t_0$, where t_0 is the actual time in the synchronous gauge. Note that t_0 is approximately $t_0 \approx 1/H_0$ in terms of the Hubble constant.

In GR, the 00 component of the Einstein equation in the LTB geometry becomes the Friedmann equation in the FLRW metric at background level. We set $\bar{a}_0 = \bar{a}(\tau = 1)$ today, and we neglect relativistic componenents in the late Universe. Hence, we obtain an analytical solution:[4] the evolution of the background scale factor in terms τ is written as

$$\bar{a}(\tau) = \left(\frac{\Omega_{m0}}{\Omega_{\Lambda 0}}\right)^{1/3} \left\{\sinh\left[\frac{3}{2}\sqrt{\Omega_{\Lambda 0}}\,(\tau - 1) + \operatorname{arcsinh}\left(\sqrt{\frac{\Omega_{\Lambda 0}}{\Omega_{m0}}}\right)\right]\right\}^{2/3} . \qquad (8)$$

We used the cosmological density parameters $\Omega_{m0} = \rho_{m0}/\rho_{c0}$, and $\Omega_{\Lambda 0} = \Lambda/\left(\chi\,\rho_{c0}\right)$ for matter and cosmological constant components, respectively, where $\rho_{c0} = 3H_0^2/\chi$ is the actual critical energy density of the Universe. We recall also that $\bar{\rho}(\tau) \propto \bar{a}(\tau)^{-3}$ for a matter component.

In Fig. 1 we plot the deceleration parameter $\bar{q}(\tau) \equiv -\ddot{\bar{a}}/\left(\bar{a}\,\bar{H}^2\right)$, fixing $\Omega_{m0} = 0.3111$ and $\Omega_{\Lambda 0} = 1 - \Omega_{m0}$ from Ref. 3. Here, the dot denotes the derivative with respect to τ. The scale factor $\bar{a}(\tau)$ increases for growing values of τ, and note that $\bar{q} \to -1$ in the future.

We point out that the solution (8) in the ΛCDM cosmological model is viable only at late times ($\tau \gg 10^{-6}$) in the Universe, otherwise we have to include relativistic components, and solve numerically the Friedmann equation.

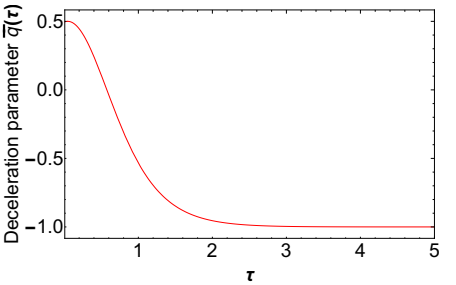


Fig. 1. Background deceleration parameter $\bar{q}(\tau)$ in the FLRW geometry in terms of a dimensionless variable τ. The background scenario in the ΛCDM model are graphically the same comparing it with the one in the $f(R)$ Hu-Sawicki model in the Jordan frame.

Now, we focus on the background field equations in the Jordan frame gravity. Starting from the 01, 00, 11 components of the gravitational field equations in the LTB metric and the scalar field equation, if we do not include local inhomogeneities, we get the respective equation system in the Jordan frame for the background FLRW metric (see the field equations in Ref. 8). We use the Hu-Sawicki model in the Jordan frame, which is characterized by the scalar field potential in Eq. (3).

To solve numerically the equation system, we fix Ω_{m0} at the same value used in the previous study for the ΛCDM model, and we choose[10] $|F_{R0}| = 10^{-7}$ to obtain $c_1 = 2.0 \cdot 10^6$ and $c_2 = 1.5 \cdot 10^5$. We set the free parameters in a way that the Hu-Sawicki model in the Jordan frame is very close to the ΛCDM scenario for the background level. In Fig. 2 note that the scalar field potential (3) exhibits a slow-roll mimicking a cosmological constant.

Finally, we solve numerically the equation system in the Jordan frame,[4] imposing the conditions at $\tau = 1 : \bar{\phi}(1) = 1 - |F_{R0}|$ and $d\bar{\phi}/d\tau\,(1) = 0$. The resulting plot of the deceleration parameter $\bar{q}(\tau)$ are graphically undistinguishable from the respective ones in the ΛCDM model (see Fig. 1). Concerning the evolution of the scalar field, we plot in Fig. 2 the deviation $\left|1 - \bar{\phi}(\tau)\right|$ from the GR limit ($\bar{\phi} = 1$) with a scale 10^{-7}. Note that the dynamics of the Jordan frame is very close to the one in ΛCDM model at the background order for any τ. Now, we can move on to the linearized solutions to highlight the differences between the two cosmological models.

3.2. *Comparing linear solutions*

To examine separately the time and radial evolution of local inhomogeneities, we use the separable variables method at the first-order perturbation theory. So, we

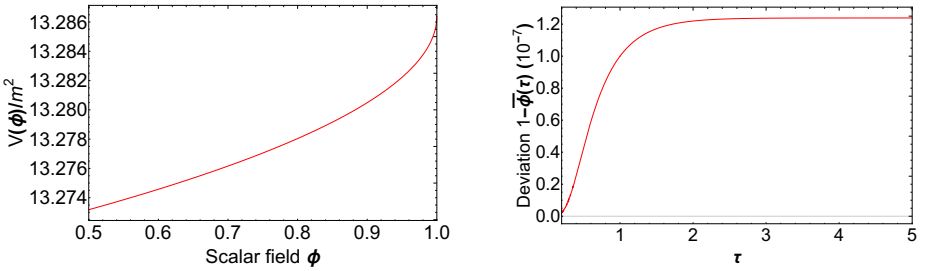


Fig. 2. Plots referred to the Hu-Sawicki model in the Jordan frame, adopting the FLRW metric (background solution). Left panel: scalar field potential of the Hu-Sawicki model in the Jordan frame. Right panel: deviation $|1-\bar{\phi}(\tau)|$ from the GR scenario.

factorize all the linear perturbations into time and radial functions. More specifically, we rewrite the perturbations of the scale factor $\delta a(\tau,r) \equiv \delta \mathfrak{a}(\tau)\, \delta \mathfrak{a}(r)$, the metric functions $\delta\alpha(\tau,r) \equiv \delta A(\tau)\, \delta\mathcal{A}(r)$ and $\delta\beta(\tau,r) \equiv \delta B(\tau)\, \delta\mathcal{B}(r)$, the energy density $\delta\rho(\tau,r) \equiv \delta P(\tau)\, \delta\varrho(r)$, and the scalar field $\delta\phi(\tau,r) \equiv \delta\Phi(\tau)\, \delta\varphi(r)$. The factorization is assumed at the linear level to find analytical solutions, but the separation of variables is not a general method, if we also include non-linear terms.

Then, we investigate the effect of inhomogeneities in the dynamics, according to the cosmological model considered. We focus on the linearized equations in the LTB metric within the framework of the ΛCDM cosmological model.

Using the factorization abovementioned, it is easy to show that the 11 component of the linearized Einstein equations in the LTB metric can be separated in two parts.[4] The radial behavior is $\delta\mathfrak{a}(r) = K^2(r)$,while we have an ordinary differential equation in terms of τ for the time evolution:

$$\delta\ddot{\mathfrak{a}}(\tau) + \frac{\dot{\bar{a}}(\tau)}{\bar{a}(\tau)}\,\delta\dot{\mathfrak{a}}(\tau) - \delta\mathfrak{a}(\tau)\left[\frac{\ddot{\bar{a}}(\tau)}{\bar{a}(\tau)} + \left(\frac{\dot{\bar{a}}(\tau)}{\bar{a}(\tau)}\right)^2\right] = 0\,. \tag{9}$$

This equation is solved numerically, once you specify the background solution (8). We impose the following conditions at $\tau = 1$ today: $\delta\mathfrak{a}(1) = 10^{-5}$ and $\delta\dot{\mathfrak{a}}(1) = 0$. In Fig. 3 we plot the numerical solution, after we have defined the ratio $\eta(\tau)$ between the perturbed scale factor and the background respective quantity, namely $\eta(\tau) \equiv |\delta\mathfrak{a}(\tau)\,/\,\bar{a}(\tau)|$. Note that the perturbations of the scale factor remain stable, since the background scale factor $\bar{a}(\tau)$ dominates for any τ in the late Universe, i.e. $\eta \ll 1$.

After that, we consider the 00 component of the linearized Einstein equations in the LTB metric, and we combine it with the continuity equation. After long but straightforward calculations,[4] we obtain $\delta\mathfrak{a} \propto 1/r^3$ in GR. According to the cosmological principle, inhomogeneities decay for increasing values of r. Furthermore, we also obtain that the perturbation of the energy density for a matter component follows the same dependence of the respective background quantity in the FLRW metric, namely $\delta P(\tau) \propto 1/\bar{a}^3(\tau)$. However, we can provide $\delta P(\tau) \ll \bar{\rho}(\tau)$ at any times, setting the constant of integration.

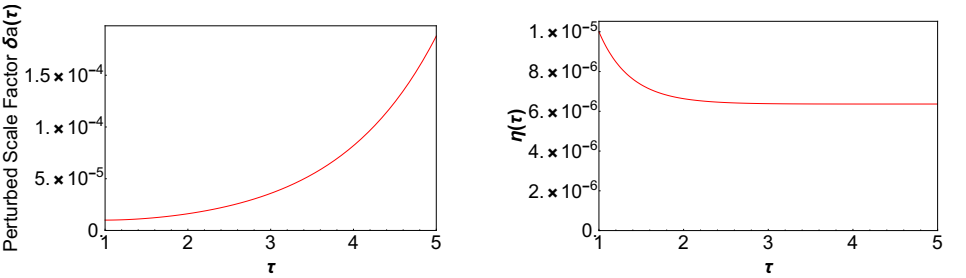


Fig. 3. Left panel: time evolution of the linear perturbation of the scale factor $\delta \mathrm{a}(\tau)$. Right panel: the ratio between the linear deviation and background scale factor.

Now, we repeat the same approach in the modified gravity dynamics in the Jordan frame, considering the corresponding linearized field equations in the LTB metric. Expanding the scalar field potential and its derivative around the background scalar field, as mentioned at the beginning of Sec. 3, it can be proved that the potential $V(\phi)$ is involved only in the time evolution of the linearized field equations.[4] As a result, we expect that the radial evolution of local inhomogeneities is completely independent of the choice of a particular $f(R)$ model, since $V(\phi)$ is directly related to $f(R)$.

In what follows, we do not focus on the time evolution of perturbations, but we just mention that we find a numerical solution using again the Hu-Sawicki model (Sec. 1), and we have checked that inhomogeneous perturbations remain smaller than respective background contribution for any time τ.

To obtain an analytical solution for the radial part with the separable variables method, we assume two simplifying conditions: $\delta \dot{A}(\tau) = \delta \dot{B}(\tau)$ and $\delta \varrho(r) \propto \delta \varphi(r)$. After long but straightforward calculations,[4] combining the linearized field equations, we obtain a Yukawa-like solution for the radial part of the linear perturbation of the scalar field

$$\delta \varphi(r) = \frac{C}{r} \exp\left(-\frac{r}{r_c}\right), \tag{10}$$

where C and r_c are constants. We also have a similar behavior for $\delta \mathcal{A}(r)$ and $\delta \mathcal{B}(r)$. It should be emphasized that modified gravity introduces a typical spatial scale, r_c, such that for $r \gg r_c$ the inhomogeneities decay faster than ones in GR. We stress again that this kind of solutions applies to any $f(R)$ model, and the radial evolution is different from the one in GR.

4. Summary and conclusions

We have investigated local inhomogeneities in the LTB metric regarded as small deviations from a flat FLRW background metric. To date, there are no modified gravity models that predict significant deviations from GR, reconciling all possible cosmological data. Hence, to try to discriminate between several cosmological

models, it is crucial to test gravity in different regimes or by using other techniques. Here, we have suggested one possible method, studying the different dynamics of inhomogeneous perturbations. Our results have pointed out a distinctive element in the evolution of local inhomogeneities of the Universe from a theoretical point of view, allowing to distinguish between the ΛCDM cosmological model and $f(R)$ modified gravity theories. We have shown that the radial evolution of inhomogeneous perturbations within the framework of the Jordan frame gravity is independent of the scalar field potential $V(\phi)$: the distinctive radial solution is a feature of any $f(R)$ model, and the radial behavior is completely different from the one in GR. This work may be an interesting arena to account for the effects of local inhomogeneities in cosmological observables,[15] when forthcoming missions such as Euclid Deep Survey,[16] will be able to test the large-scale properties of the Universe.

References

1. S. Weinberg, *Cosmology* (Oxford University Press, 2008)
2. P. J. E. Peebles, *Principles of Physical Cosmology* (Princeton University Press, 1993)
3. N. Aghanim, Y. Akrami, M. Ashdown, *et al.*, Planck 2018 results. VI. Cosmological parameters, *A&A* **641**, A6 (2020)
4. T. Schiavone, G. Montani, P. Marcoccia, Signature of $f(R)$ gravity via Lemaître-Tolman-Bondi inhomogeneous perturbations, to be submitted
5. E. Di Valentino, O. Mena, P. Supriya, *et al.*, In the realm of the Hubble tension — A review of solutions, *Class. Quantum Grav.* **38**, 153001 (2021)
6. M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, and G. Lambiase, On the Hubble Constant Tension in the SNe Ia Pantheon Sample, *ApJ* **912(2)**, 150 (2021)
7. S. Nojiri, and S. D. Odintsov, Introduction to modified gravity and gravitational alternative for dark energy, *IJGMM* **4**, 115 (2007)
8. T. P. Sotiriou, and V. Faraoni, $f(R)$ theories of gravity, *Rev. Mod. Phys.* **82**, 451 (2010)
9. S. Capozziello, and V. Faraoni, *Beyond Einstein gravity* (Springer, 2013)
10. W. Hu, and I. Sawicki, Models of $f(R)$ cosmic acceleration that evade solar system tests, *Phys. Rev. D* **76**, 064004 (2007)
11. P. Marcoccia, and G. Montani, Weakly Inhomogeneous models for the Low-Redshift Universe, arXiv:1808.01489
12. A. A. Starobinsky, Disappearing cosmological constant in $f(R)$ gravity, *JETP Lett.* **86**, 157 (2007)
13. S. Tsujikawa, Observational signatures of $f(R)$ dark energy models that satisfy cosmological and local gravity constraints, *Phys. Rev. D* **77**, 023507 (2008)
14. A. de la Cruz-Dombriz, P. K. S. Dunsby, S. Kandhai, and D. Saez-Gomez, Theoretical and observational constraints of viable $f(R)$ theories of gravity, *Phys. Rev. D* **93**, 084016 (2016)
15. G. Fanizza, B. Fiorini, and G. Marozzi, Cosmic variance of H_0 in light of forthcoming high-redshift surveys, *Phys. Rev. D* **104**, 083506 (2021)
16. L. Amendola, S. Appleby, A. Avgoustidis., *et al.*, Cosmology and fundamental physics with the Euclid satellite, *Living Rev. Relativ.* **21**, 2 (2018)

Soft dark energy and soft dark matter

Emmanuel N. Saridakis
National Observatory of Athens, Lofos Nymfon, 11852 Athens, Greece
E-mail: msaridak@noa.gr

We present "soft cosmology", namely we investigate small deviations from the usual framework due to the effective appearance of soft-matter properties in the Universe sectors. One effect of such a case would be the dark energy to exhibit a different equation-of-state parameter at large scales (which determine the universe expansion) and at intermediate scales (which determine the sub-horizon clustering and the large scale structure formation). Concerning soft dark matter, we show that it can effectively arise due to the dark-energy clustering, even if dark energy is not soft. We propose a novel parametrization introducing the "softness parameters" of the dark sectors. As we see, although the background evolution remains unaffected, due to the extreme sensitivity and significant effects on the global properties even a slightly non-trivial softness parameter can improve the clustering behavior and alleviate e.g. the $f\sigma_8$ tension.

Keywords: Dark energy; Dark matter; Cosmological perturbations; Growth rate; σ_8 tension; MG16 Proceedings.

1. Introduction

Standard cosmology has been proven very efficient, qualitatively and quantitatively, in describing the Universe evolution and properties at early and late times, as well as at large and small scales. Nevertheless, since cosmology has now become an accurate science, with the appearance of a huge amount of data of progressively increasing precision, slight disagreements, deviations and tensions between theory and observations lead to a large variety of extensions and modifications of the concordance paradigm (a procedure that might be assisted by purely theoretical investigation).

Although in the usual ways of extensions one may add various novel fields, fluids, sectors and their mutual interactions,[1,2] or alter the underlying gravitational theory,[3–5] there is a rather strong assumption that is maintained in all of them, namely that the sectors that constitute the Universe are simple, or equivalently that one can apply the physics, the hydrodynamics and thermodynamics of usual, "hard" matter. Nevertheless, in condensed matter physics it is well known that there is a large variety of "soft" matter forms, which are characterized by complexity, simultaneous co-existence of phases, entropy dominance, extreme sensitivity, viscoelasticity, etc, properties that arise effectively at intermediate scales due to scale-dependent effective interactions that are not present at the fundamental scales.[6,7]

In this work, based on,[8] we examine the possibility of "soft cosmology", namely small deviations from the usual framework due to the effective appearance of

soft-matter properties in the Universe sectors. We mention that due to the extreme sensitivity and significant effects of softness on the global properties, one does not need to consider a large deviation from standard considerations, since even a very slight departure would be adequate to improve the observed cosmological behavior at the required level. Finally, we stress that new fundamental physics is not directly needed, since the dark energy dynamical evolution and clustering, which is a widely accepted possibility in many scenarios beyond ΛCDM paradigm, is adequate to effectively induce the soft behavior.

2. Standard Cosmology

Let us briefly present the basics of cosmology.[9] The cosmological principle (the universe is homogeneous and isotropic at large scales) allows to consider the Friedmann-Robertson-Walker (FRW) metric $ds^2 = dt^2 - a^2(t)\,\delta_{ij}dx^i dx^j$. Concerning the universe content, one considers the usual baryonic matter and radiation (i.e. all Standard Model particles), the dark matter sector, as well as the dark energy sector. The microphysics of dark matter is unknown, and its source may be most probably some particle(s), however it may arise from black holes, from modified gravity, or even from a combination of the above, i.e. a multi-component dark matter.[10] The microphysics of dark energy is unknown too, and it may arise from new fields or matter forms in the framework of general relativity, or it may have an effective nature of gravitational origin due to modifications of gravity.

The next step is to consider that (at least after a particular stage of the universe evolution) cosmological scales are suitably large in order to allow one to neglect the microphysics of the universe ingredients and describe them effectively through fluid dynamics and continuum flow (at earlier stages one should use the Boltzmann equation). Hence, one can ignore the microscopic Lagrangian of the various sectors, and write their energy momentum tensors as $T^{(i)}_{\mu\nu} = (\rho_i + p_i)u_\mu u_\nu + p_i g_{\mu\nu}$, with ρ_i and p_i the energy density and pressure of the fluid corresponding to the i-th sector, u_μ the 4-velocity vector field and $g_{\mu\nu}$ the metric. Note that one can extend the above expression by including viscosity or/and heat flux.

Under the above considerations, any cosmological scenario will be determined by the two Friedmann equations

$$H^2 = \frac{\kappa^2}{3}(\rho_b + \rho_r + \rho_{dm} + \rho_{de}), \tag{1}$$

$$2\dot{H} + 3H^2 = -\kappa^2(p_b + p_r + p_{dm} + p_{de}), \tag{2}$$

with $\kappa^2 = 8\pi G$, and where $H \equiv \dot{a}/a$ is the Hubble parameter. We mention that the above equations are of general validity, holding for every model of dark matter and dark energy. Additionally, the conservation equation $\nabla^\mu T^{(tot)}_{\mu\nu} = \nabla^\mu \left[\sum_i T^{(i)}_{\mu\nu}\right] = 0$ in the case of FRW geometry and for non-interacting fluids gives rise to the separate conservation equations $\dot{\rho}_i + 3H(\rho_i + p_i) = 0$, while the extension to interacting

cosmology can be realized through phenomenological descriptors Q_i of the interaction with $\sum_i Q_i = 0$ and with $\dot{\rho}_i + 3H(\rho_i + p_i) = Q_i$.

In order for the equations to close we need to impose the equation of state for each sector. The usual consideration is to assume barotropic fluids, in which the pressure is a function of the energy density only, with the simplest case being $p_i = w_i \rho_i$ with w_i the equation-of-state parameter. Lastly, note that the above framework provides ΛCDM cosmology for $\rho_{de} = -p_{de} = \Lambda/\kappa^2$, with Λ the cosmological constant.

The above formulation of cosmological evolution allows one to proceed to a more subtle investigation, and study small perturbations around the FRW background. Focusing without loss of generality to the linear theory of scalar isentropic perturbations in the Newtonian gauge, imposing $ds^2 = -(1+2\Psi)dt^2 + a^2(t)(1-2\Phi)\gamma_{ij}dx^i dx^j$, then in a general non-interacting scenario which includes the aforementioned sectors the scalar perturbations are determined by the equations[11,12]

$$\dot{\delta}_i + (1+w_i)\left(\frac{\theta_i}{a} - 3\dot{\Psi}\right) + 3H[c_{\rm eff}^{(i)2} - w_i]\delta_i = 0, \tag{3}$$

$$\dot{\theta}_i + H\left[1 - 3w_i + \frac{\dot{w}_i}{H(1+w_i)}\right]\theta_i - \frac{k^2 c_{\rm eff}^{(i)2}\delta_i}{(1+w_i)a} - \frac{k^2\Psi}{a} = 0, \tag{4}$$

assuming zero anisotropic stress and with k the wavenumber of Fourier modes (in the case of ΛCDM paradigm the corresponding dark-energy perturbation equations are not considered). In the above equations $\delta_i \equiv \delta\rho_i/\rho_i$ are the density perturbations and θ_i is the divergence of the fluid velocity. Furthermore, $c_{\rm eff}^{(i)2} \equiv \delta p_i/\delta\rho_i$ is the effective sound speed square of the i-th sector (it determines the amount of clustering). Note that the above equations can be simplified through the consideration of the Poisson equation, which at sub-horizon scales can be written as:[11,12]

$$-\frac{k^2}{a^2}\Psi = \frac{3}{2}H^2 \sum_i \left[\left(1 + 3c_{\rm eff}^{(i)2}\right)\Omega_i\delta_i\right]. \tag{5}$$

Finally, we note here that in general the above formulation can be applied also in the cases where the dark energy sector is an effective one arising from gravitational modifications.

We close this section by mentioning that one can find a big variety and many versions of the above formalism. However, there is a rather strong assumption that is maintained in all of them, namely that the sectors that constitute the Universe are simple, or equivalently that one can apply the physics of usual matter. In particular, the underlying assumption is that the laws that determine the Universe behavior at large scales can be induced by the laws that determine the interactions between its individual constituents. Focusing on the hydrodynamic description, the use of fluid energy densities and pressures arises from the assumption that we can define fundamental "particles" of the corresponding sector, the collective flow of which gives rise to ρ_i and p_i, while all physics below the particle scale has been integrated out.

3. Soft Cosmology

The above formulation of standard cosmology is definitely correct and can provide a quite successful quantitative description of the Universe evolution. However, the question is: could it miss something on the details?

One source of possible information loss is the assumption that the Universe sectors and interactions are simple and more or less scale-independent. For instance, two regions of the dark-matter fluid will mutual interact in the same way in the intermediate- and late-time universe, or the interaction of two big clusters of dark matter can arise by the superposition of all individual interactions of their sub-clusters, etc. In other words, in the concordance cosmological formulation one assumes that the sectors that constitute the Universe behave as usual, "hard" matter.

"Soft" matter is a research field that has attracted a large amount of interest of the condensed matter community,[6,7] since it has very interesting and peculiar properties far different than those of hard matter. The problem is that there is not a definition of what is soft matter. In particular, the best definition we have is that soft matter is the one that has the properties of soft materials. Examples of soft materials are the polymers (plastic, rubber, polystyrene, lubricants etc), the colloids (paints, milk, ice-cream etc), the surfactants, granular materials, liquid crystals, gels, biological matter (proteins, RNA, DNA, viruses, etc), etc. Although these examples of soft matter are very different from each other, they have some common properties and features that distinguish them from usual, hard, matter. Amongst others these include complexity (new qualitative properties arise at intermediate scales due to interactions that are not present at the fundamental scales), co-existence of phases (they have different phase properties depending on the scale at one examines them, e.g at the same time they can be fluid at small scales and solid at large scales), entropy dominance instead of energy dominance, flexibility, extreme sensitivity to reactions, viscoelasticity (they exhibit viscous and elastic properties simultaneously) etc.

In the following we examine the possibility that the dark sectors of the universe may exhibit (intrinsically or effectively) slight soft properties, which could then lead to small corrections to the corcodance model. We mention here that the discussion below holds independently of the underlying gravitational theory, i.e it is valid both in the framework of general relativity as well as in modified gravity, nevertheless in the latter case we have richer possibilities to obtain scale-dependent interactions.

3.1. *A. Soft Dark Energy*

The nature and underlying physics of dark energy is unknown. The basic framework that has been studied in extensive detail is that the dark energy fluid has the same fluid properties at all scales at a given moment/redshift. However, as we saw, in soft matter the complexity that arises at intermediate scales may lead the material to have a different equation of state (EoS) at different scales simultaneously.

As a simple phenomenological model of soft dark energy we may consider the case where dark energy has the usual EoS at large scales, namely at scales entering the Friedmann equations, but having a different value at intermediate scales, namely at scales entering the perturbation equations. In this case the Universe's expansion history will remain identical to standard cosmology, nevertheless the large-scale structure evolution can deviate from the standard one and can be brought closer to observations. In summary one can obtain richer behavior. We mention that in the following we focus on sub-horizon scales $k \gg aH$ and thus to perturbation modes affected only by the intermediate-scale dark-energy EoS. The full analysis, in which different perturbation modes are affected by different EoS according to their scale, will be presented elsewhere.

A first approach on the subject would be to introduce the effective "softness parameter" s_{de} of the dark energy sector. This implies that while at cosmological, large scales (ls) dark energy has the usual EoS, namely w_{de-ls}, at intermediate scales (is) we have[8]

$$w_{de-is} = s_{de} \cdot w_{de-ls}, \tag{6}$$

and standard cosmology is recovered for $s_{de} = 1$.

For instance let us assume that the large-scale dark energy EoS w_{de-ls} is a constant one $w_{de-ls} = w_0$ or e.g. the CPL one $w_{de-ls} = w_0 + w_a(1-a)$. According to (6) at intermediate scales the dark energy EoS w_{de-is} is different, either constant $w_{de-is} = w_2$ or time-varying. Hence, the background Universe evolution will remain the same, however since $c_{\rm eff}^{(de)}$ will change, through the Poisson equation and (3),(4) we will acquire a different evolution for the matter overdensity δ_m. Hence, the resulting $f\sigma_8 \equiv f(a)\sigma(a)$ value, with $f(a) = d\ln\delta_m(a)/d\ln a$ and $\sigma(a) = \sigma_8\delta_m(a)/\delta_m(1)$, will be different than the corresponding one of standard cosmology with the above dark energy EoS (note that since the background behavior remains unaffected we do not need to worry about incorporating fiducial cosmology.[13] As we observe, we have a straightforward way to alleviate the σ_8 tension since we can suitably adjust w_{de-is} in order to obtain slightly lower $f\sigma_8$. As a specific example in Fig. 1 we depict the $f\sigma_8$ as a function of z. The dashed curve is for ΛCDM. The solid curve is for soft dark energy with $s_{dm} = 1.1$, i.e. with $w_{de-is} = -1.1$, and $c_{\rm eff}^{(de)} = 0.1$, while dark matter is the standard one (i.e. not soft) with $w_{dm} = 0$. Note that in principle s_{de} can be varying too and one could introduce its parametrization, or one could additionally have more complicated situations in which w_{de-is} and w_{de-ls} have different parametrizations. In this first approach on soft dark energy we consider the simplest case of (6).

Finally, note that soft materials may exhibit different EoS properties not only at different scales, but at different directions too. In this case, one may think of a dark energy sector that has a different EoS at different directions, namely an anisotropic dark energy. However, such an analysis would require to deviate from FRW and consider explicitly anisotropic geometries such as the Bianchi ones. We will study this possibility in a separate work.

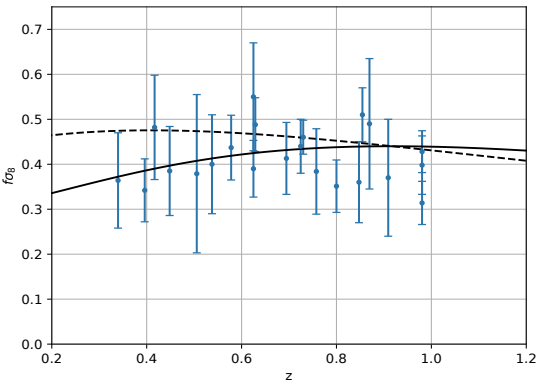


Fig. 1. The $f\sigma_8$ as a function of z. The dashed curve is for ΛCDM. The solid curve is for soft dark energy with $s_{de} = 1.1$, i.e. with $w_{de-ls} = -1$ and $w_{de-is} = -1.1$, and $c_{\rm eff}^{(de)} = 0.1$, while dark matter is standard (i.e. not soft) with $w_{dm} = 0$.

3.2. *B. Soft Dark Matter*

In this subsection we examine the possibility that the dark matter sector exhibits soft properties. In the framework of general relativity the gravitational interaction of dark matter with itself or with baryonic matter cannot produce internal complexity (unless the unknown microphysics of dark matter does impose an intrinsic soft structure). However, even if it is not intrinsic, soft behavior in the dark matter sector can still arise in an effective way due to the presence of non-trivial dark energy. Specifically, if the dark energy is clustering then, even if dark energy is not intrinsically soft, it will induce scale-dependent, qualitatively different intermediate structures in the dark matter clustering, at scales similar to the dark energy clusters. In particular, the interaction between two dark-matter clusters below the dark-energy clustering scale (i.e. two dark-matter clusters with sparse dark energy between them) will be different from the interaction between two dark-matter clusters with a dark-energy cluster between them. Hence, one will have the effective appearance of screening effects at intermediate scales, and thus of complexity (this is the standard way that complexity appears in the colloids, namely due to the non-trivial, scale-dependent structure of the bulk between them). In summary, one could have a dark matter sector which at large scales, namely at scales entering the Friedmann equations, behaves in the usual dust way, but which at intermediate scales, namely at scales entering the (sub-horizon) perturbation equations, it could slightly deviate from that.

We can introduce the dark matter softness parameter s_{dm} (standard cosmology is recovered for $s_{dm} = 1$) as:

$$w_{dm-is} + 1 = s_{dm} \cdot (w_{dm-ls} + 1), \tag{7}$$

(mind the difference in the parametrization comparing to soft dark energy, in order to handle the fact that the dark matter EoS at large scales w_{dm-ls} is 0). Similarly to the example of the previous subsection, the background evolution will remain identical with that of standard cosmology, but the perturbation behavior (at sub-horizon scales) and the large-scale structure can be improved. In order to provide a specific example, in Fig. 2 we depict the $f\sigma_8$ as a function of z, in the case of soft dark matter with $w_{dm-is} = 0.05$, i.e for dark matter with softness parameter $s_{dm} = 1.05$.

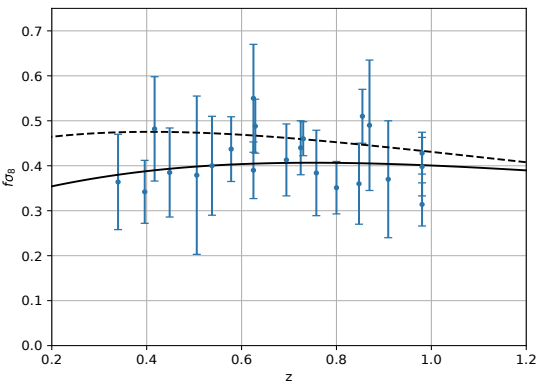


Fig. 2. The $f\sigma_8$ as a function of z. The dashed curve is for ΛCDM. The solid curve is for soft dark matter with softness parameter $s_{dm} = 1.05$, i.e. for dark matter with $w_{dm-ls} = 0$ and $w_{dm-is} = 0.05$ (note that dark energy is not soft).

In the above analysis we did not consider the dark energy sector to be soft. Definitely, proceeding to such a possibility would make the induced soft bahavior for dark matter easier. Moreover, this would be the case if one considers a mutual interaction between dark matter and dark energy too, since a different than usual dark-energy clustering behavior would be transferred to a different than usual dark-matter clustering behavior, due to the interaction. Finally, deviating from general relativity would provide additional possibilities to induce effective soft properties to the dark matter sector, since dark matter will implicitly interact in a scale-dependent way (one would have the additional screening behavior due to the extra (scalar) graviton degrees of freedom that dress the dark matter in a scale-dependent way, altering its self-interaction[16]).

Let us make a comment here on the clustering features. The clustering behavior of soft matter has been extensively studied, and indeed it has been shown that the resulting spectrum, factorial moments, fractal dimension, etc, depend on the specific intermediate-scale features. For instance the fractal dimension has been experimentally found to cover all the range from 1 to 3 according to different materials (e.g. colloids of gold nanoparticle in aqueous media give $d_f = 1.75 \pm 0.05$ for

diffusion-limited kinetics and $d_f = 2.01 \pm 0.10$ for reaction-limited kinetics)[17,18]). On the other hand, the large scale structure and the galaxy distribution in the Universe has a fractal dimension $d_f = 1.63 \pm 0.20$.[19,20] The fact that soft matter clustering exhibits naturally non-trivial dynamics due to its intermediate-scale complexity, could be useful in describing the details of the observed large-scale structure. We mention here that the non-trivial clustering of soft matter changes at scales below the intermediate ones, and hence soft dark matter could alleviate the cuspy halo problem,[21] the dwarf galaxy problem,[22] and other clustering-related problems that seem to puzzle the standard collisionless dark matter.

4. Conclusions

We examined the possibility of "soft cosmology", namely small deviations from the usual framework due to the effective appearance of soft properties in the Universe sectors. We started by considering the possibility of soft dark energy due to intermediate-scale features that could arise from its unknown microphysics. One effect of such a case would be the dark energy to exhibit a different EoS at large and intermediate scales. As we saw, although the background evolution remains unaffected, even a slight softness at intermediate scales can improve the clustering behavior and alleviate e.g. the $f\sigma_8$ tension.

We proceeded to the examination of soft dark matter, which can effectively arise just due to the dark-energy clustering even if dark energy is not soft. By considering a slightly different equation of state at large and intermediate scales we were able to improve the clustering behavior. Furthermore, in the additional incorporation of soft dark energy, and/or modified gravity, the effective soft properties of dark matter could be richer, due to the extra screening mechanisms.

We mention that in this work we incorporated softness by phenomenologically introducing a slightly different EoS at different scales. Clearly, in order to incorporate complexity and estimate the scale-dependent behavior of the equation of states from first principles one should revise and extend the cosmological perturbation theory and perform a detailed mesoscopic statistical mechanical analysis. Such a full investigation is necessary and would provide a robust argumentation in favour of soft cosmology.

References

1. E. J. Copeland, M. Sami and S. Tsujikawa, *Dynamics of dark energy*, Int. J. Mod. Phys. D **15**, 1753-1936 (2006).
2. Y. F. Cai, E. N. Saridakis, M. R. Setare and J. Q. Xia, *Quintom Cosmology: Theoretical implications and observations*, Phys. Rept. **493**, 1-60 (2010)
3. E. N. Saridakis *et al.* [CANTATA], *Modified Gravity and Cosmology: An Update by the CANTATA Network*, [arXiv:2105.12582 [gr-qc]].
4. S. Capozziello and M. De Laurentis, *Extended Theories of Gravity*, Phys. Rept. **509**, 167-321 (2011).

5. Y. F. Cai, S. Capozziello, M. De Laurentis and E. N. Saridakis, *f(T) teleparallel gravity and cosmology*, Rept. Prog. Phys. **79**, no.10, 106901 (2016).
6. R. A. L. Jones, *Soft Condensed Matter*, Oxford University Press, Oxford (2002).
7. L. M. C. Sagis, *Dynamic properties of interfaces in soft matter: Experiments and theory*, Rev. Mod. Phys. 83, 1367-1403 (2011).
8. E. N. Saridakis, *Do we need soft cosmology?*, Phys. Lett. B **822**, 136649 (2021).
9. P. J. E. Peebles, *Principles of physical cosmology*, Princeton University Press, Princeton (1993).
10. G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, Phys. Rept. **405**, 279-390 (2005).
11. V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, *Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions*, Phys. Rept. **215**, 203-333 (1992).
12. C. P. Ma and E. Bertschinger, *Cosmological perturbation theory in the synchronous and conformal Newtonian gauges*, Astrophys. J. **455**, 7-25 (1995).
13. N. Aghanim *et al.* [Planck], *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641**, A6 (2020).
14. W. J. Percival and M. White, *Testing cosmological structure formation using redshift-space distortions*, Mon. Not. Roy. Astron. Soc. **393**, 297 (2009).
15. A. P. Deshpande, J. M. Krishnan, S. Kumar, *Rheology of Complex Fluids*, Springer, Heidelberg (2010).
16. A. Joyce, B. Jain, J. Khoury and M. Trodden, *Beyond the Cosmological Standard Model*, Phys. Rept. **568**, 1-98 (2015).
17. D. A. Weitz, J. S. Huang, M. Y. Lin, and J. Sung, *Limits of the Fractal Dimension for Irreversible Kinetic Aggregation of Gold Colloids*, Phys. Rev. Lett. 54 (1985) 1416-1419.
18. S. Lazzari, L. Nicoud, B. Jaquet, M. Lattuada, M. Morbidelli, *Fractal-like structures in colloid science*, Advances in colloid and interface science 235, 1.
19. P. H. Coleman and L. Pietronero, *The fractal structure of the universe*, Phys. Rept. **213**, 311-389 (1992).
20. S. Teles, A. R. Lopes and M. B. Ribeiro, *Fractal Analysis of the UltraVISTA Galaxy Survey*, Phys. Lett. B **813**, 136034 (2021).
21. R. Teyssier, A. Pontzen, Y. Dubois and J. Read, *Cusp-core transformations in dwarf galaxies: observational predictions*, Mon. Not. Roy. Astron. Soc. **429**, 3068 (2013).
22. J. S. Bullock and M. Boylan-Kolchin, *Small-Scale Challenges to the ΛCDM Paradigm*, Ann. Rev. Astron. Astrophys. **55**, 343-387 (2017).
23. M.-a. Watanabe, S. Kanno and J. Soda, *The Nature of Primordial Fluctuations from Anisotropic Inflation*, Prog. Theor. Phys. **123**, 1041-1068 (2010).
24. M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha, *Hydrodynamics of soft active matter*, Rev. Mod. Phys. 82 (2013) 1143-1189.

A simple parametrisation for coupled dark energy

Vitor da Fonseca* and Nelson J. Nunes

Instituto de Astrofísica e Ciências do Espaço,
Faculdade de Ciências da Universidade de Lisboa,
Campo Grande, PT1749-016 Lisboa, Portugal
**E-mail: fc52156@alunos.fc.ul.pt*

Tiago Barreiro

Instituto de Astrofísica e Ciências do Espaço,
ECEO, Universidade Lusófona de Humanidades e Tecnologias,
Campo Grande, 376, 1749-024 Lisboa, Portugal

We proposes a phenomenological generalisation of the standard model with only one extra degree of freedom that parametrises the evolution of a scalar field responsible for the cosmic acceleration. The model also foresees an additional parameter in the form of a coupling between dark energy and dark matter. This model captures a large diversity of dark energy evolutions at low redshift and could usefully complement common CPL parametrisations widely used. In this context, we have been constraining the parametrisation with data from Planck and KiDS, bringing different results between the early and late universe observations.

Keywords: Quintessence, dark matter, observational cosmology.

1. Introduction

The discovery in 1998 that the Universe is accelerating[1,2] was really stunning at the time. While the source of the cosmic acceleration is yet to be unveiled, the existence of a dark energy component has been postulated to explain it.[3,4] In the standard model, dark energy takes the form of a cosmological constant Λ which by definition possesses a constant energy density and negative pressure that fights the attractive pull of matter.[5] That model is, at the moment, compatible with all observations and it is the simplest one,[6] at least from a parametrisation point of view. One of the outstanding issue in the concordance model is the so-called coincidence problem.[7] Given how the energy densities dilute with the expansion, the fact that the cosmological constant is catching up with matter at the present time necessitates initial conditions at the origin that are dramatically precise. This issue has been paving the way to study varying dark energy[8] in order to reproduce more dynamically the current abundances.

The simplest and model independent approach is phenomenological. It consists in parametrising the dark energy equation of state, being the pressure over the energy density that equals -1 for a cosmological constant. The detection of a time dependence would exclude a cosmological constant. Introducing the fewest

possible parameters is ideal to limit degeneracies between them. A simple and popular parametrisation[9] is a Taylor expansion at first order which has a very limited domain of validity. Various improvements have been introduced, like the Chevallier-Polarski-Linder (CPL) parametrisation.[10,11] Also, to account for the possibility of early dark energy, a parametrisation stepwise in redshift[12] has been proposed with a larger number of additional parameters.

Here, we explore an alternative way to parametrise dark energy by proposing to parametrise a scalar field ϕ called quintessence that would be responsible for dark energy both in the early and late Universe.[13] Instead of setting an equation of state parameter evolution, or a scalar potential, this method parametrises the dark energy scalar field itself. We argue that with a single parameter we recover the generic behaviour of the previous parametrisations. We test this approach with data from Planck and KiDS to constrain its specific parameters, identifying contrasting results between the early and late universe observations. This is a comprehensive work going from the analytical solutions to parameter constraints.

2. Scalar field parametrisation

We consider that the cosmological fluid is composed of matter (m), i.e. baryons (b) and cold dark matter (c), as well as the scalar field itself in a flat Friedmann-Lemaitre-Robertson-Walker background. The scalar field is supposed to be canonical and homogeneous at large scale,[14,15] approximated by a perfect fluid. We choose to work with the number of e-folds, $N \equiv \ln a$, as the time variable. In the following, the prime stands for the derivative with respect to the number of e-foldings, and the dot the derivative with respect to cosmic time.

By combining the Friedmann equation,

$$H^2 = \frac{\kappa^2}{3}\left(\rho_b + \rho_c + \rho_\phi\right), \tag{1}$$

with the continuity equation of the scalar field,

$$\rho'_\phi + 3H^2\phi'^2 = 0, \tag{2}$$

and the continuity equation of matter,

$$\rho'_c + 3\rho_c = 0, \tag{3}$$

one obtains[16] a differential equation for the dark energy density (where we conventionnally set $\kappa = 1$),

$$\frac{\rho'_\phi}{\rho_{m_0}} = -\phi'^2\left(a^{-3} + \frac{\rho_\phi}{\rho_{m_0}}\right), \tag{4}$$

whose general solution is our starting point,

$$\frac{\rho_\phi}{\rho_{m_0}} = e^{-\int_0^N dN\phi'^2} \times \left[\frac{\Omega_{\phi_0}}{\Omega_{m_0}} - \int_0^N dN\phi'^2 e^{-3N+\int_0^N dN\phi'^2}\right]. \tag{5}$$

Among possible parametrisations enabling us to solve the former equation analytically,

$$\phi' = \epsilon\phi, \tag{6}$$

$$\phi' = \sigma\phi^{-1}, \tag{7}$$

$$\phi' = \lambda, \tag{8}$$

where ϵ, σ and λ are given constants, we explore the simplest one in Eq. (8).[16]

The advantage of this parametrisation is three folds. Firstly, it is very simple, as we only add one single parameter to extend the standard model. It allows to reconstruct analytically the form of the potential,[16]

$$V(\phi) = A\, e^{-\frac{3}{\lambda}\phi} + B\, e^{-\lambda\phi}, \tag{9}$$

with the following mass scales,

$$A = \frac{3}{2}\frac{\lambda^2}{3-\lambda^2} H_0^2\, \Omega_{m_0}\, e^{\frac{3}{\lambda}\phi_0}, \tag{10}$$

$$B = \frac{3}{2}\frac{\lambda^2-6}{3-\lambda^2} H_0^2 \left(\frac{\lambda^2}{3} - \Omega_{\phi_0}\right) e^{\lambda\phi_0}, \tag{11}$$

where H_0 is today's Hubble expansion rate and $\Omega_{i_0} = \rho_{i_0}/3H_0^2$ today's abundance parameter of species i. Having a simple form of the potential allows a straightforward implementation of the model within the existing Boltzmann code CLASS[17,18] that can be validated by comparing the numerical results with the expected asymptotic behaviour of the analytic expressions. Given the small number of parameters involved and thanks to its simplicity, our analysis permits to effectively constrain the model parameters with current data as in the next section.

Secondly, this scalar potential, being the sum of two exponential terms, does alleviate the severe fine-tuning of the initial conditions as illustrated in Fig. 1. For small values of the parameter λ, when the first exponential term in the potential is steep enough, the scalar field energy density is attracted and temporarily scales with the background as the universe expands, be it dominated by radiation and then matter. Later, thanks to the second shallow exponential, whose slope is the inverse of the first one, the scalar field energy density eventually freezes, mimicking a cosmological constant. By freezing, dark energy overcomes matter at late time, enabling the acceleration of the universe that we are currently experiencing. This well-known attractor mechanism[19] ensures that one can start from a large domain of initial conditions for the scalar field, to reach unavoidably the tracking solution and end up with the correct order of magnitude both during matter and dark energy domination.

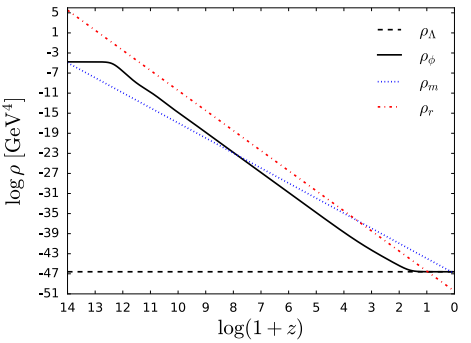


Fig. 1. Energy density dilution of a cosmological constant, as well as quintessence, matter and radiation for the scalar field parametrisation ($\lambda = 0.01$).

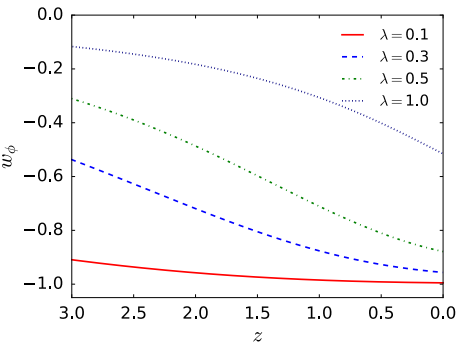


Fig. 2. Dynamics of the dark energy equation of state for given values of the parameter λ.

Last but not least, despite just one extra parameter, the model covers a wide range of possible dark energy evolution at low redshift as displayed in Fig 2. The parametrisation is thus able to capture a large variety of possible dynamics for dark energy. The observational data might even become precise enough in the future for discriminating between these different evolutions.

As a step further, we decide to complement our parametrisation by assuming a non-minimally coupling between dark energy and cold dark matter.[20] Again for the sake of simplicity, we consider a conformal and constant coupling β that parametrises the interaction between dark matter and the scalar field. The stress-energy tensor is jointly conserved within the dark sector in order to ensure the covariance of the theory,

$$\nabla_\mu \left(T^{(\phi)}{}^{\mu}_{\nu} + T^{(c)}{}^{\mu}_{\nu} \right) = 0, \tag{12}$$

$$\nabla_\mu T^{(\phi)}{}^{\mu}_{\nu} = +\beta \rho_c \nabla_\mu \phi, \tag{13}$$

$$\nabla_\mu T^{(c)}{}^{\mu}_{\nu} = -\beta \rho_c \nabla_\mu \phi. \tag{14}$$

Assuming that dark energy does not interact with baryons, the latter are separately conserved contrary to dark matter and quintessence,

$$\rho_b = \rho_{b_0}\, e^{-3N}, \tag{15}$$

$$\rho_c = \rho_{c_0}\, e^{-3N+\beta(\phi-\phi_0)}, \tag{16}$$

$$\rho'_\phi + 3H^2\phi'^2 = -\beta\phi'\rho_c. \tag{17}$$

The sign of the product $\beta\phi'$, i.e. $\beta\lambda$ in our parametrisation, determines whether energy is being pumped from the scalar field or from dark matter. Similarly to Ref. 16, we use the Friedmann equation (1) with the solution we find for the scalar field energy density,

$$\begin{aligned}\rho_\phi = &\left(\rho_{\phi_0} + \frac{\lambda^2+\beta\lambda}{\lambda^2+\beta\lambda-3}\,\rho_{c_0} + \frac{\lambda^2}{\lambda^2-3}\,\rho_{b_0}\right) e^{-\lambda^2 N} \\ &- \frac{\lambda^2+\beta\lambda}{\lambda^2+\beta\lambda-3}\,\rho_{c_0}\, e^{(\beta\lambda-3)N} - \frac{\lambda^2}{\lambda^2-3}\,\rho_{b_0}\, e^{-3N},\end{aligned} \tag{18}$$

to reconstruct, for the first time, the corresponding potential with similar properties of the dynamical system as in the minimally coupled case,

$$V(\phi) = A\, e^{\left(-\frac{3}{\lambda}+\beta\right)\phi} + B\, e^{-\lambda\phi} + C\, e^{-\frac{3}{\lambda}\phi}, \tag{19}$$

where the mass scales are now the following,

$$A = \frac{3}{2}\,\frac{\lambda^2+2\beta\lambda}{3-\lambda^2-\beta\lambda}\; H_0^2\,\Omega_{c_0}\, e^{\left(\frac{3}{\lambda}-\beta\right)\phi_0}, \tag{20}$$

$$B = \frac{6-\lambda^2}{2}\, H_0^2 \left(1 + \frac{3}{\lambda^2-3+\beta\lambda}\Omega_{c_0} + \frac{3}{\lambda^2-3}\Omega_{b_0}\right) e^{\lambda\phi_0}, \tag{21}$$

$$C = \frac{3}{2}\,\frac{\lambda^2}{3-\lambda^2}\; H_0^2\,\Omega_{b_0}\, e^{\frac{3}{\lambda}\phi_0}. \tag{22}$$

We modify the CLASS code to accommodate that potential as well as the coupled equations (16) and (17) to simulate the model with the two additional parameters λ and β. The numerical results plotted in Fig. 3 are in line with the asymptotic solution for the dark energy equation of state during matter domination when baryonic matter is neglected,

$$w_\phi \approx -\frac{\beta}{\lambda+\beta}. \tag{23}$$

One can note that today's equation of state, w_0, does not depend on the coupling, and drops close to -1 during the scalar field dominated epoch for small values of λ,

$$w_{_0} = -1 + \frac{\lambda^2}{3\Omega_{\phi_0}}. \tag{24}$$

Furthermore, in order to numerically predict observables that we use to constrain the parameters with observational datasets at the perturbative level, we again modify CLASS to implement the coupled evolution of the scalar fluctuations at linear

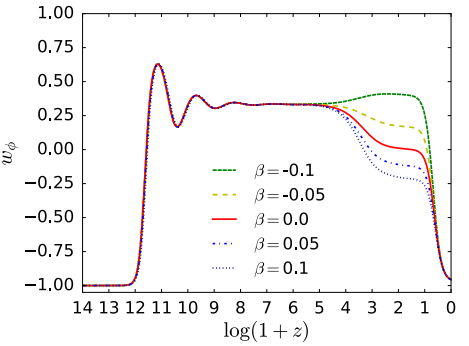


Fig. 3. Dark energy evolution in the coupled parametrisation. Beta negative corresponds to the transfer of energy from the dark matter component to the scalar field.

level in the synchronous gauge adopting usual conventions.[21] We modify the existing perturbed Klein-Gordon equation, as well as the equation of motion for the dark matter density contrast δ_c and velocity perturbations θ_c to account for the coupling, by adding the relevant sources term in the right-hand side,

$$\ddot{\varphi} + 2\mathcal{H}\dot{\varphi} + \left(k^2 + a^2 V_{,\phi\phi}\right)\varphi + \frac{\dot{h}\dot{\phi}}{2} = -\beta a^2 \bar{\rho}_c \delta_c, \tag{25}$$

$$\dot{\delta}_c + \theta_c + \frac{\dot{h}}{2} = \beta\dot{\varphi}, \tag{26}$$

$$\dot{\theta}_c + \mathcal{H}\theta_c = \beta\left(k^2\varphi - \dot{\phi}\theta_c\right), \tag{27}$$

where the overbar denotes the background quantities. Additionally, we adapt the transformation from the synchronous gauge into the comoving gauge to maintain the matter density source function δ_m gauge-invariant,

$$\delta_m^C = \frac{\delta\rho_m}{\bar{\rho}_m} + \left(3 - \frac{\beta\phi'\bar{\rho}_c}{\bar{\rho}_c + \bar{\rho}_b}\right)\frac{\mathcal{H}}{k^2}\theta_m. \tag{28}$$

In parallel, to provide for the code validation, we derive the analytic equation of motion of the dark matter fluctuations in the Newtonian limit during the domination of matter, precisely when the coupling makes a difference,

$$\delta_c'' + \frac{1}{2}\left(1 + 3\lambda\beta\right)\delta_c' - \frac{3}{2}\left(1 - \frac{\lambda^2}{3} - \frac{\lambda\beta}{3}\right)\left(1 + 2\beta^2\right)\delta_c = 0. \tag{29}$$

In the left panel of Fig 4, we note that the scalar field slows the growth of perturbations against the standard model, by decreasing the dynamical term in the equation of motion. On the other hand, the existence of the coupling brings competitive effects that further slow the fluctuation or, on the contrary, accelerate them, depending on whether energy is being transferred from the scalar field or into it. We are also able to get an analytic expression of the growth rate $m+$ under the

Newtonian approximation,

$$m_+ = \frac{1}{4}\left[-1 - 3\lambda\beta + \sqrt{24\left(1 - \frac{\lambda^2}{3} - \frac{\lambda\beta}{3}\right)(1 + 2\beta^2) + (1 + 3\lambda\beta)^2}\right]. \qquad (30)$$

One can see in the right panel of Fig. 4 that the numerical results in the matter era on a small scale are very much in line with it. The standard growth, which equals 1, is obviously recovered for vanishing values of the parameters λ and β.

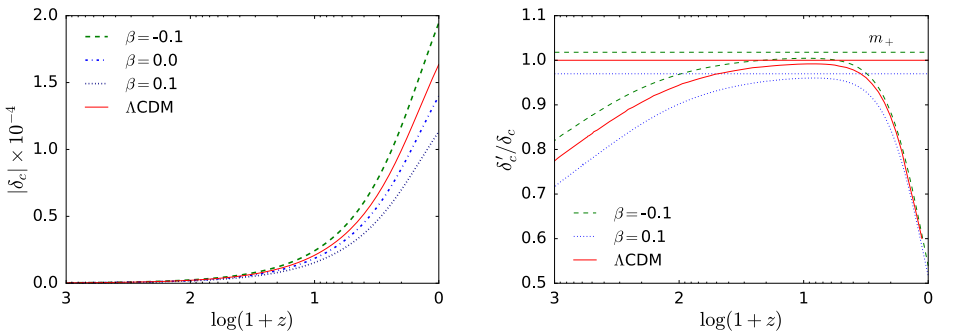


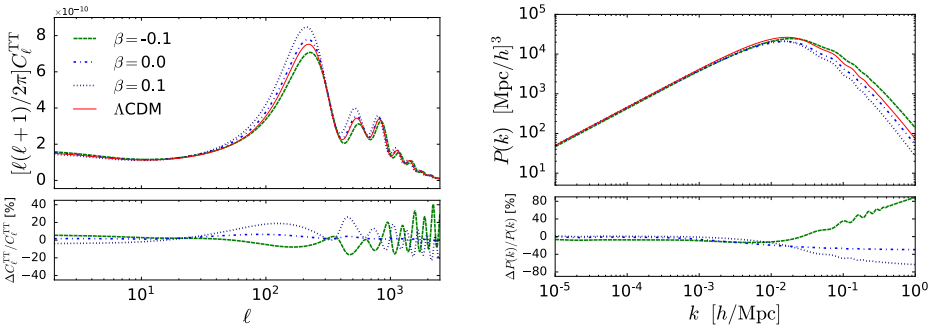
Fig. 4. $\lambda = 0.3$ on scale $k = 0.1\ h/\mathrm{Mpc}$. (a) Dark matter fluctuations. (b) Growth rate function.

3. Observational constraints

We turn now our attention to the observations in order to constrain the value of the posteriors λ and β through Bayesian inference with MontePython[22–27] and GetDist.[28] Given that it affects the growth of dark matter perturbations, our parametrisation leaves observational signatures, both on the CMB anisotropies and matter power spectrum that we can numerically predict with our adapted version of CLASS and compare with observations.

As for the CMB anisotropies, early dark energy as well as the coupling affect the amplitude and position of the acoustic peaks as illustrated on the left panel of Fig. 5. We use the data from the Planck 2018 likelihood[29] on the temperature and polarisation of the CMB in the MCMC analysis. As regards today's linear matter power spectrum, the delay or the acceleration in the growth of dark matter perturbations has consequences too (right panel of Fig. 5). We use the gravitational weak lensing measurements, cosmic shear, from the KiDS-450 survey[30] in the Bayesian fitting.

The probability contours and posterior distribution we obtain in the parameter space are displayed on Fig. 6 for the two specific degrees of freedom of the model. On the one hand, the parameter estimation done with Planck provides an upper bound for the posterior λ which is compatible with a cosmological constant, that



(a) (b)

Fig. 5. Observables predictions for $\lambda = 0.3$ (a) CMB dimensionless angular power spectrum. (b) Linear matter power spectrum at redshift $z = 0$.

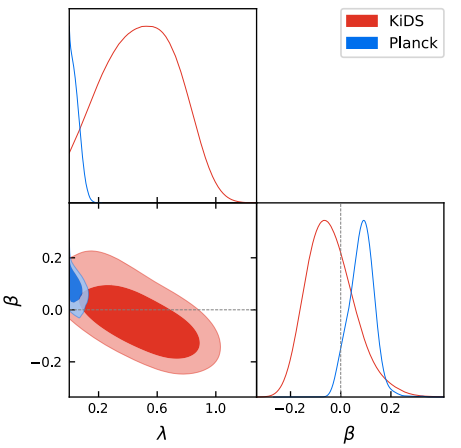


Fig. 6. Constraints on the (λ,β) posterior plane. Contours with 68% and 95% probability.

cannot be discarded. The coupling posterior corresponds to the energy injection from dark energy to dark matter. However, on the other hand, weak lensing data brings a different constraint on the posterior λ, not favouring a vanishing value. It also slightly detects energy transfer from dark matter to dark energy.

We also undertake the same likelihood analysis with the standard model as a benchmark to compare the goodness-of-fit of the parametrisation (Table 1). In light of the reduced chi-square, the two models look more or less equivalent. Nonetheless, according to the Akaike Information Criterion (AIC),[31] since it further penalises model complexity, the favourite cosmology for both probes is ΛCDM, this is particularly true for KiDS.

Table 1. Standard ΛCDM model vs scalar field parametrisation

	Planck		KiDS	
	ΛCDM	ϕ param.	ΛCDM	ϕ param.
χ^2	602.6	599.7	321.5	321.7
$\chi^2_{\rm red}$	1.0	1.0	2.6	2.7
AIC	616.6	617.7	335.5	339.7
ΔAIC	0	1.1	0	4.2

4. Concluding remarks

Despite its simplicity, the parametrisation we propose is able to cover a large span of dark energy evolution with a limited number of parameters. Constraints on more complex parametrisations are usually limited by the degeneracies between the various degrees of freedom. The model preserves the fact that dark matter and dark energy are in comparable ratios today, while relaxing the initial conditions thanks to the existence of tracking solutions. It is thus capable of reproducing the evolution of the background, as well as the CMB anisotropies and the formation of Large Scale Structures. Since the parametrisation has a ΛCDM limit for small λ and vanishing coupling β, it allows a direct test of departures from the standard model. We therefore argue that the parametrisation does help constrain the evolution of the equation of state and the departure from ΛCDM, even though the approach is purely phenomenological. Finally, the MCMC analysis carried out at perturbation level identifies tensions on the posteriors between early and late time probes. Background data is not expected to provide stronger constraints but the upcoming Euclid data will certainly improve the situation.

Acknowledgements

The authors thank Giuseppe Fanizza and Miguel Zumalacarregui for their useful comments. This research was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grants: UID/FIS/04434/2019, PTDC/FIS-OUT/29048/2017 (DarkRipple), COMPETE 2020:POCI-01-0145-FEDER-028987 & FCT:PTDC/FIS-AST/28987/2017 (CosmoESPRESSO) and IF/00852/2015.

References

1. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
2. S. Perlmutter *et al.*, Measurements of Omega and Lambda from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
3. L. Amendola and S. Tsujikawa, *Dark Energy* (Cambridge University Press, 2015).
4. T. Clifton, P. G. Ferreira, A. Padilla and C. Skordis, Modified Gravity and Cosmology, *Phys. Rept.* **513**, 1 (2012).

5. P. J. E. Peebles and B. Ratra, The Cosmological constant and dark energy, *Rev. Mod. Phys.* **75**, 559 (2003) [592(2002)].
6. Planck Collaboration, Planck 2018 results. VI. Cosmological parameters, *arXiv e-prints*, p. arXiv:1807.06209 (Jul 2018).
7. J. Martin, Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask), *Comptes Rendus Physique* **13**, 566 (2012).
8. P. J. E. Peebles and B. Ratra, Cosmology with a Time Variable Cosmological Constant, *Astrophys. J.* **325**, p. L17 (1988).
9. J. Weller and A. Albrecht, Future supernovae observations as a probe of dark energy, *Physical Review D* **65** (May 2002).
10. M. Chevallier and D. Polarski, Accelerating Universes with Scaling Dark Matter, *International Journal of Modern Physics D* **10**, 213 (2001), (uses Latex, 12 pages, 6 Figures) Minor corrections, Figures 4, 6 revised. Conclusions unchanged.
11. E. Linder, Exploring the expansion history of the universe, *Physical review letters* **90**, p. 091301 (04 2003).
12. P. S. Corasaniti and E. J. Copeland, Model independent approach to the dark energy equation of state, *Physical Review D* **67** (Mar 2003).
13. C. Wetterich, The Cosmon model for an asymptotically vanishing time dependent cosmological 'constant', *Astron. Astrophys.* **301**, 321 (1995).
14. R. R. Caldwell, R. Dave and P. J. Steinhardt, Cosmological imprint of an energy component with general equation of state, *Physical Review Letters* **80**, p. 1582–1585 (Feb 1998).
15. S. Tsujikawa, Quintessence: A Review, *Class. Quant. Grav.* **30**, p. 214003 (2013).
16. N. J. Nunes and J. E. Lidsey, Reconstructing the dark energy equation of state with varying alpha, *Phys. Rev.* **D69**, p. 123511 (2004).
17. J. Lesgourgues, The cosmic linear anisotropy solving system (class) i: Overview (2011).
18. D. Blas, J. Lesgourgues and T. Tram, The cosmic linear anisotropy solving system (class). part ii: Approximation schemes, *Journal of Cosmology and Astroparticle Physics* **2011**, p. 034–034 (Jul 2011).
19. T. Barreiro, E. J. Copeland and N. J. Nunes, Quintessence arising from exponential potentials, *Phys. Rev.* **D61**, p. 127301 (2000).
20. L. Amendola, Coupled quintessence, *Phys. Rev. D* **62**, p. 043511 (2000).
21. C.-P. Ma and E. Bertschinger, Cosmological perturbation theory in the synchronous and conformal newtonian gauges, *The Astrophysical Journal* **455**, p. 7 (Dec 1995).
22. B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, *JCAP* **1302**, p. 001 (2013).
23. T. Brinckmann and J. Lesgourgues, MontePython 3: boosted MCMC sampler and other features (2018).
24. F. Feroz and M. P. Hobson, Multimodal nested sampling: an efficient and robust alternative to markov chain monte carlo methods for astronomical data analyses, *Monthly Notices of the Royal Astronomical Society* **384**, p. 449–463 (Jan 2008).
25. F. Feroz, M. P. Hobson and M. Bridges, Multinest: An efficient and robust bayesian inference tool for cosmology and particle physics, *Monthly Notices of the Royal Astronomical Society* **398**, p. 1601–1614 (Oct 2009).
26. F. Feroz, M. P. Hobson, E. Cameron and A. N. Pettitt, Importance nested sampling and the multinest algorithm, *The Open Journal of Astrophysics* **2** (Nov 2019).
27. J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley and D. Kocevski, X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue, *Astron. Astrophys.* **564**, p. A125 (2014).

28. A. Lewis, GetDist: a Python package for analysing Monte Carlo samples (2019).
29. P. Collaboration, Planck 2018 results. v. cmb power spectra and likelihoods (2019).
30. H. Hildebrandt, M. Viola, C. Heymans, S. Joudaki, K. Kuijken, C. Blake, T. Erben, B. Joachimi, D. Klaes, L. Miller and et al., Kids-450: cosmological parameter constraints from tomographic weak gravitational lensing, *Monthly Notices of the Royal Astronomical Society* **465**, p. 1454–1498 (Nov 2016).
31. H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**, 716 (1974).

A tale of two double quasars: Hubble constant tension or biases?

L. J. Goicoechea* and V. N. Shalyapin†

GLENDAMA Project Core Team, Universidad de Cantabria,
Avda. de Los Castros 48, E-39005 Santander, Spain
gravlens.unican.es
**E-mail: goicol@unican.es*
†*Main addresses: O.Ya. Usikov Institute for Radiophysics and Electronics,*
National Academy of Sciences of Ukraine,
12 Acad. Proscury St., UA-61085 Kharkiv, Ukraine
and
Institute of Astronomy of V.N. Karazin Kharkiv National University,
Svobody Sq. 4, UA-61022 Kharkiv, Ukraine

For a flat ΛCDM (standard) cosmology, a small sample of gravitationally lensed quasars with measured time delays has recently provided a value of the Hubble constant H_0 in tension with the *Planck* flat ΛCDM result. Trying to check if this tension is real or not, we used basic observational constraints for two double quasars of the GLENDAMA sample (SBS 0909+532 and SDSS J1339+1310) to discuss the underlying value of H_0 in a standard cosmology. For SBS 0909+532, we were not able to obtain a reliable measurement of H_0. However, the current data of SDSS J1339+1310 are consistent with H_0 around 67.8 km s^{-1} Mpc^{-1} and $\sigma(H_0)/H_0 \sim 10\%$. Although the formal uncertainty is still large and mainly due to the lack of details on the mass density profile of the main lens galaxy, the central value of H_0 coincides with that of the TDCOSMO+SLACS collaboration (using gravitational lens systems) and is within the 1σ interval from *Planck* cosmic microwave background data. After getting these preliminary encouraging results through only one double quasar, we are currently planning to use several GLENDAMA systems to accurately measure the Hubble constant and put constraints on other cosmological parameters.

Keywords: Gravitational lensing: strong; quasars: individual (SBS 0909+532, SDSS J1339+1310); cosmological parameters.

1. Introduction

Optical photometric monitoring of gravitationally lensed quasars (GLQs) brings plenty of astrophysical information.[1] For example, time delays between correlated brightness variations of their multiple images are used to estimate the current expansion rate of the Universe (the so-called Hubble constant H_0), provided lensing mass distributions can be constrained by observational data.[2,3] Throughout this paper, H_0 is expressed in standard units of km s^{-1} Mpc^{-1}, so units only are explicitly given in tables and figures.

Very recently, the H0LiCOW collaboration performed a joint analysis of six GLQs with measured time delays.[4] For a flat ΛCDM standard cosmology, they obtained $H_0 = 73.3^{+1.7}_{-1.8}$, in good agreement with $H_0 = 74.03 \pm 1.42$ from SNe data by

the SH0ES collaboration,[5] but in apparent tension with Cosmic Microwave Background (CMB) data.[6] *Planck* observations of the CMB suggested a narrow 1σ interval ranging from 66.9 to 67.9, which is clearly inconsistent with H0LiCOW/SH0ES results. It is also worth noting that Freedman *et al.*[7] obtained an intermediate value of $H_0 = 69.8 \pm 1.9$.

The big question is whether the tension between early and late-Universe probes is due to systematic errors or has a physical origin. Possible systematic errors in some methods may fix this issue, avoiding hasty rejection of the standard cosmological model. Thus, we use two doubly imaged quasars of the Gravitational LENses and DArk MAtter (GLENDAMA) sample[8] to discuss the influence of observational constraints, and hypotheses and priors on the mass model in the estimation of the Hubble constant in a standard cosmology. Section 2 briefly presents the GLENDAMA project and the framework of time-delay cosmography through double quasars, while Sec. 3 and Sec. 4 include preliminary results for the GLQs SBS 0909+532 and SDSS J1339+1310, respectively. A discussion of results and future prospects appear in Sec. 5.

2. GLENDAMA project and H_0 from doubles

The GLENDAMA project[a] is aimed to accurately study a sample of ten GLQs in the Northern Hemisphere over a period of about 25 years, basically covering the first quarter of this century.[8] The sample includes seven double quasars with two images (A and B) each, and three quads having four images (A, B, C and D) each. Figure 1 shows the distribution on the sky of the selected, optically bright GLQs. The Gran Telescopio CANARIAS (GTC) is being used to obtain deep spectroscopy of the lens systems, while the optical variability of quasar images is traced from light curves mainly based on observations with the Liverpool Telescope (LT). These optical light curves are allowing us to measure the time delay $\Delta t_{\rm AB}$ in doubles, and three independent delays $\Delta t_{\rm AB}$, $\Delta t_{\rm AC}$ and $\Delta t_{\rm AD}$ in quads (see results in Table 1). Current GLENDAMA delays have been estimated to a typical accuracy of about 5%, with only two exceptions: the relative error in the delay of QSO 0957+561 is well below 5%, and the three delays of the quad HE 1413+117 have larger relative uncertainties.

The time delay between the two images of a double quasar can be expressed in terms of the so-called time-delay distance $D_{\Delta t}$, the speed of light c and a dimensionless factor $\Delta\Phi_{\rm AB}$, so that[3] $\Delta t_{\rm AB} = (D_{\Delta t}/c)\Delta\Phi_{\rm AB}$. Here, $D_{\Delta t}$ depends on the source (quasar) and deflector (main lens galaxy) redshifts, as well as cosmological parameters. Measuring the redshifts, and assuming a flat ΛCDM cosmological model with $\Omega_{\rm M} = 0.3$ (matter density) and $\Omega_\Lambda = 0.7$ (dark energy density), $D_{\Delta t}/c$ is given as a known constant divided by H_0[b]. Additionally, $\Delta\Phi_{\rm AB}$ depends on the position

[a] `https://gravlens.unican.es/`.

[b] The time-delay distance does not appreciably change when matter and dark energy densities are slightly different to 0.3 and 0.7, respectively.

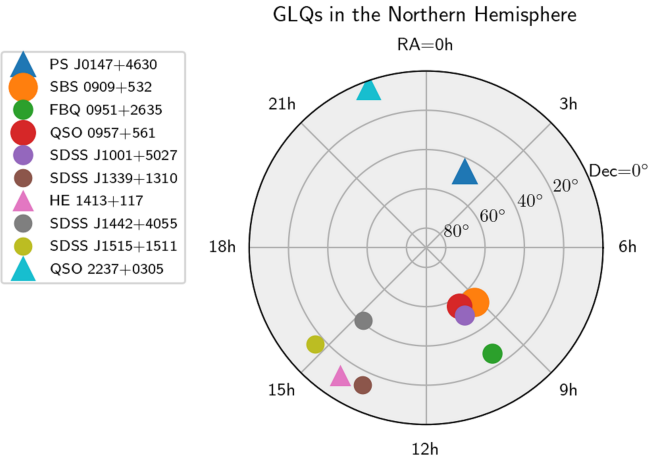


Fig. 1. GLENDAMA GLQs in the Northern Hemisphere. Triangles and circles represent quadruply and doubly imaged quasars, respectively. Larger symbols mean brighter quasars.

Table 1. GLENDAMA time delays from light curves in the SDSS r band.

GLQ	Δt_{AB} (days)	Δt_{AC} (days)	Δt_{AD} (days)	Reference
PS J0147+4630	pm	pm	pm	—
SBS 0909+532	50^{+2}_{-4}	—	—	Ref. 9
FBQ 0951+2635	pm	—	—	—
QSO 0957+561	420.6 ± 1.9	—	—	Ref. 10
SDSS J1339+1310	48 ± 2	—	—	Ref. 11
HE 1413+117	17 ± 3	20 ± 4	23 ± 4	Ref. 12
SDSS J1442+4055	25.0 ± 1.5	—	—	Ref. 13
SDSS J1515+1511	211 ± 5	—	—	Ref. 14

Note: pm = preliminary measure.

of both images and the source, and the lens potential at the image positions.[2,3] Hence, the lensing mass distribution determines the value of the multiplicative factor $\Delta\Phi_{AB}$.

We used a lens model to describe the primary lensing mass in SBS 0909+532 and SDSS J1339+1310. For each of these two double quasars, our lens model consisted of an elliptical surface mass density to account for the main lens galaxy G and an external shear γ due to the gravitational action of other galaxies around the lens system. The surface mass density of G was modeled as a singular power-law distribution since a composite model (treating baryons and dark matter individually) leads to similar results.[15,16] In this preliminar study, instead of using high-resolution imaging to put constraints on the power-law index of G, we focused on an isothermal distribution, i.e., a singular isothermal ellipsoid (SIE). Such distribution

is consistent with stellar and gas motions in the Milky Way, as well as observations of many spiral and elliptical galaxies. We also did not use the stellar kinematics of G.[17]

We considered constraints on the time delay, the relative astrometry and the flux ratio between images, along with some observationally-motivated priors on SIE+γ lens model parameters. These constraints/priors and the LENSMODEL software[18] allowed us to simultaneously fit lens model parameters, position and flux of the source quasar, and $H_0^{\rm model}$ with dof = 0, where "dof" means degrees of freedom. In addition to the mass that is explicitly included in the lens model (main deflector plus external shear), we must take the mass along the line of sight to G into account. This additional effect can be approximated as an external convergence in the lens plane $\kappa_{\rm ext}$, which may be positive or negative depending on the mass distribution along the sightline. The true time-delay distance $D_{\Delta t}^{\rm true}$ relates to that derived from the lens model and measured delay $D_{\Delta t}^{\rm model}$ by $D_{\Delta t}^{\rm true} = D_{\Delta t}^{\rm model}/(1-\kappa_{\rm ext})$ (e.g., see Eq. (4) of Ref. 4), which leads to $H_0^{\rm true} = H_0^{\rm model}(1-\kappa_{\rm ext})$. Therefore, when accounting for an external convergence, the Hubble constant decreases/increases in a factor $1-\kappa_{\rm ext}$. The two next sections deal with estimates of $H_0^{\rm model}$ from observations of SBS 0909+532 and SDSS J1339+1310.

3. SBS 0909+532

SBS 0909+532 is a doubly imaged quasar in which the background source (quasar) and the foreground early-type lens galaxy (main deflector) have redshifts $z_{\rm s} = 1.377$ and $z_{\rm d} = 0.830$, respectively.[19–21] Our first set of observational constraints consisted of the SBS 0909+532 time delay in Table 1 taking a symmetric uncertainty (50 ± 3 days)[c], the relative astrometry of B and G (their positions with respect to A at the origin of coordinates) and the flux of B in units such that the flux of A is equal to one. These last astro-photometric constraints were taken from the *HST* near-IR data in Table 3 of Ref. 22. We also considered priors on the ellipticity e and external shear of the SIE+γ lens model described in Sec. 2: $e \leq 0.5$ (see Table 3 of Ref. 23) and $\gamma \leq 0.1$ (see Table 4 of Ref. 22).

Although the data fit yielded a best solution for $H_0^{\rm model}$ of 68.4 (see Table 2), unfortunately, the *HST* relative astrometry of Lehár *et al.*[22] is not so good as would be desirable. For instance, the relative position of the faint lens galaxy G was determined with a large uncertainty of about 100 mas (1 mas = 0.″001). The insufficiently accurate astrometric measures were responsible for a broad valley in the χ^2 curve (see the black solid line in Fig. 2), so the 1σ confidence interval for $H_0^{\rm model}$ included values below 55 and above 80. If we were able to improve the Lehár *et al.*'s astrometry, e.g., reducing errors in relative positions of B and G in factors 3 and 10, respectively, the best solution of $H_0^{\rm model}$ would be practically the same, but

[c]Despite 49 ± 3 days is fully consistent with the measurement in Table 1, initially we have preferred to keep its central value and divide the error bar into two identical halves.

Table 2. Results for $H_0^{\rm model}$ using a SIE+γ lens model (see main text).

	Observational constraints						Priors on model parameters			$H_0^{\rm model}$	
GLQ	$\Delta t_{\rm AB}$ [a]	$\Delta x_{\rm AB}$ [b]	$\Delta y_{\rm AB}$ [b]	$\Delta x_{\rm AG}$ [b]	$\Delta y_{\rm AG}$ [b]	$F_{\rm B}/F_{\rm A}$ [c]	e [d]	θ_e [d]	γ [e]	best [f]	1σ [f]
SBS 0909+532	50 ± 3	-0.987 ± 0.003	-0.498 ± 0.003	-0.415 ± 0.100	-0.004 ± 0.100	0.89 ± 0.10	≤ 0.5	—	≤ 0.1	68.4	—
		-0.987 ± 0.001	-0.498 ± 0.001	-0.415 ± 0.010	-0.004 ± 0.010					68.3	68.5 ± 7.5 [g]
		-0.9868 ± 0.0006	-0.4973 ± 0.0006	-0.464 ± 0.003	-0.055 ± 0.003	0.88 ± 0.10	0.11 ± 0.08	-48.1 ± 16.9	—	38.1	38.2 ± 3.3 [h]
							≤ 0.5	—	≤ 0.1	47.5	—
SDSS J1339+1310	47.0 ± 5.5	$+1.419 \pm 0.001$	$+0.939 \pm 0.001$	$+0.981 \pm 0.010$	$+0.485 \pm 0.010$	0.175 ± 0.015	0.18 ± 0.05	32 ± 10	—	69.1	69^{+10}_{-8} [i]
	48 ± 2									67.6	67.8 ± 4.4

Note:
[a] Time delay between both images in days. Some errors have been made symmetric.
[b] Relative positions of B and G with respect to A at the origin of coordinates. Here, Δx and Δy are given in arc seconds, and their positive directions are defined by west and north, respectively. For SBS 0909+532, some errors have been conveniently approximated.
[c] Flux ratio. For SBS 0909+532, errors are enlarged to 10% to account for moderate microlensing effects.
[d] Ellipticity and position angle of the SIE. The position angle (θ_e) is measured east of north.
[e] External shear strength.
[f] Best solution and 1σ confidence interval for $H_0^{\rm model}$. We use standard units of km s^{-1} Mpc^{-1}.
[g] Plausible but not real measurement. Astrometric errors have been reduced to "achievable" values (see next row).
[h] Real measurement, but based on a biased astrometry or an inappropriate (strongly affected by microlensing) flux ratio.
[i] Measurement relying on an old, innacurate time delay.

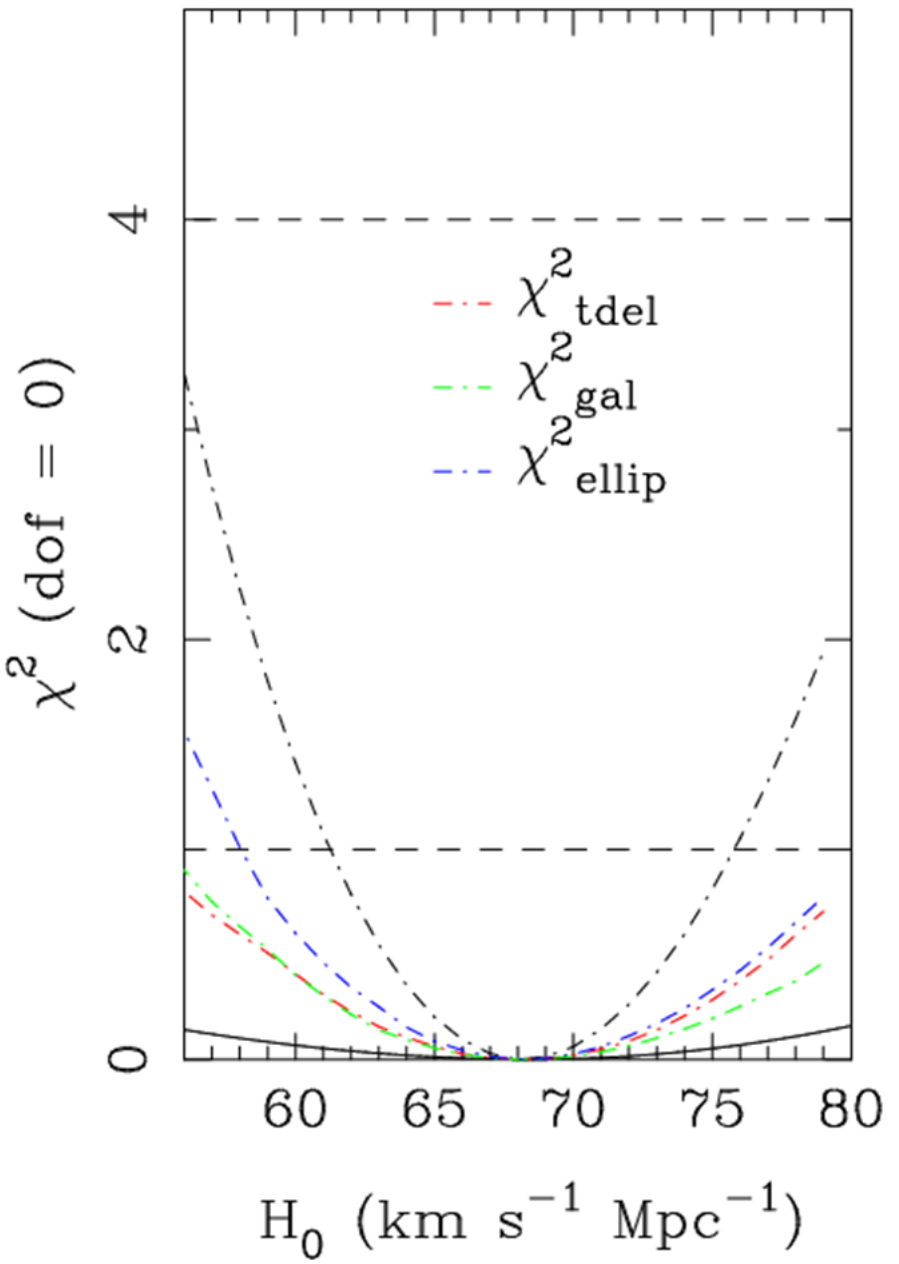


Fig. 2. Estimation of $H_0^{\rm model}$ from the SBS 0909+532 time delay and the astro-photometric constraints of Lehár *et al.*[22] We also used observationally-motivated priors on the ellipticity of the lens galaxy and the external shear. We show the χ^2 curve (black solid line) along with its 1σ and 2σ maximum thresholds (horizontal dashed lines). The black dashed-dotted line corresponds to an "improved" astrometry (see main text), with blue, green and red dashed-dotted lines describing some contributions to the total χ^2.

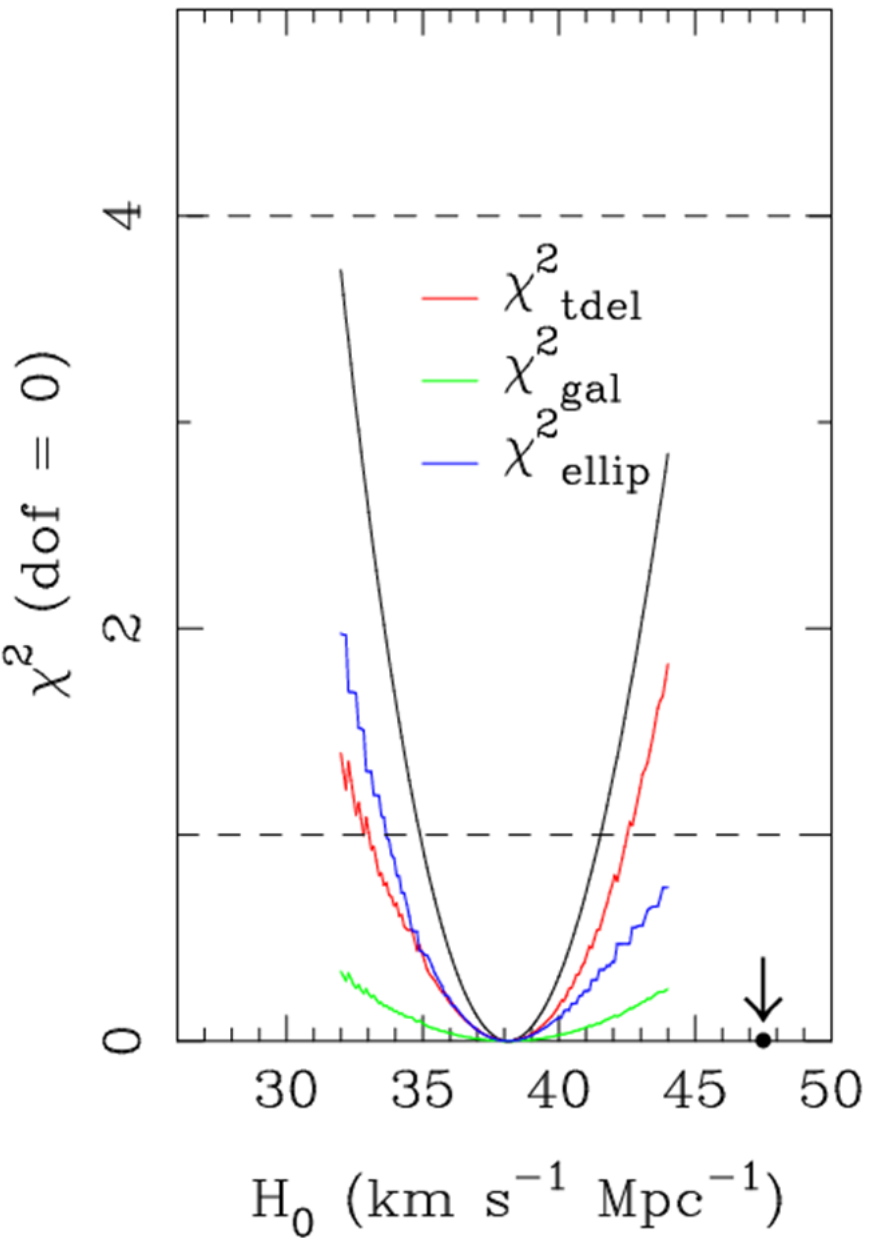


Fig. 3. Estimation of $H_0^{\rm model}$ from the SBS 0909+532 time delay and the astro-photometric constraints of Sluse *et al.*[23] The solid lines are related to priors on the shape of the lens galaxy (ellipticity and position angle; the black solid line represents the total χ^2), while the black dot and vertical arrow indicate the best solution when using priors on the ellipticity and the external shear.

its uncertainty would be dramatically decreased to about 10%. Using "achievable" uncertainties of 1 mas in B and 10 mas in G, we obtained the black dashed-dotted line in Fig. 2 and $H_0^{\rm model} = 68.5 \pm 7.5$.

In addition to new observations of the lens system, a reanalysis of the available *HST* frames of SBS 0909+532 might produce a better astrometry for the system, and thus provide an accurate measure of the Hubble constant. This is a promising task that we and other astronomers are exploring. Sluse *et al.*[23] have reanalysed the available *HST* near-IR frames, obtaining a formally improved astrometry and even details on the structure of G. The error in the relative position of G was only 3 mas; about 30−40 times smaller than the uncertainty derived by Lehár *et al.* We also considered these new constraints to measure $H_0^{\rm model}$. Using the SBS 0909+532 time delay with symmetric error (see above) and the astro-photometric solutions in Table 4 of Sluse *et al.*, along with the ellipticity and position angle of G in Table 3 of Sluse *et al.* (priors on the SIE+γ lens model), we found $H_0^{\rm model} = 38.2 \pm 3.3$ (see Table 2 and the black solid line in Fig. 3). Even the 2σ confidence interval only includes values below 45. Although a moderate increase in the best solution of $H_0^{\rm model}$ is found when taking the previous priors $e \leq 0.5$ and $\gamma \leq 0.1$ (see the black dot and vertical arrow in Fig. 3), the Sluse *et al.*'s relative astrometry leads to best solutions below 50. Hence, either such astrometry is biased or the near-IR fluxes of the quasar images (optical emission) are strongly affected by microlensing in the lens galaxy.[24]

4. SDSS J1339+1310

The gravitational lens system SDSS J1339+1310 was discovered by Inada *et al.*[25] It consists of two quasar images (A and B) at $z_{\rm s} = 2.231$ and an early-type galaxy G at $z_{\rm d} = 0.607$ acting as main deflector.[26] The first set of observational constraints included the relative astrometry of B and G in the last column of Table 1 of Ref. 27, the macrolens magnification ratio from narrow-line/line-core flux ratios and a standard extinction law (based on emission lines in GTC spectra),[26] and an old time delay from LT light curves.[26] We note that the first time delay we used (47.0 ± 5.5 days) is more inaccurate than the updated delay in Table 1. Additionally, we have taken the ellipticity and position angle of G in the last column of Table 1 of Shalyapin et al.[27] as priors on the SIE+γ lens model. The data fit led to an 1σ confidence interval $H_0^{\rm model} = 69^{+10}_{-8}$ (accuracy of $\sim$13%; see Table 2 and the black line in Fig. 4). The observational constraint on the time delay is the primary contribution to the χ^2 curve (see the red line in Fig. 4), while other constraints/priors (e.g., the position of G; see the green line in Fig. 4) play a secondary role.

Results in Fig. 4 suggest that a tighter constraint on the time delay would produce a more accurate determination of the Hubble constant. Therefore, in a second approach, we used the updated time delay with a 4% error that appears in Table 1 to more accurate estimate $H_0^{\rm model}$. The new χ^2 curve in Fig. 5(a) indicates that $H_0^{\rm model} = 67.8 \pm 4.4$ (see also Table 2). This is a quite robust measurement of

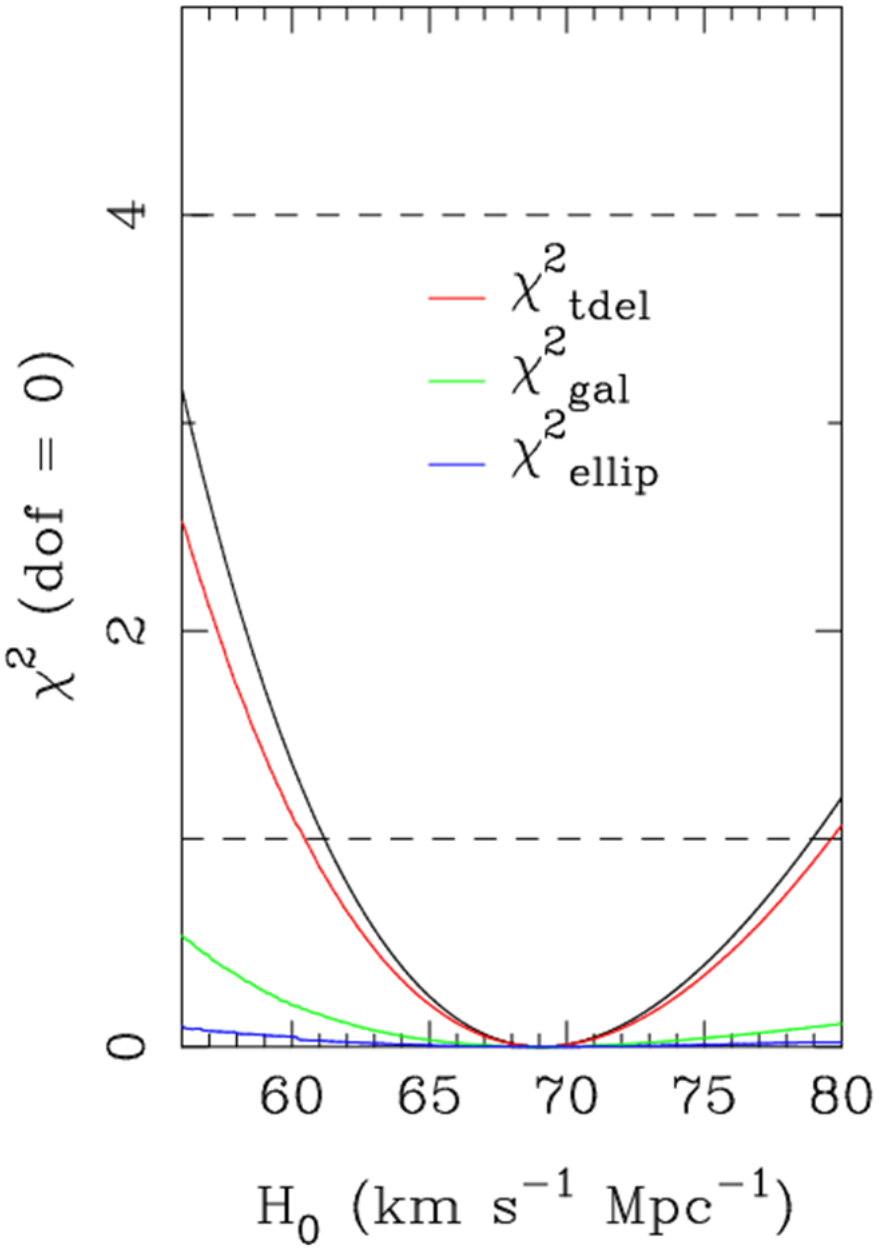


Fig. 4. Estimation of $H_0^{\rm model}$ from an old time delay of SDSS J1339+1310 with an accuracy of $\sim$12% and constraints/priors from results in Refs. 26 and 27. The black line represents the total χ^2, while the blue, green and red lines describe three different contributions to the total curve. The 1σ and 2σ maximum thresholds are also depicted (horizontal dashed lines).

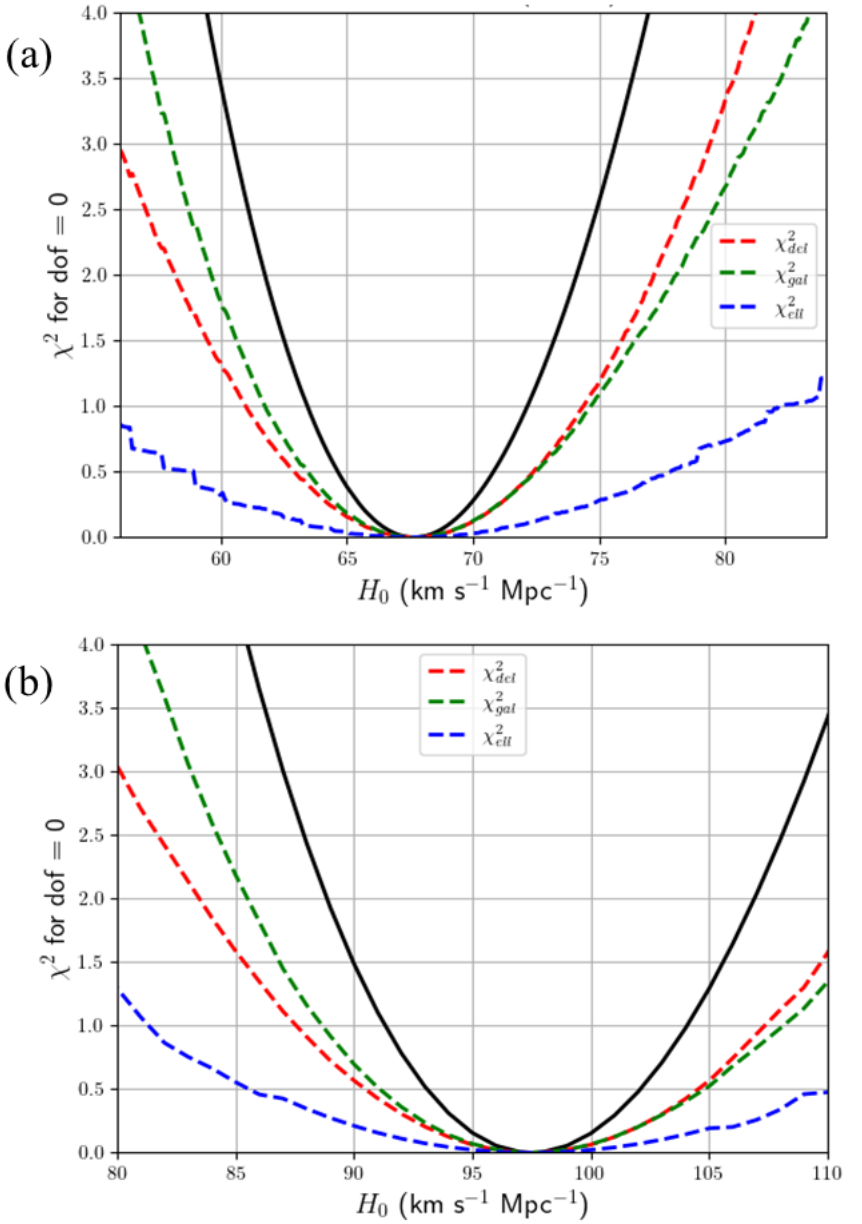


Fig. 5. Estimation of $H_0^{\rm model}$ from the updated time delay of SDSS J1339+1310 with an accuracy of ∼4% and constraints/priors from results in Refs. 26 and 27. (a) SIE+γ lens model. (b) DV+γ lens model. To obtain the χ^2 curve in this bottom panel, we have assumed that light traces mass of the main lens galaxy (see main text).

$H_0^{\rm model}$ because its relative error is small (only 6.5%), and the priors on e and θ_e do not play a relevant role (see the blue dashed line in Fig. 5(a)). In addition, the macrolens magnification ratio is not affected by microlensing/extinction effects, and only one major issue must be addressed: the hypothesis about an isothermal mass distribution for G. Using 58 gravitational lens systems from the SLACS Survey, Koopmans *et al.*[28] concluded that massive early-type galaxies have close to isothermal total density profiles, with a scatter between their logarithmic density slopes below 10%. For a particular lens galaxy, a small deviation from the isothermal power-law index is plausible, and this potential deviation can be taken into account by increasing the error in $H_0^{\rm model}$ (see a more complete discussion in Sec. 5).

It is easy to demonstrate the need for dark matter (e.g., a power-law mass distribution) and the fact that a model in which light traces mass produces biased results. To this end, we again considered the updated time delay in Table 1 and added a new prior, i.e., we worked with three priors instead two. Assuming that light traces mass of G, i.e., a de Vaucouleurs (DV) mass distribution instead a singular isothermal one, in a self-consistent way, the optical structure of G in the last column of Table 1 of Ref. 27 (effective radius, ellipticity and position angle) was used to describe the structure of its mass. This scheme led to a biased $H_0^{\rm model}$ value of about 100 (97.7 ± 6.4; see Fig. 5(b)). Even the 2σ lower limit is above 85.

5. Discussion and future prospects

Using two double quasars of the GLENDAMA sample (see Fig. 1), we focused on the role that some observational constraints and hypotheses/priors on the mass model play in estimating $H_0^{\rm model}$ in a standard cosmology. The main lens galaxies in SBS 0909+532 and SDSS J1339+1310 were modelled with a singular isothermal ellipsoid, in agreement with observations in the Milky Way and SLACS Survey results for massive early-type galaxies acting as gravitational lenses.[28] Adding the external shear γ that is caused by galaxies around a lens system, we initially considered a SIE+γ lens (mass) model.

For SBS 0909+532, there are two different astrometric solutions based on the same *HST* near-IR data. While the Lehár *et al.*'s solution[22] led to a best value of $H_0^{\rm model}$ equal to 68.4 and a broad 1σ interval for this parameter, the Sluse *et al.*'s solution[23] provided an 8.6% measurement of $H_0^{\rm model}$ around a central value of 38.2 (we derived biased results making different choices of priors). Assuming that the time delay and flux ratio between quasar images that we used are right, the last astrometry would be biased. However, the observed near-IR fluxes correspond to optical emission from the quasar accretion disk, so they could be strongly affected by microlenses (stars) in the main lens galaxy.[24] Hence, an accurate and reliable astrometric solution along with a detailed analysis of the macrolens magnification ratio (flux ratio free from extinction and microlensing effects) is required before robustly measuring $H_0^{\rm model}$ for a SIE+γ scenario.

Results for SDSS J1339+1310 are really encouraging because its current astrometry, updated time delay and macrolens magnification ratio through GTC spectra allowed us to accurately measure $H_0^{\rm model}$ (67.8 ± 4.4), with priors on the ellipticity and position angle of the SIE not playing a relevant role. It is also noteworthy that our 1σ interval is not in tension with other recent estimates from GLQs[4] and the CMB[6] (see also Ref. 7), and the central value practically coincides with the upper limit of the *Planck* collaboration. Accounting for a potential microlensing effect on the time delay[29] ($\sim$1 day) would only modify $H_0^{\rm model}$ by $\sim$2%. Additionally, the use of a main galaxy mass model with an unrealistic density profile may have a significant impact on $H_0^{\rm model}$ and be responsible for an error of about 10%.[30–32] At present, we do not know details about the mass density profile of the main deflector in SDSS J1339+1310, and thus we should adopt an uncertainty in $H_0^{\rm model}$ greater than that obtained with a SIE. Very recently, assuming that the deflectors of the H0LiCOW GLQs and the SLACS lenses share the same mass density properties, Birrer *et al.*[33] have obtained $H_0 = 67.4^{+4.1}_{-3.2}$. This new GLQ-based result is in excellent agreement with ours and the CMB-based estimation of H_0, notably reduces tension between early and late-Universe probes, and illustrates the importance of assumptions on mass distributions.

Future time-domain observations of large collections of GLQs will lead to robust constraints on H_0, and the matter and dark energy components of the Universe.[3] The GLENDAMA project includes the first initiative to robotically monitor a small sample of 10 GLQs for about 20 years.[8] This project and the associated robotic monitoring with the LT will end in 2025, after providing accurate time delays for several GLQs and discussing their cosmological implications. In next few years, other ongoing monitoring projects will also measure accurate delays for small/medium samples of GLQs (see the paper by Geoff Chih-Fan Chen in these proceedings), which will contribute to a rich database of tens of measured delays. Despite this optimistic perspective about time-domain results, some issues must be fixed before sheding light on unbiased values of cosmological parameters from such delay database. Deep spectroscopy, high-resolution imaging and other complementary observations will be required. For example, unaccounted mass along GLQ sightlines may produce overestimated/underestimated values of H_0 (see the end of Sec. 2), so accurate H_0 estimates cannot ignore external convergences. Here, although we ignored the external convergence for SDSS J1339+1310, the unaccounted mass is expected to translate to a few percent relative uncertainty in H_0,[33,34] noticeably less than that related to the mass density profile of G (see above).

Acknowledgments

We thank the organizers of Sixteenth Marcel Grossmann Meeting for planning a very interesting event and allowing us to give a talk in the parallel session "Cosmography with Gravitational Lensing". We also thank the chairs of such parallel session for

creating a pleasant environment. We acknowledge Claudio Grillo, Mimoza Hafizi and Sherry Suyu for helpful comments that have significantly contributed to prepare the final text of this contribution. Among other things, the Gravitational LENses and DArk MAtter (GLENDAMA) project aims to construct accurate optical light curves of SBS 0909+532 and SDSS J1339+1310, and measure robust time delays for both systems. Although these optical variability studies mainly rely on observations with the Liverpool Telescope (LT), we are particularly grateful to our collaborators working in several institutions, who provide us with complementary data from the Maidanak Astronomical Observatory and the US Naval Observatory, and participate actively in the project development. The LT is operated on the island of La Palma by the Liverpool John Moores University (with financial support from the UK Science and Technology Facilities Council), in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. We thank the staff of the telescope for a kind interaction before, during and after the observations. This research has been supported by the grant AYA2017-89815-P funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", and the grant PID2020-118990GB-I00 funded by MCIN/AEI/10.13039/501100011033.

References

1. P. Schneider, C. S. Kochanek and J. Wambsganss, *Gravitational Lensing: Strong, Weak and Micro* (Springer-Verlag, Berlin, 2006).
2. N. Jackson, The Hubble constant, *Living Rev. Relativ.* **18**, 2 (2015), `https://doi.org/10.1007/lrr-2015-2`.
3. T. Treu and P. J. Marshall, Time delay cosmography, *Astron. Astrophys. Rev.* **24**, 11 (2016).
4. K. C. Wong, S. H. Suyu, G. C.-F. Chen *et al.*, H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes, *Mon. Notices Royal Astron. Soc.* **498**, 1420 (2020).
5. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM, *Astrophys. J.* **876**, 85 (2019).
6. Planck Collaboration, *Planck* 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, 6 (2020).
7. W. L. Freedman, B. F. Madore, D. Hatt *et al.*, The Carnegie-Chicago Hubble program. VIII. An independent determination of the Hubble constant based on the tip of the red giant branch, *Astrophys. J.* **882**, 34 (2019).
8. R. Gil-Merino, L. J. Goicoechea, V. N. Shalyapin and A. Oscoz, New database for a sample of optically bright lensed quasars in the northern hemisphere, *Astron. Astrophys.* **616**, 118 (2018).
9. L. J. Hainline, C. W. Morgan, C. L. MacLeod *et al.*, Time delay and accretion disk size measurements in the lensed quasar SBS 0909+532 from multiwavelength microlensing analysis, *Astrophys. J.* **774**, 69 (2013).
10. V. N. Shalyapin, L. J. Goicoechea and R. Gil-Merino, A 5.5-year robotic optical monitoring of Q0957+561: substructure in a non-local cD galaxy, *Astron. Astrophys.* **540**, 132 (2012).

11. V. N. Shalyapin, L. J. Goicoechea, C. W. Morgan, M. A. Cornachione and A. V. Sergeyev, Resolving the inner accretion flow towards the central supermassive black hole in SDSS J1339+1310, *Astron. Astrophys.* **646**, 165 (2021).
12. L. J. Goicoechea and V. N. Shalyapin, Time delays in the gravitationally lensed quasar H1413+117 (Cloverleaf), *Astrophys. J.* **708**, 995 (2010).
13. V. N. Shalyapin and L. J. Goicoechea, Gravitationally lensed quasar SDSS J1442+4055: redshifts of lensing galaxies, time delay, microlensing variability, and intervening metal system at z $\sim$ 2, *Astrophys. J.* **873**, 117 (2019).
14. V. N. Shalyapin and L. J. Goicoechea, Doubly imaged quasar SDSS J1515+1511: Time delay and lensing galaxies, *Astrophys. J.* **836**, 14 (2017).
15. S. H. Suyu, T. Treu, S. Hilbert *et al.*, Cosmology from gravitational lens time delays and *Planck* data, *Astrophys. J.* **788**, L35 (2014).
16. M. Millon, A. Galan, F. Courbin *et al.*, TDCOSMO. I. An exploration of systematic uncertainties in the inference of H0 from time-delay cosmography, *Astron. Astrophys.* **639**, 101 (2020).
17. D. Paraficz and J. Hjorth, Gravitational lenses as cosmic rulers: $\Omega_{\rm m}$, Ω_Λ from time delays and velocity dispersions, *Astron. Astrophys.* **507**, L49 (2009).
18. C. R. Keeton, Computational methods for gravitational lensing, `https://arxiv.org/pdf/astro-ph/0102340.pdf` (2001).
19. C. S. Kochanek, E. E. Falco, R. Schild *et al.*, SBS 0909+532: A new double gravitational lens or binary quasar?, *Astrophys. J.* **479**, 678 (1997).
20. A. Oscoz, M. Serra-Ricart, E. Mediavilla, J. Buitrago and L. J. Goicoechea, Support for the gravitational lens interpretation of SBS 0909+532, *Astrophys. J.* **491**, L7 (1997).
21. L. M. Lubin, C. D. Fassnacht, A. C. S. Readhead, R. D. Blandford and T. Kundić, A Keck survey of gravitational lens systems. I. Spectroscopy of SBS 0909+532, HST 1411+5211, and CLASS B2319+051, *Astron. J.* **119**, 451 (2000).
22. J. Lehár, E. E. Falco, C. S. Kochanek *et al.*, *Hubble Space Telescope* observations of 10 two-image gravitational lenses, *Astrophys. J.* **536**, 584 (2000).
23. D. Sluse, V. Chantry, P. Magain, F. Courbin and G. Meylan, COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses. X. Modeling based on high-precision astrometry of a sample of 25 lensed quasars: consequences for ellipticity, shear, and astrometric anomalies, *Astron. Astrophys.* **538**, 99 (2012).
24. E. Mediavilla, J. A. Muñoz, C. S. Kochanek *et al.*, The first precise determination of an optical-far-ultraviolet extinction curve beyond the Local Group ($z = 0.83$), *Astrophys. J.* **619**, 749 (2005).
25. N. Inada, M. Oguri, M. Shin *et al.*, Five new high-redshift quasar lenses from the Sloan Digital Sky Survey, *Astron. J.* **137**, 4118 (2009).
26. L. J. Goicoechea and V. N. Shalyapin, Gravitational lens system SDSS J1339+1310: microlensing factory and time delay, *Astron. Astrophys.* **596**, 77 (2016).
27. V. N. Shalyapin and L. J. Goicoechea, Deep optical imaging and spectroscopy of the lens system SDSS J1339+1310, *Astron. Astrophys.* **568**, 116 (2014).
28. L. V. E. Koopmans, A. Bolton, T. Treu *et al.*, The structure and dynamics of massive early-type galaxies: on homology, isothermality, and isotropy inside one effective radius, *Astrophys. J.* **703**, L51 (2009).
29. S. S. Tie and C. S. Kochanek, Microlensing makes lensed quasar time delays significantly time variable, *Mon. Notices Royal Astron. Soc.* **473**, 80 (2018).
30. P. Schneider and D. Sluse, Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing, *Astron. Astrophys.* **559**, 37 (2013).

31. C. S. Kochanek, Overconstrained gravitational lens models and the Hubble constant, *Mon. Notices Royal Astron. Soc.* **493**, 1725 (2020).
32. H. R. Stacey, C. M. O'Riordan, S. Vegetti *et al.*, The inner density profile of an elliptical galaxy at $z = 1.15$ from gravitational lensing, submitted to *Mon. Notices Royal Astron. Soc.*, `https://arxiv.org/pdf/2109.10929.pdf` (2021).
33. S. Birrer, A. J. Shajib, A. Galan *et al.*, TDCOSMO IV: Hierarchical time-delay cosmography - joint inference of the Hubble constant and galaxy density profiles, *Astron. Astrophys.* **643**, 165 (2020).
34. C. E. Rusu, C. D. Fassnacht, D. Sluse *et al.*, H0LiCOW – III. Quantifying the effect of mass along the line of sight to the gravitational lens HE 0435−1223 through weighted galaxy counts, *Mon. Notices Royal Astron. Soc.* **467**, 4220 (2017).

Dark energy and dark matter unification from dynamical space time: BBN constraints

D. Benisty

DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

E-mail: benidav@post.bgu.ac.il

A unification of dark matter and dark energy based on a dynamical space time theory is suggested. By introducing a dynamical space time vector field χ_μ as a Lagrange multiplier, a conservation of an energy momentum tensor $T^{\mu\nu}_{(\chi)}$ is implemented. This Lagrangian generalizes the Unified dark energy and dark matter from a scalar field different from quintessence which did not consider a Lagrangian formulation. This generalization allows the solutions which were found previously, but in addition to that also non singular bouncing solutions that rapidly approach to the ΛCDM model. The BBN constraint is also studied.

Keywords: Dark Matter; Dark Energy

1. Introduction

Dark energy and Dark matter constitute most of the observable Universe. Yet the true nature of these two phenomena is still a mystery. One fundamental question with respect to those phenomena is the coincidence problem which is trying to explain the relation between dark energy and dark matter densities. In order to solve this problem, one approach claims that the dark energy is a dynamical entity and hope to exploit solutions of scaling or tracking type to remove dependence on initial conditions. Others left this principle and tried to model the dark energy as a phenomenological fluid which exhibits a particular relation with the scale factor,[1] Hubble constant[2] or even even the cosmic time itself.[3]

Unifications between dark energy and dark matter from an action principle were obtained from K-essence type actions,[4] or by introducing a complex scalar field.[5] Beyond those approaches, a unified description of Dark Energy and Dark Matter using a new measure of integration has been formulated.[6,7] Also a diffusive interacting of dark energy and dark matter models was introduced in[8,9] and it has been found that diffusive interacting dark energy — dark matter models can be formulated in the context of an action principle based on a generalization of those Two Measures Theories in the context of quintessential scalar fields,[10,11] although these models are not equivalent to the previous diffusive interacting dark energy — dark matter models.[8,9]

A model for a unification of dark energy and dark matter from a single scalar field ϕ, was suggested by Gao, Kunz, Liddle and Parkinson.[12] Their model is close

to traditional quintessence, and gives dynamical dark energy and dark matter, but introduces a modification of the equations of motion of the scalar field that apparently are impossible to formulate in the framework of an action principle. The basic stress energy tensor which was considered in addition to Einstein equation was:

$$T^{\mu\nu} = -\frac{1}{2}\phi^{,\mu}\phi^{,\nu} + U(\phi)g^{\mu\nu} \tag{1}$$

where ϕ is a scalar field and $U(\phi)$ is the potential for that scalar. Assuming homogeneous and isotropic behavior the scalar field should be only time dependent $\phi = \phi(t)$. Then the kinetic term $-\frac{1}{2}\phi^{,\mu}\phi^{,\nu}$ is parameterzing the dark matter because it contains only energy density with no pressure and $U(\phi)g^{\mu\nu}$ is parameterzing the dark energy. The basic requirement for this stress energy tensor is it's conservation law $\nabla_\mu T^{\mu\nu} = 0$. By assuming a constant potential $U(\phi) = \mathbf{Const}$ the model provides from the potential the traditional cosmological constant and the kinetic term of the scalar field is shown to provide, from the conservation law of the energy momentum tensor, that the kinetic term dependence has a dust like behavior.

$$-\frac{1}{2}\nabla_\mu(\phi^{,\mu}\phi^{,\nu}) = 0 \quad \Rightarrow \quad \dot{\phi}^2 \sim \frac{1}{a^3} \tag{2}$$

This simple case refers to the classical ΛCDM model. The special advantage of this model is a unification of dark energy and dark matter from one scalar field and has an interesting possibility for exploring the coincidence problem.

The lack of an action principle for this model brought us to reformulate the unification between dark energy and dark matter idea put forward by Gao, Kunz, Liddle and Parkinson[12] in the framework of a Dynamical Space Time Theory[13,14] which forces a conservation of energy momentum tensor in addition to the covariant conservation of the stress energy momentum tensor that appears in Einstein equation. In the next chapter we explore the equations of motion for these theories. In the third chapter we solve analytically the theory for constant potentials which reproduce the ΛCDM model with a bounce, which gives a possibility to solve the initial big bang singularity. In the last chapter we solve the theory for an exponential potential which gives a good possibility for solving the coincidence problem.

2. Dynamical Space Time Theory

2.1. *A basic formulation*

One from the basic fatures in the standard approach to theories of gravity is the local conservation of an energy momentum tensor. In the field theory case it's derived as a result rather than a starting point. For example, the conservation of energy can be derived from the time translation invariance principle. The local conservation of an energy momentum tensor can be a starting point rather than a derived result. Let's consider a 4 dimensional case where a conservation of a symmetric energy

momentum tensor $T^{\mu\nu}_{(\chi)}$ is imposed by introducing the term in the action:

$$\mathcal{S}_{(\chi)} = \int d^4x\sqrt{-g}\chi_{\mu;\nu}T^{\mu\nu}_{(\chi)} \tag{3}$$

where $\chi_{\mu;\nu} = \partial_\nu\chi_\mu - \Gamma^\lambda_{\mu\nu}\chi_\lambda$. The vector field χ_μ called a dynamical space time vector, because of the energy density of $T^{\mu\nu}_{(\chi)}$ is a canonically conjugated variable to χ_0, which is what we expected from a dynamical time:

$$\pi_{\chi_0} = \frac{\partial\mathcal{L}}{\partial\dot{\chi}^0} = T^0_0(\chi) \tag{4}$$

If $T^{\mu\nu}_{(\chi)}$ is being independent of χ_μ and having $\Gamma^\lambda_{\mu\nu}$ being defined as the Christoffel connection coefficients (the second order Formalism), then the variation with respect to χ_μ gives a covariant conservation law:

$$\nabla_\mu T^{\mu\nu}_{(\chi)} = 0 \tag{5}$$

From the variation of the action with respect to the metric, we get a conserved stress energy tensor $G^{\mu\nu}$ (in appropriate units), which is well known from Einstein equation:

$$G^{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta\sqrt{-g}}{\delta g^{\mu\nu}}[\mathcal{L}_\chi + \mathcal{L}_m]\,, \quad \nabla_\mu G^{\mu\nu} = 0 \tag{6}$$

where $G^{\mu\nu}$ is Einstein tensor, $\mathcal{L}_\chi$ is the Lagrangian in (3) and $\mathcal{L}_m$ is an optional action that involve other contributions.

Some basic symmetries that holds for the dynamical space time theory are two independent shift symmetries:

$$\chi_\mu \to \chi_\mu + k_\mu\,, \quad T^{\mu\nu}_{(\chi)} \to T^{\mu\nu}_{(\chi)} + \Lambda g^{\mu\nu} \tag{7}$$

where Λ is some arbitrary constant and k_μ is a Killing vector of the solution. This transformation will not change the equations of motions, which means also that the process of redefinition of the energy momentum tensor in the action (3) will not change the equations of motion. Of course such type of redefinition of the energy momentum tensor is exactly what is done in the process of normal ordering in Quantum Field Theory for instance.

2.2. *A connection to modified measures*

A particular case of the stress energy tensor with the form $T^{\mu\nu}_{(\chi)} = \mathcal{L}_1 g^{\mu\nu}$ corresponds to a modified measure theory. By substituting this stress energy tensor into the action itself, the determinant of the metric is canceled:

$$\sqrt{-g}\chi^\mu_{;\mu}\mathcal{L}_1 = \partial_\mu(\sqrt{-g}\chi^\mu)\mathcal{L}_1 = \Phi\mathcal{L}_1 \tag{8}$$

where $\Phi = \partial_\mu(\sqrt{-g}\chi^\mu)$ is like a "modified measure". A variation with respect to the dynamical time vector field will give a constraint on $\mathcal{L}_1$ to be a constant:

$$\partial_\alpha\mathcal{L}_1 = 0 \quad \Rightarrow \quad \mathcal{L}_1 = M = Const \tag{9}$$

This situation corresponds to the "Two Measures Theory"[15,16] where in addition to the regular measure of integration in the action $\sqrt{-g}$ includes another measure of integration which is also a density and a total derivative. Notable effects that can be obtained in this way are the spontaneous breaking of the scale invariance, the see saw cosmological effects,[15] the resolution of the 5th force problem in quintessential cosmology[17] and a unified picture of both inflation and of slowly accelerated expansion of the present universe.[18,19] As we mentioned before in the introduction Two Measure Theory can serve to build unified models of dark energy and dark matter.

Usually the construction of this measure is from 4 scalar fields φ_a, where $a = 1, 2, 3, 4$.

$$\Phi = \frac{1}{4!}\varepsilon^{\alpha\beta\gamma\delta}\varepsilon_{abcd}\partial_\alpha\varphi^{(a)}\partial_\beta\varphi^{(b)}\partial_\gamma\varphi^{(c)}\partial_\delta\varphi^{(d)} \tag{10}$$

and then we can rewrite an action that uses both of these densities:

$$S = \int d^4x\Phi\mathcal{L}_1 + \int d^4x\sqrt{-g}\mathcal{L}_2\,. \tag{11}$$

As a consequence of the variation with respect to the scalar fields φ_a, assuming that $\mathcal{L}_1$ and $\mathcal{L}_2$ are independent of the scalar fields φ_a, we obtain that for $\Phi \neq 0$ it implies that $\mathcal{L}_1 = M = const$ as in the dynamical time theory with the case of (9).

3. DE-DM Unified Theory from Dynamical Space-Time

A suggestion of an action which can produce DE-DM unification takes the form:

$$\mathcal{L} = -\frac{1}{2}R + \chi_{\mu;\nu}T^{\mu\nu}_{(\chi)} - \frac{1}{2}g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - V(\phi) \tag{12}$$

Consisting of an Einstein Hilbert action ($8\pi G = 1$), quintessence and Dynamical space-time action, when the original stress energy tensor $T^{\mu\nu}_{(\chi)}$ is the same as the stress energy tensor (1) Gao and colleagues used:

$$T^{\mu\nu}_{(\chi)} = -\frac{1}{2}\phi^{,\mu}\phi^{,\nu} + U(\phi)g^{\mu\nu} \tag{13}$$

The action depends on three different variables: the scalar field ϕ, the dynamical space time vector χ_μ and the metric $g_{\mu\nu}$. Therefore there are 3 sets in for the equation of motions. For the solution we assume homogeneity and isotropy, therefore we solve our theory with a FLRW metric:

$$ds^2 = -dt^2 + a(t)^2\left(\frac{dr^2}{1-Kr^2} + r^2d\Omega^2\right) \tag{14}$$

According to this ansatz, the scalar field is just a function of time $\phi(t)$ and the dynamical vector field will be taken only with a time component $\chi_\mu = (\chi_0, 0, 0, 0)$, where χ_0 is also just a function of time. A variation with respect to the dynamical space time vector field χ_μ will force a conservation of the original stress energy tensor, which in FRWM gives the relation:

$$\ddot{\phi} + \frac{3}{2}\mathcal{H}\dot{\phi} + U'(\phi) = 0 \tag{15}$$

Compared with the equivalent equation which comes from quintessence model, this model gives a different and smaller friction term, as compared to the canonical scalar field. Therefore for increasing redshift, the densities for the scalar field will increase slower than in the standard quintessence.

The second variation with respect to the scalar field ϕ gives a non-conserved current:

$$\chi^{\lambda}_{;\lambda} U'(\phi) - V'(\phi) = \nabla_{\mu} j^{\mu} \tag{16a}$$

$$j^{\mu} = \frac{1}{2}\phi_{,\nu}(\chi^{\mu;\nu} + \chi^{\nu;\mu}) + \phi^{,\mu} \tag{16b}$$

and the derivatives of the potentials are the source of this current. For constant potentials the source becomes zero, and we get a covariant conservation of this current. In a FLRW metric this equation of motion takes the form:

$$\ddot{\phi}(\dot{\chi}_0 - 1) + \dot{\phi}[\ddot{\chi}_0 + 3\mathcal{H}(\dot{\chi}_0 - 1)] = U'(\phi)(\dot{\chi}_0 + 3\mathcal{H}\chi_0) - V'(\phi) \tag{17}$$

Substituting the term of the potential derivative $U'(\phi)$ from equation (15):

$$[1 - 2\dot{\chi}_0 - 3\mathcal{H}\chi_0]\ddot{\phi} - [\ddot{\chi}_0 - 3\mathcal{H} + \frac{9}{2}\mathcal{H}(\dot{\chi}_0 + \chi_0\mathcal{H})]\dot{\phi} + V'(\phi) = 0 \tag{18}$$

The last variation, with respect to the metric, gives the stress energy tensor that is defined by the value of the Einstein's tensor. For a spatially homogeneous, cosmological case, the energy density and the pressure of the scalar field are:

$$\rho = \dot{\phi}^2(\dot{\chi}_0(1 - \frac{3}{2}\mathcal{H}) - \frac{1}{2}) + V(\phi) - \dot{\phi}\dot{\chi}_0(U'(\phi) + \ddot{\phi}) \tag{19a}$$

$$p = \frac{1}{2}\dot{\phi}^2(\dot{\chi}_0 - 1) - V(\phi) - \chi_0\dot{\phi}U'(\phi) \tag{19b}$$

Substituting the potential derivative $U'(\phi)$ from Eq. (15) into the energy density term, makes the equation simpler:

$$\rho = (\dot{\chi}_0 - \frac{1}{2})\dot{\phi}^2 + V(\phi) \tag{20}$$

which has no longer dependence on the potential $U(\phi)$ or it's derivatives. Those three variations are sufficient for building a complete solution for the theory. Let's see a few simple cases.

4. The Evolution of the Homogeneous Solutions

4.1. *A bouncing ΛCDM solution*

In order to compute the evolution of the scalar field and to check whether it is compatible with observable universe, we have to specify a form for the potentials. Let's take a simplified case of constant potentials:

$$U(\phi) = C, \quad V(\phi) = \Omega_{\Lambda} \tag{21}$$

Overall, in the equations of motions only the derivative the potential $U(\phi)$ appears, not the potential itself. Therefore a constant part of the potential $U(\phi)$ does not contribute to the solution. However $V(\phi)$, as we shall see below, gives the cosmological constant. The conservation of the stress energy tensor from equation (15) gives:

$$\dot{\phi}^2 = \frac{2\Omega_m}{a^3} \tag{22}$$

where Ω_m is an integration constant which appears from the solution. From the second variation, with respect to the scalar field ϕ, a conserved current is obtained, which from equation (18) gives the exact solution of the dynamical time vector field:

$$\dot{\chi}_0 = 1 - \kappa\, a^{-1.5} \tag{23}$$

where κ is another integration of constant. Eventually, the densities and the pressure for this potentials are given by (20). By substituting the solutions for the scalar $\dot{\phi}$ and the vector $\dot{\chi}_0$ (in units with $\rho_c = \frac{8\pi G}{3H_0^2} = 1$) we get:

$$\rho = \Omega_\Lambda - \frac{\Omega_\kappa}{a^{4.5}} + \frac{\Omega_m}{a^3} \tag{24a}$$

$$p = -\Omega_\Lambda - \frac{1}{2}\frac{\Omega_\kappa}{a^{4.5}} \tag{24b}$$

where $\Omega_\kappa = \kappa\Omega_m$. Notice that Ω_m, Ω_κ are integration constants the solution contains and Ω_Λ is parameter from the action of the theory. We can separate the result into three different "dark fluids": dark energy ($\omega = -1$), dark matter ($\omega = 0$) and an exotic part ($\omega = \frac{1}{2}$), which is the responsible for the bounce (for $\kappa > 0$). From Eq. (22) the solution produces a positive Ω_m since it's proportional to $\dot{\phi}^2$. For Ω_Λ the measurements for the late universe forces the choice of this parameter to be positive. However for another solutions (in the context of Anti de-Sitter space, for instance) this parameter could be negative from the beginning.

In addition to those solutions, there is a strong correspondence between the zero component of the dynamical space time vector field and the cosmic time. For ΛCDM there is no bouncing solution $\kappa = 0$ and therefore from equation (23) we get $\chi_0 = t$ that implies that the dynamical time is exactly the cosmic time. For bouncing ΛCDM we obtain a relation between the dynamical and the cosmic time with some delay between the dynamical time and the cosmic time for the early universe (in the bouncing region). For the late universe the dynamical time returns back to run as fast as the cosmic time again. This relation between the dynamical and the cosmic time may have interesting application in the solution to "the problem of time" in quantum cosmology which will discussed elsewhere. Notice that the dynamical time is a field variable while the cosmic time is a coordinate.

In order to constrain our model, we use the Big Bang Nucleosynthesis (BBN) speed-up factor, defined as the ratio of the expansion rate predicted in a given model versus that of the ΛCDM model at the BBN epoch ($z_{BBN} \sim 10^9$). This amounts

to the limit[20] with $(\Delta H/H_{\Lambda CDM})^2 < 10\%$, where $H_{\Lambda CDM}$ is the Hubble rate for the ΛCDM model and the ΔH is the difference between the Dynamical Spacetime Hubble rate and the ΛCDM Hubble rate, $\Delta H = H - H_{\Lambda CDM}$. The bounds gives:

$$\Omega_\kappa < 6 \cdot 10^{-10}. \tag{25}$$

With this bound the model is consistent with the cosmological observations.[21]

5. Discussion and Future Work

In this paper the "unified dark energy and dark matter from a scalar field different from quintessence" is formulated through an action principle. Introducing the coupling of a dynamical space time vector field to an energy momentum tensor that appears in the action, determines the equation of motion of the scalar field from the variation of the dynamical space time vector field or effectively from the conservation law of an energy momentum tensor, as in.[12] The energy momentum tensor that is introduced in the action is related but not in general the same as the one that appears in the right hand side of the gravitational equations, as opposed to the non Lagrangian approach of,[12] so our approach and that of[12] are not equivalent. However in many situations the solutions studied in[12] can be also obtained here, but there are other solutions, in special non singular bounce solutions which are not present in.[12]

In those simple solutions, the dynamical time behaves very close to the cosmic time. In particular in solutions which are exactly ΛCDM, the cosmic time and the dynamical time exactly coincide with each other. If there is a bounce, the deviation of the dynamical time with respect to the cosmic time takes place only very close to the bounce region. The use of this dynamical time as the time in the Wheeler de-Witt equation should also a subject of interest.

In principle we can introduce two different scalar potentials: one coupled directly to $\sqrt{-g}$ and the other appearing in the original stress energy tensor $T^{\mu\nu}_{(\chi)}$. So far, for the purposes of starting the study of the theory, we have only introduced a scalar potential coupled directly to $\sqrt{-g}$ and shown that this already leads to an interacting Dark Energy — Dark Matter model, although the full possibilities of the theory will be revealed when the two independent potentials will be introduced.

Finally, another direction for research has been started by studying models of this type in the context of higher dimensional theories, where they can provide a useful framework to study the "inflation compactication" epoch and an exit from this era to the present LCDM epoch could be further explored.[22]

Acknowledgments

D.B gratefully acknowledge the support the supports of the Blavatnik and the Rothschild fellowships. We have received partial support from European COST actions CA15117 and CA18108.

References

1. V. F. Cardone, A. Troisi and S. Capozziello, Unified dark energy models: A Phenomenological approach, *Phys. Rev. D* **69**, p. 083517 (2004).
2. G. Dvali and M. S. Turner, Dark energy as a modification of the Friedmann equation (1 2003).
3. S. Basilakos, Cosmological implications and structure formation from a time varying vacuum, *Mon. Not. Roy. Astron. Soc.* **395**, p. 2347 (2009).
4. R. J. Scherrer, Purely kinetic k-essence as unified dark matter, *Phys. Rev. Lett.* **93**, p. 011301 (2004).
5. A. Arbey, Dark fluid: A Complex scalar field to unify dark energy and dark matter, *Phys. Rev. D* **74**, p. 043516 (2006).
6. E. Guendelman, E. Nissimov and S. Pacheva, Unified Dark Energy and Dust Dark Matter Dual to Quadratic Purely Kinetic K-Essence, *Eur. Phys. J. C* **76**, p. 90 (2016).
7. E. Guendelman, E. Nissimov and S. Pacheva, Quintessential Inflation, Unified Dark Energy and Dark Matter, and Higgs Mechanism, *Bulg. J. Phys.* **44**, 015 (2017).
8. G. Koutsoumbas, K. Ntrekis, E. Papantonopoulos and E. N. Saridakis, Unification of Dark Matter - Dark Energy in Generalized Galileon Theories, *JCAP* **02**, p. 003 (2018).
9. Z. Haba, A. Stachowski and M. Szydłowski, Dynamics of the diffusive DM-DE interaction – Dynamical system approach, *JCAP* **07**, p. 024 (2016).
10. D. Benisty and E. I. Guendelman, Interacting Diffusive Unified Dark Energy and Dark Matter from Scalar Fields, *Eur. Phys. J. C* **77**, p. 396 (2017).
11. D. Benisty and E. I. Guendelman, Unified DE–DM with diffusive interactions scenario from scalar fields, *Int. J. Mod. Phys. D* **26**, p. 1743021 (2017).
12. C. Gao, M. Kunz, A. R. Liddle and D. Parkinson, Unified dark energy and dark matter from a scalar field different from quintessence, *Phys. Rev. D* **81**, p. 043520 (2010).
13. E. I. Guendelman, Gravitational Theory with a Dynamical Time, *Int. J. Mod. Phys. A* **25**, 4081 (2010).
14. D. Benisty and E. I. Guendelman, Radiation Like Scalar Field and Gauge Fields in Cosmology for a theory with Dynamical Time, *Mod. Phys. Lett. A* **31**, p. 1650188 (2016).
15. E. I. Guendelman, Scale invariance, new inflation and decaying lambda terms, *Mod. Phys. Lett. A* **14**, 1043 (1999).
16. E. I. Guendelman and A. B. Kaganovich, Transition to Zero Cosmological Constant and Phantom Dark Energy as Solutions Involving Change of Orientation of Space-Time Manifold, *Class. Quant. Grav.* **25**, p. 235015 (2008).
17. E. I. Guendelman and A. B. Kaganovich, Absence of the Fifth Force Problem in a Model with Spontaneously Broken Dilatation Symmetry, *Annals Phys.* **323**, 866 (2008).
18. E. I. Guendelman and O. Katz, Inflation and transition to a slowly accelerating phase from SSB of scale invariance, *Class. Quant. Grav.* **20**, 1715 (2003).
19. E. Guendelman, R. Herrera, P. Labrana, E. Nissimov and S. Pacheva, Emergent Cosmology, Inflation and Dark Energy, *Gen. Rel. Grav.* **47**, p. 10 (2015).
20. J.-P. Uzan, Varying Constants, Gravitation and Cosmology, *Living Rev. Rel.* **14**, p. 2 (2011).
21. F. K. Anagnostopoulos, D. Benisty, S. Basilakos and E. I. Guendelman, Dark energy and dark matter unification from dynamical space time: observational constraints and cosmological implications, *JCAP* **06**, p. 003 (2019).
22. D. Benisty and E. I. Guendelman, Inflation compactification from dynamical spacetime, *Phys. Rev. D* **98**, p. 043522 (2018).

Entropy and irreversible processes in gravity and cosmology

Llorenç Espinosa-Portalés* and Juan García-Bellido†
Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid
Cantoblanco, 28049 Madrid, Spain
**E-mail: llorenc.espinosa@uam.es*
†*E-mail: juan.garciabellido@uam.es*

General Relativity is *a priori* a theory invariant under time reversal. Its integration with the laws of thermodynamics allows for a formulation of non-equilibrium phenomena in gravity and the introduction of an arrow of time, i.e. the the breaking of such invariance. Even though most of the evolution of the universe takes place in local thermal equilibrium, the effects of irreversible processes on the expansion via entropic forces may be phenomenologically relevant. We review our previous work on the covariant formulation of non-equilibrium thermodynamics in General Relativity and the proposal to explain the recent cosmic acceleration from it.

Keywords: General Relativity, thermodynamics, non-equilibrium phenomena, entropic forces, cosmic acceleration.

1. Introduction

General Relativity is an extremely successful physical theory. More than a century after its formulation, its predictions continue to be valid at all probed scales, albeit an extension at short scales will be required in order to obtain a UV-complete quantum theory of gravity and resolve space-time singularities.

Thermodynamics is an even older discipline. Its fundamental laws seem to resist the pass of time and are still of great relevance today. On the one hand, thermodynamics may help in building the bridge between classical and quantum gravity, as the laws of black hole thermodynamics points towards the existence of a microphysical description of gravity yet to be understood.

On the other hand, the second law of thermodynamics, i.e. the growth of entropy, dictates the sign of the arrow of time. Physical laws are usually invariant under time inversion. The increase in entropy with time in out-of-equilibrium phenomena, however, allows one to distinguish the future-directed from the past directed description of a physical process.

There lacks a consistent and rigorous integration between General Relativity and the laws of thermodynamics. Understanding the very notion of the arrow of time is of particular interest for cosmology. In section 2 we argue for the need of going beyond reversible cosmology. This can be achieved for any space-time metric using variational techniques.[1] We review our main results in this new approach to non-equilibrium thermodynamics in General Relativity and present them in sections 3 to 6.

The growth of entropy associated to the causal horizon in open inflation scenarios may explain the current accelerated expansion of the universe within the general relativistic entropic acceleration (GREA) theory.[2] We briefly describe how this mechanism works in section 7 and finish with our conclusions.

2. Reversible cosmology

Let us begin the study of the problem of reversibility in gravity and cosmology by reviewing a prototypic case of reversible gravitational system: a homogeneous and isotropic universe. It is described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2 \right) , \tag{1}$$

where $a(t)$ is the scale factor, $k = -1, 0, 1$ is the curvature parameter corresponding to, respectively, an open, flat and closed universe; and $d\Omega_2^2$ is the solid angle element. This space-time is filled with a perfect fluid, described by the stress-energy tensor

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p g_{\mu\nu} , \tag{2}$$

where ρ and p are, respectively, the density and pressure of the fluid. The Einstein field equations for this metric and matter content deliver the dynamics for the scale factor, the well-known Friedmann equations

$$H^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) , \tag{3}$$

where $H = \dot{a}/a$ is the Hubble parameter. There is a constraint on the stress-energy tensor due to the Einstein field equations and the Bianchi identities, namely its covariant conservation $D_\mu T^{\mu\nu} = 0$. From this constraint one can derive the continuity equation

$$\dot{\rho} + 3H(\rho + p) = 0 . \tag{4}$$

However, one can also derive this equation from the second law of thermodynamics. Indeed, changes in entropy are related to changes in internal energy and work $TdS = \delta U + \delta W$. If we apply this to a region of fixed comoving volume $a(t)^3$ we get

$$T\frac{dS}{dt} = \frac{d}{dt}\left(\rho a^3\right) + p\frac{d}{dt}\left(a^3\right) . \tag{5}$$

If the expansion of the universe is reversible, we can set the LHS to 0 and recover the continuity equation. However, this is only true in thermodynamical equilibrium and, in general, entropy is a monotonically increasing function of time. Most of the expansion history of the universe is indeed adiabatic. However, it is out-of-equilibrium at certain key points such as (p)reheating, phase transitions or gravitational collapse. Allowing for a time-varying entropy implies the addition of a term in the continuity equation

$$\dot{\rho} + 3H(\rho + p) = \frac{T\dot{S}}{a^3} . \tag{6}$$

Combining this with the first Friedmann equation we obtain a modified, non-equilibrium second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p - \frac{T\dot{S}}{a^3 H}\right) . \tag{7}$$

In principle, this evolution equation does not seem to be compatible with the Einstein field equations. In order to achieve that, the laws of thermodynamics need to be rigorously incorporated into the computation of the equations of motion. This can be achieved by applying the variational formalism of non-equilibrium thermodynamics, developed in another context by Gay-Balmaz and Yoshimura,[3,4] to General Relativity.

3. Entropic forces in mechanics and field theory

Entropic forces emerge naturally in any physical system out of equilibrium. They are a consequence of the coarse-graining of physical degrees of freedom and the laws of thermodynamics, which impose entropy to be a monotonically increasing function of time. This breaks time reversibility.

The dynamics of the coarse-grained degrees of freedom is unknown or ignored and so they do not appear in the action of the physical system. It would seem that a variational treatment of an out-of-equilibrium system is not possible. However, both the extremal-action principle and the second law of thermodynamics can be merged consistently by imposing the latter as a constrain on the variational problem defined by the action.[3,4] On the other hand, the first law of thermodynamics is obtained from the symmetries of the problem.

3.1. *Entropic forces in mechanics*

Let us start by reviewing the emergence of entropic forces in a mechanical system. Consider the action

$$\mathcal{S} = \int dt L(q, \dot{q}, S) , \tag{8}$$

where the Lagrangian depends on the generalized coordinate $q(t)$, its time derivative $\dot{q}(t)$ and the entropy $S(t)$. The variation of the action gives

$$\delta\mathcal{S} = \int dt \left(\frac{\delta L}{\delta q}\delta q + \frac{\partial L}{\partial S}\delta S\right) . \tag{9}$$

Setting $\delta\mathcal{S} = 0$ defines the variational problem. In order to enforce the second law of thermodynamics, we need to impose the variational constraint

$$\frac{\partial L}{\partial S}\delta S = f\delta q , \tag{10}$$

which simply states the relationship between variations of the entropy and the generalized coordinate. If we plug this in the variation of the action, then we can

readily obtain the equations of motion

$$\frac{\delta L}{\delta q} = \frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = -f\,, \tag{11}$$

which is the Euler-Lagrange equation modified by the addition of the entropic force f. In order to know its precise form, we need to impose as well the phenomenological constraint, which is obtained by formally replacing the variations with time derivatives

$$\frac{\partial L}{\partial S}\dot{S} = f\dot{q}\,. \tag{12}$$

Note that usually temperature can be introduced as

$$T = -\frac{\partial L}{\partial S} > 0 \tag{13}$$

providing a clearer meaning to the variational and phenomenological constraint.

The imposition of the constraints and the emergence of an entropic force break the symmetry under time inversion. First, note that formally the Euler-Lagrange equation is invariant under the $t \to -t$ transformation, even with a non-vanishing entropic force. Next, the positiveness of the temperature and the change of entropy imposes $f\dot{q} < 0$. The entropic force has qualities of a generalized friction, as it opposes the coordinate velocity. Now, if one performs time inversion, this sign constraint becomes $f\dot{q} > 0$, as the temperature remains positive but entropy decreases with time. Before solving the equations of motion, $\dot{q}$ is still a degree of freedom and, thus, one must conclude a flip in the sign of f, which flips the overall sign of the Euler-Lagrange equation. Hence, one concludes that the emergence of an entropic force breaks symmetry under time reversion. The evolution of the system becomes irreversible.

3.2. *Entropic forces in classical field theory*

The extension of the variational formalism of non-equilibrium thermodynamics to the continuum is a bit involved. We present a short-cut derivation that relies on the introduction of an additional constraint. We refer the reader to the appendix of[1] to check its equivalence with the full variational derivation originally presented in.[4]

The action of a scalar field on Minkowski space-time contains now a dependency on a scalar function $s(t, \vec{x})$ that encodes information related to coarse-grained degrees of freedom

$$\mathcal{S} = \int d^4x \mathcal{L}\left(\phi, \partial_\mu \phi, s\right)\,. \tag{14}$$

In a similar fashion as before, the extremal-action principle needs to be supplemented by a variational constraint

$$\frac{\partial \mathcal{L}}{\partial s}\delta s = f\delta\phi\,, \tag{15}$$

so that the equation of motion becomes

$$\frac{\delta\mathcal{L}}{\delta\phi} = \frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} = -f\,, \tag{16}$$

which is nothing but the Euler-Lagrange equation of a scalar field with a new term of entropic origin. As before, one can usually introduce a temperature as

$$T = -\frac{\partial\mathcal{L}}{\partial s} > 0 \tag{17}$$

and we would like to interpret s as the entropy density. However, due to spatial fluxes entropy need not necessarily increase locally and, thus, the constraint would not have a fixed sign. Before proceeding, we inspect the corresponding phenomenological constraint

$$\frac{\partial\mathcal{L}}{\partial s}\partial_0 s = f\partial_0\phi\,. \tag{18}$$

Instead, we interpret δs and $\partial_0 s$ as local entropy production and introduce a new function $s_{\rm tot}$ which is the actual entropy density and whose changes $\delta s_{\rm tot}$ and $\partial_0 s_{\rm tot}$ are indeed total local changes of the entropy density. Both are related as

$$\partial_0 s = \partial_0 s_{\rm tot} - \partial_i j_s^i\,, \tag{19}$$

where j_s^i is the entropy flux. This latter equation is an additional constraint we impose for the variational formalism to be consistent. Now one can check that $f\partial_0\phi < 0$ and time reversibility is broken by the same argument used in the mechanics example.

3.3. *Entropic forces in presence of additional symmetries*

The generalization of the above formalism to higher order tensors or to representations of some internal symmetry group is straightforward. Let us consider a field tensor z of contravariant rank r, which is also in some representation of an internal symmetry labelled by an index A. Then one builds the variational constraint as

$$\frac{\partial\mathcal{L}}{\partial s}\delta s = f_{A;\mu_1,\ldots\mu_r}\delta z_A^{\mu_1,\ldots\mu_r}\,, \tag{20}$$

which delivers the equation of motion

$$\frac{\delta\mathcal{L}}{\delta z_A^{\mu_1,\ldots\mu_r}}, = -f_{A;\mu_1,\ldots\mu_r}\,. \tag{21}$$

4. Entropic forces in General Relativity

The previous discussion makes us ready to study entropic forces in General Relativity. However, with the introduction of a dynamical space-time the very notion of time evolution becomes non-trivial. As we will see shortly, it is possible to obtain a modification of entropic origin to Einstein's field equation in the Lagrangian formulation of General Relativity. Its proper interpretation will require, nevertheless, the use of the Hamiltonian formalism.

4.1. *Lagrangian formulation*

Let us consider the Einstein-Hilbert action plus a matter term

$$\mathcal{S} = \frac{1}{2\kappa}\int d^4x\sqrt{-g}R + \int d^4x\mathcal{L}_m(g_{\mu\nu}, s)\,, \tag{22}$$

where $\kappa = 8\pi G$ is the gravitational coupling and we allow for the dependence of the matter Lagrangian on a function $s(t, \vec{x})$, which will have a similar interpretation to the one presented in the previous section. The extremal-action principle is supplemented by the variational constraint.

$$\frac{\partial\mathcal{L}_m}{\partial s} = \frac{1}{2}\sqrt{-g}f_{\mu\nu}\delta g^{\mu\nu}\,. \tag{23}$$

From the extremal-action principle and the variational constraint we obtain the modified Einstein's field equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa\left(T_{\mu\nu} - f_{\mu\nu}\right)\,, \tag{24}$$

where there is an additional term $f_{\mu\nu}$ of entropic origin. In order to obtain an expression for this term and to check the breaking of symmetry under time inversion we need to work in the Hamiltonian formulation of General Relativity.

4.2. *Hamiltonian formulation*

General Relativity admits a Hamiltonian formulation in the Arnowitt-Deser-Misner (ADM) formalism. Space-time is foliated in constant time hypersurfaces with normal unit vector n^μ, being the its 4-metric split as

$$g_{\mu\nu} = h_{\mu\nu} - n_\mu n_\nu\,, \tag{25}$$

where $h_{\mu\nu}$ is the 3-metric induced on the hypersurfaces. Analogously, one can parametrize the 4-metric in terms of the 3-metric h_{ij} and the lapse and shift functions N and N^i

$$ds^2 = -(Ndt)^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt)\,. \tag{26}$$

Greek indices run from 0 to 3, while Latin ones do from 1 to 3 and are raised and lowered by h_{ij}. The normal vector can be written as

$$n_\mu = (-N, 0, 0, 0)\,. \tag{27}$$

Note that h_{ij} is the purely spatial part of $h_{\mu\nu}$ and is also the pull-back of $g_{\mu\nu}$ onto the hypersurface.

The Einstein-Hilbert action for this parametrization of the metric is given by the following gravitational Lagrangian:

$$\mathcal{L}_G = \sqrt{-g}R = \frac{1}{2\kappa}N\sqrt{h}\left({}^{(3)}R + K_{ij}K^{ij} - K^2\right)\,, \tag{28}$$

where K_{ij} is the extrinsic curvature of the 3-hypersurface Σ and is given by the Lie derivative along the normal vector n:

$$K_{ij} = \frac{1}{2}\pounds_n h_{ij} = \frac{1}{2N}\left(\partial_0 h_{ij} - \nabla_i N_j - \nabla_j N_i\right) \tag{29}$$

where ∇ denotes the covariant derivative on Σ with respect to the 3-metric h_{ij}. Its trace and traceless part are:

$$\begin{aligned} K &= h^{ij}K_{ij} = \frac{1}{N}\left(\partial_0 \ln\sqrt{h} - \nabla_i N^i\right) \\ \bar{K}_{ij} &= K_{ij} - \frac{1}{3}K h_{ij}\,. \end{aligned} \tag{30}$$

Unlike the intrinsic curvature, described by the Riemann tensor $R^{\rho}_{\mu\nu\lambda}$ and its contractions, the extrinsic curvature is a quantity that depends on the embedding of a surface in a larger manifold.

We are now ready to introduce the Hamiltonian formulation of the theory. Note that the only quantity whose time derivative appears in the gravitational Lagrangian is the 3-spatial metric h_{ij} and, thus, it is the only dynamical or propagating d.o.f. Correspondingly, one defines its conjugate momentum as:

$$\Pi^{ij} = \frac{\partial \mathcal{L}_G}{\partial \dot{h}_{ij}} = \sqrt{h}\left(K^{ij} - K h^{ij}\right)\,. \tag{31}$$

With this, the gravitational Lagrangian can be rewritten as:

$$\begin{aligned} \mathcal{L}_G &= N\sqrt{h}^{(3)}R - \frac{N}{\sqrt{h}}\left(\Pi_{ij}\Pi^{ij} - \frac{1}{2}\Pi^2\right) - 2\Pi^{ij}\nabla_i N_j \\ &= \Pi^{ij}\dot{h}_{ij} - N\mathcal{H} - N_i\mathcal{H}^i - 2\nabla_i\left(\Pi^{ij}N_j\right)\,, \end{aligned} \tag{32}$$

where $\Pi = h_{ij}\Pi^{ij}$ and we introduced the functions:

$$\begin{aligned} \mathcal{H} &= -\sqrt{h}^{(3)}R + \frac{1}{\sqrt{h}}\left(\Pi_{ij}\Pi^{ij} - \frac{1}{2}\Pi^2\right) \\ \mathcal{H}^i &= -2\nabla_j\left(h^{-1/2}\Pi^{ij}\right)\,. \end{aligned} \tag{33}$$

Since N and N_i are not dynamical variables, they merely enter the gravitational Lagrangian as Lagrange multipliers. One defines the gravitational Hamiltonian as:

$$\begin{aligned} \mathcal{H}_G &= \Pi^{ij}\dot{h}_{ij} - \mathcal{L}_G \\ &= N\mathcal{H} + N_i\mathcal{H}^i + \nabla_i\left(\Pi^{ij}N_j\right)\,, \end{aligned} \tag{34}$$

with the Hamiltonian and momentum constraints:

$$\begin{aligned} \frac{\delta\mathcal{H}_G}{\delta N} &= \mathcal{H} = 0 \\ \frac{\delta\mathcal{H}_G}{\delta N_i} &= \mathcal{H}^i = 0\,. \end{aligned} \tag{35}$$

The Hamiltonian evolution equations are obtained from the variations of the action with respect to the metric and conjugate momentum

$$\delta\mathcal{S} = \int d^4x \left[\left(-\dot{\Pi}^{ij} - \frac{\delta\mathcal{H}_G}{\delta h_{ij}} - \kappa N\sqrt{h}\tilde{f}^{ij} + 2\kappa\frac{\partial\mathcal{L}_m}{\partial h_{ij}} \right)\delta h_{ij} + \left(\dot{h}_{ij} - \frac{\delta\mathcal{H}_G}{\delta\Pi^{ij}} \right)\delta\Pi^{ij} \right], \tag{36}$$

where the variational constraint

$$\frac{\partial\mathcal{L}_m}{\partial s}\delta s = -\frac{1}{2}N\sqrt{h}\tilde{f}^{ij}\delta h_{ij} \tag{37}$$

was already implemented. Note that the minus sign arises from

$$\frac{\partial h_{ij}}{\partial h^{kl}} = -\frac{1}{2}\left(h_{ik}h_{jl} + h_{il}h_{jk}\right) \tag{38}$$

and that $\tilde{f}_{ij}$ is the pull-back of $f_{\mu\nu}$ onto the hypersurfaces. By setting the variation to 0 we obtain the two Hamilton equations

$$\begin{aligned} \frac{\delta\mathcal{H}_G}{\delta h_{ij}} &= -\dot{\Pi}^{ij} - \kappa N\sqrt{h}f^{ij} + 2\kappa\frac{\partial\mathcal{L}_m}{\partial h_{ij}} \\ \frac{\delta\mathcal{H}_G}{\delta\Pi^{ij}} &= \dot{h}_{ij}\,, \end{aligned} \tag{39}$$

which completes the derivation of the entropic modification to the gravitational equations of motion in the Hamiltonian formulation.

The tensor $\tilde{f}_{ij}$ can be obtained from the phenomenological constraint, which can now be stated rigorously. In the ADM formalism, a well-defined notion of time evolution is given by the flow along the normal vector n^μ. Hence, the time derivative is generalized to the Lie derivative along n^μ. Then the phenomenological constraint is given by

$$\frac{\partial\mathcal{L}}{\partial s}\pounds_n s = \frac{1}{2}N\sqrt{h}\tilde{f}_{ij}\pounds_n h^{ij} \tag{40}$$

where growth in entropy by local processes is related to total entropy density growth by

$$\pounds_n s = \pounds_n s^{tot} - \nabla_i j_s^i\,. \tag{41}$$

Entropy produced locally is expected to grow over time, i.e. with the flow along the hypersurfaces, in compliance with the second law of thermodynamics. This completes the variational formulation of entropic forces in General Relativity.

4.3. *The Raychauduri equation*

Let us explore an immediate dynamical consequence of the inclusion of entropic forces, namely its effect on a congruence of worldlines with tangent vector n^μ. The congruence is then characterized by the tensor:

$$\Theta_{\mu\nu} = D_\nu n_\mu = \frac{1}{3}\Theta h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu} - a_\mu n_\nu \tag{42}$$

where θ is the expansion rate of the congruence, $\sigma_{\mu\nu}$ is its shear or symmetric traceless part and $\omega_{\mu\nu}$ is its vorticity or antisymmetric part. If the worldline is not a geodesic, then the congruence suffers an acceleration given by:

$$a_\mu = n^\nu D_\nu n_\mu \,. \tag{43}$$

One can compute the Lie derivative of the expansion of the congruence along its tangent vector and find the Raychauduri equation:[5]

$$\pounds_n \Theta = -\frac{1}{3}\Theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu}n^\mu n^\nu + D_\mu a^\mu \,. \tag{44}$$

Let us perform the standard analysis of the sign of this equation. It is clear that $\sigma_{\mu\nu}\sigma^{\mu\nu} > 0$ and $\Theta^2 > 0$. On the other hand, if the congruence is chosen to be orthogonal to the spatial hypersurfaces, as we have been considering, then the vorticity vanishes $\omega_{\mu\nu} = 0$. Lastly, it is left to consider the term $R_{\mu\nu}n^\nu n^\nu$, which we can rewrite with the help of the field equations:

$$R_{\mu\nu}n^\mu n^\nu = 8\pi G\left(T_{\mu\nu}n^\mu n^\nu + \frac{1}{2}T - f_{\mu\nu}n^\mu n^\nu - \frac{1}{2}f\right). \tag{45}$$

If the strong energy condition is satisfied, then:

$$T_{\mu\nu}n^\mu n^\nu \geq -\frac{1}{2}T \tag{46}$$

and, in the absence of intrinsic acceleration, $a_\mu = 0$, we can establish the bound:

$$\pounds_n \Theta + \frac{1}{3}\Theta^2 \leq 8\pi G\left(f_{\mu\nu}n^\mu n^\nu + \frac{1}{2}f\right). \tag{47}$$

For a vanishing entropic force $f_{\mu\nu} = 0$, this means that an expanding congruence cannot indefinitely sustain its divergence and will eventually recollapse. On the contrary, a positive and sufficiently large entropic contribution can avoid such recollapse. This may become relevant for an expanding universe, but also to generic gravitational collapse and the singularity theorems.[6–8]

5. Sources of entropy

A main ingredient in the variational formulation of non-equilibrium thermodynamics in General Relativity is the inclusion of entropy at the Lagrangian level and the derivation of a notion of temperature from it. In this section we present two relevant examples: hydrodynamical matter, which is a prototypical case, and horizons.

5.1. *Entropy from hydrodynamical matter*

A classical fluid is the simplest matter content that can be considered in General Relativity and it is of particular relevance in Cosmology. Without paying attention to microphysical details, the Lagrangian of hydrodynamical matter can be written as

$$\mathcal{L}_m = -\sqrt{-g}\rho(g_{\mu\nu}, s)\,, \tag{48}$$

being then the temperature simply given by

$$T = -\frac{1}{\sqrt{-g}}\frac{\partial \mathcal{L}_m}{\partial s} = \frac{\partial \rho}{\partial s}\,. \tag{49}$$

This is analogous to the case of a mechanical system, where the Lagrangian is generically given by a kinetic and a potential energy

$$L = E_K(q, \dot{q}) - U(q, S) \tag{50}$$

and temperature can be defined as

$$T = \frac{\partial L}{\partial S} = -\frac{\partial U}{\partial S}\,. \tag{51}$$

Thus, the energy density of a fluid can be readily interpreted as the thermodynamic internal energy.

5.2. *Entropy from gravity and horizons*

Gravity itself has thermodynamical features. It is known since the discovery of the laws of black hole mechanics and their promotion to laws of black hole thermodynamics, allowed by the introduction of Bekenstein entropy and Hawking temperature. We propose to include the entropy associated with a horizon $\mathcal{H}$ by extending the Einstein-Hilbert actions with surface terms of Gibbons-Hawking-York (GHY) type

$$S_{GHY} = \frac{1}{8\pi G}\int_{\mathcal{H}} d^3y\sqrt{h}K\,, \tag{52}$$

where h is the determinant of the induced 3-metric on the horizon and K is the trace of its extrinsic curvature. Definitions are analogous to the ones used in the ADM formalism, but we stress that here the hypersurface of interest is a horizon and not constant-time hypersurfaces.

From the thermodynamic point of view, the GHY term contributes to the internal energy of the system. Hence, it can be rewritten as a function of the temperature and entropy of the horizon

$$S_{GHY} = -\int dt\, N(t)\, TS\,. \tag{53}$$

We have kept the lapse function $N(t)$, to indicate that the variation of the total action with respect to it will generate a Hamiltonian constraint with an entropy term together with the ordinary matter/energy terms. In order to illustrate this, let us now compute the GHY for the event horizon of a Schwarzschild black hole.

The space-time of a Schwarzschild black hole of mass M is described by the static metric

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2\,. \tag{54}$$

We foliate it with spherical hypersurfaces, i.e. their intersection with constant time hypersurfaces is a 2-sphere around the origin of coordinates. The corresponding normal vector is

$$n = -\sqrt{1 - \frac{2GM}{r}}\partial_r \,. \tag{55}$$

With this, the trace of the extrinsic curvature for such a sphere scaled by the metric determinant is

$$\sqrt{h}K = (3GM - 2r)\sin\theta \,. \tag{56}$$

Integrating over the angular coordinates and setting the 2-sphere at the event horizon, i.e. $r = 2GM$, and restoring for a moment $\hbar$ and c, the GHY becomes

$$S_{GHY} = -\frac{1}{2}\int dt Mc^2 = -\int dt T_{BH} S_{BH} \,, \tag{57}$$

where T_{BH} is the Hawking temperature and S_{BH} is the Bekenstein entropy of the Schwarzschild black hole:

$$T_{BH} = \frac{\hbar c^3}{8\pi GM} \quad S_{BH} = \frac{Ac^3}{4G\hbar} = \frac{4\pi GM^2}{\hbar c} \,. \tag{58}$$

This favors the interpretation of the GHY term of a horizon as a contribution to the internal energy in the thermodynamic sense.

6. Irreversible cosmology

We derived in section 4 a powerful, generic tool to describe non-equilibrium thermodynamic effects in gravity. In the Hamiltonian formulation of General Relativity it is possible to obtain the modified equations of motion and rigorously impose the time-evolution of the entropy as dictated by the second law of thermodynamics.

In section 2 we motivated the study of these phenomena by our interest in understanding the dynamics of irreversible cosmology and justifying its equations of motion. One can obtain them using the Hamilton equations.[1] Here, however, we present a slightly different approach. Due to the symmetries of the FLRW universe, homogeneity and isotropy, there is a preferred slicing and time evolution is well-defined even at the Lagrangian level. Therefore, we can obtain the equations of non-equilibrium cosmology by imposing these symmetries, i.e. making an ansatz for the metric

$$ds^2 = -N(t)^2 dt^2 + a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega_2^2 \,, \right) \tag{59}$$

where the lapse function $N(t)$ accounts for the freedom in choosing the time coordinate, i.e. the symmetry under $t \to f(t)$. The Ricci scalar associated to this metric is

$$R = \frac{6}{a^2}\left(\frac{a\ddot{a}}{N^2} + \frac{\dot{a}^2}{N^2} + k\right). \tag{60}$$

Let us stress that this result is imposed by symmetry, not dynamics. Without loss of generality, we can restrict the action to a region of comoving volume 1 and write it as

$$S = \int dtL = \frac{3}{8\pi G}\int dtNa\left(\frac{a\ddot{a}}{N^2} + \frac{\dot{a}^2}{N^2} + k\right) + \int dtNa^3\mathcal{L}_m(N,a,S)\,. \tag{61}$$

Effectively, this action describes a mechanical system, for the scale factor $a(t)$ has no spatial dependency and we got rid of the integral over spatial coordinates. The first term can be rewritten using integration by parts in order to get only terms with at most the first derivative of a

$$S = \frac{3}{8\pi G}\int dtNa\left(-\frac{\dot{a}^2}{N^2} + k\right) + \int dtNa^3\mathcal{L}_m(N,a,S)\,. \tag{62}$$

The variational constraint is here given by the usual expression for a mechanical system

$$\frac{\partial L}{\partial S}\delta S = f\delta a\,. \tag{63}$$

The Hamiltonian of the system is

$$H = \dot{a}\frac{\partial L}{\partial \dot{a}} - L = \frac{3}{8\pi G}\left(-\frac{\dot{a}^2 a}{N} - kaN\right) - Na^3\mathcal{L}_m\,. \tag{64}$$

For an arbitrary lapse function $N(t)$ this can be rewritten as

$$H = N^2a^3\frac{\partial \mathcal{L}_m}{\partial N}\,, \tag{65}$$

which gives the Hamiltonian constraint of the system. On the other hand, the dynamics is obtained from the equation of motion for a

$$\frac{\delta L}{\delta a} = -f\,. \tag{66}$$

Let us now consider the matter Lagrangian to be that of a perfect fluid, i.e.

$$\mathcal{L}_m = -\rho(a,S)\,. \tag{67}$$

Its stress-energy tensor is given in terms of the density ρ and pressure p by

$$T^{\mu\nu} = (\rho + p)u^{\mu\nu} + pg^{\mu\nu} \tag{68}$$

and $u^m u = (N,0,0,0)$ is the unit vector tangent to a comoving observer. Pressure is then obtained as

$$p = \frac{a^2}{3}T^{ij}\delta_{ij} = -\frac{1}{2a^4}\frac{\partial a^3\rho}{\partial a}\,. \tag{69}$$

Using the expressions for ρ and p and rearranging the terms in the Hamiltonian constraint and the equation of motion for $a(t)$ we arrive at the modified Friedmann

equations

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho$$
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p + \frac{f}{a^2}\right). \tag{70}$$

The expression for the entropic force F is obtained from the phenomenological constraint

$$\left(\frac{\partial L}{\partial S}\right)\dot{S} = -T\dot{S} = f\dot{a} < 0\,, \tag{71}$$

which determines the sign $f < 0$ whenever dealing with an expanding universe $\dot{a} > 0$. We express finally the second Friedmann equation as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p - \frac{T\dot{S}}{a^2\dot{a}}\right). \tag{72}$$

From this equation we can conclude that entropic forces generally drive an acceleration of the expansion of the universe. Whether this can dominate the dynamics of the scale factor will depend on the particular thermodynamic process. Most of the expansion history of the universe takes place in equilibrium. Out of equilibrium processes, such as (p)reheating, phase transitions or gravitational collapse are short-lived. Should their associated entropic force dominate, we still only expect a short period of accelerated expansion.

Symmetry under time inversion is broken by the same arguments presented in section 3. Hence, the Friedmann equations together with the phenomenological constraint, i.e. the second law of thermodynamics, describe cosmic irreversible dynamics.

We currently live in a universe that is undergoing an accelerated expansion. The possibility of explaining this by means of an entropic force is fascinating. In the next section we review our proposal to achieve this by means of the sustained growth of the entropy associated to a causal horizon.

7. Cosmic acceleration as an entropic force

The growth of entropy associated to the cosmic horizon may be responsible for the current observed accelerated expansion of the universe. The choice of horizon is in principle not unique. The only available one which can be defined locally in time is the cosmic apparent horizon, but it fails to significantly affect the expansion.[2]

There is another option in the framework of eternal inflation, according to which we live in an open universe nucleated by quantum tunneling from a false to a true vacuum. After nucleation the bubble universe undergoes its own inflationary era, which renders the local metric almost flat. However, due to the presence of the bubble walls, the true causal horizon is located at a finite distance. It induces an

entropic fluid via GHY term with energy density

$$\rho_H\, a^2 = \frac{T_H S_H}{a} = \frac{x_0}{2G}\,\sinh(2a_0 H_0 \eta)\,, \quad x_0 \equiv \frac{1-\Omega_0}{\Omega_0} = e^{-2N}\left(\frac{T_{\rm rh}}{T_{\rm eq}}\right)^2 (1+z_{\rm eq})\,, \tag{73}$$

where η is the conformal time, Ω_0 is the density parameter, $T_{\rm rh}$ is the reheating temperature, $T_{\rm eq}$ and $z_{\rm eq}$ are, respectively, the temperature and redshift at matter-radiation equality. Introducing $\tau = a_0 H_0 \eta$ one can write the second Friedmann equation in conformal time as

$$\left(\frac{a'}{a_0}\right)^2 = \Omega_M \left(\frac{a}{a_0}\right) + \Omega_K \left(\frac{a}{a_0}\right)^2 + \frac{4\pi}{3}\Omega_K \left(\frac{a}{a_0}\right)^2 \sinh(2\tau)\,, \tag{74}$$

where Ω_M is the matter density paremeter and Ω_K is the curvature parameter. We call this the general relatvistic entropic acceleration (GREA) theory.

By solving this equation with cosmological parameters consistent with the CMB values (Planck 2018: $\Omega_M \simeq 0.31$, $\Omega_K \simeq 0.0006$, $h_0 \simeq 0.68$) and initial conditions deep in the matter era, $a_i(\tau) = a_0\,\Omega_M \tau^2/4$, we find generic accelerating behaviour beyond the scale factor $a \sim 1/2$ (*i.e.* $z \sim 1$), see Fig. 1. This is consistent with the current observed acceleration of the universe and may even resolve the Hubble tension,[9] providing a way to obtain from the CMB a present value of H_0 that is consistent with late-universe observations, see Fig. 2.

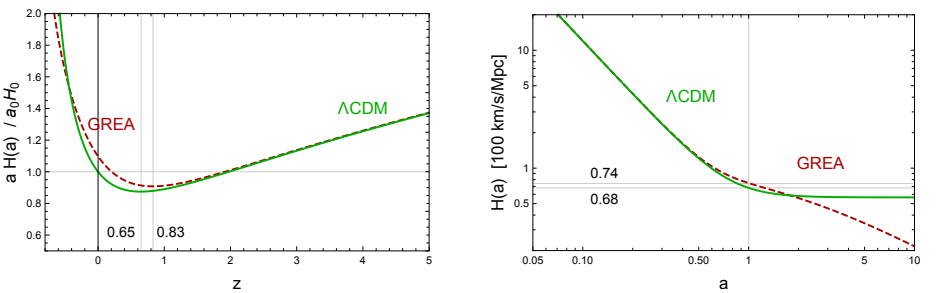


Fig. 1. The left plot shows the evolution of the inverse comoving horizon with the coasting point for each model, at $z \simeq 0.65$ for ΛCDM (in green) and $z \simeq 0.83$ for GREA (in red). The right plot shows the evolution of the rate of expansion. For GREA the present rate of expansion is approximately 74 km/s/Mpc, compared with the value of 68 km/s/Mpc predicted by ΛCDM, in agreement with the asymptotic value at the CMB.

8. Conclusions

The consistent inclusion of non-equilibrium phenomena in General Relativity leads to the modification of the Einstein field equations, as can be checked both in the Lagrangian and Hamiltonian formulations of the theory. This breaks symmetry under time inversion and allows for the introduction of an arrow of time.

In cosmology this implies the appearance of a term of entropic origin in the second Friedmann equation, which tends to accelerate the expansion of the universe as a result of the increase in entropy. Some physical processes such as (p)reheating, phase transitions or gravitational collapse may lead to phenomenologically relevant applications of this formalism. We look forward to further developments.

The sustained entropy growth associated to a causal horizon in the open universe scenario leads to an acceleration consistent with current observations and it may even solve the H_0 tension. Further research will be required to establish the full viability of the GREA theory.

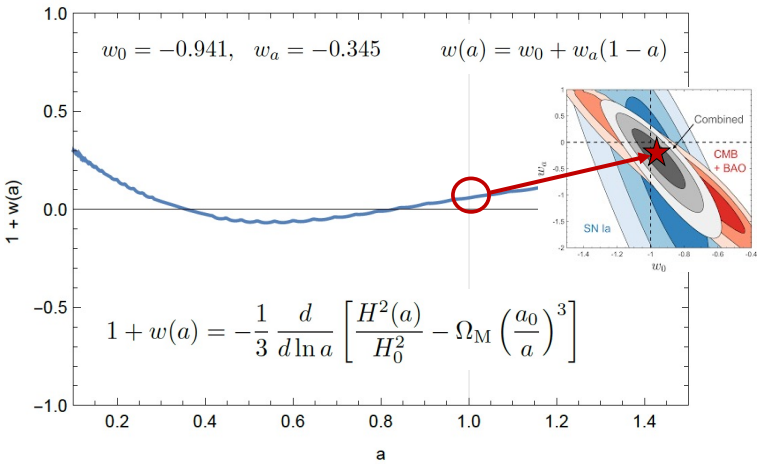


Fig. 2. The effective equation of state of the non-matter component of the GREA theory, as a function of thee sale factor. Note that the predicted effective PCL parameters (w_o, w_a) agree remarkably well with present observations.

Acknowledgments

The authors acknowledge support from the Spanish Research Project PGC2018-094773-B-C32 (MINECO-FEDER) and the Centro de Excelencia Severo Ochoa Program SEV-2016-0597. The work of LEP is funded by a fellowship from "La Caixa" Foundation (ID 100010434) with fellowship code LCF/BQ/IN18/11660041 and the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 713673.

References

1. L. Espinosa-Portales and J. García-Bellido, Covariant formulation of non-equilibrium thermodynamics in General Relativity, *Phys. Dark Univ.* **34**, p. 100893 (2021).
2. J. García-Bellido and L. Espinosa-Portales, Cosmic acceleration from first principles, *Phys. Dark Univ.* **34**, p. 100892 (2021).

3. F. Gay-Balmaz and H. Yoshimura, A lagrangian variational formulation for nonequilibrium thermodynamics. part i: Discrete systems, *Journal of Geometry and Physics* **111**, 169 (2017).
4. F. Gay-Balmaz and H. Yoshimura, A lagrangian variational formulation for nonequilibrium thermodynamics. part ii: Continuum systems, *Journal of Geometry and Physics* **111**, 194 (2017).
5. R. M. Wald, *General Relativity* (Chicago Univ. Pr., Chicago, USA, 1984).
6. R. Penrose, Gravitational collapse and space-time singularities, *Phys. Rev. Lett.* **14**, 57 (1965).
7. R. Penrose, Gravitational collapse: The role of general relativity, *Riv. Nuovo Cim.* **1**, 252 (1969).
8. S. W. Hawking and R. Penrose, The Singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. Lond. A* **314**, 529 (1970).
9. A. G. Riess, The Expansion of the Universe is Faster than Expected, *Nature Rev. Phys.* **2**, 10 (2019).

LHC experiments for long-lived particles of the dark sector

Vasiliki A. Mitsou

Instituto de Física Corpuscular (IFIC), CSIC – Universitat de València, C/ Catedrático José Beltrán 2, E-46980 Paterna (Valencia), Spain
E-mail: vasiliki.mitsou@ific.uv.es
webific.ific.uv.es/web

Dark matter scenarios are being tested at the LHC in the general-purpose experiments through promptly decaying states. In parallel, new dedicated detectors have been proposed for the LHC to probe dark matter portal theories predicting long-lived particles that decay away from the interaction point: MoEDAL-MAPP, MoEDAL-MALL, FASER, SND@LHC, CODEX-b, MATHUSLA, AL3X, ANUBIS, FACET, milliQan, FORMOSA. In addition, the SHiP beam-dump experiment is planned to operate with the SPS beam to extend the discovery reach for such particles. The detector design and expected physics sensitivity of these experiments is presented with emphasis on scenarios explaining the nature of dark matter.

Keywords: Dark matter; Portal models; Long-lived particles; Displaced vertices; LHC.

1. Introduction

In collider physics, and in particular at the Large Hadron Collider (LHC),[1] there is a growing experimental interest in long-lived particles (LLPs),[2,3] which travel a macroscopic distance before either decaying within the detector or giving rise to anomalous ionisation. In theory, long lifetimes may be due to a symmetry (leading to stable particles), narrow mass splittings, small couplings, or a heavy mediator. In this arena, a class of LLPs called Feebly interacting Particles (FIPs)[4] is characterised by indirect interactions of FIPs with Standard Model (SM) particles through low-dimensional operators. These interactions, commonly referred to as *portals*, may predict dark photons (vector portal), a light dark Higgs boson (scalar portal), axion-like particles (pseudoscalar portal), or heavy neutral leptons (fermion portal).

Scenarios beyond-the-SM (BSM) that introduce a hidden sector in addition to the visible SM sector are required to explain a number of observed phenomena in particle physics, astrophysics and cosmology such as the non-zero neutrino masses, the dark matter (DM), the baryon asymmetry of the Universe and the cosmological inflation. The mystery of DM, in particular, is quite intriguing with around 25% of our Universe invisible (dark) that only interacts through gravity and remains unaccounted for in the SM. This review focuses on the possibility to explore DM with LLPs, as a complementary way to searches performed by the main experiments. ATLAS[5] and CMS[6] look for DM in mono-X final states, in associated production and in resonances via mediators,[7] while LHCb[8] has good prospects for probing DM-portal LLPs after its Phase-I upgrade.[9]

This review is organised as follows. After a brief introduction in Sec. 1, the specially designed LHC experiments to tackle LLPs are concisely presented in Sec. 2. The dark-scalar and dark-vector portals, proposing neutral LLPs that decay in the detector volumes to known particles, are reviewed in Sec. 3 and Sec. 4, respectively. Dark photons at the limit of zero mass, leading to fractionally charged particles, are discussed in Sec. 5. In Sec. 6, other BSM scenarios pertinent to LLPs are highlighted. Finally, possible measurements of new particle properties are outlined in Sec. 7, before summarising and giving an outlook in Sec. 8.

2. Dedicated LLP Experiments at the LHC

There is a constantly growing list of experiments planned for the LHC optimised for the detection of feebly interacting particles. Their projected added value stems mostly from their different operation and geometry parameters: angle w.r.t. the beam axis, the detector volume and distance from IP — probing different ranges of lifetime, couplings to SM and boost (mass) —, their time scale and the detector design.

Besides the detectors highlighted below, there are other, operating or future, non-LHC experiments with sensitivity to DM and other portal models. Such cases include BaBar,[10] existing CERN beam-dump experiments, like the NA48/2,[11] NA62[12] and NA64,[13] the Fermilab SeaQuest[14] and the reactor experiment SoLid.[15] Proposed experiments include the Search for Hidden Particles (SHiP)[16] at the CERN Beam Dump Facility, SHADOWS[17] at the CERN North Area, LUXE-NPOD[18] at DESY, HECATE[19] for the FCC-ee or CEPC, GAZELLE[20] in Bell II at SuperKEKB, SUBMET[21] at J-PARC and FerMINI[22] at Fermilab, among others.

2.1. *MAPP – MoEDAL Apparatus for Penetrating Particles*

The MoEDAL (Monopole and Exotics Detector at the LHC)[23] experiment is mainly dedicated to searches for manifestations of new physics through highly ionising particles in a manner complementary[24] to ATLAS and CMS. It is the first dedicated *search* LHC experiment. The principal motivation for MoEDAL is the quest for magnetic monopoles[25] and dyons, as well as for any massive, stable or long-lived, slow-moving particle[26] with single or multiple electric charge arising in various extensions of the SM.[27] The MoEDAL detector[28] is deployed around the region at interaction point 8 (IP8) of the LHC in the LHCb[8] vertex locator cavern.

It is a unique and, to a large extend, passive detector based on three different detection techniques: nuclear track detectors (NTDs), magnetic monopole trackers (MMTs) and TimePix pixel devices. MoEDAL has pioneered the quest for magnetic charges, by being the only contender in the high-charge regime[29–32] and by being first in constraining dyons[33] in colliders and in investigating the Schwinger thermal production of monopoles.[34] More analyses, also on electric charges,[35] are ongoing.[36]

The last few years, MoEDAL proposes to deploy MAPP[37,38] in a gallery near IP8 shielded by an overburden of approximately 100 m of limestone from cosmic

rays to extent its reach to the *low* ionisation regime. It is envisaged that the first-stage detector, MAPP-1 will be installed during LHC Long Shutdown 2 (LS2) and Run 3.[36] The purpose of the innermost detector, the MAPP-mQP, shown in Fig. 1, is to search for particles with fractional charge as small as $0.001e$, the so called *millicharged particles (mCP)*, using plastic scintillation bars. A prototype of the mQP detector (10% of the original system) is already in place since 2017 and the data analysis is ongoing.[39] The Phase-1 MAPP-mQP is going to be deployed in the UA82 gallery in a distance 100 m from IP8.

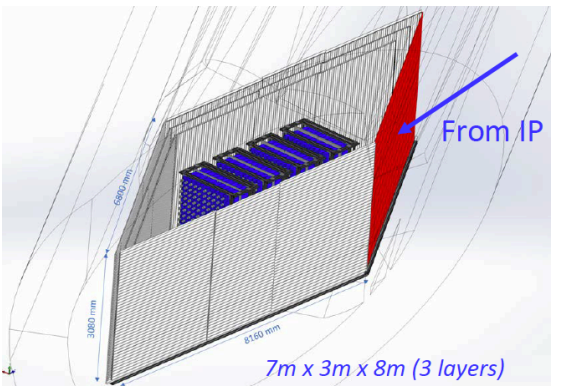


Fig. 1. Diagram of the MAPP-1 detector components: mQP at the centre and three layers of LLP around it.

Another part of the detector, the MAPP-LLP, is deployed as three nested boxes of scintillator hodoscope detectors, in a "Russian doll" configuration, following as far as possible the contours of the cavern as depicted in Fig. 1. It is designed to be sensitive to long-lived neutral particles from new physics scenarios via their interaction or decay in flight in a decay zone of size approximately 5 m (wide) $\times$ 10 m (deep) $\times$ 3 m (high). The MAPP detector can be deployed in a number of positions in the forward direction, at a distance of $\mathcal{O}(100$ m) from IP8. An upgrade plan for the MAPP-1 detector is envisaged for the High-Luminosity LHC (HL-LHC),[40] called MAPP-2, with considerably larger volume than MAPP-1.

Furthermore, the MoEDAL Apparatus for very Long Lived particles (MALL)[37] is intended to push the search for decays of new *electrically charged*, massive and *extremely* long-lived particles, with lifetimes well in excess of a year, by monitoring the exposed MMTs for decay products of trapped BSM particles.[37,38] Quite recently, an independent proposal has been presented to install such absorber volumes near the CMS IP.[41]

2.2. *FASER – ForwArd Search ExpeRiment*

FASER[42] is an approved small and inexpensive experiment designed to search for new particles produced in decays of light mesons copiously present at zero angle at the LHC in Run-3 and beyond. Such particles may be produced in large numbers and

travel for hundreds of meters without interacting, and then decay to SM particles. To search for such events, FASER will be located 480 m downstream of the ATLAS[5] IP in the unused service tunnel TI12 (cf. Fig. 12). It is planned to be constructed and installed in LS2 and collect data during Run 3.[43]

An overview of the different detector components is shown in Fig. 2.[44] The magnets are 0.55 T permanent dipole magnets based on the Halbach array design with a radius of 10 cm. There are three scintillator stations for timing, trigger and vetoing incoming charged particles. The electromagnetic calorimeter consists of four spare outer ECAL modules from LHCb. The tracker consists of three tracking stations, each containing three layers instrumented with spare ATLAS semiconductor-tracker barrel modules.

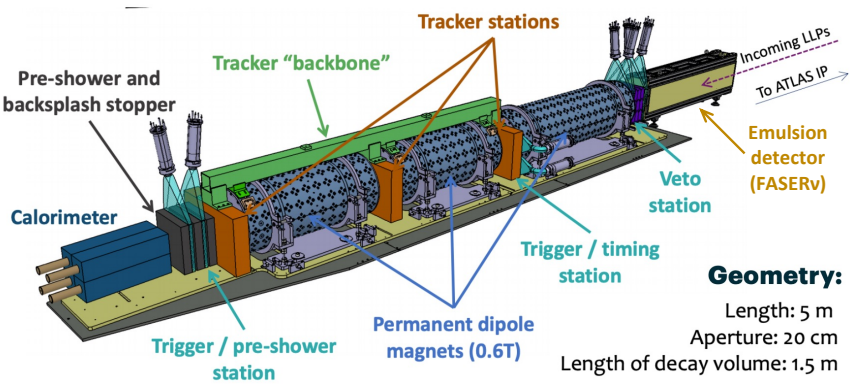


Fig. 2. Annotated schematic view of the FASER detector components.

Moreover, FASERν[45] is designed to directly detect collider neutrinos and study their cross sections at TeV energies. In 2018, a pilot detector employing emulsion films was installed in the far-forward region of ATLAS, also visible in Fig. 2, and collected 12.2 fb^{-1} of pp collision data at a $\sqrt{s} = 13$ TeV. The first candidate vertices consistent with neutrino interactions at the LHC were observed, an example of which is presented in Fig. 3, with a measured 2.7σ excess of neutrino-like signal above muon-induced backgrounds.[46]

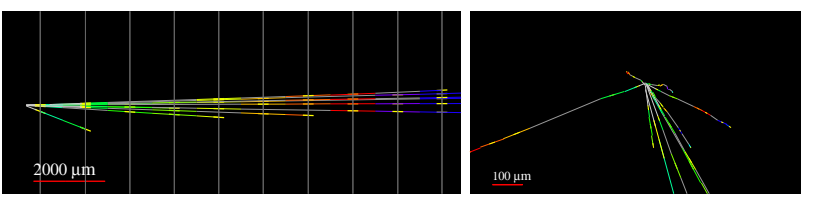


Fig. 3. FASERν event displays of a neutral vertex in the $y - z$ projection longitudinal to the beam direction (left) and in the view transverse to the beam direction (right).[46]

For the HL-LHC, FASER2,[43] a larger successor with roughly $R \sim 1$ m and $L \sim 5$ m is planned to be constructed during LS3 and installed in the LHC Forward Physics Facility (FPF) (cf. Fig. 12). Possible options for FPF locations are an expanded UJ12 cavern or a new cavern ~ 600 m downstream from IP1. Other experiments proposed to be housed in FPF include FASERν2,,[47] a detector with roughly ten times the mass of FASERν, as well as the Forward Liquid Argon Experiment (FLArE),[48] composed of a 10- or 100-tonne-scale liquid argon time projection chamber (LArTPC).

2.3. *SND@LHC – Scattering and Neutrino Detector at the LHC*

SND@LHC[49] is a recently approved, compact and stand-alone experiment designed to measure neutrinos produced at the LHC and search for FIPs in the unexplored range of $7.2 < \eta < 8.7$, where neutrinos are mostly produced from charm decay. It is a small-scale prototype of the SHiP[50] experiment Scattering and Neutrino Detector (SND). The proposed detector is hybrid, combining the nuclear emulsion technology and electronic detector; a schematic view is provided in Fig. 4. It will be installed in the TI18 tunnel, in a location off-axis with respect to the ATLAS[5] IP1.

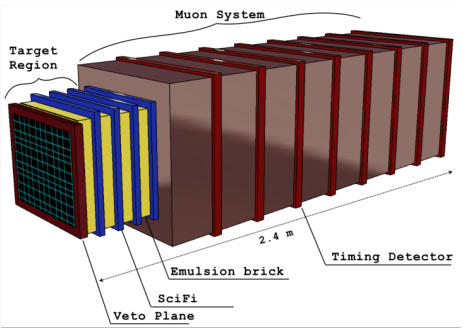


Fig. 4. Annotated schematic view of the SND@LHC detector components.[49]

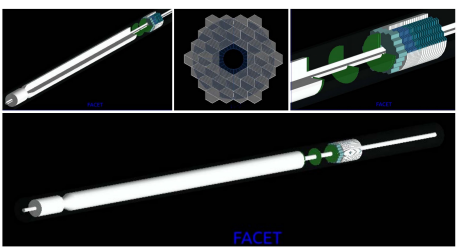


Fig. 5. Diagrams of the detector components of the FACET spectrometer.

2.4. *FACET – Forward-Aperture CMS ExTension*

FACET[51] is a multi-particle spectrometer to be located at $z \sim +100$ m from the CMS[6] IP5. The detector will have a radius of ~ 50 cm and coverage $6 < \eta < 8$, therefore it will be much closer to the IP and feature much larger decay volume than FASER. It is shielded by about $30 - 50$ m of steel in front of it, which corresponds to $190 - 300$ interaction lengths, λ_{int}. FACET will be fully integrated in CMS and can be used either a forward part of CMS or a standalone detector. Some GEANT4 modelling drawings of the FACET detector components are shown in Fig. 5. If approved, FACET will operate during HL-LHC.

2.5. *CODEX-b – COmpact Detector for EXotics at LHCb*

CODEX-b[52,53] is proposed as a cubic detector with a nominal fiducial volume of $10\ \text{m} \times 10\ \text{m} \times 10\ \text{m}$ to be situated in a transverse location ~ 25 m from the LHCb[8] IP8, as depicted in Fig. 6, corresponding to the pseudorapidity range $0.2 < \eta < 0.6$. It will be composed of six Resistive Plate Chamber (RPC) layers — being fast, precise and cheap for a large-area tracker — at 4 cm intervals on each box face with 1 cm granularity. An additional passive Pb shielding of $25\lambda_{\text{int}}$ with embedded active scintillator veto will reduce IP and secondary backgrounds. It is reminded that IP8 runs at a factor of ~ 10 less luminosity than IP1, IP2 and IP5.

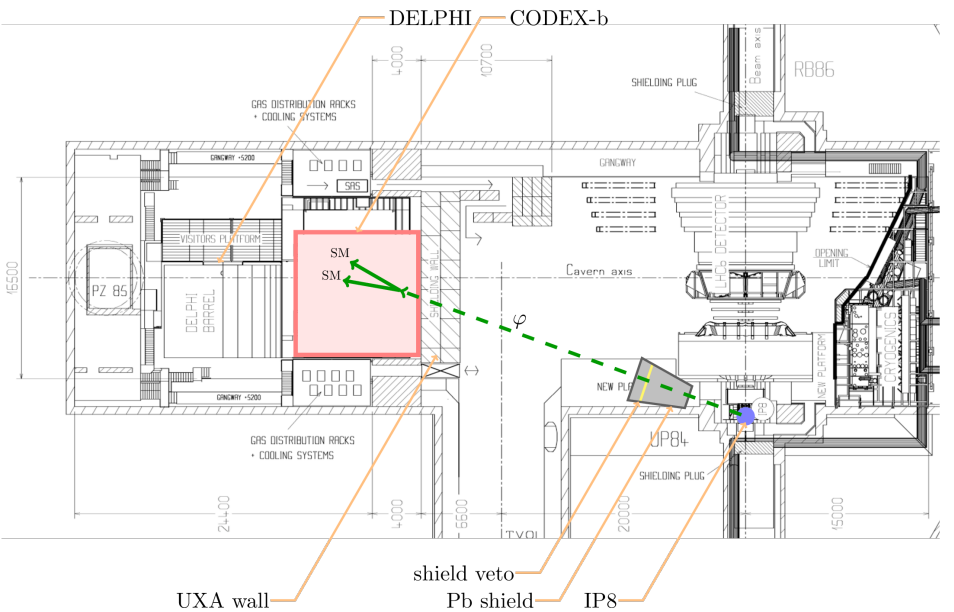


Fig. 6. Layout of the LHCb cavern UX85 at IP8, overlaid with the CODEX-b volume.[53]

CODEX-b will be interfaced with LHCb data acquisition, therefore there will have a unique trigger. Further additions to CODEX-b include calorimetry, which would significantly enhance the physics reach, e.g. photon signatures. Moreover, absorber or pre-shower layers can also perform particle identification, e.g. e/γ separation.

The CODEX-β demonstrator, with a volume of $2 \times 2 \times 2\ \text{m}^3$, is proposed to be operated during Run 3 to provide proof of concept. Its primary goal is to be integrate with LHCb online, reconstruct K_L^0 and measure background rates. It will provide competitive sensitivity to $b \to s\chi(\to \text{hadrons})$, as well.

2.6. *MATHUSLA – MAsive Timing Hodoscope for Ultra Stable neutraL pArticles*

MATHUSLA[54,55] is a proposed surface detector of a large footprint (area 100 × 100 m^2) and large decay volume (height 25 m) to be located above CMS,[6] as depicted in Fig. 7. The ∼ 90 m of rock between the IP5 and the detector decay volume provides enough shielding for MATHUSLA to work in a clean environment. The air-filled decay volume will be occupied by several detector layers for tracking in a modular way. RPCs and plastic scintillators are proven technologies that meet the specifications. Being a background-free experiment increases the sensitivity to LLPs up to decay lengths of 10^7 m and extends the sensitivity of the main detectors by orders of magnitude.

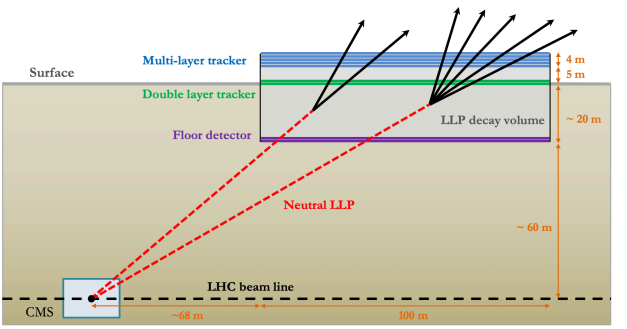


Fig. 7. The MATHUSLA detector layout positioned relative to CMS IP5.[55]

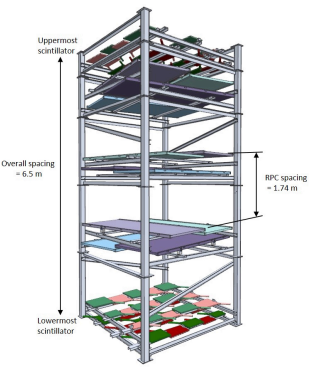


Fig. 8. The MATHUSLA test stand.[56]

A test stand following the concept of the full detector was installed in the surface area above the ATLAS[5] IP, taking data with different beam conditions during 2018.[56] The 2.5 × 2.5 × 6.5 m^3 unit detector is composed of one external layer of scintillators in the upper part and one in the lower part with six layers of RPCs between them, as shown in Fig. 8. The obtained results confirmed the background assumptions in the MATHUSLA proposal.

Moreover, MATHUSLA could act as a cosmic-ray (CR) telescope performing very precise measurements up to the PeV scale.[57,58] By integrating a device with the possibility to measure arrival times and particle densities of extensive air showers, such as an RPC, MATHUSLA can be employed as a CR detector and monitor a big portion of the sky above ($\theta < 80°$), without limitation to inclined events.

2.7. *ANUBIS – AN Underground Belayed In-Shaft*

ANUBIS[59] is an off-axis detector designed for neutral LLPs with $c\tau \gtrsim 5$ m, proposed to occupy the PX14 installation shaft of the ATLAS[5] experiment (cf. Fig. 9(left)),

which is not used during regular LHC operation. It will comprise four evenly spaced tracking stations (TS) with a cross-sectional area of 230 m^2 each, shown in Fig. 9(right). The tracking stations will use the same RPC technology as the new ATLAS layers and ATLAS can be used as an active veto of SM activity. The projective decay volume optimises the acceptance for different lifetimes. Two smaller TS prototypes are planned to be installed for Run 3 as a demonstrator.

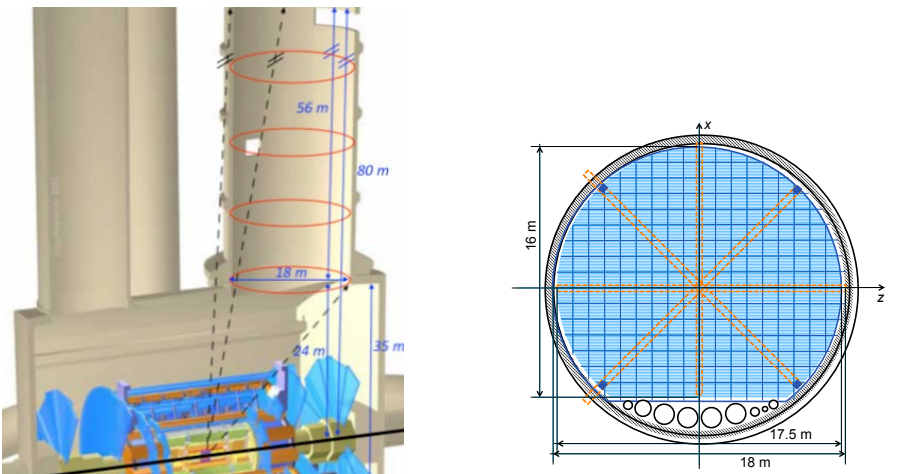


Fig. 9. Left: Location of the ANUBIS detector in the PX14 installation shaft of the ATLAS experiment. Above: One of the four ANUBIS tracking stations in the (x, z) plane. The shaft walls and the ATLAS cavern pipework are shown in gray, the TS in blue and the TS support structure in orange.[59]

2.8. *AL3X – A Laboratory for Long-Lived eXotics*

In the event that ALICE[60] finishes its physics program before the end of HL-LHC, it has been proposed[61] to reuse the L3 magnet at IP2 and the ALICE Time Projection Chamber (TPC) for LLP searches. This proposal requires to move the IP by 11.25 m outside the magnet, as shown in Fig. 10, to allow LLPs to travel before decaying. An additional thick shield and an active veto (D_1 and D_3) will reduce the background. The detector geometry corresponds to a pseudorapidity acceptance $0.9 < \eta < 3.7$. The AL3X configuration is in essence a tracking detector behind a heavy shield, which can be thought of as analogous to a calorimeter that is solely absorber. This permits AL3X to search for much rarer signals in a very low background environment compared to ATLAS and CMS, and in this sense AL3X would be complementary to the existing (and proposed upgraded) multi-purpose detectors.

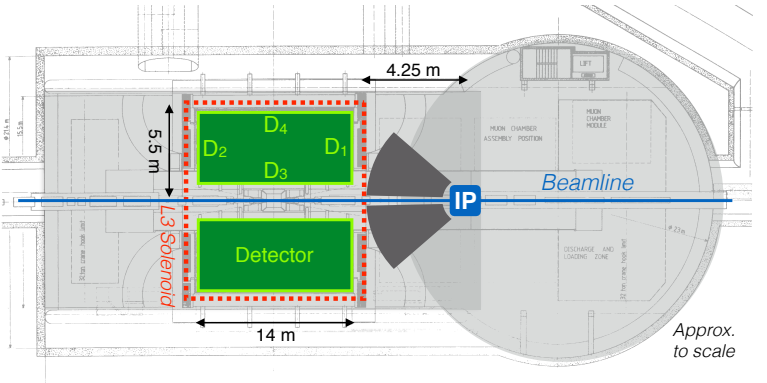


Fig. 10. Schematic view of the AL3X detector layout: the cylindrical TPC (dark green) and the veto and trigger layers D_i (light green). The current L3 magnet is shown (dashed red) for reference.[61]

2.9. *milliQan*

Just like MAPP-mQP, the milliQan detector,[62,63] to be installed near the CMS[6] IP5, will search for stable *millicharged* particles, unlike the majority of the aforementioned detectors which are sensitive to visible SM particles originating from *neutral* DM decays. The milliQan is located in an underground tunnel at a distance of 33 m from the CMS IP, with 17 m of rock between the IP and the detector that provides shielding from most particles produced in LHC collisions. In order to be sensitive to particles with charges as low as $0.001e$, a large active area of scintillator is required. For Run 3, two detector designs are planned for deployment. The bar detector is made of 0.2 m $\times$ 0.2 m $\times$ 3 m plastic scintillator bars surrounded by an active μ veto shield. The slab detector will increase the reach for heavier mCPs through 40 cm $\times$ 60 cm $\times$ 5 cm scintillator slabs.[64]

A small fraction of $\sim 1\%$ of the full detector, the *milliQan demonstrator,* shown in Fig. 11 was installed and operated to measure backgrounds and provide proof of principle. It allowed the first search for millicharged particles at a collider.[65] A data sample of 37.5 fb^{-1} pp collisions at $\sqrt{s} = 13$ TeV has been analysed and no excess over the background prediction observed. The results interpretation in terms dark photons is discussed later in Sec. 5.

2.10. *FORMOSA – FORward MicrOcharge SeArch*

FORMOSA[66] is also designed to discover millicharged particles, yet in the far-forward direction, close to the beam collision axis, where it can benefit from an enhanced mCP production cross section compared to the transverse direction. It is a scintillator-based detector to be hosted in the FPF, shown in Fig. 12. FORMOSA is proposed to start in Run 3 by moving the milliQan demonstrator to UJ12, as a phase called FORMOSA-I, with the full milliQan-type detector, FORMOSA-II, to be deployed at a later stage. The expected physics potential is discussed in Sec. 5.

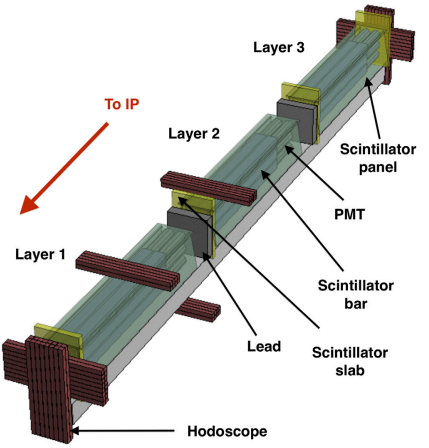


Fig. 11. Diagram of the milliQan demonstrator components.[65]

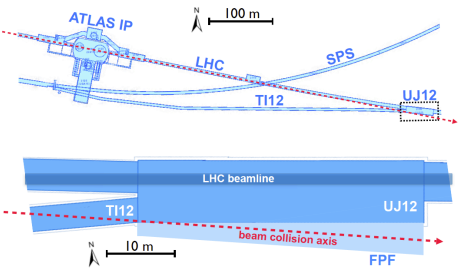


Fig. 12. Location of FORMOSA in the cavern UJ12 or side tunnel TI12 (blue) close to the beam collision axis (red). The FPF extension is shown as a light-blue area.[66]

3. The Higgs Portal

The so-called *scalar portal* involves a new dark scalar S, which mixes slightly with the SM-like Higgs, and provides a simple target for new physics searches.[67] The dark Higgs mixing portal admits exotic inclusive $B \to X\phi$ decays,[a] where ϕ is a light CP-even scalar that mixes with the SM Higgs, with a mixing angle of $\theta \ll 1$. The particle lifetime depends on the degree of mixing. One possible, simple Lagrangian which includes this new dark Higgs mixing is given by[68]

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{DS}} + \mu_S^2 S^2 - \frac{\lambda_S}{4} S^4 - \epsilon S^2 H^\dagger H, \tag{1}$$

where S is a real scalar field, H is a SM-like Higgs field, ϵ is the portal coupling, and λ_S is a free parameter. The quartic term contains the mixing between the SM Higgs and the new scalar, with the resulting physical fields: the SM Higgs h and the dark Higgs ϕ. Both fields acquire a non-zero VEV and the coupling between these two particles induces new Yukawa-like couplings between the dark Higgs and the SM fermions. In addition, there can appear a non-negligible trilinear interaction term between ϕ and h with the corresponding coupling denoted by λ, i.e. $\propto \lambda h\phi\phi$. Thus, the signal sought after in LLP experiments is two charged lepton tracks originating from dark Higgs decays $\phi \to \ell^+\ell^-$, in their fiducial volume.

Currently the best experimental limits on dark Higgs production at colliders come from CHARM[69] and LHCb,[70,71] shown in Fig. 13 in the $\sin^2\theta$-versus-m_S plane. If no trilinear coupling is assumed and the background is considered negligible, we derive the exclusion curves for MATHUSLA, SHiP, CODEX-b, MAPP-1

[a] D mesons and kaons have much smaller branching ratios into a Higgs-mixed scalar and are neglected.

and MAPP-2 that are shown in the same figure. MAPP-1 during Run 3 (30 fb^{-1}) is expected to provide considerable coverage, when the other experiments will be under construction; from the CODEX-b side, the CODEX-β demonstrator will be in operation. MAPP-2 extents significantly MAPP-1 sensitivity making it competitive with SHiP. As expected, MATHUSLA has better sensitivity in lower mixing angles, thus longer lifetimes, due to its large distance from the IP.

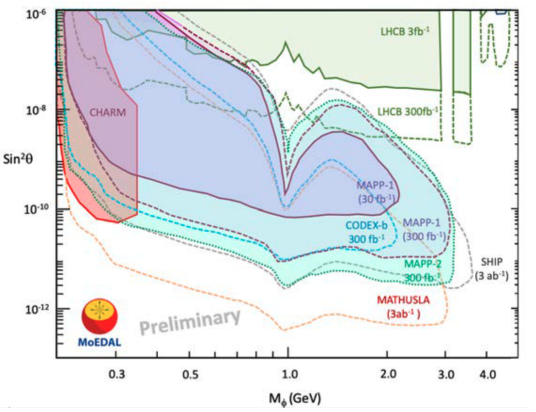


Fig. 13. 95% CL exclusion bounds Projected sensitivity of MoEDAL-MAPP and other experiments for dark Higgs bosons produced in rare B decays at $\sqrt{s} = 14$ TeV. Adopted from Ref. 61.

Assuming now a non-vanishing trilinear coupling λ, the expected sensitivity from the experiments presented in Sec. 2 is shown in Fig. 14. This time, a larger fraction of the parameters space is covered due to the contributions arising from a Higgs in the decay cascade, either virtual, e.g. the $B \to KSS$ mode, or real, e.g. the $h \to SS$. The branching fraction $BR(h \to SS) \sim 10^{-2}$ is considered to remain compatible with the LHC searches for the Higgs to invisible channels. The larger impact is provided by the bigger experiments, MATHUSLA,[72] SHiP,[73] FASER2[68] and CODEX-b,[74] which can explore the region well above the GeV mass scale in a fully uncharted range of couplings.

Dark scalar portals may have cosmological implications, contributing to the observed DM abundance. Dark Higgs bosons may mediate interactions with hidden dark matter that has the correct thermal relic density or resolves small-scale-structure discrepancies. Indeed, there are models with a dark Higgs inflaton strongly favoured by cosmological Planck[75] and BICEP/Keck Array[76] data that constrain the energy scale of inflation. One of these is expected to leave imprints on LHC experiments, such as FASER and MAPP-1.[77]

4. Dark Photons: Neutral Metastable States

A large class of BSM models includes interactions with light new *vector* particles. Such particles can result from additional gauge symmetries of BSM physics. New vector states can mediate interactions between the SM fields and extra dark-sector fields that may eventually play the role of the DM states.

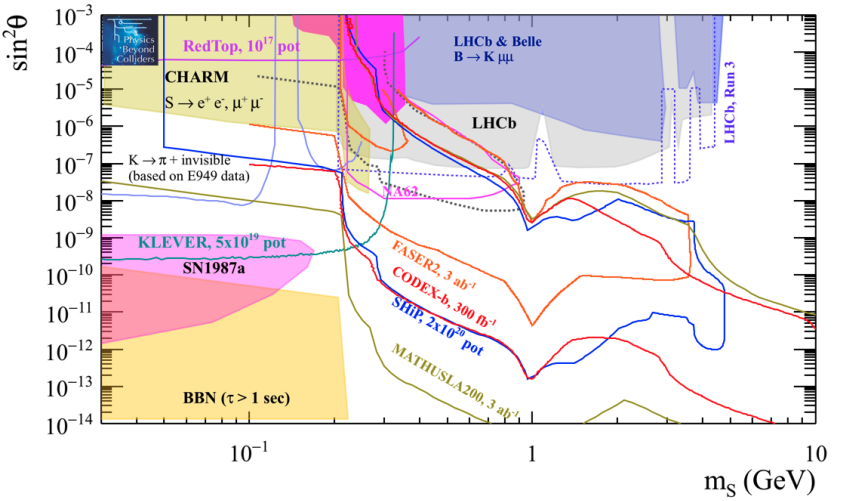


Fig. 14. Prospects on 10-15 year timescale for dark-scalar mixing with Higgs in the $(m_S, \sin^2\theta)$ plane with trilinear coupling $\lambda \neq 0$ and assuming $BR(h \to SS) = 10^{-2}$. The NA62^{++} and KLEVER curves correspond to $\lambda = 0$, so they are conservative.[74]

The most minimal vector portal interaction can be written as

$$\mathcal{L}_{\text{vector}} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{DS}} - \frac{\epsilon}{2\cos\theta_W} F'_{\mu\nu} B^{\mu\nu}, \tag{2}$$

where $\mathcal{L}_{\text{SM}}$ is the SM Lagrangian, $B_{\mu\nu}$ and $F'_{\mu\nu}$ are the field stengths of hypercharge and new $U'(1)$ gauge groups, ϵ is the so-called *kinetic mixing* parameter,[78] and $\mathcal{L}_{\text{DS}}$ stands for the dark sector Lagrangian that may include new matter fields χ charged under $U'(1)$,

$$\mathcal{L}_{\text{DS}} = -\frac{1}{4} F'_{\mu\nu} F'^{\mu\nu} + \frac{1}{2} m^2_{A'} A'_\mu A'^\mu + |(\partial_\mu + i g_D A'_\mu)\chi|^2 + \ldots \tag{3}$$

If χ is (meta)stable, it may constitute a fraction or entirety of dark matter. At low energy this theory contains a new massive vector particle, a *dark photon* state, coupled to the electromagnetic current with ϵ-proportional strength, $\epsilon A'_\mu J^\mu_{\text{EM}}$.

In the minimal dark photon model, DM is assumed to be either heavy or contained in a different sector. The dark photon, γ_{d}, decays into SM states (visible decays). The physics potential of the proposed LLP experiments as a function of the dark photon mass $m_{A'}$ and the coupling of dark photon with the SM photon ϵ is shown in Fig. 15, compared with existing bounds from several beam-dump data, e.g. Refs. 79, 80. The sensitivity for dark photons decaying into visible final states is expected to be dominated by SHiP,[73] while FASER2,[68] LDMX[81] and AWAKE[82] will be directly competing with SeaQuest, LHCb, Heavy Photon Search (HPS)[83] experiment, and others. MATHUSLA in this scenario is however not competitive, mostly due to the fact that the dark photon is produced in the forward direction.

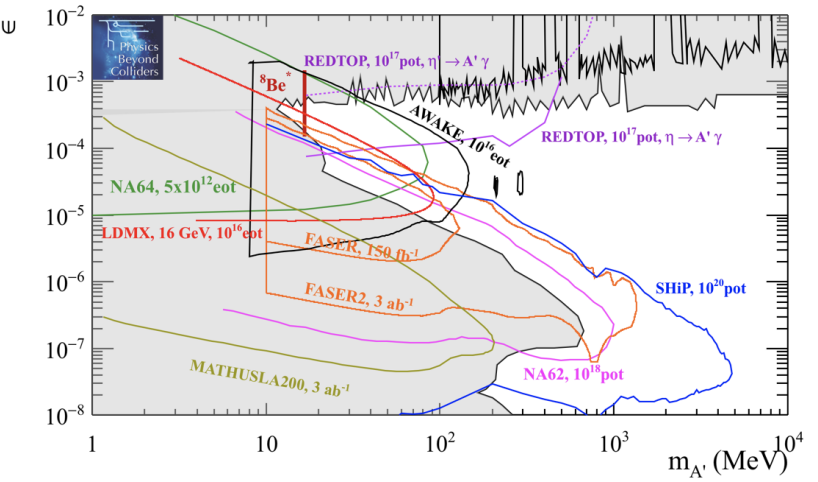


Fig. 15. Future upper limits at 90% CL for a minimal dark-photon model with visible decays in the plane mixing strength ϵ versus mass $m_{A'}$ for various projects on a ∼10–15 year timescale.[74]

In a different scenario, light (sub-GeV) DM, χ, may be coupled to a dark photon, constituting a minimally coupled weakly interacting massive particle (WIMP) model. The preferred values of the coupling g_D are such that $A' \to \chi\chi$, with χ potentially scattering further on electrons and nuclei, while $m_{A'}/m_\chi = 3$ is assumed for the masses. Recent studies on DM scattering off electrons[48] and off nuclei[84] in an emulsion detector (FASERν2) and a LArTPC (FLArE) in the FPF showed very interesting sensitivity, probing the thermal-relic region. Neutrino background may be separated from DM signal using energy and angle selection criteria. Such forward detectors open the possibility for performing a direct-detection-type DM search at the LHC.

In inelastic DM, which constitutes a viable and compelling paradigm for light thermal DM, DM couples to the SM only by interacting with a nearly degenerate dark-photon mediator heavier than $\sim$ 10 GeV. For relative mass-splittings larger than $\mathcal{O}(10^{-6})$, DM-nucleon/electron scattering at direct detection experiments is kinematically suppressed. However, at the LHC, where the DM and excited state can be directly produced, for mass-splittings above a few MeV, the excited state can decay back to DM and a pair of SM fermions, often on collider timescales. giving rise to visible displaced vertices. ATLAS, CMS, LHCb, CODEX-b, FASER, and MATHUSLA can detect such DM signals in the cosmologically motivated mass range of $\sim 1 - 100$ GeV.[85]

5. Dark Photons: Stable Millicharged Particles

Millicharged particles have been discussed in connection with the mechanism of electric charge quantisation and possible non-conservation of electric charge.[86] There are three experiments planned to run at the LHC that are sensitive to the detection

of the low ionisation coming from an mCP: milliQan (Sec. 2.9), the MAPP-mQP sub-detector (Sec. 2.1) and, the recently proposed, FORMOSA (Sec. 2.10).

A well-motivated mechanism that predicts mCPs is the introduction of a new massless $U'(1)$ gauge field, the dark photon, $A'_{\mu\nu}$, which is coupled to the SM photon field, $B^{\mu\nu}$. A new massive dark fermion ψ *(dark QED)* of mass M_{mCP}, is predicted, which is charged under the new $U'(1)$ field A' with charge e'. The Lagrangian for the model is given by[62]

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} - \frac{1}{4}A'_{\mu\nu}A'^{\mu\nu} + i\bar{\psi}(\partial\!\!\!/ + ie'A\!\!\!/' + iM_{\mathrm{mCP}})\psi - \frac{\kappa}{2}A'_{\mu\nu}B^{\mu\nu}. \tag{4}$$

The last term contains the kinetic mixing, which one can eliminate by expressing the new gauge boson as, $A'_\mu \to A'_\mu + \kappa B_\mu$. Applying this field redefinition reveals a coupling between the charged matter field ψ to the SM hypercharge. The Lagrangian (4) then becomes:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} - \frac{1}{4}A'_{\mu\nu}A'^{\mu\nu} + i\bar{\psi}(\partial\!\!\!/ + ie'A\!\!\!/' - i\kappa e'B\!\!\!/ + iM_{\mathrm{mCP}})\psi. \tag{5}$$

It is now apparent that the field ψ acts as a field charged under hypercharge with a millicharge $\kappa e'$, which couples to the photon and Z^0 boson with a charge $\kappa e' \cos\theta_W$ and $-\kappa e' \sin\theta_W$, respectively. Expressing the fractional charge in terms of electric charge thus gives $\epsilon = \kappa e' cos\theta_W / e$.

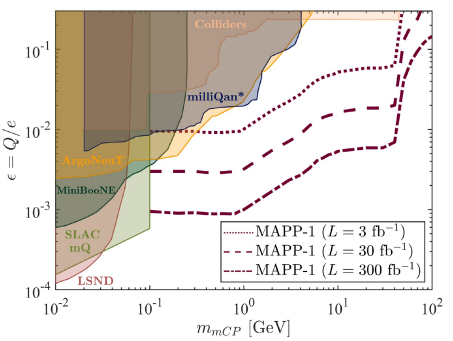


Fig. 16. 95% CL exclusion limits for dark fermion mCPs in the mass vs. charge plane obtained by the milliQan demonstator[65] compared to previous constraints and to the MAPP-mQP sensitivity for various integrated luminosity assumptions.[87]

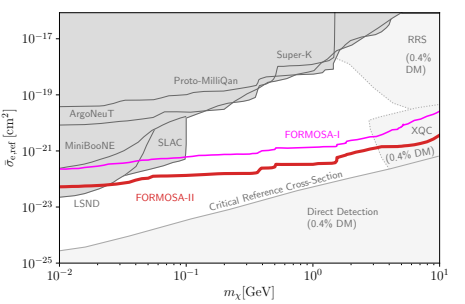


Fig. 17. Sensitivity reaches of FORMOSA in the millicharged SIDM window in terms of the reference cross-section $\bar{\sigma}_{\mathrm{e,ref}}$. In addition to accelerator constraints, constraints from direct-detection experiments (assuming 0.4% DM abundance for the direct-detection experiments) are drawn.[66]

The milliQan 1% scale demonstrator discussed in Sec. 2.9 and shown in Fig. 11 provided the first constraints on mCPs at LHC, as shown in Fig. 16. Pair-production of millicharged particles at 13 TeV is considered through the Drell-Yan process, as well as from Υ, J/ψ, $\psi(2S)$, ϕ, ρ, and ω decays into mCP pairs, and from Dalitz decays of π^0, η, η', and ω. This search excluded mCP masses of $20 - 4700$ MeV

for charges varying between $0.006e$ and $0.3e$ depending on mass.[65] MAPP-mQP, on the other hand, can extend these limits especially towards higher masses, while the reach in terms of low charges depends on the integrated luminosity; charges as low as $10^{-3}e$ will need the HL-LHC,[87] as evident from Fig. 16. Similar sensitivity is expected by the full milliQan detector during the HL-LHC run.[64]

When considering DM abundance, mCPs can account for a fraction of it that cannot be detected by direct-detection experiments due to attenuation effects, when the ambient DM has substantial cross section with SM particles making it to lose most of its kinetic energy. For some model parameters,[88,89] the DM particles cannot be detected by ground-based direct detection experiments after interacting with the atmospheric particles and the crust. These DM particles are generally referred to as strongly interacting DM (SIDM). In general, mCP-hunting experiments constrain the larger cross sections, leaving a gap, that can be filled by FORMOSA.[66] This is shown in Fig. 17, where the accessible *reference cross section* versus mCP mass is drawn for accelerator and for direct-detection experiments, together with the FORMOSA expected reach.

In addition to DM, apparent fractionally charged particles may arise as heavy neutrinos with a large enough electric dipole moment (EDM) to yield ionisation in mCP-sensitive detectors. It has been demonstrated that milliQan and MAPP-mQP have very good sensitivity in a scenario where the heavy neutrino is considered to be a member of a fourth generation lepton doublet with the EDM introduced within a dimension-five operator.[90,91]

6. More Portals and Beyond

As highlighted in the introduction in Sec. 1, besides the hidden sectors directly related to DM, namely the dark Higgs and dark photons, there are other portals, such as those predicting heavy neutral leptons and axion-like particles (ALPs). The quest for these objects is of equal interest to Particle Physics open questions, they are often related to dark matter and are similarly relevant to the above mentioned LLP experiments, as those discussed in Secs. 3, 4 and 5. Besides the portal connection to DM, LLPs can probe specific DM models, such as the case of MATHUSLA and dynamical dark matter.[92]

6.1. *Extended neutrino sector*

In the so-called *fermion portal*, new heavy neutral leptons (HNLs) are added to the SM to provide an elegant way to generate non-zero neutrino masses via the seesaw mechanism.[93–95] In type-I seesaw models, one extends the SM by adding neutral right-handed fermions (identified with HNLs) that couple to the SM neutrinos similarly to the coupling between left- and right-handed components of the charged leptons. In a considerable class of these models, the HNLs become long lived and the LLP experiments have good prospects to detect them.[74,96–100]

To highlight an example, in the case of right-handed neutrinos being produced in the decay of an additional Z' boson in the gauged $B-L$ model, which also contains a singlet Higgs that spontaneously breaks the extra $U(1)_{B-L}$ gauge symmetry, MAPP-2 will fill the gap left by CMS, LHCb, MATHUSLA, FASER2 and CODEX-b.[101]

Light sterile neutrinos can account for dark matter,[102–104] while sterile neutrinos with a broad range of masses can account for the baryon asymmetry of the Universe through leptogenesis.[105] Sterile neutrinos may be long-lived in simplified models where the SM is extended with one sterile neutrino[106] or in neutrino-extended SM Effective Field Theories, νSMEFT. Intermediate-mass can be produced in leptonic and semi-leptonic decays of charmed and bottomed mesons, decaying to leptons via neutral and charged weak currents, thus becoming detectable in LHC LLP experiments.[107]

6.2. *Axions & axion-like particles*

Axions, or more generally ALPs, are pseudoscalar pseudo-Nambu-Goldstone bosons arising from approximate Abelian global symmetries beyond the SM which are broken spontaneously at a scale much greater than the electroweak scale. Axions, in particular, were postulated by the Peccei-Quinn theory to resolve the strong CP problem in quantum chromodynamics. ALPs also provide an interesting connection to the puzzle of dark matter, because they can mediate the interactions between the DM particle and SM states and allow for additional annihilation channels relevant for the thermal freeze-out of DM. Prospects for searches for ALPs in LLP experiments are reviewed in Refs. 2, 4, 43, 53, 57, 73, 74, 108.

6.3. *Supersymmetry*

Going even further, supersymmetry (SUSY), a theoretical framework that provides a natural DM candidate, predicts the existence of LLPs. For instance, sleptons, charginos and R-hadrons (namely gluinos, top- or bottom-squarks) with unit electric charge may be detected in the MoEDAL NTDs, in the case they are sufficiently long lived to reach the detector.[109–111]

R-parity violating SUSY also predicts LLPs, such as light long-lived neutralinos $\tilde{\chi}_1^0$ decaying via λ'_{ijk} couplings to charged particles. Benchmark scenarios related to either charm or bottom mesons decaying into $\tilde{\chi}_1^0$ have been considered, in similar fashion as in sterile neutrinos, showing that these experiments can cover various meson production and decay modes and $\tilde{\chi}_1^0$ lifetimes.[53, 98, 112, 113]

7. Post-Discovery Matters

The variety of BSM scenarios predicting LLPs and the sensitivity to those by dedicated LLP experiments — and main detectors equally — begs the question regarding the power to provide qualitative and even quantitative information about the *nature*

of these LLPs in the post-discovery phase. The definition of *LLP simplified models*[2] serves as a bridge between specific theoretical models and particular experimental signatures.

The geometry of two-body γ_d/ϕ decays to massless final states can provide information about their velocity and the ability to discriminate between different γ_d/ϕ masses. CODEX-b can reconstruct the γ_d/ϕ velocity to better than 1% using spacial tracking information. If RPC timing information is used in a complementary way to discriminate between slow-moving new states, it is possible to separate between ϕ masses of 0.5 GeV and 2.0 GeV.[52]

If the LLP is pair-produced in Higgs boson decays, it is possible to measure the mass of this particle and determine the dominant decay mode with less than 100 observed events in MATHUSLA. In more general cases, the latter should be able to distinguish the production mode of the LLP and to determine its mass and spin based on the decay products of the long-lived particle.[114]

Moreover, if information from the LLP and the main experiment of the same IP, e.g. MATHUSLA and CMS, is combined, the LLP production mode topology could be determined with as few as ~ 100 observed LLP decays. Underlying theory parameters, like the LLP and parent particle masses, can also be measured with $\lesssim 10\%$ precision.[115]

8. Summary and Outlook

There is an ever increasing interest in long-lived particle searches at the LHC (and not only) to exlore the dark sector besides other unanswered questions in Particle Physics today. Besides the efforts in the main experiments ATLAS, CMS and LHCb, additional complementary experiments have been approved (MoEDAL, FASER, SND@LHC), aiming at data taking in Run 3, or have been proposed (MATHUSLA, CODEX-b, milliQan, ANUBIS, AL3X and others). In addition to these LHC detectors, the SHiP experiment at the SPS is also planned to explore hidden (dark) sectors. Several of these experiments have constructed, operated and analysed the data from small-scale demonstrators that proved the detector concept and provided the first encouraging (physics) results. Others plan to install such prototype for (part of) Run 3, before preparing a complete detector for the HL-LHC operation. The enigma of dark matter is challenged in a different perspective — the lifetime frontier — at the LHC Run 3 and beyond.

Acknowledgements

The author would like to thank the MG16 Meeting organisers for the kind invitation to present this talk. This work was supported in part by the Generalitat Valenciana via a special grant for MoEDAL and via the Project PROMETEO-II/2017/033, and by the Spanish MICIU / AEI and the European Union / FEDER via the grant PGC2018-094856-B-I00.

References

1. L. Evans and P. Bryant, *JINST* **3**, S08001 (2008).
2. J. Alimena *et al.*, *J. Phys. G* **47**, 090501 (2020), arXiv:1903.04497 [hep-ex].
3. L. Lee, C. Ohm, A. Soffer and T.-T. Yu, *Prog. Part. Nucl. Phys.* **106**, 210 (2019), arXiv:1810.12602 [hep-ph].
4. P. Agrawal *et al.* (2 2021), arXiv:2102.12143 [hep-ph].
5. ATLAS Collaboration, G. Aad *et al.*, *JINST* **3**, S08003 (2008).
6. CMS Collaboration, S. Chatrchyan *et al.*, *JINST* **3**, S08004 (2008).
7. V. A. Mitsou, *Int. J. Mod. Phys.* **A28**, 1330052 (2013), arXiv:1310.1072 [hep-ex].
8. LHCb Collaboration, A. A. Alves, Jr. *et al.*, *JINST* **3**, S08005 (2008).
9. M. Borsato *et al.* (5 2021), arXiv:2105.12668 [hep-ph].
10. BaBar Collaboration, J. P. Lees *et al.*, *Phys. Rev. Lett.* **119**, 131804 (2017), arXiv:1702.03327 [hep-ex].
11. NA48/2 Collaboration, J. R. Batley *et al.*, *Phys. Lett. B* **746**, 178 (2015), arXiv:1504.00607 [hep-ex].
12. NA62 Collaboration, E. Cortina Gil *et al.*, *JHEP* **03**, 058 (2021), arXiv:2011.11329 [hep-ex].
13. D. Banerjee *et al.*, *Phys. Rev. Lett.* **123**, 121801 (2019), arXiv:1906.00176 [hep-ex].
14. A. Berlin, S. Gori, P. Schuster and N. Toro, *Phys. Rev. D* **98**, 035011 (2018), arXiv:1804.00661 [hep-ph].
15. SoLid Collaboration, N. Roy, *PoS* **ICHEP2020**, 775 (2021).
16. SHiP Collaboration, M. Anelli *et al.* (4 2015), arXiv:1504.04956 [physics.ins-det].
17. W. Baldini *et al.* (10 2021), arXiv:2110.08025 [hep-ex].
18. Z. Bai *et al.* (7 2021), arXiv:2107.13554 [hep-ph].
19. M. Chrzaszcz, M. Drewes and J. Hajer, *Eur. Phys. J. C* **81**, 546 (2021), arXiv:2011.01005 [hep-ph].
20. S. Dreyer *et al.*, *Physics reach of a long-lived particle detector at Belle II* (5 2021), arXiv:2105.12962 [hep-ph].
21. S. Choi *et al.* (7 2020), arXiv:2007.06329 [physics.ins-det].
22. K. J. Kelly and Y.-D. Tsai, *Phys. Rev. D* **100**, 015043 (2019), arXiv:1812.03998 [hep-ph].
23. MoEDAL Collaboration, J. Pinfold *et al.*, *Technical Design Report of the MoEDAL Experiment*, CERN-LHCC-2009-006 (2009).
24. A. De Roeck, A. Katre, P. Mermod, D. Milstead and T. Sloan, *Eur. Phys. J.* **C72**, 1985 (2012), arXiv:1112.2999 [hep-ph].
25. N. E. Mavromatos and V. A. Mitsou, *Int. J. Mod. Phys. A* **35**, 2030012 (2020), arXiv:2005.05100 [hep-ph].
26. M. Fairbairn, A. C. Kraan, D. A. Milstead, T. Sjostrand, P. Z. Skands and T. Sloan, *Phys. Rept.* **438**, 1 (2007), arXiv:hep-ph/0611040 [hep-ph].
27. MoEDAL Collaboration, B. Acharya *et al.*, *Int. J. Mod. Phys.* **A29**, 1430050 (2014), arXiv:1405.7662 [hep-ph].
28. V. A. Mitsou, *PoS* **CORFU2019**, 009 (2020).
29. MoEDAL Collaboration, B. Acharya *et al.*, *JHEP* **08**, 067 (2016), arXiv:1604.06645 [hep-ex].
30. MoEDAL Collaboration, B. Acharya *et al.*, *Phys. Rev. Lett.* **118**, 061801 (2017), arXiv:1611.06817 [hep-ex].
31. MoEDAL Collaboration, B. Acharya *et al.*, *Phys. Lett.* **B782**, 510 (2018), arXiv:1712.09849 [hep-ex].
32. MoEDAL Collaboration, B. Acharya *et al.*, *Phys. Rev. Lett.* **123**, 021802 (2019), arXiv:1903.08491 [hep-ex].

33. MoEDAL Collaboration, B. Acharya *et al.*, *Phys. Rev. Lett.* **126**, 071801 (2021), arXiv:2002.00861 [hep-ex].
34. MoEDAL Collaboration, B. Acharya *et al.* (6 2021), arXiv:2106.11933 [hep-ex].
35. MoEDAL Collaboration, B. Acharya *et al.*, *Search for highly ionizing particles in pp collisions at the LHC's Run-1 using the prototype MoEDAL detector*, in preparation.
36. V. A. Mitsou, *Results and future plans of the MoEDAL experiment*, in *European Physical Society Conference on High Energy Physics 2021* (2021). arXiv:2111.03468 [hep-ex].
37. J. L. Pinfold, *Phil. Trans. Roy. Soc. Lond.* **A377**, 20190382 (2019).
38. J. L. Pinfold, *Universe* **5**, 47 (2019).
39. M. Staelens, *Recent results and future plans of the MoEDAL experiment*, in *Meeting of the Division of Particles and Fields of the American Physical Society* (10 2019). arXiv:1910.05772 [hep-ex].
40. G. Apollinari, O. Brüning, T. Nakamoto and L. Rossi, *CERN Yellow Rep.*, 1 (2015), arXiv:1705.08830 [physics.acc-ph].
41. J. Kieseler, J. Alimena, J. Simms, T. Aarrestad, M. Pierini and A. Kish (10 2021), arXiv:2110.13837 [hep-ph].
42. J. L. Feng, I. Galon, F. Kling and S. Trojanowski, *Phys. Rev. D* **97**, 035001 (2018), arXiv:1708.09389 [hep-ph].
43. FASER Collaboration, A. Ariga *et al.* (1 2019), arXiv:1901.04468 [hep-ex].
44. FASER Collaboration, M. Queitsch-Maitland, *PoS* **ICHEP2020**, 273 (2021).
45. FASER Collaboration, H. Abreu *et al.*, *Eur. Phys. J. C* **80**, 61 (2020), arXiv:1908.02310 [hep-ex].
46. FASER Collaboration, H. Abreu *et al.* (5 2021), arXiv:2105.06197 [hep-ex].
47. FASER Collaboration, H. Abreu *et al.*, *FASERν2: A forward neutrino experiment at the HL LHC,* https://www.snowmass21.org/docs/files/summaries/NF/SNOWMASS21-NF10_NF6-EF6_EF9-IF0_FASERnu2-006.pdf (7, 2020), Expression of Interest: Snowmass NF10_NF6-EF6_EF9-IF0.
48. B. Batell, J. L. Feng and S. Trojanowski, *Phys. Rev. D* **103**, 075023 (2021), arXiv:2101.10338 [hep-ph].
49. SHiP Collaboration, C. Ahdida *et al.* (2 2020), arXiv:2002.08722 [physics.ins-det].
50. SHiP Collaboration, C. Ahdida *et al.*, *JINST* **14**, P03025 (2019), arXiv:1810.06880 [physics.ins-det].
51. FACET Collaboration, D. Cerci *et al.*, *A Long-Lived Particle and Dark Matter Search at the LHC at $z = 80-127$ m* https://www.snowmass21.org/docs/files/summaries/EF/SNOWMASS21-EF9_EF8_ALBROW-111.pdf (8, 2020), Expression of Interest: Snowmass EF08+09+10.
52. V. V. Gligorov, S. Knapen, M. Papucci and D. J. Robinson, *Phys. Rev. D* **97**, 015023 (2018), arXiv:1708.09395 [hep-ph].
53. G. Aielli *et al.*, *Eur. Phys. J. C* **80**, 1177 (2020), arXiv:1911.00481 [hep-ex].
54. MATHUSLA Collaboration, H. Lubatti *et al.*, *JINST* **15**, C06026 (2020), arXiv:1901.04040 [hep-ex].
55. MATHUSLA Collaboration, C. Alpigiani *et al.* (9 2020), arXiv:2009.01693 [physics.ins-det].
56. M. Alidra *et al.* (5 2020), arXiv:2005.02018 [physics.ins-det].
57. D. Curtin *et al.*, *Rept. Prog. Phys.* **82**, 116201 (2019), arXiv:1806.07396 [hep-ph].
58. C. Alpigiani, *JINST* **15**, C09048 (2020), arXiv:2006.00788 [physics.ins-det].
59. M. Bauer, O. Brandt, L. Lee and C. Ohm (9 2019), arXiv:1909.13022 [physics.ins-det].
60. ALICE Collaboration, K. Aamodt *et al.*, *JINST* **3**, S08002 (2008).

61. V. V. Gligorov, S. Knapen, B. Nachman, M. Papucci and D. J. Robinson, *Phys. Rev. D* **99**, 015023 (2019), arXiv:1810.03636 [hep-ph].
62. A. Haas, C. S. Hill, E. Izaguirre and I. Yavin, *Phys. Lett. B* **746**, 117 (2015), arXiv:1410.6816 [hep-ph].
63. A. Ball *et al.* (7 2016), arXiv:1607.04669 [physics.ins-det].
64. milliQan Collaboration, A. Ball *et al.*, *Phys. Rev. D* **104**, 032002 (2021), arXiv:2104.07151 [hep-ex].
65. A. Ball *et al.*, *Phys. Rev. D* **102**, 032002 (2020), arXiv:2005.06518 [hep-ex].
66. S. Foroughi-Abari, F. Kling and Y.-D. Tsai, *Phys. Rev. D* **104**, 035014 (2021), arXiv:2010.07941 [hep-ph].
67. B. Patt and F. Wilczek (2006), arXiv:hep-ph/0605188 [hep-ph].
68. FASER Collaboration, A. Ariga *et al.*, *Phys. Rev. D* **99**, 095011 (2019), arXiv:1811.12522 [hep-ph].
69. CHARM Collaboration, F. Bergsma *et al.*, *Phys. Lett. B* **157**, 458 (1985).
70. LHCb Collaboration, R. Aaij *et al.*, *Phys. Rev. Lett.* **115**, 161802 (2015), arXiv:1508.04094 [hep-ex].
71. LHCb Collaboration, R. Aaij *et al.*, *Phys. Rev. D* **95**, 071101 (2017), arXiv:1612.07818 [hep-ex].
72. J. A. Evans, *Phys. Rev. D* **97**, 055046 (2018), arXiv:1708.08503 [hep-ph].
73. S. Alekhin *et al.*, *Rept. Prog. Phys.* **79**, 124201 (2016), arXiv:1504.04855 [hep-ph].
74. J. Beacham *et al.*, *J. Phys. G* **47**, 010501 (2020), arXiv:1901.09966 [hep-ex].
75. Planck Collaboration, Y. Akrami *et al.*, *Astron. Astrophys.* **641**, A10 (2020), arXiv:1807.06211 [astro-ph.CO].
76. BICEP2, Keck Array Collaboration, P. A. R. Ade *et al.*, *Phys. Rev. Lett.* **121**, 221301 (2018), arXiv:1810.05216 [astro-ph.CO].
77. L. A. Popa (10 2021), arXiv:2110.09392 [hep-ph].
78. B. Holdom, *Phys. Lett. B* **166**, 196 (1986).
79. J. Blümlein and J. Brunner, *Phys. Lett. B* **701**, 155 (2011), arXiv:1104.2747 [hep-ex].
80. J. Blümlein and J. Brunner, *Phys. Lett. B* **731**, 320 (2014), arXiv:1311.3870 [hep-ph].
81. A. Berlin, N. Blinov, G. Krnjaic, P. Schuster and N. Toro, *Phys. Rev. D* **99**, 075001 (2019), arXiv:1807.01730 [hep-ph].
82. A. Caldwell *et al.* (12 2018), arXiv:1812.11164 [physics.acc-ph].
83. HPS Collaboration, P. H. Adrian *et al.*, *Phys. Rev. D* **98**, 091101 (2018), arXiv:1807.11530 [hep-ex].
84. B. Batell, J. L. Feng, A. Ismail, F. Kling, R. M. Abraham and S. Trojanowski, *Phys. Rev. D* **104**, 035036 (2021), arXiv:2107.00666 [hep-ph].
85. A. Berlin and F. Kling, *Phys. Rev. D* **99**, 015021 (2019), arXiv:1810.01879 [hep-ph].
86. A. Y. Ignatiev, V. A. Kuzmin and M. E. Shaposhnikov, *Phys. Lett. B* **84**, 315 (1979).
87. M. Staelens, The Search for New Physics at the LHC with the MoEDAL-MAPP Detector and an Investigation of Emergent Magnetic Monopole Like Excitations in an Exotic Condensed Matter System, PhD thesis, Alberta U., (2021).
88. T. Emken, R. Essig, C. Kouvaris and M. Sholapurkar, *JCAP* **09**, 070 (2019), arXiv:1905.06348 [hep-ph].
89. R. Plestid, V. Takhistov, Y.-D. Tsai, T. Bringmann, A. Kusenko and M. Pospelov, *Phys. Rev. D* **102**, 115032 (2020), arXiv:2002.11732 [hep-ph].
90. M. Sher and J. R. Stevens, *Phys. Lett. B* **777**, 246 (2018), arXiv:1710.06894 [hep-ph].
91. M. Frank, M. de Montigny, P.-P. A. Ouimet, J. Pinfold, A. Shaa and M. Staelens, *Phys. Lett.* **B802**, 135204 (2020), arXiv:1909.05216 [hep-ph].
92. D. Curtin, K. R. Dienes and B. Thomas, *Phys. Rev. D* **98**, 115005 (2018), arXiv:1809.11021 [hep-ph].

93. R. N. Mohapatra and G. Senjanovic, *Phys. Rev. Lett.* **44**, 912 (1980).
94. T. Yanagida, *Prog. Theor. Phys.* **64**, 1103 (1980).
95. J. Schechter and J. W. F. Valle, *Phys. Rev. D* **22**, 2227 (1980).
96. F. Kling and S. Trojanowski, *Phys. Rev. D* **97**, 095016 (2018), arXiv:1801.08947 [hep-ph].
97. M. Hirsch and Z. S. Wang, *Phys. Rev. D* **101**, 055034 (2020), arXiv:2001.04750 [hep-ph].
98. D. Dercks, H. K. Dreiner, M. Hirsch and Z. S. Wang, *Phys. Rev.* **D99**, 055020 (2019), arXiv:1811.01995 [hep-ph].
99. C. Arbeláez, J. C. Helo and M. Hirsch, *Phys. Rev. D* **100**, 055001 (2019), arXiv:1906.03030 [hep-ph].
100. G. Cottin, J. C. Helo, M. Hirsch, A. Titov and Z. S. Wang, *JHEP* **09**, 039 (2021), arXiv:2105.13851 [hep-ph].
101. F. Deppisch, S. Kulkarni and W. Liu, *Phys. Rev. D* **100**, 035005 (2019), arXiv:1905.11889 [hep-ph].
102. M. Drewes, *Int. J. Mod. Phys. E* **22**, 1330019 (2013), arXiv:1303.6912 [hep-ph].
103. A. Kusenko, *Phys. Rept.* **481**, 1 (2009), arXiv:0906.2968 [hep-ph].
104. M. Drewes *et al.*, *JCAP* **01**, 025 (2017), arXiv:1602.04816 [hep-ph].
105. F. F. Deppisch, P. S. Bhupal Dev and A. Pilaftsis, *New J. Phys.* **17**, 075019 (2015), arXiv:1502.06541 [hep-ph].
106. G. Cottin, J. C. Helo and M. Hirsch, *Phys. Rev. D* **98**, 035012 (2018), arXiv:1806.05191 [hep-ph].
107. J. De Vries, H. K. Dreiner, J. Y. Günther, Z. S. Wang and G. Zhou, *JHEP* **03**, 148 (2021), arXiv:2010.07305 [hep-ph].
108. J. L. Feng, I. Galon, F. Kling and S. Trojanowski, *Phys. Rev. D* **98**, 055021 (2018), arXiv:1806.02348 [hep-ph].
109. K. Sakurai, D. Felea, J. Mamuzic, N. E. Mavromatos, V. A. Mitsou, J. L. Pinfold, R. Ruiz de Austri, A. Santra and O. Vives, *J. Phys. Conf. Ser.* **1586**, 012018 (2020), arXiv:1903.11022 [hep-ph].
110. D. Felea, J. Mamuzic, R. Masełek, N. E. Mavromatos, V. A. Mitsou, J. L. Pinfold, R. Ruiz de Austri, K. Sakurai, A. Santra and O. Vives, *Eur. Phys. J. C* **80**, 431 (2020), arXiv:2001.05980 [hep-ph].
111. B. S. Acharya, A. De Roeck, J. Ellis, D. K. Ghosh, R. Masełek, G. Panizzo, J. L. Pinfold, K. Sakurai, A. Shaa and A. Wall, *Eur. Phys. J. C* **80**, 572 (2020), arXiv:2004.11305 [hep-ph].
112. D. Dercks, J. De Vries, H. K. Dreiner and Z. S. Wang, *Phys. Rev.* **D99**, 055039 (2019), arXiv:1810.03617 [hep-ph], [arXiv:1810.03617].
113. H. K. Dreiner, J. Y. Günther and Z. S. Wang, *Phys. Rev. D* **103**, 075013 (2021), arXiv:2008.07539 [hep-ph].
114. D. Curtin and M. E. Peskin, *Phys. Rev. D* **97**, 015006 (2018), arXiv:1705.06327 [hep-ph].
115. J. Barron and D. Curtin, *JHEP* **12**, 061 (2020), arXiv:2007.05538 [hep-ph].

Constraining the interactions in the dark sector with cosmological data

Adrià Gómez-Valent[1,*], Valeria Pettorino[2] and Luca Amendola[1]

[1] *Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*
E-mail: gomez-valent@thphys.uni-heidelberg.de

[2] *AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité F-91191 Gif-sur-Yvette, France*

We provide constraints on coupled dark energy (CDE) cosmology with Peebles-Ratra (PR) potential, $V(\phi) = V_0\phi^{-\alpha}$, and constant coupling strength β. This modified gravity scenario introduces a fifth force between dark matter particles, mediated by a scalar field that plays the role of dark energy. The mass of the dark matter particles does not remain constant, but changes with time as a function of the scalar field. Here we assess the ability of the model to describe updated cosmological data sets that include the *Planck* 2018 cosmic microwave background (CMB) temperature, polarization and lensing, baryon acoustic oscillations, the Pantheon compilation of supernovae of Type Ia, data on $H(z)$ from cosmic chronometers, and redshift-space distortions. We also study the impact of the local measurement of H_0 from SH0ES and the strong-lensing time delay data from the H0LICOW collaboration on β. We find a peak corresponding to a coupling $\beta > 0$ and to a potential parameter $\alpha > 0$, more or less evident depending on the data set combination. We show separately the impact of each data set and remark that it is especially CMB lensing the one data set that shifts the peak the most towards ΛCDM. When a model selection criterion based on the full Bayesian evidence is applied, however, ΛCDM is still preferred in all cases, due to the additional parameters introduced in the CDE model. The model is not able to loosen significantly the H_0 tension. This contribution to the proceedings of the DM1 parallel session of the 16th Marcel Grossmann virtual Conference: "Interacting dark matter" is based on the paper 2004.00610.[1]

Keywords: Cosmology: observations; Cosmology: theory; cosmological parameters; dark energy; dark matter.

1. Introduction

Important observational hints in favor of the positive acceleration of the Universe appeared already more than twenty years ago, thanks to the detection of standardizable high-redshift supernovae of Type Ia (SNIa) and the measurement of their light-curves and redshifts.[2,3] Since then, many other probes have contributed to increase the evidence in favor of the late-time accelerated phase. They range e.g. from the detection of the baryon acoustic peak in the two-point correlation function of matter density fluctuations[4,5] to the very accurate measurement of the cosmic microwave background (CMB) temperature anisotropies by WMAP[6] and *Planck*.[7–9]

*Speaker. E-mail: gomez-valent@thphys.uni-heidelberg.de

At the phenomenological level, the easiest explanation for such acceleration is given by the presence of a very tiny cosmological constant in Einstein's field equations, with an associated energy density which is orders of magnitude lower than the quantum field theoretical estimates made for the vacuum energy density. Protecting such low value from radiative corrections is extremely difficult and constitutes the core of the so-called "old" cosmological constant problem, cf. e.g.[10–12] In addition, explaining why the current value of this energy density is of the same order of magnitude as the matter energy density, the so-called "coincidence problem", is considered by part of the cosmological community as another problem that needs to be addressed. The cosmological constant is a pivotal ingredient of the standard cosmological model, also known as ΛCDM or concordance model (cf. e.g. the reviews[13,14]), which can explain most of the cosmological observations with high proficiency. Nevertheless, the aforementioned theoretical conundrums, together with few persistent tensions in some relevant parameters of the model as the Hubble parameter H_0[9,15] and the root-mean-square (*rms*) of mass fluctuations at scales of $8h^{-1}$ Mpc,[16] σ_8 (or $S_8 = \sigma_8(\Omega_m^{(0)}/0.3)^{0.5}$ [a17]), with h being the reduced Hubble parameter, motivate theoretical cosmologists to look for alternative scenarios in which these problems can be solved or, at least, alleviated, see[18,19] and references therein. Wherever the solution comes from, i.e. a departure from General Relativity or some sort of new field describing dark energy (DE), it must mimic very well the behavior of a cosmological constant at low redshifts, meaning that the corresponding effective equation of state (EoS) parameter must be very close to -1, and that the new component must not be able to cluster efficiently at low scales.

In this paper we consider a scenario in which dark matter (DM) particles interact via a force mediated by a scalar field, which in turn drives cosmic acceleration. This scenario is referred to as *coupled dark energy* (CDE). It was originally proposed as a means of alleviating the coincidence problem,[20,21] considering not only a potential energy density for quintessence to generate its dynamics, but also allowing an interaction with other sectors of the theory. These interactions extended the original quintessence models.[22–25] They cannot be ruled out *a priori* and, hence, they must be duly constrained by experiments and observations.

Some works already set constraints on this model, but using older cosmological data sets, for instance CMB data from the WMAP satellite and the South Pole Telescope,[26] or considering past (2013, 2015) releases of *Planck* CMB data in combination with other data sets, as e.g. from baryon acoustic oscillations (BAO) and SNIa,[27,28] Intriguingly, these works detected a likelihood peak at a non-vanishing value of the coupling constant. One of our main goals is then to critically revisit and update these results in the light of the recent strengthening of the H_0 tension and of the rich amount of currently available data at our disposal, in particular the *Planck* 2018 CMB temperature, polarization and lensing data, but also other new cosmological data, for instance Refs.[29,30]. For constraints on other models with DM-DE

[a]The superscripts (0) will denote from now on quantities evaluated at present, i.e. at $a = 1$.

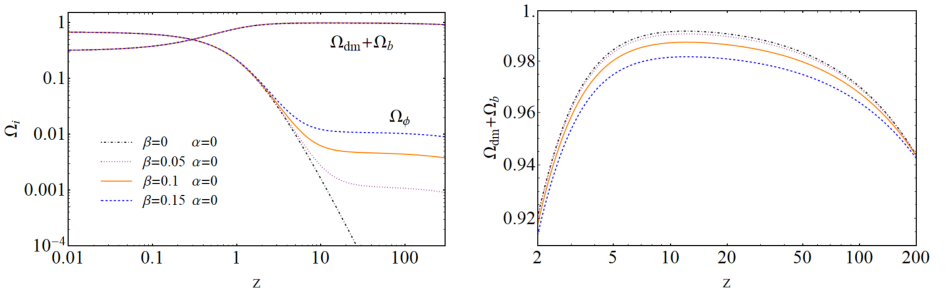


Fig. 1. *Left plot:* Normalized densities $\Omega_{dm}(z)+\Omega_b(z)$ and $\Omega_\phi(z)$ for four alternative values of β and considering a constant potential. The other parameters (including the current energy densities) have been set to the best-fit ΛCDM values from the TTTEEE+lowE *Planck* 2018 analysis.[9] *Right plot:* Here we zoom in the range $z=[2,200]$ of the $\Omega_{dm}+\Omega_b$ curves in order to better visualize their evolution during the matter-dominated epoch, when the system is near the ϕMDE fixed point. See the text for details.

interactions see e.g.,[31–44] and when the interaction is motivated in the context of the running vacuum models.[40,41,45–48]

2. Coupled dark energy

We consider a CDE scenario, as studied in,[21,49,50] to which we refer for a detailed description. We here briefly recall the main equations. This CDE model is formulated in the so-called Einstein or observational frame.[51] Apart from the Standard Model of Particle Physics and a potential extension accounting for the origin of the neutrino masses, we consider a dark sector described by the following Lagrangian density:

$$\mathcal{L}_{\rm dark}=-\partial_\mu\phi\partial^\mu\phi-V(\phi)-m(\phi)\bar{\psi}\psi+\mathcal{L}_{\rm kin}[\psi]\,, \tag{1}$$

where ϕ is the scalar field that plays the role of DE, with potential $V(\phi)$, and ψ is the DM field, considered here to be of fermionic nature, just for illustrative purposes. The DM particles interact with the DE due to the ϕ-dependent mass term appearing in (1). Such interaction introduces a fifth force that alters the trajectory in spacetime of the DM with respect to the one found in the uncoupled case. As we do not couple ϕ to the standard model sector we avoid the stringent local (Solar System) constraints on the violation of the weak equivalence principle,[52] and also on screened fifth forces that couple ϕ to non-dark matter, e.g. from Casimir experiments,[53] precision measurements of the electron magnetic moment,[54] or measurements of the Eötvös parameter.[55] They have no impact on the CDE model under study.

The variation of the total action with respect to the metric leads as usual to Einstein's equations, and the covariant energy of the joint system DM-DE is conserved. Hence, $\nabla^\mu T^\phi_{\mu\nu}=+Q_\nu$ and $\nabla^\mu T^{dm}_{\mu\nu}=-Q_\nu$, with Q_ν defined as

$$Q_\nu=\beta\kappa T^{dm}\nabla_\nu\phi\,, \tag{2}$$

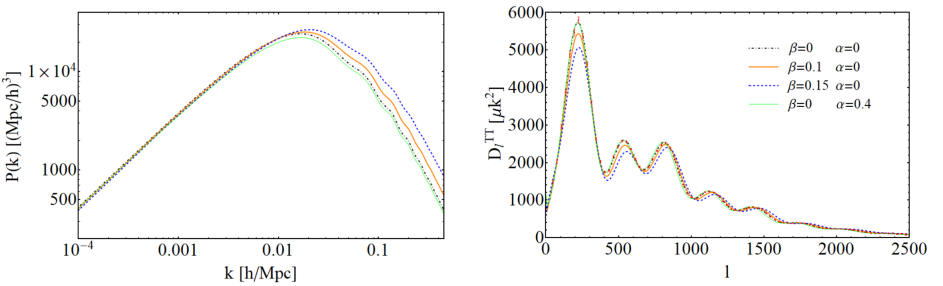


Fig. 2. Theoretical curves of the current matter power spectrum (left plot) and CMB temperature anisotropies (right plot) for the ΛCDM, two CDE models with $\beta = 0.1$, 0.15 and flat potential, and also for the uncoupled Peebles-Ratra model with $\alpha = 0.4$. We set the other parameters as in Fig. 1. In the right plot we also include the observational data from[9] (in red). These figures show: (i) the enhancement of the growth of matter perturbations caused by $\beta > 0$, and the opposite effect produced by $\alpha > 0$; and (ii) the shift to larger multipoles and the amplitude suppression of the acoustic peaks induced by increasing values of β. See the text for further details.

where $\kappa = \sqrt{8\pi G}$, T^{dm} is the trace of the DM energy-momentum tensor, and β controls the strength of the interaction and is in general a function of ϕ. If set to zero, we recover the equations of uncoupled quintessence. In this work we consider β to be a positive constant.

We assume that the Universe is spatially flat, as supported by CMB information from *Planck* 2018 when combined with BAO[9] and/or SNIa,[56] with the curvature parameter $\Omega_K^{(0)}$ constrained to be lower than $\sim 2\%$ at 68% c.l. in ΛCDM. Thus, we can make use of the Friedmann-Lemaître-Robertson-Walker metric, which at the background level reads $ds^2 = a^2(\tau)\left[-d\tau^2 + \delta_{ij}dx^i dx^j\right]$, with a being the scale factor, τ the conformal time, and x^i for $i = 1, 2, 3$ the spatial comoving coordinates. In addition, we treat DM as a pressureless perfect fluid, so the conservation equations for DE and DM can be written, respectively,

$$\beta\kappa a^2 \rho_{dm} = \phi'' + 2\mathcal{H}\phi' + a^2 \frac{\partial V}{\partial \phi}\,, \tag{3}$$

$$\rho'_{dm} + 3\mathcal{H}\rho_{dm} = -\beta\kappa\rho_{dm}\phi'\,, \tag{4}$$

with ρ_{dm} the DM energy density, $\mathcal{H} = a'/a$, and the primes denoting derivatives *w.r.t.* the conformal time. All the functions entering these equations are background quantities. If we assume the conservation of the number density of DM particles then their mass evolves as $m(\phi) = m^{(0)} e^{\beta\kappa(\phi^{(0)} - \phi)}$.

A feature of the model is that for $\beta^2 < 3/2$ it has an unstable (saddle) fixed point at $(\Omega_{dm}, \Omega_\phi) = (1 - 2\beta^2/3, 2\beta^2/3)$, where $\Omega_i = \rho_i/\rho_c$, with ρ_c the critical energy density. This fixed point (dubbed ϕMDE in[21]) cannot be reached exactly, since there is also a non-null fraction of baryons, but the system can be quite close to it, since the DM energy density is much larger than the baryonic one (cf. Fig. 1). During this phase the effective EoS parameter, i.e. the ratio of the total pressure

and the critical energy density in the Universe, is given by $w_{\rm eff} = \Omega_\phi$, and hence the deceleration parameter reads $q = \frac{1}{2}(1+3\,w_{\rm eff}) = \frac{1}{2}+\beta^2$. Thus, the coupling between DM and DE makes the Universe more decelerated with respect to the uncoupled quintessence case during the matter-dominated epoch (MDE). This fact together with the fifth force that enters now as a new source term in the Poisson equation help matter inhomogeneities to grow faster for larger values of β. We also remark that for fixed values of the present energy densities, matter becomes dominant over radiation earlier in time when $\beta > 0$, with respect to the uncoupled case. In the CDE scenario, the equation for the DM density contrast $\delta_{dm} = \delta\rho_{dm}/\rho_{dm}$ at deep subhorizon scales ($k \gg \mathcal{H}$) and when non-linear processes are unimportant, reads,

$$\delta''_{dm} + (\mathcal{H} - \ \beta\kappa\phi')\delta'_{dm} - 4\pi G a^2[\rho_b\delta_b + \rho_{dm}\delta_{dm}(1+2\beta^2)] = 0\,. \tag{5}$$

If we neglect the contribution of baryons, $\delta_m(a) \sim a^{1+2\beta^2}$. Hence, larger values of β enhance the matter power spectrum (see the left plot of Fig. 2) and leave an imprint on the CMB temperature anisotropies. First, the integrated Sachs-Wolfe effect[57] is enhanced during the MDE earlier than in the uncoupled scenario, in which such effect is only relevant after matter-domination; second, the coupling affects lensing of CMB by large scale structure; the interaction also shifts the position of the acoustic peaks to larger multipoles due to the decrease of the sound horizon at the baryon-drag epoch, which is caused by the increase of the mass of the DM particles. Finally, the amplitude is suppressed, because of the decrease of ρ_b/ρ_{dm} at recombination. These two effects explain why the coupling strength is degenerate with the Hubble parameter today,[27] whose value is related to the position and overall amplitude of the first peak. These and other aspects of the structure formation were already discussed in.[50,58–60] See therein for further details, and also the plots in Fig. 2.

The quintessence potential only rules the dynamics of ϕ in the late-time universe, after the MDE, when the interaction term appearing in the *l.h.s.* of (3) becomes subdominant. It helps to slow down structure formation processes *w.r.t.* the flat-potential scenario (for a fixed value of the current DE density). Hence, it can compensate in lesser or greater extent (depending on its steepness) the enhancement of power generated by the fifth force during the MDE (cf. the left plot of Fig. 2 and its caption).

We employ the Peebles-Ratra (PR) potential,[24,25]

$$V(\phi) = V_0\phi^{-\alpha}\,, \tag{6}$$

with V_0 and $\alpha > 0$ being constants, and the former having dimensions of mass$^{4+\alpha}$ in natural units, since ϕ has dimensions of mass. We want to update the constraints on the parameters of the CDE model with PR potential that were obtained in some past works using older CMB data, from WMAP and/or past releases of *Planck* (cf.[26–28,61]), so it is natural to stick to (6) here. Also because it has proved to be capable of improving the description of some cosmological data sets with respect to the ΛCDM model in the non-interactive case.[62–64]

The CDE model we are considering (i.e. CDE with PR potential) has three nested models, namely the ΛCDM, the PR model, and the CDE model with flat potential. They are obtained from the full CDE model with (6) in the limits $(\alpha, \beta) \to (0, 0)$, $\beta \to 0$ and $\alpha \to 0$, respectively. For constraints on these scenarios cf. appendix B of Ref.[1].

For recent studies on CDE with an exponential potential, see[31,33,34,43] and appendix C of Ref.[1]. The results are very similar to the ones obtained with the PR potential.

3. Methodology and data

We have implemented the CDE model described in Sec. 2 in our own modified version of the Einstein-Boltzmann system solver `CLASS`.[65] The Bayesian exploration of the parameter space of the model in the light of the various data sets has been carried out with the Monte Carlo sampler `Montepython`.[66] We have also used the `Python` package `GetDist`[67] to process the chains and obtain the mean values and uncertainties of the parameters reported in Table 1, as well as the contours of Figs. 3-4. Finally, we have computed the full Bayesian evidences for all the models and under the various data sets, by processing the corresponding Markov chains with the code `MCEvidence`.[68] This has allowed us to carry out a rigorous model comparison analysis, which we present in Sec. 4.

Our data set is very similar to the one used by the *Planck* collaboration in their 2018 analysis of the ΛCDM and minimal extensions of it.[9] There are some differences, though, e.g. we analyze here the effect of cosmic chronometers and the H0LICOW data, something that was not done there. We refer the reader to Sec. 3 and reference[9] for details.

This is the list of individual data sets that we employ in this work to constrain the CDE model presented in Sec. 2:

- CMB: The main results of this paper are derived making use of the full TTTEEE+lowE CMB likelihood from *Planck* 2018.[9] We also study what is the impact of also including the CMB lensing likelihood.[69]
- BAO: We use the data points reported in[30,70–75]
- SNIa: We consider 6 effective points on the Hubble rate, i.e. $E(z) \equiv H(z)/H_0$, and the associated covariance matrix. They compress the information of 1048 SNIa contained in the Pantheon compilation[76] and the 15 SNIa at $z > 1$ from the Hubble Space Telescope Multi-Cycle Treasury programs.[77]
- Cosmic chronometers (CCH): We have employed the 31 data points on $H(z_i)$ at various redshifts provided in.[78–85] More concretely, we make use of the *processed* sample provided in Table 2 of,[86] which is more conservative, since it introduces corrections accounting for the systematic errors mentioned above.

Table 1. Constraints obtained using the data set combinations described in Sec. 3 on the following parameters of the CDE model: the reduced DM and baryon energy densities, $\Omega_{dm}^{(0)}h^2$ and $\Omega_b^{(0)}h^2$; the reionization optical depth, τ; the Hubble parameter, H_0 (in units of km/s/Mpc); the power of the primordial power spectrum, n_s; the current amplitude of mass fluctuations at $8h^{-1}$ Mpc, σ_8; the coupling strength β; and the power of the PR potential (6). We provide the mean values and 68% confidence intervals for each of them. We also show the differences *w.r.t.* the ΛCDM of the minimum values of the χ^2-function, and the natural logarithm of the Bayes ratio $B_{\rm CDE,\Lambda}$, as defined in (9)-(10). The (small) negative values of $\chi^2_{min.\rm CDE} - \chi^2_{min,\Lambda}$ tell us that CDE is able to fit slightly better the data than the ΛCDM; if we use as an alternative estimator the Bayes factor, we find negative values of $\ln(B_{\rm CDE,\Lambda})$, indicating a preference for the ΛCDM model. See Sec. 4 for a thorough discussion.

Parameter/Info. criteria	P18	P18+BSC
$\Omega_{dm}^{(0)}h^2$	$0.1207^{+0.0014}_{-0.0013}$	0.1192 ± 0.0008
$\Omega_b^{(0)}h^2$	0.02237 ± 0.00015	$0.02242^{+0.00010}_{-0.00015}$
τ	0.0538 ± 0.0070	$0.0532^{+0.0075}_{-0.0087}$
H_0	$67.74^{+0.57}_{-0.66}$	68.41 ± 0.38
n_s	$0.9654^{+0.0035}_{-0.0042}$	0.9690 ± 0.0038
σ_8	0.8164 ± 0.0076	0.8104 ± 0.0076
α	< 0.50	0.52 ± 0.17
β	$0.0158^{+0.0067}_{-0.0120}$	$0.0206^{+0.0070}_{-0.0095}$
$\chi^2_{min,\rm CDE} - \chi^2_{min,\Lambda}$	-0.02	-0.28
$\ln B_{\rm CDE,\Lambda}$	-8.05	-9.95
Parameter	P18+SH0ES+H0LICOW	P18+BSC+RSD
$\Omega_{dm}^{(0)}h^2$	$0.1172^{+0.0012}_{-0.0014}$	0.1187 ± 0.0008
$\Omega_b^{(0)}h^2$	$0.02262^{+0.00016}_{-0.00014}$	$0.02253^{+0.00010}_{-0.00012}$
τ	0.0594 ± 0.0074	0.0501 ± 0.0052
H_0	$69.43^{+0.72}_{-0.53}$	$68.64^{+0.30}_{-0.38}$
n_s	0.9731 ± 0.0042	$0.9701^{+0.0029}_{-0.0033}$
σ_8	$0.8121^{+0.0065}_{-0.0080}$	0.8048 ± 0.0052
α	1.32 ± 0.18	$0.67^{+0.11}_{-0.16}$
β	$0.0294^{+0.0120}_{-0.0076}$	$0.0151^{+0.0073}_{-0.0083}$
$\chi^2_{min,\rm CDE} - \chi^2_{min,\Lambda}$	-0.58	-1.56
$\ln B_{\rm CDE,\Lambda}$	-7.57	-8.33
Parameter	P18lens+BSC+RSD	P18+BSC+SH0ES+H0LICOW
$\Omega_{dm}^{(0)}h^2$	0.1191 ± 0.0007	0.1185 ± 0.0008
$\Omega_b^{(0)}h^2$	$0.02253^{+0.00013}_{-0.00011}$	$0.02253^{+0.00011}_{-0.00013}$
τ	$0.0525^{+0.0052}_{-0.0064}$	$0.0579^{+0.0069}_{-0.0078}$
H_0	68.45 ± 0.34	$68.79^{+0.35}_{-0.40}$
n_s	0.9685 ± 0.0034	0.9705 ± 0.0034
σ_8	$0.8073^{+0.0048}_{-0.0056}$	0.8120 ± 0.0074
α	$0.25^{+0.09}_{-0.20}$	$0.73^{+0.11}_{-0.27}$
β	$0.0095^{+0.0030}_{-0.0087}$	$0.0206^{+0.0076}_{-0.0100}$
$\chi^2_{min,\rm CDE} - \chi^2_{min,\Lambda}$	-0.90	-1.34
$\ln B_{\rm CDE,\Lambda}$	-7.83	-7.95
Parameter	P18lens+SH0ES+H0LICOW	-
$\Omega_{dm}^{(0)}h^2$	$0.1182^{+0.0011}_{-0.0010}$	-
$\Omega_b^{(0)}h^2$	$0.02259^{+0.00014}_{-0.00016}$	-
τ	$0.0637^{+0.0065}_{-0.0096}$	-
H_0	68.99 ± 0.51	-
n_s	0.9713 ± 0.0037	-
σ_8	0.8160 ± 0.0068	-
α	$0.33^{+0.19}_{-0.23}$	-
β	$0.0197^{+0.0094}_{-0.0084}$	-
$\chi^2_{min,\rm CDE} - \chi^2_{min,\Lambda}$	-1.46	-
$\ln B_{\rm CDE,\Lambda}$	-8.75	-

- Redshift-spcae distortions (RSD): We also use large-scale structure measurements from the anisotropic clustering of galaxies in redshift space. Galaxy redshift surveys provide constraints on the product of the growth rate of structure, $f(z) = \frac{d \ln \delta_m(a)}{d \ln a}$, and the *rms* of mass fluctuations at scales of $8h^{-1}$ Mpc, $\sigma_8(z)$. The data points employed in this work are found in Refs.[30,71,87–95]. The internal correlations between the BAO and RSD data from[71] and[30] have been duly taken into account through the corresponding covariance matrices provided in these two references.
- SH0ES: In some of our data set combinations we include the prior on the Hubble parameter, $H_{0,\mathrm{SH0ES}} = (74.03 \pm 1.42)$ km/s/Mpc obtained by the SH0ES Team with the cosmic distance ladder method.[15] This value of the Hubble parameter is in 4.4σ tension with the TTTEEE+lowE+lensing best-fit ΛCDM model of *Planck* 2018,[9] $H_0 = 67.36 \pm 0.54$ km/s/Mpc.
- H0LICOW: In combination with the prior on H_0 from SH0ES we also use the angular diameter distances reported by the H0LICOW collaboration. They analyze six gravitationally lensed quasars of variable luminosity. After measuring the time delay between the deflected light rays and modeling the lenses they are able to measure the so-called time-delay distances $D_{\Delta t}$ (cf.[29] and references therein). We use their reported six time-delay distances (one for each lensed system), and one distance to the deflector B1608+656, which according to the authors of[29] is uncorrelated with the corresponding $D_{\Delta t}$. The relevant information for building the likelihood can be found in Tables 1 and 2 of,[29] and their captions. Assuming the concordance model, these distances lead to a value of $H_0 = (73.3^{+1.7}_{-1.8})$ km/s/Mpc, which is in 3.2σ tension with the one obtained from the TTTEEE+lowE+lensing analysis by *Planck*.[9]

For more detailed information about these data sets see Ref.[1] and the original observational works.

We proceed now to describe the data set combinations under which we have obtained the main results of this work. They are discussed in detail in Sec. 4. We put constraints using the following combinations: (i) TTTEEE+lowE CMB data from *Planck* 2018,[9] in order to see the constraining power of the CMB when used alone, and also to check whether these data lead to a higher value of H_0 than in the ΛCDM. For simplicity, we will refer to this data set as P18 throughout the paper; (ii) P18+BSC, with BSC denoting the background data set BAO+SNIa+CCH; (iii) We add on top of the latter the linear structure formation information contained in the RSD data, P18+BSC+RSD; (iv) We study the impact of the CMB lensing by also adding the corresponding likelihood, P18lens+BSC+RSD; (v) Finally, we analyze the impact of the prior on H_0 from SH0ES[15] and the H0LICOW angular diameter distances[29] by using the data sets P18+SH0ES+H0LICOW, P18lens+SH0ES+H0LICOW and P18+BSC+SH0ES+H0LICOW. The distance ladder and strong-lensing time delay measurements of the Hubble constant are

completely independent (see e.g. the reviews[96, 97]). When combined, they lead to

$$H_{0,comb} = (73.74 \pm 1.10)\,\mathrm{km/s/Mpc}\,, \tag{7}$$

in 5.2σ tension with the best-fit ΛCDM value reported by *Planck* 2018.[9] Hence, it is interesting to check what is the response of the CDE model under these concrete data sets, and to compare the results with those obtained using only the CMB likelihood.

4. Results

Our main results are presented in Table 1 and Figs. 3-4. When we only employ the CMB temperature and polarization data from *Planck* 2018[9] (i.e. the P18 data set) to constrain the CDE model, the fitting values obtained for α and β are compatible at 1σ c.l. with 0, i.e. with a cosmological constant and no interaction in the dark sector (cf. the first row, first column in Table 1). The value of H_0 remains low, roughly 4.1σ below the cosmic distance ladder measurement of.[15] Similarly, when we combine *Planck* with BSC background data or with BSC+RSD, we get a value of H_0 which is 3.8σ and 3.7σ away from the SH0ES value, respectively.

As we have explained in Sec. 2, there is a degeneracy between the strength of the fifth force, i.e. the parameter β, and the Hubble parameter. CDE is in principle able to lower the value of the sound horizon at the decoupling time, r_s, and the amplitude of the first peak of the $\mathcal{D}_l^{TT}$'s. The CMB data fix with high precision the angle $\theta_* = r_s/D_A^{(c)}(z_{dec})$, with $D_A^{(c)}(z_{dec})$ the comoving angular diameter distance to the CMB last scattering surface. This means that in order to keep this ratio constant, H_0 will tend to grow for increasing values of the coupling strength, so that $D_A^{(c)}(z_{dec})$ decreases and compensates in this way the lowering of r_s, while keeping the height of the first peak compatible with data. This positive correlation between H_0 and β can be appreciated in the left-most contour plot of Fig. 3. The latter shows 1 and 2σ posterior probabilities for a selection of cosmological parameters. As discussed, we confirm from the first plot a mild degeneracy between H_0 and β. The strength of the fifth force does not seem to be very degenerate with σ_8 nor with the potential parameter α.

The impact of adding background data on top of P18 can be grasped by looking at the one-dimensional posterior distributions of Fig. 3 (in blue), and also at the numbers of the first row/second column of Table 1. Using the P18+BSC combined data set we find that β and α are now ~ 2.5 and $\sim 3.1\sigma$ away from 0, respectively. The values of H_0 and σ_8, are however compatible at 1σ with the ones obtained using only the P18 data set. They are also fully compatible with those obtained with the ΛCDM under the same data set, which read: $H_0 = (68.29 \pm 0.37)$ km/s/Mpc, $\sigma_8 = 0.812^{+0.006}_{-0.008}$. The peaks in β and α may indicate a mild preference of low-redshift data, when combined with the CMB, for a non-null interaction in the dark sector and a running quintessence potential. As noted already in,[28] we remark that this preference does not seem to correspond to a large improvement in the minimum

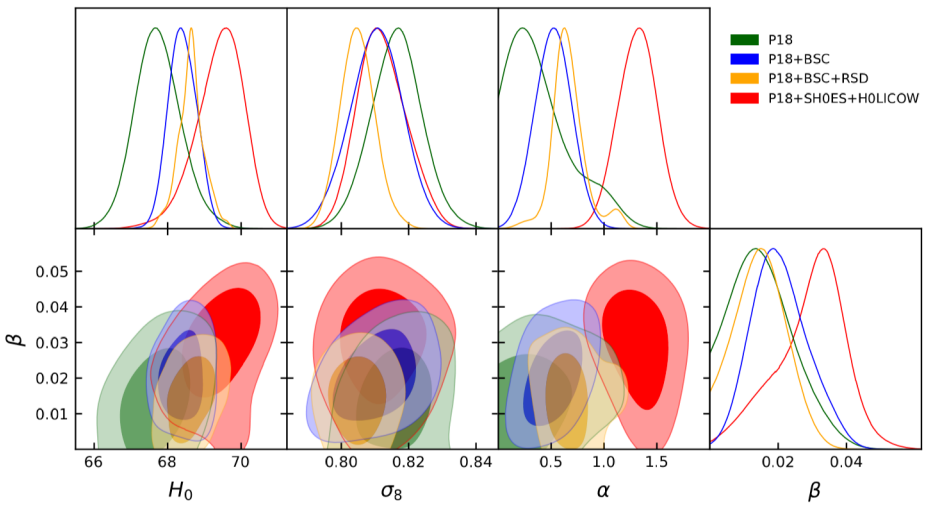


Fig. 3. 1 and 2σ confidence contours obtained using some of the combined data sets described in Sec. 3 in the (H_0, β), (σ_8, β), and (α, β) planes, together with the marginalized one-dimensional posterior distributions for these parameters. See the discussion of these results in Sec. 4.

value of χ^2 with respect to the ΛCDM: under the P18+BSC data set, $\chi^2_{min,\mathrm{CDE}} - \chi^2_{min,\Lambda}$ is negative, but very close to 0, which means that the CDE model only is able to improve the description of the data in a very marginal way.

The addition of the RSD data to the P18+BSC combined data set doesn't change much the result: there is a very small shift in the peak of the one-dimensional posterior distribution for α to larger values and the one for β to lower ones (see the yellow curves in Fig. 3). These two facts reduce a little bit the value of σ_8. The aforesaid peaks are now ~ 5 and $\sim 2\sigma$ away from 0, respectively, with a reduction of χ^2_{min} *w.r.t.* the ΛCDM of 1.56 units (cf. Table 1, second row/second column), i.e. pointing to a very small preference for CDE. The value of H_0 is almost unchanged.

We find important to highlight the specific impact of CMB lensing data with respect to the P18+BSC+RSD data set. If we include also the CMB lensing information, i.e. if we consider the P18lens+BSC+RSD combined data set, posterior probabilities squeeze, as expected, towards the ΛCDM values. This can be seen in Fig. 4, and also in the third row/first column of Table 1.

In order to further evaluate the level at which the degeneracy observed in the (H_0,β)-plane can alleviate the tension in the Hubble parameter between *Planck* and {SH0ES, H0LICOW} data, we perform a Monte Carlo analysis combining those data within the CDE model: results are shown in the second row/first column in Table 1 and correspond to red contours in Fig. 3. In this case, the best fit corresponds to a value of $\beta = 0.0294^{+0.0120}_{-0.0076}$, i.e. at 3σ from zero coupling, a value of $\alpha = 1.32 \pm 0.18$, with $\alpha > 0$ at $\sim 7\sigma$ c.l., and $H_0 = (69.43^{+0.72}_{-0.53})$ km/s/Mpc. The raise of H_0 is

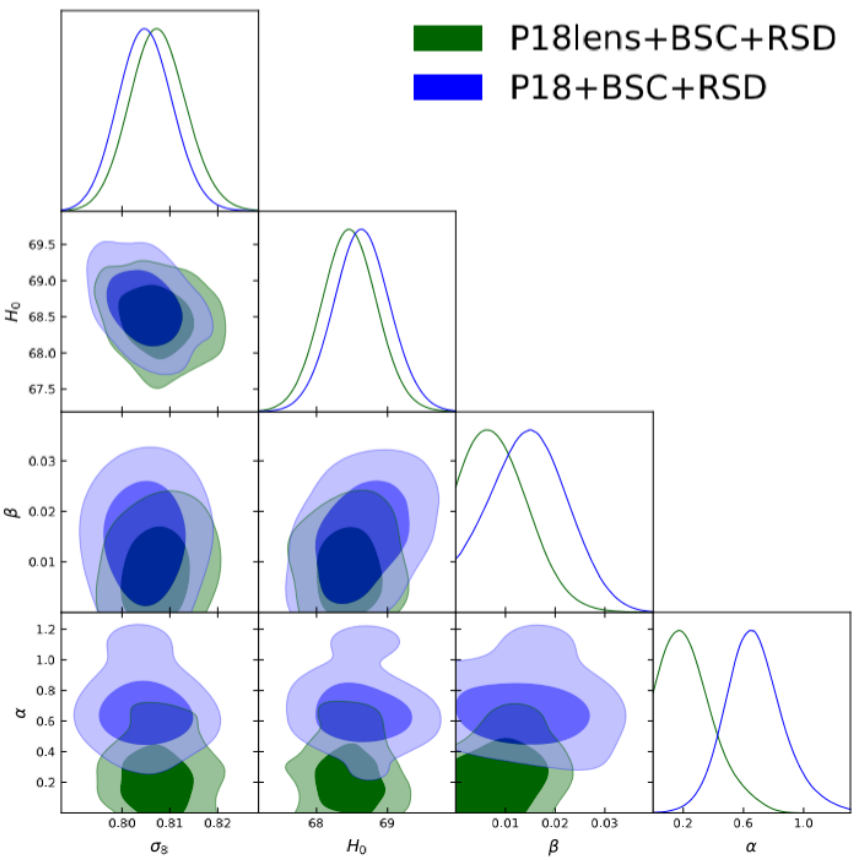


Fig. 4. 1 and 2σ confidence contours obtained with the P18+BSC+RSD and P18lens+BSC+RSD data sets in the most relevant two-dimensional planes of the CDE model parameter space. They allow us to see what is the impact of the CMB lensing on our results. We also show the corresponding marginalized one-dimensional posterior distributions for all the parameters. See the related comments in Sec. 4.

possible thanks to the increase of β, which in turn needs also larger values of α. The tension with the SH0ES+H0LICOW measurement (7) is slightly reduced from 4.8σ (when only P18 is used to constrain the model, cf. the first row/first column of Table 1) to 3.5σ (when also the SH0ES+H0LICOW data are considered). This shifts the H_0 value 1.9σ higher than the best fit using the P18 data set alone, within CDE. Combining also with background data, such as BSC, can partially break degeneracies and leads to $\alpha = 0.73^{+0.11}_{-0.27}$, with $\alpha > 0$ at 3.8σ and $H_0 = (68.79^{+0.35}_{-0.40})$ km/s/Mpc at 4.3σ from the SH0ES+H0LICOW value (7), reducing the chance of CDE to alleviate the tension, as shown in the third row/second column of the table. Finally, the impact of adding CMB lensing is shown in the last row, where now $\beta = 0.0197^{+0.0094}_{-0.0084}$ and $\alpha = 0.33^{+0.19}_{-0.23}$, with $\beta > 0$ and $\alpha > 0$ at 2.2σ and 1.6σ,

respectively, i.e. shifting back towards ΛCDM. In this case $H_0 = (68.99 \pm 0.51)$ km/s/Mpc, 3.9σ away from the SH0ES+H0LICOW value (7) and even more had we included also BSC.

Finally, we can further quantify the relative ability of the CDE model to describe the various data sets *w.r.t.* the ΛCDM cosmology using the Bayes ratio, in alternative to the more approximate χ^2 estimate we mentioned so far. Given a data set $\mathcal{D}$, the probability of a certain model M_i to be the best one among a given set of models $\{M\}$ reads,

$$P(M_i|\mathcal{D}) = \frac{P(M_i)\mathcal{E}(\mathcal{D}|M_i)}{P(\mathcal{D})}\,, \tag{8}$$

where $P(M_i)$ is the prior probability of the model M_i and $P(\mathcal{D})$ the probability of having the data set $\mathcal{D}$. Obviously, the normalization condition $\sum_{j\in\{M\}} P(M_j) = 1$ must be fulfilled. The quantity $\mathcal{E}(\mathcal{D}|M_i)$ is the so-called marginal likelihood or evidence. If the model M_i has n parameters $p_1^{M_i}, p_2^{M_i}, ..., p_n^{M_i}$, the evidence takes the following form,

$$\mathcal{E}(\mathcal{D}|M_i) = \int \mathcal{L}(\mathcal{D}|\vec{p}^{M_i}, M_i)\pi(\vec{p}^{M_i})d^n p^{M_i}\,, \tag{9}$$

with $\mathcal{L}(\mathcal{D}|\vec{p}^{M_i}, M_i)$ being the likelihood and $\pi(\vec{p}^{M_i})$ the prior of the parameters entering the model M_i. The evidence is larger for those models that have more overlapping volume between the likelihood and the prior distributions, but penalizes the use of additional parameters having a non-null impact on the likelihood. Hence, the evidence constitutes a good way of quantifying the performance of the model by implementing in practice the Occam razor principle. If we compare the CDE and ΛCDM models by assuming equal prior probability for both of them, i.e. $P(\text{CDE}) = P(\Lambda\text{CDM})$, then we find that the ratio of their associated probabilities is directly given by the ratio of their corresponding evidences, i.e.

$$\frac{P(\text{CDE}|\mathcal{D})}{P(\Lambda\text{CDM}|\mathcal{D})} = \frac{\mathcal{E}(\mathcal{D}|\text{CDE})}{\mathcal{E}(\mathcal{D}|\Lambda\text{CDM})} \equiv B_{\text{CDE},\Lambda}\,. \tag{10}$$

This is known as Bayes ratio and is the quantity we are interested in. For more details we refer the reader to.[18,98,99] Notice that the computation of (10) is not an easy task in general, since we usually work with models with a high number of (mostly nuisance) parameters, so the integrals under consideration becomes quite involved. We have computed the evidences numerically using the Markov chains obtained from the Monte Carlo analyses and with the aid of the numerical code `MCEvidence`,[68] which is publicly available (cf. Sec. 3). We report the values obtained for the natural logarithm of the Bayes ratio (10) in the last row of Table 1. For all the data sets under study we find values of $\ln(B_{\text{CDE},\Lambda}) < -5$, which point to a preference of the ΛCDM over the CDE model according to Jeffreys' scale.[18,98,99] Although the CDE model we are studying here is able to reduce slightly the value of χ^2_{min} *w.r.t.* the ΛCDM, it has two additional parameters, namely α and β. Moreover, the initial value of the scalar field, ϕ_{ini}, is also left free in the Monte Carlo analysis,

cf. Appendix A of Ref.[1] for details[b]. It turns out that the decrease in χ^2_{min} is insufficient to compensate the penalization introduced by the use of these extra parameters. If instead of using the evidences (9) and the Bayes ratio (10) to perform the model comparison we make use of e.g. the Akaike,[100] Bayesian[101] or Deviance[102] information criteria, we reach similar conclusions[c]. We want to note, though, that all these information criteria are approximations of the exact Bayesian approach. Although they allow to skip the demanding computation of the evidence (9), they are only reliable when the posterior distribution is close to a multivariate Gaussian (which is not the case under study), and the Akaike and Bayesian criteria do not take into account the impact of priors nor the existing correlations between the parameters.

Similar results and conclusions are reached using an exponential potential for the scalar field, instead of (6). See appendix C of Ref.[1].

Finally, it is worth to mention that our results are compatible with the ones obtained in the context of the Ricci running vacuum models (RVMs) of type I recently studied in Ref.[48], in which there is also an interaction in the dark sector between dark matter and a dynamical vacuum component with $p_{\rm vac} = -\rho_{\rm vac}$ and $\rho_{\rm vac}(\mathcal{R}) = \frac{3}{8\pi G_N}\left(c_0 + \frac{\nu}{12}\mathcal{R}\right)$, where $\mathcal{R} = 12H^2 + 6\dot{H}$ is the Ricci scalar. These models can be motivated using renormalization group arguments in QFT in curved space-time, see Refs.[12,103] and references therein. In Ref.[48] the authors show that these RVMs are also unable to alleviate the cosmological tensions significantly when the interaction is active around the CMB decoupling time. Allowing for a late-time activation of the interaction around $z \sim 1$, with an energy transfer from the vacuum to the matter sectors, it is possible to mitigate the σ_8 tension due to the increase of the relative amount of vacuum energy with respect to dark matter in the past. Nevertheless, the H_0 tension persists. Other Ricci RVMs (those of Type II) are able to soften both tensions at a time, see[48] for further details.

5. Conclusions

Cosmological observations help to test the dark sector, and in particular interactions between dark matter particles, mediated by a dark energy scalar field, as in CDE cosmologies. Up to a conformal transformation, this is another way of testing gravity at large scales. In this paper we carried out this task in one of the simplest and most studied models, namely, a dark energy-dark matter conformal coupling with a

[b]In the computation of the evidence (9) for the CDE model we have employed the following flat priors for the extra parameters: $0 < \beta < 0.1$, $0 < \alpha < 2$, and $0 < \kappa\phi_{ini} < 50$. Slightly broader or tighter priors can be considered, but $\ln(B_{\rm CDE,\Lambda})$ only changes logarithmically, so our conclusions are not very sensitive to them.

[c]For instance, Akaike criterion[100] is given by $\mathrm{AIC} = \chi^2_{min} + 2n$, where n is the number of parameters in the model (the degree of correlation between them is not taken into account). Considering that CDE with PR potential has an effective number of parameters between 2 and 3 we find $2.5 < \mathrm{AIC}_{\rm CDE} - \mathrm{AIC}_\Lambda < 6$ for the scenarios explored in Table 1, which leads to a positive preference for ΛCDM, again using Jeffreys' scale.[18,98,99]

Peebles-Ratra potential. CDE might probe helpful to explain the well-known tension between local and cosmological values of H_0. Any detection of a varying dark energy potential or interaction would clearly constitute a major result and it is therefore important to monitor the constraints that newer data impose. This is particularly true in view of earlier results that detected a non-zero value of the coupling β.[27,28]

We confirm the existence of a peak in the marginalized posterior distribution for β and α, more or less evident depending on the data set combination. While for P18 + SH0ES + H0LICOW $\beta > 0$ at 3σ and $\alpha > 0$ at nearly 7σ, inclusion of background data reduces the evidence to $\beta > 0$ at 2.3σ and $\alpha > 0$ at nearly 3.8σ. Inclusion of CMB lensing shifts both values to be compatible with ΛCDM within 2σ. We find it important to stress that specifically CMB lensing prefers ΛCDM. In all cases, we find that, overall, the peak does not correspond to a better Bayes ratio and ΛCDM remains the favored model when employing Bayesian model comparison, given the extra parameters introduced within the model. With regard to H_0, we find that under the P18+SH0ES+H0LICOW combined data set the simple coupled model with constant coupling investigated in this work leads to a value in 3.5σ tension with (7), or in 4.3σ tension when including further background data. The values of σ_8 are also similar to those found in the ΛCDM (i.e. $\sigma_8 \sim 0.80 - 0.82$), even when RSD data are considered together with CMB and background data. In this case we find $\beta = 0.010^{+0.003}_{-0.009}$ and $\beta = 0.015^{+0.007}_{-0.008}$, with and without CMB lensing, respectively. For the values of the coupling strength preferred by the data we find the typical increase of the mass of the DM particles to be $m(\phi_{ini})/m^{(0)} - 1 \lesssim \mathcal{O}(1)\%$.

The question that naturally arises is then, which modification of CDE can help alleviating the tensions? One can immediately suppose that a varying β can go some way towards this. Or, it could be that a model with both energy- and momentum-couplings (see e.g.[104]), which can introduce a weaker gravity, helps with the tensions. These issues will be investigated in future publications.

Acknowledgments

I want to express my most sincere gratitude to Prof. Mavromatos for his invitation to give my talk at the very interesting DM1 parallel session of the MG16 Meeting. I also want to thank Prof. Luca Amendola and Valeria Pettorino for their collaboration in the work my presentation of July 5th 2021 was based on, namely arXiv:2004.00610.[1]

References

1. A. Gómez-Valent, V. Pettorino and L. Amendola, Update on coupled dark energy and the H_0 tension, *Phys. Rev. D* **101**, p. 123513 (2020).
2. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
3. S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
4. S. Cole *et al.*, The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications, *Mon. Not. Roy. Astron. Soc.* **362**, 505 (2005).

5. D. J. Eisenstein *et al.*, Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, *Astrophys. J.* **633**, 560 (2005).
6. G. Hinshaw *et al.*, Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results, *Astrophys. J. Suppl.* **208**, p. 19 (2013).
7. P. A. R. Ade *et al.*, Planck 2013 results. I. Overview of products and scientific results, *Astron. Astrophys.* **571**, p. A1 (2014).
8. P. A. R. Ade *et al.*, Planck 2015 results. XIII. Cosmological parameters, *Astron. Astrophys.* **594**, p. A13 (2016).
9. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)].
10. S. Weinberg, The Cosmological Constant Problem, *Rev. Mod. Phys.* **61**, 1 (1989), [,569(1988)].
11. J. Martin, Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask), *Comptes Rendus Physique* **13**, 566 (2012).
12. J. Solà, Cosmological constant and vacuum energy: old and new ideas, *J. Phys. Conf. Ser.* **453**, p. 012015 (2013).
13. P. J. E. Peebles and B. Ratra, The Cosmological Constant and Dark Energy, *Rev. Mod. Phys.* **75**, 559 (2003).
14. T. Padmanabhan, Cosmological constant: The Weight of the vacuum, *Phys. Rept.* **380**, 235 (2003).
15. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *Astrophys. J.* **876**, p. 85 (2019).
16. E. Macaulay, I. K. Wehus and H. K. Eriksen, Lower Growth Rate from Recent Redshift Space Distortion Measurements than Expected from Planck, *Phys. Rev. Lett.* **111**, p. 161301 (2013).
17. H. Hildebrandt *et al.*, KiDS+VIKING-450: Cosmic shear tomography with optical+infrared data, *Astron. Astrophys.* **633**, p. A69 (2020).
18. L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations* (Cambridge Univ. Press, Cambridge, 2015).
19. A. Joyce, B. Jain, J. Khoury and M. Trodden, Beyond the Cosmological Standard Model, *Phys. Rept.* **568**, 1 (2015).
20. C. Wetterich, The Cosmon model for an asymptotically vanishing time dependent cosmological 'constant', *Astron. Astrophys.* **301**, 321 (1995).
21. L. Amendola, Coupled quintessence, *Phys. Rev.* **D62**, p. 043511 (2000).
22. R. D. Peccei, J. Solà and C. Wetterich, Adjusting the Cosmological Constant Dynamically: Cosmons and a New Force Weaker Than Gravity, *Phys. Lett.* **B195**, 183 (1987).
23. C. Wetterich, Cosmology and the Fate of Dilatation Symmetry, *Nucl. Phys.* **B302**, 668 (1988).
24. P. J. E. Peebles and B. Ratra, Cosmology with a Time Variable Cosmological Constant, *Astrophys. J.* **325**, p. L17 (1988).
25. B. Ratra and P. J. E. Peebles, Cosmological Consequences of a Rolling Homogeneous Scalar Field, *Phys. Rev.* **D37**, p. 3406 (1988).
26. V. Pettorino, L. Amendola, C. Baccigalupi and C. Quercellini, Constraints on coupled dark energy using CMB data from WMAP and SPT, *Phys. Rev.* **D86**, p. 103507 (2012).
27. V. Pettorino, Testing modified gravity with Planck: the case of coupled dark energy, *Phys. Rev.* **D88**, p. 063519 (2013).

28. P. A. R. Ade *et al.*, Planck 2015 results. XIV. Dark energy and modified gravity, *Astron. Astrophys.* **594**, p. A14 (2016).
29. K. C. Wong *et al.*, H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes (2019).
30. H. Gil-Marín *et al.*, The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: structure growth rate measurement from the anisotropic quasar power spectrum in the redshift range $0.8 < z < 2.2$, *Mon. Not. Roy. Astron. Soc.* **477**, 1604 (2018).
31. J.-Q. Xia, New Limits on Coupled Dark Energy from Planck, *JCAP* **1311**, p. 022 (2013).
32. A. Pourtsidou and T. Tram, Reconciling CMB and structure growth measurements with dark energy interactions, *Phys. Rev.* **D94**, p. 043518 (2016).
33. C. van de Bruck, J. Mifsud and J. Morrice, Testing coupled dark energy models with their cosmological background evolution, *Phys. Rev.* **D95**, p. 043513 (2017).
34. C. van de Bruck and J. Mifsud, Searching for dark matter - dark energy interactions: going beyond the conformal case, *Phys. Rev.* **D97**, p. 023506 (2018).
35. Y.-H. Li, J.-F. Zhang and X. Zhang, Exploring the full parameter space for an interacting dark energy model with recent observations including redshift-space distortions: Application of the parametrized post-Friedmann approach, *Phys. Rev.* **D90**, p. 123007 (2014).
36. Y.-H. Li, J.-F. Zhang and X. Zhang, Testing models of vacuum energy interacting with cold dark matter, *Phys. Rev.* **D93**, p. 023002 (2016).
37. E. Di Valentino, A. Melchiorri and O. Mena, Can interacting dark energy solve the H_0 tension?, *Phys. Rev.* **D96**, p. 043503 (2017).
38. E. G. M. Ferreira, J. Quintin, A. A. Costa, E. Abdalla and B. Wang, Evidence for interacting dark energy from BOSS, *Phys. Rev.* **D95**, p. 043520 (2017).
39. A. A. Costa, X.-D. Xu, B. Wang and E. Abdalla, Constraints on interacting dark energy models from Planck 2015 and redshift-space distortion data, *JCAP* **1701**, p. 028 (2017).
40. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, The H_0 tension in light of vacuum dynamics in the Universe, *Phys. Lett.* **B774**, 317 (2017).
41. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Possible signals of vacuum dynamics in the Universe, *Mon. Not. Roy. Astron. Soc.* **478**, 4357 (2018).
42. M. Martinelli, N. B. Hogg, S. Peirone, M. Bruni and D. Wands, Constraints on the interacting vacuum–geodesic CDM scenario, *Mon. Not. Roy. Astron. Soc.* **488**, 3423 (2019).
43. P. Agrawal, G. Obied and C. Vafa, H_0 tension, swampland conjectures, and the epoch of fading dark matter, *Phys. Rev. D* **103**, p. 043523 (2021).
44. S. Pan, G. S. Sharov and W. Yang, Field theoretic interpretations of interacting dark energy scenarios and recent observations, *Phys. Rev. D* **101**, p. 103533 (2020).
45. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, Dynamical dark energy vs. Λ = const in light of observations, *EPL* **121**, p. 39001 (2018).
46. A. Gómez-Valent and J. Solà Peracaula, Density perturbations for running vacuum: a successful approach to structure formation and to the σ_8-tension, *Mon. Not. Roy. Astron. Soc.* **478**, 126 (2018).
47. P. Tsiapi and S. Basilakos, Testing dynamical vacuum models with CMB power spectrum from Planck, *Mon. Not. Roy. Astron. Soc.* **485**, 2505 (2019).
48. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Perez and C. Moreno-Pulido, Running vacuum against the H_0 and σ_8 tensions, *EPL* **134**, p. 19001 (2021).

49. L. Amendola, Linear and non-linear perturbations in dark energy models, *Phys. Rev.* **D69**, p. 103524 (2004).
50. V. Pettorino and C. Baccigalupi, Coupled and Extended Quintessence: theoretical differences and structure formation, *Phys. Rev.* **D77**, p. 103003 (2008).
51. L. Amendola and V. Pettorino, Beyond self-acceleration: force- and fluid-acceleration, *Phys. Lett.* **B802**, p. 135214 (2020).
52. C. M. Will, The Confrontation between general relativity and experiment, *Living Rev. Rel.* **9**, p. 3 (2006).
53. B. Elder, V. Vardanyan, Y. Akrami, P. Brax, A.-C. Davis and R. S. Decca, The Classical Symmetron Force in Casimir Experiments, *Phys. Rev.* **D101**, p. 064065 (2020).
54. P. Brax, A.-C. Davis, B. Elder and L. K. Wong, Constraining screened fifth forces with the electron magnetic moment, *Phys. Rev.* **D97**, p. 084050 (2018).
55. J. Bergé, P. Brax, G. Métris, M. Pernot-Borràs, P. Touboul and J.-P. Uzan, MICROSCOPE Mission: First Constraints on the Violation of the Weak Equivalence Principle by a Light Scalar Dilaton, *Phys. Rev. Lett.* **120**, p. 141101 (2018).
56. G. Efstathiou and S. Gratton, The evidence for a spatially flat Universe (2020).
57. R. K. Sachs and A. M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background, *Astrophys. J.* **147**, 73 (1967), [Gen. Rel. Grav.39,1929(2007)].
58. L. Amendola, V. Pettorino, C. Quercellini and A. Vollmer, Testing coupled dark energy with next-generation large-scale observations, *Phys. Rev.* **D85**, p. 103008 (2012).
59. M. Baldi, V. Pettorino, G. Robbers and V. Springel, Hydrodynamical N-body simulations of coupled dark energy cosmologies, *Mon. Not. Roy. Astron. Soc.* **403**, 1684 (2010).
60. M. Baldi and V. Pettorino, High-z massive clusters as a test for dynamical coupled dark energy, *Mon. Not. Roy. Astron. Soc.* **412**, p. L1 (2011).
61. L. Amendola and C. Quercellini, Tracking and coupled dark energy as seen by WMAP, *Phys. Rev.* **D68**, p. 023514 (2003).
62. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, Dynamical dark energy: scalar fields and running vacuum, *Mod. Phys. Lett.* **A32**, p. 1750054 (2017).
63. J. Ooba, B. Ratra and N. Sugiyama, Planck 2015 constraints on spatially-flat dynamical dark energy models, *Astrophys. Space Sci.* **364**, p. 176 (2019).
64. J. Solà Peracaula, A. Gómez-Valent and J. de Cruz Pérez, Signs of Dynamical Dark Energy in Current Observations, *Phys. Dark Univ.* **25**, p. 100311 (2019).
65. D. Blas, J. Lesgourgues and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes, *JCAP* **1107**, p. 034 (2011).
66. B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, *JCAP* **1302**, p. 001 (2013).
67. A. Lewis, GetDist: a Python package for analysing Monte Carlo samples (2019).
68. A. Heavens, Y. Fantaye, A. Mootoovaloo, H. Eggers, Z. Hosenie, S. Kroon and E. Sellentin, Marginal Likelihoods from Monte Carlo Markov Chains (2017).
69. N. Aghanim *et al.*, Planck 2018 results. VIII. Gravitational lensing, *Astron. Astrophys.* **641**, p. A8 (2020).
70. P. Carter, F. Beutler, W. J. Percival, C. Blake, J. Koda and A. J. Ross, Low Redshift Baryon Acoustic Oscillation Measurement from the Reconstructed 6-degree Field Galaxy Survey, *Mon. Not. Roy. Astron. Soc.* **481**, 2371 (2018).

71. H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres and M. D. Olmstead, The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, *Mon. Not. Roy. Astron. Soc.* **465**, 1757 (2017).
72. E. A. Kazin *et al.*, The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature, *Mon. Not. Roy. Astron. Soc.* **441**, 3524 (2014).
73. T. M. C. Abbott *et al.*, Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1, *Mon. Not. Roy. Astron. Soc.* **483**, 4866 (2019).
74. M. Blomqvist *et al.*, Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14, *Astron. Astrophys.* **629**, p. A86 (2019).
75. V. de Sainte Agathe *et al.*, Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14, *Astron. Astrophys.* **629**, p. A85 (2019).
76. D. M. Scolnic *et al.*, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *Astrophys. J.* **859**, p. 101 (2018).
77. A. G. Riess *et al.*, Type Ia Supernova Distances at Redshift > 1.5 from the Hubble Space Telescope Multi-cycle Treasury Programs: The Early Expansion Rate, *Astrophys. J.* **853**, p. 126 (2018).
78. R. Jiménez, L. Verde, T. Treu and D. Stern, Constraints on the equation of state of dark energy and the Hubble constant from stellar ages and the CMB, *Astrophys. J.* **593**, 622 (2003).
79. J. Simon, L. Verde and R. Jiménez, Constraints on the redshift dependence of the dark energy potential, *Phys. Rev.* **D71**, p. 123001 (2005).
80. D. Stern, R. Jiménez, L. Verde, M. Kamionkowski and S. A. Stanford, Cosmic Chronometers: Constraining the Equation of State of Dark Energy. I: H(z) Measurements, *JCAP* **1002**, p. 008 (2010).
81. M. Moresco *et al.*, Improved constraints on the expansion rate of the Universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers, *JCAP* **1208**, p. 006 (2012).
82. C. Zhang, H. Zhang, S. Yuan, T.-J. Zhang and Y.-C. Sun, Four new observational $H(z)$ data from luminous red galaxies in the Sloan Digital Sky Survey data release seven, *Res. Astron. Astrophys.* **14**, 1221 (2014).
83. M. Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$, *Mon. Not. Roy. Astron. Soc.* **450**, L16 (2015).
84. M. Moresco, L. Pozzetti, A. Cimatti, R. Jiménez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro and D. Wilkinson, A 6% measurement of the Hubble parameter at $z \sim 0.45$: direct evidence of the epoch of cosmic re-acceleration, *JCAP* **1605**, p. 014 (2016).
85. A. L. Ratsimbazafy, S. I. Loubser, S. M. Crawford, C. M. Cress, B. A. Bassett, R. C. Nichol and P. Väisänen, Age-dating Luminous Red Galaxies observed with the Southern African Large Telescope, *Mon. Not. Roy. Astron. Soc.* **467**, 3239 (2017).
86. A. Gómez-Valent, Quantifying the evidence for the current speed-up of the Universe with low and intermediate-redshift data. A more model-independent approach, *JCAP* **1905**, p. 026 (2019).
87. F. Qin, C. Howlett and L. Staveley-Smith, The redshift-space momentum power spectrum — II. Measuring the growth rate from the combined 2MTF and 6dFGSv surveys, *Mon. Not. Roy. Astron. Soc.* **487**, 5235 (2019).

88. F. Shi *et al.*, Mapping the Real Space Distributions of Galaxies in SDSS DR7: II. Measuring the growth rate, clustering amplitude of matter and biases of galaxies at redshift 0.1, *Astrophys. J.* **861**, p. 137 (2018).
89. F. Simpson, C. Blake, J. A. Peacock, I. Baldry, J. Bland-Hawthorn, A. Heavens, C. Heymans, J. Loveday and P. Norberg, Galaxy and mass assembly: Redshift space distortions from the clipped galaxy field, *Phys. Rev.* **D93**, p. 023525 (2016).
90. C. Blake *et al.*, Galaxy And Mass Assembly (GAMA): improved cosmic growth measurements using multiple tracers of large-scale structure, *Mon. Not. Roy. Astron. Soc.* **436**, p. 3089 (2013).
91. C. Blake *et al.*, The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9, *Mon. Not. Roy. Astron. Soc.* **415**, p. 2876 (2011).
92. F. G. Mohammad *et al.*, The VIMOS Public Extragalactic Redshift Survey (VIPERS): Unbiased clustering estimate with VIPERS slit assignment, *Astron. Astrophys.* **619**, p. A17 (2018).
93. L. Guzzo *et al.*, A test of the nature of cosmic acceleration using galaxy redshift distortions, *Nature* **451**, 541 (2008).
94. Y.-S. Song and W. J. Percival, Reconstructing the history of structure formation using Redshift Distortions, *JCAP* **0910**, p. 004 (2009).
95. T. Okumura *et al.*, The Subaru FMOS galaxy redshift survey (FastSound). IV. New constraint on gravity theory from redshift space distortions at $z \sim 1.4$, *Publ. Astron. Soc. Jap.* **68**, p. 38 (2016).
96. L. Verde, T. Treu and A. G. Riess, Tensions between the early and late Universe, *Nature Astronomy* **3**, 891 (2019).
97. A. G. Riess, The Expansion of the Universe is Faster than Expected, *Nature Rev. Phys.* **2**, 10 (2020).
98. R. E. Kass and A. E. Raftery, The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature, *J. Amer. Statist. Assoc.* **90**, 773 (1995).
99. K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference* (Springer, New York, 2002).
100. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Autom. Control* **19**, 716 (1974).
101. G. Schwarz, Estimating the Dimension of a Model, *Ann. Stat.* **6**, 461 (1978).
102. D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde, Bayesian measures of model complexity and fit, *J. Roy. Stat. Soc.* **64**, p. 583 (2002).
103. C. Moreno-Pulido and J. Solà Peracaula, Running vacuum in quantum field theory in curved spacetime: renormalizing ρ_{vac} without $\sim m^4$ terms, *Eur. Phys. J. C* **80**, p. 692 (2020).
104. L. Amendola and S. Tsujikawa, Scaling solutions and weak gravity in dark energy with energy and momentum couplings, *JCAP* **06**, p. 020 (2020).

Running vacuum interacting with dark matter or with running gravitational coupling. Phenomenological implications

Joan Solà Peracaula[a,b]

[a] *Departament de Física Quàntica i Astrofísica,*
and
[b] *Institute of Cosmos Sciences,*
Universitat de Barcelona,
Av. Diagonal 647, E-08028 Barcelona, Catalonia, Spain
E-mail: sola@fqa.ub.edu

The cosmological term, Λ, in Einstein's equations is an essential ingredient of the 'concordance' ΛCDM model of cosmology. In this mini-review presentation, we assess the possibility that Λ can be a dynamical quantity, more specifically a 'running quantity' in quantum field theory in curved spacetime. A great deal of phenomenological works have shown in the last few years that this option (sometimes accompanied with a running gravitational coupling) may cure some of the tensions afflicting the ΛCDM. The 'running vacuum models' (RVM's) are characterized by the vacuum energy density, $\rho_{\rm vac}$, being a series of (even) powers of the Hubble rate and its time derivatives. Here we describe the technical quantum field theoretical origin of the RVM structure in FLRW spacetime, which goes well-beyond the original semi-qualitative renormalization group arguments. In particular, we compute the renormalized energy-momentum tensor using the adiabatic regularization procedure and show that it leads to the RVM form. In other words, we find that the renormalized vacuum energy density, $\rho_{\rm vac}(H)$ evolves as a (constant) additive term plus leading dynamical components $\mathcal{O}(H^2)$. There are also $\mathcal{O}(H^4)$ contributions, which can be relevant for the early universe. Remarkably enough, the renormalized $\rho_{\rm vac}(H)$ does not exhibit dangerous terms proportional to the quartic power of the masses ($\sim m^4$) of the fields. It is well-known that these terms have been the main source of trouble since they are responsible for the extreme fine tuning and ultimately for the cosmological constant problem. In its canonical form, the current $\rho_{\rm vac}(H)$ is dominated by a constant term, as it should be, but it acquires a mild dynamical component $\sim \nu H^2$ $(0 < \nu \ll 1)$ which makes the RVM to mimic quintessence.

Keywords: Running Vacuum, Dynamical Dark Energy, Inflation, Dark Matter

1. Introduction

The standard or concordance model of cosmology, the so-called ΛCDM, is a very successful theoretical framework for the description of the Universe.[1] Crucial ingredients of it are left, though, without direct observational evidence and/or a proper theoretical interpretation based on fundamental principles. Such is the situation with dark matter, but also with the cosmological term, Λ, which has been traditionally associated with the vacuum energy density (VED) of the Universe, $\rho_{\rm vac}$. In fact, the notion of VED in cosmology is a most subtle concept, which is challenging theoretical physicists and cosmologists for many decades, specially with the advent of Quantum Theory in general and the more sophisticated machinery of Quantum Field Theory (QFT). The roots of the problem reside in the interpretation of the

cosmological constant (CC) term, Λ, in Einstein's equations as a quantity being connected with the VED, which is also a fundamental concept in QFT. The proposed connection is $\rho_{\rm vac} = \Lambda/(8\pi G_N)$, where G_N is Newton's constant. Accurate measurements of $\rho_{\rm vac}$ in the last decades from distant type Ia supernovae (SnIa) and the cosmic microwave background (CMB),[2–4] have put the foundations of the concordance ΛCDM model of cosmology.[1]

The concordance model is formulated in the context of the Friedman-Lemaitre-Robertson-Walker (FLRW) framework and is deeply ingrained in the General Relativity (GR) paradigm. However, one of the most important drawbacks of GR is that it is a non-renormalizable theory. This can be considered a serious theoretical obstruction for GR to be considered a fundamental theory of gravity. This fact adversely impacts on the ΛCDM too. GR cannot properly describe the short distance effects of gravity, i.e. the ultraviolet regime (UV), only the large distance effects (or infrared regime). As a consequence, GR cannot provide by itself a framework for quantizing gravity (the spacetime metric field) along with the rest of the fundamental interactions (assuming of course that gravity is amenable to be quantized on conventional grounds). In this sense, a first (rougher but effective) approach is to treat gravity as a classical (external or background) field and quantize only the matter fields of the fundamental interactions (electroweak and strong interactions). This is the main program of the semiclassical approach, namely the point of view of QFT in curved spacetime, see[5–8] for a review. This will be our pursue here as well. We shall nevertheless still be able to compute the quantum effects from matter and their contribution to the VED. Not only so, it will allow us to renormalize these (originally UV-divergent) quantum effects and obtain a finite quantity that can be better compared with observations. The renormalized VED found in the present framework is free from the traditional $\sim m^4$ effects (proportional to the quartic powers of the masses of the matter fields). This was first shown in the literature in the work[9] and further extended in.[10] It was demonstrated that $\rho_{\rm vac}$ appears (in the current universe) as a constant (dominant) term plus a dynamical component which varies as $\sim \nu H^2 m_{\rm Pl}^2$, with ν a small (dimensionless) and QFT-computable coefficient, $m_{\rm Pl}$ being the usual Planck mass ($m_{\rm Pl} = G_N^{-1/2}$). Quite obviously, such a result is highly compatible with the ΛCDM since it involves only a small departure from the rigid vacuum term enforced in it. Notwithstanding, it makes a significant qualitative (and quantitative) prediction: the quantum vacuum does not remain static throughout the cosmic evolution but is mildly dynamical, and such dynamics can be effectively evaluated from first principles, for ν (as mentioned) can be accounted for using QFT in curved spacetime. As long as the vacuum turns out to be free from the weird $\sim m^4$ effects obtained in simplified renormalization treatments, the finite result which is proposed here appears theoretically distinct and very tantalizing. It suggests to reassess if the Cosmological Constant Problem (CCP)[11–16] is still in force in the current theoretical framework. Remember that the CCP is one of the hardest and longstanding mysteries of theoretical physics,

and proves to be a serious impediment to reconcile cosmology with particle physics and quantum field theory in general.[17]

The smoother formulation of the quantum vacuum as put forward here is not only convenient from the theoretical point of view, but also from the observational side. It has long been known that, observationally, there appear to be some discrepancies or "tensions" between the results of data analyses from the Planck Collaboration, based on the ΛCDM,[4] and the local (distance ladder) measurements of the Hubble parameter today, the so called H_0 tension,[18,19] see e.g. the various reviews.[20–23] Not only so, there exist also tensions in the large scale structure (LSS) growth data, the so-called σ_8 tension.[21,22,24] Although such discordances in the arena of the 'concordance' model might admit a variety of more mundane astrophysical explanations,[25–27] or even disappear into thin air when more data will be available in the near future, it has been conceded already that they may well be passing the point of being attributable to a fluke.[19]

As it turns out, the smooth quantum vacuum phenomenological framework described here has the capacity to deal successfully with the H_0 and the σ_8 tensions. Such a framework represents the running vacuum model of cosmology (RVM) – see[17] for a detailed review and references therein. The latter was originally motivated by semi-qualitative renormalization group arguments,[28] which were later on formulated in some cases from an action functional viewpoint [29] – and was finally substantiated from full-fledged QFT calculations in.[9,10] For related studies of vacuum energy in QFT, see e.g.[30–35] In this summarized account I will mainly review the emergence of the RVM from the mentioned QFT calculations in curved spacetime, specifically in the context of the FLRW background as well as the most recent phenomenological applications. It is remarkable that it is also possible to derive a string-inspired version of the RVM cosmological model with gravitational anomalies in the early Universe, see[36] for a detailed exposition and[37–40] for more details and spin-off possibilities.

The structure of our presentation is as follows. In sections 2 and 3 we define our basic QFT framework in curved spacetime and the adiabatic expansion procedure. In section 4 we renormalize the energy-momentum tensor in the FLRW context and make contact with the RVM. In section 5 we apply the RVM to tackle the σ_8 and H_0 tensions. Finally, section 6 contains our conclusions and outlook.

2. Non-minimally coupled scalar field in FLRW background

We consider the semiclassical calculation of the energy density of the vacuum fluctuations of a quantized scalar field in FLRW spacetime. It serves as a representative case of study for the kind of results and difficulties we expect to encounter when dealing with generic quantized fields in a curved background. Such QFT calculation implies to perform renormalization since we meet UV-divergent integrals. Conventional approaches such as e.g. the minimal subtraction scheme lead to quartic dependence on the mass of the field ($\sim m^4$) enforcing a serious fine tuning among

the parameters, see e.g.[17] and references therein. Here we avoid using such an unsuccessful method and adopt the adiabatic renormalization procedure (ARP).[5–7] However, we follow the specific approach presented in,[9] which is related to that of.[34] For a full-fledged exposition and calculation details, see.[10]

2.1. *Action and classical energy-momentum tensor*

We start from the Einstein-Hilbert (EH) action for gravity plus matter[a]:

$$S_{\rm EH} = \frac{1}{16\pi G_N}\int d^4x\sqrt{-g}\,R - \int d^4x\sqrt{-g}\,\rho_\Lambda + S_m\,. \tag{1}$$

The matter action S_m is generic at this point, but will be specified shortly. The (constant) term ρ_Λ has dimension of energy density. Because it is unrelated to the matter part (fully contained in S_m), ρ_Λ is usually called the vacuum energy density (VED). However, we will not call it that way here since it is not yet the physical VED, $\rho_{\rm vac}$, as we shall see. For us the term ρ_Λ is just a bare parameter of the EH action, as the gravitational coupling G_N itself. The physical values can only be identified after renormalizing the bare theory. The corresponding gravitational field equations emerging from the variation of the action (1) can be put in the convenient form

$$\mathcal{M}_{\rm Pl}^2\,G_{\mu\nu} = -\rho_\Lambda g_{\mu\nu} + T^{\rm m}_{\mu\nu}\,, \tag{2}$$

where $\mathcal{M}_{\rm Pl} = m_{\rm Pl}/\sqrt{8\pi} = 1/\sqrt{8\pi G_N}$ is the reduced Planck mass, the expression $G_{\mu\nu} = R_{\mu\nu} - (1/2)g_{\mu\nu}R$ is the usual Einstein tensor and $T^{\rm m}_{\mu\nu}$ is the stress-energy-momentum tensor, or just energy-momentum tensor (EMT for short) of matter:

$$T^{\rm m}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\rm m}}{\delta g^{\mu\nu}}\,. \tag{3}$$

For simplicity we will assume that there is only one (matter) field contribution to the EMT on the right hand side of (2) in the form of a real scalar field, ϕ. Such contribution will be denoted $T^{\phi}_{\mu\nu}$. We neglect for the moment the incoherent matter contributions from dust and radiation. They can be added a posteriori without altering the pure QFT aspects on which we wish to focus at this point. We assume that ϕ is non-minimally coupled to gravity. Thus, the part of the action involving ϕ reads

$$S[\phi] = -\int d^4x\sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + \frac{1}{2}(m^2+\xi R)\phi^2\right)\,. \tag{4}$$

The non-minimal coupling of ϕ to gravity is ξ. For $\xi = 1/6$, the massless ($m = 0$) action is conformally invariant. We will keep ξ general to explore its influence. We assume also that ϕ does not couple to itself and hence we shall not consider a

[a]Our conventions and other useful formulae are those of,[9,10] see the appendices of these references.

possible contribution from a classical potential for ϕ in our analysis. In this study, we wish to target mainly the zero-point energy (ZPE) of ϕ.

The classical EMT can be derived from the action (4) and reads as follows:

$$T^{\phi}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S[\phi]}{\delta g^{\mu\nu}} = (1-2\xi)\partial_\mu\phi\partial_\nu\phi + \left(2\xi - \frac{1}{2}\right) g_{\mu\nu}\partial^\sigma\phi\partial_\sigma\phi \\ - 2\xi\phi\nabla_\mu\nabla_\nu\phi + 2\xi g_{\mu\nu}\phi\Box\phi + \xi G_{\mu\nu}\phi^2 - \frac{1}{2}m^2 g_{\mu\nu}\phi^2. \tag{5}$$

Varying the action (4) with respect to ϕ we find the Klein-Gordon (KG) equation in curved spacetime:

$$(\Box - m^2 - \xi R)\phi = 0\,, \tag{6}$$

where $\Box\phi = g^{\mu\nu}\nabla_\mu\nabla_\nu\phi = (-g)^{-1/2}\partial_\mu\left(\sqrt{-g}\,g^{\mu\nu}\partial_\nu\phi\right)$. The FLRW line element for spatially flat three-dimensional geometry can be written in conformal coordinates as $ds^2 = a^2(\tau)\eta_{\mu\nu}dx^\mu dx^\nu$, where $\eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$ is the Minkowski metric in our conventions. The derivative with respect to the conformal time, τ, will be denoted $' \equiv d/d\tau$ and thus the Hubble rate in conformal time reads $\mathcal{H}(\tau) \equiv a'/a$. Since $dt = ad\tau$, the relation between the Hubble rate in cosmic and conformal times is simply $\mathcal{H}(\tau) = aH(t)$, with $H(t) = \dot{a}/a$ ($\dot{} \equiv d/dt$) the usual Hubble rate.

The KG equation (6) in conformally flat coordinates becomes

$$\phi'' + 2\mathcal{H}\phi' - \nabla^2\phi + a^2(m^2 + \xi R)\phi = 0\,, \tag{7}$$

where we used the curvature scalar of spacetime: $R = 6a''/a^3$. The separation of variables in these coordinates, namely $\phi(\tau, x) \sim \int d^3k\ A_{\mathbf{k}}\psi_k(\mathbf{x})\phi_k(\tau) + cc$, can be achieved with $\psi_k(x) = e^{i\mathbf{k}\cdot\mathbf{x}}$. However, in contrast to the Minkowski case we cannot take $\phi_k(\tau) = e^{\pm i\omega_k\tau}$ since the mode frequencies are not constant anymore. The form of the modes $\phi_k(\tau)$ in the curved spacetime case are determined by the KG equation. Starting from the Fourier expansion with separated space and time variables

$$\phi(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}}\left[A_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}}\phi_k(\tau) + A^*_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{x}}\phi^*_k(\tau)\right] \tag{8}$$

(in which $A_{\mathbf{k}}$ and their complex conjugates $A^*_{\mathbf{k}}$ are the classical Fourier coefficients) and substituting it into (7) the mode functions $\phi_k(\tau)$ are determined by solving the nontrivial differential equation

$$\phi_k'' + 2\mathcal{H}\phi_k' + \left(\omega_k^2(m) + a^2\xi R\right)\phi_k = 0\,, \tag{9}$$

where $\omega_k^2(m) \equiv k^2 + a^2m^2$. The mode functions depend only on the modulus $k \equiv |\mathbf{k}|$ of the (co-moving) momenta. In the absence of gravitation ($R = 0$) the physical frequencies are $\tilde{\omega}_k = \sqrt{\tilde{k}^2 + m^2}$, with $\tilde{k} = k/a$ the physical momenta. However, for $R \neq 0$ and $\xi \neq 0$ the frequencies become a function of the gravitational field as well, which makes the particle interpretation hard. If we perform the change of field mode variable $\phi_k = \varphi_k/a$ the above equation simplifies to $\varphi_k'' + \left(\omega_k^2(m) + a^2\left(\xi - 1/6\right)R\right)\varphi_k = 0$. For conformally invariant matter ($m = 0$

and $\xi = 1/6$), such equation boils down to the form $\varphi_k'' + k^2\varphi_k = 0$, whose positive- and negative-energy solutions are just $e^{-ik\tau}$ and $e^{+ik\tau}$, respectively. On the other hand, in the massless case with minimal coupling ($\xi = 0$) the previous equation further simplifies to $\varphi_k'' + (k^2 - a^2R/6)\varphi_k = 0$. In the radiation epoch ($a \propto \tau$, thus $R = 6a''/a^3 = 0$) we find once more the trivial modes $\varphi_k(\tau) = e^{\pm ik\tau}$. Both in the de Sitter ($a = -1/(H\tau)$, H =const.) and matter-dominated ($a \propto \tau^2$) epochs we have $a^2R = 12/\tau^2$, which leads to $\varphi_k'' + (k^2 - 2/\tau^2)\varphi_k = 0$. This equation admits an exact (positive-energy) solution in terms of Hankel functions of half integer order, hence in close analytic form. In the de Sitter case ($\tau < 0$) one may impose the Bunch-Davies vacuum limit $\sim e^{-ik|\tau|}$ in the far remote past ($\tau \to -\infty$) and one finds $\varphi(\tau) \propto \ (1 - i/(k|\tau|))e^{-ik|\tau|}$. The same solution is valid for the matter-dominated era (for which $\tau > 0$). If, however, $m \neq 0$ and/or $\xi \neq 1/6$ no analytic solution of (9) is available, and this leads us to perform a WKB (Wentzel-Kramers-Brillouin) expansion of the solution. But before tackling that method, let us consider the quantization of the scalar field ϕ.

2.2. *Quantum fluctuations*

To account for the quantum fluctuations of the scalar field ϕ we must address the expansion of the field around its background value ϕ_b:

$$\phi(\tau, x) = \phi_b(\tau) + \delta\phi(\tau, x). \tag{10}$$

One starts defining an appropriate vacuum state, called adiabatic vacuum.[41] The vacuum expectation value (VEV) of ϕ is identified with the background value, $\langle 0|\phi(\tau, x)|0\rangle = \phi_b(\tau)$, whereas the VEV of the fluctuation is zero: $\langle\delta\phi\rangle \equiv \langle 0|\delta\phi|0\rangle = 0$. This is not the case for the VEV of the bilinear products of fluctuations, e.g. $\langle\delta\phi^2\rangle \neq 0$. It is convenient to decompose $\langle T_{\mu\nu}^{\phi}\rangle = \langle T_{\mu\nu}^{\phi_b}\rangle + \langle T_{\mu\nu}^{\delta\phi}\rangle$, where $\langle T_{\mu\nu}^{\phi_b}\rangle = T_{\mu\nu}^{\phi_b}$ is the contribution from the classical background part, whereas $\langle T_{\mu\nu}^{\delta\phi}\rangle \equiv \langle 0|T_{\mu\nu}^{\delta\phi}|0\rangle$ is the genuine vacuum contribution from the field fluctuations. Taking into account that ρ_Λ is also part of the vacuum action (1), the full vacuum contribution is the sum

$$\langle T_{\mu\nu}^{\rm vac}\rangle = -\rho_\Lambda g_{\mu\nu} + \langle T_{\mu\nu}^{\delta\phi}\rangle\,. \tag{11}$$

Thus, the total vacuum part receives contributions from both the cosmological term in the action as well as from the quantum fluctuations of the field (the ZPE). However, since these quantities are formally UV-divergent, the physical vacuum can only be identified a posteriori, namely upon suitable regularization and renormalization. For this we adopt the adiabatic method along the lines of.[9,10]

Obviously the classical and quantum parts of the field (10) obey the curved spacetime KG equation (9) separately. Similarly for $\varphi = \varphi_b + \delta\varphi$ (where $\phi = \varphi/a$). Denoting the frequency modes of the fluctuating part $\delta\varphi$ by $h_k(\tau)$, we can write

$$\delta\varphi(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[A_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} h_k(\tau) + A_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} h_k^*(\tau)\right]. \tag{12}$$

Here $A_{\mathbf{k}}$ and $A^{\dagger}_{\mathbf{k}}$ are now the (time-independent) annihilation and creation operators, which satisfy the commutation relations

$$[A_{\mathbf{k}}, A^{\prime\dagger}_{\mathbf{k}}] = \delta(\mathbf{k}-\mathbf{k}'), \qquad [A_{\mathbf{k}}, A^{\prime}_{\mathbf{k}}] = 0. \tag{13}$$

The frequency modes of the fluctuations, $h_k(\tau)$, satisfy the differential equation

$$h_k'' + \Omega_k^2(\tau) h_k = 0 \qquad \Omega_k^2(\tau) \equiv \omega_k^2(m) + a^2\,(\xi - 1/6)R\,. \tag{14}$$

Except in the simple cases mentioned above, the solution of that equation requires a recursive self-consistent iteration, the WKB expansion. One start from

$$h_k(\tau) = \frac{1}{\sqrt{2W_k(\tau)}} \exp\left(i\int^{\tau} W_k(\tilde{\tau})d\tilde{\tau}\right), \tag{15}$$

where the normalization factor $1/\sqrt{2W_k(\tau)}$ insures that the Wronskian condition $h_k' h_k^* - h_k h_k^{*\prime} = i$ is satisfied. It warrants the standard equal-time commutation relations. Functions W_k in the above ansatz obey the (non-linear) equation

$$W_k^2(\tau) = \Omega_k^2(\tau) - \frac{1}{2}\frac{W_k''}{W_k} + \frac{3}{4}\left(\frac{W_k'}{W_k}\right)^2, \tag{16}$$

which is amenable to be solved using the WKB expansion. The latter is applicable only for large k, therefore short wave lengths (as e.g. in geometrical Optics), and weak gravitational fields. The mode functions $h_k(\tau)$ are no longer of the form $\varphi_k(\tau) = e^{\pm i\omega_k \tau}$, hence particles with definite frequencies cannot be strictly defined in a curved background. Notwithstanding, an approximate Fock space interpretation is still feasible if the vacuum is defined as the quantum state which is annihilated by all the operators $A_{\mathbf{k}}$ of the above Fourier expansion. This defines in a precise way the notion of the adiabatic vacuum.[5–8,41]

2.3. *WKB expansion of the mode functions*

In the gravitational context, the WKB expansion leads to the adiabatic regularization procedure (ARP). For a review, see e.g. the classic books.[5,6] The regularization involved in the ARP amounts to subtracted integrals which become UV-finite and hence one obtains direct renormalization of the physical quantities. In the two successives works[9,10] the WKB expansion was performed first up to $4th$ and subsequently up to $6th$ adiabatic order. In the latter case the calculational details are rather cumbersome, but are necessary in order to study the on-shell renormalized theory. Here, however, we will limit ourselves to describe the results up to $4th$ order, which is enough to reormalize the theory off-shell. The counting of adiabatic orders in the WKB expansion follows the number of time derivatives. Thus: k^2 and a are of adiabatic order 0; a' and $\mathcal{H}$ of adiabatic order 1; $a'', a'^2, \mathcal{H}'$ and $\mathcal{H}^2$ as well as R are of adiabatic order 2. Each additional derivative increases the adiabatic order by one unit. The expansion collects the different adiabatic orders:

$$W_k = \omega_k^{(0)} + \omega_k^{(2)} + \omega_k^{(4)} + \omega_k^{(6)} \cdots, \tag{17}$$

General covariance precludes the odd adiabatic orders. The $\omega_k^{(j)}$ can be expressed in terms of $\Omega_k(\tau)$ and its time derivatives. Following[9,10] we consider an off-shell procedure in which the frequency ω_k of a given mode is defined not at the mass m of the particle but at an arbitrary mass scale M:

$$\omega_k \equiv \omega_k(\tau, M) \equiv \sqrt{k^2 + a^2(\tau)M^2}\,. \tag{18}$$

At the moment we will use just the notation ω_k to indicate such off-shell value, and when necessary we will distinguish it from the on-shell one using the forms $\omega_k(M)$ and $\omega_k(m)$, both being of course functions of τ (which we will omit to simplify notation). Working out the second and fourth order terms of (17) one finds[9]

$$\begin{aligned}
\omega_k^{(0)} &= \omega_k\,,\\
\omega_k^{(2)} &= \frac{a^2\Delta^2}{2\omega_k} + \frac{a^2 R}{2\omega_k}(\xi - 1/6) - \frac{\omega_k''}{4\omega_k^2} + \frac{3\omega_k'^2}{8\omega_k^3}\,,\\
\omega_k^{(4)} &= -\frac{1}{2\omega_k}\left(\omega_k^{(2)}\right)^2 + \frac{\omega_k^{(2)}\omega_k''}{4\omega_k^3} - \frac{\omega_k^{(2)\prime\prime}}{4\omega_k^2} - \frac{3\omega_k^{(2)}\omega_k'^2}{4\omega_k^4} + \frac{3\omega_k'\omega_k^{(2)\prime}}{4\omega_k^3}\,.
\end{aligned} \tag{19}$$

The quadratic mass differences $\Delta^2 \equiv m^2 - M^2$ must be counted as being of adiabatic order 2 since they appear in the WKB expansion along with other terms of the same adiabatic order[b]. The on-shell result is recovered for $M = m$, for which $\Delta = 0$ and corresponds to the usual ARP procedure.[5,6] One could extend the expansion up to the next nonvanishing adiabatic order, which is order $6th$, although we refrain from quoting the result here.[10] It is easy to see that the adiabatic expansion becomes an expansion in powers of $\mathcal{H}$ and its time derivatives. For example, the first two derivatives of ω_k read

$$\omega_k' = a^2\mathcal{H}\frac{M^2}{\omega_k}, \qquad \omega_k'' = 2a^2\mathcal{H}^2\frac{M^2}{\omega_k} + a^2\mathcal{H}'\frac{M^2}{\omega_k} - a^4\mathcal{H}^2\frac{M^4}{\omega_k^3}\,, \tag{20}$$

where we recall that $\mathcal{H}$ is the Hubble function in conformal time. From these elementary differentiations one can then compute the more laborious derivatives appearing in the above expressions, such as $\omega_k^{(2)\prime}, \omega_k^{(2)\prime\prime}$ etc. Therefore, the final result appears as an expansion in powers of $\mathcal{H}$ and multiple derivatives of it.

3. Adiabatic expansion of the ZPE

We have now all the necessary ingredients to compute the zero-point energy (ZPE) associated to the quantum vacuum fluctuations in curved spacetime with FLRW metric. We closely follow the presentation of.[9] Inserting the decomposition (10) of the quantum field ϕ in the EMT as given in Eq. (5) and selecting only the fluctuating

[b]In the context of the effective action, it can be justified more formally on replacing m by M in the heat kernel expansion of the propagator, but we shall not take this path here, see[10] for details.

parts $\delta\phi$, the ZPE (which is ssociated to the 00-component) reads

$$\langle T_{00}^{\delta\phi}\rangle = \left\langle \frac{1}{2}\left(\delta\phi'\right)^2 + \left(\frac{1}{2}-2\xi\right)\left(\nabla\delta\phi\right)^2 + 6\xi\mathcal{H}\delta\phi\delta\phi' \right. \\ \left. -2\xi\delta\phi\,\nabla^2\delta\phi + 3\xi\mathcal{H}^2\delta\phi^2 + \frac{a^2m^2}{2}(\delta\phi)^2 \right\rangle . \tag{21}$$

Notice that $\delta\phi'$, the fluctuation of the differentiated field (with respect to conformal time), is given by $\delta\phi' \equiv \delta\partial_0\phi = \partial_0\delta\phi = (\delta\phi)'$. Next we substitute the Fourier expansion of $\delta\phi = \delta\varphi/a$, as given in (12), into Eq. (21) and use the commutation relations (13). At the same time we symmetrize the operator field products $\delta\phi\delta\phi'$ with respect to the creation and annihilation operators. We present the final result in Fourier space, and hence we integrate $\int \frac{d^3k}{(2\pi)^3}(...)$ over solid angles:[9]

$$\begin{aligned}
\langle T_{00}^{\delta\phi(0-4)}\rangle = \frac{1}{8\pi^2a^2}\int dk k^2 \Bigg[& 2\omega_k + \frac{a^4M^4\mathcal{H}^2}{4\omega_k^5} - \frac{a^4M^4}{16\omega_k^7}(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 + 8\mathcal{H}'\mathcal{H}^2 + 4\mathcal{H}^4) \\
& + \frac{7a^6M^6}{8\omega_k^9}(\mathcal{H}'\mathcal{H}^2 + 2\mathcal{H}^4) - \frac{105a^8M^8\mathcal{H}^4}{64\omega_k^{11}} \\
& + \left(\xi-\frac{1}{6}\right)\left(-\frac{6\mathcal{H}^2}{\omega_k} - \frac{6a^2M^2\mathcal{H}^2}{\omega_k^3} + \frac{a^2M^2}{2\omega_k^5}(6\mathcal{H}''\mathcal{H} - 3\mathcal{H}'^2 + 12\mathcal{H}'\mathcal{H}^2)\right. \\
& \left. - \frac{a^4M^4}{8\omega_k^7}(120\mathcal{H}'\mathcal{H}^2 + 210\mathcal{H}^4) + \frac{105a^6M^6\mathcal{H}^4}{4\omega_k^9}\right) \\
& + \left(\xi-\frac{1}{6}\right)^2\left(-\frac{1}{4\omega_k^3}(72\mathcal{H}''\mathcal{H} - 36\mathcal{H}'^2 - 108\mathcal{H}^4) + \frac{54a^2M^2}{\omega_k^5}(\mathcal{H}'\mathcal{H}^2 + \mathcal{H}^4)\right)\Bigg] \\
& + \frac{1}{8\pi^2a^2}\int dk k^2 \left[\frac{a^2\Delta^2}{\omega_k} - \frac{a^4\Delta^4}{4\omega_k^3} + \frac{a^4\mathcal{H}^2M^2\Delta^2}{2\omega_k^5} - \frac{5}{8}\frac{a^6\mathcal{H}^2M^4\Delta^2}{\omega_k^7}\right. \\
& \left. + \left(\xi-\frac{1}{6}\right)\left(-\frac{3a^2\Delta^2\mathcal{H}^2}{\omega_k^3} + \frac{9a^4M^2\Delta^2\mathcal{H}^2}{\omega_k^5}\right)\right] .
\end{aligned} \tag{22}$$

As expected, only even powers of $\mathcal{H}$ remain in the final result. The Minkowskian spacetime result for the on-shell ZPE ($M = m$) is obtained as a very particular case of the above expression for $a = 1$ ($\mathcal{H} = 0$):

$$\langle T_{00}^{\delta\phi}\rangle\Big|_{\text{Minkowski}} = \frac{1}{4\pi^2}\int dk k^2\omega_k = \int \frac{d^3k}{(2\pi)^3}\left(\frac{1}{2}\,\hbar\,\omega_k\right) , \tag{23}$$

where $\hbar$ has been restored only in the last expression for better cognizance. The result is quartically UV-divergent. Usual attempts (e.g. through the minimal subtraction scheme) to regularize and renormalize this quantity by e.g. cancelling the corresponding UV-divergence against the bare ρ_Λ term in the action (1) ends up with the well-known fine-tuning problem, which is considered to be the weirdest and toughest aspect of the CCP – see e.g.[17,42] and references therein. We will certainly not proceed in this way here. We seek (and will find) an alternative way.

4. Renormalization of the VED in curved spacetime: the RVM

The vacuum energy density in the expanding universe can be compared with a Casimir device in which the parallel plates slowly move apart ("expand").[17] Although the total VED cannot be measured, the distinctive effect associated to the presence of the plates, and then also to their increasing separation with time, it can. In a similar fashion, in the cosmological spacetime there is a distinctive non-vanishing spacetime curvature R as compared to Minkowskian spacetime that is changing with the expansion. We expect that the measurable VED must be that one which is associated to purely geometric contributions proportional to R, R^2, $R^{\mu\nu}R_{\mu\nu}$ etc., hence to H^2 and $\dot{H}$ (including higher powers of these quantities in the early Universe).

Following,[9,10] a subtraction of the VEV of the EMT is carried out at an arbitrary mass scale M, playing the role of renormalization point. Taking into account that the only adiabatic orders that are divergent in the case of the EMT are the first four ones, the subtraction at the scale M is performed only up to the fourth adiabatic order. The on-shell value of the EMT can be computed of course at any order. The terms beyond the 4th order are finite. The renormalized EMT in this context therefore reads

$$\langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M) = \langle T^{\delta\phi}_{\mu\nu}\rangle(m) - \langle T^{\delta\phi}_{\mu\nu}\rangle^{(0-4)}(M)\,. \tag{24}$$

Let us apply this procedure to the ZPE part of the EMT, as given by Eq. (22). To ease the presentation of the explicit result, it proves convenient to recover at least in part the more explicit notation (18) so as to distinguish explicitly between the off-shell energy mode $\omega_k(M) = \sqrt{k^2 + a^2M^2}$ (formerly denoted just as ω_k) and the on-shell one $\omega_k(m) = \sqrt{k^2 + a^2m^2}$. With this notation, lengthy but straightforward calculations from equations (22) and (24) lead to the following compact result:[9]

$$\langle T^{\delta\phi}_{00}\rangle_{\rm Ren}(M) = \frac{a^2}{128\pi^2}\left(-M^4 + 4m^2M^2 - 3m^4 + 2m^4\ln\frac{m^2}{M^2}\right) - \left(\xi - \frac{1}{6}\right)$$
$$\times\,\frac{3\mathcal{H}^2}{16\pi^2}\left(m^2 - M^2 - m^2\ln\frac{m^2}{M^2}\right) + \left(\xi - \frac{1}{6}\right)^2\frac{9\left(2\mathcal{H}''\mathcal{H} - \mathcal{H}'^2 - 3\mathcal{H}^4\right)}{16\pi^2a^2}\ln\frac{m^2}{M^2} + \ldots \tag{25}$$

where dots stand just for higher adiabatic orders. The renormalized expression for the vacuum fluctuations, $\langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M)$, is not yet the final one to extract the renormalized VED. As indicated in (11), the latter is obtained from including the contribution from the ρ_Λ-term in the Einstein-Hilbert action (1). Therefore, the renormalized vacuum EMT at the scale M is given by

$$\langle T^{\rm vac}_{\mu\nu}\rangle_{\rm Ren}(M) = -\rho_\Lambda(M)g_{\mu\nu} + \langle T^{\delta\phi}_{\mu\nu}\rangle_{\rm Ren}(M)\,. \tag{26}$$

We distinguish between VED and ZPE: the latter is caused by the vacuum fluctuations of the fields, whereas the former combines the ZPE and the parameter ρ_Λ in the action. The renormalized VED is precisely the 00th component of the above

expression:

$$\rho_{\rm vac}(M)=\frac{\langle T_{00}^{\rm vac}\rangle_{\rm Ren}(M)}{a^2}=\rho_\Lambda(M)+\frac{\langle T_{00}^{\delta\phi}\rangle_{\rm Ren}(M)}{a^2}\,, \tag{27}$$

where we have used the fact that $g_{00}=-a^2$ in the conformal metric. Explicitly, the VED comes out to be

$$\begin{aligned}\rho_{\rm vac}(M)=\rho_\Lambda(M)&+\frac{1}{128\pi^2}\left(-M^4+4m^2M^2-3m^4+2m^4\ln\frac{m^2}{M^2}\right)\\&-\left(\xi-\frac{1}{6}\right)\frac{3\mathcal{H}^2}{16\pi^2a^2}\left(m^2-M^2-m^2\ln\frac{m^2}{M^2}\right)\\&+\left(\xi-\frac{1}{6}\right)^2\frac{9\left(2\mathcal{H}''\mathcal{H}-\mathcal{H}'^2-3\mathcal{H}^4\right)}{16\pi^2a^4}\ln\frac{m^2}{M^2}+\cdots\end{aligned} \tag{28}$$

The (very) interesting aspect about this form of renormalized VED is that the first two terms of this expression (i.e. those not depending on $\mathcal{H}$) exactly cancel when we compute the difference of $\rho_{\rm vac}$ values at two scales, say M and M_0:

$$\begin{aligned}\rho_{\rm vac}(M)-\rho_{\rm vac}(M_0)=&\left(\xi-\frac{1}{6}\right)\frac{3\mathcal{H}^2}{16\pi^2a^2}\left(M^2-M_0^2-m^2\ln\frac{M^2}{M_0^2}\right)\\&+\left(\xi-\frac{1}{6}\right)^2\frac{9}{16\pi^2a^4}\left(\mathcal{H}'^2-2\mathcal{H}''\mathcal{H}+3\mathcal{H}^4\right)\ln\frac{M^2}{M_0^2}\,.\end{aligned} \tag{29}$$

To verify the cancellation of the mentioned terms, we rewrite Einstein's equations (2) using the renormalized parameters and including the higher derivative tensor $H_{\mu\nu}^{(1)}$, which is necessary for renormalization purposes.[5] We need to write only the vacuum part of the EMT since in doing the mentioned subtraction the background contribution of the field ϕ (and any other contribution, indicated below by $\ldots$) will cancel, except the (M-dependent) change of the vacuum EMT at the two scales:

$$\mathcal{M}_{\rm Pl}^2(M)G_{\mu\nu}+\alpha(M)H_{\mu\nu}^{(1)}=\langle T_{\mu\nu}^{\rm vac}\rangle_{\rm Ren}(M)+\ldots \tag{30}$$

On the *r.h.s.* we have used Eq. (26), which can be made more explicit using Eq. (25). We may now subtract side by side (30) at the scales M and M_0 and project the 00th component. Using the explicit form of G_{00} and $H_{00}^{(1)}$ in the FLRW metric we can perform the identifications on both sides of the subtracted equation. In particular, this renders specific expressions for the shifts $\mathcal{M}_{\rm Pl}^2(M)-\mathcal{M}_{\rm Pl}^2(M_0)$ and $\alpha(M)-\alpha(M_0)$, which we need not quote here[9,10] . After performing these identifications, what is left of $\langle T_{\mu\nu}^{\rm vac}\rangle_{\rm Ren}(M)-\langle T_{\mu\nu}^{\rm vac}\rangle_{\rm Ren}(M_0)$ must be zero. This is how we can prove that $\rho_\Lambda(M)-\rho_\Lambda(M_0)$ exactly cancels against the difference of the second term of (29) at the two scales (see Refs.[9,10] for more details):

$$\begin{aligned}&\rho_\Lambda(M)|_{M_0}^M+\frac{1}{128\pi^2}\left(-M^4+4m^2M^2-3m^4+2m^4\ln\frac{m^2}{M^2}\right)\Big|_{M_0}^M\\&=\rho_\Lambda(M)-\rho_\Lambda(M_0)+\frac{1}{128\pi^2}\left(-M^4+M_0^4+4m^2(M^2-M_0^2)-2m^4\ln\frac{M^2}{M_0^2}\right)=0\,.\end{aligned} \tag{31}$$

This important equality, enforced by the renormalized form of Einstein's equations, demonstrates that the renormalized EMT, and in particular the renormalized VED, is free from quartic contributions associated to mass scales. That is why the relation (29) between the VED at the two renormalization points M and M_0 is a smooth function $\sim M^2\mathcal{H}^2$. With no quartic mass term surviving in this renormalization procedure, there is no need of fine tuning. This is, of course, an extremely welcome feature, which obviously impinges on a possible solution of the CCP. Another appealing (and novel) feature of our renormalization framework is the following: Eq. (30) says that in Minkowski spacetime $\langle T_{00}^{\rm vac}\rangle_{\rm Ren}(M) = 0$, and as a result the renormalized VED in flat spacetime is also zero in this context, $\rho_{\rm vac}^{\rm Mink} = 0$.

Let us now assume that we define the renormalized VED in the context of some Grand Unified Theory (GUT) scale $M_0 = M_X$, where typically $M_X \sim 10^{16}$ GeV is associated with the inflationary scale. We denote by $\rho_{\rm vac}(M_X)$ the value of the VED at M_X. We can relate $\rho_{\rm vac}(M_X)$ with the current value of the VED, $\rho_{\rm vac}^0$, assuming that $\rho_{\rm vac}(M = H_0) = \rho_{\rm vac}^0$, where we choose the second scale at today's value of the Hubble parameter, H_0. This quantity can be used as an estimate for the energy scale of the background gravitational field associated to the FLRW universe at present. We neglect the $\sim H^4$ terms in the current universe. Then, according to Eq. (29), the connection between the two values of the VED is

$$\rho_{\rm vac}^0 = \rho_{\rm vac}(M_X) + \frac{3}{16\pi^2}\left(\frac{1}{6} - \xi\right) H_0^2 \left[M_X^2 + m^2 \ln \frac{H_0^2}{M_X^2}\right], \tag{32}$$

or swapping terms on both sides:

$$\rho_{\rm vac}(M_X) = \rho_{\rm vac}^0 - \frac{3\nu}{8\pi}\, H_0^2\, m_{\rm Pl}^2\,. \tag{33}$$

Here we have defined the 'running parameter' for the VED:

$$\nu \equiv \frac{1}{2\pi}\left(\frac{1}{6} - \xi\right) \frac{M_X^2}{m_{\rm Pl}^2}\left(1 + \frac{m^2}{M_X^2} \ln \frac{H_0^2}{M_X^2}\right). \tag{34}$$

We naturally expect $|\nu| \ll 1$, owing to the ratio $M_X^2/m_{\rm Pl}^2 \ll 1$. The vanishing of ν and hence of the dynamical $\sim H^2$ part of (32) is obtained only for conformal coupling: $\xi = 1/6$. There are, however, fermionic contributions to ν as well. They do not depend on ξ, of course, but shall not be addressed here.[43] The accurate determination of ν can only be obtained by fitting the RVM to the overall cosmological data, as it has been done e.g. in.[44–50] These analyses show that ν is positive and of order 10^{-3}. From (33) and (29) we can approximately estimate the VED near our time by taking M of order of the energy scale defined by the numerical value of H around the current epoch:[9]

$$\rho_{\rm vac}(H) \simeq \rho_{\rm vac}^0 + \frac{3\nu}{8\pi}\,(H^2 - H_0^2)\, m_{\rm Pl}^2 = \rho_{\rm vac}^0 + \frac{3\nu}{8\pi G_N}\,(H^2 - H_0^2)\,. \tag{35}$$

This expression reproduces the canonical form of the RVM.[17] A generalized form of the RVM is possible in which an additional term proportional to $\dot{H}$ (of order H^2)

can be included on the *r-h.s.* of (35) with another small coefficient $\tilde{\nu}$, see[9,10] for details. As noted, we have neglected the $\mathcal{O}(H^4)$ terms to derive the previous formulas. These terms can have an impact only for the early universe and produce inflation, but we refer once more the reader to[9,10] for an expanded discussion.

5. Phenomenological applications: H_0 and σ_8 tensions

We are now ready to produce some phenomenological output using the low-energy RVM form (35), or even the generalized one containing the $\sim \dot{H}$ term. In fact, we can use the following extended RVM structure for the VED:[9,10]

$$\rho_{\rm vac}(H) = \frac{3}{8\pi G_N}\left(c_0 + \nu H^2 + \tilde{\nu}\dot{H}\right) + \mathcal{O}(H^4)\,. \tag{36}$$

The additive constant c_0 is fixed by the boundary condition $\rho_{\rm vac}(H_0) = \rho^0_{\rm vac}$. For recent expositions of the RVM and its many phenomenological applications making use of the above form of VED, see[51,52] and references therein[c]. In what follows I will summarize the findings of the recent study,[50] which involves the most complete set of cosmological data. Since the parameters ν and $\tilde{\nu}$ are ultimately to be fitted to observations, we shall simplify our analysis here assuming one single parameter with the choice $\tilde{\nu} = \nu/2$. This does not change anything fundamental. As a result, if we neglect the higher order terms, Eq. (36) takes on the suggestive form

$$\rho_{\rm vac}(H) = \frac{3}{8\pi G_N}\left(c_0 + \frac{\nu}{12}R\right) \equiv \rho_{\rm vac}(R)\,, \tag{37}$$

where $R = 12H^2 + 6\dot{H}$ is the curvature scalar. For this reason we may call this form of the VED the 'RRVM'. Such a RRVM implementation has the double advantage of using one single parameter and provides a safe path to the early epochs of the cosmological evolution since when we approach the radiation dominated era we have $R \simeq 0$, or to be more precise: $R/H^2 \ll 1$. This fact insures that no conflict is generated with the BBN constraints. Early on the RVM has its own mechanism for inflation (as we have already mentioned), but we have no room to address these aspects here, see.[54–57] With one and the same VED given by Eq. (37) we may consider two types of RRVM scenarios, to wit: type I scenario, in which the vacuum interacts with matter; and type II, where matter is conserved at the expense of an exchange between the vacuum and a mildly evolving gravitational coupling $G(H)$. For type I models we assume that the vacuum exchanges energy with cold dark matter (CDM) only, as follows:

$$\dot{\rho}_{dm} + 3H\rho_{dm} = -\dot{\rho}_{\rm vac}\,. \tag{38}$$

[c]It is interesting to mention that at the pure cosmographic/cosmokinetic level (hence in a more model-independent way) the RVM appears also as a favoured model. For example, the form (36) has been used in[53] to study the different types of DE models in the framework of the cosmographic approach. Using the Hubble diagrams for SnIa, quasars, gamma-ray bursts as well as the data on baryonic acoustic oscillations in different combinations, it is found that the RVM fits better the cosmographic data than other DE models, including the concordance ΛCDM.

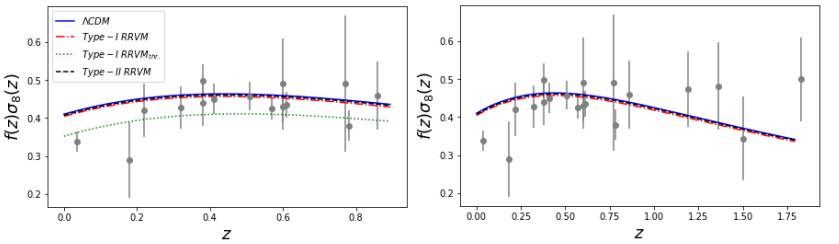


Fig. 1. Theoretically predicted curves of $f(z)\sigma_8(z)$ for the various models (type I and type II), testing the possible existence of a threshold redshift in the case of type I. The data points employed in our analysis are described in.[50] We present the results in two different redshift windows. It can be seen that the type I RRVM with threshold redshift $z_* \simeq 1$ has a most visible and favorable impact on solving the σ_8 tension.

Solving for the matter densities one finds[50]

$$\rho_m(a) = \rho_m^0 a^{-3\xi}\,, \quad \rho_{dm}(a) = \rho_m^0 a^{-3\xi} - \rho_b^0 a^{-3}\,, \tag{39}$$

where $\xi \equiv \frac{1-\nu}{1-\frac{3}{4}\nu}$. They recover the ΛCDM form for $\xi = 1$ ($\nu = 0$). The small departure of ν from zero (or ξ from one) is what permits a mild dynamical vacuum evolution:

$$\rho_{\rm vac}(a) = \rho_{\rm vac}^0 + \left(\frac{1}{\xi} - 1\right)\rho_m^0\left(a^{-3\xi} - 1\right)\,. \tag{40}$$

For type I models we admit also the possibility that the dynamics of vacuum is relatively recent (see e.g.[58]). For instance, one may assume that the vacuum became dynamical in the form (40) at a threshold redshift $z_* \simeq 1$ (so $\rho_{\rm vac}(z) = \rho_{\rm vac}^0$ for $z > z_*$). We shall compare this option with the situation when there is no such threshold. As for type II models, matter is conserved (hence no exchange with vacuum), but the vacuum can still evolve as long as the effective gravitational coupling also evolves (very mildly) with the expansion $G_{\rm eff} = G_{\rm eff}(H)$, starting from an initial value (which enters our fit). In this case, we do not consider the effect of the threshold because it proves to be much smaller. One can show that the approximate behavior of the VED in the present time is (recall that $|\nu| \ll 1$):[50]

$$\rho_{\rm vac}(a) = \frac{3c_0}{8\pi G_N}(1 + 4\nu) + \nu\rho_m^0 a^{-3} + \mathcal{O}(\nu^2)\,. \tag{41}$$

Once more, for $\nu = 0$ the VED remains constant at the value $\rho_{\rm vac} = 3c_0/(8\pi G_N) = \Lambda/(8\pi G_N)$, but otherwise it shows a moderate dynamics as in the type I case. One can also show that the effective gravitational coupling evolves approximately as $G_{\rm eff}(a) \simeq G_N(1 + \epsilon \ln a)$ in the current epoch (with $0 < \epsilon \ll 1$ of order ν), thus confirming the very mild (logarithmic) evolution of G. Since $\epsilon > 0$ the gravitational strength exhibits an asymptotically free behavior (i.e. being smaller in the past).

For an accurate comparison of the theoretical predictions and the observations, the LSS formation data are also of paramount importance, all the more if we take into account that one of the aforementioned ΛCDM tensions (the σ_8 one) stems

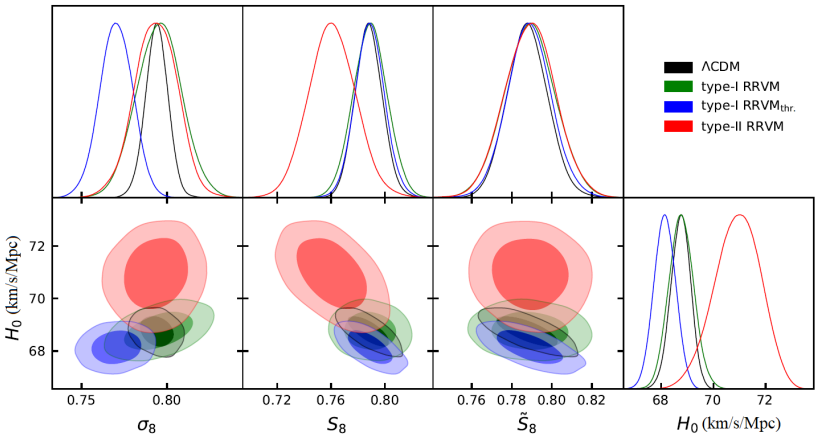


Fig. 2. The 1σ and 2σ c.l. contours in the H_0-$\sigma_8, S_8, \tilde{S}_8$ planes and the corresponding one-dimensional posteriors for the RRVM's obtained from the Baseline+H_0 data set (meaning that the local value of H_0 is included along with the remaining data, see[50]). It is apparent that the type II model alleviates the H_0 tension without spoiling the σ_8 one, whereas the type I model with threshold redshift $z_* \simeq 1$ can fully fix the latter (see also Fig. 1) but cannot reduce the former.

from it. As we shall see, allowing for some evolution of the vacuum can be the clue to alleviate the σ_8 tension since such dynamics affects significantly the cosmological perturbations.[59] We consider the perturbed, spatially flat, FLRW metric $ds^2 = -dt^2 + (\delta_{ij} + h_{ij})dx^i dx^j$, in which h_{ij} stands for the metric fluctuations. These fluctuations are nontrivially coupled to the matter density perturbations $\delta_m = \delta\rho_m/\rho_m$. We have implemented the full perturbations analysis in the context of the Einstein-Boltzmann code `CLASS`[60] (in the synchronous gauge).[50] Since baryons do not interact with the time-evolving VED the perturbed conservation equations are not directly affected. However, the corresponding equation for CDM is modified in the following way:

$$\dot{\delta}_{dm} + \frac{\dot{h}}{2} - \frac{\dot{\rho}_{\rm vac}}{\rho_{dm}}\delta_{dm} = 0\,, \tag{42}$$

with $h = h_{ii}$ denoting the trace of h_{ij}. We remark that the term $\dot{\rho}_{\rm vac}$ is nonvanishing for these models. Thus, it affects the fluctuations of CDM in a way which produces a departure from the ΛCDM. The above equation is, of course, coupled with the metric fluctuations and the combined system must be solved numerically. The analysis of the linear LSS regime is performed with the help of the weighted linear growth $f(z)\sigma_8(z)$, where $f(z)$ is the growth factor and $\sigma_8(z)$ is the rms mass fluctuation amplitude on scales of $R_8 = 8\,h^{-1}$ Mpc at redshift z. The quantity $\sigma_8(z)$ is directly provided by `CLASS`. Similarly, we can extract the (observationally measured) linear growth function $f(a)$ directly from the matter power spectrum $P_m(a, \vec{k})$, which is computed numerically by `CLASS` under adiabatic initial conditions. The results can be seen in Figures 1 and 2 for the various models.

To compare the RRVM's (types I and II) with the ΛCDM, we have defined a joint likelihood function $\mathcal{L}$. The overall fitting results are reported in Tables 1 and 2. The total χ^2 to be minimized in our case is given by

$$\chi^2_{\rm tot} = \chi^2_{\rm SnIa} + \chi^2_{\rm BAO} + \chi^2_H + \chi^2_{{\rm f}\sigma_8} + \chi^2_{\rm CMB}\,. \tag{43}$$

The above χ^2 terms are defined in the standard way from the data including the covariance matrices.[16] In particular, the χ^2_H part may contain or not the local H_0 value measured by Riess, see.[50] The local determination of H_0 (which is around 4σ away from the corresponding Planck 2018 value based on the CMB) is the origin of the so-called H_0 tension.[20–23] Taking into account that the RRVM's of type I and II have one and two more parameters, respectively, as compared to the ΛCDM, a fairer model comparison is achieved by computing the numerical differences between the Deviance Information Criterion (DIC) of the ΛCDM model against the RRVM's: $\Delta{\rm DIC} = {\rm DIC}_{\Lambda{\rm CDM}} - {\rm DIC}_{\rm RRVM}$. These differences will be (and in fact are) positive if the RRVM's fit better the overall data than the ΛCDM. The DIC is defined by[61] ${\rm DIC} = \chi^2(\overline{\theta}) + 2p_D$. Here $p_D = \overline{\chi^2} - \chi^2(\overline{\theta})$ is the effective number of parameters of the model, and $\overline{\chi^2}$ and $\overline{\theta}$ the mean of the overall χ^2 distribution and of the parameters, respectively. The posterior distributions and corresponding constraints for the various dataset combinations have been obtained with `Montepython`[62] in combination with `CLASS`.[60]

The DIC value can be obtained from the Markov chains generated with `MontePython`. If $+5 < \Delta{\rm DIC} < +10$ one should conclude strong evidence favoring the RRVM's over the ΛCDM. For $\Delta{\rm DIC} > +10$ it is said that the evidence is very strong. Such is the case when we use a threshold redshift $z_* \simeq 1$ in type I RRVM. In stark contrast, when the threshold is removed we find only moderate evidence against it (viz. $-3 < \Delta{\rm DIC} < -2$), although the fitting performance is still slightly better (smaller $\chi^2_{\rm min}$) than the ΛCDM. The effect of the threshold can be very important and suggests that a mild dynamics of the vacuum is welcome, especially if it gets activated at around the epoch when the vacuum dominance appears, namely at $z \simeq 1$. Unfortunately, type I models with fixed $G_{\rm eff} = G_N$ do not help an inch to solve the H_0 tension since the value of H_0 remains stuck around the CMB value.[50] In stark contrast, type II models can alleviate the two tensions at a time. The overall $\Delta{\rm DIC}$ value of the fit is quite significant (+5.5), still in the strong evidence region, providing values of H_0 markedly higher as compared to type I models (specifically $H_0 = 70.93^{+0.93}_{-0.87}$ Km/s/Mpc[50]) along with σ_8 and S_8 values in the needed moderate range ($\sigma_8 = 0.794^{+0.013}_{-0.012}$ and $S_8 = 0.761^{+0.018}_{-0.017}$).[50] The values of S_8 in all RRVM's are perfectly compatible with recent weak lensing and galaxy clustering measurements.[63] For type II models a related observable analogous to (but different from) S_8 is $\tilde{S}_8 = S_8\sqrt{G_{\rm eff}(0)/G_N}$. We show the corresponding contours in Fig. 2. The net outcome of this analysis is that the only model capable of alleviating the two tensions (H_0 and σ_8) is RRVM of type II, whereas the type I model can (fully) solve the σ_8 tension but has no bearing on the H_0 one.

6. Conclusions and outlook

In this short review presentation I have described the renormalization of the energy-momentum tensor (EMT) of a real quantum scalar field non-minimally coupled to classical gravity in the cosmological context. The main aim was to show that, out of the very quantum effects of matter, there emerges the running vacuum model (RVM) structure, which is the effective form of the renormalized vacuum energy density (VED). The method is based on an off-shell extension of the adiabatic regularization and renormalization procedure, which we have used in the recent works[9,10] and where for the first time we provided a calculation of the zero point energy (ZPE) of a quantum scalar field that is free from the need of extreme fine tuning. The latter is well-known to be one of the most striking and bizarre aspects of the cosmological constant problem.[11] The absence of need for fine-tuning is related to the non appearance in our framework of the terms which are proportional to the quartic mass of the fields, i.e. $\sim m^4$. In the standard model of particle physics, these terms are usually responsible for the exceedingly large contributions to the VED and requires preposterous fine-tuning with the renormalized vacuum parameter ρ_Λ in the action.[17] The calculational procedure in our approach is based on the WKB expansion of the field modes in the FLRW spacetime and the use of an appropriately renormalized EMT. The latter is obtained by performing a substraction of its on-shell value (i.e. the value defined on the mass shell m of the quantized field) at an arbitrary renormalization point M. The resulting EMT becomes finite because we subtract the first four adiabatic orders (the only ones that can be divergent). Since the off-shell renormalized EMT is a function of the arbitrary renormalization point M, we can compare the renormalized result at different epochs of the cosmic history. The introduction of such 'sliding' scale leads to the renormalization group (RG) flow, which in the case of FLRW cosmology we may associate in a natural way with Hubble's expansion rate H.

While the RG approach was actually the first qualitative idea behind the RVM,[17] with the present QFT calculations in curved spacetime we provide for the first time[9,10] a solid foundation of the RVM, in which the dynamical structure of the VED is seen to emerge from the quantum effects associated with the adiabatic renormalization of the EMT. Let us mention that even though our QFT calculation has been simplified by the use of a single (real) quantum scalar field, further investigations show that the generalization of these results for multiple fields, involving scalar as well as vector and fermionic components, lead as well to the generic RVM structure mentioned here up to (nontrivial) computation details.[43]

At the end of the day, we have been able to show that the genuine form of the VED for the current universe can be achieved from direct calculations of QFT in the FLRW spacetime. In such structure, the powers of H (and its time derivatives) are of even adiabatic order. This means that all of the allowed powers effectively carry an even number of time derivatives of the scale factor, which is essential to preserve the general covariance of the action. Linear terms (cubic, or in general of odd order)

in H are incompatible with such a covariance and in fact do not appear in the final result. To be more precise, all terms with an odd number of time derivatives of the scale factor are ruled out. The form of $\rho_{\rm vac}(H)$ predicted by the RVM at low energies is remarkably simple but it certainly goes beyond a rigid cosmological constant term. It consists of an additive constant (to be identified basically with the cosmological term) together with a small dynamical component $\sim \nu H^2$, in which the dimensionless parameter ν can be computed from the underlying QFT framework and is predicted to be small ($|\nu| \ll 1$). Ultimately, its value can only be known upon fitting the RVM to the overall cosmological data. The physical outcome is that today's cosmic vacuum is mildly dynamical. In fact, in previous works the model has been phenomenologically fitted to a large wealth of cosmological data and the running parameter ν has been found to be positive and in the ballpark of $\sim 10^{-3}$ cf.[44–47]

Recent phenomenological analyses involving a large set of updated cosmological data on SnIa+H(z)+BAO+LSS+CMB [50,64] and analyzed with the Boltzmann code `CLASS`[60] confirm to a large extent the results reported in the aforementioned works, which were carried out within an approximate treatment of the CMB. The basic result is still the same, they point to substantial evidence that a mild dynamics of the cosmic vacuum is helpful to describe the overall cosmological observations as compared to the standard cosmological model with a rigid Λ-term. From our analysis of two variants of the RVM, we have found that for type I models the σ_8 tension can be fully overcome ($\lesssim 0.4\sigma$ c.l.) provided there exists a threshold redshift $z_* \simeq 1$ where the vacuum dynamics is triggered. Solving the H_0 tension, however, proves more demanding as it requires the combination of vacuum dynamics with running G, which is the characteristic of the type II models. Interestingly enough, the two tensions can actually be dealt with at a time, the H_0 remaining at $\sim 1.6\sigma$ and the σ_8 one at $\sim 1.3\sigma$ (or at only $\sim 0.4\sigma$ if stated in terms of S_8).[63] The successful cutback of the two tensions is highly remarkable and is strongly supported by standard information criteria, such as the deviance information criterion (DIC). More work will be needed, of course, to confirm if the RVM can fully solve the σ_8 and H_0 tensions.[64] It will depend also on the upcoming data in the next few years.

Acknowledgements

It is my pleasure to thank N.E. Mavromatos for inviting me to speak in the DM1 parallel session: "Interacting Dark Matter" of the MG16 Marcel Grossmann virtual Conference, July 5-10 2021. Work partially funded by projects PID2019-105614GB-C21 and FPA2016-76005-C2-1-P (MINECO, Spain), 2017-SGR-929 (Generalitat de Catalunya) and CEX2019-000918-M (ICCUB). The author acknowledges participation in the COST Association Action CA18108 "*Quantum Gravity Phenomenology in the Multimessenger Approach (QG-MM)*". This presentation is based in part on works with A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido. I thank them warmly for the enjoyable collaboration.

References

1. P. J. E. Peebles, Astrophys. J. **284** (1984) 439; L.M. Krauss and M.S. Turner, Gen. Rel. Grav. **27** (1995) 1137; J. P. Ostriker, Paul J. Steinhardt, Nature **377** (1995) 600.
2. A. G. Riess et al., Astron. J. **116** (1998) 1009; S. Perlmutter et al. ApJ **517** (1999) 565.
3. D. M. Scolnic et al., Astrophys. J. **859** (2018) 101.
4. N. Aghanim et al. [Planck Collab.], A & A **641** (2020) A6.
5. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge U. Press (1982).
6. L. E. Parker and D. J. Toms *Quantum Field Theory in Curved Spacetime: Quantized fields and gravity*, Cambridge U. Press (2009).
7. W. Fulling, *Aspects of Quantum Field Theory in Curved Space-Time*, Cambridge U. Press, (1989).
8. V. F. Mukhanov and S. Winitzki, *Quantum Effects in Gravity*, Cambridge U. Press (2007).
9. C. Moreno-Pulido. and J. Solà Peracaula, Eur. Phys. J. C**80** (2020) 692.
10. C. Moreno-Pulido and J. Solà Peracaula, *Renormalizing the vacuum energy in cosmological spacetime: Implications for the cosmological constant problem* (in preparation).
11. S. Weinberg, Rev. Mod. Phys. **61** (1989) 1.
12. V. Sahni and A. Starobinsky, Int. J. of Mod. Phys. A**9** (2000) 373.
13. P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75** (2003) 559.
14. T. Padmanabhan, Phys. Rept. **380** (2003) 235.
15. E. J. Copeland, M. Sami and S. Tsujikawa, Int. J. Mod. Phys. D **15** (2006) 1753.
16. L. Amendola and S. Tsujikawa, *Dark Energy*, Cambridge U. Press (2010) & (2015).
17. J. Solà, J. Phys. Conf. Ser. **453** (2013) 012015 [arXiv:1306.1527]; AIP Conf. Proc. **1606** (2015) 19 [arXiv:1402.7049]; J. Phys. Conf. Ser. **283** (2011) 012033 [arXiv:1102.1815].
18. A. G. Riess et al., Astrophys. J. **826** (2016) (2016) 56; Astrophys. J. **855** (2018) 136.
19. A. G. Riess et al., Astrophys.J. **876** (2019) 85; Astrophys. J. Lett. **908** (2021) L6.
20. L. T. Verde, T. Treu T. and A. G. Riess, Nature Astron. **3** (2019) 891.
21. L. Perivolaropoulos and F. Skara (2021), *Challenges for ΛCDM: An update*, [arXiv:2105.05208 [astro-ph.CO]].
22. J. Solà Peracaula, Int.J.Mod.Phys.A**33** (2018) 1844009.
23. E. Di Valentino et al., Astropart. Phys. **131** (2021) 102605, Class. Quant. Grav. **38** (2021) 153001.
24. E. Di Valentino et al., Astropart. Phys. **131** (2021) 102604.
25. G. Efstathiou, MNRAS **505** (2021) 3866; W.L. Freedman, Nat. Astron. **1** (2017) 0121.
26. E. Mortsell et al., *The Hubble Tension Revisited: Additional Local Distance Ladder Uncertainties*, arXiv:2106.09400; *The Hubble Tension Bites the Dust: Sensitivity of the Hubble Constant Determination to Cepheid Color Calibration*, arXiv:2105.11461.
27. M. Giovanna Dainotti et al., Astrophys. J. **912** (2021) 150.
28. I. L. Shapiro and J. Solà, JHEP **02** (2002) 006; Phys. Lett. B **682** (2009) 105.
29. J. Solà, J. Phys. A**41** (2008) 164066.
30. A. Babic, B. Guberina, R. Horvat and H. Stefancic, Phys. Rev. D**71** (2005) 124041.
31. M. Maggiore, Phys. Rev. D**83** (2011) 063514; M. Maggiore, L. Hollenstein, M. Jaccard and E. Mitsou, Phys. Lett. B**704** (2011) 102.
32. N. Bilic, Phys. Rev. D**83** (2011) 105003; Rom. J. Phys. **57** (2012) 793.
33. K. Kohri and H. Matsui, JCAP **06** (2017) 006; Phys. Rev. D**98** (2018) 103521.
34. A. Ferreiro and J. Navarro-Salas, Phys. Lett. B **792** (2019) 81; Phys. Rev. D **102** (2020) 045021; A. Ferreiro, S. Nadal-Gisbert, J. Navarro-Salas, Phys. Rev. D **104** (2021) 2, 025003. J. Fernando Barbero et al., Phys. Rev. D **98** (2018) 025016.

35. I. G. Marian et al., *Renormalization of Field-Independent Term in the Cosmological Constant Problem*, arXiv:2107.06069.
36. N. E. Mavromatos and J. Solà Peracaula, Eur. Phys. J. Spec. Top. **230** (2021) 2077.
37. S. Basilakos, N. E. Mavromatos and J. Solà Peracaula, Int. J. Mod. Phys. D **28** (2019) 1944002; Phys. Rev. D **101** (2020) 045001; Phys. Lett. B **803** (2020) 135342.
38. N. E. Mavromatos and J. Solà Peracaula, *Inflationary physics and transplanckian conjecture in the Stringy Running-Vacuum-Model: From the phantom vacuum to the true vacuum,* arXiv:2105.02659.
39. N. E. Mavromatos, *Geometrical origins of the Universe dark sector: String-inspired torsion and anomalies as seeds for inflation and dark matter*, arXiv:2108.02152.
40. N. E. Mavromatos, *Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology*, arXiv:2108.03998.
41. T. S. Bunch, J. Phys. A**13** (1980) 1297.
42. E.K. Akhmedov, *Vacuum energy and relativistic invariance*, hep-th/0204048.
43. J. Solà Peracaula, C. Moreno-Pulido and S. Cheraghchi (in preparation).
44. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, Astrophys. J. Lett. **811** (2015) L14; Astrophys. J. **836** (2017) 43.
45. A. Gómez-Valent, J. Solà and S. Basilakos, JCAP **01** (2015) 004; A. Gómez-Valent and J. Solà, MNRAS **448** (2015) 2810; A. Gómez-Valent, E. Karimkhani and J. Solà JCAP **12** (2015) 048.
46. J. Solà, A. Gómez-Valent and J. de Cruz Pérez, Phys. Lett. B**774** (2017) 317.
47. J. Solà Peracaula, J. de Cruz Pérez and A. Gómez-Valent, EPL **121** (2018) 39001; MNRAS **478** (2018) 4357.
48. M. Rezaei, M. Malekjani and J. Solà Peracaula, Phys. Rev. D**100** (2019) 023539.
49. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, Class. Quant. Grav. **37** (2020) 245003; Astrophys. J. Lett. **886** (2019) L6.
50. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido, EPL **134** (2021) 19001.
51. A. Gómez-alent, *Vacuum energy in Quantum Field Theory and Cosmology* (PhD. Thesis), arXiv:1710.01978.
52. J. de Cruz Pérez, *Implications of Dynamical Dark Energy in the expansion of the Universe and the Structure Formation* (PhD. Thesis), arXiv:2105.14800.
53. M. Rezaei, J. Solà Peracaula and M. Malekjani, *Cosmographic approach to Running Vacuum dark energy models: New constraints using BAOs and Hubble diagrams at higher redshifts*, arXiv:2108.06255.
54. S. Basilakos, J. A. S. Lima and J. Solà, MNRAS **431** (2013) 923; E. L. D. Perico, J. A. S. Lima, S. Basilakos and J. Solà, Phys. Rev. D **88** (2013) 063531.
55. J. Solà and A. Gómez-Valent, Int. J. Mod. Phys. **D 24** (2015) 1541003.
56. J. Solà, Int. J. Mod. Phys. D**24** (2015) 1544027.
57. J. Solà Peracaula and H. Yu, Gen. Rel. Grav. **52** (2020) 17.
58. M. Martinelli et al., MNRAS **488** (2019) 3423; V. Salvatelli et al., Phys. Rev. Lett. **113** (2014) 181301.
59. A. Gómez-Valent and J. Solà Peracaula, MNRAS **478** (2018) 126; EPL **120** (2017) 39001.
60. D. Blas, J. Lesgourgues and T. Tram, JCAP **1107** (2011) 034.
61. A. R. Liddle, MNRAS **377** (2007) L74.
62. B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, CAP **1302** (2013) 001.
63. C. Heymans et al., A & A **646** (2021) A140.
64. J. Solà Peracaula, A. Gómez-Valent, J. de Cruz Pérez and C. Moreno-Pulido (in preparation).

Dark matter properties from the Fornax globular cluster timing: Dynamical friction and cored profiles

D. Blas

Grup de Física Teòrica, Departament de Física,
Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain, and
Institut de Fisica d'Altes Energies (IFAE),
The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra , Spain
E-mail: dblas@ifae.es

I summarize our recent results to use the orbits of globular clusters (GCs) in the Fornax dwarf spheroidal (dSph) galaxy to learn more about dark matter (DM) properties. Our focus is on clarifying how dynamical friction (DF) from the DM halo is modified from the different microscopic properties of DM, which may alter *both* the scattering processes responsible of DF and the DM profiles (in particular generating a core), which also modifies DF. We consider: (*i*) fermionic degenerate dark matter (DDM), where Pauli blocking should be taken into account in the dynamical friction computation; (*ii*) self-interacting dark matter (SIDM) and (*iii*) ultralight dark matter (ULDM), for which this problem has been addressed by a variety of methods in recent literature. We derive DF with a Fokker-Planck formalism, reproducing previous results for ULDM and cold DM, while providing new results for DDM. Furthermore, ULDM, DDM and SIDM may generate cores in dSphs, which suppress dynamical friction and prolong GC orbits. We conclude that in all these cases the modifications in the DM modelling does not easily solve the so-called timing 'problem' of Fornax GCs. We finally study this 'problem' in terms of the initial conditions, demonstrating that the observed orbits of Fornax GCs are consistent with this expectation of a cuspy DM profile with a mild 'fine-tuning' at the level of $\sim 25\%$.

Keywords: Dark matter; globular clusters; dynamical friction.

1. Introduction and motivation

The presence of dark matter (DM) in the universe is probed by several different methods. In particular, the rotation curves of *tracers* of the gravitational potential in galaxies has long been considered as one of the most pressing motivations for the existence of a halo of DM extending beyond the volume occupied by visible matter. When tracers move in this *medium*, their dynamics may also be altered by direct momemtum exchange with the DM, which may lead to further dynamical consequences beyond rotation curves, as for instance tidal disruptions, dynamical heating or dynamical friction. The latter is the focus of this contribution, and in particular using it to learn about properties of DM.

Dynamical friction is generated by the relative motion between the tracer (probe) and the DM medium. Indeed, as the probe moves in the DM halo, its gravitational interaction with the latter leaves behind an overdensity of DM particles (wake) that pulls back gravitationally from it. The net result is a friction force, first calculated

by Chandrasekhar,[1] and that can be parameterized as (for a probe of mass $m_\star$)[2]

$$\frac{d\mathbf{V}}{dt} = -\frac{4\pi G^2 m_\star \rho}{V^3}\, C\, \mathbf{V}, \tag{1}$$

where $\mathbf{V}$ is the relative velocity, G is Newton's constant, ρ is the energy density of the DM medium and C is a parameter depending on the way momentum is exchanged between the probe and the DM medium. For a classical gas of particles with mass m, and with a cold distribution in velocities v_m given by $f(v_m)$, the later reads

$$C_{\rm class} = 4\pi \ln\Lambda \int_0^V dv_m v_m^2 f_v(v_m). \tag{2}$$

In this formula, $\ln\Lambda$ represents the so-called Coulomb logarithm, and is a factor of $O(1)$. We can already understand that different models of DM may alter the previous formula through modifications of the microscopic scattering (for instance if DM is not a simple collection of free classical particles, but has some coherent or quantum properties), different DM energy densities ρ or distribution functions f_v.

The question we want to address is if these features can lead to observable consequences. For this, let us consider a system where DF is supposed to have played a relevant role in current observations: the globular clusters (GCs) of the Milky Way dwarf spheroidal (dSph) satellite galaxies. The latter are believed to be DM dominated 'compact' galaxies.[3,4] One intriguing puzzle about the dSph galaxies related to DF concerns the GCs of the Fornax dSph.[5] Fornax is a very luminous nearby (kpc away) dwarf satellite, with a stellar mass or around $\sim 4\times 10^7\, M_\odot$ and of $\sim$ kpc scale. It contains six known GCs,[4,6] with masses around $m_\star \sim 10^5\, M_\odot$, which we consider as our 'probes' moving in the DM halo. The puzzle arises because 2 of the most innermost GCs should have lost enough momentum due to DF to make them fall to the center of Fornax, while they live relatively far away from this point. In other words, from the expression (1), one can naïvely estimate the typical time scale to fall to center for GCs,

$$\tau \equiv \frac{|\mathbf{V}|}{|d\mathbf{V}/dt|} \sim 1.8 \left(\frac{V}{12\ {\rm km/s}}\right)^3 \left(\frac{10^5\, M_\odot}{m_\star}\right) \frac{2\times 10^7 \frac{M_\odot}{\rm kpc^3}}{\rho}\ {\rm Gpc}. \tag{3}$$

When applied to the six GCs of Fornax assuming the usual CDM cusp density profile (see, e.g.[7]), this time scale for the six GCs is presented in Table 2[8] (see also[9]). On the other hand, the stellar content of the GCs (and most of the stellar content of Fornax) is old, with life estimates > 10 Gyr[10,11].[12,13] Hence these GCs had enough time to fall to the center since their 'birth'.

This puzzle (in particular the time scales for GC3 and GC4) has been highlighted as a possible tension of the standard DM paradigm that may be solved by changing the properties of DM, see e.g.[9] A first important observation is that the Jeans analysis based on kinematic information of Fornax (the velocity dispersion along the line of sight of ~ 2500 stars as a function of radius) is compatible with the energy density profiles M19 NFW and M19 ISO expected from standard CDM,[7] but also

Table 1. Some details of Fornax GCs: mass, projected radius and CDM instantaneous DF time (Eq. (3)). See Ref.[8] for more details about how these number are obtained and the origin of the discrepancies with previous literature.

	$m_\star\ [10^5 M_\odot]$	$r_\perp$[kpc]	$\tau_{\rm CDM}$[Gyr]
GC1	0.42 ± 0.10	1.73 ± 0.05	119
GC2	1.54 ± 0.28	0.98 ± 0.03	14.7
GC3	4.98 ± 0.84	0.64 ± 0.02	2.63
GC4	0.76 ± 0.15	0.154 ± 0.014	0.91
GC5	1.86 ± 0.24	1.68 ± 0.05	32.2
GC6	~ 0.29	0.254 ± 0.015	5.45

with the presence of a cored distribution arising in different interesting DM models. The latter include fermionic DM of masses where quantum degeneracy may be relevant (DDM) in the dynamics of Fornax,[14–19] ultra-light DM with de Broglie wavelength large enough to affect the dynamics of Fornax[9] or self-interacting DM (SIDM) within the parameter space allowed by other observables, e.g.[20,21] A key motivation for our work is exploring whether DF may distinguish among these options, and hint towards new properties of DM.

2. Derivation of DF for different DM models with a Fokker-Planck approach

In order to include non-trivial properties of the DM microphysics into the expression (1), we followed a Fokker-Planck (FP) approach in,[8] where the probe particle (a GC in our case) correspond to species 1 traveling through a gas of spectator particles (species 2) with a certain distribution f_2 (DM particles in our case). We consider the following elastic scattering process of two particle species,

$$1(p) + 2(k) \to 1(p') + 2(k').$$

The phase space distribution function for the particle species 1 (the GC) evolves according to the Boltzmann equation,

$$\frac{df_1}{dt} = C[f_1]. \tag{4}$$

The collision integral $C[f_1]$ contains information about the elastic scattering processes, and is written as

$$C[f_1] = \frac{(2\pi)^4}{2E_p} \int d\Pi_k d\Pi_{p'} d\Pi_{k'}\, \delta^{(4)}(p + k - p' - k') |\overline{\mathcal{M}}|^2 \times \Big[f_1(p') f_2(k') (1 \pm f_1(p))(1 \pm f_2(k)) - f_1(p) f_2(k) (1 \pm f_1(p'))(1 \pm f_2(k')) \Big], \tag{5}$$

where $|\overline{\mathcal{M}}|^2$ is a squared matrix element averaged over initial and final spins, and $d\Pi_k = \frac{g}{2E_k} \frac{d^3k}{(2\pi)^3}$ is the Lorentz invariant phase element with the number of internal

degrees of freedom g. The sign in $1 \pm f_i$ refers to bosons (+) or fermions (−), respectively. The Boltzmann equation can be greatly simplified if the momentum exchange

$$q = p' - p, \tag{6}$$

is smaller than the typical momentum in the distribution function f_1. This is the adequate limit to compute DF.[2] In such cases, the Boltzmann equation reduces to the nonlinear Fokker-Planck equation,

$$\frac{df_1}{dt} = -\frac{\partial}{\partial p^i}\left[f_1(1 \pm f_1)D_i\right] + \frac{1}{2}\frac{\partial}{\partial p^i}\left[\frac{\partial}{\partial p^j}(D_{ij}f_1) \pm f_1^2\frac{\partial}{\partial p^j}D_{ij}\right], \tag{7}$$

where the diffusion coefficients are defined in.[2,8] Furthermore, the gravitational scattering of a probe particle of mass $m_\star$ and a particle in the medium with mass m is described by the spin-averaged matrix element

$$|\overline{\mathcal{M}}|^2 = \frac{1}{2s+1}\frac{(16\pi G)^2 m^4 m_\star^4}{\left[(q^0)^2 - \mathbf{q}^2\right]^2}, \tag{8}$$

where s is the spin of the particle in the medium, and $(q^0, \mathbf{q})$ is the transferred 4−momentum. In the nonrelativistic limit, we can neglect q^0 and maintain only $\mathbf{q}$ in Eq. (8).

Of particular importance for our analysis is the diffusion coefficient $D_{||}$, corresponding to the diffusion in momentum parallel to the probe object's instantaneous velocity. Indeed, the DF deceleration on a probe of mass $m_\star$ moving with velocity $\mathbf{V}$ w.r.t. the medium is[2]

$$\frac{d\mathbf{V}}{dt} = \frac{D_{||}}{m_\star}\hat{\mathbf{V}}. \tag{9}$$

From (1), we see that the dimensionless coefficient C of the DF reads,

$$C = -\frac{V^2 D_{||}}{4\pi G^2 m_\star^2 \rho}. \tag{10}$$

When computed for a gas of classical particles (where f_2 corresponds to a classical distribution), one reproduces the known results of the Chandrasekhar[1] formula, with C given by (2). This limit is also the relevant one for the SIDM case of interest here, since the corresponding cross-sections always correspond to large mean-free paths, where the approximation of the Chandrasekhar calculation holds.

Regarding DDM, when one assumes that f_2 is given by a Fermi-Dirac distribution close to the degeneracy limit[a] one finds

$$C_{\rm DDM} \to \ln\Lambda \begin{cases} 1 & V \gg v_F \\ \frac{V^3}{v_F^3} & V \ll v_F \end{cases}, \tag{11}$$

[a]Note that it is not known what is the momentum distribution for degenerate fermions interacting only gravitationally. Still, one normally assumes an equilibrium configuration as a first approximation. This assumption may also be used to constrain the mass of fermionic DM from virialized DM objects.[19,22]

where v_F is the Fermi velocity, related to the DM energy density by

$$\rho = \frac{g m_{\rm DM}^4 v_F^3}{6\pi^2}, \tag{12}$$

where g represents the number of degrees of freedom of the species (see also[23]). We used $m_{\rm DM}$ for the DM mass. This modification of the DF formula changes the falling time τ, though not parametrically. Furthermore, the DDM case also generates a *core* due to the quantum degeneracy pressure. Both effects are useful to reduce DF and hence prolong τ.

A similar calculation can be done for the bosonic case with large occupation numbers. In this case, the collision term is also modified and allows our formalism to capture the most relevant phenomena for DF in the ULDM case described in.[9,24–26] Indeed, these Bose-enhancement terms in (7) generate large-scale density fluctuations, causing additional velocity drift that can be characterised by an extra term to $C \to C + \Delta C$ as

$$\Delta C = \ln\Lambda \left(\frac{m_{\rm eff}}{m_\star}\right)\left({\rm erf}(X_{\rm eff}) - \frac{2X_{\rm eff}}{\sqrt{\pi}} e^{-X_{\rm eff}^2}\right), \tag{13}$$

where $m_{\rm eff} = \pi^{3/2}\rho/(m_{\rm DM}\sigma)^3$ is the ULDM mass enclosed in an effective de Broglie volume and $X_{\rm eff} \equiv v/\sqrt{2}\sigma_{\rm eff}$ with $\sigma_{\rm eff} = \sigma/\sqrt{2}$. Numerically, $m_{\rm eff} \approx 1.2\times10^6 \left(10^{-21}\ {\rm eV}/m_{\rm DM}\right)^3 [\rho/(3\times10^7\ M_\odot/{\rm kpc}^3)][(10\ {\rm km/s})/\sigma]^3\ {\rm M}_\odot$. With these numbers and keeping in mind a typical GC mass $m_\star \sim 10^5\ {\rm M}_\odot$, the ΔC effect becomes quantitatively important in Fornax for $m \lesssim 3\times10^{-20}$ eV. This value starts to be in tension with other constraints of ULDM, see e.g.[27,28] Notice that the previous calculation is correct in the limit $r > \lambda_{\rm db}$, where $\lambda_{\rm db}$ is the de Broglie wavelength of DM. In the opposite regime, one can have effects coming from the coherent nature of the DM waves which affect the scattering cross-section,[9] or even resonant processes that our formalism does not capture, see e.g.[26,29]

3. Timing of Fornax GCs for different DM models

From the previous calculation, one can estimate the falling time of the GCs in Fornax once the profile of DM energy density is known. For the latter, one can use a Jeans method consistent with the line-of-sight velocity (LOSV) observed for Fornax.[2] However, as already stressed, the latter is equally well fitted by a M19 NFW, M19 ISO CDM profiles, as long as DDM of mass $m_{\rm DM} \approx 135$ eV (which should generate a degenerate core of kpc) or a SIDM model with velocity averaged cross section $\langle\sigma v\rangle/m_{\rm DM} \sim 25\ {\rm cm}^2\ {\rm g}^{-1}\ {\rm km\ s}^{-1}$ (also generating a isothermal cored profile of kpc size, this time due the SIDM scatterings[20]). These cases are shown, together with the relevant data of the dispersion in the LOSV, σ_{LOS}, as a function of radius in the left panel of Fig. 1. It should be clear from this panel that this data agrees well with the models just described. The right panel shows the different energy density profiles for these cases also as a function of radius. For completeness we have also shown the energy density of the star content of Fornax, to show explicitly that

DM is needed in this system. Another CDM model (coreNFW) is also shown. This models aims at describing a DM dominated halo where the baryonic feedback is somehow taken into account.

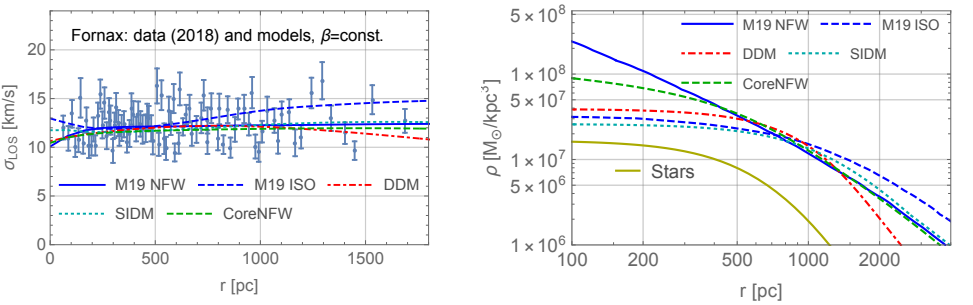


Fig. 1. **Left panel**: LOSVD data and fits for the M19 NFW and M19 ISO of,[7] DDM with mass $m \approx 135\,\mathrm{eV}$ and SIDM models with cores of order kpc. **Right panel**: Density profiles corresponding to these cases. The the energy density of the star content of Fornax is also shown.

Now that we have all the ingredients to compute the DF, we proceed to find the orbital motion of the different GCs of Fornax by solving the equation

$$\frac{d\mathbf{V}}{dt} = -\frac{GM(r)}{r^2}\hat{r} - \frac{4\pi G^2 m_\star \rho}{V^3}\, C\, \mathbf{V}, \tag{14}$$

for the relevant cases. This is a 'semi-analytical' approach that yields good agreement with simulations.[8] The results are only mildly dependent on the Coulomb logarithm $\ln\Lambda$, which we calibrate to numerical work in.[8] As an example, we show in Fig. 2 the evolution of the radius as a function of time, as compared to two profiles studied numerically in.[7]

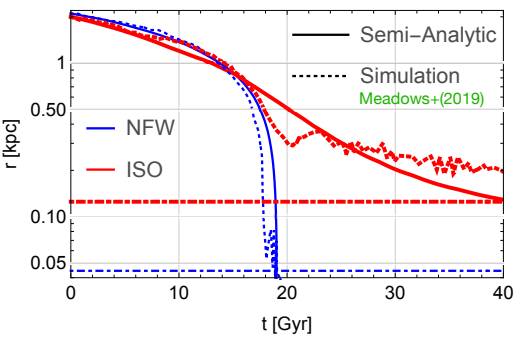


Fig. 2. Radius of an infalling orbit of a GC with mass $m_\star = 3 \times 10^5\, M_\odot$. n dotted blue (thick dotted red) we plot the simulation result in[7] for the NFW (ISO) halo. In solid lines, we plot our semi- analytic integration. Horizontal dot-dashed lines show the radii where the semi-analytic treatment breaks down.

From the previous analysis, we are now in the position to find the time τ that it takes to a GC to fall to the center of Fornax for different DM models. We will address the initial conditions momentarily. For now, let us compare the 'instantaneous' time in the DF formulae for the different GCs in CDM, DDM of mass $m_{\rm DM} \approx 135\,{\rm eV}$ and SIDM generating a kpc core. This is shown in Table 2 and Fig. 3. The conclusion is clear: both models beyond CDM prolong the plunging time of the problematic GCs, and hence alleviate the timing problem.

Table 2. Instantaneous DF times for the GCs of Fornax in the DDM and SIDM models described in the main text.

	$\tau_{\rm CDM}$[Gyr]	$\tau^{(135)}_{\rm DDM}$ [Gyr]	$\tau_{\rm SIDM}$ [Gyr]
GC1	119	122	79.3
GC2	14.7	7.12	8.82
GC3	2.63	1.48	2.21
GC4	0.91	10.7	14.8
GC5	32.2	30.1	20
GC6	5.45	16.1	22

This conclusion is confirmed by a numerical study of the orbits, where the different DM profiles and dispersion properties are taking into account for the different models. The latter analysis is vital to find the real dynamics of the GCs, since the timescales can vary by $O(1)$ factors. We will skip the details of this study in this short contribution, and the interested reader is invited to check our work[8] for details. Let us simply mention that beyond the numerical integration of Eq. (14), in[8] we also considered different projection effects that alter the radius of the different GCs and their velocities. The main lesson we learned is that the conclusions extracted from the estimates in Table 2 and Fig. 3 are robust against these uncertainties.

The main lesson from Fig. 3 is that cores (including those in coreNFW) increase the time τ, but to make GC4's settling time substantially different, one requires a

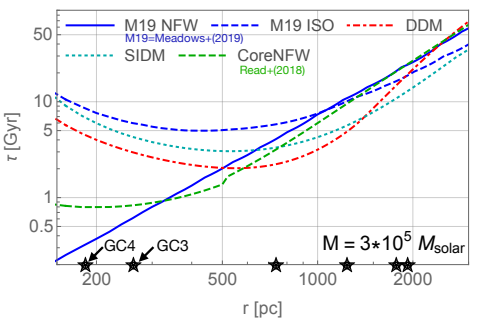


Fig. 3. Instantaneous DF time, evaluated for different DM models for $m_\star = 3 \times 10^5 M_\odot$ as a function of distance to the center of Fornax. The stars represent the positions of the different GCs.

large (kpc-size) core. Note that the microscopic properties in the DF formula are not particularly relevant in our calculations.

Before closing this section, let us notice that in Table 2 and Fig. 3 we have focused on DM models which help prolonging the orbits of GCs in Fornax. However, for the DDM case this requires models with masses already in strong tension with Ly-α observations[30] (and light tension with Jeans analysis of other systems[19]). In fact, regarding the Ly-α constraints, the interesting reader may also find in our work[8] a rather model independent bound from these observations

$$g m_{\rm DM}^4 > 2 \times (1.4 \text{ keV})^4 \,, \tag{15}$$

confirmed in the recent detailed work.[31] On the contrary, the SIDM model we used for the existence of a core at kpc distances can be read from the relation

$$r_c \sim \frac{m_{\rm DM}}{\rho \sigma} = 48 \frac{10^8 M_\odot/\text{kpc}^3}{\rho} \frac{1 \text{ cm}^2/\text{gr}}{\sigma/m_{\rm DM}} \text{kpc} \,, \tag{16}$$

related to the isothermal profile generated by scattering of DM particles with cross-section σ. The ballpark used in the previous formulae corresponds to viable models of DM,[20,21] which puts our analysis on a solid phenomenological basis. Indeed, one is tempted to conclude that the Fornax GCs favour the SIDM model.

4. Brief discussion on the late-time distribution of GCs

Our discussion has so far focused on the falling time of the innermost GCs of Fornax. The 'puzzle' we are trying to address relies on this time being smaller than what one would naively expect for arbitrary GCs that have lived in the dSph DM halo for long enough. However, given a collection of GCs, one expects some of them to be form currently at distances that could seem fine tuned for the average GC. Hence, the best-posed question to learn about the DM effects in the dynamics of GCs in Fornax (given that we observe 6 of them) is which is the long-term distribution function of these objects as a function of radius and for different initial conditions. Note first, that the problem of initial conditions has already been identified as a candidate to explain the phenomenology we are discussing.[7,32] Here we will summaryze the *analytical* treatment developed in.[8] For this, the key analytical tool is the time it takes an object to move from radius r_0 to r_f. In the case of nearly-circular orbits, this time can be estimated as

$$\Delta t \left(r_i; r_f\right) = \int_{r_f}^{r_0} \frac{dr}{2r} \left(1 + \frac{d \ln M}{d \ln r}\right) \tau \left(r, v_{\rm circ} \left(r\right)\right) . \tag{17}$$

Quite remarkably, one can show that for objects that today at at small radii $r \ll r_{\rm crit}$, with $r_{\rm crit}$ parametrizing the radius such that any object that *started* its life at $r < r_{\rm crit}$ as already fallen into the center, the conditional distribution function (CDF) of GCs as a function of radius in a NFW profile satisfies

$$F_{\Delta t}(r) \approx A \frac{\tau(r)}{\Delta t}, \tag{18}$$

independently on initial conditions. This prediction is modified in the case of cored profiles. To have a better picture of this effect, we show in Fig. 4 the number of GCs expected to be enclosed at projected radius $r_\perp$, when numerically integrating the orbital motion with our semianalytical model (no need of N-body simulations). The initial conditions we used for these plots are discussed in.[8] They are not very particularly relevant for this discussion since we found that the main features of Fig. 4 are reproduced by all reasonable initial conditions we considered.

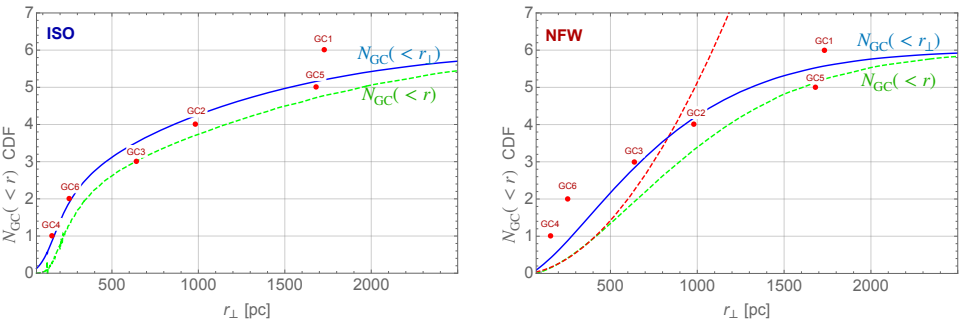


Fig. 4. CDF (number of clusters at $r_\perp$ below the value of the x axes) of projected radii using $\tau(r)$ from Fig. 3. **Left:** Cored ISO halo. **Right:** NFW halo. The solid blue line shows the CDF after projection effects are taken into account. The dashed green line shows the result before projection. For the NFW case, the small-r prediction of Eq. (18) is shown by the red dashed line. Observed Fornax GCs are also shown. The initial radial GC PDFs used to make the plot are explained in.[8]

A visual comparison of the left panel in Fig. 4 (performed with an isothermal cored for the model of SIDM described above) and the right panel (performed with NFW) shows that GC4, GC6 and GC3 tend to align better over the CDF of the second one, in particular after projection effects are taken into account (blue-solid line). As we just mentioned, these figures were produced for one of many possible initial conditions that reproduce reasonably well the distribution of final GCs, and we could not find any other reasonable initial data which could account for the position of GC4 today without destroying the agreement of the rest of GCs. Still, these figures also show that the level of 'fine-tuning' to explain GC4, GC6 and GC3 within the standard NFW profile is not very problematic, which implies that a solid conclusion about the role of DF in the final distribution of GCs requires the observation of (several) other similar systems. Another interesting observation is that any reasonable distribution agreeing with current data should have included several inner GCs which should have fallen to the center of Fornax and generated a nuclear cluster of $\mathcal{O}(10^6\, M_\odot)$. The absence of this nuclear cluster in Fornax may also have something to say about DM properties, though this is a different story.

We are fully aware that the dynamics of GCs is more complex than what we presented in.[8] Still, we believe that our simple analytical model captures some

interesting (even intriguing) part of the dynamics, that we hope to study in the future with more realistic methods.

5. Conclusions and outlook

In this presentation, we revisited the timing 'problem' of the GCs in the Fornax dSph, and use it to learn about new properties of DM. The latter are manifested in either a modification of the microscopic origin of dynamical friction (DF) or the halo morphologies, also influencing DF. Some of the 6 GCs of Fornax are placed too close to the center of this galaxy, which may pose a tuning problem in terms of typical time-scales to plunge to the center of this object. Indeed, the last part of the work described in[8] (and briefly reviewed in Sec. 4) was devoted to quantifying this degree of tuning for the standard CDM paradigm yielding NFW profiles. The observations presented in Table 2 can be accounted for as a moderate fluctuation with a Poisson probability of about around 25%.

Still, it is quite interesting that the calculation of DF in different DM models allows us to reduce this tension and find better accuracy with data. This is why a large part of this short presentation has consisted in a succinct explanation of how to derived DF from a Fokker-Planck formalism (which allows the introduction of fermionic and baryonic effects in the collision term). As summarized in Fig. 3, the presence of a small core due to baryonic feedback may slightly alleviate this tension, but not at an interesting level. Once one considers other DM models (as DDM, ULDM and SIDM) a large core (of kpc size) may be formed, which predicts enough reduction of dynamical friction to better reproduce the observed positions of the GCs in Fornax. Still, in these cases, the GC distribution depends strongly on initial conditions. From these models, it seems that SIDM with $\sigma/m_{\rm DM} \approx 1$ cm^2/gr seems favoured, since it does not contradict any other observation, but more data is required to confirm this claim.

The way forward is clear: applying our methods to more extensive data, aiming at enough statistical significance to generate robust conclusions about DM properties. The hint about SIDM we just discussed makes this project particularly exciting, since future data from GCs may finally start closing up on the fundamental nature of DM.

Acknowledgments

It is a please to thank N. Bar, K. Blum and H. Kim for all the hard work they devoted to this work. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. The research leading to these results has received funding from the Spanish Ministry of Science and Innovation (PID2020-115845GB-I00/AEI/10.13039/501100011033).

References

1. S. Chandrasekhar, Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction., *Astrophys. J.* **97**, p. 255 (Mar 1943).
2. J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition* 2008.
3. M. G. Walker, M. Mateo, E. W. Olszewski, J. Penarrubia, N. W. Evans and G. Gilmore, A universal mass profile for dwarf spheroidal galaxies?, *The Astrophysical Journal* **704**, p. 1274 (2009).
4. D. R. Cole, W. Dehnen, J. I. Read and M. I. Wilkinson, The mass distribution of the Fornax dSph: constraints from its globular cluster distribution, *Mon. Not. Roy. Astron. Soc.* **426**, p. 601 (2012).
5. S. Tremaine, The formation of the nuclei of galaxies. ii-the local group, *The Astrophysical Journal* **203**, 345 (1976).
6. M.-Y. Wang, S. Koposov, A. Drlica-Wagner, A. Pieres, T. Li, T. de Boer, K. Bechtol, V. Belokurov, A. Pace, D. Bacon *et al.*, Rediscovery of the sixth star cluster in the fornax dwarf spheroidal galaxy, *The Astrophysical Journal Letters* **875**, p. L13 (2019).
7. N. Meadows, J. F. Navarro, I. Santos-Santos, A. Benítez-Llambay and C. Frenk, Cusp or core? Revisiting the globular cluster timing problem in Fornax, *Mon. Not. Roy. Astron. Soc.* **491**, 3336 (Jan 2020).
8. N. Bar, D. Blas, K. Blum and H. Kim, Assessing the Fornax globular cluster timing problem in different models of dark matter, *Phys. Rev. D* **104**, p. 043021 (2021).
9. L. Hui, J. P. Ostriker, S. Tremaine and E. Witten, Ultralight scalars as cosmological dark matter, *Phys. Rev.* **D95**, p. 043541 (2017).
10. T. de Boer and M. Fraser, Four and one more: The formation history and total mass of globular clusters in the fornax dsph, *Astronomy & Astrophysics* **590**, p. A35 (2016).
11. A. D. Mackey and G. F. Gilmore, Surface brightness profiles and structural parameters for globular clusters in the Fornax and Sagittarius dwarf spheroidal galaxies, *Mon. Not. Roy. Astron. Soc.* **340**, p. 175 (2003).
12. A. del Pino, S. L. Hidalgo, A. Aparicio, C. Gallart, R. Carrera, M. Monelli, R. Buonanno and G. Marconi, Spatial dependence of the star formation history in the central regions of the fornax dwarf spheroidal galaxy, *Monthly Notices of the Royal Astronomical Society* **433**, 1505 (2013).
13. M.-Y. Wang, T. de Boer, A. Pieres, T. Li, A. Drlica-Wagner, S. Koposov, A. Vivas, A. Pace, B. Santiago, A. Walker *et al.*, The morphology and structure of stellar populations in the fornax dwarf spheroidal galaxy from dark energy survey data, *The Astrophysical Journal* **881**, p. 118 (2019).
14. V. Domcke and A. Urbano, Dwarf spheroidal galaxies as degenerate gas of free fermions, *JCAP* **1501**, p. 002 (2015).
15. L. Randall, J. Scholtz and J. Unwin, Cores in dwarf galaxies from fermi repulsion, *Monthly Notices of the Royal Astronomical Society* , p. stx161 (Jan 2017).
16. C. Di Paolo, F. Nesti and F. L. Villante, Phase space mass bound for fermionic dark matter from dwarf spheroidal galaxies, *Mon. Not. Roy. Astron. Soc.* **475**, 5385 (2018).
17. D. Savchenko and A. Rudakovskyi, New mass bound on fermionic dark matter from a combined analysis of classical dSphs, *Mon. Not. Roy. Astron. Soc.* **487**, 5711 (2019).
18. A. Boyarsky, O. Ruchayskiy and D. Iakubovskyi, A Lower bound on the mass of Dark Matter particles, *JCAP* **0903**, p. 005 (2009).
19. J. Alvey, N. Sabti, V. Tiki, D. Blas, K. Bondarenko, A. Boyarsky, M. Escudero, M. Fairbairn, M. Orkney and J. I. Read, New Constraints on the Mass of Fermionic Dark Matter from Dwarf Spheroidal Galaxies (10 2020).

20. M. Kaplinghat, S. Tulin and H.-B. Yu, Dark Matter Halos as Particle Colliders: Unified Solution to Small-Scale Structure Puzzles from Dwarfs to Clusters, *Phys. Rev. Lett.* **116**, p. 041302 (2016).
21. S. Tulin and H.-B. Yu, Dark Matter Self-interactions and Small Scale Structure, *Phys. Rept.* **730**, 1 (2018).
22. S. Tremaine and J. E. Gunn, Dynamical Role of Light Neutral Leptons in Cosmology, *Phys. Rev. Lett.* **42**, 407 (1979), [,66(1979)].
23. P.-H. Chavanis, Landau equation for self-gravitating classical and quantum particles: Application to dark matter, *arXiv e-prints* , p. arXiv:2012.12858 (December 2020).
24. L. Lancaster, C. Giovanetti, P. Mocz, Y. Kahn, M. Lisanti and D. N. Spergel, Dynamical Friction in a Fuzzy Dark Matter Universe, *JCAP* **2001**, p. 001 (2020).
25. B. Bar-Or, J.-B. Fouvry and S. Tremaine, Relaxation in a Fuzzy Dark Matter Halo, *Astrophys. J.* **871**, p. 28 (2019).
26. B. Bar-Or, J.-B. Fouvry and S. Tremaine, Relaxation in a Fuzzy Dark Matter Halo. II. Self-consistent kinetic equations (10 2020).
27. N. Bar, D. Blas, K. Blum and S. Sibiryakov, Galactic rotation curves versus ultralight dark matter: Implications of the soliton-host halo relation, *Phys. Rev.* **D98**, p. 083027 (2018).
28. D. J. Marsh and J. C. Niemeyer, Strong Constraints on Fuzzy Dark Matter from Ultrafaint Dwarf Galaxy Eridanus II, *Phys. Rev. Lett.* **123**, p. 051103 (2019).
29. Y. Wang and R. Easther, Dynamical Friction From Ultralight Dark Matter (10 2021).
30. J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville and M. Viel, Lyman-alpha Forests cool Warm Dark Matter, *JCAP* **08**, p. 012 (2016).
31. M. Carena, N. M. Coyle, Y.-Y. Li, S. D. McDermott and Y. Tsai, Cosmologically Degenerate Fermions (8 2021).
32. S. Shao, M. Cautun, C. S. Frenk, M. Reina-Campos, A. J. Deason, R. A. Crain, J. D. Kruijssen and J. Pfeffer, The survival of globular clusters in a cuspy Fornax (12 2020).

Growth of linear perturbations in a universe with superfluid dark matter

S. Banerjee

Inst. for Quantum Gravity, FAU Erlangen-Nuremberg, Staudtstr. 7, 91058 Erlangen, Germany
E-mail: shreya.banerjee@fau.de

S. Bera

Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India
E-mail: sayantani@iucaa.in

D. F. Mota

Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway
E-mail: d.f.mota@astro.uio.no

The Lambda-Cold Dark Matter (ΛCDM) model agrees with most of the cosmological observations, but has some hindrances from observed data at smaller scales such as galaxies. Recently, Berezhiani and Khoury proposed a new theory involving interacting superfluid dark matter with three model parameters in,[1] which explains galactic dynamics with great accuracy. In the present work, we study the cosmological behaviour of this model in the linear regime of cosmological perturbations. In particular, we compute both analytically and numerically the matter linear growth factor and obtain new bounds for the model parameters which are significantly stronger than previously found. These new constraints come from the fact that structures within the superfluid dark matter framework grow quicker than in ΛCDM, and quite rapidly when the DM-baryon interactions are strong.

Keywords: Dark matter, cosmological perturbations

1. Introduction

With the advent of precision cosmology and satellites like Planck and WMAP, we have gained new insights about the evolution of the universe. Till date, Lambda-Cold Dark Matter (ΛCDM) provides the best fit to these available data and has been widely accepted as the standard model of cosmology.[2] The hypothesis of CDM, which are assumed to be collisionless non-relativistic particles, along with baryonic matter explains the CMB temperature anisotropy, matter power spectra, large scale galaxy distributions and lensing data remarkably well. In fact, the abundance of galaxy clusters and observed large scale structure formation history strongly supports the collisionless CDM scenario as opposed to any alternative theories to ΛCDM.[3–5] However, at smaller scales, CDM faces a number of challenges that need to be addressed.[6] For example, the Baryonic Tully-Fisher relation and the

corresponding tight correlation between the mass and dispersion velocity at the high-mass end can not satisfactorily be explained by CDM halo which predicts a larger scatter due to feedback processes in the galaxy.[7] Apart from this, there is another issue with the standard CDM picture in the galactic scale, known as the *cusp-core problem*.[8] The simulations of galactic halos with CDM produce a kink (cusp) at the center of the galaxy, whereas observations of various galactic density profile suggest a flat core. With improved observations of the faint dwarf galaxies and substructures within the galaxies like Milky Way and Andromeda, new set of discrepancies arise. While the missing satellite problem in dwarf galaxies[9] has been addressed to some extent, the *Too Big To Fail Problem*, arising from the prediction of satellites that are too massive and too dense by ΛCDM, compared to those observed, still remains unresolved.[10,11]

Due to the above unresolved issues, scientists have looked into other alternative explanations through modifications of General Relativity (GR). Several models have been proposed so far with the aim to explain existing data to the same degree of accuracy as ΛCDM as well as overcome its drawbacks. Many of them have already been ruled out or are highly constrained by the ongoing observations of gravitational waves, but some theories like $f(R)$, $f(T)$, $f(\mathcal{G})$, Scalar-tensor-vector theories of gravity etc. are still consistent with the data, and new observations are required to falsify these theories.[12–18] These theories are relativistic corrections of GR which modify the dynamics of spacetime through the modified field equations. The theory of Modified Newtonian Dynamics (MOND), on the other hand, is a modification to the Newtonian force law that changes the dynamics of interaction between two massive bodies in the non-relativistic limit.[19,20] MOND was first proposed in 1983 by Milgrom to account for the flattened galaxy rotation curves near the edge of the spiral galaxies like Milky Way. There is a universal acceleration scale a_0 in MOND, whose value is obtained as $10^{-8} cm/s^2$. For accelerations much lower than this scale, the Newtonian law is modified, and this explains the flat galaxy rotation curve data for a large number of galaxies.[21] Interestingly, the Baryonic Tully-Fisher relation in galaxies can exactly be derived from MOND where $M \propto v_c^4$. MOND can also explain several other galactic observations like the planar structure of galaxies, low merger rate etc.[22] Thus, we see that MOND, with just one free parameter, is a very well-behaved theory at the galactic scale. However, despite these successes, MOND faces several challenges in extragalactic and cosmological scales. Proper relativistic extension of MOND is not available.[23] Hence it cannot be applied at cosmological scales.

The effectiveness of MOND at small scales and success of ΛCDM at cosmological scales are the main motivations for scientists to look for models which are CDM-MOND hybrids i.e., theories that include usual cold dark matter as collisionless particles at cosmological scales, but give rise to a MOND-like modified force law at galactic scales such that they satisfy both sets of observations. This class of models take into consideration the interacting dark matter-baryon picture where

a MOND-like force is mediated through this new interaction term. Based on this idea, many models have been proposed which can reproduce both CDM features as well as MOND in their respective regime of validity.[1,24–27]

In this paper we shall focus on one such model proposed recently by Berezhiani and Khoury,[1] where CDM can form condensates at galactic scales depending upon the surrounding temperature and can behave as superfluid. It has already been shown by the authors that such model can explain a number of galactic scale observations due to their MONDian behaviour, which normal CDM fails to explain.[1,28–30] Although there are recent studies which suggest that the superfluid DM model is disfavoured compared to a spherically symmetric CDM halo at galactic scales as the superfluid DM overestimates the vertical acceleration which has to be counterbalanced to match the observations.[31,32] There are two free parameters in the theory which are assumed to be temperature dependent. It has been argued that at cosmological scales, the theory behaves as usual CDM and thus the background evolution and other cosmic histories remain unchanged as compared to ΛCDM. Here, we study the cosmological evolution of the background as well as the matter perturbations. We check whether the present model remains well-behaved at cosmological scales as has been claimed by the authors and compare our results with ΛCDM.

2. Dark Matter Superfluid-Overview

The central idea of this model is that CDM is made up of particles which undergo phase transition below a particular critical temperature and becomes a superfluid. This requires that the particle CDM needs to be strongly interacting below a particular temperature. The superfluid behaviour depends on the strength of interaction and the mass of the particle. It has been shown in[1] that in order to form a Bose-Einstein Condensate (BEC) the following condition must be satisfied

$$m \lesssim \left(\frac{\rho}{v^3}\right)^{1/4} \tag{1}$$

where m and v corresponds to the mass and velocity of the particle respectively and ρ is the density of the condensate. Assuming virialization of dark matter halo at galactic scales, this gives an upper bound on the mass of the particle forming the halo

$$m \lesssim 2.3(1+z_{vir})^{3/8}\left(\frac{M}{10^{12}h^{-1}M_{\odot}}\right)^{-1/4}\text{eV} \tag{2}$$

Further assuming thermalization of CDM particles, one obtains the bound on interaction cross section as

$$\frac{\sigma}{m} \gtrsim 52(1+z_{vir})^{-7/2}\left(\frac{m}{eV}\right)^{4}\left(\frac{M}{10^{12}h^{-1}M_{\odot}}\right)^{2/3}\text{cm}^2\text{g}^{-1} \tag{3}$$

Using equipartition law, the critical temperature T_c of the CDM condensate can be obtained as

$$T_c = 6.5 \left(\frac{\text{eV}}{\text{m}}\right)^{5/3} (1 + z_{vir})^2 \text{mK} \tag{4}$$

It has been argued in[1] that the temperature of CDM at cosmological scales is much below the critical temperature ($\mathcal{O}(10^{-28})$ for $m \sim$ eV) which implies that the condensate behaves as a $T \approx 0$ superfluid at cosmological scales.

The description of superfluid dark matter is given in terms of a low energy effective theory with the Lagrangian of the form:

$$\mathcal{L} = \frac{2\Lambda(2m)^{3/2}}{3} \left(\dot{\theta} - m\Phi - \frac{(\nabla\theta)^2}{2m}\right)^{3/2} \tag{5}$$

Let us now understand the motivation of choosing such a Lagrangian. Here, θ is the phase of the wavefunction describing the superfluid phonon modes and Φ is the gravitational potential in which the DM particle sits and is given by the standard Newtonian potential in the usual non-relativistic case. This Lagrangian has a free parameter Λ which defines the strength of the superfluid (i.e. defined by the number of particles in the condensate state). The power of the Lagrangian is defined by the choice of the equation of state (EoS), and a fractional power of $5/2$ is indeed obtained in superfluids formed by ultra cold atoms. In the case of CDM superfluid, the choice of the power $3/2$ in the Lagrangian is somewhat arbitrary, but motivated by the fact that the superfluid DM should give rise to MOND-like dynamics at galactic scales when baryons are also included. This also corresponds to an equation of state $P \sim \rho^3$ which is suggestive of a dominant three-body interaction process. What kind of particles can lead to such a superfluid with this particular EoS and the physics of its formation has not been discussed earlier and is beyond the scope of this paper. For our purpose, we shall assume the form of this Lagrangian to study the characteristic features of the resultant superfluid DM model.

In the effective field theory formalism, the superfluid is described in terms of interacting phonon modes. The phonon modes can be described by the scalar field θ, which, at a constant chemical potential μ, can be expanded as,

$$\theta = \mu t + \phi$$

where ϕ denotes the excitation of the phonon modes.

The DM superfluid couples to the baryons through the phonon modes via an interaction given by the Lagrangian:

$$\mathcal{L}_{int} = -\alpha \frac{\Lambda}{M_{Pl}} \theta \rho_b \tag{6}$$

This kind of interaction ensures a MOND force. Here α is a dimensionless free parameter, which sets the interaction strength of the interaction, and ρ_b is the baryonic mass density.

Thus, the complete Lagrangian for an interacting superfluid DM is given by,

$$\mathcal{L} = \frac{2\Lambda(2m)^{3/2}}{3}\left(\dot{\theta} - m\Phi - \frac{(\nabla\theta)^2}{2m}\right)^{3/2} - \alpha\frac{\Lambda}{M_{Pl}}\theta\rho_b \tag{7}$$

It has been shown in[1] that the MONDian acceleration arises as a special case of the dynamics of the above Lagrangian. The validity of this model in solar system and Bullet cluster has also been discussed there.

In the cosmological context, although the authors in[1] discuss some general points regarding the background behaviour and the equation of state of this new superfluid dark matter, they do not shed much light on other important points such as growth of perturbations and structure formation. In the next sections, we solely focus on the cosmological aspects of this new theory.

3. Cosmological Solutions

In this section, we will study this theory in cosmological context. This is of particular interest since the theory also needs to be consistent with the present cosmological data.

3.1. *Background Solutions*

For the background cosmology, we have $\theta = \theta(t)$. In the FLRW background with a scale factor a, the equation of motion for θ can be derived from the action as,

$$\frac{d}{dt}\left[(2m)^{3/2}a^3\dot{\theta}^{1/2}\right] = -\frac{\alpha}{M_{Pl}}a^3\rho_b \tag{8}$$

Assuming the evolution of baryons i.e. $\rho_b \propto 1/a^3$ as in standard ΛCDM, we get,

$$\rho_m = -\frac{\alpha\Lambda}{M_{Pl}}m\rho_b t + \frac{m\Lambda C}{a^3} \tag{9}$$

Here C is an integration constant which has to be determined from the present DM density. The second term (ρ_{dust}) corresponds to the dust like evolution. The form of the density is similar to those obtained in dynamical space-time theories as discussed in.[33–35] For the second term to dominate (such that ρ_m behaves as dust), it can be shown that one needs to satisfy the following constraint:

$$\frac{\alpha\Lambda\rho_b}{M_{Pl}\rho_{dust}}mt_0 \leq 1 \tag{10}$$

where t_0 is the present age of the universe.

Bounds on the model parameters:

- **From the EoS**- Equation of state for the DM superfluid (assuming negligible interaction) is given by,

$$w = \frac{\rho_{dust}^2}{12\Lambda^2 m^6} \tag{11}$$

For DM to behave as dust at the background level, Λ should be bounded from below,

$$\Lambda \gg 0.1 \left(\frac{m}{\text{eV}}\right)^{-3} \text{eV} \tag{12}$$

- **From coupling to baryons**- From (10) and (12), and assuming a constant baryon-to-DM ratio ($\rho_{dust}/\rho_b = 6$), we get,

$$\alpha \ll 2.4 \times 10^{-4} \left(\frac{m}{\text{eV}}\right)^2 \tag{13}$$

These bounds are different from the bounds obtained for galaxies, as discussed in.[1,28]

3.2. *Perturbations*

Study of linear perturbation theory in the context of ΛCDM has been an important step towards understanding the evolution of the universe. CMB spectra carries information about the inhomogeneities present in the early universe. Hence, any cosmological model needs to satisfy the CMB data to a high degree of accuracy. This requires analysing the matter power spectrum resulting from the initial density perturbations. In this section, we examine the growth of cosmological perturbations in DM superfluid model at linear order.

The Lagrangian of the theory in an FLRW matter dominated universe is given as,

$$\mathcal{L} = c_1 \left(\dot{\theta} - \frac{(\nabla\theta)^2}{2m} - m\Phi\right)^{3/2} - c_2\rho_b\theta \tag{14}$$

where c_1, c_2 are constants expressed as,

$$\begin{aligned} c_1 &= \frac{2\Lambda(2m)^{3/2}}{3} \\ c_2 &= \alpha\frac{\Lambda}{M_{Pl}} \end{aligned} \tag{15}$$

Here, $\theta(x,t) = \bar{\theta}(t) + \delta\theta(x,t)$ is a scalar field which is a fucntion of both space and time.

We can find the Euler-Lagrange equation from the above equation as:

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\theta}}\right) - \frac{\partial\mathcal{L}}{\partial\theta} = 0 \tag{16}$$

This gives us the background equation of motion as shown in the previous subsection as well as the first order perturbation equation of the dark matter density. But this single governing equation is inadequate to obtain the complete numerical solution which requires a complete set of differential equations.

In order to get the full set of perturbation equations, we start with the fluid equations that govern the dynamics of the dark matter superfluid. The fluid equations, namely the continuity equation and the Navier-Stokes equation can be derived using the Hamiltonian formalism, as described in.[36] In,[36] the authors work out the fluid equations for an interacting two-component BEC dark matter. Here in this work, we follow the same prescription for a superfluid dark matter which interacts with the baryonic matter. The corresponding Lagrangian is given by (14).

From the Lagrangian, we get the conjugate momentum as,

$$\begin{aligned}\Pi_\theta &= \frac{\partial \mathcal{L}}{\partial \dot{\theta}} \\ &= \Lambda(2m)^{3/2}\left[\dot{\theta} - m\Phi - \frac{(\nabla\theta)^2}{2m}\right]^{1/2}\end{aligned} \tag{17}$$

The Hamiltonian H describing the superfluid can be obtained as,

$$H = \Pi_\theta \dot{\theta} - \mathcal{L} \tag{18}$$

Since, $\dot{\theta} = m\Phi + \frac{(\nabla\theta)^2}{2m} + \frac{\Pi_\theta^2}{\Lambda^2(2m)^3}$ from (17), we get the Hamiltonian H as follows,

$$H = \frac{\Pi_\theta^3}{3\Lambda^2(2m)^3} + \left(m\Phi + \frac{(\nabla\theta)^2}{2m}\right)\Pi_\theta + \frac{\alpha\Lambda}{M_{pl}}\rho_b\theta \tag{19}$$

3.2.1. *Hamilton's equation of motion*

The Hamilton's equations of motion are :

$$\dot{\theta} = \frac{\partial H}{\partial \Pi_\theta} \tag{20}$$

and

$$\dot{\Pi}_\theta = -\frac{\partial H}{\partial \theta} \tag{21}$$

For this model, the two equations become, respectively,

$$\dot{\theta} = \frac{\Pi_\theta^2}{\Lambda^2(2m)^3} + m\Phi + \frac{(\nabla\theta)^2}{2m} \tag{22}$$

and

$$\dot{\Pi}_\theta = \frac{1}{m}\nabla\cdot(\Pi_\theta\nabla\theta) - \frac{\alpha\Lambda}{M_{pl}}\rho_b \tag{23}$$

3.2.2. *Fluid equations*

In order to get the fluid equations from the above Hamilton's equations of motion, we identify the terms as corresponding hydrodynamical variables. We define the mass density term (as the co-efficient of Φ in the Hamiltonian) and the four-velocity of the fluid, $\vec{u}$ as

$$\rho_m = m\Pi_\theta, \quad \vec{u} = -\frac{\nabla\theta}{m}. \tag{24}$$

Using the above definitions, we get the fluid equations from equation (23) and (22) as follow,

$$\dot{\rho}_m + \nabla\cdot(\rho_m\vec{u}) = -\frac{\alpha\Lambda m}{M_{pl}}\rho_b \tag{25}$$

$$\dot{\vec{u}} + (\vec{u}\cdot\nabla)\vec{u} = -\frac{\rho_m\nabla\rho_m}{4\Lambda^2 m^6} - \nabla\Phi \tag{26}$$

These are the two fluid equations: Continuity equation and Navier-Stokes equation.

Now, the Poisson's equation can be written as

$$\nabla^2\Phi = 4\pi G(\bar{\rho} + \delta\rho) \tag{27}$$

Integrating twice and substituting the background density using Friedmann equations, we get the potential as:

$$\Phi = -\frac{1}{2}(\dot{H} + H^2)l^2 + \phi \tag{28}$$

where l is the proper distance defined as $\vec{l} = a(t)\vec{x}$ and ϕ is the potential due to inhomogeneities.

Similarly, the four-velocity $\vec{u}$ can be split into two parts, Hubble flow and a peculiar velocity $\vec{v}$ as follows:

$$\vec{u} = H\vec{l} + \vec{v} \tag{29}$$

Expressing everything in comoving co-ordinates $\vec{x}$ and using $\nabla_l = \frac{1}{a(t)}\nabla_x$, we get,

$$\dot{\rho}_m + 3H\rho_m + \frac{1}{a}\nabla\cdot(\rho_m\vec{v}) = -\frac{\alpha\Lambda m}{M_{pl}}\rho_b \tag{30}$$

and

$$\dot{\vec{v}} + H\vec{v} + \frac{1}{a}(\vec{v}\cdot\nabla)\vec{v} = -\frac{\rho_m\nabla\rho_m}{4a\Lambda^2 m^6} - \frac{\nabla\phi}{a} \tag{31}$$

These are the two fluid equations of motion that we shall use for the rest of our calculations.

3.2.3. *Evolution of perturbations*

The total DM density ρ_m and the baryonic density ρ_b can be split into two parts: background and perturbation:

$$\rho_m = \bar{\rho}_m + \delta\rho_m, \quad \rho_b = \bar{\rho}_b + \delta\rho_b$$

respectively.

We define, the relative density perturbations for these two components as,

$$\delta_m = \frac{\delta\rho_m}{\bar{\rho}_m + \bar{\rho}_b} \quad \text{and} \quad \delta_b = \frac{\delta\rho_b}{\bar{\rho}_m + \bar{\rho}_b}.$$

In the linear perturbation regime, we treat $\delta\rho_m$, $\delta\rho_b$ and $\vec{v}$ to be small, and hence, neglect the higher orders of these terms. Perturbing the two fluid equations in the linear regime gives:

$$\dot{\delta}_m + \frac{\bar{\rho}_m}{a\bar{\rho}}\nabla\cdot\vec{v} = -\frac{\alpha\Lambda m}{M_{pl}}\delta_b \tag{32}$$

and

$$\dot{\vec{v}} + H\vec{v} = -\frac{\bar{\rho}_m\nabla\delta\rho_m}{4a\Lambda^2 m^6} - \frac{1}{a}\nabla\phi \tag{33}$$

By using the above equations along with the Poisson's equation and assuming $\bar{\rho}_m \sim \bar{\rho}$, we get the evolution equation for δ_m as follows:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{\bar{\rho}_m\delta_m}{2M_{pl}^2} - \frac{\bar{\rho}_m^2\nabla^2\delta_m}{4a^2\Lambda^2 m^6} = -\frac{\alpha\Lambda m\dot{\delta}_b}{M_{pl}} - 2H\frac{\alpha\Lambda m\delta_b}{M_{pl}} + \frac{\bar{\rho}_b\delta_b}{2M_{pl}^2} \tag{34}$$

This is a second order differential equation. The coefficient of the spatial derivative ∇^2 gives the square of the sound speed c_s. Thus, we get,

$$c_s^2 = \frac{\bar{\rho}_m^2}{4\Lambda^2 m^6} \tag{35}$$

Below in Fig. 1, we show the plot for c_s^2 vs the redshift z for $m = 1$ eV and $\Lambda = 500$ eV. We take the time evolution of the background density $\bar{\rho}_m$ as

$$\bar{\rho}_m = \frac{0.4(1+z)^3}{(1+1000)^3} \tag{36}$$

where the value of $\bar{\rho}_m$ at equality ($z = 1000$) is set as 0.4 eV^4 ([36]). As evident from the plot, the sound speed is very small (compared to the speed of light $c = 1$).

3.3. *Analytical Solution*

We now have the perturbation equations (32) and (33) supplemented by the Poisson equation. In the absence of baryons, we recover the usual evolution of CDM as a non-relativistic fluid. This can be seen as follows.

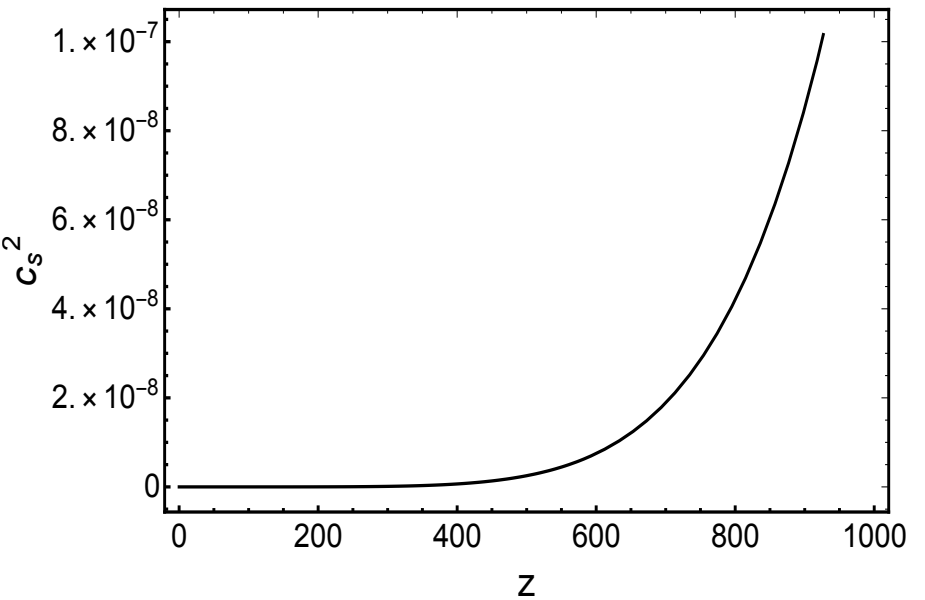


Fig. 1. Plot for c_s^2 vs z. As can be seen, the sound speed is very small compared to the speed of light ($c = 1$ in this case) at all times.

Without baryons:
In the absence of baryons, i.e. by setting $\bar{\rho}_b = 0$ and $\delta\rho_b = 0$ in equations (32) and (33), we obtain a set of equations as follows:

$$\dot{\delta}_m + \frac{1}{a}\nabla \cdot \vec{v} = 0 \tag{37}$$

and

$$\dot{\vec{v}} + H\vec{v} = -\frac{\bar{\rho}_m \nabla \delta\rho_m}{4a\Lambda^2 m^6} - \frac{1}{a}\nabla\phi \tag{38}$$

For $\Lambda \gg 1$, i.e, when the sound speed c_s is taken to be very small, the above set of equations reduce to the usual perturbation evolution equations in a ΛCDM model in the matter-dominated regime.[37] Thus, in the absence of baryons, usual CDM like evolution is recovered at the background level as well as for the perturbations. This can also be seen from the second order differential equation governing the evolution. From equation (34), in the absence of baryons, we obtain,

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{\bar{\rho}_m \delta_m}{2M_{pl}^2} - \frac{\bar{\rho}_m^2 \nabla^2 \delta_m}{4a^2\Lambda^2 m^6} = 0 \tag{39}$$

As expected, the above equation is the usual evolution equation for CDM in a flat ΛCDM universe, with a small sound speed given by (35). Since the sound speed

is very small, for all practical purposes, this term can be neglected and we end up with the following second order differential equation:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - \frac{\bar{\rho}_m \delta_m}{2M_{pl}^2} = 0 \tag{40}$$

This is the evolution equation of non-relativistic CDM in the ΛCDM model.[38] The growing solution of this equation is the usual CDM like evolution, which is $\delta \propto a$ for a matter dominated universe.[37,38] In the absence of baryons, the superfluid dark matter thus behaves exactly like CDM and the growth of perturbations follow the ΛCDM-like evolution.

With baryons:
We now investigate how the perturbations evolve when both baryons and baryonic interactions with superfluid dark matter is present. A complete solution is to be found numerically for different values of the model parameters. This will be pursued in the next section. For now, we try to find the generic nature of the DM perturbations growth with some basic assumptions regarding the baryonic density evolution and the interaction strength between baryons and the superfluid. We assume that the baryon perturbation δ_b follows the same rate of growth as in ΛCDM i.e. at late times $\delta_b \propto a$ as obtained from the observed power spectrum. We now consider equation (34) which is the single second order evolution equation. We are only interested in the temporal behaviour of δ_m. With the assumption that $c_s^2 \ll 1$, equation (34) can be analytically solved for two limiting cases. For the first case when the interaction strength is negligible, one can set $\alpha = 0$. With this condition, the only remaining term in the RHS of (34) is the last term. The resulting equation has a power law solution of the form $\delta \propto t^{2/3} \propto a$ in the matter dominated era. This is the usual CDM-like behaviour as expected when the interaction strength is negligible. The other limiting case solution can be obtained by setting a very strong interaction strength i.e. by taking a large enough α such that the first two terms in the RHS of (34) become dominant. In this case, the solution for the matter dominated universe comes out to be of the form

$$\delta_m \propto t^{5/3} \propto a^{5/2} \tag{41}$$

The time evolution of The DM perturbation growth in the two cases is thus captured by the following form:

$$\delta_m \propto a \qquad \text{for no interaction} \tag{42}$$

$$\delta_m \propto a^{5/2} \qquad \text{for strong interaction} \tag{43}$$

Important distinct features arise when we look at the time evolution of δ_m for each mode. During the matter dominated era, δ_m grows as a in ΛCDM whereas in this model, it grows as $a^{5/2}$ i.e. at a much faster rate compared to ΛCDM.

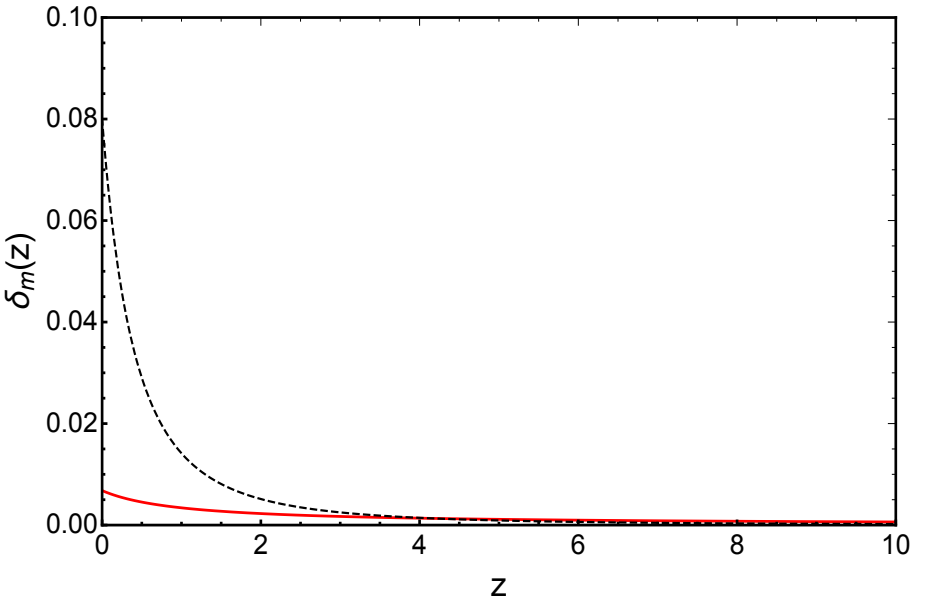


Fig. 2. Growth of δ_m with respect to z. The red solid line represents the growth for ΛCDM and the black dashed line corresponds to the growth for superfluid for $\alpha = 10^{-6}$. The growth in case of superfluid DM is dominated by the term proportional to $a^{5/2}$ and is higher compared to ΛCDM. The higher the value of α, i.e. the stronger the superfluid DM-baryonic interaction, the steeper is the growth rate (as will be discussed in the next section).

For convenience, we write the evolution of δ_m in terms of the redshift:

$$\delta_m \propto \frac{1}{(1+z)^{5/2}} \tag{44}$$

Fig. 2 shows the nature of growth in both the models (red solid curve representing ΛCDM, black dashed curve representing superfluid DM).

A proper way to find the full solution for the perturbed quantities is to solve coupled differential equations using a numerical approach. In the next section we solve the perturbation equations numerically in the linear regime and look for any possible deviations from ΛCDM.

4. Numerical Solution

In order to obtain the solutions for δ_m, we rewrite equations (32) and (33) in the Fourier domain in physical co-ordinate as,

$$\dot{\delta}_m + \frac{\bar{\rho}_m}{\bar{\rho}}(ikv) = -\frac{\alpha \Lambda m}{M_{pl}}\delta_b \tag{45}$$

and

$$ik\dot{v} + ikHv = \frac{k^2\bar{\rho}_m\bar{\rho}\delta_m}{4\Lambda^2 m^6} + \frac{a^2}{2M_{pl}^2}\left(\delta\rho_m + \delta\rho_b + \frac{3iaH\bar{\rho}v}{k}\right) \tag{46}$$

To solve the above equations, we write them in terms of redshift,

$$-H(1+z)\frac{d\delta_m}{dz} + \frac{\bar{\rho}_m}{\bar{\rho}}(ikv) = -\frac{\alpha\Lambda m}{M_{pl}}\delta_b \tag{47}$$

and

$$-ikH(1+z)\frac{dv}{dz} + ikHv = \frac{k^2\bar{\rho}_m\bar{\rho}\delta_m}{4\Lambda^2 m^6} + \frac{\bar{\rho}}{2M_{pl}^2(1+z)^2}\left(\delta_m + \delta_b + \frac{3iHv}{k(1+z)}\right) \tag{48}$$

Parameters and initial conditions:

The model parameters involved are m, Λ and α. We take $m = 1$ eV and $\Lambda = 500$ eV while keeping the parameter α as free parameter which is varied to check where the model deviates from flat ΛCDM.

We integrate the perturbation equations using the following initial conditions at the epoch of equality $z = 1000$: We set $\delta_b(z = 1000) = \delta_m(z = 1000) = 10^{-5}$ and $H(z = 1000) = m = 1$ eV.

Since $\bar{\rho_m} \gg \bar{\rho}_b$, we assume $\bar{\rho} = \bar{\rho}_m + \bar{\rho}_b \approx \bar{\rho}_m$ as given in (36).

The initial value of v at $z = 1000$ is chosen to be around 1. For the time evolution of the background density and Hubble parameter, we take the usual ΛCDM evolution of these quantities in matter-dominated era, i.e. $\bar{\rho}_m \propto 1/a^3$ and $H \propto 1/a^{3/2}$. Furthermore, we take $\delta_b \propto a$. We keep the wavenumber k fixed at 0.0001 eV, although the nature remains same for larger values of k.

Figure 3 shows the evolution of the DM density perturbation δ_m with respect to the redshift z for different values of $\alpha = 10^{-8},\ 10^{-7},\ 10^{-6},\ 10^{-4}$ and also for ΛCDM corresponding to $\alpha = 0, \Lambda \to \infty$. As expected, the smaller the value of α, the closer the resemblance with ΛCDM-like evolution. As we see in Fig. 3, the plot for $\alpha = 10^{-8}$ coincides with ΛCDM. When α is large enough, the growth is very steep. This is because a large enough α implies large interaction strength between the superfluid phonons and baryons, ensuring that structure formation takes place at an earlier epoch as compared to ΛCDM.

In Figure 4, we plot the relative differences between the perturbation growth in ΛCDM model and superfluid DM model for different values of α in terms of $\delta_{\text{superfluid}}/\delta_{\Lambda\text{CDM}}$. As expected, the ratio is very high at a lower redshift. As we go to higher redshifts, the ratio tends to 1 i.e., they eventually agree with ΛCDM at very high redshifts and matches exactly at $z = 1000$ where we set our initial conditions. The ΛCDM model corresponds to $\alpha = 0$. For $\alpha = 10^{-8}$, the deviation from ΛCDM at low redshift goes up to 0.13% at $z = 0.01$. The larger the value of α, the higher is the ratio, implying a stronger deviation from ΛCDM at low enough redshifts. As α is increased to 10^{-7}, the deviation from ΛCDM becomes much larger ($\sim 62\%$).

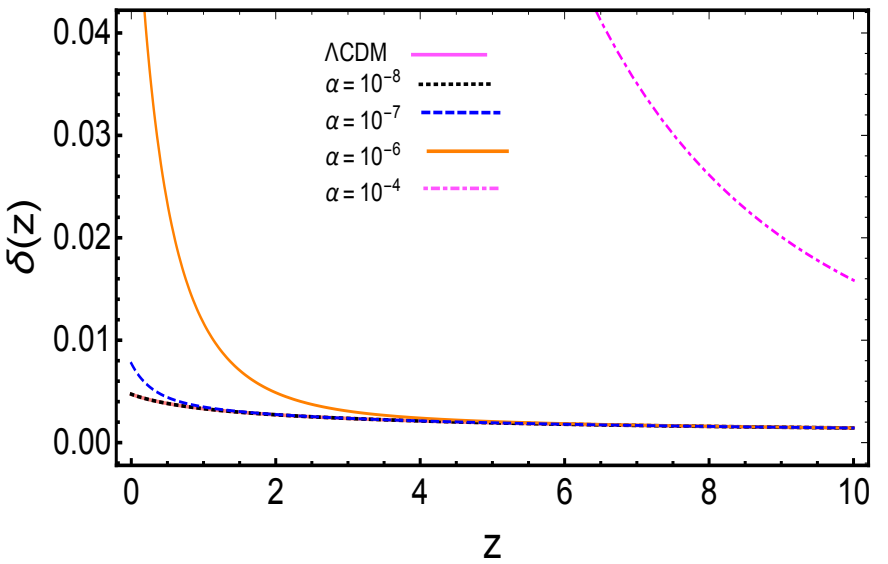


Fig. 3. Plot for δ_m vs z for different values of α as obtained numerically. The different curves correspond to different α as shown in the figure. The curve corresponding to $\alpha = 10^{-8}$ coincides with ΛCDM. For higher values of α, the deviation from ΛCDM increases gradually. The growth rate at a given redshift is maximum for $\alpha = 10^{-4}$ for the cases considered here.

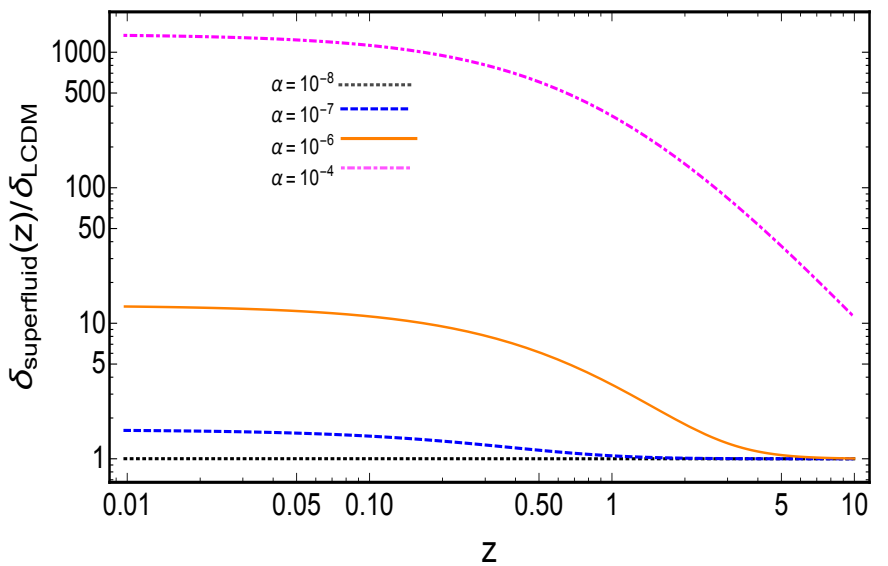


Fig. 4. Plot for $\delta_{\rm superfluid}/\delta_{\Lambda\rm CDM}$ vs z for different values of α. For an exact coincidence with ΛCDM, this ratio should be 1. Values > 1 signify larger deviation from ΛCDM. For $\alpha = 10^{-8}$, the ratio is almost nearly 1 showing a deviation only up to 0.13% at $z = 0$. For $\alpha = 10^{-7}$, the deviation increases up to 62% at the same redshift.

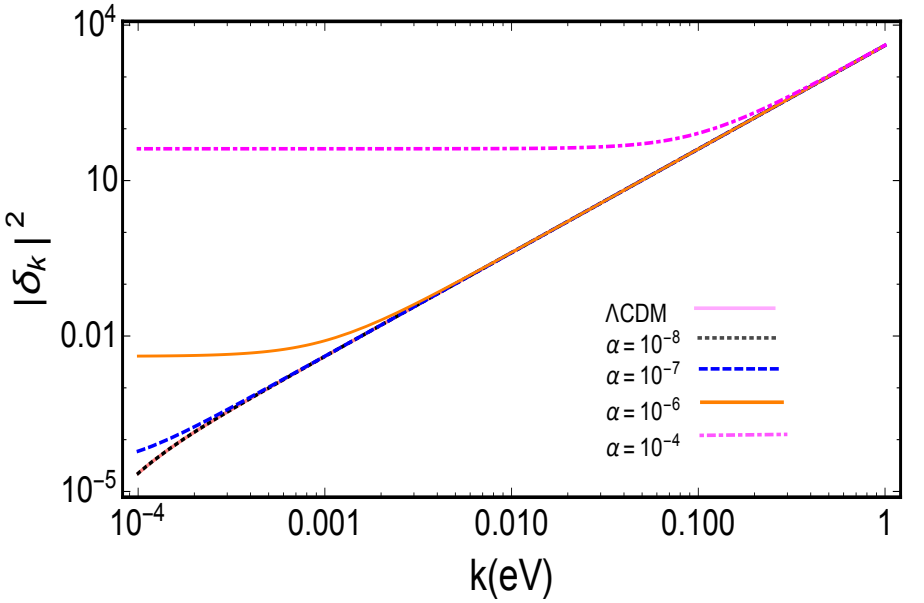


Fig. 5. Plot for $|\delta_k|^2$ vs k at $z = 0$. This is the matter power spectrum up to a constant factor. The plot reaffirms our previous results. In general the power increase approximately linearly with k. However, for large enough α, it becomes relatively flatter.

We can also plot the matter power spectrum $P(k)$ as a function of k at $z = 0$. The matter power spectrum $P(k) \propto |\delta_m(k)|^2$. In Fig. 5, we plot $|\delta_m(k)|^2$ vs. k which shows how the power varies for different values of α. As shown in the figure, the power spectrum for $\alpha = 10^{-8}$ matches with the ΛCDM prediction. As can be seen, the power increases for larger values of α at a given value of k. This is because the perturbation growth is stronger for large α as discussed earlier.

5. Results and Discussions

The superfluid dark matter model is a very promising and newly emerging model of cosmology combining together the rich physics of condensed matter, particle physics and cosmology. In view of its success in explaining a number of observations within the galaxies where ΛCDM fails to provide a satisfactory explanation, this model can be said to offer a greater understanding of the universe. In their earlier works, Khoury and his collaborators have investigated the implications of this model at galactic scales. However, a complete study of cosmological implications have not been performed earlier. In this paper, we have tried to investigate, both analytically and numerically, whether the predicted cosmology of the model tallies well with the observations and how different the predictions are from that of ΛCDM. In the realm of non-relativistic low energy effective theory of superfluid, the background cosmology agrees with the predictions of ΛCDM, and this gives a constraint on the

two model parameters α and Λ which turn out to be different than their galactic scale constraints. This result has also been discussed in.[1] At the level of first order perturbation, we find that the above constraints lead to a cosmology which differ significantly from ΛCDM. In particular, our analytical results suggests that the growth of density perturbations of dark matter superfluid roughly goes as $a^{5/2}$, which is much higher compared to the ΛCDM picture ($\delta_m \propto a$). This might be due to the strong interaction between superfluid phonons and baryonic matter. This behaviour has also been verified from the numerical solutions. For the numerical calculations, in particular, we have kept two of the model parameters m and Λ fixed at 1 eV and 500 eV respectively. This gives an upper bound on the third parameter: $\alpha \leq 10^{-8}$ corresponding to just 0.13% deviation from ΛCDM. This is different from the value quoted in.[1] The bound obtained in,[1] for $m = 1$ eV, is $\alpha \leq 10^{-4}$, which, even though predicts the correct background evolution, strongly deviates from ΛCDM in the context of perturbation growth in the present epoch. This can be seen in Figs. 3, 4 and 5. In our analysis, we have assumed the baryonic component to follow standard dust evolution ($\propto \frac{1}{a^3}$). In the absence of baryons, however, this model successfully reproduces the usual non-relativistic CDM evolution as obtained in case of a flat matter dominated ΛCDM universe, both at the background as well as first order perturbation level.

A more complete analysis of the perturbation growth should rely on the proper relativistic extension of the theory, which has not been attempted in this paper. Some relativistic models have been discussed in the original paper,[1] however a rigorous analysis is still lacking. We hope to address the same in a future work. Our work looks into the solution in the linear regime where perturbations are taken to be small. In future, we plan to extend our analysis to the non-linear regime and study the structure formation through spherical collapse. It would also be interesting to see how well this model predicts the CMB or the halo mass function.

Acknowledgments

S. Banerjee would like acknowledge funding from Israel Science Foundation and John and Robert Arnow Chair of Theoretical Astrophysics. S. Bera would like to acknowledge support from Navajbai Ratan Tata Trust and IUCAA research grant. S. Banerjee and S. Bera would also like to thank the hospitality of the Tata Institute of Fundamental Research, India, where part of the work was done. DFM thank the Research Council of Norway for their support. Computations were performed on resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway.

References

1. L. Berezhiani and J. Khoury, *Theory of dark matter superfluidity, Phys. Rev.* D **92** (2015) 103510.

2. P. A. R. Ade *et al.* [Planck Collaboration], *XIV. Dark energy and modified gravity, Astron. Astrophys.* **594** (2016) A13.
3. M. Boylan-Kolchin, *et al.*, *Resolving cosmic structure formation with the Millennium-II Simulation, MNRAS* **398** (2009) 1150.
4. F. Iocco, M. Pato and G. Bertone, *Evidence for dark matter in the inner Milky Way, Nat. Phys.* **11** (2015) 245-248.
5. M. Vogelsberger *et al.*, *Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe, MNRAS* **444** (2014) 1518.
6. D. H. Weinberg, *et al.*, *Cold dark matter: Controversies on small scale, Proc. Nat. Acad. Sci.* **112** (2015) 12249.
7. S. McGaugh, *The baryonic Tully-Fisher relation of gas-rich galaxies as a test of CDM and MOND, Astron. J.* **143** (2012) 40.
8. M. G. Walker and J. Penarrubia, *A method for measuring (slopes of) the mass profiles of dwarf spheroidal galaxies, ApJ* **742** (2011) 20.
9. A. A. Klypin, *et al.*, *Where are the Missing Galactic Satellites?, ApJ* **522** (1999) 82-92.
10. M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *Too big to fail? The puzzling darkness of massive Milky Way subhaloes, MNRAS* **415** (2011) L40.
11. M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *The Milky Way's bright satellites as an apparent failure of ΛCDM, MNRAS* **422** (2012) 1203-1218.
12. A. A. Starobinsky, *A new type of isotropic cosmological models without singularity, Phys. Lett.* B **91** (1980) 99.
13. E. V. Linder, *Einstein's Other Gravity and the Acceleration of the Universe, Phys. Rev.* D **81** (2010) 127301.
14. S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, *Is cosmic speed-up due to new gravitational physics?, Phys. Rev.* D **70** (2004) 043528.
15. G. Cognola, E. Elizalde, S. Nojiri, S. D. Odintsov and S. Zerbini, *Dark energy in modified Gauss-Bonnet gravity: Late-time acceleration and the hierarchy problem, Phys. Rev.* D **73** (2006) 084007.
16. J. P. Uzan, *Varying Constants, Gravitation and Cosmology, Living Rev. Rel.* **14**, (2011) 2.
17. J. R. Brownstein and J. W. Moffat, *Galaxy rotation curves without nonbaryonic dark matter, ApJ.* **636** (2006) 721.
18. P. Horava, *Quantum gravity at a Lifshitz point, Phys. Rev.* D. **79** (2009) 084008.
19. M. Milgrom, *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis, Astrophys. J.* **270** (1983) 365.
20. P. Kroupa, M. Pawlowski and M. Milgrom, *The failures of the standard model of cosmology require a new paradigm, Int. J. Mod. Phys.* D **21** (2012) 1230003.
21. W. J. G. de Blok and S. S McGaugh, *Testing modified newtonian dynamics with low surface brightness galaxies: Rotation curve fits, ApJ* **508** (1998) 132.
22. C. Nipoti, P. Londrillo and L. Ciotti, *Galaxy merging in MOND, MNRAS* **381** (2007) 104.
23. C. Skordis, D. F. Mota, P. G. Ferreira and C. Boehm, *Large Scale Structure in Bekenstein's theory of relativistic Modified Newtonian Dynamics, Phys. Rev. Lett.* **96** (2006) 011301.
24. J. P. Bruneton, S. Liberati, L. Sindoni and B. Famaey, *Reconciling MOND and dark matter?, JCAP* **0903** (2009) 021.
25. C. M. Ho, D. Minic and Y. J. Ng, *Cold Dark Matter with MOND Scaling, Phys. Lett.* B **693** (2010) 567.
26. S. Alexander, E. McDonough and D. N. Spergel,*Chiral Gravitational Waves and Baryon Superfluid Dark Matter, JCAP* **1805**, (2018) 003.

27. B. Famaey, J. Khoury and R. Penco, *Emergence of the mass discrepancy-acceleration relation from dark matter-baryon interactions, JCAP* **1803** (2018) 038.
28. L. Berezhiani, B. Famaey and J. Khoury, *Phenomenological consequences of superfluid dark matter with baryon-phonon coupling, JCAP*, **1809** (2018) 021.
29. A. Hodson, H. Zhao, J. Khoury and B. Famaey, *Galaxy Clusters in the Context of Superfluid Dark Matter, Astron. Astrophys.* **607** (2017) A108.
30. J. Khoury, *Another Path for the Emergence of Modified Galactic Dynamics from Dark Matter Superfluidity, Phys. Rev.* D **93** (2016) 103533.
31. M. Lisanti, M. Moschella, N. J. Outmezguine and O. Slone, *The Inconsistency of Superfluid Dark Matter with Milky Way Dynamics*, ArXiv: 1911.12365 (2019).
32. M. Lisanti, M. Moschella, N. J. Outmezguine and O. Slone, *Testing Dark Matter and Modifications to Gravity using Local Milky Way Observables, Phys. Rev.* D **100** (2019) 083009.
33. D. Benisty, E. Guendelman and Z. Haba, *Unification of dark energy and dark matter from diffusive cosmology, Phys. Rev.* D **99** (2019) 123521.
34. D. Benisty and E. Guendelman, *Interacting Diffusive Unified Dark Energy and Dark Matter from Scalar Fields, Eur. Phys.* J.C **77** (2017) 396.
35. S. Banerjee, D. Benisty and E. Guendelman, *Running Vacuum from Dynamical Space-time Cosmology*, ArXiv: 1910.03933 (2019).
36. E. G. M. Ferreira, G. Franzmann, J. Khoury and R. Brandenberger, *Unified Superfluid Dark Sector, JCAP* **1908** (2019) 027.
37. S. Dodelson, *Modern Cosmology*, Academic Press (2003).
38. R. Brandenberger, *Lectures on the Theory of Cosmological Perturbations*, arXiv:hep-th/0306071 (2003).

Interacting dark sector in the late Universe: Mapping fields and fluids, and observational signatures

Joseph P. J.* and S. Shankaranarayanan†

Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India

*E-mail: josephpj@iitb.ac.in

†E-mail: shanki@phy.iitb.ac.in

In this work, we discuss a cosmological model with dark energy – dark matter interaction. Demanding that the interaction strength Q_ν in the dark sector must have a field theory description, a unique form of interaction strength can be obtained. We show the equivalence between the fields and fluids for the $f(R, \chi)$ model where f is an arbitrary, smooth function of R and classical scalar field χ, which represents dark matter. Up to first order in perturbations, there is a one-to-one mapping between the classical field theory description and the phenomenological fluid description of interacting dark energy and dark matter, which exists only for this unique form of interaction. Different formulations of interacting dark energy models in the literature can be classified into two categories based on the field-theoretic description. Then we discuss the quantifying tools to distinguish between the interacting and non-interacting dark sector scenarios. We focus on the variation of the scalar metric perturbed quantities as a function of redshift related to structure formation, weak gravitational lensing, and the integrated Sachs-Wolfe effect and show that the difference in the evolution becomes significant for lower redshifts ($z < 20$), for all length scales.

Keywords: Dark matter - dark energy interaction, based on the work Ref. 1

1. Introduction

Dark matter dominates the galaxy mass, and dark energy forms the majority of our Universe's energy density.[2,3] However, we have little information about the properties of these two components that dominate the energy content of the Universe today.[4] The only information we have about the two components is that (i) Dark energy contributes negative pressure to the energy budget, and (ii) Dark matter has negligible, possibly zero, pressure. The above properties are based on gravitational interactions. More importantly, we do not know how they interact with each other and Baryons/Photons.

It has been shown that the dark matter-dark energy interaction can reconcile the tensions in the Hubble constant H_0.[5,6] In most interacting dark sector models, phenomenologically, the interaction is proposed between the fluid terms in the dark sector (Cf. Ref.[1]). More specifically, individually, dark matter (DM) and dark energy (DE) do not satisfy the conservation equations; however, the combined sector satisfies the energy conservation equation,[7] i. e.,

$$\nabla^\mu T^{(\mathrm{DE,DM})}_{\mu\nu} = Q^{(\mathrm{DE,DM})}_\nu, \quad Q^{(\mathrm{DE})}_\nu + Q^{(\mathrm{DM})}_\nu = 0 \tag{1}$$

where Q determines the interaction strength between dark matter and dark energy. Since the gravitational effects on dark matter and dark energy are opposite, even a small interaction can impact the cosmological evolution.[8] Since we have little information about the dark sector, in many of these models, the interaction strength Q_ν is put in by hand. However, it is unclear whether these broad classes of phenomenological models can be obtained from a field theory action.

In this work, we show that under conformal transformations, $f(R,\chi)$ is equivalent to a model with two coupled scalar fields. The dark energy - dark matter interaction, represented by the coupling between the classical scalar fields, can also be represented by the evolution equations of the dark energy (represented by a scalar field) and dark matter (represented by a fluid). We show that a one-to-one mapping exists between the field theory and fluid description for a unique interaction term.

To detect the signatures of the interacting dark sector from the observations, one needs to construct theoretical tools to distinguish interacting dark sector models from non-interacting ones. We focus on the three cosmological phenomena: (i) structure formation, (ii) weak gravitational lensing (iii) integrated Sachs Wolfe effect. We study the evolution of the relevant perturbed quantities in the redshift range $0 \leq z \leq 1500$ at different length scales. We see that there is a clear difference in the perturbed evolution in the interacting dark sector model as compared to non-interacting ones. This difference becomes significant for $z < 20$, especially at smaller length scales.

2. Dark sector interaction: Field and fluid description

Consider the following action in Jordan frame:

$$S_J = \int d^4x\sqrt{-\tilde{g}}\left[\frac{1}{2\kappa^2}f(\tilde{R},\tilde{\chi}) - \frac{1}{2}\tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\tilde{\chi}\tilde{\nabla}_\nu\tilde{\chi} - V(\tilde{\chi})\right] \tag{2}$$

where $f(\tilde{R},\tilde{\chi})$ is an arbitrary, smooth function of Ricci scalar, and scalar field $\tilde{\chi}$, and $V(\chi)$ is the self-interaction potential of the scalar field $\tilde{\chi}$. Under the conformal transformation:

$$g_{\mu\nu} = \Omega^2\tilde{g}_{\mu\nu}, \quad \text{where} \quad \Omega^2 = F(\tilde{R},\tilde{\chi}) \equiv \frac{\partial f(\tilde{R},\tilde{\chi})}{\partial\tilde{R}} \tag{3}$$

and a field redefinition, the action in the Einstein frame takes the following form

$$S = \int d^4x\sqrt{-g}\left(\frac{1}{2\kappa^2}R - \frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - U(\phi) - \frac{1}{2}e^{2\alpha(\phi)}g^{\mu\nu}\nabla_\mu\chi\nabla_\nu\chi - e^{4\alpha(\phi)}V(\chi)\right). \tag{4}$$

where

$$U = \frac{F\tilde{R} - f}{2\kappa^2F^2}\,.$$

and $\alpha(\phi)$ denotes the interaction between dark energy and dark matter.

Defining the dark matter fluid by specifying the four velocity energy density and pressure

$$u_\mu = -\left[-g^{\alpha\beta}\nabla_\alpha\chi\nabla_\beta\chi\right]^{-\frac{1}{2}}\nabla_\mu\chi \tag{5}$$

$$p_m = -\frac{1}{2}e^{2\alpha}\left[g^{\mu\nu}\nabla_\mu\chi\nabla_\nu\chi + e^{2\alpha}V(\chi)\right], \quad \rho_m = -\frac{1}{2}e^{2\alpha}\left[g^{\mu\nu}\nabla_\mu\chi\nabla_\nu\chi - e^{2\alpha}V(\chi)\right]. \tag{6}$$

Then the interaction function in the field theory and fluid descriptions are given by

$$Q_\nu^{(\mathrm{F})} = -e^{2\alpha(\phi)}\alpha_{,\phi}(\phi)\nabla_\nu\phi\left[\nabla^\sigma\chi\nabla_\sigma\chi + 4e^{2\alpha(\phi)}V(\chi)\right] = -\alpha_{,\phi}(\phi)\nabla_\nu\phi(\rho_m - 3p_m) \tag{7}$$

A one-to-one mapping between the field theory description and fluid description of the interacting dark sector described above exist *only* for this form of interaction function. A classification of interacting dark sector models based on the existence of this mapping is given below.[1]

Interacting DE-DM model	**DE-DM Interaction** $\nabla^\mu T^{(\mathrm{DE,DM})}_{\mu\nu} = Q^{(\mathrm{DE,DM})}_\nu$	**Is** $\boldsymbol{Q_\nu \propto Q_\nu^{(\mathrm{F})}}$?
Amendola - 1999[9]	$\dot{\rho}_m + 3H\rho_m = -C\rho_m\dot{\phi}$	Yes
Amendola - 1999[10]	$\dot{\rho}_m + 3H\rho_m = -C\rho_m\dot{\phi}$	Yes
Billyard & Coley -1999[11]	$\dot{\phi}(\ddot{\phi} + 3H\dot{\phi} + kV) = \frac{(4-3\gamma)}{2\sqrt{\omega+\frac{3}{2}}}\dot{\phi}\mu$	Yes
Olivares.etal - 2005[12]	$\frac{d\rho_c}{dt} + 3H\rho_c = 3Hc^2(\rho_c + \rho_x)$	No
Amendola.etal - 2006[13]	$\dot{\rho}_{DM} + 3H\rho_{DM} - \delta(a)H\rho_{DM} = 0$	No
Olivares.etal - 2007[14]	$\dot{\rho}_c + 3H\rho_c = 3Hc^2(\rho_x + \rho_c)$	No
Boehmer.etal - 2008[15]	$\dot{\rho}_c + 3H\rho_c = -\sqrt{2/3}\kappa\beta\rho_c\dot{\varphi}$ $\dot{\rho}_c + 3H\rho_c = -\alpha H\rho_c$	Yes No
Caldera-Cabral.etal - 2008[16]	$\dot{\rho}_c = -3H\rho_c + 3H(\alpha_x\rho_x + \alpha_c\rho_c)$ $\dot{\rho}_c = -3H\rho_c + 3(\Gamma_x\rho_x + \Gamma_c\rho_c)$	No No
He & Wang - 2008[17]	$\dot{\rho}_{DM} + 3H\rho_{DM} - \delta H\rho_{DM} = 0$ $\dot{\rho}_{DM} + 3H\rho_{DM} - \delta H(\rho_{DM} + \rho_{DE}) = 0$	No No
Pettorino & Baccigalupi - 2008[18]	$\phi'' + 2\mathcal{H}\phi' + a^2U_{,\phi} = a^2C_c\rho_c$	Yes
Quartin.etal - 2008[19]	$\frac{\mathrm{d}\rho_c}{\mathrm{d}N} + 3\rho_c = 3\lambda_x\rho_x + \lambda_c\rho_c$	No
Boehmer.etal - 2009[20]	$\dot{\rho}_c = -3H\rho_c - \frac{\alpha}{M_0}\rho_\varphi^2$ $\dot{\rho}_c = -3H\rho_c - \frac{\beta}{M_0}\rho_c^2$ $\dot{\rho}_c = -3H\rho_c - \frac{\gamma}{M_0}\rho_\varphi\rho_c$	No No No
Beyer.etal - 2010[21]	$\ddot{\varphi} + 3H\dot{\varphi} - \alpha M^3e^{-\alpha\varphi/M} = \frac{\beta}{M}\rho_\chi$	Yes
Lopez Honorez.etal - 2010[22]	$\dot{\rho}_{dm} + 3H\rho_{dm} = \beta(\phi)\rho_{dm}\dot{\phi}$	Yes
Avelino & Silva - 2012[23]	$\dot{\rho}_m + 3H\rho_m = \alpha Ha^\beta\rho_w$	No
Pan.etal - 2012[24]	$\dot{\rho}_m + 3H\rho_m = 3\lambda_mH\rho_m + 3\lambda_dH\rho_d$	No
Salvatelli.etal - 2013[25]	$\dot{\rho}_{dm} + 3\mathcal{H}\rho_{dm} = \xi\mathcal{H}\rho_{de}$	No
Chimento.etal - 2013[26]	$\rho'_{\mathrm{m}} + \gamma_{\mathrm{m}}\rho_{\mathrm{m}} = -\alpha\rho'\rho$	No
Amendola.etal - 2014[27]	$\dot{\rho}_\alpha + 3H\rho_\alpha = -\kappa\sum_i C_{i\alpha}\dot{\phi}_i\rho_\alpha$	Yes
Marra - 2015[28]	$\dot{\rho}_m + 3H\rho_m = \nu\delta_m^n\rho_m\dot{\phi}/M_{Pl}$	No
Bernardi & Landim - 2016[29]	$\dot{\rho}_m + 3H\rho_m = Q(\rho_\phi + \rho_m)\dot{\phi}$ $\dot{\rho}_m + 3H\rho_m = Q\rho_\phi\dot{\phi}$	No No
Pan & Sharov - 2016[30]	$\dot{\rho}_{dm} + 3\mathcal{H}\rho_{dm} = 3\lambda_mH\rho_{dm} + 3\lambda_dH\rho_d$	No
Bruck & Mifsud - 2017[31] [a]	$\nabla^\mu T^{DM}_{\mu\nu} = Q\nabla_\nu\phi$ $Q = \frac{C_{,\phi}}{2C}T_{DM} + \frac{D_{,\phi}}{2C}T^{\mu\nu}_{DM}\nabla_\mu\phi\nabla_\nu\phi$ $-\nabla_\mu\left[\frac{D}{C}T^{\mu\nu}_{DM}\nabla_\nu\phi\right]$	Yes if $D = 0$
Gonzalez & Trodden - 2018[33]	$\dot{\rho}_\chi + 3H\rho_\chi = \alpha'\dot{\phi}\rho_\chi$	Yes
Barros.etal - 2018[34]	$\dot{\rho}_c + 3H\rho_c = -\kappa\beta\dot{\phi}\rho_.$	Yes
Landim - 2019[35]	$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = -Q\rho_m$	Yes

[a] Violates causality condition ($D(\phi) > 0$) for the disformal transformations[32]

3. Cosmological evolution with dark energy – dark matter interaction

Consider the perturbed FRW metric in the Newtonian gauge

$$g_{00} = -(1+2\Phi)\,, \quad g_{0i} = 0, \quad g_{ij} = a^2(1-2\Psi)\delta_{ij}\,, \tag{8}$$

For the FRW Universe, the background evolution of the scalar fields in the field theory description is given by

$$\ddot{\overline{\chi}} + 3H\dot{\overline{\chi}} + e^{2\alpha}V_{,\chi}(\overline{\chi}) + 2\alpha_{,\phi}(\overline{\phi})\dot{\overline{\phi}}\dot{\overline{\chi}} = 0$$
$$\ddot{\overline{\phi}} + 3H\dot{\overline{\phi}} + U_{,\phi}(\overline{\phi}) + 4e^{4\alpha}\alpha_{,\phi}(\overline{\phi})V(\overline{\chi}) - e^{2\alpha}\alpha_{,\phi}(\overline{\phi})\dot{\overline{\chi}}^2 = 0\,.$$

In the fluid description, the dark energy scalar field $\bar{\phi}$ and the dark matter fluid energy density $\bar{\rho}_m$ evolve as

$$\ddot{\overline{\phi}}\dot{\overline{\phi}} + 3H\dot{\overline{\phi}}^2 + U_{,\phi}(\overline{\phi})\dot{\overline{\phi}} = \overline{Q}^{(\mathrm{F})}, \quad \dot{\overline{\rho}}_m + 3H(\overline{\rho}_m + \overline{p}_m) = -\overline{Q}^{(\mathrm{F})}$$

where the interaction term $\overline{Q}^{(\mathrm{F})}$ is given by

$$\overline{Q}^{(\mathrm{F})} = -\alpha_{,\phi}(\overline{\phi})\dot{\overline{\phi}}(\overline{\rho}_m - 3\overline{p}_m) = \alpha_{,\phi}(\overline{\phi})\dot{\overline{\phi}}e^{2\alpha(\overline{\phi})}\left[\dot{\overline{\chi}}^2 - 4e^{2\alpha}V(\overline{\chi})\right]. \tag{9}$$

Then the dark matter fluid energy density evolves as

$$\overline{\rho}_m = \overline{\rho}_{m_0}a^{-3(1+\omega_m)}e^{[\alpha(\overline{\phi})-\alpha_0](1-3\omega_m)}, \tag{10}$$

The perturbed evolution in the Newtonian gauge is given by

$$\delta\dot{\rho}_m + 3H(\delta p_m + \delta\rho_m) + (\overline{p}_m + \overline{\rho}_m)\left[\frac{\nabla^2\delta u^s}{a^2} - 3\dot{\Psi}\right] = -\delta Q \tag{11}$$

$$\dot{\overline{\phi}}\left(\dot{\delta\phi} - \frac{\nabla^2\delta\phi}{a^2} - 2\Phi\dot{\overline{\phi}} + U_{,\phi\phi}(\overline{\phi})\delta\phi\right) + \dot{\delta\phi}\left(\dot{\overline{\phi}} + 6H\dot{\overline{\phi}} + U_{,\phi}(\overline{\phi})\right)$$
$$-\frac{\dot{\overline{\phi}}^2}{2}\left(3\dot{\Psi} + \dot{\Phi} + 6H\Phi\right) = \delta Q, \tag{12}$$

where

$$\delta Q^{(\mathrm{F})} = -(\delta\rho_m - 3\delta p_m)\alpha_{,\phi}(\overline{\phi})\dot{\overline{\phi}} - (\overline{\rho}_m - 3\overline{p}_m)\left[\alpha_{,\phi\phi}(\overline{\phi})\dot{\overline{\phi}}\delta\phi + \alpha_{,\phi}(\overline{\phi})\dot{\delta\phi}\right] \tag{13}$$

The metric perturbations Φ and Ψ satisfy the equations

$$\Psi - \Phi = 0 \tag{14}$$

$$\dot{\Psi} + H\Phi = \frac{\kappa^2}{2}\left[\dot{\overline{\phi}}\delta\phi - (\overline{\rho}_m + \overline{p}_m)\,\delta u^s\right] \tag{15}$$

$$3H\dot{\Psi} - \frac{\nabla^2\Psi}{a^2} + 3H^2\Phi = -\frac{\kappa^2}{2}\left(\delta\rho_m + \dot{\delta\phi}\dot{\bar{\phi}} - \Phi\dot{\bar{\phi}}^2 + U_{,\phi}(\bar{\phi})\delta\phi\right) \tag{16}$$

$$\begin{aligned} &3\ddot{\Psi} + \frac{\nabla^2\Phi}{a^2} + 6\Phi\left(H^2 + \dot{H}\right) + 3H\left(2\dot{\Psi} + \dot{\Phi}\right) \\ &\quad = \frac{\kappa^2}{2}\left(\delta\rho_m + 3\delta p_m + 4\dot{\delta\phi}\dot{\bar{\phi}} - 4\Phi\dot{\bar{\phi}}^2 - 2U_{,\phi}(\bar{\phi})\delta\phi\right) \end{aligned} \tag{17}$$

4. Observable signatures of interacting dark sector

To detect the signatures of dark energy – dark matter interaction in cosmological observation, one needs to study the evolution of perturbed quantities related to cosmological observations. For this purpose, we consider the following cosmological phenomena and the relevant perturbed quantities

(1) Structure formation: $\delta_m(t,x,y,z) \equiv \frac{\delta\rho_m(t,x,y,z)}{\overline{\rho_m}(t)}$
(2) Weak lensing : $\Phi + \Psi$
(3) Integrated Sachs-Wolfe (ISW) effect: $\Phi' + \Psi'$

where *prime* denoted derivative with respect to $N \equiv \ln a$.

To illustrate the difference between the evolution of perturbations in the interacting and non-interacting scenarios, we consider the following dark energy scalar field potential and interaction function

$$U(\phi) \sim \frac{1}{\phi}, \quad \alpha(\phi) \sim \phi \tag{18}$$

We then study the evolution of the flowing quantities in the redshift range $0 \leq z \leq 1500$ at different length scales

$$\Delta\delta_m = \delta_{m_i} - \delta_{m_{ni}}, \quad \Delta\Phi = \Phi_i - \Phi_{ni}, \quad \Delta\Phi' = \Phi'_i - \Phi'_{ni} \tag{19}$$

Evolution these quantities are given in Figs. 1 and 2.

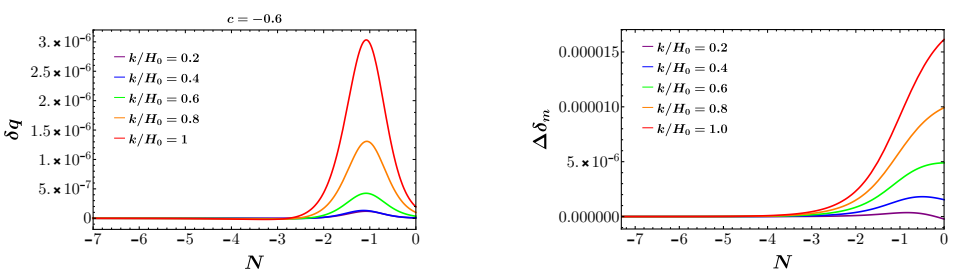


Fig. 1. Evolution of δq(left panel) and $\Delta\delta_m$(right panel) as a function of N for different values of k.

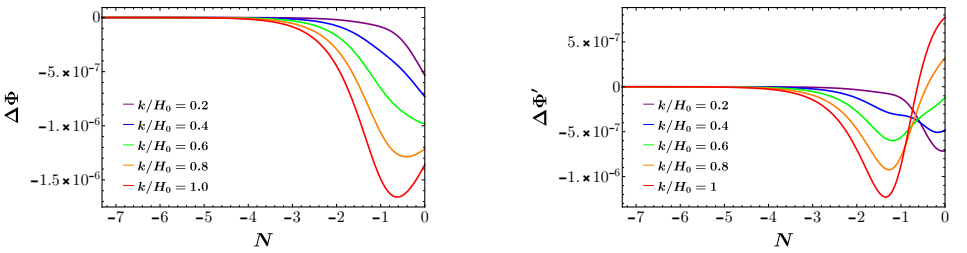


Fig. 2. Evolution of $\Delta\Phi$(left panel) and $\Delta\Phi'$(right panel) as a function of N for different values of k.

where δq is the scaled interaction function given by

$$\delta q = \frac{\delta Q}{H^3 M_{Pl}^2}. \tag{20}$$

Here we see that the evolution of perturbed quantities in the interacting dark sector models shows significant differences for $z < 20$, especially at lower length scale (larger values of k). This trend is consistent with the evolution of scaled interaction function δq.

5. Conclusion

In this work, we consider an interacting dark sector model. We demand that the energy-momentum of individual components of dark sector is conserved, but that of the individual components is not. This model is implemented through a field theory action where dark energy and dark matter are represented by classical scalar fields. This field theory description can be derived from a $f(\tilde{R}, c\tilde{h}i)$ by means of a conformal transformation and field redefinition. We show that there is a one-to-one mapping between the classical field theory description of the dark energy – dark matter interaction and the fluid description of the interacting dark sector. This mapping exist *only* for a unique interaction term $Q_\nu^{(\mathrm{F})}$. We then classified the popular interacting dark sector models found in literature based on whether or not this mapping is applicable to those models.

We then look at the evolution of first-order scalar perturbations in the interacting dark sector in the redshift range of $0 \leq z \leq 1500$. We consider an inverse power-law dark energy scalar field potential and a linear interaction function. To detect the potential signatures of dark energy - dark matter interaction from cosmological observations, we focus on three cosmological phenomena and related perturbed quantities: (i) structure formation (δ_m) (ii) weak gravitational lensing ($\Phi + \Psi$)(iii) integrated Sachs Wolfe effect ($\Phi' + \Psi'$). We see that the density perturbation δ_m grows at a faster rate in the interacting scenario, especially at the smaller length scales. The effect of the interaction becomes significant for redshift $z < 20$. This is consistent with the evolution of perturbed interaction function δq. We see a similar

trend in the case of Φ and Φ', which indicates that the future observations of weak gravitational lensing and integrated Sachs Wolfe effect can be used to distinguish between the interacting and non-interacting models of dark sector.

Acknowledgments

J.P.J. is supported by CSIR Senior Research Fellowship, India. The work is partially supported by the ISRO-Respond grant.

References

1. J. P. Johnson and S. Shankaranarayanan, Cosmological perturbations in the interacting dark sector: Mapping fields and fluids, *Phys. Rev. D* **103**, p. 023510 (2021).
2. A. G. Riess and Others, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
3. N. Aghanim *et al.*, Planck 2018 results. vi. cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
4. E. J. Copeland, M. Sami and S. Tsujikawa, Dynamics of dark energy, *Int. J. Mod. Phys.* **D15**, 1753 (2006).
5. E. Di Valentino, A. Melchiorri and O. Mena, Can interacting dark energy solve the H_0 tension?, *Phys. Rev. D* **96**, p. 043503 (2017).
6. A. Gómez-Valent, V. Pettorino and L. Amendola, Update on coupled dark energy and the h_0 tension, *Phys. Rev. D* **101**, p. 123513 (2020).
7. B. Wang, E. Abdalla, F. Atrio-Barandela and D. Pavon, Dark Matter and Dark Energy Interactions: Theoretical Challenges, Cosmological Implications and Observational Signatures, *Rept. Prog. Phys.* **79**, p. 096901 (2016).
8. Y. L. Bolotin, A. Kostenko, O. Lemets and D. Yerokhin, Cosmological Evolution With Interaction Between Dark Energy And Dark Matter, *Int. J. Mod. Phys. D* **24**, p. 1530007 (2014).
9. L. Amendola, Perturbations in a coupled scalar field cosmology, *Mon. Not. Roy. Astron. Soc.* **312**, p. 521 (2000).
10. L. Amendola, Coupled quintessence, *Phys. Rev. D* **62**, p. 043511 (2000).
11. A. P. Billyard and A. A. Coley, Interactions in scalar field cosmology, *Phys. Rev. D* **61**, p. 083503 (2000).
12. G. Olivares, F. Atrio-Barandela and D. Pavon, Observational constraints on interacting quintessence models, *Phys. Rev. D* **71**, p. 063523 (2005).
13. L. Amendola, G. Camargo Campos and R. Rosenfeld, Consequences of dark matter-dark energy interaction on cosmological parameters derived from SNIa data, *Phys. Rev. D* **75**, p. 083506 (2007).
14. G. Olivares, F. Atrio-Barandela and D. Pavon, Dynamics of Interacting Quintessence Models: Observational Constraints, *Phys. Rev. D* **77**, p. 063513 (2008).
15. C. G. Boehmer, G. Caldera-Cabral, R. Lazkoz and R. Maartens, Dynamics of dark energy with a coupling to dark matter, *Phys. Rev. D* **78**, p. 023505 (2008).
16. G. Caldera-Cabral, R. Maartens and L. Urena-Lopez, Dynamics of interacting dark energy, *Phys. Rev. D* **79**, p. 063518 (2009).
17. J.-H. He and B. Wang, Effects of the interaction between dark energy and dark matter on cosmological parameters, *JCAP* **06**, p. 010 (2008).
18. V. Pettorino and C. Baccigalupi, Coupled and Extended Quintessence: theoretical differences and structure formation, *Phys. Rev. D* **77**, p. 103003 (2008).

19. M. Quartin, M. O. Calvao, S. E. Joras, R. R. Reis and I. Waga, Dark Interactions and Cosmological Fine-Tuning, *JCAP* **05**, p. 007 (2008).
20. C. G. Boehmer, G. Caldera-Cabral, N. Chan, R. Lazkoz and R. Maartens, Quintessence with quadratic coupling to dark matter, *Phys. Rev. D* **81**, p. 083003 (2010).
21. J. Beyer, S. Nurmi and C. Wetterich, Coupled dark energy and dark matter from dilatation anomaly, *Phys. Rev. D* **84**, p. 023010 (2011).
22. L. Lopez Honorez, O. Mena and G. Panotopoulos, Higher-order coupled quintessence, *Phys. Rev. D* **82**, p. 123525 (2010).
23. P. Avelino and H. da Silva, Effective dark energy equation of state in interacting dark energy models, *Phys. Lett. B* **714**, 6 (2012).
24. S. Pan, S. Bhattacharya and S. Chakraborty, An analytic model for interacting dark energy and its observational constraints, *Mon. Not. Roy. Astron. Soc.* **452**, 3038 (2015).
25. V. Salvatelli, A. Marchini, L. Lopez-Honorez and O. Mena, New constraints on Coupled Dark Energy from the Planck satellite experiment, *Phys. Rev. D* **88**, p. 023531 (2013).
26. L. P. Chimento, M. G. Richarte and I. E. Sánchez García, Interacting dark sector with variable vacuum energy, *Phys. Rev. D* **88**, p. 087301 (2013).
27. L. Amendola, T. Barreiro and N. J. Nunes, Multifield coupled quintessence, *Phys. Rev. D* **90**, p. 083508 (2014).
28. V. Marra, Coupling dark energy to dark matter inhomogeneities, *Phys. Dark Univ.* **13**, 25 (2016).
29. F. F. Bernardi and R. G. Landim, Coupled quintessence and the impossibility of an interaction: a dynamical analysis study, *Eur. Phys. J. C* **77**, p. 290 (2017).
30. S. Pan and G. Sharov, A model with interaction of dark components and recent observational data, *Mon. Not. Roy. Astron. Soc.* **472**, 4736 (2017).
31. C. Van De Bruck and J. Mifsud, Searching for dark matter - dark energy interactions: going beyond the conformal case, *Phys. Rev. D* **97**, p. 023506 (2018).
32. J. D. Bekenstein, The Relation between physical and gravitational geometry, *Phys. Rev. D* **48**, 3641 (1993).
33. M. Carrillo González and M. Trodden, Field Theories and Fluids for an Interacting Dark Sector, *Phys. Rev. D* **97**, p. 043508 (2018), [Erratum: Phys. Rev. D 101, 089901 (2020)].
34. B. J. Barros, L. Amendola, T. Barreiro and N. J. Nunes, Coupled quintessence with a ΛCDM background: removing the σ_8 tension, *JCAP* **01**, p. 007 (2019).
35. R. G. Landim, Cosmological perturbations and dynamical analysis for interacting quintessence, *Eur. Phys. J. C* **79**, p. 889 (2019).

The role of self interactions in the cosmological evolution of warm dark matter

R. Yunis[1,2], C. R. Argüelles[1,3], D. López Nacir[4,5], C. Scóccola[3,6], N. Mavromatos[7,8], A. Krut[2]

[1] *International Center for Relativistic Astrophysics Network – ICRANet, Piazza della Repubblica 10, 65122 Pescara, Italy*

[2] *ICRA, Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy*

[3] *Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, B1900FWA La Plata, Argentina*

[4] *Departamento de Física Juan José Giambiagi, FCEyN UBA*

[5] *IFIBA CONICET-UBA, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón I, 1428 Buenos Aires, Argentina*

[6] *CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Argentina*

[7] *Theoretical Particle Physics and Cosmology Group, Physics Department, King's College London, Strand, London WC2R 2LS, UK*

[8] *Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, 9, Heroon Polytechneiou Str., Zografou Campus, Athens 157 80, Greece*

In this work we present a summary of recent studies on the effects of elastic self interactions in the evolution of Warm Dark Matter models (WDM), focusing on structure formation and the evolution of cosmological perturbations. We pay special attention to a particular class of sterile neutrino WDM known as νMSM and provide examples for the case of vector field self interactions. We calculate the effects of assuming self interacting dark matter in X-Ray astrophysical observations, in the formation of fermionic DM halos in (quasi) equilibrium states and in the evolution of DM perturbations in the early universe, assuming particle masses between $\mathcal{O}(1-100)$ keV. In the latter topic, we perform simulations using a modification to the public Boltzmann solver CLASS and compare our results with observations. We find self interactions to be an interesting addition to WDM models, which can alleviate tensions both present in standard CDM cosmology and regarding WDM itself, as well as provide an interesting avenue for DM halo formation.

Keywords: Dark Matter, Self Interactions, Cosmology

1. Introduction

Several observations at large and small scales, such as the distribution of large-scale structure, CMB anisotropies and the internal structure of DM halos has lead to a standard model of cosmology: ΛCDM.[1,2] DM in this model is assumed to be "Cold": a distribution of colissionless, non relativistic particles with negligible velocity dispersion, which forms structure in a "bottom-up" fashion. However, recent observations have challenged this paradigm[3]; namely, the so called missing satellites problem[4–6]; the too big to fail problem[7]; and the core-cusp problem,[3,8,9] among others. While these by no means rule out the standard model, it has sparked interest in the community in other families of DM models.

In one of these families of models, known as Warm Dark Matter (WDM),[10–12] the DM particles are produced while relativistic, but become non relativistic before the energy budget of the universe is dominated by matter. These particles would have a significant amount of velocity dispersion today, which causes a significant reduction of small scale structure population and an alleviation to some of the tensions we mentioned above. While traditionally these models were also expected to aid the *core-cusp* problem, it was shown[13] that the particle mass requirement for this to happen was inconsistent with existing bounds, in what was known as the "catch-22" problem in WDM. However, in recent years it has been shown that WDM halos with a core-halo distribution can exist[14] (and are indeed stable[15]) as end results of the process of violent relaxation, which fit observations from both Milky Way and dwarf galaxies. While these collisionless relaxation processes are thought to result in the formation and thermalization of these halos,[16,17] it has been argued in[18] that the process of Self Interactions can contribute to this end.

A particularly interesting particle candidate that can belong in the family of WDM models can be found in νMSM (Neutrino Minimal Standard model)[12,19] which assumes a light sterile neutrino to be the sole DM component, produced out of equilibrium via neutrino oscillations in the plasma, and requiring only a minimal extension to the Standard Model Lagrangian. Several observations heavily constrain the parameter space of this model, such as X-Ray constraints from astrophysical objects,[20–22] structure formation constraints,[23,24] production bounds[25]), among others. In recent years, Lyman-α observations have been particularly important, almost entirely ruling out the non-thermal production of WDM sterile neutrinos.[24,26] These tight constraints may be alleviated by performing simple extensions to this model: in particular, we focus here on the inclusion of DM Self Interactions, in what we name Self Interacting Warm Dark Matter (SIWDM).[27] As we will summarize in this work, these may not only contribute in the aforementioned challenges, but also allow us to potentially address tensions in ΛCDM previously unaccounted by WDM models alone.

2. Self Interactions and WDM

The addition of DM Self Interactions is a well studied problem, with many realizations mostly based on CDM models.[28,29] Its main assumption is the existence of non gravitational interaction terms between DM particles through a Standard Model or a Dark mediator. Typically, in the realm of structure formation the interactions considered are elastic, and therefore do not change the number or identity of particles. We focus here on interaction channels of these kinds in WDM and characterize their possible effects across cosmological history.

Self Interactions have been invoked in N-body simulations of structure formation, where it has been realized that the inclusion of these interactions can "flatten"

the inner cores of dwarf galaxies (as indicated by observations) via a process of thermalization in the inner regions of these halos.[30,31] Thus, a combined model of WDM with the addition of Self Interactions may overcome both structure formation ("missing satellites" and "too big to fail") as well as inner halo dynamics ("core-cusp") challenges at the same time, preserving the benefits of both WDM and Self Interacting models. While there can be many particle physics models that reflect these interactions, we focus here instead on maintaining a certain model independence: some examples of particular model realizations can be found in the literature.[32–35]

Since the first introduction of Self Interactions in WDM,[36] it was realized that a combined model could provide interesting modifications in their dynamics. Indeed, Self Interactions do not only have an effect during structure formation, but also in perturbation evolution[37,38] and DM production.[12,39] This last point was studied in the past in the context of νMSM and, in particular, as mentioned in reference,[40] the phenomenon of mediator decay can significantly relax the parameter space of the model. There it was considered the case of a WDM model with vector field interactions, previously suggested in the literature to provide interesting consequences for DM halos.[41] These studies focus on WDM core-halo distributions,[14] known in the literature as RAR profiles, extended it via a Self Interacting model and calculated its potential effects on νMSM bounds coming from X-Ray observations. Interestingly, although bounds of the same order of magnitude as previous studies (that use N-body simulation-borne DM distributions) were obtained when considering observations of the MW galactic center, it was shown that these core-halo distributions are only compatible with νMSM bounds with the inclusion of additional production channels, such as Self Interactions. We present these results here in figure 1, where we plot the parameter space for νMSM, together with the bounds obtained from X-Ray observations of the MW Galactic Center.

While it was shown in reference[40] that the DM distribution of these halos is not affected by Self Interactions, it is an interesting open question if the formation of these systems may be affected by the inclusion of DM scattering. Indeed, these profiles are constructed assuming (General Relativistic) thermal equilibrium between DM particles. While coarse-grained equilibrium can be achieved by a collisionaless fluid, as was shown in previous works,[14,17,42] Self Interactions can thermalize the inner regions of these systems[30] and, potentially, lead to the formation of a fermionic core.

3. SIWDM and Linear Cosmology

Self Interactions have already been found to have significant effects in the evolution of linear cosmological perturbations.[36–38] In particular, the effects of these interactions can be realized in the power spectrum and, in the realm of WDM, it

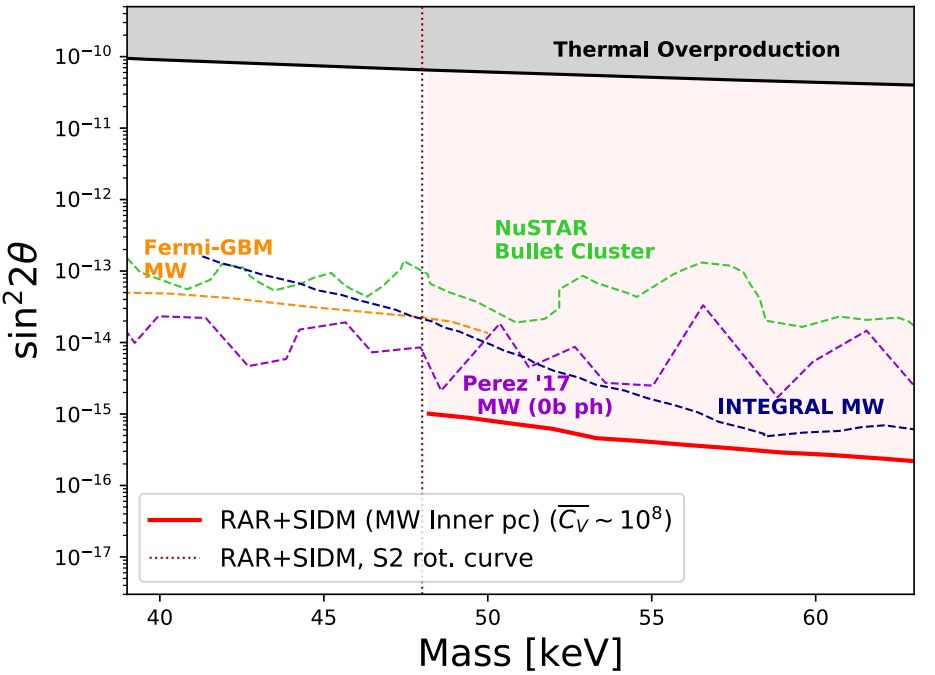


Fig. 1. Sterile neutrino parameter space limits obtained for MW GC observations using RAR+SIDM profiles (continuous red line), when assuming DM production due to self interactions and interaction strength according to Bullet Cluster constraints. The light Red shaded region above the continuous red line corresponds to RAR+SIDM limits given by X-ray bounds (i.e. indirect detection analysis), while the vertical shaded region below 48 keV labels the smallest DM mass compatible with S-cluster stars' rotation curve data that can provide a BH alternative. The upper shaded region corresponds to production mechanism bounds, while other dotted lines refer to several X-ray bounds corresponding to different observations/DM profiles.[20] Reproduced from reference.[40]

can lead to interesting scenarios such as non relativistic self decoupling (a.k.a. late kinetic decoupling). Particularly, the first challenge consists in implementing correctly a collision term in the first order Boltzmann equation for massive species.[43] While many implementations involve performing fluid approximations for the DM component,[44,45] it was realized that this approach is inaccurate in the case of light relics, such as SM interacting neutrinos.[43,46] Based on this realization, some of us obtained a reduced, kernel-based form for the Boltzmann collision term,[27] based on an ansatz for the interaction amplitude that encompasses most tree-level massive mediator models in an exact way. We refer the reader to reference[27] for a full expression of these terms.

The system of equations derived in[27] allows us to study the evolution of DM perturbations under the Relaxation Time Approximation (RTA),[47] shown to be very precise for the case of massive mediator interactions.[46] We can express the Legendre

expansion of the first order Boltzmann equation in the synchronous gauge as

$$\begin{aligned}
\dot{F}_0(k, E_q, \tau) &\simeq -\frac{qk}{E_q} F_1(k, E_q, \tau) + \frac{\dot{h}}{6} \frac{\partial f_0}{\partial \ln q} \\
\dot{F}_1(k, E_q, \tau) &\simeq \frac{qk}{3E_q} F_0(k, E_q, \tau) - \frac{2qk}{3E_q} F_2(k, E_q, \tau) \\
\dot{F}_2(k, E_q, \tau) &\simeq \frac{qk}{5E_q} \Big[2F_1(k, E_q, \tau) - 3F_3(k, E_q, \tau) \Big] - \frac{\partial f_0}{\partial \ln q} \left[\frac{1}{15} \dot{h} + \frac{2}{5} \dot{\eta} \right] - a \frac{F_2(k, E_q, \tau)}{\tau_{rel}} \\
\dot{F}_l(k, E_q, \tau) &\simeq \frac{qk}{(2l+1)E_q} \Big[l F_{(l-1)}(k, E_q, \tau) - (l+1) F_{(l+1)}(k, E_q, \tau) \Big] \\
&\quad - a \frac{F_l(k, E_q, \tau)}{\tau_{rel}} \quad l \geq 3,
\end{aligned} \tag{1}$$

where $f(\vec{k}, \vec{q}, \tau) = F(\vec{k}, \vec{q}, \tau) + f_0(q, \tau)$ is the DM distribution, $F(\vec{k}, \vec{q}, \tau)$ is the first order perturbation and we have defined the Legendre moments and the gravitational potentials h, η as in reference.[48] The collision rate $\Gamma(\tau)$ is defined as

$$\tau_{rel}^{-1} = \frac{g_i^3}{32(2\pi)^3} \frac{\int dE_q \, dE_l \, ds \, f_{eq}(E_q, \tau) \, f_{eq}(E_l, \tau) \, \chi(s)}{\int dE_q \, q \, E_q \, f_{eq}(E_q, \tau)} \ , \tag{2}$$

where this expression is defined according to the notation in references.[18,27] In the same references particular expressions for the collision kernel χ for various interaction models can be found.

These equations involve the background DM distribution f_0. While in the standard approach, this function remains in a relativistic form reminiscent of WDM, a few circumstances can alter this assumption. In particular, if the Self Interactions maintain the DM fluid in kinetic equilibrium all the way until the fluid becomes non relativistic, the background at that moment will switch into a non relativistic form, constituting the scenario known as Non Relativistic Self Decoupling. The consequences of this scenario were explored by some of us in reference.[18] There, it was found that, if one imposes continuity of the limiting expressions for the energy density, the non relativistic distribution function becomes

$$f_0(T < m) \sim 4.534 \mathcal{C}_{\rm NR} \exp\left[-1.075 q^2 / T_{0,\rm R}^2 \right] \ , \tag{3}$$

where $\mathcal{C}_{\rm NR}, T_{0,\rm R}$ are the normalization and temperature of the species in the relativistic limit.

4. SIWDM Cosmology: Simulations and Observation

In reference[18] some of us developed a numerical implementation of these Self Interacting WDM models using a modified CLASS code, a publicly available Boltzmann solver.[49] This modification is available at the following link github.com/yunis121/siwdm-class. We show there a few examples of the resulting power spectra for the Self Interacting models in figure 2.

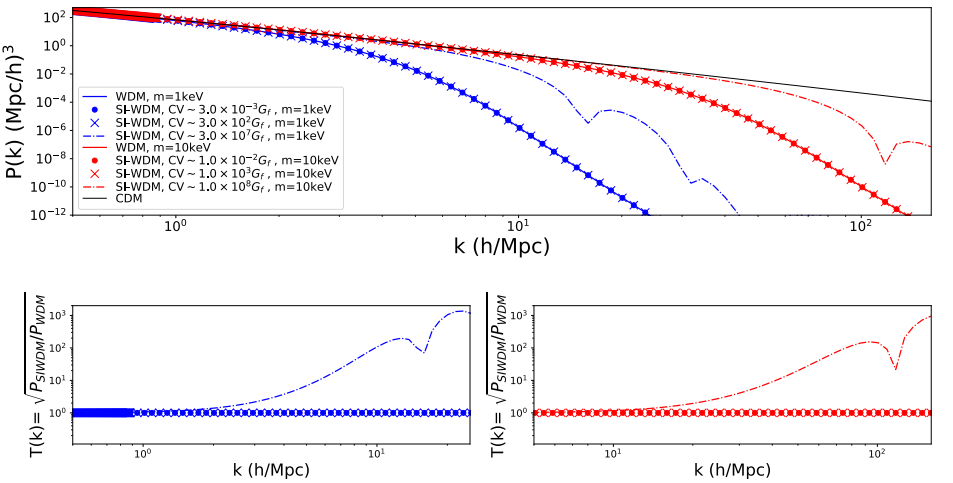


Fig. 2. Power Spectrum (*top panel*) and Transfer Functions with respect to standard WDM (*bottom panels*) for a vector field SI-WDM model, simulated using a modification to CLASS. We assume the relaxation time approximation (1), and consider two values of the DM particle mass: 1 and 10 keV. Also plotted are the power spectra of a CDM model and of WDM models with DM mass of 1 and 10 keV. All WDM and SI-WDM models consider a nonresonant production scenario (Dodelson-Widrow mechanism,[50]) with $T \sim (4/11)^{1/3} T_\gamma$. Reproduced from reference[18]

There, we see some of the particular features of the models. We see that the inclusion of Self Interactions, while modifying slightly the high k behavior of the power spectra, remain small elsewhere in the power spectrum for all models that do not undergo Non Relativistic Self Decoupling. However, for models with Non Relativistic Self Decoupling, the resulting power spectra may differ significantly from its relativistic counterpart. Indeed, we find that in this regime the models are "colder" (i. e. as if they correspond to a higher particle mass), and show even at smaller k values a distinctive oscillatory pattern. This indeed has the effect of increasing the amount of small structure formed for these models.

Most of the tensions inherent to νMSM WDM models come from structure formation, namely MW satellite counts and Lyman-α observations. These are related to the fact that the preferred parameter ranges may underproduce small structure and almost rule out the available parameter space. So, the inclusion of Self Interactions can significantly relax the existing bounds on this family of models.

Reference[18] provides an evaluation of the predictions of these models for the number of MW satellites[23] as well for the observations of the Lyman-α forest. We have found that a maximally interacting model (as allowed by the observations of the Bullet Cluster[41]) can readmit a significant portion of the νMSM parameter space, and we illustrate this in figure 3.

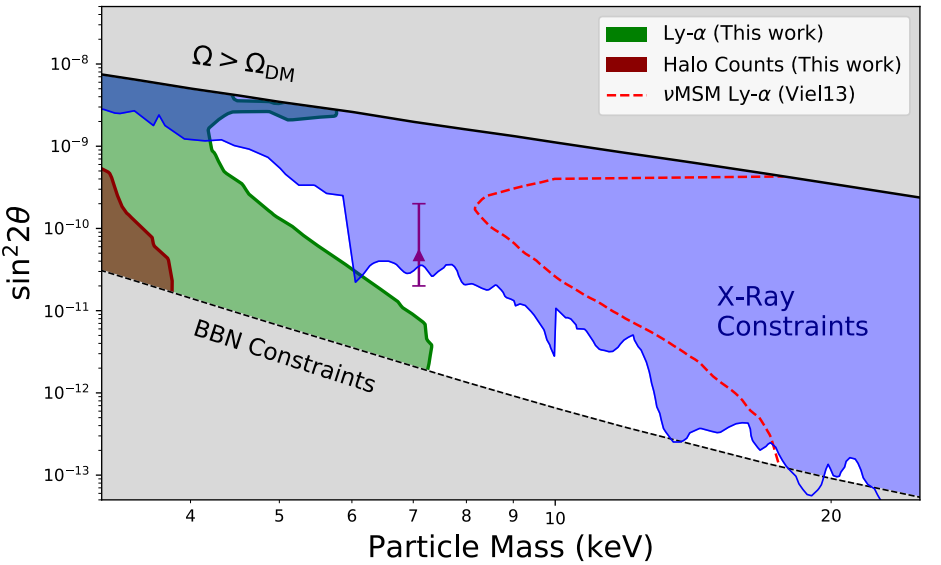


Fig. 3. Parameter space constraints for νMSM, where MW satellite halo counts and Lyman-α forest bounds are analyzed under a self interacting model as outlined above. We consider a self interacting model under a vector field mediator, with its interaction constant given by $\sigma/m \sim 0.144 C_v^2/m^3 = 0.1\text{cm}^2/g$, the upper limit given by Bullet Cluster constraints.[41] For comparison, we plot the Lyman-alpha bounds for the non interacting case for a comparable analysis, according to the results in.[26] We also plot other bounds to the νMSM parameter space for informative purposes.[12,21–23,25] Also for informative purposes, we plot a model compatible with a tentative 3.5 keV DM signal as a purple triangle[51–53] for informative purposes. Reproduced from reference[18]

5. Conclusions

We have studied in detail the evolution of Warm Dark Matter in the presence of Self Interactions. With the aim of contributing to a detailed study of this subject from both a cosmological and astrophysical perspective we provide here an overview of the results obtained so far.

Reference[40] presents and exploration of the consequences of considering WDM core halo RAR distributions[14] in the parameter space of sterile neutrino WDM. There it is suggested that the inclusion of Self Interactions can at the same time relax the WDM sterile neutrino parameter space via additional production channels, readmit RAR core-halo distributions into these models and assist with relaxation and thermalization of these systems.

Reference[27] provides a theoretical framework for the treatment of cosmological perturbations in WDM that allows to quantify the effects of these extended models in the cosmological perturbations. There, we developed a kernel-based expression for the collision terms based on an ansatz for the interaction amplitude, as well as

considered a few approximations to the resulting Boltzmann hierarchies such as the Relaxation Time Approximation. Reference[18] presents a numerical application to these models in CLASS and a comparison of the resulting spectra with MW subhalo and Lyman-α observables, both in the standard decoupling scenario as well as in the non relativistic case.

We have reached conclusions that indicate the Self Interacting models may provide an interesting extension of the sterile neutrino WDM models. In the future, we hope to contribute further in the development of a more accurate treatment of these Self Interactions at a Cosmological level, as well as complementary models for early universe histories and, possibly, a deeper insight into the formation of these core-halo models. While further research is needed if this model is to be considered a viable alternative to standard DM approaches, we believe the work summarized here may be an important stepping stone in the study of these extensions.

Acknowledgments

CRA has been supported by CONICET, Secretary of Science and Technology of FCAG and UNLP (grants G140 and G175), National Agency for the Promotion of Science and Technology (ANPCyT) of Argentina (grant PICT-2018-03743) and ICRANet. CGS is supported by ANPCyT grant PICT-2016-0081; and grants G140, G157 and G175 from UNLP. DLN has been supported by CONICET, ANPCyT and UBA. The work of NEM is supported in part by the UK Science and Technology Facilities research Council (STFC) under the research grant ST/T000759/1. NEM also acknowledges participation in the COST Association Action CA18108 "*Quantum Gravity Phenomenology in the Multimessenger Approach (QG-MM)*".

References

1. N. A. Bahcall, J. P. Ostriker, S. Perlmutter and P. J. Steinhardt, The Cosmic triangle: Assessing the state of the universe, *Science* **284**, 1481 (1999).
2. P. Collaboration, Planck 2018 results: VI. Cosmological parameters, *Astronomy and Astrophysics* **641**, p. A6 (sep 2020).
3. J. S. Bullock and M. Boylan-Kolchin, Small-Scale Challenges to the ΛcDM Paradigm, *Annual Review of Astronomy and Astrophysics* **55**, 343 (2017).
4. B. F. Griffen, A. P. Ji, G. A. Dooley, F. A. Gómez, M. Vogelsberger, B. W. O'Shea and A. Frebel, The Caterpillar Project: A Large Suite of Milky Way Sized Halos, *The Astrophysical Journal* **818**, p. 10 (2016).
5. W. Janesh, K. L. Rhode, J. J. Salzer, S. Janowiecki, E. A. K. Adams, M. P. Haynes, R. Giovanelli and J. M. Cannon, Five Gas-rich Ultrafaint Dwarf Galaxy Candidates Discovered in WIYN Imaging of ALFALFA Sources, *The Astronomical Journal* **157**, p. 183 (2019).
6. The DES Collaboration, A. Drlica-Wagner, K. Bechtol, E. S. Rykoff, E. Luque, A. Queiroz, Y. Y. Mao, R. H. Wechsler, J. D. Simon, B. Santiago, B. Yanny, E. Balbinot, S. Dodelson, A. F. Neto, D. J. James, T. S. Li, M. A. G. Maia, J. L. Marshall, A. Pieres, K. Stringer, A. R. Walker, T. M. C. Abbott, F. B. Abdalla, S. Allam,

A. Benoit-Levy, G. M. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. C. Rosell, M. C. Kind, J. Carretero, M. Crocce, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. E. Evrard, D. A. Finley, B. Flaugher, P. Fosalba, J. Frieman, E. Gaztanaga, D. W. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, K. Kuehn, N. Kuropatkin, O. Lahav, P. Martini, R. Miquel, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Tucker, V. Vikram, W. Wester, Y. Zhang, J. Zuntz and Others, Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey, *Astrophys. J.* **813**, p. 109 (aug 2015).

7. M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, Too big to fail? The puzzling darkness of massive Milky Way subhaloes, *Monthly Notices of the Royal Astronomical Society: Letters* **415**, 1 (feb 2011).
8. S.-H. Oh, D. A. Hunter, E. Brinks, B. G. Elmegreen, A. Schruba, F. Walter, M. P. Rupen, L. M. Young, C. E. Simpson, M. Johnson, K. A. Herrmann, D. Ficut-Vicas, P. Cigan, V. Heesen, T. Ashley and H.-X. Zhang, High-resolution mass models of dwarf galaxies from LITTLE THINGS, *Astronomical Journal* **149** (feb 2015).
9. J. F. Navarro, C. S. Frenk and S. D. M. White, A Universal density profile from hierarchical clustering, *Astrophys.J.* **490**, 493 (1997).
10. P. Bode, J. P. Ostriker and N. Turok, Halo Formation in Warm Dark Matter Models, *The Astrophysical Journal* **556**, 93 (2001).
11. M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky and O. Ruchayskiy, The haloes of bright satellite galaxies in a warm dark matter universe, *Monthly Notices of the Royal Astronomical Society* **420**, 2318 (2012).
12. A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens and O. Ruchayskiy, Sterile Neutrino Dark Matter, *Prog. Part. Nucl. Phys.* **104**, 1 (2019).
13. A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider and B. Moore, Cores in warm dark matter haloes: a Catch 22 problem, *Monthly Notices of the Royal Astronomical Society* **424**, 1105 (2012).
14. C. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, Novel constraints on fermionic dark matter from galactic observables II: galaxy scaling relations, *Physics of the Dark Universe* **24**, p. 100278 (2019).
15. C. R. Argüelles, M. I. Díaz, A. Krut and R. Yunis, On the formation and stability of fermionic dark matter halos in a cosmological framework, *Monthly Notices of the Royal Astronomical Society* (dec 2020).
16. D. Lynden-Bell, Statistical mechanics of violent relaxation in stellar systems, *Monthly Notices of the Royal Astronomical Society* **136**, p. 101 (1967).
17. P.-H. Chavanis, Statistical mechanics of self-gravitating systems in general relativity: I. The quantum Fermi gas, *Eur. Phys. J. Plus* **135** (aug 2020).
18. R. Yunis, C. R. Argüelles, C. G. Scóccola, D. L. Nacir and G. Giordano, Self-Interacting Dark Matter in Cosmology: accurate numerical implementation and observational constraints (aug 2021).
19. R. Adhikari, M. Agostini, N. Ky, T. Araki, M. Archidiacono, M. Bahr, J. Baur, J. Behrens, F. Bezrukov, P. Bhupal Dev, D. Borah, A. Boyarsky, A. de Gouvea, C. Pires, H. de Vega, A. Dias, P. Di Bari, Z. Djurcic, K. Dolde, H. Dorrer, M. Durero, O. Dragoun, M. Drewes, G. Drexlin, C. Düllmann, K. Eberhardt, S. Eliseev, C. Enss, N. Evans, A. Faessler, P. Filianin, V. Fischer, A. Fleischmann, J. Formaggio, J. Franse,

F. Fraenkle, C. Frenk, G. Fuller, L. Gastaldo, A. Garzilli, C. Giunti, F. Glück, M. Goodman, M. Gonzalez-Garcia, D. Gorbunov, J. Hamann, V. Hannen, S. Hannestad, S. Hansen, C. Hassel, J. Heeck, F. Hofmann, T. Houdy, A. Huber, D. Iakubovskyi, A. Ianni, A. Ibarra, R. Jacobsson, T. Jeltema, J. Jochum, S. Kempf, T. Kieck, M. Korzeczek, V. Kornoukhov, T. Lachenmaier, M. Laine, P. Langacker, T. Lasserre, J. Lesgourgues, D. Lhuillier, Y. Li, W. Liao, A. Long, M. Maltoni, G. Mangano, N. Mavromatos, N. Menci, A. Merle, S. Mertens, A. Mirizzi, B. Monreal, A. Nozik, A. Neronov, V. Niro, Y. Novikov, L. Oberauer, E. Otten, N. Palanque-Delabrouille, M. Pallavicini, V. Pantuev, E. Papastergis, S. Parke, S. Pascoli, S. Pastor, A. Patwardhan, A. Pilaftsis, D. Radford, P.-O. Ranitzsch, O. Rest, D. Robinson, P. Rodrigues da Silva, O. Ruchayskiy, N. Sanchez, M. Sasaki, N. Saviano, A. Schneider, F. Schneider, T. Schwetz, S. Schönert, S. Scholl, F. Shankar, R. Shrock, N. Steinbrink, L. Strigari, F. Suekane, B. Suerfu, R. Takahashi, N. Van, I. Tkachev, M. Totzauer, Y. Tsai, C. Tully, K. Valerius, J. Valle, D. Venos, M. Viel, M. Vivier, M. Wang, C. Weinheimer, K. Wendt, L. Winslow, J. Wolf, M. Wurm, Z. Xing, S. Zhou, K. Zuber and Others, A White Paper on keV Sterile Neutrino Dark Matter, *JCAP* **1701**, p. 25 (2017).

20. K. Perez, K. C. Y. Ng, J. F. Beacom, C. Hersh, S. Horiuchi and R. Krivonos, Almost closing the νMSM sterile neutrino dark matter window with NuSTAR, *Phys. Rev.* **D95**, p. 123002 (2017).
21. J. F. Cherry and S. Horiuchi, Closing in on Resonantly Produced Sterile Neutrino Dark Matter, *Physical Review D* **D95**, p. 83015 (jan 2017).
22. K. C. Y. Ng, B. M. Roach, K. Perez, J. F. Beacom, S. Horiuchi, R. Krivonos and D. R. Wik, New Constraints on Sterile Neutrino Dark Matter from *NuSTAR* M31 Observations, *Physical Review D* **99** (jan 2019).
23. A. Schneider, Astrophysical constraints on resonantly produced sterile neutrino dark matter, *Journal of Cosmology and Astroparticle Physics* **1604**, p. 59 (jan 2016).
24. W. Enzi, R. Murgia, O. Newton, S. Vegetti, C. Frenk, M. Viel, M. Cautun, C. D. Fassnacht, M. Auger, G. Despali, J. McKean, L. V. E. Koopmans and M. Lovell, Joint constraints on thermal relic dark matter from a selection of astrophysical probes, *MNRAS* **000**, 1 (oct 2020).
25. T. Venumadhav, F.-Y. Y. Cyr-Racine, K. N. Abazajian and C. M. Hirata, Sterile neutrino dark matter: Weak interactions in the strong coupling epoch, *Physical Review D* **94**, 1 (2016).
26. M. Viel, G. Becker, J. Bolton and M. Haehnelt, Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-alpha forest data, *Phys. Rev. D* **88**, p. 43502 (2013).
27. R. Yunis, C. R. Argüelles and D. L. Nacir, Boltzmann hierarchies for self-interacting warm dark matter scenarios, *Journal of Cosmology and Astroparticle Physics* **2020** (feb 2020).
28. M. Vogelsberger, J. Zavala, F.-Y. Y. Cyr-Racine, C. Pfrommer, T. Bringmann and K. Sigurdson, ETHOS - an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems, *Mon. Not. Roy. Astron. Soc.* **460**, 1399 (2016).
29. S. Tulin and H.-B. Yu, Dark Matter Self-interactions and Small Scale Structure, *Phys. Rept.* **730**, 1 (2018).
30. J. S. Almeida and I. Trujillo, Numerical simulations of dark matter haloes produce polytropic central cores when reaching thermodynamic equilibrium, *Monthly Notices of the Royal Astronomical Society* **504**, 2832 (apr 2021).
31. S. Balberg, S. L. Shapiro and S. Inagaki, Self-Interacting Dark Matter Halos and the Gravothermal Catastrophe, *The Astrophysical Journal* **568**, 475 (oct 2001).

32. M. Y. Khlopov, Fundamental Particle Structure in the Cosmological Dark Matter, *International Journal of Modern Physics A* **28** (nov 2013).
33. S. Tulin, H.-B. Yu and K. M. Zurek, Resonant Dark Forces and Small Scale Structure, *Physical Review Letters* **110** (oct 2012).
34. L. G. van den Aarssen, T. Bringmann and C. Pfrommer, Is dark matter with long-range interactions a solution to all small-scale problems of Λ CDM cosmology?, *Physical Review Letters* **109** (may 2012).
35. M. A. Buen-Abad, G. Marques-Tavares and M. Schmaltz, Non-Abelian dark matter and dark radiation, *Physical Review D* **92** (2015).
36. S. Hannestad and R. J. Scherrer, Selfinteracting warm dark matter, *Phys. Rev.* **D62**, p. 43522 (2000).
37. R. Huo, Matter Power Spectrum of Light Freeze-in Dark Matter: With or without Self-Interaction, *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **802** (jul 2019).
38. D. Egana-Ugrinovic, R. Essig, D. Gift and M. LoVerde, The Cosmological Evolution of Self-interacting Dark Matter, *Journal of Cosmology and Astroparticle Physics* **2021** (feb 2021).
39. A. De Gouvêa, M. Sen, W. Tangarife and Y. Zhang, Dodelson-Widrow Mechanism in the Presence of Self-Interacting Neutrinos, *Phys. Rev. Lett.* **124**, p. 81802 (2020).
40. R. Yunis, C. R. Argüelles, N. E. Mavromatos, A. Moliné, A. Krut, M. Carinci, J. A. Rueda and R. Ruffini, Galactic Center constraints on self-interacting sterile neutrinos from fermionic dark matter ("ino") models, *Physics of the Dark Universe* **30** (aug 2020).
41. C. R. Argüelles, N. E. Mavromatos, J. A. Rueda and R. Ruffini, The role of self-interacting right-handed neutrinos in galactic structure, *Journal of Cosmology and Astroparticle Physics* **1604**, p. 38 (2016).
42. P.-H. Chavanis, On the 'coarse-grained' evolution of collisionless stellar systems, *Monthly Notices of the Royal Astronomical Society* **300**, 981 (1998).
43. I. M. Oldengott, C. Rampf and Y. Y. Y. Wong, Boltzmann hierarchy for interacting neutrinos I: formalism, *JCAP* **1504**, p. 16 (2015).
44. S. Heimersheim, N. Schöneberg, D. C. Hooper and J. Lesgourgues, Cannibalism hinders growth: Cannibal Dark Matter and the S_8 tension, *Journal of Cosmology and Astroparticle Physics* **2020** (aug 2020).
45. M. Garny, T. Konstandin, L. Sagunski and S. Tulin, Lyman-α forest constraints on interacting dark sectors, *Journal of Cosmology and Astroparticle Physics* **2018** (may 2018).
46. I. M. Oldengott, T. Tram, C. Rampf and Y. Y. Y. Wong, Interacting neutrinos in cosmology: exact description and constraints, *JCAP* **1711**, p. 27 (2017).
47. P. L. Krapivsky, S. Redner and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, 2010).
48. C.-P. Ma and E. Bertschinger, Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges, *The Astrophysical Journal* **455**, p. 7 (jun 1995).
49. J. Lesgourgues and T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics, *Journal of Cosmology and Astroparticle Physics* **2011**, p. 32 (apr 2011).
50. S. Dodelson and L. M. Widrow, Sterile Neutrinos as Dark Matter, *Physical Review Letters* **72**, 17 (mar 1993).
51. E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein and S. W. Randall, Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters, *Astrophys. J.* **789**, p. 13 (2014).

52. T. E. Jeltema and S. Profumo, Discovery of a 3.5 keV line in the Galactic Centre and a critical look at the origin of the line across astronomical targets, *Mon. Not. Roy. Astron. Soc.* **450**, 2143 (2015).
53. N. Cappelluti, E. Bulbul, A. Foster, P. Natarajan, M. C. Urry, M. W. Bautz, F. Civano, E. Miller and R. K. Smith, Searching for the 3.5 keV Line in the Deep Fields with Chandra : The 10 Ms Observations, *The Astrophysical Journal* **854**, p. 179 (feb 2018).

Interaction energy between a charged medium and its electromagnetic field as a dark matter candidate

Mayeul Arminjon

Lab. 3SR (Grenoble-Alpes University, Grenoble-INP, CNRS), Grenoble, 38041 cedex 9, France
E-mail: Mayeul.Arminjon@3sr-grenoble.fr
https://www.univ-grenoble-alpes.fr/english/

In the scalar theory of gravitation with a preferred reference frame, a consistent formulation of electrodynamics in the presence of gravitation needs to introduce an additional energy tensor: the interaction energy tensor. This energy is gravitationally active and might contribute to the dark matter, because it has an exotic character and it is not localized inside matter. In order to check if that energy might form representative dark halos, one has to model the interstellar radiation field in a galaxy as a complete electromagnetic field obeying the Maxwell equations. A model has been built for this purpose, based on assuming axial symmetry and on recent results about axisymmetric Maxwell fields. Its predictions for the variation of the spectral energy density inside our Galaxy are relatively close to those of a recent radiation transfer model, except on the symmetry axis of the Galaxy, where the present model predicts extremely high values of the energy density.

Keywords: Dark matter; interstellar radiation field; Maxwell equations; Milky Way.

1. Introduction

Our initial motivation for the present work was independent of the problem of dark matter. It was to develop a consistent electrodynamics in an alternative theory of gravity: "the scalar ether theory", or SET. This is a preferred-frame theory based on a scalar field only,[1,2] that reduces to special relativity (SR) when the gravitational field vanishes. In general relativity (GR), the modification of the equations of electrodynamics in the presence of a gravitational field consists simply in rewriting the equations that are valid in SR, by using the "comma goes to semicolon" rule: $_{,\nu} \rightarrow {}_{;\nu}$, i.e.: partial derivatives are replaced by covariant derivatives based on the metric connection. (See Ref. 3 for an interesting discussion.) In particular, the dynamical equation for the energy(-momentum-stress) tensor $\boldsymbol{T}$ that is valid in SR is: $T^{\lambda\nu}_{\ \ ,\nu} = 0$. Using the rule mentioned above, that equation is modified to: $T^{\lambda\nu}_{\ \ ;\nu} = 0$, which is indeed the dynamical equation in GR and in many of its extensions or modifications. However, in the general situation, the latter equation is not equivalent to the dynamical equation of SET,[1] hence the foregoing rule cannot be used in SET.

Therefore, in that alternative theory, a different and less obvious path has to be taken for the purpose of adaptating classical electrodynamics in the presence of a gravitational field. It turns out that this leads to introduce an exotic form of energy,

and that this new form is a possible candidate for dark matter. In this conference paper, we quickly follow that path. We then summarize the Maxwell model of the interstellar radiation field, that we built to prepare the test of this candidate.

2. Necessity of an interaction tensor in SET

In SET, we assume classically that the electromagnetic field tensor $\boldsymbol{F}$ derives from a 4-potential A_μ:

$$F_{\mu\nu} := A_{\nu,\mu} - A_{\mu,\nu} = A_{\nu;\mu} - A_{\mu;\nu}. \tag{1}$$

This is (locally) equivalent to assuming that (i) $\boldsymbol{F}$ is antisymmetric ($F_{\mu\nu} = -F_{\nu\mu}$) and (ii) the first group of the Maxwell equations is satisfied:

$$F_{\lambda\mu\,,\nu} + F_{\mu\nu,\lambda} + F_{\nu\lambda,\mu} \equiv F_{\lambda\mu\,;\nu} + F_{\mu\nu;\lambda} + F_{\nu\lambda;\mu} = 0. \tag{2}$$

(The first equality in (2) is indeed an identity due to the antisymmetry of the field tensor and to the symmetry of the metric connection.) Therefore, in SET, the first group of the Maxwell equations is left unchanged. In a first version of electrodynamics in the presence of a gravitational field in SET, the second group of the Maxwell equations was got by applying the dynamical equation of SET to a charged medium in the presence of the Lorentz force, assuming that the following holds for the energy tensors, as is the case in SR and still in GR:

$$\text{(A) Total energy tensor } \boldsymbol{T} = \boldsymbol{T}_{\text{charged medium}} + \boldsymbol{T}_{\text{field}}. \tag{3}$$

(The total energy tensor $\boldsymbol{T}$ is the source of the gravitational field — more precisely, in SET, that source is the component T^{00} in the preferred reference frame of the theory; see Ref. 1 for details.) The additivity (3) leads to a form of Maxwell's second group of equations in SET.[4] But that form of Maxwell's second group in SET predicts charge production/destruction at untenable rates, therefore it has to be *discarded*.[5] The additivity assumption (3) is contingent and may be abandoned. This means introducing an "interaction" energy tensor $\boldsymbol{T}_{\text{inter}}$, such that

$$\boldsymbol{T} = \boldsymbol{T}_{\text{charged medium}} + \boldsymbol{T}_{\text{field}} \underline{+\boldsymbol{T}_{\text{inter}}}\,. \tag{4}$$

One then has to constrain the form of $\boldsymbol{T}_{\text{inter}}$ and to derive equations for it.

3. Form of the interaction tensor

In SR, the additivity (3) of the energy tensors does apply, thus $\boldsymbol{T}_{\text{inter}} = \boldsymbol{0}$. In SET we may impose that $\boldsymbol{T}_{\text{inter}}$ should be Lorentz-invariant in the situation of SR, i.e. when the metric $\boldsymbol{\gamma}$ is Minkowski's metric $\boldsymbol{\gamma}^0$ ($\gamma^0_{\mu\nu} = \eta_{\mu\nu}$ in Cartesian coordinates). This is true if and, one can prove,[6] *only if* we have:

$$T_{\text{inter}\ \mu\nu} = p\,\gamma^0_{\mu\nu} \qquad \text{(situation of SR)}, \tag{5}$$

with some scalar field p. This is equivalent to:

$$T^{\mu}_{\text{inter}\ \ \nu} = p\,\delta^\mu_\nu \qquad \text{(situation of SR)}. \tag{6}$$

The definition

$$T^{\mu}_{\text{inter}\ \ \nu} := p\,\delta^{\mu}_{\nu}, \qquad \text{or} \quad (T_{\text{inter}})_{\mu\nu} := p\,\gamma_{\mu\nu}, \tag{7}$$

thus got in a Minkowski spacetime, is in fact generally-covariant. Hence, we adopt (7) for the general case. With a general metric $\boldsymbol{\gamma}$, the tensor (7) is still pointwise Lorentz-invariant — in the sense that we have $(T_{\text{inter}})_{\mu\nu}(X) = p(X)\,\eta_{\mu\nu}$ in any coordinates that are Cartesian at a given event X, and this form remains invariant after any coordinate transformation that is Lorentz at X, i.e., such that the matrix $\left(\frac{\partial x'^{\mu}}{\partial x^{\nu}}(X)\right)$ belongs to the Lorentz group.

4. SET electrodynamics with the interaction tensor

With the additivity assumption (3) of the energy tensors, i.e., $\boldsymbol{T}_{\text{inter}} = \mathbf{0}$, the system of equations of electrodynamics of SET is closed, but violates charge conservation. With the interaction energy tensor (7) we have just one unknown more: the scalar field p. So we need just one scalar equation more. It turns out to be consistent to add *charge conservation* as the new scalar equation.[7] Then the system of equations of electrodynamics of SET is again closed, and now it satisfies charge conservation.

Based on that closed system, equations were derived that *determine the field p in a given general electromagnetic (EM) field $(\mathbf{E}, \mathbf{B})$ and in a given weak gravitational field with Newtonian potential U*:[7] the scalar field p (or more exactly, its first approximation p_1) obeys an advection equation:

$$\partial_T\, p_1 + u^j \partial_j\, p_1 = S. \tag{8}$$

That equation has given source S and given characteristic curves, the latter being the integral curves $\mathcal{C}(T_0, \mathbf{x}_0)$ of the spatial vector field $\mathbf{u}$ in Eq. (8). Here, "given" means that the source field S, as also the vector field $\mathbf{u}$ and hence the characteristic curves $\mathcal{C}(T_0, \mathbf{x}_0)$, do not depend on the unknown field p_1. It follows that p_1 can be obtained by integrating the source field S along those curves.[7]

The "medium" defined by the corresponding interaction energy tensor field $\boldsymbol{T}_{\text{inter}} = p\boldsymbol{\gamma}$ can be counted as "dark matter", for

- it is not localized inside (usual) matter: indeed, the equations for the field p_1 show that its source S is, in general, non-zero as soon as there is a general EM field: $\mathbf{E} \neq 0,\ \mathbf{B} \neq 0,\ \mathbf{E}.\mathbf{B} \neq 0$, and a variable gravitational field with $\partial_T U \neq 0$, where the time derivative $\partial_T U$ of the Newtonian potential is taken in the preferred frame;[7]
- it is gravitationally active, since, from its definition (4), it contributes to the source of the gravitational field in SET, that is the component T^{00} in the preferred frame;
- it is "exotic", i.e., it is not usual matter — as shown by the form (7) of its energy tensor, which is very different from the possible energy tensors of any fluid, solid, or EM field. The fact that it is Lorentz-invariant means

that no velocity can be defined for that medium. The energy tensor (7) depends only on one scalar field (p), hence no equation of state is needed.

The foregoing considerations are at the classical level, hence do not tell if the "matter" with the energy tensor (7) is made of quantum particles.

5. Maxwell model of the ISRF

In order to check if the interaction energy E_{inter} might be distributed in the form of dark halos and contribute significantly to the dark matter distribution, we have to compute the field p for a model of a galaxy. This needs that we have a model of the Interstellar Radiation Field in a galaxy (ISRF) that provides that field as a solution of the Maxwell equations. However, the existing models of the ISRF (e.g. Refs. 8–13) focus on the radiation transfer (mainly via absorption, reemission or scattering by dust particles). They follow the paths of light rays or photons. To the best of our knowledge, no previous model of the ISRF did consider the full EM field with its six interacting components subjected to the Maxwell equations. Therefore, we had to build a model entirely from scratch, which involved both theoretical and numerical difficulties.[17]

5.1. *Maxwell model of the ISRF: Main assumptions*

i) *Axial symmetry* is a relevant approximation for many galaxies, and is in fact often used in the existing models of the ISRF (see e.g. Refs. 13–16). We adopt cylindrical coordinates (ρ, ϕ, z) whose the z axis is the symmetry axis.

The primary source of the ISRF is made of the stars or other bright astrophysical objects. We want to describe the ISRF as a smoothed-out field at the galactic scale, not the field in the stars or in their neighborhood. Therefore:

ii) we consider the *source-free* Maxwell equations.

We proved the following result:[18]

Theorem. Any time-harmonic axisymmetric source-free Maxwell field is the sum of two simple fields of that same kind:

- **1**) one deriving from a vector potential $\mathbf{A}$ having just $A_z \neq 0$, with A_z a time-harmonic axisymmetric solution of the scalar wave equation;
- **2**) one deduced from a field of the form (**1**) by EM duality, i.e.

$$\mathbf{E}' = c\mathbf{B}, \quad \mathbf{B}' = -\mathbf{E}/c. \tag{9}$$

5.2. *Maxwell model of the ISRF: Form of the model*

We consider an EM field having a finite set of frequencies (ω_j) $(j = 1, ..., N_\omega)$. That EM field is thus the sum of N_ω time-harmonic EM fields. Using the Theorem above,

each of them is generated by potentials A_{jz}, A'_{jz}. The scalar potential A_{jz}, for the field of the form (1) above with frequency ω_j, can be a priori any time-harmonic axisymmetric solution of the scalar wave equation having frequency ω_j.[18] However, in the relevant "totally propagating" case, such a solution can be written explicitly in terms of a spectrum function $S_j = S_j(k) \quad (-K_j \leq k \leq K_j, \quad K_j := \frac{\omega_j}{c})$: $A_{jz} = \psi_{\omega_j\, S_j}$, with[19]

$$\psi_{\omega_j\ S_j}\,(t,\rho,z) := e^{-\,\mathrm{i}\,\omega_j t}\int_{-K_j}^{K_j} J_0\left(\rho\sqrt{K_j^2-k^2}\right)\ e^{\mathrm{i}\,k\,z}\,S_j(k)\ \mathrm{d}\,k, \tag{10}$$

where J_0 is the Bessel function of the first kind and of order 0. The "dual" potential A'_{jz}, for the field of the form (2) above with frequency ω_j, has just the same form (10), with, in the general case, another spectrum function, say S'_j.

5.3. *Maxwell model of the ISRF: Model of a galaxy*

We model an axisymmetric galaxy as a finite set $\{\mathbf{x}_i\}$ of point-like "stars", the azimuthal distribution of which is uniform. That set of points is obtained by pseudo-random generation of their cylindrical coordinates ρ, ϕ, z with specific probability laws, ensuring that[17]

- the distribution of ρ and z is approximately that valid for the star distribution in the galaxy considered (in the numerical application, we took our Galaxy);
- the set $\{\mathbf{x}_i\}$ is approximately invariant under azimuthal rotations of any angle ϕ.

5.4. *Maxwell model of the ISRF: Determining the potentials*

To determine the potentials $A_{j\,z}$ and $A'_{j\,z}$ $(j = 1, ..., N_\omega)$ that generate the model ISRF (Subsect. 5.2), we are fitting to the form (10) a sum of spherical potentials emanating from the "stars" at points $\mathbf{x}_i$, thus determining the unknown spectrum functions S_j and S'_j.[17] For the purpose of this fitting, every point-like "star" is indeed assumed to contribute spherical scalar waves $\psi_{\mathbf{x}_i\,\omega_j}$ having the same frequencies ω_j as has the model ISRF, and whose emission center is the spatial position $\mathbf{x}_i$ of the star:

$$\psi_{\mathbf{x}_i\,\omega_j}\ (t,\mathbf{x}) := \psi_{\omega_j}\ (t,\mathbf{x}-\mathbf{x}_i) = \frac{e^{\mathrm{i}(K_j\,r_i-\omega_j t)}}{K_j\,r_i}. \tag{11}$$

Here $r_i := |\mathbf{x}-\mathbf{x}_i|$, $K_j := \frac{\omega_j}{c}$, and the function

$$\psi_{\omega_j}\ (t,\mathbf{x}) = \frac{e^{\mathrm{i}(K_j\,r-\omega_j t)}}{K_j\,r}, \qquad r := |\mathbf{x}| \tag{12}$$

is (up to an amplitude factor) the unique time-harmonic solution of the scalar wave equation, with frequency ω_j, that has spherical symmetry around $\mathbf{x} = \mathbf{0}$ and that

is an outgoing wave. Spherical symmetry is assumed in order to ensure that all of the directions starting from the star are equivalent, of course. Of course also, the uniqueness of the solution (12) means the uniqueness of the solution translated from $\mathbf{x} = \mathbf{0}$ to $\mathbf{x} = \mathbf{x}_i$: the function $\psi_{\mathbf{x}_i\,\omega_j}$ given by Eq. (11). This implies that we cannot define different contributions of the "star" at $\mathbf{x}_i$ to the $A_{j\,z}$ potential and to the "dual" potential $A'_{j\,z}$, other than through multiplying $\psi_{\mathbf{x}_i\,\omega_j}$ by two different amplitude factors — for which there is no apparent reason. Therefore, we actually must assume that $A_{j\,z} = A'_{j\,z}$, thus $S_j = S'_j$, and our fitting problem writes

$$\sum_{i=1}^{i_{\max}} \psi_{\mathbf{x}_i\,\omega_j} \cong \psi_{\omega_j\ S_j} \quad \text{on } G \qquad (j = 1, ..., N_\omega). \tag{13}$$

Here the symbol $\cong$ indicates that the equality is in the sense of the least-squares, the two sides being evaluated on some spatio-temporal grid G. The unknown spectrum function S_j is defined (approximately) by its values $S_{nj} := S_j(k_{nj})$ at a regular discretization $k_{nj} = -K_j + n\frac{2K_j}{N}$ $(n = 0, ..., N)$ of the integration interval $[-K_j, +K_j]$ for k in the integral (10).[17] With this discretization, (13) becomes the explicit least-squares system

$$\sum_{i=1}^{i_{\max}} \psi_{\mathbf{x}_i\,\omega_j} \cong \sum_{n=0}^{N} f_{nj}\, S_{nj} \quad \text{on } G \qquad (j = 1, ..., N_\omega), \tag{14}$$

with $f_{nj}(t, \rho, z) = \exp(-\,\mathrm{i}\,\omega_j t)\, g_{nj}(\rho, z)$ a specific time-harmonic function.[20] The complex numbers S_{nj} $(n = 0, ..., N;\ j = 1, ..., N_\omega)$ are the solved-for parameters. Note that (14) defines N_ω fitting problems. Previously, a unique "grouped fitting" was done: solving the least-squares system obtained by summing on the frequency index j on both sides of (14).[17] The "separate fitting" (14) is more precise — and also less time-consuming, since actually the common harmonic time dependence can be removed from both sides of (14), thus eliminating the time variable, and hence considering only a spatial grid G' instead of a spatio-temporal grid G.[20] The computer time is indeed an important point to be considered, because a precision better than quadruple must be implemented.[17]

5.5. *Application to the spatial variation of the spectral energy density in the Galaxy*

Because we consider an EM field with a finite frequency spectrum (ω_j) $(j = 1, ..., N_\omega)$, each among its six components has the following form:

$$F^{(q)}(t, \mathbf{x}) = \mathcal{R}e\left(\sum_{j=1}^{N_\omega} C_j^{(q)}(\mathbf{x}) e^{-\,\mathrm{i}\,\omega_j t}\right) \qquad (q = 1, ..., 6). \tag{15}$$

It follows that the time-averaged volumic energy density of the field is given by:[21]

$$\overline{U}(\mathbf{x}) := \overline{\frac{\delta W}{\delta V}}(\mathbf{x}) = \sum_{j=1}^{N_\omega} u_j(\mathbf{x}), \qquad u_j(\mathbf{x}) := \frac{1}{4}\sum_{q=1}^{6} \alpha_q \left| C_j^{(q)}(\mathbf{x}) \right|^2, \tag{16}$$

where $\alpha_q = \epsilon_0$ for an electric field component, whereas $\alpha_q = \epsilon_0 c^2$ for a magnetic field component (here ϵ_0 is the vacuum permittivity, with $\epsilon_0 = 1/(4\pi \times 9 \times 10^9)$ in SI units). Thus, the spectral energy density (SED) has a discrete form.

Specializing to the present axisymmetric model, we thus have $C_j^{(q)} = C_j^{(q)}(\rho, z)$ and $u_j = u_j(\rho, z)$. The potentials $A_{jz} = A'_{jz} = \psi_{\omega_j S_j}$ are determined by the spectrum functions S_j in Eq. (10), which are given in the numerical model by the values $S_{nj} := S_j(k_{nj})$, that are the output of the fitting. These potentials generate the EM field, hence the $C_j^{(q)}(\rho, z)$ coefficients in Eq. (15) are expressed uniquely in terms of the S_{nj} 's.[21] However, in the least-squares problem (14), the scalar radiations emitted by every point-like "star" are taken to be exactly $\psi_{\mathbf{x}_i\, \omega_j}$. Clearly, we may multiply the l.h.s. of (14) by some number $\xi_j > 0$, thus obtaining now new values $S'_{nj} = \xi_j S_{nj}$ $(n = 0, ..., N)$ as the solution of (14). We determine the numbers $\xi_j > 0$ so that the SED measured at our local position $\mathbf{x}_{\rm loc}$ in the Galaxy coincides with the calculated values $u_j(\mathbf{x}_{\rm loc})$. This allows us then to make predictions: in particular, ones of the spatial variation of the SED in the Galaxy, which we may compare with the predictions of the existing models of the ISRF. Figures 1–2 show this comparison for the four positions in the Galaxy for which the predicted SED is shown in Ref. 13. The predictions of the two models are quite reasonably close, although the SED predicted by the present model has rather marked oscillations as function of the wavelength. Note that the different wavelengths are fully uncoupled due to the "separate fitting" defined by the N_ω least-square problems (14).

A surprising prediction of this model is that for the values of the maximum of the energy density,

$$u_{j\max} = \mathrm{Max}\{u_j(\rho_m, z_p);\ m = 1, ..., N_\rho,\ p = 1, ..., N_z\}, \tag{17}$$

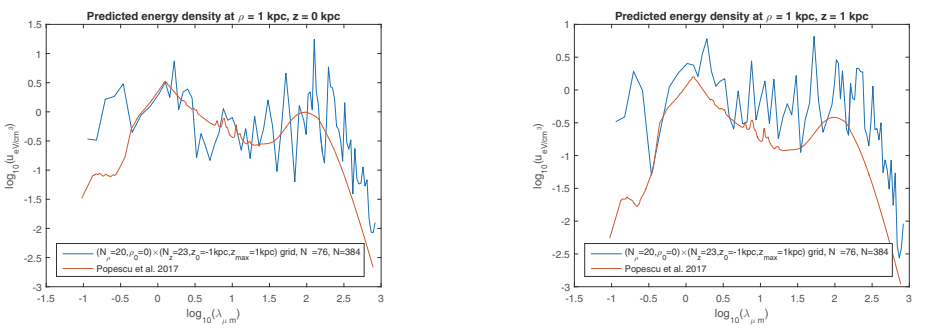


Fig. 1. SEDs at ($\rho = 1\,$kpc, $z = 0$) and at ($\rho = 1\,$kpc, $z = 1\,$kpc).

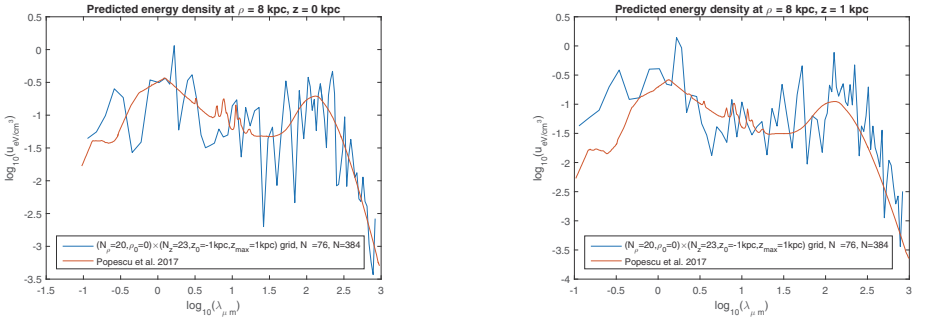


Fig. 2. SEDs at ($\rho = 8\,\mathrm{kpc}$, $z = 0$) and at ($\rho = 8\,\mathrm{kpc}$, $z = 1\,\mathrm{kpc}$).

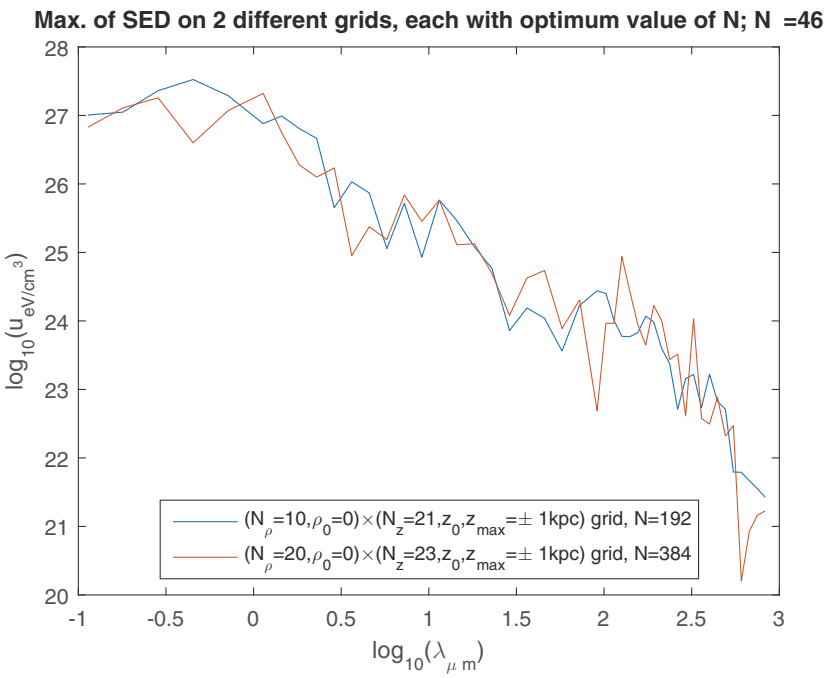


Fig. 3. Maximum of energy density; comparison between two spatial grids.

found for the different spatial grids ($N_\rho \times N_z$) investigated, all having ρ varying regularly from $\rho_0 = 0$ to $\rho_{\max} \simeq 10\,\mathrm{kpc}$ and z varying regularly from $z_0 = 0$ or $z_0 = -z_{\max}$ to $z_{\max} \leq 1\,\mathrm{kpc}$. Figure 3 compares the curves $u_{j\max} = f(\lambda_j)$ found with two spatial grids. It is seen that the two curves are quite close to one another, and both show extremely high levels of $u_{j\max}$, from $10^{27}\mathrm{eV/cm}^3$ to $10^{21}\mathrm{eV/cm}^3$. This is confirmed by a rather detailed investigation of the effects of the settings of the calculation (the spatial grid, and also the fineness of the frequency mesh: N_ω,

and that of the discretization: N).[20] The values of the maximum of u_j are always found on the axis of the Galaxy ($\rho = 0$), moreover the level of u_j decreases very rapidly when one departs from the axis.[20] This prediction of the model may be described as a kind of self-focusing effect of the ISRF in an axisymmetric galaxy.

6. Conclusion

In the "scalar ether theory" of gravity (SET), a consistent electrodynamics in a gravitational field needs the introduction of an additional energy tensor: $\boldsymbol{T}_{\text{inter}}$, with $T^{\mu}_{\text{inter}\ \ \nu} := p\,\delta^{\mu}_{\nu}$. Thus, this energy tensor was not designed to build missing mass. However, it turns out that the corresponding "medium" could contribute to dark matter, for it is not localized inside matter, it is gravitationally active, and it is "exotic". Moreover, the scalar field p, that determines $\boldsymbol{T}_{\text{inter}}$, can be in principle calculated from the data of the EM field and the gravitational field, through explicit equations. This however demands to be able to model the EM field in a galaxy, which is essentially in the form of the interstellar radiation field (ISRF).

Therefore, we built a Maxwell model of the ISRF. This was motivated by the foregoing, but it is also interesting independently of that, as the ISRF is a very important physical characteristic of a galaxy and interacts strongly with the cosmic rays. In any case, this model in itself is totally independent of the theory of gravitation and the assumption about the interaction tensor. It is based on an explicit representation of any time-harmonic axisymmetric source-free Maxwell field through a pair of scalar potentials, and on determining these potentials by fitting contributions emanating from a set of point-like "stars" schematizing a galaxy. The predictions of the model for the variation of the spectral energy distribution in the Galaxy are currently being checked. They are relatively close to the predictions of a recent radiation transfer model — except for the fact that the Maxwell model of the ISRF predicts extremely high values of the energy density on the axis of the Galaxy, that however decrease very rapidly when departing from that axis. We hope to be able in a future work to apply the model to calculate the interaction energy and to check if its distribution resembles a dark halo.

References

1. M. Arminjon, Space isotropy and weak equivalence principle in a scalar theory of gravity, *Braz. J. Phys.* **36**, 177–189 (2006).
2. M. Arminjon and R. W. Winkler, Motion of a test particle according to the scalar ether theory of gravitation and application to its celestial mechanics, *Z. Naturforsch. A* **74**, 305–316 (2019).
3. H. Stephani, *Relativity, An Introduction to Special and General Relativity*, Third edition (Cambridge University Press, Cambridge, 2004), pp. 156–158.
4. M. Arminjon, Continuum dynamics and the electromagnetic field in the scalar ether theory of gravitation, *Open Phys.* **14**, 395–409 (2016).
5. M. Arminjon, Charge conservation in a gravitational field in the scalar ether theory, *Open Phys.* **15**, 877–890 (2017).

6. M. Arminjon, Lorentz-invariant second-order tensors and an irreducible set of matrices, *J. Geom. Symmetry Phys.* **50**, 1–10 (2018).
7. M. Arminjon, On the equations of electrodynamics in a flat or curved spacetime and a possible interaction energy, *Open Phys.* **16**, 488–498 (2018).
8. B. T. Draine, Photoelectric heating of interstellar gas, *Astrophys. J. Suppl. Ser.* **36**, 595–619 (1978).
9. J. S. Mathis, P. G. Mezger and N. Panagia, Interstellar radiation field and dust temperatures in the diffuse interstellar matter and in giant molecular clouds, *Astron. Astrophys.* **128**, 212–229 (1983).
10. X. Chi and A. W. Wolfendale, The interstellar radiation field: a datum for cosmic ray physics, *J. Phys. C: Nucl. Part. Phys.* **17**, 987–998 (1991).
11. K. D. Gordon, K. A. Misselt, A. N. Witt and G. C. Clayton, The DIRTY model. I. Monte Carlo radiative transfer through dust, *Astrophys. J.* **551**, 269–276 (2001).
12. T. P. Robitaille, HYPERION: an open-source parallelized three-dimensional dust continuum radiative transfer code, *Astron. Astrophys.* **536**, A79, 17 pages (2011).
13. C. C. Popescu, R. Yang, R. J. Tuffs, G. Natale, M. Rushton and F. Aharonian, A radiation transfer model for the Milky Way: I. Radiation fields and application to High Energy Astrophysics, *Mon. Not. Roy. Astr. Soc.* **470**, no. 3, 2539–2558 (2017).
14. N. D. Kylafis and J. N. Bahcall, Dust distribution in spiral galaxies, *Astrophys. J.* **317**, 637–645 (1987).
15. T. A. Porter and A. W. Strong, A new estimate of the galactic interstellar radiation field between 0.1μm and 1000μm, in *Proc. 29th International Cosmic Ray Conference Pune* (Tata Institute of Fundamental Research, Mumbai, 2005), vol. 4, pp. 77–80.
16. C. C. Popescu, R. J. Tuffs, M. A. Dopita, J. Fischera, N. D. Kylafis and B. F. Madore, Modelling the spectral energy distribution of galaxies. V. The dust and PAH emission SEDs of disk galaxies, *Astron. Astrophys.* **527**, A109, 40 pages (2011).
17. M. Arminjon, An analytical model for the Maxwell radiation field in an axially symmetric galaxy, *Open Phys.* **19**, 77–90 (2021).
18. M. Arminjon, An explicit representation for the axisymmetric solutions of the free Maxwell equations, *Open Phys.* **18**, 255–263 (2020).
19. M. Zamboni-Rached, E. Recami and H. E. Hernández-Figueroa, Structure of non-diffracting waves and some interesting applications, in *Localized Waves*, eds. H. E. Hernández-Figueroa, M. Zamboni-Rached, E. Recami (John Wiley & Sons, Hoboken, 2008), pp. 43–77.
20. M. Arminjon, *Interstellar radiation as a Maxwell field: improved numerical scheme and application to the spectral energy density*, Preprint HAL-03341905 (2021).
21. M. Arminjon, Spectral energy density in an axisymmetric galaxy as predicted by an analytical model for the Maxwell field, *Adv. Astron.* **2021**, 5524600, 13 pages (2021).

The maximum mass of dilute axion stars

Pierre-Henri Chavanis

Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France
E-mail: chavanis@irsamc.ups-tlse.fr

We consider the possibility that dark matter is made of bosons in the form of Bose-Einstein condensates. We establish the mass-radius relation $M(R)$ of nonrelativistic self-gravitating Bose-Einstein condensates with repulsive or attractive self-interaction. If the self-interaction is repulsive there exists an equilibrium state for any value of the mass but if the self-interaction is attractive, as in the case of axions, equilibrium states exist only below a maximum mass $M_{\rm max}^{\rm NR}=5.073\, M_P/\sqrt{|\lambda|}$ [P.H. Chavanis, Phys. Rev. D **84**, 043531 (2011)]. This is the maximum mass of dilute axion stars. Above that mass, the star collapses leading to a bosenova, a black hole, a dense axion star or axion drops. We consider how the maximum mass changes with the dimension of space, the presence of a central black hole, a cosmological constant, and relativistic effects. We apply these results to dark matter halos. We establish the general expression of the core mass – halo mass relation $M_c(M_h)$. For a repulsive self-interaction, we show that the core mass M_c is always much below the maximum mass $M_{\rm max}^{\rm GR}$ set by general relativity so the core cannot collapse towards a black hole. For an attractive self-interaction, we show that the core mass M_c can reach in principle the maximum mass $M_{\rm max}^{\rm NR}$ in sufficiently large dark matter halos, leading to core collapse, and we discuss if this situation can happen in practice.

Keywords: Self-gravitating Bose-Einstein condensates; axion stars; maximum mass

1. Introduction

The concept of boson stars was introduced by Kaup[1] and Ruffini and Bonazzola[2] in the 1960s. They considered the $T=0$ limit in which bosons form Bose-Einstein condensates (BECs). In that case, all the bosons are in the same quantum state described by a unique complex wave function $\varphi(x^\mu)$ satisfying the Klein-Gordon-Einstein (KGE) equations. Boson stars can be regarded as macroscopic quantum states that are only prevented from collapsing gravitationally by the Heisenberg uncertainty principle. Kaup[1] and Ruffini and Bonazzola[2] showed that equilibrium states can exist only below a maximum mass $M_{\rm max}^{\rm GR}=0.633M_P^2/m$, where $M_P=(\hbar c/G)^{1/2}$ is the Planck mass, set by general relativity. Above that mass, boson stars are expected to collapse and form a black hole. These results are remarkably similar to those obtained for neutron stars which also display a maximum mass $M_{\rm max}^{\rm GR}=0.384\, M_P^3/m^2$ as found by Oppenheimer and Volkoff.[3] However, there also exist crucial differences between boson and fermion stars. In particular, boson stars are stopped from collapsing by the Heisenberg uncertainty principle while gravitational collapse in fermion stars is avoided by the Pauli exclusion principle. This difference is reflected in the scaling with m of the maximum mass of stable configurations. The maximum mass of boson stars scales as $M_{\rm max}^{\rm GR}\sim M_P^2/m$ instead

of $M_{\rm max}^{\rm GR} \sim M_P^3/m^2$ for fermions. They differ by a factor $m/M_P \ll 1$. This leads to the concept of "mini-boson stars" or "mini-soliton stars" with extremely high densities.

Colpi *et al.*[4] considered the case of bosons stars with a repulsive $\frac{\lambda}{4\hbar c}|\varphi|^4$ ($\lambda > 0$) self-interaction and found that the resulting configurations differ markedly from the noninteracting case. In the Thomas-Fermi (TF) limit, the maximum mass set by general relativity is $M_{\rm max}^{\rm GR} = 0.06\sqrt{\lambda} M_P^3/m^2$ (see Refs.[4,5]). For $\lambda \sim 1$ it exhibits the same scaling as the maximum mass of fermion stars. This leads to much bigger structures than in the noninteracting case, making them much more astrophysically interesting. They are called "massive boson stars". In a sense, the repulsive self-interaction for bosons plays a role similar to the quantum pressure arising from the Pauli exclusion principle for fermions.

The case of bosons with an attractive $\frac{\lambda}{4\hbar c}|\varphi|^4$ ($\lambda < 0$) self-interaction, like the QCD axion with a mass $m \sim 10^{-4}\,{\rm eV}/c^2$ and a negative self-interaction constant $\lambda \sim -7.39 \times 10^{-49}$, was considered by Kolb and Tkachev[6] in a cosmological context. In the early Universe, self-gravity can be neglected. Because of the attractive self-interaction, axions can form miniclusters of mass $M_{\rm axiton} \sim 10^{-12}\,M_\odot$ and size $R_{\rm axiton} \sim 10^9\,{\rm m}$ called axitons. Tkachev[7] took self-gravity into account and considered the possibility to form axion stars by Jeans instability. However, he assumed a repulsive self-interaction ($\lambda > 0$). When the self-interaction between bosons is attractive, there exists a maximum mass $M_{\rm max}^{\rm NR} = 5.073\,M_P/\sqrt{|\lambda|}$ which was first identified by Chavanis.[8] This is the maximum mass of dilute axion stars. This is a purely nonrelativistic result, contrary to the maximum mass of boson and fermion stars discussed above which is due to general relativity. For QCD axion stars, one finds $M_{\rm max}^{\rm NR} = 6.46 \times 10^{-14}\,M_\odot$ and a corresponding radius $(R_{99}^*) = 227\,{\rm km}$ which are of the order of the mass and size of asteroids. This leads to the notion of "axteroids". For $M > M_{\rm max}^{\rm NR}$ there is no equilibrium state and the axion star collapses.[9] This leads to a bosenova[10] or a black hole[11,12] if we take special or general relativity into account, or to the formation of a dense axion star[13] if we take into account higher order terms in the expansion of the self-interaction potential, e.g., a repulsive $|\phi|^6$ self-interaction[14,15] that can stabilize the star against gravitational collapse.[a] The axion star can also fragment into several stable pieces (axion "drops") of mass $M' < M_{\rm max}$, thereby preventing its complete collapse.[19,20]

If bosons are ultralight,[21] with a mass $m \sim 10^{-22}\,{\rm eV/c^2}$, they can form compact objects of galactic size.[b] Therefore, ultralight axions have been invoked in models

[a] Visinelli *et al.*[16] and Eby *et al.*[17] argue that relativistic effects are crucial on the branch of dense axion stars while self-gravity is negligible. As a result, dense axion stars correspond to "pseudobreathers" or "oscillons" which are described by the sine-Gordon equation. For a real scalar field (SF), these objects are known to be unstable and to decay via emission of relativistic axions on a timescale much shorter than any cosmological timescale. This conclusion is, however, contested by Braaten and Zhang.[18]

[b] More massive bosons with a mass $10^{-22}\,{\rm eV/c^2} \leq m \leq 10^{-3}\,{\rm eV/c^2}$ can also form compact objects of galactic size provided that they are self-interacting (with $\lambda \leq 10^{-15}$).[22]

of dark matter (DM) halos. Boson stars or BECDM halos can be formed from a dissipationless relaxation process called gravitational cooling[23] or violent relaxation.[24] By this process, large DM halos acquire a "core-halo" structure with a quantum core (soliton) in its ground state surrounded an extended halo which results from the quantum interferences of the excited states. This "core-halo" structure has been evidenced in numerical simulations of the Schrödinger-Poisson equations.[25–33] The quantum core may solve the core-cusp problem of the cold dark matter (CDM) model and the halo (similar to an isothermal of NFW profile) accounts for the flat rotation curves of the galaxies. A general expression of the core mass-halo mass relation $M_c(M_h)$ in BECDM halos with no, repulsive or attractive self-interaction has been obtained in Refs.[22,34–36]. For an attractive self-interaction, and in sufficiently large DM halos, the core mass can overcome the maximum mass $M_{\rm max}^{\rm NR}$ and collapse. The conditions for this collapse require, however, slightly stronger self-interactions that those commonly allowed by particle physics and cosmology.

In these Proceedings, we review some aspects of our work on self-gravitating BECs. In Sec. 2, we discuss self-gravitating BECs in the nonrelativistic regime described by the Gross-Pitaevskii-Poisson (GPP) equations. In Sec. 3, we discuss the exact mass-radius relation of self-gravitating BECs. In Sec. 4, we obtain an approximation analytical expression of the mass-radius relation of self-gravitating BECs based on a Gaussian ansatz. In Sec. 5, we discuss the maximum mass of dilute axion stars (self-gravitating BECs with attractive self-interaction) and some of its generalizations. In Sec. 6, we apply our results to DM halos and discuss the core mass – halo mass relation and the possibilty or not to have core collapse. In Appendices A and B, we consider general relativistic BECs described by the KGE equations for a complex and a real SF.

2. Self-gravitating BECs

2.1. *Gross-Pitaevskii-Poisson equations*

We assume that DM is made of bosons (like the axion) in the form of BECs at $T = 0$. We use a nonrelativistic approach based on Newtonian gravity. The evolution of the wave function $\psi(\mathbf{r}, t)$ of a self-gravitating BEC is governed by the GPP equations[8,37]

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi + m\Phi\psi + m\frac{dV}{d|\psi|^2}\psi, \tag{1}$$

$$\Delta\Phi = 4\pi G|\psi|^2, \tag{2}$$

where m is the mass of the bosons, $\Phi(\mathbf{r}, t)$ is the gravitational potential and $\rho(\mathbf{r}, t) = |\psi|^2$ is the mass density of the BEC. The first term in Eq. (1) is the kinetic term which accounts for the Heisenberg uncertainty principle. The second term accounts for the self-gravity of the BEC. The third term takes into account the self-interaction of the bosons. We shall consider a standard $|\psi|^4$ potential of the form

$$V(|\psi|^2) = \frac{2\pi a_s\hbar^2}{m^3}|\psi|^4, \tag{3}$$

where a_s is the scattering length of the bosons. The interaction between the bosons is repulsive when $a_s > 0$ and attractive when $a_s < 0$.

The GPP equations conserve the mass $M = \int |\psi|^2 \, d\mathbf{r}$ and the energy

$$E_{\rm tot} = \frac{\hbar^2}{2m^2} \int |\nabla\psi|^2 \, d\mathbf{r} + \frac{1}{2} \int |\psi|^2 \Phi \, d\mathbf{r} + \int V(|\psi|^2) \, d\mathbf{r}, \tag{4}$$

which is the sum of the kinetic energy Θ, the gravitational energy W, and the internal energy U (i.e. $E_{\rm tot} = \Theta + W + U$).

2.2. *Madelung transformation*

Writing the wave function as

$$\psi(\mathbf{r}, t) = \sqrt{\rho(\mathbf{r}, t)} e^{iS(\mathbf{r},t)/\hbar}, \tag{5}$$

where $\rho(\mathbf{r}, t)$ is the mass density and $S(\mathbf{r}, t)$ is the action, and making the Madelung transformation

$$\rho(\mathbf{r}, t) = |\psi|^2 \qquad \text{and} \qquad \mathbf{u} = \frac{\nabla S}{m}, \tag{6}$$

where $\mathbf{u}(\mathbf{r}, t)$ is the velocity field, the GPP equations (1) and (2) can be written under the form of hydrodynamic equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{7}$$

$$\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + m\Phi + mV'(\rho) + Q = 0, \tag{8}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{m} \nabla Q - \frac{1}{\rho} \nabla P - \nabla\Phi, \tag{9}$$

$$\Delta\Phi = 4\pi G\rho, \tag{10}$$

where

$$Q = -\frac{\hbar^2}{2m} \frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} = -\frac{\hbar^2}{4m} \left[\frac{\Delta\rho}{\rho} - \frac{1}{2} \frac{(\nabla\rho)^2}{\rho^2} \right] \tag{11}$$

is the quantum potential taking into account the Heisenberg uncertainty principle, and $P(\rho)$ is the pressure arising from the self-interaction of the bosons given by

$$P(\rho) = \rho V'(\rho) - V(\rho). \tag{12}$$

For the standard self-interaction potential (3) we get

$$P = \frac{2\pi a_s \hbar^2}{m^3} \rho^2. \tag{13}$$

This is the equation of state of a polytrope of index $n = 1$. The squared speed of sound is $c_s^2 = P'(\rho) = \rho V''(\rho) = 4\pi a_s \hbar^2 \rho / m^3$. The hydrodynamic equations (7)-(10) are called the quantum Euler-Poisson equations. Equation (7), corresponding to the

imaginary part of the GP equation, is the continuity equation. Equation (8), corresponding to the real part of the GP equation, is the quantum Hamilton-Jacobi (or Bernoulli) equation. Equation (9), obtained by taking the gradient of Eq. (8), is the momentum equation. Equation (10) is the Poisson equation. In the TF limit where the quantum potential can be neglected (formally $\hbar \to 0$ with $4\pi a_s\hbar^2/m^3$ finite), they become equivalent to the classical Euler-Poisson equations for a barotropic gas.

The quantum Euler equations conserve the mass $M = \int \rho\, d\mathbf{r}$ and the energy

$$E_{\rm tot} = \int \rho \frac{\mathbf{u}^2}{2}\, d\mathbf{r} + \frac{1}{m}\int \rho Q\, d\mathbf{r} + \frac{1}{2}\int \rho\Phi\, d\mathbf{r} + \int V(\rho)\, d\mathbf{r}, \tag{14}$$

which is the sum of the classical kinetic energy Θ_c, the quantum kinetic energy Θ_Q, the gravitational energy W, and the internal energy U.

2.3. *Equilibrium states*

The condition of quantum hydrostatic equilibrium, corresponding to a steady state of the quantum Euler equation (9), writes

$$\frac{\rho}{m}\nabla Q + \nabla P + \rho\nabla\Phi = \mathbf{0}. \tag{15}$$

Combining Eq. (15) with the Poisson equation (10), and considering a standard BEC described by the potential (3), we obtain the fundamental differential equation of quantum hydrostatic equilibrium[8,37]

$$\frac{\hbar^2}{2m^2}\Delta\left(\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) - \frac{4\pi a_s\hbar^2}{m^3}\Delta\rho = 4\pi G\rho. \tag{16}$$

This equation describes the balance between the quantum potential taking into account the Heisenberg uncertainty principle, the pressure due to the self-interaction of the bosons, and the self-gravity. It corresponds to a stationary state of the GPP equations.[8,37] This equation can also be obtained from an energy principle. Indeed, one can show (see Appendix B of Ref.[22]) that an equilibrium state of the GPP equations is an extremum of energy $E_{\rm tot}$ at fixed mass M and that an equilibrium state is dynamically stable if, and only if, it is a minimum of energy at fixed mass.

3. Exact mass-radius relation of self-gravitating BECs

The fundamental equation of hydrostatic equilibrium of self-gravitating BECs, Eq. (16), has been solved numerically (exactly) in Ref.[38] for an arbitrary self-interaction (repulsive or attractive). The nodeless solution describes a compact gravitational quantum object (soliton/BEC) in its ground state. From this solution we have determined the exact mass-radius relation of self-gravitating BECs.

3.1. *Noninteracting bosons*

For noninteracting bosons ($a_s = 0$), the mass-radius relation is given by[38,39]

$$M = 9.95\frac{\hbar^2}{Gm^2R_{99}}, \tag{17}$$

where R_{99} represents the radius containing 99% of the mass (the density profile has not a compact support). The mass decreases as the radius increases. The equilibrium states are all stable.

3.2. *Bosons with a repulsive self-interaction*

For bosons with a repulsive self-interaction ($a_s > 0$), the exact mass-radius relation is represented in Fig. 1. The mass decreases as the radius increases. In the TF limit $M \gg \hbar/\sqrt{Gma_s}$ where the quantum potential can be neglected, the density profile can be obtained analytically since the system is equivalent to a polytrope of index $n = 1$. We get

$$\rho(r) = \frac{\rho_0 R_{\rm TF}}{\pi r}\sin\left(\frac{\pi r}{R_{\rm TF}}\right). \tag{18}$$

The equilibrium states have a unique radius given by[7,8,40–44]

$$R_{\rm TF} = \pi\left(\frac{a_s\hbar^2}{Gm^3}\right)^{1/2}, \tag{19}$$

which is independent of their mass M. This is the minimum radius of self-gravitating BECs with a repulsive self-interaction. In the noninteracting limit $M \ll \hbar/\sqrt{Gma_s}$ and $R \gg R_{\rm TF}$, we recover Eq. (17).

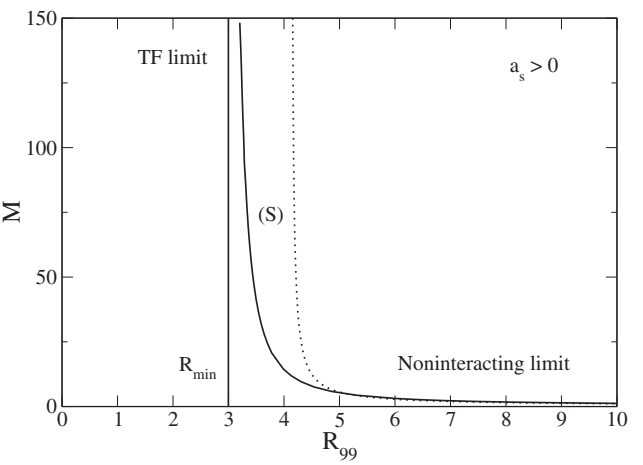


Fig. 1. Mass-radius relation of self-gravitating BECs with $a_s > 0$ (full line: exact;[38] dotted line: Gaussian ansatz[8]). The mass is normalized by $M_a = \hbar/\sqrt{Gma_s}$ and the radius by $R_a = (a_s\hbar^2/Gm^3)^{1/2}$.

3.3. *Bosons with an attractive self-interaction*

For bosons with an attractive self-interaction ($a_s < 0$), the exact mass-radius relation is represented in Fig. 2. It displays a maximum mass[8,38]

$$M_{\rm max}^{\rm NR} = 1.012\,\frac{\hbar}{\sqrt{Gm|a_s|}} \quad \text{at} \quad R_{99}^* = 5.5\,\left(\frac{|a_s|\hbar^2}{Gm^3}\right)^{1/2}. \tag{20}$$

There is no equilibrium state with $M > M_{\rm max}^{\rm NR}$. In that case, the BEC is expected to collapse.[9] The outcome of the collapse (dense axion star, black hole, bosenova, axion drops...) is discussed in Refs.[9–20]. For $M < M_{\rm max}$ there are two possible equilibrium states with the same mass. The equilibrium states with $R > R_{99}^*$ are stable and the equilibrium states with $R < R_{99}^*$ are unstable. This can be shown by using the Poincaré criterion, the Wheeler theorem, or by computing the squared pulsation.[8,9,15,38] In the nongravitational limit $M \ll M_{\rm max}^{\rm NR}$ and $R \ll R_{99}^*$, the mass-radius relation is given by[38]

$$M = 0.275\,\frac{mR_{99}}{|a_s|}. \tag{21}$$

These equilibrium states are unstable. In the noninteracting limit $M \ll M_{\rm max}^{\rm NR}$ and $R \gg R_{99}^*$, we recover Eq. (17). These equilibrium states are stable.

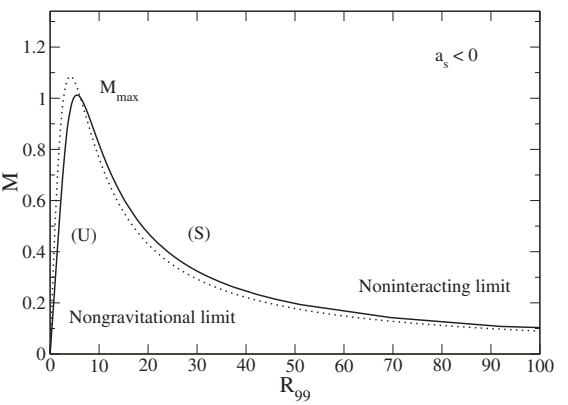


Fig. 2. Mass-radius relation of self-gravitating BECs with $a_s < 0$ (full line: exact;[38] dotted line: Gaussian ansatz[8]). The mass is normalized by $M_a = \hbar/\sqrt{Gm|a_s|}$ and the radius by $R_a = (|a_s|\hbar^2/Gm^3)^{1/2}$.

4. Approximate mass-radius relation of self-gravitating BECs from a Gaussian ansatz

Stable BEC stars correspond to minima of energy $E_{\rm tot}$ at fixed mass M. We can obtain an approximate analytical expression of the mass-radius relation of self-gravitating BECs by making a Gaussian ansatz for the wave function leading to the

density profile[c]

$$\rho(\mathbf{r}) = \frac{M}{R^3}\frac{1}{\pi^{3/2}}e^{-r^2/R^2}. \tag{22}$$

The radius containing 99% of the mass is $R_{99} = 2.38167\,R$. With this ansatz the total energy can be written as[8,37]

$$E_{\rm tot}(R) = \sigma\frac{\hbar^2 M}{m^2R^2} - \nu\frac{GM^2}{R} + \zeta\frac{2\pi a_s\hbar^2M^2}{m^3R^3} \tag{23}$$

with $\sigma = 3/4$, $\nu = 1/\sqrt{2\pi}$ and $\zeta = 1/(2\pi)^{3/2}$. At equilibrium, the condition $E'_{\rm min}(R) = 0$ (extremum of energy) gives the mass-radius relation[8]

$$M = \frac{\frac{2\sigma}{\nu}\frac{\hbar^2}{GmR}}{1-\frac{6\pi\zeta}{\nu}\frac{a_s\hbar^2}{Gm^3R^2}}. \tag{24}$$

A more general expression of the mass-radius relation is given in Eq. (292) of Ref.[37]. The BEC is stable provided that $E''_{\rm tot}(R) > 0$ which corresponds to the requirement that the equilibrium state is a minimum of energy at fixed mass or, equivalently, that the squared pulsation is positive ($\omega^2 > 0$).

For noninteracting bosons ($a_s = 0$), the mass-radius relation from Eq. (24) reduces to

$$M = \frac{2\sigma}{\nu}\frac{\hbar^2}{Gm^2R}. \tag{25}$$

For bosons with a repulsive self-interaction ($a_s > 0$), in the TF limit ($\hbar \to 0$ with fixed $g = 4\pi a_s\hbar^2/m^3$), the mass-radius relation from Eq. (24) reduces to

$$R = \left(\frac{6\pi\zeta}{\nu}\right)^{1/2}\left(\frac{a_s\hbar^2}{Gm^3}\right)^{1/2}. \tag{26}$$

For bosons with an attractive self-interaction ($a_s < 0$), the approximate values of the maximum mass and of the corresponding radius obtained from Eq. (24) are

$$M_{\rm max} = \left(\frac{\sigma^2}{6\pi\zeta\nu}\right)^{1/2}\frac{\hbar}{\sqrt{Gm|a_s|}}, \qquad R_* = \left(\frac{6\pi\zeta}{\nu}\right)^{1/2}\left(\frac{|a_s|\hbar^2}{Gm^3}\right)^{1/2}. \tag{27}$$

In the nongravitational limit, the mass-radius relation from Eq. (24) reduces to

$$M = \frac{\sigma}{3\pi\zeta}\frac{mR}{|a_s|}. \tag{28}$$

5. Generalizations of the maximum mass of axion stars

In this section, we discuss some generalizations of the maximum mass of axion stars.

[c]Here, we restrict ourselves to the equilibrium state, so we just need to make an ansatz for the density profile. See Sec. 8 of Ref.[37] for a more general study.

5.1. *Dimension of space*

The previous results are valid for spherically symmetric systems in $d = 3$ dimensions. It is possible to generalize them in a space of arbitrary dimension d.[46] For an attractive self-interaction, there is a maximum mass when $2 \leq d \leq 4$.[d] Using a Gaussian ansatz we find that

$$M_{\rm max} = (4-d)^{\frac{4-d}{2}}(d-2)^{\frac{d-2}{2}}\frac{\sigma}{\nu^{\frac{d-2}{2}}(d\zeta)^{\frac{4-d}{2}}}\frac{\hbar^{d-2}}{(2\pi|a_s|)^{\frac{4-d}{2}}G^{\frac{d-2}{2}}m^{\frac{3d-8}{2}}}, \tag{29}$$

$$R_* = \left(\frac{d-2}{4-d}\right)^{1/2}\left(\frac{d\zeta}{\nu}\right)^{1/2}\left(\frac{2\pi|a_s|\hbar^2}{Gm^3}\right)^{1/2} \tag{30}$$

with $\sigma = d/4$, $\nu = 1/[\Gamma(d/2)2^{d/2}]$ and $\zeta = 1/(2\pi)^{d/2}$. These equations define $M_{\rm max}(d)$ and $R_*(d)$. In $d = 2$ dimensions, we obtain

$$M_{\rm max} = \frac{\sigma}{2\pi\zeta}\frac{m}{|a_s|}, \qquad R_* = 0, \tag{31}$$

which corresponds to the nongravitational limit ($G = 0$). The prefactor is equal to $1/2$. This expression can be compared with the exact solution $M_{\rm max}^{\rm exact} = 0.465638...\frac{m}{|a_s|}$ associated with the Townes[47] soliton obtained by solving the ordinary GP equation in $d = 2$. In $d = 4$ dimensions, we obtain

$$M_{\rm max} = \frac{2\sigma}{\nu}\frac{\hbar^2}{Gm^2}, \qquad R_* \to +\infty, \tag{32}$$

which corresponds to the noninteracting limit ($a_s = 0$).

5.2. *Central black hole*

The effect of a central black hole, creating an external potential $\Phi_{\rm BH} = -GM_{\rm BH}/r$, on the value of the maximum mass has been investigated in Ref.[48]. It is found that

$$\frac{M_{\rm max}}{M_{\rm max}^{(0)}} = \frac{R_*}{R_*^{(0)}} = -\frac{\mu}{2} + \frac{1}{2}\sqrt{\mu^2+4}, \tag{33}$$

where

$$\mu = \frac{\lambda}{\nu}\frac{M_{\rm BH}}{M_{\rm max}^{(0)}} \tag{34}$$

is the normalized mass of the black hole. These equations define $M_{\rm max}(M_{\rm BH})$ and $R_*(M_{\rm BH})$. In these expressions, $M_{\rm max}^{(0)}$ and $R_*^{(0)}$ denote the maximum mass and the corresponding radius in the absence of a central black hole [see Eq. (27)] and $\lambda = 2/\sqrt{\pi}$.

[d] When $d < 2$ there is a stable equilibrium state for any mass M. When $d > 4$ there is an equilibrium state for any mass M but it is unstable. When $2 \leq d \leq 4$ there is a maximum mass. The equilibrium states with $R < R_*$ are unstable and the equilibrium states with $R > R_*$ are stable. For a repulsive self-interaction there is a minimum mass when $d \geq 4$ given by an expression similar to Eq. (29). The equilibrium states are stable if $R < R_*$ and unstable if $R > R_*$. Noninteracting bosons stars are unstable when $d \geq 4$. These results show that the dimension $d = 3$ of our universe is very special (see also Ref.[45] for fermion stars).

5.3. *Harmonic potential*

The effect of an external harmonic potential $\Phi_{\rm H}=\frac{1}{2}\omega_0^2r^2$ on the value of the maximum mass has been investigated in Ref.[46]. When $\omega_0^2>0$ the potential is confining and can take into account tidal effects from neighboring systems. When $\omega_0^2<0$ the potential mimics a solid rotation $\Omega=\sqrt{-\omega_0^2}$ or the effect of the cosmological constant $\Lambda=-3\omega_0^2$ (dark energy). It is found that

$$M_{\rm max}=\frac{8R_*}{3R_*^2+5},\qquad \omega_0^2=\frac{2(1-R_*^2)}{R_*^4(3R_*^2+5)}, \tag{35}$$

where $M_{\rm max}$ is normalized by $M_{\rm max}^{(0)}$, R_* is normalized by $R_*^{(0)}$, and ω_0 is normalized by the inverse of the dynamical time $t_D=(6\pi\zeta/\nu)(\alpha/\sigma)^{1/2}|a_s|\hbar/(Gm^2)$. These equations define $M_{\rm max}(\omega_0)$ and $R_*(\omega_0)$ in parametric form. In the nongravitational case ($G=0$) and $\omega_0^2>0$ we find that

$$M_{\rm max}^{\rm NG}=\frac{4\sigma}{15\pi\zeta}\left(\frac{2\sigma}{5\alpha}\right)^{1/4}\left(\frac{m\hbar}{\omega_0 a_s^2}\right)^{1/2},\qquad R_*^{\rm NG}=\left(\frac{2\sigma}{5\alpha}\right)^{1/4}\left(\frac{\hbar}{m\omega_0}\right)^{1/2} \tag{36}$$

with $\alpha=3/2$. The prefactor of the maximum mass is equal to 0.6705.... This result can be compared with the exact solution $M_{\rm max}^{\rm NG,exact}=0.573...(m\hbar/\omega_0 a_s^2)^{1/2}$ obtained by numerically solving the ordinary GP equation with a confining harmonic potential.[49]

5.4. *Relativistic corrections*

For boson stars with an attractive self-interaction ($a_s<0$), the relativistic corrections to the maximum mass and the corresponding radius are[46]

$$M_{\rm max}=1.012\,\frac{\hbar}{\sqrt{Gm|a_s|\left(1+2.56\frac{Gm}{|a_s|c^2}\right)}}, \tag{37}$$

$$R_*=5.5\left[\frac{|a_s|\hbar^2\left(1+2.56\frac{Gm}{|a_s|c^2}\right)}{Gm^3}\right]^{1/2}. \tag{38}$$

These formulae interpolate between the exact expressions from Eq. (20) when $|a_s|\to+\infty$ and the exact expression of the maximum mass $M_{\rm max}^{\rm GR}=0.633\,\hbar c/Gm$ set by general relativity when $a_s=0$. We see that relativistic corrections tend to reduce the maximum mass and increase the minimum radius of axion stars with respect to the nonrelativistic limit. Relativistic corrections become important when $|a_s|\sim r_S$, where $r_S=Gm/c^2$ is the effective Schwarzschild radius of the bosons. Relativistic corrections are negligible when $|a_s|\gg r_S$. The transition scale $r_S\sim Gm/c^2$ can be obtained by equating $M_{\rm max}^{\rm NR}\sim\hbar/\sqrt{Gm|a_s|}$ and $M_{\rm max}^{\rm GR}\sim\hbar c/Gm$.

6. Application to DM halos

We now apply the results of the previous sections to DM halos. DM halos typically have a "core-halo" structure made of a quantum core (soliton) in its ground state surrounded by an approximately isothermal atmosphere arising from quantum interferences of excited states.[34,37] This "core-halo" structure results from a process of gravitational cooling[23] and violent relaxation[24] as evidenced in numerical simulations of the Schrödinger-Poisson equations.[25–33] The results of Secs. 2-5 describe the ground state of a self-gravitating BEC,[8] so they apply to the quantum core of DM halos. The "minimum halo" of mass $(M_h)_{\rm min}$ (the smallest halo observed in the universe), which is a purely condensed object without atmosphere, corresponds to the condition $M_c = M_h$. For bigger halos of mass $M_h \geq (M_h)_{\rm min}$, the mass M_c of the quantum core increases with the halo mass M_h. For noninteracting bosons and for bosons with a repulsive self-interaction, we may wonder if the core mass M_c can reach the maximum mass $M_{\rm max}^{\rm GR}$ set by general relativity and collapse towards a supermassive black hole (SMBH). For bosons with an attractive self-interaction, we may wonder if the core mass M_c can reach the maximum mass $M_{\rm max}^{\rm NR}$ of Ref.[8] and collapse. The outcome of the collapse in that case would be a dense axion "star", a black hole, a bosenova, or axion drops.[9–20]

6.1. *Core mass–halo mass relation*

In Refs.[22,34–36], we have derived the core mass–halo mass relation of DM halos (without or with the presence of a central black hole) from a thermodynamic approach. We have obtained a general relation $M_c(M_h)$ valid for noninteracting bosons as well as for bosons with a repusive or an attractive self-interaction (and for fermions not considered here). To obtain this relation, we have proceeded as follows:

(i) Using some approximations, we have calculated the Lynden-Bell entropy $S(M_c)$ as a function of the core mass M_c. The equilibrium core mass is then obtained by maximizing the Lynden-Bell entropy at fixed total mass and total energy. We have shown[34,35] that this maximization problem is equivalent to the "velocity dispersion tracing" relation according to which the velocity dispersion in the core $v_c^2 \sim GM_c/R_c$ is of the same order as the velocity dispersion in the halo $v_h^2 \sim GM_h/r_h$. This relation can be written as

$$v_c \sim v_h \qquad \Rightarrow \qquad \frac{M_c}{R_c} \sim \frac{M_h}{r_h}. \tag{39}$$

(ii) To relate the core radius R_c to the core mass M_c, we have used the analytical expression of the mass-radius relation $M_c(R_c)$ of self-gravitating BECs obtained from the Gaussian ansatz [see Eq. (24)] yielding[8]

$$M_c = \frac{3.76\,\frac{\hbar^2}{Gm^2R_c}}{1-3\frac{a_s\hbar^2}{Gm^3R_c^2}} \tag{40}$$

or, equivalently,

$$R_c = 1.87\,\frac{\hbar^2}{Gm^2M_c}\left(1 \pm \sqrt{1+0.849\,\frac{Gma_sM_c^2}{\hbar^2}}\right). \tag{41}$$

(iii) Assuming that the atmosphere is isothermal, we have shown that the halo mass-radius relation $M_h(r_h)$ is given by[34]

$$M_h = 1.76\,\Sigma_0 r_h^2, \tag{42}$$

where $\Sigma_0 = 141\,M_\odot/\mathrm{pc}^2$ is the universal surface density of DM halos infered from the observations.[e]

Combining Eqs. (39)-(42), we obtain the general expression of the core mass–halo mass relation $M_c(M_h)$ under the form[22,35]

$$M_c = 2.23\,\frac{\hbar\Sigma_0^{1/4}M_h^{1/4}}{G^{1/2}m}\left(1+1.06\,\frac{a_s}{m}\Sigma_0^{1/2}M_h^{1/2}\right)^{1/2}. \tag{43}$$

Writing $M_c = M_h$ for the minimum halo, we obtain

$$(M_h)_{\min}^{3/2} = 4.99\,\frac{\hbar^2\Sigma_0^{1/2}}{Gm^2}\left(1+1.06\,\frac{a_s}{m}\Sigma_0^{1/2}(M_h)_{\min}^{1/2}\right). \tag{44}$$

This equation determines the mass $(M_h)_{\min}$ of the minimum halo as a function of m and a_s. Inversely, for given $(M_h)_{\min}$, it determines a consistency relation between m and a_s (see the universal curve plotted in Fig. 19 of Ref.[35]).

6.2. *Noninteracting bosons*

For noninteracting bosons ($a_s = 0$), the core mass – halo mass relation (43) reduces to[22,35]

$$M_c = 2.23\left(\frac{\hbar^4\Sigma_0 M_h}{G^2m^4}\right)^{1/4}. \tag{45}$$

On the other hand, according to Eq. (44), the mass of the boson is given by $m = 2.23\,(\hbar^4\Sigma_0/G^2(M_h)_{\min}^3)^{1/4} = 2.25\times10^{-22}\,\mathrm{eV}/c^2$ [in the numerical applications we assume that $(M_h)_{\min} = 10^8\,M_\odot$]. For a DM halo of mass $M_h = 10^{12}\,M_\odot$ similar to the one that surrounds our Galaxy, we obtain a core mass $M_c = 10^9\,M_\odot$ and a core radius $R_c = 63.5\,\mathrm{pc}$. The quantum core represents a bulge or a nucleus. It cannot mimic a SMBH because it is too much extended ($Rc^2/GM \sim 10^6 \gg 1$).

The maximum mass and the minimum radius of a noninteracting boson star at $T = 0$ set by general relativity are[1,2]

$$M_{\max}^{\mathrm{GR}} = 0.633\,\frac{\hbar c}{Gm}, \qquad R_{\min}^{\mathrm{GR}} = 9.53\,\frac{GM_{\max}}{c^2}. \tag{46}$$

[e]We note that the halo mass M_h used in our papers[22,34–36] is related to the virial halo mass M_v used in Ref.[26] by $M_h/M_\odot = 6.01\times10^{-6}\,(M_v/M_\odot)^{4/3}$ [see Eq. (146) of Ref.[35]]. As a result, the scalings of the core mass – halo mass relations $M_c(M_h)$ and $M_c(M_v)$ are different.

These scalings can be qualitatively obtained as explained in Appendix B.2 of Ref.[8]. For a boson of mass $m = 2.25 \times 10^{-22}\,\mathrm{eV}/c^2$, we obtain $M_{\rm max}^{\rm GR} = 3.76 \times 10^{11}\,M_\odot$ and $R_{\rm min}^{\rm GR} = 0.171\,\mathrm{pc}$. The maximum mass is much larger than the typical quantum core mass of a DM halo. According to Eq. (45), the mass of the soliton would be equal to the maximum mass ($M_c = M_{\rm max}^{\rm GR}$) in a DM halo of mass $M_h = 6.49 \times 10^{-3} c^4/(G^2\Sigma_0) = 2.01 \times 10^{22}\,M_\odot$ (remarkably, this expression is independent of the boson mass and is of the order of the mass of the Universe[36]). Such a large halo mass is clearly unrealistic (the biggest DM halos observed in the Universe have a mass $M_h \sim 10^{14}\,M_\odot$). Therefore, the soliton present at the center of a noninteracting BECDM halo can never collapse towards a SMBH. This conclusion was first reached in Appendix C of Ref.[36]. Since $M_c \ll M_{\rm max}^{\rm GR}$ in all realistic DM halos, a nonrelativistic approach is justified.

6.3. *Repulsive self-interaction in the TF limit*

For bosons with a repulsive self-interaction in the TF limit, the core mass–halo mass relation (43) reduces to[22,35]

$$M_c = 2.30 \left(\frac{\hbar^2 \Sigma_0 a_s M_h}{G m^3}\right)^{1/2}. \tag{47}$$

On the other hand, according to Eq. (44), the ratio a_s/m^3 is given by $a_s/m^3 = 0.189\,G(M_h)_{\rm min}/\hbar^2\Sigma_0 = 4.35 \times 10^3\,\mathrm{fm}/(\mathrm{eV}/\mathrm{c}^2)^3$. For a DM halo of mass $M_h = 10^{12}\,M_\odot$ similar to the one that surrounds our Galaxy, we obtain a core mass $M_c = 10^{10}\,M_\odot$ and a core radius $R_c = 635\,\mathrm{pc}$. The quantum core represents a bulge or a nucleus. It cannot mimic a SMBH because it is too much extended ($Rc^2/GM \sim 10^6 \gg 1$).

The maximum mass and the minimum radius of a self-interacting boson star at $T = 0$ in the TF limit set by general relativity are[4,5]

$$M_{\rm max}^{\rm GR} = 0.307\,\frac{\hbar c^2 \sqrt{a_s}}{(Gm)^{3/2}}, \qquad R_{\rm min}^{\rm GR} = 6.25\,\frac{G M_{\rm max}}{c^2}. \tag{48}$$

These scalings can be qualitatively obtained as explained in Appendix B.3 of Ref.[8]. For a ratio $a_s/m^3 = 4.35 \times 10^3\,\mathrm{fm}/(\mathrm{eV}/\mathrm{c}^2)^3$, we obtain $M_{\rm max}^{\rm GR} = 2.35 \times 10^{15}\,M_\odot$ and $R_{\rm min}^{\rm GR} = 703\,\mathrm{pc}$. The maximum mass is much larger than the typical quantum core mass of a DM halo. According to Eq. (48), the mass of the soliton would be equal to the maximum mass ($M_c = M_{\rm max}^{\rm GR}$) in a DM halo of mass $M_h = 0.0178 c^4/(G^2\Sigma_0) = 5.51 \times 10^{22}\,M_\odot$ (remarkably, this expression is independent of the ratio a_s/m^3 and is of the order of the mass of the Universe[36]). Such a large halo mass is clearly unrealistic (the biggest DM halos observed in the Universe have a mass $M_h \sim 10^{14}\,M_\odot$). Therefore, the quantum core present at the center of a BECDM halo with repulsive self-interactions can never collapse towards a SMBH. This conclusion was first reached in Appendix C of Ref.[36]. Since $M_c \ll M_{\rm max}^{\rm GR}$ in all realistic DM halos, a nonrelativistic approach is justified.

6.4. *Attractive self-interaction*

For bosons with an attractive self-interaction, the core mass–halo mass relation (see Fig. 3) presents a maximum when the core mass reaches the critical value [see Eq. (27)]

$$(M_c)_{\rm max}=1.085\frac{\hbar}{\sqrt{Gm|a_s|}}=10.9\left(\frac{f^2\hbar}{c^3m^2G}\right)^{1/2} \tag{49}$$

at which it becomes unstable and collapses.[8,9] The collapse of the core, leading to a dense axion "star" (soliton), a black hole, a bosenova or to axion drops,[9–20] occurs in a DM halo of mass[22,35]

$$(M_h)_{\rm max}=0.223\frac{m^2}{a_s^2\Sigma_0}=2255\frac{f^4}{\hbar^2c^6\Sigma_0}. \tag{50}$$

We have expressed these results in terms of the decay constant f (see Appendix B) which is related to the mass m and the scattering length a_s by Eq. (B.13). We note that the maximum halo mass $(M_h)_{\rm max}$ depends only on f while the maximum core mass $(M_c)_{\rm max}$ depends on f and m.

The maximum mass of a self-gravitating BEC made of ultralight axions with mass $m_{\rm th}=2.92\times10^{-22}\,{\rm eV}/c^2$ and scattering length $(a_s)_{\rm th}=-3.18\times10^{-68}\,{\rm fm}$ (corresponding to $f_{\rm th}=1.34\times10^{17}\,{\rm GeV}$ and $\lambda_{\rm th}=-1.18\times10^{-96}$) is $M_{\rm max}=5.10\times10^{10}\,M_\odot$ and the corresponding radius is $R^*_{99}=1.09\,{\rm pc}$. The minimum halo ($(M_h)_{\rm min}=10^8\,M_\odot$, $(r_h)_{\rm min}=1\,{\rm kpc}$) has a mass much smaller than the maximum mass, so it is stable ($(M_h)_{\rm min}<(M_c)_{\rm max}$). The halo mass at which the core mass would become unstable ($M_c=(M_c)_{\rm max}$) and collapse is $(M_h)_{\rm max}=1.01\times10^{20}\,M_\odot$. Since the largest DM halos observed in the Universe have a much smaller mass, of the order of $M_h\sim10^{14}\,M_\odot\ll(M_h)_{\rm max}$, we conclude that the quantum cores of BECDM halos with an attractive self-interaction are always stable ($M_c<(M_c)_{\rm max}$). Furthermore, since $M_h\ll(M_h)_{\rm max}$, the attractive self-interaction is negligible. This corroborates our claims[22,35,36] that the attractive self-interaction of the bosons can be neglected for what concerns the structure of DM halos in the nonlinear regime: Everything happens *as if* the bosons were noninteracting.

Remark: We recall that the theoretical value $f_{\rm th}=1.34\times10^{17}\,{\rm GeV}$ of the decay constant used above comes from the constraints from particle physics and cosmology.[21] It may be interesting to relax these constraints and consider arbitrary values of f. In that case, we find that an attractive self-interaction would be important in realistic DM halos of mass $M_h\leq10^{14}\,M_\odot$, and lead to the collapse of their quantum core when $M_h=(M_h)_{\rm max}$ [corresponding to $M_c=(M_c)_{\rm max}$], if $f<4.22\times10^{15}\,{\rm GeV}$ [the bound corresponds to $(M_h)_{\rm max}=10^{14}\,M_\odot$ in Eq. (50)]. For example, for $f=1.34\times10^{15}\,{\rm GeV}$ and $m=10^{-22}\,{\rm eV}/c^2$, we find that the quantum core of a DM halo of mass $(M_h)_{\rm max}\sim10^{12}\,M_\odot$ reaches its critical mass $(M_c)_{\rm max}\sim10^9\,M_\odot$ [with $R^{99}_*\sim300\,{\rm pc}$] and collapses. The collapse of the quantum

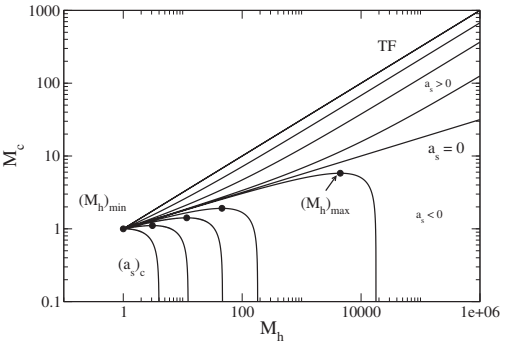


Fig. 3. Core mass M_c as a function of the halo mass M_h for different values of a_s and m such that the minimum halo mass $(M_h)_{\rm min}$ is fixed [see Eq. (44)]. The mass is normalized by $(M_h)_{\rm min}$ (typically $(M_h)_{\rm min} \sim 10^8\, M_\odot$). We have indicated the position of the minimum halo mass $(M_h)_{\rm min}$ (common origin) and the position of the maximum halo mass $(M_h)_{\rm max}$ (bullet) above which the quantum core becomes unstable and collapses (when $a_s < 0$) [see Eqs. (49) and (50)]. We have highlighted the curve corresponding to the noninteracting case $a_s = 0$ [see Eq. (45)] and the curve corresponding to the TF limit $a_s/a_*' \gg 1$ [see Eq. (47)].

core leads to a dense axion "star",[13–15] a bosenova,[10] or axion drops.[19,20f] However, we recall that values of f smaller than $4.22 \times 10^{15}\,$GeV are outside of the range $10^{16}\,{\rm GeV} \leq f \leq 10^{18}\,{\rm GeV}$ predicted by particle physics and cosmology ($10^{16}\,$GeV is the Grand Unified Theory scale and $10^{18}\,$GeV is the Planck scale),[21] so that, according to these constraints, we do not expect the collapse of the core to occur in realistic DM halos of mass $M_h < 10^{14}\, M_\odot$ (as discussed above and in Refs.[22,35,36]).

7. Conclusion

In these Proceedings, we have emphasized the importance of the maximum mass of dilute axion stars (i.e. nonrelativistic self-gravitating BECs with an attractive self-interaction) first identified in Ref.[8]. Above this maximum mass the star collapses leading to a dense axion "star", a black hole, a bosenova, or axion drops.[9–20] We have discussed some applications of these results in relation to DM halos.[22,35,36] Other applications have been discussed in Refs.[50–55].

Appendix A. Relativistic complex scalar field

In this Appendix, we discuss the main properties of a relativistic complex self-interacting SF, establish its hydrodynamic representation, consider the TF approximation, and determine its equation of state $P(\epsilon)$ for an arbitrary self-interaction

[f]It does not lead to a SMBH which would be obtained for much larger values of f (above the reduced Planck mass $10^{18}\,$GeV)[11,12] for which our previous assumptions do not apply.

potential $V(|\varphi|^2)$. For a $|\varphi|^4$ self-interaction, we justify the equation of state introduced by Colpi *et al.*[4] (see also Refs.[56,57]). We also justify the GPP equations (1) and (2) in the nonrelativistic limit.

Appendix A.1. *Klein-Gordon-Einstein equations*

We consider a relativistic complex SF $\varphi(x^\mu) = \varphi(x, y, z, t)$ which is a continuous function of space and time. It can represent the wavefunction of a relativistic BEC.[56,58] The total action of the system, which is the sum of the Einstein-Hilbert action of general relativity + the action of the SF, can be written as

$$S = \int \left(\frac{c^4}{16\pi G} R + \mathcal{L} \right) \sqrt{-g}\, d^4x, \tag{A.1}$$

where R is the Ricci scalar curvature, $\mathcal{L} = \mathcal{L}(\varphi, \varphi^*, \partial_\mu\varphi, \partial_\mu\varphi^*)$ is the Lagrangian density of the SF, and $g = \det(g_{\mu\nu})$ is the determinant of the metric tensor. We consider a canonical Lagrangian density of the form

$$\mathcal{L} = \frac{1}{2} g^{\mu\nu} \partial_\mu\varphi^* \partial_\nu\varphi - V_{\rm tot}(|\varphi|^2), \tag{A.2}$$

where the first term is the kinetic energy and the second term is minus the potential energy. The potential energy can be decomposed into a rest-mass energy term and a self-interaction energy term:

$$V_{\rm tot}(|\varphi|^2) = \frac{m^2c^2}{2\hbar^2}|\varphi|^2 + V(|\varphi|^2). \tag{A.3}$$

The least action principle $\delta S = 0$ with respect to variations $\delta\phi$ (or $\delta\phi^*$), which is equivalent to the Euler-Lagrange equation

$$D_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu\varphi)^*} \right] - \frac{\partial \mathcal{L}}{\partial \varphi^*} = 0, \tag{A.4}$$

yields the KG equation

$$\Box\varphi + 2\frac{dV_{\rm tot}}{d|\varphi|^2}\varphi = 0, \tag{A.5}$$

where $\Box = D_\mu\partial^\mu = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\, g^{\mu\nu}\partial_\nu)$ is the d'Alembertian operator. For a free massless SF ($V_{\rm tot} = 0$), the KG equation reduces to $\Box\varphi = 0$.

The least action principle $\delta S = 0$ with respect to variations $\delta g^{\mu\nu}$ yields the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{A.6}$$

where $R_{\mu\nu}$ is the Ricci tensor and $T_{\mu\nu}$ is the energy-momentum (stress) tensor of the SF given by

$$T^\nu_\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\nu\varphi)}\partial_\mu\varphi + \frac{\partial \mathcal{L}}{\partial(\partial_\nu\varphi^*)}\partial_\mu\varphi^* - g^\nu_\mu \mathcal{L}. \tag{A.7}$$

For the Lagrangian (A.2), we get

$$T_{\mu\nu} = \frac{1}{2}(\partial_\mu\varphi^*\partial_\nu\varphi + \partial_\nu\varphi^*\partial_\mu\varphi) - g_{\mu\nu}\mathcal{L}. \tag{A.8}$$

Equations (A.5) and (A.6) with Eq. (A.8) form the KGE equations. The conservation of the energy-momentum tensor $D_\mu T^{\mu\nu} = 0$ is automatically included in the Einstein equations through the contracted Bianchi identities.

Appendix A.2. *Hydrodynamic representation*

We can write the KG equation (A.5) under the form of hydrodynamic equations by using the de Broglie transformation (see, e.g., Refs.[56,58]). To that purpose, we write the SF as

$$\varphi = \frac{\hbar}{m}\sqrt{\rho}e^{im\theta/\hbar}, \tag{A.9}$$

where ρ is the pseudo rest-mass density[g] and $\theta = S_{\rm tot}/m$ is the action by unit of mass ($\sim$ phase). They satisfy

$$\rho = |\psi|^2 = \frac{m^2}{\hbar^2}|\varphi|^2 \quad \text{and} \quad \theta = \frac{\hbar}{2mi}\ln\left(\frac{\varphi}{\varphi^*}\right). \tag{A.10}$$

Substituting Eq. (A.9) into the Lagrangian density (A.2), we obtain

$$\mathcal{L} = \frac{1}{2}g^{\mu\nu}\rho\partial_\mu\theta\partial_\nu\theta + \frac{\hbar^2}{8m^2\rho}g^{\mu\nu}\partial_\mu\rho\partial_\nu\rho - V_{\rm tot}(\rho) \tag{A.11}$$

with

$$V_{\rm tot}(\rho) = \frac{1}{2}\rho c^2 + V(\rho). \tag{A.12}$$

The Euler-Lagrange equations for θ and ρ, resulting from the least action principle, are

$$D_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\theta)}\right] - \frac{\partial\mathcal{L}}{\partial\theta} = 0, \quad D_\mu\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\rho)}\right] - \frac{\partial\mathcal{L}}{\partial\rho} = 0. \tag{A.13}$$

They yield the equations of motion[56,58]

$$D_\mu\left(\rho\partial^\mu\theta\right) = 0, \tag{A.14}$$

$$\frac{1}{2}\partial_\mu\theta\partial^\mu\theta - \frac{\hbar^2}{2m^2}\frac{\Box\sqrt{\rho}}{\sqrt{\rho}} - V'_{\rm tot}(\rho) = 0. \tag{A.15}$$

The same equations are obtained by substituting the de Broglie transformation from Eq. (A.9) into the KG equation (A.5), and separating the real and the imaginary parts. Equation (A.14) can be interpreted as a continuity equation $D_\mu J^\mu = 0$ accounting for the conservation of the charge (or boson number) and Eq. (A.15)

[g]We stress that ρ is *not* the rest-mass density. It is only in the nonrelativistic regime $c \to +\infty$ that ρ coincides with the rest-mass density.

can be interpreted as a quantum relativistic Hamilton-Jacobi (or Bernoulli) equation with a relativistic covariant quantum potential

$$Q_{\rm dB} = \frac{\hbar^2}{2m}\frac{\Box\sqrt{\rho}}{\sqrt{\rho}}. \tag{A.16}$$

Appendix A.3. *TF approximation*

In the classical limit or in TF approximation ($\hbar \to 0$), the Lagrangian from Eq. (A.11) reduces to

$$\mathcal{L} = \frac{1}{2}g^{\mu\nu}\rho\partial_\mu\theta\partial_\nu\theta - V_{\rm tot}(\rho). \tag{A.17}$$

The Euler-Lagrange equations (A.13) yield the equations of motion

$$D_\mu\left(\rho\partial^\mu\theta\right) = 0, \tag{A.18}$$

$$\frac{1}{2}\partial_\mu\theta\partial^\mu\theta - V'_{\rm tot}(\rho) = 0. \tag{A.19}$$

The same equations are obtained by making the TF approximation in Eq. (A.15), i.e., by neglecting the quantum potential. Equation (A.18) can be interpreted as a continuity equation and Eq. (A.19) can be interpreted as a classical relativistic Hamilton-Jacobi (or Bernoulli) equation.

Assuming $V'_{\rm tot} > 0$, and using Eq. (A.19), we introduce the fluid quadrivelocity

$$u_\mu = -\frac{\partial_\mu\theta}{\sqrt{2V'_{\rm tot}(\rho)}}c, \tag{A.20}$$

which satisfies the identity $u_\mu u^\mu = c^2$. The energy-momentum tensor is given in the hydrodynamic representation, by

$$T^\nu_\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\nu\theta)}\partial_\mu\theta + \frac{\partial\mathcal{L}}{\partial(\partial_\nu\rho)}\partial_\mu\rho - g^\nu_\mu\mathcal{L}. \tag{A.21}$$

For the Lagrangian (A.17) we get

$$T_{\mu\nu} = \rho\partial_\mu\theta\partial_\nu\theta - g_{\mu\nu}\mathcal{L}. \tag{A.22}$$

This expression can also be obtained from Eq. (A.8) by using the de Broglie representation (A.9) and making the TF approximation. Using Eq. (A.20), we get

$$T_{\mu\nu} = 2\rho V'_{\rm tot}(\rho)\frac{u_\mu u_\nu}{c^2} - g_{\mu\nu}\mathcal{L}. \tag{A.23}$$

The energy-momentum tensor (A.23) can be written under the perfect fluid form

$$T_{\mu\nu} = (\epsilon + P)\frac{u_\mu u_\nu}{c^2} - Pg_{\mu\nu}, \tag{A.24}$$

where ϵ is the energy density and P is the pressure, provided that we make the identifications

$$P = \mathcal{L}, \qquad \epsilon + P = 2\rho V'_{\rm tot}(\rho). \tag{A.25}$$

Therefore, the Lagrangian plays the role of the pressure of the fluid. Combining Eq. (A.17) with the Bernoulli equation (A.19), we get

$$\mathcal{L} = \rho V'_{\rm tot}(\rho) - V_{\rm tot}(\rho). \tag{A.26}$$

Therefore, according to Eqs. (A.25) and (A.26), the energy density and the pressure derived from the Lagrangian (A.17) are given by[56,59]

$$\epsilon = \rho V'_{\rm tot}(\rho) + V_{\rm tot}(\rho) = \rho c^2 + \rho V'(\rho) + V(\rho), \tag{A.27}$$

$$P = \rho V'_{\rm tot}(\rho) - V_{\rm tot}(\rho) = \rho V'(\rho) - V(\rho), \tag{A.28}$$

where we have used Eq. (A.12) to get the second equalities. Eliminating ρ between Eqs. (A.27) and (A.28), we obtain the equation of state $P(\epsilon)$. On the other hand, Eq. (A.28) can be integrated into[56]

$$V(\rho) = \rho \int \frac{P(\rho)}{\rho^2}\, d\rho. \tag{A.29}$$

Equation (A.28) determines $P(\rho)$ as a function of $V(\rho)$ while Eq. (A.29) determines $V(\rho)$ as a function of $P(\rho)$.

Appendix A.4. *Nonrelativistic limit*

In the weak-field gravity limit of general relativity $\Phi/c^2 \ll 1$, using the simplest form of the Newtonian gauge, the line element is given by

$$ds^2 = c^2\left(1 + 2\frac{\Phi}{c^2}\right)dt^2 - \left(1 - 2\frac{\Phi}{c^2}\right)\delta_{ij}dx^i dx^j, \tag{A.30}$$

where $\Phi(\mathbf{r}, t)$ is the Newtonian potential. Making the Klein transformation (see, e.g., Ref.[58])

$$\varphi(\mathbf{r}, t) = \frac{\hbar}{m} e^{-imc^2 t/\hbar}\psi(\mathbf{r}, t) \tag{A.31}$$

in the KGE equations (A.5) and (A.6), where ψ is the pseudo wave function satisfying $\rho = |\psi|^2$, we obtain after simplification the general relativistic GP equation[58,60]

$$i\hbar c\,\partial^0\psi - \frac{\hbar^2}{2m}\Box\psi + \frac{1}{2}mc^2(g^{00} - 1)\psi + i\frac{\hbar c^2}{2}\Box t\,\psi - m\frac{dV}{d|\psi|^2}\psi = 0, \tag{A.32}$$

coupled to the Einstein equations expressed in terms of ψ. We note that $\Box t$ can be written as $\Box t = -\frac{1}{c}g^{\mu\nu}\Gamma^0_{\mu\nu}$, where $\Gamma^\sigma_{\mu\nu}$ are the Christoffel symbols.[58] Taking the nonrelativistic limit $c \to +\infty$ of these equations with the Newtonian gauge (A.30), we obtain the GPP equations (1) and (2) [see Refs.[58,61,62] for the details of the derivation]. For a complex SF, the potential $V(|\psi|^2)$ which occurs in the GP equation (1) coincides with the potential $V(|\varphi|^2)$ which occurs in the KG equation (A.5) with Eq. (A.3) up to the change of function from Eq. (A.31) leading to

$$|\varphi|^2 = \frac{\hbar^2}{m^2}|\psi|^2. \tag{A.33}$$

Appendix A.5. $|\varphi|^4$ *potential*

We consider a $|\varphi|^4$ (quartic) potential of the form

$$V(|\varphi|^2) = \frac{\lambda}{4\hbar c}|\varphi|^4, \tag{A.34}$$

where λ is the dimensionless self-interaction constant. For nonrelativistic BECs, the potential that occurs in the GP equation (1) is usually written as in Eq. (3). Substituting Eq. (A.33) into Eq. (3), we get

$$V(|\varphi|^2) = \frac{2\pi a_s m}{\hbar^2}|\varphi|^4. \tag{A.35}$$

Comparing Eqs. (A.34) and (A.35), we obtain[9]

$$\frac{\lambda}{8\pi} = \frac{a_s mc}{\hbar} = \frac{a_s}{\lambda_C}, \tag{A.36}$$

where $\lambda_C = \hbar/mc$ is the Compton wavelength of the bosons. For the $|\varphi|^4$ model the KG and GP equations can be written as

$$\Box\varphi + \frac{m^2c^2}{\hbar^2}\varphi + \frac{8\pi a_s m}{\hbar^2}|\varphi|^2\varphi = 0, \tag{A.37}$$

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi + m\Phi\psi + \frac{4\pi a_s\hbar^2}{m^2}|\psi|^2\psi = 0. \tag{A.38}$$

In terms of the pseudo rest-mass density [see Eqs. (A.9) and (A.10)], the potential (A.35) reads

$$V(\rho) = \frac{2\pi a_s\hbar^2}{m^3}\rho^2. \tag{A.39}$$

Substituting Eq. (A.39) into Eqs. (A.27) and (A.28) we obtain

$$\epsilon = \rho c^2\left(1 + \frac{6\pi a_s\hbar^2}{m^3c^2}\rho\right), \tag{A.40}$$

$$P = \frac{2\pi a_s\hbar^2}{m^3}\rho^2. \tag{A.41}$$

Eliminating ρ between these two equations, we get

$$P = \frac{m^3c^4}{72\pi a_s\hbar^2}\left(\sqrt{1 + \frac{24\pi a_s\hbar^2}{m^3c^4}\epsilon} \mp 1\right)^2. \tag{A.42}$$

This relativistic equation of state $P(\epsilon)$ was introduced by Colpi *et al.*[4] in the context of boson stars. It was studied in detail by Chavanis and Harko[5] in connection with general relativistic BEC stars and by Li *et al.*[63] and Suárez and Chavanis[59] in a BECDM cosmological context. This equation of state reduces to that of an $n = 1$ polytrope [see Eq. (A.41) with $\epsilon \sim \rho c^2$] at low densities $\rho \ll m^3c^2/|a_s|\hbar^2$

(nonrelativistic limit) and to the linear law $P \sim \epsilon/3$ similar to the equation of state of the radiation at high densities (ultrarelativistic limit).

Remark: Sometimes, a $|\varphi|^4$ self-interaction potential is written as

$$V(|\varphi|^2) = \frac{m^2}{2\hbar^4}\lambda_s|\varphi|^4, \tag{A.43}$$

where λ_s is the dimensional self-interaction constant. Comparing Eq. (A.43) with Eqs. (A.34) and (A.35) we get

$$\lambda_s = \frac{4\pi a_s\hbar^2}{m} = \frac{\lambda\hbar^3}{2m^2c}. \tag{A.44}$$

Appendix B. Relativistic real scalar field

In this Appendix, we discuss the main properties of a relativistic real SF, consider the instantonic potential of axions, take the nonrelativistic limit, and justify the GPP equations (1)-(3) with $a_s < 0$.

Appendix B.1. *Klein-Gordon-Einstein equations*

For a real SF described by a canonical Lagrangian, the KGE equations write

$$\Box\varphi + \frac{m^2c^2}{\hbar^2}\varphi + \frac{dV}{d\varphi} = 0, \tag{B.1}$$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{B.2}$$

where the energy-momentum tensor is given by

$$T_{\mu\nu} = \partial_\mu\varphi\partial_\nu\varphi - g_{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\partial_\rho\varphi\partial_\sigma\varphi - \frac{m^2c^2}{2\hbar^2}\varphi^2 - V(\varphi)\right]. \tag{B.3}$$

Appendix B.2. *Nonrelativistic limit*

In the nonrelativistic limit $c \to +\infty$ where the SF displays rapid oscillations, the KGE equations can be simplified by averaging over the oscillations. To that purpose, we write

$$\varphi(\mathbf{r},t) = \frac{1}{\sqrt{2}}\frac{\hbar}{m}\left[\psi(\mathbf{r},t)e^{-imc^2t/\hbar} + \psi^*(\mathbf{r},t)e^{imc^2t/\hbar}\right], \tag{B.4}$$

where the complex wave function $\psi(\mathbf{r},t)$ is a slowly varying function of time (the fast oscillations $e^{imc^2t/\hbar}$ of the SF have been factored out). This transformation allows us to separate the fast oscillations of the SF with pulsation $\omega = mc^2/\hbar$ caused by its rest mass from the slow evolution of $\psi(\mathbf{r},t)$. Using the simplest form of the Newtonian gauge [see Eq. (A.30)], substituting Eq. (B.4) into the KGE equations (B.1)-(B.3), averaging over the oscillations and taking the nonrelativistic limit

$c \to +\infty$, we obtain the GPP equations (1) and (2) [see Secs. II and III of Ref.[15] and Appendix A of Ref.[64] for the details of the derivation]. For a real SF, the (effective) potential $V(|\psi|^2)$ that occurs in the GP equation (1) is obtained from the potential $V(\varphi)$ that occurs in the KG equation (B.1) by first substituting Eq. (B.4) into $V(\varphi)$, then averaging over the oscillations. It is different from the potential obtained by directly replacing φ^2 by $(\hbar/m)^2|\psi|^2$ in the potential $V(\varphi)$.

Remark: General relativistic boson stars decribed by a real SF have a naked singularity at the origin and are always unstable.[65h] It is possible to construct regular solutions which are periodic in time[66,67] when $M < M_{\rm max}^{\rm GR} = 0.606\, M_P^2/m$ but, on a long timescale (which can exceed the age of the universe), these "oscillatons" are unstable and disperse to infinity or form a black hole. Their instability is basically due to the fact that the charge (boson number) is not conserved for a real SF. However, in the nonrelativistic limit, particle number conservation is approximately restored and Newtonian boson stars (like axion stars) can be stable (see Sec. 3).

Appendix B.3. *Instantonic potential of axions*

Axions are hypothetical pseudo-Nambu-Goldstone bosons of the Peccei-Quinn phase transition associated with a $U(1)$ symmetry that solves the strong charge parity (CP) problem of quantum chromodynamics (QCD). The axion is a spin-0 particle with a very small mass $m = 10^{-4}\,{\rm eV}/c^2$ and an extremely weak self-interaction (with a decay constant $f = 5.82 \times 10^{10}\,{\rm GeV}$) arising from nonperturbative effects in QCD. Axions have huge occupation numbers so they can be described by a classical relativistic quantum field theory with a real SF $\varphi(\mathbf{r}, t)$ whose evolution is governed by the Klein-Gordon-Einstein (KGE) equations. The instantonic potential of axions is

$$V(\varphi) = \frac{m^2 c f^2}{\hbar^3}\left[1 - \cos\left(\frac{\hbar^{1/2}c^{1/2}\varphi}{f}\right)\right] - \frac{m^2c^2}{2\hbar^2}\varphi^2, \tag{B.5}$$

where m is the mass of the axion and f is the axion decay constant. For this potential, the KG equation (B.1) takes the form

$$\Box\varphi + \frac{m^2 c^{3/2} f}{\hbar^{5/2}} \sin\left(\frac{\hbar^{1/2}c^{1/2}\varphi}{f}\right) = 0. \tag{B.6}$$

This is the general relativistic sine-Gordon equation. Considering the dilute regime $\varphi \ll f/\sqrt{\hbar c}$ (which is valid in particular in the nonrelativistic limit $c \to +\infty$)[i] and expanding the cosine term of Eq. (B.5) in Taylor series, we obtain at leading order the φ^4 potential

$$V(\varphi) = -\frac{m^2c^3}{24 f^2 \hbar}\varphi^4. \tag{B.7}$$

[h] This result is in agreement with the cosmic censorship conjecture which excludes spacetimes with naked singularities.

[i] According to Eq. (B.10), the axion decay constant f scales as $c^{3/2}$.

In that case, the KG equation (B.1) takes the form

$$\Box\varphi + \frac{m^2c^2}{\hbar^2}\varphi - \frac{m^2c^3}{6f^2\hbar}\varphi^3 = 0. \tag{B.8}$$

In the nonrelativistic limit, the effective potential $V(|\psi|^2)$ appearing in the GP equation (1) is given by[14,15]

$$V(|\psi|^2) = \frac{m^2cf^2}{\hbar^3}\left[1 - \frac{\hbar^3 c}{2f^2m^2}|\psi|^2 - J_0\left(\sqrt{\frac{2\hbar^3 c|\psi|^2}{f^2m^2}}\right)\right], \tag{B.9}$$

where J_0 is the Bessel function of zeroth order. If we keep only the first term in the expansion of Eq. (B.9), we obtain the $|\psi|^4$ potential

$$V(|\psi|^2) = -\frac{\hbar^3c^3}{16f^2m^2}|\psi|^4. \tag{B.10}$$

This approximation is valid for dilute axion stars satisfying $|\psi|^2 \ll f^2m^2/\hbar^3c$. We note that $V(|\psi|^2) \equiv \overline{V(\varphi)}$ is different from the expression that one would have naively obtained by directly substituting $\varphi^2 = (\hbar/m)^2|\psi|^2$ into Eq. (B.5). The difference is already apparent in the first term of the expansion of the potential which involves a coefficient $-1/16$ [see Eq. (B.10)] instead of $-1/24$ [see Eq. (B.7)]. They differ by a factor 2/3. This is because φ is a real SF. Therefore, substituting φ (exact) from Eq. (B.4) into $V(\varphi)$, *then* averaging over the oscillations, is different from substituting $\varphi^2 = (\hbar/m)^2|\psi|^2$ (already averaged over the oscillations) into $V(\varphi)$.

In general, a quartic potential is written as

$$V(\varphi) = \frac{\lambda}{4\hbar c}\varphi^4, \tag{B.11}$$

where λ is the dimensionless self-interaction constant. Comparing Eqs. (B.7) and (B.11), we find that

$$\lambda = -\frac{m^2c^4}{6f^2}. \tag{B.12}$$

On the other hand, comparing Eq. (B.10) with Eq. (3), we obtain

$$a_s = -\frac{\hbar c^3 m}{32\pi f^2}. \tag{B.13}$$

Equations (B.12) and (B.13) then yield

$$\frac{\lambda}{8\pi} = \frac{2a_s mc}{3\hbar}. \tag{B.14}$$

We note that the relation between λ and a_s is different for a real SF and for a complex SF (see Appendix A). They differ by a factor 2/3 for the reason indicated previously.

We note that the self-interaction constant λ or the scattering length a_s is negative, so that the φ^4 self-interaction term for axions is *attractive*. This attraction

is responsible for the collapse of dilute axion stars above the maximum mass from Eq. (20).[8,9] The next order φ^6 term in the expansion of the potential (B.5) has been considered in Refs.[14,15] and turns out to be repulsive. This repulsion, that occurs at high densities and which has a relativistic origin, may stop the collapse of dilute axion stars and lead to the formation of dense axion stars[13] (see, however, footnote a concerning their possible instability with respect to relativistic decay).

References

1. D.J. Kaup, Phys. Rev. **172**, 1331 (1968).
2. R. Ruffini, S. Bonazzola, Phys. Rev. **187**, 1767 (1969).
3. J.R. Oppenheimer, G.M. Volkoff, Phys. Rev. **55**, 374 (1939).
4. M. Colpi, S.L. Shapiro, I. Wasserman, Phys. Rev. Lett. **57**, 2485 (1986).
5. P.H. Chavanis, T. Harko, Phys. Rev. D **86**, 064011 (2012).
6. E.W. Kolb, I.I. Tkachev, Phys. Rev. D **49**, 5040 (1994).
7. I.I. Tkachev, Sov. Astron. Lett. **12**, 305 (1986).
8. P.H. Chavanis, Phys. Rev. D **84**, 043531 (2011).
9. P.H. Chavanis, Phys. Rev. D **94**, 083007 (2016).
10. D.G. Levkov, A.G. Panin, I.I. Tkachev, Phys. Rev. Lett. **118**, 011301 (2017).
11. T. Helfer *et al.*, JCAP **03**, 055 (2017).
12. F. Michel, I.G. Moss, Phys. Lett. B **785**, 9 (2018).
13. E. Braaten, A. Mohapatra, H. Zhang, Phy. Rev. Lett. **117**, 121801 (2016).
14. J. Eby, M. Leembruggen, P. Suranyi, L.C.R. Wijewardhana, JHEP **12**, 066 (2016).
15. P.H. Chavanis, Phys. Rev. D **98**, 023009 (2018).
16. L. Visinelli, S. Baum, J. Redondo, K. Freese, F. Wilczek, Phys. Lett. B **777**, 64 (2018).
17. J. Eby, M. Leembruggen, L. Street, P. Suranyi, L.C.R. Wijewardhana, Phys. Rev. D **100**, 063002 (2019).
18. E. Braaten, H. Zhang, Rev. Mod. Phys. **91**, 041002 (2019).
19. S. Davidson, T. Schwetz, Phys. Rev. D **93**, 123509 (2016).
20. E. Cotner, Phys. Rev. D **94**, 063503 (2016).
21. L. Hui, J. Ostriker, S. Tremaine, E. Witten, Phys. Rev. D **95**, 043541 (2017).
22. P.H. Chavanis, Phys. Rev. D **103**, 123551 (2021).
23. E. Seidel, W.M. Suen, Phys. Rev. Lett. **72**, 2516 (1994).
24. D. Lynden-Bell, Mon. Not. R. Astron. Soc. **136**, 101 (1967).
25. H.Y. Schive, T. Chiueh, T. Broadhurst, Nature Physics **10**, 496 (2014).
26. H.Y. Schive *et al.*, Phys. Rev. Lett. **113**, 261302 (2014).
27. B. Schwabe, J. Niemeyer, J. Engels, Phys. Rev. D **94**, 043513 (2016).
28. P. Mocz *et al.*, Mon. Not. R. Astron. Soc. **471**, 4559 (2017).
29. P. Mocz, L. Lancaster, A. Fialkov, F. Becerra, P.H. Chavanis, Phys. Rev. D **97**, 083519 (2018).
30. J. Veltmaat, J.C. Niemeyer, B. Schwabe, Phys. Rev. D **98**, 043509 (2018).
31. P. Mocz *et al.*, Phys. Rev. Lett. **123**, 141301 (2019).
32. P. Mocz *et al.*, Mon. Not. R. Astron. Soc. **494**, 2027 (2020).
33. J. Veltmaat, B. Schwabe, J.C. Niemeyer, Phys. Rev. D **101**, 083518 (2020).
34. P.H. Chavanis, Phys. Rev. D **100**, 083022 (2019).
35. P.H. Chavanis, Phys. Rev. D **100**, 123506 (2019).
36. P.H. Chavanis, Phys. Rev. D **101**, 063532 (2020).
37. P.H. Chavanis, Eur. Phys. J. Plus **132**, 248 (2017).
38. P.H. Chavanis, L. Delfini, Phys. Rev. D **84**, 043532 (2011).

39. M. Membrado, A.F. Pacheco, J. Sanudo, Phys. Rev. A **39**, 4207 (1989).
40. M. Membrado, J. Abad, A.F. Pacheco, J. Sañudo, Phys. Rev. D **40**, 2736 (1989).
41. J.W. Lee, I. Koh, Phys. Rev. D **53**, 2236 (1996).
42. J. Goodman, New Astronomy **5**, 103 (2000).
43. A. Arbey, J. Lesgourgues, P. Salati, Phys. Rev. D **68**, 023511 (2003).
44. C.G. Böhmer, T. Harko, J. Cosmol. Astropart. Phys. **06**, 025 (2007).
45. P.H. Chavanis, Phys. Rev. D **76**, 023004 (2007).
46. P.H. Chavanis, preprint.
47. R.Y. Chiao, E. Garmire, C.H. Townes, Phys. Rev. Lett. **13**, 479 (1964).
48. P.H. Chavanis, Eur. Phys. J. Plus **134**, 352 (2019).
49. P.A. Ruprecht, M.J. Holland, K. Burnett, M. Edwards, Phys. Rev. A **51**, 4704 (1995).
50. V. Desjacques, A. Kehagias, A. Riotto, Phys. Rev. D **97**, 023529 (2018).
51. K. Fujikura, M. Hertzberg, E. Schiappacasse, M. Yamaguchi, arXiv:2109.04283.
52. L.E. Padilla, T. Rindler-Daller, P. Shapiro, T. Matos, J.A. Vázquez, Phys. Rev. D **103**, 063012 (2021).
53. J. Chen, X. Du, E. Lentz, D. Marsh, J. Niemeyer, Phys. Rev. D **104**, 083022 (2021).
54. N. Glennon, C. Prescod-Weinstein, Phys. Rev. D **104**, 083532 (2021).
55. J. Eby, S. Shirai, Y.V. Stadnik, V. Takhistov, arXiv:2106.14893.
56. P.H. Chavanis, arXiv:2109.05963.
57. D.R. Karkevandi, S. Shakeri, V. Sagun, O. Ivanytskyi, arXiv:2109.03801.
58. P.H. Chavanis, T. Matos, Eur. Phys. J. Plus **132**, 30 (2017).
59. A. Suárez, P.H. Chavanis, Phys. Rev. D **95**, 063515 (2017).
60. T. Matos, A. Avilez, T. Bernal, P.H. Chavanis, Gen. Rel. Grav. **51**, 159 (2019).
61. A. Suárez, P.H. Chavanis, Phys. Rev. D **92**, 023510 (2015).
62. A. Suárez, P.H. Chavanis, J. Phys.: Conf. Series **654**, 012008 (2015).
63. B. Li, T. Rindler-Daller, P.R. Shapiro, Phys. Rev. D **89**, 083536 (2014).
64. P.H. Chavanis, Phys. Rev. D **102**, 083531 (2020).
65. P. Jetzer, D. Scialom, Phys. Lett. A **169**, 12 (1992).
66. E. Seidel, W.-M. Suen, Phys. Rev. Lett. **66**, 1659 (1991).
67. M. Alcubierre *et al.*, Class. Quantum Grav. **20**, 2883 (2003).

A dark matter solution for the XENON1T electron excess and the galactic center 511 keV line

Yasaman Farzan

School of Physics, Institute for Research in Fundamental Sciences (IPM)
P.O.Box 19395-5531, Tehran, Iran
E-mail: yasaman@theory.ipm.ac.ir
http://physics.ipm.ac.ir/

The excess of the 511 keV line from the Milky Way galactic bulge, confirmed by the INTEGRAL detector, is a longstanding mystery. The morphology of the line appears to be proportional to the square of the dark matter density, hinting towards a dark matter origin. On the other hand, in 2020, XENON1T has reported an excess of electrons with a recoil energy of $2-4$ keV. We present a model based on a dark matter of a few MeV mass that decays into a pair of pico-charged particles with a lifetime much larger than the age of the Universe. The magnetic field of the galaxy accumulates these relativistic pico-charged particles whose scattering on the electrons can explain the signal reported by XENON1T. The annihilation of the pico-charged particles in the galactic bulge leads to e^-e^+ production and therefore to an excess of the 511 keV line. We review the present observational bounds and the strategies to test the model.

Keywords: Dark matter, pico-charged particles, XENON1T electron excess, 511 keV line, galactic bulge

1. Introduction

Various observations such as galactic rotation curves, structure formation, dynamics of the galaxy clusters, collision of bullet clusters and anisotropies of Cosmic Mircrowave Background (CMB) point towards existence of a new form of matter all over the universe known as Dark Matter (DM). The constituent particles of DM should be either electrically neutral or their electric charge should be much smaller than the electron electric charge. Otherwise, they will emit photons and will not remain dark. The dark matter lifetime should be much longer than the age of the Universe. The observation of bullet clusters constrain the self-interaction of the dark matter particles indicating that they cannot have strong interactions. Last but not least, at the onset of structure formation, the dark matter should be non-relativistic. None of the Standard Model (SM) elementary particles satisfies these conditions. As a result, the observational hints for dark matter opens a window towards new physics beyond the standard model by introducing a new neutral (meta)stable massive particle that can play the role of dark matter.[a]

[a]The primordial blackholes as dark matter candidates have received popularity after observation of gravitational waves by LIGO and VIRGO. We will not however pursue this idea in this letter.

The models for dark matter candidates are quite diverse. Depending on the details of the model, different strategies for dark matter detection have been developed. In a wide class of models known as WIMP, where dark matter has a mass in the wide range of MeV to 100 TeV and it weakly interacts with the Standard Model (SM) particles[b], there are three approaches to dark matter detection: (1) direct dark matter detection, (2) indirect dark matter detection and (3) dark matter production in colliders. No conclusive evidence for dark matter detection has been reported from any of these experiments so far but every now and then, a signal is reported by indirect search experiments which may be interpreted as dark matter. Some of these signals disappear by collecting more data or are eventually proved to be artifacts. There are however signals that pass the test of time but after some period of time simply go out of fashion because the simplistic dark matter models developed to explain them become ruled out by independent observations. However, the dark sector does not need to be simplistic. Indeed, judging from the rich and dazzling structure of the SM, simplicity and minimalism is not the taste of the physics of the elementary particles. Going beyond simplistic dark matter models, it may be possible to resurrect dark matter explanation for the signal.

One example of the signals that has survived the test of time is the 511 keV line observed from galactic center which must come from the positronium (the e^-e^+ atom) decay: $e^- + e^+ \to \gamma + \gamma$. The intensity of the line indicates an excess of e^+ in the galactic center. Although the excess may come from the SM sources such as pulsars but entertaining a DM origin is too tantalizing to dismiss, especially that the morphology of the intensity of the line follows a ρ_{DM}^2 distribution, further pointing towards DM pair annihilation as the origin of the excess. In Ref.[1], it had been shown that dark matter with mass of a few MeV can explain the excess of the 511 keV line observed from the galactic center by pair annihilation into pairs of electron positron with a cross section of $\langle \sigma(\mathrm{DM} + \mathrm{DM} \to e^-e^+)v \rangle \sim O(10^{-38})$ cm^2 . This model implies huge energy pump into the plasma at the era of recombination when the DM density is high, through annihilation into e^-e^+. Such entropy pump can distort the CMB fluctuation pattern. The Planck data therefore rules this solution out.[2] Moreover, this model is disfavored by the lack of observation of the 511 keV line from dwarf galaxies.[3]

Ref.[4] proposes a dark matter model for the 511 keV line from the center of the Milky Way (MW) that can avoid the bounds from CMB and explain the lack of the 511 keV line from satellite dwarf galaxies. In this model, dark matter is composed of a scalar particle, X, with a mass of a few MeV which can be identified with the SLIM particle.[5,6] X decays into a pair of pico-charged particle $(C + \bar{C})$ with a lifetime much longer than the age of Universe: $\Gamma(X \to C + \bar{C})t_0 \ll 1$ in which t_0 is the age of the Universe. The magnetic field of the MW accumulates the C and

[b]Here, by weak interactions, we do not necessarily mean nuclear weak interaction, characterized by the SM W and Z gauge boson exchange.

$\bar{C}$ particles. The $C\bar{C}$ pairs eventually annihilate with each other producing e^-e^+ pairs. At the recombination and during the dark ages, the density of $C\bar{C}$ would be too small to pump energy to the plasma so the CMB bound can be avoided. Moreover, the magnetic field in the dwarf galaxies may be too weak to accumulate $C\bar{C}$ so the lack of the 511 keV line from them can be explained.

Ref.[4] predicted a signal at direct dark matter search experiments with low energy threshold. The relativistic C particles wandering around can impart a recoil energy of a few keV via Coulomb interaction to the electrons. In 2020, XENON1T reported an excess of electrons with recoil energy of 2–4 keV.[7] In Ref.[8], we discussed the parameter space where the model can explain both signals. Refs.[9] have also attempted to explain the XENON1T signal via DM.

This paper is organized as follows. In Sect. 2, we describe the model and the bounds that constrain its parameter space. In Sect. 6, we show how the model can explain the 511 keV line from the galactic bulge. In Sect. 4, we describe how the model can explain the XENON1T data. We show that the XENON1T electron recoil spectrum can also set the strongest upper bound on $f = \Gamma(X \to C\bar{C})t_0$. A summary is given in Sect. 5. In Sect. 5, we also discuss how the model can be tested with further collection of the 511 keV data from satellite galaxies and studying the correlation of the intensity with the magnetic field of the host [dwarf] galaxy. We also discuss the signatures in diffuse gamma ray background from the inflight annihilation of positron. We also comment on the positron flux to be detected by Voyager.

2. The model and its constraints

As mentioned in the introduction, the model is based on a scalar dark matter, X with a mass of a few MeV which decays into a pair of pico-charged particles, $C\bar{C}$. The X particle can be identified with the SLIM particle introduced in Ref.[5]. In the SLIM scenario, the DM is thermally produced via interaction with neutrinos and its abundance is set *a la* freeze-out scenario through lepton number violating annihilation into neutrino or antineutrino pair, $\langle\sigma(X+X \to \nu+\nu)v\rangle = \langle\sigma(X+X \to \bar{\nu}+\bar{\nu})v\rangle \sim pb$. Within the SLIM scenario, the annihilation to neutrino or antineutrino pair takes place via the t-channel exchange of a new Majorana neutrino which has a Yukawa coupling with the SM neutrino and X. A Z_2 symmetry, under which only the new Majorana neutrino and X are odd, guarantees the stability of X which is the lightest among the Z_2 odd particles. The same Z_2 symmetry forbids Yukawa coupling between the leptons, the Higgs and the new Majorana fermions and as a result, neutrinos cannot obtain a Dirac mass at the tree level. The SM neutrinos obtain a Majorana mass at one loop level within the SLIM scenario. The smallness of the neutrino mass is explained, thanks to the loop suppression despite a relatively large Yukawa coupling between, X and neutrinos of order of 10^{-3}. The model is testable by searching for the three body decay of Kaon and pion where along with the charged leptons, these new particles are emitted and

appear as missing energy.[10] Ref.[6] shows how the SLIM scenario can be embedded within a UV complete model. Combining the condition $\langle\sigma(X+X)v\rangle \sim pb$ (to obtain the required DM relic abundance via the freeze-out mechanism) with the one-loop contribution to the neutrino mass to be equal to $\sim \sqrt{\Delta m^2_{atm}}$ sets an upper bound of $\sim$ 10 MeV on the dark matter mass. On the other hand, X, being in thermal equilibrium with neutrinos at the time of the neutrino decoupling, cannot be lighter than a few MeV;[11] otherwise, it will lead to too many extra relativistic degrees of freedom. If X is not the SLIM and its production is not thermal, these lower and upper bounds on its mass can be relaxed.

Taking the pico-charged particles C and $\bar{C}$ to be scalars, the decay $X \to C\bar{C}$ can take place via a soft Z_2 symmetry breaking trilinear term of form $AX\bar{C}C$ with a very small A making X metastable. Let us now discuss how C and $\bar{C}$ acquire their minuscule electric charge. For this purpose, we need a new Abelian gauge symmetry such that the C and $\bar{C}$ are charged under the new $U_X(1)$ gauge symmetry. We denote the new gauge boson with A_μ. The kinetic and mass mixings between A_μ and the SM hypercharge gauge boson B_μ lead to a small electric charge for the C and $\bar{C}$ particles. We follow the notation of Ref.[12] and write

$$-\frac{A_{\mu\nu}A^{\mu\nu}}{4} - \frac{B_{\mu\nu}B^{\mu\nu}}{4} - \frac{\delta}{2}A_{\mu\nu}B^{\mu\nu} - \frac{1}{2}\left(\partial_\mu\sigma + M_1 A_\mu + \epsilon M_1 B_\mu\right)^2 \ , \qquad (1)$$

where $\delta, \epsilon \ll 1$. In the kinetic and mass basis, the three neutral gauge bosons are the SM γ, Z bosons and a new gauge boson A'_μ which we shall call dark photon. To the first approximation in ϵ and δ, the mass of A' is decoupled from m_Z and is equal to M_1. Up to $O(\epsilon^2, \delta^2)$, the coupling between the SM charged fermions, f and A' can be written as the following:

$$q' \bar{f}\gamma^\mu f A'_\mu \qquad \text{where} \qquad q' = e\cos\theta_W(\epsilon - \delta)Q_f, \qquad (2)$$

in which θ_W is the weak (Weinberg) mixing angle. The coupling of $C\bar{C}$ to A' is given by $g_X J^\mu_C A'_\mu$ where J^μ_C is the current of the C particles. Then, the C particle obtains an electric charge of

$$q_C = -g_X \epsilon \cos\theta_W .$$

From Eq. (2), we observe that in the limit $\epsilon \to \delta$, the SM fermions do not couple to A'. In this limit, if A' is lighter than the $C\bar{C}$ pair with a mass of order of a few keV, it will become metastable with a lifetime much greater than the age of the Universe.[4] This means the A' particles produced in the early universe remain as relics today and contribute as a subdominant DM component. In Ref.[8], we do not however consider this limit and allow for a faster decay of A'. As shown in Ref.[8], the $U_X(1)$ symmetry can be identified with the $L_\mu - L_\tau$ symmetry with a gauge coupling of $g_{\tau-\mu}$ which opens up the possibility of fast decay $A' \to \bar{\nu}_\mu\nu_\mu, \bar{\nu}_\tau\nu_\tau$. The decay can relax the bound from supernova cooling, opening up the window $m_{A'} \sim 100$ keV, $q' \sim 10^{-10} - 10^{-9}$ and $g_{\tau-\mu} \sim 10^{-9}$. Such values of coupling are too small to produce A' with significant abundance in the early universe so the bounds on the

extra relativistic degrees of freedom from CMB and BBN can be relaxed. On the other hand, the A' particles produced in the supernova core decay into neutrino and antineutrino pairs in the outer layers of supernova. This process can severely change the supernova evolution. It can even facilitate the shock revival. Thus, in future with better observation of supernovae and more advanced simulations, this part of the parameter range of the model can be tested. For the time being, however, this range of the parameter space is accepted within the uncertainties. The A' particles escaping the core decay into $\nu\bar{\nu}$ pairs so similarly to the standard picture, the binding energy of the star will be emitted in the form of neutrinos to be observed by detectors on the Earth such as Super-Kamiokande.

The C particles produced by the X decay will be relativistic at the production with a velocity far exceeding the escape velocity from the galaxy. They are kept in the galactic disc by the magnetic field. To accumulate the C and $\bar{C}$ particles, their Larmour radius,

$$r_c = (100\text{ pc})\frac{3\times 10^{-11}}{q_C}\frac{m_X}{10\text{ MeV}}\frac{1\ \mu\text{Gauss}}{B}$$

should be smaller than the thickness of the galactic disc. This requirement sets a lower bound on q_C: $q_C > 10^{-11}$, see Refs.[4,8]. The strongest upper bound on q_C in this mass range comes from supernova cooling consideration and is $q_C < 10^{-9}$, see Ref.[13].

As shown in Ref.[14], expanding supernova remnants can pump energy to the charged particles, enhancing their Larmor radius and expelling them from the galactic disc with a rate of $(100\text{ M}yr)^{-1}$ unless there is an efficient mechanism for energy loss for the charged particles. As discussed in Ref.[4], in the limit $\delta \to \epsilon$ where A' is (meta)stable, the relic background of A' can behave as a coolant. That is the C and $\bar{C}$ particles can scatter off the background A', losing energy. As discussed in Ref.[8], in general case $\delta \neq \epsilon$ where there is no A' background, another coolant should be introduced. Following Ref.[8], we denote the coolant with Y and take $10\text{ eV} < m_Y < 10\text{ keV}$, $n_Y|_{local} = (\rho_{DM}|_{local}/\langle\rho_{DM}\rangle)\langle n_Y\rangle$ with a coupling of g_Y to the C particles leading to the scattering cross section of $\sigma_S \sim g_Y^4/(4\pi E_C^2)$. The energy loss at each collision can be written as $\Delta E_C = m_Y v^2 (E_C/m_C)^2$.

The energy of the C particle at the production from $X \to C\bar{C}$ is equal to $m_X/2$. In the absence of energy pump from supernova, the time scale for the energy loss from $m_X/2$ to $m_C(1+v_f^2/2)$ is

$$\tau_E = \int_{m_C(1+v_f^2/2)}^{m_X/2} \frac{dE_C}{\Delta E_C}\frac{1}{\sigma_S v n_Y} \sim \frac{4\pi m_C^3}{g_Y^4 n_Y m_Y}\left(\frac{1}{v_f}-\frac{1}{v_i}\right). \tag{3}$$

Equating τ_E with the time scale of the energy gain from expanding supernova remnants (*i.e.,* 100 Myr), we find

$$v_f = 0.08\left(\frac{0.25}{g_Y}\right)^4\left(\frac{m_C}{3\text{MeV}}\right)^3\frac{0.1\times\rho_{DM}|_{local}}{n_Y m_Y}. \tag{4}$$

Let us now discuss the bounds on $\Gamma_X = \Gamma(X \to C\bar{C})$ or equivalently on $f = \Gamma_X t_0$. A rather strong bound comes from the requirement that annihilation $C\bar{C} \to A'A'$ does not deplete the C abundance of the galactic disc: That is $\sigma(C\bar{C} \to A'A')n_C v_f t_0 < 1$ which leads to

$$f < 10^{-2} \frac{m_X}{10 \text{ MeV}} \left(\frac{0.15}{g_X}\right)^4 \left(\frac{m_C}{5 \text{ MeV}}\right)^2 \frac{0.06}{v_f}.$$

A stronger bound comes from the requirement that the process of cooling during the history of the galaxy does not repel all the coolants. We should bear in mind that the coolant is a subdominant DM component which contributes less than 10 % to the whole dark matter content. Ref.[8] shows that this requirement implies

$$f < 10^{-3} \frac{m_X}{10 \text{ MeV}} \frac{m_Y}{10 \text{ keV}} \frac{10\rho_Y}{\rho_{DM}}.$$

As we shall see in Sect. 4, the bound from XENON1T detector on f is much stronger.

In order to produce e^-e^+ from $C\bar{C}$ annihilation yet another interaction is required. We will discuss this coupling in Sect. 6.

3. Explaining the 511 keV line

The 511 keV photon line has been detected from the galactic bulge of the MW for more than 40 years.[21] The INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL) is a space telescope launched in 2002. The data from the SPI spectrometer at INTEGRAL over the years has confirmed the 511 keV signal at 56 σ C.L. The narrow photon line peaked at 511 keV, whose energy coincides with the rest mass of electron or positron, should come from the annihilation of the electron positron pairs at rest which form positronium atoms. The source of such large positron abundance in the bulge of MW is subject to debate. An intriguing possibility can be annihilation of DM leading to the e^-e^+ production. Astrophysical sources such as pulsars or X-ray binaries may also contribute to the positron flux but their characteristic spectral shapes do not fit the observations.[15] Moreover, the morphology of the observed 511 keV line is better compatible with the ρ^2_{DM} ditribution than what is predicted from these point sources, hinting further toward a DM annihilation origin.[16]

Ref.[1] proposes DM with mass of few MeV annihilating to e^-e^+ as a solution to the 511 keV line. With an annihilation $\langle\sigma(DM + DM \to e^- + e^+)v\rangle \sim 10^{-2}$ pb, MeVish DM can account for the intensity of the 511 keV line. However, if the annihilation to e^-e^+ is a S-wave process, the reioinzing energy dump at recombination will cause delayed recombination which is in tension with the CMB data.[2] If the annihilation is P-wave, at the freeze-out temperatures, the annihilation cross section will be $\sim$10 pb, washing out the relic density of DM. Thus, we should go beyond this minimal scenario. In the following, we show that the model described in Sect. 2 can provide a reliable explanation.

The X DM particles in the galaxy decay into $C\bar{C}$ with a rate of Γ_X. If these C particles are accumulated close to their production point, their number density after t^0 will be

$$n_C = \frac{\rho_X}{m_X} f \quad \text{with} \quad f = \Gamma_X t^0 \tag{5}$$

in which f is the fraction of X particles that have decayed. According to the INTEGRAL findings,[21] the intensity of 511 keV line from galactic center is $\Phi_{511} = (0.96 \pm 0.07) \times 10^{-3}$ph cm^{-2} sec^{-1}. If each $C\bar{C}$ annihilation produces N_{pair} pairs of e^-e^+ and subsequently N_{pair} pairs of photons, the flux from the galaxy bulge will be

$$\Phi_{511} \simeq 2N_{pair} \frac{\int_0^{r_b} (n_X(r))^2 f^2 \langle \sigma_{C\bar{C}} v \rangle 4\pi r^2 dr}{4\pi r_{sol}^2} \tag{6}$$

where $r_{sol} \simeq 8$ kpc is the distance between us and the center of MW, $r_b \simeq 1$ kpc is the galactic bulge radius and $n_X = \rho_X(r)/m_X$. In the minimalistic dark matter scenario in which the dark matter pair annihilates directly into e^-e^+, the flux is given by the formula in Eq. (6), replacing $N_{pair}f^2$ with 1 and $\sigma_{C\bar{C}}$ with $\sigma(\mathrm{DM} + \mathrm{DM} \to e^-e^+)$. In this scenario, the cross section should be of order of $10^{-39} - 10^{-38}$ cm^2 to account for the 511 keV line intensity which implies that in our model:

$$\sigma_{C\bar{C}} \sim 2\ nb \left(\frac{1.5 \times 10^{-4}}{f} \right)^2 \left(\frac{m_X}{5\ \mathrm{MeV}} \right)^2 \frac{1}{N_{pair}}. \tag{7}$$

Let us now discuss the couplings that leads to the e^-e^+ production from the $C\bar{C}$ annihilation. Taking effective coupling $(\bar{e}e)(\bar{C}C)/\Lambda$, the cross section of 2 nb implies $\Lambda < 100$ GeV which would be within the reach of the LEP collider and other accelerator experiments. We therefore invoke the possibility of intermediate light neutral particles, ϕ, to avoid the bounds from the LEP. Then, the production of e^-e^+ takes place in two steps: First, $C\bar{C} \to \phi\bar{\phi}$ and subsequently $\phi \to e^-e^+$. As a result, $N_{pair} = 2$. The $\phi \to e^-e^+$ decay can take place through

$$g_\phi \phi \bar{e} e. \tag{8}$$

As long as $g_\phi < 10^{-11}$, the ϕ production in the supernova core will be negligible. We need $g_\phi > 10^{-15}$ in order for the ϕ decay length to be shorter than 10 pc. If ϕ travels less than ∼100 pc in the bulge (i.e., of order of the resolution of the INTEGRAL observatory) before decay, the n_{DM}^2 dependence of the 511 keV line morphology will be maintained, as favored by observation. Thus, $0.3 \times 10^{-15} < g_\phi < 10^{-11}$. ϕ is mainly an electroweak singlet mixed with the SM Higgs via

$$a_\phi \phi |H|^2$$

Then,

$$g_\phi = \sqrt{2} \frac{a_\phi v}{m_h^2} Y_e \tag{9}$$

in which Y_e is the Yukawa coupling of the SM Higgs to the electron. Fortunately, the hierarchy between the H and ϕ masses does not require unnatural fine tuned cancellation at tree level because $a_\phi \lesssim m_\phi$. In the early universe, ϕ can be produced via $e^-e^+ \to \phi\gamma$ with a number density $n_\phi \sim \langle\sigma_\phi v\rangle n_e^2 H^{-1}|_{T=m_\phi}$ where $\sigma_\phi = e^2 g_\phi^2/8\pi m_\phi^2$. Thus, during the period between $T = m_\phi$ and Γ_ϕ^{-1}, we can write $n_\phi/n_\gamma = 8\times10^{-8}(g_\phi/10^{-11})^2$. As a result, n_ϕ at $T \sim 1$ MeV would be small enough to satisfy the BBN bounds. In the range of g_ϕ in which we are interested, the decay of the ϕ particles takes place long before the recombination. For more details, see Ref.[4].

4. Direct detection of the C particles in the galaxy

The C particles, having a tiny electric charge, can have Coulomb interaction with the electrons and protons. It can also interact with matter fields by the t-channel exchange of dark photon, A'. The velocity of C arriving at the detector is larger than the typical velocity of the DM particles (*i.e.,* velocity of $C \gg 10^{-3}$). As a result, the recoil energy imparted by the C collision can be much larger than that expected from the collision of a DM particle of mass of MeV. This rises the hope to search for C particles by the direct dark matter search experiments, despite C being light. In this section, we shall show how the electron excess with recoil energy of $2-4$ keV at XENON1T can be explained within our model. We will then show that XENON1T sets strongest upper bound on the fraction of the X particle decaying into $C\bar{C}$, f.

According to Eq. (4), the velocity of C particles around us is $v_f \sim 0.08$. The Larmor radius in the Earth magnetic field is 5×10^8 km$(10^{-11}/q_C)(m_C/3$ MeV). Since the Larmor radius is much larger than the Earth radius, the Earth magnetic field cannot accumulate the C particles so their density around the Earth will be similar to anywhere in the solar system (i.e., anywhere in a distance of 8 kpc from the MW galactic center): $(\rho_{DM}|_{local}/m_X)f$ in which $\rho_{DM}|_{local} \sim 0.4$ GeV/cm^3.

Consider a C particle with a velocity of v_f scattering off a non-relativistic particle of mass, m. Up to corrections of $O(v_f^4)$, the maximum recoil energy, which corresponds to the backward scattering, can be written as

$$E_{max} = 2mv_f^2 \frac{m_C^2}{(m+m_C)^2}. \tag{10}$$

For $m_C \sim 1-5$ MeV, $v_f = 0.08$ and $m = m_e$, $E_{max} = 3-5.5$ keV which is tantalizingly within the range of electron recoil excess observed by XENON1T.[7] Taking m equal to the mass of Xenon or of Argon, E_{max} turns out to be much below the detection energy threshold of the experiments such as XENON1T,[17] DarkSide[18] and even the CRESST detector.[19]

The spectrum of recoiled electrons per detector mass will be

$$\frac{dN}{dE_r} = \frac{Z_{out}}{m_N}\left(2f\frac{\rho_X}{m_X}\right)\int f_C(v)\frac{d\sigma}{dE_r}vdv$$

where m_N is the mass of the nuclei inside the detector; *i.e.,* $m_N = 131$ GeV for the Xenon. Z_{out} denotes the number of the electrons in the outer orbitals of the detector atoms whose binding energies are smaller than E_r. Ref.[20] shows that to a reasonable approximation, the binding energy can be treated with a step function in computing the scattering cross section. For Xenon, $Z_{out} = 44$ which is the number of the electrons in the $n = 3, 4, 5$ orbitals. These outer electrons have velocities smaller than v_f so we neglect their velocity in our analysis. The velocity distributions of the C and $\bar{C}$ particles are given by $f_C(v)$. In our computation, we take $f_C(v) = \delta(v - v_f)$ for simplicity. As far as $|\epsilon|/2m_eE_r \gg |\epsilon - \delta|/(2m_eE_r + m_{A'}^2)$, the t-channel photon exchange gives the dominant contribution to the C particles scattering off the electrons. The differential cross section can be written as[24]

$$\frac{d\sigma}{dE_r} = \frac{e^2 q_C^2}{8\pi m_e v^2} \frac{1}{E_r^2} \left(1 + \frac{2m_eE_r}{2m_eE_r + m_{A'}^2} \frac{\delta - \epsilon}{\epsilon}\right)^2 \tag{11}$$

where $E_r < E_{max} = 2m_e v^2 \left(\frac{m_C}{m_C + m_e}\right)^2$. The first (second) term in the parenthesis comes from the contribution from t channel photon (dark photon, A') exchange.

Analyzing the XENON1T data, we take $v_f = 0.08$, $q_C = 10^{-11}$ and $m_X = 10$ MeV and use the data presented in Fig. 4 of Ref.[7]. Let us define χ^2 as follows

$$\chi^2 = \sum_{bins} \frac{[N_i^{obs} - N_i^{pred}]^2}{\sigma_i^2} \tag{12}$$

where N_i^{obs} is the observed number of events at bin i and N_i^{pred} is the prediction for the sum of the background and the signal from the C scattering in the ith bin. The values of the uncertainty (σ_i), the background and N_i^{obs} are obtained from Fig. 4 of Ref.[7]. We only focus on the first seven bins because the excess appears only at $E_r < 8$ keV. The variation of χ^2 with m_C seems to be only mild.

In Sect. 2, we discussed that for 100 keV $< m_{A'} <$ few MeV, identifying the $U_X(1)$ with the $L_\mu - L_\tau$ gauge symmetry, the bounds on $q' \propto (\delta - \epsilon)$ from supernova cooling can be relaxed because A' can decay into neutrino pairs relatively fast. For $m_{A'} \sim$ few MeV, $2m_eE_r \ll m_{A'}^2$ and we can treat $(\delta - \epsilon)/(\epsilon m_{A'}^2)$ and f as two free parameters to be fitted with the data. Taking $m_C = 5$ MeV, we find that the minimum of χ^2 lies at $f = 10^{-7}$ and $(\delta - \epsilon)/\epsilon = -2 \times 10^3 (m_{A'}/\text{MeV})^2$. For $\epsilon = -q_C/(g_X \cos\theta_W) \sim 10^{-11}/(g_X \cos\theta_W)$, this corresponds to $q' \sim (10^{-8}/g_X)(m_{A'}/\text{MeV})^2$. At this point, $\chi^2 = 4.7$ with $5 = 7 - 2$ degrees of freedom which corresponds to a one-sided goodness of fit of 45 %. The results are shown in Fig. 1 with a red curve. In the first bin, contribution from the dark photon and the SM photon cancel each other. In case of the green curve shown in Fig. 1, we have taken $2m_eE_r(\delta/\epsilon - 1)/m_{A'}^2 \to 0$ so there is no cancellation between dark and SM photons. As seen in the figure, in this case the spectrum diverges for small E_r.

For $m_{A'} \sim 100$ keV, $2m_eE_r$ is comparable to $m_{A'}^2$ and we should use the whole formula to fit the data. Taking $m_C = 5$ MeV and $m_{A'} = 100$ keV ($m_{A'} = 200$ keV),

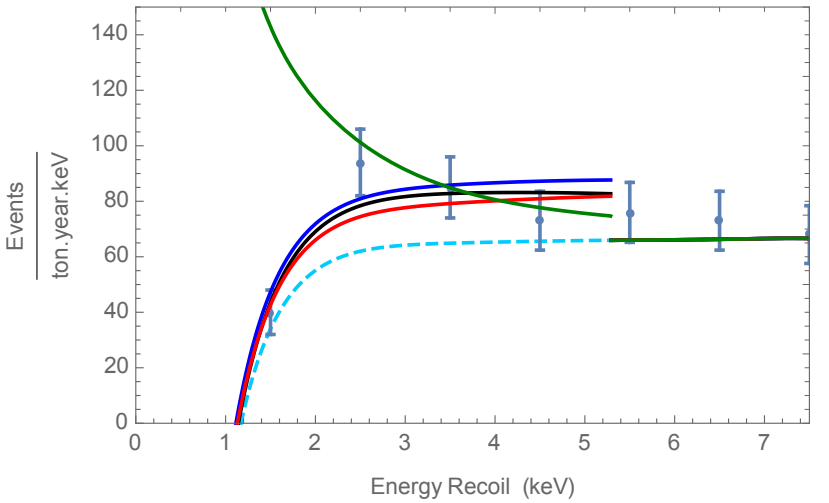


Fig. 1. XENON1T datapoints superimposed on our predictions. We have set $m_C = 5$ MeV corresponding to $E_{max} = 5.3$ keV. The red curve shows the prediction for $m_{A'} > 1$ MeV, $f = 10^{-7}$ and $(\delta-\epsilon)/\epsilon = -2\times10^3(m_{A'}/\text{MeV})^2$. The black curve depicts the prediction for $m_{A'} = 0.1$ MeV, $(\delta-\epsilon)/\epsilon = -13.9$ and $f = 6.6\times10^{-7}$. The blue curve illustrates the prediction for $m_{A'} = 0.2$ MeV, $(\delta-\epsilon)/\epsilon = -76$ and $f = 2\times10^{-7}$. The green curve shows the SM photon dominant regime with $f = 5\times10^{-6}$. The SM background is shown by the cyan dashed curve.

the best fit will correspond to $(\delta-\epsilon)/\epsilon = -13.9$ and $f = 6.6\times10^{-7}$ ($(\delta-\epsilon)/\epsilon = -76$ and $f = 2\times10^{-7}$) and to the minimum $\chi^2 = 4.11$ ($\chi^2 = 4.5$). From Fig. 1, we observe that both fits result in very close prediction. At this fit, q' is right above the supernova bound ($q' \sim 10^{-10} - 10^{-9}$) which can be ruled in by opening the possibility of decay of A' into neutrino pairs in the outer layers of the supernova as discussed in Sect. 2. The effects of such couplings can be tested by studying the flavor composition, the neutrino energy spectrum and the duration of the neutrino emission. The q' coupling can lead to the production of the A' particles in the early universe, contributing to the extra relativistic degrees of freedom, $\delta N_{eff} < 1$. More accurate δN_{eff} determination by future CMB and BBN studies can test this model with $q' \sim 10^{-10} - 10^{-9}$.

Let us now discuss the bound that can be derived on f from the XENON1T data. To do so, we consider the bins with energy less than 8 keV. Fig. 2 shows χ^2 as a function of f. Since the first bin is close to the detection threshold and it suffers from large uncertainty in the evaluation of the background from ^{214}Pb, we present results both including and excluding the first bin. We have taken $m_C = 4$ MeV, but the results are rather insensitive to the value of m_C. To draw the blue and green lines, we have set $\delta = \epsilon$ which means the dark photon decouples and does not contribute to the scattering. Drawing the orange line, we have chosen the value of $2m_e(\delta-\epsilon)/(m_{A'}^2\epsilon)$ such that the contributions from SM photon and dark photon cancel each other at $E_r = 1.5$ keV. From the figure we find an upper bound of

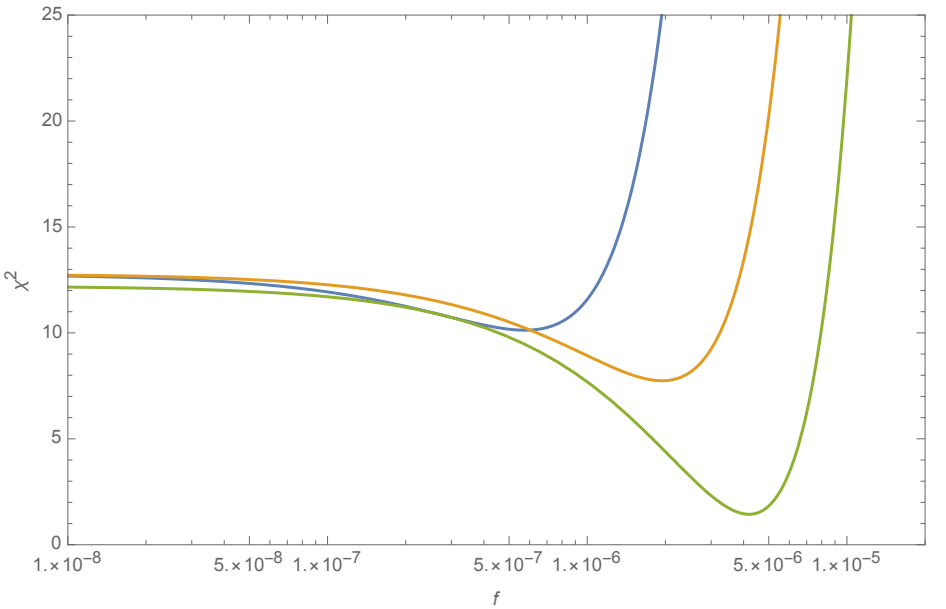


Fig. 2. χ^2 vs. f. We have taken $v_f = 0.08$, $q_c = 10^{-11}$, $m_X = 10$ MeV and $m_C = 4$ MeV. The blue and orange curves (the green curve) show χ^2 for the seven (six) bins between 1 keV$< E_r <$ 8 keV (2 keV$< E_r <$ 8 keV). To draw the blue and green curves, we have assumed that the photon exchange gives the dominant contribution to the scattering, $(m_e E_r/m_{A'}^2)[(\delta-\epsilon)/\epsilon] \ll 1$. To draw the orange curve we have set $(2m_e/m_{A'}^2)[(\delta-\epsilon)/\epsilon] = -1/(1.5\text{ keV})$.

1.5×10^{-6} on f at 3 σ, taking all the bins with $E_r < 8$ keV:

$$f < 1.5 \times 10^{-6} \quad \text{at } 3\sigma.$$

Excluding the first energy bin relaxes the bound to 10^{-5}. Notice that these bounds are much stronger than the constraints discussed in Sect. 2. Since $f = \Gamma_X t_0$, the upper bound on f is equivalent to a lower bound on the dark matter lifetime. The lifetime of X should be larger than $10^{15} - 10^{16}$ years.

Let us revisit the explanation for the 511 keV line taking into account this very stringent upper bound on f. To explain the intensity of the 511 keV line from the galactic bulge, the annihilation cross section of $C\bar{C} \to \phi\bar{\phi}$ should be $\text{mb}(10^{-7}/f)^2(m_X/5\text{ MeV})^2$. Taking a quartic coupling between these scalars of form $\lambda_{\phi C}|\phi|^2|C|^2$, we need $\lambda_{\phi C} \sim 0.02$ to obtain an annihilation cross section of mb.

5. Conclusion and discussion

We have presented a dark matter model that can simultaneously explain the 511 keV line from the galactic bulge and the electron excess with recoil energy of $2-4$ keV reported by XENON1T. The model is based on introducing a dark matter particle with a mass of $O(1-10)$ MeV that decays into a pair of pico-charged particles.

Although the velocities of the C and $\bar{C}$ particles are much larger than the escape velocity from the galaxy, the galactic magnetic field can accumulate them inside the galactic disk. The C particles with velocities of order of $0.08c$ can scatter off the electrons of the direct dark matter search experiments, imparting a recoil energy of a few keV which can explain the electron excess at the low energy bins reported by XENON1T. Moreover, the $C\bar{C}$ accumulated in the galactic bulge can annihilate with each other, giving rise to the excess of the 511 keV line. We have shown that the XENON1T data also sets strong upper bound of 10^{-6} on the fraction of X particles that have decayed into $C\bar{C}$ since early universe until today, $f < 10^{-6}$. The density of $C\bar{C}$ at the time of recombination and during the dark ages would be too low to pump any significant energy and ionize the atoms so the bounds from the CMB on the minimalistic DM explanation for the 511 keV line do not apply here.

The model enjoys a rich phenomenology and can be tested by a myriad of methods. Most obvious test is further data from upcoming direct dark matter search experiments, especially the XENONnT detector which is an upgrade of XENON1T. The XENONnT detector is in the construction phase.

The C particles obtain their electric charge through the mixing of the photon and the dark photon, A' which is the gauge boson of a new $U_X(1)$ symmetry under which C and $\bar{C}$ are charged. Simultaneous explanation for the excess of events with recoil energy between $2-4$ keV and for the absence of an excess at the 1 keV bin requires a cancellation between the contributions from the t-channel exchanges of the photon and A'. Such a cancellation is possible with a light A' with a mass of 100 keV–1 MeV. We have shown that by identifying the $U_X(1)$ symmetry with the $L_\mu - L_\tau$ symmetry and therefore allowing for $A' \to \nu_\mu \bar{\nu}_\mu, \nu_\tau \bar{\nu}_\tau$, it is possible to avoid the present bounds from the supernova cooling consideration. The A' decaying to $\nu\bar{\nu}$ in the outer layers of supernova can however alter the evolution of supernova, leaving its imprint on the duration of neutrino emission, flavor composition, energy spectrum and shock revival. All these signatures provide alternative methods to test the model.

The intensity of the 511 keV line from dwarf galaxies is proportional to the product of the number of C and $\bar{C}$ particles. As a result, there should be a correlation between the magnitude of the magnetic field in dwarf galaxies and the intensity of the 511 keV photon line emitted from them which can be tested in the future.

The positrons from the $C\bar{C}$ annihilation at the production are relativistic. A fraction of these positrons may annihilate in flight giving rise to a continuous γ ray spectrum on which there is a bound, implying that the C particles should be lighter than $6-15$ MeV depending on the models for ionization fraction in the galaxy.[21–23] In future with more data, lighter C can be probed by this consideration. Moreover, there will be an excess cosmic positron at few MeV range. At these energies, the solar wind renders measurements of cosmic positron flux in the MeV range obsolete within the solar system. Fortunately, there are detectors on board the Voyager spacecraft that have set an upper bound on the cosmic positrons outside the solar

system. The bound can be interpreted as $m_C < 20$ MeV. Further data from Voyager can test smaller values of m_C or hopefully detect a positron excess which will be a hint in favor of the present model.

Acknowledgments

The article is prepared for the proceedings of the sixteenth Marcel Grossmann meeting (MG16). The author would like to thank the organizers and conveners of this meeting for the kind invitation. She is also grateful for ICRANET whose support made this participation possible.

This project has received funding/support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement No 860881-HIDDeN. The author has also received partial financial support from Saramadan under contract No. ISEF/M/98223, No. ISEF/M/9916 and No. ISEF/M/400188.

References

1. C. Boehm, D. Hooper, J. Silk, M. Casse and J. Paul, Phys. Rev. Lett. **92** (2004) 101301 doi:10.1103/PhysRevLett.92.101301 [astro-ph/0309686].
2. R. J. Wilkinson, A. C. Vincent, C. Boehm and C. McCabe, Phys. Rev. D **94** (2016) no. 10, 103525 doi:10.1103/PhysRevD.94.103525 [arXiv:1602.01114 [astro-ph.CO]].
3. T. Siegert, R. Diehl, A. C. Vincent, F. Guglielmetti, M. G. H. Krause and C. Boehm, Astron. Astrophys. 595 (2016) A25 doi:10.1051/0004-6361/201629136 [arXiv:1608.00393 [astro-ph.HE]].
4. Y. Farzan and M. Rajaee, JHEP **12** (2017), 083 doi:10.1007/JHEP12(2017)083 [arXiv:1708.01137 [hep-ph]].
5. C. Boehm, Y. Farzan, T. Hambye, S. Palomares-Ruiz and S. Pascoli, Phys. Rev. D **77** (2008) 043516 doi:10.1103/PhysRevD.77.043516 [hep-ph/0612228].
6. Y. Farzan, Phys. Rev. D **80** (2009) 073009 doi:10.1103/PhysRevD.80.073009 [arXiv:0908.3729 [hep-ph]].
7. E. Aprile *et al.* [XENON], Phys. Rev. D **102** (2020) no. 7, 072004 doi:10.1103/PhysRevD.102.072004 [arXiv:2006.09721 [hep-ex]].
8. Y. Farzan and M. Rajaee, Phys. Rev. D **102** (2020) no. 10, 103532 doi:10.1103/PhysRevD.102.103532 [arXiv:2007.14421 [hep-ph]].
9. G. Choi, T. T. Yanagida and N. Yokozaki, [arXiv:2007.04278 [hep-ph]]; J. Hisano, A. Ibarra and R. Nagai, [arXiv:2007.03216 [hep-ph]]; D. Aristizabal Sierra, V. De Romeri, L. J. Flores and D. K. Papoulias, [arXiv:2006.12457 [hep-ph]]; G. Choi, T. T. Yanagida and N. Yokozaki, [arXiv:2007.04278 [hep-ph]]; S. Karmakar and S. Pandey, [arXiv:2007.11892 [hep-ph]]; L. A. Anchordoqui, I. Antoniadis, K. Benakli and D. Lust, [arXiv:2007.11697 [hep-th]]; J. Cao, X. Du, Z. Li, F. Wang and Y. Zhang, [arXiv:2007.09981 [hep-ph]]; C. W. Chiang and B. Q. Lu, [arXiv:2007.06401 [hep-ph]]; N. Okada, S. Okada, D. Raut and Q. Shafi, [arXiv:2007.02898 [hep-ph]]; W. DeRocco, P. W. Graham and S. Rajendran, [arXiv:2006.15112 [hep-ph]]; M. Lindner, Y. Mambrini, T. B. d. Melo and F. S. Queiroz, [arXiv:2006.14590 [hep-ph]]; K. Nakayama and Y. Tang, [arXiv:2006.13159 [hep-ph]]; G. Alonso-Álvarez, F. Ertas, J. Jaeckel, F. Kahlhoefer and L. J. Thormaehlen, [arXiv:2006.11243 [hep-ph]]; K. Kannike, M. Raidal, H. Veermäe, A. Strumia and D. Teresi, [arXiv:2006.10735 [hep-ph]];

P. Ko and Y. Tang, [arXiv:2006.15822 [hep-ph]]; L. Delle Rose, G. Hütsi, C. Marzo and L. Marzola, [arXiv:2006.16078 [hep-ph]]; U. K. Dey, T. N. Maity and T. S. Ray, [arXiv:2006.12529 [hep-ph]]; H. Alhazmi, D. Kim, K. Kong, G. Mohlabeng, J. C. Park and S. Shin, [arXiv:2006.16252 [hep-ph]]; P. Ko and Y. Tang, [arXiv:2006.15822 [hep-ph]]; L. Delle Rose, G. Hütsi, C. Marzo and L. Marzola, [arXiv:2006.16078 [hep-ph]]; Y. Jho, J. C. Park, S. C. Park and P. Y. Tseng, [arXiv:2006.13910 [hep-ph]]; B. Fornal, P. Sandick, J. Shu, M. Su and Y. Zhao, [arXiv:2006.11264 [hep-ph]]; Q. H. Cao, R. Ding and Q. F. Xiang, [arXiv:2006.12767 [hep-ph]]; Y. Chen, J. Shu, X. Xue, G. Yuan and Q. Yuan, [arXiv:2006.12447 [hep-ph]]; W. Cho, K. Y. Choi and S. M. Yoo, [arXiv:2007.04555 [hep-ph]]; H. J. He, Y. C. Wang and J. Zheng, [arXiv:2007.04963 [hep-ph]]; H. Davoudiasl, P. B. Denton and J. Gehrlein, [arXiv:2007.04989 [hep-ph]]; H. An and D. Yang, [arXiv:2006.15672 [hep-ph]]; J. Smirnov and J. F. Beacom, [arXiv:2002.04038 [hep-ph]]; N. F. Bell, J. B. Dent, B. Dutta, S. Ghosh, J. Kumar and J. L. Newstead, [arXiv:2006.12461 [hep-ph]]; W. Chao, Y. Gao and M. j. Jin, [arXiv:2006.16145 [hep-ph]]; S. Baek, J. Kim and P. Ko, [arXiv:2006.16876 [hep-ph]]; I. M. Bloch, A. Caputo, R. Essig, D. Redigolo, M. Sholapurkar and T. Volansky, [arXiv:2006.14521 [hep-ph]]; J. Bramante and N. Song, [arXiv:2006.14089 [hep-ph]]; H. M. Lee, [arXiv:2006.13183 [hep-ph]]; M. Baryakhtar, A. Berlin, H. Liu and N. Weiner, [arXiv:2006.13918 [hep-ph]]; D. Choudhury, S. Maharana, D. Sachdeva and V. Sahdev, [arXiv:2007.08205 [hep-ph]]; H. An and D. Yang, [arXiv:2006.15672 [hep-ph]]; Y. Ema, F. Sala and R. Sato, [arXiv:2007.09105 [hep-ph]]; S. Chigusa, M. Endo and K. Kohri, [arXiv:2007.01663 [hep-ph]]; A. Das and M. Sen, [arXiv:2104.00027 [hep-ph]]; S. Vagnozzi, L. Visinelli, P. Brax, A. C. Davis and J. Sakstein, [arXiv:2103.15834 [hep-ph]]; H. J. He, Y. C. Wang and J. Zheng, [arXiv:2012.05891 [hep-ph]]; A. Aboubrahim, M. Klasen and P. Nath, JHEP **02** (2021), 229 doi:10.1007/JHEP02(2021)229 [arXiv:2011.08053 [hep-ph]]; R. Harnik, R. Plestid, M. Pospelov and H. Ramani, Phys. Rev. D **103** (2021) no. 7, 075029 doi:10.1103/PhysRevD.103.075029 [arXiv:2010.11190 [hep-ph]]; B. Salehian, M. A. Gorji, H. Firouzjahi and S. Mukohyama, Phys. Rev. D **103** (2021) no. 6, 063526 doi:10.1103/PhysRevD.103.063526 [arXiv:2010.04491 [hep-ph]]; A. N. Khan, Phys. Lett. B **819** (2021), 136415 doi:10.1016/j.physletb.2021.136415 [arXiv:2008.10279 [hep-ph]]; A. Karozas, S. F. King, G. K. Leontaris and D. K. Papoulias, Phys. Rev. D **103** (2021) no. 3, 035019 doi:10.1103/PhysRevD.103.035019 [arXiv:2008.03295 [hep-ph]]; V. Brdar, A. Greljo, J. Kopp and T. Opferkuch, JCAP **01** (2021), 039 doi:10.1088/1475-7516/2021/01/039 [arXiv:2007.15563 [hep-ph]].

10. Y. Farzan, Mod. Phys. Lett. A **25** (2010), 2111-2120 doi:10.1142/S0217732310034018 [arXiv:1009.1234 [hep-ph]].
11. N. Sabti, J. Alvey, M. Escudero, M. Fairbairn and D. Blas, JCAP **01** (2020), 004 doi:10.1088/1475-7516/2020/01/004 [arXiv:1910.01649 [hep-ph]].
12. D. Feldman, Z. Liu and P. Nath, Phys. Rev. D **75** (2007), 115001 doi:10.1103/PhysRevD.75.115001 [arXiv:hep-ph/0702123 [hep-ph]].
13. S. Davidson, S. Hannestad and G. Raffelt, JHEP **05** (2000), 003 doi:10.1088/1126-6708/2000/05/003 [arXiv:hep-ph/0001179 [hep-ph]].
14. L. Chuzhoy and E. W. Kolb, JCAP **07** (2009), 014 doi:10.1088/1475-7516/2009/07/014 [arXiv:0809.0436 [astro-ph]].
15. R. M. Bandyopadhyay, J. Silk, J. E. Taylor and T. J. Maccarone, Mon. Not. Roy. Astron. Soc. **392** (2009) 1115 doi:10.1111/j.1365-2966.2008.14113.x [arXiv:0810.3674 [astro-ph]].
16. R. Diehl, D. Hartmann, P. Hoppe and N. Prantzos. (2002). Astronomy with Radioactivities, Springer.

17. E. Aprile *et al.* [XENON], Phys. Rev. Lett. **121** (2018) no. 11, 111302 doi:10.1103/PhysRevLett.121.111302 [arXiv:1805.12562 [astro-ph.CO]].
18. P. Agnes te al. (DarkSide-50), Phys. Rev. Lett. 121 (2018) 081307.
19. J. Schieck, G. Angloher, A. Bento, C. Bucci, L. Canonica, X. Defay, A. Erb, F. v. Feilitzsch, N. Ferreiro Iachellini, P. Gorla, A. Gütlein, D. Hauff, J. Jochum, M. Kiefer, H. Kluck, H. Kraus, J. C. Lanfranchi, J. Loebell, M. Mancuso, A. Münster, C. Pagliarone, F. Petricca, W. Potzel, F. Pröbst, R. Puig, F. Reindl, K. Schäffner, S. Schönert, W. Seidel, M. Stahlberg, L. Stodolsky, C. Strandhagen, R. Strauss, A. Tanzke, H. H. T. Thi, C. Türkoglu, M. Uffinger, A. Ulrich, I. Usherov, S. Wawoczny, M. Willers, M. Wüstrich and A. Zöller, PoS **ICHEP2016** (2016), 217 doi:10.22323/1.282.0217 [arXiv:1611.02113 [astro-ph.CO]].
20. C. C. Hsieh, L. Singh, C. P. Wu, J. W. Chen, H. C. Chi, C. P. Liu, M. K. Pandey and H. T. Wong, Phys. Rev. D **100** (2019) no. 7, 073001 doi:10.1103/PhysRevD.100.073001 [arXiv:1903.06085 [hep-ph]].
21. T. Siegert, R. Diehl, G. Khachatryan, M. G. H. Krause, F. Guglielmetti, J. Greiner, A. W. Strong and X. Zhang, Astron. Astrophys. **586** (2016), A84 doi:10.1051/0004-6361/201527510 [arXiv:1512.00325 [astro-ph.HE]].
22. J. F. Beacom and H. Yuksel, Phys. Rev. Lett. **97** (2006), 071102 doi:10.1103/PhysRevLett.97.071102 [arXiv:astro-ph/0512411 [astro-ph]].
23. P. Sizun, M. Casse and S. Schanne, Phys. Rev. D **74** (2006), 063514 doi:10.1103/PhysRevD.74.063514 [arXiv:astro-ph/0607374 [astro-ph]].
24. R. Harnik, Z. Liu and O. Palamara, JHEP **07** (2019), 170 doi:10.1007/JHEP07(2019)170 [arXiv:1902.03246 [hep-ph]].

Preliminary results of rich galaxy clusters' spatial distribution analysis on CfA2 Redshift Survey data: Compact objects or dark matter presence at redshift less 0.032

I. V. Arkhangelskaja

Division of Nuclear Physics and Technologies, National Research Nuclear University MEPhI, Kashirskoe shosse, 31, Moscow, 115409, Russia
E-mail: IVArkhangelskaya@mephi.ru
https://home.mephi.ru/en/users/1525/public

A. M. Galper

Nuclear Physics and Astrophysics Division, Lebedev Physical Institute, Russian Academy of Sciences, Leninskiy prospekt 53, Moscow, 119991, Russia
Division of Nuclear Physics and Technologies, National Research Nuclear University MEPhI, Kashirskoe shosse, 31, Moscow, 115409, Russia

L. N. Khanh and D. N. Dorosheva

Division of Nuclear Physics and Technologies, National Research Nuclear University MEPhI, Kashirskoe shosse, 31, Moscow, 115409, Russia

Preliminary results of the investigation of the properties of 13 clusters of galaxies from CfA2 redshift survey are discussed in the presented article. The distributions on absolute magnitude and luminosity represent two areas for clusters ##88, 1101, 1046, 142, 933, 1242, 1652, 107, 150, 316, 317, 961, 977. Redshifts of these clusters are in the region 0.002 – 0.032. The distributions on groups members position, absolute magnitude and luminosity represent two areas for these clusters. Galaxies from these areas are paired accordingly its spectral characteristics and position. Also several anomalies of spatial dynamic of galaxies in these clusters were separated. Such structure could be caused by dark matter presence inside cluster in configuration similar to Zeldovich pancake or gravitational lensing on compact object or dark matter blob located between galaxy cluster and observer. Several peculiarities have found on the spatial distributions of galaxies in clusters ##933, 142, 1046, and 1652. Moreover, these groups reveals associations with high-energy gamma-emission sources on Fermi/LAT 10-Year Point Source Catalog 4FGL DR2 data (4FGLJ1144.9+1937, 4FGLJ0152.2+3714, 4FGLJ1230.8+1223 and 4FGLJ1653.8+3945 correspondingly).These sources are active galaxies 3C 264, B2 0149+37, M87 and MRC 501. Furthermore, 3C 264 and M87 observed in subTeV energy band by VERITAS data. Joint observations of such clusters by orbital gamma-ray observatories with high angular resolution and ground-based Cherenkov air-shower experiments could possibly clarify the type of gravitational lensing and processes of particle acceleration in these objects especially highest energy of emitted gammas. Thus we propose including these and similar clusters in the programs of observations of the planned experiment GAMMA-400 (Gamma Astronomical Multifunctional Modular Apparatus) with angular resolution $\sim 0.01^\circ$ at $E_\gamma = 100\,GeV$ and several TeV upper energy band. Also now it is discussed coordination of multiwavelength observations program of Cherenkov Telescope Array (CTA) and GAMMA-400 objects list for observations.

Keywords: Clusters of galaxies; CfA2 redshift survey.

1. Introduction

Galaxy clusters are the biggest gravitationally bound systems[1,2] reveal an independent approach to the problems related to the distribution of matter at various redshift ranges – particulars given in Ref. 4 and Ref. 5. It possible to use the results of its characteristics investigation to understand processes of formation of such systems and galaxies contained its possibly connecting dark matter halos (DMH) and to constrain cosmological models – see, for example Ref. 6 and Ref. 5. The differences of galaxies were identified by their luminosity, morphology, colour, etc. Moreover, following to Ref. 3 and Ref. 2, it allowed studying the relation between the topology and properties of galaxies at the large-scaling structure and concluding both cosmological parameters and galaxy formation mechanisms. Also the investigation of galaxies clusters relates to the dependence of the characteristics of galaxies on the surrounding medium properties, and the large-scale structure formation. Thus the cluster system dynamic investigation also should estimate possible DM presence due to analysis of several catalogues – for instance, CfA2 Redshift Survey.[8,14,17] Therefore, the large-scale distribution of matter in the Metagalaxy will be described by analysing quantities of galaxies to various limiting magnitudes and the redshift surveys – see, for example Ref. 5 and Ref. 16.

2. The Second CfA Redshift Survey data analysis

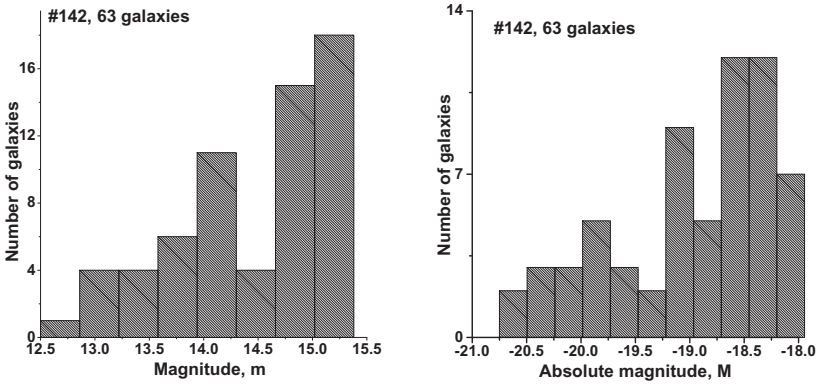
Preliminary results of the investigation of the properties of clusters of galaxies from The Second CfA Redshift Survey are discussed in the presented article.

The second CfA redshift survey was started between 1985 and 1995 due measurements of relative distances via redshifts for about 18000 bright galaxies in the northern sky. It contain data of 1971 galaxies groups (totally 6787 members) at galactic latitudes $b \geq 20^\circ$ – see Ref. 17.

Early data from this catalogue were analysed in Ref. 15 and Ref. 14 and it were found several peculiarities for clusters ##933, 1242, 88, 142, 1046, 1101 in distributions of group members on absolute magnitude, angular velocity, etc.

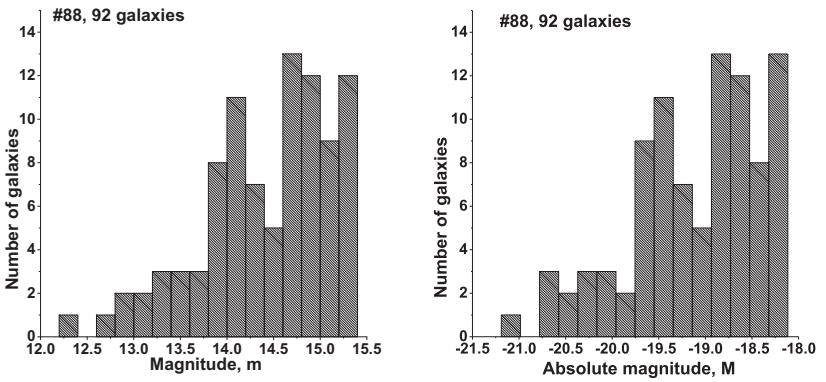
The results of additional analysis of these clusters characteristics are discussed in the presented article. The mean values of it's characteristics are listed in the Table 1 (rows 1 – 5). Several additional peculiarities have found on the distributions of galaxies inside these clusters on magnitude and absolute magnitude on CfA2 data. The examples of such distributions for clusters of galaxies #142 and #88 are presented at figures Fig. 1(a), Fig. 1(b) and Fig. 2(a), Fig. 2(b) correspondingly. Dual structure was separated at distributions of galaxies inside cluster on magnitude and absolute magnitude for these two clusters. Bifurcation points are $m_{142} = 14.5 \pm 0.2$ and $M_{142} = 19.5 \pm 0.2$ for group #142, also $m_{88} = 14.5 \pm 0.2$ and $M_{88} = 19.0 \pm 0.2$ for one #88.

Than we have separate clusters of galaxies contain more than 20 members for subsequent analysis. The mean values of characteristics for other 8 clusters from this subsample are listed in the Table 1 (rows 6 – 13). Fig. 3(a), Fig. 4(a) and



(a) (b)

Fig. 1. Several peculiarities on the distributions of galaxies inside cluster #142 on CfA2 data. (a) Graph for distribution of galaxies on magnitude. (b) Graph for distribution of galaxies on absolute magnitude.



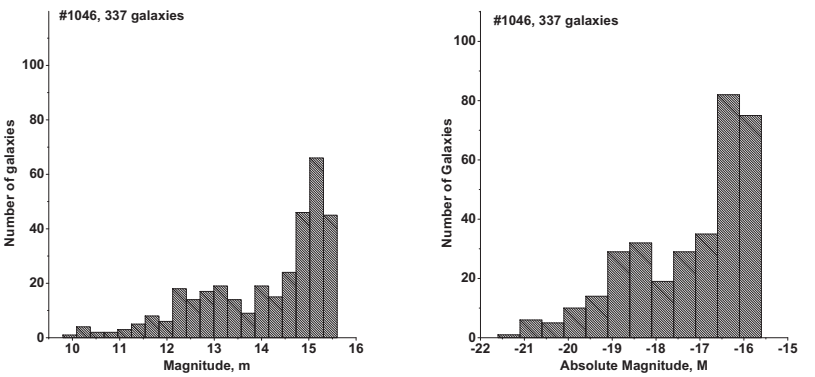
(a) (b)

Fig. 2. Several peculiarities on the distributions of galaxies inside cluster #88 on CfA2 data. (a) Graph for distribution of galaxies on magnitude. (b) Graph for distribution of galaxies on absolute magnitude.

Fig. 6(a) shows the other examples of distributions of galaxies inside cluster with members numbers both than 50 and 20 on magnitude for groups #1046, #933 and #1652 correspondingly. Graphs Fig. 3(b), Fig. 4(b) and Fig. 6(b) represents the distributions of galaxies in these clusters on absolute magnitude. Dual structure also was separated at distributions of galaxies inside cluster on magnitude and absolute magnitude for all investigated clusters without any correlation with members numbers. The obtained bifurcation points in the distributions of galaxies in these clusters on magnitude are $m_{1046} = 13.8 \pm 0.2$, $m_{933} = 14.3 \pm 0.2$, $m_{1652} = 14.9 \pm 0.2$ and ones on absolute magnitude are $M_{1046} = -17.9 \pm 0.2$, $M_{933} = -19.8 \pm 0.2$, $M_{1652} = -20.0 \pm 0.2$.

Table 1. The mean values of characteristics for several clusters of galaxies on CfA2 data.

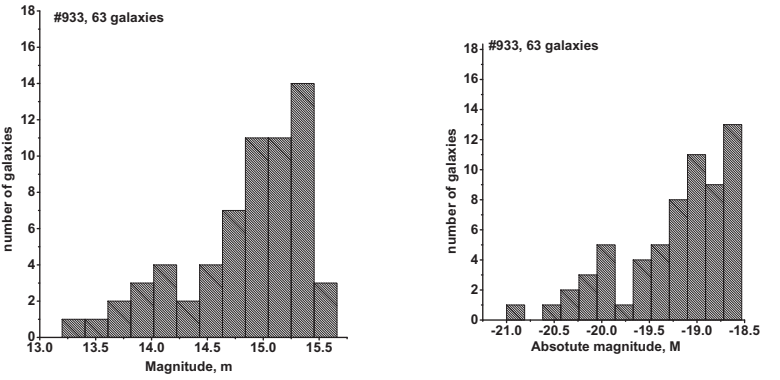
Cluster number	Amount of galaxies	RQF – Relative quantities false (%)	Heliocentric velocity (km/s)	Mean distance of member of group from its centre (Mpc)
1046	337	2.328	1847 ± 519	1.423
1101	118	0.784	7433 ± 751	1.006
88	92	1.426	5040 ± 440	1.357
933	63	0.372	6656± 703	0.497
142	63	0.721	4868± 496	0.667
1242	26	0.406	2750 ± 190	0.384
1652	28	1.561	9324 ± 427	1.394
107	34	0.303	1530 ± 395	0.354
150	20	1.629	4964 ± 323	0.821
316	21	0.294	6068 ± 269	0.450
317	23	0.504	6152 ± 288	0.564
961	27	2.422	1264 ± 167	0.481
977	20	0.485	6288 ± 402	0.407



(a) (b)

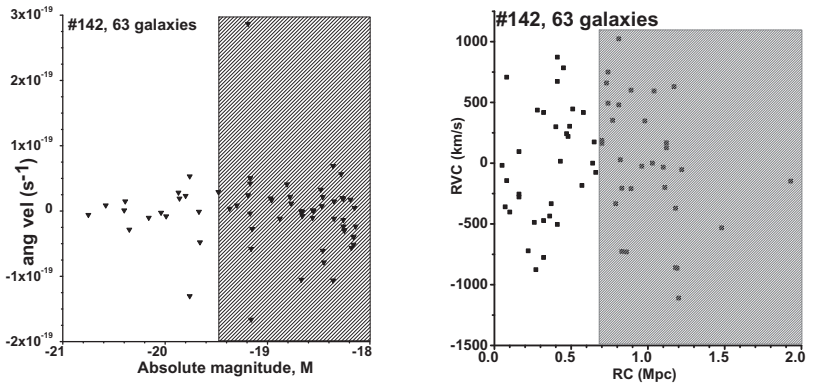
Fig. 3. Several peculiarities on the distributions of galaxies inside cluster #1046 on CfA2 data. (a) Graph for distribution of galaxies on magnitude. (b) Graph for distribution of galaxies on absolute magnitude.

Moreover several peculiarities have found on the distributions on angular velocity and absolute magnitude together with distributions on velocity and distance to cluster centre. Such distributions for clusters of galaxies #142 and #88 are presented at Fig. 7(a), Fig. 7(b) and Fig. 8(a), Fig. 8(b) correspondingly. Dual structure was separated on the distributions of galaxies inside cluster on angular velocity and absolute magnitude for these two clusters with bifurcation points similar to results of analysis of Fig. 1(b) and Fig. 2(b). Also dual structure appear on the distributions on velocity and distance to cluster centre with bifurcation points $Rc_{142} = 0.70 \pm 0.05$ and $Rc_{88} = 1.00 \pm 0.05$ for clusters #142 and #88.



(a) (b)

Fig. 4. Several peculiarities on the distributions of galaxies inside cluster #933 on CfA2 data. (a) Graph for distribution of galaxies on magnitude. (b) Graph for distribution of galaxies on absolute magnitude.

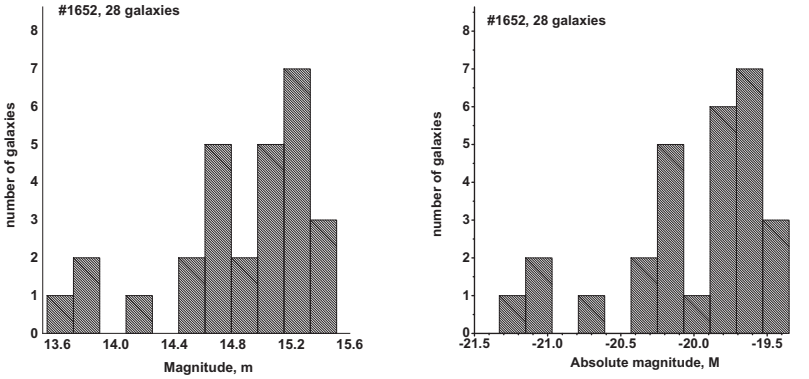


(a) (b)

Fig. 5. Several peculiarities on the distributions of galaxies inside cluster #142 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre.

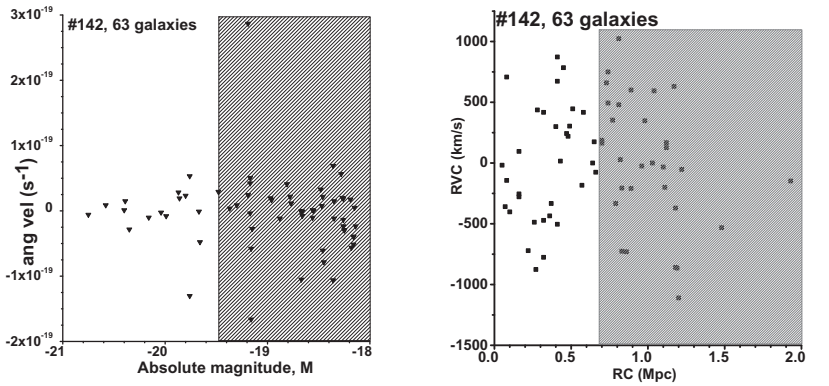
The distributions on angular velocity and absolute magnitude groups #1046, #933 and #1652 are presented at Fig. 9(a), Fig. 10(a) and Fig. 6(a) correspondingly. Dual structure reveals on these figures with bifurcation points similar to results of analysis of Fig. 3(a), Fig. 4(a) and Fig. 6(a) look like clusters #142 and #88. Moreover dual structure appear on the distributions on velocity and distance to cluster centre with bifurcation points $Rc_{1046} = 1.7 \pm 0.1$, $Rc_{933} = 0.52 \pm 0.05$ and $Rc_{1652} = 1.5 \pm 0.1$ for clusters #1046, #933 and #1652 correspondingly.

The spatial distribution of galaxies in listed in the Table 1 clusters were analysed. Figure 12(a) represents such distribution for cluster #1046. Dual structure



(a) (b)

Fig. 6. Several peculiarities on the distributions of galaxies inside cluster #1652 on CfA2 data. (a) Graph for distribution of galaxies on magnitude. (b) Graph for distribution of galaxies on absolute magnitude.

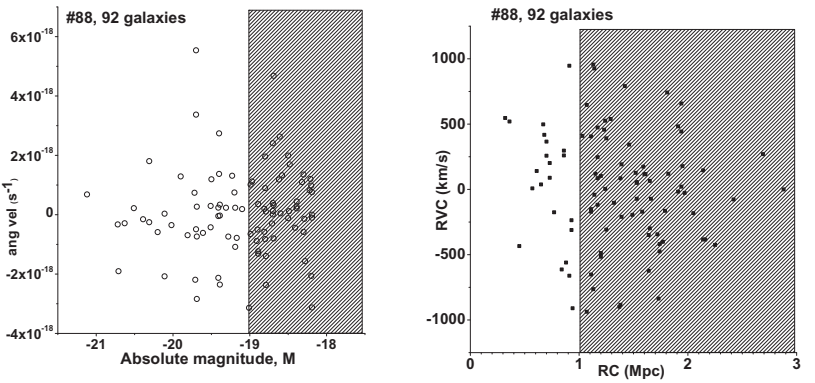


(a) (b)

Fig. 7. Several peculiarities on the distributions of galaxies inside cluster #142 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre. Shaded regions correspond areas from right side of bifurcation points on Fig. 1.

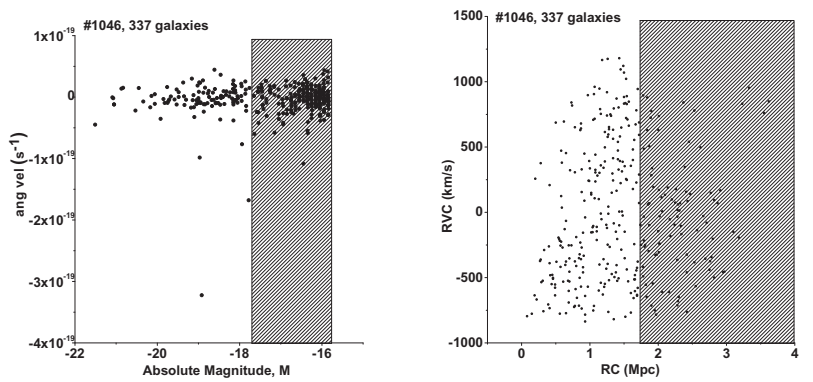
were separated inside this cluster accordingly to preliminary analysis rezults — see Fig. 12(b). Similar peculiarities have found on the spatial distributions for other clusters. The examples are presented at Fig. 13, Fig. 14 and Fig. 15.

The results of preliminary data analysis have shown that the distributions on groups members position, magnitude, absolute magnitude, angular velocity and absolute magnitude also with ones on velocity and distance to cluster centre and represent two areas for clusters ##88, 1101, 1046, 142, 933, 1242, 1652, 107, 150, 316, 317, 961 and 977. Redshifts of these clusters are in the region 0.002 – 0.032.



(a) (b)

Fig. 8. Several peculiarities on the distributions of galaxies inside cluster #88 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre. Shaded regions correspond areas from right side of bifurcation points on Fig. 2.

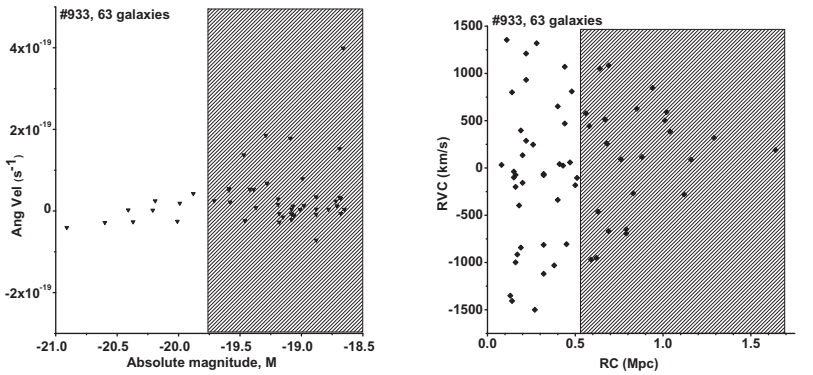


(a) (b)

Fig. 9. Several peculiarities on the distributions of galaxies inside cluster #1046 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre. Shaded regions correspond areas from right side of bifurcation points on Fig. 3.

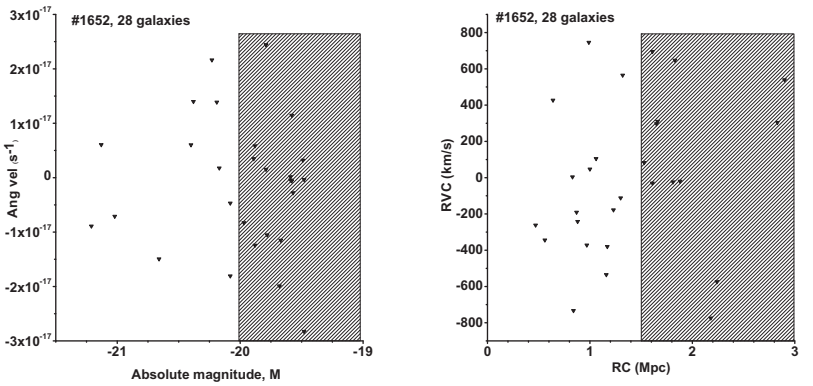
Galaxies from these areas are paired accordingly its spectral characteristics and position. Unfortunately direct data analysis not allow understanding of reality of such dual structure because it could be caused by several selection effects.

Let's try to investigate dynamic of these systems using Nonlinear Time Series Analysis methods – see, for instance, Ref. 9. Unfortunately real time variables are absent in the analysed systems. But dynamic behavior of cluster's members strongly depend of its velocities and distances to cluster centre.[10] We have construct phase space cluster of galaxies #88 using values of redshift, coordinates, magnitude,



(a) (b)

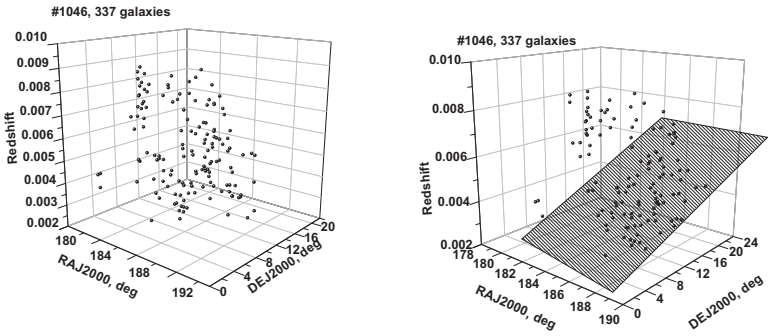
Fig. 10. Several peculiarities on the distributions of galaxies inside cluster #933 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre. Shaded regions correspond areas from right side of bifurcation points on Fig. 4.



(a) (b)

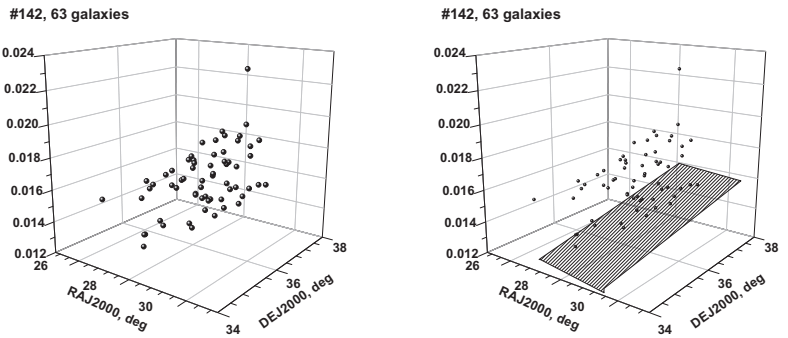
Fig. 11. Several peculiarities on the distributions of galaxies inside cluster #1652 on CfA2 data. (a) Graph for distribution of galaxies on angular velocity and absolute magnitude. (b) Graph for distribution of galaxies on velocity and distance to cluster centre. Shaded regions correspond areas from right side of bifurcation points on Fig. 6.

absolute magnitude and distance to centre. We have use ratio between distance to centre and tangential velocity as time similar variable in the presented article and constructed simple attractors in this phase space for first attempt of investigation of dynamic of system. Let's separate basins of attractors correspondingly to bifurcation points on the analysable distributions. Accordingly to preliminary results of analysis we have obtained two attractors in the phase space of cluster #88 which concludes real dual structure of this system. Figure 16 represents three dimension projection of obtained attractors.



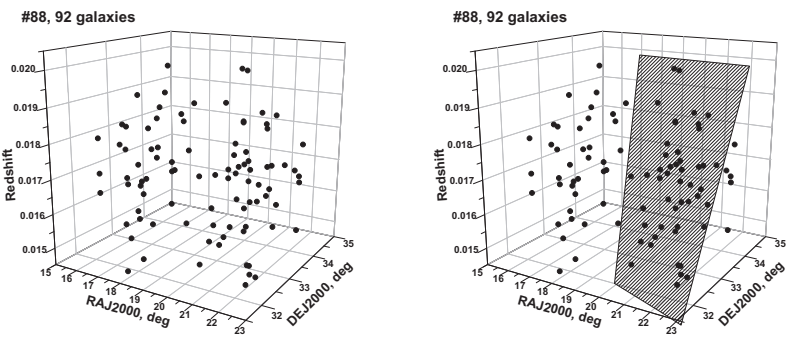
(a) (b)

Fig. 12. The spatial distributions of galaxies for cluster #1046 on CfA2 data (a) and dual structure appearance on this distribution (b).



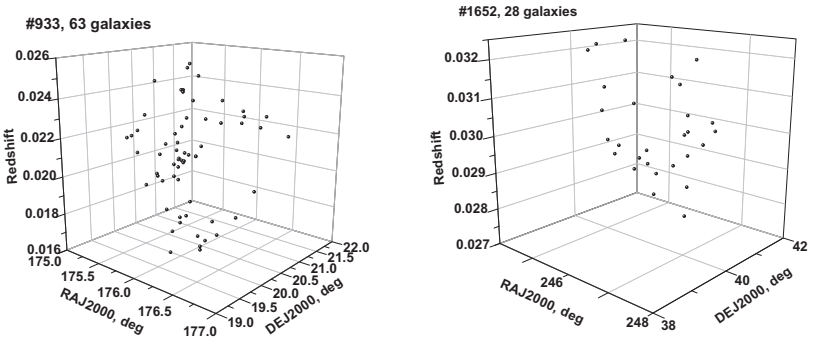
(a) (b)

Fig. 13. The spatial distributions of galaxies for cluster #142 on CfA2 data (a) and dual structure appearance on this distribution (b).



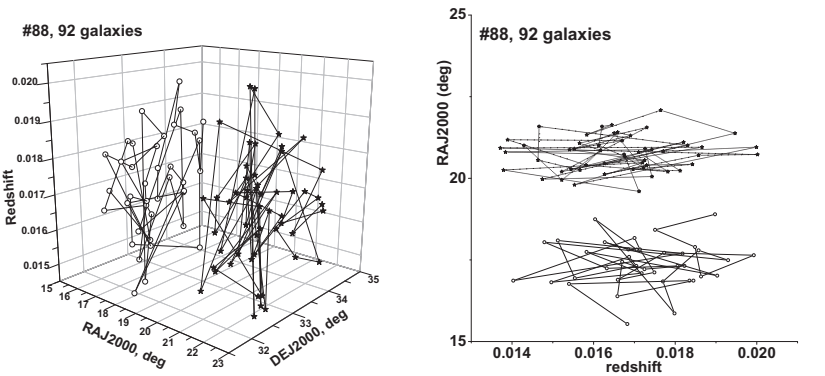
(a) (b)

Fig. 14. The spatial distributions of galaxies for cluster #142 on CfA2 data (a) and dual structure appearance on this distribution (b).



(a) (b)

Fig. 15. The spatial distributions of galaxies for several clusters on CfA2 data. (a) Graph for cluster #933. (b) Graph for cluster #1652.



(a) (b)

Fig. 16. The projections of six-dimensional system of attractors for cluster #88. (a) Projection in three-dimension space with axes corresponds to coordinates and redshift. (b) Two dimension projection with axes corresponds to RA and redshift.

3. Conclusions

Preliminary results of the properties of 13 clusters of galaxies from CfA2 redshift survey (##88, 1101, 1046, 142, 933, 1242, 1652, 107, 150, 316, 317, 961, 977) are presented. Redshifts of these clusters are in the region 0.002 – 0.032. The distributions on magnitude, absolute magnitude and angular velocity, etc represent two areas for these clusters. Galaxies from these areas are paired accordingly its spectral

characteristics and position. Also several anomalies of spatial dynamic of galaxies in these clusters were separated.

The presence of such structure allows conclude two alternatives. In the first one dark matter presence inside cluster in configuration similar to Zeldovich pancake. Second case is gravitational lensing on compact object or dark matter blob located between galaxy cluster and observer.

Also several peculiarities have found on the spatial distributions of galaxies in clusters##933, 142, 1046, and 1652. Moreover, groups ##933, 142, 1046, and 1652 reveals associations with high-energy gamma-emission sources on Fermi/LAT 10–Year Point Source Catalog 4FGL DR2 data (4FGLJ1144.9+1937, 4FGLJ0152.2+3714, 4FGLJ1230.8+1223 and 4FGLJ1653.8+3945 correspondingly).

Furthermore, 4FGLJ1144.9+1937 and 4FGLJ1230.8+1223 observed in subTeV energy band by VERITAS data.

Joint observations of such clusters by orbital gamma-ray observatories with high angular resolution and ground-based Cherenkov air-shower experiments could possibly clarify the type of influence to groups characteristics (gravitational lensing or object inside cluster) and processes of particle acceleration in these objects especially highest energy of emitted gammas. Thus we propose including these and similar clusters in the programs of observations of the planned experiment GAMMA-400 (Gamma Astronomical Multifunctional Modular Apparatus) with angular resolution $\sim 0.01^\circ$ at $E_\gamma = 100\,GeV$ and upper energy band boundary about several TeV. Also now the coordination of multiwavelength observations program of Cherenkov Telescope Array (CTA) and GAMMA-400 is discussed.

Acknowledgments

Authors thank for the support from National Research Nuclear University MEPhI in the framework of the Russian Academic Excellence Project (contract No. 02.a03.21.0005, 27.08.2013).

References

1. F. M. Rieger and A. Levinson, Radio Galaxies at VHE Energies, *Galaxies 6*, **116** (2018).
2. Chervin F. P. Laporte and Simon D. M. White, The re-distribution of matter in the cores of galaxy clusters, *MNRAS 451* **1177–1189** (2015).
3. Y.-Y. Choi, C. Park, J. Kim, J. R. Gott, D. H. Weinberg, M. S. Vogeley, and S. S. Kim (for the SDSS Collaboration), Galaxy clustering topology in the Sloan Digital Sky Survey main galaxy sample: A test for galaxy formation models. Cosmology and Nongalactic Astrophysics, astro-ph.CO.arXiv:1005.0256 (2018).

4. G. O. Abell, Ap. J. Suppl, The Distribution of Rich Clusters of Galaxies, *Astrophysical Journal Supplement, vol. 3* **211** (1958).
5. Z. L.Wen, J. L. Han, and F. Yang, A catalogue of clusters of galaxies identified from all sky surveys of 2MASS, WISE, and SuperCOSMOS, *MNRAS 475* **343** (2018).
6. A. P. Mahtessian and V. G. Movsessian, List of groups of galaxies based on the CfA2 redshift survey, *Astrophys, 53* **70** (2010).
7. John Huchra, (Harvard-Smithsonian Center for Astrophysics) `https://www.cfa.harvard.edu/~dfabricant/huchra/` .
8. A. P. Mahtessian, V. H. Movsessian, and V. M. Mahtessian, Relationship of Markarian galaxies with their surroundings, *Astrophys, 40* **285** (1997).
9. H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, *Cambridge University Press, 1st edn.*, (Cambridge, UK, 1997).
10. T. Buzug and G.Pfister, Optimal delay time and embedding dimension for delay-time coordinates by analysis of the glocal static and local dynamic behavior of strange attractors, *Phys. Rev. A 45* **7073–7084** (1992).
11. F. Takens, Detecting strange attractors in fluid turbulence, (In Rand, D. & Young, L. S. eds.) *Dynamical Systems and Turbulence, vol. 898 of Lecture Notes in Mathematics*, **366–381** (Springer-Verlag, Berlin, 1981).
12. J. Ballet and Fermi-LAT Collab, Fermi Large Area Telescope Fourth Source Catalog, *ApJS 247, 33* (2020).
13. R. Mukherjee (for the VERITAS Collab), Observing the Energetic Universe at Very High Energies with the VERITAS Gamma Ray Observatory, *Advances in Space Research* (2018).
14. L. N. Khanh, I.V. Arkhangelskaja, A. M. Galper and D. N. Dorosheva, Rich Galaxy Clusters from CfA2 Redshift Survey: Spatial Dynamic and High-Energy Gamma-Emission, *Physics of Atomic Nuclei Vol. 83, No. 2* **300–306** (2020).
15. I. V. Arkhangelskaja, L. N. Khanh, A. M. Galper and D. N. Dorosheva, The results of analysis of Rich Galaxy Clusters from CfA2 Redshift Survey spatial distribution/I. V. Arkhangelskaja, L. N. Khanh, A. M. Galper and D. N. Dorosheva *J. Phys.: Conf. Ser. vol. 1690* , ID:012025 (2020).
16. T. E. Jeltema, J. Kehayias, and S. Profumo, Gamma Rays from Clusters and Groups of Galaxies: Cosmic Rays versus Dark Matter, *Phys. Rev. D 80, 023005* (2009).
17. A. P. Mahtessian V. H. Movsessian, and V. M. Mahtessian, Relationship of Markarian galaxies with their surroundings, *Astrophys vol. 40* **285** (1997).
18. A. A. Leonov, A. M. Galper, N. P. Topchiev, A. V. Bakaldin, O. D. Dalkarov, E. A. Dzhivelikyan, A. E. Egorov, M. D. Kheymits, V. V. Mikhailov, P. Picozza, R. Sparvoli, S. I. Suchkov, Yu. T. Yurkin, and V. G. Zverev, Multiple Coulomb scattering method to reconstruct low-energy gamma–ray direction in the GAMMA-400 space-based gamma–ray telescope, *Adv. Space Res 63*, **3420-3427** (2019).
19. N. P. Topchiev, A. M. Galper, I. V. Arkhangelskaja, A. I. Arkhangelskiy, A. V. Bakaldin, I. V. Chernysheva, O. D. Dalkarov, A. E. Egorov, Yu. V. Gusakov, M. D. Kheymits, A. A. Leonov, P. Yu. Naumov, N. Yu. Pappe, M. F. Runtso, Yu. I. Stozhkov, S. I. Suchkov, Yu. T. Yurkin and V. G. Zverev, Space-based GAMMA-400 mission for direct gamma- and cosmic-ray observations, *IOP Conf. Series: Journal of Physics: Conf. Series 1181* (2019).

20. I. V. Arkhangelskaja, A. I. Arkhangelskiy, E. N. Chasovikov, A. M. Galper, M. D. Kheymits, A. E. Murchenko and Y. T. Yurkin, The Counting and Triggers Signals Formation System for Gamma-telescope GAMMA-400, *Physics Proce: Conf. Ser. 675, 032015* **212-219** (2016).
21. R. A. Ong and CTA Consortium, *in Proceedings of the 7th Roma International Conference on Astroparticle Physics* (Rome, Italy, 2019), (Ed. by M. De Vincenzi, A. Capone and A. Morselli, EPJ Web Conf. 209, 01038) (2019).

Probing the nature of dark matter with Milky Way subhaloes

M. R. Lovell
Science Institute, University of Iceland,
107 Reykjavík, Iceland
E-mail: lovell@hi.is

The nature of dark matter is one of the most pressing questions in modern physics. Efforts to answer this question have centred on observations that determine its impact on the properties of galaxies plus attempts to detect rare events in which the dark matter indirectly generates photons. In this talk I presented work on using the properties of Milky Way satellite galaxies to constrain models of self-interacting dark matter (SIDM) and warm dark matter (WDM). First I used N-body simulations of an SIDM model with a very large self-interaction at low relative velocities ($\sigma > 100$ cm^{2}gr^{-1}) to demonstrate that subhaloes in such models undergo gravothermal collapse and increase the diversity of satellite matter distributions in line with inferences from observations. Second, I considered the radial distribution of subhaloes in WDM simulations from the point-of-view of stellar stream gap constraints. I showed that the subhalo population within 50 kpc of the host halo centre is comprised of massive satellites that have sunk under dynamical friction. Given that the abundance and structure of massive satellites in WDM is very similar to the fiducial cold dark matter (CDM) model, the differences between the subhalo populations in the two models in the inner halo are much smaller than for the halo as a whole, and therefore the stellar stream gaps statistics, which are only available out to 30 kpc, are not as powerful at discriminating between the models as previously anticipated. I ended with a brief presentation of estimates for the production rate of X-ray photons in dark matter decay, and predicted that the velocity dispersion of X-ray decay lines to be probed in nearby galaxy clusters will have values in the range [500,800] kms^{-1}.

Keywords: Dark matter

1. Introduction

Observations of a wide variety of astrophysical objects have returned evidence for the existence of an invisible substance that constitutes $\sim 82\%$ of the mass of the Universe. These observations include the anisotropies of the cosmic microwave background radiation (CMB)[1] and the rotation curves of galaxies (e.g. Ref 2). All of these observations are consistent with the presence of matter that is dynamically cold, at least on the scales of massive galaxies and larger, and does not interact with light; this material is therefore dubbed cold dark matter (CDM).

The CMB and rotation curves studies probe the properties of dark matter on scales typically larger than 10 kpc. However, on scales of dwarf galaxies, the close match between observations on the one hand and predictions for gwalaxy abundance and structure from CDM N-body simulations start to break down.[3–5] In part, such discrepancies can be explained by astrophysical processes that change the properties

of dark matter haloes, through reionisation[6,7] and supernova feedback.[8,9] However, it is not clear that astrophysics alone provides the best match to observations.[10–12] In the context of the non-detection of CDM-type dark matter candidates with either direct detection[13,14] or γ-ray indirect detection experiments,[15,16] it is important to consider dark matter candidates that behave somewhat differently, both in their astrophysical impact on structure formation and in their detection abilities.

In this talk I covered three topics: two in the properties of Milky Way (MW) satellites and a third on the potential for detection of dark matter decay in X-rays. The two MW subhaloes topics concern very different dark matter models: first, an extreme velocity dependent self-interacting dark matter (SIDM) model in the context of MW satellite densities, and second, the effect of the warm dark matter (WDM) model consequences on the spatial distribution of both satellites and non-luminous subhaloes. The dark matter decay results are applicable to any model in which dark matter undergoes two-body decay that does not affect the galaxy kinematics significantly.

These proceedings are organised into three sections. I describe the SIDM work in Section. 2, the WDM work in Section. 3, and the dark matter decay results in Section. 4. The papers from which these three sections were derived are Ref. 17, Ref. 18 and Ref. 19 respectively.

2. Gravothermal collapse in SIDM and the densities of Milky Way satellites

The masses of MW satellites have been a popular probe of dark matter and astrophysics models for at least 10 years. Refs. 5, 20 showed that, at least at a first glance, CDM simulations predicted the existence of satellites more massive and dense than any of the observed MW satellites at the same time that other studies reported evidence for cored dark matter density profiles.[21] A successful dark matter model will have to explain both of these observations simultaneously. Ref. 22 argued that cored SIDM dark matter profiles were insufficiently dense to explain the very high reported masses of the MW's ultrafaint dwarf galaxies, and therefore in order for these models to be viable a fraction of SIDM subhaloes below a given mass scale would need to undergo the process of gravothermal collapse, by which material starts rapidly flowing towards the halo centre and creates a density profile much steeper than in CDM.[23]

For Ref. 22 we had run an extreme model of SIDM that would cause gravothermal collapse to occur in MW satellites within a Hubble time. It featured a strong cross section-velocity dependence: a cross section $\sigma \sim 100$ cm^{2}g^{-1} at interactions velocities < 30 kms^{-1}. This cross-section was chosen in order to induce gravothermal collapse without relying on stripping of the satellite outskirts[24] and to avoid violating published bounds on high velocity interaction cross-sections from cluster mergers;[25] we do not attempt to match the reported measurements of ~ 1 cm^{2}g^{-1} cross-sections at $\sim$1500-2000 kms^{-1} by Refs. 26, 27. Our goal in Ref. 17 was to

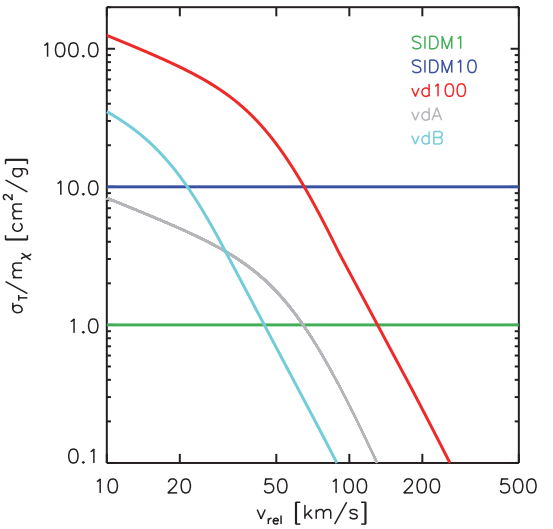


Fig. 1. Self-interaction cross section as a function of particle relative velocity for five SIDM models. The primary model of interest in our study was vd100, which is shown in red. We also show two further velocity-dependent SIDM models, labelled vdA (grey) and vdB (cyan), plus two velocity independent models with cross-sections ~ 1 cm^{2}g^{-1} (SIDM1, green) and ~ 10 cm^{2}g^{-1} (SIDM10, blue). Reproduced from Ref. 17.

investigate the abundance and density of subhaloes in this model, which we dubbed 'vd100'. We present the velocity dependence of the cross section for this model and other SIDM models in Fig. 1.

We explored the properties of the vd100 model using the same simulation that we ran for Ref. 22. This was zoomed resimulation of the Aquarius Aq-A3 MW halo-analogue[28] using the same initial conditions as for the original CDM simulation but with the inclusion of vd100 model self-interactions. The simulation was run with the Arepo code.[29] The host halo virial mass $M_{200} = 1.9 \times 10^{12}$ M$_\odot$, the simulation particle mass was 4.9×10^4 M$_\odot$, and the gravitational softening length $\epsilon = 120.5$ pc. The cosmological parameters were consistent with the WMAP first year results.[30] The same code, simulation parameters and cosmological parameters were used in the other four SIDM simulations.

We looked for evidence of gravothermal collapse through the following methodology. We computed circular velocity curves for subhaloes within each of these simulations that were located within 300 kpc of the centre of the host halo, and parametrised the structure of velocity curves with two numbers: the maximum value of the circular velocity, $V_{\rm max}$, and the radius $r_{\rm max}$ at which $V_{\rm max}$ is attained. $V_{\rm max}$ is often interpreted as a proxy for the subhalo mass, and $r_{\rm max}$ as a measure of concentration at fixed $V_{\rm max}$, with larger $r_{\rm max}$ corresponding to a less concentrated halo. We present contour plots for the $V_{\rm max} - r_{\rm max}$ distributions of haloes in the CDM, SIDM10, vdB, and vd100 simulation subhaloes in Fig. 2.

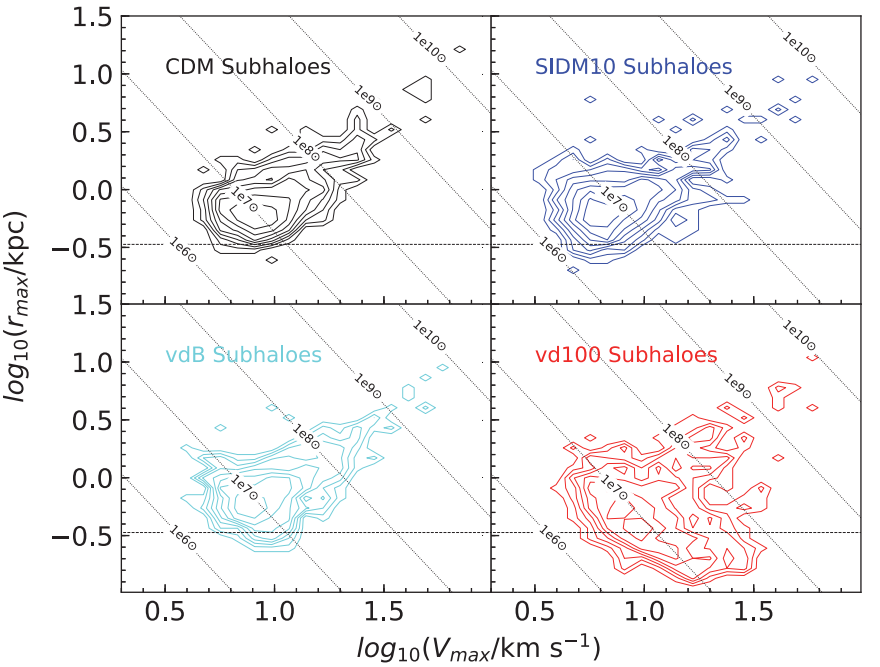


Fig. 2. Contour plots of $V_{\rm max}$ - $r_{\rm max}$ values for subhaloes in four models: CDM (top left panel), SIDM10 (top right), vdB (bottom left) and vd100 (bottom right). Diagonal lines denote parameter space along which the mass enclosed with $r_{\rm max}$ is constant. The horizontal lines denote the radius corresponding to 2.8× the softening length of the simulation and therefore marks the hard resolution limit. Reproduced from Ref. 17.

The vdB subhaloes follow the same relation in $V_{\rm max}$ - $r_{\rm max}$ as their CDM equivalents despite their very high cross-sections at low velocities, > 20 cm^{2}gr^{-1}. By contrast, a large population of vd100 subhaloes have $r_{\rm max}$ well below the CDM relation: the subhaloes are therefore denser than CDM and have undergone gravothermal collapse. We explore this trend further for a sample of subhalo circular velocity curves that have $V_{\rm max}$ values in the range $V_{\rm max} = [25, 35]$ kms^{-1} from the CDM and vd100 runs, which we present in Fig. 3.

The number of vd100 subhaloes with $r_{\rm max} > 1$ kpc is approximately the same as the number of CDM subhaloes in the same $V_{\rm max}$ bin, since the implementation of velocity-dependent SIDM has only a minor impact on the halo mass function[a]. From radii < 2 kpc these vd100 subhaloes' circular velocity profiles decline much faster than their CDM counterparts as a consequence of their cored profiles. However, there also exists a population of approximately twice as many vd100 subhaloes with $r_{\rm max} < 1$ kpc. The accurate values of $r_{\rm max}$ and $V_{\rm max}$ for these subhaloes

[a]In models for which the cross-section is large at high velocities, as is common for velocity independent models, interactions between subhaloes and the host can evaporate subhaloes and suppress the mass function.[11,17]

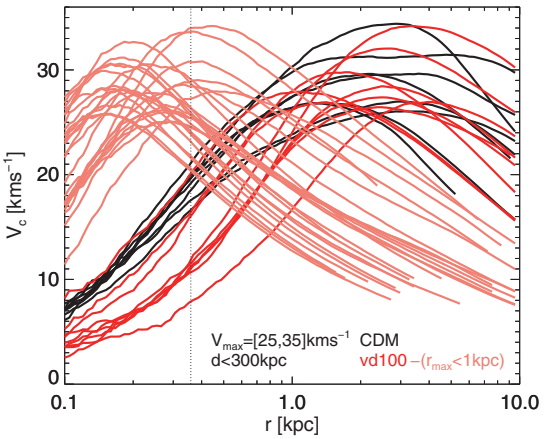


Fig. 3. Circular velocity curves for subhaloes in the CDM and vd100 simulations that have $V_{\rm max}$ =[25, 35] kms^{-1}. CDM subhaloes are shown in black. vd100 subhaloes that have $r_{\rm max} >$ 1 kpc are shown in red, and vd100 subhaloes with $r_{\rm max} < 1$ kpc are shown in pink. Reproduced from Ref. 17.

are unknown because the circular velocity curves are still rising at the softening resolution limit; we therefore just state that these haloes have indeed undergone collapse and are concentrated enough to explain the ultrafaint dwarf masses at the same time that the uncollapsed cored haloes can fit the brighter satellites.[22] We also show in Ref. 17 that the density profiles of these objects is a power law of slope -3.

3. The WDM subhalo distribution and implications for gaps in streams studies

For the second part of the talk I switched to MW satellites in the WDM cosmology. The motivation for this model arises from clues that it might, under some circumstances, be a better match to observations than CDM[12] and also that its primary particle physics candidate, the sterile neutrino[31,32] may have been detected in X-ray decay.[33,34] This model suppresses the number of dwarf mass haloes[35] and lowers their concentrations[10] relative to CDM. The WDM model can therefore be constrained by observations that measure the number and concentration of dwarf haloes, especially dark matter haloes so small that they never form stars.

One such method is to attempt to detect gaps in stellar streams. Globular clusters and satellite galaxies experience stripping of their outer layers as they orbit within the gravitational field of their host galaxies, leaving behind a smooth stream of stars. The structure of the stream will be disturbed by interactions with other structures, such as dark matter subhaloes. The effect of dark matter physics on the abundance and density of subhaloes can therefore be inferred from the presence of the gaps in streams.[36–38]

Several studies have claimed strong limits on WDM in particular from the apparent detection of such gaps in streams,[37,38] although the nature of dark matter substructures as a cause is disputed. For the paper from which this talk section was derived, I examined the subhaloes that populate the region where streams can be probed for gaps given current observational constraints, i.e. $\lesssim$30 kpc from the MW centre, in CDM and in WDM and analysed the difference between these inner-host subhaloes on the one hand and subhaloes at all radii on the other.

For this project, we estimated the subhalo population properties using two N-body simulations, one of CDM and one of a WDM . Our two companion simulations were CDM and WDM variations of the COCO volume;[39,40] a zoomed simulation of a spherical region $\sim$28 Mpc in radius with a dark matter simulation particle mass of $\sim 2 \times 10^5$ $M_\odot$. The WDM model was a 3.3 keV thermal relic; this was chosen to be consistent with a set of Lyman-α forest constraints[41] and is also a reasonable approximation to the matter power spectrum of the sterile neutrino that could explain the reported 3.55 keV line.[12] We identified 26 MW-halo analogues: these we selected to be in the mass range $[1.0, 1.5] \times 10^{12}$ $M_\odot$, and also to be isolated in that the distance to any larger halo was > 1 Mpc[b].

We first considered the abundance of subhaloes in two mass bins: $[10^7, 10^8]$ $M_\odot$ and $[10^8, 10^9]$ $M_\odot$. We computed the number of subhaloes in these mass bins for CDM and WDM in two radius bins: < 40 kpc from the halo centre to represent stream gap-generating subhaloes and < 300 kpc for the subhalo population at large. Any changes in these quantities will be driven by two factors: the larger number of haloes generated in CDM plus the additional stripping of subhaloes in WDM due to their lower concentrations. In order to ascertain the contribution of these two factors we therefore performed a 'hybrid' calculation, in which for each CDM subhalo we selected a WDM subhalo with the same infall mass and infall time and assigned this WDM subhalo final time mass to the CDM subhalo. This was equivalent to the mass that the CDM subhalo may have had if it had experienced a WDM stripping rate. Our hybrid calculation therefore factors out the contribution to CDM–WDM differences from the change in infall mass function and retains the differences from stripping.

With these results we compute the ratio of WDM-to-CDM subhalo abundances in the two mass bins plus the ratios of the hybrid measurements and present the results in the left-hand set of panels in Fig. 4. Beginning with the $[10^7, 10^8]$ $M_\odot$ bin, we find strong differences in the relative properties of the models between radius bins. Within the full 300 kpc distance from the host halo centre the number of WDM subhaloes is suppressed to 10% of the CDM value on average, whereas for the 40 kpc bin it is instead 30% in the median, with a much larger scatter. We argue in the paper that this result is due to the inner halo being dominated by the remnants of

[b]The MW is located only 750 kpc from the more massive Andromeda galaxy; we expect that this will not affect the results.

massive haloes that have sunk under dynamical friction; the abundance of massive haloes is much more similar between the two models than the mass function down to 10^7 $M_\odot$, therefore the value of this ratio is much closer to 1. A similar result is found for the more massive bin, where the suppression is a median 40% of CDM for the < 300 kpc bin but 60% of CDM.

One possible complication to this picture is the issue of extra stripping in WDM subhaloes as discussed above, and we can tease out the contributions of changes in stripping and changes in the mass function by comparing the hybrid calculation results to the full WDM results: if the hybrid results are the same as the full WDM then the change between WDM and CDM is entirely due to extra stripping, whereas if the hybrid results instead straddle the ratio = 1 line then the change was entirely due to the input mass function. In the lower mass bin the < 300 kpc hybrid ratio is much closer to a ratio of 1 than to 0.1, therefore the change in mass function drives the difference, whereas in the < 40 kpc radius the hybrid suppression ratio is lower than at < 300 kpc as per the switch from unstripped low-infall mass subhaloes to high-mass stripped ones.

The difference in subhalo infall mass at different radii will also have an impact on the concentration–mass relations at these radii, with the subhaloes in the central region more likely to be the dense cores of massive progenitors. The concentration of these objects will also have an effect on the gaps in streams measurements. Obtaining an accurate concentration–mass relation for subhaloes close to the host halo centre with N-body simulations is challenging for a number of reasons, including the possibility that mass from the subhalo is inaccurately ascribed to the host halo and that the mass distribution of the host will be strongly impacted by the presence of the stellar disc, both from the disc itself and from the contraction of the halo. We therefore developed an estimate for the mass and size of subhaloes in the centre using the infall mass / structure plus a measurement of the observed MW mass profile. We used the measured MW mass profile of Ref 42 to calculate the tidal radius of a subhalo at a distance of 20 kpc from the halo centre based on its infall mass and infall density profile, then assumed that the halo layers outside the tidal radius were removed while the mass inside the tidal radius was not affected. We thus calculated a new mass based on the mass interior to the tidal radius, and a new size as the half-mass radius of the remaining (unstripped) material. We computed median mass–radius relations for subhaloes within 50 kpc of the host centre using these estimates of stripped mass and size, and also median relations across the subhalo using the infall mass and infall size. We present these results in the right hand panel of Fig. 4.

Beginning with the whole subhalo population, the CDM and WDM sizes are identical at $6 \times 10^9\,M_\odot$ but quickly diverge towards lower masses, with WDM subhaloes of mass $6 \times 10^7\,M_\odot$ on average 50% larger than CDM subhaloes of the same mass. The subhaloes in the inner halo instead show much smaller sizes, such that

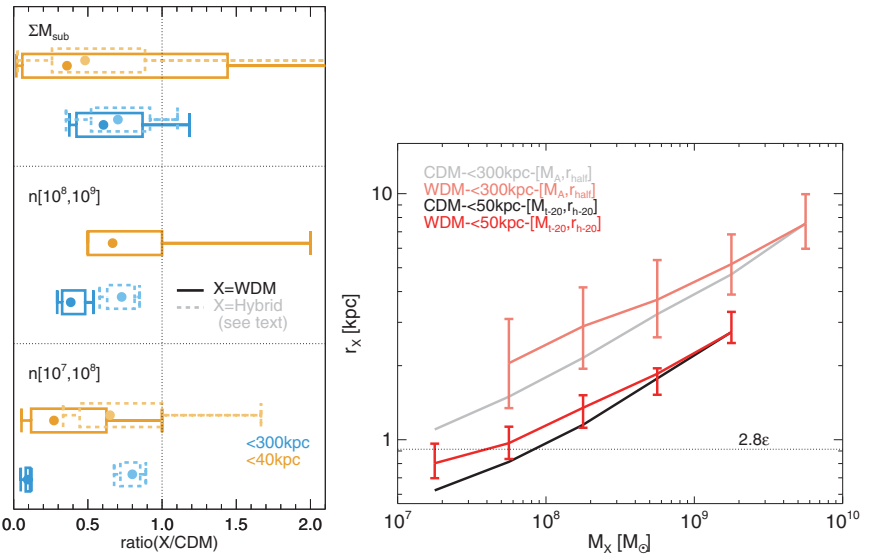


Fig. 4. *Left-hand panels:* ratio of quantities between WDM and CDM versions of the 24 volumes. In the top panel we show the ratio of the mass in WDM and CDM; in the middle panel the fraction of the number of subhaloes in the mass bin $[10^8, 10^9]$ $M_\odot$, and in the bottom panel the ratio of $[10^7, 10^8]$ $M_\odot$ subhalo number. Ratios for subhaloes within 40 kpc of the host centre are shown in orange and within 300 kpc in blue. Filled dots denote median ratios, boxes 68% of the data and error bars 95% of the data. Solid lines are used for the measured WDM subhalo fractions and dashed lines for the hybrid calculation as described in the text; the hybrid calculation is omitted for the mass= $[10^8, 10^9]$ $M_\odot$, distance< 30 kpc bin due to poor statistics. *Right-hand panel:* the mass–radius relation for subhaloes in two radius bins. Data for subhaloes within 300 kpc are shown using infall masses and sizes in grey (CDM) and pink (WDM); data for subhaloes in the < 50 kpc bin are calculated using estimates for the stripping radius at 20 kpc as discussed in the text, and are shown in black (CDM) and red (WDM). Error bars on the WDM relations denote 58% of the data; error bars are omitted for CDM on clarity grounds, and are of the same size as for WDM. Reproduced from Ref. 18.

the WDM subhaloes within 50 kpc are more concentrated the CDM subhalo population. The difference in ratio between the stripped estimates for these inner-host subhaloes is approximately 20% at $10^8\,M_\odot$, and the majority of such WDM subhaloes are less than 2 kpc in size. We therefore argue that the subhaloes likely to generate stream gaps would have to have stream impact parameters less than 2 kpc in order for the difference in concentrations to be imprinted on the gap properties.

In conclusion, I showed in this part of the talk that the properties of the subhaloes that generate gaps are very different to the properties of subhaloes generally, in that they are the stripped, concentrated remnants of massive haloes. Therefore the ability of stream-gap analyses to discern between CDM and WDM is weaker than one would infer from the subhalo population at large.

4. Predictions for the X-ray emission signal from decaying dark matter in galaxy clusters

For the final part of the talk I shifted to indirect detection opportunities for dark matter decay, and presented a prediction for the decay to be measured in nearby galaxy clusters by the XRISM X-ray experiment.

Decay is possible for dark matter particles that have a lifetime much longer than the age of the Universe but are never the less not completely stable. Such particles are typically light and therefore decay in the X-ray band. The prime dark matter particle candidate that has a mass in this range is the resonantly produced sterile neutrino, which in addition to dark matter explains baryogenesis and the oscillations of standard model neutrinos;[31,32,43] it also behaves as WDM,[44,45] and is thus accessible to the constraints from the previous section. The decay of this

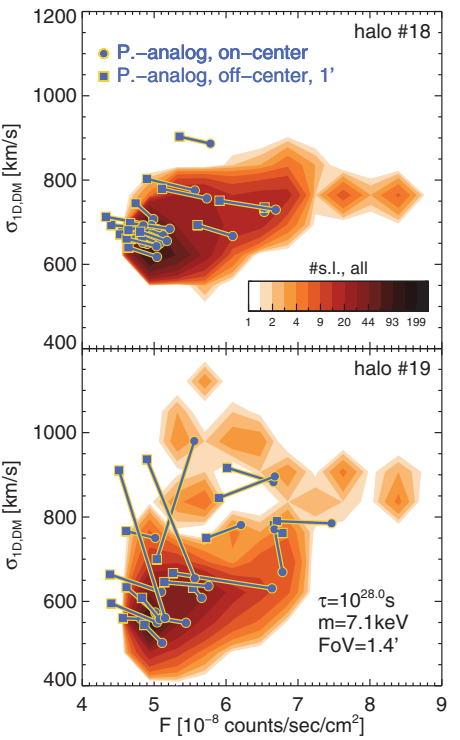


Fig. 5. Predictions for the properties of a putative dark matter X-ray decay line in the Perseus cluster to be measured by the XRISM experiment. We use simulations of two Perseus-mass clusters, and each cluster is represented in one panel and is subject to 500 sightlines. We show contours of the one dimensional velocity dispersion versus the total line flux: the velocity dispersion is independent of the particle physics properties, and the flux assumes $\tau = 10^{28}$ s and $m = 7.1$ keV. The blue circles denote on-centre sightlines for which the velocity dispersion of the cluster member galaxies is the measured velocity dispersion of Perseus cluster members, and the attached blue squares offset 1 arcminute from the centre. Reproduced from Ref. 19.

particle takes the form of a monochromatic X-ray line at an energy half the sterile neutrino rest mass energy. The width of the line is determined by the macroscopic dark matter one dimensional velocity dispersion, and is therefore independent of the particle physics parameters.

The XRISM experiment, to be launched in 2023, will have sufficient spectral resolution to measure the width of the claimed dark matter decay line in nearby clusters of galaxies. I therefore used the C-EAGLE simulations[46,47] of galaxy clusters to perform mock observations with the XRISM field-of-view of dark matter decay lines in Perseus-mass clusters.[19] The flux calculation assumed a particle lifetime $\tau = 10^{28}$ s and mass $m = 7.1$ keV, as inferred from the claimed detections of the 3.55 keV line.[33,34] We identified two of the 30 C-EAGLE clusters as having same mass as Perseus within some tolerance, and computed 500 sightlines through the halo centre from randomized viewing angles. The results are shown in Fig. 5.

The flux of the line can vary by a factor of 50%, with most of the sightlines bunching up towards lower fluxes in the allowed range. The velocity dispersion is concentrated in the [500,800] kms^{-1} range, and exceptionally at [400,1000] kms^{-1}. This velocity dispersion is significantly narrower than the measured velocity dispersion of the member stars, which is typically 1300 kms^{-1}, and is still much broader than astrophysical emission lines (< 200 kms^{-1}). Given that the velocity dispersion is also independent of the particle physics parameters for the sterile neutrino model, the XRISM observatory offers arguably the best opportunity to detect dark matter within the next five years.

5. Conclusions

In this talk I summarised three leads for constraining and identifying the dark matter: the density of the Milky Way (MW) ultrafaint satellites, the properties of subhaloes that carve gaps in stellar streams, and the parameters of X-ray decay lines in galaxy clusters.

First, I investigated how the distribution of ultrafaint satellite densities could be explained by a velocity-dependent self-interacting dark matter model (SIDM) in which the self-interaction cross-section is of order 100 $\mathrm{cm}^2\mathrm{gr}^{-1}$ for relative velocities below 20 kms^{-1}. I showed that this model induces gravothermal collapse in a fraction of the dark matter subhaloes. The density profiles of these collapsed subhaloes exhibit a power law of index -3. The combination of collapsed and cored SIDM subhaloes can potentially explain the diversity of MW satellites masses, in which bright satellites could be hosted in massive, cored satellites whereas ultrafaint dwarfs are hosted in less massive, gravothermally collapsed subhaloes.[17]

For the second part of the talk I switched to comparing the subhaloes that carve gaps in stellar streams between cold dark matter (CDM) and warm dark matter (WDM), for which I extracted subhaloes from the COCO simulations that are located within the same region of configuration space that stream gaps can be detected (< 30 kpc from the host centre). I showed that these central-region

subhaloes are the remnants of massive progenitors that have sunk to the halo centre under dynamical friction. I therefore argued that, since the differences between CDM and WDM massive subhaloes are small, the discriminating power of stream gaps as a dark matter probe is weaker than one would infer from comparing all subhaloes associated with each host.[18]

I ended the talk with a brief presentation of work on dark matter indirect detection. I showed calculations for the expected width of putative lines from dark matter decay in a series of nearby galaxy clusters. Given the uncertainty between viewing angles, I showed that the line velocity dispersion would have a value in the range $[500, 800]$ kms^{-1}, which is narrower than the velocity dispersion of the member galaxies but crucially still significantly broader than X-ray emission lines from hot gas.[19] The upcoming XRISM observatory will therefore have an excellent opportunity to determine whether the line detected at 3.55 keV in galaxies and clusters[33,34] is from an astrophysical source or is indeed from dark matter decay.

Acknowledgments

MRL acknowledges support by a Grant of Excellence from the Icelandic Research Fund (grant number 206930). This work used the DiRACDurham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1 and ST/R002371/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

References

1. Planck Collaboration, R. Adam, N. Aghanim, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, E. Calabrese, J.-F. Cardoso, J. Carron, H. C. Chiang, L. P. L. Colombo, C. Combet, B. Comis, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, E. Di Valentino, C. Dickinson, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, Y. Fantaye, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, A. Gruppuso, J. E. Gudmundsson, F. K. Hansen, G. Helou, S. Henrot-Versillé, D. Herranz, E. Hivon, Z. Huang, S. Ilić, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J.-M. Lamarre, M. Langer, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. López-Caniego, Y.-Z. Ma, J. F. Macías-Pérez, G. Maggio, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M.-A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss,

P. Naselsky, P. Natoli, C. A. Oxborrow, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, L. Patrizii, O. Perdereau, L. Perotto, V. Pettorino, F. Piacentini, S. Plaszczynski, L. Polastri, G. Polenta, J.-L. Puget, J. P. Rachen, B. Racine, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, M. Rossetti, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, G. Sirri, R. Sunyaev, A.-S. Suur-Uski, J. A. Tauber, M. Tenti, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, A. Zacchei and A. Zonca, Planck intermediate results. XLVII. Planck constraints on reionization history, *A&A* **596**, p. A108 (December 2016).

2. X. X. Xue, H. W. Rix, G. Zhao, P. Re Fiorentin, T. Naab, M. Steinmetz, F. C. van den Bosch, T. C. Beers, Y. S. Lee, E. F. Bell, C. Rockosi, B. Yanny, H. Newberg, R. Wilhelm, X. Kang, M. C. Smith and D. P. Schneider, The Milky Way's Circular Velocity Curve to 60 kpc and an Estimate of the Dark Matter Halo Mass from the Kinematics of ˜2400 SDSS Blue Horizontal-Branch Stars, *ApJ* **684**, 1143 (September 2008).
3. A. Klypin, A. V. Kravtsov, O. Valenzuela and F. Prada, Where Are the Missing Galactic Satellites?, *ApJ* **522**, 82 (September 1999).
4. B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel and P. Tozzi, Dark Matter Substructure within Galactic Halos, *ApJ* **524**, L19 (October 1999).
5. M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, Too big to fail? The puzzling darkness of massive Milky Way subhaloes, *MNRAS* **415**, L40 (July 2011).
6. J. S. Bullock, A. V. Kravtsov and D. H. Weinberg, Reionization and the Abundance of Galactic Satellites, *ApJ* **539**, 517 (August 2000).
7. A. J. Benson, C. S. Frenk, C. G. Lacey, C. M. Baugh and S. Cole, The effects of photoionization on galaxy formation - ii. satellite galaxies in the local group, *MNRAS* **333**, 177 (jun 2002).
8. F. Governato, C. Brook, L. Mayer, A. Brooks, G. Rhee, J. Wadsley, P. Jonsson, B. Willman, G. Stinson, T. Quinn and P. Madau, Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows, *Nature* **463**, 203 (January 2010).
9. T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, T. Theuns, R. G. Bower, R. A. Crain, M. Furlong, A. Jenkins, M. Schaller and J. Schaye, The chosen few: the low-mass haloes that host faint galaxies, *MNRAS* **456**, 85 (February 2016).
10. M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky and O. Ruchayskiy, The haloes of bright satellite galaxies in a warm dark matter universe, *MNRAS* **420**, 2318 (March 2012).
11. M. Vogelsberger, J. Zavala and A. Loeb, Subhaloes in self-interacting galactic dark matter haloes, *MNRAS* **423**, 3740 (July 2012).
12. M. R. Lovell, V. Gonzalez-Perez, S. Bose, A. Boyarsky, S. Cole, C. S. Frenk and O. Ruchayskiy, Addressing the too big to fail problem with baryon physics and sterile neutrino dark matter, *MNRAS* **468**, 2836 (July 2017).
13. D. S. Akerib, S. Alsum, H. M. Araújo, X. Bai, A. J. Bailey, J. Balajthy, P. Beltrame, E. P. Bernard, A. Bernstein, T. P. Biesiadzinski, E. M. Boulton, R. Bramante, P. Brás, D. Byram, S. B. Cahn, M. C. Carmona-Benitez, C. Chan, A. A. Chiller, C. Chiller, A. Currie, J. E. Cutter, T. J. R. Davison, A. Dobi, J. E. Y. Dobson, E. Druszkiewicz, B. N. Edwards, C. H. Faham, S. Fiorucci, R. J. Gaitskell, V. M. Gehman, C. Ghag, K. R. Gibson, M. G. D. Gilchriese, C. R. Hall, M. Hanhardt, S. J. Haselschwardt, S. A. Hertel, D. P. Hogan, M. Horn, D. Q. Huang, C. M. Ignarra, M. Ihm, R. G. Jacobsen, W. Ji, K. Kamdin, K. Kazkaz, D. Khaitan, R. Knoche, N. A. Larsen, C. Lee, B. G. Lenardo, K. T. Lesko, A. Lindote, M. I. Lopes, A. Manalaysay, R. L. Mannino, M. F. Marzioni, D. N. McKinsey, D. M. Mei, J. Mock, M. Moongweluwan, J. A. Morad,

A. S. J. Murphy, C. Nehrkorn, H. N. Nelson, F. Neves, K. O'Sullivan, K. C. Oliver-Mallory, K. J. Palladino, E. K. Pease, P. Phelps, L. Reichhart, C. Rhyne, S. Shaw, T. A. Shutt, C. Silva, M. Solmaz, V. N. Solovov, P. Sorensen, S. Stephenson, T. J. Sumner, M. Szydagis, D. J. Taylor, W. C. Taylor, B. P. Tennyson, P. A. Terman, D. R. Tiedt, W. H. To, M. Tripathi, L. Tvrznikova, S. Uvarov, J. R. Verbus, R. C. Webb, J. T. White, T. J. Whitis, M. S. Witherell, F. L. H. Wolfs, J. Xu, K. Yazdani, S. K. Young and C. Zhang, Results from a Search for Dark Matter in the Complete LUX Exposure, *Phys. Rev. Lett.* **118**, p. 021303 (January 2017).

14. E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, F. D. Amaro, M. Anthony, F. Arneodo, L. Baudis, B. Bauermeister, M. L. Benabderrahmane, T. Berger, P. A. Breur, A. Brown, A. Brown, E. Brown, S. Bruenner, G. Bruno, R. Budnik, C. Capelli, J. M. R. Cardoso, D. Cichon, D. Coderre, A. P. Colijn, J. Conrad, J. P. Cussonneau, M. P. Decowski, P. de Perio, P. Di Gangi, A. Di Giovanni, S. Diglio, A. Elykov, G. Eurin, J. Fei, A. D. Ferella, A. Fieguth, W. Fulgione, A. Gallo Rosso, M. Galloway, F. Gao, M. Garbini, C. Geis, L. Grandi, Z. Greene, H. Qiu, C. Hasterok, E. Hogenbirk, J. Howlett, R. Itay, F. Joerg, B. Kaminsky, S. Kazama, A. Kish, G. Koltman, H. Landsman, R. F. Lang, L. Levinson, Q. Lin, S. Lindemann, M. Lindner, F. Lombardi, J. A. M. Lopes, J. Mahlstedt, A. Manfredini, T. Marrodán Undagoitia, J. Masbou, D. Masson, M. Messina, K. Micheneau, K. Miller, A. Molinario, K. Morå, M. Murra, J. Naganoma, K. Ni, U. Oberlack, B. Pelssers, F. Piastra, J. Pienaar, V. Pizzella, G. Plante, R. Podviianiuk, N. Priel, D. Ramírez García, L. Rauch, S. Reichard, C. Reuter, B. Riedel, A. Rizzo, A. Rocchetti, N. Rupp, J. M. F. dos Santos, G. Sartorelli, M. Scheibelhut, S. Schindler, J. Schreiner, D. Schulte, M. Schumann, L. Scotto Lavina, M. Selvi, P. Shagin, E. Shockley, M. Silva, H. Simgen, D. Thers, F. Toschi, G. Trinchero, C. Tunnell, N. Upole, M. Vargas, O. Wack, H. Wang, Z. Wang, Y. Wei, C. Weinheimer, C. Wittweg, J. Wulf, J. Ye, Y. Zhang and T. Zhu, Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, *Physical Review Letters* **121**, p. 111302 (September 2018).
15. J. M. Gaskins, A review of indirect searches for particle dark matter, *Contemporary Physics* **57**, 496 (October 2016).
16. T. R. Slatyer, TASI Lectures on Indirect Detection of Dark Matter, *ArXiv e-prints* (October 2017).
17. H. C. Turner, M. R. Lovell, J. Zavala and M. Vogelsberger, The onset of gravothermal core collapse in velocity-dependent self-interacting dark matter subhaloes, *MNRAS* **505**, 5327 (August 2021).
18. M. R. Lovell, M. Cautun, C. S. Frenk, W. A. Hellwing and O. Newton, The spatial distribution of Milky Way satellites, gaps in streams, and the nature of dark matter, *MNRAS* **507**, 4826 (November 2021).
19. M. R. Lovell, D. Iakubovskyi, D. Barnes, S. Bose, C. S. Frenk, T. Theuns and W. A. Hellwing, Simulating the Dark Matter Decay Signal from the Perseus Galaxy Cluster, *ApJ* **875**, p. L24 (April 2019).
20. M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, The Milky Way's bright satellites as an apparent failure of ΛCDM, *MNRAS* **422**, 1203 (May 2012).
21. M. G. Walker and J. Peñarrubia, A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies, *ApJ* **742**, p. 20 (November 2011).
22. J. Zavala, M. R. Lovell, M. Vogelsberger and J. D. Burger, Diverse dark matter density at sub-kiloparsec scales in Milky Way satellites: Implications for the nature of dark matter, *Phys. Rev. D* **100**, p. 063007 (Sep 2019).
23. S. Balberg, S. L. Shapiro and S. Inagaki, Self-Interacting Dark Matter Halos and the Gravothermal Catastrophe, *ApJ* **568**, 475 (April 2002).

24. H. Nishikawa, K. K. Boddy and M. Kaplinghat, Accelerated core collapse in tidally stripped self-interacting dark matter halos, *Phys. Rev. D* **101**, p. 063009 (March 2020).
25. A. H. G. Peter, M. Rocha, J. S. Bullock and M. Kaplinghat, Cosmological Simulations with Self-Interacting Dark Matter II: Halo Shapes vs. Observations, *Mon. Not. Roy. Astron. Soc.* **430**, p. 105 (2013).
26. M. J. Jee, H. Hoekstra, A. Mahdavi and A. Babul, Hubble Space Telescope/Advanced Camera for Surveys Confirmation of the Dark Substructure in A520, *ApJ* **783**, p. 78 (March 2014).
27. F. Kahlhoefer, K. Schmidt-Hoberg, J. Kummer and S. Sarkar, On the interpretation of dark matter self-interactions in Abell 3827, *MNRAS* **452**, L54 (September 2015).
28. V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk and S. D. M. White, The Aquarius Project: the subhaloes of galactic haloes, *MNRAS* **391**, 1685 (December 2008).
29. V. Springel, E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh, *MNRAS* **401**, 791 (January 2010).
30. D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack and E. L. Wright, First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters, *ApJS* **148**, 175 (September 2003).
31. T. Asaka and M. Shaposhnikov, The @nMSM, dark matter and baryon asymmetry of the universe [rapid communication], *Physics Letters B* **620**, 17 (July 2005).
32. M. Laine and M. Shaposhnikov, Sterile neutrino dark matter as a consequence of νMSM-induced lepton asymmetry, *J. Cosmology Astropart. Phys.* **6**, p. 31 (June 2008).
33. A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi and J. Franse, Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster, *Physical Review Letters* **113**, p. 251301 (December 2014).
34. E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein and S. W. Randall, Detection of an Unidentified Emission Line in the Stacked X-Ray Spectrum of Galaxy Clusters, *ApJ* **789**, p. 13 (July 2014).
35. P. Bode, J. P. Ostriker and N. Turok, Halo Formation in Warm Dark Matter Models, *ApJ* **556**, 93 (July 2001).
36. D. Erkal and V. Belokurov, Properties of dark subhaloes from gaps in tidal streams, *MNRAS* **454**, 3542 (December 2015).
37. N. Banik, J. Bovy, G. Bertone, D. Erkal and T. J. L. de Boer, Novel constraints on the particle nature of dark matter from stellar streams, *arXiv e-prints*, p. arXiv:1911.02663 (Nov 2019).
38. A. Bonaca, D. W. Hogg, A. M. Price-Whelan and C. Conroy, The Spur and the Gap in GD-1: Dynamical Evidence for a Dark Substructure in the Milky Way Halo, *ApJ* **880**, p. 38 (July 2019).
39. W. A. Hellwing, C. S. Frenk, M. Cautun, S. Bose, J. Helly, A. Jenkins, T. Sawala and M. Cytowski, The Copernicus Complexio: a high-resolution view of the small-scale Universe, *MNRAS* **457**, 3492 (April 2016).
40. S. Bose, W. A. Hellwing, C. S. Frenk, A. Jenkins, M. R. Lovell, J. C. Helly and B. Li, The Copernicus Complexio: statistical properties of warm dark matter haloes, *MNRAS* **455**, 318 (January 2016).
41. M. Viel, G. D. Becker, J. S. Bolton and M. G. Haehnelt, Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data, *Phys. Rev. D* **88**, p. 043502 (August 2013).

42. M. Cautun, A. Benítez-Llambay, A. J. Deason, C. S. Frenk, A. Fattahi, F. A. Gómez, R. J. J. Grand, K. A. Oman, J. F. Navarro and C. M. Simpson, The milky way total mass profile as inferred from Gaia DR2, *MNRAS* **494**, 4291 (May 2020).
43. A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, The Role of Sterile Neutrinos in Cosmology and Astrophysics, *Annual Review of Nuclear and Particle Science* **59**, 191 (November 2009).
44. M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-α forest, *Phys. Rev. D* **71**, p. 063534 (March 2005).
45. A. Boyarsky, J. Lesgourgues, O. Ruchayskiy and M. Viel, Lyman-α constraints on warm and on warm-plus-cold dark matter models, *J. Cosmology Astropart. Phys.* **5**, p. 12 (May 2009).
46. Y. M. Bahé, D. J. Barnes, C. Dalla Vecchia, S. T. Kay, S. D. M. White, I. G. McCarthy, J. Schaye, R. G. Bower, R. A. Crain, T. Theuns, A. Jenkins, S. L. McGee, M. Schaller, P. A. Thomas and J. W. Trayford, The Hydrangea simulations: galaxy formation in and around massive clusters, *MNRAS* **470**, 4186 (October 2017).
47. D. J. Barnes, S. T. Kay, Y. M. Bahé, C. Dalla Vecchia, I. G. McCarthy, J. Schaye, R. G. Bower, A. Jenkins, P. A. Thomas, M. Schaller, R. A. Crain, T. Theuns and S. D. M. White, The Cluster-EAGLE project: global properties of simulated clusters with resolved galaxies, *MNRAS* **471**, 1088 (October 2017).

Addressing classical cosmological back-reaction with multiple scales

Yonadav Barry Ginat
Faculty of Physics, Technion – Israel Institute of Technology, Haifa, 3200003, Israel
E-mail: ginat@campus.technion.ac.il

This is a summary of my talk at the 16th Marcel Grossmann conference, given on-line in July 2021. Various aspects of the averaging problem – the problem of finding an effective large-scale cosmological solution of the Einstein field equations, when small-scale perturbations are present — are discussed, and treated with the multiple-scales technique of singular perturbation theory. This allows one to show that a split between a background metric that varies only on large scales, and perturbations to it, is consistent, provided certain conditions are met. I finish by giving an explicit example of the back-reaction of a perturbation consisting of a single small-scale mode, and point at possible future directions.

Keywords: Cosmological perturbation theory; gravity.

1. Introduction

The Universe is usually described, on large scales, by a Friedmann-Lemaître-Robertson-Walker (FLRW) metric[1]

$$\mathrm{d}s^2 = a^2(\eta)\left(-\mathrm{d}\eta^2 + \gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j\right), \tag{1}$$

which is a solution of the Einstein equations

$$R_{ab}(g_{ab}) - \Lambda g_{ab} = 8\pi G\rho_{ab} \equiv 8\pi G\left(T_{ab} - \frac{1}{2}Tg_{ab}\right), \tag{2}$$

for a homogeneous and isotropic energy-momentum tensor T_{ab}. However, in reality, the matter distribution is inhomogeneous on small scales, and becomes homogeneous and isotropic only on scales of about[2] $\sim$ 100 Mpc so an averaging procedure is required, to find a homogeneous $\overline{\rho}_{ab}$. Such averaging does not, for a general metric, commute with the non-linear operation of computing the Ricci tensor from the metric, so $\overline{R}_{ab} \neq R_{ab}(\overline{g}_{ab})$. Hence, the equation, whose solution is the FLRW metric, does not, strictly speaking, correspond to a realistic universe. An averaged equation should rather be

$$\overline{R}_{ab} - \Lambda\overline{g}_{ab} = 8\pi G\overline{\rho}_{ab}. \tag{3}$$

This discrepancy is a consequence of the averaging, or back-reaction problem.[3–6]

Another instance of the same problem arises in the context of splitting the space-time metric g_{ab} into an FLRW 'background' piece, g^{FLRW}_{ab} and a perturbation, h_{ab}, *viz.*

$$g_{ab} = g^{\mathrm{FLRW}}_{ab} + h_{ab}, \tag{4}$$

where in standard cosmological perturbation theory, h_{ab} is taken to be small.[1] However, its magnitude depends on scale and, in general, on the co-ordinate system one uses;[7] this calls into question the very validity of such a background-perturbation split, which rests on the consistency of even defining a large-scale background.

Several proposals were put forward in the past to address this problem, including special averaging techniques[8] or special perturbative expansions,[9] and here I wish to describe another proposal, laid down in Ginat (2021),[10] which can address some of the problems with less assumptions than the other techniques.

Below I choose units such that $c = 8\pi G = 1$. The entire discussion here is classical, and quantum aspects of back-reaction are not discussed at all.

2. Multiple-Scales Treatment of the Einstein Equations

2.1. *The Multiple-Scales Method*

The multiple-scales technique consists in treating small scales and large scales as separate variables, interacting via the non-linearity in the dynamical equations that one wishes to study.[11] If ε denotes a certain small parameter involved in the equations, and x denotes the independent variable, then one defines $X = x/\varepsilon$, and changes

$$\frac{\partial}{\partial x^a} \mapsto \frac{\partial}{\partial x^a} + \frac{1}{\varepsilon}\frac{\partial}{\partial X^a} \tag{5}$$

in the dynamical equations; then, one treats x and X as independent variables. This split is done in conjunction with an asymptotic expansion of the dependent variable, the metric $g_{ab}(x, X, \varepsilon)$, in our case, in ε, as $g_{ab}(x, X) \sim g^0_{ab} + \varepsilon g^1_{ab} + \ldots$, and a similar one for the energy-momentum tensor (with the convention that any X-independent terms enter into ρ^0_{ab}). This expansion is consistent, of course, only if the condition $\varepsilon^n \left|g^n_{ab}\right| \gg \varepsilon^{n+1} \left|g^{n+1}_{ab}\right|$ is satisfied.

The addition of an independent variable, without changing the dynamics, implies that there aren't enough equations to determine the solution fully, leaving certain integration constants, or solutions of homogeneous equations, free. However, in the general case, small-scale resonances yield diverging large-scale behaviour, which potentially violates the above-mentioned consistency requirement. To ensure that it is not, one imposes additional consistency conditions, which determine the undetermined functions or integration constants, in precisely such a manner as to remove the resonances, and ensure the asymptoticity of the expansion. This adds exactly enough equations to fix the solution fully.

2.2. *Application to The Einstein Equations*

Working in harmonic co-ordinates renders the Einstein equations hyperbolic,[12] *viz.*

$$R_{ab} \equiv R^{(h)}_{ab} = -\frac{1}{2} g^{cd} \partial^2_{cd} g_{ab} + P^{cdefgh}_{ab}(g) \partial_c g_{ef} \partial_d g_{gh}, \tag{6}$$

where $P_{ab}^{cdefgh}(g)$ is some function of the metric g, and not of its derivatives. Let us now set x as the cosmological scale and X as some small scale, e.g. a galactic scale; then ε is the ratio between these two scales, and indeed $\varepsilon \ll 1$.

The Einstein equations are second order, so, upon expanding them in ε, the leading order is ε^{-2}, i.e., in harmonic co-ordinates,

$$-\frac{1}{2}g_0^{cd}\partial^2_{X^cX^d}g^0_{ab} + P_{ab}^{cdefgh}(g_0)\partial_{X^c}g^0_{ef}\partial_{X^d}g^0_{gh} = \rho^{-2}_{ab}. \tag{7}$$

Here we can already address the question of the existence of a consistent background-perturbation split for the metric: if $\rho^{-2}_{ab} \neq 0$, i.e. if there are small-scale, relativistic sources, like black holes, then necessarily g^0_{ab} is X *dependent*, and therefore any good definition of a large-scale background is impossible. On the other hand, if one only considers small-scale, 'Newtonian' sources, for which the gravitational potential is $\ll 1$, then ρ^{-2}_{ab} does vanish, and the solution to equation (7) is an X-independent g^0_{ab}, a true background.

2.3. *Consistency Conditions*

To establish the background-perturbation split, all that's left to do is to establish the consistency condition for εg^1_{ab}, that is, to show that it remains small relative to g^0_{ab}. The order ε^{-1} equation is

$$\rho^{-1}_{\mathrm{rel},ab} + \rho^{-1}_{\mathrm{nr},ab} = -\frac{1}{2}g_0^{cd}\partial^2_{X^cX^d}g^1_{ab}, \tag{8}$$

where the subscript 'rel' ('nr') is there, to remind one that this is a relativistic (Newtonian) contribution to the energy-momentum tensor, one which arises, upon Fourier transforming in X, from matter modes which have a relativistic (non-relativistic) dispersion relation. In fact, because $\partial_X g^0 = 0$, in-so-far-as the X-co-ordinates are concerned, the space-time over which g^1_{ab} is defined *qua* a tensor-valued function, is flat: equation (8) is a wave equation for g^1_{ab}, with an X-independent metric $g^0_{ab}\mathrm{d}X^a\mathrm{d}X^b$, so its solutions are waves in flat space-time, parameterised by the large-scale co-ordinate x. This also means that a Fourier transform in X is unambiguous.

Since $\rho^{-1}_{\mathrm{rel},ab}$ is in resonance with the wave-equation operator, the consistency of the asymptotic expansion requires that, in fact, it vanish. If it does, then εg^1_{ab} remains smaller than g^0_{ab}.[10] So far there have been two non-trivial consistency conditions, which limit the applicability to Newtonian sources. Relativistic ones are allowed, though, at all orders, from ε^0 and above. One can generalise this to conclude that whenever $\rho^n_{\mathrm{rel},ab} \neq 0$, one needs an X-dependent g^{n-1}_{ab}, that will precisely cancel the resonances that arise. For example, at $n = 2$, this is the case, and after a few manipulations, one finds, for $k^2 = 0$ (where k is the Fourier conjugate of X)

the following consistency condition:

$$-k^c\partial_{x^c}\hat{g}^1_{\mathrm{hom},ab}(x,k)+P^{cdefgh}_{ab}(g_0)\left(k_c\hat{g}^1_{\mathrm{hom},ef}(x,k)\partial_{x^d}g^0_{gh}+k_d\hat{g}^1_{\mathrm{hom},gh}(x,k)\partial_{x^c}g^0_{ef}\right)$$
$$-\mathrm{i}\int \mathrm{d}^4X e^{\mathrm{i}k\cdot X}\left\{P_{ab}(g_0)\partial_X g_1\partial_X g_1-\frac{1}{2}g_1^{cd}\partial^2_{X^cX^d}g^1_{ab}\right\}_{\mathrm{osc}}=-8\pi\mathrm{i}G\hat{\rho}^0_{ab}(x,k), \quad (9)$$

where a circumflex denotes a Fourier transform in X, a subscript 'osc' means that one removes the X average of the term in the curly brackets from it, and $g^1_{\mathrm{hom},ab}$ is the homogeneous part of the solution of equation (8).

This is the key element of the multiple-scales expansion: the additional degrees of freedom, that are created by treating x and X separately, are fixed by requiring the asymptoticity of the expansion in ε – by removing resonances order-by-order. Indeed, as there is an un-fixed homogeneous part in g^1_{ab}, relativistic sources in ρ^0_{ab} are not a problem.

The X-independent part of the Einstein equation at order ε^0 also gives an effective Einstein equation, governing the large-scale dependence of g^0_{ab}:[10]

$$-\frac{1}{2}g_0^{cd}\partial^2_{x^cx^d}g^0_{ab}+P_{ab}(g_0)\partial_x g_0\partial_x g_0-\Lambda g^0_{ab}-B_{ab}=8\pi G\langle\rho^0_{ab}\rangle_X. \quad (10)$$

When computed in a co-ordinate system of a freely-falling observer in the x-spacetime, the back-reaction tensor B_{ab} corresponds, roughly, to the X-averaged Newtonian gravitational potential energy of the sources in $\rho^{-1}_{\mathrm{nr},ab}$. The averaged equation agrees with the one derived by Green & Wald (2011)[9] under different assumptions.

2.4. *Single Short Mode*

Let me also give a simple example of a situation where $B_{ab}\neq 0$. Consider a cosmological setting, where, in addition to the uniform matter density $\mu_0(\eta)$, the matter distribution has a component $\delta\mu(\eta,\mathbf{x})=2\varepsilon\mu_0(\eta)\rho_0\cos(\mathbf{k}\cdot\mathbf{x}/\varepsilon)$, where $\mathbf{k}$ is a constant of order one, and $\rho_0\mu_0$ is much smaller than the critical density of the Universe.[a] Here, η, $\mathbf{x}$ are conformal, co-moving co-ordinates. The back-reaction tensor, in this case, is given by[10]

$$B^{ij}=\frac{\mu_0^2\rho_0^2}{2a^2k^4}\left(k^ik^j+k^2\delta_{ij}\right), \quad (11)$$

with all other components equal to zero, to leading order in ρ_0. This back-reaction, albeit small, introduces, unsurprisingly, anisotropic stress perturbations.

3. Conclusions

In conclusion, the multiple-scales technique enables one to show that there is a well-defined background metric that depends only on the large-scale, as well as to derive an effective Einstein equation where small-scale gravitational potential

[a]This latter assumption is not strictly necessary, but simplifies matters considerably, in that it allows one to expand in $\rho_0\mu_0/\rho_{\mathrm{crit}}$, which is an expansion in large-scale cosmological perturbations, and then proceed as one would do for standard cosmological perturbation theory.

energy acts as a source of gravity on the large scale. Additionally, one can determine the consistency conditions under which these conclusions obtain, and, in particular, that they do obtain when small-scale over-densities are chiefly due to non-relativistic matter, to leading order in ε.

In the future it would be worthwhile to try to expand this technique from asymptoticity to a convergence property, e.g. two-scale convergence,[13] and also, perhaps more importantly for practical applications, to attempt to include small-scale relativistic objects, such as neutron stars or black holes, perhaps by utilising some extension of Birkhoff's theorem.[14] It is plausible that this theorem would allow one to include isolated black holes in the multiple-scales framework, by relaxing the requirement $\varepsilon \left|g^1_{ab}\right| \ll \left|g^0_{ab}\right|$, and allowing g^1_{ab} to take extremely large values in some very restricted areas of space-time. This should enable one to conclude from Birkhoff's theorem or something similar that an observer far from any isolated black holes will not observe any more back-reaction than if these black holes were normal stars with the same masses.

References

1. S. Weinberg, *Cosmology* (Oxford University Press, Oxford, 2008).
2. A. A. Coley and G. F. R. Ellis, Theoretical cosmology, *Classical and Quantum Gravity* **37**, p. 013001 (January 2020).
3. T. Futamase, Averaging of a locally inhomogeneous realistic universe, *Phys. Rev. D* **53**, 681 (Jan 1996).
4. R. M. Zalaletdinov, Averaging problem in general relativity, macroscopic gravity and using Einstein's equations in cosmology, *Bull. Astron. Soc. India* **25**, 401 (1997).
5. C. Clarkson, G. Ellis, J. Larena and O. Umeh, Does the growth of structure affect our dynamical models of the universe? The averaging, backreaction and fitting problems in cosmology, *Rept. Prog. Phys.* **74**, p. 112901 (2011).
6. T. Buchert and S. Räsänen, Backreaction in Late-Time Cosmology, *Annual Review of Nuclear and Particle Science* **62**, 57 (Nov 2012).
7. J. Adamek, C. Clarkson, D. Daverio, R. Durrer and M. Kunz, Safely smoothing spacetime: backreaction in relativistic cosmological simulations, *Classical and Quantum Gravity* **36**, p. 014001 (dec 2018).
8. T. Buchert, On Average Properties of Inhomogeneous Fluids in General Relativity: Dust Cosmologies, *General Relativity and Gravitation* **32**, 105 (Jan 2000).
9. S. R. Green and R. M. Wald, New framework for analyzing the effects of small scale inhomogeneities in cosmology, *Phys. Rev. D* **83**, p. 084020 (Apr 2011).
10. Y. B. Ginat, Multiple-scales approach to the averaging problem in cosmology, *JCAP* **2021**, p. 049 (February 2021).
11. G. A. Pavliotis and A. M. Stuart, *Multiscale methods*, Texts in Applied Mathematics, Vol. 53 (Springer, New York, 2008), Averaging and homogenization.
12. Y. Choquet-Bruhat, *General relativity and the Einstein equations*Oxford Mathematical Monographs, Oxford Mathematical Monographs (Oxford University Press, Oxford, 2009).
13. G. Allaire, Homogenization and two-scale convergence, *SIAM J. Math. Anal.* **23**, 1482 (1992).
14. A. Einstein and E. G. Straus, The Influence of the Expansion of Space on the Gravitation Fields Surrounding the Individual Stars, *Reviews of Modern Physics* **17**, 120 (April 1945).

Imaging formation process for DM profiles

Omar de J. Cabrera-Rosas* and Tonatiuh Matos

Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, México Distrito Federal, A.P. 14-740, 07000, México

**E-mail: omar.cabrera@cinvestav.mx*

The imaging formation process in halos for some dark matter profiles is studied. Approaching these models on a small scale, we analyze the images generated on the lens plane by obtaining the analytical surface mass densities $\Sigma(x)$ and their corresponding deflection angles $\alpha(x)$. We identify the presence of Einstein rings, by mapping fringes that represent possible sources (such as other galaxies), placed on the source plane. We approach the simplest case, where lines parallel to the x axis are mapped onto the lens plane, to find out how are the solutions of the $\mathbf{X}$ vector field, which is in this case, the geometrical equivalent to the usual lens mapping.

Keywords: Dark Matter, Gravitational Lensing

1. Introduction

The dynamics of galaxies is a topic of great interest in cosmology. In recent years, this dynamic has been studied with the observation and analysis of the rotation curves of several samples of galaxies. From these observations an acceleration has been measured, and the presence of Dark Matter (DM) is necessary to explain their dynamics.[1,2]

The restrictions belonging to the central surface density of the halo, determine the value of the quantity μ_{DM} which is always a constant, that it seems to be an universal invariant, and there are evidences of this restriction in spiral, dwarf irregulars and ellipticals galaxies to name some types.[3–6] It is important to observe that μ_{DM} is constructed as the product of the characteristic length r_s and the central density value ρ_s of each galaxy, and this is related with the *soliton* region (the core of the galaxy). This is an important zone, because the extra galactic components do not alter this region and the invariant objects present there, are very helpful to understand the complete behavior of the galaxy.[7] In fact, certain models for Scalar Field Dark Matter (SFDM) have predicted that the total mass of these systems is $M_{DM}(300\text{pc}) \sim 10^7 M_\odot$ with a characteristic size $r_s = 300\text{pc}$; that is, the soliton region is perfectly delimited for its analysis, and it is very important for eventual tests that improves the physics interpretation of the data. Therefore, to study the dynamical process for translating this information to the imaging information, this soliton zone must be analyzed, with the corresponding considerations for that region.

In fact, the universal DM profile (the Navarro-Frenk-White profile NFW) has been exhaustively studied, because it fits properly with observational data[8,9];

however, it presents problems on a small scale, which implicates that there must be another models, that explain the core zone.

For the above reasons, a complete optical analysis of this region is in order, using the proper data that comprehends all the physical information encoded in each model, but at the same time with a simplification of the studied equations. Hence, we use the usual gravitational lensing formalism,[10] that contains all the mathematical restrictions to fully understand such optical processes. Hence, we compute the images on the lens plane, as functions of the positions of the sources, the source plane and the redshift, which is encoded into the critical surface mass density Σ_{cr}, that appears in the volumetric density $\rho(r)$ (and thus, into the deflection angle $\alpha(r)$ generated in each physical situation). Therefore, the principal goal for using this approach, is to describe the images generated for recent models of DM, and test which of them behave like the observed data. In this stage of the work, we only study the images generated by objects (fringes) very near to the observation axis, by finding the roots of the equation that represents the mapping in that zone (see Eq. 13 below). This eventually means to translate the optical information obtained from the corresponding images, and link it with the information of the galactic dynamics.

This work is organized as follows. In section 2 we set the basic equations for gravitational lensing, and we obtain the surface mass density and the deflection angle analytically for each profile. In section 3, we establish the $\mathbf{X}$ vector field that describes the ray tracing of the optical process; later we use the method provided in,[11] to study the images that will appear on the lens plane. Finally in section 4, we present the conclusions of this work.

2. Basics equations for gravitational lensing

The configuration for gravitational lensing is as usual[10] , where the source plane, the lens plane and the observer are shown. D_s, D_l and D_{ls} are the distances between observer and source, observer and lens and between lens and source planes, respectively.[10] Now, assume that the DM density profile is given in the form $\rho(x) = \rho_s f(x)$, with ρ_s, r_s and $x = r/r_s$ are the characteristic density, the characteristic radius and the scale radius, respectively; $f(x)$ is any given function of its argument.[2] Hence, it is necessary to find the deflection angle in each case, and because the systems considered here are axially symmetric, the surface mass density and the deflection angle are given by[10]

$$\begin{aligned} \Sigma(\xi) &= \int_{-\infty}^{\infty} \rho(\xi, z) dz, \\ \alpha(x) &= \frac{m(x)}{x}, \end{aligned} \tag{1}$$

Name	$f(x)$	$\Sigma(x)$	$\alpha(x) = m(x)/x$
(1)	(2)	(3)	(4)
Burkert	$\frac{1}{(1+x)(1+x^2)}$	$\beta_B F(x)_B$	$\frac{\sigma_B}{x} g(x)_B$
M-SFDM	$\frac{\sin^2 x}{x^2}$	$\beta_{MsT} F(x)_{MSt}$	$\sigma_{MsT}\left[J_1(2x) + \frac{x^3}{3}\right]$
PI	$\frac{1}{1+x^2}$	$\frac{\beta_{PI}}{\sqrt{1+x^2}}$	$\frac{\sigma_{PI}}{x}(\sqrt{x^2+1} - 1)$
Spano	$\frac{1}{(1+x^2)^{3/2}}$	$\frac{\beta_{Sp}}{1+x^2}$	$\frac{\sigma_{Sp}}{x}\ln(1+x^2)$
Wave DM	$\frac{1}{(1+x^2)^8}$	$\frac{\beta_W}{(x^2+1)^{15/2}}$	$\frac{\sigma_W}{x}\left(1 - \frac{1}{(x^2+1)^{13/2}}\right)$
NFW	$\frac{1}{x(1+x)^2}$	$\frac{2\mu_{DM}}{x^2-1} F(x)_{NFW}$	$\frac{4\kappa_s}{x}\left(\ln\frac{x}{2} + \frac{2}{\sqrt{1-x^2}}\operatorname{arctanh}\sqrt{\frac{1-x}{1+x}}\right)$

where the radial coordinate r is related to cylindrical polar coordinates by $r = \sqrt{\xi^2 + z^2}$ and $m(x)$ is the dimensionless mass defined by

$$m(x) = 2\int_0^x \kappa(\xi)\xi d\xi. \tag{2}$$

The shear of the system is

$$\kappa(x) = \Sigma(x)/\Sigma_{cr} \tag{3}$$

and $\Sigma_{cr} = c^2 D_s/(4\pi G D_l D_{ls})$ is the critical surface mass density.

Using Eqs. (1) and (2), it is obtained the above table, by using the Burkert, Multistate-SFDM, PI, Spano, Wave DM and NFW dark matter profiles given in,[2,7] where β and σ are constants that determine the scaling of such core, and they depend on the values of μ_{DM} and κ_s.

For the Burkert profile the $F(x)_B$ function for the surface mass density is given by

$$\begin{aligned} F(x)_B &= \frac{1}{\sqrt{1+x^2}}\left[\frac{\pi}{2} + \operatorname{arctanh}\left(\frac{1}{\sqrt{1+x^2}}\right)\right] \\ &-2\begin{cases} \frac{1}{\sqrt{1-x^2}}\operatorname{arctanh}\sqrt{\frac{1-x}{1+x}} & (x<1), \\ \frac{1}{\sqrt{x^2-1}}\arctan\sqrt{\frac{x-1}{x+1}} & (x>1), \end{cases} \\ &= \frac{1}{\sqrt{2}}\left[\frac{\pi}{2} + \operatorname{arctanh}\left(\frac{1}{\sqrt{2}}\right)\right] - 1 \quad (x=1), \end{aligned} \tag{4}$$

and the $g_B(x)$ function for the deflection angle is

$$g(x)_B = \frac{1}{2}\left[x^2\left(\frac{\pi}{2} + \operatorname{arctanh}\left(\frac{1}{\sqrt{1+x^2}}\right)\right) + \sqrt{1+x^2}\right] + \ln x$$
$$+2\begin{cases} \sqrt{1-x^2}\operatorname{arctanh}\sqrt{\dfrac{1-x}{1+x}} & (x<1), \\ -\sqrt{x^2-1}\arctan\sqrt{\dfrac{x-1}{x+1}} & (x>1), \end{cases} \tag{5}$$
$$= \frac{\pi}{4} + \frac{1}{\sqrt{2}} + \frac{1}{2}\operatorname{arctanh}\frac{1}{\sqrt{2}} \quad (x=1).$$

For the Multi-State profile, the F_{MSt} function for the surface mass density is given by

$$F(x)_{MSt} = J_0(2x) + \frac{\pi}{2}[J_1(2x)\mathbf{H}_0(2x) - J_0(2x)\mathbf{H}_1(2x)], \tag{6}$$

where J_ν are the Bessel functions of the first kind of ν-th order and $\mathbf{H}_\mu$ the Struve functions of μ-th order[12,13]; the F_{NFW} function and the corresponding deflection angle for the NFW profile are given in.[8,9]

3. Imaging formation

Knowing the deflection angle for each case, it is possible to develop the imaging formation process. It will be helpful to express the quantities for describing the optical process, in a cylindrical coordinate basis $\{\hat{\boldsymbol{\varrho}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{z}}\}$. From this information, the imaging mapping is given by the vector field

$$\mathbf{X} = \mathbf{r} + l\hat{\mathbf{R}}_G, \tag{7}$$

where $\hat{\mathbf{R}}_G$ is the gravitational refraction ray given by[11]

$$\hat{\mathbf{R}}_G = \cos\alpha\hat{\boldsymbol{z}} - \sin\alpha\hat{\boldsymbol{\varrho}}, \tag{8}$$

and l is the length from the lens plane $z=0$, to a point on a given deflected light ray.

Using Eq. (7) the deflected light rays are

$$\mathbf{X} = (\varrho - l\sin\alpha)\hat{\boldsymbol{\varrho}} + l\cos\alpha\hat{\boldsymbol{z}}. \tag{9}$$

Consider now a family of one dimensional sources in the region $z>0$, locally described by[11]

$$\mathbf{X} = \mathbf{X}(\eta, n), \tag{10}$$

where n denotes the source and η labels the points on that source. From Eqs. (9) and (10), the images that the observer sees on the lens plane $z=0$, given by all the

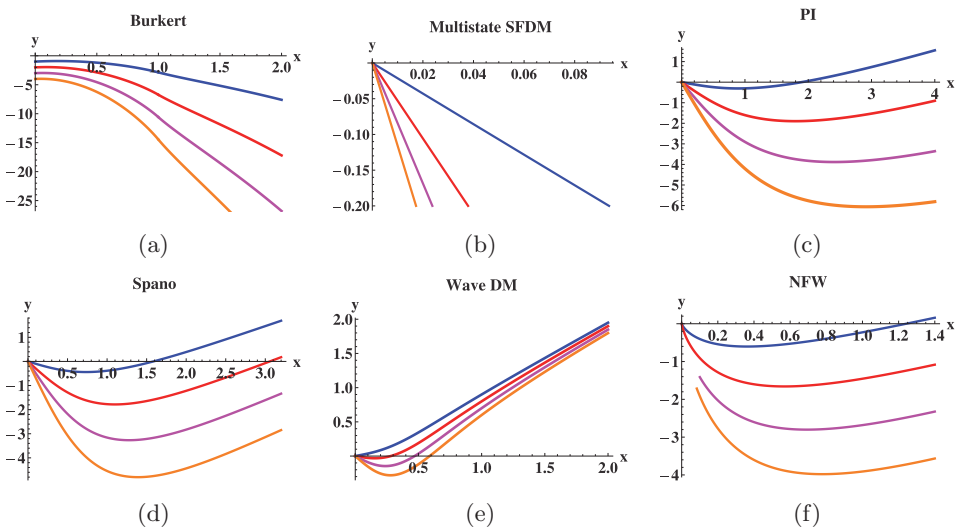


Fig. 1. Plots of the mapping $y = x - z_0\alpha$ for $z_0 = 1, 2, 3, 4$ and $n = 0$. The solutions for $x = 0$ in each profile denote the possible Einstein rings produced in each case. Notice that because in the Burkert and the multistate profiles, the curves does not intersect the horizontal axis, these profiles don't have Einstein rings.

points $(x\cos\phi, x\sin\phi, 0)$, where x and ϕ are solutions to[11]

$$\begin{aligned} X_s(\eta, n) &= [x - Z_s(\eta, n)\alpha(x)]\cos\phi, \\ Y_s(\eta, n) &= [x - Z_s(\eta, n)\alpha(x)]\sin\phi. \end{aligned} \tag{11}$$

Taking the simplest case for line sources that coincide with the x axis, the equations for obtaining the image for each fringe is determined by all the points in the lens plane such that[11]

$$X_s = [x - z_0\alpha(x)]\cos\phi, \tag{12}$$

$$n = [x - z_0\alpha(x)]\sin\phi, \tag{13}$$

with $-L \leq X_s \leq L$.

Now, Eq. (7) encodes the information related with the imaging formation process of the optical system, because describes how is the mapping of points from the lens plane to the source plane; the $\mathbf{X}$ is related with the usual lens mapping.

Because of the axial symmetry of the optical system, the caustic generated by the vector field (7), has a central branch (along the observation axis z). If this branch has contact with the source plane, the observed image is a ring; these are the *Einstein rings*.

In fact, for $n = 0$ (a source aligned exactly with the observation axis), these rings correspond to the roots of Eq. (13), and this is illustrated in Figs. 1. For example,

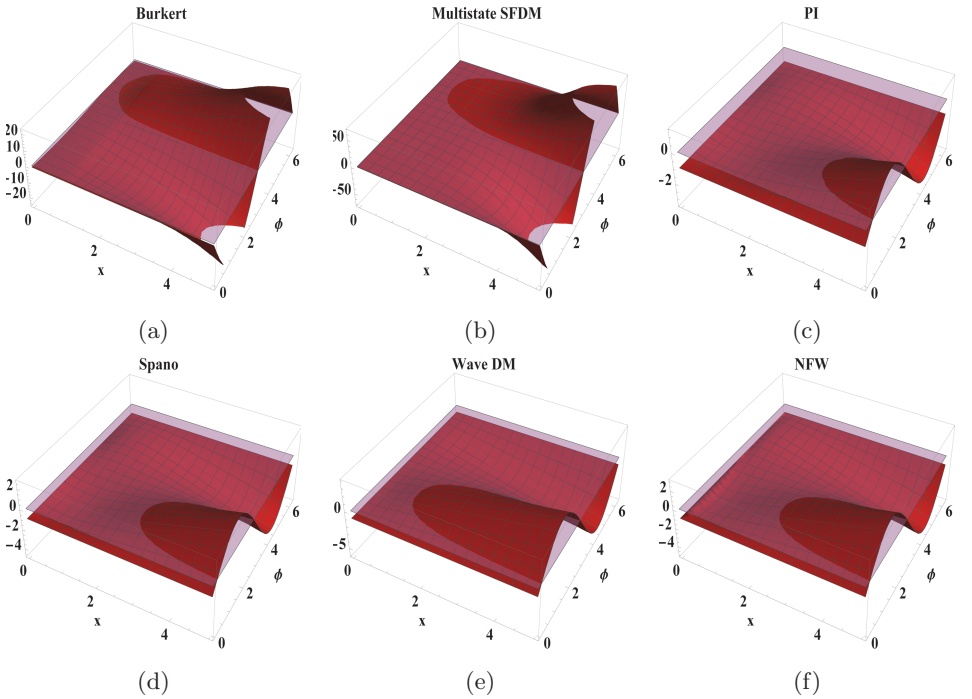


Fig. 2. Schematic of the surfaces representing the solutions of Eq. (13) for each profile on $z_0 = 1$ and $n = 1$. The pink plane intersecting each surface is the $P = 0$ plane. Observe that although the profiles present similar behaviour, the imaging formation process does not begin in the same values of x (neither the values of ϕ are the same).

observe that for the Burkert and Multistate profiles, this function have no roots, hence, they have no presence of Einstein rings for that positions of the source plane.

However, it is important to notice that these rings, are not the only images that could appear in this process. There is another branch of the caustic: a surface of revolution wrapped along the observation axis,[11] that also contributes to the images present on the lens plane, because a change to the source position does not lead to the change of the number of images unless a source moves along a caustic branch,[10] and this situation will appear, if the source and the lens plane are close enough, or in a position where this branch has contact with the source.

In this stage, we are only studying the simplest case, where we are not assuming that the source and the lens are close.

On the other hand, for $n \neq 0$, notice that in general, the deflection angle α is not an invertible function, therefore, it is necessary to find the solutions of Eq. (13) numerically. The above means to analyze Eq. (13) as a family of surfaces labeled by the n index, that are functions of x and ϕ (see Figs. 2).

By analyzing the intersections of these surfaces with the $P = 0$ plane, the image multiplicity according to the deflection angle α (see Fig. 2) can be found. The above

means to be able of identifying the differences on the values of x where the profile begins to generate images, or for the values of ϕ on each quadrant, where the images appear.

An important observation is that it has been shown that the DM central surface density μ_{DM} for all these profiles, is a constant (its values lie in the range $\mu_{DM} = (575 - 648) M_{\odot} \mathrm{pc}^{-2})$[2,7] and this fact will be very useful for eventual calculations on imaging formation, because this means that the r_s and ρ_s parameters are not independent from each other, and hence by fixing the conditions for one of them, we can establish the correct values for the β and σ parameters.

In fact, because the Navarro-Frenk-White (NFW) profile has been studied in several works,[8,9] it will be important to compare it with the other profiles, and these results must fit with the description of the physical behaviour in DM halos, for later comparing such information with that obtained in the NFW case based on experimental data, but only in the moment where all the profiles can be expressed in a comparable scale.

4. Conclusions

In this work, the analysis for the imaging formation process for dark matter halos has been performed. The goal is to understand if the optical information provided by the lens mapping, gives information about the physical processes that lies in the core of some types of galaxies.

The motivation behind this analysis, was to obtain the analytic equation for the optical mapping from the source plane to the lens plane, and translate directly such information into the images generated on the lens plane.

So far, we have find out that the formation of Einstein rings, depend on the solutions of Eq. (13), that is a projection of the $\mathbf{X}$ vector field (the equivalent of the usual lens mapping) on the lens plane. These calculations are helpful to differentiate the behavior of each profile, and this was carried out by obtaining the deflection angle, encoded in the $\mathbf{X}$ mapping, and later by analyzing the behavior of this function.

This was achieved by plotting the n surfaces that represent the solutions of Eq. (13) to identify the values that correspond to the zones where the imaging formation process begin.

An important observation is that, the ultimate goal for studying such process is to eventually address the problems of the ΛCDM on a small scale, because the NFW profile has problems in this region, and alternative models must be tested there. This means to have at hand an observational tool for testing the possible wave nature of the DM or the species diversity, basing the calculations in the soliton region, because it seems to be that this zone is established by means of an universal invariant μ_{DM} (with the correct values of r_s and ρ_s, that in turn determine our β and σ constants).

Finally, as a future goal based in this work, we consider that we will be able to "draw" the corresponding images mapped from sources (in this case lines) on the lens plane, that represent galaxies or other kind of cosmological objects, by making use of the **X** field. In fact, we already have used this process in the description of a Schwarzschild lens, and the plots allowed us to explicitly identify the positions of the images on the lens plane,[11] that is, by generating images to later compare them with visual data. We believe that the above, can be used as an efficient geometrical way to identify optical configurations, for later interpret the physical data encoded in such images in a nimbly way.

References

1. S. S. McGaugh, F. Lelli and J. M. Schombert, *Radial acceleration relation in rotationally supported galxies*, Phys. Rev. Lett., **117**, 201101 (2016).
2. L. A. Ureña-López, V. H. Robles and T. Matos, *Mass discrepancy-acceleration relation: A universal maximum dark matter acceleration and implications for the ultralight scalar dark matter model*, Phys. Rev. D. **96**, 043005 (2017).
3. M. Spano, M. Marcelin, P. Amram, C. Carignam, B. Epinat and O. Hernández, *GHASP: An H_α kinematic survey of spiral and irregular galaxies — V. Dark matter distribution in 36 nearby spiral galaxies*, MNRAS, 383, 297-316 (2008).
4. F. Donato, G. Gentile, P. Salucci et al,*A constant dark matter halo surface density in galxies* , MNRAS, 397, 1169-1176 (2009).
5. P. Salucci and A. Burkert, *Dark Matter scaling relations*, ApJ **537**, L9 (2000).
6. A. Burkert, *The structure and dark halo core properties of dwarf spheroidal galaxies*, ApJ **808** 158 (2015).
7. L. E. Padilla, J. Solís-López, T. Matos and A. A. Aviléz-López, *Consequences for the Scalar Field Dark Matter Model from the McGaugh Observed-baryon Acceleration Correlation*, The Astrophysical Journal, 909: 162 (16pp), (2021).
8. M. Meneghetti, M. Bartelmann and L. Moscardini, *Cluster cross-sections for strong lensing: analytic and numerical lens models*, Mon. Not. R. Astron. Soc., **340**, 105-114 (2003).
9. G. Golse and J. P. Kneib, *Pseudo elliptical lensing mass model: Application to the NFW mass distribution*, A&A **390**, 821-827 (2002).
10. P. Schneider, C. Kochanek and J. Wambsganss, *Gravitational lensing: strong, weak and micro*, Springer (2006).
11. N. Bretón, O. de J. Cabrera-Rosas, E. Espíndola-Ramos, S. A. Juárez-Reyes, I. Julián-Macías, A. Montiel, P. Ortega-Vidals, E. Román-Hernández, G. Silva-Ortigoza, C. T. Sosa-Sánchez and R. Suárez-Xique, *Towards the Ronchi test for gravitational lenses: the gravitoronchigram*, J. Opt. **19** (2017) 065602 (19pp).
12. I. Gradshteyn and I. Ryzhik, *Table of integrals, series and products*, Academic Press, 7th. Ed. (2007).
13. M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*, 9th printing (1970).

The self-gravitating Fermi gas in Newtonian gravity and general relativity

Pierre-Henri Chavanis

Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France
E-mail: chavanis@irsamc.ups-tlse.fr

We review the history of the self-gravitating Fermi gas in Newtonian gravity and general relativity. We mention applications to white dwarfs, neutron stars and dark matter halos. We describe the nature of instabilities and phase transitions in the self-gravitating Fermi gas as energy (microcanonical ensemble) or temperature (canonical ensemble) is reduced. When $N < N_{\rm OV}$, where $N_{\rm OV}$ is the Oppenheimer-Volkoff critical particle number, the self-gravitating Fermi gas experiences a gravothermal catastrophe at E_c stopped by quantum mechanics (Pauli's exclusion principle). The equilibrium state has a core-halo structure made of a quantum core (degenerate fermion ball) surrounded by a classical isothermal halo. When $N > N_{\rm OV}$, a new turning point appears at an energy E_c'' below which the system experiences a gravitational collapse towards a black hole [P.H. Chavanis, G. Alberti, Phys. Lett. B **801**, 135155 (2020)]. When $N_{\rm OV} < N < N_*'$, the self-gravitating Fermi gas experiences a gravothermal catastrophe at E_c leading to a fermion ball, then a gravitational collapse at E_c'' leading to a black hole. When $N > N_*'$, the condensed branch disappears and the instability at E_c directly leads to a black hole. We discuss implications of these results for dark matter halos made of massive neutrinos.

Keywords: Fermi-Dirac statistics; White dwarfs; Neutron stars; Dark matter halos; Black holes

1. Introduction

The self-gravitating Fermi gas can have applications in different astrophysical systems ranging from white dwarfs and neutron stars to dark matter halos, where the fermions are electrons, neutrons and massive neutrinos respectively. The study of the self-gravitating Fermi gas is also of fundamental conceptual importance as it combines quantum mechanics and general relativity. Initially, fermionic models were developed at zero temperature ($T = 0$) but they have been later generalized at nonzero temperature, especially in the case of dark matter halos. In these Proceedings, we provide a brief history of the self-gravitating Fermi gas. A more detailed historical account of the statistical mechanics and thermodynamics of self-gravitating systems (classical and quantum) in Newtonian gravity and general relativity can be found in Refs.[1–4].

The statistical equilibrium state of a system of self-gravitating fermions can be determined from a maximum entropy principle. For systems with long-range interactions the mean field approximation becomes exact in an appropriate thermodynamic limit.[3,5] The most probable distribution of an isolated system of self-gravitating fermions at statistical equilibrium is obtained by maximizing the Fermi-Dirac

entropy S at fixed mass-energy $\mathcal{E} = Mc^2$ and particle number N:

$$\max \left\{ S \quad | \quad \mathcal{E} = Mc^2, \quad N \quad \text{fixed} \right\}. \tag{1}$$

The variational problem for the first variations reads

$$\delta S/k_B - \beta_\infty \delta \mathcal{E} + \alpha \delta N = 0, \tag{2}$$

where $\beta_\infty = 1/k_B T_\infty$ and $\alpha = \mu_\infty / k_B T_\infty$ are Lagrange multipliers associated with the conservation of mass-energy and particle number. Here, T_∞ and μ_∞ represent the temperature and the chemical potential measured by an observer at infinity. The maximization problem (1) is associated with the microcanonical ensemble. If the system is in contact with a thermal bath fixing the temperature T_∞ the statistical equilibrium state is obtained by minimizing the free energy $F = \mathcal{E} - T_\infty S$ at fixed particle number N:

$$\min \left\{ F = \mathcal{E} - T_\infty S \quad | \quad N \quad \text{fixed} \right\}. \tag{3}$$

This minimization problem is associated with the canonical ensemble. At $T = 0$ the equilibrium state is obtained by minimizing the mass-energy $\mathcal{E} = Mc^2$ at fixed particle number N. The equilibrium states in the microcanonical and canonical ensembles are the same. They are determined by the variational principle (2). However, their stability may differ in the two ensembles. This is the notion of ensemble inequivalence for systems with long-range interactions.[3,5] Microcanonical stability implies canonical stability but the converse is wrong.

The equilibrium state of a gas of self-gravitating fermions results from the balance between the repulsion due to the quantum pressure (Pauli's exclusion principle) and the gravitational attraction. The variational principle (2) yields all the equations that we need to determine the equilibrium state of the self-gravitating Fermi gas: (i) the Fermi-Dirac distribution function; (ii) the ideal equation of state of fermions; (iii) the Oppenheimer-Volkoff equations determining the condition of hydrostatic equilibrium in general relativity; (iv) the Tolman-Klein relations expressing how the local temperature $T(r)$ and the local chemical potential $\mu(r)$ are affected by the metric. We can solve these equations numerically and plot the caloric curve $T_\infty(\mathcal{E})$ relating the temperature to the energy. When $T > 0$ we need to enclose the system within a spherical box of radius R in order to prevent its evaporation and have equilibrium states with a finite mass. In the general case, the caloric curve depends on N and R. For convenience, instead of $T_\infty(\mathcal{E})$, we shall plot $\beta_\infty(-E)$ where $E = (M - Nm)c^2$ is the binding energy which reduces to the usual energy $E = K + W$ (kinetic + potential) in the nonrelativistic limit $c \to +\infty$. At $T = 0$, the system is self-confined (without the need of a box) and we shall plot the mass-radius relation $M(R)$ where R denotes here the radius where the density vanishes. The maximum entropy formalism for classical and quantum self-gravitating systems in Newtonian gravity and general relativity is reviewed in Refs.[3,4] where all the equations are derived and an exhaustive list of references is given.

2. Self-gravitating fermions at $T = 0$

The study of a self-gravitating gas of fermions started in the context of white dwarf stars when Fowler[6] first realized that these compact objects owe their stability to the quantum pressure of the degenerate electron gas. Indeed, the quantum pressure arising from the Pauli exclusion principle is able to counteract the gravitational attraction and explain the very high densities of white dwarf stars. Early studies were devoted to determining the ground state ($T = 0$) of the system. Nonrelativistic white dwarf stars are equivalent to a polytropic gas of index $n = 3/2$. Their density profile can be obtained by solving the Lane-Emden equation numerically. The density profile of white dwarf stars at $T = 0$ has a compact support, i.e., the density vanishes at a finite radius. The mass-radius relation of nonrelativistic white dwarf stars was first obtained by Stoner,[7] Milne[8] and Chandrasekhar.[9a] They showed that the radius of the star decreases as the mass increases according to the law $M = 91.9\,\hbar^6/(G^3m^8R^3)$ (see Fig. 1-a).[11] All the configurations of the series of equilibria are stable.

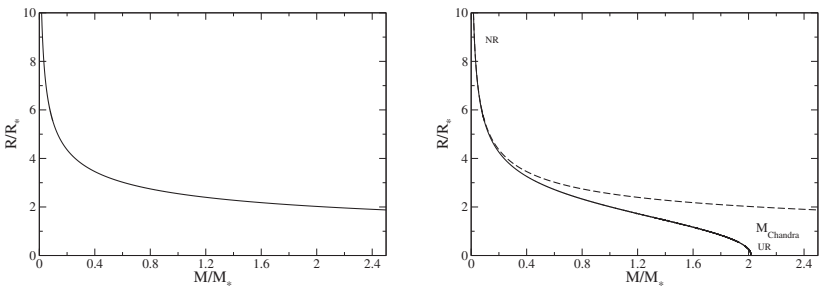


Fig. 1. Mass-radius relation of self-gravitating fermions at $T = 0$ in Newtonian gravity (the mass is normalized by $M_* = (3\pi/4)^{1/2}\hbar^{3/2}c^{3/2}/(G^{3/2}m^2)$ and the radius by $R_* = (3\pi/4)^{1/2}\hbar^{3/2}/(c^{1/2}G^{1/2}m^2)$). (a) Nonrelativistic white dwarf stars. (b) Special relativistic white dwarf stars.

The fact that relativistic effects become important in white dwarf stars whose mass is of the order of the solar mass was first reported by Frenkel[12] in a not well-known paper. However, he did not consider the implications of this result. Special relativistic effects in white dwarf stars were studied in detail by Anderson,[13] Stoner,[14] Chandrasekhar,[15] and Landau.[16] They found that no equilibrium state is possible above a maximum mass, now known as the Chandrasekhar limit.[b] These authors considered the equation of state of a relativistic Fermi gas at $T = 0$ and

[a]Stoner[7] developed an analytical approach based on a uniform density approximation for the star while Milne[8] and Chandrasekhar[9] developed a numerical approach based on the Lane-Emden theory of polytropes.[10]

[b]See the introduction of Ref.[17] for a short history of the discovery of the maximum mass of white dwarf stars.

used Newtonian gravity appropriate to white dwarf stars.[c] An ultrarelativistic Fermi gas at $T = 0$ is equivalent to a polytrope of index $n = 3$. Its density profile is obtained by solving the corresponding Lane-Emden equation and it has a compact support. For a polytrope $n = 3$, the mass-radius relation degenerates and indicates that different configurations with an arbitrary radius can exist at the same mass $M_{\rm Chandra} = 3.1\, M_P^3/m^2 = 1.5\, M_\odot$, where $M_P = (\hbar c/G)^{1/2}$ is the Planck mass and m is the proton mass. This argument immediately implies the existence of a critical mass.[15] In a more detailed study, Chandrasekhar[22] considered partially relativistic configurations and numerically obtained the complete mass-radius relation of white dwarf stars, valid for arbitrary densities, joining the nonrelativistic limit to the ultrarelativistic one (see Fig. 1-b).[d] As M approaches $M_{\rm Chandra}$ the radius of the star tends to zero while its density tends to infinity, leading to a Dirac peak. This study unambiguously shows the absence of equilibrium state above a maximum mass. Therefore, the quantum pressure arising from the Pauli exclusion principle cannot balance the gravitational attraction anymore when the star becomes sufficiently relativistic (or when its mass is too large). This is a striking effect of relativity combined with quantum mechanics and gravity. However, the result of Chandrasekhar[22] was severely criticized by Eddington[23] who argued that the absence of equilibrium states above a maximum mass leads to a *reductio ad absurdum* of the formula of relativistic degeneracy. Although the arguments of Eddington were entirely unfounded, his enormous prestige led to an early rejection of Chandrasekhar's work by many in the astronomical community. This pushed Chandrasekhar to abandon the subject, and delayed the discovery of the phenomenon of gravitational collapse and the concept of black hole.

In the following years, similar results were found by Oppenheimer and Volkoff[24] in connection to neutron stars. They solved the equations of general relativity with the relativistic equation of state for fermions at $T = 0$ and found that the mass-radius relation of neutron stars presents a turning point of mass (see Fig. 2).[e] As a result, no equilibrium state exists above a maximum mass $M_{\rm OV} = 0.384\, M_P^3/m^2 = 0.710\, M_\odot$, where m is the neutron mass, called the Oppenheimer-Volkoff limit (note that the density profile with the maximum mass $M_{\rm OV}$ is *not* singular contrary to the Newtonian density profile at the maximum mass $M_{\rm Chandra}$). They argued that, above that mass, the star undergoes gravitational collapse. This problem was specifically studied by Oppenheimer and Snyder[25] who obtained an analytical solution of the Einstein equations describing the collapse of a pressureless gas up to its

[c] In principle, general relativistic effects become important close to the Chandrasekhar maximum mass[18,19] but other phenomena like Coulomb corrections to the electron pressure and the formation of neutrons by inverse beta decay destabilize the star before general relativistic effects come into play.[20,21]

[d] A similar mass-radius relation was obtained earlier by Stoner[14] from an approximate analytical model based on uniform density stars.

[e] It was shown later that the mass-radius relation of neutron stars forms a spiral and that a mode of stability is lost at each turning point of mass.[21]

Schwarzschild radius. Strangely enough, these important results did not receive much attention until the 1960's. At that epoch, detailed models of compact objects with more realistic equations of state taking into account the repulsive effect of nuclear forces and connecting white dwarfs to neutron stars were constructed and the fundamental discoveries of Chandrasekhar, Landau and Oppenheimer and Volkoff were confirmed (unfortunately, the early contributions of Anderson and Stoner were forgotten).[21] Pulsars were discovered by Hewish *et al.*[26] in 1968. The same year, Gold[27,28] proposed that pulsars are rotating neutron stars, and this is generally accepted today. It is also at that moment that the name "black hole" was used by Wheeler[29] to designate the object resulting from gravitational collapse, and became popular.

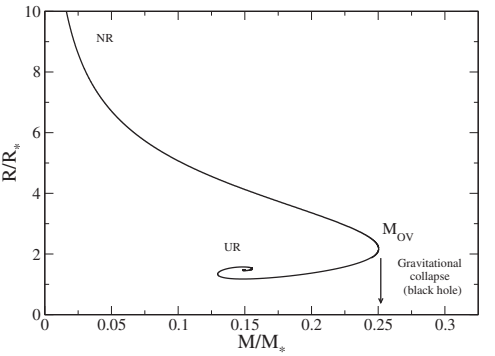


Fig. 2. Mass-radius relation of self-gravitating fermions at $T = 0$ in general relativity (neutron stars).

3. Nonrelativistic classical particles at $T > 0$ and self-gravitating radiation

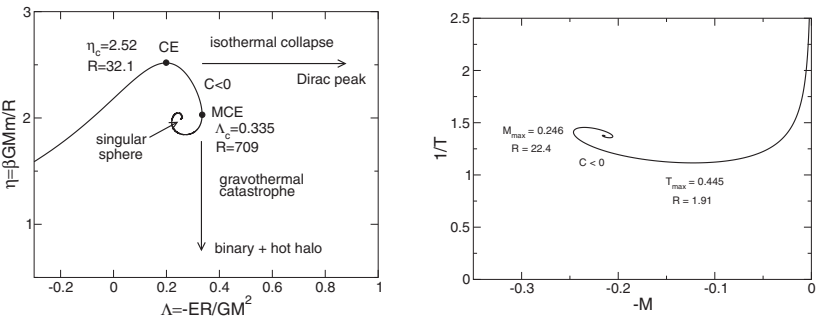
The thermodynamics of self-gravitating systems is a fascinating subject.[30–32] Its study started with the pioneering work of Antonov[33] who considered an isolated system of nonrelativistic classical particles in gravitational interaction. He used a microcanonical ensemble description in which the mass and the energy are conserved. This situation applies approximately to stellar systems like globular clusters. Their equilibrium state (most probable state) can be obtained by maximizing the Boltzmann entropy S at fixed mass M and energy E.[34] This leads to the mean field Boltzmann distribution which is self-consistently coupled to the Poisson equation. The Boltzmann-Poisson equation was previously introduced and studied in the context of isothermal stars.[10,11] It can be reduced to the Emden equation that has to be solved numerically. Antonov[33] observed that no maximum entropy state exists in an infinite domain (the solution of the Emden equation has an infinite mass), so he proposed to confine the particles within a spherical box of radius R.

This artifice prevents the evaporation of the system and leads to a well-defined mathematical problem. By computing the second variations of entropy, Antonov[33] showed that equilibrium states with a density contrast $\mathcal{R} = \rho_0/\rho(R) < 709$, where ρ_0 is the central density and $\rho(R)$ the density on the edge of the box, are thermodynamically stable (entropy maxima) while equilibrium states with a density contrast $\mathcal{R} > 709$ are thermodynamically unstable (saddle points of entropy). Lynden-Bell and Wood[35] rediscussed the results of Antonov[33] in more physical terms. They plotted the series of equilibria $E(\mathcal{R})$ and showed that it displays damped oscillations. As a result, there is no equilibrium state with an energy $E < E_c = -0.335\,GM^2/R$, where E_c corresponds to the first turning point of energy (with a density contrast $\mathcal{R}_c = 709$). Invoking the Poincaré turning point criterion,[36] they concluded that the series of equilibria becomes unstable at the minimum energy E_c. In this manner, they recovered the critical density contrast $\mathcal{R}_c = 709$ found by Antonov.[33] They also interpreted the Antonov instability in terms of a "gravothermal catastrophe" caused by the negative specific heat of the system in its densest parts. When this instability occurs, the system undergoes core collapse. This ultimately leads to a binary star surrounded by a hot halo. Lynden-Bell and Wood[35] considered other statistical ensembles, notably the canonical ensemble in which the temperature and the mass are fixed. In that case, the equilibrium state is obtained by minimizing the Boltzmann free energy $F = E - TS$ at fixed mass M. The series of equilibria $T(\mathcal{R})$ displays damped oscillations. No equilibrium state exists with a temperature $T < T_c = 0.397\,GMm/(k_B R)$, where T_c corresponds to the first turning point of temperature (with a density contrast $\mathcal{R}'_c = 32.1$).[f] Using the Poincaré turning point criterion,[36] they concluded that equilibrium states with a density contrast $\mathcal{R} < 32.1$ are thermodynamically stable (free energy minima) while equilibrium states with a density contrast $\mathcal{R} > 32.1$ are thermodynamically unstable (saddle points of free energy). Below T_c the system undergoes an "isothermal collapse" leading to a Dirac peak containing all the particles. Since the stability limits in the microcanonical and canonical ensembles differ, Lynden-Bell and Wood[35] encountered for the first time in statistical mechanics a situation of ensemble inequivalence. This is a peculiarity of systems with long-range interactions.[5] In the present context, it is related to the fact that negative specific heats are allowed in the microcanonical ensemble while they are forbidden in the canonical ensemble.[35] Similar results were obtained independently by Thirring.[37] Katz[38] plotted the caloric curve $\beta(E)$ of isothermal self-gravitating spheres and exhibited its spiral behavior (see Fig. 3-a).[g] He also extended the Poincaré theory on linear series of equilibria[36] to the case where there are several turning points and developed a general method to determine the thermodynamical stability of the equilibrium states from the topology of the caloric

[f] These results were first found by Emden.[10]

[g] This spiral behavior was previously observed for self-gravitating isothermal stars in other representations.[11,39–44] In the present case, it is associated with the damped oscillations of energy and temperature as a function of the density contrast.

curve $\beta(E)$. A change of stability can only occur at a turning point of energy in the microcanonical ensemble or at a turning point of temperature in the canonical ensemble. A mode of stability is lost if the curve $\beta(-E)$ turns clockwise and gained if it turns anticlockwise. In this manner, one can determine the thermodynamical stability of the system by simply plotting the caloric curve (series of equilibria). The seminal works of Antonov,[33] Lynden-Bell and Wood,[35] and Katz[38] were followed by many other studies (see, e.g., Refs.[45–52]).



(a) (b)

Fig. 3. (a) Caloric curve of nonrelativistic classical self-gravitating particles (cold spiral).[32] We have plotted the normalized inverse temperature $\eta = \beta GMm/R$ as a function of minus the normalized binding energy $\Lambda = -ER/GM^2$. (b) Caloric curve of the self-gravitating radiation in general relativity (hot spiral).[54] We have plotted the normalized inverse temperature $\hbar^{3/4}c^{7/4}/(k_B T_\infty G^{1/4}R^{1/2})$ as a function of minus the normalized energy $-GM/Rc^2$.

The statistical mechanics of the self-gravitating black-body radiation (photon star) confined within a cavity in general relativity was investigated by Sorkin *et al.*[53] and, more recently, by Chavanis.[54] They showed that the caloric curve $\beta_\infty(\mathcal{E})$ forms a spiral (see Fig. 3-b). There is no equilibrium state above a maximum mass-energy $M_{\max}c^2 = 0.246\, Rc^4/G$ (corresponding to a density contrast 22.4) or above a maximum temperature $(T_\infty)_{\max} = 0.445\, \hbar^{3/4}c^{7/4}/(k_B G^{1/4}R^{1/2})$ (corresponding to a density contrast 1.91). In that case, the system is expected to collapse towards a black hole. We note that the "hot spiral" (see Fig. 3-b) of the self-gravitating radiation in general relativity (ultrarelativistic limit) is inverted with respect to the "cold spiral" of the nonrelativistic classical self-gravitating gas (see Fig. 3-a).

4. Classical particles at $T > 0$ in general relativity

The statistical mechanis of classical particles in general relativity has been considered by Roupas[55] and, independently, by Alberti and Chavanis.[1] The caloric curve depends on one parameter, the particle number N (more precisely N/R). Generically, the caloric curve $\beta_\infty(E)$ has the form of a double spiral (see Fig. 4) which combines the aspects of the "cold spiral" corresponding to a nonrelativistic gas and the aspects of the "hot spiral" corresponding to an ultrarelativistic gas discussed

in Sec. 3.[h] There is no equilibrium state below a minimum energy (resp. minimum temperature) and above a maximum energy (resp. maximum temperature) in the microcanonical (resp. canonical) ensemble. When the number of particles N increases, the two spirals approach each other, merge, form a loop, and finally disappear (by reducing to a point) at $N_{\max} = 0.1764\, Rc^2/Gm$. For $N > N_{\max}$, there is no equilibrium state whatever the value of mass-energy and temperature.

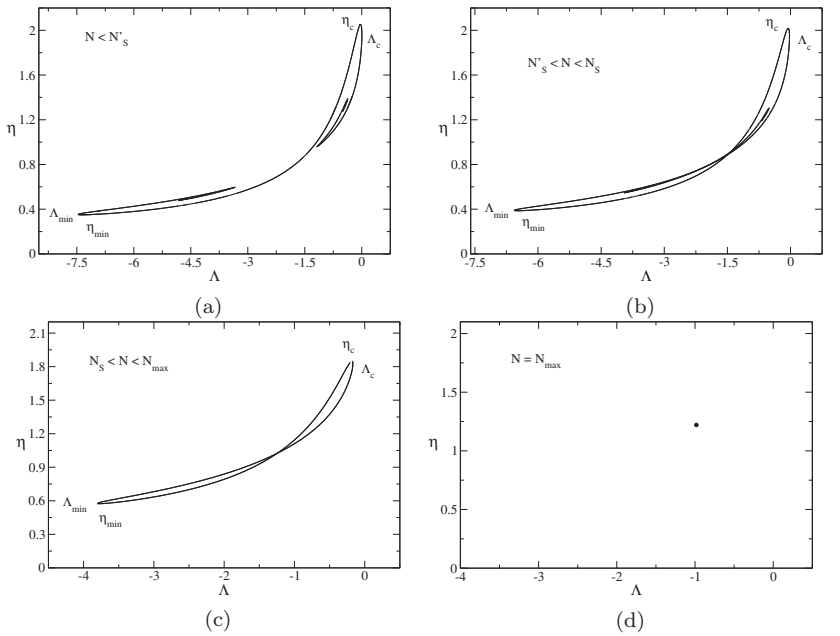


Fig. 4. Caloric curve of classical particles in general relativity. We have plotted the normalized inverse temperature $\eta = \beta_\infty GNm^2/R$ as a function of minus the normalized binding energy $\Lambda = -ER/GN^2m^2$ for different values of the normalized particle number $\nu = GNm/Rc^2$. (a) Double spiral (b) Merging (c) Loop (d) Point.[1]

5. Nonrelativistic fermions at $T > 0$

The statistical mechanics of self-gravitating fermions at nonzero temperature was first studied by Hertel and Thirring[56,57] as a simple (nonrelativistic) model of neutron stars. In that case, the density profile decreases with the distance as r^{-2} and we need to confine the system within a box in order to have an equilibrium state with a finite mass. Hertel and Thirring[56] rigorously proved that the mean field approximation (which amounts to neglecting correlations among the particles) and the Thomas-Fermi approximation (which amounts to neglecting the quantum potential)

[h]The hot spiral of ultrarelativistic classical particles in general relativity is similar, but not identical, to the hot spiral of the self-gravitating radiation.[1,3,54]

become exact in a proper thermodynamic limit where $N \to +\infty$. This leads to the Fermi-Dirac-Poisson equations, also known as the temperature-dependent Thomas-Fermi equations. Hertel and Thirring[57] solved these equations numerically and plotted the caloric curve. The caloric curve depends on one parameter, the box radius R (more precisely NR^3). They found that, for sufficiently large systems ($R > R_{\rm CCP} = 12.8\,\hbar^2/(N^{1/3}Gm^3)$), a negative specific heat region occurs in the microcanonical ensemble (see Fig. 5). Since negative specific heats are forbidden in the canonical ensemble, this implies that the statistical ensembles are inequivalent. The region of negative specific heats which is allowed in the microcanonical ensemble is replaced by a first order phase transition in the canonical ensemble. This phase transition is expected to take place at a transition temperature T_t connecting the gaseous phase to the condensed phase through a horizontal Maxwell plateau in the caloric curve $T(E)$. This is accompanied by a discontinuity of energy. There is also a lower critical temperature T_c (spinodal point) at which the metastable gaseous phase disappears and the system collapses (zeroth order phase transition). This collapse is associated with the isothermal collapse of classical self-gravitating systems.[10,49] However, for self-gravitating fermions, the collapse stops when quantum degeneracy comes into play. In that case, the system achieves a "core-halo" configuration made of a quantum core (fermion ball) containing almost all the mass surrounded by a tenuous isothermal atmosphere. There is also a higher critical temperature T_* (spinodal point) at which the metastable condensed phase disappears and the system explodes. These results were exported by Bilic and Viollier[58] to the context of dark matter.

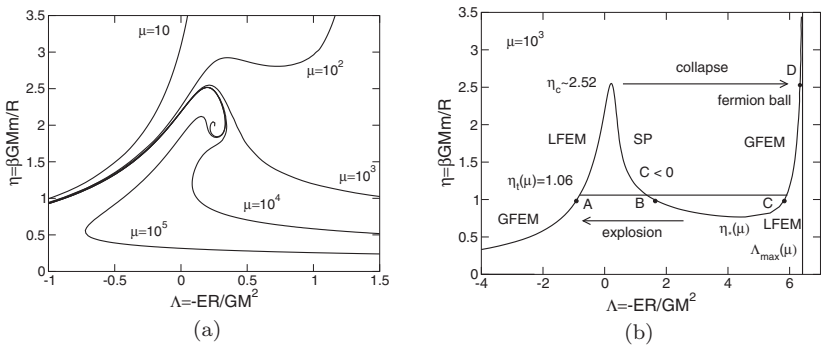


Fig. 5. Caloric curve of nonrelativistic self-gravitating fermions.[32] We have plotted the normalized inverse temperature $\eta = \beta GMm/R$ as a function of minus the normalized energy $\Lambda = -ER/GM^2$ for different values of the normalized radius, or degeneracy parameter $\mu = \eta_0\sqrt{512\pi^4 G^3 MR^3}$, where $\eta_0 = 2m^4/h^3$ is the maximum possible value of the distribution function fixed by the Pauli exclusion principle. (a) Dependence of the series of equilibria on the size of the system. (b) For small systems, the caloric curve has an N-shape structure.

An exhaustive study of phase transitions in the self-gravitating Fermi gas was made by Chavanis[32,59] in both canonical and microcanonical ensembles.

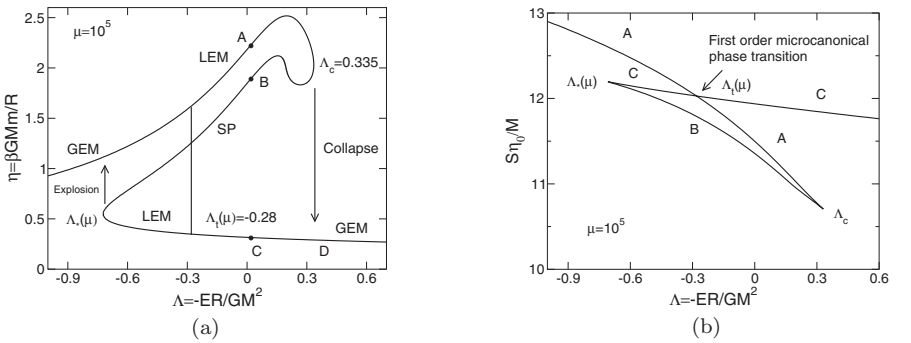


Fig. 6. Caloric curve of nonrelativistic self-gravitating fermions.[32] (a) For large systems, the caloric curve has a Z-shape structure resembling a dinosaur's neck. (b) Illustration of the microcanonical first order phase transition on the $S(E)$ curve.

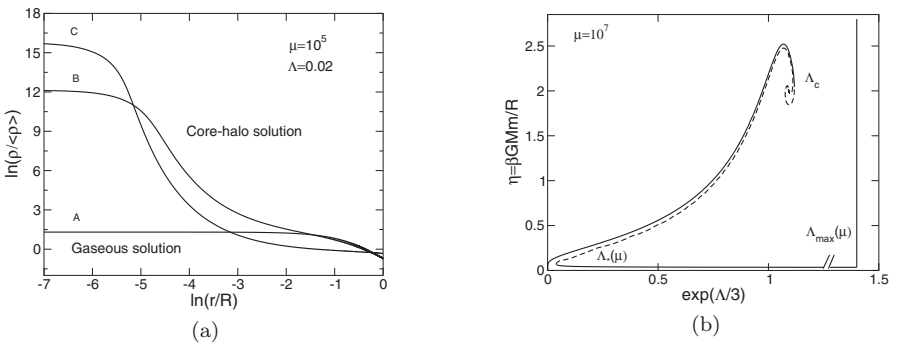


Fig. 7. (a) Density profiles corresponding to the gaseous and core-halo solutions.[32] (b) The classical limit $\mu \to +\infty$ (very large systems).[32] According to the Poincaré-Katz criterion,[36,38] the equilibrium states are unstable in the microcanonical (resp. canonical) ensemble between the first and the last turning points of energy (resp. temperature).

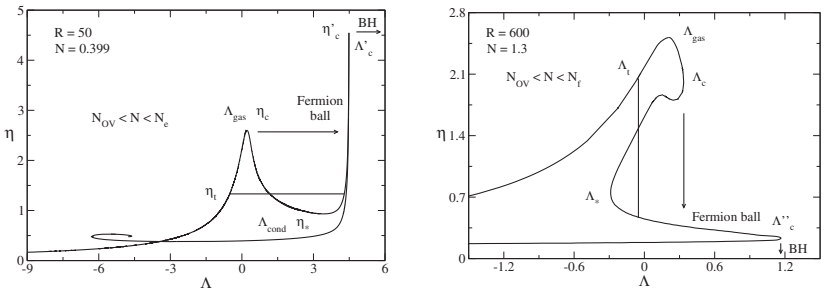
He confirmed the phase transition in the canonical ensemble previously found by Hertel and Thirring[57] and evidenced, for sufficiently large systems ($R > R_{\rm MCP} = 130\,\hbar^2/(N^{1/3}Gm^3)$), a first order phase transition in the microcanonical ensemble (see Fig. 6). This phase transition is expected to take place at a transition energy E_t connecting the gaseous phase to the condensed phase through a vertical Maxwell plateau in the caloric curve $T(E)$. This is accompanied by a discontinuity of temperature. There is also a lower critical energy E_c (spinodal point) at which the metastable gaseous phase disappears and the system collapses (zeroth order phase transition). This collapse is associated with the gravothermal catastrophe of classical self-gravitating systems.[33,35] However, for self-gravitating fermions, the collapse stops when quantum degeneracy comes into play. In that case, the system achieves a core-halo configuration[60] made of a quantum core (fermion ball) containing a moderate fraction of the total mass surrounded by a massive isothermal atmosphere

(see Fig. 7-a). There is also a higher critical energy E_* (spinodal point) at which the metastable condensed phase disappears and the system explodes. As a result, there exist two distinct critical points in the self-gravitating Fermi gas, one in each ensemble, above which phase transitions occur. For small systems ($R < R_{\rm CCP}$), there is no phase transition, for intermediate size systems ($R_{\rm CCP} < R < R_{\rm MCP}$) a phase transition takes place in the canonical ensemble but not in the microcanoncal ensemble, and for large systems ($R > R_{\rm MCP}$) a phase transition takes place in both ensembles. When $R \to +\infty$, the series of equilibria rotates several times before unwinding (see Fig. 7-b) and we recover the classical spiral from Fig. 3-a. Chavanis[32] determined the phase diagram of the nonrelativistic self-gravitating Fermi gas. He also argued that the lifetime of metastable states is extremely long, scaling as e^N where N is the number of particles, so that the first order phase transitions do not take place in practice.[61] Only zeroth order phase transitions associated with the spinodal points are physically meaningful.

6. General relativistic fermions at $T > 0$

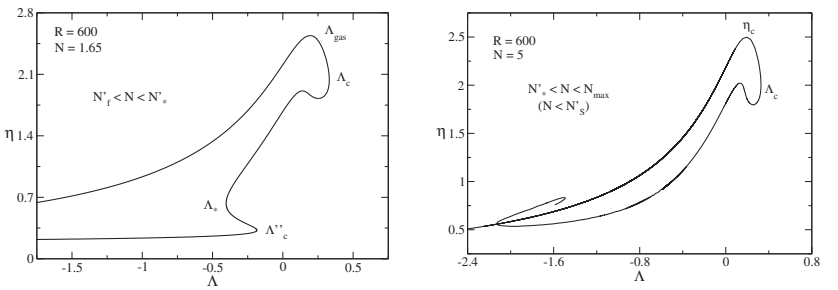
The statistical mechanics of self-gravitating fermions at nonzero temperature in general relativity was first considered by Bilic and Viollier.[62] As before, the system has to be confined within a spherical box of radius R in order to prevent its evaporation. The caloric curve depends on two parameters, the box radius R and the particle number N. Bilic and Viollier[62] studied the case where R is relatively small and $N < N_{\rm OV}$. In that case, the results are qualitatively similar to those obtained for nonrelativistic fermions (see Sec. 5). There is a first order phase transition in the canonical ensemble which replaces the region of negative specific heat present in the microcanonical ensemble. An equilibrium state, which is either "gaseous" (corresponding to the classical isothermal sphere) or "condensed" (made of a fermion ball surrounded by a classical isothermal envelope), exists for any value of temperature T_∞ and binding energy E. General relativistic effects only introduce a small correction to the Newtonian results.

A more general study was performed by Alberti, Chavanis and Roupas[2,63,64] who considered arbitrary values of R and N. When $N > N_{\rm OV}$, they identified the existence of a new turning point of temperature in the canonical ensemble and a new turning point of binding energy in the microcanonical ensemble below which the system collapses and forms a black hole of mass $M_{\rm OV}$ (see Fig. 8). This is the finite temperature generalization of the result originally found by Oppenheimer and Volkoff[24] at $T = 0$. These results lead to the following scenario (we restrict ourselves to the microcanonical ensemble which is the most relevant). At high energies, the system is in the gaseous phase. Below a critical energy E_c it becomes thermodynamically unstable and experiences a gravothermal catastrophe. However, core collapse stops when quantum mechanics (Pauli's exclusion principle) comes into play. This leads to the formation of a fermion ball surrounded by a hot halo. Below E_c'', the condensed phase becomes thermodynamically and dynamically unstable



(a) (b)

Fig. 8. Caloric curve of self-gravitating fermions in general relativity when $N > N_{\rm OV}$.[2] We have plotted the normalized inverse temperature $\eta = \beta_\infty GNm^2/R$ as a function of minus the normalized binding energy $\Lambda = -ER/GN^2m^2$ for different values of the normalized radius $R/R_{\rm OV}$ and normalized particle number $N/N_{\rm OV}$. (a) Small systems: As T decreases the system experiences an isothermal collapse at T_c leading to a fermion ball, then a gravitational collapse at T_c' leading to a black hole. (b) Large systems: As E decreases the system experiences a gravothermal catastrophe at E_c leading to a fermion ball surrounded by a hot halo, then a gravitational collapse at E_c'' leading to a black hole.



(a) (b)

Fig. 9. Caloric curve of self-gravitating fermions in general relativity.[2] When $N > N_*'$ the condensed branch disappears completely so that only the collapse at E_c towards a black hole is possible.

(in a general relativistic sense) and collapses towards a black hole. Alberti and Chavanis[2] also evidenced a critical particle number N_*' above which the condensed phase completely disappears (see Fig. 9). In that case, there is no possibility to form a fermion ball. The gravothermal catastrophe at E_c directly leads to a black hole. In conclusion, for $N < N_{\rm OV}$ the system forms a fermion ball; for $N_{\rm OV} < N < N_*'$ the system generically forms a fermion ball, then (possibly) a black hole; for $N > N_*'$ the system directly forms a black hole. Alberti and Chavanis[2] emphasized the core-halo structure of the equilibrium states in the microcanonical ensemble and mentioned the relation to red-giants (leading to white dwarfs) and supernovae (leading to neutron stars and black holes) suggested in Refs.[64–67]. They also provided the complete phase diagram of the general relativistic Fermi gas.

7. Truncated models

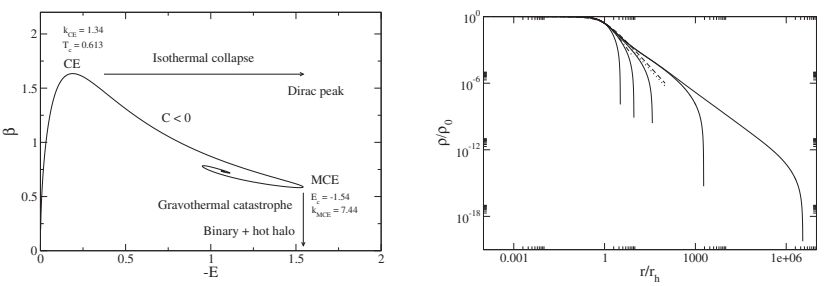
The study of phase transitions in the self-gravitating Fermi gas can be extended to the fermionic King model described by the distribution function

$$f = A\frac{e^{-\beta(\epsilon-\epsilon_m)}-1}{1+\frac{A}{\eta_0}e^{-\beta(\epsilon-\epsilon_m)}} \qquad (\epsilon \leq \epsilon_m). \tag{4}$$

The fermionic King model was introduced heuristically by Ruffini and Stella[68] as a generalization of the classical King model[69]

$$f = A\left[e^{-\beta(\epsilon-\epsilon_m)}-1\right] \qquad (\epsilon \leq \epsilon_m). \tag{5}$$

The fermionic King model was also introduced independently by Chavanis[70] who derived it from a kinetic theory based on the fermionic Landau equation. The fermionic King model is more realistic than the usual fermionic model because it avoids the need of an artificial box to confine the system. The nonrelativistic fermionic King model was studied by Chavanis *et al.*[71,72] who showed that the density profiles generically have a core-halo structure with a quantum core (fermion ball) and a tidally truncated isothermal halo leading to flat rotation curves.[i] They also studied the caloric curves, the thermodynamical stability of the equilibrium states, and the phase transitions between gaseous and condensed states (see Figs. 10 and 11). They showed that the phenomenology of phase transitions in the fermionic King model is the same as in the case of box-confined systems obtained in Refs.[32,59]. The results of Chavanis *et al.*[72] have been generalized by Argüelles *et al.*[73] to the fermionic King model in general relativity. They also found that the phenomenology of phase transitions in the general relativistic fermionic King model is the same as in the case of box-confined systems obtained in Refs.[2,63,64].



(a) (b)

Fig. 10. (a) Caloric curve of the nonrelativistic classical King model.[71] (b) Density profiles along the series of equilibria.[71] The marginal (critical) King profile at the point of gravothermal instability is relatively close to the Burkert[74] profile (dashed line) which provides a good fit of the density profile of dark matter halos.

[i]The name "fermionic King model" was introduced in Ref.[72].

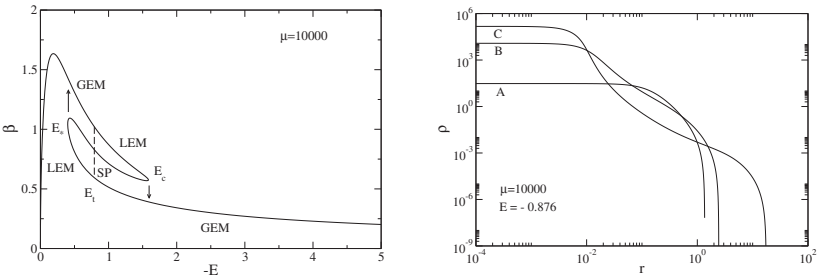


Fig. 11. (a) Caloric curve of the nonrelativistic fermionic King model for large systems.[72] (b) Density profiles corresponding to the stable gaseous phase (A), the stable condensed phase (C), and the unstable intermediate solution (B).[72]

8. Application to dark matter halos

In addition to white dwarfs and neutron stars,[75] the self-gravitating Fermi gas model has been applied in the context of dark matter halos made of massive neutrinos. The suggestion that dark matter is made of massive neutrinos was originally proposed by Markov[76] and Cowsik and McClelland.[77,78] This suggestion has been developed by numerous authors (see the introduction of Ref.[3] for an exhaustive list of references). The first models decribed dark matter halos at $T = 0$ using the equation of state of a completely degenerate fermion gas either in the nonrelativistic limit or in general relativity. Subsequent models considered dark matter halos at nonzero temperature showing that they have a core-halo structure consisting of a dense core (fermion ball) solving the core-cusp problem of classical cold dark matter surrounded by a dilute isothermal atmosphere leading to flat rotation curves. Most models were based on the ordinary Fermi-Dirac distribution in Newtonian gravity or general relativity. Some models were based on the more realistic fermionic King model (describing tidally truncated fermionic dark matter halos). The self-gravitating Fermi gas was also studied in relation to the violent relaxation of collisionless self-gravitating systems described by the Lynden-Bell[79] distribution which is formally similar to the Fermi-Dirac distribution. As argued in Refs.[71,72], the theory of violent relaxation may justify how a collisionless gas of self-gravitating fermions, such as a dark matter halo, can reach a statistical equilibrium state described by the Fermi-Dirac distribution on a timescale smaller than the age of the universe.[j]

The detailed study of the motion of S-stars near the Galactic center has revealed the presence of a very massive central object, Sagittarius A* (Sgr A*). This central object is usually associated with a supermassive black hole (SMBH) of mass

[j]The relaxation time due to close gravitational encounters exceeds the age of the universe by many orders of magnitude. The collisional relaxation time may be shorter if the fermions are self-interacting.[80]

$M = 4.2 \times 10^6\, M_\odot$ and Schwarzschild radius $R_S = 4.02 \times 10^{-7}\,\mathrm{pc}$. Whatever the object may be, its radius must be smaller than $R_\mathrm{P} = 6 \times 10^{-4}\,\mathrm{pc}$ ($R_\mathrm{P} = 1492\, R_S$), the S2 star pericenter. Similar objects are expected to reside at the center of most spiral and elliptical galaxies, in active galactic nuclei (AGN). Although it is commonly believed that these objects are SMBHs, this is not yet established on a firm observational basis in all cases. Some authors have proposed that such objects could be fermion balls or boson stars that could mimic a SMBH. Let us consider this possibility in the framework of the fermionic dark matter model. More precisely, let us investigate if a fermion ball can mimic a SMBH at the center of the Galaxy.

Bilic *et al.*[81] developed a general relativistic model of fermionic dark matter halos at nonzero temperature with a fermion mass $m = 15\,\mathrm{keV/c^2}$ that describes both the center and the halo of the Milky Way in a unified manner. The density profile has a core-halo structure made of a quantum core (fermion ball) surrounded by a classical isothermal atmosphere. The core and the halo are separated by an extended plateau. By using the usual Fermi-Dirac distribution and choosing parameters so as to fit observational data at large distances, they found a fermion ball of mass $M_c = 2.27 \times 10^6\, M_\odot$ and radius $R_c = 18\,\mathrm{mpc}$ and argued that this fermion ball can mimic a SMBH like Sgr A*. Unfortunately, its radius is larger by a factor 100 than the bound $R_P = 6 \times 10^{-4}\,\mathrm{pc}$ set by later observations. This is why Bilic and coworkers abandoned this fermion ball scenario (R. Viollier, private communication). The same problem was encountered later by Ruffini *et al.*[82] who developed a similar model with a fermion mass $m \sim 10\,\mathrm{keV/c^2}$.

More recently, Argüelles *et al.*[83] considered the general relativistic fermionic King model accounting for a tidal confinement. They applied this model to the Milky Way and determined the parameters by fitting the core-halo profile to the observations. For a fermion mass $m = 48\,\mathrm{keV/c^2}$ they obtained a fermion ball of mass $M_c = 4.2 \times 10^6\, M_\odot$ and radius $R_c = R_P = 6 \times 10^{-4}\,\mathrm{pc}$ which, this time, is consistent with the observational constraints of Sgr A*. Therefore, in order to obtain accurate results, it is important to use the fermionic King model[72,83] instead of the usual fermionic model.[32,81,82] Argüelles *et al.*[83] managed to fit the entire density profile and the entire rotation curve of the Milky Way with the fermionic King distribution and argued that a fermion ball can mimic the effect of a SMBH like Sgr A*. This scenario is very attractive because it can explain the whole structure of the galaxy, the supermassive central object and the isothermal halo, by a single distribution (the fermionic King model[68,70]).

Developing the theory of phase transitions in the self-gravitating Fermi gas, Chavanis *et al.*[72] argued that the core-halo solution considered by Bilic *et al.*,[81] Ruffini *et al.*,[82] and Argüelles *et al.*[83] with a small fermion ball mimicking a SMBH surrounded by a classical isothermal atmosphere, which was claimed to reproduce the structure of the Milky Way, is thermodynamically unstable because it lies in the intermediate branch of the caloric curve between the first and the last turning points of energy (see Fig. 7-b). As a result, it is a saddle point of entropy, not an entropy maximum. Therefore, Chavanis *et al.*[72] concluded that this type of solution

is not likely to result from a natural evolution and, consequently, they questioned the possibility that a fermion ball could mimic a central SMBH.

Following this study,[72] Argüelles *et al.*[73] computed the caloric curve of the fermionic King model in general relativity (see Fig. 12). They showed that the core-halo solution of Ref.[83] which provides a good agreement with the structure of the Milky Way lies just *after* the turning point (b) of energy (see the inset of Fig. 12), so that it is thermodynamically stable in the microcanonical ensemble which is the correct ensemble to consider.[k] This is a very interesting result because it shows that this core-halo structure *may* arise from a natural evolution in the sense of Lynden-Bell. This gives further support to the scenario according to which a fermion ball could mimic a SMBH at the center of the galaxies.

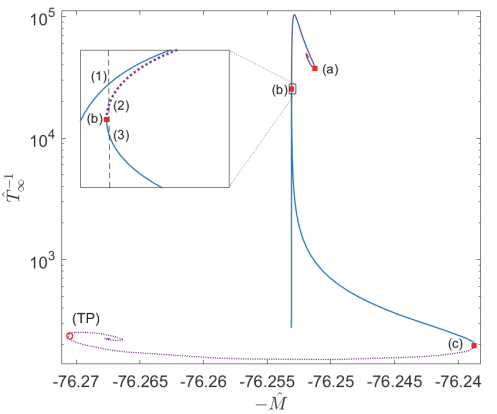


Fig. 12. Caloric curve of the general relativistic fermionic King model (from Ref.[73]).

However, it does not prove that this core-halo structure with a very high central density will necessarily arise from a natural evolution. The reason is that violent relaxation is in general incomplete.[79] The fluctuations of the gravitational potential that are the engine of the collisionless relaxation can die out before the system has reached statistical equilibrium. Therefore, it is not clear if violent relaxation can produce this type of structures.[l] In order to vindicate this scenario, the next step would be to perform direct numerical simulations of collisionless fermionic matter to

[k]Chavanis *et al.*[72] did not focus on the stable branch of condensed states located after point (b) because they argued that these solutions are not astrophysically relevant. Indeed, by considering particular solutions of the condensed branch, they observed that these solutions have a too extended envelope that is not consistent with the structure of dark matter halos (see solution C in Fig. 11-b and Figs. 38-45 of Ref.[72]). Although this claim is correct for most of the solutions on the condensed branch, it turns out that the solutions located just after point (b) *are* astrophysically relevant and correspond to the core-halo solutions studied by Argüelles *et al.*[83]

[l]It may be easier to form core-halo configurations with a very high central density if the fermions are self-interacting and if the Fermi-Dirac equilibrium state results from a collisional relaxation of nongravitational origin instead of a collisionless relaxation as suggested in Ref.[80].

see if they spontaneously generate fermion balls with the characteristics of SMBHs. A purely gaseous solution without quantum core, which is also a maximum entropy state, may be easier to reach through a violent relaxation process and is consistent with the observations. However, it does not account for a massive central object at the center of the galaxies. In that case, we either have to introduce a primordial SMBH "by hand" or advocate a scenario of gravitational collapse such as the one described below.[80]

For a fermion mass $m = 48\,\mathrm{keV/c^2}$, the mass $M_h = 10^{11}\,M_\odot$ of the Milky Way is larger than the OV mass $M_{\rm OV} = 2.71 \times 10^8\,M_\odot$, so we have to take into account general relativity in the caloric curve. As first shown by Alberti and Chavanis[2,64] for box-confined fermionic systems, and recovered by Argüelles *et al.*[73] for tidally truncated models, relativistic effects create a new turning point of energy in the caloric curve at which the condensed branch terminates (see Figs. 8, 9 and point (c) in Fig. 12). Below E_c'' the system collapses towards a black hole.

Suppose that violent relaxation selects the gaseous solution. On a secular timescale, because of collisions, the system follows the upper branch of the series of equilibria (gaseous states) up to the point of minimum energy E_c. At that point, it becomes thermodynamically unstable and undergoes a gravothermal catastrophe. However, core collapse is stopped by quantum mechanics, leading to the formation of a fermion ball. Then, if the energy keeps decreasing, the system follows the lower branch of the series of equilibria up to the point of minimum energy E_c'' where it becomes thermodynamically and dynamically unstable (in a general relativistic sense) and collapses towards a SMBH of mass $M_{\rm OV}$. A similar outcome arises if violent relaxation selects the core-halo solution where the fermion ball mimics a SMBH. On a secular timescale, the system follows the lower branch of the series of equilibria up to the point of minimum energy E_c'' at which it collapses towards a SMBH. In the two cases, the ultimate fate of the system is to form a SMBH of mass $M_{\rm OV}$ surrounded by an envelope. However, the formation of a SMBH may take time (more than the age of the universe) so that the two objects (fermion ball or SMBH) are possible in practice.

For a fermion mass $m = 48\,\mathrm{keV/c^2}$, the OV mass $M_{\rm OV} = 2.71 \times 10^8\,M_\odot$ is too large to account for the mass of a SMBH like Sgr A* at the center of the Milky Way. This suggests that the object at the center of the Galaxy (Sgr A*) is a fermion ball instead of a SMBH as argued by Argüelles *et al.*[73] However, for very large halos ($N > N_*'$), it is shown by Alberti and Chavanis[2] that the condensed branch disappears (see panel (b) of Fig. 9). In that case, there is no stable solution with a fermion ball and the system necessarily collapses towards a SMBH. Therefore, medium size galaxies like the Milky Way should harbor a fermion ball of mass $M_c = 4.2 \times 10^6\,M_\odot$ while very large galaxies like ellipticals should harbor a SMBH of mass $M_{\rm OV} = 2.71 \times 10^8\,M_\odot$ that could even grow by accretion. This scenario[80] could account for the mass of SMBHs in AGNs like the one recently photographed in M87 ($M_h \sim 10^{13}\,M_\odot$ and $M_{\rm BH} \sim 10^{10}\,M_\odot$).

On the other hand, for a much larger fermion mass $m = 386\,\mathrm{keV/c^2}$, the OV mass $M_{\rm OV} = 4.2\times10^6\,M_\odot$ is comparable to the mass of Sgr A*. Furthermore, when applied to the Milky Way, the caloric curve corresponding to $m = 386\,\mathrm{keV/c^2}$ is similar to the one reported in panel (b) of Fig. 9 so there is no possibility to form a fermion ball. In that case, the Milky Way could have undergone a gravitational collapse at E_c leading to a SMBH of mass $M_{\rm OV} = 4.2\times10^6\,M_\odot$.[80] In this process, the halo surrounding the SMBH is left undisturbed and could correspond to a marginal classical King profile, which is known[71,72] to give a good agreement with the empirical Burkert[74] profile of observed dark matter halos (see Fig. 10-b).

Different scenarios are possible depending on the value of the fermion mass m. Argüelles *et al.*[73,83] determined the mass of the fermionic dark matter particle in such a way that the fermion ball that would be at the center of a large fermionic dark matter halo, like the one that surrounds the Milky Way, mimics the effect of a SMBH of mass $M_c = 4.2\times10^6\,M_\odot$ and radius $R_c = 6\times10^{-4}\,\mathrm{pc}$ like Sgr A*. This leads to a fermion mass $m = 48\,\mathrm{keV/c^2}$.[m] Alternatively, Chavanis[80,86] determined the mass of the fermionic dark matter particle by arguing that the smallest halos observed in the universe (dSphs like Fornax) with a typical mass $M\sim10^8\,M_\odot$ and a typical radius $R\sim1\,\mathrm{kpc}$ represent the ground state of the self-gravitating Fermi gas at $T = 0$. This yields a much smaller fermion mass $m = 165\,\mathrm{eV/c^2}$. When this model is applied to the Milky Way,[80] it leads to a large fermion ball of mass $M_c = 9.45\times10^9\,M_\odot$ and radius $R_c = 240\,\mathrm{pc}$. Therefore, it predicts the existence of a large dark matter bulge at the center of the Galaxy instead of a compact fermion ball mimicking a SMBH.[n] A large dark matter bulge is not inconsistent with the observations and may even solve some issues. For example, De Martino *et al.*[87] have argued that the presence of a bosonic dark matter bulge (soliton) of mass $M_c\simeq10^9\,M_\odot$ and radius $R_c\simeq100\,\mathrm{pc}$ at the center of the Galaxy may account for the dispersion velocity peak observed in the Milky Way. A large dark matter bulge made of fermions should have the same effect.[80]

Finally, we mention potential difficulties or, alternatively, potentially important predictions associated with the model of Argüelles *et al.*[73,83] If the fermion mass is $m = 48\,\mathrm{keV/c^2}$, dark matter halos of mass $M_h = 10^8\,M_\odot$ such as dSphs like Fornax should have a very pronounced core-halo structure since they do not correspond to the ground state of the self-gravitating Fermi gas (unlike the model of Ref.[80] with $m = 165\,\mathrm{eV/c^2}$). More precisely, the fermionic dark matter model with a fermion mass $m = 48\,\mathrm{keV/c^2}$ predicts that dSphs of mass $M_h = 10^8\,M_\odot$ should contain a fermion ball of mass $M_c = 1.57\times10^4\,M_\odot$ and radius $R_c = 5.42\,\mathrm{mpc}$ possibly

[m] In very recent works, Becerra-Vergara *et al.*[84,85] showed that the gravitational potential of a fermion ball (with a particle mass $m = 56\,\mathrm{keV/c^2}$) leads to a better fit of the orbits of all the 17 best resolved S-stars orbiting Sgr A* (including the S2 and G3 objects) with respect to the one obtained by the central SMBH model.

[n] In that case, a primordial SMBH has to be introduced "by hand" at the center of the Galaxy in order to account for the presence of Sgr A*.

mimicking an intermediate mass black hole.[80] This result is consistent with the detailed work of Argüelles *et al.*[88] who obtained dense cores of mass between $M_c = 10^3\,M_\odot$ and $M_c = 10^6\,M_\odot$ depending on the central effective temperature of the fermions. This is either a very important prediction (if confirmed by observations) or the evidence that this model is incorrect (if invalidated by observations). It would be extremely important to clarify this issue by confronting the model of Argüelles *et al.*[73,83] to ultracompact halos in order to determine which of the two models (the model of Argüelles *et al.*[73,83] with $m = 48\,\mathrm{keV/c^2}$ or the one developed by Chavanis[80] with $m = 165\,\mathrm{eV/c^2}$ or $m \sim 1\,\mathrm{keV}/c^2$) is the most relevant for dark matter halos.

9. Conclusion

In these Proceedings, we have provided a brief history of the self-gravitating Fermi gas in Newtonian gravity and general relativity. We have focused exclusively on papers that discuss the caloric curves and the mass-radius relations of the self-gravitating Fermi gas. We have shown how these curves become more and more complex, displaying various types of phase transitions and instabilities, when gravity effects, thermal effects, quantum effects, relativity effects and tidal effects are progressively taken into account. Of course, there are many more interesting papers on self-gravitating fermions that are not reviewed here. A detailed bibliography on the subject can be found in Refs.[3,80] and in standard textbooks of astrophysics.

We have applied the self-gravitating Fermi gas model to dark matter halos. The Fermi-Dirac distribution may be justified either from the theory of collisionless violent relaxation[72,79] or from a collisional relaxation of nongravitational origin if the fermions are self-interacting.[80] If the fermions have a small mass ($m \lesssim 1\,\mathrm{keV}/c^2$), the caloric curve applied to the Milky Way has an N-shape structure (see Fig. 5-b) and the equilibrium states display a large quantum bulge of mass $M_c \sim 10^{10}\,M_\odot$ and radius $R_c \sim 100\,\mathrm{pc}$ surrounded by an isothermal atmosphere similar to the Burkert profile.[80] If the fermions have a large mass ($m \sim 50\,\mathrm{keV}/c^2$), the caloric curve has a Z-shape structure (see Fig. 6-a). It displays a nonrelativistic turning point of energy at E_c triggering the gravothermal catastrophe. For nonrelativistic fermions, the gravothermal catastrophe is stopped by quantum degeneracy (Pauli's exclusion principle).[60] This may lead to a compact fermion ball of mass $M_c \sim 4.2\times 10^6\,M_\odot$ and radius $R_c \sim 6\times 10^{-4}\,\mathrm{pc}$ mimicking a SMBH surrounded by an isothermal atmosphere.[73] When $N > N_{\rm OV}$, which is the case for the Milky Way, a new turning point of energy appears at E_c'' due to general relativity (see Figs. 8-b and 12). It triggers a gravitational collapse towards a SMBH. This new turning point of energy was first evidenced in Refs.[2,63,64] for box-confined fermions and confirmed in Ref.[73] for the fermionic King model. The possibility to form either a fermion ball or a SMBH at the center of the galaxies depends on the size of the galaxy. In medium size galaxies like the Milky Way (when $N < N_*'$) we expect to form a fermion ball of mass $M_c \sim 4.2\times 10^6\,M_\odot$ but in large galaxies (when $N > N_*'$) the condensed branch

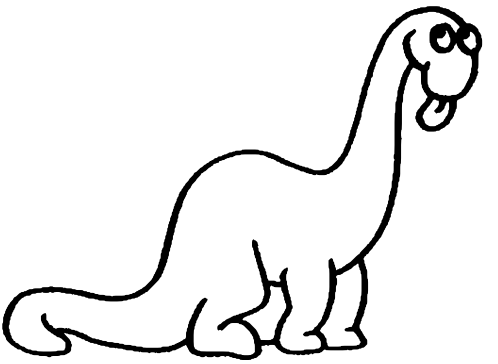


Fig. 13. Dinosaur (artistic view) similar to Fig. 6-a.

disappears (see Fig. 9-b) and the gravothermal catastrophe necessarily results in the formation of a SMBH of mass $M_{\rm OV} \sim 10^8\, M_\odot$.

It is interesting to study the effect of the dimension of space d on phase transitions in the self-gravitating Fermi gas. This is done in Refs.[17, 89–91]. In particular, it is shown that fermion stars are unstable in a universe with $d \geq 4$ dimensions. In that case, quantum mechanics cannot stabilize matter against gravitational collapse even in the nonrelativistic regime.[17, 89, 90] This is similar to a result found by Ehrenfest[92] who considered the effect of the dimension of space on the laws of physics and showed that planetary motion and the Bohr atom would not be stable in a space of dimension $d \geq 4$. Therefore, the dimension $d = 3$ of our Universe is very particular with possible implications regarding the Anthropic Principle.

Finally, it is interesting to compare the results obtained for fermion stars with those obtained for boson stars and self-gravitating Bose-Einstein condensates (BECs) (see our contribution[93] in these Proceedings). Similarly to fermionic dark matter halos, BEC dark matter halos also have a core-halo structure in which the "fermion ball" is replaced by a "soliton". The analogy between fermionic and bosonic dark matter halos is discussed in Refs.[80, 86, 94].

References

1. G. Alberti, P.H. Chavanis, Phys. Rev. E **101**, 052105 (2020).
2. G. Alberti, P.H. Chavanis, Eur. Phys. J. B **93**, 208 (2020).
3. P.H. Chavanis, Eur. Phys. J. Plus **135**, 290 (2020).
4. P.H. Chavanis, Eur. Phys. J. Plus **135**, 310 (2020).
5. A. Campa, T. Dauxois, D. Fanelli, S. Ruffo, *Physics of Long-Range Interacting Systems* (Oxford University Press, 2014).
6. R.H. Fowler, Mon. Not. R. Astron. Soc. **87**, 114 (1926).
7. E.C. Stoner, Phil. Mag. **7**, 63 (1929).
8. E.A. Milne, Mon. Not. R. Astron. Soc. **91**, 4 (1930).
9. S. Chandrasekhar, Phil. Mag. **11**, 592 (1931).
10. R. Emden, *Gaskugeln* (Leipzig, 1907).

11. S. Chandrasekhar, *An Introduction to the Theory of Stellar Structure* (University of Chicago Press, 1939).
12. J. Frenkel, Z. Phys. **50**, 234 (1928).
13. W. Anderson, Zeit. f. Phys. **56**, 851 (1929).
14. E.C. Stoner, Phil. Mag. **9**, 944 (1930).
15. S. Chandrasekhar, Astrophys. J. **74**, 81 (1931).
16. L.D. Landau, Phys. Zeit. Sow. **1**, 285 (1932).
17. P.H. Chavanis, Phys. Rev. D **76**, 023004 (2007).
18. S.A. Kaplan, Naukovy Zapiski (Sci. Notes Univ. Lwow) **15**, 109 (1949).
19. S. Chandrasekhar, R.F. Tooper, Astrophys. J. **139**, 1396 (1964).
20. T. Hamada, E.E. Salpeter, Astrophys. J. **134**, 683 (1961).
21. B.K. Harrison, K.S. Thorne, M. Wakano, J.A. Wheeler, *Gravitation Theory and Gravitational Collapse*, (Chicago University Press, Chicago, 1965).
22. S. Chandrasekhar, Mon. Not. R. Astron. Soc. **95**, 207 (1935).
23. A.S. Eddington, Mon. Not. R. Astron. Soc. **95**, 194 (1935).
24. J.R. Oppenheimer, G.M. Volkoff, Phys. Rev. **55**, 374 (1939).
25. J.R. Oppenheimer, H. Snyder, Phys. Rev. **56**, 455 (1939).
26. A. Hewish, S.J. Bell, J.D.H. Pilkington, P.F. Scott, R.A. Collins, Nature **217**, 709 (1968).
27. T. Gold, Nature **218**, 731 (1968).
28. T. Gold, Nature **221**, 25 (1969).
29. J.A. Wheeler, American Scientist **56**, 1 (1968).
30. T. Padmanabhan, Phys. Rep. **188**, 285 (1990).
31. J. Katz, Found. Phys. **33**, 223 (2003).
32. P.H. Chavanis, Int. J. Mod. Phys. B **20**, 3113 (2006).
33. V.A. Antonov, Vest. Leningr. Gos. Univ. **7**, 135 (1962).
34. K.F. Ogorodnikov, *Dynamics of Stellar Systems* (Pergamon, 1965).
35. D. Lynden-Bell, R. Wood, Mon. Not. R. Astron. Soc. **138**, 495 (1968).
36. H. Poincaré, Acta Math. **7**, 259 (1885).
37. W. Thirring, Z. Physik **235**, 339 (1970).
38. J. Katz, Mon. Not. R. Astron. Soc. **183**, 765 (1978).
39. S. Chandrasekhar, Astrophys. J. **87**, 535 (1938).
40. L.R. Henrich, S. Chandrasekhar, Astrophys. J. **94**, 525 (1941).
41. M. Schönberg, S. Chandrasekhar, Astrophys. J. **96**, 161 (1942).
42. R. Ebert, Z. Astrophys. **37**, 217 (1955).
43. W.B. Bonnor, Mon. Not. R. Astron. Soc. **116**, 351 (1956).
44. W.H. McCrea, Mon. Not. R. Astron. Soc. **117**, 562 (1957).
45. T. Padmanabhan, Astrophys. J. Supp. **71**, 651 (1989).
46. H.J. de Vega, N. Sanchez, F. Combes, Phys. Rev. D **54**, 6008 (1996).
47. H.J. de Vega, N. Sanchez, Nucl. Phys. B **625**, 409 (2002).
48. H.J. de Vega, N. Sanchez, Nucl. Phys. B **625**, 460 (2002).
49. P.H. Chavanis, Astron. Astrophys. **381**, 340 (2002).
50. P.H. Chavanis, Astron. Astrophys. **401**, 15 (2003).
51. J. Katz, I. Okamoto, Mon. Not. R. Astron. Soc. **317**, 163 (2000).
52. P.H. Chavanis, Astron. Astrophys. **432**, 117 (2005).
53. R.D. Sorkin, R.M. Wald, Z.Z. Jiu, Gen. Relat. Grav. **13**, 1127 (1981).
54. P.H. Chavanis, Astron. Astrophys. **483**, 673 (2008).
55. Z. Roupas, Class. Quantum Grav. **32**, 135023 (2015).
56. P. Hertel and W. Thirring, Commun. Math. Phys. **24**, 22 (1971).

57. P. Hertel and W. Thirring, Thermodynamic Instability of a System of Gravitating Fermions. In: H.P. Dürr (Ed.): *Quanten und Felder* (Brauschweig: Vieweg 1971).
58. N. Bilic, R.D. Viollier, Phys. Lett. B **408**, 75 (1997).
59. P.H. Chavanis, Phys. Rev. E **65**, 056123 (2002).
60. P.H. Chavanis, J. Sommeria, Mon. Not. R. Astron. Soc. **296**, 569 (1998).
61. P.H. Chavanis, Astron. Astrophys. **432**, 117 (2005).
62. N. Bilic, R.D. Viollier, Eur. Phys. J. C **11**, 173 (1999).
63. Z. Roupas, P.H. Chavanis, Class. Quant. Grav. **36**, 065001 (2019).
64. P.H. Chavanis, G. Alberti, Phys. Lett. B **801**, 135155 (2020).
65. Y. Pomeau, M. Le Berre, P.H. Chavanis and B. Denet, Eur. Phys. J. E **37**, 26 (2014).
66. P.H. Chavanis, B. Denet, M. Le Berre and Y. Pomeau, Eur. Phys. J. B **92**, 271 (2019).
67. P.H. Chavanis, B. Denet, M. Le Berre and Y. Pomeau, Eur. Phys. Lett. **129**, 30003 (2020).
68. R. Ruffini, L. Stella, Astron. Astrophys. **119**, 35 (1983).
69. I.R. King, Astron. J. **70**, 376 (1965).
70. P.H. Chavanis, Mon. Not. R. Astron. Soc. **300**, 981 (1998).
71. P.H. Chavanis, M. Lemou, F. Méhats, Phys. Rev. D **91**, 063531 (2015).
72. P.H. Chavanis, M. Lemou, F. Méhats, Phys. Rev. D **92**, 123527 (2015).
73. C.R. Argüelles, M.I. Díaz, A. Krut, R. Yunis, Mon. Not. R. astr. Soc. **502**, 4227 (2021).
74. A. Burkert, Astrophys. J. **447**, L25 (1995).
75. S.L. Shapiro, S.A. Teukolsky *Black Holes, White Dwarfs, and Neutron Stars* (Wiley Interscience, 1983).
76. M.A. Markov, Phys. Lett. **10**, 122 (1964).
77. R. Cowsik and J. McClelland, Phys. Rev. Lett. **29**, 669 (1972).
78. R. Cowsik and J. McClelland, Astrophys. J. **180**, 7 (1973).
79. D. Lynden-Bell, Mon. Not. Roy. Astr. Soc. **136**, 101 (1967).
80. P.H. Chavanis, *Predictive model of fermionic dark matter halos with a quantum core and an isothermal atmosphere*, preprint
81. N. Bilic, G.B. Tupper, R.D. Viollier, Lect. Notes Phys. **616**, 24 (2003).
82. R. Ruffini, C.R. Argüelles, J.A. Rueda, Mon. Not. R. Astron. Soc. **451**, 622 (2015).
83. C.R. Argüelles, A. Krut, J.A. Rueda, R. Ruffini, Phys. Dark Univ. **21**, 82 (2018).
84. E.A. Becerra-Vergara, C.R. Argüelles, A. Krut, J.A. Rueda, R. Ruffini, Astron. Astrophys. **641**, A34 (2020).
85. E.A. Becerra-Vergara, C.R. Argüelles, A. Krut, J.A. Rueda, R. Ruffini, Mon. Not. Roy. Astr. Soc. **505**, L64 (2021).
86. P.H. Chavanis, Phys. Rev. D **100**, 123506 (2019).
87. I. De Martino, T. Broadhurst, S.H. Henry Tye, T. Chiueh, H.Y. Schive, Phys. Dark Univ. **28**, 100503 (2020).
88. C.R. Argüelles, A. Krut, J.A. Rueda, R. Ruffini, Phys. Dark Univ. **24**, 100278 (2019).
89. P.H. Chavanis, Phys. Rev. E **69**, 066126 (2004).
90. P.H. Chavanis, C. R. Physique **7**, 331 (2006).
91. M. Kirejczyk, G. Müller, P.H. Chavanis, arXiv:2110.01044
92. P. Ehrenfest, Proc. Amst. Acad. **20**, 200 (1917).
93. P.H. Chavanis, *The maximum mass of dilute axion stars*, preprint.
94. P.H. Chavanis, Phys. Rev. D **100**, 083022 (2019).

The dark matter: DAMA/LIBRA and its perspectives

R. Bernabei, P. Belli*, V. Caracciolo, R. Cerulli and V. Merlo

Dip. Fisica, Università di Roma "Tor Vergata", 00133 Rome, Italy
INFN sezione di Roma "Tor Vergata", 00133 Rome, Italy
** E-mail: pierluigi.belli@roma2.infn.it*

F. Cappella, A. d'Angelo and A. Incicchitti

Dip. Fisica, Università di Roma "La Sapienza", 00185 Rome, Italy
INFN sezione di Roma, 00185 Rome, Italy

C.J. Dai, X.H. Ma and X.D. Sheng

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, 100049 Beijing, PR China

F. Montecchia

INFN sezione di Roma "Tor Vergata", 00133 Rome, Italy
Dip. Ingegneria Civile e Ingegneria Informatica, Università di Roma "Tor Vergata", 00133 Rome, Italy

Z.P. Ye

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, 100049 Beijing, PR China
University of Jinggangshan, Ji'an, Jiangxi, PR China

Experimental observations and theoretical arguments point out that Dark Matter (DM) particles are one of the most prominent component of the Universe. This motivated the pioneer DAMA experiment to investigate the presence of these particles in the galactic halo, by exploiting the model independent signature of the DM annual modulation of the rate and very highly radio-pure apparatus in underground site. In this paper the results obtained by other two annual cycles of DAMA/LIBRA–phase2 are presented and the long-standing model-independent annual modulation effect measured by DAMA deep underground at the Gran Sasso National Laboratory (LNGS) of the I.N.F.N. with different experimental configurations is summarized. In particular, the DAMA/LIBRA–phase2 apparatus, $\simeq$ 250 kg highly radio-pure NaI(Tl), profits from a second generation high quantum efficiency photomultipliers and of new electronics with respect to DAMA/LIBRA–phase1. The improved experimental configuration has also allowed to lower the software energy threshold. Including the results of these other two annual cycles presented here, the total exposure of DAMA/LIBRA–phase2 over 8 annual cycles is 1.53 ton $\times$ yr. DAMA/LIBRA–phase2 confirms the evidence of a signal that meets all the requirements of the model independent Dark Matter annual modulation signature, at 11.8 σ C.L. in the energy region (1–6) keV. In the energy region between 2 and 6 keV, where data are also available from DAMA/NaI and DAMA/LIBRA–phase1 (14 additional annual cycles), the achieved C.L. for the full exposure (2.86 ton $\times$ yr) is 13.7 σ; the modulation amplitude of the *single-hit* scintillation events is: (0.01014 $\pm$ 0.00074) cpd/kg/keV,

the measured phase is (142.4 ± 4.2) days and the measured period is (0.99834 ± 0.00067) yr, all these values are well in agreement with those expected for DM particles. No systematics or side reaction able to mimic the exploited DM signature (i.e. to account for the whole measured modulation amplitude and to simultaneously satisfy all the requirements of the signature), has been found or suggested by anyone throughout some decades thus far.

Keywords: Scintillation detectors; elementary particle processes; Dark Matter; annual modulation.

1. Introduction

The DAMA/LIBRA[1–23] experiment, as the pioneer DAMA/NaI,[24–51] has the main aim to investigate the presence of DM particles in the galactic halo by exploiting the DM annual modulation signature (originally suggested in Ref.[52,53]). In addition, the developed highly radio-pure NaI(Tl) target-detectors[1,6,9,54] ensure sensitivity to a wide range of DM candidates, interaction types and astrophysical scenarios (see e.g. Refs.[2,14,16–18,25–32,35–42], and in literature).

The origin of the DM annual modulation signature and of its peculiar features is due to the Earth motion with respect to the DM particles constituting the Galactic Dark Halo. In fact, as a consequence of the Earth's revolution around the Sun, which is moving in the Galaxy with respect to the Local Standard of Rest towards the star Vega near the constellation of Hercules, the Earth should be crossed by a larger flux of DM particles around $\simeq$ 2 June and by a smaller one around $\simeq$ 2 December. In the former case the Earth orbital velocity is summed to that of the solar system with respect to the Galaxy, while in the latter the two velocities are subtracted. The DM annual modulation signature is very distinctive since the effect induced by DM particles must simultaneously satisfy all the following requirements: the rate must contain a component modulated according to a cosine function (1) with one year period (2) and a phase that peaks roughly $\simeq$ 2 June (3); this modulation must only be found in a well-defined low energy range, where DM particle induced events can be present (4); it must apply only to those events in which just one detector of many actually "fires" (*single-hit* events), since the DM particle multi-interaction probability is negligible (5); the modulation amplitude in the region of maximal sensitivity must be $\lesssim 7\%$ of the constant part of the signal for usually adopted halo distributions (6), but it can be larger in case of some proposed scenarios such as e.g. those in Ref.[55–59] (even up to $\simeq 30\%$). Thus this signature has many peculiarities and, in addition, it allows to test a wide range of parameters in many possible astrophysical, nuclear and particle physics scenarios.

This DM signature might be mimicked only by systematic effects or side reactions able to account for the whole observed modulation amplitude and to simultaneously satisfy all the requirements given above.

The full description of the DAMA/LIBRA set-up and the adopted procedures during the phase1 and phase2 and other related arguments have been discussed in details e.g. in Refs.[1–6,19–21,23].

At the end of 2010 all the photomultipliers (PMTs) were replaced by a second generation PMTs Hamamatsu R6233MOD, with higher quantum efficiency (Q.E.) and with lower background with respect to those used in phase1; they were produced after a dedicated R&D in the company, and tests and selections.[6,54] The new PMTs have Q.E. in the range 33-39% at 420 nm, wavelength of NaI(Tl) emission, and in the range 36-44% at peak. The commissioning of the DAMA/LIBRA–phase2 experiment was successfully performed in 2011, allowing the achievement of the software energy threshold at 1 keV, and the improvement of some detector's features such as energy resolution and acceptance efficiency near software energy threshold.[6]

The adopted procedure for noise rejection near software energy threshold and the acceptance windows are the same unchanged along all the DAMA/LIBRA–phase2 data taking, throughout the months and the annual cycles. The typical behaviour of the overall efficiency for *single-hit* events as a function of the energy is also shown in Ref.[6]; the percentage variations of the efficiency follow a gaussian distribution with $\sigma = 0.3\%$ and do not show any modulation with period and phase as expected for the DM signal (for a partial data release see Ref.[21]).

At the end of 2012 new preamplifiers and special developed trigger modules were installed and the apparatus was equipped with more compact electronic modules.[60] Here we just remind that the sensitive part of DAMA/LIBRA–phase2 set-up is made of 25 highly radio-pure NaI(Tl) crystal scintillators (5-rows by 5-columns matrix) having 9.70 kg mass each one; quantitative analyses of residual contaminants are given in Ref.[1]. In each detector two 10 cm long UV light guides (made of Suprasil B quartz) act also as optical windows on the two end faces of the crystal, and are coupled to two low background PMTs working in coincidence at single photoelectron level. The detectors are housed in a sealed low-radioactive copper box installed in the center of a low-radioactive Cu/Pb/Cd-foils/polyethylene/paraffin shield; moreover, about 1 m concrete (made from the Gran Sasso rock material) almost fully surrounds (mostly outside the barrack) this passive shield, acting as a further neutron moderator. The shield is decoupled from the ground by a metallic structure mounted above a concrete basement; a neoprene layer separates the concrete basement and the floor of the laboratory. The space between this basement and the metallic structure is filled by paraffin for several tens cm in height.

A threefold-level sealing system prevents the detectors from contact with the environmental air of the underground laboratory and continuously maintains them in HP (high-purity) Nitrogen atmosphere. The whole installation is under air conditioning to ensure a suitable and stable working temperature. The huge heat capacity of the multi-tons passive shield ($\approx 10^6$ cal/^{o}C) guarantees further relevant stability of the detectors' operating temperature. In particular, two independent systems of air conditioning are available for redundancy: one cooled by water refrigerated by a dedicated chiller and the other operating with cooling gas. A hardware/software monitoring system provides data on the operating conditions. In particular, several probes are read out and the results are stored with the production data.

Moreover, self-controlled computer based processes automatically monitor several parameters, including those from DAQ, and manage the alarms system. All these procedures, already experienced during DAMA/LIBRA–phase1,[1–5] allow us to control and to maintain the running conditions stable at a level better than 1% also in DAMA/LIBRA–phase2 (see e.g. Ref.[21,23]).

The light response of the detectors during phase2 typically ranges from 6 to 10 photoelectrons/keV, depending on the detector. Energy calibration with X-rays/γ sources are regularly carried out in the same running condition down to few keV (for details see e.g. Ref.[1]; in particular, double coincidences due to internal X-rays from ^{40}K (which is at ppt levels in the crystals) provide (when summing the data over long periods) a calibration point at 3.2 keV close to the software energy threshold. The DAQ system records both *single-hit* events (where just one of the detectors fires) and *multiple-hit* events (where more than one detector fires) up to the MeV region despite the optimization is performed for the lowest energy.

The radio-purity and details are discussed e.g. in Refs.[1–5,54] and references therein. The adopted procedures provide sensitivity to large and low mass DM candidates inducing nuclear recoils and/or electromagnetic signals.

The data of the former DAMA/NaI setup and, later, those of the DAMA/LIBRA-phase1 have already given (with high confidence level) positive evidence for the presence of a signal that satisfies all the requirements of the exploited DM annual modulation signature.[2–5,35,36]

In this paper the model independent result of eight annual cycles of DAMA/LIBRA–phase2 is presented. The total exposure of DAMA/LIBRA–phase2 is: 1.53 ton $\times$ yr with an energy threshold at 1 keV; when including also that of the first generation DAMA/NaI experiment and DAMA/LIBRA–phase1 the cumulative exposure is 2.86 ton $\times$ yr, corresponding to twenty-two independent annual cycles.

2. The DAMA/LIBRA–phase2 annual cycles

The details of the annual cycles of DAMA/LIBRA–phase2 are reported in Table 1. The first annual cycle was dedicated to the commissioning and to the optimizations towards the achievement of the 1 keV software energy threshold.[6] This period has: i) no data before/near Dec. 2, 2010 (the expected minimum of the DM signal); ii) data sets with some set-up modifications; iii) $(\alpha - \beta^2) = 0.355$ well different from 0.5 (i.e. the detectors were not being operational evenly throughout the year). Thus, this period cannot be used for the annual modulation studies; however, it has been used for other purposes.[6,13] Therefore, as shown in Table 1 the considered annual cycles of DAMA/LIBRA–phase2 are eight (exposure of 1.53 ton$\times$yr). The cumulative exposure, also considering the former DAMA/NaI and DAMA/LIBRA–phase1, is 2.86 ton$\times$yr.

The total number of events collected for the energy calibrations during the eight annual cycles of DAMA/LIBRA–phase2 is about 1.6×10^8, while about

Table 1. Details about the annual cycles of DAMA/LIBRA–phase2. The mean value of the squared cosine is $\alpha = \langle cos^2\omega(t-t_0)\rangle$ and the mean value of the cosine is $\beta = \langle cos\omega(t-t_0)\rangle$ (the averages are taken over the live time of the data taking and $t_0 = 152.5$ day, i.e. June 2^{nd}); thus, the variance of the cosine, $(\alpha - \beta^2)$, is $\simeq 0.5$ for a detector being operational evenly throughout the year.

DAMA/LIBRA–phase2 annual cycle	Period	Mass (kg)	Exposure (kg×day)	$(\alpha - \beta^2)$
1	Dec. 23, 2010 – Sept. 9, 2011	commissioning of phase2		
2	Nov. 2, 2011 – Sept. 11, 2012	242.5	62917	0.519
3	Oct. 8, 2012 – Sept. 2, 2013	242.5	60586	0.534
4	Sept. 8, 2013 – Sept. 1, 2014	242.5	73792	0.479
5	Sept. 1, 2014 – Sept. 9, 2015	242.5	71180	0.486
6	Sept. 10, 2015 – Aug. 24, 2016	242.5	67527	0.522
7	Sept. 7, 2016 – Sept. 25, 2017	242.5	75135	0.480
8	Sept. 25, 2017 – Aug. 20, 2018	242.5	68759	0.557
9	Aug. 24, 2018 – Oct. 3, 2019	242.5	77213	0.446
DAMA/LIBRA–phase2	Nov. 2, 2011 – Oct. 3, 2019	557109 kg×day $\simeq$ 1.53 ton×yr		0.501
DAMA/NaI + DAMA/LIBRA–phase1 + DAMA/LIBRA–phase2:			2.86 ton×yr	

1.7×10^5 events/keV have been collected for the evaluation of the acceptance window efficiency for noise rejection near the software energy threshold.[1,6]

As it can be inferred from Table 1, the duty cycle of the experiment is high, ranging between 76% and 86%. The routine calibrations and, in particular, the data collection for the acceptance windows efficiency mainly affect it.

3. The annual modulation of the residual rate

Fig. 1 shows the time behaviour of the experimental residual rates of the *single-hit* scintillation events in the (1–3), and (1–6) keV energy intervals for DAMA/LIBRA–phase2. The residual rates are calculated from the measured rate of the *single-hit* events after subtracting the constant part, as described in Refs.[2–5,35,36]. The null modulation hypothesis is rejected at very high C.L. by χ^2 test: $\chi^2 = 176$ and 202, respectively, over 69 d.o.f.. The P-values are P = 2.6×10^{-11}, and P = 5.6×10^{-15}, respectively. The residuals of the DAMA/NaI data (0.29 ton × yr) are given in Ref.[2,5,35,36], while those of DAMA/LIBRA–phase1 (1.04 ton × yr) in Ref.[2–5].

The former DAMA/LIBRA–phase1 and the new DAMA/LIBRA–phase2 residual rates of the *single-hit* scintillation events are reported in Fig. 2. The energy interval is from 2 keV, the software energy threshold of DAMA/LIBRA–phase1, up to 6 keV. The null modulation hypothesis is rejected at very high C.L. by χ^2 test: $\chi^2/d.o.f. = 240/119$, corresponding to P-value = 3.5×10^{-10}.

The *single-hit* residual rates of the DAMA/LIBRA–phase2 (Fig. 1) have been fitted with the function: $A\cos\omega(t-t_0)$, considering a period $T = \frac{2\pi}{\omega} = 1$ yr and a phase $t_0 = 152.5$ day (June 2^{nd}) as expected by the DM annual modulation signature; this can be repeated for the only case of (2-6) keV energy interval also including the former DAMA/NaI and DAMA/LIBRA–phase1 data. The goodness of the fits

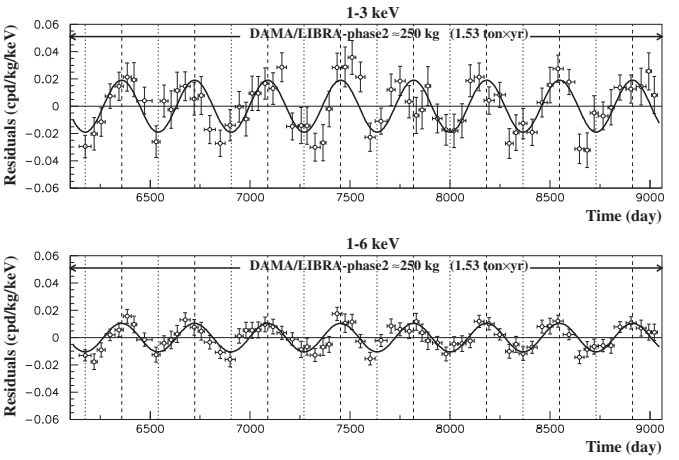


Fig. 1. Experimental residual rate of the *single-hit* scintillation events measured by DAMA/LIBRA–phase2 over eight annual cycles in the (1–3), and (1–6) keV energy intervals as a function of the time. The time scale is maintained the same of the previous DAMA papers for consistency. The data points present the experimental errors as vertical bars and the associated time bin width as horizontal bars. The superimposed curves are the cosinusoidal functional forms $A\cos\omega(t-t_0)$ with a period $T=\frac{2\pi}{\omega}=1$ yr, a phase $t_0=152.5$ day (June 2^{nd}) and modulation amplitudes, A, equal to the central values obtained by best fit on the data points of the entire DAMA/LIBRA–phase2. The dashed vertical lines correspond to the maximum expected for the DM signal (June 2^{nd}), while the dotted vertical lines correspond to the minimum.

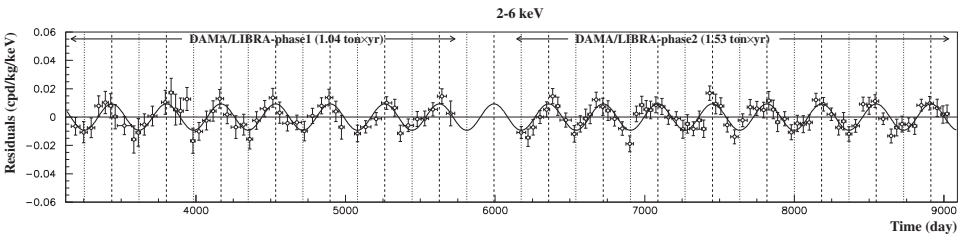


Fig. 2. Experimental residual rate of the *single-hit* scintillation events measured by DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 in the (2–6) keV energy intervals as a function of the time. The superimposed curve is the cosinusoidal functional forms $A\cos\omega(t-t_0)$ with a period $T=\frac{2\pi}{\omega}=1$ yr, a phase $t_0=152.5$ day (June 2^{nd}) and modulation amplitude, A, equal to the central value obtained by best fit on the data points of DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2. For details see Fig. 1.

is well supported by the χ^2 test; for example, $\chi^2/d.o.f. = 81.6/68, 66.2/68, 130/155$ are obtained for the (1–3) keV and (1–6) keV cases of DAMA/LIBRA–phase2, and for the (2–6) keV case of DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2, respectively. The results of the best fits in the different cases are summarized in Table 2. Table 2 also reports the cases when the period and the phase are kept free in the fitting procedure. The period and the phase are well compatible with expectations for a DM annual modulation signal. In particular, the phase is consistent with

Table 2. Modulation amplitude, A, obtained by fitting the *single-hit* residual rate of DAMA/LIBRA–phase2, as reported in Fig. 1, and also including the residual rates of the former DAMA/NaI and DAMA/LIBRA–phase1. It was obtained by fitting the data with the formula: $A\cos\omega(t-t_0)$. The period $T=\frac{2\pi}{\omega}$ and the phase t_0 are kept fixed at 1 yr and at 152.5 day (June 2^{nd}), respectively, as expected by the DM annual modulation signature, and alternatively kept free. The results are well compatible with expectations for a signal in the DM annual modulation signature.

	A (cpd/kg/keV)	$T=\frac{2\pi}{\omega}$ (yr)	t_0 (days)	C.L.
DAMA/LIBRA–phase2:				
1-3 keV	(0.0191±0.0020)	1.0	152.5	9.7 σ
1-6 keV	(0.01048±0.00090)	1.0	152.5	11.6 σ
2-6 keV	(0.00933±0.00094)	1.0	152.5	9.9 σ
1-3 keV	(0.0191±0.0020)	(0.99952±0.00080)	149.6±5.9	9.6 σ
1-6 keV	(0.01058±0.00090)	(0.99882±0.00065)	144.5±5.1	11.8 σ
2-6 keV	(0.00954±0.00076)	(0.99836±0.00075)	141.1±5.9	12.6 σ
DAMA/LIBRA–phase1 + phase2:				
2-6 keV	(0.00941±0.00076)	1.0	152.5	12.4 σ
2-6 keV	(0.00959±0.00076)	(0.99835±0.00069)	142.0±4.5	12.6 σ
DAMA/NaI + DAMA/LIBRA–phase1 + phase2:				
2-6 keV	(0.00996±0.00074)	1.0	152.5	13.4 σ
2-6 keV	(0.01014±0.00074)	(0.99834±0.00067)	142.4±4.2	13.7 σ

about June 2^{nd} and is fully consistent with the value independently determined by Maximum Likelihood analysis (see later). For completeness, we recall that a slight energy dependence of the phase could be expected (see e.g. Refs.[38, 58, 59, 61–63]), providing intriguing information on the nature of Dark Matter candidate and related aspects.

4. Absence of modulation of the background

As done in previous data releases, absence of any significant background modulation in the energy spectrum has also been verified in the present data taking for energy regions not of interest for DM. In fact, the background in the lowest energy region is essentially due to "Compton" electrons, X-rays and/or Auger electrons, muon induced events, etc., which are strictly correlated with the events in the higher energy region of the spectrum. Thus, if a modulation detected in the lowest energy region were due to a modulation of the background (rather than to a signal), an equal or larger modulation in the higher energy regions should be present.

For example, the measured rate integrated above 90 keV, R_{90}, as a function of the time has been analysed. Fig. 3 shows the distribution of the percentage variations of R_{90} with respect to the mean values for all the detectors in DAMA/LIBRA–phase2. It shows a cumulative gaussian behaviour with $\sigma \simeq 1\%$, well accounted for by the statistical spread expected from the used sampling time.

Moreover, fitting the time behaviour of R_{90} including a term with phase and period as for DM particles, a modulation amplitude $A_{R_{90}}$ compatible with zero has

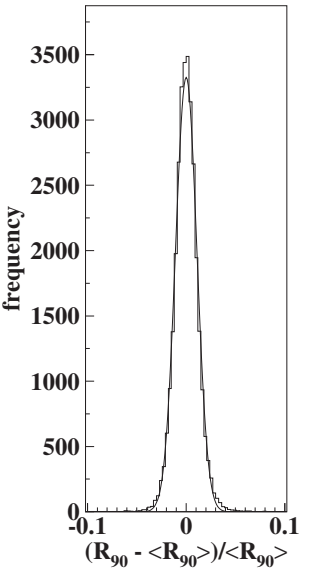


Fig. 3. Distribution of the percentage variations of R_{90} with respect to the mean values for all the detectors in the DAMA/LIBRA–phase2 (histogram); the superimposed curve is a gaussian fit.

been found for all the annual cycles (see Table 3). This also excludes the presence of any background modulation in the whole energy spectrum at a level much lower than the effect found in the lowest energy region for the *single-hit* scintillation events. In fact, otherwise – considering the R_{90} mean values – a modulation amplitude of order of tens cpd/kg would be present for each annual cycle, that is $\simeq 100\ \sigma$ far away from the measured values.

Table 3. Modulation amplitudes, $A_{R_{90}}$, obtained by fitting the time behaviour of R_{90} in DAMA/LIBRA–phase2, including a term with a cosine function having phase and period as expected for a DM signal. The obtained amplitudes are compatible with zero, and incompatible ($\simeq 100\ \sigma$) with modulation amplitudes of tens cpd/kg. Modulation amplitudes, $A_{(6-14)}$, obtained by fitting the time behaviour of the residual rates of the *single-hit* scintillation events in the (6–14) keV energy interval. In the fit the phase and the period are at the values expected for a DM signal. The obtained amplitudes are compatible with zero.

DAMA/LIBRA–phase2 annual cycle	$A_{R_{90}}$ (cpd/kg)	$A_{(6-14)}$ (cpd/kg/keV)
2	(0.12±0.14)	(0.0032±0.0017)
3	-(0.08±0.14)	(0.0016±0.0017)
4	(0.07±0.15)	(0.0024±0.0015)
5	-(0.05±0.14)	-(0.0004±0.0015)
6	(0.03±0.13)	(0.0001±0.0015)
7	-(0.09±0.14)	(0.0015±0.0014)
8	-(0.18±0.13)	-(0.0005±0.0013)
9	(0.08±0.14)	-(0.0003±0.0014)

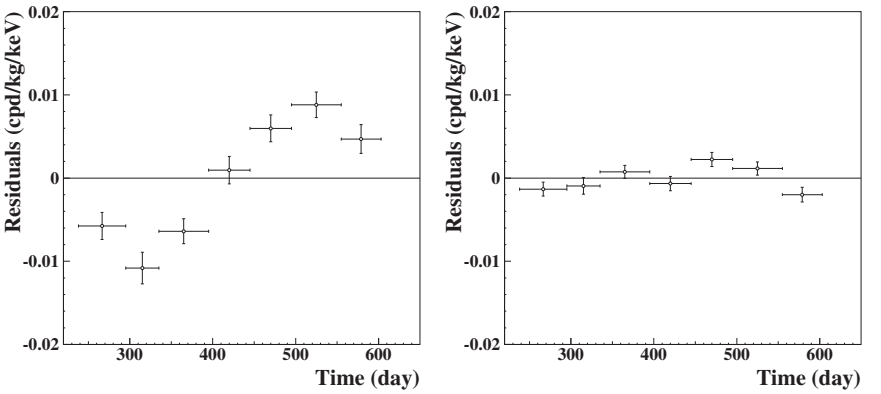


Fig. 4. Experimental *single-hit* residuals in the (1–6) keV and in the (10–20) keV energy regions for DAMA/LIBRA–phase2 as if they were collected in a single annual cycle (i.e. binning in the variable time from the January 1st of each annual cycle). The data points present the experimental errors as vertical bars and the associated time bin width as horizontal bars. The initial time of the figures is taken at August 7^{th}. A clear modulation satisfying all the peculiarities of the DM annual modulation signature is present in the lowest energy interval with A=(0.00956 ± 0.00090) cpd/kg/keV, while it is absent just above: A=(0.0007 ± 0.0005) cpd/kg/keV.

Similar results are obtained when comparing the *single-hit* residuals in the (1–6) keV with those in other energy intervals; for example Fig. 4 shows the *single-hit* residuals in the (1–6) keV and in the (10–20) keV energy regions for DAMA/LIBRA–phase2 as if they were collected in a single annual cycle (i.e. binning in the variable time from the January 1st of each annual cycle).

Moreover, Table 3 shows the modulation amplitudes obtained by fitting the time behaviour of the residual rates of the *single-hit* scintillation events in the (6–14) keV energy interval for the DAMA/LIBRA–phase2 annual cycles. In the fit the phase and the period are at the values expected for a DM signal. The obtained amplitudes are compatible with zero.

A further relevant investigation on DAMA/LIBRA–phase2 data has been performed by applying the same hardware and software procedures, used to acquire and to analyse the *single-hit* residual rate, to the *multiple-hit* one. Since the probability that a DM particle interacts in more than one detector is negligible, a DM signal can be present just in the *single-hit* residual rate. Thus, the comparison of the results of the *single-hit* events with those of the *multiple-hit* ones corresponds to compare the cases of DM particles beam-on and beam-off. This procedure also allows an additional test of the background behaviour in the same energy interval where the positive effect is observed.

In particular, in Fig. 5 the residual rates of the *single-hit* scintillation events collected during DAMA/LIBRA–phase2 are reported, as collected in a single cycle, together with the residual rates of the *multiple-hit* events, in the considered energy intervals. While, as already observed, a clear modulation, satisfying all the

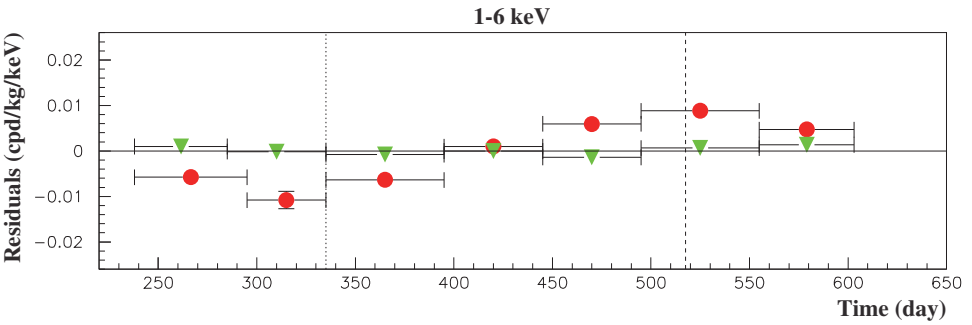


Fig. 5. Experimental residual rates of DAMA/LIBRA–phase2 *single-hit* events (filled red on-line circles), class of events to which DM events belong, and for *multiple-hit* events (filled green on-line triangles), class of events to which DM events do not belong. They have been obtained by considering for each class of events the data as collected in a single annual cycle and by using in both cases the same identical hardware and the same identical software procedures. The initial time of the figure is taken on August 7^{th}. The experimental points present the errors as vertical bars and the associated time bin width as horizontal bars. Analogous results were obtained for DAMA/NaI (two last annual cycles) and DAMA/LIBRA–phase1.[2–5, 36]

peculiarities of the DM annual modulation signature, is present in the *single-hit* events, the fitted modulation amplitude for the *multiple-hit* residual rate is well compatible with zero: (0.00030 ± 0.00032) cpd/kg/keV in the (1–6) keV energy region. Thus, again evidence of annual modulation with proper features as required by the DM annual modulation signature is present in the *single-hit* residuals (events class to which the DM particle induced events belong), while it is absent in the *multiple-hit* residual rate (event class to which only background events belong). Similar results were also obtained for the two last annual cycles of DAMA/NaI[36] and for DAMA/LIBRA–phase1.[2–5] Since the same identical hardware and the same identical software procedures have been used to analyse the two classes of events, the obtained result offers an additional strong support for the presence of a DM particle component in the galactic halo.

In conclusion, no background process able to mimic the DM annual modulation signature (that is, able to simultaneously satisfy all the peculiarities of the signature and to account for the measured modulation amplitude) has been found or suggested by anyone throughout some decades thus far (see also discussions e.g. in Ref.[1–5, 7, 8, 19–21, 23, 34–36]).

5. The analysis in frequency

To perform the Fourier analysis of the DAMA/LIBRA–phase1 and phase2 data in a wider region of considered frequency, the *single-hit* events have been grouped in 1 day bins. Due to the low statistics in each time bin, a procedure detailed in Ref.[64] has been followed. The whole power spectra up to the Nyquist frequency and the zoomed ones are reported in Fig. 6. A clear peak corresponding to a period of

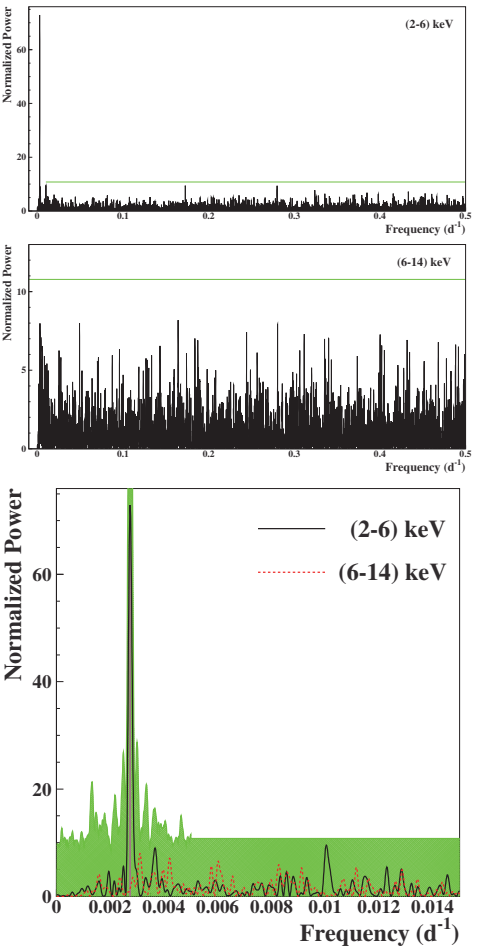


Fig. 6. Power spectra of the time sequence of the measured *single-hit* events for DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 grouped in 1 day bins. From top to bottom: spectra up to the Nyquist frequency for (2–6) keV and (6–14) keV energy intervals and their zoom around the 1 y^{-1} peak, for (2–6) keV (solid line) and (6–14) keV (dotted line) energy intervals. The main mode present at the lowest energy interval corresponds to a frequency of 2.74×10^{-3} d^{-1} (vertical line, purple on-line). It corresponds to a period of $\simeq 1$ year. A similar peak is not present in the (6–14) keV energy interval. The shaded (green on-line) area in the bottom figure – calculated by Monte Carlo procedure – represents the 90% C.L. region where all the peaks are expected to fall for the (2–6) keV energy interval. In the frequency range far from the signal for the (2–6) keV energy region and for the whole (6–14) keV spectrum, the upper limit of the shaded region (90% C.L.) can be calculated to be 10.8 (continuous lines, green on-line).

1 year is evident for the lowest energy interval; the same analysis in the (6–14) keV energy region shows only aliasing peaks instead. Neither other structure at different frequencies has been observed.

As to the significance of the peaks present in the periodogram, we remind that the periodogram ordinate, z, at each frequency follows a simple exponential

distribution e^{-z} in the case of the null hypothesis or white noise.[65] Therefore, if M independent frequencies are scanned, the probability to obtain values larger than z is: $P(>z) = 1 - (1 - e^{-z})^M$.

In general M depends on the number of sampled frequencies, the number of data points N, and their detailed spacing. It turns out that M is very nearly equal to N when the data points are approximately equally spaced, and when the sampled frequencies cover the frequency range from 0 to the Nyquist frequency.[66,67]

The number of data points used to obtain the spectra in Fig. 6 is $N = 5047$ (days measured over the 5479 days of the 15 DAMA/LIBRA–phase1 and phase2 annual cycles) and the full frequencies region up to Nyquist frequency has been scanned. Therefore, assuming $M = N$, the significance levels $P = 0.10$, 0.05 and 0.01, correspond to peaks with heights larger than $z = 10.8$, 11.5 and 13.1, respectively, in the spectra of Fig 6.

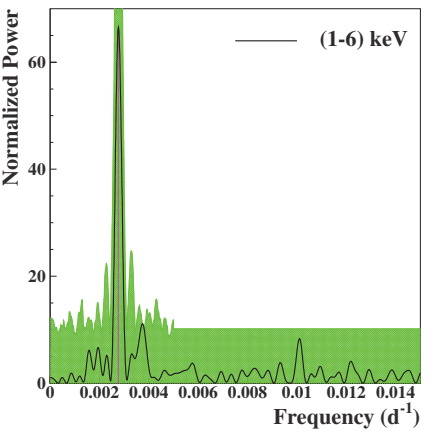


Fig. 7. Power spectrum of the time sequence of the measured *single-hit* events in the (1–6) keV energy interval for DAMA/LIBRA–phase2 grouped in 1 day bin. The main mode present at the lowest energy interval corresponds to a frequency of 2.77×10^{-3} d^{-1} (vertical line, purple on-line). It corresponds to a period of $\simeq$ 1 year. The shaded (green on-line) area – calculated by Monte Carlo procedure – represents the 90% C.L. region where all the peaks are expected to fall for the (1–6) keV energy interval.

In the case below 6 keV, a signal is present; thus, to properly evaluate the C.L. the signal must be included. This has been done by a dedicated Monte Carlo procedure where a large number of similar experiments has been simulated. The 90% C.L. region (shaded, green on-line) where all the peaks are expected to fall for the (2–6) keV energy interval is reported in Fig 6. Several peaks, satellite of the one year period frequency, are present.

In conclusion, apart from the peak corresponding to a 1 year period, no other peak is statistically significant either in the low and high energy regions.

Moreover, for each annual cycle of DAMA/LIBRA–phase1 and phase2, the annual baseline counting rates have been calculated for the (2–6) keV energy interval. Their power spectrum in the frequency range $0.00013 - 0.0019$ d^{-1} (corresponding to a period range 1.4–21.1 year) has been calculated according to Ref.[5]. No statistically-significant peak is present at frequencies lower than 1 y^{-1}. This implies that no evidence for a long term modulation in the counting rate is present.

Finally, the case of the (1–6) keV energy interval of the DAMA/LIBRA–phase2 data is reported in Fig. 7. As previously the only significant peak is the one corresponding to one year period. No other peak is statistically significant being below the shaded (green on-line) area obtained by Monte Carlo procedure.

6. The modulation amplitudes by the maximum likelihood approach

The annual modulation present at low energy can also be pointed out by depicting the energy dependence of the modulation amplitude, $S_m(E)$, obtained by maximum likelihood method considering fixed period and phase: T =1 yr and $t_0 = 152.5$ day. For such purpose the likelihood function of the *single-hit* experimental data in the $k-$th energy bin is defined as: $\mathbf{L_k} = \mathbf{\Pi}_{ij} e^{-\mu_{ijk}} \frac{\mu_{ijk}^{N_{ijk}}}{N_{ijk}!}$, where N_{ijk} is the number of events collected in the i-th time interval (hereafter 1 day), by the j-th detector and in the k-th energy bin. N_{ijk} follows a Poisson's distribution with expectation value $\mu_{ijk} = [b_{jk} + S_i(E_k)]\, M_j \Delta t_i \Delta E \epsilon_{jk}$. The b_{jk} are the background contributions, M_j is the mass of the $j-$th detector, Δt_i is the detector running time during the i-th time interval, ΔE is the chosen energy bin, ϵ_{jk} is the overall efficiency. The signal can be written as:

$$S_i(E) = S_0(E) + S_m(E) \cdot \cos \omega (t_i - t_0),$$

where $S_0(E)$ is the constant part of the signal and $S_m(E)$ is the modulation amplitude. The usual procedure is to minimize the function $y_k = -2ln(\mathbf{L_k}) - const$ for each energy bin; the free parameters of the fit are the $(b_{jk} + S_0)$ contributions and the S_m parameter.

The modulation amplitudes for the whole data sets: DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 (total exposure 2.86 ton×yr) are plotted in Fig. 8; the data below 2 keV refer only to the DAMA/LIBRA-phase2 exposure (1.53 ton×yr). It can be inferred that positive signal is present in the (1–6) keV energy interval, while S_m values compatible with zero are present just above. All this confirms the previous analyses. The test of the hypothesis that the S_m values in the (6–14) keV energy interval have random fluctuations around zero yields $\chi^2/d.o.f.$ equal to 20.3/16 (P-value = 21%).

For the case of (6–20) keV energy interval $\chi^2/d.o.f. = 42.2/28$ (P-value = 4%). The obtained χ^2 value is rather large due mainly to two data points, whose centroids are at 16.75 and 18.25 keV, far away from the (1–6) keV energy interval. The P-values obtained by excluding only the first and either the points are 14% and 23%.

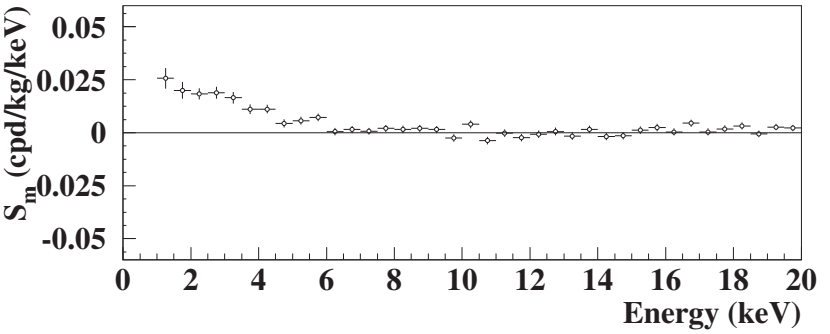


Fig. 8. Modulation amplitudes, S_m, for the whole data sets: DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 (total exposure 2.86 ton×yr) above 2 keV; below 2 keV only the DAMA/LIBRA-phase2 exposure (1.53 ton × yr) is available and used. The energy bin ΔE is 0.5 keV. A clear modulation is present in the lowest energy region, while S_m values compatible with zero are present just above. In fact, the S_m values in the (6–20) keV energy interval have random fluctuations around zero with $\chi^2/d.o.f.$ equal to 42.2/28 (P-value is 4%).

6.1. *The S_m distributions*

The method also allows the extraction of the S_m values for each detector. In particular, the modulation amplitudes S_m integrated in the range (2–6) keV for each of the 25 detectors for the DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 periods can be produced. They have random fluctuations around the weighted averaged value confirmed by the χ^2 analysis. Thus, the hypothesis that the signal is well distributed over all the 25 detectors is accepted.

As previously done for the other data releases,[2–5,19–21,23] the S_m values for each detector for each annual cycle and for each energy bin have been obtained. The S_m are expected to follow a normal distribution in absence of any systematic effects. Therefore, the variable $x = \frac{S_m - \langle S_m \rangle}{\sigma}$ has been considered to verify that the S_m are statistically well distributed in the 16 energy bins ($\Delta E = 0.25$ keV) in the (2–6) keV energy interval of the seven DAMA/LIBRA–phase1 annual cycles and in the 20 energy bins in the (1–6) keV energy interval of the eight DAMA/LIBRA–phase2 annual cycles and in each detector. Here, σ are the errors associated to S_m and $\langle S_m \rangle$ are the mean values of the S_m averaged over the detectors and the annual cycles for each considered energy bin.

Defining $\chi^2 = \Sigma x^2$, where the sum is extended over all the 272 (192 for the 16^{th} detector[4]) x values, $\chi^2/d.o.f.$ values ranging from 0.8 to 2.0 are obtained, depending on the detector.

The mean value of the 25 $\chi^2/d.o.f.$ is 1.092, slightly larger than 1. Although this can be still ascribed to statistical fluctuations, let us ascribe it to a possible systematics. In this case, one would derive an additional error to the modulation amplitude measured below 6 keV: $\leq 2.4 \times 10^{-4}$ cpd/kg/keV, if combining quadratically the errors, or $\leq 3.6 \times 10^{-5}$ cpd/kg/keV, if linearly combining them. This possible

additional error: $\leq 2.4\%$ or $\leq 0.4\%$, respectively, on the DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 modulation amplitudes is an upper limit of possible systematic effects coming from the detector to detector differences.

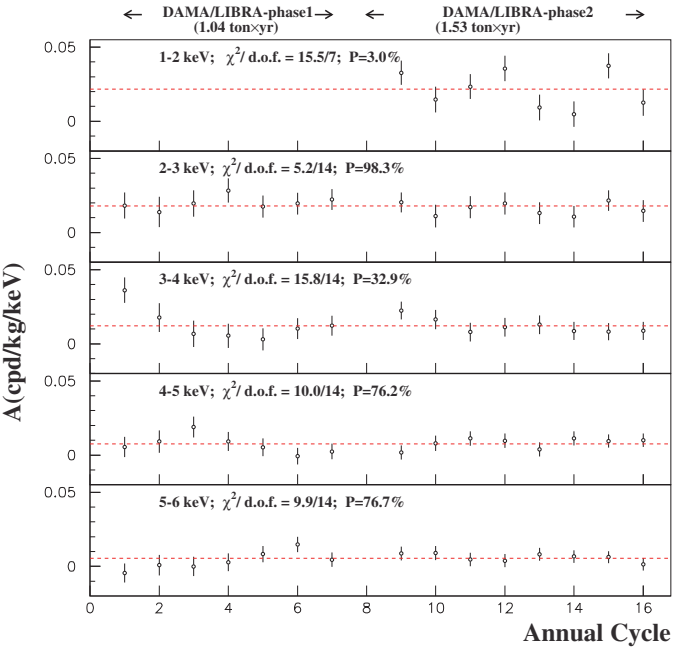


Fig. 9. Modulation amplitudes of each single annual cycle of DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2. The error bars are the 1σ errors. The dashed horizontal lines show the central values obtained by best fit over the whole data set. The χ^2 test and the *run test* accept the hypothesis at 95% C.L. that the modulation amplitudes are normally fluctuating around the best fit values.

Among further additional tests, the analysis of the modulation amplitudes as a function of the energy separately for the nine inner detectors and the remaining external ones has been carried out for DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2, as already done for the other data sets.[2–5, 19–21, 23] The obtained values are fully in agreement; in fact, the hypothesis that the two sets of modulation amplitudes belong to same distribution has been verified by χ^2 test, obtaining e.g.: $\chi^2/d.o.f. = 1.9/6$ and $36.1/38$ for the energy intervals (1–4) and (1–20) keV, respectively ($\Delta E = 0.5$ keV). This shows that the effect is also well shared between inner and outer detectors.

Fig. 9 shows the modulation amplitudes singularly calculated for each annual cycle of DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2. To test the hypothesis that the amplitudes are compatible and normally fluctuating around their mean values, the χ^2 test has been performed. The $\chi^2/d.o.f.$ values and the P-values are also shown in Fig. 9. In addition to the χ^2 test, another independent statistical test has been applied: the *run test* (see e.g. Ref.[69]); it verifies the hypothesis that the

positive (above the mean value) and negative (under the mean value) data points are randomly distributed. The lower (upper) tail probabilities obtained by the *run test* are: 89(37)%, 87(30)%, 17(94)%, 17(94)% and 30(85)%, respectively. This analysis confirms that the data collected in all the annual cycles with DAMA/LIBRA–phase1 and phase2 are statistically compatible and can be considered together.

7. Investigation of the annual modulation phase

Let us, finally, release the assumption of the phase $t_0 = 152.5$ day in the procedure to evaluate the modulation amplitudes. In this case the signal can be alternatively written as:

$$\begin{aligned} S_i(E) &= S_0(E) + S_m(E)\cos\omega(t_i - t_0) + Z_m(E)\sin\omega(t_i - t_0) \\ &= S_0(E) + Y_m(E)\cos\omega(t_i - t^*). \end{aligned} \tag{1}$$

For signals induced by DM particles one should expect: i) $Z_m \sim 0$ (because of the orthogonality between the cosine and the sine functions); ii) $S_m \simeq Y_m$; iii) $t^* \simeq t_0 = 152.5$ day. In fact, these conditions hold for most of the dark halo models; however, as mentioned above, slight differences can be expected in case of possible contributions from non-thermalized DM components (see e.g. Refs.[38, 58, 59, 61–63]).

Considering cumulatively the data of DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 the obtained 2σ contours in the plane (S_m, Z_m) for the (2–6) keV and (6–14) keV energy intervals are shown in Fig. 10–*left* while in Fig. 10–*right* the obtained 2σ contours in the plane (Y_m, t^*) are depicted. Moreover, Fig. 10 also shows only for DAMA/LIBRA–phase2 the 2σ contours in the (1–6) keV energy interval.

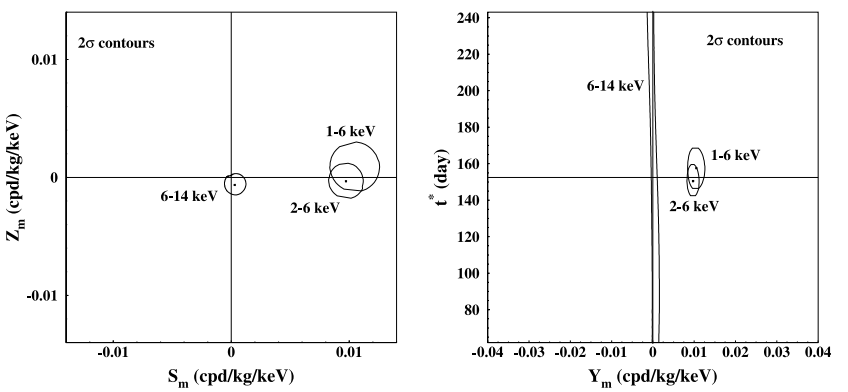


Fig. 10. 2σ contours in the plane (S_m, Z_m) (*left*) and in the plane (Y_m, t^*) (*right*) for: i) DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 in the (2–6) keV and (6–14) keV energy intervals (light areas, green on-line); ii) only DAMA/LIBRA–phase2 in the (1–6) keV energy interval (dark areas, blue on-line). The contours have been obtained by the maximum likelihood method. A modulation amplitude is present in the lower energy intervals and the phase agrees with that expected for DM induced signals.

The best fit values in the considered cases (1σ errors) for S_m versus Z_m and Y_m versus t^* are reported in Table 4.

Table 4. Best fit values (1σ errors) for S_m versus Z_m and Y_m versus t^*, considering: i) DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 in the (2–6) keV and (6–14) keV energy intervals; ii) only DAMA/LIBRA–phase2 in the (1–6) keV energy interval. See also Fig. 10.

E (keV)	S_m (cpd/kg/keV)	Z_m (cpd/kg/keV)	Y_m (cpd/kg/keV)	t^* (day)
DAMA/NaI+DAMA/LIBRA–phase1+DAMA/LIBRA–phase2:				
2–6	(0.0097 ± 0.0007)	-(0.0003 ± 0.0007)	(0.0097 ± 0.0007)	(150.5 ± 4.0)
6–14	(0.0003 ± 0.0005)	-(0.0006 ± 0.0005)	(0.0007 ± 0.0010)	undefined
DAMA/LIBRA–phase2:				
1–6	(0.0104 ± 0.0007)	(0.0002 ± 0.0007)	(0.0104 ± 0.0007)	(153.5 ± 4.0)

Finally, setting S_m in eq. (1) to zero, the Z_m values as function of the energy have also been determined by using the same procedure. The Z_m values as a function of the energy for DAMA/NaI, DAMA/LIBRA–phase1, and DAMA/LIBRA–phase2 data sets are expected to be zero. The χ^2 test applied to the data supports the hypothesis that the Z_m values are simply fluctuating around zero; in fact, in the (1–20) keV energy region the $\chi^2/d.o.f.$ is equal to 40.6/38 corresponding to a P-value = 36%.

The energy behaviors of the Y_m and of the phase t^* are produced for the cumulative exposure of DAMA/NaI, DAMA/LIBRA–phase1, and DAMA/LIBRA–phase2. As in the previous analyses, an annual modulation effect is present in the lower energy intervals and the phase agrees with that expected for DM induced signals. No modulation is present above 6 keV and the phase is undetermined.

8. Perspectives

To further increase the experimental sensitivity of DAMA/LIBRA and to disentangle some of the many possible astrophysical, nuclear and particle physics scenarios in the investigation on the DM candidate particle(s), an increase of the exposure in the lowest energy bin and a further decreasing of the software energy threshold are needed. This is pursued by running DAMA/LIBRA–phase2 and upgrading the experimental set-up to lower the software energy threshold below 1 keV with high acceptance efficiency.

Firstly, particular efforts for lowering the software energy threshold have been done in the already-acquired data of DAMA/LIBRA–phase2 by using the same technique as before with dedicated studies on the efficiency. As consequence, a new data point has been added in the modulation amplitude as function of energy down to 0.75 keV, see Fig. 11. A modulation is also present below 1 keV, from 0.75 keV.

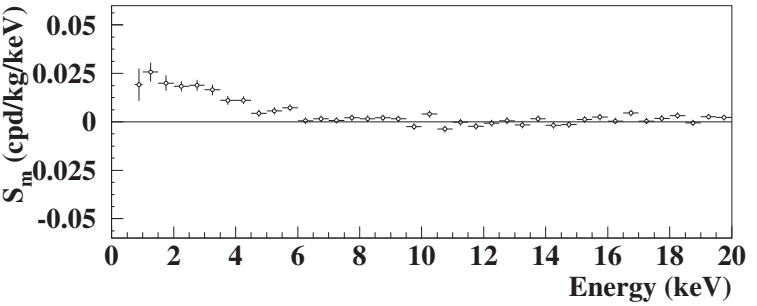


Fig. 11. As Fig. 8; the new data point below 1 keV, with software energy threshold at 0.75 keV, shows that an annual modulation is also present below 1 keV. This preliminary result confirms the necessity to lower the software energy threshold by a hardware upgrade and to improve the experimental error on the first energy bin.

This preliminary result confirms the necessity to lower the software energy threshold by a hardware upgrade and an improved statistics in the first energy bin.

This dedicated hardware upgrade of DAMA/LIBRA–phase2 is underway. It consists in equipping all the PMTs with miniaturized low background new concept preamplifier and HV divider mounted on the same socket, and related improvements of the electronic chain, mainly the use of higher vertical resolution 14-bit digitizers.

9. Conclusions

DAMA/LIBRA–phase2 confirms a peculiar annual modulation of the *single-hit* scintillation events in the (1–6) keV energy region satisfying all the many requirements of the DM annual modulation signature; the cumulative exposure by the former DAMA/NaI, DAMA/LIBRA–phase1 and DAMA/LIBRA–phase2 is 2.86 ton $\times$ yr.

As required by the exploited DM annual modulation signature: 1) the *single-hit* events show a clear cosine-like modulation as expected for the DM signal; 2) the measured period is well compatible with the 1 yr period as expected for the DM signal; 3) the measured phase is compatible with the roughly $\simeq 152.5$ days expected for the DM signal; 4) the modulation is present only in the low energy (1–6) keV interval and not in other higher energy regions, consistently with expectation for the DM signal; 5) the modulation is present only in the *single-hit* events, while it is absent in the *multiple-hit* ones as expected for the DM signal; 6) the measured modulation amplitude in NaI(Tl) target of the *single-hit* scintillation events in the (2–6) keV energy interval, for which data are also available by DAMA/NaI and DAMA/LIBRA–phase1, is: (0.01014 ± 0.00074) cpd/kg/keV (13.7 σ C.L.). No systematic or side processes able to mimic the signature, i.e. able to simultaneously satisfy all the many peculiarities of the signature and to account for the whole measured modulation amplitude, has been found or suggested by anyone throughout some decades thus far. In particular, arguments related to any possible role of

some natural periodical phenomena have been discussed and quantitatively demonstrated to be unable to mimic the signature (see references; e.g. Refs.[7,8]). Thus, on the basis of the exploited signature, the model independent DAMA results give evidence at 13.7σ C.L. (over 22 independent annual cycles and in various experimental configurations) for the presence of DM particles in the galactic halo.

The new data, released in this Conference, determine the modulation parameters with increasing precision and allow us to disentangle with larger C.L. among different DM candidates, DM models and astrophysical, nuclear and particle physics scenarios.

The DAMA model independent evidence is compatible with a wide set of astrophysical, nuclear and particle physics scenarios for high and low mass candidates inducing nuclear recoil and/or electromagnetic radiation, as also shown in various literature. Moreover, both the negative results and all the possible positive hints, achieved so-far in the field, can be compatible with the DAMA model independent DM annual modulation results in many scenarios considering also the existing experimental and theoretical uncertainties; the same holds for indirect approaches. For a discussion see e.g. Ref.[5] and references therein.

Finally, we stress that to efficiently disentangle among at least some of the many possible candidates and scenarios an increase of exposure in the new lowest energy bin and the decrease of the software energy threshold below the present 1 keV is important. The experiment is collecting data and the hardware efforts towards the lowering of the software energy threshold is in progress.

References

1. R. Bernabei et al., *Nucl. Instr. and Meth. A* **592**, 297 (2008).
2. R. Bernabei et al., *Eur. Phys. J. C* **56**, 333 (2008).
3. R. Bernabei et al., *Eur. Phys. J. C* **67**, 39 (2010).
4. R. Bernabei et al., *Eur. Phys. J. C* **73**, 2648 (2013).
5. R. Bernabei et al., *Int. J. of Mod. Phys. A* **28**, 1330022 (2013).
6. R. Bernabei et al., *J. of Instr.* **7**, P03009 (2012).
7. R. Bernabei et al., *Eur. Phys. J. C* **72**, 2064 (2012).
8. R. Bernabei et al., *Eur. Phys. J. C* **74**, 3196 (2014).
9. DAMA coll., issue dedicated to DAMA, *Int. J. of Mod. Phys. A* **31** (2016) and refs therein.
10. R. Bernabei et al., *Eur. Phys. J. C* **74**, 2827 (2014).
11. R. Bernabei et al., *Eur. Phys. J. C* **62**, 327 (2009).
12. R. Bernabei et al., *Eur. Phys. J. C* **72**, 1920 (2012).
13. R. Bernabei et al., *Eur. Phys. J. A* **49**, 64 (2013).
14. R. Bernabei et al., *Eur. Phys. J. C* **75**, 239 (2015).
15. P. Belli et al., *Phys. Rev. D* **84**, 055014 (2011).
16. A. Addazi et al., *Eur. Phys. J. C* **75**, 400 (2015).
17. R. Bernabei et al., *Int. J. of Mod. Phys. A* **31**, 1642009 (2016).
18. R. Cerulli et al., *Eur. Phys. J. C* **77**, 83 (2017).
19. R. Bernabei et al., *Universe* **4**, 116 (2018).
20. R. Bernabei et al., *Nucl. Phys. At. Energy* **19**, 307 (2018).
21. R. Bernabei, *Bled Workshops in Physics* **19** n. 2, 27 (2018).

22. R. Bernabei et al., *Nucl. Phys. At. Energy* **20(4)**, 317 (2019).
23. R. Bernabei et al., *Prog. Part. Nucl. Phys.* **114**, 103810 (2020).
24. P. Belli, R. Bernabei, C. Bacci, A. Incicchitti, R. Marcovaldi, D. Prosperi, DAMA proposal to INFN Scientific Committee II, April 24^{th} 1990.
25. R. Bernabei et al., *Phys. Lett. B* **389**, 757 (1996).
26. R. Bernabei et al., *Phys. Lett. B* **424**, 195 (1998).
27. R. Bernabei et al., *Phys. Lett. B* **450**, 448 (1999).
28. P. Belli et al., *Phys. Rev. D* **61**, 023512 (2000).
29. R. Bernabei et al., *Phys. Lett. B* **480**, 23 (2000).
30. R. Bernabei et al., *Phys. Lett. B* **509**, 197 (2001).
31. R. Bernabei et al., *Eur. Phys. J. C* **23**, 61 (2002).
32. P. Belli et al., *Phys. Rev. D* **66**, 043503 (2002)
33. R. Bernabei et al., *Il Nuovo Cim. A* **112**, 545 (1999).
34. R. Bernabei et al., *Eur. Phys. J. C* **18**, 283 (2000).
35. R. Bernabei el al., *La Rivista del Nuovo Cimento* **26** n.1, 1-73 (2003), and refs. therein.
36. R. Bernabei et al., *Int. J. Mod. Phys. D* **13**, 2127 (2004) and refs. therein.
37. R. Bernabei et al., *Int. J. Mod. Phys. A* **21**, 1445 (2006).
38. R. Bernabei et al., *Eur. Phys. J. C* **47**, 263 (2006).
39. R. Bernabei et al., *Int. J. Mod. Phys. A* **22**, 3155 (2007).
40. R. Bernabei et al., *Eur. Phys. J. C* **53**, 205 (2008).
41. R. Bernabei et al., *Phys. Rev. D* **77**, 023506 (2008).
42. R. Bernabei et al., *Mod. Phys. Lett. A* **23**, 2125 (2008).
43. R. Bernabei et al., *Phys. Lett. B* **408**, 439 (1997).
44. P. Belli et al., *Phys. Lett. B* **460**, 236 (1999).
45. R. Bernabei et al., *Phys. Rev. Lett.* **83**, 4918 (1999).
46. P. Belli et al., *Phys. Rev. C* **60**, 065501 (1999).
47. R. Bernabei et al., *Il Nuovo Cimento A* **112**, 1541 (1999).
48. R. Bernabei et al., *Phys. Lett. B* **515**, 6 (2001).
49. F. Cappella et al., *Eur. Phys. J.-direct C* **14**, 1 (2002).
50. R. Bernabei et al., *Eur. Phys. J. A* **23**, 7 (2005).
51. R. Bernabei et al., *Eur. Phys. J. A* **24**, 51 (2005).
52. K.A. Drukier et al., *Phys. Rev. D* **33**, 3495 (1986).
53. K. Freese et al., *Phys. Rev. D* **37**, 3388 (1988).
54. R. Bernabei and A. Incicchitti, *Int. J. Mod. Phys. A* **32**, 1743007 (2017).
55. D. Smith and N. Weiner, *Phys. Rev. D* **64**, 043502 (2001).
56. D. Tucker-Smith and N. Weiner, *Phys. Rev. D* **72**, 063509 (2005).
57. D. P. Finkbeiner et al, *Phys. Rev. D* **80**, 115008 (2009).
58. K. Freese et al., *Phys. Rev. D* **71**, 043516 (2005).
59. K. Freese et al., *Phys. Rev. Lett.* **92**, 111301 (2004).
60. P. Belli et al., *Int. J. of Mod. Phys. A* **31**, 1642005 (2016).
61. P. Gondolo et al., *New Astron. Rev.* **49**, 193 (2005).
62. G. Gelmini and P. Gondolo, *Phys. Rev. D* **64**, 023504 (2001).
63. F.S. Ling, P. Sikivie and S. Wick, *Phys. Rev. D* **70**, 123503 (2004).
64. G. Ranucci and M. Rovere, *Phys. Rev. D* **75**, 013010 (2007).
65. J.D. Scargle, *Astrophys. J.* **263**, 835 (1982).
66. W.H. Press et al., *Numerical recipes in Fortran 77: the art of scientific computing*, Cambridge University Press, Cambridge, England 1992, section 13.8.
67. J.H. Horne and S.L. Baliunas, *Astrophys. J.* **302**, 757 (1986).
68. R. Bernabei et al., *Bled Workshop in Physics* **15**, no. 2, 19 (2014).
69. W.T. Eadie et al., *Statistical methods in experimental physics*, ed. American Elsevier Pub. (1971).

Dark matter directionality approach

R. Bernabei, P. Belli, V. Caracciolo*, R. Cerulli, V. Merlo
Dip. di Fisica, Università di Roma "Tor Vergata", Rome, Italy
INFN, sez. di Roma "Tor Vergata", Rome, Italy
**E-mail: vincenzo.caracciolo@roma2.infn.it*

F. Cappella, A. Incicchitti
Dip. di Fisica, Università di Roma "La Sapienza", Rome, Italy
INFN, sez. di Roma, Rome, Italy

N. Cherubini, E. Piccinelli
ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, C.R: Casaccia, Roma 00123, Italy

F. A. Danevich, D. V. Kasperovych, O. G. Polischuk, V. I. Tretyak
Institute for Nuclear Research of NASU, 03028 Kyiv, Ukraine

Dark Matter candidate particles able to induce nuclear recoils can also be studied using the so-called directionality approach. In this case the correlation between the nuclear recoils direction and the Earth motion in the galactic frame is studied. Several experimental techniques to explore the directionality approach have been proposed. In this talk, a review of such experimental techniques will be addressed.

Keywords: Dark Matter; Directionality; Nuclear Recoil.

1. Introduction

Astrophysical observations have pointed out the presence of Dark Matter (DM) on all astrophysical scales and many arguments have been suggested that a large fraction of it should be in form of relic DM particles.

Currently, several approaches are exploited to investigate DM particles in the galactic halo; here, we will focus our attention on a particular direct strategy: the directionality approach.

In direct detection experiments, a model-independent signature is a powerful tool to provide a DM signal identification with respect to the background. In fact, as originally pointed out in Refs. 1, 2 and discussed in Refs. 3, 4, independently of the considered DM scenario, the events induced by a physical reaction in a target detector depend on the product between the cross-section of the process and the relative velocity between the incident DM particle (DMp) and the target. Such a product is expected to have a characteristic time behaviour in case of DM annual modulation: annual variation of the interaction rate due to Earth motion around the Sun which is moving in the Galaxy, successfully exploited by DAMA collaboration.[5–9]

Besides this main one, other possible signatures are expected: a diurnal modulation, due to the Earth revolution around its axis;[10] a daily variation of the interaction rate due to the different Earth depth crossed by the DMp[11] and the directionality signature, due to the correlation of DM impinging direction with Earth's galactic motion, in case of DMp able to induce nuclear recoils.[12] In fact, the dynamics of the rotation of the Milky Way galactic disc through the halo of DM causes the Earth to experience a wind of DM particles seemingly flowing along a direction opposite to that of the solar motion relative to the DM halo.

Therefore, the observation of an anisotropy in the distribution of nuclear recoil directions could give further evidence and information for such DM candidate particles and the related astrophysics scenario. Thus, a direction-sensitive detector is needed.

2. Dark matter signal in case of elastic DMp–nucleus scattering

Direct detection experiments, using the directionality approach, look for the signal of DMp via their elastic scattering interactions with detector nuclei. Therefore, due to kinematic consideration, the recoil nucleus direction is correlated with the direction of the impinging DMp (see Fig. 1). Assuming an elastic DMp–nucleus scattering (see Fig. 1), the nucleus kinetic energy after the interaction with the impinging DM particle is:

$$E_R = \frac{4M_A M_{DM}}{(M_{DM} + M_A)^2} \cos^2(\theta)\, E_{DM} \tag{1}$$

where, M_A is the mass of target nucleus, M_{DM} and E_{DM} are the incoming DMp mass and kinetic energy (laboratory system), respectively, E_R is the recoil nucleus kinetic energy (laboratory frame) and θ is the scattering angle of the recoil nucleus in the lab frame. According to the equation (1), depending on the value of $M_A \sim 1{-}200$ GeV, considering a very simple isothermal halo model with $v \sim 300$ km/s $\cong 10^{-3}c$, thus with $E_{DM} \sim 0.5 \times 10^{-6}\, M_{DM}$, the maximum E_R can be $\ll 200$ keV up to $\sim$ 200 keV (without considering the quenching effect in the direct detection of

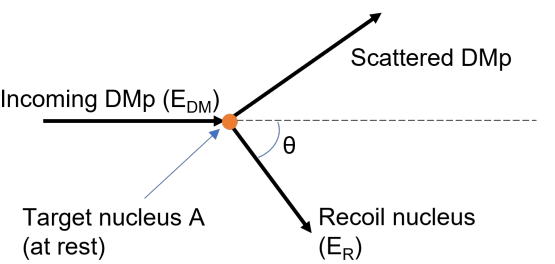


Fig. 1. DMp–nucleus elastic scattering diagrams for the laboratory coordinate systems.

the nuclear recoiling energy). Considering this information, the maximum range of nuclear recoiling is of the order of mm in low pressure gas medium, μm in liquid and less in solid medium.[13–15]

Thus, Low-Pressure Time Projection Chamber looks advantaged, but some drawbacks are present and will be discussed later.

To evaluate the direction of DMp in the laboratory frame, we can determine the detector velocity in the galactic rest frame, $\overrightarrow{v_d}(t)$, which varies with time. To study $\overrightarrow{v_d}(t)$ it is possible to consider various coordinate frames; in particular, following the formalism in,[12] we consider the horizontal coordinate frame located at the North pole, described by the "polar–zenith", $\Theta_z(t)$, and by the "polar–azimuth", $\phi_a(t)$. In this frame the area described in the sky by the direction of $\overrightarrow{v_d}(t)$ is shown in Fig. 2 (green band). Considering an experiment performed at a latitude similar to, e.g., the latitude of the Gran Sasso National Laboratory (LNGS) of I.N.F.N. ($42°27'12''$ N latitude and $13°34'26''$ E longitude), we can assume that Θ_z is always near 40°. In fact, at a certain time of the day ($\approx$ 21:00 h LST) the DMp come mainly from the top, while 12 h later they come near the horizon and from North (see Fig. 2). Thus, assuming $\overrightarrow{v}$ to be the velocity of the DMp in the laboratory frame, the DMp velocity in the galactic rest frame is $\overrightarrow{v} + \overrightarrow{v_d}(t)$; finally we can indicate with $f(\overrightarrow{v} + \overrightarrow{v_d}(t))$ the DM velocity distribution in the galactic rest frame (having assumed a specific DM halo model) and $\frac{\xi\rho_0}{M_{DM}} v f(\overrightarrow{v} + \overrightarrow{v_d}(t))$ is the flux of the considered DMp candidate, where ξ is the fraction of the local DM halo density (ρ_0) of the considered candidate; M_{DM} is the mass of the DM candidate particle.

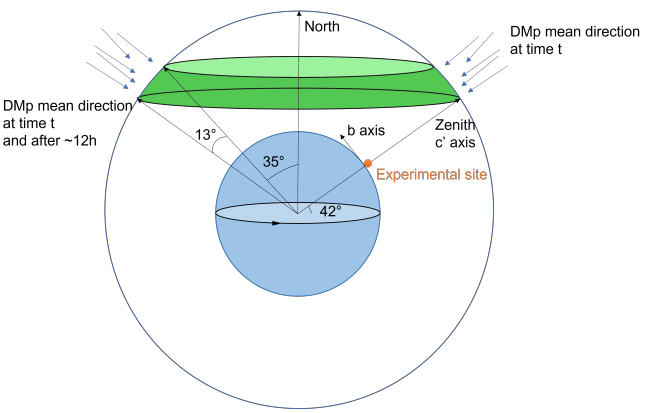


Fig. 2. Schematic representation of the experimental approach mentioned in the text. The detector is considered as placed at the Gran Sasso National Laboratory (LNGS) with the axis in the vertical direction and another axis is pointing to the North. The area in the sky from which the DM particles are preferentially expected is highlighted.

The expected counting rate of the signal, in the window (E_1,E_2) of the detected energy spectrum, as a function of the time t is:

$$R(E_1, E_2, t) = \int d^3\overrightarrow{v} \int d\Omega_{cm} \frac{\xi \rho_0 N_A}{M_{DM}} v f(\overrightarrow{v} + \overrightarrow{v_d}(t)) \frac{d\sigma_A}{d\Omega_{cm}} + \frac{1}{2} erf\left(\frac{q_A(\Omega_{out})E_R - E_1}{\sqrt{2}\Delta}\right) - \frac{1}{2} erf\left(\frac{q_A(\Omega_{out})E_R - E_2}{\sqrt{2}\Delta}\right)$$

where N_A is the number of target–nuclei of the A–th species per mass unit; Ω_{cm} is the nuclear recoil direction in the center of mass (c.m.) frame; $\frac{d\sigma_A}{d\Omega_{cm}}$ is the differential cross section in the c.m. frame, which is assumed to be isotropic; $q_A(\Omega_{out})$ is the quenching factor (Q.F.) which can depend on the direction of the nuclear recoil in the laboratory frame; and Δ is the detector energy resolution.

3. Detectors attempts

Current directional detectors are at the research and development stage. The experimental strategies are classifiable in two main groups: tracking detectors and detectors using anisotropic features. The first one bases the detection principle on the possibility of evaluating the direction of the nuclear recoils; the second one, instead, uses an anisotropic response of the detector with respect to the direction of the nuclear recoil strongly correlated with the impinging direction of the DMp.

Experiments based on Low-Pressure TPC aims to develop a TPC for directional DM searches and solar neutrino spectroscopy. Some of big drawbacks are the energy threshold, angular and spatial resolutions. Moreover, the total mass is often modest (in the literature are present liquid TPCs very big as ICARUS, DUNE, etc., but they are not applicable to DM sector). Another of the most critical issues is the low angular resolution at very low energy. Other important drawbacks are: low background techniques are not mature; some systematic effects with respect to the direction of the electric field are present in case of high charge gain; the data analysis is based on a strong software gamma rejection, introducing systematics.

The NEWSdm collaboration uses a nuclear emulsion-based detector acting both as a target and as a solid tracking device,[16] declaring a spatial threshold of (120 ± 5) nm. The radioactive contaminants of nuclear emulsion are mainly dominated by contamination of ^{14}C and U/Th chains. The main drawback is to reach enough spatial and angle resolution to be sensible to a range of nuclear recoiling, that at maximum is a few 100 nm, without compromising the emulsion path and minimizing the systematics during the data analysis and with a reasonable duty cycle.

We also mention two other ideas. A proposed method consists in identifying the direction of an incident DMp using the spectroscopy of quantum defects in macroscopic solid-state crystals as diamonds.[17] Instead, in Ref. 18 an odd type of DM detector made of DNA (or RNA) is proposed.

The second main group of detectors is based on the anisotropic response, showing a signal that varies with the direction of impinging particles. In particular, in literature, there are attempts to use (i) carbon nanotubes, (ii) columnar recombination in a TPC dual-phase, (iii) a realistic and very promising technique using anisotropic crystal scintillators.

The first one is an idea based on the use of a large number of carbon nanotubes vertically aligned (VA-CNT). There are weak indications that an electron is extracted from this VA-CNT forrest with a higher probability if the impinging direction of a particle is parallel to the VA-CNT, as compared to the impinging direction perpendicular to the VA-CNT. This possible effect could be implemented to develop an anisotropic sensor.[19]

The second case exploits a dual-phase time projection chamber (TPC) is filled with a liquid noble gas. In those devices, the scintillation light (S1) in the liquid phase and the ionization charge signals (generated in the liquid phase), which is drifted to the gas phase and there converted into a proportional scintillation signal (S2), are measured. Along the drift path, the electrons, due to several reasons, are partially recombined with the medium. It is expected that in case the initial electrons are generated along a direction perpendicular to the electric field (due to the direction of the nuclear recoil scattered from the impinging DMp), the columnar recombination could be less than in case the electrons are generated along the direction of the electron field. If detectable, the S2 signal, in the first case, is expected higher than in the second case as supposed in Ref. 20. The only study, using a very small LAr-TPC (SCENE detector), is published in Ref. 21, with a very poor indication of this phenomenon (see Fig. 21 of Ref. 21). For more information, a synoptic survey on liquid noble gases for DM searches is reported in Ref. 22, where the very low energy application of double phase noble gas detectors is analysed and compared; the main technical aspects of the existing experimental applications have been discussed and some implications have been outlined.

All these practical limitations and problems about these ideas can be overcome using the third approach, i.e. anisotropic scintillation detectors in a suitable low background experimental set-up located deep underground.

4. Anisotropic crystal scintillator

In anisotropic scintillators, the detector response to heavy particles depends on their impinging direction with respect to crystal axes. Therefore, in the case of nuclear recoils induced by DMp, the measured energy spectrum at low energy is expected to change its shape during the sidereal day as a consequence of the changing orientation of the crystal axes with respect to the DM wind. This effect can be pointed out as a peculiar variation of the counting rate measured in a given low energy window and offers the possibility to highlight the DM signal with respect to the electromagnetic background (for details, see, e.g., in Refs. 3, 4, 12).

The use of anisotropic scintillators to study the directionality signature was proposed for the first time in Ref. 3 and revisited in Ref. 4, where the case of anthracene detector was preliminarily analysed and several practical difficulties in the feasibility of such an experiment were underlined. Nevertheless, the authors suggested that competitive sensitivities could be reached with new devoted R&D's of anisotropic scintillators having significantly larger size, higher light response, better stiffness, higher atomic weights and anisotropic features similar as – or even better than – those of the anthracene scintillator.

4.1. *$ZnWO_4$ anisotropic scintillator*

The potential offered by the anisotropic properties of $ZnWO_4$ crystal scintillator in the search for DM with the directionality approach was indicated for the first time in Ref. 23. Recent measurements and R&D show that the $ZnWO_4$ detector can offer suitable features for the purpose.[12,24–27] In the light of this, the Anisotropic detectors for DArk Matter Observation (ADAMO) project[12,28] was considered and R&Ds have been progressed.

In particular, in the last years various $ZnWO_4$ detectors have been developed in the framework of the collaboration between the DAMA group of INFN (Istituto Nazionale di Fisica Nucleare) and the INR-Kyiv group.[26–36] Formerly, some crystals were produced by the Institute for Scintillation Materials (ISMA, Kharkiv, Ukraine) and, later on, a collaboration with the Nikolaev Institute of Inorganic Chemistry (Novosibirsk, Russia) has started and an R&D to produce ultra-radiopure $ZnWO_4$ by using the low-thermal gradient Czochralski technique in a platinum crucible is ongoing.[37] The produced crystals have been put in measurement in the DAMA/R&D underground facility at the Gran Sasso laboratory (LNGS) of INFN.[26,27,35,36] Measurements and R&D works have shown the competitiveness of $ZnWO_4$ scintillators for a DM experiment based on the directionality. In fact, the light output and the time profile of the scintillation pulse for heavy particles (p, α, nuclear recoils) depends on the direction of such particles with respect to the crystal axes while no difference is observed for γ/β radiation.[23] The shape of the scintillation pulse is also different for $\gamma(\beta)$ radiation and heavy particles (α). This pulse shape discrimination capability can potentially be of interest not only for a DM experiment but also for double beta decay searches. The $ZnWO_4$ also offers a high atomic weight and the possibility to realize single crystals with masses of a few kg.[37] Moreover, the presence of three target nuclei with very different masses (Zn, W and O) makes these scintillators sensitive to both small and large mass DM candidates, as also the NaI(Tl) is. The recently developed $ZnWO_4$ scintillators have a very high level of radiopurity. The upper limits for ^{228}Th are < 0.17 μBq/kg – < 1.3 μBq/kg, for ^{40}K are < 20 μBq/kg and for ^{226}Ra are < 2 μBq/kg.[24,26] A further radio-purification of $ZnWO_4$ crystal scintillators is feasible. The R&D is still ongoing[25,38,39] and very good results have also been obtained in improving the light transmission. The next step will be to suitably increase the size of crystals.

As confirmed by the measurements performed at LNGS, the crystals have relatively high light output at room temperature, being about 20% of the NaI(Tl) scintillator. It appears feasible to further improve the light output of the crystal considering that it can increase when working at low temperature.[40] To study this feature a small cryostat is currently under test at LNGS to realize and optimize the cooling system; this system will allow to reach a stable working temperature around -50°C.

In order to employ the $ZnWO_4$ for DM investigation with the directionality approach it is crucial to measure and quantify the anisotropy of the scintillator for nuclear recoils. For this purpose a campaign of measurements was performed by using a 7.99 g mass $ZnWO_4$ crystal scintillator of $(10 \times 10 \times 10.4)$ mm^3, in the framework of the ADAMO project. The crystal has been obtained by a second crystallization procedure using the low-thermal gradient Czochralski technique from zinc tungstate crystals made from tungsten oxide additionally purified by double sublimation of tungsten chlorides.[41] The crystallographic axes were identified by the producer and experimentally verified.

Before measuring the response of the $ZnWO_4$ crystal with a monochromatic neutron source at ENEA-CASACCIA, the crystal was irradiated with α particles at LNGS. In the following the obtained results are briefly reported.

4.2. *Measurements with α particles*

The small $ZnWO_4$ crystal was coupled to a HAMAMATSU H11934-200 PMT (Ultra Bialkali photocathode with an effective area of (23×23) mm^2 and quantum efficiency $\simeq 43\%$ at 400 nm and $\simeq 25\%$ at 500 nm). The scintillation pulses coming from the detector were recorded by a LeCroy WaveSurf24X-sA oscilloscope (4 chn, 2.5 GSamples/s, 200 MHz) in a time window of 100 μs. The measurements were performed by using an ^{241}Am source and various sets of thin mylar films as absorbers to decrease the α energy. The beam of α particles was collimated before reaching the crystal face. The energies of α particles were measured with a CANBERRA Alpha Spectrometer (model 7401VR). The energy scale of the crystal for each measurement was calibrated by using ^{137}Cs and ^{22}Na γ sources.

The typical energy distributions of the α particles impinging along the three axes of the crystal are shown in Fig. 3-left. The $ZnWO_4$ crystal was irradiated in the directions perpendicular to the (100), (001) and (010) crystal planes: hereafter crystal axes I (blue on-line), II (green on-line) and III (red on-line), respectively in Fig. 3.

In the Fig. 3-right the dependence of the α/β ratio as a function of the energy for the three different directions of the α beam relatively to the crystal axes is shown. In particular, the quenching factor for α particles measured along the crystal axis III is about 1.2 times larger than that measured along the crystal axes I and II. Instead, the quenching factors measured along the crystal axes I and II are quite similar. The error bars are mainly due to the uncertainty of the alpha energy which is slightly degraded in air.

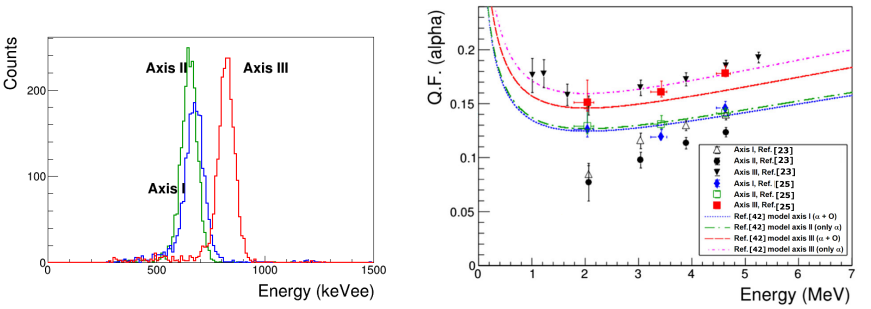


Fig. 3. Left: energy spectra of 4.63 MeV α particles impinging along the three axes of the crystal. Right: dependence of the α/β ratio on the energy of the α particles measured with a $ZnWO_4$ scintillator in Ref. 23 (black points) compared with those reported in Ref. 25 (colored points). The anisotropic behavior of the crystal is evident. The models for each crystal axis, obtained as global fits on all the data (from α's and from recoils; see also later) of Ref. 25 following the prescription of Ref. 42, is also reported.

The quenching factors and the anisotropic effect reported in Ref. 25 are in reasonable agreement with those of Ref. 23, as shown in Fig. 3-right; in the figure the behavior of the α/β ratio as expected for each crystal axis in the model of Ref. 42 are also reported.

Therefore, the data confirm the anisotropic features of $ZnWO_4$ crystal scintillator in case of α particles.

4.3. *Measurements with neutrons*

Monochromatic neutrons were generated by the Thermo Scientific portable generator MP 320. Neutrons are produced in the $d(t,\alpha)n$ reaction with energy around 14.7 MeV by accelerating deuterons toward a tritium target in electric potential. For the requirements of the experiment a configuration with beam acceleration voltage of 60 kV and a beam current of 40 μA was adopted in order to optimize the neutron yield and the stability of the beam operation; the production rate was around 10^7 n/s. In the experimental setup, neutrons leaving the target at 90 degrees to the forward direction were used; at this angle, simple kinematics predicts a value $E_n = 14.05$ MeV for a beam acceleration voltage of 60 kV.

The scattered neutron in the crystal is detected by two neutron detectors placed at a given scattering angle. In such a configuration the energy and the direction of the recoiling nucleus are fixed and, by measuring the energy released in the $ZnWO_4$ detector – in keV electron equivalent (keVee) – it is possible to determine the quenching factor. By changing the crystal axes orientation the quenching factor for the different axes can be measured.

A scheme of the set-up is shown in Fig. 4. The $ZnWO_4$ detector is placed in front of the neutron channel on a revolving platform around a vertical axis (O axis), which allows us to fix the direction of the crystal axis with respect to the impinging neutrons: the α angle as depicted in Fig. 4. The $ZnWO_4$ crystal is centered on

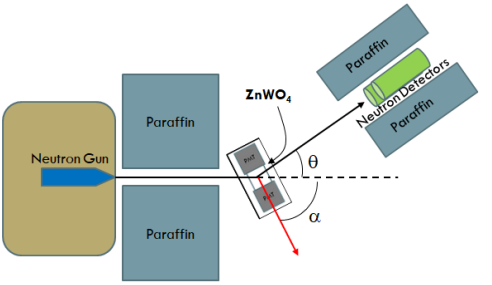


Fig. 4. Schematic view from the top of the experimental set-up. The two neutron detectors are one above the other.

the O axis and is optically coupled to two HAMAMATSU H11934-200 PMTs on opposite faces. In the set-up, the PMTs are on the faces perpendicular to the crystal axis III; thus, α identifies the angle between the crystal axis III and the impinging neutrons direction. The crystal axis I is also on the horizontal plane, while the axis II is vertical. The neutrons scattered off the $ZnWO_4$ crystal are tagged by two neutron detectors by Scionix employing EJ-309 liquid scintillator. The EJ-309 scintillator has a very high capability to discriminate neutrons interactions from the gamma background by Pulse Shape Discrimination (PSD). The neutron detectors are placed 82 cm far from O axis, one above the other to maintain the same neutron scattering angle θ (see Fig. 4) and to improve the solid angle acceptance. They are held by an arm and free to rotate around the O axis.

The signals from the two PMTs coupled to the $ZnWO_4$ detector are summed and recorded by a CAEN DT5720 transient digitizer with 250 MSamples/s. The trigger is obtained by the coincidence between a signal in the $ZnWO_4$ detector and in the EJ-309 detectors within ± 500 ns.

The energy detected in the $ZnWO_4$ detector is evaluated from the digitized pulse area. ^{133}Ba and ^{137}Cs sources were used to calibrate the detector energy scale. The typical energy resolution was $\sigma/E = 4.4\%$ at the 662 keV γ peak of ^{137}Cs. The energy calibration was performed before and after the neutrons irradiation.

Neutron events in the EJ-309 detectors were selected by a PSD data analysis based on: i) head/tail analysis; ii) analysis of the mean time, τ, of the time profile of the pulse. A typical example of the separation between gamma and neutrons exploiting the two techniques is reported in Fig. 5-left and in Fig. 5-right, respectively.

To select the nuclear recoils induced by the elastic scattering of neutrons over the background an important quantity is the TOF between the $ZnWO_4$ detector and the neutron detectors EJ-309. The expectation value of TOF due to the typical velocity of the scattered neutrons is around 15 ns. It can be defined the variable $\Delta t = t_0^{EJ} - t_0^{ZWO}$, where t_0^{EJ} (t_0^{ZWO}) is the starting time of the EJ-309 ($ZnWO_4$) pulse. The transit time of the neutron detectors PMT is ≈ 40 ns while it is ≈ 6 ns for the HAMAMATSU H11934-200. The Δt variable is shifted with respect to TOF

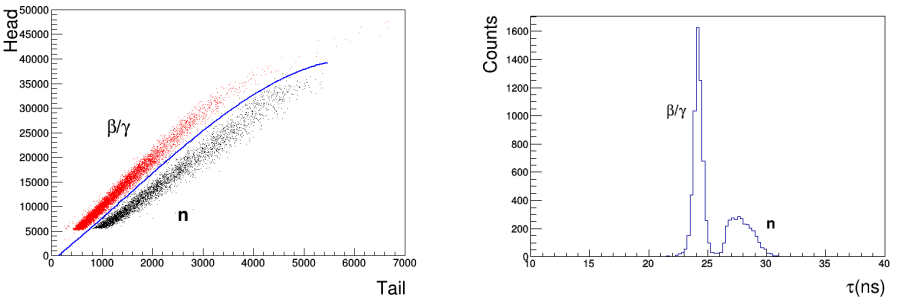


Fig. 5. Example of gamma/neutron separation by PSD in EJ-309 liquid scintillator. Left: head/tail analysis. Right: distribution of the mean time, τ, variable.

and it has been calibrated considering the coincidences between high energy events in the $ZnWO_4$ detector and gamma events in the neutron detectors. A $\Delta t \approx 34$ ns has been obtained from the data in agreement with the expectation. This TOF calibration has been taken into account in the analysis.

For a fixed scattering angle θ and crystal axis, the event pattern searched for was represented by a scintillation pulse in the $ZnWO_4$ detector in coincidence with a neutron in the EJ-309 detectors.

In Fig. 6-left the TOF distribution is depicted for the case of $\theta = 70°$ and crystal axis I, while in Fig. 6-right the bi-dimensional plot TOF vs $ZnWO_4$ energy is presented (E_{ZWO} is in keVee). These plots show a continuum due to random coincidences, a clear excess of events with energy around 80 keVee and TOF in agreement with the expectation for scattered neutrons on target nuclei. The peak in the TOF variable shows a tail on the left due to the first photoelectron delay in $ZnWO_4$ (effective average scintillation decay time ≈ 24 μs[23]). The observed excess can be ascribed to the O recoils in the $ZnWO_4$ detector and its position provides the quenching factor of this nucleus for the used scattering angle and crystal axis. The Zn and W recoils are expected to be well below the energy threshold that has been used in the present experimental conditions.

The behavior of the TOF distribution can be explained by considering that events belong to the sum of two main contributions: 1) a peak due to the coincidences caused by the elastic scatterings of neutrons off the $ZnWO_4$ nuclei, and 2) the random flat coincidences (f_{rnd}). To build a model for the TOF behavior some assumptions can be considered: i) the signal of the neutron detectors is "prompt" (that is the time delay between the neutron interaction and the starting time of the EJ-309 pulse is mostly equal to the PMT transit time); ii) the $ZnWO_4$ time delay between the neutron interaction and the starting time of the pulse is given by the PMT transit time plus the delay of the first photoelectron in $ZnWO_4$ detector; iii) the probability to have the first photoelectron in $ZnWO_4$ at the time t is $\frac{1}{\tau}e^{-t/\tau}$, where τ is the characteristic time decay of the process; iv) the fluctuations of the

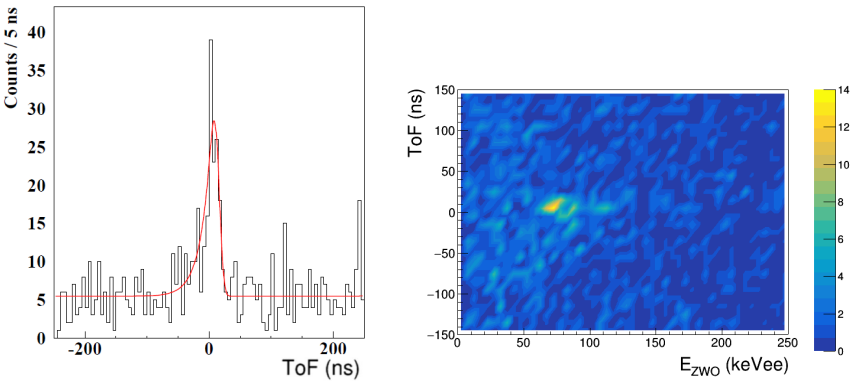


Fig. 6. TOF distribution and bi-dimensional plot TOF vs $ZnWO_4$ energy for coincidences obtained after selecting neutrons in the neutron detectors, for the case of scattering angle $\theta = 70°$ and axis I. Left: The TOF distribution shows a continuum due to random coincidences and a clear peak, that is in agreement with the expected TOF for neutrons after elastic scattering off the nuclei of $ZnWO_4$ detector. The peak shows a tail on the left part due to the first photoelectron delay in $ZnWO_4$. The fit with the function of eq. 3 is superimposed (See text). Right: A clear peak is present at the proper value of TOF and at the given energy. The peak position in the E_{ZWO} distribution provides the quenching factor of the O nucleus for the used scattering angle and crystal axis.

transit times of the PMTs of the $ZnWO_4$ and of the EJ-309 detectors, and the time resolution of the used digitizer (2 ns bin size for the neutron detectors and 4 ns for the $ZnWO_4$ detector) are taken into account by making a convolution of the time distribution with a Gaussian function having a characteristic spread σ. Thus, the model function of TOF can be written as:

$$f(TOF|f_{rnd}, A, \tau, \sigma, TOF_0) = f_{rnd} + A \int_0^\infty e^{-x/\tau} e^{-\frac{(x+TOF-TOF_0)^2}{2\sigma^2}} dx \qquad (2)$$

where TOF_0 represents the expectation value of the neutrons time of flight and A is a normalization factor. By solving the integral one gets:

$$f(TOF) = f_{rnd} + Be^{-\frac{TOF_0-TOF}{\tau}} \left[1 - erf\left(-\frac{TOF_0 - TOF}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\tau}\right)\right] \qquad (3)$$

where B includes only constant factors: $B = A\sqrt{\frac{\pi}{2}}\sigma e^{\frac{\sigma^2}{2\tau^2}}$. The result of the fit of the TOF distribution with the function given in eq. 3 is plotted in Fig. 6-left; it gives: $f_{rnd} \approx 5.5$ counts/ 5 ns, $B \approx 21.6$ counts/ 5 ns, $\tau \approx 16.9$ ns, $\sigma \approx 5.8$ ns and $TOF_0 \approx 15.4$ ns. The model reproduce well the experimental data and is in a good agreement with the expectations. A Monte Carlo simulation gives similar results.

For the estimation of the quenching factors three scattering angles were considered. Since the responses of the crystal axes I and II are rather similar, only the

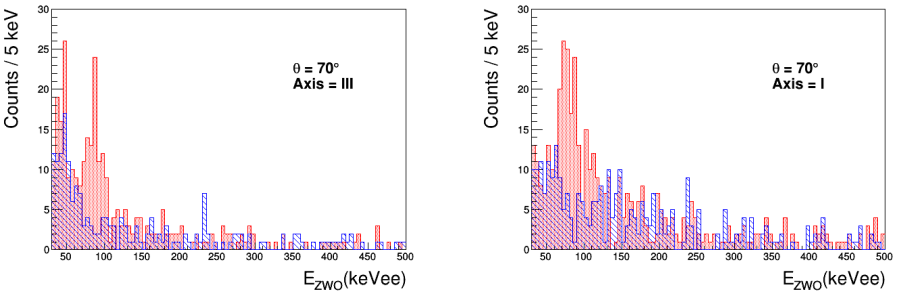


Fig. 7. Energy distributions of the detected energy, E_{ZWO}, in the $ZnWO_4$ detector, measured in keV electron equivalent (keVee) for the case of $\theta = 70°$ and axes III and I. Only events identified as recoils due to neutrons scattering, with subsequent registration of the scattered neutron in the neutrons detectors, are selected. Light (red on-line) histogram: events selected in the proper window of the TOF variable (-20 ns $< TOF <$ 30 ns). Dashed (blue on-line) histogram: events selected in the off-window (60 ns $< TOF <$ 110 ns).

case of axis I was considered. Figure 7 shows an example of the energy distributions of the detected energy, E_{ZWO}, in the $ZnWO_4$ detector, measured in keV electron equivalent ($\theta = 70°$ and axes III and I). Two distributions are reported for each plot: one is obtained by selecting events in the time window expected for the TOF of neutrons elastically scattered off the nuclei of $ZnWO_4$ detector, -20 ns $< TOF <$ 30 ns (light red histogram); the other is presented by selecting off-window events (random coincidences), 60 ns $< TOF <$ 110 ns (blue histogram). The off-window events are related to random coincidences and, therefore, their distribution is the background distribution in the histogram of the in-window events. The peaks are evident and can be ascribed to the oxygen nuclear recoils. The positions of the peaks are obtained by fitting them with a Gaussian curve plus an exponential function that simulates the background of the random coincidences. The energy resolutions, σ, of the peaks are between 8 and 12 keVee; the values are well in agreement with the energy dependence: $\sigma \propto \sqrt{E}$.

In Table 1 the summary of the peak positions measured for the three considered angles is reported. The expected recoil energies for the oxygen nucleus, $E_{R,O}$, are also given. The obtained quenching factors are calculated as the ratio $E_{ZWO}/E_{R,O}$ and the last column reports the degree of anisotropy for each scattering-angle/recoil energy.

In Fig. 8 the obtained values for the quenching factors are plotted together with the models for the considered crystal axes derived from Ref. 42 for a single kB parameter. This parameter has been estimated, for each axis, taking into account both the response of $ZnWO_4$ to α particles and to oxygen recoils ($kB = 12.06$ mg MeV^{-1} cm^{-2} for axis I, $kB = 11.80$ mg MeV^{-1} cm^{-2} for axis II – only α data are considered – and $kB = 9.98$ mg MeV^{-1} cm^{-2} for axis III, respectively).

Table 1. Summary table of the peak position due to oxygen nuclear recoils. For each scattering angle, θ, and for the different axes of the $ZnWO_4$ crystal, the peak position, E_{ZWO}, the energy resolution, σ, the expected recoil energies for the oxygen nucleus, $E_{R,O}$, the quenching factors, Q, and the anisotropy, Q_{III}/Q_I ratio, are reported.

Scattering angle, θ	Crystal axis	E_{ZWO} (keVee)	σ (keVee)	$E_{R,O}$ (keV)	Quenching factor, Q	Q_{III}/Q_I
80°	III	99.3 ± 2.5	9	1402	0.0708 ± 0.0018	1.174 ± 0.051
	I	84.5 ± 2.9	12		0.0603 ± 0.0021	
70°	III	86.5 ± 2.0	7	1128	0.0767 ± 0.0018	1.121 ± 0.038
	I	77.2 ± 1.9	10		0.0684 ± 0.0017	
60°	III	75.4 ± 1.8	9	866	0.0871 ± 0.0021	1.166 ± 0.059
	I	64.7 ± 2.9	10		0.0747 ± 0.0033	

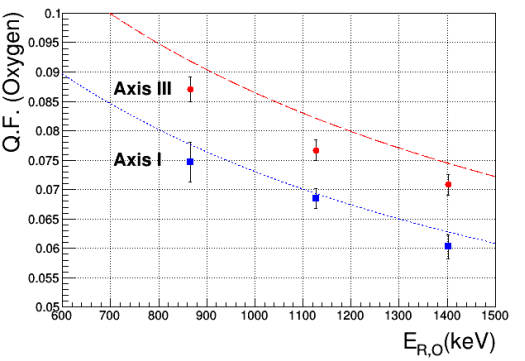


Fig. 8. Quenching factors for oxygen nuclear recoils in $ZnWO_4$ for the crystal axes I and III as function of the expected recoil energies $E_{R,O}$. In the plot the expected behavior of quenching factor for the considered crystal axes is also shown; it has been obtained by the global fits on the α and oxygen recoil data, following the prescription of Ref. 42 (for more details see Ref. 25).

5. Conclusion

The directionality investigates the correlation between the direction of the nuclear recoils induced by DMp and the motion of the Earth in the galactic rest frame. In fact, due to the solar motion in the DM halo, a wind of DMp is expected on the Earth with an seemingly opposite direction with respect to the Earth velocity in the galactic rest frame. As a consequence of the Earth's rotation, an observer on the Earth's surface experiences a change in the DM average arrival direction during a sidereal day. The induced nuclear recoils' directions are expected to be strongly correlated with that of the DM particles. Thus, the study of the nuclear recoils directions can also be pursued to investigate the class of DM candidate particles interacting via elastic-scattering on target nuclei.

A brief review of the R&Ds present in the literature has been addressed. In particular, we have briefly discussed the case of Low-Pressure TPC, the emulsion-based detector, columnar recombination, and some other ideas like the use of DNA based detector, diamonds or VA-CNT. However, such techniques suffer of several big problems, such as the short range of nuclear recoil, high radioactivity level, stability during the running condition, mass limitation etc., making this measurement really difficult. On the other hand, the anisotropic crystal scintillators offer a unique possibility of overcoming all these problems and giving a realistic technique to exploit the directionality approach (see Refs. 12, 25 for details). In fact, the results confirm the anisotropic response of the $ZnWO_4$ crystal scintillator to α particles and to oxygen nuclear recoils. The presence of a good anisotropic response also in the lower energy region is supported by the trend of the measured quenching factors at lower energy. In particular, for nuclear recoils induced by neutrons, the anisotropic response of the $ZnWO_4$ crystal scintillator has been determined at 5.4 standard deviations. This opens a possibility to realize a pioneer experiment to investigate the mentioned DM candidates by means of the directionality with the help of $ZnWO_4$ anisotropic scintillators. The features and the potentiality of these detectors can achieve – in some of the many possible scenarios and for some DM candidates able to induce a nuclear recoil – sensitivities not far from that of the DAMA/LIBRA positive result:[5–7,43] a low background pioneer experiment named ADAMO (Anisotropic detectors for DArk Matter Observation) located deep underground.[12,28]

For the near future, the DAMA, INR–Kyiv and ENEA/Casaccia collaboration plans to further investigate the light response of $ZnWO_4$ concerning the operation temperature (to increase the light-yield eventually), the light response at very low energy in case of the electromagnetic excitation and the Q.F. for nuclear recoils of W and Zn and other scattered angles of oxygen nucleus.

References

1. A.K. Drukier, K. Freese, D.N. Spergel. *Detecting cold dark-matter candidates. Phys. Rev. D* **33**, 3495 (1986).
2. K. Freese, J.A. Frieman, A. Gould. *Signal modulation in cold-dark-matter detection. Phys. Rev. D* **37**, 3388 (1988).
3. P. Belli et al. Identifying a "dark matter" signal by nonisotropic scintillation detector. *Il Nuovo Cimento C* **15**, 473–479 (1992).
4. R. Bernabei et al. *Anisotropic scintillators for WIMP direct detection: revisited. Eur. Phys. J. C* **28**, 203-209 (2003).
5. R. Bernabei et al. *The DAMA project: Achievements, implications and perspectives. Prog. Particle Nucl. Phys.* **114**, 103810 (2020).
6. R. Bernabei et al. *DAMA/LIBRA–phase2 results and implications on several dark matter scenarios.* Int. J. Mod. Phys. A **35**(36), 2044023 (2020).
7. R. Bernabei et al. *New model-dependent analyses including DAMA/LIBRA-phase2. Nuovo Cim. C* **43**(2-3), 23 (2020).
8. R. Bernabei et al. *The Future Role of Inorganic Crystal Scintillators in Dark Matter Investigations. Instruments* **5(2)**, 16 (2021).

9. R. Belli et al. *The dark matter: DAMA/LIBRA and its perspectives.* To appear on the proceedings of Sixteenth Marcel Grossmann Meeting - MG16 (2021).
10. R. Bernabei et al. *Model independent result on possible diurnal effect in DAMA/LIBRA-phase1. Eur. Phys. J. C* **74**(3), 2827 (2014).
11. R. Bernabei, et al. *Investigating Earth shadowing effect with DAMA/LIBRA-phase1. Eur. Phys. J. C* **75**(5), 239 (2015).
12. F. Cappella et al. *On the potentiality of the $ZnWO_4$ anisotropic detectors to measure the directionality of Dark Matter. Eur. Phys. J. C* **73**, 2276 (2013).
13. B.G. Harvey. *Recoil techniques in nuclear reaction and fission studies. Ann. Rev. Nucl. Part. Sci.* **10**, 235-258 (1960).
14. M. Ismail. *Measurement of excitation functions and mean projected recoil ranges of nuclei in α-induced reactions on F, Al, V, Co and Re nuclei. Pramana - J. Phys.* **40**, 227 (1993).
15. L. Bryde et al. *Ranges of recoil ions from alpha-reactions. Mat. Fys. Medd. Dan. Vid. Selsk.* **33**, no. 8 (1962).
16. N. D'Ambrosio et al. *Directional detection of Dark Matter with nuclear emulsion based detector. PoS EPS-HEP2017* 059 (2017).
17. S. Rajendran et al. *A method for directional detection of dark matter using spectroscopy of crystal defects. Phys. Rev. D* **96**, 035009 (2017).
18. A. Drukier et al. *New Dark Matter Detectors using DNA or RNA for Nanometer Tracking. MCTP–12-14.* arXiv: 1206.6809 (2012).
19. G. Cavoto et al. *Carbon nanotubes as anisotropic target for dark matter. J. Phys.: Conf. Ser.* **1468**, 012232 (2020).
20. D. R. Nygren. *Columnar recombination: a tool for nuclear recoil directional sensitivity in a xenon-based direct detection WIMP search. J. Phys.: Conf. Ser.* **460**, 012006 (2013).
21. H. Cao et al. *Measurement of scintillation and ionization yield and scintillation pulse shape from nuclear recoils in liquid argon. Phys. Rev. D* **91**, 092007 (2015).
22. R. Bernabei et al. *Liquid noble gases for dark matter searches: An updated survey. Int. J. Mod. Phys. A* **30**(26), 1530053 (2015).
23. F.A. Danevich et al. *$ZnWO_4$ crystals as detectors for 2β decay and dark matter experiments. Nucl. Instrum. Meth. A* **544**, 553–564 (2005).
24. P. Belli et al. *New development of radiopure $ZnWO_4$ crystal scintillators. Nucl. Instrum. Meth. A* **935**, 89–94 (2019).
25. P. Belli et al. *Measurements of $ZnWO_4$ anisotropic response to nuclear recoils for the ADAMO project. Eur. Phys. J. A* **56**, 83 (2020).
26. P. Belli et al. *Radioactive contamination of $ZnWO_4$ crystal scintillators. Nucl. Instrum. Meth. A* **626-627**, 31 (2011).
27. P. Belli et al. *Final results of experiment to search for 2β processes in zinc and tungsten with the help of radiopure $ZnWO_4$ crystal scintillators. J. Phys. G: Nucl. Part. Phys.* **38**, 115107 (2011).
28. V. Caracciolo et al. *The ADAMO Project and developments. J. Phys.: Conf. Ser.* **718**, 042011 (2016).
29. R. Cerulli for DAMA/INR-Kyiv coll., talk at IPRD 2019, October 14-17 2019, Siena, Italy (http://people.roma2.infn.it/dama/pdf/cerulli_IPRD2019.pdf).
30. A. Di Marco for DAMA/INR-Kyiv coll., talk at CYGNUS 2019 workshop on directional DM detection, July 10-12 2019, Rome, Italy (http://people.roma2.infn.it/~dama/pdf/cygnus_2019.pdf).
31. V. Caracciolo for DAMA/INR-Kyiv coll., talk at Particle Physics at the Silver Jubilee of Lomonosov Conferences, 464 (2019).

32. R. Bernabei et al. *$ZnWO_4$ anisotropic scintillator for Dark Matter investigation with the directionality technique. EPJ Web of Conferences* **136**, 05002 (2017).
33. V. Caracciolo for DAMA/INR-Kyiv coll., talk at CYGNUS-TPC kick-off meeting: a mini-workshop on directional Dark Matter searches and coherent neutrino scattering, LNF, April 7 2016, Italy.
34. R. Bernabei et al. *Crystal scintillators for low background measurements. AIP Conference Proceedings* **1549**, 189 (2013).
35. P. Belli et al. *Search for 2β processes in ^{64}Zn with the help of $ZnWO_4$ crystal scintillator. Phys. Lett. B* **658**, 193 (2008).
36. P. Belli et al. *Search for double beta decay of zinc and tungsten with low background $ZnWO_4$ crystal scintillators. Nucl. Phys. A* **826**, 256 (2009).
37. E.N. Galashov et al. *Growing of $ZnWO_4$ single crystals from melt by the low thermal gradient Czochralski technique. Functional Materials* **16**, 65 (2009).
38. A.S. Barabash et al. *Low background scintillators to investigate rare processes. JINST* **15** 07, C07037 (2020).
39. P.Belli et al. *Developments and improvements of radiopure $ZnWO_4$ anisotropic scintillators. JINST* **15** 05, C05055 (2020).
40. L.L. Nagornaya et al. *Tungstate and molybdate scintillators to search for dark matter and double beta decay. IEEE Trans. Nucl. Sci.* **56**, 2513 (2009).
41. V.N. Shlegel et al. *Recent progress in oxide scintillation crystals development by low-thermal gradient Czochralski technique for particle physics experiments. JINST* **12**, C08011 (2017).
42. V.I. Tretyak, *Semi-empirical calculation of quenching factors for ions in scintillators. Astropart. Phys.* **33**, 40 (2010).
43. R. Bernabei et al. *First results from DAMA/LIBRA and the combined results with DAMA/NaI. Eur. Phys. J. C* **67**, 39–49 (2010).

Collapse models under test by high sensitivity γ-ray and X-ray measurements

Catalina Curceanu[1,2], Kristian Piscicchia[3,1], Massimiliano Bazzi[1], Mario Bragadireanu[2,1], Michael Cargnelli[4,1], Alberto Clozza[1], Luca De Paolis[1], Raffaele Del Grande[5,1], Carlo Guaraldo[1], Mihail Iliescu[1], Matthias Laubenstein[6], Johann Marton[4,1], Marco Miliucci[1], Fabrizio Napolitano[1], Alessio Porcelli[4,1], Alessandro Scordo[1], Francesco Sgarmella[1], Hexi Shi[4], Diana Laura Sirghi[1,2], Florin Sirghi[1,2], Oton Vazquez Doce[1] and Johann Zmeskal[4,1]

[1] *Laboratori Nazionali di Frascati, INFN, Italy*
[2] *IFIN-HH, Institutul National pentru Fizica si Inginerie Nucleara Horia Hulubei, Romania*
[3] *Centro Ricerche Enrico Fermi – Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", Italy*
[4] *Stefan-Meyer-Institute for subatomic physics, Austrian Academy of Science, Austria*
[5] *Excellence Cluster Universe, Technische Universität München, Germany*
[6] *Laboratori Nazionali del Gran Sasso, INFN, Italy*

The article reviews our recent experimental results on the Continuous Spontaneous Localization (CSL) model and on the gravity related collapse model developed by Diósi and Penrose (DP). These models of dynamical reduction of the wave function consist in non-linear and stochastic modifications of the Schröedinger equation, which lead to a progressive breakdown of the superposition principle, as the size of the system increases. We performed a high sensitivity survey of the spontaneous radiation phenomenon, predicted by the collapse models, in a dedicated experiment operated in the extremely low background of the Gran Sasso underground National Laboratory of INFN in Italy. Our studies set the strongest bounds on the CSL parameters, in a broad region of the parameters space, and rule out the DP in its present formulation.

Keywords: Collapse Models; Spontaneous Radiation; Germanium Detectors.

1. Dynamical collapse models and spontaneous radiation

Quantum Theory (QT) is the basis of our understanding of the physical world. Since its inception QT successfully described plenty of puzzling experimental phenomena, it explained the spectrum of the black body radiation and the atomic structure, it was the cornerstone of the development of modern chemistry, of nuclear physics and of quantum field theory, just to give some examples, and presently fuels the growth of vanguard technologies. Despite its success and the outstanding precision of the experimental validations, QT still contains a conundrum in its grounding pillars. Why the superposition principle, characterizing the evolution of microscopic systems, does not carry over to macroscopic objects? Why in the act of measuring the deterministic dynamics is replaced by a probabilistic behaviour governed by the Born rule?

Models of dynamical reduction of the wave function represent phenomenological, and testable, concrete solutions to the problem (see e.g.,[1–8] for a review and references see also[9]). They consist in non-linear and stochastic modifications of the

Schröedinger equation, which preserve the QT predictions in the microscopic regime, and go over to classical mechanics in the macroscopic limit, by breaking down the quantum linear evolution proportionally to the growing size of the system.

Besides interferometric experiments (see e.g.[10–14]), which aim to measure the interference pattern of the spatial superposition which is created in an interferometer, collapse models can be also probed with indirect tests (see e.g.[15–27]). Common denominator of this second class of experiments is to exploit the random motion associated to the collapse mechanism, which allows to test the effect of the models predictions on macroscopic objects. Clear advantage is the magnification effect, which leads to the stronger constraints on the collapse models, this is the case of micrometer cantilever,[16] gravitational wave detectors[28,29] and X/γ-ray measurements,[25–27] the latter being the subject of this paper.

We will review in this paper our latest experimental results on the Continuous Spontaneous Localization (CSL) and on the Diósi-Penrose (DP) models. In the CSL model[4–6] the non-linear and stochastic terms are characterized by the interaction with a continuous set of independent noises, with zero average, Gaussian correlation in space and, in the simplest version, white correlation in time. The model is defined in terms of two phenomenological parameters, denoted by λ and r_C. λ has the dimensions of a rate and sets the strength of the collapse. r_C is a correlation length which determines the spatial resolution of the collapse, the collapse is weak if the superposition size is much smaller than r_C, while becomes effective for delocalizations which are much larger than r_C. Different theoretical considerations lead to alternative choices for the parameters: Ghirardi, Rimini and Weber[3] proposed $\lambda = 10^{-17}\mathrm{s}^{-1}$ and $r_C = 10^{-7}\mathrm{m}$, Adler[30] proposed $\lambda = 10^{-8\pm2}\mathrm{s}^{-1}$ for $r_C = 10^{-7}\mathrm{m}$, and $\lambda = 10^{-6\pm2}\mathrm{s}^{-1}$ for $r_C = 10^{-6}\mathrm{m}$.

Roger Penrose argued[7,8] that when a system is found in a spatial quantum superposition, a corresponding superposition of two different space-times is generated. The superposition is unstable and decays in time. The more massive the system in the superposition, the larger the difference in the two space-times and the faster the wave-function collapse. The average collapse time τ would then be given by the expression $\tau \approx \hbar/E_g$, where $\hbar$ is the reduced Planck's constant and E_g is the gravitational self-energy of the difference between two (stationary) mass distributions of the superposition. Lajos Diósi developed a dynamical theory of gravity-related wave function collapse[1,2] which predicts the same form for the collapse time. Considered that the gravitational self-interaction energy diverges for point-like constituents, Diósi introduced[31] a minimum length R_0, which limits the spatial resolution of the mass density. E_g is then a function of R_0, the smaller R_0 the faster the collapse.

An unavoidable consequence of the dynamics of both the CSL and the gravity related collapse developed by Diósi, is that the non-linear interaction with the noise-filed induces a Brownian-like diffusion motion for the particles which, if charged, emits radiation. This phenomenon, which is not predicted in the context of QT, is usually called *spontaneous radiation*, and represents the observable which was investigated in our experimental surveys. The spontaneous radiation rate, due to

the emission of protons in the atomic nuclei, was calculated in Refs.[26,27] and is given by:

$$\left.\frac{d\Gamma}{dE}\right|_t = N_{atoms} \cdot N_A^2 \cdot \frac{\hbar\, e^2}{4\,\pi^2\,\epsilon_0\,c^3\,m_0^2} \cdot \frac{\lambda}{r_C^2}\frac{1}{E}\,, \tag{1}$$

for the CSL model, and by:

$$\left.\frac{d\Gamma}{dE}\right|_t = N_{atoms} \cdot N_A^2 \cdot \frac{2}{3}\frac{G\,e^2}{\pi^{3/2}\,\epsilon_0\,c^3} \cdot \frac{1}{R_0^3 E}\,, \tag{2}$$

for the DP model. N_{atoms} is the number of atoms in the system with atomic number N_A, c is the speed of light, ϵ_0 is the vacuum permittivity, m_0 is the nucleon mass, G is the universal gravitational constant, E and t are the energy and the time. Electrons are relativistic in the energy range which was considered in our analyses, hence their contribution to the spontaneous radiation emission can not be considered.

The paper is organized as follows: in Section 2 the experimental apparatus is described, in Section 3 is given a brief description of the statistical analysis and the constraints on the characteristic parameters of the CSL and DP models are summarized, while concluding remarks and the future developments of our studies are outlined in Section 4.

2. The experimental setup

The experimental apparatus was based on a coaxial p-type High Purity Germanium detector (HPEGe) surrounded by multiple shielding layers: the inner shielding consisting of 5cm thick electrolytic copper, the external part of lead (30 cm from the bottom and 25 cm from the sides). Both the shielding and the cryogenic system were enclosed in an air tight steel housing, flushed with boil-off nitrogen, in order to suppress radon contamination (see Fig. 2 in Ref.[26] for a schematic representation of the setup, further details on the setup structure can be found in Ref.[26] and therein references). The experiment was operated in the extremely low background environment of the Gran Sasso underground National Laboratory of INFN in Italy.

The goal of the measurement was to disentangle a faint contribution of the spontaneous radiation emission process, to the measured spectrum, from environmental background. To this aim an accurate characterization of the whole apparatus was performed with a validated Monte Carlo (MC) code[32] based on the GEANT4 software library.[33]

The measured spectrum (corresponding to an exposure of 124 kg $\cdot$ day) is shown in black in Fig. 1, in the range $\Delta E = (1000 \div 3800)$keV (this is a reproduction of the original Fig. 1 in Ref.[27]). The energy range fulfills the theoretical requirements for the validity of the calculated rates (Eqs. (1) and (2)). In ΔE the main contribution to the background was found to be originated by residual radionuclides, present

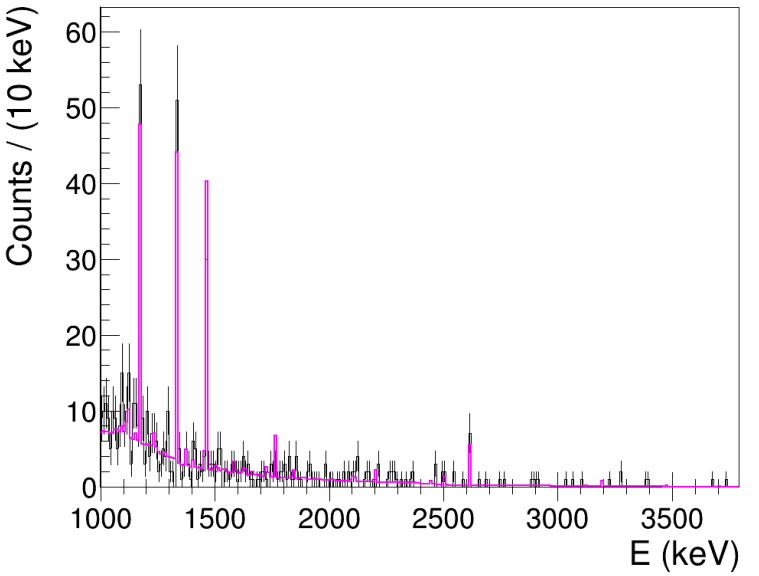


Fig. 1. The measured X-ray spectrum is shown in black in the selected energy range $\Delta E = (1000 - 3800)$ keV. The simulated background distribution is also shown in magenta.

in the materials of the setup, whose measured activities represented the inputs of the MC simulations. The magenta distribution in Fig. 1 represents the simulated background, 88% of the measured spectrum can be described in terms of known emission processes.

The simulation also allowed to compute the efficiencies, as a function of the energy, for the detection of spontaneously emitted photons in each component of the setup. To this end 10^8 photons, with uniform spatial distribution, were generated in each material in steps of 200 keV (i.e. 15 points in the ΔE). The efficiency functions $\epsilon_i(E)$ (i labelling the material of the detector) were then estimated by means of polynomial fits of the corresponding distributions, for each component of the detector which gives an appreciable contribution. The results of this analysis are summarized in Fig. 2 and Table 1 of Ref.[27].

3. Summary of the data analyses and results

The strategy of the Bayesian statistical analysis was to perform a comparison of the theoretically predicted spontaneous emission rate, generated by each component of the apparatus and weighted by the experimental efficiency and acceptance, with the measured distribution, accounting for the estimated background.

Given the calculated efficiency functions, and the theoretical rate (Eqs. (1) and (2)), the expected number of measured events, due to the spontaneous emission by

protons belonging to the i-th material, during the acquisition time T is:

$$\int_{\Delta E} \left.\frac{d\Gamma}{dE}\right|_t T\,\epsilon_i(E)\,dE. \tag{3}$$

Summation over i yields the total expected signal contribution, which is a function of the phenomenological parameters (λ, r_C) or R_0, depending on the selected collapse model. The measured counts were assumed to fluctuate according to a Poissonian distribution and the probability density functions for the expected number of counts were derived, from which the following constraints on the models parameters were calculated, corresponding to a probability of 0.95:

$$\frac{\lambda}{r_C^2} < 52\,\mathrm{s}^{-1}\mathrm{m}^{-2}\,; \qquad R_0 > 0.54\cdot 10^{-10}\mathrm{m}\,, \tag{4}$$

see Refs.[26,27] for more details concerning the data analysis.

The first limit in Eq. (4) represents the stronger existing bound on the CSL model for values of the correlation length $r_C \leq 10^{-6}$m. The limit on R_0 (right side in Eq. (4)) is about three orders of magnitude stronger than previous bounds in the literature.[29] If R_0 is chosen as the size of the nucleus's wave function (as suggested by Penrose), our result is to be compared with the square-root of the mean square displacement of a nucleus in the lattice. For the Germanium crystal, cooled down at the liquid Nitrogen temperature, this would amount to an R_0 value of about $0.05\cdot 10^{-10}$m, which is more than one order of magnitude less than the lower limit set by our experiment.

4. Conclusions and perspectives

We summarized the results of a dedicated measurement, performed at the Gran Sasso underground National Laboratory of INFN. The study was devoted to the search of spontaneous radiation emission, predicted by the CSL and the DP models of wave function collapse. We set the strongest bounds on the CSL phenomenological parameters in the region $r_C \leq 10^{-6}$m. Penrose's proposal for a gravity-related collapse of the wave function, in the present formulation, is ruled out. Our result indicates that the idea of gravity-related wave function collapse, which remains very appealing, will probably require a new approach. Indeed new theoretical developments are seeking for non-Markovian and/or dissipative versions of the collapse models, predicting a lower rate of spontanoeus radiation depending on the photon frequency. Both Penrose and Diósi are also pursuing the idea of a radiation free gravity-related collapse.

This pushes our efforts through further refinements our experimental techniques and data analyses methods. We are presently expanding our sensitive energy region from the MeV to few keV, with experimental setups based on Broad Energy Germanium detectors and ultra-radio pure targets. We are investigating the application of Machine Learning algorithms to identify the faint signal of the dynamical collapse, towards a deeper understanding of the foundations of Quantum Mechanics.

Acknowledgments

This publication was made possible through the support of Grant 62099 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. We acknowledge support from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation (Grants No. FQXi-RFP-CPW-2008 and FQXi-MGB-2011), and from the H2020 FET TEQ (Grant No. 766900) and INFN (VIP). We thank the Austrian Science Foundation (FWF) which supports the VIP2 project with the grants P25529-N20, project P 30635-N36 and W1252-N27 (doctoral college particles and interactions). K. P. acknowledges support from the Centro Ricerche Enrico Fermi - Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi" (Open Problems in Quantum Mechanics project).

References

1. L. Diosi, A universal master equation for the gravitational violation of quantum mechanics, *Physics letters A* **120**, 377 (1987).
2. L. Diósi, Models for universal reduction of macroscopic quantum fluctuations, *Physical Review A* **40**, p. 1165 (1989).
3. G. C. Ghirardi, A. Rimini and T. Weber, Unified dynamics for microscopic and macroscopic systems, *Physical review D* **34**, p. 470 (1986).
4. P. Pearle, Combining stochastic dynamical state-vector reduction with spontaneous localization, *Physical Review A* **39**, p. 2277 (1989).
5. G. C. Ghirardi, P. Pearle and A. Rimini, Markov processes in hilbert space and continuous spontaneous localization of systems of identical particles, *Physical Review A* **42**, p. 78 (1990).
6. P. Pearle and E. Squires, Bound state excitation, nucleon decay experiments and models of wave function collapse, *Physical Review Letters* **73**, p. 1 (1994).
7. R. Penrose, On gravity's role in quantum state reduction, *Gen. Rel. Grav. 28: 581-600* (1996).
8. R. Penrose, On the gravitization of quantum mechanics 1: Quantum state reduction, *Foundations of Physics 44, 557-575* (2014).
9. A. Bassi and G. Ghirardi, Dynamical reduction models, *Physics Reports* **379**, 257 (2003).
10. T. Kovachy, P. Asenbaum, C. Overstreet, C. Donnelly, S. Dickerson, A. Sugarbaker, J. Hogan and M. Kasevich, Quantum superposition at the half-metre scale, *Nature* **528**, 530 (2015).
11. S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor and J. Tüxen, Matter–wave interference of particles selected from a molecular library with masses exceeding 10000 amu, *Physical Chemistry Chemical Physics* **15**, 14696 (2013).
12. M. Toroš and A. Bassi, Bounds on quantum collapse models from matter-wave interferometry: calculational details, *Journal of Physics A: Mathematical and Theoretical* **51**, p. 115302 (2018).
13. K. C. Lee, M. R. Sprague, B. J. Sussman, J. Nunn, N. K. Langford, X.-M. Jin, T. Champion, P. Michelberger, K. F. Reim, D. England *et al.*, Entangling macroscopic diamonds at room temperature, *Science* **334**, 1253 (2011).

14. S. Belli, R. Bonsignori, G. D'Auria, L. Fant, M. Martini, S. Peirone, S. Donadi and A. Bassi, Entangling macroscopic diamonds at room temperature: Bounds on the continuous-spontaneous-localization parameters, *Physical Review A* **94**, p. 012108 (2016).
15. T. Kovachy, J. M. Hogan, A. Sugarbaker, S. M. Dickerson, C. A. Donnelly, C. Overstreet and M. A. Kasevich, Matter wave lensing to picokelvin temperatures, *Physical review letters* **114**, p. 143004 (2015).
16. A. Vinante, R. Mezzena, P. Falferi, M. Carlesso and A. Bassi, Improved noninterferometric test of collapse models using ultracold cantilevers, *Physical review letters* **119**, p. 110401 (2017).
17. O. Usenko, A. Vinante, G. Wijts and T. Oosterkamp, A superconducting quantum interference device based read-out of a subattonewton force sensor operating at millikelvin temperatures, *Applied Physics Letters* **98**, p. 133105 (2011).
18. A. Vinante, A. Collaboration *et al.*, Present performance and future upgrades of the auriga capacitive readout, *Classical and Quantum Gravity* **23**, p. S103 (2006).
19. B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Gw150914: The advanced ligo detectors in the era of first discoveries, *Physical review letters* **116**, p. 131103 (2016).
20. B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Observation of gravitational waves from a binary black hole merger, *Physical review letters* **116**, p. 061102 (2016).
21. M. Armano, H. Audley, G. Auger, J. Baird, M. Bassan, P. Binetruy, M. Born, D. Bortoluzzi, N. Brandt, M. Caleno *et al.*, Sub-femto-g free fall for space-based gravitational wave observatories: Lisa pathfinder results, *Physical review letters* **116**, p. 231101 (2016).
22. M. Armano, H. Audley, J. Baird, P. Binetruy, M. Born, D. Bortoluzzi, E. Castelli, A. Cavalleri, A. Cesarini, A. Cruise *et al.*, Beyond the required lisa free-fall performance: New lisa pathfinder results down to 20 μ hz, *Physical review letters* **120**, p. 061101 (2018).
23. S. L. Adler and A. Vinante, Bulk heating effects as tests for collapse models, *Physical Review A* **97**, p. 052119 (2018).
24. M. Bahrami, Testing collapse models by a thermometer, *Physical Review A* **97**, p. 052118 (2018).
25. K. Piscicchia, A. Bassi, C. Curceanu, R. D. Grande, S. Donadi, B. C. Hiesmayr and A. Pichler, CSL collapse model mapped with the spontaneous radiation, *Entropy* **19**, p. 319 (2017).
26. S. Donadi, K. Piscicchia, C. Curceanu, L. Diósi, M. Laubenstein and A. Bassi, Underground test of gravity-related wave function collapse, *Nature Physics* **17**, 74 (2021).
27. S. Donadi, K. Piscicchia, R. Del Grande, C. Curceanu, M. Laubenstein and A. Bassi, Novel CSL bounds from the noise-induced radiation emission from atoms, *The European Physical Journal C* **81**, 1 (2021).
28. M. Carlesso, A. Bassi, P. Falferi and A. Vinante, Experimental bounds on collapse models from gravitational wave detectors, *Physical Review D* **94**, p. 124036 (2016).
29. B. Helou, B. Slagmolen, D. E. McClelland and Y. Chen, Lisa pathfinder appreciably constrains collapse models, *Physical Review D* **95**, p. 084054 (2017).
30. S. L. Adler, Lower and upper bounds on csl parameters from latent image formation and IGM heating, *Journal of Physics A: Mathematical and Theoretical* **40**, p. 2935 (2007).
31. L. Diósi, Notes on certain newton gravity mechanisms of wave function localization and decoherence, *J. Phys. A 40, 2989-2995* (2007).

32. M. Boswell, Y.-D. Chan, J. A. Detwiler, P. Finnerty, R. Henning, V. M. Gehman, R. A. Johnson, D. V. Jordan, K. Kazkaz, M. Knapp *et al.*, MaGe-a Geant4-based Monte Carlo application framework for low-background germanium experiments, *IEEE Transactions on Nuclear Science* **58**, 1212 (2011).
33. S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand *et al.*, Geant4—A simulation toolkit, *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**, 250 (2003).

Leptophilic dark matter at linear colliders

P. S. Bhupal Dev

Department of Physics and McDonnell Center for the Space Sciences, Washington University, St. Louis, MO 63130, USA
E-mail: bdev@wustl.edu

Leptophilic dark matter (LDM) could naturally arise in many beyond the Standard Model scenarios and could address certain experimental anomalies. We discuss some model-independent collider constraints on the LDM effective couplings with the Standard Model sector, considering its production at a future electron-positron linear collider (with polarized and unpolarized beam options) in both mono-photon and mono-Z channels.[1]

Keywords: Dark Matter, Effective Field Theory, Lepton Collider

1. Introduction

Many of the existing experimental constraints on dark matter (DM) crucially rely on the DM interactions with nucleons, and therefore, can be largely weakened if the DM predominantly interacts with the Standard Model (SM) leptons, but not quarks at tree-level. Such *leptophilic* DM (LDM) could arise naturally in many beyond the Standard Model (BSM) scenarios,[2–23] some of which could even explain various experimental anomalies, such as the muon anomalous magnetic moment,[24] DAMA/LIBRA annual modulation,[25] anomalous cosmic ray positron excess,[26–29] the galactic center gamma-ray excess,[30] and XENON1T electron excess.[31] Dedicated searches for LDM in direct detection[32–34] and beam dump[35,36] experiments have also been discussed.

In this proceedings based on Ref.[1], we focus on the LDM searches at lepton colliders, which are complementary to the direct and indirect detection searches. We adopt an effective field theory (EFT) approach, which has been widely used in the context of collider searches for DM following the early works of Refs.[37–45] The same interactions responsible for DM pair-annihilation in the early universe leading to their thermal freeze-out guarantee their direct production at colliders, as long as kinematically allowed. This will give a characteristic mono-X signature, where the large missing transverse momentum carried away by the DM pair is balanced by a visible sector particle X (which can be either a photon, jet, W, Z, or Higgs, depending on the model) emitted from an initial, intermediate or final state (see Refs.[46,47] for reviews). Specifically, the mono-jet signature has become emblematic for LHC DM searches.[48–51] However, for an LDM with loop-suppressed interactions to the SM quarks, the hadron colliders like the LHC are not expected to provide a better limit than the existing constraints from indirect searches, such as from AMS-02,[52,53] at least within the EFT framework with contact interactions.

On the other hand, lepton colliders provide an ideal testing ground for the direct production of LDM and its subsequent detection via either mono-photon[42,45,54–63] or mono-Z[64–67] signatures. We go beyond the existing literature and perform a comprehensive and comparative study of both mono-photon and mono-Z signatures of LDM at future e^+e^- colliders in a model-independent, EFT approach.[1] Our analysis is generically applicable to all future e^+e^- colliders, such as the ILC,[68] CLIC,[69] CEPC[70] and FCC-ee,[71] but for concreteness, we have taken the $\sqrt{s} = 1$ TeV ILC as our case study for numerical simulations. We also assume the DM to be fermionic and limit ourselves to the dimension-6 operators, but taking into consideration all possible dimension-6 operators of scalar-pseudoscalar (S-P), vector-axialvector (V-A) and tensor-axialtensor (T-AT) type as applicable for the most general DM-electron coupling. Within the minimal EFT approach, the only relevant degrees of freedom in our analysis are the DM mass and an effective cut-off scale Λ which determines the strength of the four-Fermi operators. This enables us to derive model-independent ILC sensitivities on LDM in the (m_χ, Λ) plane in both mono-photon and mono-Z (leptonic and hadronic) channels, after taking into account all relevant backgrounds and systematic uncertainties. We consider both unpolarized and polarized beam options,[68,72] and find that with the proper choice of polarizations for the e^- and e^+ beams (which depends on the operator type), the DM sensitivities could be significantly enhanced.

2. Effective operators

Our primary assumptions are (i) the DM particle χ couples directly only to the SM leptons but not to the quarks (hence leptophilic), and (ii) the energy scale of the associated new physics is large compared to the collider energies under consideration, thus allowing us to integrate out the heavy mediators and parametrize the DM-SM interactions using effective higher-dimensional operators. For concreteness, we assume that the DM particles are Dirac fermions, and therefore, the leading order DM-SM interactions are the dimension-six four-Fermi interactions, with the most-general effective Lagrangian given by[37]

$$\mathcal{L}_{\text{eff}} = \frac{1}{\Lambda^2} \sum_j \left(\overline{\chi}\Gamma^j_\chi \chi\right) \left(\overline{\ell}\Gamma^j_\ell \ell\right) , \tag{1}$$

where Λ is the cut-off scale for the EFT description and the index j corresponds to different Lorentz structures, as shown below. Since our main focus is on e^+e^- colliders, we will just set $\ell = e$ in Eq. (1) and assume this to be the only leading-order coupling, but our discussion below could be easily extended to other cases, e.g. future muon colliders[73] by setting $\ell = \mu$.

A complete set of Lorentz-invariant operators consists of scalar (S), pseudoscalar (P), vector (V), axial-vector (A), tensor (T) and axial-tensor (AT) currents.

We classify them as follows:

$$\begin{aligned}
&\text{S-P type}: && \Gamma_\chi = c_S^\chi + ic_P^\chi\gamma_5\,, && \Gamma_e = c_S^e + ic_P^e\gamma_5\,,\\
&\text{V-A type}: && \Gamma_\chi^\mu = (c_V^\chi + c_A^\chi\gamma_5)\,\gamma^\mu\,, && \Gamma_{e\mu} = (c_V^e + c_A^e\gamma_5)\,\gamma_\mu\,,\\
&\text{T-AT type}: && \Gamma_\chi^{\mu\nu} = (c_T^\chi + ic_{AT}^\chi\gamma_5)\,\sigma^{\mu\nu}\,, && \Gamma_{e\mu\nu} = \sigma_{\mu\nu}\,,
\end{aligned} \tag{2}$$

where $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ is the spin tensor and $c_j^{\chi,e}$ are dimensionless, real couplings. For simplicity, in Eq. (1) we have used a common cut-off scale Λ for all Lorentz structures. Furthermore, in our subsequent numerical analysis, we will consider one type of operator at a time, by setting the corresponding couplings $c_j^{\chi,e} = 1$ without loss of generality and all other couplings equal to zero, unless otherwise specified. For instance, setting $c_S^\chi = c_P^\chi = c_S^e = c_P^e = 1$ and all other couplings equal to zero gives us the (S+P)-type operator, which we will simply refer to as the **SP**-type in the following discussion. Similarly, we will denote the $c_V^\chi = c_A^\chi = c_V^e = c_A^e = 1$ case simply as the **VA**-type, and $c_T^\chi = c_{AT}^\chi = 1$ as the **TAT**-type for presenting our numerical results in the (m_χ, Λ) plane. For other choices of the couplings, our results for the sensitivity on Λ can be easily scaled accordingly.

We will impose a theoretical limit of $\Lambda > \sqrt{s}$ for the EFT validity. For relatively larger DM mass, we must also have $\Lambda > 2m_\chi$ in order to describe DM pair annihilation by the EFT. In fact, using $\Lambda = 2m_\chi$ induces 100% error in the EFT prediction for s-channel UV completions. Therefore, we will use $\Lambda > \max\{\sqrt{s}, 3m_\chi\}$ as a conservative lower bound[74] to ensure the validity of our EFT approach.

3. Mono-photon channel

For the mono-photon signal $e^+e^- \to \chi\overline{\chi}\gamma$, the χ's will contribute to the missing transverse energy at the detector. The dominant irreducible SM background to this process comes from neutrino pair production with an associated ISR photon, i.e. $e^+e^- \to \nu\overline{\nu}\gamma$. Since neutrinos are practically indistinguishable from DMs on an event-by-event basis, the majority of this background survives the event selection cuts. However, as we will show later, this background is highly polarization-dependent, and therefore, can be significantly reduced by the proper choice of polarized beams, without affecting the signal much.

Apart from the neutrino background, any SM process with a single photon in the final state can contribute to the total background if all other visible particles escape detected. The SM processes containing either jets or charged particles are relatively easy to distinguish from a DM event, so their contribution to the total background is negligible.[58] The only exception is the *Bhabha scattering* process associated with an extra photon (either from initial or final state radiation), i.e., $e^+e^- \to e^+e^-\gamma$, which has a large cross section, is polarization-independent, and can significantly contribute to the total background whenever the final-state electrons and positrons go undetected, e.g. along beam pipes. In our following analysis, we consider both neutrino and radiative Bhabha backgrounds.

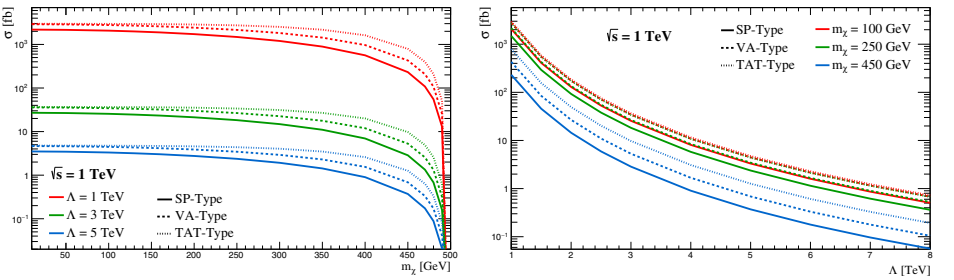


Fig. 1. Variation of mono-photon signal cross-section with the DM mass (left) and the cut-off scale (right) at $\sqrt{s} = 1$ TeV ILC. The solid, dashed and dotted lines are for the SP, VA and TAT-type operators respectively. In the left panel, the red, green and blue curves respectively correspond to different values of the cut-off scale $\Lambda = 1$ TeV, 3 TeV and 5 TeV, while in the right panel, they correspond to different values of the DM mass $m_\chi = 100$ GeV, 250 GeV and 450 GeV.

3.1. *Cross-sections*

The cross-sections for the mono-photon signal $e^+e^- \to \chi\overline{\chi}\gamma$ and the radiative neutrino background $e^+e^- \to \nu\overline{\nu}\gamma$ at $\sqrt{s} = 1$ TeV ILC are estimated using `CalcHEP`[75] with proper implementation of ISR and beamsstrahlung effects, which significantly affect the width and position of the neutrino Z-resonance. For this purpose, the EFT Lagrangian (1) is implemented in `FeynRules`[76] to generate the CHO library required for `CalcHEP`. To avoid collinear and infrared divergences, we limit the phase space in the event generation with the following cuts on the outgoing photon energy E_γ and its polar angle θ_γ:

$$8 \text{ GeV} < E_\gamma < 500 \text{ GeV}, \quad |\cos\theta_\gamma| \leq 0.995\,. \tag{3}$$

The *radiative* Bhabha scattering events are generated using `WHIZARD`[77] (to better handle the singularities) with the same set of cuts as in Eq. (3) to the matrix element photon (i.e., excluding the ISR and beamsstrahlung photons). Also, some additional cuts are implemented for the Bhabha process to take care of the soft and collinear divergences:

$$M_{e^\pm_{\rm in},e^\pm_{\rm out}} < 2m_e,\ M_{e^\pm_{\rm out},e^\pm_{\rm out}} < 5 \text{ GeV},\ P_T^\gamma > 1 \text{ GeV},\ \Delta R_{e^\pm,\gamma} > 0.2,\ \Delta R_{e^\pm,e^\pm} > 0.4. \tag{4}$$

After generating the signal and background events, we perform a fast detector simulation of the `SiD` detector of ILC[78] using `Delphes3`[79] with the configuration card validated in Ref.[80]. The variations of the unpolarized signal cross section as a function of the DM mass and the cut-off scale are shown in Figure 1 left and right panels respectively for all three operator types, namely, SP (solid), VA (dashed) and TAT (dotted)-type. We find that the cross-section is the smallest (largest) for the SP (TAT)-type operator at any given DM mass. In the left panel, the sudden drop in the cross-section as m_χ approaches $\sqrt{s}/2$ is due to phase-space suppression. Otherwise, for smaller DM masses, the cross-section for a given operator type and

Table 1. Comparison of the SM backgrounds and signal cross sections for the mono-photon channel with different choices of beam polarization at $\sqrt{s} = 1$ TeV. For the signal, we have chosen benchmark values of $m_\chi = 100$ GeV and $\Lambda = 3$ TeV. The numbers in bold highlight the optimal polarization choice for a given operator type.

Process type	**Unpolarized cross-section (fb)**	**Polarization $P(e^-, e^+)$**	**Polarized cross-section (fb)**			
			$(+,+)$	$(+,-)$	$(-,+)$	$(-,-)$
$\nu\bar{\nu}\gamma$	4782	$(80, 0)$	1106	1106	8506	8506
		$(80, 20)$	1268	963	10160	6793
		$(80, 30)$	1393	860	10993	5931
$e^-e^+\gamma$	68439	$(80, 0)$	67920	67920	68867	68867
		$(80, 20)$	67909	68386	69285	68297
		$(80, 30)$	67809	68566	69502	68181
SP-type	25.5	$(80, 0)$	25.5	25.5	25.5	25.5
		$(80, 20)$	29.6	21.4	21.4	29.6
		$(80, 30)$	**31.6**	19.4	19.4	31.6
VA-type	34.3	$(80, 0)$	61.7	61.7	6.9	6.9
		$(80, 20)$	49.4	74.1	5.5	8.2
		$(80, 30)$	43.2	**80.3**	4.8	8.9
TAT-type	36.5	$(80, 0)$	36.5	36.5	36.5	36.5
		$(80, 20)$	42.3	30.6	30.6	42.3
		$(80, 30)$	**45.2**	27.7	27.7	45.2

a given cut-off scale is almost independent of the DM mass. In the right panel, we see that for a given DM mass the cross-section drops as Λ^{-4}, as expected.

As for the background, we find that the neutrino background cross section at $\sqrt{s} = 1$ TeV is 4.8 pb, while the radiative Bhabha background is 68.4 pb (though it is substantially reduced after the baseline selection). On the other hand, the DM signal cross section is found to be much smaller, as shown in Table 1 for a benchmark DM mass of $m_\chi = 100$ GeV and the cut-off scale $\Lambda = 3$ TeV.

3.2. *Effect of polarization*

One important advantage of lepton colliders is that the incoming beams can be polarized. This helps to reduce the neutrino background considerably, as shown in Table 1. To utilize the full advantage of the beam polarization, we investigate the effect of different choices of polarization on the signal and background. At the ILC, the baseline design foresees at least 80% electron beam polarization at the interaction point, whereas the positron beam can be polarized up to 30% for the undulator positron source (up to 60% may be possible with the addition of a photon collimator).[68] For comparison, we show our results for three different nominal absolute values of polarization: $|P(e^-, e^+)| = (80, 0)$, (80,20) and (80,30). In each case, we can also have four different polarization configurations, namely, $\text{sign}(P(e^-), P(e^+)) = (+,+)$, $(+,-)$, $(-,+)$ and $(-,-)$, where $+$ and $-$ denote the right- and left-handed helicities respectively.

In Table 1, we show the effect of different schemes of polarizations and helicity orientations on the mono-photon signal and background cross-sections. It is clear that the radiative Bhabha background remains almost unchanged. On the other

hand, electron beam polarization is very effective in reducing the neutrino background, as a 80% *right-handed* electron beam can reduce the neutrino background to 23% of the unpolarized case, even without any polarization on the positron beam. The effect is further enhanced by a *left-handed* positron beam. We see that for 20% and 30% left-handed positron beam polarization, the neutrino background is reduced to 20% and 18% of its unpolarized value, respectively.

The signals are also affected to some extent by beam polarization and the optimal helicity configuration depends on the operator type. For SP- and TAT-type operators we see no effect of electron-beam polarization, but a 20% (30%) *right-handed* positron beam can enhance the signal by 16% (24%). The VA-type signal, on the other hand, prefers the $(+,-)$ helicity configuration – the same choice for which the neutrino background is minimized. With the $(+80\%, -30\%)$ configuration, the VA-type signal is enhanced by a factor of 2.3, whereas the $(+80\%, +30\%)$ configuration enhances it by a modest 26%.

Overall, although the $(+80\%, -30\%)$ configuration minimizes the background the most, looking at the different signal to background ratio, we find that the $(+80\%, +30\%)$ configuration is the best for the SP- and TAT-type operators. For direct comparison between the results for different operators, we choose to work with the $(+80\%, +30\%)$ configuration democratically for all the operator types, unless otherwise specified.

3.3. *Cut-based analysis*

Now we analyze various kinematic distributions and perform a cut-based analysis to optimize the signal-to-background ratio. This of course depends on the DM mass, so in Table 2, we list three benchmark points (BPs) with $m_\chi = 100$ GeV, 250 GeV and 350 GeV respectively, and present the corresponding selection cuts optimized for each case. Here we fix $\Lambda = 3$ TeV for illustration, but in the next subsection, we will vary both m_χ and Λ to obtain the 3σ sensitivity limits. As for the choice of the DM mass values, since it was seen from Figure 1 that the signal cross-sections are barely sensitive to the DM mass up to around 100 GeV, our BP1 essentially captures the light DM scenario. Similarly, our BP3 is chosen moderately close to the kinematic limit of $\sqrt{s}/2$ (going too close to $\sqrt{s}/2$ will result in cross-section values too low too low to give sizable event counts after all the selection cuts). The BP2 is chosen for an intermediate mass DM in between BP1 and BP2.

We define our mono-photon signals by those events that pass through the baseline selection criteria as defined below, in addition to the cuts given in Eq. (3):

$$E_\gamma > 10 \text{ GeV}, \;\; |\eta_\gamma| < 2.45 \;\; \text{and} \;\; P_T^{\text{miss}} > 10 \text{ GeV}\,, \tag{5}$$

where the hardest photon in an event is considered as the signal photon. For the radiative Bhabha background, we define the selection criteria for electrons as $P_{T,e} > 10$ GeV, $|\eta_e| < 2.5$, and have kept only those events which contain no electrons (and positrons) passing these criteria, which means they have escaped detection.

Table 2. Mono-photon selection cuts for different BPs across all operator types.

	BP1	**BP2**	**BP3**
Definition	$m_\chi = 100$ GeV, $\Lambda = 3$ TeV	$m_\chi = 250$ GeV, $\Lambda = 3$ TeV	$m_\chi = 350$ GeV, $\Lambda = 3$ TeV
Baseline selection	$E_\gamma > 10$ GeV, $\|\eta_\gamma\| < 2.45$, $P_T^{\rm miss} > 10$ GeV		
SP-type			
Cut-1	$E_\gamma < 450$ GeV	$E_\gamma < 340$ GeV	$E_\gamma < 250$ GeV
Cut-2	$\|\eta_\gamma\| < 1.6$		
Cut-3	$P_T^{\rm miss} < 450$ GeV	$P_T^{\rm miss} < 340$ GeV	$P_T^{\rm miss} < 240$ GeV
Cut-4	$P_T^{\rm frac} < 1.3$		
Cut-5	$1.1 < \Delta R_{\gamma,\rm MET} < 4.5$		
VA-type			
Cut-1	$E_\gamma < 440$ GeV	$E_\gamma < 350$ GeV	$E_\gamma < 250$ GeV
Cut-2	$\|\eta_\gamma\| < 1.7$		
Cut-3	$P_T^{\rm miss} < 400$ GeV	$P_T^{\rm miss} < 340$ GeV	$P_T^{\rm miss} < 250$ GeV
Cut-4	$P_T^{\rm frac} < 1.2$		
Cut-5	$1.1 < \Delta R_{\gamma,\rm MET} < 4.5$		
TAT-type			
Cut-1	$E_\gamma < 460$ GeV	$E_\gamma < 360$ GeV	$E_\gamma < 230$ GeV
Cut-2	$\|\eta_\gamma\| < 1.7$		
Cut-3	$P_T^{\rm miss} < 450$ GeV	$P_T^{\rm miss} < 350$ GeV	$P_T^{\rm miss} < 230$ GeV
Cut-4	$P_T^{\rm frac} < 1.2$		
Cut-5	$1.1 < \Delta R_{\gamma,\rm MET} < 4.4$		

After implementing these baseline selection cuts, we find that the signal and the neutrino background are reduced to about 60% of their original values in Table 1. Similarly, the actual Bhabha-induced background relevant for our signal is found to be only about 13% of its original value quoted in Table 1 after the baseline selection cuts, taking into account only the missed electron events. To further enhance our signal-to-background ratio, we then examine the signal versus background distributions of some relevant kinematic variables and devise further cuts, which are dynamic with respect to different BPs, as summarized in Table 2. See Ref.[1] for details.

Even after implementing the baseline and analysis cuts 1 through 5, the neutrino background can only be reduced to about 40% of its original value in Table 1. Similarly, the radiative Bhabha background, although substantially reduced to about 4% of its original value in Table 1 after the baseline selection and analysis cuts, still remains sizable and comparable to the neutrino background. However, an electromagnetic calorimeter in the very forward direction of the beamline (BeamCal)[81] can further suppress the Bhabha background to the per mille level. To properly incorporate the effect of BeamCal, we have used the selection efficiencies obtained from a full detector simulation performed in Ref.[61] by modeling the complete instrumented region in a realistic way. According to this analysis, the selection efficiency of the Bhabha background after the BeamCal veto only is 2.7%, while that of the neutrino background is between 98% and 99.6%. As for the DM signal, we expect it to be

Table 3. Signal significance in the mono-photon channel for the three BPs at $\sqrt{s} = 1$ TeV with $\mathcal{L}_{\rm int} = 1000$ fb^{-1} integrated luminosity. The values in the parenthesis denote the significances with a 1% background systematic uncertainty.

Operator type	**Signal significance for $\mathcal{L}_{\rm int} = 1000\,{\rm fb}^{-1}$**					
	Unpolarized beams			Polarized beams		
	BP-1	BP-2	BP-3	BP-1	BP-2	BP-3
SP-type	8.1 (0.6)	5.8 (0.4)	3.5 (0.3)	18.1 (2.4)	13.0 (1.7)	7.8 (1.0)
VA-type	10.9 (0.8)	8.5 (0.6)	5.6 (0.4)	24.9 (3.2)	19.4 (2.5)	12.9 (1.7)
TAT-type	11.8 (0.8)	10.8 (0.8)	8.5 (0.6)	26.2 (3.5)	24.1 (3.2)	19.2 (2.6)

basically unaffected (just like the neutrino background) by the BealCal veto, as it does not contain highly energetic charged particles in the longitudinal direction.

For the polarized case, after the baseline selection cuts, the Bhabha background remains almost same as in the unpolarized case. The neutrino background, on the other hand, is significantly reduced in the polarized case to about 28% of its unpolarized value. The other cut efficiencies are also slightly better for the neutrino background in the polarized case.

As for the signals, from Table 1, we see that the TAT-type operator has the largest cross section to start with, both for the unpolarized as well as for the $(+80\%, +30\%)$ polarized cases. Even after the baseline selection and the specialized cuts discussed above, the TAT-type signal retains the largest efficiency among the three types. This will be reflected in our signal significance results below.

3.4. *Signal significance*

After implementing all the cuts mentioned above, we calculate the final signal significance for our benchmark scenarios using the definition

$$\text{Sig} = \frac{S}{\sqrt{S + B + (\epsilon B)^2}}\,, \tag{6}$$

where S and B are the number of signal and total background events respectively for a given integrated luminosity, and ϵ is the background systematic uncertainty. Our results are given in Table 3 for the three BPs. We show the numbers for an ideal case with zero systematics and also for a more realistic case with 1% systematics, i.e. with $\epsilon = 0.01$ (in parentheses). The results are significantly weakened in the latter case because of the relatively large background compared to the signal.

From Table 3, we see that the significance enhances as we go to lower DM mass regions, as expected because of kinematic reasons. Operator-wise we see that TAT and VA-type operators perform better than the SP-type. We also find substantial (around 50%) increase in significance on application of optimal beam polarization.

Going beyond the three BPs, we now vary the DM mass and calculate the signal significance following the same cut-based analysis procedure outlined above. Our results for the 3σ sensitivity contours in the (m_χ, Λ) plane are shown in Figure 2 for all the operator types. The solid (dashed) contours are for the unpolarized (optimally

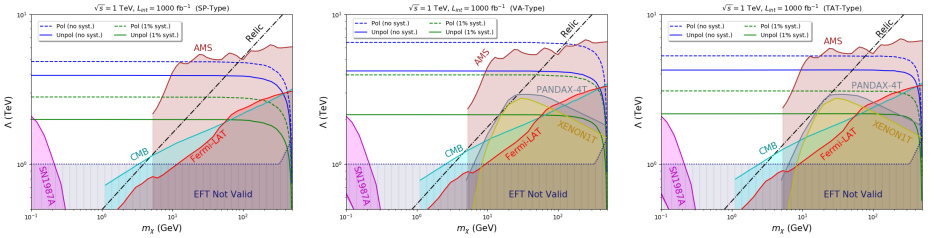


Fig. 2. 3σ sensitivity contours in the mono-photon channel for the SP (left), VA (middle) and TAT (right)-type operators with *unpolarized* (solid lines) and *polarized* (dashed lines) e^+e^- beams at $\sqrt{s} = 1$ TeV center-of-mass energy and with $\mathcal{L}_{\rm int} = 1000$ fb^{-1} integrated luminosity. The blue (green) contours are assuming zero (1%) background systematics. The various shaded regions are excluded by direct detection (XENON1T, PANDAX-4T), indirect detection (Fermi-LAT, AMS), astrophysics (SN1987A) and cosmology (CMB) constraints. In the shaded region below $\Lambda = \max\{\sqrt{s}/2, 3m_\chi\}$, our EFT framework is not valid. Along the dot-dashed line, the observed DM relic density is reproduced for a thermal DM assuming only DM-electron effective coupling.

polarized) case, and the blue (green) contours are assuming zero (1%) background systematics. The shaded regions are excluded by various constraints. First of all, for $\Lambda < \max\{\sqrt{s}/2, 3m_\chi\}$, our EFT framework is not valid (cf. Section 2). This is shown by the navy blue-shaded regions in Figure 2. For $\sqrt{s} = 1$ TeV as considered here, this EFT validity limit supersedes the previous LEP limit.[42]

The same effective operator given in Eq. (1) also gives rise to DM scattering with electrons $\chi e^- \to \chi e^-$. The exact analytic expressions for these cross sections in our EFT framework can be found in Appendix C of Ref.[82] for all the operator types. Comparing these with the experimental upper limits on $\sigma_{\chi e}$ from dedicated direct detection experiments,[32,33] we can derive a *lower* limit on the cut-off scale Λ as a function of the DM mass m_χ. However, the current best limit on $\sigma_{\chi e}$ from XENON1T is at the level of $\mathcal{O}(10^{-39})$ cm^2,[33] which translates into a very weak bound on Λ and is not relevant for our study. Even the future ambitious proposals like DARKSPHERE can only reach up to $\mathcal{O}(10^{-42})$cm^2,[83] still 5 orders of magnitude weaker than that needed to probe a TeV-scale Λ value.

However, more stringent limits can be derived from DM-nucleon scattering searches. Even for an LDM as in our case, DM-nucleon couplings are necessarily induced at loop level from photon exchange between virtual leptons and the quarks. In fact, as shown in Ref.[37], the loop-induced DM-nucleon scattering almost always dominates over the DM-electron scattering. The analytic expressions for the one and two-loop DM-nucleon scattering cross sections can be found in Ref.[37]. We have translated the experimental upper limits from XENON1T[84] and PANDAX-4T[85] onto the (m_χ, Λ) plane, as shown by the yellow and grey-shaded regions respectively in Figure 2. Note that these limits are only applicable for the vector and tensor lepton currents, i.e. $\Gamma_\ell = \gamma_\mu,\ \sigma_{\mu\nu}$ in Eq. (1). For the scalar lepton current, $\Gamma_\ell = 1$, the one-loop DM-nucleon coupling vanishes, and one has to go to two loops

which is suppressed by $\alpha_{\rm em}^2$ for the S-S type coupling and $\alpha_{\rm em}^2 v^2$ (where $v \sim 10^{-3}$ is the DM velocity) for the P-S type coupling. In contrast, for pseudo-scalar and axial-vector lepton currents, i.e. $\Gamma_\ell = \gamma_5,\ \gamma_\mu\gamma_5$, the DM-nucleon coupling vanishes to all orders. Therefore, we have not shown the XENON1T and PANDAX-4T limits for the SP-type operator on the top left panel of Figure 2.

The same effective operator given in Eq. (1) also enables DM annihilation into electrons $\chi\overline{\chi} \to e^+e^-$. The exact analytic expressions for these cross sections in our EFT framework can be found in Appendix C of Ref.[82] for all the operator types. Using these, we calculate the thermal-averaged cross section times relative velocity $\langle\sigma v\rangle$ which goes as m_χ^2/Λ^4 and compare it with the existing indirect detection upper limits on $\langle\sigma v\rangle$ in the e^+e^- channel to put a lower bound on Λ as a function of the DM mass. This is shown in Figure 2 by the red and brown-shaded regions respectively for the Fermi-LAT[86] and AMS-02[53] constraints on $\langle\sigma v\rangle$. Similar constraints on $\langle\sigma v\rangle$ can be derived using CMB anisotropies,[86] which is shown by the cyan-shaded region in Figure 2, assuming an s-wave annihilation (for p-wave annihilation, the CMB bound will be much weaker).

Along the dot-dashed line in Figure 2, the observed relic density can be reproduced for a DM. In principle, the region to the left and above of this line is disfavored for a thermal DM, because in this region $\langle\sigma v\rangle$ is smaller than the observed value of $\sim (2-5)\times 10^{-26}{\rm cm}^3{\rm sec}^{-1}$ (depending on the DM mass[87]), which leads to an overabundance of DM, since $\Omega_\chi h^2 \propto 1/\langle\sigma v\rangle$. However, this problem can be circumvented by either opening up additional leptonic annihilation channels (like $\mu^+\mu^-$, $\tau^+\tau^-$ and $\nu\bar{\nu}$) or even going beyond the DM paradigm and invoking e.g., the freeze-in mechanism.[88] This will not affect our main results, since the collider phenomenology discussed here only depends on the DM coupling to electrons.

Also shown in Figure 2 is the supernova constraint, which excludes the magenta-shaded region from consideration of energy-loss and optical depth criteria from the observation of SN1987A.[82] Here we have used an average supernova core temperature of 30 MeV. Note that the supernova bound is only applicable for DM mass below $\sim$ 200 MeV or so, and for a certain range of Λ values, above which the DM particles cannot be efficiently produced in the supernova core, and below which they will no longer free-stream.

From Figure 2, we find that in spite of a large irreducible background, the accessible range of the cut-off scale Λ at $\sqrt{s} = 1$ TeV ILC looks quite promising in the mono-photon channel, especially for low mass DM, where the collider sensitivity is almost flat, whereas the existing direct and indirect detection constraints are much weaker. This complementarity makes the collider searches for DM very promising. With unpolarized beams, the 3σ-reach for the SP-type operator can be up to 3.9 TeV, while for the VA and TAT-type operators, it can be up to 4.2 TeV. With optimally polarized beams, i.e. with $(+80\%, +30\%)$ for the SP and TAT-types and $(+80\%, -30\%)$ for the VA type, the sensitivity reaches can be extended to 4.8 TeV (SP), 6.5 TeV (VA) and 5.3 TeV (TAT), as shown in Figure 2.

4. Mono-Z channel

In addition to the mono-photon channel discussed in the previous section, another useful channel for LDM search at lepton colliders is the mono-Z channel, where the Z-boson is emitted from one of the initial states. Depending on the subsequent decay of the Z-boson to either leptonic or hadronic final states, we perform a dedicated cut-based signal and background analysis, as discussed below.

4.1. *Leptonic mode*

For the leptonic decay of the Z-boson, we examine the process $e^+e^- \to \chi\overline{\chi}Z(\to \ell^-\ell^+)$. We will only consider $\ell = e, \mu$ for simplicity and use the lepton pair as the visible particles for tagging. The main SM background for this channel is $e^+e^- \to \nu\overline{\nu}\,\ell^+\ell^-$, and it is polarization-dependent.

4.1.1. *Unpolarized and polarized cross-sections*

For the signal and background simulation, we generated the UFO library for our EFT framework using `FeynRules`[76] and then generated events for both signal and background using `MadGraph 5`[89] with the following basic baseline cuts:

$$P_T(\ell) > 10 \text{ GeV}, \quad |\eta_\ell| \leq 2.5, \quad \Delta R_{\ell\ell} \geq 0.4\,. \tag{7}$$

For the signal, the Z-bosons are decayed into the charged lepton pairs via the `MadSpin`[90,91] package which is implemented in `MadGraph 5`, to take care of the spin-correlation effects of the lepton pairs. A fast detector simulation to these events is done using `Delphes 3`[79] with the same configuration card[80] as in Section 3.1.

With unpolarized beams, we find that the neutrino background cross section at $\sqrt{s} = 1$ TeV is 420.5 fb, whereas the DM signal cross section is much smaller, as shown in Table 4 for a benchmark DM mass of $m_\chi = 100$ GeV and the cut-off scale $\Lambda = 3$ TeV. Similar to the mono-photon case, we also examine the effect of polarization on the signal and background cross-sections, as shown in Table 4. The neutrino background can be reduced to 28% of its original value by making the electron beam +80% polarized, and further reduced to 21% of its original value by additionally making the positron beam −30% polarized. The $(+80\%, -30\%)$ polarization configuration also enhances the VA-type signal by a factor of 2.4. However, the $(+80\%, +30\%)$ configuration is better for the SP and TAT-type signals. For ease of comparison between different operator types, we choose to work with the $(+80\%, +30\%)$ configuration democratically for all operator types, as well as for the background, unless otherwise specified.

4.1.2. *Cut-based analysis*

We define our signals by those events that pass through the baseline selection criteria as defined below: $P_{T,\ell} > 20$ GeV, $|\eta_\ell| < 2.45$, where the Z-boson is reconstructed

Table 4. Comparison of the leptonic mono-Z background and signal cross-sections for different choices of beam polarization for $m_\chi = 100$ GeV and $\Lambda = 3$ TeV at $\sqrt{s} = 1$ TeV ILC. The numbers in bold highlight the optimal polarization choice for a given operator type.

Process type	**Unpolarized cross-section (fb)**	**Polarization $P(e^-, e^+)$**	**Polarized cross-section (fb)**			
			$(+,+)$	$(+,-)$	$(-,+)$	$(-,-)$
$\nu\bar{\nu}\ell^-\ell^+$	420	$(80, 0)$	116	116	723	723
		$(80, 20)$	135	98	856	590
		$(80, 30)$	145	88	926	523
SP-Type	0.28	(80, 0)	0.26	0.26	0.25	0.25
		(80,20)	0.29	0.22	0.21	0.29
		(80, 30)	**0.32**	0.19	0.19	0.32
VA-Type	0.08	(80, 0)	0.15	0.15	0.02	0.02
		(80,20)	0.12	0.18	0.01	0.02
		(80, 30)	0.11	**0.19**	0.01	0.02
TAT-Type	0.68	(80, 0)	0.62	0.62	0.62	0.62
		(80,20)	0.72	0.52	0.52	0.72
		(80, 30)	**0.77**	0.47	0.47	0.77

by the condition that all final state lepton-pairs are oppositely charged and of same flavor (OSSF). Other selection criteria are dynamic with respect to different BPs. We have taken the same three BPs as in the mono-photon case to probe different regions of the parameter space, namely, BP1 essentially represents all light DM region, BP3 represents the region close to the kinematic limit of $\sqrt{s}/2 - m_Z$, whereas BP2 captures the intermediate DM mass region.

After implementing the baseline selection cuts, we find that the background is reduced to about 40% of its original value in Table 4 for the unpolarized (polarized) case, whereas the signals are reduced to about 60%-70% of their original values. We then consider various kinematic distributions for the signal and background, and devise some specialized selection cuts.[1] We find that after applying all these cuts, we can still retain about 35%-45% of the signal, whereas the background is reduced to below percent level of the original values given in Table 4.

4.1.3. *Results*

After implementing all these cuts, we calculate the final signal significance for the three BPs using Eq. (6). Our results are given in Table 5 for an integrated luminosity of $\mathcal{L}_{\rm int} = 1000$ fb^{-1}. We see that as we go higher up in the DM mass the signal significance drops. We also find that the best-performing operator type is the TAT-type, for which more than 97% of the background events are removed after all the selection cuts. For the signal we retain $58\% - 61\%$ of the events, although for BP3 only 48% remains. The SP-type operator also gives good results, where we retain $50\% - 66\%$ of the signal across BPs and polarization choices, while removing more than 96% of the background events. Even for VA-type we retain more than 50% of the signal events and are able to cut down the background event yields to 11%. We also notice the positive effect of the beam polarization by which we achieve an enhancement of signal significance by more than 2 times compared to the ones with

Table 5. Signal significance in the mono-Z leptonic channel at $\sqrt{s} = 1$ TeV and $\mathcal{L}_{\rm int} = 1000\,{\rm fb}^{-1}$. The values in the parenthesis correspond to 1% background systematic uncertainty.

Operator type	**Signal significance for $\mathcal{L}_{\rm int} = 1000\,{\rm fb}^{-1}$**					
	Unpolarized beams			Polarized beams		
	BP-1	BP-2	BP-3	BP-1	BP-2	BP-3
SP-type	1.7 (1.3)	0.7 (0.5)	0.1 (0.1)	3.8 (3.6)	1.7 (1.5)	0.3 (0.3)
VA-type	0.2 (0.1)	0.1 (0.1)	0.1 (0.1)	0.5 (0.4)	0.4 (0.3)	0.2 (0.2)
TAT-type	4.5 (3.9)	2.4 (1.9)	0.6 (0.5)	9.4 (9.1)	5.4 (5.1)	1.3 (1.3)

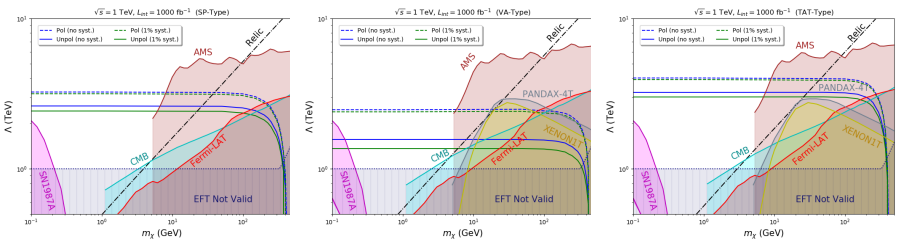


Fig. 3. 3σ sensitivity contours in the mono-Z leptonic channel. Labels are same as in Figure 2.

unpolarized beams. For VA-type though the significance can be further increased for the polarized beam case by choosing the left-handed positron beams as is evident from Table 4.

Going beyond the three BPs, we now vary the DM mass and present the 3σ sensitivity reach for this channel in Figure 3 for all the operators. The labels and shaded regions are the same as in the mono-photon case (cf. Figure 2). We see that the accessible range of the cut-off scale Λ for the unpolarized beams can reach up to 3.2 TeV for the TAT-type operator, whereas for the SP and VA-type, it can reach up to 2.6 TeV and 1.6 TeV respectively. But with the application of optimally polarized beams as discussed earlier, we see an increase by about 25% of the 3σ reach on the Λ scale, up to 3.2 TeV, 2.5 TeV and 4 TeV for the for SP, VA and TAT-type operators, respectively.

4.2. *Hadronic mode*

Next we study $e^+e^- \to \chi\overline{\chi}Z(\to jj)$, where $j \equiv u, d, c, s, b$ quarks. The relevant SM background processes for this channel are $e^+e^- \to \nu\overline{\nu}jj$ and $e^+e^- \to jj\ell\nu$ (with one charged lepton escaping the detector) where the jets and leptons in the final state can come from any possible source (not necessarily from an on-shell Z).

4.2.1. *Unpolarized and polarized cross-sections*

We use the same UFO library as before which is implemented using `FeynRules`[76] and simulate the events for the signal and backgrounds via `MadGraph 5`[89] with the

Table 6. Comparison of the hadronic mono-Z background and signal cross-sections for different choices of beam polarization with $m_\chi = 100$ GeV and $\Lambda = 3$ TeV at $\sqrt{s} = 1$ TeV ILC. The numbers in bold highlight the optimal polarization choice for a given operator type.

Process type	**Unpolarized cross-section (fb)**	**Pol. $P(e^-, e^+)$**	**Polarized cross-section (fb)**			
			$(+,+)$	$(+,-)$	$(-,+)$	$(-,-)$
$\nu\bar{\nu}jj$	798	$(80,0)$	178	178	1415	1415
		$(80,20)$	206	151	1689	1134
		$(80,30)$	219	136	1833	989
$jj\ell\nu$	1186	$(80,0)$	302	302	2061	2061
		$(80,20)$	359	246	2446	1685
		$(80,30)$	386	216	2635	1492
SP-Type	2.78	$(80,0)$	2.57	2.57	2.58	2.58
		$(80,20)$	2.98	2.15	2.15	2.97
		$(80,30)$	**3.17**	1.95	1.95	3.17
VA-Type	0.83	$(80,0)$	1.35	1.35	0.15	0.15
		$(80,20)$	1.07	1.61	0.12	0.18
		$(80,30)$	0.94	**1.76**	0.10	0.19
TAT-Type	6.77	$(80,0)$	6.22	6.22	6.21	6.21
		$(80,20)$	7.23	5.23	5.24	7.21
		$(80,30)$	**7.72**	4.73	4.73	7.69

following basic cuts to the parameter space:

$$P_T(j,\ell) > 10 \text{ GeV}, \quad |\eta_j| \leq 3.0, \;\; |\eta_\ell| \leq 2.5, \;\; \Delta R_{jj,\ell j} \geq 0.4\,. \tag{8}$$

For the signals, as in the leptonic case, the on-shell Z-bosons are decayed into the pairs of jets using the `MadSpin` package,[90,91] implemented in `MadGraph 5`. Both the signal and background samples are hadronized using `Pythia8.2`[92] and then the final state jets are reconstructed with with `anti-`kT[93] clustering algorithm with a minimum P_T of 10 GeV and a cone radius (R) of 0.4 using the `FastJet`[94] package. The fast detector simulation to these events are done using `Delphes 3`[79] with the same configuration card[80] as discussed in Section 3.1.

With unpolarized beams, we find the neutrino-pair background is 798 fb, whereas the $jj\ell\nu$ background is 1186 fb. On the other hand, the DM signal is only at a few fb level, as shown in Table 6 for a benchmark DM mass of $m_\chi = 100$ GeV and the cut-off scale $\Lambda = 3$ TeV. We then examine different choices of beam-polarization on both the event samples for this channel, as shown in Table 6. We find that both backgrounds are polarization-dependent and fall off significantly for right-handed electron beam and with increasing degree of polarization. We choose the polarization configuration $P(e^-, e^+) = (+80\%, +30\%)$ democratically over all the operator types.

4.2.2. *Cut-based analysis*

After obtaining the signal and background cross-sections as reported in Table 6, we proceed with our cut-based analysis to optimize the signal significance. We select the events that contain at least two jets with the following transverse momentum and pseudorapidity requirements: $P_{T,j} > 20$ GeV, $|\eta_j| < 2.45$. The hardest two jets

Table 7. Signal significances of the mono-Z hadronic channel at $\sqrt{s} = 1$ TeV and $\mathcal{L}_{\rm int} = 1000$ fb^{-1}. The values in the parenthesis denote the significance with a 1% background systematic uncertainty.

Operator types	**Signal significance for $\mathcal{L}_{\rm int} = 1000\,{\rm fb}^{-1}$**					
	Unpolarized beams			Polarized Beam		
	BP-1	BP-2	BP-3	BP-1	BP-2	BP-3
SP-type	4.7 (3.3)	1.5 (1.0)	0.3 (0.2)	10.3 (9.1)	3.6 (3.1)	0.8 (0.6)
VA-type	0.4 (0.2)	0.3 (0.1)	0.1 (0.1)	0.2 (0.1)	0.1 (0.1)	0.1 (0.04)
TAT-type	14.2 (10.4)	5.8 (3.7)	1.2 (0.7)	27.7 (25.6)	12.9 (11.1)	2.8 (2.3)

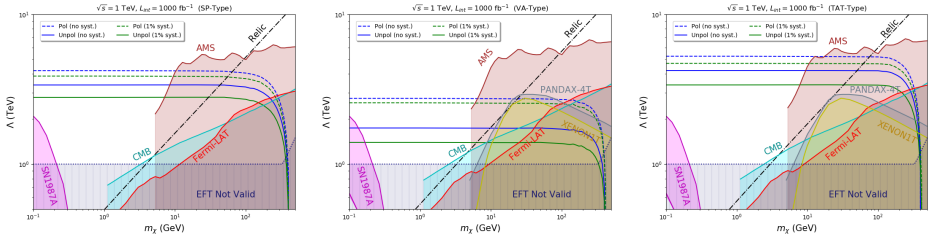


Fig. 4. 3σ sensitivity contours in the mono-Z hadronic channel. Labels are same as in Figure 2.

are required to reconstruct the Z-boson. Further selection cuts are applied some of which depend on the DM mass. So, as in the leptonic channel, we have taken the same three BPs with varying DM mass and impose dynamic cuts.[1]

4.2.3. *Results*

The signal significances calculated using Eq. (6) are tabulated in Table 7. We see similar behavior for the different BPs as in the previously discussed channels, i.e. enhanced signal significance with decreasing mass of the DM. The selection cuts are most efficient for SP- and TAT-type operators. For BP-1 we remove more than 98% of the background events while keeping at least 21% of the signal events for the two operator types, yielding a large signal significance especially for the TAT-type operator with polarized beams.

Varying the DM mass, we display the 3σ sensitivity contours for all three operators in Figure 4. It is clear that the TAT-type operator has the best sensitivity, which reaches up to 4.2 TeV with unpolarized beams and 5.2 TeV with optimally polarized beams. The SP-type operator also has a sensitivity comparable to the mono-photon channel, and can reach up to 3.4 (4.2) TeV with unpolarized (polarized) beams. The VA-type operator has a modest sensitivity in this channel, only up to 1.7 (2.7) TeV with unpolarized (polarized) beams.

5. Conclusion

We have explored the physics potential of the future e^+e^- colliders in probing such *leptophilic* DM in a model-independent way. As a case study, we have taken the

Table 8. Summary of our results for the 3σ sensitivity reach of the cut-off scale Λ in the three different channels discussed in the text. Here we have fixed the DM mass at 1 GeV. The numbers in parentheses are with 1% background systematics. The numbers in bold show the highest Λ value that can be probed for a given operator.

Process type	**Beam configuration**	**3σ sensitivity reach of Λ (TeV)**		
		SP	VA	TAT
Mono-γ	Unpolarized	3.91 (1.99)	4.19 (2.14)	4.25 (2.17)
	Polarized	**4.84** (2.81)	**6.49 (3.94)**	**5.28** (3.08)
Mono-Z leptonic	Unpolarized	2.62 (2.42)	1.57 (1.36)	3.22 (3.00)
	Polarized	3.24 (3.16)	2.47 (2.39)	4.02 (3.93)
Mono-Z hadronic	Unpolarized	3.38 (2.79)	1.74 (1.39)	4.22 (3.38)
	Polarized	4.21 (**3.87**)	2.75 (2.57)	5.25 (**4.71**)

$\sqrt{s} = 1$ TeV ILC with an integrated luminosity of 1000 fb^{-1} and have analyzed the pair-production of fermionic DM using leptophilic dimension-6 operators of all possible bilinear structures, namely, scalar-pseudoscalar, vector-axialvector and tensor-axialtensor. We have performed a detailed cut-based analysis for each of these operators in three different channels based on the tagged particle, namely, mono-photon, mono-Z leptonic and hadronic.

We have taken into account one of the most important and powerful features of lepton colliders, i.e., the possibility of beam-polarization with different degrees of polarization and helicity orientations. We find that the sign$(P(e^-), P(e^+)) = (+,+)$ beam configuration is optimal for the SP and TAT-type operators, while the $(+,-)$ configuration is better for probing the VA-type operators. The maximum value of the cut-off scale Λ that can be probed in each channel at 3σ is given in Table 8. We find that without any systematics, the mono-photon channel provides the best sensitivity across all operator types, while in presence of background systematic effects, the mono-Z hadronic channel provides better sensitivity for the SP and TAT-type operators.

We also demonstrate the complementarity of our lepton collider study with other existing direct and indirect detection searches for LDM (cf. Figures 2, 3 and 4). In particular, we show that lepton colliders will be able to provide the best-ever sensitivity in the still unexplored light DM regime.

Acknowledgments

The work of BD is supported in part by the US Department of Energy under Grant No. DE-SC0017987 and by a Fermilab Intensity Frontier Fellowship.

References

1. S. Kundu, A. Guha, P. K. Das and P. S. B. Dev, A model-independent analysis of leptophilic dark matter at future electron-positron colliders in the mono-photon and mono-Z channels (10 2021).
2. L. M. Krauss, S. Nasri and M. Trodden, A Model for neutrino masses and dark matter, *Phys. Rev. D* **67**, p. 085002 (2003).

3. E. A. Baltz and L. Bergstrom, Detection of leptonic dark matter, *Phys. Rev. D* **67**, p. 043516 (2003).
4. E. Ma, Verifiable radiative seesaw mechanism of neutrino mass and dark matter, *Phys. Rev. D* **73**, p. 077301 (2006).
5. T. Hambye, K. Kannike, E. Ma and M. Raidal, Emanations of Dark Matter: Muon Anomalous Magnetic Moment, Radiative Neutrino Mass, and Novel Leptogenesis at the TeV Scale, *Phys. Rev. D* **75**, p. 095003 (2007).
6. R. Bernabei *et al.*, Investigating electron interacting dark matter, *Phys. Rev. D* **77**, p. 023506 (2008).
7. M. Cirelli, M. Kadastik, M. Raidal and A. Strumia, Model-independent implications of the e+-, anti-proton cosmic ray spectra on properties of Dark Matter, *Nucl. Phys. B* **813**, 1 (2009), [Addendum: Nucl.Phys.B 873, 530–533 (2013)].
8. C.-R. Chen and F. Takahashi, Cosmic rays from Leptonic Dark Matter, *JCAP* **02**, p. 004 (2009).
9. X.-J. Bi, P.-H. Gu, T. Li and X. Zhang, ATIC and PAMELA Results on Cosmic e+- Excesses and Neutrino Masses, *JHEP* **04**, p. 103 (2009).
10. A. Ibarra, A. Ringwald, D. Tran and C. Weniger, Cosmic Rays from Leptophilic Dark Matter Decay via Kinetic Mixing, *JCAP* **08**, p. 017 (2009).
11. P. S. B. Dev, D. K. Ghosh, N. Okada and I. Saha, Neutrino Mass and Dark Matter in light of recent AMS-02 results, *Phys. Rev. D* **89**, p. 095001 (2014).
12. S. Chang, R. Edezhath, J. Hutchinson and M. Luty, Leptophilic Effective WIMPs, *Phys. Rev. D* **90**, p. 015011 (2014).
13. P. Agrawal, Z. Chacko and C. B. Verhaaren, Leptophilic Dark Matter and the Anomalous Magnetic Moment of the Muon, *JHEP* **08**, p. 147 (2014).
14. N. F. Bell, Y. Cai, R. K. Leane and A. D. Medina, Leptophilic dark matter with Z' interactions, *Phys. Rev. D* **90**, p. 035027 (2014).
15. A. Freitas and S. Westhoff, Leptophilic Dark Matter in Lepton Interactions at LEP and ILC, *JHEP* **10**, p. 116 (2014).
16. Q.-H. Cao, C.-R. Chen and T. Gong, Leptophilic dark matter confronts AMS-02 cosmic-ray positron flux, *Chin. J. Phys.* **55**, 10 (2017).
17. B.-Q. Lu and H.-S. Zong, Leptophilic dark matter in Galactic Center excess, *Phys. Rev. D* **93**, p. 083504 (2016), [Addendum: Phys.Rev.D 93, 089910 (2016)].
18. G. H. Duan, L. Feng, F. Wang, L. Wu, J. M. Yang and R. Zheng, Simplified TeV leptophilic dark matter in light of DAMPE data, *JHEP* **02**, p. 107 (2018).
19. E. Madge and P. Schwaller, Leptophilic dark matter from gauged lepton number: Phenomenology and gravitational wave signatures, *JHEP* **02**, p. 048 (2019).
20. S. Junius, L. Lopez-Honorez and A. Mariotti, A feeble window on leptophilic dark matter, *JHEP* **07**, p. 136 (2019).
21. S. Ghosh, A. Dutta Banik, E. J. Chun and D. Majumdar, Leptophilic-portal Dark Matter in the Light of AMS-02 positron excess (3 2020).
22. S. Chakraborti and R. Islam, Implications of dark sector mixing on leptophilic scalar dark matter, *JHEP* **03**, p. 032 (2021).
23. S.-I. Horigome, T. Katayose, S. Matsumoto and I. Saha, Leptophilic fermion WIMP: Role of future lepton colliders, *Phys. Rev. D* **104**, p. 055001 (2021).
24. B. Abi *et al.*, Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm, *Phys. Rev. Lett.* **126**, p. 141801 (2021).
25. R. Bernabei *et al.*, The DAMA project: Achievements, implications and perspectives, *Prog. Part. Nucl. Phys.* **114**, p. 103810 (2020).
26. S. Abdollahi *et al.*, Cosmic-ray electron-positron spectrum from 7 GeV to 2 TeV with the Fermi Large Area Telescope, *Phys. Rev. D* **95**, p. 082007 (2017).

27. G. Ambrosi *et al.*, Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons, *Nature* **552**, 63 (2017).
28. O. Adriani *et al.*, Extended Measurement of the Cosmic-Ray Electron and Positron Spectrum from 11 GeV to 4.8 TeV with the Calorimetric Electron Telescope on the International Space Station, *Phys. Rev. Lett.* **120**, p. 261102 (2018).
29. M. Aguilar *et al.*, The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II — Results from the first seven years, *Phys. Rept.* **894**, 1 (2021).
30. M. Ajello *et al.*, Fermi-LAT Observations of High-Energy γ-Ray Emission Toward the Galactic Center, *Astrophys. J.* **819**, p. 44 (2016).
31. E. Aprile *et al.*, Excess electronic recoil events in XENON1T, *Phys. Rev. D* **102**, p. 072004 (2020).
32. E. Aprile *et al.*, Exclusion of Leptophilic Dark Matter Models using XENON100 Electronic Recoil Data, *Science* **349**, 851 (2015).
33. E. Aprile *et al.*, Light Dark Matter Search with Ionization Signals in XENON1T, *Phys. Rev. Lett.* **123**, p. 251801 (2019).
34. D. S. Akerib *et al.*, Projected sensitivities of the LUX-ZEPLIN (LZ) experiment to new physics via low-energy electron recoils (2 2021).
35. C.-Y. Chen, J. Kozaczuk and Y.-M. Zhong, Exploring leptophilic dark matter with NA64-μ, *JHEP* **10**, p. 154 (2018).
36. L. Marsicano, M. Battaglieri, A. Celentano, R. De Vita and Y.-M. Zhong, Probing Leptophilic Dark Sectors at Electron Beam-Dump Facilities, *Phys. Rev. D* **98**, p. 115022 (2018).
37. J. Kopp, V. Niro, T. Schwetz and J. Zupan, DAMA/LIBRA and leptonically interacting Dark Matter, *Phys. Rev. D* **80**, p. 083502 (2009).
38. M. Beltran, D. Hooper, E. W. Kolb, Z. A. C. Krusberg and T. M. P. Tait, Maverick dark matter at colliders, *JHEP* **09**, p. 037 (2010).
39. J. Goodman, M. Ibe, A. Rajaraman, W. Shepherd, T. M. P. Tait and H.-B. Yu, Constraints on Light Majorana dark Matter from Colliders, *Phys. Lett. B* **695**, 185 (2011).
40. Y. Bai, P. J. Fox and R. Harnik, The Tevatron at the Frontier of Dark Matter Direct Detection, *JHEP* **12**, p. 048 (2010).
41. J. Goodman, M. Ibe, A. Rajaraman, W. Shepherd, T. M. P. Tait and H.-B. Yu, Constraints on Dark Matter from Colliders, *Phys. Rev. D* **82**, p. 116010 (2010).
42. P. J. Fox, R. Harnik, J. Kopp and Y. Tsai, LEP Shines Light on Dark Matter, *Phys. Rev. D* **84**, p. 014028 (2011).
43. P. J. Fox, R. Harnik, J. Kopp and Y. Tsai, Missing Energy Signatures of Dark Matter at the LHC, *Phys. Rev. D* **85**, p. 056011 (2012).
44. A. Rajaraman, W. Shepherd, T. M. P. Tait and A. M. Wijangco, LHC Bounds on Interactions of Dark Matter, *Phys. Rev. D* **84**, p. 095013 (2011).
45. Y. J. Chae and M. Perelstein, Dark Matter Search at a Linear Collider: Effective Operator Approach, *JHEP* **05**, p. 138 (2013).
46. F. Kahlhoefer, Review of LHC Dark Matter Searches, *Int. J. Mod. Phys. A* **32**, p. 1730006 (2017).
47. B. Penning, The pursuit of dark matter at colliders—an overview, *J. Phys. G* **45**, p. 063001 (2018).
48. V. Khachatryan *et al.*, Search for dark matter, extra dimensions, and unparticles in monojet events in proton–proton collisions at $\sqrt{s} = 8$ TeV, *Eur. Phys. J. C* **75**, p. 235 (2015).
49. G. Aad *et al.*, Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector, *Eur. Phys. J. C* **75**, p. 299 (2015), [Erratum: Eur. Phys. J. C 75, 408 (2015)].

50. A. M. Sirunyan *et al.*, Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **97**, p. 092005 (2018).
51. G. Aad *et al.*, Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector, *Phys. Rev. D* **103**, p. 112006 (2021).
52. L. A. Cavasonza, H. Gast, M. Krämer, M. Pellen and S. Schael, Constraints on leptophilic dark matter from the AMS-02 experiment, *Astrophys. J.* **839**, p. 36 (2017), [Erratum: Astrophys.J. 869, 89 (2018)].
53. I. John and T. Linden, Cosmic-Ray Positrons Strongly Constrain Leptophilic Dark Matter (7 2021).
54. J. Abdallah *et al.*, Photon events with missing energy in e+ e- collisions at s**(1/2) = 130-GeV to 209-GeV, *Eur. Phys. J. C* **38**, 395 (2005).
55. A. Birkedal, K. Matchev and M. Perelstein, Dark matter at colliders: A Model independent approach, *Phys. Rev. D* **70**, p. 077701 (2004).
56. P. J. Fox and E. Poppitz, Leptophilic Dark Matter, *Phys. Rev. D* **79**, p. 083528 (2009).
57. P. Konar, K. Kong, K. T. Matchev and M. Perelstein, Shedding Light on the Dark Sector with Direct WIMP Production, *New J. Phys.* **11**, p. 105004 (2009).
58. C. Bartels, M. Berggren and J. List, Characterising WIMPs at a future e^+e^- Linear Collider, *Eur. Phys. J. C* **72**, p. 2213 (2012).
59. H. Dreiner, M. Huck, M. Krämer, D. Schmeier and J. Tattersall, Illuminating Dark Matter at the ILC, *Phys. Rev. D* **87**, p. 075015 (2013).
60. Z. Liu, Y.-H. Xu and Y. Zhang, Probing dark matter particles at CEPC, *JHEP* **06**, p. 009 (2019).
61. M. Habermehl, M. Berggren and J. List, WIMP Dark Matter at the International Linear Collider, *Phys. Rev. D* **101**, p. 075053 (2020).
62. J. Kalinowski, W. Kotlarski, K. Mekala, P. Sopicki and A. F. Zarnecki, Sensitivity of future e^+e^- colliders to processes of dark matter production with light mediator exchange (7 2021).
63. B. Barman, S. Bhattacharya, S. Girmohanta and S. Jahedi, Catch 'em all: Effective Leptophilic WIMPs at the $e^+ e^-$ Collider (9 2021).
64. N. Wan, M. Song, G. Li, W.-G. Ma, R.-Y. Zhang and J.-Y. Guo, Searching for dark matter via mono-Z boson production at the ILC, *Eur. Phys. J. C* **74**, p. 3219 (2014).
65. Z.-H. Yu, X.-J. Bi, Q.-S. Yan and P.-F. Yin, Dark matter searches in the mono-Z channel at high energy e^+e^- colliders, *Phys. Rev. D* **90**, p. 055010 (2014).
66. S. Dutta, D. Sachdeva and B. Rawat, Signals of Leptophilic Dark Matter at the ILC, *Eur. Phys. J. C* **77**, p. 639 (2017).
67. B. Grzadkowski, M. Iglicki, K. Mekala and A. F. Zarnecki, Dark-matter-spin effects at future e^+e^- colliders, *JHEP* **08**, p. 052 (2020).
68. P. Bambade *et al.*, The International Linear Collider: A Global Project (3 2019).
69. M. J. Boland *et al.*, Updated baseline for a staged Compact Linear Collider (8 2016).
70. M. Dong *et al.*, CEPC Conceptual Design Report: Volume 2 - Physics & Detector (11 2018).
71. A. Abada *et al.*, FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2, *Eur. Phys. J. ST* **228**, 261 (2019).
72. T. Barklow, J. Brau, K. Fujii, J. Gao, J. List, N. Walker and K. Yokoya, ILC Operating Scenarios (6 2015).
73. J. P. Delahaye, M. Diemoz, K. Long, B. Mansoulié, N. Pastrone, L. Rivkin, D. Schulte, A. Skrinsky and A. Wulzer, Muon Colliders (1 2019).

74. S. Matsumoto, S. Mukhopadhyay and Y.-L. S. Tsai, Effective Theory of WIMP Dark Matter supplemented by Simplified Models: Singlet-like Majorana fermion case, *Phys. Rev. D* **94**, p. 065034 (2016).
75. A. Belyaev, N. D. Christensen and A. Pukhov, CalcHEP 3.4 for collider physics within and beyond the Standard Model, *Comput. Phys. Commun.* **184**, 1729 (2013).
76. A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, FeynRules 2.0 - A complete toolbox for tree-level phenomenology, *Comput. Phys. Commun.* **185**, 2250 (2014).
77. W. Kilian, T. Ohl and J. Reuter, WHIZARD: Simulating Multi-Particle Processes at LHC and ILC, *Eur. Phys. J. C* **71**, p. 1742 (2011).
78. H. Abramowicz *et al.*, The International Linear Collider Technical Design Report - Volume 4: Detectors (6 2013).
79. J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens and M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *JHEP* **02**, p. 057 (2014).
80. C. T. Potter, DSiD: a Delphes Detector for ILC Physics Studies (2 2016).
81. H. Abramowicz *et al.*, Forward Instrumentation for ILC Detectors, *JINST* **5**, p. P12002 (2010).
82. A. Guha, P. S. B. Dev and P. K. Das, Model-independent Astrophysical Constraints on Leptophilic Dark Matter in the Framework of Tsallis Statistics, *JCAP* **02**, p. 032 (2019).
83. L. Hamaide and C. McCabe, Fuelling the search for light dark matter-electron scattering (10 2021).
84. E. Aprile *et al.*, Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, *Phys. Rev. Lett.* **121**, p. 111302 (2018).
85. Y. Meng *et al.*, Dark Matter Search Results from the PandaX-4T Commissioning Run (7 2021).
86. R. K. Leane, T. R. Slatyer, J. F. Beacom and K. C. Y. Ng, GeV-scale thermal WIMPs: Not even slightly ruled out, *Phys. Rev. D* **98**, p. 023016 (2018).
87. G. Steigman, B. Dasgupta and J. F. Beacom, Precise Relic WIMP Abundance and its Impact on Searches for Dark Matter Annihilation, *Phys. Rev. D* **86**, p. 023506 (2012).
88. L. J. Hall, K. Jedamzik, J. March-Russell and S. M. West, Freeze-In Production of FIMP Dark Matter, *JHEP* **03**, p. 080 (2010).
89. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *JHEP* **07**, p. 079 (2014).
90. S. Frixione, E. Laenen, P. Motylinski and B. R. Webber, Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations, *JHEP* **04**, p. 081 (2007).
91. P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, *JHEP* **03**, p. 015 (2013).
92. T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, An introduction to PYTHIA 8.2, *Comput. Phys. Commun.* **191**, 159 (2015).
93. M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm, *JHEP* **04**, p. 063 (2008).
94. M. Cacciari, G. P. Salam and G. Soyez, FastJet User Manual, *Eur. Phys. J. C* **72**, p. 1896 (2012).

DM6 session: Dark matter and rare processes
– *Short summary* –

Conveners: R. Bernabei[a] and Z. Berezhiani[b]

[a] *Dipartimento di Fisica, Universitá di Roma Tor Vergata, 00133 Rome, Italy*
INFN sezione di Roma Tor Vergata, 00133 Rome, Italy
E-mail: rita.bernabei@roma2.infn.it
[b] *Dipartimento di Fisica e Chimica, Universitá di L'Aquila, 67100 Coppito, AQ, Italy*
INFN, Laboratori Nazionali del Gran Sasso, 67010 Assergi, AQ, Italy
E-mail: zurab.berezhiani@lngs.infn.it

Keywords: Dark matter; direct detection; rare processes.

This session has been dedicated to several topics of Dark Matter (DM) and Rare Processes (RP) field. Some talks have review character while others present some specific experimental results and developments. Mainly the contributions refer to Dark Matter (DM) topics while the last contribution discusses a specific rare process.

The first talk presented by M. Ricci (INFN-LNF) discusses the indirect signatures of DM from Space Satellite missions. Among the indirect searches for DM (complementary to the direct ones), the speaker has reviewed the possibilities of observing the DM signatures in the Cosmic Rays and in the gamma-ray sector, considering space missions like PAMELA, AMS, FERMI, DAMPE and the planned projects like GAPS (on stratospheric balloon) and HERD. An exhaustive overview on the current status and on the perspective of this field has been given.

The second talk presented by P. Belli (INFN-Roma Tor Vergata) is instead related to a direct detection on DM present at galactic scale. In fact, P. Belli has presented a detailed overview of the results of DAMA/LIBRA experiment and new results of two additional annual cycles, which further confirm the evidence of a signal that meets all the requirements of the model-independent DM annual modulation signature, at very high C.L.

A. Messina (INFN-Roma La Sapienza) in his talk discussed in details the known impact of the Migdal effect on the DM detection, mainly for candidates with GeV range masses, and pointed to some new ideas how one could measure such an effect with detectors based on Time Projection Chamber exposed to high neutron flux.

Talk by V. Caracciolo, (INFN-Roma Tor Vergata) was focused on the directionality approach which holds for DM candidates inducing just nuclear recoils. This approach is based on the study of the correlation between the nuclear recoils direction

and the Earth motion in the galactic rest frame. Several experimental techniques to explore the directionality approach have been proposed. In this talk, a review of such experimental approaches have been addressed with particular regard to the ADAMO project based on the use of anisotropic scintillators.

Talk of A. D'Angelo (INFN-Roma Tor Vergata) described the Heavy Photon Search (HPS) experiment which searches for an electro-produced dark photon using an electron beam provided by the CEBAF accelerator at the Thomas Jefferson National Accelerator Facility. It has been shown in some details that HPS looks for dark photons through two distinct methods, a resonance search in the e+e- invariant mass distribution above the large QED background (large dark photon-SM particles coupling region) and a displaced vertex search for long-lived dark photons (small coupling region). HPS employs a compact spectrometer, matched to the forward kinematic characteristics of A' electro- production. Three data taking periods took place in 2015, 2016 and 2019, while a fourth run has been scheduled in summer 2021. Results from the available data have been presented together with future projects.

The talk by A. Addazi (Sichuan University) has been focused in some details on the investigation in the laboratory and in the sky of the Mirror Dark Matter candidates. In particular, he discussed prospectives of laboratory search of the neutron to mirror neutron oscillations and its astrophysical implications, in particular for the neutron stars.

A. Zhitnitsky (University of British Columbia) discussed the Axion Quark Nugget Dark Matter Model, suggesting possible explanation of Telescope Array anomaly and for the positive dark matter signal of DAMA/LIBRA via the neutrons produced by quark nuggets in the rock. P. Belli has commented that already existing results disfavour such a hypothesis. It seems that other efforts will be pursued on that model.

B. Dev (Washington University in St. Louis) presented a detailed discussion on the possibilities to detect Leptophillic Dark Matter (LDM) at Linear Colliders. Such candidates could naturally arise in many beyond the Standard Model scenarios and might address certain experimental anomalies. Thus, some collider constraints on the LDM effective couplings with the Standard Model sector, considering its production at a future electron-positron linear collider (with polarized and unpolarized beam options) in the mono-photon and mono-Z channels.

Finally, C Curceanu (INFN-LNF) discussed experiment testing quantum mechanics in the Gran Sasso National Laboratories (LNGS): collapse models and spin-statistics. She has reported that they are experimentally investigating possible departures from the standard quantum mechanics' predictions, and - in particular - with radiation detectors they are searching signals predicted by the collapse models (spontaneous emission of radiation) which were proposed to solve the "measurement problem" in quantum physics and signals coming from a possible violation of the Pauli Exclusion Principle. Recent results published in Nature Physics under the title "Underground test of gravity-related wave function collapse", where they ruled out the parameter-free version of the gravity-related collapse model. More generic

results on testing Continuous Spontaneous Localization collapse models have been mentioned and future perspectives discussed. Finally, the VIP experiment looking for possible violations of the Pauli Exclusion Principle by searching for forbidden atomic transitions has been discussed.

This session was held in two subsequent afternoons and in both days the talks have been very interesting, the discussions were constructive and proceeded in exciting and serene atmosphere. The merit goes to the speakers and listeners and we give an account here. Finally, we remind that all the presentations are available on Youtube and that in most cases they refer to previously existing publications. Only some of the participants have sent the written contribution; we thank them.

Primordial black holes as dark matter candidates in the Galactic halo

Lindita Hamolli*, Mimoza Hafizi,

Department of Physics, University of Tirana, Tirana, 1016, Albania
**E-mail: lindita.hamolli@fshn.edu.al*

Francesco De Paolis, Achille A. Nucita

Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Via Arnesano, 73100 Lecce, Italy

INFN, Sezione di Lecce, Via Arnesano, 73100 Lecce, Italy

Gravitational microlensing is a powerful method to constrain the abundance of massive dark objects in the Milky Way halo. We calculate the optical depth and the microlensing rate for events caused by Primordial Black Holes (PBHs) eventually distributed in the Milky Way halo, towards some selected directions of observation, as the Galactic bulge, the Large and the Small Magellanic Clouds and the M31 galaxy. The capability of the Euclid space telescope to constraint the abundance of PBHs with mass $\geq 10^{-7}\ M_{\odot}$ in observation towards the Galactic bulge is also discussed.

Keywords: Primordial black holes; Gravitational microlensing; Galactic halo.

1. Introduction

Primordial black holes (PBHs) are a hypothetical type of black holes that formed soon after the Big Bang. Their existence was proposed in 1966 by Zel'dovich and Novikov[1] and by Stephen Hawking[2] in 1971 who studied in depth their origins. In the early universe, high densities and heterogeneous conditions could have led sufficiently dense regions to undergo gravitational collapse, forming black holes. There are several mechanisms able to produce such inhomogeneities. For all of these, the increased cosmological energy density at early times plays a major role, yielding a rough connection between the PBH mass and the formation time t (see Ref. 3):

$$M_{PBH} \simeq \frac{c^3 t}{G} \simeq 10^{15} \left(\frac{t}{10^{-23}\,s}\right) g \simeq 10^5 \left(\frac{t}{1\,s}\right) M_{\odot}. \tag{1}$$

Hence PBHs could span an enormous mass range: those formed at the Planck time ($10^{-43}s$) would have the Planck mass 10^{-5} g, whereas those formed at 1 s would be as large as 10^5 $M_{\odot}$, comparable to the mass of the holes thought to reside in galactic nuclei. By contrast, black holes forming at the present epoch (e.g. in the final stages of stellar evolution) could never be smaller than about $1M_{\odot}$.

Then, depending on the PBH formation epoch, black holes of different masses can form in the early universe.

However, primordial black holes originally having mass below about $10^{-18}\ M_\odot$ would not have survived to the present due to Hawking radiation, which causes complete evaporation in a time shorter than the age of the Universe.[3] Hawking thought that PBHs would have formed before the electroweak phase transition, which occurred at $t \simeq 10^{-12}$ s. In this case one finds $M_{PHB} \leq 10^{-7}\, M_\odot$. However, there is no reason in principle which forbids the formation of PBHs later on. In this case, the PBH mass can rise of orders of magnitude, even up to $\sim 10^5\, M_\odot$ or so. Certainly, only observations might be able to settle this open issue.

PBHs might lead to various interesting astrophysical consequences, such as providing seeds for supermassive black holes in galactic nuclei,[4,5] influence the generation of large-scale structure through Poisson fluctuations[6] and cause important effects on the thermal and ionization history of the Universe.[7] But, perhaps, the most exciting possibility is that they could provide a non-negligible fraction of the dark matter, which is known to constitute about 25% of the critical density of the Universe.[8]

There are some observational conundra that may be explained by PBHs. They are associated with lensing,[9] accretion,[10] dynamical[11] and gravitational wave (GW) effects.[12] Various constraints on their abundance allowed to conclude that PBHs in three ranges of masses could provide the dark matter:[13] the asteroid mass range ($10^{-17} - 10^{-16}\ M_\odot$), the sublunar mass range ($10^{-13} - 10^{-9}\ M_\odot$) and what is sometimes called the intermediate-mass black hole (IMBH) range ($10 - 10^3\ M_\odot$). In general, constraints from these observables have been computed under the assumption that all PBHs have the same mass, assuming a single Dirac delta function. However, an extended (non-monochromatic) mass function, which will correctly combine constraints across all masses, is necessary.

Concerning the mass-fraction f of PBHs in the Galactic halo, the currently available limits are summarized in Fig. 1 of Carr & Kuhnel.[14] The main constraints are derived from PBH evaporation, different gravitational-lensing experiments, numerous dynamical effects and PBH accretion. Even if many different methods have been proposed in the literature for detecting and constraining PBHs (see, e.g. Refs. [8, 15] and references therein)[a], gravitational microlensing offers the most powerful and robust method of constraining PBHs. By the Kepler microlensing observations of Galactic sources[17] a limit in the planetary mass range: $f < 0.3$ for $2 \times 10^{-9} M_\odot < M < 10^{-7} M_\odot$ was found, while the MACHO project established that $f \sim 0.1$ in the $[0.1, 1] M_\odot$ mass range.[18]

Recently, Niikura *et al.*,[9] using the largest sample of microlensing events for stars in the Galactic bulge (2622 microlensing data) obtained from the 5-years OGLE observation, provided that PBHs constitute about 1% of the DM in the MW. Moreover, from a seven hour lasting observation with the Subaru Hyper Suprime-Cam (HSC) towards the $M31$ galaxy, a single candidate event was identified, which

[a]In addition to the methods discussed in that paper we also mention that lensing by fast radio bursts (FRBs) can allow constraining PBHs with mass larger than about $2\, M_\odot$.[16]

translates into a rather stringent upper bounds on the abundance of PBHs in the mass range[19] $[10^{-11} \div 10^{-6}]\ M_{\odot}$.

Gravitational microlensing is the best method to detect objects that do not emit light since this phenomenon may occur when a compact object (lens) passes close enough to the line of sight toward a source star.

In the simplest case, when both the lens and the source are considered as the point-object[b], individual images cannot be resolved (the image separation is the order of $\mu arcsec$), but the total brightness of the images increases with respect to that of the unlensed source, leading to a specific time-dependent amplification of the source star brightness, that is

$$A = \frac{2 + u^2}{u\sqrt{4 + u^2}}. \tag{2}$$

Here, $u(t)$ is the distance, in units of the Einstein radius, between the source star and the lens at an observation epoch t, and is given by

$$u(t) = \sqrt{u_0^2 + \left(\frac{t - t_0}{t_E}\right)^2}, \tag{3}$$

where t_0 is the peak time of the event (at the time of closest approach) and u_0 is the impact parameter, whereas $t_E = R_E/v_T$ is the Einstein timescale, which is defined as the time required for the lens to transit the Einstein radius. The Einstein radius is the radius of the ring image formed when the observer, lens and source are perfectly aligned. It is given by, $R_E = \sqrt{\frac{4GMD_S}{c^2}x(1-x)}$, where M is the mass of the lens and $x = D_L/D_S$ is the normalized lens distance, whereas D_S and D_L are the source-observer and lens-observer distances, respectively. We also note that v_T is the relative transverse velocity between the lens and the source. During a microlensing event, when the source position projected in the lens plane encounters the Einstein ring, the projected separation is $u_T = 1$ and the source threshold amplification is $A_T = 1.34$ (a microlensing event occurs when the source trajectory crosses the Einstein ring). The event amplification takes the maximum value for the smallest separation between the source and lens, $u = u_0$. By photometric measurements of the event lightcurve, three parameters can be defined: t_0, t_E and u_0. However, among these parameters, only t_E contains information about the lens and this gives rise to the so-called parameter degeneracy problem, which does not allow to infer the lens parameters uniquely. To break this degeneracy, second-order effects, as the finite source effects[21] and the parallax effect,[22] can be considered. Also, it is well known that a gravitational microlensing event gives rise to an astrometric shift of the source,[23] which may be extremely useful to break, at least partially, the parameter degeneracy problem in microlensing observations.[24]

[b]In our analysis we also assume that the relative motion among the observer, lens, and source is uniform (with constant velocity) and linear (on a straight line). See the discussion on this issue in Ref. 20 and references therein.

The goal of this work is to trace the PBHs in Galactic halo (if PBHs contribute to the Galactic halo DM) via microlensing observations. For this purpose we selected four directions (Galactic bulge, LMC, SMC and $M31$) and calculate the optical depth and microlensing rate towards them considering a wide range of possible masses for the PBHs. Moreover, we forecast the number of events caused by PBHs and observable by the Euclid telescope. It is a Medium Class mission of the ESA (European Space Agency), scheduled to be launched in 2022 at the L2 Lagrangian point. For it an additional survey is planned towards Galactic bulge, although not definitely approved as yet.[25] The photometric error of Euclid is expected to be less then[26] $\sim 0.1\%$, so the threshold amplification turns out to be $A_T \simeq 1.001$ and from Eq. (2) one can find $u_T \simeq 6.54$.

Until now, several microlensing surveys have been undertaken towards the Galactic bulge (in particular MOA and $OGLE$), $M31$ ($MEGA$,[27] $AGAPE$,[28] $WeCAPP$,[29] $PLAN$[30]) and the Large and the Small Magellanic Clouds (in particular $MACHO$, $EROS$,[31] $SuperMACHO$[32]). However, due to the way these surveys were designed, they were not able to give reliable constraints on PBHs with mass $\leq 10^{-6}\ M_\odot$ in the Galactic halo.

The plan of the paper is as follows: in Section 2, we review the models used in our analysis, for PBHs in Galactic halo and for the source stars in the MW bulge, LMC, SMC and M31. In Section 3 we give some basics for microlensing quantities. In Section 4 we present and discuss the results obtained for the optical depth and microlening rate towards four selected targets and Euclid field of view. Finally, our conclusions are drawn in Section 5.

2. MODEL

The microlensing quantities, the optical depth, the microlensing rate and Einstein timescale depend on the underlying astrophysical model. Below we provide some details for the mass density of the halo dark matter in our Galaxy[c] and the distribution of the stellar components in some favorable directions as the MW bulge, LMC, SMC and the $M31$ galaxy.

2.1. *Distribution of dark matter in Galactic halo*

For the DM in the MW halo region we assume the standard halo model (S), which consists of a cored isothermal sphere with dark matter density given by:

$$\rho_{DM}(r) = \rho_0 \frac{r_c^2 + r_0^2}{r_c^2 + r^2}, \tag{4}$$

[c]Since we are interested in microlensing observations with sources in our Galaxy we neglect[30] the presence of dark matter associated to the targets. In the case of the M31 galaxy this choice, of course, will results in a lower limit of the optical depth, microlensing rate and expected number of events.

where r is the galactocentric radius, $\rho_0 \simeq 0.0079\, M_\odot\, pc^{-3}$ is the local dark matter density, $r_c \simeq 5.6$ kpc is the halo dark matter core radius and $r_0 \simeq 8.5$ kpc is the galactocentric distance of the Sun.[33] Considering a scenario in which PBHs constitute some mass-fraction f of the DM, the PBH mass density distribution is $\rho_{PHB}(r) = f\rho_{DM}(r)$.

2.2. *Source star distributions*

To trace the microlensing events caused by PBHs in Galactic halo we selected some favorable directions for such observations, as the MW bulge, LMC, SMC and the $M31$ galaxy.

2.2.1. *Sources in Galactic bulge*

For the mass density distribution of the stellar population in the Galactic bulge we use the triaxial bulge model given by

$$\rho_b = \frac{M_b}{8\pi abc} exp\left(-\frac{s^2}{2}\right). \tag{5}$$

Here $s^4 = (x^2/a^2 + y^2/b^2)^2 + z^4/c^4$, the scale length values are $a = 1.49$ kpc, $b = 0.58$ kpc and $c = 0.40$ kpc (we remind that the coordinates x and y span the galactic disk plane, whereas z is perpendicular to it) and the bulge mass is $M_b \approx 2 \times 10^{10}\, M_\odot$.[34]

2.2.2. *Sources in the LMC*

The source star distribution in the LMC is generally described as a luminous disk and a bar.[35] The distance between the center of the LMC and the observer is $D_{LMC} \simeq 50$ kpc and the stellar disk is modeled with a double exponential profile:

$$\rho_{disk} = \frac{M_d}{4\pi z_d R_d^2} \exp(-R/R_d - |z|/z_d), \tag{6}$$

where $R = \sqrt{x^2 + y^2}$. $R_d \simeq 1.6$ kpc is the radial scale length, $z_d \simeq 0.3$ kpc is the vertical scale height and $M_d = [2.4, 4.8] \times 10^9\, M_\odot$ is the LMC disk mass. The galactic coordinates for the center of the LMC disk are $l = 297.7°$, $b = -33.5°$. The LMC disk is inclined by an angle $i = 30°$ with respect to our line of sight and its position angle is $\phi = 170°$. In addition to the disk, the LMC possess a stellar bar described by a triaxial Gaussian density profile:

$$\rho_{bar} = \frac{M_{bar}}{(2\pi)^{3/2} x_b y_b z_b} \exp\left(-\frac{1}{2}\left[\left(\frac{x}{x_b}\right)^2 + \left(\frac{y}{y_b}\right)^2 + \left(\frac{z}{z_b}\right)^2\right]\right), \tag{7}$$

where x, y, z are coordinates along the principal axes of the bar. Here, $x_b = 1.0$ kpc, $y_b = 0.3$ kpc, $z_b = 0.3$ kpc are the scale lengths along the three axes and $M_{bar} = [0.6, 1.2] \times 10^9 M_\odot$ is the total mass of the bar. The bar is inclined by an angle $i = 30°$ and has a position angle $\phi = 120°$. The galactic coordinates for the center of the LMC stellar bar are $l = 280.5°, b = -32.8°$.

2.2.3. *Sources in the SMC*

Accordingly to the star formation history of the SMC, two stellar components can be distinguished: an old star (OS) and a young star (YS) population (see Ref. 36). For the YS mass distribution, we adopt a spheroidal model with a fully Gaussian profile:

$$\rho_{SMC}^{(YS)} = \rho_0^{(YS)} \exp\left(-\frac{1}{2}\left[\left(\frac{\xi}{\xi_{YS}}\right)^2 + \left(\frac{\eta}{\eta_{YS}}\right)^2 + \left(\frac{\zeta}{\zeta_{YS}}\right)^2\right]\right). \tag{8}$$

As far as the OS mass distribution is concerned, we keep the Gaussian profile along the line of sight, and a smoother exponential profile in the orthogonal plane:

$$\rho_{SMC}^{(OS)} = \rho_0^{(OS)} \exp\left[-\sqrt{\left(\frac{\Xi}{\Xi_{OS}}\right)^2 + \left(\frac{\Upsilon}{\Upsilon_{OS}}\right)^2}\right] \exp\left[-\frac{1}{2}\left(\frac{Z}{Z_{OS}}\right)^2\right]. \tag{9}$$

The central density values for the YS and OS population are $\rho_0^{(YS)} = 8.5 \times 10^6\, M_\odot\, kpc^{-3}$ and $\rho_0^{(OS)} = 3.9 \times 10^7\, M_\odot\, kpc^{-3}$, respectively. The value of the position angle, with respect to the north direction, is fixed at $\phi_{YS} = 66°$ and $\phi_{OS} = 83°$ for the YS and OS populations, respectively. The reference frames (ξ, η, ζ) and (Ξ, Υ, Z) are directed along the principal axes of the YS and OS spheroid, respectively. For the YS population, we fix $(\xi_{YS}, \eta_{YS}, \zeta_{YS}) = (0.8, 3.5, 1.8)$ kpc and for the OS population, $Z_{OS} = 2.1$ kpc, $(\Xi_{OS}, \Upsilon_{OS}) = (0.8, 2.1)$ kpc. The galactic coordinates for the SMC are $l = 307°, b = -46°$. The YS population is inclined by the angle $i = 74°$, while the OS population is not inclined. The OS and YS populations have a relative distance shift between them, since the centre of the YS population is at a distance of about 2 kpc behind that of the OS component, which is at a distance of $\simeq 63.5$ kpc from the Earth.

2.2.4. *Sources in the M31 galaxy*

The morphology of M31 is similar to that of the Milky Way, with a central bulge and a disc. The distance between the center of the M31 and the observer is $D_{M31} \simeq 778$ kpc. The M31 bulge is parameterized by a flattened power law of the form

$$\rho_B(R, z) = \rho_B(0)\left[1 + \left(\frac{R}{a}\right)^2 + q^{-2}\left(\frac{z}{a}\right)^2\right]^{-s/2}, \tag{10}$$

where the coordinates x and y span the M31 disk plane (z is perpendicular to it), $\rho_B(0) \simeq 4.5 \times 10^9 M_\odot kpc^{-3}$, $q \simeq 0.6$ is the ratio of the minor to major axes, $a \simeq 1.0$ kpc is the core radius and $s = 3.8$ is the power-law index.[37] The mass density of the M31 disk stars adopting the parameters of the Reference model in Ref. 38 is described by a sech-squared profile as follows:

$$\rho_D(R, z) = \rho_D(0) \exp(-R/h) \operatorname{sech}^2(z/H), \tag{11}$$

where R is the distance to the center of the M31 disk, $H \simeq 0.3$ kpc and $h \simeq 6.4$ kpc are, respectively, the scale height and scale lengths of the disk and $\rho_D(0) \simeq 3.5 \times 10^8\, M_\odot kpc^{-3}$ is the central mass density. The M31 disk is inclined by the angle $i = 77^\circ$ to our line of sight and has position angle $\phi = 38.6^\circ$.

3. Microlensing events by PBHs

The microlensing events are unpredictable phenomena, and to quantify their rarity we present the optical depth and the microlensing rate.

3.1. *The optical depth*

The microlensing optical depth τ is the probability that any given star is being significantly lensed at any given time by a foreground compact object. For a star and a given model of the lens density distribution one can calculate τ through the equation

$$\tau = \int_0^{D_S} n(D_l)\pi u_T^2 R_E^2 dD_l \tag{12}$$

which is just the probability that a lens object is inside the microlensing tube (see, e.g., Refs. 39 and 40). Here, $n = \rho/M$ is the number density of the lenses (ρ is their mass density) and M is the lens mass. Since $R_E^2 \propto M$, it turns out that the optical depth is simply a geometrical quantity and does not depend on the lens mass.

The sources are not all at the same distance D_S, but are distributed along the line of sight, so we define the average optical depth as:[41]

$$< \tau >= \left[\int_{D_{S,min}}^{D_{S,max}} dD_S n(D_S) D_S^2\right]^{-1} \int_{D_{S,min}}^{D_{S,max}} dD_S n(D_S) D_S^2 \tau(D_S) \tag{13}$$

where $D_{S,min}$ and $D_{S,max}$ are the minimum and maximum distances (up to the boundary of the observation region) from the observer's position and $n(D_S)$ is the number density distribution of source stars. Considering the above mentioned scenario, that PBHs constitutes some mass fraction f of the halo DM in the MW halo region, where $f = \frac{\Omega_{PHB}}{\Omega_{DM}} = \frac{\rho_{PHB}}{\rho_{DM}}$, the optical depth due to the PBHs can be written in the form $< \tau >_{PBH} = f < \tau >_{DM}$.

3.2. *The microlensing rate*

The event rate is the rate at which a given background star undergoes a microlensing event due to a foreground lensing object. The microlensing event rate for any given source is extremely small, but it can be maximized toward regions with a high surface density of sources, as the Galactic bulge, LMC, SMC and the $M31$ galaxy.

In order to define the microlensing rate, we consider the geometry and variables defined in Fig. 4 of Griest,[39] which gives the differential event rate ($d\Gamma$) of a lensing

object entering a volume element along the line-of-sight, where the lens causes a microlensing with magnification above a certain threshold value. Integrating it along the line of sight to the source, the microlensing rate per background star is defined as

$$\Gamma = \int d\Gamma = \int \frac{n(D_l) f(\vec{v}) d^3x d^3v}{dt}, \tag{14}$$

where $n(D_l)$ is the number density of the lenses, $f(\vec{v})$ is the lens velocity distribution, $d^3x = v_\perp cos(\theta) dt R_E u_T d\alpha d(D_l)$ is the volume element and $d^3v = d^2v_\perp dv_\parallel$ the velocity element (see the Ref. 35 for details). The perpendicular component of the velocity element in cylindrical coordinates can be then written as $d^2v_\perp = v_\perp dv_\perp d\theta$. Therefore, Eq. (14) can be rewritten as:

$$d\Gamma = \frac{\rho}{M} u_T R_E dD_l f(v_\perp, v_\parallel) d^2v_\perp dv_\perp d\theta \cos(\theta) d\alpha dv_\parallel. \tag{15}$$

We note that the parameters $\theta, \alpha, v_\perp$ in the equation above vary in the range $[-\pi/2, \pi/2]$, $[0, 2\pi]$ and $[0, \infty)$, respectively. Considering the velocity distribution of the lenses and source to be a Maxwellian distribution with dispersion velocity σ_l and σ_s and including the transverse velocity of the microlensing tube $v_{t\perp}$ one gets

$$\begin{aligned}\Gamma(D_S) = \frac{2u_T D_S}{M} \int_0^1 dx \rho(x) R_E(x) \times \int_0^\infty \frac{v_{s\perp}}{\pi\sigma_s^2} exp(-\frac{v_{s\perp}^2}{\sigma_s^2}) dv_{s\perp} \\ \times \int_0^\infty \frac{v_\perp^2}{\pi\sigma_l^2} exp(-\frac{v_\perp^2 + v_{t\perp}^2}{\sigma_l^2}) dv_{l\perp} \times \int_0^{2\pi} exp(-\frac{2v_\perp v_{t\perp} cos\beta}{\sigma_l^2}) d\beta \times \int_0^{2\pi} d\phi.\end{aligned} \tag{16}$$

Using the variables $y = v_{l\perp}/\sigma_l$, $\eta = v_{t\perp}/\sigma_l$ and $z = v_{s\perp}/\sigma_s$, the microlensing rate is

$$\begin{aligned}\Gamma(D_S) = 4u_T D_S \sqrt{\frac{4GMD_s}{c^2}} \frac{\sigma_l}{\pi M} \int_0^1 dx \rho(x) \sqrt{x(1-x)} \times \int_0^\infty z exp(-z^2) dz \\ \times \int_0^\infty y^2 exp(-y^2 - \eta^2) dy \times \int_0^{2\pi} exp(-2y\eta cos\beta) d\beta.\end{aligned} \tag{17}$$

Finally, the average microlensing rate per star can be rewritten as (see, e.g.,[41] for further details)

$$< \Gamma >= \left[\int_{D_{S,min}}^{D_{S,min}} dD_S n(D_S) D_S^2 \right]^{-1} \int_{D_{S,min}}^{D_{S,min}} dD_S n(D_S) D_S^2 \Gamma(D_S), \tag{18}$$

In the same way, the microlensing rate due to the PBHs can be written in the form $< \Gamma >_{PBH} = f < \Gamma >_{DM}$.

The median microlensing timescale is related to the optical depth and the microlensing rate[40] through the relation

$$t_E = \frac{2}{\pi} \frac{\tau}{\Gamma}. \tag{19}$$

Since τ does not depend on the lens mass and Γ scales with the PBH mass as $M^{-1/2}$, the microlensing timescale t_E turns out to be proportional to the square root of the PBHs mass.

4. Model results

Let us now consider that a mass-fraction f of the Galactic halo DM is in the form of PBHs, assumed to have a mass density distribution exactly as that of the halo DM.

First, we consider a monochromatic mass distribution for the PBHs, so we assume that all PBHs have the same mass. We calculate the optical depth and the microlensing rate for this model, varying masses in the range $10^{-14}\,M_\odot \leq M \leq 10^2\,M_\odot$ (see, e.g. Ref. 42 and references therein). Using the Galactic celestial coordinate variables (l, b), an object at the distance d from the observer (Sun's position) is at a distance $r = \sqrt{r_0^2 - 2r_0 d\cos l\cos b + d^2}$ from the Galactic center, where $r_0 \simeq 8.5$ kpc is the distance between the Sun and Galactic center, $d = \sqrt{x'^2 + y'^2 + z'^2}$ and $x' = d\cos b\cos l$, $y' = d\cos b\sin l$, $z' = d\sin b$. In the Cartesian coordinate system (x', y', z'), the observer is at the coordinate origin. Since we are interested in estimating the optical depth and microlensing rate due to the PBHs in the Galactic halo towards three nearby galaxies, i.e. LMC, SMC and $M31$, we introduce the coordinate system (x_0, y_0, z_0), which has the origin in the center of these galaxies at (l_0, b_0, D_0) and has the $z_0 - axis$ directed toward the observer: the $x_0 - axis$ is assumed to be anti parallel to the galactic longitude axis and the $y_0 - axis$ parallel to the galactic latitude axis. For an object at distance D from the observer, a point with galactic coordinates (l, b) can be defined through the coordinates (x_0, y_0, z_0) as:

$$\begin{aligned} x_0 &= -D\cos b\sin(l - l_0) \\ y_0 &= D\sin b\cos b_0 - D\cos b\sin b_0\cos(l - l_0) \\ z_0 &= D_0 - D\cos b\cos b_0\cos(l - l_0) - D\sin b\sin b_0 \end{aligned} \tag{20}$$

Therefore, in order to find the coordinates (x, y, z) according to the selected systems, in the case of LMC and M31 they are obtained by rotating around $z_0 - axis$ with position angle ϕ counterclockwise and around the new $x - axis$ with the inclination angle i clockwise. In the case of SMC, the coordinates (Ξ, Υ, Z) for the OS population and (ξ, η, ζ) for the YS population are obtained by rotating around $z_0 - axis$ with the position angle ϕ counterclockwise and around the new $x - axis$ with the inclination angle i counterclockwise.

In particular, the optical depth depends uniquely on the lens (and source) population spatial density, whereas the microlensing rate depends also on the lens mass function and the lens–source relative velocity.

4.1. *Optical depth*

Based on Eq. (13) we estimate the average optical depth due to the PBHs towards the Galactic bulge, LMC, SMC and the M31 galaxy. In Table 1 are given the values of the optical depth towards the Galactic bulge, LMC, SMC and M31 galaxy. The galactic coordinate are: $(l_0, b_0) = (0^\circ, 0^\circ)$ for the Galactic bulge, $(l_0, b_0) = (280.5^\circ, -32.8^\circ)$ for LMC, $(l_0, b_0) = (307^\circ, -46^\circ)$ for SMC and $(l_0, b_0) = (121.17^\circ, -21.57^\circ)$ for M31. For definiteness, the values are given for $f_{PBH} = 1$ and $u_T = 1$. In case of other values, the rescaling is obviously a multiplication by f_{PBH} and by u_T^2.

As one can see from Table 1, the largest values of the optical depth are found towards M31 ($\tau \sim 10^{-6}$), with a factor about 20 times larger than those towards the Galactic bulge. This can be certainly attributed to the enormous volume and large mass content between the Earth and M31.

Table 1. The average optical depth, microlensing rate (per source per year) and the Einstein timescale due to PBHs in Galactic halo. From top to bottom, the values are given towards the MW bulge, LMC disk and bar, SMC young and old stellar populations and M31 bulge and disk. For the microlening rate and microlensing timescale t_E are selected two values of the PBH mass M, i.e. $M = 10^{-14}\, M_\odot$ and $M = 10^2\, M_\odot$.

	$<\tau>$ ($\times 10^{-7}$)	$<\Gamma>(10^{-14}\, M_\odot)$	$<\Gamma>(10^2\, M_\odot)$ ($\times 10^{-8}$)	$t_E(10^{-14}\, M_\odot)$ (s)	$t_E(10^2\, M_\odot)$ (year)
MW_{bulge}	1.30	8.14	8.14	0.32	1.02
LMC_{disk}	4.84	9.96	9.96	0.98	3.10
LMC_{bar}	4.84	9.97	9.97	0.98	3.09
SMC_{YS}	7.26	15.7	15.7	0.93	2.95
SMC_{OS}	7.11	16.4	16.4	0.87	2.76
$M31_{bulge}$	14.25	20.64	20.64	1.39	4.40
$M31_{disk}$	14.25	20.35	20.35	1.41	4.46

One can also remark that the optical depth towards SMC is significantly larger than that towards LMC. This has to be attributed mainly to the larger value of the SMC center longitude with respect to the corresponding value of LMC, but also to its larger distance. The ratio of the optical depth towards SMC and LMC turns out to be $<\tau>_{SMC} \simeq 1.4\ <\tau>_{LMC}$, consistent with the results to Ref. 43. The LMC optical depth is about 5×10^{-7}, in agreement to the result in Ref. 39, in which the adopted values of the model parameters are slightly different.

We also remark that the optical depth towards the Galactic bulge is compatible with the results obtained in Ref. 9 (see e.g., Table I).

Concerning SMC, there is a more visible difference between the role of the two star populations, mainly due to the fact that their centers are shifted by about 2 kpc one with respect to the other.

4.2. *Microlensing rate*

Based on Eq. (18), we estimate the average microlensing rate varying the PBH masses in the range $10^{-14}\,M_\odot \leq M \leq 10^2\,M_\odot$ considering the same conditions as in the previous section. Concerning the transverse velocity dispersion, we keep the value $\sigma_l = 210$ km/s as it is considered for the lenses in the MW halo.[35] For the source star populations in the MW bulge, the velocity dispersion is assumed to be $\sigma = 156$ km/s, whereas in the case of the M31 galaxy, we consider[37] $\sigma_{bulge} = 100$ km/s and $\sigma_{disk} = 30$ km/s. For LMC, we have $\sigma_{disk} = 20.2$ km/s and $\sigma_{bar} = 24.7$ km/s, while for SMC,[44] $\sigma_{YS} = 20$ km/s and $\sigma_{OS} = 30$ km/s.

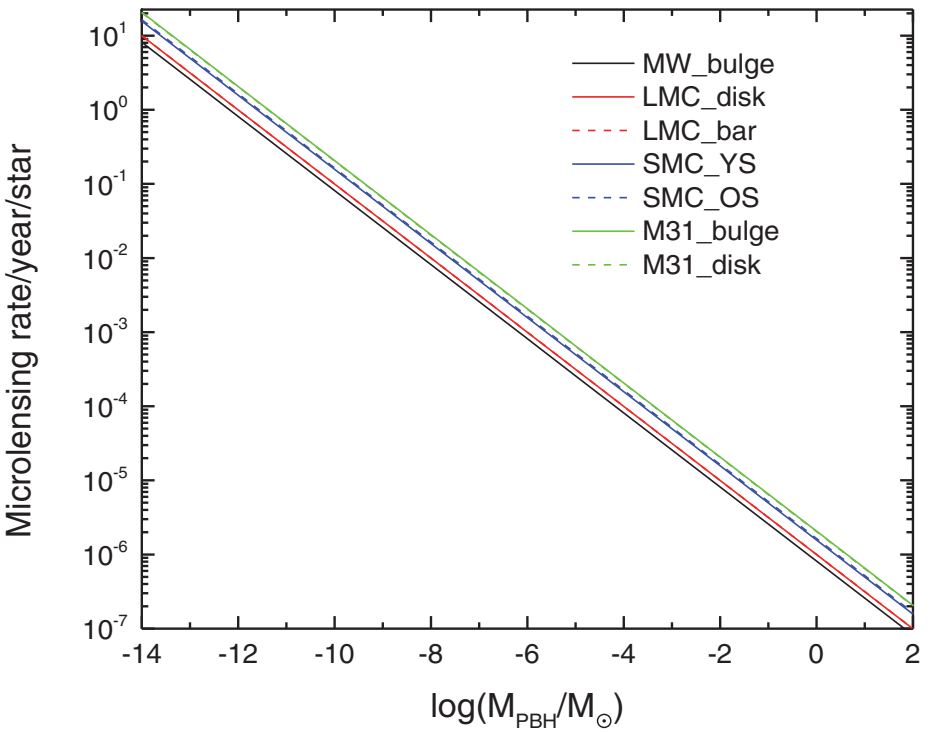


Fig. 1. The microlensing rate caused by our Galactic halo PBHs to the stars in MW bulge (black line), LMC (red line), SMC (blue line), and M31 (green line) is plotted as a function of the PBH mass.

In Fig. 1 we plot the PBH microlensing rate per star per year towards the MW bulge, LMC, SMC and M31 as a function of the PBH mass. We also remind that the plots are given for $f_{PBH} = 1$ and $u_T = 1$. In case of other values, the rescaling is obviously obtained with a multiplication by f_{PBH} and u_T. As one can see, the value of the microlensing rate vary from 10^{-7} for $M \simeq 10^2\,M_\odot$ up to

10^1 for $M \simeq 10^{-14}\, M_\odot$. The highest value of the microlensing rate is obtained in the case of observations towards the M31 galaxy, this has to be attributed to the dark matter distribution function as a cored isothermal sphere. Indeed, microlensing observations towards M31 are able to allow a full map of the Galactic halo, whereas the other considered targets are immersed within the Galactic halo.

Based on the Eq. (19) we define the microlensing timescale of the events caused by the PBHs in Galactic halo towards four selected targets. In Table 1 we present the Einstein timescale for events caused by PBHs towards the center of the MW bulge, LMC, SMC and M31 for two extremal cases of $M = 10^{-14}\, M_\odot$ and $M = 10^2\, M_\odot$ of the PBH mass. As expected, due to the wide mass interval in consideration, the t_E values vary by several orders of magnitude, from a fraction of a second up to several years. For that reason, since a sufficiently high image sampling is requested in order to recognize a microlensing event (let's say at least 5-10 images within t_E are necessary for that aim), different observing strategies and facilities are needed in order to probe the full PBH mass range.

4.3. *PBHs by Euclid*

Since an additional survey of the Euclid telescope, although not definitely approved as yet, is to observe for a few months the Galactic bulge, we devote a special attention to its survey towards the Galactic bulge.[25] Based on the galactic coordinates of the line of sight $(l, b) = (1.1^\circ, -1.7^\circ)$ and the distributions of the PBHs and bulge stars we find the optical depth to be $< \tau > \simeq 1.30 \times 10^{-7}$ and the microlensing rate per stars per year to be $< \Gamma > = 8.12 \times 10^{-7}$ for the PBH mass, $M = 1\, M_\odot$. Based on the magnitude limit of the Euclid telescope, which is foreseen to be $m = 24$, the mass-luminosity relation $\frac{L}{L_\odot} = (\frac{M}{M_\odot})^{3.5}$ and considering the middle point 8.5 kpc away from us, we find the lowest value $M_{min} = 0.31 M_\odot$ of the star's mass to be detected by this telescope. Using a Salpeter mass function $\frac{dN}{dM} \propto M^{-2.4}$, for stars in the Galactic bulge, we calculate their mean mass value $< M > = 0.97\, M_\odot$. Considering the distance of the bulge stars to be within the range $[7, 10]$ kpc with mass density distribution given by Eq. (5) and since the Euclid's field of view is 0.54 square degree we obtain that the number N_{ED} of detectable source stars in the Euclid microlensing observations is $N_{ED} = 6.45 \times 10^7$. By multiplying $< \Gamma > N_{ED} t_{obs}$, where t_{obs} is the observation timescale in years, we find about 340 microlensing events per year (for $f_{PBH} = 1$ and $u_T = 6.54$), with a microlensing timescale $t_E \sim 8$ months in the case of a PBH mass of $1\, M_\odot$. In Fig. 2 we plot the PBH microlensing rate and microlensing timescale towards the Euclid field of view as a function of the PHB mass. We remind that the plot of the microlensing rate is given for $f_{PBH} = 1$, $u_T = 6.54$ and the microlensing timescale for $u_T = 6.54$, the last one is independent of f_{PBH}.

Considering a cadence of 20 minutes[25] and no less than 5 points within t_E, we find that Euclid could detect PBHs down to $M = 10^{-7} M_\odot$.

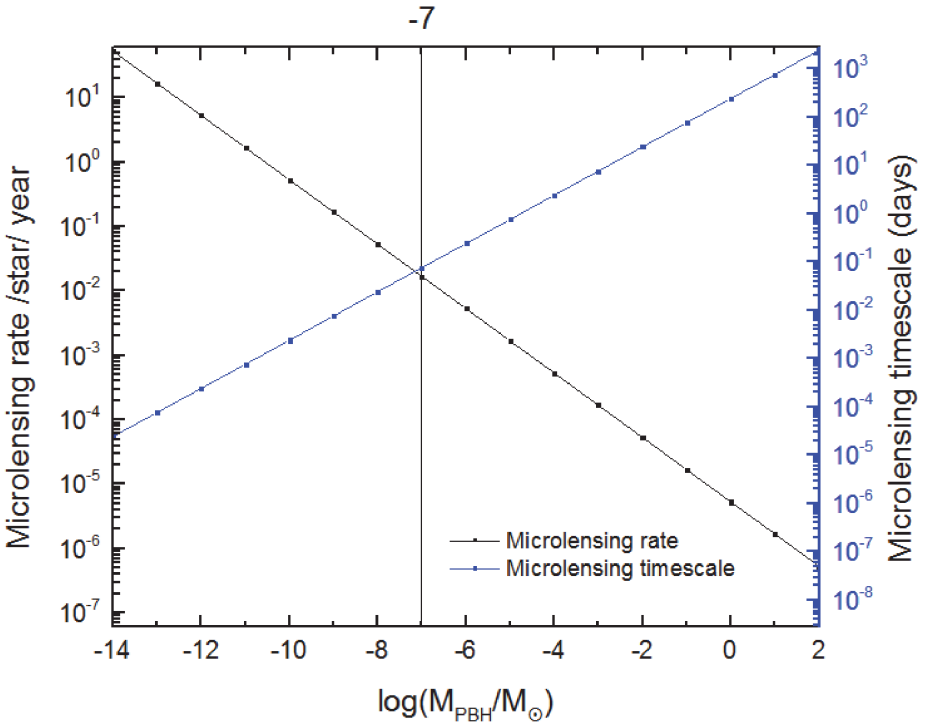


Fig. 2. The microlensing rate (black line) and the microlensing timescale (blue line) caused by our Galactic halo PBHs to the stars in Euclid field view as a function of the PBH mass. The vertical line at $M_{PBH} = 10^{-7}\ M_{\odot}$ represents the lower mass limit for observations towards the Galactic bulge by the Euclid telescope.

5. Conclusions

To obtain valuable information about the mass function and the spatial distribution of PBHs, if they contribute to a high enough fraction of the Galactic halo dark matter, the gravitational microlensing is the best method. We use this method to investigate the traces of MW halo's PBHs towards four selected targets: the MW bulge, Large and Small Magellanic Clouds and the M31 galaxy.

We estimate the optical depth and the microlensing rate towards these targets, assuming for PBHs a standard halo model and for the target sources, their mass density distributions. The largest values of both the optical depth and the microlensing rate are found for observations towards the M31 galaxy, due to its larger distance.

We calculate and discuss the microlensing timescale, which changes by many orders of magnitude, for the reason of the high mass range $[10^{-14}, 10^{2}]M_{\odot}$ of PBHs, taken into consideration. It varies from less than 1 seconds up to about 4.5 years. Therefore, completely different observation strategies are necessary in order to

detect PBHs in different bins of masses. We also mention that the present paper is a summary of a more detailed work.[45]

As far as the Euclid observations towards the Galactic bulge is concerned, and based on its features, we define that it will observe and will detect about 340 microlensing events per year (for mass-fraction $f_{PBH} = 1$ and $u_T = 6.54$), with a microlensing timescale $t_E \sim 8$ months in the case of a PBH mass of $1\,M_\odot$. Also, we find that it could detect PBHs with mass down to $M = 10^{-7} M_\odot$. Below this mass limit, the expected number of points in the lightcurve is insufficient to perform a reliable analysis. For PBH masses around this value we estimated a number of about 10^6 events per year, since the microlensing rate varies as $1/\sqrt{M}$.

Since a confusing element in the observed stars is the stellar intrisic variability, it needs a special attention. In fact, this can be avoided by carefully looking at different bands: gravitational microlensing is an achromatic effect, while stellar variability appears different in different spectral bands.

Acknowledgements

FDP and AAN acknowledge the support by the INFN projects TAsP and Euclid.

References

1. Y. B. Zel'dovich and I. Novikov, The hypothesis of cores retarded during expansion and the hot cosmological model, *Astronomicheskii Zhurnal* **43**, p. 758 (1966).
2. S. Hawking, Gravitationally collapsed objects of very low mass, *Monthly Notices of the Royal Astronomical Society* **152**, 75 (1971).
3. B. J. Carr and S. W. Hawking, Black holes in the early universe, *Monthly Notices of the Royal Astronomical Society* **168**, 399 (1974).
4. R. Bean and J. Magueijo, Could supermassive black holes be quintessential primordial black holes?, *Physical Review D* **66**, p. 063505 (2002).
5. J. García-Bellido, Massive primordial black holes as dark matter and their detection with gravitational waves, in *Journal of Physics: Conference Series*, (1)2017.
6. N. Afshordi, P. McDonald and D. Spergel, Primordial black holes as dark matter: the power spectrum and evaporation of early structures, *The Astrophysical Journal Letters* **594**, p. L71 (2003).
7. M. Ricotti, J. P. Ostriker and K. J. Mack, Effect of primordial black holes on the cosmic microwave background and cosmological parameter estimates, *The Astrophysical Journal* **680**, p. 829 (2008).
8. B. Carr, F. Kühnel and M. Sandstad, Primordial black holes as dark matter, *Physical Review D* **94**, p. 083504 (2016).
9. H. Niikura, M. Takada, S. Yokoyama, T. Sumi and S. Masaki, Constraints on earth-mass primordial black holes from ogle 5-year microlensing events, *Physical Review D* **99**, p. 083503 (2019).
10. N. Cappelluti, A. Kashlinsky, R. Arendt, A. Comastri, G. Fazio, A. Finoguenov, G. Hasinger, J. Mather, T. Miyaji and S. Moseley, Cross-correlating cosmic infrared and x-ray background fluctuations: Evidence of significant black hole populations among the cib sources, *The Astrophysical Journal* **769**, p. 68 (2013).

11. S. Clesse, J. García-Bellido and S. Orani, Detecting the stochastic gravitational wave background from primordial black hole formation, *arXiv preprint arXiv:1812.11011* (2018).
12. B. Abbott, S. Jawahar, N. Lockerbie and K. Tokmakov, Ligo scientific collaboration and virgo collaboration (2016) directly comparing gw150914 with numerical solutions of einstein's equations for binary black hole coalescence. physical review d, 94 (6). issn 1550-2368, http://dx.doi.org/10.1103/physrevd.94.064035, *PHYSICAL REVIEW D Phys Rev D* **94**, p. 064035 (2016).
13. A. Barrau, D. Blais, G. Boudoul and D. Polarski, Peculiar relics from primordial black holes in the inflationary paradigm, *Annalen der Physik* **13**, 115 (2004).
14. B. Carr and F. Kühnel, Primordial black holes as dark matter: recent developments, *Annual Review of Nuclear and Particle Science* **70**, 355 (2020).
15. A. M. Green and B. J. Kavanagh, Primordial black holes as a dark matter candidate, *Journal of Physics G: Nuclear and Particle Physics* **48**, p. 043001 (2021).
16. R. Laha, Lensing of fast radio bursts: Future constraints on primordial black hole density with an extended mass function and a new probe of exotic compact fermion and boson stars, *Physical Review D* **102**, p. 023016 (2020).
17. K. Griest, A. M. Cieplak and M. J. Lehner, Experimental limits on primordial black hole dark matter from the first 2 yr of kepler data, *The Astrophysical Journal* **786**, p. 158 (2014).
18. C. Alcock, R. Allsman, D. R. Alves, T. Axelrod, A. C. Becker, D. Bennett, K. H. Cook, N. Dalal, A. J. Drake, K. Freeman *et al.*, The macho project: microlensing results from 5.7 years of large magellanic cloud observations, *The Astrophysical Journal* **542**, p. 281 (2000).
19. H. Niikura, M. Takada, N. Yasuda, R. H. Lupton, T. Sumi, S. More, T. Kurita, S. Sugiyama, A. More, M. Oguri *et al.*, Microlensing constraints on primordial black holes with subaru/hsc andromeda observations, *Nature Astronomy* **3**, 524 (2019).
20. A. Gould, A natural formalism for microlensing, *The Astrophysical Journal* **542**, p. 785 (2000).
21. L. Hamolli, M. Hafizi, F. De Paolis and A. A. Nucita, Estimating finite source effects in microlensing events due to free-floating planets with the euclid survey, *Advances in Astronomy* **2015** (2015).
22. L. Hamolli, F. De Paolis, M. Hafizi and A. Nucita, Predictions on the detection of the free-floating planet population with k2 and spitzer microlensing campaigns, *Astrophysical Bulletin* **72**, 73 (2017).
23. L. Hamolli, M. Hafizi, F. De Paolis and A. A. Nucita, The astrometric signal of microlensing events caused by free floating planets, *Astrophysics and Space Science* **363**, 1 (2018).
24. L. Hamolli, M. Hafizi, F. De Paolis and A. Nucita, Free-floating planets in the milky way, *Arabian Journal of Mathematics* **8**, 305 (2019).
25. R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, Euclid definition study report, *arXiv preprint arXiv:1110.3193* (2011).
26. N. Regnault, Photometric calibration of wide field imagers, PhD thesis, Université Pierre et Marie Curie-Paris VI2013.
27. G. Ingrosso, S. C. Novati, F. De Paolis, P. Jetzer, A. Nucita, G. Scarpetta and F. Strafella, A new analysis of the mega m 31 microlensing events, *Astronomy & Astrophysics* **462**, 895 (2007).
28. V. Belokurov, J. An, N. Evans, P. Hewett, P. Baillon, S. C. Novati, B. Carr, M. Crézé, Y. Giraud-Héraud, A. Gould *et al.*, The point-agape survey—ii. an unrestricted search

for microlensing events towards m31, *Monthly Notices of the Royal Astronomical Society* **357**, 17 (2005).
29. A. Riffeser, S. Seitz and R. Bender, The m31 microlensing event wecapp-gl1/point-agape-s3: evidence for a macho component in the dark halo of m31?, *The Astrophysical Journal* **684**, p. 1093 (2008).
30. S. C. Novati, V. Bozza, I. Bruni, M. Dall'Ora, F. De Paolis, M. Dominik, R. Gualandi, G. Ingrosso, P. Jetzer, L. Mancini *et al.*, The m31 pixel lensing plan campaign: Macho lensing and self-lensing signals, *The Astrophysical Journal* **783**, p. 86 (2014).
31. P. Tisserand, L. Le Guillou, C. Afonso, J. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. Beaulieu, X. Charlot *et al.*, Limits on the macho content of the galactic halo from the eros-2 survey of the magellanic clouds, *Astronomy & Astrophysics* **469**, 387 (2007).
32. A. Rest, C. Stubbs, A. C. Becker, G. Miknaitis, A. Miceli, R. Covarrubias, S. Hawley, R. C. Smith, N. B. Suntzeff, K. Olsen *et al.*, Testing lmc microlensing scenarios: The discrimination power of the supermacho microlensing survey, *The Astrophysical Journal* **634**, p. 1103 (2005).
33. A. M. Green, Astrophysical uncertainties on stellar microlensing constraints on multi-solar mass primordial black hole dark matter, *Physical Review D* **96**, p. 043020 (2017).
34. M. Hafizi, F. De Paolis, G. Ingrosso and A. A. Nucita, Microlensing signature of a white dwarf population in the galactic halo, *International Journal of Modern Physics D* **13**, 1831 (2004).
35. P. Jetzer, L. Mancini and G. Scarpetta, Microlensing towards the large magellanic cloud, *Astronomy & Astrophysics* **393**, 129 (2002).
36. S. Calchi Novati, S. Mirzoyan, P. Jetzer and G. Scarpetta, Microlensing towards the smc: a new analysis of ogle and eros results, *Monthly Notices of the Royal Astronomical Society* **435**, 1582 (2013).
37. F. De Paolis, G. Ingrosso, A. Nucita and A. Zakharov, Influence of magnification threshold on pixel lensing optical depth, event rate and time scale distributions towards m 31, *Astronomy & Astrophysics* **432**, 501 (2005).
38. E. Kerins, M. Darnley, J. Duke, A. Gould, C. Han, Y.-B. Jeon, A. Newsam, B.-G. Park and (A. Collaboration), The angstrom project: a microlensing survey of the structure and composition of the bulge of the andromeda galaxy, *Monthly Notices of the Royal Astronomical Society* **365**, 1099 (2006).
39. K. Griest, Galactic microlensing as a method of detecting massive compact halo objects, *The astrophysical journal* **366**, 412 (1991).
40. B. S. Gaudi, Microlensing surveys for exoplanets, *Annual Review of Astronomy and Astrophysics* **50**, 411 (2012).
41. M. Kiraga and B. Paczynski, Gravitational microlensing of the galactic bulge stars, *The Astrophysical Journal* **430**, L101 (1994).
42. J. Sureda, J. Magaña, I. J. Araya and N. D. Padilla, Press–schechter primordial black hole mass functions and their observational constraints, *Monthly Notices of the Royal Astronomical Society* **507**, 4804 (2021).
43. C. Afonso, J. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. Beaulieu, G. Blanc, X. Charlot, F. Couchot *et al.*, Limits on galactic dark matter with 5 years of eros smc data, *Astronomy & Astrophysics* **400**, 951 (2003).
44. S. C. Novati, F. De Luca, P. Jetzer and G. Scarpetta, Microlensing towards the large magellanic cloud: a study of the lmc halo contribution, *Astronomy & Astrophysics* **459**, 407 (2006).
45. L. Hamolli, M. Hafizi, F. De Paolis and A. A. Nucita, Gravitational microlensing constraints on primordial black holes by euclid, *Astrophysics and Space Science* **366**, 1 (2021).

Giant cosmic ray halos around M31 and the Milky Way

S. Recchia

Dipartimento di Fisica, Universitá di Torino,
via P. Giuria 1, 10125 Torino, Italy
E-mail: sarah.recchia@unito.it

Istituto Nazionale di Fisica Nucleare, Sezione di Torino,
Via P. Giuria 1, 10125 Torino, Italy

S. Gabici

Université de Paris, CNRS, Astroparticule et Cosmologie,
F-75006 Paris, France

F. A. Aharonian

Dublin Institute for Advanced Studies,
31 Fitzwilliam Place, Dublin 2, Ireland

Max-Planck-Institut für Kernphysik,
Postfach 103980, D-69029 Heidelberg, Germany

V. Niro

Université de Paris, CNRS, Astroparticule et Cosmologie,
F-75006 Paris, France

Recently, a diffuse γ-rays emission in the energy range 1-100 GeV has been detected around M31, that extends up to 120-200 kpc from its center. Such extended emission is difficult to be explained in the typical scenario of cosmic rays produced in the galactic disk or in the galactic center (GC) and diffusing in the galactic halo. We show that a cosmic ray origin, either hadronic or leptonic, of the emission is viable if non-standard cosmic ray transport scenarios are considered, or if particles are accelerated directly in the galactic halo (*in situ* acceleration). The cosmic ray halo can be powered by the accretion of intergalactic gas or by the activity of galaxy's central black hole. If giant cosmic ray halos are common around galaxies, the interactions of cosmic ray protons and nuclei with the circumgalactic gas surrounding Milky Way could explain the isotropic diffuse flux of neutrinos observed by Icecube.

Keywords: Galaxies: halos; cosmic rays; gamma rays; neutrinos.

1. Introduction

The Andromeda Galaxy (M31) and the Milky Way (MW) present many similarities. Both have a bulge, a disk, an extended gaseous halo, a central supermassive black hole and a dark matter halo with a radial extension of $200 - 300$ kpc and a total mass of $\sim 10^{12} M_{\odot}$.[1] The integrated $\gamma-$ray luminosity of the Andromeda galaxy,

as reported by *Fermi*-LAT,[2,3] above 100 MeV is of $\sim 6.6 \times 10^{41}$ erg s^{-1}, less than a factor ~ 2 that of the MW.[2] Fermi Bubbles-like structure emanating from the central region of M31 may also exist, as recently reported.[4] However, while in the MW the $\gamma-$ray emission correlates spatially with the gaseous disk, in M31 it appears to be concentrated within the inner ~ 5 kpc region.[3]

Recently the *Fermi*-LAT data have been analyzed in detail in Ref. 1, from a region that corresponds to a projected radius of ~ 200 kpc from the center of M31 and found an excess that extends up to $\sim 120 - 200$ kpc. The analysis was performed with an accurate modeling of the MW foreground emission and considering a spherically symmetric template, centered on M31 and including three regions:

- the inner galaxy (IG)
  a ~ 5.5 kpc radius region that contains the emission from the inner galaxy[3]
- the spherical halo (SH)
  a ring of $\sim 5.5 - 120$ kpc
- the outer halo (OH)
  a ring of $\sim 120 - 200$ kpc.

For each region the authors provide the $\gamma-$ray spectra. Moreover, given the contamination from the MW disk emission, especially relevant for the northern part of the OH, the spectra for the southern and northern part of the SH and OH are also reported. The north/south regions' spectra are quite similar for the SH, while they exhibit relevant differences for the OH, where the spectrum of the northern part shows a bumpy profile likely associated to a contamination from the MW. Thus, the authors concluded that, while the excess from the SH region is likely associated to the halo of M31, the origin of the emission from the OH is much less clear and may be rather related to the MW, or to something else. Here we focus on the SH region.

Two possible spectral fits for the SH region are reported by Ref. 1, one with a power-law plus exponential cut-off parametrization given by

$$I_{SH} \approx 9.8 \times 10^{-11}\, E_{\rm GeV}^{-1.9}\, e^{-E_{\rm GeV}/11.6} {\rm MeV}^{-1}\, {\rm cm}^{-2} {\rm s}^{-1}\, {\rm sr}^{-1} \qquad (1)$$

and one with pure power-law, as shown in Fig. 1. Here $E_{\rm GeV}$ is the photon energy in GeV. The total γ-ray luminosity (northern plus southern part) of SH is:

$$L_\gamma \approx 1.7 - 1.9 \times 10^{39} {\rm erg/s}\ . \qquad (2)$$

where the lower(upper) value corresponds to the power law plus cutoff(pure power law) fit to data. Based on the large extension of $\gamma-$ray halo, on the intensity and spectrum of the various components, the authors of Ref. 1 suggest a dark matter interpretation for such emission[5] and, based on typical CR propagation scenarios,

conclude that it is unlikely that a major fraction of the emission is due to cosmic rays (CR) in the halo of M31.

Instead, we illustrate that the $\sim 100 - 200$ kpc extended emission from Andromeda could be naturally explained as the result of CR interactions in the halo provided that: *i)* a non-standard mechanism of CR transport into galactic halos, by means of buoyant bubbles, is considered; *ii)* particles are directly accelerated in the galaxy's halo. In what follows we summarize our results reported in Ref. 6 and we refer the reader to that paper for more details.

Interestingly, if the MW is surrounded by a halo similar to that reported for M31, it could be possible to explain the diffuse Icecube neutrino flux, as previously proposed by Ref. 7. Moreover, such large halo around M31 was previously proposed by Ref. 8, which showed that it could contribute to the isotropic gamma-ray background.

2. Hadronic and leptonic origin of the M31 γ-ray halo: Energetics

2.1. *Leptonic scenario*

In this scenario the γ-ray emission from the SH is the result of inverse Compton scattering (ICS) of relativistic electrons off the interstellar radiation field (IRF). At the distances from the disk of the SH region, the IRF is dominated by CMB photons and the galactic magnetic field is likely below $\sim 1\,\mu$G, so that synchrotron losses are probably negligible compared to ICS losses. Thus, $\gamma-$rays in the range $\sim 1 - 100$ GeV would be produced by electrons of energy $\sim 0.6 - 6$ TeV, and the cutoff at $E_\gamma = 11.6$ GeV would correspond to electrons of energy $E_{max,e} \approx 1.9$ TeV. The lepton energy loss time in the CMB field is given by (see Ref. 6 and references therein)

$$\tau_{CMB} \sim 1.3 \times 10^6 E_{\rm TeV}^{-1} \text{ yr}, \tag{3}$$

where $E_{\rm TeV}$ is the particle energy in TeV.

Such short timescale implies that, for any realistic interstellar CR diffusion, electrons coming from the disk/GC would lose all their energy before reaching traveling a few kpc in the halo. Thus, if the SH emission of M31 is produced by electrons they have to be directly accelerated in that region. The minimal energy requirement is obtained in the assumption that the $\gamma-$ray production takes place in a calorimetric regime (namely the ICS timescale is shorter than any other relevant dynamical scale). In this regime the observed γ-ray luminosity (Eq. 2) equals the electron luminosity:

$$L_e = L_\gamma \approx 1.7 - 1.9 \times 10^{39} \text{erg/s}. \tag{4}$$

Notice that this luminosity refers only to electrons in the energy band $\approx 0.6 - 6$ TeV, and is of the same order than the total estimated power of CR electrons in the disk of the MW.[9]

2.2. *Hadronic scenario*

In this scenario $\gamma-$rays are produced in proton-proton interactions between CRs and the ambient gas, where photons of energy E_γ are generated by CR protons of $E_p \approx 10\, E_\gamma$. In order to account for the observed γ-ray luminosity L_γ, the required CR proton luminosity reads:[6,7]

$$L_p = 3\, L_\gamma / f, \quad f = 1 - e^{-\tau_{res}/\tau_{pp}}, \tag{5}$$

where τ_{res} is the residence time of CRs in the halo and τ_{pp} is the timescale for pp interactions. The latter depends on the density of the ambient medium and is given by

$$\tau_{pp} \sim 7.1 \times 10^{10} n_{H,-3}^{-1} \text{ yr} \tag{6}$$

where $n_{H,-3}$ is the halo hydrogen density in units of 10^{-3} cm^{-3}. At a distance of ~ 100 kpc from the Galactic disk, the gas density of the MW halo is probably of the order $\sim 10^{-4} - 10^{-3}$cm^{-3} (see Ref. 6 and references therein), for which the loss time is larger than the age of the Universe. Given the similarities with M31, we may expect that a similar gaseous halo is also present around Andromeda. Taking a reference value of $\tau_{res} = 10^9 \tau_{res,9}$ yr for the CR residence time, the CR proton luminosity reads:

$$L_p \approx 1.8 \times 10^{41}\, \tau_{res,9}^{-1}\, n_{H,-3}^{-1} \text{erg/s}\,. \tag{7}$$

Notice that for $\tau_{res,9} \gtrsim 1$ and $n_{H,-3} \sim 1$, such luminosity is of the same order of that invoked to explain the population of CR protons observed in the disk of the MW.[9]

3. Origin of the radiating particles

The most popular hypothesis for the origin of Galactic CRs is that particles are accelerated in sources located in the disk (e.g. supernova remnants, pulsars, stellar winds, the GC) and escape form the Galaxy undergoing a diffusive-advective transport, the former due to the scattering of CRs on magnetic inhomogeneities and the latter to the possible presence of a galactic wind.[9] A similar situation is likely present also in M31 and one may wonder whether the emission from the SH results from the transport of particles from the disk/GC of Andromeda. However this hypothesis is very problematic. Indeed, for typical values of the interstellar CR diffusion coefficient, TeV electrons would lose energy due to ICS and sychrotron emission before reaching distances of few kpc from the disk. As for protons, typical models of diffusive-advective propagation from the disk/GC predict a substantial decrease of the CR density with the distance from the disk. Moreover, also the halo gas density decreases moving away from the disk, with a density of $\approx 1\,\text{cm}^{-3}$ in the disk and $\approx 10^{-4} - 10^{-3}\,\text{cm}^{-3}$ at a distance of ~ 100 kpc. This imply that, in order to fit the γ-ray flux observed from the SH, the CR proton intensity in the disk of M31 should be large, which would result in a bright γ-ray emission from the disk,

which is not observed. In particular, the observed ratio between the average $\gamma-$ray emissivity of the IG and SH regions is

$$\frac{\langle \epsilon_{SH} \rangle}{\langle \epsilon_{IG} \rangle} \sim 10^{-3}. \tag{8}$$

while in standard CR propagation scenarios would be (see Ref. 6 for a detailed discussion)

$$\frac{\langle \epsilon_{SH} \rangle}{\langle \epsilon_{IG} \rangle} \lesssim 10^{-5} \frac{n_{SH,-3}}{n_{IG,0}}, \tag{9}$$

where we assumed a typical gas density of $n_{IG} \sim 1\,\mathrm{cm}^{-3}$ and $n_{SH} \sim 10^{-3}\,\mathrm{cm}^{-3}$ for the IG and SH region respectively.

3.1. *Cosmic ray transport by means of buoyant bubbles*

The Fermi Bubbles[10] in the MW are probably the result of past nuclear activity of the GC, and similar structures are possibly present also in the M31.[4] Motivated by these observations, we envisage a scenario in which CR protons produced in periodic episodes of nuclear activity are transported into the SH within buoyant bubbles.

Buoyant bubbles are often observed in the central regions of galaxy clusters,[11,12] but also in galaxies.[13] Such bubbles are generally found to rise at a speed of the order of a fraction of the sound speed (≈ 100 km/s in the hot diffuse circumgalactic gas[14]), before being disrupted due to plasma instabilities. Their lifetime could reach order of $\sim 10^9$ yrs if the possible stabilizing action of a magnetic field (see Ref. 14 and references therein).

Let us assume that bubbles are generated at a rate of $\nu_B = 10^{-2}\nu_{B,-2}\mathrm{Myr}^{-1}$. The effective rate at which CR protons may be injected into the halo is $L_p = \eta E_B \nu_B \sim 3.2 \times 10^{41} \eta E_{B,57} \nu_{B,-2}\mathrm{erg/s}$, where $E_{B,57} = E_B/10^{57}$ erg[15,16] and η is a CR conversion efficiency that also accounts for possible adiabatic energy losses due to the bubble expansion during its rise in the halo. Comparing this expression with the energy requirement in Eq. 7 one finds

$$\eta \approx 0.56\ \tau_{res,9}^{-1} n_{H,-3}^{-1} E_{B,57}^{-1} \nu_{B,-2}^{-1}. \tag{10}$$

CRs can be transported into the galactic halo within such bubbles before the bubble is disrupted. When the bubble is disrupted CRs are released in the SH and spread diffusively to fill a region of ≈ 100 kpc radius. If the contributions of several bubbles overlap in the SH, a quasi-stationary CR population can be established in this region, while in the IG, an intermittent Fermi Bubble like structure could be observed. In this way, it is possible to populate the SH with CRs without having a large CR density in the IG and in the disk, contrary to the case in which CRs diffuse/advect directly from the galactic disk or the GC.

If the CR diffusion coefficient in the SH increases with the particle energy, similarly to what happens in the near-disk region of the MW, a quasi-stationary CR population can be maintained in the SH for energies such that the diffusion

time $\tau_{res} \sim R_{SH}^2/(6\ D)$ over a region of size $R_{SH} \sim 100$ kpc is shorter than the typical time between episodes of nuclear activity $1/\nu_B$. At large enough energy, this condition is no longer valid and protons will populate the halo intermittently, which would result in a spectral cut-off at the particle energy, E_c, such that $D(E_c) = 5 \times 10^{30}\ R_{100}^2 \nu_{B,-2}$ cm^2/s. This would explain the power law plus cut-off fit to $\gamma-$ray data (black points in Fig. 1), provided that $E_c \approx 100$ GeV (see Ref. 6 and references therein).

Instead, the pure power law fit to the data (see Fig. 1) could be explained in the case in which the diffusion coefficient is rather flat in energy, and/or if CRs injected during the entire lifetime of M31 are confined in the SH remain confined in the SH.

3.2. *In situ acceleration*

In the case of a leptonic origin, the acceleration of CRs have to take place directly in the SH. Thus, let us assume that a giant spherical shock of radius R_s (whose possible origin is discussed below) is present in the SH. The maximum energy $E_{max,e}$ of accelerated electrons can be estimated: *i*) by assuming that the energy losses due to ICS are the most relevant limitation to the acceleration; *ii*) the particle acceleration at shocks is accompanied by an amplification of the magnetic field. Under such assumptions, he maximum energy reads (see Ref. 6 for the details)

$$E_{max,e} \approx 24\ u_{s,3}^3 n_{0,-4}^{1/4} \left(\frac{\xi_B}{0.035}\right)^{1/4} \text{TeV}, \tag{11}$$

where $n_0 = 10^{-4} n_{0,-4}$ cm^{-3} is the number density of the intergalactic gas, $u_s = 10^3 u_{s,3}$ km/s is the velocity at which matter flows into the shock, $\xi_B \approx 3.5\%$ of the shock ram pressure is converted into magnetic field energy.

Given that the observed γ-ray spectrum extends up to at least a photon energy $E_\gamma \approx 10$ GeV, the electron spectrum should extend at least up to an energy $E_{max,e} \approx 2$ TeV, which requires a shock speed

$$u_{s,3} \gtrsim 0.43\ n_{0,-4}^{-1/12} \left(\frac{\xi_B}{0.035}\right)^{-1/12}. \tag{12}$$

The possible origin of such a shock can be envisaged by noticing that the radius of the SH, $R_{SH} = 100 R_{SH,2}$ kpc, is of the same order of the virial radius of the galaxy. Moreover, the shock velocity inferred above is close to the free fall velocity at the edge of the SH $v_{ff} \sim 0.29 \times 10^3 M_{12}^{1/2} R_{SH,2}^{-1/2}$ km/s, where $M = 10^{12} M_{12} M_\odot$ is the total mass of M31. Thus, it is possible to speculate that particles are accelerated at a spherical accretion shock. However it is important to remark that the existence of accretion shocks around galaxies is matter of debate.[6] The energy flow across the shock would be approximately:

$$L_s \approx (4\pi R_{SH}^2) \frac{\varrho_0 v_{ff}^3}{2} \sim 3.4 \times 10^{42} R_{SH}^{1/2} n_{0,-4} M_{12}^{3/2} \text{erg/s} \tag{13}$$

which could accommodate the energy requirement given by Eq. 4.

A giant spherical shock may also be formed as a result of the GC nuclear activity. In this case an estimate of the shock velocity v_s for a shock radius $R_s \sim R_{SH}$ and a overall duration of the nuclear activity of $\tau_{GC} = 10^9 \tau_{GC,9}$ is given by (see Ref. 6 and references therein)

$$u_s \approx 0.2 \times 10^3 \; L_{GC,43}^{1/5} n_{0,-4}^{-1/5} \tau_{GC,9}^{-2/5} \mathrm{km/s}, \tag{14}$$

where $L_{GC} = 10^{43} L_{GC,43}$ erg/s is the time averaged rate of energy injection due to galactic nuclear activity. Interestingly, a typical value of 10^{43} erg/s for the energy injection rate is much smaller than the Eddington luminosity of the central supermassive black hole of mass M_{BH}, $L_{Edd} \sim 1.3 \times 10^{46} \left(\frac{M_{BH}}{10^8 \; M_\odot}\right)$ erg/s.

In both scenarios for the origin of the shock, the CR electron luminosity needed to explain observations (Eq. 4) can be explained with an acceleration efficiency of the order of $\eta \approx 10^{-4}$.

From the study of Galactic CRs[9] it emerges that interstellar shocks, together with electrons, also accelerate protons, likely with a much larger efficiency. In this case, the maximum energy $E_{max,p}$ is most likely limited either by the age of the system, $\tau_s = 10^9 \tau_{s,9}$ yr, since in the dilute medium that form the SH the proton-proton loss time is typically larger than the age of the Universe (see Eq. 6), either by the particle confinement time in the accelerator, namely the diffusion length at energy $E_{max,p}$ should be smaller than the size of the accelerator ($\sim R_{SH}$). The two conditions provides similar values of $E_{max,p}$, so we consider here the former, which gives $E_{max,p} \approx 4.6 \times 10^2 \; u_{s,3}^3 \tau_{s,9} n_{0,-4}^{1/2} \left(\frac{\xi_B}{0.035}\right)^{1/2}$ PeV. Thus, the acceleration of CR protons can proceed well beyond the $\sim$ 10-100 TeV needed to explain the γ-ray observations. The latter can be accounted for by an acceleration efficiency at the level of $\approx 1\%$.

4. A giant CR halo around the MW and the origin of Icecube neutrinos

The possible existence of a $\sim 100 - 200$ kpc CR halo surrounding the MW was proposed by Ref. 7 for explaining the diffuse flux of multi-TeV neutrinos detected by Icecube. The differential isotropic flux of astrophysical neutrinos (all flavors, neutrinos plus antineutrinos) measured at Earth can be fitted with a power law[17]

$$\Phi_\nu^{IC}(E_\nu) \sim 6.37 \times 10^{-18} \left(\frac{E_\nu}{100 \text{ TeV}}\right)^{-2.87} \mathrm{GeV^{-1} cm^{-2} s^{-1} sr^{-1}}. \tag{15}$$

The integrated isotropic flux is

$$F_\nu(> 100 \text{ TeV}) \sim 1.2 \times 10^{-10} \, \mathrm{erg/cm^2/s/sr}. \tag{16}$$

Let us assume that such neutrino flux is produced in the MW halo due proton-proton interactions at a typical distance of $R_H = 10^2 R_{H,2}$ kpc. The corresponding

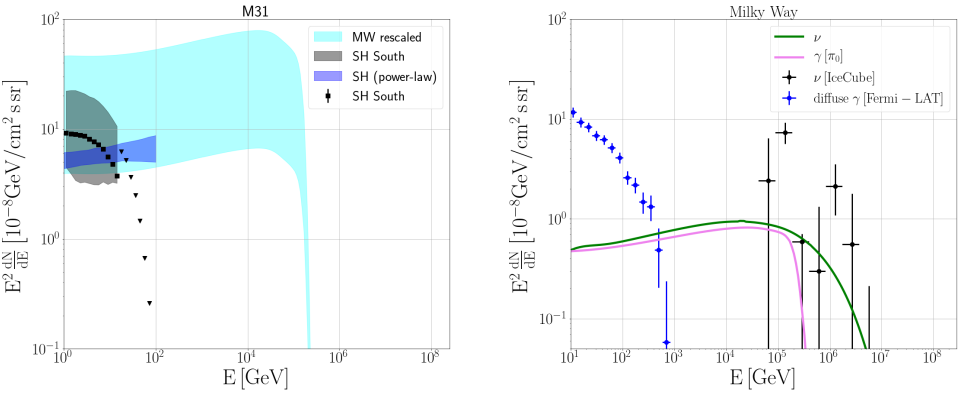


Fig. 1. *left:* Expected γ-ray emission from M31 (shaded cyan region) obtained re-scaling the predictions for the halo of the Milky Way. The black data points and the shaded gray region show *Fermi*-LAT observations of M31[1] *right:* Isotropic diffuse neutrino and γ-ray emission observed by Icecube[17] (black) and *Fermi*-LAT[18] (blue data points). Solid lines are predictions for the neutrino (green) and γ-ray (pink) resulting from the interactions of CR protons with ambient gas in a ~ 100 kpc halo surrounding the MW.

differential neutrino emissivity from the entire MW would be

$$Q_\nu^{MW}(E_\nu) = (4\pi)^2 \Phi_\nu^{IC}(E_\nu) R_H^2. \tag{17}$$

Then, a neutrino luminosity $L_\nu^{MW}(> 100 \text{ TeV}) \approx 1.8 \times 10^{39} R_{H,2}^2 \text{erg/s}$ is expected, together with an isotropic diffuse γ-ray emission above ~ 100 TeV with a similar luminosity.

The isotropic diffuse fluxes of γ-rays and neutrinos measured by *Fermi*-LAT[18] and Icecube,[17] respectively, are shown in Fig. 1, together with the predictions from the γ-ray and neutrino emission from CR proton-proton interactions in the MW halo, estimated assuming a proton spectrum $\propto E_p^{-2} \exp(-E_p/20 \text{ PeV})$ normalized in such a way to contain a total energy equal to $W_p \sim 4.6 \times 10^{57} R_{H,2}^2 n_{H,-3}^{-1}$ erg. A cutoff in the γ-ray spectrum appears due to the absorption of very high energy photons by pair production in the CMB.

Since the MW and M31 are quite similar galaxies, and the gas density in their halos are inferred from observations to be comparable (see e.g. Refs. 19, 20), it appears natural to compare the MW halo origin proposed for the Icecube neutrino flux and the M31 γ-ray SH emission. In the proposed scenarios the CR proton density in the halo would scale linearly with the central SMBH mass (if CR protons are powered by the galactic nuclear activity) or with the total mass of the galaxy to the power 3/2 (if CRs are accelerated at an accretion shock). The ratio between the mass of M31 and of the MW is ~ 2, while the ratio between the SMBH masses is ~ 33 (see Ref. 6 and references therein), which means that, if the CR confinement time in the halo is similar for the two galaxies (order of Gyr), one may suppose that the γ-ray and neutrino luminosities of M31 are a factor of $\approx 2.8 - 33$ than those of

the MW. In Fig. 1 we show the rescaled MW neutrino flux together with the $\gamma-$ray spectrum of M31 observed by *Fermi*-LAT for a power law plus exponential cutoff modeling of data (black points and shaded gray region) and for a power law model (blue line and shaded region).[1]

Remarkably, in the absence of a cutoff/steepening in the γ-ray spectrum of M31, we find a good agreement of our predictions and *Fermi*-LAT data, that suggests a possible common origin of the GeV γ-ray emission from the halo of M31 and the neutrino emission from the halo of the MW.

5. Conclusions

Recently, *Fermi*-LAT discovered a giant γ-ray halo of size $\sim$ 100-200 kpc around M31. We investigated a possible CR origin of the emission and we concluded that it could be explained by CR protons interacting with the circumgalactic gas. Particles can be accelerated directly in the SH of M31, or may be produced in the GC of M31 and then transported into the halo by means of buoyant bubbles. The morphology of the emission is expected to be spherically symmetric in the former case and similar to that of Fermi Bubbles, but much more extended, in the latter case. If CR protons can be confined for long times in the SH of M31 (order of gigayears), the emission measured by *Fermi*-LAT could be explained with a time averaged luminosity of CRs injected in the halo of the order of $\gtrsim 10^{40} - 10^{41}$ erg/s, comparable to the estimated luminosity of CR sources in the Galactic disk. Such luminosity can be accounted for by the galaxy nuclear activity or by an accretion shock.

Moreover, if a similar giant CR halo is present also around the MW, which appears plausible given the similarities between the MW and M31 and the observational evidence of similar extended gaseous halos around both galaxies, also the isotropic diffuse neutrino emission observed by Icecube could be explained as originating from the halo of our own Galaxy (as proposed by Ref. 7).

If gaseous halos are a common feature of MW and M31-like galaxies they might all emit both γ-rays and neutrinos, which however are beyond the capabilities of current instruments for distances significantly larger than M31 (see Ref. 6 for a discussion).

Alternatively, the M31 $\gamma-$ray emission could be of leptonic origin, namely the result of ICS of electrons on the CMB. In this case the particles have to be accelerated directly in the SH and no neutrinos would be expected from M31.

References

1. C. M. Karwin, S. Murgia, S. Campbell and I. V. Moskalenko, Fermi-LAT Observations of γ-Ray Emission toward the Outer Halo of M31, *ApJ* **880**, p. 95 (August 2019).
2. A. A. Abdo *et al.*, Fermi Large Area Telescope observations of Local Group galaxies: detection of M 31 and search for M 33, *A&A* **523**, p. L2 (November 2010).
3. M. Ackermann *et al.*, Observations of M31 and M33 with the Fermi Large Area Telescope: A Galactic Center Excess in Andromeda?, *ApJ* **836**, p. 208 (February 2017).

4. M. S. Pshirkov, V. V. Vasiliev and K. A. Postnov, Evidence of Fermi bubbles around M31, *MNRAS* **459**, L76 (June 2016).
5. C. Karwin, S. Murgia, I. Moskalenko, S. Fillingham, A.-K. Burns and M. Fieg, Dark Matter Interpretation of the Fermi-LAT Observations Toward the Outer Halo of M31, *arXiv e-prints* , p. arXiv:2010.08563 (October 2020).
6. S. Recchia, S. Gabici, F. A. Aharonian and V. Niro, Giant Cosmic-Ray Halos around M31 and the Milky Way, *ApJ* **914**, p. 135 (June 2021).
7. A. M. Taylor, S. Gabici and F. Aharonian, Galactic halo origin of the neutrinos detected by IceCube, *Phys. Rev. D* **89**, p. 103003 (May 2014).
8. R. Feldmann, D. Hooper and N. Y. Gnedin, Circum-galactic Gas and the Isotropic Gamma-Ray Background, *ApJ* **763**, p. 21 (January 2013).
9. S. Gabici, C. Evoli, D. Gaggero, P. Lipari, P. Mertsch, E. Orlando, A. Strong and A. Vittino, The origin of Galactic cosmic rays: Challenges to the standard paradigm, *International Journal of Modern Physics D* **28**, 1930022 (January 2019).
10. M. Ackermann *et al.*, The Spectrum and Morphology of the Fermi Bubbles, *ApJ* **793**, p. 64 (September 2014).
11. T. W. Jones and D. S. De Young, Magnetohydrodynamic simulations of relic radio bubbles in clusters, *The Astrophysical Journal* **624**, p. 586–605 (May 2005).
12. E. Churazov, M. Brüggen, C. R. Kaiser, H. Böhringer and W. Forman, Evolution of Buoyant Bubbles in M87, *ApJ* **554**, 261 (June 2001).
13. A. Finoguenov, M. Ruszkowski, C. Jones, M. Brüggen, A. Vikhlinin and E. Mandel, In-Depth Chandra Study of the AGN Feedback in Virgo Elliptical Galaxy M84, *ApJ* **686**, 911 (October 2008).
14. C. Zhang, E. Churazov and A. A. Schekochihin, Generation of internal waves by buoyant bubbles in galaxy clusters and heating of intracluster medium, *Monthly Notices of the Royal Astronomical Society* **478**, 4785 (05 2018).
15. F. Guo and W. G. Mathews, Thefermibubbles. i. possible evidence for recent agn jet activity in the galaxy, *The Astrophysical Journal* **756**, p. 181 (Aug 2012).
16. H. Y. K. Yang, M. Ruszkowski, P. M. Ricker, E. Zweibel and D. Lee, The Fermi Bubbles: Supersonic Active Galactic Nucleus Jets with Anisotropic Cosmic-Ray Diffusion, *ApJ* **761**, p. 185 (December 2012).
17. R. Abbasi *et al.*, The IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data, *arXiv e-prints* , p. arXiv:2011.03545 (November 2020).
18. M. Ackermann, M. Ajello, A. Albert, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini and et al., The spectrum of isotropic diffuse gamma-ray emission between 100 mev and 820 gev, *The Astrophysical Journal* **799**, p. 86 (Jan 2015).
19. A. Gupta, S. Mathur, Y. Krongold, F. Nicastro and M. Galeazzi, A Huge Reservoir of Ionized Gas around the Milky Way: Accounting for the Missing Mass?, *ApJ* **756**, p. L8 (September 2012).
20. Z. Qu, R. Huang, J. N. Bregman and J.-T. Li, An X-ray and SZ bright diffuse source toward M31: A Local Hot Bridge, *arXiv e-prints* , p. arXiv:2011.02125 (November 2020).

A nearly complete census of intergalactic gas using the kinematic Sunyaev-Zel'dovich effect

Chaves-Montero, Jonás

Donostia International Physics Centre, Paseo Manuel de Lardizabal 4, 20018 Donostia-San Sebastian, Spain
E-mail: jonas.chaves@dipc.org
Webpage: https://jchavesmontero.github.io/jchavesmontero/

A complete census of baryons in the late universe is a long-standing challenge due to the intermediate temperate and rarefied character of the majority of cosmic gas. To gain insight into this problem, we extract measurements of the kinematic Sunyaev-Zel'dovich (kSZ) effect from the cross-correlation of angular redshift fluctuations maps, which contain precise information about the cosmic density and velocity fields, and CMB maps high-pass filtered using aperture photometry; we refer to this technique as ARF-kSZ tomography. Remarkably, we detect significant cross-correlation for a wide range of redshifts and filter apertures using 6dF galaxies, BOSS galaxies, and SDSS quasars as tracers, yielding a 11 sigma detection of the kSZ effect. We then leverage these measurements to set constraints on the location, density, and abundance of gas inducing the kSZ effect, finding that this gas resides outside dark matter haloes, presents densities ranging from 10 to 250 times the cosmic average, and comprises half of cosmic baryons. Taken together, these findings indicate that ARF-kSZ tomography provides a nearly complete census of intergalactic gas from $z = 0$ to 5. This contribution is a summary of the work already published in Ref. 1.

Keywords: Diffuse radiation; cosmic background radiation; large-scale structure of Universe; cosmology: observations; cosmology: theory

1. Introduction

Over the last decades, precise observations of primordial CMB anisotropies[2] and Big Bang nucleosynthesis studies[3] have set strict constraints on the abundance and distribution of baryons in the early universe. However, a complete census of baryons at late times remains elusive; this is principally due to the intermediate temperature and rarefied character of nearly all cosmic gas, which hinders the detection of baryons outside high-density regions and leaves invisible the majority of cosmological volume. Such is the case that until recently, low redshift studies only detected 70% of the expected amount of baryons.[4–6]

Nonetheless, recent studies have successfully detected baryons outside high-density areas conducting kinematic Sunyaev-Zel'dovich (kSZ) effect observations,[7,8] thermal Sunyaev-Zel'dovich (tSZ) studies,[9,10] low-redshift Lyman-α surveys,[11] and deep X-ray campaigns.[12,13] Despite their success, these works only set constraints on the distribution of baryons at either a few specific redshifts or across a reduced

number of line-of-sights, failing to provide a complete picture of cosmic gas in the late universe.

Of the different strategies listed above, throughout this work we set constraints on the properties of intergalactic gas using measurements of the kSZ effect, which refers to the Doppler boosting of CMB photons as these scatter off free electrons moving relative to the CMB rest frame.[14,15] The motivation of using these measurements is that the kSZ effect is sensitive to free electrons independently of the temperature and density of the medium in which these reside, and thus it is uniquely suited to study the large-scale distribution of baryons at low redshift.

Even though the kSZ effect presents significant advantages to observe cosmic gas, it is challenging to extract this signal from observations because the amplitude of kSZ fluctuations is approximately two orders of magnitude smaller than that of primordial CMB fluctuations and the spectral shape of both signals is practically the same. Furthermore, most extraction methods require estimating the peculiar velocity field of intervening matter,[16,17] which is also challenging and adds notable uncertainties. Other approaches circumvent such estimation but require either using cosmological simulations for calibration or modelling and subtracting other effects,[8,18–20] which also introduces substantial uncertainties.

In this scenario, the cross-correlation of angular redshift fluctuations (ARF),[21] which encode precise information about the cosmic density and velocity fields, and CMB observations provides a clean window towards a tomographic detection of the kSZ effect given that systematic uncertainties affecting either of these observables do not present significant correlation. This approach, which we refer to as ARF-kSZ tomography, requires redshift information from either spectroscopic or spectro-photometric surveys as well as theoretical predictions for the large-scale cross-correlation between ARF and the kSZ effect.

2. Results

To extract measurements of the kSZ effect, we start by generating ARF using galaxies from the 6dF Galaxy Survey (6dF),[22] galaxies from the Baryon Oscillation Spectroscopic Survey (BOSS),[23,24] and quasars from the extended Baryon Oscillation Spectroscopic survey (eBOSS)[25] selected at 16 different redshift shells between $z = 0$ and 5. Then, we high-pass filter foreground-reduced *Planck* maps by applying aperture photometry filters of different sizes on the positions given by the aforementioned galaxies and quasars, thereby extracting kSZ signal induced by gas surrounding these sources. This technique also reduces contamination from primordial CMB anisotropies as aperture photometry removes signals that are constant over the size of the aperture. Then, we cross-correlate these maps, finding significant correlation for a wide range of redshifts and filter apertures. After accounting for

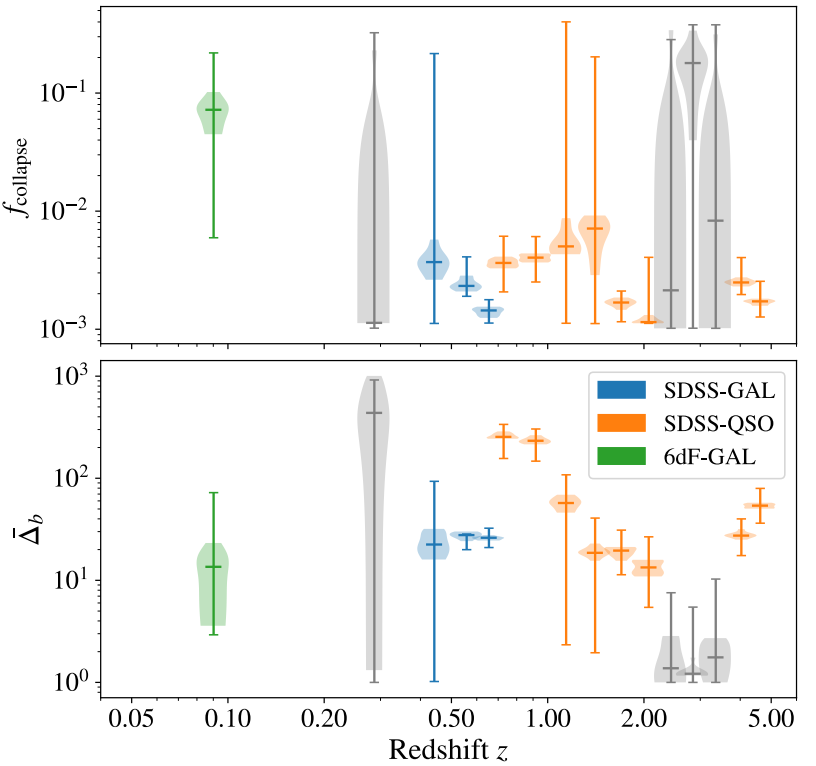


Fig. 1. Fraction of kSZ gas to within dark matter haloes (top panel) and average density of kSZ gas outside these structures (bottom panel). The horizontal bar and size of the shaded regions indicate the average and scatter of results computed using different *Planck* foreground-reduced maps, respectively; green, blue, and orange colours indicate the results for shells containing 6dF galaxies, BOSS galaxies, and SDSS quasars; grey colours denote shells with no significant detection of the kSZ effect; and error bars show 1σ uncertainties for *Planck* COMMANDER map. We find that more than 99% of kSZ gas resides outside haloes and that the density of this gas ranges from 10 to 250 times the cosmic baryon density, which is in agreement with the density of the intergalactic medium according to hydrodynamical simulations.

correlations between different measurements, we find that our measurements yield an 11σ detection of the kSZ effect, which is the highest significance detection of the kSZ effect up to date.

We leverage these measurements to set constraints on the location, density, and abundance of kSZ gas from redshift $z = 0$ to 5. In Fig. 1, we show that more than 99% of kSZ gas resides outside haloes and that the density of this gas ranges from 10 to 250 times the cosmic average, which is in agreement with the density of the gas in filaments and sheets according to hydrodynamical simulations. Then, in Fig. 2 we show that our kSZ measurements are compatible with detecting 50% of

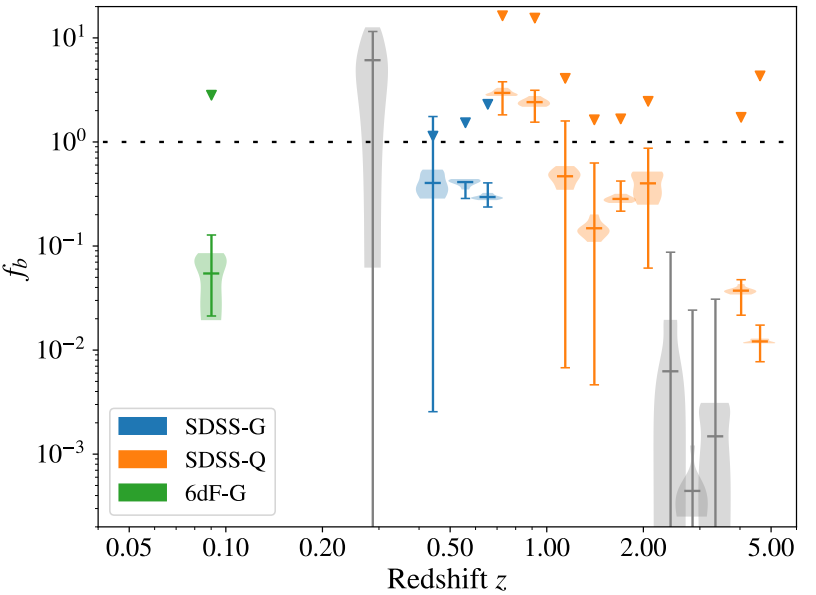


Fig. 2. Fraction of the total abundance of cosmic baryons detected at different redshifts. We use the same coding as Fig. 1. Remarkably, ARF-kSZ tomography is sensitive to approximately 50% of cosmic baryons, which highlights the efficiency of this technique detecting intergalactic gas. The results are well below 50% for the first and last two shells because the number density of sources at these redshifts is very low; in fact, forecasts for f_b when considering an infinite number density of sources (triangles) are compatible among all shells.

cosmic baryons. Taken together, these findings indicate that ARF-kSZ tomography provides a nearly complete census of intergalactic gas from $z = 0$ to 5.

References

1. J. Chaves-Montero, C. Hernández-Monteagudo, R. E. Angulo and J. D. Emberson, Measuring the evolution of intergalactic gas from $z = 0$ to 5 using the kinematic Sunyaev–Zel'dovich effect, *Monthly Notices of the Royal Astronomical Society* **503**, 1798 (March 2021).
2. Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley,

F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wand elt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei and A. Zonca, Planck 2018 results. VI. Cosmological parameters, *A&A* **641**, p. A6 (September 2020).

3. R. J. Cooke, M. Pettini and C. C. Steidel, One Percent Determination of the Primordial Deuterium Abundance, *ApJ* **855**, p. 102 (March 2018).
4. M. Fukugita and P. J. E. Peebles, The Cosmic Energy Inventory, *ApJ* **616**, 643 (December 2004).
5. F. Nicastro, S. Mathur and M. Elvis, Missing baryons and the warm-hot intergalactic medium, *Science* **319**, 55 (2008).
6. J. M. Shull, B. D. Smith and C. W. Danforth, The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing, *ApJ* **759**, p. 23 (November 2012).
7. C. Hernández-Monteagudo, Y.-Z. Ma, F. S. Kitaura, W. Wang, R. Génova-Santos, J. Macías-Pérez and D. Herranz, Evidence of the Missing Baryons from the Kinematic Sunyaev-Zeldovich Effect in Planck Data, *Physical Review Letters* **115**, p. 191301 (November 2015).
8. J. C. Hill, S. Ferraro, N. Battaglia, J. Liu and D. N. Spergel, Kinematic Sunyaev-Zel'dovich Effect with Projected Fields: A Novel Probe of the Baryon Distribution with Planck, WMAP, and WISE Data, *Phys. Rev. Lett.* **117**, p. 051301 (Jul 2016).
9. A. de Graaff, Y.-C. Cai, C. Heymans and J. A. Peacock, Probing the missing baryons with the Sunyaev-Zel'dovich effect from filaments, *A&A* **624**, p. A48 (April 2019).
10. H. Tanimura, G. Hinshaw, I. G. McCarthy, L. Van Waerbeke, N. Aghanim, Y.-Z. Ma, A. Mead, A. Hojjati and T. Tröster, A search for warm/hot gas filaments between pairs of SDSS Luminous Red Galaxies, *MNRAS* **483**, 223 (Feb 2019).
11. S. G. Gallego, S. Cantalupo, S. Lilly, R. A. Marino, G. Pezzulli, J. Schaye, L. Wisotzki, R. Bacon, H. Inami, M. Akhlaghi, S. Tacchella, J. Richard, N. F. Bouche, M. Steinmetz and M. Carollo, Stacking the Cosmic Web in fluorescent Ly α emission with MUSE, *MNRAS* **475**, 3854 (April 2018).
12. F. Nicastro, J. Kaastra, Y. Krongold, S. Borgani, E. Branchini, R. Cen, M. Dadina, C. W. Danforth, M. Elvis, F. Fiore, A. Gupta, S. Mathur, D. Mayya, F. Paerels, L. Piro, D. Rosa-Gonzalez, J. Schaye, J. M. Shull, J. Torres-Zafra, N. Wijers and L. Zappacosta, Observations of the missing baryons in the warm-hot intergalactic medium, *Nature* **558**, 406 (Jun 2018).

13. O. E. Kovács, Á. Bogdán, R. a. K. Smith, R. P. Kraft and W. R. Forman, Detection of the Missing Baryons toward the Sightline of H1821+643, *ApJ* **872**, p. 83 (Feb 2019).
14. R. A. Sunyaev and Y. B. Zeldovich, The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies, *Comments on Astrophysics and Space Physics* **4**, p. 173 (November 1972).
15. R. A. Sunyaev and I. B. Zeldovich, The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement, *MNRAS* **190**, 413 (February 1980).
16. Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, E. Aubourg, J. Aumont, C. Baccigalupi, A. J. Banday and R. B. Barreiro, Planck intermediate results. XXXVII. Evidence of unbound gas from the kinetic Sunyaev-Zeldovich effect, *A&A* **586**, p. A140 (Feb 2016).
17. E. Schaan, S. Ferraro, M. Vargas-Magaña, K. M. Smith, S. Ho, S. Aiola, N. Battaglia, J. R. Bond, F. De Bernardis and E. Calabrese, Evidence for the kinematic Sunyaev-Zel'dovich effect with the Atacama Cosmology Telescope and velocity reconstruction from the Baryon Oscillation Spectroscopic Survey, *Phys. Rev. D* **93**, p. 082002 (Apr 2016).
18. P. G. Ferreira, R. Juszkiewicz, H. A. Feldman, M. Davis and A. H. Jaffe, Streaming Velocities as a Dynamical Estimator of Ω, *ApJ* **515**, L1 (April 1999).
19. N. Hand, G. E. Addison, E. Aubourg, N. Battaglia, E. S. Battistelli, D. Bizyaev, J. R. Bond, H. Brewington, J. Brinkmann and B. R. Brown, Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect, *Phys. Rev. Lett.* **109**, p. 041101 (Jul 2012).
20. S. Ferraro, J. C. Hill, N. Battaglia, J. Liu and D. N. Spergel, Kinematic Sunyaev-Zel'dovich effect with projected fields. II. Prospects, challenges, and comparison with simulations, *Phys. Rev. D* **94**, p. 123526 (Dec 2016).
21. C. Hernandez-Monteagudo, J. Chaves-Montero and R. E. Angulo, Angular Redshift Fluctuations: A New Cosmological Observable, *arXiv e-prints* , p. arXiv:1911.12056 (November 2019).
22. D. H. Jones, W. Saunders, M. Colless, M. A. Read, Q. A. Parker, F. G. Watson, L. A. Campbell, D. Burkey, T. Mauch, L. Moore, M. Hartley, P. Cass, D. James, K. Russell, K. Fiegert, J. Dawe, J. Huchra, T. Jarrett, O. Lahav, J. Lucey, G. A. Mamon, D. Proust, E. M. Sadler and K.-i. Wakamatsu, The 6dF Galaxy Survey: Samples, observational techniques and the first data release, *MNRAS* **355**, 747 (December 2004).
23. D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot and et al., SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems, *AJ* **142**, p. 72 (September 2011).
24. K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos, J. E. Gunn, H. Guo, D. Haggard, P. B. Hall, J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A. Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee,

D. C. Long, C. P. Loomis, B. Lundgren, R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDonald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza, M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, A. G. Sánchez, C. Sayres, S. J. Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Shelden, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S. A. Smee, S. A. Snedden, F. Stauffer, O. Steele, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake, M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M. Wood-Vasey, C. Yeche, I. Zehavi, G.-B. Zhao and Z. Zheng, The Baryon Oscillation Spectroscopic Survey of SDSS-III, *AJ* **145**, p. 10 (January 2013).

25. A. D. Myers, N. Palanque-Delabrouille, A. Prakash, I. Pâris, C. Yeche, K. S. Dawson, J. Bovy, D. Lang, D. J. Schlegel, J. A. Newman, P. Petitjean, J.-P. Kneib, P. Laurent, W. J. Percival, A. J. Ross, H.-J. Seo, J. L. Tinker, E. Armengaud, J. Brownstein, E. Burtin, Z. Cai, J. Comparat, M. Kasliwal, S. R. Kulkarni, R. Laher, D. Levitan, C. K. McBride, I. D. McGreer, A. A. Miller, P. Nugent, E. Ofek, G. Rossi, J. Ruan, D. P. Schneider, B. Sesar, A. Streblyanska and J. Surace, The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Quasar Target Selection, *ApJS* **221**, p. 27 (December 2015).

Searching for Intermediate Mass Black Holes in the Milky Way's galactic halo

A. Franco*, A.A. Nucita, F. De Paolis, F. Strafella and M. Maiorano

Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Via per Arnesano, I-73100, Italy, INFN, Sezione di Lecce, Via per Arnesano, I-73100, Italy

**E-mail: antonio.franco@le.infn.it*

Intermediate Mass Black Holes (IMBHs) are a class of black holes with masses in the range $10^2 \div 10^5\ M_\odot$, which cannot directly derive from stellar evolution. Looking for these objects and estimating their abundance is important not only for a deeper understanding of their origin but also for unveiling the nature and distribution of the dark matter in the galactic halo. Since February 2018 to January 2020, the Large and Small Magellanic Cloud have been intensively monitored by the DECAM instrument, installed on the 4m V. Blanco Telescope (CTIO, Chile) with the main objective to discover microlensing events possibly due to IMBHs.

Here we outline the developed data analysis pipeline. We have tested it versus known variable sources finding many not previously known variables objects. A few sources show a light curve similar to that expected for a microlensing event, but further analysis is required to confirm the microlensing nature of these events.

For these sources, and in particular for the uncatalogued variable stars, we try to determine if they are periodic or not via a periodogram analysis.

Keywords: Intermediate Mass Black Holes, ISIS Subtraction package, Large Magellanic Cloud, Small Magellanic Cloud, DECAM, Microlensing, Variable stars.

1. Introduction

During the last years of XVIII century, the English pastor John Michell and the mathematician Pierre Simon de Laplace imagined the existence of some dark objects, which were called *dark stars*.[1] The given name derived from the fact that if the radius of an object with mass M turns out to be $R = 2GM/c^2$ (where G is Newton's gravitational constant and c is the light speed), it cannot be visible from outside and therefore it would "appear" as *dark*.

The first actual mathematical description was provided in 1916 by Karl Schwarzschild with his solution of Einstein's field equations, representing the gravitational field around a non-rotating, spherical, and electrically neutral mass.

Half a century later, in 1967, John Wheeler coined the term *black hole* (BH) to indicate an object whose gravitational field is so large that does not let anything, even light, escape from its gravitational attraction. Unlike the classical formalization, General Relativity predicts the existence of an event horizon located at $r = r_S = 2GM/c^2$, which is called the Schwarzschild radius.

2. Black Holes Classification

There are three main kinds of black holes: stellar, intermediate, and supermassive BHs. The most common ones are stellar and supermassive BHs: the first one, with a mass less than $30\,M_{\odot}$, are produced by stellar death, when very massive stars with $M > 25 - 30\,M_{\odot}$ reach the and of their lives.

Supermassive BHs, with $M \simeq 10^5 \div 10^{10}\,M_{\odot}$, are located at the center of almost every galaxy in the Universe and are also responsible for the AGN activity.

The most unknown and mysterious intermediate-mass black holes (IMBHs) have a hypothetical mass in the range $30 \div 10^5\,M_{\odot}$ and might provide the Galactic Halo's dark matter contribution. Their formation and evolution are not well known as we are not able to observe many of these objects. There are four main possible mechanisms for IMBHs[a] formation[11]:

- Evolution of Population III stars
- Stellar collisions in dense star clusters
- Primordial Black Holes[12–17]

3. Gravitational Microlensing

Gravitational microlensing may offer an efficient way to detect BHs if they populate the Galactic Halo. This phenomenon shows up when an object (the lens) stands in the way between the observer and a faraway source. Photons coming from the source are deflected by the gravitational field of the lens (see Ref. 18 for a review). Depending on the phenomenon geometry, we might observe virtual images of the source. If the images are separated by large enough angles, as happens in the case of lensed quasars, one can detect double or multiple images of the same source. In the case of unresolved images, we have a microlensing event. It might be possible to observe the luminosity variation of the source, that is called *amplification*, which follows the *Paczyński law*[19]:

$$\mu(t) = \frac{u(t)^2 + 2}{u(t)\sqrt{u(t)^2 + 4}} \tag{1}$$

where $u(t) = \beta(t)/\theta_E$ is the Lens-Source angular separation ($\beta(t)$) in Einstein's angle (θ_E) unit.

4. Searching for Intermediate Mass Black Holes with DECAM

Searching for IMBHs might be important to obtain significant information regarding the evolution of black holes and constrain their contribution to the halo dark

[a]Searches for IMBHs have been attempted recently in dwarf spheroidal galaxies (dSph) where these objects are expected to form (see, e.g., Ref. 2–9). In fact, extrapolating the fundamental M_{BH} - M_{Bulge} relation (see, e.g., Ref. 10, for the supermassive BHs case) down to typical dSph masses, black holes in the typical IMBH mass range arise.

matter. This study could give major contributions about PBHs as dark matter candidates.[20,21] This study can be carried out using the microlensing phenomenon and monitoring some regions in the sky with a very high stellar density, such as the Magellanic Clouds.

The microlensing rate toward the Large Magellanic Cloud (LMC in the following) is estimated to be[22]:

$$\Gamma \approx \frac{1.66 \times 10^{-6}}{(M/M_{\odot})^{1/2}} \text{ event/year} \tag{2}$$

and, given $N_0 \approx 10^7$ the number of LMC stars, the number of expected microlensing events for solar mass lenses is:

$$N_{ev} = N_0\, \Gamma(\mathrm{M})\, t_{exp}\, \varepsilon(\mathrm{M}) \approx 50 \text{ events} \tag{3}$$

where $t_{exp} \simeq 2$ years is the duration of the survey and $\varepsilon \approx 0.1$ is the detection efficiency. Taking into account microlensing events with a duration up to two years, is possible to estimate the detectable mass limit that is $\lesssim 300\, M_{\odot}$: we can therefore observe lenses in the IMBH mass range.

4.1. *DECAM instrument*

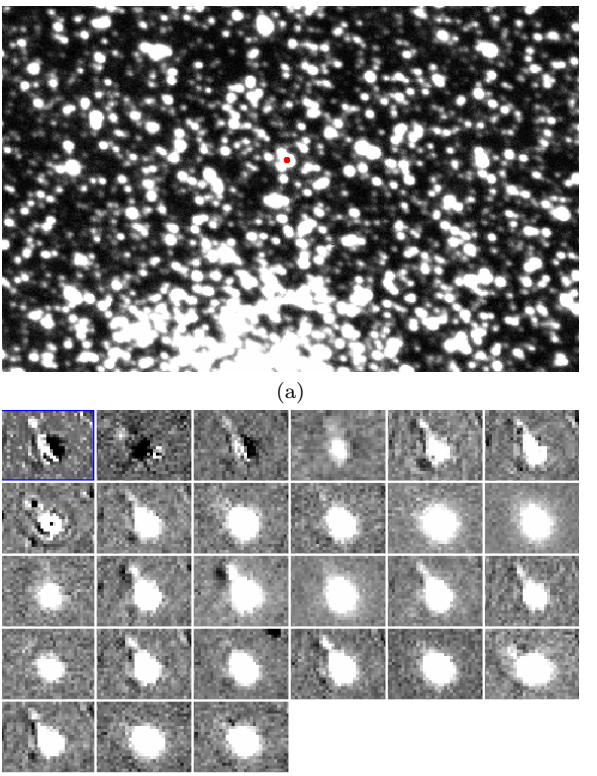
DECAM[23] (Dark Energy CAMera) is a high-performance, wide-field CCD imager mounted at the prime focus of the Victor M. Blanco 4m Telescope (CTIO, Chile). This camera consists of a grid of 62 CCDs (2048×4096 Pixels) which cover a field of view of $\sim 3\,\mathrm{deg}^2$ (~ 2.2 degree wide) with 0.263 arcsecond/pixel resolution. It uses a set of filters similar to those employed in the Sloan Digital Sky Survey (g,r,i,z,Y filters). It is active since September 2012 on behalf of the Dark Energy Survey (DES), finished in January 2019.

4.2. *Blanco/DECAM survey*

The proposal considered for this work, whose Principal Investigator is William Dawson, is related to the current rate of LIGO events that favors such a large abundance that IMBHs would make up the majority of dark matter. A Blanco/DECAM microlensing survey is useful to search for similar IMBHs in our Milky Way.[24] The survey was designed to intensively observe the Magellanic Clouds for two years, from February 2018 to January 2020, acquiring several images toward 23 LMC fields and 6 SMC fields. In particular, the LMC survey, currently under study, has continuously monitored a portion of sky with $4\mathrm{h}\,40\mathrm{m} < \alpha < 6\mathrm{h}\,20\mathrm{m}$ and $-73° < \delta < -65°$. The images have been acquired with a sampling of 7-10 days between each observation night, and a sub-sampling of few hours, acquiring 3-4 images during the same night. The final result is about 40 images acquired for each field and for each filter used (mainly in g and r bands, with just few images obtained in the i band).

In particular, for our aim, we have used 36 resampled images (bias, flat and dark subtracted) in the g band acquired toward the Large Magellanic Cloud.

4.3. *Detection method*



(a)

(b)

Fig. 1. (a) Small frame within the observed field of view toward LMC. The red dot indicates an identified microlensing candidate at coordinates $\alpha = 84.126\,738°, \delta = -68.797\,069°$. (b) Differential luminosity over time for the microlensing candidate event.

In order to search for microlensing events, it is essential the detection of variable sources in the images. This goal can be achieved using the ISIS 2.2 subtraction package.[25,26] The ISIS software, developed by Christophe Alard since the last years of the 90s, allows to perform the photometry of different objects through images difference, especially in very crowded field like LMC and SMC. The software works in multiple steps:

- Interpolation: images alignment;
- Creation of a reference image (from N selected "good" images from the hole sample) to compute the subtraction process;
- Subtraction between the reference image and each image of the sample, highlighting the brightness difference for variables;
- Detection of variable sources;
- Execution of the photometry for these objects.

Once a variable and potentially interesting source is identified, we use the Schwarzenberg-Czerny algorithm,[27] providing the phased light curve, and attempting an estimate of the period of the source. This algorithm gives more reliable results, in particular for high signal/noise ratio values, where the Lomb-Scargle algorithm is not efficient enough.[28]

4.4. *Microlensing candidate analysis*

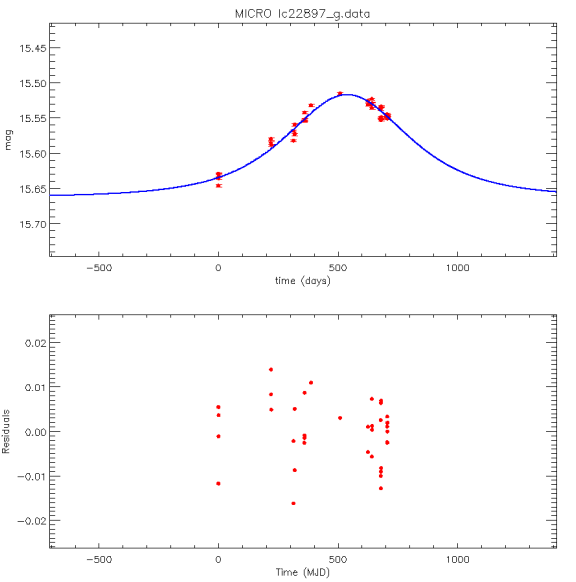


Fig. 2. The upper box shows the magnitude of the variable sources over two years of data (red dots) and the fitted Paczyński curve. The bottom box shows the residuals.

One of the most interesting variable sources detected so far is a microlensing like curve that seems to be associated with an object in the direction of LMC ($\alpha = 84.126\,738°, \delta = -68.797\,069°$). The field of view containing this source is shown in Fig. 1(a).

Using the ISIS software it appears that this source is clearly variable, as shown in Fig. 1(b). The corresponding light curve is shown in Fig. 2 from which it is apparent a Paczynski-like behavior, typical of a microlensing event. The fit and the residuals are shown in the upper and lower panels of Fig. 2, respectively.

The fit returns an estimate of the event Einstein time $t_E = (246.4 \pm 31.7)$ days and the baseline magnitude $m_0 = 15.66 \pm 0.01$. We note that the indicated magnitude represents the instrumental magnitude of the camera, without any photometric correction. In addition, it is important to understand the geometry of the phenomenon in order to estimate the lens mass. For these reasons new images toward

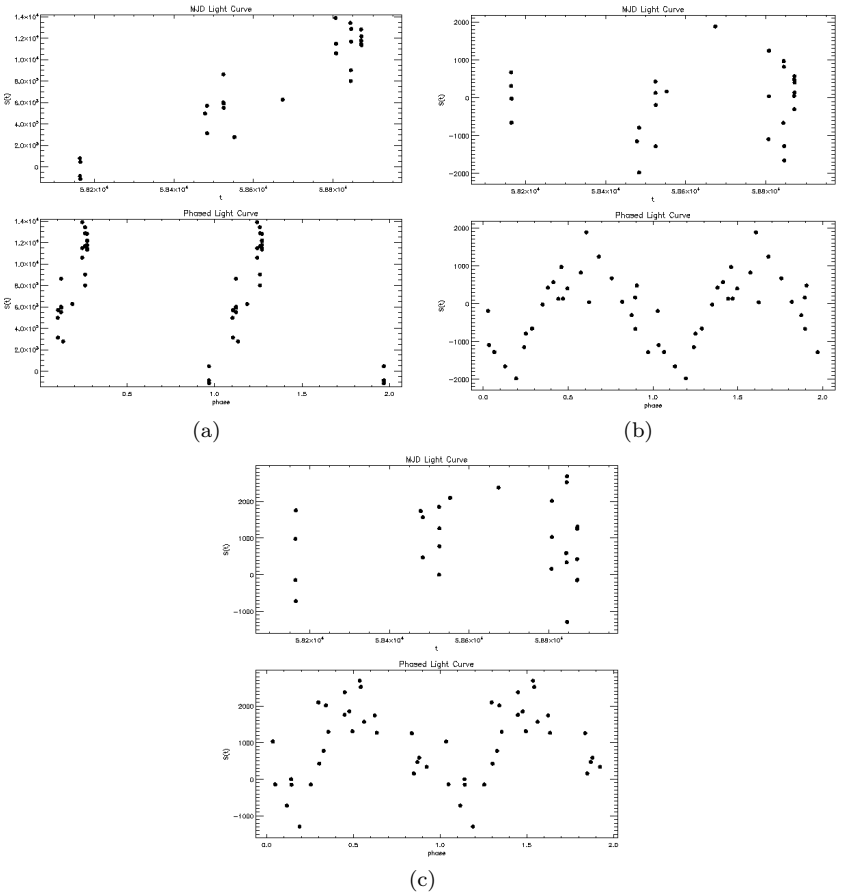


Fig. 3. The MJD light curves (upper panel) and phased light curve (lower panel) for three different variable sources. (a) Non-periodic variable and (b), (c) eclipsing binary candidates.

the particularly interesting fields will be acquired in the next future in order to provide a better and complete results, that will be published after completion of the work.

4.5. *Periodic variable sources candidate analysis*

ISIS 2.2 and the Schwarzenberg-Czerny algorithm enable the recognition of a few periodic variables and allow to obtain an estimate of their period. This can be used both to confirm previously observed sources and to detect new ones. The observed data for three variable sources toward the LMC are shown in Fig. 3. Panel (a) shows the light curve for non-periodic variable source, at coordinates $\alpha_a = 84.530\,329°, \delta_a = -68.791\,993°$, displaying the increasing flux over time and highlighting the absence of a particular periodicity in the phased light curve. Panels (b)

and (c) show two sources, at coordinates $\alpha_b = 84.486\,772°, \delta_b = -68.799\,780°$ and $\alpha_c = 84.303\,917°, \delta_c = -68.741\,760°$, displaying the variation of the flux with respect to the time (upper) and over two phases (bottom), enabling the estimate of the period that is $P_b \simeq 1.69$ days and $P_c \simeq 0.31$ days.

5. Conclusions

The detection of periodic and non-periodic variable sources has allowed confirming the goodness of the developed pipeline. In particular, the two years time window, the sampling and the efficiency of the DECAM instrument have shown the possibility to detect microlensing events. So far, many known variable sources have been identified in four DECAM fields of view - toward LMC - and some unknown variables are waiting to be studied more deeply. A few possible microlensing event candidates (e.g. that shown in Fig. 2), which wait for confirmation, have been also found in the data.

The contents presented in this paper will be discussed in more details in the the paper "*Searching for galactic halo Intermediate Mass Black Holes through gravitational microlensing*", by A. Franco et al., currently under preparation.

Acknowledgments

We acknowledge the support by the Euclid and TAsP (Theoretical Astroparticle Physics) projects of the Istituto Nazionale di Fisica Nucleare (INFN).

References

1. S. J. Crothers, *"A Brief History of Black Holes"*, Progress in Physics, Volume 2, Issue 2, pp.54-57 (2006).
2. L. Manni et al., *"A XMM-Newton observation of a sample of four close dwarf spheroidal galaxies"*, Monthly Notices of the Royal Astronomical Society, Volume 451, Issue 3, p.2735-2749 (2015).
3. A. Nucita et al., *"Hint for a faint intermediate mass black hole in the Ursa Minor dwarf galaxy"*, New Astronomy, Volume 23, p. 107-112 (2013).
4. A. Nucita et al., *"An XMM-Newton search for X-ray sources in the Fornax dwarf galaxy"*, Astronomy and Astrophysics, Volume 550, id.A18, 13 pp. (2013).
5. A. Nucita et al., *"The high energy search for IMBHs in close dSph Milky Way satellites"*, Memorie della Societa Astronomica Italiana, v.84, p.645 (2013).
6. A. Nucita et al., *"The high energy view of NGC 6388: hints for an IMBH?"*, Memorie della Societa Astronomica Italiana, v.84, p.643 (2013).
7. A. Nucita et al., *"The XMM-Newton slew view of IGRJ17361-4441: A transient in the globular cluster NGC 6388"*, New Astronomy, Volume 17, Issue 6, p. 589-593 (2012).
8. A. E. Reines et al., *"Dwarf Galaxies with Optical Signatures of Active Massive Black Holes"*, The Astrophysical Journal, Volume 775, Issue 2, article id. 116, 24 pp. (2013).
9. S. M. Lemons et al., *"An X-Ray Selected Sample of Candidate Black Holes in Dwarf Galaxies"*, The Astrophysical Journal, Volume 805, Issue 1, article id. 12, 10 pp. (2015).

10. J. Magorrian et al., *"The Demography of Massive Dark Objects in Galaxy Centers"*, The Astronomical Journal, Volume 115, Issue 6, pp. 2285-2305 (1998).
11. J. E. Greene, J. Strader, L. C. Ho, *"Intermediate-Mass Black Holes"*, Annual Review of Astronomy and Astrophysics, vol. 58, p.257-312 (2020).
12. Y. B. Zel'dovich and I. Novikov, *The hypothesis of cores retarded during expansion and the hot cosmological model*, Astronomicheskii Zhurnal, Vol. 43, p.758 (1966).
13. S. Hawking, *Gravitationally collapsed objects of very low mass*, Monthly Notices of the Royal Astronomical Society, Vol. 152, p. 75 (1971).
14. B. J. Carr, S. W. Hawking, *"Black holes in the early Universe"*, Monthly Notices of the Royal Astronomical Society, Vol. 168, p. 399-416 (1974).
15. P. H. Frampton, *"The Primordial Black Hole Mass Range"*, Modern Physics Letters A, Volume 31, Issue 12, id. 1650064 (2016).
16. J. Garcìa-Bellido, *Massive primordial black holes as dark matter and their detection with gravitational waves*, Journal of Physics: Conference Series, (1) 2017.
17. M. Yu. Khlopov, *"Primordial Black Holes"*, Research in Astronomy and Astrophysics, Volume 10, Issue 6, pp. 495-528 (2010).
18. F. De Paolis et al., *"The Scales of Gravitational Lensing"*, Universe, vol. 2, issue 1, p. 6, https://doi.org/10.3390/universe2010006 (2016).
19. B. Paczyński, *"Gravitational microlensing by the galactic halo"*, Astrophysical Journal v.304, p.1-5 (1986).
20. N. Afshordi, P. McDonald and D. Spergel, *Primordial black holes as dark matter:the power spectrum and evaporation of early structures*, The Astrophysical Journal, Volume 594, Issue 2, pp. L71-L74 (2003).
21. B. Carr, F. Kühnel and M. Sandstad, *Primordial black holes as dark matter*, Physical Review D, Volume 94, Issue 8, id.083504 (2016).
22. K. Griest, *"Galactic microlensing as a method of detecting massive compact halo objects"*, Astrophysical Journal v.366, p.412-421 (1991).
23. R. A. Shaw, *"NOAO Data Handbook (Version 2.2)"*, Tucson: National Optical Astronomy Observatory (2015).
24. W. Dawson, *"Program Information for 2018A-0273"*, Blanco/DECAM survey proposal, https://www.noao.edu/noaoprop/abstract.mpl?2018A-0273.
25. C. Alard, R. H. Lupton, *"A method for optimal image subtraction"*, The Astrophysical Journal, Volume 503, Issue 1, pp. 325-331 (1998).
26. C. Alard, *"Image subtraction using spacially variable Kernel"*, Astronomy & Astrophysics, Astron. Astrophys. Suppl. Ser. 144, 363–370 (2000).
27. A. Schwarzenberg-Czerny, *"Accuracy of period determination"*, Monthly Notices of the Royal Astronomical Society (ISSN 0035-8711), vol. 253, Nov. 15, 1991, p. 198-206 (1991).
28. A. Schwarzenberg-Czerny, *"Fast and Statistically Optimal Period Search in Uneven Sampled Observations"*, Astrophysical Journal Letters v.460, p.L107 (1996).

Virial clouds evolution from the last scattering up to the formation of first stars

Noraiz Tahir[1,2,*], Asghar Qadir[3,†], Muhammad Sakhi[4,‡], and Francesco De Paolis[1,2,§]

[1] *Department of Mathematics and Physics "E. De Giorgi", University of Salento, Via per Arnesano, I-73100 Lecce, Italy*

[2] *INFN, Sezione di Lecce, Via per Arnesano, I-73100 Lecce, Italy*

[3] *Abdus Salam School of Mathematics, G.C. University, Lahore, Pakistan*

[4] *Department of Physics, Quaid-e-Azam University, Islamabad, Pakistan*

[*] *noraiz.tahir@le.infn.it*

[†] *asgharqadir46@gmail.com*

[‡] *sakhi.cosmos@gmail.com*

[§] *francesco.depaolis@le.infn.it*

The asymmetry in the cosmic microwave background (CMB) towards several nearby galaxies detected by *Planck* data is probably due to the rotation of "cold gas" clouds present in the galactic halos. In 1995 it had been proposed that galactic halos are populated by pure molecular hydrogen clouds which are in equillibrium with the CMB. More recently, it was shown that the equillibrium could be stable. Nevertheless, the cloud chemical composition is still a matter to be studied. To investigate this issue we need to trace the evolution of these virial cloud from the time of their formation to the present, and to confront the model with the observational data. The present paper is a short summary of a paper.[1] Here we only concentrate on the evolution of these clouds from the last scattering surface (LSS) *up to* the formation of first generation of stars (population-III stars).

Keywords: Dark matter; cosmic microwave background radiations; spiral galaxies; molecular clouds; galactic halos.

1. Introduction

The study of the nature of galactic halos and their dynamics is a task that is difficult to address. Here we present a summary of a more detailed analysis addressing this issue.[1]

The Λ cold dark matter (ΛCDM) model entails that $\sim 5\%$ of our Universe is made of baryonic matter,[2] of which $\sim 60\%$ is detected,[3–7] but $\sim 40\%$ is still undetected at present. This is the so-called "missing baryon problem".

In 1995 it was proposed that a fraction, f, of these missing baryons is present in the galactic halos in the form of *pure* molecular hydrogen (H_2) clouds which are in equillibrium with the CMB.[8] The difficutly was on the observation of such "chameleons" megerd with the background. One of the suggestions was, to look for a Doppler shift effect due to the rotation of galaxies, assuming that the rotation of these clouds is synchronized with the rotation of galactic halos, and hence they should be Doppler shifted, those clouds roatating towards us should give a blue-shift effect while those rotating away from us would give a red-shifted contribution.

In 2011 WMAP data was analyzed for M31. The analysis revelaed a temperature asymmetry in the CMB which was almost frequency independent,[10] which was a strong indication of the Doppler shift effect due to the galactic halo rotation. This opens up a window to observe these clouds and to study the baryonic content of the galactic halos.[11] Soon after WMAP, in 2014 *Planck* data towards M31 was analyzed and the asymmetry was seen at a more precise level.[12] A temperature asymmetry was also detected towards several nearby spiral galaxies.[13–16]

There was more than one item of evidence of the predicted Doppler shift effect, but observing that there is a Doppler shift due to the halo rotation does not reveal the true nature of the effect, i.e. if it is partially or fully due to the molecular clouds in the halos, or, if is there anything else that could give a masking or mimiking effect in the asymmetry. Another unanswered question is the chemical composition of these clouds: are they pure H_2 clouds, or there is some contamination of dust or heavier molecules in them? It is quite obvious that galactic halos contain a significant fraction of dust that *should* contaminate these clouds,[18] so one needs to model the clouds.

As these clouds should survive on acount of the virial theorem, they were called "virial clouds". These clouds were modeled and it was seen that at the current CMB temperature the centeral density of pure H_2 clouds was $\approx 1.60 \times 10^{-18}$kg m^{-3}, similarly the mass and radius were $\approx 1.93 \times 10^{-4} M_{\odot}$, and ≈ 0.032 pc.[19] The change in their physical parameters with the contamination of heavier molecules and dust were also estimated. It was seen that as the contamination of dust and hevier molecules were increased in the clouds, they became denser, and their mass and radius decreased.[19–21] An objection was raised on the stability of these clouds, as it was believed that molecular clouds can not be stable at this, very low, CMB temperature, as there would be no mode that could be excited by the photons and the cloud might collapse to form stars or other planetary objects,[22] but it was demonstrated that this equilibrium *does* arise on account of the translational mode, despite its extremely small probability, because of the size of the virial clouds and the time scales available for thermal equilibrium to be reached, so that the time required for thermalization is much less than that required for collapse.[19]

Modeling the virial clouds and estimating the change in physical parameters with the contamination of heavier molecules and dust, and observing the CMB temperature asymmetry by *Planck* data still does not answer the question on the nature of virial clouds and the exact cause of the observed asymmetry in the CMB. One has to run the clock back and trace the evolution of virial clouds when they were formed at the LSS to the present. This task needs to be done in two phases: (*i*) from LSS up to the formation of population-III stars; and (*ii*) from the formation and explosion of population-III stars to the present. Here we discuss the first part of evolution, as various qualitative changes took place during the formation of population-III stars.[23] Hence there will be significant changes during the second step, which will be studied more clearly and in more detail later.

2. First epoch of virial clouds evolution

Virial clouds would have formed at $z = 1100$, the last scattering time (LSS) and they evolved in their chemical composition and physical parameters, but in order to maintain their stability they survived the collapse and stayed in quasi-static equilibrium with the CMB since then. It is quite obvious that when they formed they should have had the primordial chemical composition, i.e. $\sim 75\%$ atomic hydrogen and $\sim 25\%$ helium. In addition to H and He there were other atoms and molecules like deuterium, helium-3, lithium and molecular hydrogen, which could have contributed to the virial clouds, but their fraction was negligilble as compared to H and He. Hence these molecules and atoms could not have any significant effect on the virial cloud physical parameters. As a result the ratio of atomic hydrogen and helium would remain the same and these would be the main component to form the virial clouds during the period, but there should be a fast change in their chemical composition after the formation and explosion of population-III stars.

2.1. *Hydrogen-Helium virial clouds*

Since virial clouds must be considered to be in thermal equilibrium because they are embedded in the heat bath of the CMB, we need to use the canonical distribution function for a fixed temperature and use the cooling of the heat bath to provide a quasi-equilibrium. Moreover, these clouds should start to form in the potential well of cold dark matter (CDM). As the clouds are thermalized the potential well will not cause them to collapse to form population-III stars,[24] but will modify the physical parameters of virial clouds.

To obtain a general expressions we consider a virial cloud composed of an arbitrary mixture of H and He, with mass fractions α and β. Then, we use the primordial cosmological fractions of H and He for the final computation. The total mass of the cloud is, obviously, $M_{cl}(r) = \alpha M_H(r) + \beta M_{He}(r)$, with the condition $\alpha + \beta = 1$. The density distribution for two fluids is given by[19]

$$\rho_{cl}(r) = \sqrt{\frac{64}{27}} \frac{(G\rho_{c_H}\rho_{c_{He}})^{3/2}}{(k_B T)^{9/2}} (m_H m_{He})^{5/2} \\ exp\left[-\frac{1}{2}\left(\frac{\alpha G M_H(r) m_H}{r k_B T} + \frac{\beta G M_{He}(r) m_{He}}{r k_B T}\right)\right]. \quad (1)$$

and the corresponding differential equation can be written as[19]

$$r\frac{d\rho_{cl}(r)}{dr} - r^2\left(\frac{2\pi G}{k_B T}\right)[\rho_{cl}(r)(\alpha\rho_{c_H} m_H \\ + \beta\rho_{c_{He}} m_{He})] - \rho_{cl}(r)\ln\left(\frac{\rho_{cl}(r)}{\tau}\right) = 0, \quad (2)$$

where, $\tau = (8/3\sqrt{3})[(G\rho_{c_H}\rho_{c_{He}})^{3/2}/(k_B T)^{9/2}][m_H m_{He}]^{5/2}$, ρ_{c_H} is the central density of H cloud, $\rho_{c_{He}}$ is the central density of He cloud, m_H, the mass of single atom of hydrogen, and m_{He}, the mass of single atom of helium.

We use eq. (2) to estimate the central density of the clouds. We estimated the central density, Jeans mass and radius of the two-fluid virial clouds with primordial fraction of H, and He, i.e. $\alpha = 0.75$, and $\beta = 0.25$. In order to solve eq. (1) numerically, we assumed a guess value of ρ_c, at a fixed temperature, and estimate where the density becomes *exactly* zero at the boundary. We then compare the Jeans radius with that central density with the value available to us. We adjust the central density so that the density becomes zero exactly at the Jeans radius. In this way we get a self-consistent solution of the differential equation subject to the given boundary conditions. Next, we decrease the temperature and repeat the process.

It was seen that with the decrease in temperature and redshift, the density of virial clouds increased, at $z = 1100$ these clouds were $\approx 8.53 \times 10^{-20} kgm^{-3}$ and at $z = 50$ the density of the clouds was $\approx 6.05 \times 10^{-19}$kg m^{-3}. On the other hand the mass decreased with the decrease in temperature and so as the radius of the clouds. These clouds were massive ≈ 50 pc, with a mass $\approx 3.05 \times 10^5 M_\odot$ at $z = 1100$, then at $z = 50$ the clouds were ≈ 3.54 pc with a mass of $\approx 7.54 \times 10^3 M_\odot$.

3. Results and discussion

The first stage of the virial cloud evolution turns out to be quite simple. We have seen that virial clouds became denser with time, and they lost mass and shrunk in size. The physics of the second stage of evolution will not be that simple since there will be cooling due to the formation of molecular hydrogen, turbulence and the angular momentum effects in the virial cloud after the formation and explosion of population-III stars. We need to consider all these things while tracing the second epoch of evolution, and this will be done in the later paper.

One could expect molecular hydrogen to be formed in the first epoch, but during the era of study in the current summarized paper any H_2 molecules formed in the clouds will be unstable and dissociate due to the radiation pressure at that time, since H_2 needs dust particles to remain stable. The molecular cooling of such clouds played a vital role after the recombination and thir cooling rate has been analyzed from $T \sim 120 - 3$ K.[25] Hence, we do not need to consider the effect of cooling from molecular hydrogen till population-III stars exploded. It may also be expected that the electronic transition mode in hydrogen and helium atoms by the CMB photons. One can check that most of the CMB photons at 3000 K do not have enough energy to excite the electrons of a hydrogen atom from its ground state, but the higher energy tail does have sufficient energy to do so. The percentage of such photons is 0.016, which is quite low. Hence, there will be a negligible effect of this mode, but these effects will be more significant in the later stage of evolution.

Acknowledgements

FDP and NT acknowledge support from the TAsP and Euclid INFN projects.

References

1. N. Tahir et al., Evolution of virial clouds-I: From the surface of last scattering up to the formation of population-III stars, *Europ. Phys. J. C* **81**, 827 (2021).
2. A.R.P. Ade, et al., Planck 2015 results-xiii. cosmological parameters, *Astron. Astrophys.* **594**, A13 (2015).
3. A. S. Kassin et al., Dark and baryonic matter in bright spiral galaxies. II. Radial distributions for 34 galaxies, *Astrophys. J.* **643**, 804 (2006).
4. F. Nicastro et al., A decade of warm-hot intergalactic medium searches, where do we stand and where do we go?, *Astron. Nachr.* **338**, 281-286 (2017).
5. R. Dave et al., Baryons in the warm-hot intergalactic medium, *Astrophys. J.* **552**, 473 (2010).
6. J.T. Li et al., Baryon budget of the hot circumgalactic medium of massive spiral galaxies, *Astrophys. J. Letters* **855**, L24 (2018).
7. R. Cen et al., Where are the baryons?, *Astrophys. J.* **514**, 1 (1999).
8. F. De Paolis et al., A Case for a baryonic dark halo, *Pys. Rev. Letters* **74**, 14 (1995).
9. F. De Paolis et al., Observing molecular hydrogen clouds and dark massive objects in galactic halos, *Astron. Astrophys.* **299**, 647 (1995).
10. F. De Paolis et al., Possible detection of the M 31 rotation in WMAP data, *Astron. Astrophys.* **534**, L8 (2011).
11. F. De Paolis et al., CMB as a possible new tool to study the dark baryons in galaxies, *J. Phys.: Conference Series* **354**, 1 (2012).
12. F. De Paolis et al., Planck confirmation of the disk and halo rotation of M 31, *Astron. Astrophys.* **565**, L3 (2014).
13. F. De Paolis et al., Planck revealed bulk motion of Centaurus A lobes, *Astron. Astrophys.* **580**, L8 (2015).
14. V. G. Gurzadyan et al., Planck view of the M 82 galaxy, *Astron. Astrophys.* **582**, A77 (2015).
15. F. De Paolis et al., Triangulum galaxy viewed by Planck, *Astron. Astrophys.* **593**, A57 (2016).
16. V. G. Gurzadyan et al., TMessier 81's Planck view versus its halo mapping, *Astron. Astrophys.* **609**, A131 (2018).
17. F. De Paolis et al., Rotating baryonic dark halos, *Astron. Astrophys.* **629**, A87 (2019).
18. V.N. Yershov et al., Distant foreground and the Planck-derived Hubble constant, *Mon. Not. R. Astron. Soc.* **492**, 5052-5056 (2020).
19. A. Qadir et al., Virial clouds explaining the observed rotational asymmetry in the galactic halos, *Phys. Rev. D* **100**, 4043028 (2019).
20. N. Tahir et al., Constraining Baryons in the M31 galactic halo by Planck data, *Int. J. Mod. Phys.* **28**, 7 (2019).
21. N. Tahir et al., Seeing the halo rotation of nearby spiral galaxies using Planck data, *Arab. J. Math.* **8**, 193-199 (2019).
22. T. Padmanabhan, Statistical mechanics of gravitating systems, *Phys. Rep.* **188**, 285-362 (1990).
23. M. Jeon et al., Recovery from Population III supernova explosions and the onset of second-generation star formation, *Mon. Not. R. Astron. Soc.* **444**, 3288-3300 (2014).
24. R. Barkana et al., In the beginning: the first sources of light and the reionization of the universe, *Phys. Rep.* **349**, 125-238 (2001).
25. P., Denis et al., Thermal equilibrium of molecular clouds in cooling flow clusters, *Astron. Astrophys.* **345**, 723-732 (1999).

Testing Weyl-modified gravity on M31 and Milky Way

Muhammad Bilal* and Asghar Qadir†

Abdus Salam School of Mathematical Sciences, 68-B Model Town, Government College University, Pakistan

**Email: bilalchughtai799@gmail.com*

†*Email: asgharqadir46@gmil.com*

Lee and Qadir,[1] on the basis of various arguments, suggested using a gravitational Lagrangian with a coupling between the Weyl tensor and two copies of the stress-energy tensor, without testing the suggestion for any actual galaxy. As a first step, they used a constant density in their analysis. However unrealistic the constant density approximation is, here we follow up their first step by using their calculations to obtain the coupling constant for the extra term that fits for M31 and the Milky Way, to check if they are consistent within the crude approximation of the model.

Keywords: Dark matter; flat rotation curves; Weyl-induced gravity.

1. Introduction

While the cosmological constant explains the so-called "dark energy problem",[2] providing the Λ of the standard model of cosmology (SMC), the ΛCDM model, The standard model of particle physics (SMPP) ceased to explain the dark matter required to explain the observed dynamics of galaxies. Various suggestions to modify the law of gravity had emerged since the observations of Vera Rubin,[3] but all needed many parameters to solve a single problem, which is no use. Qadir, Lee and Kim[4] had suggested a modification of the Einstein-Hilbert Lagrangian to include an explicit interaction term, RT, on the grounds that it might also solve the problem of quantum gravity. However, the gravitational field is given by the Weyl tensor, $C^{\mu}{}_{\nu\rho\pi}$ and not the Ricci scalar, which gives the mass-energy.

Lee and Qadir[1] suggested a modification of the Lagrangian by including the interaction term, $\lambda C_{\mu\nu\rho\pi}T^{\mu\rho}T^{\nu\pi}$. They obtained the rotational velocity of particles at arbitrary distances from the centre if the density is known as a function of the radial distance. They went on to approximate it as a constant function as a first step, to simplify the calculation. The idea had been that the same interaction coupling must explain the dark matter for all galaxies and systems, or clusters, of galaxies, and manage to avoid the problem of a non-renormalizable theory of quantum gravity. There had been no basis for the latter, just a pious hope. For the former, on the other hand, it should reasonably work out. The calculations were not applied to any specific galaxy. In this brief note, we compare the results for the value of λ required to obtain the rotational curves for M31 and the Milky Way.

The model takes the Schwarzschild interior metric

$$ds^2 = -e^{\nu(r)}c^2dt^2 + e^{\mu(r)}dr^2 + r^2d\Omega^2 \ , \tag{1}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. Since the rotation curves are obtained by seeing stars in the outer reaches of the galaxies, the relevant radius is the radius of the visible part of the galaxy, the mass is the total mass contained within that radius and the test particle mass, m, is the mass of a typical star. For definiteness, we take it to be a solar mass.

Using the field equations obtained from the modified Lagrangian[1] we get

$$\mu(r) = F(r) - \ln\left[\left\{\frac{r}{1-\lambda rc^4\rho\rho'}\left(\int\frac{e^{F(r)}}{r^2}dr - \kappa c^2\int\rho(r)e^{F(r)}dr\right)\right\}\right] \ , \tag{2}$$

where

$$F(r) = \frac{1}{2}\int\frac{4+\lambda rc^4\left(r\rho^2\nu'^2 - 2r\rho\rho'\nu' - 2r\rho'^2 - 2r\rho\rho'' - 8\nu'\rho\rho'\right)}{r\left(1-\lambda rc^4\rho\rho'\right)}dr \tag{3}$$

and

$$\nu'(r) = \frac{1+\sqrt{1+2\lambda\rho^2c^4\left(e^{\mu(r)}-1\right)}}{2r\lambda\rho^2c^4} \ . \tag{4}$$

The square of the rotational velocity is

$$v^2(r) = \frac{1+\sqrt{1+2\lambda\rho^2c^4\left(e^{\mu(r)}-1\right)}}{2\lambda\rho^2c^2} \ , \tag{5}$$

and hence,

$$\lambda = \frac{2\left(\frac{v}{c}\right)^2 + e^{\mu(r)} - 1}{2\rho^2v^4} \ . \tag{6}$$

Following[1] we assume a constant density. Thus

$$e^{\mu(r)} = \frac{e^{F_{const}(r)}}{r\left(\int\frac{e^{F_{const}(r)}}{r^2}dr - \kappa\rho c^2\int e^{F_{const}(r)}dr\right)} \ , \tag{7}$$

where,

$$F_{const}(r) = 2\ln\left(\frac{rc^2}{2GM(r)}\right) + \frac{\lambda\rho^2c^4}{2}\int r\nu'^2dr \ , \tag{8}$$

where, M(r) is the mass interior to the radius r. It is worth mentioning that in[1] an r is missing inside the integral which has been corrected here. The square of the rotational velocity is determined numerically, taking r to be the radius of the visible galaxy and thus $M(r)$ to be the mass of the visible galaxy. As required, for positive λ, the "gravitational force" is effectively enhanced.

For M31, the Andromeda galaxy,[5] the radius of the visible halo is $\sim 34.7\ kpc$, and the mass $\sim 3.43 \times 10^{11}\ M_\odot$. The constant density comes out to be $1.97 \times 10^{-3}\ M_\odot/pc^3$. For the Milky Way[6] the radius of the visible halo is $\sim 31.5\ kpc$ and the mass $\sim 6.06 \times 10^{10}\ M_\odot$, yielding a density $\sim 4.66 \times 10^{-4}\ M_\odot/pc^3$. The rotational velocity of the visible halo of M31 is $\sim 226\ km/s$ and of the Milky Way is $\sim 220\ km/s$.

The rotational velocity curve does not come out flat for either galaxies, see Fig. 1. This is due to the crude assumption of a constant density. However the curves for both galaxies match for $\lambda = 1.14 \times 10^{16}\ km^2 s^4/kg^2$, thus verifying that both require the same interaction coupling.

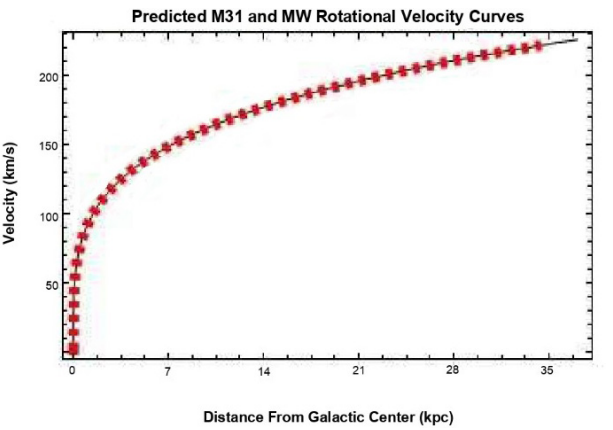


Fig. 1. The black line gives the predicted rotation curve for M31, and the red dotted line the predicted rotation curve for the Milky Way, for $\lambda = 1.14 \times 10^{16}\ km^2 s^4/kg^2$, having taken the rotational velocity at the edge of M31 and the Milky Way as 226 km/s, 220 km/s respectively.

For the purpose of checking that the proposal of the Weyl-interaction modification of gravity is viable, it would be necessary to check the proposal for a more realistic model (variable density) for a number of different galaxies, taking into account the baryonic dark matter in that region and distinguishing between the density for the disc matter and the bulge matter.

References

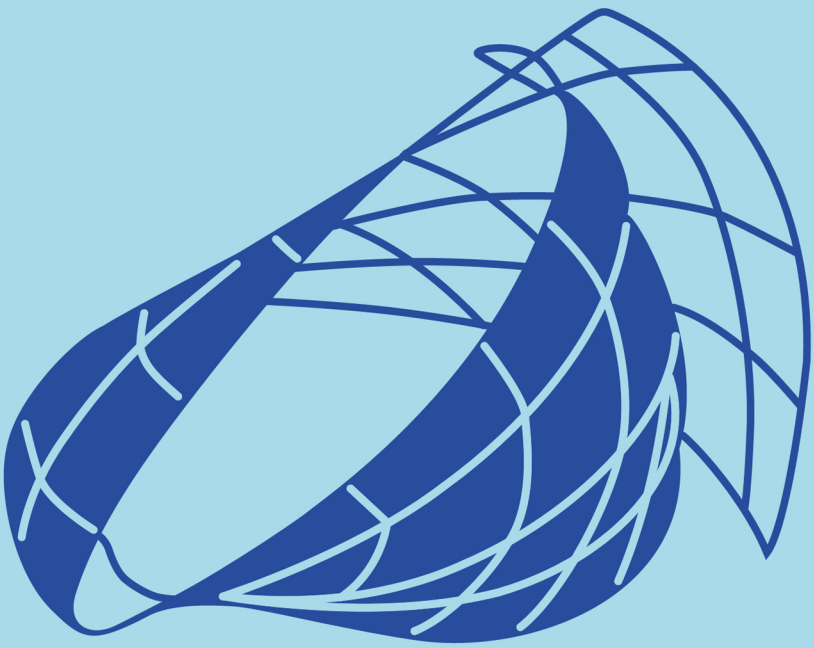
1. A. Qadir, H. W. Lee, *Int. J. Mod. Phys. D* **28**, (2019) 2040014.
2. A. Qadir, *Proc. MG16*, Eds. R. Ruffini, R. Jantzen and G. Vereshchigan (World Scientific 2021).
3. V. C. Rubin, *Ap. J* **451** (1995).
4. A. Qadir, H. W. Lee and K. Y. Kim, *Int. J. Mod. Phys. D* **26** (2017) 1741001.
5. C. Carignan, L. Chemin, W. K. Huchtmeier and F. J. Lockman *ApJ* **641** (2006) L109.
6. P. J. Mcmillan, *MNRAS* **465** (2017) 76.

This page intentionally left blank

PART C

PROCEEDINGS OF THE SIXTEENTH

MARCEL GROSSMANN MEETING ON GENERAL RELATIVITY



Editors

Remo Ruffini

Gregory Vereshchagin

World Scientific

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

This page intentionally left blank

PART C

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

Proceedings of the MG16 Meeting
on General Relativity
Online, 5–10 July 2021

Editors

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Gregory Vereshchagin
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Series Editor

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Names: Marcel Grossmann Meeting on General Relativity (16th : 2021 : Online) | Ruffini, Remo, editor. | Vereshchagin, Gregory, editor.

Title: The sixteenth Marcel Grossmann meeting on recent developments in theoretical and experimental general relativity, astrophysics, and relativistic field theories : proceedings of the MG16 meeting on general relativity online, 5–10 July 2021 / editors: Remo Ruffini, University of Rome "La Sapienza", Rome, Italy, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy, Gregory Vereshchagin, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy.

Other titles: Proceedings of the MG16 meeting on general relativity online, 5–10 July 2021

Description: New Jersey : World Scientific, [2023] | Includes bibliographical references.

Identifiers: LCCN 2022047088 | ISBN 9789811269769 (set ; hardcover) | ISBN 9789811266584 (v. 1) | ISBN 9789811266591 (v. 2) | ISBN 9789811266607 (v. 3) | ISBN 9789811266614 (v. 4) | ISBN 9789811269776 (set ; ebook)

Subjects: LCSH: General relativity (Physics)--Congresses. | Gravitation--Congresses. | Quantum gravity--Congresses. | Cosmology--Congresses. | Astrophysics--Congresses.

Classification: LCC QC173.6 .M37 2021 | DDC 523.01--dc23/eng/20220929

LC record available at https://lccn.loc.gov/2022047088

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2023 by Editors

All rights reserved.

This is an Open Access volume published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 (CC BY-NC) License. Further distribution of this work is permitted, provided the original work is properly cited.

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/13149#t=suppl

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

THE MARCEL GROSSMANN MEETINGS

Series Editor: REMO RUFFINI

Publications in the Series of Proceedings

Proceedings of the Sixteenth Marcel Grossmann Meeting on General Relativity
(Virtual Meeting, 2021)
Edited by G. Vereshchagin, R. Ruffini
World Scientific, 2022

Proceedings of the Fifteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2018)
Edited by E.S. Battistelli, R.T. Jantzen, R. Ruffini
World Scientific, 2022

Proceedings of the Fourteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2015)
Edited by M. Bianchi, R.T. Jantzen, R. Ruffini
World Scientific, 2017

Proceedings of the Thirteenth Marcel Grossmann Meeting on General Relativity
(Stockholm, Sweden, 2012)
Edited by K. Rosquist, R.T. Jantzen, R. Ruffini
World Scientific, 2015

Proceedings of the Twelfth Marcel Grossmann Meeting on General Relativity
(Paris, France, 2009)
Edited by T. Damour, R.T. Jantzen, R. Ruffini
World Scientific, 2012

Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity
(Berlin, Germany, 2006)
Edited by H. Kleinert, R.T. Jantzen, R. Ruffini
World Scientific, 2007

Proceedings of the Tenth Marcel Grossmann Meeting on General Relativity
(Rio de Janiero, Brazil, 2003)
Edited by M. Novello, S. Perez-Bergliaffa, R. Ruffini
World Scientific, 2005

Proceedings of the Ninth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2000)
Edited by V.G. Gurzadyan, R.T. Jantzen, R. Ruffini
World Scientific, 2002

Proceedings of the Eighth Marcel Grossmann Meeting on General Relativity
(Jerusalem, Israel, 1997)
Edited by T. Piran
World Scientific, 1998

Proceedings of the Seventh Marcel Grossmann Meeting on General Relativity
(Stanford, USA, 1994)
Edited by R.T. Jantzen and G.M. Keiser
World Scientific, 1996

Proceedings of the Sixth Marcel Grossmann Meeting on General Relativity
(Kyoto, Japan, 1991)
Edited by H. Sato and T. Nakamura
World Scientific, 1992

Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity
(Perth, Australia, 1988)
Edited by D.G. Blair and M.J. Buckingham
World Scientific, 1989

Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 1985)
Edited by R. Ruffini
World Scientific, 1986

Proceedings of the Third Marcel Grossmann Meeting on General Relativity
(Shanghai, People's Republic of China, 1982)
Edited by Hu Ning
Science Press – Beijing and North-Holland Publishing Company, 1983

Proceedings of the Second Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1979)
Edited by R. Ruffini
North-Holland Publishing Company, 1982

Proceedings of the First Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1975)
Edited by R. Ruffini
North-Holland Publishing Company, 1977

SPONSORS

International Center for Relativistic Astrophysics Network (ICRANet)
International Center for Relativistic Astrophysics (ICRA)

FREEDOM OF MOVEMENT FOR SCIENTISTS

The Marcel Grossmann Meetings were founded with the premise that scientists of all nations have a right to meet to exchange knowledge independent of national borders.

ACKNOWLEDGEMENTS

We acknowledge the outstanding job done before, during and after the meeting by the ICRANet/ICRA administrative and secretarial staff: Cristina Adamo, Silvia Latorre, Elisabetta Natale, and Cinzia di Niccolo. Finally this meeting and its proceedings could not have functioned without the dedicated IT support of the ICRANet system manager Gabriele Brandolini, with some temporary assistance from Domenico La Selva and Damiano Verzulli. We would like to thank Linda Kwan from World Scientific for valuable assistance during preparation of these proceedings.

ORGANIZING BODIES OF THE SIXTEENTH MARCEL GROSSMANN MEETING

INTERNATIONAL ORGANIZING COMMITTEE

Blair David, Choquet Bruhat Yvonne, Damour Thibault, De Bernardis Paolo, Everitt C. W. Francis, Fryer Chris, Haensch Theodor, Henneaux Marc, Jones Christine, Kerr Roy, Kleinert Hagen, Kunz Jutta, Laemmerzahl Claus, Longair Malcolm, Mirabel Felix, Mirzoyan Razmik, Piran Tsvi, Rueda Jorge, Ruffini Remo (chair), Sasaki Misao, Sato Humitaka, Sunyaev Rashid, 't Hooft Gerard, Weinberg Steven, Yau Shing-Tung, Zhang Bing

LOCAL ORGANIZING COMMITTEE

Adamo Cristina, Bianco Carlo Luciano, Brandolini Gabriele A., di Niccolo Cinzia, Latorre Silvia, La Selva Domenico, Li Liang, Loppini Alessandro, Natale Elisabetta, Verzulli Damiano, Vereshchagin Gregory (chair), Wang Yu

INTERNATIONAL COORDINATING COMMITTEE

ALBANIA: Hafizi M. - ARGENTINA: Arguelles C., Scoccola C., Reula O., Romero G.E. - ARMENIA: Sahakyan N. - AUSTRALIA: Blair D., Ju L., Lun A., Manchester D., Melatos A., Quinn P., Scott S.M., Steele J.D. - AUSTRIA: Aichelburg P.C., Schindler S. - BELARUS: Kilin S., Prakapenia M., Siutsou I. - BELGIUM: Henneaux M. - BOLIVIA: Aguirre C.B. - BOSNIA: Pasic V. - BRAZIL: Barres de Almeida U., Coelho Goulart J., Dalmolin F.T., de Lima Rafael C.R., Guzzo M., Maia C., Malheiro M., Romero Filho C.A., Shellard R.C., Zen Vasconcellos C. - BULGARIA: Yazadjiev S. - CANADA: Singh D., Smolin L., Turok N. - CHILE: Bauer F., Bunster W.C., Giacomini A. - CHINA (MAINLAND): Cai R., Cai Y., Cao Z., Chang J., Chen J., Chen X., Dai Z., Feng L.-L., Han W., Jing Y., Li T.-P., Lin W., Lou Y.-Q., Luo J., Mei J., Tam T., Wang A., Wang Y., Wu X.-P., Wu Y.-L., Yuan Y.-F., Zhang B.-B., Zhang S.-N., Zhao G. - CHINA (TAIWAN): Chen Chiang-Mei, Chen Pisin, Lee Da-Shin, Lee Wo-Lung, Ni Wei-Tou - COLOMBIA: Bargueño de Retes P., Gonzalez G., Higuera Garzon M.A., Núñez L., Romano A.E., Valenzuela Toledo C.A., Zuluaga J.I. - CROATIA: Dominis Prester D., Karlica M., Milekovic M., Smolcic V., Smolic I., Suric T. - CUBA: Perez Martinez A., Pérez Rojas H. - CZECH REPUBLIC: Bicak J., Stuchlik Z. - DENMARK: Naselsky P. - ECUADOR: Contreras E. - EGYPT: Tawfik A.N., Wanas M.I. - ESTONIA: Einasto J., Saar E. - FINLAND: Poutanen J., Volovik G. - FRANCE: Brillet A., Buchert T., Chardonnet P., Coullet P., de Freitas Pacheco J.A., Deruelle N., Iliopoulos J., Lamanna G., Mignard F. - GEORGIA: Lavrelashvili George, Machabeli Giorgi - GERMANY: Biermann P., Blumlein J., Di Piazza A., Fritzsch H., Genzel R., Gilmozzi R., Hasinger G., Hehl F., Keitel C., Kiefer C., Mirzoyan R.,

Neugebauer G., Nicolai H., Renn J., Ringwald A., Ruediger A. - GREECE: Batakis N.A., Cotsakis S., Vagenas E.C. - HUNGARY: Fodor G., Levai P. - ICELAND: Bjornsson G., Jakobsson P. - INDIA: Chakrabarti S.K., Iyer B., Padmanabhan T., Souradeep T. - IRAN: Baghram S., Bavarsad E., Eslam Panah B., Firouzjahi H., Haghighat M., Mansouri R., Mashhoon B., Shakeri S., Sobouti Y., Taghi Mirtorabi M. - IRELAND: O'Murchada N. - ISRAEL: Milgrom M., Nakar E., Pe'er A., Piran T. - ITALY: Belinski V., Bianchi M., Bianco C.L., Cherubini C., Della Valle M., Falciano S., Filippi S., Haardt F., Menotti P., Merafina M., Pani P., Ricci F., Treves A., Vereshchagin G.V., Vitale S., Xue S.- S. - JAPAN: Fujimoto M.-K., Makishima K., Nakamura T., Sato K., Shibata M. - KAZAKHSTAN: Abishev M., Aimuratov Y., Boshkayev K., Mychelkin E.G., Spitaleri C. - KOREA (PYEONGYANG): Kim J.S. - KOREA (SEOUL): Kim S.P., Kim S.-W., Lee H.K., Lee H.-W., van Putten M. - KYRGYZSTAN: Gurovich V.Ts. - LIBYA: Gadri M. - MEXICO: Breton N., Cervantes-Cota J.L., Fraija Cabrera N.I., García-Diaz A.A., Macías Alvarez A., Mielke Eckehard W., Quevedo H., Rodriguez L.F. - NETHERLANDS: Slagter R. - NEW ZEALAND: Visser M., Wiltshire D. - NORWAY: Elgaroy O., Fonseca Mota D., Knutsen H. - PAKISTAN: Qadir A., Qamar S. - PERU: Vargas T. - POLAND: Belczynski K., Demianski M., Lewandowski Jerzy, Nurowski P., Sokolowski L. - PORTUGAL: Costa M., Da Silva A., Lemos J.P.S., Lobo F., Moniz P., Silva L.O. - ROMANIA: Visinescu M. - RUSSIA: Aksenov A., Arkhangelskaja I., Bisnovatyi-Kogan G., Blinnikov S., Chechetikin V.M., Cherepaschuk A.M., Khriplovich I., Lipunov V.M., Lukash V.N., Novikov I., Rudenko V.N., Starobinsky A.A. - SERBIA: Djordjevic G., Jovanovic P., Knezevic Z., Pankov-Hzvojevic M., Popovic L., Prodanovic T., Sijacki D., Simic S. - SLOVAKIA: Balek V. - SLOVENIA: Cadez A., Gomboc A., Zavrtanik D. - SOUTH AFRICA: Larena J., Maharaj S. - SPAIN: Elizalde E., Ibanez J., Perez M.J., Verdaguer E. - SWEDEN: Abramowicz M.A., Marklund M., Ryde F. - SWITZERLAND: Durrer R., Jetzer P. - TURKEY: Aliev A., Gurses M. - UKRAINE: Novosyadlyj B., Zaslavskii O., Zhuk A. - UNITED ARAB EMIRATES: Fernini I. - UNITED KINGDOM: Cruise A.M., Frenk Carlos S., Green M., Mavromatos N., Perry M., Willingale R. - USA: Abel T., Ashtekar A., Bardeen J., Carlstrom J., Cornish N., Dermer C., Fan X., Flanagan E., Fraschetti F., Fryer C., Incera V., Jantzen R.T. (chairperson), Kolb R., Laguna P., Lousto C., Madau Piero, Mathews Grant, Matzner Richard, Melia Fulvio, Mester John, Michelson Peter, Nordtvedt Kenneth, Parker Leonard, Pretorius F., Pullin J., Shapiro I., Shapiro S., Shoemaker D., Smoot G., Stiavelli M., Teukolsky S., van Nieuwenhuizen P., Zhang B. - UZBEKISTAN: Ahmedov B., Zalaletdinov R.M. - VATICAN CITY: Gionti G. - VENEZUELA: Fuenmayor E. - VIETNAM: Long H.N.

MARCEL GROSSMANN AWARDS

Sixteenth Marcel Grossmann Meeting

Institutional Awards

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

Individual Awards

DEMETRIOS CHRISTODOULOU
"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

GERARD 't HOOFT
"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

TSVI PIRAN
"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".

STEVEN WEINBERG
"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Each recipient is presented with a silver casting of the TEST sculpture by the artist A. Pierelli. The original casting was presented to His Holiness Pope John Paul II on the first occasion of the Marcel Grossmann Awards.

15th Marcel Grossmann Meeting
July 2018, Rome, Italy

Institutional Awards

PLANCK SCIENTIFIC COLLABORATION (ESA)
"for obtaining important constraints on the models of inflationary stage of the Universe and level of primordial non-Gaussianity; measuring with unprecedented sensitivity gravitational lensing of Cosmic Microwave Background fluctuations by large-scale structure of the Universe and corresponding B-polarization of CMB, the imprint on the CMB of hot gas in galaxy clusters; getting unique information about the time of reionization of our Universe and distribution and properties of the dust and magnetic fields in our Galaxy"

- presented to Jean-Loup Puget, the Principal Investigator of the High Frequency Instrument (HFI)

HANSEN EXPERIMENTAL PHYSICS LABORATORY AT STANFORD UNIVERSITY
"to HEPL for having developed interdepartmental activities at Stanford University at the frontier of fundamental physics, astrophysics and technology"

- presented to Research Professor Leo Hollberg, HEPL Assistant Director

Individual Awards

LYMAN PAGE
"for his collaboration with David Wilkinson in realizing the NASA Explorer WMAP mission and as founding director of the Atacama Cosmology Telescope"

RASHID ALIEVICH SUNYAEV
"for the development of theoretical tools in the scrutinising, through the CMB, of the first observable electromagnetic appearance of our Universe"

SHING-TUNG YAU
"for the proof of the positivity of total mass in the theory of general relativity and perfecting as well the concept of quasi-local mass, for his proof of the Calabi conjecture, for his continuous inspiring role in the study of black holes physics"

14th Marcel Grossmann Meeting
July 2015, Rome, Italy

Institutional Award

EUROPEAN SPACE AGENCY (ESA)
"for the tremendous success of its scientific space missions in astronomy, astrophysics, cosmology and fundamental physics which have revolutionized our knowledge of the Universe and hugely benefited science and mankind"

- presented to its Director General Johann-Dietrich Woerner

Individual Awards

KEN'ICHI NOMOTO
"for heralding the role of binary systems in the evolution of massive stars"

MARTIN REES
"for fostering Research in black holes, gravitational waves and cosmology"

YAKOV G. SINAI
"for applying the mathematics of chaotic systems to physics and cosmology"

SACHIKO TSURUTA
"for pioneering the physics of hot neutron stars and their cooling"

FRANK C.N. YANG
"for deepening Einstein's geometrical approach to physics in the best tradition of Paul Dirac and Hermann Weyl"

T.D. LEE (award received by Yu-Qing Lou on behalf of Prof. T.D. Lee)
"for his work on white dwarfs motivating Enrico Fermi's return to astrophysics and guiding the basic understanding of neutron star matter and fields"

13th Marcel Grossmann Meeting
July 2012, Stockholm, Sweden

Institutional Award

ALBANOVA
for its innovative status as a joint institute established by Stockholm University and the Royal Institute of Technology and for fostering contributions to cosmology and astrophysics in the profound scientific tradition established by Oskar Klein.

- presented to the Rector of Stockholm University, Prof. Kåre Bremer.

Individual Awards

DAVID ARNETT
for exploring the nuclear physics and yet unsolved problems of the endpoint of thermonuclear evolution of stars, leading to new avenues of research in physics and astrophysics.

VLADIMIR BELINSKI and I.M. KHALATNIKOV
for the discovery of a general solution of the Einstein equations with a cosmological singularity of an oscillatory chaotic character known as the BKL singularity.

FILIPPO FRONTERA
for guiding the Gamma-ray Burst Monitor Project on board the BeppoSAX satellite, which led to the discovery of GRB X-ray afterglows, and to their optical identification.

12th Marcel Grossmann Meeting
July 2009, Paris, France

Institutional Award

INSTITUT DES HAUTES ÉSTUDES SCIENTIFIQUE (IHÉS)
for its outstanding contributions to mathematics and theoretical physics, and notably for having renewed basic geometrical concepts, and having developed new mathematical and physical aspects of spacetime.

- presented to Prof. Jean-Pierre Bourguignon

Individual Awards

JAAN EINASTO
for pioneering contributions in the discovery of dark matter and cosmic web and fostering research in the historical Tartu Observatory.

CHRISTINE JONES
for her fundamental contributions to the X-ray studies of galaxies and clusters tracing their formation and evolution and for her role in collaborations using clusters to study dark matter and in analyzing the effects of outbursts from supermassive black holes on the intracluster gas.

MICHAEL KRAMER
for his fundamental contributions to pulsar astrophysics, and notably for having first confirmed the existence of spin-orbit precession in binary pulsars.

11th Marcel Grossmann Meeting
July 2006, Berlin, Germany

Institutional Award

FREIE UNIVERSITÄT BERLIN
for the successful endeavor of re-establishing — in the spirit of the Humboldt tradition — freedom of thinking and teaching within a democratic society in a rapidly evolving cosmos

- presented to Dr. Dieter Lenzen, President of FUB

Individual Awards

ROY KERR
for his fundamental contribution to Einstein's theory of general relativity: "The gravitational field of a spinning mass as an example of algebraically special metrics."

GEORGE COYNE
for his committed support for the international development of relativistic astrophysics and for his dedication to fostering an enlightened relationship between science and religion.

JOACHIM TRUMPER
for his outstanding scientific contributions to the physics of compact astrophysical objects and for leading the highly successful ROSAT mission which discovered more than 200,000 galactic and extragalactic X-ray sources: a major step in the observational capabilities of X-ray astronomy and in the knowledge of our universe.

10th Marcel Grossmann Meeting
July 2003, Rio de Janeiro, Brazil

Institutional Award

CBPF (Brazilian Center for Research in Physics)
for its role as a teaching and research institution and as a place originating fundamental physics ideas in the exploration of the universe.

- presented to its founders Cesar Lattes, Josè Leite Lopez and Jayme Tiomno

Individual Awards

YVONNE CHOQUET-BRUHAT AND JAMES W. YORK, JR.
for separate as well as joint work in establishing the mathematical framework for proving the existence and uniqueness of solutions to Einstein's gravitational field equations.

YUVAL NE'EMAN
for his contributions to science, epistimology, mathematics and physics from subnuclear to space sciences.

9th Marcel Grossmann Meeting
July 2000, Rome, Italy

Institutional Award

SOLVAY INSTITUTES
for identifying and recording in discussions by the protagonists the crucial developments of physics and astrophysics in the twentieth century.

- presented to Jacques Solvay

Individual Awards

CECILLE AND BRYCE DEWITT
for promoting General Relativity and Mathematics research and inventing the "summer school" concept.

RICCARDO GIACCONI
for opening, five successive times, new highways for exploring the Universe.

ROGER PENROSE
for extending the mathematical and geometrical foundations of General Relativity.

8th Marcel Grossmann Meeting
June 1997, Jerusalem

Institutional Award

HEBREW UNIVERSITY
for its role as a cradle of Science and Humanities and for hosting the manuscripts of Albert Einstein.

- presented to M. Magidor, President of the Hebrew University of Jerusalem

Individual Awards

TULLIO REGGE
for his contributions to the interface between mathematics and physics leading to new fields of research of paramount importance in relativistic astrophysics and particle physics.

FRANCIS EVERITT
for leading the development of extremely precise space experiments utilizing superconducting technology to test General Relativity and the Equivalence Principle.

7th Marcel Grossmann Meeting
June 1994, Stanford, USA

Institutional Award

SPACE TELESCOPE SCIENCE INSTITUTE
for its critical role in the direction and operation of the Hubble Space Telescope, a truly unique international laboratory for the investigation and testing of general relativity in the context of modern astrophysics and cosmology.

- presented to Peter Stockman

Individual Awards

SUBRAHMANYAN CHANDRASEKHAR
for his contributions to the analysis of gravitational phenomena from Newton to Einstein and especially for leading the way to relativistic astrophysics with the concept of critical mass for gravitational collapse.

JIM WILSON
for having built on his experience in nuclear physics, thermonuclear reactions, and extensive numerical simulation to create a new testing ground for the novel concepts of relativistic astrophysics.

6th Marcel Grossmann Meeting
June 1991, Kyoto, Japan

Institutional Award

RITP
for keeping alive first in Hiroshima and them in Kyoto research in relativity, cosmology, and relativistic field theory and the development of a school of international acclaim.

- presented to Professor K. Tomita

Individual Awards

MINORU ODA
for participating in the pioneering work of the early sixties in X-ray astronomy and for his subsequent molding of an agile and diversified Japanese scientific space program investigating the deepest aspects of relativistic astrophysics.

STEPHEN HAWKING
for his contributions to the understanding of spacetime singularities and of the large scale structure of the Universe and of its quantum origins.

5th Marcel Grossmann Meeting
August 1988, Perth, Australia

Institutional Award

THE UNIVERSITY OF WESTERN AUSTRALIA
for its contributions to relativistic astrophysics.

- presented to the Vice Chancellor, Professor Robert Smith

Individual Awards

SATIO HAYAKAWA
for his contributions to research in gamma, X-ray and infrared radiation as well as cosmic rays.

JOHN ARCHIBALD WHEELER
for his contributions to geometrodynamics and Einstein's visions.

4th Marcel Grossmann Meeting
July 1985, Rome, Italy

Institutional Award

THE VATICAN OBSERVATORY
for its contributions to the origin and development of astrophysics.

- presented to His Holiness Pope John Paul II

Individual Awards

WILLIAM FAIRBANK
for his work in gravitation and low temperature physics.

ABDUS SALAM
for his work in unifying fundamental interactions.

Institutional Awards for the Spektrum-Roentgen-Gamma (SRG) mission

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

On Tuesday June 29, 2021, the following 31 astro-ph appeared:

1. https://arxiv.org/abs/2106.14517
2. https://arxiv.org/abs/2106.14518
3. https://arxiv.org/abs/2106.14519
4. https://arxiv.org/abs/2106.14520
5. https://arxiv.org/abs/2106.14521
6. https://arxiv.org/abs/2106.14522
7. https://arxiv.org/abs/2106.14523
8. https://arxiv.org/abs/2106.14524
9. https://arxiv.org/abs/2106.14525
10. https://arxiv.org/abs/2106.14526
11. https://arxiv.org/abs/2106.14527
12. https://arxiv.org/abs/2106.14528
13. https://arxiv.org/abs/2106.14529
14. https://arxiv.org/abs/2106.14530
15. https://arxiv.org/abs/2106.14531
16. https://arxiv.org/abs/2106.14532
17. https://arxiv.org/abs/2106.14533
18. https://arxiv.org/abs/2106.14534
19. https://arxiv.org/abs/2106.14535
20. https://arxiv.org/abs/2106.14536
21. https://arxiv.org/abs/2106.14537
22. https://arxiv.org/abs/2106.14541
23. https://arxiv.org/abs/2106.14542
24. https://arxiv.org/abs/2106.14543
25. https://arxiv.org/abs/2106.14544
26. https://arxiv.org/abs/2106.14545
27. https://arxiv.org/abs/2106.14546
28. https://arxiv.org/abs/2106.14547
29. https://arxiv.org/abs/2106.14548
30. https://arxiv.org/abs/2106.14549
31. https://arxiv.org/abs/2106.14550

S.A. LAVOCHKIN ASSOCIATION
presented to its Designer General **Alexander Shirshakov**

Dr Alexander Shirshakov

S.A. Lavochkin Association created the Navigator space platform carrying German eRosita and Russian ART-XC X-Ray Telescopes, organized the launch of SRG Orbital X-Ray Observatory to the second Lagrangian point of the Sun-Earth system at a distance of 1.5 million km from the Earth and managed the observatory flight and the daily reception of its scientific data on Earth for 23.5 months.

Dr Alexander Shirshakov, Designer General of the S.A. Lavochkin Association, is specialized in design, manufacture, testing, launch and control of S/C for scientific purposes. Among those S/C launched, there are the «Radiostron» Astrophysical Observatory (2011) and the «Spektr-RG» space observatory (2019), while the planned S/C launches are «Luna-25» and «Exomars».

Dr Shirshakov started his career in 1973, working as an engineer of the State Unitary Enterprise «NPO named by S.A. Lavochkin» in Khimki (Russian Federation). Starting from 1989 he has played multiple roles within the Lavochkin Association, been appointed head of the group, head of the sector, head of department, deputy head of the complex, head of the branch, director of the center, deputy head of the Design Bureau, deputy General Designer and deputy General Director.

Dr Shirshakov is an editorial board Member of the reviewed edition of «Vestnik of Lavochkin Association». Since 2017, he is also member of the General Designer council. He has been awarded Honored Mechanical engineer of the Russian Federation as well as Agency-level award of the Russian Federal Space Agency.

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

Professor Peter Predehl

eROSITA is the soft X-ray telescope on-board the Russian-German Spektr-RG mission which was successfully launched from Baikonur on July 13, 2019 and placed in a halo orbit around the L2 point. 30 years after ROSAT, eROSITA performs an all-sky survey with an unprecedented sensitivity, spectral and angular resolution. Clusters of galaxies are the largest collapsed objects in the Universe. Their formation and evolution is dominated by gravity, i.e. Dark Matter, while their large scale distribution and number density depends on the geometry of the Universe, i.e. Dark Energy. X-ray observations of clusters of galaxies provide information on the rate of expansion of the Universe, the fraction of mass in visible matter, and the amplitude of primordial fluctuations which are the origin of clusters of galaxies and the whole structure of the universe. eROSITA has been designed to detect at least 100.000 clusters of galaxies and to detect systematically more than 3 million obscured accreting Black Holes. eROSITA will also allow to study the physics of galactic X-ray source

populations, like pre-main sequence stars, supernova remnants and X-ray binaries. The eROSITA telescope consists of seven identical Wolter-1 mirror modules. A novel detector system has been developed by MPE on the basis of the successful XMM-Newton pn-CCD technology. MPE is the scientific lead institute of eROSITA, responsible for the development of the instrument, the operation, the analysis software and data archive. Peter Predehl led this development as Principal Investigator of eROSITA and German lead scientist of the SRG mission for more than 15 years until the completion of the first of eight surveys in 2020. At this time eROSITA has already discovered more than 1 million X-ray sources, more than all X-ray observatories of the last 50 years together. This demonstrates that the design goals of the mission will easily be fulfilled.

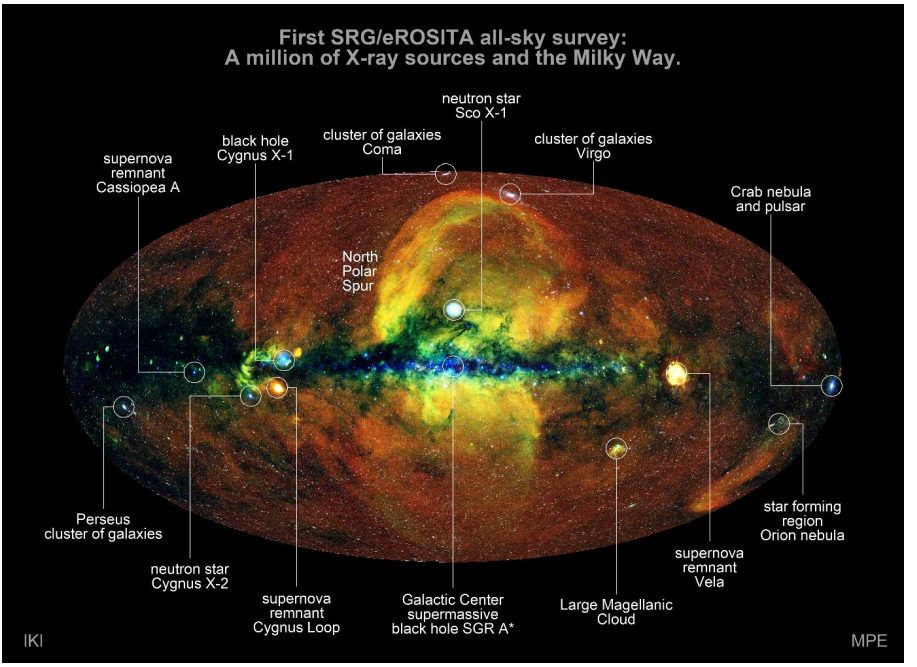
SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
presented to Professor **Rashid Sunyaev**

Professor Rashid Sunyaev

Space Research Institute (IKI) of the Russian Academy of Sciences was responsible for developing the overall concept and scientific program of the SRG Orbital observatory and played a leading role in developing the ART-XC telescope and the entire SRG observatory as part of the Russian space science program carried out by Roskosmos Corporation in the interests of the Russian Academy of Sciences.

During the flight to the L2 point of the Sun-Earth system, SRG with German (eRosita) and Russian (ART-XC named after Mikhail Pavlinsky) X-ray Telescopes aboard performed calibrations and long duration Performance Verification observations of a dozen of targets and deep fields. Starting in the middle of December 2019, the SRG scanned the whole sky three times. During these scans, SRG discovered two million point X-ray sources: mainly quasars, stars with hot and bright coronae, and more than 30 thousand clusters of galaxies. There is a competition and synergy in the search for clusters of galaxies between SRG and the ground-based Atacama Cosmology and South Pole Telescopes, which are searching for clusters of galaxies in microwave spectral band using Sunyaev-Zeldovich effect. SRG provided the X-Ray map of the whole sky in hard and soft bands, the last is now the best among existing. The huge samples of the X-ray selected quasars at the redshifts up to z = 6.2 and clusters of galaxies will be used for well-known cosmological tests and detailed study of the growth of the large scale structure of the Universe during and after reionization. SRG/eRosita is discovering every day several extragalactic objects which increased or decreased their brightness more than 10 times during half of the year after the previous scan of the same one-degree wide strip on the sky. A significant part of these objects has observational properties similar to the Events of Tidal Disruption of a star orbiting in the vicinity of the supermassive black hole. ART-XC discovered a lot of bright galactic and extragalactic transients.

Rashid Sunyaev is the Principal Investigator of SRG mission in Russia, director-emeritus of the Max-Planck Institute for Astrophysics and Maureen and John Hendricks distinguished visiting professor of the Institute for Advanced Study, Princeton.



Individual Awards

Professor **DEMETRIOS CHRISTODOULOU**

"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

Professor Demetrios Christodoulou

It was back in 1967 that Achille Papapetrou mentioned the case of the 16-year-old Demetrios Christodoulou to John Archibad Wheeler. Wheeler interviewed Demetrios in Paris and brought him immediately to Princeton where he was registered as an undergraduate at the university. After one year he entered the graduate school and started collaborating with me. At the time I was working with Wheeler on the effective potential approach to geodesics co-rotating and counter-rotating (see e.g. reference in The Classical Theory of Fields (Landau and Lifshitz, 1980) in the Kerr metric (later renamed as ISCO; see e.g. (Gravitation Misner, Thorne, Wheeler. 1973). In parallel, Frank Zerilli was working on the gravitational radiation emitted by the fall of a test particle in a Schwarzschild black hole (Zerilli 1970). From these limited conceptual arena Charles Misner and later Kip Thorne launched a program for the detection of gravitational waves on the Earth; see e.g. Misner 1974, Abbott et al. 2016, Abbott et al. 2017. See however Davis et al. 1972, Rodriguez et al. 2018 and J.A. Rueda et al. 2018.

A new approach started with the arrival of Demetrios: he was just creating mathematics following his needs. We identified the reversible and irreversible transformations of a Kerr black hole. Wheeler advanced a thermodynamic analogy. I addressed the need of identifying the concept of irreducible mass (from the Italian "irriducibile"), and was Demetrios's contribution to integrate, overnight, the differential equation for infinitesimal reversible transformations which led to the finite mass-energy formula of a Kerr black hole. That evening, while walking back home through IAS woods, I expressed to Wheeler the great relevance of the newly found formula by Demetrios and proposed to let Demetrios be the single author of this article, admiring his great mathematical talent. Wheeler agreed. The Editor of PRL objected since in that two pages article the Fig. 2 by Wheeler and myself was still unpublished. Actually that Fig. 2 followed a discussion I previously had with Penrose in Florence (Penrose 1961) which allowed us to present there, for the first time, a "Penrose Process". Some difficulties in achieving this process were obvious from the example in Fig. 2, which Roger later recognized himself (Penrose & Floyd 1971). The Editor finally agreed on our written request and the paper appeared on September 17, 1970 (Christodoulou, 1970). On January 1971 appeared my article with Johnny introducing the Black Hole (Ruffini & Wheeler, 1971), with the new physics we were developing in Princeton, including the concept of the "ergosphere". On march 1 1971 we submitted the mass formula of the Kerr Newmann metric, including the relation between the surface area of the horizon and the irreducible (Christodoulou & Ruffini, 1971). On March 11, 1971 the same results were independently confirmed by Steven Hawking, extending further the applicability of our equation (Hawking 1971).

The thesis was successfully discussed by a committee including Eugene Wigner (see Fig. 1), one of the closest collaborators of Albert Einstein and David Wilkinson (see Fig. 2), the head of the NASA WMAP mission, and Johnny and myself as supervisors. The new message was clear: Black Holes, far from being a sink of energy, were energy sources emitting "in principle" 50% of their mass energy, being extractable (Christodoulou & Ruffini, 1971).

Fig. 1 and Fig. 2: Demetrios during his thesis presentation with Eugene Wigner (Fig. 1) and David Wilkinson (Fig.2). Johnny and I were supervisors, ready to intervene in case of need, but no need of intervention was necessary! Wigner elaborated the aphorism of Niels Bohr "Interesting = wrong" in the most definite "very interesting if true = totally wrong".

Demetrios turned soon to the study of partial differential equations and mathematical physics, to which he dedicated for the rest of his career and results were published in four monographs: (Christodoulou and Klainerman 1994, Christodoulou 2007, Christodoulou 2009, Christodoulou 2019). In 1968, Johnny proposed to Demetrios the collapse of a "geon" composed of massless scalar field as a second topic for his thesis. It took almost forty years for him to solve this problem, extended by Demetrios to the focusing of gravitational waves leading to black hole formation (Christodoulou 2009).

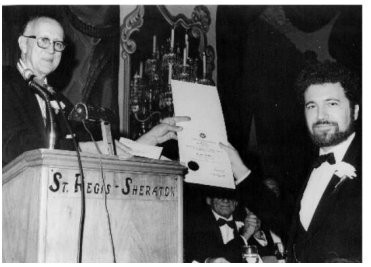


Fig. 3: Prof. Remo Ruffini receiving the Cressy Morrison Award of the New York Academy of Sciences, 1972 for the discovery of the first Black Hole in our galaxy Cygnus X1.

A "long march" started on 12 December 1970 with the launch of the Uhuru satellite by Riccardo Giacconi. Early in 1971 an almost daily conversation with him and Herb Gursky at the Smithsonian Astrophysical Observatory, leading to the discovery of binary X-ray sources. This was soon followedby the announcement of Cygnus X1 identified as the first black hole in our galaxy (Ruffini 1973); see e.g. Gursky & Ruffini 1975, which contained as well the first publicannouncement of the Discovery of Gamma Ray burst, as well as Giacconi & Ruffini 1980, 2009; see Figs. 3 and 4).

Fig. 4: In the second row, from left to right, there are, among others: E. T. Newman, S. Chandrasekhar (Nobel 1983), R. Giacconi (Nobel 2002), R. Ruffini, A. Treves, A. Hewish (Nobel 1974), D. Arnett, J.H. Taylor (Nobel 1993), J. Wilson, R. Penrose (Nobel 2020), as well as J. Bahcall, T. Damour, T. Piran et al.

Today, after fifty years, this "long march" has reached a definite result: through the grandest observational multi-wavelength effort in the history of mankind, from space, ground and underground observatories, we are finally finding evidence that black holes are "alive" and their "extractable energy" in our mass formula (Christodoulou & Ruffini, 1971), is the energy source of the most energetic cosmological sources: gamma ray bursts (GRBs),the active galactic nuclei (AGNs) as well as the ultra-high energy cosmic rays (UHECRs) (Ruffini et al. 2021 and references therein). Their "inner engine", has three independent components: 1) a Kerr black hole which is neither in a stationary state nor in vacuum, 2) a background magnetic field aligned with the black hole rotation axis, and 3) an extremely diluted fully ionized plasma (Moradi et al. 2021).There is no role in this inner engine for ISCO. Indeed a new electro dynamical field equations describe the synchrotron radiation emitted close to the black hole horizon, they point to a discrete and repetitive emission of "blackholic quanta" in the MeV and in the GeV. The magnitudes and the emission time scales of these quanta, for M87 and GRB 130427A, are expressed as a function of the above three parameters (Rueda & Ruffini, 2021). A long lasting GeV emission with a luminosity decreasing as a temporal power law, allows for the first time in GRBs, the determination of the black hole mass and spin as well as their time evolution perfectly fulfilling our mass energy formula (Christodoulou & Ruffini, 1971): a long lasting emission process profoundly different from the traditional process of continued gravitational contraction.

Remo Ruffini

Professor **GERARD 't HOOFT**

"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

Professor Gerard 't Hooft

Prof. Gerard 't Hooft has been a full Professor at the Utrecht University (the Netherlands), since 1977. Nowadays, he is an Emeritus Professor at that University. During his career, he has paid extended scientific visits to CERN (Geneva), Harvard, Stanford, Princeton and Duke University, NC. In 1999, together with M. Veltman, he received the Nobel Prize in Physics, awarded by The Royal Swedish Academy of Sciences, *"For elucidating the quantum structure of electroweak interactions in physics"*.

Prof. 't Hooft's main subjects of research includes:

– Gauge Theories for the sub-atomic particles and forces, various aspects and ingredients of what is now called "The Standard Model of the sub-atomic particles: renormalizability, topological features such as magnetic monopoles and instantons, 1/N expansions.

– Theories for the quantization of the gravitational force and black holes: producing models for the quantum properties of a black hole, as derived from Standard Model and General Relativity alone; its topological features such as antipodal identification.

– Fundamental theories underlying quantum mechanics, in particular returning determinism and reality to the dynamics of the tiniest material entities in his universe.

Prof. 't Hooft has been awarded the Wolf Prize of the State of Israel (1982), the Pius XI Medal (Vatican City, 1983), the Lorentz Medal (KNAW Amsterdam, 1986) as well as the Spinoza Premium (Netherlands Organization for Scientific Research NWO, 1995).

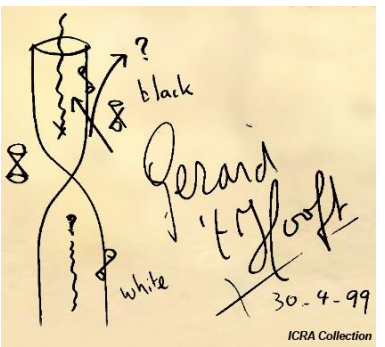


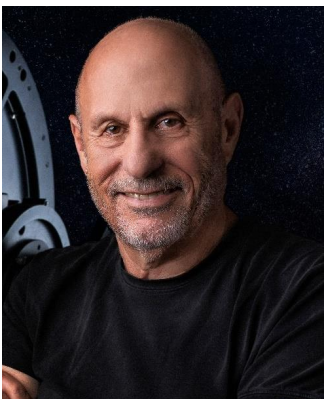
Fig. 2: The signature of Gerard 't Hooft on the wall of ICRA Room 301 (April 4, 1999).

A special event took place at ICRA on April 30, 1999. Prof. Ruffini invited Gerard 't Hooft to Rome to discuss a boundary condition for a quantum field on the black hole horizon, a topic Prof. Ruffini discussed in a previous article "Black-hole evaporation in the Klein-Sauter-Heisenberg-Euler formalism" with Thibault Damour (Phys. Rev. D 14, 332, 1976), but which needed to be examined in more detail. Prof. Ruffini planned to direct Gerard's attention to some specific aspects of this problem. Because we have traditionally been very attentive in spending ICRA travel funds, ICRA offered Gerard to come to Rome on a reduced fare weekend ticket arriving Friday and departing Monday. He had a great relaxing weekend together with Prof. Ruffini following his seminar, which among other things allowed Gerard to sign the wall in our ICRA Room (see Fig. 2), and during this splendid Rome spring weekend he also was able to find a missing factor of 2 in a formula in Prof. Ruffini's 1971 paper with Demetri Christodoulou on the black hole mass formula. The following October, Gerard received the Nobel prize, which meant that we could no longer get away with bringing him to Rome on a cheap ticket! Ever since Gerard has been in our MG IOC helping us with the preparation of the meetings. We are very happy to announce this MG16 Award to Gerard 't Hooft with the motivating phrase *"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon"*.

Remo Ruffini

Professor **TSVI PIRAN**

"for extending relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".



Professor Tsvi Piran

Tsvi Piran is the emeritus Schwartzmann professor at the Hebrew University of Jerusalem. He obtained his PhD in Physics, in 1976 from the Hebrew University working on the collisional Penrose process. Piran returned to the Hebrew University at 1981after being a post doc at Oxford and Texas and a long-term member at the IAS at Princeton. In 1982 he initiated and directed the first ever summer school on Gravitational Waves that took place at Les Houches. Piran was a visiting professor at Harvard, Columbia and New York and a Moore scholar at Caltech.

Piran's research deals with numerous aspects of relativistic astrophysics, ranging from the foundation of numerical relativity to modeling of observer relativistic phenomena and analytic work on the fate of gravitational collapse. Piran's research work focuses mostly on black holes and in particular on gamma-ray bursts. He was among the first to point out their cosmological origin and their association with merging neutron stars and heavy r-process nucleo synthesis. Piran's achievements were recognized in the 2019 EMET prize for Physics.

Professor **STEVEN WEINBERG**

"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Professor Steven Weinberg. Photo courtesy of Matt Valentine.

Steven Weinberg is a member of the Physics and Astronomy Departments at The University of Texas at Austin. His research has covered a broad range of topics in quantum field theory, elementary particle physics and cosmology. He has been honored with numerous awards, including the Nobel Prize in Physics, the National Medal of Science, the Heinemann Prize in Mathematical Physics and in 2020, the Breakthrough Prize. He is a member of the US National Academy of Sciences, Britain's Royal Society, and other academies in the USA and abroad. The American Philosophical Society awarded him the Benjamin Franklin Medal, with a citation that said he is "considered by many to be the preeminent theoretical physicist alive in the world today." His books for physicists include *Gravitation and Cosmology*, the three-volume work *The Quantum Theory of Fields*, *Cosmology* and published in April of 2021, *Foundations of Modern Physics*. Educated at Cornell, Copenhagen, and Princeton, he also holds honorary degrees from sixteen other universities. He taught at Columbia, Berkeley, M.I.T., and Harvard, where he was Higgins Professor of Physics, before coming to Texas in 1982.

Fig. 1: Chuo Pei Yuan and Cheng Ning Yang at MG2 in Trieste, Italy (1979).

The Sixteenth Marcel Grossmann Meeting (MG16) is a very special one in many respects: it will take place during a pandemic and in spite of the many difficulties, we have decided not to postpone it but to organize it as a virtual meeting. As described on the MG series webpage, these meetings started in 1975 with the first meeting at the International Centre for Theoretical Physics (ICTP) in Trieste (Italy) that I organized with Nobel Prize winner Abdus Salam. A second meeting followed in 1979, with a significantly larger participation including Nobel Laurate Cheng Ning Yang and a Chinese delegation led by Chuo Pei Yuan (see Fig. 1), including Fang Li-Zhi who had accompanied me during my entire first visit to China in 1979. The first truly international MG meeting followed in 1982 in Shanghai (China):this represented an especially important step forward both for the meeting and for China. A multi-millennia *"motto"* in China, which was then proclaimed on banners everywhere, read *"Friends from all over the world are welcomed"*.

We were soon at an impasse over the participation of scientists from Israel, since no diplomatic relations existed between China and Israel at that time and the Israeli scientists were not to be allowed to attend the meeting. A long negotiation began. The boundary conditions were clearly set by Steven Weinberg, a member of the present MG16 IOC: no MG meetings on Einstein's theory of general relativity could occur without the participation of Israeli scientists. The intervention of Yuval Ne'emann, also a member of the MG IOC then as well as the Minister of Science of Israel (see Fig.2), proposed a compromise that would admit at least one Israeli scientist. I went to Beijing alone, meeting every morning for a week with 12 Chinese representatives led by Chuo Pei Yuan going over all possible options. I stayed in an isolated villa not far from Tiananmen Square, accompanied by the 3 volumes of Matteo Ricci (RI MA TO) to keep me company. No solution was in sight the entire week. At the last moment, just before my departure, an agreement was finally reached allowing two Israeli scientists into China. The historic compromise would admit Gerard Tauber and Tsvi Piran into China using a special ICRA travel document I had proposed for them to be able to participate in the meeting, accepted by the Chinese Ambassador in Rome. This modified the thousand-year Chinese *"motto"* to read *"Scientists from all over the world are welcomed"*. The event was extremely beneficial for China and signaled the truly international nature of the MG meetings.

Fig. 2: From right to left: Chaim Weizmann, President of Israel; Yuval Ne'emann, Minister of Science of Israel; R. Ruffini.

I kept on meeting Tauber in the years which followed (see Fig. 3). Soon after, Yuval Ne'emann visited China. The development of bilateral relations, including military cooperation and economical tights, grow exponentially until the establishment of normal diplomatic relations between Israel and China in 1992.

Fig. 3: From right to left: Arrigo Finzi, Remo Ruffini, Gerard Tauber and Konrad Bleuler.

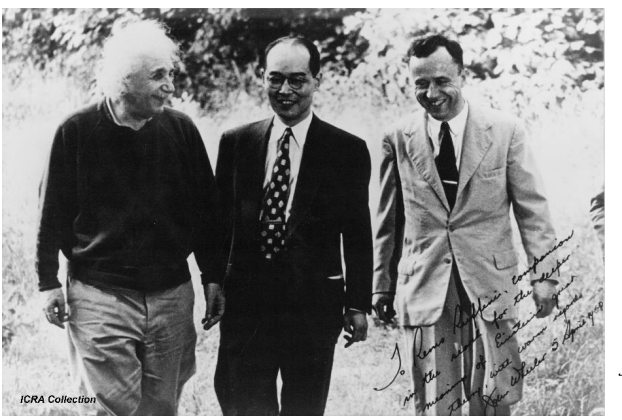


Fig. 4: Albert Einstein, Hideki Yukawa and John. A. Wheeler with a hand-written dedication to Remo Ruffini "To Remo Ruffini, companion in the search for the deeper meaning of Einstein great theory. With warm regards, John Wheeler 5 April 1968".

Given their key role played in the foundations of the MG meetings, I am very happy to propose on behalf of the MG16 IOC, two special Marcel Grossmann Individual Awards: one to Steven Weinberg for *"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory"* and another one to Tsvi Piran, *"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory"*, in the words of John A. Wheeler's photo dedication to me (see Fig. 4).

Remo Ruffini

PREFACE

Since 1975, the Marcel Grossmann Meetings on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories have been organized in order to provide opportunities for discussing recent advances in gravitation, general relativity and relativistic field theories, emphasizing mathematical foundations, physical predictions and experimental tests. The objective of these meetings is to elicit exchange among scientists that may deepen our understanding of spacetime structures as well as to review the status of ongoing experiments aimed at testing Einstein's theory of gravitation either from the ground or from space. Previous meetings have been held in Trieste (MG1: 1975) and (MG2: 1979), Shanghai (MG3: 1982), Rome (MG4: 1985, MG9: 2000), Perth (MG5: 1988), Kyoto (MG6: 1991), Stanford (MG7: 1994), Jerusalem (MG8: 1997), Rio (MG10: 2003), Berlin (MG11: 2006), Paris (MG12: 2009), Stockholm (MG13: 2012), MG14 in 2015 and MG15 in 2018 both in Rome.

Due to the COVID-19 pandemic spreading in the last two years the decision was taken to organize the Sixteenth Marcel Grossmann meeting for the first time in history entirely online. Despite numerous challenges, related to the organization of large worldwide event, MG16 showed the strongest ever interest from the scientific community with a record-breaking number of almost 1200 registered participants and of more than 1000 speakers.

The traditional six-day schedule has been modified to account for different time zones of the speakers and each day the program of the meeting was divided in three blocks with the reference to the Central European Summer Time. The first block was starting at 06:30 in the morning, allowing comfortable time for speakers from Asia and Oceania. The second block was held in the daytime in Europe and Africa. The third block was starting in the afternoon and ending at 19:30 allowing accommodation of the speakers from the Americas. Each day the blocks of plenary sessions were interchanging with the blocks of about 30 parallel sessions each, making this one of the most intense MG meetings ever. All this was possible thanks to recent developments in communication technologies. The Indico open-source software was selected as a web platform for this meeting, while Zoom platform was adopted for the video-conferencing. The meeting was streamed on ICRANet YouTube channel.

The meeting started on Monday July 5 with the Award ceremony. The individual awards went to Demetrios Christodoulou, Tsvi Piran, Gerard 't Hooft and Steven Weinberg, while the Institutional Awards went to the S.A. Lavochkin Association, to the Max Planck Institute for Extraterrestrial Physics – MPE and to the Space Research Institute IKI of the Russian Academy of Sciences. Overall there were 54 plenary talks, 4 public lectures and 5 roundtables and about 90 parallel sessions. The plenary session "Events in Relativistics Astrophysics" on Monday have seen the contributions from Rashid Sunyaev, Michael Kramer, James Miller-Jones, Felix Mirabel. The public lectures were delivered by Razmik Mirzoyan, Asghar Qadir

and Mohammad Bagheri. Plenary talks on Tuesday session "Black holes and the Quantum" by Juan Maldacena, Ahmed Almheiri, Gerard 't Hooft, Mihalis Dafermos, Sergiu Klainerman, Abhay Ashtekar and Frank Wilczek were bracketed by two roundtables on "New results from SRG/eRosita" with the participation of Andrea Merloni, Prof. Rashid Sunyaev, Alexander Lutovinov, Chandreyee Maitra, Esra Bulbul and "Solar neutrinos and Borexino" with the participation of Gianpaolo Bellini and Wick Haxton. Plenary talks on Wednesday in the session "Lambda CDM tensions" by George Efstathiou, Scolnic Daniel, Marc Kamionkowski, Wendy Freedman, Priya Natarajan and Licia Verde were followed by the roundtable "Precision cosmology" with the participation of Licia Verde, Marc Kamionkowski, Piero Rosati, and the public lecture by Francis Halzen. Two blocks of Thursday plenary sessions "Black holes in GRBs" and "Precision tests" included the talks by Roy Kerr, Yuan Ha, Lorenzo Amati, Elena Pian, Carlos Raúl Argüelles, Di Li, Jianglai Liu, Claus Lämmerzahl, Gerhard Heinzel and Ignazio Ciufolini and were followed by the roundtable "GRB 170817A and GRB 190829A" with the participation of Eleonora Troja, Liang Li, Rahim Moradi, Jorge Armando Rueda Hernandez. Two plenary blocks on Friday "Massive stars" and "Physics behind stellar collapse" included the talks by Selma de Mink, Norbert Langer, Jiri Bicak and Tomáš Ledvinka, Ivan De Mitri, Rahim Moradi and Giancarlo Cella. Finally, two plenary blocks on Saturday "Current and future missions" have seen the talks by Shuang-Nan Zhang, Weimin Yuan, Makoto Tashiro, Ruoyu Liu, Jean-Luc Atteia, Jim Hinton and Nicholas White and were followed by the roundtable "What is in our Galactic center" with the participation of Reinhard Genzel, Carlos Raúl Argüelles, Andreas Krut, Jorge Armando Rueda Hernandez, Eduar Becerra Vergara. The program of the meeting can be found at the official website http://www.icra.it/mg/mg16 and at ICRANet Indico website https://indico.icranet.org/event/1/.

These proceedings include about 400 papers containing the results presented at the Sixteenth Marcel Grossmann meeting. The plenary papers from the meeting have been published in International Journal of Modern Physics D as they were submitted. The table of contents includes also the links to YouTube videos with talks given at the meeting and cover plenary talks, public lectures, roundtables and all parallel sessions. The general link to the videos from MG16 is: https://www.youtube.com/watch?v=QFe1lsSid-o&list=PLr5RLbSWSonsaOnZukBDs0qsNIWM8AvRF.

As the editors we would like to express our gratitude to all the chairpersons of the parallel sessions at MG16, who peer-reviewed the papers submitted for these proceedings, as well as to the ICRANet secretariat office and in particular to Cinzia di Niccolo, Elisabetta Natale and Yasmina Di Domizio, as well as to ICRANet system manager Gabriele Brandolini for their help in preparation of this publication.

Remo Ruffini and Gregory Vereshchagin
November 2021

This page intentionally left blank

Contents

Publications in this Series v

Sponsors and Acknowledgements vii

Organizing Committees ix

Marcel Grossmann Awards xi

Preface xxxvi

PART A
PLENARY SESSIONS

New results from testing relativistic gravity with radio pulsars
Michael Kramer 3

Dragging of inertial frames by matter and waves
Jiří Bičák and Tomáš Ledvinka 22

Probes of the progenitors, engines and physics behind stellar collapse
Chris L. Fryer 39

The observation of high-energy neutrinos from the cosmos: Lessons learned for multimessenger astronomy
Francis Halzen 59

The first results of PandaX-4T
Jianglai Liu on behalf of the PandaX Collaboration 85

XRISM: X-ray imaging and spectroscopy mission
Makoto S. Tashiro and the XRISM team 95

The SVOM mission
J.-L. Atteia, B. Cordier and J. Wei on behalf of the SVOM Collaboration 104

Quantum field theory with boundary conditions at the horizons
Gerard 't Hooft 133

The development of general relativity and the cosmological constant
Asghar Qadir 143

The irreducible mass of Christodoulou-Ruffini-Hawking mass formula
Yuan K. Ha 154

Reshaping our understanding on structure formation with the quantum nature of the dark matter
C. R. Argüelles, E. A. Becerra-Vergara, A. Krut et al. 164

First results of LHAASO
Ruo-Yu Liu for the LHAASO Collaboration 180

On the MG16 awards 2021
Remo Ruffini 196

The white dwarf binary merger model of GRB 170817A
J. A. Rueda, R. Ruffini, Liang Li et al. 217

PARALLEL SESSIONS

Accretion

• MHD Processes Near Compact Objects
Chairperson: Sergey Moiseenko
YouTube link: https://youtu.be/2WMTg06ZmV8

A semi-implicit multidimensional unstructured gas dynamical solver for astrophysical applications
Ilya A. Kondratyev and Sergey G. Moiseenko 242

Magnetized neutron stars propagating through a non-uniform ISM
O. D. Toropina, M. M. Romanova and R. V. E. Lovelace 255

Calculation of the kinetic coefficients of arbitrary degenerate electrons in magnetized dense matter
M. V. Glushikhina and G. S. Bisnovatyi-Kogan 264

Modeling of magnetic fields of accretion discs, using no-z- and RZ-approximations
M. V. Pashentseva and E. A. Mikhailov 272

• Accretion Discs and Jets
Chairpersons: Audrey Trova and Shokoufe Faraji
YouTube link: https://youtu.be/29Wj9RCVEKw

Limiting effects in tori clusters
D. Pugliese and Z. Stuchlik 280

Hydrodynamical transport of angular momentum in accretion disks in the presence of nonlinear perturbations due to noise
Subham Ghosh and Banibrata Mukhopadhyay 295

Properties of accretion disc models in the background quadrupole
Shokoufe Faraji 307

Magnetized tori around a uniformly accelerating black hole
Shokoufe Faraji and Audrey Trova 317

Multifrequency behaviour of high mass X-ray binaries (Time lag between optical and X-ray outbursts)
Franco Giovannelli 321

Active Galactic Nuclei

• The Black Hole in M87
Chairpersons: Brian Punsly and Jorge Rueda
YouTube link: https://youtu.be/l1lTgksyJag

Rotation of the crescent image of M87* and polarization of its ESE hotspot
Krzysztof Nalewajko 339

Magnetic reconnection in jet-accretion disk systems
Chandra B. Singh, Elisabete M. de Gouveia Dal Pino, Luis H. S. Kadowaki et al. 344

• Machine Learning in Astronomy: AGN, Transient Events, Cosmology and Others
Chairpersons: Rahim Moradi and Yu Wang
YouTube link: https://youtu.be/ErqrmMZQsBk

Exact fractal model of the universe and possible machine learning methods for the verification of the fractality
A. A. Kirillov, E. P. Savelova and P. O. Vladykina 352

Estimating the photometric redshifts of galaxies and QSOs using regression techniques in machine learning
A. Momtaz, M. H. Salimi and S. Shakeri 368

Deep learning in quasar physics
F. Rastegar Nia, M. T. Mirtorabi, R. Moradi, Y. Wang and A. Vafaei Sadr 382

Cosmological density field emulation and gravitational wave inference based on dimensionality reduction and supervised machine learning
Miguel Conceição, António da Silva and Alberto Krone-Martins 391

Unsupervised photometric detection of galaxy cluster candidates in large surveys
Ana Carvalho, Alberto Krone-Martins and Antonio da Silva 409

• Multiwavelength and Multi-Messenger Observations of Active Galactic Nuclei
Chairpersons: Paolo Giommi and Narek Sahakyan

YouTube link: https://youtu.be/-Hyu2NQsExg

Time-dependent lepto-hadronic modeling of the emission processes in blazar jets
S. Gasparyan, D. Bégué and N. Sahakyan 429

Multiwavelength study of high-redshift blazars
G. Harutyunyan and D. Israyelyan 445

Alternative Theories

• Extended Theories of Gravity and Quantum Cosmology
Chairpersons: Yi-Fu Cai and Wentao Luo

YouTube links: https://youtu.be/ADRr9DfV5zM
https://youtu.be/eOzpiC1cFkU
https://youtu.be/kYzJds_JIp8

Quantum gravity phenomenology from thermodynamics of spacetime
A. Alonso-Serrano and M. Liška 462

Gauge theory of gravity based on the correspondence between the 1^{st} and the 2^{nd} order formalisms
David Benisty 479

$U(1)$ local strings in hybrid metric-Palatini gravity
Tiberiu Harko, Francisco S. N. Lobo and Hilberto M. R. da Silva 485

Inflationary supersymmetric FRLW quantum cosmology
N. E. Martínez-Pérez, C. Ramirez and V. M. Vázquez-Báez 499

Effective $f(R)$ actions for modified loop quantum cosmologies
Ana Rita Ribeiro, Daniele Vernieri and Francisco S. N. Lobo 517

Probing multiverse using gravitational wave observations
Moe Kukihara and Kazuhiro Hayama 531

Operator ordering ambiguity in observables of quantum cosmology
Harkirat Singh Sahota and Kinjalk Lochan 538

Decoupled quark stars in self-interacting Brans-Dicke gravity
M. Sharif and Amal Majid 548

Big-bounce in projectively invariant Nieh-Yan models: The Bianchi I case
Flavio Bombacigno, Simon Boudet, Gonzalo J. Olmo and Giovanni Montani 561

Late time cosmology with derivatives of matter Lagrangian
Shahab Shahidi 576

On the semiclassical and quantum picture of the Bianchi I polymer dynamics
E. Giovannetti, G. Montani and S. Schiattarella 588

Quantum corrections to the Bianchi II transition under local rotational invariance
Sara F. Uria, David Brizuela and Ana Alonso-Serrano 597

• **Mathematical Problems of Relativistic Physics: Classical and Quantum**
Chairpersons: A. Shadi Tahvildar-Zadeh and Michael Kiessling

YouTube links: https://youtu.be/9Dr3M9Kb2jo
https://youtu.be/gMOykpapJ5A

The hypercomplex medium as storage of physical equations
Alexander P. Yefremov 605

The Maxwell-Bopp-Landé-Thomas-Podolsky-Einstein system for a static point source
Érik Amorim 619

On recent developments in the theory of relativistic dissipative fluids
V. Hoang 627

Adiabatic solutions in general relativity as null geodesics on the space of boundary diffeomorphisms
Emine Şeyma Kutluk 635

The point spectrum of the Dirac Hamiltonian on the zero-gravity Kerr-Newman spacetime
M. Kiessling, E. Ling and A. S. Tahvildar-Zadeh 648

Causal fermion systems: Classical gravity and beyond
Felix Finster 661

Newman-Penrose-Debye formalism for fields of various spins in pp-wave backgrounds
Aleksandr Kulitskii and Elena Yu Melkumova 679

Gravitational geometric phase
Banibrata Mukhopadhyay, Tanuman Ghosh and Soumya Kanti Ganguly 689

Retarded potentials and radiation in odd dimensions
D. V. Gal'tsov and M. Khlopunov 699

Wave propagation in the anti-deSitter optical metric
D. García-Peláez and C. S. López-Monsalvo 713

New approaches to constrained dynamics and Hamilton-Jacobi procedures in general relativity
D. Salisbury, J. Renn and K. Sundermeyer 719

Orientability of space from electromagnetic quantum fluctuations
N. A. Lemos and M. J. Rebouças 725

Essential self-adjointness of Dirac operators under the influence of general-relativistic gravity
Michael K.-H. Kiessling, A. Shadi Tahvildar-Zadeh and Ebru Toprak 736

• **Wormholes, Energy Conditions and Time Machines**
Chairpersons: Francisco Lobo and Diego Rubiera-Garcia

YouTube links: https://youtu.be/tu_3Wqcd9Ys
https://youtu.be/NqN1c-2fv8Y

Relic magnetic wormholes as possible source of toroidal magnetic fields in galaxies
A. A. Kirillov and E. P. Savelova 743

Wormhole geometries induced by action-dependent Lagrangian theories
Ismael Ayuso, Francisco S. N. Lobo and José P. Mimoso 756

Gravitational lensing by wormholes in binary systems
S. Pietroni 774

Hyper-fast positive energy warp drives
E. W. Lentz 779

From black-bounce to traversable wormhole, and beyond
Alex Simpson 787

Tractor beams, pressor beams, and stressor beams within the context of general relativity
Matt Visser, Jessica Santiago and Sebastian Schuster 808

A singularity theorem for evaporating black holes
E.-A. Kontou, B. Freivogel and D. Krommydas 822

Circularly symmetric thin-shell wormholes in F(R) gravity with (2+1)-dimensions
Cecilia Bejarano, Ernesto F. Eiroa and Griselda Figueroa-Aguirre 831

Warp drive dynamic solutions considering different fluid sources
Osvaldo L. Santos-Pereira, Everton M. C. Abreu and Marcelo B. Ribeiro 840

Symmetries and geometry of spacetime: Towards a new paradigm
Francisco Cabral, Francisco S. N. Lobo and Diego Rubiera-Garcia 856

• Theories of gravity: Alternatives to the cosmological and particle standard models
Chairpersons: Stefano Bellucci and Orlando Luongo

YouTube links: https://youtu.be/-aQcrYoQfBM
https://youtu.be/WsCwPb5OQhY
https://youtu.be/JMnrUfBgqVU
https://youtu.be/pPhZY-bbsew

Thermodynamics of scalar-tensor gravity: A new approach
Valerio Faraoni 876

Two body dynamics in a quadratic modification of general relativity
Soham Bhattacharyya 883

Alternatives to Λ: Torsion, generalized couplings, and scale invariance
C. J. A. P. Martins, C. M. J. Marques, C. B. D. Fernandes et al. 907

Model-independent test of scalar-tensor gravity theory by reconstructing scalar mode of GW170817
Yuya Gushima and Kazuhiro Hayama 921

Cosmology in the novel scalar-tensor representation of $f(R,T)$ gravity
Tiago B. Gonçalves, João Luís Rosa and Francisco S. N. Lobo 932

On the interaction between electromagnetic, gravitational, and plasma related perturbations on LRS class II spacetimes
P. Semrén 943

Condition for expansion-collapse duality between Einstein and Jordan frames
Dipayan Mukherjee, H. K. Jassal and Kinjalk Lochan 958

The model of dark energy based on the quantum-mechanical uncertainty relation
Yu. V. Dumin 967

• Conformal Dilaton Gravity and Related Issues
Chairperson: Reinoud Jan Slagter

YouTube link: https://youtu.be/A3Ygi3YBs5A

Conformal dilaton gravity, antipodal mapping and black hole physics on a warped spacetime
R. J. Slagter 978

From neutrino masses to the full size of the universe — Some intriguing aspects of the tetron model
B. Lampe 999

Summary parallel session AT5
Reinoud Jan Slagter 1014

• Horava-Lifshitz Gravity
Chairperson: Anzhong Wang

YouTube links: https://youtu.be/vPT1dH1zITE
https://youtu.be/5z7zhpiDOpw

Boundary conditions for the Klein-Gordon field on Lifshitz spacetimes
Lissa de Souza Campos 1017

Dynamical system analysis of Bianchi-I spacetimes in $f(R)$ gravity
Saikat Chakraborty, Kazuharu Bamba and Alberto Saa 1026

Cosmological implications in modified Hořava-Lifshitz gravity
Abdul Jawad, Kazuharu Bamba and Farwa Khurshid 1038

Finite action principle and wormholes
Jan Chojnacki and Jan Kwapisz 1046

Strange quark stars in Hořava gravity
Grigoris Panotopoulos 1054

Shadows of Kerr-like black holes in $4D$ Einstein–Gauss–Bonnet gravity and constraints from EHT observations
Sushant G. Ghosh and Rahul Kumar Walia 1069

Wormhole interaction in 2d Horăva-Lifshitz quantum gravity
Jan Ambjørn, Yuki Hiraga, Yoshiyasu Ito and Yuki Sato 1084

Nature of singularities in vector-tensor theories of gravity
V. H. Satheeshkumar 1095

Hořava-Lifshitz and Einstein-Æther gravity in the light of Event Horizon Telescope observations of M87*
Emmanuel N. Saridakis 1104

Hořava-Lifshitz gravity in $(3 + 1)$ dimensions coupled with anisotropic matter and possible constraints from GRB 170817A
Tao Zhang and Fu-Wen Shu 1112

Summary of the parallel session AT6
Anzhong Wang 1119

• Ghost-Free Models of Modified Gravity: Massive Gravity, Horndeski and DHOST Theories, Other Related Models; Their Properties and Solutions
Chairpersons: Dmitry Gal'tsov and Michael Volkov

YouTube links: https://youtu.be/l8KHUPnT2D8
https://youtu.be/48OHIKpgNqs

Non-local R^2-like inflation, gravitational waves and Non-Gaussianities
K. Sravan Kumar 1124

Palatini kinetic scalar-tensor theory: Analytical and numerical solutions
D. V. Gal'tsov and D. S. Bushuev 1136

PART B

Black Holes: Theory and Observations/Experiments

• Theoretical and Observational Studies of Astrophysical Black Holes
Chairperson: Alexander Zakharov

YouTube links: https://youtu.be/fiv_MH-N2kw
https://youtu.be/GYoOb17GvE8

Reconstruction of a star motion in the vicinity of black hole from the redshift of the electromagnetic spectrum
S. O. Komarov and A. K. Gorbatsievich 1151

Shadows of hairy Kerr black holes and constraints from M87*
Sushant G. Ghosh and Misba Afrin 1167

Displacement memory and BMS symmetries
Shailesh Kumar 1179

Physical black holes in semiclassical gravity
Sebastian Murk and Daniel R. Terno 1196

• Black Hole Thermodynamics
Chairperson: Hernando Quevedo

YouTube links: https://youtu.be/XmZDf5mrXQk
https://youtu.be/amseL2qykfk

Information recovery from evaporating rotating charged black holes
Zhi-Wei Wang, Samuel L. Braunstein and Saurya Das 1212

Black hole thermodynamics from entanglement mechanics
S. Mahesh Chandran and S. Shankaranarayanan 1223

Thermodynamics of charged black hole
M. Sharif and Amjad Khan 1238

Rindler trajectories and Rindler horizons in the Schwarzschild spacetime
Kajol Paithankar and Sanved Kolekar 1250

Linear growth of the two-point function for the Unruh state in $1+1$ dimensional black holes
Paul R. Anderson, Zachary P. Scofield and Jennie Traschen 1255

Stress-energy tensor for a quantized scalar field in a four-dimensional black hole that forms from the collapse of a null shell
Shohreh Gholizadeh Siahmazg, Paul R. Anderson, Raymond D. Clark and Alessandro Fabbri 1265

Microscopic model building for black hole membranes from constraints of symmetry
Swastik Bhattacharya and S. Shankaranarayanan 1275

Einstein-Maxwell-Dilaton-Axion mass formulas for black holes with struts and strings
Dmitri Gal'tsov, Gérard Clément and Igor Bogush 1291

• Black Holes in Alternative Theories of Gravity
Chairpersons: Jutta Kunz and Kamal Hajian

YouTube links: https://youtu.be/FRNGJKhiw7c
https://youtu.be/Tjfmuut1Eo0

Holography for rotating black holes in $f(T)$ gravity
Masoud Ghezelbash 1308

Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity
Semin Xavier, Jose Mathew and S. Shankaranarayanan 1319

Universe in a black hole with spin and torsion
Nikodem Popławski 1327

Asymptotically flat black hole solution in modified gravity
Surajit Kalita and Banibrata Mukhopadhyay 1337

Shadow of a charged black hole surrounded by an anisotropic matter field
Javier Badía and Ernesto F. Eiroa 1343

Constraining modified gravity theories with physical black holes
Sebastian Murk 1351

Penrose suggestion as to pre-Planck-era black holes showing up in present universe data sets discussed, with a possible candidate as to GW radiation which may provide initial CMBR data
A. W. Beckwith 1359

Summary of the parallel session BH3
Kamal Hajian and Jutta Kunz 1372

Binaries

• **Explosive Events Associated with Compact-Object Binary Mergers**
Chairpersons: Chris Belczynski and Jorge Rueda

YouTube links: https://youtu.be/Dwq1ZU3gKrg
https://youtu.be/nw02ylI6R2M

Uncertainties in kilonova modeling
C. L. Fryer, C. J. Fontes, O. Korobkin et al. 1391

• **Post-Newtonian and Post-Minkowskian Corrections for Binary Gravitating Systems**
Chairperson: Johannes Bluemlein

YouTube link: https://youtu.be/wfiLG5r08yE

Tutti-Frutti method: Recent developments in the PN/PM/SF treatment of the gravitational two-body problem
Donato Bini and Andrea Geralico 1405

• **Multichannel Studies of Nonstationary Relativistic Stars**
Chairperson: Vladimir Lipunov

YouTube link: https://youtu.be/usn2PlU_qFA

GRB observations on cubesate satellites in the Universat–SOCRAT project
Sergey I. Svertilov, Michail I. Panasyuk, Vitaly V. Bogomolov et al. 1412

Multiwavelength observations of GRB160625B by MASTER, Lomonosov, Konus-Wind and three stage collapse
V. M. Lipunov, V. A. Sadovnichy, M. I. Panasyuk et al. 1429

The role of the magnetic fields in GRB outflows
N. Jordana-Mitjans, C. G. Mundell, S. Kobayashi et al. 1449

MASTER optical observations of the blazar TXS0506+056 during the IC170922A
V. M. Lipunov, K. Zhirkov, V. G. Kornilov et al. 1467

Boson Stars

• Scalar Fields in Cosmology
Chairpersons: Carlos Herdeiro and Alfredo Macias

YouTube links: https://youtu.be/SyLoguueGKk
https://youtu.be/f9nuo8Jvw-w

A short review on nonlinear perturbation theory of structure formation for modified gravity
Jorge L. Cervantes-Cota and Alejandro Aviles 1474

Testing modified gravity theories with marked statistics
Alejandro Aviles 1494

Dark matter as condensed phase of generic bosons
Elías Castellanos and Jorge Mastache 1513

Cosmic Microwave Background

• Cosmic Backgrounds from Radio to Far-IR
Chairperson: Carlo Burigana

YouTube link: https://youtu.be/4e3Cj5wahck

New Planck tSZ map and its cosmological analysis
H. Tanimura, M. Douspis and N. Aghanim 1527

The CMB dipole: Eppur si muove
R. M. Sullivan and D. Scott 1532

High angular resolution Sunyaev Zel'dovich observations: The case of MISTRAL
E. S. Battistelli, E. Barbavara, P. de Bernardis et al. 1542

Cosmological and astrophysical results exploiting magnification bias with high-z sub-millimetre galaxies
L. Bonavera, M. M. Cueli and J. Gonzalez-Nuevo 1557

The impact of the Lorentz symmetry violation on the CMB polarization
Seddigheh Tizchang, Rohoollah Mohammadi and She-Sheng Xue 1571

Cosmic backgrounds from the radio to the far-infrared: Recent results and perspectives from cosmological and astrophysical surveys
Carlo Burigana, Elia Sefano Battistelli, Laura Bonavera et al. 1579

• **New Horizons in Cosmology with CMB Spectral Distortions**
Chairpersons: Jens Chluba and Andrea Ravenni

YouTube links: https://youtu.be/uBYLO4Smw3o
https://youtu.be/5oAPfzAe35k

CMB μT cross-correlations as a probe of PBH scenarios
Ogan Özsoy and Gianmassimo Tasinato 1609

Theoretical and numerical aspects of CMB spectral distortions from non-thermal energy injections
Sandeep Kumar Acharya, Jens Chluba and Abir Sarkar 1628

BISOU: A balloon project to measure the CMB spectral distortions
B. Maffei, M. H. Abitbol, N. Aghanim et al. 1633

Cosmic microwave background spectral distortions constraints on decaying dark matter particles and axion-like particles using *COBE/FIRAS* and *EDGES*
Boris Bolliet 1645

The COSmic Monopole Observer (COSMO)
S. Masi, E. Battistelli, P. de Bernardis et al. 1654

• **Status of the $\mathbf{H_0}$ and $\boldsymbol{\sigma_8}$ Tensions: Theoretical Models and Model-Independent Constraints**
Chairpersons: Joan Solà Peracaula and Adrià Gómez-Valent

YouTube links: https://youtu.be/VNWZ1Bzjus4
https://youtu.be/zEBOCwvetKE
https://youtu.be/nzgC7qV9H_Y

Measuring the Hubble constant H_0 from gravitational lensing
Liliya L. R. Williams 1672

Extra components consistency in the Hubble tension and BBN
Osamu Seto and Yo Toda 1686

Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology
Nick E. Mavromatos 1693

Early and not so early dark energy. What do cosmological observations tell us about them?
Adrià Gómez-Valent, Ziyang Zheng, Luca Amendola, Valeria Pettorino and Christof Wetterich 1713

Renormalized $\rho_{\rm vac}$ without m^4 terms
Cristian Moreno-Pulido and Joan Solà Peracaula 1733

BD-ΛCDM and running vacuum models: Theoretical background and current observational status
Javier de Cruz Pérez, Joan Solà Peracaula, Adrià Gómez-Valent and Cristian Moreno-Pulido 1752

Cosmological tensions: Hints for a new concordance model?
E. Di Valentino 1770

Solving both H_0 and σ_8 tensions in $f(T)$ gravity
Emmanuel N. Saridakis 1783

Precision Cosmology and Hubble tension in the era of LSS surveys
G. Fanizza 1792

• **Effects of Primordial Perturbations Enhancement: From Black Holes Formation to CMB Anomalies**
Chairpersons: Antonio Enea Romano and Krzysztof Turzynski
YouTube link: https://youtu.be/frjjONXbd1M

Primordial black holes arise when the inflaton falls
Keisuke Inomata 1803

Effects of the modification of gravity on the production of primordial black holes
Sergio Andrés Vallejo Peña 1809

Cosmic Strings

• **Cosmic Strings**
Chairpersons: Reinoud Jan Slagter and Batool Imtiaz
YouTube link: https://youtu.be/ZpaU82ZUHzM

$U(1)$ local strings in generalized hybrid metric-Palatini gravity
Hilberto M. R. da Silva, Tiberiu Harko, Francisco S. N. Lobo and João Luís Rosa 1820

New evidence of the azimuthal alignment of quasars spin vector in Large Quasar Groups and cosmic strings
R. J. Slagter 1835

Summary parallel session cosmic strings I
Reinoud Jan Slagter and Batool Imtiaz 1848

• **From Cosmic Strings to Superstrings**
Chairpersons: Carlos Martins and Ivan Rybak
YouTube link: https://youtu.be/LJFtV_4aSAg

Scaling solutions of wiggly cosmic strings
A. R. R. Almeida and C. J. A. P. Martins 1851

High resolution calibration of string network evolution
J. R. C. C. C. Correia and C. J. A. P. Martins 1871

Radiation from Global Cosmic Strings using adaptive mesh refinement
Amelia Drew and E. P. S. Shellard 1891

Analysing the scaling density of axion strings
A. Lopez-Eiguren 1898

Electroweak axion string and superconductivity
Yu Hamada, Yoshihiko Abe and Koichi Yoshioka 1912

Dark Energy and Large Scale Structure

• **Dark Energy and the Accelerating Universe**
Chairpersons: Alexei Starobinky and David Polarski
YouTube links: https://youtu.be/JZEPRS_rqbE
https://youtu.be/3aG_AT4UzWE
https://youtu.be/ZlTWquG-pFk

Hints for the $H_0 - r_d$ tension in uncorrelated Baryon Acoustic Oscillations dataset
Denitsa Staicova 1923

Observational constraints on nonlinear matter extensions of general relativity
E.-A. Kolonia and C. J. A. P. Martins 1935

Constraining the dark energy-dark matter interaction model using low-redshift observations
Archana Sangwan, Joseph P. J. and S. Shankaranarayanan 1948

On the evolution of inhomogeneous perturbations in the ΛCDM model and $f(R)$ modified gravity theories
T. Schiavone and G. Montani 1961

Soft dark energy and soft dark matter
Emmanuel N. Saridakis 1970

A simple parametrisation for coupled dark energy
Vitor da Fonseca, Nelson J. Nunes and Tiago Barreiro 1979

• **Cosmography with Gravitational Lensing**
Chairpersons: Claudio Grillo and Mimoza Hafizi
YouTube link: https://youtu.be/FXplIAvUDBM

A tale of two double quasars: Hubble constant tension or biases?
L. J. Goicoechea and V. N. Shalyapin 1990

Dark Matter

• **Interacting Dark Matter**
Chairpersons: Nikolaos Mavromatos
YouTube links: https://youtu.be/zOshKJlwD-Y
https://youtu.be/JR44dh2GYik

Dark energy and dark matter unification from dynamical space time: BBN constraints
D. Benisty 2005

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2013

LHC experiments for long-lived particles of the dark sector
Vasiliki A. Mitsou 2029

Constraining the interactions in the dark sector with cosmological data
Adrià Gómez-Valent, Valeria Pettorino and Luca Amendola 2050

Running vacuum interacting with dark matter or with running gravitational coupling. Phenomenological implications
Joan Solà Peracaula 2069

Dark matter properties from the Fornax globular cluster timing: Dynamical friction and cored profiles
D. Blas 2089

Growth of linear perturbations in a universe with superfluid dark matter
S. Banerjee, S. Bera and D. F. Mota 2101

Interacting dark sector in the late Universe: Mapping fields and fluids, and observational signatures
Joseph P. J. and S. Shankaranarayanan 2119

The role of self interactions in the cosmological evolution of warm dark matter
R. Yunis, C. R. Argüelles, D. López Nacir et al. 2127

Interaction energy between a charged medium and its electromagnetic field as a dark matter candidate
Mayeul Arminjon 2139

• Dark Matter Searches with Liquid Xenon and Argon Detectors and Self Gravitating Systems and Dark Matter
Chairpersons: Marco Merafina and Soroush Shakeri and She-Sheng Xue

YouTube link: https://youtu.be/H9oGYnGq9pI

The maximum mass of dilute axion stars
Pierre-Henri Chavanis 2149

A dark matter solution for the XENON1T electron excess and the galactic center 511 keV line
Yasaman Farzan 2174

Preliminary results of rich galaxy clusters' spatial distribution analysis on CfA2 Redshift Survey data: Compact objects or dark matter presence at redshift less 0.032
I. V. Arkhangelskaja, A. M. Galper, L. N. Khanh and D. N. Dorosheva 2189

• Dark Matter: Beyond ΛCDM
Chairpersons: Carlos Argüelles and Andreas Krut

YouTube links: https://youtu.be/hdKeo5L7pYE
https://youtu.be/i0IPHXzmV-s

Probing the nature of dark matter with Milky Way subhaloes
M. R. Lovell 2202

Addressing classical cosmological back-reaction with multiple scales
Yonadav Barry Ginat 2217

Imaging formation process for DM profiles
Omar de J. Cabrera-Rosas and Tonatiuh Matos 2222

The self-gravitating Fermi gas in Newtonian gravity and general relativity
Pierre-Henri Chavanis 2230

• Dark Matter and Rare Processes
Chairpersons: Rita Bernabei and Zurab Berezhiani

YouTube links: https://youtu.be/wheVrbETP_0
https://youtu.be/dVdZ4TIDxt0

The dark matter: DAMA/LIBRA and its perspectives
R. Bernabei, P. Belli, V. Caracciolo et al. 2252

Dark matter directionality approach
R. Bernabei, P. Belli, V. Caracciolo et al. 2272

Collapse models under test by high sensitivity γ-ray and X-ray measurements
C. Curceanu, Kristian Piscicchia, Massimiliano Bazzi et al. 2288

Leptophilic dark matter at linear colliders
P. S. Bhupal Dev 2296

DM6 session: Dark matter and rare processes
R. Bernabei and Z. Berezhiani 2316

• **The Nature of Galactic Halos**
Chairpersons: Francesco De Paolis and Asghar Qadir
YouTube link: https://youtu.be/qIwVnNxi0n0

Primordial black holes as dark matter candidates in the Galactic halo
Lindita Hamolli, Mimoza Hafizi, Francesco De Paolis and Achille A. Nucita 2319

Giant cosmic ray halos around M31 and the Milky Way
S. Recchia, S. Gabici, F. A. Aharonian and V. Niro 2335

A nearly complete census of intergalactic gas using the kinematic Sunyaev-Zel'dovich effect
Chaves-Montero, Jonás 2345

Searching for Intermediate Mass Black Holes in the Milky Way's galactic halo
A. Franco, A. A. Nucita, F. De Paolis, F. Strafella and M. Maiorano . . . 2352

Virial clouds evolution from the last scattering up to the formation of first stars
Noraiz Tahir, Asghar Qadir, Muhammad Sakhi and Francesco De Paolis 2360

Testing Weyl-modified gravity on M31 and Milky Way
Muhammad Bilal and Asghar Qadir 2365

PART C

Education

• **Teaching Einsteinian Physics to School Students**
Chairpersons: David Blair and Matteo Luca Ruggiero
YouTube links: https://youtu.be/W-WV6J8kprg
https://youtu.be/UHYpwKQ09SU

Teaching relativity: A paradigm change
F. Herrmann and M. Pohlig 2371

Teaching relativity: Computer-aided modeling
F. Herrmann and M. Pohlig 2381

Solstices and Equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, and the stellar aberration of Sirius
Costantino Sigismondi and Silvia Pietroni 2388

Positional astrometry at arcsecond accuracy using historical instruments, with light equipment
Costantino Sigismondi and Lorenzo Ricciardi 2398

Daily, seasonal, and equinoctial solar paths on a school soccer field
Costantino Sigismondi 2411

Teaching relativity at the AstroCamp
C. J. A. P. Martins 2415

Sungrazing comets as General Relativistic gravitational probes
Silvia Pietroni and Costantino Sigismondi 2424

The three Summer solstice's markers of 1721 unveiled in the Basilica of Santa Maria degli Angeli in Rome
Costantino Sigismondi 2428

Einstein-First: Bringing children our best understanding of reality
A. Popkova, K. Adams, S. Boublil et al. 2438

Exact Solutions

• **Exact Solutions in Four and Higher Dimensions**
Chairpersons: David Blair and Matteo Luca Ruggiero

YouTube link: https://youtu.be/GHZUS5-4gVQ

Kundt spacetimes in the Einstein–Gauss–Bonnet theory
R. Švarc, J. Podolský and O. Hruška 2453

Exact decoupled solutions in curvature-matter coupled gravity
M. Sharif and Fizza Furqan 2464

Tolman-Oppenheimer-Volkov conditions beyond spherical symmetry
José P. Mimoso, Alan Maciel and Morgan Le Delliou 2479

A spherically symmetric stiff fluid spacetime in light of cosmic structure formation
Daniele Gregoris 2497

• **Exact Solutions (Including Higher Dimensions)**
Chairperson: Susan Scott

YouTube link: https://youtu.be/-hJjJvmmOew

Three-parameter solution for the null-surface formulation in 2+1 dimensions
Tina A. Harriott and J. G. Williams 2510

New exact stationary cylindrical anisotropic fluid solution of GR
M.-N. Célérier 2522

Early Universe

• **Quantum Fields**
Chairperson: Andrei Lebed

YouTube links: https://youtu.be/BLaTp0r0TkQ
https://youtu.be/ZifgaSDy5Vc

Hydrodynamic representation and energy balance for the Dirac and Weyl fermions in curved space-times
Tonatiuh Matos, Omar Gallegos and Pierre-Henri Chavanis 2533

Breakdown of the Equivalence Principle for a composite quantum body
A. G. Lebed 2551

Extended DeWitt-Schwinger subtraction scheme, heavy fields and decoupling
Antonio Ferreiro and Jose Navarro-Salas 2557

Renormalization and decoupling for the Yukawa model in curved spacetime
Sergi Nadal-Gisbert, Antonio Ferreiro and José Navarro-Salas 2562

Trace anomaly and evaporation of spherical black holes
P. Meda 2573

On decay of the false Unruh vacuum
A. Shkerin 2587

Behaviour of noise kernel in de Sitter and FRW space-times
Ankit Dhanuka and Kinjalk Lochan 2600

Breaking Buchdahl: Ultracompact stars in semiclassical gravity
Julio Arrechea, Carlos Barceló, Raúl Carballo-Rubio and Luis J. Garay 2608

Einstein anomaly with tensors of odd order in six dimensional curved space
Kohei Yamamoto and Satoshi Yajima 2619

Quantum memory and BMS symmetries
Sanved Kolekar and Jorma Louko 2623

• Topological Methods, Global Existence Problems, and Spacetime Singularities
Chairperson: Spiros Cotsakis

YouTube link: https://youtu.be/H8Itnc_C10c

Gravitational singularities, scattering maps for bouncing, and structure-preserving algorithms
Philippe G. LeFloch 2630

Brane-world asymptotics in a nonlinear fluid bulk
I. Antoniadis, S. Cotsakis and Ifigeneia Klaoudatou 2645

Primordial synchronization of Mixmaster spatial points
Spiros Cotsakis 2657

• The Early Universe
Chairperson: Stefano Ansoldi

YouTube links: https://youtu.be/m80hHWgOlFs
https://youtu.be/6tyaZ8MMVtw

Quintessential inflation from Lorentzian slow roll
David Benisty 2663

Condensed light, quantum black holes and L-CDM cosmology: Experimentally suggested and tested unified approach to dark matter, dark energy, cosmogenesis and two-stage inflation
Victor Borsevici 2672

Helical magnetic fields lead to baryogenesis
Ashu Kushwaha and S. Shankaranarayanan 2692

Polymer Quantization of the Isotropic Universe: Comparison with the Bounce of Loop Quantum Cosmology
G. Barca, E. Giovannetti, F. Mandini and G. Montani 2700

General relativistic evolution equations for density perturbations in open, flat and closed FLRW universes and the problem of structure formation
Pieter G. Miedema 2708

Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data
M. Berti 2726

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2737

Fundamental Interactions and Stellar Evolution

• Why and How the Sun and the Stars Shine: The Borexino Experiment
Chairpersons: Giampaolo Bellini, Dmitry Naumov, Gioacchino Ranucci, Gemma Testera

YouTube links: https://youtu.be/hh5wDnM8miU
https://youtu.be/TG-HgBf7W4s
https://youtu.be/HLiis2LFeEs

Experimental detection of the CNO cycle
B. Caccianiga, N. Rossi, G. Testera et al. 2753

Borexino detector performances
A. Caminata, M. Agostini, K. Altenmüller et al. 2765

Study of antineutrinos from the Earth and the Cosmos with the Borexino detector
Sandra Zavatarelli, M. Agostini, K. Altenmuller et al. 2774

Unveiling the engine of the Sun: Measurements of the *pp*-chain solar neutrinos with Borexino
D. Guffanti, A. C. Re and O. Smirnov 2785

Electron neutrino survival probability in the energy range 200 keV–15 MeV
Marco Pallavicini on behalf of the Borexino Collaboration 2804

The relevance of pp-chain and CNO-cycle neutrino measurements for solar physics
F. L. Villante and A. M. Serenelli 2815

Role of the CNO cycles in stars
A. Ianni 2835

Geoneutrino observation
Tadao Mitsui for the KamLAND Collaboration 2840

Synthesis of the session: Why and how the Sun and the stars shine
Gianpaolo Bellini 2845

• **Rotation in Stellar Evolution**
Chairperson: Georges Meynet

YouTube link: https://youtu.be/9LNoiwa0nv8

The internal rotation of low-mass stars from solar and stellar seismology
G. Buldgen and P. Eggenberger 2848

The rotation of supermassive stars
L. Haemmerlé 2865

Fast Transients

• **What Can We Learn from a Growing Sample of Fast Radio Bursts?**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/yo4n1SgfUrQ
https://youtu.be/NwGooPauhjU

Cosmology with high-redshift FRBs
A. Fialkov 2880

• **Non Standard Cosmological Probes**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/EEFUgiFeMck
https://youtu.be/Ryb15AINfMs

Closing the cosmological loop with the redshift drift
C. J. A. P. Martins, C. S. Alves, J. Esteves, A. Lapel and B. G. Pereira 2890

Gamma-Ray Bursts as potential cosmological probes
L. Izzo 2906

Surface brightness fluctuations: The method and future applications
Michele Cantiello 2915

Preliminary results of analysis of Ia supernovae redshift distributions on data of the Asiago Supernova and Open Supernova Catalogues
I. V. Arkhangelskaja 2930

• **Photospheric Emission in GRBs**
Chairpersons: Gregory Vereshchagin and Damien Bégué

YouTube links: https://youtu.be/ZifnzoUXIFc
https://youtu.be/n7vWAVhFXiU

Understanding prompt emission: Where do we stand?
Asaf Pe'er 2946

On explaining prompt emission from GRB central engines with photospheric emission model
M. Bhattacharya and P. Kumar 2957

Monte Carlo simulations of photospheric emission in Gamma Ray Bursts
T. M. Parsotan and D. Lazzati 2972

The photosphere emission spectrum of hybrid relativistic outflow for Gamma-Ray Bursts
Yan-Zhi Meng, Jin-Jun Geng and Xue-Feng Wu 2982

On diffusive photospheres in Gamma-Ray Bursts
G. V. Vereshchagin 2989

Summary of the parallel session GB3
G. V. Vereshchagin and D. Bégué 3002

• High and Very High Energy Emission from Gamma-Ray Bursts
Chairpersons: Francesco Longo and Fabian Schüssler

YouTube links: https://youtu.be/ZUkLyyYowaM
https://youtu.be/zwN1mNVeUzA

Synchrotron and synchrotron self-Compton emission components in GRBs detected at very high energies
Jagdish C. Joshi, Vikas Chand and Soebur Razzaque 3009

The VERITAS gamma-ray burst follow-up program
D. Ribeiro for the VERITAS Collaboration 3017

MAGIC view of Gamma-Ray Bursts at very high energies
A. Berti on behalf of the MAGIC Collaboration 3030

Prospects for VHE monitoring of gamma-ray bursts with SWGO
G. La Mura, U. Barres de Almeida, R. Conceição et al. 3041

Theoretical implications on the very high energy emission from GRB 190114C
D. Miceli, A. Berti, Z. Bosnjak et al. on behalf of the MAGIC Collaboration 3052

AGILE and GRBs: 13 years of observations
A. Ursi on behalf of the AGILE Team 3062

Searching for Gamma-Ray Bursts with the High-Altitude Water Cherenkov (HAWC) observatory
K. L. Engel for the HAWC Collaboration 3074

• **Electromagnetic Counterparts of Compact Binary Mergers**
Chairpersons: Jonathan Granot and Paz Beniamini

YouTube link: https://youtu.be/AS7bwaT48us

CALET search for gamma-ray counterparts of gravitational wave events
Masaki Mori for the CALET Collaboration 3084

• **Unusual and New Types of Gamma-Ray Bursts**
Chairperson: Binbin Zhang

YouTube link: https://youtu.be/VrF0iU8Q6us

Off-axis jet scenario for early afterglow emission of low-luminosity Gamma-Ray Burst GRB 190829A
Yuri Sato, Kaori Obayashi, Ryo Yamazaki, Kohta Murase and Yutaka Ohira 3095

• **Gamma-Ray Burst Correlations: Observational Challenges and Theoretical Interpretation**
Chairpersons: Maria Giovanna Dainotti and Liang Li

YouTube links: https://youtu.be/gcEy2h6y1jg
https://youtu.be/m2pCT6RAmI0
https://youtu.be/GTCvcVhBjN4

GRB prompt phase spectra under backscattering dominated model
Mukesh Kumar Vyas, Asaf Pe'er and David Eichler 3101

Applying models of pulsar wind nebulae to explain X-ray plateaux following short Gamma-Ray Bursts
L. C. Strang and A. Melatos 3107

Searching for strange quark planets
Xu Wang, Yong-Feng Huang and Bing Li 3118

Probe the universe by using Gamma-Ray Bursts with X-ray plateaus
Fan Xu and Yong-Feng Huang 3124

A new perspective on cosmology through Supernovae Ia and Gamma Ray Bursts
B. De Simone, V. Nielson, E. Rinaldi and M. G. Dainotti 3130

Theory of plateau phase in Gamma-Ray Bursts
Asaf Pe'er 3141

Exploring the canonical behaviour of long Gamma-Ray Bursts using an intrinsic multi-wavelength afterglow correlation
S. R. Oates, J. L. Racusin, M. De Pasquale et al. 3150

• **GRB 170817A and Binary Models**
Chairpersons: Marica Branchesi and Giulia Stratta
YouTube link: https://youtu.be/FRBvkaLX5WU

Kilonova emission observed so far: A comparison with AT2017gfo
A. Rossi 3162

• **Binary-Driven Hypernovae of Type 1, 2 and 3**
Chairpersons: Carlo Luciano Bianco, Christian Cherubini and Simonetta Filippi
YouTube link: https://youtu.be/8IFLfget3C0

Neutrinos and gamma-ray production from proton-proton interactions in binary-driven hypernovae framework
S. Campion, J. D. Melon Fuksman and J. A. Rueda Hernandez 3172

General relativistic turbulence in spherically symmetric core-collapse supernovae simulations
L. Boccioli, G. J. Mathews and E. P. O'Connor 3184

Gravitational Waves

• **Sources of Gravitational Waves**
Chairperson: Andrew Melatos
YouTube links: https://youtu.be/Lr8b9nKsFm4
https://youtu.be/xuqJOXPDyVY
https://youtu.be/qzOrgojHsMg
https://youtu.be/KCJoWMd71pE

Mountain formation by repeated, inhomogeneous crustal failure in a neutron star
A. D. Kerin and A. Melatos 3194

Gravitational waves from neutrino mass generating phase transitions
Nobuchika Okada and Osamu Seto 3206

Efficiency of registration of chirp bursts and signals of collapsing stars by the Euro-Asian network of GW interferometers
V. N. Rudenko, S. L. Andrusenko, D. P. Krichevskiy and G. D. Manucharyan 3219

Joint analysis method on gravitational waves and low-energy neutrinos to detect core-collapse supernovae
O. Halim, C. Casentini, M. Drago et al. 3228

• **Mid-frequency Gravitational Waves (0.1–10 Hz): Sources and Detection Methods**
Chairperson: Wei-Tou Ni
YouTube link: https://youtu.be/sJ6A7a73Vxw

A cryogenic and superconducting inertial sensor for the Lunar Gravitational–Wave Antenna, the Einstein Telescope and Selene-physics
F. Badaracco, J. V. van Heijningen, E. C. Ferreira and A. Perali 3245

Space gravitational wave antenna DECIGO and B-DECIGO
S. Kawamura and the DECIGO working group 3254

Summary of the parallel session GW2
Dongfeng Gao, Wei-Tou Ni, Jin Wang et al. 3261

• **Numerical Relativity and Gravitational Wave Observations**
Chairperson: Nigel Bishop

YouTube links: https://youtu.be/eZPytAU4Zmk
https://youtu.be/th-3KqkxDnU

Salient features of the optimised PyCBC IMBH search
Koustav Chandra, Archana Pai, V. Villa-Ortega et al. 3277

Matter shells modifying gravitational wave signals
Monos Naidoo, Nigel T. Bishop and Petrus J. van der Walt 3286

Odd-dimensional gravitational waves from a binary system on a three-brane
D. V. Gal'tsov and M. Khlopunov 3301

Developments in numerical relativity and gravitational wave observations
Nigel T. Bishop 3309

High Energy

• **Very High Energy Gamma Rays**
Chairpersons: Razmik Mirzoyan and Alessandro De Angelis

YouTube link: https://youtu.be/jHM1RH20ZyM

Insights into the Galactic Center environment from VHE gamma-ray observations with ground-based facilities
C. Fruck 3316

The TAIGA experiment
M. Tluczykont, I. I. Astapov, A. K. Awad et al. 3324

Science perspectives of the Southern Wide-field Gamma-ray Observatory (SWGO)
K. L. Engel for the SWGO Collaboration 3343

• **Future Missions for High-Energy Astrophysics**
Chairpersons: Filippo Frontera and Shaolin Xiong

YouTube link: https://youtu.be/JgDsZX6RUkU

Laue lenses: Focusing optics for hard X/soft Gamma-ray astronomy
L. Ferro, M. Moita, P. Rosati et al. 3355

ASTENA: A mission concept for a deep study of the transient gamma-ray sky and for nuclear astrophysics
E. Virgilli, F. Frontera, P. Rosati et al. on behalf of the ASTENA Collaboration 3368

Polarimetric prospects of a new hard X-soft gamma-ray space mission for next decades
M. Moita, L. Ferro, F. Frontera et al. 3385

• **The SRG Mission: First Results from eROSITA and ART-XC**
Chairperson: Andrea Merloni
YouTube link: https://youtu.be/l0t1B716UcM

Prospect for WHIM detection in the cosmic web by *SRG/eROSITA*
H. Tanimura and N. Aghanim 3400

• **eXTP – Enhanced X-Ray Timing and Polarimetry Mission**
Chairpersons: Marco Feroci and Fangjun Lu
YouTube link: https://youtu.be/3d08eKmsImI

The role of eXTP in the multi-messenger astronomy era
G. Stratta and Gor Oganesyan 3403

• **Observations of HE and UHE Cosmic Rays**
Chairpersons: Ivan De Mitri and Fabio Gargano
YouTube link: https://youtu.be/l9DZyBMWO5c

CALET on the ISS: The first 5 years
Pier Simone Marrocchesi for the CALET Collaboration 3427

The fluxes of charged cosmic rays as measured by the DAMPE satellite
Paolo Bernardini on behalf of the DAMPE Collaboration 3442

Recent results from the Pierre Auger Observatory
E. Roulet for the Pierre Auger Collaboration 3449

The HERD space mission
F.C.T. Barbato on behalf of the HERD Collaboration 3455

PART D

History of Relativity

• **The "Fall and Rise" of Betelgeuse**
Chairperson: Costantino Sigismondi
YouTube link: https://youtu.be/VmbrE2gYmOM

The observation of the stars in daytime and near the horizon
Costantino Sigismondi and Paolo Ochner 3471

Fall and Rise of Betelgeuse: The summary of HR1 session
Costantino Sigismondi and Paolo Ochner 3475

Photometry of Betelgeuse at daylight
Otmar Nickel 3479

Evidence for dynamical changes in Betelgeuse using multi-wavelength data
Sneha Kachhara, Sandip V. George, Ranjeev Misra and G. Ambika 3485

The curious case of Betelgeuse
Jacco Th. van Loon 3494

Variable stars observed from city sites: The 2500 AAVSO-SGQ database
Costantino Sigismondi and Paolo Ochner 3501

Betelgeuse: An introductory course to observational astronomy
Costantino Sigismondi and Tiziana Pompa 3507

The meridian line of the Vatican obelisk to study the stellar aberration
Costantino Sigismondi and Lorenzo Ricciardi 3513

Betelgeuse, Sirius and the stars in the roman *Settecento*
Costantino Sigismondi 3519

• **History of Relativity, Gravitation and Cosmology**
Chairperson: Luis Crispino

YouTube link: https://youtu.be/RNbPUSp95PQ

On Einstein's last bid to keep a stationary cosmology
Salvador Galindo-Uribarri and Jorge L. Cervantes-Cota 3536

Jayme Tiomno: Relativity, gravity, cosmology, and the Marcel Grossmann Meetings
William D. Brewer 3547

A look inside Feynman's route to gravitation
M. Di Mauro, S. Esposito and A. Naddeo 3563

Towards detecting gravitational waves: A contribution by Richard Feynman
M. Di Mauro, S. Esposito and A. Naddeo 3576

Stellar gravitational collapse, singularity formation and theory breakdown
Kiril Maltsev 3596

The Hamilton-Jacobi analysis by Peter Bergmann and Arthur Komar of classical general relativity
D. Salisbury 3626

• Time and Philosophy in Physics
Chairperson: Shokoufe Faraji

YouTube link: https://youtu.be/986v-V5JJEk

The passage of time and top-down causation
Barbara Drossel 3631

Explaining time's passage
Jonathan J. Dickau 3646

A glimpse to Feynman's contributions to the debate on the foundations of quantum mechanics
M. Di Mauro, S. Esposito and A. Naddeo 3657

Summary of the parallel session HR3
Shokoufe Faraji 3671

Neutron Stars

• Dense Matter in Compact Stars
Chairpersons: Alessandro Drago and Jorge Rueda

YouTube links: https://youtu.be/U5Yr0oDVhqY
https://youtu.be/MvMDFh-1_bc

Massive compact stars in the two-families scenario
P. Char, A. Drago and G. Pagliara 3677

Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond
J. Soldateschi, N. Bucciantini and L. Del Zanna 3684

Screening and elastic properties of the NS crust in the OCP approximation
D. Barba González, C. Albertus Torres and M. A. Pérez-García 3703

Tidal deformability as a probe of dark matter in neutron stars
D. Rafiei Karkevandi, S. Shakeri, V. Sagun and O. Ivanytskyi 3713

Binary neutron star mergers with quark matter equations of state
Atul Kedia, Hee Il Kim, Grant Mathews and In-Saeng Suh 3732

Probing dense matter physics with transiently-accreting neutron stars: The case of source MXB 1659-29
Melissa Mendes, Andrew Cumming, Charles Gale and Farrukh J. Fattoyev 3736

• **Compact Stars as Laboratories for Testing Strong Gravity**
Chairpersons: Aurora Perez Martinez and César Augusto Zen Vasconcellos

YouTube link: https://youtu.be/C1KGecdNEfs

Vacuum properties and astrophysical implications
A. Pérez Martínez, M. Pérez-Garcia, E. Rodríguez Querts and A. Romero Jorge 3756

Testing extended theories of gravity with GRBs
L. Mastrototaro 3762

• **Pulsar Power in Physics and Astrophysics and Pulsars and Pulsar Systems at High Energies**
Chairpersons: Andrea Possenti and Pak-Hin Tam

YouTube links: https://youtu.be/gAG29DZwUbM
https://youtu.be/ytdUBFrViHI

News and views regarding PSR J1757–1854, a highly-relativistic binary pulsar
A. D. Cameron, M. Bailes, V. Balakrishnan et al. 3774

On the origin of the unique isolated X-Ray pulsar 1E 161348-5055 with 6.7 hr. spin period
V. Yu. Kim 3785

Advantages of including globular cluster millisecond pulsars in Pulsar Timing Arrays
M. Maiorano, F. De Paolis, A. A. Nucita and A. Franco 3791

Searching for pulsars in globular clusters with the MeerKAT Radio Telescope
F. Abbate on behalf of the MeerTIME/TRAPUM Collaboration 3799

Precision Tests

• **Gravitational Lensing and Shadows**
Chairpersons: Perlick Volker and Oleg Tsupko

YouTube links: https://youtu.be/6DXXWpMJ3IQ
https://youtu.be/F-K1gXn71_Y
https://youtu.be/B-V8r2HMztw

Gravitational lensing by rotating Simpson–Visser black holes
Sushant G. Ghosh and Shafqat Ul Islam 3812

Killing tensors in foliated spacetimes and photon surfaces
Igor Bogush, Kirill Kobialko and Dmitri Gal'tsov 3827

Decoding black hole metrics from the interferometric pattern of relativistic images
V. Bozza 3839

Symplectic evolution of an observed light bundle
N. Uzun 3844

Shadow of black holes with a plasma environment in 4D Einstein-Gauss-Bonnet gravity
Javier Badía and Ernesto F. Eiroa 3856

Aspects of neutrino mass hierarchy in gravitational lensing
Himanshu Swami 3865

Photon regions in stationary axisymmetric spacetimes and umbilic conditions
K. V. Kobialko and D. V. Gal'tsov 3874

Gravitational lensing by charged accelerating black holes
Torben C. Frost 3885

• Experimental Gravitation
Chairpersons: Angela di Virgilio and Claus Lammerzahl

YouTube link: https://youtu.be/mXcxztQ0nyk
https://youtu.be/tE3gIUBviTc

A manmade experiment aimed to clarify the gravity law in the Solar system
Alexander P. Yefremov and Alexandra A. Vorobyeva 3905

Gravitomagnetic field generation using high permittivity materials in superconducting magnetic energy storage devices
G. V. Stephenson 3910

Large ring laser gyroscopes: Geometry stabilization and laser control
U. Giacomelli, N. Beverini, G. Carelli et al. 3920

Dark gravitomagnetism with LISA and gravitational waves space detectors
A. Tartaglia, M. Bassan, G. Pucacco, V. Ferroni and D. Vetrugno 3929

Light rays in the Solar system experiments: Phases and displacements
Pravin Kumar Dahal and Daniel R. Terno 3942

The Ginger project – preliminary results
C. Altucci, F. Bajardi, A. Basti et al. 3956

• **Variation of the Fundamental Constants, Tests of the Fundamental Symmetries and Probes of the Dark Sector**
Chairpersons: Angela Victor Flambaum and Yevgeny Stadnik

YouTube links: https://youtu.be/NAZA-0tWHak
https://youtu.be/juhDGBJ12Lg
https://youtu.be/zxGgDP2sn60

Varying fundamental constants and dark energy in the ESPRESSO era
C. J. A. P. Martins 3963

• **Dragging is Never Draggy: MAss and CHarge Flows in GR**
Chairperson: Oldrich Semerak

YouTube link: https://youtu.be/ho31IgLNxu8

Testing the general relativistic nature of the Milky Way rotation curve with Gaia DR2
Mariateresa Crosta 3970

Spinning cylinders in general relativity: A canonical form for the Lewis metrics of the Weyl class
L. Filipe O. Costa, José Natário and N. O. Santos 3982

Magnetized black holes: The role of rotation, boost, and accretion in twisting the field lines and accelerating particles
Ondřej Kopáček and Vladimír Karas 3999

Spinning particle: Is Newton-Wigner the only way?
V. Witzany 4010

Gravitomagnetic resonance and gravitational waves
Matteo Luca Ruggiero and Antonello Ortolan 4019

Quantum Gravity

• **Loop Quantum Gravity**
Chairpersons: Marcin Kisielowski and Jerzy Lewandowski

YouTube links: https://youtu.be/VaLPseYWh9E
https://youtu.be/WFIgMqrQ07U

Studying the EPRL spinfoam self-energy
Pietropaolo Frisoni 4026

A spin foam framework for the black-to-white hole transition
Farshid Soltani 4045

Holographic properties of the bulk-to-boundary transmission of information in regions of quantum space
Eugenia Colafranceschi 4062

• **Quantum Gravity Phenomenology**
Chairpersons: Giovanni Amelino-Camelia and Jerzy Kowalski-Glikman

YouTube links: https://youtu.be/vYnVb2zNl0o
https://youtu.be/icK8z80Hm-Y
https://youtu.be/svXLa0yhyYY

Minimal length discretization and properties of modified metric tensor and geodesics
Abdel Nasser Tawfik, Fady T. Farouk, F. Salah Tarabia and Muhammad Maher 4074

The structure of the multiverse from the entanglement entropy
Samuel Barroso Bellido 4082

Effective field theory from relativistic Generalized Uncertainty Principle
Vasil N. Todorinov, Saurya Das and Pasquale Bosso 4088

Stelle gravity as the limit of quantum gravity with maximal momentum
V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das 4107

Baryon asymmetry and minimum length
Saurya Das, Mitja Fridman, Gaetano Lambiase and Elias C. Vagenas . . . 4114

On quantum gravity and quantum gravity phenomenology
Douglas Edmonds, Djordje Minic and Tatsu Takeuchi 4126

WKB approach to the gravity-matter dynamics: A cosmological implementation
G. Maniccia and G. Montani 4146

Natural evidence for fuzzy sphere noncommutative geometry: Super-Chandrasekhar white dwarfs
Surajit Kalita, T. R. Govindarajan and Banibrata Mukhopadhyay 4159

A model of polymer gravitational waves: Theory and some possible observational consequences
Angel Garcia-Chung, James B. Mertens, Saeed Rastgoo, Yaser Tavakoli and Paulo Vargas Moniz 4166

• Loop Quantum Gravity: Cosmology and Black Holes
Chairpersons: Jorge Pullin and Parampreet Singh

YouTube links: https://youtu.be/CAuAK31Ukho
https://youtu.be/JlYnEzRiRR0

Primordial power spectrum from a matter-ekpyrotic scenario in loop quantum cosmology
Bao-Fei Li, Sahil Saini and Parampreet Singh 4178

The primordial power spectra in modified loop quantum cosmology
Bao-Fei Li, Javier Olmedo, Parampreet Singh and Anzhong Wang 4188

Primordial perturbations in kinetically dominated regimes of classical and quantum cosmology
B. Elizaga Navascués, R. Jiménez-Llamas and G. A. Mena Marugán 4193

Revisiting the Hamiltonian formalism of the Ashtekar–Olmedo–Singh black hole model
Alejandro García-Quismondo and Guillermo A. Mena Marugán 4211

A comparison of different choices of clocks in a reduced phase space quantization in loop quantum cosmology with an inflationary potential using effective techniques
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4228

Initial conditions in LQC/mLQCs
Bao-Fei Li, Parampreet Singh and Anzhong Wang 4234

Holonomy corrections in effective midisuperspace models
A. Alonso-Bardaji and D. Brizuela 4239

Infrared signatures of quantum bounce in collapsing geometry
Harkirat Singh Sahota and Kinjalk Lochan 4247

Effective black hole interior and the Raychadhuri equation
Keagan Blanchette, Saurya Das, Samantha Hergott and Saeed Rastgoo . . . 4256

Effect of loop quantization prescriptions on the physics of non-singular gravitational collapse
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4267

Summary of the parallel session QG3
Jorge Pullin and Parampreet Singh 4272

Strong Field

• Strong Electromagnetic and Gravitational Field Physics: From Laboratories to Early Universe

Chairpersons: Sang Pyo Kim and She-Sheng Xue

YouTube links: https://youtu.be/kexqTayqFiU
https://youtu.be/laymGp6x9Hg

Reliable equations of state of viscous strong and electroweak matter
A. Tawfik 4277

Neutral fermion pair production by Sauter-like magnetic step
T. C. Adorno, Zi-Wang He, S. P. Gavrilov and D. M. Gitman 4290

On the magnetic field screening in strong crossed electromagnetic field
S. Campion, J. A. Rueda Hernandez, S.-S. Xue and R. Ruffini 4306

Particle creation by strong fields and quantum anomalies
José Navarro-Salas 4317

Constraints on the non-minimal coupling of electromagnetic fields from astrophysical observations
Susmita Jana and S. Shankaranarayanan 4326

New partial resummation of the QED effective action
Silvia Pla and José Navarro-Salas 4337

Can a detector detect soft photons
Sanved Kolekar and Jorma Louko 4347

Breaking of the adiabatic invariance in the production of particles by strong fields
P. Beltrán-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla 4352

Dynamics of relativistic electrons in non-uniform magnetic fields and its applications in quantum computing and astrophysics
Srishty Aggarwal and Banibrata Mukhopadhyay 4362

Validity of the semiclassical approximation in 1+1 electrodynamics: Numerical solutions to the linear response equation
Ian M. Newsome 4374

On Kerr black hole perfect MHD processes in Doran coordinates
C. Cherubini, S. Filippi, A. Loppini et al. 4387

Tadpole contribution to magnetic photon-graviton conversion
N. Ahmadiniaz, F. Bastianelli, F. Karbstein and C. Schubert 4393

• **The Effects of (Non)Linear Electrodynamics on the Properties of Astrophysical/Gravitational Compact Objects**
Chairperson: Seyed Hossein Hendi

YouTube link: https://youtu.be/SAIAXtQhavE

Correspondence of gamma radiation coming from GRBs and magnetars based on the effects of nonlinear vacuum electrodynamics
Tursynbek Yernazarov, Medeu Abishev and Yerlan Aimuratov 4401

Absorption of massless scalar waves by electrically charged regular black holes
Marco A. A. de Paula, Luiz C. S. Leite and Luís C. B. Crispino 4410

White Dwarfs

• **White Dwarf Explosions**
Chairpersons: Robert Fisher and María Pilar Ruiz Lapuente

YouTube links: https://youtu.be/ndaW6u2xuOo
https://youtu.be/IwjYaQaqJeI

Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass
Friedrich K. Röpke, Florian Lach, Sabrina Gronow et al. 4420

Type Ia supernovae and their explosive nucleosynthesis: Constraints on progenitors
Shing-Chi Leung and Ken'ichi Nomoto 4427

Charged polarized white dwarfs with finite temperature as a possible source of type Ia supernovae
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4447

• **White Dwarfs, Magnetic Compact Stars and Nuclear Astrophysics**
Chairpersons: Manuel Malheiro and Jaziel Goulart Coelho

YouTube link: https://youtu.be/onicFElJQnA

CTCV J2056-3014 and other fast-spinning white dwarfs
C. V. Rodrigues, R. Lopes de Oliveira, A. Bruch et al. 4453

Gravitational waves from fast-spinning white dwarfs
M. F. Sousa, J. C. N. de Araujo and J. G. Coelho 4461

Highly magnetized white dwarfs: Implications and current status
B. Mukhopadhyay, M. Bhattacharya, A. J. Hackett et al. 4475

Electron captures and stability of white dwarfs
N. Chamel, L. Perot, A. F. Fantina et al. 4488

Massive hot white dwarfs: Consequences of finite temperature in the structure and on the onset of instabilities
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4508

A study of the infrared emission of SGR/AXPs in a disk scenario and its implications for their origin
Sarah Villanova Borges 4514

Particle acceleration and high energy emission in the white dwarf binaries AE Aquarii and AR Scorpii
P. J. Meintjes, S. T. Madzime, Q. Kaplan et al. 4522

Study the effects of anisotropy on the highly magnetized white dwarfs
Debabrata Deb, Banibrata Mukhopadhyay and Fridolin Weber 4532

List of Participants 4545

PART C

This page intentionally left blank

Teaching relativity: A paradigm change

F. Herrmann* and M. Pohlig
Institute for Theoretical Solid State Physics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Baden-Württemberg, Germany
**E-mail: f.herrmann@kit.edu*
www.kit.edu

The teaching of relativity usually starts with kinematics: The invariance of the speed of light, clock synchronization, time dilatation and length contraction, the relativity of simultaneity, Lorentz transformation and the Minkowski diagram. The change of the reference frame is a central topic. Only afterwards problems of relativistic dynamics are discussed. Such an approach closely follows the historical development of the Special Theory of Relativity.

We believe that this access to relativity is unnecessarily complicated, and unsuitable for beginners. We present the basics of a teaching approach in which the initial postulate of relativity is the identity of energy and relativistic mass. Reference frame changes are largely avoided.

Keywords: Special relativity, additional postulate, reference frame change, relativistic dynamics.

1. Introduction

Relativity, and we mean for the moment only special relativity, is more than 100 years old, but still has not found its place in school. Just compare: Faraday-Maxwell electromagnetism, which is certainly not simpler than special relativity, would still not be included in the curricula 120 years after its creation, i.e. in 1980. There are several reasons for this deficiency. Here we want to discuss only one of them: Relativity is still taught today as it has originated historically. One starts with a very special relativistic effect, and works through with great effort to the more important and useful general statements. It is as if one enters a splendid palace not by the beautiful main portal, but by some shabby servants' entrance.

We describe the basics of a course on special relativity. The concept has been tested and is used at numerous secondary schools: parts of it already in the lower secondary school, the complete program in the upper secondary school. The course is part of the *Karlsruhe Physics Course*.[1] It can be downloaded from the Internet in various languages. A bilingual English-Chinese version was published recently.[2]

We do not describe the details of this course. We are merely presenting some ideas that underlie its development. Some of the topics we address are also discussed in articles of the column *Historical burdens on physics*.[3]

We justify our paradigm change in Section 2. We choose a different "entrance" to relativity. We substantiate our choice in section 2.1. In sections 2.2 and 2.3 we begin with a critical discussion of two popular topics: the reference frame change,

and the role of the observer. Section 2.4 is about naming. How is the word mass used and how do we want to use it. Section 2.5 deals with the way to write Einstein's famous equation $E = mc^2$. We are thus prepared for our main topic, which will be treated in section 3: relativistic dynamics.

2. Paradigm change

2.1. *Additional postulate of the special theory of relativity*

The laws of the special theory of relativity, or special relativity for short, are obtained from those of classical mechanics by adding one extra postulate. Traditionally and for historical reasons, the choice was made for the invariance of the speed of light upon a change of the reference frame.

Once one has become aware that this choice as a starting point is only one of several possibilities, one discovers that completely new perspectives arise for the development of a teaching concept. We have decided to introduce the mass-energy equivalence as an additional postulate instead of the invariance of the speed of light.

With this choice, we arrive more quickly at that part of special relativity that we consider being the most important one, namely relativistic dynamics.

The traditional choice of the invariance of the speed of light has a rather incidental cause: when special relativity came into being, light was the only known system that behaved relativistically. Einstein's work – both his famous publication of 1905 *Zur Elektrodynamik bewegter Körper*[4] and his textbook *Grundzüge der Relativitätstheorie*[5] – begins with a detailed, and one can say somewhat tiring part on relativistic kinematics.

One might imagine what the course of history would have been if the first relativistic observation had been that a cup of hot coffee is heavier than a cup of cold coffee (or that the corresponding observation had been made with particles in an accelerator). The presentation of the theory of relativity in our textbooks would certainly be very different from what it is actually.

Of course, the mass-energy equivalence is not supported by our everyday experience (neither is the invariance of the speed of light). But one can make it easily plausible and discuss its consequences even in beginners' classes. It leads to surprising and at first unbelievable statements; but it does not lead to the cognitive conflicts one has to deal with in the traditional approach to relativistic kinematics, which questions our basic convictions about space and time.

2.2. *Reference frames and reference frame changes*

In the traditional approach to special relativity, the following topics are dealt with before relativistic dynamics is addressed:

The invariance of the speed of light
Clock synchronization

The relativity of simultaneity
Time dilatation and length contraction
Velocity addition
Lorentz transformation
Minkowski diagram

The problem with such an approach is that one begins with the most confusing part of the theory: the relationship between space and time.

Certainly, students can learn a lot of physics by analyzing the same process in different reference frames. But we should not forget that we are dealing with beginners, and it is better to stick to the old rule: Choose a suitable reference frame right at the beginning, i.e. a reference frame in which the description of your problem becomes as simple as possible, and don't change it anymore.

And above all, don't change the reference frame in the middle of dealing with your problem (as is usually done when discussing the twin paradox). By the way, in classical mechanics and electromagnetism, too, one can create the greatest confusion if one chooses the reference frame improperly or if one changes it in the middle of the discussion.

This is why our decision was not to make reference frame changes the main topic of our lessons and to avoid them as far as possible. Above all, the impression should not be created that special relativity is essentially a theory of reference frame changes – an impression that some presentations certainly arouse. Even the name relativity gives that impression.

2.3. *The observer*

Closely related to the question of the choice of the reference frame is the problem of the so-called observer. The observer seems to be particularly important in two areas of physics: in quantum physics (where the observer always appears as the one making a "measurement") and in the theory of relativity.

An observation is always made from a certain perspective. It thereby emphasizes something that does not play a particular role in the phenomenon to be described.

We believe the observation should not be in the foreground as long as the understanding of a process is the objective. This is especially true when teaching at school, i.e. beginners.

It is true that we get all the information about the world by observing and measuring. But the idea we form of the world is quite different from what we observe. So, if we wanted to explain the shape of the earth to someone, we would certainly not start with the shadow of the obelisk in Alexandria, but simply say: The earth is a sphere.

In our opinion when teaching physics we should primarily give a picture of what nature is like – not how it is perceived by an observer.

2.4. *The use of the term mass*

Mass is a physical quantity that until not so long ago did not cause any problems. It was known for which properties it is a measure.

With Einstein's special theory of relativity, this only changed insofar as the mass of a body became dependent on its velocity, temperature and other variables. It was no longer a quantity that had a characteristic value for a body or a particle.

Thus, a body, a particle, a field, or any other structure, has a mass that depends, among other things, on its velocity. The value that the mass assumes when the centre of mass of the particle or body is at rest is called its rest mass (symbol m_0). Even more appropriate would actually be the less common term *proper mass*, because when the centre of mass is at rest, this does not mean that the parts or particles of the system are at rest.

It is that simple, or, unfortunately, one must say: it could be that simple.

For there is an area of physics in which another use of the term mass has established itself: Particle physics. A particle has a well-defined rest mass. The rest mass is characteristic of the particle species. Among the various other parameters, such as electric charge, spin, lepton number, etc., it is considered the main characteristic. It seems to constitute the identity of the particle. For this property, a compact, plausible name was needed, and particle physicists simply called it mass. Thus, in particle physics, the term mass refers to only part of the quantity that describes the inertia of a particle.

However, this custom also spread beyond particle physics, and this results in several misunderstandings and ambiguities. What is to be understood by the mass of a macroscopic body that is at rest? Is it the mass that would be measured with a (very accurate) scale, or is it the sum of the (rest) masses of the particles that constitute the body? This is a question that particle physicists probably don't ask, but we teachers do.

We have therefore decided to use the term mass (symbol m) exclusively for the quantity that measures gravity and inertia, no matter what kind of object is considered and in what state it is. Thus, a hot cup of coffee has a larger mass than the same coffee when it is cold. A photon has a mass and a magnetic field has a mass (a liter of magnetic field near a neutron star has a mass of some hundred grams).

By the way, if one follows this use of the term mass, it makes no sense to say that mass is a "form of energy" or that mass can be converted into energy.

2.5. *The identity of mass and energy*

First, let us look at the term *mass-energy equivalence*. It is a pity that a simple fact is expressed so unclearly. The word equivalence is certainly not wrong, but why not say directly: Mass and energy are the same physical quantity.

If one were to ask someone who has never seen the equation $E = mc^2$ to express this fact in a formula, he would probably write something like this:

$$E = k \cdot m \ . \tag{1}$$

The factor k tells us how the units joule and kilogram are converted into each other. As the definitions of the units kilogram and joule are independent of the choice of the reference frame, k is a universal constant.

Its value is obtained by a measurement. One finds

$$k = 9 \cdot 10^{16} \mathrm{J/kg} \ . \tag{2}$$

But what is wrong with writing

$$E = mc^2 \ ? \tag{3}$$

Every student learns in mathematics that a linear relationship between the variables x and y is written as

$$y = a \cdot x \ . \tag{4}$$

On the right side first the factor of proportionality a, and second the independent variable. The unbiased student might interpret the famous equation (3) this way: The energy is proportional to the square of the speed of light – and not: energy and mass are the same physical quantity. One might object: This can easily be explained to the students. Of course it can. But doesn't the statement become clearer if one writes $E = k \cdot m$? Would the iconic character of equation (3) survive if it were formulated in this way?

3. The laws of dynamics

In the *Karlsruhe Physics Course*[1], the extensive quantities energy, momentum, electric charge and entropy are introduced as basic quantities. Especially momentum and entropy have a very direct and vivid interpretation. Momentum is a measure of the "amount of motion", that is, what one would colloquially call "impetus" or "drive". Entropy measures almost perfectly what would colloquially be called the amount of heat (not to be confused with the rather difficult concept of heat that has established itself in physics).

Therefore, in the context of relativity, it is natural to ask in the first place for the dependence of different quantities on momentum. Momentum is our independent variable. We give momentum to a body or particle and ask how it reacts to it: How does its mass (= energy) behave? What happens to its velocity? In other words: We ask for the functions $E(p)$ and $v(p)$.

3.1. *The energy momentum relationship*

To derive $E(p)$, we take over as much as possible from non-relativistic physics. In addition, we only require the identity of mass and energy, i.e. we assume the

validity of equation (1). We start with the change dE of the energy, that results from a change of the momentum dp.

$$\mathrm{d}E = v\mathrm{d}p \ . \tag{5}$$

With $p = m \cdot v$ we obtain

$$\mathrm{d}E = \frac{p}{m}\mathrm{d}p \ . \tag{6}$$

Replacing m with E/k, and reordering returns

$$E\mathrm{d}E = kp\mathrm{d}p \ . \tag{7}$$

We thus obtain

$$\mathrm{d}E^2 = k\mathrm{d}p^2 \tag{8}$$

and

$$E^2 = kp^2 + C \ , \tag{9}$$

where C is the constant of integration.

The value of C can easily be determined, because for $p = 0$ the energy E assumes the value of the rest energy E_0. Thus, C must be equal to ${E_0}^2$. We therefore get

$$E^2 = {E_0}^2 + kp^2 \tag{10}$$

and for the sought-after relationship between energy and momentum we get:

$$E(p) = \sqrt{{E_0}^2 + kp^2} \ . \tag{11}$$

The red line in Figure 1 shows the graphic representation of relation (11). Two limiting cases are of particular interest.

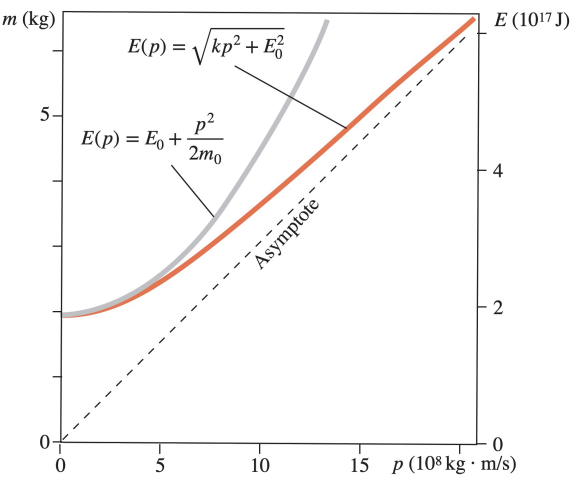


Fig. 1. Relationship between mass/energy and momentum (red line). For large values of the momentum the curve approaches the asymptote (dashed line), for small values the classical quadratic relation (grey line).

For small momentum values, equation (11) changes to

$$E(p) = E_0 + \frac{kp^2}{2E_0} = E_0 + \frac{p^2}{2m_0} \ . \tag{12}$$

We obtain the classical kinetic energy, increased by the rest energy (grey line in Figure 1).

If the momentum is very large, so that ${E_0}^2$ can be neglected in comparison with kp^2, equation (11) turns into

$$E(p) = \sqrt{k}p \ , \tag{13}$$

see the dashed line in Figure 1. For bodies whose rest mass is 0 kg, equation (13) applies for all values of the momentum, not only for large values (Figure 2). Thus, in the highly relativistic limiting case, energy and momentum are proportional to each other. This shows that there is a similarity between these quantities, which becomes even clearer when we solve equation (11) according to ${E_0}^2$

$${E_0}^2 = E^2 - kp^2 \ . \tag{14}$$

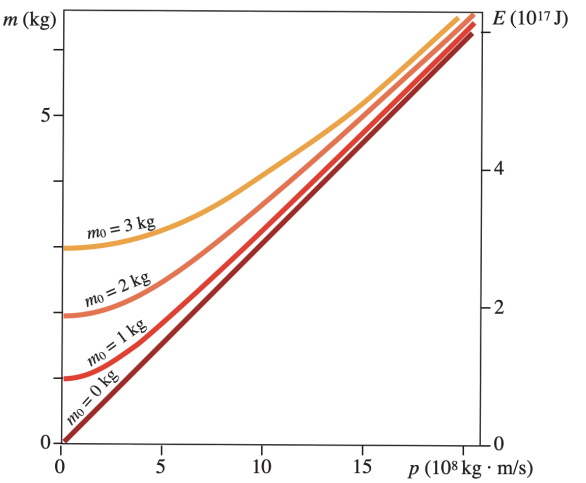


Fig. 2. Energy momentum relationship for four different rest masses. For photons (rest mass zero) the relation is linear.

We thus have the rule: If the momentum of a body changes, its energy changes in such a way that the difference $E^2 - kp^2$ retains its value. This value is the square of the rest energy E_0.

3.2. *The velocity momentum relationship*

Now our second question: How does the velocity of a body depend on its momentum? We solve $p = mv$ for v, then apply equations (1) and (11) and obtain

$$v(p) = \frac{p}{m} = \frac{kp}{E} = \frac{kp}{\sqrt{{E_0}^2 + kp^2}} \,. \tag{15}$$

If we replace the rest energy with the rest mass we get

$$v(p) = \frac{kp}{\sqrt{k^2{m_0}^2 + kp^2}} \,. \tag{16}$$

Figure 3 shows the dependence of the velocity on the momentum for different rest-masses. From equation (16) follows that the velocity of a body approaches a terminal value as the momentum increases. It is

$$\lim_{p\to\infty} v(p) = \lim_{p\to\infty} \frac{kp}{\sqrt{k^2{m_0}^2 + kp^2}} = \sqrt{k} \,. \tag{17}$$

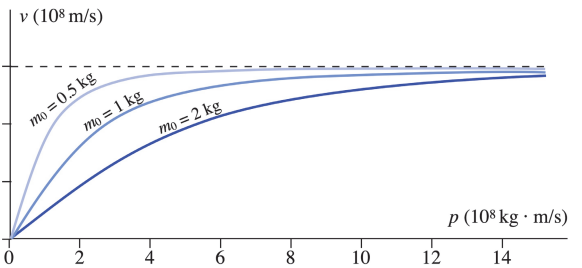


Fig. 3. Dependence of the velocity on the momentum for different rest masses.

Up to now, k only played the role of a conversion factor, but now it gets a physical meaning. Its value is the square of the terminal speed. Since k is a universal constant, its square root, i.e. the terminal speed, is also a universal constant. The terminal speed is the same for all bodies and particles and is independent of the reference frame.

This can be seen in Figure 3. The diagram also shows: the smaller the rest mass of a body is, the "faster" it approaches the terminal speed.

Let us come back to equation (16). We see: If one supplies momentum to a body, its velocity initially increases linearly with the momentum, while its mass almost does not change. This is the Newtonian limiting case. When its momentum has become very large, its velocity no longer changes, but its mass increases.

But what is the value of k and thus the value of the terminal speed? So far, nothing has been said about it. The answer to this question can only be obtained by a measurement. There are several ways to do that: Either one increases the momentum of a particle until its velocity no longer changes (in a particle accelerator)

and then measures its velocity, or one measures the velocity of photons, i.e. particles of rest mass zero. Photons always move with the terminal speed.

Because of the great importance of the terminal speed, one gives it its own symbol

$$c := \sqrt{k} \ . \tag{18}$$

The measurement results in

$$c = 3 \cdot 10^8 \mathrm{m/s} \tag{19}$$

and therefore

$$k = 9 \cdot 10^{16} \mathrm{J/kg} \ . \tag{20}$$

The constant c is also called speed of light. But our derivation shows that light does not play a particular role in special relativity. That is why we prefer to call c terminal speed.

3.3. *Mass and inertia*

From classical physics we are used to consider mass as a measure of inertia. Let us first clarify what is meaningfully understood by inertia.

To determine the inertia of an object, we supply a certain amount of momentum to the object and we look at the resulting change in velocity. The more momentum $\mathrm{d}p$ is needed to achieve a desired change in velocity $\mathrm{d}v$, the greater the inertia.

Therefore we can define the inertia as

$$T := \frac{\mathrm{d}p}{\mathrm{d}v} \ . \tag{21}$$

We first consider a classical motion, i.e. a motion with $v \ll c$. We know the $p - v$ relationship to be

$$p = m \cdot v \ . \tag{22}$$

This results in

$$T = m \ , \tag{23}$$

which is no surprise.

If however the movement is relativistic, i.e. if no longer $v \ll c$, things become more complicated. From equation (15) we obtain

$$p(v) = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{24}$$

and

$$T(v) = \frac{\mathrm{d}p}{\mathrm{d}v} = \frac{m_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{3}{2}}} \ . \tag{25}$$

The inertia now depends on the velocity. It can no longer be described by a single number. By the way, it is also not identical with the so-called relativistic mass.

We know a similar behavior from other contexts. The current-voltage relationship of an ohmic resistor can be characterized by a single number, its resistance. In general, however, the resistive behavior of an electrical component cannot be characterized by a single number. What we need is the $U - I$ characteristic. The situation is like that of inertia. In general, one cannot say that the mass is a measure for the inertia of a body. Rather, the inertial behavior of a body is characterized by a characteristic curve, equation (25). Sometimes the quantity defined by equation (25) is called the longitudinal mass. We think this is rather clumsy. The simple facts are thereby somewhat obscured.

4. Conclusion

The development of a teaching concept for the school, in our case for the secondary school, is a balancing act.

On one hand, teaching at school differs fundamentally from popular science presentations. The latter can limit themselves to showcasing the spectacular, the impressive and the surprising of the scientific results – one can almost say: to exhibit them like objects in a museum.

School teaching has to meet other requirements. The statements must be logically coherent. They have to fit into the previous teaching and form a foundation for the future teaching, for example, at the university.

On the other hand, we must make sure that we do not treat high-school students like university students, that we do not overburden them. Let us not forget: One can calculate and prove without generating understanding.

It should also be borne in mind that most high school students a priori have no particular interest in physics.

We have tried to develop a course under these constraints. We would like to emphasize once again that the above remarks do not represent the content of our course. We have presented only what we believe is different in our approach from that of other textbooks.

References

1. F. Herrmann et al., *The Karlsruhe Physics Course* (2016), `http://www.physikdidaktik.uni-karlsruhe.de/`.
2. F. Herrmann et al., *The Karlsruhe Physics Course* (Guangzhou: Guangdong Education Publishing House , 2018), `http://www.physikdidaktik.uni-karlsruhe.de/`.
3. F. Herrmann, G. Job, *Historical burdens on physics* (2019), `http://www.physikdidaktik.uni-karlsruhe.de/`.
4. A. Einstein, *Zur Elektrodynamik bewegter Körper,* **Vol 10 (322)** , (Annalen der Physik, 1905), pp. 891–921.
5. A. Einstein, *Grundzüge der Relativitätstheorie*, 5th edn. (Berlin: Akademie-Verlag, 1970).

Teaching relativity: Computer-aided modeling

F. Herrmann and M. Pohlig*
Institute for Theoretical Solid State Physics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Baden-Württemberg, Germany
**E-mail: pohlig@kit.edu*
www.kit.edu

Mathematical derivations alone do not necessarily lead to physical understanding. Tools that can replace the mathematical treatment of a physical process and at the same time increase the physical understanding are computer-aided modeling programs, also called system dynamics software. Examples of such software are Stella, Berkeley Madonna, Wensim, Dynasys, Powersim or COACH 7. They solve differential equations and systems of differential equations with numerical methods. One works with a graphical user interface. We want to show how such a software can be used to get from a non-relativistic model to a relativistic model with only minimal modifications. Equating mass and energy alone, ensures that the model provides essential statements of relativistic dynamics: the existence of a terminal velocity for all physical motions, the relativistic dependence of the velocity of a body on its momentum, the relativistic relation between momentum and energy of a body.

Keywords: Teaching relativity, computer-aided modeling, system dynamic software, Coach 7.

1. Introduction

There are two reasons why learning relativity is difficult. First, there is a widespread belief that relativity is essentially a physics of reference frame changes. Second, mathematics, which seems to be indispensable for a first approach to relativity according to common usage, is a major, even often insurmountable, hurdle for many students. Therefore, the structure of our course follows two didactic recommendations: 1. avoid reference frame changes 2. reduce the use of mathematics to the most necessary. We fulfill the first recommendation by teaching dynamics before kinematics. Instead of postulating that the speed of light is invariant under reference frame changes, we declare from the very beginning that energy and mass are the same physical quantity. We call it energy when its value is measured in joules and mass when measured in kilograms. The second recommendation can be fulfilled if we use a suitable modeling software, a so-called system dynamics software (SDS). In a SDS, physical quantities and their relations are described by graphical symbols and are intuitively understandable. Since its usage is self-explanatory to a large extent, we do not need to go into details of the handling here.

As a representative of such a software we use COACH 7.[1]

2. Teaching relativity using a system dynamics software

2.1. *Momentum flow into a body*

We create a model for a body which is accelerated from rest. Its momentum, which is zero at the beginning, increases as a constant momentum current flows into it (as a force is exerted on it). In a first step we treat this model non-relativistically. In the second step, the model is transformed into a relativistic model by a small modification.

In a SDS model, momentum p and other substance-like (extensive) quantities such as electric charge Q, energy E, or entropy S are represented by boxes. All these quantities change their values by an inflow or an outflow. (Entropy can also change its value by being produced). In the model, these currents are represented by thick arrows, Figure 1. Since in our lessons momentum is not introduced as a derived quantity, it gets its own unit of measurement, the Huygens, abbreviated Hy. The unit of its current is Huygens per second (Hy/s). The unit Hy is SI-compatible. We thus have 1Hy/s = 1 N.

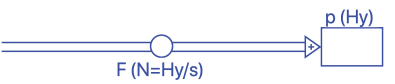


Fig. 1. A temporally constant momentum current F leads to a linear increase of the momentum p.

Since the body is accelerated from rest, its momentum at the beginning of its motion is 0 Hy. We assume the momentum current to be constant in time, we set e.g. $F = 1$ N. When the model is started, the simulation begins running and the actual momentum is calculated for previously defined time steps.

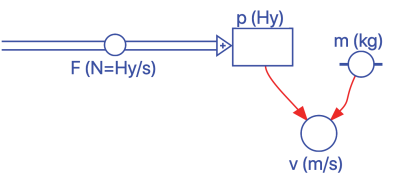


Fig. 2. The velocity is calculated from momentum and mass.

In order for our model to be able to tell us something about the velocity of the body, we must add the mass m of the body to the model, in Figure 2 represented by a circle. We choose $m = 1$ kg for a first simulation and $m = 2$ kg for the second one. The velocity is calculated using $v = p/m$. This relationship is stored in circular symbol for the velocity v. Thin, red arrows are pointing from the symbols for p and m to the symbol of v. This ensures that the actual values of p and m are available

for the calculation of v at any instant of time during the simulation. In concrete terms, this means that while the simulation is running, the momentum is divided by the mass at fixed time steps, and thus the actual velocity v is calculated.

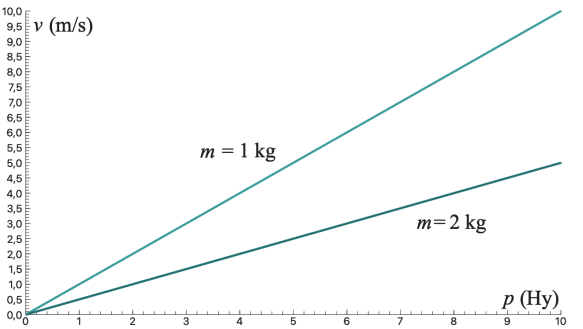


Fig. 3. $v - p$ diagrams or bodies of masses 1kg and 2 kg.

The dependencies of the various variables occurring in the model can be displayed graphically in output windows. Figure 3 shows the v-p diagrams for two bodies of masses 1 kg and 2 kg.

2.2. *Together with momentum energy is flowing into the body*

We now add the kinetic energy of the body to our model, Figure 4.

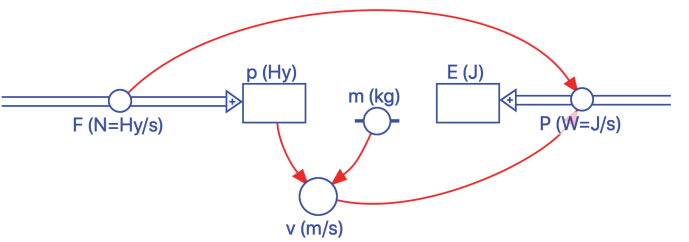


Fig. 4. The energy current P into the body is calculated from the velocity v of the body and the momentum current F (force).

Also the energy E is represented by a box. Since the body is accelerated from rest, not only the momentum but also the kinetic energy must have the initial value $E_{\text{int}} = 0$. Like the momentum current, the energy current (power) into the body is represented by a thick arrow. It can be calculated as $P = vF$. The software provides the energy-momentum diagrams for $m = 1$ kg and $m = 2$ kg, Figure 5.

These graphs could be described by the formula

$$E_{\text{kin}} = \frac{p^2}{2m} \,. \tag{1}$$

However, this equation has not been used by the SDS.

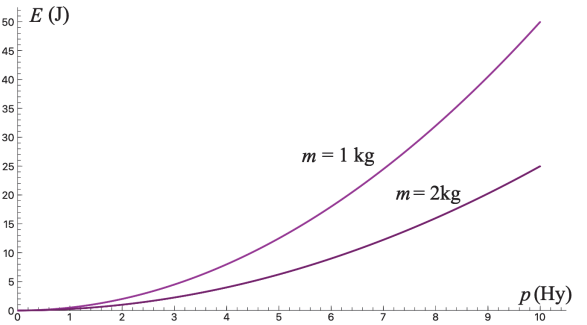


Fig. 5. $E - p$ diagrams generated with the SDS for $m = 1$ kg and $m = 2$ kg.

2.3. *From the classical to the relativistic model*

The two models described up to now serve as basic models for others, such as for "free fall in the gravitational field of the earth", for "falling with friction" and others.

If momentum is replaced by other substance-like quantities such as electric charge or entropy, analogous models are obtained from electricity and thermodynamics, respectively. In this way, students learn to use the modeling software as a tool in other subfields of physics. So when they create relativistic models, they are already familiar with the tool. We now want to create a relativistic model by modifying our non-relativistic model appropriately, Figure 4. For this purpose, only a small modification has to be made. The mass, which was originally a constant,

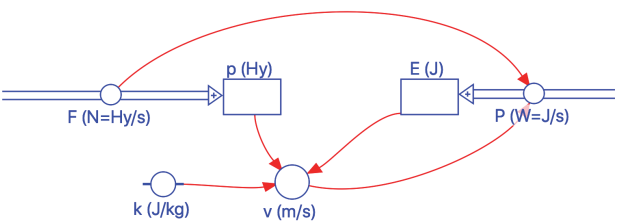


Fig. 6. Relativistic model: The constant mass is replaced by the total energy of the body.

is now identified with the energy[2] according to

$$E = k \cdot m \tag{2}$$

Here, the energy is no longer the kinetic energy but the total energy of the body. Figure 6 shows the modified model. The initial value of the energy is no longer 0 J, but it is equal to the value of the energy that the body has at rest. We choose $E_{\text{int}} = 10$ J and for another simulation $E_{\text{int}} = 5$ J. To transform the unit kilograms into joules, we have to multiply the mass by k. We first set its value arbitrarily to $k = 16$ J/kg. The actual true value of k will be discussed later. The velocity of the body is now calculated according to

$$v = \frac{p}{m} = \frac{kp}{E} \ . \tag{3}$$

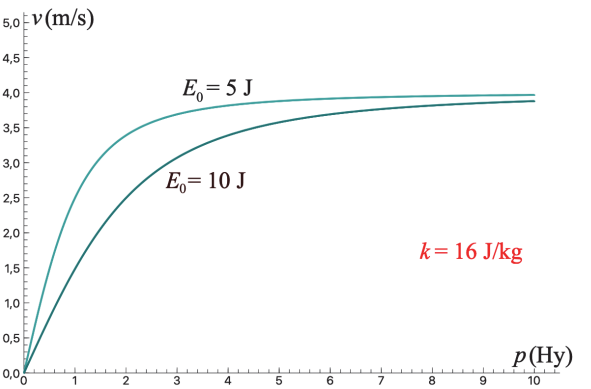


Fig. 7. $v - p$ diagrams obtained with the relativistic model.

As Figure 7 shows, the v-p diagrams are different from those in Figure 3. It can be seen that the velocities of both bodies approach a common terminal velocity of 4 m/s. One "plays" with further, freely chosen values for the rest energy and always finds the same terminal velocity. Furthermore, the v-p diagrams show that light bodies reach the terminal velocity quicker than heavy ones. Finally, one recognizes that for sufficiently small momentum values the velocity of a body increases linearly with momentum. This confirms what is already known from non-relativistic mechanics.

The diagrams are graphs of the relation

$$v(p) = \frac{p}{m} = \frac{kp}{E} = \frac{kp}{\sqrt{{E_0}^2 + kp^2}} \ . \tag{4}$$

However, we don't need to know this equation, or to derive it, in order to be able to read the important properties of this relationship from the diagrams.

The diagrams show that for relativistic motions the following rules hold:

- small momentum values: $v \sim p$
- large momentum values: $v = \sqrt{k}$

By choosing other initial values for the energy and other values for k, one easily gets convinced of the generality of these rules. For the diagrams in Figure 8, the values 9 J/kg, 16 J/kg and 25 J/kg were selected for k. The terminal velocity results to be 3 m/s, 4 m/s and 5 m/s, respectively.

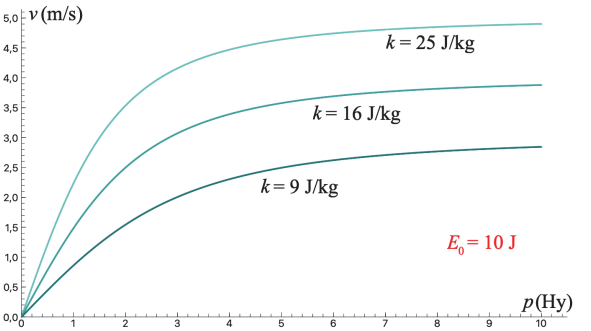


Fig. 8. $v - p$ diagrams for different k values.

We see that the value of the terminal velocity is just the square root of k. While k used to be merely a conversion factor between kilogram and joule, it now acquires a physical meaning. Moreover, from the fact that a conversion factor for units is universal, the terminal velocity also has a universal value. It must be a universal constant. For this reason, we give the square root of k, i.e. the terminal velocity, a symbol of its own, namely

$$c := \sqrt{k} \tag{5}$$

Up to now, we had chosen the value of k and thus also that of c arbitrarily. But which value has nature given to k resp. c? Experiments show that $k = 9 \cdot 10^{16}$ J/kg and therefore $\sqrt{k} = c = 3 \cdot 10^8$ m/s.

Another relationship that our model provides in the form of a diagram is that between energy and momentum, Figure 9. The diagrams shown are graphs of the equation:

$$E(p) = \sqrt{{E_0}^2 + kp^2} = \sqrt{{E_0}^2 + c^2 p^2} \tag{6}$$

For small momentum values we get the non-relativistic relation between kinetic energy and momentum, shifted by the rest energy of the body:

$$E(p) = \frac{p^2}{2m} + E_0 \tag{7}$$

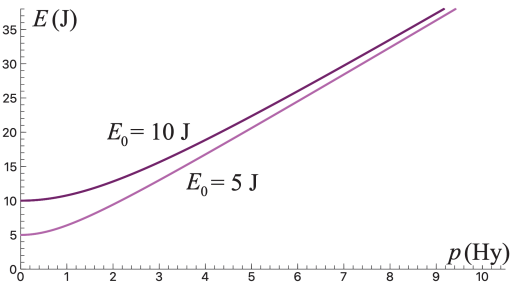


Fig. 9. $E - p$ relation in the relativistic model. The rest energy was set to 5 J and in the second run to 10 J.

Moreover, we see that the ratio of energy and momentum is the same for large values of the momentum.

$$E(p) = \sqrt{k} \cdot p = c \cdot p \tag{8}$$

3. Conclusion

A non-relativistic SDS model which describes the behavior of a body whose momentum increases linearly, i.e. into which a constant momentum current is flowing, becomes a relativistic model by a small modification. The change consists in equating the quantities mass and energy. The new model then provides diagrams that are well-known from relativistic physics. The software provides them without using the respective equations. From these diagrams, important results of relativity can be interpreted and understood. The use of computer-aided modeling in teaching has the advantage of getting rid of the mathematical ballast and that one can concentrate on the physical content. Our experience shows that even younger students can easily learn a software like COACH 7 and use it to work out challenging results in relativistic dynamics.

References

1. Coach 7 for Desktop and Tablet: CMA (Centre for Microcomputer Application `https://cma-science.nl/downloads_en` accessed July 26 2021.
2. Herrmann, F. et al., *The Karlsruhe Physics Course: Mechanics: for the upper secondary school*, (2019). `http://www.physikdidaktik.uni-karlsruhe.de/index_en.html` accessed August 2021.

Solstices and Equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, and the stellar aberration of Sirius

Costantino Sigismondi

ICRA, Sapienza University of Rome, Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

Silvia Pietroni

Dipartimento di Fisica "E.R. Caianiello", Università di Salerno
Via Giovanni Paolo II 132, I-84084 Fisciano, Italy
Istituto Nazionale di Fisica Nucleare, Sezione di Napoli
Via Cintia, 80126, Napoli, Italy
E-mail: spietroni@unisa.it

The 1703 was the first year of full operation of the meridian line in the Basilica of St. Maria degli Angeli in Rome. The instants of solstices and equinoxes, the *Anni Cardines*, obtained by comparing transit timings of Sun and Sirius, also in daytime, are affected either by the East deviation of the meridian line of about 5' Eastward (geometrical effect), either by the stellar aberration of Sirius (relativistic effect). Similarly the seasonal shifts of Sirius' declination observed by Bianchini are here firstly recognized as depending on the stellar aberration in ecliptic latitude. The eccentricity of the Earth's orbit and the orientation of its axis in the space can be measured, since Ptolemy, by knowing the exact timing of the solstices and the equinoxes. The dates of 1703 equinoxes and solstices have been published by Francesco Bianchini in local roman mean time, referred to the local (roman) mean noon, i.e. after the solar meridian mean transit time. By using the observations of the three lunar eclipses of 1703 we found the equation of time for that year, and the UTC corresponding timings of the beginning of the seasons. This operation lead to find the contribution of the Sirius aberration to the instant calculated by Bianchini with respect to the current celestial mechanical models of IMCCE.

Keywords: St. Maria degli Angeli, Meridian Line, Stellar Aberration, Sirius, 1703, Solstices, Equinoxes, Ecliptic Longitude, UTC, UT1, TDT, TT, Roman mean time.

1. Stellar aberration measures in 1701 before Bradley's discovery

The stellar aberration was discovered by James Bradley in 1727, and the same astronomer discovered the nutation of the Earth's axis in 1737. At the meridian line of St. Maria degli Angeli, the giant Clementine Gnomon, built by astronomer Francesco Bianchini (1662-1729) and funded by the Cardinal Gianfrancesco Albani (1649-1721), elected pope on 23 November 1700 with the name of Clement XI.

The effects of stellar aberration have been detected on the Polaris (2006), and on Sirius (2021, present work).

The aberration of the Polaris influenced the measure of the latitude of the pinhole, the gnomon of the meridian line, made by Bianchini on January 1-8, 1701.

The aberration on Sirius affected the instant of the equinoxes and the solstices, calculated with the difference between solar and stellar meridian transit, the latter observed also in full daylight. To obtain such evidence of the first special relativistic effect observable, a complete calibration of the meridian line has been carried since 2018. With the IGEA observational campaign, Informatized Geometric Ephemerides for Astrometry: all the reference points present of the 45 meters long meridian line have been calibrated, by comparing solar observations and ephemerides.

At the Marcel Grossmann Meeting XI (Berlin, 2006) the evidences of Polaris' aberration were discovered in the latitude, from the measures by Bianchini at the Gnomon, and now at the Marcel Grossmann XVI edition we can afford the evidences on Sirius' stellar aberration with the increased precision on the calibration of this giant instrument, the second *Heliometer*, or solar telescope, of its times, after the one of Cassini in St. Petronio, Bologna.

2. Managing dates and timing of 1703, in local mean time

The equation of time shifts the instants of the meridian transit day by day with respect to an absolute time reference such as UTC. The eclipses' ephemerides are based upon the terrestrial dynamical time TDT, while the terrestrial rotational time UT1 includes the slowing down effect of the Earth's rotation in act since 1700. To translate 1703 local time to modern UTC to the nearest second we exploited the observations of the three lunar eclipses occurred in 1703, all made by Bianchini. The equation of time of that year come out from the comparison between the observed instants of eclipse's beginning and end -in local mean time- and the corresponding instants in TDT computed by the NASA ephemerides.

3. The dates and timing for 1703 lunar eclipses from NASA ephemerides

The three lunar eclipses of 1703 have been observed by Francesco Bianchini and published in the same book (Bianchini, 1703) where the great meridian line, the *Clementine Gnomon* was presented. The instants of beginning and end of lunar eclipses are independent of the position of the observers on the Earth, and they can be calculated with a suitable celestial mechanical model of the lunar orbit (provided by NASA). They occurred very close to the solstices, making the procedure of their absolute timing (expressed in TDT and then in UTC) particularly accurate.

The maximum phase of the eclipse is indicated in TD, terrestrial dynamical time to the nearest minute in the figures, and to second accuracy in the website. The maximum of the eclipse is recovered in local time after doing the average between starting time and ending time of the eclipse, to the nearest second. In the details of the *Canon of five millennia of lunar eclipses*, also the starting time and the ending time of the eclipses are provided. The eclipses have been observed by Bianchini at the telescope, improving the accuracy of the measurements.

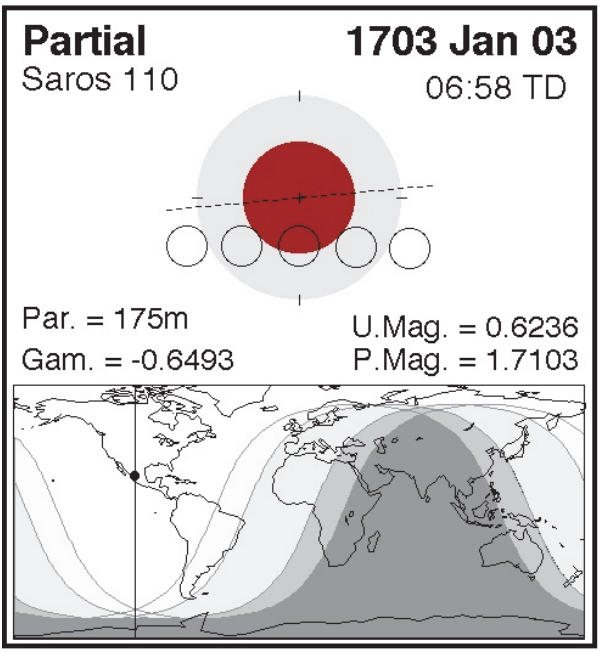


Fig. 1. The lunar eclipse of January 3^{rd} 1703

4. Solstices and equinoxes and the equation of time in the ephemerides

The dynamical model for the Earth's orbit includes the perturbations of all planets, and they are implemented in the most advanced ephemerides. IMCCE provides an online service back to 4000 B. C. for the dates of the astronomical beginning of the seasons. The comparison with the corresponding instant of 1703 as calculated by Bianchini required a conversion into UTC of the local timings. It is available online a service, *Planetcalc*, with the equations of time, from 500 B. C.

We need a self-consistent derivation of the UTC timing of the equinoxes and the solstices in 1703, being the Equation of Time itself depending on the Earth's orbit eccentricity, which is derived from the the observed beginning of the seasons, namely the equinoxes and solstices.

The obliquity component goes to zero four times per year at the equinoxes and the solstices, while the eccentricity component vanishes twice per year at the apsides (aphelion and perihelion). This is well shown in *Planetcalc* website.

5. Synchronization of the equation of time with lunar eclipses in 1703

The dates of the apsides (as shown in Sigismondi and Pietroni, 2019) and the obliquity of the year are used to compute the equation of time for that year,

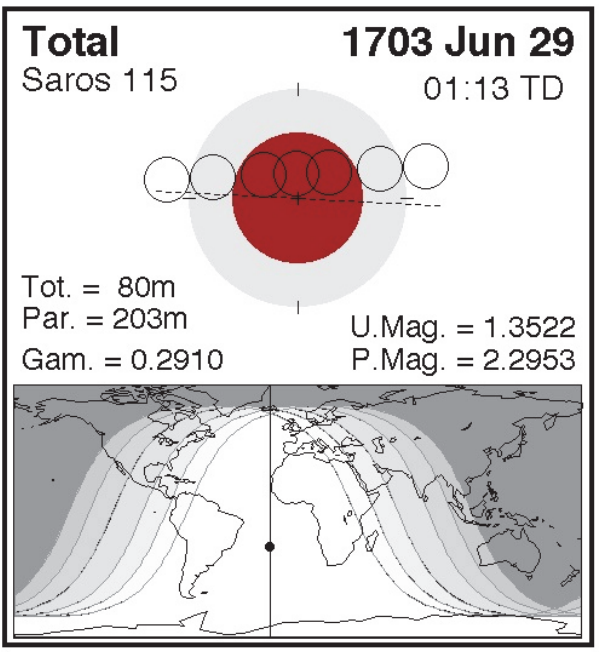


Fig. 2. The lunar eclipse of June 29^{th} 1703

but an independent synchronization with UTC is needed, to verify the IMCCE ephemerides, and the lunar eclipses were used exactly to assure that synchronization. On June 22^{nd} 1703 the equation of time of Planetcalc is -1.11 minutes, and in 28^{th} June, last solar transit before the lunar eclipse, it is -2.38 minutes. It means that this local noon should occur 2 m 22 s before a mean solar noon fixed (e.g. CET Central European Time). For Rome the mean noon is at 12:10 CET, then 12:12:22 the transit time for 28^{th} June, 1703. The eclipse was computed as maximum at 1:13 TD on 29^{th} June, so 2:13 CET. Bianchini's beginning of totality is at 13:18:15 after solar meridian transit of 28^{th} June, and the end (emersion) is 14:38:15. The middle of totality (maximum phase) is at 13:28:15, obtained by averaging. The solar transit in the same publication occurred at 12:07:36 so that the eclipse's maximum was at 02:05:51. The terrestrial dynamical time for the eclipse has been computed (Espenak and Meeus, 2009) on 02:12:47 (1:13 TD in figure 1). The TD, is ahead with respect to Bianchini's roman mean time of 6 min 56 s.

6. Including $\Delta UT1$ in the final timing

The procedure shown in detail for June 29^{th} has been applied to the other eclipses of 1703, that Bianchini observed also with Filippo Maraldi (January 3^{rd}) and alone (December 23^{rd}), recovering the Roman average time adopted by Bianchini for his observations, with respect to the Terrestrial Dynamical Time. The adopted value

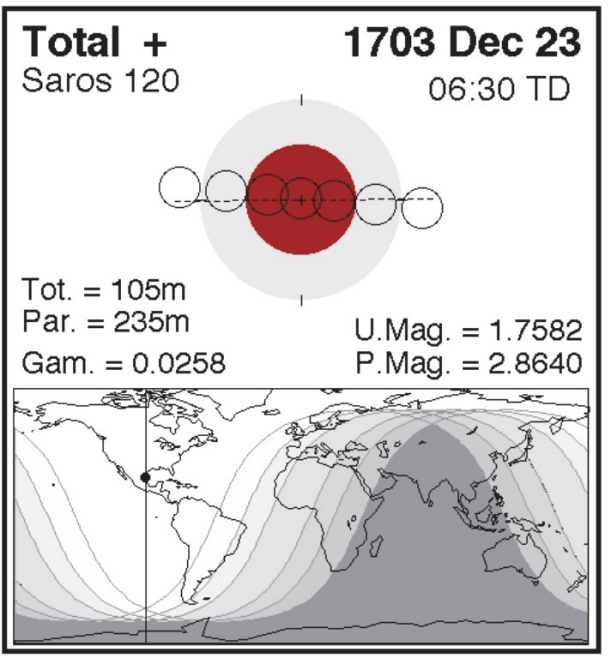


Fig. 3. The lunar eclipse of December 23^{rd} 1703

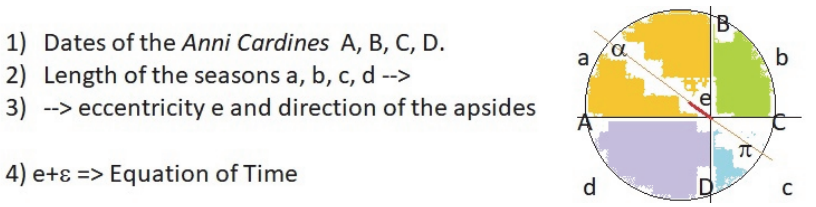


Fig. 4. The connection between the beginning of seasons' times and the eccentricity, and the equation of time. Aphelion α and perihelion π, the apsides, are also sketched.

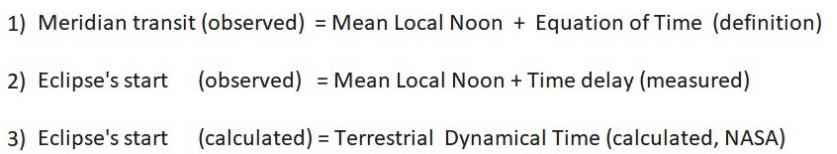


Fig. 5. The relations existing between Mean Noon, Equation of time, solar transit time and terrestrial dynamical time TD.

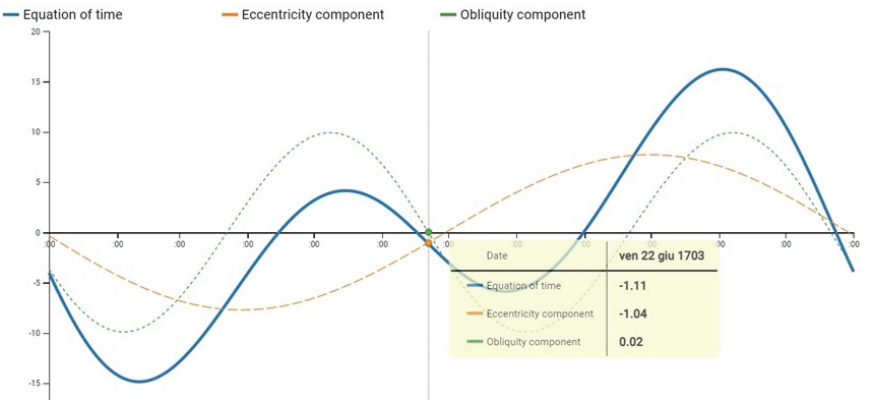


Fig. 6. The equation of time for 1703, as computed with Planetcalc.

for ΔUT1, of 7-8 seconds for the 1703, and due to the de-rotation effect of the Earth upon continental uplift after the ice age, is added to TD to obtain the coordinate universal time UTC for these observations, and for the dates and timing of solstices and equinoxes in 1703. These instants of seasons' beginning dates have been used for studying the aberration effect on the solstices and equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, demonstrating that Sirius' aberration and the deviation from Celestial North of the meridian line have caused the departure from IMCCE calculated times for 1703 seasons' starts (Sigismondi, 2021).

7. The dates and timing for 1703 *anni cardines*

Bianchini used the term *Anni Cardines* for the solstices and the equinoxes, because they are related to:

- Inclination of the Earth's axis on the ecliptic plane (the plane of the orbit), called *obliquity*;
- Position of the apsides in the Earth's orbit, which determines the *eccentricity*.

They were published in the book *De Nummo et Gnomone Clementino* of 1703, as well as on a marble epigraph in the presbytherium.

These instants are expressed in mean local time, after/before the mean meridian transit or midnight (anti-)transit. So to recover their instants as expressed in UTC it was necessary to recover three lunar eclipses observed in the same year and published by Bianchini in the same book (Bianchini, 1703): January 3, June 29 and December 23 (see below). After this operation of synchronization, we made the comparison with the IMCCE ephemerides for 1703.

Fig. 7. The epigraph of 1703 *Anni Cardines* in the presbytherium of St. Maria degli Angeli in Rome.

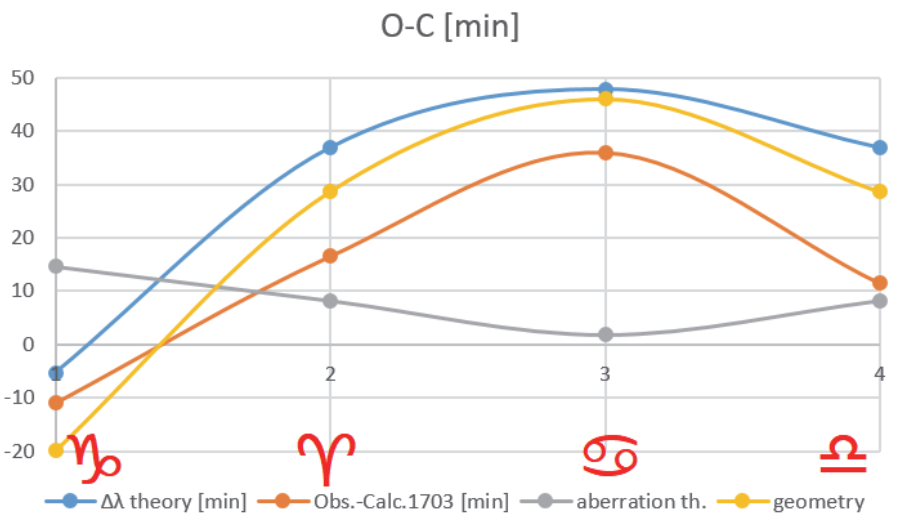


Fig. 8. The epigraph's timings translated into UTC and compared with IMCCE ephemerides.

8. Sirius' aberration and the East deviation of the line

Before the present analysis it was not clear that Bianchini used only Sirius for timing the solar meridian transits, also in daytime. The reason can be practical: Sirius is the brightest star in the sky and it can be seen in full daylight with a telescope through the open window 60 cm × 40 cm located over the pinhole. I made personally observations of meridian transits of Antares, Ras Alhague and Venus in daytime, outdoor at the meridian line of the Vatican obelisk, and the brightness of the star is crucial. Nevertheless on the meridian line of St. Maria degli Angeli there are reported 22 stars, with their ecliptic coordinates of 1701. This lead us to intend that all the 22 stars were used during the measurements (Bedinsky and Nastasi,

2007), while only Sirius was certainly used, after the examination of the documents of Bianchini.

The *Anni Cardines* are with ecliptic longitude 0°, 90°, 180°, 270°.

Sirius in 1701 had ecliptic longitude 97°57′53" (fig. 3), and it was supposed to change only for the precession, until James Bradley discovered the aberration in 1727.

9. Seasonal aberration in ecliptic latitude

Francesco Bianchini (1703) considered also the meridian shift of Sirius, as of seasonal and of meteorological cause. Here I have demonstrated that it was due to the stellar aberration in ecliptic latitude (being Sirius not on the ecliptic plane but nearly at −40°).

Fig. 9. The coordinates of Sirius in 1701 from the Stellar Atlas of Philippe de la Hire, as reported on the meridian line of St. Maria degli Angeli.

10. Discussion

The measurements reported by Bianchini on the epigraph are plotted in orange, and they are in very good agreement with the blue line which includes the effects of the line's deviation (geometry in yellow) and of the stellar aberration in grey. The geometrical effect is much larger, nevertheless the aberration's contribution matches exactly as needed, especially in the cases of opposition (Sun in Capricorn) and conjunction (Sun in Cancer) with the star. The accuracy of the method applied

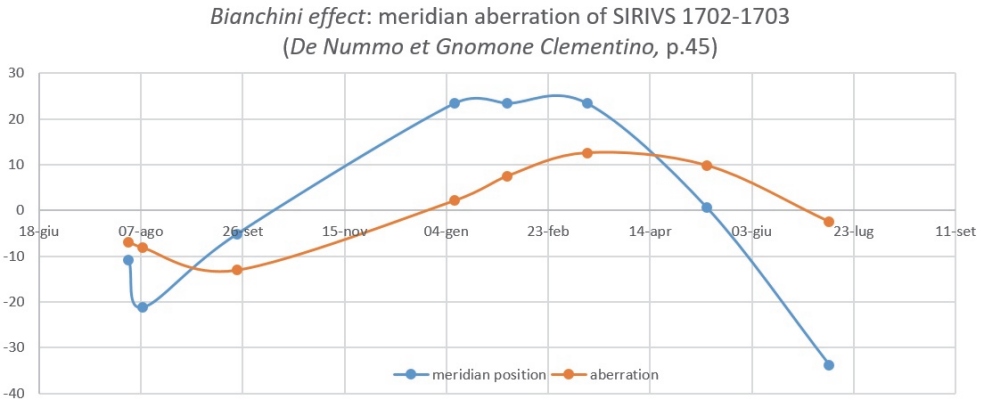


Fig. 10. The meridian position of Sirius along the year. Bianchini thought it was a meteorological effect, while it is due to aberration in ecliptic latitude.

at the Clementine meridian line is enough to recover the *Anni Cardines* within 10 minutes, as Bianchini declared in his book. The departure from this value, up to 30 minutes, is due to a systematic error (the deviation of the whole line toward East of 5.2', or 70 mm at its Northern extreme, 45 meters from the pinhole's vertical) combined with the special relativistic effect of the stellar aberration of Sirius, used in all measurements.

11. Conclusions

An important contribution of this paper is the verification that Sirius was the star always used in the meridian measurements, to compute the ecliptic coordinates of the Sun with respect to a fixed star. Finally the effect of meridian shift of Sirius along the year, was clearly recognized by Bianchini as seasonal, and it is, not because of seasonal humidity variations, but of special relativistic origin: due to the stellar aberration in ecliptic latitude, which is a rather big effect being Sirius of ecliptic latitude $b = -40°$.

The operation of synchronizing the mean solar time of Rome in 1703 with the Terrestrial Dynamical Time and then with the Universal Time Coordinated, has made possible recovering the Equation of Time of 1703 and the comparison between modern ephemerides and these data, to distinguish clearly the stellar aberration's contribution of Sirius, from the one due to the Eastward deviation of the line.

References

1. Bianchini, Francesco, Roma (1703), *De Nummo et Gnomone Clementino*, `https://archive.org/details/bub_gb_IYtGjIE_X64C`.
2. Bradley, James, (1727), *A New Discovered Motion of the Fix'd Stars*, `https://www.if.ufrj.br/~marta/caronte-hipertexto/movimentosdaterra/bradley.pdf`.

3. Sigismondi, Costantino, Proc. of MG X (2008), https://ui.adsabs.harvard.edu/abs/2008mgm..conf.2470S/abstract.
4. Bedinsky, Magdalena, Nastasi, Monica, (2007), *Le Stelle e la Meridiana*, http://www.santamariadegliangeliroma.it/paginamastersing.html?codice_url=posizione_delle_stelle&ramo_home=La_Meridiana.
5. Sigismondi, Costantino, *Earth's Obliquity and Stellar Aberration Detected at the Clementine Gnomon (Rome, 1703)* (2021), https://www.mdpi.com/2673-9984/2/1/49.
6. Bianchini, Francesco (1703), *Lunar eclipse of 29 June 1703 observed with Gabriele Manfredi*, https://archive.org/details/bub_gb_IYtGjIE_X64C/page/n325/mode/2upl
7. Espenak, Fred & Meeus, Jaan (2009), *Five Millennium Canon of Lunar Eclipses*, https://eclipse.gsfc.nasa.gov/SEpubs/5MCLE.html
8. IMCCE, Paris (2021), *DATES DES SAISONS POUR UNE ANNEE DONNEE*, https://promenade.imcce.fr/fr/pages4/439.html
9. Planetcalc LLC, Moscow (2021), *PLANETCALC Online calculators*, https://planetcalc.com/9198/.
10. Sigismondi, Costantino, *Earth's Obliquity and Stellar Aberration Detected at the Clementine Gnomon (Rome, 1703)* (2021), https://www.mdpi.com/2673-9984/2/1/49.
11. Sigismondi, Costantino and Silvia Pietroni, *L'eccentricità ed apogeo dell'orbita della Terra misurata a Santa Maria degli Angeli nel 1703*, Gerbertus **12**, 115 (2019). http://www.icra.it/gerbertus/2019/Gerb-12-2019-Sigismondi-Pietroni-eccen-115-120.pdf.

Positional astrometry at arcsecond accuracy using historical instruments, with light equipment

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

Lorenzo Ricciardi
University of Rome 3,
Rome, Italy

It is possible to make solar astrometry at arcsecond level using a meter, a watch and a smartphone at the Clementine Gnomon (1702, a giant pinhole-camera obscura). Similarly at the Vatican obelisk's meridian line (1817) the transit timing's accuracy is better than one second with the same instruments, and a small monocular. The differential positions with respect to the zodiacal marble disks allow to reach the arcsecond accuracy even outdoor.

Keywords: Video-timing, Equation of time, Spherical Astronomy, St. Maria degli Angeli, Vatican Obelisk, Meridian Line, Solstices, Equinoxes, Zodiacal signs, UTC, Ephemerides, Altazimuthal coordinates.

1. The giant pinhole camera obscura of the Clementine Gnomon (1702)

Measures of solar astrometry at arcsecond level have been done with a pencil, a meter, a watch and a smartphone at the Clementine Gnomon (1702). The Clementine meridian line is 45 meter long. The new calibration tables of the reference points on the meridian line in the Basilica of St. Maria degli Angeli, have been recently published (2021). Each measure of the solar image position is referred to its nearby marker with known position with respect to the pinhole, within ± 0.5 mm. The solar images limbs can be located on the floor within ± 1 mm. 2 arcseconds of angular position accuracy is obtained for winter observations with focal length of nearly 50 meters, while 5 arcsec is reached in summer with 21 m length involved. The pencil is used for signing the limbs' positions on the meridian line, the meter is sufficient to measure the distances of these limbs from the nearest reference point (there are 42 of them along the line), the watch synchronized with UTC is video recorded along with the solar transit to get the single frame accuracy on contact's times. A timing accuracy down to 0.3 s is reached with the average air turbulence conditions. This result is observed by N parallel transits averaged.

Examples on application of Cassini's and Laplace's corrections for air refraction on the observed data to obtain the real celestial coordinates of the date are presented (Sigismondi, 2019). This research was aimed to understand the Bianchini's effect, claimed in the book De Nummo et Gnomone Clementino (Bianchini, 1703). According to his observations Sirius changed its meridian position during the seasons because of some meteorological effect. We proved that the relativistic stellar aberration was the cause, thanks to this whole series of accurate observations of the Sun during all kind of weather conditions. This historical instrument gives the possibility to see the arcsecond, as half of a millimetre during winter, as no other instrument in the World, at the same time to many people. A set of more than 100 transits video is also presented for remote observations during all year's seasons.

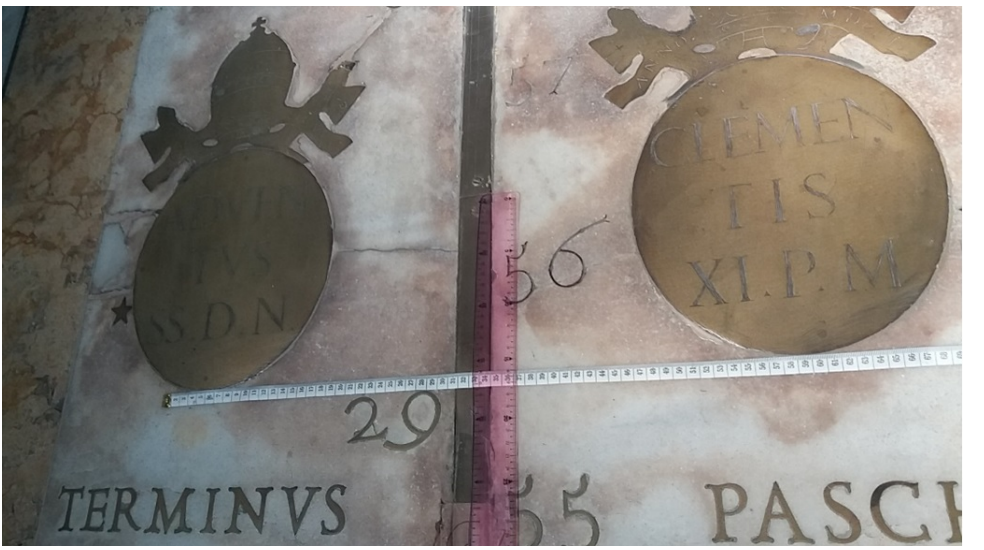


Fig. 1. The ellipses of the Sun on 20 August 1702 on the meridian line of St. Maria degli Angeli. They are an example of measurement of the solar image on the floor with respect to the references on the meridian line [e.g. the number 55's marker].

2. The outdoor Vatican obelisk's meridian line (1817)

At the Vatican obelisk's meridian line solar positions as well as stellar and planetary transits have been measured. The obelisk of St. Peter's square has a meridian line since 1817. This line is two times longer than the Clementine line, and it allows potentially accurate measurements, despite the low contrast obtainable outdoor.

A scheme of the correspondence between the features on the top of the obelisk and the marble zodiacal disks on the square is published in figure 2.

The limit of the penumbra of the obelisk's shadow has been recovered with an accuracy of 2 mm during the Summer solstices of 2020 and 2021, repeatedly, by

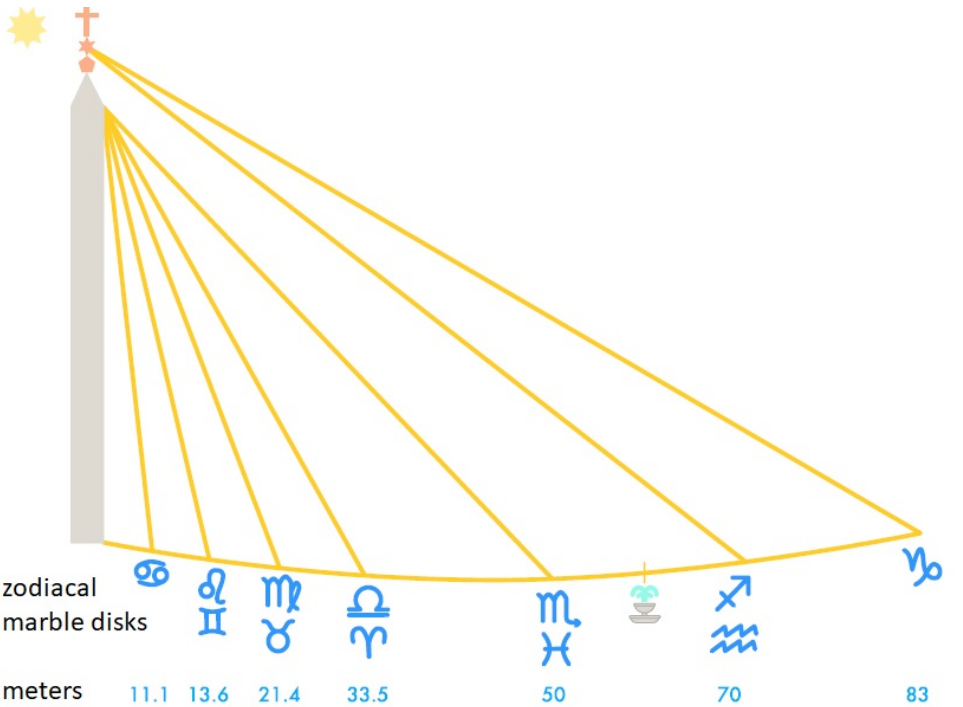


Fig. 2. The zodiacal signs on the meridian line of St. Peter's square as connected with the top of the obelisk. The distances from the axis of the obelisk are reported in meters from all the zodiacal marble disks.

moving the smartphone used for the video making, up and down near the limit and observing its cast shadow. The shadow's limits measured during 20 minutes of observations (fig. 3) on the Cancer and Leo's disks showed the rotation of the obelisk with respect to the East-West direction of 24 mm around its axis. This is the first indication on the astronomical accuracy of the re-collocation of the obelisk in September 1586 (Sigismondi, 2016). Observing the shadow's penumbra the massive presence of Sahara's sand in the upper roman atmosphere has been detectable. The contrast of the shadow almost vanished at noon time during the most sandy days. These observations were confirmed by the aerobiological measurements made in Tor Vergata University in the same period, and by the following presence of dry sand all over the city (Travaglini, 2021).

Moreover the projection of the image of the Sun when being occulted by the obelisk, obtained by a monocular 7×18, on the marble zodiacal disks (or on movable cardboard projection screen) has been video recorded during the meridian transits, attaining a timing accuracy of 1 second. The 18 mm-lens telescope projects on the marble a sharp image of the Sun, of about 30-45 mm of length, screened by the direct sunlight (fig. 4).

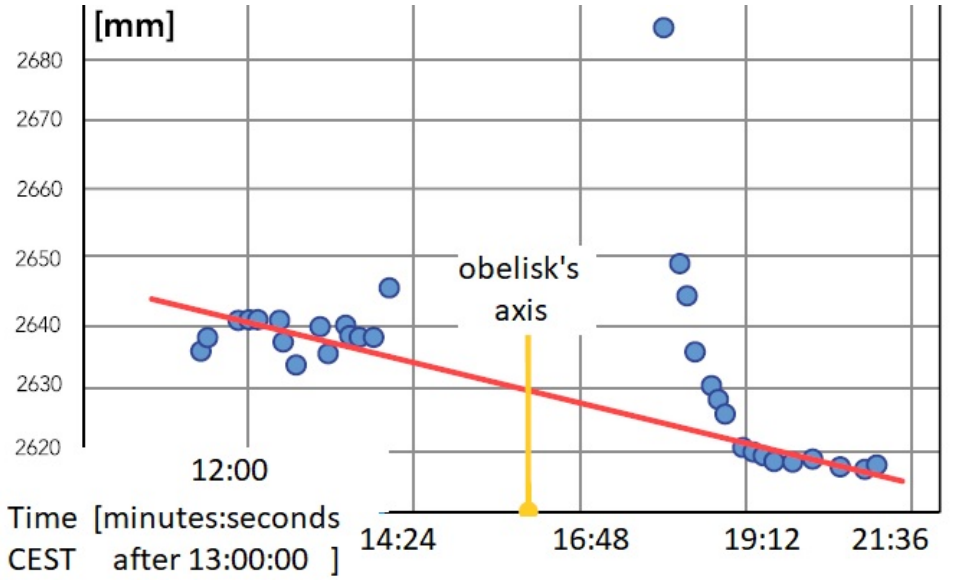


Fig. 3. The rotation of the obelisk with respect to East-West direction determines the variation of the shadow's length detected around the meridian time (24 July 2021) on the Leo disk, measured from the Cancer's black star.

2.1. *Parallax error*

The crucial problem of the all direct observations (with or without telescope) is the on-axis location of the lens or of the eye. It is a parallax error.

A mm shift from the meridian axis, with 4 mm/s velocity of the shadow corresponds to 0.25 s of timing shift. From the summer to the winter solstice the ground velocity of the shadow ranges from 2 to 6 mm/s, making the corresponding accuracy on the timing ranging from 0.5 to 0.2 s per mm.

To obtain a timing absolute accuracy ranging from 1 s to 3 s the parallax problem has to be solved. The solar shadow have not this problem, but their contrast makes the accuracy ±3 s. Using a telescope to project the eclipses on the marble disks we improved the accuracy to ±0.5s. The naked eye measurements of the stellar and planetary alignments while standing on the meridian line show an accuracy of ±1, reduced to ±0.5 s by using a ground mirror posed exactly in the middle of the meridian line and aiming at it. and the planetary/stellar transits observed with the mini-telescope standing on the meridian line. Stellar transits with the naked eye at the monocular are easier to be seen and audio-recorded because of the strong illumination of the square by night.

The achievable accuracy on many measurements from the same point, as for Antares' transits on the Capricorn disk, has been ±10 arcsec on the central azimuth of the meridian transit behind the obelisk. The first video record of a full daytime transit (October 28, 2021 of Venus) confirms that result within a σ, as well as the transit of Fomalhaut observed on November 4, 2021 from the same disk.

Fig. 4. The approach of the Sun to Scorpio black star in 2021 on the marble disk in S. Peter's square: the three dimensional star on top of the obelisk is projected next to the black star. That black star is 4 cm wide.

3. Solstices, zodiacal signs and equinoxes

In the solstices the solar altitude above the horizon reach a maximum (June) or a minimum (December) well represented by a parabola. In St. Maria degli Angeli the influence of the aphelion (4-6 July) and of the perihelion (2-4 January) on the measured altitude of the Sun is well visible as the minimum/maximum's shift of the fitting parabola on dates symmetrical to the solstice and more and more widely spaced. For St. Peter's meridian line the approach to Scorpio zodiacal sign shows

already well the change in solar meridian altitude day by day, projected on the marble disk (fig. 5).

Fig. 5. The approach of the Sun to Scorpio from 16 to 23 October 2021, looking at the obelisk of St. Peter's from the Scorpio's black star on the marble disk. The solar profile is reported with light yellow circles.

For a more precise measurement of the winter solstice a solid metal holder has been prepared for the mini-telescope. The principle is the same of the tripod with fixed angle, with a larger stability to guarantee high precision measurements, with the image shifting of $4 - 5\%$ of its length of 40-50 mm during the solstice's days.

4. Equinoxes and linear interpolation

The parabolae for the solstices and the linear interpolation for all zodiacal signs' ingresses are required to fit the observed data. This is a very good training for data analysis practice.

The observations of the equinoxes require the accurate knowledge of the Gnomon's height and geographical coordinates, and two observations: one before and another after the equinox, are necessary to have the possibility of interpolation.

Fig. 6. The fixed-angle monocular-holder built as Winter solstice meter.

These measurements confirmed the accuracy of ±10 minutes claimed by Bianchini, provided the systematic deviation of the line from the Celestial North. Similar result is obtained at St. Peter's square meridian line.

5. Level function of the mini-telescope at fixed angle

The mini-telescope with a fixed inclination can be used to test the horizontality of the seven disks. The Virgo's one shows a larger inclination because there the meridian line is reaching its minimum level in the square, before the equinox and toward the Scorpio's disk. The St. Peter's square is concave, allowing to convey the rainwater to the sewage system.

6. Special measurements

The meridian line of St. Maria degli Angeli allows to measure the position of the Sun with a great precision. The solstices can be measured also off-meridian as differential measures, always from the same reference point on the floor of the Basilica (e.g. the Summer solstice "royal" marker along the 173° azimut). The precision of 1 mm yields an angular accuracy in the range of a few arcseconds. The zodiacal marble disk of St. Peter's square meridian line have been used as fixed references, to make differential measurements of meridian transits. These differential measures

Fig. 7. The measurements of the Fall equinox on September 23, 2019 in the Basilica of St. Maria degli Angeli.

are used to recover their relative alignment with North and the obelisk's axis, and the alignment between the Sun entering the sign, the black star of the disk (see fig. 4) and the ending top features of the obelisk. The accuracy here is within 2 mm, as the one attained with the penumbral limit measured directly (fig. 3).

The measures of planetary and stellar transits have been:

- Antares from July 21 to August 21 (before sunset), Capricorn
- Ras Alhague, August 21-September 6, Virgo
- Fomalhaut, November 4 2021, Capricorn
- Saturn, Sagittarius, October 2021
- Venus in full daytime, 19-28 October 2021, Capricorn
- Jupiter, October 2021, Scorpio
- δ Capricorni, October 2021, in Scorpio
- μ and ϵ Aquarii, respectively of magnitude 4.91 and 3.96 under an unusual clear sky for Rome on October 11, 2021.

Fig. 8. The measurements of the Fall equinox on September 23, 2021 in St. Peter's square, on the meridian line. The solar image is the ellipse on the intersection of the two meters-tapes.

The lunar transits have been also recorded, and the lunar limb position precisely located. The planetary transits timing with a precision of a second allow to detect directly the nature of their motion (direct or retrograde) by comparing with the previous transit and the sidereal day length. The stellar and planetary transit paralleled the solar ones to measure the local western deviations of the meridian line on all the zodiacal disks. In particular Antares, Venus and Fomalhaut were used for the last point in the Capricorn solstice. The stellar transits allow to measure the sidereal day and the stellar aberration with a 0.01 s precision obtained over several days, either with naked eye video recorded measures (Sigismondi, 2021, these Proceedings).

6.1. *Distances on the meridian line in St. Peter's square*

The accuracy on the measures made on the floor of the square would require the use of metal meters at least 50 m long, or total stations, but due to security requirements in the square this is not possible. So we adopted a technique used in Athletics Track & Field to measure the long jump run-up length with the feet. The length of 10 adjacent feet with shoes is accurately measured to find its own unit with a mm accuracy. Distances up to 300 feet are accurate within 1%, as we could verify from the center of the colonnade of St. Peter's square, by measuring the distances of some columns from their centre.

Fig. 9. The mini-telescope and its tripode, used also as level-meter. The image of the Sun is evidenced inside the cloudlet.

7. The marble zodiacal disks and the Gregorian cycle of four centuries

There is a 4-years periodicity in the ingress' dates into the zodiacal signs, due to the Julian leap year. The 29th of February resets almost completely the 5 h 49 m 11 s of average excess of the tropical year versus the civil one of 365 days. But only over 4 centuries the Gregorian reformation of the calendar is complete. Then the sequence of the dates represented in figure 10 with the addition of 22 December, Solstice in Capricorn (fig. 6, logo), is the result of a rounding to the day of these instants. The complete sequence corresponding to the dates represented on the marble disks happened in 1815. Here the procedure applied to the IMCCE calculated dates:

- Spring in Aries 21 March 1815 11h31m UT become 21 March
- Summer in Cancer 22 June 1815 8h59m UT become 22 June
- Autumn in Libra 23 September 1815 à 22h43m UT become 23 September
- Winter in Capricorn 22 December 1815 à 15h44m UT become 22 December

This sequence will be the same for all 12 zodiacal signs ingresses in the year 2215.

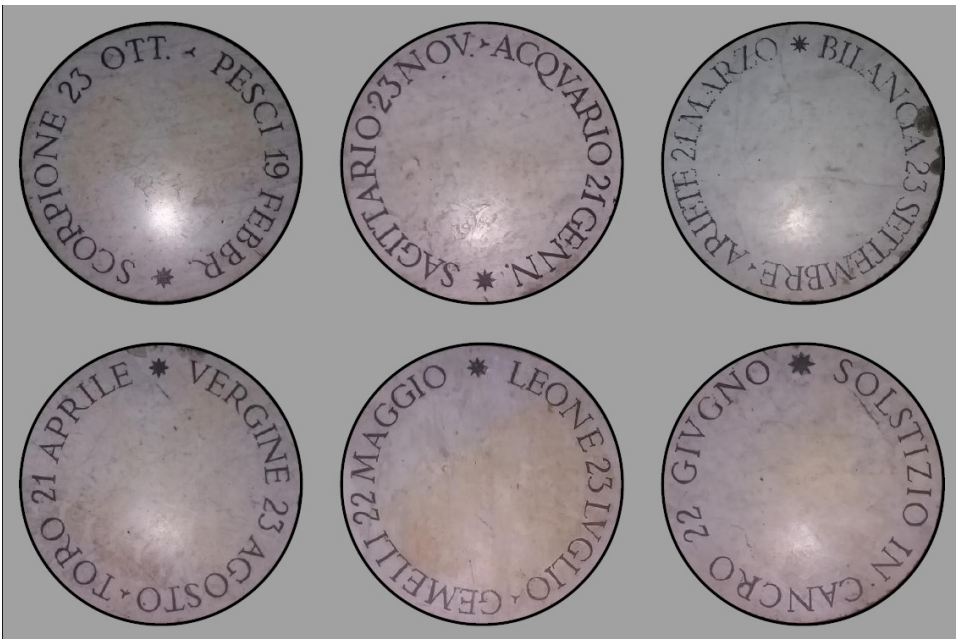


Fig. 10. Six of seven zodiacal disks: the Solstice in Capricorn 22 December is in the circular logo of figure 6. The dates of the zodiacal ingresses are the ones of 1815, and they will repeat in 2215.

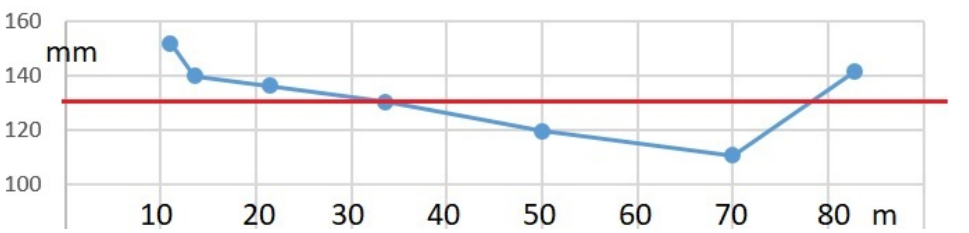


Fig. 11. On the seven zodiacal disks the meridian line is always West of the axis of the obelisk. The cross above the obelisk is also West of its axis, and it was reliably the sighting scope for the alignment of the austral zodiacal disks.

8. Deviation of the meridian line of St. Peter from Celestial North

The deviation of the meridian line has been measured from the obelisk's axis, and it is Westward.

The cross does not cast its shadow on the black stars of Leo and Cancer's disks; while it does from the Virgo-Taurus disk on.

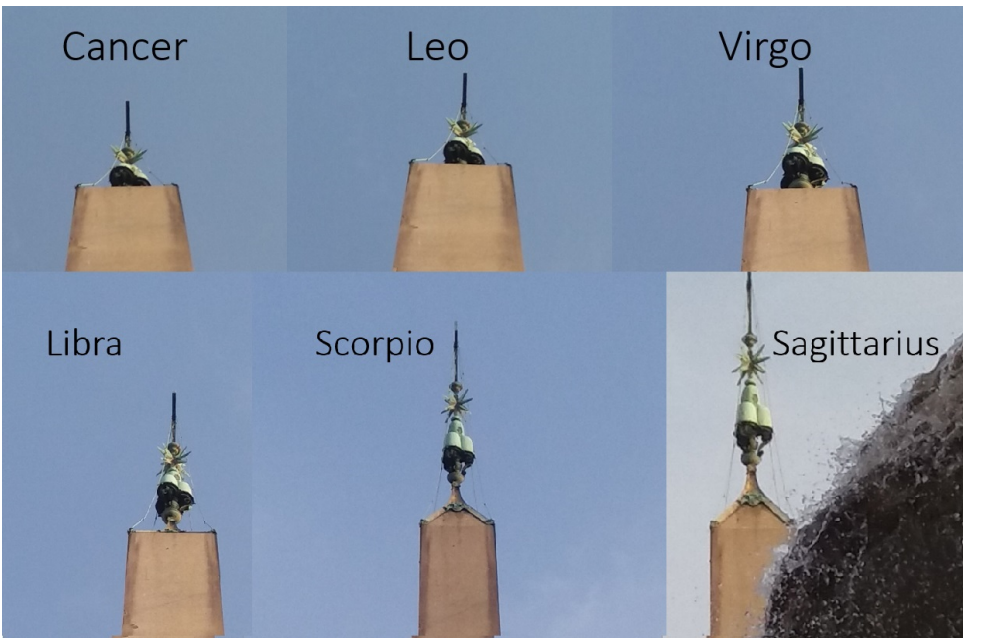


Fig. 12. The top of the obelisk as seen from all zodiacal disks.

9. Conclusions and perspectives

The positions of the celestial bodies is at the basis of Celestial Mechanics, and every gravitational theory: either Newtonian or Einstenian. To distinguish the latter are necessaries data with arcsecond level of accuracy. The measurements made at the historical meridian lines of St. Maria degli Angeli and St. Peter's square, with light equipment included:

- small meter tape
- radio synchronized watch
- a smartphone used for video and/or audio records
- "calibrated" feet
- a monocular 7×18 on small tripode

All these measurements on field allowed to know:

- the absolute calibration of the line's orientation
- the motion of the Sun, Moon and planets with respect to reference stars
- the stellar aberration with respect to the atomic time UTC
- Sun and planets' motion controlled with ephemerides (Stellarium 0.20.2)
- the variation of the Earth's rotation time $\Delta UT1$

Young students 14-15 years old understand better the season's rhythm as ruled by the solar rays' inclination at noon, as well as the astronomical observation considered as a real physics experiment, requiring calibration and errorbar evaluations, and long time basis (like the whole month needed for measuring the Sidereal Day up to 0.01 s). Also the data analysis experience with linear and parabolic fits has been mentioned. Finally, last but not least, the relevance of astronomy in the cultural heritage, owing the importance of the historical places were they can experience it: St. Maria degli Angeli, Diocletian's baths transformed into church by Michelangelo, and St. Peter's square realized along the centuries by the contributions of several popes, cardinals artists and scientists, namely Sixtus V, Domenico Fontana, Egnazio Danti, Pietro Maccarani and Filippo Luigi Gigli for the meridian line and Gianlorenzo Bernini for the colonnade's astronomical orientation from solstices' sunrises and sunsets.

References

1. Stellarium Astronomical Software (2021), http://stellarium.org/.
2. Solar transit on the Clementine Gnomon of September 23rd 2020, https://youtu.be/bhuH5FOi2tk.
3. Solar transit on St. Peter's meridian line of September 23rd 2021, https://youtu.be/eQU4B85DV7A.
4. Solar transit on St. Peter's meridian line of October 1st 2021, https://youtu.be/zfOwtO6pOl8.
5. Dates of seasons for any year, IMCCE, Paris, 2021, https://promenade.imcce.fr/en/pages4/439.html.
6. Venus transit of 27 October 2021 in St. Peter's square, https://drive.google.com/drive/u/2/folders/1ugl5TkhzQnE9qJcfJgx2BkiNnUIfKRAX.
7. Youtube Transits in St. Maria degli Angeli and Ostia's sunsets, https://docs.google.com/document/d/1TnbBfweA_-2_q-ZoC5Oi7HHOLKIyINUD8Dm5nnHqzMc/edit?usp=sharing.
8. Francesco Bianchini, De Nummo et Gnomone Clementino, Roma (1703) p. 45, https://archive.org/details/bub_gb_IYtGjIE_X64C/page/n289/mode/2up.
9. Costantino Sigismondi, Gerbertus **9**, 27 (2016).
10. Alessandro Travaglini, Pollens bulletin, Rome, Tor Vergata University (2020-2021), https://drive.google.com/drive/u/2/folders/17q3TgxmNp9L81P7_5Xy-VqKKJjokr7NI.
11. Costantino Sigismondi, *The meridian line of the Vatican obelisk to study the stellar aberration*, MG16 Proceedings (2021).

Daily, seasonal, and equinoctial solar paths on a school soccer field

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

The shadow of a wall cast by the Sun on a soccer field has been used to measure the change of the altitude and azimuth of the Sun along the day and through the seasons. The change of curvature parameter of the daily hyperbola shadow's path, has been recovered in the days around the spring equinox 2021. The study of the daily paths has been possible with chalk and meter during school days, on this field, and the data have been reported on a spreadsheet. The students were impressed by this whole-year experiment, while normally at school the experimental sessions are one-hour long.

Keywords: Solstices, Equinoxes, Shadows, Altitude, Azimuth.

1. Sun's paths on a soccer field

The shadow of a 10 meters wall (its vertex) cast on the soccer field of the Technical Institute Galileo Ferraris in Rome, 41.878 415° N, 12.454 662° E, coordinates WGS84 of the gnomon (fig. 2), has been used to perform daily measures with an accuracy up to a few millimeters. The gnomon's shadow, measured from the bottom and from the side of the field with a tape meter, in the best sets of measurements, presented rms below 1 mm. The hyperbolae drawn by the Sun through these shadows changed their aperture coefficient from Winter to Spring, passing to zero at the equinox. The students learned literally "on field" what means accuracy to the millimeter level, during several months. They learned how to do linear interpolations and extrapolations to predict the positions of the shadows in the next 10 minutes... or in the days without data (weekends and holidays).

The curvature radius of the hyperbolae is also measurable with three points along 20 minutes, repeated 3 times in a hour lesson.

All these achievements are possible during a single hour of lesson. The students remember well past situations, because they made practical actions.

The experience presented here with 14 and 15 years old students, started in February through May 2021.

The spring equinox was calculated by interpolation of aperture's parameters these months, with good agreement with the actual date.

The activity in open air and the use of the chalk on the field to mark the shadows' limits, the use of the meter and the care for the zero positions, have been appreciated alternatives to the normal indoor teaching.

2. Procedure

The positions of the shadow's cusp (e.g. fig. 3) are taken with respect the corner of the field as x and y. The curve connecting the daily positions (about ten each time) is plotted to see immediately its form. The meridian point (a, b) is measured and made center of a new coordinates system, with the appropriate translation. The meridian axis has an equation computed in the first days of observations, by marking the position of the shadow at the very instant of the meridian transit. Its angle α with respect to the axis x and y of the soccer field is used to apply the following rotation.

$$\begin{cases} x' = (x-a)\cos(\alpha) - (y-b)\sin(\alpha) + a \\ y' = (x-a)\sin(\alpha) + (y-b)\cos(\alpha) + b \end{cases}$$

Fig. 1. The equation of the roto-translation around the meridian point (a,b), to obtain a curve symmetric to the y' axis, with its vertex on the new origin (0,0).

A parabola fits with fair approximation the curve so obtained, by using a common spreadsheet (e.g Excel, LibreOffice, OpenOffice). The hyperbolic fit requires more passages, not automatically implemented in the spreadsheets.

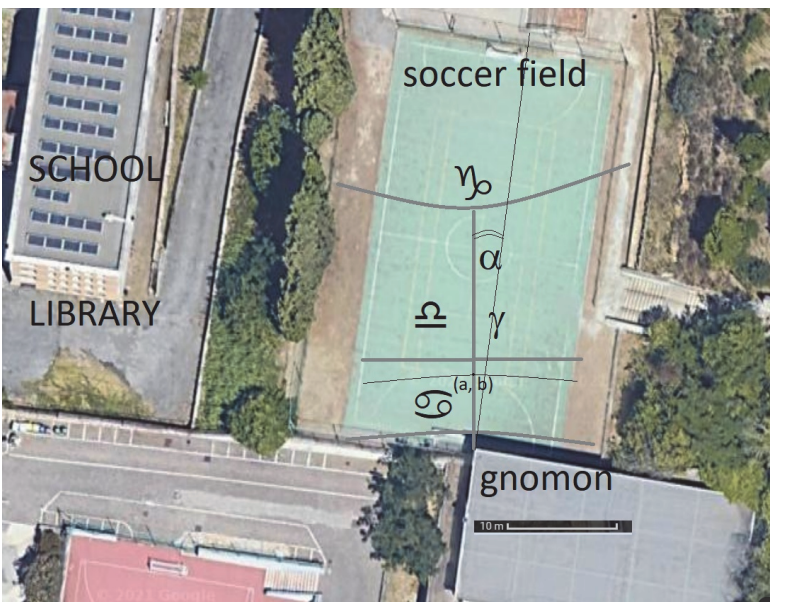


Fig. 2. The school's soccer field as seen from satellite, with the meridian line scheme superimposed.

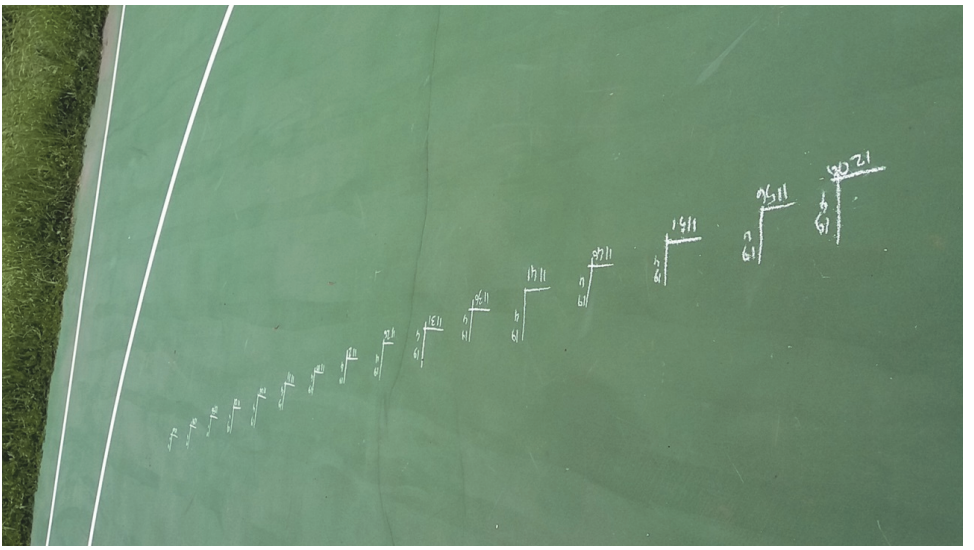


Fig. 3. The hourly path of the shadow's limit from 10:46 to 12:06 of 9 April 2021: the progress of the gnomon's shadow is visible.

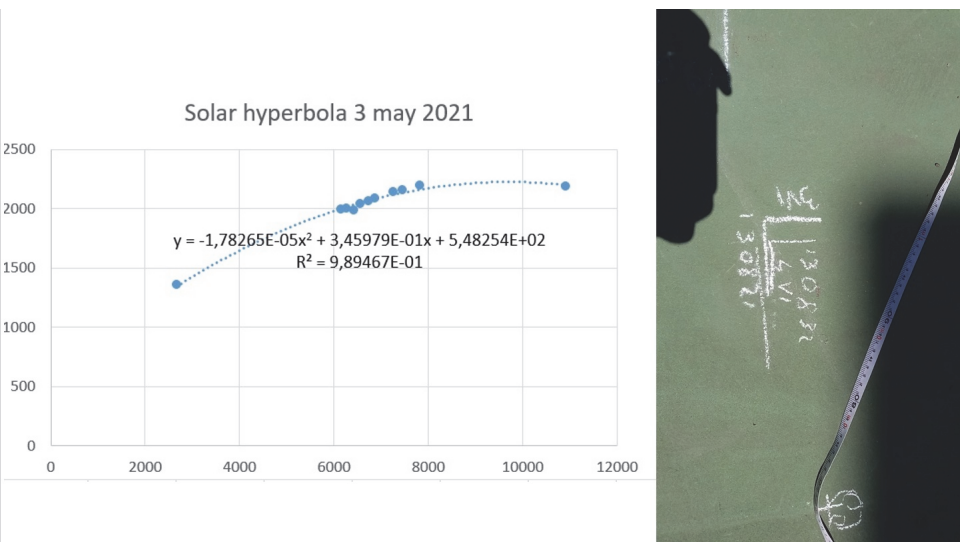


Fig. 4. The shadow on June 3rd 2021 and the plot of daily positions of May 3rd 2021, before the roto-translation.

3. Conclusions

The method presented is to fit the hyperbola of the shadow's positions of the wall's cusp on the soccer field, with the equation of a parabola. The accuracy achieved

is within 20 arcseconds, or 2 mm for each position of the shadow on the field. The path of two consecutive days, even near the summer solstice (where the positions are very close one to another), are well distinguishable each day from the other one. This procedure has been adapted to Excel spreadsheet, and the hyperbolic fits have been realized with the parabolae. The results obtained in the soccer field allowed to check the curvature's change during the equinox, and they have been published in detail (2021).

This is a suitable introduction to the trigonometrical methods of Celestial Mechanics, affordable by many schools and all students. These practical activities, made with increasing accuracy, may pave the way to understand the problems in experimental general relativity.

Reference

1. Sigismondi, C., Gerbertus, vol. 14, pp. 33-38 (2021) *L'equinozio e le parabole osculatrici*, http://www.icra.it/gerbertus/2021/Gerb-14-2021-Sigismondi-Parabole_osculatrici-33-38.pdf.

Teaching relativity at the AstroCamp

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

The AstroCamp is an academic excellence program in the field of astronomy and physics for students in the last 3 years of pre-university education, which often includes a course (or a significant part thereof) on Relativity. After an introduction to the principles, goals and structure of the camp, I describe the approach followed by camp lecturers (myself and others) for teaching Special and General Relativity, and some lessons learned and feedback from the students. I also provide some thoughts on the differences between the physics and mathematics secondary school curricula in Portugal and in other countries, and on how these curricula could be modernized.

Keywords: AstroCamp; Secondary education; Summer schools; Relativity.

1. The AstroCamp vision

The AstroCamp[a] is an academic excellence program in the field of astronomy and physics, organized by CAUP and the Paredes de Coura municipality (with the support of several national and international partners) for students in the last 3 years of pre-university education, i.e. roughly 15-18 year old students.

Our key goals are

(a) Promote scientific knowledge, with high-quality training in a secluded and tranquil setting
(b) Stimulate student curiosity and skills of critical thinking, team work and group responsibility
(c) Stimulate interactions between students with different backgrounds and life experiences but common interests

The camp was created in 2012, and now accepts applications from 42 eligible countries. Participation is strictly by invitation, after an application period (in April) and a selection phase (in May) which includes an interview in English. The camp itself takes place in August, in the Corno de Bico Protected Landscape area of the Paredes de Coura municipality (in the northwest of Portugal) and lasts 15 days. One of our points of principle is that for students in Portuguese schools the camp has no costs. Some students from other countries will also have support, others

[a] `https://www.astro.up.pt/astrocamp/`

may need to pay part of the costs or find their own support locally. In any case, significant efforts are made to keep costs low, and the maximum cost for foreign students is 400 Euro.

The main academic activity consists of two courses (each with 15h of lectures and a 2h written exam), given by currently active researchers with a PhD in a relevant area. Depending on the year, more than two courses are offered, but each student will only take two of them. In most of the 10 editions so far, one course or a significant part thereof was devoted to Relativity (both Special and General). This is to a large extent driven by the camp students themselves: one of the principles of the camp is that students are involved in the choice of the courses offered in the camp, and Relatively is clearly the prime example of a topic that the students feel is not satisfactorily covered in their school classes (if it is at all), and for which their school teachers are often unable to provide further information.

Other camp activities include observational and/or computational projects, stargazing sessions and documentaries, community service projects and evening talks which in early editions were open to the local public, but are now webcast live. There are also several other recreational activities such as hiking (including an overnight hike coinciding with the peak of the Perseids meteor shower). Interested students can also engage in several post-camp projects at CAUP and elsewhere. Our guiding vision is to provide the students with unique opportunities beyond the standard school systems, also after the camp. The extent to which students make use of all of these is of course up to them.

The camp is residential, with all students, teachers and camp monitors (university students, which most often are AstroCamp alumni) staying in the camp for the full 15 days. (Exceptionally, teachers may stay for reduced periods.) Due to COVID-19, the 2020 edition exceptionally had a hybrid format, with students living in Portugal in residence at the camp, and students living in other countries joining virtually in all activities where this was feasible. In the 2021 edition only one student had to participate remotely, and we naturally hope that the 2022 edition will be fully residential again.

2. Some statistics and outcomes

In the first four editions the camp only accepted applications from Portuguese students studying in Portuguese schools. Starting in 2016 the list of eligible countries has expanded, and currently the camp can accept applications from students matriculated in one of the last 3 years of pre-university education in any of 42 eligible countries[b], provided they are national citizens from one of these countries.

[b]These are: Andorra, Argentina, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, Norway, Poland, Portugal, Romania, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States of America and Uruguay.

In the 10 editions so far the camp had a total of 133 accepted students with 15 different nationalities. Some basic statistical information on the participants can be found in Table 1. It is worthy of note that the camp does not have quotas of any kind; the outcomes are therefore the result of the merit and academic potential of the candidates.

Table 1. Basic statistical information on the 2012-2021 AstroCamp students.

Accepted Students	2012-2015 Editions	2016-2021 Editions
Portuguese	100%	53%
Foreign	N/A	47%
Grade 10	24%	31%
Grade 11	52%	47%
Grade 12	24%	22%
Boys	47%	40%
Girls	53%	60%

Note: Grades 10, 11 and 12 are the names of the last three years of secondary education in the Portuguese school system. The student ages corresponding to these grades are country-specific, but broadly speaking Grade 10 students are 15 or 16 years old, while Grade 12 students are 17 or 18 years old.

One of the most rewarding deliverables of the AstroCamp is our Solar System Trail. This is the second scale model of the Solar System, in the Iberian Peninsula (and one of only sixteen in Europe, as far as we know), that is accurate both in terms of the sizes and of the distances of the objects. It was built during the first four AstroCamps, and a full hiking trail which follows the objects in the system—the Sun, the planets up to Neptune (including the Earth's Moon), and Ceres—was officially opened on 13 August 2016. The trail has recently been included among the teaching resources for Grade 7 students, so for a new generation of Portuguese students this may well be providing their first contact with astronomy. The trail also has an accompanying website[c] which at the time of writing only has a Portuguese version but is due to be expanded and replicated in at least 10 different languages in early 2022.

Most editions of the camp also included a computational project, lasting about 20 hours, aiming to provide the students with an introduction to scientific programming and statistics and data analysis in astrophysics. In several cases, the students continued working on these projects after the camp, and contributed to peer-reviewed publications in top academic journals.[1–3] At least one more such publication is expected soon.

[c] `https://www.astro.up.pt/trilhosistemasolar/`

3. Teaching special and general relativity

As previously mentioned, One AstroCamp principle is that students are involved in the choices of the courses taught in the camp, as well as in the ones they actually take. This is a multi-stage process, which we run in parallel with the student applications. A call for courses is issued at the beginning of each calendar year, and the submitted proposals undergo a scientific and pedagogical check, leading to a courses shortlist. The students shortlisted for the interviews are then asked to rank the shortlisted courses according to their preferences; this is done in a double-blind way—students don't know who would teach each course, and organizers don't know individual student preferences until after the student places are assigned. The courses to be offered are then decided and assignment to students, and a final tweaking of course contents can be done if there are overlaps between them. As an example, in the 2021 edition the shortlist included 6 courses, of which 3 were taught in the camp (with each student taking 2 of them).

So far, courses on Relativity/Cosmology (or containing at least ca. 5 hours of content on this) have always been selected by the students when offered. Various lecturers have been involved (in addition to the author), from Canada, Germany, Italy, The Netherlands, Portugal, Spain and the UK, some of them teaching twice.

Topics are covered both at the conceptual and (at least for some of them) also at the more detailed mathematical level. In our experience, most students also enjoy learning about the relevant historical context and background.

Galilean relativity topics covered include

(a) Galilean mechanics and acceleration
(b) Inertial frames and Galilean Transformations
(c) Historical developments from Galileo to Newton (Gassendi, Huygens, Descartes, etc)
(d) Newton's Laws and Equivalence Principle (inertial and gravitational masses)
(e) The XIX Century background to Relativity: Mechanics versus Electromagnetism

Special relativity topics covered include

(a) Derivation and interpretation of the Lorentz-Fitzgerald transformations
(b) Minkowski geometry and spacetime diagrams
(c) Special Relativity and the Principle of Relativity
(d) Invariance of the speed of light and its consequences
(e) Time dilation and Lorentz contraction: Cosmic rays and Twin paradox
(f) Physical meaning of inertial frames

Finally, topics pertaining to applications of relativity covered include

(a) Relativity and the GPS
(b) The Schwarzschild solution and black hole properties

(c) Derivation of the Friedmann and Raychaudhuri equations, and simple solutions thereof
(d) Cosmology, including Hubble's law and simple FLRW universes
(e) Gravitational waves, including their detection

Examples of conceptual questions on these topics that have been previously used in AstroCamp course exams are listed in the Appendix. Student feedback clearly shows that the Newtonian physics curriculum is unattractive, and often containing significant misconceptions. Camp students usually follow conceptual aspects very well (even when finding some of them counter-intuitive); any difficulties usually pertain to their mathematical background, which is partially correlated with their age.

4. Lessons learned, and the COVID-19 impact

The experience of ten camp editions (of which six included foreign students) reveals significant differences between the secondary school curricula in different countries, and sometimes in different regions of the same country. There are also country-specific differences between public and private (or international) schools. In Portugal, it is clear that private schools (and even a few public schools) routinely apply grade inflation. This further reinforces the prior expectation that grades alone do not provide a fair selection process. This is why, from the first edition, the application procedure includes a student motivation letter, a teacher recommendation letter and, importantly, an interview to shortlisted candidates.

Since the camp accepts students from the last three years of secondary education, one challenge for the camp teachers is that the mathematical background will be quite diverse. Specifically, derivatives introduced at different ages in different countries, and integrals are not a given even for Grade 12 students, who have finished secondary education and are about to start university. However, in our experience this is not a problem for bright students: through direct interaction with teachers or camp monitors, they can quickly gain a working knowledge, e.g. of differentiation rules, which allows then the follow the course contents that rely on them.

One rather surprising lesson is that schools can actually be a bottleneck in disseminating information on the AstroCamp to students. Specifically, from our experience less than 20% of Portuguese schools that receive direct camp information from us (including, in some editions, A4 or A3 sized posters and leaflets) actually transmit it to their students, and the fraction of students that find out about the camp through a school teacher is small—while the number is not trivial to estimate, it is clearly below 50%.

A recent challenge was provided by COVID-19. This led to a hybrid format edition in 2020, with foreign students interacting with camp teachers, monitors (including some online tutors) and students in residence at the camp through online collaboration tools such as Zoom, Discord and Colaboratory online collaboration.

This was was nevertheless very successful despite the fact that foreign students participating in this edition had a very limited experience of the camp, and clearly demonstrates that remote collaboration tools enable new teaching and mentoring opportunities, which we intend to pursue.

A more surprising realization emerged in the 2021 edition, in that the COVID-19 impact on students' cognitive and social skills and emotional maturity was clearly visible. Inter alia, although the level of the courses and exams has been kept constant in recent editions (roughly at the level of a course from first year undergraduate physics degree at a top Portuguese university), the average grades of the students were lower and, in particular, for the first time in ten editions two students failed their exams. It is likely that this impact will be noticeable for several years to came. This is all the more scary since these are bright students—the impact on average ones may be much more serious.

As the AstroCamp starts its second decade, this is a good time to revise and update its format, structure and rules, also taking into account the most recent challenges. There is clearly some room for growth—more in extension and scope than in numbers—making the camp an international reference for events of this kind.

5. Epilogue: Improving Portuguese secondary scientific education

The experience of 10 years of organizing the AstroCamp (together with teaching one course in each edition, and leading other scientific activities therein) also leads to the conclusion that the Portuguese secondary education system, in scientific areas, is clearly lagging behind those of other developed countries, and needs urgent revision and modernization.

In an effort to contribute to an urgent and much needed debate, I may suggest four very specific measures

(a) It is clear that Portuguese students are, on average, far less familiar with scientific programming that their foreign peers. (The Portuguese school system current offers a 'Computational Applications' subject, but this is a total waste of time[d].) One must urgently introduce a compulsory scientific programming subject in Grades 10 to 12 grades, using languages such as Python, Octave or Julia, and connected to other subjects, especially Physics, Chemistry, Biology and Mathematics.
(b) In the Portuguese school system, Physics and Chemistry are taught as a single subject in Grades 10 and 11, and are only separate subject in Grade 12.

[d] As well as an insult to the intelligence of most students, were it not for the fact that it allows an easy high grade to be obtained, which will count towards the students' university application.

Moreover, very few students take Grade 12 Physics, either because the school does not offer it or because the students consider it a hard subject (form the perspective of getting a good grade) and prefer softer subjects such as English, Psychology or the infamous Computational Applications. This is a crass mistake. Physics and Chemistry should be separate subjects from Grade 10 (and, arguably, even earlier).

(c) The Portuguese school system has national exams which, depending on the subject, are either at the end of Grade 11 or Grade 12. The former, which currently include the Physics exam, is a clear mistake. The Physics national exam (and indeed all national exams) should be cumulative and at the end of Grade 12. Moreover, Grade 12 Physics should be made compulsory for students intending to apply for university physics and engineering degrees

(d) Last but not least, these is a vast disconnect between the secondary school environment and that of university (not to mention that of research, business, etc). A way to bridge the gap would be to reward students with top results in national exams with summer internships at research labs or tech companies (with the student matching done by the hosts). As a way to foster equal opportunities, these internships should not be merely awarded to the students with the numerically highest grades but, for example, to the student with the highest grade in each district (or, more ambitiously, each municipality in Portugal—in the former case this kind of program would reach about 20 students, while in the latter it would reach about 300.

All of these are, at least in principle, easy to implement and would have dramatic and comparatively rapid (as well as quantifiable) impacts. The main bottlenecks, other than political will are our blatant 'teaching to the test approach' (manifest in the blind acceptance of school rankings, but partially stemming from the university numerus clausus system), and the pathological reluctance of teacher unions to any kind of change. The latter can be, at least in part, ascribed to the fact that most secondary school science teachers are in their 50s or 60s. This is especially relevant in the context of the first of the above points, since these teachers and have little computing literacy, and none in coding and scientific programming. Nevertheless, delaying the modernization of the system, with these and other measures, simply increases the personal and financial costs for future generations.

Acknowledgments

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017.

Appendix A. Conceptual question examples

The following is a partial list of relevant conceptual questions that have been used in at least one AstroCamp exam. Except for the last one (with is a proposed essay topic) they are all multiple choice questions.

Two inertial observers measure the speed of the same beam of light and obtain the same finite value. This would surprise Newton because

(a) According to him the speed of anything should depend on the frame of reference.
(b) He expected the speed of light to be infinite in inertial frames.
(c) In inertial reference frames he believed light to be stationary.
(d) He believed light should have an infinite speed in all frames.

The Michelson-Morley experiments were an attempt to

(a) Measure the speed of light.
(b) Test the Lorentz-Fitzgerald transformations.
(c) Measure the Earth's speed relative to the Sun.
(d) Detect the Earth's movement relative to the aether.

Relative to an object at rest, a moving object is

(a) Shorter and younger.
(b) Longer and older.
(c) Shorter and older.
(d) Longer and younger.

Special relativity (SR) and general relativity (GR) corrections must be taken into account for the GPS system to work. If these satellites were moved to LEO,

(a) Both corrections would decrease.
(b) The SR correction would decrease and the GR one would increase.
(c) The SR correction would increase and the GR one would decrease.
(d) Both corrections would increase.

If an astronaut in a windowless rocket feels pressed against the back of her seat, she won't be able to tell

(a) Whether the rocket is accelerating or moving at constant speed.
(b) Whether she is accelerating or resting on the surface of some planet.
(c) Whether the rocket it moving at constant speed in space or at rest in space.
(d) Whether she is at rest in space or freely falling towards some planet.

A green animal lives on an extremely dense planet. When observed from space,

(a) It will look green.
(b) It will look white.

(c) It will look blue.
(d) It will look red.

Which of these is not a direct test of the Einstein Equivalence Principle?

(a) Measurements of gravitational redshift.
(b) Tests of the stability of fundamental couplings.
(c) Measurements of the precession of Mercury's perihelion.
(d) Tests of the universality of free fall.

Essay topic: Describe the relation between physical reality and the mathematical models with which we describe it. In particular, comment on how this relation may differ in the contexts of the pure and applied physical sciences (e.g., theoretical physics and engineering).

References

1. C. J. A. P. Martins, A. M. M. Pinho, R. F. C. Alves, M. Pino, C. I. S. A. Rocha and M. von Wietersheim, Dark energy and Equivalence Principle constraints from astrophysical tests of the stability of the fine-structure constant, *JCAP* **08**, p. 047 (2015).
2. C. S. Alves, A. C. O. Leite, C. J. A. P. Martins, T. A. Silva, S. A. Berge and B. S. A. Silva, Current and future constraints on extended Bekenstein-type models for a varying fine-structure constant, *Phys. Rev. D* **97**, p. 023522 (2018).
3. M. C. F. Faria, C. J. A. P. Martins, F. Chiti and B. S. A. Silva, Low redshift constraints on energy-momentum-powered gravity models, *Astron. Astrophys.* **625**, p. A127 (2019).

Sungrazing comets as General Relativistic gravitational probes

Silvia Pietroni

Dipartimento di Fisica "E.R. Caianiello", Università di Salerno
Via Giovanni Paolo II 132, I-84084 Fisciano, Italy
Istituto Nazionale di Fisica Nucleare, Sezione di Napoli
Via Cintia, 80126, Napoli, Italy
spietroni@unisa.it

Costantino Sigismondi

ICRA - International Center for Relativistic Astrophysics, Sapienza Università di Roma
Piazzale Aldo Moro 5, 00185 Roma, Italy
sigismondi@icra.it

Among the celestial bodies of the solar system, the ones passing closer to the solar mass are the sungrazing comets. The sungrazers can approach the Sun with a perihelion more than 60 times closer than Mercury, experiencing a local effect of perihelion relativistic precession 60 times larger. Mercury requires 15 years, or sixty orbits, to totalize the same effect. The sungrazers have generally no more than a single orbit, because they vanish at the perihelion passage. The observability of such phenomena with SOHO coronagraphs is discussed.

Keywords: Sungrazing comets, perihelion relativistic precession

1. Generalities on sungrazing comet

The sungrazing comets are a class of comets whose known number significantly increased in the last 26 years, since the SOHO coronagraphes entered in function, at the rhythm of about hundred per year. The gravitational studies on such comets, starting from Newton are presented in the references.[1–21]

Their orbital parameters suggest their grouping as related to a few parent bodies, and we recall the statistical work on comets made by Enrico Fermi in 1922 as thesis for the Normal School of Pisa.

The sungrazing comets are the celestial bodies which approach more closely the mass of the Sun: then their orbital parameters are influenced by both General Relativity and non-gravitational effects, like mass loss and outgassing. The General Relativity effects and the non-gravitational effects are compared to understand the observational accuracy required on Mercury's perihelion. The studies carried by Le Verrier and Newcomb in the second part of XIX century, were possible because the orbit of Mercury is rather eccentric e=0.2056. When some objects may graze the solar surface, a very high eccentricity e=0.99 is considered to produce some visible effect. The accuracy on the determination of the orbital parameters of sungrazing comets is also investigated, to enforce with observations all these theoretical

concepts. The identification of new sungrazing comets is possible online, at the SOHO website, and it has been realized by the high school students of Galileo Ferraris Institute, Rome, as curricular activity, with great enthusiasm, especially with the real time discoveries of January 18, 2021 (fig. 3).

2. Measurability of the relativistic precession

$$\delta\theta = 6\pi \cdot GM_{\odot}a/c^2b^2$$

Planet	Planet (mean)	a AU	e	Perturbation [arcsec/cy]	$e \cdot \delta\theta$ [arcsec/cy]
Mercury	Mercury	0.38709893	0.20563069	43.03	8.847
Venus	Venus	0.72333199	0.00677323	8.63	0.059
Earth	Earth	1.00000011	0.01671022	3.84	0.064
Mars	Mars	1.52366231	0.09341233	1.35	0.126
Jupiter	Jupiter	5.20336301	0.04839266	0.06	0.003

Fig. 1. The relative observability from Mercury to Jupiter, as indicated by Dennis Sciama (1972).

The eccentricity larger than 0.99, that the sungrazer can present, and a perihelion's advancement of $60 \times 43''$ produces an effective visibility of 43', larger than the apparent diameter of the Sun.

The problem is that the comet is rarely visible after the perihelion, because it vanishes, and when it is visible the effect is small due to a less close encounter with the photosphere.

3. Conclusions and perspectives

The sungrazers are comets belonging to groups with similar geometrical parameters, originated by a close encounter of a big comet with the solar photosphere. They experienced the largest General Relativistic effect of perihelion precession possible in our solar system, even if for only one passage. The amplitude of such effect can be such as 0.75°, but its measurability with the SOHO coronagraph is very difficult due to the relative inclination angles with Earth's orbit; moreover the sungrazers vanish after the passage to the perihelion due to the closeness with the photosphere. The possibility to "discover" averagely a new sungrazer each week makes interesting the discussion of such effect in secondary school's physics class, where the students become protagonist of the discoveries through inspecting the SOHO movie theatre, in real time. The introduction to observational General Relativity can be effective through this approach, for the young students.

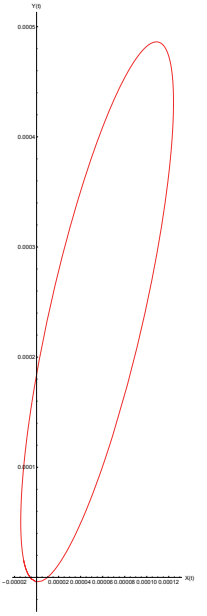


Fig. 2. One complete orbit of a sungrazer comet calculated with a computer code developed by Bozza and Pietroni (2021), distance given in parsecs; orbital parameters: semimajor axes $a \simeq 52AU \simeq 7.64 \times 10^{12}$ meters, eccentricity $e = 0.98$, inclination $i = 160°$, period $P = 360$ ys. The relativistic Lagrangian has been used to evaluate the perihelion advancement, and compared with the non-relativistic one. The interaction with the solar atmosphere is not taken into account, to close the orbit.

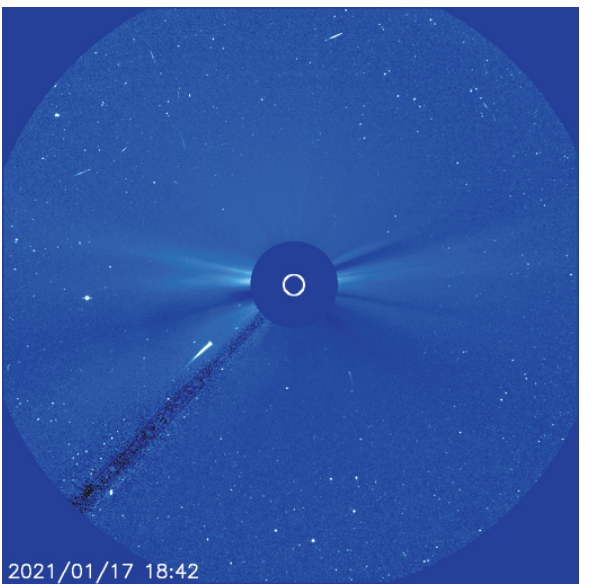


Fig. 3. Two sungrazer comets at the same time on January 17, 2021, observed by SOHO C3 coronagraphs, and "discovered" at school in real time.

References

1. T. Scarmato, Sungrazer comet c/2012 s1 (ison): Curve of light, nucleus size, rotation and peculiar structures in the coma and tail (2014).
2. R. Ruffini and C. Sigismondi, *Nonlinear Gravitodynamics* (World Scientific, 2003).
3. M. M. Knight and K. J. Walsh, Will comet ison (c/2012 s1) survive perihelion?, *The Astrophysical Journal* **776**, p. L5 (Sep 2013).
4. S. Giordano, J. C. Raymond, P. Lamy, M. Uzzo and D. Dobrzycka, Probing the solar wind acceleration region with the sun-grazing comet c/2002 s2, *The Astrophysical Journal* **798**, p. 47 (Dec 2014).
5. Z. Sekanina and R. Kracht, Strong erosion-driven nongravitational effects in orbital motions of the kreutz sungrazing system's dwarf comets, *The Astrophysical Journal* **801**, p. 135 (Mar 2015).
6. A. Sekhar, D. J. Asher, S. C. Werner, J. Vaubaillon and G. Li, Change in general relativistic precession rates due to lidov–kozai oscillations in solar system, *Monthly Notices of the Royal Astronomical Society* **468**, p. 1405–1414 (Mar 2017).
7. O. Ivanova, V. Reshetnyk, Y. Skorov, J. Blum, Z. S. Krišandová, J. Svoreň, P. Korsun, V. Afanasiev, I. Luk'yanyk and M. Andreev, The optical characteristics of the dust of sungrazing comet c/2012 s1 (ison) observed at large heliocentric distances, *Icarus* **313**, p. 1–14 (Oct 2018).
8. J. C. Raymond and S. Giordano, Probing coronal magnetic fields with sungrazing comets: H i ly from pickup ions, *The Astrophysical Journal* **887**, p. 45 (Dec 2019).
9. M.-T. Hui and Q.-Z. Ye, Observations of disintegrating long-period comet c/2019 y4 (atlas): A sibling of c/1844 y1 (great comet), *The Astronomical Journal* **160**, p. 91 (Jul 2020).
10. D. Jewitt, Systematics and consequences of comet nucleus outgassing torques, *The Astronomical Journal* **161**, p. 261 (May 2021).
11. B. Gundlach, J. Blum, Y. V. Skorov and H. U. Keller, A note on the survival of the sungrazing comet c/2011 w3 (lovejoy) within the roche limit (2012).
12. I. Haranas, O. Ragos and I. Gkigkitzis, The lens-thirring effect in the anomalistic period of celestial bodies (2013).
13. Z. Sekanina, Comments on a recent review paper on near-sun comets (2019).
14. K. Battams and M. M. Knight, Soho comets: 20 years and 3000 objects later, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**, p. 20160257 (May 2017).
15. C. Sigismondi, Astrometry and relativity, *Nuovo Cimento B Serie* **120**, p. 1169 (October 2005).
16. C. Sigismondi and F. Maiolino, Enrico Fermi and the statistics of comets, *Nuovo Cimento B Serie* **117**, p. 1207 (September 2002).
17. C. Sigismondi, The Eddington's Eclispe and a Possible Replica of the Experiment of Light Bending, *arXiv e-prints*, p. arXiv:1507.03879 (July 2015).
18. C. Sigismondi, The Eddington's eclispe replicated with the images of the SOHO coronographs, in *Fourteenth Marcel Grossmann Meeting — MG14*, eds. M. Bianchi, R. T. Jansen and R. RuffiniJanuary 2018.
19. C. Sigismondi, L'algoritmo di correzione per l'estinzione atmosfera a occhio nudo e studio su foto da webcam e da SOHO, *Gerbertvs, International Academic Publication on History of Medieval Science* **13**, 17 (July 2020).
20. D. W. Sciama, *La relatività generale. Fondamenti fisici della teoria.* 1972.
21. Soho movie theater website `https://soho.nascom.nasa.gov/data/Theater/`.

The three Summer solstice's markers of 1721 unveiled in the Basilica of Santa Maria degli Angeli in Rome

Costantino Sigismondi
ICRA, Sapienza University of Rome, Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

The solar image is projected through a pinhole inside a camera obscura in the great meridian line in the Basilica of Santa Maria degli Angeli in Rome: the daily path of 1721 summer solstice has been recovered as passing through three markers on the floor of the Basilica, near the summer part of the meridian line. The function of these three markers was considered ornamental until now, because no written documents have been found on them. Moreover two of the three markers today are no longer illuminated by the Sun, after 1750.

The same technique has been applied to the three solar markers in St. Maria degli Angeli, finding that they belong to the same 1721 solstice of three centuries ago, within a few millimeters of accuracy.

Keywords: St. Maria degli Angeli, Meridian Line, 1721, Solstices, Equinoxes, Maria Clementyna Sobieska, James III Stuart.

1. The 1721 Summer solstice perpetuated in St. Maria degli Angeli

The Meridian Line of Santa Maria degli Angeli in Rome has been realized in 1700-1702 upon the will of Pope Clement XI, since when he was Cardinal, after the project of Francesco Bianchini (1662-1729) to measure with unprecedented accuracy the tropical year and the variation of the obliquity of the ecliptic. At this Great Gnomon the seasonal effects on the atmospheric refraction and the effects of stellar aberrations were measured, but while the first were recognized the latter become systematic effects included in the errors of measurements.

At the Marcel Grossmann Meeting of Berlin (2006) this discovery was firstly presented; after other historical data were analyzed (2014), while in International Astronomical Union General Assembly of 2018 in Vienna the recognition of two brass markers as solstice markers was announced.

The two markers belong to the summer solstice because they are on the same daily hyperbola of the solar center entering Cancer in 1721. One of the two brass markers, named after this paper the King's marker, is no longer illuminated by the Sun since 1750 Vanvitelli's renovations of the Basilica, which modified the pinhole's visibility from all the floor.

A third marble marker, considered as ornamental until now, it is on the same hyperbola and it is on the floor of the Basilica, about 7 meters far from the other.

The different material of this third marker delayed our discovery of the strict correlation of the three markers. This marble was dedicated to the King James III Stuart in 1721; the well known inscription says: FELIX TEMPORVM REPARATIO meaning "Return to good times", very probably after his marriage in 1719 1st of September (Francesca Ceci, 2021), with Maria Klementyna Sobieska and the birth of their son on 31 December 1720.

These markers are offset with respect to the great meridian line, because on the Cancer's marble no more modifications were possible. This chronological sequence explains also why in his book of 1703, Bianchini did not mention these markers.

Fig. 1. Two markers: King James' (Jacobus in Latin) III in marble and Klementyna Sobieska's in brass, the one still illuminated by the Sun.

2. Observations and inspections

The three solstice's ellipses, two in brass (half ellipses, as the only one still illuminated by the Sun in the right-side of figure 1) and one in marble (whole ellipse, left-side figure 1) with the dedicational epigraph containing the date (adding together the capital letters as Latin numbers) are no longer hit by the Sun's image, excepted the last one, before the meridian line.

In 2018 the two brass decorations have been recognized as half ellipses of the Summer solstice's Sun. This year 2021, exactly three centuries after their realization, the all three markers have been recognized as belonging to the same "memorial addition" to the great meridian line.

With this addition, made very reliably by Francesco Bianchini, the royal couple Maria Clementyna Sobieska (1701-1735) and James III Stuart (1688-1766), king without land, married since 1719 and living in Rome under the protection of the pope, has been celebrated by realizing three summer solstice markers.

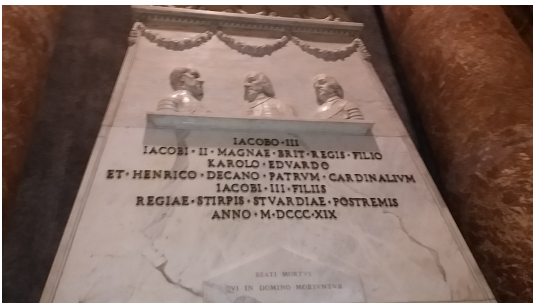


Fig. 2. The cenotaph of King James III in St. Peter's Basilica, were he is buried.

Fig. 3. Maria Klementyna Sobieska: the same image used in her cenotaph, in St. Peter's Basilica, in front of James III (painter Martin van Meytens, according to Wikipedia 2021.

3. Fitting the three positions with one hyperbola

The giant pinhole meridian line has been somewhat reproduced in a school environment, with the possibility to mark the soil with chalk and to bring many measurements with the meter, that are impossible in the Basilica. The hyperbolic paths of the Sun's shadow on the floor showed clearly the change of curvature from North to South at the spring equinox. The same method developed to fit the equation of the hyperbola to the shadow's positions of the cusp of a wall on the soccer field (Sigismondi, 2021, these Proceedings) has been applied on these three positions, marked by the historical ellipses.

These three ellipses have been considered as ancient observations, marked on the floor of the Basilica at their exact position.

The hyperbolic fits has been realized either with a parabola, either with an hyperbola: a millimeter accuracy in the positions of the Sun on the 1721 Summer solstice's path has been found in Santa Maria degli Angeli. The equation of the hyperbola of the 1721 summer solstice has been recovered with respect to the current solstice (2021) at 33.38 centesimal parts on the meridian line (on the body of Cancer in fig. 4). The meridian line is the y- axis, the origin is the 33.38 point on it, and the x-axis is perpendicular to the meridian line.

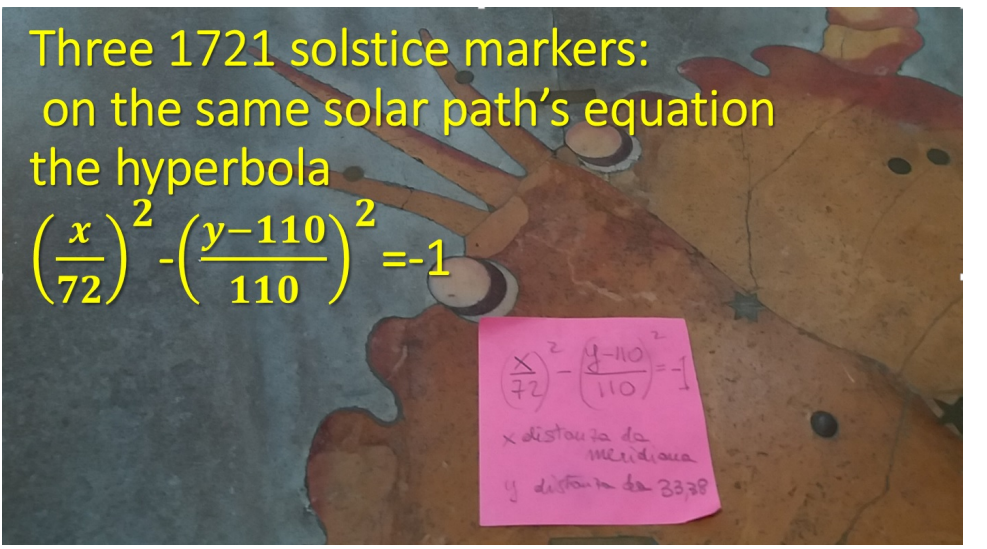


Fig. 4. The equation of the 1721 summer solstice's path as derived from the three markers. The parameters in the equation are expressed in meters.

4. Obliquity variation in three centuries

In 2021 the difference of Earth orbit's obliquity has been evident after 3 centuries, as a shift of 15 mm (fig. 5). The brass' profile is the 1721 one, while the luminous disk is the 2021 solar position. The new pinhole installed for the 2018 IGEA[a] campaign[9] has the same position along the meridian axis of 1702.

Fig. 5. Obliquity meter: differential measure of June 20, 2020 compared with 1721 at the Summer solstice. The image of the Sun is formed over the Queen's marker, the second in the sequence.

[a]IGEA standing for Informatized Geometric Ephemerides for Astrometry, started on October 27, 2018 and was named in honor of Igea Contessa (1918 Oct 27-2009 Dec 11)

Fig. 6. The marble marker with the inscription. A pencil shows its dimension.

5. Conclusions: High science in *Roman Settecento*

It is natural that the policy of the popes during the life of James III Stuart was oriented toward the restauration of the catholic dinasty at the throne, and Francesco Bianchini as close friend of Clement XI collaborated to this policy. The title of the new work of J. L. Heilbron (to appear in 2022) dedicated to Bianchini seems not to depart significantly from his approach in *The Sun in the Church* of 1999, with *tranchant* sentences on the papal policy of 1700. This paper along with the other one concerning the aberration of Sirius and the Equation of time in 1703 complete a 20 years' cycle of measurements and studies on the meridian line of St. Maria degli Angeli, inspired by Heilbron's book (1999), but in the direction of showing the prestige and the accuracy of roman astronomy, on the cutting edge of celestial mechanics of the eigtheenth century, instead of an easy *bavardage* on the court's intrigues as the forthcoming new title suggests.

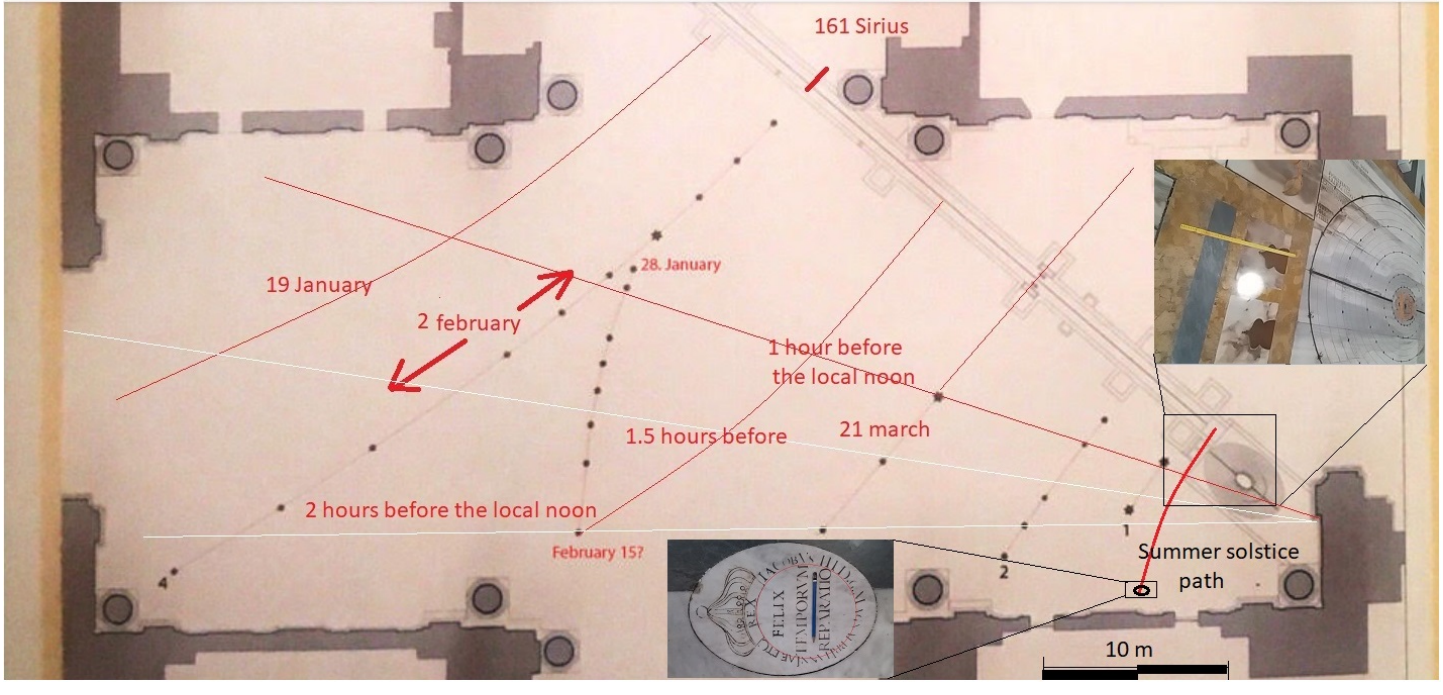


Fig. 7. Location on the map of the Basilica of the three markers.

Fig. 8. The Sun between the two brass markers, and the boreal line with the ellipses of the Jubilee years.The great meridian line starts on the left of this photo, taken on 4th July 2021 at 13:01. The Sun moves left in this photo, from the King's to the Queen's marker.

Fig. 9. The Sun's image appeared just after the King's marker on July 3rd 2021 at 12:58. It is the King's one because it was hit first by the Sun. Nowadays the Sun does not illuminates parts of the floor of the Basilica from the pinhole, after the intervents of Luigi Vanvitelli in 1750, which obscurated with a decorative arch the pinhole. Roger J. Boscovich cut the arch in correspondence of the meridian line, and the above image of the Sun is appearing from this cut.

References

1. Bianchini, F. (1703), *De Nummo et Gnomone Clementino* https://archive.org/details/dekalendarioetcy00bian
2. Sigismondi, C., et al., IAU General Assembly, Vienna (2018) *The rediscovery of the obliquity meter in the meridian line of St. Maria degli Angeli in Rome* http://homepages.vub.ac.be/~csterken/C3_Science_meeting.pdf
3. Sigismondi, C., (2008) *MG XI Proceedings* , https://ui.adsabs.harvard.edu/abs/2008mgm..conf.2470S/abstract.
4. Sigismondi, C., et al., *Measures of the Earth obliquity during the 1701 winter solstice at the Clementine meridian line in Rome*Proceedings of the Journées 2014 "Systèmes de référence spatio-temporels" https://ui.adsabs.harvard.edu/abs/2015jsrs.conf..116A/abstract.
5. Ceci, F. and G. Breccola *1 settembre 1719 a Trecento anni dalle nozze regali a Montefiascone* Edizioni, ArcheoAres, Terni (2021) https://www.facebook.com/accademia.polacca/photos/a.416643071829557/1709674122526439/?type=3&theater
6. Heilbron, J. L. *The Incomparable Monsignor: Francesco Bianchini's world of science, history, and court intrigue* (to appear in 2022) https://www.barnesandnoble.com/w/the-incomparable-monsignor-jl-heilbron/1140143257#
7. Heilbron, J. L. *The Sun in the Church* Harvard University Press, 1999.
8. Sigismondi, C., Gerbertus, vol. 14, pp. 33-38 (2021) *L'equinozio e le parabole osculatrici* http://www.icra.it/gerbertus/2021/Gerb-14-2021-Sigismondi-Parabole_osculatrici-33-38.pdf.
9. Sigismondi, C., *Earth's Obliquity and Stellar Aberration Detected at the Clementine Gnomon (Rome, 1703)* Phys. Sci. Forum 2021, **2(1)**, 49 https://www.mdpi.com/2673-9984/2/1/49/htm
10. Catamo, M. and C. Lucarini, *Il Cielo in Basilica*, Arpa-Agami Roma, 2011 2nd ed.

Einstein-First: Bringing children our best understanding of reality

A. Popkova*, K. Adams, S. Boublil, R. K. Choudhary, E. Horne, L. Ju, T. Kaur, D. McGoran, D. Wood, M. Zadnik and D. G. Blair

Department of Physics, The University of Western Australia, 35 Stirling Highway, Crawley Perth, WA 6009, Australia
**23134632@student.uwa.edu.au*

D. F. Treagust

STEM Education Research Group, School of Education, Curtin University, Kent Street, Bentley, Perth WA 6102, Australia School of Education, Curtin University
d.treagust@curtin.edu.au

The Einstein-First project is designed to resolve a conflict between modern science and the science taught at school, which is a significant cause of students' negative attitudes to STEM. Our program resolves these contradictions by teaching our best understanding of the universe, dubbed Einsteinian Physics, from an early age. We use models and group activities in a carefully crafted 8-year learning progression, to give students a basic understanding of the language and concepts describing our physical universe, from quarks to the big bang. Einstein-First works with teachers to develop courses, lesson plans and training workshops. These components all contribute to curriculum trials where student learning and attitude outcomes can be assessed. Every trial of curriculum modules or short intervention programs has yielded exciting, positive outcomes including surprising gender equalising effects and benefits for less academic and disadvantaged students. There are multiple classroom trials in place in local partner schools. In this paper we present an overview of the Einstein-First program and give examples of the ability of students from age 8 to 12 to comprehend modern scientific concepts.

Keywords: Einsteinian Physics; Science Education

1. Introduction

The experimental educational project called Einstein-First is a collaboration of scientists and educators who are concerned about the current school science curriculum which: a) includes misconceptions b) is irrelevant c) does not meet needs of people in the modern world. The existing science curriculum completely ignores the modern, awesome discoveries of relativistic astrophysics and makes teaching of modern concepts impossible.[1] For example, the first direct observation of gravitational waves in 2015 and the rapid development of gravitational wave astronomy, have provided a new spectrum for understanding the Universe as well as providing evidence to support the theory of relativity predicted by Einstein a century ago.[2] The discovery of gravitational waves has inspired new developments in areas such as optomechanics, the quantum mechanics of gravitational wave detectors, as well as observational investigations of black holes, neutron stars and heavy element nucleosynthesis.[3] Also, in the last century, science has given us knowledge of the tiniest things in

the universe and vast distances to the edge of the visible universe, and with it, the modern technologies on which our lives depend. Recently, in 2021, the Nobel prize in physics was awarded for modelling of complex systems related to climate change and the discovery of the interplay of disorder and fluctuations in physical systems from atomic to planetary scales.[4] These are all examples of modern discoveries that are not included in the current Australian science curriculum.

Primary and secondary school science curricula still rely on outdated scientific paradigms of traditional Newtonian physics based on classical concepts, research methods, postulates, and standards. Thus, the revelations of modern science and even the vocabulary of modern physics is denied to early and middle school students. The consequences of this situation may be that students are not able to digest the core concepts related to new discoveries during higher level of schooling with the subsequent loss of students' interest in the science presented as being irrelevant to real life events across the world.[5] Einsteinian physics, encompassing both general relativity and quantum mechanics provides a better understanding of modern reality and is relevant to students' everyday life. Unfortunately, general relativity and quantum mechanics have typically been reserved for specialist high school physics subjects and mainly taught at the tertiary level. For the overwhelming majority of students, modern physics may be inaccessible. Year 10 is the last year of compulsory science in Australian schools and can be the last opportunity for the education system to foster scientific literacy on our school population.

One reason for keeping children away from Einsteinian physics is because these concepts are considered to be difficult to comprehend at an early age. In 2004 Kathleen E. Metz stated "Science frameworks have frequently limited the content targeted for the early grades to such simple and concrete ideas that they offer little explanatory power. Even pre-schoolers seek explanation".[6] Einstein-First has demonstrated, through numerous studies and peer reviewed published results that Einsteinian physics can be successfully integrated into a curriculum at all levels of schooling from Year 3 to Year 10[7–11] when taught at the appropriate level. These results concur with the statement of the distinguished Harvard and Oxford educator Jerome Bruner that: "Any subject can be taught effectively in some intellectually honest form to any child at any stage of development".[12] In a spirit of Bruner, this project is designed to test the introduction of general relativity and quantum mechanics to children in a manner that will be powerful, fun and mind-expanding.

Another reason for the opposition to the teaching of Einsteinian physics in primary and lower secondary school is the false belief that advanced mathematics is a prerequisite for understanding modern physics. While advanced mathematics is not required, new mathematical skills are important. To allow children to understand the scales of the universe from atoms to galaxies, we need to also rethink mathematics education. Understanding the main concepts of Einsteinian Physics requires the development of students' mathematical concepts in three important ways: a) understanding the vast range of scales of the universe, which necessitates the development

of logarithmic thinking[13] b) the development of probabilistic thinking,[14] and c) the development of vector understanding.[15] These skills also have importance for understanding the financial world, global issues such as environmental problems, gambling and risks, and skills such as drone navigation.

Overall, education needs to remain relevant in order to impart knowledge of the modern world to the new generation. Furthermore, physics practitioners, who are at the forefront of modern investigations and discoveries, should be part of the educational partnership to ensure that contemporary scientific trends find their way into the school curriculum at every level. For that reason, "Sharing our best understanding of reality with kids" is the driving motivation of the Einstein -First project.

Our vision is that every student, by Year 10, the last compulsory year of science in Australia, will have knowledge of our best understanding of the nature of space, time and gravity, the quantum world of atoms and photons, and the universe from quarks to the big bang. We want to give students the language of science so that no one need suffer from science illiteracy. We want students to know the epic stories of the struggle to understand the universe, which we teach through plays, songs and games. We want two common phrases to become obsolete: "I can't understand science" and "You must be a genius to understand that stuff".

In 2019 the Australian Research Council funded a five-year project to develop an integrated Einsteinian Physics-focused curriculum across primary and secondary schools. An early outcome has been the book *Teaching Einsteinian Physics in Schools*, published in August of 2021.[16] To produce this book, we had to find out if young children could understand Einsteinian concepts. To do this we developed activities that would make Einsteinian concepts intuitively understandable. Then we conducted trials and gathered evidence, which showed that young children comprehended and enjoyed learning Einsteinian concepts.[17] The successes of the Einstein-First program has been demonstrated also through studies on; long-term impact,[18] gender benefits,[19] advantages for underachieving students[7,9] and positive teacher response.[20] Particularly with further trials we made three specific exciting discoveries:

(1) the less-academic students developed knowledge and understanding on a par with the more-academic students[7,9],
(2) students' attitudes to science improved strongly[17] and
(3) girls entered programs with less knowledge than boys but on completion had performance on a par with boys when learning more difficult concepts.[19]

The goal of Einstein-First is to modernize school science so that all students by Year 10, should have a basic understanding of the science behind modern technology, and the universe, from its tiniest components like atoms to vast structures like our Milky Way galaxy.

2. Einstein–First Overview

2.1. *Einsteinian physics*

In the early 1900's, discoveries by Albert Einstein and others revolutionised our understanding of physical reality. Evidence of the existence of photons led to quantum theory, and evidence that space is curved confirmed Einstein's general theory of relativity. Today, the theories we term Einsteinian Physics have been tested to exquisite precision, with the crowning glory being the recent detections of gravitational waves from colliding black holes and neutron stars, the 2017 Nobel Prize announcement and the three million dollar Breakthrough Prize shared by the gravitational waves discovery team, including four members of this project. The Einsteinian conceptions of space, time, gravity, light, energy and matter represent our deepest understanding of the Universe. Einsteinian Physics unleashed the technological revolution that underpins modern civilisation, forming the basis for computers, mobile phones, GPS and more. Gravitational wave detectors themselves harness concepts of space as a rippling, stretchable fabric, as well as the quantum properties of light and quantum uncertainty, which underpins their exquisite measurement technology.

The Einstein-First project aims to introduce Einsteinian Physics starting with Year 3. For this age, we focus on Einsteinian concepts of understanding the atomic nature of matter (Brownian motion), Einstein's proof that light comes as photons, photons have momentum and that heat in solids fluctuates as phonons (1917).

2.2. *Educational contradictions*

Einstein and his collaborators revolutionized our understanding of space, time, light, matter and gravity. Einsteinian Physics, developed about 100 years ago, has extraordinary explanatory power, and has enabled the development of technologies that revolutionised our lives such as mobile phones, cameras, solar panels, and medical imaging. Despite the power and importance of Einsteinian physics, children from Primary to Year 10 are taught 19^{th} century physics concepts in school, sometimes implicitly and sometimes explicitly. Hence the modern understanding of physical reality is withheld from the majority of students. The few students who do specialise in physics in Year 11 and Year 12 have to learn a new paradigm. This situation is grossly inefficient and leads to the general view that science is a domain for experts alone.

2.3. *Children's awareness*

Children pick up bits of modern science from media, the internet, and children's books but at school almost all of this relevant, modern and exciting science is withheld. From an early age most have heard of time warps, black holes and $E = mc^2$, but school learning contradicts all this. They currently teach that time is absolute, geometry is Euclidean, and mass does not change with energy. Children are aware that most of their school science is "old stuff" because there is more

modern science on *The Simpsons* than there is in the classroom. Hence students see school science as being boring and irrelevant.

Many teachers and schools do wonderful things with robots, coding and more, to enhance STEM education but usually within the outdated curriculum. We realized that we needed to address the fundamentals and start at primary school when students are best able to assimilate the new language of Einsteinian physics.

2.4. *How the contradictions in science education arose*

New theories that change our way of thinking are always hard to get used to. When the new physics was discovered, people claimed that the new theories were much too complex for ordinary people to understand. Physicists did little to dispel these views and curriculum designers accepted them. This combined with early views of education theorists (now dispelled by recent research[21,22]) that children cannot learn abstract concepts. In the 1930s atoms, photons, electrons and curved space were all newly discovered concepts that were inferred but not directly observed.

However, in the modern era all the things that were conceptual have changed their status: they are now observed, imaged and part of observational reality, and a few clicks away on the internet. Because physical science underpins everything from biology to plate tectonics, the lack of conceptual foundations means that the key motivational content of science — the ongoing stream of scientific discoveries – is not taught and hence students adopt a negative attitudes towards existing science curricula. Also, the few students who specialise must suffer an abrupt shift in paradigm from Newtonian to Einsteinian physics that makes their learning more difficult.

2.5. *International partnerships*

There is widespread concern that school curricula do not meet the needs of modern children. The development of the Einsteinian Physics curriculum started with a collaboration of Australian and multinational teams. Currently, there is a lot evidence from different groups around the world, including our partner researchers, that middle school and even primary school children can understand concepts of mass and gravity,[23] theory of relativity,[24] theory of atoms[25,26] and the connection between wavelength and momentum.[27] However there is no evidence in the literature that any school jurisdictions in the world are embracing the need for modernising the curriculum. This modernization includes both the physics concepts and the associated mathematical concepts needed to understand the Universe.

2.6. *A contradiction-free curriculum*

We are working to overcome the deep-seated conflict between children's learning in schools and the current science that underpins all modern technology. During the years of rapid childhood learning when minds are most fluid, children are learning obsolete 19^{th} century concepts that contradict modern science. Awesome discoveries

are withheld because both the concepts and language are inaccessible to young learners. Here are some examples:

(1) In Year 3 students learn about heat without reference to atoms because atoms are considered too conceptual, and not to be taught until children are 12-14 years old.
(2) Children are taught that mass is always conserved in chemical reactions despite the fact that recent conceptually simple experiments are able to measure the mass loss of an atom when it emits a photon, a consequence of Einstein's $E = mc^2$.
(3) Students learn implicitly that time is absolute even though our GPS navigators must correct for the time warp between Earth and the satellites' orbits.

The modernization of the school curriculum requires the integration of new knowledge with innovative approaches. To identify a new learning methods and activities, the Einstein-First project encompasses two key ideas: a) "never underestimate the child's abilities to absorb new information and cope with new ideas",[28] and b) the central concept of the theory of constructivism in education that is "complex ideas can be taught at a simplified level first, then re-visited at more complex levels later on".[12]

Our learning framework is divided into the three stages.

1. Primary School: Opening minds to Einsteinian concepts and creating expectations. We use qualitative and experiential activity-based learning to emphasise the language of Einsteinian Physics.
2. Middle School: Basic understanding of Einsteinian concepts, using semi-quantitative activity-based learning, including the use of graphs and simple calculations.
3. High School: Developing a quantitative understanding of Einsteinian concepts, as well as a quantitative understanding, that includes calculations.

Einstein-First is creating a seamless learning progression that begins in early primary school and is designed to ensure that every student develops an intuitive understanding of physical reality. Currently, Einstein-First is developing curricula for every school level from Year 3 to Year 10 as well as introducing the language of modern science in Year 2. The modules of curriculum are presented below:

Year 3	Atom frenzy and Hot stuff
Year 4	May the forces be with you
Year 5	Fantastic photons and Our place in the universe
Year 6	Electric world
Year 7	Warp spacetime: Gravity from the Earth to black holes
Year 8	From mc^2 to renewable energy
Year 9	Quantum world: quarks to photons
Year 10	Climate change to cosmology

2.7. *How have we made Einsteinian physics accessible? Models, toys, analogies and drama*

Einsteinian physics is imagined by most primary and secondary school teachers to be impossibly difficult. This is because those who learnt the 19^{th} century paradigm cannot envisage ideas like space having a shape, and matter having properties of both waves and particles. Children, however, accept Einstenian concepts with ease when taught with appropriate models and activities.

In order to make modern physics understandable we use multiple, interactive methods that are discussed in detail here.

2.7.1. *Activities and toys*

Most of the universe is beyond the reach of our normal senses it is either, too big, too small or invisible. We need to find creative analogies and models so that concepts and ideas can become tangible. We understand and remember things not just with abstract thought, but with our bodies, by doing things. Furthermore, for all children toys are critical means of learning.

Our diversity of toys and models are simple, robust, and inexpensive. They allow students to learn complex concepts intuitively, mainly through group activities. For example, students roll steel balls around a central mass on a stretched lycra sheet (often called a space-time simulator). The 2D lycra sheet is our toy space-time. The balls orbit the central mass due to the curvature of the membrane caused by the central mass. These orbits resemble the orbit of planets around stars. The greater the central mass, the greater the curvature of space-time and the stronger the force of gravitation exerted on the orbiting bodies.

2.7.2. *Crossing boundaries, playing Einstein*

The discovery of Einsteinian physics was the culmination of centuries of enquiry by scientists all over the world. There were intellectual battles between scientists, notably between holists and atomists in the 19th century. The holists who believed that matter was smooth and continuous forever, lost the battle when conclusive evidence was found for electrons (1897), photons (1905) and atoms (1911). Bringing these human stories to life helps students understand the progress and process of science and shows science as an epic human endeavour. In our reading plays, often performed in drama lessons, inquisitive students mix with scientists from across the ages, enacting the process of discoveries. Characters such as Einstein, sing songs and recite aphorisms that help cement the Einsteinian vocabulary with fun and humor.

2.7.3. *Mathematics for the modern world*

To allow children to understand the scales of the universe from atoms and galaxies, we need to rethink mathematics education. Modern science understanding depends

on fundamental mathematical areas that go beyond primary school arithmetic. Einstein-First has introduced early activity-based learning in three crucial areas: huge and tiny (powers of ten), maths of arrows (graphical vectors), and maths of chance. In our use of mathematics we are creating seamless learning progressions that begins in early primary school, designed to ensure that everyone develops an intuitive understanding of concepts crucial for understanding, for example, the size of the whole observable universe, the electromagnetic spectrum and the quantum world. This mathematics also has social importance for understanding money, weather, risks and gambling. Starting at the youngest ages, we teach these topics with fun activities designed to open children's minds to ideas that mathematics is more than numbers, maths is easy, and maths is fun.

3. Methods of Einstein-First

In this section we present examples of four activities of Einstein-First methods to deliver concepts of modern science and associated mathematics concepts of probability, big and small numbers, and vectors.

3.1. *Early learning of powers of 2 and powers of ten by cutting up tape measure*

Exponential notation is essential to our ability to comprehend the scale of the Universe from the tiniest component to the Hubble volume. To introduce exponential notations we use a paper cutting activity. Students cut a one meter paper tape measure into repeated halves. Then they cut another one meter paper tape measure into tenth to learnt that ten powers of two are approximately three powers of ten. Then this exercise is mentally extrapolated to find how many times this process needs to be continued to reach the size of the atom.

3.2. *The entire Universe in a single book*

To introduce big and small numbers and the scale of the Universe we have developed a novel long-term activity consisting of the creation of a 130 page book called *Powers of the Universe*. All numbers describing the known Universe can be placed in this book. Each page is a power of ten. Line position on each page is designated for various quantities, all expressed in SI units. For example there are line bands for distances, masses, energy etc. Students use all possible methods (e.g., using Google) to discover numbers at all scales. For example, they start the book at the page labelled -37, while they indicate, that 10^{-37} is the rest mass of the neutrino. Page +90 shows the total number of photons in the Universe and page +2 is the mass of a human.

Through this activity, students investigate the scale of our universe and expand their knowledge and their ability to envisage big and small numbers.

3.3. *Children pushing a car to discover vector addition*

In order to introduce the concept of vectors at early age we use activity in which students push their teacher's car. Students in the class altogether push a car on a flat car park under the supervision of a class teacher. Teachers use creative questioning to encourage all students to contribute answers about what happened and why it happened. In this way, students represent the pushing forces by vectors and learn rules of head to tail adding vectors. Then we introduce the concept of the resultant. To reinforce the learning, students use magnetic arrows on magnetic boards to investigate the sum of the vectors from where they find out that the resultant is not dependent on an order of adding arrows.

3.4. *A boy and his atom. Feeling hidden atoms pattern*

The world's smallest movie *A boy and his atom* is used to introduce atoms and molecules, and how we are able to see and move them. Students watch a short video from IMB[29] and discuss how the Atomic Force Microscopy works. Students work in pairs. One student makes the image using playdough and toy atoms in a form of marbles and covers it with a piece of cloth. Then the other student determines the hidden pattern by feeling in the same way that an atomic force microscope "feel" when it scans a surface. Finally, the student draws a picture of the pattern.

All of these activities mentioned above along with others were included in recent short programs for students from Year 3 to 8.

4. Four Recent Programs

The Einstein-First project has held short programs in schools over the previous six months to teach students about atoms and molecules, heat, forces, light and gravity. Group activities, hands-on learning and team competitions were all parts of the programs. Students' attitudes and knowledge were assessed using pre- and post-questionnaires that students completed on the first and last day of the programs. Assessments of the attitudes and knowledge were administrated for 15 minutes before and after the programs. Every lesson was delivered using the following method:

a) The first 15 minutes were used to review what they had learnt in the previous lesson and to introduce the idea they would be learning that day;
b) The main part of the lesson was dedicated to the associated activity and concepts; and
c) The final 15 minutes were set out for discussion or completion of the worksheets.

4.1. *Discovering black holes*

This four-week, eight-hour program was designed for motivated Year 7 and Year 8 students from various schools. This program had a total of 20 students, including a

6-year-old gifted student). The program was presented by current PhD students involved in the Einstein-First project. Each two hour session consisted of Einsteinian physics concepts as well as the mathematics that accompanied them. By doing activities on a space-time simulator (see section 2.7.1), students visualized the concept of gravity, black holes, and deflection of light. Students summarised their learning with the following phrase *"Space-time tells matter how to move; matter tells space-time how to curve."*

All of the above activities, as well as others in the program, require the development of mathematical skills to comprehend both big and small numbers. Understanding big and small numbers aids in comprehending the nature of supermassive black holes and how little displacements in the interferometer arms are essential to detect gravitational waves. For example the activity described in Section 3.1 was used for understanding powers of ten for these numbers.

4.2. *How big is an atom? How small is a light year?*

Twenty six Year 5 and 6 students participated in a four-session program on big and small numbers. The lessons were delivered by a mathematics teacher in the presence of the PhD student who developed the program. Each session lasted 1 hour, teaching mainly mathematics associated with Einsteinian physics. The main goal of the program was to teach students about the scale of the universe and how to use powers of ten to understand big and small numbers as described in Sections 3.1 and 3.2. Students also took part in a reading play called "Ten times Alice" in which the main character eats magic cookies or a magic liquid to grow ten times smaller or bigger than her companions. This play was created for students to practice proportional reasoning and multiplication by 10.

4.3. *Vectors and forces*

This short program was aimed at Year 4 students. The main goal of the program was to give an understanding of how vectors are used in physics and addition of vectors. There were 2.5 sessions that were presented to 27 students. The program was presented by the leader of the Einstein-First project with assistance of a current PhD student. Each session lasted 90 minutes and encompassed concepts of forces and geometry associated with them. We used the activity presented in Section 3.3 to practice vector addition and the concept of a resultant. After, children applied the ideas of vectors pulling wooden blocks with rubber bands in different direction to find the final movement (Fig. 1). The program also included activities with balance, for example, balancing a big wooden ruler horizontally on one finger to understand conception of the resultant.

4.4. *Atom frenzy*

Twenty-four Year 3 students participated in the 12-session atoms and molecules program delivered by the project's senior member and Phd student. Each session was

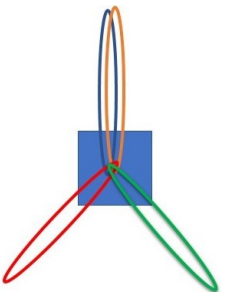


Fig. 1. Each coloured loop above represents a rubber band being pulled by an external force. Students are tasked with finding the resultant of four forces applied to wooden block with rubber bands as detailed in Section 4.3.

90 minutes and delivered through engaging, hands-on activities. We used various methods of understanding of atoms from children acting as atoms to boxes of balls that represent atoms, states of matter and heat. Students learnt three molecules H_2O, CO_2, O_2, N_2, using tennis and ping-pong balls with magnets inside representing toy atoms.

5. Results and Discussion

The section presents some outcomes of the programs described in Section 4.

5.1. *Results from the programs "Discovering Black Holes" and "How big is an atom? How small is a light year?"*

Students' understanding of the exponential function, writing numbers in scientific notations, and calculations significantly improved. Fig. 2 presents the results from the programs described in the sections 4.1 and 4.2. We evaluated students' choice to write big numbers such as *a) one hundred b) one million c) one trillion and d) one billion* before and after interventions. We discovered that just 26% of students chose powers of ten to write the numbers before the program. However, after the program, all students were able to write the numbers in powers of ten accurately. According to the research findings from program 4.2, 73% of participants chose to write numbers in powers of ten, compared to only 12 percent at the start.

Overall, we found that primary school students were capable of using powers of ten in reasoning about big numbers.

5.2. *Results from the programs "Forces and vectors"*

Results from the program 4.3 is shown in Fig. 3. A question was asked about force vectors and resultant to evaluate what they had learnt from the car-pushing activity (3.3). More than 80% of students accurately added vectors after the program.

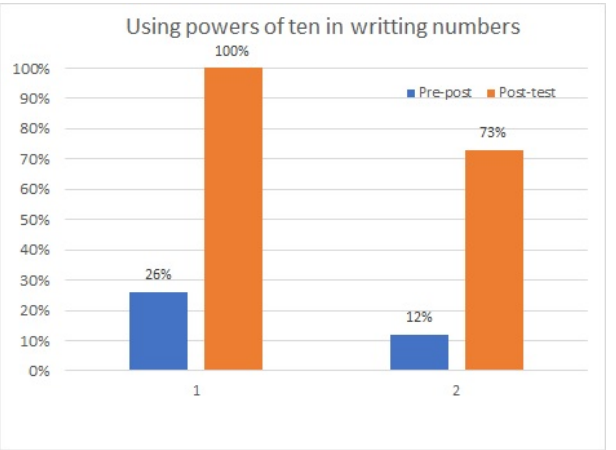


Fig. 2. The percentage of students using powers of ten in writing numbers before and after programs 1) 4.1 and 2) 4.2.
Q: Using powers of ten (or otherwise) write the numbers below: a)one hundred b) one million b) one billion c) one trillion.

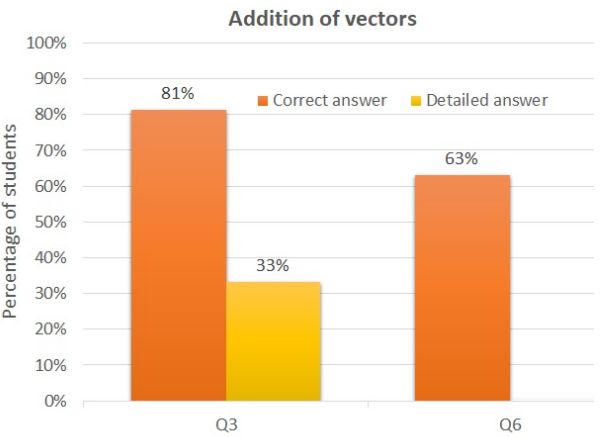


Fig. 3. The percentage of students, giving correct answers on the questions about arrows addition after program Vectors and Forces.
Q3: Eight children pushed the teacher's car, but seven children pushed in the opposite directions. Draw the resultant arrow? What was the total Force?
(Orange light indicates that one answer is correct. Yellow indicates that both answers are correct and detailed picture is provided.)
Q6: Four kids pull the wooden block in different direction using four rubber bands (the picture was provided) In what direction will the block move? Draw an arrow for the direction.

However, only 33% of students were able to determine the resultant of two opposing forces.

Q6 is an extension of Q3, indicating that 63% of students can add four arbitrary vectors to determine a resultant applied to wooden block with rubber bands (Fig. 1)

and its final movement. The graph shows that the vast majority of student can grasp the notion of vectors and add them together.

5.3. *Results from the programs "Atom Frenzy"*

The results from the program 4.4 are shown in Fig. 4. Despite the fact that the Australian curriculum stipulated that atoms should not be introduced until the age of 14, 40% of students had a basic understanding of atoms prior to the program. Figure 3 indicates that 83% of students were able to answer *what is an atom?* after the program, compared to only 40% of students before the program. In addition, 75% of students expanded their answers by referencing atom structure, naming subatomic particles, or thinking them to be a part of molecules. Also 87.5% of students correctly identified atom number six as a carbon atom, drew its structure, and presented it as a component of carbon dioxide.

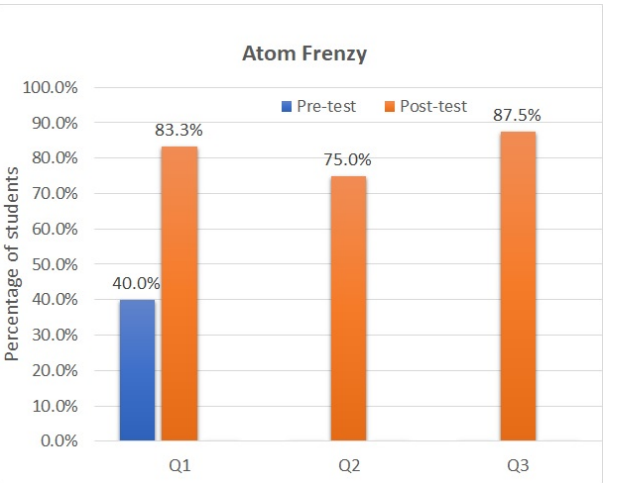


Fig. 4. The percentage of students, giving corrected answers on the questions about atoms before and after program Atom Frenzy?
Q1: Describe what you think an atom is.
Q2: Write down three things you know about atoms and the things they are made of.
Q3: Draw a picture of atom number 6.

Overall, outcomes of programs show that, the students demonstrated a considerable gain in their grasp of atoms, vectors, and the use of powers of ten as a technique of manipulating big and small numbers.

6. Conclusion

We have emphasized the vital necessity to rethink the science curriculum and mathematics associated with it. It has become obvious to introduce ideas that students

find interesting and engaging especially because of the widespread decline of students' attitude to science. Einstein-First is an innovative experimental education project rethinking the physical science curriculum by developing an innovative approach for introducing Einsteinian Physics in school starting at an early age. We have given the general principle of our approach and examples of methods used to make the physics accessible to teachers and students alike. We have presented some results from recent short interventions, that have shown that students aged 11 to 14 years have shown competency with power of ten notation. Secondly, we have demonstrated that students aged 10 obtain high scores in understanding the concept of vectors, including vectors addition. Within the larger Einstein-First project, learning progressions of curricula have been created, starting with Year 3 and going up to Year 10. Einstein-First also considers Year 2 to be the pre-learning stage for modern physics vocabulary. Models and analogies, drama plays, and mathematical support modules are suggested to make concepts of modern physics digestible and fun to learn. It has been demonstrated that even short interventions can bring high outcomes in learning Einsteinian physics in both primary and early-secondary schools.

Acknowledgments

We very much thank the anonymous teachers, who facilitated and taught in the programs presented here. This research was supported by Australian Research Council Linkage Grant LP 180100859.

References

1. https://k10outline.scsa.wa.edu.au/home/teaching/curriculum-browser/science-v8.
2. R. Weiss, *Reviews of modern Physics* **90** (2018) 1.
3. B. Abbott et al., *Physical Review Letters* (2016) 131103.
4. https://www.nobelprize.org/prizes/physics/2021/press-release/.
5. P. Potvin, A. Hasni, *Journal of Science Education and Technology* **23** (2014) 784.
6. K. E. Metz, *Cognition and Instruction* **22** (2004) 219.
7. M. Pitts, G. Venville, D. Blair, M. Zadnik *Research in Science Education* **44** (2014) 363.
8. T. Kaur, D. Blair, J. Moschilla, W. Stannard, M. Zadnik, *Physics Education* **52** (2017) 065014.
9. T. Kaur, D. Blair, W. Stannard, D. Treagust, G. Venville, M. Zadnik, W. Mathews, D. Perks, *Research in Science Education* **50** (2019) 2505.
10. R. Choudhary, U. Kraus, M. Kersting, D. Blair, C. Zahn, Marjan Zadnik, *The Physics Educator* **1** (2019) 1950016.
11. R. Choudhary, D. Blair, *European Journal of Physics* **42** (2021) 035408.
12. Jerome Bruner, *The Process of education* (Harvard University Press, Cambridge, Manchester, London 1976).
13. S. Mahajan, *American Journal of Physics* **86** (2018) 859.
14. D. I. Sari, I. K. Budayasa, D. Juniati, Probabilistic Thinking of Elementary School Students in Solving Probability Tasks Based on Math Ability, *AIP Conference Proceedings 1867*, 2017, 1867, p. 020028.

15. W. Widada, D. Herawaty, Y. Beka, R. M. Sari, R. Riyani, K. Umam, Z. Nugroho, *Journal of Physics Conference* **1470** (2020) 012071.
16. M. Kersting, D. Blair, *Teaching Einsteinian Physics in Schools: An Essential Guide for Teachers in Training and Practice* (Taylor and Francis Group, 2021).
17. R. K Choudhary, A.Foppoli, T. Kaur, D. G. Blair, M. Zadnik, R. Meagher, *Physics Education* **53** (2018) 065020.
18. K. Adams, R. Dattatri, T. Kaur, D. Blair, *Physics Education* **56** (2021) 055031.
19. T. Kaur, D. Blair, R. K. Choudhary, Y. S. Dua, A. Foppoli, D. Treagust, M. Zadnik, *Physics Education* **55** (2020) 035029.
20. A. Foppoli, R. Choudhary, D. Blair, T. Kaur, J. Moschilla, M. Zadnik, *Physics Education* **54** (2019) 015001.
21. K. E. Metz, *Review of Educational Research* **65** (1995) 93.
22. National Research Council, *Taking science to school: Learning and teaching science in grades K-8*, (National Academies Press, 2007).
23. M. Luca Ruggiero, S. Mattiello, M. Leone, *Physics Education* **56** (2019) 065011.
24. M. Kersting, G. Schrocker, S. Papantoniou, *International Journal of Science Education* **43** (2021) 2044.
25. C. Haeusler, J. Donovan, *Research in Science Education* **50**, 2 (2020).
26. A. Mandrikas, E. Michailidi, D. Stavrou, *Research in Science Technological Education* **38** (2020) 337.
27. P. Bitzenbauer, J.-P. Meyn, *Physics Education* **55** (2020) 055031.
28. D. Whitebread, P. Coltman, *Teaching and Learning in the Early Years* (Taylor and Francis, United Kingdom, 2015).
29. https://research.ibm.com/articles/madewithatoms.shtml.
30. https://www.einsteinianphysics.com/.

Kundt spacetimes in the Einstein–Gauss–Bonnet theory

R. Švarc*, J. Podolský† and O. Hruška

Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University, Prague, V Holešovičkách 2, 180 00 Praha 8, Czech Republic
*E-mail: *robert.svarc@mff.cuni.cz, †podolsky@mbox.troja.mff.cuni.cz*
www.mff.cuni.cz

Based on our recent results we present the complete class of vacuum solutions in the Einstein–Gauss–Bonnet gravity which admit non-expanding, shear-free and twist-free null geodesic congruence and thus form the Kundt family of geometries. We explicitly derive the field equations and classify their solutions into three distinct subfamilies. Algebraic structure of the Weyl tensor is determined and using the corresponding scalars entering the invariant form of geodesic deviation equation we discuss the specific local physical properties of the gravitational field constrained by the EGB theory. Moreover, we analyze interesting subclasses of such vacuum solutions, namely all geometries with constant-curvature transverse space, and the class of *pp*-waves admitting a covariantly constant null vector field.

Keywords: Einstein–Gauss Bonnet theory; Kundt spacetimes; exact solutions; *pp*-waves.

1. Introduction

The Einstein–Gauss–Bonnet (EGB) theory represents the first and simultaneously the simplest non-trivial contribution of the Lovelock gravities[1] extending the classic Einstein–Hilbert action into higher dimensions $D \geq 5$ and keeping the second order field equations. This theory also arises in various scenarios, e.g., as the heterotic string theory limit for low energies.[2,3]

In particular, the EGB vacuum action in D-dimensions reads

$$S = \int \left[\kappa^{-1}\,(R - 2\Lambda_0) + \gamma\, L_{GB}\right] \sqrt{-g}\;\mathrm{d}^D\mathrm{x}\,, \tag{1}$$

with the Gauss–Bonnet term defined as

$$L_{GB} \equiv R_{cdef}\,R^{cdef} - 4\,R_{cd}\,R^{cd} + R^2\,. \tag{2}$$

Here R stands for the Ricci scalar, R_{cdef} and R_{cd} are Riemann and Ricci tensors, respectively, Λ_0, κ and γ are theory constants. Obviously, for $\gamma = 0$ the standard Einstein–Hilbert action of the general relativity is restored. The least action principle, i.e., variation $\delta S = 0$, leads to the field equations in the form

$$R_{ab} - \tfrac{1}{2} R\, g_{ab} + \Lambda_0\, g_{ab} + 2k\, H_{ab} = 0 \qquad \text{where} \qquad k \equiv \kappa\gamma\,, \tag{3}$$

and

$$H_{ab} \equiv R\,R_{ab} - 2R_{acbd}\,R^{cd} + R_{acde}\,R_b{}^{cde} - 2R_{ac}\,R_b{}^{c} - \tfrac{1}{4}\,g_{ab}\,L_{GB}\,. \tag{4}$$

The aim of this proceeding contribution is to briefly summarize main results of Ref. 4, where we explicitly derived and analyzed the second order field equations (3)

in the case of so-called Kundt geometries,[5,6] see e.g. Ref. 7 for detail list of references, and compared properties of obtained EGB solutions with those extensively studied within general relativity.[8–11]

The Kundt geometries are defined via geometric assumption on existence of a specific affinely parametrized null geodesic congruence. In general, transverse deformations of such null congruence, generated by null vector field $\boldsymbol{k}$, are described by the optical scalars which characterize its twist, shear, and expansion, namely

$$A^2 = -k_{[a;b]}k^{a;b}\,, \qquad \sigma^2 = k_{(a;b)}k^{a;b} - \tfrac{1}{D-2}(k^{a}_{\ ;a})^2\,, \qquad \Theta = \tfrac{1}{D-2}k^{a}_{\ ;a}\,. \tag{5}$$

Then, the Kundt spacetimes are those which admit the congruence with all above optical scalars vanishing, i.e., we deal with the non-twisting, shear-free, and non-expanding geometries. Simultaneously, presence of the non-twisting vector field $\boldsymbol{k}$ implies existence of the spacetime null foliation[8] with $\boldsymbol{k}$ being normal (tangent). In such a case, it is straightforward to introduce adapted set of coordinates $\{r,\, u,\, x^p\}$ in D dimensions, where $u = \text{const}$ labels null hypersurfaces generated by $\boldsymbol{k}$, r is an affine parameter along the non-twisting null congruence, i.e., $\boldsymbol{k} = \partial_r$, and the $(D-2)$-dimensional transverse subspace with $u = \text{const}$ and $r = \text{const}$ is covered by spatial coordinates x^p with $p = 2, \ldots, D-1$. Using these adapted coordinates the general non-twisting line element can be written as

$$\mathrm{d}s^2 = g_{pq}(r,u,x)\,\mathrm{d}x^p\mathrm{d}x^q + 2g_{up}(r,u,x)\,\mathrm{d}x^p\mathrm{d}u - 2\mathrm{d}u\mathrm{d}r + g_{uu}(r,u,x)\,\mathrm{d}u^2\,, \tag{6}$$

where g_{pq} represents metric of the transverse space. Moreover, the shear-free condition implies

$$g_{pq} = \rho^2(r,u,x)\,g_{pq}(u,x) \qquad \text{with} \qquad \rho = \exp\Big(\int \Theta(r,u,x)\,\mathrm{d}r\Big)\,. \tag{7}$$

Finally, the non-expanding condition leads to the r-independent transverse space metric, i.e., $g_{pq} = g_{pq}(u,x)$. The most general Kundt line element can thus be expressed as

$$\mathrm{d}s^2 = g_{pq}(u,x)\,\mathrm{d}x^p\mathrm{d}x^q + 2g_{up}(r,u,x)\,\mathrm{d}x^p\mathrm{d}u - 2\mathrm{d}u\mathrm{d}r + g_{uu}(r,u,x)\,\mathrm{d}u^2\,. \tag{8}$$

Moreover, in our contribution[4] we assume simplification corresponding to $g_{up} = 0$, i.e., our metric ansatz takes the form

$$\mathrm{d}s^2 = g_{pq}(u,x)\,\mathrm{d}x^p\mathrm{d}x^q - 2\mathrm{d}u\mathrm{d}r + g_{uu}(r,u,x)\,\mathrm{d}u^2\,, \tag{9}$$

where the metric functions g_{pq} and g_{uu} have to be specified by the field equations (3).

Finally, let us briefly mention the most important members of the Kundt family of spacetimes studied within Einstein's general relativity. Definitely, the prominent role is played by spacetimes admitting a covariantly constant null vector field, which can be associated with $\boldsymbol{k}$, leading to all optical scalars vanishing. These *pp*-wave spacetimes were introduced by Brinkmann[12] in the context of conformal mapping. On the level of Kundt line element (8) this corresponds to all metric functions being r-independent. The spacetimes providing vanishing scalar curvature invariants of all

orders (VSI spacetimes)[13] also belong to the Kundt class and their metric functions have to be of the form

$$g_{pq} = \delta_{pq}\,, \quad g_{up} = e_p(u,x) + r\, f_p(u,x)\,, \quad g_{uu} = a(u,x)\, r^2 + b(u,x)\, r + c(u,x)\,. \tag{10}$$

Moreover, the Kundt geometries include also various direct product spacetimes such as Bertotti—Robinson, Nariai, and Plebanski—Hacyan backgrounds, see e.g. Chap. 7 in Ref. 9, or toy models for null particles with internal spin called gyratons independently discovered by Bonnor[14] and Frolov.[15]

2. EGB field equations in the Kundt case

Let us analyze the specific constraints on the Kundt geometry (9) implied by the EGB field equations (3). These can be rewritten in alternative form

$$R_{ab} = \frac{2\Lambda_0}{D-2} g_{ab} - 2k\left(H_{ab} - \frac{g_{ab}}{D-2} H\right) \qquad \text{with} \qquad H = g^{ab} H_{ab}\,. \tag{11}$$

Calculating all the curvature tensors we find that the rr and rp components of (11) are identically satisfied and the only conditions we have to discuss are

$$R_{ru} = -\tfrac{2\Lambda_0}{D-2} - 2k\left(H_{ru} + \tfrac{1}{D-2} H\right), \qquad R_{up} = -2k\, H_{up}\,,$$
$$R_{pq} = \tfrac{2\Lambda_0}{D-2}\, g_{pq} - 2k\left(H_{pq} - \tfrac{g_{pq}}{D-2} H\right), \qquad R_{uu} = \tfrac{2\Lambda_0}{D-2}\, g_{uu} - 2k\left(H_{uu} - \tfrac{g_{uu}}{D-2} H\right).$$

In particular, we obtain:

- ru-component connects the transverse space geometry described by the metric $g_{pq}(u,x)$ with the theory constants Λ_0 and k, namely

$$ {}^S\!R - 2\Lambda_0 + k\left({}^S\!R^2_{klmn} - 4\, {}^S\!R^2_{mn} + {}^S\!R^2 \right) = 0\,, \tag{12}$$

  where the superscript ^{S} denotes quantities calculated using the $(D-2)$-dimensional metric g_{pq}. Moreover, in the general relativity limit corresponding to $k = 0$ the transverse space Ricci scalar ${}^S\!R$ has to be a constant, i.e., ${}^S\!R = 2\Lambda_0$.

- pq-component generically provides relation between the transverse space geometry and the r-dependence of the metric function g_{uu} via relation

$$\begin{aligned} &Q_{pq}\, g_{uu,rr} + {}^S\!R_{pq} \\ &\quad + 2k\big({}^S\!R_{pq}\, {}^S\!R - 2\, {}^S\!R_{pmqn}\, {}^S\!R^{mn} \\ &\qquad + {}^S\!R_{pklm}\, {}^S\!R_q{}^{klm} - 2\, {}^S\!R_{pm}\, {}^S\!R_q{}^m \big) = 0\,, \end{aligned} \tag{13}$$

  where Q_{pq} is a *fundamental quantity* important for the further discussion defined as

$$Q_{pq} \equiv -\tfrac{1}{2}\, g_{pq} + k\left(2\, {}^S\!R_{pq} - {}^S\!R\, g_{pq}\right). \tag{14}$$

  To obtain a particular constraint we evaluate the *trace* of (13), namely

$$-Q\, g_{uu,rr} = 4\Lambda_0 - {}^S\!R \tag{15}$$

with

$$Q \equiv g^{pq}\, Q_{pq} = -\left[\tfrac{1}{2}(D-2) + k(D-4)\, {}^{S}R\right]. \tag{16}$$

Obviously, the further discussion has to be split into three distinct cases, namely $Q \neq 0$, and $Q = 0$ with $Q_{pq} \neq 0$, or $Q_{pq} = 0$. In Einstein's theory we have $Q \neq 0$ and the condition (15) can be simply integrated to obtain the specific r-dependence of g_{uu}.

- *up*-component represents constraint on the spatial dependence of the metric function g_{uu} combined with the admitted u-dependence of the transverse space metric g_{pq},

$$\begin{aligned} Q_{pn}\, g^{nm}\big(g_{uu,rm} - 2g^{kl} g_{k[m,u||l]}\big) \\ + 2k\big(-2\, {}^{S}R^{kl}\, \delta^{m}_{p} + {}^{S}R_{p}{}^{kml}\big) g_{k[m,u||l]} = 0\,, \end{aligned} \tag{17}$$

where the symbol $_{||}$ denotes covariant derivative on the transverse space.

- *uu*-component restricts the possible amplitudes of Kundt gravitational waves typically encoded in the second covariant derivative $g_{uu||pq}$. The condition becomes

$$\begin{aligned} Q^{pq}\big(g_{uu||pq} + g_{pq,uu} - \tfrac{1}{2} g_{uu,r}\, g_{pq,u} - \tfrac{1}{2} g^{kl} g_{kp,u}\, g_{lq,u}\big) \\ + 2k\left(g^{ko} g^{ls} - 2\, g^{kl} g^{os}\right) g^{pq}\, g_{k[p,u||l]}\, g_{o[q,u||s]} = 0\,. \end{aligned} \tag{18}$$

The above equations represent a complete set of the constraints on the Kundt spacetime geometry implied by the EGB theory which have to be satisfied. In the following subsections we will discuss three particular cases defined with respect to the values of quantities Q, and Q_{pq}, respectively.

2.1. *Case $Q \neq 0$ – generic case*

In this non-degenerate case the pq-equation determines r-dependence of g_{uu} to be polynomial. Subsequently, the remaining equations are decoupled with respect to power in r. In particular,

- pq-equation trace (15) can be integrated to obtain

$$g_{uu}(r,u,x) = b(u,x)\, r^2 + c(u,x)\, r + d(u,x)\,, \tag{19}$$

where

$$b = \frac{4\Lambda_0 - {}^{S}R}{(D-2) + 2k\,(D-4)\, {}^{S}R}\,. \tag{20}$$

With g_{uu} in hand the pq-equation restrict the transverse space geometry,

$$\begin{aligned} &\big[(D-2) + 4k(D-4)\big(1 + k\, {}^{S}R\big)\, {}^{S}R + 16k\Lambda_0\big]\, {}^{S}R_{pq} \\ &- \big(1 + 2k\, {}^{S}R\big)\big(4\Lambda_0 - {}^{S}R\big) g_{pq} \\ &- 2k\big[(D-2) + 2k(D-4)\, {}^{S}R\big]\big(2\, {}^{S}R_{pmqn}\, {}^{S}R^{mn} - {}^{S}R_{pklm}\, {}^{S}R_{q}{}^{klm} \\ &\qquad + 2\, {}^{S}R_{pm}\, {}^{S}R_{q}{}^{m}\big) = 0\,. \end{aligned} \tag{21}$$

In general relativity ($k = 0$) this condition implies that the transverse space has to be the Einstein space, i.e.,

$$ {}^{S}R_{pq} = \frac{2\Lambda_0}{D-2}\, g_{pq}\,. \tag{22} $$

- *up*-component has a non-trivial part which determines the spatial dependence of $c(u,x)$ in (19) and couples it to the u-dependence of the spatial metric g_{km},

$$ \begin{aligned} & Q_{pn}\, g^{nm}\left(c_{,m} - 2g^{kl} g_{k[m,u||l]}\right) \\ & \qquad + 2k\left(-2\,{}^{S}R^{kl}\,\delta^{m}_{p} + {}^{S}R_{p}{}^{kml}\right) g_{k[m,u||l]} = 0\,. \end{aligned} \tag{23} $$

  Basically, in general relativity this equation plays very similar role.
- *uu*-component gives two independent conditions implied by its parts linear in r, and r-independent, respectively,

$$ Q^{pq}\big(c_{||pq} - b\, g_{pq,u}\big) = 0\,, \tag{24} $$

$$ \begin{aligned} & Q^{pq}\big(d_{||pq} - \tfrac{1}{2}\, c\, g_{pq,u} + g_{pq,uu} - \tfrac{1}{2} g^{kl} g_{kp,u}\, g_{lq,u}\big) \\ & \qquad + 2k\big(g^{mo} g^{ns} - 2\, g^{mn} g^{os}\big) g^{pq} g_{m[p,u||n]}\, g_{o[q,u||s]} = 0\,. \end{aligned} \tag{25} $$

  These constraints can be understood as restrictions on the spatial dependence of g_{uu}. In particular, it gives a possible amplitudes $d_{||pq}$ of the Kundt gravitational waves.

2.2. *Case* $Q = 0$ *with* $Q_{pq} \neq 0$

In principal, there may occur a peculiar situation in which $Q_{pq} \neq 0$ and simultaneously $Q = 0$. In such a case the condition $Q = 0$ implies a strict constraint on the transverse-space scalar curvature, namely

$$ {}^{S}R = -\frac{D-2}{2k(D-4)}\,, \tag{26} $$

which is a non-vanishing and constant. This case is clearly *not allowed* in the Einstein theory. Moreover, the trace of pq-equation gives

$$ {}^{S}R = 4\Lambda_0\,. \tag{27} $$

These two conditions combined together represent an interesting coupling of all theory parameters in the form

$$ 8(D-4)k\Lambda_0 = -(D-2)\,. \tag{28} $$

We can thus conclude that for any Gauss–Bonnet parameter $\gamma = k/\kappa$ there is a unique value of Λ_0. Simultaneously, parameters k and Λ_0 must have opposite signs and none of them can be vanishing. Finally, since $Q_{pq} \neq 0$, the pq-equation has to be satisfied for every pair p and q which again implies at most quadratic r-dependence of g_{uu} and thus all remaining equations are similar to those of the previous case.

2.3. *Case $Q_{pq} = 0$ implying $Q = 0$*

As in the previous case $Q = 0$ implies relation between the theory constants,

$$8(D-4)k\Lambda_0 = -(D-2) \quad \text{and} \quad {}^{S}R = 4\Lambda_0\,. \tag{29}$$

However, the condition $Q_{pq} = 0$ strongly constraints the transverse-space geometry,

$${}^{S}R_{pq} = \frac{1}{4k}\,g_{pq} + \frac{1}{2}\,{}^{S}R\,g_{pq}\,, \quad \text{i.e.} \quad {}^{S}R_{pq} = \frac{4\Lambda_0}{D-2}\,g_{pq} \equiv -\frac{1}{2k(D-4)}\,g_{pq}\,. \tag{30}$$

In this degenerate case the $(D-2)$-dimensional transverse space has to be the Einstein space. Interestingly, this subclass of EGB vacuum solutions does not allowed a straightforward limit to the Einstein theory. The constraints implied by the field equations become:

- pq-equation restricts the transverse-space Riemann tensor as

$${}^{S}R_{pklm}\,{}^{S}R_{q}{}^{klm} = 2\big[2k(D-4)\big]^{-2}\,g_{pq} \equiv -\frac{1}{k(D-4)}\,{}^{S}R_{pq}\,, \tag{31}$$

  and the ru-equation becomes identically satisfied.
- up-equation significantly simplifies to the form

$$\big(-\tfrac{8\,\Lambda_0}{D-2}\,\delta^{m}_{p}\,g^{kl} + {}^{S}R_{p}{}^{kml}\big)\,g_{k[m,u||l]} = 0\,. \tag{32}$$

- uu-component constraints only the transverse metric g_{pq} and its u-dependence as

$$\big(g^{ko}g^{ls} - 2\,g^{kl}g^{os}\big)g^{pq}\,g_{k[p,u||l]}\,g_{o[q,u||s]} = 0\,. \tag{33}$$

Interestingly, the metric component $g_{uu}(r,u,x)$ remains *arbitrary function of all spacetime variables*. Moreover, for the spatial metric g_{pq} being u-independent the equations (32) and (33) are identically satisfied.

3. Analysis of the solution

In this section we summarize two methods employed in analysis of the general solution derived above. In particular, we discuss the structure of the Weyl tensor and describe relative motion of free test particles, respectively.

3.1. *Algebraic structure*

The higher-dimensional algebraic classification can be formulated in terms of the boost-weight irreducible components of any tensor with respect to a suitable null frame.[7,16] Such a frame in the case of Kundt geometries (9) takes the form

$$\boldsymbol{k} = \partial_r\,, \qquad \boldsymbol{l} = \tfrac{1}{2}\,g_{uu}\partial_r + \partial_u\,, \qquad \boldsymbol{m}_i = m_i^p\,\partial_p\,. \tag{34}$$

The Weyl tensor components sorted by their boost weights with respect to the generic null frame $\{\boldsymbol{k}, \boldsymbol{l}, \boldsymbol{m}_i\}$ are

$$\begin{aligned}
\Psi_{0^{ij}} &= C_{abcd}\, k^a\, m_i^b\, k^c\, m_j^d\,, & & \\
\Psi_{1^{ijk}} &= C_{abcd}\, k^a\, m_i^b\, m_j^c\, m_k^d\,, & \Psi_{1T^i} &= C_{abcd}\, k^a\, l^b\, k^c\, m_i^d\,, \\
\Psi_{2^{ijkl}} &= C_{abcd}\, m_i^a\, m_j^b\, m_k^c\, m_l^d\,, & \Psi_{2S} &= C_{abcd}\, k^a\, l^b\, l^c\, k^d\,, \\
\Psi_{2^{ij}} &= C_{abcd}\, k^a\, l^b\, m_i^c\, m_j^d\,, & \Psi_{2T^{ij}} &= C_{abcd}\, k^a\, m_i^b\, l^c\, m_j^d\,, \\
\Psi_{3^{ijk}} &= C_{abcd}\, l^a\, m_i^b\, m_j^c\, m_k^d\,, & \Psi_{3T^i} &= C_{abcd}\, l^a\, k^b\, l^c\, m_i^d\,, \\
\Psi_{4^{ij}} &= C_{abcd}\, l^a\, m_i^b\, l^c\, m_j^d\,. & &
\end{aligned} \tag{35}$$

Moreover, their irreducible components identifying specific algebraic subtypes are defined[16] as

$$\begin{aligned}
\tilde{\Psi}_{1^{ijk}} &= \Psi_{1^{ijk}} - \tfrac{2}{D-3}\delta_{i[j}\Psi_{1T^{k]}}\,, \\
\tilde{\Psi}_{2T^{(ij)}} &= \Psi_{2T^{(ij)}} - \tfrac{1}{D-2}\delta_{ij}\Psi_{2S}\,, \\
\tilde{\Psi}_{2^{ijkl}} &= \Psi_{2^{ijkl}} - \tfrac{2}{D-4}\big(\delta_{ik}\tilde{\Psi}_{2T^{(jl)}} + \delta_{jl}\tilde{\Psi}_{2T^{(ik)}} - \delta_{il}\tilde{\Psi}_{2T^{(jk)}} - \delta_{jk}\tilde{\Psi}_{2T^{(il)}}\big) \\
&\quad - \tfrac{4\,\delta_{i[k}\delta_{l]j}}{(D-2)(D-3)}\Psi_{2S}\,, \\
\tilde{\Psi}_{3^{ijk}} &= \Psi_{3^{ijk}} - \tfrac{2}{D-3}\delta_{i[j}\Psi_{3T^{k]}}\,.
\end{aligned} \tag{36}$$

The above quantities evaluated for the Kundt geometry (9) without employing any particular theory constraints imply:

- with respect to the natural null frame (34) the +2 and +1 boost-weight components Ψ_0 and Ψ_1 vanish identically and the analyzed spacetimes are *at least* of algebraic type II with $\boldsymbol{k} = \partial_r$ being a *double* degenerate Weyl aligned null direction (WAND),
- since $\Psi_{2^{ij}} = 0$, the geometry (9) is of the algebraic *subtype* II(d).

Moreover, the specific EGB theory restrictions derived in the previous section enter the discussion and may affect the Weyl tensor structure. We observe:

- in the generic case $Q \neq 0$ the type II(d) remains in general unchanged,
- the case $Q = 0$ with $Q_{pq} \neq 0$ specializes to II(ad) $\Leftrightarrow \Psi_{2S} = 0$, i.e.,

$$b = -\frac{{}^{S}R}{(D-2)(D-3)} = -\frac{4\Lambda_0}{(D-2)(D-3)} = \frac{1}{2k(D-3)(D-4)}\,, \tag{37}$$

- in the class $Q_{pq} = 0$ both the highest and the lowest non-trivial boost weights Ψ_2 and Ψ_4 contain *arbitrary* metric function g_{uu} and are thus in general non-vanishing. However, since the transverse space has to be the Einstein space we get $\tilde{\Psi}_{2T^{(ij)}} = 0$ which leads to the subtype II(bd).

Analogous discussion can be performed also for the Ricci tensor, see Ref. 4.

3.2. *Geodesic deviation*

To observe a physical nature of a given solution its useful to study relative motion of free test particles encoded in the geodesic deviation.[17] This is described the Jacobi equation

$$\frac{\mathrm{D}^2 Z^a}{\mathrm{d}\,\tau^2} = R^a{}_{bcd}\, u^b u^c Z^d\,, \tag{38}$$

where Z^a are components of the separation vector of nearby geodesics, $R^a{}_{bcd}$ is the Riemann curvature tensor, and u^b are components of the fiducial observer 4-velocity, i.e., $\boldsymbol{u} = \dot{r}\,\partial_r + \dot{u}\,\partial_u + \dot{x}^p\partial_p$. To obtain coordinate-independent information it is natural to project (38) onto a suitable orthonormal frame $\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(i)}\}$ associated with some particular observer where we assume $\boldsymbol{e}_{(0)} \equiv \boldsymbol{u}$. This leads to $\ddot{Z}^{(a)} = R^{(a)}{}_{(0)(0)(b)}\, Z^{(b)}$ with $\ddot{Z}^{(a)} \equiv e^{(a)}_b \frac{\mathrm{D}^2 Z^b}{\mathrm{d}\,\tau^2}$ and $Z^{(b)} \equiv e^{(b)}_a Z^a$, where $a, b = 0, 1, \ldots, D-1$, which gives $\ddot{Z}^{(0)} = 0$. Moreover, using decomposition of the Riemann tensor we obtain the invariant form of the geodesic deviation equation,

$$\ddot{Z}^{(\mathrm{i})} = \left[C_{(\mathrm{i})(0)(0)(\mathrm{j})} + \frac{1}{D-2}\left(R_{(\mathrm{i})(\mathrm{j})} - \delta_{\mathrm{ij}}\, R_{(0)(0)}\right) - \frac{R\,\delta_{\mathrm{ij}}}{(D-1)(D-2)}\right] Z^{(\mathrm{j})}\,, \tag{39}$$

where $\mathrm{i}, \mathrm{j} = 1, 2, \ldots, D-1$.

To explicitly analyze the specific effects of the EGB theory we restrict ourselves to the transversally static observers, i.e., we set $\sqrt{2}\,\dot{u} = 1$, $\dot{x}^p = 0$, which corresponds to the direct relation between observer's orthonormal frame $\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(i)}\}$ and the null frame (34), namely

$$\boldsymbol{k} = \tfrac{1}{\sqrt{2}}(\boldsymbol{u} + \boldsymbol{e}_{(1)})\,, \qquad \boldsymbol{l} = \tfrac{1}{\sqrt{2}}(\boldsymbol{u} - \boldsymbol{e}_{(1)})\,, \qquad \boldsymbol{m}_i = \boldsymbol{e}_{(i)}\,. \tag{40}$$

The straightforward evaluation of all projections in (39) and expression of the Ricci tensor components using the EGB field equations give the decomposed form of the geodesic deviation equation,

$$\begin{aligned}
\ddot{Z}^{(1)} =& \frac{2\,(\Lambda_0 + 2kH/D)}{(D-1)(D-2)} Z^{(1)} + \Psi_{2S}\, Z^{(1)} - \frac{1}{\sqrt{2}}\,\Psi_{3T^j}\, Z^{(j)} \\
&+ \frac{k}{D-2}\left(4\mathcal{H}_{ru}\, Z^{(1)} + \sqrt{2} m^p_j\, \mathcal{H}_{up}\, Z^{(j)}\right), \qquad (41) \\
\ddot{Z}^{(i)} =& \frac{2\,(\Lambda_0 + 2kH/D)}{(D-1)(D-2)} Z^{(i)} - \Psi_{2T^{(ij)}}\, Z^{(j)} - \frac{1}{\sqrt{2}}\,\Psi_{3T^i}\, Z^{(1)} - \frac{1}{2}\Psi_{4^{ij}}\, Z^{(j)} \\
&+ \frac{k}{D-2}\Big(- 2m^p_i m^q_j\, \mathcal{H}_{pq}\, Z^{(j)} + \sqrt{2} m^p_i\, \mathcal{H}_{up}\, Z^{(1)} \\
&\qquad + \left[(g_{uu} + 2)\mathcal{H}_{ru} + \mathcal{H}_{uu}\right] Z^{(i)}\Big), \qquad (42)
\end{aligned}$$

where the components of the traceless Gauss–Bonnet part $\mathcal{H}_{ab} \equiv H_{ab} - \frac{1}{D} H g_{ab}$ satisfy $2\mathcal{H}_{ru} = g^{pq}\,\mathcal{H}_{pq}$.

4. Example: Solutions with a constant-curvature transverse space

To explicitly demonstrate application of the above completely general constraints on the Kundt geometries within the EGB theory together with analysis of their properties we present a particular important case of (9) with an additional assumption that $(D-2)$-dimensional transverse space with the metric $g_{pq}(u,x)$ is of a *constant curvature*. In such a case the Riemann tensor can be written as

$$ {}^{S}R_{pqmn} = \frac{{}^{S}R}{(D-3)(D-2)} \left(g_{pm}g_{qn} - g_{pn}g_{qm}\right), \tag{43} $$

and the transverse metric becomes conformally flat, namely

$$ g_{pq} = \frac{\delta_{pq}}{P^2}, \quad \text{where} \quad P = 1 + \frac{{}^{S}R}{4(D-3)(D-2)} \left[\left(x^2\right)^2 + \cdots + \left(x^{D-1}\right)^2\right]. \tag{44} $$

The scalar curvature ${}^{S}R$ is then determined via the ru-equation (12) by generic theory parameters $k = \kappa\gamma$ and Λ_0,

$$ {}^{S}R = \frac{(D-2)(D-3)}{2k(D-4)(D-5)} \left(\pm\sqrt{1 + 8k\Lambda_0 \frac{(D-4)(D-5)}{(D-2)(D-3)}} - 1\right). \tag{45} $$

Immediately we can see that $D = 5$ and $D = 4$ are exceptional cases for which we have ${}^{S}R = 2\Lambda_0$ corresponding to the pure Einstein theory. Moreover, there are *two* branches of generic solutions in the EGB gravity where the " $+$ " case has a GR limit while the " $-$ " case is peculiar.

For generic values of the theory parameters we have $Q_{pq} \neq 0$ and the resulting line element becomes

$$ \begin{aligned} \mathrm{d}s^2 = {} & \left[1 + \frac{{}^{S}R\,\delta_{mn}\,x^m x^n}{4(D-2)(D-3)}\right]^{-2} \delta_{pq}\,\mathrm{d}x^p\,\mathrm{d}x^q - 2\,\mathrm{d}u\,\mathrm{d}r \\ & + \left[\frac{4\Lambda_0 - {}^{S}R}{(D-2) + 2k\,(D-4)\,{}^{S}R}\,r^2 + c(u)\,r + d(u,x)\right] \mathrm{d}u^2, \end{aligned} \tag{46} $$

where the scalar curvature ${}^{S}R$ is given by (45), $c(u)$ is arbitrary function, and $d(u,x)$ has to satisfy the Laplace equation,

$$ g^{pq} d_{||pq} = 0. \tag{47} $$

The Weyl tensor algebraic type of such solution is II(bcd). However, for the flat transverse space, i.e., with ${}^{S}R = 0$, $g_{pq} = \delta_{pq}$, we have $\Lambda_0 = 0$ and all the Gauss–Bonnet corrections vanish which means that we effectively deal with the Einstein theory (subclass of VSI spacetimes of the Weyl-type N). These spacetimes can be interpreted as exact type II gravitational waves propagating on the type D(bcd) background corresponding to the direct-product (anti-)Nariai universe.

This interpretation is simply supported by analysis of relative accelerations in terms of the geodesic deviation. For the transversally static observer associated with

the privileged null congruence we get a simple result,

$$\ddot{Z}^{(1)} = \frac{4\Lambda_0 - {}^{S}R}{(D-2) + 2k\,(D-4)\,{}^{S}R}\, Z^{(1)}\,, \qquad \ddot{Z}^{(i)} = -\frac{1}{2}\Psi_{4^{ij}}\, Z^{(j)}\,, \tag{48}$$

where the longitudinal spatial acceleration in the direction $\boldsymbol{e}_{(1)}$ is determined by the direct-product background corresponding to the non-trivial coefficient of r^2 term in g_{uu}, The transverse accelerations in $D-2$ directions $\boldsymbol{e}_{(i)}$ are given by $\Psi_{4^{ij}} = -\frac{1}{2} m_i^p m_j^q\, d_{||pq}$ and cause symmetric and traceless deformations representing the exact Kundt–EGB gravitational waves.

Finally, as a special case of the above solution we can set ${}^{S}R = 4\Lambda_0$ which makes the resulting metric r-independent and we are thus dealing with the *pp*-wave solution. Subsequently, the parameter Λ_0 can be vanishing, which leads to the classic solution analogous to the general relativity one, or it has to become

$$\Lambda_0 = -\frac{(D-2)(D-3)}{8k(D-4)(D-5)}\,. \tag{49}$$

This *new* non-trivial class is only allowed in the EGB theory with $D > 5$ and its metric takes the form

$$\mathrm{d}s^2 = \left(1 - \frac{\delta_{mn}\, x^m x^n}{8k(D-4)(D-5)}\right)^{-2} \delta_{pq}\, \mathrm{d}x^p\, \mathrm{d}x^q - 2\,\mathrm{d}u\,\mathrm{d}r + d(u,x)\,\mathrm{d}u^2\,, \tag{50}$$

where the Laplace equation (47) for the metric function $d(u,x)$ reflects the non-trivial transverse-space geometry corresponding to the Weyl-type II(bcd) solutions.

5. Conclusions

In this contribution we have summarized the main results of Ref. 4. In particular, we have described explicit class of the Kundt solutions to the Einstein–Gauss–Bonnet gravity. As the first step, the explicit form of the EGB field equations has been derived and geometrically distinct subclasses identified. We have also indicated differences in comparison with the classic solutions to the Einstein's general relativity. As a suitable tool characterizing the solutions we have briefly discussed the algebraic structure of the Weyl tensor and geodesic deviation. Finally, we have presented a particular example of the Kundt solutions with the constant-curvature transverse space including the EGB *pp*-waves.

Acknowledgments

This work was supported by the Czech Science Foundation Grant No. GAČR 20-05421S.

References

1. D. Lovelock, The Einstein tensor and its generalizations, *J. Math. Phys.* **12**, 498 (1971).
2. D. J. Gross and J. H. Sloan, The quartic effective action for the heterotic string, *Nucl. Phys. B* **291**, 41 (1987).
3. M. C. Bento and O. Bertolami, Maximally symmetric cosmological solutions of higher-curvature string effective theories with dilatons, *Phys. Lett. B* **368**, 198 (1996).
4. R. Švarc, J. Podolský, and O. Hruška, Kundt spacetimes in the Einstein-Gauss-Bonnet theory, *Phys. Rev. D* **102**, 084012 (2020).
5. W. Kundt, The plane-fronted gravitational waves, *Z. Physik* **163**, 77 (1961).
6. W. Kundt, Exact solutions of the field equations: Twist-free pure radiation fields, *Proc. Roy. Soc. A* **270**, 328 (1962).
7. J. Podolský and R. Švarc, Explicit algebraic classification of Kundt geometries in any dimension, *Class. Quantum Grav.* **30**, 125007 (2013).
8. H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge University Press, Cambridge, 2003).
9. J. B. Griffiths and J. Podolský, *Exact Space-Times in Einstein's General Relativity* (Cambridge University Press, Cambridge, 2009).
10. J. Podolský and M. Žofka, General Kundt spacetimes in higher dimensions, *Class. Quantum Grav.* **26**, 105008 (2009).
11. A. Coley, S. Hervik, G. Papadopoulos, and N. Pelavas, Kundt spacetimes, *Class. Quantum Grav.* **26**, 105016 (2009).
12. H. W. Brinkmann, Einstein spaces which are mapped conformally on each other, *Math. Annal.* **94**, 119 (1925).
13. A. Coley, R. Milson, V. Pravda, and A. Pravdová, Vanishing scalar invariant spacetimes in higher dimensions, *Class. Quantum Grav.* **21**, 5519 (2004).
14. W. B. Bonnor, Spinning null fluid in general relativity, *Int. J. Theor. Phys.* **3**, 257 (1970).
15. V. P. Frolov and D. V. Fursaev, Gravitational field of a spinning radiation beam pulse in higher dimensions, *Phys. Rev. D* **71**, 104034 (2005).
16. M. Ortaggio, V. Pravda, and A. Pravdová, Algebraic classification of higher dimensional spacetimes based on null alignment, *Class. Quantum Grav.* **30**, 013001 (2013).
17. J. Podolský and R. Švarc, Interpreting spacetimes of any dimension using geodesic deviation, *Phys. Rev. D* **85**, 044057 (2012).

Exact decoupled solutions in curvature-matter coupled gravity

M. Sharif* and Fizza Furqan
Department of Mathematics, University of the Punjab, Quaid-e-Azam Campus, Lahore-54590, Pakistan
** E-mail: msharif.math@pu.edu.pk*
www.pu.edu.pk

This paper explores the extended gravitational decoupling procedure for a static sphere in the context of $f(\mathbb{R},T)$ theory where $\mathbb{R}$ denotes the scalar curvature and T represents the trace of the energy-momentum tensor. This method extends the domain of a seed solution by including a new gravitational source. Deformations in radial and temporal metric potentials split the set of field equations into two subsystems associated with isotropic and additional matter sources. We utilize the Korkina-Orlyanskii spacetime as a solution for the system describing the seed source and use some physical constraints to extend it to anisotropic domain. A linear gravity model, $f(\mathbb{R},T)=\mathbb{R}+2\chi T$ (where χ couples geometry to matter) is employed to interpret the influence of the decoupling parameter on the developed solutions. It is found that physically acceptable solutions can be formulated in the background of $f(\mathbb{R},T)$ gravity through the decoupling approach.

Keywords: $f(\mathbb{R},T)$ gravity; Extended gravitational decoupling; Anisotropy.

1. Introduction

Numerous fascinating astronomical discoveries have opened the gateway to the mysteries of cosmic origin and evolution. In this regard, one of the significant developments is Hubble's discovery of an expanding universe.[1] Observations based upon redshifts of Type Ia supernovae provide evidence that the rate of this expansion is accelerating.[2] An unknown energy field (known as dark energy) possessing large negative pressure is assumed to be the source of this acceleration.[3] Einstein's general theory of relativity has successfully been employed to describe different astrophysical phenomena. However, this theory supports a universe experiencing a receding rate of expansion owing to gravitational attraction. Consequently, the cosmological constant was added in the field equations of general relativity (GR) to construct the ΛCDM model which has been used to justify the current cosmic behavior.[4] Nevertheless, this model is unable to solve the problems of fine-tuning[5] and cosmic coincidence.[6] The fine-tuning problem is concerned with the large discrepancy between the observed and predicted values of the cosmological constant whereas the cosmic coincidence problem questions the equal densities of dark matter (nonbaryonic matter with strong gravitational field) and dark energy detected in the current era.

Modifications of GR are proposed to find a suitable explanation of these cosmic issues. The simplest modification replaced the Ricci scalar in the Einstein-Hilbert

action by its generic function and was put forward by Buchdal[7] as the $f(\mathbb{R})$ theory. Harko et al.[8] introduced $f(\mathbb{R},T)$ gravity by generalizing the $f(\mathbb{R})$ theory. The $f(\mathbb{R},T)$ theory incorporates a non-zero covariant derivative of the trace of the energy-momentum tensor. This causes test particles to execute non-geodesic motion and produces acceleration. Various models of $f(\mathbb{R},T)$ gravity have been developed to explore different cosmological events. The most widely used model is $f(\mathbb{R},T) = f_1(\mathbb{R}) + f_2(T)$ where $f_2(T)$ is often set equal to $2\chi T$.[8] Houndjo[9] inspected the matter-dominated phase and late-time acceleration of the universe for this minimally coupled $f(\mathbb{R},T)$ model. This theory also aids in understanding of gravitational effects at the quantum level.[10] Das et al.[11] used different equations of state to examine the properties of gravastar in the context of $f(\mathbb{R},T)$ theory. Effects of the trace on the state variables of a white dwarf were discussed by Carvalho et al.[12] Moreover, the viability of relativistic interiors has been analyzed in this theory.[13]

Solutions of the Einstein field equations represent feasible stellar interiors. Therefore, they help in investigating the essential properties of compact objects. Models portraying physically acceptable astrophysical structures with extremely dense matter have been suggested to exhibit anisotropy. Ruderman[14] studied the existence of anisotropy in high density pure neutron matter. Canuto[15] reviewed the effects of deviations from local isotropy on various stellar characteristics. Impact of anisotropy generated during the phase transition from normal to pion-condensed matter was analyzed by Lovas et al.[16] Gleiser and Dev[17] revealed a stability range for anisotropic stars that was wider in contrast to their isotropic counterparts. Paul and Deb[18] constructed a class of viable relativistic models composed of anisotropic fluid in hydrostatic equilibrium.

However, extraction of analytic anisotropic solutions of the non-linear field equations is often difficult. Therefore, new methods to formulate viable solutions have been developed. In this respect, the decoupling approach allows us to extend known solutions of the field equations to more complex domains. This minimizes the requirement of additional constraints and simplifies the system of equations. Ovalle[19] introduced decoupling through the minimal geometric deformation (MGD) approach by addition of new matter sources. The MGD technique splits the original system of equations into two sets by deforming the radial metric coefficient. Both sets are solved individually and the complete solution is obtained by the principle of superposition.

The MGD technique has been employed extensively for the derivation of anisotropic extensions of well-known solutions. Ovalle et al.[20] utilized this technique to devise stellar models from the isotropic Tolman IV metric. Sharif and Sadiq[21] found well-behaved solutions for spherically symmetric matter configurations. Singh et al.[22] considered the MGD approach in relation to the embedding class-one condition. Furthermore, Sharif and collaborators[23] adopted this procedure to obtain realistic solutions in modified theories of gravity. The use of MGD

decoupling has yielded several solutions of the $f(\mathbb{R},T)$ field equations. Maurya and Tello-Ortiz[24] computed charged anisotropic extensions of the isotropic Durgapal-Fuloria model. Recently, Maurya et al.[25] utilized a linear $f(\mathbb{R},T)$ model to develop a compact structure through the Korkina-Orlyanskii line element and ascertained its viability as well as stability through energy conditions and speed of sound criteria, respectively.

Despite the effectiveness of the MGD method, the scheme deforms only one metric function leading to a purely gravitational interaction between matter sources. Thus, Ovalle[26] proposed a technique incorporating transformations of both radial and temporal metric components known as the extended geometric deformation (EGD) approach. The continuity equation obtained through this procedure holds regardless of the absence or presence of matter sources. Contreras and Bargueño[27] developed a charged BTZ solution by implementing this technique in (1+2)-dimensional spacetime. Maurya et al.[28] extended an embedding class-one solution through the EGD approach. Sharif and Ama-Tul-Mughani[29] constructed viable solutions from the isotropic Krori-Barua metric under the influence of the electromagnetic field. Moreover, the EGD technique has also been employed with modified theories to produce feasible stellar interiors.[30]

In this work, we provide two anisotropic extensions of the Korkina-Orlyanskii spacetime using the EGD approach in connection to $f(\mathbb{R},T)$ gravity. The salient features of the generated models are analyzed through energy conditions and stability criteria. This paper has been arranged as follows. In section **2**, we apply the EGD technique to decouple the field equations of $f(\mathbb{R},T)$ theory. An anisotropic solution corresponding to certain constraints on the extra source is evaluated in section **3**. Physical relevance of the solution is also discussed in the same section. The last section summarizes the main results.

2. Basic Formalism in $f(\mathbb{R},T)$ Theory

The action for $f(\mathbb{R},T)$ theory incorporating the Lagrangian density (L_ϑ) of an additional source (ϑ) turns out to be[8]

$$S=\int\sqrt{-g}(\frac{f(\mathbb{R},T)}{16\pi}+L_m+\alpha L_\vartheta)d^4x, \tag{1}$$

where L_m, g and α indicate the matter Lagrangian, determinant of the metric tensor and dimensionless decoupling parameter, respectively. The energy-momentum tensors of the seed and additional sources are written as

$$T_{\omega\beta}=-\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}L_m)}{\delta g^{\omega\beta}}=g_{\omega\beta}L_m-2\frac{\partial L_m}{\partial g^{\omega\beta}}, \tag{2}$$

$$\vartheta_{\omega\beta}=-\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}L_\vartheta)}{\delta g^{\omega\beta}}=g_{\omega\beta}L_\vartheta-2\frac{\partial L_\vartheta}{\partial g^{\omega\beta}}. \tag{3}$$

while the field equations for $f(\mathbb{R},T)$ gravity are derived as

$$8\pi(T_{\omega\beta}+\alpha\vartheta_{\omega\beta})-f_T(\mathbb{R},T)(T_{\omega\beta}+\Theta_{\omega\beta})=(R_{\omega\beta}-\nabla_\omega\nabla_\beta+g_{\omega\beta}\Box)f_{\mathbb{R}}(\mathbb{R},T) - \frac{1}{2}g_{\omega\beta}f(\mathbb{R},T). \tag{4}$$

Here ∇_ω symbolizes the covariant derivative operator while $f_{\mathbb{R}}(\mathbb{R},T)$ and $f_T(\mathbb{R},T)$ are derivatives of the generic function with respect to $\mathbb{R}$ and T, respectively. Moreover, $\Box=g^{\omega\beta}\nabla_\omega\nabla_\beta$ and $\Theta_{\omega\beta}=g_{\omega\beta}L_m-2T_{\omega\beta}-2g^{\delta\gamma}\frac{\partial^2 L_m}{\partial g^{\omega\beta}\partial g^{\delta\gamma}}$.

Covariant differentiation of Eq.(4) yields

$$\nabla^\omega T_{\omega\beta}=\frac{f_T(\mathbb{R},T)}{8\pi-f_T(\mathbb{R},T)}\left[(T_{\omega\beta}+\Theta_{\omega\beta})\nabla^\omega lnf_T(\mathbb{R},T)+\nabla^\omega\Theta_{\omega\beta}-\frac{1}{2}g_{\omega\beta}\nabla^\omega T - \frac{8\pi\alpha}{f_T(\mathbb{R},T)}\nabla^\omega\vartheta_{\omega\beta}\right], \tag{5}$$

which shows the non-conserved energy-momentum tensor in $f(\mathbb{R},T)$ theory. The isotropic matter distribution is characterized by the following energy-momentum tensor

$$T_{\omega\beta}=(\rho+p)u_\omega u_\beta-pg_{\omega\beta}, \tag{6}$$

with isotropic pressure (p), matter density (ρ), four-velocity (u_ω) and $u_\omega u^\omega=1$. Since the matter Lagrangian appears explicitly in the field equations, in order to obtain the corresponding dynamics we set $L_m=-p$ corresponding to a perfect fluid.[31] This further leads to $\Theta_{\omega\beta}=-2T_{\omega\beta}-pg_{\omega\beta}$. We consider a minimally coupled $f(\mathbb{R},T)$ model of the form

$$f(\mathbb{R},T)=\mathbb{R}+2\chi T, \tag{7}$$

which provides

$$\begin{aligned}G_{\omega\beta}&=8\pi(T_{\omega\beta}+\alpha\vartheta_{\omega\beta})+\chi Tg_{\omega\beta}+2\chi(T_{\omega\beta}+pg_{\omega\beta})\\&=8\pi\widetilde{T}_{\omega\beta}+8\pi\alpha\vartheta_{\omega\beta}=8\pi T^{(eff)}_{\omega\beta},\end{aligned} \tag{8}$$

where $\widetilde{T}_{\omega\beta}=T_{\omega\beta}(1+\frac{\chi}{4\pi})+\frac{\chi}{8\pi}(T+2p)g_{\omega\beta}$.

To describe the interior configuration of the stellar object, we choose a line element given by

$$ds^2=e^\nu dt^2-e^\eta dr^2-r^2(d\theta^2+\sin^2\theta d\phi^2), \tag{9}$$

where ν and η are functions of the radial coordinate. The field equations can now be rewritten as

$$\frac{1}{r^2}-e^{-\eta}\left(\frac{1}{r^2}-\frac{\eta'}{r}\right)=8\pi(\bar{\rho}+\alpha\vartheta^t_t), \tag{10}$$

$$-\frac{1}{r^2}+e^{-\eta}\left(\frac{1}{r^2}+\frac{\nu'}{r}\right)=8\pi(\bar{p}-\alpha\vartheta^r_r), \tag{11}$$

$$\frac{e^{-\eta}}{4}\left(2\nu''+\nu'^2-\eta'\nu'+2\frac{\nu'-\eta'}{r}\right)=8\pi(\bar{p}-\alpha\vartheta^\theta_\theta). \tag{12}$$

Here prime denotes derivative with respect to r while $\bar{\rho}$ and $\bar{p}$ are identified as follows

$$\bar{\rho} = \rho + \frac{\chi}{8\pi}(3\rho - p), \tag{13}$$

$$\bar{p} = p - \frac{\chi}{8\pi}(\rho - 3p). \tag{14}$$

The physical variables (total density and radial/tangential pressure) are defined as

$$\rho^{(tot)}(r) = \rho(r) + \alpha\vartheta_t^t, \tag{15}$$

$$p_r^{(tot)}(r) = p(r) - \alpha\vartheta_r^r, \tag{16}$$

$$p_t^{(tot)}(r) = p(r) - \alpha\vartheta_\theta^\theta, \tag{17}$$

whereas the anisotropy generated from involvement of an extra source is obtained as

$$\Delta = p_t^{(tot)}(r) - p_r^{(tot)}(r) = \alpha(\vartheta_r^r - \vartheta_\theta^\theta). \tag{18}$$

Seven unknowns appear in the system of equations (10)-(12). Therefore, to reduce the degrees of freedom and close the undetermined system, we adopt the EGD scheme.

2.1. *The EGD Approach*

The EGD approach disintegrates the field equations into two distinct sets through deformations in the temporal and radial metric potentials. To proceed with this technique, we consider a line element corresponding to perfect fluid

$$ds^2 = e^{\xi(r)}dt^2 - \frac{dr^2}{\mu(r)} - r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{19}$$

Taking non-zero values of the parameter α incorporates the effects of the source $\vartheta_{\omega\beta}$ in $T_{\omega\beta}$. The metric components linked to (ν, η)-frame are segregated as

$$\nu = \xi + \alpha h(r), \tag{20}$$

$$e^{-\eta(r)} = \mu + \alpha f(r), \tag{21}$$

where $h(r)$ and $f(r)$ represent temporal and radial deformation functions, respectively. Substitution of above transformations in the field equations decomposes the system into two sets. We acquire an array for the seed distribution given as

$$8\pi\bar{\rho} = -\frac{\mu'}{r} - \frac{\mu}{r^2} + \frac{1}{r^2} = 8\pi\rho + \chi(3\rho - p), \tag{22}$$

$$8\pi\bar{p} = \frac{\mu\xi'}{r} + \frac{\mu}{r^2} - \frac{1}{r^2} = 8\pi p - \chi(\rho - 3p), \tag{23}$$

$$8\pi\bar{p} = \frac{\mu}{4}(2\xi'' + (\xi')^2 + 2\frac{\xi'}{r}) + \frac{\mu'}{4}(\xi' + \frac{2}{r}) = 8\pi p - \chi(\rho - 3p). \tag{24}$$

The isotropic density and pressure are evaluated from the above equations as

$$\rho=\frac{\chi(\mu\xi' r+\mu-1)+(8\pi+3\chi)(-\mu' r-\mu+1)}{r^2(8\pi+2\chi)(8\pi+4\chi)}, \tag{25}$$

$$p=\frac{\chi(-\mu' r-\mu+1)+(8\pi+3\chi)(\mu\xi' r+\mu-1)}{r^2(8\pi+2\chi)(8\pi+4\chi)}. \tag{26}$$

The second system, related to the source $\vartheta_{\omega\beta}$, is obtained as

$$8\pi\vartheta^t_t=-\frac{f'}{r}-\frac{f}{r^2}, \tag{27}$$

$$8\pi\vartheta^r_r=-\frac{\mu h'}{r}-f(\frac{\xi'}{r}+\frac{\alpha h'}{r}+\frac{1}{r^2}), \tag{28}$$

$$\begin{aligned}8\pi\vartheta^\theta_\theta=&-\frac{f}{4}\left(2\xi''+2\alpha h''+(\xi')^2+\alpha^2(h')^2+2\xi'\alpha h'+2\frac{\xi'}{r}+2\frac{\alpha h'}{r}\right)\\&-\frac{f'}{4}(\frac{2}{r}+\xi'+\alpha h')-\mu\left(\frac{h''}{2}+\frac{\alpha(h')^2}{4}+\frac{\xi' h'}{2}+\frac{h'}{2r}\right)-\frac{\mu' h'}{4}.\end{aligned} \tag{29}$$

The energy-momentum tensor corresponding to the line element (19) satisfies the Bianchi identity as

$$\nabla^{(\xi,\mu)}_\omega\widetilde{T}^\omega_\beta=-\frac{dp}{dr}-\frac{\xi'}{2}(\rho+p)+\frac{\chi}{2(4\pi+\chi)}\frac{d}{dr}(\rho-p)=0, \tag{30}$$

while the divergence $\nabla_\omega\widetilde{T}^\omega_\beta$ associated with the system (22)-(24) yields

$$\nabla_\omega\widetilde{T}^\omega_\beta=\nabla^{(\xi,\mu)}_\omega\widetilde{T}^\omega_\beta+\alpha\frac{h'}{2}(\rho+p)\delta^1_\beta. \tag{31}$$

Furthermore, conservation of the additional source $\vartheta_{\omega\beta}$ is provided as

$$\nabla_\omega\vartheta^\omega_\beta=\frac{d\vartheta^r_r}{dr}-\frac{\xi'}{2}(\vartheta^t_t-\vartheta^r_r)-\frac{2}{r}(\vartheta^\theta_\theta-\vartheta^r_r)=-\frac{h'}{2}(\rho+p)\delta^1_\beta. \tag{32}$$

Consequently, the continuity equation for the effective energy-momentum tensor is constructed as follows

$$-\frac{dp}{dr}-\frac{\xi'}{2}(\rho+p)+\frac{\chi}{2(4\pi+\chi)}\frac{d}{dr}(\rho-p)+\alpha\left[\frac{d\vartheta^r_r}{dr}-\frac{\xi'}{2}(\vartheta^t_t-\vartheta^r_r)-\frac{2}{r}(\vartheta^\theta_\theta-\vartheta^r_r)\right]=0. \tag{33}$$

An exchange of energy between the two sources ($\widetilde{T}_{\omega\beta}$ and $\vartheta_{\omega\beta}$) is evident from Eqs.(31) and (32) which allows successful implementation of the decoupling process. Meanwhile, decoupling in the case of MGD is possible only without a transfer of energy. Note that the system (22)-(24) consists of four unknowns (μ, ξ, ρ, p) whereas the system (27)-(29) comprises seven unknowns (μ, ξ, ρ, p, f, h, ϑ^t_t, ϑ^r_r, ϑ^θ_θ). The extraction of solutions can be simplified substantially by specifying a seed solution to the system of equations connected to the original energy-momentum tensor. As a result, the unknowns in the second array of equations decrease from seven to five. Hence, the EGD scheme facilitates the evaluation of anisotropic solutions by reducing the degrees of freedom.

3. Interior Anisotropic Solution

In this section, we use the Korkina-Orlyanskii solution for the system of equations (22)-(24). The metric coefficients of the Korkina-Orlyanskii spacetime have been previously utilized for the development of spherically symmetric static solutions to the Einstein-Maxwell system.[32] Moreover, this ansatz has helped examine the role of pressure anisotropy on charged stellar objects.[33] The metric components and corresponding physical variables in the framework of $f(\mathbb{R},T)$ theory read

$$\mu(r)=\frac{1+Ar^2}{1+2Ar^2}, \tag{34}$$

$$e^{\xi(r)}=C(1+Ar^2)\left[D+\frac{\sqrt{1+2Ar^2}}{1+Ar^2}-\sqrt{2}\ln\left(\sqrt{1+2Ar^2}+\sqrt{2+2Ar^2}\right)\right]^2, \tag{35}$$

$$\rho(r)=\frac{A}{8\chi^2+48\pi\chi+64\pi^2}\left[\frac{\chi}{2Ar^2+1}+\frac{(3\chi+8\pi)\left(2Ar^2+3\right)}{\left(2Ar^2+1\right)^2}-\frac{4\chi\left(A^2r^4+Ar^2\left(\sqrt{Ar^2+1}+2\right)+1\right)}{\left(2A^2r^4+3Ar^2+1\right)^{3/2}\left(D+\psi(r)\right)}\right], \tag{36}$$

$$p(r)=\frac{A}{8\chi^2+48\pi\chi+64\pi^2}\left[\frac{\chi\left(2Ar^2+3\right)}{\left(2\mathrm{Ar}^2+1\right)^2}+\frac{3\chi+8\pi}{2Ar^2+1}-\frac{4(3\chi+8\pi)\left(A^2r^4+Ar^2\left(\sqrt{Ar^2+1}+2\right)+1\right)}{\left(2A^2r^4+3Ar^2+1\right)^{3/2}\left(D+\psi(r)\right)}\right], \tag{37}$$

with $\psi(r)=\frac{\sqrt{1+2Ar^2}}{1+Ar^2}-\sqrt{2}\ln(\sqrt{1+2Ar^2}+\sqrt{2+2Ar^2})$. Here A, C and D are constants that are calculated by matching with the exterior Schwarzschild line element expressed as

$$ds^2=\left(1-\frac{2M}{r}\right)dt^2-\frac{1}{(1-\frac{2M}{r})}dr^2-r^2(d\theta^2+\sin^2\theta d\phi^2), \tag{38}$$

where M is the total mass of the exterior geometry.

To ensure a smooth junction between the interior and exterior regions at $r=R$, the matching conditions were introduced by Darmois.[34] These conditions are given as

$$[ds^2]_{r=R}=0, \tag{39}$$

$$[p_r^{(tot)}(r)]_{r=R}=0. \tag{40}$$

which provide

$$A=\frac{2M}{R^2(R-4M)}, \tag{41}$$

$$C=\frac{1-\frac{2M}{R}}{(AR^2+1)}\left[D-\sqrt{2}\ln\left(\sqrt{2AR^2+1}+\sqrt{2AR^2+2}\right)\right.\quad\left.+\frac{\sqrt{2AR^2+1}}{AR^2+1}\right]^{-2}, \tag{42}$$

$$D=\frac{\sqrt{2AR^2+1}\,(D_2(R)\chi+4\pi D_1(R))+D_3(R)}{(AR^2+1)^{3/2}\,(4AR^2\chi+8\pi AR^2+3\chi+4\pi)}, \tag{43}$$

where

$$\begin{aligned}
D_1(R)&=4+4A^2R^4-\sqrt{1+AR^2}+2AR^2(4+\sqrt{1+AR^2}),\\
D_2(R)&=6+6A^2R^4-3\sqrt{1+AR^2}+2AR^2(6+\sqrt{1+AR^2}),\\
D_3(R)&=(3\chi+4\pi+4\chi AR^2+8\pi AR^2)(1+AR^2)^{3/2}\sqrt{2}\ln\left[\sqrt{1+2AR^2}\right.\\
&\quad\left.+\sqrt{2+2AR^2}\right].
\end{aligned}$$

We now produce a solution of Eqs.(27)-(29) corresponding to the extra source ϑ_ω. Because the number of unknowns (f, h, ϑ_t^t, ϑ_r^r, ϑ_θ^θ) is greater than the number of equations, two additional constraints need to be applied to close the system and extend the isotropic sector to anisotropic domain. The matching condition in Eq.(40) is fulfilled if

$$\vartheta_r^r(r)=p(r). \tag{44}$$

Thus, the above equation is selected as the first constraint. Accordingly, the physical entities of the anisotropic configuration are acquired through Eqs.(15)-(17) as

$$\rho^{(tot)}(r)=\rho(r)+\frac{\alpha}{8\pi}(-\frac{f'}{r}-\frac{f}{r^2}), \tag{45}$$

$$p_r^{(tot)}(r)=p(r)(1-\alpha), \tag{46}$$

$$\begin{aligned}
p_t^{(tot)}(r)&=p(r)-\frac{\alpha}{8\pi}\left[-\frac{f}{4}\left(2\xi''+2\alpha h''+(\xi')^2+\alpha^2(h')^2+2\xi'\alpha h'+2\frac{\xi'}{r}\right.\right.\\
&\quad\left.+\,2\frac{\alpha h'}{r}\right)-\frac{f'}{4}\left(\frac{2}{r}+\xi'+\alpha h'\right)-\mu\left(\frac{h''}{2}+\frac{\alpha(h')^2}{4}+\frac{\xi' h'}{2}+\frac{h'}{2r}\right)\\
&\quad\left.-\,\frac{\mu' h'}{4}\right].
\end{aligned} \tag{47}$$

In the subsequent discussion, the respective anisotropic extension is generated by means of a second constraint.

3.1. *Solution*

Along with the mimic constraint stated in Eq.(44), we impose a condition on the anisotropy of the system developed through the conservation equation by Bowers

and Liang.[35] The Bowers-Liang constraint takes the following precise form

$$\vartheta_\theta^\theta - \vartheta_r^r = jr\frac{\xi'}{2}(\vartheta_t^t + \vartheta_r^r). \tag{48}$$

Here j is an arbitrary constant for which we have defined a value of 2. This constraint has been used extensively with gravitational decoupling in search of feasible star models.[36,37] The geometric deformations $f(r)$ and $h(r)$ are obtained from Eqs.(44) and (48), respectively as

$$\begin{aligned} f(r) = {} & \frac{1}{(\chi+2\pi)(\chi+4\pi)\left(\alpha r h(r) + r\xi'(r) + 1\right)}\left(-r\chi^2 h(r)\mu(r)\right. \\ & + \pi r\chi\mu'(r) - 6\pi r\chi h(r)\mu(r) - 8\pi^2 r h(r)\mu(r) - 3\pi r\chi\mu(r)\xi'(r) \\ & \left. - 8\pi^2 r\mu(r)\xi'(r) - 2\pi\chi\mu(r) - 8\pi^2\mu(r) + 2\pi\chi + 8\pi^2\right), \end{aligned} \tag{49}$$

$$\begin{aligned} & \frac{1}{r}\left(r\left(h(r)\left(r\left(\alpha f'(r) + \mu'(r)\right) - 2\mu(r)\left((j-1)r\xi'(r) + 1\right)\right) + f'(r)\left((r\right.\right.\right. \\ & - 2jr)\xi'(r) + 2) + 2r\mu(r)h'(r) + \alpha r h(r)^2\mu(r)\big) + f(r)\left(2\alpha r^2 h'(r)\right. \\ & - 2\alpha r h(r)\left((j-1)r\xi'(r) + 1\right) + \alpha^2 r^2 h(r)^2 - 2jr^2\xi'(r)^2 - 4jr\xi'(r) \\ & \left. + 2r^2\xi''(r) + r^2\xi'(r)^2 - 2r\xi'(r) - 4\right)\big) = 0. \end{aligned} \tag{50}$$

A numerical solution of the deformation function $h(r)$ is revealed by applying the initial condition $h(0) = 0$ to Eq.(50).

The deformed temporal and radial metric components are derived as

$$\begin{aligned} \nu(r) = {} & \ln\left(C(1+Ar^2)\left[\frac{\sqrt{1+2Ar^2}}{1+Ar^2} - \sqrt{2}\ln\left(\sqrt{1+2Ar^2}\right.\right.\right. \\ & \left.\left.\left. + \sqrt{2+2Ar^2}\right) + D\right]^2\right) + \alpha h(r), \end{aligned} \tag{51}$$

$$\begin{aligned} e^{-\eta} = {} & \frac{1+Ar^2}{1+2Ar^2} + \alpha\left[\frac{1}{(\chi+2\pi)(\chi+4\pi)\left(\alpha r h(r) + r\xi'(r) + 1\right)}\left(\pi r\chi\mu'(r)\right.\right. \\ & - r\chi^2 h(r)\mu(r) - 6\pi r\chi h(r)\mu(r) - 8\pi^2 r h(r)\mu(r) - 3\pi r\chi\mu(r)\xi'(r) \\ & \left.\left. - 8\pi^2 r\mu(r)\xi'(r) - 2\pi\chi\mu(r) - 8\pi^2\mu(r) + 2\pi\chi + 8\pi^2\right)\right], \end{aligned} \tag{52}$$

while the matter variables are formulated through Eqs.(45)-(47). After the inclusion of anisotropy in the solution, matching of interior and exterior line elements yields

the following expressions

$$\ln\left(1-\frac{2M}{R}\right)=\ln\left(C(1+AR^2)\left[\frac{\sqrt{1+2AR^2}}{1+AR^2}-\sqrt{2}\ln\left(\sqrt{1+2AR^2}\right.\right.\right.$$
$$\left.\left.\left.+\sqrt{2+2AR^2}\right)+D\right]^2\right)+\alpha h(R), \tag{53}$$

$$1-\frac{2M}{R}=\frac{1+AR^2}{1+2AR^2}+\alpha\left[\frac{1}{(\chi+2\pi)(\chi+4\pi)\left(\alpha Rh(R)+R\xi'(R)+1\right)}\right.$$
$$\times\left(\pi R\chi\mu'(R)-R\chi^2h(R)\mu(R)-6\pi R\chi h(R)\mu(R)-8\pi^2\mu(R)\right.$$
$$-\left.8\pi^2Rh(R)\mu(R)-3\pi R\chi\mu(R)\xi'(R)-8\pi^2R\mu(R)\xi'(R)\right.$$
$$-\left.\left.2\pi\chi\mu(R)+2\pi\chi+8\pi^2\right)\right]. \tag{54}$$

which cannot determine the constants A and C. Hence, they are considered as free parameters and acquired from Eqs.(41) and (42). On the other hand, the condition $[p_r^{(tot)}(r)]_{r=R}=0$ provides the constant D as computed in Eq.(43).

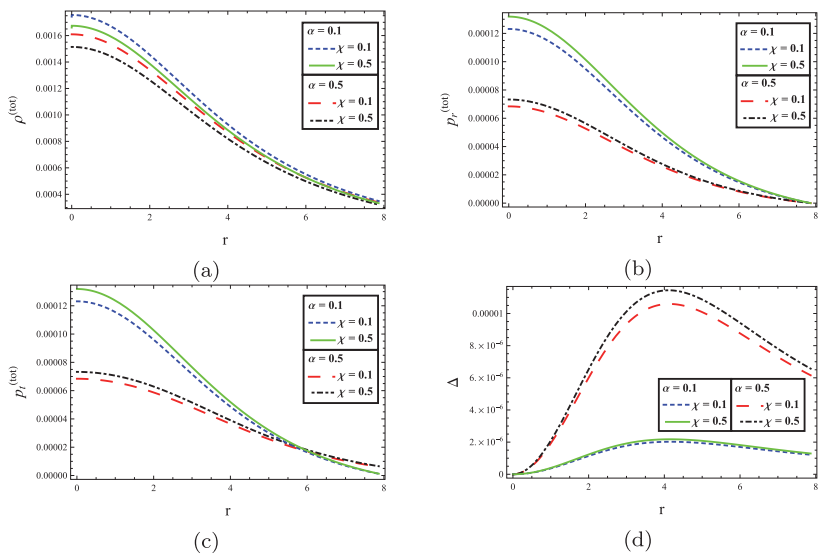


Fig. 1. Behavior of (a) density, (b) radial pressure, (c) tangential pressure and (d) anisotropy factor relative to radial coordinate.

The physical entities of the constructed setup are analyzed graphically for the star 4U 1538-52 with a mass of 1.28325km and a radius 7.866km.[38] The density and pressure components must be positive and decreasing from a maximum at the core. Figure 1 exhibits variations of these physical attributes relative to the radial coordinate. It is evident that the total energy density as well as both pressure components decrease continually from the center towards the surface. Moreover, the radial pressure vanishes at $r=R$ while the anisotropy displays a null value at

Table 1. State determinants of the stellar model corresponding to 4U 1538-52 ($M = 1.28325$ km and $R = 7.866$ km).

α	χ	$\rho_{\mathcal{C}}^{(tot)}$ (gm/cm^3)	$\rho_{\mathcal{S}}^{(tot)}$ (gm/cm^3)	$p_{\mathcal{C}}^{(tot)}$ $(dyne/cm^2)$	A	C	D
0.1	0.1	2.3466×10^{15}	4.6142×10^{14}	1.4790×10^{35}	0.0152	0.0179	5.3580
0.1	0.5	2.2355×10^{15}	4.3801×10^{14}	1.5824×10^{35}	0.0152	0.0189	5.2508
0.5	0.1	2.1526×10^{15}	4.5165×10^{14}	8.2173×10^{34}	0.0152	0.0179	5.3580
0.5	0.5	2.0268×10^{15}	4.2771×10^{14}	8.8221×10^{34}	0.0152	0.0189	5.2508

$r = 0$. The positive anisotropy factor suggests the presence of an outward-directed force. Higher values of χ cause the radial/tangential pressure and anisotropy to increase. Furthermore, the energy density declines as α and χ increase. Table 1 shows values of A, C and D along with the state determinants for different values of α and χ. The subscripts $\mathcal{C}$ and $\mathcal{S}$ denote that the quantity has been evaluated at the center and surface of the star, respectively.

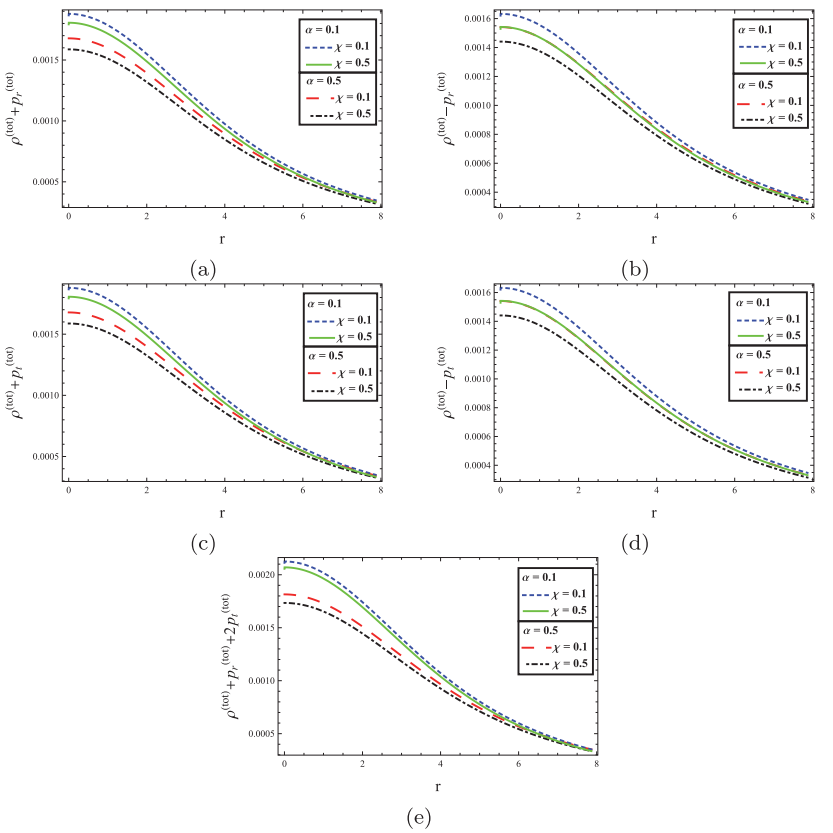


Fig. 2. Plots of energy conditions ((a)-(e)) against radial coordinate.

To verify the viability of the constructed solution, we place some bounds on the physical parameters called the energy conditions. These conditions, classified as null (NEC), weak (WEC), strong (SEC) and dominant (DEC), are presented as

$$\text{NEC: } \rho^{(tot)}+p_r^{(tot)}\geq 0, \quad \rho^{(tot)}+p_t^{(tot)}\geq 0,$$
$$\text{WEC: } \rho^{(tot)}\geq 0, \quad \rho^{(tot)}+p_r^{(tot)}\geq 0, \quad \rho^{(tot)}+p_t^{(tot)}\geq 0,$$
$$\text{SEC: } \rho^{(tot)}+p_r^{(tot)}\geq 0, \quad \rho^{(tot)}+p_t^{(tot)}\geq 0, \quad \rho^{(tot)}+p_r^{(tot)}+2p_t^{(tot)}\geq 0,$$
$$\text{DEC: } \rho^{(tot)}-p_r^{(tot)}\geq 0, \quad \rho^{(tot)}-p_t^{(tot)}\geq 0.$$

Figure 2 demonstrates positive behavior of above inequalities. Therefore, the solution is viable for the considered values of α and χ.

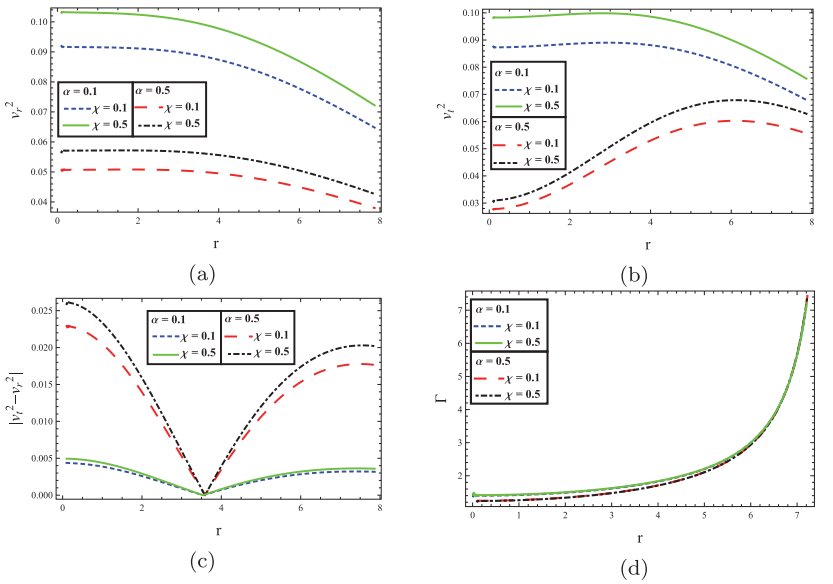


Fig. 3. Graphical representation of (a) v_r^2, (b) v_t^2, (c) $|v_t^2-v_r^2|$ and (d) Γ.

Another important feature of anisotropic stellar models is stability to fluctuations in the physical variables. The causality constraint is one of the stability criteria which limits the radial ($v_r^2=\frac{dp_r^{(tot)}}{d\rho^{(tot)}}$) and tangential ($v_t^2=\frac{dp_t^{(tot)}}{d\rho^{(tot)}}$) sound speeds inside the static configuration as follows

$$0\leq v_r^2\leq c, \quad 0\leq v_t^2\leq c,$$

where $c=1$ is the speed of light. Herrera[39] devised another approach to examine stability known as cracking. Cracking is caused by a sudden change in the direction of the radial force within a stellar configuration as a result of perturbations in density and anisotropy. Later, Abreu et al.[40] defined the potentially stable and

unstable regions within a compact object through differences of sound speeds given as

$$\begin{aligned} -1 \leq & v_t^2 - v_r^2 \leq 0 \Rightarrow \text{Potentially stable}, \\ 0 < & v_t^2 - v_r^2 \leq 1 \Rightarrow \text{Potentially unstable}. \end{aligned}$$

The above inequalities can be combined in to a single relation of the form $0 \leq |v_t^2 - v_r^2| \leq 1$.

In addition to the above constraints, an alternative tool to establish the stability of compact spheres is the adiabatic index which measures the resistance offered by a stellar configuration to perturbations in density. It is obtained as

$$\Gamma = \frac{p_r^{(tot)} + \rho^{(tot)}}{p_r^{(tot)}} v_r^2.$$

An adiabatic index greater than $4/3$ is proposed to indicate stable relativistic structures.[41] Figure 3 displays graphs of radial and temporal sound speeds as well as the adiabatic index for the developed solution. The sound speeds lie between 0 and 1 whereas the adiabatic index remains greater than $4/3$. Hence, we deduce that the resulting stellar model is stable.

4. Conclusions

In this paper, we have derived an anisotropic solution depicting the interior of a self-gravitating system in connection to $f(\mathbb{R}, T)$ gravity. The set of field equations (10)-(12) has been decoupled into two sets through the EGD scheme. These arrays separately describe the seed and anisotropic distributions. The Bianchi identity has been employed to examine energy transfer between the matter sources. We have considered the Korkina-Orlyanskii metric as a solution to the array corresponding to the isotropic configuration. Matching conditions on the hypersurface have been used to calculate the constants appearing in the solution. In order to extend the seed solution to anisotropic domain, we require two additional constraints. Thus, we have implemented the mimic constraint on ϑ_r^r along with the Bowers-Liang constraint. Physical relevance of the acquired solution has been investigated by employing the mass and radius of the star 4U 1538-53. For this purpose, we have performed graphical examination of the state determinants corresponding to $\alpha = 0.1,\ 0.5$ with $\chi = 0.1,\ 0.5$.

Plots for effective energy density, radial and tangential pressures have been found to monotonically decrease from the core. The anisotropy increases steadily from the center where $\Delta = 0$. The solution is viable and stable since all required quantities are in agreement with the energy bounds and speed of sound criteria. Furthermore, smaller values of the parameters α and χ give rise to denser, more compact and stable stellar models. Maurya et al.[25] evaluated an anisotropic, stable interior solution corresponding to a mimic constraint on pressure through the MGD technique. When compared to the solution obtained via the MGD approach, the EGD scheme

has yielded a denser star model with higher anisotropy. Hence, we conclude that the construction of anisotropic and ultra-dense stellar interiors is attainable using the EGD approach in the framework of $f(\mathbb{R}, T)$ theory.

References

1. E. Hubble, A relation between distance and radial velocity among extra-galactic nebulae, *Proc. Nat. Acad. Sci.* **15**, 168 (1929).
2. A. G. Riess et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astrophys. J.* **116**, 1009 (1998).
3. E. J. Copeland, M. Sami and S. Tsujikawa, Dynamics of dark energy, *Int. J. Mod. Phys. D* **15**, 1753 (2006).
4. D. Parkinson et al., The WiggleZ dark energy survey: Final data release and cosmological results, *Phys. Rev. D* **86**, 103518 (2012).
5. J. Martin, Everything you always wanted to know about the cosmological constant problem, *Comptes Rendus Physique* **13**, 566 (2012).
6. N. Sivanandam, Is the cosmological coincidence a problem?, *Phys. Rev. D* **87**, 083514 (2013).
7. H. A. Buchdahl, Non-linear Lagrangians and cosmological theory, *Mon. Not. R. Astron. Soc.* **150**, 1 (1970).
8. T. Harko et al., $f(\mathbb{R}, T)$ gravity, *Phys. Rev. D* **84**, 024020 (2011).
9. M. J. S. Houndjo, Reconstruction of $f(\mathbb{R}, T)$ gravity describing matter dominated and accelerated phases, *Int. J. Mod. Phys. D* **21**, 1250003 (2012).
10. M. X. Xu, T. Harko and S. D. Liang, Quantum Cosmology of $f(\mathbb{R}, T)$ gravity, *Eur. Phys. J. C* **76**, 449 (2016).
11. A. Das et al., Gravastars in $f(\mathbb{R}, T)$ gravity, *Phys. Rev. D* **95**, 124011 (2017).
12. G. A. Carvalho et al., Stellar equilibrium configurations of white dwarfs in the $f(\mathbb{R}, T)$ gravity, *Eur. Phys. J. C* **77**, 871 (2017).
13. D. Deb et al., Anisotropic strange stars under simplest minimal matter-geometry coupling in the $f(\mathbb{R}, T)$ gravity, *Phys. Rev. D* **97**, 084026 (2018); S. Hansraj and A. Banerjee, Dynamical behavior of the Tolman metrics in $f(\mathbb{R}, T)$ gravity, *Phys. Rev. D* **97**, 104020 (2018).
14. A. Ruderman, Pulsars: Structure and dynamics, *Annu. Rev. Astron. Astrophys.* **10**, 427 (1972).
15. V. Canuto, Equation of state at ultrahigh densities, *Annu. Rev. Astron. Astrophys.* **12**, 167 (1974).
16. I. Lovas, J. Németh and K. Sailor, Equilibrium between anisotropic normal and pion-condensed nuclear matter, *Nucl. Phy. A* **430**, 731 (1984).
17. M. Gleiser and K. Dev, Anistropic stars: Exact solutions and stability, *Int. J. Mod. Phys. D* **13**, 1389 (2004).
18. B. C. Paul and R. Deb, Relativistic solutions of anisotropic compact objects, *Astrophys. Space Sci.* **354**, 421 (2014).
19. J. Ovalle, Decoupling gravitational sources in general relativity: From perfect to anisotropic fluids, *Phys. Rev. D* **95**, 104019 (2017).
20. J. Ovalle et al., Anisotropic solutions by gravitational decoupling, *Eur. Phys. J. C* **78**, 122 (2018).
21. M. Sharif and S. Sadiq, Gravitational decoupled charged anisotropic spherical solutions, *Eur. Phys. J. C* **78**, 410 (2018).
22. K. N. Singh et al., Minimally deformed anisotropic model of class one space-time by gravitational decoupling, *Eur. Phys. J. C* **79**, 851 (2019).

23. M. Sharif and A. Waseem, Anisotropic spherical solutions by gravitational decoupling in $f(\mathbb{R})$ gravity, *Ann. Phys.* **405**, 14 (2019); M. Sharif and A. Majid, Decoupled anisotropic spheres in self-interacting Brans-Dicke gravity, *Chin. J. Phys.* **68**, 406 (2020).
24. S. K. Maurya and F. Tello-Ortiz, Charged anisotropic compact star in $f(\mathbb{R}, T)$ gravity: A minimal geometric deformation gravitational decoupling approach, *Phys. Dark Universe* **27**, 100442 (2020).
25. S. K. Maurya et al., Gravitational decoupling minimal geometric deformation model in modified gravity theory, *Phys. Dark Universe* **30**, 100640 (2020).
26. J. Ovalle, Decoupling gravitational sources in general relativity: The extended case, *Phys. Lett. B* **788**, 213 (2019).
27. E. Contreras and P. Bargueño, Extended gravitational decoupling in 2+1 dimensional space-times, *Class. Quantum Grav.* **36**, 215009 (2019).
28. S. K. Maurya, F. Tello-Ortiz and M. K. Jasim, An EGD model in the background of embedding class I space-time, *Eur. Phys. J. C* **80**, 918 (2020).
29. M. Sharif and Q. Ama-Tul-Mughani, Extended gravitational decoupled charged anisotropic solutions, *Chin. J. Phys.* **65**, 207 (2020).
30. M. Sharif and A. Majid, Extended gravitational decoupled solutions in self-interacting Brans-Dicke theory, *Phys. Dark Universe* **30**, 100610 (2020); M. Sharif and S. Saba, Extended gravitational decoupling approach in $f(G)$ gravity, *Int. J. Mod. Phys. D* **29**, 2050041 (2020).
31. J. D. Brown, Action functionals for relativistic perfect fluids, *Class. Quantum Grav.* **10**, 1579 (1993).
32. S. K. Maurya and Y. K. Gupta, Exact well behaved solutions of Einstein-Maxwell equations for relativistic charged superdense star models, *Astrophys. Space Sci.* **340**, 323 (2012).
33. S. K. Maurya, A. Banerjee and S. Hansraj, Role of pressure anisotropy on relativistic compact stars, *Phys. Rev. D* **97**, 044022 (2018).
34. G. Darmois, *Memorial Des Sciences Mathematiques* (Gauthier-Villars, Paris, 1927).
35. R. Bowers and E. Liang, Anisotropic spheres in general relativity, *Astrophys. J.* **188**, 657 (1974).
36. M. Sharif and A. Majid, Effects of charge on decoupled solutions in self-interacting Brans-Dicke theory, *Phys. Dark Universe* **32**, 100803 (2021).
37. G. Abellán et al., Regularity condition on the anisotropy induced by gravitational decoupling in the framework of MGD, *Eur. Phys. J. C* **80**, 177 (2020); E. Contreras, F. Tello-Ortiz and S. K. Maurya, Regular decoupling sector and exterior solutions in the context of MGD, *Class. Quantum Grav.* **37**, 155002 (2020).
38. M. L. Rawls et al., Refined neutron star mass determinations for six eclipsing x-ray pulsar binaries, *Astrophys. J.* **730**, 25 (2011).
39. L. Herrera, Cracking of self-gravitating compact objects, *Phys. Lett. A* **165**, 206 (1992).
40. H. Abreu, H. Hernandez and L. A. Nunez, Sound Speeds, cracking and stability of self-gravitating anisotropic compact objects, *Class. Quantum Gravit.* **24**, 4631 (2007).
41. H. Heintzmann and W. Hillebrandt, Neutron stars with an anisotropic equation of state-Mass, redshift and stability, *Astron. Astrophys.* **38**, 51 (1975).
42. S. K. Maurya, F. Tello-Ortiz and S. Ray, Decoupling gravitational sources in $f(\mathbb{R}, T)$ gravity under class I spacetime, *Phys. Dark Universe* **31**, 100753 (2021).

Tolman-Oppenheimer-Volkov conditions beyond spherical symmetry

José P. Mimoso

Departamento de Física and Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Ed. C8 1749-016 Lisboa, Portugal
E-mail: jpmimoso@fc.ul.pt

Alan Maciel

Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Avenida dos Estados 5001, CEP 09210-580, Santo André, São Paulo, Brazil
E-mail: alan.silva@ufabc.edu.br

Morgan Le Delliou

Institute of Theoretical Physics, School of Physical Science and Technology, Lanzhou University, No. 222, South Tianshui Road, Lanzhou, Gansu 730000, P R China
E-mail: delliou@lzu.edu.cn, delliou@ift.unesp.br

The TOV equation is usually interpreted as the relativistic counterpart of the classical condition for hydrostatic equilibrium, and characterises the static equilibrium of bound, spherical distributions of matter such as stars. In the present work we aim at showing that a generalised TOV equation also determines the equilibrium of models endowed with other symmetries besides spherical. We resort to the dual null formalism applied to spacetimes with two dimensional spherical, planar and hyperbolic symmetries, and consider a perfect fluid as the source. Static configurations assume the existence of a time-like Killing vector field orthogonal to the surfaces of symmetry, and homogeneous dynamical solutions arise when the Killing field is space-like. In order to treat equally all the aforementioned cases, we discuss the definition of a quasi-local energy for the spacetimes with planar and hyperbolic foliations, since the Hawking-Hayward definition only applies to compact foliations. This procedure enables us to translate our geometrical formalism to the fluid dynamics language in a unified way, to find the generalised TOV equation, for the three cases when the solution is static, and to obtain the evolution equation, for the homogeneous spacetime cases. Remarkably, we show that the static solutions which are not spherically symmetric violate the weak energy condition (WEC). We also show that the counterpart of the TOV equation $\rho + P = 0$, defines a cosmological constant-type behaviour, both in the hyperbolic and spherical cases. This implies a violation of the strong energy condition in both cases, added to the above mentioned violation of the weak energy condition in the hyperbolic case. We illustrate our unified treatment obtaining analogs of Schwarzschild interior solution, for an incompressible fluid $\rho = \rho_0$ constant.

Keywords: Gravitational collapse; hydrostatic equilibrium.

1. Introduction

The Tolman-Oppenheimer-Volkoff (TOV) equation[1,2] characterises spherical objects in equilibrium in the framework of General Relativity. It appears as the

relativistic counterpart of the classical condition for hydrostatic equilibrium. Given the relevance of the spherical configurations in astrophysics, many solutions were investigated,[3–15] and more specifically, different formalisms[16–20] and solutions generating techniques have been developed.[21,22] Extensions of the TOV equation have also been investigated in the framework of modified gravity theories.[23–29]

Yet a unified characterization of the underlying features of the TOV equation has attracted little attention, and this is what concerns us in the present work. There is a widespread perception of the TOV equation as being restricted to the spherically symmetric case, but this is a misled assumption. So, in the present work we derive a generalised TOV equation that also characterises the equilibrium of models endowed with other symmetries besides the spherical.[a]

We adopt the dual null formalism[b] that offers a description of the spacetime based on the properties of the optical flow, and has the significant advantage of being a coordinate free formalism. In the literature it has been considered in connection with the behaviour of dynamical black holes,[35,36] the definition of energy in more general geometries,[37] the gravitational collapse of fluids,[38] and even with the definition of generalized horizons in modified gravity.[39] The dual null formalism has also been useful to explicit the "linear" behavior of gravity for sources that satisfy the hypotheses of the Birkhoff theorem.[31]

Here, we apply the dual null formalism to analyse, in a unified way, the spacetimes which admit a codimension-two foliation with constant curvature leaves.[40] This comprises the spherical, planar and hyperbolic symmetries, sourced by a perfect fluid. Aside from the Killing vector fields that are tangent to those surfaces of symmetry, we assume the existence of an additional symmetry generated by a Killing vector field orthogonal to those surfaces at each event. A particular case of this setup, where the symmetry is spherical and the Killing vector is timelike, corresponds to the spherically symmetric perfect fluid in hydrostatic equilibrium, which leads us to the well-known TOV equation. As we will show this celebrated equation arises most naturally in the dual-null framework which, moreover, allows us to generalize it for the planar and hyperbolic cases. This generalisation of the geometry underlying the TOV equation, stepping beyond spherical symmetry, is novel and leads to consequences which are far from trivial.[40]

Since in planar and hyperbolic geometries the spatial hypersurfaces are open, this extension requires the novel introduction of a mass-energy "parameter". In fact one needs to promote a generalization of the Misner-Sharp/Hawking-Hayward definitions of the mass-energy distribution, which overcomes the problem of the divergence of the latter quantities due to the natural threading with infinite surfaces.

[a] A similar prejudice applying to the Birkhoff theorem, has been addressed with its generalization to more general geometries by C. Bona,[30] and more recently by us in.[31]

[b] We follow here the nomenclature coined by Sean A. Hayward[32–34] although in some references this formalism is referred to as double-null.

Finally, our investigation of the planar and hyperbolic geometries reveals the violation of the weak energy condition in order to maintain hydrostatic equilibrium. This takes the form of negative effective mass, physically translating repulsive curvature effects, which suggest a link to repulsive source models, as those proposed to mimic dark energy, generate bouncing universes, or support classical wormholes.[41,42]

For completness, when the metric is characterised by a spacelike, rather than a timelike Killing vector, we have spatially homogeneous spacetimes that correspond to some of the Bianchi spacetimes, or Kantowski-Sachs, as expected. The hydrostatic equilibrium on those spacetimes is only possible when their source is a cosmological constant in the non flat cases, also implying the violation of energy conditions.

2. The dual null formalism

We consider metrics that have a codimension-two maximally symmetric foliation, and can be written as

$$\mathrm{d}s^2 = N_{ab}\mathrm{d}x^a\mathrm{d}x^b + Y^2(x^c)\left(\mathrm{d}\theta^2 + S_\epsilon^2\mathrm{d}\phi^2\right) , \tag{1}$$

where

$$S_\epsilon = \begin{cases} \sin\theta, \text{ for } \epsilon = 1 \\ 1, \text{ for } \epsilon = 0 \\ \sinh\theta, \text{ for } \epsilon = -1 \end{cases} ,$$

and where we divide the tangent space $\mathcal{T}$ at each event in two orthogonal subspaces $\mathcal{T} = \mathcal{N} \oplus \mathcal{S}$. Here $\mathcal{S}$ is the subspace generated by the orbits of (θ, ϕ) and $\mathcal{N}$, the subspace of $\mathcal{T}$ orthogonal to $\mathcal{S}$. The x^a coordinates are chosen orthogonal to $\mathcal{S}$, which gives the metric in the warped sum form of Eq. (1).

We denote $s_{ab} = Y^2\gamma_{ab}$ the induced metric in each leaf of the foliation where $Y(x^c)$ is the warp factor. Evidently, $\gamma_{ab} := \delta^\theta_a\delta^\theta_b + S^2_\epsilon\delta^\phi_a\delta^\phi_b$ has constant curvature and does not depend on the coordinates x^a which identify each leaf Σ_{x^c}, defined as the locus spanned by the orbits of θ and ϕ for fixed x^c. We define an orthonormal two dimensional basis $(n^a\,, e^a)$ for $\mathcal{N}$, whose induced metric is N_{ab}, according to Eq. (1). This basis satisfies

$$-n^a n_a = e^a e_a = 1\,, \quad n^a e_a = n^a s_{ab} = e^a s_{ab} = 0\,. \tag{2}$$

We may also define a dual null basis for the same subspace from n^a and e^a by

$$\begin{aligned} k^a &= \frac{1}{2}\left(n^a + e^a\right) , \quad l^a = \frac{1}{2}\left(n^a - e^a\right) , \\ n^a &= k^a + l^a\,, \quad e^a = k^a - l^a\,, \end{aligned} \tag{3}$$

which satisfies

$$k^a k_a = l^a l_a = 0\,, \quad k^a l_a = -\frac{1}{2}\,. \tag{4}$$

The metric g_{ab} can be written as

$$g_{ab} = \frac{2}{k^c l_c} k_{(a} l_{b)} + s_{ab} \,. \tag{5}$$

We associate the null expansion for each null vector as follows

$$\Theta_k = \frac{1}{2} s^{ab} \mathcal{L}_k s_{ab} = \frac{1}{2} Y^{-2} \gamma^{ab} \mathcal{L}_k Y^2 \gamma_{ab} = \frac{2}{Y} k^a \partial_a Y \,. \tag{6}$$

We may extend the definition of null expansion to timelike and spacelike vectors in $\mathcal{N}$, calling it the two-expansion, since it measures the rate of variation of area, as in the null case. We may define the mean curvature form $\mathcal{K}_a = \partial_a \ln Y^2$, such that, we obtain for the two-expansion $\Theta_{(u)}$ of any vector u^a in $\mathcal{N}$

$$\Theta_{(u)} = u^a \mathcal{K}_a \,. \tag{7}$$

We describe our spacetimes by means of the behaviour of the null expansion, casting the Einstein equations, $G_{ab} = 8\pi T_{ab}$ in terms of expansions, i.e., by writing the Raychaudhuri equations and constraint equations[35,43–45]

$$\mathcal{L}_k \Theta_{(k)} = \nu_k \Theta_{(k)} - \frac{\Theta^2_{(k)}}{2} - 8\pi T_{ab} k^a k^b \,, \tag{8a}$$

$$\mathcal{L}_l \Theta_{(l)} = \nu_l \Theta_{(l)} - \frac{\Theta^2_{(l)}}{2} - 8\pi T_{ab} l^a l^b \,, \tag{8b}$$

$$\mathcal{L}_k \Theta_{(l)} + \mathcal{L}_l \Theta_{(k)} = -\Theta_{(l)} \nu_k - \Theta_{(k)} \nu_l - 2\,\Theta_{(k)}\,\Theta_{(l)} + \epsilon \frac{2\,k^a l_a}{Y^2} + 16\pi\, T_{ab} k^a l^b \,, \tag{8c}$$

where we included the inaffinities ν_k and ν_l, defined as

$$\nu_k = \frac{1}{k^c l_c} l^b k^a \nabla_a k_b \qquad \nu_l = \frac{1}{k^c l_c} k^b l^a \nabla_a l_b \,. \tag{9}$$

In the latter equations, for T_{ab}, if we take the source to be a perfect fluid[c], then the energy momentum tensor reduces to

$$T_{ab} = \rho n_a n_b + P(e_a e_b + s_{ab}) \,. \tag{10}$$

where we adapt our vector basis to the fluid source, such that n^a gives its flow. In the latter decomposition ρ is the energy density, and P is the isotropic pressure, both measured by an observer moving with 4-velocity n^a. By construction, the flow n^a is always orthogonal to the surfaces of symmetry and will be characterized by two quantities

$$\mathcal{A} = e^a \dot{n}_a = e^a n^b \nabla_b n_a \,, \qquad \mathcal{B} = e^a n'_a = e^a e^b \nabla_b n_a \,. \tag{11}$$

The scalar $\mathcal{A}$ gives us the acceleration of the flow, a positive sign meaning that the acceleration is outwards in the spherical, compact case. The scalar $\mathcal{B}$ gives the

[c]The equations in the case of a general fluid are given by us in Ref.[46].

change of direction of n^a as we travel along e^a. It is the $e-e$ component of the extrinsic curvature K_{ab} of the 3-space orthogonal to this flow, since

$$K_{ab} = \frac{1}{2}\mathcal{L}_n h_{ab}\,, \tag{12}$$

where $h_{ab} = g_{ab} + n_a n_b$. We may also write

$$h_{ab} = e_a e_b + Y^2 \gamma_{ab}\,, \tag{13}$$

which gives

$$K_{ab} = \mathcal{B}\, e_a e_b + \frac{\Theta_{(n)}}{2}\, Y^2 \gamma_{ab}\,. \tag{14}$$

The trace of Eq. (14) gives us the flow volumetric expansion $\Theta_3 = \nabla_a n^a = K^a_a$ as

$$\Theta_3 = \mathcal{B} + \Theta_{(n)}\,. \tag{15}$$

In order to relate our quantities with the flow scalars, we compute the shear scalar σ, by taking the symmetric traceless part of K_{ab}. We obtain

$$\sigma = \frac{\Theta_{(n)}}{6} - \frac{\mathcal{B}}{3}\,, \tag{16}$$

which implies

$$\frac{\Theta_3}{3} + \sigma = \frac{\Theta_{(n)}}{2}\,, \tag{17}$$

in agreement with the result obtained in Ref.[38].

Using the inaffinities of the null basis vectors, $\mathcal{A}$ and $\mathcal{B}$ can be expressed as

$$\mathcal{A} = \nu_k - \nu_l\,, \qquad \mathcal{B} = \nu_k + \nu_l\,. \tag{18}$$

Contracting the conservation of the energy-momentum tensor with e_b (Euler equation in[47]) we get

$$e_b \nabla_a T^{ab} = (\rho + P)\dot{n}^b e_b + e^a \nabla_a P = 0 \Rightarrow$$
$$\mathcal{A} = -\frac{e^a \partial_a P}{\rho + P}\,. \tag{19}$$

3. Orthogonal Killing vector

We now assume that our metric has a Killing vector orthogonal to maximally symmetric surfaces, as we are interested in static configurations. Our symmetry requirements imply that it commutes with the symmetry generators on the foliation. We denote this hypersurface orthogonal Killing vector field χ^a. It satisfies the Killing equation,

$$\mathcal{L}_\chi g_{ab} = 0\,. \tag{20}$$

Proposition 1. If a spacetime is described by a metric of the form (1) and admits an orthogonal Killing vector $\chi^a \in \mathcal{N}$, then $\Theta_\chi = 0$.

Proof. We may write, from Eq. (6), $\Theta_\chi = \frac{1}{2}s^{ab}\mathcal{L}_\chi s_{ab} = \frac{1}{2}Y^{-2}\gamma^{ab}\mathcal{L}_\chi Y^2\gamma_{ab}$, and

$$g_{ab} = N_{ab} + Y^2\gamma_{ab}\,. \tag{21}$$

Then

$$\begin{aligned} 0 = Y^{-2}\gamma^{ab}\mathcal{L}_\chi g_{ab} &= Y^{-2}\gamma^{ab}\mathcal{L}_\chi N_{ab} + 2\Theta_\chi \\ &= -Y^{-2}N^{ab}\mathcal{L}_\chi\gamma_{ab} + 2\Theta_\chi\,. \end{aligned} \tag{22}$$

However

$$\mathcal{L}_\chi\gamma_{ab} = 0\,, \tag{23}$$

since χ^a does not admit components in $\mathcal{S}$ and γ_{ab} doesn't depend on coordinates along $\mathcal{N}$. Therefore, Eq. (22) implies that $\Theta_\chi = 0$. □

Consequently, if there is an extra symmetry with orbits orthogonal to those of the maximally symmetric leaves of the foliation, the two-expansion of its generator vanishes. This also implies that if $\mathrm{d}Y$ is spacelike, then χ_a is timelike and vice-versa. If $\mathrm{d}Y$ is null, the Killing vector will also be null.

4. Extension of the mass-energy parameter to open geometries

There is a widely known mass-energy definition suitable to the spherically symmetric case, namely the Misner-Sharp mass-energy,[48,49] that is defined regardless of asymptotic assumptions. However, as we also intend to analyze nonspherical spacetimes in this work, we are led to consider a more general mass-energy definition, namely the Hawking-Hayward's (hereafter HH).[37,50] The HH mass-energy gives the mass-energy content inside a closed compact surface in terms of an integral over that surface, in a manner similar to the Gauss law in Newtonian gravity. This quasilocal mass-energy has been explored in different contexts, such as seen in Refs.[51–53].

Under our symmetry assumptions the Hawking-Hayward mass-energy enclosed by Σ is reduced to

$$M_\Sigma = \frac{1}{8\pi}\sqrt{\frac{A}{16\pi}}\int_\Sigma\left[\mathcal{R} - \frac{1}{k^a l_a}\Theta_{(k)}\Theta_{(l)}\right]\mathrm{d}\Sigma \tag{24}$$

where $\mathcal{R}$ is the two-dimensional Ricci scalar and A is the area of Σ. We have included a factor of $-\dfrac{1}{k^a l_a}$ in the optical scalars part of the mass-energy, compared with the formula present in Ref.[37], in order to take account of our different normalization of the null normals. Our symmetry assumptions imply that the only non-vanishing optical scalar on Σ is the null expansion.

Since we assume that Σ is maximally symmetric, we have $\mathcal{R} = \frac{2\epsilon}{Y^2}$. We also have

$$\Theta_{(k)}\Theta_{(l)} = k^{(a}l^{b)}\partial_a \ln Y^2 \partial_b \ln Y^2 = \frac{k^c l_c}{2} g^{ab}\partial_a \ln Y^2 \partial_b \ln Y^2 =$$

$$\frac{1}{2}k^c l_c ||\mathrm{d}\ln Y^2||^2 = k^c l_c \frac{2}{Y^2}||\mathrm{d}Y||^2 \Rightarrow \frac{\Theta_{(k)}\Theta_{(l)}}{k^c l_c} = \frac{2}{Y^2}||\mathrm{d}Y||^2\,, \tag{25}$$

where we used Eq. (5) in the second step.

For the spherical case $\epsilon = 1$ and $A = 4\pi Y^2$, we obtain the known interpretation of $||\mathrm{d}Y||$ in terms of the Misner-Sharp mass-energy, which coincides with the Hawking-Hayward one

$$M_\Sigma = \frac{Y}{2}\left(1 - ||\mathrm{d}Y||^2\right) \Leftrightarrow ||\mathrm{d}Y||^2 = 1 - \frac{2M_\Sigma}{Y}\,. \tag{26}$$

In this work, we aim at treating all three symmetry types in the same manner. In the planar and hyperbolic cases ($\epsilon = 0$ and $\epsilon = -1$, respectively), the Hawking-Hayward mass is not conveniently defined for the integration domain set by our preferred foliation, as it requires a closed compact surface which is absent in the latter cases. Therefore, we need a mass-energy definition which might be equivalent to the HH mass-energy, but suitable to deal with (instead of adapted for) non compact domains in order to take advantage of the planar or hyperbolic symmetry.

We can make such an extension of the HH mass-energy, as long as their boundary correspond to a pair of symmetric two-surfaces of symmetry Σ corresponding to the same warp factor Y. Of course, those domains are infinite and have an infinite mass-energy content in general. However, as they are homogeneous along the surfaces of symmetry, we can successfully adapt the HH mass-energy definition in order to obtain a finite *mass-energy parameter* with those cases. They then describe an infinite mass-energy distribution, homogeneous along the surfaces of symmetry, with a finite density.

We proceed by first making the replacement

$$\frac{1}{8\pi}\sqrt{\frac{A}{16\pi}} \to \frac{Y}{4\pi\kappa}, \tag{27}$$

in order to keep its dimensionality, and eliminating the explicit dependence on the area of Σ. Evidently, by setting $\kappa = 4$ we recover the Hawking-Hayward mass-energy in the spherical case. This step is justified by the fact that originally this factor was introduced to correct the dimensionality of the mass-energy, and to make it match the Arnowitt-Deser-Misner (ADM) mass,[54] where both are well defined. Since our symmetric spacetimes allow an "areal scalar" as the warp factor Y, we can replace $\sqrt{A}$ by Y as the quantity with dimension of length associated to each surface of symmetry.

We then define the quasi-local mass-energy parameter $\mu(Y)$ by

$$M_\Sigma = \frac{\mu(Y)}{4\pi} \int S_\epsilon(\theta)\, \mathrm{d}\theta \mathrm{d}\phi\,, \tag{28}$$

and we write

$$\frac{Y}{4\pi\kappa}\left[\mathcal{R} - \frac{\Theta_{(k)}\Theta_{(l)}}{k^c l_c}\right]\int_\Sigma \mathrm{d}\Sigma = \frac{Y}{4\pi\kappa}\left[2\epsilon - 2||\mathrm{d}Y||^2\right]\int S_\epsilon(\Theta)\mathrm{d}\theta\mathrm{d}\phi\,. \tag{29}$$

Equating Eqs. (28) and (29) and eliminating the improper area integral on both sides we derive

$$||\mathrm{d}Y||^2 = \epsilon - \frac{\kappa\mu(Y)}{2Y}\,. \tag{30}$$

Equation (30) coincides with the known mass function [see Eq. (15.7a) in[55]] which appears as we integrate the Einstein equations of specific spacetimes with metrics of the form (1) for planar and hyperbolic symmetries. From now on, we will consider Eq. (30) with the choice $\kappa = 4$ as the mass-energy definition.

An alternative route to Eq. (30) can be obtained by computing the HH mass-energy in a finite domain, symmetric with respect to the central plane or wire, $Y = 0$, and taking the limit where the domain tends to be the whole surface. The finite integration domain consist of the union of

(1) a subset of the Σ_Y, that we denote Γ_r, bounded by a circle γ_r of radius r on the (θ, ϕ) coordinate plane and
(2) a compact surface given by the surfaces Δ_r defined by γ_r transported along Y orbits.

It forms a closed surface, corresponding to a part of a cylinder bounded by $Y =$ constant surfaces in the space of coordinates (Y, θ, ϕ). Therefore, the HH mass-energy enclosed by those surfaces will by finite, and given by

$$M_r = \frac{1}{8\pi}\sqrt{\frac{A_r}{16\pi}}\left(\int_{\Gamma_r}(\dots)S_\epsilon \mathrm{d}\theta\mathrm{d}\phi + \int_{\Delta_r}(\dots)\mathrm{d}\Delta\right) \tag{31}$$

where (...) replaces the integrand of Eq. (24). In the limit $r \to \infty$, the first integral in Eq. (31) scales as r^2 while the second one scales as r. This means that, in the limit $r \to \infty$, and repeating the replacement in Eq. (27) we obtain

$$\frac{M_r}{A_r} \to \frac{\mu(Y)}{4\pi Y^2}\,. \tag{32}$$

This relation is particularly adequate to obtain the extension of the TOV equation for the open geometries, as it will become apparent in the following section.

5. Evolution equations

5.1. *Timelike Killing vector*

We assume $\chi^a\chi_a < 0$. In this case, the spacetime is static, and $n^a \sim \chi^a$. Therefore, from Proposition 1, $\Theta_{(n)} = 0$ everywhere, and $\mathrm{d}Y$ is spacelike, since it is orthogonal to n^a. If $\Theta_{(n)}$ vanishes everywhere, this means that the fluid has no radial velocity, therefore we are dealing with a static fluid with a flow parallel to the Killing vector field.

In order to characterize its static equilibrium, we realise the derivative of the flow 2-expansion along the flow itself is vanishing

$$\mathcal{L}_n\Theta_{(n)} = 0\,, \tag{33}$$

since $\Theta_{(n)} = 0$ everywhere.

We may write $\mathcal{L}_n\Theta_{(n)}$ in terms of the null expansions as

$$\mathcal{L}_n\Theta_{(n)} = \mathcal{L}_k\Theta_{(k)} + \mathcal{L}_l\Theta_{(l)} + \mathcal{L}_l\Theta_{(k)} + \mathcal{L}_k\Theta_{(l)}\,. \tag{34}$$

Substituting the Eqs. (8a), (8b), and (8c), we obtain

$$\begin{aligned}\mathcal{L}_n\Theta_{(n)} = &-\frac{(\Theta_{(k)} + \Theta_{(l)})^2}{2} - \\ &\Theta_{(k)}\Theta_{(l)} - \frac{\epsilon}{Y^2} - 8\pi T_{ab}e^a e^b + \mathcal{A}(\Theta_{(k)} - \Theta_{(l)})\,.\end{aligned} \tag{35}$$

Recall that we are assuming $\Theta_{(n)} = \Theta_{(k)} + \Theta_{(l)} = 0$, and that $\Theta_{(k)} - \Theta_{(l)} = \Theta_{(e)}$, using Eqs. (7) and (3). We identify here $\Theta_{(k)}\Theta_{(l)}$ as the mass term, since it equals $\frac{2}{Y^2}||\mathrm{d}Y||^2 = \frac{2}{Y^2}\left(\epsilon - \frac{2\mu(Y)}{Y}\right)$.

Taking the source to be a perfect fluid, (10), and contracting the divergence of the energy-momentum tensor with e_b (Euler equation in[47]) we get

$$\begin{aligned}e_b\nabla_a T^{ab} = (\rho + P)\dot{n}^b e_b + e^a\nabla_a P = 0 \Rightarrow \\ \mathcal{A} = -\frac{e^a\partial_a P}{\rho + P}\,.\end{aligned} \tag{36}$$

Since $\Theta_{(n)} = 0$, e^a is proportional to ∂_Y, and as e^a is normalized, we have

$$e_a = \frac{1}{||\mathrm{d}Y||}\partial_a Y\,, \quad e^a = ||\mathrm{d}Y||(\partial_Y)^a\,, \tag{37}$$

which gives us

$$\mathcal{A}\Theta_e = -||\mathrm{d}Y||^2\,\frac{2}{Y}\frac{\partial_Y P}{\rho + P}\,. \tag{38}$$

Therefore, replacing $||\mathrm{d}Y||^2$ by its meaning in terms of mass, $\mathcal{L}_n\Theta_{(n)} = 0$ corresponds to[40]

$$\left(\frac{\epsilon}{Y^2} - \frac{2\mu(Y)}{Y^2}\right)\frac{\partial_Y P}{\rho + P} = -\frac{1}{Y}\frac{\mu(Y)}{Y^2} - 4\pi P\,, \tag{39}$$

that reveals that the generalization of the concept of mass-energy given by (32) is particularly suited to avoid the caveats of the divergence of that quantity if it were defined in accordance to the canon of the spherically symmetric models.

Of course the latter equation (39) can alternatively be cast under a more familiar form as

$$\frac{\partial_Y P}{\rho + P} = -\left(\frac{\mu(Y)}{Y^2} + 4\pi P Y\right)\left(\epsilon - \frac{2\mu(Y)}{Y}\right)^{-1}, \tag{40}$$

which is what we denote as *the unified TOV equation*. It reduces to the well-know TOV equation for spherically symmetric spacetimes when $\epsilon = 1$, and it corresponds to the equation of hydrostatic equilibrium for planar and hyperbolic geometries, in the cases where $\epsilon = 0$ and $\epsilon = -1$, respectively. This underlines the fact that the TOV equation is a hydrostatic equilibrium equation, and not an equation of state, as it is sometimes erroneously stated.

In order to determine $\mu(Y)$ we consider the $\mathcal{L}_e\Theta_{(e)}$ Raychaudhuri equation

$$\begin{aligned}\mathcal{L}_e\Theta_{(e)} = \mathcal{B}\Theta_{(n)} - \frac{\Theta^2_{(e)}}{2} + \frac{1}{4}\left(\Theta^2_{(n)} - \Theta^2_{(e)}\right) \\ + \frac{\epsilon}{Y^2} - 8\pi T_{ab} n^a n^b\,,\end{aligned} \tag{41}$$

which, by using $\Theta_{(n)} = 0$, and Eq. (37) lead us to

$$||\mathrm{d}Y||\partial_Y\left(\frac{2}{Y}||\mathrm{d}Y||\right) = -\frac{3}{Y^2}||\mathrm{d}Y||^2 + \frac{\epsilon}{Y^2} - 8\pi\rho\,. \tag{42}$$

Substituting Eq. (30) into Eq. (42), we obtain

$$\partial_Y \mu = 4\pi\rho Y^2\,, \tag{43}$$

which looks like the mass-energy equation of spherical symmetry. Here, it should be interpreted as the mass-energy equation in the spherical case, and as a mass-energy parameter equation in the planar and hyperbolic cases. Furthermore, Eqs. (43) and (30) imply that, if the *weak energy condition (WEC)*[56] *holds*, only the spherically symmetric case admits static regular solutions. Indeed, as those solutions require $||\mathrm{d}Y||^2 > 0$, that implies $\mu < 0$ for $\epsilon \leq 0$ and, as in regular spacetimes,

$$\mu(Y) = 4\pi\int_0^Y \rho(y) y^2 \mathrm{d}y\,, \tag{44}$$

this imposes $\rho < 0$.

With Eq. (43), the last requirement to solve Eq. (40) is the equation of state of the fluid, $f(\rho, P) = 0$ which should come from specific physical considerations.

5.2. *Spacelike Killing vector: The cosmological cases*

In the case of a spacelike Killing vector, $\mathrm{d}Y$ is timelike, the flow n_a is orthogonal to the Killing vector, and the unitary base vector e^a is parallel to it. This imposes no constraint on the sign of μ according to Eq. (30), and thus there is no need to

violate energy conditions in order to consider these solutions, thorougly studied in cosmology.[57]

Combining the dynamical equation given by $\mathcal{L}_e\Theta_{(e)} = 0$, and Proposition 1, we have $\Theta_{(e)} = 0$. Replacing Eq. (16) in Eq. (41), we derive

$$\frac{3}{4}\Theta^2_{(n)} - 3\sigma\Theta_{(n)} + \frac{\epsilon}{Y^2} = 8\pi\rho\,, \tag{45}$$

and, using Eq. (17), we can express this Eq. (45) in terms of the volume expansion Θ_3 obtaining

$$\frac{\Theta_3^2}{3} - 3\sigma^2 = 8\pi\rho - \frac{\epsilon}{Y^2}\,, \tag{46}$$

which corresponds to the generalised Friedmann constraint equation for the evolution of a homogeneous and anisotropic universe. In the case $\sigma = 0$, we may identify $\Theta_3 = 3H$, and we recover the usual Friedmann equation for the flat ($\epsilon = 0$) and open ($\epsilon = -1$) spatially isotropic universes. Notice though that $\sigma = 0$ also yields anisotropic, cosmological solutions when the matter content is not a perfect fluid.[58]

The $\mathcal{L}_n\Theta_{(n)}$ Raychaudhuri equation gives the evolution of $\Theta_{(n)}$. Using Eq. (35) we obtain

$$\mathcal{L}_n\Theta_{(n)} = -\frac{3}{4}\Theta^2_{(n)} - \frac{\epsilon}{Y^2} - 8\pi T_{ab}e^a e^b\,, \tag{47}$$

so that subtracting Eq. (45) from Eq. (47), we further derive

$$\mathcal{L}_n\Theta_{(n)} = -3\sigma\Theta_{(n)} - 8\pi\left(\rho + P\right)\,, \tag{48}$$

which, together with an equation of state relating ρ and P closes our system. By adding half of the Eq. (48) with one third of Eq. (46), we obtain:

$$\begin{aligned}\mathcal{L}_n\left(\frac{\Theta_3}{3}\right) + \left(\frac{\Theta_3}{3}\right)^2 = \\ - 2\sigma\left(\frac{\Theta_3}{3} + \sigma\right) - \frac{\epsilon}{3Y^2} - \frac{4\pi}{3}\left(\rho + 3P\right)\,.\end{aligned} \tag{49}$$

Those homogeneous and anisotropic spacetimes belong to a subclass of Bianchi models,[59,60] with the case $\epsilon = 0$ corresponding to Bianchi type I universes, $\epsilon = -1$ corresponding to the Bianchi type III models, and $\epsilon = 1$ to the Kantowski-Sachs spacetimes.[61]

In this work we are mainly focused on the hydrostatic equilibrium situations. Thus, our interest will be directed to understanding whether it is possible to find a correspondence between the TOV equation of static equilibrium, and some condition applying to the spatially homogeneous models.

Imposing staticity amounts in the present case to have $\Theta_3 = 0$, $\mathcal{L}_n\Theta_3 = 0$, and $\sigma = 0$ in Eqs. (46) and (49). Reconciling the reduced equations simply requires

$$\rho + P = 0\,, \tag{50}$$

in the $\epsilon = \pm 1$ cases, and has no realisation when $\epsilon = 0$, as $\rho = 0$ from Eq.(46). Hence we conclude that the TOV condition interpreted as a cornerstone of stability yields the well-known equation of state characterising a cosmological constant in the non flat cases. In hindsight, one could have anticipated this result which emerges here in a self-consistent way. Moreover we see that strong energy condition (SEC) is violated for both cases $\epsilon = \pm 1$, whilst the WEC is additionally violated for the $\epsilon = -1$ case, as follows from Eq. (46).

6. An illustration: Incompressible fluid solutions

Using our unified TOV equation, Eq. (40), we may look for static perfect fluid solutions for all three symmetries considered here. By choosing a timelike coordinate T along the flow, making $n_a = -\alpha(Y)\mathrm{d}T$, and the warp factor Y, we obtain the following line element in the (T, Y) coordinates:

$$\mathrm{d}s^2 = -\alpha^2(Y)\mathrm{d}T + \frac{\mathrm{d}Y^2}{\epsilon - \frac{2\mu(Y)}{Y}} + Y^2\mathrm{d}\Omega_\epsilon\,, \tag{51}$$

where $\mathrm{d}\Omega_\epsilon = \left(\mathrm{d}\theta^2 + S_\epsilon^2\mathrm{d}\phi^2\right)$ and the functions α and μ will be given by solving Einstein equations, i.e., Eqs. (40) and (30).

Here, we will apply our unified treatment to find the analogs of Schwarzschild interior solution, that is, we will use the equation of state of an incompressible fluid $\rho = \rho_0$ constant. It is important to note that, as we have discussed in Sec. 5.1, the static solutions with $\epsilon \neq 1$ violate the WEC, therefore we should take $\rho_0 < 0$ in those cases.

Equation (36) implies

$$\frac{\alpha'}{\alpha} = -\frac{P'}{\rho + P} \Rightarrow \alpha = \frac{c_0}{\rho_0 + P}\,, \tag{52}$$

where c_0 is an integration constant that can be set by rescaling the time coordinate and the prime denotes Y differentiation.

Equation (43) gives us

$$\mu(Y) = \frac{4\pi\rho_0 Y^3}{3}\,, \tag{53}$$

which we replace in Eq. (40) to obtain

$$P(Y) = \rho_0 \left(\frac{2\sqrt{|\epsilon - \frac{Y_s}{Y_g}|}}{3\sqrt{|\epsilon - \frac{Y_s}{Y_g}|} - \sqrt{|\epsilon - \frac{Y_s Y^2}{Y_s^3}|}} - 1 \right)\,. \tag{54}$$

where Y_g is the analog of the radius of the object and is the least positive number that satisfy $P(Y_g) = 0$, $Y_s = \dfrac{8\pi\rho_0 Y_g^3}{3}$ is the analog of the Schwarzschild radius,

although it can not be interpreted as a location since it will be a negative number. This gives

$$\alpha = \frac{1}{2}\left(3\sqrt{\left|\epsilon - \frac{Y_s}{Y_g}\right|} - \sqrt{\left|\epsilon - \frac{Y_s Y^2}{Y_g^3}\right|}\right) \tag{55}$$

which has a similar form to the interior Schwarzschild solution, where we only change the sign of the mass-energy parameter and change the value of ϵ in the formula. Of course the physical properties are very distinct, since the solutions violate the WEC.

In Fig. 1 we compare the pressure for the three cases. From the slope of the curves, we notice that only the hyperbolic case presents $P' > 0$, compensating the repulsive gravity force in this setup. This is the opposite of the more familiar situations presented in the spherical and planar cases, where gravity is attractive, with $P' < 0$ sustaining the weight of the configuration. We can also see that the planar case admits a positive pressure for $0 < Y < Y_g$. That means that, as long as mass-energy is negative, we may have static plane configurations over a finite Y interval. On the other hand, the hyperbolic solution only admits positive pressure for $Y > Y_g$, so there is no analog of the Schwarzschild interior solution for this foliation, although it can be interpreted as an exterior fluid solution to an internal void. It can thus be matched to a hyperbolic vacuum solution for $Y < Y_g$, found as a particular case in Ref.[31]:

$$\mathrm{d}s^2 = -\left(\frac{2m}{Y} - 1\right)\mathrm{d}t^2 + \frac{\mathrm{d}Y^2}{\frac{2m}{Y} - 1} + Y^2(\mathrm{d}\theta^2 + \sinh^2\theta \mathrm{d}\phi^2)\,, \tag{56}$$

where the parameter $m = |\mu|$. The peculiarities of the hyperbolic solutions with regard to the energy conditions are also found in the work.[42]

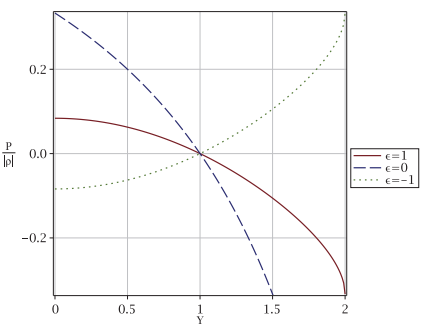


Fig. 1. Pressure as function of Y for $Y_g = 1$ and $|Y_s| = 0.25$ for $\epsilon = 1$, $\epsilon = 0$ and $\epsilon = -1$.

The equation of state consisting of a negative energy density with a positive pressure might be achieved by some kind of phantom field, but a Lagrangian description of the fluid is beyond the scope of this work. However, our simple incompressible

model, with constant energy density, but varying pressure, is reminiscent of a constant time surface of a McVittie or Shah-Vaidya spacetime, which admits a Lagrangian scalar field as source.[62,63] This suggests the possibility that there exist field models in the literature which can source the solutions presented in this paper.

We notice that the planar solutions may also represent a subclass of cylindrical solutions (see Refs.[64,65]) if we select one coordinate along the plane to be periodic. Thus, our planar solution may be interpreted as a static cylinder of fluid with boundary given by $Y = Y_g$. At this surface, the solution must be matched with a vacuum solution.

Actually, all fluid configurations found can be matched with the corresponding static solutions presented in Ref.[31] which arise from applying Birkhoff theorem for external fluid sources that satisfy the hypotheses of the theorem. In those cases there is a matter content present outside, the most common examples being an electromagnetic field, and a cosmological constant. Therefore the matching surface will correspond to a surface where $P(Y)$ matches the pressure of the exterior solution, in a manner similar to the way in which an uncompressible charged sphere is matched to a Reissner-Nordström solution in Ref.[66].

7. Conclusion

In this paper we analysed spacetimes with a two-dimensional maximally symmetric foliation sourced by a perfect fluid. We proved that in those cases, if there is a Killing vector orthogonal to the leaves, its two-expansion vanishes, which allows us to simplify our dynamical equations in terms of the two-expansion of a unit vector orthogonal to the Killing field.

When the Killing vector χ^a is timelike, we find that the flow lines must be tangent to χ^a, and as this is true at all times, the equations describe a hydrostatic equilibrium, governed by a (generalised) TOV equation. When the Killing vector is spacelike, we have instead a spatially homogeneous dynamical spacetime. The result is a subclass of Bianchi universes, with only one shear degree of freedom. The corresponding equation gives the evolution of expansion and shear scalars.

Our approach relates the mass parameter to the geometrically defined quasi-local mass-energies of Misner-Sharp and Hawking-Hayward by slightly changing its definition in order to apply it to our infinite mass-energy distributions. This innovation is in itself a step towards addressing the open issue of defining mass/energy in gravitation and cosmology, c.f the recent works of,[67] and others[68,69] on this subject.

Using these concepts we could recover the physical interpretation of the geometrical quantities appearing in the equilibrium/evolution equations, translate the dual null formalism to the more usual relativistic fluid dynamics framework, and show that the TOV equation arises as a particular case of those equations. Henceforth the generalizations of the TOV equation appear automatically by just setting $\epsilon = 0$ or -1 accordingly.

From this treatment it emerges the fact that the only static fluid solutions that satisfy the WEC are the spherical ones, as the other two cases require a negative energy density.

In what regards the spatially homogeneous spacetimes, the hydrostatic equilibrium condition also implies a violation of the SEC[70, 71] for the non planar solutions, constraining the equation of state for the perfect fluid to be that of a cosmological constant[d].

In order to illustrate the analogy between the planar, hyperbolic and spherical cases we studied the static solutions for an incompressible fluid. We found that, besides the known case of spherically symmetric spacetimes, we can obtain a static interior fluid configuration only in the case of planar symmetric spacetimes. In the hyperbolic case, the static configuration is an exterior solution that can surround an inner vacuum region.

Our unified way to describe three classes of spacetimes foliated with codimension-two leaves of constant curvature leads the way to further generalizations that we will address elsewhere.

Acknowledgements

The work of M.Le D. has been supported by Lanzhou University starting fund and by the Fundamental Research Funds for the Central Universities (Grant No.lzujbky-2019-25). The work of JPM was supported by FCT/MCTES through national funds (PIDDAC) by the grants UID/FIS/04434/2019, UIDB/04434/2020 and UIDP/04434/2020. as well as by CERN/FIS-PAR/0037/2019 and PTDC/FIS-OUT/29048/2017. The authors wish to thank Xu Yumeng for helpful discussions.

References

1. R. C. Tolman, Effect of inhomogeneity on cosmological models, *Proc. Natl. Acad. Sci. U.S.A.* **20**, 169 (March 1934).
2. J. R. Oppenheimer and G. M. Volkoff, On massive neutron cores, *Phys. Rev.* **55**, 374 (February 1939).
3. W. H. Zurek and D. N. Page, Black-hole thermodynamics and singular solutions of the Tolman-Oppenheimer-Volkoff equation, *Phys. Rev.* **D29**, 628 (1984).
4. M. S. R. Delgaty and K. Lake, Physical acceptability of isolated, static, spherically symmetric, perfect fluid solutions of Einstein's equations, *Comput. Phys. Commun.* **115**, 395 (1998).
5. S. Berger, R. Hojman and J. Santamarina, General Exact Solutions of Einstein Equations for Static Perfect Fluids With Spherical Symmetry, *J. Math. Phys.* **28**, p. 2949 (1987).
6. G. Fodor, Generating spherically symmetric static perfect fluid solutions (2000).
7. K. Dev and M. Gleiser, Anisotropic stars: Exact solutions, *Gen. Rel. Grav.* **34**, 1793 (2002).

[d]There are several arguments in the literature suggesting the canonical energy conditions should be abandoned as a criterion of viability.[72]

8. S. Rahman and M. Visser, Space-time geometry of static fluid spheres, *Class. Quant. Grav.* **19**, 935 (2002).
9. V. Gorini, U. Moschella, A. Yu. Kamenshchik, V. Pasquier and A. A. Starobinsky, Tolman-Oppenheimer-Volkoff equations in presence of the Chaplygin gas: stars and wormhole-like solutions, *Phys. Rev.* **D78**, p. 064064 (2008).
10. L. Zou, F.-Y. Li and H. Wen, Exact solutions of the Einstein-Maxwell equations in charged perfect fluid spheres for the generalized Tolman-Oppenheimer-Volkoff equations, *Int. J. Mod. Phys.* **D22**, p. 1350009 (2013).
11. L. Herrera and W. Barreto, General relativistic polytropes for anisotropic matter: The general formalism and applications, *Phys. Rev.* **D88**, p. 084022 (2013).
12. F. Özel and P. Freire, Masses, Radii, and the Equation of State of Neutron Stars, *Ann. Rev. Astron. Astrophys.* **54**, 401 (2016).
13. L. Herrera and A. Di Prisco, Self-similarity in static axially symmetric relativistic fluids, *Int. J. Mod. Phys.* **D27**, p. 1750176 (2017).
14. V. Faraoni, Embedding black holes and other inhomogeneities in the universe in various theories of gravity: a short review, *Universe* **4**, p. 109 (2018).
15. Y. X. Martins, L. F. A. Campos, D. d. S. P. Teixeira and R. J. Biezuner, Existence and Classification of Pseudo-Asymptotic Solutions for Tolman-Oppenheimer-Volkoff Systems, *Annals Phys.* **409**, p. 167929 (2019).
16. J. D. V. Arbañil, J. P. S. Lemos and V. T. Zanchin, Polytropic spheres with electric charge: compact stars, the Oppenheimer-Volkoff and Buchdahl limits, and quasiblack holes, *Phys. Rev.* **D88**, p. 084023 (2013).
17. S. Carloni and D. Vernieri, Covariant Tolman-Oppenheimer-Volkoff equations. I. The isotropic case, *Phys. Rev.* **D97**, p. 124056 (2018).
18. S. Carloni and D. Vernieri, Covariant Tolman-Oppenheimer-Volkoff equations. II. The anisotropic case, *Phys. Rev.* **D97**, p. 124057 (2018).
19. A. A. Isayev, Comment on "Covariant Tolman-Oppenheimer-Volkoff equations. II. The anisotropic case", *Phys. Rev.* **D98**, p. 088503 (2018).
20. P. Papadopoulos and J. A. Font, Relativistic hydrodynamics on space - like and null surfaces: Formalism and computations of spherically symmetric space-times, *Phys. Rev.* **D61**, p. 024015 (2000).
21. K. Lake, All static spherically symmetric perfect-fluid solutions of Einstein's equations, *Phys. Rev. D* **67**, p. 104015 (May 2003).
22. P. Boonserm, M. Visser and S. Weinfurtner, Solution generating theorems for the TOV equation, *Phys. Rev.* **D76**, p. 044024 (2007).
23. G. J. Olmo, H. Sanchis-Alepuz and S. Tripathi, Stellar Structure Equations in Extended Palatini Gravity, *Phys. Rev.* **D86**, p. 104039 (2012).
24. K. Glampedakis, G. Pappas, H. O. Silva and E. Berti, Post-Tolman-Oppenheimer-Volkoff formalism for relativistic stars, *Phys. Rev.* **D92**, p. 024056 (2015).
25. R. Jain, B. G. Sidharth and C. Corda, On the Jeans Theorem and the 'Tolman-Oppenheimer-Volkoff Equation' in R^2 Gravity, *Adv. High Energy Phys.* **2016**, p. 2601741 (2016).
26. H. Velten, A. M. Oliveira and A. Wojnar, A free parametrized TOV: Modified Gravity from Newtonian to Relativistic Stars, *PoS* **MPCS2015**, p. 025 (2016).
27. A. Wojnar and H. Velten, Equilibrium and stability of relativistic stars in extended theories of gravity, *Eur. Phys. J.* **C76**, p. 697 (2016).
28. K. A. Bronnikov, J. C. Fabris, O. F. Piattella and E. C. Santos, Static, spherically symmetric solutions with a scalar field in Rastall gravity, *Gen. Rel. Grav.* **48**, p. 162 (2016).

29. A. V. Astashenok, S. D. Odintsov and A. de la Cruz-Dombriz, The realistic models of relativistic stars in $f(R) = R + \alpha R^2$ gravity, *Class. Quant. Grav.* **34**, p. 205008 (2017).
30. C. Bona, A new proof of the generalized birkhoff theorem, *Journal of Mathematical Physics* **29**, 1440 (1988).
31. A. Maciel, M. Le Delliou and J. P. Mimoso, Revisiting the Birkhoff theorem from a dual null point of view, *Phys. Rev.* **D98**, p. 024016 (2018).
32. S. A. Hayward, Dual - null dynamics, *Ann. Inst. H. Poincare Phys. Theor.* **59**, 399 (1993).
33. S. A. Hayward, Dual-null dynamics of the einstein field, *Classical and Quantum Gravity* **10**, 779 (apr 1993).
34. S. A. Hayward, The general solution to the einstein equations on a null surface, *Classical and Quantum Gravity* **10**, 773 (apr 1993).
35. S. A. Hayward, General laws of black-hole dynamics, *Phys. Rev. D* **49**, 6467 (June 1994).
36. J. M. Senovilla, Trapped surfaces, *Int. J. Mod. Phys. D* **20**, p. 2139 (2011).
37. S. A. Hayward, Quasilocal gravitational energy, *Phys. Rev. D* **49**, 831 (January 1994).
38. A. Maciel, M. Le Delliou and J. P. Mimoso, Dual null formalism for the collapse of fluids in a cosmological background, *Phys. Rev.* **D92**, p. 083525 (2015).
39. A. Maciel, Quasilocal approach to general universal horizons, *Phys. Rev. D* **93**, p. 104013 (2016).
40. A. Maciel, M. Le Delliou and J. P. Mimoso, New perspectives on the TOV equilibrium from a dual null approach, *Classical and Quantum Gravity* **37**, p. 125005 (June 2020).
41. C. Cattoen and M. Visser, Necessary and sufficient conditions for big bangs, bounces, crunches, rips, sudden singularities, and extremality events, *Class.Quant.Grav.* **22**, 4913 (2005).
42. F. S. N. Lobo and J. P. Mimoso, Possibility of hyperbolic tunneling, *Phys. Rev.* **D82**, p. 044034 (2010).
43. R. M. Wald, *General Relativity* (The University of Chicago Press, Chicago, 1984).
44. S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).
45. S. Carroll, *Spacetime and Geometry* (Addison-Wesley Publishing Company, San Francisco, 2004).
46. A. Maciel, M. Le Delliou and J. P. Mimoso, New perspectives on the TOV equilibrium from a dual null approach, *Class. Quant. Grav.* **37**, p. 125005 (2020).
47. J. P. Mimoso, M. Le Delliou and F. C. Mena, Separating expansion from contraction in spherically symmetric models with a perfect fluid: Generalization of the tolman-oppenheimer-volkoff condition and application to models with a cosmological constant, *Phys. Rev. D* **81**, p. 123514 (June 2010).
48. C. W. Misner and D. H. Sharp, Relativistic equations for adiabatic, spherically symmetric gravitational collapse, *Phys. Rev.* **136**, B571 (October 1964).
49. S. A. Hayward, Gravitational energy in spherical symmetry, *Phys. Rev. D* **53**, 1938 (1996).
50. V. Faraoni, Is the Hawking quasilocal energy "Newtonian"?, *Symmetry* **7**, 2038 (2015).
51. S. A. Hayward, Unified first law of black-hole dynamics and relativistic thermodynamics, *Classical Quantum Gravity* **15**, 3147 (1998).
52. A. Prain, V. Vitagliano, V. Faraoni and M. Lapierre-Léonard, Hawking–Hayward quasi-local energy under conformal transformations, *Class. Quant. Grav.* **33**, p. 145008 (2016).

53. V. Faraoni and J. Coté, Quasilocal mass and multipole expansion in scalar-tensor gravity (2019).
54. R. L. Arnowitt, S. Deser and C. W. Misner, Dynamical structure and definition of energy in general relativity, *Phys. Rev.* **116**, 1322 (1959).
55. H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers and E. Herlt, *Exact Solutions of Einstein's Field Equations*, Cambridge Monographs on Mathematical Physics, 2nd edn. (Cambridge University Press, Cambridge, England, 2009).
56. P. Martin-Moruno and M. Visser, Classical and semi-classical energy conditions, *Fundam. Theor. Phys.* **189**, 193 (2017).
57. G. F. R. Ellis and H. van Elst, Cosmological models: Cargese lectures 1998, *NATO Sci. Ser. C* **541**, 1 (1999).
58. J. P. Mimoso and P. Crawford, Shear - free anisotropic cosmological models, *Class. Quant. Grav.* **10**, 315 (1993).
59. L. Bianchi, On the three-dimensional spaces which admit a continuous group of motions, *General Relativity and Gravitation* **33**, 2171 (2001).
60. G. F. R. Ellis and M. A. H. MacCallum, A Class of homogeneous cosmological models, *Commun. Math. Phys.* **12**, 108 (1969).
61. R. Kantowski and R. K. Sachs, Some spatially homogeneous anisotropic relativistic cosmological models, *J. Math. Phys.* **7**, p. 443 (1966).
62. E. Abdalla, N. Afshordi, M. Fontanini, D. C. Guariento and E. Papantonopoulos, Cosmological black holes from self-gravitating fields, *Phys. Rev. D* **89**, p. 104018 (May 2014).
63. D. C. Guariento, A. Maciel, M. M. C. Mello and V. T. Zanchin, Charged cosmological black holes: a thorough study of a family of solutions, *Phys. Rev.* **D100**, p. 104050 (2019).
64. F. C. Mena, Cylindrically symmetric models of gravitational collapse to black holes: A short review, *Int. J. Mod. Phys.* **D24**, p. 1542021 (2015).
65. K. Bronnikov, N. O. Santos and A. Wang, Cylindrical Systems in General Relativity (2019).
66. J. D. V. Arbañil, J. P. S. Lemos and V. T. Zanchin, Incompressible relativistic spheres: Electrically charged stars, compactness bounds, and quasiblack hole configurations, *Phys. Rev.* **D89**, p. 104054 (2014).
67. V. Faraoni, Quasilocal energy in modified gravity, *Classical Quantum Gravity* **33**, p. 015007 (2016).
68. H. Barzegar, P. T. Chrusciel and M. Hörzinger, Energy in higher-dimensional spacetimes, *Phys. Rev.* **D96**, p. 124002 (2017).
69. M.-T. Wang and S.-T. Yau, Quasilocal mass in general relativity, *Phys. Rev. Lett.* **102**, p. 021101 (2009).
70. S. Capozziello, F. S. N. Lobo and J. P. Mimoso, Energy conditions in modified gravity, *Phys. Lett.* **B730**, 280 (2014).
71. S. Capozziello, F. S. N. Lobo and J. P. Mimoso, Generalized energy conditions in Extended Theories of Gravity, *Phys. Rev.* **D91**, p. 124019 (2015).
72. C. Barcelo and M. Visser, Twilight for the energy conditions?, *Int. J. Mod. Phys.* **D11**, 1553 (2002).

A spherically symmetric stiff fluid spacetime in light of cosmic structure formation

Daniele Gregoris

School of Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China
E-mail: danielegregoris@libero.it

We will report here a critical inspection of the Penrose conjecture according to which the gravitational entropy should be quantified via the Weyl curvature, with the Clifton-Ellis-Tavakol entropy being one specific realization of this proposal. In fact, we will show that in some exact inhomogeneous and anisotropic cosmological models which arise as exact solutions in general relativity with either closed and open topologies, the Clifton-Ellis-Tavakol gravitational entropy is increasing in time despite the decrease of the magnitude of the Weyl curvature: this is possible thanks to the growth of the spatial shearing effects. The matter content driving the dynamics of this class of models comes in the form of a stiffened fluid which can be relevant in the early universe. We choose the values of the free parameters entering the metric tensor consistently with the holographic principle and the second law of thermodynamics. Our study can be of interest in light of the modeling of the formation of some primordial structures, like the Large Quasar Groups, as suggested by the growth of gravitational entropy, and whose existence cannot be accounted for by standard perturbation methods over a homogeneous background.

Keywords: Exact solution; stiff matter; gravitational entropy.

1. Introduction

Our Universe contains a plethora of different astrophysical structures with different sizes such as galaxies, clusters of galaxies, filaments, voids,... The formation process for cosmic structures of sizes smaller than 150 Mpc, which is the length scale above which the universe is considered to be homogeneous,[1] can be described with perturbation schemes applied to a homogeneous background as the Friedman-Lemaître-Robertson-Walker one.[2,3] However, there is observational evidence of the existence of astrophysical structures like the one known as the "axis of evil"[4] which challenge the Copernican principle because they come with an alignment of matter along a preferred direction and their sizes are larger than 150 Mpc. Other examples are the Large Quasar Groups with sizes in the range of 70-350 Mpc.[5–8] Accounting for their existence by means of perturbation methods is problematic because the cosmic material would not cluster quickly enough for forming a bound system of this size: this is one among few other open questions affecting the standard model of cosmology.[12] Inhomogeneous cosmology may offer an alternative route for developing theoretical frameworks which can describe the evolution of matter perturbations undergoing a local collapse which can lead to the formation of an astrophysical structure. For example, in a dust Szekeres spacetime the density contrast grows eight times quicker

than in the linear perturbation regime making the process of formation of astrophysical structures more efficient.[13] There is also observational evidence of exotic astrophysical structures with non-standard size which should have formed in the early universe, as a supermassive black hole at redshift $z \sim 10$,[9,10] and the Huge Quasar Group at redshift $z \sim 1.27$,[11] whose existence cannot be explained by invoking effects played by an inhomogeneous spacetime supported by dust (which can be interpreted as pressureless dark matter). Therefore, the issue becomes to provide an appropriate theoretical framework in which spatial shear effects and tidal effects can trigger a local collapse in an overall expanding background supported by a matter content consistent with the early stages of evolution of the universe.

In the primordial universe shearing effects may play an important role, as compared to the one of the matter content, even though they are almost negligible at the present day.[14] Moreover, it has been argued that in early cosmic epochs the matter content may come in the form of a stiff fluid[a],[15–19] which is the effective hydrodynamical realization of a massless scalar field according to the canonical formalism.[20,21] More in detail, in[18,19] it has been argued that the early universe should be filled with cold baryons which fulfill the equation of state $p = \rho$, and that if cosmic matter is described as a relativistic self-gravitating Bose-Einstein condensate, then the cosmic evolution would experience a stiff-matter dominated era. Indeed the development of specific algorithms for integrating the Einstein field equations when the matter content is a stiff fluid has received attention in the literature;[22–28] then a number of exact and analytical solutions expressible in terms of elementary functions has been found obeying to a variety of symmetry groups,[29–32] just to mention a few examples.

One specific solution to the Einstein equations that we have adopted in our own research,[33] and which was found and investigated from the mathematical perspective by a number of different authors,[34–38] reads as (see also page 261 in[39]):

$$ds^2 = -\left(\frac{r}{2}\right)^2 dt^2 + \frac{dr^2}{\epsilon + Cr^2} + r^2\left[\frac{\epsilon}{2} + h(t)\right](d\theta^2 + \sin^2\theta d\phi^2), \tag{1}$$

where the "generalized scale factor" can either be

$$h(t) = A\sin(t) + B\cos(t) \qquad \text{if} \qquad \epsilon = -1\,, \tag{2}$$

$$h(t) = -\left(\frac{t}{2}\right)^2 + 2At + B \qquad \text{if} \qquad \epsilon = 0\,, \tag{3}$$

$$h(t) = Ae^t + Be^{-t} \qquad \text{if} \qquad \epsilon = 1\,, \tag{4}$$

according to the topology of the spacetime. We can note that the model (1) depends on three free parameters A, B, and C, and that only its angular part, and not the

[a] A stiff fluid is a perfect fluid, e.g. it is fully described by its adiabatic pressure p and energy density ρ, whose equation of state in natural units is $p = \rho$, and with matter-energy tensor $T^{\mu}{}_{\nu} = \mathrm{diag}[-\rho, p, p, p]$.

radial one, is expanding. The spacetime is supported by a fluid obeying to a stiffened equation of state

$$p = \rho + \frac{3C}{4\pi}\,. \tag{5}$$

which reduces to the case of a stiff fluid, i.e. of a massless scalar field, in the limit of $C \to 0$. Thus, this latter parameter, which does not affect directly the time evolution of the generalized scale factor $h(t)$ as we can appreciate from (2)-(3)-(4), would instead make the cosmic fluid pressure to be non-zero but equal to a constant in an empty space for which $\rho = 0$; more in general th equation of state (5) can be written in the form $p = \omega(\rho)\rho$ in which the density-dependent equation of state parameter can be interpreted as a chameleon field.[40,41] The Hubble function, invariant shear, cosmic energy density and Weyl curvature of the spacetime (1) are given by

$$H = \frac{4\dot{h}(t)}{3(\epsilon + 2h(t))r}\,, \tag{6}$$

$$\sigma^2 = \frac{4\dot{h}(t)^2}{3(\epsilon + 2h(t))^2 r^2} = \frac{3H^2}{4}\,, \tag{7}$$

$$\rho = -\frac{c^2(R + 18C)}{16\pi G}\,, \qquad R = -2\frac{\epsilon + \mathcal{R} + 6Cr^2(\epsilon + 2h(t))^2}{r^2(\epsilon + 2h(t))^2}\,, \tag{8}$$

$$\Psi_2 = -\frac{\mathcal{R} + \epsilon}{3r^2(2h(t) + \epsilon)^2}\,, \tag{9}$$

where an overdot denotes a derivative with respect to the cosmic time. In the computation of the latter quantity we have applied the Newman-Penrose formalism,[42] and the following notation has been introduced:

$$\begin{aligned} &\mathcal{R} = 4(A^2 + B^2) \qquad \text{for} \qquad \epsilon = -1\,, \\ &\mathcal{R} = 4(4A^2 + B) \qquad \text{for} \qquad \epsilon = 0\,, \\ &\mathcal{R} = -16AB \qquad \text{for} \qquad \epsilon = 1\,. \end{aligned} \tag{10}$$

Since we have a nonzero Weyl scalar Ψ_2, some tidal forces are present in the cosmological model (1),[43] and they can trigger a local gravitational collapse whose evolution then experiences the non-standard spatial shearing effects described by σ^2, potentially taming the previously mentioned problems of accounting for the existence of some astrophysical structures observed in the early universe. For exploring this topic, in our research[33] we have computed the gravitational entropy for the cosmology (1) by following the Clifton-Ellis-Tavakol proposal,[44] in which we have implemented some numerical values for the free parameters A, B, and C which are consistent with the thermodynamical requirements formulated through the cosmological holographic principle and the second law of thermodynamics. We have obtained that the gravitational entropy can be increasing even in the time intervals in which the strength of the Weyl curvature is decreasing, and therefore we have claimed that *Thermodynamics of shearing massless scalar field spacetimes*

is inconsistent with the Weyl curvature hypothesis according to which the conformal curvature can be adopted as a measure for the gravitational entropy.[45] Our result is due to the effects played by the spatial shear which accounts for the spatial anisotropies.

This *MG16 conference proceeding* is organized as follows: in Sect. 2 we impose the holographic principle and the second law of thermodynamics to the spacetime (1) showing how they can set some constraints on the behavior of various relevant physical quantities. In Sect. 3 we introduce the Clifton-Ellis-Tavakol entropy, discuss its importance in the modeling of the formation of astrophysical structures, and report on our specific results for the spacetime under investigation. In Sect. 4 we formulate the two messages that according to us the reader should take at home regardless the technicalities involved in this project. Finally in Sect. 5 we put our research into the larger perspective of the studies of gravitational entropy in inhomogeneous cosmology.

2. Imposing the thermodynamical requirements

According to the cosmological holographic principle, as formulated by Bousso, the matter entropy S_m inside a region bounded by a "horizon" should be smaller than the area A_H of the horizon itself, or more precisely $S_m \leq A_H/4$.[46] We can see that the celebrated Bekenstein-Hawking entropy for static black holes[47,48] constitutes the limiting case of this more general principle. For a dynamical spacetime we consider appropriate to work with the dynamical apparent horizon as the boundary of the region we are interested in; by introducing the areal radius

$$\tilde{r} = r\sqrt{\frac{\epsilon}{2} + h(t)}\,, \tag{11}$$

the location $\tilde{r}_H$ of the dynamical apparent horizon is such that[49]

$$||\nabla\tilde{r}||^2_{\tilde{r}=\tilde{r}_H} = 0\,. \tag{12}$$

In a time-evolving configuration, the dynamical apparent horizon is not the unique choice for the boundary of the region to which the cosmological holographic principle can be imposed; however in the homogeneous and isotropic Friedman universe assuming the space to be filled with some form of dark energy the first and second law of thermodynamics have been shown to hold for the dynamical apparent horizon, but not if one works with the cosmic horizon,[50] providing a motivation for choosing our route.[51,52]

Since the fluid energy density should be non-negative all along the time evolution of the system, we found that C should be non-positive; under this restriction, equation (12) admits a solution only for the topology $\epsilon = 1$. Setting the parameters A and B to be positive guarantees both a well-defined cosmic energy density and the existence of a solution for the equation determining the location of the dynamical apparent horizon. Taking into account that $A_H = 4\pi\tilde{r}_H^2$ and that $S_m = \alpha\tilde{r}_H^3$, with

α an overall proportionality constant, it can be seen after some computations that the holographic principle is further constraining the value of C, i.e. of the deviations of the equation of state of the cosmic matter from a stiff fluid. The cosmological consequences of these restrictions on the values of the arbitrary model parameters are:

- If we compute the deceleration parameter as in,[53] it would be negative;
- Since $\epsilon + 2h(t) \neq 0$, there would be no initial singularity in this cosmology because there would not exist a time t_B at which the energy density (8) would diverge.

On the other hand, imposing the second law of thermodynamics to the matter content would require $\dot{S}_m > 0$, which can be translated into the equivalent condition $\frac{d\tilde{r}_H}{dt} > 0$. After some algebraic manipulations, we could recast this condition as $Ae^t - Be^{-t} > 0$; thus, the cosmological consequences within the model (1) of imposing the second law are:

- The lower limit on the size of the universe $h(t) > 2Be^{-t}$ is established;
- The lower limit on the age of universe $t > \frac{1}{2}\ln\frac{B}{A}$ is established;
- The Weyl curvature Ψ_2 is a time-decreasing quantity.

3. Computing the gravitational entropy

"Entropy" can be naively regarded as a measure of the disorder within a system (as related to the existence of an arrow of time), or as a quantification of the number of different microscopic realizations compatible with the same macroscopic system (statistical entropy). Therefore, in cosmology, as astrophysical structures form in one specific spatial regions rather than in others with the whole universe remaining homogeneous on large enough length scales, the gravitational entropy should increase. Since when an observer looks at the universe from its location it may or may not detect an astrophysical structure, the growth of gravitational entropy should be accompanied by an increase of the spatial anisotropy. One specific way of computing the density $s_{\rm grav}$ of gravitational entropy in Petrov type D spacetime, as for example (1), is due to Clifton-Ellis-Tavakol[44]:

$$T_{\rm grav}\dot{s}_{\rm grav} = -dV\sigma_{\mu\nu}\left(\pi^{\mu\nu}_{\rm grav} + \frac{(\rho+p)}{3\rho_{\rm grav}}E^{\mu\nu}\right). \tag{13}$$

The ingredients of this formula are the following:

- The temperature of the free gravitational field defined as

$$T_{\rm grav} = \frac{|u_{a;b}l^a n^b|}{\pi} = \frac{r}{8\pi\sqrt{Cr^2+\epsilon}}, \tag{14}$$

  where u^a is the matter-comoving observer four-velocity, and l^a and n^a are part of the null coframe containing the dt and dr contributions.

- The volume element
$$dV = \frac{r^2 \sin\theta(\epsilon + 2h(t))}{2\sqrt{Cr^2+\epsilon}}\, dr d\theta d\phi\,. \tag{15}$$
- The shear tensor $\sigma_{\mu\nu}$ which describes the rate of distortion of a given region with fixed volume during the cosmic evolution.
- The gravitational energy defined in terms of the Weyl curvature as:
$$\rho_{\rm grav} := 16\pi|\Psi_2|\,. \tag{16}$$
- The gravitational anisotropic pressure also defined in terms of the Weyl curvature as:
$$\pi^{\rm grav}_{\mu\nu} := \frac{|\Psi_2|}{16\pi}(-x_a x_b + y_a y_b + z_a z_b + u^a u^b)\,, \tag{17}$$
in which x_a, y_a, z_a constitute the spacelike vectors in the orthonormal base for the metric (1).
- The gravito-electric tensor $E_{\mu\nu}$ which accounts for tidal effects, which are indeed necessary for having a gravitational collapse.

We have considered the approach (13) to be a sound proposal for computing the gravitational entropy because, as pointed out in,[44] it delivers a non-negative entropy, it is consistent with the Hawking-Bekenstein entropy of black holes, it provides a vanishing entropy in and only in conformally flat spacetimes (as for example in the homogeneous and isotropic Friedman-Lemaître-Robertson-Walker universe in which astrophysical structures cannot form due to the lack of any matter density contrast), and it is related to the growth of anisotropies in the spacetime (as they indeed increase during the structure formation phase).

Thus, it does not come as a surprise that the conjecture of a connection between gravitational entropy and Weyl curvature has been regarded as a useful tool for the description of the formation of astrophysical structures in various inhomogeneous models. For example, in[54] a relationship between the growth of gravitational entropy and the amplitude of initial fluctuations of spatial curvature at the last scattering time has been found, while the possible saturation of its value at a certain time has been investigated in[55] for the Lemaître-Tolman-Bondi universe, in[56] more in general for the class of silent universes for which the gravitomagnetic part of the Weyl curvature vanishes, and in[57] for a perturbed Friedman spacetime. Furthermore, a gravitational entropy quantified via the Weyl curvature is consistent with the generalized second law in the void Lemaître-Tolman-Bondi universe,[58] and the specific Clifton-Ellis-Tavakol entropy has been shown to be a time increasing function in a number of cosmological models.[59]

Our explicit computations for the Clifton-Ellis-Tavakol gravitational entropy delivered the result
$$T_{\rm grav}\dot{s}_{\rm grav} = dV\frac{64\pi\dot{h}(t)(1-16AB)}{3(2h(t)+\epsilon)^3 r^3}\,. \tag{18}$$

Here we have restricted ourselves to the spacetime (4) because, as mentioned in Sect. 2, it is the only case within (1) for which a thermodynamical investigation based on the cosmological holographic principle was possible. Then, implementing the constraints on the free parameters obtained from such analysis and those arising from the second law of thermodynamics, we have found that a challenge to the Weyl curvature conjecture proposed by Penrose[45] is posed. In fact, the gravitational entropy is increasing even when the gravito-electric curvature as measured by Ψ_2 in (9) is not, and thus the latter cannot be considered too naively as a measure of the former.

4. Discussion

Quantifying the effects that the interactions between small-scale inhomogeneities (e.g. galaxies, galaxy clusters, voids, filaments, etc...) have on the large-scale evolution of the Universe is not a trivial task, and the debate in the literature about their importance is still open.[61] There exists in fact a range of completely different claims from that they have no effects at all to the one that they can bring an energy budget which can be responsible for the observed accelerated expansion of the universe without the need of a cosmological constant; see[62] for a review of different approaches and their predictions. Once endowed with a cosmological model formulated in terms of a metric tensor solution of the field equations of the underlying gravitational theory, it is necessary to investigate its physical applicability. Some widely applied cosmological tests which can assess the suitability of a mathematical solution of the Einstein equations as a cosmological model rely on supernovae data,[63–65] cosmic microwave background data,[66,67] or baryon acoustic oscillations data,[68,69] and clearly require data-analysis skills. For example, the inhomogeneous dust Lemaître-Tolman-Bondi model has been shown to be in tension with the kinematic Sunyaev-Zel'dovich effect[70] although it is not with observations related to the luminosity distance of supernovae, without the need of dark energy but assuming that the observer is located inside a giant void;[71] more in general the development of "model independent" tests for inhomogeneous cosmology constitutes a line of current research.[72–74] However, in Sect. 2 we have explained how "theoretical" requirements can also be exploited for constraining the values of the free model parameters and for extracting cosmological information about the model under investigation. We had previously applied the same way of thinking to the Stephani cosmological model by estimating the strength of spatial inhomogeneities, the matter abundance and the size of the universe by imposing the cosmological holographic principle and the second law of thermodynamics[75] (however this latter spacetime is conformally flat and therefore it does not offer any viable possibility for an investigation of the gravitational entropy). Thus, the first take at home message from our research is:

- By working with solely pen and paper, we can constrain a given cosmological model known in terms of its metric tensor by imposing the

thermodynamical requirements which follow from the cosmological holographic principle and the second law of thermodynamics.

In Sect. 3 we have mentioned that the Clifton-Ellis-Tavakol gravitational entropy is increasing in time for the spacetime (1) although the Weyl curvature is not thanks to the presence of a term given by shear tensor which is inserted by hands in its formulation. The fact that spatial shearing effects are developing in the same time interval in which the gravitational entropy is increasing is a fundamental consistency check for the mathematical formula adopted for the latter. In fact, both these quantities are expected to increase during the structure formation phase because the universe is becoming anisotropic and inhomogeneous, with the former being measured by the shear tensor while the latter can be regarded as an increase of the level of "disorder" of the system which should be quantified via an appropriate entropy quantity. Thus, our second take at home message, whose consequences will be further explored in the next Sect. 5 is:

- Implementing the Weyl curvature conjecture according to which the Weyl curvature should serve as a measure for the gravitational entropy is not as simple as it may look like at a first sight and other ingredients may be necessary in the construction of a mathematical formula for the gravitational entropy.

5. Outlook

The search for an appropriate inhomogeneous cosmological model which can provide a viable framework for addressing the issue of the formation of astrophysical structures with sizes larger than 150 Mpc (and which is also in agreement with other datasets), and the formulation of an appropriate measure for the "gravitational entropy" in the spirit of the Weyl curvature conjecture are two related problems. Therefore, we consider appropriate to investigate various notions of gravitational entropy as the following:

- $$S = \frac{C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}}{R_{\mu\nu}R^{\mu\nu}} \tag{19}$$

  where $C_{\mu\nu\rho\sigma}$ and $R_{\mu\nu}$ are respectively the Weyl and Ricci tensor, and its spatially integrated version

  $$\tilde{S} = \int dV S\,. \tag{20}$$

  This formula has arisen in the context of cosmologies exhibiting an anisotropic singularity.[76] However this proposal cannot be applied to vacuum spacetimes because of the vanishing of the Ricci tensor.

- Eq. (19) but multiplied by the square root of the determinant of the spatial metric h as suggested in[77]:

$$S = \sqrt{h}\frac{C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}}{R_{\mu\nu}R^{\mu\nu}}\,, \tag{21}$$

  which would be something in between the two proposals of the previous step.
- Applying a spatially averaging scheme and writing the entropy as[78,79]

$$S = V_{\mathcal{D}}\left\langle \rho \ln \frac{\rho}{\langle\rho\rangle_{\mathcal{D}}} \right\rangle_{\mathcal{D}}\,, \tag{22}$$

  where angular brackets denote an average over the spatial domain of interest $\mathcal{D}$.
- The Clifton-Ellis-Tavakol entropy that we have already talked about in Sect. 3.

These inequivalent ways of computing the gravitational entropy should be applied to the general line element describing a spherically symmetric dynamical spacetime (see page 251 in[39])

$$ds^2 = -e^{2\nu(t,r)}dt^2 + e^{2\lambda(t,r)}dr^2 + Y^2(t,r)(d\theta^2 + \sin^2\theta d\phi^2) \tag{23}$$

in a set of future projects. Intuitively, we expect that demanding the simultaneous growth of shear, Weyl curvature and gravitational entropy would provide some information on the increasing/decreasing and concavity properties of the functions $\nu(t,r)$, $\lambda(t,r)$ and $Y(t,r)$, which then will be implemented into the Einstein equations for understanding which types of matter content permit to achieve this goal.

All the previously mentioned proposals for computing the gravitational entropy are directly sensitive to the matter content of the spacetime. In fact in some of them the energy density (Hosoya-Buchert-Morita) and possibly also the pressure (Clifton-Ellis-Tavakol) appear explicitly; in other cases as in (19)-(20)-(21) they are inserted via the Ricci tensor, which is connected to the matter content of the universe as we can see from the Einstein equations written in the form

$$R_{\mu\nu} = 8\pi\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)\,. \tag{24}$$

A more conservative proposal which has been formulated recently in the context of black hole physics argues that it is possible to write the density s of gravitational entropy just in terms of the frame component of the Weyl tensor W and of its Newman-Penrose derivative DW as[80]

$$s = \left|\frac{DW}{W}\right|. \tag{25}$$

The applicability of such formalism in inhomogeneous cosmology and in particular the investigation of its relationship to the spatial anisotropies quantified by the shear tensor would constitute a subject of future studies.

Acknowledgments

The author thanks the organizers of the online 16 Marcel Grossmann conference.

References

1. V. Springel, et al., Simulating the joint evolution of quasars, galaxies and their large-scale distribution, *Nature* **435**, 629 (2005).
2. E. Lifshitz, Republication of: On the gravitational stability of the expanding universe, *J. Phys. (USSR)* **10** (1946) 116; *Gen. Rel. Grav.* **49**, 18 (2017).
3. P. J. E. Peebles, and J. T. Yu, Primeval adiabatic perturbation in an expanding universe, *Astrophys. J.* **162**, 815 (1970).
4. K. Land, and J. Magueijo, Examination of evidence for a preferred axis in the cosmic radiation anisotropy, *Phys. Rev. Lett.* **95**, 071301 (2005).
5. M. Scrimgeour, et al., The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity, *Mon. Not. Roy. Astron. Soc.* **425**, 116 (2012).
6. J. K. Yadav, J. S. Bagla, and N. Khandai, Fractal dimension as a measure of the scale of homogeneity, *Mon. Not. Roy. Astron. Soc.* **405**, 2009 (2010).
7. F. Sylos Labini, Super-homogeneity and inhomogeneities in the large scale matter distribution, *PoS Cosmology* **002**, 2009 (2009).
8. F. Sylos Labini, N. L. Vasilyev, Y. V. Baryshev, and M. Lopez-Corredoira, Absence of anti-correlations and of baryon acoustic oscillations in the galaxy correlation function from the Sloan Digital Sky Survey DR7, *Astron. Astrophys.* **505**, 981 (2009).
9. D. J. Mortlock, et al., A luminous quasar at a redshift of $z = 7.085$, *Nature* **474**, 616 (2011).
10. Y. Matsuoka et al., Subaru high-z exploration of lowluminosity quasars (SHELLQs). I. Discovery of 15 quasars and bright galaxies at $5.7 < z < 6.9$, *Astrophys. J.* **828**, 26 (2016).
11. R. G. Clowes, et al., A structure in the early universe at $z \sim 1.3$ that exceeds the homogeneity scale of the R-W concordance cosmology, *Mon. Not. Roy. Astron. Soc.* **429**, 2910 (2013).
12. A. D. Dolgov, Beasts in Lambda-CDM Zoo, *Physics of Atomic Nuclei* **80**, 987 (2017).
13. K. Bolejko, Structure formation in the quasispherical Szekeres model, *Phys. Rev. D* **73**, 123508 (2006).
14. A. Coley, and W. C. Lim, Spikes and matter inhomogeneities in massless scalar field models, *Class. Quantum Grav.* **33**, 015009 (2016).
15. E. M. Lifshitz, and I. M. Khalatnikov, Investigations in relativistic cosmology, *Adv. Phys.* **12**, 185 (1963).
16. V. A. Belinski, I. M. Khalatnikov, and E. M. Lifshitz, Oscillatory approach to the singular point in relativistic cosmology, *Adv. Phys.* **19**, 525 (1970).
17. V. A. Belinski, I. M. Khalatnikov, and E. M. Lifshitz, A general solution of the Einstein equations with a time singularity, *Adv. Phys.* **31**, 639 (1982).
18. J. B. Zel'dovic, A hypothesis, unifying the structure and the entropy of the Universe, *Mont. Not. Roy. Astron. Soc.* **160**, 1 (1972).
19. P. -H. Chavanis, Cosmology with a stiff matter era, *Phys. Rev. D* **92**, 103004 (2015).
20. D. H. Lyth, and A. Riotto, Particle physics models of inflation and the cosmological density perturbation, *Phys. Rep.* **314**, 1 (1999).
21. C. Armendariz-Picon, T. Damour, and V. F. Mukhanov, k-Inflation, *Phys. Lett. B* **458**, 209 (1999).
22. R. Geroch, A method for generating solutions of Einstein's equations, *J. Math. Phys.* **12**, 918 (1971).

23. R. Geroch, A method for generating new solutions of Einstein's equation. II., *J. Math. Phys.* **13**, 394 (1972).
24. D. Garfinkle, E. N. Glass, and J. P. Krisch, Solution generating with perfect fluids, *Gen. Rel. Grav.* **29**, 467 (1997).
25. H. Stephani, Symmetries of Einstein's field equations with a perfect fluid source as examples of Lie-Bäcklund symmetries, *J. Math. Phys.* **29**, 1650 (1988).
26. R. Hirota, Direct method of finding exact solutions of nonlinear evolution equations, in "Bäcklund transformations, the inverse scattering method, solitons, and their applications", *Lect. Not. Math. R.M. Miura eds.* **515** (1976).
27. V. A. Belinski, and V. E. Zakharov, Steady-state gravitational solitons with axial symmetry, *Sov. Phys. JETP* **50**, 1 (1979); *Zh. Eksp. Teor. Fiz.* **77**, 3 (1979).
28. J. Wainwright, W. C. W. Ince, and B. J. Marshman, Spatially homogeneous and inhomogeneous cosmologies with equation of state $p = \mu$, *Gen. Rel. Grav.* **10**, 259 (1979).
29. D. Gregoris, W. C. Lim, and A. Coley, Stiff fluid spike solutions from Bianchi type V seed solutions, *Class. Quantum Grav.* **34**, 235013 (2017).
30. M. Marklund, and M. Bradley, Invariant construction of solutions to Einstein's field equations - LRS perfect fluids II, *Class. Quantum Grav.* **16**, 1577 (1999).
31. A. Coley, D. Gregoris, and W. C. Lim, On the first G_1 stiff fluid spike solution in general relativity, *Class. Quantum Grav.* **33**, 215010 (2016).
32. D. Gregoris, On the connection between cosmological parameters and peculiar motion in a G_2 massless scalar field spacetime, *Int. J. Mod. Phys. D* 2150099 (2021).
33. D. Gregoris, Y. C. Ong, and B. Wang, Thermodynamics of shearing massless scalar field spacetimes is inconsistent with the Weyl curvature hypothesis, *Phys. Rev. D* **102** 023539 (2020).
34. C. Leibovitz, Time-dependent solutions of Einstein's equations, *Phys. Rev. D* **4**, 2949 (1971).
35. K. Lake, Remark concerning spherically symmetric nonstatic solutions to the Einstein equations in the comoving frame, *Gen. Rel. Grav.* **15**, 357 (1983).
36. N. van den Bergh and P. Wils, Exact solutions for nonstatic perfect fluid spheres with shear and an equation of state, *Gen. Rel. Grav.* **17**, 223 (1985).
37. B. Collins and J. M. Lang, A class of self-similar perfect fluid spacetimes, and a generalisation, *Class. Quantum Grav.* **4**, 61 (1987).
38. S. D. Maharaj, R. Maartens, and M. Maharaj, A note on a class of spherically symmetric solutions, *Nuovo Cimento B* **108**, 1 (1993).
39. H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt, *Exact solutions of Einstein's field equations* (Cambridge University Press, Cambridge, England, 2002).
40. J. Khoury, and A. Weltman, Chameleon fields: Awaiting surprises for tests of gravity in space, *Phys. Rev. Lett.* **93**, 171104 (2004).
41. J. Khoury, and A. Weltman, Chameleon cosmology, *Phys. Rev. D* **69**, 044026 (2004).
42. E. Newman, and R. Penrose, An approach to gravitational radiation by a method of spin coefficients, *J. Math. Phys* **3**, 566 (1962).
43. J. C. Baez, and E. F. Bunn, The meaning of Einstein's equation, *Amer. Jour. Phys.* **73**, 644 (2005).
44. T. Clifton, G. F. R. Ellis, and R. Tavakol, A gravitational entropy proposal, *Class. Quantum Grav.* **30**, 125009 (2013).
45. R. Penrose, Singularities and time-asymmetry, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
46. R. Bousso, The holographic principle, *Rev. Mod. Phys.* **74**, 825 (2002).

47. S. W. Hawking, Particle creation by black holes, *Comm. Math. Phys.* **43**, 199 (1975).
48. J. D. Bekenstein, Black holes and entropy, *Phys. Rev. D* **7**, 2333, (1973).
49. A. Ashtekar, and B. Krishnan, Isolated and dynamical horizons and their applications, *Living Rev. Relativity* **7**, 10 (2004).
50. B. Wang, Y. Gong, and E. Abdalla, Thermodynamics of an accelerated expanding universe, *Phys. Rev. D* **74**, 083520 (2006).
51. V. Faraoni, Cosmological apparent and trapping horizons, *Phys. Rev. D* **84**, 024003 (2011).
52. D. Bak, and S.-J. Rey, Cosmic holography, *Class. Quantum Grav.* **17**, L83 (2000).
53. G. F. R. Ellis, and H. van Elst, Cosmological models, *Cargése lectures 1998*, [arXiv:gr-qc/9812046].
54. R. A. Sussman, and J. Larena, Gravitational entropy of local cosmic voids, *Class. Quantum Grav.* **32**, 165012 (2015).
55. R. A. Sussman, and J. Larena, Gravitational entropies in LTB dust models, *Class. Quantum Grav.* **31**, 075021 (2014).
56. K. Bolejko, Gravitational entropy and the cosmological no-hair conjecture, *Phys. Rev. D* **97**, 083515 (2018).
57. G. Marozzi, J.-P. Uzan, O. Umeh, and C. Clarkson, Cosmological evolution of the gravitational entropy of the large-scale structure, *Gen. Rel. Grav.* **47**, 114 (2015).
58. P. Mishra, and T. P. Singh, Thermodynamics and Lemaitre-Tolman-Bondi void models, *Phys. Rev. D* **89**, 123007 (2014).
59. S. Chakraborty, S. Guha, and R. Goswami, An investigation on gravitational entropy of cosmological models, *Int. J. Mod. Phys. D* **30**, 2150051 (2021).
60. T. Buchert, et al., Is there proof that backreaction of inhomogeneities is irrelevant in cosmology?, *Class. Quantum Grav.* **32**, 215021 (2015).
61. T. Buchert, et al., Is there proof that backreaction of inhomogeneities is irrelevant in cosmology?, *Class. Quantum Grav.* **32**, 215021 (2015).
62. C. Clarkson, G. Ellis, J. Larena, and O. Umeh, Does the growth of structure affect our dynamical models of the universe? The averaging, backreaction, and fitting problems in cosmology, *Rep. Prog. Phys.* **74**, 112901 (2011).
63. A. G. Riess, et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
64. S. Perlmutter, et al. (The Supernova Cosmology Project), Measurements of Omega and Lambda from 42 high-redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
65. D. M. Scolnic, The complete light-curve sample of spectroscopically confirmed type Ia supernovae from Pan-STARRS1 and cosmological constraints from the combined pantheon sample, *Astrophys. J.* **859**, 101 (2018).
66. E. Komatsu, et al., Seven-year Wilkinson microwave anisotropy probe (WMAP) observations: cosmological interpretation, *Astrophys. J. Suppl.* **192**, 18 (2011).
67. R. Keisler, et al., A measurement of the damping tail of the cosmic microwave background power spectrum with the South Pole Telescope, *Astrophys. J.* **743**, 28 (2011).
68. D. J. Eisenstein, et al., Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, *Astrophys. J.* **633**, 560 (2005).
69. S. Cole, et al. (The 2dFGRS Team), The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications, *Mon. Not. R. Astron. Soc.* **362**, 505 (2005).
70. P. Bull, T. Clifton, and P. G. Ferreira, The kSZ effect as a test of general radial inhomogeneity in LTB cosmology, *Phys. Rev. D* **85**, 024002 (2012).
71. T. Clifton, P. G. Ferreira, and K. Land, Living in a void: Testing the Copernican principle with distant supernovae, *Phys. Rev. Lett.* **101**, 131302 (2008).

72. L. H. Dam, A. Heinesen, and D. L. Wiltshire, Apparent cosmic acceleration from type Ia supernovae, *Mon. Not. R. Astron. Soc.* **472**, 835 (2017).
73. A. Heinesen, C. Blake, Y.-Z. Li, and D. L. Wiltshire, Baryon acoustic oscillation methods for generic curvature: Application to the SDSS-III baryon oscillation spectroscopic survey, *J. Cosmol. Astropart. Phys.* **03**, 003 (2019).
74. J. Trost Nielsen, A. Guffanti, and S. Sarkar, Marginal evidence for cosmic acceleration from Type Ia supernovae, *Sci. Rep.* **6**, 35596 (2016).
75. D. Gregoris, Y. C. Ong, B. Wang, Holographic principle and the second law in Stephani cosmology revisited, *Eur. Phys. J. Plus* **135**, 246 (2020).
76. J. Wainwright, and P. J. Anderson, Isotropic singularities and isotropization in a class of Bianchi Type-VI$_h$ cosmologies, *Gen. Rel. Grav.* **16**, 609 (1984).
77. Ø. Grøn, and S. Hervik, Gravitational entropy and quantum cosmology, *Class. Quantum Grav.* **18**, 601 (2001).
78. A. Hosoya, T. Buchert, M. Morita, Information entropy in cosmology, *Phys. Rev. Lett.* **92**, 141302 (2004).
79. R. Sussman, Weighed scalar averaging in LTB dust models: Part I. Statistical fluctuations and gravitational entropy, *Class. Quantum Grav.* **30**, 065015 (2013).
80. D. Gregoris, and Y. C. Ong, Understanding Gravitational Entropy of Black Holes: A New Proposal via Curvature Invariants, arXiv:gr-qc/3945089.

Three-parameter solution for the null-surface formulation in 2+1 dimensions

Tina A. Harriott

Department of Mathematics and Statistics, Mount Saint Vincent University, Halifax, Nova Scotia, B3M 2J6, Canada
E-mail: tina.harriott@msvu.ca
www.msvu.ca

J. G. Williams

Department of Mathematics and Computer Science, and the Winnipeg Institute for Theoretical Physics, Brandon University, Brandon, Manitoba, R7A 6A9, Canada
E-mail: williams@brandonu.ca

The null-surface formuation (NSF) of general relativity is equivalent to standard general relativity but uses families of null surfaces rather than the metric or connection. The NSF can be constructed in dimension 3+1, in any dimension higher, and also in dimension 2+1, which is a special case: In 2+1 dimensions, the main NSF field equation is equivalent to Cartan's metricity condition. The latter arose in differential equation theory to address the problem of classifying solutions of 4 third-order ordinary differential equations. Solving the NSF/Cartan equation has proved challenging, and only three solutions are known to date. This talk presents a fourth solution, which depends upon three independent parameters. Two of the previously known solutions are included as special cases. Energy conditions and possible source terms are examined. The physical interpretation is discussed in detail.

Keywords: Null-surface formulation; NSF; metricity condition; low dimensional gravity; imperfect fluid.

1. Introduction

The NSF[1–3] uses families of null surfaces, rather than a metric, and can be presented in terms of a function Z that specifies the families. More commonly, the NSF is described in terms of a function Λ that is related to Z by differentiation. If required, a metric can be constructed up to a conformal factor from a knowledge of either Z or Λ and an auxiliary function Ω. Forni, Iriondo, Kozameh and Parisi,[4,5] Tanimoto,[6] and Silva-Ortigoza[7] have constructed a (2+1)-dimensional version of the NSF. A brief summary of the 2+1 NSF is given in the remainder of this section. The sections that follow will present a new solution and list its properties.

The equation $u = Z(x^a; \varphi)$ defines a surface, with x^a $(a = 0, 1, 2)$ being spacetime coordinates and $\varphi \in S^1$ being an angular variable. The role of φ is to label the surfaces. For the surfaces to be *null* with respect to a spacetime metric $g_{bc}(x^a)$,

the gradient of Z must satisfy

$$g^{bc}(x^a)Z_{,b}(x^a;\varphi)\,Z_{,c}(x^a;\varphi) \;=\; 0, \tag{1}$$

where $Z_{,a} \equiv \partial_a Z \equiv \partial Z/\partial x^a$. Equation (1) and its derivatives with respect to φ lead to the *metricity conditions*, which ensure that the nullness requirement is satisfied. The metricity conditions are most easily expressed in terms of coordinates that are naturally adapted to the surfaces.[4–6] Such coordinates are called *intrinsic* coordinates,[2] and are denoted

$$\begin{aligned} u &:= Z(x^a;\varphi), \\ \omega &:= \partial u \;\equiv\; \partial Z(x^a;\varphi), \\ \rho &:= \partial\omega \;=\; \partial^2 u \;\equiv\; \partial^2 Z(x^a;\varphi), \end{aligned}$$

where $\partial := \partial/\partial\varphi$ is the derivative with respect to φ when the x^a are held fixed. In principle, the above equations can be inverted to give

$$x^a \;=\; x^a(u,\omega,\rho,\varphi).$$

The third derivative of Z (i.e. of u) is denoted by Λ:

$$\Lambda(u,\omega,\rho,\varphi) := \partial^3 Z(x^a(u,\omega,\rho,\varphi);\varphi).$$

If Λ is known, then Z can be determined by integration. The intrinsic coordinates u, ω and ρ are φ-dependent and it can be shown that the action of the differential operator ∂ on a function $f(u,\omega,\rho,\varphi)$ is given by[4,6]

$$\partial \;=\; \partial' + \omega\,\partial_u + \rho\,\partial_\omega + \Lambda\,\partial_\rho,$$

where ∂' denotes the derivative with respect to φ when u, ω and ρ are held fixed.

Equation (1) leads to the following form for the inverse metric g^{ij},[4,6] expressed in terms of the u,ω,φ coordinates, with the signature of g^{ab}, and hence of g^{ij}, being chosen to be $(-++)$:

$$[\,g^{ij}\,] = \Omega^2\,[\,\gamma^{ij}\,] = \Omega^2 \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & \frac{1}{3}\partial_\rho\Lambda \\ -1 & \frac{1}{3}\partial_\rho\Lambda & \frac{1}{3}\partial(\partial_\rho\Lambda) - \frac{1}{9}(\partial_\rho\Lambda)^2 - \partial_\omega\Lambda \end{pmatrix}.$$

It follows that the components of the metric g_{ij} are

$$[\,g_{ij}\,] = \Omega^{-2}\,[\,\gamma_{ij}\,] = \Omega^{-2} \begin{pmatrix} -\frac{1}{3}\partial(\partial_\rho\Lambda) + \frac{2}{9}(\partial_\rho\Lambda)^2 + \partial_\omega\Lambda & \frac{1}{3}\partial_\rho\Lambda & -1 \\ \frac{1}{3}\partial_\rho\Lambda & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \tag{2}$$

where γ_{ij} is called the *unphysical* metric. Equations (1) and (2) can be used to derive the following metricity conditions.[4,6]

$$2[\partial(\partial_\rho\Lambda) - \partial_\omega\Lambda - \tfrac{2}{9}(\partial_\rho\Lambda)^2]\partial_\rho\Lambda - \partial^2(\partial_\rho\Lambda) + 3\,\partial(\partial_\omega\Lambda) - 6\,\partial_u\Lambda = 0, \tag{3}$$

$$3\,\partial\Omega \;=\; \Omega\,\partial_\rho\Lambda. \tag{4}$$

Equations (3) and (4) ensure that Λ will determine a null surface with respect to *some* spacetime metric $g_{bc}(x^a)$. Equation (3) is the main metricity condition. Equation (4) is the secondary metricity condition. If the two metricity conditions are satisfied, then the Einstein equations, $G_{ij} = \kappa T_{ij}$, will be satisfied if the following equation holds:[4]

$$\partial_\rho^2 \Omega \;=\; \kappa T_{\rho\rho} \Omega. \tag{5}$$

Finding a solution in the NSF approach requires solving the coupled set of NSF field equations, Eqs (3), (4) and (5), for Λ and Ω. In 2+1 dimensions, only three nontrivial solutions of the NSF equations have been found to date.[8–10] The present paper introduces a new family of solutions.

2. Solution

The form of the differential operator ∂ suggests looking for solutions that are additively separable. Consider solutions that take the form

$$\Lambda(u, \omega, \rho) = -a\omega + h(\rho + au) = -a\omega + h(x), \tag{6}$$

where a is a constant. Writing $x := \rho + au$, will convert the main metricity condition, Eq. (3), into an ordinary differential equation with x as the independent variable:

$$h^2 \frac{d^3h}{dx^3} - h \frac{dh}{dx} \frac{d^2h}{dx^2} + \frac{4}{9} \left(\frac{dh}{dx} \right)^3 + 4a \frac{dh}{dx} = 0. \tag{7}$$

The solution of Eq. (7), together with Eq. (6), represents a solution of the main metricity condition. After introducing a new dependent variable $y := h^{2/3}$, and replacing h by $y^{3/2}$, Eq. (6) becomes

$$\Lambda(u, \omega, \rho) \;=\; -a\omega + [y(\rho + au)]^{3/2} \;\equiv\; -a\omega + [y(x)]^{3/2}, \tag{8}$$

and Eq. (7) becomes

$$\frac{d^3y}{dx^3} + 4a\, y^{-3} \frac{dy}{dx} \;=\; 0,$$

which can be rewritten

$$\frac{d}{dx} \left(\frac{d^2y}{dx^2} - 2ay^{-2} \right) \;=\; 0,$$

and leads to

$$\frac{d^2y}{dx^2} - 2ay^{-2} \;=\; k,$$

where k is a constant. A further integration can be achieved by multiplying by dy/dx to give

$$\frac{dy}{dx} \frac{d^2y}{dx^2} - 2ay^{-2} \frac{dy}{dx} - k \frac{dy}{dx} \;=\; 0,$$

which leads to

$$\left(\frac{dy}{dx}\right)^2 + 4ay^{-1} - 2ky \;=\; -A,$$

where A is a constant. The minus sign in front of A has been inserted for later convenience. It follows that

$$\frac{dy}{dx} \;=\; \pm\left(2ky - A - 4ay^{-1}\right)^{1/2},$$

and so

$$x = \pm \int \frac{dy}{\sqrt{2ky - A - 4ay^{-1}}},$$

$$\pm \int \frac{y\,dy}{\sqrt{y\,(2ky^2 - Ay - 4a)}}. \tag{9}$$

Equation (9) expresses y as an implicit function of x and, together with Eq. (8), represents an exact solution of the main metricity condition, Eq. (3). Although the right side of Eq. (9) can be evaluated in terms of elliptic integrals of the first and second kinds, the result is too complicated for the equation to be inverted to give y as an explicit function of x. Simplifications arise when some of the constants, A, a, and k, are zero. For example, if $A \neq 0, k \neq 0$, and $a = 0$, the solution becomes the perfect fluid, Petrov-type D, solution that the present authors reported in an earlier paper.[8] The special case where $k \neq 0, a \neq 0$, and $A = 0$ was also reported in an earlier paper, where the spacetime was determined to be of Petrov-type I and the source was represented by a minimally coupled scalar field.[9]

Now consider the secondary metricity condition, Eq. (4),

$$3\,\partial\Omega \;=\; \Omega\,\partial_\rho\Lambda.$$

As with Λ, assume that Ω depends upon u and ρ through the combination $\rho + au$, but assume that Ω is independent of ω. Thus $\Omega = \Omega(\rho + au)$. Using $\Lambda = -a\omega + h = -a\omega + y^{3/2}$, it follows that

$$\Omega = h^{1/3} = y^{1/2}.$$

The third NSF equation, Eq. (5),

$$\partial_\rho^2\Omega \;=\; \kappa T_{\rho\rho}\,\Omega,$$

then implies

$$T_{\rho\rho} = \frac{1}{2\kappa\,y^2}\left\{y\,\frac{d^2y}{dx^2} - \frac{1}{2}\left(\frac{dy}{dx}\right)^2\right\}.$$

It follows that

$$T_{\rho\rho} = \frac{1}{4\kappa\,y^2}\left(A + 8ay^{-1}\right) = \frac{1}{4\kappa\,y^2}\,W,$$

where W is defined by

$$W := A + 8ay^{-1}. \tag{10}$$

For any null vector V, the null energy condition requires $R_{ij}\,V^iV^j \geq 0$, which is equivalent to $G_{ij}\,V^iV^j \geq 0$. The vector $V^\rho = \partial_\rho \equiv (0,0,1)$ is null and so $T_{\rho\rho} \geq 0$, or equivalently $W \geq 0$, is a *necessary* condition for the null energy condition to hold. Henceforth, it will be assumed that $W \geq 0$.

3. Metric and Curvature

The form of the solution given in Section 2 implies

$$\partial_\rho^2 y = 2a\,y^{-2} + k, \tag{11}$$

and

$$(\partial_\rho y)^2 = 2k\,y - A - 4a\,y^{-1}. \tag{12}$$

Equations (11) and (12), together with the fact that $\Omega = y^{1/2}$, lead to the following expression for the metric g_{ij}:

$$[\,g_{ij}\,] = \Omega^{-2}\,[\,\gamma_{ij}\,] = \begin{pmatrix} -\left(\frac{1}{4}A + 3ay^{-1}\right) & \frac{1}{2}y^{-1/2}\,\partial_\rho y & -y^{-1} \\ \frac{1}{2}y^{-1/2}\,\partial_\rho y & y^{-1} & 0 \\ -y^{-1} & 0 & 0 \end{pmatrix}. \tag{13}$$

Hence $g := \det g_{ij} = -y^{-3}$. The inverse metric is

$$[\,g^{ij}\,] = \Omega^2\,[\,\gamma^{ij}\,] = \begin{pmatrix} 0 & 0 & -y \\ 0 & y & \frac{1}{2}y^{3/2}\,\partial_\rho y \\ -y & \frac{1}{2}y^{3/2}\,\partial_\rho y & \frac{1}{2}ky^3 + 2ay \end{pmatrix}.$$

Equations (11), (12), and (13) can be used to find the Christoffel symbols and the Ricci tensor R_{ij}, which are listed in the appendix. The scalar curvature is constant:

$$R = \frac{1}{32}\,A^2 + ka. \tag{14}$$

If the constants A. a, and k, are all assumed to be positive (which is consistent with $W > 0$), then it is straightforward to show that the matrix $[\,R_{ij}\,]$ is positive definite. This implies that the null, weak, and strong energy conditions are all satisfied.

Being a theory of null surfaces,the NSF does not distinguish between conformally related spacetimes. Thus NSF solutions that correspond to conformally flat spacetimes are considered to be trivial. In 2+1 dimensions, the Cotton-York tensor is defined by[11]

$$C^i_{\ j} \;:=\; \varepsilon^{ikp}\left(R_{pj} - \frac{1}{4}R\,g_{pj}\right)_{;\,k}$$

and provides a test for conformal flatness: A (2+1)-dimensional spacetime is conformally flat if and only if $C^i_j = 0$. For the solution considered in the present paper, $C^i_j \neq 0$ (except for the trivial case where the constants A, k, and a are all zero). Thus the spacetime is not conformally flat. The components C^i_j are listed in the appendix.

4. Fluid Source: Kinematics

Begin by assuming that the source is a fluid. Equation (10) and the discussion around the null-energy condition gave $W := A + 8ay^{-1} \geq 0$. Make the further assumption that $W + 4ay^{-1} > 0$ so that

$$\frac{1}{4}\,A + 3ay^{-1} > 0.$$

It follows from Eq. (13) that $g_{uu} < 0$. The form of the Ricci tensor, R_{ij}, given in the appendix then suggests the following convenient choice for the fluid velocity vector U,

$$(U^u,\, U^\omega,\, U^\rho) := a^{-1/2} y^{1/2}\,(1,\, 0,\, -\tfrac{1}{8} W y),$$

$$(U_u,\, U_\omega,\, U_\rho) \;=\; a^{-1/2} y^{-1/2}\,(-\tfrac{1}{8} W y - a,\, \tfrac{1}{2} y^{1/2}\,\partial_\rho y,\, -1),$$

which, using the R_{ij} given in the appendix, leads to

$$R_{ij} U^i U^j = \frac{1}{8} A R a^{-1} y. \tag{15}$$

Note that $U^i U_i = -1$, as required. The covariant derivative $U_{i\,;\,j}$ can be written as a sum[12] (p. 85)

$$U_{i\,;\,j} \;=\; \omega_{ij} + \sigma_{ij} + \frac{1}{2}\,\theta\, h_{ij} - \dot{U}_i\, U_j,$$

where the projection tensor h_{ij}, the acceleration vector $\dot{U}_i$, the vorticity tensor ω_{ij}, the expansion θ, and the shear tensor σ_{ij} are defined by

$$h_{ij} := g_{ij} + U_i U_j,$$

$$\dot{U}_i := U_{i\,;\,j}\, U^j,$$

$$\omega_{ij} := \frac{1}{2}\,(U_{i\,;\,k}\, h^k_j - U_{j\,;\,k}\, h^k_i),$$

$$\theta := U^i_{\,;\,i},$$

$$\sigma_{ij} := \frac{1}{2}\,(U_{i\,;\,k}\, h^k_j + U_{j\,;\,k}\, h^k_i) - \frac{1}{2}\,\theta\, h_{ij}.$$

The coefficient of θ, both in the expression for $U_{i\,;\,j}$ and in the expression for σ_{ij}, is $\frac{1}{2}$ (instead of $\frac{1}{3}$, as would be the case in 3+1 dimensions). This follows because the trace $h^i_{\ i}$ of the projection tensor is 2 (instead of 3). The components of $U_{i\,;\,j}$ and other kinematic quantities, including the acceleration vector $\dot{U}^i$, are listed in the appendix. The divergence of the acceleration vector, $\dot{U}^i_{\ ;\,i}$, is

$$\dot{U}^i_{\ ;\,i} = \frac{1}{8}(A - 4ky)Ra^{-1}y. \tag{16}$$

The expansion scalar, θ, measures the isotropic volume expansion of the fluid and is found to be zero:

$$\theta = 0. \tag{17}$$

The scalar vorticity ω and the scalar shear σ are given by

$$\omega^2 := \frac{1}{2}\omega^{ij}\,\omega_{ij} \ = \ \frac{1}{4}R^2a^{-2}y^2, \tag{18}$$

$$\sigma^2 := \frac{1}{2}\sigma^{ij}\,\sigma_{ij} \ = \ \frac{1}{128}A^2Ra^{-2}y^2. \tag{19}$$

Raychaudhuri's equation can be written[13] (p. 5295)

$$\dot{\theta} = -\frac{1}{2}\theta^2 - R_{ij}U^iU^j - 2\left(\sigma^2 - \omega^2\right) + \dot{U}^i_{\ ;\,i}.$$

Using Eqs (15) to (19), it is straightforward to verify that Raychaudhuri's equation holds.

5. Fluid Source: Dynamics

In analysing the Einstein field equations, $G_{ij} = R_{ij} - \frac{1}{2}R\,g_{ij}$, the velocity U^i and the projection tensor h_{ij} can be used to write the stress energy tensor as a sum of separate terms[12] (p. 91)

$$T_{ij} \ = \ \mu\,U_iU_j + p\,h_{ij} + q_iU_j + U_iq_j + \pi_{ij},$$

where $q_iU^i = 0$, $\pi_{ij}U^j = 0$, and $\pi^i_{\ i} = 0$. The function μ denotes the total mass-energy density, p the isotropic pressure (which may include a contribution from the bulk viscosity), q^i the heat-flux vector, and π_{ij} the trace-free anisotropic pressure due to processes such as dynamic viscosity. It follows that

$$T^i_{\ i} = 2p - \mu,$$

and

$$\kappa\,(2p - \mu) = -\frac{1}{2}\,R.$$

Hence the Einstein field equations can be rewritten

$$R_{ij} = \kappa\,[\,T_{ij} - (2p - \mu)g_{ij}\,]$$

and then decomposed using U^i and $h^j{}_k$ to produce the following equivalent set of equations:

$$R_{ij}\,U^iU^j = 2\kappa\,p,$$

$$R_{ij}\,U^ih^j{}_m = -\kappa\,q_m,$$

$$R_{ij}\,h^i{}_m\,h^j{}_n = \kappa\,(\mu - p)\,h_{mn} + \kappa\,\pi_{mn}.$$

It follows that the pressure p is given by

$$p = \frac{1}{16\kappa}a^{-1}yAR,$$

and that the mass-energy density μ is given by

$$\mu = \frac{1}{8\kappa}a^{-1}yAR + \frac{1}{2\kappa}\,R.$$

The components q_i of the heat-flux vector are given by

$$q_u = \frac{1}{64\kappa}a^{-3/2}y^{3/2}AWR,$$

$$q_\omega = -\frac{1}{16\kappa}a^{-3/2}yAR\,\partial_\rho y,$$

$$q_\rho = \frac{1}{8\kappa}a^{-3/2}y^{1/2}AR.$$

Heat flow is related to temperature and acceleration by the temperature gradient law[12] (p. 96),

$$q^i \;=\; -\,\lambda\,h^{ij}\left(T_{,\,j} + T\,\dot{U}_j\right),$$

where λ denotes the coefficient of thermal conductivity. It can be checked that this law is satisfied identically. However, the usual phenomenological equation of state relating anisotropic pressure to the shear is *not* satisfied: $\pi_{ij} \neq -2\eta\sigma_{ij}$ (η being the coefficient of shear viscosity). Situations where this equation of state may not hold have been considered by MacCallum, Stewart, and Schmidt.[14]

6. Summary and Conclusions

A family of (2+1)-dimensional null-surface formulation solutions has been introduced and involves three parameters, A, a, and k. Since the solutions that make up the family all satisfy the NSF main metricity condition, they automatically satisfy Cartan's metricity condition.[15–17] The family of solutions led to a metric and to the Einstein equations being satisfied. An imperfect fluid was presented as a possible source. The mass-energy density and the pressure were shown to be positive. The scalar expansion was found to be zero. The scalar vorticity and scalar shear were

found to be nonzero, and simple formulas were derived. The temperature gradient law was satisfied. It was pointed out that the special choices $\{A \neq 0, k \neq 0, a = 0\}$ and $\{k \neq 0, a \neq 0, A = 0\}$ correspond to two solutions that were reported earlier by the present authors.[8,9]

Acknowledgements

The authors would like to thank Dr. Alan Coley for a number of helpful comments. We would also like to acknowledge the generous hospitality of Dr. Ted Newman during our visits to the University of Pittsburgh.

Appendix

The Christoffel symbols are

$$\Gamma^{u}_{uu} = -a\Gamma^{u}_{\omega\omega} = -a\Gamma^{\omega}_{\omega\rho} = 2^{-1}ay^{-1}\,\partial_\rho y,$$

$$\Gamma^{u}_{u\omega} = 8^{-1}Ay^{-1/2} + ay^{-3/2},$$

$$\Gamma^{u}_{u\rho} = \Gamma^{u}_{\omega\rho} = \Gamma^{u}_{\rho\rho} = \Gamma^{\omega}_{\rho\rho} = 0,$$

$$\Gamma^{\omega}_{uu} = 2^{-1}Aay^{-1/2} + 3a^2y^{-3/2} - 2^{-1}kay^{1/2},$$

$$\Gamma^{\omega}_{u\omega} = -(16^{-1}A + ay^{-1})\,\partial_\rho y,$$

$$\Gamma^{\omega}_{u\rho} = 8^{-1}Ay^{-1/2} + ay^{-3/2},$$

$$\Gamma^{\omega}_{\omega\omega} = -4^{-1}Ay^{-1/2} - ay^{-3/2} + 2^{-1}ky^{1/2},$$

$$\Gamma^{\rho}_{uu} = -2^{-1}a\,(-4^{-1}A + 3ay^{-1} + 2^{-1}ky)\,\partial_\rho y,$$

$$\Gamma^{\rho}_{u\omega} = -(16^{-1}Aky^{3/2} + a^2y^{-3/2} + kay^{1/2}),$$

$$\Gamma^{\rho}_{u\rho} = (16^{-1}A - ay^{-1})\,\partial_\rho y,$$

$$\Gamma^{\rho}_{\omega\omega} = (2^{-1}\,ay^{-1} + 4^{-1}\,ky)\,\partial_\rho y,$$

$$\Gamma^{\rho}_{\omega\rho} = 8^{-1}Ay^{-1/2} - 2^{-1}ky^{1/2},$$

$$\Gamma^{\rho}_{\rho\rho} = -y^{-1}\,\partial_\rho y.$$

Using the abbreviation $W := A + 8a\,y^{-1}$, the Ricci tensor can be written as a matrix, $[R_{ij}]$:

$$[\,R_{ij}\,] = \begin{pmatrix} \frac{1}{256}\,(A^3+W^3) + \frac{1}{8}\,Aka & -\frac{1}{64}\,AW\,y^{-1/2}\,\partial_\rho y & \frac{1}{32}\,W^2\,y^{-1} \\ -\frac{1}{64}\,AW\,y^{-1/2}\,\partial_\rho y & -\frac{1}{32}\,A^2\,y^{-1} + \frac{1}{8}\,Ak & -\frac{1}{8}\,A\,y^{-3/2}\,\partial_\rho y \\ \frac{1}{32}\,W^2\,y^{-1} & -\frac{1}{8}\,A\,y^{-3/2}\,\partial_\rho y & \frac{1}{4}\,W\,y^{-2} \end{pmatrix}.$$

The scalar curvature, R, is given in Eq. (14).

The covariant derivatives of the velocity can be written as a matrix, $[\,U_{i\,;\,j}\,]$:

$$\begin{pmatrix} -\frac{1}{8}a^{1/2}y^{-1/2}\,W\,\partial_\rho y & -\frac{1}{64}a^{-1/2}\,W^2 & -\frac{1}{16}a^{-1/2}y^{-1/2}\,W\,\partial_\rho y \\ a^{-1/2}\left(\frac{1}{2}ka - \frac{1}{64}W^2 + a^2y^{-2}\right) & \frac{1}{16}a^{-1/2}y^{-1/2}\,A\,\partial_\rho y & \frac{1}{2}a^{-1/2}y^{-1}\left(ky - \frac{1}{4}A\right) \\ -a^{1/2}y^{-3/2}\,\partial_\rho y & -\frac{1}{8}a^{-1/2}y^{-1}\,W & -\frac{1}{2}a^{-1/2}y^{-3/2}\,\partial_\rho y \end{pmatrix}$$

The acceleration vector, $\dot{U}_i$, is defined by $\dot{U}_i := U_{i\,;\,j}\,U^j$, and its components are

$$\dot{U}^u = \left(1 - \frac{1}{16}Wa^{-1}y\right)\,\partial_\rho y,$$

$$\dot{U}^\omega = \frac{1}{4}Wy^{1/2}\left(1 - \frac{1}{16}Wa^{-1}y\right) - \frac{1}{2}Ra^{-1}y^{3/2},$$

$$= a^{-1}y^{3/2}\left(-R + \frac{1}{2}ka + a^2y^{-2}\right),$$

$$\dot{U}^\rho = -a\left(1 - \frac{1}{16}Wa^{-1}y\right)\,\partial_\rho y - \frac{1}{4}Ra^{-1}y^2\,\partial_\rho y.$$

The only nonzero components of the vorticity tensor, ω_{ij}, are

$$\omega_{u\omega} = -\omega_{\omega u} = -\frac{1}{16}WRa^{-3/2}y,$$

$$\omega_{\omega\rho} = -\omega_{\rho\omega} = \frac{1}{2}Ra^{-3/2}.$$

The components of the shear tensor, σ_{ij}, are

$$\sigma_{uu} = -\frac{1}{1024}AW^2a^{-3/2}y^{3/2}\,\partial_\rho y,$$

$$\sigma_{u\omega} = \sigma_{\omega u} = \frac{1}{128}W^2ka^{-3/2}y^2 - \frac{1}{16}WRa^{-3/2}y,$$

$$\sigma_{u\rho} = \sigma_{u\rho} = -\frac{1}{128} AWa^{-3/2} y^{1/2}\, \partial_\rho y,$$

$$\sigma_{\omega\omega} = -\frac{1}{32} Aka^{-3/2} y^{3/2}\, \partial_\rho y,$$

$$\sigma_{\omega\rho} = \sigma_{\rho\omega} = \frac{1}{16} Wka^{-3/2} y - \frac{1}{2} Ra^{-3/2},$$

$$\sigma_{\rho\rho} = -\frac{1}{16} Aa^{-3/2} y^{-1/2}\, \partial_\rho y.$$

The components of the Cotton-York tensor, $C^i_{\ j}$, are

$$C^u_{\ u} = -\frac{1}{4} W \left(R + \frac{1}{32} AW \right),$$

$$= -\frac{1}{4} A \left(R + \frac{1}{32} W^2 \right) - 2Ray^{-1},$$

$$C^u_{\ \omega} = y^{-1/2} \left(\frac{1}{2} R + \frac{1}{32} A^2 \right) \partial_\rho y,$$

$$C^u_{\ \rho} = -Ry^{-1} - \frac{1}{16} AWy^{-1},$$

$$C^\omega_{\ u} = ay^{-1/2} \left(\frac{1}{2} R + \frac{1}{32} AW \right) \partial_\rho y,$$

$$C^\omega_{\ \omega} = \frac{1}{128} A^2 W + Ray^{-1},$$

$$C^\omega_{\ \rho} = \frac{1}{4} Aay^{-3/2}\, \partial_\rho y,$$

$$C^\rho_{\ u} = \frac{1}{128} AW^2 a - 2Ra^2 y^{-1} + \frac{1}{2} RWa,$$

$$C^\rho_{\ \omega} = -ay^{-1/2} \left(\frac{1}{2} R + \frac{1}{32} A^2 \right) \partial_\rho y,$$

$$C^\rho_{\ \rho} = \frac{1}{2} Aa^2 y^{-2} - 2ka^2 y^{-1} + \frac{1}{4} AR + 3Ray^{-1},$$

$$= \frac{1}{16} AWay^{-1} + \frac{1}{4} AR + Ray^{-1}.$$

References

1. S. Frittelli, C. N. Kozameh and E. T. Newman, Lorentzian metrics from characteristic surfaces, *J. Math. Phys.* **36**, 4975 (1995).
2. S. Frittelli, C. N. Kozameh and E. T. Newman, GR via characteristic surfaces, *J. Math. Phys.* **36**, 4984 (1995).
3. S. Frittelli, C. N. Kozameh and E. T. Newman, Linearized Einstein theory via null surfaces, *J. Math. Phys.* **36**, 5005 (1995).
4. D. M. Forni, N. Iriondo and C. N. Kozameh, Null surfaces formulation in three dimensions, *J. Math. Phys.* **41**, 5517 (2000).
5. D. M. Forni, N. Iriondo, C. N. Kozameh and M. F. Parisi, Understanding singularities in Cartan's and null surface formulation geometric structures, *J. Math. Phys.* **43**, 1584 (2002).
6. M. Tanimoto, On the Null Surface Formalism, gr-qc/9703003, (1997).
7. G. Silva-Ortigoza, Null surfaces and their singularities in three-dimensional Minkowski space-time, *Gen. Rel. Grav.* **32**, 2243 (2000).
8. T. A. Harriott and J. G. Williams, Solution for the null-surface formulation of general relativity in 2+1 dimensions, *Gen. Rel. Grav.* **46**, 1666 (2018).
9. T. A. Harriott and J. G. Williams, Three-variable solution in the (2 + 1)-dimensional null-surface formulation, *Gen. Rel. Grav.* **50**, 39 (2018).
10. T. A. Harriott and J. G. Williams, Petrov type-N solution for the null-surface formulation in 2 + 1 dimensions, *Gen. Rel. Grav.* **51**, 98 (2019).
11. A. A. Garcia-Diaz, *Exact Solutions in Three-Dimensional Gravity* (Cambridge University Press, 2017).
12. G. F. R. Ellis, R. Maartens and M. A. H. MacCallum, *Relativistic Cosmology* (Cambridge University Press, 2021).
13. J. D. Barrow, D. J. Shaw and C. G. Tsagas, Cosmology in three dimensions: Steps towards the general solution, *Class. Q. Grav.* **23**, 5291 (2006).
14. M. A. H. MacCallum, J. M. Stewart and B. Schmidt, Anisotropic stresses in homogeneous cosmologies, *Commun. Math. Phys.* **17**, 343 (1970).
15. E. Cartan, Les espaces generalises et l'integration de certaines classes d'equations differentielles, *C. R. Acad. Sci* **206**, 1425 (1938).
16. E. Cartan, La geometria de las ecuaciones diferenciales de tercer orden, *Rev. Mat. Hispano-Amer.* **4**, 1 (1941).
17. E. Cartan, Sur une classe d'espaces de Weyl, *Ann. Sc. Ec. Norm. Sup. (3)* **60**, 1 (1943).

New exact stationary cylindrical anisotropic fluid solution of GR

M.-N. Célérier

Laboratoire Univers et THéories, Observatoire de Paris,
Université PSL, CNRS, Université de Paris,
5, Place Jules Janssen, Meudon, 92190, France
E-mail: marie-noelle.celerier@obspm.fr

In a previous paper, the properties of interior spacetimes sourced by stationary cylindrical anisotropic fluids have been analytically studied for both nonrigid and rigid rotation. The gravito-electromagnetic features of different classes of such GR solutions have been described. Their regularity conditions and those for their junction to a vacuum exterior have also been provided. A new class of rigidly rotating exact solutions to Einstein's field equations satisfying a physically consistent equation of state for anisotropic fluids is displayed here. Its physical properties are discussed.

Keywords: General Relativity; exact solution; cylindrical symmetry; anisotropic fluids

1. Introduction

Cylindrically symmetric spacetimes have attracted much attention and have been extensively investigated in General Relativity (GR) for a number of purposes.[1,2] A recent review has been displayed in.[3]

The results presented here have been obtained as a continuation of a couple of previous works. The rigid rotation of stationary spacetimes sourced by a cylindrically symmetric anisotropic fluid has been considered in.[4] This study has been subsequently extended to nonrigidly rotating fluids,[5] while the analysis of the rigid case proposed in[4] has been supplemented, with a focuss on its Weyl tensor gravito-electromagnetic properties. Both were matched to a stationary vacuum exterior of the Lewis type.[6]

The new results displayed here consist of a class of exact interior solutions to the Einstein field equations for a rigidly rotating cylindrically symmetric anisotropic fluid bounded by a cylindrical hypersurface through which it is matched to an exterior Lewis vacuum of the Weyl class.[7] The particular case of an axially directed principal stress is considered here as a first step towards the analysis of anisotropy in this framework.[8] The main physical properties of these spacetimes are also discussed.

This contribution is organized as follows: in Section 2, the stress-energy tensor and the stationary cylindrically symmetric line element to be used in the following are set up. In Section 3, the new exact solution is constructed, from the field equations first displayed, through its regularity and junction conditions, up to its final form. A number of mathematical and physical properties pertaining to this solution are analyzed in Section 4. The conclusions are displayed in Section 5.

2. Cylindrical spacetime inside the source

A rigidly rotating stationary cylindrically symmetric anisotropic nondissipative fluid bounded by a cylindrical surface Σ whose principal stresses P_r, P_z and P_ϕ obey the equation of state $P_r = P_\phi = 0$ is considered. Its stress-energy tensor can be written as

$$T_{\alpha\beta} = \rho V_\alpha V_\beta + P_z S_\alpha S_\beta, \tag{1}$$

where ρ is the energy density of the fluid, V_α is its timelike 4-velocity, and S_α is a spacelike 4-vector, satisfying

$$V^\alpha V_\alpha = -1, \quad S^\alpha S_\alpha = 1, \quad V^\alpha S_\alpha = 0. \tag{2}$$

It is assumed, for the inside Σ spacetime, that the spacelike ∂_z Killing vector is hypersurface orthogonal, such as to ease its subsequent matching to the exterior Lewis metric. Hence, its stationary cylindrically symmetric line element reads

$$\mathrm{d}s^2 = -f\mathrm{d}t^2 + 2k\mathrm{d}t\mathrm{d}\phi + \mathrm{e}^\mu(\mathrm{d}r^2 + \mathrm{d}z^2) + l\mathrm{d}\phi^2, \tag{3}$$

where f, k, μ and l are real functions of the radial coordinate r only. Owing to cylindrical symmetry, the coordinates satisfy the following ranges

$$-\infty \leq t \leq +\infty, \quad 0 \leq r, \quad -\infty \leq z \leq +\infty, \quad 0 \leq \phi \leq 2\pi. \tag{4}$$

Now, still due to cylindrical symmetry, the two limits of the ϕ coordinate are topologically identified. These coordinates are numbered $x^0 = t$, $x^1 = r$, $x^2 = z$ and $x^3 = \phi$.

3. Rigidly rotating new solution

3.1. *Equation of state and 4-velocity of the fluid*

Choosing a corotating frame for the stationary fluid source,[4,5,8] the fluid 4-velocity can be written as

$$V^\alpha = v\delta^\alpha_0, \tag{5}$$

where v is a function of r only. The timelike condition for V^α, provided in (2), thus becomes

$$fv^2 = 1. \tag{6}$$

The spacelike 4-vector verifying conditions (2) can be chosen as

$$S^\alpha = \mathrm{e}^{-\mu/2}\delta^\alpha_2, \tag{7}$$

and a function $D(r)$, which will proved useful for the calculations, is defined as

$$D^2 = fl + k^2. \tag{8}$$

3.2. *Field equations*

With the above choices, and using (5)-(8) into (1), one obtains the stress-energy tensor components corresponding to the five non-vanishing Einstein tensor components and the five field equations for the inside Σ spacetime follow.

A new function $h(r)$ defined as $h(r) \equiv P_z(r)/\rho(r)$, allows to write them as

$$G_{00} = \frac{\mathrm{e}^{-\mu}}{2}\left[-f\mu'' - 2f\frac{D''}{D} + f'' - f'\frac{D'}{D} + \frac{3f(f'l' + k'^2)}{2D^2}\right] = \kappa\rho f, \tag{9}$$

$$G_{03} = \frac{\mathrm{e}^{-\mu}}{2}\left[k\mu'' + 2k\frac{D''}{D} - k'' + k'\frac{D'}{D} - \frac{3k(f'l' + k'^2)}{2D^2}\right] = -\kappa\rho k, \tag{10}$$

$$G_{11} = \frac{\mu' D'}{2D} + \frac{f'l' + k'^2}{4D^2} = 0, \tag{11}$$

$$G_{22} = \frac{D''}{D} - \frac{\mu' D'}{2D} - \frac{f'l' + k'^2}{4D^2} = \kappa\rho h\mathrm{e}^{\mu}, \tag{12}$$

$$G_{33} = \frac{\mathrm{e}^{-\mu}}{2}\left[l\mu'' + 2l\frac{D''}{D} - l'' + l'\frac{D'}{D} - \frac{3l(f'l' + k'^2)}{2D^2}\right] = \kappa\rho\frac{k^2}{f}, \tag{13}$$

where the primes stand for differentiation with respect to r.

3.3. *Bianchi identity*

The stress-energy tensor conservation is analogous to the Bianchi identity that reads

$$T^{\beta}_{1;\beta} = 0. \tag{14}$$

From (1), we have

$$T^{\alpha\beta} = \rho V^{\alpha} V^{\beta} + P_z S^{\alpha} S^{\beta}, \tag{15}$$

with V^{α} given by (5), and the space-like vector S^{α} given by (7), inserted into (15). Using (3) and (6), the Bianchi identity (14) reduces to

$$T^{\beta}_{1;\beta} = \frac{1}{2}\rho\frac{f'}{f} - \frac{1}{2}P_z\mu' = 0, \tag{16}$$

which, when the $h(r)$ function defined above is inserted, becomes

$$\frac{f'}{f} - h\mu' = 0. \tag{17}$$

3.4. *Solving the field equations*

Since only five independent differential equations are available for six unknowns, i.e., the four metric functions f, k, e^{μ} and l, the energy density ρ, and the pressure defined by h, the set of equations needs to be closed by an additional one. The particular assumption displayed in[5] as Eq. (111), and amounting to make a particular choice of coordinates issued from an arbitrary constraint on the Weyl tensor components, is chosen for this purpose as

$$\mu' = \frac{2h'}{1-h} + \frac{2h'}{h}, \tag{18}$$

which can be integrated as

$$\mathrm{e}^{\mu} = c_{\mu}\frac{h^2}{(1-h)^2}, \tag{19}$$

where c_{μ} is an integration constant. Inserting (18) into the Bianchi identity (17), one obtains

$$\frac{f'}{f} = \frac{2h'}{1-h}, \tag{20}$$

which can be integrated as

$$f = \frac{c_f}{(1-h)^2}, \tag{21}$$

c_f being another integration constant.

Now, (9) combined with (10), gives

$$\left(\frac{kf' - fk'}{D}\right)' = 0, \tag{22}$$

which can be integrated as[4]

$$kf' - fk' = 2cD, \tag{23}$$

where $2c$ is an integration constant, the factor 2 being chosen here for further convenience. Considered as a first order ordinary differential equation for $k(r)$, (23) possesses as a general solution

$$k = f\left(c_0 - 2c\int_{r_0}^{r}\frac{D(v)}{f(v)^2}\mathrm{d}v\right), \tag{24}$$

where c_0 and r_0 are new integration constants. With (21) inserted, it can be written as

$$k = \frac{c_f}{(1-h)^2}\left[c_0 - \frac{2c}{c_f^2}\int_{r_0}^{r}(1-h(v))^4\,D(v)\mathrm{d}v\right]. \tag{25}$$

The last metric function l follows from (8) as

$$l = \frac{(1-h)^2}{c_f}\left\{D^2 - \frac{c_f^2}{(1-h)^4}\left[c_0 - \frac{2c}{c_f^2}\int_{r_0}^{r}(1-h(v))^4\,D(v)\mathrm{d}v\right]^2\right\}. \tag{26}$$

3.5. *Interim results*

The field equations (9), (11) and (12), or, equivalently, (10), (11) and (12), give an expression for D'/D as a function of h and of its first and second derivatives. Then, inserting (19), (21), (25), (26) and the expression found for D'/D into (11) gives an expression for ρ depending on the same function h and its derivatives. Then inserting the same expressions into (9), or equivalently into (13), one obtains another expression for ρ still using h and derivatives

Equalizing both expressions for ρ yields a second order ordinary differential equation for h whose general solution allows one to calculate both the first and second derivatives of this h function. They are thus inserted into the previously obtained expressions for ρ and D'/D which can be integrated as

$$D^2 = c_f^2 c_k^2 \frac{(-2\ln h + 4h - h^2 + c_\beta)}{(1-h)(1+h)}. \tag{27}$$

Now (27) inserted into (25) gives, after integration,

$$k = \frac{c_f}{(1-h)^2}\left[c_0 + c_k(-2\ln h + 4h - h^2 + c_\beta)\right]. \tag{28}$$

Finally, the last metric function l emerges from inserting f, k and D into (8), which gives

$$l = c_f c_k^2 \frac{(1-h)}{(1+h)}(-2\ln h + 4h - h^2 + c_\beta) - \frac{c_f}{(1-h)^2}\left[c_0 + c_k(-2\ln h + 4h - h^2 + c_\beta)\right]^2. \tag{29}$$

3.6. *Regularity conditions*

The regularity conditions on the symmetry axis for metric (3) have already been displayed in.[4,5] However, since they will be needed in the following, their main outcomes are recalled briefly here.

To impose elementary flatness in the rotation axis vicinity, the norm X of the Killing vector ∂_ϕ must satisfy[2]

$$\lim_{r\to 0} \frac{g^{\alpha\beta} X_{,\alpha} X_{,\beta}}{4X} = 1, \tag{30}$$

where $X = g_{\phi\phi}$. Eqs. (3) and (30) yield

$$\lim_{r\to 0} \frac{e^{-\mu} l'^2}{4l} = 1. \tag{31}$$

The requirement that $g_{\phi\phi}$ vanishes on the axis implies

$$l \overset{0}{=} 0, \tag{32}$$

where $\overset{0}{=}$ means that the values are taken at $r = 0$.

Since there cannot be singularities along the axis, spacetime is forced to tend there to flatness. The coordinates are thus scaled such that, for $r \to 0$, the metric should be written as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + 2\omega r^2 \mathrm{d}t\mathrm{d}\phi + \mathrm{d}r^2 + \mathrm{d}z^2 + r^2\mathrm{d}\phi^2, \tag{33}$$

implying

$$f \overset{0}{=} \mathrm{e}^{\mu} \overset{0}{=} 1, \qquad k \overset{0}{=} 0, \tag{34}$$

giving

$$D \overset{0}{=} 0, \tag{35}$$

and, from (31) and (33),

$$l' \overset{0}{=} 0. \tag{36}$$

Then, the addition of the requirement that the Einstein tensor components in (9)-(13) do not diverge, implies

$$f' \overset{0}{=} k' \overset{0}{=} k'' - k'\frac{D'}{D} \overset{0}{=} 0, \tag{37}$$

Inserting (19), (21) and (28) into (34) one obtains

$$c_f \overset{0}{=} (1-h)^2, \tag{38}$$

$$c_\mu \overset{0}{=} \frac{c_f}{h^2}, \tag{39}$$

$$c_0 = 0, \tag{40}$$

$$c_\beta \overset{0}{=} 2\ln h - 4h + h^2. \tag{41}$$

All the other regularity conditions are verified provided (40) and (41) are satisfied.

3.7. *Junction conditions*

These conditions have also been displayed in[4,5] for metric (3). For completeness, and also since they will be partially needed further on, they are recalled here briefly.

Since the system is stationary, the Lewis vacuum metric[6] is used to represent the exterior spacetime, and the Weyl class[7] is here chosen for real junction condition purpose. Its metric is

$$\mathrm{d}s^2 = -F\mathrm{d}t^2 + 2K\mathrm{d}t\mathrm{d}\phi + \mathrm{e}^M(\mathrm{d}R^2 + \mathrm{d}z^2) + L\mathrm{d}\phi^2, \tag{42}$$

where

$$F = aR^{1-n} - a\delta^2 R^{1+n}, \tag{43}$$

$$K = -(1-ab\delta)\delta R^{1+n} - abR^{1-n}, \tag{44}$$

$$\mathrm{e}^{M} = R^{(n^2-1)/2}, \tag{45}$$

$$L = \frac{(1-ab\delta)^2}{a} R^{1+n} - ab^2 R^{1-n}, \tag{46}$$

with

$$\delta = \frac{c}{an}, \tag{47}$$

where a, b, c and n are the real constants determining the Weyl class of the metrics.

In accordance with Darmois' junction conditions,[9] metric (3) and metric (42)'s coefficients and their derivatives must be continuous across the Σ surface. It has been shown in[4] that, for the most general anisotropic fluid, these conditions imply $P_r \overset{\Sigma}{=} 0$, which is in agreement with the equation of state of the here studied fluid which imposes $P_r = 0$ everywhere.

Hence the conditions for a smooth junction to a Lewis vacuum of the Weyl class are fulfilled by this solution.

3.8. *Final form of the solution*

The constraints (38)–(41) on the integration constants, imposed by the regularity conditions, and inserted into the metric functions as given in (21), (19), (28) and (29) lead to their final form which can be displayed as

$$f = \left(\frac{1-h_0}{1-h}\right)^2, \tag{48}$$

$$\mathrm{e}^{\mu} = \left(\frac{1-h_0}{h_0}\right)^2 \left(\frac{h}{1-h}\right)^2, \tag{49}$$

$$k = (1-h_0)^2 c_k \frac{\left[2\ln\frac{h_0}{h} + 4(h-h_0) - (h^2-h_0^2)\right]}{(1-h)^2}, \tag{50}$$

$$l = (1-h_0)^2 c_k^2 \left[2\ln\frac{h_0}{h} + 4(h-h_0) - (h^2-h_0^2)\right] \times \left\{\frac{1-h}{1+h} - \frac{\left[2\ln\frac{h_0}{h} + 4(h-h_0) - (h^2-h_0^2)\right]}{(1-h)^2}\right\}. \tag{51}$$

An implicit integral equation for $h(r)$ reads

$$r = \frac{(1-h_0)^2}{c} \int_{h_0}^{h} \left\{\frac{1+u}{u^2(1-u)^3\left[2\ln\frac{h_0}{u} + 4(u-h_0) - (u^2-h_0^2)\right]}\right\}^{\frac{1}{2}} \mathrm{d}u, \tag{52}$$

where h_0 denotes the h value on the symmetry axis, i.e., $h_0 \equiv h(r=0)$. And the expression for ρ follows as

$$\rho = \frac{2c^2h_0^2}{\kappa(1-h_0)^4}\frac{(1-h)^4}{h^2(1+h)^2}\left\{\frac{h\left[2\ln\frac{h_0}{h}+4(h-h_0)-(h^2-h_0^2)\right]}{(1+h)}+2(1-h)^2\right\}. \tag{53}$$

Notice that the constant c_k appearing in (50) and (51) can be absorbed by a rescaling of the ϕ coordinate.

4. Physical properties of the solution

4.1. *Hydrodynamical scalars, vectors and tensors*

The timelike 4-vector V_α can be invariantly decomposed through the genuine tensor $V_{\alpha;\beta}$ into three independent parts such as

$$V_{\alpha;\beta} = -\dot{V}_\alpha V_\beta + \omega_{\alpha\beta} + \sigma_{\alpha\beta}, \tag{54}$$

where

$$\dot{V}_\alpha = V_{\alpha;\beta}V^\beta, \tag{55}$$

$$\omega_{\alpha\beta} = V_{[\alpha;\beta]} + \dot{V}_{[\alpha}V_{\beta]}, \tag{56}$$

$$\sigma_{\alpha\beta} = V_{(\alpha;\beta)} + \dot{V}_{(\alpha}V_{\beta)}. \tag{57}$$

These three quantities are, respectively, the acceleration vector, the rotation tensor, and the shear tensor. For the timelike 4-vector given by (5), their nonzero components[5] are

$$\dot{V}_1 = \frac{1}{2}\frac{f'}{f}, \tag{58}$$

which becomes, with (48) inserted,

$$\dot{V}_1 = \frac{c}{(1-h_0)^2}h\left\{\frac{(1-h)}{(1+h)}\left[2\ln\frac{h_0}{h}+4(h-h_0)-(h^2-h_0^2)\right]\right\}^{\frac{1}{2}} \tag{59}$$

and

$$2\omega_{13} = -(2kv'+k'v). \tag{60}$$

From (58), the acceleration vector modulus is therefore

$$\dot{V}^\alpha\dot{V}_\alpha = \frac{1}{4}\frac{f'^2}{f^2}\mathrm{e}^{-\mu}, \tag{61}$$

which becomes, with (48) and (49) inserted, and with (52) differentiated with respect to r

$$\dot{V}^\alpha\dot{V}_\alpha = \frac{c^2h_0^2}{(1-h_0)^6}\frac{(1-h)^3}{(1+h)}\left[2\ln\frac{h_0}{h}+4(h-h_0)-(h^2-h_0^2)\right]. \tag{62}$$

The rotation scalar, ω, defined by

$$\omega^2 = \frac{1}{2}\omega^{\alpha\beta}\omega_{\alpha\beta}, \tag{63}$$

follows as

$$\omega^2 = \frac{1}{4\mathrm{e}^{\mu}D^2}\left(k\frac{f'}{f} - k'\right)^2. \tag{64}$$

Inserting (23) into (64), one obtains

$$\omega^2 = \frac{c^2}{f^2\mathrm{e}^{\mu}}, \tag{65}$$

As already stressed in,[4,5] the shear tensor vanishes for any rigidly rotating fluid.

4.2. *Singularities*

The solution exhibits three possible singularities.

One would occur for $h = +1$ where the whole metric function set diverges, which is not the case for the density ρ. Indeed, from its expression (53), the energy density is merely seen to vanish. Since a more complete analysis should be needed to conclude definitively, the question whether this locus is or not a mere coordinate singularity remains open.

For $h = -1$, l is the only metric function which diverges. However, at this location, ρ also happens to diverge and would change sign if h was allowed to reach values below -1, which is not the case as it is shown elsewhere.[8] However, this behaviour would imply a curvature singularity if such a value of h should be reachable.

A third singularity might occur for $h = 0$. Here, the density diverges and is no more defined if the h sign happens to become different from the h_0 sign, owing to the presence of $\ln h_0/h$ in its expression. We would be therefore confronted to another curvature singularity. However, as it is shown elsewhere,[8] such a singularity lies outside the domain of definition of these spacetimes

4.3. *Integration constants and metric signature*

Among the three integration constants, c, c_k and h_0, one, c_k, can be suppressed by a rescaling of the azimuthal coordinate ϕ.

Now, h_0 possesses an obvious physical interpretation from the consideration of the expression for $h(r)$ given by (52). Since the integral in the right hand side vanishes for $h = h_0$, such does r on the left hand side, and therefore, h_0 must be interpreted as the value of h on the axis. Notice that $h_0 = 1$, implying the vanishing of the whole set of metric functions, lies thus outside the domain of definition of the solution.

The parameter c is shown elsewhere[8] to represent the amplitude of the rotation scalar of the fluid on the axis.

Finally, the signature of the metric $(-+++)$ implies sign constraints on the components of the metric tensor, which, in turn, impose restrictions on the validity ranges of $h(r)$.[8]

5. Conclusions

Following the investigations of stationary cylindrical anisotropic gravitational fluids initiated in,[4,5] the case of rigidly rotating fluids with particular equation of state $P_r = P_\phi = 0$ has been considered here. A new exact solution of the field equations has thus been exhibited under the form of $h(r)$ functions for the metric and the density, with h defined as $h(r) = P_z(r)/\rho(r)$, and of an integral expression for h as a r function. As usual, this solution is valid in a given coordinate system which has been chosen such as to ease a direct physical interpretation. To allow any future use for astrophysical purposes, this solution has been forced to satisfy the regularity conditions on the symmetry axis and it has been matched to the Weyl class of the Lewis vacuum solution through a cylindrical Σ hypersurface.

A number of important physical properties have been analysed. The hydrodynamical scalars, vectors and tensors of the fluid have been computed such as to be used for further interpretation purpose. Three singularities have been identified and discussed. However, they will prove to be located outside the definition domain of the solution, which will thus appear as singularity-free.

The solution depends on two independent parameters, h_0, the value of the function $h(r)$ on the axis of symmetry, and c which represents the amplitude of the rotation scalar of the fluid on the axis.[8] It has been stressed that the signature of the metric implies constraints on the signs of the components of the metric tensor and this has been shown, in turn, to restrict the validity domain of the solution.

In,[8] the properties of the constant parameters h_0 and c have been analysed and the different classes of solutions which might be associated to these properties have been identified.

Now, the present display of this first particular class of solutions for a fluid with axially directed stress will be followed, in the future, by the study of other complementary cases which would hopefully provide us with a deeper understanding of stationary rotating cylinders of matter in GR.

References

1. J. B. Griffiths and J. Podolský, *Exact spacetimes in Einstein's General Relativity,* Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2009).
2. H. Stephani, D. Kramer, M. MacCallum, C. Honselaers, and E. Herlt, *Exact Solutions to Einstein's Field Equations,* Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2009).
3. K. A. Bronnikov, N. O. Santos, and A. Wang, Cylindrical systems in general relativity, Classical Quantum Gravity **37**, 113002 (2020).

4. F. Debbasch, L. Herrera, P. R. C. T. Pereira, and N. O. Santos, Stationary cylindrical anisotropic fluid, Gen. Relativ. Gravit. **38**, 1825 (2006).
5. M.-N. Célérier and N. O. Santos, Stationary cylindrical anisotropic fluid and new purely magnetic GR solutions, Phys. Rev. D **102**, 044026 (2020).
6. T. Lewis, Some special solutions of the equations of axially symmetric gravitational fields, Proc. R. Soc. A **136**, 176 (1932).
7. M. F. A. da Silva, L. Herrera, F. M. Paiva, and N.O. Santos, The parameters of the Lewis metric for the Weyl class, Gen. Relativ. Gravit. **27**, 859 (1995).
8. M.-N. Célérier, New classes of exact interior nonvacuum solutions to the GR field equations for spacetimes sourced by a rigidly rotating stationary cylindrical anisotropic fluid, Phys. Rev. D **104**, 064040 (2021).
9. G. Darmois, Les équations de la gravitation einsteinienne, *Mémorial des Sciences Mathématiques* fasc. XXV (Gauthier-Villars, Paris, 1927).

Hydrodynamic representation and energy balance for the Dirac and Weyl fermions in curved space-times

Tonatiuh Matos* and Omar Gallegos

Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, A.P. 14-740, 07000 CDMX, México

**E-mail: tonatiuh.matos@cinvestav.mx*

Pierre-Henri Chavanis

Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France

Using a generalized Madelung transformation, we derive the hydrodynamic representation of the Dirac equation in arbitrary curved space-times coupled to an electromagnetic field. We obtain Dirac-Euler equations for fermions involving a continuity equation and a first integral of the Bernoulli equation. Using the comparison of the Dirac and Klein-Gordon equations we obtain the balance equation for fermion particles. We also use the correspondence between fermions and bosons to derive the hydrodynamic representation of the Weyl equation which is a chiral form of the Dirac equation.

Keywords: Dirac equation; Weyl equation; Fermions; Hydrodynamics

1. Introduction

The Standard Model of elementary particles establishes that there exist two kinds of particles, fermions and bosons. In previous works,[1,2] the energy balance for bosons was derived starting from the general relativistic Klein-Gordon (KG) equation. In the present work, we study a system of fermions described by the Dirac equation in arbitrary curved space-times taking into account electromagnetic effects. We also use the Weyl equation which is a chiral form of the Dirac equation due to the relationship between the Lie algebras of the symmetry groups for both systems of particles. We give the hydrodynamic representation of the Dirac and Weyl equations for fermions using previous results obtained for boson particles.

Many examples of fermion particles in strong gravitational fields can be found in nature. Indeed, the curvature of space-time plays an important role in a neutron star, in the early Universe, or in a fermion cloud (e.g. a dark matter halo) in the vicinity of a black hole. We need to develop a general framework to identify what are the different energy contributions in such systems. In this work we use the geometrical decomposition of the metric in 3+1 slices and the tetrad formalism to study the particle spin in an arbitrary space-time. We define the gamma matrices in curved space-times and derive the generalized Dirac and Weyl equations. Then, using the Madelung transformation, we introduce a hydrodynamic representation of the Dirac and Weyl spinors. This hydrodynamic representation can help us to describe the fermionic system in a general framework.

2. Field Equations

We use the tetrad formalism for the space-time geometry, and the canonical expansion of the space-time in a 3+1 ADM decomposition,[3–8] such that the coordinate t is the parameter of evolution. The 3+1 metric reads

$$\mathrm{d}s^2 = N^2c^2\mathrm{d}t^2 - h_{ij}\left(\mathrm{d}x^i + N^i c\,\mathrm{d}t\right)\left(\mathrm{d}x^j + N^j c\,\mathrm{d}t\right), \tag{1}$$

where N represents the lapse function which measures the proper time of the observers traveling along the world line, N^i is the shift vector that measures the displacement of the observers between the spatial slices and h_{ij} is the 3-dimensional slice-metric. In what follows $i,j,k,l = 1,2,3$; $a,b,c = 0,1,2,3$ are the internal indices and $\mu,\nu,\alpha = 0,1,2,3$ the space-time indices. We write eq. (1) in the tetrad formalism as $\mathrm{d}s^2 = \eta_{ab}e^a_{\ \mu}e^b_{\ \nu}dx^\mu dx^\nu$, where $\eta_{ab} = \mathrm{diag}(1,-1,-1,-1)$. Here $e^a = e^a_{\ \mu}dx^\mu$ is the set of one-forms base of the cotangent space at the space-time manifold given by

$$\begin{aligned} e^0 &= N\mathrm{d}t, \\ e^k &= \hat{e}^k_{\ i}\left(\mathrm{d}x^i + N^i c\,\mathrm{d}t\right), \end{aligned} \tag{2}$$

with inverse

$$\begin{aligned} e_0 &= \frac{1}{N}\left(\frac{\partial}{c\,\partial t} - N^j\frac{\partial}{\partial x^j}\right), \\ e_k &= \hat{e}_k^{\ j}\frac{\partial}{\partial x^j}, \end{aligned} \tag{3}$$

where $\hat{e}^k = \hat{e}^k_{\ i}\mathrm{d}x^i$ are the one-form base to the three-dimensional slice of the cotangent manifold, such that $h_{ij} = \delta_{kl}\hat{e}^k_{\ i}\hat{e}^l_{\ j}$. We can also define the set of vectors base of the tangent-space to the space-time as $e_a = e_a^{\ \mu}\partial_\mu$, such that $e^a e_b = \delta^a_{\ b}$. We will use the tetrad formalism[3,6–10] to describe the space-time geometry where the fermion particles are located.

The action of a fermion system in curved space-time coupled to an electromagnetic field A_μ is given by $S[\psi,\partial_\mu\psi] = \int \mathcal{L}(\psi,\partial_\mu\psi(x^\mu))\,d^4x$, where $\mathcal{L} = \mathcal{L}(\psi,\partial_\mu\psi(x^\mu))$ is the Lagrangian density[11–13]

$$\mathcal{L} = \sqrt{-g}\frac{i\hbar c}{2}\left[\psi^\dagger B\gamma^\mu\left(D_\mu\psi\right) - \left(D_\mu\psi\right)^\dagger B\gamma^\mu\psi + \frac{2imc}{\hbar}\psi^\dagger B\psi\right]. \tag{4}$$

Here $D_\mu = \nabla_\mu + \dfrac{iq}{\hbar c}A_\mu$ is the total covariant derivative accounting for electromagnetic effects. The covariant derivative of a spinor $\psi = (\psi_{\dot{\nu}})$ is given by $\nabla_\mu(\psi_{\dot{\nu}}) = \partial_\mu(\psi_{\dot{\nu}}) + \Gamma^{\dot{\alpha}}_{\mu\dot{\nu}}(\psi_{\dot{\alpha}})$, where $\Gamma^{\dot{\alpha}}_{\mu\dot{\nu}}$ is the spin connection.[3,14] Using the least action principle it is possible to obtain from eq. (4) the corresponding Dirac equation. This equation is given by

$$\left[i\hbar\gamma^\mu\left(\nabla_\mu + iqA_\mu\right) - mc\right]\psi = 0, \tag{5}$$

where $\hbar$, c are the Planck constant and the speed of light respectively, while q,m are the charge and mass of the fermion particle and ψ is its spinor. Besides, the

gamma matrices γ^μ are related to the spin and space-time geometry. They can be written as $\gamma^\mu = e^\mu_{\ a}\tilde{\gamma}^a$, where $\tilde{\gamma}^a$ are the gamma matrices in flat space-time, which are well-know from Quantum Field Theory (QFT).[15–17] Therefore,

$$\begin{aligned}\gamma^0 &= N\tilde{\gamma}^0,\\ \gamma^k &= \hat{e}^k_{\ j}(\tilde{\gamma}^j + N^j\tilde{\gamma}^0).\end{aligned} \tag{6}$$

In general, these matrices fulfill the following anti-commutation relation[3,18]

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \tag{7}$$

where $g_{\mu\nu}$ represents the metric that describes the space-time geometry. Furthermore, as we know, the gamma matrices in flat space-time are related to the Pauli matrices, which describe the spin of the fermion particles. In general, the gamma matrices obey the following relation[11–13,19]

$$(\gamma^\mu)^\dagger = B\gamma^\mu B^{-1}, \tag{8}$$

where B is a hermitian matrix, i.e. $B^\dagger = B$, that is uniquely determined by the gamma matrices γ^μ. As usual, we denote by $B^\dagger$ the conjugate (or Hermitian) transpose of B. We note that in QFT the relation (8) is fulfilled when $B = \tilde{\gamma}^0$ and the gamma matrices are in flat space-time. From the action (4) of the fermion system we can find the equation for the transpose conjugated spinor by making an infinitesimal variation of this action with respect to ψ. Another way of getting this equation of motion is to take the transpose conjugate of the Dirac equation (5) and using (8). In this manner we find that the transpose conjugated Dirac equation in curved space-time is given by

$$i\left(\nabla_\mu\bar{\psi}\right)\gamma^\mu - i\psi^\dagger\nabla_\mu\left(B\gamma^\mu\right) + i\bar{\psi}\nabla_\mu\gamma^\mu + \bar{\psi}A_\mu\gamma^\mu + m\bar{\psi} = 0. \tag{9}$$

To simplify the notations, here and in the following we use $mc/\hbar \to m$ in natural units ($c = \hbar = 1$). We consider that $(\nabla_\mu\psi)^\dagger = \nabla_\mu\psi^\dagger$, and we denote the adjoint spinor as $\bar{\psi} = \psi^\dagger B$. Using the gamma matrices in flat space-time and the fact that $B = \tilde{\gamma}^0$ we recover the definition of $\bar{\psi}$ in QFT and the transpose conjugated Dirac equation. However, in an arbitrary space-time $\nabla_\mu\gamma^\mu$ is distinct from zero, since $\gamma^\mu = e^\mu_{\ a}\tilde{\gamma}^a$, so in general $\nabla_\mu e^\mu_{\ a}$ is non-zero.

We can get the conserved charge from the Noether theorem.[20] The Dirac current is

$$J^\mu = \bar{\psi}\gamma^\mu\psi = \psi^\dagger B\gamma^\mu\psi. \tag{10}$$

To obtain the continuity equation

$$\nabla_\mu J^\mu = 0 \tag{11}$$

for the Dirac current, we take the covariant derivative of eq. (10). This gives

$$\nabla_\mu J^\mu = (\nabla_\mu\bar{\psi})\gamma^\mu\psi + \bar{\psi}\left(\nabla_\mu\gamma^\mu\right)\psi + \bar{\psi}\gamma^\mu\nabla_\mu\psi. \tag{12}$$

If we multiply the Dirac equation (5) by $\bar{\psi}$ and its transpose conjugate (9) by ψ and sum both equations, it follows that

$$\nabla_\mu J^\mu = \psi^\dagger \nabla_\mu \left(B\gamma^\mu\right)\psi. \tag{13}$$

If we require that the continuity equation (11) is fulfilled, i.e., that the number of particles is conserved, then we need $\nabla_\mu \left(B\gamma^\mu\right) = 0$, or equivalently

$$(\nabla_\mu B)\gamma^\mu = -B\nabla_\mu\gamma^\mu. \tag{14}$$

At this point, we want to emphasize the consistency conditions for the continuity equation (11). Some authors[21] impose $\nabla_\mu\gamma^\nu = 0$ while others[22] impose $\nabla_\mu B = 0$. These conditions are independent of each other, i.e., in general the condition of Ref.[21] is not fulfilled in Ref.[22] and *vice versa*. In Refs.[12,13] the authors conclude that the condition $\nabla_\mu(B\gamma^\nu) = 0$ is the most convenient because it is implied by $\nabla_\mu\gamma^\nu = 0$ and $\nabla_\mu B = 0$.

In addition, we can note that the matrix B can be obtained for a general metric (1) by solving the differential equation

$$\left(\nabla_0(BN) + \nabla_j(B\hat{e}_i^j N^i)\right)\tilde{\gamma}^0 - \nabla_j(B\hat{e}_i^j)\tilde{\gamma}^i = 0, \tag{15}$$

which follows from eq. (14). Using the condition (14), it is possible to rewrite the transpose conjugated Dirac equation (9) as

$$i\left(\nabla_\mu\bar{\psi}\right)\gamma^\mu + i\bar{\psi}\nabla_\mu\gamma^\mu + \bar{\psi}A_\mu\gamma^\mu + m\bar{\psi} = 0. \tag{16}$$

In order to find the conserved quantity resulting from the continuity equation, we take an arbitrary surface $\mathcal{S}$ enclosing the volume $\mathcal{V}$ which contains the whole system. Let k^j be an orthonormal vector to $\mathcal{S}$ such that

$$\int_{\mathcal{V}} \nabla_\mu J^\mu dV = \int_{\mathcal{V}} \nabla_0 J^0 dV + \int_{\mathcal{S}} k_j J^j dS = 0. \tag{17}$$

We assume that far away from the source, J^μ is negligible. Then, the surface integral in eq. (17) vanishes, and we obtain

$$\frac{dQ}{dt} = \int_{\mathcal{V}} \nabla_0 J^0 dV = 0, \tag{18}$$

where $Q = \int_{\mathcal{V}} J^0 dV$ is the conserved charge and dV is the curved volume element $dV = \sqrt{-g}d^4x$. In QFT this charge is identified with the number of fermions or with the electric charge of the system. In flat space-time we have $B = \tilde{\gamma}^0$, so that $J^0 = \psi^\dagger\psi = n$ represents the number density of fermion particles. In curved space-time J^0 (which is determined by γ^0 and by the generalized gamma matrices) has a different interpretation. The form of B given by eqs. (8) and (14) for each metric is related to the gamma matrices and to the tetrad formalism.

With the Maxwell four-potential we can define the Faraday tensor

$$F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu. \tag{19}$$

In the electromagnetic theory, the Faraday tensor $F_{\mu\nu}$ satisfies the Maxwell field equations

$$\nabla_\nu F^{\nu\mu} = J^{E\mu}, \tag{20}$$

where $J^{E\mu}$ is the four-electromagnetic current.

The problem of the Energy Balance for boson particles in a curved space-time is studied in Ref.[1], where the conserved 4-current associated with the KG equation describing the evolution of a complex scalar field $\Phi(x^\mu)$ is defined. We can generalize this idea by defining a new 4-current J_μ^{KG}, changing the scalar field by a spinor and the complex conjugate scalar field by the conjugate transpose of the spinor. Namely, the KG current is redefined as

$$J_\mu^{KG} = i\frac{q}{2m^2}\left[\psi\left(\nabla_\mu - iqA_\mu\right)\psi^\dagger - \psi^\dagger\left(\nabla_\mu + iqA_\mu\right)\psi\right]. \tag{21}$$

3. Dirac Hydrodynamic Representation

Analogously to the hydrodynamic representation of the Schrödinger equation, which was introduced by Madelung,[23] we derive the hydrodynamic representation of the Dirac equation. We carry out the following generalized Madelung transformation for each component of the spinor $\psi = \psi(x^\mu)$ as follows

$$\psi = \exp(i\theta)R, \tag{22}$$

where R is a spinor and θ is a function. For the case where we consider a Dirac electron-like fermion, $\theta = \theta(x^\mu)$, the spinor ψ reads

$$\psi = \begin{pmatrix} R_{\dot{1}} \\ R_{\dot{2}} \\ R_{\dot{3}} \\ R_{\dot{4}} \end{pmatrix} \exp(i\theta) = R\exp(i\theta), \tag{23}$$

where we use the notation $\dot{\mu}$, $\dot{\nu}$, ...$= \dot{1}, \cdots, \dot{4}$ for the spinor indices such that

$$R = \begin{pmatrix} R_{\dot{1}} \\ R_{\dot{2}} \\ R_{\dot{3}} \\ R_{\dot{4}} \end{pmatrix} = \begin{pmatrix} \sqrt{n_{\dot{1}}} \\ \sqrt{n_{\dot{2}}} \\ \sqrt{n_{\dot{3}}} \\ \sqrt{n_{\dot{4}}} \end{pmatrix}. \tag{24}$$

In the appendix we show some exact solutions of the Dirac equation with this ansatz in flat space-time. Here the dot indices represent elements of each component and we do not use the sum convention when the indices are up and down. On the other hand, we assume this notation $R_{\dot{\mu}} = \sqrt{n_{\dot{\mu}}}$ to (24), where $n_{\dot{\mu}}$ is the number density, represents the modulus of $\psi_{\dot{\mu}}$ and θ is its phase (both are real variables). In general, $n_{\dot{\mu}}$ is different for each component of the spinor. Note that the covariant derivative of the spinor ψ in terms of its decomposition (23) is $\nabla_\mu(\psi_{\dot{\nu}}) = \partial_\mu(R_{\dot{\nu}}e^{i\theta}) + \Gamma^{\dot{\alpha}}_{\mu\dot{\nu}}(R_{\dot{\alpha}}e^{i\theta}) = (\partial_\mu R_{\dot{\nu}})e^{i\theta} + i(\partial_\mu\theta)R_{\dot{\nu}}e^{i\theta} + \Gamma^{\dot{\alpha}}_{\mu\dot{\nu}}(R_{\dot{\alpha}}e^{i\theta})$, implying that $\nabla_\mu\theta = \partial_\mu\theta$.

Using the transformation (23) in eq. (5), the Dirac equation in terms of the variables R and θ reads

$$\exp(i\theta)\gamma^{\mu}\left(i\nabla_{\mu}R-(\nabla_{\mu}\theta)R-qA_{\mu}R-\frac{m}{4}\gamma_{\mu}R\right)=0. \tag{25}$$

To get the last term, we used the property of the gamma matrices that $\gamma_{\mu}\gamma^{\mu}=4\mathbb{I}$, where $\mathbb{I}$ is the 4×4 identity matrix. This property results from the anti-commutation relation of the gamma matrices. Similarly, the continuity equation (11) with (10) can be written with these new variables as

$$\left(\nabla_{\mu}R^{T}\right)K^{\mu}R+R^{T}K^{\mu}\left(\nabla_{\mu}R\right)=0, \tag{26}$$

where R^{T} denotes the transpose of R and $K^{\mu}=B\gamma^{\mu}$. Observe that K^{μ} is hermitian ($K^{\mu\dagger}=K^{\mu}$).

In conclusion, we have introduced the hydrodynamic representation of the Dirac equation (25) and its conjugate transpose equation by making the change of variables from eq. (22).

4. Dirac-Euler Equation

As for the Klein-Gordon equation,[1,2] we define the 4-velocity v_{μ} by

$$mv_{\mu\dot{\nu}}=\nabla_{\mu}S_{\dot{\nu}}+qA_{\mu}. \tag{27}$$

Here, $S(x^{\mu})$ is a phase with components $S_{\dot{\alpha}}=\theta_{\dot{\nu}}\delta^{\dot{\nu}}{}_{\dot{\alpha}}-\omega_{\dot{\nu}}\delta^{\dot{\nu}}{}_{\dot{\alpha}}t$ where $\omega_{\dot{\nu}}$ are constants that can be related to the mass of the fermion particle by $\omega_{\dot{\nu}}=mc^{2}/\hbar$ and $\theta_{\dot{\nu}}=\theta$. In this manner we can write

$$\nabla_{\mu}\theta_{\dot{\alpha}}\delta^{\dot{\alpha}}{}_{\dot{\nu}}=(mv_{\mu\dot{\alpha}}-\omega_{\dot{\alpha}}\delta^{0}{}_{\mu})\delta^{\dot{\alpha}}{}_{\dot{\nu}}-qA_{\mu}. \tag{28}$$

We interpret $n_{\dot{\nu}}$ as the density number of fermions and $v_{\mu\dot{\nu}}$ as its velocity. Actually, eq. (25) can be interpreted as the first integral of the Bernoulli equation for fermions in an arbitrary space-time. To see this, we apply the operator $i\gamma^{\mu}D_{\mu}=i\gamma^{\mu}\nabla_{\mu}-q\gamma^{\mu}A_{\mu}$ to the Dirac equation (5) written under the form $i\gamma^{\mu}\nabla_{\mu}\psi=q\gamma^{\mu}A_{\mu}\psi-m\psi$. This yields

$$-\gamma^{\mu}\gamma^{\nu}\left(\nabla_{\mu}\nabla_{\nu}\psi+iq(\nabla_{\mu}A_{\nu})\psi+iqA_{\nu}(\nabla_{\mu}\psi)+iqA_{\mu}(\nabla_{\nu}\psi)-q^{2}A_{\mu}A_{\nu}\psi\right)- m^{2}\psi-\gamma^{\mu}(\nabla_{\mu}\gamma^{\nu})(\nabla_{\nu}\psi+iqA_{\nu}\psi)=0. \tag{29}$$

Using the relation (7) in eq. (29) we obtain

$$\Box_{E}\psi+m^{2}\psi+\frac{i}{2}q\gamma^{\mu}\gamma^{\nu}F_{\mu\nu}\psi+\gamma^{\mu}(\nabla_{\mu}\gamma^{\nu})(D_{\nu}\psi)=0, \tag{30}$$

where we have defined the D'Alambertian operator in the presence of an electromagnetic field by $\Box_{E}=(\nabla_{\mu}+iqA_{\mu})(\nabla^{\mu}+iqA^{\mu})$ and the anti-symmetric Faraday tensor by $F_{\mu\nu}=\nabla_{\mu}A_{\nu}-\nabla_{\nu}A_{\mu}$. Eq. (30) is similar to the Klein-Gordon equation with an electromagnetic source except that here ψ is a spinor instead of a complex SF. The last term of eq. (30) contains the covariant derivative of γ^{μ} which vanishes

in a flat space-time. According to Refs.[1,2] if we apply the transformation (22) to eq. (30), we could expect to obtain the continuity equation for the imaginary part and the Bernoulli equation for the real part. However, in the case of the Dirac equation the four components are mixed by the presence of the four dimensional spinor ψ. Hence, we obtain the following expression

$$i\left[2(mv^\mu-\omega\delta_0^\mu)\nabla_\mu R-qA_\mu+q\nabla_\mu(A^\mu R)+\nabla_\mu(mv^\mu-\omega\delta_0^\mu-qA^\mu)R\right]+\left(m^2v_\mu v^\mu+2m\omega v^0+\frac{\omega^2}{N^2}+m^2\right)R-\Box R+\frac{i}{2}q\gamma^\mu\gamma^\nu F_{\mu\nu}R+\gamma^\mu(\nabla_\mu\gamma^\nu)(i(mv_\nu+\omega\nabla_\nu t)R+D_\nu R)=0. \quad (31)$$

Here, we have defined $\Box=\nabla^\nu\nabla_\nu$ and we have introduced the diagonal matrices $v_\mu=v_{\mu\dot{\nu}}\delta^{\dot{\nu}}_{\dot{\alpha}}$ and $\omega=\omega_{\dot{\nu}}\delta^{\dot{\nu}}_{\dot{\alpha}}$. For bosons, the real and imaginary parts separate into two independent equations, namely, the continuity equation and the Bernoulli equation.[1,2] But in the spinor case, the last line of equation (31) mixes both the imaginary and real parts and there is no natural separation into real and imaginary parts. The system remains coupled.

5. Weyl Representation

The Dirac equation for 1/2-spin particles is associated with the $SO(1,3)$ symmetry group. Nevertheless, we can introduce a new representation as in standard QFT, since there exists a surjective homomorphism between the $SO(1,3)$ and $SU(2)\otimes SU(2)$ Lie groups.

In terms of the Pauli matrices the 4×4 gamma matrices γ^μ can be written as two 2×2 block matrices

$$\gamma^0=N\tilde{\gamma}^0=N\begin{pmatrix}0&\mathbb{I}\\\mathbb{I}&0\end{pmatrix}, \quad (32)$$

$$\gamma^j=\hat{e}^j{}_i(\tilde{\gamma}^i+N^i\tilde{\gamma}^0)=\begin{pmatrix}0&-\hat{e}^j{}_i(\tilde{\sigma}^i-N^i\mathbb{I})\\\hat{e}^j{}_i(\tilde{\sigma}^i+N^i\mathbb{I})&0\end{pmatrix}, \quad (33)$$

where $\tilde{\sigma}^i$ are the 2×2 Pauli matrices in flat space-time and $\mathbb{I}$ is the 2×2 identity matrix. The γ^μ matrices satisfy $\left(\gamma^0\right)^\dagger=\gamma^0$ and $\left(\gamma^j\right)^\dagger=-\gamma^j+2N^j\gamma^0/N$.

As we know, the special unitary group $SU(2)$ is formed by the set of 2×2 complex matrices A, which satisfy $\det(A)=1$. Explicitly, we have

$$A=\begin{pmatrix}a&-\bar{b}\\b&\bar{a}\end{pmatrix}, \quad (34)$$

with $\det(A)=|a|^2+|b|^2=1$, where a and b are complex parameters. Equivalently, we have the identity $A^\dagger=A^{-1}$.

The Lie algebra $\mathfrak{su}(2)$ associated to the $SU(2)$ Lie group is given by the exponential map

$$\exp(\mathfrak{su}(2))\to SU(2). \quad (35)$$

For any element X of the Lie algebra, we have $\exp(X)\exp(X)^\dagger = \mathbb{I}$, implying that $X + X^\dagger = 0$. In what follows, we will indistinctly use $\exp(X)$ and e^X as the exponential map.

In the Weyl representation we can write a Dirac fermion as a four-spinor ψ made of two spinors each of which having two components, for instance

$$\psi = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix}, \tag{36}$$

where ψ_R and ψ_L are the right- and the left- handed Weyl spinors, respectively. If we write the adjoint spinor $\bar{\psi}$ and use the Weyl representation, it follows that

$$\bar{\psi} = \psi^\dagger B = \left(\psi_R^\dagger, \psi_L^\dagger\right) B, \tag{37}$$

where B is the matrix from eqs. (8) and (14). If we use the relation (8) it is straightforward to see that the matrix B must have the following form

$$B = \begin{pmatrix} 0 & B_\zeta \\ B_\zeta & 0 \end{pmatrix}, \tag{38}$$

where the 2×2 matrix B_ζ is a diagonal matrix, $B_\zeta = b\mathbb{I}$, with $b = b(x^\mu)$. Therefore, we get $B = b\tilde{\gamma}^0$ and eq. (15) transforms into

$$\nabla_0(Nb) + \nabla_j(\hat{e}_i^j N^i b) = 0, \tag{39}$$

$$\nabla_j(\hat{e}_i^j b)\tilde{\sigma}^i = 0. \tag{40}$$

Using the definition of the spinor and its adjoint we can write the Dirac quadricurrent J^μ from eq. (10) as

$$J^\mu = \left(\psi_R^\dagger, \psi_L^\dagger\right) B\gamma^\mu \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix}, \tag{41}$$

where the gamma matrices are defined by eqs. (32) and (33) and, in general, B is given by the previously mentioned conditions. This yields

$$J^0 = Nb(\psi_R^\dagger \psi_R + \psi_L^\dagger \psi_L) = Nbn, \tag{42}$$

$$J^j = b\hat{e}^j{}_i(\psi_R^\dagger(\tilde{\sigma}^i + N^i\mathbb{I})\psi_R - \psi_L^\dagger(\tilde{\sigma}^i - N^i\mathbb{I})\psi_L). \tag{43}$$

In order to simplify the notation, we now define the vectors of 2×2 matrices $\mathbb{S}^a = (\mathbb{I}, \tilde{\sigma}^j + N^j\mathbb{I})$ and $\bar{\mathbb{S}}^a = (-\mathbb{I}, \tilde{\sigma}^j - N^j\mathbb{I})$ in terms of the Pauli matrices. $\mathbb{S}^a$ and $\bar{\mathbb{S}}^a$ are the (generalized) Pauli matrices in flat space-time. In terms of these new definitions, the density currents read

$$\begin{aligned} J^j &= b\hat{e}^j{}_i(\psi_R^\dagger \mathbb{S}^i \psi_R - \psi_L^\dagger \bar{\mathbb{S}}^i \psi_L) \\ &= b(\psi_R^\dagger \sigma^i \psi_R - \psi_L^\dagger \bar{\sigma}^i \psi_L), \end{aligned} \tag{44}$$

where we have defined the 2×2 Pauli matrices in a curved space-time by $\sigma^\mu = e^\mu{}_a \mathbb{S}^a$ and $\bar{\sigma}^\mu = e^\mu{}_a \bar{\mathbb{S}}^a$. With this definition, the matrices γ^j read

$$\gamma^j = \begin{pmatrix} 0 & -\bar{\sigma}^j \\ \sigma^j & 0 \end{pmatrix}. \tag{45}$$

Furthermore, observe that the σ^j matrices follow the same commutation relations as the flat space-time Pauli matrices. This means that $[\sigma^i, \bar{\sigma}^j] = -\hat{e}^i_k \hat{e}^j_l [\tilde{\sigma}^k, \tilde{\sigma}^l]$. For the Weyl representation we have to obtain two equations for each Dirac fermion. Thus, we need to redefine the covariant derivative ∇_μ and the spinor affine connection Γ_μ,[14,24] which can be written as $\nabla_\mu = \partial_\mu + \Gamma_\mu$ and $\Gamma_\mu = \frac{1}{4}\bar{\sigma}_\nu \sigma^\nu_{;\mu}$ where $\sigma^\mu_{;\nu} = \partial_\nu \sigma^\mu + \Gamma^\mu_{\alpha\nu}\sigma^\alpha$. Nevertheless, in this representation we need to introduce two other notations due to the presence of $\bar{\sigma}^\mu$. Let $\bar{\nabla}_\mu$ and $\tilde{\Gamma}_\mu$ be the bar covariant derivative and the bar spinor affine connection, respectively, defined by $\bar{\nabla}_\mu = \partial_\mu + \tilde{\Gamma}_\mu$ where $\tilde{\Gamma}_\mu = \frac{1}{4}\sigma_\nu \bar{\sigma}^\nu_{;\mu}$ (we stress that we use the greek indices for denoting the objects in curved space-time as the gamma and Pauli matrices).

We can now apply the Weyl representation to rewrite the Dirac equation (5) for a spinor with four components as

$$\begin{pmatrix} i\sigma^\mu \left(\bar{\nabla}_\mu + iqA_\mu\right)\psi_R - m\psi_L \\ i\bar{\sigma}^\mu \left(\nabla_\mu + iqA_\mu\right)\psi_L - m\psi_R \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{46}$$

These are the Weyl equations for a spinor in a curved space-time coupled to an electromagnetic field. If we apply the Weyl representation to the transpose conjugated Dirac equation (16), it is straightforward to obtain the Weyl equation for the adjoint spinor (37). However, we shall not write the adjoint spinor equation explicitly because the results are analogous to the spinor equation as we have seen in the previous sections.

If we set $B = b\tilde{\gamma}^0$, the current density now reads

$$J^\mu = b\left(\psi^\dagger_R \sigma^\mu \psi_R - \psi^\dagger_L \bar{\sigma}^\mu \psi_L\right). \tag{47}$$

Explicitly, we have

$$J^0 = Nbn, \tag{48}$$

$$J^j = b\hat{e}^j_i \left(\psi^\dagger_R \tilde{\sigma}^i \psi_R - \psi^\dagger_L \tilde{\sigma}^i \psi_L + \frac{N^i}{Nb^2} J^0\right). \tag{49}$$

On the other hand, the last line of eq. (31) can be obtained from the identities

$$\gamma^\mu \gamma^\nu F_{\mu\nu}\psi = \begin{cases} (2NN^k F_{0k} + i\hat{F}_{ij}\epsilon^{ij}{}_k \tilde{\sigma}^k)\psi_R \\ -(2NN^k F_{0k} - i\hat{F}_{ij}\epsilon^{ij}{}_k \tilde{\sigma}^k)\psi_L \end{cases}, \tag{50}$$

and using definition (45), we find that

$$\begin{aligned} \gamma^\mu(\nabla_\mu \gamma^\nu)(D_\nu \psi) &= \begin{cases} -\bar{\mathbb{S}}^a \mathbb{S}^b (\hat{\nabla}_a \hat{e}^\nu_b)(D_\nu \psi_R) \\ -\mathbb{S}^a \bar{\mathbb{S}}^b (\hat{\nabla}_a \hat{e}^\nu_b)(D_\nu \psi_L) \end{cases} \\ &= \begin{cases} (N(\nabla_0 N) - \bar{\sigma}^j(\nabla_j N))(D_0\psi_R) + (N(\nabla_0 \sigma^i) - \bar{\sigma}^j(\nabla_j \sigma^i))(D_i \psi_R) \\ (N(\nabla_0 N) + \sigma^j(\nabla_j N))(D_0\psi_L) - (N(\nabla_0 \bar{\sigma}^i) - \sigma^j(\nabla_j \bar{\sigma}^i))(D_i \psi_L) \end{cases} \\ &= \begin{cases} (\hat{\nabla}_0 N - \bar{\mathbb{S}}^k(\hat{\nabla}_k N))(D_0\psi_R) + (\mathbb{S}^k \hat{\nabla}_0 \hat{e}^i_k - \bar{\mathbb{S}}^k \mathbb{S}^l \hat{\nabla}_k \hat{e}^i_l))(D_i \psi_R) \\ (\hat{\nabla}_0 N + \mathbb{S}^k(\hat{\nabla}_k N))(D_0\psi_L) - (\bar{\mathbb{S}}^k \hat{\nabla}_0 \hat{e}^i_k - \mathbb{S}^k \bar{\mathbb{S}}^l (\hat{\nabla}_k \hat{e}^i_l))(D_i \psi_L), \end{cases} \end{aligned} \tag{51}$$

where $\epsilon^{ij}{}_k$ is the usual Levi-Civita tensor, $\hat{F}_{ij} = \hat{e}_i^l \hat{e}_j^m F_{lm}$ is the directional Maxwell tensor $\hat{F}_{ij} = (\hat{e}_i^l \hat{\nabla}_j - \hat{e}_j^l \hat{\nabla}_i) A_l$, and $\hat{\nabla}_a = \hat{e}_a^\alpha \nabla_\alpha$ is the directional covariant derivative which defines the Cartan connection $\hat{\nabla}_c \hat{e}_b^\nu = \Gamma_{bc}^a \hat{e}_a^\nu$. The Cartan connection $\Gamma_{bc}^a = \hat{e}_\nu^a \hat{\nabla}_c \hat{e}_b^\nu$ determines the Cartan first fundamental form $d\hat{e}^a + \Gamma_b^a \wedge \hat{e}^b$ for the connections $\Gamma_b^a = \Gamma_{bd}^a \hat{e}^d$ with the property that $\Gamma_{ab} + \Gamma_{ba} = 0$, where $\Gamma_{ab} = \eta_{ad} \Gamma_b^d$.

6. Weyl Hydrodynamic Representation

We now have all the ingredients to propose a hydrodynamic representation for the Weyl fermions, following the same procedure as the one developed for the Schrödinger and KG equations in Refs.[1,2].

We start to propose our Madelung transformation in the Weyl spinor, using the exponential map, that is

$$\Psi = \begin{pmatrix} \psi_R \\ \psi_L \end{pmatrix} = \begin{pmatrix} R_R \\ R_L \end{pmatrix} e^{i\theta}. \tag{52}$$

Since ψ_R and ψ_L are two spinors, we observe that R_R and R_L are two two-dimensional vectors. The Weyl representation of the adjoint spinor $\bar{\Psi}$ when $B = b\tilde{\gamma}^0$ is

$$\bar{\Psi} = b \left(\psi_R^\dagger, \psi_L^\dagger \right) \tilde{\gamma}^0 = \left(R_R^\dagger, R_L^\dagger \right) e^{-i\theta}. \tag{53}$$

Since R is a real vector, the transposed conjugate is equal to the transposed, that is $R^\dagger = R^T$.

Using the Madelung transformation (52) in the Weyl equations (46) and applying the Lie algebra and the Lie group, we can get the following expression

$$\begin{pmatrix} -\sigma^\mu \left(\bar{\nabla}_\mu \theta_R \right) R_R + i\sigma^\mu \left(\bar{\nabla}_\mu R_R \right) - q\sigma^\mu A_\mu R_R \\ -\bar{\sigma}^\mu \left(\nabla_\mu \theta_L \right) R_L + i\bar{\sigma}^\mu \left(\nabla_\mu R_L \right) - q\bar{\sigma}^\mu A_\mu R_L \end{pmatrix} = \begin{pmatrix} mR_L \\ mR_R \end{pmatrix}. \tag{54}$$

These are the Weyl equations in curved space-time with the Madelung transformation. We can also apply the Madelung transformation (52) and (53) to the current density (47), thereby obtaining

$$J^\mu = b \left(R_R^T \bar{\sigma}^\mu R_R - R_L^T \sigma^\mu R_L \right). \tag{55}$$

Its components are

$$J^0 = Nb(R_R^T R_R + R_L^T R_L) = Nbn, \tag{56}$$

$$J^j = b \left(\hat{e}_3^j (n_{\dot{1}} - n_{\dot{2}} - n_{\dot{3}} + n_{\dot{4}}) + 2\hat{e}_1^j (\sqrt{n_{\dot{1}} n_{\dot{2}}} - \sqrt{n_{\dot{3}} n_{\dot{4}}}) + \hat{e}_i^j N^i n \right). \tag{57}$$

We note that the zero component, where $n = \sum_{\dot{\nu}=\dot{1}}^{\dot{4}} n_{\dot{\nu}}$ is the density number of fermions in the system, gives the number of both right- and left-handed particles. We can write the following expressions $|\psi_R|^2 = \psi_R^\dagger \psi_R = R_R^T R_R = n_R$ and $|\psi_L|^2 = \psi_L^\dagger \psi_L = R_L^T R_L = n_L$ for the right- and left-handed spinors, as in the Dirac case. Thus, n_R, n_L are the right- and left- handed particle number and $n = n_R + n_L$ is the total density number.

Furthermore, eq. (31) becomes

$$
\begin{aligned}
i\,[2(mv_R^\mu-\omega_R\delta_0^\mu)\nabla_\mu R_R - qA_\mu + q\nabla_\mu(A^\mu R_R) + \nabla_\mu(mv_R^\mu-\omega_R\delta_0^\mu - qA^\mu)R_R] + \\
\left(m^2 v_{R\mu}v_R^\mu + 2m\omega_R v_R^0 + \frac{\omega^2}{N^2} + m^2\right)R_R - \Box R_R + \\
(2NN^k F_{0k} + i\epsilon^{ij}{}_k\hat{F}_{ij}\tilde{\sigma}^k)R_R + \\
(N(\nabla_0 N) - \bar{\sigma}^j(\nabla_j N))((mv_{R0}-\omega_R)R_R + D_0 R_R) + \\
(N(\nabla_0\sigma^i) - \bar{\sigma}^j(\nabla_j\sigma^i))(imv_{Ri}R_R + D_i R_R) = 0.
\end{aligned}
\tag{58}
$$

A similar equation is obtained for the left-handed spinor R_L with the substitution $R \longrightarrow L$ and $\mathbb{S} \longleftrightarrow \bar{\mathbb{S}}$ in eq. (58). Simplifying the first line in this equation for $\dot{\nu} = 1, 2$ corresponding to right-handed components, we get

$$
\begin{aligned}
i\frac{m}{\sqrt{n_{\dot{\nu}}}}\left[-\frac{\omega_{\dot{\nu}}}{m}\nabla_0 n_{\dot{\nu}} + \nabla_\mu(n_{\dot{\nu}}v_{\dot{\nu}}^\mu) + \frac{\omega_{\dot{\nu}}}{m}\Box t\right] + \\
\sqrt{n_{\dot{\nu}}}\left[m^2 v_{\mu R}v_{\dot{\nu}}^\mu + 2m\omega_{\dot{\nu}}v_{\dot{\nu}}^0 + \frac{\omega_R^2}{N^2} + m^2 - \frac{\Box\sqrt{n_{\dot{\nu}}}}{\sqrt{n_{\dot{\nu}}}}\right] + \\
(2NN^k F_{0k} + i\epsilon^{lj}{}_k\hat{F}_{lj}\tilde{\sigma}^k)R_R - \\
(\hat{\nabla}_a\hat{e}_b^\alpha)\bar{\mathbb{S}}^a\mathbb{S}^b((mv_{R\alpha}-\omega_{\dot{\nu}}\delta_\alpha^0)R_R + D_\alpha R_R) = 0.
\end{aligned}
\tag{59}
$$

The equation for the left-handed components $\dot{\nu} = 3, 4$ is obtained by changing $R_R \longrightarrow R_L$ and $\mathbb{S} \longleftrightarrow \bar{\mathbb{S}}$. Here, we have introduced the two-dimensional vectors $v_\mu = (v_{R\mu}, v_{L\mu})$ and $\omega = (\omega_R, \omega_L)$. Explicitly, they are given by

$$
R = \begin{pmatrix} R_R \\ R_L \end{pmatrix} = \begin{pmatrix} R_{\dot{1}} \\ R_{\dot{2}} \\ R_{\dot{3}} \\ R_{\dot{4}} \end{pmatrix} = \begin{pmatrix} \sqrt{n_{\dot{1}}} \\ \sqrt{n_{\dot{2}}} \\ \sqrt{n_{\dot{3}}} \\ \sqrt{n_{\dot{4}}} \end{pmatrix}, \tag{60}
$$

$$
v_\mu = \begin{pmatrix} v_{R\mu} & 0 \\ 0 & v_{L\mu} \end{pmatrix} = \begin{pmatrix} v_{\mu\dot{1}} & 0 & 0 & 0 \\ 0 & v_{\mu\dot{2}} & 0 & 0 \\ 0 & 0 & v_{\mu\dot{3}} & 0 \\ 0 & 0 & 0 & v_{\mu\dot{4}} \end{pmatrix}, \tag{61}
$$

$$
\omega = \begin{pmatrix} \omega_R & 0 \\ 0 & \omega_L \end{pmatrix} = \begin{pmatrix} \omega_{\dot{1}} & 0 & 0 & 0 \\ 0 & \omega_{\dot{2}} & 0 & 0 \\ 0 & 0 & \omega_{\dot{3}} & 0 \\ 0 & 0 & 0 & \omega_{\dot{4}} \end{pmatrix}. \tag{62}
$$

We now write the last two lines of eq. (59) explicitly and separate them into imaginary and real parts, respectively. Using the Pauli representation of the $\tilde{\sigma}^j$ matrices

$$
\tilde{\sigma}^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tilde{\sigma}^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tilde{\sigma}^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{63}
$$

we obtain for $\dot{\nu} = \dot{1}$:

$$\frac{m}{\sqrt{n_{\dot{1}}}}\left[-\frac{\omega_{\dot{1}}}{m}\nabla_0 n_{\dot{1}} + \nabla_\mu(n_{\dot{1}}v^\mu_{\dot{1}}) + \frac{\omega_{\dot{1}}}{m}\Box t\right] =$$
$$F_{12}\sqrt{n_{\dot{1}}} + F_{23}\sqrt{n_{\dot{2}}} - 2\Gamma^a_{21}((m\hat{v}_{a\dot{1}} - \omega_{\dot{1}}\hat{\delta}^0_a)\sqrt{n_{\dot{1}}} + \hat{D}_a\sqrt{n_{\dot{1}}}) -$$
$$2(\Gamma^a_{21}N^1 - \Gamma^a_{32}N^3 + \Gamma^a_{20} + \Gamma^a_{32})((m\hat{v}_{a\dot{2}} - \omega_{\dot{2}}\hat{\delta}^0_a)\sqrt{n_{\dot{2}}} + \hat{D}_a\sqrt{n_{\dot{2}}}),$$
$$\sqrt{n_{\dot{1}}}\left[m^2 v_{\mu\dot{1}}v^\mu_{\dot{1}} + 2m\omega_{\dot{1}}v^0_{\dot{1}} + \frac{\omega^2_{\dot{1}}}{N^2} + m^2 - \frac{\Box\sqrt{n_{\dot{1}}}}{\sqrt{n_{\dot{1}}}}\right] =$$
$$2N(F_{01}N^1 + F_{02}N^2 + F_{03}N^3)\sqrt{n_{\dot{1}}} - F_{13}\sqrt{n_{\dot{2}}} +$$
$$\left[\Gamma^a_{11}(1-(N^1)^2) + \Gamma^a_{22}(1-(N^2)^2) + \Gamma^a_{33}(1-(N^3)^2) +\right.$$
$$\left.2\Gamma^a_{31}N^1 + 2\Gamma^a_{32}N^2 - \Gamma^a_{00} + 2\Gamma^a_{30}\right]((m\hat{v}_{a\dot{1}} - \omega_{\dot{1}}\hat{\delta}^0_a)\sqrt{n_{\dot{1}}} + \hat{D}_a\sqrt{n_{\dot{1}}}) +$$
$$(-2\Gamma^a_{21}N^2 - 2\Gamma^a_{31}N^3 + 2\Gamma^a_{10} + 2\Gamma^a_{31})((m\hat{v}_{a\dot{2}} - \omega_{\dot{2}}\hat{\delta}^0_a)\sqrt{n_{\dot{2}}} + \hat{D}_a\sqrt{n_{\dot{2}}}), \tag{64}$$

for $\dot{\nu} = \dot{2}$:

$$\frac{m}{\sqrt{n_{\dot{2}}}}\left[-\frac{\omega_{\dot{2}}}{m}\nabla_0 n_{\dot{2}} + \nabla_\mu(n_{\dot{2}}v^\mu_{\dot{2}}) + \frac{\omega_{\dot{2}}}{m}\Box t\right] =$$
$$-F_{12}\sqrt{n_{\dot{2}}} + F_{23}\sqrt{n_{\dot{1}}} + 2\Gamma^a_{21}((m\hat{v}_{a\dot{2}} - \omega_{\dot{2}}\hat{\delta}^0_a)\sqrt{n_{\dot{2}}} + \hat{D}_a\sqrt{n_{\dot{2}}}) +$$
$$2(\Gamma^a_{21}N^1 - \Gamma^a_{32}N^3 + \Gamma^a_{20} - \Gamma^a_{32})(m\hat{v}_{a\dot{1}} - \omega_{\dot{1}}\hat{\delta}^0_a)\sqrt{n_{\dot{1}}} + \hat{D}_a\sqrt{n_{\dot{1}}}),$$
$$\sqrt{n_{\dot{2}}}\left[m^2 v_{\mu\dot{2}}v^\mu_{\dot{2}} + 2m\omega_{\dot{2}}v^0_{\dot{2}} + \frac{\omega^2_{\dot{2}}}{N^2} + m^2 - \frac{\Box\sqrt{n_{\dot{2}}}}{\sqrt{n_{\dot{2}}}}\right] =$$
$$2N(F_{01}N^1 + F_{02}N^2 + F_{03}N^3)\sqrt{n_{\dot{2}}} + F_{13}\sqrt{n_{\dot{1}}} +$$
$$\left[\Gamma^a_{11}(1-(N^1)^2) + \Gamma^a_{22}(1-(N^2)^2) + \Gamma^a_{33}(1-(N^3)^2) +\right.$$
$$\left.-2\Gamma^a_{31}N^1 - 2\Gamma^a_{32}N^2 - \Gamma^a_{00} - 2\Gamma^a_{30}\right](m\hat{v}_{a\dot{2}} - \omega_{\dot{2}}\hat{\delta}^0_a)\sqrt{n_{\dot{2}}} + \hat{D}_a\sqrt{n_{\dot{2}}}) +$$
$$(-2\Gamma^a_{21}N^2 - 2\Gamma^a_{31}N^3 + 2\Gamma^a_{10} - 2\Gamma^a_{31})((m\hat{v}_{a\dot{1}} - \omega_{\dot{1}}\hat{\delta}^0_a)\sqrt{n_{\dot{1}}} + \hat{D}_a\sqrt{n_{\dot{1}}}), \tag{65}$$

for $\dot{\nu} = \dot{3}$:

$$\frac{m}{\sqrt{n_{\dot{3}}}}\left[-\frac{\omega_{\dot{3}}}{m}\nabla_0 n_{\dot{3}} + \nabla_\mu(n_{\dot{3}}v^\mu_{\dot{3}}) + \frac{\omega_{\dot{3}}}{m}\Box t\right] =$$
$$F_{12}\sqrt{n_{\dot{3}}} + F_{23}\sqrt{n_{\dot{4}}} - 2\Gamma^a_{21}((m\hat{v}_{a\dot{3}} - \omega_{\dot{3}}\hat{\delta}^0_a)\sqrt{n_{\dot{3}}} + \hat{D}_a\sqrt{n_{\dot{3}}}) +$$
$$2(\Gamma^a_{21}N^1 - \Gamma^a_{32}N^3 + \Gamma^a_{20} - \Gamma^a_{32})((m\hat{v}_{a\dot{4}} - \omega_{\dot{4}}\hat{\delta}^0_a)\sqrt{n_{\dot{4}}} + \hat{D}_a\sqrt{n_{\dot{4}}}),$$
$$\sqrt{n_{\dot{3}}}\left[m^2 v_{\mu\dot{3}}v^\mu_{\dot{3}} + 2m\omega_{\dot{3}}v^0_{\dot{3}} + \frac{\omega^2_{\dot{3}}}{N^2} + m^2 - \frac{\Box\sqrt{n_{\dot{3}}}}{\sqrt{n_{\dot{3}}}}\right] =$$
$$2N(F_{01}N^1 + F_{02}N^2 + F_{03}N^3)\sqrt{n_{\dot{3}}} - F_{13}\sqrt{n_{\dot{4}}} +$$
$$\left[\Gamma^a_{11}(1-(N^1)^2) + \Gamma^a_{22}(1-(N^2)^2) + \Gamma^a_{33}(1-(N^3)^2) +\right.$$
$$\left.-2\Gamma^a_{31}N^1 - 2\Gamma^a_{32}N^2 - \Gamma^a_{00} - 2\Gamma^a_{30}\right]((m\hat{v}_{a\dot{3}} - \omega_{\dot{3}}\hat{\delta}^0_a)\sqrt{n_{\dot{3}}} + \hat{D}_a\sqrt{n_{\dot{3}}}) +$$
$$(2\Gamma^a_{21}N^2 + 2\Gamma^a_{31}N^3 - 2\Gamma^a_{10} + 2\Gamma^a_{31})((m\hat{v}_{a\dot{4}} - \omega_{\dot{4}}\hat{\delta}^0_a)\sqrt{n_{\dot{4}}} + \hat{D}_a\sqrt{n_{\dot{4}}}), \tag{66}$$

and for $\dot{\nu}=\dot{4}$:

$$\frac{m}{\sqrt{n_{\dot{4}}}}\left[-\frac{\omega_{\dot{4}}}{m}\nabla_0 n_{\dot{4}}+\nabla_\mu(n_{\dot{4}}v^\mu_{\dot{4}})+\frac{\omega_{\dot{4}}}{m}\Box t\right]=$$
$$-F_{12}\sqrt{n_{\dot{4}}}+F_{23}\sqrt{n_{\dot{3}}}+2\Gamma^a_{21}((m\hat{v}_{a\dot{4}}-\omega_{\dot{4}}\hat{\delta}^0_a)\sqrt{n_{\dot{4}}}+\hat{D}_a\sqrt{n_{\dot{4}}})-$$
$$2(\Gamma^a_{21}N^1-\Gamma^a_{32}N^3+\Gamma^a_{20}+\Gamma^a_{32})((m\hat{v}_{a\dot{3}}-\omega_{\dot{3}}\hat{\delta}^0_a)\sqrt{n_{\dot{3}}}+\hat{D}_a\sqrt{n_{\dot{3}}}),$$
$$\sqrt{n_{\dot{4}}}\left[m^2v_{\mu\dot{4}}v^\mu_{\dot{4}}+2m\omega_{\dot{4}}v^0_{\dot{4}}+\frac{\omega^2_{\dot{4}}}{N^2}+m^2-\frac{\Box\sqrt{n_{\dot{4}}}}{\sqrt{n_{\dot{4}}}}\right]=$$
$$2N(F_{01}N^1+F_{02}N^2+F_{03}N^3)\sqrt{n_{\dot{4}}}+F_{13}\sqrt{n_{\dot{3}}}+$$
$$\left[\Gamma^a_{11}(1-(N^1)^2)+\Gamma^a_{22}(1-(N^2)^2)+\Gamma^a_{33}(1-(N^3)^2)\right.+$$
$$\left.2\Gamma^a_{31}N^1+2\Gamma^a_{32}N^2-\Gamma^a_{00}+2\Gamma^a_{30}\right]((m\hat{v}_{a\dot{4}}-\omega_{\dot{4}}\hat{\delta}^0_a)\sqrt{n_{\dot{4}}}+\hat{D}_a\sqrt{n_{\dot{4}}})+$$
$$(2\Gamma^a_{21}N^2+2\Gamma^a_{31}N^3-2\Gamma^a_{10}-2\Gamma^a_{31})((m\hat{v}_{a\dot{3}}-\omega_{\dot{3}}\hat{\delta}^0_a)\sqrt{n_{\dot{3}}}+\hat{D}_a\sqrt{n_{\dot{3}}}),\qquad(67)$$

where we have used that $\Gamma_{ab}+\Gamma_{ba}=0$ and defined the directional quantities $v_{a\dot{\nu}}=v_{R\alpha\dot{\nu}}\hat{e}^\alpha_a$, $\hat{\delta}^0_a=\delta^0_\alpha\hat{e}^\alpha_a=N\delta^0_a$ and $\hat{D}_a=\hat{e}^\alpha_aD_\alpha$.

The first line of eq. (59) represents the continuity equation of the fermionic fluid. The second line is the Bernoulli equation. In this respect, we note that eq. (25) is the first integral of this equation. Finally, the last three lines of eq. (59) are the source of the fermionic fluid, something that is not present in the case of bosons. This is because the Dirac equation was introduced in Ref.[25] in order to eliminate the negative probability problem of the KG equation. As a result, the Dirac equation involves only first derivatives while the KG equation is a second order equation.

Observe that the structure of equations (64)-(67) is

$$\frac{m}{\sqrt{n_{\dot{\nu}}}}\left[-\frac{\omega_{\dot{\nu}}}{m}\nabla_0 n_{\dot{\nu}}+\nabla_\mu(n_{\dot{\nu}}v^\mu_{\dot{\nu}})+\frac{\omega_{\dot{\nu}}}{m}\Box t\right]=$$
$$e_{1\dot{\nu}}F_{12}\sqrt{n_{\dot{\nu}}}+F_{23}\sqrt{n_{\ddot{\nu}}}-2e_{1\dot{\nu}}\Gamma^a_{21}((m\hat{v}_{a\dot{\nu}}-\omega_{\dot{\nu}}\hat{\delta}^0_a)\sqrt{n_{\dot{\nu}}}+\hat{D}_a\sqrt{n_{\dot{\nu}}})-$$
$$2(\Gamma^a_{21}N^1-\Gamma^a_{32}N^3+\Gamma^a_{20}+e_{2\dot{\nu}}\Gamma^a_{32})((m\hat{v}_{a\ddot{\nu}}-\omega_{\ddot{\nu}}\hat{\delta}^0_a)\sqrt{n_{\ddot{\nu}}}+\hat{D}_a\sqrt{n_{\ddot{\nu}}}),$$
$$\sqrt{n_{\dot{\nu}}}\left[m^2v_{\mu\dot{\nu}}v^\mu_{\dot{\nu}}+2m\omega_{\dot{\nu}}v^0_{\dot{\nu}}+\frac{\omega^2_{\dot{\nu}}}{N^2}+m^2-\frac{\Box\sqrt{n_{\dot{\nu}}}}{\sqrt{n_{\dot{\nu}}}}\right]=$$
$$2N(F_{01}N^1+F_{02}N^2+F_{03}N^3)\sqrt{n_{\dot{\nu}}}-e_{1\dot{\nu}}F_{13}\sqrt{n_{\ddot{\nu}}}+$$
$$\left[\Gamma^a_{11}(1-(N^1)^2)+\Gamma^a_{22}(1-(N^2)^2)+\Gamma^a_{33}(1-(N^3)^2)\right.+$$
$$\left.2e_{2\dot{\nu}}(\Gamma^a_{31}N^1+\Gamma^a_{32}N^2+\Gamma^a_{30})-\Gamma^a_{00}\right]((m\hat{v}_{a\dot{\nu}}-\omega_{\dot{\nu}}\hat{\delta}^0_a)\sqrt{n_{\dot{\nu}}}+\hat{D}_a\sqrt{n_{\dot{\nu}}})+$$
$$(-2e_{3\dot{\nu}}(\Gamma^a_{21}N^2+\Gamma^a_{31}N^3-\Gamma^a_{10})+2e_{1\dot{\nu}}\Gamma^a_{31})((m\hat{v}_{a\ddot{\nu}}-\omega_{\ddot{\nu}}\hat{\delta}^0_a)\sqrt{n_{\ddot{\nu}}}+\hat{D}_a\sqrt{n_{\ddot{\nu}}}),\quad(68)$$

where the coefficients $e_{i\dot{\nu}}$ are ±1 with $e_{1\dot{\nu}}=(+,-,+,-)$, $e_{2\dot{\nu}}=(+,-,+,-)$ and $e_{3\dot{\nu}}=(+,+,-,-)$, and the sub-index $\ddot{\nu}$ is the conjugate of the sub-index $\dot{\nu}$ such that $\ddot{1}=\dot{2}$, $\ddot{2}=\dot{1}$, $\ddot{3}=\dot{4}$ and $\ddot{4}=\dot{3}$.

7. Energy Balance

We can also write equation (30) as a Schrödinger-like equation. If we perform the transformation $\psi = \Psi e^{i\omega_0 t}$, where Ψ is a four spinor that depends on all the variables x^μ, equation (30) becomes

$$i\nabla^0\Psi - \frac{1}{2\omega_0}\Box_E\Psi + \frac{m^2}{2\omega_0}\Psi + \left(-\frac{\omega_0}{N^2} - 2qA^0 + i\Box t\right)\Psi + \frac{1}{2\omega_0}\begin{pmatrix} 2NN^kF_{0k} + i\hat{F}_{ij}\epsilon^{ij}{}_k\tilde{\sigma}^k & 0 \\ 0 & -2NN^kF_{0k} + i\hat{F}_{ij}\epsilon^{ij}{}_k\tilde{\sigma}^k \end{pmatrix}\Psi - \frac{1}{2\omega_0}\begin{pmatrix} \bar{\mathbb{S}}^a\mathbb{S}^b & 0 \\ 0 & \mathbb{S}^a\bar{\mathbb{S}}^b \end{pmatrix}\Gamma^d_{ba}(\hat{D}_d\Psi + i\omega_0 N\delta^0_d\Psi) = 0. \tag{69}$$

Equation (69) is the generalization of the Schrödinger equation for fermions with electromagnetic field interaction in an arbitrary space-time.

Finally, from equation (59) we can identify the different energy contributions to the Fermi gas, and obtain an energy balance equation for fermions analogous to the one obtained for bosons in Refs.[1,2]. In order to simplify the notations, we can re-write equation (59) in terms of the $\dot{\nu}$ coefficients with the understanding that the subindex R refers to each component $R = \dot{1}, \dot{2}$ individually. We get

$$i\left[-\omega_{\dot{\nu}}\nabla_0\ln(n_{\dot{\nu}}) + \frac{m\nabla_\mu(n_{\dot{\nu}}v^\mu_{\dot{\nu}})}{n_{\dot{\nu}}} + \frac{\omega_{\dot{\nu}}}{n_{\dot{\nu}}}\Box t\right] + 2m^2\left(K_{\dot{\nu}} + \frac{1}{m}\omega_{\dot{\nu}}v^0_{\dot{\nu}} + \frac{1}{2}U^N_{\dot{\nu}} + U^Q_{\dot{\nu}}\right) + E_{\dot{\nu}} + U^S_{\dot{\nu}} = 0. \tag{70}$$

This equation is valid for right handed fermions. The result is the same for left handed fermions changing $R_R \longrightarrow R_L$ in the first line, and $\mathbb{S} \longleftrightarrow \bar{\mathbb{S}}$ in the second line.

The first line in eq. (70) describes the free density evolution of the fermions, while the contribution of the different energy terms appears in the second line. The first one is the kinetic energy $K_{\dot{\nu}}$ defined as

$$K_{\dot{\nu}} = \frac{1}{2}v_{\dot{\nu}\mu}v^\mu_{\dot{\nu}}. \tag{71}$$

The lapse potential $U^N_{\dot{\nu}}$ is given by

$$U^N_{\dot{\nu}} = \frac{\omega^2_{\dot{\nu}}}{m^2}\frac{1}{N^2} + 1. \tag{72}$$

It represents the energy contribution due to the chosen lapse function N. The quantum potential $U^Q_{\dot{\nu}}$ is defined as

$$U^Q_{\dot{\nu}} = -\frac{1}{2m^2}\frac{\Box\sqrt{n_{\dot{\nu}}}}{\sqrt{n_{\dot{\nu}}}}. \tag{73}$$

The contribution of the electromagnetic interaction $E_{\dot\nu}$ is given by

$$\begin{aligned} E_{\dot\nu} &= (2NN^k F_{0k} + i\epsilon^{lj}{}_k \hat{F}_{lj}\tilde{\sigma}^k)|_{\dot\nu}, \\ &= 2N(F_{01}N^1 + F_{02}N^2 + F_{03}N^3) - e_{1\dot\nu}F_{13}\sqrt{\frac{n_{\ddot\nu}}{n_{\dot\nu}}} + i\left(e_{1\dot\nu}F_{12} + F_{23}\sqrt{\frac{n_{\ddot\nu}}{n_{\dot\nu}}}\right). \end{aligned} \tag{74}$$

It depends on the Faraday tensor, shift vector and lapse function that are related to the Pauli matrices. This relationship is due to the interaction between the electromagnetic field and the fermionic spin. Finally, the potential $U^S_{\dot\nu}$ describes the interaction between the spin and the geometry of space-time. It is given by

$$\begin{aligned} U^S_{\dot\nu} &= -\left((m\hat{v}_{Rd} - \omega_{\dot\nu}\hat{\delta}^0_d) + \frac{\hat{D}_\alpha\sqrt{n_{\dot\nu}}}{\sqrt{n_{\dot\nu}}}\right)\Gamma^d_{ba}\bar{\mathbb{S}}^a\mathbb{S}^b|_{\dot\nu}, \\ &= \left[\Gamma^a_{11}(1-(N^1)^2) + \Gamma^a_{22}(1-(N^2)^2) + \Gamma^a_{33}(1-(N^3)^2)\right. \\ &\quad \left. + 2e_{2\dot\nu}(\Gamma^a_{31}N^1 + \Gamma^a_{32}N^2 + \Gamma^a_{30}) - \Gamma^a_{00}\right]\left((m\hat{v}_{a\dot\nu} - \omega_{\dot\nu}\hat{\delta}^0_a) + \frac{\hat{D}_a\sqrt{n_{\dot\nu}}}{\sqrt{n_{\dot\nu}}}\right) \\ &\quad + (-2e_{3\dot\nu}(\Gamma^a_{21}N^2 + \Gamma^a_{31}N^3 - \Gamma^a_{10}) + 2e_{1\dot\nu}\Gamma^a_{31})\left((m\hat{v}_{a\ddot\nu} - \omega_{\ddot\nu}\hat{\delta}^0_a)\sqrt{\frac{n_{\ddot\nu}}{n_{\dot\nu}}} + \frac{\hat{D}_a\sqrt{n_{\ddot\nu}}}{\sqrt{n_{\dot\nu}}}\right) \\ &\quad + i\left[-2e_{1\dot\nu}\Gamma^a_{21}\left((m\hat{v}_{a\dot\nu} - \omega_{\dot\nu}\hat{\delta}^0_a) + \frac{\hat{D}_a\sqrt{n_{\dot\nu}}}{\sqrt{n_{\dot\nu}}}\right)\right. \\ &\quad \left. - 2(\Gamma^a_{21}N^1 - \Gamma^a_{32}N^3 + \Gamma^a_{20} + e_{2\dot\nu}\Gamma^a_{32})\left((m\hat{v}_{a\ddot\nu} - \omega_{\ddot\nu}\hat{\delta}^0_a)\sqrt{\frac{n_{\ddot\nu}}{n_{\dot\nu}}} + \frac{\hat{D}_a\sqrt{n_{\ddot\nu}}}{\sqrt{n_{\dot\nu}}}\right)\right] \end{aligned} \tag{75}$$

for $\dot\nu = \dot1, \dot2$, and by making the substitution $\mathbb{S} \longleftrightarrow \bar{\mathbb{S}}$ for $\dot\nu = \dot3, \dot4$. In the foregoing equations, the notation $|_{\dot\nu}$ means that we have to evaluate the quantity at the corresponding $\dot\nu$. Note that $U^S_{\dot\nu}$ disappears if we assume a flat space-time or if we consider particles without spin. Furthermore, $U^S_{\dot\nu}$ is constructed with the generalized gamma matrices (45), which are related to the spin (the Pauli matrices) and to the space-time geometry (tetrads).

8. Conclusions

The main difference between the hydrodynamic representation of bosons[1,2] and fermions concerns the form of the Bernoulli equation. For bosons, after making the Madelung transformation, we can separate the KG equation into real and imaginary parts. By contrast, for fermion particles we have to work with the complete equations of motion because the real and imaginary parts cannot be separated easily. This is related to the fact that the gamma matrices are a representation of the SO(1, 3) group.

The spin is a fundamental outcome of the Dirac equation[25] which combines both elements of special relativity and quantum mechanics and was introduced to solve the problem of negative probability present in the KG equation – first proposed as a relativistic generalization of the Schrödinger equation. Here, we observe that the general relativistic Dirac equation involves an additional contribution due to geometry and spin through the generalized gamma and Pauli matrices. These terms arise from endowing a quantum field with a curvature (geometry) given by a metric in General Relativity. Such a contribution is absent in a flat space-time and in a system without spin as for a scalar field.

With this work we open the possibility of studying in detail the behavior of fermions in different situations (such as massive stars or dark matter halos harboring a central black hole) where general relativity effects may be important. We solved the problem of energy balance for both bosons and fermions. In this manner, we can compare the result of the hydrodynamic representation for classical and quantum fluids in the various geometries mentioned above.

Acknowledgments

This work was partially supported by CONACyT México under grants: A1-S-8742, 304001, 376127; Project No. 269652 and Fronteras Project 281;Xiuhcoatl and Abacus clusters at Cinvestav, IPN; I0101/131/07 C-234/07 of the Instituto Avanzado de Cosmología (IAC) collaboration (http:// www.iac.edu.mx). O.G. acknowledges financial support from CONACyT doctoral fellowship and appreaciates Angelica C. Aguirre Castañón for her valuable review and support. Works of T.M. are partially supported by Conacyt through the Fondo Sectorial de Investigación para la Educación, grant CB-2014-1, No. 240512.

Appendix A. Solutions to the Dirac equation in flat space-time

Equation (5) in flat space-time, using the Pauli matrices (63), reads

$$\begin{bmatrix} \frac{\partial}{\partial t}\psi_y - \frac{\partial}{\partial x}\psi_z + i\frac{\partial}{\partial y}\psi_z - \frac{\partial}{\partial z}\psi_y - m\psi_t \\ \frac{\partial}{\partial t}\psi_z - \frac{\partial}{\partial x}\psi_y - i\frac{\partial}{\partial y}\psi_y + \frac{\partial}{\partial z}\psi_z - m\psi_x \\ \frac{\partial}{\partial t}\psi_t + \frac{\partial}{\partial x}\psi_x - i\frac{\partial}{\partial y}\psi_x + \frac{\partial}{\partial z}\psi_t - m\psi_y \\ \frac{\partial}{\partial t}\psi_x + \frac{\partial}{\partial x}\psi_t + i\frac{\partial}{\partial y}\psi_t - \frac{\partial}{\partial z}\psi_x - m\psi_z \end{bmatrix} = 0, \tag{A.1}$$

where we have defined the spinor as $\psi = (\psi_{\dot{\mu}}) = (\psi_x, \psi_y, \psi_z, \psi_t)^T$. In order to find an exact solution of the previous equation, we use the ansatz $\psi_{\dot{\mu}} = R_{0\dot{\mu}} \exp(i(x_0 x + y_0 y + z_0 z + t_0 t))$, where $x_0 \cdots t_0$ and $R_{0\dot{\mu}}$ are constants. Here we have the simplest solutions of the Dirac equation where the exponential is the same for all components.

We obtain four linear equations

$$\begin{aligned} iR_{0z}\zeta_0^* + iR_{0y}\eta_0 + mR_{0t} &= 0, \\ iR_{0y}\zeta_0 - iR_{0z}\xi_0 + mR_{0x} &= 0, \\ R_{0x}\zeta_0^* + R_{0t}\xi_0 + imR_{0y} &= 0, \\ R_{0t}\zeta_0 - R_{0x}\eta_0 + imR_{0z} &= 0, \end{aligned} \tag{A.2}$$

where $\zeta_0 = x_0 + iy_0$, $\eta_0 = z_0 - t_0$, and $\xi_0 = z_0 + t_0$. The solution of these equations is

$$\begin{aligned} R_{0t} &= -\frac{1}{m}(iR_{0y}\eta_0 + iR_{0z}\zeta_0), \\ R_{0x} &= \frac{1}{m}(iR_{0z}\xi_0 - iR_{0y}\zeta_0^*), \end{aligned} \tag{A.3}$$

where $x_0^2 + y_0^2 + z_0^2 - t_0^2 = m^2$.

Now we use the ansatz $\psi_\mu = R_{0\mu}\exp(i\theta)$, where θ is an arbitrary function of the coordinates. Substituting this ansatz into (A.1) we obtain

$$\begin{aligned} iR_{0z}Z_0^* + iR_{0y}E_0 + mR_{0t} &= 0, \\ iR_{0y}Z_0 - iR_{0z}F_0 + mR_{0x} &= 0, \\ R_{0x}Z_0^* + R_{0t}F_0 + imR_{0y} &= 0, \\ R_{0t}Z_0 - R_{0x}E_0 + imR_{0z} &= 0, \end{aligned} \tag{A.4}$$

where $Z_0 = \theta_{,x} + i\theta_{,y}$, $E_0 = \theta_{,z} - \theta_{,t}$, and $F_0 = \theta_{,z} + \theta_{,t}$. The solution of the previous system of differential equations is

$$\begin{aligned} \theta &= F(X) - it \\ &\quad + \frac{m}{2R_{0t}R_{0z} + 2R_{0x}R_{0y}}\left(i\zeta_0^*(R_{0x}^2 - R_{0z}^2) - i\zeta_0(R_{0y}^2 - R_{0t}^2)\right), \end{aligned} \tag{A.5}$$

where $F(X)$ is an arbitrary function of

$$X = \frac{R_{0t}(-\zeta R_{0y} - \zeta^* R_{0x} + \xi R_{0y} - \eta R_{0z})}{2R_{0t}R_{0z} + 2R_{0x}R_{0y}}. \tag{A.6}$$

References

1. T. Matos, A. Avilez, T. Bernal, P.H. Chavanis, Gen. Rel. Grav. **51**, 159 (2019).
2. P.H. Chavanis, T. Matos, Eur. Phys. J. Plus **132**, 30 (2017).
3. P. Collas, D. Klein, *The Dirac Equation in Curved Spacetime: A Guide for Calculations* (Springer Briefs in Physics, Springer, 2019).
4. R.L. Arnowitt, S. Deser, C.W. Misner, Gen. Rel. Grav. **40**, 1997 (2008).
5. M. Alcubierre, *Introduction to* 3+1 *Numerical Relativity* (International Series of Monographs on Physics, Oxford, 2008).
6. S.W. Hawking, G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2011).
7. M. Nakahara, *Geometry, topology and physics* (IOP Publishing, 2003).
8. O. Gallegos, T. Matos, Gen. Rel. Grav. **53**, 50 (2021).

9. R. Penrose, W. Rindler, *Spinors and Space-time. Vol. 2: Spinor and Twistor Methods in Space-time Geometry* (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1988).
10. R. M. Wald, *General Relativity* (The University of Chicago Press, 1984).
11. M. Arminjon, F. Reifler, Braz. J. Phys. **40**, 242 (2010).
12. M. Arminjon, F. Reifler, Braz. J. Phys. **43**, 64 (2013).
13. M. Arminjon, F. Reifler, Int. J. Geom. Meth. Mod. Phys. **09**, 1250026 (2012).
14. N.J. Poplawski, arXiv:0710.3982.
15. S. Weinberg, *The Quantum Theory of Fields* (Press Syndicate of the University of Cambridge, 1995).
16. I.J.R. Aitchison, A.J.G. Hey, *Gauge Theories in Particles Physics Volume 1: From Relativistic Quantum Mechanics to QED* (CRC Press, 2003).
17. T. Lancaster, S.J. Blundell, *Quantum Field Theory for the Gifted Amateur* (Oxford University Press, 2014).
18. N.D. Birrell, P.C.W. Davies, *Quantum Fields in Curved Space* (Cambridge Monographs on Mathematical Physics, Cambridge Univ. Press 1984).
19. M.D. Pollock, Acta Physica Polonica B **41**, 8 (2010).
20. E. Noether, Nachr. d. Kgl. Ges d. Wiss (Math. phys. Klasse), Göttingen, 235 (1918).
21. D.R. Brill, J.A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).
22. L. Parker, Phys. Rev. D **22**, 1922 (1980).
23. E. Madelung, Zeitschrift fr̈ Physik **40**, 322 (1927).
24. Y-Q Gu, gr-qc/0610001.
25. P.A.M. Dirac, Proc. Roy. Soc. Lond. A **117**, 610 (1928).

Breakdown of the Equivalence Principle for a composite quantum body

A. G. Lebed

Department of Physics, University of Arizona, Tucson, Arizona 85721, USA and
L.D. Landau Institute for Theoretical Physics, Kosygina Street 2, Moscow 117334, Russia
E-mail: lebed@arizona.edu
www.university of arizona.edu

We derive the semiclassical Hamiltonian for electron in a hydrogen atom in a weak gravitational field, which takes into account the quantum effects of electron motion in the atom. We show that this Hamiltonian predicts a breakdown of the equivalence between electron passive gravitational mass and its energy in the absence of the gravitational field. More strictly, we demonstrate that quantum measurement of electron mass, which initially is in a ground state and has energy E_1, can give the following quantized values: $m_n = m_e + \frac{E_n}{c^2}$, where E_n is energy of electron orbit nS in a hydrogen atom. We correct some drawbacks of our early pioneering papers and discuss some difficulties in the possible observations of the above mentioned mass quantization phenomenon.

Keywords: Equivalence Principle; Quantum gavity; Equivalence betweem mass and energy

1. Introduction

Einstein's Equivalence Principle (EEP) between gravitational and inertial masses is known to be a keystone of the classical General Relativity. In the current scientific literature, there exists a wide discussion if it can survive in the possible quantum theory of gravity. Since the quantum gravitation theory has not been elaborated yet, the EEP is often studied in framework of the so-called semiclassical approach to quantum gravity, where gravitational field is not quantized, but the matter is quantized.[1–5] Note that gravitational mass of a composite body is not a trivial notion even in classical General Relativity.[6–8] The reason for this is that gravitational field is not directly coupled to the total energy, E, but is coupled to combination, $E + V(\mathbf{r})$, where $V(\mathbf{r})$ is the so-called virial term.[6–8] In quantum case, as shown by us,[1,2,4,5] the existence of the virial term results in breakdown of the equivalence between active and passive gravitational masses and energy of a body with internal degrees of freedom. The goal of our current paper is to suggest new method which demonstrates that the virial term is responsible for inequivalence of passive gravitational mass and energy. Below, we demonstrate the above mentioned inequivalence for a hydrogen atom, which is the simplest quantum object with internal degrees of freedom. Of course our conclusions, obtained in the paper, are applied to any composite quantum body.

2. Quantization of Passive Electron Gravitational Mass

2.1. *Electron wave function with a definite energy in the absence of gravitational field*

Let us consider the case, where at $t < 0$ gravitational field is zero and electron occupies ground state in a hydrogen atom, characterizing by the wave function:

$$\Psi_1(r,t) = \exp\left(\frac{-im_e c^2 t}{\hbar}\right)\exp\left(\frac{-iE_1 t}{\hbar}\right)\Psi_1(r)\ . \tag{1}$$

Note that the wave function (1) corresponds to the $1S$ electron orbit and is known to be a ground state solution of the following Schrödinger equation:

$$i\hbar\frac{\partial\Psi_1(r,t)}{\partial t} = \left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right) - \frac{e^2}{r}\right]\Psi_1(r,t). \tag{2}$$

[Here E_1 is electron ground state energy, r is a distance between the electron with coordinates (x,y,z) and proton; m_e is the bare electron mass, $\hbar$ is the Planck constant, and c is the velocity of light.]

2.2. *Electron wave function in the presence of gravitational field*

Let us consider the following Gedanken experiment: we turn on a weak (e.g., the Earth's) gravitational field at $t = 0$. As known,[9] the interval, characterizing a curved spacetime in the so-called weak field approximation, can be written as

$$ds^2 = -\left(1+2\frac{\phi}{c^2}\right)(cdt)^2 + \left(1-2\frac{\phi}{c^2}\right)(dx^2+dy^2+dz^2),\ \phi = -\frac{GM}{R}, \tag{3}$$

where M is mass of a source of gravity, G is the gravitational constant. Below we use the local Lorentz invariance property of a spacetime in general relativity.[9] To this end, it is convenient to introduce the proper local coordinates,

$$\tilde{t} = \left(1+\frac{\phi}{c^2}\right)t,\ \tilde{x} = \left(1-\frac{\phi}{c^2}\right)x,\ \tilde{y} = \left(1-\frac{\phi}{c^2}\right)y,\ \tilde{z} = \left(1-\frac{\phi}{c^2}\right)z, \tag{4}$$

where the interval (3) has the Minkowski's form,

$$d\tilde{s}^2 = -(cd\tilde{t})^2 + (d\tilde{x}^2+d\tilde{y}^2+d\tilde{z}^2). \tag{5}$$

Due to the local Lorentz invariance, if we disregard the tidal terms (i.e., if we don't differentiate the gravitational potential ϕ), the wave functions, written in the local proper coordinates (4), obeys the same Schrödinger equation at $t, \tilde{t} > 0$:

$$i\hbar\frac{\partial\Psi(\tilde{\mathbf{r}},\tilde{t})}{\partial\tilde{t}} = \left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial\tilde{x}^2}+\frac{\partial^2}{\partial\tilde{y}^2}+\frac{\partial^2}{\partial\tilde{z}^2}\right) - \frac{e^2}{\tilde{r}}\right]\Psi(\tilde{\mathbf{r}},\tilde{t}). \tag{6}$$

Nevertheless, it is important that the wave function (1) is not a solution of the Schrödinger equation (6) and, thus, is not anymore characterized by definite energy and weight in the gravitational field (3). To clarify this point, we rewrite Eq. (6) in

initial, (t, x, y, z), spacetime coordinates:

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = \left\{\left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}\right]\right.$$
$$\left.+\left(\frac{\phi}{c^2}\right)\left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}+\hat{V}(\mathbf{r})\right]\right\}\Psi(\mathbf{r},t), \quad (7)$$

where the virial operator[10] is equal to

$$\hat{V}(\mathbf{r}) = -2\frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}. \quad (8)$$

From Eq. (8), it directly follows that the external gravitational field (3) is coupled not only to Hamiltonian (2) but also to the virial operator (8). Note that the Hamiltonian (7),(8) can be also obtained from the results,[11] where some different physical problem is considered, if we disregard all tidal terms.

It is important that the gravitational field (3) can be considered as a sudden perturbation, $\hat{H}_1(\mathbf{r},t)$, to Hamiltonian (2) in the absence of gravitational field:

$$\hat{H}_1(\mathbf{r},t) = \left(\frac{\phi}{c^2}\right)\left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}+\hat{V}(\mathbf{r})\right]\Theta(t), \quad (9)$$

where $\Theta(t)$ is the step-like function. Then, a general solution of Eq. (7) can be written in the following way:

$$\Psi(r,t) = \exp\left(\frac{-i\tilde{m}_e c^2 t}{\hbar}\right)\Psi_1^1(r)\exp\left(\frac{-i\tilde{E}_1 t}{\hbar}\right)$$
$$+\exp\left(\frac{-im_e c^2 t}{\hbar}\right)\sum_{n>1}^{\infty} a_n\Psi_n(r)\exp\left(\frac{-iE_n t}{\hbar}\right), \quad (10)$$

where the wave functions $\Psi_n(r)$ are solutions for the nS orbits in a hydrogen atom and are normalized,

$$\int[\Psi_1^1(r)]^2\, d^3r \simeq 1; \quad \int\Psi_n^2(r)\, d^3r = 1, \quad n>1. \quad (11)$$

[It is easy to show that perturbation (9) can results only in non-zero quantum transitions between $1S$ and nS electron orbits, therefore, we keep in Eq. (10) only $\Psi_n(r)$ wave functions, which are real.]

2.3. *Matrix elements and probabilities*

According with the standard time-dependent perturbation theory,[10] the corrected wave-function of ground state, $\Psi_1^1(r)$, as well as the corrections to mass and energy of ground state in Eq. (10) can be written as:

$$\Psi_1^1(r) = \Psi_1(r) + \left(\frac{\phi}{c^2}\right)\sum_{n>1}^{\infty}\frac{V_{n,1}}{E_1 - E_n}\Psi_n(r),$$
$$\tilde{m}_e = \left(1+\frac{\phi}{c^2}\right)m_e, \quad \tilde{E}_1 = \left(1+\frac{\phi}{c^2}\right)E_1, \quad (12)$$

where $V_{n,1}$ is matrix element of the virial operator (8):

$$V_{n,1} = \int \Psi_n(r)\left[-2\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}\right]\Psi_1(r)d^3\mathbf{r}. \tag{13}$$

Note that the very last term in Eq. (12) corresponds to the so-called red shift in gravitational field. It is due to the expected contribution to passive gravitational mass from electron binding energy in the atom. As to the coefficients a_n with $n \neq 1$ in Eq. (10), they can be also written in terms of the virial operator matrix elements,

$$a_n = -\left(\frac{\phi}{c^2}\right)\left(\frac{V_{n,1}}{E_1 - E_n}\right). \tag{14}$$

It is important that the wave function (10)-(14), which corresponds to electron energy levels in the presence of the gravitational field (3) at $t > 0$, is a series of eigenfunctions of electron energy operator, taken in the absence of the field. Therefore, if we measure energy, in electron quantum state (10)-(14), we obtain the following quantized values for electron gravitational mass:

$$m_n = m_e + \frac{E_n}{c^2}, \tag{15}$$

where we omit the red shift effect. From Eqs.(10)-(15), we can state that the expected Einstein's equation, $m_1 = m_e + \frac{E_1}{c^2}$, survives in our case with probability close to 1, whereas with the following small probabilities,

$$P_n = |a_n|^2 = \left(\frac{\phi}{c^2}\right)^2 \frac{V_{n,1}^2}{(E_n - E_1)^2}, \quad n \neq 1, \tag{16}$$

it is broken. The reason for this breakdown is that, the virial term (8) does not commute with the Hamiltonian in the absence of gravitational field (2). As a result, electron wave functions with definite passive gravitational masses are not characterized by definite energies in the absence of gravitational field. Let us calculate probability P_2 near the Earth's surface, where $\frac{\phi^2}{c^4} \approx 0.49 \times 10^{-18}$. From Eq. (16), we obtain the following probability for $n = 2$ in a hydrogen atom:

$$P_2 = |a_2|^2 = 1.5 \times 10^{-19}, \tag{17}$$

where

$$\frac{V_{2,1}}{E_2 - E_1} = 0.56. \tag{18}$$

3. Experimental Aspects

Here, let us describe another Gedanken experiment, where gravitational field is adiabatically switched on. To this end, we consider wave function (1) to be valid at $t \to -\infty$ and apply the following perturbation due to gravitation for the Hamiltonian (2):

$$\hat{H}_2(\mathbf{r},t) = \left(\frac{\phi}{c^2}\right)\left[m_e c^2 - \frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{r}+\hat{V}(\mathbf{r})\right]\exp(\lambda t),\ \lambda \to 0. \tag{19}$$

Then, at $t \simeq 0$ (i.e., in the presence of the field), the electron wave function can be written as

$$\Psi(r,t) = \exp\left(\frac{-i\tilde{m}_e c^2 t}{\hbar}\right)\Psi_1^1(r)\exp\left(\frac{-i\tilde{E}_1 t}{\hbar}\right)$$
$$+\exp\left(\frac{-im_e c^2 t}{\hbar}\right)\sum_{n>1}^{\infty}\tilde{a}_n\Psi_n(r)\exp\left(\frac{-iE_n t}{\hbar}\right). \qquad (20)$$

Application of the standard time-dependent perturbation theory[10] in the case of adiabatic switching on gravitational field gives:

$$\Psi_1^1(r) = \Psi_1(r) + \left(\frac{\phi}{c^2}\right)\sum_{n>1}^{\infty}\frac{V_{n,1}}{E_1 - E_n}\Psi_n(r),$$
$$\tilde{m}_e = \left(1+\frac{\phi}{c^2}\right)m_e, \quad \tilde{E}_1 = \left(1+\frac{\phi}{c^2}\right)E_1, \qquad (21)$$

and

$$\tilde{a}_n = 0, \quad \tilde{P}_n = 0. \qquad (22)$$

Thus, in adiabatic limit, the phenomenon of quantization of passive gravitational mass (15),(16) disappears. This means that the possible experimental observations of the above mentioned phenomenon has to be done in quickly changing gravitational field. It is important that step-like function, $\Theta(t)$, which was used to derive Eqs.(15),(16) does not mean motion of an atom or a source of gravity with velocity higher than the speed of light. We can use step-like function if significant change of gravitational field happens quicker that the period of quasiclassical rotation of electron in a hydrogen atom. In the case under consideration, we need the time about $\delta t \leq t_0 = \frac{2\pi\hbar}{E_2 - E_1} \sim 10^{-15}s$. We hope that there exist much more convenient quantum systems with higher values of the parameter t_0, where discussed above phenomenon could be observed. We recall that all excited energy levels are quasistationary and, thus, decay with time by emitting photons. Therefore, it is much more efficient to detect emitted photons than to directly measure a weight. Here, we also stress that we did not calculate the gravitational mass and weight of the atom with an accuracy ϕ^2/c^4 since we did not take in account terms of the order of ϕ^2/c^4 in the metric (3). As to the relatively small probability of the quantization (16), it is not too small and can be compensated by large value of Avogadro number, $N_A \approx 6 \times 10^{23}$. In other words, for macroscopic number of the atoms, we may have large number of emitted photons. For instance, the number of excited electrons (i.e., emitted photons) for 1000 moles of the atoms is estimated as

$$N_n = 2.95 \times 10^8 \times \left(\frac{V_{n,1}}{E_n - E_1}\right)^2, \quad N_2 = 0.9 \times 10^8. \qquad (23)$$

4. Conclusion

In the paper, we have shown that energy is not equivalent to passive gravitational mass of the such simplest composite body as a hydrogen atom. This happens due to the existence in the electron Hamiltonian in gravitational field of the so-called virial term, which does not commute with the Hamiltonian, taken in the absence of the field. It is easy to show that this is a general property of any quantum system with internal degrees of freedom in General Relativity. Therefore, we can conclude that application of the basic quantum laws to General Relativity results in the breakdown of the equivalence between passive gravitational mass and energy. It is important that the non-zero matrix elements (16),(18) of the virial operator (8) are also responsible, as shown by us,[1,2,4] for breakdown of the equivalence between active gravitational mass and energy for a quantum body with internal degrees of freedom. Therefore, formulation of the EEP has to be significantly reinvestigated and changed in semiclassical gravity.

Acknowledgments

We are thankful to N.N. Bagmet (Lebed), V.A. Belinski, Steven Carlip, Fulvio Melia, Douglas Singleton, and V.E. Zakharov for fruitful and useful discussions.

References

1. A. G. Lebed, Breakdown of the equivalence between active gravitational mass and energy for a quantum body, *J. Phys.: Conf. Ser.* **738**, 012036 (2016).
2. A. G. Lebed, Inequivalence between gravitational mass and energy due to quantum effects at microscopic and macroscopic levels, *Int. J. of Mod. Phys. D* **26**, 1730022 (2017).
3. M. Zych and C. Brukner, Quantum formulation of the Einstein equivalence principle, *Nature Phys.* **14**, 1027 (2018).
4. A. G. Lebed, Breakdown of the equivalence between gravitational mass and energy due to quantum effects, *Int. J. of Mod. Phys. D* **28**, 1930020 (2019).
5. A. G. Lebed, Breakdown of the Einstein's Equivalence Principle for a quantum body, *Mod. Phys. Lett. A* **35**, 2030010 (2020).
6. C. W. Misner and P. Putnam, Active gravitational mass, *Phys. Rev.* **116**, 1045 (1959).
7. K. Nordtvedt, Post-Newtonian gravity: Its theory-experimental interface, *Class. Quantum Grav.* **11**, A119 (1994).
8. S. Carlip, Kinetic energy and the equivalence principle, *Am. J. Phys.* **66**, 409 (1998).
9. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W.H. Freeman and Co, New York, USA, 1973).
10. D. Park, *Introduction to the Quantum Theory*, 3rd edn. (Dover Publications, New York, USA, 2005).
11. E. Fischbach, B. S. Freeman, and W.-K. Cheng, General-relativistic effects in hydrogen systems, *Phys. Rev. D* **23**, 2157 (1981).

Extended DeWitt-Schwinger subtraction scheme, heavy fields and decoupling*

Antonio Ferreiro

Centre for Astrophysics and Relativity, School of Mathematical Sciences, Dublin City University, Glasnevin, Dublin 9, Rep. of Ireland
E-mail: antonio.ferreiro@dcu.ie

Jose Navarro-Salas

Departamento de Física Teórica and IFIC, Centro Mixto Universidad de Valencia-CSIC Facultad de Física, Universidad de Valencia, Burjassot-46100, Valencia, Spain
E-mail: jnavarro@ific.uv.es

Keywords: Renormalization techniques; quantum field theories.

1. Introduction

A standard result of quantum field theory is the existence of divergences when computing observables. This requires the construction of techniques, called regularization and renormalization schemes, to overcome these infinities and produce well defined physical observables. A standard example is the computation of vacuum polarization effects in perturbative Quantum Electrodynamics (QED) which results in the well-known Lamb Shift an the anomalous magnetic momentum of the electron. As a byproduct of these methods, an arbitrary mass scale parameter μ present in the calculations implies a dependence under this scale of the different coupling constants of the theory.[1] Instead of becoming an obstacle it offers a very useful tool, the renormalization group equation,[1,2] to understand the behaviour of the theories at some particular ranges, without knowing the exact details of the theory. The relevant magnitudes that encode the information of this equation are the beta functions, $\beta_O = \mu \frac{dO_R}{d\mu}$, where O_R is any possible renormalized coupling constant of the theory.

At the level of quantum field theory in curved spacetime, similar divergences arise. These appear even when computing observables in the vacuum state since the standard method of regularizing the vacuum divergences in Minkowski, known as *normal ordering*, is not available anymore since it strictly depends on the definition of the vacuum state, which is not unique in this context. Nevertheless, many regularization methods, including Dimensional Regularization and Pauli-Villars have been inherit from the Minkowski spacetime case, and have obtained finite renormalized

*Talk given by A. Ferreiro at the Sixteenth Marcel Grossmann Meeting (2021).

quantities, e.g. the Feynman propagator $G(x,x')$ and the expectation value of the stress energy tensor T_{ab} for scalar, Dirac or gauge fields in curved spacetime.[3–5]

Most methods rely on the convenient expression of the propagator in terms of an integral in the DeWitt-Schwinger proper time

$$G(x,x') = \frac{\Delta^{1/2}(x,x')}{(4\pi)^2}\int_0^\infty \frac{ds}{(is)^2} e^{-im^2 s + \frac{\sigma(x,x')}{2is}} F(x,x',is)\ . \tag{1}$$

The advantage of expression (1) is that it admits an expansion

$$F(x,x',is) = \sum_{j=0}^{\infty} a_j(x,x')(is)^j. \tag{2}$$

Note that the effective action and consequently the stress-energy tensor can be derived from $G(x,x')$ in the limit $x' \to x$ where this expression becomes divergent. In the case of four dimensions $d = 4$, the divergences are encoded and isolated in the first two terms of the expansion (the first three terms in case of the effective action). There are several ways to regulate expression (1). For instance, it can be analytically extended to higher dimension $d > 4$, via dimensional regularization or it can be expanded in terms of point splitting parameter σ. The infinities are recovered in the case of $d \to 4$ and $\sigma \to 0$ respectively. It can be shown[3] that all the divergent pieces of the one loop corrections of a free field can be consistently reabsorbed in the original action

$$S_G = \int d^4x\sqrt{-g}\left(-\Lambda + \frac{1}{2\kappa}R + \alpha_1 C^2 + \alpha_2 R^2 + \alpha_3 E + \alpha_4 \Box R\right), \tag{3}$$

where $\kappa = 8\pi G$. It is in principle possible to use the DeWitt-Schwinger decomposition (1) to compute the exact one-loop contribution.[3,5] In this case, as a byproduct of the DeWitt-Schwinger decomposition we can construct a natural subtraction scheme without invoking any artificial regulator

$$G_{\rm ren}(x,x) = \lim_{x'\to x} G(x,x') - \frac{1}{(4\pi)^2}\sum_{j=0}^{1} a_j(x)\int_0^\infty \frac{ds}{(is)^{(2-j)}} e^{-im^2 s}\ . \tag{4}$$

We can also construct subtraction schemes from dimensional regularization or point splitting. In any case the different regularization and subtraction schemes differ in terms that can always be reabsorbed in the action (3).[4]

Note that the subtracted contribution of (4) needs to include an small imaginary $m^2 - i\epsilon$ to avoid an infrared singularity. In the mass-less case this is an obstacle since the infrared divergence is present in the limit $s \to 0$.[4] In order to overcome this problem we can insert an artificial parameter μ^2 as a mass-scale instead of m^2 in (4) and then take the limit $\mu^2 \to 0$. This has been done for instance to obtain the conformal anomaly.[3] However, another interesting approach is to maintain in all the potential calculations μ^2 different to zero and to modify consistently the coefficients $a_j(x)$ of (4) in order to ensure that no extra divergences appear in the

case of $\mu^2 \neq 0$. In the most general case, without assuming $m^2 = 0$, the extended DeWitt-Schwinger subtraction scheme would be[6]

$$\bar{G}_{\rm ren}(x,x) = \lim_{x' \to x} G(x,x') - \frac{1}{(4\pi)^2} \sum_{j=0}^{1} \bar{a}_j(x) \int_0^\infty \frac{ds}{(is)^{(2-j)}} e^{-i\left(m^2+\mu^2\right)s} \ . \qquad (5)$$

Here the the new coefficients are $\bar{a}_0 = 1$, $\bar{a}_1 = a_1 + \mu^2$ and $\bar{a}_2 = a_1\mu^2 + \frac{1}{2}\mu^4$ [a]. The introduction of this extra parameter is required for the massless limit. However, it also offers a possible running coupling interpretation analogue to the Minimal subtraction scheme in dimensional regularization. We will analyze this in the next section.

An important property of the regularization and renormalization program is that it should not spoil the decoupling of heavy massive fields in the low energy limit. For instance, we should not have to compute the vacuum polarization of the quark top mass when performing a low energy (below the top mass) scattering process in QED. The decoupling of heavy field in renormalizable theories has been well stated by the Appelquist-Carazzone theorem.[8] In the case of QED, Minimal Subtraction is not compatible with decoupling[9] and a different subtraction scheme has to be taken in order to have a physical interpretation of the μ parameter at low energy regime. A possible scheme is known as *Momentum subtraction scheme*,[9] which not only decouples the heavy fields, i.e., $\beta_O \to 0$ when $m^2 \to \infty$ but also recovers the Minimal Subtraction beta functions in the limit of $\mu^2 \to \infty$ or $m^2 \to 0$.

In the case of quantum fields in curved spacetime an analog to the Momentum subtraction scheme has been elusive. A notable result of the subtraction scheme of (5) is that it is consistent with the decoupling of heavy massive fields,[6] as we will show in the next section.

2. Running of the couplings and decoupling of massive fields

In order to compute the running of the couplings with the parameter μ it is useful to use the extended DeWitt-Schwinger subtraction scheme for the effective action, which can be constructed in a similar manner to (5).[7] It is also useful for pedagogical purposes to consider a complex scalar field interacting with an electromagnetic field obeying the Klein-Gordon equation

$$\left(g^{\mu\nu} D^x_\mu D^x_\nu + m^2 + \xi R\right) G(x,x') = -|g(x)|^{-1/2} \delta(x'-x), \qquad (6)$$

where $D_\mu = \nabla_\mu + iqA_\mu$. In this case, the extended DeWitt-Schwinger subtraction scheme takes the form[7]

$$W^{(1)}_{\rm ren} = W^{(1)} - \frac{2i}{2(4\pi)^2} \int d^4x \sqrt{-g} \sum_{j=0}^{2} \bar{a}_j(x) \int_0^\infty \frac{ds}{(is)^{(3-j)}} e^{-i\left(m^2+\mu^2\right)s} \qquad (7)$$

[a] Note here that in terms of adiabatic order, the μ^2 in the exponential of (5) is of adiabatic order zero while the μ^2 of the coefficients are of adiabatic order two. This is equivalent to the methodology taken in[7] for Parker-Fulling adiabatic regularization.

where $W^{(1)}$ is the divergent one-loop effective action constructed from (1) and the coefficients are

$$\bar{a}_0(x) = 1\ , \qquad \bar{a}_1(x) = \left(\frac{1}{6} - \xi\right) R + \mu^2$$
$$\bar{a}_2(x) = \frac{1}{180} R_{\alpha\beta\gamma\delta} R^{\alpha\beta\gamma\delta} - \frac{1}{180} R^{\alpha\beta} R_{\alpha\beta} - \frac{1}{6}\left(\frac{1}{5} - \xi\right) \Box R$$
$$+ \frac{1}{12}\left(\xi - \frac{1}{6}\right)^2 R^2 - \frac{q^2}{12} F_{\mu\nu} F^{\mu\nu} + \left(\frac{1}{6} - \xi\right) R\mu^2 + \frac{1}{2}\mu^4\ . \tag{8}$$

The subtraction terms of (7) can be absorbed into the original action

$$S^c = S_G + \int d^4x \sqrt{-g}\left(-\frac{1}{4} Z_A \tilde{F}_{\mu\nu} \tilde{F}^{\mu\nu} + (\tilde{D}_\mu \phi)^\dagger \tilde{D}^\mu \phi + m^2 |\phi|^2 + \xi R |\phi|^2\right) \tag{9}$$

with $\tilde{D}_\mu = \nabla_\mu + iq Z_A^{1/2} \tilde{A}_\mu$, where we have re-scaled the potential to introduce an extra coupling constant Z_A. The complete effective action would be

$$S^{(1)}_{\rm ren} = W^{(1)}(m, \xi, q Z_A^{1/2}) + S^c(m, \xi, q, Z_A, \Lambda, \kappa, \alpha_i), \tag{10}$$

where we have made visible the explicit dependence of the bare couplings and fields. The complete action in terms of the physical renormalized couplings O_R is

$$S^{(1)}_{\rm ren} = W^{(1)}_{\rm ren}(m_R, \xi_R, q_R Z_R^{1/2}) + S^c_{\rm ren}(Z_R, \Lambda_R, \kappa_R, \alpha_{iR}) \tag{11}$$

where $S^c_{\rm ren}$ is the original S^c upgrading the bare divergent coupling constants to the corresponding renormalized finite ones.In the particular case of a free complex scalar field, it is not difficult to see that expression (7) does not involve any divergent piece that need to be reabsorbed neither in m^2, nor ξ and the combination $qZ^{1/2}$ and therefore we can already write

$$\xi = \xi_R \qquad m = m_R \qquad q Z_A^{1/2} = q_R Z_{AR}^{1/2}. \tag{12}$$

This is of course not a general result and indeed this couplings constants could be divergent in theories including other interactions.[10] We will from now on drop the subscript $_R$ from these terms.

In order to compute the beta functions $\beta_O = \mu \frac{dO_R}{d\mu}$ we will impose the invariance of the renormalized action with respect to the arbitrary parameter μ

$$\mu \frac{d}{d\mu} S^{(1)}_{\rm ren} = 0. \tag{13}$$

From (7), it is easy to check

$$\mu \frac{d}{d\mu} W^{(1)}_{\rm ren} = \int d^4x \sqrt{-g} \frac{1}{16\pi^2 (m^2 + \mu^2)} \left(2\mu^2 a_2(x) - 2\mu^4 a_1(x) - \mu^6\right). \tag{14}$$

Using (12), (13) and (14), we finally obtain the following beta functions

$$\beta_{\alpha 1} = -\frac{1}{960\pi^2}\frac{\mu^2}{m^2+\mu^2} \qquad \beta_{\alpha 2} = -\frac{\left(\xi-\frac{1}{6}\right)^2}{16\pi^2}\frac{\mu^2}{m^2+\mu^2}$$
$$\beta_{\Lambda} = \frac{1}{16\pi^2}\frac{\mu^6}{m^2+\mu^2} \qquad \beta_{\kappa^{-1}} = \frac{\xi-\frac{1}{6}}{4\pi^2}\frac{\mu^4}{m^2+\mu^2} \qquad \beta_{q} = \frac{q_R^3}{48\pi^2}\frac{\mu^2}{m^2+\mu^2} \quad . \tag{15}$$

It can be easily checked that the running of dimensionless coupling constants coincides with the Minimal Subtraction beta functions,[11] but with the advantage that the structure $\frac{\mu^2}{m^2+\mu^2}$, similar to the cutoff Wilsonian scheme[12] makes a vanishing contribution when $m^2 \to \infty$ thus making this renormalization scheme compatible with decoupling of heavy massive fields. The remarkable result is that also the running of both the Newton constant and the cosmological constant also decouple in this limit, thus enforcing the fact that higher massive fields do not contribute to the low energy regime, even when the gravitational field is present.

References

1. M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading MA, 1995).
2. J.C. Collins, *Renormalization* (Cambridge University Press, Cambridge, England 1984).
3. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England 1982).
4. R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics* (University of Chicago Press, Chicago, 1994).
5. L. Parker and D. J. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Fields and Gravity* (Cambridge University Press, Cambridge, England 2009).
6. A. Ferreiro and J. Navarro-Salas, *Phys. Rev. D* **102**, 045021 (2020).
7. A. Ferreiro and J. Navarro-Salas, *Phys. Lett. B* **792**, 81 (2019).
8. T. Appelquist and J. Carazzone, *Phys. Rev. D* **11**, 2856 (1975).
9. L. Alvarez-Gaume and M. A. Vazquez-Mozo, *An Invitation to Quantum Field Theory* (Springer-Verlag, Berlin, 2012).
10. A. Ferreiro, S. Nadal-Gisbert and J. Navarro-Salas, *Phys. Rev. D* **104**, 025003 (2021)
11. J. Solà, *J. Phys. Conf. Ser.* **453**, 012015 (2013).
12. T. J. Hollowood, *Renormalization Group and Fixed Points in Quantum Field Theory* (Springer, Heidelberg, 2013).

Renormalization and decoupling for the Yukawa model in curved spacetime

Sergi Nadal-Gisbert*, Antonio Ferreiro† and José Navarro-Salas‡

Departamento de Física Teórica and IFIC, Centro Mixto Universidad de Valencia-CSIC Facultad de Física, Universidad de Valencia, Burjassot-46100, Valencia, Spain

*E-mail: †antonio.ferreiro@ific.uv.es, *sergi.nadal@uv.es, ‡jnavarro@ific.uv.es*

We consider the renormalization of the one-loop effective action for the Yukawa interaction. We compute the beta functions in the generalized DeWitt-Schwinger subtraction scheme. For the quantized scalar field we obtain that all the beta functions exhibit decoupling for heavy fields as stated by the Appelquist-Carazzone theorem including also the gravitational couplings. For the quantized Dirac field, decoupling appears for almost all of them. We obtain the atypical result that the mass parameter of the background scalar field, does not decouple.[a]

Keywords: Decoupling, renormalization, running couplings, beta functions, one-loop effective Lagrangian, DeWitt-Schwinger expansion

1. Introduction

Regularization and renormalization of infinities of quantum fields in curved spacetime is a subtle subject. Infinite quantities emerge at the vacuum level and can not be erased due to the non-trivial structure of the classical gravitational background field. There are well established covariant and pragmatic methods to evaluate the vacuum expectation value of the stress-energy tensor in physically reasonable states.[2–7] For a general spacetime, the DeWitt-Schwinger technique[8–10] allows us to evaluate and renormalize the stress-energy tensor as well as the one-loop effective action which constitute a practical tool that encodes all the relevant information about the quantum effects of fields.

The regularization machinery usually incorporates an arbitrary, non physical, mass-scale parameter. Studying how the coupling constants depend on this scale, i.e., the running of the couplings, is very useful in physical situations where more formal computations are very involved. This is done by requiring that physical quantities should not depend on spurious parameters.[11–13] As a representative case, in dimensional regularization an arbitrary mass scale μ is introduced to compensate the fictitious dimensions. In the case of the DeWitt-Schwinger proper-time expansion a mass scale μ is usually introduced in the short distance logarithmic term $\log \mu^2\sigma/2$ to overcome an infrared divergence for massless fields. Changing the

[a]Expanded version of the talk given by S. Nadal-Gisbert in the Sixteenth Marcel Grossmann Meeting (2021).

mass scale μ to μ' allows to obtain an effective running for the parameters of the Lagrangian.

Another important feature of effective field theories is that massive fields should decouple at low energies[14,15] as stated by the Appelquist-Carazzone theorem[16b]. This idea which lies at the heart of the effective field theories states that a field of mass m can not influence the physics at scales larger than m^{-1}. In a physical renormalization scheme this idea should be manifested in the beta functions of the theory if they want to describe both infrared and ultraviolet regimes. Moreover, in gravitational physics decoupling is crucial to get a proper physical interpretation in the cosmic infrared regime. This can be particularly relevant for the cosmological constant problem and for the running of the Newton's gravitational constant.[19–22] A partial list of works dealing with this issue in a curved space are.[23–28]

Recently, an arbitrary scale parameter μ has been introduced in the DeWitt-Schwinger subtraction scheme in order to consistently construct the subtraction terms avoiding the infrared divergences[29] (see also[30–32]). This was done for the free complex scalar field in the Einstein-Maxwell theory. This approach has the benefit of producing the decoupling of heavy fields. In this work we further extend this approach by including Yukawa interactions with spinor fields and scalars also coupled to gravity. This model has been broadly studied in dimensional regularization[24,33–35] also by including gauge fields.[36,37] We will concentrate in the simplest case of a scalar background coupled to a quantized dirac field and also to a scalar field. The aim of this work is to compute the beta functions in the generalized DeWitt-Schwinger renormalization approach and observe whether decoupling explicitly appears.

2. Interaction with a quantized scalar field and renormalization

Consider a quantized real scalar field φ coupled to a real scalar background ϕ via the Yukawa interaction $\frac{h^2}{2}\phi^2\varphi^2$

$$S = \int d^4x\sqrt{-g}\left\{-\Lambda + \frac{R}{16\pi G} + \frac{1}{2}\nabla^\mu\varphi\nabla_\mu\varphi - \frac{1}{2}\left(m^2+\xi R\right)\varphi^2 \right.$$
$$\left. -\frac{h^2}{2}\phi^2\varphi^2 + \frac{1}{2}\nabla^\mu\phi\nabla_\mu\phi - V(\phi)\right\}\ , \tag{1}$$

where m^2 is the mass parameter for the quantized scalar, ξ is the coupling of φ^2 to the Ricci scalar, and $V(\phi)$ is a general potential that can contain interactions between the background field and the curvature but is independent of the quantized scalar field φ. The Feynman propagator G_F for φ satisfies

$$\left(\Box_x + m^2 + \xi R + h^2\phi^2\right) G_{\rm F}(x,x') = -|g(x)|^{-1/2}\delta(x-x'). \tag{2}$$

[b]There are some situations were decoupling is violated. This typically happens on theories with spontaneously broken gauge symmetries.[17,18]

The effective action can be generated from this propagator by $S_{\rm eff} = -i\frac{1}{2}\,{\rm Tr}\log(-G_{\rm F})$. The ultraviolet divergences of the one-loop effective Lagrangian can be explicitly manifested if one expresses the Feynman propagator as an integral in the proper time s,

$$G_{\rm F}(x,x') = -i\int_0^\infty ds\; e^{-im^2 s}\langle x,s|x',0\rangle\;, \tag{3}$$

where m^2 is understood to have an infinitesimal negative imaginary part ($m^2 \equiv m^2 - i\epsilon$). The heat kernel $\langle x,s|x',0\rangle$ can be expanded in powers of the proper time as follows

$$\langle x,s|x',0\rangle = i\frac{\Delta^{1/2}(x,x')}{(4\pi)^2(is)^2}\exp\frac{\sigma(x,x')}{2is}\sum_{j=0}^\infty a_j(x,x')(is)^j\;, \tag{4}$$

where $\Delta(x,x')$ is the Van Vleck-Morette determinant and $\sigma(x,x')$ is the proper distance along the geodesic from x' to x. Therefore, the effective Lagrangian, defined as $S_{\rm eff} = \int d^4x\sqrt{-g}L_{\rm eff}$, has the following asymptotic expansion

$$L_{\rm eff} = \frac{i}{2(4\pi)^2}\sum_{j=0}^\infty a_j(x)\int_0^\infty e^{-ism^2}(is)^{j-3}ds\;. \tag{5}$$

The first coefficients $a_n(x,x')$ are given, in the coincidence limit $x\to x'$, by[5]

$$\begin{aligned} a_0(x) =& 1\;, \qquad a_1(x) = \frac{1}{6}R - Q \\ a_2(x) =& \frac{1}{180}R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} - \frac{1}{180}R^{\alpha\beta}R_{\alpha\beta} - \frac{1}{30}\Box R + \frac{1}{72}\;R^2 \\ &+ \frac{1}{2}Q^2 - \frac{1}{6}RQ + \frac{1}{6}\Box Q + \frac{1}{12}W_{\mu\nu}W^{\mu\nu}\;, \end{aligned} \tag{6}$$

where for the case of a scalar field we have $W_{\mu\nu} \equiv [\nabla_\mu,\nabla_\nu] = 0$ and $Q = \xi R + h^2\phi^2$. Expansion (5) shows that the ultraviolet divergent terms of the effective Lagrangian are localized in the first three terms of the DeWitt-Schwinger expansion in the limit $s\to 0$ of the integral. The renormalization procedure can be performed by directly subtracting these divergent terms to the total one-loop contribution.

$$L_{\rm ren} = \int_0^\infty ds\left[\frac{e^{-im^2 s}}{is}\langle x,s|x,0\rangle - \frac{i}{2(4\pi)^2}\sum_{j=0}^2 a_j(x)\frac{e^{-ism^2}}{(is)^{-j+3}}\right], \tag{7}$$

where the notation of $L_{\rm ren}$ has to be understood as the finite one-loop correction to the background Lagrangian. One can notice that the massless case inherits an infrared divergence ($s\to\infty$ limit). We avoid it by introducing a mass scale μ^2 in the exponential term of the DeWitt-Schwinger expansion by writing $\sum_j a_j(x)e^{-ism^2} \to \sum_j \bar{a}_j(x)e^{-is(m^2+\mu^2)}$ in (5). As mentioned in,[29] this is the unique way of introducing the μ parameter if we want the exponential form to remain. The main point of the introduction of the arbitrary parameter μ at the same level of the mass, is to obtain decoupling in the infrared behaviour of the beta functions, as we will see.

The DeWitt coefficients a_j are redefined by consistency with the adiabatic orders, $\bar{a}_0(x) = 1, \bar{a}_1(x) = a_1(x) + \mu^2, \bar{a}_2(x) = a_2(x) + a_1(x)\mu^2 + \frac{1}{2}\mu^4$, and renormalization is now performed by taking the difference $L_{\rm ren}(\mu) = L_{\rm eff} - L_{\rm div}(\mu)$, as for the standard situation with $\mu = 0$, (7)

$$L_{\rm ren}(\mu) = \int_0^\infty ds \left[\frac{e^{-im^2 s}}{is} \langle x, s | x, 0 \rangle - \frac{i}{2(4\pi)^2} \sum_{j=0}^{2} \bar{a}_j(x) \frac{e^{-is(m^2+\mu^2)}}{(is)^{-j+3}} \right] . \qquad (8)$$

We note that the renormalized one-loop Lagrangian depends now on μ. This dependece in the arbitrary parameter μ has to be compensated by a running of the couplings of the background Lagrangian, thus the physical one-loop renormalized Lagrangian $L_{\rm phys} = L_{\rm B}(\mu) + L_{\rm ren}(\mu)$ is μ-independent, where $L_{\rm B}(\mu)$ refers to the background Lagrangian.

2.1. *Running of the coupling constants and decoupling*

The background Lagrangian required for renormalization is given by

$$L_{\rm B} = L_{\rm grav} + \frac{1}{2} Z \nabla^\mu \phi \nabla_\mu \phi - \frac{M^2}{2} Z \phi^2 - \frac{\xi_\phi}{2} R Z \phi^2 - \frac{\lambda}{4!} Z^2 \phi^4 + \gamma_1 \Box Z \phi^2 , \qquad (9)$$

where

$$L_{\rm grav} = -\Lambda + \frac{1}{2}\kappa R + \alpha_1 R^2 + \alpha_2 R_{\mu\nu} R^{\mu\nu} + \alpha_3 R_{\mu\nu\alpha\beta} R^{\mu\nu\alpha\beta} + \alpha_4 \Box R . \qquad (10)$$

We have defined $\kappa = 1/8\pi G$. The new terms are required to absorb the divergences of the one-loop correction. The coupling Z will not receive any contribution from the scalar quantum fluctuations, therefore we can canonically normalize it to 1. The remaining couplings $\lambda(\mu), \kappa(\mu), \alpha_i(\mu), M(\mu)$, etc will inherit a dependence on the mass scale μ.

The beta functions and the running couplings are built from the renormalization process, therefore they get only contributions from the divergent terms. For this reason, it is enough to approximate (8) with the first three terms of the asymptotic expansion for the heat kernel.

$$L_{\rm ren}(\mu) \approx \frac{i}{2(4\pi)^2} \int_0^\infty \frac{ds}{s^3} \Big\{ e^{-im^2 s} \left[1 + (is) a_1(x) + (is)^2 a_2(x) \right] - e^{-is(m^2+\mu^2)} \left[1 + (is)\bar{a}_1(x) + (is)^2 \bar{a}_2(x) \right] \Big\} . \qquad (11)$$

This is a finite integral that can now be computed. Its result depends on the DeWitt coeficients $a_0(x)$, $a_1(x)$ and $a_2(x)$ and also with the arbitrary μ parameter. Requiring that the physical one-loop renormalized Lagrangian $L_{\rm phys} = L_{\rm B}(\mu) + L_{\rm ren}(\mu)$ has to be μ-independent, leads to the running of the couplings and the beta functions.

$$\frac{dL_{\rm phys}}{d\mu} = 0 \quad \rightarrow \quad \beta_i \frac{\partial L_B}{\partial q_i} = -\mu \frac{\partial L_{\rm ren}}{\partial \mu} , \qquad (12)$$

where q_i refers to all the parameters of the Lagrangian and $\beta_i = \mu \frac{\partial q_i}{\partial \mu}$. A representative of the beta functions that we obtain are

$$\beta_\lambda = \frac{3h^4}{4\pi^2}\frac{\mu^2}{m^2+\mu^2} \qquad \beta_\Lambda = \frac{1}{32\pi^2}\frac{\mu^6}{m^2+\mu^2}$$
$$\beta_\kappa = \frac{\bar{\xi}}{8\pi^2}\frac{\mu^4}{m^2+\mu^2} \qquad \beta_{M^2} = -\frac{h^2}{8\pi^2}\frac{\mu^4}{m^2+\mu^2}\ , \tag{13}$$

where we have defined $\bar{\xi} = \left(\xi - \frac{1}{6}\right)$. The rest of beta functions can be found in Appendix A. Several remarks are worth of mention in these results. For the dimensionless parameters, like $\lambda(\mu)$, we obtain a factor of the form $\mu^2/(m^2+\mu^2)$, this factor also arises in the hierarchy of beta functions in the Wilsonian renormalization approach for a scalar field theory.[38] Analyzing this factor it is easy to see that for $m^2 \gg \mu^2$ the dimensionless beta functions decouple as expected in the infrared regime. Moreover for $\mu^2 \gg m^2$ the ultraviolet regime is recovered as expected from dimensional regularization with MS (minimal subtraction).[33] Another important feature concerns the dimensional couplings. For the Newton and cosmological constants we recover the same result as in the free field theory. Furthermore, all the dimensional couplings also decouple when $m^2 \gg \mu^2$.

3. Renormalization for Dirac fields

In this section we consider a quantized Dirac field ψ coupled via Yukawa interaction $g_Y \phi \bar{\psi}\psi$ with a classical scalar background ϕ. The action is given by

$$S = \int d^4x\sqrt{-g}\left(-\Lambda + \frac{R}{16\pi G} + \bar{\psi}\left(i\gamma^\mu\nabla_\mu - m\right)\psi - g_Y\phi\bar{\psi}\psi + \frac{1}{2}\nabla^\mu\phi\nabla_\mu\phi - V(\phi)\right), \tag{14}$$

where the covariant derivative ∇_μ acting on the Dirac field is defined as the ordinary derivative plus the spin connection term. $\gamma^\mu(x)$ are the curved space Dirac matrices $\gamma^\mu(x) = e_a{}^\mu\gamma^a$, defined in terms of the usual Dirac matrices in Minkowski space γ^a and the vierbein $e_a{}^\mu$.

In order to build the DeWitt-Schwinger expansion one needs the Klein-Gordon equation for the Dirac field

$$(\Box_x + m^2 + Q)G_{\rm F}(x,x') = -|g(x)|^{-1/2}\delta(x-x')\ . \tag{15}$$

Consequently, the one-loop effective action takes the form $S^{(1)}_{\rm eff} = \frac{1}{2}i\,{\rm Tr}\log(-G_F)$. One can expand the heat kernel in the proper time asymptotic series as for the scalar field (4) and make use of the same coefficients (6). In this case Q is given by

$$Q(x) = \frac{1}{4}R(x) + ig_Y\gamma^\mu\nabla_\mu\phi(x) + g_Y^2\phi^2(x) + 2g_Y m\phi(x)\ . \tag{16}$$

It is important to stress the appearance of the last term in the above expression which is proportional to m. This contrast with the obtained expression for Q in the scalar case ($Q = \xi R + h^2\phi^2$). Generically, all DeWitt coefficients a_n (or $\bar{a}_n$) are local geometrical quantities independent of the mass of the field. The Yukawa interaction for Dirac fermions introduces a mass-dependent term in the expression for $Q(x)$. This term will be the responsible of the violation of the decoupling characteristic for the background scalar mass parameter.

As in the previous section one gets the expression for the subtraction terms with the asymptotic expansion

$$L_{\rm div}(\mu) = \frac{-i}{2(4\pi)^2}\sum_{j=0}^{2}\operatorname{tr}\bar{a}_j(x)\int_0^\infty e^{-is(m^2+\mu^2)}(is)^{j-3}ds\ . \tag{17}$$

The main difference with the scalar case is that one needs to take into account the trace of the spinor indices acting on the coefficients $\operatorname{tr}\bar{a}_j(x)$. We still have the modified DeWitt coefficients $\bar{a}_0(x) = 1$, $\bar{a}_1(x) = a_1(x) + \mu^2$, and $\bar{a}_2(x) = a_2(x) + \left(\frac{1}{6}R - Q\right)\mu^2 + \frac{1}{2}\mu^4$ related to the DeWitt coefficients (6) but now with Q given by (16) and $W_{\mu\nu} = [\nabla_\mu, \nabla_\nu] = -\frac{1}{8}R_{\mu\nu ab}\left[\gamma^a, \gamma^b\right]$.

3.1. *Running of the coupling constant and decoupling*

The background Lagrangian terms are required by renormalization in order to absorb the divergent terms coming from the one-loop correction (17)

$$\begin{aligned} L_{\rm B} = {} & L_{\rm grav} + \frac{1}{2}Z\nabla^\mu\phi\nabla_\mu\phi - \frac{M^2}{2}Z\phi^2 - \frac{\xi}{2}RZ\phi^2 - \tau Z^{1/2}\phi - \frac{\eta}{3!}Z^{3/2}\phi^3 \\ & - \frac{\lambda}{4!}Z^2\phi^4 - \xi_1 RZ^{1/2}\phi + \gamma_1\Box Z\phi^2 + \gamma_2\Box Z^{1/2}\phi\ , \end{aligned} \tag{18}$$

where $L_{\rm grav}$ was given in (10). In the Dirac case, Z gets a contribution from the quantum fluctuations of the Dirac field. Therefore, it has a running that can be related to a running of the field ϕ by a reparametrization. We write $Z = 1 + \delta Z$ as usual, to take into account canonical normalization and the one-loop correction. For simplicity, we introduce new primed couplings to absorb Z, except for the kinetic term, where we leave Z explicitly

$$\frac{1}{2}Z\nabla^\mu\phi\nabla_\mu\phi - \frac{M'^2}{2}\phi^2 - \frac{\xi'}{2}R\phi^2 - \tau'\phi - \frac{\eta'}{3!}\phi^3 - \frac{\lambda'}{4!}\phi^4 - \xi_1' R\phi + \gamma_1'\Box\phi^2 + \gamma_2'\Box\phi\ . \tag{19}$$

As for the scalars, we just need to consider the divergent part of the one-loop effective action (11) to compute the running of the parameters. Again, we impose that the physical one-loop renormalized Lagrangian $L_{\rm phys} = L_{\rm B}(\mu) + L_{\rm ren}(\mu)$ must be independent of the value of μ. Therefore considering (12) for this background Lagrangian we can obtain the beta functions for the primed parameters. By direct differentiation and keeping one-loop order $\mathcal{O}(\hbar)$, it is straightforward to obtain the

beta functions of the unprimed original couplings

$$\begin{aligned}
\beta_{M^2} &= \beta_{M'^2} - M^2\beta_Z & \beta_\xi &= \beta_{\xi'} - \xi\beta_Z \\
\beta_{\xi_1} &= \beta_{\xi_1'} - \frac{1}{2}\xi_1\beta_Z & \beta_\tau &= \beta_{\tau'} - \frac{\tau}{2}\beta_Z \\
\beta_\eta &= \beta_{\eta'} - \frac{3}{2}\eta\beta_Z & \beta_\lambda &= \beta_{\lambda'} - 2\lambda\beta_Z \\
\beta_{\gamma_1} &= \beta_{\gamma_1'} - \gamma_1\beta_Z & \beta_{\gamma_2} &= \beta_{\gamma_2'} - \frac{1}{2}\gamma_2\beta_Z \ .
\end{aligned}$$

The result for all the beta functions can be seen in Appendix A. Let us analyze the different regimes of the scalar wave function Z as a representative of the dimensionless parameters.

$$\beta_Z = -\frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2+\mu^2} \ . \tag{20}$$

In the ultraviolet regime, $\mu \gg m$ we recover the result from dimensional regularization with MS.[33,34]

$$\beta_Z = -\frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2+\mu^2} \to_{\mu\gg m} -\frac{g_Y^2}{4\pi^2} \ , \tag{21}$$

while for the infrared regime, $\mu \ll m$, we find decoupling

$$\beta_Z = -\frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2+\mu^2} \to_{\mu\ll m} -\frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2} \ . \tag{22}$$

The other beta functions for the dimensionless parameters exhibit the same behaviour. Decoupling of massive fields is manifested in the infrared regime and the ultraviolet regime agrees with the results of dimensional regularization with MS.[33,34]

Things are more involved for the dimensional constants. We observe different behaviours with the parameter μ. For example for the couplings $\Lambda(\mu)$, $\kappa(\mu)$ and $M'^2(\mu)$ we get

$$\begin{aligned}
\beta_\Lambda &= -\frac{1}{8\pi^2}\frac{\mu^6}{m^2+\mu^2} & \beta_\kappa &= -\frac{1}{24\pi^2}\frac{\mu^4}{m^2+\mu^2} \\
\beta_{M^2} &= \frac{g_Y^2}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(+4\mu^2 - 8m^2 + 2M^2\right) \ .
\end{aligned} \tag{23}$$

A common feature is that decoupling appears when $m^2 \gg \mu^2$ for all the dimensional couplings except for the background mass parameter M^2. In this case the beta function essentially reproduces the value obtained via dimensional regularization with MS times a factor of order μ^2/m^2. The origin of the term proportional to $-8m^2$ can be retrieved from the term $2g_Y m\phi$ in Q. This can be linked to the fact that when building the DeWitt-Schwinger expansion all the mass dependence is assumed to be encoded in the exponential term[39] in the form m^2 and this expansion does not work properly if Q inherits a mass dependence. The latter result shows that finding decoupling for all the coupling constants, of a given theory, although it is

crucial to define a cosmic infrared regime, it is indeed a nontrivial task, as already emphasized in.[25]

Finally it is worth showing the finite expression for the running of the Newton gravitational constant G, the cosmological constant Λ and the scalar mass term M'^2 coming from the Dirac field at two different scales μ and μ_0.

$$\Lambda(\mu) = \Lambda_0 - \frac{1}{32\pi^2}\left((\mu^4-\mu_0^4) - 2m^2(\mu^2-\mu_0^2) + 2m^4\log\left(\frac{m^2+\mu^2}{m^2+\mu_0^2}\right)\right), \tag{24}$$

$$G(\mu) = \frac{G_0}{1-\frac{G_0}{6\pi}\left(\mu^2-\mu_0^2 - m^2\log\left(\frac{m^2+\mu^2}{m^2+\mu_0^2}\right)\right)}\ , \tag{25}$$

$$M'^2(\mu) = M_0'^2 + \frac{g_Y^2}{4\pi^2}\left((\mu^2-\mu_0^2) - 3m^2\log\left(\frac{m^2+\mu^2}{m^2+\mu_0^2}\right)\right), \tag{26}$$

where G_0, Λ_0 and $M_0'^2$ are the parameters at the scale μ_0. Note the appearance of μ^4 and μ^2 terms. In more conventional approaches only the logarithmic terms are present. However these non-logarithmic terms can be traced to the non trivial quartic and quadratic divergences and are indeed responsible of the appearance of decoupling. Expanding $\Lambda(\mu)$ and $G(\mu)$ for $\mu^2 \gg m^2$ one can check that the runnings are suppressed by a factor μ^6/m^2 and μ^4/m^2. However for $M'^2(\mu)$ we still get a non suppressed term proportional to μ^2.

4. Conclusions and final comments

We have considered the one-loop effective action for a quantized scalar field and Dirac field coupled to a background scalar field and gravity. Using the upgraded DeWitt-Schwinger subtraction scheme we have renormalized the theory and computed the beta functions for all the parameters. As a remarkable result, for the dimensionless couplings we obtain the beta functions that satisfy both ultraviolet and infrared regimes. In the ultraviolet regime we recover the beta function from MS scheme. In the infrared regime the beta functions explicitly exhibit the decoupling property. For the dimensional parameters new μ^4 and μ^2 terms appear in the runnings. These terms signal the presence of quadratic and quartic divergences and are indeed the responsible for the decoupling in the running couplings. Therefore, the obtained beta functions of the dimensional parameters exhibit the property of decoupling in the low energy regime, including the Newton's and cosmological constant. The exception is for the contribution of the Dirac field to the scalar mass parameter. As stress before the reason is localized in the linear term $2g_y m\phi$ in Q.

Acknowledgments

This work has been partially supported by Spanish Ministerio de Economia, Industria y Competitividad Grants No. FIS2017-84440-C2-1-P (MINECO/FEDER,

EU) and No. FIS2017-91161-EXP, and also by the project PROMETEO/2020/079 (Generalitat Valenciana). A. F. is supported by the Severo Ochoa Ph.D. fellowship, Grant No. SEV-2014-0398-16-1, and the European Social Fund. S. N. is supported by the Universidad de Valencia, within the Atracció de Talent Ph.D fellowship No. UV-INV- 506 PREDOC19F1-1005367.

Appendix A

The obtained beta functions for the scalar one-loop correction are

$$\begin{aligned}
\beta_{\xi_\phi} &= \frac{h^2\bar{\xi}}{8\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta_{\alpha_1} &= -\frac{\bar{\xi}^2}{32\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_{\alpha_4} &= -\frac{\xi-\frac{1}{5}}{96\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta_{\alpha_2} &= \frac{1}{2880\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_{\alpha_3} &= -\frac{1}{2880\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta\gamma_1 &= -\frac{h^2}{96\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_\lambda &= \frac{3h^4}{4\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta_\Lambda &= \frac{1}{32\pi^2}\frac{\mu^6}{m^2+\mu^2}\\
\beta_\kappa &= \frac{\bar{\xi}}{8\pi^2}\frac{\mu^4}{m^2+\mu^2} & \beta_{M^2} &= -\frac{h^2}{8\pi^2}\frac{\mu^4}{m^2+\mu^2}\ .
\end{aligned} \tag{A.1}$$

Dimensionless beta functions for the Dirac field are:

$$\begin{aligned}
\beta_Z &= -\frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_\xi &= -\frac{g_Y^2}{24\pi^2}\frac{\mu^2}{m^2+\mu^2}(1-6\xi) & \beta_\lambda &= -\frac{g_Y^2}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(24g_Y^2-4\lambda\right)\\
\beta_{\alpha_1} &= \frac{1}{1152\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta_{\alpha_2} &= -\frac{1}{720\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_{\alpha_3} &= -\frac{7}{5760\pi^2}\frac{\mu^2}{m^2+\mu^2} & \beta_{\alpha_4} &= \frac{1}{480\pi^2}\frac{\mu^2}{m^2+\mu^2}\\
\beta_{\gamma_1} &= \frac{g_Y^2}{4\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(\frac{1}{6}+\gamma_1\right) & \beta_{\gamma_2} &= \frac{g_Y}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(\frac{2}{3}m+\gamma_2\right)\ .
\end{aligned} \tag{A.2}$$

Dimensional beta functions for the Dirac field:

$$\begin{aligned}
\beta_\Lambda &= -\frac{1}{8\pi^2}\frac{\mu^6}{m^2+\mu^2} & \beta_{\xi_1} &= \frac{g_Y}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(-\frac{m}{3}+g_Y\xi_1\right)\\
\beta_\kappa &= -\frac{1}{24\pi^2}\frac{\mu^4}{m^2+\mu^2} & \beta_\tau &= -\frac{g_Y}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(4m\mu^2-\tau g_Y\right)\\
\beta_{M^2} &= \frac{g_Y^2}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(+4\mu^2-8m^2+2M^2\right) & \beta_\eta &= -\frac{g_Y^3}{8\pi^2}\frac{\mu^2}{m^2+\mu^2}\left(24g_Y m-3\eta\right)\ .
\end{aligned} \tag{A.3}$$

References

1. A. Ferreiro, S. Nadal-Gisbert and J. Navarro-Salas, *Phys. Rev. D* **104**, 025003 (2021).
2. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England 1982).
3. S. Fulling, *Aspects of Quantum Field Theory in Curved Space-Time* (Cambridge University Press, Cambridge, England 1989).
4. R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics* (University of Chicago Press, Chicago, 1994).
5. L. Parker and D. J. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Fields and Gravity* (Cambridge University Press, Cambridge, England 2009).
6. B.-L. B. Hu and E. Verdaguer, *Semiclassical and Stochastic Gravity* (Cambridge University Press, Cambridge, England 2020).
7. I. L. Buchbinder and I. L. Shapiro, *Introduction to Quantum Field Theory with Applications to Quantum Gravity* (Oxford University Press, Oxford, England, 2021).
8. J. Schwinger, *Phys. Rev.* **82**, 664 (1951).
9. B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York 1965).
10. B. S. DeWitt, *Phys. Rep.* **19**, 295 (1975).
11. S. Weinberg, *The Quantum Theory of Fields* (Cambridge University Press, Cambridge, England 1995), Vol. 1, 2.
12. M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading MA, 1995).
13. L. Alvarez-Gaume and M. A. Vazquez-Mozo, *An Invitation to Quantum Field Theory* (Springer-Verlag, Berlin, 2012).
14. I. L. Shapiro and J. Solà, *J. High Energy Phys.* 02, (2002) 006.
15. A. Babić, B. Guberina, R. Horvat and H. Štefančić, *Phys. Rev. D* **65**, 085002 (2002).
16. T. Appelquist and J. Carazzone, *Phys. Rev. D* **11**, 2856 (1975).
17. J.C. Collins, *Renormalization* (Cambridge University Press, Cambridge, England 1984).
18. F. Feruglio, *Acta Phys. Polon. B* **25**, 1279 (1994).
19. S.M. Carroll, *Living Rev. Relativ.* **4**, 1 (2001).
20. J. Martin, *C.R. Phys* **13**, 566 (2012).
21. J. Solà, *J. Phys. Conf. Ser.* **453**, 012015 (2013).
22. M. Niedermaier and M. Reuter, *Living Rev. Rel.* **9**, 5 (2006). M. Reuter, arXiv: hep-th/0012069.
23. E.V. Gorbar and I. L. Shapiro, *J. High Energy Phys.* 02, (2003) 021; *J. High Energy Phys.* 06, (2003) 004.
24. E.V. Gorbar and I. L. Shapiro *J. High Energy Phys.* 02, (2004) 060.
25. S.A. Franchino-Viñas, T. de Paula Netto and O. Zanusso, *Universe*, **5**, 67 (2019).
26. S.A. Franchino-Viñas, T. D. Netto, I.L. Shapiro and O. Zanusso, *Phys. Lett. B* **790**, 229 (2019).
27. O. Antipin and B. Melic, *Eur. Phys. J. C* **77**, 583 (2017).
28. T. Markkanen, *Phys. Rev. D* **91**, 124011 (2015).
29. A. Ferreiro and J. Navarro-Salas, *Phys. Rev. D* **102**, 045021 (2020).
30. A. Ferreiro and J. Navarro-Salas, *Phys. Lett. B* **792**, 81 (2019).
31. C. Moreno-Pulido and J. Solà, *Eur. Phys. J. C* **80**, 692 (2020).
32. P. Beltran-Palau, A. del Río, S. Nadal-Gisbert and J. Navarro-Salas, *Phys. Rev. D* **103**, 105002 (2021).
33. D.J. Toms, *J. High Energy Phys.* 05 (2018) 139.

34. V.F. Barra, I.L. Buchbinder, J.G. Joaquim, A.R. Rodrigues, and I. Shapiro, *Eur. Phys. J.C* **79**, 458 (2019).
35. I.L. Buchbinder, A.R. Rodrigues, E.A. dos Reis and I. Shapiro, *Eur. Phys. J.* C **79**, 1002 (2019).
36. D.J. Toms, *Phys. Rev. D* **98**, 025015 (2018).
37. D.J. Toms, arXiv:1906.02515.
38. T. J. Hollowood, *Renormalization Group and Fixed Points in Quantum Field Theory* (Springer, Heidelberg, 2013).
39. S. M. Christensen, *Phys. Rev. D* **14**, 2490 (1976).

Trace anomaly and evaporation of spherical black holes

P. Meda

Dipartimento di Fisica, Università degli Studi di Genova,
Istituto Nazionale di Fisica Nucleare - Sezione di Genova,
Genova, 16146, Italia
E-mail: paolo.meda@ge.infn.it

The evaporation of four-dimensional spherically symmetric black holes is presented in the framework of quantum field theory in curved spacetimes and semiclassical gravity. It is discussed how the evaporation process can be sourced by the presence of the trace anomaly of a massless, conformally coupled scalar field outside the apparent horizon of the black hole.

Keywords: Evaporation; black hole; trace anomaly; semiclassical gravity; Vaidya spacetime

1. Quantum fields in curved spacetimes

From the time of the discovery made by Hawking that black holes can emit radiation,[1,2] black hole evaporation has become of great interest in analyzing the interplay between matter and gravity. A theoretical framework where modelling black hole evaporation is represented by quantum field theory in curved spacetimes,[3] which consists of describing the propagation of a quantum field over a classical curved background. On one side, the background is assumed to be a four-dimensional globally hyperbolic spacetime $(\mathcal{M}, g)$, where $\mathcal{M}$ is a smooth manifold and g a Lorentzian metric with signature $(-,+,+,+)$. On the other side, the quantum field is the fundamental observable, which is viewed as an operator-valued distribution living in a $*$-algebra $\mathcal{A}(\mathcal{M})$. For a real, free Klein-Gordon field ϕ the $*$-algebra is generated by the set of linear smeared fields $\{\phi(f), f \in \mathcal{D}(\mathcal{M})\}$ satisfying the following relations:

$$\phi(Pf) = 0, \qquad \phi(f)^* = \phi(\bar{f}), \qquad [\phi(f), \phi(h)] = \mathrm{i}\Delta(f,h).$$

Here, $P = -\Box_g + m^2 + \xi R$ is the Klein-Gordon operator constructed from the d'Alambertian operator $\Box_g = g^{\mu\nu}\nabla_\mu\nabla_\nu$ on $(\mathcal{M}, g)$, and $\Delta = \Delta_R - \Delta_A$ is the causal propagator of the spacetime, which is defined as the difference between the unique retarded and advanced fundamental solutions of P. A generic quantum state ω is a linear, positive, normalized functional over $\mathcal{A}(\mathcal{M})$, and it is determined by the n-point functions

$$\omega_n(f_1, \dots, f_n) = \langle \phi(f_1) \dots \phi(f_n) \rangle_\omega \in \mathcal{D}'(\mathcal{M}^n).$$

Actually, for Gaussian pure states it is sufficient to known the kernel of the two-point function $\omega_2(x', x) = \langle \phi(x')\phi(x) \rangle_\omega$.

Different from the flat case, where the Poincaré invariance selects a vacuum state uniquely as preferred reference state, in a curved spacetime there is not in principle a preferred notion of "vacuum" dictated by the symmetries of the spacetime. However, a guiding principle to select physically reasonable states in curved spacetimes is that Wick normal-ordered fields like $:\phi^2:$ and $:T_{\mu\nu}:$ should acquire finite expectation values when evaluated on ω. In globally hyperbolic spacetimes, the set of Hadamard states represents the right class where computing normal-ordered fields, because they share the same ultraviolet singular structure of the Minkowski vacuum two-point function $\langle 0|\phi(x')\phi(x)|0\rangle$. Given $\sigma(x', x)$ the one half of the signed geodesic distance between x' and x, the two-point function of a Hadamard state is always locally of the form

$$\omega_2(x', x) = \frac{1}{8\pi^2}\left(\mathscr{H}_{0^+}(x', x) + w(x', x)\right).$$

Here,

$$\mathscr{H}_{0^+}(x', x) = \lim_{\varepsilon\to 0^+} \frac{u(x', x)}{\sigma_\varepsilon} + \sum_{n\geq 0} v_n(x', x)\sigma(x', x)^n \log\left(\frac{\sigma_\varepsilon}{\mu^2}\right) \tag{1}$$

is the universal singularity called Hadamard parametrix viewed in the distributional sense, where $\sigma_\varepsilon(x', x) = \sigma(x', x) + \mathrm{i}\epsilon(t(x') - t(x))$, t is any time function and $\mu > 0$ is a length scale. The coefficients $u(x', x)$, $v_n(x', x)$ are real-valued bi-scalars fixed by the geometry and the equation of motion, while $w(x', x)$ characterizes the state and must be chosen in such a way that ω_2 is a bi-solution of the Klein-Gordon equation. The Hadamard point-splitting regularization consists of subtracting the divergences contained in $\mathscr{H}_{0^+}(x', x)$ before computing the coinciding point limits: it generalizes the normal-ordering prescription in the flat spacetime, and it is also local and covariant, because only local geometry enters in the subtraction. This procedure, and more generally any prescription to obtain Wick observables, is always unique up to a fixed combination of local terms depending on the mass and on the curvature, whose coefficients represent the renormalization freedom of the quantum theory.[4,5]

1.1. *Semiclassical gravity, the quantum stress-energy tensor, and the trace anomaly*

Strictly related to a formulation of quantum fields propagating in curved backgrounds is the semiclassical theory of gravity, which aims to analyze the back-reaction of quantum fields on the spacetime geometry using the semiclassical Einstein equations

$$G_{\mu\nu} = 8\pi \langle :T_{\mu\nu}: \rangle_\omega \tag{2}$$

in units convention $G = c = \hbar = 1$. In this equations, the classical Einstein tensor is equated to the renormalized quantum stress-energy tensor evaluated in a certain quantum state ω. Any pair formed by a spacetime and a quantum state satisfying

these equations constitutes a solution in semiclassical gravity: such a solution models a curved spacetime which incorporates quantum effects due to the propagation of the matter field, and it is expected to describe physical phenomena where the interplay between quantum matter and curvature becomes significant, like in the early Universe or in the vicinity of black holes. On the other hand, any semiclassical model represents an approximation of a more fundamental theory of quantum gravity, and thus it should be only valid in a regime where quantum gravity effects are negligible, and when fluctuations of the stress-energy tensor $:T_{\mu\nu}:$ are small.

A conserved quantum stress-energy tensor which can enter Eqs. (2) can be constructed in perturbative interacting quantum field theory in curved spacetimes according to the Hadamard point-splitting procedure.[5–7] For the stress-energy tensor of a free Klein-Gordon field

$$T_{\mu\nu} = \nabla_\mu\phi\nabla_\mu\phi - \frac{1}{2}g_{\mu\nu}\left(\nabla_\rho\phi\nabla^\rho\phi + m^2\phi^2\right) + \xi\left(G_{\mu\nu}\phi^2 - \nabla_\mu\nabla_\nu\phi^2 + g_{\mu\nu}\nabla_\rho\nabla^\rho\phi^2\right),$$

the Hadamard renormalization procedure yields the following expectation value:

$$\langle :T_{\mu\nu}: \rangle_\omega (x) = \lim_{x'\to x} D_{\mu\nu} w(x', x),$$

where

$$\begin{aligned} D_{\mu\nu} =& (1-2\xi) g_\mu^{\mu'} \nabla_{\mu'}\nabla_\nu - 2\xi\nabla_\mu\nabla_\nu + \xi G_{\mu\nu} + \\ & g_{\mu\nu}\left\{2\xi\Box_g + \left(2\xi - \frac{1}{2}\right) g_\rho^{\rho'}\nabla_{\rho'}\nabla^\rho - \frac{1}{2}m^2\right\} + \frac{1}{3}g_{\mu\nu}P \end{aligned}$$

is the bi-differential operator which realizes the point-splitting of $T_{\mu\nu}$, and $w(x', x)$ is the smooth part of the Hadamard state ω. In this expression, unprimed and primed indices denote covariant derivatives in the points x and x', respectively, while $g_\nu^{\nu'}$ is the bitensor of parallel transport. The last term proportional to $g_{\mu\nu}$, which vanishes in the classical case, must be added to obtain $\nabla^\mu \langle :T_{\mu\nu}: \rangle_\omega (x) = 0$. However, any construction of a covariantly conserved expectation value of $:T_{\mu\nu}:$ leads to the violation of the classical conformal invariance of $T_{\mu\nu}$, and it gives rise to the so-called trace anomaly; namely, the trace of the quantum stress-energy tensor does not vanish for a massless, conformally coupled field, i.e., when $m = 0$ and $\xi = 1/6$, different from its classical counterpart. This anomalous trace does not depend on the choice of the quantum state, but it is a local contribution depending on the geometry and on the linear equation of motion of the matter field. In the Hadamard point-splitting regularization scheme and for a free scalar field ϕ, it is proportional to the Hadamard coefficient v_1 evaluated in the coinciding limit, and it reads in four dimensions as

$$\langle :T_\rho{}^\rho: \rangle_\omega = \lambda\left(C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} + R_{\mu\nu}R^{\mu\nu} - \frac{1}{3}R^2\right), \qquad \lambda = \frac{1}{2880\pi^2}, \tag{3}$$

where $C_{\alpha\beta\gamma\delta}$ is the Weyl tensor, $R_{\mu\nu}$ is the Ricci tensor and R is the Ricci scalar. The $\Box_g R$ usually appearing in $\langle :T_\rho{}^\rho: \rangle_\omega$ was removed by a judicious choice of the

renormalization constants, according to the previous discussion about local and covariant Wick observables in curved spacetimes. In the case of $\langle{:}T_\rho{}^\rho{:}\rangle_\omega$, the renormalization scalar freedom is[5]

$$Q = \tilde{\beta}_1 m^4 + \tilde{\beta}_2 m^2 R + \tilde{\beta}_3 R^2 + \tilde{\beta}_4 \Box_g R,$$

where $\tilde{\beta}_i \in \mathbb{R}$ are renormalization constants. Hence, in some special cases like the massless, conformally coupled field, the higher-order derivative term proportional to $\Box_g R$ can be always cancelled by choosing properly $\tilde{\beta}_4$, thus meeting the requirement that curvature terms do not contain non-classical higher order derivatives of the metric.[8]

2. Evaporation of spherically symmetric black holes

A model of evaporation of black holes which incorporates the back-reaction of the matter fields on the geometry can be fully developed in the framework of four-dimensional spherically symmetric spacetimes. In this class of geometries, invariants like the mass of the black hole and the area of the horizon can be provided thanks to the spherical symmetry[9]: the area of each two-sphere is $\mathcal{A} = 4\pi r^2$, where r is the radius, while the mass

$$m = \frac{r}{2}\left(1 - \nabla^\rho r \nabla_\rho r\right)$$

represents the Misner-Sharp energy enclosed inside the sphere. The metric of every spherically symmetric spacetime can be always represented in double-null coordinates with respect to the null normal directions ∂_V and ∂_U as

$$\mathrm{d}s^2 = -2A(V,U)\mathrm{d}V\mathrm{d}U + r^2(V,U)\mathrm{d}\Omega, \tag{4}$$

where $\mathrm{d}\Omega = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2$ denotes the line-element of the two-sphere of unit radius $\mathbb{S}^2$. The orientation of the spacetime can be fixed in such a way that $A(V,U) > 0$ and at spatial infinity $\partial_V r > 0$, $\partial_U r < 0$. To study evaporation, another convenient choice of parametrization is the Bardeen-Vaidya metric[10]

$$\mathrm{d}s^2 = -\mathrm{e}^{2\Psi(v,r)}\left(1 - \frac{2m(v,r)}{r}\right)\mathrm{d}v^2 + 2\mathrm{e}^{\Psi(v,r)}\mathrm{d}v\mathrm{d}r + r^2\mathrm{d}\Omega, \tag{5}$$

where v denotes the advanced time coordinate and parametrizes any ingoing radial null geodesic.

The notion of horizon in spherically symmetric black holes is encoded in the idea that no light rays can escape from this hypersurface, whereas they can fall inside. Namely, it is a future outer local trapping horizon.[11] To formulate this idea on mathematical grounds, one introduces a pair of null radial geodesic congruences with respect to the outgoing and ingoing directions ∂_V and ∂_U, respectively, and the relative expansion parameters

$$\theta_+ = \frac{2}{Ar}\partial_V r, \qquad \theta_- = \frac{2}{Ar}\partial_U r.$$

Each expansion parameter measures the rate of variation of the cross-sectional area of the geodesic congruence: $\theta_{\pm} = A^{-1}\partial_{V,U}\log(\mathcal{A})$, and it determines if the congruence is expanding or focusing along that null direction. Thus, the apparent horizon of a spherically symmetric black hole is defined as a three-manifold $\mathcal{H}$ foliated by closed surfaces such that

$$\theta_+ \overset{\mathcal{H}}{=} 0, \qquad \theta_- \overset{\mathcal{H}}{<} 0, \qquad \partial_U\theta_+ \overset{\mathcal{H}}{<} 0, \tag{6}$$

and it is located in $\mathcal{H} : r - 2m = 0$ (the subscript $\mathcal{H}$ labels the evaluation on the apparent horizon). The third condition captures the property that this trapping horizon is outer and it is referred to a black hole. In coordinates (v, r), the hypersurface $\mathcal{H}$ is a one-dimensional line $r_{\mathcal{H}}(v)$ in the plane (v, r), and hence the mass of an evaporating black hole is defined by the relation

$$M(v) = m\left(v, r_{\mathcal{H}}(v)\right) = \frac{r_{\mathcal{H}}(v)}{2}. \tag{7}$$

On spherically symmetric black holes, the mechanism of evaporation can be described in terms of the variation of the mass

$$\Delta M = \int_{\delta\mathcal{H}} \mathrm{d}m, \tag{8}$$

along a small portion $\delta\mathcal{H}$ of the apparent horizon. In coordinates (v, r), $\delta\mathcal{H}$ corresponds to the line enclosed between two points $P = (v_P, r_{\mathcal{H}}(v_P))$ and $Q = (v_Q, r_{\mathcal{H}}(v_Q))$ in the (v, r) plane, and the dynamics of $M(v)$ is fully determined by the rate of evaporation $\dot{M}(v) = \partial_v M(v)$. If one considers a generic stress-energy tensor $T_{\mu\nu}$, then the evolution of $M(v)$ is governed by the vv-component of the (semiclassical) Einstein equations evaluated on $\mathcal{H}$, which reads

$$\dot{M}(v) = \mathcal{A}_{\mathcal{H}}(v)T_{vv}(v, r_{\mathcal{H}}(v)), \tag{9}$$

where $\dot{M}(v)$ and $\mathcal{A}_{\mathcal{H}} = 4\pi r_{\mathcal{H}}^2$ denote the rate of evaporation and the area of the horizon, respectively. Thus,

$$\Delta M = 4\pi \int_{v_P}^{v_Q} \frac{M}{\kappa} T_{vv}(r_{\mathcal{H}})\mathrm{d}v. \tag{10}$$

As $T_{vv} \overset{\mathcal{H}}{=} T_{VV}$, both $\dot{M}(v)$ and ΔM depend on the ingoing flux of matter which is falling inside the black hole from the horizon. Whenever quantum matter is involved in the back-reaction process, it may happen that the quantum stress-energy tensor $\langle{:}T_{\mu\nu}{:}\rangle_\omega$ violates the classical null energy condition on the horizon, and hence the ingoing quantum component $\langle{:}T_{VV}{:}\rangle_\omega$ may be negative on $\delta\mathcal{H}$. In the end, the presence of a negative ingoing flux on the horizon makes the black hole mass to evaporate with a negative rate $\dot{M}(v) < 0$.

2.1. *The static black hole*

The vacuum spherically symmetric spacetime is the Schwarzschild solution of the Einstein equations according to Birkhoff's theorem, and it describes a static spherical black hole of constant mass M.[12] A double-null parametrization for arbitrary $U, V > 0$ is the maximally extension of the Schwarzschild eternal black hole in the so-called Kruskal-Szekeres coordinates

$$\mathrm{d}s^2 = -\frac{32M^3}{r}\mathrm{e}^{-r/2M}\mathrm{d}V\mathrm{d}U + r^2(V,U)\mathrm{d}\Omega,$$

where $r(V,U)$ is implicitly defined by

$$UV = \left(1 - \frac{r}{2M}\right)\mathrm{e}^{r/2M}.$$

On the other hand, the Bardeen-Vaidya metric corresponds to the advanced Eddington-Finkelstein coordinates, with $\Psi(v,r) = 0$ and $m(v,r) = M$.

The static symmetry of the Schwarzschild spacetime reflects the existence of a Killing vector field $\xi = \frac{1}{4M}(V\partial_V - U\partial_U)$, which is unit timelike at infinity and it is hypersurface orthogonal. Due to the static symmetry, the apparent horizon coincides with the null event horizon of the black hole, which is located in the Schwarzschild radius $r = 2M$ (see Figure 1[a]). The surface gravity of the black hole is defined by the equation $\nabla^\mu\xi^2 = -2\kappa\xi^\mu$ and it constant along each null generator of ξ. Physically, the surface gravity represents the gravitational acceleration detected along the black hole horizon, and it is strictly positive on the event horizon.

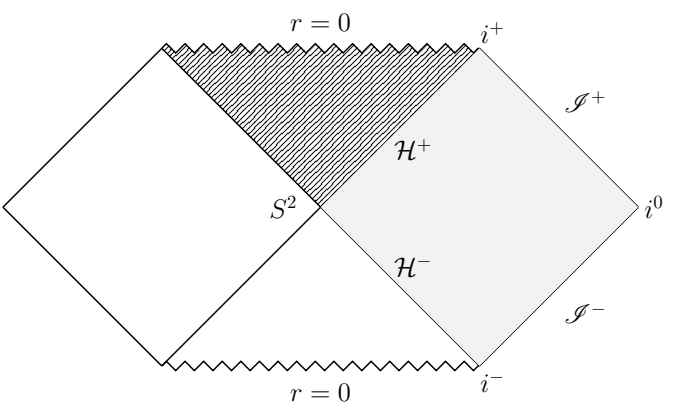


Fig. 1. Penrose diagram of the maximally extended Schwarzschild spacetime. The bifurcated Killing horizon $\mathcal{H} = \mathcal{H}^+ \cup \mathcal{H}^-$ associated to ξ is composed by two null surfaces $\mathcal{H}^\pm$ located in $U = 0$ and $V = 0$, respectively. The event horizon $\mathcal{H}^+$ corresponds to the future branch of $\mathcal{H}$, and it separates the outer region (grey region) from the black hole region (dashed region); the null surfaces intersects each other in the sphere of bifurcation S^2 located in $U = 0$, $V = 0$.

[a]Penrose diagrams were made in TikZ and are based on the Latex code posted in StackExchange.[13]

According to Hawking's works, a static black hole can emits thermal radiation at infinity in form of a flux of quantum particles, if the back-reaction is not taken into account and after assuming a vacuum state in the asymptotic past. The temperature of the radiation is the so-called Hawking temperature and it is proportional to the surface gravity of the black hole: in the units convention adopted in this paper it reads

$$T_H = \frac{\kappa}{2\pi} = \frac{1}{8\pi k_B M},$$

with k_B being the Boltzmann constant. The correspondence between the temperature of the radiation and the surface gravity lets to interpret the first law of black hole mechanics

$$\mathrm{d}M = \frac{\kappa}{8\pi}\mathrm{d}\mathcal{A} + \Omega \mathrm{d}J$$

from a thermodynamic point of view, with M and $\mathcal{A}/4$ playing the role of energy and entropy of the black hole, respectively. More recently,[14] it was shown that in case of a static black hole formed in a gravitational collapse, the Hawking radiation with temperature T_H always appears at future infinity whenever the state ω is a Hadamard state approaching the ground state near spatial infinity; for instance, in the case of the Unruh state,[15,16] which is regular on the horizon and respects the spherical symmetries of the spacetime.

Thus, the power of the radiation emitted from the black hole at future infinity is regarded as describing the evaporation of the black hole, because it can be instantaneously equated to the rate of loss of mass in the adiabatic approximation, namely neglecting how the spacetime is dynamically changing during the evaporation. Starting from this assumption, the negative ingoing flux on the horizon, that is, $\langle{:}T_{vv}{:}\rangle_\omega \overset{\mathcal{H}}{=} \langle{:}T_{VV}{:}\rangle_\omega$, can be related to the positive outgoing flux $\langle{:}T_{uu}{:}\rangle_\omega$ detected at infinity, where u denotes the retarded Eddington-Finkelstein coordinate parametrizing outgoing null geodesics. In the Unruh state, it holds that

$$\frac{\dot{M}(v)}{4\pi} = \lim_{r\to r_{\mathcal{H}}} \left(\langle{:}T_{VV}{:}\rangle_\omega\, r^2\right) = -\lim_{r\to\infty} \left(\langle{:}T_{uu}{:}\rangle_\omega\, r^2\right) = -\frac{L_{\mathcal{H}}}{4\pi}, \tag{11}$$

where $L_{\mathcal{H}} > 0$ denotes the luminosity of the black hole, namely the radial flux of particles measured by a static observer at large distance from the black hole.[17] Indeed, Eq. (11) loses meaning outside the adiabatic approximation, since the rate $\dot{M}(v)$ always vanishes in case of constant mass. It turns out, in fact, that the Schwarzschild spacetime is never a solution of the semiclassical Einstein equations, and hence an eternal black hole cannot be in equilibrium with the back-reaction of any quantum matter field in the semiclassical approach.

2.2. *Dynamical black holes*

The direct correspondence given in Eq. (11) between the ingoing negative flux and the positive outgoing flux holds in Schwarzschild spacetime thanks to the existence

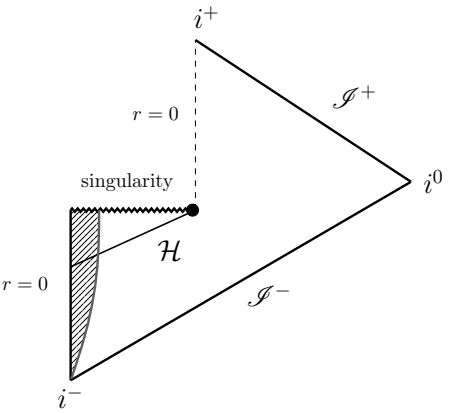


Fig. 2. Penrose diagram of a Schwarzschild gravitational collapse followed by evaporation and leaving an empty space at the end of the evolution. The black dot represents the end of the evaporation, and the dashed line denotes the new regular centre of the spacetime after the evaporation.[2]

of a global Killing field which is timelike outside the event horizon, but it is missing in arbitrary dynamical backgrounds. In this case, the mechanism of evaporation and the following emission of power radiated should be dealt separately as causally related processes: from one hand, evaporation is a local effect induced by the negative ingoing flux and affecting the horizon; on the other hand, the expected Hawking radiation is associated to the positive outgoing flux emitted to infinity, and it depends globally on the entire history of spacetime.

Moreover, the adiabatic approximation is unable to describe evaporations when the spacetime is treated with its full dynamics, since it does not give informations about the evolution of the apparent horizon. In this respect, the Penrose diagram shown in Figure 2 should be modified in case of dynamical backgrounds outside that approximation, because the dynamics of $\mathcal{H}$ depends sharply on the negative flux $\langle{:}T_{VV}{:}\rangle_\omega$, and the causal structure of the spacetime may be significantly modified in the back-reaction process. Conversely, a semiclassical model of evaporation should necessarily rely on the semiclassical Einstein equations (2) to study the variation of the mass $M(v)$ induced by the back-reaction of the quantum matter content, without further assumptions.

Contrary to Schwarzschild spacetimes modelling static black holes and event horizons, a general spherically symmetric spacetime describes a dynamical black hole, whose evolution is dictated by the dynamics of the apparent horizon (6). To define properties of spherically symmetric black holes like the surface gravity, there is a preferred choice of a vector field, which replaces the standard Killing vector field in stationary spacetimes. It is the Kodama vector[18] $K = g^{-1}(*_n \mathrm{d}r)$, where $*_n$ denotes the Hodge operator in the two-dimensional spacetime normal to the two-sphere. This vector is timelike in the region outside the horizon, it becomes lightlike on $\mathcal{H}$, and eventually it is spacelike in the interior of the black hole. It also reduces

to the Killing vector ξ in stationary spacetimes. Furthermore, it is divergenceless: $\nabla_\mu K^\mu = 0$, and it can be used to define conserved currents $T_{\mu\nu}K^\nu$ given any symmetric tensor $T_{\mu\nu}$ which respects the spherical symmetry. Finally, from the definition of K it holds that

$$\mathscr{L}_K K_\mu = K^\nu \left(\nabla_\nu K_\mu - \nabla_\mu K_\nu\right) \overset{\mathcal{H}}{=} \pm\kappa K_\mu, \tag{12}$$

where $\mathscr{L}_K$ denotes the Lie derivative along K, and

$$\kappa \doteq \frac{1}{2}\Box_n r, \tag{13}$$

where the d'Alambertian operator is evaluated in the two-dimensional normal spacetime. Namely, this quantity corresponds to the definition of the surface gravity for a dynamical black hole and it reduces to the standard one in the case of a Killing vector field. In particular, $\partial_U \theta_+ < 0$ implies that $\kappa > 0$ on $\mathcal{H}$.

In analogy with the static case, Hayward showed[9] that the four laws of black hole mechanics can be generalized in presence of spherical symmetry to apparent horizons using the properties of the Kodama vector and the definition of the surface gravity given in Eq. (13). In this class of spacetimes, the first law reads

$$m' = \frac{\kappa}{8\pi}\mathcal{A}' + wV', \tag{14}$$

where the prime denotes the derivative along the horizon, i.e., $f' = \ell \cdot \nabla f$, and ℓ is any tangent vector to $\mathcal{H}$. Here, the second contribution in the right-hand side represents the variation of the work done by the matter on the horizon, where $V = \frac{4}{3}\pi r^3$ and $w = A^{-1}T_{UV}$. Furthermore, the same thermodynamic interpretation already seen in the static case can be given in terms of the Kodama surface gravity, with $\kappa/2\pi$ playing the role of dynamical local temperature of the apparent horizon with respect to the time parameter defined of the Kodama vector. This correspondence was shown in the dynamical case by studying the tunnelling probability of quantum matter across apparent horizons using the Hamilton-Jacobi method,[19,20] and by analyzing the scaling limit of the Hadamard two-point function on such horizons.[21]

According the previous section, *cf.* Eqs. (9) and (10), evaporation of spherically symmetric black holes is fully determined by the ingoing flux located on the horizon. However, in the absence of further symmetries it is very challenging to evaluate the component $\langle{:}T_{VV}{:}\rangle_\omega$, and, more generally, the renormalized quantum stress-energy tensor $\langle{:}T_{\mu\nu}{:}\rangle_\omega$, in a state ω which was a vacuum state in the past. On the other hand, the variation of the mass of the black hole can be constrained by the matter-geometry content in the causal past and outside the black hole horizon.[22] This is accomplished by applying the divergence theorem (Stokes' theorem) to the currents J_r and J_k obtained by contracting the stress-energy tensor with the gradient ∇r and with the Kodama vector K, respectively. The domain over which the divergence theorem is spherically symmetric and it has the form $\mathrm{D} \times \mathbb{S}^2$, where D is a suitable region in the two-dimensional normal spacetime intersecting a portion of

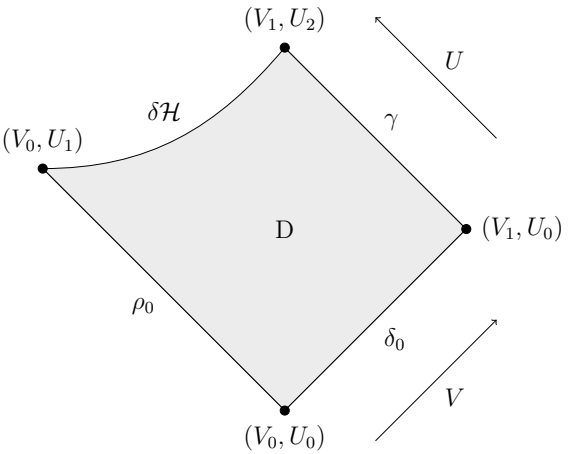


Fig. 3. Picture of the domain D. For spacelike $\delta\mathcal{H}$, $\mathrm{D} \times \mathbb{S}^2 = J^+((V_0, U_0) \times \mathbb{S}^2) \cap J^-(\delta\mathcal{H} \times \mathbb{S}^2)$, with $U_2 < U_1$. The points $P = (V_0, U_1)$ and $Q = (V_1, U_2)$ denote the extreme values of $\delta\mathcal{H}$ in the (V, U) plane. In describing evaporation, the initial conditions are posed on $S_0 = \{(V, U, \theta, \varphi) \in \mathcal{M} \mid V = V_0,\ U = U_0\}$ and on the curves ρ_0 and δ_0 of the boundary $\partial\mathrm{D}$.[22]

the apparent horizon. A representation of D for a spacelike portion of the horizon is pictured in Figure 3.

If one considers a stress-energy tensor $T_{\mu\nu}$ which respects the spherical symmetry and which satisfies vacuum-like initial condition at the boundary of $J^+(S_0)$, namely

$$T_{\mu\nu}(p) = 0, \qquad p \in \partial J^+(S_0), \tag{15}$$

then applying the divergence theorem to J_r yields

$$\Delta M = -\left(\mathcal{S} + \mathcal{W}\right), \tag{16}$$

where

$$\mathcal{S} = 2\pi \int_{\mathrm{D}} \nabla \cdot J_r \mathrm{d}V_{\mathrm{D}}, \tag{17}$$

$$\mathcal{W} = 4\pi \int_{\gamma} \frac{T_{UV} r^2}{A} (-\partial_U r) \mathrm{d}U. \tag{18}$$

Here, $\mathrm{d}V_{\mathrm{D}} \doteq A r^2 \mathrm{d}V \wedge \mathrm{d}U$ is the volume form on D in the (V, U) plane. Eq. (17) denotes the matter source inside the domain D which influences the variation of the mass, while Eq. (18) is related to the work done by the matter when evaluated on $\mathcal{H}$, in view of Hayward's first law (14). Moreover, applying the divergence theorem to J_K in spacetimes where $\mathcal{W} = 0$ yields

$$\left.\frac{\Delta M}{4\pi}\right|_{\mathcal{W}=0} = -\int_{\gamma} \frac{T_{UU} r^2}{A} \partial_V r \mathrm{d}U. \tag{19}$$

Namely, in the quantum case the variation of the mass is related to the outgoing flux $\langle{:}T_{UU}{:}\rangle_\omega$ across γ, which can interpreted as Hawking radiation sourced by $\mathcal{S}$ and emitted from the evaporating black hole to infinity. In this view, the relation (19) generalizes Eq. (11) to more general non-static, spherically symmetric black holes outside the adiabatic approximation.

2.3. *The trace anomaly as source of evaporation*

It is well known that the evaporation of a black hole, viewed as emission of Hawking radiation at infinity, can be ascribed to the presence of an anomalous trace in the stress-energy tensor of a quantum matter field in the vicinity of the horizon.[23–26] However, such analysis are often provided after making some approximations on the spacetime, like the adiabatic one. On the contrary, a full semiclassical approach to study the evaporation sourced by the trace anomaly should only rely on the semiclassical Einstein equations in four dimensions, without making further assumptions or referring to asymptotic effects which need global informations about the spacetime.

Based on the results presented in the previous section, the variation of the black hole mass can be related to the source term $\mathcal{S}$ given in Eq. (17), and to the boundary term $\mathcal{W}$ given in Eq. (18). Thus, it was proved[22] that the evaporation of a four-dimensional dynamical black hole can be induced by the trace anomaly (3) of a massless, conformally coupled scalar field affecting the horizon from the outside of the black hole, provided a weighted version of Eq. (16). This result is obtained using only the semiclassical Einstein equations (2), once the initial condition (15) is assumed on the state ω and a quantum averaged energy condition on $\langle{:}T_{UV}{:}\rangle_\omega$ is posed. Such a condition is of the form

$$\int_\gamma \frac{\langle{:}T_{UV}{:}\rangle_\omega\, r^2}{A} f(V,U) A \mathrm{d}U \geq 0, \tag{20}$$

where γ is one of the boundary hypersurface of the domain $\mathrm{D} \times \mathbb{S}^2$ (see Figure 3), and $f(V,U)$ is a certain strictly positive exponential function constructed from the geometry of the spacetime, whose role is to tame the contributions inside $\langle{:}T_\rho{}^\rho{:}\rangle_\omega$ which do not contribute to a negative variation of the mass. The form of the function f is similar to the ones entering usual quantum averaged weak energy conditions, which were shown to be valid both in flat and curved spacetimes for several values of the coupling-to-curvature parameter.[27]

In principle, it is difficult to verify if such an energy condition can be fulfilled by the quantum collapsing matter during the evaporation process, because the lack of a quantum state in spherically symmetric spacetimes which was a vacuum in the past prevents to evaluate $\langle{:}T_{\mu\nu}{:}\rangle_\omega$ on and outside the horizon. However, one can expect that this condition is satisfied in an approximate way by a background geometry which is not too different from the one obtained in a classical model of gravitational collapse. Indeed, classical solutions are approximately valid in semiclassical gravity,

because quantum corrections are usually small outside the horizon, and thus one can assume that solutions of the full semiclassical equations maintain their classical form outside the horizon. To provide a simple application, consider the Vaidya spacetime

$$\mathrm{d}s^2 = -\left(1 - \frac{2M(v)}{r}\right)\mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r + r^2\mathrm{d}\Omega,$$

which describes a null radiating spherically symmetric star with time-dependent mass $M(v)$. If one supposes that the quantum state fulfils $\langle{:}T_{UV}{:}\rangle_\omega = 0$, namely it makes the above metric a solution of Eqs. (2), then the negative ingoing flux on the horizon induced by the anomalous trace (3) can be evaluated explicitly, and it reads

$$\langle{:}T_{VV}{:}\rangle_\omega \, r^2 \overset{\mathcal{H}}{=} -\frac{3\lambda}{40M(v)^2}.$$

Hence,

$$\dot{M}(v) = -\frac{3\pi\lambda}{10M(v)^2}, \qquad \frac{\mathrm{d}S_{\mathcal{H}}(v)}{\mathrm{d}v} = -\frac{12\pi^2\lambda}{5M(v)},$$

where $S_{\mathcal{H}} = \mathcal{A}_{\mathcal{H}}/4$ denotes the Wald-Kodama entropy.

3. Conclusions

The evaporation of four-dimensional spherically symmetric black holes in semiclassical gravity can be fully ascribed to the presence of a negative ingoing flux on the horizon, which determines the global behaviour of the decreasing mass of the black hole. Whereas the evaporation of any static black hole can be described directly in terms of Hawking radiation and black hole luminosity, for dynamical black holes the variation of the mass can be constrained by the quantum matter content outside and in the causal past of the black hole, without directly referring to any expected radiation detected at infinity. When conformally coupled quantum scalar fields are involved, the evaporation can be sourced by the trace anomaly of the quantum stress-energy tensor, assuming vacuum-like initial conditions in the past, and an auxiliary quantum energy condition outside the horizon. It is, however, very challenging to extend these results to more generic quantum matter contents, due to the difficulties in evaluating the renormalized stress-energy tensor in a quantum state. In order to overcome this issue, a semiclassical analysis which takes advantage of some quantum energy conditions similar to the one presented in Eq. (20) might be viable in the future.

Acknowledgments

The author would like to thank Simone Roncallo for his careful reading of an earlier version of this proceeding.

References

1. S. W. Hawking, Black hole explosions?, *Nature* **248**, 30 (1974).
2. S. W. Hawking, Particle creation by black holes, *Communications in Mathematical Physics* **43**, p. 199 (1975).
3. S. Hollands and R. M. Wald, Quantum fields in curved spacetime, *Physics Reports* **574**, 1 (2015).
4. S. Hollands and R. M. Wald, Local wick polynomials and time ordered products of quantum fields in curved space-time, *Communications in Mathematical Physics* **223**, 289 (2001).
5. S. Hollands and R. M. Wald, Conservation of the stress tensor in perturbative interacting quantum field theory in curved spacetimes, *Reviews in Mathematical Physics* **17**, 227 (2005).
6. V. Moretti, Comments on the stress-energy tensor operator in curved spacetime, *Communications in Mathematical Physics* **232**, 189 (2003).
7. T. P. Hack, *Cosmological Applications of Algebraic Quantum Field Theory in Curved Spacetimes* (Springer International Publishing, 2016).
8. R. M. Wald, The back reaction effect in particle creation in curved spacetime, *Communications in Mathematical Physics* **54**, 1 (1977).
9. S. A. Hayward, Unified first law of black hole dynamics and relativistic thermodynamics, *Classical and Quantum Gravity* **15**, 3147 (1998).
10. J. M. Bardeen, Black holes do evaporate thermally, *Physical Review Letter* **46**, p. 382 (1981).
11. S. A. Hayward, General laws of black-hole dynamics, *Physical Review D* **49**, 6467 (1994).
12. R. M. Wald, *General Relativity* (Chicago University Press, Chicago, USA, 1984).
13. Nice scientific pictures show off `https://tex.stackexchange.com/questions/158668/nice-scientific-pictures-show-off/209201`, [Online; 05-February-2014].
14. K. Fredenhagen and R. Haag, On the derivation of hawking radiation associated with the formation of a black hole, *Communications in Mathematical Physics* **127**, p. 273 (1990).
15. W. G. Unruh, Notes on black-hole evaporation, *Physical Review D* **14**, 870 (1976).
16. C. Dappiaggi, V. Moretti and N. Pinamonti, Rigorous construction and Hadamard property of the Unruh state in Schwarzschild spacetime, *Advances in Theoretical and Mathematical Physics* **15**, 355 (2011).
17. P. Candelas, Vacuum polarization in schwarzschild space-time, *Physical Review D* **21**, p. 2185 (1980).
18. H. Kodama, Conserved energy flux for the spherically symmetric system and the backreaction problem in the black hole evaporation, *Progress of Theoretical Physics* **63**, p. 1217 (1980).
19. R. Di Criscienzo, M. Nadalini, L. Vanzo, S. Zerbini and G. Zoccatelli, On the hawking radiation as tunneling for a class of dynamical black holes, *Physics Letters B* **657**, 107 (2007).
20. L. Vanzo, G. Acquaviva and R. Di Criscienzo, Tunnelling methods and hawking's radiation: achievements and prospects, *Classical and Quantum Gravity* **28**, p. 183001 (2011).
21. F. Kurpicz, N. Pinamonti and R. Verch, Temperature and entropy-area relation of quantum matter near spherically symmetric outer trapping horizons, *Letters in Mathematical Physics* **111** (2021).

22. P. Meda, N. Pinamonti, S. Roncallo and N. Zanghì, Evaporation of four-dimensional dynamical black holes sourced by the quantum trace anomaly, *Classical and Quantum Gravity* **38**, p. 195022 (2021).
23. P. C. W. Davies, S. A. Fulling and W. G. Unruh, Energy momentum tensor near an evaporating black hole, *Physical Review D* **13**, 2720 (1976).
24. S. M. Christensen and S. A. Fulling, Trace anomalies and the hawking effect, *Physical Review D* **15**, 2088 (1977).
25. R. Balbinot and A. Barletta, The back reaction and the evolution of quantum black holes, *Classical and Quantum Gravity* **6**, 195 (1989).
26. F. Canfora and G. Vilasi, Trace anomaly and black holes evaporation (2003), arXiv:gr-qc/0302036.
27. E. A. Kontou and K. Sanders, Energy conditions in general relativity and quantum field theory, *Classical and Quantum Gravity* **37**, p. 193001 (2020).

On decay of the false Unruh vacuum

A. Shkerin

William I. Fine Theoretical Physics Institute, School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA
E-mail: ashkerin@umn.edu

We provide a method to calculate the rate of false vacuum decay induced by a black hole. The method uses complex tunneling solutions and consistently takes into account the structure of quantum vacuum in the black hole metric via boundary conditions. We illustrate the technique by computing the rate of decay of the Unruh vacuum in a toy two-dimensional model of a scalar field with inverted Liouville potential in a background of a dilaton black hole.

Keywords: Black holes; false vacuum decay; Unruh vacuum; quantum fields and gravity

1. Introduction and motivation

There has been significant interest recently in the problem of black hole catalysis of vacuum decay.[1–15] A part of the motivation driving these studies comes from the fact that the Standard Model electroweak vacuum may not be absolutely stable.[16–22] Requiring the lifetime of the vacuum exceed the age of the Universe puts constraints on the parameters of the Standard Model, of its possible extensions, and of systems and environments which catalyze vacuum decay, including black holes.[23]

The catalyzing effect of a black hole (BH) is two-fold. First, it is a local spacetime inhomogeneity. Hence, as for any impurity, one expects for the BH to facilitate nucleation of bubbles of true vacuum in its vicinity. Second, associated with BHs is the quantum vacuum effect manifesting itself as Hawking radiation. As any field excitation, this radiation is expected to increase the decay rate. For small enough BHs, the catalyzing effects due to curved geometry and quantum vacuum may be equally important.

In the semiclassical regime, the vacuum decay is described by a (complex) classical solution of field equations. The solution — bounce — saturates the amplitude of transition from the false to the true vacuum regions. It is important to note that the vacuum is defined not only by classical field expectation values but also by the state of quantum fluctuations around these values. Different false vacuum states reveal themselves through different boundary conditions imposed on the bounce.

It is well-known how to look for the bounce solution in equilibrium systems.[24–27] Realistic BHs, however, are not in thermal equilibrium with their environment. In particular, this is true for hypothetical small primordial BHs that could exist in the early Universe but are completely evaporated by now.[28] It is desirable to have an approach to vacuum decay that can handle such non-equilibrium systems as a hot isolated BH in a comparatively low energy background. This system is described

by the Unruh state.[29] In the Unruh vacuum, the BH emits Hawking radiation but does not receive anything from asymptotic infinity.

In this report, based on Ref. 30, we outline the recent progress made in the study of BH catalysis of false vacuum decay. In Sec. 2 we provide a general method to calculate the decay probability of a false vacuum state characterized by an arbitrary (Gaussian) density matrix. The method employs complex tunneling solutions and, when applied to a BH, allows one to discriminate between different vacuum states associated with the BH. In particular, it allows one to compute the rate of decay of the Unruh vacuum for which equilibrium methods are not applicable. In Sec. 3 we discuss the application of the method to a toy two-dimensional model of vacuum decay in the presence of a BH. Sec. 4 is dedicated to conclusions and outlook.

2. The method

To illustrate the method, we consider a real scalar field in the background of a dilaton BH in two dimensions.[31] The only region that is relevant for vacuum decay is the external (physical) patch of the BH spacetime.[30] It is convenient to introduce tortoise coordinates (t, x) covering this patch. Then the metric is read off from the line element

$$ds^2 = \Omega(x)(-dt^2 + dx^2) \,, \tag{1}$$

where the conformal factor takes the form

$$\Omega(x) = \frac{1}{1 + e^{-2\lambda x}} \,. \tag{2}$$

The parameter λ is related to the BH temperature $T_{\rm BH}$ as $\lambda = 2\pi T_{\rm BH}$. For the sake of brevity, we will refer to λ itself as temperature. The horizon is located at $x \to -\infty$. Near the horizon, the conformal factor behaves as $\Omega(x) \approx e^{2\lambda x}$ and the metric (1) is approximately the Rindler one.[32] The physical size of the near-horizon region is

$$l_h \sim \int_{-\infty}^{0} dx \sqrt{\Omega} \sim \frac{1}{\lambda} \,. \tag{3}$$

In the opposite limit, $x \to \infty$, the metric is asymptotically flat. The vacuum boundary conditions are imposed in the remote past, $t \to -\infty$, on both sides of the physical patch, $x \to \pm\infty$.

Next, we take the following scalar field action

$$S = \frac{1}{\mathrm{g}^2} \int d^2x \sqrt{-g} \left(-\frac{1}{2} g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - \frac{m^2 \varphi^2}{2} - V_{\rm int}(\varphi) \right) \,. \tag{4}$$

Here $V_{\rm int}(\varphi)$ is the interaction part of the tunneling potential chosen such that the false vacuum is located at $\varphi = 0$, and $V_{\rm int}(0) = 0$. For convenience, we factored out the small coupling constant g that controls the semiclassical expansion. In the

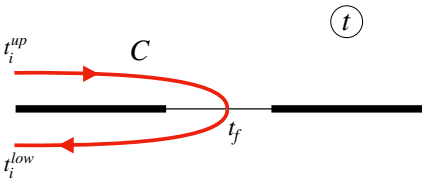


Fig. 1. Contour C in the complex time plane for the calculation of the vacuum decay probability. Black lines denote the branch-point singularities of the bounce in a theory with scalar potential unbounded from below.

background specified by eqs. (1) and (2), the action (4) becomes

$$S = \frac{1}{\mathrm{g}^2} \int dt dx \left(-\frac{1}{2} \eta^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - \frac{1}{2} \Omega m^2 \varphi^2 - \Omega V_{\mathrm{int}}(\varphi) \right) , \tag{5}$$

where $\eta^{\mu\nu} = \mathrm{diag}(-1, 1)$ is the Minkowski metric. We see that the dependence on the background is contained entirely in the (position-dependent) potential of the field φ.

Let us study linear perturbations around the false vacuum $\varphi = 0$. To this end, we neglect the self-interaction part of the potential V_{int}, and decompose φ using a complete set of positive- and negative-frequency modes:

$$\varphi_\omega^+(t, x) = f_\omega(x) e^{-i\omega t} , \quad \varphi_\omega^-(t, x) = f_\omega^*(x) e^{i\omega t} , \quad \omega > 0 . \tag{6}$$

The equation for f_ω follows from the linearized field equation for φ and reads

$$-f_\omega'' + U_{\mathrm{eff}}(x) f_\omega = \omega^2 f_\omega , \tag{7}$$

where prime denotes derivative with respect to x. This is the Schrödinger equation with the potential

$$U_{\mathrm{eff}}(x) = m^2 \Omega . \tag{8}$$

Next, we turn to the construction of the bounce solution. Bounce φ_{b} is a regular solution of fields equations that saturates the transition amplitude from the false vacuum initial state to the basin of attraction of true vacuum. In our case, the field equation of motion is

$$\Box \varphi - U_{\mathrm{eff}}(x) \varphi - \Omega V_{\mathrm{int}}'(\varphi) = 0 . \tag{9}$$

The bounce lives on a contour C in the complex time plane. The contour runs from the initial moment of time in the asymptotic past, $t = t_i^{up}$, as shown in Fig. 1, to the final moment $t = t_f$ and back to the asymptotic past, $t = t_i^{low}$. The contour must bypass singularities of the bounce. Assuming that the bounce is unique, its values on the upper and lower sides of the contour are complex conjugated, hence at $t = t_f$ it is real and can be analytically continued along the real time axis where it describes the evolution of the field after tunneling.

In the limits $t \to t_i^{up}$ and $t \to t_i^{low}$ the bounce must satisfy boundary conditions imposed by the false vacuum state. These boundary conditions turn out to be the

same as for the time-ordered Green's function in the corresponding vacuum X.[30] The latter is defined as a time-ordered average of the field operators $\hat{\varphi}$ in the state X,

$$\mathcal{G}_{\rm X}(t,x;t',x') = {}_X\langle T(\hat{\varphi}(t,x)\hat{\varphi}(t',x'))\rangle_X \ . \tag{10}$$

In turn, the field operator is composed out of the complete set of modes (6) in the standard way.[33]

A Green's function $\mathcal{G}$ satisfies the equation

$$(\Box - U_{\rm eff}(x))\,\mathcal{G}(t,x;t',x') = i\delta(t-t')\delta(x-x') \ . \tag{11}$$

Using eq. (11), the field equation (9) can be recast into the integral form. To select a particular solution—bounce—of the integral equation, one should, first, adopt the time-integration contour C in the complex time plane and, second, choose a particular Green's function $\mathcal{G}_{\rm X}$. Thus, we arrive at

$$\varphi_{\rm b}(t,x) = -i\int_C dt' \int_{-\infty}^{\infty} dx' \mathcal{G}_{\rm X}(t,x;t',x')\Omega(x')V'_{\rm int}(\varphi_{\rm b}(t',x')) \ . \tag{12}$$

This form of the bounce equation will be useful in what follows.

Finally, let us discuss the tunneling rate Γ. The latter is defined as the probability of tunneling per unit time. We are interested in the main exponential dependence and write

$$\Gamma \sim e^{-B} \ . \tag{13}$$

Up to boundary terms, the coefficient B equals the imaginary part of the bounce action computed along the contour C. One can show that[30]

$$B = -\frac{i}{\mathrm{g}^2}\int_C dt \int_{-\infty}^{\infty} dx \left(\frac{1}{2}\varphi_{\rm b}V'_{\rm int}(\varphi_{\rm b}) - V_{\rm int}(\varphi_{\rm b})\right) \ . \tag{14}$$

3. The toy model

3.1. *Tunneling in flat spacetime*

In general, solving eq. (12) (or eq. (9) on the contour C with the vacuum boundary conditions) requires a numerical procedure. A big simplification of the problem happens in theories where the nonlinear core of the bounce, where it probes the true vacuum region, is much smaller in size than the Compton wavelength of the free field $\propto m^{-1}$. Then the source in the integral (12) is essentially point-like, and the solution outside the core is simply proportional to the Green's function $\mathcal{G}_{\rm X}$. On the other hand, the core itself can be found in neglecting the mass term in the field equation (9). The full solution is constructed by matching the long-distance asymptotics of the core with the short-distance asymptotics of the Green's function.

To meet the above requirements, we consider the following interaction potential[30]

$$V_{\rm int}(\varphi) = -2\kappa\,(e^{\varphi} - 1) \tag{15}$$

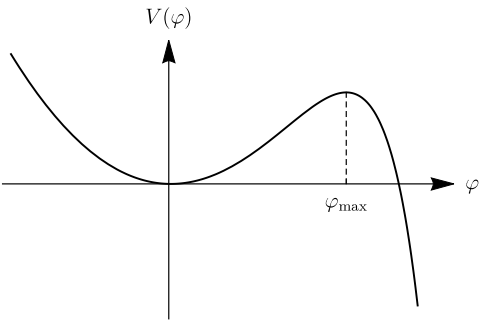


Fig. 2. The scalar field potential.

with $\kappa > 0$. In flat spacetime, the full scalar field potential $V(\varphi) = \frac{1}{2}m^2\varphi^2 + V_{\rm int}(\varphi)$ is shown in Fig. 2. To ensure the applicability of the split-and-match procedure, the following relation between the parameters is adopted

$$\ln \frac{m}{\sqrt{\kappa}} \gg 1 \, . \tag{16}$$

Thanks to this hierarchy, the theory possesses two intrinsic energy scales: the mass scale m and the scale associated with the potential barrier $m \ln \frac{m}{\sqrt{\kappa}}$. Generally speaking, the first controls the width of the linear tail of the bounce, while the second determines the size of its core. The barrier separating the regions of false and true vacua has its maximum at

$$\varphi_{\max} \approx 2 \ln \frac{m}{\sqrt{\kappa}} \, , \tag{17}$$

where only the leading logarithmic term is taken into account.

To illustrate the matching procedure outlined above, let us discuss bounce in the flat-space Minkowski vacuum. Assume that the only singularities of the bounce are located on the real-time axis. Then, the contour C can be deformed into the contour C' that runs along the Euclidean time axis, see Fig. 3(a) for illustration. The vacuum boundary condition at C, which is provided by the Feynman Green's function, becomes the vanishing boundary condition at C'. Hence, the standard Euclidean approach is reproduced.[24,25]

To find the core of the bounce, we neglect the mass term in eq. (9) and, using eq. (15), obtain

$$\Box \, \varphi_{\rm b}\big|_{\rm core} + 2\kappa e^{\,\varphi_{\rm b}|_{\rm core}} = 0 \, . \tag{18}$$

This is the Liouville equation, and its general solution is known. Next, the linear tail of the bounce is proportional to the (massive scalar) Feynman Green's function

$$\mathcal{G}_{\rm F}(-i\tau, x; 0, 0) = \frac{1}{2\pi} K_0 \left(m\sqrt{\tau^2 + x^2 + i\epsilon} \right) \, , \tag{19}$$

where $\tau = it$ is the Euclidean time coordinate and K_0 is the modified Bessel function of the second kind. The core and the tail are matched in the region where, on the

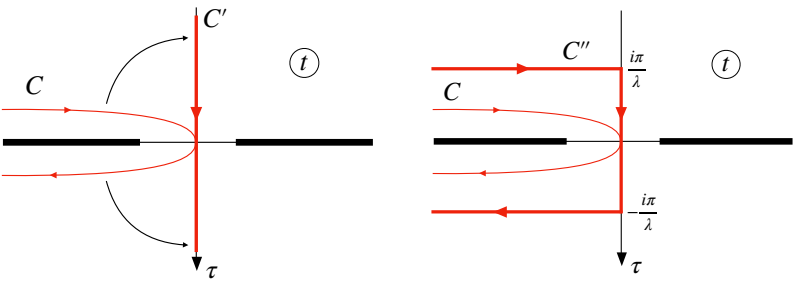


Fig. 3. (a) Deformation of the contour C into the Euclidean time contour C' used for the calculation of the flat-space Minkowski bounce. Black lines denote singularities of the bounce. (b) Deformation of C into C'' containing the Euclidean time segment which is used for the calculation of the Unruh bounce.

one hand, the solution to the Liouville equation is linearized and, on the other hand, the Green's function is approximated by its short-distance asymptotics. We obtain

$$\varphi_{\rm b}|_{\rm core} = \ln\left[\frac{4C_M^2}{(1+\kappa C_M^2(\tau^2+x^2))^2}\right] \tag{20}$$

$$\varphi_{\rm b}|_{\rm tail} = 8\pi\mathcal{G}_{\rm F}(-i\tau,x;0,0)\ , \tag{21}$$

where $C_M = m^2 e^{2\gamma_E}/(2\kappa)$ and γ_E is the Euler constant. The matching region is defined by $(C_M\sqrt{\kappa})^{-1} \ll \sqrt{\tau^2+x^2} \ll m^{-1}$. Its existence is ensured by eq. (16). We see that the bounce is real for real τ, vanishes at infinity and has zero time derivative at $\tau = 0$. Hence, it is a valid tunneling solution. Moreover, it has no singularities apart from the ones on the real-time axis, which justifies the deformation of the contour C into C'.

The tunneling suppression is given by eq. (14) where one should substitute the core of the bounce (20). To the leading-log approximation, this gives

$$B_M = \frac{16\pi}{\mathrm{g}^2}\ln\frac{m}{\sqrt{\kappa}}\ . \tag{22}$$

We observe that the suppression is enhanced by the large logarithm.

In the BH background (2), the linearized field equation (7) is still exactly solvable, and this allows us to compute explicitly the Green's functions of interest. The solvability of the equation for the core of the bounce is lost, but can be recovered in the two regions: near the horizon where the metric is approximately the Rindler one, and far away from the BH where the spacetime is asymptotically flat. From eq. (2) we see that the two regions are defined by $x < 0$, $|x| \gg \lambda^{-1}$ and $x > 0$, $|x| \gg \lambda^{-1}$, respectively. Although we do not have the explicit bounce solution in the transition region $|x| \lesssim \lambda^{-1}$, we will still be able to draw a qualitative picture of the evolution of the bounce across this region.

3.2. *Decay of the Unruh vacuum near horizon*

As discussed in introduction, the Unruh vacuum corresponds to a flux of thermal radiation propagating from the BH, whose spectrum is regularized by the m^2-asymptotics of the mode potential (8). In the absence of area growth in one spatial dimension, the flux enhances the vacuum decay rate at any distance from the BH. In the near-horizon region, the decay rate is additionally enhanced due to the accumulation of field excitations and curved geometry. Let us discuss the near-horizon case first.

To find the Unruh bounce in the BH vicinity, we have to integrate eq. (9) with $\Omega = e^{2\lambda x}$ along the contour C shown in Fig. 1 and satisfy the boundary condition imposed by the Unruh vacuum. The contour, in fact, can be deformed partially to the Euclidean time domain, as shown in Fig. 3(b). In the core region of the bounce the field equation reduces to

$$\Box\, \varphi_{\rm b}\big|_{\rm core} + 2\kappa e^{2\lambda x + \varphi_{\rm b}|_{\rm core}} = 0\,. \tag{23}$$

This again admits a general solution. The tail of the bounce is determined by the time-ordered Unruh Green's function $\mathcal{G}_U$ computed in the BH vicinity. Overall, we write

$$\varphi_{\rm b}\big|_{\rm core} = \ln\left[\frac{4\lambda^2 b_{U1}}{\kappa\left(-2\lambda(v-v_1)\sinh\left(\frac{\lambda}{2}(u-u_1)\right) + b_{U1}e^{\frac{\lambda}{2}(u-u_1)}\right)^2}\right] - 2\lambda x \tag{24}$$

$$\varphi_{\rm b}\big|_{\rm tail} = 8\pi\mathcal{G}_U(t,x;0,x_1)\,. \tag{25}$$

Here we defined $u = t - x$, $v = t + x$, and $u_1 = t - x_1$, $v_1 = t + x_1$. The parameter x_1 determines the position of the center of the bounce[a] and b_{U1} is found from matching the core and the tail. The expression for the Unruh Green's function at $m|x - x_1| \ll 1$ is given by eq. (A.2) from the appendix. Using it, we get

$$b_{U1} = \bar{b}_{U1}e^{-2\lambda x_1}\,, \qquad \bar{b}_{U1} = \frac{\kappa}{m^2}e^{\frac{32\lambda}{3\pi m} - 2\gamma_E - \frac{3}{2}}\,, \tag{26}$$

where we additionally assumed $\lambda \gg m$. We obtain a one-parameter family of solutions parametrized by x_1. This is an artefact of our approximation of the near-horizon geometry by the Rindler spacetime. Upon taking into account BH corrections to the Rindler metric, a unique tunneling solution is selected.

The bounce (24), (25) is real on the real-time axis and describes the tun-away toward the true vacuum $\varphi \to +\infty$ at positive time. Hence, it is a valid tunneling solution.

Investigating the applicability of the matching procedure, one arrives at the following condition[30]

$$\bar{b}_{U1} \lesssim 1\,. \tag{27}$$

[a]We used the time translational invariance of the theory to put the center of the bounce at $t = 0$.

This inequality translates into the upper bound on the BH temperature, $\lambda \lesssim \lambda_{U1}$, where to the leading-log accuracy

$$\lambda_{U1} = \frac{3\pi m}{16} \ln \frac{m}{\sqrt{\kappa}} \,. \tag{28}$$

At $\lambda > \lambda_{U1}$, the bound (27) is violated and the matching procedure breaks down. This can be given a physical interpretation: at $\lambda \simeq \lambda_{U1}$ the nonlinear core of the bounce stops fitting the near-horizon region, as its physical size exceeds l_h, see eq. (3). The subsequent evolution of the bounce happens outside the BH vicinity, and we will discuss it below.

To compute the exponential suppression of the decay rate, we use the general expression (14) with the contour C'' depicted in Fig. 3(b). The result is[30]

$$B_{U1} = \frac{16\pi}{\mathrm{g}^2} \left(\ln \sqrt{\frac{\lambda m}{\kappa}} - \frac{8\lambda}{3\pi m} \right) \,, \quad \lambda \lesssim \lambda_{U1} \,, \tag{29}$$

where for the sake of simplicity we omitted the order-one terms in the parentheses. We see that the suppression decreases with the BH temperature and becomes approximately half of the vacuum suppression at $\lambda \approx \lambda_{U1}$.

3.3. *Decay of the Unruh vacuum away from the black hole*

Far from the BH, the vacuum decay proceeds in the flat spacetime, and the only difference between the Unruh and Minkowski vacua is the presence of out-of-equilibrium field excitations. The question of how such excitations catalyze vacuum decay is interesting by itself.

Solving eq. (18) on the contour C'' and matching with the Unruh Green's function in the corresponding region, we arrive at the following solution

$$\varphi_{\mathrm{b}}\big|_{\mathrm{core}} = \ln \left[\frac{4\lambda^2 b_{U2}}{\kappa \left(-2\lambda (v - v_2) \sinh \left(\frac{\lambda}{2} (u - u_2) \right) + b_{U1} e^{\frac{\lambda}{2}(u-u_2)} \right)^2} \right] \tag{30}$$

$$\varphi_{\mathrm{b}}\big|_{\mathrm{tail}} = 8\pi \mathcal{G}_U(t, x; 0, x_2) \,, \tag{31}$$

where we denoted $u_2 = t - x_2$, $v_2 = t + x_2$, and the expression for $\mathcal{G}_U$ in the limit $|x - x_2| \ll m^{-1}$ and $\lambda \gg m$ is given in eq. (A.3) from the appendix. The parameter b_{U2} is found upon matching the core and the tail of the bounce and equals

$$b_{U2} = \frac{\kappa}{m^2} e^{\frac{8\lambda}{3\pi m} - 2\gamma_E + \frac{1}{2}} \,. \tag{32}$$

Similarly to the case studied above, one can take

$$b_{U2} \lesssim 1 \tag{33}$$

as a conservative condition for the validity of the bounce solution (30), (31). The bound (33) translates into the condition $\lambda \lesssim \lambda_{U2}$, where, to the leading-log accuracy,

$$\lambda_{U2} = \frac{3\pi m}{4} \ln \frac{m}{\sqrt{\kappa}} \,. \tag{34}$$

Finally, for the exponential suppression of the decay rate we get

$$B_{U2} = \frac{16\pi}{\mathrm{g}^2}\left(\ln\sqrt{\frac{\lambda m}{\kappa}} - \frac{2\lambda}{3\pi m}\right), \qquad \lambda \lesssim \lambda_{U2}\ . \tag{35}$$

We see that the coefficient in the linear in λ term is smaller than in eq. (29). Thus, the tunneling far from the horizon is more suppressed, as expected. Again, the suppression becomes half of the flat space one when the critical temperature is reached.

What happens with the decay rate at $\lambda > \lambda_{U2}$? We could not find the relevant tunneling solution in this regime. Therefore, let us proceed with a different method. At $\lambda \gg m$, the dominant contribution to the Unruh Green's function comes from the modes with $\omega \sim m$. The occupation number of modes of the dominant frequency is large, being enhanced by λ/m. Hence, the flux of particles in the Unruh state is essentially classical. It is these modes that are relevant for vacuum decay. Therefore, the vacuum decay probability can be estimated as the probability of the classical field fluctuation to reach beyond the maximum of the potential barrier at $\varphi = \varphi_{\max}$, see eq. (17) and Fig. 2. To find the variance of classical fluctuations $(\delta\varphi)^2_U$, we take the Unruh Green's function at coincident points, upon renormalizing it by subtraction of the Green's function in empty space,

$$(\delta\varphi)^2_U = \mathrm{g}^2 \lim_{t,x\to 0}\left[\mathcal{G}_U(t,x;0,0) - \mathcal{G}_F(t,x;0,0)\right]\ . \tag{36}$$

Since the interaction quickly dies out at $\varphi < \varphi_{\max}$, one can take the field fluctuations to be Gaussian. Then we obtain the following estimation of the decay rate due to classical field fluctuations

$$\Gamma_{U2,\mathrm{high}-\lambda} \sim \exp\left(-\frac{\varphi^2_{\max}}{2(\delta\varphi)^2_U}\right) \sim \exp\left(-\frac{6\pi^2 m}{\mathrm{g}^2\lambda}\left(\ln\frac{m}{\sqrt{\kappa}}\right)^2\right)\ . \tag{37}$$

As a consistency check, at $\lambda \approx \lambda_{U2}$ the stochastic estimate coincides with the bounce suppression (35) in the leading-log approximation.

Finally, let us go back to the near-horizon tunneling channel and see what happens with the decay rate after the critical temperature (28) is reached. At this point we can use again the stochastic estimate. We use the expression for the Unruh Green's function (A.3) to the right side from the barrier but not far from it and obtain

$$\Gamma_{U1,\mathrm{high}-\lambda} \sim \exp\left(-\frac{3\pi^2 m}{2\mathrm{g}^2\lambda}\left(\ln\frac{m}{\sqrt{\kappa}}\right)^2\right)\ . \tag{38}$$

We observe that in the leading-log approximation the suppression is continuous across the transition from the low- to high-temperature regime.

4. Conclusions and outlook

In this report we discussed the method to compute the rate of decay of the false Unruh vacuum. We applied the method to a toy model of a real scalar field in

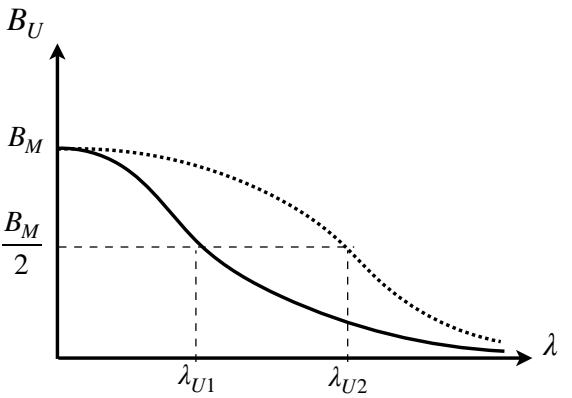


Fig. 4. Suppression of the Unruh vacuum decay as a function of BH temperature. The solid line represents the decay channel in the BH vicinity, and the dotted line shows the decay channel far from the BH. B_M, λ_{U1} and λ_{U2} are given in eqs. (22), (28) and (34), respectively.

the background of a dilaton BH in two dimensions. The results are summarized in Fig. 4. We see, in particular, that the decay channel far from the BH is suppressed as compared to the one close to the BH. This is in agreement with expectations. We also see that at high BH temperature the decay becomes unsuppressed.

Let us discuss to what extent our results are generalized to a realistic case such as a Schwarzschild BH in four dimensions. One should understand which features of the four-dimensional geometry are captured by the toy model in two dimensions, and which are not. For this, we can limit our discussion to the linear part of the theory. This is reasonable, since, as discussed above, properties of linear modes (or, more precisely, of Green's functions) are crucial for finding vacuum boundary conditions and, eventually, the decay rate. Then, one can readily compare the linear sector of the toy model with that of the spherical reduction of the free four-dimensional theory. More precisely, we can compare the greybody factors for linear perturbations. The four-dimensional greybody factor includes the barrier whose height depends on the size of the BH (and on its temperature). There is no such barrier in two dimensions. Its presence in four dimensions can change the results concerning the Unruh vacuum decay at high temperatures. Indeed, the barrier reduces the flux of Hawking radiation far from the BH and promotes the accumulation of Hawking quanta in the near-horizon region. In principle, our method allows us to take this effect into account, and we leave this study for future work.

Next, we found that the Unruh bounce does not exist above the certain temperature, both in the BH vicinity and in the asymptotically-flat region. Here, the situation is similar to the decay from a state in thermal equilibrium. It is known that a finite-temperature periodic instanton, which describes tunneling from such a state, yields to a static configuration—sphaleron—once the temperature exceeds a critical value.[34] The sphaleron corresponds to classical thermal jumps of the field

over the barrier separating the false and true vacua. One can expect that in the case of Unruh vacuum a similar transition happens. We could not find the analog of sphaleron in the Unruh state analytically. However, thanks to the properties of our model, we were able to use the simple stochastic estimate and compute the decay rate above the critical temperature. Let us reiterate that the validity of the estimate relies on the two facts. First, (almost) any relevant for decay field fluctuation is Gaussian, as it follows from the scalar field potential. Second, the field excitations of relevant frequency, $\omega \sim m$, have large occupation numbers and, hence, can be treated as classical. In more general situations, the stochastic approach to vacuum decay can still be applicable, but would require a full-fledged numerical simulation. We also leave this for future work.

Acknowledgments

The author thanks S. Sibiryakov for useful discussions. The work was partially supported by the Department of Energy Grant DE-SC0011842.

Appendix A. Unruh Green's function

Here we quote the expressions for the (massive scalar) Unruh Green's function in two dimensions. We focus on the two regions: the near-horizon one ("left"), defined by the condition $x, x' < 0$, $|x|, |x'| \gg \lambda^{-1}$, and the asymptotically-flat one ("right"), determined by $x, x' > 0$, $|x|, |x'| \gg \lambda^{-1}$. Given that the metric (1) is static, we can set $t' = 0$ in the Green's function. Next, we consider the limit

$$|x - x'| \ll m^{-1} , \qquad |t| \ll m^{-1} . \tag{A.1}$$

Under these assumptions, the Unruh Green's function reads as follows[30]

$$\begin{aligned} \mathcal{G}_U(t, x; 0, x')\big|_{\text{left}} = &-\frac{1}{4\pi} \ln \left[2 \sinh \left(\frac{\lambda}{2} (x - x' - t) \right) m(x - x' + t) + i\epsilon \right] \\ &- \frac{\lambda(x + x')}{4\pi} + \frac{4\lambda}{3\pi^2 m} + \frac{\ln 2 - \gamma_E}{4\pi} - \frac{3}{16\pi} \end{aligned} \tag{A.2}$$

and

$$\begin{aligned} \mathcal{G}_U(t, x; 0, x')\big|_{\text{right}} = &-\frac{1}{4\pi} \ln \left[2 \sinh \left(\frac{\lambda}{2} (x - x' - t) \right) m(x - x' + t) + i\epsilon \right] \\ &+ \begin{cases} -\frac{\lambda}{4\pi}(x + x') + \frac{4\lambda}{3\pi^2 m} + \frac{\ln 2 - \gamma_E}{4\pi} - \frac{3}{16\pi} , & x + x' \ll m^{-1} \\ \frac{\lambda}{3\pi^2 m} + \frac{\ln 2 - \gamma_E}{4\pi} + \frac{1}{16\pi} , & x + x' \gg m^{-1} \end{cases} \end{aligned} \tag{A.3}$$

References

1. W. A. Hiscock, Can black holes nucleate vacuum phase transitions?, *Phys. Rev.* **D35**, 1161 (1987).

2. V. A. Berezin, V. A. Kuzmin and I. I. Tkachev, O(3) Invariant Tunneling in General Relativity, *Phys. Lett.* **B207**, 397 (1988).
3. P. B. Arnold, Gravity and false vacuum decay rates: O(3) solutions, *Nucl. Phys.* **B346**, 160 (1990).
4. V. A. Berezin, V. A. Kuzmin and I. I. Tkachev, Black holes initiate false vacuum decay, *Phys. Rev.* **D43**, 3112 (1991).
5. R. Gregory, I. G. Moss and B. Withers, Black holes as bubble nucleation sites, *JHEP* **03**, p. 081 (2014).
6. P. Burda, R. Gregory and I. Moss, Gravity and the stability of the Higgs vacuum, *Phys. Rev. Lett.* **115**, p. 071303 (2015).
7. P. Burda, R. Gregory and I. Moss, Vacuum metastability with black holes, *JHEP* **08**, p. 114 (2015).
8. P. Burda, R. Gregory and I. Moss, The fate of the Higgs vacuum, *JHEP* **06**, p. 025 (2016).
9. N. Tetradis, Black holes and Higgs stability, *JCAP* **1609**, p. 036 (2016).
10. D. Canko, I. Gialamas, G. Jelic-Cizmek, A. Riotto and N. Tetradis, On the Catalysis of the Electroweak Vacuum Decay by Black Holes at High Temperature, *Eur. Phys. J.* **C78**, p. 328 (2018).
11. D. Gorbunov, D. Levkov and A. Panin, Fatal youth of the Universe: Black hole threat for the electroweak vacuum during preheating, *JCAP* **1710**, p. 016 (2017).
12. K. Mukaida and M. Yamada, False Vacuum Decay Catalyzed by Black Holes, *Phys. Rev.* **D96**, p. 103514 (2017).
13. K. Kohri and H. Matsui, Electroweak Vacuum Collapse induced by Vacuum Fluctuations of the Higgs Field around Evaporating Black Holes, *Phys. Rev.* **D98**, p. 123509 (2018).
14. T. Hayashi, K. Kamada, N. Oshita and J. Yokoyama, On catalyzed vacuum decay around a radiating black hole and the crisis of the electroweak vacuum, *JHEP* **08**, p. 088 (2020).
15. T. Miyachi and J. Soda, False vacuum decay in a two-dimensional black hole spacetime, *Phys. Rev. D* **103**, p. 085009 (2021).
16. R. A. Flores and M. Sher, Upper Limits to Fermion Masses in the Glashow-Weinberg-Salam Model, *Phys. Rev. D* **27**, p. 1679 (1983).
17. M. Sher, Electroweak Higgs Potentials and Vacuum Stability, *Phys. Rept.* **179**, 273 (1989).
18. G. Isidori, G. Ridolfi and A. Strumia, On the metastability of the standard model vacuum, *Nucl. Phys. B* **609**, 387 (2001).
19. F. Bezrukov, M. Yu. Kalmykov, B. A. Kniehl and M. Shaposhnikov, Higgs Boson Mass and New Physics, *JHEP* **10**, p. 140 (2012).
20. G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori and A. Strumia, Higgs mass and vacuum stability in the Standard Model at NNLO, *JHEP* **08**, p. 098 (2012).
21. D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio and A. Strumia, Investigating the near-criticality of the Higgs boson, *JHEP* **12**, p. 089 (2013).
22. A. V. Bednyakov, B. A. Kniehl, A. F. Pikelner and O. L. Veretin, Stability of the Electroweak Vacuum: Gauge Independence and Advanced Precision, *Phys. Rev. Lett.* **115**, p. 201802 (2015).
23. T. Markkanen, A. Rajantie and S. Stopyra, Cosmological Aspects of Higgs Vacuum Metastability, *Front. Astron. Space Sci.* **5**, p. 40 (2018).
24. S. R. Coleman, The Fate of the False Vacuum. 1. Semiclassical Theory, *Phys. Rev.* **D15**, 2929 (1977), [Erratum: Phys. Rev. D16, 1248 (1977)].

25. C. G. Callan, Jr. and S. R. Coleman, The Fate of the False Vacuum. 2. First Quantum Corrections, *Phys. Rev.* **D16**, 1762 (1977).
26. S. R. Coleman, The Uses of Instantons, *Subnucl. Ser.* **15**, p. 805 (1979).
27. S. R. Coleman and F. De Luccia, Gravitational Effects on and of Vacuum Decay, *Phys. Rev.* **D21**, p. 3305 (1980).
28. B. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, Constraints on Primordial Black Holes (2020).
29. W. G. Unruh, Notes on black hole evaporation, *Phys. Rev.* **D14**, p. 870 (1976).
30. A. Shkerin and S. Sibiryakov, Black hole induced false vacuum decay from first principles (5 2021).
31. C. G. Callan, Jr., S. B. Giddings, J. A. Harvey and A. Strominger, Evanescent black holes, *Phys. Rev.* **D45**, p. R1005 (1992).
32. W.-Y. Ai, Correspondence between Thermal and Quantum Vacuum Transitions around Horizons, *JHEP* **03**, p. 164 (2019).
33. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics (Cambridge Univ. Press, Cambridge, UK, 1984).
34. A. D. Linde, Decay of the False Vacuum at Finite Temperature, *Nucl. Phys. B* **216**, p. 421 (1983), [Erratum: Nucl. Phys. B 223, 544 (1983)].

Behaviour of noise kernel in de Sitter and FRW space-times

Ankit Dhanuka* and Kinjalk Lochan†

Department of Physical Sciences,
Indian Institute of Science Education & Research (IISER)
Mohali, Sector 81 SAS Nagar, Manauli PO 140306 Punjab India
**E-mail: ph17006@iisermohali.ac.in*
†E-mail: kinjalk@iisermohali.ac.in

Validity of the results of semiclassical analysis relies upon the assumption that the 2nd order or higher order corrections are negligible compared to 1st order semiclassical results which are based only on expectation values of operators. However, if the quantum fluctuations are large, then one may need to supplement semiclassical analysis with corrections coming from second order calculations. In the stochastic gravity paradigm, these fluctuations are quantified by noise kernel which are then supposed to act as a source of geometric fluctuations. In this work, we study the behaviour of noise kernel for a scalar field in de Sitter spacetime. We also carry out a similar analysis for some other FRW spacetimes invoking an equivalence that exists between scalar fields in de Sitter and FRW spacetimes.

Keywords: Stochastic gravity, Noise kernel etc.

1. Introduction

In semiclassical approach, we take matter fields to be quantum fields evolving over the classical geometry of a spacetime.[1] However, in this approach, we consider only the 1st order effects, for example, we replace classical stress energy tensor in Einstein's equations by the expectation value of the corresponding operator w.r.t. some quantum state in which the matter field is supposed to be. Now, conclusions drawn on the basis of such an analysis always rely on the implicit assumption that the 2nd or higher order quantum effects are relatively small and will not significantly affect the 1st order results. Though, in situations where higher order effects are appreciable, one may need to revisit the conclusions made on the basis of these 1st order calculations and may have to supplement them with 2nd or higher order calculations. For these considerations, one may resort to stochastic gravity[2,3] paradigm where one tries to incorporate the effects of quantum fluctuations, inherent in quantum matter fields, on the dynamics of classical background. These quantum fluctuations are quantified by what is called noise kernel which is given as follows:

$$N_{abcd} = \frac{1}{8}\{\hat{t}_{ab}, \hat{t}_{cd}\}, = \frac{1}{8}(\langle \hat{t}_{abcd}(x,x')\rangle + \langle \hat{t}_{cdab}(x,x')\rangle) \tag{1}$$

where $\hat{t}_{ab} = \hat{T}_{ab} - \langle \hat{T}_{ab} \rangle$, and

$$\begin{aligned} \langle \hat{t}_{abcd}(x, x') \rangle &\equiv \langle \hat{t}_{ab}(x) \hat{t}_{cd}(x') \rangle \\ &= \langle 0 | \hat{T}_{ab}(x) \hat{T}_{cd}(x') | 0 \rangle - \langle 0 | \hat{T}_{ab}(x) | 0 \rangle \langle 0 | \hat{T}_{cd}(x') | 0 \rangle \,. \end{aligned} \tag{2}$$

Stress energy quantum operator, $\hat{T}_{ab}$, is obtained by replacing classical fields by their quantum counterparts and the classical expression for stress energy tensor is obtained by taking the metric variation of the matter action, more precisely, $T_{ab}(x) = -\frac{2}{\sqrt{-g}} \frac{\delta S_M}{\delta g^{ab}(x)}$.

In this work, we are interested in studying the behaviour of noise kernel for scalar fields in a wide class of FRW spacetimes.[4,5] The line element for a general FRW spacetime is given by

$$ds^2 = a^2(\eta)(-d\eta^2 + d\vec{x}^2) \,, \tag{3}$$

where the scale factor, $a(\eta)$, characterizes that spacetime. In particular, we are interested in de Sitter spacetime, with $a(\eta) = \frac{-1}{H\eta}$. de Sitter spacetime is a maximally symmetric spacetime and present day observations[6] suggest that the early universe went through a period of nearly de Sitter phase. Similarly, universe is also believed to have undergone through other FRW type periods/epochs. Therefore, studying these second order quantum effects in FRW settings is an important cosmological task and may be helpful in better understand the cosmological data. We make use of an equivalence that exists between scalar fields in de Sitter and FRW spacetimes, to study the noise kernel in FRW spacetimes in terms of the noise kernel in de Sitter spacetime. Equally important is the study of backreaction of the noise kernel on the classical geometry of spacetime, though in the present work, we largely ignore this aspect of the analysis.

2. Point-Splitting technique and Bunch Davies vacuum

Let's consider a minimally coupled massive scalar field in general space-time metric i.e.,

$$S[g_{\alpha\beta}, \phi] = -\frac{1}{2} \int d\eta d^3\vec{x} \sqrt{-g} \left(g^{\alpha\beta} \nabla_\alpha \phi \nabla_\beta \phi + m^2 \phi^2 \right) . \tag{4}$$

We find that its stress energy operator is given by:

$$T_{\alpha\beta}(x) = \nabla_\alpha \phi \nabla_\beta \phi - \frac{1}{2} g_{\alpha\beta} (g^{\gamma\delta} \nabla_\gamma \phi \nabla_\delta \phi + m^2 \phi^2) \,. \tag{5}$$

One can write the above stress energy expression in the following point-split form[7]

$$T_{ab}(x) = \lim_{y\to x} P_{ab}(x,y)\phi(x)\phi(y)\,, \tag{6}$$

where

$$P_{ab}(x,y) = \left(\delta^c_{(a}\delta^d_{b)} - \frac{1}{2}\left(\frac{g_{ab}(x)+g_{ab}(y)}{2}\right)\left(\frac{g^{cd}(x)+g^{cd}(y)}{2}\right)\right)\nabla^x_c\nabla^y_d$$
$$- \frac{1}{2}\left(\frac{g_{ab}(x)+g_{ab}(y)}{2}\right)m^2\,. \tag{7}$$

Now substituting the above point split form in the formula for stress-energy two point correlator w.r.t the vacuum of the field, we find that

$$\langle t_{abcd}(x,x')\rangle = \lim_{\substack{y\to x\\ y'\to x'}} P_{ab}(x,y)P_{cd}(x',y')\langle 0|\phi(x)\phi(y)\phi(x')\phi(y')|0\rangle$$
$$- \lim_{\substack{y\to x\\ y'\to x'}} P_{ab}(x,y)P_{cd}(x',y')\langle 0|\phi(x)\phi(y)|0\rangle\langle 0|\phi(x')\phi(y')|0\rangle\,. \tag{8}$$

After some manipulations, this becomes

$$\langle t_{abcd}(x,x')\rangle = 2\lim_{\substack{y\to x\\ y'\to x'}} P_{ab}(x,y)P_{cd}(x',y')G(x,x')G(y,y')\,, \tag{9}$$

where $G(x,x')$ is the Wightman function for the scalar field in the considered vacuum,

$$G(x,x') = \langle 0|\,\phi(x)\phi(x')\,|0\rangle\ . \tag{10}$$

For de Sitter spacetime, we determine the behaviour of noise kernel for Bunch Davies vacuum[8,9] which is given by

$$G(Z) = (H^2/16\pi^2)\Gamma(a)\Gamma(b)_2F_1\left(a,b,2,\frac{1+Z}{2}\right), \tag{11}$$

where $Z(x,x') = 1+\frac{(\eta-\eta')^2-(\vec{x}-\vec{x}')^2}{2\eta\eta'}$ and $a,b = \frac{3}{2}\pm\sqrt{\frac{9}{4}-\frac{m^2}{H^2}}$. We see that the Bunch Davies vacuum depends on the coordinates only through the invariant distance $Z(x,x')$ between them.

3. Noise Kernel in de Sitter universe

Since for de Sitter spacetime, $g_{ab} = \frac{1}{(H\eta)^2}\eta_{ab}$, we find that

$$\begin{aligned}\langle t_{abcd}(x,x')\rangle_{dS} = \Bigg(& \nabla_b\nabla'_c G(x,x')\nabla_a\nabla'_d G(x,x') + \nabla_b\nabla'_d G(x,x')\nabla_a\nabla'_c G(x,x') \\ & - \eta_{cd}\eta^{\rho\sigma}\nabla_a\nabla'_\rho G(x,x')\nabla_b\nabla'_\sigma G(x,x') \\ & - \frac{1}{H^2\eta'^2}m^2\eta_{cd}\nabla_a G(x,x')\nabla_b G(x,x') \\ & - \eta_{ab}\eta^{\gamma\delta}\nabla_\gamma\nabla'_c G(x,x')\nabla_\delta\nabla'_d G(x,x') \\ & + \frac{1}{2}\eta_{ab}\eta^{\gamma\delta}\eta_{cd}\eta^{\rho\sigma}\nabla_\gamma\nabla'_\rho G(x,x')\nabla_\delta\nabla'_\sigma G(x,x') \\ & + \frac{1}{2H^2\eta'^2}m^2\eta_{ab}\eta^{\gamma\delta}\eta_{cd}\nabla_\gamma G(x,x')\nabla_\delta G(x,x') \\ & - \frac{1}{H^2\eta^2}m^2\eta_{ab}\nabla'_c G(x,x')\nabla'_d G(x,x') \\ & + \frac{1}{2H^2\eta^2}m^2\eta_{ab}\eta_{cd}\eta^{\rho\sigma}\nabla'_\rho G(x,x')\nabla'_\sigma G(x,x') \\ & + \frac{1}{2H^4\eta^2\eta'^2}m^4\eta_{ab}\eta_{cd}G(x,x')G(x,x')\Bigg). \end{aligned} \tag{12}$$

In order to see whether the primordial fluctuations are significant or not in an expanding universe, we consider constant time sheets i.e., we take $\eta = \eta'$ and look at the behaviour of noise kernel in the limit in which $a(\eta) \to \infty$ i.e., $\eta \to 0$ limit. Therefore, for Bunch Davies vacuum, the $(a = b = c = d = 0)$ component of the stress energy correlator behaves in the following manner[10]:

$$\langle t_{00}(\eta,\vec{x})t_{00}(\eta,\vec{x}')\rangle_{dS}\big|_{\text{late time}} = \frac{H^4\Gamma^2(\nu)\Gamma^2(\frac{5}{2}-\nu)}{\pi^5}\left[\frac{9\eta^{2-4\nu}}{32(\Delta\vec{x})^{6-4\nu}} + O(\eta^{4-4\nu})\right], \tag{13}$$

where $\nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}$.

From the above expression, we see that the above considered component of stress energy correlator vanishes in the $\eta \to 0$ limit when ν lies between $[0, \frac{1}{2})$ while it diverges in the same limit for ν lying between $(\frac{1}{2}, \frac{3}{2}]$. At $\nu = \frac{1}{2}$, it becomes independent of ν and is given by $\frac{9H^4}{32\pi^4(\Delta\vec{x})^4}$. The range for ν is decided by the requirement that the mass of the field, m, cannot take imaginary values. Since energy density at any spacetime point, x, is given by $T_{ab}(x)t^a t^b$, where t^a is tangent vector to a comoving observer at that point, we find that the energy-energy density is given by:

$$\frac{< \hat{T}_{00}(\eta,\vec{x})\hat{T}_{00}(\eta',\vec{x}') >_{dS}}{(a(\eta)a(\eta'))^2}. \tag{14}$$

Using the above formula and the fact that $a(\eta) = -\frac{1}{H\eta}$, we find that the energy-energy correlator vanishes for all values of ν lying between $[0, \frac{3}{2}]$ except at $\frac{3}{2}$ for

which it attains a constant value equal to $\frac{9H^4}{128\pi^4}$. One can also extend the above results to include all other components of the stress energy correlator as well as the non-minimal coupling.[10]

4. Noise Kernel for Friedmann spaces

As has already been stated, in order to study the behaviour of noise kernel in FRW spacetimes we make use of an equivalence between scalar fields in de Sitter and FRW spacetimes. So the idea is as follows[11]:

We consider a massless scalar field in a generic FRW universe i.e., $g_{\alpha\beta} = a^2(\eta)\eta_{\alpha\beta}$ with $a(\eta) = (H\eta)^{-q}$, with the action given by

$$S = -\frac{1}{2}\int d^4x\, a^4\big(a^{-2}\,\eta^{\alpha\beta}\partial_\alpha\phi\,\partial_\beta\phi\big)\,. \tag{15}$$

Under the transformation, $\phi(x) = (H\eta)^{-1+q}\,\psi(x)$, the action becomes

$$S = -\frac{1}{2}\int d^4x\, b^4\big(b^{-2}\,\eta^{\alpha\beta}\partial_\alpha\psi\,\partial_\beta\psi - m_{eff}^2\psi^2\big)\,, \tag{16}$$

where $b(\eta) = (H\eta)^{-1}$ and $m_{eff}^2 = H^2(1-q)(2+q)$. Therefore, with the above transformation, one can map a massless scalar field in a Friedmann universe with scaling factor, $a(\eta) = (H\eta)^{-q}$, to a massive scalar field in a de Sitter universe. From the relation, $m_{eff}^2 = H^2(1-q)(2+q)$, we conclude that the range $[0, \frac{3}{2}]$ of $\frac{m^2}{H^2}$ gets mapped to q between the range $[-2, 1]$. And since we are interested in the limit of η for which $a(\eta) \to \infty$, we find that, for $q \to [-2, 0]$, the late time limit corresponds to $\eta \to \infty$ while for $q \to [0, 1]$, the late time limit corresponds to $\eta \to 0$. From the above transformation between fields, we see that the Wightman functions in FRW spacetime and de Sitter spacetime are related by $G^{F.R.W.}(x, x') = (H\eta)^{q-1}(H\eta')^{q-1}G^{d.S.}(x, x')$. Using this relation between Wightman functions, we find that the stress energy correlator is given by:

$$\begin{aligned}
&\langle t_{00}(\eta,\vec{x})t_{00}(\eta,\vec{x'})\rangle_{F.R.W.}\\
&= (H\eta)^{4(q-1)}\Bigg[\frac{G^2}{2\eta^4}(q-1)^4 \;+ GG'\Big[\frac{(2q^3-7q^2+8q-3)(\Delta x)^2}{2\eta^6} - \frac{(q-1)^2}{\eta^4}\Big]\\
&+ GG''\frac{(q-1)^2(\Delta\vec{x})^4}{4\eta^8} + G'G''\Big[\frac{(q-\frac{3}{2})(\Delta\vec{x})^6}{4\eta^{10}} + \frac{(q-\frac{9}{4})(\Delta\vec{x})^4}{\eta^8} - \frac{(\Delta\vec{x})^2}{\eta^6}\Big]\\
&+ (G')^2\Big[\frac{2}{\eta^4} + \frac{(q^2-5q+\frac{11}{2})(\Delta\vec{x})^2}{\eta^6} + \frac{(2q^2-6q+\frac{9}{2})(\Delta\vec{x})^4}{4\eta^8}\Big]\\
&+ (G'')^2\Big[\frac{(\Delta\vec{x})^6}{4\eta^{10}} + \frac{(\Delta\vec{x})^8}{32\eta^{12}} + \frac{(\Delta\vec{x})^4}{2\eta^8}\Big]\Bigg]\,,
\end{aligned} \tag{17}$$

where $G(\eta, \vec{x}, \vec{x}')$ is the Wightman function in de Sitter universe[a]. Again taking the Wightman function to be the one corresponding to Bunch Davies vacuum, the

[a]We have suppressed the superscript d.S. in this expression.

behaviour of noise kernel for different cases can be summarized as follows[10]:

- $q = 1$: This is just massless scalar field case in de Sitter spacetime for which results are already given in the last section.
- $q \in (0, 1)$: In this case, the late time universe correspond to $\eta \to 0$ limit. And we find that, in this limit, $\langle t_{00}(\eta, \vec{x}) t_{00}(\eta, \vec{x'})\rangle_{P.L.}\big|_{\eta\to 0} = \frac{H^{4q}(\Delta\vec{x})^{4q-8}}{8\pi^5}\Big((11 - 12q{+}4q^2)(\Gamma(2{-}q))^2(\Gamma(0.5{+}q))^2\Big)$ i.e., it is time independent and it decreases with spatial distance.
- $q \in (-2, 0)$: In this case, the late time universe corresponds to $\eta \to \infty$. For this case, we have $\langle t_{00}(\eta, \vec{x}) t_{00}(\eta, \vec{x'})\rangle_{F.R.W.}\big|_{\text{late time}} = \lim_{\eta\to\infty}(H\eta)^{4q-4}\left[\frac{3H^4\eta^4}{2\pi^4(\Delta\vec{x})^8} + lower\ powers\ of\ \eta\right]$. This implies that there are no late time correlations in this case.
- $q = -2$: This is a slightly interesting case as in this case massless field in FRW spacetime goes to a massless scalar field in de Sitter spacetime. And because of this the stress energy correlator shows the infrared divergence as is present in Wightman function for a massless scalar in de Sitter spacetime.

Using the relation between the exponent, q, of the scale factor for a given FRW universe and the equation of state parameter, w, for the corresponding ideal fluid, i.e, $q = -2/(1 + 3w)$, we see that, for Friedmann universes with $q \in (0, 1]$, the equation of state parameter $w \in (-\infty, -1]$ whereas $w \in [0, \infty)$ for $q \in [-2, 0)$. Therefore, the results of the previous section with corresponding fluids can be summarized by the following diagram:

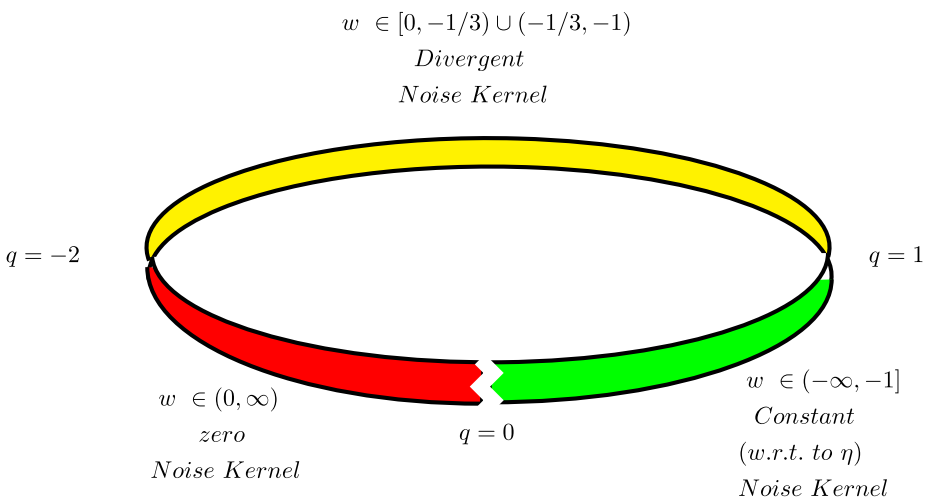


Fig. 1. Relation between different types of fluid (and the corresponding Friedmann space-times) and the behaviour of stress energy correlator in these regions.

Results for different epochs of the universe can be summarized in the following table[10]:

Table 1. Behaviour of the $(a = 0, b = 0, c = 0, d = 0)$ component of the stress-energy correlator and energy-energy correlator in the late time universe.[10]

	Radiation dominated $q = -1$	Dust dominated $q = -2$	de Sitter $q = 1$ & $\nu \in [0, \frac{1}{2})$	de Sitter $q = 1$ & $\nu \in (\frac{1}{2}, \frac{3}{2})$
$< t_{00}(\eta, \vec{x}) t_{00}(\eta, \vec{x'}) >$	0	∞	0	∞
$\frac{< t_{00}(\eta, \vec{x}) t_{00}(\eta, \vec{x'}) >}{(a(\eta))^4}$	0	∞	0	0

In the light of above results, we conclude that the semiclassical analyses alone may not be sufficient to draw any robust predictions about physical observables and questions related to the stability of the spacetimes. For example, for universe with positive equation of state parameter, the semiclassical analysis can still be trusted, at least in this stochastic approach, however, one may need to seriously relook at first order results for fluids with negative equation of state parameter and need to supplement them with, at least, 2nd order calculations. Similarly, for very light scalar fields, stochastic calculations may render the semiclasical conclusions invalid.[1,12]

References

1. L. E. Parker and D. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Field and Gravity*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 8 2009).
2. N. G. Phillips and B. L. Hu, Noise kernel in stochastic gravity and stress energy bitensor of quantum fields in curved spacetimes, *Phys. Rev. D* **63**, p. 104001 (Apr 2001).
3. B. L. Hu and E. Verdaguer, Stochastic Gravity: Theory and Applications, *Living Rev. Rel.* **11**, p. 3 (2008).
4. G. Perez-Nadal, A. Roura and E. Verdaguer, Stress tensor fluctuations in de Sitter spacetime, *JCAP* **1005**, p. 036 (2010).
5. L. H. Ford and T. A. Roman, Minkowski vacuum stress tensor fluctuations, *Phys. Rev. D* **72**, p. 105010 (2005).
6. Y. Akrami *et al.*, Planck 2018 results. X. Constraints on inflation, *Astron. Astrophys.* **641**, p. A10 (2020).
7. M. B. Fröb, Fully renormalized stress tensor correlator in flat space, *Phys. Rev. D* **88**, p. 045011 (2013).
8. B. Allen, Vacuum states in de sitter space, *Phys. Rev. D* **32**, 3136 (Dec 1985).
9. M. Spradlin, A. Strominger and A. Volovich, Les Houches lectures on de Sitter space, in *Les Houches Summer School: Session 76: Euro Summer School on Unity of Fundamental Physics: Gravity, Gauge Theory and Strings*, 10 2001.

10. A. Dhanuka and K. Lochan, Stress energy correlator in de Sitter spacetime: Its conformal masking or growth in connected Friedmann universes, *Phys. Rev. D* **102**, p. 085009 (2020).
11. K. Lochan, K. Rajeev, A. Vikram and T. Padmanabhan, Quantum correlators in Friedmann spacetimes: The omnipresent de Sitter spacetime and the invariant vacuum noise, *Phys. Rev. D* **98**, p. 105015 (2018).
12. L. H. Ford, Quantum instability of de sitter spacetime, *Phys. Rev. D* **31**, 710 (Feb 1985).

Breaking Buchdahl: Ultracompact stars in semiclassical gravity

Julio Arrechea*, Carlos Barceló†

Institute of Astrophysics of Andalusia (IAA-CSIC),
Glorieta de la Astronomía, 18005 Granada, Spain
**E-mail: arrechea@iaa.es, †E-mail: carlos@iaa.es*

Raúl Carballo-Rubio

Florida Space Institute, University of Central Florida,
12354 Research Parkway, Partnership 1, 32826 Orlando, FL, USA
E-mail: Raul.CarballoRubio@ucf.edu

Luis J. Garay

Universidad Complutense de Madrid, 28040 Madrid, Spain
and
Instituto de Estructura de la Materia (IEM-CSIC),
Serrano 121, 28006 Madrid, Spain
E-mail: luisj.garay@ucm.es

The semiclassical approximation takes into account the gravitational contribution of zero-point energies. We model this contribution via the renormalized stress-energy tensor (RSET) of a massless scalar field, which we compute in a cutoff-regularized version of the Polyakov approximation. When the field is in the Boulware vacuum state (the natural vacuum for stellar geometries), the RSET works in favor of violating the Buchdahl compactness limit. We review the family of classical constant-density stellar solutions, paying particular attention to the notion of criticality—the presence of offsets in the mass function—and use it as a warm up for the analysis of the semiclassical set of solutions. For stars that surpass Buchdahl limit by far, the critical solution has an irregular pressure. This divergence in pressure moves inward by introducing a negative offset in the mass. In the semiclassical theory we find something rather different, namely that the critical configuration already displays a pressure that diverges exactly at the center of the structure. This drastic difference between the classical and semiclassical space of solutions suggests that semiclassical gravity could potentially allow for the existence of ultracompact stellar objects.

Keywords: Quantum fields in curved spacetimes, relativistic stars, semiclassical gravity

1. Introduction

In recent times, new windows have opened towards probing the nature of astrophysical black holes through both gravitational wave[1] and electromagnetic[2] observations. As observations aspire to unveiling the nature of these dark and compact objects, the issue of guessing what lurks behind their event horizons, assuming these long-lived surfaces exist in the first place, becomes more pressing than ever[a]. Among all

[a]Expanded version of the contributed talk by J.A. for the Sixteenth Marcel Grossmann Meeting (2021).

the different theories that, in principle, manage to change our notion of theoretical (Schwarzschild) black holes, semiclassical gravity stands out, mainly because it exhibits a conceptually well-stablished framework that has led to the discovery of effects such as cosmological particle creation[3] and Hawking evaporation.[4]

The main idea behind semiclassical gravity is that an effective geometrical description of the spacetime is retained, while the sources of this classical spacetime are of quantum origin. Therefore, this approximate scheme could be regarded as a conservative approach towards a theory of quantum gravity, in the sense that it captures the first corrections to classical general relativity originated by the zero-point energies of quantum matter fields. The presence of spacetime curvature modifies these zero-point energies in a way that cannot be renormalized away, the subsequent contribution from these zero-point energies to spacetime curvature being captured by the semiclassical field equations

$$G_{\mu\nu} = 8\pi \left(T_{\mu\nu} + \hbar \langle \hat{T}_{\mu\nu} \rangle \right), \tag{1}$$

where $\langle \hat{T}_{\mu\nu} \rangle$ is the vacuum expectation value of the renormalized stress-energy tensor (RSET) of quantized fields. Obtaining analytical expressions for the RSET is an arduous task,[5] so it is typical to resort to highly symmetrical or dimensionally reduced models where the RSET can be computed explicitly.[6,7]

2. The Polyakov RSET: Boulware vs Buchdahl

For simplicity, we restrict our analysis to static and spherically symmetric spacetimes with line element

$$ds^2 = -e^{2\phi(r)} dt^2 + \left[1 - C(r)\right]^{-1} dr^2 + r^2 d\Omega^2, \tag{2}$$

where $d\Omega^2$ is the line element of 2-spheres and ϕ, C are the redshift and compactness functions, respectively. The former measures the amount of redshift suffered by outgoing null rays and the latter is a measure of the amount of mass contained within concentrical spheres, namely $C(r) = 2m(r)/r$, with $m(r)$ the Misner-Sharp mass.[8]

An appealing method for obtaining a particularly simple RSET is via the s-wave Polyakov approximation.[9] This approximation models the RSET as that of a massless scalar field propagating in a $1+1$ dimensional spacetime [the t, r sector of (2)]. The resulting RSET is a function of the components of the metric[10] and has the form

$$\begin{aligned} \langle \hat{T}_{rr} \rangle^{(2)} &= -\frac{l_{\mathrm{P}}^2 \psi^2}{2} + \langle \mathrm{SDT} \rangle, \quad \langle \hat{\mathrm{T}}_{\mathrm{rt}} \rangle^{(2)} = \langle \hat{\mathrm{T}}_{\mathrm{tr}} \rangle^{(2)} = 0, \\ \langle \hat{T}_{tt} \rangle^{(2)} &= \frac{l_{\mathrm{P}}^2 e^{2\phi}}{2} \left[2\psi' \left(1 - C\right) + \psi^2 \left(1 - C\right) - \psi C' \right] + \langle \mathrm{SDT} \rangle, \end{aligned} \tag{3}$$

where $\langle \mathrm{SDT} \rangle$ stands for state dependent terms, $l_{\mathrm{P}} = 1/\sqrt{12\pi}$, $\psi = \phi'$ and the $'$ denotes derivatives with respect to the r coordinate. These components are then

identified with those of a $3+1$ dimensional tensor

$$\langle \hat{T}_{\mu\nu} \rangle^{(\mathrm{P})} = F(r) \delta^a_\mu \delta^b_\nu \langle \hat{T}_{ab} \rangle^{(2)}, \tag{4}$$

where Latin indices take $0, 1$ values and Greek indices take $0, 3$ values.

The Polyakov RSET provides an appropriate balance between simplicity and accuracy (see[11] for an extended discussion). There is an ambiguity in the $3+1$-dimensional Polyakov RSET given by the radial function $F(r)$. Typically, this function is fixed to be $F(r) = 1/r^2$ to yield a covariantly conserved RSET with vanishing angular pressures. In turn, the resulting RSET is singular at $r = 0$, even computed over entirely regular stellar spacetimes. This poses a problem when backreaction is taken into account and it is the main motivation behind exploring other choices of regulator function. Following a conservative logic, we supplied the regulator with a cut-off of the form $F_{\mathrm{CP}}(r) = 1/(r^2 + \alpha l_{\mathrm{P}}^2)$, with $\alpha > 1$. This distortion of the regulator function induces a non-conservation of the RSET, which can be compensated by the introduction of angular pressures.[12]

In previous works[12,13] we explored the sets of semiclassical vacuum and electrovacuum solutions using the cut-off regulator Polyakov RSET. In what follows we will consider the Boulware vacuum state, obtained by taking $\langle \mathrm{SDT} \rangle = 0$, as it is the sole state compatible with staticity (absence of stationary fluxes) and asymptotic flatness. The Boulware state is singular at the Schwarzschild event horizon, hence backreaction will affect it in a non-perturbative manner. The semiclassical counterpart of the Schwarzschild black hole has its horizon replaced by the throat of an asymmetric wormhole. At finite affine distance inside this throat lives an undressed singularity. This result motivates the introduction of a classical matter fluid as the only semiclassically-consistent possibility of obtaining geometries devoid from singularities. We will explore whether these geometries are potentially able to mimic astrophysical black holes.

This potentiality comes from realizing that semiclassical physics provides several ways of surpassing the Buchdahl compactness bound[14] that applies to relativistic stars in hydrostatic equilibrium. The conditions for deriving this limit involve that (i) the star has a Schwarzschild exterior, (ii) pressures in the angular directions that do not surpass the pressure in the radial direction, and (iii) a density profile that is non-increasing outwards. Semiclassical corrections modify the exterior (vacuum) geometry so that it is no longer Schwarzschild but an asymmetric wormhole. The RSETs are anisotropic by construction, although this anisotropy is underestimated in regularized Polyakov-like approximations. Lastly, the violations of energy conditions provided by the RSET generate regions inside the star where the energy density decreases inwards.

In this contribution, we report on several results concerning the self-consistent solutions to (1). We will point out the remarkable differences that exist between constant-density stellar solutions in the classical and semiclassical theories. We will elaborate on the notion of criticality—a criteria for identifying stellar geometries that exhibit deficit or excess of mass that come from singular sources—and use

it to obtain a family of *quasi-regular* semiclassical geometries. These quasi-regular geometries are characterized by a central (singular) core such that at its boundary the mass function is not trans-Planckian and pressures are finite.

3. Criticality in classical stars

First we give a brief summary of classical stellar equilibrium applied to the constant-density perfect fluid. In this simplified setting, the concept of criticality can be introduced in a clear fashion, paving the path toward the more involved analysis involving semiclassical corrections. The stress-energy tensor (SET) of the isotropic perfect fluid is

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}, \tag{5}$$

with p and ρ denoting the pressure and energy density measured by an observer comoving with the fluid with 4-velocity u^μ. In the following we will consider that the fluid obeys an equation of state of the form

$$\rho = \text{const.} \tag{6}$$

There are several reasons behind considering a constant-density fluid. First, this equation of state leaves the pressure of the fluid free to arrange itself in whichever form necessary to attain equilibrium. Second, the classical field equations admit (in some cases) analytical solutions. This characteristic is lost once we introduce semiclassical corrections. Finally, the constant-density assumption saturates one of the conditions necessary for the Buchdahl limit to hold.[15]

Covariant conservation of the classical SET yields

$$p' = -(\rho + p)\, \psi. \tag{7}$$

The rr and tt components of the classical field equations (1) take the form

$$C = \frac{-8\pi r^2 p + 2r\psi}{1 + 2r\psi}, \quad C' = 8\pi r\rho - \frac{C}{r}. \tag{8}$$

Expressions (6, 7, 8) form a closed system. The differential equation for C can be directly integrated, yielding

$$C = \frac{8}{3}\pi r^2 \rho + \frac{2m_0}{r}. \tag{9}$$

Here, m_0 is an arbitrary integration constant that accounts for a constant shift in the Misner-Sharp mass $m(r)$. The following relation

$$M_{\text{ADM}} = \int_0^R 4\pi r^2 \rho \, \mathrm{d}r + m_0, \tag{10}$$

calls for interpreting m_0 as a term that compensates the disagreement between the mass generated by the whole fluid sphere and the ADM mass M_{ADM}. Therefore, regularity in the compactness of a classical star enforces $\rho = \rho_{\text{crit}} = 3C(R)/8\pi R^2$, where ρ_{crit} stands for the critical density. We use this value to distinguish three

families of solutions

- Sub-critical regime ($\rho < \rho_{\rm crit}$): These solutions are characterized by a positive m_0 and can be interpreted as a fluid sphere with a mass excess whose gravitational pull cannot be withstood by finite pressures. Sub-critical solutions have infinite pressure singularities at some $r > 0$. For a fixed $C(R)$, the radius where this singularity appears nears the surface as ρ decreases.
- Critical regime ($\rho = \rho_{\rm crit}$): These solutions have a vanishing mass at $r = 0$ ($m_0 = 0$) and can be regular or irregular depending on whether the Buchdahl compactness bound is being surpassed. Critical solutions that surpass Buchdahl limit ($C(R) = 8/9$) display a pressure divergence at

$$R_{\rm div} = 3R\sqrt{1 - \frac{8}{9C(R)}}. \tag{11}$$

- Super-critical regime ($\rho > \rho_{\rm crit}$): Stellar geometries characterized by a negative m_0. The gravitational repulsion exherted by this negative mass tames the growing of the pressure of the fluid but at the cost of producing a curvature singularity at $r = 0$. As m_0 becomes more negative, the pressure at the core decreases, eventually becoming finite at $r = 0$ for some separatrix density.

Despite its apparent simplicity, this space of solutions reveals an important characteristic: for stars that surpass the Buchdahl limit, $C(R) = 8/9$, the respective separatrices for the pressure and mass functions do not overlap in any region. Fig. 1 shows integrations surrounding the critical solution, the only one with a regular compactness function[b]. The associated pressure profiles all develop singularities significantly far from $r = 0$. On the other hand, Fig. 2 shows integrations surrounding the solution separatrix in pressure, that is, the solution which is super-critical enough as to have a pressure that diverges exactly at $r = 0$. However, the price to pay for generating finite-pressure configurations is that the mass goes highly super-critical. At this point, enforcing the mass function to vanish at $r = 0$ would require introducing an additional positive mass contribution inside a small core. This sort of regularization cannot come without additional pathologies as the density of such sphere would necessarily be trans-Planckian. In what follows we will show that semiclassical corrections naturally provide this regularization mechanism for the mass, although they fail in achieving a strictly regular behavior at $r = 0$.

4. Semiclassical corrections in stellar spacetimes

The classical equations (8) guarantee the behavior of the compactness is independent from that of the pressure, although the converse is not true. The addition

[b]Recall that $C = 2m/r$, so only a Misner-Sharp mass that vanishes at the origin returns a non-singular compactness.

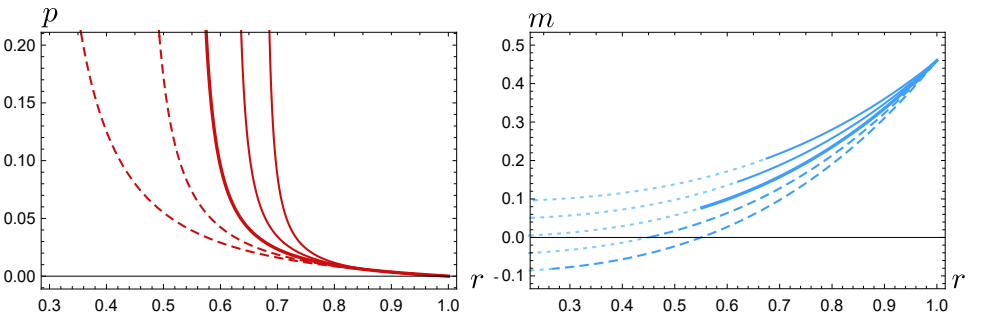


Fig. 1. Pressure (left panel) and Misner-Sharp mass (right panel) profiles surrounding the critical solution of a star with $C(R) = 0.92$ and $R = 1$. The thick, continuous line denotes the critical solution, while thin lines are sub-critical and dashed lines are super-critical stars. The only solution regular in the mass (critical) has a highly irregular pressure. The mass curves have been analytically extended for pictorical purposes, but the associated pressure profiles are singular. From right to left (top to bottom) $\rho/\rho_{\rm crit} = \{0.8, 0.9, 1.0, 1.1, 1.2\}$.

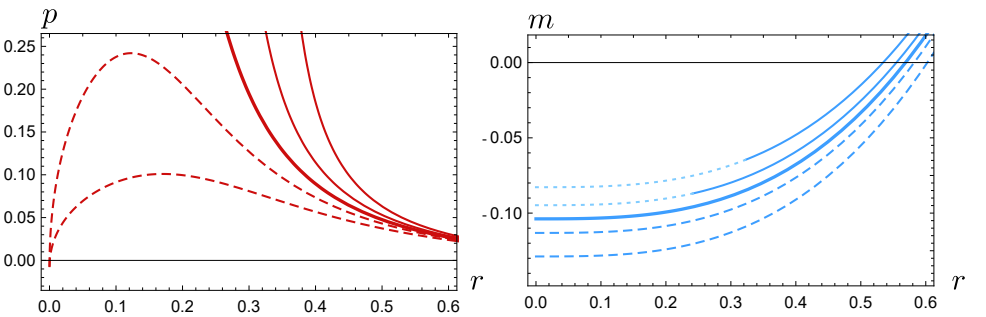


Fig. 2. Pressure (left panel) and Misner-Sharp mass (right panel) profiles surrounding the pressure-separatrix solution of a star with $C(R) = 0.92$ and $R = 1$. The thick, continuous line denotes the separatrix solution, while thin lines are infinite-pressure solutions and dashed lines are finite-pressure ones. Note that making the pressure everywhere finite enforces a negative central mass. All integrations are super-critical because $m_0 < 0$. From right to left (top to bottom) $\rho/\rho_{\rm crit} \simeq \{1.18, 1.21, 1.23, 1.25, 1.28\}$.

of semiclassical corrections intertwines the behavior of both functions. As a consequence, when we analyze the semiclassical equations, the notion of criticality becomes more involved. This is because the differential equation for the compactness cannot be integrated analytically, so we lack an explicit expression that informs about the value of $\rho_{\rm crit}$. The most relevant departure of semiclassical stars with respect to their classical counterparts is that, in the semiclassical theory, the separatrix solutions in mass and pressure overlap in a narrow region of the parameter space. In the following we will catalogue the space of semiclassical constant-density solutions and elaborate on this aspect.

Introducing the cut-off regularized Polyakov RSET (4) as a curvature source in the field equations (1), they result in

$$C = \frac{-8\pi r^2 p + 2r\psi + l_{\rm P}^2 r^2\psi^2/\left(r^2+\alpha l_{\rm P}^2\right)}{1 + 2r\psi + l_{\rm P}^2 r^2\psi^2/\left(r^2+\alpha l_{\rm P}^2\right)}, \tag{12}$$

$$C' = \frac{8\pi r^2\rho - C + l_{\rm P}^2 r^2\left(\psi^2 + 2\psi'\right)(1-C)/\left(r^2+\alpha l_{\rm P}^2\right)}{r + l_{\rm P}^2 r^2\psi/\left(r^2+\alpha l_{\rm P}^2\right)}. \tag{13}$$

Semiclassical corrections to stellar spacetimes are typically negligible (they are proportional to $\hbar$) unless the geometries under consideration are able to explore and stimulate the energetic contributions of the Boulware state as to compensate their suppression of order $\hbar$. The scale that can potentially compensate this screening of semiclassical effects is the compactness of the star. In the following, we will be considering spheres of fluid whose surface is located very close and above the throat of the vacuum wormhole geometry [$C(R) \to 1$]. Such a configuration is defined by the following boundary conditions

$$p(R) = 0, \quad \phi(R) = \phi_{\rm S}, \quad \psi(R) = \frac{R^2+\alpha l_{\rm P}^2}{R l_{\rm P}^2}\left[\sqrt{1 + \frac{R^2}{R^2+\alpha l_{\rm P}^2}\frac{C(R)}{1-C(R)}} - 1\right], \tag{14}$$

and a given value of ρ and $C(R)$. Here, the parameter $\phi_{\rm S}$ is obtained through a numerical integration of the vacuum equations from the asymptotic region inwards.

Plugging the boundary conditions (14) in the field equations (12, 13) we obtain expressions for $\psi'(R)$ and $C'(R)$. Then, the semiclassical energy density and radial pressure at $r = R$ are

$$\begin{aligned}
\rho_{\rm sc} &= -\langle\hat{T}_t^t\rangle^{(\rm P)} = -\frac{1}{8\pi R^2} + \rho + \mathcal{O}\left(\sqrt{1-C}\right), \\
p_{\rm sc}^r &= \langle\hat{T}_r^r\rangle^{(\rm P)} = -\frac{1}{8\pi R^2} + \mathcal{O}\left(\sqrt{1-C}\right), \\
p_{\rm sc}^\theta &= \langle\hat{T}_\theta^\theta\rangle^{(\rm P)} = -\frac{\alpha l_{\rm P}^2}{8\pi R^2\left(R^2+\alpha l_{\rm P}^2\right)} + \mathcal{O}\left(\sqrt{1-C}\right).
\end{aligned} \tag{15}$$

The RSET components are comparable in magnitude to the classical SET at the surface. In fact, as long as $\rho \lesssim 1/8\pi R^2$ the total RSET violates energy conditions at the surface. Additionally, for more realistic equations of state with a classical energy density that vanishes at the surface, the total RSET will violate all pointwise energy conditions there.

Vacuum polarization provides an extra contribution to the Misner-Sharp mass of the spacetime. This contribution is negative in the vacuum region and permeates it entirely. In the interior of the star it can be overall positive or negative depending on the value of ρ. The equivalent to formula (10) in the semiclassical theory is

$$M_{\rm ADM} = \int_0^\infty 4\pi r^2\left[\Theta\left(R-r\right)\rho + \rho_{\rm sc}\right] dr + m_0, \tag{16}$$

where Θ is the Heaviside step function. Since the total energy density is now intertwined to the pressure and redshift of the geometry through Eq. (13), the critical

value of the classical density (recall, the value that corresponds to a vanishing m_0) cannot be determined analytically. In turn, a numerical exploration of the space of solutions is needed. We dedicate the rest of this note to highlight the chief results from our exploration[c].

Concerning the Buchdahl limit, we have found that the classical Buchdahl bound of $C(R) = 8/9$ gets perturbatively modified by semiclassical corrections. These correction makes the most compact configuration which is regular in p and C have $C(R) < 8/9$. This result follows from $\rho_{\rm sc}$ in (16) amounting to an overall positive contribution throughout the interior of the structure. In fact, the greatest positive contribution comes from the region $r \sim \sqrt{\alpha} l_{\rm P}$, precisely where the physics of the Polyakov RSET is driven by the regulator function. Using a local, more refined approximation to the RSET[16] the energetic contribution near the center of constant-density stars that are approaching Buchdahl limit turns out to be negative. The cutoff-Polyakov RSET could be incorrectly estimating the magnitude and sign of the energetic contribution from the vacuum near the center of compact fluid spheres.

In our exploration of the semiclassical space of solutions (see Fig. 3 for a pictorical representation) we found the following regimes

- Sub-critical regime ($\rho \ll \rho_{\rm crit}$, regions I and III of the diagram in Fig. 3): These solutions are asymmetric wormholes with characteristics akin to the vacuum solution, but filled with a classical fluid of diverging pressures. As ρ is increased, the throat of the wormhole progressively shrinks. When the throat is very small compared to the overall size of the star, we identify a new regime not present in the classical theory.
- Quasi-regular regime ($\rho \sim \rho_{\rm crit}$, narrow orange band in Fig. 3): This subfamily is comprised by super-critical solutions that lie below Buchdahl limit and sub-critical solutions of any compactness with an extremely tiny wormhole throat. In this regime, we can always identify a central core of radius $r_{\rm core} \ll R$. At the boundary of this core the pressure remains finite and the Misner-Sharp mass can be made as small as desired.
- Critical regime ($\rho = \rho_{\rm crit}$, black line in Fig. 3): Sub-Buchdahl stars amount to a perturbative correction over the classical geometries and are regular. Super-Buchdahl stars are irregular but extend all the way to $r = 0$, contrary to their classical counterparts. These solutions have infinite positive pressure and infinite negative mass at $r = 0$.
- Super-critical regime ($\rho > \rho_{\rm crit}$, regions II and IV of the diagram in Fig. 3): They correspond to perturbative corrections over the classical solutions, with the caveat that the Misner-Sharp mass is now infinite and negative at $r = 0$, which strengthens the central singularity. Pressure is finite everywhere.

[c]We refer the reader to[11] for details.

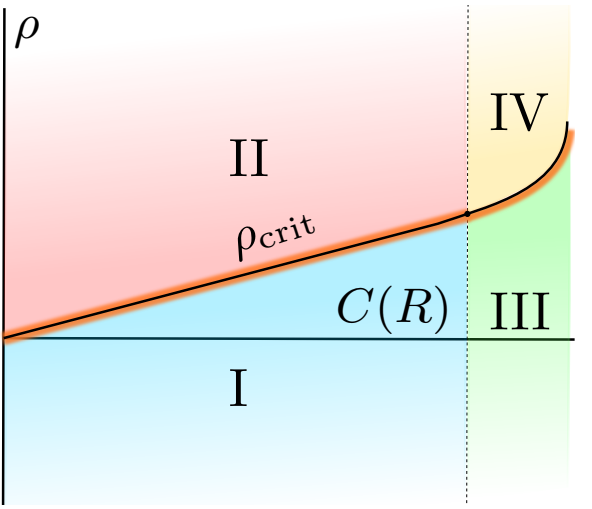


Fig. 3. Pictorial representation of an $R = \text{const.} \gg l_{\rm P}$ slice of the semiclassical constant-density stellar space of solutions. The vertical and horizontal axes represent the energy density and surface compactness of stars. The black curve denotes critical configurations. Its intersection with the vertical, dotted line represents the Buchdahl limit, or the most compact structure regular in both p and C. We distinguish four regions: regions I and II for sub- and super-critical stars below Buchdahl limit and regions III and IV for sub- and super-critical stars above Buchdahl limit. The narrow orange band that surrounds the $\rho_{\rm crit}$ line denotes the quasi-regular regime which spans through regions I, II and III.

Figure 4 shows the pressure and mass profiles for various configurations surrounding the critical solution of a star that surpasses the Buchdahl limit (rightmost portion of diagram 3). There are several crucial differences between these diagrams and those from Figures 1 and 2. Firstly, the critical solution (thick line), albeit singular, coincides with the separatrix solution for the pressure. In the classical case, the critical solution and the separatrix in the pressure were distant in the space of parameters. Here, because of the way quantum corrections operate, we find a coincidence between these two solutions. Secondly, quasi-regular configurations (thin lines) for which the Misner-Sharp mass vanishes near the center of spherical symmetry are very different from their classical counterparts. Enforcing the mass function to vanish at $r = 0$ no longer requires the introduction of arbitrarily high masses inside small cores: Semiclassical corrections already provides a significant part of this contribution.

5. Conclusions

In this work we have reviewed the classical stellar solutions of constant density under our particular scope, with the aim of understanding properly the solutions that incorporate vacuum polarization contributions. We have not been able to find strictly regular configurations that surpass a Buchdahl limit (which receives perturbative

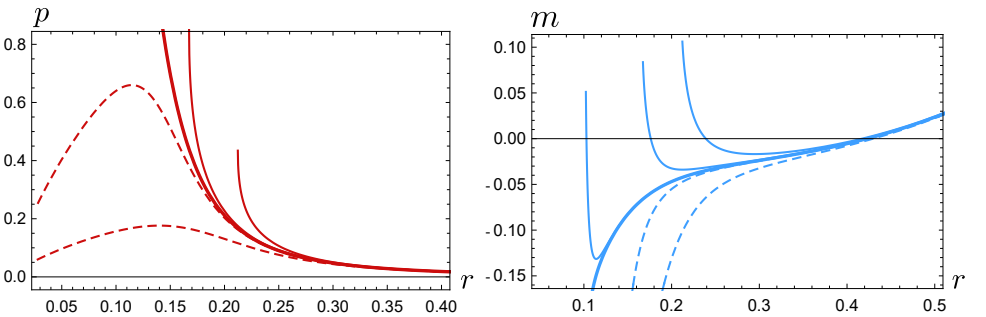


Fig. 4. Pressure (left panel) and Misner-Sharp mass (right panel) profiles surrounding the critical solution of a semiclassical star with $C(R) = 0.92, R = 1$ and $\alpha = 1.01$. The thick line denotes the critical solution (the separatrix between distinct compactness behaviors), while the thin lines are sub-critical and the dashed lines are super-critical. In the right panel the quasi-regular regime can be observed. Quasi-regular solutions are characterized by a mass function that vanishes in the innermost regions of the star. By a suitable choice of ρ, this vanishing can be achieved as close to $r = 0$ as desired. From right to left (top to bottom) the densities utilized are $\rho/\rho_{\rm crit} = \{0.9990, 0.9998, 0.9999, 1, 1.0001, 1.0010\}$, where the critical density $\rho_{\rm crit} \simeq 0.1214$ is determined numerically within a precision of 10^{-6}.

corrections). However, the study of singular solutions hints towards semiclassical gravity being able to regularize at the same time the pressure and mass of ultracompact configurations. This notion is what we have attempted to capture in the notiont of quasi-regular configurations.

We believe that the crudity of the approximations used here for the semiclassical source could be preventing the quasi-regular solutions from appearing as regular in the first place. Ideally, these analyses should be revisited when more accurate approximations to the RSET are available. As an intermediate step in this direction, it is possible to ask whether a deformation of the regulator function from its cutoff-regulator form $F_{\rm CP}$ just inside the central core of the structure is sufficient to generate entirely regular and ultracompact configurations. The results of this ongoing investigation will appear in a forthcoming publication.

References

1. B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, Affeldt and et. al., Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, p. 061102 (Feb 2016).
2. K. Akiyama *et al.*, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, p. L1 (2019).
3. L. Parker, Particle creation in expanding universes, *Phys. Rev. Lett.* **21**, 562 (Aug 1968).
4. S. W. Hawking, Breakdown of predictability in gravitational collapse, *Phys. Rev. D* **14**, 2460 (Nov 1976).
5. R. M. Wald, Axiomatic Renormalization of the Stress Tensor of a Conformally Invariant Field in Conformally Flat Space-Times, *Annals Phys.* **110**, 472 (1978).

6. M. R. Brown and A. C. Ottewill, Effective actions and conformal transformations, *Phys. Rev. D* **31**, 2514 (May 1985).
7. V. P. Frolov and A. I. Zel'nikov, Killing approximation for vacuum and thermal stress-energy tensor in static space-times, *Phys. Rev. D* **35**, 3031 (May 1987).
8. C. W. Misner and D. H. Sharp, Relativistic equations for adiabatic, spherically symmetric gravitational collapse, *Phys. Rev.* **136**, B571 (1964).
9. A. M. Polyakov, Quantum Geometry of Bosonic Strings, *Phys. Lett.* **B103**, 207 (1981).
10. P. C. W. Davies and S. A. Fulling, Quantum vacuum energy in two dimensional space-times, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **354**, 59 (1977).
11. J. Arrechea, C. Barceló, R. Carballo-Rubio and L. J. Garay, Semiclassical constant-density spheres in a regularized Polyakov approximation, *Phys. Rev. D* **104**, p. 084071 (2021).
12. J. Arrechea, C. Barceló, R. Carballo-Rubio and L. J. Garay, Schwarzschild geometry counterpart in semiclassical gravity, *Phys. Rev. D* **101**, p. 064059 (Mar 2020).
13. J. Arrechea, C. Barceló, R. Carballo-Rubio and L. J. Garay, Reissner-Nordström geometry counterpart in semiclassical gravity (2 2021).
14. H. A. Buchdahl, General relativistic fluid spheres, *Phys. Rev.* **116**, 1027 (Nov 1959).
15. A. Urbano and H. Veermäe, On gravitational echoes from ultracompact exotic stars, *JCAP* **04**, p. 011 (2019).
16. W. A. Hiscock, Gravitational vacuum polarization around static spherical stars, *Phys. Rev. D* **37**, 2142 (Apr 1988).

Einstein anomaly with tensors of odd order in six dimensional curved space

Kohei Yamamoto* and Satoshi Yajima

Department of Physics, Kumamoto University,
2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan
**E-mail: 191d9021@st.kumamoto-u.ac.jp*

By applying the covariant Taylor expansion method to the heat kernel, Einstein anomaly associated with the Weyl fermion of spin 1/2 interacting with tensor fields of 1 and 3 order in six dimensional curved space are given. From the relation between Einstein and Lorentz anomalies, which are the gravitational anomalies, all terms of the Einstein anomaly should form total derivatives.

Keywords: Einstein anomaly

1. Introduction

The gravitational anomalies are the Lorentz anomaly that appears from the local Lorentz transformation and the general coordinate anomaly that appears from the covariant general coordinate transformation, in the model for the Weyl fermion in even dimensional curved space. The Lorentz anomaly can be canceled by appropriately replacing the expectation value of the energy-momentum tensor, and then the general coordinate anomaly, called the Einstein anomaly, remains.

The Lagrangian in $N = 1$ supergravity coupled to supersymmetric Yang-Mills theory has some four-Fermi interactions, in addition to the minimal couplings of the fermion with the gravitational and gauge fields. The gravitino field with an irreducible element of spin $\frac{3}{2}$ exists in the supergravity. The kinetic part of the gravitino is treated as that for a fermion of spin $\frac{1}{2}$, fixing the field by a suitable gauge, and the spacetime index of the gravitino may be regarded as that of the representation matrix elements of a gauge group in vector and odd order tensor type couplings. Then the tensor may not commute with the vector.

2. Heat kernel

In this article, in order to apply for the above model, we consider the covariant gravitational anomalies for a Weyl fermion with spin $\frac{1}{2}$, interacting with both the vector field V_μ and the totally antisymmetric tensor $C_{\mu\alpha\beta}$, in even $2n$-dimensional Riemannian space. The concrete form of the anomalies may directly be calculated by using the heat kernel $K^{(2n)}(x, x'; t)$ for the fermion, which is defined as[1]

$$\begin{aligned} -\frac{\partial}{\partial t}K^{(2n)}(x,x';t) &= \not{D}^2 K^{(2n)}(x,x';t), \\ K^{(2n)}(x,x';0) &= \mathbf{1}|h(x)|^{-\frac{1}{2}}|h(x')|^{-\frac{1}{2}}\delta^{(2n)}(x-x'), \end{aligned} \tag{1}$$

where $\mathbf{1}$ is the spinor unit matrix, $\delta^{(2n)}(x-x')$ denotes the delta function, and $h(x)$ stand for the determinant $\det h^m{}_\mu$ of a vielbein $h^m{}_\mu$. Note that the metric tensor is $g_{\mu\nu} = h^m{}_\mu h^n{}_\nu \eta_{mn}$ with $\eta_{mn} = -\delta_{mn}$ in flat tangent space. Moreover, the square of the Dirac operator $\not{D}$ for fermion ψ is written as follows,

$$\begin{aligned}
&\not{D}^2 = D_\mu D^\mu + X,\ \not{D} = \gamma^\mu \nabla_\mu + Y,\ D_\mu = \nabla_\mu + Q_\mu,\ Q_\mu = \frac{1}{2}\{\gamma_\mu, Y\},\\
&X = Z - \nabla_\mu Q^\mu - Q_\mu Q^\mu,\ \nabla_\mu \psi = \partial_\mu \psi + \frac{1}{4}\gamma_{mn}\omega^{mn}{}_\mu \psi,\ \gamma_{\mu_1\cdots\mu_n} = \gamma_{[\mu_1}\cdots\gamma_{\mu_n]},\\
&Z = \frac{1}{2}\gamma^{\mu\nu}[\nabla_\mu, \nabla_\nu] + \gamma^\mu(\nabla_\mu Y) + Y^2,\ [D_\mu, D_\nu]\psi = \Lambda_{\mu\nu}\psi,\\
&Y = \gamma^\mu V_\mu + \gamma^\mu \gamma^{\alpha\beta} C_{\mu\alpha\beta},\ V_\mu \equiv V_\mu^a T^a,\ C_{\mu\alpha\beta} \equiv C_{\mu\alpha\beta}^a T^a,
\end{aligned} \tag{2}$$

where $\omega^{mn}{}_\mu$ is the spin connection. The interaction term Y of the boson in $\not{D}$ is given by vector and tensor fields, and T^a is the matrix representation of the nonabelian gauge group. The γ-matrix valued Q_μ, X and $\Lambda_{\mu\nu}$ in (2) are rewritten in the following form,

$$\begin{aligned}
Q_\mu &= V_\mu + 3\gamma^{\alpha\beta} C_{\mu\alpha\beta},\ F_{\mu\nu} \equiv \partial_\mu V_\nu - \partial_\nu V_\mu + [V_\mu, V_\nu],\\
X &= -\frac{1}{4}R + 12 C^{\alpha\beta\gamma} C_{\alpha\beta\gamma} + \gamma^{\mu\nu}\left(\frac{1}{2}F_{\mu\nu} + 18 C_\mu{}^{\alpha\beta} C_{\nu\alpha\beta}\right)\\
&\quad - \gamma^{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma;\delta} + \gamma^{\mu\nu\alpha\beta\gamma\delta} C_{\mu\alpha\beta} C_{\nu\gamma\delta},\\
\Lambda_{\mu\nu} &= \frac{1}{4}\gamma^{\rho\sigma} R_{\rho\sigma\mu\nu} + F_{\mu\nu} - 18[C_\mu{}^{\alpha\beta}, C_{\nu\alpha\beta}] - \gamma^{\alpha\beta}(6 C_{\alpha\beta[\mu;\nu]} + 72 C^\rho{}_{\alpha[\mu|} C_{\rho\beta|\nu]})\\
&\quad + 18\gamma^{\alpha\beta\gamma\delta} C_{[\mu|\alpha\beta} C_{|\nu]\gamma\delta},\ \gamma_{2n+1} = i^n \gamma^1\gamma^2\cdots\gamma^{2n},
\end{aligned} \tag{3}$$

where the semi-colon ';μ' denotes the covariant differentiation $\nabla_\mu + V_\mu$ for the vector gauge transformation in curved space, $R_{\rho\sigma\mu\nu}$ is the Riemann curvature tensor, and $R = g^{\rho\mu} g^{\sigma\nu} R_{\rho\sigma\mu\nu}$. The antisymmetric product $\gamma^{\alpha\beta\gamma\delta}$ and $\gamma^{\mu\nu\alpha\beta\gamma\delta}$ of X can be replaced by $-\frac{i}{2}\gamma_7\gamma^{\rho\sigma}\epsilon_{\rho\sigma\alpha\beta\gamma\delta}$ and $i\gamma_7\epsilon^{\mu\nu\alpha\beta\gamma\delta}$ in 6 dimensions, respectively.

It is difficult to obtain the exact form of the heat kernel. In order to perform the concrete calculation, the following ansatz by DeWitt is applied to the heat kernel.[2]

$$K^{(2n)}(x, x'; t) \sim \frac{\Delta^{\frac{1}{2}}(x, x')}{(4\pi t)^n} \exp\left\{\frac{\sigma(x, x')}{2t}\right\} \sum_{q=0}^{\infty} a_q(x, x') t^q, \tag{4}$$

where $\sigma(x, x')$ is half of the square of the geodesic distance between x and x', $\Delta(x, x') = |h(x)|^{-1}|h(x')|^{-1} \det \sigma_{;\mu\nu'}$, and $a_q(x, x')$ is a bispinor. The coincidence limit of a_n appears in formal expression of the anomaly, and is defined by $\lim_{x'\to x} a_n(x, x') \equiv [a_n](x)$.[3] The orthonormal bases $|n\rangle \equiv |\mu_1\cdots\mu_n\rangle$ being the

eigenfunctions belonging to the eigenvalue n for $\sigma^\nu D_\nu$ consist of n products of $\sigma_{\mu'}(=\nabla_{\mu'}\sigma)$, and a_q can be expanded by the bases,

$$\begin{aligned} a_q &= \sum_{n=0}^{\infty} |n\rangle\langle n|a_q\rangle \ = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\sigma^{\mu'_1}\cdots\sigma^{\mu'_n} \lim_{x\to x'}[D_{(\mu_1}\cdots D_{\mu_n)}a_q], \\ \langle n|a_q\rangle &= [a_{q!(\mu_1\cdots\mu_n)}] = \lim_{x\to x'}[D_{(\mu_1}\cdots D_{\mu_n)}a_q],\ \langle 0|a_q\rangle = [a_q], \end{aligned} \tag{5}$$

where the exclamation mark '$!\mu$' means the covariant differentiation D_μ modified in (2).

3. Einstein anomaly

The formal forms of the general coordinate anomaly $\mathcal{A}^{(2n)}_\nu$ and the Lorentz anomaly $\mathcal{A}^{(2n)}_{\mu\nu}$ are obtained from the path integral measure of left-handed Weyl fermion in 2n dimensions, in Fujikawa's method,[4] and are expressed by the heat kernel after regularized by the Gaussian cut-off differential operator.

$$\begin{aligned} &D^\mu \langle T_{\mu\nu}\rangle = \mathcal{A}^{(2n)}_\nu, \quad \langle T_{\mu\nu}\rangle_{\mathrm{A}} \equiv \frac{1}{2}\left(\langle T_{\mu\nu}\rangle - \langle T_{\nu\mu}\rangle\right) = \mathcal{A}^{(2n)}_{\mu\nu}, \\ &\mathcal{A}^{(2n)}_\nu(x) = \frac{1}{2}\lim_{t\to 0}\lim_{x'\to x}\mathrm{Tr}\left\{\gamma_{2n+1}\left(D_\nu - D_{\nu'}\right)K^{(2n)}\left(x,x';t\right)\right\}, \\ &\mathcal{A}^{(2n)}_{\mu\nu}(x) = \frac{1}{4}\lim_{t\to 0}\lim_{x'\to x}\mathrm{Tr}\left\{\gamma_{2n+1}\gamma_{\mu\nu}K^{(2n)}\left(x,x';t\right)\right\}, \end{aligned} \tag{6}$$

where 'Tr' runs over both indices of γ-matrix and T^a. These anomalies seem to break the general covariance and the local Lorentz symmetry simultaneously, but one of the symmetries is restored by using the relation $2D^\mu\mathcal{A}^{(2n)}_{\mu\nu} = \mathcal{A}^{(2n)}_\nu$ and redefining the expectation value of the energy-momentum tensor.[5] Einstein anomaly $G^{(2n)}_\nu$ is given by $\langle T'_{\mu\nu}\rangle = \langle T_{\mu\nu}\rangle - \mathcal{A}^{(2n)}_{\mu\nu}$. Thus, the local Lorentz symmetry is preserved.

$$D^\mu\left\langle T'_{\mu\nu}\right\rangle_S = G^{(2n)}_\nu = D^\mu\mathcal{A}^{(2n)}_{\mu\nu} = \frac{1}{2}\mathcal{A}^{(2n)}_\nu, \quad \left\langle T'_{\mu\nu}\right\rangle_A = 0, \tag{7}$$

where $\left\langle T'_{\mu\nu}\right\rangle_S$ is the symmetric part of $\left\langle T'_{\mu\nu}\right\rangle$. From (7), it is expected that the concrete form of the Einstein anomaly becomes the total derivative of the Lorentz anomaly. Substituting (4) into (6), $G^{(2n)}_\nu$ is rewritten by the coincidence limit of a_n and its first order derivative in (5),

$$G^{(2n)}_\nu(x) = \frac{1}{4(4\pi)^n}\mathrm{Tr}\left\{2\gamma_{2n+1}[a_{n!\nu}](x) - \gamma_{2n+1}[a_n]_{!\nu}(x)\right\}. \tag{8}$$

The tensorial form of the anomaly in 6 dimensions is obtained from

$$
\begin{aligned}
G_\nu^{(6)} &= \frac{1}{256\pi^3}\,\mathrm{Tr}\left\{\gamma_7\left(2\left[a_{3!\nu}\right]-\left[a_3\right]_{!\nu}\right)\right\} \\
&= \frac{1}{256\pi^3}\,\mathrm{Tr}\left\{\gamma_7\left[\frac{1}{6}\Lambda_{\mu\nu}\left(\frac{1}{6}R+X\right)^2+\frac{1}{45}J_{[\mu}X_{!\nu]}-\frac{1}{60}J_{[\mu!\nu]}X\right.\right. \\
&\quad +\frac{1}{15}\Lambda_{\mu\nu}X_{!\rho}{}^{\rho}+\frac{2}{45}\Lambda_{\mu\nu!\rho}X^{!\rho}+\frac{1}{40}\Lambda_{\mu\nu!\rho}{}^{\rho}X+\frac{1}{180}\left[\Lambda_{\mu\rho},\Lambda_\nu{}^{\rho}\right]X \\
&\quad +\frac{1}{180}R^{\rho}{}_{[\mu}\Lambda_{\nu]\rho}X+\frac{17}{360}R_{\mu\nu\rho\sigma}\Lambda^{\rho\sigma}X+\frac{1}{36}\Lambda_{\mu\nu}\Lambda_{\rho\sigma}\Lambda^{\rho\sigma} \\
&\quad \left.\left.+\frac{1}{45}\Lambda_{[\mu}{}^{\rho}\Lambda_{\nu]}{}^{\sigma}\Lambda_{\rho\sigma}-\frac{1}{90}\Lambda_{\mu\nu!\rho}J^{\rho}-\frac{1}{45}\Lambda_{\rho[\mu}J_{\nu]}{}^{!\rho}\right]^{;\mu}\right\}, \qquad (9)
\end{aligned}
$$

where $J_\mu \equiv \Lambda^{\rho}{}_{\mu!\rho}$. Since $\mathrm{Tr}\gamma_7 X = \mathrm{Tr}\{\gamma_7\Lambda_{\mu\nu}\} = 0$, terms containing X and $\Lambda_{\mu\nu}$ of one degree do not contribute to the anomaly. The concrete form of $G_\nu^{(6)}$ becomes some total derivatives before the trace operation of γ-matrices. Substituting (3) into (9) and performing the trace operation, $G_\nu^{(6)}$ yields many terms in tensorial form.

References

1. J. Schwinger, Phys. Rev. **82**, 664 (1951).
2. B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, 1965).
3. I. G. Avramidi, Nucl. Phys. **B355**, 712 (1991).
4. K. Fujikawa, Phys. Rev. **D21**, 2848 (1980).
5. K. Fujikawa, M. Tomiya and O. Yasuda, Z. Phys. **28**, 289 (1985).

Quantum memory and BMS symmetries

Sanved Kolekar

Indian Institute of Astrophysics,
2nd Block, Koramangala, Bangalore 560034, India
E-mail: sanved.kolekar@iiap.res.in
www.iiap.res.in

Jorma Louko

School of Mathematical Sciences, University of Nottingham,
Nottingham NG7 2RD, UK
E-mail: jorma.louko@nottingham.ac.uk

In this talk, we discuss the effect of BMS symmetries on quantum entanglement and its implications in the context of the black hole information paradox. In particular, we illustrate the gravitational memory effect for linear uniformly accelerated observers in a physical process involving a BMS shock-wave without planar/spherical symmetry. This classical memory is accompanied by a quantum memory that modulates the quantum entanglement between the opposing Rindler wedges in quantum field theory. A corresponding phenomenon across the Schwarzschild black hole horizon suggests that the Negativity measure of entanglement between infalling and outgoing Hawking pair should be degraded due to an infalling BMS shockwave while there should be linear order generation of Negativity between two outgoing Hawking particles. Implications are discussed.

Keywords: Gravitational memory effect, BMS, Negativity

1. Introduction

Bondi-Metzner-Sachs (BMS) showed that asymptotically flat spacetimes possess an infinite number of symmetries now known as the BMS supertranslations.[1] These diffeomorphisms extend the symmetry group of the standard four global translations at the asymptotic boundary. The BMS supertranslations are related to the gravitational memory effect and also the Weinberg's soft theorems via the so called Universal triad relations.[4–8] Based on these relations Hawking et. al. conjectured that applying these relations to an asymptotically flat spacetime with a black hole in the interior would imply the existence of an infinite number of soft hairs[2,3] for the black hole which can have possible implications in the context of the black hole information paradox.

One can analyse the process of implanting soft hairs on a black hole through a physical process involving a BMS type shock wave without spherical symmetry infalling into the horizon.[3] In this talk, we consider a simplified and analogous setting consisting of the Rindler horizon and the linearly uniformly accelerated (LUA) observers, but, in the presence of a BMS type shock wave, having planar asymmetry, and impinging onto the Rindler horizon; so as to analyse the effect of Rindler BMS

type supertranslations on LUA observers and on the Unruh thermal bath. The former constitutes the classical version of the gravitational memory effect on uniformly accelerated observers[9] while the effect on the latter is the quantum memory effect which modulates the quantum entanglement between the opposing Rindler wedges in quantum field theory.[10] A corresponding phenomenon across the Schwarzschild black hole horizon suggests that the Negativity measure of entanglement between infalling and outgoing Hawking pair should be degraded due to an infalling BMS shockwave while there should be linear order generation of Negativity between two outgoing Hawking particles. Implications are discussed.

2. Implanting supertranslational hair: Physical process

We work with the advanced Bondi-type co-ordinates (v, r, x, y) and write the metric of the Rindler spacetime as

$$ds^2 = -2\kappa r dv^2 + 2dvdr + \delta_{AB} dx^A dx^B \tag{1}$$

where $\kappa > 0$, $-\infty < v < \infty$ and $-\infty < r < \infty$. The BMS-type supertranslation diffeomorphism vector for the Rindler horizon which preserves the Bondi-type gauge and also the structure of the Rindler horizon is given by[11–15]

$$\Xi^a = \frac{1}{\kappa}[f(x,y), 0, -r\partial^A f(x,y)] \tag{2}$$

where $f(x, y)$ depends only on the transverse co-ordinates. The Rindler spacetime with a BMS type null shock wave at $v = v_0$ having non-planar symmetry infalling on the Rindler horizon at $r = 0$ is then defined to be[9]

$$\begin{aligned} ds^2 = &-2\kappa r dv^2 + 2dvdr + 4rh(v-v_0)\partial_A f dv dx^A \\ &+\left(\delta_{AB} + 2h(v-v_0)\frac{r}{\kappa}\partial_A\partial_B f\right)dx^A dx^B \end{aligned} \tag{3}$$

where $h(v - v_0)$ is the Heaviside step function. Essentially, the above metric is obtained by joining the Rindler supertranslated metric to the one described in Eq.(1) along the shock wave at $v = v_0$. The stress energy tensor including a surface energy term satisfying the null energy conditions is non zero along the shockwave at $v = v_0$ while it vanishes elsewhere for $v \neq v_0$.

3. Gravitational memory for LUA observers

The linearly uniformly accelerated (LUA) observers are the generalisation of the flat spacetime Rindler trajectories to the case of a curved background spacetime. As the name suggests, the LUA trajectories have constant magnitude of 4 vector acceleration and are also linear, in the sense of the Letaw-Frenet equations in a curved spacetime with vanishing torsion and hyper-torsion.[9] The linearity is essential since it guarantees that the LUA trajectory is locally hyperbolic in all local inertial frames at each point on the trajectory. Such trajectories satisfy the condition

$$u^b \nabla_b a^a = w^a = |a|^2 u^a \tag{4}$$

where u^i and a^i are the 4-velocity and the uniform magnitude 4-acceleration vectors respectively. The constraint in Eq.(4) guarantees that any change in the velocity and acceleration vectors lie in the same plane along the trajectory.

To address the question of finding the LUA trajectories for the case of infalling supertranslated BMS type shock wave on the Rindler horizon as described in section 2, we choose a representative family of orbits of the boost Killing vector $\bar{\xi}^a = \kappa^{-1}(1,0,0,0)$ in the advanced Bondi-type co-ordinates for $v < v_0$, the region before the wave. The worldline for such a family of trajectories is given by $x^a(\tau) = [\tau/\sqrt{2\kappa r_c}, r_c, x_c^A]$ where τ is the proper time and r_c, x_c^A are some fixed constants. These worldline can be extended over the BMS tyoe shock wave at $v = v_0$ by demanding that the constraint in Eq.(4) holds all along the trajectory. Working to linear order in the perturbation, the solution for the 4 velocity can be found to be

$$u^a = \left[\frac{1}{\sqrt{2\kappa r}}, 0, -h(v-v_0)\cosh[\kappa(v-v_0)]\frac{\sqrt{2\kappa r}}{\kappa}\partial^A f\right] \tag{5}$$

while the trajectories are found by integrating with respect to τ to be

$$x^a(\tau) = \left[\frac{\tau}{\sqrt{2\kappa r_c}}, r_c, x_c^A - h\left(\frac{\kappa(\tau-\tau_0)}{\sqrt{2\kappa r_c}}\right)\partial^A f(x_c^A) 2r_c \sinh\left(\frac{\kappa(\tau-\tau_0)}{\sqrt{2\kappa r_c}}\right)\right] \tag{6}$$

where τ_0 is the proper time at $v = v_0$. When the velocity vector in Eq.(5) is transformed to a set of standard Minkowski co-ordinates (T, X, Y^A), it takes the form

$$\begin{aligned} U^a = \Bigg[&\cosh\left(\frac{\kappa\tau}{\sqrt{2\kappa r_c}} - \frac{\log[\kappa r_c]}{2}\right), \sinh\left(\frac{\kappa\tau}{\sqrt{2\kappa r_c}} - \frac{\log[\kappa r_c]}{2}\right), \\ &h\left(\frac{\kappa(\tau-\tau_0)}{\sqrt{2\kappa r_c}}\right)\alpha^A \cosh\left(\frac{\kappa\tau}{\sqrt{2\kappa r_c}} - \kappa v_0\right)\Bigg] \end{aligned} \tag{7}$$

where $\alpha^A = \frac{\sqrt{2\kappa r_c}}{\kappa}\partial^A f(x_c^A)$. Interestingly one can check that a trajectory with given r_c, x_c^A is an integral curve of a boost Killing vector that is obtained by applying to $\bar{\xi}^a$ the Lorentz boost

$$\Lambda^a{}_b = \begin{pmatrix} 1 & 0 & \alpha^Y \cosh\beta & \alpha^Z \cosh\beta \\ 0 & 1 & -\alpha^Y \sinh\beta & -\alpha^Z \sinh\beta \\ \alpha^Y \cosh\beta & \alpha^Y \sinh\beta & 1 & 0 \\ \alpha^Z \cosh\beta & \alpha^Z \sinh\beta & 0 & 1 \end{pmatrix} \tag{8}$$

where $\beta = (1/2)\log[\kappa r_c] - \kappa v_0$. Note that the magnitude and direction of the boost (8) depend on r_c and x_c^A. Hence, we can conclude that even though, before the wave, we started with a family of solutions with the same boost killing vector $\bar{\xi}^a$, after the wave, each trajectory is an orbit of a different boost Killing vector as determined by the Lorentz boost in Eq.(8); thus implanting a supertranslational hair on the Rindler horizon by the non-planar BMS type matter shock wave and boosting a family of Rindler trajectories in a way that differs from trajectory to trajectory through their initial conditions. The trajectory-dependence essentially carries a memory of the

planar inhomogeneity of the wave and constitutes the gravitational memory effect for uniformly linearly accelerated observers.

4. Quantum memory

We consider a real massless Klein Gordon scalar field in the background of the shockwave metric of section 2. We only need to write down the solution of the perturbed Klein Gordon field upto the first order in the supertranslation field f. The Klein-Gordon field equation, $\nabla_a \nabla^a \phi = 0$ for the metric in Eq.(3), becomes

$$\begin{aligned} 0 = {} & 2\kappa r \partial_r^2 \phi + 2\partial_r \partial_v \phi + \partial_x^2 \phi + \partial_y^2 \phi + 2\kappa \partial_r \phi \\ & - 4rh(\partial^A f)\partial_r \partial_A \phi - 2\kappa^{-1} rh(\partial^A \partial^B f)\partial_A \partial_B \phi - 2h(\partial^A f)\partial_A \phi \\ & - 2\kappa^{-1} rh(\partial^A \partial_B \partial^B f)\partial_A \phi + \kappa^{-1} h(\partial_B \partial^B f)\partial_v \phi + \kappa^{-1} rh'(\partial_B \partial^B f)\partial_r \phi \ . \end{aligned} \tag{9}$$

The solution before the wave can be written as a complete set for positive frequency with respect to $\bar{\xi}$ as

$$\phi^0_{\omega,\mathbf{k}} = \sqrt{\frac{\sinh(\pi\omega/\kappa)}{4\pi^4\kappa}}\, e^{-i\omega v} \left(\frac{2k\sqrt{r}}{\sqrt{2\kappa}}\right)^{i\omega/\kappa} K_{i\omega/\kappa}\left(\frac{2k\sqrt{r}}{\sqrt{2\kappa}}\right) e^{ik_A x^A} \ , \tag{10}$$

where $\omega > 0$ and $\mathbf{k} = (k_x, k_y)$ and the mode functions satisfying the usual Klein Gordon inner product rules. For the $v > v_0$ region, we exploit the fact that the the metric in Eq.(3) is obtained as a diffeomorphism generated by the Rindler supertranslation vector field Ξ to obtain the solution in linear order of perturbation using the supertranslated vector as

$$\begin{aligned} \phi^1_{\omega,\mathbf{k}} &= (1 - \Xi^a \partial_a)\, \phi^0_{\omega,\mathbf{k}} \\ &= \left(1 + i\kappa^{-1}\omega f + i\kappa^{-1} r k_A \partial^A f\right) \phi^0_{\omega,\mathbf{k}} \ . \end{aligned} \tag{11}$$

We can now expand the solution at $v > v_0$ in the basis $\left\{\phi^1_{\omega,\mathbf{k}}\right\}$ to get

$$\widehat{\phi}_{\omega,\mathbf{k}} = \begin{cases} \phi^0_{\omega,\mathbf{k}} & \text{for } v < v_0; \\ \displaystyle\int_0^\infty d\omega' \int d^2\mathbf{k}' \left(\alpha_{\omega,\mathbf{k};\omega',\mathbf{k}'} \phi^1_{\omega',\mathbf{k}'} + \beta_{\omega,\mathbf{k};\omega',\mathbf{k}'} \phi^{1*}_{\omega',\mathbf{k}'}\right) & \text{for } v > v_0, \end{cases} \tag{12}$$

where the αs and βs are the Bogoliubov coefficients between the $\left\{\phi^0_{\omega,\mathbf{k}}\right\}$ basis and the $\left\{\phi^1_{\omega,\mathbf{k}}\right\}$ basis. Imposing the proper the matching conditions at $v = v_0$, one can evaluate the Bogoliubov coefficients in a perturbation series, the linear order form of which is presented in.[10] The explicit form shall not be needed here except for one significant detail that both the Bogoliubov coefficients are non-vanishing. We can then quantise the field in the standard way which would then imply that the Rindler vacuum in the $v < v_0$ region before the wave is not same as the vacuum in the $v > v_0$ region, that is, the presence of the asymmetric supertranslated shock wave changes the background Rindler vacuum. We next address the question how

the non-zero Bogoliubov coefficients affect the entanglement between the modes in the left and right Rindler wedge.

Before the shock wave we consider a two mode entangled state comprising of the left and right Rindler modes given by

$$|\Phi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \ , \tag{13}$$

where

$$|\phi_1\rangle = \frac{1}{\sqrt{1+p^2}} \left(|0\rangle_A \otimes |0\rangle_B + p|1\rangle_A \otimes |1\rangle_B\right) \ , \tag{14a}$$

$$|\phi_2\rangle = \frac{1}{\sqrt{1+q^2}} \left(|0\rangle_C \otimes |0\rangle_D + q|1\rangle_C \otimes |1\rangle_D\right) \ , \tag{14b}$$

and p and q are real-valued parameters. We have labelled the two modes as 1 and 2 and their corresponding left and right mode pairs as (A, B) and (C, D). The choice of such a particular state is motivated from the form of low frequency expansion of the entanglement between left and right Rindler modes in the Minkowski vacuum state and hence is also related to the low energy Hawking spectra in Black holes. We use the entanglement monotone of Negativity[16–23] to quantify the entanglement between different pairs of modes. The initial nonvanishing negativities are

$$\mathcal{N}_{A\leftrightarrow B} = \frac{p}{1+p^2} \ , \tag{15a}$$

$$\mathcal{N}_{C\leftrightarrow D} = \frac{q}{1+q^2} \ . \tag{15b}$$

After the wave, One can compute the Negativity between the different pairs of particles using the Bogoliubov coefficients in terms of their perturbative orders to get

Negativity	**Before wave** $v < v_0$	**After wave** $v > v_0$
A $\leftrightarrow$ B	$\frac{p}{1+p^2}$	$\downarrow \mathrm{O}(\lambda^2)$
B $\leftrightarrow$ D	0	$\uparrow \mathrm{O}(\lambda)$
B $\leftrightarrow$ C	0	$\uparrow \mathrm{O}(\lambda^2)$

The complete expressions are provided in.[10] One can observe from the above table that the negativity between the left and right modes A and B respectively is decreased by $\mathrm{O}(\lambda^2)$ while there is linear order generation of Negativity between the right Rindler modes B and D. We have further checked that this is consistent with the quantum monogamy relation for Negativity.

5. Discussion

We have investigated the gravitational memory effect for the Rindler family of uniformly linearly accelerated observers using the BMS-type symmetries of the Rindler horizon. We described a physical process analogous to that proposed by HPS[3] that

implants supertranslational hair on a Rindler horizon by a matter shock wave without planar symmetry. Before the wave, a family of observers are considered which are orbits of a single boost Killing vector. After the wave has passed, each observer satisfying the constraint of local hyperbolicity in Eq.(4) is found to be an orbit of a different boost Killing vector which differs from trajectory to trajectory. This is the classical memory effect for Rindler observers since the different boosts retains a memory of the planar inhomogeneity of the wave. The Classical memory is also accompanied by a supertranslation quantum memory which is found to modulate the entanglement between the opposing Rindler wedges in quantum field theory. A corresponding phenomenon across the Schwarzschild black hole horizon indicates that the Negativity measure of entanglement between infalling and outgoing Hawking pair should be degraded due to an infalling soft hair implanting shockwave while there should be linear order generation of Negativity between two outgoing Hawking particles Fig. 1.

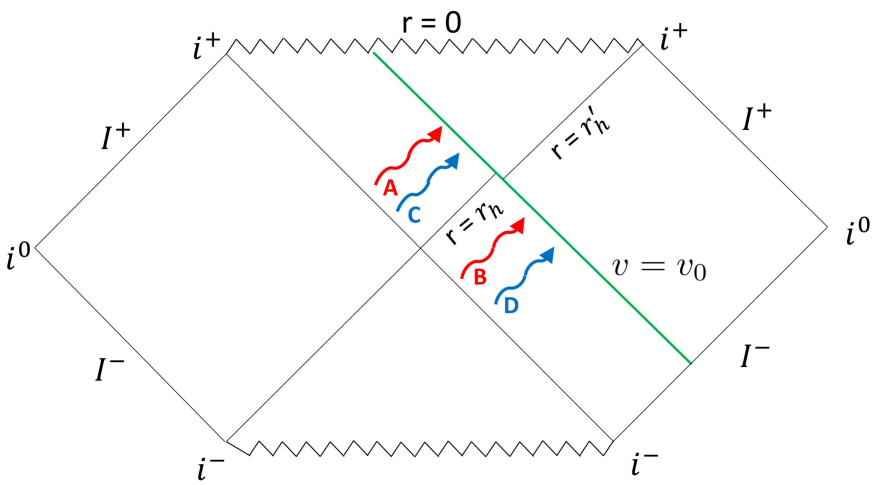


Fig. 1. The extended Schwarzschild spacetime with an infalling, supertranslation-implanting shock wave at $v = v_0$. Two Hawking pairs are shown, $A \leftrightarrow B$ (red) and $C \leftrightarrow D$ (blue).

References

1. H. Bondi, M. G. J. van der Burg and A. Metzner, Proc. Roy. Soc. Lond. **A269**, 21-52 (1962); R. Sachs, Proc. Roy. Soc. Lond. **A270**, 103-126 (1962); R. Sachs, Phys. Rev. **128**, 2851-2864 (1962).
2. S. W. Hawking, M. J. Perry and A. Strominger, Phys. Rev. Lett. **116**, 231301 (2016) [arXiv:1601.00921 [hep-th]].
3. S. W. Hawking, M. J. Perry and A. Strominger, JHEP **1705**, 161 (2017) [arXiv:1611.09175 [hep-th]].
4. A. Strominger, JHEP **1407**, 152 (2014) [arXiv:1312.2229].
5. T. He, V. Lysov, P. Mitra and A. Strominger, [arXiv:1401.7026 hep-th].

6. J. Winicour, Class. Quant. Grav. **31**, 205003 (2014) [arXiv:1407.0259].
7. A. Ashtekar, Surveys in Differential Geometry **20**, 99 (2015) [arXiv:1409.1800].
8. A. Strominger and A. Zhiboedov, JHEP **1601**, 086 (2016) [arXiv:1411.5745].
9. S. Kolekar and J. Louko, Phys. Rev. D **96**, 024054 (2017) [arXiv:1703.10619 [hep-th]].
10. S. Kolekar and J. Louko, Phys. Rev. D **97**, 085012 (2018) [arXiv:1709.07355 [hep-th]].
11. L. Donnay, G. Giribet, H. A. Gonzalez and M. Pino, Phys. Rev. Lett. **116**, 091101 (2016) [arXiv:1511.08687 [hep-th]].
12. L. Donnay, G. Giribet, H. A. Gonzalez and M. Pino, JHEP **1609**, 100 (2016) [arXiv:1607.05703 [hep-th]].
13. C. Eling and Y. Oz, JHEP **1607**, 065 (2016) [arXiv:1605.00183 [hep-th]].
14. R. G. Cai, S. M. Ruan and Y. L. Zhang, JHEP **1609**, 163 (2016) [arXiv:1609.01056 [gr-qc]].
15. M. Hotta, J. Trevison and K. Yamaguchi, Phys. Rev. D **94**, 083001 (2016) [arXiv:1606.02443 [gr-qc]].
16. A. Peres, Phys. Rev. Lett. **77**, 1413 (1996) [arXiv:quant-ph/9604005].
17. M. Horodecki, P. Horodecki and R. Horodecki, Phys. Lett. A **223**, 1 (1996) [arXiv:quant-ph/9605038].
18. G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002) [arXiv:quant-ph/0102117].
19. M. B. Plenio, Phys. Rev. Lett. **95**, 090503 (2005) [arXiv:quant-ph/0505071].
20. K. Audenaert, M. B. Plenio, and J. Eisert, Phys. Rev. Lett. **90**, 027901 (2003) [arXiv:quant-ph/0207146].
21. M. B. Plenio and S. Virmani, Quant. Inf. Comput. **7**, 1 (2007) [arXiv:quant-ph/0702059].
22. R. Simon, Phys. Rev. Lett. **84**, 2726 (2000).
23. G. Adesso and F. Illuminati, J. Phys. A **40**, 7821 (2007) [arXiv:quant-ph/0701221].

Gravitational singularities, scattering maps for bouncing, and structure-preserving algorithms

Philippe G. LeFloch

Laboratoire Jacques-Louis Lions and Centre National de la Recherche Scientifique
Sorbonne Université, 4 Place Jussieu, 75252 Paris, France
Email: contact@philippelefloch.org. Blog: philippelefloch.org

This note emphasizes the role of multi-scale wave structures and junction conditions in many fields of physics, from the dynamics of fluids with non-convex equations of state to the study of gravitational singularities and bouncing cosmologies in general relativity. Concerning the definition and construction of bouncing spacetimes, we review the recent proposal in collaboration with B. Le Floch and G. Veneziano based on the notion of singularity scattering maps. We also present recent numerical investigations of small-scale phenomena arising in compressible fluid flows on FRLW or Kasner geometries for which we developed structure-preserving algorithms.

Keywords: Gravitational singularity; scattering map; small-scale phenomenon; structure-preserving algorithm.

1. Introduction

Recent advances on the mathematical modeling of complex fluid flows containing nonlinear waves and gravitational singularities lead to new challenges for numerical relativity. New ideas and techniques to analyze and compute the asymptotic behavior of multi-physics and multi-scale nonlinear waves have been introduced in the context of classical fluid dynamics but is relevant also for the dynamics of self-gravitating fluids. In this short note, we show some numerical results based on structure-preserving numerical algorithms for a few model problems describing multi-scale nonlinear waves on a fixed background. We also present a recent proposal for defining bouncing cosmologies, which is based on the notion of universal scattering maps for gravitational singularities.

This presentation uses material from joint collaborations, especially with B. Le Floch and G. Veneziano (scattering maps) and with F. Beyer (Fuchsian algorithm). It is based on a lecture entitled *"On the scattering laws of bouncing universes"* given at the *Sixteenth Marcel Grossmann Meeting* in July 2021. It is also based on a lecture entitled *"Gravitational singularities, massive fields, and asymptotic localization"* given at the Workshop *"Computational Challenges in Multi-Messenger Astrophysics",* held at IPAM, University of California at Los Angeles, in October 2021.

Only few references are included in this short note and, for our motivations to study bouncing cosmologies and relativistic fluids with complex equation of state,

we refer to the review papers by Ashtekar,[2] Blaschke,[7] Font,[11] as well as the recent papers by Ibánez et al.[10] and Wilson-Ewing.[23]

An outline is as follows. In Section 2 our standpoint concerning problems involving viscosity-capillarity waves and other multi-scale interfaces is discussed. In Section 3, numerical observations obtained with structure-preserving algorithms are overviewed for fluid problems in the vicinity of gravitational singularities. In Section 4, the scattering laws for gravitational singularities are presented and we distinguish between their universal and model-dependent properties.

2. Structure of multi-physics and multi-scale waves

2.1. *Multi-scale wave phenomena*

Systems of balance laws. The models of interest are formulated from first principles of continuum physics (that is, conservation laws or more generally balance laws) when suitable physics features are taken into account.

- Several parameters are often relevant in order to fully describe the fine dynamics of the fluid flows of interest: viscosity, surface tension, heat conduction, Hall effect, friction, etc. Various competitive effects take place, due to the presence of several very different scales in, both, the fluid unknowns and the geometry unknowns of the problem.
- These nonlinear systems of partial differential equations, in the course of the time evolution, exhibit a formation of interfaces, shocks, oscillating patterns, etc. Fluids, gases, plasmas, solid materials, etc. all exhibit fine structures that form dynamically. This is the case of liquid-vapor flows, thin liquid films, combustion wave problems, bores in shallow water, astrophysical flows, neutron stars, phase transformations (austenite-martensite), etc.
- Very often, a *fine-scale structure* (oscillations, turbulence) arises asymptotically (for instance in the vicinity of a singularity hypersurface) which is challenging to analyze mathematically and to compute numerically.

Asymptotic analysis. The ideas and techniques developed in the context of fluid dynamics are relevant for the modeling of the global dynamics of massive fields (Klein-Gordon), complex fluids, as well as models of modified gravity beyond Einstein gravity. In addition to the shock waves observed in fluid dynamics one must also encompass impulsive gravitational waves and cosmological singularities.

In order to understand the global dynamics of these *scale-sensitive nonlinear waves,* a novel perspective is required which is based on an asymptotic analysis of the physical models.

- We restrict attention to the regime where we can extract variables with well-defined limits, even though persistent oscillations in some other variables may be observed as the singularity is approached.

- We seek junction laws or scattering laws, which in the case of fluid take the form of jump laws for (subsonic) liquid-gas boundaries, for combustion waves, etc.
- The interfaces that require additional junction laws are typically *undercompressive waves* and turn out to enjoy (saddle) stability properties.

2.2. *Regime of small viscosity and capillarity*

Non-convex equation of state. One of the simplest partial differential equation that exhibits small-scale dependent interfaces is the following model in plane-symmetry formulated as a hyperbolic conservation law with non-convex equation of state $p(\rho) = \rho^3$

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x}\rho^3 = \varepsilon\,\frac{\partial^2}{\partial x^2}\rho + \kappa\,\frac{\partial^3}{\partial x^3}\rho, \tag{1}$$

for a fluid density unknown $\rho = \rho(t,x)$ (with $t \geq 0$ and $x \in \mathbb{R}$, say). This equation describes certain phase transition phenomena, but also arises as a simplified model of magnetohydrodynamics. Here, a small viscosity coefficient $\varepsilon > 0$ and a surface tension/capillarity coefficient $\kappa > 0$ are given, the latter representing the intermolecular forces between a liquid and its surroundings.

The Riemann problem consists of an initial value problem associated with a single initial discontinuity (also called the dam breaking problem). Interestingly, in the limit of arbitrarily small $\varepsilon, \kappa \to 0$, the equation (1) exhibits complex wave patterns, which are more involved than the ones usually observed with (polytropic perfect, say) compressible fluids. See the illustration in Figure 1. Many scalar wave models exhibit the same features, for instance the so-called thin liquid film model and the so-called Benjamin–Bona–Mahony model. Let us for instance give here a class of fourth-order models (with $\varepsilon, \kappa, \lambda > 0$):

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x}p(\rho) = \varepsilon\,\frac{\partial^2}{\partial x^2}\rho + \kappa\,\frac{\partial^3}{\partial x^3}\rho - \lambda\,\frac{\partial^4}{\partial x^4}\rho. \tag{2}$$

Three possible asymptotic regimes. In the limit $\varepsilon \to 0$ in (1), we should distinguish between three behaviors depending upon the capillarity parameter κ:

- $\kappa << \varepsilon^2$: the viscosity is then dominant, and no oscillations are observed while a single limit is reached. The solutions (sometimes referred to as classical weak solutions) are characterized by standard Rankine-Hugoniot relations and entropy criteria.
- $\kappa = \alpha\,\varepsilon^2$: this is the *balanced regime* of main interest in the present discussion. The numerical solutions converge in a strong sense, despite some mild oscillations in the limit. (See again Figure 1.) Interestingly, the limit (now sometimes referred to as a nonclassical weak solution) *depends upon the ratio* $\alpha = \kappa/\varepsilon^2$ of the viscosity and capillarity parameters. (See also Figure 3 for van der Waals fluids.)

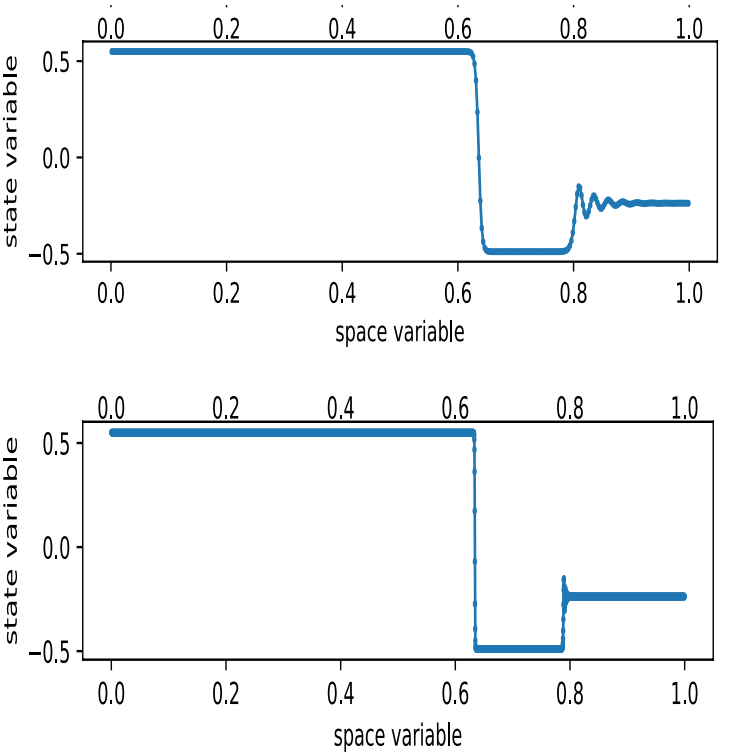


Fig. 1. The Riemann problem for a non-convex equation of state.

- $\kappa >> \varepsilon^2$: in this regime, the effect of the surface tension is dominant and wild oscillations arise as $\varepsilon \to 0$. There is no well-defined limit, unless weak convergence techniques or a suitable turbulence-like theory are introduced.

Wave structure of van der Waals fluids. Let us provide a further illustration with the system of isothermal compressible fluids which is closest to the interest of the GR community. We consider a Van der Waals fluid, in a non-relativistic setup, whose dynamics is modeled by the following two conservation laws (continuity equation, conservation of momentum) :

$$\frac{\partial}{\partial t}\tau - \frac{\partial}{\partial x}u = 0, \qquad \frac{\partial}{\partial t}u + \frac{\partial}{\partial x}p(v) = \varepsilon\,\frac{\partial^2}{\partial x^2}u - \kappa\,\frac{\partial^3}{\partial x^3}\tau, \tag{3}$$

in which $\tau = 1/\rho$ represents the specific volume and u denotes the velocity component for plane-symmetric solutions to the Euler equations. The viscosity parameter ε and the capillarity parameter κ are taken to be small. We assume that the temperature T remains constant (in a suitable approximation) and we model liquid-vapor phase transitions by the non-convex pressure law (after normalization)

$$p(\tau) = \frac{8T}{3\tau - 1} - \frac{3}{\tau^2}. \tag{4}$$

The equations (3)-(4) are nonlinear hyperbolic when T is sufficiently large, but of mixed (hyperbolic-elliptic) type otherwise. The typical Riemann wave structure for van der Waal fluids is shown in Figure 2.

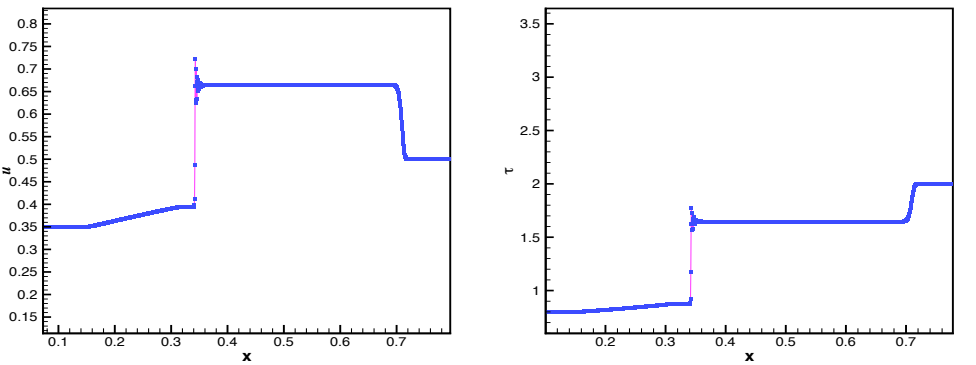


Fig. 2. Van der Waals fluids: velocity (left-hand) and specific volume (right-hand).

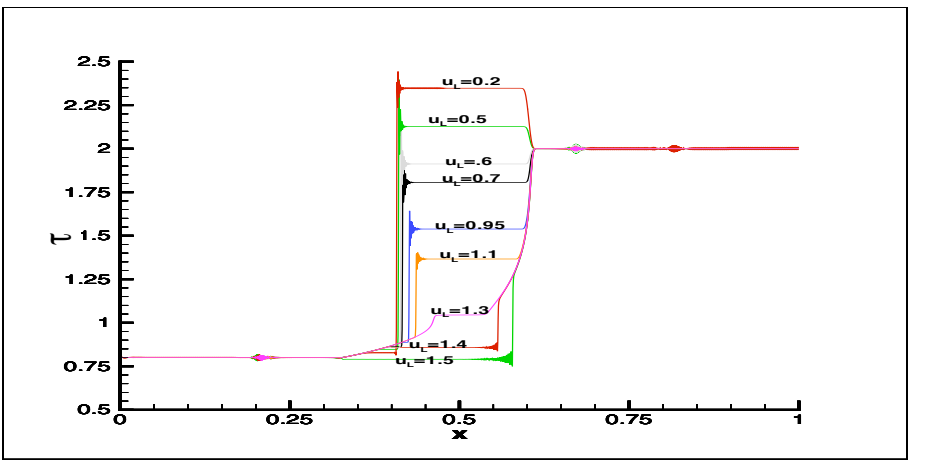


Fig. 3. Van der Waals fluids: specific volume for different ratios of the viscosity and capillarity.

Proposed standpoint: hidden junction conditions. Small-scale dependent, shock waves enjoy a structure that is not properly captured by the standard Rankine-Hugoniot junction conditions of fluid dynamics, even when a suitable entropy condition is enforced. Further analytical work is required which shows that

the wave structure strongly depends upon varying the ratio surface tension/viscosity $\kappa = \alpha\,\varepsilon^2$ (in the balanced regime). (See again Figure 3 for van der Waals fluids.) Suitable rules for connecting left- and right-hand state values from both sides of the interface is necessary, beyond the standard Rankine-Hugoniot relations. These *scattering laws* take the form, in this context, of *kinetic functions for interfaces*. It is outside the scope of this short Note to explain this notion here, and we refer the reader to the textbook[13] and the lecture notes[12,14] and the historical references therein. Importantly, based on these scattering maps we need to design structure-preserving algorithms, which are typically of front tracking type, or of shock capturing type with well-controlled dissipation.[15] The algorithms are designed in order to *mimic analytical properties at the discrete level* and this requirement may take very different flavors: divergence form, spacelike decay, timelike decay shock-capturing, energy balance laws, asymptotics on singularities, etc.

3. Structure-preserving algorithms for fluids near gravitational singularities

3.1. *Preserving the asymptotic structure on a Kasner background: evolution from the singularity*

Evolution in the vicinity of a gravitational singularity. In a joint work with F. Beyer,[6] I have studied the evolution of a compressible fluid on a Kasner geometry, namely on a spatially homogeneous, anisotropic vacuum background described by the metric

$$g = t^{(K^2-)/2}\big(-dt^2 + dx^2 + t^{1-K}dy^2 + t^{1+K}dz^2\big), \qquad M = (0,+\infty)\times\mathbb{T}^3 \tag{5}$$

Here, we set $t \in (0,\infty)$ and $x, y, z \in (0, 2\pi)$ and we introduce the parameter $K \in \mathbb{R}$ and the Kasner exponents

$$p_1 = \frac{K^2-1}{K^2+3}, \quad p_2 = \frac{2(1-K)}{K^2+3}, \quad p_3 = \frac{2(1+K)}{K^2+3}. \tag{6}$$

The free parameter $K \in \mathbb{R}$ is sometimes referred to as the *asymptotic velocity*. Except for the three flat Kasner cases given by $K = 1$, $K = -1$, and (formally) $|K| \to \infty$, the Kasner metrics g have a curvature singularity in the limit $t \searrow 0$. We treat a compressible fluid flow with pressure law

$$p = (\gamma-1)\rho \quad \text{ with } \gamma \in (1,2). \tag{7}$$

Fuchsian analysis for the Euler equations of compressible fluids. In[6] we discovered a *characteristic fluid exponent* defined as

$$\Gamma = \frac{1}{4}\big(3\gamma - 2 - K^2(2-\gamma)\big) \in (0,1), \tag{8}$$

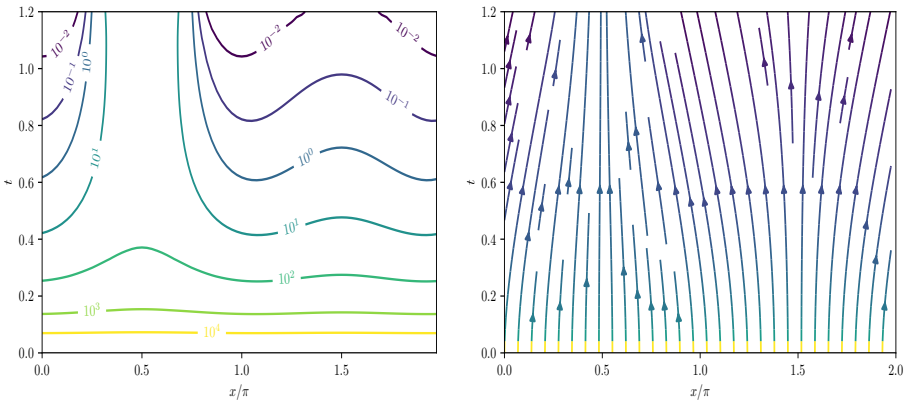


Fig. 4. Evolution of a compressible fluid **from** a gravitational singularity: Kasner background.

which compares the geometry and fluid behaviors:

- $\Gamma > 0$: this is a sub-critical regime which is dynamically stable.
- $\Gamma \leq 0$: this is a super-critical (or critical) regime which is dynamically unstable.

This (positivity/negativity) condition arises by plugging an expansion in power of t for the fluid unknown variables of the Euler equations, and attempting to validate this expansion in t via the Fuchsian method.

The Fuschian method concerns the evolution problem *from* the cosmological singularity $t = 0$ (rather than *towards* it). We formulate a singular initial value problem of the form

$$B^0(U,t,x)\frac{\partial}{\partial t}U + B^1(U,t,x)\frac{\partial}{\partial x}U = f(U,t,x) \tag{9}$$

with suitable "singular initial data" prescribed at $t = 0$. Fuchsian-type expansions are then derived from an ODE approximation near the cosmological singularity. In the present work, we focus on the sufficiently regular, shock-free regime for the Euler equations.

Algorithm preserving the Fuchsian structure. We approximate the singular Cauchy problem of Fuchsian type by a sequence of regular Cauchy problems, which we next discretize by the pseudo-spectral method of lines, say $U(t,x) \simeq V(t) = (V_j(t))$ with

$$\partial_t V - AV = h(V,t). \tag{10}$$

A high-order Runge-Kutta discretization is used in time. Importantly, we introduce *suitably rescaled variables* and next careful study of the numerical error. To this

end, we take into account the Fuchsian expansion available near the singularity. We observe that there are two sources of approximation error, namely continuum and discrete.

Our proposal in order to get an efficient algorithm is to *keep the two error sources asymptotically in balance*. With this numerical strategy, we can demonstrate the *nonlinear stability of the flow* near the cosmological singularity in the sub-critical regime. See Figure 4 for a numerical simulation on a Kasner background for a typical evolution from the singularity. The fluid density is shown in contour plot while the normalized velocity field is shown in flow lines.

The density ρ unbounded as the time $t \to 0$, and we carefully checked the numerical error in order to produce quantitative error control. In turn we arrive at reliable and accurate algorithm, despite the solutions being *highly singular*.

3.2. *Preserving the asymptotic structure on an inhomogeneous FLRW background*

Evolution of fluid with shock waves. In a joint work[8,9] with Y. Cao and M. Ghazizadeh, I have considered the evolution of a compressible fluid *toward* a cosmological singularity. Therein, we consider the Cauchy problem i isothermal, relativistic compressible flow with linear pressure law $p(\rho) = k^2\rho$. The background geometry is taken to be a contracting (or expanding) inhomogeneous background with torus topology $\mathbb{T}^2$, corresponding to a FLRW-type cosmological background with small inhomogeneities. For isothermal fluids, the Euler equations take the form of a system of two nonlinear hyperbolic balance laws, which we treat in one and in two space dimensions under symmetry assumptions. Let us focus here on the future-contracting direction ($t < 0$ with $t \to 0$) and we observe that the mass energy density $\rho \to +\infty$. The expanding direction is also treated in.[8,9] The challenge is to design a scheme that preserves the structure of the solutions as the singularity is approached and the unknown variables become singular.

We analyze the asymptotic behavior of the fluid variables toward the cosmological singularity, by rewriting first the equations as a nonlinear hyperbolic systems on a curved geometry, say

$$\partial_t U + \partial_x F(t, x, U) = H(t, x, U). \tag{11}$$

Two *competitive effects* take place in this problem which concern the contracting geometry and the shock propagation. The nonlinear wave interactions give rise to a small-scale structure, which in the present setup is simply driven by the background geometry. Yet, we can keep in mind the analogy with phase transition dynamics which involves multiple scales as explained in Section 2. Extensive numerical experiments were performed in one and in two space dimensions are performed in order to investigate the fine behavior of the solutions in the vicinity of the singularity.

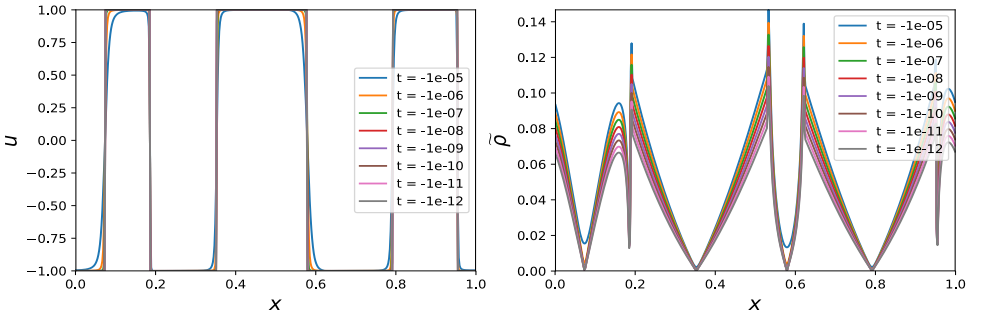


Fig. 5. Evolution of a compressible fluid **toward** a gravitational singularity. Left-hand: velocity scalar. Right-hand: rescaled mass energy density.

Structure-preserving methodology. We formulate a shock-capturing scheme in divergence form, based on the finite volume methodology. The shock-capturing property ensures that the shock waves will propagate with the correct speed given by the Rankine-Hugoniot condition. We require high accuracy and use a discretization which is 4th-order in time and 2nd-order in space, hence ensure that the numerical solution is oscillation-free. In order to guarantee the *well-balanced property* we proceed as follows:

- We introduce suitably *rescaled unknowns* which are directly motivated by a Fuchsian analysis of the problem.
- The discrete form of the balance laws is designed in order to enforce the correct *asymptotic state equations at the discrete level*.
- In turn , we enforce the *commutation property*

$$\lim_{t\to 0}\lim_{\Delta x\to 0} U = \lim_{\Delta x\to 0}\lim_{t\to 0} U \tag{12}$$

relating the limits of the numerical solution $U_{\Delta x}(t,x)$

The typical behavior that we observed *toward* the cosmological singularity are sharp transitions with spikes, and these spikes turn out to be well capured by our numerical method in the asymptotic limit $t \to 0$. See the plots of the rescaled velocity component u and rescaled density $\widetilde{\rho}$. See Figure 5 for the evolution of a compressible fluid **toward** a gravitational singularity.

4. On the scattering laws of bouncing universes

4.1. *Seeking a formulation of bouncing laws*

Bouncing cosmologies. In a joint work in collaboration with B. Le Floch (Paris) and G. Veneziano (Geneva) I considered gravitational singularities in the context of the Einstein equations coupled to a self-gravitating scalar field. More generally, our ideas apply to a stiff fluid or a compressible fluid. We are interested in analyzing

bouncing cosmologies involving a contracting phase and an expanding phase of evolution, connected across a singularity hypersurface. Our analysis takes place in the vicinity of such a singularity hypersurface, and we focus on the jump relations that should hold at that interface. Recall that in presence of strong fields or away from the near-Minkowski regime of evolution, Penrose and Hawking's singularity theorems (or incompleteness theorems) suggest that gravitational singularities (or other incompleteness phenomena) are expected to occur, although very little is known on the actual asymptotic behavior beyond the classical BKL conjecture.

Physics literature on bouncing. We build upon a large literature on bouncing cosmologies by Penrose, Tod, Lübbe, Turok, Barrow, and many others. We emphasize that existing proposals are based on the analysis of symmetric spacetimes and/or the formulation of special junctions. In order to define the bouncing, a large number of proposals are found in the literature, including the pre-Big Bang scenario (Gasperini, Veneziano, etc.), a variety of models of modified gravity-matter (Brandenberger, Chamseddine, Cotsakis, Mukhanov, Peter, Steinhardt, Turok, etc.), and the theory of loop quantum cosmology (Asthekar, de Cesare, Gupt, Pawlowski, Singh, Wilson-Ewing, etc.).

Proposed standpoint. Our aim is to propose a formulation of the problem of junction at singularities and to perform a systematic investigation of such junction conditions. This is motivated by the ideas discussed in Section 2. We expect to derive classes of physically meaningful junction conditions which depend upon certain (limited) degrees of freedom and constraints at the singularity. Interestingly, based on the scattering maps we propose we can also define and construct classes of cyclic spacetimes. Here, we only outline a selection of our results and refer the reader to[18–20] for the complete definitions and further material.

We study the class of (past, future) *singularity data* denoted by $(g^{\pm}, K^{\pm}, \phi_0^{\pm}, \phi_1^{\pm})$ and defined after a suitable rescaling (see below). We then introduce a *singularity scattering map* $(g^-, K^-, \phi_0^-, \phi_1^-) \mapsto (g^+, K^+, \phi_0^+, \phi_1^+)$ in order to cross the singularity hypersurface. Our main contribution is a full classification of the possible bouncing conditions. We are able to distinguish between *universal* scattering laws and *model-dependent* scattering maps.

We thus provide a flexible framework and a classification which uncovers all possible classes of junctions that are geometrically and physically meaningful. We also exhibit several sub-classes of particular interest, namely the conformal/non-conformal maps, the spacelike/null/timelike maps, and we encompass different models of matter (scalar field, stiff fluid, compressible fluid). In the course of establishing our classification, we also discover three universal laws that put constrains on the macroscopic aspects of spacetime bounces, *regardless of their origin from different microscopic corrections*. Our classification provides one with a guide in order to uncover relevant structures, described by scattering maps associated with specific theories.

4.2. *Proposed formulation of the bouncing problem*

The field equations. We consider a local ADM formulation near a singularity hypersurface, based on a Gaussian foliation (local patch) $\mathcal{M}^{(4)} = \bigcup_{\tau\in[\tau_{-1},\tau_1]} \mathcal{H}_\tau$ by spacelike hypersurfaces (diffeomorphic to $\mathcal{H}_0$). We write

$$g^{(4)} = \big(g^{(4)}_{\alpha\beta}\big) = -d\tau^2 + g(\tau) \qquad g(\tau) = g_{ij}(\tau)dx^i dx^j, \tag{13}$$

and we impose Einstein's evolution equations for the unknown metric g and the unknown extrinsic curvature K

$$\partial_\tau g_{ij} = -2\,K_{ij} \qquad \partial_\tau K^i_j = \mathbf{Tr}(K)K^i_j + R^i_j - 8\pi\, M^i_j. \tag{14}$$

Here, $M^i_j = \frac{1}{2}\rho g^i_j + T^i_j - \frac{1}{2}\mathbf{Tr}(T)g^i_j$ is the matter contribution. These equations are supplemented with Einstein's constraint equations

$$R + |K|^2 - \mathbf{Tr}(K^2) = 16\pi\rho \qquad \nabla_i K^i_j - \nabla_j(\mathbf{Tr}K) = 8\pi J_j, \tag{15}$$

which are nonlinear elliptic equations. Finally, we assume that the energy-momentum tensor is given by a massless scalar field ϕ whose evolution is given by the wave equation

$$\Box_{g^{(4)}}\phi = 0. \tag{16}$$

Near a singularity hypersurface, it is standard to rely on the Fuchsian method (Baouendi, Goulaouic, Rendall, Isenberg, Moncrief, etc.):

- We solve locally from $\tau = 0$ toward the past ($\tau < 0$) or the future ($\tau > 0$).
- We apply the so-called "velocity dominated" Ansatz, which tells us that in the gauge under consideration all spatial derivatives (except for the momentum equations) can be neglected.
- In turn, we must solve a system of nonlinear coupled differential equations in the time variable with nonlinear source-terms (containing spatial derivatives) and by a suitable iteration scheme we generate solutions to the full Einstein equations.

Singularity data and asymptotic profiles. A 3-manifold $\mathcal{H}$ is fixed throughout and represents the singularity hypersurface.

1. An **asymptotic profile** associated with data $(g^-, K^-, \phi_0^-, \phi_1^-) \in \mathbf{I}(\mathcal{H})$ is the flow on $\mathcal{H}$
$$\begin{aligned} \tau \in (-\infty, 0) \mapsto \big(g^*, K^*, \phi^*\big)(\tau), \qquad & g^*(\tau) = |\tau|^{2K^-} g^-, \\ K^*(\tau) = \frac{-1}{\tau} K^-, \qquad & \phi^*(\tau) = \phi_0^- \log|\tau| + \phi_1^-, \end{aligned} \tag{17}$$
in which $|\tau|^{2K^-}$ is defined by exponentiation.
2. A **singularity initial data set** $(g^-, K^-, \phi_0^-, \phi_1^-)$, consists of two tensor fields (rescaled metric and extrinsic curvature) and two scalar fields defined on $\mathcal{H}$.

3. The **asymptotic version of the Einstein constraints** reads as follows:

$$\begin{array}{ll} \text{Riemannian metric} & g^- = (g^-_{ij}) \text{ on } \mathcal{H}. \\ \text{CMC symmetric } (1,1)\text{-tensor} & K^- = (K_i^{-j}) \text{ with } \mathbf{Tr}(K^-) = 1 \text{ on } \mathcal{H}. \\ \text{Hamiltonian constraint} & 1 - |K^-|^2 = 8\pi\,(\phi_0^-)^2 \text{ on } \mathcal{H}. \\ \text{momentum constraints} & \mathbf{Div}_{g^-}(K^-) = 8\pi\,\phi_0^-\, d\phi_1^- \text{ on } \mathcal{H}. \end{array} \tag{18}$$

We denote by $\mathbf{I}(\mathcal{H})$ the set of space of all singularity data.

Junction conditions. The notion of a *scattering map* arises naturally from earlier considerations about fluid interfaces. We think of a singularity hypersurface as a fluid-like interface between two "phases", across which the geometry and the matter field encounter a "jump". Small-scale physics phenomena are not directly modeled at this stage. We are only interested in the "average" effect rather than the detailed physics that may take place "within" this interface. This is a standard strategy in fluid dynamics and material science in presence of phase transition phenomena, especially when some (micro-scale) parameters (viscosity, surface tension, heat conduction, etc.,) can be neglected in the modeling. Macro-scale effects are captured by jump conditions such as Rankine-Hugoniot, kinetic relations, Dal Maso-LeFloch-Murat's paths, etc. For instance, kinetic relations in material science (martensite-austenite) and in two-phase liquid-vapor flows have been extensively studied. In this context, hypersurfaces are timelike; in the present discussion of gravitational singularities we focus our attention on spacelike hypersurfaces and refer the reader to[19] for the treatment of timelike singularity hypersurfaces.

Bounce based on a singularity scattering map. A (past-to-future) **singularity scattering map** on $\mathcal{H}$ is defined as a diffeomorphism-covariant map on $\mathbf{I}(\mathcal{H})$

$$\mathbf{S} : \mathbf{I}(\mathcal{H}) \ni \left(g^-, K^-, \phi_0^-, \phi_1^-\right) \mapsto \left(g^+, K^+, \phi_0^+, \phi_1^+\right) \in \mathbf{I}(\mathcal{H}) \tag{19}$$

satisfying the *ultra-locality property*: for all $x \in \mathcal{H}$

$$\mathbf{S}(g^-, K^-, \phi_0^-, \phi_1^-)(x) \text{ depends only on } (g^-, K^-, \phi_0^-, \phi_1^-)(x) \tag{20}$$

Thanks to the locality condition, it is natural to identify singularity scattering maps $\mathbf{S}$ on all 3-manifolds and suppress the dependence on $\mathcal{H}$.

The map $\mathbf{S}$ is said to be a **tame-preserving map** if it preserves positivity, in the sense that

$$\text{if } K^- > 0 \text{ then } K^+ > 0, \text{ where } K^+ \text{ is defined from the image of } \mathbf{S}. \tag{21}$$

It is said to be a **rigidly-conformal map** if

$$g^+ \text{ and } g^- \text{ only differ by a conformal factor.} \tag{22}$$

Observations. We emphasize that the asymptotic profiles with $K^-, K^+ > 0$ describe a "bounce" at which

- volume element decreases to zero as $\tau \to 0^-$, and
- then increases back to finite values for $\tau > 0$.

For further notions and constructions of *cyclic spacetimes* containing many singularity hypersurfaces, we refer to[19] and, in this short Note, we focus on the issue of defining a proper junction at the bouncing.

The regime of quiescent singularities $K > 0$ is motivated by the absence of BKL oscillations in this case (named after Belinsky, Khalatnikov, and Lifshitz). This quiescent regime exhibits a monotone behavior which has received a lot of attention by mathematicians in recent years (Rendall, Andersson, Lott, Fournodavlos, Luk, Rodnianksi, Speck, etc.).

4.3. *Classification of bouncing laws*

Rigidly conformal maps. We discover that *only two classes* of ultra-local spacelike and rigidly conformal, singularity scattering maps exist for self-gravitating scalar fields, namely:

- The maps describing **isotropic rigidly conformal bounces** $\mathbf{S}^{\text{iso, conf}}_{\lambda,\varphi}$:

$$g^+ = \lambda^2 g^-, \qquad K^+ = \delta/3, \qquad \phi_0^+ = 1/\sqrt{12\pi}, \qquad \phi_1^+ = \varphi, \tag{23}$$

  parametrized by a conformal factor $\lambda = \lambda(\phi_0^-, \phi_1^-, \det K^-) > 0$ and a constant φ.
- The maps describing **non-isotropic rigidly conformal bounces** $\mathbf{S}^{\text{ani, conf}}_{f,c}$:

$$\begin{aligned} g^+ &= c^2\mu^2 g^-, & K^+ &= \mu^{-3}(K^- - \delta/3) + \delta/3, \\ \phi_0^+ &= \mu^{-3}\phi_0^-/F'(\phi_1^-), & \phi_1^+ &= F(\phi_1^-), \end{aligned} \tag{24}$$

  parametrized by a constant $c > 0$ and a function $f\colon \mathbb{R} \to [0, +\infty)$:

$$\mu(\phi_0, \phi_1) = \big(1 + 12\pi(\phi_0)^2 f(\phi_1)\big)^{1/6}, \; F(\phi_1) = \int_0^{\phi_1} (1 + f(\varphi))^{-1/2} d\varphi. \tag{25}$$

General classification. *Only two classes* of ultra-local spacelike singularity scattering maps exist for self-gravitating scalar fields, which we denote by $\mathbf{S}^{\text{iso}}_{\lambda,\varphi}$ (isotropic bounce) and $\mathbf{S}^{\text{ani}}_{\Phi,c}$ (non-isotropic bounce). Now, λ is a two-tensor, Φ a "canonical transformation", c a constant. For the detailed expressions we refer the reader to.[19]

Three universal laws of quiescent bouncing cosmology. Interestingly our conclusions can be restated in terms of universal laws obeyed by any ultra-local bounce, as follows.

- **First law. Scaling of Kasner exponents**:

$$\begin{aligned} &\text{There exists a (dissipation) constant } \gamma \in \mathbb{R} \text{ such that} \\ &|g^+|^{1/2}\mathring{K}^+ = -\gamma\,|g^-|^{1/2}\mathring{K}^- \end{aligned} \tag{26}$$

  for the spatial metric g in synchronous gauge with volume factor $|g|^{1/2}$, where $\mathring{K}$ denotes the traceless part of the extrinsic curvature (as a $(1,1)$ tensor).
- **Second law. Canonical transformation of matter:** by denoting the conjugate matter momentum by $\pi_\phi \sim \phi_0$,

$$\begin{aligned} &\text{there exists a nonlinear map } \Phi\colon (\pi_\phi, \phi)^- \mapsto (\pi_\phi, \phi)^+ \\ &\text{preserving the volume form in the phase space } d\pi_\phi \wedge d\phi \\ &\text{and depending solely on the scalar invariant } \mathbf{det}(\mathring{K}_-). \end{aligned} \tag{27}$$

- **Third law. Directional metric scaling:**

$$\begin{aligned} &g^+ = \exp\bigl(\sigma_0 + \sigma_1 K + \sigma_2 K^2\bigr) g^-, \\ &\text{which is a nonlinear scaling in each proper direction of } K. \end{aligned} \tag{28}$$

  We have either $\gamma = 0$ in the isotropic scattering and no restriction $\sigma_0, \sigma_1, \sigma_2$, or else $\gamma \neq 0$ for non-isotropic scattering and explicit formulas are available in terms of Φ, γ.

4.4. *Ongoing work*

Our next task will be the derivation of scattering maps associated with specific theories taking small-scale physical modeling into account. Mathematical investigations as well as numerical investigations are required. We emphasize that our methodology is based on a class of *general* spacetimes without symmetry restrictions. Earlier approaches relied on classes of symmetric spacetimes or on special junction proposals, and therefore could not single out the three universal laws we have derived. The recent developments overviewed in the present Note suggest to perform numerical simulations of cyclic spacetimes with a bounce using some of the techniques discussed in Sections 2 and 3. We refer the reader to[18–20] and to.[17]

References

1. P. Amorim, C. Bernardi, and P.G. LeFloch, Computing Gowdy spacetimes via spectral evolution in future and past directions, *Class. Quantum Grav.* 26 (2009), 1–18.
2. A. Ashtekar, Singularity resolution in loop quantum cosmology: a brief overview, *J. Phys. Conf. Ser.* 189 (2009), 012003.
3. F. Beyer and P.G. LeFloch, Second–order hyperbolic Fuchsian systems and applications, *Class. Quantum Grav.* 27 (2010), 245012–245033.
4. F. Beyer and P.G. LeFloch, Second–order hyperbolic Fuchsian systems. Asymptotic behavior of geodesics in Gowdy spacetimes, *Phys. Rev. D.* 84 (2011), 084036.

5. F. Beyer and P.G. LeFloch, Dynamics of self–gravitating fluids in Gowdy-symmetric spacetimes near cosmological singularities, *Comm. Part. Diff. Equa.* 42 (2017), 1199–1248.
6. F. Beyer and P.G. LeFloch, A numerical algorithm for Fuchsian equations and fluid flows on cosmological spacetimes, *J. Comput. Phys.* 431 (2021), 110145.
7. D. Blaschke, D.E. Alvarez-Castillo, A. Ayriyan, H. Grigorian, N.K. Lagarni, and F. Weber, Astrophysical aspects of general relativistic mass twin stars, Chapter 1, Topics on Strong Gravity, 2019.
8. Y.-Y. Cao, M.A. Ghazizadeh, and P.G. LeFloch, Asymptotic structure of cosmological Burgers flows in one and two space dimensions: A numerical study, *Communications in Computational Physics* 29 (2021), 472–509.
9. Y.-Y. Cao, M.A. Ghazizadeh, and P.G. LeFloch, Asymptotic structure of cosmological fluid flows in one and two space dimensions: A numerical study, *Commun. Appl. Math. Comput. Sci.* (2021), submitted. See also ArXiv:1912.13439.
10. J.M. Ibánez, A. Marquina, S. Serna, and M.A. Aloy, Anomalous dynamics triggered by a non-convex equation of state in relativistic flows, *MNRAS* 000 (2018), 1–22.
11. J.A. Font, Numerical hydrodynamics and magnetohydrodynamics in general relativity, *Living Reviews in Relativity* 11 (2008).
12. P.G. LeFloch, An introduction to nonclassical shocks of systems of conservation laws, *Lect. Notes Comput. Eng.,* Vol. 5, Springer Verlag, 1999, pp. 28–72.
13. P.G. LeFloch, *Hyperbolic Systems of Conservation Laws. The theory of classical and nonclassical shock waves, Lectures in Mathematics, ETH Zürich,* Birkhäuser, 2002.
14. P.G. LeFloch, Kinetic relations for undercompressive shock waves. Physical, mathematical, and numerical issues, *Contemporary Mathematics,* vol. 526, Amer. Math. Soc., Providence, RI, 2010, pp. 237–272.
15. P.G. LeFloch, Structure-preserving shock-capturing methods: late-time asymptotics, curved geometry, small-scale dissipation, and nonconservative products, in *"Lecture Notes of the XV Jacques-Louis Lions Spanish-French School"*, Ed. C. Parès, C. Vázquez, and F. Coquel, Springer Verlag, Switzerland, 2014, pp. 179–222.
16. B. Le Floch and P.G. LeFloch, Compensated compactness and corrector stress tensor for the Einstein equations in T2 symmetry, *Portugaliae Math.* 77 (2020), 409–421.
17. B. Le Floch and P.G. LeFloch, On the global evolution of self-gravitating matter. Scattering maps for interfaces, in preparation.
18. B. Le Floch, P.G. LeFloch, and G. Veneziano, Universal scattering laws for quiescent bouncing cosmology, *Physical Rev. D.* 8 (2021), 083531. See also ArXiv:2006.08620.
19. B. Le Floch, P.G. LeFloch, and G. Veneziano, Cyclic spacetimes through singularity scattering maps. The laws of bouncing cosmology. Preprint ArXiv:2005.11324.
20. B. Le Floch, P.G. LeFloch, and G. Veneziano, Cyclic spacetimes through singularity scattering maps. Plane-symmetric gravitational collisions. Preprint ArXiv:2106.09666.
21. M. Mars and J.M. Senovilla, Geometry of general hypersurfaces in spacetime: Junction conditions, *Class. Quantum Grav.* 10 (1993), 1865–1897.
22. A.D. Rendall, *Partial differential equations in general relativity,* Oxford University Press, Oxford, 2008.
23. E. Wilson-Ewing, The loop quantum cosmology bounce as a Kasner transition, *Classical Quant. Grav.* 35 (2018), 065005.

Brane-world asymptotics in a nonlinear fluid bulk

I. Antoniadis

Laboratoire de Physique Théorique et Hautes Energies - LPTHE, Sorbonne Université, CNRS 4 Place Jussieu, 75005 Paris, France, and
Institute for Theoretical Physics, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium
E-mail: antoniad@lpthe.jussieu.fr

S. Cotsakis

Institute of Gravitation and Cosmology, RUDN University ul. Miklukho-Maklaya 6, Moscow 117198, Russia, and
Research Laboratory of Geometry, Dynamical Systems and Cosmology, University of the Aegean, Karlovassi 83200, Samos, Greece
E-mail: skot@aegean.gr

Ifigeneia Klaoudatou

Research Laboratory of Geometry, Dynamical Systems and Cosmology, University of the Aegean, Karlovassi 83200, Samos, Greece
E-mail: iklaoud@aegean.gr

We present recent results on the asymptotics of a brane-world that consists of a flat 3-brane embedded in a five-dimensional bulk. The bulk matter is modelled by a fluid that satisfies a nonlinear equation of state of the form $p = \gamma\rho^{\lambda}$, where p is the 'pressure' and ρ is the 'density' of the fluid. We show that for appropriate ranges of the parameters γ and λ, it is possible to construct a regular solution, compatible with energy conditions, that successfully localizes gravity on the brane. These results improve significantly previous findings of the study of a bulk fluid with a linear equation of state.

Keywords: Brane-worlds, singularities

1. Introduction

In this paper, we review very recent work, presented in,[1] on the asymptotic behaviors of a class of brane-worlds. The models consist of a flat 3-brane embedded in a five-dimensional bulk filled with a fluid that satisfies a nonlinear equation of state $p = \gamma\rho^{\lambda}$, where p is the 'pressure' and ρ is the 'density' of the fluid and γ, λ are constants.

Such an equation of state, has been studied in cosmology for its contribution in avoiding big-rip singularities during late time asymptotics,[2–4] in obtaining inflationary models with special properties,[5] in unifying models of dark energy and dark matter,[6,7] as well as in studies of singularities.[8–10]

Our goal in studying these brane-world models, is to find a solution that is regular and meets physical requirements set by energy conditions and localization of gravity on the brane. In our previous work in,[11–16] which extended work done

in brane-worlds with scalar fields,[17–19] we studied a 3-brane (flat or curved) and a fluid with a linear equation of state $p = \gamma\rho$. We found that such a solution becomes possible only for the special value of $\gamma = -1$ and only for a flat brane. These results gave us the motivation to explore whether the generalization of the linear equation of state to a nonlinear one, could have a decisive impact in finding a solution with all required properties and valid for a wide range of γ and λ. This exploration turned out to be very fruitful, and we were able to find in[1] such solutions for a flat brane and a fluid with $\gamma < 0$ and $\lambda > 1$.

As the analysis of the nonlinear equation of state is quite complicated, for the most part of this paper, we will focus on the study of a specific value of $\lambda = 3/2$, which is simpler for illustration and offers important insight on the behavior of solutions for general $\lambda > 1$.

The structure of this paper is the following: In Section 2, we set up our model and derive the field equations. In Section 3, we formulate the null energy condition which we embody in our solutions later in Section 4. In Section 5, we derive a matching solution for the specific value of $\lambda = 3/2$ and based on this solution, we calculate, in Section 6, the four-dimensional Planck mass. Finally, in Section 7, we discuss briefly the asymptotic behaviors of solutions for general λ.

2. Setup of brane-world

To set up our model, we start by considering a flat 3-brane embedded in a five-dimensional bulk with a metric of the form,

$$g_5 = a^2(Y)g_4 + dY^2, \tag{1}$$

where $a(Y)$ is the warp factor and g_4 is the four-dimensional flat metric, *i.e.*,

$$g_4 = -dt^2 + dx_1^2 + dx_2^2 + dx_3^2. \tag{2}$$

The timelike coordinate is denoted by t and the spacelike ones by (x_1, x_2, x_3, Y). In our notation, Capital Latin indices are taken as $A, B, \cdots = 1, 2, 3, 4, 5$ while lowercase Greek indices range as $\alpha, \beta, \cdots = 1, 2, 3, 4$. The 5-dimensional Riemann tensor is defined by the formula,

$$R^A{}_{BCD} = \partial_C \Gamma^A{}_{BD} - \partial_D \Gamma^A{}_{BC} + \Gamma^M_{BD}\Gamma^A_{MC} - \Gamma^M_{BC}\Gamma^A_{MD} \tag{3}$$

the Ricci tensor is the contraction,

$$R_{AB} = R^C{}_{ACB}, \tag{4}$$

and the five-dimensional Einstein equations on the bulk space are given by,

$$G_{AB} = R_{AB} - \frac{1}{2}g_{AB}R = \kappa_5^2 T_{AB}. \tag{5}$$

Next, we set up the bulk matter component by considering a bulk fluid with an energy-momentum tensor of the form,

$$T_{AB} = (\rho + p)u_A u_B - p g_{AB}, \tag{6}$$

where p is the 'pressure' and ρ is the 'density' which we take as functions only of the extra dimension Y. In (6), the velocity vector field is $u_A = (0, 0, 0, 0, 1)$, that is $u_A = \partial/\partial Y$, parallel to the Y-dimension. We assume that the pressure and density are interconnected by the following nonlinear equation of state

$$p = \gamma \rho^{\lambda}, \tag{7}$$

with γ and λ constants.

The Einstein equations can be written as

$$\frac{a'^2}{a^2} = \frac{\kappa_5^2}{6}\rho, \tag{8}$$

$$\frac{a''}{a} = -\frac{\kappa_5^2}{6}(2\gamma\rho^{\lambda} + \rho), \tag{9}$$

with the prime ($'$) denoting differentiation with respect to Y. The equation of energy-momentum conservation, on the other hand,

$$\nabla_A T^{AB} = 0,$$

becomes

$$\rho' + 4(\gamma\rho^{\lambda} + \rho)\frac{a'}{a} = 0. \tag{10}$$

At this point, it is useful to clarify the terminology that we use in what follows. We use the term *finite-distance singularities* to refer to the following singular behaviors of a, occurring within a finite distance Y_s:

- $a \to 0^+$, as $Y \to Y_s$ (collapse singularity),
- $a \to \infty$, as $Y \to Y_s$ (big-rip singularity).

With Y_s, we denote a finite value of Y designating the position of the singularity. The above behaviors may be accompanied, in general, by a divergence in the density, or, even in the pressure of the fluid. Note that, these singularities are not related to geodesic incompleteness as in standard cosmology, but rather on a pathological behavior of the warp factor. In the absence of finite-distance singularities, we call the solutions regular and include in this category the behaviors of the warp factor given above, provided that these occur *only* at infinite distance, *i.e* $Y \to \pm\infty$.

We are interested to see, if it is possible to find solutions of Eqs. (8)-(10) that rectify previous findings of solutions of brane-worlds with linear bulk fluids studied in.[11–16] As mentioned in the introduction, excepting the case of a flat brane and a fluid with $\gamma = -1$, the main problem that we faced in the study of a linear bulk fluid, is that regular solutions that exist in the case of curved branes, or even singular ones out of which we can construct a regular matching solution for both flat and curved branes, lead to a challenging compromise: they either satisfy energy conditions, or, localize gravity on the brane. Therefore, a solution arising from the consideration of this different and more complicated type of fluid would be preferable and worth

examining in detail, if it has each one of the following characteristics:

- it is regular (no finite-distance singularities)
- it satisfies physical conditions, such as energy conditions
- it leads to a finite four-dimensional Planck mass, thus, it localizes gravity on the brane.

We can simplify the situation of finding such a solution, by incorporating, for example, the energy conditions a priori. As we show in the next Sections this is indeed possible. The regular nature as well as the requirement of a finite four-dimensional Planck mass can be checked as a subsequent step, since to accomplish that, we first need to find the exact form of the warp factor and density.

3. The null energy condition

We are going to derive the null energy condition for our type of nonlinear bulk fluid described by (6) and (7) and transform it to a condition imposed on the density ρ and parameters γ and λ of the equation of state.

To achieve this, we follow the technique analysed in.[14,16] We start by recalling that the metric (1) and the bulk fluid (6) and (7) depend only on the extra dimension Y and they are therefore static with respect to the time coordinate t. Such fluid can be viewed as an anisotropic one with energy-momentum tensor

$$T_{AB} = (\rho^0 + p^0)u^0_A u^0_B - p^0 g_{\alpha\beta}\delta^\alpha_A \delta^\beta_B - p_Y g_{55}\delta^5_A \delta^5_B, \tag{11}$$

where $u^0_A = (a(Y), 0, 0, 0, 0)$, $A, B = 1, 2, 3, 4, 5$ and $\alpha, \beta = 1, 2, 3, 4$. As a next step, we compare (6) with (11) and find that the following relations should hold

$$p_Y = -\rho \tag{12}$$

$$\rho^0 = -p \tag{13}$$

$$p^0 = p. \tag{14}$$

We note that once we combine the last two relations we get

$$p^0 = -\rho^0, \tag{15}$$

which means that this type of matter satisfies a cosmological constant-like equation of state. Substituting (12)-(15) in (11), we find that

$$T_{AB} = -p g_{\alpha\beta}\delta^\alpha_A \delta^\beta_B + \rho g_{55}\delta^5_A \delta^5_B. \tag{16}$$

Here, we focus only on the null energy condition according to which, every future-directed null vector k^A should satisfy[20]

$$T_{AB}k^A k^B \geq 0. \tag{17}$$

This means that we should have

$$p + \rho \geq 0, \tag{18}$$

from which we find

$$\gamma\rho^{\lambda} + \rho \geq 0, \tag{19}$$

after inputting the equation of state $p = \gamma\rho^{\lambda}$. Now, Eq. (19) can be written as

$$\rho^{\lambda}(\gamma + \rho^{1-\lambda}) \geq 0 \tag{20}$$

and since $\rho \geq 0$ from Eq. (8), we arrive at the the final form of the null energy condition, which reads

$$\gamma + \rho^{1-\lambda} \geq 0. \tag{21}$$

In the next Section, we are going to incorporate this condition in the process of deriving our solutions.

4. Regular solution

We start solving the system of Eqs. (8)–(10), by first integrating the continuity equation (10) to find the relation between the warp factor and the density. In the integration process we arrive at a logarithmic term of the form $\ln|\gamma + \rho^{1-\lambda}|$. To incorporate from the beginning the null energy condition (21), we choose to ignore the absolute value and simply put this term equal to $\ln(\gamma + \rho^{1-\lambda})$. The resulting relation between ρ and a is

$$\rho = \left(-\gamma + c_1 a^{4(\lambda-1)}\right)^{1/(1-\lambda)}, \tag{22}$$

where

$$c_1 = \frac{\gamma + \rho_0^{1-\lambda}}{a_0^{4(\lambda-1)}}, \tag{23}$$

with $\rho_0 = \rho(Y_0), a_0 = a(Y_0)$ being the initial conditions. According to (21) this translates to $c_1 \geq 0$.

It is crucial to note that there is a singularity in the density ($\rho \to \infty$) for $\lambda > 1$ and

$$a^{4(\lambda-1)} = \frac{\gamma}{c_1}. \tag{24}$$

However, it is possible to avoid this singularity, by restricting the range of the parameter γ, only to negative values, which we do in the analysis that follows.

Substituting Eq. (22) in Eq. (8) and integrating, we obtain,

$$\int \frac{a}{(c_1 - \gamma a^{4(1-\lambda)})^{1/(2(1-\lambda))}}\, da = \pm\frac{\kappa_5}{\sqrt{6}} \int dY. \tag{25}$$

The complicated integral on the LHS of Eq. (25), can be greatly simplified for those values of λ that make $1/(2(1-\lambda))$ a negative integer. This is possible when $\lambda = (n+1)/n$, with $n = 2k$ and k a positive integer.

Clearly, the simplest case is for $n = 2$ which corresponds to $\lambda = 3/2$. Then $1/(2(1-\lambda)) = -1$ and we can directly integrate Eq. (25) to get the solution in the following implicit form

$$\pm Y + c_2 = \frac{\sqrt{6}}{\kappa_5}\left(\frac{c_1}{2}a^2 - \gamma \ln a\right), \tag{26}$$

where c_2 is an integration constant. It is straightforward to see that this solution is regular: all possible singular behaviors of a, namely $a \to 0$ and $a \to \infty$ happen only for infinite values of Y with p and ρ either vanishing there, or, approaching finite nonzero values. In particular, we have

$$a \to \infty, \quad \rho \to 0, \quad p \to 0, \quad \text{as} \quad Y \to \pm\infty \tag{27}$$

$$a \to 0^+, \quad \rho \to 1/\gamma^2, \quad p \to -1/\gamma^2, \quad \text{as} \quad Y \to \pm\infty. \tag{28}$$

Summarizing, the solution for $\lambda = 3/2$ has already two out of the three good qualities mentioned in the previous Section: it is regular and at the same time satisfies the null energy condition. In the following section, we construct an appropriate matching solution out of this solution, which leads to a finite four-dimensional Planck mass, thus fitting perfectly the profile of a plausible solution.

As for the asymptotics for general values of λ, it is interesting to note here that even though the case of $\lambda = 3/2$ is the simplest possible one for the nonlinear fluid, further analysis performed in,[1] shows that this case is significant in the sense that it can act as a guide for the whole range of values of $\lambda > 1$. This follows from the fact that all of its good physical properties are, in a way, inherited to all cases with $\lambda > 1$.

5. Matching solution for $\lambda = 3/2$

For $\lambda = 3/2$, a matching solution that leads to a finite four-dimensional Planck mass, can be constructed in the following way: First, we write down the two different branches of solution of Eq. (26),

$$Y^+ = h_+(a) = \frac{\sqrt{6}}{\kappa_5}\left(\frac{c_1^+}{2}a^2 - \gamma \ln a\right) + C_2^+ \tag{29}$$

$$Y^- = h_-(a) = \frac{\sqrt{6}}{\kappa_5}\left(-\frac{c_1^-}{2}a^2 + \gamma \ln a\right) + C_2^-, \tag{30}$$

using the notation $Y^\pm$ to describe the solutions for the $(\pm)$ choice of sign in Eq. (26). Similarly, the notation $c_1^\pm$, $C_2^\pm$ describes the values of c_1 and $\mp c_2$, respectively, on the $(\pm)$ branch of Y.

Second, in order to gain a complete idea of the behavior of $h_+(a)$, we study the shape of the curve of $h_+(a)$ by calculating the first and second derivative of $h_+(a)$.

These read

$$h'_+(a) = \frac{\sqrt{6}}{\kappa_5}\left(\frac{c_1^+ a^2 - \gamma}{a}\right) \tag{31}$$

$$h''_+(a) = \frac{\sqrt{6}}{\kappa_5}\frac{c_1^+}{a^2}\left(a^2 + \frac{\gamma}{c_1^+}\right). \tag{32}$$

For our choice of integration constants and parameters, $c_1^+ > 0$ and $\gamma < 0$, Eq. (31) shows that h_+ is strictly increasing for all values of a. Also, Eq. (32) shows that the graph of h_+ changes from concave downward on $(0, \sqrt{-\gamma/c_1^+})$ to concave upward on $(\sqrt{-\gamma/c_1^+}, \infty)$, making the point $(\sqrt{-\gamma/c_1^+}, -\sqrt{6}/\kappa_5(\gamma/2 + \gamma \ln \sqrt{-\gamma/c_1^+}) + C_2^+)$, an inflection point of the graph of h_+.

On the other hand, the first two derivatives of $h_-(a)$ are exactly the opposites of the ones of h_+ given in Eq. (31) and Eq. (32). This means that h_- is strictly decreasing and the graph of h_- changes from concave upward on $(0, \sqrt{-\gamma/c_1^-})$ to concave downward on $(\sqrt{-\gamma/c_1^-}, \infty)$, with $(\sqrt{-\gamma/c_1^-}, \sqrt{6}/\kappa_5(\gamma/2 + \gamma \ln \sqrt{-\gamma/c_1^-}) + C_2^-)$ being an inflection point of the graph of h_-.

We can match the two branches at their common inflection point by making an appropriate choice of the constants $c_1^\pm$ and $C_2^\pm$. This is a third step in the matching process. Naturally, we assume that the warp factor is continuous

$$\sqrt{\frac{-\gamma}{c_1^+}} = \sqrt{\frac{-\gamma}{c_1^-}}, \quad \text{or,} \quad c_1^+ = c_1^- = c_1 \tag{33}$$

and that, of course, the inflection point has the same Y coordinate through the two branches, which leads to

$$h_+\left(\sqrt{\frac{-\gamma}{c_1}}\right) = h_-\left(\sqrt{\frac{-\gamma}{c_1}}\right) \tag{34}$$

and yields the following relation between C_2^+ and C_2^-,

$$-\frac{\sqrt{6}}{\kappa_5}\left(\frac{\gamma}{2} + \gamma \ln \sqrt{\frac{-\gamma}{c_1}}\right) + C_2^+ = \frac{\sqrt{6}}{\kappa_5}\left(\frac{\gamma}{2} + \gamma \ln \sqrt{\frac{-\gamma}{c_1}}\right) + C_2^-, \tag{35}$$

or,

$$C_2^+ = \frac{\sqrt{6}}{\kappa_5}\left(\gamma + 2\gamma \ln \sqrt{\frac{-\gamma}{c_1}}\right) + C_2^-. \tag{36}$$

There is an axis of symmetry at $Y = Y_s$, with Y_s given by

$$Y_s = h_\pm(\sqrt{-\gamma/c_1}) = -\frac{\sqrt{6}}{\kappa_5}\left(\frac{\gamma}{2} + \gamma \ln \sqrt{\frac{-\gamma}{c_1}}\right) + C_2^+ = \frac{\sqrt{6}}{\kappa_5}\left(\frac{\gamma}{2} + \gamma \ln \sqrt{\frac{-\gamma}{c_1}}\right) + C_2^-. \tag{37}$$

Since the embedding of the brane introduces a $Y \to -Y$ symmetry in the bulk, it is natural to position the brane at $Y = Y_s$. Without loss of generality and in order to simplify our expressions, we can place the brane at $Y = 0$, that is we take $Y_s = 0$ in (37). This will set the values of $C_2^{\pm}$ to

$$C_2^+ = \frac{\sqrt{6}}{\kappa_5}\left(\frac{\gamma}{2} + \gamma \ln\sqrt{\frac{-\gamma}{c_1}}\right) = -C_2^-. \tag{38}$$

The matching solution described by Eq. (29) and (30) that satisfies the above boundary condition can be written as

$$|Y| = \frac{\sqrt{6}}{\kappa_5}\left(-\frac{c_1}{2}a^2 + \gamma \ln a - \frac{\gamma}{2} - \gamma \ln\sqrt{\frac{-\gamma}{c_1}}\right), \quad 0 < a \leq \sqrt{\frac{-\gamma}{c_1}}. \tag{39}$$

It is essential that the density is well defined and continuous at the position of the brane $Y = 0$. Here, we find

$$\rho\left(\sqrt{\frac{-\gamma}{c_1}}\right) \equiv \rho(0) = \frac{1}{4\gamma^2}, \tag{40}$$

where $\rho(0)$ denotes $\rho(Y = 0)$ and it is an abbreviation of $\rho(h_{\pm}^{-1}(\sqrt{-\gamma/c_1}))$.

The graph of h_+ is depicted in Fig. 1 with a thick solid and dashed line, while the graph of h_-, in the same Figure, is depicted with a thin solid and dashed line. The matching solution is build up only from the solid parts of both graphs, as this is the only choice that leads to a solution with a finite four-dimensional Planck mass. In Fig. 2, we have rotated the axes of Fig. 1, for a more convenient view of the evolution of the warp factor as a function of Y.

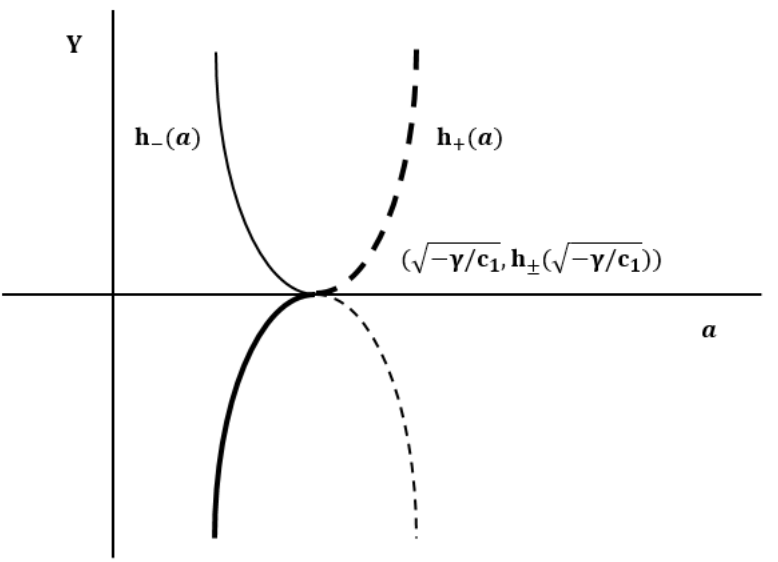


Fig. 1. Matching graphs of $h_{\pm}(a)$ for $\lambda = 3/2$.

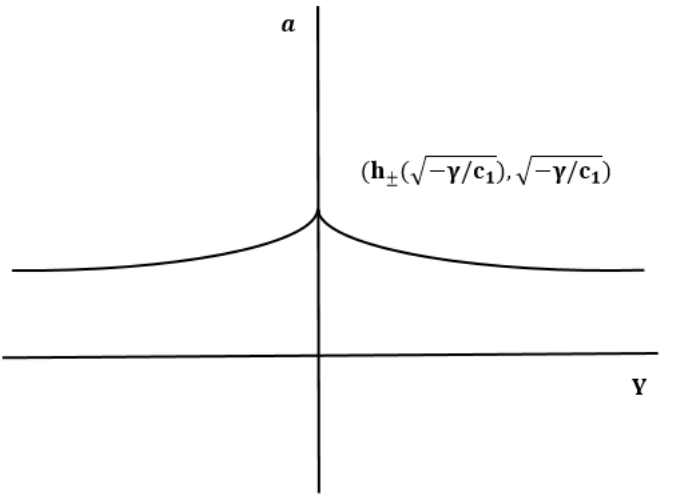


Fig. 2. Graph of $a(Y)$ for the matching solution for $\lambda = 3/2$.

Finally, we take into account the jump of the derivative of the warp factor across the brane and find, for our type of geometry, the following junction condition

$$a'(0^+) - a'(0^-) = 2a'(0^+) = -\kappa_5^2 \frac{f(\rho(0))a(0)}{3}, \tag{41}$$

where $f(\rho(0))$ is the tension of the brane. Using Eq. (39) in the above equation we can deduce the form of the brane tension, this reads

$$f(\rho(0)) = -\frac{\sqrt{6}}{2\kappa_5\gamma} = \frac{\sqrt{6\rho(0)}}{\kappa_5}, \tag{42}$$

where the last equality is derived by using Eq. (40).

This completes the construction of a viable matching solution for $\lambda = 3/2$. In the next Section, we are going to show that this solution also leads to a finite four-dimensional Planck mass.

6. Localization of gravity for $\lambda = 3/2$

To calculate the value of the four-dimensional Planck mass, $M_p^2 = 8\pi/\kappa$, we use the following integral (see[1, 14, 16, 21]),

$$\frac{\kappa_5^2}{\kappa} = \int_{-Y_c}^{Y_c} a^2(Y)dY. \tag{43}$$

We can deduce the behaviour of a^2 as $Y \to -\infty$ from Eq. (39). We find

$$a^2 \sim e^{-(\sqrt{6}\kappa_5/(3\gamma))Y}, \tag{44}$$

where the symbol $\sim$ is used to signify 'behaves according to', and using the symmetry of Eq. (39), we can write the above integral in the following form

$$\int_{-Y_c}^{Y_c} a^2 dY = 2\int_{-Y_c}^{0} a^2 dY \sim 2\int_{-Y_c}^{0} e^{-(\sqrt{6}\kappa_5/(3\gamma))Y} dY = -\sqrt{6}\frac{\gamma}{\kappa_5}(1-e^{(\sqrt{6}\kappa_5/(3\gamma))Y_c}). \tag{45}$$

Taking $Y_c \to \infty$ and keeping in mind that $\gamma < 0$, it follows that the four-dimensional Planck mass remains finite and is proportional to

$$-\sqrt{6}\frac{\gamma}{\kappa_5}.$$

In conclusion, we have seen a specific example of the nonlinear equation of state $p = \gamma\rho^{3/2}$, for which it is possible to construct a regular matching solution compatible with the null condition and which also localizes gravity on the brane. In the discussion below, we outline what happens for extended ranges of values of λ.

7. Discussion

We briefly review in this Section, the behavior of solutions for general λ, a detailed analysis can be found in.[1]

The key factor that determines whether a solution of Eqs. (8)-(10), is regular, or, singular, is the ordering of λ with respect to the value one: for $\lambda < 1$, we encounter singular solutions, whereas, for $\lambda > 1$ the solutions are regular.

Before presenting the solutions for general λ, we emphasize that for $\lambda \neq 1 + 1/(2k)$, where k is a positive integer, the integral on the LHS of Eq. (25) cannot in general be calculated directly. However, we can still deduce the behaviors of solutions by expressing them in terms of the Gaussian hypergeometric function ${}_2F_1(\alpha, b, c; z)$.

In particular, for $\lambda < 1$, the solution is given implicitly in terms of a hypergeometric function and reads

$$\pm Y + c_2 = \frac{\sqrt{6}}{2\kappa_5} c_1^{1/(2(\lambda-1))} a^2 \, {}_2F_1\left(\frac{1}{2(1-\lambda)}, \frac{1}{2(1-\lambda)}, \frac{1}{2(1-\lambda)} + 1; \frac{\gamma}{c_1} a^{4(1-\lambda)}\right). \tag{46}$$

As shown in,[1] this solution has a finite-distance singularity at $Y \to \pm c_2$, with

$$a \to 0^+, \quad \rho \to \infty, \quad p \to 0, \quad \text{if} \quad \lambda < 0 \tag{47}$$

$$a \to 0^+, \quad \rho \to \infty, \quad p \to \infty, \quad \text{if} \quad 0 < \lambda < 1. \tag{48}$$

The behaviors of p and ρ above, have been deduced from Eq. (7) and Eq. (22).

For $\lambda > 1$, on the other hand, we have the following types of solution. First, we have the general form of solution for $\lambda = 1 + 1/(2k)$, with k a positive integer, which is

$$\pm Y + c_2 = \frac{\sqrt{6}}{\kappa_5}\left(\sum_{s=0}^{k-1} \frac{k!}{(k-s)!s!} \frac{c_1^{k-s}}{2-2s/k} a^{2-2s/k}(-\gamma)^s + (-\gamma)^k \ln a\right). \tag{49}$$

Second, for $\lambda > 3/2$ we have the following solution

$$\pm Y + c_2 = \frac{\sqrt{6}}{\kappa_5}\left(\frac{a^2}{2}(c_1 - \gamma a^{4(1-\lambda)})^{1/(2(\lambda-1))} - \frac{\gamma c_1^{(3-2\lambda)/(2(\lambda-1))}}{2(3-2\lambda)}a^{2(3-2\lambda)} \times {}_2F_1\left(\frac{3-2\lambda}{2(1-\lambda)}, \frac{3-2\lambda}{2(1-\lambda)}, \frac{3-2\lambda}{2(1-\lambda)}+1; \frac{\gamma}{c_1}a^{4(1-\lambda)}\right)\right), \tag{50}$$

Finally, the case of $1 < \lambda < 3/2$ is more complicated: we have forms of solutions valid for $1+1/(2k) < \lambda < 1+1/2(k-1)$, with k a positive integer such that $k \geq 2$, and given by

$$\pm Y + c_2 = \frac{\sqrt{6}}{\kappa_5}\left(\sum_{s=0}^{n/2-1}\frac{(-\gamma)^s}{2(1-2s(\lambda-1))}\left(c_1 a^{4(\lambda-1)} - \gamma\right)^{1/(2(\lambda-1))-s} + \frac{(-\gamma)^{n/2}(c_1)^{((n+1)-n\lambda)/(2(\lambda-1))}}{2((n+1)-n\lambda)}a^{2((n+1)-n\lambda)} \times {}_2F_1\left(\frac{(n+1)-n\lambda}{2(1-\lambda)}, \frac{(n+1)-n\lambda}{2(1-\lambda)}, \frac{(n+1)-n\lambda}{2(1-\lambda)}+1; \frac{\gamma}{c_1}a^{4(1-\lambda)}\right)\right). \tag{51}$$

All solutions for $\lambda > 1$ are free from finite-distance singularities, and their asymptotic behaviors are

$$a \to \infty, \quad \rho \to 0, \quad p \to 0, \quad \text{as} \quad Y \to \pm\infty \tag{52}$$

$$a \to 0^+, \quad \rho \to (-\gamma)^{1/(1-\lambda)}, \quad p \to -(-\gamma)^{1/(1-\lambda)}, \quad \text{as} \quad Y \to \pm\infty. \tag{53}$$

The procedure for deriving a matching solution with a finite four-dimensional Planck mass, presented in the previous Sections, can still be performed for all solutions with $\lambda > 1$ and can be found in.[1]

Summarizing, the effect of the nonlinear equation of state is catalytic in the process of creating a regular solution with essential physical properties and valid for a wide range of parameters. It is left to be seen from a physics point of view, what are the possible field realizations that could lead to such an equation of state.

Acknowledgments

Work partially performed by I.A. as International professor of the Francqui Foundation, Belgium.

References

1. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Regular branewords with nonlinear bulk-fluids, *Eur. Phys. J. C* 81, (2021). [arXiv:2106.15669v2 [hep-th]].
2. S. K. Srivastava, Future Universe With $w < -1$ Without Big Smash, *Phys. Lett. B* 619, 1-4 (2005). [arXiv:0407048v4 [astro-ph]].

3. P. F. Gonzalez-Diaz, You need not be afraid of phantom energy, *Phys. Rev. D* 68, 021303 (2003). [arXiv:0305559 [astro-ph]].
4. A. V. Astashenok, S. Nojiri, S. D. Odintsov, A. V. Yurov. Phantom Cosmology without Big Rip Singularity, *Phys. Lett. B* 709, 396-403 (2012). [arXiv:1201.4056 [gr-qc]].
5. J. D. Barrow, Graduated inflationary universes, *Phys. Lett. B* 235, 40-43 (1990).
6. A. Y. Kamenshchik, U. Moschella, V. Pasquier, An alternative to quintessence, *Phys. Lett. B* 511, 265-268 (2001). [arXiv:0103004 [gr-qc]].
7. M. C. Bento, O. Bertolami, A. A. Sen., Generalized Chaplygin gas, accelerated expansion and dark energy-matter unification, *Phys. Rev. D* 66, 043507 (2002). [arXiv:0202064 [gr-qc]].
8. S. Cotsakis, I. Klaoudatou, Future Singularities of Isotropic Cosmologies, *J. Geom. Phys.* 55, 306-315 (2005). [arXiv:0409022 [gr-qc]].
9. S. Cotsakis, I. Klaoudatou, Cosmological Singularities and Bel-Robinson Energy, *J. Geom. Phys.* 57, 1303-1312 (2007). [arXiv:0604029 [gr-qc]].
10. S. Nojiri, S. D. Odintsov, S. Tsujikawa, Properties of singularities in (phantom) dark energy universe, *Phys. Rev. D* 71, 063004 (2005). [arXiv:0501025 [hep-th]].
11. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Braneworld cosmological singularities, *Proceedings of the 11th Marcel Grossmann Meeting on General Relativity* 3, 2054-2056 (2007). [arXiv:0701033 [gr-qc]].
12. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Brane singularities and their avoidance, *Class. Quant. Grav.* 27, 235018 (2010). [arXiv:1010.6175 [gr-qc]].
13. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Brane singularities with mixtures in the bulk, *Fortschr. Phys.* 61, 20-49 (2013). [arXiv:1206.0090 [hep-th]].
14. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Enveloping branes and braneworld singularities, *Eur. Phys. J. C* 74, 3192 (2014). [arXiv:1406.0611v2 [hep-th]].
15. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Dynamics and asymptotics of brane-worlds, *Proceedings of the 13th Marcel Grossmann Meeting on General Relativity* 3, 1859-1861 (2015).
16. I. Antoniadis, S. Cotsakis, I. Klaoudatou, Curved branes with regular support, *Eur. Phys. J. C* 76, 511 (2016). [arXiv:1606.09453 [hep-th]].
17. N. Arkani-Hamed, S. Dimopoulos, N. Kaloper, R. Sundrum, A small cosmological constant from a large extra dimension, *Phys. Lett. B* 480, 193-199 (2000). [arXiv:0001197v2 [hep-th]].
18. S. Kachru, M. Schulz, E. Silverstein. Bounds on curved domain walls in 5d gravity, *Phys. Rev. D* 62, 085003 (2000). [arXiv:0002121 [hep-th]].
19. S. S. Gubser, Curvature singularities: The good, the bad, and the naked, *Adv. Theor. Math. Phys.* 4, 679-745 (2000). [arXiv:0002160 [hep-th]].
20. E. Poisson, *A Relativist's Toolkit* (Cambridge University Press, 2004).
21. S. Forste, H. P. Nilles and I. Zavala, Nontrivial Cosmological Constant in Brane Worlds with Unorthodox Lagrangians, *JCAP* 07, 007 (2011). [arXiv:1104.2570 [hep-th]].

Primordial synchronization of Mixmaster spatial points

Spiros Cotsakis

Institute of Gravitation and Cosmology, RUDN University
ul. Miklukho-Maklaya 6, Moscow 117198, Russia
and
Research Laboratory of Geometry, Dynamical Systems and Cosmology
University of the Aegean, Karlovassi 83200, Samos, Greece
E-mail: skot@aegean.gr

We review recent work on the possibility of primordial synchronization of different Mixmaster regions in generic inhomogeneous spacetime. It is shown that inhomogeneous domains undergoing chaotic oscillations may synchronize themselves exponentially fast and proceed in perfect symphony asymptotically in the past direction. Implications of this result for the structure and evolution of an early phase of the universe are briefly discussed.

Keywords: Nature of the cosmological singularity; structure of primordial states; chaotic synchronization

1. Introduction

Recently, we introduced a dynamical mechanism via chaotic synchronization of different spatial point of Mixmaster type as a new factor governing the global structure and evolution of primordial cosmology.[1] We have shown that a pair of spatial Mixmaster points in generic inhomogeneous spacetime coupled to each other and one of them ('the transmitter') communicating to the other ('the receiver') only part of the information necessary to determine its state, evolves in such a way that the receiver becomes able to completely reconstruct the remaining information and synchronize its chaotic evolution with that of the transmitter point.

This mechanism radically changes the structure of the early universe and provides a novel setting for the consideration of a number of fundamental issues in cosmology, such as the homogeneity ('horizon') problem, and the unexplained behaviour of the entropy in the early universe. In this contribution, we shall briefly recall the main features of the proposed cosmological sync mechanism.[1]

2. Evolution of inhomogeneous Mixmaster points

Our results concern the so-called G_0 cosmologies that admit no isometries and may be described in vacuum by the dimensionless state vector field X along the orbit of a fixed spatial point A. This is given by,[2]

$$X_A(\tau) = (\Sigma_+, \Sigma_-, N_1, N_2, N_3), \quad \Sigma_\pm = \frac{\sigma_\pm}{H}, \; N_\alpha = \frac{n_\alpha}{H}, \tag{1}$$

where H denotes the Hubble scalar function equal to $1/3\Theta$ and is taken with respect to the fundamental 4-velocity timelike vector field $u^a, a = 0, 1, 2, 3$. The expansion scalar is defined by $\Theta = \nabla_a u^a$. Derivatives are taken with respect to the dimensionless τ-time related to the proper (clock) time by $dt/d\tau = 1/H$, the variables $\Sigma_\pm$ describe the anisotropy in the Hubble flow, while the N_α's are related to the spatial curvature and the Bianchi type of the isometry group associated with the evolution of the spatial point.

We assume that different spatial points evolve in the past direction as separate homogeneous Mixmaster universes coupled to each other in inhomogeneous spacetime. For any pair of spatial points A, B, we have a pair of dynamical equations, namely, for the point A we have equations for the (N, Σ) variables, and for the point B similar equations for the different (M, Π) variables as follows[3–5]:

System A:

$$N_1' = (q - 4\Sigma_+)N_1, \tag{2}$$

$$N_2' = (q + 2\Sigma_+ + 2\sqrt{3}\Sigma_-)N_2, \tag{3}$$

$$N_3' = (q + 2\Sigma_+ - 2\sqrt{3}\Sigma_-)N_3, \tag{4}$$

$$\Sigma_+' = -(2 - q)\Sigma_+ - 3S_+, \tag{5}$$

$$\Sigma_-' = -(2 - q)\Sigma_- - 3S_-, \tag{6}$$

with the constraint,

$$\Sigma_+^2 + \Sigma_-^2 + \frac{3}{2}\left(N_1^2 + N_2^2 + N_3^2 - 2(N_1N_2 + N_2N_3 + N_3N_1)\right) = 1, \tag{7}$$

where,

$$q = 2(\Sigma_+^2 + \Sigma_-^2), \tag{8}$$

$$S_+ = \frac{1}{2}\left((N_2 - N_3)^2 - N_1(2N_1 - N_2 - N_3)\right), \tag{9}$$

$$S_- = \frac{\sqrt{3}}{2}(N_3 - N_2)(N_1 - N_2 - N_3), \tag{10}$$

and a prime denoting differentiation with respect to the τ-time.

System B: We have the unknowns $y = (M, \Pi)$, the variables $M = (M_1, M_2, M_3)$ satisfy similar equations to the Eqns. (2)-(4), and the shear variables $\Pi = (\Pi_+, \Pi_-)$ satisfy the system,

$$\Pi_+' = -(2 - p)\Pi_+ - 3Q_+, \tag{11}$$

$$\Pi_-' = -(2 - p)\Pi_- - 3Q_-, \tag{12}$$

with $p = 2(\Pi_+^2 + \Pi_-^2)$, the Q's are like the variables S's in the Eqns. (9), (10) but with the M's in the corresponding places of N's. The constraint is identical to (7) but with the (M, Π)'s in the places of the (N, Σ)'s.

3. The problem of cosmological sync

We shall be interested in the possibility that the spatial points A, B, with A in the causal past of B, can influence each other by sending signals from A to B and sync with each other in the past direction, while chaotically oscillating in inhomogeneous Mixmaster spacetime.

We imagine that the A-system in the form $x' = f(x)$ for the variables $x = (N, \Sigma)$, breaks into two 'subsystems', with the N variables satisfying the first three equations ('sys-1') and the Σ's the last two ('sys-2'), namely,

$$N' = g(N, \Sigma), \quad \Sigma' = h(N, \Sigma), \tag{13}$$

where $N = (N_1, N_2, N_3)$, $\Sigma = (\Sigma_+, \Sigma_-)$. We then set the B-system (receiver) variables M simply equal to the corresponding A-transmitter variables N (that is to those in sys-1). Then the B-system equations break into the pair consisting of the sys-1 equations, and those identical to sys-2 but for the Π variables, the 'sys-3' system:

$$M_i = N_i, \quad i = 1, 2, 3 \tag{14}$$

$$\Pi'_+ = -(2-p)\Pi_+ - 3S_+, \tag{15}$$

$$\Pi'_- = -(2-p)\Pi_- - 3S_-. \tag{16}$$

Following,[6,7] we introduce the synchronization (error) function $\Omega = (\Omega_+, \Omega_-)$, which in the present case we take it to be: $\Omega = \Sigma - \Pi$, that is,

$$\Omega_+ = \Sigma_+ - \Pi_+, \quad \Omega_- = \Sigma_- - \Pi_-. \tag{17}$$

Then we say that there is complete synchronization of the two Mixmaster oscillating spatial points (that is of systems A and B) provided we can show that,

$$\Omega \to (0, 0), \quad \text{as} \quad \tau \to -\infty. \tag{18}$$

If this happens, we say that system A is synchronized with system B. Otherwise, the two oscillating spatial points corresponding to systems A and B will evolve autonomously, out of sync.

In general, we expect the appearance of extra coupling terms in the dynamical equations which describe the evolution of the sync function Ω during the past evolution the Mixmaster oscillating spatial points, the latter evolution given by Eqns. (5), (6), and (11), (12). The question then arises as to whether or not a system of spatial points chaotically oscillating in the past direction may sync and evolve in unison thereafter.

This is the problem of cosmological sync for the case of spatial points satisfying Mixmaster-type evolution equations. As shown in,[1] there are indeed different types of sync possible in this problem depending on the form of the coupling between different spatial points.

4. Generic sync

Let us suppose that there is a linear in the sync function Ω coupling between the two spatial points. In this case, the evolution equations become,

$$\Sigma' = -(2-q)\Sigma - 3S + \alpha(\Pi - \Sigma) \tag{19}$$

$$\Pi' = -(2-p)\Pi - 3Q + \alpha(\Sigma - \Pi), \tag{20}$$

where Q is the analogous expression of S for the Π system, and α is some coupling between the two oscillators.

Then if the orbit Σ has Lyapunov exponent λ, we find that the associated Ω-evolution equation leads to the result,[1]

$$|\Omega| \leq Ce^{(\lambda - 2\alpha)\tau}, \tag{21}$$

with C an integration constant.

Therefore if we set $\alpha_c = \lambda/2$, for the critical coupling strength, we conclude that when $\alpha > \alpha_c$, we have complete synchronization of the Mixmaster oscillating regions.

A typical value of the Lyapunov exponents for Mixmaster orbits is (see, e.g.,[8]) $\lambda \sim 0.45$, so that when $\alpha > 0.225$, we have synchronization between the two spatial Mixmaster points A, B.

5. Lyapunov function sync

As is further shown in,[1] there is in fact a Lyapunov function for the dynamics of sync, for a general coupling $g(\Omega)$ satisfying a positivity condition, and under the further condition that,

$$\Omega' \cdot \nabla g \leq 0, \tag{22}$$

where Ω' is the orbital derivative of Ω. The Lyapunov function is given by,

$$V(\Omega_+, \Omega_-) = \frac{1}{2}\left(\Omega_+^2 + \Omega_-^2\right) + g(\Omega), \tag{23}$$

This implies that the state $\Omega = 0$ is asymptotically stable, making synced spatial regions globally asymptotically stable in the past direction.

As an example, consider the simplest, uncoupled, case. The sync equations of the dynamics take the form,

$$\Omega_+' = -(2-q)\Omega_+ \tag{24}$$

$$\Omega_-' = -(2-q)\Omega_-. \tag{25}$$

Then the function,

$$V(\Omega_+, \Omega_-) = \frac{1}{2}\left(\Omega_+^2 + \Omega_-^2\right), \tag{26}$$

is a Lyapunov function for the dynamics and satisfies,

$$V \leq V_0 e^{-4\tau}. \tag{27}$$

where V_0 is a constant. We conclude that Mixmaster regions in inhomogeneous spacetime synchronize exponentially fast.

6. Phase sync

A different characterization of the possibility of cosmological sync in the assumed primordial chaotic state of the universe may be effected when considering any number of oscillating domains having their own frequencies. This leads to different dynamical evolution of their phases, and the question arises as to whether we can have phase sync between them in this situation.

This problem was shown to have an affirmative answer in two steps. In the first step,,[9] Barrow wrote down the equation describing the phase evolution of a spatial Mixmaster point as a response of the system in terms of the mean field amplitude and phase r, ψ respectively,

$$\theta' = \frac{3}{\rho}\sqrt{S_+^2 + S_-^2}\,\sin(\theta - \psi), \quad \tan\psi = S_-/S_+, \tag{28}$$

with ρ being the polar coordinate in the pair (ρ, θ) in the (Σ_+, Σ_-) plane. A further elaboration using these dynamical equations shows that sync is indeed possible.

It can be further shown[1] that this equation leads to a direct coupling between the oscillating domains, precisely given by a time-dependent Kuramoto-like coupling of spontaneous synchronization. For any partition of N regions, we can show that the i-th domain responds directly to the j-th domain in the partition via the equation,[1]

$$\theta_i' = \frac{3}{N\rho_i}\sqrt{S_+^2 + S_-^2}\sum_{j=1}^{N}\sin(\theta_j - \theta_i). \tag{29}$$

7. Physical meaning of sync

We are led to the following physical interpretation[1] of the process of sync in the present context. The spatial point B will synchronize with the point A lying in B's past after receiving a signal from A. This will be so because their sync function $\Omega \to 0$, and sync will become progressively apparent when their corresponding BKL[10–12] parameter u values for A and B become equal with each other in the sense that the numbers of Kasner epochs and eras for the two evolutions of the spatial points A, B become identical to each other exponentially fast. Thereafter the two points will start oscillating in perfect unison.

This process will lead to the simplest possible initial state of the universe, characterized by a perfect sync between its different but correlated spatial regions. This will be so because, according to the results reported here, every spatial point will continually send and receive signals from other spacetime points shifting their state

of oscillation, adjusting them to that of the other points, and resetting their motion, until sync organizes them all in perfect harmony. In turn, this may have important implications some of which will be discussed elsewhere.

References

1. S. Cotsakis, *Onset of synchronization in coupled Mixmaster oscillators*, to appear in the Phil. Trans. Roy. Soc. A.; arXiv: 2010.00298.
2. C. Uggla, H. van Elst, J. Wainwright and G.F.R. Ellis, Phys. Rev. D 68 (2003) 103502; arXiv: gr-qc/0304002.
3. G. F. R. Ellis and M. MacCallum, Commun. Math. Phys. 12, 108 (1969).
4. J. Wainwright and L. Hsu, Class. Quantum Gravity 6, 1409 (1989).
5. J. Wainwright and G. F. R. Ellis, *Dynamical systems in cosmology* (CUP. 1993).
6. L. M. Pecora, T. L. Carroll, Phys. Rev. Lett. 64, 821 (1990).
7. L. M. Pecora, *et al*, Chaos 7, 520 (1997).
8. B. K. Berger, Gen. Rel. Grav. 23, 1385 (1991).
9. J. D. Barrow, Phys. Rev. D 102, 024017 (2020); arXiv: gr-qc/2006.01562.
10. V. A. Belinski, I. M. Khalatnikov, and E. M. Lifshitz, Adv. Phys. 12, 185 (1963).
11. V. A. Belinski, I. M. Khalatnikov, and E. M. Lifshitz, Adv. Phys. 19, 525 (1970).
12. V. A. Belinski, I. M. Khalatnikov, and E. M. Lifshitz, Adv. Phys. 31, 639 (1982).

Quintessential inflation from Lorentzian slow roll

David Benisty

DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK
Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

From the assumption that the slow roll parameter ϵ has a Lorentzian form as a function of the e-folds number N, a successful model of a quintessential inflation is obtained. The form corresponds to the vacuum energy both in the inflationary and in the dark energy epochs. The form satisfies the condition to climb from small values of ϵ to 1 at the end of the inflationary epoch. At the late universe ϵ becomes small again and this leads to the Dark Energy epoch. The observables that the models predicts fits with the latest Planck data: $r \sim 10^{-3}, n_s \approx 0.965$. Naturally a large dimensionless factor that exponentially amplifies the inflationary scale and exponentially suppresses the dark energy scale appears, producing a sort of *cosmological see saw mechanism*. We find the corresponding scalar Quintessential Inflationary potential with two flat regions - one inflationary and one as a dark energy with slow roll behavior.

Keywords: Dark Matter; Dark Energy

1. Introduction

The inflationary paradigm is considered as a necessary part of the standard model of cosmology, since it provides the solution to the fundamental puzzles of the old Big Bang theory, such as the horizon, the flatness, and the monopole problems.[1–9] It can be achieved through various mechanisms, for instance through the introduction of a scalar inflaton field.[10–25] Almost twenty years after the observational evidence of cosmic acceleration the cause of this phenomenon, labeled as dark energy", remains an open question which challenges the foundations of theoretical physics: why there is a large disagreement between the vacuum expectation value of the energy momentum tensor which comes from quantum field theory and the observable value of dark energy density.[26–28] One way to parametrize dynamical dark energy uses a scalar field, the so-called quintessence model for canonical scalar fields.[29–31] In such a way that the cosmological constant gets replaced by a dark energy fluid with a nearly constant density today.[32–36] For the slow roll approximation the scalar field behaves as an effective dark energy. The form of the potential is clearly unknown and many different potentials have been studied and confronted to observations.

These two regimes of accelerated expansion are treated independently. However, it is both tempting and economical to think that there is a unique cause responsible for a quintessential inflation[37–42, 42–48] which refers to unification of both concepts

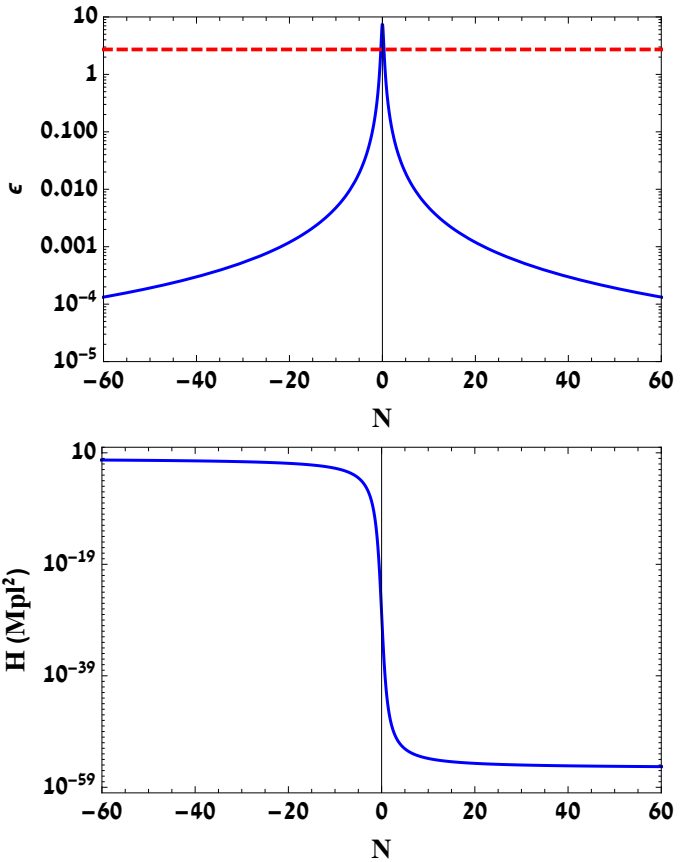


Fig. 1. The upper panel shows the slow roll parameter ϵ vs. the number of e-folds for the ansatz (1), in a logarithmic scale. The lower panel shows the corresponding Hubble function of the vacuum vs. the number of e-folds.

using a single scalar field. Consistency of the scenario demands that the new degree of freedom, namely the scalar field, should not interfere with the thermal history of the Universe, and thereby it should be "invisible" for the entire evolution and reappear only around the present epoch giving rise to late-time cosmic acceleration.

2. Lorentzian Ansatz

In order to formulate an ansatz for the Hubble function that treats symmetrically both the early and late times we use the Lorentzian function for the slow roll parameter:

$$\epsilon(N) = \frac{\xi}{\pi} \frac{\Gamma/2}{N^2 + (\Gamma/2)^2} \tag{1}$$

as a function of the number of e-folds $N = \log(a/a_i)$, where a_i is the scale parameter at some time (which we may choose as the initial state of the inflationary phase). ξ is the amplitude of the Lorentzian, Γ is the width of the Lorentzian. In that way the ϵ parameter increases from the initial value to 1 at the end of inflation, then continues to increase, peak and then decreases until it gets down to the value 1 and this represents the beginning of a the new Dark Energy phase that will eventually dominate the late evolution of the Universe. The upper panel of Fig. 1 presents the qualitative shape of this behavior.

The strong energy condition yields another bound on the coefficients. The equation of states w is in the range $|w| \leq 1$. From the relation $\epsilon = \frac{3}{2}(w+1)$ we obtain the bound $0 \leq \epsilon \leq 3$. The ansatz for the vacuum energy evolution (1) positive always, hence the lower bound is preserved. The largest value of the ansatz (1) is $2\xi/\pi\Gamma$. From the the upper bound of ϵ we obtain the condition:

$$\Gamma < 2\xi/3\pi. \tag{2}$$

In general, the calculation of the above observables demands a detailed perturbation analysis. Nevertheless, one can obtain approximate expressions by imposing the slow-roll assumptions, under which all inflationary information is encoded in the slow-roll parameters. In particular, one first introduces[49]

$$\epsilon_{n+1} = \frac{d}{dN} \log |\epsilon_n|, \tag{3}$$

where $\epsilon_0 \equiv H_i/H$ and n a positive integer. The slow roll parameters read:

$$\epsilon \equiv \epsilon_1 = -\frac{H'}{H}, \quad \epsilon_2 = \frac{H''}{H'} - \frac{H'}{H},$$

and so on. From the first slow roll parameter definition with the ansatz (1), we obtain the solution:

$$H = \sqrt{\frac{\Lambda_0}{3}} \exp\left[-\frac{\xi}{\pi} \tan^{-1}\left(\frac{2N}{\Gamma}\right)\right], \tag{4}$$

where Λ_0 is an integration constant. The Hubble function interpolates from the inflationary values $H_{-\infty}$ to the dark energy value $H_{+\infty}$ that corresponds to:

$$H_{\pm} = \sqrt{\frac{\Lambda_0}{3}} \exp^{\mp \xi/2}. \tag{5}$$

The magnitude of the vacuum energy at the inflationary phase reads $10^{-8} Mpl^4$, while the magnitude of the vacuum energy at the present slowly accelerated phase of the universe is $10^{-120} Mpl^4$. From the Friedmann equations the values of the energy density is $3H^2$ in the Planck scale. Therefore, the coefficients of the model are:

$$\xi \approx 129, \quad \Lambda_0 = 1.7 \cdot 10^{-32} Mpl^4. \tag{6}$$

We calculate the other slow roll parameters using (3):

$$\epsilon_2 = -\frac{8N}{\Gamma^2 + 4N^2}, \quad \epsilon_3 = \frac{1}{N} - \frac{8N}{\Gamma^2 + 4N^2}. \tag{7}$$

For $\Gamma \to 0$ all of the slow roll parameters with $n \geq 3$ yields the value $-1/N$. However in the general case, all of the slow parameters have small values if the ϵ_2 is small.

As usual inflation ends at a scale factor a_f where $\epsilon_1(a_f) = 1$ and the slow-roll approximation breaks down. Therefore the end of inflation takes place when the number of e-folds reads:

$$N_f = \pm\sqrt{\frac{\Gamma}{4\pi}(2\xi - \pi\Gamma)}. \tag{8}$$

Notice that with the condition (2) it gets a definite value. In order to have an inflationary phase the condition $2\xi > \pi\Gamma$ must be satisfied. The negative value of N_f is the final state of the inflationary phase, while the positive value of N_f is the initial value of the slow rolling Dark Energy at the late universe. Therefore, in order to calculate the inflationary observables, we must take the minus sing of N_f. Consequently the initial N_i satisfies the condition: $N_f - N_i = \mathcal{N} \approx 50 - 60$, where we impose 60 e-folds for the inflationary phase. Hence, the initial state of the inflationary phase reads:

$$N_i = -\sqrt{\frac{\Gamma}{4\pi}(2\xi - \pi\Gamma)} - \mathcal{N}. \tag{9}$$

The inflationary observables are expressed as[49]

$$r \approx 16\epsilon_1, \quad n_{\rm s} \approx 1 - 2\epsilon_1 - \epsilon_2, \quad \alpha_{\rm s} \approx -2\epsilon_1\epsilon_2 - \epsilon_2\epsilon_3, \quad n_{\rm T} \approx -2\epsilon_1, \tag{10}$$

where all quantities are calculated at N_i. Therefore the tensor to scalar ratio and the primordial tilt give:

$$r = \frac{32\Gamma\xi}{\pi\Gamma^2 + 4\pi N_i^2}, \quad n_s = \frac{\pi\left(\Gamma^2 + 4N_i(N_i+2)\right) - 4\Gamma\xi}{\pi\left(\Gamma^2 + 4N_i^2\right)}. \tag{11a}$$

For 60 e-folds and $\Gamma = 0.1$ the observables read:

$$r = 0.0076, \quad n_s = 0.961754. \tag{12}$$

These values in agreement with the latest 2018 Planck data[50,51]:

$$0.95 < n_s < 0.97, \quad r < 0.064. \tag{13}$$

Fig. 2 shows the predicted distribution of the observables.[52] We assume a uniform prior: $N \in [50; 70]$, $\xi \in [100; 200]$, $\Gamma \in [0; 1]$, with 10^7 Markov Chain Monte Carlo samples. We find the posterior yields:

$$r = 0.045^{+0.065}_{-0.053}, \tag{14}$$

$$n_s = 0.9624^{+0.0087}_{-0.011}, \tag{15}$$

$$\alpha_s = -\left(33^{+27}_{-30}\right) \cdot 10^{-5}, \tag{16}$$

in good agreement with the recent Planck values.

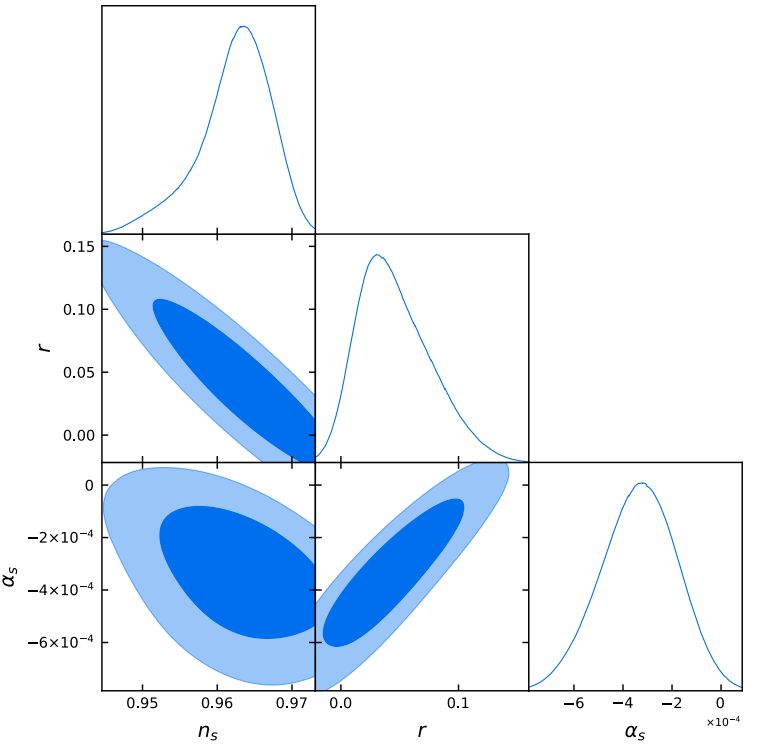


Fig. 2. The predicted scalar to tensor ratio vs. the primordial tilt of the model.

3. Scalar field dynamics

The above ansatz is of general applicability in any inflation realization, whether this is driven by a scalar field, or it arises effectively from modified gravity, or from any other mechanism. In order to provide a more transparent picture let us consider a realization of these ideas in the context of a canonical scalar field theory ϕ moving in a potential $V(\phi)$. The Friedmann equations are

$$H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right], \quad \dot{H} = -4\pi G\dot{\phi}^2, \tag{17}$$

while that for the scalar field is

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \tag{18}$$

Let us apply the ansatz in order to reconstruct a physical scalar-field potential that can generate the desirable inflationary observables. From the Friedmann equation (17) that holds in every scalar-field inflation, we extract the following solutions

$$\phi = \int_0^N \sqrt{-2\frac{H'}{H}}\, dN, \quad V(\phi) = HH' + 3H^2, \tag{19}$$

with $8\pi G = 1$. From the integration of the Hubble parameter we get:

$$N = \frac{\Gamma}{2}\sinh\left(\sqrt{\frac{\pi}{\xi\Gamma}}\phi\right), \quad V(N) = \Lambda_0 e^{-\frac{2\xi}{\pi}\tan^{-1}\left(\frac{2N}{\Gamma}\right)}\left(1 - \frac{2\Gamma\xi}{3\pi\Gamma^2 + 12\pi N^2}\right). \quad (20)$$

Relations (20) can be used in order to write V in terms of ϕ analytically

$$V(\phi) = \Lambda_0 e^{-\frac{2\xi}{\pi}\tan^{-1}(\sinh x)}\left(1 - \frac{2\xi}{3\pi\Gamma}\mathrm{sech}^2 x\right), \quad (21)$$

with $x \equiv \sqrt{\pi/\Gamma\xi}\phi$. Fig. 3 shows the scalar potential $V(\phi)$. The universe in this picture begins with $\phi \to \infty$ with a slow roll behavior and goes to the left-hand side. After approaching the minimum the universe evolves with another slow roll behavior that corresponds to the dark energy epoch when $\phi \to -\infty$. The asymptotic values of the potential are:

$$V_{+\infty} = \Lambda_0 e^{\xi}, \quad V_{-\infty} = \Lambda_0 e^{-\xi}. \quad (22)$$

Notice that this represents a see saw cosmological effect, that is if Λ_0 represents an intermediate scale, we see that in order to make the inflationary scale big forces the present vacuum energy to be small. Λ_0 represents the geometric average of the inflationary vacuum energy and the present Dark Energy vacuum energies.

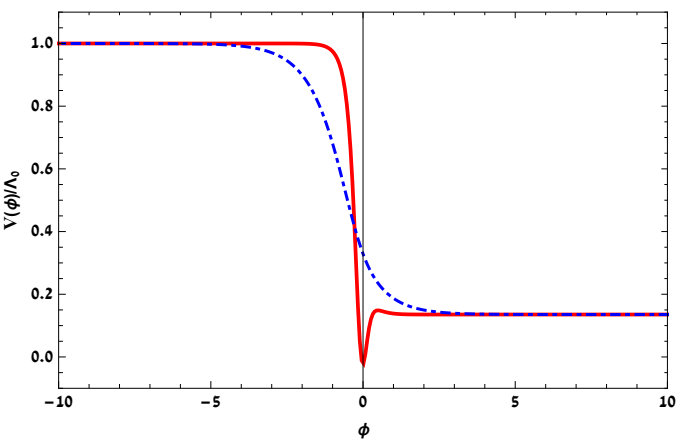


Fig. 3. The corresponding scalar field potential for the Lorenzian ansatz, with different values of Γ: 0.1 red smooth line, 1 blue dashed line.

4. Discussion

This essay introduces a model where we start with an ansatz for the slow roll parameter ϵ for the whole history of the Universe. We choose a Lorentzian form for ϵ, which peaks at some point and goes to zero for the early and late Universe, so these two epochs have an accelerated phase. The magnitude of the vacuum energies

at the early and late Universe obeys a see saw mechanism: since the asymptotic values of the potential are $\Lambda_0 e^{\pm\xi}$ this represents a see saw cosmological effect, where the requirement that one scale (the inflationary scale) be large pushes the Dark Energy scale to be very low. See saw cosmological effects in modified measure theories with spontaneously broken scale invariance have been studied in.[53–55] For the situation presented in this paper to work, we must choose Λ_0 as an intermediate scale, and indeed then we see that in order to make the inflationary scale big, this forces the present vacuum energy to be small. Λ_0 represents the geometric average of the inflationary vacuum energy and the present Dark Energy vacuum energies.

The model formulates the vacuum energies both in the inflationary epoch and in the dark energy epoch. However to compare the basis of the model with the whole history of universe, we have to take into account particle creation models with temperature, as well as radiation production.

Acknowledgments

D.B. gratefully acknowledge the support the supports of the Blavatnik and the Rothschild fellowships. We have received partial support from European COST actions CA15117 and CA18108.

References

1. A. H. Guth, *Phys. Rev.* **D23** (1981) 347, [Adv. Ser. Astrophys. Cosmol.3,139(1987)].
2. A. H. Guth and S. Y. Pi, *Phys. Rev. Lett.* **49** (1982) 1110.
3. A. A. Starobinsky, *JETP Lett.* **30** (1979) 682 [767(1979)].
4. D. Kazanas, *Astrophys. J.* **241** (1980) L59.
5. A. A. Starobinsky, *Phys. Lett.* **91B** (1980) 99 [771(1980)].
6. A. D. Linde, *Phys. Lett.* **108B** (1982) 389, [Adv. Ser. Astrophys. Cosmol.3,149(1987)].
7. A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48** (1982) 1220, [Adv. Ser. Astrophys. Cosmol.3,158(1987)].
8. J. D. Barrow and A. C. Ottewill, *J. Phys.* **A16** (1983) 2757.
9. S. K. Blau, E. I. Guendelman and A. H. Guth, *Phys. Rev.* **D35** (1987) 1747.
10. J. D. Barrow and A. Paliathanasis, *Phys. Rev.* **D94** (2016) 083518, arXiv:1609.01126 [gr-qc].
11. J. D. Barrow and A. Paliathanasis, *Gen. Rel. Grav.* **50** (2018) 82, arXiv:1611.06680 [gr-qc].
12. K. A. Olive, *Phys. Rept.* **190** (1990) 307.
13. A. D. Linde, *Phys. Rev.* **D49** (1994) 748, arXiv:astro-ph/9307002 [astro-ph].
14. A. R. Liddle, P. Parsons and J. D. Barrow, *Phys. Rev.* **D50** (1994) 7222, arXiv:astro-ph/9408015 [astro-ph].
15. C. Germani and A. Kehagias, *Phys. Rev. Lett.* **105** (2010) 011302, arXiv:1003.2635 [hep-ph].
16. T. Kobayashi, M. Yamaguchi and J. Yokoyama, *Phys. Rev. Lett.* **105** (2010) 231302, arXiv:1008.0603 [hep-th].
17. C.-J. Feng, X.-Z. Li and E. N. Saridakis, *Phys. Rev.* **D82** (2010) 023526, arXiv:1004.1874 [astro-ph.CO].

18. C. Burrage, C. de Rham, D. Seery and A. J. Tolley, *JCAP* **1101** (2011) 014, arXiv:1009.2497 [hep-th].
19. T. Kobayashi, M. Yamaguchi and J. Yokoyama, *Prog. Theor. Phys.* **126** (2011) 511, arXiv:1105.5723 [hep-th].
20. J. Ohashi and S. Tsujikawa, *JCAP* **1210** (2012) 035, arXiv:1207.4879 [gr-qc].
21. Y.-F. Cai, J.-O. Gong, S. Pi, E. N. Saridakis and S.-Y. Wu, *Nucl. Phys.* **B900** (2015) 517, arXiv:1412.7241 [hep-th].
22. V. Kamali, S. Basilakos and A. Mehrabi, *Eur. Phys. J.* **C76** (2016) 525, arXiv:1604.05434 [gr-qc].
23. D. Benisty and E. I. Guendelman, *Int. J. Mod. Phys.* **A33** (2018) 1850119, arXiv:1710.10588 [gr-qc].
24. I. Dalianis, A. Kehagias and G. Tringas, *JCAP* **1901** (2019) 037, arXiv:1805.09483 [astro-ph.CO].
25. I. Dalianis and G. Tringas, *Phys. Rev.* **D100** (2019) 083512, arXiv:1905.01741 [astro-ph.CO].
26. S. Weinberg, *Rev. Mod. Phys.* **61** (1989) 1.
27. L. Lombriser, *Phys. Lett.* **B797** (2019) 134804, arXiv:1901.08588 [gr-qc].
28. D. Merritt, *Stud. Hist. Phil. Sci.* **B57** (2017) 41, arXiv:1703.02389 [physics.hist-ph].
29. B. Ratra and P. J. E. Peebles, *Phys. Rev.* **D37** (1988) 3406.
30. R. R. Caldwell, R. Dave and P. J. Steinhardt, *Phys. Rev. Lett.* **80** (1998) 1582, arXiv:astro-ph/9708069 [astro-ph].
31. D. Benisty and E. I. Guendelman, *Phys. Rev.* **D98** (2018) 023506, arXiv:1802.07981 [gr-qc].
32. I. Zlatev, L.-M. Wang and P. J. Steinhardt, *Phys. Rev. Lett.* **82** (1999) 896, arXiv:astro-ph/9807002 [astro-ph].
33. R. R. Caldwell, *Phys. Lett.* **B545** (2002) 23, arXiv:astro-ph/9908168 [astro-ph].
34. T. Chiba, T. Okabe and M. Yamaguchi, *Phys. Rev.* **D62** (2000) 023511, arXiv:astro-ph/9912463 [astro-ph].
35. M. C. Bento, O. Bertolami and A. A. Sen, *Phys. Rev.* **D66** (2002) 043507, arXiv:gr-qc/0202064 [gr-qc].
36. S. Tsujikawa, *Class. Quant. Grav.* **30** (2013) 214003, arXiv:1304.1961 [gr-qc].
37. C. Wetterich, *Phys. Rev.* **D89** (2014) 024005, arXiv:1308.1019 [astro-ph.CO].
38. M. W. Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Phys. Rev.* **D90** (2014) 023512, arXiv:1402.6661 [gr-qc].
39. E. Guendelman, E. Nissimov and S. Pacheva, *Bulg. J. Phys.* **44** (2017) 015, arXiv:1609.06915 [gr-qc].
40. E. Guendelman, E. Nissimov and S. Pacheva, *Bulg. J. Phys.* **45** (2018) 152, arXiv:1709.03786 [gr-qc].
41. E. Guendelman, E. Nissimov and S. Pacheva, *Bulg. J. Phys.* **42** (2015) 249, arXiv:1505.07680 [gr-qc].
42. M. W. Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Phys. Rev.* **D90** (2014) 023512, arXiv:1402.6661 [gr-qc].
43. M. W. Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Phys. Lett.* **B737** (2014) 191, arXiv:1405.7491 [gr-qc].
44. M. Wali Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Int. J. Mod. Phys.* **D24** (2015) 1530014, arXiv:1410.6100 [gr-qc].
45. C.-Q. Geng, M. W. Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Phys. Rev.* **D92** (2015) 023522, arXiv:1502.03597 [gr-qc].

46. C.-Q. Geng, C.-C. Lee, M. Sami, E. N. Saridakis and A. A. Starobinsky, *JCAP* **1706** (2017) 011, arXiv:1705.01329 [gr-qc].
47. A. B. Kaganovich, *Phys. Rev.* **D63** (2001) 025022, arXiv:hep-th/0007144 [hep-th].
48. M. W. Hossain, R. Myrzakulov, M. Sami and E. N. Saridakis, *Phys. Rev.* **D89** (2014) 123513, arXiv:1404.1445 [gr-qc].
49. J. Martin, C. Ringeval and V. Vennin, *Phys. Dark Univ.* **5-6** (2014) 75, arXiv:1303.3787 [astro-ph.CO].
50. Planck Collaboration (N. Aghanim *et al.*) (2018) arXiv:1807.06209 [astro-ph.CO].
51. Planck Collaboration (Y. Akrami *et al.*) (2018) arXiv:1807.06211 [astro-ph.CO].
52. A. Lewis (2019) arXiv:1910.13970 [astro-ph.IM].
53. E. I. Guendelman, *Mod. Phys. Lett.* **A14** (1999) 1397, arXiv:hep-th/0106084 [hep-th].
54. E. I. Guendelman, *Mod. Phys. Lett.* **A14** (1999) 1043, arXiv:gr-qc/9901017 [gr-qc].
55. E. Guendelman, R. Herrera, P. Labrana, E. Nissimov and S. Pacheva, *Gen. Rel. Grav.* **47** (2015) 10, arXiv:1408.5344 [gr-qc].

Condensed light, quantum black holes and L-CDM cosmology: Experimentally suggested and tested unified approach to dark matter, dark energy, cosmogenesis and two-stage inflation

Victor Borsevici

International Informatization Academy, branch Rep. of Moldova, Chisinau, Rep. of Moldova
E-mail: borsevici@mail.ru

Based on the fundamental last discoveries (such as gravitational waves, Higgs field, gravitating Bose-Einstein photon condensates etc.), guided by the last A. Einstein's conjecture (about algebraic approach to the combined quantum and continuum description of reality) and directed by the J. Bekenstein-V. Mukhanov approach to the quantum black holes physics and astrophysics, the L. Susskind's "one photon – one bit" principle and V. Gribov-S. Hawking tunneling effect, one can easily find the simple but consistent algebraic relations and physical laws that govern birth and death of black holes in the Early Universe, its accretion and two-particle emitting (both outside black hole) accompanied with gravitational radiation.

From these basic relations in a natural way derive the Planck units system that confirm the theoretical consistence of presented approach and directly leads to the Planck Scale physics where standard models of the particle physics and cosmology are "blinded". Presented approach elucidates the physical background of dark matter (black holes), dark energy (gravitational waves), hot Big Bang and two-stage inflation of the Universe. Moreover, it solves many old unsolved problems, such as "cosmic censorship" conjecture, information loss paradox, black hole "end-point" problem, "cosmological constant" problem and so on.

The great importance for the validation of presented approach and it results had the comparison with "Planck-2018" data and other relevant information from the sources of observational cosmology, astrophysics, gravitational waves astronomy and numerical relativity.

Keywords: Planck scale physics, dark matter, dark energy, Bose–Einstein photon condensates, Cosmology.

To the blessed memory of Jacob Bekenstein and Steven Weinberg

Introduction

"Raffiniert ist der HerrGott, aber boshaft ist er nicht."
(A. Einstein)

What "banged"? What is dark matter? What is dark energy? What caused two-stage inflation? *etc.* Such problematic questions that provoked a lot of troubles for L-CDM cosmology are a consequence of unsolved problems in physics. L. Smolin called this stagnant situation "Einstein's unfinished revolution" and the more critical theoreticist S. Hossenfelder called it "Lost in math".

But over the last decade, experimental physics and observational cosmology have made many fundamental discoveries: gravitational waves (LIGO), Higgs boson (LHC), photon condensates with rest energy and rest mass trapped in "mirror cavities" (Bonn University).[1,2]

Through these remarkable results, Nature suggests that at pressures and temperatures well above the Higgs field level (246 GeV), only the 2-d and 3-d photon condensates trapped in their own "gravitational cavities" should be unique sources of gravity during the Early Universe. Moreover, we conclude that the real *"prima materia"* at the "beginning of the times" was a 3-dimensional Planck photonic condensate as "explosive" accompanied by Planck fluctuations as "fuse". This unambiguously postulates the fundamental role of the laws of conservation of energy and momentum in cosmology, as well as reflect the decisive role of quantum fluctuation in the formation of the modern Universe (see V. Mukhanov-G. Chibisov and S. Hawking – B. Carr approaches).

Note that after the discovery of gravitational waves, the famous Nobel Prize winner P. W. Anderson, in his prophetic "Four Last Conjectures" (2018), argues that "dark energy" is gravitational radiation which causes irreversible loss of mass in the Universe.

It is quite clear that in the swift initial process of Hot Bing Bang we have only three kinds of "actors" at the Planck scale: the 2-d spherical self-gravitating photonic condensates (primordial quantum black holes as "dark matter"), free hot photons and gravitational waves (both acting as massless stretched "dark energy" forces). Note that only gravitational radiation represents one irreversible repulsive force. According to our calculations, only a part of relic gravitational waves with Planck energies E_p and $E_p/2$ make up 93.38% of the dark energy in the modern Universe. As you can see, S. Weinberg was absolutely right: only the energy of Hot Bing Bang causes the universal expansion process.

Moreover, the irreversibility of the gravitational radiation energy prohibits all kind of "big crunch" and "bouncing" (cyclic) cosmological theories, proving instead the Hot Big Bang approach to the cosmogenesis.

We must point out separately that at the end of the initial phase of the Hot Big Bang process (at about $50\pi \times 5.391 \times 10^{-44}$ sec), when the strongest inequality $\Omega_M \ll \Omega_\Lambda$ was reached and stage I of hyperinflation began, a part of the surviving primordial black holes became "seed" of the primordial Cosmic Web, in which baryogenesis occurs and all astrophysical objects are formed.

Where do these conclusions follow? Taking into account the quantum nature of the 2-d spherical photon condensates, trapped in their own gravitational fields, guided by the J. Bekenstein's fundamental dependence between the Compton wavelength and the gravitational radius,[3] L. Susskind's principle "one photon — one bit",[4] the V. Mukhanov's quantum black holes "degeneracy" principle[5] and the modified V. Gribov–S. Hawking tunneling effect,[6,7] one can easily find any exact and adequate laws that govern birth and death, rise and downfall, accretion of light

and quantum two-particle emission (both outside black holes). All these processes are accompanied by the emission of gravitational waves.[8]

In what follows, we will demonstrate that the most important thing that distinguishes such natural black holes modelling from all other "membrane models" (see K. S. Thorne, F. Wilczek and M. Maggiore approaches) is that tremendous gravitational squeezed force is perfectly equilibrated by repulsive ("antigravitational") quantum-mechanical one.

Note that only the 2-d photon condensate "construct" with its tremendous "band" of possible wavelengths can explain the incredibly wide range of sizes, lifetimes and masses of black holes, ranging from the Planck scale to tens of Pluto orbits and tens of billions of masses of the Sun (see the quasar *TON 618*). We will demonstrate that this approach is free from the "cosmic censorship" principle, the "no hair" theorem and the information loss paradox.

Using this unified approach, we can calculate that at the end of the epoch of baryogenesis, when stage I of inflation ends, the rest energies of black holes (dark matter) and energy of gravitational waves (dark energy) are 28.62% and 66.42%, respectively. After this epoch, there came a non-accelerated expansion. But 6–8 billion years ago, the II stage of inflation began. Now, accordingly to "Planck-2018" data, we find the 26.57% and 68.47%, respectively. Comparison with the end of the stage I clearly shows that the increase in dark energy is caused by a decrease in the energy associated with dark matter. This leads to the unambiguous conclusion that stage II of inflation is provided by binary coalescences of black holes.

The greatest discoveries of L-CDM cosmologies are namely such epiphenomenons as "dark matter", "dark energy", "Hot Big Bang" and "two-stage inflation". It remains to find the deep physical background of these unsolved phenomenologies. The "Einstein's Unfinished Revolution" must be continued. However, it is quite clear that this is impossible without taking into account the new discoveries of modern physics and cosmology, including the self-gravitating Bose–Einstein photonic condensates. We must demonstrate the new approach to the Planck Scale Physics, where standard models of physics and cosmology was "blinded".

The new theoretical approach is consistent, when from it basic relations and equations directly derive any fundamental constants. In what follows, we will demonstrate that proposed basic equations generate the fundamental system of Planck units. This directly leads to the Terra Incognita of Planck Scale Physics and solves many puzzles of black hole physics, astrophysics and cosmology, such as "cosmic censorship" problem, information loss paradox, black hole "end-point" problem, "cosmological constant" problem, dark matter and dark energy problems and so on.

1. Principia

In 1916 A. Einstein wrote: "... das die Quantentheorie nicht nur die Maxwellische Elektrodynamik, ... auch die neue Gravitationstheorie wird modifizieren mussen".[9]

Already in those distant times, under the influence of Bohr's theory of hydrogen atom, he came to the conclusion that a spherical system should emit not only quanta of light, but also any portions of gravitational waves. This led him to believe that the new theory of gravity should be modified taking into account quantum electrodynamics.

Many years later, in 1954, in his last work, he wrote the closing lines that sound like a testament: "... a finite system with its finite energy can be described in full by a finite set of numbers (quantum numbers). Seemingly, it cannot be compliant with the continuum theory and requires a purely algebraic theory for a reality description. However now nobody knows, how to find a basis for such a theory".[10] Actually, A. Einstein explicitly clamed in his "last will" that just the algebraic approach must be the magistral way to unify quantum physics and continuum general relativity.

Only two decades later, it was the young Israeli physicist J. Bekenstein who turned out to be the "nobody" who found a way to do this. In his revolutionary work "Black Holes and Entropy"[3] he gave an amazing formula for the entropy of a black hole

$$S_s \text{ (Bekenstein)} = \frac{\ln 2}{8\pi} k c^3 \hbar^{-1} G^{-1} A_S \ ,$$

which unites all world constants: Boltzmann's constant (k), speed of light (c), reduced Plank constant ($\hbar$), gravitational constant (G), as well as the area of the Schwarzschild spherical singularity

$$A_S = 4\pi R_S^2 \ ,$$

where R_S is the Schwarzschild gravitational radius.

Bekenstein's formula was striking in that it linked together thermodynamics, quantum physics, the theory of gravity, and geometry. But the most surprising thing was that he connected all this within the framework of a unified information approach to the physics and astrophysics of black holes. In his famous formula, he associated the multiplier $\ln 2$ with Shannon's "bit". And this is not accidental, because he followed the famous principle of his great teacher J. A. Wheeler: "It from bit".

In 1986, the young Russian physicist V. Mukhanov, guided (as A. Einstein) by hydrogen atom analogy, came to the irrefutable conclusion that the emission mechanism of quantum black holes must obey the "degeneracy" principle.[11] He turns to J. Bekenstein and in 1995 their common outstanding work "Spectroscopy of the quantum black hole"[12] is published. In this work, it was shown that the true emission spectrum of quantum black holes *is discrete*.

With the advent of new fundamental experimental discoveries, this work of J. Bekenstein and V. Mukhanov began to acquire fundamental significance for physics and cosmology. Moreover, it turned out to be not just a "window" (as its authors was believed), but a whole "gateway" to *Terra Incognita* of Planck Scale Physics.

2. Basic Relations

With the discovery of 2-d photonic Bose-Einstein condensate[1] it became clear that the rejected Schwarzshild spherical singularity was becoming an obvious alternative to the black hole model in the form of a "point singularity" surrounded by an "event horizon" of Schwarzschild radius.

The biggest problem of all "membrane models" was their physical inconsistence. No one has been able to prove their resilience when facing the tremendous gravitational squeezed forces. From the "relativistic" point of view, this seemed to be an unsolvable problem. That is why, first of all, after justifying our approach, we will prove the strong equilibrium between squeezed gravitational forces and repulsive "antigravitational" quantum mechanical forces of quantum black hole. Note, that this strong equilibrium may be broken only by V. Mukhanov's "degeneracy" effect.

Based on the Bekenstein–Mukhanov approach, we will show how all these unsolved problems are solved in terms of the original basic system of equations for spherical 2-d photon condensates, trapped in their own gravitational field:

$$\begin{cases} \bar{\lambda}_{sm} = R_s \ , & (1) \\ \bar{\lambda}_{sm} = \dfrac{\hbar}{M_{sm}c} = \dfrac{\hbar c}{E_{sm}} \ , & (2) \\ R_s = \dfrac{2G}{c^2} M_s = \dfrac{2G}{c^4} E_s \ , & (3) \\ N_s = \dfrac{M_s}{M_{sm}} = \dfrac{E_s}{E_{sm}} \ , & (4) \\ N_s(min) = 2 \ . & (5) \end{cases}$$

Equations (1)–(3) bind together the reduced Compton wavelength, $\bar{\lambda}_{sm}$, which is associated to one photon with rest mass M_{sm} and rest energy E_{sm}, and Schwarzschild radius, R_s, which is expressed in terms of M_s (total rest mass of the condensate) and E_s (total rest energy).

Note that relation of the type of Eq. (1) was first time discovered by Bekenstein in his remarkable work Ref. 3 to determine the largest Compton wavelength of a particle absorbable by a black hole. In essence, general Eq. (1) can be called "Bekenstein relation". This is the only one equation known to us that organically interconnects Quantum Physics and General Relativity. In fact, it reflect the deepest equivalence between the "quantum wavelength geometry" ($\bar{\lambda}_{sm}$) and the "relativistic geometry" (R_s) that occurs in black holes as spherical 2-d photon condensates trapped in their own gravitational fields.

Equation (4) organically make links between the number of photons in the condensate, N_S, and the quantities expressing its "individual" rest mass, M_{sm}, and rest energy, E_{sm}, as well as with "total" rest mass, M_s, and rest energy, E_s. We note in particular that N_s is a natural integer, that reflect a quantum nature of black holes physics.

Equation (5) sets a natural limit on the minimum number of photons in the condensate, equal to two: $N_S(min) = 2$.

Note that for the evident lower boundary condition, system Eqs. (1)–(5) gives us (it is easy to verify) the following solutions:

$$M_s(min) = \sqrt{\frac{\hbar c}{G}} = M_p\,, \quad E_s(min) = \sqrt{\frac{\hbar c^5}{G}} = E_p\,, \quad R_s(min) = 2\sqrt{\frac{\hbar G}{c^3}} = 2l_p\,, \tag{6}$$

where M_p and E_p are the Planck mass and energy, respectively, l_p — Planck length.

It is striking that we unexpectedly found ourselves in the field of Planck Scale Physics and absolutely accurately described the Planck black hole without any preliminary assumptions, hypotheses and conjectures. We have used only the most basic information from Quantum Physics and General Relativity enclosed in the base system Eqs. (1)–(5).

Note that we do not use the "quantum gravity" approach based on the "gravitons" hypothesis because of the gravitational waves (the "ripples" of spacetime) radically differ from the quantum-mechanical waves (the "waves of probability").

3. Equilibrium

In order to make certain that a quantum black hole taken as spherical self-gravitating 2-d photon condensate is a quasi-stable astrophysical object, from all the solutions of the system Eqs. (1)--(5) we will pick the following one:

$$M_{sm} = \frac{\hbar}{cR_s} = \frac{\hbar c}{2GM_s} = \frac{M_p^2}{2M_s}\,, \tag{7}$$

$$E_{sm} = \frac{\hbar c}{\overline{\lambda}_{sm}} = \frac{\hbar c}{R_s} = \frac{\hbar c^5}{2GE_s} = \frac{E_p^2}{2E_s}\,. \tag{8}$$

As we can see, according to Birkhoff's Theorem the attractive gravitational force for one photon with rest mass M_{sm} is equal to

$$F_{sm,\,gr} = \frac{2GM_{sm}M_s}{R_s^2} = \frac{\hbar c}{R_s^2}\,. \tag{9}$$

But this attractive gravitational force must be perfectly equilibrated by quantum-mechanical "antigravitational" force:

$$F_{sm,\,qu} = \frac{\mathrm{d}E_{sm}\left(\overline{\lambda}_{sm}\right)}{\mathrm{d}\overline{\lambda}_{sm}} = \frac{\mathrm{d}\left({}^{\hbar c}/_{R_s}\right)}{\mathrm{d}R_s} = -\frac{\hbar c}{R_s^2}\,. \tag{10}$$

4. Quantum Information and Entropy: The Downfall of the "Point-Centered" Model

In his remarkable book "The Cosmic Landscape",[4] L. Susskind convincingly demonstrates that every photon is a carrier of one bit of quantum information. The point is that a photon can be in two spin states, $|0\rangle$ or $|1\rangle$.

In the case of a quantum black hole consisting of N_s photons, this means that the power of the set of its possible quantum states is 2^{N_s}. From the point of view

of information theory, a quantum black hole can be represented as a binary text consisting of N_s positions filled with symbols from the alphabet $\{|0\rangle, |1\rangle\}$. To each such text, θ_i, a probability, p_i, can be attributed. The maximum of C. E. Shannon's entropy is achieved for an uniform distribution $p_i = 2^{-N_s}$:

$$H_s \text{ (Shannon)} \underset{max}{=} -\sum_{i=1}^{2^{N_s}} 2^{-N_s} \log_2 2^{-N_s} = N_s \text{ (bit)} \ . \tag{11}$$

According to Shannon's approach, the maximum amount of quantum information, I_s, contained in a quantum black hole does not exceed this value. Moreover, it is equal to it:

$$I_s = H_s \text{ (Shannon)}_{\max} = N_s \text{ (bit)} \ . \tag{12}$$

Now let's calculate the thermodynamic entropy of the black hole. To do this, we first need to calculate the temperature of the quantum black hole from the relation

$$E_{sm} = 2 \times kT_s \ , \tag{13}$$

where kT_s is the energy corresponding to vibrational freedom degree, 2 is the multiplier due to the two spin states, $|0\rangle$ and $|1\rangle$. From Eq. (13) directly follows the wanted relationship for the temperature

$$T_s = \frac{E_{sm}}{2k} = \frac{E_p^2}{4kE_s} = \frac{\hbar c^3}{4kGM_s} \ . \tag{14}$$

It is easy to see that we have obtained an expression very close to the one of S. Hawking:

$$T_s \text{ (Hawkinhg)} = \frac{1}{2\pi} T_s \ . \tag{15}$$

However, the essential difference lies in the fact that the temperature T_s (Hawkinhg) is the typical temperature of the black hole taken as a "black body", whereas we are dealing with a quantum black hole. On the other hand, this similitude of the results suggests that we are acting in the right direction.

From the Clausius relation $\mathrm{d}S = \mathrm{d}E/T$ and Eq. (14), by integration, we easily find the required expression for the thermodynamic entropy:

$$S_s = \int_0^{E_s} \frac{4k}{E_p^2} E \, \mathrm{d}E = k2\frac{E_s^2}{E_p^2} = kN_s \ . \tag{16}$$

In other words, we get a direct connection between the entropy of a black hole and the Shannon's entropy:

$$S_s = kH_s \text{ (Shannon)}_{\max} = kN_s \ . \tag{17}$$

From Eqs. (3), (4) and (16) we can get another important relation:

$$S_s = kN_s = k\frac{R_s^2}{2l_p^2} = \frac{1}{8\pi} kc^3\hbar^{-1}G^{-1}A_s \ , \tag{18}$$

where A_s is the area of the Schwarzschild spherical singularity. This immediately implies a connection with the cited above Bekenstein formula:

$$S_s \text{ (Bekenstein)} = \ln 2 \cdot S_s \ . \tag{19}$$

The remarkable physicist S. Hawking relied on a model of a black hole in the form of a "point-singularity" surrounded by a "event-horizon". Actually, this means that in the center of symmetry of any black hole there should be matter with the limiting Planck density. Only 3-d Planck photon condensate can have such a density. It is not difficult to calculate the amount of quantum information, I_{s*}, in this model:

$$I_{s*} = N_{s*} = \frac{M_s}{M_p} \ . \tag{20}$$

For a 2-d model, as we already know, this quantity will be equal to

$$I_s = N_s = 2M_s^2/M_p^2 \ .$$

Let's compare these amounts of quantum information for a black hole of solar mass $M_\odot$:

$$I_s/I_{s*} = 2M_\odot/M_p = 1.827 \times 10^{38} \ (sic!) \ .$$

As we can see, in the light of modern discoveries, this difference in the amounts of quantum information, is simply monstrous. That clearly shows the invalidity of "point-model".

5. Natural Quantum Black Hole Description

We can continue J. Bekenstein–V. Mukhanov approach that in fact derive from the A. Einstein "testament" that was cited above.

From (4) and (5) we can conclude that the number of photons in quantum black hole taken as a spherical self-gravitating 2-d photon condensate must be presented in the following form:

$$N_s = 2n_s, \quad n_s = 1, 2, 3, \ldots \tag{21}$$

where n_s is *the principal quantum number*.

Then it is easy to show that all our previous results can be represented as follows:

$$\overline{\lambda}_{sm} = R_s = 2l_p\sqrt{n_s} \ ; \tag{22}$$

$$M_s = M_p\sqrt{n_s} \ , \qquad E_s = E_p\sqrt{n_s} \ ; \tag{23}$$

$$M_{sm} = M_p/2\sqrt{n_s} \ , \qquad E_{sm} = E_p/2\sqrt{n_s} \ ; \tag{24}$$

$$\omega_{sm} \text{ (ciclic frequence)} = E_{sm}/\hbar = \omega_p/2\sqrt{n_s} \ , \tag{25}$$

where $\omega_p = 1/t_p$ is the Planck frequency, $t_p = \sqrt{\hbar G/c^5}$ is the Planck time;

$$A_s = 16\pi l_p^2 n_s \ ; \tag{26}$$

$$I_s = N_s = 2n_s \text{ (bit)} \ ; \tag{27}$$

$$S_s = kI_s = 2kn_s \ ; \tag{28}$$

$$T_s = E_p/4k\sqrt{n_s} = T_p/4\sqrt{n_s} \ , \tag{29}$$

where $T_p = E_p/k$ is the Planck temperature;

$$\rho^I_{sh}\ (\text{surface information density}) = I_s/A_s = 1/8\pi l_p^2 = \frac{c}{\hbar\varkappa}\ \left(\frac{\text{bit}}{\text{m}^2}\right)\ , \tag{30}$$

where $\varkappa = 8\pi G/c^2$ is the Einstein constant;

$$g_{sh}\ (\text{surface gravity}) = G\frac{M_s}{R_s^2} = a_p/4\sqrt{n_s}\ , \tag{31}$$

$a_p = F_p/M_p = c/t_p$ is the Planck gravity *etc.*

Note, that this natural "emergence" of Planck Scale Physics, that was "automatically" arised from basics system Eqs. (1)–(5), clearly shows, that Planck Units System as such was organically and naturally emerge from experimental physics and observational cosmology. Moreover, as we see, the "great" unity between Quantum Physics and General Relativity was reached in natural way.

The famous physicist J. A. Wheeler once said: "Mass tells spacetime how to curve; spacetime tells mass how to move." Inspired by this example, we can coin a new one: "Light quanta tells spacetime how to quantize; spacetime tells light quanta how to curve."

6. Quantum Black Hole "Evaporation": Two-Particle Emission and Gravitational Waves Radiation

As shown above, quantum black hole has the property of strong equilibrium between squeezed gravitational force and repulsive ("antigravitational") quantum-mechanical force. This state of strong equilibrium can be broken only by the V. Mukhanov's degeneracy effect, which consists in spontaneous quantum jump from n_s state to $n_s - 1$ state. In our case this means that in all relations Eqs. (21)–(31) the argument n_s must be replaced by $n_s - 1$. It's obvious that quantum energy corresponding to the level $E_p\sqrt{n_s}$ must reduce to the one corresponding to the level $E_p\sqrt{n_s - 1}$. The difference

$$\Delta E_{s,s-1} = E_p\left(\sqrt{n_s} - \sqrt{n_s - 1}\right) = \frac{E_p}{2\sqrt{n_s}}\left(1 + \frac{1}{4n_s} + \frac{1}{8n_s^2} + ...\right) \tag{32}$$

is equal to the sum of radiance energy of two-particle emission

$$\Delta E_{s,s-1\,qu} = \frac{E_p}{2\sqrt{n_s}}$$

and gravitational waves radiation energy

$$\Delta E_{s,s-1\,gr} = \Delta E_{s,s-1} - \Delta E_{s,s-1\,qu} = \frac{E_p}{2\sqrt{n_s}}\left(\frac{1}{4n_s} + \frac{1}{8n_s^2} + ...\right)\ . \tag{33}$$

Note that energy of the quadrupolar mode is represented by term:

$$\left|\frac{\hbar}{c}\frac{\mathrm{d}g_{sh}(n_s)}{\mathrm{d}n_s}\right| = \frac{E_p}{2\sqrt{n_s}} \times \frac{1}{4n_s}\ . \tag{34}$$

From this result we easily understand that the next term ${}^{1}/_{8n_s^2}$ from the sequence in parentheses from Eq. (33) corresponds to octupolar mode and so on.

We come to the conclusion that in the parenthesis

$$\left(1+\frac{1}{4n_s}+\frac{1}{8n_s^2}+...\right)$$

of the main expansion of polynomial in Eq. (32), the first term refers to the dipolar mode, the second term — to the quadrupolar mode, the third — to the octupolar and so on. In other words, all the terms, except the first one, are directly related to the gravitational radiation of a quantum black hole. It only remains to find out the physical nature of the dipolar mode. For this, we turn to the data of experimental physics. In 2011, physicists from Chalmers University of Technology experimentally proved the existence of Casimir–Unruh dynamical effect in the squeezed vacuum excited by the "moving mirror" with frequency ω, that produce a pair of non-entangled photons with associated frequency $\omega/2$ each.[13]

By this experiment Nature suggests that in general case the energy $\Delta E_{s,s-1\,qu}$ of two-particle emission, $\left(X_s,\bar{X}_s\right)$, must be presented by the following form:

$$\Delta S_{s,s-1\,qu}=\hbar\omega_{sm}=\hbar\omega_{sx}+\hbar\omega_{s\bar{x}}=E\left(X_s\right)+E\left(\bar{X}_s\right)\ , \tag{35}$$

where ω_{sm} is the cyclic frequency Eq. (36),

$$\omega_{sx}=\omega_{s\bar{x}}={}^{\omega_{sm}}/_{2}\ ,\qquad E\left(X_s\right)=E\left(\bar{X}_s\right)={}^{E_{sm}}/_{2}\ , \tag{36}$$

where ω_{sx}, $\omega_{s\bar{x}}$ are the frequencies associated with emitted particles.

The following quantum black hole "degeneracy" diagram can illustrate all these results:

$$BH_s\longrightarrow BH_{s-1}+X_s+\bar{X}_s+\Gamma_s\ , \tag{37}$$

where Γ_s is the generated gravitational wave with energy Eq. (33).

Note that the pair $\left(X_s,\bar{X}_s\right)$ of non-entangled particles, depending on their energy, can be a pair of hot photons, this being especially important at energies well above the Higgs field level (246 GeV).

Remark 6.1. Using the expressions obtained above, it is easy to calculate the energy structure of quantum black hole at its "end-point" ($n_s=1$):

$$BH_p\longrightarrow 2\gamma_{p,0}+\Gamma_{p,0} \tag{38}$$

with energies:

$$E\left(BH_p\right)=E_p\sqrt{1}=E_p\ ,\qquad 2E\left(\gamma_{p,0}\right)={}^{E_p}/_{2\sqrt{1}}={}^{E_p}/_{2}\ , \tag{39}$$

$$E\left(\Gamma_{p,0}\right)=E_p\left(\sqrt{1}-\sqrt{1-1}-{}^{1}/_{2\sqrt{1}}\right)={}^{E_p}/_{2}\ . \tag{40}$$

As it can be observed, at the "end-point" of the existence of any black hole, all ends with a spectacular explosion with emission of half of energy, $\left({}^{E_p}/_{2}\right)$, as quantum electromagnetic radiation and another half, $\left({}^{E_p}/_{2}\right)$ — as gravitational

radiation. Note that these last two photons carry the last two bits of quantum information. As we can see, in the presented unified approach so called "end-point" problem is solved.

In general case we can find that the total energy radiated by a quantum black hole "evaporating" from the state n_s to its "end-point" is equal to

$$E_{s\,qu}\ (\text{rad}) = \sum_{n=1}^{n_s} E_p/2\sqrt{n}\ , \tag{41}$$

$$E_{s\,gr}\ (\text{rad}) = \sum_{n=1}^{n_s} E_p \left(\sqrt{n} - \sqrt{n-1} - \frac{1}{2\sqrt{n}}\right) = E_p\sqrt{n_s} - \sum_{n=1}^{n_s} \frac{E_p}{2\sqrt{n_s}}\ , \tag{42}$$

$$E_s\ (\text{total rad.}) = E_{s\,qu}\ (\text{rad}) + E_{s\,gr}\ (\text{rad}) = E\sqrt{n_s} = E_s\ . \tag{43}$$

Note that with increase of n_s the total gravitational radiation quickly tends to its maximal value E_p. As we see, the proportionality

$$^{E_{s\,gr}\ (\text{rad})}\!/_{E_{s\,qu}\ (\text{rad})} = {}^{1}\!/_{4n_s} + {}^{1}\!/_{8n_s^2} + ... \tag{44}$$

clearly shows the great importance of gravitational radiance when the temperatures of black holes are near to Planckian scale.

7. Real Mechanism of Quantum-Gravitational Radiation and Removal of "Information Loss" Problem

Above we have showed that, in accordance with the V. Mukhanov's degeneracy effect, the strong equilibrium of quantum black hole can be broken by quantum jump from n_s state to the $n_s - 1$ one. We have found, in Eq. (36), that each of the two emitted particles have the energy equal to $E_{sm}/2$. From this, we can easily calculate their non-reduced Compton wavelength:

$$\lambda_c\,(X_s) = \lambda_c\left(\bar{X}_s\right) = \frac{2\pi\hbar c}{E_{sm}/2} = 4\pi R_s\ . \tag{45}$$

If we compare this value with quantum black hole gravitational radius R_s, we find that these two particles will immediately "escape" from tremendous gravitational "well" of quantum black hole via the V. Gribov–S. Hawking tunneling effect. However, the "breaking news" is: this quantum tunneling effect of two emitted particles is accompanied with radiation of gravitational waves that play an important role in the Planck Scale Physics. Note that quantum black hole radiates two bit of quantum information during "jumping" process:

$$\Delta I_{s,s-1\,qu} = I_s - I_{s-1} = 2n_s - 2(n_s - 1) = 2\ (\text{bit})\ , \tag{46}$$

and these two bits were carried away by two "escaped" particles.

As we see above Eq. (52), the last, "remnant" Planck hole collapsed emitting the last two photons carrying the last two bits of quantum information. These

conclusions can represented by the strict mathematical relation:

$$I_s \text{ (radiated)} = \sum_{n=1}^{n_s} \Delta I_{n,n-1\,qu} = 2n_s = I_s \ , \tag{47}$$

that confirms the unitarity principle of quantum physics in the quantum black hole theory in the form of the law of quantum information conservation, that definitely solves the so-called "information loss" problem.

8. Entropy Production by Radiation

In conformity with Clausius relation, we can easily find the entropy production by quantum black hole radiation:

$$\begin{aligned} \Delta S_{s,s-1} = \frac{\Delta E_{s,s-1}}{T_s} = \frac{\Delta E_{s,s-1\,qu}}{T_s} + \frac{\Delta E_{s,s-1\,gr}}{T_s} = \\ = \Delta S_{s,s-1\,qu} + \Delta S_{s,s-1\,gr} = 2k + 2k\left(\frac{1}{4n_s} + \frac{1}{8n_s} + ...\right) \ . \end{aligned} \tag{48}$$

The quantum entropy term

$$\Delta S_{s,s-1\,qu} = 2k = k\Delta I_{s,s-1} \tag{49}$$

is closely connected with two bit of radiated information Eq. (46). But what represent the gravitational radiance term $\Delta S_{s,s-1\,gr}$? The form of the term

$$\Delta S_{s,s-1\,gr} = 2k\left(\frac{1}{4n_s} + \frac{1}{8n_s^2} + ...\right) \tag{50}$$

suggests that it is closely connected with quadrupolar, octupolar etc. modes of gravitational radiation that is the source of a dissipative and irreversible dark energy. We may call this term "gravitational waves entropy production".

It is interesting that famous physicist T. Jacobson in his important work "Non-equilibrium Thermodynamics of Spacetime" named a similar term "bulk viscosity entropy production term" and connected it with Ricci flow. But what is more fascinating: the famous mathematician Grisha Perelman presented his extraordinary work "The entropy formula for Ricci flow and its geometric application", in which argue that the entropy is an effective tool for mathematical investigation.

9. Dynamics: Radiance Power and Lifetime

Before the V. Mukhanov's "degeneracy" jump, quantum black hole must relax to gain the spherical form. Note that inside the 2-d photon condensate the "time flow" is the same as the "time flow" for external observer. This is not a deliberative assumption but a result of LIGO measurement in which the duration of ring-down phase is only a small part of a second. This is because in each point of the 2-d photon condensate all the tangential forces are equilibrated. The "stopped time-flow" effect take place only in the nearest vicinity, but not inside the "stretched horizon".

It is clear that the relaxation process needs a causal synchronism, which means that all the pairs of points located, one from another, at a distance of half a circle (πR_s) must be casually connected. This requires fulfillment of a lot of time at n_s state

$$\Delta t_s = \frac{\pi R_s}{c} = 2\pi t_p \sqrt{n_s} \; . \tag{51}$$

From this formula, we can easily calculate the lifetime of an "evaporated" quantum black hole:

$$\tau_s = \sum_{n=1}^{n_s} 2\pi t_p \sqrt{n} \; . \tag{52}$$

Using Bohr's correspondence principle, a very simple relation is found:

$$\tau_s = \int_0^{n_s} 2\pi t_p \sqrt{n} \, \mathrm{d}n = \frac{4\pi}{3} t_p n_s^{3/2} = \frac{4\pi}{3} t_p \left(\frac{E_s}{E_p} \right)^3 \; . \tag{53}$$

If we compare it with the S. Hawking's formula, we can see that

$$\tau_s = \frac{1}{3840} \tau_s \text{ (Hawking)} \; . \tag{54}$$

This big difference appears because S. Hawking associated black hole radiation with the blackbody radiation and with the Stefan-Boltzmann law, which means a continuous thermal spectrum. In contrast to this, following J. Bekenstein and V. Mukhanov, we have showed that this spectrum is discrete. Hence such a strong difference occurs.

Let's find the radiance power:

$$P_s = \frac{\Delta E_{s,s-1}}{\Delta t_s} \cong \frac{E_p}{4\pi t_p n_s} = \frac{\hbar c^6}{4\pi G^2 M_s^2} \; . \tag{55}$$

Comparing this with the S. Hawking's result, we find, respectively

$$P_s = 3840 P_s \text{ (Hawking)} \; . \tag{56}$$

The reason for such a large difference is the same.

We have a rare opportunity to qualitatively compare our results. It is known that the "baryonic epoch" lasted for about 400 thousand years. According to our calculations, in conformity with Eq. (53), the lifetime of the black hole, able of generating baryons, is about 240 thousand years, that quite "fits" into the mentioned interval of 400 thousand years. Whereas by S. Hawking model Eq. (54) it take as long as 920 millions (!) years, which is completely inconsistent with the data of observational cosmology. In what follows, we can find more details.

Moreover, such quantum black holes can produce pairs of ordinary neutrons via broken chiral symmetry that may solve the "old" problems of baryogenesis (followed by nucleosynthesis and leptogenesis) and matter–antimatter asymmetry in the Universe.

10. Dynamical Parameters of Particle Production

Whereas with him, the results obtained above can be easily extended to the processes of pair-production of any kind. To do this, it is enough to implement the condition for their generation

$$E_{s,s-1\,qu} \geqslant 2E_x \ , \tag{57}$$

where $E_x = m_x c^2$ is the rest energy of each particle, m_x being the rest mass. This condition is easily transformed into

$$E_s \leqslant \frac{E_p^2}{4E_x} \ . \tag{58}$$

The main result can be obtained by substituting Eq. (58) into Eq. (53):

$$\tau_s(x) \leqslant \frac{\pi}{48} t_p \left(\frac{E_p}{E_x}\right)^3 \ . \tag{59}$$

Condition Eq. (58) is easily modified into

$$M_s(x) \leqslant \frac{M_p^2}{4m_x} \ , \tag{60}$$

which determines the extremely large mass of a black hole able to produce a pair of particles $\left(X_s, \bar{X}_s\right)$ of a given type.

If it is known that, in the table of rest masses, a particle of type X_1 of mass m_{X1} is immediately followed by a particle of type X_2 of a larger mass m_{X2}, then we can determine the generation time of a particle of type X_1 by a black hole using the formula

$$\Delta\tau_1(X_1) = \tau_s(X_1) - \tau_s(X_2) \ , \tag{61}$$

whereas the total number of generated particles and antiparticles of the type X_1 is easily found from relation

$$N_s(X_1) = \frac{M_p^2}{8}\left(\frac{1}{m_{X1}^2} - \frac{1}{m_{X2}^2}\right) \ . \tag{62}$$

11. Birth and Rise of Quantum Black Holes at The Beginning of the Universe and at the Gravitational Collapse of Astrophysical Objects

"What I cannot create, I do not understand."
(R. Feynman)

The natural "suppression" of the Higgs field at the energies well above 246 GeV exclude all Feynman's diagrams that describe scattering processes in Standard Model. Even two photon scattering "box-diagram" do not work because it needs the presence of a non-zero Higgs field. Despite that quantum electromagnetic field do not

depends on Higgs field, two-photon scattering require intermediate virtual particle-antiparticle pair that is impossible when Higgs field is suppressed by energies of photons. Note that photons do not interact directly. Only gravitational field may save this situation but photons are massless particles. With one notable exception: photons do have rest masses in 2-d and 3-d photon condensates.

With these arguments, we conclude that Planckian black hole may be created only through the mediation of the gravitational field that acted in the 3-d Planck photon condensate. Note that a minimal photon condensate is represented by binary Planck system, (γ_p, γ_p), with two Planck photons in a close contact. These distinct binary Planck systems were formed when local homogeneity and isotropy were broken through the Planckian fluctuations The distance between the "centers" of this "fuzzballs" is equal to Planck length l_p. The total mass of Planck binary system is $2M_p$ and its rest energy is $2E_p$. The initial internal gravitational force is equal to the Planckian one:

$$F(\gamma_p, \gamma_p) = G\frac{M_p M_p}{l_p^2} = F_p \ . \tag{63}$$

This tremendous gravitational force provokes coalescence of this binary Planckian system. This tremendous Planckian coalescence inevitably generates a Planckian gravitational wave Γ_p that break the continuity of spacetime and forms a two-photon spacetime "buble". The energy of this Planckian gravitational wave is equal to Planckian work W_p:

$$E(\Gamma_p) = W_p = F_p l_p = E_p \ . \tag{64}$$

But what is the rest energy of this "buble"? This rest energy is equal to difference between initial rest energy of binary Planck system and energy of radiated Planck gravitational wave:

$$E(\gamma_p, \gamma_p) - E(\Gamma_p) = 2E_p - E_p = E_p \ .$$

In Sec. "2. Basic Relations" we have found that a minimal spherical 2-d photon condensate with two photons have Planck rest mass M_p and Planck rest energy E_p, Eq. (6). Note that this result was obtained without any presumptions and hypotheses about the importance of the Planckian Units System. We may say that it emerged in natural way starting from basic system of relations Eqs. (1)–(5) for quantum black holes. This means that the produced spacetime "buble" is nothing but the Planck black hole BH_p.

Thus, we can write the following fundamental scheme (diagram) of the birth of Planck black hole accompanied by Planck gravitational wave:

$$(\gamma_p, \gamma_p) \longrightarrow BH_p + \Gamma_p \tag{65}$$

with

$$E(BH_p) = E(\gamma_p, \gamma_p) - E(\Gamma_p) = E_p \ . \tag{66}$$

Expressions Eqs. (65) and (66) represent the fundamental conditions for the nucleation of Planck black holes, which, through accretion, can reach very significant sizes.

A very important question arises: where and when conditions of formation of Planck binary systems (γ_p, γ_p) could exist?

There are only two mutually reinforcing answers to this question:

- at the Beginning of the Universe, when *primordial* quantum black holes were created;
- in the nuclei of gravitationally collapsed astrophysical objects, such as neutron stars, where did the *secondary* quantum black holes emerged from.

Let us define the natural condition for accretion in the general form

$$\lambda_\gamma \leqslant \lambda_{sm} = 2\pi R_s \ , \tag{67}$$

which is equivalent to the condition

$$E_\gamma \geqslant E_{sm} \ . \tag{68}$$

Here E_γ is the energy of the absorbed photon, γ, and λ_γ is its wavelength.

Note that the velocity of accretion can be well above velocity of emission because it depends only on the intensity of captured photons flow.

Note one more very important circumstance: black holes consist of light and "feed" only on light. A clear understanding of the function of accretion disks and jets accompanying black holes immediately follows from this. Accretion discs perform the function of "thermal extraction" of light from the ordinary matter, while jets perform the functions of "evacuators" of rested plasma as a sort of "production wast". The tremendous two jets from *Sagittarius A**, which is called "Fermi bubbles", are the good illustrations and demonstration of these statements.

12. Cosmogenesis: Hot Big Bang and Pre-Inflation Stage

According to the law of conservation of energy, it accounts for an energy equal to the value of the energy of the modern Universe of the order of $10^{61}\, E_p$. Since in the future we will operate with energy densities, the exact value of this quantity is not important for us, only important being the fact that the total number of Planck photons in this condensate is tremendous large and has a value of the order of 10^{61}, that allows us to use in the following the theory of probabilities and the law of large numbers. At the same time, we will consider our diagrams as probabilistic events which should be connected in the casual and energetical sense.

Remark 12.1. Considering primordial Planck photon condensate trapped in its gravitational field as a *"prima materia"* at the "beginning of times" we automatically remove the synchronization problem, since the speed of physical processes

is the same in all domains of primordial condensate. Also, we are removing the homogeneity and isotropy problems due to the fact that the 3-d Planck condensate has the densest possible packing, close to the fuzzy spherical one, in which each Planck photon is densely "covered" by 12 other ones (interference is impossible due to the impossibility of exceeding the Planck energy density E_p/l_p^3). The "time flow" was homogenic and was determined by the Planckian time t_p. Note that it was this fact that caused the initial synchronization of all processes that occurred at the beginning of Hot Bing Bang.

Remark 12.2. The law of conservation of energy in the observable Universe guarantees us that it is the energy of primordial Planckian 3-d photon condensate trapped in its own gravitational field that cause all physical processes that occurred and will occur in the Universe.

Vacuum cannot serve as a *"perpetuum mobile"* of the Universe. All attempts to explain the so called "cosmology constant problem" by the cosmic vacuum energy was failed. However, it is the Planckian quantum fluctuations of vacuum in the primordial 3-d condensate that are the natural sources of "breaking" of its gravitational homogeneity and anisotropy.

We must consider one very important problem: not all Planckian photons in the primordial 3-d condensate can form the primordial binary Planckian systems (γ_p, γ_p). A part becames free.

The probability of such event to occur, due to its exceptional importance for cosmology, we will note as $\aleph$ (*alef*: "divine"). This probability can be computed using the relation

$$\aleph = (1 - \aleph)^{12} \ , \tag{69}$$

where $(1 - \aleph)^{12}$ is the probability for all its twelve surrounding photons to become non-free, *i.e.* to take part at events of the type Eq. (65). Solving the Eq. (69) we obtain

$$\aleph = 0.14745 \ldots . \tag{70}$$

If we use the presented above diagrams Eq. (65) as events from the point of view of probability theory and laws of large numbers we can easily find the following numbers at the beginning of the Universe as a beginning of the Hot Big Bang:

$$\Omega_L = \aleph \cong 0.1474 \ , \qquad \Omega_{BH} = \Omega_{GR} = \frac{(1 - \aleph)}{2} \cong 0.4263 \ , \tag{71}$$

where Ω_L — energy density of free light (free Planck photons, γ_p), Ω_{BH} — rest energy density of Planck black holes (BH_p), Ω_{GR} — energy density of Planck gravitational waves (Γ_p).

After about $50\pi t_p$ (*i.e.* $\sim 10^{-41}$ sec) the preinflation stage of Hot Big Bang is finished with following results:

$$\Omega_L \ = (7\aleph - 1)\sum_{n=1}^{7}\frac{1}{2\sqrt{n}} + (1 - 6\aleph)\sum_{n=1}^{9}\frac{1}{2\sqrt{n}} \cong 0.3358\,, \tag{72}$$

$$\Omega_{BH} \cong 0 \text{ (a very small value)}\,, \tag{73}$$

$$\Omega_{GR} = 1 - \Omega_L - \Omega_{BH} \cong 0.6642\ (66.42\%)\ . \tag{74}$$

Note that this irreversible density of energy (66.42%) remains practically unchanged till the beginning of II stage of inflation, when the binary collisions of black holes will become the source of the increase of this density (as we can see below).

From Eqs. (72)–(74) we find the strongest inequality

$$\Omega_{BH} \ll \Omega_L + \Omega_{GR}\ , \tag{75}$$

that denote the end of the pre-inflation stage of the Hot Big Bang and the beginning of the hyperinflation stage of the Universe.

The strong inequality Eq. (75) shows that at the end of the pre-inflation stage the weak energy of the squeezed gravitational forces were totally suppressed by the tremendous common energy of the quantum electromagnetic and gravitational radiation stretched forces (close to the whole energy of the Universe!).

Moreover, the very small value of Ω_{BH} is backed by the enormous absolute number of primordial black holes that were the "seeds" of the modern Cosmic Web.

As far as we know, all these results are the only physically substantiated confirmation of Hot Bing Bang and inflation theories.

In conformity with "Planck-2018" observational data[14] the energy density of dark energy Ω_Λ represents 0.6847 (68.47%) of the total energy of the Universe. This means that the relic gravitational waves emitted in the pre-inflation phases of Hot Big Bang represent

$$\frac{\Omega_{GR}}{\Omega_\Lambda} \times 100\% = \frac{0.6642}{0.6847} \times 100\% \cong 97\%\ (sic!) \tag{76}$$

of the nowadays dark energy.

Note that the only two kinds of primordial gravitational waves Γ_p and $\Gamma_{p,0}$ with E_p and $E_p/2$ energies that accompanied birth and death of Planck black holes taken together represents 93.39% of the nowadays dark energy.

These and others relic gravitational waves form the homogenic and anisotropic relic Cosmic Gravitational Waves Background (CGWsB) manifested only as the most of the irreversible and omnipresent dark energy shown in the "Planck-2018" data. Now these waves are of infinitesimal amplitudes and of frequency range of few kilohertz.

In fact, these results represent the veritable solutions of so called "cosmological constant problem",[15] dark matter and dark energy problems.

13. Two-stage Inflation

At the end of the baryogenesis, leptogenesis and nucleosynthesis epoch, when the stage I of inflation was ended, the remaining energy of ordinary matter reached the 4.96% of the total energy of the observable Universe (note that this value presented by "Planck-2018" data is unchanged in time). If we neglect the small energy density of light, we can determine the remaining energy density of black holes (dark matter) at the end of I-st stage of inflation:

$$\Omega_{BH}(\%) = 100\% - 4.96\% - 66.42\% = 28.62\% \, .$$

After this epoch, there came the non-accelerated stage of the expanse of the Universe. But 6–8 billions of years ago, the II-nd stage of small inflation began. Now, from "Planck-2018" data we know that the distribution of the energy in the nowadays Universe is roughly represented by the following numbers:

$$\begin{aligned} 4.96\% \quad & \text{(for baryonic matter)}, \\ 26.57\% \quad & \text{(for dark matter)}, \\ 68.47\% \quad & \text{(for dark energy)}. \end{aligned}$$

As we can see, a comparison with the end of the I stage of inflation clearly shows that the increase in the dark energy is caused by the decrease of the energy associated to dark matter. This leads to the unambiguous conclusion that the stage II of inflation is provided by the binary coalescences of black holes.

Conclusions

Presented approach to the Planck scale physics that governed the cosmological processes at the beginning of the Universe is based on the simple but consistent algebraic relations and schemes (diagrams) with such specific terms as Planck binary systems (γ_p, γ_p) quantum black holes BH_s, gravitational waves Γ_s *etc.* These physico-mathematical tools work in the area where the tools of standard models of physics and cosmology do not work because of tremendous temperatures and pressures that blocked the Higgs field activity. In a such terrible environment only 3-d photon condensates and 2-d spherical photon condensates (quantum black holes) are unique sources of gravity at the beginning of the Universe.

As it was showed above, presented results elucidate the decisive role of primordial black holes (as a dominated part of dark matter) in the generation of gravitational waves (a dominant part of dark energy) and clearly explain the physical background of Hot Big Bang and two-stage inflation. Moreover, these results solve the "old" unsolved problem of physics and cosmology such as "cosmic censure" problem, information loss paradox, black hole "end-point" problem, so called "cosmological constant" problem and so on.

From the theoretical point of view, presented approach derive from the last A. Einstein's conjecture about algebraic combined quantum and continuum description of reality and from the deepest principles and conjectures of such

distinguished physicists and cosmologists as J. Bekenstein, S. Weinberg, S. Hawking, V. Mukhanov, L. Susskind, G. 't Hooft, W. Unruh, T. Jacobson, K. S. Thorne, F. Wilczek, M. Maggiore., J. Peebles, B. Carr, P. W. Anderson, L. Smolin, C. Rovelly, F. Vidotto ("Planck Star" concept), A. Kashlinsky, S. Hossenfelder and many other eminent theorists. But the decisive role was played the theoretical consequences from the last fundamental discoveries of experimental physics and observational cosmology.

Acknowledgments

The author expresses his sincere gratitude to prof. Gr. Vereshchagin, prof. S. Ansoldi and prof. A. Pittelli for the opportunity to present these results at the "Early Universe" session of the 16-th Marcel Grossmann Meeting (Rome, 2021).

References

1. J. Klaers, J. Schmitt, F. Vewinger and M. Weitz, Bose-Einstein condensation of photons in an optical microcavity, *Nature* **468**, 545-548 (2010).
2. Eberhard E. Muller, Note on Bose-Einstein condensation of photons, *arXiv: 1801.05220 (quant-ph).*
3. J. D. Bekenstein, Black Holes and Entropy, *Phys Rev* **D7 No. 8**, 2333 (1973).
4. L. Susskind, *The Cosmic Landscape* (Little, Brown and Co. NY, 2006).
5. V. Mukhanov, Quantum Black Holes, *arXiv:1810.03525 (physics).*
6. L. V. Okun, "Vladimir Naumovich Gribov (1930-1991)", *Physics Today* **51, 3** 104 (1998). ("... Gribov insisted, long before Stephen Hawking proposal, that due to quantum tunneling, black holes must emit particles.")
7. S. W. Hawking, Particle creation by black holes, *Commun Math Phys* **43**, 199 (1975).
8. V. Borsevici, Black holes and Genesis of the Universe: Physics and Astrophysics of Spherical Photon Condensates in Their Own Gravitational Fields (Chisinau (Rep. Moldova): Stratum Plus, 2019), 179 p. (ISBN 978-9975-3198-5-0).
9. A. Einstein, Naherungsweise Integration der Feldgleichungen der Gravitation, *Sitzungsber preuss. Akad. Wiss* **1**, 688-696 (1916).
10. A. Einstein, *Relativistic Theory of the non-symmetric Field. The Meaning of Relativity.* 5th edn. (Princeton, 1955).
11. V. Mukhanov, Evaporation and Entropy of Quantized Black Hole, *Lebedev Institute preprint* **N 163** (1986).
12. J. D. Bekenstein, V. F. Mukhanov, Spectroscopy of the quantum black hole, *arXiv:gr-qc/9505012 (gr-qc).*
13. C. M. Wilson, G. Johansson, A. Pourkabirian, M. Simoen, J. R. Johansson, T. Duty, F. Nori and P. Delsing, Observation of the dynamical Casimir effect in a superconducting circuit, *Nature* **479**, 376-379 (2011).
14. Planck 2018 results. VI. Cosmological parameters. *arXiv:1807.06209 (astro-ph).*
15. S. Weinberg, The cosmological constant problem. *Reviews of Modern Physics* **Vol. 61, No. 1**, (January 1989).
16. L. Smolin, *The Trouble with Physics* (Houghton Mifflin, 2006).
17. S. Hossenfelder, *Lost in Math: How Beauty Leads Physics Astray.* (Basic Books, 2018).
18. A. Kashlinski *et al.*, Electromagnetic probes of primordial black holes as dark matter, *arXiv:1903.04424v2 (astro-ph.CO).*

Helical magnetic fields lead to baryogenesis

Ashu Kushwaha and S. Shankaranarayanan
Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India
E-mail: ashu712@iitb.ac.in, shanki@phy.iitb.ac.in

The origin of primordial magnetic fields and baryon asymmetry of the Universe are still unresolved issues and require physics beyond the standard models of cosmology and particle physics. Since both require physics beyond the standard model, there is a possibility that the same new physics can solve both. In this talk, I will discuss our model, where non-minimal coupling to the Riemann tensor generates sufficient primordial helical magnetic fields at all observable scales during inflation. Interestingly, the generation of helical magnetic fields leads to baryogenesis, and the model predicts the observed amount of baryon asymmetry of the Universe for a range of reheating temperatures consistent with the observations.

Keywords: Helical magnetic fields, baryogenesis, based on the works[1,2]

1. Introduction

The present Universe contains only matter and no antimatter, except for the rare antiparticles produced by cosmic rays. This asymmetry between baryons and antibaryons, also known as Baryon Asymmetry of the Universe (BAU), is characterized by the parameter $\eta_B \sim 10^{-10}$. Yet, we still do not completely understand the genesis of this asymmetry. However, in a remarkable paper in 1967, Sakharov[4] pointed out that to have observed baryon asymmetry following conditions must be satisfied by any theory: (1) baryon number violation (2) charge (C) and charge parity (CP) violation and (3) departure from thermal equilibrium.[6]

Observations from Faraday rotation and synchrotron radiation show the presence of micro-Gauss strength magnetic fields in the galaxies and the clusters of galaxies.[8] While the magnetic field measurements from Faraday rotation and synchrotron radiation provide upper bounds of the magnetic fields, the FERMI measurement of gamma-rays emitted by blazars provides a lower bound of the order of 10^{-15} G in intergalactic voids.[9] The origin of the magnetic field in these regions is still an open problem in modern cosmology; there is no compelling theoretical model to explain the generation of these large-scale magnetic fields.

The origin of primordial magnetic fields and the origin of the baryon asymmetry of the Universe are the unresolved issues in modern cosmology and particle physics models. Both require physics beyond the standard model and pose an exciting question— are these processes cosmological or particle physics or both? It seems impossible to generate the observed amount of baryon asymmetry within the Standard Model of particle physics framework. Since both require physics beyond the

standard model, there is a possibility that the same physics can solve both problems. This talk discusses a mechanism that leads to the generation of primordial helical magnetic fields and baryon asymmetry at the beginning of the radiation-dominated epoch.

The electromagnetic field has two transverse degrees of freedom, i.e., left and right circular polarization, associated with the left and right-handed helicity modes. Furthermore, modes with the same evolution (or dispersion relation) lead to the non-helical electromagnetic field. In contrast, differently propagating modes with non-zero net helicity imbalance give a helical magnetic field. In order to create this helicity imbalance we consider the following scenario: under parity transformation electric and magnetic fields transform as $\mathbf{E} \longrightarrow -\mathbf{E}, \mathbf{B} \longrightarrow \mathbf{B}$, since standard electromagnetic action, $F_{\mu\nu}F^{\mu\nu} = (B^2 - E^2)/2$ is parity invariant, but the Chern-Simon term $F_{\mu\nu}\tilde{F}^{\mu\nu} = -4\mathbf{E}\cdot\mathbf{B}$ is parity non-invariant. Therefore, introducing this term in the action will lead to the different evolution of the modes. There is an interesting quantity related to the $F_{\mu\nu}\tilde{F}^{\mu\nu}$ called magnetic helicity density, which in a given volume, is proportional to the difference between the number of left and right-handed photons, and is given by: $\mathcal{H}_M = \frac{1}{V}\int_V d^3x \mathbf{A}\cdot\mathbf{B}$, where $\mathbf{A}$ is electromagnetic vector potential. Helical magnetic fields are very interesting for the following reasons: (1) the decay rate to energy density and coherence length is slow for these fields, which differs from the non-helical, (2) helical magnetic fields leave a very distinct signature as they violate parity symmetry which leads to observable effects, for example, correlations between the anisotropies in the temperature and B-polarisation or the E- and the B-polarisations in the CMB, and (3) if we could measure helical magnetic fields, it would provide evidence of CP violation in the early Universe.

It is important to note that both kinds of magnetogenesis models, i.e., early time and late time models, have problems. Inflation provides a causal mechanism to generate the magnetic fields of large coherence length. The problem with magnetic field generation during inflation is that standard electromagnetic action is conformally flat. Hence, the inflationary mechanism can not amplify quantum fluctuations of electromagnetic fields. Therefore one needs to break the conformal symmetry of the action. In the literature, models breaking conformal invariance by introducing non-minimal coupling terms in the electromagnetic action are proposed or by the time-dependent coupling with the gauge field.[8] These models produce non-helical magnetic fields. On the other hand, the generation of helical magnetic fields baryogenesis in the early Universe requires a parity-violating source.

A more popular scenario of helical field generation (and baryogenesis) is by adding the interaction term $f(\phi)F_{\mu\nu}\tilde{F}^{\mu\nu}$ which is responsible for creating the imbalance between left-handed and right-handed photons.[10] However, these scalar field coupled models suffer from strong coupling and backreaction problems and sensitivity to the reheating dynamics. In this talk, I will discuss our model where the electromagnetic field is coupled to the Riemann curvature, which breaks the conformal invariance and parity symmetry of the theory and leads to the production

of primordial helical fields at all length scales. Interestingly, as we will see that the same helical modes will give a non-zero contribution to Chern-Simon (CS) number density and hence lead to the baryogenesis.

1.1. *Baryogenesis in the presence of primordial helical magnetic field*

Almost two decades back, Davidson[3] pointed out some very interesting relation between primordial magnetic field and Sakharov's conditions of the baryon asymmetry of the Universe. The presence of background magnetic fields in the early Universe may break the $C, CP, SO(3)$ and thermal equilibrium. Specifically, she argued that there are symmetries that are broken due to the presence of magnetic field: (1) there should be some kind of out-of- thermal-equilibrium dynamics because in equilibrium, the photon distribution is thermal, and there are no particle currents to sustain a "long-range" field, (2) because $\mathbf{B}$ is odd (changes sign) under CP, the presence of magnetic field will lead to CP violation, (3) since the magnetic field is a vector quantity, it chooses a particular direction hence breaks the isotropy (rotational invariance). This implies that one could see the presence of a magnetic field to provide the CP violation and out of thermal equilibrium conditions required in baryogenesis models. However, Davidson's conditions are necessary but not sufficient. One key missing ingredient is the requirement of *primordial helical magnetic fields* which can provide evidence of CP violation in the early Universe if we could measure them. Interestingly, as we show next, the presence of primordial helical fields leads to the non-zero Chern-Simons number and, eventually, the change in the Fermion number.

In the early Universe, the generation of the non-zero primordial helical magnetic fields leads to a chiral anomaly resulting from the imbalance between left and right-handed fermions. In the presence of an electromagnetic field in curved space-time, the chiral anomaly is given by the following equation[5]:

$$\nabla_\mu J^\mu_A = -\frac{1}{384\pi^2}\epsilon^{\mu\nu\rho\sigma} R_{\mu\nu\alpha\beta} R^{\alpha\beta}{}_{\rho\sigma} + \frac{e^2}{16\pi^2}\epsilon^{\mu\nu\alpha\beta} F_{\mu\nu} F_{\alpha\beta} \tag{1}$$

where J^μ_A is the chiral current, $R_{\rho\sigma}{}^{\alpha\beta}$ is the Riemann tensor and A_μ is the four-vector potential of the electromagnetic field, $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$ and $\epsilon^{\mu\nu\rho\sigma} = \frac{1}{\sqrt{-g}}\eta^{\mu\nu\rho\sigma}$ is a fully antisymmetric tensor, and we set $\eta^{0123} = 1 = -\eta_{0123}$.

In the case of flat FRW background in conformal time (η):

$$ds^2 = a^2(\eta)\left(d\eta^2 - \delta_{ij} dx^i dx^j\right), \tag{2}$$

the first non-zero contribution of the first term in the RHS of Eq. (1) comes only at second order and hence will not play any role in our analysis. However, due to the presence of the magnetic fields, the second term in the RHS of Eq.(1) is the non-zero only and hence leads to a net chiral current. Thus, if we consider only

up to the first-order in perturbations, *only* the second term in the RHS of Eq. (1) contributes, and the chiral anomaly equation reduces to:

$$\partial_\mu \left(\sqrt{-g} J^\mu_A\right) = \frac{e^2}{4\pi^2} \partial_\mu \left(\sqrt{-g} K^\mu\right) \tag{3}$$

where

$$K^\mu = \frac{\eta^{\mu\nu\alpha\beta}}{\sqrt{-g}} A_\nu \partial_\alpha A_\beta$$

is the topological current. Using Eq. (6), the net baryon number density, $n_B = a(\eta)\langle 0|J^0_A|0\rangle$ is related to Chern-Simon number density $n_{CS} = \langle 0|K^0|0\rangle$ as,[7]

$$n_B \equiv \frac{e^2}{4\pi^2} a(\eta) n_{CS}\,, \tag{4}$$

where

$$n_{CS} = \frac{1}{a^4} \epsilon_{ijk} \langle 0|A_i\, \partial_j A_k|0\rangle = \frac{1}{a^4} \int_\mu^\Lambda \frac{dk}{k} \frac{k^4}{2\pi^2} \left(|A_+|^2 - |A_-|^2\right)\,, \tag{5}$$

where Λ, and μ set the possible energy range (or epoch) of the baryogenesis after inflation, and $A_\pm$ refer to the positive and negative modes of the helical magnetic fields which are generated during inflation. Note that due to the absence of standard model particles $n_B = 0$ during inflation. As we can see from Eq.(8) that all the modes re-entering the horizon at the beginning of the radiation-dominated epoch contribute significantly, and hence the value of n_{CS} depends on the upper cut-off Λ. It is important to note that n_{CS} vanishes if the primordial magnetic fields are non-helical, i. e. $|A_+| = |A_-|$, and hence will not lead to baryogenesis. Thus, the key missing ingredient of Davidson's conditions is the requirement of helical magnetic fields. non-zero only and hence leads to a net chiral current. Thus, if we consider only up to the first-order in perturbations, *only* the second term in the RHS of Eq. (1) contributes, and the chiral anomaly equation reduces to:

$$\partial_\mu \left(\sqrt{-g} J^\mu_A\right) = \frac{e^2}{4\pi^2} \partial_\mu \left(\sqrt{-g} K^\mu\right) \tag{6}$$

where

$$K^\mu = \frac{\eta^{\mu\nu\alpha\beta}}{\sqrt{-g}} A_\nu \partial_\alpha A_\beta$$

is the topological current. Using Eq. (6), the net baryon number density, $n_B = a(\eta)\langle 0|J^0_A|0\rangle$ is related to Chern-Simon number density $n_{CS} = \langle 0|K^0|0\rangle$ as,[7]

$$n_B \equiv \frac{e^2}{4\pi^2} a(\eta) n_{CS}. \tag{7}$$

where

$$n_{CS} = \frac{1}{a^4} \epsilon_{ijk} \langle 0|A_i\, \partial_j A_k|0\rangle = \frac{1}{a^4} \int_\mu^\Lambda \frac{dk}{k} \frac{k^4}{2\pi^2} \left(|A_+|^2 - |A_-|^2\right)\,, \tag{8}$$

where Λ, and μ set the possible energy range (or epoch) of the baryogenesis after inflation, and $A_\pm$ refer to the positive and negative modes of the helical magnetic fields which are generated during inflation. Note that due to the absence of standard model particles $n_B = 0$ during inflation. As we can see from Eq.(8) that all the modes re-entering the horizon at the beginning of the radiation-dominated epoch contribute significantly, and hence the value of n_{CS} depends on the upper cut-off Λ. It is important to note that n_{CS} vanishes if the primordial magnetic fields are non-helical, i. e. $|A_+| = |A_-|$, and hence will not lead to baryogenesis. Thus, the key missing ingredient of Davidson's conditions is the requirement of helical magnetic fields.

2. Generation and evolution of primordial helical magnetic field

We consider the following action:

$$S = -\frac{M_{\rm P}^2}{2}\int d^4x\sqrt{-g}\,R + \int d^4x\sqrt{-g}\left[\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right] - \frac{1}{4}\int d^4x\,\sqrt{-g}\,F_{\mu\nu}F^{\mu\nu} - \frac{\sigma}{M^2}\int d^4x\,\sqrt{-g}\,R_{\rho\sigma}{}^{\alpha\beta}F_{\alpha\beta}\,\tilde{F}^{\rho\sigma} \tag{9}$$

where $R_{\rho\sigma}{}^{\alpha\beta}$ is the Riemann tensor, A_μ is the four-vector potential of the electromagnetic field, $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$ and $\tilde{F}^{\rho\sigma} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}$ is the dual of $F_{\mu\nu}$. $\epsilon^{\mu\nu\rho\sigma} = \frac{1}{\sqrt{-g}}\,\eta^{\mu\nu\rho\sigma}$ is fully antisymmetric tensor, $\eta^{\mu\nu\rho\sigma}$ is Levi-Civita symbol whose values are ±1 and we set $\eta^{0123} = 1 = -\eta_{0123}$.

Note that in Eq. (9), the first three terms are the standard terms (the Einstein-Hilbert action, scalar field action, and standard electrodynamics, respectively). However, the presence of the Riemann tensor term breaks the conformal (and parity) invariance of the action. We assume that the scalar field (ϕ) dominates the energy density in the early Universe (during inflation) and leads to $60 - 70$ e-foldings of inflation with $H \simeq 10^{14}$GeV. M is the energy scale, which sets the scale for the breaking of conformal invariance. We assume that $10^{-3} \leq (H_{\rm Inf}/M) \leq 1$ where $H_{\rm Inf} \sim 10^{14}$ GeV is the Hubble scale during inflation. It is important to note that due to Riemann coupling, M appears as a time-dependent coupling in the FRW background i.e. $\frac{1}{M_{\rm eff}} \sim \frac{1}{M}\frac{a'}{a^2} = \frac{H}{M}$. Using $H_0 \approx 10^{-44}$GeV and $M \approx 10^{17}$GeV, we can see that at current epcoh $\frac{H_0}{M} \sim 10^{-59}$. Therefore, the Riemann coupling is tiny, and the non-minimal coupling term in the electromagnetic action will have a significant contribution only in the early Universe.

Our model has the following salient features: First, our model does not require the coupling of the electromagnetic field with the scalar field. Hence, there are no extra degrees of freedom, which will not lead to a strong-coupling problem. Second, the conformal invariance is broken due to the coupling to the Riemann tensor. Since the curvature is significant in the early Universe, the coupling term will introduce non-trivial corrections to the electromagnetic action. However, as mentioned above, the Riemann coupling term will not contribute at late-times, and the model is

identical to standard electrodynamics. Third, as we show explicitly, our model is free from backreaction for a range of scale-factor during inflation. This is different from other models where a specific coupling function is chosen to avoid any back-reaction.[10]

2.1. *Estimating the strength of the helical magnetic fields*

We estimate the total electromagnetic energy density at the horizon exit and to identify whether the modes lead to back-reaction on the metric, we define the quantity, $R = \frac{(\rho_B+\rho_E)|_{\mathcal{H}\sim k_*}}{6H^2M_{\rm Pl}}$, which is the ratio of the total energy density of the fluctuations at horizon exit and background energy density during inflation. Since $R << 1$ for any $\alpha \leq -\frac{1}{2}$ ($\alpha = -1/2$ corresponds to de-Sitter universe), the model does not have back-reaction.

α	$\rho_{\rm total}$ (in GeV^4)	R
$-\frac{1}{2}-\epsilon$	$\sim 10^{64}$	$\sim 10^{-4}$
$-\frac{3}{4}$	$\sim 10^{62}$	$\sim 10^{-6}$
-1	$\sim 10^{61}$	$\sim 10^{-7}$
-3	$\sim 10^{59}$	$\sim 10^{-9}$

To estimate the current value of the helical fields, we assume instantaneous reheating and the fact that Universe becomes radiation-dominated after inflation. Due to flux conservation, the magnetic energy density will decay as $1/a^4$, i.e. $\rho_B(0) = \rho_B^{(f)}\left(\frac{a_f}{a_0}\right)^4$ where a_0 is the present day scale-factor, $\rho_B^{(f)}$ and a_f refer to the magnetic energy density and the scale-factor at the end of inflation, respectively. Using the fact that the relevant modes exited Hubble radius around 30 e-foldings of inflation, with energy density $\rho_B \approx 10^{64}$GeV^4, the primordial helical fields at GPc scales is: $B_0 \approx 10^{-20}$G where we have used $1G = 1.95 \times 10^{-20}$GeV^2. Our model predicts the following primordial helical fields that re-entered the horizon at two different epochs: $B|_{50\ \rm MPc} \sim 10^{-18}\ G$ ($z \sim 20$) and $B|_{1\ \rm MPc} \sim 10^{-14}\ G$ ($z \sim 1000$). Thus, the model generates sufficient primordial helical magnetic fields at all observable scales. Also, our model predicts the present-day helical magnetic field of strength 10^{-15}G on Mpc scales.

3. Baryon asymmetry of the Universe

Now, we compute the baryon asymmetry created by the helical modes generated during inflation. To do that, we need to obtain the ranges of Λ and μ. The largest scales observed in the CMB are produced around 40 - 60 e-foldings before the end of inflation, and hence we have considered these modes for the generation of the helical magnetic field. However, taking advantage that modes are generated at all length scales in our model, we focus on the helical fields that renter the horizon very early

(at the beginning of the radiation-dominated epoch) to generate the required BAU. This means that the modes that left the horizon around the last 5 to 10 e-foldings of inflation are only relevant.

Since the entropy density per comoving volume is conserved, the quantity n_B/s is better suited for theoretical calculations. Assuming that there was no significant entropy production after reheating phase, entropy density in the radiation-dominated epoch is:

$$s \simeq \frac{2\pi^2}{45} g\, T_{\rm RH}^3\,, \tag{10}$$

where $T_{\rm RH}$ is the reheating temperature and the effective relativistic degrees of freedom $g \sim 100$ at reheating. The BAU parameter is given by:

$$\eta_B = \frac{n_B}{s} \approx 10^{-29}|C|^2 \frac{\Lambda^3}{T_{\rm RH}^3} \approx 10^{-2} \left(\frac{M}{M_{\rm Pl}}\right)^3 \left(\frac{\Lambda}{T_{\rm RH}}\right)^3 \tag{11}$$

where $|C| \approx \frac{M^{3/2}\eta_0}{\sqrt[4]{\eta_{end}10^{45}GeV^3}}$ and $M_{\rm Pl} \approx 10^{19}$GeV. Assuming the exit of inflation at 10^{14}GeV, and for the observed baryon asymmetry 10^{-10}, the reheating temperature $T_{\rm RH}$ is in the range 10^{12}—10^{14} GeV, which is is consistent with the constraints on the reheating temperature. This suggests that our model does not prefer a very low-energy reheating temperature

4. Conclusions

This talk focuses on a viable scenario for the generation of helical magnetic fields during inflation which lead to the baryogenesis in the early Universe. The model does not require coupling to the scalar field and hence does not lead to a strong coupling problem, and also, there is no backreaction on the background dynamics. The generation of the helical fields is due to the coupling of the electromagnetic fields with the dual of the Riemann tensor, which breaks the parity symmetry of the action. Interestingly, our model generates the present-day magnetic field of strength $10^{-15}G$ Mpc scales.

We have explicitly shown that Davidson's conditions are necessary *but not* sufficient. The key missing ingredient is the requirement of *primordial helical magnetic fields*. We have shown that the BAU parameter predicted by our model is independent of any specific inflation model and reheating dynamics; however, it depends on the scale at which inflation ends and reheating temperature. Furthermore, focussing on the modes that leave the horizon around 5 to 10 e-foldings the BAU parameter obtained in our model is inversely proportional to reheating temperature. Assuming the exit of inflation at 10^{14} GeV, for the observed amount of baryon asymmetry $\eta_B \sim 10^{-10}$, we obtained that the reheating temperature should be in the range $10^{12} - 10^{14}$ GeV, which is consistent with the constraints on the reheating temperature.

Acknowledgments

The MHRD fellowship at IIT Bombay financially supports AK. This work is supported by the ISRO-Respond grant.

References

1. Ashu Kushwaha and S. Shankaranarayanan. Helical magnetic fields from riemann coupling. *Phys. Rev. D*, 102:103528, Nov 2020.
2. Ashu Kushwaha and S. Shankaranarayanan. Helical magnetic fields from Riemann coupling lead to baryogenesis. *Phys. Rev. D*, 104(6):063502, 2021.
3. Sacha Davidson. Ingredients and equations for making a magnetic field in the early universe. *Phys. Lett. B*, 380:253–256, 1996.
4. A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Sov. Phys. Usp.*, 34(5):392–393, 1991.
5. Leonard E. Parker and D. Toms. *Quantum Field Theory in Curved Spacetime: Quantized Field and Gravity*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 8 2009.
6. Antonio Riotto and Mark Trodden. Recent progress in baryogenesis. *Ann. Rev. Nucl. Part. Sci.*, 49:35–75, 1999.
7. Neil D. Barrie and Archil Kobakhidze. Inflationary Baryogenesis in a Model with Gauged Baryon Number. *JHEP*, 09:163, 2014.
8. Kandaswamy Subramanian. The origin, evolution and signatures of primordial magnetic fields. *Rept. Prog. Phys.*, 79(7):076901, 2016.
9. Andrii Neronov and Ievgen Vovk. Evidence for strong extragalactic magnetic fields from fermi observations of tev blazars. *Science*, 328(5974):73–75, 2010.
10. Ramkishor Sharma, Kandaswamy Subramanian, and T. R. Seshadri. Generation of helical magnetic field in a viable scenario of inflationary magnetogenesis. *Phy. Rev. D*, D97(8):083503, 2018.

Polymer Quantization of the Isotropic Universe: Comparison with the Bounce of Loop Quantum Cosmology

G. Barca*, E. Giovannetti and F. Mandini

Department of Physics, La Sapienza University of Rome, Rome, Italy
**Speaker. E-mail: gabriele.barca@uniroma1.it*

G. Montani

Department of Physics, La Sapienza University of Rome, Rome, 00185 Italy
Fusion and Nuclear Safety Department, ENEA, Frascati (RM), Italy

We implement Polymer Quantum Mechanics on the Hamiltonian formulation of the isotropic Universe in both the representations of the standard Ashtekar-Barbero-Immirzi connection and of a new generalized coordinate conjugate to the Universe volume. The resulting dynamics is a bouncing cosmology; when quantizing the volume-like variable the Big Bounce is an intrinsic cut-off on the cosmological dynamics, while when using the standard connection the Bounce density results to be dependent on the initial conditions of the prepared wave packet. Then we compare the nature of the resulting Bounce with what emerges in Loop Quantum Cosmology, where the dependence of the critical density on the initial conditions is present when the minimum area eigenvalue is implemented in a comoving representation instead of the physical one. We conclude that, if one hand the preferable scenario should be a Big Bounce whose density depends on initial conditions in view of the privileged SU(2) character that the Ashtekar-Barbero-Immirzi connection possesses in the full Loop Quantum Gravity, on the other hand the equivalence demonstrated in the context of polymer cosmology can be a hint in favour of the viability of the improved scheme of Loop Quantum Cosmology even though it is not expressed through the privileged set of variables.

Keywords: Quantum Cosmology; Polymer Quantum Mechanics; Early Universe.

1. Introduction

Loop Quantum Cosmology (LQC)[1–5] is the implementation of Loop Quantum Gravity (LQG)[6] in the symmetry-reduced cosmological minisuperspaces. It predicts the emergence of a Bouncing scenario that removes the classical singularity, but has the intrinsic limitation that the $SU(2)$ symmetry of LQG is lost.

Here we address the nature of the Bounce through the implementation of Polymer Quantum Mechanics (PQM)[7] on the flat isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) model in two different representations: the Ashtekar-Barbero-Immirzi variables (the basic $SU(2)$ variables of LQG and LQC) and the volume variables.[8] Then, we compare our results with LQC in order to gain some insight on both frameworks. A more detailed analysis is presented in Ref. 9.

This work is organized as follows: in section 2 we introduce the framework of PQM; then we implement PQM on the FLRW model firstly expressed in the Ashtekar variables in section 3, and then in the volume variables in section 4; in

section 5 we compare our results with LQC; finally in section 6 we summarize our work. We use the natural units $\hbar = c = 8\pi G = 1$.

2. Polymer Quantum Mechanics

PQM[7] is an alternative representation that is non-unitarily connected to the standard Schrödinger representation. It implements a fundamental scale in the Hilbert space through the introduction of a lattice structure. Its aim is to reproduce effects similar to those of LQG through an independent framework that allows for more freedom in the choice of the preferred configurational variable and that is easily and reliably applicable to many Hamiltonian systems.

Let us consider a generic Hamiltonian system with canonical variables (Q, P). The kinematical framework of PQM consists in assigning to one of the variables, usually the position Q, a discrete character. As a consequence, in the quantization procedure, the conjugate momentum P cannot be promoted to a well-defined operator because the translational operator $\hat{T}(\lambda) = e^{i\lambda P}$ is not weakly continuous.

The procedure to regularize the momentum P and to construct a well-defined operator consists in the introduction of a lattice on position Q with constant spacing $\beta_0 \in \mathbb{R}$. The Hilbert space is then constructed as the one that contains all those states $|\psi\rangle = \sum_i b_i \, |\beta_n\rangle$, with $\beta_n = n\beta_0$ and $\sum_i |b_i|^2 < \infty$. Now the translational operator must be restricted to act only by discrete steps to remain on γ_{β_0}: this is achieved simply by setting $\zeta = \beta_0$ in order to obtain $\hat{T}(\beta_0) \, |\beta_n\rangle = |\beta_{n+1}\rangle$.

When the condition $P\beta_0 \ll 1$ is satisfied, we can approximate P with a sine function, so that the regulated momentum operator can be constructed as the incremental ratio of the translational operator on the lattice:

$$P \approx \frac{\sin(\beta_0 P)}{\beta_0} = \frac{e^{i\beta_0 P} - e^{-i\beta_0 P}}{2i\beta_0}, \tag{1a}$$

$$\widehat{P_{\beta_0}} \, |\beta_n\rangle = \frac{\hat{T}(\beta_0) - \hat{T}(-\beta_0)}{2i\beta_0} \, |\beta_n\rangle = \frac{|\beta_{n+1}\rangle - |\beta_{n-1}\rangle}{2i\beta_0}. \tag{1b}$$

For the squared momentum operator there are infinitely many approximations, but they usually amount to just a rescaling of the lattice parameter β_0; the simplest one in order to avoid discrepancies is obtained by composing the operator (1b) with itself, corresponding to the approximation $P^2 \approx \sin^2(\beta_0 P)/\beta_0^2$.

When performing the quantization of a system using the momentum polarization of the polymer representation, the regulated momentum operator must be used together with the differential coordinate operator. Alternatively, it is possible to perform a semiclassical analysis by using the formal substitution (1a) in the classical Hamiltonian, thus including quantum modifications in the classical dynamics.[8,10,11]

3. Polymer FLRW Dynamics in Ashtekar Variables

We will now show the dynamics of the flat isotropic FLRW model expressed in the Ashtekar variables. Given the symmetries of the model, the connection and

the triads reduce to a very simple form: $A^i_a \to c = \gamma\dot{a}$, $E^a_i \to p$, with $|p| = a^2$, $\{c, p\} = \gamma/3$, where $a = a(t)$ is the scale factor of the Universe, the dot represents a derivative with respect to time t and γ is the Immirzi parameter.

3.1. *Semiclassical Dynamics*

We consider the FLRW model filled with a free massless scalar field ϕ; we discretize the area p, so we use the sine substitution (1a) on the connection c. We have a four-dimensional phase space $(c, p; \phi, P_\phi)$, where P_ϕ is the momentum conjugate to the scalar field. The modified Hamiltonian constraint for this model is

$$\mathcal{C}_{\text{poly}} = -\frac{3}{\gamma^2\beta_0^2}\sqrt{p}\,\sin^2(\beta_0 c) + \rho_\phi\, p^{\frac{3}{2}} = 0. \tag{2}$$

Then the dynamics is derived through the standard Hamilton's equations.

Through the constraint, we obtain a modified Friedmann equation:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{p}}{2p}\right)^2 = \frac{\rho_\phi}{3}\left(1 - \frac{\rho_\phi}{\rho_\beta}\right), \quad \rho_\beta = \frac{3}{\gamma^2\beta_0^2 p}; \tag{3}$$

the correction factor in parentheses contains a regularizing density ρ_β that depends on time through p; the minus sign introduces a critical point for the dynamics.

Since P_ϕ is a constant of motion, it is possible to use the scalar field as internal time; this makes the Friedmann equation easily solvable, and its solution $p(\phi)$ is

$$p(\phi) = \frac{\gamma\beta_0 P_\phi}{\sqrt{6}}\,\cosh\left(\sqrt{\frac{2}{3}}\,\phi\right). \tag{4}$$

The hyperbolic cosine has a non-zero minimum: the singularity is avoided and the Big Bang is replaced by a Big Bounce. However, the critical density, i.e. the energy density at the Bounce, is $\rho_{\text{crit}} = \rho_\beta(p|_{\phi_B}) = 3^{\frac{3}{2}}\sqrt{2}\,/P_\phi\gamma\beta_0$: it is dependent on the constant of motion P_ϕ that must be therefore set through initial conditions.

3.2. *Quantum Dynamics*

In the quantum picture, the fundamental variables are promoted to operators so that the quantum Hamiltonian constraint yields a Wheeler–De Witt (WDW) equation:

$$\hat{c} = \frac{\sin(\beta_0 c)}{\beta_0}, \quad \hat{p} = -i\frac{\gamma}{3}\frac{\mathrm{d}}{\mathrm{d}c}, \quad \widehat{P_\phi} = -i\frac{\mathrm{d}}{\mathrm{d}\phi}, \tag{5a}$$

$$\hat{\mathcal{C}}_{\text{poly}}\Psi(c,\phi) = \left[-\frac{2}{3\beta_0^2}\left(\sin(\beta_0 c)\,\frac{\mathrm{d}}{\mathrm{d}c}\right)^2 + \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\right]\Psi(c,\phi) = 0; \tag{5b}$$

through a substitution, the latter is recast as a Klein-Gordon-like (KG) equation:

$$x = \sqrt{\frac{3}{2}}\,\ln\left[\tan\left(\frac{\beta_0 c}{2}\right)\right] + x_0 \quad \Longrightarrow \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2}\Psi(x,\phi) = \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\Psi(x,\phi). \tag{6}$$

We construct its solution as a Gaussian-like wavepacket over eigenvalues k_ϕ of $\hat{P}_\phi$:

$$\Psi(x,\phi) = \int_0^\infty \frac{\mathrm{d}k_\phi}{\sqrt{4\pi\sigma}}\, e^{-\frac{(k_\phi-\overline{k}_\phi)^2}{2\sigma^2}}\, k_\phi\, e^{ik_\phi x}\, e^{ik_\phi \phi}; \tag{7}$$

then we evolve it backwards in time, and study the behaviour of the expectation value of the energy density operator through a KG scalar product of the form

$$\langle\widehat{\rho_\phi}\rangle = \int_{-\infty}^{\infty} \mathrm{d}x \left(\Psi^* \partial_\phi(\widehat{\rho_\phi}\Psi) - (\widehat{\rho_\phi}\Psi)\partial_\phi\Psi^*\right), \quad \widehat{\rho_\phi} = \frac{\hat{P}_\phi^2}{2|\hat{p}|^3}. \tag{8}$$

The result is a function of time ϕ. Figure 1 shows the energy density increasing towards the singularity, reaching a finite maximum and then decreasing again; in view of its scalar nature, this behaviour confirms the presence of a Big Bounce also at the quantum level. However, the critical density at the Bounce (i.e. the value of the peak) is dependent on the initial condition $\overline{k}_\phi$, as in the semiclassical dynamics.

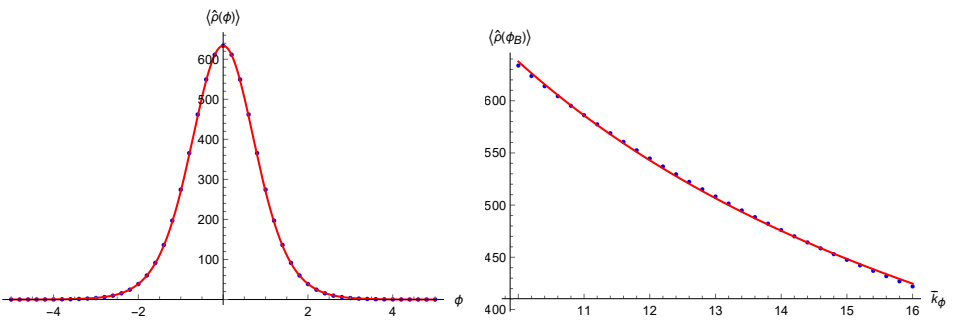


Fig. 1. Behaviour of the expected value of the energy density operator $\langle\widehat{\rho_\phi}\rangle$ as function of time ϕ (left panel) and dependence of the value of the energy density at the Bounce $\langle\widehat{\rho_\phi}\rangle\big|_{\phi_B}$ on the initial condition for the wavepacket $\overline{k}_\phi$ (right panel). The blue dots are the quantum expectation values, fitted with red continuous functions in accordance with semiclassical evolution.

4. Polymer FLRW Dynamics in Volume Variables

4.1. *Specialization of Polymer Cosmology*

In order to find the suitable variable to obtain a fixed Bounce density,[8] we perform a canonical transformation on the classical Hamiltonian to a generic function F of the scale factor: $p \to F(a)$, $c \to P_F = 2\sqrt{p}\, c\, /F'(a)$ with $F'(a) = \partial F/\partial a$. By using the polymer substitution (1a) on the new momentum P_F, we obtain a modified Hamiltonian constraint and a corresponding modified Friedmann equation:

$$\mathcal{C}_{\mathrm{poly}} = -\frac{3}{4\gamma^2\beta_0^2}\frac{\left(F'(a)\right)^2}{\sqrt{p}}\sin^2(\beta_0 P_F) + \rho_\phi\, p^{\frac{3}{2}} = 0, \quad \rho_\phi = \frac{P_\phi^2}{2p^3}, \tag{9a}$$

$$H^2 = \frac{\rho_\phi}{3}\left(1 - \frac{\rho_\phi}{\tilde{\rho}_\beta}\right), \quad \tilde{\rho}_\beta = \frac{3}{4\gamma^2\beta_0^2}\frac{\left(F'(a)\right)^2}{p^2} \propto \left(\frac{F'(a)}{p}\right)^2. \tag{9b}$$

To obtain a universal regularizing density $\tilde{\rho}_\beta$, we must have $F(a) \propto a^3$; we choose to use as new fundamental variables $v = a^3$ and $\tilde{c} = 2c/3\sqrt{p} \propto \dot{a}/a$.

4.2. *Semiclassical Dynamics*

We implement PQM on the system by considering v as discrete and using the formal substitution (1a) on the new generalized coordinate $\tilde{c}$:

$$\mathcal{C}_{\text{poly}} = -\frac{27}{4\gamma^2\beta_0^2}\, v\, \sin^2(\beta_0\tilde{c}) + \rho_\phi\, v = 0, \quad \rho_\phi = \frac{P_\phi^2}{2v^2}. \tag{10}$$

This new representation yields a modified Friedmann equation with a fixed regularizing density, as expected:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{v}}{3v}\right)^2 = \frac{\rho_\phi}{3}\left(1 - \frac{\rho_\phi}{\tilde{\rho}_\beta}\right), \quad \tilde{\rho}_\beta = \frac{27}{4\gamma^2\beta_0^2} = \tilde{\rho}_{\text{crit}}. \tag{11}$$

The quantity $\tilde{\rho}_\beta = \tilde{\rho}_{\text{crit}}$ depends only on fundamental constants and the Immirzi parameter, and can therefore be already considered a critical density. Using ϕ as time, the new modified Friedmann equation is easily solvable for $v(\phi)$:

$$v(\phi) = \frac{2\gamma\beta_0}{\sqrt{54}}\, P_\phi \cosh\left(\sqrt{\frac{3}{2}}\,\phi\right). \tag{12}$$

We still have a hyperbolic cosine, indicating a Bounce that replaces the Big Bang, but this time the value of the energy density at which this happens is universal.

4.3. *Quantum Dynamics*

In the quantization procedure the fundamental variables are promoted to operators and the quantum Hamiltonian constraint yields a WDW equation:

$$\hat{v} = -i\,\frac{\mathrm{d}}{\mathrm{d}\tilde{c}}, \quad \hat{\tilde{c}} = \frac{\sin(\beta_0\tilde{c})}{\beta_0}, \quad \hat{P}_\phi = -i\,\frac{\mathrm{d}}{\mathrm{d}\phi}; \tag{13}$$

$$\hat{\tilde{\mathcal{C}}}_{\text{poly}}\Psi(\tilde{c},\phi) = \left[-\frac{3}{2\beta_0^2}\left(\sin(\beta_0\tilde{c})\,\frac{\mathrm{d}}{\mathrm{d}\tilde{c}}\right)^2 + \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\right]\Psi(\tilde{c},\phi) = 0. \tag{14}$$

Through a suitable substitution, the latter is recast as a KG-like equation in the variable $\tilde{x}$ where ϕ plays the role of time:

$$\tilde{x} = \sqrt{\frac{2}{3}}\,\ln\left[\tan\left(\frac{\beta_0\tilde{c}}{2}\right)\right] + \tilde{x}_0 \quad \Longrightarrow \quad \frac{\mathrm{d}^2}{\mathrm{d}\tilde{x}^2}\Psi(\tilde{x},\phi) = \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\Psi(\tilde{x},\phi). \tag{15}$$

We construct the solution as a Gaussian-like wavepacket of the form

$$\Psi(\tilde{x},\phi) = \int_0^\infty \frac{\mathrm{d}k_\phi}{\sqrt{4\pi\sigma}}\, e^{-\frac{(k_\phi - \overline{k}_\phi)^2}{2\sigma^2}}\, k_\phi\, e^{ik_\phi\tilde{x}}\, e^{ik_\phi\phi}. \tag{16}$$

Finally, we evolve it backwards in time and study the behaviour of the expected value of relevant operators (the volume $\hat{V} = \hat{v}$ and the energy density $\widehat{\rho_\phi} = \hat{P}_\phi^2/2\hat{v}^2$) through the same KG scalar product (8) in terms of $\tilde{x}$. The Bounce is present also

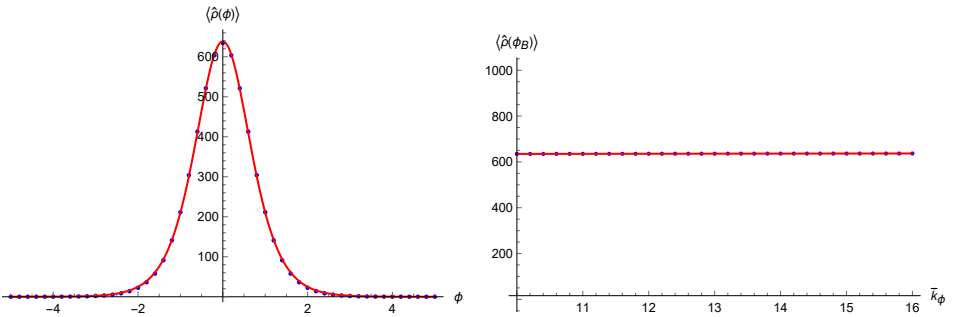


Fig. 2. Expected value of the energy density operator as function of time ϕ (left panel) and dependence of the critical density (the value of the energy density at the Bounce) on the initial condition for the wavepacket $\overline{k}_\phi$. The blue dots are the quantum expectation values that have been fitted with red continuous functions in accordance with semiclassical evolution.

in this quantum setting; however in this case the energy density, depicted in the left panel of figure 2, has a value at the Bounce that is constant and independent on $\overline{k}_\phi$, as shown in the right panel of the same figure.

5. Discussion of the Results

5.1. *The Nature of the Bounce in Polymer Cosmology*

The quantization procedure of PQM implements on a Hamiltonian system a lattice that has a constant step by construction. Therefore the canonical changes of variables must be performed on the classical Hamiltonian constraint before discretization. The nature of the resulting Bounce in the two pictures is not the same, and the two dynamics are inequivalent.

To recover the equivalence, we can perform the canonical transformation *after* the implementation of the lattice. We consider the system with the discretized volume and change the variables to Ashtekar ones; the canonical transformation must preserve the Poisson brackets and therefore must satisfy the condition $\beta_0 \tilde{c} = \beta' c$. We obtain a discretized system in the variables (c, p) that has a new lattice spacing β' dependent on the variable p. The new equation of motion for p is mapped to the original equation of motion in the Ashtekar variables, but the lattice parameter β_0 is replaced by the new one β'; the resulting dynamics is a Big Bounce scenario with a fixed, universal critical density, the same obtained by discretizing the volume with a constant lattice. Thus we conclude that the nature of the Bounce is decided by the variable to which a constant lattice spacing is assigned.

We stress that this result can be demonstrated only at a semiclassical level; on the quantum level this comparison is impossible, because a quantization procedure with a non-constant translational parameter has not been developed yet.

5.2. *Comparison with Loop Quantum Cosmology*

One of the most important results of LQG is the quantization of the kinematical geometrical operators of area and volume,[12] with the Area Gap Δ being the smallest non-zero area eigenvalue. In the construction of the quantum Hamiltonian constraint of LQC, the area gap is introduced from the full theory somewhat ad hoc, and the character of its implementation defines different schemes of LQC. In the first formulation, the area gap was implemented as a kinematical feature (as in full LQG) on the comoving area: this is equivalent to introducing a fixed lattice and yields the original μ_0 scheme of LQC, a scenario very similar to our polymer cosmology with the Ashtekar variables of section 3, that presents a Big Bounce dependent on initial conditions.[2,3] Then, in the improved dynamics, the area gap is implemented as a dynamical feature on the physical area, i.e. rescaled by the squared scale factor; this leads to a lattice with a non-constant spacing $\bar{\mu}(p)$, and prevents quantization. Therefore a change of basis is performed to a volume-like variable $\nu \propto v$ to make the lattice parameter constant (note how this is the inverse change of variables that we performed above). Now the quantization can be completed, and the resulting dynamics is a Big Bounce with a critical density that is a universal feature.[4,5] This scenario, dubbed $\bar{\mu}$ scheme, is analogous to the polymer FLRW cosmology in volume variables presented here in section 4.

The improved scheme avoids the dependence of quantum gravitational effects on initial conditions, a feature that does not have a very clear physical interpretation, and therefore seems to be more appealing than the original formulation. One criticism is that the new $\bar{\mu}$ scenario takes LQC further away from LQG because it does not use the Ashtekar variables anymore, that are privileged in the full theory in view of their $SU(2)$ character. However the dynamical similarities of PQM and LQC may justify the change of variables: at least on a semiclassical effective level, the dynamics of the model expressed in volume variables with a constant spacing is equivalent to that of the same model expressed in Ashtekar variables with a p-dependent lattice step.

6. Summary and Conclusions

We have shown how the implementation of PQM on the flat isotropic model solves the singularity by replacing the Big Bang with a Big Bounce, but its nature depends on the fundamental variable used to describe the model: with the area p from the Ashtekar variables the energy density at the Bounce is dependent on the initial conditions, while if the volume $v = a^3$ is used instead, the Bounce is a universal feature with a fixed energy density; therefore the two pictures yield two different

and inequivalent dynamics. A possibility to recover the equivalence is to perform the canonical transformation after the implementation of the lattice and to allow for a non-constant step; this does not affect the dynamics, meaning that the nature of the Bounce is decided by the variable that has a lattice with constant spacing.

In LQC two analogous pictures emerge: the original μ_0 scheme yields a Bounce dependent on initial conditions, while the improved $\bar{\mu}$ scheme prevents this feature through a change of basis and presents a universal Bounce. These similarities with polymer cosmology may be used to gain insight into the validity of the two scenarios: if on one hand the improved scheme, although physically more appealing, can be criticized for not using the privileged $SU(2)$ variables of LQG, on the other hand the parallelism with the minisuperspace implementation of PQM justifies the $\bar{\mu}$ scenario in view of the equivalence demonstrated in semiclassical polymer cosmology.

References

1. M. Bojowald, Loop Quantum Cosmology, *Living Reviews in Relativity* **8** (Dec 2005).
2. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the Big Bang, *Physical Review Letters* **96** (Apr 2006).
3. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the Big Bang: An analytical and numerical investigation, *Physical Review D* **73** (Jun 2006).
4. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the Big Bang: Improved dynamics, *Physical Review D* **74** (Oct 2006).
5. A. Ashtekar, A. Corichi and P. Singh, Robustness of key features of Loop Quantum Cosmology, *Physical Review D* **77** (Jan 2008).
6. C. Rovelli and L. Smolin, Loop space representation of quantum General Relativity, *Nuclear Physics B* **331**, 80 (1990).
7. A. Corichi, T. Vukašinac and J. A. Zapata, Polymer Quantum Mechanics and its continuum limit, *Physical Review D* **76** (Aug 2007).
8. G. Montani, C. Mantero, F. Bombacigno, F. Cianfrani and G. Barca, Semiclassical and quantum analysis of the isotropic Universe in the polymer paradigm, *Physical Review D* **99** (Mar 2019).
9. F. Mandini, G. Barca, E. Giovannetti and G. Montani, Polymer quantum dynamics of the isotropic Universe in the Ashtekar-Barbero-Immirzi and in the volume variables (2021).
10. R. Moriconi and G. Montani, Behavior of the Universe anisotropy in a Big Bounce cosmology, *Physical Review D* **95** (Jun 2017).
11. S. Antonini and G. Montani, Singularity-free and non-chaotic inhomogeneous Mixmaster in polymer representation for the volume of the Universe, *Physics Letters B* **790**, p. 475–483 (Mar 2019).
12. C. Rovelli and L. Smolin, Discreteness of area and volume in quantum gravity, *Nuclear Physics B* **442**, 593 (1995).

General relativistic evolution equations for density perturbations in open, flat and closed FLRW universes and the problem of structure formation

Pieter G. Miedema

Independent Scientist
Breda, The Netherlands
E-mail: PG.Miedema@ProtonMail.com

Due to the general covariance of the Einstein equations and conservation laws, the linearized equations have solutions which are gauge-dependent and have, therefore, no physical significance.

In this contribution I will show that the decomposition theorems for symmetric second-rank tensors of the maximally symmetric subspaces of constant time imply that there are exactly two, unique, gauge-invariant quantities which describe the true, physical perturbations to the energy density and particle number density. In the limit of zero spatial fluid velocity, and hence zero pressure, the set of linearized Einstein equations and conservation laws, combined with the new gauge-invariant quantities reduce to the Poisson equation of the Newtonian Theory of Gravity and the energy-mass relation of the Special Theory of Relativity. The relativistic gauge transformation reduces to the Newtonian gauge transformation in which time and space are decoupled.

The cosmological perturbation theory for open, flat and closed Friedmann-Lemaître-Robertson-Walker (FLRW) universes consists of a second-order ordinary differential equation (with source term entropy perturbations) which describes the evolution of perturbations in the total energy density, and a first-order ordinary differential equation which describes the evolution of entropy perturbations.

The cosmological perturbation theory is applied to a flat FLRW universe. For large-scale perturbations the outcome is in accordance with treatments in the literature. In the radiation-dominated era small-scale perturbations grew proportional to the square root of time and perturbations in the CDM particle number density were, due to gravitation, coupled to perturbations in the total energy density. Therefore, structure formation could have begun successfully only after decoupling of matter and radiation. After decoupling density perturbations exchanged heat with their environment. This heat exchange may have enhanced the growth rate of their mass sufficiently to explain structure formation in the early universe, a phenomenon which cannot be understood from adiabatic density perturbations.

Keywords: Cosmology; Perturbation Theory; Structure Formation.

1. Introduction

The theory related to the linearized Einstein equations and conservation laws is important in cosmology because it describes the growth of all kinds of structures in the expanding universe, such as stars, galaxies and microwave background fluctuations. Lifshitz[1] and Lifshitz and Khalatnikov[2] were the first researchers to derive a cosmological perturbation theory. They encountered the problem that the solutions of the linearized Einstein equations and conservation laws have no physical significance. Due to the linearity of the equations, the solutions can be changed

by a linear coordinate transformation, i.e., a gauge transformation. Therefore, the solutions are *gauge-dependent*. This is the well-known *gauge problem of cosmology* which is up till now not solved.

In this contribution I outline the solution of the gauge problem. Next, I briefly give the derivation of a new perturbation theory for cosmological perturbations in open, flat and closed FLRW universes. The perturbation theory is applied to a flat FLRW universe to study the evolution of density perturbations. The derivation of the perturbation theory is given in full detail in Ref. 3, whereas the problem of structure formation in a flat FLRW universe is discussed in detail in Ref. 4.

2. Cosmological Perturbation Theory

We consider FLRW universes filled with a perfect fluid. The energy momentum tensor is then given by

$$T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu - pg^{\mu\nu}. \tag{1}$$

From thermodynamics it is known that both the energy density ε and the pressure p depend on the particle number density n and the absolute temperature T, i.e.,

$$\varepsilon = \varepsilon(n, T), \quad p = p(n, T). \tag{2}$$

Since T can, in principle, be eliminated from these equations of state, a computationally more convenient equation of state for the pressure is used, namely

$$p = p(n, \varepsilon). \tag{3}$$

Using (1) and (3) one can write down the Einstein equations and conservation laws into a set of background equations

$$G^{\mu\nu}_{(0)} = \kappa T^{\mu\nu}_{(0)}, \quad T^{\mu\nu}_{(0)\,;\nu} = 0, \tag{4}$$

which govern the *global* evolution of FLRW universes, and a system of perturbation equations

$$G^{\mu\nu}_{(1)} = \kappa T^{\mu\nu}_{(1)}, \quad T^{\mu\nu}_{(1)\,;\nu} = 0. \tag{5}$$

These equations govern the evolution of *local* perturbations in FLRW universes.

2.1. *General Covariance*

It is well-known that the General Theory of Relativity is invariant under general coordinate transformations $x^\mu \to x'^\mu(x^\nu)$. In particular, the linearized Einstein equations and conservation laws are invariant under a general linear coordinate transformation, i.e., a gauge transformation

$$x^\mu \to x'^\mu = x^\mu - \xi^\mu(t, \boldsymbol{x}), \tag{6}$$

where $\xi^\mu(t, \boldsymbol{x})$ are four arbitrary infinitesimal functions of time, $x^0 = ct$, and space, $\boldsymbol{x} = (x^1, x^2, x^3)$, coordinates, the so-called gauge functions. As a consequence, there

are no preferred coordinate systems (see Weinberg,[5] Appendix B), so that one may choose any system of reference which makes the calculations tractable.

However, because of the linearity of the perturbation equations and conservation laws, the so-called gauge problem of cosmology emerges.

2.2. *Gauge Problem of Cosmology*

Consider a closed, flat or open FLRW universe filled with a perfect fluid, with an equation of state for the pressure (3). The evolution of the energy density $\varepsilon_{(0)}$, the particle number density $n_{(0)}$ and the expansion scalar $\theta_{(0)} = 3H$ of the background universe are governed by the usual Einstein equations and conservation laws for a FLRW universe

$$3H^2 = \tfrac{1}{2}R_{(0)} + \kappa\varepsilon_{(0)} + \Lambda, \quad \kappa = 8\pi G_{\mathrm{N}}/c^4, \tag{7a}$$

$$\dot{R}_{(0)} = -2HR_{(0)}, \tag{7b}$$

$$\dot{\varepsilon}_{(0)} = -3H\varepsilon_{(0)}(1+w), \qquad w := p_{(0)}/\varepsilon_{(0)}, \tag{7c}$$

$$\dot{n}_{(0)} = -3Hn_{(0)}. \tag{7d}$$

From this system of equations it follows that the evolution of FLRW universes is characterized by the three scalars $\varepsilon_{(0)}$, $n_{(0)}$ and $\theta_{(0)}$.

The general solution of the linearized Einstein equations and conservation laws contain, among other quantities, the perturbed quantities $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$. Since the linearized equations are invariant under linear coordinate transformations (6) one can generate new, equivalent, solutions $\varepsilon'_{(1)}$, $n'_{(1)}$ and $\theta'_{(1)}$, i.e.,

$$\varepsilon'_{(1)} = \varepsilon_{(1)} + \xi^0\dot{\varepsilon}_{(0)}, \quad n'_{(1)} = n_{(1)} + \xi^0\dot{n}_{(0)}, \quad \theta'_{(1)} = \theta_{(1)} + \xi^0\dot{\theta}_{(0)} \tag{8}$$

where $\xi^0\dot{\varepsilon}_{(0)}$, $\xi^0\dot{n}_{(0)}$ and $\xi^0\dot{\theta}_{(0)}$ are the so-called gauge modes. Therefore, $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$ are *gauge-dependent* and have, as a consequence, no physical significance.

3. Previous Attempts to Solve the Gauge Problem of Cosmology

Bardeen[6] was the first to demonstrate that, using gauge-invariant quantities, one can recast the linearized Einstein equations and conservation laws into a new set of evolution equations that are free of non-physical gauge modes. Bardeen's definition of a gauge-invariant density perturbation is such that it becomes equal to the gauge-dependent density perturbation in the limit of small scales. The underlying assumption is that on small scales the gauge modes vanish so that gauge-dependent quantities become gauge-independent. However, the physical quantities $\varepsilon^{\mathrm{phys}}_{(1)}$, $n^{\mathrm{phys}}_{(1)}$ and $\theta^{\mathrm{phys}}_{(1)}$ do not become equal to the gauge-dependent quantities $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$ in the non-relativistic limit, since the gauge modes $\hat{\varepsilon}_{(1)}$, $\hat{n}_{(1)}$ and $\hat{\theta}_{(1)}$ never become zero. This is a direct consequence of the facts that the universe is not static in the non-relativistic limit and that the relativistic space-time gauge transformations reduce in the non-relativistic limit to Newtonian gauge transformations with space

and time decoupled. Consequently, the quantities defined by Bardeen are not equal to the real physical energy density perturbation.

The article of Bardeen has inspired the pioneering works of Ellis *et al.*,[7–9] Mukhanov *et al.*[10,11] and Bruni *et al.*[12] These researchers proposed alternative perturbation theories using gauge-invariant quantities which differ from the ones used by Bardeen. The theory proposed by Mukhanov *et al.*[10,11] yields the Poisson equation of the Newtonian Theory of Gravity only if the Hubble function, i.e., the expansion scalar, vanishes. This, however, violates the background Friedmann equation. In fact, the background Friedmann equation is needed to derive the Poisson equation from the linearized Friedmann equation in the non-relativistic limit. Consequently, the gauge-invariant quantities defined in Refs. 10 and 11 are not equal to the true physical perturbations.

In the approach in Refs. 7–9 a density perturbation is defined using gradients, i.e., a gauge-invariant and covariant function is defined which 'closely corresponds to the intention of the usual gauge-dependent density contrast function'. However, a perturbation theory based on this definition does not yield the Poisson equation in the non-relativistic limit.

The above mentioned pioneering treatments caused an abundance of other important works, e.g., Refs. 13–18. In his review article, Ellis[19] discusses the work of Lifshitz and Khalatnikov, and also provides an overview of other research on the subject. From the vast literature on cosmological density perturbations, it must be concluded that there is still no agreement on which gauge-invariant quantities are the true energy density perturbation and particle number density perturbation. None of the cosmological perturbation theories in the literature has a correct non-relativistic limit, so that there is still no solution to the gauge problem of cosmology.

4. Real, Measurable Density Perturbations

In contrast to the gauge-dependent quantities $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$ the real, measurable density perturbations $\varepsilon^{\rm phys}_{(1)}$, $n^{\rm phys}_{(1)}$ and $\theta^{\rm phys}_{(1)}$ are independent of the choice of a coordinate system, i.e., are *gauge-invariant*. The physics of density perturbations is hidden in the general solution of the linearized Einstein equations and conservation laws. Therefore, the physical quantities $\varepsilon^{\rm phys}_{(1)}$, $n^{\rm phys}_{(1)}$ and $\theta^{\rm phys}_{(1)}$ can be expressed as linear combinations, with time-dependent coefficients, of gauge-dependent solutions of these equations, such that the gauge modes are eliminated. Furthermore, $\varepsilon^{\rm phys}_{(1)}$ and $n^{\rm phys}_{(1)}$ have the property that in the non-relativistic limit the linearized Einstein equations and conservation laws combined with expressions for $\varepsilon^{\rm phys}_{(1)}$and $n^{\rm phys}_{(1)}$ reduce, in a *non-static* flat FLRW universe, to the time-independent Poisson equation with $\varepsilon^{\rm phys}_{(1)}$ as source term, and the well-known Einstein mass-energy relation $\varepsilon^{\rm phys}_{(1)} = n^{\rm phys}_{(1)} mc^2$ of the Special Theory of Relativity. Finally, the relativistic gauge transformation (6) reduces in the non-relativistic limit to the Newtonian gauge transformation. Consequently, the gauge problem is, in fact, the problem of finding expressions for $\varepsilon^{\rm phys}_{(1)}$, $n^{\rm phys}_{(1)}$ and $\theta^{\rm phys}_{(1)}$.

5. Solution of the Gauge Problem

The gauge-dependent quantities $\varepsilon_{(1)}$, $n_{(1)}$ and $\theta_{(1)}$ transform under the gauge transformation (6) in exactly the same way, namely as (8). Therefore, one can combine two of these quantities such that the gauge modes are eliminated. We have three choices and in each choice one of the gauge-invariant quantities vanishes. The obvious, non-trivial choice is

$$\varepsilon^{\rm phys}_{(1)} := \varepsilon_{(1)} - \frac{\dot{\varepsilon}_{(0)}}{\dot{\theta}_{(0)}}\theta_{(1)}, \quad n^{\rm phys}_{(1)} := n_{(1)} - \frac{\dot{n}_{(0)}}{\dot{\theta}_{(0)}}\theta_{(1)}, \quad \theta^{\rm phys}_{(1)} := \theta_{(1)} - \frac{\dot{\theta}_{(0)}}{\dot{\theta}_{(0)}}\theta_{(1)} = 0, \tag{9}$$

where $\theta^{\rm phys}_{(1)} = 0$ implies that local density perturbations have no influence on the global expansion $\theta_{(0)} = 3H$ of the universe.

Now, the question arises, are $\varepsilon^{\rm phys}_{(1)}$, $n^{\rm phys}_{(1)}$ and $\theta^{\rm phys}_{(1)}$ the real, measurable perturbations? Since the linear combinations (9) are the only possible ones, it is plausible that this is indeed the case. However, to be sure, we have to study the quantities (9) together with the background equations (4) and the perturbation equations (5) in the non-relativistic limit. This is the subject of Sec. 8. To that end, we have to rewrite the linearized Einstein equations and conservation laws (5) into a more convenient form. This will be done in Sec. 7 with the help of the decomposition theorems in the next section.

6. Decomposition Theorems of Symmetric Second-rank 3-tensors

Since the General Theory of Relativity is invariant under coordinate transformations, we may choose our coordinates such that the calculations are tractable.

We choose a synchronous system of reference, since this facilitates the non-relativistic limit: in the Newtonian Theory of Gravity all coordinate systems are synchronous. A second reason is that in synchronous coordinates, the decomposition theorems in Refs. 20–22 for symmetric second rank tensors of the maximally symmetric three-spaces of constant time can be used.

In synchronous coordinates the metric tensor $g_{\mu\nu}(t, \boldsymbol{x})$ of FLRW universes has the form

$$g_{00} = 1, \quad g_{0i} = 0, \quad g_{ij} = -a^2(t)\tilde{g}_{ij}(\boldsymbol{x}), \tag{10}$$

where $a(t)$ is the scale factor of the universe, $g_{00} = 1$ indicates that coordinate time is equal to proper time, $g_{0i} = 0$ is the global synchronicity condition (see the textbook of Landau and Lifshitz,[23] § 84) and $\tilde{g}_{ij}$ is the metric tensor of the three-dimensional maximally symmetric sub-spaces of constant time[a]. From the Killing equations $\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0$ and (10) it follows that the functions $\xi^\mu(t, \boldsymbol{x})$ in (6) become

$$\xi^0 = \psi(\boldsymbol{x}), \quad \xi^i = \tilde{g}^{ik}(\boldsymbol{x})\frac{\partial\psi(\boldsymbol{x})}{\partial x^k}\int^{ct}\frac{{\rm d}\tau}{a^2(\tau)} + \chi^i(\boldsymbol{x}), \tag{11}$$

[a]Quantities with a tilde, i.e., $\tilde{q}$, belong to the three-dimensional maximally symmetric subspaces of constant time.

if only transformations between synchronous coordinates are allowed. The four functions $\psi(\boldsymbol{x})$ and $\chi^i(\boldsymbol{x})$ cannot be fixed since the four coordinate conditions $g_{00} = 1$ and $g_{0i} = 0$ have already exhausted all four degrees of freedom (see Weinberg,[24] Section 7.4 on coordinate conditions).

In Refs. 20–22 it is shown that *any* symmetric second-rank three-tensor, and hence the perturbed metric tensor h_{ij} and the perturbed Ricci tensor $R_{(1)ij}$, can uniquely be decomposed into three parts. For the perturbed metric tensor, one has

$$h^i{}_j = h^i_{\parallel j} + h^i_{\perp j} + h^i_{*j}, \qquad h^k_{\perp k} = 0, \quad h^k_{*k} = 0, \quad h^k_{*i|k} = 0, \tag{12}$$

where tensor, vector and scalar perturbations are denoted by $*$, $\perp$ and $\parallel$, respectively. The perturbed Ricci tensor, $R_{(1)ij}$, being a symmetric second-rank three-tensor, can in the same way be decomposed

$$R^i_{(1)j} = R^i_{(1)\parallel j} + R^i_{(1)\perp j} + R^i_{(1)*j}, \qquad R^k_{(1)\perp k} = 0, \quad R^k_{(1)*k} = 0, \quad R^k_{(1)*i|k} = 0. \tag{13}$$

Finally, York and Stewart demonstrated that the components $h^i_{\parallel j}$ can be written in terms of two independent potentials $\phi(t, \boldsymbol{x})$ and $\zeta(t, \boldsymbol{x})$, namely

$$h^i_{\parallel j} = \frac{2}{c^2}(\phi\delta^i{}_j + \zeta^{|i}{}_{|j}). \tag{14}$$

The decompositions (12) and (13) and the expression (14) imply that the decomposition of the spatial part $\boldsymbol{u}_{(1)} = (u^1_{(1)}, u^2_{(1)}, u^3_{(1)})$ of the perturbed fluid four-velocity

$$\boldsymbol{u}_{(1)} = \boldsymbol{u}_{(1)\parallel} + \boldsymbol{u}_{(1)\perp}, \tag{15}$$

where the components $\boldsymbol{u}_{(1)\parallel}$ and $\boldsymbol{u}_{(1)\perp}$ have the properties

$$\tilde{\boldsymbol{\nabla}} \cdot \boldsymbol{u}_{(1)} = \tilde{\boldsymbol{\nabla}} \cdot \boldsymbol{u}_{(1)\parallel}, \quad \tilde{\boldsymbol{\nabla}} \times \boldsymbol{u}_{(1)} = \tilde{\boldsymbol{\nabla}} \times \boldsymbol{u}_{(1)\perp}, \tag{16}$$

is consistent with the three linearized momentum constraint equations $G^{0i}_{(1)} = \kappa T^{0i}_{(1)}$. In expressions (16) $\tilde{\boldsymbol{\nabla}}$ is the generalized vector differential operator, defined by $\tilde{\nabla}_i v^k := v^k{}_{|i}$.

7. Evolution Equations for Scalar Perturbations

The general solution of the linearized Einstein equations and conservation laws (5) is a linear combination of scalar, vector and tensor perturbations. From the decomposition theorems it follows that only the scalar perturbations are coupled to density perturbations. Therefore, in the linearized equations (5) we may replace $h^i{}_j$ by $h^i_{\parallel j}$ and $u^i_{(1)}$ by $u^i_{(1)\parallel}$, to obtain perturbation equations which exclusively describe the evolution of scalar perturbations. For the Ricci tensor we find $R^i_{(1)j} \to R^i_{(1)\parallel j}$, so that the local perturbation to the global spatial curvature $R_{(0)}$ due to local density perturbations is given by

$$R_{(1)} := R^k_{(1)k} = g^{ij}_{(0)}(h^k{}_{k|i|j} - h^k{}_{i|k|j}) + \tfrac{1}{3}R_{(0)}h^k{}_k. \tag{17}$$

Since the rotational part of the local spatial fluid velocity is not coupled to scalar perturbations we may for scalar perturbations replace the three components of the fluid velocity by its divergence, i.e.,

$$\vartheta_{(1)} := u^k_{(1)|k}. \tag{18}$$

Finally, the perturbation to the expansion is given by

$$\theta_{(1)} = \vartheta_{(1)} - \tfrac{1}{2}\dot{h}^k{}_k. \tag{19}$$

Using the substitutions given above, we can rewrite the full set of perturbed Einstein equations and conservation laws (5) in the form

$$2H(\theta_{(1)} - \vartheta_{(1)}) = \tfrac{1}{2}R_{(1)} + \kappa\varepsilon_{(1)}, \tag{20a}$$

$$\dot{R}_{(1)} = -2HR_{(1)} + 2\kappa\varepsilon_{(0)}(1+w)\vartheta_{(1)} - \tfrac{2}{3}R_{(0)}(\theta_{(1)} - \vartheta_{(1)}), \tag{20b}$$

$$\dot{\varepsilon}_{(1)} = -3H(\varepsilon_{(1)} + p_{(1)}) - \varepsilon_{(0)}(1+w)\theta_{(1)}, \tag{20c}$$

$$\dot{\vartheta}_{(1)} = -H(2-3\beta^2)\vartheta_{(1)} - \frac{1}{\varepsilon_{(0)}(1+w)}\frac{\tilde{\nabla}^2 p_{(1)}}{a^2}, \quad \beta^2 := \frac{\dot{p}_{(0)}}{\dot{\varepsilon}_{(0)}}, \tag{20d}$$

$$\dot{n}_{(1)} = -3Hn_{(1)} - n_{(0)}\theta_{(1)}, \tag{20e}$$

which describe exclusively scalar perturbations. The system (20) consists of the energy density constraint equation or linearized Friedmann equation (20a), the momentum constraint equation (20b), the energy density conservation law (20c), the momentum conservation law (20d) and the particle number density conservation law (20e). These equations are the perturbed counterparts of Eqs. (7). Since the system of equations (20) includes all constraint equations and conservation laws, the *dynamical* equations $G^{ij}_{(1)} = \kappa T^{ij}_{(1)}$ are automatically satisfied by the solution of the system (20).

The background equations (7) combined with the system (20) determines the evolution of $\varepsilon^{\text{phys}}_{(1)}$ and $n^{\text{phys}}_{(1)}$ given by (9). However, the system (20) has also gauge modes as solution, i.e.,

$$\hat{\varepsilon}_{(1)} = \psi\dot{\varepsilon}_{(0)}, \quad \hat{n}_{(1)} = \psi\dot{n}_{(0)}, \quad \hat{\theta}_{(1)} = \psi\dot{\theta}_{(0)}, \tag{21a}$$

$$\hat{\vartheta}_{(1)} = -\frac{\tilde{\nabla}^2\psi}{a^2}, \quad \hat{R}_{(1)} = 4H\left[\frac{\tilde{\nabla}^2\psi}{a^2} - \tfrac{1}{2}R_{(0)}\psi\right], \tag{21b}$$

where $\xi^0 = \psi(\boldsymbol{x})$ in synchronous coordinates, see (11). Therefore, I have derived a new system of equations which is free from the gauge modes and determines exclusively the evolution of $\varepsilon^{\text{phys}}_{(1)}$ and $n^{\text{phys}}_{(1)}$. This will be the subject of Sec. 9. First, we show in the next section that $\varepsilon^{\text{phys}}_{(1)}$ and $n^{\text{phys}}_{(1)}$ are the real, measurable density perturbations.

8. Non-relativistic Limit

The standard non-relativistic limit is defined by four requirements, see, e.g., the textbook of Carroll,[25] page 153:

(1) The gravitational field should be weak, i.e., can be considered as a perturbation of a flat, i.e., $R_{(0)} = 0$, FLRW universe.

(2) The particles are moving slowly with respect to the speed of light.

(3) The transformation (6) with (11) is reduced to transformations for which the Newtonian Theory of Gravity is invariant.

(4) The gravitational field of a density perturbation should be static, i.e., it does not change with time.

Requirement (3), which is not used in the literature, is needed because we have to deal with gauge transformations (6) with (11) in an expanding universe.

In the Newtonian Theory of Gravity space is flat. Therefore, we consider a flat, $R_{(0)} = 0$, FLRW universe, so that spatial covariant derivatives reduce to ordinary derivatives. Combining expressions (14) and (17), we get

$$R_{(1)} = \frac{4}{c^2}\phi^{|k}{}_{|k} = -\frac{4}{c^2}\frac{\nabla^2\phi}{a^2}, \tag{22}$$

where ∇^2 is the usual Laplace operator.

The non-relativistic limit follows from

$$u^i_{(1)\mathrm{phys}} \to 0, \quad \Leftrightarrow \quad p^{\mathrm{phys}}_{(1)} \to 0, \tag{23}$$

i.e., the *physical* part of the fluid velocity is small with respect to the speed of light. Note that non-zero pressure differences imply fluid flow and vice versa. Using the background equations (7), the quantities (9) and the perturbation equations (20) we can draw the following conclusions:

(1) The momentum constraint equation (20b) implies with (22) and (23) that $\dot{\phi} = 0$, i.e., in the non-relativistic limit the potential ϕ is independent of time.

(2) The perturbed Friedmann equation (20a) and the time-derivative of the background Friedmann equation (7a) imply the Poisson equation

$$\nabla^2\varphi(\boldsymbol{x}) = \frac{4\pi G_{\mathrm{N}}}{c^2}\varepsilon^{\mathrm{phys}}_{(1)}(\boldsymbol{x}), \tag{24}$$

where $\varphi(\boldsymbol{x}) := \phi(\boldsymbol{x})/a^2(t_0)$.

(3) In the limit (23) the masses of the particles in a density perturbation are equal to their rest masses, so that the definitions (9) imply with (20c) and (20e)

$$n^{\mathrm{gi}}_{(1)}(\boldsymbol{x}) = \frac{\varepsilon^{\mathrm{gi}}_{(1)}(\boldsymbol{x})}{mc^2}, \tag{25}$$

which is the well-known energy mass relation of the Special Theory of Relativity.

(4) The relativistic gauge transformation (6) reduces in the non-relativistic limit to the Newtonian gauge transformation

$$x^0 \to x'^0 = x^0 - C, \quad x^i \to x'^i = x^i - \chi^i(\boldsymbol{x}), \tag{26}$$

where $\chi^i(\boldsymbol{x})$ are three arbitrary functions of the spatial coordinates and C is an arbitrary constant. Thus, in the non-relativistic limit time and space transformations are decoupled: time coordinates may be shifted and spatial coordinates may be chosen arbitrarily.

(5) Finally, the perturbed conservation laws (20c) and (20e) have only the gauge modes as solutions. Therefore, these equations have no physical significance and need not be considered anymore.

We can now draw the conclusion that $\varepsilon^{\text{phys}}_{(1)}$, $n^{\text{phys}}_{(1)}$ and $\theta^{\text{phys}}_{(1)} = 0$ defined by the expressions (9) are the unique, measurable, physical perturbations to the energy density, the particle number density and the global expansion, respectively.

9. Evolution Equations for Density Perturbations

I have used the computer algebra system to rewrite the system (20) into a new system of equations for the density contrast functions $\delta_\varepsilon := \varepsilon^{\text{phys}}_{(1)}/\varepsilon_{(0)}$ and $\delta_n := n^{\text{phys}}_{(1)}/n_{(0)}$. The result is the system of evolution equations for density perturbations in an open, flat and closed FLRW universe:

$$\ddot{\delta}_\varepsilon + b_1\dot{\delta}_\varepsilon + b_2\delta_\varepsilon = b_3\left[\delta_n - \frac{\delta_\varepsilon}{1+w}\right], \tag{27a}$$

$$\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t}\left[\delta_n - \frac{\delta_\varepsilon}{1+w}\right] = \frac{3Hn_{(0)}p_n}{\varepsilon_{(0)}(1+w)}\left[\delta_n - \frac{\delta_\varepsilon}{1+w}\right], \tag{27b}$$

where the coefficients b_1, b_2 and b_3 are determined by the background equations (7) and the equation of state for the pressure $p = p(n, \varepsilon)$.

9.1. *Diabatic Perturbations*

Density perturbations are diabatic ('non-adiabatic') if and only if the equation of state for the pressure depends on both the particle number density n and the energy density ε.

The combined First and Second Law of Thermodynamics is given by

$$T_{(0)}s^{\text{phys}}_{(1)} = -\frac{\varepsilon_{(0)}(1+w)}{n_{(0)}}\left[\delta_n - \frac{\delta_\varepsilon}{1+w}\right]. \tag{28}$$

Thus, the evolution of density perturbations is determined by both their internal gravity and heat exchange with their environment.

For diabatic perturbations the pressure perturbation is given by

$$p^{\text{phys}}_{(1)} = \beta^2\varepsilon_{(0)}\delta_\varepsilon + n_{(0)}p_n\left[\delta_n - \frac{\delta_\varepsilon}{1+w}\right], \tag{29}$$

where the first term is the *adiabatic* part of the pressure perturbation and the second term is the *diabatic* part.

9.2. *Adiabatic Perturbations*

For barotropic equations of state, i.e., $p = p(\varepsilon)$, we have $p_n := (\partial p/\partial n)_\varepsilon = 0$, implying that $b_3 = 0$. In this case Eq. (27a) is a homogeneous equation and, as a consequence, describes adiabatic perturbations. Since for adiabatic perturbations $T_{(0)}s^{\rm phys}_{(1)} = 0$, Eq. (27b) reduces in this case to an algebraic equation

$$\delta_n - \frac{\delta_\varepsilon}{1+w} = 0, \tag{30}$$

so that for adiabatic perturbations the perturbations in the energy density are coupled to perturbations in the particle number density. Adiabatic perturbations evolve only through their internal gravity.

10. Structure Formation in a Flat FLRW Universe

In this section we consider a flat FLRW universe. The coefficients of Eq. (27a) are in this case given by

$$b_1 = H(1-3w-3\beta^2) - 2\frac{\dot\beta}{\beta}, \tag{31a}$$

$$b_2 = \kappa\varepsilon_{(0)}\left[2\beta^2(2+3w) - \tfrac{1}{6}(1+18w+9w^2)\right] + 2H\frac{\dot\beta}{\beta}(1+3w) - \beta^2\frac{\nabla^2}{a^2}, \tag{31b}$$

$$b_3 = \left\{\frac{-2}{1+w}\left[\varepsilon_{(0)}p_{\varepsilon n}(1+w) + \frac{2p_n}{3H}\frac{\dot\beta}{\beta} + p_n(p_\varepsilon - \beta^2) + n_{(0)}p_{nn}\right] + p_n\right\}\frac{n_{(0)}}{\varepsilon_{(0)}}\frac{\nabla^2}{a^2}, \tag{31c}$$

where the partial derivatives of the equation of state for the pressure $p(n,\varepsilon)$ are given by

$$p_n := \left(\frac{\partial p}{\partial n}\right)_\varepsilon, \quad p_\varepsilon := \left(\frac{\partial p}{\partial \varepsilon}\right)_n, \quad p_{\varepsilon n} := \frac{\partial^2 p}{\partial\varepsilon\,\partial n}, \quad p_{nn} := \frac{\partial^2 p}{\partial n^2}. \tag{32}$$

The symbol ∇^2 denotes the Laplace operator. The quantity $\beta(t)$ is defined by $\beta^2 := \dot p_{(0)}/\dot\varepsilon_{(0)}$. Using that $\dot p_{(0)} = p_n\dot n_{(0)} + p_\varepsilon\dot\varepsilon_{(0)}$ and the conservation laws (7c) and (7d) one gets

$$\beta^2 = p_\varepsilon + \frac{n_{(0)}p_n}{\varepsilon_{(0)}(1+w)}. \tag{33}$$

From the definitions $w := p_{(0)}/\varepsilon_{(0)}$ and $\beta^2 := \dot p_{(0)}/\dot\varepsilon_{(0)}$ and the energy conservation law (7c), one finds for the time-derivative of w

$$\dot w = 3H(1+w)(w-\beta^2). \tag{34}$$

This expression holds true *independent* of the precise form of the equation of state.

10.1. *Radiation Dominated Universe*

In this era the primordial fluid can be treated as radiation-dominated with equations of state

$$\varepsilon = a_{\mathrm{B}} T_\gamma^4, \quad p = \tfrac{1}{3} a_{\mathrm{B}} T_\gamma^4, \tag{35}$$

where a_{B} is the black body constant and T_γ the radiation temperature. The equations of state (35) imply the equation of state for the pressure $p = \frac{1}{3}\varepsilon$, so that, with (32),

$$p_n = 0, \quad p_\varepsilon = \tfrac{1}{3}. \tag{36}$$

Therefore, one has from (33),

$$\beta^2 = w = \tfrac{1}{3}. \tag{37}$$

Using (36) and (37), the perturbation equations (27) reduce to

$$\ddot{\delta}_\varepsilon - H\dot{\delta}_\varepsilon - \left[\frac{1}{3}\frac{\nabla^2}{a^2} - \tfrac{2}{3}\kappa\varepsilon_{(0)}\right]\delta_\varepsilon = 0, \tag{38a}$$

$$\delta_n - \tfrac{3}{4}\delta_\varepsilon = 0. \tag{38b}$$

Using the solutions of the background equations (7) and using that $\delta_\varepsilon(t, \boldsymbol{x}) = \delta_\varepsilon(t, \boldsymbol{q}) \exp(\mathrm{i}\boldsymbol{q} \cdot \boldsymbol{x})$, with $|\boldsymbol{q}| = 2\pi/\lambda$, Eq. (38a) can be written in the form

$$\delta_\varepsilon'' - \frac{1}{2\tau}\delta_\varepsilon' + \left[\frac{\mu_{\mathrm{r}}^2}{4\tau} + \frac{1}{2\tau^2}\right]\delta_\varepsilon = 0, \quad \mu_{\mathrm{r}} := \frac{2\pi}{\lambda_0}\frac{1}{H(t_0)}\frac{1}{\sqrt{3}}, \quad \lambda_0 := \lambda a(t_0), \tag{39}$$

where a prime denotes differentiation with respect to the dimensionless time $\tau := t/t_0 \geq 1$. The general solution of Eq. (39) is

$$\delta_\varepsilon(\tau, \boldsymbol{q}) = \left[A_1(\boldsymbol{q}) \sin\left(\mu_{\mathrm{r}}\sqrt{\tau}\right) + A_2(\boldsymbol{q}) \cos\left(\mu_{\mathrm{r}}\sqrt{\tau}\right)\right]\sqrt{\tau}, \tag{40}$$

where $A_1(\boldsymbol{q})$ and $A_2(\boldsymbol{q})$ are the 'constants' of integration. For large-scale ($\lambda \to \infty$) perturbations, we find from Eq. (40) the well-known[2, 26–29] solution

$$\delta_\varepsilon = c_1\tau + c_2\tau^{1/2}. \tag{41}$$

We can now draw the following conclusions. From Eq. (40) it follows that density perturbations oscillate. However, in contrast to the outcome of the standard perturbation theory in the literature, density perturbations oscillate with an *increasing* amplitude. Moreover, baryons were tightly coupled to radiation through Thomson scattering, i.e., baryons obey $\delta_{n,\,\mathrm{baryon}} = \frac{3}{4}\delta_\varepsilon$. Thus, for baryons (38b) is identically satisfied. In contrast to baryons, CDM is *not* coupled to radiation through Thomson scattering. However, equation (38b) follows from the General Theory of Relativity. As a consequence, equation (38b) should be obeyed by *all* kinds of particles that interact through gravitation. In other words, equation (38b) holds true for baryons as well as CDM. Since CDM interacts only via gravity with baryons and radiation, the fluctuations in CDM are coupled through gravitation to fluctuations in the energy density.

10.2. *Era after Decoupling of Matter and Radiation*

Once protons and electrons combined to yield hydrogen, the radiation pressure was negligible, and the equations of state have become those of a non-relativistic monatomic perfect gas with three degrees of freedom

$$\varepsilon(n,T) = nmc^2 + \tfrac{3}{2}nk_{\mathrm{B}}T, \quad p(n,T) = nk_{\mathrm{B}}T, \tag{42}$$

where k_{B} is Boltzmann's constant, m the mean particle mass, and T the temperature of the matter. Eliminating T from (42) yields the equation of state for the pressure

$$p(n,\varepsilon) = \tfrac{2}{3}(\varepsilon - nmc^2), \tag{43}$$

so that with (32) one has

$$p_n = -\tfrac{2}{3}mc^2, \quad p_\varepsilon = \tfrac{2}{3}. \tag{44}$$

Substituting p_n, p_ε, and ε given by (42) into (33) on finds, using that $mc^2 \gg k_{\mathrm{B}}T$,

$$\beta \approx \frac{v_{\mathrm{s}}}{c} = \sqrt{\frac{5}{3}\frac{k_{\mathrm{B}}T_{(0)}}{mc^2}}, \quad w \approx \frac{k_{\mathrm{B}}T_{(0)}}{mc^2} \lesssim 2.7\times 10^{-10}, \tag{45}$$

with v_{s} the adiabatic speed of sound and $T_{(0)}$ the matter temperature. Using that $\beta^2 \approx \frac{5}{3}w$ and $w \ll 1$, expression (34) reduces to $\dot{w} \approx -2Hw$, so that with $H := \dot{a}/a$ one has $w \propto a^{-2}$. This implies with (45) that the matter temperature decays as

$$T_{(0)} \propto a^{-2}. \tag{46}$$

This, in turn, implies with (45) that $\dot{\beta}/\beta = -H$. The system (27) can now be rewritten as

$$\ddot{\delta}_\varepsilon + 3H\dot{\delta}_\varepsilon - \left[\beta^2\frac{\nabla^2}{a^2} + \tfrac{5}{6}\kappa\varepsilon_{(0)}\right]\delta_\varepsilon = -\frac{2}{3}\frac{\nabla^2}{a^2}\left(\delta_n - \delta_\varepsilon\right), \tag{47a}$$

$$\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t}\left(\delta_n - \delta_\varepsilon\right) = -2H\left(\delta_n - \delta_\varepsilon\right), \tag{47b}$$

where $w \ll 1$ and $\beta^2 \ll 1$ have been neglected with respect to constants of order unity. From equation (47b) it follows with $H := \dot{a}/a$ that

$$\delta_n - \delta_\varepsilon \propto a^{-2}. \tag{48}$$

Combining (46) and (48) it follows that the fluctuation in the kinetic energy density is given by

$$\delta_{\mathrm{kin}}(t,\boldsymbol{x}) \approx \delta_\varepsilon(t,\boldsymbol{x}) - \delta_n(t,\boldsymbol{x}) \approx \frac{3}{2}\frac{k_{\mathrm{B}}T_{(0)}(t)}{mc^2}\delta_T(t_{\mathrm{dec}},\boldsymbol{x}), \tag{49}$$

where $\delta_T(t_{\mathrm{dec}},\boldsymbol{x})$ is a constant, i.e., a free parameter of the perturbation theory and t_{dec} is the time of decoupling of matter and radiation. Moreover, it is found for the entropy perturbation and the fluctuation in the pressure, respectively,

$$s^{\mathrm{phys}}_{(1)} \approx \tfrac{3}{2}k_{\mathrm{B}}\delta_T, \quad \delta_p \approx \tfrac{5}{3}\delta_\varepsilon + \delta_T, \tag{50}$$

as follows from (28), (29) and (42).

Using that $\delta_\varepsilon(t,\boldsymbol{x}) = \delta_\varepsilon(t,\boldsymbol{q})\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x})$, $\delta_n(t,\boldsymbol{x}) = \delta_n(t,\boldsymbol{q})\exp(\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{x})$ with $|\boldsymbol{q}| = 2\pi/\lambda$ and the solutions of the background equations (7), the system (47) can be rewritten as one ordinary differential equation

$$\delta_\varepsilon'' + \frac{2}{\tau}\delta_\varepsilon' + \left[\frac{4}{9}\frac{\mu_\mathrm{m}^2}{\tau^{8/3}} - \frac{10}{9\tau^2}\right]\delta_\varepsilon = -\frac{4}{15}\frac{\mu_\mathrm{m}^2}{\tau^{8/3}}\delta_T(t_\mathrm{dec},\boldsymbol{q}), \quad \tau \geq 1, \tag{51}$$

where a prime denotes differentiation with respect to the dimensionless time $\tau := t/t_\mathrm{dec}$. The parameter μ_m is expressed in observable quantities at the present time t_p

$$\mu_\mathrm{m} = \frac{2\pi}{\lambda_\mathrm{dec}}\frac{1}{cH(t_\mathrm{p})}\frac{1}{\left[z(t_\mathrm{dec})+1\right]}\sqrt{\frac{5}{3}\frac{k_\mathrm{B}T_{(0)\gamma}(t_\mathrm{p})}{m}} = \frac{16.48}{\lambda_\mathrm{dec}}, \tag{52}$$

where λ_dec is expressed in parsec.

For large scales $\lambda_\mathrm{dec} \to \infty$, the solution of Eq. (51) is given by

$$\delta_\varepsilon = c_1\tau^{2/3} + c_2\tau^{-5/3}. \tag{53}$$

The first term is a well-known solution, whereas the gauge mode τ^{-1} is absent.

In order to solve Eq. (51) for all finite scales λ_dec, we have to specify the initial values $\delta(t_\mathrm{dec},\boldsymbol{q})$ and its derivative $\delta'(t_\mathrm{dec},\boldsymbol{q})$. From measurements of the Planck satellite[30] it is found that for $\tau = 1$, i.e., at $t = t_\mathrm{dec}$,

$$|\delta_\varepsilon(t_\mathrm{dec},\boldsymbol{q})| \lesssim 10^{-5}. \tag{54}$$

The initial growth rate is unknown. Therefore, we assume that there is initially no growth at all, i.e., we take

$$\delta_\varepsilon'(t_\mathrm{dec},\boldsymbol{q}) \approx 0. \tag{55}$$

The initial value of the entropy perturbation (50) is a random variable and depends on how a density perturbation undergoes the very fast transition from a very high pressure epoch just before decoupling to the very low pressure era just after decoupling. Some small areas end up with positive entropy perturbations and others with negative entropy perturbations. In fact, it is found that

(1) $s_{(1)}^\mathrm{phys} \approx \frac{3}{2}k_\mathrm{B}\delta_T > 0 \;\Rightarrow$ large voids

(2) $s_{(1)}^\mathrm{phys} \approx \frac{3}{2}k_\mathrm{B}\delta_T < 0 \;\Rightarrow$ structure formation

The results of the integration of Eq. (51) are depicted in Fig. 1. The curves are labeled by the diabatic pressure perturbation δ_T (50). Density perturbations start to grow at the time of decoupling. Each curve shows the time (or, equivalently, the redshift) when a density perturbation becomes non-linear, i.e., $\delta_\varepsilon = 1$. The growth of density perturbations is governed by both gravity and heat exchange. At a scale of 6.5 pc = 21 ly (the peak value) gravity and heat exchange work perfectly together, resulting in very early star formation. Because of the steepness of the curves to the left of the peak, I will refer to this value as the *relativistic* Jeans scale. The corresponding relativistic Jeans mass is $4.3\times10^3\,M_\odot$. For scales larger than the relativistic Jeans scale, gravity is strong but heat exchange is low because of the large

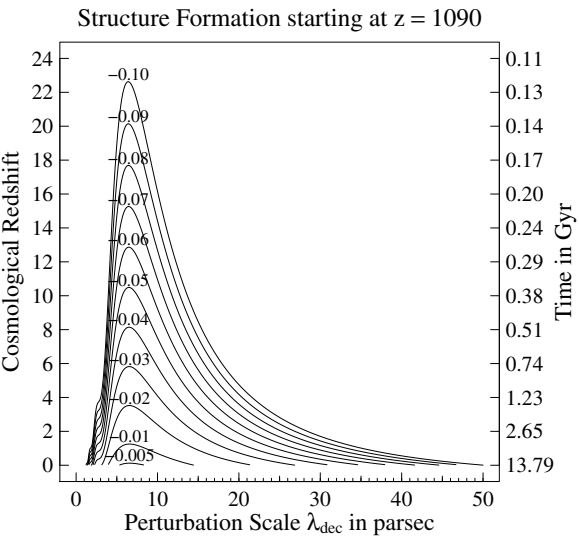


Fig. 1. The curves give the redshift and time, as a function of $\lambda_{\rm dec}$, when a linear perturbation in the energy density with initial values $\delta_\varepsilon(t_{\rm dec}, \boldsymbol{q}) \lesssim 10^{-5}$ and $\delta'_\varepsilon(t_{\rm dec}, \boldsymbol{q}) \approx 0$ starting to grow at an initial redshift of $z(t_{\rm dec}) = 1090$ has become non-linear, i.e., $\delta_\varepsilon(t, \boldsymbol{q}) = 1$. The curves are labeled with the initial values of the relative perturbations $\delta_T(t_{\rm dec}, \boldsymbol{q})$ in the diabatic part of the pressure. For each curve, the Jeans scale, i.e., the peak value, is at 6.5 pc.

scale, resulting in a later collapse time. The larger the scale, the later the collapse time. For very large scales the collapse time is proportional to $\tau^{2/3}$, the classic result for adiabatic density perturbations. On the other hand, for scales smaller than the relativistic Jeans scale, gravity becomes rapidly very weak, resulting also in a much later collapse time.

Figure 2 shows the growth rates of density perturbations as a function of their initial scale $\lambda_{\rm dec}$ at decoupling. The fastest growth occurs within the first 16 Myr. Density perturbations at the relativistic Jeans scale of 6.5 pc grow fast until non-linearity. Density perturbations with initial scale 4.1 pc grow initially faster than perturbations at the relativistic Jeans scale. Eventually, the rapid growth slows down and the perturbations grow slowly into the non-linear phase. Density perturbations which are initially smaller than 4.1 pc oscillate towards non-linearity. Finally, perturbations smaller than approximately 1 pc and larger than 50 pc do not become non-linear within the age of the universe.

11. Standard Perturbation Theory

The standard equation which is supposed to describe the evolution of density perturbations $\delta := \varepsilon_{(1)}/\varepsilon_{(0)}$ and which is ubiquitous in the literature reads

$$\ddot{\delta} + 2H\dot{\delta} - \left[\beta^2 \frac{\nabla^2}{a^2} + \tfrac{1}{2}\kappa\varepsilon_{(0)}(1+w)(1+3w)\right]\delta = 0. \tag{56}$$

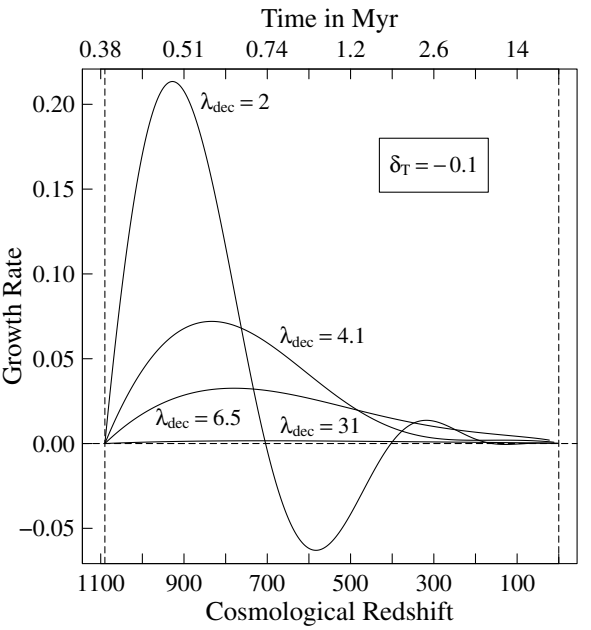


Fig. 2. The curves give the growth rates δ'_ε, as function of the redshift z, or time in million of years. The evolution of perturbations started at $z = 1090$. The initial dimensions λ_{dec} of the perturbations are measured in parsec.

See, for example, the classic article of Bonnor.[31] In this equation $\beta^2 = w = \frac{1}{3}$, (37), in the radiation-dominated universe, whereas $\beta = v_{\mathrm{s}}/c$ and $w = 0$, (45), after decoupling of matter and radiation. The standard equation (56) differs substantially from the new equations (38) and (47), since the latter two equations determine the evolution of the true, measurable energy density fluctuation $\delta_\varepsilon := \varepsilon_{(1)}^{\mathrm{phys}}/\varepsilon_{(0)}$. In this section we briefly explain why the standard equation differs from the new equations.

Since equations (38) and (47) are derived from the system (20), we try to derive Eq. (56) also from the system of equations (20), this time not using the definitions (9). In this procedure Eq. (20e) is not needed, since the equation of state for the pressure in Eq. (56) is of the form $p = p(\varepsilon)$.

11.1. *Radiation-dominated Universe*

The result for the radiation-dominated era is

$$\ddot{\delta} + 2H\dot{\delta} - \left[\frac{1}{3}\frac{\nabla^2}{a^2} + \tfrac{4}{3}\kappa\varepsilon_{(0)}\right]\delta = -\tfrac{4}{3}H\vartheta_{(1)}, \tag{57a}$$

$$\dot{\vartheta}_{(1)} + H\vartheta_{(1)} + \frac{1}{4}\frac{\nabla^2\delta}{a^2} = 0. \tag{57b}$$

The problem with this system is that it admits, next to physical solutions, also the gauge modes

$$\hat{\delta}(t,\boldsymbol{x}) := -4H(t)\psi(\boldsymbol{x}), \quad \hat{\vartheta}(t,\boldsymbol{x}) := -\frac{\nabla^2\psi(\boldsymbol{x})}{a^2(t)}. \tag{58}$$

Since these solutions have no physical significance, one cannot impose *physical* initial conditions to solve the system (57) since that would imply that the arbitrary gauge function $\psi(\boldsymbol{x})$ could be fixed by physical conditions, which is impossible. Furthermore, the gauge modes displace the divergence $\vartheta_{(1)}$ of the fluid velocity to the right-hand side, Eq. (57a). Since $\vartheta_{(1)}$ is an intrinsic property of an evolving density perturbation, it should be incorporated in the left-hand side of the evolution equation, as is the case for Eq. (38a). That is why Eq. (38a) predicts oscillating density perturbations with an *increasing* amplitude, whereas the standard equation (56) yields oscillating perturbations with a *constant* amplitude. Because of the presence of gauge modes in its solution, the system (57) is inadequate to study the evolution of density perturbations in the universe. In contrast to the system (57) the standard equation (56) lacks the divergence $\vartheta_{(1)}$. Since, in addition, the gauge mode $\hat{\delta}$ is a solution of Equation (56), the standard equation is simply useless.

11.2. *Era after Decoupling of Matter and Radiation*

In the era after decoupling we find

$$\ddot{\delta} + 2H\dot{\delta} - \left[\beta^2\frac{\nabla^2}{a^2} + \tfrac{1}{2}\kappa\varepsilon_{(0)}\right]\delta = -3H\beta^2\vartheta_{(1)}, \tag{59a}$$

$$\dot{\vartheta}_{(1)} + 2H\vartheta_{(1)} + \beta^2\frac{\nabla^2\delta}{a^2} = 0. \tag{59b}$$

In this case the gauge function $\psi(\boldsymbol{x})$ reduces to a constant C, (26), in accordance with the fact that after decoupling the cosmological fluid is described by a non-relativistic fluid (42). Therefore, the only surviving gauge mode is given by

$$\hat{\delta}(t,\boldsymbol{x}) := -3H(t)C, \tag{60}$$

and $\vartheta_{(1)} = \vartheta_{(1)}^{\text{phys}}$ is a physical quantity. Due to the presence of a gauge mode, the system (59) is still inadequate to study the evolution of density perturbations.

Finally, we consider a pressure-less fluid, also known as 'dust'. Substituting $p = 0$, i.e., $\beta = 0$, into the system (59) we find

$$\ddot{\delta} + 2H\dot{\delta} - \tfrac{1}{2}\kappa\varepsilon_{(0)}\delta = 0, \tag{61a}$$

$$\dot{\vartheta}_{(1)} + 2H\vartheta_{(1)} = 0. \tag{61b}$$

These equations are decoupled, implying that the divergence of the fluid velocity has no influence whatsoever on the growth of a perturbation. However, if there are no pressure differences in a fluid, then there is no fluid flow, $\vartheta_{(1)}^{\text{phys}} = 0$, so that density perturbations do not evolve. Since Eq. (61b) has only the gauge mode $\hat{\vartheta}_{(1)}$ as solution this equation is not needed anymore.

Equation (61a) follows also from the Newtonian Theory of Gravity if one takes the expansion into account. Since it has next to the gauge mode $\propto t^{-1}$ also a physical solution $\propto t^{2/3}$ regardless of the fact that there is no fluid flow, it should be clear that the Newtonian perturbation theory is inadequate to describe the evolution of density perturbations in the universe.

12. Conclusion

In this article I have sketched the derivation of a perturbation theory for open, flat and closed FLRW universes filled with perfect fluid described by a non-barotropic equation of state for the pressure $p(n, \varepsilon)$. The new theory has a Newtonian limit in a non-static universe. As a consequence, the gauge problem of cosmology is finally solved. The new perturbation theory consists of a second-order ordinary differential equation for the density contrast $\delta_\varepsilon := \varepsilon^{\text{phys}}_{(1)}/\varepsilon_{(0)}$, and a first-order ordinary differential equation for $\delta_n - \delta_\varepsilon/(1+w)$, where the latter expression is proportional to entropy perturbations. A first conclusion that can be drawn from the new perturbation theory is that local density perturbations do not affect the global expansion of the universe.

The new perturbation theory has been applied to a flat FLRW universe. The results are as follows. In the radiation-dominated universe density perturbations oscillate with an amplitude which is proportional to the square root of time. Both CDM as well as ordinary matter are coupled to perturbations in the radiation. As a consequence, structure formation can start only after decoupling of matter and radiation. After decoupling the growth of structure depends on both its internal gravity and its heat exchange with its environment. Heat exchange depends on the initial value of the entropy perturbation as well as the initial scale of a density perturbation. The entropy perturbation is a free parameter of the perturbation theory and can, therefore, take on random values. For sufficiently large negative values of the entropy perturbation I have found that the first stars are formed at $z \approx 24$–12, or, equivalently 110–300 Myr after the Big Bang. The relativistic Jeans scale is $6.5\,\text{pc} \approx 21\,\text{ly}$, and the corresponding relativistic Jeans mass is $4.3 \times 10^3\, M_\odot$.

The research in Refs. 3 and 4 has revealed that the General Relativistic perturbation theory explains structure formation in the early universe, without the need of CDM. The standard equation (56) of cosmology should not be used anymore.

References

1. E. M. Lifshitz, On the Gravitational Stability of the Expanding Universe, *J. Phys.* **X**, 116 (1946).
2. E. M. Lifshitz and I. M. Khalatnikov, Investigations in Relativistic Cosmology, *Adv. Phys.* **12**, 185 (1963).
3. P. G. Miedema, General Relativistic Evolution Equations for Density Perturbations in Closed, Flat and Open FLRW Universes, *eprint arXiv:1410.211* (2014).
4. P. G. Miedema, Structure Formation in the Early Universe, *ArXiv e-prints* (January 2016).

5. S. Weinberg, *Cosmology*, first edn. (Oxford University Press Inc., New York, 2008).
6. J. M. Bardeen, Gauge-invariant Cosmological Perturbations, *Phys. Rev.* **D22**, 1882 (1980).
7. G. F. R. Ellis and M. Bruni, Covariant and Gauge-invariant Approach to Cosmological Density Fluctuations, *Phys. Rev. D* **40**, 1804 (1989).
8. G. F. R. Ellis, J. Hwang and M. Bruni, Covariant and Gauge-independent Perfect-Fluid Robertson-Walker Perturbations, *Phys. Rev. D* **40**, 1819 (1989).
9. G. F. R. Ellis and H. van Elst, Cosmological Models (Cargèse lectures 1998) (1998).
10. V. F. Mukhanov, H. A. Feldman and R. H. Brandenberger, Theory of Cosmological Perturbations, *Phys. Rep.* **215**, 203 (1992).
11. V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, 2005).
12. M. Bruni, P. K. Dunsby and G. F. Ellis, Cosmological perturbations and the physical meaning of gauge invariant variables, *Astrophys.J.* **395**, 34 (1992).
13. K. A. Malik and D. Wands, Cosmological Perturbations, *Phys.Rept.* **475**, 1 (2009).
14. C. Knobel, An Introduction into the Theory of Cosmological Structure Formation, *eprint arXiv:1208.5931* (2012).
15. P. Peter, Cosmological Perturbation Theory, *eprint 1303.2509* (2013).
16. K. A. Malik and D. R. Matravers, Comments on gauge-invariance in cosmology, *General Relativity and Gravitation* **45**, 1989 (October 2013).
17. H. Kodama and M. Sasaki, Cosmological Perturbation Theory, *Progress of Theoretical Physics Supplements* **78**, 1 (1984).
18. C.-P. Ma and E. Bertschinger, Cosmological perturbation theory in the synchronous and conformal newtonian gauges, *Astrophys. J.* **455**, 7 (December 1995).
19. G. F. R. Ellis, Editorial note to: E. lifshitz, On the gravitational stability of the expanding universe, *General Relativity and Gravitation* **49** (2017).
20. J. W. York, jr., Covariant Decomposition of symmetric Tensors in the Theory of Gravitation, *Annales de l' Institut Henry Poincaré - Section A: Physique théorique* **XXI**, 319 (1974).
21. J. M. Stewart and M. Walker, Perturbations of Space-times in General Relativity, *Proc. R. Soc. London* **A 341**, 49 (1974).
22. J. M. Stewart, Perturbations of Friedmann-Robertson-Walker Cosmological Models, *Class. Quantum Grav.* **7**, 1169 (1990).
23. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Course of Theoretical Physics, Vol. 2, 4 edn. (Oxford: Butterworth-Heinemann (Elsevier), 2010).
24. S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (John Wiley & Sons, Inc. New York, 1972).
25. S. M. Carroll, *Space Time and Geometry*, first edn. (Benjamin Cummings, 2003).
26. P. J. Adams and V. Canuto, Exact Solution of the Lifshitz Equations Governing the Growth of Fluctuations in Cosmology, *Physical Review D* **12**, 3793 (1975).
27. D. W. Olson, Density Perturbations in Cosmological Models, *Physical Review D* **14**, 327 (1976).
28. E. W. Kolb and M. S. Turner, *The Early Universe* (Westview Press, 1994).
29. W. H. Press and E. T. Vishniac, Tenacious Myths about Cosmological Perturbations larger than the Horizon Size, *The Astrophysical Journal* **239**, 1 (July 1980).
30. Planck Collaboration, Planck 2018 results. VI. Cosmological parameters, *A&A* **641**, p. A6 (September 2020).
31. W. B. Bonnor, Jeans' Formula for Gravitational Instability, *Monthly Notices of the Royal Astronomical Society* **117**, 104 (02 1957).

Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data

M. Berti

SISSA – International School for Advanced Studies,
Via Bonomea 265, Trieste, 34136, Italy
INFN – National Institute for Nuclear Physics,
Via Valerio 2, Trieste, 34127, Italy
IFPU – Institute for Fundamental Physics of the Universe,
Via Beirut 2, Trieste, 34151, Italy
E-mail: mberti@sissa.it
https://www.sissa.it/

Observations of the 21cm signal through intensity mapping techniques are expected in the near future. This new observable will allow probing the evolution of the Universe in a very wide redshift range, from the dark ages, through the epoch of reionization up to the present time. We constrain cosmological parameters from forecast measurements of the 21 cm signal power spectrum $P_{21}(k, z)$ combined with the latest CMB data from Planck 2018 observations. In the same framework, we test also modified gravity models to unveil beyond ΛCDM features coupling information from primordial probes, such as the CMB, to lower redshift ones. We extend the codes EFTCAMB/EFTCosmoMC to compute the likelihood function for $P_{21}(k, z)$ and we construct a mock data set of forecast intensity mapping observations. At the time being, we are bound by the experimental state-of-the-art to consider the redshift bin $z = 0.39$. However, in the future, it may be possible to study also wider redshift ranges. We describe our likelihood implementation and present the results we obtained from the statistical Monte-Carlo Markov-Chain analysis we conducted.

Keywords: Cosmology; 21cm Intensity Mapping; Dark Energy; Modified Gravity; Forecast constraints.

1. Introduction

According to the Standard Model of Cosmology, the Λ - Cold Dark Matter (ΛCDM) model, the Universe we live in is currently going to a phase of accelerated expansion. The cause of acceleration is assigned to an enigmatic component, called Dark Energy (DE), which constitutes $\sim 70\%$ of the Universe. From the birth of the precision cosmology era, the ΛCDM model has been extensively tested and it proved to be in remarkable agreement with several, independent observations.[1–4] Despite its success, tensions on cosmological parameters[5–8] and open theoretical problems, regarding the nature of the Cosmological Constant (Λ),[9–11] motivated the emergence of a plethora of models alternative to ΛCDM.[12,13]

At the state-of-the-art, constraints on Modified Gravity (MG) or DE models are all broadly compatible with ΛCDM. However, observations from future experiments,

such as Euclid[a] and the Square Kilometer Array Observatory (SKAO),[b] are expected to be crucial in the quest for selecting the correct model.

In this work, we test the effect of the detection on the 21cm signal on beyond ΛCDM models, i.e. $P_{21}(k)$. We forecast intensity mapping observations of $P_{21}(k)$, as it will be measured by the MeerKAT radio-telescope,[c] the SKAO precursor and pathfinder in South Africa.[14–16] We build a realistic mock data set for $P_{21}(k)$ at redshift $z = 0.39$ and conduct a Monte Carlo Markov Chain (MCMC) analysis to constrain beyond ΛCDM theories. We adopt a model-independent approach by studying MG/DE within the Effective Field Theory (EFT) formalism for Cosmic Acceleration.[17,18] To this end we modify the codes `EFTCAMB/EFTCosmoMC`,[19–21] to include a new module for the $P_{21}(k)$ likelihood computation. We compute constraints on cosmological and EFT parameters from $P_{21}(k)$ alone and combined with the latest Planck 2018 CMB data.[1]

This work is intended as an introduction to a more extensive analysis, that can be found in Ref.[22]. The structure of the paper is the following. First, we review general aspects of the theoretical and experimental framework in section 1.1. We outline the main aspects of our methodology in section 2. Results are presented in section 3 and summarized in section 4.

1.1. *Theoretical and experimental framework*

The Effective Field Theory (EFT) technique was first used in a cosmological scenario to describe inflation[23,24] and then applied to the description of Cosmic Acceleration.[17,18] The idea behind EFT for DE is to construct the most general action to be *effective*, i.e. easily interfaced with data, and *unifying*, in order to include as many DE/MG models as special cases. Up to second order in perturbation theory and in the Jordan frame, the resulting action is

$$\begin{aligned} S = \int \mathrm{d}^4x\sqrt{-g}\Bigg\{ & \frac{m_0^2}{2}\left[1+\Omega^{\rm EFT}(\tau)\right]R + \Lambda(\tau) - c(\tau)a^2\delta g^{00} \\ & + \frac{M_2^4(\tau)}{2}\left(a^2\delta g^{00}\right)^2 - \frac{\bar{M}_1^3(\tau)}{2}a^2\delta g^{00}\delta K \\ & - \frac{\bar{M}_2^2(\tau)}{2}\left(\delta K\right)^2 - \frac{\bar{M}_3^2(\tau)}{2}\delta K^\mu_\nu \delta K^\nu_\mu \\ & + m_2^2(\tau)(g^{\mu\nu}+n^\mu n^\nu)\partial_\mu\left(a^2g^{00}\right)\partial_\nu\left(a^2g^{00}\right) \\ & + \frac{\hat{M}^2(\tau)}{2}a^2\delta g^{00}\delta R^{(3)} + \dots \Bigg\} + S_m, \end{aligned} \tag{1}$$

where the nine functions of the conformal time τ, $\{\Omega^{\rm EFT}, \Lambda, c\}$ and $\{M_2^4, \bar{M}_1^3, \bar{M}_2^2, \bar{M}_3^2, m_2^2, \hat{M}^2\}$, are referred to as EFT functions. We recover the

[a] See `https://sci.esa.int/web/euclid`.
[b] See `https://www.skatelescope.org/`.
[c] See `https://www.sarao.ac.za/gallery/meerkat/`.

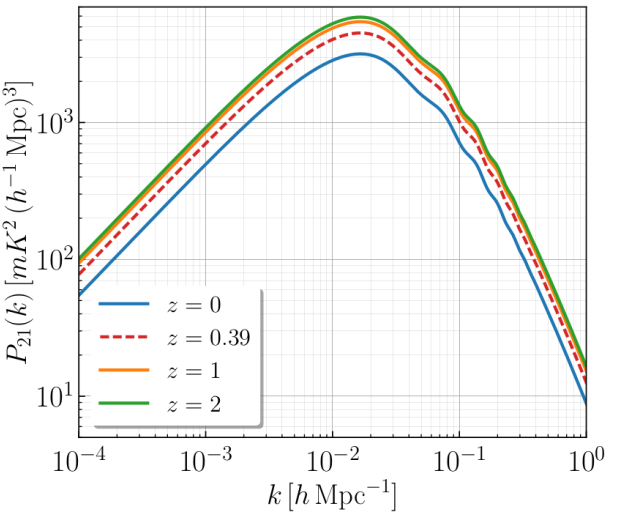


Fig. 1. Linear power spectra for the 21cm signal at different redshifts. Results are computed with the model in Eq. (2) for a ΛCDM fiducial cosmology, i.e. Planck 2018 best fit results.[1]

ΛCDM limit when all the EFT functions are set to zero. By choosing the evolution of each EFT function one can either test the specific effect of the single operators on observables (*pure* EFT approach) or map the EFT functions into well known DE/MG models. We will adopt the *pure* EFT approach and we will study a subset of models described by the functions $\{\Omega^{\rm EFT}, \Lambda, c\}$ on a ΛCDM background.

EFT models can be studied numerically by means of the Einstein-Boltzmann solver `EFTCAMB`,[19,21] an extension of the code `CAMB`,[25] and the MCMC sampler `EFTCosmoMC`,[20] an extension of `CosmoMC`.[26,27] *Pure* EFT models have been studied in the literature,[1,20,28] however, at the state-of-the-art, all the models are compatible with ΛCDM. For a complete review on EFT for DE and the latest constraints, we refer to Ref.[29].

Although no significant proof supporting a specific beyond ΛCDM model has yet been found, in this work we quantify the impact of future observations on EFT models. We focus on the 21cm signal detection through line intensity mapping techniques, which allows sampling the neutral hydrogen (HI) distribution in a wide redshift range.[30–35] These observations will open an important window on DE/MG theories.[36–42] Currently, several purpose-built radio-telescopes, such as the Canadian Hydrogen Intensity Mapping Experiment (CHIME),[d] the Hydrogen Intensity and Real-Time Analysis experiment (HIRAX),[e] the Five-hundred-meter Aperture Spherical Telescope (FAST),[f] are already taking data or are under construction.

[d]See `https://chime-experiment.ca/en`.
[e]See `https://hirax.ukzn.ac.za/`.
[f]See `https://fast.bao.ac.cn/`.

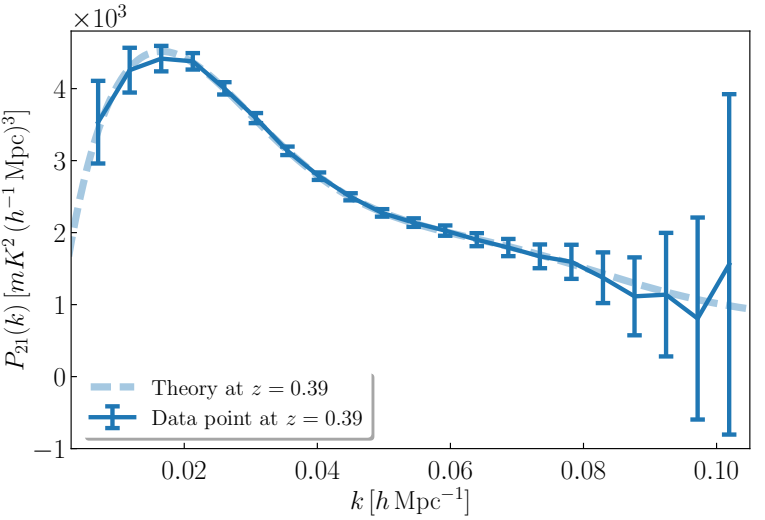


Fig. 2. Mock data set observations and theory prediction at $z = 0.39$ for the 21cm signal power spectrum. Error bars are generated considering a realistic intensity mapping survey with the MeerKAT telescope. Central values are instead obtained from points predicted by the theory for ΛCDM and the fiducial cosmology we assume. Each point from the theory is then displaced randomly according to a gaussian distribution with a fifth of the error on that point as standard deviation.

Moreover, radio cosmology and the study of DE/MG theories is one of the main science goals of the world's largest interferometer, the Square Kilometer Array Observatory (SKAO),[g] in particular with intensity mapping techniques.[43]

The MeerKAT telescope,[h] the SKAO precursor in South Africa, can be successfully exploited for 21cm intensity mapping[14,15] and has attained encouraging results.[16] In this work, we mimick observations with the MeerKAT telescope, which operates at an interesting redshift range for DE and for which real data will be available in the near future.

2. Likelihood implementation

To compute the likelihood function $\mathcal{L}(D|\boldsymbol{\alpha})$, which we assume to be a multivariate Gaussian, we need the theoretical expectations for the set of parameters $\boldsymbol{\alpha}$, describing the considered model, and the observed set of data D. To compute $\boldsymbol{\alpha}$ we extend the codes `EFCAMB/ETCosmoMC` and model the linear power spectrum of the 21cm signal as[35,43,44]

$$P_{21}(z,\,k) = \bar{T}_b^2(z)(b_{\mathrm{HI}}(z) + f(z))^2 P_{\mathrm{m}}(z,k), \tag{2}$$

[g]See `https://www.skatelescope.org/`.
[h]See `https://www.sarao.ac.za/gallery/meerkat/`.

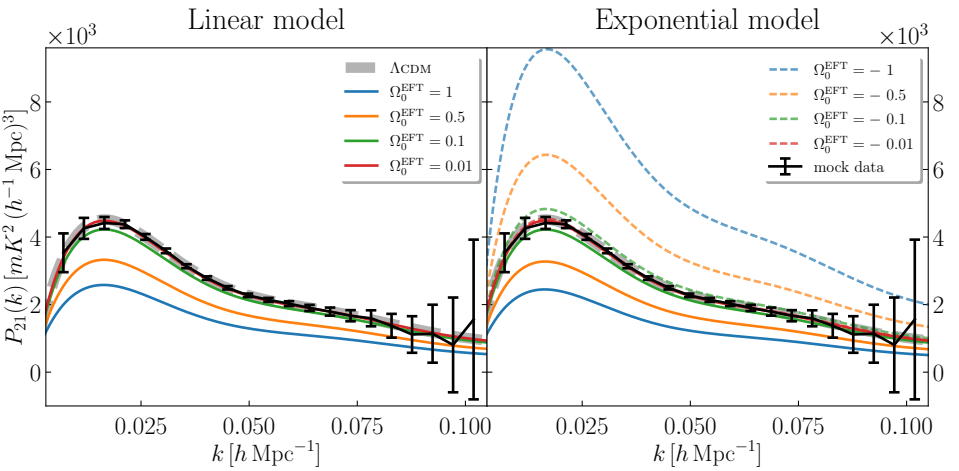


Fig. 3. Linear power spectra for the 21cm signal for ΛCDM and several *pure* EFT models, described by different values of the parameter $\Omega_0^{\rm EFT}$. We show theoretical predictions at redshift $z = 0.39$ for a linear (left panel) and an exponential (right panel) evolution of the function $\Omega_0^{\rm EFT}$. For the linear case we consider $\Omega_0^{\rm EFT} = 0.01, 0.1, 0.5, 1$, while for the exponential $\Omega_0^{\rm EFT} = \pm 0.01, \pm 0.1, \pm 0.5, \pm 1$. Black points with errorbars are the mock data set we construct in this work, as described in Sec. 2.

where T_b is the brightness temperature, $b_{\rm HI}$ is the HI bias, f is the linear growth rate and $P_{\rm m}$ is the linear matter power spectrum. We stress that the model dependence here enters through f and $P_{\rm m}$. Our predictions for $P_{21}(z, k)$ at different redshifts are shown in Fig. 1.

Given that real data are not available yet, we construct a mock data set of forecast observations of $P_{21}(k)$. We model the errors taking into account the technical specifications of the MeerKAT telescope, while we generate central points from the theory. Our mock data set is presented in Fig. 2. Here we focus on the redshift bin $z = 0.39$, which will be measured by MeerKAT. We explore also the effects of tomography in Ref.[22].

3. Results

In this section, we present our main results. We begin by studying a ΛCDM reference scenario. Then, we study a *pure* EFT model on a ΛCDM background, where the function $\Omega^{\rm EFT}(a)$ evolves as a linear function of the scale factor a, i.e. $\Omega^{\rm EFT}(a) = \Omega_0^{\rm EFT}\, a$. In Fig. 3, we compare theoretical predictions of $P_{21}(k)$ for different values of $\Omega_0^{\rm EFT}$ with the mock data set we constructed. One can see that the $P_{21}(k)$ observable is sensitive to variations in $\Omega_0^{\rm EFT}$, thus we expect that the $P_{21}(k)$ likelihood could help in constraining such model. However, we anticipate that this single data set alone is still not powerful enough to constrain $\Omega^{\rm EFT}$ alone. When combined with CMB data, instead, $\Omega_0^{\rm EFT}$ is already constrained by CMB to be < 0.1 (green line),

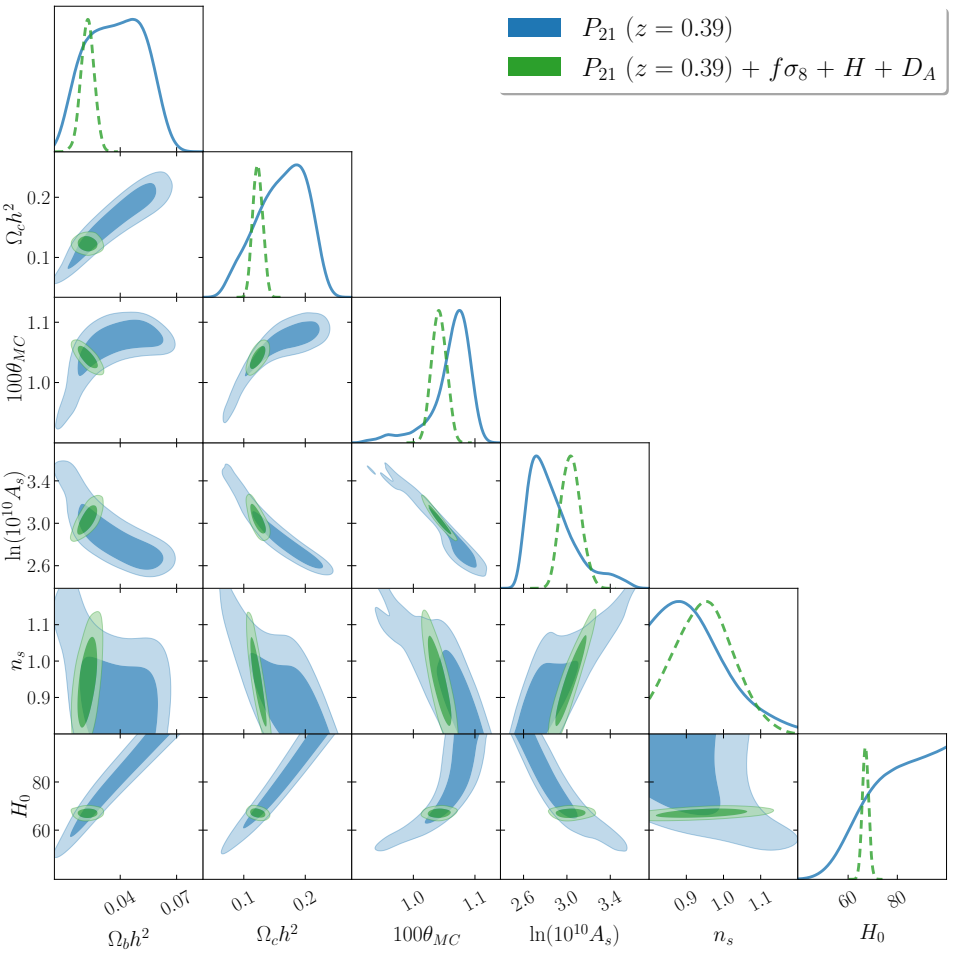


Fig. 4. Joint constraints (68% and 95% confidence regions) and marginalized posterior distributions on cosmological parameters for a ΛCDM model. Here the label $P_{21}(z = 0.39)$ stands for the 21cm power spectrum likelihood at redshift $z = 0.39$, while $f\sigma_8 + H + D_A$ represents the additional *background* likelihood computed at the same redshift $z = 0.39$. We consider the parameter τ fixed to a Planck 2018 fiducial cosmology value,[1] i.e. $\tau = 0.0543$.

therefore we will observe a small improvement in the constraining power when adding $P_{21}(k)$.

We analyse the effect of the $P_{21}(k)$ likelihood both alone and combined with CMB data. We consider the following data sets:

- *Planck 2018*: CMB data from Ref.[1], i.e. TT, TE, EE power spectra + low polarization (lowE) data + lensing;[i]

[i]See `http://pla.esac.esa.int/pla/\#home`.

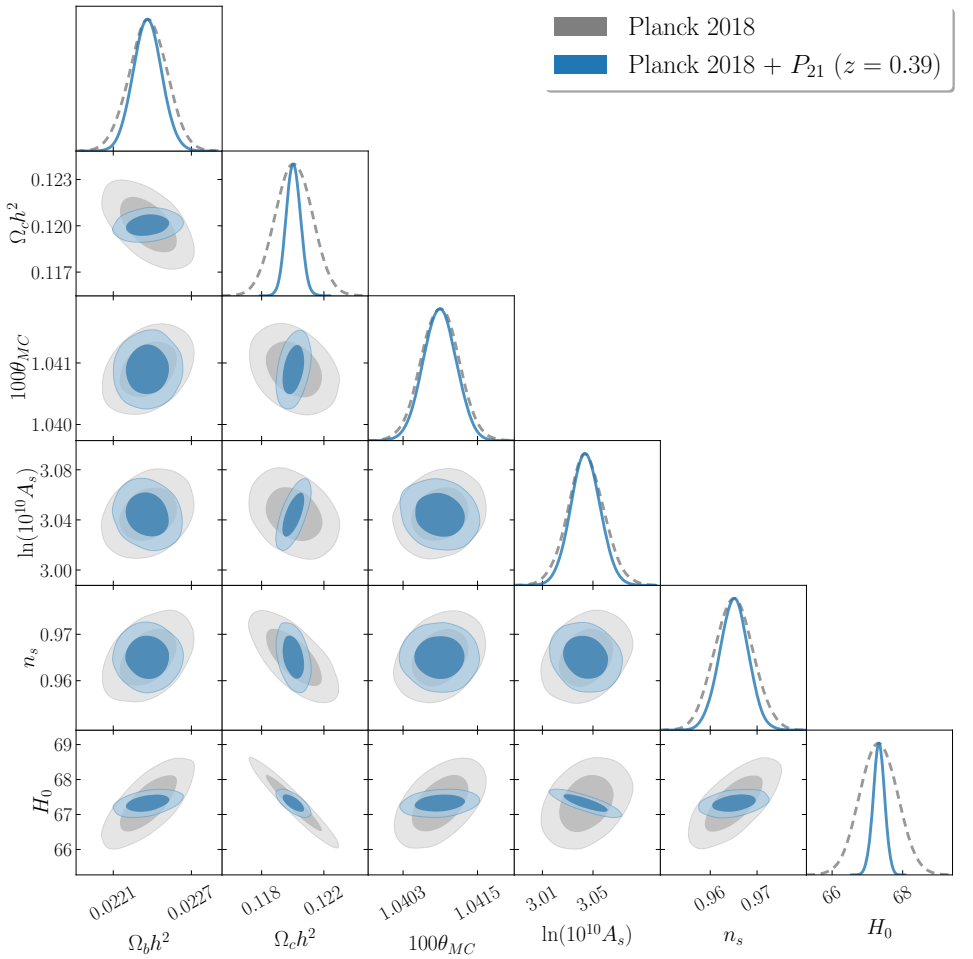


Fig. 5. Joint constraints (68% and 95% confidence regions) and marginalized posterior distributions on cosmological parameters for a ΛCDM model. Here the label Planck 2018 stands for TT, TE, EE + lowE + lensing while the label $P_{21}(z = 0.39)$ stands for the 21cm power spectrum likelihood at redshift $z = 0.39$.

Table 1. Marginalized constraints on cosmological parameters at the 68% confidence level for a ΛCDM model. Deviations in the error with respect to Planck 2018 results are shown in brackets.

Parameter	Planck 2018	$+ P_{21}(z = 0.39)$	$P_{21}(z = 0.39)$	$+ f\sigma_8 + H + D_A$
$\Omega_b h^2$	0.02237 ± 0.00014	0.02236 ± 0.00011 (−24%)	0.038 ± 0.015	0.0226 ± 0.0035
$\Omega_c h^2$	0.1201 ± 0.0012	0.12004 ± 0.00046 (−61%)	$0.162^{+0.050}_{-0.033}$	0.1227 ± 0.0081
n_s	0.9650 ± 0.0041	0.9651 ± 0.0031 (−25%)	< 0.959	$0.951^{+0.072}_{-0.085}$
H_0	67.32 ± 0.53	67.32 ± 0.16(−69%)	> 73.6	67.1 ± 1.3

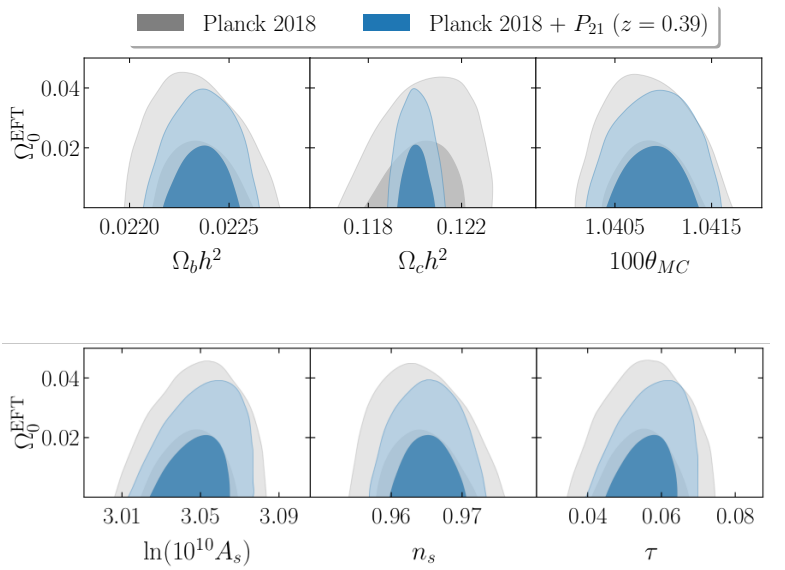


Fig. 6. Joint constraints (68% and 95% confidence regions) on cosmological and EFT parameters for a *pure* linear EFT model on a ΛCDM background. Here the label Planck 2018 stands for TT, TE, EE + lowE + lensing while the label $P_{21}(z = 0.39)$ stands for the 21cm power spectrum likelihood at redshift $z = 0.39$.

- $P_{21}(k)(z = 0.39)$: mock data set of MeerKAT forecast observations at redshift $z = 0.39$, constructed as described in Sec. 2;
- *background*: additional background and structure formation mock data sets for $f\sigma_8$, H, D_A at $z = 0.39$, inferred from higher redshift intensity mapping forecasts.[45]

For the ΛCDM case, we observe that the $P_{21}(k)$ likelihood alone has a weak constraining power on the cosmological parameters, as shown in Fig. 4. We find a strong correlation between H_0 and $\Omega_c h^2$, a feature that results to be pivotal when combining the 21cm signal data set with CMB data. The constraining power is significantly improved when we add the *background* data set. With $P_{21}(k)$ + *background* we are able to constrain the Hubble constant to $H_0 = 67.1 \pm 1.3$ km s^{-1}Mpc^{-1}, with a competitive error compared with results from other probes.[8] Combining CMB data with $P_{21}(k)$, we obtain an improvement on the errors on the cosmological parameters, as shown in Fig. 5 and Table 1. The effect is maximized for H_0 and $\Omega_c h^2$, for which we obtain a reduction on the error with respect to Planck estimate at the level of $\sim 60-70\%$. The correlation between H_0 and $\Omega_c h^2$ obtained with CMB data alone is perpendicular to the one found with $P_{21}(k)$. Therefore, combining the two data sets removes the correlation almost completely, thus significantly improving the constraints on H_0 and $\Omega_c h^2$.

Table 2. Marginalized constraints on cosmological parameters at the 68% confidence level for a *pure* linear EFT model on a ΛCDM background. Deviations in the error with respect to Planck 2018 results are shown in brackets.

Parameter	Planck 2018	$+ P_{21}(z = 0.39)$
$\Omega_b h^2$	0.02234 ± 0.00015	0.02237 ± 0.00011 (-26%)
$\Omega_c h^2$	0.1203 ± 0.0012	0.12003 ± 0.00046 (-63%)
H_0	67.21 ± 0.55	67.33 ± 0.17 (-70%)
$\Omega_0^{\rm EFT}$ ...	< 0.0147	< 0.0135 (-11%)

As anticipated, for the beyond ΛCDM case, we find that $P_{21}(k)$ alone has a weak constraining power on the EFT parameters. Thus, here we discuss only the results for $P_{21}(k)$ combined with CMB data, which are presented in Fig. 6 and Table 2. We find that adding $P_{21}(k)$ to CMB data reduces the error on the EFT parameter $\Omega_0^{\rm EFT}$ at the level of 11%. From the 2D contours, we observe that $P_{21}(k)$ also helps in removing some of the correlation between $\Omega_0^{\rm EFT}$ and the other cosmological parameters. Moreover, the constraints we obtain are comparable with results in the literature.[20,28] We stress that opening the parameter space to EFT models does not spoil the results on the cosmological parameters described above.

4. Conclusions

In this work we extended the `EFTCAMB/EFTCosmoMC` codes by implementing a new likelihood module fully integrated with original codes, to test forecast 21cm Intensity Mapping observations. We constructed a realistic data set at $z = 0.39$ from future MeerKAT-like observation settings. We tested the impact of the $P_{21}(k)$ likelihood on DE/MG models in the EFT framework.

We find a significant improvement on the $\Omega_c h^2$, H_0 constraints, from P_{21} combined with Early Universe probes, i.e. Planck 2018 CMB data.[1] The impact of $P_{21}(k)$ on beyond ΛCDM models is at the level of 11%. Thus, from this analysis, we conclude that 21cm power spectrum observations could help in increasing our knowledge of DE. However, we believe that adding tomographic data set of $P_{21}(k)$ observations could significantly improve the constraining power on EFT parameters. We investigate this claim and further expand this analysis in Ref.[22].

Acknowledgments

The author would like to thank her supervisor Prof. Matteo Viel and her collaborators Marta Spinelli, Sandeep B. Haridasu, and Alessandra Silvestri. We thank José Fonseca, for his valuable help in the construction of the MeerKAT-like $P_{21}(k)$ mock data, and Marco Raveri for useful discussions and technical help. This work is supported by the INFN INDARK grant.

References

1. Planck 2018 results, Vi. cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
2. T. M. C. Abbott *et al.*, Dark Energy Survey Year 1 Results: Constraints on Extended Cosmological Models from Galaxy Clustering and Weak Lensing, *Phys. Rev. D* **99**, p. 123505 (2019).
3. H. Hildebrandt *et al.*, KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing, *Mon. Not. Roy. Astron. Soc.* **465**, p. 1454 (2017).
4. A. G. Riess *et al.*, Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant, *Astrophys. J.* **861**, p. 126 (2018).
5. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *Astrophys. J.* **876**, p. 85 (2019).
6. K. C. Wong *et al.*, H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes, *Mon. Not. Roy. Astron. Soc.* **498**, 1420 (2020).
7. W. L. Freedman *et al.*, The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch (7 2019).
8. L. Verde, T. Treu and A. G. Riess, Tensions between the Early and the Late Universe, *Nature Astron.* **3**, p. 891 (7 2019).
9. S. M. Carroll, The Cosmological Constant, *Living Rev. Rel.* **4**, p. 1 (2001).
10. S. Weinberg, The Cosmological Constant Problem, *Rev. Mod. Phys.* **61**, 1 (1989).
11. P. Bull *et al.*, Beyond ΛCDM: Problems, solutions, and the road ahead, *Phys. Dark Univ.* **12**, 56 (2016).
12. P. G. Ferreira, Cosmological Tests of Gravity, *Ann. Rev. Astron. Astrophys.* **57**, 335 (2019).
13. M. Ishak, Testing General Relativity in Cosmology, *Living Rev. Rel.* **22**, p. 1 (2019).
14. M. G. Santos *et al.*, Cosmology from a SKA HI intensity mapping survey, *PoS* **AASKA14**, p. 019 (2015).
15. M. G. Santos *et al.*, MeerKLASS: MeerKAT Large Area Synoptic Survey (9 2017).
16. J. Wang *et al.*, HI intensity mapping with MeerKAT: Calibration pipeline for multi-dish autocorrelation observations (11 2020).
17. G. Gubitosi, F. Piazza and F. Vernizzi, The Effective Field Theory of Dark Energy, *JCAP* **02**, p. 032 (2013).
18. J. Bloomfield, É. É. Flanagan, M. Park and S. Watson, Dark energy or modified gravity? An effective field theory approach, *Journal of Cosmology and Astro-Particle Physics* **2013**, p. 010 (Aug 2013).
19. B. Hu, M. Raveri, N. Frusciante and A. Silvestri, *Effective Field Theory of Cosmic Acceleration: An implementation in CAMB*, *Phys. Rev. D* **89**, p. 103530 (2014).
20. M. Raveri, B. Hu, N. Frusciante and A. Silvestri, *Effective Field Theory of Cosmic Acceleration: Constraining dark energy with CMB data*, *Phys. Rev. D* **90**, p. 043513 (2014).
21. B. Hu, M. Raveri, N. Frusciante and A. Silvestri, *EFTCAMB/EFTCosmoMC: Numerical Notes v3.0* (5 2014).
22. M. Berti, M. Spinelli, B. S. Haridasu, M. Viel and A. Silvestri, Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data (9 2021).

23. P. Creminelli, M. A. Luty, A. Nicolis and L. Senatore, Starting the Universe: Stable Violation of the Null Energy Condition and Non-standard Cosmologies, *JHEP* **12**, p. 080 (2006).
24. C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan and L. Senatore, The Effective Field Theory of Inflation, *JHEP* **03**, p. 014 (2008).
25. A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538**, 473 (2000).
26. A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: A Monte Carlo approach*, *Phys. Rev. D* **66**, p. 103511 (2002).
27. A. Lewis, Efficient sampling of fast and slow cosmological parameters, *Phys. Rev. D* **87**, p. 103529 (2013).
28. P. . results, XIV. Dark energy and modified gravity, *Astron. Astrophys.* **594**, p. A14 (2016).
29. N. Frusciante and L. Perenon, Effective field theory of dark energy: A review, *Phys. Rept.* **857**, 1 (2020).
30. S. Bharadwaj, B. B. Nath, B. B. Nath and S. K. Sethi, Using HI to probe large scale structures at z $\sim$ 3, *J. Astrophys. Astron.* **22**, p. 21 (2001).
31. S. Bharadwaj and S. K. Sethi, HI Fluctuations at Large Redshifts. 1. Visibility correlation, *J. Astrophys. Astron.* **22**, p. 293 (2001).
32. R. A. Battye, R. D. Davies and J. Weller, Neutral hydrogen surveys for high redshift galaxy clusters and proto-clusters, *Mon. Not. Roy. Astron. Soc.* **355**, 1339 (2004).
33. M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist and S. R. Furlanetto, Cosmological parameter estimation using 21 cm radiation from the epoch of reionization, *Astrophys. J.* **653**, 815 (2006).
34. E. D. Kovetz *et al.*, Line-Intensity Mapping: 2017 Status Report (9 2017).
35. F. Villaescusa-Navarro *et al.*, *Ingredients for 21 cm Intensity Mapping*, *Astrophys. J.* **866**, p. 135 (2018).
36. T.-C. Chang, U.-L. Pen, J. B. Peterson and P. McDonald, Baryon Acoustic Oscillation Intensity Mapping as a Test of Dark Energy, *Phys. Rev. Lett.* **100**, p. 091303 (2008).
37. A. Hall, C. Bonvin and A. Challinor, Testing General Relativity with 21-cm intensity mapping, *Phys. Rev. D* **87**, p. 064026 (2013).
38. N. A. Lima, V. Smer-Barreto and L. Lombriser, Constraints on decaying early modified gravity from cosmological observations, *Phys. Rev. D* **94**, p. 083507 (2016).
39. P. Brax, S. Clesse and A.-C. Davis, Signatures of Modified Gravity on the 21-cm Power Spectrum at Reionisation, *JCAP* **01**, p. 003 (2013).
40. G. Zhao, D. Bacon, R. Maartens, M. Santos and A. Raccanelli, Model-independent constraints on dark energy and modified gravity with the SKA, *PoS* **AASKA14**, p. 165 (2015).
41. C. Heneka and L. Amendola, General Modified Gravity With 21cm Intensity Mapping: Simulations and Forecast, *JCAP* **10**, p. 004 (2018).
42. C. B. V. Dash and T. G. Sarkar, Constraining dark energy using the cross correlations of weak lensing with post-reionization probes of neutral hydrogen, *JCAP* **02**, p. 016 (2021).
43. S. C. SWG, *Cosmology with Phase 1 of the Square Kilometre Array: Red Book 2018: Technical specifications and performance forecasts*, *Publ. Astron. Soc. Austral.* **37**, p. e007 (2020).
44. N. Kaiser, *Clustering in real space and in redshift space*, *Mon. Not. Roy. Astron. Soc.* **227**, 1 (1987).
45. A. Obuljen, E. Castorina, F. Villaescusa-Navarro and M. Viel, High-redshift post-reionization cosmology with 21cm intensity mapping, *JCAP* **05**, p. 004 (2018).

Entropy and irreversible processes in gravity and cosmology

Llorenç Espinosa-Portalés* and Juan García-Bellido†
Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid
Cantoblanco, 28049 Madrid, Spain
**E-mail: llorenc.espinosa@uam.es*
†*E-mail: juan.garciabellido@uam.es*

General Relativity is *a priori* a theory invariant under time reversal. Its integration with the laws of thermodynamics allows for a formulation of non-equilibrium phenomena in gravity and the introduction of an arrow of time, i.e. the the breaking of such invariance. Even though most of the evolution of the universe takes place in local thermal equilibrium, the effects of irreversible processes on the expansion via entropic forces may be phenomenologically relevant. We review our previous work on the covariant formulation of non-equilibrium thermodynamics in General Relativity and the proposal to explain the recent cosmic acceleration from it.

Keywords: General Relativity, thermodynamics, non-equilibrium phenomena, entropic forces, cosmic acceleration.

1. Introduction

General Relativity is an extremely successful physical theory. More than a century after its formulation, its predictions continue to be valid at all probed scales, albeit an extension at short scales will be required in order to obtain a UV-complete quantum theory of gravity and resolve space-time singularities.

Thermodynamics is an even older discipline. Its fundamental laws seem to resist the passage of time and are still of great relevance today. On the one hand, thermodynamics may help in building the bridge between classical and quantum gravity, as the laws of black hole thermodynamics point towards the existence of a microphysical description of gravity yet to be understood.

On the other hand, the second law of thermodynamics, i.e. the growth of entropy, dictates the sign of the arrow of time. Physical laws are usually invariant under time inversion. The increase in entropy with time in out-of-equilibrium phenomena, however, allows one to distinguish the future-directed from the past directed description of a physical process.

There lacks a consistent and rigorous integration between General Relativity and the laws of thermodynamics. Understanding the very notion of the arrow of time is of particular interest for cosmology. In section 2 we argue for the need of going beyond reversible cosmology. This can be achieved for any space-time metric using variational techniques.[1] We review our main results in this new approach to non-equilibrium thermodynamics in General Relativity and present them in sections 3 to 6.

The growth of entropy associated to the causal horizon in open inflation scenarios may explain the current accelerated expansion of the universe within the general relativistic entropic acceleration (GREA) theory.[2] We briefly describe how this mechanism works in section 7 and finish with our conclusions.

2. Reversible cosmology

Let us begin the study of the problem of reversibility in gravity and cosmology by reviewing a prototypic case of reversible gravitational system: a homogeneous and isotropic universe. It is described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2 \right) , \tag{1}$$

where $a(t)$ is the scale factor, $k = -1, 0, 1$ is the curvature parameter corresponding to, respectively, an open, flat and closed universe; and $d\Omega_2^2$ is the solid angle element. This space-time is filled with a perfect fluid, described by the stress-energy tensor

$$T_{\mu\nu} = (\rho + p) u_\mu u_\nu + p g_{\mu\nu} , \tag{2}$$

where ρ and p are, respectively, the density and pressure of the fluid. The Einstein field equations for this metric and matter content deliver the dynamics for the scale factor, the well-known Friedmann equations

$$H^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho , \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) , \tag{3}$$

where $H = \dot{a}/a$ is the Hubble parameter. There is a constraint on the stress-energy tensor due to the Einstein field equations and the Bianchi identities, namely its covariant conservation $D_\mu T^{\mu\nu} = 0$. From this constraint one can derive the continuity equation

$$\dot{\rho} + 3H(\rho + p) = 0 . \tag{4}$$

However, one can also derive this equation from the second law of thermodynamics. Indeed, changes in entropy are related to changes in internal energy and work $TdS = \delta U + \delta W$. If we apply this to a region of fixed comoving volume $a(t)^3$ we get

$$T\frac{dS}{dt} = \frac{d}{dt}\left(\rho a^3\right) + p\frac{d}{dt}\left(a^3\right) . \tag{5}$$

If the expansion of the universe is reversible, we can set the LHS to 0 and recover the continuity equation. However, this is only true in thermodynamical equilibrium and, in general, entropy is a monotonically increasing function of time. Most of the expansion history of the universe is indeed adiabatic. However, it is out-of-equilibrium at certain key points such as (p)reheating, phase transitions or gravitational collapse. Allowing for a time-varying entropy implies the addition of a term in the continuity equation

$$\dot{\rho} + 3H(\rho + p) = \frac{T\dot{S}}{a^3} . \tag{6}$$

Combining this with the first Friedmann equation we obtain a modified, non-equilibrium second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p - \frac{T\dot{S}}{a^3 H}\right) . \tag{7}$$

In principle, this evolution equation does not seem to be compatible with the Einstein field equations. In order to achieve that, the laws of thermodynamics need to be rigorously incorporated into the computation of the equations of motion. This can be achieved by applying the variational formalism of non-equilibrium thermodynamics, developed in another context by Gay-Balmaz and Yoshimura,[3,4] to General Relativity.

3. Entropic forces in mechanics and field theory

Entropic forces emerge naturally in any physical system out of equilibrium. They are a consequence of the coarse-graining of physical degrees of freedom and the laws of thermodynamics, which impose entropy to be a monotonically increasing function of time. This breaks time reversibility.

The dynamics of the coarse-grained degrees of freedom is unknown or ignored and so they do not appear in the action of the physical system. It would seem that a variational treatment of an out-of-equilibrium system is not possible. However, both the extremal-action principle and the second law of thermodynamics can be merged consistently by imposing the latter as a constrain on the variational problem defined by the action.[3,4] On the other hand, the first law of thermodynamics is obtained from the symmetries of the problem.

3.1. *Entropic forces in mechanics*

Let us start by reviewing the emergence of entropic forces in a mechanical system. Consider the action

$$\mathcal{S} = \int dt L(q, \dot{q}, S) , \tag{8}$$

where the Lagrangian depends on the generalized coordinate $q(t)$, its time derivative $\dot{q}(t)$ and the entropy $S(t)$. The variation of the action gives

$$\delta\mathcal{S} = \int dt \left(\frac{\delta L}{\delta q}\delta q + \frac{\partial L}{\partial S}\delta S\right) . \tag{9}$$

Setting $\delta\mathcal{S} = 0$ defines the variational problem. In order to enforce the second law of thermodynamics, we need to impose the variational constraint

$$\frac{\partial L}{\partial S}\delta S = f\delta q , \tag{10}$$

which simply states the relationship between variations of the entropy and the generalized coordinate. If we plug this in the variation of the action, then we can

readily obtain the equations of motion

$$\frac{\delta L}{\delta q} = \frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = -f\,, \tag{11}$$

which is the Euler-Lagrange equation modified by the addition of the entropic force f. In order to know its precise form, we need to impose as well the phenomenological constraint, which is obtained by formally replacing the variations with time derivatives

$$\frac{\partial L}{\partial S}\dot{S} = f\dot{q}\,. \tag{12}$$

Note that usually temperature can be introduced as

$$T = -\frac{\partial L}{\partial S} > 0 \tag{13}$$

providing a clearer meaning to the variational and phenomenological constraint.

The imposition of the constraints and the emergence of an entropic force break the symmetry under time inversion. First, note that formally the Euler-Lagrange equation is invariant under the $t \to -t$ transformation, even with a non-vanishing entropic force. Next, the positiveness of the temperature and the change of entropy imposes $f\dot{q} < 0$. The entropic force has qualities of a generalized friction, as it opposes the coordinate velocity. Now, if one performs time inversion, this sign constraint becomes $f\dot{q} > 0$, as the temperature remains positive but entropy decreases with time. Before solving the equations of motion, $\dot{q}$ is still a degree of freedom and, thus, one must conclude a flip in the sign of f, which flips the overall sign of the Euler-Lagrange equation. Hence, one concludes that the emergence of an entropic force breaks symmetry under time reversal. The evolution of the system becomes irreversible.

3.2. *Entropic forces in classical field theory*

The extension of the variational formalism of non-equilibrium thermodynamics to the continuum is somewhat involved. We present a short-cut derivation that relies on the introduction of an additional constraint. We refer the reader to the appendix of Ref. 1 to check its equivalence with the full variational derivation originally presented in Ref. 4.

The action of a scalar field on Minkowski space-time contains now a dependency on a scalar function $s(t, \vec{x})$ that encodes information related to coarse-grained degrees of freedom

$$\mathcal{S} = \int d^4x \mathcal{L}\left(\phi, \partial_\mu \phi, s\right)\,. \tag{14}$$

In a similar fashion as before, the extremal-action principle needs to be supplemented by a variational constraint

$$\frac{\partial \mathcal{L}}{\partial s}\delta s = f\delta\phi\,, \tag{15}$$

so that the equation of motion becomes

$$\frac{\delta\mathcal{L}}{\delta\phi} = \frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} = -f\,, \tag{16}$$

which is nothing but the Euler-Lagrange equation of a scalar field with a new term of entropic origin. As before, one can usually introduce a temperature as

$$T = -\frac{\partial\mathcal{L}}{\partial s} > 0 \tag{17}$$

and we would like to interpret s as the entropy density. However, due to spatial fluxes entropy does not need to necessarily increase locally and, thus, the constraint would not have a fixed sign. Before proceeding, we inspect the corresponding phenomenological constraint

$$\frac{\partial\mathcal{L}}{\partial s}\partial_0 s = f\partial_0\phi\,. \tag{18}$$

Instead, we interpret δs and $\partial_0 s$ as local entropy production and introduce a new function $s_{\rm tot}$ which is the actual entropy density and whose changes $\delta s_{\rm tot}$ and $\partial_0 s_{\rm tot}$ are indeed total local changes of the entropy density. Both are related as

$$\partial_0 s = \partial_0 s_{\rm tot} - \partial_i j_s^i\,, \tag{19}$$

where j_s^i is the entropy flux. This latter equation is an additional constraint we impose for the variational formalism to be consistent. Now one can check that $f\partial_0\phi < 0$ and time reversibility is broken by the same argument used in the mechanics example.

3.3. *Entropic forces in presence of additional symmetries*

The generalization of the above formalism to higher order tensors or to representations of some internal symmetry group is straightforward. Let us consider a field tensor z of contravariant rank r, which is also in some representation of an internal symmetry labelled by an index A. Then one builds the variational constraint as

$$\frac{\partial\mathcal{L}}{\partial s}\delta s = f_{A;\mu_1,\ldots\mu_r}\delta z_A^{\mu_1,\ldots\mu_r}\,, \tag{20}$$

which delivers the equation of motion

$$\frac{\delta\mathcal{L}}{\delta z_A^{\mu_1,\ldots\mu_r}}, = -f_{A;\mu_1,\ldots\mu_r}\,. \tag{21}$$

4. Entropic forces in General Relativity

The previous discussion makes us ready to study entropic forces in General Relativity. However, with the introduction of a dynamical space-time the very notion of time evolution becomes non-trivial. As we will see shortly, it is possible to obtain a modification of entropic origin to Einstein's field equation in the Lagrangian formulation of General Relativity. Its proper interpretation will require, nevertheless, the use of the Hamiltonian formalism.

4.1. *Lagrangian formulation*

Let us consider the Einstein-Hilbert action plus a matter term

$$\mathcal{S} = \frac{1}{2\kappa}\int d^4x\sqrt{-g}R + \int d^4x\mathcal{L}_m(g_{\mu\nu}, s)\,, \tag{22}$$

where $\kappa = 8\pi G$ is the gravitational coupling and we allow for the dependence of the matter Lagrangian on a function $s(t, \vec{x})$, which will have a similar interpretation to the one presented in the previous section. The extremal-action principle is supplemented by the variational constraint.

$$\frac{\partial\mathcal{L}_m}{\partial s} = \frac{1}{2}\sqrt{-g}f_{\mu\nu}\delta g^{\mu\nu}\,. \tag{23}$$

From the extremal-action principle and the variational constraint we obtain the modified Einstein's field equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa\left(T_{\mu\nu} - f_{\mu\nu}\right)\,, \tag{24}$$

where there is an additional term $f_{\mu\nu}$ of entropic origin. In order to obtain an expression for this term and to check the breaking of symmetry under time inversion we need to work in the Hamiltonian formulation of General Relativity.

4.2. *Hamiltonian formulation*

General Relativity admits a Hamiltonian formulation in the Arnowitt-Deser-Misner (ADM) formalism. Space-time is foliated in constant time hypersurfaces with normal unit vector n^μ, being the 4-metric split as

$$g_{\mu\nu} = h_{\mu\nu} - n_\mu n_\nu\,, \tag{25}$$

where $h_{\mu\nu}$ is the 3-metric induced on the hypersurfaces. Analogously, one can parametrize the 4-metric in terms of the 3-metric h_{ij} and the lapse and shift functions N and N^i

$$ds^2 = -(Ndt)^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt)\,. \tag{26}$$

Greek indices run from 0 to 3, while Latin ones do from 1 to 3 and are raised and lowered by h_{ij}. The normal vector can be written as

$$n_\mu = (-N, 0, 0, 0)\,. \tag{27}$$

Note that h_{ij} is the purely spatial part of $h_{\mu\nu}$ and is also the pull-back of $g_{\mu\nu}$ onto the hypersurface.

The Einstein-Hilbert action for this parametrization of the metric is given by the following gravitational Lagrangian:

$$\mathcal{L}_G = \sqrt{-g}R = \frac{1}{2\kappa}N\sqrt{h}\left({}^{(3)}R + K_{ij}K^{ij} - K^2\right)\,, \tag{28}$$

where K_{ij} is the extrinsic curvature of the 3-hypersurface Σ and is given by the Lie derivative along the normal vector n

$$K_{ij} = \frac{1}{2}\pounds_n h_{ij} = \frac{1}{2N}\left(\partial_0 h_{ij} - \nabla_i N_j - \nabla_j N_i\right) , \tag{29}$$

where ∇ denotes the covariant derivative on Σ with respect to the 3-metric h_{ij}. Its trace and traceless part are:

$$\begin{aligned} K &= h^{ij}K_{ij} = \frac{1}{N}\left(\partial_0 \ln\sqrt{h} - \nabla_i N^i\right) \\ \bar{K}_{ij} &= K_{ij} - \frac{1}{3}K h_{ij} . \end{aligned} \tag{30}$$

Unlike the intrinsic curvature, described by the Riemann tensor $R^{\rho}_{\mu\nu\lambda}$ and its contractions, the extrinsic curvature is a quantity that depends on the embedding of a surface in a larger manifold.

We are now ready to introduce the Hamiltonian formulation of the theory. Note that the only quantity whose time derivative appears in the gravitational Lagrangian is the 3-spatial metric h_{ij} and, thus, it is the only dynamical or propagating d.o.f. Correspondingly, one defines its conjugate momentum as:

$$\Pi^{ij} = \frac{\partial \mathcal{L}_G}{\partial \dot{h}_{ij}} = \sqrt{h}\left(K^{ij} - K h^{ij}\right) . \tag{31}$$

With this, the gravitational Lagrangian can be rewritten as

$$\begin{aligned} \mathcal{L}_G &= N\sqrt{h}^{(3)}R - \frac{N}{\sqrt{h}}\left(\Pi_{ij}\Pi^{ij} - \frac{1}{2}\Pi^2\right) - 2\Pi^{ij}\nabla_i N_j \\ &= \Pi^{ij}\dot{h}_{ij} - N\mathcal{H} - N_i\mathcal{H}^i - 2\nabla_i\left(\Pi^{ij}N_j\right) , \end{aligned} \tag{32}$$

where $\Pi = h_{ij}\Pi^{ij}$ and we introduced the functions:

$$\begin{aligned} \mathcal{H} &= -\sqrt{h}^{(3)}R + \frac{1}{\sqrt{h}}\left(\Pi_{ij}\Pi^{ij} - \frac{1}{2}\Pi^2\right) \\ \mathcal{H}^i &= -2\nabla_j\left(h^{-1/2}\Pi^{ij}\right) . \end{aligned} \tag{33}$$

Since N and N_i are not dynamical variables, they merely enter the gravitational Lagrangian as Lagrange multipliers. One defines the gravitational Hamiltonian as:

$$\begin{aligned} \mathcal{H}_G &= \Pi^{ij}\dot{h}_{ij} - \mathcal{L}_G \\ &= N\mathcal{H} + N_i\mathcal{H}^i + \nabla_i\left(\Pi^{ij}N_j\right) , \end{aligned} \tag{34}$$

with the Hamiltonian and momentum constraints:

$$\begin{aligned} \frac{\delta\mathcal{H}_G}{\delta N} &= \mathcal{H} = 0 \\ \frac{\delta\mathcal{H}_G}{\delta N_i} &= \mathcal{H}^i = 0 . \end{aligned} \tag{35}$$

The Hamiltonian evolution equations are obtained from the variations of the action with respect to the metric and conjugate momentum

$$\delta\mathcal{S} = \int d^4x \left[\left(-\dot{\Pi}^{ij} - \frac{\delta\mathcal{H}_G}{\delta h_{ij}} - \kappa N\sqrt{h}\tilde{f}^{ij} + 2\kappa\frac{\partial\mathcal{L}_m}{\partial h_{ij}} \right)\delta h_{ij} + \left(\dot{h}_{ij} - \frac{\delta\mathcal{H}_G}{\delta\Pi^{ij}} \right)\delta\Pi^{ij} \right], \tag{36}$$

where the variational constraint

$$\frac{\partial\mathcal{L}_m}{\partial s}\delta s = -\frac{1}{2}N\sqrt{h}\tilde{f}^{ij}\delta h_{ij} \tag{37}$$

was already implemented. Note that the minus sign arises from

$$\frac{\partial h_{ij}}{\partial h^{kl}} = -\frac{1}{2}\left(h_{ik}h_{jl} + h_{il}h_{jk}\right) \tag{38}$$

and that $\tilde{f}_{ij}$ is the pull-back of $f_{\mu\nu}$ onto the hypersurfaces. By setting the variation to 0 we obtain the two Hamilton equations

$$\begin{aligned} \frac{\delta\mathcal{H}_G}{\delta h_{ij}} &= -\dot{\Pi}^{ij} - \kappa N\sqrt{h}f^{ij} + 2\kappa\frac{\partial\mathcal{L}_m}{\partial h_{ij}} \\ \frac{\delta\mathcal{H}_G}{\delta\Pi^{ij}} &= \dot{h}_{ij}\,, \end{aligned} \tag{39}$$

which completes the derivation of the entropic modification to the gravitational equations of motion in the Hamiltonian formulation.

The tensor $\tilde{f}_{ij}$ can be obtained from the phenomenological constraint, which can now be stated rigorously. In the ADM formalism, a well-defined notion of time evolution is given by the flow along the normal vector n^μ. Hence, the time derivative is generalized to the Lie derivative along n^μ. Then the phenomenological constraint is given by

$$\frac{\partial\mathcal{L}}{\partial s}\pounds_n s = \frac{1}{2}N\sqrt{h}\tilde{f}_{ij}\pounds_n h^{ij}\,, \tag{40}$$

where growth in entropy by local processes is related to total entropy density growth by

$$\pounds_n s = \pounds_n s^{tot} - \nabla_i j_s^i\,. \tag{41}$$

Entropy produced locally is expected to grow over time, i.e. with the flow along the hypersurfaces, in compliance with the second law of thermodynamics. This completes the variational formulation of entropic forces in General Relativity.

4.3. *The Raychauduri equation*

Let us explore an immediate dynamical consequence of the inclusion of entropic forces, namely its effect on a congruence of worldlines with tangent vector n^μ. The congruence is then characterized by the tensor

$$\Theta_{\mu\nu} = D_\nu n_\mu = \frac{1}{3}\Theta h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu} - a_\mu n_\nu\,, \tag{42}$$

where θ is the expansion rate of the congruence, $\sigma_{\mu\nu}$ is its shear or symmetric traceless part and $\omega_{\mu\nu}$ is its vorticity or antisymmetric part. If the worldline is not a geodesic, then the congruence suffers an acceleration given by:

$$a_\mu = n^\nu D_\nu n_\mu \,. \tag{43}$$

One can compute the Lie derivative of the expansion of the congruence along its tangent vector and find the Raychauduri equation[5]:

$$\pounds_n \Theta = -\frac{1}{3}\Theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu}n^\mu n^\nu + D_\mu a^\mu \,. \tag{44}$$

Let us perform the standard analysis of the sign of this equation. It is clear that $\sigma_{\mu\nu}\sigma^{\mu\nu} > 0$ and $\Theta^2 > 0$. On the other hand, if the congruence is chosen to be orthogonal to the spatial hypersurfaces, as we have been considering, then the vorticity vanishes $\omega_{\mu\nu} = 0$. Lastly, it is left to consider the term $R_{\mu\nu}n^\nu n^\nu$, which we can rewrite with the help of the field equations:

$$R_{\mu\nu}n^\mu n^\nu = 8\pi G\left(T_{\mu\nu}n^\mu n^\nu + \frac{1}{2}T - f_{\mu\nu}n^\mu n^\nu - \frac{1}{2}f\right) . \tag{45}$$

If the strong energy condition is satisfied, then

$$T_{\mu\nu}n^\mu n^\nu \geq -\frac{1}{2}T \tag{46}$$

and, in the absence of intrinsic acceleration, $a_\mu = 0$, we can establish the bound:

$$\pounds_n \Theta + \frac{1}{3}\Theta^2 \leq 8\pi G\left(f_{\mu\nu}n^\mu n^\nu + \frac{1}{2}f\right) . \tag{47}$$

For a vanishing entropic force $f_{\mu\nu} = 0$, this means that an expanding congruence cannot indefinitely sustain its divergence and will eventually recollapse. On the contrary, a positive and sufficiently large entropic contribution can avoid such recollapse. This may become relevant for an expanding universe, but also to generic gravitational collapse and the singularity theorems.[6–8]

5. Sources of entropy

A main ingredient in the variational formulation of non-equilibrium thermodynamics in General Relativity is the inclusion of entropy at the Lagrangian level and the derivation of a notion of temperature from it. In this section we present two relevant examples: hydrodynamical matter, which is a prototypical case, and horizons.

5.1. *Entropy from hydrodynamical matter*

A classical fluid is the simplest matter content that can be considered in General Relativity and it is of particular relevance in Cosmology. Without paying attention to microphysical details, the Lagrangian of hydrodynamical matter can be written as

$$\mathcal{L}_m = -\sqrt{-g}\rho(g_{\mu\nu}, s) \,, \tag{48}$$

the temperature being then simply given by

$$T = -\frac{1}{\sqrt{-g}}\frac{\partial \mathcal{L}_m}{\partial s} = \frac{\partial \rho}{\partial s}\,. \tag{49}$$

This is analogous to the case of a mechanical system, where the Lagrangian is generically given by a kinetic and a potential energy

$$L = E_K(q, \dot{q}) - U(q, S) \tag{50}$$

and temperature can be defined as

$$T = \frac{\partial L}{\partial S} = -\frac{\partial U}{\partial S}\,. \tag{51}$$

Thus, the energy density of a fluid can be readily interpreted as the thermodynamic internal energy.

5.2. *Entropy from gravity and horizons*

Gravity itself has thermodynamical features. It is known since the discovery of the laws of black hole mechanics and their promotion to laws of black hole thermodynamics, allowed by the introduction of Bekenstein entropy and Hawking temperature. We propose to include the entropy associated with a horizon $\mathcal{H}$ by extending the Einstein-Hilbert actions with surface terms of Gibbons-Hawking-York (GHY) type

$$S_{GHY} = \frac{1}{8\pi G}\int_{\mathcal{H}} d^3y\sqrt{h}K\,, \tag{52}$$

where h is the determinant of the induced 3-metric on the horizon and K is the trace of its extrinsic curvature. Definitions are analogous to the ones used in the ADM formalism, but we stress that here the hypersurface of interest is a horizon and not constant-time hypersurfaces.

From the thermodynamic point of view, the GHY term contributes to the internal energy of the system. Hence, it can be rewritten as a function of the temperature and entropy of the horizon

$$S_{GHY} = -\int dt\, N(t)\, TS\,. \tag{53}$$

We have kept the lapse function $N(t)$, to indicate that the variation of the total action with respect to it will generate a Hamiltonian constraint with an entropy term together with the ordinary matter/energy terms. In order to illustrate this, let us now compute the GHY for the event horizon of a Schwarzschild black hole.

The space-time of a Schwarzschild black hole of mass M is described by the static metric

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega_2^2\,. \tag{54}$$

We foliate it with spherical hypersurfaces, i.e. their intersection with constant time hypersurfaces is a 2-sphere around the origin of coordinates. The corresponding normal vector is

$$n = -\sqrt{1 - \frac{2GM}{r}}\partial_r \,. \tag{55}$$

With this, the trace of the extrinsic curvature for such a sphere scaled by the metric determinant is

$$\sqrt{h}K = (3GM - 2r)\sin\theta \,. \tag{56}$$

Integrating over the angular coordinates and setting the 2-sphere at the event horizon, i.e. $r = 2GM$, and restoring for a moment $\hbar$ and c, the GHY becomes

$$S_{GHY} = -\frac{1}{2}\int dt Mc^2 = -\int dt T_{BH} S_{BH} \,, \tag{57}$$

where T_{BH} is the Hawking temperature and S_{BH} is the Bekenstein entropy of the Schwarzschild black hole:

$$T_{BH} = \frac{\hbar c^3}{8\pi GM} \,, \quad S_{BH} = \frac{Ac^3}{4G\hbar} = \frac{4\pi GM^2}{\hbar c} \,. \tag{58}$$

This favors the interpretation of the GHY term of a horizon as a contribution to the internal energy in the thermodynamic sense.

6. Irreversible cosmology

We derived in section 4 a powerful, generic tool to describe non-equilibrium thermodynamic effects in gravity. In the Hamiltonian formulation of General Relativity it is possible to obtain the modified equations of motion and rigorously impose the time-evolution of the entropy as dictated by the second law of thermodynamics.

In section 2 we motivated the study of these phenomena by our interest in understanding the dynamics of irreversible cosmology and justifying its equations of motion. One can obtain them using the Hamilton equations.[1] Here, however, we present a slightly different approach. Due to the symmetries of the FLRW universe, homogeneity and isotropy, there is a preferred slicing and time evolution is well-defined even at the Lagrangian level. Therefore, we can obtain the equations of non-equilibrium cosmology by imposing these symmetries, i.e. making an ansatz for the metric

$$ds^2 = -N(t)^2 dt^2 + a(t)^2\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2\right) \,, \tag{59}$$

where the lapse function $N(t)$ accounts for the freedom in choosing the time coordinate, i.e. the symmetry under $t \to f(t)$. The Ricci scalar associated to this metric is

$$R = \frac{6}{a^2}\left(\frac{a\ddot{a}}{N^2} + \frac{\dot{a}^2}{N^2} + k\right) \,. \tag{60}$$

Let us stress that this result is imposed by symmetry, not dynamics. Without loss of generality, we can restrict the action to a region of comoving volume 1 and write it as

$$S = \int dtL = \frac{3}{8\pi G}\int dtNa\left(\frac{a\ddot{a}}{N^2} + \frac{\dot{a}^2}{N^2} + k\right) + \int dtNa^3\mathcal{L}_m(N,a,S)\,. \tag{61}$$

Effectively, this action describes a mechanical system, for the scale factor $a(t)$ has no spatial dependency and we got rid of the integral over spatial coordinates. The first term can be rewritten using integration by parts in order to get only terms with at most the first derivative of a

$$S = \frac{3}{8\pi G}\int dtNa\left(-\frac{\dot{a}^2}{N^2} + k\right) + \int dtNa^3\mathcal{L}_m(N,a,S)\,. \tag{62}$$

The variational constraint is here given by the usual expression for a mechanical system

$$\frac{\partial L}{\partial S}\delta S = f\delta a\,. \tag{63}$$

The Hamiltonian of the system is

$$H = \dot{a}\frac{\partial L}{\partial \dot{a}} - L = \frac{3}{8\pi G}\left(-\frac{\dot{a}^2 a}{N} - kaN\right) - Na^3\mathcal{L}_m\,. \tag{64}$$

For an arbitrary lapse function $N(t)$ this can be rewritten as

$$H = N^2a^3\frac{\partial \mathcal{L}_m}{\partial N}\,, \tag{65}$$

which gives the Hamiltonian constraint of the system. On the other hand, the dynamics is obtained from the equation of motion for a

$$\frac{\delta L}{\delta a} = -f\,. \tag{66}$$

Let us now consider the matter Lagrangian to be that of a perfect fluid, i.e.

$$\mathcal{L}_m = -\rho(a,S)\,. \tag{67}$$

Its stress-energy tensor is given in terms of the density ρ and pressure p by

$$T^{\mu\nu} = (\rho + p)u^{\mu\nu} + pg^{\mu\nu} \tag{68}$$

and $u^m u = (N,0,0,0)$ is the unit vector tangent to a comoving observer. Pressure is then obtained as

$$p = \frac{a^2}{3}T^{ij}\delta_{ij} = -\frac{1}{2a^4}\frac{\partial a^3\rho}{\partial a}\,. \tag{69}$$

Using the expressions for ρ and p and rearranging the terms in the Hamiltonian constraint and the equation of motion for $a(t)$ we arrive at the modified Friedmann

equations

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho$$
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p + \frac{f}{a^2}\right). \tag{70}$$

The expression for the entropic force F is obtained from the phenomenological constraint

$$\left(\frac{\partial L}{\partial S}\right)\dot{S} = -T\dot{S} = f\dot{a} < 0\,, \tag{71}$$

which determines the sign $f < 0$ whenever dealing with an expanding universe $\dot{a} > 0$. We express finally the second Friedmann equation as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p - \frac{T\dot{S}}{a^2\dot{a}}\right). \tag{72}$$

From this equation we can conclude that entropic forces generally drive an acceleration of the expansion of the universe. Whether this can dominate the dynamics of the scale factor will depend on the particular thermodynamic process. Most of the expansion history of the universe takes place in equilibrium. Out of equilibrium processes, such as (p)reheating, phase transitions or gravitational collapse are short-lived. Should their associated entropic force dominate, we still only expect a short period of accelerated expansion.

Symmetry under time inversion is broken by the same arguments presented in section 3. Hence, the Friedmann equations together with the phenomenological constraint, i.e. the second law of thermodynamics, describe cosmic irreversible dynamics.

We currently live in a universe that is undergoing an accelerated expansion. The possibility of explaining this by means of an entropic force is fascinating. In the next section we review our proposal to achieve this by means of the sustained growth of the entropy associated to a causal horizon.

7. Cosmic acceleration as an entropic force

The growth of entropy associated to the cosmic horizon may be responsible for the current observed accelerated expansion of the universe. The choice of horizon is in principle not unique. The only available one which can be defined locally in time is the cosmic apparent horizon, but it fails to significantly affect the expansion.[2]

There is another option in the framework of eternal inflation, according to which we live in an open universe nucleated by quantum tunneling from a false to a true vacuum. After nucleation the bubble universe undergoes its own inflationary era, which renders the local metric almost flat. However, due to the presence of the

bubble wall, the true causal horizon is located at a finite distance. It induces an entropic fluid via GHY term with energy density

$$\rho_H\, a^2 = \frac{T_H S_H}{a} = \frac{x_0}{2G}\,\sinh(2a_0 H_0 \eta)\,, \quad x_0 \equiv \frac{1-\Omega_0}{\Omega_0} = e^{-2N}\left(\frac{T_{\rm rh}}{T_{\rm eq}}\right)^2 (1+z_{\rm eq})\,, \tag{73}$$

where η is the conformal time, Ω_0 is the density parameter, $T_{\rm rh}$ is the reheating temperature, $T_{\rm eq}$ and $z_{\rm eq}$ are, respectively, the temperature and redshift at matter-radiation equality. Introducing $\tau = a_0 H_0 \eta$ one can write the second Friedmann equation in conformal time as

$$\left(\frac{a'}{a_0}\right)^2 = \Omega_M \left(\frac{a}{a_0}\right) + \Omega_K \left(\frac{a}{a_0}\right)^2 + \frac{4\pi}{3}\Omega_K \left(\frac{a}{a_0}\right)^2 \sinh(2\tau)\,, \tag{74}$$

where Ω_M is the matter density paremeter and Ω_K is the curvature parameter. We call this the general relatvistic entropic acceleration (GREA) theory.

By solving this equation with cosmological parameters consistent with the CMB values (Planck 2018: $\Omega_M \simeq 0.31$, $\Omega_K \simeq 0.0006$, $h_0 \simeq 0.68$) and initial conditions deep in the matter era, $a_i(\tau) = a_0\,\Omega_M \tau^2/4$, we find generic accelerating behaviour beyond the scale factor $a \sim 1/2$ (*i.e.* $z \sim 1$), see Fig. 1. This is consistent with the current observed acceleration of the universe and may even resolve the Hubble tension,[9] providing a way to obtain from the CMB a present value of H_0 that is consistent with late-universe observations, see Fig. 2.

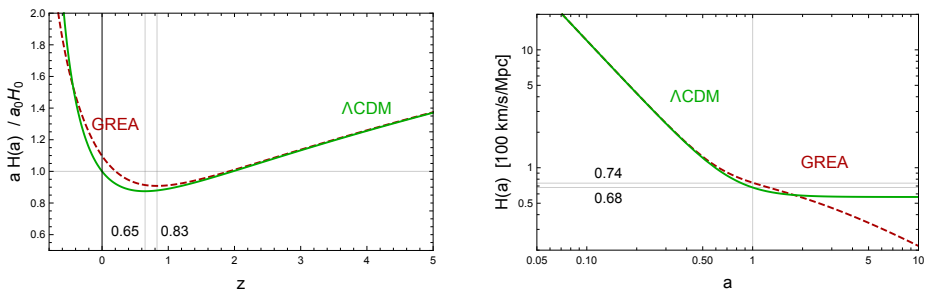


Fig. 1. The left plot shows the evolution of the inverse comoving horizon with the coasting point for each model, at $z \simeq 0.65$ for ΛCDM (in green) and $z \simeq 0.83$ for GREA (in red). The right plot shows the evolution of the rate of expansion. For GREA the present rate of expansion is approximately 74 km/s/Mpc, compared with the value of 68 km/s/Mpc predicted by ΛCDM, in agreement with the asymptotic value at the CMB.

8. Conclusions

The consistent inclusion of non-equilibrium phenomena in General Relativity leads to the modification of the Einstein field equations, as can be checked both in the Lagrangian and Hamiltonian formulations of the theory. This breaks symmetry under time inversion and allows for the introduction of an arrow of time.

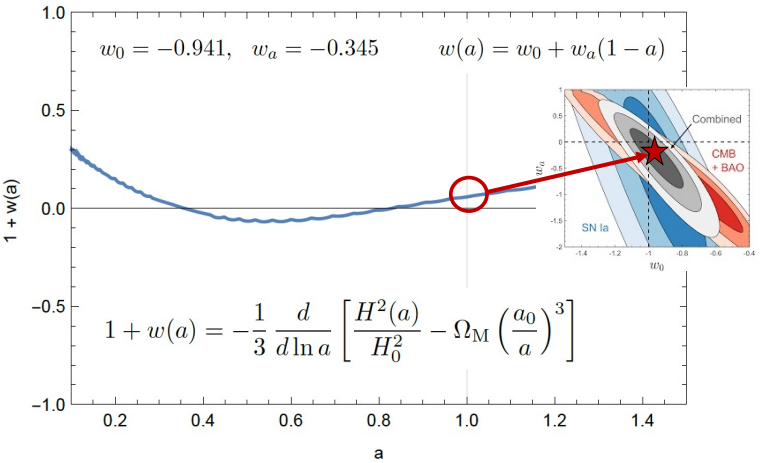


Fig. 2. The effective equation of state of the non-matter component of the GREA theory, as a function of thee sale factor. Note that the predicted effective PCL parameters (w_o, w_a) agree remarkably well with present observations.

In cosmology this implies the appearance of a term of entropic origin in the second Friedmann equation, which tends to accelerate the expansion of the universe as a result of the increase in entropy. Some physical processes such as (p)reheating, phase transitions or gravitational collapse may lead to phenomenologically relevant applications of this formalism. We look forward to further developments.

The sustained entropy growth associated to a causal horizon in the open universe scenario leads to an acceleration consistent with current observations and it may even solve the H_0 tension. Further research will be required to establish the full viability of the GREA theory.

Acknowledgments

The authors acknowledge support from the Spanish Research Project PGC2018-094773-B-C32 (MINECO-FEDER) and the Centro de Excelencia Severo Ochoa Program SEV-2016-0597. The work of LEP is funded by a fellowship from "La Caixa" Foundation (ID 100010434) with fellowship code LCF/BQ/IN18/11660041 and the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 713673.

References

1. L. Espinosa-Portales and J. Garcia-Bellido, Covariant formulation of non-equilibrium thermodynamics in General Relativity, *Phys. Dark Univ.* **34**, p. 100893 (2021).
2. J. García-Bellido and L. Espinosa-Portales, Cosmic acceleration from first principles, *Phys. Dark Univ.* **34**, p. 100892 (2021).

3. F. Gay-Balmaz and H. Yoshimura, A lagrangian variational formulation for nonequilibrium thermodynamics. part i: Discrete systems, *Journal of Geometry and Physics* **111**, 169 (2017).
4. F. Gay-Balmaz and H. Yoshimura, A lagrangian variational formulation for nonequilibrium thermodynamics. part ii: Continuum systems, *Journal of Geometry and Physics* **111**, 194 (2017).
5. R. M. Wald, *General Relativity* (Chicago Univ. Pr., Chicago, USA, 1984).
6. R. Penrose, Gravitational collapse and space-time singularities, *Phys. Rev. Lett.* **14**, 57 (1965).
7. R. Penrose, Gravitational collapse: The role of general relativity, *Riv. Nuovo Cim.* **1**, 252 (1969).
8. S. W. Hawking and R. Penrose, The Singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. Lond. A* **314**, 529 (1970).
9. A. G. Riess, The Expansion of the Universe is Faster than Expected, *Nature Rev. Phys.* **2**, 10 (2019).

Experimental detection of the CNO cycle

B. Caccianiga[1], N. Rossi[2], G. Testera[3]
M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, R. Biondi, D. Bravo, B. Caccianiga, F. Calaprice, P. Cavalcante, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X.F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, A. Formozov, D. Franco, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, A.S. Göttel, M. Gromov, D. Guffanti, Aldo Ianni, Andrea Ianni, A. Jany, D. Jeschke, V. Kobychev, G. Korga, S. Kumaran, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, M. Nieslony, R. Nugmanov, L. Oberauer, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, L. Pelicci, Ö. Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M.T. Ranalli, G. Ranucci, A. Razeto, A. Re, M. Redchuk, A. Romani, N. Rossi, S. Schönert, D. Semenov, G. Settanta, M. Skorokhvatov, A. Singhal, O. Smirnov, A. Sotnikov, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, F. Villante, A. Vishneva, R.B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, S. Zavatarelli, K. Zuber, and G. Zuzel

[1] *Dipartimento di Fisica, Università degli Studi e INFN, 20133 Milano, Italy*

[2] *INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy*

[3] *Dipartimento di Fisica, Università degli Studi e INFN, 16146 Genova, Italy*

The Borexino has recently reported the first experimental evidence of neutrinos from the CNO cycle. Since this process accounts only for about 1% of the total energy production in the Sun, the associated neutrino flux is extremely low as compared with the one from the pp-chain, the dominant process of hydrogen burning. This experimental evidence of the CNO neutrinos was obtained using the highly radio-pure liquid scintillator of Borexino. Improvements in the thermal stabilization of the detector over the last five years enabled us to exploit a method to constrain the rate of ^{210}Bi background. Since the CNO cycle is dominant in massive stars, this result gives the first experimental proof of the primary mechanism for stellar conversion of hydrogen into helium in the Universe.

Keywords: Solar neutrinos, CNO cycle, Liquid organic scintillators, solar models

Introduction

Stars are fuelled by nuclear fusion of light elements with a release of an enormous amount of energy. In particular the *pp* chain and the CNO cycle, active also in the Sun, produces a rich spectrum of electron-flavour neutrinos detectable on the Earth.[1,2] Those two mechanisms basically converts four hydrogen nuclei into a helium nucleus: the first using direct process throughout a big variety of branches, while the second using a closed loop starting and ending with ^{12}C acting as catalyst.

The contribution of these two mechanisms in the energy production is related to mass, and then to the core temperature, of the stars and to abundance of elements heavier than helium in the core (in jargon, "metallicity").

For stars heavier than 1.3 $M_{\odot}$ the CNO cycle is the dominant process of hydrogen fusion,[3] while in the Sun, with a core temperature of about 15×10^6K, the CNO accounts only for 1% of the total luminosity, with a large uncertainty related to poorly known metallicity.[4] The Borexino experiment, starting from mid-2007, has published a comprehensive study of the neutrino from the *pp* chain [5] and recently has reported the first direct detection of the CNO.[26] The CNO quest, started by Borexino, could in principle create a situation that makes it possible the solution of the long standing "solar metallicity problem" ,[2] i.e. the disagreement between the metallicity predicted by solar models using updated (low) metal abundances from spectroscopy (SSM-LZ) [6] and the one obtained from the helioseismology, which foresees a higher metal content (SSM-HZ).

In Sec. 1 a description of the detector and the main physics results of Borexino are reported. At the end of the section the hardware improvement for the CNO detection in Phase-III is also reported. In Sec. 2 the main strategy for the CNO detection is described and in Sec. 3 the independent constraint of the ^{210}Bi (the main background for the CNO neutrino interaction rate) is largely detailed. Finally, in Sec. 4 the spectral analysis for the extraction of the CNO rate in Borexino is presented.

1. The Borexino detector

Borexino experiment is the only experiment capable to detect solar neutrinos with a threshold as low as about 150 keV and to reconstruct the position and the energy of each event in real-time.

Borexino is located in the Hall C of Laboratori Nazionali Gran Sasso (INFN).[12] The detector is made of concentric shells with increasing radiopurity (see e.g. Ref.[13]): the innermost core, enclosed in a 125 μm thick ultra-pure nylon vessel of radius 4.25 m, is made of 300 tons of liquid scintillator (1,2,4-Trimethylbenzene with 1.5 g/l of PPO wavelength shifter). The active core is contained by a stainless steel sphere filled up with $\sim$ 1000 of buffer liquid (1,2,4-Trimethylbenzene with DMP quencher), whose internal surface is instrumented with more than 2000 PMTs for detecting the scintillation light. Finally, the SSS is located inside a 2000 ton water Cherenkov detector, equipped with about 200 PMTs. In Fig. 1 all the parts described above are represented in a graphical scheme. Thanks to an intense calibration campaign carried out in 2010, the Borexino detector is able to reconstruct the event position with an accuracy of $\sim$ 10 (at 1 MeV) and with energy resolution of about $\sigma(E)/E = 5\%/\sqrt{(E/[MeV])}$.

The Borexino data-set is traditionally divided in three Phases, determined by the hardware milestones: Phase-I, from mid-2007 to beginning of 2010, ended with the calibration campaign, in which the first measurement of the ^{7}Be solar neutrino interaction rate[8–10] was performed; Phase-II, from the-beginning of 2012 to mid-2016, started after an intense purification campaign, based on water extraction, with unprecedented suppression of the radioactive contaminants, in which the first

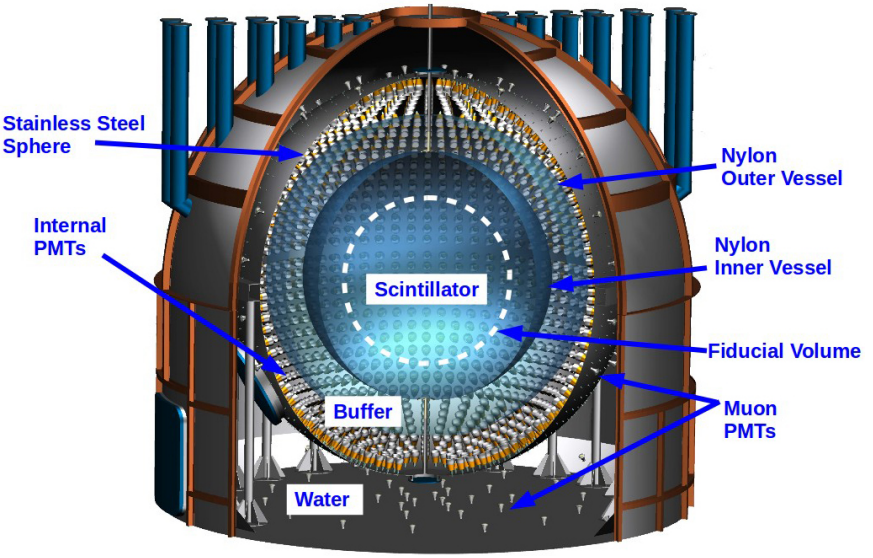


Fig. 1. Scheme of the Borexino detector.

evidence of the *pep* neutrinos[15] and a 10% measurement of the *pp* neutrinos[16] were published, later updated in the solar neutrino comprehensive analysis;[17–19] Phase-III, from mid-2016 to 2021 (still in progress), after the thermal insulation program, in which the first detection of the CNO neutrinos[11] discussed here was performed. Table 1 summarizes the most important results concerning solar neutrinos interaction rates measured by Borexino. Thanks to its unprecedented radio-purity, Borexino has also set a lot of limits on rare processes (see e.g.,[20–24] and performed other neutrino physics studies, as e.g. geo-neutrino detection (for review, see e.g.[25])

The CNO detection has required, as it is explained in Sec. 2, an independent constraint of the ^{210}Bi contaminant from the ^{210}Po through the secular equilibrium starting from the out-of-equilibrium parent ^{210}Pb, present in the liquid scintillator

Table 1. Solar neutrino interaction rates and fluxes measured by Borexino. Rates are reported in cpd/100t, while fluxes are reported in $\mathrm{cm^{-2}s^{-1}}$.

Species	Rate [cpd/100t]	Flux [$\mathrm{cm^{-2}\ s^{-1}}$]
pp	$(134 \pm 10)^{+6}_{-10}$	$(6.1 \pm 0.5)^{+0.3}_{-0.5} \times 10^{10}$
^{7}Be	$(48.3 \pm 1.1)^{+0.4}_{-0.7}$	$(4.99 \pm 0.11)^{+0.06}_{-0.08} \times 10^{9}$
pep (HZ)	$(2.7 \pm 0.4)^{+0.1}_{-0.2}$	$(1.3 \pm 0.3)^{+0.1}_{0.1} \times 10^{8}$
^{8}B($>$ 3 MeV)	$0.223^{+0.021}_{-0.022}$	$5.68^{+0.42}_{-0.44} \times 10^{6}$
CNO	$7.2^{+3.0}_{-1.7}$	$7.0^{+3.0}_{-2.0} \times 10^{8}$
hep	< 0.002 (90% CL)	$< 1.8 \times 10^{5}$ (90% CL)

after the purification carried out at the end of Phase-I. That constraint required the contaminants, and then the whole scintillator, to be stable against the the convective motions activated by the seasonal variation of the temperature in the experimental hall hosting the detector. In mid-2014 Borexino was equipped with a detailed temperature mapping. In mid-2015 the detector was covered with a thermal insulation and, from the beginning of 2016, a series of active temperature control system were installed around the water tank and at the end on the experimental hall. See for details Ref.[26].

The thermal insulation program was aimed at the realization of a vertical temperature gradient from the cold floor, made of constant temperature rock with a huge thermal capacitance, and the detector top, warmed up by the air flushed into the underground laboratory coming from the external environment out of the Gran Sasso mountain. This challenging enterprise was successful and the strongest convective motions, mixing the liquid scintillator and then the contaminants from the outermost regions into the center, were actually stopped and the conditions for the ^{210}Bi-^{210}Po link, described in Sec. 3 were finally met. Figure 2 shows the ^{210}Po contaminant evolution during the thermal insulation program in 59 small cubes enumerated from the bottom to the top inside a radius of 3 m from the center of the Inner Vessel. The red superimposed curve represents the average temperature in the same region. The dashed vertical lines show the most important milestones of the temperature stabilisation program: 1. Beginning of insulation; 2. Turning off of the water recirculation system in the Water Tank; 3. First operation of the active temperature control system; 4. Change of the active control set point; 5. Installation and commissioning of the Hall C temperature control system. The white vertical bands represent different technical DAQ breaks. The ^{210}Po shown in the figure is a

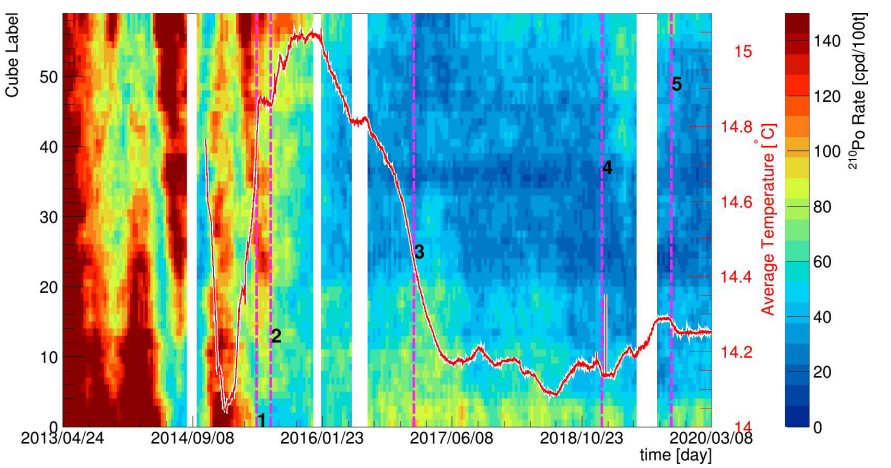


Fig. 2. Evolution of the ^{210}Po contamination during the thermal insulation program.

tracker of the convective motions, that are drastically damped as a consequence of the insulation program.

2. CNO detection strategy

Neutrinos emitted in the CNO cycle, see Fig. 3, have a continuous spectrum with a maximum energy of 1.74 MeV (see Fig. 4, left). In the Sun, the flux of CNO neutrinos is two orders of magnitude smaller than the one due to proton-proton neutrinos: this makes the expected rate in Borexino very low, of the order of few counts per day. Besides the low rate, the main difficulty in detecting CNO neutrinos with Borexino is the presence of two radioactive backgrounds which produce signal in approximately the same energy region: ^{11}C and ^{210}Bi. Furthermore, also solar neutrinos from the *pep* reaction are a significant background for the CNO analysis.

Fig. 4 (right) shows a MonteCarlo simulation of the energy spectrum of electrons scattered by neutrinos from the CNO cycle (red line) in comparison with the signal induced by the radioactive decay of ^{11}C (purple line) and ^{210}Bi (green line). The blue line is the *pep* signal. The relative proportion of the four contributions is approximately the one expected in reality. From this plot the difficulty of extracting the CNO signal from the background becomes evident: it is like looking for a needle in a haystack!

The cosmogenic ^{11}C is an isotope produced by residual muons crossing the scintillator which decays emitting a positron with a lifetime of approximately 30 minutes. It can be very effectively tagged by the Three Fold Coincidence (TFC) method described in,[27] which looks for the triple coincidence of the parent muon, the cosmogenic neutron produced together with ^{11}C and the positron emitted in the ^{11}C decay. This tagging method allows us to divide the data-set in two complementary parts, one rich and the other depleted in ^{11}C which are then simultaneously fitted in the multivariate analysis described in Section 4. This is a powerful tool to help the fit isolate the ^{11}C signal contribution, breaking the correlation with the CNO neutrino signal.

Separating the CNO signal from the ^{210}Bi and *pep* one is more tricky since their spectral shapes are quite similar and the multivariate fit has trouble in distinguishing between them, if no additional external contraints are imposed.

For what concerns *pep* neutrinos, we exploited the external and independent information coming from the available solar neutrino data, solar luminosity and the most recent oscillation parameters to put a constraint to their rate in the fit with a 1.4% uncertainty.

The remaining obstacle to the CNO neutrino detection is the contamination from ^{210}Bi. This isotope belongs to the ^{210}Pb chain, which in turn is the last segment of the ^{238}U chain. In Borexino, we found the ^{210}Pb chain to be out of equilibrium with respect to the ^{238}U chain and with a significantly higher activity, of the order of tens of counts per day per 100 tons of scintillator.

The ^{210}Pb chain is composed of three isotopes: the long lived ^{210}Pb father of the chain ($\tau \sim 32\,\mathrm{y}$), ^{210}Bi which has a relatively short lifetime ($\tau \sim 7.23\,\mathrm{d}$) and ^{210}Po which decays into ^{206}Pb (stable) emitting an alpha particle ($\tau \sim 199.1\,\mathrm{d}$). If secular equilibrium holds, the activities of ^{210}Bi and ^{210}Po are the same (apart for negligible corrections due to the finite ^{210}Pb lifetime). Measuring the ^{210}Po rate is in principle straightforward since an alpha peak is clearly separable from the continuous beta-like background underneath. Furthermore, alpha/beta discrimination techniques provide a powerful additional tool to isolate the ^{210}Po events.

Based on these premises, our strategy for the CNO detection relies on the ^{210}Bi-^{210}Po link to determine independently the ^{210}Bi content from ^{210}Po and constrain it in the multivariate fit. The next Section will describe in details the difficulties we encountered to apply this technique and how we were able of overcoming them.

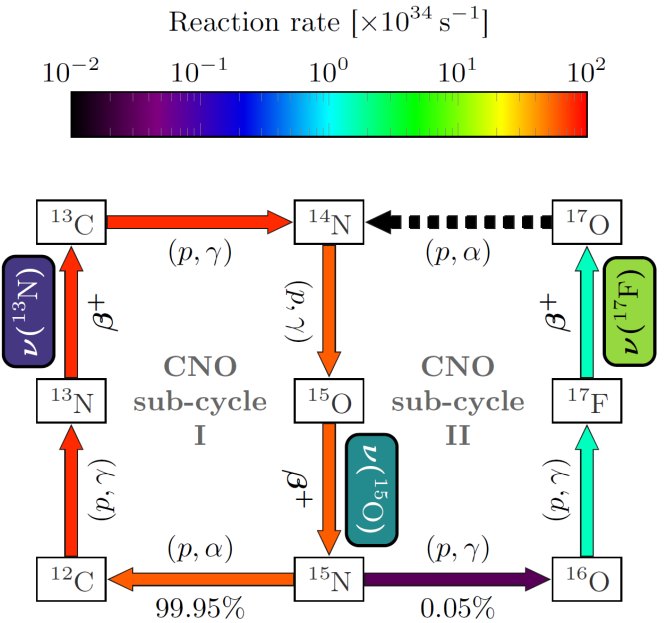


Fig. 3. Neutrinos in CNO cycle.

3. The ^{210}Bi–^{210}Po link

In order to exploit the ^{210}Bi–^{210}Po link to obtain an external constraint on the ^{210}Bi rate, we carefully studied the evolution in time and the spatial distribution

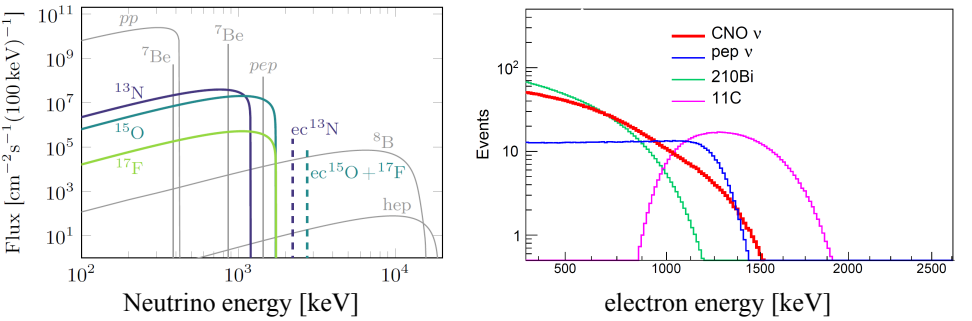


Fig. 4. (Left): Energy spectrum of neutrinos emitted in the CNO cycle (colored lines); (right): energy spectrum in Borexino for events of CNO (red), *pep* (blue), $^{@10}$Bi (geen) and ^{11}C (purple) obtained from MonteCarlo simulations.

of the polonium contamination in the detector during the Borexino data-taking. We soon realized that the ^{210}Po present in the fiducial region used in the analysis (an innermost subset of the scintillator volume) is actually the sum of two contributions: ^{210}Po from the intrinsic ^{210}Pb scintillator contamination (in equilibrium with ^{210}Bi); ^{210}Po from the nylon vessel containing the scintillator. This last contribution was found to be changing in time and highly correlated with variations of the temperature gradient in the detector, which induce convective motions strong enough to bring ^{210}Po from the vessel surface into the fiducial region. In order to reduce these convective motions, we thermally insulated the detector, as discussed in Section 1.

After the completion of the thermal insulation (2016), the temperature of the detector has become very stable and the convective currents inside the scintillator have significantly reduced. This has favoured the formation of an innermost region of the detector, called Low Polonium Field (LPoF), where the contribution to the ^{210}Po rate coming from the vessel is very small. A pictorial view of this region is shown in Fig. 5. This region exhibits an effective migration profile with a minimum in the ^{210}Po rate above the detector equator (dark blue region). The qualitative shape and approximate position of the LPoF is reproduced by fluid dynamical numerical simulations reported in .[28]

Assuming azimuthal symmetry around the detector z-axis (which was confirmed by a 3D analysis), the ^{210}Po minimum activity is determined by fitting LPoF with a 2D paraboloidal function:

$$\frac{d^2R(^{210}\mathrm{Po})}{d(\rho^2)dz} = \left[R(^{210}\mathrm{Po}_{\mathrm{min}})\epsilon_{\mathrm{E}}\epsilon_{\mathrm{MLP}} + R_\beta\right] \times \\ \times \left(1 + \frac{\rho^2}{a^2} + \frac{(z-z_0)^2}{b^2}\right), \tag{1}$$

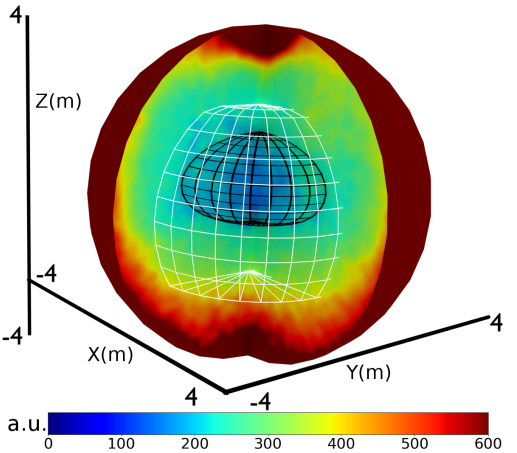


Fig. 5. Pictorial view of the ^{210}Po activity in the Borexino scintillator. The color scale goes from red (highest activity) to blue (lowest activity). The LPoF is the innermost blue region where the contribution of ^{210}Po from the vessel is the smallest (black grid). The white grid indicates the software defined fiducial volume used for the analysis to extract the CNO neutrinos.

Here, $\rho^2 = x^2 + y^2$, a and b are the paraboloid axes, z_0 is the position of the minimum along the z axis, ϵ_{E} and ϵ_{MLP} are the efficiencies of the cuts used to select ^{210}Po events, and R_β is the residual rate of β events after the selection. First, we performed the fit integrating data from 2 months (compatible results are obtained also changing the time bins) and we found that the minimum position slowly moves along the z direction by less than 20 cm per month. In order to perform a better estimation of the ^{210}Po minimum, we finally sum up all the time bins after aligning the 3D distributions with respect to z_0.

From this fit we extract the ^{210}Po minimum. This value might still have a small contribution from the vessel component, therefore this method provides only an upper limit for the ^{210}Bi rate. In order to cross-check the results, we performed the fit in different ways: adopting a standard likelihood fit with ROOT or a Bayesian tool called MultiNest. In addition to the 2D paraboloidal fit, we also performed the fit with a 3D ellipsoidal function or alternatively with a cubic spline. The differences in the results were included in the systematic error associated to the upper limit on ^{210}Bi. In the analysis to extract the CNO neutrino rate which will be described in Section 4, we would like to costraint the value of the ^{210}Bi to the upper limit found in the low polonium field region, extrapolating the results to the whole fiducial volume. In order to do so, we need to verify that ^{210}Bi is uniform in this volume during the time period over which the estimation is performed. For this reason, we studied the spatial distribution of β-like events in an energy window where the bismuth contribution is maximized. This distribution was found to be uniform within errors: we include in the systematic error budget a contribution of 0.78 cpd/100t accounting for residual non uniformity.

The final upper limit on ^{210}Bi including statistical and systematical errors added in quadrature is:

$$\text{Rate}(^{210}\text{Bi}) \leq (11.5 \pm 1.3)\,\text{cpd/100t}$$

4. Spectral analysis

We performed a multivariate analysis, simultaneously fitting the energy spectra in the window between 320 keV and 2640 keV and the radial distribution of the selected events, using the data collected from June 2016 to February 2020 (Borexino Phase-III). The fiducial volume (FV) is defined as r<2.8 m and z in the interval (-1.8 m, 2.2 m) (r and z being the reconstructed radial and vertical position, respectively). The total exposure corresponds to 1072 days · 71.3 tonnes. Free parameters in the fit procedure are: the interaction rate of CNO neutrinos, ^{85}Kr, ^{11}C, internal and external ^{40}K, external ^{208}Tl and ^{214}Bi, and ^{7}Be neutrinos. The pep neutrino rate is constrained to to 2.74 ± 0.04 cpd per 100 t by multiplying the standard likelihood with a symmetric Gaussian term. The upper limit to the ^{210}Bi rate is the one reported in previous section. It is enforced asymmetrically by multiplying the likelihood with a half-Gaussian term, i.e., leaving the ^{214}Bi rate unconstrained between 0 and 11.5 cpd per 100 t .

The reference spectral and radial distributions of each signal and background species are obtained with a complete GEANT4-based Monte Carlo simulation which models all physics processes occurring in the scintillator, including energy deposition, photon emission, propagation, and detection, generation and processing of the electronic signal. The simulation takes into account the evolution in time of the detector response and produces data that are reconstructed and selected following the same pipeline of real data. The results of the simultaneous multivariate fit of the energy spectra are given in Fig. 6, showing the TFC subtracted and TFC-tagged data. The fit is performed in the energy estimator N_{hits} (defined as the sum of all photons triggering a PMT, normalized to 2000 active PMTs) and the results are reported also in keV. The p-value of the fit is 0.3, demonstrating fair agreement between data and the underlying fit model. The fit clearly prefers a non-zero CNO neutrino rate as also shown in a detailed study of log-likelihood profile.

The best fit value of the CNO neutrinos interaction rate is 7.2 cpd per 100 t with an asymmetric confidence interval of -1.7 cpd per 100 t and +2.9 cpd per 100 t (68% C.L.), considering the statistical error only. The many details of the analysis, including the list and evaluation of the systematic effects are reported in.[26] Folding the systematic uncertainty over the log-likelihood profile we determine the final CNO interaction rate to be $7.2^{+3}_{-1.7}$ cpd per 100 t. The rate can be converted to of CNO neutrinos on Earth of $7.2^{+3.0}_{-2.0} \cdot 10^8$ cm^{-2} s^{-1} assuming MSW conversion in matter and a density of electrons in the scintillator of $(3.307 \pm 0.015) \cdot 10^{31}$ e$^-$per 100 t.

In order to evaluate the significance of our result in rejecting the no-CNO hypothesis, we performed a frequentist hypothesis test using the profile likelihood.

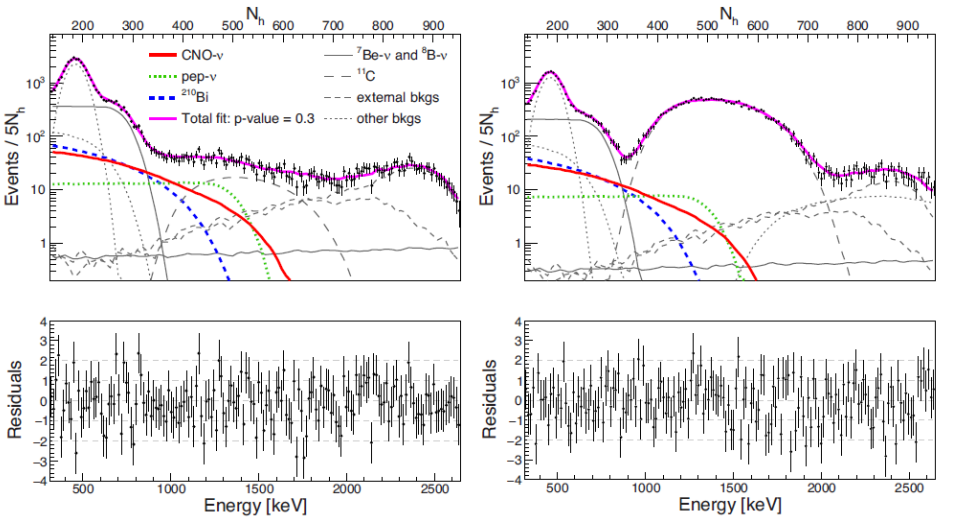


Fig. 6. Energy distributions from a multivariate fit of the Borexino data. Full multivariate fit results for the TFC-subtracted (left) and the TFC-tagged (right) energy spectra with corresponding residuals. In both graphs the magenta lines represent the resulting fit function, the red line is the CNO neutrino electron recoil spectrum, the green dotted line is the pep neutrino electron recoil spectrum, the dashed blue line is the ^{210} β spectrum, and in grey we report the remaining background (bkgs) contributions.

We defined a test statistics q, as the ratio between the maximum likelihood obtained by keeping the CNO rate fixed to zero or free. Figure 7 shows the q distribution obtained from 13.8 millions of pseudo-data sets simulated with deformed PDFs to include systematics effects, and no-CNO injected (q_0, grey curve). In the same plot, the theoretical q_0 distribution in case of no PDF deformation is shown (blue curve). The result on data obtained from the fit is the black line ($q_{data} = 30.05$). The results reported in Figure 7 allow us to reject the CNO = 0 hypothesis with a significance better than 5.0σ at 99.0% C.L. In the figure we also provide as reference the q distribution (red) obtained with one million pseudo-data sets including systematic deformations and injected CNO rate equal to 7.2 cpd per 100 t, i.e., our best fit value.

The results of the analysis based on the fit are also confirmed by that obtained with a nearly independent method (counting analysis), in which we simply count events in an optimized energy window (region of interest, ROI) and subtract the contributions due to known backgrounds. This method reveals a non zero CNO signal and its consistency with the multivariate analysis is remarkable. This method is simpler, albeit less powerful, with respect to the multivariate fit and is less prone to possible correlations between different species. However, while the multivariate analysis implicitly checks the validity of the background model by the goodness of the fit, the counting analysis relies completely on the assumption that there are no unknown backgrounds contributing to the ROI.

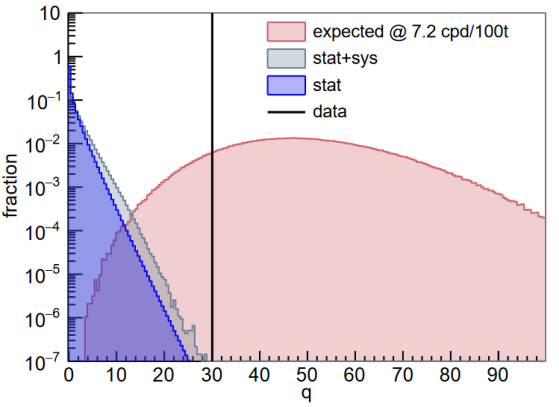


Fig. 7. Distribution of the test statistic q, as discussed in the text.

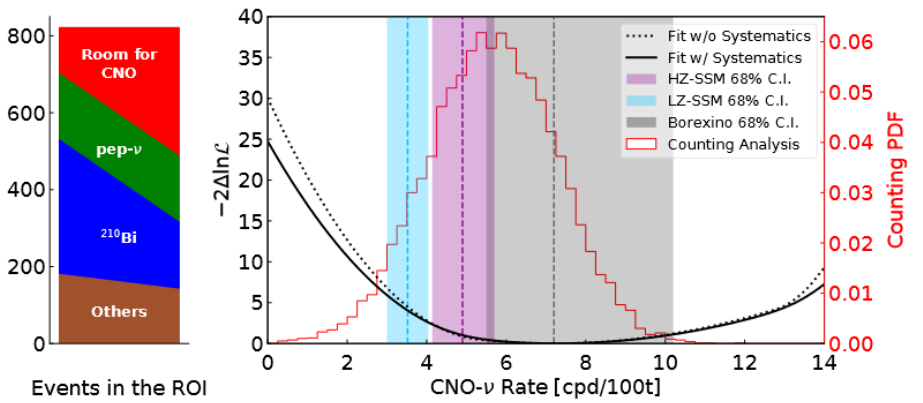


Fig. 8. Results of the CNO counting and spectral analyses. Left. Counting analysis bar chart. The height represents the number of events allowed by the data for CNO neutrinos and backgrounds in the ROI; on the left, the CNO signal is minimum and backgrounds are maximum, while on the right, CNO is maximum and backgrounds are minimum. It is clear from this figure that CNO cannot be zero. Right. CNO neutrinos rate negative log-likelihood profile directly from the multivariate fit (dashed black line) and after folding in the systematic uncertainties (black solid line). Histogram in red: CNO neutrino rate obtained from the counting analysis. Finally, the blue, violet, and grey vertical bands show 68% confidence intervals (C.I.) for the SSM-LZ (3.52±0.52 cpd per 100 t) and SSM-HZ (4.92±0.78 cpd per 100 t) predictions and the Borexino result (corresponding to black solid-line log-likelihood pro- file), respectively.

Conclusions

We have reported the results about the first direct detection of neutrinos from the CNO cycle in the Sun. This result proves the occurrence of the primary mechanism

for the stellar conversion of hydrogen into helium in the Universe. The observation of CNO neutrinos experimentally confirms the overall solar picture and shows that a direct measurement of the metallicity of the Sun's core is within reach of an improved, future measurement.

References

1. J. N. Bahcall, Neutrino Astrophysics, *Cambridge University Press*, 1989.
2. N. Vinyoles, A.M. Serenelli, F.L. Villante, S. Basu, J. Bergström, M.C. Gonzalez-Garcia, M. Maltoni, C. Peña-Garay, and N. Song, *Astrophys. J.*, **835**(2):202, 2017.
3. M. Salaris and S. Cassisi, Evolution of Stars and Stellar Populations, *John Wiley & Sons Ltd.*, 2005.
4. C. Angulo et al., *Nuclear Physics A*, **656**(1):3–183, 1999.
5. M. Agostini et al. (Borexino Collaboration), *Nature*, **562**(7728):505–510, 2018.
6. A. M. Serenelli, W. C. Haxton, and C. Peña-Garay, *The Astrophysical Journal*, **743**(1):24, 2011.
7. G. Bellini et al. [Borexino], Phys. Rev. D **89** (2014).
8. C. Arpesella et al. [Borexino], Phys. Lett. B **658** (2008), 101-108.
9. C. Arpesella et al. [Borexino], Phys. Rev. Lett. **101**(2008), 091302.
10. G. Bellini, et al. [Bore ino], Phys. Rev. Lett. **107** (2011), 141302.
11. M. Agostini et al. [Borexino], Nature **587** (2020), 577-582.
12. Website: `https://www.lngs.infn.it`
13. G. Alimonti et al. [Borexino], Astropart. Phys. **16** (2002), 205-234.
14. H. Back et al. [Borexino], JINST **7** (2012), P10018.
15. G. Bellini et al. [Borexino], Phys. Rev. Lett. **108** (2012), 051302.
16. G. Bellini et al. [Borexino], Nature **512** (2014) no.7515, 383-386.
17. M. Agostini et al. [Borexino], Nature **562** (2018) no.7728, 505-510.
18. M. Agostini et al. [Borexino], Phys. Rev. D **100** (2019) no.8, 082004.
19. M. Agostini et al. [Borexino], Phys. Rev. D **101** (2020) no.6, 062001.
20. A. Vishneva et al. [Borexino], J. Phys. Conf. Ser. **888** (2017), 012193.
21. S. K. Agarwalla et al. [Borexino], JHEP **02** (2020), 038.
22. M. Agostini et al. [Borexino], Astropart. Phys. **125** (2021), 102509.
23. M. Agostini et al. [Borexino], Phys. Rev. D **96** (2017), 091103.
24. G. Bellini et al. [Borexino], Phys. Rev. D **88** (2013), 072010.
25. M. Agostini et al. [Borexino], Phys. Rev. D **101** (2020), 012009.
26. The Borexino Collaboration, *Nature* **587**, 577 (2020).
27. The Borexino Collaboration, *arXiV:2106.1097*. Submitted for publication on *EPJ C*.
28. V. Di Marcello *et al*, *Nucl. Instrum. Meth. A*, **964**, 163801 (2020).

Borexino detector performances

A. Caminata*, M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, R. Biondi, D. Bravo, B. Caccianiga, F. Calaprice, P. Cavalcante, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X.F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, A. Formozov, D. Franco, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, A.S. Göttel, M. Gromov, D. Guffanti, Aldo Ianni, Andrea Ianni, A. Jany, D. Jeschke, V. Kobychev, G. Korga, S. Kumaran, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, M. Nieslony, R. Nugmanov, L. Oberauer, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, L. Pelicci, Ö. Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M.T. Ranalli, G. Ranucci, A. Razeto, A. Re, M. Redchuk, A. Romani, N. Rossi, S. Schönert, D. Semenov, G. Settanta, M. Skorokhvatov, A. Singhal, O. Smirnov, A. Sotnikov, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, F. Villante, A. Vishneva, R.B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, S. Zavatarelli, K. Zuber, and G. Zuzel

**INFN Sezione di Genova, Genova, 16146, Italy*

Borexino, a large volume detector for low energy neutrino spectroscopy, is currently taking data underground since 2007 at the Laboratori Nazionali del Gran Sasso, Italy. The main goal of the experiment is the real-time measurement of solar neutrinos, especially the low energy part of the spectrum. Neutrinos are detected via neutrino-electron scattering in an ultra-pure organic liquid scintillator. The light generated by the interaction is detected by 2212 phototubes. During many years of data taking the experiment provided several remarkable results as the first evidence of pep neutrinos, the real-time detection of the pp neutrinos, the evidence of CNO neutrinos, and the detection of antineutrinos from the Earth. All these results are based on an accurate modelling of the detector's response and performances. The contribution shows the design, the modelling of the detector's response, and the performances. Moreover it will be discussed how the performances and the response were studied by means of extensive calibration campaigns.

Keywords: Borexino; Neutrino; Solar neutrinos; calibrations

1. The Borexino detector

Borexino is a liquid-scintillator experiment at the hall C of the Laboratori Nazionali del Gran Sasso in Italy. Borexino looks for solar neutrinos scattering off electrons in its active volume. Given the tiny cross-section of neutrino interactions with electrons ($\sigma \sim 10^{-44}\, cm^2$ to $10^{-45}\, cm^2$ for the solar-neutrino energy range), the rates expected in Borexino are small, ranging from less than one to a few tens of counts per day per 100 tons for different solar-neutrino components. To cope with such a low event rate, Borexino has a large target mass (about 300 t) and it is housed deep underground, under 3,800 m water equivalent of dolomitic rock that suppresses the flux of cosmic radiation by a factor of approximately one million. The active core of the detector consists of about 300 t of pseudocumene (1,2,4-trimethylbenzene)

doped with 1.5 g per litre of PPO (2,5-diphenyloxazole) and contained in a spherical nylon inner vessel (radius R = 4.25 m). The scintillator is surrounded by a non-scintillating pseudocumene-based buffer liquid which serves as a passive shield against external radioactivity (see Figure 1). The scintillator fluorescence light is collected by 2,212 photomultiplier tubes mounted on the Stainless Steel Sphere (radius R = 6.9 m). The entire detector is enclosed in a domed, cylindrical tank filled with high-purity water, equipped with 208 photomultiplier tubes, which provides extra shielding against external radioactivity (photons and neutrons), and also serves as an active water Cherenkov veto against residual cosmic muons.[1] Together with the lack of directionality information from the scintillation light, the low rate demands a high detector radiopurity, a deep understanding of the backgrounds, and an accurate modeling of the detector response.

To achieve the extreme sensitivity required by the experiment, very stringent procedures of material selection and cleaning were adopted to limit radioactive background in the experiment. Since usual materials are contaminated at the order of Bq/kg, material selection played a key role in the success of the experiment. One example is the choice of oil coming from deep wells to produce pseudocumene. In this way a low content of ^{14}C is achieved in the detector. Several cycles of scintillation purification were performed during the years achieving the unprecedented contamination of $^{238}U < 9.4 \cdot 10^{-20}$ g/g (95% C.L.) and $^{232}Th < 5.7 \cdot 10^{-19}$ g/g (95% C.L.). Further details on the purification strategy can be found here.[2]

The position and pulse-shape of each event are reconstructed by exploiting the number of detected photons and their detection times. The information about the event energy is carried by the number of detected photoelectrons or just the number of hit PMTs, as in our energy range the PMTs mainly work in a single photoelectron regime. Moreover, from the time distribution of the emitted light it is possible to statistically discriminate between α and β^- events[3] and between β^+ and β^- events.[4]

Due to a small leak in the nylon vessel, the vessel shape changes with time. At analysis stage it is updated every three weeks looking at radioactive decays happening on in (mostly due to ^{210}Bi, ^{40}K, ^{208}Tl) as shown by Figure 2. This method enables to reconstruct the vessel shape with a ± 5 cm uncertainty.

2. The Borexino calibration campaign

Calibrations in Borexino are performed inserting radioactive sources inside the liquid scintillator (internal calibrations) and placing a ^{232}Th source close to the stainless steel sphere (external calibrations). The former is used to map the detector's response while the latter is used to better understand and modeling the backgrounds that are coming from the construction materials. To insert sources inside the detector, a complex system of rods and cameras is exploited. This ensures knowing the position of the source at $\sim$cm level.[6] Table 1 shows the used sources and their purpose.

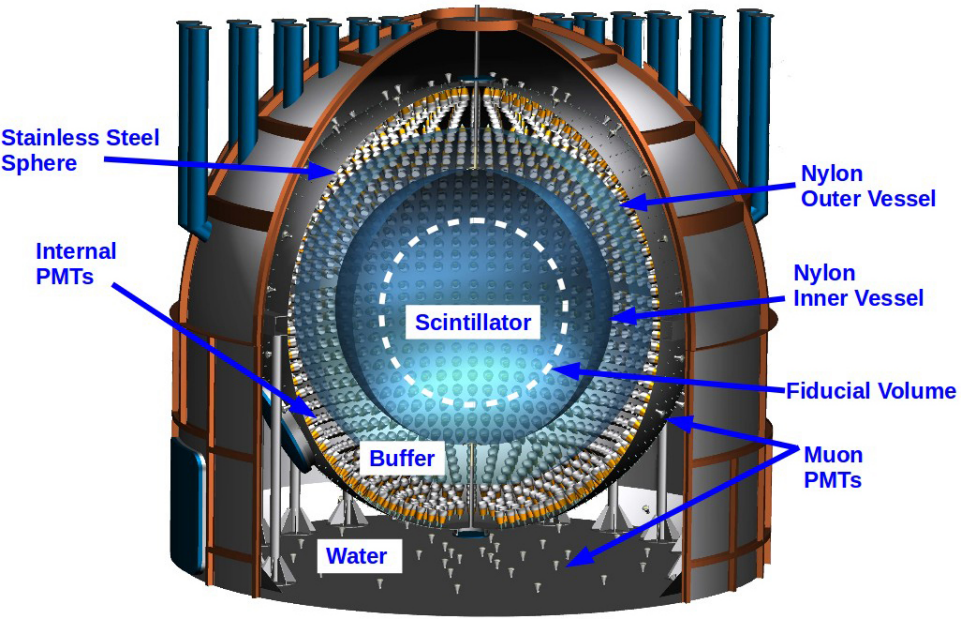


Fig. 1. The Borexino detector: Schematic view of the structure of the Borexino apparatus. From inside to outside: the liquid scintillator, the buffer liquid, the stainless steel sphere with the photomultipliers, and the water tank.[5]

Table 1. Used sources and placement during the internal calibration campaign.[6]

Source	Type	E [MeV]	Position	Motivations
^{57}Co	γ	0.122	in IV volume	Energy scale
^{139}Ce	γ	0.165	in IV volume	Energy scale
^{203}Hg	γ	0.279	in IV volume	Energy scale
^{85}Sr	γ	0.514	z-axis + sphere R=3 m	Energy scale + FV
^{54}Mn	γ	0.834	along z-axis	Energy scale
^{65}Zn	γ	1.115	along z-axis	Energy scale
^{60}Co	γ	1.173, 1.332	along z-axis	Energy scale
^{40}K	γ	1.460	along z-axis	Energy scale
^{222}Rn+^{14}C	β,γ	0-3.20	in IV volume	FV+uniformity
	α	5.5, 6.0, 7.4	in IV volume	FV+uniformity
^{241}Am^{9}Be	n	0-9	sphere R=4 m	Energy scale + FV
394 nm laser	light	-	center	PMT equalization

The position of an event is calculated by the time of arrival of the first photon on photomultipliers corrected by the hit multiplicity and minimizing a proper likelihood function. Figure 3 shows the resolution as function of the energy of the event. The reconstruction of calibration events and their comparison with the know source position shows a bias of maximum 4 cm in the vertical direction for events in the

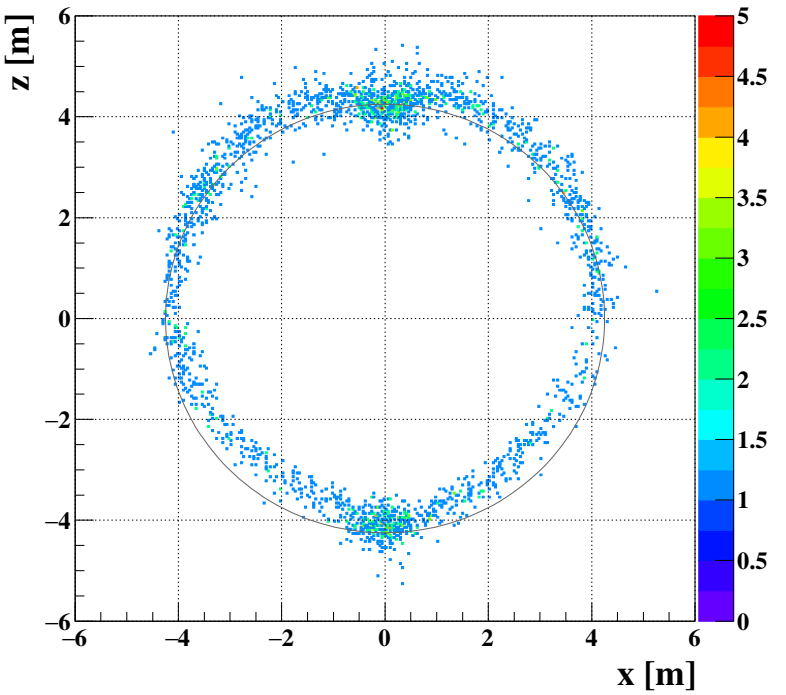


Fig. 2. A cross section (z-x plane, $|y| < 0.5$ m) view of the distribution of 2478 events, acquired during a 3 week period, and selected for the IV shape reconstruction. The color axis represents the number of events per cubic pixel of 0.04 m of size. This distribution reveals the IV shift and deformation with respect to its nominal spherical position shown in solid black line.[3]

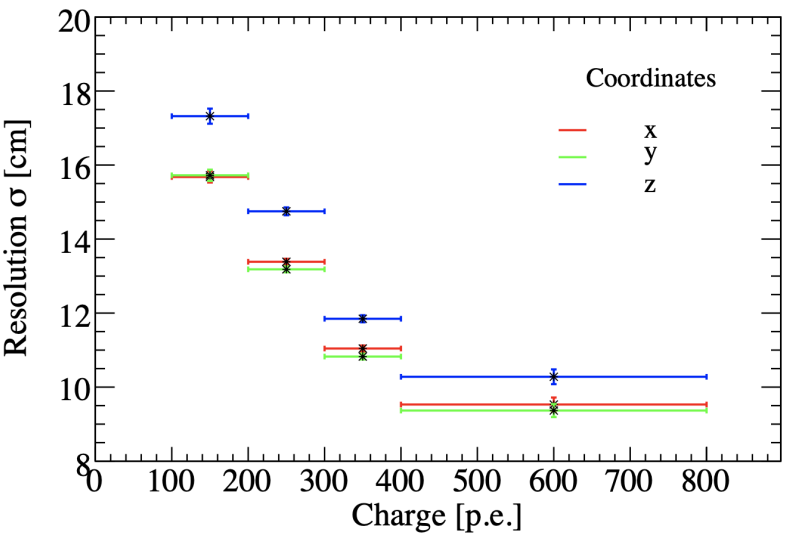


Fig. 3. Position resolution in Borexino as function of event energy.[6]

center. This bias has a small (less than 0.2%) impact on the fiducial volume determination. Since Borexino is a self triggering experiment, calibrations with laser and radioactive sources were performed to test the triggering efficiency, which is 100% at the energies of interest for solar neutrino detection.[7]

3. Modeling the detector response

The collaboration developed two independent approaches for modeling the detector's response. One is based on a particle physics simulation,[8] the other on an analytical model.[4]

The MC code developed for Borexino is a customized Geant4-based simulation package, which can simulate all processes following the interaction of a particle in the detector (energy loss including ionization quenching in the scintillator; scintillation and Cherenkov light production; optical photon propagation and interaction in the scintillator modelling absorption and re-emission, Rayleigh scattering, interaction of the optical photons with the surface of the materials; photon detection on the PMTs, and response of the electronics chain) including all known characteristics of the apparatus (geometry, properties of the materials, variable number of the working channels over the duration of the experiment as in the real data) and their evolution in time. The code thus produces a fully simulated detector response function because it provides a simulated version of all the measured physical variables. All the MC input parameters have been chosen or optimized using samples of data independent from the specific analysis (laboratory measurements and Borexino calibrations with radioactive sources.

In the analytical approach, a PDF for the energy estimator under consideration and analytical expressions for its mean value and variance are introduced. This PDF describes the detector's energy response function to monoenergetic events and, in brief, it is mainly influenced by the number of scintillation and Cherenkov photons and effects due to the nonuniformity of the light collection. The energy spectra of each species are transformed into the corresponding distributions of the energy estimators. Effects like the ionization quenching in the scintillator, the contribution of the Cherenkov light, the spatial dependence of the reconstructed energy and its resolution are accounted for through some parameters, part of which are fixed, while others are free to vary during the fit on the dataset.

4. Conclusion

Borexino is a neutrino detector able to measure the flux of all the neutrinos emitted by the Sun and the existence of antineutrinos emitted by the Earth.[3–5,9] Thanks to a wide calibration campaign the performances of the detector were assessed: for a 1 MeV event the energy resolution is ~ 50 keV and the position resolution is ~ 10 cm. The triggering efficiency is 100% for energies typical of solar neutrinos events. Moreover the Monte Carlo and analytical models were tested on calibration data with agreement better than 1%.

References

1. G. Alimonti, C. Arpesella, H. Back, M. Balata, D. Bartolomei, A. de Bellefon, G. Bellini, J. Benziger, A. Bevilacqua, D. Bondi, S. Bonetti, A. Brigatti, B. Caccianiga, L. Cadonati, F. Calaprice, C. Carraro, G. Cecchet, R. Cereseto, A. Chavarria, M. Chen, A. Chepurnov, A. Cubaiu, W. Czech, D. D'Angelo, F. Dalnoki-Veress, A. De Bari, E. De Haas, A. Derbin, M. Deutsch, A. Di Credico, A. Di Ludovico, G. Di Pietro, R. Eisenstein, F. Elisei, A. Etenko, F. von Feilitzsch, R. Fernholz, K. Fomenko, R. Ford, D. Franco, B. Freudiger, N. Gaertner, C. Galbiati, F. Gatti, S. Gazzana, V. Gehman, M. Giammarchi, D. Giugni, M. Goeger-Neff, T. Goldbrunner, A. Golubchikov, A. Goretti, C. Grieb, C. Hagner, T. Hagner, W. Hampel, E. Harding, S. Hardy, F. Hartmann, R. von Hentig, T. Hertrich, G. Heusser, M. Hult, A. Ianni, A. Ianni, L. Ioannucci, K. Jaenner, M. Joyce, H. de Kerret, S. Kidner, J. Kiko, T. Kirsten, V. Kobychev, G. Korga, G. Korschinek, Y. Kozlov, D. Kryn, P. La Marche, V. Lagomarsino, M. Laubenstein, C. Lendvai, M. Leung, T. Lewke, E. Litvinovich, B. Loer, F. Loeser, P. Lombardi, L. Ludhova, I. Machulin, S. Malvezzi, A. Manco, J. Maneira, W. Maneschg, I. Manno, D. Manuzio, G. Manuzio, M. Marchelli, A. Martemianov, F. Masetti, U. Mazzucato, K. McCarty, D. McKinsey, Q. Meindl, E. Meroni, L. Miramonti, M. Misiaszek, D. Montanari, M. Monzani, V. Muratova, P. Musico, H. Neder, A. Nelson, L. Niedermeier, S. Nisi, L. Oberauer, M. Obolensky, M. Orsini, F. Ortica, M. Pallavicini, L. Papp, R. Parcells, S. Parmeggiano, M. Parodi, N. Pelliccia, L. Perasso, A. Pocar, R. Raghavan, G. Ranucci, W. Rau, A. Razeto, E. Resconi, P. Risso, A. Romani, D. Rountree, A. Sabelnikov, P. Saggese, R. Saldhana, C. Salvo, R. Scardaoni, D. Schimizzi, S. Schönert, K. Schubeck, T. Shutt, F. Siccardi, H. Simgen, M. Skorokhvatov, O. Smirnov, A. Sonnenschein, F. Soricelli, A. Sotnikov, S. Sukhotin, C. Sule, Y. Suvorov, V. Tarasenkov, R. Tartaglia, G. Testera, D. Vignaud, S. Vitale, R. Vogelaar, V. Vyrodov, B. Williams, M. Wojcik, R. Wordel, M. Wurm, O. Zaimidoroga, S. Zavatarelli and G. Zuzel, The borexino detector at the laboratori nazionali del gran sasso, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **600**, 568 (2009).
2. J. Benziger, L. Cadonati, F. Calaprice, M. Chen, A. Corsi, F. Dalnoki-Veress, R. Fernholz, R. Ford, C. Galbiati, A. Goretti, E. Harding, A. Ianni, A. Ianni, S. Kidner, M. Leung, F. Loeser, K. McCarty, D. McKinsey, A. Nelson, A. Pocar, C. Salvo, D. Schimizzi, T. Shutt and A. Sonnenschein, A scintillator purification system for the borexino solar neutrino detector, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **587**, 277 (2008).
3. M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, B. Caccianiga, F. Calaprice, A. Caminata, L. Cappelli, P. Cavalcante, F. Cavanna, A. Chepurnov, K. Choi, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X. F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, G. Fiorentini, A. Formozov, D. Franco, F. Gabriele, C. Galbiati, M. Gschwender, C. Ghiano, M. Giammarchi, A. Goretti, M. Gromov, D. Guffanti, C. Hagner, E. Hungerford, A. Ianni, A. Ianni, A. Jany, D. Jeschke, S. Kumaran, V. Kobychev, G. Korga, T. Lachenmaier, T. Lasserre, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, F. Mantovani, G. Manuzio, S. Marcocci, J. Maricic, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, M. Montuschi, V. Muratova, B. Neumair, M. Nieslony, L. Oberauer, A. Onillon, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, O. Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M. T. Ranalli, G. Ranucci,

A. Razeto, A. Re, M. Redchuk, B. Ricci, A. Romani, N. Rossi, S. Rottenanger, S. Schönert, D. Semenov, M. Skorokhvatov, O. Smirnov, A. Sotnikov, V. Strati, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, A. Vishneva, M. Vivier, R. B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, O. Zaimidoroga, S. Zavatarelli, K. Zuber and G. Zuzel, Comprehensive geoneutrino analysis with borexino, *Phys. Rev. D* **101**, p. 012009 (Jan 2020).

4. M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, B. Caccianiga, F. Calaprice, A. Caminata, S. Caprioli, M. Carlini, P. Cavalcante, A. Chepurnov, K. Choi, L. Collica, D. D'Angelo, S. Davini, A. Derbin, X. F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, K. Fomenko, A. Formozov, D. Franco, F. Gabriele, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, M. Gromov, D. Guffanti, C. Hagner, T. Houdy, E. Hungerford, A. Ianni, A. Ianni, A. Jany, D. Jeschke, V. Kobychev, D. Korablev, G. Korga, D. Kryn, M. Laubenstein, E. Litvinovich, F. Lombardi, P. Lombardi, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, G. Manuzio, S. Marcocci, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, L. Oberauer, B. Opitz, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, Ö. Penek, N. Pilipenko, A. Pocar, A. Porcelli, G. Raikov, G. Ranucci, A. Razeto, A. Re, M. Redchuk, A. Romani, R. Roncin, N. Rossi, S. Schönert, D. Semenov, M. Skorokhvatov, O. Smirnov, A. Sotnikov, L. F. F. Stokes, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, M. Toropova, E. Unzhakov, F. L. Villante, A. Vishneva, R. B. Vogelaar, F. von Feilitzsch, H. Wang, S. Weinz, M. Wojcik, M. Wurm, Z. Yokley, O. Zaimidoroga, S. Zavatarelli, K. Zuber, G. Zuzel and T. B. Collaboration, Comprehensive measurement of pp-chain solar neutrinos, *Nature* **562**, 505 (2018).
5. M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, R. Biondi, D. Bravo, B. Caccianiga, F. Calaprice, A. Caminata, P. Cavalcante, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X. F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, A. Formozov, D. Franco, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, A. S. Göttel, M. Gromov, D. Guffanti, A. Ianni, A. Ianni, A. Jany, D. Jeschke, V. Kobychev, G. Korga, S. Kumaran, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, M. Nieslony, R. Nugmanov, L. Oberauer, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, L. Pelicci, Ö. Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M. T. Ranalli, G. Ranucci, A. Razeto, A. Re, M. Redchuk, A. Romani, N. Rossi, S. Schönert, D. Semenov, G. Settanta, M. Skorokhvatov, A. Singhal, O. Smirnov, A. Sotnikov, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, F. L. Villante, A. Vishneva, R. B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, S. Zavatarelli, K. Zuber, G. Zuzel and T. B. Collaboration, Experimental evidence of neutrinos produced in the cno fusion cycle in the sun, *Nature* **587**, 577 (2020).
6. H. Back, G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, M. B. Avanzini, B. Caccianiga, L. Cadonati, F. Calaprice, C. Carraro, P. Cavalcante, A. Chavarria, A. Chepurnov, D. D. Angelo, S. Davini, A. Derbin, A. Etenko, F. von Feilitzsch, G. Fernandes, K. Fomenko, D. Franco, C. Galbiati, S. Gazzana, C. Ghiano, M. Giammarchi, M. Goeger-Neff, A. Goretti, L. Grandi, E. Guardincerri, S. Hardy, A. Ianni, A. Ianni, A. Kayunov, S. Kidner, V. Kobychev, D. Korablev, G. Korga, Y. Koshio, D. Kryn, M. Laubenstein, T. Lewke, E. Litvinovich, B. Loer, F. Lombardi, P. Lombardi, L. Ludhova, I. Machulin, S. Manecki, W. Maneschg, G. Manuzio, Q. Meindl,

E. Meroni, L. Miramonti, M. Misiaszek, D. Montanari, P. Mosteiro, V. Muratova, L. Oberauer, M. Obolensky, F. Ortica, K. Otis, M. Pallavicini, L. Papp, L. Perasso, S. Perasso, A. Pocar, R. S. Raghavan, G. Ranucci, A. Razeto, A. Re, A. Romani, N. Rossi, D. Rountree, A. Sabelnikov, R. Saldanha, C. Salvo, S. Schönert, H. Simgen, M. Skorokhvatov, O. Smirnov, A. Sotnikov, S. Sukhotin, Y. Suvorov, R. Tartaglia, G. Testera, D. Vignaud, R. B. Vogelaar, J. Winter, M. Wojcik, A. Wright, M. Wurm, J. Xu, O. Zaimidoroga, S. Zavatarelli and G. Zuzel, Borexino calibrations: hardware, methods, and results, *Journal of Instrumentation* **7**, P10018 (oct 2012).

7. G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, M. Buizza Avanzini, B. Caccianiga, L. Cadonati, F. Calaprice, P. Cavalcante, A. Chavarria, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Empl, A. Etenko, K. Fomenko, D. Franco, F. Gabriele, C. Galbiati, S. Gazzana, C. Ghiano, M. Giammarchi, M. Göger-Neff, A. Goretti, L. Grandi, M. Gromov, C. Hagner, E. Hungerford, A. Ianni, A. Ianni, V. Kobychev, D. Korablev, G. Korga, D. Kryn, M. Laubenstein, T. Lewke, E. Litvinovich, B. Loer, F. Lombardi, P. Lombardi, L. Ludhova, G. Lukyanchenko, I. Machulin, S. Manecki, W. Maneschg, G. Manuzio, Q. Meindl, E. Meroni, L. Miramonti, M. Misiaszek, M. Montuschi, P. Mosteiro, V. Muratova, L. Oberauer, M. Obolensky, F. Ortica, K. Otis, M. Pallavicini, L. Papp, C. Pena-Garay, L. Perasso, S. Perasso, A. Pocar, G. Ranucci, A. Razeto, A. Re, A. Romani, N. Rossi, R. Saldanha, C. Salvo, S. Schönert, H. Simgen, M. Skorokhvatov, O. Smirnov, A. Sotnikov, S. Sukhotin, Y. Suvorov, R. Tartaglia, G. Testera, D. Vignaud, R. B. Vogelaar, F. von Feilitzsch, J. Winter, M. Wojcik, A. Wright, M. Wurm, J. Xu, O. Zaimidoroga, S. Zavatarelli and G. Zuzel, Final results of borexino phase-i on low-energy solar neutrino spectroscopy, *Phys. Rev. D* **89**, p. 112007 (Jun 2014).
8. M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, D. Bick, G. Bonfini, L. Borodikhina, D. Bravo, B. Caccianiga, F. Calaprice, A. Caminata, M. Canepa, S. Caprioli, M. Carlini, P. Cavalcante, A. Chepurnov, K. Choi, D. D'Angelo, S. Davini, A. Derbin, X. Ding, L. Di Noto, I. Drachnev, K. Fomenko, A. Formozov, D. Franco, F. Froborg, F. Gabriele, C. Galbiati, C. Ghiano, M. Giammarchi, M. Goeger-Neff, A. Goretti, M. Gromov, C. Hagner, T. Houdy, E. Hungerford, A. Ianni, A. Ianni, A. Jany, D. Jeschke, V. Kobychev, D. Korablev, G. Korga, D. Kryn, M. Laubenstein, E. Litvinovich, F. Lombardi, P. Lombardi, L. Ludhova, G. Lukyanchenko, I. Machulin, M. Magnozzi, G. Manuzio, S. Marcocci, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, L. Oberauer, B. Opitz, F. Ortica, M. Pallavicini, L. Papp, A. Pocar, G. Ranucci, A. Razeto, A. Re, A. Romani, R. Roncin, N. Rossi, S. Schönert, D. Semenov, P. Shakina, M. Skorokhvatov, O. Smirnov, A. Sotnikov, L. Stokes, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, M. Toropova, E. Unzhakov, A. Vishneva, R. Vogelaar, F. von Feilitzsch, H. Wang, S. Weinz, M. Wojcik, M. Wurm, Z. Yokley, O. Zaimidoroga, S. Zavatarelli, K. Zuber and G. Zuzel, The monte carlo simulation of the borexino detector, *Astroparticle Physics* **97**, 136 (2018).
9. G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, B. Caccianiga, L. Cadonati, F. Calaprice, A. Caminata, P. Cavalcante, A. Chavarria, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Empl, A. Etenko, K. Fomenko, D. Franco, F. Gabriele, C. Galbiati, S. Gazzana, C. Ghiano, M. Giammarchi, M. Göger-Neff, A. Goretti, M. Gromov, C. Hagner, E. Hungerford, A. Ianni, A. Ianni, V. Kobychev, D. Korablev, G. Korga, D. Kryn, M. Laubenstein, B. Lehnert, T. Lewke, E. Litvinovich, F. Lombardi, P. Lombardi, L. Ludhova, G. Lukyanchenko, I. Machulin, S. Manecki, W. Maneschg,

S. Marcocci, Q. Meindl, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, M. Montuschi, P. Mosteiro, V. Muratova, L. Oberauer, M. Obolensky, F. Ortica, K. Otis, M. Pallavicini, L. Papp, L. Perasso, A. Pocar, G. Ranucci, A. Razeto, A. Re, A. Romani, N. Rossi, R. Saldanha, C. Salvo, S. Schönert, H. Simgen, M. Skorokhvatov, O. Smirnov, A. Sotnikov, S. Sukhotin, Y. Suvorov, R. Tartaglia, G. Testera, D. Vignaud, R. B. Vogelaar, F. von Feilitzsch, H. Wang, J. Winter, M. Wojcik, A. Wright, M. Wurm, O. Zaimidoroga, S. Zavatarelli, K. Zuber, G. Zuzel and B. Collaboration, Neutrinos from the primary proton–proton fusion process in the sun, *Nature* **512**, 383 (2014).

Study of antineutrinos from the Earth and the Cosmos with the Borexino detector

Sandra Zavatarelli*, M. Agostini, K. Altenmuller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini,J. Benziger, R. Biondi, D. Bravo, B. Caccianiga, A. Caminata, F. Calaprice, P. Cavalcante, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X.F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, A. Formozov, D. Franco, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, A.S. Gottel, M. Gromov, D. Guffanti, Aldo Ianni, Andrea Ianni, A. Jany, D. Jeschke, V. Kobychev, G. Korga, S. Kumaran, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, M. Nieslony, R. Nugmanov, L. Oberauer, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, L. Pelicci, O Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M.T. Ranalli, G. Ranucci, A. Razeto, A. Re, M. Redchuk, A. Romani, N. Rossi, S. Schonert, D. Semenov, G. Settanta, M. Skorokhvatov, A. Singhal, O. Smirnov, A. Sotnikov, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, F. Villante, A. Vishneva, R.B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, S. Zavatarelli, K. Zuber, and G. Zuzel (Borexino Collaboration)

**Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Via Dodecaneso 33, 16146 Genoa, Italy*
E-mail: sandra.zavatarelli@ge.infn.it

The largest amount of antineutrinos detected about the Earth is emitted by the natural radioactive decays of ^{232}Th and ^{238}U chains isotopes and of ^{40}K. Other flux components are yielded by cosmic rays interactions in the atmosphere or by possible extra-terrestrial sources such as supernovae explosions, gamma ray bursts, GW events and solar flares. This contribution is aimed to summarise the results obtained by the Borexino experiment about antineutrinos from the Earth and from extraterrestrial sources.

Keywords: Antineutrinos; Geonetrinos; solar flares; diffuse supernovae background.

1. Introduction

Large underground ultrapure liquid scintillators are very suitable to antineutrinos studies. Low energy neutrinos such as solar neutrinos are in general detected through the elastic scattering process on electrons but for electron antineutrinos a specific interaction channel can be exploited, the Inverse Beta-Decay (IBD) mechanism on the free proton:

$$\bar{\nu}_e + p \rightarrow n + e^+ \tag{1}$$

The emitted positron deposits its kinetic energy and annihilates almost immediately, inducing a prompt signal whose visible energy E_{prompt} is directly correlated with the incident antineutrino energy $E_{\bar{\nu}_e}$:

$$E_{prompt} = E_{\bar{\nu}_e} - 0.784\,\mathrm{MeV}. \tag{2}$$

The neutron thermalizes and is captured by a proton with the mean capture time: $\sim \tau = 254.5 \pm 1.8 \mu$s.[1] The IBD cross section is on average a factor ~ 100 higher than elastic scattering at few MeV. Moreover the fast time coincidence between the positron annihilation and the neutron capture provides an almost background free signature, that allows to investigate also tiny flux components. By exploiting this detection mechanism, Borexino[2] has robustly detected the geo-neutrino signal and begun to place constraints on the amount of radiogenic heating in the Earth's interior: the null-hypothesis of observing a geoneutrino signal from the mantle has been excluded at a 99.0% C.L. and the overall production of radiogenic heat constrained to $38.2\ ^{+13.6}_{-12.7}$ TW.[3] The extreme radiopurity of the Borexino detector has also allowed to set new limits on diffuse supernova antineutrino background for in the previously unexplored energy region below 8 MeV,[4] and to obtain the best upper limits on all flavor antineutrino fluences in the few MeV energy range from solar flares,[4] gamma-ray bursts[5] and from gravitational wave events.[6] In the following sections we detail the analysis procedures and the results obtained by Borexino about geoneutrinos, diffuse supernovae background and possible time correlated signals with solar flares.

2. Geo-neutrinos

Geoneutrinos are electron antineutrinos ($\bar{\nu}_e$) emitted in the β decays of long-lived isotopes, which are present in the Earth, such as ^{40}K and the ^{238}U and ^{232}Th chain unstable isotopes.[11] The decay schemes and the released heat are summarised in the following equations:

$$^{238}\mathrm{U} \rightarrow ^{206}\mathrm{Pb} + 8\alpha + 8e^- + 6\bar{\nu}_e + 51.7 \quad \mathrm{MeV} \tag{3}$$

$$^{235}\mathrm{U} \rightarrow ^{207}\mathrm{Pb} + 7\alpha + 4e^- + 4\bar{\nu}_e + 46.4\ \mathrm{MeV} \tag{4}$$

$$^{232}\mathrm{Th} \rightarrow ^{208}\mathrm{Pb} + 6\alpha + 4e^- + 4\bar{\nu}_e + 42.7\ \mathrm{MeV} \tag{5}$$

$$^{40}\mathrm{K} \rightarrow ^{40}\mathrm{Ca} + e^- + \bar{\nu}_e + 1.31\ \mathrm{MeV}\ (89.3\%) \tag{6}$$

$$^{40}\mathrm{K} + \mathrm{e} \rightarrow ^{40}\mathrm{Ar} + \nu_e + 1.505\ \mathrm{MeV} \quad (10.7\%) \tag{7}$$

Since the present-day isotopic abundance of ^{235}U is small (0.7%), the overall contribution of ^{238}U, ^{232}Th, and ^{40}K is largely predominant. Information about the deep Earth's composition is mainly based on indirect probes: seismologic data constrains the density profile, while geochemistry provides predictions based on chemical compositions of upper Earth's rocks, chondritic meteorites, and the Sun photosphere. The surface of the Earth is emitting an heat flow at an estimated rate of 47 ± 2 TW.[12] This result is based on $\sim$ 40,000 observations of heat conduction in surface rocks and corrected for hydrothermal circulation and volcanism. Measurements of Th, U and K in the continental crust rocks prove that at least $\sim$8 TW of this heat

is from radiogenic decays but is not at all known what proportion of the overall surface heat flow comes from radioactive decays and how much is from primordial sources such as the gravitational energy released after core formation or the kinetic energy of accretion bombardment.

Geo-neutrinos give us a way of measuring the amount of this radiogenic heat directly: the heat released is in a well fixed ratio with the total mass of Heat Producing Elements (HPE's) inside the Earth. Thus, it is possible to extract from the measured geo-neutrino fluxes several geological information unreachable by other means. The idea of studying geo-neutrinos dates back to the sixties[7,8] but only in 1998 it was proposed the idea of using solar neutrino and reactor neutrino detectors[9,10]: presently, two large-volume, liquid-scintillator neutrino experiments, KamLAND in Japan[13–15] and Borexino in Italy,[3,16–18] have been able to measure the geo-neutrino signal. The typical fluxes are of the order of 10^6 cm^{-2}s^{-1} and with cross section values of 10^{-43} cm^{-2} only a hand-full number of interactions, few tens per year are expected with the current-size detectors. This means, underground laboratories are crucial to shield the detector from cosmic radiation and make possible their measurement. Since the inverse beta decay (IBD) on protons has a threshold of 1.8 MeV (Eq. 1) and both the ^{40}K and ^{235}U geo-neutrinos spectra are below (Fig. 1), they cannot be detected by this process. However, the elemental abundances ratios are much better known than the absolute abundances. Therefore, by measuring the absolute content of ^{238}U and ^{232}Th, also the overall amount of ^{40}K and ^{235}U can be inferred with an improved precision.

Geo-neutrino spectrum extends up to 3.26 MeV and the ^{238}U and ^{232}Th contributions can be distinguished according to their different end-points (Fig. 1).

The most recent Borexino analysis is based on 3263 days of data taking in the period December 2007 - April 2019.

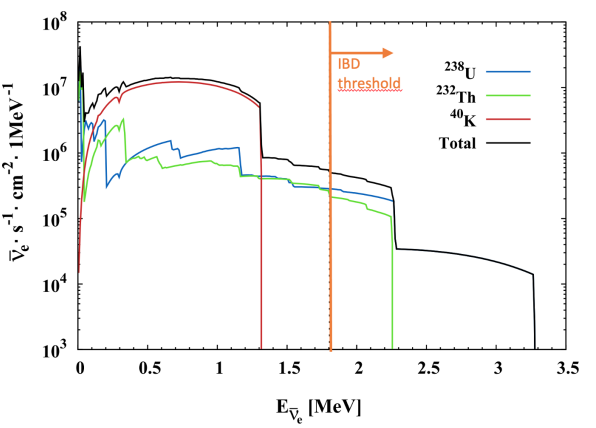


Fig. 1. Geoneutrino fluxes[3] from different isotopes and the overall signal at LNGS calculated with geophysical and geochemical inputs from[24] for the far-field lithosphere and from[23] for the local crust.

To select the $\bar{\nu}_e$ candidates we apply the following criteria: we discard events occurring within 2 ms of every muon crossing the outer detector to reject cosmogenic neutrons. In case of muons crossing the inner detector long-lived cosmogenic nuclides such a ^{9}Li may be produced: to reduce this possible background a different veto ranging from 0.2 to 2 s is applied after the muon depending on the deposited energy, the track reconstruction quality and the eventually associated neutrons. Independently from a detected muon, we ask that selected events are neither preceded or followed by neutron-like events within a 2 ms window. The coincidences are searched for at a distance between the reconstructed positions Δr < 1.3 m, that accounts for the uncertainty of the spatial reconstruction algorithm and the free path of the 2.2 MeV γ's and with Δt in the range 20–1280 μs ($\sim$5 neutron capture time) for single events, and 2.5–12.5 μs for double cluster events. We remind that if a decay triggering the Borexino DAQ is followed within 16.5 μs by other decays, it is classified as a unique event with multiple clusters. To reduce the external γ background from radioactive decays in the detector materials, we accept only candidates having the prompt event position reconstructed inside the inner detector at distance larger than 10 cm with respect to the the time-varying IV surface: $D_{IV,prompt} > 10\,cm$. The energy of the prompt event is required to be above the value corresponding to the IBD threshold, considering the energy resolution ($N_{p.e.,prompt} > 408$ p.e.), while for the delayed event, the energy cut was tuned to cover the gamma peak from neutron capture on proton and on carbon ($700 < N_{p.e.,delayed} < 3000$ p.e.). The lower limit is justified because photons at the edge of the scintillator can escape, depositing only a fraction of their total energy. This limit was increased to 860 p.e. during the detector purifications in 2010-2011 to remove the time-correlated $\beta + (\alpha + \gamma)$ decays of ^{214}Bi-^{214}Po, having a time constant close to the neutron capture time. Such background was relevant only during the purifications as a consequence of the increased radon contamination. Finally a pulse shape $\alpha - \beta$ discrimination is applied based on a trained MultiLayerPerceptron binary classifier [3] : $\mathrm{MLP}_D > 0.8$ for delayed signals.

The combined efficiency of the cuts is determined by Monte Carlo to be (85$\pm$1.0)%. The total efficiency-corrected exposure for the data set is (1.29 $\pm$ 0.05) $\times$ 10^{32} protons $\times$ year with an increase by a factor of two over the previous Borexino analysis reported in 2015. We have identified 154 candidates passing all the selection cuts. The estimated background, that mimics $\bar{\nu}_e$ candidates, amounts to 8.28 $\pm$ 1.01 events, mainly due to random coincidences, residual ^{9}Li decays and ^{13}C(α,n)^{16}O reactions in the scintillator.

Antineutrinos from nuclear reactor power plants are the main background to the geo-neutrino measurements. The $\bar{\nu}_e$ flux comes primarily from the beta decays of neutron-rich fragments produced in the fission of four isotopes: ^{235}U, ^{238}U, ^{239}Pu, and ^{241}Pu. The expected fluxes can be estimated from the knowledge of the monthly energy production at each reactor site, including the neutrino propagation effects. At present, there are about 440 nuclear power reactors in the world, providing,

nominally, a total amount of about 1200 Thermal GW, corresponding to about 400 Electrical GW. Since there are no nuclear power plants close by, the Gran Sasso laboratory is well suited for geo-neutrino studies. According to a detailed computation[22] based on the information retrieved by the International Atomic Energy Agency database[19] we expect 91-98 events in our sample whereas the main source of uncertainty in the prediction comes from the knowledge of the energy spectrum for each fuel component at the source. The lowest value (91.9 ± 1.6) is obtained when the spectra are normalised to the measurements of reactor neutrino experiments[20,21] and they include also the so called "5 MeV excess".

A smaller background can also be induced by atmospheric neutrinos. Because of the uncertainty of the relevant fluxes and cross sections, this contribution is poorly constrained: we have estimated 9.2 ± 4.6 event in our sample.

Both reactor and atmospheric $\bar{\nu}_e$ are more energetic respect to the geo-neutrino; therefore they can be experimentally disentangled thanks to a fit of prompt event energies. An unbinned likelihood fit of the energy spectrum of selected prompt events has been performed.

In Fig. 2 the selected candidates and the fit results are shown. Using the value ratio for the masses of Th and U, m(Th)/m(U)=3.9, suggested by the chondritic metheorithes, our best fit yield $N_{geo} = 52.6^{+9.4}_{-8.6}(stat)^{+2.7}_{-2.1}(sys)$ events, with 18% precision. A compatible result was found when contributions from ^{238}U and ^{232}Th were both fit as free parameters. Antineutrino background from reactors is fit unconstrained and found compatible with the expectations. The measured geo-neutrino signal corresponds to fluxes at the detector from decays in the U and Th chains of $\Phi_U = 2.8^{+0.6}_{-0.4} \cdot 10^6$ cm^{-2}s^{-1} and $\Phi_{Th} = 2.6^{+0.6}_{-0.4} \cdot 10^6$ cm^{-2}s^{-1} respectively.

To connect the number of geo-neutrinos detected by an experiment with the overall radioactivity in the crust and in the mantle, we have to rely on our knowledge

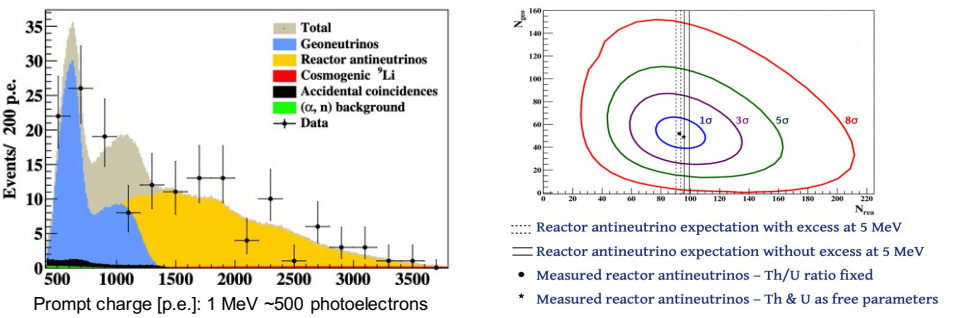


Fig. 2. **(Left)** Spectral fit of the data (black points assuming the chondritic Th/U ratio. Geoneutrinos (blue) and reactor antineutrinos (yellow) were kept as free fit parameters. Other non antineutrino backgrounds were constrained in the fit. **(Right)** The best fit point (black dot) and the 2D coverage contours for N_{geo} vs N_{rea}.

of the local geological conditions. Geoscientist developed models of the crust close to the Borexino site, based on geological surveys and on 3D models of all the layers up to the mantle boundary.[23,24] Both U and Th are lithophile i.e. "rock-loving" elements so they are believed to be concentrated in the crust and mantle and absent in the iron-made core. According to these studies 28.8 ± 5.6 events should have been induced by the crust radioactivity among the 154 candidates: we have therefore repeated the fit with separate contribution of mantle and crust geoneutrinos, the latter being constrained to the expectations within the error. A small difference in the relative amount of U and Th fraction in the two signals has been introduced according to geological predictions. The fit returns $N_{geo} = 23.7^{+10.7}_{-10.0}(stat)^{+1.2}_{-1.0}(sys)$ events: with a sensitivity study the null-hypothesis of observing geoneutrino signal from the mantle is excluded at a 99.0% C.L..

It is possible to extract from the measured geo-neutrino signal, the Earth's radiogenic heat power that is a fundamental question for understanding the plate tectonics and mantle convection. The uncertain distribution of HPE's in the deep Earth has to be included: two rather extreme cases are represented by the homogeneous mantle or by the assumption that all the U and Th are placed at the mantle core boundary. By following this conservative approach one can demonstrate that our measured mantle signal corresponds to the production of a radiogenic heat of 24.6 ± 11.1 TW (68% interval) from ^{238}U and ^{232}Th in the mantle. Assuming 18% contribution of ^{40}K in the mantle and 8.1 ± 1.9 TW of total radiogenic heat of the lithosphere, the Borexino estimate of the total radiogenic heat of the Earth is 38.2 ± 13.6 TW, which corresponds to ∼81% of the overall Earth's heat flux (47 TW). These values are compatible with different geological predictions, in the particular with the geodynamical BSE models,[25] however there is a $\sim 2.4\sigma$ tension with those Earth models which predict the lowest concentration of heat-producing elements in the mantle, such as cosmochemical ones.[25] In addition, by constraining the number of expected reactor antineutrino events, the existence of a hypothetical georeactor at the center of the Earth having power greater than 2.4 TW is excluded at 95% C.L..[26,27] In conclusion, Borexino confirms the feasibility of geoneutrino measurements as well as the validity of different geological models predicting the U and Th abundances in the Earth. This is an enormous success of both neutrino physics and geosciences. The next generation of large volume liquid scintillator detectors (such as JUNO and SNO+) has a strong potential to provide fundamental information about our planet and geoneutrinos may be the key tool to support the new discoveries about the deep Earth.

3. Diffuse supernovae neutrino background

The Diffuse Supernova Neutrino Background (DSNB) is formed by the stars that collapsed during the evolution of the Universe and it is made of neutrinos and antineutrinos of all flavours. The flux spectrum at the Earth can be parametrised

as[28]:

$$\frac{d\phi_\nu}{dE_\nu} = \frac{c}{H_0}\int_0^{z_{max}} \frac{dN_\nu(E'_\nu)}{dE'_\nu} \frac{R_{SN}(z)dz}{\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}}, \tag{8}$$

where c is the speed of light, z is the red shift, H_0 is Hubble constant, $R_{SN}(z)$ is the supernova rate at the distance z to the observer $\frac{dN_\nu(E'_\nu)}{dE'_\nu}$ is the neutrino emission spectrum for individual supernova, Ω_m and Ω_Λ are the relative densities of matter and dark energy in the Universe. The spectrum is sensitive to particular cosmological model through Ω_m and Ω_Λ and reflects the expansion of the Universe through the dependence on H_0.

The possible signal due to tiny, still undisclosed, extraterrestrial $\bar{\nu}_e$ fluxes, such as supernovae relic neutrinos, can be put in evidence as an excess of events with respect to the backgrounds and the known sources of $\bar{\nu}_e$. Borexino made a search for this possible signal, based on the data collected between December 2007–October 2017. The events selection is based on an approach similar to the one adopted in the geo-neutrino analysis, the only differences being more conservative muon and FV cuts. In detail we applied a veto of 2 s after each muon crossing the inner detector and we asked for a prompt event distance from the inner vessel larger than 25 cm: in the overall statistics we have selected 101 candidates.

As outlined in the previous sections, the most relevant sources of $\bar{\nu}_e$ events below 10 MeV are the Earth's radioactive isotopes and the nuclear reactors, while at higher energies the atmospheric neutrino background dominates the energy spectrum. Borexino made a comprehensive study of geo-neutrinos but the achieved precision, in particular on the mantle signal is still poor ($\sim$ 40%), With the aim to quote conservative limits, the minimal expected number of events for each background has been considered. For the geoneutrino signal we have therefore chosen the Minimal Radiogenic Earth model, which only includes the radioactivity from the crust and that, in our case, corresponds to 17.9 ± 2.1 events in our data sample.

The spectrum of reactor $\bar{\nu}_e$ is significant till $\sim$10 MeV. As discussed before, the flux calculation that includes the normalisation to reactor neutrino experiments data provides the lowest signal. We have taken this option being the most conservative: 61.1 ± 1.7 events among our candidates have been attributed to reactor $\bar{\nu}_e$. Finally the most serious background to the detection of DSNB fluxes at energies above 10 MeV is induced by the atmospheric neutrinos, i.e., the ν's and $\bar{\nu}$'s generated in the decay of secondary particles produced in the interactions of primary cosmic rays with Earth's atmosphere. After the simulation of detector response and by applying the same selection cuts as for real data, we have estimated 6.5 ± 3.2 IBD-like events in the present analysis statistics and in the equivalent $\bar{\nu}_e$ energy window 1.8–16.8 MeV. In the light of the large uncertainty on this background source, a conservative choice would be to not consider it at all in the upper limit calculations: this is the approach followed in Fig. 3. The model-independent limit for electron antineutrino

flux ($\Phi_{\bar{\nu}_e}$) in each energy bin (i) have been computed according to the equation:

$$\Phi_{\bar{\nu}_e,i} = \frac{N_{90,i}}{<\sigma> \cdot \varepsilon \cdot N_p \cdot T} \tag{9}$$

where N_{90} is the 90% C.L. upper limit for the number of antineutrino interactions obtained by following the Feldman-Cousins approach,[29] $<\sigma>$ is the mean cross-section of Inverse Beta Decay calculated according to[30] for each energy bin, ε=(0.850 $\pm$ 0.015) is the average detection efficiency, $N_p = (1.32 \pm 0.06) \times 10^{31}$ is the number of protons in the Borexino average fiducial volume mass and $T = 2485$ days is the total live-time. The other limits existing in literature are quoted on the same plot.

Borexino limits are the only existing below 8 MeV, thanks to the high energy resolution, the low intrinsic backgrounds, and the small reactor $\bar{\nu}_e$ flux at the Gran Sasso site.

4. Search for time correlated signals with solar flares

Solar flares are induced by the rearrangement of the solar magnetic field which also bring to the acceleration of charged particles. Pions, eventually produced in the flare's region by *pp*- and *pα*-collisions, could then decay by emitting neutrinos with a mean neutrino energy expected around $\sim 10\,\mathrm{MeV}$.[35]

In the eighties R. Davis[36,37] advanced for the first time the possibility of neutrino emissions correlated with solar flares as an explanation for the excess of events in several runs of the Homestake Cl-Ar experiment. In general the production of $\bar{\nu}_e$ is expected to be smaller respect to ν_e due to the higher threshold of π^- generation in *pp*-collisions: for this reason we have searched for ν_x and $\bar{\nu}_x$ ($x = e, \mu, \tau$)

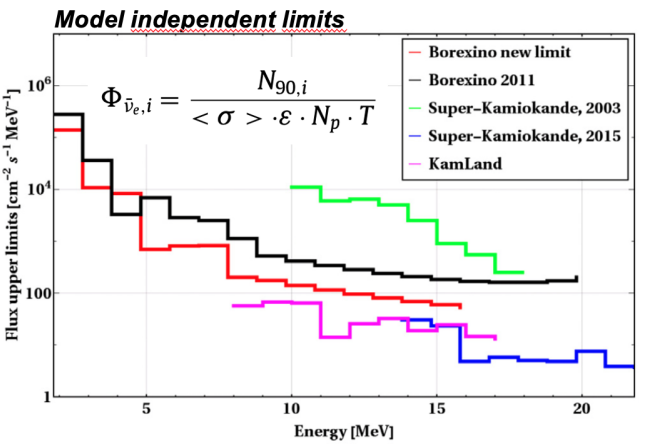


Fig. 3. Borexino model-independent limits on electron $\bar{\nu}_e$ fluxes from unknown sources compared with the results of other experiments (Super-Kamiokande,[31,32] KamLAND[33]) and older Borexino limits.[34]

signals correlated with solar flares by looking their elastic scattering on electrons in the Borexino scintillator. According to the assumption that the neutrino flux is proportional to the flare's intensity, we have only considered the most intense flares of M and X classes according to the GOES database.[38]

The analysis is based on the statistics collected between November 2009 and October 2017 : data acquired both by the primary and the FADC DAQ systems have been used in the energy interval of 1–15 MeV, while in the interval 0.25 – 1 MeV only the primary DAQ data have been considered. After removing muons and muon daughters, single events are selected without any fiducial volume cut in the interval of 1–15 MeV, while at lower energies a fiducial volume cut of 145 tons (75 cm from the IV) is applied to reduce the external radioactivity. In our study, an excess of single events above the measured background at the time of a flare is searched for. For the flare signal we choose a time window equal to the flare's duration according to the database, while the background is calculated in a time window of the same length but opened before the flare time. By requiring at least 95% of Borexino's up time for both windows we restrict our analysis to a 472 flares sample. No statistically significant excess of events is observed in correlation with the flares. The fluence limit for neutrinos of energy E_ν is calculated according to the equation:

$$\Phi_\nu(E_\nu) = \frac{N_{90}(E_\nu)}{N_e \sigma_{\text{eff}}(E_\nu)}, \tag{10}$$

where N_e is the number of electrons in the Borexino scintillator, that is, $N_e = 9.2 \cdot 10^{31}$ for the whole IV and $N_e = 4.8 \cdot 10^{31}$ for the 145 tons FV. The scattering of monoenergetic neutrinos with energy E_ν off electrons leads to recoil electrons with a Compton-like continuous energy spectrum with maximum energy $T_\nu^{max} = 2E_\nu^2/(m_e + 2E_\nu)$. For each neutrino energy bin, a corresponding recoiled electrons energy interval is computed by considering the cross section and the detector response function.

In order to set the fluence limits for flare-correlated neutrinos (antineutrinos) of electron and $(\mu + \tau)$ flavors individually, the corresponding cross section in Eq. 10 was set to σ_{ν_e} ($\sigma_{\bar{\nu}_e}$) and $\sigma_{\nu_{\mu,\tau}}$ ($\sigma_{\bar{\nu}_{\mu,\tau}}$), respectively. Figure 4 quotes Borexino limits obtained from the primary DAQ ($E_\nu < 3.5$ MeV) and the FADC DAQ ($E_\nu > 3.5$ MeV).

Limits for ν_e quoted by SNO[39] and an allowed band for the neutrino fluence that would have explained the Homestake run 117 excess of events are also shown for comparison.

As of today, Borexino sets the strongest limits on fluences of all neutrino flavors from the solar flares below 3–7 MeV. Under the hyphothesis that neutrino flux is proportional to the flare's intensity, Borexino's data excludes an intense solar flare occurred during run 117 of the Cl-Ar Homestake experiment as a possible source for the observed excess of events.

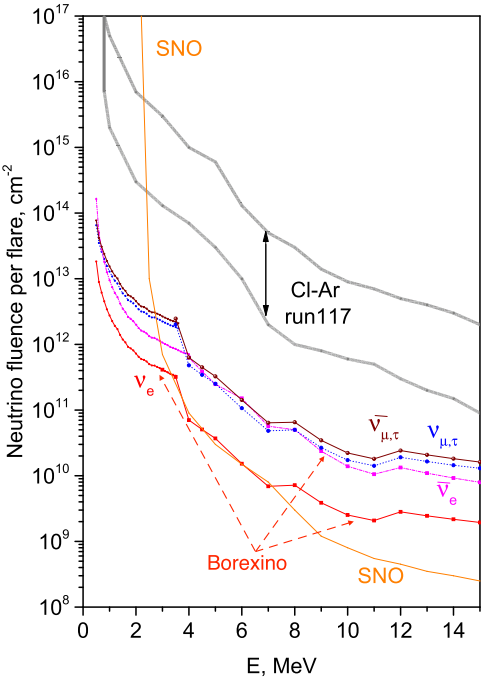


Fig. 4. Borexino 90% C.L. fluence upper limits obtained through elastic scattering for ν_e, $\bar{\nu}_e$, $\nu_{\mu,\tau}$, and $\bar{\nu}_{\mu,\tau}$. The limits by SNO[39] for ν_e are reported and the fluences that would have explained the Cl-Ar Homestake excess in run 117.

5. Conclusions

Large ultrapure liquid scintillators in underground laboratories have a strong potential for the comprehension of our planet energetics and in the various fields of experimental neutrino astronomy. Collecting more data in the incoming years with a network of detectors will be crucial to probe the very foundations of our understanding of the Universe and of our planet.

References

1. G. Bellini et al. (Borexino Collaboration), *J. Instrum.* **6**, P05005 (2011).
2. G. Alimonti et al. (Borexino Collaboration), *Nucl. Instr. and Meth. A* **600**, 58 (2009).
3. M. Agostini et al. (Borexino Collaboration), *Phys. Rev. D* **101**, 012009 (2020).
4. M. Agostini et al. (Borexino Collaboration), *Astropart. Phys.* **125**, 102509 (2021).
5. M. Agostini et al. (Borexino Collaboration), *Astropart. Phys.* **86**, 11 (2017).
6. M. Agostini et al. (Borexino Collaboration), *Ap. J.* **850**, 21 (2017).
7. G. Eder, *Nucl. Phys.* **78** (1966)
8. G. Marx, *Czech J. Phys. B* **19** (1969).
9. R. S. Raghavan et al., *Phys. Rev. Lett.* **80**, 635 (1998).
10. C. G. Rotschild, et al., *Geo. Res. Lett.* **25**, 1083 (1998).

11. G. Fiorentini, M. Lissia, and F. Mantovani, *Phys. Rep.* **453**, 117 (2007).
12. J. H. Davies and D. R. Davies, *Solid Earth* **1**, 5 (2010).
13. T. Araki et al. (KamLAND Collaboration), *Nature* **436**, 499 (2005).
14. S. Abe et al. (KamLAND Collaboration), *Phys. Rev. Lett.* **100**, 221803 (2008).
15. A. Gando et al. (KamLAND Collaboration), *Phys. Rev. D* **88**, 033001 (2013).
16. G. Bellini et al. (Borexino Collaboration), *Phys. Lett. B* **687**, 299 (2010).
17. G. Bellini et al. (Borexino Collaboration), *Phys. Lett. B* **722**, 295 (2013).
18. M. Agostini et al. (Borexino Collaboration), *Phys. Rev. D* **92**, 031101 (2015).
19. Nuclear Power Engineering Section, IAEA-PRIS database, http://www.iaea.org/pris/
20. F. P. An et al. (Daya Bay Collaboration), *Phys. Rev. Lett.* **116**, 061801 (2016).
21. F. P. An et al. (Daya Bay Collaboration), *Phys. Rev. Lett.* **118**, 099902(E) (2017).
22. M. Baldoncini, I. Callegari, G. Fiorentini, F. Mantovani, B. Ricci, V. Strati, and G. Xhixha, *Phys. Rev. D* **91**, 065002 (2015).
23. M. Coltorti et al., *Earth Planet. Sci. Lett.* **293**, 259 (2010).
24. Y. Huang, V. Chubakov, F. Manotavani, R. L. Rudnick, and W. F. McDonough, *Geochemistry, Geophysics, Geosystems* **14**, 2003 (2013).
25. O. Sramek, W. F. McDonough, E. S. Kite, V. Lekic, S. T. Dye, and S. Zhong, *Earth and Planet. Sci. Lett.* **361**, 356 (2013).
26. J. M. Herndon, *Proc. Natl. Acad. Sci. U.S.A. 93(2)*, 646 (1996).
27. J. M. Herndon and D. A. Edgerley, *arXiv:hep-ph/0501216.*
28. S. Ando and K. Sato, *New J. Phys* **6**, 170 (2004).
29. G. J. Feldman, R. D. Cousins, *Phys. Rev. D* **57**, 3873(1998).
30. A. Strumia, F. Vissani, *Phys. Lett. B* **564**, 42 (2003).
31. Y. Gando, et al. (Super-Kamiokande Collaboration), *Phys. Rev. Lett.* **90**, 171302 (2003).
32. H. Zhang, et al. (Super-Kamiokande Collaboration), *Astropart. Phys.* **60**, 41 (2015).
33. A. Gando, et al. (KamLAND Collaboration), *Astroph. J.* **745**, 193 (2012).
34. G. Bellini, et al. (Borexino Collaboration), *Phys. Lett. B* **696**, 191 (2011).
35. G. E. Kocharov, G. A. Kovaltsov, I. G. Usoskin, *Il nuovo cimento* **14 C**, 417 (1991).
36. R. Davis Jr., *Prog. Part. Nucl. Phys.* **32**, 13 (1994).
37. R. Davis Jr., *Nucl. Phys. Proc. Suppl.* **48**, 284 (1996).
38. GOES Database: https://hesperia.gsfc.nasa.gov/goes/goeseventlistings
39. B. Aharmim et al. (SNO Collaboration), *Astropart. Phys.* **551137**, 1 (2014).

Unveiling the engine of the Sun: Measurements of the *pp*-chain solar neutrinos with Borexino

D. Guffanti[1,†], A.C. Re[2], O. Smirnov[3]

[1] *Institute of Physics and Excellence Cluster PRISMA+, Johannes Gutenberg-Universität Mainz, 55099, Mainz, Germany*
daniele.guffanti@uni-mainz.de

[2] *Dipartimento di Fisica "A. Pontremoli", Università degli Studi and INFN Milano, 20133 Milano, Italy*
alessandra.re@mi.infn.it

[3] *Joint Institute for Nuclear Research, 141980 Dubna, Russia*
oleg.smirnov@lngs.infn.it

on behalf of the Borexino Collaboration:
M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, Z. Bagdasarian, D. Basilico, G. Bellini, J. Benziger, R. Biondi, D. Bravo, B. Caccianiga, F. Calaprice, A. Caminata, P. Cavalcante, A. Chepurnov, D. D'Angelo, S. Davini, A. Derbin, A. Di Giacinto, V. Di Marcello, X.F. Ding, A. Di Ludovico, L. Di Noto, I. Drachnev, A. Formozov, D. Franco, C. Galbiati, C. Ghiano, M. Giammarchi, A. Goretti, A.S. Göttel, M. Gromov, Aldo Ianni, Andrea Ianni, A. Jany, D. Jeschke, V. Kobychev, G. Korga, S. Kumaran, M. Laubenstein, E. Litvinovich, P. Lombardi, I. Lomskaya, L. Ludhova, G. Lukyanchenko, L. Lukyanchenko, I. Machulin, J. Martyn, E. Meroni, M. Meyer, L. Miramonti, M. Misiaszek, V. Muratova, B. Neumair, M. Nieslony, R. Nugmanov, L. Oberauer, V. Orekhov, F. Ortica, M. Pallavicini, L. Papp, L. Pelicci, Ö. Penek, L. Pietrofaccia, N. Pilipenko, A. Pocar, G. Raikov, M.T. Ranalli, G. Ranucci, A. Razeto, M. Redchuk, A. Romani, N. Rossi, S. Schönert, D. Semenov, G. Settanta, M. Skorokhvatov, A. Singhal, A. Sotnikov, Y. Suvorov, R. Tartaglia, G. Testera, J. Thurn, E. Unzhakov, F. Villante, A. Vishneva, R.B. Vogelaar, F. von Feilitzsch, M. Wojcik, M. Wurm, S. Zavatarelli, K. Zuber, and G. Zuzel.

About 99% of solar energy is produced through sequences of nuclear processes that convert Hydrogen into Helium in the so-called *pp*-chain. The neutrinos emitted in five of these reactions represent a unique probe of the Sun's internal working and, at the same time, offer an intense natural neutrino beam for fundamental physics research.

The Borexino experiment consists of a large-volume liquid-scintillator detector designed and constructed for real-time detection of low energy solar neutrinos. It is installed at the underground INFN Laboratori Nazionali del Gran Sasso (L'Aquila, Italy) and started taking data in May 2007. Borexino has been the only experiment so far capable of performing a complete study of the *pp*-chain by directly measuring the neutrino-electron

[†] Current address: Dipartimento di Fisica "G. Occhialini", Università degli Studi and INFN Milano Bicocca, 20126 Milano, Italy.

elastic scattering rates for the neutrinos produced in four of its reactions: the initial proton–proton (*pp*) fusion, the electron capture of Beryllium-7, the proton–electron–proton (*pep*) fusion, and the Boron-8 β^+ decay. A limit on the neutrino flux produced in the helium-proton fusion (hep) was also set. This set of measurements further probes the solar fusion mechanism via the direct determination of the relative intensity of the two primary terminations of the *pp*-chain, and the computation of the solar neutrino luminosity. Moreover, the Beryllium-7 and Boron-8 fluxes are indicative of the Sun's core temperature, and their measurement shows a mild preference for the higher temperature expected from the high-metallicity Standard Solar Model scenario. Finally, the experimental survival probability of these solar electron neutrinos allows to simultaneously probe the MSW neutrino flavor conversion paradigm, both in vacuum and in matter-dominated regimes, at different energies.

The details of the strategy adopted by the Borexino collaboration for successfully isolating the spectral components of the *pp*-chain neutrinos signal from residual backgrounds in the total energy spectrum will be presented.

Keywords: Solar neutrinos, proton-proton chain, Borexino

1. Introduction

The study of neutrinos generated by nuclear processes in the Sun has been producing outstanding scientific results for more than 50 years in both the fields of astrophysics and particle physics. The measurements of solar neutrinos provided the first direct evidence of the nuclear origin of stars' luminosity, as well as the first indication of flavour transition in the neutrino sector. In the years, solar neutrino experiments have progressively improved their accuracy, bringing new insights on the functioning of the Sun and on the properties of neutrinos.

The Borexino experiment is a key player in this field, being the most sensitive detector so-far for low-energy solar neutrinos. One of the milestones of the Borexino physics program, which significantly exceeded the expected physics reach, is the (almost) complete investigation of neutrinos produced throughout the entire proton-proton fusion chain, the sequence of nuclear reactions responsible for $\approx 99\%$ of the Sun luminosity. This fundamental result, partially achieved already during the first phase of the experiment (2007-2010), has been further improved in the phase II (2011-2016) and made it possible to test our knowledge of the Sun and to constrain exotic properties of neutrinos.

We present in this contribution the latest results of Borexino regarding solar neutrinos emitted throughout the *pp*-chain and we discuss their implications for both neutrino and solar physics. Section 2 describes the proton-proton fusion chain and the neutrinos produced therein, while the Borexino detector is presented in Sec. 3. The spectrum of solar neutrinos spans across a wide energy range extended up to 19 MeV and their fluxes vary of several orders of magnitude, therefore we setup two separate analyses targeting the low- and high-energy region of the solar neutrino spectrum that are described in Secs. 4 and 5 respectively along with the main results of Borexino.

2. Solar neutrinos from the *pp*-chain

The Sun, as most of other stars, produces energy by fusing four protons into a helium nucleus. This idea, originally hypothesized by Perrin and Eddington,[1,2] was later formalized by Bethe and Critchfield and independently by Weizsäcker who proposed two mechanisms for the hydrogen burning, known as the proton-proton chain[3] (*pp*-chain) and the Carbon-Nitrogen-Oxygen (CNO) cycle.[4–6] The relative efficiency of these two processes depends heavily on the features of the star under consideration, specifically on its core temperature and on its chemical composition. While the CNO-cycle was initially believed to be the dominant energy-production mechanism in the Sun, in light of a more accurate estimate of the Sun core temperature it became clear that the *pp*-chain drives the energy production in the Sun. The relative weight of the two mechanisms is estimated by a detailed model of the Sun, also known as Standard Solar Model (SSM), which -provided an accurate description of all the physical processes driving the evolution of a star- is able to produce a comprehensive snapshop of the Sun by forcing the modelled star to match the current solar mass, radius, luminosity and surface composition. The most recent update of the SSM[7] indicates that only $\approx 1\%$ of the Sun energy is produced through the CNO-cycle, while the real "engine" of the Sun is the *pp*-chain, accounting for $\approx 99\%$ of the Sun luminosity.

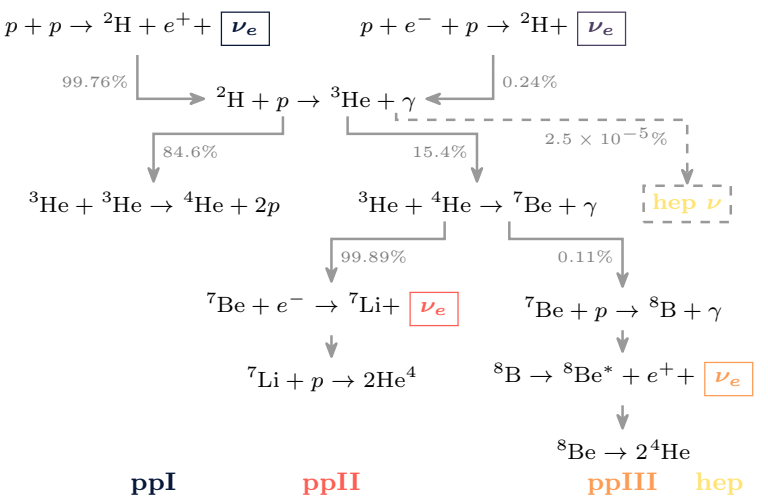


Fig. 1. The sequence of reactions composing the proton-proton fusion chain.

The sequence of reactions composing the *pp*-chain is shown in Fig. 1. Neutrinos are emitted in five of such processes, starting with the fusion of two protons into a deuteron which begins the chain (*pp* neutrinos). An alternative way of producing a deuteron, although $\approx$ 400 times less likely than the proton-proton fusion,

consists in the reaction between two protons and an electron, which releases a 1.44 MeV monoenergetic neutrino (*pep* neutrino). The *pp*-chain can progress into different branches, namely ppI, ppII, ppIII and hep. Depending on the evolution of the chain, different neutrinos can be emitted: while the most prominent branch (ppI) terminates the chain without releasing any additional neutrino, the ppII branch is characterized by monoenergetic neutrinos produced by electron-capture onto ^{7}Be, with energy of 0.86 MeV (branching ratio 89.5%) or 0.38 MeV (B.R. 10.5%) depending if the ^{7}Li nucleus in the final state is produced at its ground state or on the first excited state. The high-energy neutrinos emitted in the decay of ^{8}B (Q-value 15.5 MeV) tag the ppIII branch, and the faintest branch of all, the helium-proton fusion (hep) branch, produces neutrinos up to an energy of 19.8 MeV. The expected solar neutrino spectrum is shown in Fig. 2.

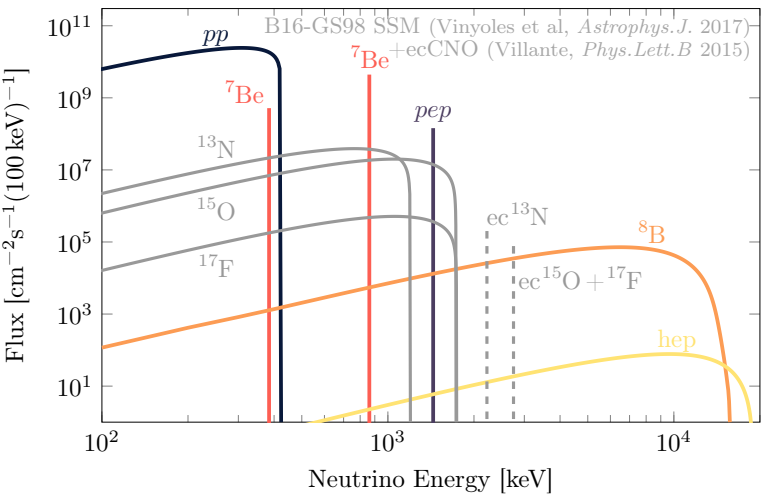


Fig. 2. Expected solar neutrino spectrum, with the neutrinos originating from the *pp*-chain shown with coloured lines and those produced in the CNO-cycle in grey. The fluxes are taken from[7,8] (note that the flux of monoenergetic neutrinos is reported in $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$).

The fluxes of *pp*-chain solar neutrinos depends on the abundance of heavy elements in the Sun, which is currently not fully understood and affects the temperature profile thus impacting on the *pp*-chain nuclear reactions rate. Recent estimates of the photosphere composition[9] (AGSS09met) show a consistent reduction of the metallicity respect to previous studies[10] (GS98), but this new "low-metallicity" (LZ) SSM shows a worst agreement with observation of the sound speed profile as inferred by helioseismology measurements respect to the old "high-metallicity" (HZ) predictions.[7] The precision measurement of solar neutrino fluxes from the *pp*-chain and the CNO-cycle is a promising way to solve the controversy about solar metallicity and the impact of Borexino results in this connection is discussed in Sec. 6.

3. The Borexino experiment

The Borexino experiment can detect solar neutrinos which interact via elastic scattering with the electrons of a liquid scintillator (LS). This detection technique, which provides a low-energy threshold and a good energy resolution, makes it possible to separate the different components of the solar neutrino flux exploiting their specific signature in the energy spectrum, but requires an extremely careful strategy to reduce the indistinguishable radioactive background since the expected interaction rate is ≈ 9 orders of magnitude smaller than the average activity of ordinary water.

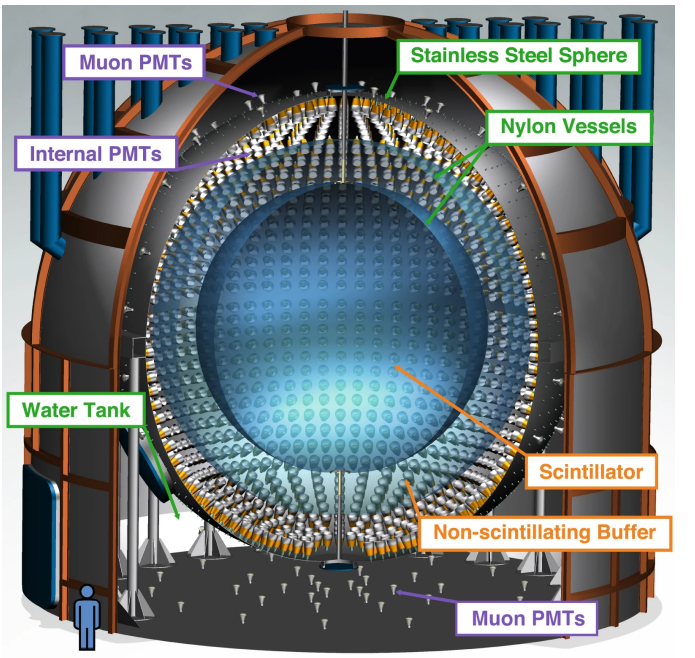


Fig. 3. Schematic view of the Borexino detector. Figure from.[11]

A cutaway of the detector is shown in Fig. 3. The experiment[12] is located in the INFN Gran Sasso National Laboratories, under the Gran Sasso massif in L'Aquila (Italy). The underground laboratories are shielded from cosmic radiation by 1400 m or rock (3800 m.w.e.), which attenuates the cosmic-ray flux by 6 orders of magnitude. Residual muons still reaching the experimental hall[13] ($3.432 \pm 0.003\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$) are tagged using a water Cherenkov detector equipped with 208 PMTs which surrounds the entire inner detector also providing an effective shielding against environmental neutrons and γ-rays. The detector active volume consists of ≈ 300 ton of pseudocumene mixed with 1.5 g/l of (PPO) as a wavelength shifter and is contained into a 125 µm-thick, 4.25 m radius spherical nylon vessel (Inner Vessel, IV). The light

produced by the energy deposition of charged particles is detected by 2212 8 in PMTs (1800 of which are equipped with a light collector cone) mounted onto a 6.85 m radius Stainless Steel Sphere. The space between the IV and the PMTs is filled with quenched LS in order to shield the active volume from the radiation coming from radioactive contamination in the PMTs' glass and in the steel of the support structure.

With the underground location and shielding strategies fighting the external background, the key of Borexino results has been the successful suppression of internal background, *i.e.* those radioactive contamination of the detector itself. This has been achieved thanks to a careful selection of materials and to the implementation of effective purification methods for the LS (primarily distillation and water extraction) which reduced the contamination of ^{238}U and ^{232}Th down to $< 9.5 \times 10^{-20}$ g/g and $< 7.2 \times 10^{-19}$ g/g respectively in the Phase II of the experiment.[14]

The reconstruction of the solar neutrino interaction rate and its separation from the residual background is based primarily (although not exclusively) on the signatures in the visible energy spectrum reconstructed from the scintillation light, therefore the understanding of the detector energy response is fundamental. The energy released in the detector is reconstructed by measuring the isotropic scintillation light produced by recoiling electrons in the LS; the information about the detected light is expressed using three different energy observable that are used in the analysis: the number of hit PMTs in a 230(400) time window ($N_p^{dt_{1(2)}}$), the number of reconstructed hits (N_h) and the total charge measured at the PMTs anode, *i.e.* the number of detected photoelectrons (N_{pe}).

Borexino can rely on two different methods to model the detector response function: the first is based on a full Monte Carlo (MC) simulation of the entire experimental apparatus,[15] which reproduces all the physical processes following the interaction of a particle in the detector (including the electronic chain response) and keeps track of evolution of the detector properties in time, such as the number of operating channels. The simulation has been tuned using both lab measurements and data collected in an extensive calibration campaign,[16] resulting in a sub-percent accuracy for the variables of interest for solar neutrino analysis.[15] An alternative method for modelling the detector energy response consists in an analytical model encapsulating the effect of ionization quenching, the contribution of Cherenkov light as well as the spatial dependence of the reconstructed energy and of the energy resolution.[14] With such model it is possible to reproduce the observed energy spectrum of a given signal by computing the convolution of the visible energy spectrum with the detector response function. More details about this procedure can be found in.[14]

The performance of the detector and the reliability of the response models have been carefully studied during the calibration campaign[16] and are constantly monitored using some known background (such as the ^{14}C, cosmogenic ^{11}C or the characteristic peak of ^{210}Po, see Sec. 4). During the Phase II of Borexino data taking, the energy resolution for 1 MeV deposited energy was close to 50 keV, while the

position of the event –reconstructed via a time-of-flight method– had an accuracy of $\approx 10\,\mathrm{cm}$

4. Low-Energy solar neutrino spectroscopy

The observation of low-energy solar neutrinos has been one of the main goal of the experiment, given their fundamental importance for both neutrino- and solar physics. Differently from the previous generation of low-energy solar neutrino experiments, which were able to measure the integrated neutrino flux over a given energy threshold, the key advantage of Borexino is the possibility to separate the individual contribution of neutrinos produced in different reactions thanks to their specific signature in the visible energy spectrum.

Already during the first phase[17] of the experiment Borexino was able to measure the interaction rate of ^{7}Be, ^{8}B and *pep* neutrinos,[18–20] while imposing an upper limit to the rate of CNO neutrinos.[20] We present in this section the result obtained from the analysis of low-energy neutrinos obtained in the second phase of the Borexino data taking, started in 2011 following a very successful purification campaign which further reduced the already low contamination level in the liquid scintillator. In Sec. 4.1 we discuss the measurement of *pp*-chain low-energy neutrinos reported in,[11,14] while in 4.2 we show the evidence of the time modulation of solar neutrinos interaction rate expected from the Earth's orbit eccentricity. In the second part of this section we report the search of exotic properties of neutrinos performed using low-energy solar neutrinos: in particular Sec. 4.3 presents a limit on the neutrino magnetic moment, while we discuss in Sec. 4.4 the search for neutrino non-standard interactions.

Given their low-flux and different energy scale, neutrinos from the ^{8}B decay require a separate analysis, which is presented in Sec. 5.

4.1. *Simultaneous measurement of low-energy solar neutrino fluxes*

The latest measurements of low-energy solar neutrinos produced by the *pp*-chain performed by Borexino is based on a dataset collected between December 14^{th}, 2011 and May 21^{st}, 2016 (1291.51 days). The increased exposure ($\approx$ 1.6 times larger than Phase I), combined with the lower contamination background and the improved understanding of the detector made it possible to simultaneously extract all the interaction rates of the low-energy components of the *pp*-chain in a single fit over a large portion of the energy spectrum (0.19–2.93 MeV), differently from previous Borexino results where the different solar neutrino species were studied separately optimising the choice of the energy window.

Data recorded by the detector are filtered to remove residual cosmic muons and short-living cosmogenic isotopes by applying a 300 ms veto.[14] In order to remove "external" γ-ray background generated in the material surrounding the liquid

scintillator (especially in the PMTs and in the nylon of the inner vessel) we select only those events reconstructed in a 71.3 tonnes fiducial volume defined in the innermost part of the detector.[14]

The reconstruction of solar neutrinos interaction rates from the energy spectrum is made more challenging by the presence of multiple backgrounds that still contaminates the dataset even after the successful purification of the scintillator and the data selection outlined above: in particular, the sensitivity of Borexino in the low-energy part of the spectrum is limited by the presence of ^{14}C in the LS molecules (β^-, $Q = 0.16\,\mathrm{MeV}$), which decay in the detector at an much higher rate than solar neutrino interactions. Other contamination affecting the energy window are ^{85}Kr (β^-, $Q = 0.69\,\mathrm{MeV}$), ^{210}Po (α, $E_\alpha = 5.3\,\mathrm{MeV}$, light quenched by a factor ≈ 10 in LS) and ^{210}Bi (β^-, $Q = 1.16\,\mathrm{MeV}$). At higher energies, the relevant background is mostly due to residual "external" γ-rays and to cosmogenic ^{11}C (β^+, $Q = 0.96\,\mathrm{MeV}$). The latter can be efficiently tagged by exploiting the fact that it is often produced together with one or more neutrons following muon spallation onto ^{12}C, therefore one can evaluate the likelihood for an event to be a ^{11}C decay on the basis of a set of observables such as the distance and time interval from a crossing muon, the distance of a detected neutron capture and the muon energy deposition profile.[21] Using this selection it has been possible to tag $(92 \pm 4)\,\%$ of ^{11}C events retaining $(64.28 \pm 0.01)\,\%$ of the total exposure, and those events classified as likely to be ^{11}C decays are used to build a separate, ^{11}C-tagged dataset that is analysed jointly with ^{11}C suppressed one.

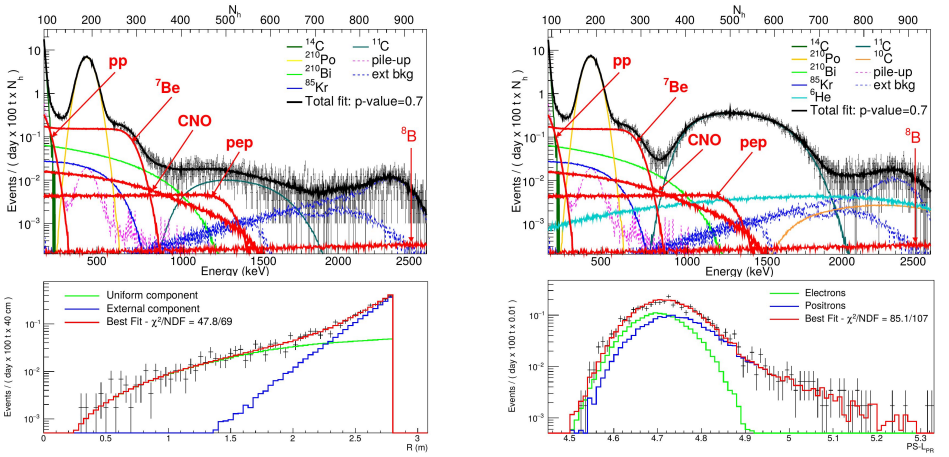


Fig. 4. Result of the simultaneous measurement of low-energy *pp*-chain solar neutrinos. *Top panel*: visible energy spectra in the ^{11}C-subtracted (left) and -tagged datasets. *Bottom left*: Distribution of the events distance from the detector centre. *Bottom right*: Distribution of the pulse shape parameter. All figures from.[14]

To help constraining some of these backgrounds, in some specific cases it is possible to rely on some distinctive features of the signal or on some independent measurements: this is the case of the rate of ^{14}C, which has been measured using a separate dataset to be 40 ± 1 Bq/100 t and that was constrained in the analysis together with the rate of random coincidences of ^{14}C decays. Differently, the signal from external γ-rays shows a clear exponential decrease towards the detector centre due to the shielding effect of LS, while energy release of positrons emitted in the β^+ decay of ^{11}C and its subsequent annihilation results in a slightly different time profile, that we used to define a Pulse Shape (PS) parameter[14] that improves the separation between β^+ and β^-events. In order to take into account these features, we implemented a multivariate analysis which include the events' distance from the detector centre and pulse shape parameter in addition to the energy observables in the region where they can help the analysis.

The analysis procedure has been tested running a detailed sensitivity study[14] based on simulated experiments. This study showed a significant correlation between the reconstructed rates of *pep*, CNO end ^{210}Bi events, which led to a reduced sensitivity for such parameters. For this reason, we decided to constrain in the analysis the rate of CNO neutrinos to the value expected from the prediction of the SSM (both HZ and LZ). The results of the fit[11,14] are reported in Tab. 1 next to the expectations from the SSM,[7] assuming the MSW-LMA oscillation parameters.[22] Similarly, the lower part of Tab. 1 reports the fluxes of solar netrinos inferred from Borexino measurements assuming the MSW-LMA solution. Figure 4 shows the different contributions of solar neutrinos and backgrounds to the energy spectrum as well as to the radial distance and pulse shape distributions.

Systematic uncertainties in the measurement account for possible deviations of the computed exposure, uncertainties in the detector response model as well as specific effects of the analysis methods.[11,14]

The results in Tab. 1 represents the most accurate experimental measurements of low-energy solar neutrinos from the *pp*-chain, improving Borexino previous findings: in particular, the interaction rate of ^{7}Be neutrinos is constrained at the 3 % level, twice as accurate as the expectation of the SSM, while the absence of *pep* neutrinos is disfavoured with a 5σ significance.

4.2. *Time modulations of neutrino signal*

The solar neutrino signal should follow the annual variations of the Sun-to-Earth distance, the effect appears as a 6.7% peak-to-peak seasonal amplitude modulation, with a maximum at the perihelion. The observation of seasonal variation would confirm the solar origin of the signal. Another possible modulation could be induced due to the Earth matter: neutrino are passing through the dense matter in the night and this could cause flavour regeneration with probability dependent on

Table 1. Results of Borexino-Phase II on the *pp*, ^{7}Be and *pep* solar neutrinos interaction rate. The first term of the associated uncertainty represents the statistical error obtained with a likelihood profile under the Wilk's approximation, while the second indicates the systematic uncertainty. The *pep* neutrino rate depends on the different SSM used to constrain the CNO neutrino flux, and therefore two results are quoted. The fluxes reported in the lower part of the table are computed assuming the MSW-LMA oscillation parameters[22]

Solar ν	**Borexino results** Rate (cpd/100 t)	**HZ SSM** Rate (cpd/100 t)	**LZ SSM** Rate (cpd/100 t)
pp	$134 \pm 10^{+6}_{-10}$	131.0 ± 2.4	132.1 ± 2.3
^{7}Be	$48.3 \pm 1.1^{+0.4}_{-0.7}$	47.8 ± 2.9	43.7 ± 2.6
pep (HZ)	$2.43 \pm 0.36^{+0.15}_{-0.22}$	2.74 ± 0.05	2.78 ± 0.05
pep (LZ)	$2.65 \pm 0.36^{+0.15}_{-0.24}$		
	Flux cm^{-2}s^{-1}	Flux cm^{-2}s^{-1}	Flux cm^{-2}s^{-1}
pp	$(6.1 \pm 0.5^{+0.3}_{-0.5}) \times 10^{10}$	$5.98(1 \pm 0.006) \times 10^{10}$	$6.03(1 \pm 0.005) \times 10^{10}$
^{7}Be	$(4.99 \pm 0.13^{+0.07}_{-0.10}) \times 10^9$	$4.93(1 \pm 0.06) \times 10^9$	$4.50(1 \pm 0.06) \times 10^9$
pep (HZ)	$(1.27 \pm 0.19^{+0.08}_{-0.12}) \times 10^8$	$1.44(1 \pm 0.009) \times 10^8$	$1.46(1 \pm 0.009) \times 10^8$
pep (LZ)	$(1.39 \pm 0.19^{+0.08}_{-0.13}) \times 10^8$		

the oscillation parameters. The fact of observation or non-observation of diurnal modulation can be used to restrict the neutrino oscillation parameters.

The seasonal modulation of the ^{7}Be neutrino signal was studied by Borexino collaboration in,[17,23] The earlier work was performed with Borexino Phase-I data acquired from May 2007 to May 2010; the period and phase were found to be consistent with a solar origin of the signal. More advanced analysis of 1456 astronomical days of Phase-II data can be found in more recent paper.[23] The analysis of neutrino signal variations on the many years time-scale is complicated by the presence of time-dependent background components. Three methods have been used to extract the time modulations: the analytical fit to event rate, and two more elaborated techniques: Lomb-Scargle and the Empirical Mode Decomposition. The results obtained with these methods are in excellent agreement. The duration of astronomical year measured with neutrino by using fit to the event rate is 367 ± 10 days with phase shift of $T_0 = 14 \pm 22$ days, i.e. in good agreement with the expected values. Also, the observe eccentricity $\epsilon = 0.0174 \pm 0.0045$ corresponds to the modulation amplitude of $(7.1 \pm 1.9)\%$.

In contrast to the annual signal modulation, no variation of day/night solar neutrino signal at the level of 1% is observed in Borexino data.[24] The measured asymmetry $A_{dn} = 0.001 \pm 0.012(stat) \pm 0.007(syst)$ agrees with the prediction of LMA solution for MSW mechanism. Moreover, this result disfavours MSW oscillations with mixing parameters in the LOW region at more than 8.5σ. The LOW

solution was previously excluded by using the reactor antineutrino data, which involves the assumption of CPT symmetry.

4.3. *Neutrino magnetic moment*

One of the explanation of recent puzzling result of the XENON1T collaboration at few keV electronic recoils could be due to the scattering of solar neutrinos with large Majorana transition magnetic moments[25] (MM). The required transition MM strengths lie within the range that can be probed by current or near future experiments.

Borexino is spectroscopic detector with a relatively good energy resolution at low energies of about 5% at 1 MeV. The spectroscopic feature provides a good possibility to search for the deviations from the standard shape of recoils electrons scattered off monoenergetic ^{7}Be neutrinos. The manifestation of the large neutrino magnetic moment in the electron recoils spectrum would be an additional term in the cross section with a characteristic $1/\mathrm{T}_e$ (T is electron recoil kinetic energy) signature. A sensitive search for the deviations of the electron recoil spectra from the predicted by the Standard Model was performed using Borexino Phase-II data in.[26] A model-independent limit on the effective moment of the solar neutrino of $\mu_{\text{eff}} < 2.810^{-11}\,\mu_B$ at 90% C.L. was obtained. This could be compared to the best direct limit on the magnetic moment of electron antineutrino $\mu_{\text{eff}} < 2.910^{-11}\,\mu_B$ at 90% C.L. obtained using Ge detectors in the most sensitive reactor experiment GEMMA.[27] The Borexino limit is free from uncertainties associated with predictions of the SSM neutrino flux and with the detector's fiducial volume, and is obtained by constraining the sum of the solar neutrino fluxes using the results from radiochemical (gallium) experiments.

Since neutrinos are a mixture of mass eigenstates the effective magnetic moment for neutrino-electron scattering is:

$$\mu^2_{\text{eff}} = \sum_j |\sum_k \mu_{kj} A_k(E_\nu, L)|^2, \tag{1}$$

where μ_{jk} is an element of the neutrino electromagnetic moments matrix and $A_k(E_\nu, L)$ is the amplitude of the k-mass state at the point of scattering. For the Majorana neutrino, only the transition moments are nonzero, while the diagonal elements of the matrix are equal to zero due to CPT-conservation. For the Dirac neutrino, all matrix elements may have nonzero values. The effective magnetic moment can be expanded both in terms of the mass eigenstates or the flavor eigenstates.

In the general case the expression for the effective magnetic moment in the mass eigenstate basis will have a complex form consisting of interference terms $\propto \mu_{jk}\mu_{ik}$. Without significant omissions the solar neutrinos arriving at the Earth can be considered as an incoherent mixture of mass eigenstates. In the case of Dirac

neutrinos assuming that only diagonal magnetic moments μ_{ii} are nonvanishing:

$$\mu^2_{\rm eff} = P^{3\nu}_{e1}\mu^2_{11} + P^{3\nu}_{e2}\mu^2_{22} + P^{3\nu}_{e3}\mu^2_{33} \tag{2}$$

where $P^{3\nu}_{ei} = |A_i(E,L)|^2$ is the probability of observing the i-mass state at the scattering point for an initial electron flavor.

In the case of Majorana transition magnetic moments the effective moment is:

$$\mu^2_{\rm eff} = P^{3\nu}_{e1}(\mu^2_{12} + \mu^2_{13}) + P^{3\nu}_{e2}(\mu^2_{21} + \mu^2_{23}) + P^{3\nu}_{e3}(\mu^2_{31} + \mu^2_{32}) \tag{3}$$

For the well-known approximation of three- to two- neutrino oscillation probabilities for solar neutrinos: $P^{3\nu}_{e1} = \cos^2\theta_{13}P^{2\nu}_{e1}$, $P^{3\nu}_{e2} = \cos^2\theta_{13}P^{2\nu}_{e2}$ and $P^{3\nu}_{e3} = \sin^2\theta_{13}$ – one can get the effective magnetic moment expressed in well-established oscillation parameters in the mass eigenstate basis. Equation (2) can be rewritten as:

$$\mu^2_{\rm eff} = C^2_{13}P^{2\nu}_{e1}\mu^2_{11} + C^2_{13}P^{2\nu}_{e2}\mu^2_{22} + S^2_{13}\mu^2_{33} \tag{4}$$

where $C^2_{13} \equiv \cos^2\theta_{13}$ and $S^2_{13} \equiv \sin^2\theta_{13}$, and $P^{2\nu}_{e1} + P^{2\nu}_{e2} = 1$. Similarly, assuming CPT-conservation ($\mu_{jk} = \mu_{kj}$) relation (3) for the transition moments can be rewritten as:

$$\mu^2_{\rm eff} = C^2_{13}P^{2\nu}_{e1}\mu^2_{12} + (1 - C^2_{13}P^{2\nu}_{e2})\mu^2_{13} + (1 - C^2_{13}P^{2\nu}_{e1})\mu^2_{23} \tag{5}$$

In general, $P^{2\nu}_{e1}$ and $P^{2\nu}_{e2}$ (and $P^{2\nu}_{ee}$) depend on the neutrino energy, but in the energy region below 1 MeV the probabilities can be assumed constant.

Since $\mu^2_{\rm eff}$ is the sum of positively defined quantities, one can constrain any term in (4) and (5). By using the most probable values of $P^{2\nu}_{ee}$, θ_{12} and θ_{13} one can obtain the following limits from the relation $\mu_{\rm eff} \leq 2.8 \times 10^{-11}\mu_B$:

$$|\mu_{11}| \leq 3.4; \quad |\mu_{22}| \leq 5.1; \quad |\mu_{33}| \leq 18.7; \tag{6}$$

$$|\mu_{12}| \leq 2.8; \quad |\mu_{13}| \leq 3.4; \quad |\mu_{23}| \leq 5.0; \tag{7}$$

all measured in units of $10^{-11}\mu_B$ and for 90% C.L.

The effective magnetic moment for the LMA-MSW solution in flavour basis assuming the survival probability of *pp* and ^{7}Be solar neutrinos is the same:

$$\mu^2_{\rm eff} = P^{3\nu}\mu^2_e + (1 - P^{3\nu})(\cos^2\theta_{23}\cdot\mu^2_\mu + \sin^2\theta_{23}\cdot\mu^2_\tau), \tag{8}$$

where $P^{3\nu} = \sin^4\theta_{13} + \cos^4\theta_{13}P^{2\nu}$ is the probability that ν_e is detected in its original flavor (survival probability), with $P^{2\nu}$ calculated in the "standard" 2-neutrino scheme, θ_{13} and θ_{23} are the corresponding mixing angles. Though $P^{2\nu}$ depends on E_ν, the difference between $P^{2\nu}(400) = 0.57$ for a neutrino energy close to the *pp*-neutrino spectrum end point of 420 keV (only a small fraction of the total *pp*-neutrino spectrum close to the end point contributes to the sensitive region in analysis) and $P^{2\nu}(862) = 0.55$ for ^{7}Be-neutrinos (higher energy line) is negligible. An estimate of $P^{2\nu} = 0.55$ was used in calculations.

The limits on the flavor magnetic moment can be obtained from (8) because individual contributions are positive. With $\mu_{\rm eff} < 2.8 \cdot 10^{-11}\mu_B$ and for $\sin^2\theta_{13} = 0.0210 \pm 0.0011$ and $\sin^2\theta_{23} = 0.51 \pm 0.04$ for normal hierarchy (or $\sin^2\theta_{23} = 0.50 \pm 0.04$ for inverted hierarchy) the following limits were obtained: $\mu_e < 3.9 \cdot 10^{-11}\mu_B$, $\mu_\mu < 5.8 \cdot 10^{-11}\mu_B$ and $\mu_\tau < 5.8 \cdot 10^{-11}\mu_B$, all at 90% C.L.

The values above were calculated for the choice of hierarchy providing a more conservative limit.

In another paper[28] the Borexino collaboration performed a search for the hidden antineutrino sources in the data. The presence of antineutrinos in original neutrino fluxes can be a consequence of neutrino electromagnetic interactions induced by the non-zero neutrino magnetic moment. The limit derived for solar antineutrino flux is 384 cm^{-2} s^{-1} (90% C.L.), assuming an undistorted solar ^{8}B neutrinos energy spectrum, corresponding to a neutrino-antineutrino transition probability $P_{\nu\to\overline{\nu}} < 7.2 \times 10^{-5}$ (90% C.L.) for E_ν above inverse beta-decay threshold of 1.8 MeV. Based on the upper limit of the antineutrino admixure in the Solar neutrino flux the following limit on the magnetic moment was obtained:

$$\mu_\nu < 6.0 \cdot 10^{-9} B_\perp[\text{kG}] \cdot \mu_B \ (90\%\ \text{CL}) \tag{9}$$

where $B_\perp$ is transverse component of the strength of the solar magnetic field. The estimates of B in the solar core varies by some orders of magnitude ranging from $B < 600$ G to $B < 7$ MG. These estimations correspond to upper limits of the neutrino magnetic moment between $1.0 \cdot 10^{-8}\mu_B$ and $8.5 \cdot 10^{-13}\mu_B$ (90% C.L.). The latter value is stronger than the current limit from astrophysical observations while the former limit is overlaid by other measurements.

4.4. *Neutrino non-standard interactions*

The observed neutrino contribution in Borexino spectra are determined by the fraction of the electron neutrino in the total neutrino flux at the surface of the Earth (or electron neutrino survival probability P_{ee}(E)), and the neutrino-electron scattering cross section defined by the chiral couplings of the neutrino and electron. Some theories beyond the Standard Model postulate the existence of Non-Standard Interactions (NSI's) which can modify the chiral couplings and P_{ee}. Phenomenologically, NSI's of the neutral current (NC) type are described by the Lagrangian density

$$\mathcal{L}_{\rm NSI} = -2\sqrt{2}G_F\varepsilon_{\alpha\beta}^{ff'C}(\bar{\nu}_\alpha\gamma^\mu\nu_\beta)(\bar{f}\gamma_\mu C f')$$

where $\alpha, \beta = e, \mu, \tau$ label the neutrino flavor, f and f' are leptons or quarks of the same charge but not necessarily the same flavor, C is the chirality of the ff' current (L or R), and $\varepsilon_{\alpha\beta}^{ff'C}$ is a dimensionless coupling constant parametrizing the strength of the NSI interaction normalized to G_F. In the Borexino data analysis

the attention was paid to the flavor-diagonal case $f = f' = e$ and $\alpha = \beta$, so it is convenient to denote $\varepsilon^{fC}_{\alpha\beta}$ as ε^{C}_{α}. Borexino, relying on neutrino-electron elastic scattering, is particularly sensitive to this type.

At detection, ε_α shifts the coupling constants that appear in the expression for the differential cross section for the weak interactions

$$g^C_\alpha \to \tilde{g}^C_\alpha = g^C_\alpha + \varepsilon^C_{\alpha\beta}$$

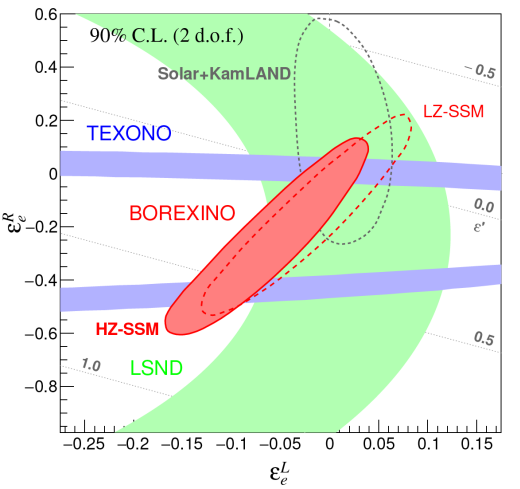


Fig. 5. Allowed region for NSI parameters in $\varepsilon^{L/R}_e$ plane obtained in.[29] The parameters $\varepsilon^{L/R}_\tau$ are fixed at zero. Both HZ- (filled red) and LZ- (dashed red) SSM's were considered. The bounds from the LSND and TEXONO experiments are provided for comparison. Besides, the contour obtained from the global analysis of solar neutrino experiments is presented by dashed black line,[30] NSI's are included in detection and propagation. All contours correspond to 90% C.L. (2 d.o.f.). The dotted gray lines represent the corresponding range of ε' parameter, relevant for NSI's at propagation.

The Borexino Phase II data data were used for the search of the flavor-diagonal neutral current interactions responsible for the ν_{ee} and and $\nu_{\tau e}$ couplings not excluded by other experiments. The solar luminosity constraints were implied to improve the sensitivity to the effects under study, both high- and low-metallicity (HZ and LZ) Standard Solar Model (SSM) predictions of the solar neutrino fluxes were considered. No indications of new physics were found at the level of sensitivity of the detector. The constraints on the parameters of the NSI's were obtained[29] compatible with or better than the existing ones. The results for $\varepsilon^{L/R}_e$ and $\varepsilon^{L/R}_\tau$ parameters are shown in Fig. 5 and Fig. 6 correspondingly.

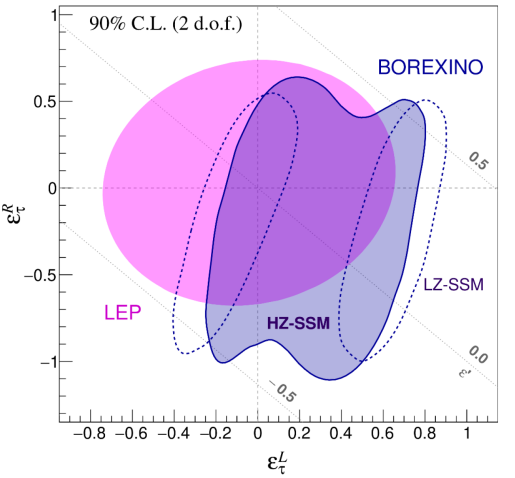


Fig. 6. Allowed region for NSI parameters in $\varepsilon_\tau^{L/R}$ plane obtained in.[29] The $\varepsilon_e^{L/R}$ are fixed at zero. Both HZ- (filled dark blue) and LZ- (dashed dark blue) SSM'sare considered. The contour from LEP experiment is provided for comparison. Both contours correspond to 90% C.L. (2 d.o.f.). The dotted gray lines represent the corresponding range of ε' parameter, relevant for NSI's at propagation.

5. High-energy solar neutrinos in Borexino

As already discussed, Borexino's original goal was the measurement of low energy solar neutrinos and this objective was fully centered. However, through the years, Borexino turned out to be a quite versatile experiment and the collaboration succeeded in measuring also neutrinos belonging to the above-MeV part of the solar neutrino spectrum. In particular, in 2020, Borexino presented a refinement of the measurement of the ^{8}B solar neutrino interaction rate,[31] also setting a new limit on the interaction rate of solar *hep* neutrinos.

The High-Energy Region (HER) work is based on 2062.4 live days of data, collected between January 2008 and December 2016 (detector's purifications and calibrations period excluded). In order to better handling the backgrounds, the analysis is performed with MonteCarlo radial fits, independently performed on two energy subranges: the HER-I (1650, 2950 p.e., roughly 3.2-5 MeV) and the HER-II (2950, 8500 p.e., roughly 5-16 MeV). The HER-I sector includes events originated from natural radioactivity while the HER-II sector is characterized by intense external γ rays following the neutron capture processes that happens on the Stainless Steel Sphere. In Borexino, the energy depositions of natural radioactivity never exceeds 5 MeV (Q-value of the ^{208}Tl β decay) since the α scintillation signals are strongly quenched and end below this analysis threshold. As of 2020, the ^{8}B neutrinos are detected via their elastic scattering off electrons on a fiducial mass of 300 t of liquid scintillator (the entire Borexino active mass) resulting in a total exposure of

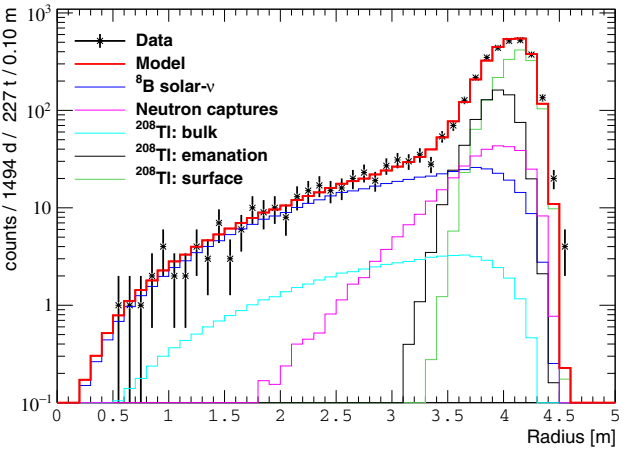


Fig. 7. Fit of the event radial distribution in the HER-I range[31] [1650,2950]p.e.

1.5 ktonne x year: an impressive 11.5-fold increase with respect to the previous Borexino ^{8}B measurement.[19]

In the HER analysis, the selection of data is performed via software cuts that tag/remove the primary/secondary backgrounds associated with the interaction of cosmic muons in the detector. The identification of the ^{8}B neutrino signal is then based on its different radial distribution with respect to that of background: neutrino events are expected to be uniformly distributed in the active volume while the remaining background event distribution is strongly radial dependent (Fig. 7).

The radial fit estimator is a binned likelihood and the χ^2/dof of the radial fits are 31.3/36 (HER-I), and 30.4/35 (HER-II) excluding empty bins.
The best-fit rates for the ^{8}B solar neutrinos are found to be:

$$R_{\mathrm{HER-I}} = 0.136\,^{+0.013}_{-0.013}\,(\mathrm{stat})\,^{+0.003}_{-0.003}\,(\mathrm{syst})\ \ \mathrm{cpd}/\,100\ \mathrm{t}$$
$$R_{\mathrm{HER-II}} = 0.087\,^{+0.080}_{-0.010}\,(\mathrm{stat})\,^{+0.005}_{-0.005}\,(\mathrm{syst})\ \ \mathrm{cpd}/\,100\ \mathrm{t}$$
$$R_{\mathrm{HER}} = 0.223\,^{+0.015}_{-0.016}\,(\mathrm{stat})\,^{+0.006}_{-0.006}\,(\mathrm{syst})\ \ \mathrm{cpd}/\,100\ \mathrm{t}$$

The precision of the HER ^{8}B rate analysis is then about 8% and it improves more than a factor 2 our previous result.[19] The equivalent ^{8}B neutrino flux inferred from the HER analysis[31] is $\Phi(^{8}\mathrm{B}) = 2.57\,^{+0.17}_{-0.18}\,(\mathrm{stat})\,^{+0.07}_{-0.07}\,(\mathrm{syst}) \times 10^{6}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ well in agreement with the previous Borexino result[19] and with the high-precision measurement performed by Super Kamiokande.[32] The B16 HZ-SSM[7] expected ^{8}B neutrino flux is $5.46 \pm 0.66 \times 10^{6}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ but the apparently missing flux is fully compatible the MSW-LMA neutrino oscillation scenario.

The Borexino experiment is not optimized for a clear detection of neutrino fluxes as low as those predicted by the SSM for the *hep* neutrinos ($\sim 10^{3}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$). Nevertheless, this elusive type of neutrinos is searched via their elastic scattering

on electrons as well as via their neutral current-mediated inelastic scattering on Carbon nuclei. The Borexino collaboration was able to set a limit on the number of expected *hep* events of 4.37 (90% C.L.), corresponding to a limit on the flux of $< 1.8 \times 10^5$ cm^{-2} s^{-1} which is about 1.2 times stronger than the one we previously reported in 2018.[11]

6. Conclusion and discussion

Solar neutrinos are an essential tool to understand stellar physics, being able at the same time to reveal some of their fundamental properties. The Borexino experiment, with its 14-years-long data taking, has been one of the key players in the investigation of these particles. Thanks to its outstanding radiopurity and to the deep understanding of the detector behaviour, Borexino has been able to overcome the intrinsic limitations of the previous generations of radiochemical experiments and to provide a precision measurement of all the components of the solar neutrino spectrum with the only exception of the very low flux of neutrinos from the Helium-proton fusion for which an upper limit has been set. In particular, in this contribution we report the Borexino results on neutrinos produced in the proton-proton fusion chain, which is by far the main energy production mechanism in our star. With a separate analysis of the low- and high-energy region of the energy spectrum, the individual fluxes of *pp*, *pep*, ^{7}Be and ^{8}B neutrinos were directly measured for the first time by a single experiment.

This result has significant implications for solar and particle physics: by measuring the flux of *pp* neutrinos with a 11% accuracy and the ^{7}Be flux with an unprecedented 3% precision, we could compute the luminosity of the Sun using neutrinos,[11] finding an excellent agreement with the photon-inferred result ($L_\odot^{\rm ph} = (3.846 \pm 0.015) \times 10^{33}\,{\rm erg\,s^{-1}}$ vs $L_\odot^{\nu} = (3.89^{+0.35}_{-0.42}) \times 10^{33}\,{\rm erg\,s^{-1}}$), thus confirming the nuclear origin of the solar power and the thermodynamical equilibrium of the Sun over the timescale required for radiation to flow from the core to the surface of the Sun, which is in the order of 10^5 years. Using the measurements of *pp* and ^{7}Be neutrino fluxes we could also study the relative weight of the two main terminations of the *pp*-chain (ppI and ppII in Fig. 1), which is the ratio $R_{\rm I/II}$ between the ^{3}He$-^4$He and ^{3}He$-^3$He fusion rates. Neglecting the small contribution of the ppIII branch and assuming the local kinetic equilibrium of ^{2}H and ^{3}He, we find $R_{\rm I/II} = 0.178^{+0.027}_{-0.023}$, in good agreement with the predictions of both the HZ and LZ SSM[11] ($R^{\rm HZ}_{\rm I/II} = 0.180 \pm 0.011$, $R^{\rm LZ}_{\rm I/II} = 0.161 \pm 0.010$). Furthermore, we could exploit the strong dependence of ^{7}Be and ^{8}B fluxes upon the Sun core temperature to test the predictions of the SSM implementing a different chemical composition. Changes in the Sun metallicity will indeed lead to different expectations of the temperature profile, affecting the expected neutrino fluxes. Borexino *pp*-chain results are compatible with the predictions of both HZ and LZ SSM,[11] but an hypothesis test show a mild preference for the higher core temperature expected from the HZ SSM.[11]

In its long scientific life, Borexino exceeded the expectations of its physics reach and is now among the leading solar neutrino experiments. This extraordinary journey recently culminated with the first observation of solar neutrinos from the CNO-cycle,[33] adding a new fundamental piece to our understanding of the Sun and of all stars.

Acknowledgements

The Borexino collaboration acknowledges the hospitality and support of the Laboratori Nazionali del Gran Sasso (Italy). The Borexino program is made possible by funding from Istituto Nazionale di Fisica Nucleare (INFN) (Italy), National Science Foundation (NSF) (USA), Deutsche Forschungsgemeinschaft (DFG) and Helmholtz-Gemeinschaft (HGF) (Germany), Russian Science Foundation (RSF) (grant number 21-12-00063), and Narodowe Centrum Nauki (NCN) (grant number UMO 2017/26/M/ST2/00915) (Poland).

References

1. A. Eddington, The Internal Constitution of the Stars, *Nature* **106** (1920).
2. A. Eddington, The internal constitution of the stars, *Observatory* **43** (1920).
3. H. Bethe and C. Critchfield, The Formation of Deuterons by Proton Combination, *Phys. Rev.* **54** (1938).
4. H. Bethe, Energy Production in Stars, *Phys. Rev.* **55** (1939).
5. C. von Weizsäcker, Über Elementumwandlungen in Innern der Sterne I, *Phys. Z.* (1937).
6. C. von Weizsäcker, Über Elementumwandlungen in Innern der Sterne II, *Phys. Z.* (1938).
7. N. Vinyoles, A. M. Serenelli, F. L. Villante, S. Basu, J. Bergström, M. C. Gonzalez-Garcia, M. Maltoni, C. Peña Garay and N. Song, A new Generation of Standard Solar Models, *Astrophys. J.* **835**, p. 202 (2017).
8. F. L. Villante, ecCNO Solar Neutrinos: A Challenge for Gigantic Ultra-Pure Liquid Scintillator Detectors, *Phys. Lett. B* **742**, 279 (2015).
9. M. Asplund, A. Grevesse, N.and Sauval and P. Scott, The chemical composition of the sun, *Annu. Rev. Astron. Astrophys.* **47** (2009).
10. N. Grevesse and A. Sauval, Standard solar composition, *Space Sci. Rev.* **85** (1998).
11. M. Agostini *et al.*, Comprehensive measurement of *pp*-chain solar neutrinos, *Nature* **562**, 505 (2018).
12. G. Alimonti *et al.*, The Borexino detector at the Laboratori Nazionali del Gran Sasso, *Nucl. Instrum. Meth. A* **600**, 568 (2009).
13. M. Agostini *et al.*, Modulations of the Cosmic Muon Signal in Ten Years of Borexino Data, *JCAP* **02**, p. 046 (2019).
14. M. Agostini *et al.*, First Simultaneous Precision Spectroscopy of *pp*, ^{7}Be, and *pep* Solar Neutrinos with Borexino Phase-II, *Phys. Rev. D* **100**, p. 082004 (2019).
15. M. Agostini *et al.*, The Monte Carlo simulation of the Borexino detector, *Astropart. Phys.* **97**, 136 (2018).
16. H. Back *et al.*, Borexino calibrations: Hardware, Methods, and Results, *JINST* **7**, p. P10018 (2012).

17. G. Bellini *et al.*, Final results of Borexino Phase-I on low-energy solar neutrino spectroscopy, *Phys. Rev. D* **89**, p. 112007 (Jun 2014).
18. C. Arpesella *et al.*, Direct Measurement of the Be-7 Solar Neutrino Flux with 192 Days of Borexino Data, *Phys. Rev. Lett.* **101**, p. 091302 (2008).
19. G. Bellini *et al.*, Measurement of the solar 8B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector, *Phys. Rev. D* **82**, p. 033006 (2010).
20. G. Bellini *et al.*, First evidence of *pep* solar neutrinos by direct detection in Borexino, *Phys. Rev. Lett.* **108**, p. 051302 (2012).
21. M. Agostini *et al.*, Identification of the cosmogenic ^{11}C background in large volumes of liquid scintillators with Borexino (6 2021).
22. I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler and T. Schwetz, Updated fit to three neutrino mixing: Exploring the accelerator-reactor complementarity, *JHEP* **01**, p. 087 (2017).
23. M. Agostini *et al.*, Seasonal modulation of the ^{7}Be solar neutrino rate in Borexino, *Astropart. Phys.* **92**, 21 (2017).
24. G. Bellini *et al.*, Absence of a day–night asymmetry in the ^{7}Be solar neutrino rate in Borexino, *Phys. Lett. B* **707**, 22 (2012).
25. E. Aprile *et al.*, Excess electronic recoil events in XENON1T, *Phys. Rev. D* **102**, p. 072004 (Oct 2020).
26. M. Agostini *et al.*, Limiting neutrino magnetic moments with Borexino Phase-II solar neutrino data, *Phys. Rev. D* **96**, p. 091103 (Nov 2017).
27. A. G. Beda *et al.*, Gemma experiment: The results of neutrino magnetic moment search, *Physics of Particles and Nuclei Letters* **10**, 139 (2013).
28. M. Agostini *et al.*, Search for low-energy neutrinos from astrophysical sources with borexino, *Astroparticle Physics* **125**, p. 102509 (2021).
29. M. Agostini *et al.*, Constraints on flavor-diagonal non-standard neutrino interactions from Borexino Phase-II, *J. High Energ. Phys.* **38** (2020).
30. A. Bolaños, O. G. Miranda, A. Palazzo, M. A. Tórtola and J. W. F. Valle, Probing nonstandard neutrino-electron interactions with solar and reactor neutrinos, *Phys. Rev. D* **79**, p. 113012 (Jun 2009).
31. M. Agostini *et al.*, Improved measurement of ^{8}B solar neutrinos with $1.5 kty$ of Borexino exposure, *Phys. Rev. D* **101**, p. 062001 (2020).
32. K. Abe *et al.*, Solar neutrino measurements in Super-Kamiokande-IV, *Phys. Rev. D* **94**, p. 052010 (Sep 2016).
33. M. Agostini *et al.*, Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun, *Nature* **587**, 577 (2020).

Electron neutrino survival probability in the energy range 200 keV–15 MeV

Marco Pallavicini
Dipartimento di Fisica - Università degli Studi di Genova
and
INFN - Istituto Nazionale di Fisica Nucleare, Sezione di Genova
Via Dodecaneso 33, 16146 Genova - Italy
E-mail: marco.pallavicini@unige.it

on behalf of the Borexino Collaboration:
M. Agostini K. Altenmüller S. Appel V. Atroshchenko Z. Bagdasarian D. Basilico G. Bellini J. Benziger R. Biondi D. Bravo B. Caccianiga F. Calaprice A. Caminata P. Cavalcante A. Chepurnov D. D'Angelo S. Davini A. Derbin A. Di Giacinto V. Di Marcello X.F. Ding A. Di Ludovico L. Di Noto I. Drachnev A. Formozov D. Franco C. Galbiati C. Ghiano M. Giammarchi A. Goretti A.S. Göttel M. Gromov D. Guffanti Aldo Ianni Andrea Ianni A. Jany D. Jeschke V. Kobychev G. Korga S. Kumaran M. Laubenstein E. Litvinovich P. Lombardi I. Lomskaya L. Ludhova G. Lukyanchenko L. Lukyanchenko I. Machulin J. Martyn E. Meroni M. Meyer L. Miramonti M. Misiaszek V. Muratova B. Neumair M. Nieslony R. Nugmanov L. Oberauer V. Orekhov F. Ortica M. Pallavicini L. Papp L. Pelicci Ö. Penek L. Pietrofaccia N. Pilipenko A. Pocar G. Raikov M.T. Ranalli G. Ranucci A. Razeto A. Re M. Redchuk A. Romani N. Rossi S. Schönert D. Semenov G. Settanta M. Skorokhvatov A. Singhal O. Smirnov A. Sotnikov Y. Suvorov R. Tartaglia G. Testera J. Thurn E. Unzhakov F. Villante A. Vishneva R.B. Vogelaar F. von Feilitzsch M. Wojcik M. Wurm S. Zavatarelli K. Zuber G. Zuzel

The Borexino experiment, located at the Laboratori Nazionali del Gran Sasso in Italy, has been the first and so far unique experiment capable to measure the interaction rate of all solar neutrino components produced by the Sun through the so called *pp*-chain and CNO-cycle fusion mechanisms. Particularly, Borexino has measured the rate of *pp*, ^{7}Be, *pep* and ^{8}B neutrinos, which span a wide energy range from a few hundreds keV up to almost 15 MeV. This capability offered Borexino the unique opportunity to experimentally test the expected electron neutrino survival probability predicted by the theory. The paper briefly summarises this important achievement and discusses possible future developments.

Keywords: Solar neutrinos; electron neutrino survival probability; neutrino oscillations; LMA-MSW effect

1. Introduction

The Sun is a powerful source of low energy neutrinos that are a unique tool to test the solar theory and to probe neutrino physics. A remarkable set of discoveries have been made by means of solar neutrinos, among which the most important is the discovery electron neutrino oscillations into muon and tau neutrinos made by the SNO[1,2] experiment, which led to the Nobel prize of Art McDonald in 2015.[3]

It is quite well known that the stellar theory has long predicted two mechanisms for hydrogen fusion into helium in the Sun, known as the *pp*-chain and the

CNO cycle.[4] Both mechanism are based on a sequence of nuclear reactions, some of which produce electron neutrinos. The *pp*-chain is initiated by the fusion of two protons and produce neutrinos through four detectable mechanisms[a]. The CNO cycle bears this name because carbon, nitrogen and oxygen act as catalysers and produce neutrinos as well.[4] Fig. 1 shows the electron neutrino spectrum predicted by[4] without any neutrino oscillation effect.

The two fusion reactions occur in the Sun's core, and because of the relatively low solar temperature, they contribute very unequally to energy production: about 99% of the solar luminosity is due to the *pp*-chain, while the rest is due to the CNO cycle. The CNO cycle is dominant for more massive stars.

Borexino is the first experiment that was able to measure directly all individual solar neutrino components emitted by the *pp*-chain reactions[5] and by the CNO cycle in the Sun.[6] This fact is by itself a remarkable experimental achievement, which was obtained thanks to almost 30 years of activity.[7] It has also the consequence that the

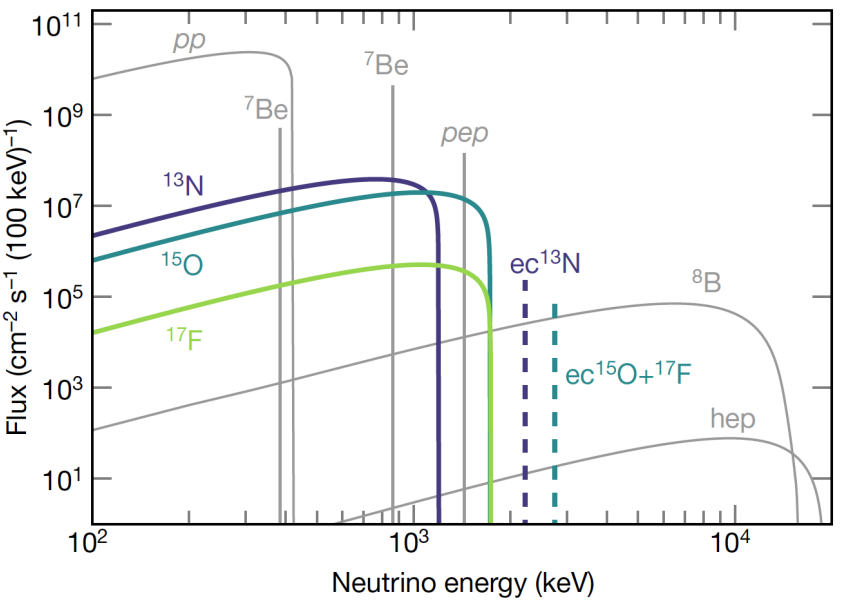


Fig. 1. The spectrum of solar neutrinos as predicted by Vinyoles et al. Hep neutrinos, while at high energy, have a flux that is too low to be detected by current generation experiments. All other components have been measured directly by Borexino through spectral analysis, exploiting the very low radioactive background, the low energy threshold and the good energy resolution of the liquid scintillator detector. Borexino has been so far the only experiment being able to perform such a comprehensive spectral measurement.

[a] A fifth mechanism due to the fusion of ^{3}He with a proton (hep neutrinos) is too small to be detected.

experiment was able to measure the survival probability of neutrinos as a function of the energy and to test directly the LMA-MSW[8] theory of neutrino oscillations. LMA refers to Large Mixing Angle solution, while MSW to the Mikheyev-Smirnov-Wolfenstein mechanism.[9–11]

According to MSW theory, at energy above a few MeV the different refractive index affecting the propagation of electron neutrinos with respect to that affecting the muon and tau neutrinos propagation introduce an additional phase shift that can change the survival probability of electron neutrinos from the central core to the Earth. At some energies the effect can be very large because of a resonance.[10,11] As a consequence, the LMA-MSW theory predicts that solar neutrino oscillations are strongly energy dependent and particularly are dominated by the matter effects above above a few MeV. Vacuum oscillations are dominant at lower energies.

When the LMA-MSW theory is taken into account, the neutrino fluxes predicted by the Standard Solar Model[4] are translated into a well defined prediction of the neutrino interaction rate in Borexino, which depends both on the total electron neutrino flux and on the fraction of electron neutrinos and other neutrino types reaching the Earth.

In this paper we show how the precision measurement pp, ^{7}Be, pep and ^{8}B neutrino rate made by Borexino translates into the first direct measurement of the electron neutrino survival probability as a function of the neutrino energy.

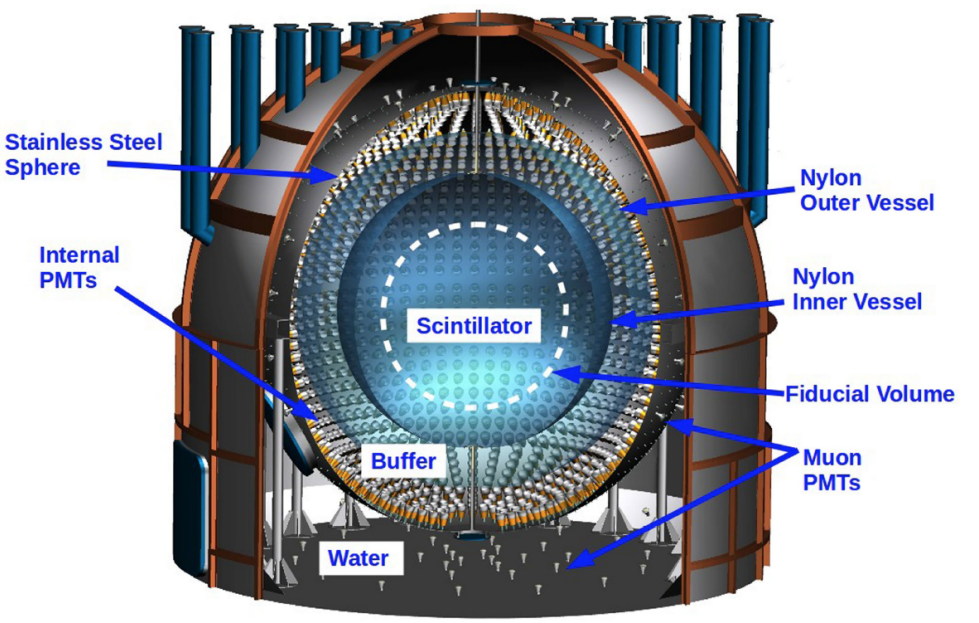


Fig. 2. Drawing of the Borexino detector. The various components are labelled in the picture. To give a feeling of the size, we quote the diameter of the Stainless Steel Sphere, which is 13.7 m.

2. The Borexino experimental setup

Borexino is a very pure liquid scintillator detector[12–15] designed for the real–time observation of low energy solar neutrinos. It is located in Italy deep underground ($\simeq$3800 meters of water equivalent, m w.e.) in the Hall C of the Laboratori Nazionali del Gran Sasso, where the muon flux is suppressed by a factor of $\approx 10^6$ with respect to sea level.

The detector is schematically depicted in Fig. 2. The inner part is an unsegmented stainless steel sphere (SSS) that is both the container of the scintillator and the mechanical support of the photomultipliers. Within this sphere, two nylon vessels separate the scintillator volume in three shells of radii 4.25 m, 5.50 m and 6.85 m, the latter being the radius of the SSS itself. The inner nylon vessel (IV) contains the liquid scintillator solution, namely PC (pseudocumene, 1,2,4-trimethylbenzene $C_6H_3(CH_3)_3$) as a solvent and the fluor PPO (2,5-diphenyloxazole, $C_{15}H_{11}NO$) as a solute at a concentration of approximately 1.5 g/l (0.17 % by weight). The second and the third shell contain PC with a small amount (few g/l) of DMP (dimethylphthalate, $C_6H_4(COOCH_3)_2$) that is added as a light quencher in order to further reduce the scintillation yield of pure PC.[12]

The Inner Vessel is made of 125 μm thick Nylon-6 carefully selected and handled in order to achieve maximum radio–purity. Since the PC/PPO solution is slightly lighter (about 0.4 %) than the PC/DMP solution, the Inner Vessel is anchored to the bottom (south pole of the SSS) with a set of nylon strings. The outer nylon vessel (OV) has a diameter of 11 m and is built with the same material as the inner one. The OV is a barrier that prevents ^{222}Rn emanated from the external materials (steel, glass, photomultiplier materials) to di?use into the fiducial volume.

The buffer fluid between the Inner Nylon Vessel and the SSS (PC/DMP solution) is the last shielding against external backgrounds. The use of PC as a buffer is convenient because it matches both the density and the refractive index of the scintillator, thus reducing the buoyancy force for the nylon vessel and avoiding optics aberrations that would spoil the spatial resolution.

The addition of the DMP quenches the scintillation yield of the buffer fluid by a factor of 20. This is important in order to avoid the unacceptable trigger rate due to the radioactivity of the photomultipliers.

The scintillation light is collected by nominally 2212 photomutipliers (PMTs) that are uniformly attached to the inner surface of the SSS. All but 384 photomultipliers are equipped with light concentrators that are designed to reject photons not coming from the active scintillator volume, thus reducing the background due to radioactive decays originating in the buffer liquid or γ's from the PMTs. The PMTs without concentrators are used to measure this background, and to help identify muons that cross the buffer, but not the Inner Vessel.

The number of active PMTs has changed over time during the almost 15 years of data taking, so that collection efficiency is a function of time and run–dependent.

The SSS is supported by 20 steel legs and enclosed within a large tank that is filled with ultra-pure water. The tank has a cylindrical base with a diameter of 18 m and a hemispherical top with a maximum height of 16.9 m. The Water Tank (WT) is a powerful shielding against external background (γ rays and neutrons from the rock) and is also used as a Cherenkov muon counter and muon tracker. The muon flux, although reduced by a factor of 10^6 by the 3800 m w.e. depth of the Gran Sasso Laboratory, is of the order of 1 m^{-2} h^{-1}, corresponding to about 4000 muons per day crossing the detector. This flux is well above Borexino requirements and a strong additional reduction factor (about 10^4) is necessary and was indeed achieved[13] by the combined use of WT signal and inner detector signal. More details on the detector can be found in.[12,13]

3. Neutrino detection

Low energy neutrinos of all flavours are detected by means of their elas- tic scattering off electrons or, in the case of electron anti–neutrinos, by means of their inverse beta decay on protons or carbon nuclei. The electron (positron) recoil energy is converted into scintillation light, which is then collected by a set of pho- tomultipliers. Neutron detection is often possible as well, through its capture on proton and emission of energetic gamma. A tiny amount of Cherenkov light is also produced, though it is mostly absorbed and re–emitted.

Scintillation light offers several advantages with respect to both the water Cherenkov detectors and the older radiochemical detectors used before Borexino in solar neutrino experiments. Water Cherenkov detectors, in fact, can not effectively detect solar neutrinos whose energy is below approximately 3.5 MeV, both because the Cherenkov light yield is low and because the intrinsic radioactive background cannot be pushed down to suffciently low levels. On the other hand, radiochemical experiments cannot intrinsically perform spectral measurements and do not detect events in real time. Cherenkov detector, however, can effciently detect the direction of the scattered electron and therefore of the incoming neutrino, which is not possible yet in liquid scintillators.

A liquid scintillator, on the other hand, solves these problems: the detection of low energy neutrino becomes possible possible because the light yield is high, offering an energy threshold as low as a few tens ofkeV[b]; the organic nature of the scintillator, and its liquid form at ambient temperature, provide very low solubility of ions and metal impurities, and yield the technical possibility to purify the material as required. However, no measurement of the direction of the incoming neutrino is possible[c] and, even more importantly, the neutrino induced events are intrinsically

[b]However, the unavoidable contamination of ^{14}C that is present in any organic liquid practically limits the "neutrino window" above ?200 keV

[c]Very recently Borexino has shown that some of this prompt Cherenkov light can be detected and identified. Future detector might be able to make useful use of this additional information.

indistinguishable from β and γ radioactivity, posing formidable requirements in terms of radio–purity of the scintillator and of the detector materials.

The total number of detected photons and their arrival times are used to reconstruct the electron recoil energy and the interaction point in the detector, respectively. The energy and spatial resolution of Borexino is not constant because it has slowly deteriorated over time owing to the steady loss of photomultiplier tubes (on average 1238 channels were active during the recent CNO analysis, with larger numbers available at the time of *pp*-chain analysis), with current values of $\sigma_E/E \approx 6\%$ and $\sigma_{x;y;z} \approx 11$ cm for 1 MeV events at the centre of the detector. The time profile of the scintillation light provides also a powerful way to distinguish between different particle types α, β and β^+ via pulse-shape discrimination methods.[16,17]

The final radio–purity levels achieved by Borexino on ^{238}U and ^{232}Th chains, and on several other radioactive contaminants, are many orders of magnitude lower than those either set as design goals or achieved by any other low–background experiment. In this short paper it is simply not possible to give an account of the amount of work done, nor a detailed description of the final numbers obtained. We invite the interested reader to look at the bibliography and particularly to[16,17] and references therein.

The total cross section of electron neutrinos on electrons depends on both charged and neutral currents weak interactions, while that of muon and tau neutrinos is induced by neutral currents only. The interaction rate depends therefore on the neutrino type at the target, a fact that makes the experiment sensitive to neutrino oscillations occurring along the path from the production site in the Sun's up to the detector. This is the key element that allows the direct measurement of the energy dependence of the flavour conversion probability through the Sun induced by the LMA-MSW effect.

The extraction of the survival probability can be obtained from the measured rate using the total fluxes predicted by the Standard Solar Model through the formula:

$$R_x(E) = \Phi_x \left[P_{ee}(E)\sigma_e + (1 - P_{ee}(E))\,\sigma_\mu\right] \tag{1}$$

where R_x is the measured rate of the x neutrino component with average energy E (x can be *pp*, ^{7}Be, *pep*, ^{8}B), Φ_x is the predicted electron neutrino flux of the x component at the Sun's core, P_{ee} is the electron neutrino survival probability, and σ_e and σ_μ are the neutrino-electron total cross sections for electron and muon neutrinos respectively. It is clear from the equation above that for any given measured rate the corresponding value of P_{ee} can be directly measured using as inputs the very well known standard model cross sections (which introduce a negligible uncertainty) and the solar model predictions, which drive most of the uncertainty beside experimental errors.

At the time of the proposal, the main goal of Borexino was the precise measurement of the rate induced by the monochromatic electron neutrinos (0.862 keV)

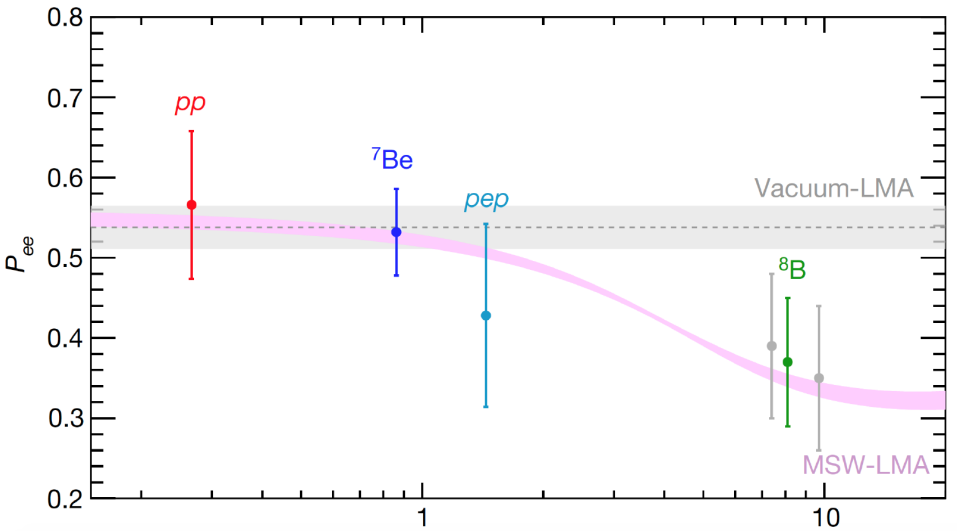


Fig. 3. The measured survival probability for the four neutrino components of the *pp*-chain. The pink band is the theoretical prediction of the P_{ee} with its uncertainty. The error on each value of P_{ee} depends on both theoretical errors and experimental error. See text for more details.

produced by the electron capture decay of ^{7}Be in the Sun. However, the very high radio purity of the scintillator offered new unexpected results, such as a clear evidence of the *pep* solar neutrinos,[18] a low energy threshold (3 MeV) detection of ^{8}B neutrinos,[19] the first direct measurement of *pp* neutrinos,[20] an unambiguous detection of geo-neutrinos,[21–23] and, last but definitely not least, the first prove of the existence of the CNO cycle in the Sun.[6]

After a careful detector calibration with internal sources[15] and an extensive purification campaign performed in 2010-2011, the complete *pp*-chain has been studied. This chain includes four reactions that emit four electron neutrino components of quite different energies: the *pp* neutrinos, whose average energy is around 200 keV, the ^{7}Be neutrinos, which offer a monochromatic line at 0.862 keV, the *pep* neutrinos, which offer another monochromatic line at 1.44 MeV, and ^{8}B neutrinos, which have a broad β^+ energy spectrum reaching about 15 MeV. These four neutrino components have very different survival probabilities in the Sun matter because of MSW effect, as we will see in the next section.

Although not relevant for LMA-MSW measurement, it is worth recalling that Borexino, after a substantial effort devoted to the thermal stabilisation of the detector developed in 2016-2018, has been able to measure the small and diffcult CNO neutrino signal.

In next section we focus on the complete measurement of the *pp*-chain neutrinos and on the consequence of these measurements to the knowledge of LMA-MSW effect in the Sun.

4. Electron neutrino survival probability measurement

The neutrinos produced in the *pp*-chain fusion reactions are all electron neutrinos because they originate from reactions with either an electron in the initial state (^{7}Be and *pep*) or with a positron in the final state (^{8}B and *pp*). However, neutrinos may undergo flavour transitions through the propagation inside solar matter, in the vacuum between the Sun and the Earth and, in the case of ^{8}B neutrinos, also while propagating at night through the Earth itself.

Without matter effects the conversion probability would be just a constant that depends on the vacuum mixing angle θ_{12}, with very little dependence on the other mixing angles and no dependence at all on mass splittings, because for solar neutrinos quantum coherence is completely lost. Indeed, coherence is washed out by the fact that the core of the Sun (diameter approximately 200.000 km) is comparable to the wavelength induced by mass-splitting at solar energies $(2\pi\ E/(1.27 \cdot \delta m^2_{12}) \simeq (2\pi\ 1\ MeV)/(1.27 \cdot 7.\ 10^{-5}\ eV^2) \simeq 70.000$ km) and also because the relative high temperature inside the Sun induce full decoherence of the neutrino beam ($k_B T \gg \delta m_{12}$).

However, since the late '80s and early '90s it was noticed that neutrino flavour oscillations may explain the solar neutrino problem only by assuming that the survival probability is energy dependent, chiefly because the neutrino deficit observed by Gallex[24] and SAGE[25] was different from that observed by Homestake[26] and also from that observed by the SuperKamiokande experiment.[27] The LMA-MSW effect provides an elegant mechanism to recover the energy dependence of the survival probability and Borexino is the first experiment which has been able to test it directly by measuring the survival probability of several neutrino components of various energies.

The LMA-MSW theory predicts a very well defined value of P_{ee} which is shown in Fig. 3 as a pink band whose thickness corresponds to the theoretical error. The same figure also shows the result of the P_{ee} computed using Borexino data using the measurements of *pp*, ^{7}Be, *pep* and ^{8}B neutrinos by means of formula (2). Neutrinos from *pp* initial fusion into deuteron are very low energy and their flavour conversion is dominated by vacuum oscillations. The grey band labelled as Vacuum-LMA corresponds to:

$$P_{ee}^{vac} = 1 - \frac{1}{2}\sin^2 2\theta_{12}\,, \tag{2}$$

which is the expected value for vacuum oscillations in case of completely incoherent propagation.

The neutrinos produced by ^{7}Be electron capture and *pep* three body fusion are mono–chromatic, so they bear no uncertainty in horizontal axis. *pep* is well within the transition region and shows a hint of that, with a large statistical uncertainty. ^{7}Be energy is still in the vacuum regime region and its corresponding P_{ee} is well consistent with that. Finally, ^{8}B neutrinos, on the other side, are well within the matter dominated LMA-MSW regime. Being a broad β^+ spectrum it is possible

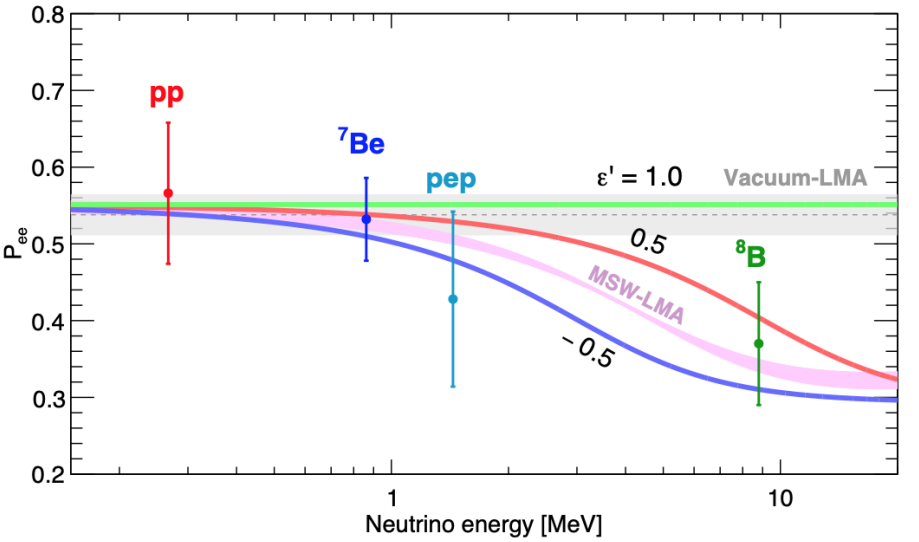


Fig. 4. Electron neutrino survival probability P_{ee} as a function of neutrino energy in LMA-MSW solution with uncertainties of oscillation parameters taken into account (pink band), and LMA-MSW + NSI solutions for ϵ'= -0.5, 0.5, 1.0 and average values of oscillation parameters. Vacuum oscillations scenario with LMA-MSW parameters is also shown (grey band). The definition of the non–standard interaction parameter is that of reference.[28]

to bin in energy, offering the data points shown in grey, while the green one is the average value. The error are too large, however, to show any hint of the expected rise of P_{ee} at lower energies (the slight uprise of the points has no real statistical significance, though consistent).

Globally, Borexino data clearly are well consistent with the theoretical expectation, confirming LMA-MSW effect, though the error are still large. Unfortunately, there is no room for Borexino to improve further.

In order to improve the result significantly two elements are important, depending on the energy. The ^{7}Be point has an error dominated by the theoretical uncertainty on the expected flux, while other points have also a relevant error coming from the rate measurement. Therefore, a better test of LMA-MSW theory requires an improvement in the Standard Solar Model predictions. For a discussion of the main sources of uncertainties see.[5]

The result obtained on the P_{ee} can be used to test the existence of low energy non–standard neutrino interactions beyond those predicted by the Standard Model. We have carefully studied this opportunity, parametrising the effect of non–standard interactions in terms of a single parameter ϵ'.[28] Fig. 4 shows that with the current theoretical and experimental uncertainties there is still quite a lot of room for new physics and that the future better measurement of this curve might be worth pursuing.

5. Conclusions

Borexino is the first and sole experiment able to measure all neutrino components from *pp*-chain reactions in the Sun and probe the electron neutrino survival probability as a function of energy.

The results confirm the LMA-MSW scenario with nice agreement with Standard Model physics. The precision is still low and room for new physics exists.

Borexino has opened the era of precise solar neutrino spectroscopy, but even better experiments are desirable to complete the job.

Acknowledgements

I thank the organisers of the Sixteenth Marcel Grossman meeting for the beautiful conference and for offering me the opportunity to give this talk. I am also grateful to the whole Borexino Collaboration for the long and beautiful journey we made all together. Finally, I warmly thank the Laboratori Nazionali del Gran Sasso for the continual support given to Borexino over a period of more than thirty years.

References

1. Q. R. Ahmad *et al.*, Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **89**, p. 011301 (2002).
2. Q. R. Ahmad *et al.*, Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **87**, p. 071301 (2001).
3. A. B. McDonald, Nobel Lecture: The Sudbury Neutrino Observatory: Observation of flavor change for solar neutrinos, *Rev. Mod. Phys.* **88**, p. 030502 (2016).
4. N. Vinyoles, A. M. Serenelli, F. L. Villante, S. Basu, J. Bergström, M. C. Gonzalez-Garcia, M. Maltoni, C. Peña Garay and N. Song, A new Generation of Standard Solar Models, *Astrophys. J.* **835**, p. 202 (2017).
5. M. Agostini *et al.*, Comprehensive measurement of *pp*-chain solar neutrinos, *Nature* **562**, 505 (2018).
6. M. Agostini *et al.*, Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun, *Nature* **587**, 577 (2020).
7. G. Alimonti *et al.*, Science and technology of BOREXINO: A Real time detector for low-energy solar neutrinos, *Astropart. Phys.* **16**, 205 (2002).
8. G. L. Fogli, E. Lisi, A. Marrone, D. Montanino, A. Palazzo and A. M. Rotunno, Global analysis of neutrino masses, mixings and phases: entering the era of leptonic CP violation searches, *Phys. Rev. D* **86**, p. 013012 (2012).
9. L. Wolfenstein, Neutrino Oscillations in Matter, *Phys. Rev. D* **17**, 2369 (1978).
10. S. P. Mikheev and A. Y. Smirnov, Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy, *Nuovo Cim. C* **9**, 17 (1986).
11. S. P. Mikheyev and A. Y. Smirnov, Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos, *Sov. J. Nucl. Phys.* **42**, 913 (1985).
12. G. Alimonti *et al.*, The Borexino detector at the Laboratori Nazionali del Gran Sasso, *Nucl. Instrum. Meth. A* **600**, 568 (2009).
13. G. Bellini *et al.*, Muon and Cosmogenic Neutron Detection in Borexino, *JINST* **6**, p. P05005 (2011).

14. G. Alimonti *et al.*, The liquid handling systems for the Borexino solar neutrino detector, *Nucl. Instrum. Meth. A* **609**, 58 (2009).
15. M. Agostini *et al.*, The Monte Carlo simulation of the Borexino detector, *Astropart. Phys.* **97**, 136 (2018).
16. G. Bellini *et al.*, Final results of Borexino Phase-I on low energy solar neutrino spectroscopy, *Phys. Rev. D* **89**, p. 112007 (2014).
17. M. Agostini *et al.*, First Simultaneous Precision Spectroscopy of *pp*, ^{7}Be, and *pep* Solar Neutrinos with Borexino Phase-II, *Phys. Rev. D* **100**, p. 082004 (2019).
18. G. Bellini *et al.*, First evidence of pep solar neutrinos by direct detection in Borexino, *Phys. Rev. Lett.* **108**, p. 051302 (2012).
19. G. Bellini *et al.*, Measurement of the solar 8B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector, *Phys. Rev. D* **82**, p. 033006 (2010).
20. G. Bellini *et al.*, Neutrinos from the primary proton–proton fusion process in the Sun, *Nature* **512**, 383 (2014).
21. G. Bellini *et al.*, Observation of Geo-Neutrinos, *Phys. Lett. B* **687**, 299 (2010).
22. G. Bellini *et al.*, Measurement of geo-neutrinos from 1353 days of Borexino, *Phys. Lett. B* **722**, 295 (2013).
23. M. Agostini *et al.*, Spectroscopy of geoneutrinos from 2056 days of Borexino data, *Phys. Rev. D* **92**, p. 031101 (2015).
24. W. Hampel *et al.*, GALLEX solar neutrino observations: Results for GALLEX IV, *Phys. Lett. B* **447**, 127 (1999).
25. J. N. Abdurashitov *et al.*, Solar neutrino flux measurements by the Soviet-American Gallium Experiment (SAGE) for half the 22 year solar cycle, *J. Exp. Theor. Phys.* **95**, 181 (2002).
26. B. T. Cleveland, T. Daily, R. Davis, Jr., J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain and J. Ullman, Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *Astrophys. J.* **496**, 505 (1998).
27. S. Fukuda *et al.*, Determination of solar neutrino oscillation parameters using 1496 days of Super-Kamiokande I data, *Phys. Lett. B* **539**, 179 (2002).
28. S. K. Agarwalla *et al.*, Constraints on flavor-diagonal non-standard neutrino interactions from Borexino Phase-II, *JHEP* **02**, p. 038 (2020).

The relevance of pp-chain and CNO-cycle neutrino measurements for solar physics

F. L. Villante[1,2]

[1] *Dipartimento di Scienze Fisiche e Chimiche, Universitá degli Studi dell'Aquila, Via Vetoio, L'Aquila, I-67100, Italy*

[2] *Laboratori Nazionali del Gran Sasso (LNGS), Istituto Nazionale di Fisica Nucleare (INFN), Via G. Acitelli 22, Assergi (AQ), I-67100, Italy*

E-mail: villante@lngs.infn.it

A. M. Serenelli[3,4]

[3] *Institute of Space Sciences (ICE, CSIC), Carrer de Can Magrans s/n, Cerdanyola del Valles, E-08193, Spain*

[4] *Institut d'Estudis Espacials de Catalunya, Carrer Gran Capita 2.4, Barcelona, E-08034, Spain*

E-mail: a.serenelli@csic.es

We discuss the present status of Standard Solar Models in order to provide the scientific framework to fully understand the relevance of the latest solar neutrino flux measurements performed by Borexino. After reviewing recent Standard Solar Model calculations, we discuss the solar abundance problem and the composition-opacity degeneracy. We then focus on hydrogen burning and neutrino production in Sun, discussing an analytical model that helps understanding the dependence of neutrino fluxes on solar core temperature and composition and the possibility they offer for determining physical characteristics of the solar interior.

Keywords: Solar neutrino fluxes, solar physics, solar models

1. Introduction

Being the closest star to us, the Sun can be investigated by a variety of different approaches/messengers. We can e.g. study the electromagnetic radiation that is emitted from its photosphere, determining its luminosity $L_\odot$, surface temperature $T_{\rm eff}$ and radius $R_\odot$. By looking at the absorption lines that are formed in the solar atmosphere, we are also able to infer the surface composition through spectroscopic techniques, more precisely the present-day abundance of chemical elements relative to hydrogen, usually summarized in the metal-to-hydrogen ratio $(Z/X)_\odot$. At the same time, the properties of the solar core, in particular its temperature and composition, can be probed by measuring the fluxes of neutrinos produced by nuclear reactions responsible for hydrogen burning. Finally, the properties of the solar interior, namely the sound speed stratification and the depth and helium abundance of

the external convective envelope, can be inferred with high accuracy by helioseismology, i.e. by looking at non-radial oscillations of the solar surface. As a result of these observations, we have a redundancy of experimental information that make the Sun an over-constrained system. This allows us to use the Sun as a laboratory for fundamental physics and a benchmark for stellar astrophysics.

As a remarkable example, the development of increasingly more precise Standard Solar Models[1,2] (SSMs), i.e. self calibrated theoretical models of the solar interior that automatically implements the observational values of $L_\odot$, $R_\odot$ and $(Z/X)_\odot$ and have no free parameters, and the fact that they were able to successfully reproduce the tight constraints imposed by helioseismology,[3] suggested that solution of the solar neutrino problem had to be found outside the realm of astrophysics, as had originally been suggested a few years earlier.[4] Simultaneously, Super-Kamiokande[5,6] led to the precise measurement of ^{8}B neutrino flux which, in combination with the results of radiochemical experiments Homestake,[7] Gallex[8] and SAGE[9] strongly hinted at the existence of solar neutrino oscillations, result confirmed just a few years later by SNO results.[10,11]

The discovery of neutrino oscillations and the accurate and precise determination of neutrino fluxes from individual reactions changed the focus of interest of solar physics. Starting from 2007, Borexino has obtained an almost complete characterization of the spectrum of neutrinos from the pp-chain.[12,13] Together with the very precise measurement of the ^{8}B flux from SNO[14] and Super-Kamiokande,[15] we have come full circle and results from solar neutrino experiments can now be used to learn about the properties of the Sun. This is timely. There is a lingering dispute about which is the detailed chemical composition of the Sun, the solar abundance problem, that is intimately linked to the uncertainties in our knowledge of radiative opacities in the solar interior. Solar neutrino data can in principle be used to disentangle this problem,[16–18] in particular if the promising results by Borexino on solar CN neutrinos[19] can be further improved.

The goal of this work is to discuss the present status of SSMs and to provide the scientific framework to fully understand the relevance of the recent solar neutrino flux measurements performed by Borexino. After reviewing recent SSM calculations (Sect. 2), we discuss the solar abundance problem (Sect. 3) and the composition-opacity degeneracy (Sect. 4). We then focus on hydrogen burning (Sect. 5) and solar neutrino production (Sect. 6), discussing an analytical model (Sect. 7) that helps understanding the dependence of neutrino fluxes on solar core temperature, composition and nuclear cross sections, and the possibility they offer for determining physical characteristics of the solar interior (Sect. 8).

2. Standard solar models

SSMs are a snapshot in the evolution of a $1\,M_\odot$ star, calibrated to match present-day surface properties of the Sun. Two basic assumptions in SSM calculations are: 1) after the phase of star formation the Sun was chemically homogenized as a

result of the fully convective phase during its contraction along the Hayashi track and before nuclear reactions start altering its initial composition and, 2) at all moments during its evolution up to the present solar age $\tau_\odot = 4.57\,\mathrm{Gyr}$ mass loss is negligible. The calibration is done by adjusting the mixing length parameter (α_{MLT}) and the initial helium and metal mass fractions (Y_{ini} and Z_{ini} respectively) in order to satisfy the constraints imposed by the present-day solar luminosity $\mathrm{L}_\odot = 3.8418 \times 10^{33}\,\mathrm{erg\,s^{-1}}$, radius $\mathrm{R}_\odot = 6.9598 \times 10^{10}\mathrm{cm}$,[20] and surface metal to hydrogen abundance ratio $(\mathrm{Z/X})_\odot$, see sect.3. As a result of this procedure, SSM has no free parameters and completely determines the physical properties of the Sun. It can be then validated (or falsified) by other observational constraints, in particular by those provided by solar neutrino fluxes measurements and helioseismic frequencies determinations.

The physics input in the SSM is rather simple and it accounts for: convective and radiative transport of energy, chemical evolution driven by nuclear reactions, microscopic diffusion of elements which comprises different processes but among which gravitational settling dominates. For over more than 25 years, since the modern version of the SSM was established with the inclusion of microscopic diffusion,[21,22] the continuous improvement of the constitutive physics has brought about the changes and the evolution of SSMs. In particular, a lot of effort has gone into experimental and theoretical work on nuclear reaction rates. But changes in radiative opacities and the equation of state were also relevant. We take here as a reference the results of recent SSM calculations,[23] the so-called Barcelona 2016 (B16, for short) SSMs, which are based on the following state of the art ingredients. The equation of state is calculated consistently for each of the compositions used in the solar calibrations by using FreeEOS.[24] Atomic radiative opacities are from the Opacity Project (OP),[25] complemented at low temperatures with molecular opacities from Ref. 26. Nuclear reaction rates for the pp-chain and CNO-bicycle are from the Solar Fusion II compilation[27] with important updates for the rates of $\mathrm{p(p},e^+\nu_e)\mathrm{d}$, $^7\mathrm{Be(p},\gamma)^8\mathrm{B}$ and $^{14}\mathrm{N(p},\gamma)^{15}\mathrm{O}$ reactions, see Refs. 28–32. Microscopic diffusion coefficients are computed as described in Ref. 33. Convection is treated according to the mixing length theory.[34] The atmosphere is grey and modeled according to a Krishna-Swamy $T-\tau$ relationship.[35]

3. The solar composition problem

The solar surface composition, determined with spectroscopic techniques, is a fundamental input in the construction of SSMs. The development of three dimensional hydrodynamic models of the solar atmosphere, of techniques to study line formation under non-local thermodynamic conditions and the improvement in atomic properties (e.g. transition strengths) have led since 2001 to a complete revision of solar abundances. Table 1 lists the abundances determined by different authors for the most relevant metals in solar modeling: GN93,[36] GS98,[37] AGSS09,[38] C11,[39] AGSS15,[40–42] and AAG20.[43]

Table 1. Solar photospheric composition through time and authors for most relevant metals in solar modeling. Abundances are given in the standard astronomical scale $\epsilon_i = \log_{10}(n_i/n_{\rm H}) + 12$, where n_i is the number density of a given atomic species.

El.	GN93	GS98	AGSS09	C11	AGSS15	AAG20
C	8.55	8.52	8.43	8.50	—	8.46
N	7.97	7.92	7.83	7.86	—	7.83
O	8.87	8.83	8.69	8.76	—	8.69
Ne	8.08	8.08	7.93	8.05	7.93	8.06
Mg	7.58	7.58	7.60	7.54	7.59	7.55
Si	7.55	7.55	7.51	7.52	7.51	7.51
S	7.33	7.33	7.13	7.16	7.13	7.12
Fe	7.50	7.50	7.50	7.52	7.47	7.46
$(Z/X)_\odot$	0.0245	0.0230	0.0180	0.0209	—	0.0187

Note that only abundances relative to hydrogen can be obtained from spectroscopy because the intensity of spectroscopic lines is measured relative to a continuum that is determined by the hydrogen abundance in the solar atmosphere. The last row in the table gives the total photospheric present-day metal-to-hydrogen ratio $(Z/X)_\odot$ and it is the quantity used as observational constraint to construct a solar model. In fact, the solar composition set used in solar models determines not only $(Z/X)_\odot$ but also the relative abundances of metals in the models. In this sense, $Z_{\rm ini}$ acts as a normalization factor that, together with $Y_{\rm ini}$ and the relation $X_{\rm ini} + Y_{\rm ini} + Z_{\rm ini} = 1$, determines completely the initial composition of the model.

There is no complete agreement among authors, and some controversy still remains as to what the best values for the new spectroscopic abundances are. See in particular recent work on solar oxygen,[44] based on the same 3D atmosphere models as used by the Asplund group, but improved atomic data and methods, leads to $\epsilon_{\rm O} = 8.75 \pm 0.03$, i.e. very close to the C11 value. However, at the moment there seems to be consensus in that all determinations of the solar metallicity based on the new generation of spectroscopic studies yield a solar metallicity lower than older spectroscopic results,[36,37] in particular for the volatile and most abundant C, N, and O. For refractories elements, like Fe, Si, Mg and S that have important role in solar modeling being important contributors to the radiative opacity, meteorites offer a very valuable alternative method (see e.g. Ref. 45) and, in fact, elemental abundances determined from meteorites have been historically more robust than spectroscopic ones.

Considering that uncertainties of element abundances are difficult to quantify, it has become customary to consider two canonical sets of abundances to which we refer to as high metallicity (HZ) and low metallicity (LZ) solar admixtures as reference assumptions for SSM calculations. These are obtained by using the photospheric (volatiles) + meteoritic (refractories) abundances from GS98 and AGSS09 respectively, and are reported in Table 2. In the last column, we give the fractional differences $\delta z_i \equiv z_i^{\rm HZ}/z_i^{\rm LZ} - 1$ where $z_i \equiv Z_i/X$ is the ratio of the $i-$element

Table 2. The two canonical HZ and LZ solar mixtures given as $\epsilon_i = \log_{10}(n_i/n_{\rm H}) + 12$. The two compilations are obtained by using the photospheric (volatiles) + meteoritic (refractories) abundances from GS98 and AGSS09 respectively, and correpond to the admixture labelled as GS98 and AGSS09met in Ref. 23.

El.	High-Z (HZ)	Low-Z (LZ)	δz_i
C	8.52 ± 0.06	8.43 ± 0.05	0.23
N	7.92 ± 0.06	7.83 ± 0.05	0.23
O	8.83 ± 0.06	8.69 ± 0.05	0.38
Ne	8.08 ± 0.06	7.93 ± 0.10	0.41
Mg	7.58 ± 0.01	7.53 ± 0.01	0.12
Si	7.56 ± 0.01	7.51 ± 0.01	0.12
S	7.20 ± 0.06	7.15 ± 0.02	0.12
Ar	6.40 ± 0.06	6.40 ± 0.13	0.00
Fe	7.50 ± 0.01	7.45 ± 0.01	0.12
$(\mathrm{Z/X})_\odot$	0.02292	0.01780	0.29

abundance with that of hydrogen, to facilitate comparison among the two admixtures. Even if GS98 abundances are presumably surpassed by the more recent determinations, they are still considered as a valid option to construct solar models because they lead to a temperature stratification that well reproduces the helioseismic constraints.

This can be better appreciated by considering Tab. 3 and Fig. 1 where we compare theoretical predictions of SSMs implementing HZ and LZ surface composition with helioseismic determinations of the surface helium abundance $Y_{\rm s}$, of the convective envelope depth $R_{\rm CZ}$ and the solar sound speed $c_\odot(r)$. We see that solar models implementing the LZ abundances fail to reproduce all helioseismic probes of solar properties. This disagreement constitutes the so-called *solar abundance problem*[46–48] that has defied a complete solution. All proposed modifications to physical processes in SSMs offer, at best, only partial improvements in some helioseismic probes (e.g.[49–53]). An alternative possibility is to consider modifications to the physical inputs of SSMs at the level of the constitutive physics, radiative opacities in particular (see next section).

4. The opacity-composition degeneracy

The interpretation of the solar abundance problem is complicated by the degeneracy between effects produced by a modification of the radiative opacity $\kappa(\rho, T, Y, Z_i)$ and effects induced by a change of the heavy element admixture $\{z_i\}$, expressed here in terms of the quantities $z_i \equiv Z_i/X$ where Z_i is the surface abundance of the i-element and X is that of hydrogen.

This degeneracy was discussed in quantitative terms in Ref. 55 by using the linear solar model (LSM) approach introduced in Ref. 56. By neglecting the role of

Table 3. Main characteristics of SSMs with different surface composition.[23] The observational values for Y_s and R_{CZ} are taken from Ref. 46 and 54, respectively. The quantity $\delta c/c = (c_\odot - c_{mod})/c_{mod}$ is the fractional difference between sound speed helioseismic determination and model prediction.

Qnt.	B16-HZ	B16-LZ	Solar
Y_s	0.2426 ± 0.0059	0.2317 ± 0.0059	0.2485 ± 0.0035
$R_{CZ}/R_\odot$	0.7116 ± 0.0048	0.7223 ± 0.0053	0.713 ± 0.001
$\langle \delta c/c \rangle$	$0.0005^{+0.0006}_{-0.0002}$	0.0021 ± 0.001	-
α_{MLT}	2.18 ± 0.05	2.11 ± 0.05	-
Y_{ini}	0.2718 ± 0.0056	0.2613 ± 0.0055	-
Z_{ini}	0.0187 ± 0.0013	0.0149 ± 0.0009	-
Z_s	0.0170 ± 0.0012	0.0134 ± 0.0008	-
Y_c	0.6328 ± 0.0053	0.6217 ± 0.0062	-
Z_c	0.0200 ± 0.0014	0.0159 ± 0.0010	-

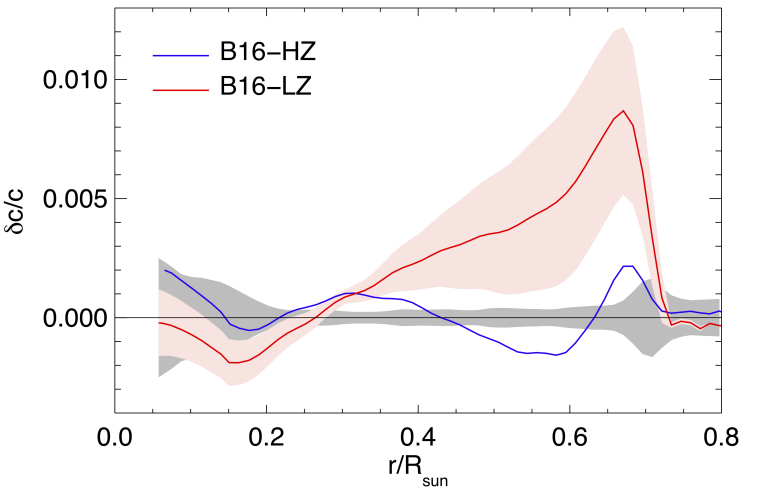


Fig. 1. Fractional sound speed difference in the sense $\delta c/c = (c_\odot - c_{mod})/c_{mod}$. Grey shaded regions corresponds to errors from helioseismic inversion procedure. Red shaded region corresponds to uncertainties in SSM predictions which we chose to plot around the B16-LZ central value (solid red line). An equivalent relative error band holds around the central value of the B16-HZ central value (solid blue line) which we do not plot for the sake of clarity.

metals in the equation of state and in the energy generation coefficient, it was shown that the source term $\delta\kappa(r)$ that drives the modification of the solar properties and that can be constrained by observational data can be written as the sum of two contributions:

$$\delta\kappa(r) = \delta\kappa_I(r) + \delta\kappa_Z(r) \quad (1)$$

The first term $\delta\kappa_{\rm I}(r)$, which we refer to as *intrinsic* opacity change, represents the fractional variation of the opacity along the solar profile and it is given by:

$$\delta\kappa_{\rm I}(r) = \frac{\kappa(\overline{\rho}(r), \overline{T}(r), \overline{Y}(r), \overline{Z}_i(r))}{\overline{\kappa}(\overline{\rho}(r), \overline{T}(r), \overline{Y}(r), \overline{Z}_i(r))} - 1 \tag{2}$$

where the notation $\overline{Q}$ indicates, here and in the following, the value for the generic quantity Q in reference SSM calculation. The second term $\delta\kappa_{\rm Z}(r)$, which we refer to as *composition* opacity change, describes the effects of a variation of $\{z_{\rm i}\}$. It takes into account that a modification of the photospheric admixture implies a different distribution of metals inside the Sun and, thus, a different opacity profile, even if the function $\kappa(\rho, T, Y, Z_i)$ is unchanged. The contribution $\delta\kappa_{\rm Z}(r)$ is given by:

$$\delta\kappa_{\rm Z}(r) = \frac{\overline{\kappa}(\overline{\rho}(r), \overline{T}(r), \overline{Y}(r), Z_i(r))}{\overline{\kappa}(\overline{\rho}(r), \overline{T}(r), \overline{Y}(r), \overline{Z}_i(r))} - 1 \tag{3}$$

where $Z_i(r) \simeq \overline{Z}_i(r)\,(z_i/\overline{z}_i)$ and can be calculated as:

$$\delta\kappa_{\rm Z}(r) \simeq \sum_i \left.\frac{\partial \ln \overline{\kappa}}{\partial \ln Z_i}\right|_{\rm SSM} \delta z_i \tag{4}$$

where δz_i represents the fractional variation of z_i and the symbol $|_{\rm SSM}$ indicates that we calculate the derivatives along the density, temperature and chemical composition profiles predicted by the reference SSM. Equation (1), although being approximate, is quite useful because it makes explicit the connection (and the degeneracy) between the effects produced by a modification of the radiative opacity and of those produced by a modification of the heavy element admixture.

Early works[57,58] suggested that a localized increase in opacities could solve or, at least, alleviate the disagreement of low-Z solar models with helioseismology. Refs. 59 and 55 have concluded that a tilted increase in radiative opacities, with a few percent increase in the solar core and a larger (15-20%) increase at the base of the convective envelope could lead to low-Z SSMs that would satisfy helioseismic probes equally as well as SSMs based on the older, higher, metallicities. Incidentally, the above conclusion is confirmed by the analysis presented in Ref. 18 where helioseismic and solar neutrino data are used to infer the optimal composition of the Sun (see their Fig. 10).

Recent years have seen a surge of activity in theoretical calculations of atomic radiative opacities. Updated calculations[25] by the Opacity Project have led the way, followed by OPAS,[60,61] STAR[62] and a new version of OPLIB, the opacities from Los Alamos.[63] For conditions in solar interiors, all theoretical opacities agree with each other within few %. Interestingly, Ref. 64 have presented the first ever measurement of opacity under conditions close to those present in the solar interior. Four measurements were carried out to measure the iron opacity, each at different temperature and electron densities. The most extreme conditions used in the experiment approach those at the base of the solar convective envelope. While the experiment has been carried out only for iron, their conclusion is that all theoretical

calculations predict a too low Rosseland mean opacity, at a level of $7 \pm 4\%$, for the temperature and density combinations realized in the experiment. Further experimental work on chromium and nickel opacities[65] helps to evaluate discrepancies between experimental and theoretical results on iron opacity. Results point towards shortcomings that affect models, particularly in the case of open electronic L-shell configurations such as is present in iron at the base of the convective envelope. Also, the disagreement between theoretical and measured line shapes for the three elements indicates shortcomings in the theoretical understanding of atomic interaction with the plasma. On the other hand, the results also indicate that the quasicontinuum opacity determined experimentally agrees well with the chromium and nickel experiments, contrary to results from the iron experiment. However, the chromium and nickel experiments were carried out at lower temperatures than those used in the extreme conditions of the original iron experiment, which suggests that the problem of missing quasicontinuum opacity might have an unknown temperature dependence, or that a systematic error affected the high temperature iron measurements. Moreover,[66] in a recent theoretical analysis of line broadening modeling in opacity calculations, have found that uncertainties linked to this are larger at the base of the convective envelope than in the core. These arguments suggest that opacity calculations are more accurate in the solar core than in the region around the base of the convective envelope.

5. Hydrogen burning in the Sun

The Sun is powered by nuclear reactions that transform hydrogen into helium-4. The overall effect of these reactions is the conversion:

$$4\,\mathrm{p} + 2\,e^- \rightarrow {}^4\mathrm{He} + 2\,\nu_\mathrm{e} \tag{5}$$

with the production of a fixed amount of energy $Q = 4\,m_\mathrm{p} + 2m_\mathrm{e} - m_{^4\mathrm{He}} = 26.7\,\mathrm{MeV}$ per synthesized ^{4}He nucleus. Most of this energy is released in the solar plasma and slowly diffuses toward the solar surface supporting the radiative luminosity of the Sun. A small fraction of it, that depends on the specific channel by which hydrogen burning proceeds, is emitted in neutrinos. According to SSM calculations, the two neutrinos carry away about 0.6 MeV on the average.

The SSM predicts that most of the solar energy ($> 99\%$) is produced by the pp-chain, i.e. the hydrogen fusion reaction chain displayed in the left panel of Fig. 2. The pp-chain is is mostly initiated by $\mathrm{p}(\mathrm{p}, e^+\nu_e)\mathrm{d}$ reaction and, to a minor extent, by electron capture reaction $\mathrm{p}(\mathrm{p}e^-, \nu_e)\mathrm{d}$ and has several possible terminations that depend on the specific mechanism by which helium-3 nuclei, which are produced by $\mathrm{d}(\mathrm{p}, \gamma)^3\mathrm{He}$ reaction, are converted to heavier elements. In the Sun, the dominant mechanism is $^3\mathrm{He}(^3\mathrm{He}, 2p)^4\mathrm{He}$ that corresponds to the so-called pp-I termination of the pp-chain. Alternatively, helium-3 can lead to production of ^{7}Be through the reaction $^3\mathrm{He}(^4\mathrm{He}, \gamma)^7\mathrm{Be}$. Depending on the destiny of ^{7}Be, that can be processed either by the electron capture $^7\mathrm{Be}(e^-, \nu_e)^7\mathrm{Li}$ or by the (largely sub-dominant)

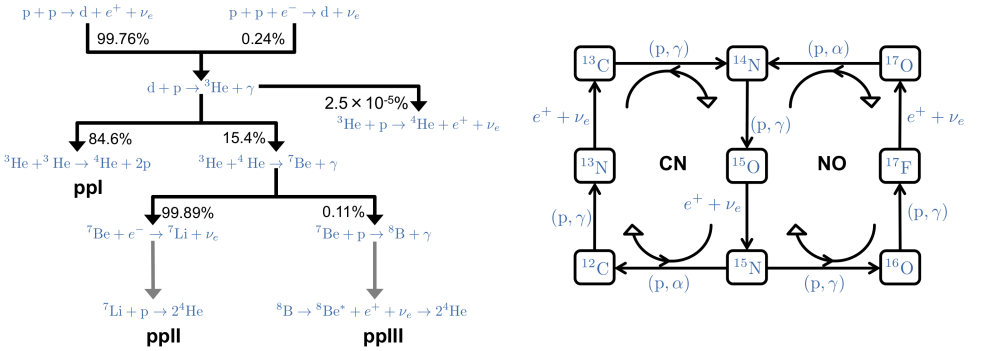


Fig. 2. Left Panel: The pp-chain; Right Panel: The CNO-bicycle.

proton capture reaction $^7\mathrm{Be}(\mathrm{p},\gamma)^8\mathrm{B}$, one obtains the pp-II or the pp-III terminations of the chain. Finally, a very small amount of helium-4 nuclei is produced by $^3\mathrm{He}(\mathrm{p},e^+\nu_e)^4\mathrm{He}$ reaction. The relative importance of the different branches of the pp-chain depends primarily on the core temperature of the Sun and on the cross section of specific reactions, as will be discussed in next section. The numbers given in Fig. 2 show the branching ratios in the present Sun.

An alternative hydrogen burning mechanism is provided by the CNO-bicycle that is displayed in the right panel of Fig. 2. The CNO-bicycle uses carbon, nitrogen and oxygen nuclei that are present in the core of the Sun as catalysts for hydrogen fusion. It is composed by two different branches, i.e. the CN-cycle and the NO-cycle, whose relative importance depends on the outcome of proton capture reaction on $^{15}\mathrm{N}$. In the Sun, the $^{15}\mathrm{N}(\mathrm{p},\alpha)^{12}\mathrm{C}$ channel is largely dominant and so, in practice, the CNO-bicycle is reduced to the CN-cycle with a marginal contribution by the NO-cycle. Note that the CN-cycle conserves the total number of C and N nuclei in the core of the Sun, but alters their distribution as it burns into equilibrium, eventually achieving equilibrium abundances proportional to the inverse of the respective rates. The reactions controlling conversion of $^{12}\mathrm{C}$ and $^{14}\mathrm{N}$ in the solar core and the approach to equilibrium are $^{12}\mathrm{C}(\mathrm{p},\gamma)^{13}\mathrm{N}$ and $^{14}\mathrm{N}(\mathrm{p},\gamma)^{15}\mathrm{O}$: these are the next-to-slowest and slowest rates in the CN-cycle, respectively. The temperature above which the $^{12}\mathrm{C}$ burning time through $^{12}\mathrm{C}(\mathrm{p},\gamma)^{13}\mathrm{N}$ is smaller than the Sun's lifetime is $T \sim 10^7\,\mathrm{K}$. In the SSM, the entire energy-producing core, $r \lesssim 0.2R_\odot$ and $m \lesssim 0.3M_\odot$ is at temperature larger than this value, so that nearly all of the core's $^{12}\mathrm{C}$ is converted to $^{14}\mathrm{N}$. The slower $^{14}\mathrm{N}(\mathrm{p},\gamma)^{15}\mathrm{O}$ reaction determines whether equilibrium is achieved. The $^{14}\mathrm{N}$ burning time is shorter than the age of the Sun for $T \gtrsim 1.3 \times 10^7\,\mathrm{K}$. Therefore equilibrium for the CN cycle is reached only for $R \lesssim 0.1R_\odot$, corresponding to the central 7% of the Sun by mass. Consequently, over a significant portion of the outer core, $^{12}\mathrm{C}$ is converted to $^{14}\mathrm{N}$, but further reactions are inhibited by the $^{14}\mathrm{N}(\mathrm{p},\gamma)^{15}\mathrm{O}$ bottleneck.

Table 4. Solar neutrino fluxes predicted by SSMs with different surface composition.[23] Units are: 10^{10} (pp), 10^{9} (^{7}Be), 10^{8} (pep, ^{13}N, ^{15}O), 10^{6} (^{8}B,^{17} F), 10^{5} (eN, eO) and 10^{3} (hep, eF) cm^{-2}s^{-1}.

Flux	B16-HZ	B16-LZ
Φ(pp)	$5.98(1 \pm 0.006)$	$6.03(1 \pm 0.005)$
Φ(pep)	$1.44(1 \pm 0.01)$	$1.46(1 \pm 0.009)$
Φ(hep)	$7.98(1 \pm 0.30)$	$8.25(1 \pm 0.30)$
$\Phi(^{7}\mathrm{Be})$	$4.93(1 \pm 0.06)$	$4.50(1 \pm 0.06)$
$\Phi(^{8}\mathrm{B})$	$5.46(1 \pm 0.12)$	$4.50(1 \pm 0.12)$
$\Phi(^{13}\mathrm{N})$	$2.78(1 \pm 0.15)$	$2.04(1 \pm 0.14)$
$\Phi(^{15}\mathrm{O})$	$2.05(1 \pm 0.17)$	$1.44(1 \pm 0.16)$
$\Phi(^{17}\mathrm{F})$	$5.29(1 \pm 0.20)$	$3.26(1 \pm 0.18)$
Φ(eN)	$2.20(1 \pm 0.15)$	$1.61(1 \pm 0.14)$
Φ(eO)	$0.81(1 \pm 0.17)$	$0.57(1 \pm 0.16)$
Φ(eF)	$3.11(1 \pm 0.20)$	$1.91(1 \pm 0.18)$

6. Solar neutrinos

A very effective tool to investigate the properties of the Sun is provided by neutrinos which are necessarily produced along with ^{4}He nuclei during hydrogen burning, in order to satisfy lepton number conservation. Neutrinos free stream in the solar plasma and reach the Earth in about 8 minutes where they can be detected by solar neutrino experiments. While the total amount of neutrinos produced in the Sun can be easily estimated from the solar luminosity constraint (see below), the evaluation of their spectrum requires the knowledge of the individual rates of neutrino producing reactions and thus the construction of a complete solar model. We report in Fig. 3 and Tab. 4, the SSM predictions for the different components of the solar neutrino flux, named according to the specific reaction by which they are produced.[23] We also include, for completeness, ecCNO neutrinos, i.e. neutrinos produced by electron capture reaction in the CNO-bicycle (in addition to the "standard" CNO neutrinos produced by β decays of ^{13}N, ^{15}O and ^{17}F) that were originally calculated in Refs. 67 and 68 and recently reevaluated in Ref. 69[a]. The two columns "B16-HZ" and "B16-LZ" reported in Tab. 4 are obtained by considering two different options for the solar surface composition, as it discussed in Sect. 3. During the last few decades, solar neutrino experiments have allowed us to determine with great accuracy most of the components of the solar flux. As an example, ^{7}Be and ^{8}B neutrino fluxes are measured with accuracy better than $\sim 3\%$ by Borexino,[13] Super-Kamiokande[15] and SNO.[14] The pp and pep-neutrino flux can be determined

[a]In order to take into account the new inputs in B16-SSM calculations, the ecCNO fluxes given in Tab. 4 have been scaled with respect to the values quoted in Ref. 69 proportionally to the corresponding β-decay fluxes. This follows from the assumption that the ratio of electron capture and beta decay processes in the Sun is equal to what evaluated in Ref. 69

with $\lesssim 1\%$ accuracy by assuming the solar luminosity constraint, see e.g. Ref. 70. These fluxes, however, have been also directly measured by Borexino[13,71,72] with $\sim 10\%$ and $\sim 17\%$ accuracy, respectively. Finally, Borexino has recently obtained the experimental identification of CNO neutrinos,[19] providing the first direct evidence that CNO-bicycle is active in the Sun.

The solar luminosity constraint

SSM calculations predict that the Sun is very close to thermal equilibrium and that evolutionary timescales are much longer than the heat diffusion time across the solar structure. In this case, the amount of energy which is produced (in unit time) by nuclear reactions in the solar core almost exactly counterbalance the luminosity radiated at the surface of the Sun, with the contribution of gravothermal energy to the solar luminosity being $< 0.1\%$. Since a fixed amount of energy Q and two electron neutrinos are produced for synthesized ^{4}He nucleus, and neglecting the small gravothermal contribution, this argument implies a relationship between the measured solar luminosity and the solar neutrino fluxes that can be expressed as

$$\frac{\mathrm{L}_\odot}{4\pi D^2} = \sum_i \left(\frac{Q}{2} - \langle E_\nu \rangle_i \right) \Phi(i) \tag{6}$$

where the index i indicates the different components of the neutrino flux and the quantity $\langle E_\nu \rangle_i$ is the average energy of neutrinos produced by a specific reaction. This argument was illustrated and discussed e.g. in Ref. 73,74 and has been recently reviewed in Refs. 75, 76.

It is clearly important to experimentally confirm the validity of the solar luminosity constraint because this represents a test of fundamental assumptions in the SSM paradigm, e.g. that the Sun is stable on timescales much longer than the heat diffusion time ($\sim 10^5$y), and that the are no additional energy losses or production mechanisms besides those normally included in solar model calculations. The recent Borexino direct identification of the pp-neutrino flux,[13] which provides the dominant term in the r.h.s of Eq. (6), has finally made this test possible, confirming the energy balance employed by SSMs at the $\sim 10\%$ level. Future more accurate measurements will be important to further constrain non standard scenarios.

7. The dependence of neutrino fluxes on the central temperature of the Sun

Even a small modification of the central temperature of the Sun reflects into large variations of solar neutrino fluxes. The dependence of solar neutrino fluxes on the core temperature of the Sun was recently re-analyzed in Ref. 77 (see also Refs. 78 and 74 for previous discussions on the subject) where an analytical model that helps understanding the possibility they offer for determining physical characteristics of the solar interior was also provided. We summarize the main results of Ref. 77.

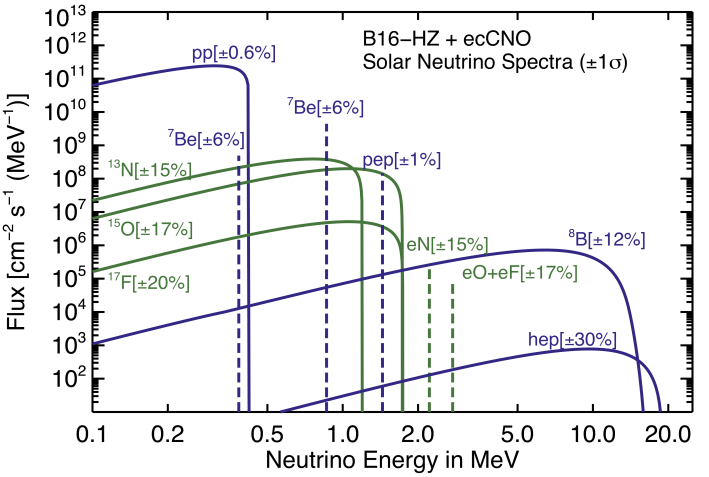


Fig. 3. The solar neutrino spectrum.

The pp-neutrino flux:
The flux $\Phi(\mathrm{pp})$ is basically fixed by the solar luminosity constraint and it is expected to be mildly dependent on the central temperature of the Sun and nuclear reaction cross sections. By taking into account that the two largest contributions to the r.h.s of Eq. 6 are provided by pp and ^{7}Be-neutrinos, one arrives at the conclusions that $\delta\Phi(\mathrm{pp}) \simeq -\eta\,\delta\Phi(^7\mathrm{Be})$, where $\delta\Phi(i)$ represent the fractional variation of the i-neutrino flux and $\eta \equiv \Phi(^7\mathrm{Be})/\Phi(\mathrm{pp}) \simeq 0.08$. By considering the dependence of $\Phi(^7\mathrm{Be})$ on solar central temperature and nuclear cross sections (discussed below), one obtains the following relationship

$$\delta\Phi(\mathrm{pp}) = -\eta\,\delta S_{34} - \frac{\eta}{2}\,(\delta S_{11} - \delta S_{33}) + \beta_{\mathrm{pp}}\,\delta T_{\mathrm{c}} \tag{7}$$

where $\beta_{\mathrm{pp}} = -\eta\beta_{\mathrm{Be}} \simeq -0.9$, δT_{c} is the fractional variation of the core temperature and δS_{ij} indicates the fractional variations of the astrophysical factor S_{ij}.

The pep-neutrino flux:
The pep-neutrinos are produced by electron capture reaction $\mathrm{p}(\mathrm{p}e^-,\nu_e)\mathrm{d}$ which is linked to the β-decay process $\mathrm{p}(\mathrm{p},e^+\nu_e)\mathrm{d}$ by well-known nuclear physics. The ratio between their rates is roughly proportional to $T_{\mathrm{c}}^{-1/2}\,n_e$ (see e.g. Ref. 27 for a review). We can thus assume $\Phi(\mathrm{pep}) \propto T_{\mathrm{c}}^{1/2}\,\Phi(\mathrm{pp})$, allowing us to conclude:

$$\delta\Phi(\mathrm{pep}) = -\eta\,\delta S_{34} - \frac{\eta}{2}\,(\delta S_{11} - \delta S_{33}) + \beta_{\mathrm{pep}}\,\delta T_{\mathrm{c}} \tag{8}$$

where $\beta_{\mathrm{pep}} = \beta_{\mathrm{pp}} - 1/2 \simeq -1.4$.

The ^{7}Be-neutrino flux:
The Be-neutrino flux is linked to the rate of the $^3\mathrm{He}(^4\mathrm{He},\gamma)^7\mathrm{Be}$ reaction, since the electron capture reaction $^7\mathrm{Be}(e^-,\nu_e)^7\mathrm{Li}$ largely dominates over the competing

proton-capture reaction. By considering that ^{3}He abundance in a large part of the solar core can be estimated by using the equilibrium condition $X_3 \simeq X_{3,\rm eq} \propto (\langle\sigma v\rangle_{11}/\langle\sigma v\rangle_{33})^{1/2}$, one obtains the expression:

$$\delta\Phi(^7\mathrm{Be}) = \delta S_{34} + \frac{1}{2}\,(\delta S_{11} - \delta S_{33}) + \beta_{\rm Be}\,\delta T_{\rm c} \tag{9}$$

where $\beta_{\rm Be} = \gamma_{34} + (\gamma_{11} - \gamma_{33})/2 \sim 11$ and we considered that $\langle\sigma v\rangle_{ij} \propto S_{ij}\, T_{\rm c}^{\gamma_{ij}}$.

The ^{8}B-neutrino flux:
The ^{8}B neutrinos constitute a largely subdominant component of the solar flux which is produced when ^{7}Be nuclei capture a proton (instead of an electron) producing ^{8}B (instead of ^{7}Li). The ^{8}B-neutrino flux is thus given by $\Phi(^8\mathrm{B}) = r\,\Phi(^7\mathrm{Be})$ where r is the ratio between proton and electron capture rates on ^{7}Be. The parameter r scales as $r \propto (S_{17}/S_{e7}) \cdot T_{\rm c}^{\alpha}$ where $\alpha = \gamma_{17} + (1/2)$ and we have considered that $\langle\sigma v\rangle_{e7} \propto S_{e7}\, T_{\rm c}^{-1/2}$ for electron capture reaction. Taking this into account, we obtain the following equation:

$$\delta\Phi(^8\mathrm{B}) = (\delta S_{17} - \delta S_{e7}) + \delta S_{34} + \frac{1}{2}\,(\delta S_{11} - \delta S_{33}) + \beta_{\rm B}\,\delta T_{\rm c} \tag{10}$$

with $\beta_{\rm B} = \beta_{\rm Be} + \gamma_{17} + 1/2 \simeq 24$. The large value of $\beta_{\rm B}$ indicates that ^{8}B neutrinos are a very sensitive probe of the core temperature of the Sun.

The CNO neutrino fluxes:
The neutrino fluxes produced in the CN-cycle by β-decay (and electron capture reactions) of ^{13}N and ^{15}O nuclei, besides depending on the solar central temperature, are approximately proportional to the stellar-core number abundance of CN elements. This dependence is extremely important because, as it is discussed in Refs. 16, 17, 79, it permits us to use CNO neutrinos, in combination with other neutrino fluxes, to directly probe the chemical composition of the Sun. By considering an analytical model to calculate the abundance of carbon and nitrogen in the core of the Sun, Ref. 77 obtained the following expressions;

$$\delta\Phi(^{15}\mathrm{O}) = \beta_{\rm O}\,\delta T_{\rm c} + (1-a)\,\delta X_{12,\,\rm s} + a\,\delta X_{14,\,\rm s} + b\,\left(\Delta^{(\rm cs)} - 0.16\right) + \delta S_{114} \tag{11}$$

$$\delta\Phi(^{13}\mathrm{N}) = \beta_{\rm N}\,\delta T_{\rm c} + (1-a')\,\delta X_{12,\,\rm s} + a'\,\delta X_{14,\,\rm s} + b\,\left(\Delta^{(\rm cs)} - 0.16\right) + f\,\delta S_{114}$$

with $\beta_{\rm O} = \gamma_{114} \simeq 20$, $f = \Phi(^{15}\mathrm{O})/\Phi(^{13}\mathrm{N}) = 0.74$, $a = 6\xi/(6\xi + 7) \simeq 0.20$ and $\xi = (X_{14,\,\rm s}/X_{12,\,\rm s}) \simeq 0.30$, $b = 0.86$, $\beta_{\rm N} \equiv f\,\beta_{\rm O} = 15$ and $a' \equiv f\,a = 0.15$. In the previous relations, $\delta X_{12,\,\rm s}$ and $\delta X_{14,\,\rm s}$ are the fractional variations of the carbon-12 and nitrogen-14 surface abundances. The quantity $\Delta^{(\rm cs)}$ takes into account the effects of elemental diffusion that enhances heavy element abundances in the solar core with respect to surface. It is defined by:

$$\mathcal{N}_{\rm c} = \mathcal{N}_{\rm s}\left[1 + \Delta^{(\rm cs)}\right] \tag{12}$$

where $\mathcal{N}_{\rm c}$ ($\mathcal{N}_{\rm s}$) is the central (surface) value of the total number abundance of CN nuclei defined as $\mathcal{N} \equiv X_{12}/12 + X_{13}/13 + X_{14}/14$. Reference SSM calculations[23] gives $\Delta^{(\rm cs)} = 0.16$ but this value can be different in non standard scenarios and/or if diffusion coefficients are modified.

8. Using CNO neutrinos to probe the solar core composition

In Refs. 16 and 17, it was suggested to combine the CN-neutrino fluxes with the boron neutrino flux that, due to the exquisite precision of current experimental results and the large temperature sensitivity can be efficiently used as solar thermometer. As can be understood by considering Eqs.(10) and (12), the following combinations can be formed:

$$\delta\Phi(^{15}\mathrm{O}) - x\,\delta\Phi(^{8}\mathrm{B}) = (1-a)\,\delta X_{12,\,\rm s} + a\,\delta X_{14,\,\rm s} + b\,\left(\Delta^{(\rm cs)} - 0.16\right) + \delta S_{114} - x\left(\frac{\delta S_{11}}{2} - \frac{\delta S_{33}}{2} + \delta S_{34} + \delta S_{17} - \delta S_{e7}\right) \quad (13)$$

$$\delta\Phi(^{13}\mathrm{N}) - x'\,\delta\Phi(^{8}\mathrm{B}) = (1-a')\,\delta X_{12,\,\rm s} + a'\,\delta X_{14,\,\rm s} + b\,\left(\Delta^{(\rm cs)} - 0.16\right) + \delta S_{114} - x'\left(\frac{\delta S_{11}}{2} - \frac{\delta S_{33}}{2} + \delta S_{34} + \delta S_{17} - \delta S_{e7}\right) \quad (14)$$

where $x = \beta_{\rm O}/\beta_{\rm B} \simeq 0.8$ and $x' = f\,x \simeq 0.6$, that are independent from $\delta T_{\rm c}$. This possibility is extremely important because it allows us to cancels out the dependence on the radiative opacity (implicit in $\delta T_{\rm c}$). The uncertainty of available opacity calculations is indeed not easily quantified and may be potentially underestimated. Moreover, it breaks the degeneracy between composition and opacity effects on solar observable properties. Indeed, the considered flux combinations only depend on the carbon and nitrogen abundance in the solar core allowing us to test the chemical composition and evolution of the Sun. The first two terms in the r.h.s. of Eqs. (13,14) quantify the effects of a variation of the surface C and N abundances. A change of the diffusion efficiency is instead described in terms of a variation of $\Delta^{(\rm cs)}$ from the SSM value, i.e. by assuming $\Delta^{(\rm cs)} - 0.016 \neq 0$. It should be remarked that the ability to probe solar composition by using this approach is only limited by experimental accuracy of flux determinations and by nuclear cross section uncertainties.

While the above relationships are based on the simplified arguments discussed in the previous section, the optimal combinations $\delta\Phi(\nu_1) - x_{12}\delta\Phi(\nu_2)$, or equivalently weighted ratios $\Phi(\nu_1)/\Phi(\nu_2)^{x_{12}}$, can be determined by using the power-law coefficients $\alpha(\nu, I)$ given in Ref. 23. The parameter x_{12} is obtained by minimizing the residual

$$\rho = \sum_{I=1}^{N} \left[\alpha(\nu_1, I) - x_{12}\,\alpha(\nu_1, I)\right]^2 \sigma_I^2 \quad (15)$$

where the sum extends to the N input parameters whose dependence we want to cancel out and σ_I are the corresponding uncertainties. The minimal value for ρ gives

the intrinsic error in the considered approach. This method, originally proposed by Refs. 16 and 17, has been recently adapted to Borexino in Ref. 80. By taking into account that the measured CNO neutrino signal in Borexino is basically probing $\delta\phi^{\rm BX}_{\rm CNO} \equiv \xi\,\delta\Phi(^{15}{\rm O}) + (1-\xi)\,\delta\Phi(^{13}{\rm N})$ with $\xi = 0.764$, it was concluded that the surface composition of the Sun can be probed by the combination:

$$\delta R^{\rm BX}_{\rm CNO} - 0.716\,\delta\Phi(^8{\rm B}) = 0.814\,\delta X_{12,\,\rm s} + 0.191\,\delta X_{14,\,\rm s} \pm 0.5\%\,({\rm env}) \pm 9.1\%\,({\rm nucl}) \pm 2.8\%\,({\rm diff}) \qquad (16)$$

where $\delta R^{\rm BX}_{\rm CNO}$ is the fractional difference of the observed CNO signal with respect to SSM expectations and the quoted uncertainties are obtained by propagating errors of SSM input parameters. The error budget is presently dominated by the uncertainty of the CNO signal Borexino measurement. However, a relevant error ($\sim 10\%$) is also provided by nuclear reactions, with the largest contributions coming from S_{114} (7.6%), S_{34} (3.4%), and S_{17} (3.5%). In the perspective of future improvements of the CNO signal determination, it is evidently important to have reliable and accurate determinations of these cross sections.

9. Final remarks

The standard solar model establishes a well-defined framework for studying solar interior physics. Despite all its success, still faces difficulties, mainly the solar abundance problem, as diagnosed by helioseismic constraints. The advancements in solar neutrino experiments, and in particular those by Borexino in the last 10 years, have added solar neutrinos as an independent probe of the solar interior properties, complementary to helioseismology. In addition to the now seemingly routine accurate measurement of the ^{7}Be flux, Borexino has opened up the possibilities of quantitative and model independent tests of the origin of solar luminosity and has detected solar CN-neutrinos for the first time (the first direct experimental evidence of the CN-cycle operating in stars). Future improvements in the CN flux measurement by Borexino might help in the quest for the solar abundances, a fundamental open problem in solar physics. This is a prerequisite for developing robust and physically sound solar models beyond the standard model paradigm. Non-standard solar physics requires unfortunately the introduction of free parameters (e.g. to treat non-standard mixing processes) that can be calibrated robustly only once the standard solar model is set on firmer grounds.

Finally, we would like to express that the value of Borexino for solar physics does not only lie in the direct results obtained by the experiment, but also in the development of experimental techniques, which represent a very valuable legacy for future neutrino experiments.

References

1. J. N. Bahcall, M. H. Pinsonneault and G. J. Wasserburg, Solar models with helium and heavy-element diffusion, *Reviews of Modern Physics* **67**, 781 (October 1995).

2. J. Christensen-Dalsgaard, W. Dappen, S. V. Ajukov, E. R. Anderson, H. M. Antia, S. Basu, V. A. Baturin, G. Berthomieu, B. Chaboyer, S. M. Chitre, A. N. Cox, P. Demarque, J. Donatowicz, W. A. Dziembowski, M. Gabriel, D. O. Gough, D. B. Guenther, J. A. Guzik, J. W. Harvey, F. Hill, G. Houdek, C. A. Iglesias, A. G. Kosovichev, J. W. Leibacher, P. Morel, C. R. Proffitt, J. Provost, J. Reiter, J. Rhodes, E. J., F. J. Rogers, I. W. Roxburgh, M. J. Thompson and R. K. Ulrich, The Current State of Solar Modeling, *Science* **272**, 1286 (May 1996).
3. D. O. Gough, A. G. Kosovichev, J. Toomre, E. Anderson, H. M. Antia, S. Basu, B. Chaboyer, S. M. Chitre, J. Christensen-Dalsgaard, W. A. Dziembowski, A. Eff-Darwich, J. R. Elliott, P. M. Giles, P. R. Goode, J. A. Guzik, J. W. Harvey, F. Hill, J. W. Leibacher, M. J. P. F. G. Monteiro, O. Richard, T. Sekii, H. Shibahashi, M. Takata, M. J. Thompson, S. Vauclair and S. V. Vorontsov, The Seismic Structure of the Sun, *Science* **272**, 1296 (May 1996).
4. Y. Elsworth, R. Howe, G. R. Isaak, C. P. McLeod and R. New, Evidence from solar seismology against non-standard solar-core models, *Nature* **347**, 536 (October 1990).
5. Y. Fukuda *et al.*, Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days, *Phys. Rev. Lett.* **81**, 1158 (1998), [Erratum: Phys.Rev.Lett. 81, 4279 (1998)].
6. S. Fukuda *et al.*, Solar B-8 and hep neutrino measurements from 1258 days of Super-Kamiokande data, *Phys. Rev. Lett.* **86**, 5651 (2001).
7. B. Cleveland, T. Daily, J. Davis, Raymond, J. R. Distel, K. Lande, C. Lee, P. S. Wildenhain and J. Ullman, Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *Astrophys. J.* **496**, 505 (1998).
8. W. Hampel *et al.*, GALLEX solar neutrino observations: Results for GALLEX IV, *Phys. Lett. B* **447**, 127 (1999).
9. J. N. Abdurashitov, V. N. Gavrin, S. V. Girin, V. V. Gorbachev, T. V. Ibragimova, A. V. Kalikhov, N. G. Khairnasov, T. V. Knodel, I. N. Mirmov, A. A. Shikhin, E. P. Veretenkin, V. M. Vermul, V. E. Yants, G. T. Zatsepin, T. J. Bowles, W. A. Teasdale, D. L. Wark, M. L. Cherry, J. S. Nico, B. T. Cleveland, R. Davis, K. Lande, P. S. Wildenhain, S. R. Elliott and J. F. Wilkerson, Measurement of the solar neutrino capture rate with gallium metal, *Phys. Rev. C* **60**, p. 055801 (1999).
10. Q. R. Ahmad *et al.*, Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **87**, p. 071301 (2001).
11. Q. R. Ahmad *et al.*, Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **89**, p. 011301 (2002).
12. C. Arpesella *et al.*, Direct Measurement of the Be-7 Solar Neutrino Flux with 192 Days of Borexino Data, *Phys. Rev. Lett.* **101**, p. 091302 (2008).
13. M. Agostini *et al.*, Comprehensive measurement of *pp*-chain solar neutrinos, *Nature* **562**, 505 (2018).
14. B. Aharmim *et al.*, Combined Analysis of all Three Phases of Solar Neutrino Data from the Sudbury Neutrino Observatory, *Phys. Rev. C* **88**, p. 025501 (2013).
15. K. Abe *et al.*, Solar Neutrino Measurements in Super-Kamiokande-IV, *Phys. Rev. D* **94**, p. 052010 (2016).
16. W. C. Haxton and A. M. Serenelli, CN-Cycle Solar Neutrinos and Sun's Primordial Core Metalicity, *Astrophys. J.* **687**, 678 (2008).
17. A. Serenelli, C. Peña-Garay and W. C. Haxton, Using the standard solar model to constrain solar composition and nuclear reaction S factors, *Phys. Rev.* **D87**, p. 043001 (2013).

18. F. L. Villante, A. M. Serenelli, F. Delahaye and M. H. Pinsonneault, The chemical composition of the Sun from helioseismic and solar neutrino data, *Astrophys. J.* **787**, p. 13 (2014).
19. M. Agostini *et al.*, Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun, *Nature* **587**, 577 (2020).
20. J. N. Bahcall, A. M. Serenelli and S. Basu, 10,000 standard solar models: A Monte Carlo simulation, *Astrophys. J. Suppl.* **165**, 400 (2006).
21. J. N. Bahcall and M. H. Pinsonneault, Standard solar models, with and without helium diffusion, and the solar neutrino problem, *Reviews of Modern Physics* **64**, 885 (October 1992).
22. J. Christensen-Dalsgaard, C. R. Proffitt and M. J. Thompson, Effects of Diffusion on Solar Models and Their Oscillation Frequencies, *Astrophysical Journal Letters* **403**, p. L75 (February 1993).
23. N. Vinyoles, A. M. Serenelli, F. L. Villante, S. Basu, J. Bergström, M. C. Gonzalez-Garcia, M. Maltoni, C. Peña-Garay and N. Song, A new Generation of Standard Solar Models, *Astrophys. J.* **835**, p. 202 (2017).
24. S. Cassisi, M. Salaris and A. W. Irwin, The initial helium content of galactic globular cluster stars from the r-parameter: Comparison with the cmb constraint, *Astrophys. J.* **588**, p. 862 (2003).
25. N. R. Badnell, M. A. Bautista, K. Butler, F. Delahaye, C. Mendoza, P. Palmeri, C. J. Zeippen and M. J. Seaton, Up-dated opacities from the Opacity Project, *Mon. Not. Roy. Astron. Soc.* **360**, 458 (2005).
26. J. W. Ferguson, D. R. Alexander, F. Allard, T. Barman, J. G. Bodnarik, P. H. Hauschildt, A. Heffner-Wong and A. Tamanai, Low temperature opacities, *Astrophys. J.* **623**, 585 (2005).
27. E. G. Adelberger, A. García, R. G. H. Robertson, K. A. Snover, A. B. Balantekin, K. Heeger, M. J. Ramsey-Musolf, D. Bemmerer, A. Junghans, C. A. Bertulani, J. W. Chen, H. Costantini, P. Prati, M. Couder, E. Uberseder, M. Wiescher, R. Cyburt, B. Davids, S. J. Freedman, M. Gai, D. Gazit, L. Gialanella, G. Imbriani, U. Greife, M. Hass, W. C. Haxton, T. Itahashi, K. Kubodera, K. Langanke, D. Leitner, M. Leitner, P. Vetter, L. Winslow, L. E. Marcucci, T. Motobayashi, A. Mukhamedzhanov, R. E. Tribble, K. M. Nollett, F. M. Nunes, T. S. Park, P. D. Parker, R. Schiavilla, E. C. Simpson, C. Spitaleri, F. Strieder, H. P. Trautvetter, K. Suemmerer and S. Typel, Solar fusion cross sections II: The pp chain and CNO cycles, *Rev. Mod. Phys.* **83**, p. 195 (2011).
28. L. E. Marcucci, R. Schiavilla and M. Viviani, Proton-Proton Weak Capture in Chiral Effective Field Theory, *Phys. Rev. Lett.* **110**, p. 192503 (2013).
29. E. Tognelli, S. Degl'Innocenti, L. E. Marcucci and P. G. Prada Moroni, Astrophysical implications of the proton–proton cross section updates, *Phys. Lett.* **B742**, 189 (2015).
30. B. Acharya, B. D. Carlsson, A. Ekström, C. Forssén and L. Platter, Uncertainty quantification for proton–proton fusion in chiral effective field theory, *Phys. Lett.* **B760**, 584 (2016).
31. X. Zhang, K. M. Nollett and D. R. Phillips, Halo effective field theory constrains the solar ^{7}Be + p $\rightarrow$ ^{8}B + gamma rate, *Phys. Lett.* **B751**, 535 (2015).
32. M. Marta, A. Formicola, D. Bemmerer, C. Broggini, A. Caciolli, P. Corvisiero, H. Costantini, Z. Elekes, Z. Fülöp, G. Gervino, A. Guglielmetti, C. Gustavino, G. Gyürky, G. Imbriani, M. Junker, A. Lemut, B. Limata, C. Mazzocchi, R. Menegazzo, P. Prati, V. Roca, C. Rolfs, C. Rossi Alvarez, E. Somorjai, O. Straniero, F. Strieder, F. Terrasi, H. P. Trautvetter and A. Vomiero, The ^{14}N(p,γ)^{15}O reaction studied with a composite germanium detector, *Phys. Rev.* **C83**, p. 045804 (2011).

33. A. A. Thoul, J. N. Bahcall and A. Loeb, Element diffusion in the solar interior, *Astrophys. J.* **421**, 828 (1994).
34. R. Kippenhahn and A. Weigert, *Stellar Structure and Evolution* 1990.
35. K. S. Krishna Swamy, Profiles of Strong Lines in K-Dwarfs, *Astrophys. J.* **145**, p. 174 (July 1966).
36. N. Grevesse and A. Noels, Cosmic abundances of the elements., in *Origin and Evolution of the Elements*, eds. N. Prantzos, E. Vangioni-Flam and M. Casse, 1993.
37. N. Grevesse and A. J. Sauval, Standard Solar Composition, *Space Sci. Rev.* **85**, 161 (1998).
38. M. Asplund, N. Grevesse, A. J. Sauval and P. Scott, The chemical composition of the Sun, *Ann. Rev. Astron. Astrophys.* **47**, 481 (2009).
39. E. Caffau, H.-G. Ludwig, M. Steffen, B. Freytag and P. Bonifacio, Solar Chemical Abundances Determined with a CO5BOLD 3D Model Atmosphere, *Solar Phys.* **268**, p. 255 (2011).
40. P. Scott, N. Grevesse, M. Asplund, A. J. Sauval, K. Lind, Y. Takeda, R. Collet, R. Trampedach and W. Hayek, The elemental composition of the Sun I. The intermediate mass elements Na to Ca, *Astron. Astrophys.* **573**, p. A25 (2015).
41. P. Scott, M. Asplund, N. Grevesse, M. Bergemann and A. J. Sauval, The elemental composition of the Sun II. The iron group elements Sc to Ni, *Astron. Astrophys.* **573**, p. A26 (2015).
42. N. Grevesse, P. Scott, M. Asplund and A. J. Sauval, The elemental composition of the Sun III. The heavy elements Cu to Th, *Astron. Astrophys.* **573**, p. A27 (2015).
43. M. Asplund, A. M. Amarsi and N. Grevesse, The chemical make-up of the Sun: A 2020 vision, *Astronomy & Astrophysics* **653**, p. A141 (September 2021).
44. M. Bergemann, R. Hoppe, E. Semenova, M. Carlsson, S. A. Yakovleva, Y. V. Voronov, M. Bautista, A. Nemer, A. K. Belyaev, J. Leenaarts, L. Mashonkina, A. Reiners and M. Ellwarth, Solar oxygen abundance, *Monthly Notices of the Royal Astronomical Society* **508**, 2236 (December 2021).
45. K. Lodders, H. Palme and H.-P. Gail, Abundances of the Elements in the Solar System, *Landolt Börnstein* , p. 712 (2009).
46. S. Basu and H. M. Antia, Constraining solar abundances using helioseismology, *Astrophys. J.* **606**, p. L85 (2004).
47. J. N. Bahcall, S. Basu, M. Pinsonneault and A. M. Serenelli, Helioseismological implications of recent solar abundance determinations, *Astrophys. J.* **618**, 1049 (2005).
48. F. Delahaye and M. Pinsonneault, The solar heavy element abundances. 1. Constraints from stellar interiors, *Astrophys. J.* **649**, 529 (2006).
49. J. A. Guzik, L. Watson and A. N. Cox, Can enhanced diffusion improve helioseismic agreement for solar models with revised abundances?, *Astrophys. J.* **627**, 1049 (2005).
50. M. Castro, S. Vauclair and O. Richard, Low abundances of heavy elements in the solar outer layers: Comparisons of solar models with helioseismic inversions, *Astron. Astrophys.* **463**, 755 (2007).
51. S. Basu and H. M. Antia, Helioseismology and Solar Abundances, *Phys. Rept.* **457**, 217 (2008).
52. J. A. Guzik and K. Mussack, Exploring mass loss, low-Z accretion, and convective overshoot in solar models to mitigate the solar abundance problem, *Astrophys. J.* **713**, 1108 (2010).
53. A. M. Serenelli, W. C. Haxton and C. Pena-Garay, Solar models with accretion. I. Application to the solar abundance problem, *Astrophys. J.* **743**, p. 24 (2011).
54. S. Basu and H. M. Antia, Seismic measurement of the depth of the solar convection zone, *Mon. Not. Roy. Astron. Soc.* **287**, 189 (May 1997).

55. F. L. Villante, Constraints on the opacity profile of the sun from helioseismic observables and solar neutrino flux measurements, *Astrophys. J.* **724**, 98 (2010).
56. F. L. Villante and B. Ricci, Linear Solar Models, *Astrophys. J.* **714**, 944 (2010).
57. J. N. Bahcall, A. M. Serenelli and S. Basu, New solar opacities, abundances, helioseismology, and neutrino fluxes, *Astrophys. J. Lett.* **621**, L85 (2005).
58. J. Montalban, A. Miglio, A. Noels, N. Grevesse and M. Di Mauro, Solar model with CNO revised abundances, *ESA Spec. Publ.* **559**, p. 574 (2004).
59. J. Christensen-Dalsgaard, M. P. Di Mauro, G. Houdek and F. Pijpers, On the opacity change required to compensate for the revised solar composition, *Astron. Astrophys.* **494**, p. 205 (2009).
60. C. Blancard, P. Cossé and G. Faussurier, Solar Mixture Opacity Calculations Using Detailed Configuration and Level Accounting Treatments, *Astrophys. J.* **745**, p. 10 (January 2012).
61. G. Mondet, C. Blancard, P. Cossé and G. Faussurier, Opacity Calculations for Solar Mixtures, *Astrophys. J. Supp.* **220**, p. 2 (September 2015).
62. M. Krief, A. Feigel and D. Gazit, Solar opacity calculations using the super-transition-array method, *Astrophys. J.* **821**, p. 45 (2016).
63. J. Colgan, D. P. Kilcrease, N. H. Magee, M. E. Sherrill, J. Abdallah, J., P. Hakel, C. J. Fontes, J. A. Guzik and K. A. Mussack, A New Generation of Los Alamos Opacity Tables, *Astrophys. J.* **817**, p. 116 (February 2016).
64. J. E. Bailey, T. Nagayama, G. P. Loisel, G. A. Rochau, C. Blancard, J. Colgan, P. Cosse, G. Faussurier, C. J. Fontes, F. Gilleron, I. Golovkin, S. B. Hansen, C. A. Iglesias, D. P. Kilcrease, J. J. Macfarlane, R. C. Mancini, S. N. Nahar, C. Orban, J.-C. Pain, A. K. Pradhan, M. Sherrill and B. G. Wilson, A higher-than-predicted measurement of iron opacity at solar interior temperatures, *Nature* **517**, 56 (January 2015).
65. T. Nagayama, J. E. Bailey, G. P. Loisel, G. S. Dunham, G. A. Rochau, C. Blancard, J. Colgan, P. Cossé, G. Faussurier, C. J. Fontes, F. Gilleron, S. B. Hansen, C. A. Iglesias, I. E. Golovkin, D. P. Kilcrease, J. J. MacFarlane, R. C. Mancini, R. M. More, C. Orban, J. C. Pain, M. E. Sherrill and B. G. Wilson, Systematic Study of L -Shell Opacity at Stellar Interior Temperatures, *Physics Review Letters* **122**, p. 235001 (June 2019).
66. M. Krief, A. Feigel and D. Gazit, Line Broadening and the Solar Opacity Problem, *Astrophys. J.* **824**, p. 98 (June 2016).
67. J. N. Bahcall, Line versus continuum solar neutrinos, *Phys. Rev.* **D41**, p. 2964 (1990).
68. L. C. Stonehill, J. A. Formaggio and R. G. H. Robertson, Solar neutrinos from CNO electron capture, *Phys. Rev.* **C69**, p. 015801 (2004).
69. F. L. Villante, ecCNO Solar Neutrinos: A Challenge for Gigantic Ultra-Pure Liquid Scintillator Detectors, *Phys. Lett.* **B742**, 279 (2015).
70. J. Bergstrom, M. C. Gonzalez-Garcia, M. Maltoni, C. Pena-Garay, A. M. Serenelli and N. Song, Updated determination of the solar neutrino fluxes from solar neutrino data, *JHEP* **03**, p. 132 (2016).
71. G. Bellini *et al.*, First evidence of pep solar neutrinos by direct detection in Borexino, *Phys. Rev. Lett.* **108**, p. 051302 (2012).
72. G. Bellini *et al.*, Neutrinos from the primary proton–proton fusion process in the Sun, *Nature* **512**, 383 (2014).
73. J. N. Bahcall, The Luminosity constraint on solar neutrino fluxes, *Phys. Rev.* **C65**, p. 025801 (2002).
74. S. Degl'Innocenti, W. A. Dziembowski, G. Fiorentini and B. Ricci, Helioseismology and standard solar models, *Astropart. Phys.* **7**, 77 (1997).

75. F. Vissani, Luminosity constraint and entangled solar neutrino signals, in *5th International Solar Neutrino Conference*, 2019.
76. D. Vescovi, C. Mascaretti, F. Vissani, L. Piersanti and O. Straniero, The luminosity constraint in the era of precision solar physics, *J. Phys. G* **48**, p. 015201 (2020).
77. F. L. Villante and A. Serenelli, The relevance of nuclear reactions for Standard Solar Models construction, *Front. Astron. Space Sci.* **7**, p. 112 (2021).
78. J. N. Bahcall and A. Ulmer, The Temperature dependence of solar neutrino fluxes, *Phys. Rev.* **D53**, 4202 (1996).
79. W. C. Haxton, R. G. Hamish Robertson and A. M. Serenelli, Solar Neutrinos: Status and Prospects, *Ann. Rev. Astron. Astrophys.* **51**, 21 (2013).
80. M. Agostini *et al.*, Sensitivity to neutrinos from the solar CNO cycle in Borexino, *Eur. Phys. J. C* **80**, p. 1091 (2020).

Role of the CNO cycles in stars

A. Ianni

I.N.F.N. Laboratori Nazionali del Gran Sasso,
Via Giovanni Acitelli 22,
67100 Assergi (AQ), Italy
E-mail: aldo.ianni@lngs.infn.it

Borexino has recently observed CNO solar neutrinos. This measurement confirms an energy production mechanism in stars predicted about a century ago. The CNO cycle in the Sun is sub-dominant with respect to the pp-chain energy production. However, it is definitely important in more massive stars. We describe the main characteristics of the CNO cycle in the Sun and in massive stars.

Keywords: Solar neutrinos; CNO cycle; stars; energy production in stars.

1. Energy production in stars and the role of the CNO bi-cycle

The fundamental paradigm for energy production in the Sun and similar stars is the transformation of hydrogen into helium through the process: $4p \rightarrow^4 He + 2e^+ + 2\nu_e + 26.73\,\mathrm{MeV}$.[1] This mechanism makes electron neutrinos of mainly sub-MeV energy. Due to the fact that neutrinos can travel through matter affected only by weak interactions, the observation of solar neutrinos allows to probe the interior of the Sun. Therefore, the Sun becomes a laboratory to understand the physics in stars. Solar neutrino measurements have been carried out since 1968.[2]

The main source of energy in the Sun is the so-called pp-chain shown in Figure 1. The pp-chain has three terminations. Each termination produces 26.2 MeV effective thermal energy and it burns 3.7×10^{38} hydrogen/sec, which corresponds to 612 ton/sec. Therefore, assuming 10% of solar mass involved in energy production, the timescale of the pp-chain for the Sun is of the order of 10^{10} years. This mechanism is dominant in first generation stars. In second or third generation stars a different mechanism is also at work. This latter involves light elements, such as carbon and nitrogen, generated after the production of helium. The idea of a second energy production mechanism was independently introduced by von Weizsaker and Bethe between 1937[3] and 1939.[4] Second and third generation stars contain some "heavy" elements, such as nitrogen and carbon. These elements allow the transformation of hydrogen into helium. The process can start from the reaction $p + {}^{12}_{6}\mathrm{C} \rightarrow {}^{13}_{7}\mathrm{N} + \gamma$, which is followed by the beta decay of ${}^{13}_{7}\mathrm{N}$ to ${}^{13}_{6}\mathrm{C}$ accompanied with the emission of an electron neutrino of maximum energy equal to 1.2 MeV. Afterward ${}^{14}_{7}\mathrm{N}$ is formed by $p + {}^{13}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + \gamma$. ${}^{14}_{7}\mathrm{N}$ allows the production of ${}^{15}_{8}\mathrm{O}$, which decays beta to ${}^{15}_{7}\mathrm{N}$ with the emission of another electron neutrino of maximum energy equal to 1.7 MeV. At this point ${}^{15}_{7}\mathrm{N}$ allows the production of ${}^{12}_{6}\mathrm{C}$ through the process

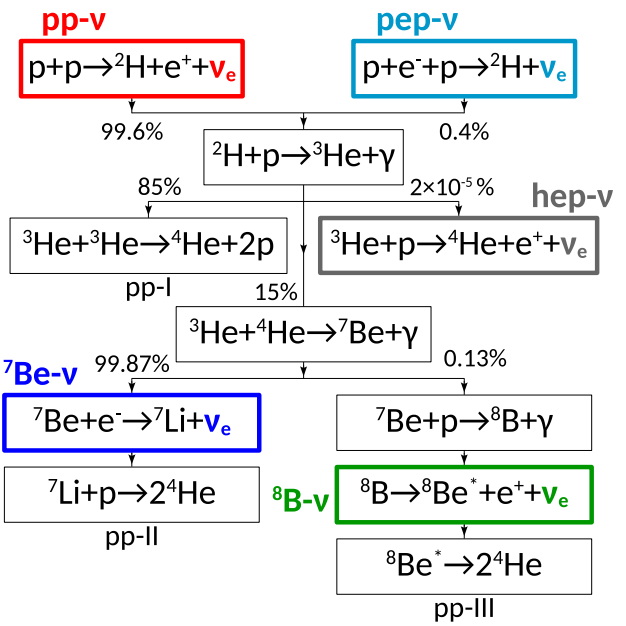


Fig. 1. Set of reactions in the pp-chain, which is the main source of energy in the Sun and in low mass H-burning stars with mass smaller than $1.3 M_{\odot}$.

$p + {}^{15}_{7}\mathrm{N} \rightarrow {}^{12}_{6}\mathrm{C} + \alpha$. With ${}^{12}_{6}\mathrm{C}$ the sequence of reactions can start over again. This so-called CN cycle is shown in Figure 2. ${}^{12}_{6}\mathrm{C}$ is used as a catalyst. This cycle produces two electron neutrinos, helium, and the same amount of energy produced by the pp-chain. The CN cycle till 1950 was considered the main source of energy in the Sun. The slowest reaction in the CN cycle is ${}^{14}_{7}\mathrm{N}(p,\gamma){}^{15}_{8}\mathrm{O}$, which has the highest Coulomb barrier.

The CN cycle might develop into a CNO bi-cycle as show in Figure 2. As a matter of fact, with a probability of order 10^{-3}, ${}^{16}_{8}\mathrm{O}$ can be produced by ${}^{15}_{7}\mathrm{N}(p,\gamma){}^{16}_{8}\mathrm{O}$. This second branch produces ${}^{17}_{9}\mathrm{F}$, which decays beta with the emission of an electron neutrino with maximum energy equal to 1.74 MeV. The CNO bi-cycle produces three electron neutrinos from beta decays of ${}^{13}\mathrm{N}$, ${}^{15}\mathrm{O}$ and ${}^{17}\mathrm{F}$. These are the so-called CNO solar neutrinos. The flux of these neutrinos is related to the abundance of these elements in the Sun's core. Therefore, a measurement of CNO solar neutrinos is a probe of the Sun's metallicity.

Measurements of stellar interaction rates have shown that the CN cycle is not dominant in the Sun. This information is used by the Solar Standard Model (SSM),[1] which is the theoretical framework to describe the evolution of the Sun. According to the SSM the CNO bi-cycle in the Sun is responsible for only 1% of the energy production. The CNO bi-cycle is important for the nucleosynthesis of ${}^{16}_{8}\mathrm{O}$, ${}^{17}_{8}\mathrm{O}$, and other light elements. The CNO bi-cycle is also referred to as the "cold" CNO cycle[5] with respect to the "hot" cycle which is discussed in the next Section.

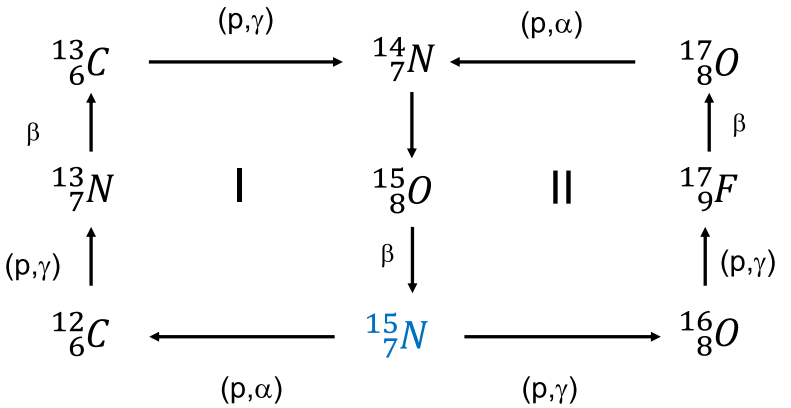


Fig. 2. The CNO bi-cycle divided in its two sub-cycles. The dominant one in the Sun is the CN cycle on the left side.

2. The CNO cycle in massive stars

The Sun's core temperature is 15×10^6 K, which we write as $15T_6$. As predicted by the SSM at this temperature the energy production for the pp-chain and the CNO bi-cycle scales as $\propto T^4$ and $\propto T^8$, respectively. For stellar masses smaller than $1.3M_\odot$, the pp-chain dominates the energy production. In more massive stars with a higher core temperature, the CNO cycle is by far the dominant source of energy. In this case one has to take into account two competing processes, namely $^{17}_{8}\mathrm{O}(p,\alpha)^{14}_{7}\mathrm{N}$ and $^{17}_{8}\mathrm{O}(p,\gamma)^{18}_{9}\mathrm{F}$. For some temperature ranges the reaction rates for these processes are comparable.[6] In this case the CNO cycle is tri-cycling as shown in Figure 3.[7] In addition, for temperature ranging up to 10^{10} K the interaction rate for $^{18}_{8}\mathrm{O}(p,\alpha)^{15}_{7}\mathrm{N}$ and $^{18}_{8}\mathrm{O}(p,\gamma)^{19}_{9}\mathrm{F}$ is of order 100 with a flat minimum between T_9 and $7T_9$.[5] Therefore, with a probability of order 1% a fourth CNO cycle can develop. This fourth cycle is also shown in Figure 3. The third and fourth CNO cycles are referred to as the "hot" CNO cycles. It turns out that when $^{19}_{9}\mathrm{F}(p,\gamma)^{20}_{10}\mathrm{Ne}$ is dominat the CNO catalytic cycling stops.

One fundamental empirical properties of stars is the relationship between their total mass and surface luminosity:[8] $L/L_\odot = (M/M_\odot)^\alpha$, with $3 \lesssim \alpha \lesssim 4$[9] and $L_\odot$ and $M_\odot$ the solar surface luminosity and mass, respectively. This property is confirmed by basic stellar theory. One can make some general arguments for the role of the CNO cycle in stars in the rage of masses $[1,10]M_\odot$. In this range it turns out that the luminosity scales as $\propto M^{\sim 4}$.[9] This implies that more massive stars have a smaller lifetime, for the lifetime scales as $\propto M^{-3}$. In addition, considering that the stars number density scales as $\propto M^{-2.5}$,[10] the energy production in a given bin of mass $[M_i, M_{i+1}]$ is written as:

$$\epsilon \propto \int_{M_i}^{M_{i+1}} M^{1.5} dM \tag{1}$$

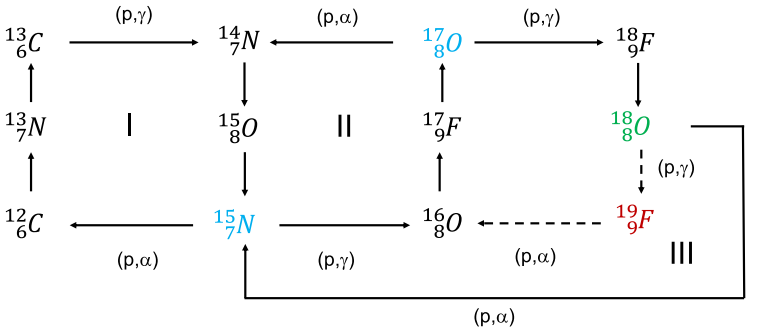


Fig. 3. The CNO tri-cycle in massive stars.

This implies that stars where the CNO energy production is dominant, namely $M > 1.3M_{\odot}$, are producing about 100 times more energy in a much shorter timescale.

3. The role of Borexino

In 2020 Borexino has reported the first ever observation of CNO neutrinos from the Sun.[11] The null hypothesis (no CNO neutrinos) has been rejected at 5σ level. Indeed, Borexino has shown that this energy production mechanism predicted about a century ago is at work in stars. Although the accuracy of the measurement at present is still large ($\sim 35\%$), this measurement allows to probe the SSM in terms of neutrino emission, metallicity, and energy production. The solar luminosity, $L_{\odot}$ can be related to the solar neutrino fluxes:[12,13]

$$\frac{L_{\odot}}{4\pi AU^2} = \sum_i \alpha_i \phi_i \tag{2}$$

where AU is one astronomical unit, ϕ_i are the neutrino fluxes, and α_i are coefficients related to the energy production in the reactions taking place in the pp-chain and CNO bi-cycle, respectively. The Borexino measurement implies that:

$$\frac{L_{CNO}}{L_{\odot}} = 1.0^{+0.4}_{-0.3}\% \tag{3}$$

This is an important confirmation of the SSM prediction. Previously Borexino had measured all the main neutrinos produced by the pp-chain,[14] proving that this mechanism is the main energy production in the Sun with an experimental accuracy of the order of 10%. Therefore, at present, solar neutrino observations have demonstrated that the pp-chain and the CNO bi-cycle are producing energy in stars as it was predicted about a century ago.

References

1. J. N. Bahcall (1989). Neutrino Astrophysics. Cambridge University Press.
2. L. Oberauer, A. Ianni, A. Serenelli (2020). Solar Neutrino Physics. The interplay between particle physics and astronomy. Wiley-VCH

3. C. F. Weizsacker, On Elementary Transmutations in the Interior of Stars: Paper II, Physik. Zeit., (38), 1937.
4. H. A. Bethe., Energy production in stars, Physical Review, 55(5): 434-456, 1939.
5. M. Wiescher et al., Annu. Rev. Nucl. Part. Sci. 2010, 60: 381-404.
6. C. Rolfs and W. S. Rodney, Astrophys. J. 194: L63-L66, 1974.
7. J. J. Cowan and W. K. Rose, Astrophys. J. 201: L45-L48, 1975.
8. M. Salaris and S. Cassisi, Evolution of Stars and Stellar Populations, John Wiley and Sons, Ltd, 2005.
9. M. Harwit, Astrophysical Concepts, 2nd Ed., Springer, 1988.
10. G. Srinivasan, Neutron Stars and Black Holes (Lecture 02: White Dwarf Stars).
11. Borexino collaboration, Nature 587 (2020) 577-582.
12. J. N. Bahcall, 2002 Phy. Rev. C 65 025801.
13. D. Vescovi et al., J. Phys. G 48 (2020) 1, 015201.
14. Borexino collaboration, Nature 562 (2018) 7728, 505-510.

Geoneutrino observation

Tadao Mitsui
Research Center for Neutrino Science,
Tohoku University,
Sendai 980-8578, Japan
E-mail: mitsui@awa.tohoku.ac.jp

for the KamLAND Collaboration

Neutrinos from the Earth's interior "geoneutrinos" have been observed to reveal Earth's composition, heat budget, and the origin of the Earth. Observations by KamLAND and Borexino experiments for 15 years are reviewed.

Keywords: Neutrino; Geoneutrino

1. Geoneutrino

Recent progress on the research about "why and how the sun and the stars shine" is quite notable, which has been summarized in this parallel session. Here I would like to review neutrinos from another familiar celestial body, the Earth. Neutrinos from the Earth ("geoneutrinos") are low-energy electron antineutrinos ($\bar{\nu}_e$) emitted in the decay chains of uranium (U), thorium (Th), and potassium (K) in the Earth's interior. Geoneutrino is expected as a probe of Earth's chemical composition. Also the decay of those radioactive elements contributes to the Earth's heat production. Then geoneutrino is a tool to reveal the Earth's energy budget. Although the existence of geoneutrino is predicted more than 50 years ago[1–3] experimental research has been difficult. Since Borexino and KamLAND experiments were planned, they have been hoped as the first experimental research of geoneutrinos.

With those liquid scintillator detectors, geoneutrinos (geo-$\bar{\nu}_e$'s) are detected via "inverse β decay" mode: $\bar{\nu}_e + p \to e^+ + n$. Considering the difference of masses of neutron (m_n) and proton (m_p), $m_n - m_p = 1.3$ MeV, the mass of positron ($m_e = 0.5$ MeV), and neglecting the small neutron recoil, the threshold of this reaction is $E_{th} \simeq m_n - m_p + m_e = 1.8$ MeV, above which neutrino energy E_ν and positron energy E_{e^+} (the kinetic energy of the positron plus the electron-positron annihilation energy) are related as $E_\nu \simeq E_{e^+} + E_{th} - 2m_e = E_{e^+} + 0.8$ MeV. The potassium neutrino is below E_{th}, so liquid scintillator detectors observe only uranium and thorium neutrinos. With energy spectrum (E_{e^+} spectrum which reflects E_ν spectrum), uranium (U) and thorium (Th) neutrinos can be separately measured. To obtain total geoneutrino flux with less statistical uncertainty, the relative abundance of U and Th is sometimes assumed to be Th/U = 3.9, based on geological consideration.

As Borexino and KamLAND plans started, calculation of expected flux of geoneutrinos, and expected signals of these detectors have also been performed. One of the examples is Ref. 4, in which expected "peaks" of geoneutrinos around $E_{e+} \simeq 1$ to 3 MeV are shown. Also shown are the expected background spectra of nuclear reactor $\bar{\nu}_e$, showing that reactor $\bar{\nu}_e$ of Borexino is much less than that of KamLAND.

2. Experimental Research

In 2005, the first experimental research of geoneutrinos has been done,[5] in which KamLAND measured low-energy spectrum of $\bar{\nu}_e$ and fitted it to geoneutrino, reactor neutrino and other background spectra. As shown in Fig. 1, geoneutirno contribution is indicated although the statistical power is limited.

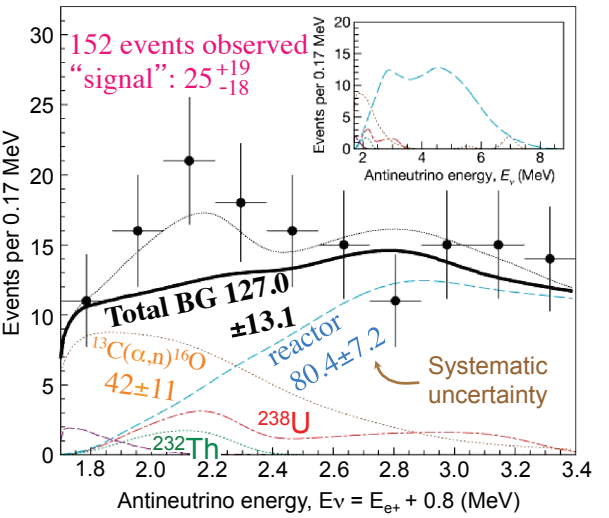


Fig. 1. First experimental research of geoneutrinos. Points are data, lines with labels and numbers are estimated background (thin line around 1.8 MeV is random coincidence). Lines labeled as ^{238}U and ^{232}Th are best-fit geoneutrino signals. Thick black line is the estimated total background, and thin black line is that with the best-fit signals added.

In 2010, Borexino collaboration reported their first result of geoneutrino observation,[6] with which the observation with two detectors on the Earth started. Borexino is an ultra pure detector and reactor neutrino flux is much less than KamLAND site. Thanks to those low background environment, the existence of geoneutrino was confirmed at 4.2 σ confidence level.

In an updated result of KamLAND,[7] it has been shown that the radiogenic heat source inferred from geoneutrino flux is significantly less than the total heat flow from the Earth's interior. With this result they concluded that there is contribution of secular cooling of the Earth to the total heat source of the Earth.

In the updated result of Borexino[8] released in 2013, the signal of geoneutrino became clearer and the probability for null geoneutrino measurement was shown to be 6×10^{-6}. The latest data of KamLAND was also shown[9] in 2013, which included the data of the period with most of Japanese nuclear reactors off. With these data, low background observation of geoneutrino was done also with KamLAND. The reactor $\bar{\nu}_e$ background was also unambiguously estimated using reactor on and off data. In those reports in 2013,[8,9] existence of mantle geoneutrino contribution was indicated by using both data in a cooperating way because crust contributions are different between KamLAND and Borexino sites.

In 2015, Borexino collaboration released new result[10] and reported that the probability for null geoneutrino measurement is 3.6×10^{-9} corresponding to 5.9 σ confidence level. This should be a conclusive result for geoneutrino existence. The latest result of Borexino was shonw in Ref. 11. This comprehensive report includes all the aspects of geoneutrino research, i.e., physics, geology, experiment, and data analysis of geoneutrino. All the researchers involved in geoneutrino study are recommended to read it carefully so that they are stimulated for new research. Also in this report, multiple Earth models are carefully tested with their latest data, including cosmochemical, geochemical, and geodynamical models.

3. Research in Geology and Theory

With those experimental data, research in geology and theory also became active. Combined analysis of KamLAND and Borexino data has been done by theorists. Ref. 12 is one of combined analysis, in which common and independent uncertainties of KamLAND and Borexino are treated properly and carefully with considering correlation.

New calculations of geoneutrino flux have also been done using up-to-date data and more precise and refined method of calculation. For example, such up-to-date "reference models" are reviewed in Refs. 11–13.

4. Future Experiments

Although the observation with two detectors has lasted for more than 10 years, new detectors will be ready and expected to start their observations. SNO+[14] is a large liquid scintillator detector in Canada to search for neutrinoless double beta decay. It can also observe geoneutrinos. Large flux of geoneutrino is expected from the continent crust. In April 2021, SNO+ completed liquid scintillator filling. We have to stay tuned.

JUNO[15] is a multipurpose neutrino detector planned to be constructed in China. Their main purpose is to measure reactor neutrino oscillation to further improve Daya Bay achievements. When completed it will be the largest liquid scintillator detector (20 kton) in the world. The construction already started.

Jinping Neutrino Experiment[16] is planned in China. When completed, it will be the deepest detector in the world, with 2400 m overburden of rocks. It aims at

observing large flux of geoneutrinos from Himalaya, which is a part of one of the largest continental crust of the present-day Earth.

For more research, especially for mantle geoneutrino detection, large liquid scintillator in the deep ocean[17] is one possibility. Directional measurement of electron antineutrinos is another future possibility in order to get more information about geoneutrino source distribution. Although the directional measurement was done by Chooz experiment,[18] application for large detector is not easy and under development. Considering the observation at a lot of points are essentially important, design of cheap detector[19] may be useful to construct as many detector as possible.

Acknowledgments

The author thanks organizers for their kind hospitality under the current unusual situation. The KamLAND experiment is supported by JSPS KAKENHI, the World Premier International Research Center Initiative (WPI Initiative), MEXT Japan, Netherlands Organization for Scientific Research (NWO), the U.S. Department of Energy (DOE), and the National Science Foundation (NSF). The Kamioka Mining and Smelting Company has provided service for activities in the mine. We acknowledge the support of NII for SINET4.

References

1. G. Eder, Terrestrial neutrinos, *Nucl. Phys.* **78** 657 (1966).
2. G. Marx, Geophysics by neutrinos, *Czech. J. Phys. B* **19** 1471 (1969).
3. G. Gamow, Letter to F. Reines (1953).
4. R. S. Raghavan *et al.*, Measuring the global radioactivity in the Earth by multidetector antineutrino spectroscopy, *Phys. Rev. Lett.* **80** 635 (1998).
5. T. Araki *et al.* (KamLAND Collab.), Experimental investigation of geologically produced antineutrinos with KamLAND, *Nature* **436** 499 (2005).
6. Borexino Collab. (G. Bellini *et al.*), Observation of geo-neutrinos, *Phys. Let. B* **687** 299 (2010).
7. KamLAND Collab. (A. Gando *et al.*), Partial radiogenic heat model for Earth revealed by geoneutrino measurements, *Nature Geosci.* **4** 647 (2011).
8. Borexino Collab. (G. Bellini *et al.*), Measurement of geo-neutrinos from 1353 days of Borexino, *Phys. Let. B* **722** 295 (2013).
9. A. Gando *et al.* (KamLAND Collab.), Reactor on-off antineutrino measurement with KamLAND, *Phys. Rev. D* **88** 033001 (2013).
10. M. Agostini *et al.* (Borexino Collab.), Spectroscopy of geoneutrinos from 2056 days of Borexino data *Phys. Rev. D* **92** 031101(R) (2015).
11. M. Agostini *et al.* (Borexino Collab.), Comprehensive geoneutrino analysis with Borexino, *Phys. Rev. D* **101** 012009 (2020).
12. G. L. Fogli, E. Lisi, A. Palazzo, and A. M. Rotunno, Combined analysis of KamLAND and Borexino neutrino signals from Th and U decays in the Earth's interior, *Phys. Rev. D* **82** 093006 (2010).
13. O. Šrámek, W. F. McDonough, J. G. Learned, Geoneutrinos, *Adv. High Energy Phys.* **2012** 235686 (2012).
14. The SNO+ experiment, `https://snoplus.phy.queensu.ca/`.

15. JUNO, http://juno.ihep.cas.cn/.
16. Jinping Neutrino Experiment, http://jinping.hep.tsinghua.edu.cn/.
17. J. G. Learned, S. T. Dye, S. Pakvasa, Hanohano: A deep ocean anti-neutrino detector for unique neutrino physics and geophysics studies, in *Proc. 12th Int. Workshop on Neutrino Telescopes*, (Venice, Italy, 2007), https://arxiv.org/abs/0810.4975.
18. M. Apollonio *et al.*, Determination of neutrino incoming direction in the CHOOZ experiment and its application to supernova explosion location by scintillator detectors, *Phys. Rev. D* **61** 012001 (1999).
19. T. Mitsui, Neutron background and possibility for shallow experiments, *Earth, Moon, and Planets* **99** 265 (2006).

Synthesis of the session: Why and how the Sun and the stars shine

Gianpaolo Bellini

Physics Department, Università degli Studi and Istituto Nazionale di Fisica Nucleare Milano, via Celoria 16, 20133, Italy
E-mail: gianpaolo.bellini@mi.infn.it

This article is the synthesis of the MG16 session "Why and how the Sun and the stars shine". A brief scientific history of the Borexino experiment and its discoveries are traced.

Keywords: Solar neutrinos, neutrinos physics, geo-neutrinos

In this session the technical and scientific results obtained with the Borexino experiment have been presented. In 1989 we began a feat that seemed absolutely impossible, that is to build a detector able to study, in real time, solar neutrinos from few tens of keV of energy by measuring separately the fluxes produced by the various nuclear fusion reactions that take place in the Sun. The difficulty was the natural radioactivity present in any material as well as the emissions from the rocks of the underground laboratory, the Gran Sasso lab., where it was decided to install the detector. Just owing to the natural radioactivity, the experiments before Borexino set a threshold between 4 and 5 MeV, then measuring only a tail towards the largest energies of the neutrino energy spectrum, corresponding to about 0.01% of the total flux.

The activity for Borexino started in 1990 and the first five years (1990-1995) were spent to develop innovative methods that would be able to break down the radioactivity of liquids up to ultra-traces. In order to measure the reached level of radio-purity and to understand if it was satisfying the very high level requested by the project, a special detector was needed; therefore a reduced and simplified version of Borexino has been constructed, the Counting Test Facility, which functioned as a bench mark. At 1995 this detector was able to demonstrate that our methods were able to reduce the radioactivity up to about 10^{-16} g of contaminants/g of pure material, as per project requests.

The construction, installation and tuning of the detector took a long time and only in 2007 we started the data taking; this time has been necessary because nothing is standard in Borexino: all the materials were carefully selected, all the parts of the detector and the components were developed on purpose also in collaboration with external companies, most of the activities were carried out in clean room, the detector itself has been equipped as a 10,000 level clean room. All these special cares were aimed to an exceptional suppression of the radioactive background.

Borexino is designed like an onion: the more internal a shell, the greater the radiopurity. At the center, a 125 micron-thick nylon vessel -IV, with a 4.5 m radius,

contains about 100 t of scintillator, with pseudocumene as solvent and PPO as solute. It is enclosed in a 6.75 m radius stainless steel sphere, SSS, which supports 2212 photomultipliers: the space between IV and SSS is filled with about 1000 t of pseudocumene added a quencher. A 5.50 m-radius nylon balloon functions as a barrier against to ^{222}Rn emitted by the SSS steel and the photomultipliers. All is inside a water tank (radius 9 m, height 16.9 m), which contains 2800 t of highly purified water and 208 PMTs, which, by means of the Cherenkov light, signal the muons crossing the detector. The water shields the IV from rocks and environment γs and ns.

The radio purification effort was very successful: for ^{238}U and ^{232}Th and their progeny the radiopurity reached some 10^{-20} and 10^{-19} g/g, respectively, and per the ubiquitous ^{222}Rn, < 1 count per day for 100 t of scintillator. Borexino is the unique detector which until now reached this unprecedented radiopurity.

With these premises, Borexino was able to identify the fusion reactions of the pp chain emitting neutrinos via the separate measurement of their fluxes. The problem of this analysis is to disentangle solar neutrinos signals from the residues of radio-contaminants, that survived the radio-purification. In addition, a software strategy had to be developed against the ^{11}C, which cannot be purified because it is continuously produced in the scintillator by the muons which survived the crossing of the Gran Sasso overburden (1.2 μ/m^2h). The results are summarized in the attached table where the measured rates and fluxes are compared with the SSM predictions for high and low metallicity: the comparison shows a good agreement.

The pp chain produces 99% of the entire solar energy; therefore the solar luminosity measured through solar neutrinos can be compared to the luminosity measured via photons: the good agreement found between them demonstrates that the Sun is in thermodynamic equilibrium over 10^4 years scale because, while neutrinos take a few seconds to leave the Sun, photons need more than 100,000 years.

reaction	**Borexino rates** (cpd/100t)	**Borexino fluxes** ($cm^{-2}s^{-1}$)	**SSM HZ Fluxes** ($cm^{-2}s^{-1}$)	**SSM LZ Fluxes** ($cm^{-2}s^{1}$)	**Global fit Fluxes** ($cm^{-2}s^{1}$)
pp	$134\pm10^{+6}_{-10}$	$(6.1\pm0.5^{+0.3}_{-0.5})$ x10^{10}	5.98(1± 0-006) x 10^{10}	6.03(1±0.005) x 10^{10}	$5.97^{+0.037}_{-0.033}$ x 10^{10}
^{7}Be	$48.3\pm1.1^{+0.4}_{-0.7}$	$(4.99\pm0.11^{+0.06}_{-0.12})$ x10^{9}	4.93 (1± 0.06) x 10^{9}	4.50(1±0.06) x 10^{9}	$4.80^{+0.24}_{-0.22}$ x 10^{9}
pep (HZ)	$2.43\pm0.36^{+0.15}_{-0.22}$	$(1.27\pm0.19^{+0.08}_{+0.12})$ x 10^{8}	1.44 (1± 0.009) x 10^{8}	1.46 (1± 0.009) x 10^{8}	1.448±0.08 x 10^{8}
Pep (LZ)	$2.65\pm0.36^{+0.15}_{-0.24}$	$(1.39\pm0.19^{+0.08}_{-0.13})$ x 10^{8}	1.44 (1± 0.009) x 10^{8}	1.46 (1± 0.009) x 10^{8}	
^{8}B	$0.220^{+0.015+0.006}_{-0.016-0.006}$	$5.68^{+0.39+0.03}_{-0.41-0.03}$	5.46 (1±0.12) x 10^{6}	4.50(1±0.12) x 10^{6}	$5.16^{+0.13}_{-0.09}$ x 10^{6}
hep	<0.002 (90% C.L.)	< 1.8 x 10^{5} (90% C.L.)	7.98 (1±0.30) x10^{3}	8.25(1±0.12) x 10^{3}	

Fig. 1. SSM-HZ and SSM-LZ fluxes: from Aldo Serenelli, F.Villante et al.. Global fit fluxes from J. Bergstroem et al.

The measurement of the various neutrino fluxes emitted by the pp chain nuclear fusions allows to determine the electron-neutrino survival probability in an energy range from a few tens of keV up to about 15 MeV, that is, according to the MSW model, from vacuum regime up to the matter enhanced oscillation. The experimental results are in agreement with the MSW-LMA predictions within the experimental and the Standard Solar Model uncertainties.

In the Sun the pp chain produces a temperature able to avoid its implosion caused by the gravitational forces acting among the solar matter. In massive stars, with a mass at least 30% greater that the Sun, a higher temperature is required to avoid implosion. Already in the '30s of the last century, Bethe and Von Weizsäcker hypothesized the existence of a cycle (CNO) dominating the massive stars where it reaches a core temperature of a few x 10^8 K; this hypothesis before Borexino never has been tested and demonstrated experimentally.

Fortunately, the CNO cycle is present in the Sun and produces 1% of its energy, and then Borexino was able to study the CNO cycle undertaking the analysis of the solar neutrinos produced by this cycle. The faced difficulties are due to the presence of CNO, pep and a residue of ^{210}Bi in the same energy window and to the tagging lack of the CNO energy spectrum. pep is not a problem because its shape and its rate are very well known, while to extract the CNO neutrinos flux, a strong constraint on the ^{210}Bi rate is needed. An important tool to this end is the presence of ^{210}Po in the scintillator and its secular equilibrium with ^{210}Bi.

Two components contribute to the ^{210}Po rate in the scintillator, one in secular equilibrium and one out of equilibrium: the latter decays with a half-life of 138.376 days and then, after a proper time, reaches a plateau corresponding to the ^{210}Po continuous production through the chain: ^{210}Pb- ^{210}Bi-^{210}Po and finally the stable ^{208}Pb. However, this plateau is not stable, but shows minor oscillations, due to convective currents in the scintillator which carry ^{210}Po nuclides, present on the IV walls, in the fiducial volume. Then a temperature stabilization is needed and it has been reached with a precision of 0.07 °C by thermally insulating the water tank and installing heating copper coils under the insulation.

As a result, the ^{210}Po and then the ^{210}Bi rates have been obtained, and, through a multivariate fit, the CNO neutrinos flux results to be:

$7.2^{+0.3}_{-1.7} \cdot 10^8 cm^{-2} s^{-1}$ (the uncertainty includes statistical and systematic errors), with no CNO hypothesis disfavored at 5σ C.L. In this way Borexino provided the direct experimental proof of the CNO cycle existence.

Borexino also observed with 5σ statistical significance the geoneutrinos, antineutrinos from the Earth interior. Unlike the solar neutrinos, Borexino is not the only experiment capable of making this measurement, because even the Japanese Kamland has studied geoneutrinos, with greater statistics and with greater background. This study allowed to investigate the presence of radioactive nuclides in the Earth's mantle, the contribution of radiogenic origin to the terrestrial heat and to obtain Earth's geo-physical and geo-chemical information.

The author thanks Lino Miramonti for his editing help.

The internal rotation of low-mass stars from solar and stellar seismology

G. Buldgen* and P. Eggenberger
Astronomy Department, University of Geneva, Versoix, 1290, Switzerland
**E-mail: Gael.Buldgen@unige.ch*

The possibility of measuring the internal rotation of the Sun and stars thanks to helio- and asteroseismology offers tremendous constraints on hydro- and magnetohydrodynamical processes acting in stellar interiors. Understanding the processes responsible for the transport of angular momentum in stellar interiors is crucial as they will also influence the transport of chemicals and thus the evolution of stars. Here we present some of the key results obtained in both fields and how detailed seismic analyses can provide stringent constraints on the physics of angular momentum transport in the interior of low mass stars and potentially rule out some candidates.

Keywords: Stellar physics; Stellar evolution; Helioseismology; Asteroseismology

1. Introduction

Low mass stars ($\approx 1M_{\odot}$) are mostly characterized by the presence in their structure of an outer convective envelope during their main-sequence. As a result, they will exhibit efficient braking through the effect of magnetised winds and they will most often exhibit long rotation periods of a few days. As a result of their low mass, they exhibit long evolutionary timescales of typically a few Gy on the main sequence. This implies that, if a given dynamical process is at work in stellar interiors, it will have time to leave a durable imprint on the evolution of the star, even if its characteristic timescale is relatively long. A good example of such a process is microscopic diffusion, that plays a key role in both solar and stellar modelling.

Another consequence of the existence of an outer convective envelope in low mass stars is that they will exhibit solar-like oscillations. These oscillations are global acoustic modes of the star that have been measured in both the solar case,[1] giving birth to the field of helioseismology, and in the stellar case,[2–4] paving the way for the field of asteroseismology of solar-like oscillators. With the advent of space-based photometry missions, the field of asteroseismology experienced an exponential growth, with the detection of global acoustic oscillations in thousands of stars.

Moreover, the quality of the observations enabled to perform similar analysis to those of the Sun, but on distant stars, increasing the sample of objects for which the internal rotation could be measured. The detection of mixed oscillation modes also gave access to the deep core rotation of evolved stars, providing laboratories for angular momentum transport process at various evolutionary stages.

In addition, the fact that most solar-like oscillators exhibited slow rotation of their surface allowed modellers to use perturbative approaches applied in helioseismology.

These measurements proved to be very challenging to explain using classical self-consistent rotating models including the effects of meridional circulation and shear-induced turbulence. While there exists multiple candidates to explain the internal rotation of low-mass stars on the main-sequence, the internal rotation of post-main sequence stars is still impossible to fully reproduce. Behind these difficulties hide the limitations of stellar models, emphasizing the need to improve our physical representation of stellar interiors.

In this brief review, we will discuss how the internal rotation of low mass stars is measured using a perturbative formalism in the fields of helio- and asteroseismology in Section 2. We will discuss the implications for angular momentum transport of the seismic constraints obtained on the main-sequence and will discuss the link with the transport of chemicals in Section 3. We will focus here on the constraints from acoustic oscillations only. More complete reviews, discussing also the results of massive gravity mode pulsators can also be found in the litterature[5] and classical textbooks also provide a more detailed view and additional references.[6] In Section 4, we will briefly present the main results using the mixed oscillation modes observed in post-main sequence stars and briefly discuss in Section 5 the implications for the main candidates foreseen as the "missing angular momentum transport process".

2. Measuring the internal rotation of low mass stars

The surface rotation of stars is accessible in photometry through the observations of spots on the stellar surface, or in spectroscopy through the determination of the rotational broadening of spectral lines. This allows us to test for example the formalisms used for the magnetic braking of low mass stars or the correlation between the depletion of light elements such as lithium or beryllium and the rotation period of stars in clusters. However, to study the internal rotation of stars, the only method available is by using the constraints derived from global oscillation modes. Indeed, their properties are intimately linked to the internal structure and dynamics of stellar interiors.

In the case of a non-rotating, non-magnetic, isolated star, the global oscillation modes will exhibit spherical symmetry and their angular dependencies can be decomposed on the basis of spherical harmonics. The eigenfunction of an oscillation mode, denoted $\vec{\xi}$, will thus be written in spherical coordinates as

$$\vec{\xi} = a(r) Y_{\ell,m}(\theta,\phi)\vec{e}_r + b(r)\left[\frac{\partial Y_{\ell,m}(\theta,\phi)}{\partial\theta}\vec{e}_\theta + \frac{imY_{\ell,m}(\theta,\phi)}{\sin\theta}\vec{e}_\phi\right], \qquad (1)$$

with r,θ and ϕ the radial and angular coordinates of the system, with their respective unit vectors $\vec{e}_r$, $\vec{e}_\theta$ and $\vec{e}_\phi$, $Y_{\ell,m}$ is the spherical harmonic of degrees ℓ and m, and $a(r)$ and $b(r)$ are the functions expressing the radial dependencies of the radial and angular components of the eigenfunctions.

There exists actually a degeneracy in the quantum numbers describing the eigenvalues, as they are identified by three quantum numbers, ℓ, the spherical degree, n, the radial order and m, the azimuthal degree in the general case, but only two are needed to describe the eigenvalue in the non-rotating case.

In the case of low-mass stars, we can assume that we are studying a slow rotator. Consequently, the effects of rotation will thus be treated as a perturbation of the operator describing the global acoustic modes, in a similar fashion to the treatment of a weak magnetic field in the case of the hydrogen atom in quantum mechanics. Just as the degeneracy in some eigenstates is lifted by the presence of the magnetic field in the hydrogen atom, the addition of rotation to the eigenvalue problem of stellar oscillations lifts the degeneracy of the eigenvalues, namely the oscillation frequencies of the star. They appear now as multiplets in the oscillation spectrum defined by their three quantum numbers, this time with m taking values between $-\ell$ and ℓ. This breaking of symmetry stems from the fact that adding rotation to the problem leads to the definition of an equator to the star and that one can now differentiates between the lines of the eigenfunction crossing the equator and those parallel to the equator, that define the quantum numbers ℓ and m in the base of spherical harmonics.

In other words, the eigenvalues will now be described as

$$\nu_{n,\ell,m} = \nu_{n,\ell,0} + \delta\nu_{n,\ell,m} \tag{2}$$

Mathematically, the perturbative approach applied to the stellar oscillation problem leads to a simple integral relation linking the so-called rotational splittings to the internal rotation as function of the radial position r and the latitude θ[7]

$$\delta\nu_{n,\ell,m} = m \int_0^R \int_0^\pi K_{n,\ell,m}(r,\theta)\Omega(r,\theta) dr d\theta \tag{3}$$

with $\delta\nu_{n,\ell,m}$ the rotational splitting, $\Omega(r,\theta)$ the internal rotation profile and $K_{n,\ell,m}(r,\theta)$ the so-called kernel function, depending on the internal structure and the eigenfunction of the oscillation mode.

This integral relation can be solved for the internal rotation profile, using specific numerical techniques (see e.g.[8]). Given that the linear perturbative approach is valid, the internal rotation profile can be determined independently from the stellar model used to perform the inversion. In other words, a non-rotating stellar model is sufficient to perform an analysis of the internal rotation of the star, if rotation is slow enough to be treated as a perturbation.

Equation 3 gives the general $2D$ form of the perturbative relations, but the ability to carry out $2D$ inversions is limited to a few exceptionnal targets, and was only possible for the Sun until very recently. In most seismic analyses, the hypothesis of spherical symmetry will be made and the rotation profile will be inferred as a function of r only. The rotation kernels are then a function of r only and show no explicit dependency in m, as shown for example in classical textbooks.[9] Moreover, the availability of only low ℓ modes for asteroseismic targets will also limit the

resolution of the inferences. The positions at which the rotational profile can be inferred are intrinsically bound to the physical nature of the observed oscillation modes and the behaviour of their kernel functions.

3. Results on the main sequence

Main-sequence low mass stars are known to exhibit purely acoustic modes called solar-like oscillations. As indicated by their name, these oscillations have been observed for the first time in the Sun. The main feature of these modes is that they have a much higher amplitude in the outer layers, as indicated in Fig 1 by the $1D$ rotational kernels computed here for a standard solar model.

Consequently, most of the inferences we will have for all main-sequence solar-like oscillators will focus on the upper radiative layers and the outer convective envelope. The inner 20% of the star will not be accessible to the pressure modes and thus will not be constrained by the inversion procedure. This statement is also true for the Sun, despite the wealth of seismic data accumulated in the last 30 years and has motivated the search for the elusive solar gravity modes.

3.1. *The solar case*

In the solar case, thousands of oscillation modes with degrees ranging from $\ell = 0$ to $\ell = 1000$ have been observed.[10–12] This tremendous amount of data has allowed a full scan of the internal solar rotation profile, down to $\approx 0.2 R_\odot$ (see e.g[13]).

The obtained rotation profile exhibits three main features. First, approximately solid-body rotation in the inner radiative layers of the Sun. Second, latitudinal differential rotation in the outer convective zone. Third, these two zones are connected by a narrow region, the so-called tachocline,[14,15] where the transition between the two rotation profiles occur.

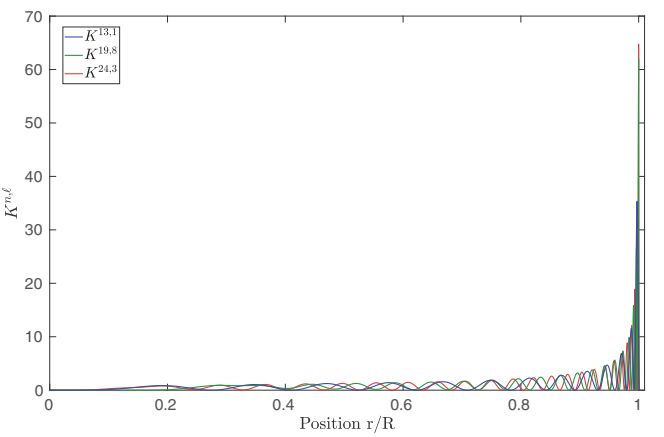


Fig. 1. $1D$ rotational kernels for various acoustic oscillaton modes of a standard solar model.

Here, we will focus mostly on the properties of the rotation profile in the inner solar radiative zone. The observed nearly uniform rotation of the inner solar layers is actually in complete contradiction with rotating models taking into account the effects of meridional circulation and shear-induced turbulence. Despite attempts at revising the prescriptions for these hydrodynamical processes,[16] it appears that their efficiency is too low to prevent the appearence of the characteristic strong differential rotation observed in these models.

Such a strong differential rotation is in disagreement with observations of the internal rotation of low-mass stars and also leads to a transport of chemicals by the shear-instability in disagreement with helioseismic inversions of the structure of the Sun and the depletion of light elements in low-mass stars. Therefore, additional processes have to been invoked to explain the discrepancies between rotating models and observations. The transport process involved should be very efficient at transporting angular momentum, but not chemical elements.

The first candidate is a large scale fossil magnetic field,[17] the second one are magnetic instabilities[18–20] and the third one are internal gravity waves.[21] We will further discuss the properties of these candidates in Section 5. They remain, to this day, the main suspects to explain the efficient angular momentum transport in stellar interiors.

3.2. *Results for Kepler and CoRoT targets*

With the advent of the CoRoT and *Kepler* missions, inferences on the internal rotation of main sequence stars could be carried out for a small sample of the best asteroseismic targets. Various authors[22–24] have shown that in the case of solar-like oscillators, the measurement of the average rotation seen by acoustic modes was consistent with the surface rotation derived from spots or spectroscopic measurements. In other words, no significant degree of differential rotation was found inside solar-like stars. This again accredited the existence of a very efficient transport of angular momentum acting in these stars, allowing to flatten their internal rotation profile.

However, a word of caution is required when stating that solar-like stars rotate as "solid-bodies". The constraints derived from asteroseismic data are not able to fully resolve the internal rotation of distant stars, but rather provide an average measurement of their internal rotation. Just as for the solar case, the purely acoustic oscillations observed for main-sequence solar-like oscillators do not allow to probe the deep core of these objects. As we will see in Section 5, this has important consequences to select one of the three candidates for additional transport mentioned above[a]. Recently, two studies[25,26] have also shown that solar-like oscillators also exhibited latitudinal differential rotation in their convective envelopes.

[a]We note that a similar situation is present also in more massive main-sequence stars such as γ Doradus stars, where the observed gravity modes only constrain the near core region, but not the convective core itself, as they do not propagate in these regions. Getting constraints on this region requires the analysis of inertial modes, recently observed.

3.3. *Link with the transport of chemicals*

An important consequence of the presence of an efficient angular momentum transport mechanism is its potential impact on the transport of chemical elements inside the star. For example, the effects of shear-induced instability in the presence of strong radial differential rotation in the radiative layers will be to erase the effects of microscopic diffusion and inhibit the settling of heavier elements from the convective layers to the deep interior.[27,28] While this effect should be limited as a result of the very low amplitude of rotation gradient in stellar radiative zones, it might not be fully negligible. In addition, a certain degree of additional mixing is required to reproduce the lithium abundance of the Sun[29] and solar-like stars.[30–33]

This is illustrated in Fig. 2 for the solar case and young solar-like stars.[34] The results show that mixing by the shear instability and the effects of magnetic instabilities will lead to a certain degree of mixing at the base of the convective zone thus changing both the helium and lithium abundances.

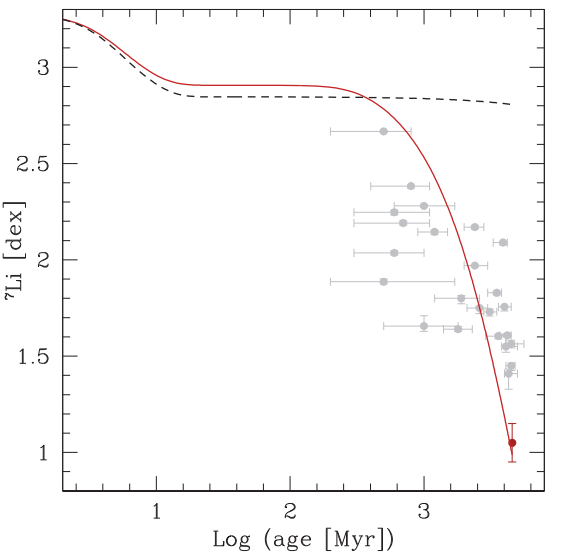


Fig. 2. Lithium abundance evolution as a function of the logarithm of age. The red dot correspond to the solar Lithium abundance[35] and the grey dots correspond to the abundance of solar-like stars.[34]The dashed line indicates the evolution of the surface Li abundance for a standard solar model, while the continuous red line corresponds to a model with hydrodynamic and magnetic instabilities.

Some of the stars observed by the *Kepler* satellite offer excellent testbeds for such effects, as for example the 16Cyg binary system,[36–39] where the binarity also allows to test the effects of planetary formation on the abundances of light elements. Moreover, seismic analyses of the helium abundance in the envelope of other *Kepler*

targets[40] have shown the need for a mecanism reducing the efficiency of microscopic diffusion, that could be due to the effects of small rotation gradients.

4. Results for evolved stars

The determination of the internal rotation of subgiant and red giant stars[41–44] has been one of the main successes of the space-based photometry missions. These results have been achievable thanks to the observation of so-called mixed oscillation modes. The peculiarity of these modes is that they are of double nature and have high amplitudes both in the deep core and in the outer layers of post-main sequence solar-like oscillators. They exhibit a gravity-like character in the core and an acoustic nature in the envelope. Therefore they can be easily detected and carry information on the deepest layers of evolved stars, regions that are even unaccessible in the Sun due to the purely acoustic nature of the observed oscillations, despite the quality of helioseismic data.

From the analysis of these oscillation modes, it has been shown that subgiant and red giant stars exhibited radial differential rotation.[41–44] Namely, that their cores rotated faster than their outer layers. While this was expected, as a result of the contraction of the core and the expansion of the convective envelope undergone as the star leaves the main sequence, the degree of differential rotation observed is far lower than what is predicted by theoretical models.[45–47]

Recent results[48] have also indicated that young subgiant stars seem to exhibit almost solid-body rotation. Later on in the subgiant phase, they show some degree of differential rotation that develops but it appears that on the red giant phase, the stellar core does not spin up with the evolution.[49,50] In later stages of evolution such as the red clump or the secondary clump, similar conclusions are drawn regarding the need for a very efficient transport process.

A crucial point to note here is that to study the efficiency of the missing angular momentum transport process, core rotation is not entirely sufficient, and a measurement of the surface rotation is also required. This is available for a few targets, either from seismology, or from spot modulations. However in the latter case, it seems that the sample is biased towards active, fast rotators, who could perhaps not be quite representative of the bulk of red giants.

The main issue posed by the results obtained for post-main sequence stars is that there is, to this day, no single mechanism capable of explaining all the observations. Analyses of the average efficiency of the missing transport process[51,52] have shown that it should exhibit a drop in efficiency during the subgiant phase but reincrease as the star ascends the red giant branch. This could possibly point at either a change of regime in the transport process or at multiple transport processes acting at different evolutionary stages. For example, it has been shown that mixed modes themselves would lead to an efficient transport of angular momentum on the upper red giant branch and could potentially take over at stages where another process sees its efficiency reducing.[53]

5. Transport mechanism candidates

Apart from the transport by mixed modes, the main suspects under investigation are the processes already invoked to explain the solar rotation profile. Namely, these are large scale fossil magnetic fields, internal gravity waves and magnetic instabilities.

However, all of them needed to be revised to explain the results observed in post-main sequence stars. We will here briefly summarize the main results and remaining issues with each of these processes.

5.1. *Fossil magnetic fields*

The first candidate introduced to explain solar rotation is a large-scale fossil field present in the radiative zone of the Sun.[17] It was demonstrated that the field, if properly confined, could lead to a rigid rotation profile in the interior of the Sun.

The main difficulty for the fossil field solution in the solar case is indeed linked to the confinement of the field, as it should not extent in the solar convective zone where differential rotation in latitude is observed. This requires the field to reach a very specific configuration for which simulations give conflictual results regarding its confinement to the radiative zone.[54–57] In addition, recent work[58] has shown that, in addition to the confinement problem, simulations required a significant increase in the viscosity of the plasma to avoid a dead zone of the field where the profile would not reproduce at all the observed data.

Nevertheless, a second issue with the fossil field configuration used to explain the solar rotation profile is that it enforces a strictly solid-body rotation in radiative zones. As shown before, this is not what is found in later evolutionary stages and thus another configuration, or another physical process, is required to explain asteroseismic data. One possible configuration for red giants is to have a solid-body rotation profile in radiative zones, combined to a profile in the power law of the radius in convective zones,[59,60] of the form $\Omega(r) \propto r^{-\alpha}$, with $\alpha \in [1.0, 1.5]$. These studies[59,60] showed that such a rotation profile could explain the core rotation rate of red giant branch stars inferred from seismic data.[49,50]

However, these results are in contradiction with recent studies[61–63] who seem to indicate that the transition from the slow rotating envelope to the fast rotating core in red giant stars seems to be located in the radiative zone, close to the hydrogen burning shell. A detailed analysis of the case of Kepler 56 was recently performed,[64] including a full dedicated modelling of the internal structure of the star and an MCMC analysis of the rotational splittings using simple parametric profiles, including the power law solution provided by large scale fossil fields. Their results show that such a rotation profile cannot be used to explain the individual rotational splittings unless α is allowed to reach higher values of the order of 3.5, in contradiction with theoretical values.[59] This suggests that fossil magnetic fields are not the explanation for the efficient angular momentum transport process acting in evolved stars, and other processes must also be invoked.

5.2. *Internal gravity waves*

The effect of internal gravity waves has been studied already early on as being of potential interest to the modelling of stellar interiors and applied to the solar case.[21,65]

Later studies[66,67] indicated that the process was not efficient enough to actually flatten the solar rotation profile and pointed that the effect of waves excited by Reynold's stresses would leave a strong imprint of the wavefront on the internal solar rotation profile that is incompatible with observations. When applied to later evolutionary stages, the gravity wave model used on the main-sequence[21] is also found to be not efficient enough to reproduce the internal rotation of subgiants and red giant stars.[68]

However, hydrodynamical simulations[69] underlined the importance of the effects of convective plumes in the generation of gravity waves. These effects were later studied, showing that gravity waves induced by the impact of convective plumes could be much more efficient at transporting angular momentum both in the Sun and subgiant stars.[70,71] They also showed that it would likely not be able to operate efficiently on the red giant branch, where another process would then be needed. It would then be particularly interesting to compute full evolutionary models with plume-induced gravity waves to investigate their effects on the rotation profile of solar-type and subgiant stars.

5.3. *Magnetic instabilities*

The last main candidate to explain the efficient transport of angular momentum in stellar radiatives zone are magnetic instabilities. Work by H. Spruit[72] showed that the first instability likely set in stellar conditions would be the Tayler instability. In the presence of a differentially rotating fluid, the winding up of a seed magnetic field together with the Tayler instability has been shown to lead to an efficient angular momentum transport mechanism, the so-called Tayler-Spruit dynamo.[73] This transport mechanism has been shown to be able to reproduce the internal solar rotation profile.[19,20]

While a promising candidate, numerical simulations have shown contrasting results about the apparition of the Tayler-Spruit dynamo.[74,75] The main difficulty with such conclusions is that such simulations are never in fully realistic stellar conditions.

Nevertheless, in its original form, the Tayler-Spruit dynamo still proves not efficient enough to counteract the spin up of the core at the end of the main-sequence. Models computed taking it into account still predict faster core rotation rates than the seismically inferred values.[76,77]

Recently, a modification of the Tayler-Spruit instability was proposed that led to a more efficient braking of the stellar cores.[78] Further analyses[79] showed however that this variant could not reproduce all the observational constraints. Namely, it was found to be too efficient for subgiants if one wished to reproduce the red giant

cores, or not efficient enough for red giants if the calibration parameter was set to reproduce subgiants, as shown in Fig 3. This calibration parameter, denoted α has to be varied from 0.5 to 1.5 to reproduce respectively the subgiants and the red giants. However, as it is set to the third power in the revised formalism, such a variation is far from anecdotic. A similar issue was found for secondary clump stars and white dwarfs.[80]

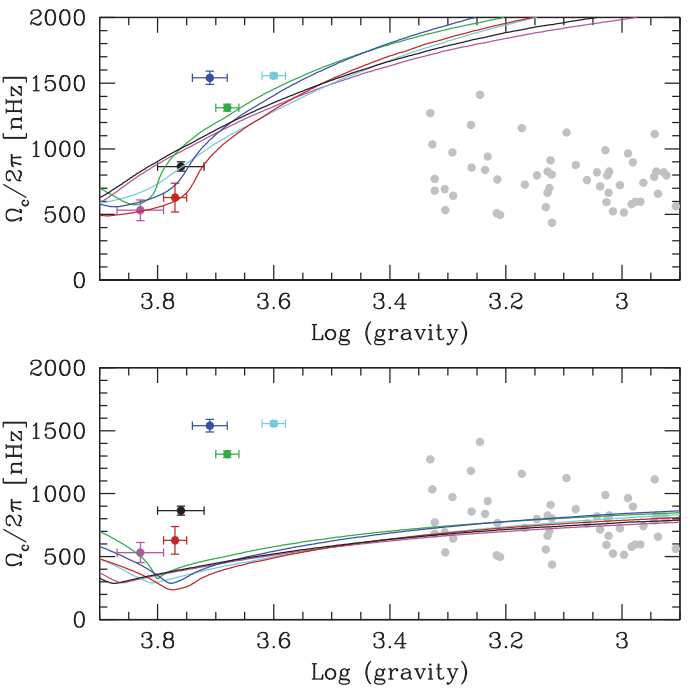


Fig. 3. Core rotation of subgiant and red giant stars as a function of the logarithm of the surface gravity. The colored points correspond to subgiant stars[43] and the grey dots to red giant stars.[49] In the upper and lower panels, the calibration parameter α in the revised version of the Tayler-Spruit dynamo is set to 0.5 and 1.5, respectively.

6. Conclusion

In this brief review, we have described the current state of the vision solar and stellar seismology brought on the internal rotation of low-mass stars in various evolutionary stages. We have focussed on the inferences brought by acoustic and mixed oscillations modes only, trying to provide a complete vision of the current issues at hand regarding the missing angular momentum transport mechanism found to act in stellar radiative zones.

The formalism for seismic inferences within the first-order perturbative formalism has been introduced in Section 2, while results for the main-sequence and post-

main sequence stars have been discussed in Sections 3 and 4. We have also briefly introduced the three main candidates currently under extensive investigations in Section 5.

The main conclusion to be drawn from the current situation is a sort of stalemate between multiple processes, none of them providing a unifying satisfactory solution. In that respect, further investigations on the effiency of the missing transport process are required, analysing its dependencies with stellar properties such as mass and metallicity. To this end, better characterizing the location of rotation gradients inside stars is crucial, as well as providing both surface and core rotation measurements from a consistent seismic inference technique to fully constrain the efficiency of the transport process.

It should also be noted that the potential strong influence of chemical composition gradients in the development and efficiency of instabilities (magnetic or hydrodynamic) also calls for a better depiction of the chemical structure of stars. Therefore, the potential solution to the angular momentum transport problem is also tightly linked to the reliability of our vision of the internal structure of stars, that also relies on improving seismic inference techniques. At stakes is likely the recipe for a new generation of solar and stellar models, which would take into account both the effects of rotation, the depletion of light elements and better reproduce seismic and spectroscopic constraints. Moreover, while a unifying solution is tempting, the variations seen at different evolutionary stages might be pointing at various processes taking over during the evolution. To provide a clearer picture, further efforts must be made in the improvement of seismic inference techniques and in the exploitation of the current and future datasets from the CoRoT,[81] *Kepler*,[82] TESS[83] and Plato[84] missions.

References

1. R. B. Leighton, R. W. Noyes and G. W. Simon, Velocity Fields in the Solar Atmosphere. I. Preliminary Report., *ApJ* **135**, p. 474 (March 1962).
2. H. Kjeldsen, T. R. Bedding, M. Viskum and S. Frandsen, Solarlike Oscillations in eta Boo, *AJ* **109**, p. 1313 (March 1995).
3. M. Martić, J. Schmitt, J. C. Lebrun, C. Barban, P. Connes, F. Bouchy, E. Michel, A. Baglin, T. Appourchaux and J. L. Bertaux, Evidence for global pressure oscillations on Procyon, *A&A* **351**, 993 (November 1999).
4. F. Bouchy and F. Carrier, The acoustic spectrum of alpha Cen A, *A&A* **390**, 205 (July 2002).
5. C. Aerts, S. Mathis and T. M. Rogers, Angular Momentum Transport in Stellar Interiors, *Annu. Rev. Astron. Astrophys.* **57**, 35 (August 2019).
6. A. Maeder, *Physics, Formation and Evolution of Rotating Stars* 2009.
7. P. Ledoux, The Nonradial Oscillations of Gaseous Stars and the Problem of Beta Canis Majoris., *ApJ* **114**, p. 373 (November 1951).
8. J. Schou, H. M. Antia, S. Basu, R. S. Bogart, R. I. Bush, S. M. Chitre, J. Christensen-Dalsgaard, M. P. Di Mauro, W. A. Dziembowski, A. Eff-Darwich, D. O. Gough, D. A. Haber, J. T. Hoeksema, R. Howe, S. G. Korzennik, A. G. Kosovichev, R. M. Larsen, F. P. Pijpers, P. H. Scherrer, T. Sekii, T. D. Tarbell, A. M. Title, M. J. Thompson

and J. Toomre, Helioseismic Studies of Differential Rotation in the Solar Envelope by the Solar Oscillations Investigation Using the Michelson Doppler Imager, *ApJ* **505**, 390 (September 1998).
9. W. Unno, Y. Osaki, H. Ando, H. Saio and H. Shibahashi, *Nonradial oscillations of stars* 1989.
10. S. G. Korzennik, M. C. Rabello-Soares, J. Schou and T. P. Larson, Accurate Characterization of High-degree Modes Using MDI Observations, *ApJ* **772**, p. 87 (August 2013).
11. T. P. Larson and J. Schou, Global-Mode Analysis of Full-Disk Data from the Michelson Doppler Imager and the Helioseismic and Magnetic Imager, *Solar Physics* **293**, p. 29 (February 2018).
12. J. Reiter, J. Rhodes, E. J., A. G. Kosovichev, P. H. Scherrer, T. P. Larson and I., S. F. Pinkerton, A Method for the Estimation of f- and p-mode Parameters and Rotational Splitting Coefficients from Un-averaged Solar Oscillation Power Spectra, *ApJ* **894**, p. 80 (May 2020).
13. M. J. Thompson, J. Toomre, E. R. Anderson, H. M. Antia, G. Berthomieu, D. Burtonclay, S. M. Chitre, J. Christensen-Dalsgaard, T. Corbard, M. De Rosa, C. R. Genovese, D. O. Gough, D. A. Haber, J. W. Harvey, F. Hill, R. Howe, S. G. Korzennik, A. G. Kosovichev, J. W. Leibacher, F. P. Pijpers, J. Provost, J. Rhodes, E. J., J. Schou, T. Sekii, P. B. Stark and P. R. Wilson, Differential Rotation and Dynamics of the Solar Interior, *Science* **272**, 1300 (May 1996).
14. E. A. Spiegel and J.-P. Zahn, The solar tachocline, *A&Ap* **265**, 106 (November 1992).
15. D. W. Hughes, R. Rosner and N. O. Weiss, *The Solar Tachocline* May 2007.
16. S. Mathis, V. Prat, L. Amard, C. Charbonnel, A. Palacios, N. Lagarde and P. Eggenberger, Anisotropic turbulent transport in stably stratified rotating stellar radiation zones, *A&A* **620**, p. A22 (December 2018).
17. D. O. Gough and M. E. McIntyre, Inevitability of a magnetic field in the Sun's radiative interior, *Nature* **394**, 755 (August 1998).
18. H. C. Spruit, Dynamo action by differential rotation in a stably stratified stellar interior, *A&A* **381**, 923 (January 2002).
19. P. Eggenberger, A. Maeder and G. Meynet, Stellar evolution with rotation and magnetic fields. IV. The solar rotation profile, *A&A* **440**, L9 (September 2005).
20. P. Eggenberger, G. Buldgen and S. J. A. J. Salmon, Rotation rate of the solar core as a key constraint to magnetic angular momentum transport in stellar interiors, *A&A* **626**, p. L1 (Jun 2019).
21. C. Charbonnel and S. Talon, Influence of Gravity Waves on the Internal Rotation and Li Abundance of Solar-Type Stars, *Science* **309**, 2189 (September 2005).
22. M. N. Lund, M. S. Miesch and J. Christensen-Dalsgaard, Differential Rotation in Main-sequence Solar-like Stars: Qualitative Inference from Asteroseismic Data, *ApJ* **790**, p. 121 (August 2014).
23. O. Benomar, M. Takata, H. Shibahashi, T. Ceillier and R. A. García, Nearly uniform internal rotation of solar-like main-sequence stars revealed by space-based asteroseismology and spectroscopic measurements, *MNRAS* **452**, 2654 (September 2015).
24. M. B. Nielsen, H. Schunker, L. Gizon and W. H. Ball, Constraining differential rotation of Sun-like stars from asteroseismic and starspot rotation periods, *A&A* **582**, p. A10 (October 2015).
25. O. Benomar, M. Bazot, M. B. Nielsen, L. Gizon, T. Sekii, M. Takata, H. Hotta, S. Hanasoge, K. R. Sreenivasan and J. Christensen-Dalsgaard, Asteroseismic detection of latitudinal differential rotation in 13 Sun-like stars, *Science* **361**, 1231 (September 2018).

26. M. Bazot, O. Benomar, J. Christensen-Dalsgaard, L. Gizon, S. Hanasoge, M. Nielsen, P. Petit and K. R. Sreenivasan, Latitudinal differential rotation in the solar analogues 16 Cygni A and B, *A&A* **623**, p. A125 (March 2019).
27. P. Eggenberger, G. Meynet, A. Maeder, A. Miglio, J. Montalban, F. Carrier, S. Mathis, C. Charbonnel and S. Talon, Effects of rotational mixing on the asteroseismic properties of solar-type stars, *A&A* **519**, p. A116 (September 2010).
28. M. Deal, M. J. Goupil, J. P. Marques, D. R. Reese and Y. Lebreton, Chemical mixing in low mass stars. I. Rotation against atomic diffusion including radiative acceleration, *A&A* **633**, p. A23 (January 2020).
29. O. Richard, S. Vauclair, C. Charbonnel and W. A. Dziembowski, New solar models including helioseismological constraints and light-element depletion., *A&Ap* **312**, 1000 (August 1996).
30. S. Talon, C. Charbonnel, The Li dip: A probe of angular momentum transport in low mass stars, *A&A*, **335**, p. 959 to 968 (1998).
31. P. Eggenberger, A. Maeder and G. Meynet, Effects of rotation and magnetic fields on the lithium abundance and asteroseismic properties of exoplanet-host stars, *A&A* **519**, p. L2 (September 2010).
32. F. Thévenin, A. V. Oreshina, V. A. Baturin, A. B. Gorshkov, P. Morel and J. Provost, Evolution of lithium abundance in the Sun and solar twins, *A&A* **598**, p. A64 (February 2017).
33. M. Carlos, J. Meléndez, L. Spina, L. A. dos Santos, M. Bedell, I. Ramirez, M. Asplund, J. L. Bean, D. Yong, J. Yana Galarza and A. Alves-Brito, The Li-age correlation: The Sun is unusually Li deficient for its age, *MNRAS* **485**, 4052 (May 2019).
34. M. Carlos, J. Meléndez, L. Spina, L. A. dos Santos, M. Bedell, I. Ramirez, M. Asplund, J. L. Bean, D. Yong, J. Yana Galarza and A. Alves-Brito, The Li-age correlation: The Sun is unusually Li deficient for its age, *MNRAS* **485**, 4052 (May 2019).
35. M. Asplund, N. Grevesse, A. J. Sauval and P. Scott, The Chemical Composition of the Sun, *ARA&A* **47**, 481 (September 2009).
36. M. Deal, O. Richard and S. Vauclair, Accretion of planetary matter and the lithium problem in the 16 Cygni stellar system, *A&A* **584**, p. A105 (December 2015).
37. G. Buldgen, D. R. Reese and M. A. Dupret, Constraints on the structure of 16 Cygni A and 16 Cygni B using inversion techniques, *A&A* **585**, p. A109 (January 2016).
38. G. Buldgen, S. J. A. J. Salmon, D. R. Reese and M. A. Dupret, In-depth study of 16CygB using inversion techniques, *A&A* **596**, p. A73 (December 2016).
39. M. Bazot, Uncertainties and biases in modelling 16 Cygni A and B, *A&A* **635**, p. A26 (March 2020).
40. K. Verma and V. Silva Aguirre, Helium settling in F stars: Constraining turbulent mixing using observed helium glitch signature, *MNRAS* **489**, 1850 (October 2019).
41. P. G. Beck, J. Montalban, T. Kallinger, J. De Ridder, C. Aerts, R. A. García, S. Hekker, M.-A. Dupret, B. Mosser, P. Eggenberger, D. Stello, Y. Elsworth, S. Frandsen, F. Carrier, M. Hillen, M. Gruberbauer, J. Christensen-Dalsgaard, A. Miglio, M. Valentini, T. R. Bedding, H. Kjeldsen, F. R. Girouard, J. R. Hall and K. A. Ibrahim, Fast core rotation in red-giant stars as revealed by gravity-dominated mixed modes, *Nature* **481**, 55 (January 2012).
42. S. Deheuvels, R. A. García, W. J. Chaplin, S. Basu, H. M. Antia, T. Appourchaux, O. Benomar, G. R. Davies, Y. Elsworth, L. Gizon, M. J. Goupil, D. R. Reese, C. Regulo, J. Schou, T. Stahn, L. Casagrande, J. Christensen-Dalsgaard, D. Fischer, S. Hekker, H. Kjeldsen, S. Mathur, B. Mosser, M. Pinsonneault, J. Valenti, J. L. Christiansen, K. Kinemuchi and F. Mullally, Seismic Evidence for a Rapidly Rotating Core in a Lower-giant-branch Star Observed with Kepler, *ApJ* **756**, p. 19 (September 2012).

43. S. Deheuvels, G. Doğan, M. J. Goupil, T. Appourchaux, O. Benomar, H. Bruntt, T. L. Campante, L. Casagrande, T. Ceillier, G. R. Davies, P. De Cat, J. N. Fu, R. A. García, A. Lobel, B. Mosser, D. R. Reese, C. Regulo, J. Schou, T. Stahn, A. O. Thygesen, X. H. Yang, W. J. Chaplin, J. Christensen-Dalsgaard, P. Eggenberger, L. Gizon, S. Mathis, J. Molenda-Żakowicz and M. Pinsonneault, Seismic constraints on the radial dependence of the internal rotation profiles of six Kepler subgiants and young red giants, *A&A* **564**, p. A27 (April 2014).
44. S. Deheuvels, J. Ballot, P. G. Beck, B. Mosser, R. Østensen, R. A. García and M. J. Goupil, Seismic evidence for a weak radial differential rotation in intermediate-mass core helium burning stars, *A&A* **580**, p. A96 (August 2015).
45. P. Eggenberger, J. Montalbán and A. Miglio, Angular momentum transport in stellar interiors constrained by rotational splittings of mixed modes in red giants, *A&A* **544**, p. L4 (August 2012).
46. J. P. Marques, M. J. Goupil, Y. Lebreton, S. Talon, A. Palacios, K. Belkacem, R.-M. Ouazzani, B. Mosser, A. Moya, P. Morel, B. Pichon, S. Mathis, J.-P. Zahn, S. Turck-Chièze and P. A. P. Nghiem, Seismic diagnostics for transport of angular momentum in stars. I. Rotational splittings from the pre-main sequence to the red-giant branch, *A&A* **549**, p. A74 (January 2013).
47. T. Ceillier, P. Eggenberger, R. A. García and S. Mathis, Understanding angular momentum transport in red giants: The case of KIC 7341231, *A&A* **555**, p. A54 (July 2013).
48. S. Deheuvels, J. Ballot, P. Eggenberger, F. Spada, A. Noll and J. W. den Hartogh, Seismic evidence for near solid-body rotation in two Kepler subgiants and implications for angular momentum transport, *A&A* **641**, p. A117 (September 2020).
49. B. Mosser, M. J. Goupil, K. Belkacem, J. P. Marques, P. G. Beck, S. Bloemen, J. De Ridder, C. Barban, S. Deheuvels, Y. Elsworth, S. Hekker, T. Kallinger, R. M. Ouazzani, M. Pinsonneault, R. Samadi, D. Stello, R. A. García, T. C. Klaus, J. Li, S. Mathur and R. L. Morris, Spin down of the core rotation in red giants, *A&A* **548**, p. A10 (December 2012).
50. C. Gehan, B. Mosser, E. Michel, R. Samadi and T. Kallinger, Core rotation braking on the red giant branch for various mass ranges, *A&A* **616**, p. A24 (Aug 2018).
51. P. Eggenberger, N. Lagarde, A. Miglio, J. Montalbán, S. Ekström, C. Georgy, G. Meynet, S. Salmon, T. Ceillier, R. A. García, S. Mathis, S. Deheuvels, A. Maeder, J. W. den Hartogh and R. Hirschi, Constraining the efficiency of angular momentum transport with asteroseismology of red giants: The effect of stellar mass, *A&A* **599**, p. A18 (March 2017).
52. P. Eggenberger, S. Deheuvels, A. Miglio, S. Ekström, C. Georgy, G. Meynet, N. Lagarde, S. Salmon, G. Buldgen, J. Montalbán, F. Spada and J. Ballot, Asteroseismology of evolved stars to constrain the internal transport of angular momentum. I. Efficiency of transport during the subgiant phase, *A&A* **621**, p. A66 (January 2019).
53. K. Belkacem, J. P. Marques, M. J. Goupil, B. Mosser, T. Sonoi, R. M. Ouazzani, M. A. Dupret, S. Mathis and M. Grosjean, Angular momentum redistribution by mixed modes in evolved low-mass stars. II. Spin-down of the core of red giants induced by mixed modes, *A&A* **579**, p. A31 (July 2015).
54. A. S. Brun and J.-P. Zahn, Magnetic confinement of the solar tachocline, *A&A* **457**, 665 (October 2006).
55. A. Strugarek, A. S. Brun and J.-P. Zahn, Magnetic confinement of the solar tachocline: II. Coupling to a convection zone, *A&A* **532**, p. A34 (August 2011).

56. L. A. Acevedo-Arreguin, P. Garaud and T. S. Wood, Dynamics of the solar tachocline - III. Numerical solutions of the Gough and McIntyre model, *MNRAS* **434**, 720 (September 2013).
57. T. S. Wood and N. H. Brummell, A Self-consistent Model of the Solar Tachocline, *ApJ* **853**, p. 97 (February 2018).
58. F. Spada, A. C. Lanzafame and A. F. Lanza, A semi-analytic approach to angular momentum transport in stellar radiative interiors, *MNRAS* **404**, 641 (May 2010).
59. Y. Kissin and C. Thompson, Rotation of Giant Stars, *ApJ* **808**, p. 35 (July 2015).
60. K. Takahashi and N. Langer, Modeling of magneto-rotational stellar evolution. I. Method and first applications, *A&A* **646**, p. A19 (February 2021).
61. M. P. Di Mauro, R. Ventura, D. Cardini, D. Stello, J. Christensen-Dalsgaard, W. A. Dziembowski, L. Paternò, P. G. Beck, S. Bloemen, G. R. Davies, K. De Smedt, Y. Elsworth, R. A. García, S. Hekker, B. Mosser and A. Tkachenko, Internal Rotation of the Red-giant Star KIC 4448777 by Means of Asteroseismic Inversion, *ApJ* **817**, p. 65 (January 2016).
62. H. Klion and E. Quataert, A diagnostic for localizing red giant differential rotation, *MNRAS* **464**, L16 (January 2017).
63. M. P. Di Mauro, R. Ventura, E. Corsaro and B. Lustosa De Moura, The Rotational Shear Layer inside the Early Red-giant Star KIC 4448777, *ApJ* **862**, p. 9 (July 2018).
64. L. Fellay, G. Buldgen, P. Eggenberger, S. Khan, S. J. A. J. Salmon, A. Miglio and J. Montalbán, Asteroseismology of evolved stars to constrain the internal transport of angular momentum. IV. Internal rotation of Kepler 56 from an MCMC analysis of the rotational splittings, *arXiv e-prints* , p. arXiv:2108.02670 (August 2021).
65. S. Talon, P. Kumar and J.-P. Zahn, Angular Momentum Extraction by Gravity Waves in the Sun, *ApJL* **574**, L175 (August 2002).
66. T. M. Rogers, Numerical Simulations of Gravity Wave Driven Shear Flows in the Solar Tachocline, in *Unsolved Problems in Stellar Physics: A Conference in Honor of Douglas Gough*, eds. R. J. Stancliffe, G. Houdek, R. G. Martin and C. A. Tout, American Institute of Physics Conference Series, Vol. 948November 2007.
67. P. A. Denissenkov, M. Pinsonneault and K. B. MacGregor, What Prevents Internal Gravity Waves from Disturbing the Solar Uniform Rotation?, *ApJ* **684**, 757 (September 2008).
68. J. Fuller, D. Lecoanet, M. Cantiello and B. Brown, Angular Momentum Transport via Internal Gravity Waves in Evolving Stars, *ApJ* **796**, p. 17 (November 2014).
69. T. M. Rogers and G. A. Glatzmaier, Angular Momentum Transport by Gravity Waves in the Solar Interior, *ApJ* **653**, 756 (December 2006).
70. C. Pinçon, K. Belkacem and M. J. Goupil, Generation of internal gravity waves by penetrative convection, *A&A* **588**, p. A122 (April 2016).
71. C. Pinçon, K. Belkacem, M. J. Goupil and J. P. Marques, Can plume-induced internal gravity waves regulate the core rotation of subgiant stars?, *A&A* **605**, p. A31 (September 2017).
72. H. C. Spruit, Differential rotation and magnetic fields in stellar interiors, *A&A* **349**, 189 (September 1999).
73. H. C. Spruit, Dynamo action by differential rotation in a stably stratified stellar interior, *A&A* **381**, 923 (January 2002).
74. J. Zahn, A. S. Brun and S. Mathis, On magnetic instabilities and dynamo action in stellar radiation zones, *A&A* **474**, 145 (October 2007).
75. J. Braithwaite and H. C. Spruit, Magnetic fields in non-convective regions of stars, *Royal Society Open Science* **4**, p. 160271 (February 2017).

76. M. Cantiello, C. Mankovich, L. Bildsten, J. Christensen-Dalsgaard and B. Paxton, Angular Momentum Transport within Evolved Low-mass Stars, *ApJ* **788**, p. 93 (June 2014).
77. J. W. den Hartogh, P. Eggenberger and R. Hirschi, Constraining transport of angular momentum in stars. Combining asteroseismic observations of core helium burning stars and white dwarfs, *A&A* **622**, p. A187 (February 2019).
78. J. Fuller, A. L. Piro and A. S. Jermyn, Slowing the spins of stellar cores, *MNRAS* **485**, 3661 (May 2019).
79. P. Eggenberger, J. W. den Hartogh, G. Buldgen, G. Meynet, S. J. A. J. Salmon and S. Deheuvels, Asteroseismology of evolved stars to constrain the internal transport of angular momentum. II. Test of a revised prescription for transport by the Tayler instability, *A&A* **631**, p. L6 (November 2019).
80. J. W. den Hartogh, P. Eggenberger and S. Deheuvels, Asteroseismology of evolved stars to constrain the internal transport of angular momentum. III. Using the rotation rates of intermediate-mass stars to test the Fuller-formalism, *A&A* **634**, p. L16 (February 2020).
81. A. Baglin, M. Auvergne, L. Boisnard, T. Lam-Trong, P. Barge, C. Catala, M. Deleuil, E. Michel and W. Weiss, CoRoT: A high precision photometer for stellar ecolution and exoplanet finding, in *36th COSPAR Scientific Assembly*, COSPAR Meeting Vol. 362006.
82. W. J. Borucki, D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, E. W. Dunham, A. K. Dupree, T. N. Gautier, J. C. Geary, R. Gilliland, A. Gould, S. B. Howell, J. M. Jenkins, Y. Kondo, D. W. Latham, G. W. Marcy, S. Meibom, H. Kjeldsen, J. J. Lissauer, D. G. Monet, D. Morrison, D. Sasselov, J. Tarter, A. Boss, D. Brownlee, T. Owen, D. Buzasi, D. Charbonneau, L. Doyle, J. Fortney, E. B. Ford, M. J. Holman, S. Seager, J. H. Steffen, W. F. Welsh, J. Rowe, H. Anderson, L. Buchhave, D. Ciardi, L. Walkowicz, W. Sherry, E. Horch, H. Isaacson, M. E. Everett, D. Fischer, G. Torres, J. A. Johnson, M. Endl, P. MacQueen, S. T. Bryson, J. Dotson, M. Haas, J. Kolodziejczak, J. Van Cleve, H. Chandrasekaran, J. D. Twicken, E. V. Quintana, B. D. Clarke, C. Allen, J. Li, H. Wu, P. Tenenbaum, E. Verner, F. Bruhweiler, J. Barnes and A. Prsa, Kepler Planet-Detection Mission: Introduction and First Results, *Science* **327**, 977 (February 2010).
83. G. R. Ricker, J. N. Winn, R. Vanderspek, D. W. Latham, G. Á. Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. De Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. M. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry and J. Villasenor, Transiting Exoplanet Survey Satellite (TESS), *Journal of Astronomical Telescopes, Instruments, and Systems* **1**, p. 014003 (January 2015).
84. H. Rauer, C. Catala, C. Aerts, T. Appourchaux, W. Benz, A. Brandeker, J. Christensen-Dalsgaard, M. Deleuil, L. Gizon, M. J. Goupil, M. Güdel, E. Janot-Pacheco, M. Mas-Hesse, I. Pagano, G. Piotto, D. Pollacco, Ċ. Santos, A. Smith, J. C. Suárez, R. Szabó, S. Udry, V. Adibekyan, Y. Alibert, J. M. Almenara, P. Amaro-Seoane, M. A.-v. Eiff, M. Asplund, E. Antonello, S. Barnes, F. Baudin, K. Belkacem, M. Bergemann, G. Bihain, A. C. Birch, X. Bonfils, I. Boisse, A. S. Bonomo,

F. Borsa, I. M. Brandão, E. Brocato, S. Brun, M. Burleigh, R. Burston, J. Cabrera, S. Cassisi, W. Chaplin, S. Charpinet, C. Chiappini, R. P. Church, S. Csizmadia, M. Cunha, M. Damasso, M. B. Davies, H. J. Deeg, R. F. Díaz, S. Dreizler, C. Dreyer, P. Eggenberger, D. Ehrenreich, P. Eigmüller, A. Erikson, R. Farmer, S. Feltzing, F. de Oliveira Fialho, P. Figueira, T. Forveille, M. Fridlund, R. A. García, P. Giommi, G. Giuffrida, M. Godolt, J. Gomes da Silva, T. Granzer, J. L. Grenfell, A. Grotsch-Noels, E. Günther, C. A. Haswell, A. P. Hatzes, G. Hébrard, S. Hekker, R. Helled, K. Heng, J. M. Jenkins, A. Johansen, M. L. Khodachenko, K. G. Kislyakova, W. Kley, U. Kolb, N. Krivova, F. Kupka, H. Lammer, A. F. Lanza, Y. Lebreton, D. Magrin, P. Marcos-Arenal, P. M. Marrese, J. P. Marques, J. Martins, S. Mathis, S. Mathur, S. Messina, A. Miglio, J. Montalban, M. Montalto, M. J. P. F. G. Monteiro, H. Moradi, E. Moravveji, C. Mordasini, T. Morel, A. Mortier, V. Nascimbeni, R. P. Nelson, M. B. Nielsen, L. Noack, A. J. Norton, A. Ofir, M. Oshagh, R. M. Ouazzani, P. Pápics, V. C. Parro, P. Petit, B. Plez, E. Poretti, A. Quirrenbach, R. Ragazzoni, G. Raimondo, M. Rainer, D. R. Reese, R. Redmer, S. Reffert, B. Rojas-Ayala, I. W. Roxburgh, S. Salmon, A. Santerne, J. Schneider, J. Schou, S. Schuh, H. Schunker, A. Silva-Valio, R. Silvotti, I. Skillen, I. Snellen, F. Sohl, S. G. Sousa, A. Sozzetti, D. Stello, K. G. Strassmeier, M. Švanda, G. M. Szabó, A. Tkachenko, D. Valencia, V. Van Grootel, S. D. Vauclair, P. Ventura, F. W. Wagner, N. A. Walton, J. Weingrill, S. C. Werner, P. J. Wheatley and K. Zwintz, The PLATO 2.0 mission, *Experimental Astronomy* **38**, 249 (November 2014).

The rotation of supermassive stars

L. Haemmerlé

Département d'Astronomie, Université de Genève, chemin des Maillettes 51, CH-1290 Versoix, Switzerland
E-mail: lionel.haemmerle@unige.ch

Supermassive stars (SMSs), with masses $> 10^5$ $M_\odot$, have been proposed as the possible progenitors of the most extreme supermassive black holes observed at redshifts $z > 6-7$. In this scenario ('direct collapse'), a SMS **accretes** at rates > 0.1 $M_\odot\,yr^{-1}$ until it collapses to a black hole via the general-relativistic (GR) instability. Rotation plays a crucial role in the formation of such supermassive black hole seeds. The centrifugal barrier appears as particularly strong in this extreme case of star formation. Moreover, rotation impacts sensitively the stability of SMSs against GR, as well as the subsequent collapse. In particular, it might allow for gravitational wave emission and ultra-long gamma-ray bursts at black hole formation, which represents currently the main observational signatures proposed in the literature for the existence of such objects. Here, I present the latest models of SMSs accounting for accretion and rotation, and discuss some of the open questions and future prospects in this research line.

Keywords: Stars: massive – stars: rotation – stars: black holes

1. Introduction

Supermassive stars (SMSs), with masses $\gtrsim 10^5$ $M_\odot$, were originally invoked to explain the high luminosity of quasars,[1,2] later understood as a consequence of accretion in the deep potential well of supermassive black holes (SMBHs).[3] The formation of these SMBHs remains, however, one of the main open problems in galaxy formation, and SMSs have been proposed as their possible progenitors.[4,5] In the last decade, a number of quasars powered by SMBHs with masses 10^9-10^{10} $M_\odot$ have been detected at redshifts 6 – 7. The most extreme case known to date is J0313-1806 a quasar at $z = 7.6$ that hosts a SMBH with mass $1-2\times10^9$ $M_\odot$.[6] At redshifts 6 – 7, the age of the Universe is a fraction of a Gyr, so that objects of $10^9 - 10^{10}$ $M_\odot$ that exist at these redshifts must have accumulated their mass at average rates $\dot{M} \gtrsim 10$ $M_\odot\,yr^{-1}$. The direct formation of SMBHs during the process of galaxy formation ('direct collapse' black holes) provides the most rapid pathway to form SMBHs with masses $\gtrsim 10^9$ $M_\odot$.[5,6] In this scenario, the direct progenitor of the SMBH is a SMS undergoing accretion at rates $\dot{M} \gtrsim 0.1$ $M_\odot\,yr^{-1}$, until it collapses due to the general-relativistic (GR) instability.[7,8]

The direct collapse of $\sim 10^9$ $M_\odot$ into a compact object requires special conditions. Fragmentation is inhibited in primordial gas due to the absence of heavy elements, but the cooling by molecular hydrogen remains efficient enough to maintain temperatures 100 – 1000 K, which is thought to result in the formation of a cluster

of stars with individual masses $M \lesssim 100$ M$_\odot$.[9] Black hole seeds formed from the collapse of such primordial clusters remain $\lesssim 1000$ M$_\odot$, and hardly explain the large masses of the earliest quasars.[6,10] Only in the presence of a strong Lyman-Werner flux the molecular hydrogen of primordial gas can be dissociated, which leads to temperatures $\sim 10^4$ K set by atomic cooling.[11] Primordial, atomically cooled halos are found to collapse isothermally with inflows 0.01 – 1 M$_\odot$ yr^{-1} sustained down to sub-parsec scales, allowing for the direct formation of a supermassive compact object in a dynamical time.[12,13] However, this scenario implies tight synchronisation between halo pairs, since only star formation in a neighbouring halo can provide the required Lyman-Werner flux.[14] Another channel of direct collapse is provided by the merger of the most massive galaxies expected to have formed at redshifts 8 – 10, with masses $\sim 10^{12}$ M$_\odot$.[15] The merger is found to trigger inflows of $\sim 10^4 - 10^5$ M$_\odot$ yr^{-1} down to a fraction of a parsec, which allows to accumulate several 10^9 M$_\odot$ in a few tens of kyr. No special chemical composition is required, since the collapse is too rapid for star formation and radiative feedback to start, and metallicities up to solar are expected in these galaxies. However, SMBH formation by direct collapse in metal-rich gas requires a SMS with mass $M \gtrsim 10^6$ M$_\odot$ in order to avoid thermonuclear explosion.[16]

The strong contraction that characterises a gravitationnal collapse leads in general to rapidly rotating structures, by a simple argument of angular momentum conservation. In fact, star formation by accretion requires strong mechanisms to remove angular momentum from the collapsing gas in order to circumvent the centrifugal barrier.[17] This angular momentum problem represents one of the main bottleneck in models of massive star formation[18–20] and plays naturally a critical role in the context of SMS and SMBH formation.[21,22] Here, I review the latest theoretical results concerning SMSs in the context of SMBH formation, focusing on the importance of rotation. In Section 2, I review the main properties of SMSs formed by accretion; I discuss the angular momentum problem in section 3; section 4 focuses on the GR instability and the maximum mass of rotating SMSs; the implication for SMBH formation by direct collapse are discussed in section 5. A summary is given in section 6.

2. Rotation in rapidly accreting SMSs

The evolution of SMSs under rapid accretion has been addressed with analytical and numerical models in the last decade.[23–25] SMSs are always close to the Eddington limit, which implies that their H-burning time is of the order of the Myr, independently of their mass. Thus, forming a SMS of mass $M \gtrsim 10^5$ M$_\odot$ by accretion before fuel exhaustion requires rates $\dot{M} \gtrsim 0.1$ M$_\odot$ yr^{-1}. Under such accretion rates, SMSs are found to evolve as 'red supergiant protostars',[24] following upwards the Hayashi limit in the Hertzsprung-Russel diagram. An example is shown in figure 1, with numerical models of Population III (Pop III) SMSs computed with the stellar evolution code GENEC.[25] The effective temperature on the Hayashi limit remains $\lesssim 10^4$ K,

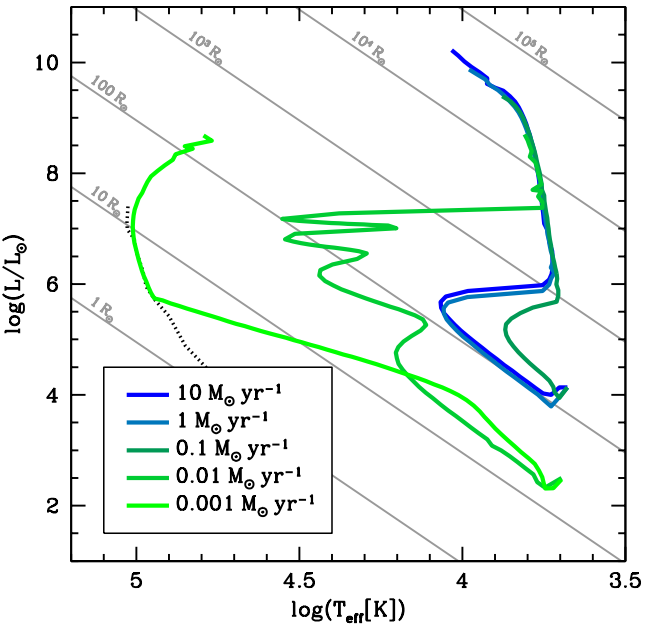


Fig. 1. Hertzsprung-Russel diagram of the GENEC models of Pop III, rapidly accreting SMSs. (Figure from.[25])

which implies negligible ionising feedback and allows for continuous accretion up to at least $M \sim 10^5$ M$_\odot$. As a consequence of the evolution along both the Eddington and the Hayashi limits, the mass-radius relation of rapidly accreting SMSs can be approximated by a simple power-law:

$$R = 260\,\mathrm{R}_\odot \left(\frac{M}{\mathrm{M}_\odot}\right)^{1/2} \tag{1}$$

The increase of the radius with the growing mass results from the fact that, at rates $\gtrsim 0.01$ M$_\odot$ yr^{-1}, the entropy advected by an element of mass ΔM at accretion cannot be radiated efficiently in the corresponding accretion time $\Delta M/\dot{M}$. Thus, the Kelvin-Helmholtz contraction of the stellar interior is too slow compared to the increase in radius that results from the accretion of new mass. As a consequence, most of the stellar interior is not thermally relaxed and maintains an outwards positive entropy gradient that stabilises the gas against convection. This is illustrated in figure 2 with the entropy profiles of the Pop III GENEC models at $\dot{M} = 1 - 100$ M$_\odot$ yr^{-1}, taken at successive stages of the evolution. The highly centralised energy production by H-burning drives convection in a core of $\sim 10\%$ of the stellar mass, that stays isentropic. An **outer** envelope is also convective (non-adiabatic), due to the high opacities at the low temperatures of the Hayashi limit. This **outer** envelope contains negligible mass ($\sim 1\%M$), but represents a significant fraction of the stellar radius. For accretion rates $\dot{M} \gtrsim 10$ M$_\odot$ yr^{-1}, the

evolutionary timescale becomes so short that the radiative envelope contracts adiabatically, compressed by the weight of the layers newly accreted.[26] In this regime, the entropy profile converges towards a simple power-law of the mass-coordinates ('hylotrope'[23]):

$$s \propto M_r^{1/2} \tag{2}$$

This power-law is shown by a grey dotted line in figures 2.

For stars near the Eddington limit, the maximum rotation velocity consistent with hydrostatic equilibrium is given by the ΩΓ-limit, that accounts for the role of radiation pressure:[28]

$$\Omega_\Gamma \simeq \Omega_K \sqrt{1-\Gamma} \tag{3}$$

where Ω_K is the Keplerian angular velocity and Γ the Eddington factor. SMSs have $\Gamma \sim 0.99$, which implies rotation velocities of $\lesssim 10\%\ \Omega_K$; in other words, SMSs must be slow rotators.[22] The maximum rotation velocity set by the ΩΓ-limit is shown in figure 3 as a function of the SMS's mass. The ratio of centrifugal to gravitational forces in the star, which is of the same order as the ratio of rotational to gravitational energies, can be estimated by

$$\frac{\Omega^2}{\Omega_K^2} \sim 1-\Gamma \lesssim 1\% \tag{4}$$

For such slow rotations, the deformation of the star by the anisotropic centrifugal force remains negligible. Moreover, due to the short evolutionary timescales, meridional currents and shear diffusion have no significant impact on the angular momentum distribution. Thus, the radiative layers contract with local angular momentum conservation, which translates into highly differential rotation, with a rotation frequency several orders of magnitude larger in the core than at the surface.[22] The break-up velocity is not reached in the core, thanks to the large Keplerian velocity in the densest regions.

3. The angular momentum problem

Because of the slow rotations imposed by the ΩΓ-limit, SMS formation by accretion implies low angular momentum in the accreted gas. This is illustrated in figure 4, that shows the evolution of a Pop III SMS accreting at $\dot{M} = 1\ \mathrm{M_\odot\,yr^{-1}}$ under the assumption of maximal rotation. Rotation is post-processed from the non-rotating GENEC structures shown in the upper panel. Angular momentum is advected at accretion in order to satisfy the constraint from the ΩΓ-limit (third panel) at each time-step, and the resulting angular momentum accretion history is shown in the fifth panel as a blue line. It remains 2 – 3 orders of magnitude below the Keplerian angular momentum (grey line), which means that, for the ΩΓ-limit to be satisfied all along the accretion phase, the angular momentum advected at accretion cannot exceed a fraction $f \sim 0.1 - 1\%$ of the Keplerian angular momentum. In other

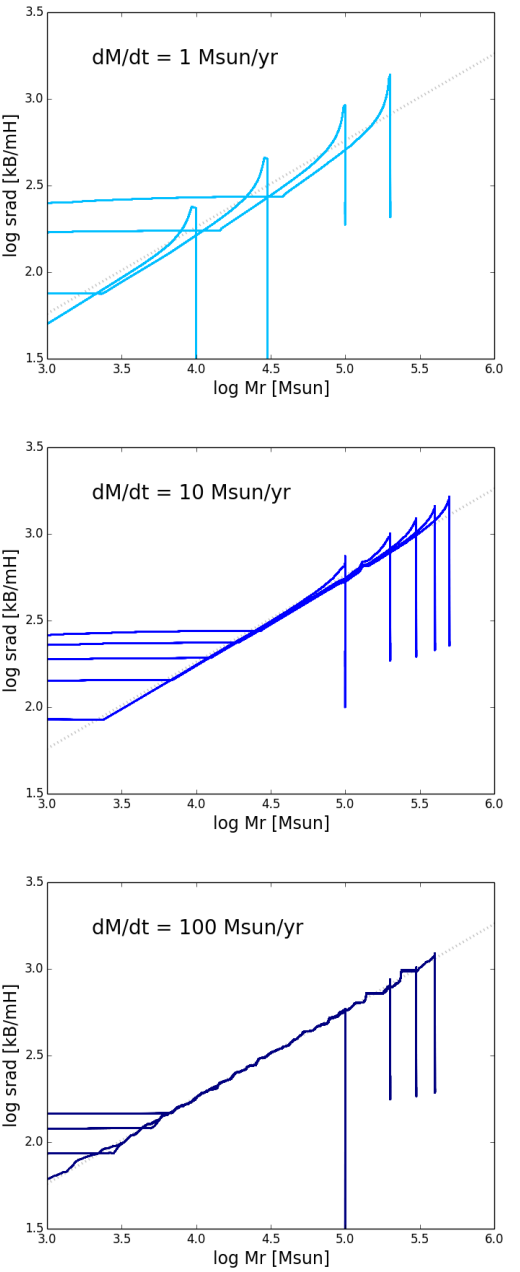


Fig. 2. Entropy profiles of the GENEC models of Pop III SMSs accreting at $\dot{M} = 1{-}100\ \mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$. The grey dotted line indicates the power-law (2). (Figure from.[27])

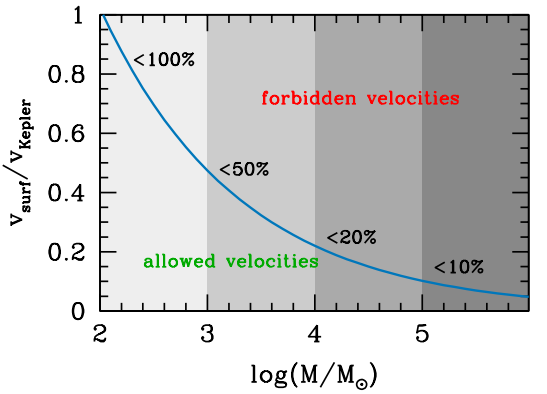


Fig. 3. Maximum rotation velocity of SMSs as a function of their mass (figure from[22]).

words, SMS formation by accretion from a Keplerian disc requires mechanisms efficient enough to remove more than 99% of the angular momentum from the disc. Notice that the fraction f is $1-2$ orders of magnitude smaller than the ratio Ω/Ω_K imposed by the $\Omega\Gamma$-limit, which reflects the instantaneous redistribution of angular momentum in the **outer** convective envelope visible in the upper panel of figure 4. **Indeed, without convection, the ratio Ω/Ω_K in the surface layer would be simply given by the fraction f of angular momentum advected by this layer. But here, the surface velocity is given by the angular momentum accreted plus that received by convection from the deeper layers of the envelope. Thus, a given constraint on the ratio Ω/Ω_K (e.g. $\lesssim 10\%$) translates into a tighter constraint on the ratio f (e.g. $\lesssim 0.1-1\%$).**

The angular momentum problem is general to star formation.[17] For low-mass stars, the problem is easily solved by the magnetic coupling of the star with its disc and winds, thanks to the convective dynamo in the stellar envelope. But the problem becomes more difficult for massive stars, since these stars have no convective envelope, and only a few percent of them have detected magnetic fields, thought to be fossil fields.[30] The most efficient processes to extract angular momentum in the context of massive star formation rely on the gravitationnal instabilities that develop in massive accretion discs.[31] Since these gravitationnal instabilities are enhanced by high densities, their efficiency increases when the centrifugal barrier is reached, as gas accumulates in the vicinity of the star. Thus, these processes allow for a self-regulated, efficient angular momentum transport all along the collapse.

In the case of rapidly accreting, maximally rotating SMSs, the surface magnetic field required to remove the angular momentum excess at accretion from a Keplerian disc by star-winds coupling is of the order of 10 kG, as shown in the bottom panel of figure 4. The corresponding Alfvén radius is shown as a red dashed line in the top panel. According to estimates based on the Rossby number, a convective dynamo in the envelope of the SMS cannot produce a field larger than 1 G. Only fossil

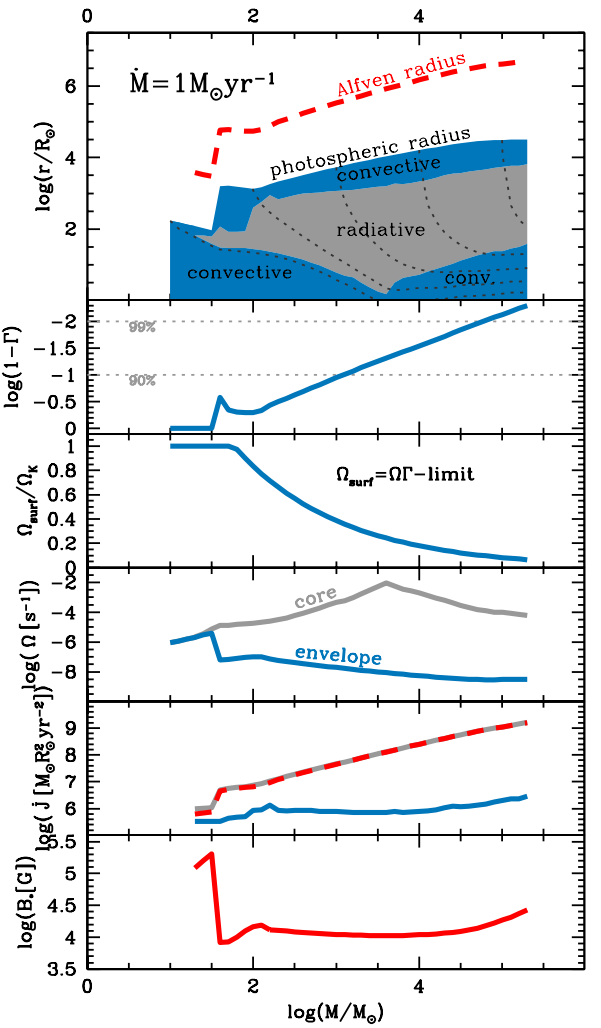


Fig. 4. GENEC model of a maximally rotating Pop III SMS accreting at $\dot{M} = 1\ \mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$. The horizontal axis represents the stellar mass, which is a time coordinate in case of constant accretion. The first four panels show the stellar structure in radial coordinates, the Eddington factor, the ratio of the surface velocity to the Keplerian velocity, and the angular velocity in the core and the envelope. The fifth panel shows the accreted angular momentum (in blue), compared to the Keplerian angular momentun (in grey). The difference is shown by the red dashed curve. The bottom panel shows the surface magnetic field required to remove this difference of angular momentum by star-wind coupling. (Figure from.[29])

fields from primordial origin could reach the required values. On the other hand, gravitational instabilities are particularly strong in the massive accretion discs of SMSs, and are thought to be efficient enough to circumvent the centrifugal barrier in the context of SMBH formation.[21,32]

4. GR instability

The formation of a SMBH by direct collapse occurs when the progenitor SMS becomes dynamically unstable. SMSs are expected to collapse via the GR instability,[33] which is a pulsation instability that arises from the GR corrections to the equation of radial momentum. For a non-rotating post-Newtonian SMS, the eigenfrequency ω of adiabatic pulsations can be obtained with[34]

$$\omega^2 I = \int \beta P \mathrm{d}V - \int \left(\frac{2GM_r}{rc^2} + \frac{8}{3}\frac{P}{\rho c^2} \right) \frac{GM_r}{r} \mathrm{d}M_r \tag{5}$$

where r is the radial distance from the stellar centre, M_r the mass-coordinates (i.e. the mass contained in all the layers below r), ρ the mass density, P the pressure, β the ratio of gas pressure to total pressure, I the moment of inertia, V the volume of the sphere r, G the gravitational constant and c the speed of light. Notice that in the post-Newtonian approximation, distinctions between rest-mass and relativistic-mass are irrelevant.

The GR instability is reached when ω becomes imaginary, which implies that exponentials are solutions to the equations of motion. Equation (5) expresses this condition in terms of the two integrals at the right-hand side. The first one is Newtonian and positive, and scales with β, that is with the departures from the Eddington limit, at which all the pressure support is provided by radiation. As the SMS grows in mass, it approaches the Eddington limit asymptotically with $\beta \to 0$, but $\beta > 0$. Thus, without any additional term, equation (5) would imply that the star remains always stable. However, the second term, that scales with the GR corrections $2GM_r/rc^2$ and $P/\rho c^2$, is negative. As the SMS grows in mass, these corrections grow until the sum of the integrals become negative. At this point, the star is dynamically unstable, which leads in general to the direct formation of a SMBH in a dynamical time.[35,36]

The exact mass at which the SMS reaches the GR instability depends on its structure, in particular on the mass of its convective core. The maximum masses of non-rotating Pop III SMSs as a function of the mass of their convective core are shown in figure 5. The black-blue-cyan tracks are the evolutionary models of Pop III SMSs computed with GENEC under accretion at the indicated rates.[25,26] These models were computed under the assumption of hydrostatic equilibrium, which makes them insensitive to dynamical instabilities, and the ends of the runs are arbitrary, imposed by the numerical stability of the models. The onset of the GR instability on these hydrostatic models, determined with equation (5), is indicated by star-like symbols. It is reached before the end of the run only for rates 1 and 10 $\mathrm{M}_\odot\,\mathrm{yr}^{-1}$. The limitations due to the numerical instability of hydrostatic models with rapid accretion can be circumvented with the use of semi-analytical models built on the hylotropic law of equation (2).[23] With these structures, the limit of stability in the (M_{core}, M)-diagram is uniquely given by the central temperature of the star and its chemical composition. The central temperature of SMSs is well constrained

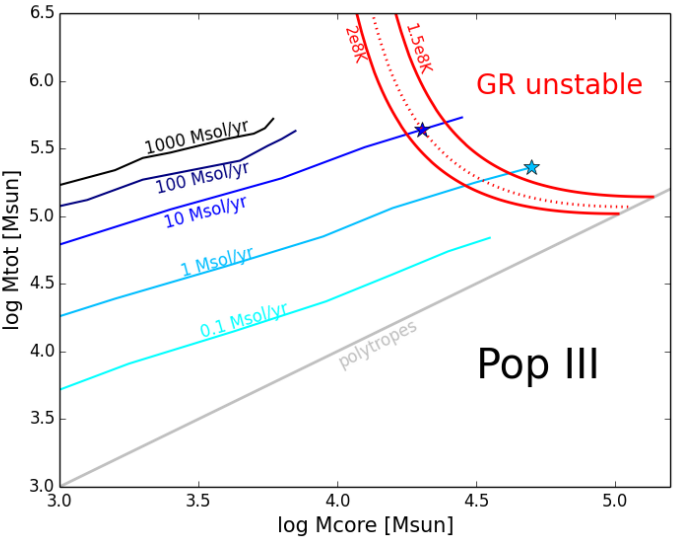


Fig. 5. Maximum masses of non-rotating Pop III SMSs consistent with GR stability, as a function of the mass of their convective core. The red curves show the limit of stability derived from hylotropic structures with the indicated central temperatures. The black-blue-cyan tracks are Pop III GENEC models computed with the indicated accretion rates. The star-like symbols indicate the point of GR instability of these models, when it is reached. (Figure from.[27])

by numerical models, and remains in a narrow range due to the thermostatic effect of H-burning ($T_c = 1.5 - 2 \times 10^8$ K for Pop III SMSs; $T_c = 0.6 - 0.9 \times 10^8$ K for Pop I SMSs). The limits of stability for Pop III SMSs derived from the hylotropic models are shown as red curves in figure 5. The dotted red curve is the stability limit for the exact central temperature of the last stable model of the GENEC track at $\dot{M} = 10$ M$_\odot$ yr^{-1}. We see that the hylotropic limit is already very precise for this rate, in spite of the small departures from the hylotropic profiles visible in figure 2. The entropy profile in the inner part, where the density is high and GR corrections are significant, is already well approximated by the hylotropic law (2) for this rate. Only for rates $\lesssim 1$ M$_\odot$ yr^{-1} the GENEC models departs significantly from the hyloptropic law. But even in this case the discrepancies in the final mass remains of ~ 0.1 dex, that is ~ 20 %. Thus, the limit of stability given by the hylotropic models remains a good approximation for any accretion rate. Moreover, the numerical simplicity of these models allows for a broader view of the parameter space. In particular, we see that for the central temperatures set by H-burning the GR instability cannot be reached if the stellar mass does not exceed $\gtrsim 10^5$ M$_\odot$. Moreover, the GR instability always requires the convective core to contain a mass $\gtrsim 10^4$ M$_\odot$.

Rotation plays a prominent role in the stability of SMSs. Near the GR instability, the dimensionless ratios β, $2GM_r/rc^2$ and $P/\rho c^2$ are all of the order of a percent.

It implies that the positive and negative integrals at the right-hand side of equation (5) represent typically a percent of the total internal and gravitational energies of the star, respectively. Thus, other small effects, that have negligible impact on the hydrostatic structure of SMSs, can play a critical role on the dynamical stability of these structures. This is the case for rotation: the slow rotation velocities ($\sim 0.1\ \Omega_K$) consistent with the $\Omega\Gamma$-limit correspond to rotational energies of the order of $\Omega^2/\Omega_K^2 \sim 1\%$ of the gravitational energy of the star (equation 4). The rotational corrections to equation (5) are thus of the same order as the β- and GR-terms. For these slow rotations and small GR corrections, however, the relativistic rotation terms are expected to be of second order, and can be neglected in the determination of the stability limit. Including the Newtonian rotation term to equation (5), we obtain[34]

$$\omega^2 I = \int \beta P \mathrm{d}V + \frac{4}{9}\int \frac{\Omega^2}{\Omega_K^2}\frac{GM_r}{r}\mathrm{d}M_r - \int\left(\frac{2GM_r}{rc^2} + \frac{8}{3}\frac{P}{\rho c^2}\right)\frac{GM_r}{r}\mathrm{d}M_r \qquad (6)$$

The positive rotation term contributes to the stability of the star, and allows it to increase its mass by orders of magnitude beyond the non-rotating limit.[34,37]

The limits of stability for rotating Pop III and Pop I SMSs are shown in figures 6 and 7, respectively.[34] The limits have been derived from hylotropic structures, on which rotation profiles are defined by local angular momentum conservation in the radiative envelope and solid rotation in the convective core. Each layer of the radiative envelope is assumed to keep the angular momentum it advected at accretion, defined as a faction f of the Keplerian angular momentum at the accretion radius given by equation (1). The angular momentum of the core is then given by that advected from the envelope. We see that the limits start to be shifted by rotation as soon as $f \gtrsim 0.1\%$, while for $f \sim 1\%$ they are already shifted by 3 orders of magnitudes towards larger masses. The lower central temperatures of the Pop I models allows for slightly larger masses when $f \sim 0.1\%$, but when $f \sim 1\%$ the limits do not depend anymore on the thermal properties of the star, and are uniquely given by f. In this regime, indicated by the red areas on the diagrams, the gas term in equation (6) becomes negligible compared to the rotation and GR terms, which means that rotation becomes the main stabilising agent against the destabilising GR corrections. An interesting consequence is that the profile of the spin parameter at the onset of the instability, which is key regarding black hole formation, becomes 'universal', that is, uniquely given by the mass-fraction of the convective core.

For typical conditions of atomically cooled halos (Pop III, $\dot{M} \lesssim 1\ \mathrm{M_\odot\,yr^{-1}}$), the $\Omega\Gamma$-limit imposes $f \lesssim 0.1 - 0.2\%$ (figure 4). In this case, rotation can increase the final stellar mass by up to a factor ~ 2, which does not allow to exceed $10^6\ \mathrm{M_\odot}$. As mentioned in section 3, this low value of f results from the instantaneous angular momentum transport in the deep convective envelope. Interestingly, the depth of the envelope decreases significantly for larger rates, and values $f \sim 1\%$ appear consistent with the conditions of merger-driven direct collapse. In this formation channel, masses up to $10^8 - 10^9\ \mathrm{M_\odot}$ could be reached before the GR instability.

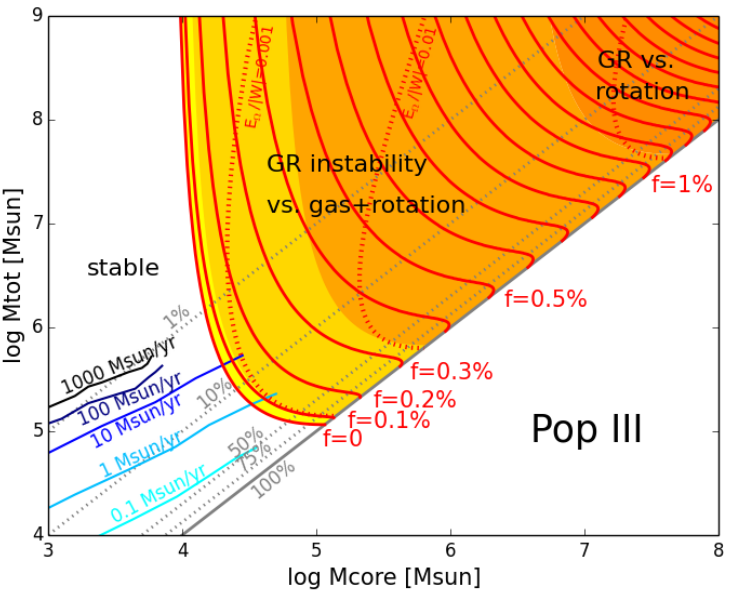


Fig. 6. Limits of GR stability for rotating Pop III SMSs, derived from the hylotropic models with $T_c = 1.8 \times 10^8$ K. The diagram represents the mass of the SMS as a function of the mass of its convective core. The various limits are obtained with the indicated values of the fraction f of accreted angular momentum over the Keplerian angular momentum (solid red lines). The limits are also shown for constant ratios of rotational to gravitational energies (dotted red lines). The coloured areas denote the relative importance of the β term (gas term) and the rotation term in the stability of the star (equation 6). The black-blue-cyan tracks are the same GENEC models as in figure 5. (Figure from.[34])

Since masses $> 10^6$ $M_\odot$ are required for SMBH formation in the Pop I case, the black hole seeds in merger-driven direct collapse must have masses $10^6 - 10^9$ $M_\odot$, which implies that the different channels of direct collapse lead to distinct ranges of seed's mass.

5. SMBH formation and observational signatures of direct collapse

The final collapse of SMSs has been followed with a large number of GR magneto-hydrodynamical simulations.[16,35,36] In the Pop III case, the collapse is found to lead to the direct formation of a SMBH, that contains $\gtrsim 90\%$ of the total stellar mass. A fraction $\sim 1 - 10$ % of the mass can remain in orbit outside the horizon at the formation of the black hole, provided rapid enough rotation. This gas further collapses through the horizon, with possible gravitational wave emission detectable by future space-based observatories. Moreover, in the presence of magnetic fields, relativistic jets could be launched and trigger ultra-long gamma-ray bursts. Thus, the final collapse of SMSs provides the most promising possibilities for observational signatures of direct collapse and the existence of SMSs.

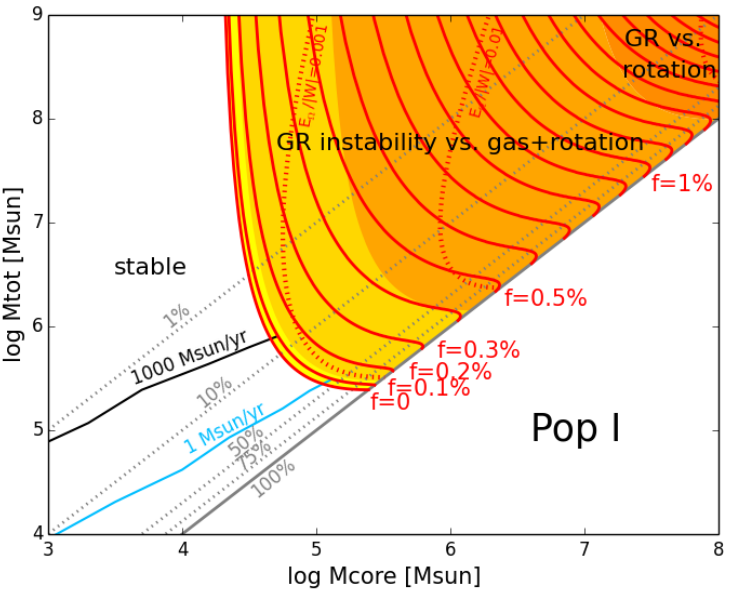


Fig. 7. Same as figure 6 for the Pop I case ($T_c = 0.85 \times 10^8$ K). (Figure from.[34])

Rotation is key since only the centrifugal barrier allows for the survival of a slowly decaying torus, and prevents the direct engulfment of the whole stellar mass by the black hole in a dynamical time. Figures 8 and 9 show the profiles of the spin parameter cJ_r/GM_r^2 (where J_r is the angular momentum of the mass M_r) for the hylotropic models of figures 6 and 7, taken at the onset of the instability. They correspond to typical conditions of atomically cooled halos and merger-driven direct collapse, respectively. In each case, the profiles are shown for three different core mass. The convective core is recognisable as the only region where the profiles departs from each other, as a result of the angular momentum transport by convection. In contrast, the profiles match each other in the radiative envelope, reflecting the angular momentum accretion law. We see that the solid rotation of the core maintains the angular momentum distribution decentralised. Notice that the profiles of figure 9 correspond to the 'universal' profiles obtained when rotation becomes the dominant stabilising agent.

From these rotation profiles, the fraction of the **outer** mass that has enough angular momentum to remain in orbit at the formation of the black hole can be estimated analytically by comparison with the angular momentum of the innermost stable circular orbits (ISCOs).[38] It is found to be significant only for the conditions of merger-driven direct collapse, and for fully convective SMSs, as indicated by the red squares in figures 8 and 9. The low angular momentum ($f \sim 0.2\%$) required in the context of atomically cooled halos prevents the formation of an orbiting structure, and even for $f \sim 1\%$ convection is required in the whole stellar interior for the

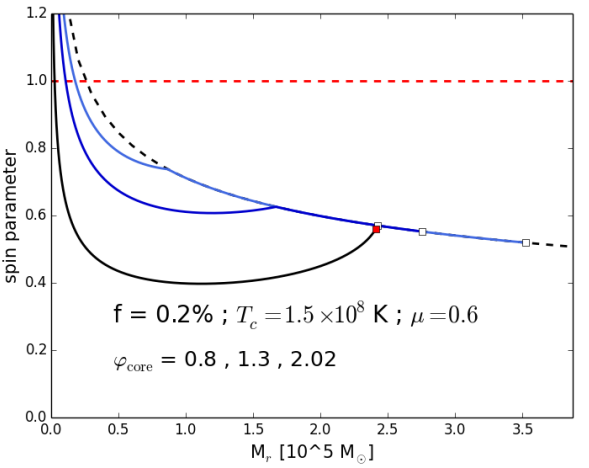


Fig. 8. Profiles of the spin parameter cJ_r/GM_r^2 at the onset of GR instability, for typical conditions of atomically cooled halos. The profiles (solid lines) are shown for three hylotropes with different **core mass-fraction $M_{\rm core}/M$: the black profile corresponds to a fully convective star $M_{\rm core} = M$ (polytrope), the dark blue profile corresponds to $M_{\rm core}/M = 60\%$ and the light blue profile to $M_{\rm core}/M = 25\%$.** The white squares indicate the surface **(i.e. the total mass)** of the three models **(the polytrope has the lowest mass, $M_{\rm core}/M = 25\%$ the highest)**. Only for the fully convective model a comparison with ISCO angular momentum suggests that a tiny fraction of the **outer** mass ($\sim 1\%M$) remains in orbit outside the horizon at black hole formation, which is indicated by the red square. The black dashed line shows the angular momentum accretion law. The red dashed line indicates the maximum value of the spin parameter consistent with Kerr black holes. (Figure from.[34])

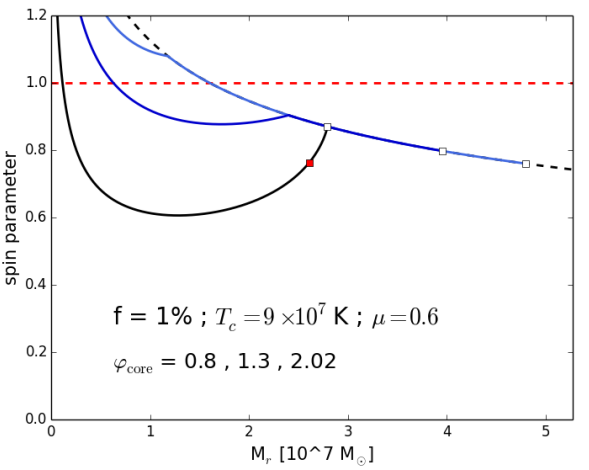


Fig. 9. Same as figure 8 for typical conditions of merger-driven direct collapse. Here, the **outer** mass fraction of the fully convective model that remains in orbit at black hole formation represents $\sim 10\%$ of the stellar mass (red square). (Figure from.[34])

angular momentum to be sufficiently decentralised. Overall, these models suggest that gravitational wave emission and ultra-long gamma-ray bursts at black hole formation are consistent with the conditions of galaxy mergers, but not with those of atomically cooled halos. In both scenarios, the centrifugal barrier is inefficient to prevent the direct formation of a massive black hole seed at the collapse of the supermassive stellar progenitor.

6. Summary and conclusions

SMSs represent promising candidates for the progenitors of the SMBHs that power the most massive quasars detected at redshifts $6-7$, since direct collapse appears as the most efficient scenario to form SMBHs in short timescales. This scenario faces several bottlenecks, however, but solutions have been proposed to all of them. The angular momentum problem represents certainly one of the strongest, as discussed in section 3.

In spite of the low spins imposed by the $\Omega\Gamma$-limit (section 2, figure 3), rotation impacts critically the life and death of SMSs. As an outstanding effect, it allows rapidly accreting SMSs to increase their final masses by several orders of magnitude compared to the non-rotating case, before they reach the GR instability and collapse into a SMBH (section 4, figures 6-7). According to hylotropic models, the masses of the black hole seeds belong to distinct ranges depending on the direct collapse channel:

- $10^5\ \mathrm{M}_\odot \lesssim M \lesssim 10^6\ \mathrm{M}_\odot$ for atomically cooled halos;
- $10^6\ \mathrm{M}_\odot \lesssim M \lesssim 10^9\ \mathrm{M}_\odot$ for merger-driven direct collapse.

Moreover, the conditions of merger-driven direct collapse are expected to be more favourable than those of atomically cooled halos for gravitational wave emission and ultra-long gamma-ray bursts at the collapse of the SMS (section 5, figures 8-9).

Many uncertainties remain, however, in these scenarios, most of them relying on the lack of consistency of the models for the formation of the black hole seed itself. Due to their hypothetical nature, SMSs have been much less studied than other classes of stars, and the picture given here deserves to be refined by additional models accounting for full stellar evolution up to the largest masses and accretion rates, including a self-consistent treatment rotation, multiplicity and radiative feedback. In particular, reliable theoretical predictions for the observational signatures of SMBH formation by direct collapse require precise determination of the properties of the stellar progenitor. In that respect, the future generation of space-based telescopes and gravitational wave observatories could unveil the most massive stars ever formed in Universe's history. Interestingly, according to the models described here, the maximum mass of these stars is set by **the interplay between general relativity and** rotation.

References

1. F. Hoyle, W. A. Fowler, 1963a, Nature , 197, 533.
2. F. Hoyle, W. A. Fowler, 1963b, MNRAS , 125, 169.
3. D. Lynden-Bell, 1969, Nature , 223, 690.
4. M. J. Rees, 1984, ARA&A , 22, 471.
5. M. Volonteri, M. C. Begelman, 2010, MNRAS , 409, 1022.
6. F. Wang, J. Yang, X. Fan, et al. 2021, ApJ , 907, L1.
7. T. E. Woods, B. Agarwal, V. Bromm, et al., 2019, PASA , 36, e027.
8. L. Haemmerlé, L. Mayer, R. S. Klessen, et al., 2020, SSR, 216, 48.
9. R. Klessen, 2019, Formation of the first stars, ed. M. Latif & D. Schleicher, 67–97.
10. A. Lupi, M. Colpi, B. Devecchi, G. Galanti, M. Volonteri, 2014, MNRAS , 442, 3616.
11. K. Omukai, 2001, ApJ , 546, 635.
12. M. A. Latif, D. R. G. Schleicher, W. Schmidt, J. C. Niemeyer, 2013, MNRAS , 436, 2989.
13. S. Chon, T. Hosokawa, N. Yoshida, 2018, MNRAS , 475, 4104.
14. J. A. Regan, E. Visbal, J. H. Wise, et al., 2017, Nature Astronomy, 1, 0075.
15. L. Mayer, S. Bonoli, 2019, Reports on Progress in Physics, 82, 016901.
16. P. J. Montero, H.-T. Janka, E. Müller, 2012, ApJ , 749, 37.
17. P. Bodenheimer, 1995, ARA&A , 33, 199.
18. H. Lee, S.-C. Yoon, 2016, ApJ , 820, 135.
19. L. Haemmerlé, P. Eggenberger, G. Meynet, et al., 2017, A&A , 602, A17.
20. S. Z. Takahashi, K. Omukai, 2017, MNRAS , 472, 532.
21. M. C. Begelman, M. Volonteri, M. J. Rees, 2006, MNRAS , 370, 289.
22. L. Haemmerlé, T. E. Woods, R. S. Klessen, A. Heger, D. J. Whalen, 2018a, ApJ , 853, L3.
23. M. C. Begelman, 2010, MNRAS , 402, 673.
24. T. Hosokawa, H. W. Yorke, K. Inayoshi, K. Omukai, N. Yoshida, 2013, ApJ , 778, 178.
25. L. Haemmerlé, T. E. Woods, R. S. Klessen, A. Heger, D. J. Whalen, 2018, MNRAS , 474, 2757.
26. L. Haemmerlé, G. Meynet, L. Mayer, et al., 2019, A&A , 632, L2.
27. L. Haemmerlé, 2020, A&A , 644, A154.
28. A. Maeder, G. Meynet, 2000, A&A , 361, 159.
29. L. Haemmerlé, G. Meynet, 2019, A&A , 623, L7.
30. J. H. Grunhut, G. A. Wade, C. Neiner, et al., 2017, MNRAS , 465, 2432.
31. T. Hosokawa, S. Hirano, R. Kuiper, et al., 2016, ApJ , 824, 119.
32. J. H. Wise, M. J. Turk, T. Abel, 2008, ApJ , 682, 745.
33. S. Chandrasekhar, 1964, ApJ , 140, 417.
34. L. Haemmerlé, 2021, A&A , 650, A204.
35. L. Sun, V. Paschalidis, M. Ruiz, M., S. L. Shapiro, 2017, Phys. Rev. D , 96, 043006.
36. L. Sun, M. Ruiz, M., S. L. Shapiro, 2018, Phys. Rev. D , 98, 103008.
37. W. A. Fowler, 1966, ApJ , 144, 180.
38. T. W. Baumgarte, S. L. Shapiro, 1999, ApJ , 526, 937.

Cosmology with high-redshift FRBs

A. Fialkov

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
E-mail: afialkov@ast.cam.ac.uk

The quest for high redshift Fast Radio Bursts (FRBs) is ongoing with telescopes such as the Five-hundred-meter Aperture Telescope (FAST) and the Green Bank Telescope (GBT) looking for highly dispersed events. If FRB-producing systems exist at early times, such sources would provide new unique ways to probe cosmic dawn and Reionization. On one hand, FRB dispersion would allow us to probe the history and topology of Reionization. On the other hand, number counts of high redshift FRBs would indirectly probe galaxy formation at high-redshifts. Here we consider prospects of advancing our understanding of the first billion years of cosmic history with high-redshift FRBs.

Keywords: High redshift Universe, Fast Radio Bursts

1. Introduction

The field of cosmology experienced breathtaking growth in the past several decades. Owing to the measurements of the Cosmic Microwave Background (CMB) by telescopes such as WMAP,[1] Planck[2] and others, which reveal a snapshot of the Universe shortly after the Big Bang, we now know that the Universe is well-described by a rather simple six-parameter model (ΛCDM) in which the total energy balance is share between dark energy (Λ), dark matter (that is usually assumed to be cold), baryons and radiative degrees of freedom.

The CMB is also sensitive to the process of reionization. The Epoch of Reionization (EoR) is often referred to as the last phase transition in the history of the evolution of the Universe.[3,4] During the EoR the ultraviolet light produced by young stars in early populations of galaxies ionized the previously neutral intergalactic medium (IGM) turning it into hot plasma. The EoR imprints faint traces in the CMB via the scattering of CMB photon off the ionized gas. Therefore, although the effect is very weak, the CMB observations can be used to constrain this epoch. The redshift of reionization, as inferred from the CMB data, is $\approx 7.5 \pm 0.82$.[2]

The CMB observations have greatly advanced our understanding of cosmology. Given the richness of information extracted from the CMB and the amount of scientific discoveries stemming from these data, it is surprising to realize that the CMB probes only a thin (nearly 2-dimensional) volume of the Universe located at $z \sim 1100$ when the Universe was only 0.4 million years old (for comparison it is 13.8 billion years old today). However, the CMB observations are only weakly sensitive to the processes at later times. Located between the CMB last scattering

surface and us are multiple layers of cosmic history harbouring events such as the formation of first dark matter halos massive enough to accret and cool gas, formation of the first stars and galaxies, and other processes, which cannot be probed by the CMB.

By probing the late-time Universe, large scale structure and galaxy surveys provide complimentary information to what is found in the CMB. These measurements are of the evolved cosmic web and mature populations of galaxies (for example, the Dark Energy Survey measured ~ 300 million galaxies[5]). Much like the CMB these data helped to advance our understanding of the Universe. We now know that the expansion of the Universe is accelerated, low-redshift observations provide ample opportunities to test General Relativity across a wide range of regimes, and we can test galaxy formation theories with data.

However, despite this impressive effort, we do not have the whole picture of how the Universe transitioned from the pristine primordial state observed in the CMB to the evolved state in which it is found today. Even the state-of-the-art observations are limited and can only probe a small part of cosmic history. Large surveys mostly cover our cosmic neighbourhood (out to $z \sim 3$) with only the brightest objects observed at higher redshifts (although using clusters of galaxies as gravitational lenses can help to grasp a more representative picture of the high-redshift Universe[6]). In addition, some of the low redshift observations are inconsistent with the CMB data (e.g. the H_0 discrepancy[7]), and so the true state of the Universe remains ambiguous. Moreover, the elegant and simple ΛCDM theory itself suffers from several shortcomings. To name a few: the nature of dark matter and dark energy, which together compose $\sim 95\%$ of the total energy density in the Universe, is undetermined; not all baryons that are measured from the CMB are accounted for at lower redshifts (*the missing baryon problem*); we have already mentioned the infamous H_0 tension; galaxy formation is not fully understood as well as the structure and composition of the IGM. Finally, the first billion years of cosmic history remain poorly explored.

FRBs, transient radio signals of yet unestablished nature, might result to be unique probes of the EoR, thus advancing our understanding of the first billion years of cosmic history. High-redshift FRBs, if exist, would probe the EoR in a unique way compared to the established tracers such as the high-redshift quasars, galaxies and the CMB. Thus, if a sample of high-redshift FRBs is collected, they will offer new ways to probe cosmology.

FRBs are significantly dispersed by the ionized medium distributed along the path between the sources and the observer. For most of the detected FRBs, the observed dispersion of FRB pulses, quantified in terms of dispersion measure (DM) is much larger than the contribution of our local environment. Therefore, it is now known that the overwhelming majority of FRBs are of extra-Galactic origin. For a typical FRB the largest contribution to the dispersion is from the ionized intergalactic medium, $DM^{\rm IGM}$, which depends on the number density of free electrons along the line of sight. For FRBs coming from redshift z, the mean IGM contribution is

given by

$$\overline{DM}^{\rm IGM}(z) = \int_0^z \frac{n_e(z')}{1+z'} dl = \int_0^z \frac{c n_e(z')}{H(z')(1+z')^2} dz', \quad (1)$$

where the first integral is along the line of sight with dl being the line element. $n_e(z')$ is the mean number density of free electrons at redshift z', c the speed of light and $H(z')$ is the Hubble expansion rate. The electron density directly depends on the ionized fraction of hydrogen $x_i(z)$, which at high redashifts traces the EoR. We will refer to the redshift dependence of the ionized fraction, $x_i(z)$, as the *EoR history*. Dependence of $DM^{\rm IGM}$ on redshift is illustrated in Figure 1 (image credit: A. Tocher).

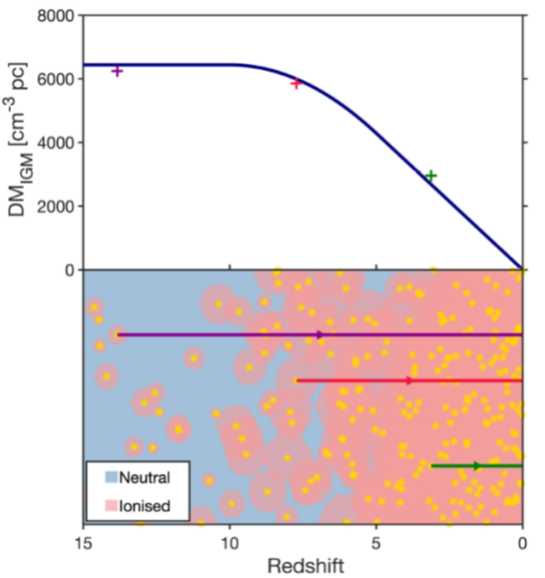


Fig. 1. Dependence of $DM^{\rm IGM}$ on redshift (top) and cartoon of the corresponding EoR history (bottom). Blue regions are the neutral patches, red regions are ionized, yellow markers represent star forming regions. Coloured crosses (top) correspond to three randomly generated FRBs and coloured lines (bottom) are the corresponding lines of sight through the EoR. Illustration credit: A. Tocher.

At low redshifts FRBs have already proven to be valuable cosmological probes. Dependence of $DM^{\rm IGM}$ on the cosmological redshift, z, of the FRB hosts (e.g., see Figure 1) could pin down properties of the IGM[9,11,12] and probe cosmological parameters.[10,11,13,15–21] To give an example, with a sample of localized low-redshift FRBs (host redshifts of $z \lesssim 0.66$), Macquart et al. (2020)[11] showed that the dispersion measure of these transients depends on the baryonic content of the Universe (along the line of sight) and, thus, can help identifying the missing baryons in the IGM by counting the total number of baryons along the line of sight. For

higher redshift FRBs, theoretical predictions show that a population of FRBs at $z \sim 3-4$ could be used to constrain helium reionization occurring around redshift $z \sim 3.5$,[22–26] while signals from $z > 5$ would primarily trace the evolution of the hydrogen ionization state, and, thus, probe the EoR.[10]

Constraining the high-redshift Universe with FRBs might be observationally plausible as the existing telescopes such as the Five-hundred-meter Aperture Telescope (FAST) and the Green Bank Telescope (GBT) can in principle detect bright signals out to $z \sim 10$, i.e. deep into the EoR.[8,14] The future Square Kilometre Array (SKA) will be capable to map the FRB sky detecting thousands of sources from early times.[10,33]

2. Tracing reionization with FRBs

If found, high redshifts FRBs could be used as probes of the EoR owing to the dependence of $DM^{\rm IGM}$ on $n_e(z')$ and, through it, on the EoR history, $x_i(z)$.

Existing constraints on the EoR are very weak. The CMB optical depth, defined as

$$\tau = \int \sigma n_e dl, \tag{2}$$

where σ is the scattering cross-section, is measured from the CMB data (e.g. state-of-the-art constraints from *Planck*[2]) with the relative error being at the level of $6-10\%$. This large uncertainty in τ degrades precision with which other cosmological parameters are measured (e.g. see the discussion by Liu et al. (2015)[27]). For example, the effect of the EoR on the CMB is degenerate with the amplitude of primordial fluctuations from inflation A_s. Additional constraints on the EoR come from quasar observations. For instance, using the dark fraction of Lyman α and β forest observations in tandem with Planck measurements Greig & Mesinger (2017)[28] constrain the redshift of the midpoint of reionization to $\sim 12\%$ (at 68% confidence level). Observations of the hydrogen 21-cm line are predicted to eventually provide precise measurement of τ as well as map the entire history of reionization. However, currently, the EoR constraints derived from the first upper limits on the 21-cm signal are very weak and only rule out the most extreme scenarios.[29,30]

Prospects of mapping the EoR with high-redshift FRBs has been discussed in literature.[10,31–34] Both $DM^{\rm IGM}$ and τ depend on the same physical quantity, namely the mean number density of free electrons weighted differently with redshift (compare between Eqs. 1 and 2). Therefore, it should be plausible to extract reionization history from the dispersion measure of high-redshift FRBs if a large enough sample with host localizations is obtained. Existing works on this topic (apart from the recent publication by Heimersheim et al. (2021)[35] which we highlight here), make assumptions about the reionization history that either lead to biases in the results or limit the constraining power of the data. In an earlier work we[10] were the first to realize that FRBs could be used as tracers of the reionization history EoR, and

might allow us to significantly tighten the constraints on τ. We performed a simplified estimation of the relative error in τ showing that the expected accuracy is much better than the current state-of-the-art, e.g. from *Planck*. This work stimulated further studies (a similar analysis was later carried out by Beniamini et al. (2020)[31]). Other examples of EoR constraints include the works by Beniamini et al. (2020)[31] and Hashimoto et al. (2021)[33] who proposed probing the EoR history by grouping FRBs into redshift bins and considering the difference between mean $DM^{\rm IGM}$ in adjacent bins. Due to the binning procedure, information on each individual FRB is lost and with it the possibility to smoothly trace the reionization history. In addition, in the absence of a defined $x_i(z)$ relation, this approach does not allow us making use of dependences (such as the monotonic growth of $x_i(z)$ with time) across cosmic time.

Details of reionization history $x_i(z)$ affect the dependence of FRB dispersion measure on redshift,[10] i.e. the functional form of $DM^{\rm IGM}$ (example shown in Figure 1). This dependence can, therefore, be used to extract the EoR history if a sample of high redshift FRBs (from before the end of the EoR at $z \sim 5$) is observed and localized. When extracting the EoR history from a sample of high-redshift FRBs it is important to keep a model-independent approach and stay agnostic about the true history of reionization. In Heimersheim et al. (2021)[35] we showed that assuming specific shape of $x_i(z)$ leads to biases in the derived EoR parameters. For example, we find that fixing the reionization history to the symmetric tanh function, parameterized by the duration and midpoint of the EoR, underestimates the value of τ by $\sim 10\%$.

2.1. *Bias-free constraints on Reionization*

Inspired by the model-independent approach adopted in the CMB community, in Heimersheim et al. (2021)[35] we have investigated the possibility of deriving the EoR history from the mock data without assuming a-priory knowledge of its redshift dependence. We constructed a mock catalogue of $DM^{\rm IGM}$, host redshifts and associated uncertainties (see details in Heimersheim et al. (2021)[35]). An illustration of our mock FRB catalogue is shown in Figure 2 (illustration credit: S. Heimersheim).

In the base of our analysis pipeline is the free-form "FlexKnot" method[36] where the entire EoR history is defined by a series of free-moving knots (rather than by a specific functional form). For each such FlexKnot parameterization we model the expected dispersion measure relation, $DM^{\rm IGM}$, and calculate its likelihood given the mock data (e.g. Figure 2). Finally, we employ Bayesian inference to derive constraints on the parameter space (including input cosmological parameters as well as the reionization history specified in terms of the positions and number of knots). This method allows us to extract the reionization history as well as cosmological parameters from the mock data. Importantly, the FexKnot parametrization allows

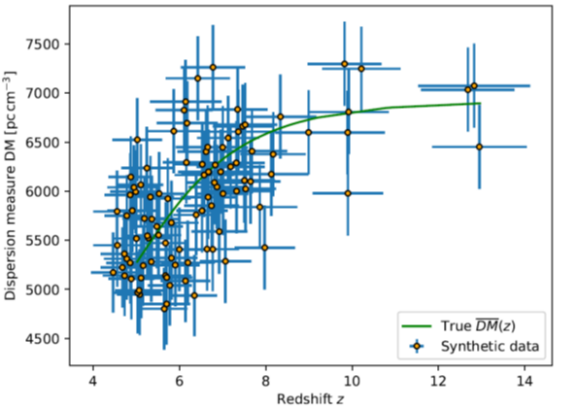


Fig. 2. Illustration of our mock FRB catalogue. Illustration credit: S. Heimersheim.

us to go over the entire space of plausible $x_i(z)$s without making assumptions about their functional shapes. Thus our results are bias-free.

In Heimersheim et al. (2021),[35] we apply the described method to our synthetic FRB catalogue and find that strong constraints on reionization history can be obtained. Figure 3 (Fig. 1 of Heimersheim et al. (2021)) shows the constraints on the reionization history marginalized over individual knot locations and cosmological parameters (only the parameters that have effect on DM^{IGM} were varied). In addition, we include the CMB polarization constraint on the high-redshift portion of the total optical depth τ from $z = 15 - 30$ (based on Planck measurements[37]).

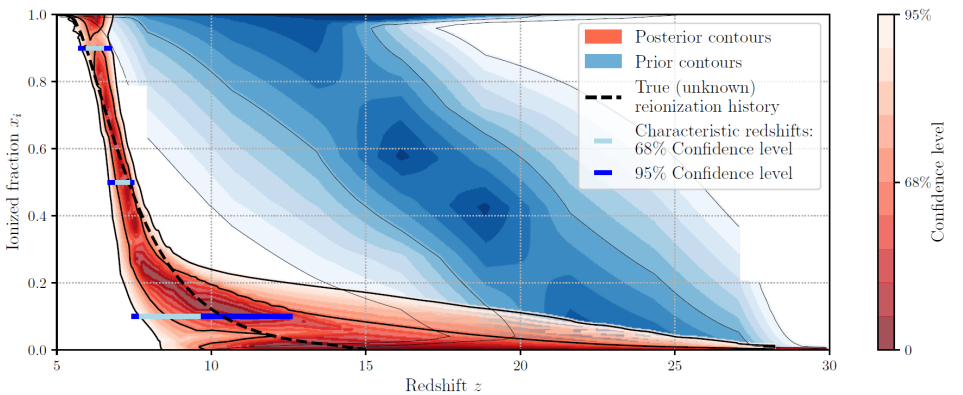


Fig. 3. Illustration of the prior (purple) and extracted (orange) reionization histories. Blue error bars show the constraints on the milestones of the EoR history, namely the start ($x_i = 0.1$), midpoint ($x_i = 0.5$) and end ($x_i = 0.9$) of reionization. Illustration credit: S. Heimersheim. Based on Figure 1 from Heimersheim et al., (2021).

In Figure 3 we demonstrate how the method works by showing the posterior and prior distributions of $x_i(z)$ (indicated by the orange and purple contours respectively, with 68% and 95% confidence intervals marked by the solid lines). We see that the posterior provides tight constraint on the reionization history, and encompasses the true reionization history used to generate the mock data (dashed cyan). Note that the analysis pipeline was not informed about this input history.

We find that, owing to the model assumption that FRB number counts trace star formation rate, the FRB data can constrain the reionization history especially well at low redshift ($z < 10$) where the vast majority (97%) of the FRB sources in our synthetic catalogue are located. Because there is only a small number of FRBs at higher redshifts, the EoR constraints at $z = 10 - 30$ are much weaker.

Owing to the flexibility of the FlexKnot method and because we have the full information on the EoR history for every parameter sample, we can directly constrain key points of the EoR history. To demonstrate this capability, we calculate posterior probabilities of the start ($x_i = 0.1$), midpoint ($x_i = 0.5$) and end ($x_i = 0.9$) of reionization. The error bars are shown in Figure 3 indicating 68% and 95% confidence limits in start, midpoint and end of the EoR, and the numerical results can be found in Table 1 of Heimersheim et al. (2021). We find that with 100 FRBs we constrain the end, midpoint and start of reionization to an accuracy of 7%, 6% and 13% respectively at 68% confidence level, achieving a better accuracy of 5%, 3% and 12% respectively for the sample with 1000 FRBs. The FRB constraints on the midpoint of the EoR would be much stronger than the limits from quasars.[28]

Marginalizing over all other parameters, including all the possible parametrizations of reionization histories, in Heimersheim et al. (2021)[35] we also derive posterior probability distributions for τ from the mock samples of 100 and 1000 FRBs (see Figure 2 in Heimersheim et al. (2021)) and quote the detailed 68% and 95% limits in Table 1 of Heimersheim et al. (2021). In summary, we find that 100 and 1000 FRBs can constrain the optical depth to 11% and 9% accuracy respectively (68% confidence). Our results for 100 FRBs are competitive with the *Planck*[2] measurement, while the accuracy for 1000 FRBs surpasses that of *Planck*. Moreover, the FRB constraint on τ is antisymmetric, owing to the sample of FRBs skewed towards the lower redshifts. Therefore, FRBs provide a much tighter lower limit on τ compared to that from the CMB. We also show the constraints on the amplitude of the primordial power spectrum.

Further details of the method and scientific results can be found in Heimersheim et al. (2021).[35]

Finally, we note that, in addition to serving as cosmological probes, high-redshift FRBs could result to be very useful probes of high-redshift stellar populations (Population III and Population II). The ambiguity regarding the sources of FRBs and the nature of their hosts remains with magnetars being one of the most popular candidates.[38,39] It is also possible that either magnetars or FRBs themselves, are produced as a result of stellar mergers (see discussion in Heimersheim et al. 2021).

In this case, the anticipated number of FRBs would strongly depend on the unknown initial mass function of the early populations of stars. Therefore, mapping the population of FRBs from the first billion years of cosmic history could shed light on the properties of high-redshift stars.

3. Summary

With this work we would like to encourage observational search for FRBs from the first billion years of cosmic history. Observational capability is nearly there with existing telescopes such as FAST and GBT sensitive enough to detect bright FRBs out to redshift of ~ 10, ASKAP excelling in localizing the FRBs and CHIME detecting large numbers of FRBs (although still at relatively low redshifts). If a sample of $100 - 1000$ FRBs from the first billion years of cosmic history can be measured and localized, these events will advance our understanding of the evolution of the Universe in this poorly constrained regime. As we showed in Heimersheim et al. (2021), high redshift FRBs could offer precision constraints on the EoR history and optical depth as well as tightening constraints on the primordial power spectrum and the sum of neutrino masses. In addition, they could serve as probes of high-redshift astrophysics, Population III and Population II stars and binaries. FRBs would probe the high-redshift Universe in new way independent from all the other known probes (high-redshift quasars, the CMB, and the 21-cm signal of hydrogen) thus providing an unparalleled view of this era.

Acknowledgments

This work was supported by the Royal Society University Research Fellowship of Fialkov.

References

1. Lewis, A. Cosmological parameters from WMAP 5- year temperature maps. Phys. Rev. D 78, 023002. arXiv: 0804.3865 [astro-ph] (July 2008).
2. Planck Collaboration et al. Planck 2018 results. VI. Cosmological parameters. arXiv e-prints, arXiv:1807.06209. arXiv: 1807.06209 [astro-ph.CO] (July 2018).
3. Barkana R., 2018, The Encyclopedia of Cosmology. Volume 1: Galaxy Formation and Evolution Rennan Barkana Tel Aviv University, doi:10.1142/9496-vol1.
4. Mesinger A., 2019, The Cosmic 21-cm Revolution; Charting the first billion years of our universe, doi:10.1088/2514- 3433/ab4a73
5. Sevilla-Noarbe I., et al. Dark Energy Survey Year 3 Results: Photometric Data Set for Cosmology. ApJS, 254, 24 (2021).
6. Atek, H. The extreme faint end of the UV luminosity function at $z \sim 6$ through gravitational telescopes: A comprehensive assessment of strong lensing uncertainties. MNRAS, 479, 5184 (2018).
7. Freedman, W. L. Measurements of the Hubble Constant: Tensions in Perspective. ApJ, 919, 16 (2021).

8. Agarwal, D. et al. Initial results from a real-time FRB search with the GBT. MNRAS 497, 352–360. arXiv: 2003.14272 [astro-ph.IM] (Sept. 2020).
9. Deng W. & Zhang B. Cosmological Implications of Fast Radio Burst/Gamma-Ray Burst Associations. ApJ 783, L35. arXiv: 1401.0059 [astro-ph.HE] (Mar. 2014).
10. Fialkov, A. & Loeb, A. Constraining the CMB optical depth through the dispersion measure of cosmological radio transients. JCAP 2016, 004. arXiv: 1602.08130 [astro-ph.CO] (May 2016).
11. Macquart, J. P. et al. A census of baryons in the Universe from localized fast radio bursts. Nature 581, 391–395. arXiv: 2005.13161 [astro-ph.CO] (May 2020).
12. McQuinn, M. Locating the "Missing" Baryons with Extragalactic Dispersion Measure Estimates. ApJ 780, L33. arXiv: 1309.4451 [astro-ph.CO] (Jan. 2014).
13. Yang Y.-P. & Zhang B. Extracting Host Galaxy Dispersion Measure and Constraining Cosmological Parameters using Fast Radio Burst Data. ApJ 830, L31. arXiv: 1608.08154 [astro-ph.HE] (Oct. 2016).
14. Zhang B. Fast Radio Burst Energetics and Detectability from High Redshifts. ApJ 867, L21. arXiv: 1808.05277 [astro-ph.HE] (Nov. 2018).
15. Zhou, B., Li, X.,Wang, T., Fan, Y.-Z. & Wei, D.-M. Fast radio bursts as a cosmic probe? Phys. Rev. D 89, 107303. arXiv: 1401.2927 [astro-ph.CO] (May 2014).
16. Gao, H., Li, Z. & Zhang, B. Fast Radio Burst/Gamma-Ray Burst Cosmography. ApJ 788, 189. arXiv: 1402.2498 [astro-ph.CO] (June 2014).
17. Walters, A., Weltman, A., Gaensler, B. M., Ma, Y.-Z. & Witzemann, A. Future Cosmological Constraints From Fast Radio Bursts. ApJ 856, 65. arXiv: 1711.11277 [astro-ph.CO] (Mar. 2018).
18. Jaroszynski, M. Fast radio bursts and cosmological tests. MNRAS 484, 1637–1644. arXiv: 1812.11936 [astro-ph.CO] (Apr. 2019).
19. Madhavacheril, M. S., Battaglia, N., Smith, K. M. & Sievers, J. L. Cosmology with kSZ: breaking the optical depth degeneracy with Fast Radio Bursts. arXiv e-prints, arXiv:1901.02418. arXiv: 1901.02418 [astro-ph.CO] (Jan. 2019).
20. Wu, Q., Yu, H. & Wang, F. Y. A New Method to Measure Hubble Parameter H(z) Using Fast Radio Bursts. ApJ 895, 33. arXiv: 2004.12649 [astro-ph.CO] (May 2020).
21. Hagstotz, S., Reischke, R. & Lilow, R. A new measurement of the Hubble constant using Fast Radio Bursts. arXiv e-prints, arXiv:2104.04538. arXiv: 2104.04538 [astro-ph.CO] (Apr. 2021).
22. Zheng, Z., Ofek, E. O., Kulkarni, S. R., Neill, J. D. & Juric, M. Probing the Intergalactic Medium with Fast Radio Bursts. ApJ 797, 71. arXiv: 1409.3244 [astro-ph.HE] (Dec. 2014).
23. Caleb, M., Flynn, C. & Stappers, B. W. Constraining the era of helium reionization using fast radio bursts. MNRAS 485, 2281–2286. arXiv: 1902.06981 [astro-ph.HE] (May 2019).
24. Linder, E. V. Detecting helium reionization with fast radio bursts. Phys. Rev. D 101, 103019. arXiv: 2001.11517 [astro-ph.CO] (May 2020).
25. Lau, A. W. K., Mitra, A., Shafiee, M. & Smoot, G. Constraining HeII reionization detection uncertainties via fast radio bursts. New Astron. 89, 101627 (Nov. 2021).
26. Bhattacharya, M., Kumar, P. & Linder, E. V. Fast radio burst dispersion measure distribution as a probe of helium reionization. Phys. Rev. D 103, 103526. arXiv:2010.14530 [astro-ph.CO] (May 2021).
27. Liu, A., Pritchard, J. R., Mortonson, M., Parsons, A. & HERA Collaboration. Improving Cosmic Microwave Background Constraints with 21cm Cosmology in American Astronomical Society Meeting Abstracts 225 (Jan. 2015), 318.01.

28. Greig, B. & Mesinger, A. The global history of reionization. MNRAS 465, 4838–4852. arXiv: 1605.05374 [astro-ph.CO] (Mar. 2017).
29. Mondal, R. et al. Tight constraints on the excess radio background at z = 9.1 from LOFAR. MNRAS, 498, 4178 (2020).
30. The HERA Collaboration. HERA Phase I Limits on the Cosmic 21-cm Signal: Constraints on Astrophysics and Cosmology During the Epoch of Reionization. arXiv:2108.07282 (2021).
31. Beniamini P., Kumar P., Ma X. & Quataert E. Exploring the epoch of hydrogen reionization using FRBs. MNRAS 502, 5134–5146. arXiv: 2011.11643 [astro-ph.CO] (Apr. 2021).
32. Zhang, Z. J., Yan, K., Li, C. M., Zhang, G. Q. & Wang, F. Y. Intergalactic Medium Dispersion Measures of Fast Radio Bursts Estimated from IllustrisTNG Simulation and Their Cosmological Applications. ApJ 906, arXiv: 2011.14494 [astro-ph.CO] (Jan. 2021).
33. Hashimoto, T. et al. Revealing the cosmic reionization history with fast radio bursts in the era of Square Kilometre Array. MNRAS502, 2346–2355. arXiv: 2101.08798 [astro-ph.CO] (Apr. 2021).
34. Pagano, M. & Fronenberg, H. Constraining the epoch of reionization with highly dispersed fast radio bursts. MNRAS 505, 2195–2206. arXiv: 2103.03252 [astro-ph.CO] (Aug. 2021).
35. Heimersheim S., et al. What it takes to measure reionization with fast radio bursts. arXiv:2107.14242 (2021).
36. Millea M. & Bouchet F. Cosmic microwave background constraints in light of priors over reionization histories. A&A 617, A96. arXiv: 1804.08476 [astro-ph.CO] (Sept. 2018).
37. Heinrich C. & Hu W. Reionization effective likelihood from Planck 2018 data. Phys. Rev. D 104, 063505. arXiv: 2104.13998 [astro-ph.CO] (Sept. 2021).
38. Petroff, E. et al. FRBCAT: The Fast Radio Burst Catalogue. Publ. Astron. Soc. Aust. 33, e045. arXiv: 1601.03547 [astro-ph.HE] (Sept. 2016).
39. The CHIME/FRB Collaboration et al. A bright millisecond-duration radio burst from a Galactic magnetar. Nature 587, 54–58 (Nov. 2020).

Closing the cosmological loop with the redshift drift

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

C. S. Alves

Department of Physics and Astronomy, University College London,
Gower Street, London WC1E 6BT, United Kingdom, and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

J. Esteves

Université de Montpellier,
163 rue Auguste Broussonnet, 34090 Montpellier, and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

A. Lapel

Institut d'Astrophysique de Paris,
98bis Bd Arago, 75014 Paris, France and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

B. G. Pereira

Faculdade de Ciências, Universidade do Porto,
Rua do Campo Alegre, 4150-007 Porto, Portugal, and
Centro de Astrofísica da Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal

The redshift drift (also known as the Sandage Test) is a model-independent probe of fundamental cosmology, enabling us to watch the universe expand in real time, and thereby to confirm (or not) the recent acceleration of the universe without any model-dependent assumptions. On the other hand, by choosing a fiducial model one can also use it to constrain the model parameters, thereby providing a consistency test for results obtained with other probes. The drift can be measured by the Extremely Large Telescope and also by the full SKA. Recently two alternative measurement methods have been proposed: the cosmic accelerometer, and the differential redshift drift. Here we summarize a comparative analysis of the various methods and their possible outcomes, using both Fisher Matrix and MCMC techniques. We find that no single method is uniformly better than the others. Instead, their comparative performance depends both on experimental parameters (including the experiment time and redshift at which the measurement is made) and also on the scientific goal (e.g., detecting the drift signal with high statistical

significance, constraining the matter density, or constraining the dark energy properties). In other words, the experiment should be optimized for the preferred scientific goal.

Keywords: Observational cosmology; Redshift drift; Extremely Large Telescope.

1. Introduction

The observational evidence for the acceleration of the universe shows that our canonical theories of cosmology and particle physics are at least incomplete, and possibly incorrect. New physics is out there, waiting to be discovered; we must search for, identify and characterize this new physics. But so far, the interpretation of all our extant cosmological observations is model-dependent, in the sense that it requires a number of assumptions on an underlying cosmological model. Ideally, one would like some fully model-independent observables, enabling consistency tests of these underling assumptions. And indeed there is one conceptually simple, but operationally very challenging, such observable: the redshift drift of objects following the cosmological expansion.

The CosmoESPRESSO team uses the universe as a laboratory to address, with precision spectroscopy and other observational, computational and theoretical tools, grand-challenge questions including the universality of physical laws, the characterization of the large-scale behaviour of gravity (and specifically of the source of the recent acceleration of the universe) and the study of fossil relics or early stages of the universe's evolution, together with contributions to ESO and ESA next-generation facilities and to education and public understanding of science. In what follows we highlight recent contributions of the CosmoESPRESSO team to this fundamental quest, pertaining to the cosmological impact of forthcoming redshift drift measurements. Further details can be found in Ref. 1 and Ref. 2.

2. The redshift drift and its sensitivity

The idea that the redshift of objects following the cosmological expansion changes with time—known as the redshift drift— is many decades old.[3,4] Conceptually, a measurement of the redshift drift is fundamentally different from our other astrophysical observables. In our observations at cosmological distances done so far, we are mapping our present-day past light cone. Instead, the redshift drift allows us to compare different past light cones. More crudely, this corresponds to watching the Universe expand in real time. Other than being a different probe of the universe, its conceptual importance is that it is a model-independent probe of the expansion of the universe, making no assumption on geometry, clustering or the behaviour of gravity, and therefore of crucial importance for fundamental cosmology.

The practical difficulty is that cosmologically relevant timescales are orders of magnitude larger than human ones: a measurement of the redshift drift therefore requires exquisite sensitivity, several orders of magnitude better than currently available: the best currently available bound,[5] is about three orders of magnitude larger

than the signal expected for the standard ΛCDM cosmology with reasonable choices of its model parameters, and contains systematic errors comparable to or larger than the statistical ones. Nevertheless, measuring this drift is a key science and design driver for the development of the Extremely Large Telescope (ELT), and specifically for its high-resolution spectrograph currently known as ELT-HIRES.[6]

A detailed feasibility study of high-redshift measurements by the ELT, on a decade timescale, was done in Ref. 7, who also discussed possible targets. It is important to note that although ELT redshift drift measurements, on their own, lead to cosmological parameter constraints that are not tighter than those available by more classical probes (such as supernovae or the CMB) they do probe regions of parameter space that are different from (and sometimes actually orthogonal to) those of other probes, enabling the breaking of degeneracies and therefore leading to more stringent combined constraints;[1] we will provide an explicit demonstration later in this contribution. More recently, it has also been pointed out that measurements at low redshifts can in principle be made by the full Square Kilometre Array (SKA),[8] while measurements at intermediate redshifts can be done by intensity mapping measurements such as CHIME[9] and HIRAX.[10] Measurements of all three of these will require appropriate hardware configurations.

Redshift drift measurements are a key part of what is commonly called real-time cosmology.[11] Usually they rely on the first derivative of the redshift. Recently the role of second derivatives of the redshift, which should also be within the reach of the full SKA, has been studied by Ref.[12] Fisher Matrix techniques[13,14] are well suited for a comparative study of the cosmological impact of redshift drift measurements by these facilities; such a study is described in detail in Ref. 1.

The redshift drift of an astrophysical object following the cosmological expansion, for an observer looking at it over a time span Δt, is

$$\frac{\Delta z}{\Delta t} = H_0 \left[1 + z - E(z)\right] , \tag{1}$$

where for convenience we have defined the dimensionless Hubble parameter $E(z) = H(z)/H_0$, although the actual astrophysical observable is usually a spectroscopically measured velocity

$$\Delta v = \frac{c\Delta z}{1+z} = (cH_0\Delta t)\left[1 - \frac{E(z)}{1+z}\right] . \tag{2}$$

The dependence on the Hubble parameter naturally leads to a model-dependent redshift dependence of the drift. Broadly speaking, in a universe that is currently accelerating but was deccelerating in the past the drift will be positive at low redshifts and negative for higher redshifts, while in a universe that always deccelerates the redshift drift would always be negative. To give a specific example, for a flat ΛCDM model, apart from trivially vanishing at $z = 0$, the signal will also vanish at

$$z_{zero} = \frac{1 - 3\Omega_m + \sqrt{1 + 2\Omega_m - 3\Omega_m^2}}{2\Omega_m} , \tag{3}$$

while the redshift of maximum spectroscopic velocity is

$$z_{v,max} = \left[\frac{2(1-\Omega_m)}{\Omega_m}\right]^{1/3} - 1 \tag{4}$$

and the maximum of the drift itself is obtained by solving the quartic equation

$$9\Omega_m^2(1+z)^4 = 4\Omega_m(1+z)^3 + 4(1-\Omega_m)\,. \tag{5}$$

More specifically, if we choose $\Omega_m = 0.3$ (in agreement with contemporary cosmological data) we obtain $z_{zero} =\sim 2.09$, $z_{v,max} \sim 0.67$ and $z_{z,max} \sim 0.95$. It is interesting to note that for flat ΛCDM with $\Omega_m = 0.3$, the acceleration phase does start at redshift $z_{v,max}$, while the cosmological constant only starts dominating the Friedmann equation at

$$z_\Lambda = \left[\frac{(1-\Omega_m)}{\Omega_m}\right]^{1/3} - 1\,, \tag{6}$$

which for Ω_m leads to $z_\Lambda \sim 0.33$.

We can analytically explore the sensitivity of the redshift drift to the cosmological parameters, for a fiducial flat CPL model.[15,16] In this case we can write the Friedmann equation as

$$E^2(z) = \Omega_m(1+z)^3 + \Omega_\phi(1+z)^{3(1+w_0+w_a)}\exp\left[-\frac{3w_a z}{1+z}\right]. \tag{7}$$

It is also convenient to define a dimensionless redshift drift

$$S_z = \frac{1}{H_{100}}\frac{\Delta z}{\Delta t} = h\left[1+z-E(z)\right]\,, \tag{8}$$

where we have defined $H_0 = hH_{100}$ with $H_{100} = 100$ km/s/Mpc; the corresponding observable spectroscopic velocity will be denoted

$$S_v = \Delta v = kh\left[1 - \frac{E(z)}{1+z}\right]\,. \tag{9}$$

In these we have introduced $k = cH_{100}\Delta t$, which is a constant parameter for a given observation time, with units of cm/s: for $\Delta t = 1$ year, $k = 3.064$ cm/s. This provides a rough estimate of the magnitude of the signal in the relevant time span, and therefore also an estimate of the required sensitivity of the spectroscopic measurements. Then for each cosmological parameter p_i the sensitivities of S_z and S_v are related via

$$\frac{\partial S_v/\partial p_i}{\partial S_z/\partial p_i} = \frac{k}{1+z}\,. \tag{10}$$

For our fiducial model we can then find the following sensitivities

$$\frac{\partial S_z}{\partial h} = 1 + z - E(z) \tag{11}$$

$$\frac{\partial S_z}{\partial \Omega_m} = -\frac{h(1+z)^3}{2E(z)}\left[1-(1+z)^{3(w_0+w_a)}\exp\left[\frac{-3w_a z}{1+z}\right]\right] \tag{12}$$

$$\frac{\partial S_z}{\partial w_0} = -\frac{3h(1-\Omega_m)}{2E(z)}(1+z)^{3(1+w_0+w_a)}\ln(1+z)\exp\left[\frac{-3w_a z}{1+z}\right] \tag{13}$$

$$\frac{\partial S_z}{\partial w_a} = -\frac{3h(1-\Omega_m)}{2E(z)}(1+z)^{3(1+w_0+w_a)}\left[\ln(1+z) - \frac{z}{1+z}\right]\exp\left[\frac{-3w_a z}{1+z}\right]. \tag{14}$$

Note that the sign of the $\partial S_z/\partial h$ term will depend on redshift, while those of the other derivatives are always negative for observationally reasonable values of the model parameters.

Some sensitivity ratios are also illuminating. Consider that of the two dark energy equation of state parameters

$$\frac{\partial S_z/\partial w_a}{\partial S_z/\partial w_0} = 1 - \frac{z}{(1+z)\ln(1+z)}\,; \tag{15}$$

as one would expect this approaches zero as $z \to 0$ and unity as $z \to \infty$. On the other hand, comparing the sensitivities to the matter density and the present-day dark energy equation of state one finds

$$\frac{\partial S_z/\partial \Omega_m}{\partial S_z/\partial w_0} = \frac{1-(1+z)^{3(w_0+w_a)}\exp\left[\frac{-3w_a z}{1+z}\right]}{3(1-\Omega_m)\ln(1+z)(1+z)^{3(w_0+w_a)}\exp\left[\frac{-3w_a z}{1+z}\right]}\,, \tag{16}$$

which tends to $(-w_0)/(1-\Omega_m)$ as $z \to 0$ and to infinity as $z \to \infty$ (though note that in this we are neglecting the radiation density). Considering the specific case of a fiducial ΛCDM model with $\Omega_m = 0.3$, at $z = 1$, the ratios are respectively

$$\frac{\partial S_z/\partial w_a}{\partial S_z/\partial w_0} = 1 - \frac{1}{2\ln 2} \sim 0.28 \tag{17}$$

$$\frac{\partial S_z/\partial \Omega_m}{\partial S_z/\partial w_0} \sim \frac{7}{2.1\ln 2} \sim 4.8\,. \tag{18}$$

Finally, it is also illuminating to consider the low-redshift limits of the various derivative terms. To obtain them, one uses the fact that as $z \to 0$ we have

$$E(z) = 1 + \frac{3}{2}\left[1 + (1-\Omega_m)w_0\right]z + \mathcal{O}(z^2)\,. \tag{19}$$

For the derivatives with respect to h, Ω_m and w_0 we then find, respectively

$$\frac{\partial S_z}{\partial h} \longrightarrow -\frac{1}{2}\left[1 + 3(1-\Omega_m)w_0\right]z \tag{20}$$

$$\frac{\partial S_z}{\partial \Omega_m} \longrightarrow \frac{3}{2}hw_0 z \tag{21}$$

$$\frac{\partial S_z}{\partial w_0} \longrightarrow -\frac{3}{2}h(1-\Omega_m)z \tag{22}$$

while that with respect to w_a (whose value, interestingly, does not affect any of the other three) is of higher order.

Figure 1 depicts the redshift dependence of the two observational sensitivities, $\partial S_v/\partial p_i$ and $\partial S_z/\partial p_i$, for a flat ΛCDM fiducial model with $\Omega_m = 0.3$ and $h = 0.7$.

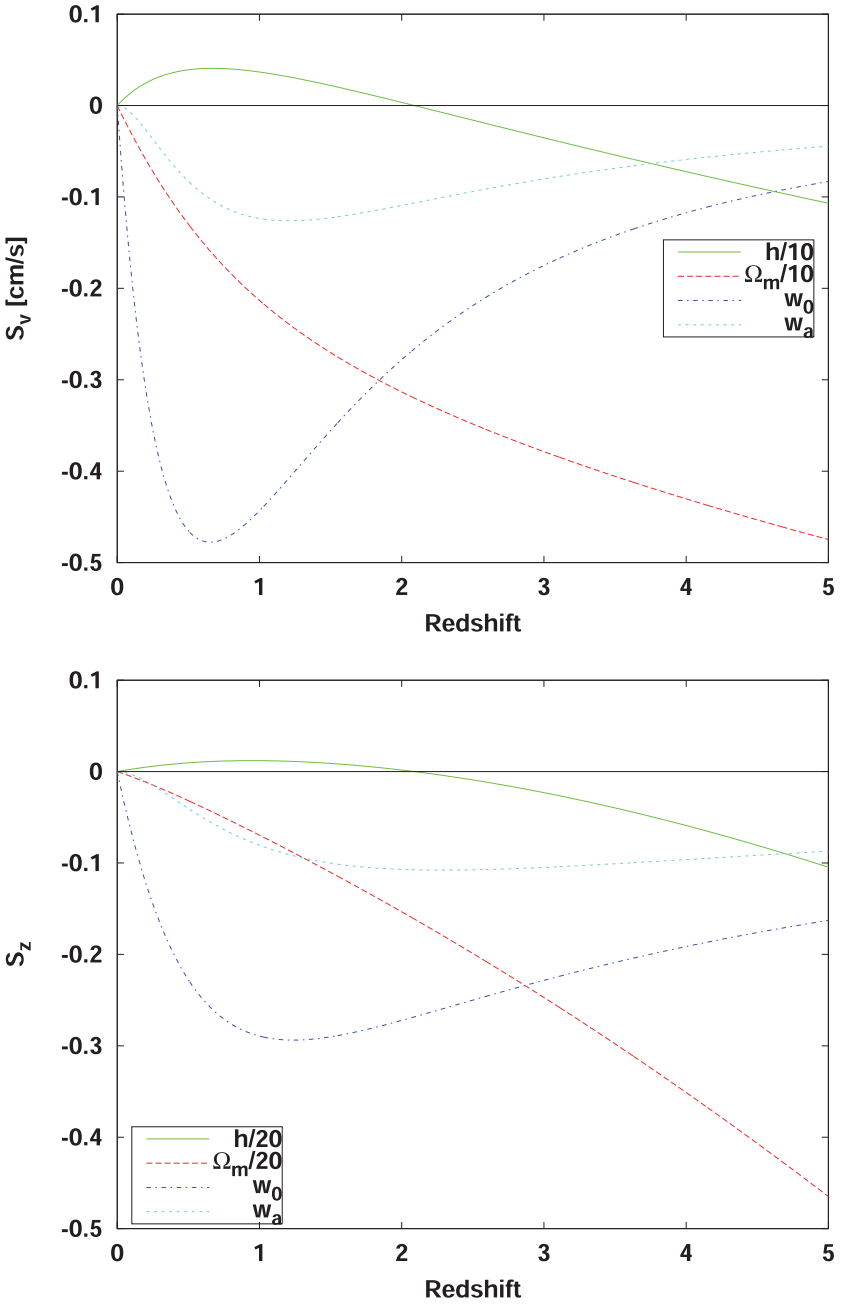


Fig. 1. Sensitivities of the spectroscopic velocity (top panel) and of the redshift drift signal (bottom panel) to the cosmological parameters in the CPL parametrization, for a flat ΛCDM fiducial. For plotting convenience the sensitivities to h and Ω_m have been divided by a factor of 10 in the top panel and by a factor of 20 in the bottom panel. The zero sensitivity line is shown in black. Similar plots can be found in Figure 4 of Ref. 1.

Note that the two are slightly different, being related by a redshift-dependent factor. Also note the different units in the vertical axes of the two panels: $\partial S_z/\partial p_i$ is dimensionless, while $\partial S_v/\partial p_i$ has units of cm/s.

As expected the sensitivity on the dimensionless Hubble constant h (which is proportional to the redshift drift signal itself) is the only one that changes sign; this and the sensitivity on Ω_m are also larger than those on the dark energy equation of state parameters w_0 and w_a. Nevertheless, the most noteworthy point is that the various sensitivity curves have different redshift dependencies. While the matter density sensitivity increases monotonically (in absolute value) with redshift, those of w_0 and w_a are maximal at around the onset of acceleration (just below $z = 1$, with the value being slightly different for S_v and S_z), and the former has a stronger redshift dependence than the latter. This is important because it implies that as long as one is able to do these measurements at sufficiently broad redshift ranges there should not be strong covariances between the parameters.[1]

3. Three experimental strategies

As previously mentioned, redshift drift measurements are unique in comparing different past light cones in a fully model-independent way, rather than mapping our present-day past light cone. Nevertheless, these measurements can complement traditional probes in constraining specific cosmological models. Indeed they have an important role to play here, since they do probe different regions of parameter space, enabling the breaking of key degeneracies upon combination. One may therefore ask how an experimental strategy aimed at maximizing the cosmological constraining power of the redhift drift (on its own or in combination with traditional probes) compares with a strategy aimed at maximizing the statistical significance of the detection of the drift signal itself.

This is a particularly pressing question since two other approaches to detecting the redshift drift have been recently proposed. The Cosmic Accelerometer[17] aims to lower the cost of the experiment, while the Acceleration Programme[18] proposes to measure the differential redshift drift between two non-zero redshifts, rather than measuring the drift with respect to today. In Ref. 2 we again used Fisher Matrix techniques in a comparative study of the cosmological impact of the three approaches. The fiducial model choices were also those described above, to maintain consistency with Ref. 1. Herewith we introduce these alternative strategies and summarize the results of this analysis.

The case study for redshift measurements by the ELT has been done in Ref. 7. The measurement relies on absorption features in the Lyman-α forest, complemented by metal absorption lines redwards of it. The main practical result of this study is that the spectroscopic velocity uncertainty is well described by

$$\sigma_v = \sigma_e \left(\frac{2370}{S/N}\right) \sqrt{\frac{30}{N_{QSO}}} \left(\frac{1+z_{QSO}}{5}\right)^{-\lambda} \; cm/s\,, \tag{23}$$

where the last exponent is $\lambda = 1.7$ up to $z = 4$ and $\lambda = 0.9$ for higher redshifts. Consistently with this and also with the latest top-level requirements for the ELT-HIRES spectrograph,[6] we assume that each measurement has a signal to noise ratio $S/N = 3000$ and uses data from $N_{QSO} = 6$ quasars; as baseline values we take $\sigma_e = 1.35$ and an experiment time $\Delta t = 20$ years, but we also explore different choices for these two parameters. The ELT can measure the redshift drift in the approximate redshift range $2.0 \leq z \leq 4.5$. The main bottleneck, in addition to the spectrograph stability which is thought not to be a bottleneck,[19] is the availability of bright quasars providing the required signal to noise in reasonable amounts of telescope time. The discovery of additional bright quasars will improve the experiment feasibility,[20] so our previously reported ELT analysis[2] might be conservative.

The first alternative, dubbed the Cosmic Accelerometer[17] (henceforth CAC) can be crudely thought of as a low-cost version of the ELT experiment, relying on commercial off the shelf equipment. The proposal is part of the Astro2020 decadal survey, and to our knowledge no detailed feasibility study has been carried out so far. We therefore take at face value the spectroscopic velocity uncertainty given in the proposal white paper

$$\sigma_v = \sigma_c \sqrt{\frac{6}{t_{exp}}}\, cm/s\,, \tag{24}$$

with the baseline value of $\sigma_c = 1.5$ and that of the experiment time $t_{exp} = 6$ years; the authors of the decadal survey proposal state that a further advantage of this approach would be a detection of the redshift drift on a shorter time scale than that of the ELT. In any case, we will again explore the impact of different values of these parameters. The redshift ranges probed are, in principle, similar to those for the ELT.

The second alternative is the Acceleration Programme[18] (henceforth APR), which uses the ELT as described above but proposes to measure the drift between sources at two different non-zero redshifts along the same line of sight. This is to be compared to the standard approach, where one of these redshifts is always $z = 0$. The measured quantity in this case will be a spectroscopic velocity difference, which making use of previously introduced definitions we can write as

$$\Delta v_{ir} = (cH_0\Delta t)\left[\frac{E(z_r)}{1+z_r} - \frac{E(z_i)}{1+z_i}\right], \tag{25}$$

where z_r and z_i are the redshifts of the reference and intervening sources, implying that $z_i < z_r$. The rationale for this proposal is that for some choices of redshifts the absolute value of the detected drift signal (or, more precisely, that of the corresponding spectroscopic velocity) can be significantly larger than in the standard approach. Since the absolute value of the spectroscopic velocity uncertainty of the measurement, σ_v, is determined by the spectrograph's technical specifications, maximizing

the value of the signal to be detected leads to an increased statistical significance of the detection.

4. Cosmological constraints based experimental trade-offs

Our main comparison diagnostic is the derived uncertainty in the cosmological parameters being considered, specifically the matter density fraction and the CPL dark energy equation of state parameters. Our fiducial model for the comparison is flat ΛCDM, with $\Omega_m = 0.3$, $H_0 = 70$ km/s/Mpc, $w_0 = -1$ and $w_a = 0$.

To compare the ELT and CAC it is convenient to assume a single measurement of the redshift drift. Clearly, even for a flat ΛCDM fiducial model a single redshift drift measurement can't simultaneously constrain the Hubble constant and the matter density, so one adopts ample uniform priors for both of these. We can then explore the impact of choices of normalization of the spectroscopic velocity uncertainties (σ_e and σ_c), of the experiment time, and of the redshift at which the measurement is made. Our specific metric in this case be the posterior matter density constraint, σ_m.

The crucial point for this analysis, already discussed and illustrated in Fig. 1 is that the cosmological parameter sensitivities of the redshift drift are redshift-dependent.[1] This sensitivity must therefore be convolved with the spectroscopic velocity uncertainty of each experiment, bearing in mind that the ELT one is redshift dependent (cf. Eq. 23) while the CAC one is not (cf. Eq. 24), at least if one takes the CAC proposal at face value.

In a nutshell, the main result of this comparison is that for identical experiment times the CAC does better, though the difference decreases for larger redshifts. A possible advantage of the CAC (in addition to cost) would be a lower experiment time. The top panel of Figure 2 compares the constraints on σ_m for a 6 year CAC and a 20 year ELT experiment, and one finds that in these circumstances the ELT does clearly better at the redhsifts of interest—up to 62% better. The reason for this is clear. While absolute uncertainties σ_v may be smaller for the CAC, the relative ones will be larger due to smaller experiment times, since a key feature of the redshift drift is that the signal grows linearly in time. The bottom panel of the same figure compares both experiments at 20 years; in this case the CAC is more constraining—by up to a factor of six for measurements around redshift $z \sim 2$.

We therefore see that an experiment optimized to detect the redshift drift signal is not necessarily optimized to constrain particular cosmological models, or indeed specific model parameters. In the example of Figure 2 the constraints on the matter density can be substantially different, but those on the Hubble constant never differ by more than a few percent.[2]

Analogous results ensue when comparing the parameter uncertainties in the canonical redshift drift and the differential redshift drift, for the CPL parameterization. To remain consistent with the work on the canonical redshift drift,[1] in Ref. 2 we used a set of five redshift drift measurements at reference redshifts

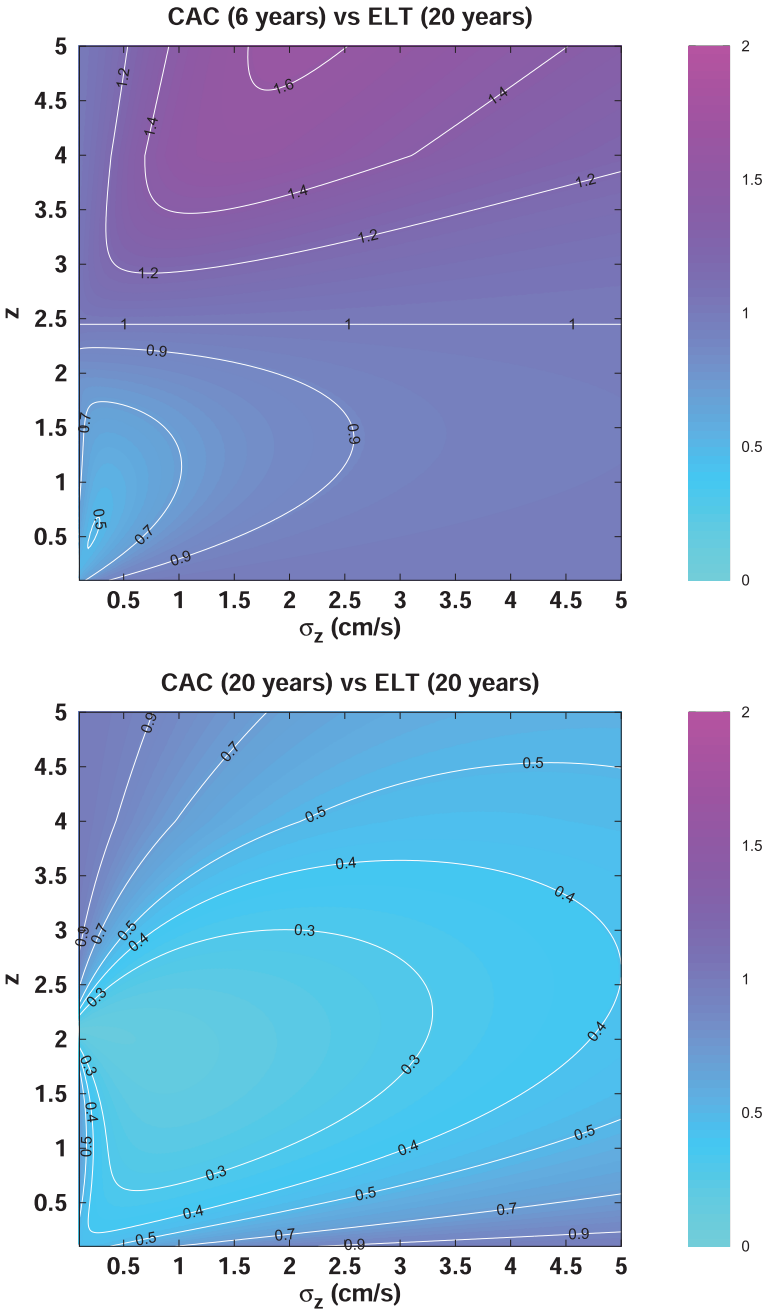


Fig. 2. Ratio of the constraints on the matter density (σ_m) from a single redshift drift measurement by the CAC and ELT, as a function of the redshift of the measurement (z) and the velocity uncertainty (σ_z), with a CAC experiment time of 6 or 20 years (top and bottom panels), and 20 years for ELT in both cases. A ratio larger than unity implies that the ELT measurement is more constraining. An alternative depiction of this comparison can be found in Figure 2 of Ref. 2.

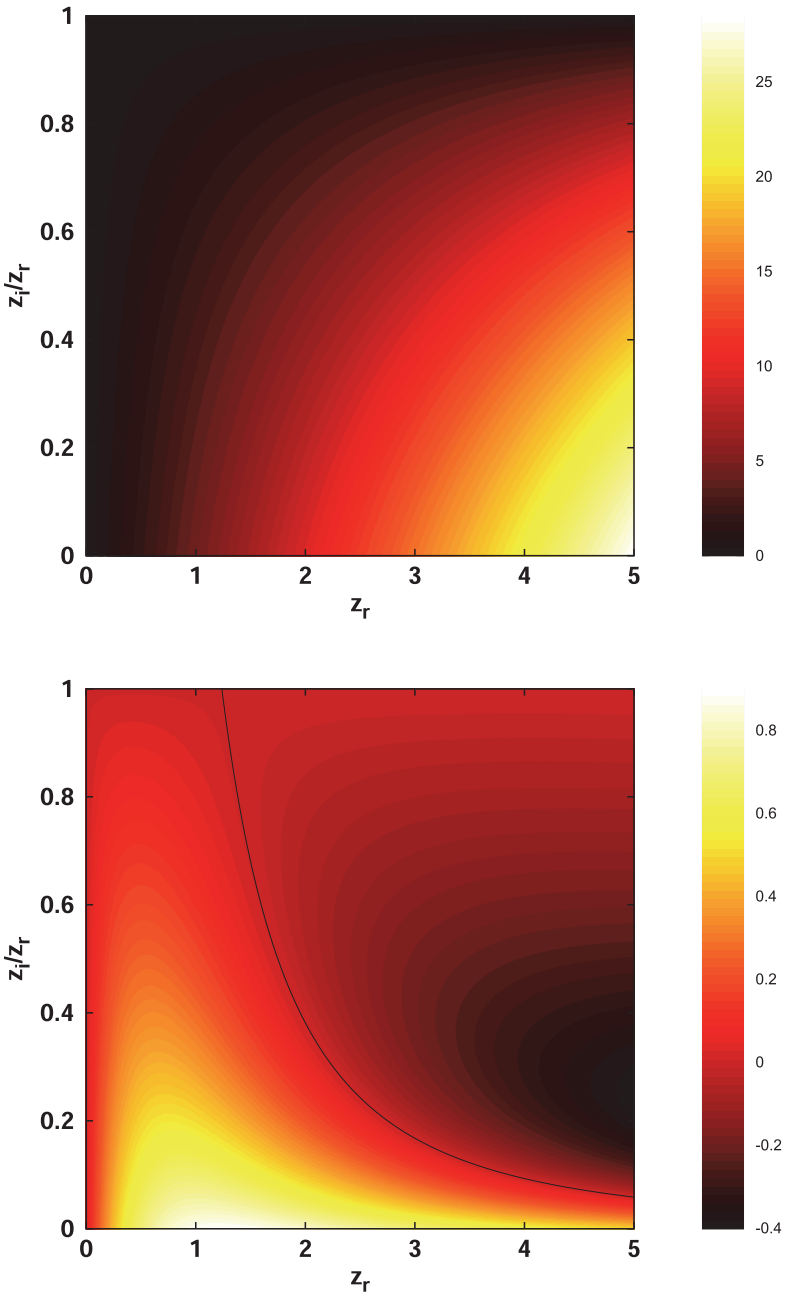


Fig. 3. The sensitivities of the differential redshift drift to the cosmological parameters Ω_m (top) and w_0 (bottom), as a function of the reference redshift z_r and the ratio of intervening and reference redshifts z_i/z_r. The colormaps show the sensitivity in cm/s per year of observation, and the black curves show the locus of zero sensitivity (for Ω_m there is no such non-trivial locus). An alternative depiction of these sensitivities, also for h and w_a, can be found in Figure 3 of Ref. 2.

$z_r = \{2.0, 2.5, 3.0, 3.5, 4.5\}$, and a 20 years experiment. As for the choice of intervening redshifts one may consider two representative choices one with all $z_1 = 0.67$ (this being the redshift with maximum positive spectroscopic velocity for our choice of fiducial model) and another numerically determined with the goal of maximizing the Figure of Merit for the dark energy equation of state parameters, w_0 and w_a, defined as the inverse of the area of the one-sigma confidence contour in the two-dimensional parameter space.

The results, further detailed in Ref. 2, confirm that amplifying the signal for the redshift drift does not necessarily improve the uncertainties of any parameters. However, the numerically determined set of redshifts shows that in ideal circumstances the uncertainties for w_0 and w_a can be improved by one to two orders of magnitude with respect to the canonical redshift drift, while the uncertainties for h and Ω_m were also improved. Nevertheless this advantage is diluted when priors are added, due to relatively high non-diagonal terms in the Fisher matrices. (In particular, we see that constraints on the dark energy equation of state parameters are dominated by the priors, making the ratios of the corresponding uncertainties become unity.) It will be important to validate these results though a MCMC analysis.

These results can be understood by calculating the sensitivity coefficients of the differential redshift drift to our four cosmological parameters, $\partial(\Delta v_{ir})/\partial p$, where p is one of (h, Ω_m, w_0, w_a), These are analogous to the ones shown in Fig. 1 for the standard redshift drift, except that they now depend on both the reference and intervening redshifts. As expected, one finds that the sensitivity to Ω_m is much larger than the others. Crucially, the sensitivity to Ω_m is maximized for $z_i = 0$ and maximal z_r: in other words, if the goal is to constrain Ω_m, the canonical redshift drift measurement is always the optimal strategy. On the other hand, for the other model parameters the differential redshift drift can provide tighter constraints.

5. Synergy with classical cosmological probes

Throughout this analysis, it has become clear that optimal constraints on the background quantities dynamically involved in the redshift-drift require wide redshift intervals to take advantage of the individual parameter sensitivities over the expansion history. This justifies, inter alia, the complementarity of redshift drift experiments, whose main candidates are embodied by the phase-2 SKA for low-redshift spiral galaxies and the ELT for distant quasars. Indeed, as previously pointed out, not only do these experiments complement each other, but there are also hints that they present promising synergies with traditional cosmological probes, potentially allowing to achieve more stringent joint constraints over cosmological parameters of the concordance model and its extensions.[1] Specifically, the fundamentally different nature of the redshift-drift experiment with respect to CMB, SNe, galaxy clustering, or weak lensing experiments, allows for novel correlations within the relevant parameter space, witnessed by degeneracy breaking. As a consistency check and to gain further insights on the degeneracies of the dark energy parameters, it is of

great interest to complement current Fisher forecasts with a more thorough MCMC analysis for accurate sampling of the distributions.

Similarly to the previous section, we account for a possible dynamical behavior of dark energy with two degrees of freedom, w_0 and w_a, through the CPL parametrization. The mock dataset for the ELT is generated according to the prescriptions of the third section for 30 QSOs populating 5 redshift bins between $z = 2$ and $z = 4.5$. As for the total uncertainty over the simulated data, Eq. (23) is used with a somewhat more optimistic value of unity for σ_e, reflecting the fact that redshifts are measured only at two distinct times (the underlying motivation is to compare the full potential of the redshift-drift against tight CMB and SNe constraints). Moreover, random noise is generated from a Gaussian distribution around the theoretical redshift-drift value, with a standard deviation of σ_v to mimic genuine measurements from Lyman-α forests. As for the phase-2 SKA galaxy sample, 5 bins are generated between redshifts 0.2 and 1.0 with redshift-dependent relative uncertainties of 10% following recommendations of previous estimations,[12] to which is added the same artificial Gaussian noise over the theoretical signal.

We implemented the theoretical redshift-drift signal to the dynamical landscape at the background level within the Boltzmann solver CLASS.[21] The theoretical drift is constrained to follow Planck 2018's cosmology[22] not to bias the relative distributions and to ease the comparison. The true posterior is then sampled with the Bayesian sampler MontePython[23] in which we integrated the simulated datasets and the scripts to generate the associated Gaussian likelihoods. The priors over all relevant cosmological parameters are consistent with the ones used in the Planck analysis.

The most noteworthy results from the forecast illustrated in Fig. 4 are first that there is no significant deviation from Gaussianity within the distribution, which further legitimates previous Fisher analyses. More importantly, one can note the remarkable degeneracy breaking of the redshift drift probability density function on most of the parameters with respect to the other probes. In particular, the clear degeneracy within the w_0–w_a space tends to suggest that the joint analysis of classical probes with future redshift drift datasets could further improve the constraints on the first-order term of the CPL parametrization and therefore work toward a better characterization of the possible dynamical behavior of dark energy. Here, the contribution of the low redshift SKA2 sample is important to this endeavor as it probes the dark energy-dominated era and allows to reveal a positive correlation between the two parameters whereas w_0 and w_a are anticorrelated for the other current cosmological observables. Nevertheless, one should keep in mind that such a combination of datasets requires particular attention to the possible correlations between probes and individual systematics.

That being said, in the light of these results, we should expect a consistent gain in the constraints on Ω_m, w_0, and w_a with this type of configuration for SKA and ELT from the joint analysis of the redshift drift measurements with current

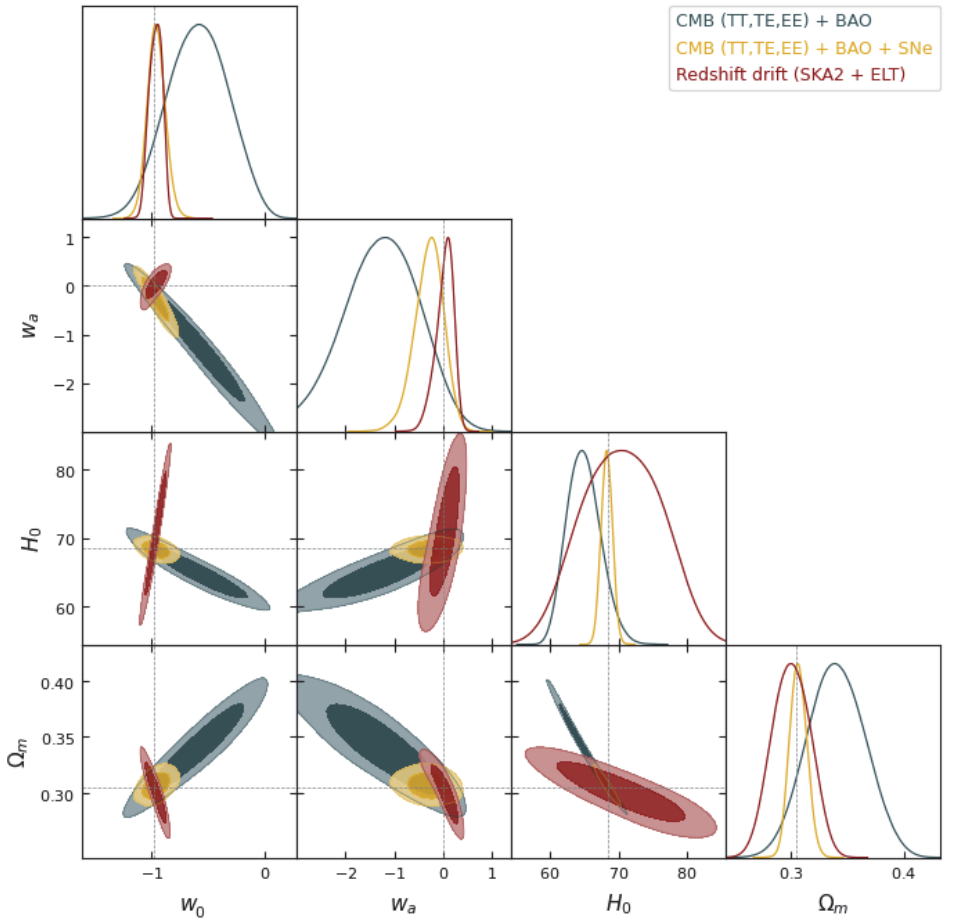


Fig. 4. Marginalized posterior distributions in the space $\{\Omega_m, H_0, w_0, w_a\}$ for the joint contribution of redshift drift experiments (red), CMB experiment with BAO measurement (teal) and CMB + BAO + supernovae (yellow). The data considered for the redshift drift experiments are the mock SKA2 and ELT samples whereas classical probes (CMB, BAO, SNe) are respectively taken from Planck, BOSS DR12, and the Pantheon supernovae samples.

cosmological probes. In the near future, a better instrumental characterization of these facilities is likely to refine the present conservative forecasts, and the CAC program, as well as intensity mapping experiments, will provide more data samples and homogenize the covered redshift range. Furthermore, by the time of the first observations of the redshift drift, next-generation galaxy surveys (Euclid, DESI), supernovae surveys (Vera C. Rubin Observatory) but also CMB experiments (Simons observatory, CMB-S4) will improve individual constraints, although in absence of new observable to present different correlations, these experiments will still highly benefit from the redshift drift degeneracy breaking.

6. Conclusions and outlook

The era of redshift drift measurements (rather than upper limits) is coming. These experiments need to be optimized for a specific purpose, which might be the detection of the drift signal itself, or constraining specific model parameters, such as the matter density or the energy equation of state, or providing a definitive consistency test for the model-dependent analysis of traditional cosmological datasets. Our analysis shows that optimized observational strategies for each of these are significantly different.

We note that an assumption of our analysis is that the measurement redshifts are free parameters, thus ignoring instrumental and observational limitations. In practice one can only do the measurements with the known bright quasars and their absorption systems. The number of astrophysical targets for these measurements remains relatively small despite recent progress.[7, 20] The Phase B of construction of the ELT-HIRES spectrograph, which will likely do the first redshift drift measurements, is about to start, and it will include a detailed feasibility study that will further explore the issues raised herein.

Acknowledgments

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017. This work was partially enabled by funding from the UCL Cosmoparticle Initiative.

References

1. C. S. Alves, A. C. O. Leite, C. J. A. P. Martins, J. G. B. Matos and T. A. Silva, Forecasts of redshift drift constraints on cosmological parameters, *Mon. Not. Roy. Astron. Soc.* **488**, 3607 (2019).
2. J. Esteves, C. J. A. P. Martins, B. G. Pereira and C. S. Alves, Cosmological impact of redshift drift measurements, *Monthly Notices of the Royal Astronomical Society: Letters* **508**, L53 (08 2021).
3. A. Sandage, The Change of Redshift and Apparent Luminosity of Galaxies due to the Deceleration of Selected Expanding Universes., *Ap. J.* **136**, p. 319 (1962).
4. G. C. McVittie, Appendix to The Change of Redshift and Apparent Luminosity of Galaxies due to the Deceleration of Selected Expanding Universes., *Ap. J.* **136**, p. 334 (1962).
5. J. Darling, Toward a Direct Measurement of the Cosmic Acceleration, *Astrophys. J.* **761**, p. L26 (2012).
6. J. Liske *et al.*, *Top Level Requirements For ELT-HIRES*, tech. rep., Document ESO 204697 Version 1 (2014).
7. J. Liske *et al.*, Cosmic dynamics in the era of Extremely Large Telescopes, *Mon. Not. Roy. Astron. Soc.* **386**, 1192 (2008).

8. H.-R. Klockner, D. Obreschkow, C. Martins, A. Raccanelli, D. Champion, A. L. Roy, A. Lobanov, J. Wagner and R. Keller, Real time cosmology - A direct measure of the expansion rate of the Universe with the SKA, *PoS* **AASKA14**, p. 027 (2015).
9. H.-R. Yu, T.-J. Zhang and U.-L. Pen, Method for Direct Measurement of Cosmic Acceleration by 21-cm Absorption Systems, *Phys. Rev. Lett.* **113**, p. 041303 (2014).
10. L. B. Newburgh *et al.*, HIRAX: A Probe of Dark Energy and Radio Transients, *Proc. SPIE Int. Soc. Opt. Eng.* **9906**, p. 99065X (2016).
11. C. Quercellini, L. Amendola, A. Balbi, P. Cabella and M. Quartin, Real-time Cosmology, *Phys. Rept.* **521**, 95 (2012).
12. C. J. A. P. Martins, M. Martinelli, E. Calabrese and M. P. L. P. Ramos, Real-time cosmography with redshift derivatives, *Phys. Rev.* **D94**, p. 043001 (2016).
13. A. Albrecht *et al.*, Report of the Dark Energy Task Force, arXiv: astro-ph/0609591, (2006).
14. A. Albrecht *et al.*, Findings of the Joint Dark Energy Mission Figure of Merit Science Working Group, arXiv: astro-ph/0901.0721, (2009).
15. M. Chevallier and D. Polarski, Accelerating universes with scaling dark matter, *Int. J. Mod. Phys.* **D10**, 213 (2001).
16. E. V. Linder, Exploring the expansion history of the universe, *Phys. Rev. Lett.* **90**, p. 091301 (2003).
17. S. S. Eikenberry *et al.*, Astro2020 Project White Paper: The Cosmic Accelerometer (7, 2019).
18. R. Cooke, The ACCELERATION programme: I. Cosmology with the redshift drift, *Mon. Not. Roy. Astron. Soc.* **492**, 2044 (2020).
19. D. Milaković, L. Pasquini, J. K. Webb and G. Lo Curto, Precision and consistency of astrocombs, *Mon. Not. Roy. Astron. Soc.* **493**, 3997 (2020).
20. K. Boutsia, A. Grazian, G. Calderone, S. Cristiani, G. Cupani, F. Guarneri, F. Fontanot, R. Amorin, V. D'Odorico, E. Giallongo and et al., The spectroscopic follow-up of the qubrics bright quasar survey, *The Astrophysical Journal Supplement Series* **250**, p. 26 (Sep 2020).
21. J. Lesgourgues, The cosmic linear anisotropy solving system (class) i: Overview (2011).
22. N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo and et al., Planck 2018 results, *Astronomy & Astrophysics* **641**, p. A6 (Sep 2020).
23. T. Brinckmann and J. Lesgourgues, Montepython 3: boosted mcmc sampler and other features (2018).

Gamma-Ray Bursts as potential cosmological probes

L. Izzo

DARK, Niels Bohr Institute, University of Copenhagen,
Niels Bohr Building, Jagtvej 128, 2200, Copenhagen, Denmark
E-mail: luca.izzo@nbi.ku.dk

Gamma-ray bursts are among the most luminous transients in the Universe, a characteristic that permits us to observe them at very high redshifts. For this reason, many efforts have been made to identify a method to use GRBs as cosmological distance indicators through the use of luminosity correlations between their high-energy observable quantities. Some of the most promising methods proposed so far to standardize GRBs and their possible contribution to cosmology are presented and discussed, with a special emphasis on the Combo relation.

Keywords: Cosmology: distance scale; Gamma-ray bursts: general; Cosmology: cosmological parameters.

1. Introduction

Gamma-ray bursts (GRBs) are quick flashes of high-energy radiation initially observed in gamma-rays, where their temporal emission, also called ther "prompt" radiation is quite diversified.[1,2] The energy distribution of the photons emitted during the prompt emission can be successfully modelled with a phenomenological function, the Band model,[3] which can be described by two power-law functions smoothly connected at a given energy value, namely the peak energy of the prompt emission E_{peak}. GRBs can also be classified in two families, the *short* and *long* events, based on the T_{90} duration of the prompt emission.[4] This classification is also reflected in the hardness of the prompt photon energy spectrum, with short bursts having higher E_{peak} values than long events.[5] This classification is not only phenomenological, but it can be related to the physical progenitors of these powerful explosions: short bursts originate in the merger of two compact neutron stars,[6,7] an evidence confirmed by the recent discovery of a double neutron stars merger detected by gravitational wave detectors associated with a short GRB,[8] while long GRBs emitted during the collapse of highly-rotating massive stars.[9,10] The detection of an accompanying supernova few days after nearby long GRBs has given support in the last years to this model.[11,12]

As the time goes on, the bulk of the radiation emitted by GRBs moves to lower energies, such as in the X-rays,[13–15] at optical[16] and radio frequencies.[17] The emission observed in this so-called "afterglow" component looks at this stage more homogeneous, in particular at X-rays where it can be categorized in few subclasses.[18] The associated afterglow optical emission, detected in a good fraction majority of GRBs, has also permitted to identify the cosmological nature of these

explosions, through the identification of the redshift, and then to quantify with good accuracy their energetics and luminosities.[19–21]

2. GRBs luminosity correlations

GRBs are very energetic and luminous events. This property permits to observe them at very far distances: the most distant GRB was observed at a redshift of $z = 9.4$,[22] few hundred millions after the Big Bang and well inside the reionization epoch, when the first galaxies begun to form.[23] Then, it was almost natural to search a method to render GRBs a possible standardizable distance indicator to scrutiny the far Universe, and to search for possible exotic behavior of the dark energy component. The most used method consists in using some correlations between high-energy GRB observables, where at least one of the parameters involved is dependent on the background cosmology while the remaining ones do not. Gamma-ray photons ($\nu > 5keV$) can travel in the empty space without being affected by inter-stellar and inter-galactic absorbers, such as dust extinction, which however affects the use of optical distance indicators such as type-Ia SNe,[24] and this represents one important advantage for using GRBs as distance tools.

One of the most promising and very tested GRB correlation is the E_{peak}-E_{iso} relation, also known as the Amati relation.[25] Here, the intrinsic peak energy of the GRB prompt emission estimated within the T_{90} was found to be related with the isotropic energy emitted in the same time interval. This correlation is satisfied by all long GRBs within a large interval for the energy (almost six orders of magnitude). To illustrate the concept of GRBs as standardizable candles, the E_{peak} is the quantity that represents our cosmology-independent observable while E_{iso} is our "candle": the latter quantity is indeed related to the underlying cosmology through the luminosity distance d_l, and quantified by the bolometric fluence, which is given by the integrated GRB prompt flux in a specific energy range.[26,27] Once we derive the distance for a large sample of GRBs, we can start to build an Hubble diagram and then test the cosmological model.[28] The use of the Amati relation has provided for the first time an interesting estimate of the cosmological parameters with GRBs,[29,30] if compared with other non-standard probes.

Observing these events for more than 20 years with space-based satellites such as *Swift*[31] and *Fermi*,[32] has permitted to build large dataset of GRBs. Moreover, other correlations between GRB observed quantities started to appear frequently in literature.[33] The majority of these correlations have also been used to test cosmological models with some interesting applications. A very promising method that has provided some very interesting results is based on the use of a correlation that connects prompt and afterglow quantities, namely the Combo relation.[34] This relation originates from two existing proposed correlations, the previously-cited Amati relation and its extension to X-rays.[18,35] The core of the Combo relation is represented by a light-curve fitting procedure to model the entire X-ray afterglow emission: the use of a code able to model the X-ray light curve and that permits to get rid of

over-imposed flares by using a very efficient iterative method[35] allows to measure the "candle". From this procedure, the X-ray afterglow observables, namely the luminosity of the afterglow L_0, the time when the late time decay starts t_a and the power-law index of the late light curve decay α are computed and have been found in a tight relation with the GRB prompt peak energy E_{peak}.

The final result is a correlation that is satisfied by all long GRBs, e.g. no outliers,[36] for which there exist a known value for E_{peak} and with a complete observed X-ray afterglow by the X-ray Telescope (XRT) instrument on-board *Swift*. About a physical interpretation of the correlation, the connection between prompt and afterglow properties has been investigated only in recent times[37] but its use to constrain the cosmological models and some exotic variant has provided very interesting results (see also Fig. 1). In recent times, Muccino et al.[36] have found a possible indication of an evolution of the dark energy equation of state parameter w with redshift. However, the standard ΛCDM model with the constant solution $w = -1$ is still consistent with the statistical uncertainties derived by their analysis, which suggests the possibility that unaccounted systematics have a great influence on GRB luminosity correlations.

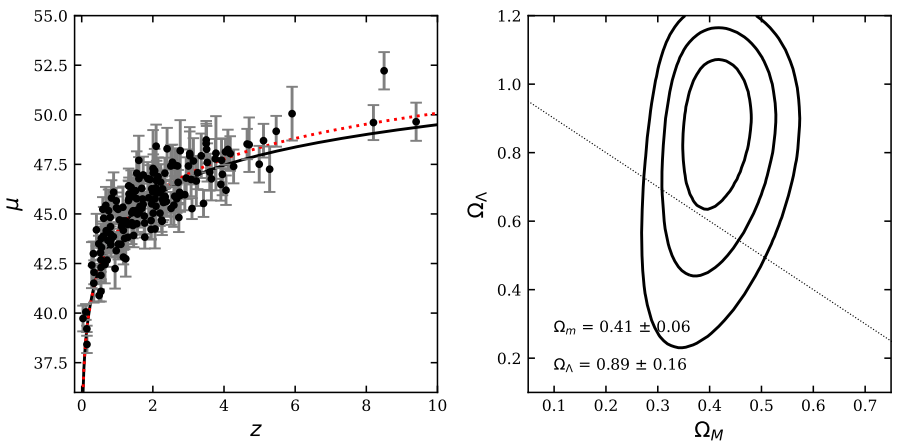


Fig. 1. *(Left panel)* The Hubble diagram built with 174 GRBs satisfying the Combo relation. The black curve corresponds to the best fit model in the context of a non-flat ΛCDM cosmological scenario, while the dotted curve represents a fiducial ΛCDM model with $H_0 = 70$ km/s/Mpc, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. *Right panel* The best-fit confidence regions obtained by fitting the Combo GRB sample of 174 events within the context of a non-flat ΛCDM cosmological scenario. Here H_0 was fixed to 70 km/s/Mpc. Extra scatter uncertainties and systematics have not been included in the analysis.

3. Drawbacks

So far so good! However, it is mandatory to discuss the main sources of uncertainties that affect the use of GRBs as standardized candles. First, the lack of a considerable

number of GRBs at low redshifts ($z < 0.02$) prevents us to use accurate primary distance indicators to calibrate GRB luminosity correlations. The only GRB event observed at very close distances, GRB 980425,[11] does not have a great coverage in X-rays, given the lack of rapidly-slewing detectors such as *Swift* at the epoch of its discovery. Moreover, its very faint emission and luminosity do not assure that GRB 980425 is really a "cosmological" GRB but it can be originated in other energetic events such as the break-out of the shock wave following the collapse of a massive star.[38]

This problem led several authors to use alternative approaches to calibrate GRB relations, such as 1) the use of an underlying cosmology as explained in the review by Schaefer,[27] a method that includes an intrinsic circularity; 2) the use of alternative techniques that approximate the behavior of the luminosity distance evolution with the use of series expansion for the Hubble-Lemaitre law,[39–41] which however lead to serious problems at high-redshifts, or 3) the use of secondary distance indicators such as type-Ia SNe.[40,42] The latter method is very promising, given the high level of accuracy that type-Ia SN cosmology has reached in the last years.[43] However, in this approach the procedure of calibrating GRB luminosity relation is actually a seven-step method (type-Ia SNe are indeed calibrated using other primary distance indicators and so on). Moreover, given that the most used sample of type-Ia SNe, such as the Pantheon[44] spans a large distance range, any cosmological inference obtained using GRBs calibrated with type-Ia SNe will be strongly biased by the cosmology traced by these SNe, being their luminosity correlations characterised by a great accuracy. In this light, GRBs can only be considered as a mere extension of type-Ia SN Hubble Diagram and not an independent probe.

Selection effects and systematics errors also represent a big source of uncertainty when computing the accuracy (e.g. the scatter) of luminosity relations, but they have been rarely taken into account. In literature, it is common to find GRB samples that have been built using data coming from different detectors, mostly at high-energies (hard X-rays), which are characterised by different energy sensitivities, and then sensitive to specific GRB events. Finally, Malmquist bias also has a strong impact on the detection of the brightest events, and then on the final construction of the sample considered. This also affects the GRB redshift identification, which then becomes possible only for the brightest events. Correcting for the Malmquist bias is not easy, given that the distribution of GRB properties (luminosity function, physical characteristics of their progenitors, etc.) can vary with redshift, and then the use of incomplete GRB samples can lead to ambiguous results.[45]

4. Perspectives

Apart from getting more GRBs at high redshifts, a mission that will be the goal of future space missions, attempting to solve the above issues represents a mandatory step to render more reliable the use of GRBs as distance indicators. The scatter of all GRB luminosity correlations is also not competitive if compared with type-Ia

SNe. However, incoming surveys are currently discovering SNe at redshift comparable to the low end of the GRB distance distribution, and with the Rubin Legacy Survey of Space and Time (LSST) project, starting in few years, we can easily arrive to discover several type-Ia SNe at redshift $z \sim 2$. This will permit to compare GRBs and SNe correlations in a wide redshift range, allowing us to refine the calibration of GRB correlations. An important support on this line comes from the use of machine learning algorithms to reconstruct the distance luminosity function with the redshift , independently on the underlying cosmology. Similar approaches have been already developed, e.g.,[46,47] with very promising results. However, also this approach includes some hidden uncertainties that depend on the model used to train the calibrating sample and the sample itself: uncertainties affecting the training sample[48] will propagate on the calibration of GRB correlations, leading to a completely wrong posterior distributions when using GRBs as cosmological tools. Finally, without a correct treatment of the systematics affecting all GRB correlations, all the results should be treated with large caution.

5. Conclusions

GRB luminosity correlations represent a powerful tool to investigate the high redshift Universe, but their use is still affected by several issues that their application as accurate distance indicators is still far to come. The lack of a complete treatment of uncertainties, in particular of the systematics affecting their observables, and of an unbiased calibration technique represents the main barrier to overcome in order to take advantage of the information that these powerful explosions can give us on the early Universe.

Acknowledgments

Li is extremely grateful to Marco Muccino, Elena Zaninoni and Giovanni Battista Pisani for their support and collaboration. LI was supported by grants from VILLUM FONDEN (project num-ber 16599 and 25501).

References

1. B. Zhang, *The Physics of Gamma-Ray Bursts* 2018.
2. A. Levan, *Gamma-Ray Bursts* 2018.
3. D. Band, J. Matteson, L. Ford, B. Schaefer, D. Palmer, B. Teegarden, T. Cline, M. Briggs, W. Paciesas, G. Pendleton, G. Fishman, C. Kouveliotou, C. Meegan, R. Wilson and P. Lestrade, BATSE Observations of Gamma-Ray Burst Spectra. I. Spectral Diversity, *ApJ* **413**, p. 281 (August 1993).
4. C. Kouveliotou, C. A. Meegan, G. J. Fishman, N. P. Bhat, M. S. Briggs, T. M. Koshut, W. S. Paciesas and G. N. Pendleton, Identification of Two Classes of Gamma-Ray Bursts, *ApJ* **413**, p. L101 (August 1993).

5. J. Hjorth, D. Watson, J. P. U. Fynbo, P. A. Price, B. L. Jensen, U. G. Jørgensen, D. Kubas, J. Gorosabel, P. Jakobsson, J. Sollerman, K. Pedersen and C. Kouveliotou, The optical afterglow of the short γ-ray burst GRB 050709, *Nature* **437**, 859 (October 2005).
6. D. Eichler, M. Livio, T. Piran and D. N. Schramm, Nucleosynthesis, neutrino bursts and γ-rays from coalescing neutron stars, *Nature* **340**, 126 (July 1989).
7. R. Perna and K. Belczynski, Short Gamma-Ray Bursts and Mergers of Compact Objects: Observational Constraints, *ApJ* **570**, 252 (May 2002).
8. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A, *ApJ* **848**, p. L14 (October 2017).
9. S. E. Woosley, Gamma-Ray Bursts from Stellar Mass Accretion Disks around Black Holes, *ApJ* **405**, p. 273 (March 1993).
10. A. I. MacFadyen and S. E. Woosley, Collapsars: Gamma-Ray Bursts and Explosions in "Failed Supernovae", *ApJ* **524**, 262 (October 1999).
11. T. J. Galama, P. M. Vreeswijk, J. van Paradijs, C. Kouveliotou, T. Augusteijn, H. Böhnhardt, J. P. Brewer, V. Doublier, J. F. Gonzalez, B. Leibundgut, C. Lidman, O. R. Hainaut, F. Patat, J. Heise, J. in't Zand, K. Hurley, P. J. Groot, R. G. Strom, P. A. Mazzali, K. Iwamoto, K. Nomoto, H. Umeda, T. Nakamura, T. R. Young, T. Suzuki, T. Shigeyama, T. Koshut, M. Kippen, C. Robinson, P. de Wildt, R. A. M. J. Wijers, N. Tanvir, J. Greiner, E. Pian, E. Palazzi, F. Frontera, N. Masetti, L. Nicastro, M. Feroci, E. Costa, L. Piro, B. A. Peterson, C. Tinney, B. Boyle, R. Cannon, R. Stathakis, E. Sadler, M. C. Begam and P. Ianna, An unusual supernova in the error box of the γ-ray burst of 25 April 1998, *Nature* **395**, 670 (October 1998).
12. J. Hjorth, J. Sollerman, P. Møller, J. P. U. Fynbo, S. E. Woosley, C. Kouveliotou, N. R. Tanvir, J. Greiner, M. I. Andersen, A. J. Castro-Tirado, J. M. Castro Cerón, A. S. Fruchter, J. Gorosabel, P. Jakobsson, L. Kaper, S. Klose, N. Masetti, H. Pedersen, K. Pedersen, E. Pian, E. Palazzi, J. E. Rhoads, E. Rol, E. P. J. van den Heuvel, P. M. Vreeswijk, D. Watson and R. A. M. J. Wijers, A very energetic supernova associated with the γ-ray burst of 29 March 2003, *Nature* **423**, 847 (June 2003).
13. B. Zhang, Y. Z. Fan, J. Dyks, S. Kobayashi, P. Mészáros, D. N. Burrows, J. A. Nousek and N. Gehrels, Physical Processes Shaping Gamma-Ray Burst X-Ray Afterglow Light Curves: Theoretical Implications from the Swift X-Ray Telescope Observations, *ApJ* **642**, 354 (May 2006).
14. J. A. Nousek, C. Kouveliotou, D. Grupe, K. L. Page, J. Granot, E. Ramirez-Ruiz, S. K. Patel, D. N. Burrows, V. Mangano, S. Barthelmy, A. P. Beardmore, S. Campana, M. Capalbi, G. Chincarini, G. Cusumano, A. D. Falcone, N. Gehrels, P. Giommi, M. R. Goad, O. Godet, C. P. Hurkett, J. A. Kennea, A. Moretti, P. T. O'Brien, J. P. Osborne, P. Romano, G. Tagliaferri and A. A. Wells, Evidence for a Canonical Gamma-Ray Burst Afterglow Light Curve in the Swift XRT Data, *ApJ* **642**, 389 (May 2006).
15. R. Willingale, P. T. O'Brien, J. P. Osborne, O. Godet, K. L. Page, M. R. Goad, D. N. Burrows, B. Zhang, E. Rol, N. Gehrels and G. Chincarini, Testing the Standard Fireball Model of Gamma-Ray Bursts Using Late X-Ray Afterglows Measured by Swift, *ApJ* **662**, 1093 (June 2007).

16. D. A. Kann, S. Klose, B. Zhang, D. Malesani, E. Nakar, A. Pozanenko, A. C. Wilson, N. R. Butler, P. Jakobsson, S. Schulze, M. Andreev, L. A. Antonelli, I. F. Bikmaev, V. Biryukov, M. Böttcher, R. A. Burenin, J. M. Castro Cerón, A. J. Castro-Tirado, G. Chincarini, B. E. Cobb, S. Covino, P. D'Avanzo, V. D'Elia, M. Della Valle, A. de Ugarte Postigo, Y. Efimov, P. Ferrero, D. Fugazza, J. P. U. Fynbo, M. Gålfalk, F. Grundahl, J. Gorosabel, S. Gupta, S. Guziy, B. Hafizov, J. Hjorth, K. Holhjem, M. Ibrahimov, M. Im, G. L. Israel, M. Jeĺinek, B. L. Jensen, R. Karimov, I. M. Khamitov, Ü. Kiziloğlu, E. Klunko, P. Kubánek, A. S. Kutyrev, P. Laursen, A. J. Levan, F. Mannucci, C. M. Martin, A. Mescheryakov, N. Mirabal, J. P. Norris, J. E. Ovaldsen, D. Paraficz, E. Pavlenko, S. Piranomonte, A. Rossi, V. Rumyantsev, R. Salinas, A. Sergeev, D. Sharapov, J. Sollerman, B. Stecklum, L. Stella, G. Tagliaferri, N. R. Tanvir, J. Telting, V. Testa, A. C. Updike, A. Volnova, D. Watson, K. Wiersema and D. Xu, The Afterglows of Swift-era Gamma-ray Bursts. I. Comparing pre-Swift and Swift-era Long/Soft (Type II) GRB Optical Afterglows, *ApJ* **720**, 1513 (September 2010).
17. D. A. Frail, S. R. Kulkarni, R. Sari, G. B. Taylor, D. S. Shepherd, J. S. Bloom, C. H. Young, L. Nicastro and N. Masetti, The Radio Afterglow from GRB 980519: A Test of the Jet and Circumstellar Models, *ApJ* **534**, 559 (May 2000).
18. M. G. Bernardini, R. Margutti, E. Zaninoni and G. Chincarini, A universal scaling for short and long gamma-ray bursts: $E_{X,iso}$ - $E_{\gamma,iso}$ - E_{pk}, *MNRAS* **425**, 1199 (September 2012).
19. N. Kawai, G. Kosugi, K. Aoki, T. Yamada, T. Totani, K. Ohta, M. Iye, T. Hattori, W. Aoki, H. Furusawa, K. Hurley, K. S. Kawabata, N. Kobayashi, Y. Komiyama, Y. Mizumoto, K. Nomoto, J. Noumaru, R. Ogasawara, R. Sato, K. Sekiguchi, Y. Shirasaki, M. Suzuki, T. Takata, T. Tamagawa, H. Terada, J. Watanabe, Y. Yatsu and A. Yoshida, An optical spectrum of the afterglow of a γ-ray burst at a redshift of z = 6.295, *Nature* **440**, 184 (March 2006).
20. J. Greiner, C. Clemens, T. Krühler, A. von Kienlin, A. Rau, R. Sari, D. B. Fox, N. Kawai, P. Afonso, M. Ajello, E. Berger, S. B. Cenko, A. Cucchiara, R. Filgas, S. Klose, A. Küpcü Yoldaş, G. G. Lichti, S. Löw, S. McBreen, T. Nagayama, A. Rossi, S. Sato, G. Szokoly, A. Yoldaş and X. L. Zhang, The redshift and afterglow of the extremely energetic gamma-ray burst GRB 080916C, *A&A* **498**, 89 (April 2009).
21. N. R. Tanvir, D. B. Fox, A. J. Levan, E. Berger, K. Wiersema, J. P. U. Fynbo, A. Cucchiara, T. Krühler, N. Gehrels, J. S. Bloom, J. Greiner, P. A. Evans, E. Rol, F. Olivares, J. Hjorth, P. Jakobsson, J. Farihi, R. Willingale, R. L. C. Starling, S. B. Cenko, D. Perley, J. R. Maund, J. Duke, R. A. M. J. Wijers, A. J. Adamson, A. Allan, M. N. Bremer, D. N. Burrows, A. J. Castro-Tirado, B. Cavanagh, A. de Ugarte Postigo, M. A. Dopita, T. A. Fatkhullin, A. S. Fruchter, R. J. Foley, J. Gorosabel, J. Kennea, T. Kerr, S. Klose, H. A. Krimm, V. N. Komarova, S. R. Kulkarni, A. S. Moskvitin, C. G. Mundell, T. Naylor, K. Page, B. E. Penprase, M. Perri, P. Podsiadlowski, K. Roth, R. E. Rutledge, T. Sakamoto, P. Schady, B. P. Schmidt, A. M. Soderberg, J. Sollerman, A. W. Stephens, G. Stratta, T. N. Ukwatta, D. Watson, E. Westra, T. Wold and C. Wolf, A γ-ray burst at a redshift of z~8.2, *Nature* **461**, 1254 (October 2009).
22. A. Cucchiara, A. J. Levan, D. B. Fox, N. R. Tanvir, T. N. Ukwatta, E. Berger, T. Krühler, A. Küpcü Yoldaş, X. F. Wu, K. Toma, J. Greiner, F. E. Olivares, A. Rowlinson, L. Amati, T. Sakamoto, K. Roth, A. Stephens, A. Fritz, J. P. U. Fynbo, J. Hjorth, D. Malesani, P. Jakobsson, K. Wiersema, P. T. O'Brien, A. M. Soderberg, R. J. Foley, A. S. Fruchter, J. Rhoads, R. E. Rutledge, B. P. Schmidt, M. A. Dopita, P. Podsiadlowski, R. Willingale, C. Wolf, S. R. Kulkarni and P. D'Avanzo, A Photometric Redshift of z ~9.4 for GRB 090429B, *ApJ* **736**, p. 7 (July 2011).

23. N. R. Tanvir, A. J. Levan, A. S. Fruchter, J. P. U. Fynbo, J. Hjorth, K. Wiersema, M. N. Bremer, J. Rhoads, P. Jakobsson, P. T. O'Brien, E. R. Stanway, D. Bersier, P. Natarajan, J. Greiner, D. Watson, A. J. Castro-Tirado, R. A. M. J. Wijers, R. L. C. Starling, K. Misra, J. F. Graham and C. Kouveliotou, Star Formation in the Early Universe: Beyond the Tip of the Iceberg, *ApJ* **754**, p. 46 (July 2012).
24. K. S. Mandel, G. Narayan and R. P. Kirshner, Type Ia Supernova Light Curve Inference: Hierarchical Models in the Optical and Near-infrared, *ApJ* **731**, p. 120 (April 2011).
25. L. Amati, F. Frontera, M. Tavani, J. J. M. in't Zand, A. Antonelli, E. Costa, M. Feroci, C. Guidorzi, J. Heise, N. Masetti, E. Montanari, L. Nicastro, E. Palazzi, E. Pian, L. Piro and P. Soffitta, Intrinsic spectra and energetics of BeppoSAX Gamma-Ray Bursts with known redshifts, *A&A* **390**, 81 (July 2002).
26. L. Amati, The $E_{p,i}$-E_{iso} correlation in gamma-ray bursts: updated observational status, re-analysis and main implications, *MNRAS* **372**, 233 (October 2006).
27. B. E. Schaefer, The Hubble Diagram to Redshift ∼6 from 69 Gamma-Ray Bursts, *ApJ* **660**, 16 (May 2007).
28. L. Izzo, S. Capozziello, G. Covone and M. Capaccioli, Extending the Hubble diagram by gamma ray bursts, *A&A* **508**, 63 (December 2009).
29. L. Amati, C. Guidorzi, F. Frontera, M. Della Valle, F. Finelli, R. Landi and E. Montanari, Measuring the cosmological parameters with the $E_{p,i}$-E_{iso} correlation of gamma-ray bursts, *MNRAS* **391**, 577 (December 2008).
30. L. Amati and M. Della Valle, Measuring Cosmological Parameters with Gamma Ray Bursts, *International Journal of Modern Physics D* **22**, p. 1330028 (December 2013).
31. N. Gehrels, E. Ramirez-Ruiz and D. B. Fox, Gamma-Ray Bursts in the Swift Era, *ARA&A* **47**, 567 (September 2009).
32. W. S. Paciesas, C. A. Meegan, A. von Kienlin, P. N. Bhat, E. Bissaldi, M. S. Briggs, J. M. Burgess, V. Chaplin, V. Connaughton, R. Diehl, G. J. Fishman, G. Fitzpatrick, S. Foley, M. Gibby, M. Giles, A. Goldstein, J. Greiner, D. Gruber, S. Guiriec, A. J. van der Horst, R. M. Kippen, C. Kouveliotou, G. Lichti, L. Lin, S. McBreen, R. D. Preece, A. Rau, D. Tierney and C. Wilson-Hodge, The Fermi GBM Gamma-Ray Burst Catalog: The First Two Years, *ApJS* **199**, p. 18 (March 2012).
33. M. Dainotti, R. Del Vecchio and M. Tarnopolski, Gamma Ray Burst Prompt correlations, *arXiv e-prints*, p. arXiv:1612.00618 (December 2016).
34. L. Izzo, M. Muccino, E. Zaninoni, L. Amati and M. Della Valle, New measurements of Ω_m from gamma-ray bursts, *A&A* **582**, p. A115 (October 2015).
35. E. Zaninoni, M. G. Bernardini, R. Margutti, S. Oates and G. Chincarini, Gamma-ray burst optical light-curve zoo: Comparison with X-ray observations, *A&A* **557**, p. A12 (September 2013).
36. M. Muccino, L. Izzo, O. Luongo, K. Boshkayev, L. Amati, M. Della Valle, G. B. Pisani and E. Zaninoni, Tracing Dark Energy History with Gamma-Ray Bursts, *ApJ* **908**, p. 181 (February 2021).
37. G. Oganesyan, S. Ascenzi, M. Branchesi, O. S. Salafia, S. Dall'Osso and G. Ghirlanda, Structured Jets and X-Ray Plateaus in Gamma-Ray Burst Phenomena, *ApJ* **893**, p. 88 (April 2020).
38. E. Nakar and R. Sari, Relativistic Shock Breakouts—A Variety of Gamma-Ray Flares: From Low-luminosity Gamma-Ray Bursts to Type Ia Supernovae, *ApJ* **747**, p. 88 (March 2012).
39. M. Visser, Cosmography: Cosmology without the Einstein equations, *General Relativity and Gravitation* **37**, 1541 (September 2005).
40. S. Capozziello and L. Izzo, Cosmography by gamma ray bursts, *A&A* **490**, 31 (October 2008).

41. V. Vitagliano, J.-Q. Xia, S. Liberati and M. Viel, High-redshift cosmography, **2010**, p. 005 (March 2010).
42. M. Demianski, E. Piedipalumbo, D. Sawant and L. Amati, Cosmology with gamma-ray bursts. I. The Hubble diagram through the calibrated $E_{p,I}$-E_{iso} correlation, *A&A* **598**, p. A112 (February 2017).
43. A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn and D. Scolnic, Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM, *ApJ* **908**, p. L6 (February 2021).
44. D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak M. McCrum, K. W. Smith, The Complete Light-curve Sample of Spectroscopically Confirmed Type la Supernovae from Pan-STARRS1 and Cosmological Constraints from The Combined Pantheon Sample, *ApJ*, **859**, p. 101 (2018).
45. A. Pescalli, G. Ghirlanda, R. Salvaterra, G. Ghisellini, S. D. Vergani, F. Nappo, O. S. Salafia, A. Melandri, S. Covino and D. Götz, The rate and luminosity function of long gamma ray bursts, *A&A* **587**, p. A40 (March 2016).
46. O. Luongo and M. Muccino, Model-independent calibrations of gamma-ray bursts using machine learning, *MNRAS* **503**, 4581 (May 2021).
47. L. Tang, X. Li, H.-N. Lin and L. Liu, Model-independently Calibrating the Luminosity Correlations of Gamma-Ray Bursts Using Deep Learning, *ApJ* **907**, p. 121 (February 2021).
48. A. Ahlstrom Kjerrgren and E. Mortsell, On the use of galaxies as clocks and the universal expansion, *arXiv e-prints*, p. arXiv:2106.11317 (June 2021).

Surface brightness fluctuations: The method and future applications

Michele Cantiello

INAF – Astronomical Observatory of Abruzzo,
Via Maggini, 64100, Teramo, Italy
E-mail: michele.cantiello@inaf.it

The surface brightness fluctuations method is one of the most robust extragalactic distance indicators: with an accuracy comparable to type Ia supernovae Ia and Cepheids ($\sim 5\%$ per galaxy) it is used in a wide range of distances: from very local, up to values relevant for measuring H_0.

In these proceedings, I present an overview of the technique and of the main results it allowed to obtain, further discussing the great potential with future applications based on forthcoming ground and space-based facilities.

Keywords: Cosmological Parameters; Hubble-Lemaitre Constant; Galaxy Distances

1. Introduction

The idea of using stellar counts to study the distance or the star content of a stellar population dates back to the first half of the previous century.[1,2] With the advent of CCDs, Ref. 3 introduced the surface brightness fluctuations (SBF hereafter) method as a modern formalism to obtain distances to stellar systems relying on the discrete nature of star counts. The SBF definition by Ref. 3 uses the Poisson statistics of star counts to measure a quantity that is closely linked to the mean brightness of red giant branch (RGB) stars in the target galaxy or, more in general, in the observed stellar system.

To date, the SBF method is one of the most robust extragalactic distance indicators: with an accuracy comparable to type Ia supernovae (SNe Ia) or Cepheids ($\sim 5\%$ per galaxy) without the limitations of the serendipitousness intrinsic to SNe, or the long observing campaigns needed for Cepheids. Moreover, it is used in a wide range of distances: from the Local Group, up to values relevant for measuring H_0.

Qualitatively, the idea behind the SBF is relatively simple: the stars which appear as resolved sources at relatively small distances, gradually "blend" in a smooth brightness profile when the stellar system moves to larger and larger distances (Figure 1).

Quantitatively, *SBFs are defined as the ratio of the second to the first moment of the stellar luminosity function in a population of stars.* From this basic definition some of the most relevant characteristics of the method follow: how SBF can be measured and what are the inherent complexities, what are the preferred targets, the opportunity of calibrating SBF magnitudes using empirical methods or stellar population synthesis techniques, etc.

In what follows, I will overview the technique, providing a detailed description of the method and on how measurements are carried out, what are the main limitations and sources of uncertainty, and present its great potential with applications based on forthcoming ground and space-based telescopes.

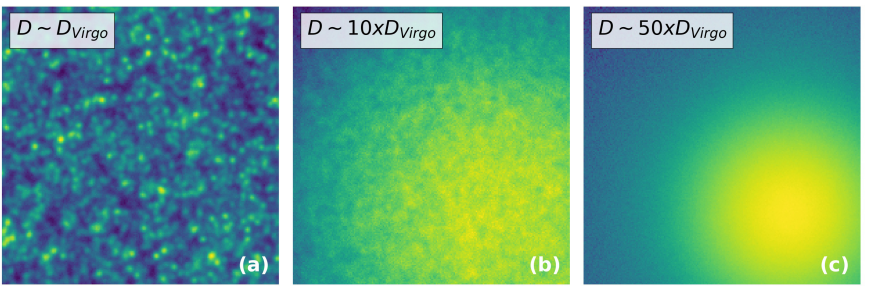


Fig. 1. Qualitative illustration of SBF signal. Panel (a): Simulation of the stellar population in a spheroidal galaxy at the distance of the Virgo cluster ($d \simeq 16.5$ Mpc, Ref. 4) as will be observed with the E-ELT in ∼1 hour exposure time (Cantiello et al., 2022, in prep.). (b) Same as in panel (a), but for a galaxy ten times more distant. Panel (c): Same as in panel (a), but for a galaxy fifty times more distant. Stars, which appear marginally resolved in panel (a), blend into a smooth brightness profile at larger distances.

2. From SBF analysis to galaxy distances

The SBF signal is not easy visualize as it might be for various other classical standard rulers (variable stars, parallaxes, SNe Ia etc.) and, consequently, it does not appear trivial how star-counts fluctuations can be translated into a distance estimate. For this reason, below I provide a concise and detailed description on how the SBF magnitudes $\overline{m}$ are measured, what are the preferred targets, how distances are then derived, and what are the dominant sources of errors.

2.1. *Measurements*

2.1.1. *Determining SBF magnitudes*

The SBF signal, in the absence of instrumental blurring and of other sources of fluctuations but the stellar counts, would simply be the Poisson variance between pixels due to the varying number of stars per pixel, normalized to the galaxy local mean flux. The instrumental point spread function (PSF) blurring forces a correlation between adjacent resolution elements, thus the SBF signal results to be convolved with the PSF, and its measure needs to be carried out in the Fourier domain.

A further complication is that there are sources of *spurious* non-stellar fluctuations. The presence of globular clusters (GCs) in the galaxy, of background galaxies in the observed field, of morphological irregularities due to dust contamination, or

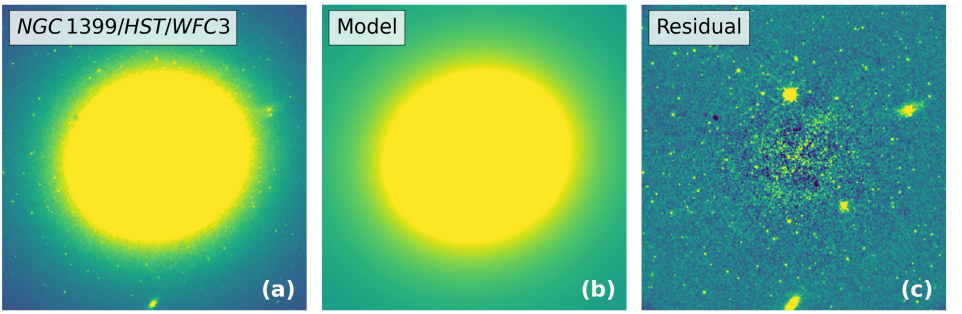


Fig. 2. Illustration of the SBF measurements. (*a*) Near-infrared image of NGC 1399 from the HST WFC3 camera. (*b*) Model of NGC 1399's surface brightness distribution derived from the WFC3/IR image. (*c*) Residual frame, obtained from the galaxy image (*a*) minus the model (*b*).

of obvious regions with markedly different stellar populations (e.g. star forming regions, overlapping regions of companion bright galaxies, etc.), affect the Poisson stellar fluctuation counts. Hence, to estimate fluctuation magnitude $\overline{m}$ all these possible sources of contamination need to be taken into account and removed.

The details of the procedures adopted for measuring SBFs are described in details in a number of papers.[5–12] Here, we summarize the main steps of the procedure, which consists of deriving the:

i) *average galaxy surface brightness profile*: this corresponds to the first moment of the host stellar luminosity function (model frame in Fig. 2 (*b*));

ii) *residual frame*: the galaxy subtracted frame. This image basically corresponds to the second moment of the luminosity function (Fig. 2 Panel (*c*));

iii) *mask frame*: all sources of non-stellar variance (dust, GCs, background galaxies, etc.) are identified and masked out;

iv) *residual, masked frame normalized to the model frame*: this image contains the bumpiness from the stellar fluctuations, plus the fluctuations of un-excised sources;

iv) *power spectrum of the total fluctuations frame*. The fluctuations are convolved with the instrumental PSF, thus in the Fourier domain they are multiplied to the Fourier transform of the PSF. Once derived the PSF and its power spectrum, the SBF amplitude is obtained as the constant term P_0 giving the best match between the power spectrum of the total fluctuations frame, $P(K)$, to the power spectrum of the PSF, $E(k)$, plus a white noise P_1 component: $P(k) = P_0 \times E(k) + P_1$. Figure 3 (left panel) shows a schematic example of the fitting procedure;

v) *correction for spurious residual fluctuations*. The masking process is effective down to a given magnitude limit: the deeper the magnitude limit the higher the detection limit of contaminating sources. However, the spurious fluctuations arising from GCs and background galaxies fainter than the detection limit need to be evaluated and subtracted out. A fit of the combined luminosity functions (LF hereafter)

of GCs and background galaxies above the limiting magnitude is used to obtain by extrapolation the extra-fluctuations term, P_r, due to sources fainter than the detection limit. An example of the total LF fit, derived as the sum of the Gaussian LF for GCs in the galaxy (GCLF) and a power law LF for background galaxies, is shown in Fig. 3 (right panel).

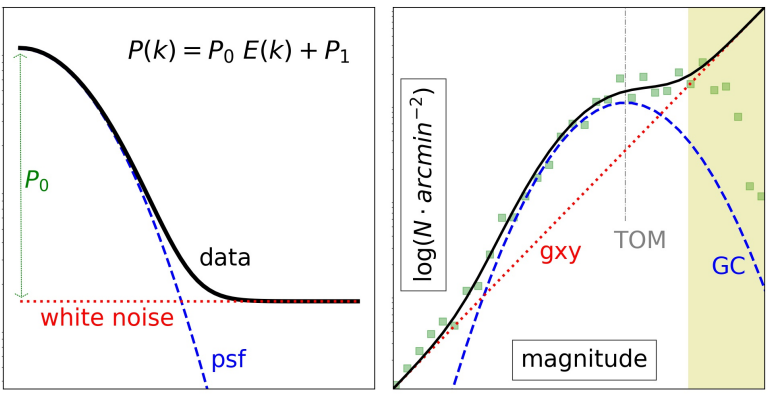


Fig. 3. Left: A schematic illustration of the SBF power spectrum analysis. Right panel: Typical luminosity function analysis for estimating the residual variance P_r due to contaminating sources: green squares show the data, the blue curve and red line show the fits to the GCs and background galaxies LFs, respectively, and the solid black line is the combined model luminosity function (data and fits are from Ref. 13). The vertical grey dashed line indicates the GCLF turnover magnitude, TOM, and the shaded area shows the magnitude interval where the detection is incomplete.

The residual variance is then evaluated from the integral of the second moment of the total luminosity function in the flux interval $[0, f_{lim}]$:

$$\sigma_r^2 = \int_0^{f_{lim}} n_{obj}(f) \times f^2 df \tag{1}$$

where f_{lim} is the flux corresponding to the limiting magnitude, m_{lim}, and $n_{obj}(f)$ is the GC and background galaxies combined LF (Fig. 3, right panel). The residual variance P_r is then derived from σ_r^2 normalized by the galaxy surface brightness;

vi) *SBF magnitudes*: once measured the total fluctuation amplitude, P_0, and estimated the spurious fluctuations term P_r, the stellar fluctuations follows as $P_f = P_0 - P_r$, and then the SBF magnitude as $\bar{m} = -2.5 \log(P_f) + mag_{z.p.}$, where the $mag_{z.p.}$ term is the instrumental zero point magnitude.

Finally, to convert $\overline{m}$ into a distance, the characteristic absolute SBF magnitude of the population, $\bar{M}$, needs to be known. By combining the absolute and apparent SBF magnitudes, the distance modulus μ_0 follows as usual: $\mu_0 = \bar{m} - \bar{M}$ and, then, the distance d is derived by inverting relation $\mu_0 = 5 \log(d/Mpc) + 25$.

2.1.2. *Calibrating SBF magnitudes*

To obtain a calibration of $\bar{M}$, to be used for deriving μ_0, two different approaches are typically adopted: empirical calibrations or population synthesis models.

$\bar{M}$ can be described as a one-parameter family, with the integrated color of the galaxy (or, more in general, of the stellar population) used as main correlation parameter, i.e. $\overline{M}=\alpha + \beta \times color$; at fixed pass-band $\bar{M}$ is a universal function of the color, with residual intrinsic dispersion which can be as ~ 0.05 mag depending on the pass-band.[4,9]

Using theoretical calibrations the SBF method is a primary distance indicator: it does not carry over the systematic uncertainty from the first steps of the calibration. Despite the generally good match between predictions from independent stellar population synthesis models, this is mostly limited to the optical bands. The matching of SBF magnitudes from different models in the blue passbands (more sensitive to stellar population properties), or in the near-IR (where the stellar models are yet not as robust is in the optical bands), is of the order of ~ 0.2 mag. The difference between models is then significantly larger than the estimated SBF cosmic scatter derived empirically.

Finally, although there are obvious advantages from using calibrations based on stellar population models, to date most of the available SBF distances rely on empirical calibrations.[4,11,14,15]

2.1.3. *Caveats in measuring SBF*

One of the major problems in measuring SBF is the contamination from GCs. As the distance of the galaxy increases, the detection of GCs gets more complicated by the combination of GCs crowding and flux limit, consequently the contamination from GCs increases.

Since the estimate of the contamination from unidentified GCs is made by extrapolating the observed GCLF at magnitude levels fainter than the limiting magnitude (Fig. 3, right panel), the brighter the limiting magnitude is, the greater the uncertainty due to the extrapolation is, and the larger is the P_r correction. The relative error on P_r is typically assumed to be of the order of 20-25%,[5,16] hence a reliable SBF measurement requires $P_f >> P_r$. A P_r correction less than 10% P_f, or lower, is an essential condition for obtaining reasonably robust estimates of SBF and, therefore, reliable distances. In optical bands, this condition is typically satisfied when the limiting magnitude is at least one magnitude fainter than the GCLF peak magnitude (TOM in Fig. 3, right). In near-infrared bands this condition is less stringent, as the contrast between M_{TOM} and $\bar{M}$ is more favourable to SBF than in optical bands. Even a limiting magnitude m_{lim} close to the TOM, or one magnitude brighter then that, can be enough for accurate measurements.

The present distance limit for SBF is then due to GCs crowding confusion and detection limit. The level of such limit improves with space-based data, where the PSF is typically sharper ($\sim 0''.1$ for HST observations).

The contamination due to background galaxies is normally a negligible problem compared to GCs. GCs have a spatial distribution strongly concentrated on the galaxy, while the density of background galaxies per angular unit is roughly constant over the observed field, and substantially smaller than GCs in the bright galaxy regions where the SBF signal is measured.

2.2. *Preferred targets and pass-bands*

Currently, the bands considered for SBF analysis range from the $\sim I$ (central wavelength $\sim 7.5 - 8.0\mu m$) to $\sim K$ (at $\sim 2.2\mu m$). SBF at shorter wavelengths are not of great interest, due to the non-negligible dependence on the properties of stellar populations in these bands,[7,16] combined with the more dramatic effect of the contamination of GCs and dust.

The small intrinsic scatter of $\overline{m}$, empirically constrained to ~ 0.05 mag, comes from optical studies in the I and z bands for massive ellipticals. To date, the intrinsic dispersion of $\overline{m}$ in the near-IR bands is much less constrained, but it appears to be a factor of two larger than in the optical bands, likely because of the larger scatter arising from bright giant branch stars.[17] Hence, the preference would seem to be for bands like I or z due to the advantages from the smaller irreducible intrinsic scatter, there are undeniable advantages in near-IR bands. At longer wavelengths SBFs are intrinsically brighter, there is a larger and more favourable contrast with respect to GCs –hence the contamination from these sources is less dramatic–, and the presence of residual dust contamination is typically negligible.

In addition to such advantages, it is worth recalling that some of the key facilities of the next decade will have their highest efficiency at near-IR wavelengths (Sect. 3).

Given the peculiarities of SBF, the ideal target for the measurements is any elliptical galaxy free of dust, with a well identifiable GC system and with a relatively homogeneous population of stars: in other words a passively evolving, red and dead galaxy.

Empirical calibrations show that for bright red galaxies, the linear correlations between $\overline{M}$ and colors have a markedly lower scatter then in the fainter blue galaxies regime. In fact, SBF distances for massive galaxies reach very small statistical uncertainty, 5% or lower.[4,9,10,12] The analysis of integrated colors and SBF for intermediate-mass galaxies, shows that they host a more complex stellar population, consequently the intrinsic scatter of the $\bar{M}$-vs-color relations is larger than for red massive galaxies. Finally, bluer dwarf ellipticals can host stars with low metallicities and/or young ages, or a combination of both. This results in an observed scatter $\gtrsim 0.2$ mag, and statistical uncertainties on distances $\gtrsim 10\%$ due to the calibration effects alone.[12]

In conclusion, red massive galaxies are the ideal target for SBF, yet red bulges of spiral galaxies, or low mass dwarf ellipticals, or any other non-ideal SBF target, represent a plausible object for SBF measurements in all cases where there are no reliable methods to estimate distances.

2.3. *Uncertainties*

2.3.1. *Main Statistical Errors*

We can group statistical uncertainties affecting SBF in two main categories: uncertainties from the measurement of the fluctuations and uncertainties related to the calibration.

For the uncertainties directly linked to the SBF measurement, we mention the accuracy of the galaxy model (item i in section 2.1.1), the fit of the external sources (item v), the fit of the spatial power spectrum (item iv), the determination of the correct normalization to the PSF (which is a systematic uncertainty for a given data set), and the presence of undetected dust.

Concerning the calibration, excluding the zero-point, the total uncertainty of $\bar{M}$ stems from the combination of the uncertainty on the integrated color of the galaxy and of the intrinsic dispersion of the calibration equation (irreducible and wavelength dependent) due to the dispersion of the properties of the stellar population in the galaxy.

Although numerous, statistical uncertainties can be individually minimized by means of suitable observing and data analysis strategies. As reported in the various works,[11,18,19] the total statistical uncertainty on $\overline{m}$ can be 0.05 mag or smaller, while that on $\bar{M}$ is limited by cosmic scatter which, in some bands, can reach levels of 0.06 mag.[4]

In summary, statistical uncertainties in well-conceived observations can be kept at levels below 3% on the distance of the single galaxy.

2.3.2. *Systematic Uncertainties*

For data-sets from different instruments, the only systematic source of error on SBF arises from the calibration zero-point.

Most of the empirical calibrations reported in the literature are based on the zero-point from the SBF ground-based survey by Tonry and collaborators,[14] which obtained SBF for bulges of spiral galaxies with known distance estimates from the Cepheids P-L relations.

In Ref. 11 a total systematic uncertainty of 0.10 mag on SBF distance moduli is cited. This includes a contribution of ~ 0.08 mag from the tie between SBF and Cepheids,[20] and a contribution of ~ 0.06 mag for the Cepheid zero-point, dominated by uncertainty in the distance of the LMC.[21]

Ref. 22 have recently revised the LMC distance using a sample of 20 detached eclipsing binaries, the new distance is $\sim$1% lower than the one used Tonry & coll. and, more important than that, the LMC distance is obtained to 1% precision. The increased accuracy of the LMC distance brings the overall Cepheid zero-point error to 0.028 mag.[23] Consequently, besides a minor effect on the absolute calibration of $\bar{M}$, this revision lowers the systematic error on $\bar{M}$ to ~ 0.085 mag, slightly more than $\sim 4\%$ on the distance.

An alternative to further lower the systematic error derives from the consideration that the ideal target for SBF are elliptical or lenticular galaxies (Sect. 2.2), which have no distance estimates via Cepheid P-L relations. Classical Cepheid stars are high-mass young stars, preferentially found in the dust rich environments of irregular and spiral galaxies, namely the type of target of no interest for SBF. As an alternative to Cepheids, the RGB Tip method (TRGB[24]) is the ideal candidate for measuring distances of early-type galaxies and obtaining an independent calibration of SBF.

A first attempt to calibrate SBF with TRGB, by Ref. 25, used a sample of 16 galaxies within 10 Mpc dominated by relatively blue objects. The authors found negligible changes on distances. However, obtaining a significant sample of SBF and TRGB distances for massive red elliptical galaxies requires reaching the Virgo cluster, and there are few TRGB distances to early-type galaxies at this distance.

Ref. 19 ([B21] hereafter) analyzed the few galaxies for which these measures exist: the massive bright ellipticals M 60 and M 87 in Virgo, and the dusty merger remnant NGC 1316 in Fornax. Excluding the problematic calibrator NGC 1316, the authors find that the mean offset between the SBF and TRGB distances of the two galaxies in Virgo is approximately -0.01 mag. Although preliminary, this result shows no significant difference between Cepheid- and TRGB-based SBF zero-points.

3. Results and future applications

3.1. *Recent results*

The SBF method has been in use for more than three decades. By counting the survey works with more than ~ 50 targets measured, the ~ 600 distances in the range from Local Group to 130 Mpc[a] have proven to be a valuable resource for numerous studies. The reliability and precision of these surveys is quite heterogeneous, with errors on the derived distances typically of 4-5% from HST observations, and of $\sim$10% or more for ground-based SBF (ignoring systematic zero-point uncertainty).

These studies covered, or allowed to cover, a wide range of scientific topics in which the measurement of the distance plays a key role. Among the others, we cite *i*) the measurement of distances for massive galaxies to obtain robust relationship between the mass of the black hole and the bulk dynamics of the galaxy;[18] *ii*) the distance to the first observed dark-matter deficient ultra-diffuse galaxy;[26] *iii*) the detailed structure of the Virgo cluster and the measure of the relative Fornax to Virgo cluster distance with exquisite and still unpaired precision;[4] *iv*) the most precise distance to the host galaxy of the binary neutron star merger event GW170817, i.e. the only GW event also observed in the EM spectrum.

[a]Without checking for duplicates, we included 300 galaxies from Ref. 14, 130 from Ref. 15 and Ref. 18, 140 from Ref. 4 from the ACSFCS and ACSVCS (see also Ref. 9), 80 from Ref. 11 based on the NGVS survey soon to be complemented with 200 more (Cantiello et al., 2022, in prep.).

The study of the Hubble-Lemaitre parameter H_0 deserves a more detailed discussion. Various estimates of this key cosmological parameter have been obtained,[14,27–29] where SBF are either used to calibrate other distance indicators, like SNe Ia, or directly adopted to estimate the local H_0 value. Here, we focus on two recent studies from Ref. 30 and B21.

Ref. 30 presented a re-calibration of the peak magnitude for a sample of 24 local SNe Ia in galaxies based on SBF distances. The authors adopted a rather heterogeneous sample of SBF-based distances, collected from sets of observations spanning over twenty years –combined with a similarly heterogeneous sample of SNeIa. In spite of the difficulties of renormalizing the SBF-distances to a common reference, the heterogeneity of the sample brings with it some advantages, like the possibility to dramatically lower the impact of any instrument- or data-analysis driven systematic effect (e.g. linked to the PSF templates used). Adopting a hierarchical Bayesian approach, the authors extended the calibration to a sample of SNe Ia at redshift $0.02 < z < 0.08$, thus obtaining $H_0 = 71.25 \pm 2.37(\text{stat.}) \pm 3.38(\text{sys.})\text{km s}^{-1}\text{ Mpc}^{-1}$[b]. Such estimate lies in between the value by Ref. 23 based on SNe Ia and Cepheids ($H_0 = 74.03 \pm 1.42$ km s^{-1} Mpc^{-1}including systematics), and the value by Ref. 31 ($H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1}obtained assuming the standard Λ CDM Cosmology). Hence, the estimate by Ref. 30 lies at $\sim 1\sigma$ from the H_0 estimates which started the debate on the "tension" in the Hubble-Lemaitre constant.[23,31]

B21, using a catalog of SBF distances for 63 early-type galaxies out to 100 Mpc obtained from WFC3/IR HST data,[18] revised the Hubble-Lemaitre parameter to $H_0 = 73.3 \pm 0.7(\text{stat.}) \pm 2.4(\text{sys.})$ km s^{-1} Mpc^{-1}. An estimate that agrees within $\sim 0.5\sigma$ with the value from SNe Ia, and a discrepancy larger than $\sim 5\sigma$ with respect to the estimates from Cosmological models and CMB anisotropy measurements (we ignored the systematic uncertainties on SBF). The H_0 by B21 is from SBF distance estimates for galaxies in the range of 15 and 100 Mpc: a distance where peculiar velocities can have a significant impact on the recessional velocities of the galaxies. Hence a regime where accurate velocities are equally important as accurate distances. In their analysis, B21 included as a test the analysis of four different treatments of the galaxy velocities: group velocities, individual velocities, and two flow-corrected model velocities to verify the robustness of their result within the reported errors. It is also worth noting that, as mentioned in Sect. 2.3, B21 adopts a joint calibration averaging the results for the Cepheid and TRGB which reduced uncertainty on H_0.

The work by B21 demonstrates that, in the next future, using a richer sample of massive galaxies with near-IR SBF within 100 Mpc and with improved constraints on the calibrations and on the cosmic-flow model, it will be possible to obtain SBF-based H_0 estimates at 1% or lower statistical uncertainty. Lowering

[b]The final value adopted by Ref. 30, $H_0 = 70.50$km s^{-1} Mpc^{-1}uses the old LMC zero-point; using the revised LMC distance, reported in the references cited in section 2.3, there is a solid 0.023 mag shift toward fainter distance moduli, which leads to an H_0 value of 71.25 km s^{-1} Mpc^{-1}.

the residual 3-4% systematic uncertainty, requires improving zero-point calibration effort as described in section 2.3 and in the next section.

It should be noted that the estimates of Ref. 30 & B21 agree within $\sim 1\sigma$ the respective statistical errors, also due to the relatively large uncertainty of the former estimate because of the heterogeneity of the sample and to the fact that Ref. 30 has an extra calibration step (SNe Ia calibration with SBF) compared to B21.

3.2. *Future applications*

The perspectives of this method can be divided into two three categories: "going wide" with the next generation survey telescopes, "going deep" with the JWST and AO-assisted facilities at 30-40m class telescopes, and "going precise" with programs aimed at lowering systematic errors and with new stellar population models.

3.2.1. *Large and deep sky surveys*

The forthcoming large sky surveys, including the LSST by the Vera Rubin Observatory and the Euclid wide survey, will produce a breakthrough in many fields of Astronomy. The expectations for SBF measurements, hence for mapping the 3D distribution of galaxies in the low-redshift Universe, are rather high.

At the time of writing, the LSST is expected to produce multi-band optical *ugrizy* observations, with typical seeing of 0.7", and 5σ point source depth of $i \sim 26.8$ mag for stationary sources after the 10 years of the main survey. Adopting the i as reference pass-band, we predict LSST will allow reaching bright massive galaxies at ~ 70 Mpc, that is the distance where the TOM+1 mag is reached at $\sim 5\sigma$ point source depth[c].

The Euclid satellite has about one fourth the collecting area of HST, but compared to LSST it has the advantage of the near-IR observations[d] and the sharp PSF (~ 0.2") allowed by space-based observations. Taking as reference the Euclid/NISP H-band, which is predicted reaching 5σ point source depth of $H \sim 24$ mag, and adopting $H_{TOM} \sim -8.3$ mag,[33] the TOM+1 magnitude will be reached for all bright galaxies at a distance of ~ 30 Mpc. However, in the H-band, the TOM+1 depth rule can be relaxed and a depth at 5σ the TOM could be sufficient in this band for reliable SBF measurement (Sect. 2.1.1), corresponding to a distance limit of ~ 70 Mpc.

Tests on the feasibility of SBF measurements with either Euclid instruments and with LSST precursor data are being carried out. If both instruments will be successful for SBF applications, we will have ~ 18.000 deg^2 of the LSST covering the

[c]Assuming an $i_{TOM} \sim -8.5$ mag,[32] we obtain a limiting distance modulus of $(m-M)_{limit} \sim 34.3$, or $D \sim 72.5$ Mpc.

[d]The wide passband of the optical instrument, VIS, basically covers the entire optical spectrum, thus we expect that the intrinsic scatter of SBF due to stellar population effects will be too large for precise distances from this instrument.

southern skies and $\sim 15.000\ deg^2$ of Euclid in the North, that will potentially allow obtaining the most accurate and complete 3-D map of massive red galaxies in the sky out to ~ 70 Mpc. Certainly a large number of lower mass, bluer galaxies at lower distances will also be mapped. The combination of these two data sets will provide exquisitely accurate cosmic flow models to be used for peculiar-velocity corrected flows, and an incredibly large sample of galaxies with SBF distances, useful for H_0 and for the needs of any near-field Cosmology study.

A further instrument of great impact will be the Roman telescope: with the same aperture of HST, similar PSF sharpness, ~ 100 times the HST field of view, near-IR coverage, and 5σ point-source depth in 1 hr of 28 mag in J and H band, this telescope will combine survey to depth capabilities, allowing to measure SBF for galaxies in clusters or groups at distances out to 100 Mpc and more.

3.2.2. *JWST and 40m ground-based telescopes*

The JWST and the 30-40m class telescopes –like the E-ELT– will have near-IR imaging capabilities at a resolution like HST or better, combined with substantially larger collecting areas, a field of view of the order of one to a few arcminutes, more modern and efficient detectors: all these characteristics combined make it plausible to predict that these telescopes will allow to overcome the 150 Mpc limit, currently "held" by SBF with the HST.[28, 29]

JWST is the natural heir of HST: with a similarly sharp (FWHM$\sim$ 0.1") and stable PSF, it has $\sim 7.5\times$ the HST aperture, possibly sufficient to enable us measuring SBF distances for galaxies at 300 Mpc limit or beyond. The limiting factors at such large distances will be, in this case, the contaminating GCs and obtaining reliable estimates for k-corrections.

The perspectives for ground-based 30-40m telescopes are at the same time exciting and challenging. The next-generation adaptive-optic (AO) modules at near-IR imagers will provide PSF reaching the limit of FWHM$\sim$0.01", thus one order of magnitude better than the best resolution presently available (even better than the JWST), combined with a collecting area at least 150 times larger than HST, and increased detector efficiency. Would this mean that the potential of the SBF with AO assisted imaging data is to reach distances even one order of magnitude farther than present limits, breaking the 1 Gpc limit (z$\sim$0.2)? To date, no answer can be given to the question, because of the lack of studies on the measurement of fluctuations from data taken with AO modules. Although some studies have been made of this topic,[34] quantitative demonstrations are lacking, because of the complexity of combining AO observations for extended objects with the issues of deriving reliable SBF amplitudes from AO supported imaging data. Let us take as example the complexities with the PSF: AO observations are characterized by a PSF variable in space and time, the lack of good PSF candidates within the galaxy field of view because of the small field of the present generation of AO systems, and by the unknown effect on SBF from the large spread of the PSF in diffraction-limited

conditions. In spite of such non-trivial issues, the forthcoming generation of 30-40m ground-based telescopes can be expected to produce breakthroughs in this field. Currently, cosmological studies based on direct distance indicators mostly rely on SNe Ia, which, although calibrated with high precision, are serendipitous events that present a zoo of peculiarities. Furthermore, SNe Ia distances rely on purely empirical information: we still lack a detailed modeling of the phenomena involved in SNe Ia explosion, useful for numerical calibrations of distances. On the other hand, the SBF measurement requires only data of suitable quality (no surreptitiousness), and relies on a signal whose origin is well known (star counts). In other words, though extremely more challenging than with the ideal JWST or Roman telescopes, SBF with 30-40m ground-based AO-assisted telescopes, could lead to SBF estimates out to redshift $z > 0.1$. At such distances, cosmological effects are predicted producing variations measurable with a distance indicator characterized by an accuracy of $\sim$5%. Hence, a dozen of well selected targets could be used to back up the SNe Ia on an independent route to precise cosmological parameters.

3.2.3. *Stellar population models and calibrations*

Possible refinements on the accuracy and precision of SBF distances will depend on the availability of new stellar population models and on improved $\overline{M}$calibrations.

At large distances, i.e. $d > 100$ Mpc, magnitudes (both "standard" and SBF magnitudes) and colors require k-corrections, which can only be derived from stellar population models. The most recent estimates for k-corrections on SBF magnitudes, date back to the 2000,[35] and predict a correction of ~ 0.1 mag for the HST/NICMOS $F160W$ filter ($\sim H$ band) at 200 Mpc. The correction changes significantly with wavelength (as an example it is significantly smaller in the J band, ~ 0.012 mag) and needs to be accounted for also in the colors, which are used for estimating $\bar{M}$.

Hence, any detailed study of cosmological parameters from SBF magnitudes for galaxies at distances where the peculiar velocities are negligible with respect to the Hubble-flow, cannot ignore using reliable k-corrections which, to date, are still poorly constrained or completely unavailable for some filters or distances.

Concerning the SBF calibration, there are here too opportunities for improvements. At present, the dominant systematic error comes from the $\sim$0.08 mag tie between SBF and Cepheids (Sec. 2.3). The results from Gaia will certainly have a positive impact on that: both because of the direct detection of Cepheids in M 31 (one of the SBF calibrators), and thanks to the more accurate and precise results on P-L relations from the data of thousands Galactic and Local Group Cepheids with measured parallaxes.[36]

A further opportunity could be deriving new SBF measurements for the same six calibrators used by Tonry & coIs. (based on imaging data taken prior to 1997) or from new calibrators.

Finally, as also pointed out in B21, the complete independence the Cepheid and the TRGB-based calibrations could also be used as a further factor to reduce the

systematic uncertainty on a joint Cepheid/TRGB calibration, leading to a combined zero-point uncertainty of $\lesssim 2$ % in distance (assuming the uncertainty on the TRGB zero-point TRGB is similar to that of Cepheids).

The upcoming launch of the James Webb Space Telescope (JWST) will allow for more numerous TRGB distances to the Virgo early-type galaxies in only modest observation times. Ultimately, we hope to anchor the SBF method using a much larger sample of giant ellipticals with TRGB distances calibrated using Gaia's parallaxes.

4. Conclusions

The SBF technique, with direct distances on single galaxies reaching an accuracy of $\sim 5\%$, is at present one of the most accurate extragalactic distance indicator in the distance regime between a few Mpc to ~ 150 Mpc.

In this contribution, I have described the basic characteristics of the method, how SBFs are measured, calibrated and used to estimate distances. I also provided a summary of the main scientific results obtained.

The next-generation of ground- and space-based telescopes will provide a variety of data that will generate thrilling results in many fields of Astronomy, not least the study of Cosmology based on direct distances from SBF. The extensive use of near-IR observations with high-resolution cameras –from JWST or E-ELT–, the deep surveys that will cover a large fraction of the sky –from the Vera Rubin Observatory or the Euclid satellite–, the improvements on SBF calibrations –from Gaia and from new stellar population models– will push forward this technique. By combining accuracy with depths and sky coverage, new data will allow using SBF distances for constraining cosmological models using a new path, independent from other direct distance methods, like SNe Ia.

References

1. W. Baade, The Resolution of Messier 32, NGC 205, and the Central Region of the Andromeda Nebula., *The Astroph. Journ.* **100**, p. 137 (September 1944).
2. B. J. Bok, *The Distribution of the Stars in Space* 1937.
3. J. Tonry and D. P. Schneider, A new technique for measuring extragalactic distances, *The Astroph. Journ.* **96**, 807 (September 1988).
4. J. P. Blakeslee, A. Jordán, S. Mei, P. Côté, L. Ferrarese, L. Infante, E. W. Peng, J. L. Tonry and M. J. West, The ACS Fornax Cluster Survey. V. Measurement and Recalibration of SBF and a Precise Value of the Fornax-Virgo Relative Distance, *The Astroph. Journ.* **694**, 556 (March 2009).
5. J. L. Tonry, E. A. Ajhar and G. A. Luppino, Observations of SBF in Virgo, *The Astroph. Journ.* **100**, p. 1416 (November 1990).
6. J. B. Jensen, J. L. Tonry and G. A. Luppino, Measuring Distances Using Infrared SBF, *The Astroph. Journ.* **505**, 111 (September 1998).
7. J. P. Blakeslee, A. Vazdekis and E. A. Ajhar, Stellar populations and SBF: New observations and models, *Mon. Not. Roy. Ast. Soc.* **320**, p. 193 (January 2001).

8. M. Cantiello, J. P. Blakeslee, G. Raimondo, S. Mei, E. Brocato and M. Capaccioli, Detection of Radial SBF and Color Gradients in Elliptical Galaxies with the Advanced Camera for Surveys, *The Astroph. Journ.* **634**, p. 239 (November 2005).
9. J. P. Blakeslee, M. Cantiello, S. Mei, P. Côté, R. Barber DeGraaff, L. Ferrarese, A. Jordán, E. W. Peng, J. L. Tonry and G. Worthey, SBF in the Hubble Space Telescope ACS/WFC F814W Bandpass and an Update on Galaxy Distances, *The Astroph. Journ.* **724**, 657 (November 2010).
10. J. B. Jensen, J. P. Blakeslee, Z. Gibson, H.-c. Lee, M. Cantiello, G. Raimondo, N. Boyer and H. Cho, Measuring Infrared SBF Distances with HST WFC3: Calibration and Advice, *The Astroph. Journ.* **808**, p. 91 (July 2015).
11. M. Cantiello, J. B. Jensen, J. P. Blakeslee, E. Berger, A. J. Levan, N. R. Tanvir, G. Raimondo, E. Brocato, K. D. Alexander, P. K. Blanchard, M. Branchesi, Z. Cano, R. Chornock, S. Covino, P. S. Cowperthwaite, P. D'Avanzo, T. Eftekhari, W. Fong, A. S. Fruchter, A. Grado, J. Hjorth, D. E. Holz, J. D. Lyman, I. Mandel, R. Margutti, M. Nicholl, V. A. Villar and P. K. G. Williams, A Precise Distance to the Host Galaxy of the Binary Neutron Star Merger GW170817 Using SBF, *The Astroph. Journ. Letters* **854**, p. L31 (February 2018).
12. S. G. Carlsten, R. L. Beaton, J. P. Greco and J. E. Greene, Using SBF to Study Nearby Satellite Galaxy Systems: Calibration and Methodology, *The Astroph. Journ.* **879**, p. 13 (July 2019).
13. M. Cantiello, I. Biscardi, E. Brocato and G. Raimondo, VLT optical BVR observations of two bright supernova Ia hosts in the Virgo cluster. SBF analysis, *Astron. & Astroph.* **532**, p. A154 (August 2011).
14. J. L. Tonry, A. Dressler, J. P. Blakeslee, E. A. Ajhar, A. B. Fletcher, G. A. Luppino, M. R. Metzger and C. B. Moore, The SBF Survey of Galaxy Distances. IV. SBF Magnitudes, Colors, and Distances, *The Astroph. Journ.* **546**, p. 681 (January 2001).
15. J. B. Jensen, J. L. Tonry, B. J. Barris, R. I. Thompson, M. C. Liu, M. J. Rieke, E. A. Ajhar and J. P. Blakeslee, Measuring Distances and Probing the Unresolved Stellar Populations of Galaxies Using Infrared SBF, *The Astroph. Journ.* **583**, 712 (February 2003).
16. M. Cantiello, G. Raimondo, E. Brocato and M. Capaccioli, New Optical and Near-Infrared SBF Models: A Primary Distance Indicator Ranging from Globular Clusters to Distant Galaxies?, *The Astroph. Journ.* **125**, p. 2783 (June 2003).
17. G. Raimondo, E. Brocato, M. Cantiello and M. Capaccioli, New Optical and Near-Infrared SBF Models. II. Young and Intermediate-Age Stellar Populations, *The Astronomical Journal* **130**, 2625 (December 2005).
18. J. B. Jensen, J. P. Blakeslee, C.-P. Ma, P. A. Milne, P. J. Brown, M. Cantiello, P. M. Garnavich, J. E. Greene, J. R. Lucey, A. Phan, R. B. Tully and C. M. Wood, Infrared SBF Distances for MASSIVE and Type Ia Supernova Host Galaxies, *arXiv e-prints*, p. arXiv:2105.08299 (May 2021).
19. J. P. Blakeslee, J. B. Jensen, C.-P. Ma, P. A. Milne and J. E. Greene, The hubble constant from infrared sbf distances, *The Astroph. Journ.* **911**, p. 65 (apr 2021).
20. J. L. Tonry, J. P. Blakeslee, E. A. Ajhar and A. Dressler, The SBF Survey of Galaxy Distances. II. Local and Large-Scale Flows, *The Astroph. Journ.* **530**, 625 (February 2000).
21. W. L. Freedman and B. F. Madore, The Hubble Constant, *Ann. Rev. Astron. & Astroph.* **48**, 673 (September 2010).
22. G. Pietrzyński, D. Graczyk, A. Gallenne, W. Gieren, I. B. Thompson, B. Pilecki, P. Karczmarek, M. Górski, K. Suchomska, M. Taormina, B. Zgirski, P. Wielgórski,

Z. Kołaczkowski, P. Konorski, S. Villanova, N. Nardetto, P. Kervella, F. Bresolin, R. P. Kudritzki, J. Storm, R. Smolec and W. Narloch, A distance to the Large Magellanic Cloud that is precise to one per cent, *Nature* **567**, 200 (March 2019).

23. A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM, *The Astroph. Journ.* **876**, p. 85 (May 2019).
24. J. Soltis, S. Casertano and A. G. Riess, The Parallax of ω Centauri Measured from Gaia EDR3 and a Direct, Geometric Calibration of the Tip of the Red Giant Branch and the Hubble Constant, *The Astroph. Journ. Letters* **908**, p. L5 (February 2021).
25. J. Mould and S. Sakai, The Extragalactic Distance Scale Without Cepheids. II. SBF, *The Astroph. Journ.* **694**, 1331 (April 2009).
26. P. van Dokkum, S. Danieli, Y. Cohen, A. J. Romanowsky and C. Conroy, The Distance of the Dark Matter Deficient Galaxy NGC 1052-DF2, *The Astroph. Journ. Letters* **864**, p. L18 (September 2018).
27. E. A. Ajhar, J. L. Tonry, J. P. Blakeslee, A. G. Riess and B. P. Schmidt, Reconciliation of the SBF and Type Ia Supernova Distance Scales, *The Astroph. Journ.* **559**, 584 (October 2001).
28. J. B. Jensen, J. L. Tonry, R. I. Thompson, E. A. Ajhar, T. R. Lauer, M. J. Rieke, M. Postman and M. C. Liu, The Infrared SBF Hubble Constant, *The Astroph. Journ.* **550**, 503 (April 2001).
29. I. Biscardi, G. Raimondo, M. Cantiello and E. Brocato, Optical SBF of Shell Galaxies toward 100 Mpc, *The Astroph. Journ.* **678**, 168 (May 2008).
30. N. Khetan, L. Izzo, M. Branchesi, R. Wojtak, M. Cantiello, C. Murugeshan, A. Agnello, E. Cappellaro, M. Della Valle, C. Gall, J. Hjorth, S. Benetti, E. Brocato, J. Burke, D. Hiramatsu, D. A. Howell, L. Tomasella and S. Valenti, A new measurement of the Hubble constant using Type Ia supernovae calibrated with SBF, *Astron. & Astroph.* **647**, p. A72 (March 2021).
31. Planck Collaboration, Planck 2018 results. VI. Cosmological parameters, *Astron. & Astroph.* **641**, p. A6 (September 2020).
32. W. E. Harris, Globular cluster systems, in *Saas-Fee Advanced Course 28: Star Clusters*, 2001.
33. J. B. Nantais, J. P. Huchra, P. Barmby, K. A. G. Olsen and T. H. Jarrett, Nearby Spiral Globular Cluster Systems. I. Luminosity Functions, *The Astronomical Journal* **131**, 1416 (March 2006).
34. D. Gouliermis, W. Brandner, D. Butler and S. Hippler, SBF: A Case for Extremely Large Telescopes, in *Science with Adaptive Optics*, eds. W. Brandner and M. E. KasperJanuary 2005.
35. M. C. Liu, S. Charlot and J. R. Graham, Theoretical Predictions for SBF and Implications for Stellar Populations of Elliptical Galaxies, *The Astroph. Journ.* **543**, 644 (November 2000).
36. G. Clementini, V. Ripepi, R. Molinaro, A. Garofalo, T. Muraveva, L. Rimoldini, L. P. Guy, G. Jevardat de Fombelle, K. Nienartowicz, O. Marchal, M. Audard, B. Holl, S. Leccia, M. Marconi, I. Musella, N. Mowlavi, I. Lecoeur-Taibi, L. Eyer, J. De Ridder, S. Regibo, L. M. Sarro, L. Szabados, D. W. Evans and M. Riello, Gaia Data Release 2. Specific characterisation and validation of all-sky Cepheids and RR Lyrae stars, *Astron. & Astroph.* **622**, p. A60 (February 2019).

Preliminary results of analysis of Ia supernovae redshift distributions on data of the Asiago Supernova and Open Supernova Catalogues

I.V. Arkhangelskaja

Division of Nuclear Physics and Technologies, National Research Nuclear University MEPhI, Kashirskoe shosse, 31, Moscow, 115409, Russia
E-mail: IVArkhangelskaya@mephi.ru
https://home.mephi.ru/en/users/1525/public

The shape of redshift distribution for uniform sources set in our Metagalaxy defined by cosmological parameters and properties of space. The suggestion that type Ia supernovae might be used as standard candles for cosmological measurements allow considered these objects as homogeneous subsample. This mention occurs since the earliest studies of supernovae in 1938. Firstly the parameters of our Metagalaxy Ω and Λ were determine due sample of Ia supernovae from the Supernova Cosmology Project analysis in 1998. It was found due SN1a characteristics analysis that space in our Metagalaxy is Euclidean at small redshifts and de-Sitter at high ones. Now several tens of thousands supernovae' characteristics analyzed in various catalogues.

The preliminary results of the redshift distribution analysis for SNIa from the Asiago Supernova and Open Supernova Catalogues are discussed in this work. Firstly it was shown that several peculiarities are presented in Ia supernovae redshift distribution. The deviation in the band $0.015 < z < 0.13$ accordingly Open Supernova Catalogue (OSC) data contain more faint supernovae. Two peculiarities also were found in the region $0.25 < z < 0.45$ on data of this catalogue. One of it's contain more faint events, other contain more bright supernovae.

Also faint_m and bright_m supernovae areas could be separated by $m_{fb} \sim 20$ in distribution of object amount on apparent magnitude on both catalogues data. Furthermore, the distribution of SNIa on apparent magnitudes and angular distance to host centre on OSC data also reveal two areas (faint_{dist} and bright_{dist} objects) and ratio between these regions populations is different for different subsamples in OSC (Pan-STARRS1 (PS1) Medium Deep Survey, ASASSN and so on).

Moreover, SNIa apparent magnitudes could differ for $\Delta m_{onehost} \sim 2.5$ within one host galaxy (for example, in UGC03432) and this difference bigger that systematic uncertainties of SN1a characteristics observations. Also there are no specific absorption lines in the energy spectra of these SNIa which exclude influence of surrounding media.

The separated peculiarities and areas could not be explained due 2 groups of type Ia SNe explosions scenarios and dimming of flux due interaction of surrounding media. But it could affect at power in dependence of distance modulus μ on redshift, and, correspondingly, the H_0 tension between various subsamples of SN1a data. May be such tension was caused by different ratio of two SNIa classes objects in analysed catalogues and value of H_0 defined from supernovae subsets analysis became lower after correction for results of dependence of apparent magnitude on distance to host centre.

Thus such peculiarities presence could be caused by several unknown aspects of SNIa explosions scenarios or really changing of the parameters of our Metagalaxy. Further conclusions required subsequent OSC database treatment in combination with high redshift datasets, for example, addition of Dark Energy Survey Supernova Program catalogue into data analysis.

Keywords: SNIa redshift distribution

1. Introduction

The homogeneous subsample characteristics understanding is necessary for the investigation of any astrophysical objects redshift distribution, for example, gamma-ray bursts. The shape of redshift distribution for uniform sources set in our Metagalaxy defined by cosmological parameters and properties of spacetime. Type Ia supernovae considered as such standatd candels since 1995.[1] The effect of a cosmological constant leads to dimming of observed brightnesses of highredshift supernovae than would have been the case with $\Omega_\Lambda = 0$ and allowed to estimate cosmological parameters due analysis of Hubble diagram contain dependence of on redshift — see, for example, Ref. 2.

The interest of the scientific community on supernovae (SNe) has increased in the recent years for several reasons. Mostly they are the advances in the understanding of the SN phenomena obtained with the intensive study of nearby SNe, first of all SN 1987A, which have raised new more fundamental questions with regard to progenitor evolution, explosion mechanism and nucleosynthesis — see, for instance, Refs. 3, 4. In addition, the at least two facts have renewed the interest for the using of SNIa as distance indicators up to cosmological distances. The first one was the calibration of the absolute magnitudes of a few SNIa obtained using the Cepheid variables found in their parent galaxies.[5] The second fact was the discovery of empirical relations between the absolute magnitudes at maximum and the shape of the light curves of SN Is.[6] Other exciting advances are expected for the association of some SNe with the other interest objects, for example, GRBs. Another reason is the necessaries of homogeneous subsample characteristics understanding for the investigation of any astrophysical objects redshift distribution, for instance, also GRB — see Ref. 7. Such wide interest has activate new deep SNe searches and creation or revision of catalogues based of its results. Now the amount of SNe listed in various catalogues is about one hundred thousands which mentioned in several tens of catalogues — see, for instance, Refs. 8–11. SN1a are approximately 25% of this subset volume — see, for instance, Refs. 9, 10, 12.

Firstly the parameters of our Metagalaxy Ω and Λ were determine due sample of Ia supernovae from the Supernova Cosmology Project analysis in 1997 — see Refs. 2, 13. Figure 1 represents the results of tis investigation. The combination of Hubble diagrams for 42 high-redshift type Ia supernovae from the Supernova Cosmology Project and and 18 low-redshift ones from the Supernova Calalan/Tololo Survey was approximated by power law fit. It was found that this fit corresponds to following composition of our Metagalaxy: 70% of dark energy, 25% of dark matter and only 5% of ordinary baryonic matter.[2,13] Later it was found that several objects from this sample are not SNIa, but a valid asymptotic estimations still remain.[14]

2. Asiago Supernova Catalogue

The history of the Asiago SN catalogue began in 1984 with the publication of data for 568 objects.[15] This was compiled starting from the Palomar Supernova Master

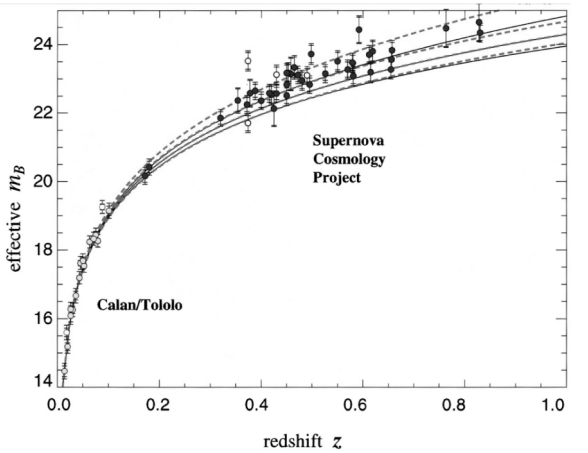


Fig. 1. Hubble diagram for 42 high-redshift type Ia supernovae from the Supernova Cosmology Project and 18 low-redshift type Ia supernovae from the Supernova Calalan/Tololo Survey. Adopted from Ref. 13.

List which since 1958 from time to time appeared in the literature see, for example, Refs. 8, 9, 16. During the same period two other SN listing have been published,[17,18] the Ref. 18 also contains the complete bibliography for each object. The 1984 Asiago SN Catalogue was superseded by a new edition in 1989[19] which listed information for the 661 supernovae discovered up to December 31, 1988.

More recently, van den Bergh at Ref. 20 published a list containing the 203 supernovae discovered between January 1, 1989 and April 3, 1994. Also the Catalogue of extragalactic Supernovae, complete up to 1993, was published in volume V of the General Catalogue of Variable Stars Ref. 21. In the last few years the Asiago Supernova Catalogue was widely used through the literature and available at the Ref. 22. Other supernova listings are available electronically, e.g. the list at the CBAT,[23] and Sternberg Astronomical Institute Supernova Catalogue.[24] The Asiago Supernova Catalogue consists of following datasets contain more than 100 supernovae:

(1) SDSS-II,
(2) LOSS,
(3) LOTOSS,
(4) ESSENCE,
(5) Puckett,
(6) SCP,
(7) HZSST.

Other 27 supernovae datasets included in this catalogue consist of less amount of members. For example, ones BAOSS and CFHT-LSSP contain only 34 and 13 objects correspondingly.

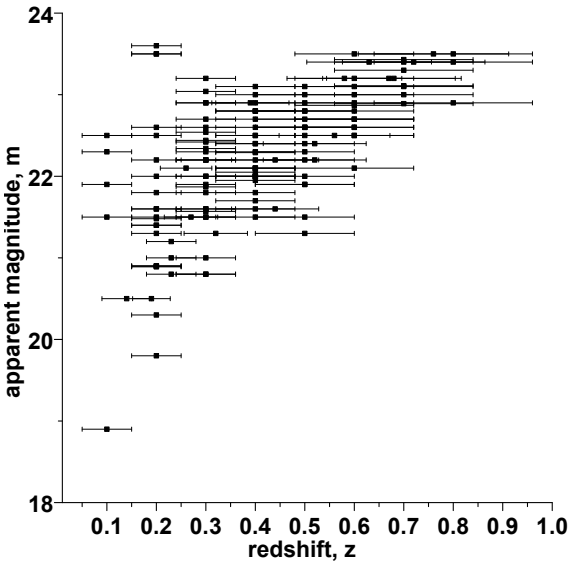


Fig. 2. Plot of SN Ia distribution on apparent magnitude dependence and redshift for ESSENCE experiment data accordingly to Asiago Supernova Catalogue.

Unfortunately the redshift data of several experiments (for example, ESSENCE – see Figure 2) contain big errors. This effect leads to several "lines" looks to areas with z=const occurs in plots of distributions on redshift. Such artefacts makes difficult to process these data and such datasets were excluded from analysis.

The Asiago Supernova Catalogue was last updated on 2 October 2017 and consist of data about 6530 supernovae.[22] The subsample of SN1a contains 3135 objects but redshifts with sufficient for analysis accuracy are known only for 2336 ones.

3. Open Supernova Catalogue

The Open Supernova Catalogue[11,12] is online collection of observations and metadata for presently 92632 supernovae and related candidates. The catalog is freely available on the web.[11] Its main interface having been designed to be a searchable table accessible on desktop and mobile devices due GitHub.[27] It represents an ongoing community driven project to collect and clean supernova metadata and observations spanning X-ray, ultraviolet, optical, infrared, and radio frequencies. At present, ∼14000 events in the catalog include light curves with at least 5 photometric points, and in total the catalogue contains ∼678000 individual photometric detections. Within the catalogue there are ∼4250 supernovae that include spectra, with ∼21000 spectra in total. In addition to the primary catalogue table containing supernova metadata, an individual page is generated for each supernova which displays its available metadata, light curves, and spectra from X-ray to radio frequencies. The information presented in the catalog is automatically rebuilt

on a daily basis and is constructed due parsing several sources, including the data presented in the supernova literature and from secondary sources such as other web-based catalogs — see, for example, Refs. 28–30. The objects included in the catalogue are intended to be entirely supernovae, i.e. the complete destruction of a star by an explosive event that may or may not leave behind a compact remnant, and objects that have been definitively identified as other transient types were actively removed. One difference between Open Supernova Catalogue and some other supernova catalogues is that authors augment the known supernovae with known supernova remnants,[31] which are thought to be supernovae but (currently) with no known associated transient. Absolute magnitudes calculated in the Open Supernova Catalogue take into account luminosity distance and redshift decrements. The Open Supernova Catalogue consists of following lists contain more than 1000 supernovae:

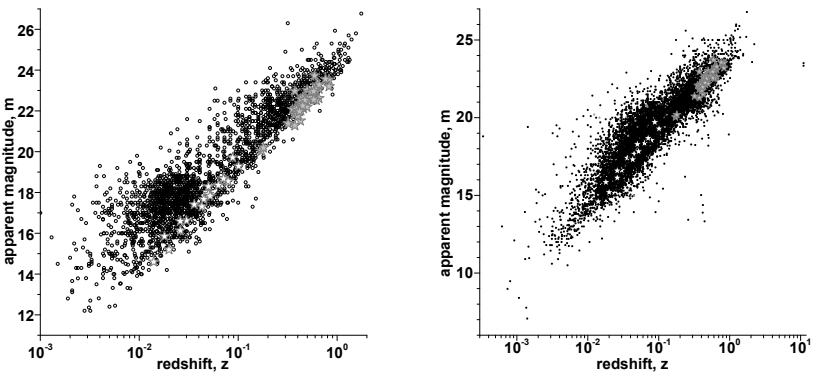
(1) The Open Supernova Catalogue,
(2) Sloan Digital Sky Survey,
(3) Latest Supernovae,
(4) JLA supernovae,
(5) SIMBAD astronomical database,
(6) Transient Name Server,
(7) THE SLOAN DIGITAL SKY SURVEY-II SUPERNOVA SURVEY,
(8) Pan-STARRS 3Pi,
(9) Asiago Supernova Catalogue,
(10) A unified supernova catalogue,
(11) NED-D,
(12) WISeREP—An Interactive Supernova Data Repository,
(13) Catalina Sky Survey,
(14) NASA/IPAC Extragalactic Database,
(15) Sternberg Astronomical Institute Supernova Light Curve Catalogue.

Other 828 supernovae lists included in this catalogue consist of 5 – 919 objects. For example, ones UCB Filippenko Group's Supernova Database (SNDB) and DES Bright Transients contain only 134 and 17 objects correspondingly. Also information about characteristics of individual supernova included in this catalogue due more than 130000 circulars, such as CBET, IAUC, ATel, etc. The Open Supernova Catalogue updated up to now[11] and its subsample of SNIa contains 16428 objects but redshifts are known only for 14607 ones.

4. The preliminary results of data analysis

This section presents the preliminary results of analysis of subsets of Ia type supernovae with known redshifts from the Asiago Supernova (ASC) and the Open Supernova (OSC) catalogues.

The distribution of SNIa on redshift and apparent magnitudes on data of these catalogues are shown on Fig. 3. Both distributions are more wide than mentioned on Fig. 1 and contain several peculiarities. Moreover, the sample OSC has deviation in the band $0.015 < z < 0.13$ contain more faint supernovae and two peculiarities in the region $0.25 < z < 0.45$. One of its contain more dim events, other contain more bright supernovae. But both analysable subsets contain data from several catalogues and at first we should exclude selection effects. These characteristic features appears at distribution on absolute magnitude and redshift too — see Fig. 4.



(a) (b)

Fig. 3. The distribution of SNIa amount on redshift and apparent magnitudes on data of following catalogues: (a) Asiago Supernova and (b) Open Supernova. Gray stars show Supernova Cosmology Project data, open ones represent subsample from the Supernova Calalan/Tololo Survey mentioned on Fig. 1.

The uncertainties in recent SN Ia cosmology analyses usually excluded due following methods[32]: probing relations between luminosity and properties of the host galaxies of the SNe (see Refs. 33, 34) and analyses of the light-curve parameters of SNe and how these parameters relate to luminosity.[35,36] Many of the associated systematic uncertainties of these effects are on the 1% level, and considering a typical SN distance modulus is measured with roughly 15% precision.[32] Also K-corrections should be used to provide a transformation between an observed-frame magnitude and a rest-frame one and account for the SN redshift (see, for example, Refs. 37–40), but object magnitude was maximally shifted for value less then 1.2 and no systematic effects were found for such corrections. Thus all uncertainties could not influence to shapes of observed distributions.

Several peculiarities in the plots of the distributions of SNIa on redshift and apparent magnitudes for HZSST and SCP experiments within Asiago Supernova Catalogue are are presented on Fig. 5. Also some peculiarities in the plots of the distributions of SNIa on redshift and apparent magnitudes for Sternberg Astronomical

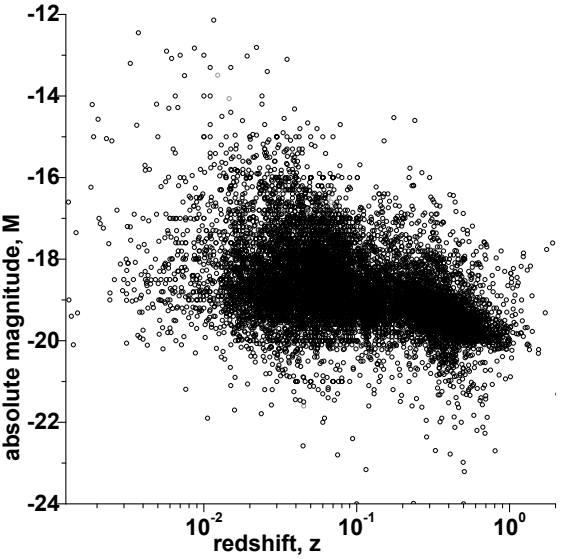


Fig. 4. The distribution of SNIa on absolute magnitudes and redshift on data of Open Supernova catalogue.

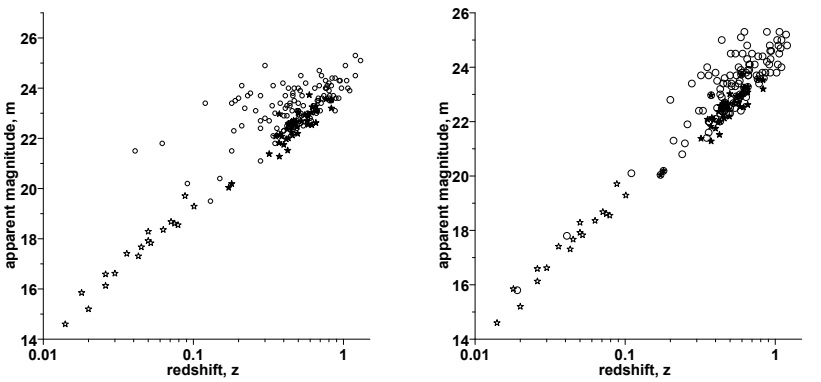
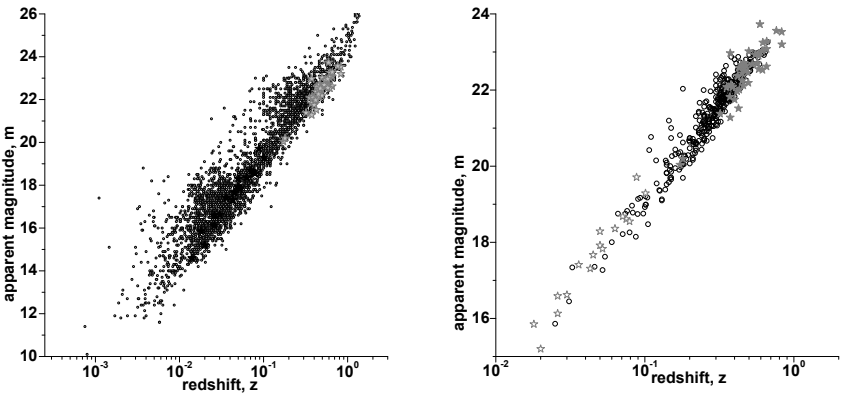


Fig. 5. Several peculiarities in the plots of SNIa apparent magnitude dependence on redshift for HZSST (a) and SCP (b) experiments. Black stars show Supernova Cosmology Project data, open ones represent subsample from the Supernova Calalan/Tololo Survey. Black stars show Supernova Cosmology Project data, open ones represent subsample from the Supernova Calalan/Tololo Survey mentioned on Fig. 1.

Institute Supernova Light Curve Catalogue[24] and Pan-STARRS1 (PS1) Medium Deep Survey[29,32] within Open Supernova Catalogue are are presented on Fig. 6. The discussed in the previous paragraph uncertainties could not caused such features too.

The distributions of SNIa amount on redshift on data of the Asiago Supernova and the Open Supernova catalogues are shown on Fig. 7. It is seen feature at $z > 0.2$ on Open Supernova dataset possible connected with peculiarities in the region $0.25 < z < 0.45$ on Fig. 3. Than we consider the distributions of these catalogues data on apparent magnitude — see Fig. 8. Preliminary analysis results have shown the presence of two subsets (faint_m and bright_m supernovae) at each distributions separated by $m_{fb} \sim 20$. These dual structure appear in the same distributions of catalogues composing OSC — see, for example, Fig. 9.

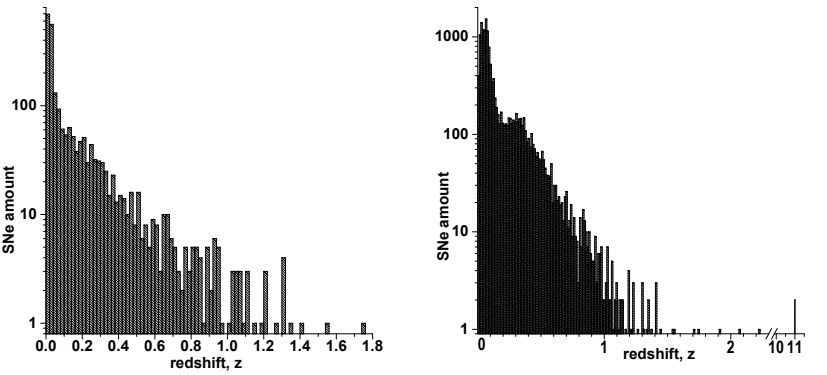


(a) (b)

Fig. 6. Several peculiarities in the plots of the distributions of SNIa on redshift and apparent magnitudes for: (a) Sternberg Astronomical Institute Supernova Light Curve Catalogue, (b) Pan-STARRS1 (PS1) Medium Deep Survey within Open Supernova Catalogue. Stars show data, mentioned on Fig. 1. Grey ones show Supernova Cosmology Project data, open and black stars represent subsample from the Supernova Calalan/Tololo Survey at panels (a) and (b) correspondingly.

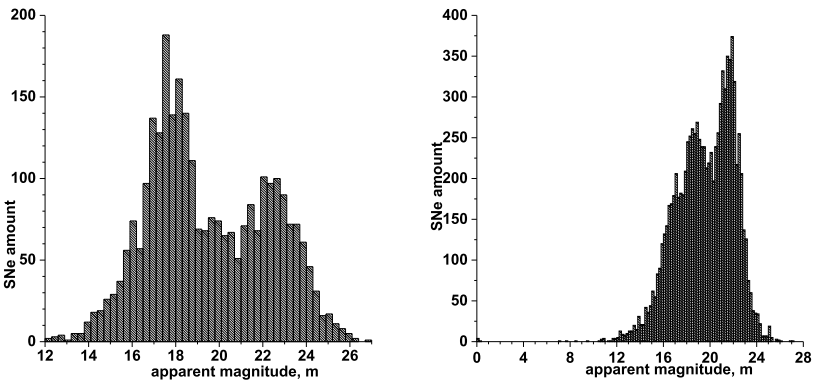
The distribution of SNIa on apparent magnitudes and angular distance to host centre for OSC objects is presented at Fig. 10. Also two areas (faint_{dist} and bright_{dist} SNIa) separated at this distribution. Preliminary analysis results allow conclude several differences between these regions and peculiarities on distributions of SNIa amount on apparent magnitude and redshift also with one on redshift. Moreover, these area not fully correspond to subsets on the distributions of apparent magnitudes from Fig. 8b and Fig. 9.

Let's consider two SNIa at $z_1 = 0.0039$ and $z_2 = 0.0037$: 1998aq and 2004W. The difference in the apparent magnitudes of these objects is ~ 6: $m1_{max} = 12.5$ (bright SNIa at Fig. 10) and $m2_{max} = 18.5$ (faint SNIa at Fig. 10) correspondingly accordingly Ref. 11. Also the same differences observed at various redshifts — see Table 1. Moreover, some hosts contain several supernovae but its apparent magnitudes could differ for $\Delta m_{onehost} \sim 2.5$. For example, galaxy UGC03432 with



(a) (b)

Fig. 7. The distribution of SNIa amount on redshift on data of following catalogues: (a) Asiago Supernova and (b) Open Supernova.

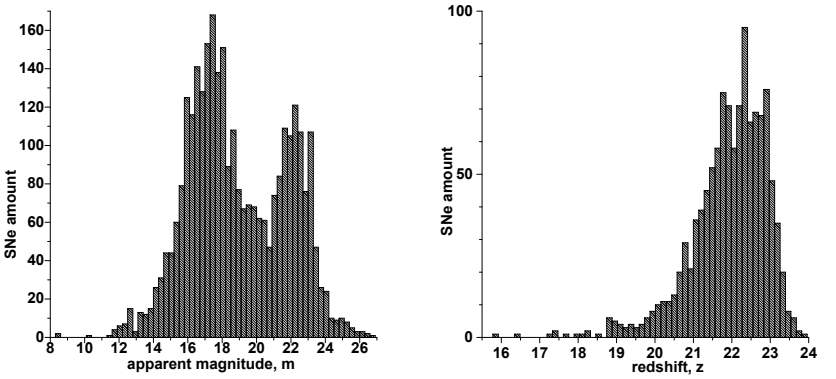


(a) (b)

Fig. 8. The distribution of SNIa amount on apparent magnitudes on data of following catalogues: (a) Asiago Supernova and (b) Open Supernova.

$m_{gal} = 15.51$ at z=0.016667 contain two SNIa 1996bv has $m_1 = 15.5$ and 2003kb has $m_2 = 18$. But usually supernovae has the same characteristics within one host. For instance, two SNIa 2004br and 1994M had m=16.3 within galaxy NGC4493 with $m_{gal} = 14.951$ at z=0.02315.

Up to now there are no evidence for an astrophysical origin of the apparent faintness of SNe Ia has been found (see Refs. 41, 42). Early only SFR (star-formation rate) of the parent galaxies was considered as influence factor for diminution of SN apparent magnitudes. Thus, dependence of distant SN Ia light-curve shapes on star-formation in the host galaxy was studied in Ref. 34, and it was found that passive galaxies without star-formation, preferentially host dimmer SNe Ia with faster-declining light-curves, while brighter events with slower-declining ones only



(a) (b)

Fig. 9. The distribution of SNIa amount on apparent magnitudes on data of: (a) Sternberg Astronomical Institute Supernova Light Curve Catalogue, (b)Pan-STARRS1 (PS1) Medium Deep Survey within Open Supernova Catalogue.

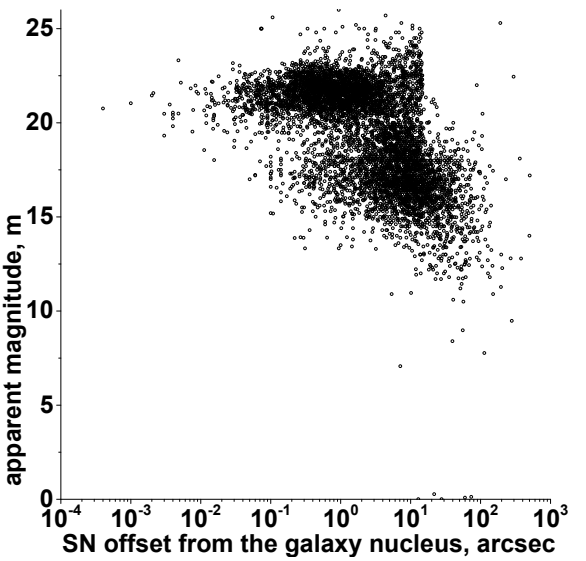


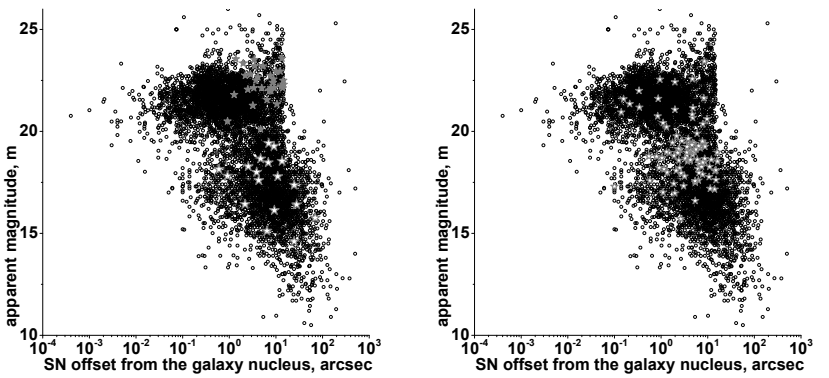
Fig. 10. The distribution of SNIa on apparent magnitudes and SN offset from the galaxy nucleus on data of Open Supernova catalogue.

originated from systems with ongoing star-formation. The mean stretch smaller for 0.08 against the typical precision on the stretch measure is $\pm 0.01 - 0.02$ at 95% confidence, which change distance modulus μ less than for 1% and could not explain the observed difference in the apparent magnitudes of SNIa $\Delta m_{onehost}$. All mentioned objects were observed by LEDA and there are no any evidences

to suppose additional matter presence between observer and SN##1–4 because additional absorption lines are absent in the spectra of these supernovae — see, for example, Refs. 25, 26.

Table 1. The examples of difference in the apparent magnitudes of SNIa at various redshifts.

#	SN	redshift z	m_{max}	Δm_{max}	m_{host}
1	2004W	0.0037	18.5	~ 6	9.8
2	1993c	0.012	18.0	~ 4.4	13.3
3	2007if	0.04	20.5	~ 4.4	16.3
4	1999G	0.10	21.4	~ 3.5	18.4
5	1998aq	0.0039	12.5	~ 6	10.9
6	1999cw	0.013	14.3	~ 4.4	14.2
7	2009do	0.04	16.1	~ 4.4	16.1
8	2012X	0.10	17.9	~ 3.5	17.7



(a) (b)

Fig. 11. The illustration of different ratio of two classes of SNIa on the distribution of SNIa on apparent magnitudes and offset from the galaxy nucleus on data of Open Supernova catalogues (open circles) in: (a) subsamples of Supernova Cosmology Project (grey stars) and Supernova Calalan/Tololo Survey (open stars) mentioned on Fig. 1 and (b) Pan-STARRS1 (PS1) Medium Deep Survey data (open stars).

Thus we have found several areas on distributions of SNIa from OSC on apparent magnitude and apparent magnitude vs offset from the galaxy nucleus. However the type Ia supernovae models not explain such subtypes. These models can be divided into two broad categories (Single Degenerate and Double Degenerate scenarios of Type Ia SNe explosions), but the authors of these scenarios concluded that both scenarios are able to correctly predict the bulk properties of Type Ia observations.[43–45] One type of progenitors are a $1.4M_{\odot}$ white dwarves that are likely created in an accretion process and self-ignite due to high central pressure/density. Other ones are white dwarves too but of considerably lesser mass ($1M_{\odot}$) can be ignited in shell

detonations or in the merging process that formed them. Both scenarios are able to correctly predict the bulk properties of Type Ia observations and widely discussed difficulties in recognition of models of concrete SNIa.

Such areas presence in various ratio in subsamples contain OSC — see Fig. 11 and could affect at power in dependence of distance modulus μ on redshift, and, correspondingly, the H_0 tension in various subsamples of SN1a data. The H_0 tension[46–48] occurs due comparison of estimation of the current expansion rate of the Netagalaxy H_0 due CMB data[50] and results of observations of astrophysical objects such as SNIa and so on — see, for example, Ref. 51. Now this tension is more than 5σ: accordingly to CMB measurements[50] a value for $H_0 = 67.36 \pm 0.54$ km s^{-1} Mpc^{-1}, but the best-fits of the SNIa data[51,52] gave the values of $H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1} and $H_0 = 75.4 \pm 1.7$ km s^{-1} Mpc^{-1} correspondingly. May be such discrepancy were caused by different ratio of two SNIa classes objects in analysed catalogues and value of H_0 defined from supernovae subsets analysis became lower.

5. Conclusions

The preliminary results of the redshift distribution analysis for SNIa from the Asiago Supernova and Open Supernova Catalogues are discussed in this work. The Asiago Supernova Catalogue (ASC) was last updated on 2 October 2017 and consist of data about 6530 supernovae.[22] The subsample of SN1a contains 3135 objects but redshifts with sufficient for analysis accuracy are known only for 2336 ones. The Open Supernova Catalogue (OSC) updated up to now[11] and its subsample of SNIa contains 16428 objects but redshifts are known only for 14607 ones. It was concluded that not all experiments from the catalogue could analysed (small statistics or big errors in the determination of z caused difficulties in data representation and analysis), for example, ESSENCE experiment data within Asiago Supernova Catalogue.[22]

The shape of redshift distribution for uniform sources set in our Metagalaxy defined by cosmological parameters and properties of space. The suggestion that type Ia supernovae might be used as standard candles[1] for cosmological measurements allow considered these objects as homogeneous subsample. Firstly the parameters of our Metagalaxy Ω and Λ were determine due sample of Ia supernovae from the Supernova Cosmology Project analysis in 1998.[2] It was found due SN1a characteristics analysis that space in our Metagalaxy is Euclidean at small redshifts and de-Sitter at high ones. Now several tens of thousands supernovae' characteristics analyzed in various catalogues.

The various redshifts distributions of SNIa from ASC and OSC have analysed. Firstly it was shown that several peculiarities are presented in Ia supernovae redshift distribution. The deviation in the band $0.015 < z < 0.13$ accordingly Open Supernova Catalogue (OSC) data contain more faint supernovae. Two peculiarities also were found in the region $0.25 < z < 0.45$ on data of this catalogue. One of it's contain more faint events, other contain more bright supernovae.

Also faint$_m$ and bright$_m$ supernovae areas could be separated by $m_{fb} \sim 20$ in distribution of object amount on apparent magnitude on both catalogues data. Furthermore, the distribution of SNIa on apparent magnitudes and angular distance to host centre on OSC data also reveal two areas (faint$_{dist}$ and bright$_{dist}$ objects) and ratio between these regions populations is different for different subsamples in OSC (Pan-STARRS1 (PS1) Medium Deep Survey, ASASSN and so on).

Moreover, SNIa apparent magnitudes could differ for $\Delta m_{onehost} \sim 2.5$ within one host galaxy (for example, in UGC03432) and this difference bigger that systematic uncertainties of SN1a characteristics observations. Also there are no specific absorption lines in the energy spectra of these SNIa which exclude influence of surrounding media.

The separated peculiarities and areas could not be explained due 2 groups of type Ia SNe explosions scenarios (see, for example, Refs. 43–45) and dimming of flux due interaction of surrounding media. But it could affect at power in dependence of distance modulus μ on redshift, and, correspondingly, the H_0 tension between various subsamples of SN1a data. May be such tension was caused by different ratio of two SNIa classes objects in analysed catalogues and value of H_0 defined from supernovae subsets analysis became lower after correction for results of dependence of apparent magnitude on distance to host centre

Thus such peculiarities presence could be caused by several unknown aspects of SNIa explosions scenarios or really changing of the parameters of our Metagalaxy. Further conclusions required subsequent OSC database treatment in combination with high redshift datasets, for example, addition of Dark Energy Survey Supernova Program catalogue[53] into data analysis.

Acknowledgments

Author thank for the support from National Research Nuclear University MEPhI in the framework of the Russian Academic Excellence Project (contract No. 02.a03.21.0005, 27.08.2013).

References

1. A. Goobar and S. Perlmutter, Feasibility of Measuring of the Cosmological Constant Λ and Mass Density Ω Using Ype Ia Supernovae, *The Astrophisical journal* **450** 14 (1995).
2. S. Perlmutter, G. Aldering, S. Deustua, S. Fabbro1, G. Goldhaber, D. E. Groom, A. G. Kim, M. Y. Kim, R.A. Knop, P. Nugent, C. R. Pennypacker, et al., Cosmology from Type Ia Supernovae, *Bulletin of the American Astronomical Society* **29** 1351 (1997).
3. H. A. Bethe, P. Pizzochero, Mass-Energy Relation for SN 1987A from Observations, *ApJL*, **350**, L33 (1990).
4. N. Soker, Possible post-kick jets in SN 1987A, *New Astronomy*, **84**, article id. 101548 (2021).
5. A. Saha, A. Sandage, G. A. Tammann, L. Labhardt, F. D. Macchetto, N. Panagia, Cepheid Calibration of the Peak Brightness of SNe Ia – IX. SN 1989B in NGC 3627, *The Astrophysical Journal* **522** 802 (1999).

6. M. M. Phillips, A. G. Riess, W. H. Press, R. Kirshner, Improved Distances to Type Ia Supernovae with Multicolor Light-Curve Shapes: MLCS2k2, *The Astrophysical Journal* **473** 88 (1996).
7. I. V. Arkhangelskaja, The sources of long GRBs: Population inhomogeneity or possibility its using as standard candles, *Journal of Physics: Conf. Ser.* **1181** article id. 012050 (2019).
8. C. T. Kowal, W. L. W. Sargent, Supernovae discovered since 1885, *Astronomical Journal* **76** 756 (1971).
9. W. L. W. Sargent, L. Searle, C. T. Kowal, Supernovae and Supernova Remnants, *Cosmovici C.B. D. Reidel Publ. Dordrecht* 33 (1974).
10. Browse Software Development Team, Asiago Supernova Catalog `http://heasarc.gsfc.nasa.gov/W3Browse/star-catalog/asiagosn.html#sn_type`.
11. J. Guillochon, J. Parrent, The Open Supernova Catalog `https://sne.space/`.
12. Guillochon, J. Parrent, L. Z. Kelley, R. Margutti, An Open Catalog for Supernova Data, *ApJ*, **835**, article id. 64 (2017).
13. S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker and R. Quimby et al, Measurements of Ω and Λ from 42 High-Redshift Supernovae, *The Astrophysical Journal* **517** 565 (1999).
14. T. J. Bronder, I. M. Hook, P. Astier, *et al.*, SNLS spectroscopy: Testing for evolution in type la supernovae, *A&A* **477**, 717–734 (2008).
15. R. Barbon, E. Cappellaro, F. Ciatti, M. Turatto, T. Kowal, The supernova 1984A in NGC4419, *Astronomy and Astrophysics* **220** 83 (1989).
16. F. Zwicky, Clusters of Galaxies. in *Handbuch der Physik*, eds. S. Flugge, *Springer, Verlag, Berlin* **51** 390 (1959).
17. M. Karpowicz, K. Rudnicki, Preliminary catalogue of supernovae, discovered till the end of 1967. *Publ. Astron. Obs. Warsaw*, **15** 189 (1968).
18. P. Flin, M. Karpowicz, W. Murawski, K. Rudnicki, Catalogue of supernovae, *Acta Cosmologica* **8**, 5 (1979).
19. R. Barbon, E. Cappellaro, M. Turatto, Photometric and spectroscopic observations of four supernovae, *Astronomy and Astrophysics* 410 (1989).
20. S. van den Bergh, A Catalog of Recent Supernovae, *Astroph. J. Suppl.* **92** 219 (1994).
21. N. N. Samus, General Catalogue of Variable Stars, fourth edition, *Kosmosinform* **5** (Moscow, 1995).
22. `https://heasarc.gsfc.nasa.gov/W3Browse/all/asiagosn.html`.
23. `http://www.cbat.eps.harvard.edu/lists/RecentSupernovae.html`.
24. D. Yu. Tsvetkov, N. N. Pavlyuk, O. S. Bartunov, Yu. P. Pskovskii, SAI Supernova Catalog `http://stella.sai.msu.ru/~pavlyuk/sncat/pages.php?results_per_page=500&resorder=0&ascdesc=1&resnewwin=0`.
25. Moore, M.; Li, W.; Filippenko, A. V.; Chornock, R.; Foley, R. J. Supernovae 2004T, 2004U, and 2004W IAU Circ., No. 8286, #2 (2004).
26. D. Branch, P. Garnavich, T. Matheson, E. Baron, R. C. Thomas, K. Hatano, P. Challis, S. Jha, R. P. Kirshner, Optical Spectra of the Type Ia Supernova 1998aq, *The Astrophysical Journal* **126** 1489 (2003).
27. astrocatalogs/supernovae `https://github.com/astrocatalogs/supernovae/issues`.
28. P. J. Brown, A. Breeveld, P. W. Roming and M. Siegel, Swift Supernovae, *The Astrophysical Journal*, **152(4)**, article id. 102 (2016).
29. The Pan-STARRS1 data archive home page, `https://panstarrs.stsci.edu/`.
30. List of Recent Supernovae, `http://www.cbat.eps.harvard.edu/lists/RecentSupernovae.html`.

31. D. A. Green, A catalogue of 294 Galactic supernova remnants, *BASI*, **42**, 47 (2014).
32. D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan- STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *The Astrophysical Journal* **859** 101 (2018).
33. Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., Kirshner, R. P., Hubble Residuals of Nearby Type Ia Supernovae are Correlated with Host Galaxy Masses, ApJ, 715, 743 (2010).
34. Sullivan, M., Conley, A., Howell, D. A., Neill, J. D.; Astier, P.; Balland, C.; Basa, S.; Carlberg, R. G.; Fouchez, D.; Guy, J.; Hardin, D.; Hook, I. M. search by orcid; Pain, R.; Palanque-Delabrouille, N.; Perrett, K. M.; Pritchet, C. J.; Regnault, N.; Rich, J.; Ruhlmann-Kleider, V.; Baumont, S. Hsiao, E.; Kronborg, T.; Lidman, C.; Perlmutter, S.; Walker, E. S., The dependence of Type Ia Supernovae luminosities on their host galaxies, *MNRAS*, **406** 782 (2010).
35. D. Scolnic and R. Kessler, *The Astrophysical Journal Letters* **822** L35 (2016).
36. K. S. Mandel, D. Scolnic, H. Shari, R. J. Foley, R. P. Kirshner, ArXiv e-prints, arXiv:1609.04470 (2016).
37. J. Guy, M. Sullivan, A. Conley, et al., The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints, *Astronomy and Astrophysics* **523** id:A7 (2010).
38. E. F. Schlafly and D. P. Finkbeiner, MEASURING REDDENING WITH SLOAN DIGITAL SKY SURVEY STELLAR SPECTRA AND RECALIBRATING SFD, *The Astrophysical Journal* **737** id:103 (2011).
39. A. Kim, A. Goobar, S. Perlmutter, A Generalized Correction for Type la Supernovae: Comparing R-band Photometry Beyond z=0.2 with B, V, and R-band Nearby Photometry, *Publications of the Astronomical Society of the Pacific* **108** 190 (1996).
40. W. D. Kenworthy, D. O. Jones, M. Dai,R. Kessler,D. Scolnic, D. Brout, M. R. Siebert, J. D. R. Pierel,K. G. Dettman, G. Dimitriadis, R. J. Foley, S. W. Jha, Y.-C. Pan, A. Riess, S. Rodney, and C. Rojas-Bravo, SALT3: An Improved Type Ia Supernova Model for Measuring Cosmic Distances, arXiv:2104.07795v1 (2021).
41. A. G. Riess, Kinematics and Dark Energy from Supernovae at $z > 1$, *ASP Conference Series*, **339** 39 (2005).
42. M. Sullivan, R. S. Ellis, G. Aldering, R. Amanullah, P. Astier, G. Blanc, M. S. Burns, A. Conley, S. E. Deustua, M. Doi, S. Fabbro, G. Folatelli, A. S. Fruchter, G. Garavini, R. Gibbons, G. Goldhaber, A. Goobar, D. E. Groom, D. Hardin, I. Hook, D. A. Howell, M. Irwin, A. G. Kim, R. A. Knop, C. Lidman, R. McMahon, J. Mendez, S. Nobili, P. E. Nugent, R. Pain, N. Panagia, C. R. Pennypacker, S. Perlmutter, R. Quimby, J. Raux, N. Regnault, P. Ruiz-Lapuente, B. Schaefer, K. Schahmaneche, A. L. Spadafora, N. A. Walton, L. Wang, W. M. Wood-Vasey and N. Yasuda, The Hubble diagram of type Ia supernovae as a function of host galaxy morphology, *MNRAS*, **340** 1057 (2003).
43. F. Hoyle and W. A. Fowler, Nucleosynthesis in Supernovae., *The Astrophysical Journal* **132** 565 (1960).
44. W. Hillebrandt, M. Kromer, F. K. Röpke and A. J. Ruiter, Towards an understanding of Type Ia supernovae from a synthesis of theory and observations, *Frontiers of Physics* **8** 116 (2013).
45. O. G. Benvenuto, J. A. Panei, K. Nomoto, H. Kitamura, I. Hachisu, Final Evolution and Delayed Explosions of Spinning White Dwarfs in Single Degenerate Models for Type Ia Supernovae, *The Astrophysical Journal Letters* **809** id:L6 (2015).
46. S. Heimersheim, N. Schoneberg, D. C. Hooper, J. Lesgourgues, Cannibalism hinders growth: Cannibal Dark Matter and the S8 tension, *arXiv:2008.08486v2* (7 Sep 2020).

47. L. A. Anchordoqui, E. Di Valentino, S. Pan, W. Yang, Dissecting the H_0 and S_8 tensions with Planck + BAO + supernova type Ia in multi-parameter cosmologies, *Journal of High Energy Astrophysics* **32**, 28 (2021).
48. R. C. Nunes, E. Di Valentino, Dark sector interaction and the supernova absolute magnitude tension, *Physical Review D* **104** id:063529 (2021).
49. G. Fanizza, Precision Cosmology and Hubble tension in the era of LSS surveys, CERN preprint # CERN-TH-2021-173 (2021), arXiv:2110.15272v2.
50. N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, p. A6 (2020).
51. A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn and D. Scolnic, Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM, *Astrophys. J. Lett.* **908**, L6 (2021).
52. D. Camarena and V. Marra, Local determination of the Hubble constant and the deceleration parameter, *Phys. Rev. Res.* **2** id: 013028 (2020).
53. J. Elvin-Poole, M. Crocce, A. J. Ross, T. Giannantonio, Dark Energy Survey Year 1 Results: Galaxy clustering for combined probes *Monthly Notices of the Royal Astronomical Society* **478** 592 (2018) `https://des.ncsa.illinois.edu/releases/sn`.

Understanding prompt emission: Where do we stand?

Asaf Pe'er
Department of Physics, Bar Ilan University,
Ramat-Gan 52900, Israel
E-mail: asaf.peer@biu.ac.il
asafpeer2.ph.biu.ac.il/

In recent years, there is a renewed debate about the origin of the observed prompt emission signal. Some authors found that synchrotron emission can dominate the spectra of several long bursts, and a recent analysis show that it may be possible to overcome the famous 'line of death' argument by a direct fitting procedure. On the other hand, several recent works showed that non-dissipative photosphere is preferred as the dominant emission model in at least 1/4 of long and 1/3 of short GRB population. Here I critically review the arguments given as well as their physical consequences. I present some recent results that show a connection between the prompt spectra and the early afterglow emission, thereby argue for an independent method of discriminating the physical conditions that result in the different dominant radiative processes.

Keywords: Gamma-ray: bursts; Radiative mechanisms: non-thermal; photosphere; polarization

1. Introduction

After more than three decades of an extensive study, there is a broad consensus about many of the observational properties defining gamma-ray bursts (GRBs), yet there is still a strong debate about the correct way of interpreting the data. One of the key problems is that when interpreting the observed data, one may conclude about the leading radiative process that produces the observed signal. However, the radiative process is the last in a long chain of physical processes, that include gravitational energy extraction, relativistic motion, kinetic energy dissipation, particle acceleration and radiative process that eventually result in the observed signal. It is therefore impossible to disentangle the radiative process from the rest of this chain. Furthermore, the underlying physics of most of these processes is still only partially understood, and is the subject of an intensive research. As a result of this complexity, the leading radiative process is still uncertain.

When looking at the data, common to all GRBs are the following. The spectra has a peak in the sub-MeV range; at lower energies (below this peak) the spectral slopes are steep, $F_\nu \propto \nu^\alpha$, with $\alpha \approx 0$, but with considerable variation between bursts, and sensitive to the fitting procedure; In several bursts a higher energy emission, at the GeV-TeV range is seen; and when making spectral fittings for bursts with strong enough signal, multiple components are identified.[1]

The lightcurve of the prompt emission, which lasts for a few- few tens of seconds generally varies at a sub-second level. The prompt phase itself is distinguished from later emission ("afterglow") by a sharp decay in the flux. In a large fraction - 10's of % of GRBs, an X-ray plateau is detected, lasting for a few thousands seconds following the prompt phase. At later phase, a decaying "afterglow" which is well modeled by a self-similar expansion phase (the "Blandford-McKee" solution) is frequently detected. Furthermore, to many nearby long GRBs, evidence for a coinciding supernovae explosion exist. Finally, the connection between GRB170817 and the gravitational wave (GW) event clearly indicates that this particular event is associated with the merger of binary neutron stars.[2]

Following these plethora of data, already in the early 1990's the basic "fireball" picture had emerged,[3–6] according to which the gravitational energy released during the collapse of a massive star (or merger of binaries) is partially converted to kinetic energy in the form of a relativistic jet. At a second stage, part of this kinetic energy is released by internal dissipation (e.g., shock waves). Part of this energy is used in a accelerating particles, which then emit the observed signal. This model could explain both the high energy ($>$ MeV) emission, which is above the threshold for pair production, as well as the high variability detected in the lightcurve. It further obtained the afterglow as a prediction, to be later confirmed after the launch of the Beppo-SAX satellite.

Despite this success, many of the basic questions remain unanswered in the framework of this model. These include: (i) the nature of the progenitor star, and the event that led to its explosion and death; (ii) the nature of the launching mechanism of the relativistic jet[7,8]; (iii) The composition of the jet, and in particular its magnetization; (iv) The structure (geometry), dynamics (velocity field) and evolution of the propagating jet; (v) The nature of the kinetic energy dissipation mechanism - while initially internal shocks were considered, it became evident that they are not very efficient energy converters,[9] and additional mechanisms were proposed, mainly involving strong magnetic fields[10,11]; This directly related to the question of particle acceleration, which is associated with this dissipation. (vi) The nature of the radiative processes involved in producing the observed signal. (vii) Addressing these questions is important in understanding the connection of GRBs to other objects of interest and field of active research, such as stellar evolution, star formation, pop-III stars, supernovae, binary star evolution, cosmic rays, energetic neutrinos, and more. Finally, complete understanding of the nature of GRBs could be useful for addressing fundamental questions in cosmology (e.g., about the expansion of the universe) and Lorentz violation.

The key problem remains that while many theories exist for each of these open questions, in order to test the theoretical ideas one has to rely on the observed signal, which is the result of the long chain of physical processes. Thus, one has to find clever ways of discriminating between the models based on the final outcome, which is not directly related to the processes themselves, and in many cases is very confusing.

2. The leading radiative processes

Very broadly speaking, there are three competing models that are in wide use in explaining the observed prompt emission signal. These are synchrotron as a leading radiative process, hadronic and photospheric model.

The idea that synchrotron emission (possibly accompanied by inverse-Compton [IC] scattering at higher energies) is the leading radiative process during the prompt phase of GRBs was in wide use in the 1990's,[12,13] and regain interest in recent years.[14–16] Synchrotron emission is indeed very ubiquitous in many different astronomical objects, and in many GRBs it provides very good fitting to the lightcurve and spectra seen during the afterglow phase. Indeed, it is a very efficient way of extracting energy from energetic electrons via radiation. Furthermore, the conditions for this process to occur, namely the existence of energetic electrons and a magnetic field are both believed to exist as part of the dissipation process of the relativistic jet's kinetic energy. This is despite the fact that a complete model for the dissipation process still does not exist.

A second possibility that had been discussed is that of a hadronic origin, namely that the observed signal is due to emission from energetic protons, e.g., via synchrotron emission,[17] or possibly photo-meson (pion production) or Bethe-Heitler (pair production) interaction of the protons with the ambient photon field. In order for this process to take place, existence of energetic protons and a magnetic field are required. While synchrotron emission from protons is considerably less efficient than emission from the electrons due to the smaller cross-section, the advantage is that observations of energetic cosmic rays provide a direct evidence that proton acceleration to high energies necessarily exist in astronomical objects, while evidence for acceleration of electrons exist only indirectly. Thus, as long as the theoretical understanding of particle acceleration is incomplete, we do have certainty that this process does exist, though of course it is not clear whether the conditions that exist during GRB prompt emission are sufficient.

The third radiative process discussed is emission from the photosphere.[18] As opposed to the previous alternatives, this can be considered as a more complete model. As part of the GRB "fireball" model, the flow begins its relativistic expansion while being optically thick. Thus, at a certain radius, it must become optically thin - this is the photospheric radius, where photons escape. Thus, a photospheric component naturally exists in expanding, (initially) optically-thick flows, and in particular as part of the GRB "fireball" model. Furthermore, as no additional kinetic energy dissipation is required, this process can be highly efficient, with 10s of % of energy being released as photons at the photosphere.[19–21] On the other hand, in order to gain high efficiency, one requires to avoid adiabatic losses, namely that the photospheric radius r_{ph} is not much above the coasting radius, which marks the end of the acceleration of the jetted material.

3. Confronting synchrotron emission with the data

Already in the 1990's, synchrotron emission was suggested as a leading radiative model to explain the prompt emission from GRBs. This is based on the fact that for relativistic expansion with $\Gamma \gtrsim 100$ and dissipated kinetic energy that is spread roughly equally between accelerating electrons and generating magnetic fields (i.e., equipartition), the observed peak energy, $\nu_m^{ob.} \sim 500$ KeV is in good agreement with the peak observed energy.

However, a closer look reveals that under such a strong magnetic field, the electrons will rapidly cool (namely, will be in the "fast cooling" regime), in which case the expected low energy spectral slope is much shallower than the observed one. This is the famous "synchrotron line of death" problem,[22] which made this model less appealing for many years.

In recent years, as the quality of data increased, several attempts were made to overcome this problem. For example, Refs.[14,15,23] argue that the low energy spectral slope (below the peak) cannot be considered as a single power low, but rather there is an intermediate break at several tens of keV. This break enables to fit the spectra with a smaller than equipartition value of the magnetic field, thereby relaxing the 'line of death' restriction. Recently, Ref.[16] correctly argued that one should not attempt to fit the synchrotron spectra to the already fitted "Band" function, as is commonly the case, but rather directly to the raw data. When doing so, they managed to provide acceptable fits to the data within the framework of the synchrotron emission model.

Despite this success, a close look reveals that in providing these fits, the magnetic field assumed was extremely weak, ~ 1 G, and the emission radius obtained from the fit was $\sim 10^{17}$ cm. These require a ratio of burst energy to ambient density of $E_{53}/n_0 \sim 10^8$, which is 8 orders of magnitude larger than inferred from afterglow data (here and below, $Q_x = Q/10^x$). Thus, while the shape of the spectra can be fitted with such a model, the values of parameters obtained make it physically unacceptable.

Similarly, the idea of synchrotron emission from protons was investigated by several authors.[17,24] The allowed parameter space in this scenario, $B \sim 10^6$ G and emission radius of $R \sim 10^{13}$ cm, are much in line with the parameter estimate from afterglow data. However, due to the much less efficient radiative energy extraction from the protons (as compared to the electrons), this scenario requires that the total amount of energy given to the accelerated protons, E_p exceeds the jet kinetic energy, namely $E_p > E_k$. Thus, this scenario challenges models of particle acceleration.

4. Confronting photospheric emission with the data

The drawbacks of the synchrotron model mentioned above, led several authors to consider photospheric emission as an alternative.[18,19,25] However, this alternative

was less appealing for many years, due to the fact that the observed spectra never appears as a "pure black body" (Planck), but rather is much broader.

This situation changes with the observation of GRB090902B.[26] This was a particularly bright burst, which showed clear evidence for a dominant narrow peak, Planck-like component. Furthermore, its extreme brightness enabled a time-resolved analysis during the onset of the prompt phase. This analysis proved that the spectrum change during the prompt phase: it started as a very narrow peak, close to a black body, and then gradually widened to resemble the common "Band" spectra. This had demonstrated the importance of **time-resolved** analysis, as an essential tool needed to understand the origin of the prompt phase.[27]

Recently, a similar analysis done on a large number of bursts[28] showed that this behavior is in fact common among pulses seen in many bright GRBs: initially, many pulses are narrow (hard spectrum) which become softer in time. This may be interpreted as a change in the leading radiative process with time: an initial thermal (or quasi-thermal) emission decays, and a later synchrotron emission becomes dominate at later times.

Furthermore, theoretical works[29–31] proved that in fact the initial expectation for a "Planck" spectrum was misleading: due to light aberration from the relativistically expanding jet, even a "pure" photospheric emission (so called "non-dissipative photosphere", or NDP) would be detected as a broader spectra, with spectral slope $dN/dE \propto E^{-\alpha}$ and $\alpha \approx 0.4$ below the peak[30] (compare with the Rayleigh-Jeans slope of $\alpha = 1$). Even this, though, is an asymptotic limit: when considering the finite detector's energy range, and the "curvature" of the spectrum near the energy peak, it is found that the expected observed values of the low energy spectral slope are even smaller than this.[32]

After considering these effects, a recent analysis show that in fact more than 1/4 of long GRBs, and 1/3 of short GRBs are consistent with having a pure thermal origin.[33,34] When adding a possible sub-photospheric energy dissipation that can potentially broaden the spectra, these fraction of course becomes much larger.

5. Applications

The realization that a photospheric emission does not appear as a textbook-solution blackbody, but rather is modified, enabled a refined analysis of the GRB properties. For example, Ref.[34] showed that the short GRB population that are consistent with having a pure thermal origin can be divided into two separate populations, clearly divided by their peak energy, E_{pk}, duration (T_{90}) and inferred Lorentz factor. One of these populations forms a continuation of the properties of the long GRB population, while the other one is separated. This result thus suggests that the classical separation of the GRB population to "long" and "short", which is based nearly entirely on the duration (T_{90}) is likely insufficient, and additional criteria are needed. Part of the "short" GRB population is in fact a continuation of the long GRB population.

One of the most puzzling questions is the observed isotropic energy-peak energy ("Amati") or luminosity - peak energy ("Yonetoku") relations reported. One appealing advantage of the photospheric emission model is that it may provide a natural explanation to the observed correlation, as due to viewing angle effect. Observers at different angles see the photons with different Doppler shifts, as $\nu^{ob.} = D\nu'$ (here, $D = [\Gamma(1 - \beta\cos\theta)]^{-1}$ is the Doppler shift, and primed quantities are in the comoving frame). Using the invariance of L'/ν^2, the observed luminosity is related to the peak energy via $L^{ob.} \propto (\nu^{ob.})^2 \sim E_{pk}^2$, providing a natural explanation to the Yonetoku relation.[35]

While this may provide a very elegant explanation to the observed relation, this analysis relies on the underlying assumption that the comoving luminosity and comoving temperature are related via $L' \propto (T')^2$, while one expect the comoving temperature to decay with radius.[36] Thus, an underlying assumption here is that the photospheric radius is roughly similar in many different GRBs, which is unknown. I further discuss this issue in section 7 below.

6. Broadening of the photospheric emission

Numerical simulations of jets propagating inside and outside the collapsing star[21,37–39] clearly demonstrate that part of the jet kinetic energy is dissipated at various radii, in many cases below the photosphere. Such a dissipation can result from recollimation shock with the collapsing stellar material, magnetic energy dissipation or internal collisions. The resulting spectra was studied by various authors, e.g., Refs.[19,25,40–42] to name a few.

In this scenario, the dissipated energy is used to heat the plasma, which is then characterized by 2 temperatures: an electron temperature T_e which is greater than the photon temperature. Under these conditions, the resulting spectrum above the thermal peak is mainly due to inverse Compton (IC) scattering of the photons inside the plasma. It depends mainly on 2 parameters: the number of scattering (namely, the optical depth at the location of the energy dissipation) and the available energy, namely the ratio between the electron and the photon field energies, u_{el}/u_{ph}.

Below the thermal peak, one may expect an additional modification due to synchrtron emission from the (newly) heated electrons. Above the thermal peak, multiple scattering broadens the peak; if the optical depth is high enough, a new thermal peak would emerge, at higher temperatures. But in the intermediate case, $\tau \lesssim 100$, there is no simple analytical solution, and one has to refer to numerical works,[25,30,43–45] which show the plethora of possible spectra.

Another, independent broadening mechanism to be considered is geometrical broadening, which is the relativistic version of the "limb darkening" effect. Even for a spherical explosion, the photospheric radius is a function of the angle to the line of sight, θ, via $r_{ph}(\theta) \propto \left(\Gamma^{-2} + \theta^2/3\right)$, where Γ is the Lorentz factor of the flow.[29] This implies that photons emerging from angles $\theta > \Gamma^{-1}$ escape the expanding plasma at large radii. Furthermore, the photons have a finite probability of

escaping the expanding plasma at different radii, leading to the concept of "vague photosphere".[29,30] This distribution inherently results in broadening of the emerging signal.

The explosion, though is not expected to be spherical, but rather a jet-like, with a finite jet opening angle. The flow properties may best be approximated as $\Gamma(\theta) \propto \Gamma_0/\sqrt{(1+(\theta/\theta_j)^{2p}}$ (see simluation by[38]), where θ_j is the jet opening angle, and p represents the "steepness" of the velocity profile. In this case, the observed signal depends, in addition, on the viewing angle, θ_v.[46] This scenario naturally leads to an enhancement of the limb darkening effect, as the Doppler boost of photons emitted from larger angles and larger radii is considerably smaller than that of photons emitted from the central jet regions. As a result, these photons are observed at much lower energies, and the observed spectrum – purely from the photosphere – is nearly flat for a wide energy range, for different jet parameters and viewing angles.[46]

This scenario naturally results in a polarized signal, for an off-axis observer. The scattered photons are 100% polarized, the photon field is anisotropic near the photosphere, and the rotational symmetry is broken for $\theta_v > 0$ (observer off the jet axis). The resulting polarization can reach 10's of % ($Q/I \lesssim 40\%$ for an off-axis observer, with flux that is still detectable). Another prediction of the theory is $\pi/2$ shift in polarization angle,[47] with spectra that is so smeared it looks very different from a pure "Planck". Indeed, such a polarization angle change was recently detected by the ASTROSAT mission,[48] though its exact origin is still debated.

7. Back scattering photospheric model

An underlying assumption of the discussion above is that the source of photons is the expanding plasma inside the jet: either via acceleration of particles that radiate synchrotron, or that the photons are coupled to the plasma below the photosphere. However, one may envision a different scenario, in which, following an initial expansion of the jet which clears a cavity inside the collapsing star,[39,49] continuous production of pairs via, e.g., neutrino annihilation occurs close to the newly formed black hole. These pairs then annihilate to produce photons. The photons, in turn, propagate inside the empty funnel until being scattered by the expanding stellar "cork" (made of stellar material cleared by the jet and by the surrounding cocoon).

In this scenario, therefore, the photons are back scattered by the relativistically expanding cork ahead of them from behind. Due to the relativistic motion of the cork, in the frame of an observer at infinity these photons are scattered in the forward direction, at some angle to the cork propagation direction. Due to the different Doppler boosts, photons scattered at different angles are seen at different energies. The obtained spectra integrated over different viewing angles, $dN/d\epsilon \propto \epsilon^1$ for energies below the peak, resembles the observed one.[50] For hot cork, multiple scattering inside the cork leads to photon energy gain,and the high energy spectral slope obtained is also similar to the observed.[51]

Perhaps the greatest achievement of this model is its ability to naturally explain the $E_{pk} - E_{iso}$ correlation ("Amati" relation) as a natural outcome of observers located at different angles, without the need to invoke any additional underlying assumptions, as in the "classical" photospheric model discussed above.

8. Summary

Understanding the origin of GRB prompt emission remains one of the big puzzles in the study of gamma-ray bursts. It is of particular importance, due to the fact that understanding it directly enables to probe the key physical ingredients involved in producing the GRB phenomena.

In recent years, there is a healthy debate between different ideas of interpreting the observed signal. As it turned out, different models both seem to produce key parts of the spectra. However, the physical parameters required for the synchrotron radiation to be the dominant mechanism seem at odds with afterglow observations.

Photospheric emission model turned out to be much more complicated than initially thought. Various effects and mechanisms act to modify the observed signal, which does not resemble the textbook "Planck" function. Among them are the finite detector's bandwidth, the jet velocity profile and sub-photospheric energy dissipation. For jet viewed off-axis, as in the vast majority of jets, the observed signal may be highly polarized. Another major advantage is the ability to explain the observed $E_{pk} - E_{iso}$ correlation, which is a natural outcome of a model in which the source of photons is not relativistically moving, but rather the photons are back-scattered from the expanding cork.

Acknowledgments

I wish to thank Damien Begue, Husne Dereli-Begue, Liang Li, Felix Ryde, Vidushi Sharma and Mukesh Vyas for many helpful discussions. This talk is attributed to the memory of Prof. David Eichler, a collaborator and friend, who sadly passed away earlier this year.

References

1. S. Iyyani, F. Ryde, J. M. Burgess, A. Pe'er and D. Bégué, Synchrotron emission in GRBs observed by Fermi: Its limitations and the role of the photosphere, *Mon. Not. R. Astron. Soc.* **456**, 2157 (February 2016).
2. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., Multi-messenger Observations of a Binary Neutron Star Merger, *Astrophys. J.* **848**, p. L12 (October 2017).
3. B. Paczynski, Gamma-ray bursters at cosmological distances, *Astrophys. J.* **308**, L43 (September 1986).
4. J. Goodman, Are gamma-ray bursts optically thick?, *Astrophys. J.* **308**, L47 (September 1986).

5. M. J. Rees and P. Meszaros, Relativistic fireballs - Energy conversion and time-scales, *Mon. Not. R. Astron. Soc.* **258**, 41P (September 1992).
6. M. J. Rees and P. Meszaros, Unsteady outflow models for cosmological gamma-ray bursts, *Astrophys. J.* **430**, L93 (August 1994).
7. R. D. Blandford and R. L. Znajek, Electromagnetic extraction of energy from Kerr black holes, *Mon. Not. R. Astron. Soc.* **179**, 433 (May 1977).
8. R. Popham, S. E. Woosley and C. Fryer, Hyperaccreting Black Holes and Gamma-Ray Bursts, *Astrophys. J.* **518**, 356 (June 1999).
9. S. Kobayashi, T. Piran and R. Sari, Can Internal Shocks Produce the Variability in Gamma-Ray Bursts?, *Astrophys. J.* **490**, p. 92 (November 1997).
10. G. Drenkhahn, Acceleration of GRB outflows by Poynting flux dissipation, *Astron. Astrophys.* **387**, 714 (May 2002).
11. G. Drenkhahn and H. C. Spruit, Efficient acceleration and radiation in Poynting flux powered GRB outflows, *Astron. Astrophys.* **391**, 1141 (September 2002).
12. P. Meszaros, P. Laguna and M. J. Rees, Gasdynamics of relativistically expanding gamma-ray burst sources - Kinematics, energetics, magnetic fields, and efficiency, *Astrophys. J.* **415**, 181 (September 1993).
13. M. Tavani, A Shock Emission Model for Gamma-Ray Bursts. II. Spectral Properties, *Astrophys. J.* **466**, p. 768 (August 1996).
14. M. E. Ravasio, G. Oganesyan, G. Ghirlanda, L. Nava, G. Ghisellini, A. Pescalli and A. Celotti, Consistency with synchrotron emission in the bright GRB 160625B observed by Fermi, *Astron. Astrophys.* **613**, p. A16 (May 2018).
15. G. Oganesyan, L. Nava, G. Ghirlanda, A. Melandri and A. Celotti, Prompt optical emission as a signature of synchrotron radiation in gamma-ray bursts, *Astron. Astrophys.* **628**, p. A59 (August 2019).
16. J. M. Burgess, D. Bégué, J. Greiner, D. Giannios, A. Bacelj and F. Berlato, Gamma-ray bursts as cool synchrotron sources, *Nature Astronomy* **4**, 174 (February 2020).
17. N. Gupta and B. Zhang, Prompt emission of high-energy photons from gamma ray bursts, *Mon. Not. R. Astron. Soc.* **380**, 78 (September 2007).
18. M. J. Rees and P. Mészáros, Dissipative Photosphere Models of Gamma-Ray Bursts and X-Ray Flashes, *Astrophys. J.* **628**, 847 (August 2005).
19. A. Pe'er, P. Mészáros and M. J. Rees, Peak Energy Clustering and Efficiency in Compact Objects, *Astrophys. J.* **635**, 476 (December 2005).
20. K. Ioka, K. Toma, R. Yamazaki and T. Nakamura, Efficiency crisis of swift gamma-ray bursts with shallow X-ray afterglows: prior activity or time-dependent microphysics?, *Astron. Astrophys.* **458**, 7 (October 2006).
21. O. Gottlieb, A. Levinson and E. Nakar, High efficiency photospheric emission entailed by formation of a collimation shock in gamma-ray bursts, *Mon. Not. R. Astron. Soc.* **488**, 1416 (September 2019).
22. R. D. Preece, M. S. Briggs, R. S. Mallozzi, G. N. Pendleton, W. S. Paciesas and D. L. Band, The Synchrotron Shock Model Confronts a "Line of Death" in the BATSE Gamma-Ray Burst Data, *Astrophys. J.* **506**, L23 (October 1998).
23. M. E. Ravasio, G. Ghirlanda, L. Nava and G. Ghisellini, Evidence of two spectral breaks in the prompt emission of gamma-ray bursts, *Astron. Astrophys.* **625**, p. A60 (May 2019).
24. G. Ghisellini, G. Ghirlanda, G. Oganesyan, S. Ascenzi, L. Nava, A. Celotti, O. S. Salafia, M. E. Ravasio and M. Ronchi, Proton-synchrotron as the radiation mechanism of the prompt emission of gamma-ray bursts?, *Astron. Astrophys.* **636**, p. A82 (April 2020).

25. A. Pe'er, P. Mészáros and M. J. Rees, The Observable Effects of a Photospheric Component on GRB and XRF Prompt Emission Spectrum, *Astrophys. J.* **642**, 995 (May 2006).
26. A. A. Abdo, M. Ackermann, M. Ajello, K. Asano, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, M. G. Baring and et. al., Fermi Observations of GRB 090902B: A Distinct Spectral Component in the Prompt and Delayed Emission, *Astrophys. J.* **706**, L138 (November 2009).
27. F. Ryde, A. Pe'er, T. Nymark, M. Axelsson, E. Moretti, C. Lundman, M. Battelino, E. Bissaldi, J. Chiang, M. S. Jackson, S. Larsson, F. Longo, S. McGlynn and N. Omodei, Observational evidence of dissipative photospheres in gamma-ray bursts, *Mon. Not. R. Astron. Soc.* **415**, 3693 (August 2011).
28. L. Li, F. Ryde, A. Pe'er, H.-F. Yu and Z. Acuner, Bayesian Time-resolved Spectroscopy of Multipulse GRBs: Variations of Emission Properties among Pulses, *Astrophys. J.* **254**, p. 35 (June 2021).
29. A. Pe'er, Temporal Evolution of Thermal Emission from Relativistically Expanding Plasma, *Astrophys. J.* **682**, 463 (July 2008).
30. A. M. Beloborodov, Collisional mechanism for gamma-ray burst emission, *Mon. Not. R. Astron. Soc.* **407**, 1033 (September 2010).
31. A. Pe'er and F. Ryde, A Theory of Multicolor Blackbody Emission from Relativistically Expanding Plasmas, *Astrophys. J.* **732**, 49 (May 2011).
32. Z. Acuner, F. Ryde and H.-F. Yu, Non-dissipative photospheres in GRBs: Spectral appearance in the Fermi/GBM catalogue, *Mon. Not. R. Astron. Soc.* **487**, 5508 (August 2019).
33. Z. Acuner, F. Ryde, A. Pe'er, D. Mortlock and B. Ahlgren, The Fraction of Gamma-Ray Bursts with an Observed Photospheric Emission Episode, *Astrophys. J.* **893**, p. 128 (April 2020).
34. H. Dereli-Bégué, A. Pe'er and F. Ryde, Classification of Photospheric Emission in Short GRBs, *Astrophys. J.* **897**, p. 145 (July 2020).
35. H. Ito, J. Matsumoto, S. Nagataki, D. C. Warren, M. V. Barkov and D. Yonetoku, The photospheric origin of the Yonetoku relation in gamma-ray bursts, *Nature Communications* **10**, p. 1504 (April 2019).
36. A. Pe'er, B.-B. Zhang, F. Ryde, S. McGlynn, B. Zhang, R. D. Preece and C. Kouveliotou, The connection between thermal and non-thermal emission in gamma-ray bursts: General considerations and GRB 090902B as a case study, *Mon. Not. R. Astron. Soc.* **420**, 468 (February 2012).
37. M. A. Aloy, E. Müller, J. M. Ibáñez, J. M. Martí and A. MacFadyen, Relativistic Jets from Collapsars, *Astrophys. J.* **531**, L119 (March 2000).
38. W. Zhang, S. E. Woosley and A. I. MacFadyen, Relativistic Jets in Collapsars, *Astrophys. J.* **586**, 356 (March 2003).
39. D. López-Cámara, B. J. Morsony, M. C. Begelman and D. Lazzati, Three-dimensional Adaptive Mesh Refinement Simulations of Long-duration Gamma-Ray Burst Jets inside Massive Progenitor Stars, *Astrophys. J.* **767**, p. 19 (April 2013).
40. D. Giannios, Prompt emission spectra from the photosphere of a GRB, *Astron. Astrophys.* **457**, 763 (October 2006).
41. I. Vurm, Y. Lyubarsky and T. Piran, On Thermalization in Gamma-Ray Burst Jets and the Peak Energies of Photospheric Spectra, *Astrophys. J.* **764**, p. 143 (February 2013).
42. H. Ito, S. Nagataki, J. Matsumoto, S.-H. Lee, A. Tolstov, J. Mao, M. Dainotti and A. Mizuta, Spectral and Polarization Properties of Photospheric Emission from Stratified Jets, *Astrophys. J.* **789**, p. 159 (July 2014).

43. I. Vurm, A. M. Beloborodov and J. Poutanen, Gamma-Ray Bursts from Magnetized Collisionally Heated Jets, *Astrophys. J.* **738**, 77 (September 2011).
44. D. Lazzati and M. C. Begelman, Non-thermal Emission from the Photospheres of Gamma-ray Burst Outflows. I. High-Frequency Tails, *Astrophys. J.* **725**, 1137 (December 2010).
45. D. Giannios, The peak energy of dissipative gamma-ray burst photospheres, *Mon. Not. R. Astron. Soc.* **422**, 3092 (June 2012).
46. C. Lundman, A. Pe'er and F. Ryde, A theory of photospheric emission from relativistic, collimated outflows, *Mon. Not. R. Astron. Soc.* **428**, 2430 (January 2013).
47. C. Lundman, A. Pe'er and F. Ryde, Polarization properties of photospheric emission from relativistic, collimated outflows, *Mon. Not. R. Astron. Soc.* **440**, 3292 (June 2014).
48. V. Sharma, S. Iyyani, D. Bhattacharya, T. Chattopadhyay, S. V. Vadawale and V. B. Bhalerao, Spectropolarimetric analysis of prompt emission of GRB 160325A: Jet with evolving environment of internal shocks, *Mon. Not. R. Astron. Soc.* **493**, 5218 (April 2020).
49. O. Bromberg, Z. Mikolitzky and A. Levinson, Sub-photospheric Emission from Relativistic Radiation Mediated Shocks in GRBs, *Astrophys. J.* **733**, p. 85 (June 2011).
50. M. K. Vyas, A. Pe'er and D. Eichler, A Backscattering-dominated Prompt Emission Model for the Prompt Phase of Gamma-Ray Bursts, *Astrophys. J.* **908**, p. 9 (February 2021).
51. M. K. Vyas, A. Pe'er and D. Eichler, Predicting Spectral Parameters in the Backscattering-dominated Model for the Prompt Phase of GRBs, *Astrophys. J.* **918**, p. L12 (September 2021).

On explaining prompt emission from GRB central engines with photospheric emission model

M. Bhattacharya[1,*] and P. Kumar[2]

[1] *Department of Physics, Pennsylvania State University, University Park, PA 16802, USA*
[*] *E-mail: mmb5946@psu.edu*
[2] *Department of Astronomy, University of Texas, Austin, TX 78712, USA*

Although the observed spectra for gamma-ray burst (GRB) prompt emission is well constrained, the underlying radiation mechanism is still not very well understood. We explore photospheric emission in GRB jets by modelling the Comptonization of fast cooled synchrotron photons whilst the electrons and protons are accelerated to highly relativistic energies by repeated energy dissipation events as well as Coulomb collisions. In contrast to the previous simulations, we implement realistic photon-to-particle number ratios of $N_\gamma/N_e \sim 10^5$ or higher, that are consistent with the observed radiation efficiency of relativistic jets. Using our Monte Carlo radiation transfer (MCRaT) code, we can successfully model the prompt emission spectra when the electrons are momentarily accelerated to highly relativistic energies (Lorentz factor $\sim 50-100$) after getting powered by $\sim 30-50$ episodic dissipation events in addition to their Coulomb coupling with the jet protons, and for baryonic outflows that originate from moderate optical depths $\sim 20-30$. We also show that the resultant shape of the photon spectrum is practically independent of the initial photon energy distribution and the jet baryonic energy content, and hence independent of the emission mechanism.

Keywords: Gamma-ray burst: general; methods: numerical; radiation mechanisms: thermal; radiative transfer; scattering.

1. Introduction

The radiation mechanism responsible for the prompt emission of long-duration GRBs has remained elusive ever since their discovery. The observed photon spectrum has a distinctly non-thermal shape and is generally modelled using the Band function,[1] which is a smoothly connected broken power-law with peak energy $E_{\rm peak} \sim 300\,{\rm keV}$ and the low/high energy dependence[2,3] given by $f_\nu \propto \nu^0/f_\nu \propto \nu^{-1.2}$. A robust radiation mechanism should explain the non-thermal features observed in the prompt emission spectrum in a self-consistent manner. The two most widely explored models to this end are the synchrotron and photospheric models.[4–6]

In the synchrotron model, the electrons accelerated by either internal shocks[4] or magnetic reconnection[7] produce the prompt radiation via synchrotron emission process.[8] Although this model accounts for the non-thermal nature of the photon spectrum, it cannot explain the high observed radiation efficiencies up to few tens of percent.[9] Moreover, the observed hard GRB spectra at low energies cannot be explained with synchrotron emission process.[10,11] Owing to these shortcomings of the synchrotron model, many researchers have considered photospheric emission model

in more detail.[12,13] Unlike the synchrotron model, photospheric model naturally explains the observed high radiation efficiencies and spectral shape is completely determined by photon-matter interactions irrespective of the dissipation mechanism involved. Although there have been many successful attempts to explain the high-energy non-thermal behaviour using sub-photospheric dissipation,[7,14,15] explaining the low-energy non-thermal tails has turned out to be really challenging.[16]

Here we study the sub-photospheric Comptonisation of synchrotron seed photons[5,17] to find the plausible conditions under which both low/high-energy non-thermal behaviour and observed peak energy can be explained self-consistently. The particles in the jet are accelerated to relativistic energies by repeated dissipation events such as internal shocks[4,18] or magnetic reconnection.[7,19] The photons continue to scatter electrons while the outflow is optically thick and gain energy until either the average photon energy matches that of the electrons or the outflow becomes optically thin so that the photons escape the photosphere. We determine a correlation between the number of reheating events and initial optical depth, and perform an exhaustive parametric space search to obtain a Band-like photon spectrum. We also perform analytical calculations to examine the evolution of photon energy spectrum with multiple scatterings and validate the numerical results.

2. Energy requirement for photons and particles

We first estimate the average energy $E_{\gamma,\rm avg}$ that the jet photons need to have to generate a Band-like output spectrum. As most of this energy is transferred by hot electrons through Comptonisation, this also places a threshold energy requirement $\gamma_{e,\rm crit}$ on the electrons. As the electrons are maintained at $\gamma_e \sim \gamma_{e,crit}$ by energy gained from either Coulomb collisions or repeated dissipation events and subsequent cooling due to IC, we further constrain the injected energy and the number of episodic dissipation events required.

2.1. *Analytical estimate for* $E_{\gamma,\rm avg}$ *and* $\gamma_{e,\rm crit}$

The observed photon spectrum has a Band-like shape with a low/high-energy dependence, $f_\nu \propto \nu^0/\nu^{-1.2}$ in the energy range $\sim 10\,\rm keV - 300\,keV/\sim 300\,keV - 10\,MeV$, where f_ν denotes the photon flux per unit frequency. The average observed energy of each photon in the lab frame is then $E'_{\gamma,\rm avg} \sim 300\,\rm eV$, assuming a jet bulk Lorentz factor $\Gamma = 300$. Equating the total energy content of the photons with the maximum energy that the electrons can deposit via Comptonisation gives

$$N_\gamma E'_{\gamma,\rm avg} = N_e(\gamma_{e,\rm crit} - 1)m_e c^2 \left(\tau_{\rm in} t'_{\rm dyn}/t'_{\rm IC}\right), \quad (1)$$

where τ_{in} is the initial optical depth, $t'_{\rm dyn/IC}$ is the dynamical/inverse-Compton (IC) timescale and $\tau_{\rm in} t'_{\rm dyn}/t'_{\rm IC} \sim$ number of times the electrons interact with photons during jet expansion. The characteristic dynamical timescale is $t'_{\rm dyn} = R_{\rm in}/\Gamma c$ while the IC timescale is $t'_{\rm IC} = 3(\gamma_e - 1)m_e c/4U'_\gamma \sigma_T \gamma_e^2 \beta_e^2$. Here, $R_{\rm in}$ is the photon injection

radius, σ_T is the Thomson cross section, γ_e is the electron energy and U'_γ is the radiation energy density. Substituting typical GRB parameters gives $\gamma_{e,crit} = 1.352$.

We now estimate the electron super-Coulomb efficiency parameter η which leads to $\gamma_e = 1.352$. For electrons to remain in equilibrium as a result of Coulomb heating and IC cooling over the jet expansion timescale,

$$\frac{(\gamma_e - 1)m_e c^2}{5 \times 10^{-19} n'_e} \cdot \frac{(8.3 \times 10^{-15} T_e'^{3/2} + \beta_p^3)}{\beta_p^2} \cdot \frac{1}{\eta} = \frac{3}{4} \frac{(\gamma_e - 1)m_e c}{U'_\gamma \sigma_T \gamma_e^2 \beta_e^2}, \tag{2}$$

where electron density $n'_e = L/(4\pi R^2 m_p c^3 \Gamma^2) = 4.17 \times 10^{15}$ cm^{-3}, radiation energy density $U'_\gamma = L_\gamma/(4\pi R^2 \Gamma^2 c) = 2 \times 10^{11}$ erg/cm^3 and $T'_e = \frac{1}{k_B}(\gamma_{e,ad} - 1)(\gamma_e - 1)m_e c^2 = 1.98 \times 10^9\,(\gamma_e - 1/\gamma_e)$ is the electron temperature for a Maxwellian distribution. Here, $\gamma_{e,ad} = (4\gamma_e + 1)/(3\gamma_e)$ is the adiabatic index of the electrons and β_p is the speed of protons divided by the speed of light. Substituting $\gamma_p \sim 1.123$ for $\tau_{in} = 8$ and $\gamma_e = \gamma_{e,crit} = 1.352$, we obtain $\eta = 4.5$.[15]

The electrons cannot be continuously heated by Coulomb collisions if the protons cool down to energies comparable to that of electrons within the dynamical timescale. While the protons cool down due to Coulomb collisions and adiabatic expansion, the electrons gain energy through Coulomb and get cooled due to adiabatic cooling and IC. The electrons cannot be heated any further when

$$\begin{aligned} &N_p(\gamma_p - 1)m_p c^2 - N_p \int_0^{t'} \frac{5 \times 10^{-19} n'_e \beta_p^2}{[0.73(\gamma_e - 1/\gamma_e)^{3/2} + \beta_p^3]} \eta dt' - N_p \int_0^{t'} \frac{(\gamma_p - 1)m_p c^2}{R/\Gamma c} dt' \\ &= N_e(\gamma_e - 1)m_e c^2 + N_e \int_0^{t'} \frac{5 \times 10^{-19} n'_e \beta_p^2}{[0.73(\gamma_e - 1/\gamma_e)^{3/2} + \beta_p^3]} \eta dt' \\ &\quad - N_e \int_0^{t'} \frac{(\gamma_e - 1)m_e c^2}{R/\Gamma c} dt' - N_e \int_0^{t'} (4/3) U'_\gamma \sigma_T (\gamma_e^2 - 1) c dt'. \end{aligned} \tag{3}$$

which simplifies for a neutral jet to, $1.5 \times 10^{-3}(1 - \ln\lambda) + 3.21 \times 10^{-5}(1 - 1/\lambda) = 1.15 \times 10^{-4}\eta(1 - 1/\lambda)\tau_{in}$. For $\lambda \sim 2$ i.e. for protons to cool down to electron energies in $t' = 2t'_{dyn}$, $\eta\tau_{in} \sim 8.28$. This means that for $\tau_{in} \gtrsim 10$, the protons cool down too fast and super-Coulomb interaction cannot keep the electrons hot beyond $t' = 2t'_{dyn}$. This necessitates the heating of electrons by some alternate subphotospheric dissipation mechanism. Even though the electrons tend to cool down rapidly due to Comptonization, their energy can still be maintained at $\gamma_e \gtrsim \gamma_{e,crit}$ provided the episodic heating events are frequent enough.

2.2. *Electron heating by repeated dissipation events*

As the total energy gained by the photons is the total energy that is transferred by the electrons through Comptonisation

$$N_\gamma \tau_{in}^{2/3} \left(E'_{\gamma,avg,obs} - E'_{\gamma,avg,i}\right) = N_e \int_0^{\tau_{in} t'_{dyn}} (4/3) U'_\gamma \tau_{in}^{-2/3} \sigma_T \gamma_e^2 \beta_e^2 c\, dt'.$$

which simplifies to

$$10^5\tau_{in}^{2/3}E'_{\gamma,avg,obs} = 8.72\times10^6 L_{\gamma,50}\tau_{in}^{-2/3}\int_{R_{in}}^{\tau_{in}R_{in}}(\gamma_e^2-1)\frac{dR}{R^2}$$
$$\sim \frac{8.72\times10^6 L_{\gamma,50}\tau_{in}^{-2/3}(\gamma_e^2-1)}{R_{in}}.$$

Substituting $R_{in}\sim 2.17\times10^{11}\ \tau_{in}^{-1}$ cm, yields $\gamma_{e,crit}\sim\sqrt{1+1.2\ L_{\gamma,50}^{-1}\ \tau_{in}^{1/3}}$. The critical electron energy obtained here is similar to $\gamma_{e,crit}\sim 1.352$ obtained earlier, especially for small initial optical depths $\tau_{in}\sim 1$.

We consider repeated sub-photospheric dissipation events that can re-accelerate the electrons as well as protons to their initial energies.

Using the fact that the electrons remain in equilibrium with energy $\gamma_{e,\rm crit}\sim\sqrt{1+1.2\ L_{\gamma,50}^{-1}\ \tau_{\rm n}^{1/3}}$, we constrain the energy injected per electron $E_{inj}=N_{rh}(\gamma_{e,in}-1)m_ec^2$. In particular,

$$N_e\tau_{in}^{4/3}(\gamma_e-1)m_ec^2 = N_e\int_0^{\tau_{in}t'_{dyn}}\frac{5\times10^{-19}n'_e\beta_p^2\tau_{in}^{-4/3}}{[0.73(\gamma_e-1/\gamma_e)^{3/2}+\beta_p^3]}dt'$$
$$+N_eN_{rh}(\gamma_{e,in}-1)m_ec^2 - N_e\int_0^{\tau_{in}t'_{dyn}}(4/3)U'_\gamma\tau_{in}^{-2/3}\sigma_T\gamma_e^2\beta_e^2c\ dt', \qquad (4)$$

where $\gamma_e\approx\gamma_{e,crit}$ and $\tau_{in}^{-4/3}/\tau_{in}^{-2/3}$ is the adiabatic cooling factor for relativistic electrons/photons. We consider the episodic dissipation events to be equally spaced over the jet expansion timescale $\tau_{in}t'_{dyn}$ and to supply fixed energy (equal to initial energy, $\gamma_{e,in}/\gamma_{p,in}$) to the electrons/protons at each instance. Substituting $R_{\rm in}=2.17\times10^{11}\ \tau_{\rm in}^{-1}$ cm and simplifying yields

$$\tau_{in}^{4/3}(\gamma_e-1)m_ec^2 = \frac{10^{-4}\tau_{in}^{-4/3}}{[0.73(\gamma_e-1/\gamma_e)^{3/2}+(2/\tau_{in})^{3/2}]}+E_{\rm inj,cr}-3.21\times10^{-5}\tau_{in}^{-2/3}, \qquad (5)$$

which constrains the critical injected energy $E_{inj,cr}(\tau_{in})=N_{rh}(\gamma_{e,in}-1)m_ec^2$ per electron in terms of τ_{in}. Equation (5) is only a necessary and not sufficient condition to obtain Band-like photon spectrum as it determines the average photon energy but does not impose any constraints on the general shape of the photon spectrum. For large E_{inj}, photon peak energy $E_{\gamma,peak}\gg E_{\gamma,obs}\sim 300$ keV while $E_{\gamma,peak}\ll E_{\gamma,obs}$ for large τ_{in}, due to significant energy loss from adiabatic cooling.

3. MCRaT code description

We discuss here the implementation of our MCRaT code and give an overview of the basic physics included. We first list the jet parameters and describe the initial distributions of the particles and the photons. We then discuss how the jet particles and photons are affected by the physical processes. Lastly, we describe the algorithm of our photospheric MCRaT code.

3.1. *Relativistic jet parameters*

The jet parameters provided as input for the code are:

- *Isotropic equivalent luminosity, L*: The bulk of the jet luminosity is contributed by the protons. We consider $L = 10^{51}, 10^{52}$ erg/s.[20]
- *Jet bulk Lorentz factor,* Γ: The bulk Lorentz factor is related to L and $E_{\gamma,peak}$. We consider $\Gamma = 30, 100, 300$ in this work.[21]
- *Initial optical depth,* $\tau_{\rm in}$: The optical depth is measured relative to $R = L\sigma_T/(8\pi m_p c^3 \beta \Gamma^3 \tau)$ which is the photon radial distance from central engine in the observer frame. $\tau_{\rm in}$ corresponds to the radial distance from central engine where all particles and photons are injected. We consider $\tau_{\rm in} = 10, 20, 40$.

3.2. *Initial distributions*

We describe here the initial energy and velocity distributions of the jet electrons, protons and photons.

- *Electrons and protons:* We consider a charge-neutral jet with particle number $N_e = N_p = 2 \times 10^2$. The initial particle velocities are distributed randomly in the jet-comoving frame. All particles are uniformly distributed in the jet-comoving frame at initial time. The initial energy of the electrons are determined from the Maxwellian distribution with temperature $k_B T'_{e,in} = (\gamma_{e,ad,in} - 1)(\gamma_{e,in} - 1) m_e c^2$ while the protons are mono-energetic with $\gamma_p = \gamma_{p,in}$. We consider $\gamma_{e,in} = 2, 10, 30, 100$ and $\gamma_{p,in} = 1.01, 1.1$ for our simulations.
- *Photons:* In order to maintain $N_\gamma/N_e = 10^5$, we consider $N_\gamma = 2 \times 10^7$ for our simulations.[15] The initial photon velocities are randomly distributed in the jet-comoving frame and photon positions are uniformly distributed within a cone of solid angle $1/\Gamma$ pointing towards the observer. The initial photon energy distribution is given by the synchrotron distribution for fast cooling electrons[5]

$$f_\nu = \begin{cases} \left(\frac{\nu_{ac}}{\nu_{sa}}\right)^{11/8} \left(\frac{\nu}{\nu_{ac}}\right)^2, & \nu_{min} < \nu < \nu_{ac} \\ \left(\frac{\nu}{\nu_{sa}}\right)^{11/8}, & \nu_{ac} < \nu < \nu_{sa} \\ \left(\frac{\nu}{\nu_{sa}}\right)^{-1/2}, & \nu_{sa} < \nu < \nu_m \\ \left(\frac{\nu_m}{\nu_{sa}}\right)^{-1/2} \left(\frac{\nu}{\nu_m}\right)^{-p/2}, & \nu_m < \nu < \nu_{max} \end{cases} \tag{6}$$

  where f_ν is the peak normalised photon flux per unit frequency and $p = 2.5$ is the spectral index at high energies.[6]

3.3. *Physical processes in the jet*

Here we discuss the physical interactions between the electrons, protons and photons that can further affect the output photon spectrum.

- *Adiabatic cooling:* As the relativistic jet expands outward, the photon and particle energies drop considerably as

$$(\gamma_{e,f}-1)/(\gamma_{e,i}-1) = \left(R_{\rm in}+\beta c\Gamma t'_f/R_{\rm in}+\beta c\Gamma t'_i\right)^{-2(\gamma_{\rm ad,e}-1)},$$
$$(\gamma_{p,f}-1)/(\gamma_{p,i}-1) = \left(R_{\rm in}+\beta c\Gamma t'_f/R_{\rm in}+\beta c\Gamma t'_i\right)^{-2(\gamma_{\rm ad,p}-1)},$$
$$E_{\gamma,f}/E_{\gamma,i} = \left(R_{\rm in}+\beta c\Gamma t'_f/R_{\rm in}+\beta c\Gamma t'_i\right)^{-2/3}, \quad (7)$$

  where the subscript i/f denotes the initial/final value of the physical quantity and $\gamma_{ad,e/p} = (4\gamma_{e/p}+1)/(3\gamma_{e/p})$ is the electron/proton adiabatic index.
- *Coulomb collisions:* While the electrons are continuously heated by the protons in the jet, they also interact with each other to quickly attain thermal equilibrium. The corresponding timescales are given by[22]

$$\dot{E}_{e-p} = \frac{5\times10^{-19}n'_e\beta^2_{\rm p,avg}}{8.3\times10^{-15}T'^{3/2}_{\rm e,avg}+\beta^3_{\rm p,avg}},$$
$$\dot{E}_{e-e} = \frac{5\times10^{-19}n'_e\beta^2_{\rm e,avg}}{8.3\times10^{-15}T'^{3/2}_{\rm e,avg}+\beta^3_{\rm e,avg}}, \quad (8)$$

  where $\beta_{p,avg}$, $\beta_{e,avg}$ and $T'_{e,avg}$ are number-averaged quantities. The electron distribution is re-initialized to Maxwellian distribution on a timescale $t'_{e-e} = (\gamma_{e,avg}-1)m_ec^2/\dot{E}_{e-e} \ll t'_{e-p}$.
- *IC scattering:* The distance s' that a photon travels before scattering an electron is given by probability density $p(s') \propto \exp(-s'/l'_{\rm mfp})$, where $l'_{\rm mfp} = 1/(n'_e\sigma_T)$ is the photon mean free path. The scattering probability of a particular electron with a photon is[15]

$$P_{sc}(\beta_e,\theta'_e) = \frac{1}{4\pi\beta_e^2}(1-\beta_e\cos\,\theta'_e), \quad (9)$$

  where β_e is the electron speed divided by speed of light and θ'_e is the angle between electron and photon velocities before scattering. Average number of scatterings that a photon experiences before escaping is $\sim 2\tau_{in}$.[23]
- *Pair production/annihilation:* Due to the episodic jet dissipation events, the electrons are often accelerated to highly relativistic energies $\gamma_e = \gamma_{e,in} \sim 100$ and can scatter energetic photons with $E'_\gamma \gtrsim 10\ E'_{\gamma,peak}$ to energies $\gtrsim 4E'_{\gamma,avg}\gamma_e^2 \sim m_ec^2 \sim 5\times10^5$ eV, before cooling down rapidly to non-relativistic energies. For synchrotron photons that we consider (see equation 6), a considerable fraction $\sim 30\%$ have sufficient energy to generate electron-positron pairs which can affect the shape of the output photon spectrum, especially for large $\tau_{in} \gtrsim 10$.

3.4. *Code implementation*

The travel distances are first drawn for all photons depending on their mean free path and the photons are propagated. The new photon positions are evaluated to check if any photon escapes the photosphere, in which case the energy is calculated

and stored. Other photons are stored in a priority queue where they are ordered based on travel distances. Next, the photon at the top of the queue is propagated, a proton is randomly selected and an electron is selected using the scattering probability, P_{sc}. The particle and photon energies are then updated due to adiabatic cooling and Coulomb collisions. The outgoing velocities and energies of the photon and electron are calculated if IC scattering occurs. Subsequently, the next photon in the queue is drawn and electron-positron pair production cross section is evaluated. If the cross section is large, a new electron and positron are generated and photons are not placed back in the queue. If positron number is non-zero, a positron is drawn randomly and the pair annihilation cross section with the electron is calculated. Two new photons are created and added to the queue, if the cross section is significant. The method described above is repeated until a third of the total photons in the jet escape and a time-averaged output photon spectrum is obtained.

4. Photospheric simulation results

Here we present the results of our photospheric MCRaT simulations. The photon energy spectrum and the electron kinetic energy spectrum are shown in the lab frame at the end of each simulation in all the figures. The photon and electron energy spectrum are Doppler boosted from the jet-comoving frame to the lab frame in all the figures. The electron kinetic energy spectra are peaked at significantly larger energies compared to the photon spectra for all the simulations as shown in the figures. In the rest of this paper, we denote the low/high energy photon spectral index by α/β and the observed photon peak energy by $E_{\gamma,obs}$.

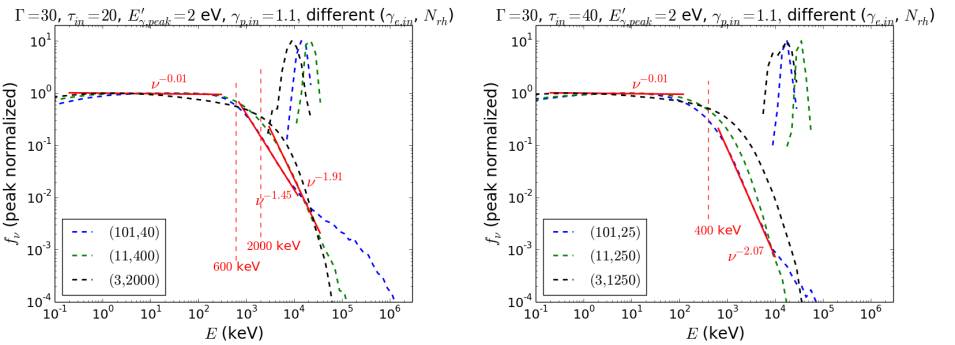


Fig. 1. **Effect of $\gamma_{e,\rm in}$ at constant $E_{\rm inj} = E_{\rm inj,cr}(\tau_{\rm in})$ and for different $\tau_{\rm in}$.** The input parameters used are $E'_{\gamma,\rm peak} = 2$ eV, $\gamma_{p,\rm in} = 1.1$, $L = 10^{52}$ erg/s and $\Gamma = 30$. *Left panel:* $\tau_{\rm in} = 20$, $E_{\rm inj} = 4000\ m_e c^2$ and $(N_{\rm rh}, \gamma_{e,\rm in}) = (40, 101)/(400, 11)/(2000, 3)$. *Right panel:* $\tau_{\rm in} = 40$, $E_{\rm inj} = 2500\ m_e c^2$ and $(N_{\rm rh}, \gamma_{e,\rm in}) = (25, 101)/(250, 11)/(1250, 3)$.

In Figure 1, we present the simulation results for fixed injected energy $E_{\rm inj,cr} = N_{\rm rh,cr}(\gamma_{e,in} - 1)\ m_e c^2 = 4000/2500\ {\rm m_e c^2}$ at $\tau_{in} = 20/40$ and different $\gamma_{e,in} = 3, 11, 101$. The photons/protons are initialized with $E'_{\gamma,peak} = 2$ eV/$\gamma_{p,in} = 1.1$

with jet parameters, $L = 10^{52}$ erg/s and $\Gamma = 30$. We find that $\alpha \sim 0$ is practically unaffected by decrease in electron energy $\gamma_{e,in}$ (irrespective of τ_{in}) and is solely determined by the critical injected energy $E_{inj,cr}$. As $\gamma_{e,in}$ increases for a given $E_{inj,cr}$, the photons tend to have lower peak energy $E_{\gamma,peak}$ and there are fewer/more photons with $E_\gamma \sim 1-10$ MeV/$\gtrsim 100$ MeV. This is expected as the electrons with $\gamma_{e,in} = 101$ are accelerated much less frequently compared to those with $\gamma_{e,in} \sim 3-11$ and then subsequently cool down very rapidly to non-relativistic γ_e after being considerably hotter for a shorter duration $\sim 10^{-3}\ t_{dyn}$ when they accelerate many photons to $E_\gamma \gtrsim 100$ MeV. The high energy bump in f_ν and deviation from power-law behaviour for large $\gamma_{e,in}$ is seen only at moderate $\tau_{in} \lesssim 20$ and is not appreciable for larger $\tau_{in} \gtrsim 40$ as the high energy photons cool down rapidly from adiabatic losses.

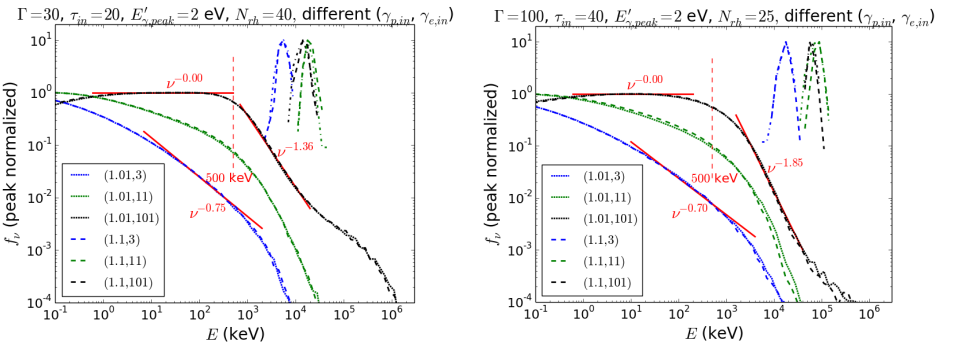


Fig. 2. **Effect of $\gamma_{p,in}$ and $\gamma_{e,in}$ for $N_{rh} = 40$ ($\tau_{in} = 20, \Gamma = 30$) and $N_{rh} = 25$ ($\tau_{in} = 40, \Gamma = 100$).** The input parameters used are $E'_{\gamma,peak} = 2$ eV and $L = 10^{51}$ erg/s. *Left panel:* $\Gamma = 30$, $\tau_{in} = 20$, $N_{rh} = 40$: $\gamma_{p,in} = (1.01, 1.1)$ and $\gamma_{e,in} = (3, 11, 101)$. *Right panel:* $\Gamma = 100$, $\tau_{in} = 40$, $N_{rh} = 25$: $\gamma_{p,in} = (1.01, 1.1)$ and $\gamma_{e,in} = (3, 11, 101)$.

We find an increase in $|\beta|$ with decrease in $\gamma_{e,in}$ for a fixed $E_{inj,cr}(\tau_{in})$ as well as with increase in τ_{in}. Moreover, $\beta \sim \beta_{obs}$ for $\gamma_{e,in} \sim$ few 10s $-$ 100 and $\tau_{in} \lesssim 20$ while the high-energy spectrum is much steeper, $f_\nu \propto \nu^{-2}$ for $\tau_{in} \gtrsim 40$, almost independent of N_{rh}. While relatively continuous energy injection (small $\gamma_{e,in} \sim$ few and large $N_{rh,cr} \sim$ few 1000s) results in steeper high energy spectra $|\beta| > |\beta_{obs}|$ along with $E_{\gamma,peak}/E_{\gamma,obs} \gtrsim 10$, episodic energy injection (large $\gamma_{e,in} \sim 100$ and small $N_{rh,cr} \sim$ few 10s) gives a high energy power-law spectrum consistent with observations for moderate optical depths $\tau_{in} \lesssim 20$. In order to have both $E_{\gamma,peak} \sim 500$ keV and $|\beta| \sim 1.2-1.5$, the particles and photons have to be initialized at $\tau_{in} \sim 20-40$ and $E_{inj,cr}(\tau_{in}) \sim 2500-4000\ m_e c^2$ energy needs to be injected into electrons with $\gamma_{e,in} \sim$ few 10s.

In Figure 2, we present the simulation results for fixed $N_{\rm rh,cr}(\tau_{\rm in}) = 40/25$ at $\tau_{\rm in} = 20$ ($\Gamma = 30$)/40 ($\Gamma = 100$) with different combinations of $\gamma_{e,\rm in} = 3, 11, 101$ and $\gamma_{p,\rm in} = 1.01, 1.1$. The seed photons have energy $E'_{\gamma,\rm peak} = 2$ eV with jet luminosity

$L = 10^{51}$ erg/s. We see that $\gamma_{p,\rm in}$ does not affect the photon output spectra irrespective of the optical depth, which is expected as the Coulomb collisions timescale $t'_{e-p} = (\gamma_{e,avg} - 1)m_e c^2/\dot{E}_{e-p}$ is considerably longer than the Comptonization timescale t'_{IC}. A minimum electron energy $\gamma_e \gtrsim 11$ is needed in order to have photons with $E_\gamma \gtrsim 10$ MeV and peak energy $E_{\gamma,peak} \sim 1$ MeV for both τ_{in} considered. The output photon spectrum does not show a power-law dependence at both low and high energies when the electron initial energy is small $\gamma_{e,in} \lesssim 11$. While the output photon spectrum shows $\alpha \sim \alpha_{obs}$ and $E_{\gamma,peak} \sim E_{\gamma,obs}$ at both optical depths for electrons with $\gamma_{e,in} = 101$ only, the high energy power-law spectral index $|\beta| \gg |\beta_{obs}|$ for $\tau_{in} = 40$ and $\sim |\beta_{obs}|$ for $\tau_{in} = 20$.

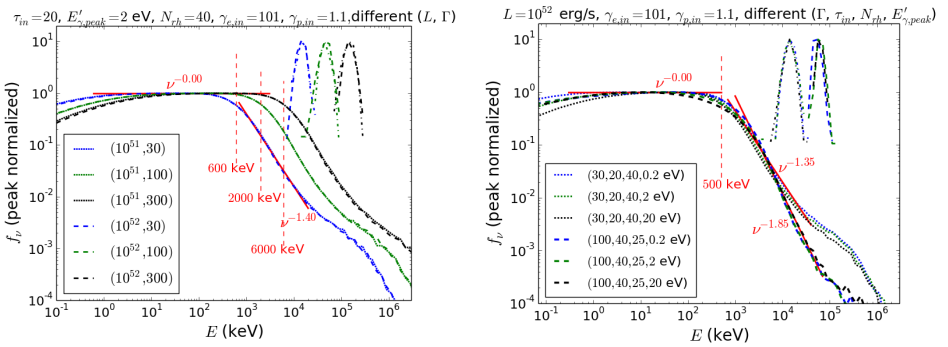


Fig. 3. *Left panel:* MCRaT simulations showing the effect of $L = (10^{51}, 10^{52})$ erg/s and $\Gamma = 30, 100, 300$ for constant $N_{rh} = 40$, $\gamma_{e,in} = 101$ and $\tau_{in} = 20$. For these simulations, we consider input parameters $E'_{\gamma,peak} = 2$ eV and $\gamma_{p,in} = 1.1$. *Right panel:* MCRaT simulations showing the effect of $E'_{\gamma,peak} = 0.2, 2, 20$ eV for $N_{rh} = 40$ ($\tau_{in} = 20, \Gamma = 30$) and $N_{rh} = 25$ ($\tau_{in} = 40, \Gamma = 100$). For these simulations, we consider input parameters $\gamma_{e,in} = 101$, $\gamma_{p,in} = 1.1$ and $L = 10^{52}$ erg/s.

In the left panel of Figure 3, we show the simulation results for fixed $E_{\rm inj,cr} = 4000\ m_e c^2$ at $\tau_{in} = 20$ and for different combinations of $L = 10^{51}, 10^{52}$ erg/s and $\Gamma = 30, 100, 300$. The photons/electrons/protons are initialized with energies $E'_{\gamma,peak} = 2$ eV/$\gamma_{e,in} = 101/\gamma_{p,in} = 1.1$ at optical depth $\tau_{in} = 20$. While the jet luminosity L has no noticeable effect on the output photon spectrum, increase in bulk Lorentz factor Γ shifts the photon peak energy to higher values. We find that even though Γ does not affect α and β, it rescales photon peak energy as $E_{\gamma,peak} \propto \Gamma$. The output photon spectrum shows $E_{\gamma,peak} \sim E_{\gamma,obs}$ only for smaller $\Gamma \sim 30$ values. While larger $\Gamma \sim 100$ can also reproduce $E_{\gamma,peak} \sim 500$ keV and $\alpha \sim 0$ at $\tau_{in} \gtrsim 40$ in agreement with the observations, it cannot explain the observed high energy spectral index (see right panel of Figure 2).

In the right panel of Figure 3, we show the simulation results for fixed $N_{rh,cr}(\tau_{in}) = 40/25$ at $\tau_{in} = 20$ ($\Gamma = 30$)/40 ($\Gamma = 100$) and different seed photon energies $E'_{\gamma,peak} = 0.2, 2, 20$ eV. The electrons/protons are initialized with energies $\gamma_{e,in} = 101/\gamma_{p,in} = 1.1$ with jet luminosity $L = 10^{52}$ erg/s. We find that for

both $E_{inj,cr}(\tau_{in}) = 2500\ m_e c^2$ and $4000\ m_e c^2$, the low/high energy spectral index α/β and the output photon peak energy $E_{\gamma,peak}$ are practically unaffected by the choice of $E'_{\gamma,peak}$. The specific photon flux f_ν falls off considerably at energies less than $\Gamma E'_{\gamma,peak}$ as most of the photons gain energy and do not populate the low energy tail after getting scattered by the electrons. For larger τ_{in}, the photons get scattered multiple times thereby increasing the probability of differential number of scatterings before escaping the photosphere and subsequent broadening of the spectrum. As a result, more photons populate the low energy tail and the spectra with different initial energies $\Gamma E'_{\gamma,peak}$ become indistinguishable for $\tau_{in} \gtrsim 40$.

5. Photon spectra for repeated scatterings

Assuming that Comptonization is the dominant process influencing the output photon spectrum, we will first evaluate the energy spectrum of synchrotron photons after they experience single scattering with the electrons. Then we extend our formalism to find the photon energy spectrum for the realistic case when they undergo repeated scatterings with the electrons in the jet before exiting the photosphere. The energy distribution of the scattered photons depends mainly on the incident photon spectrum and the electron energy distribution.

5.1. *Photon distribution after one scattering*

For our calculation, we consider electrons and incident photons with isotropic distributions in the jet-comoving frame. In this case, the scattered photons are also distributed isotropically in the jet-comoving frame. For simplicity, we only consider Thomson scattering in the rest frame of the electron and assume that all scattering events are elastic in nature. For incident photons with energy ϵ scattering off electrons with energy $\gamma m_e c^2$, the total scattered power per energy per volume is[24]

$$\frac{dE}{dV dt d\epsilon_1} = \frac{3}{4} c\sigma_T \int_{\epsilon_1/4\gamma^2}^{\infty} d\epsilon \frac{\epsilon_1}{\epsilon^2} f(\epsilon) \int_1^{\infty} \frac{d\gamma}{\gamma^2} n_e(\gamma) g_{\rm iso}\left(\frac{\epsilon_1}{4\gamma^2\epsilon}\right), \tag{10}$$

where, ϵ_1 is the scattered photon energy, $f(\epsilon)$ is the photon distribution function, $n_e(\gamma)$ is the electron distribution function and $g_{\rm iso}(x) = \frac{2}{3}(1-x)$ for isotropic photon distribution in the jet-comoving frame. We consider the simple case in which the incident photons have a synchrotron/piecewise power-law energy distribution,

$$f_{in}(\epsilon) = f_0 \begin{cases} (\epsilon/\epsilon_0)^a, & \epsilon < \epsilon_0 \\ (\epsilon/\epsilon_0)^{-b}, & \epsilon > \epsilon_0 \end{cases} \tag{11}$$

and the electrons are mono-energetic with $n_e(\gamma) = n_0 \delta(\gamma - \gamma_0)$. Substituting $x = 4\gamma_0^2 \epsilon/\epsilon_1$ for relativistic electrons yields

$$\frac{dE}{dV dt d\epsilon_1} = 2c\sigma_T n_0 \int_1^{\infty} \frac{dx}{x^2} \left(1 - \frac{1}{x}\right) f\left(\frac{\epsilon_1 x}{4\gamma_0^2}\right). \tag{12}$$

For photons below peak energy $\epsilon_1 < 4\gamma_0^2\epsilon_0$, we can further define $\epsilon_1/4\gamma_0^2 = \eta\epsilon_0$ with $\eta < 1$. For low energy photons with $\eta \ll 1$, if the incident photons have a hard spectrum with $0 < a < 1$, the scattered photon distribution $f_{sc}(\epsilon_1) \propto dE/(dVdtd\epsilon_1) \propto \eta^a \propto \epsilon_1^a/(\gamma_0^{2a}\epsilon_0^a)$ is the same as that of the incident photons. However, for a softer low energy incident photon spectrum with $a \geq 1$ and $\eta \ll 1$, we obtain $f_{sc}(\epsilon_1) \propto dE/(dVdtd\epsilon_1) \propto \epsilon_1$. Therefore, after single scattering of synchrotron photons with broken power-law energy distribution, the low energy spectrum is unaffected for hard spectra with $a < 1$ whereas $f_{sc}(\epsilon) \propto \epsilon$ for softer spectra. For photons above peak energy $\epsilon_1 > 4\gamma_0^2\epsilon_0$ and $\eta \gtrsim 1$. This gives $dE/(dVdtd\epsilon_1) \propto \epsilon_1^{-b}/(\gamma_0^{-2b}\epsilon_0^{-b})$, same as the incident photon spectrum. For fast cooled synchrotron photon spectrum, we have $a = 2$ and $b = -1$, and the scattered photon distribution after single scattering is

$$f_1(\epsilon) = f_{sc}(\epsilon) \propto f_0 \begin{cases} (\epsilon/\epsilon_0)^1, & \epsilon < \epsilon_0 \\ (\epsilon/\epsilon_0)^{-1}, & \epsilon > \epsilon_0. \end{cases} \tag{13}$$

In reality, however, each photon experiences $\sim 2\tau_{in}$ scatterings on an average before escaping the photosphere.

5.2. *Photon distribution after repeated scatterings*

With $f_1(\epsilon)$ as the incident photon distribution, we can now extend the same formalism to calculate the photon spectrum after subsequent scattering events assuming that the electron and photon distributions remain isotropic in the jet-comoving frame. The low and the high energy spectrum after each photon in the jet has undergone exactly two scatterings is

$$f_{2,l}(\epsilon_1) = 2c\sigma_T n_0 f_0 \left(-\eta \ln\eta + \frac{2}{3}\eta^2 - \frac{1}{2}\eta\right) \propto \frac{\epsilon_1}{4\gamma_0^2\epsilon_0} \ln\left(\frac{\epsilon_1}{4\gamma_0^2\epsilon_0}\right),$$

$$f_{2,u}(\epsilon_1) = 2c\sigma_T n_0 f_0 \int_1^\infty dx \frac{1}{x^2}\left(1 - \frac{1}{x}\right)\eta^{-1}x^{-1} \propto \gamma_0^2\epsilon_0/\epsilon_1.$$

The scattered photon spectrum is then

$$f_2(\epsilon) \propto f_0 \begin{cases} (\epsilon/\epsilon_0)\ln(\epsilon/\epsilon_0), & \epsilon < \epsilon_0 \\ (\epsilon/\epsilon_0)^{-1}, & \epsilon > \epsilon_0. \end{cases} \tag{14}$$

After each photon has undergone exactly three scatterings, the low and high energy spectra are given as

$$f_{3,l}(\epsilon_1) \propto \frac{\epsilon_1}{4\gamma_0^2\epsilon_0}\left[\ln\left(\frac{\epsilon_1}{4\gamma_0^2\epsilon_0}\right)\right]^2,$$

$$f_{3,u}(\epsilon_1) = 2c\sigma_T n_0 f_0 \int_1^\infty dx \frac{1}{x^2}\left(1 - \frac{1}{x}\right)\eta^{-1}x^{-1} \propto \gamma_0^2\epsilon_0/\epsilon_1,$$

and the scattered photon spectrum is

$$f_3(\epsilon) \propto f_0 \begin{cases} (\epsilon/\epsilon_0)\left[\ln(\epsilon/\epsilon_0)\right]^2, & \epsilon < \epsilon_0 \\ (\epsilon/\epsilon_0)^{-1}, & \epsilon > \epsilon_0 \end{cases} \tag{15}$$

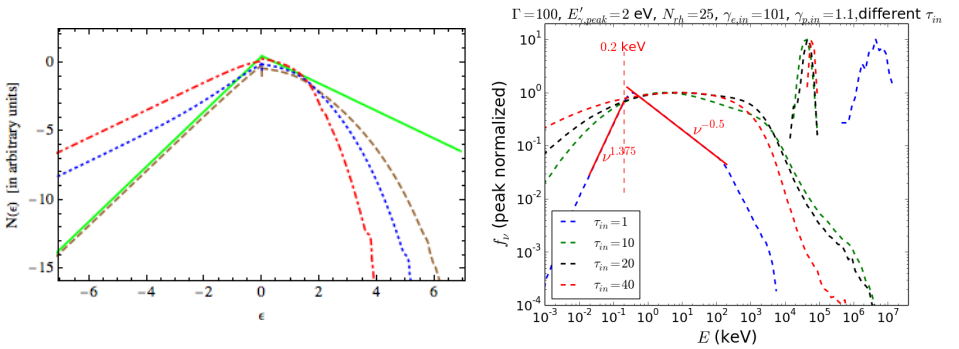


Fig. 4. **Effect of geometrical broadening on the photon spectrum for increasing optical depth.** *Left panel:* IC spectrum for synchrotrom photons with energy $\epsilon_0 = 1$ and Maxwellian electrons with peak energy $\gamma_{e,0} = 1.1$. The solid green line, brown dashed line, blue dotted line and red dot-dashed lines are the scattered photon spectra after $N = 0, 1, 2$ and 5 scatterings, respectively. *Right panel:* MCRaT simulations showing the effect of geometrical broadening on the photon spectrum at $\tau_{in} = 1, 10, 20, 40$ for $(N_{rh}, \gamma_{e,in}) = (25, 101)$ and $\gamma_{p,in} = 1.1$. For these simulations, we consider input parameters $L = 10^{52}$ erg/s, $E'_{\gamma,peak} = 2$ eV and $\Gamma = 100$.

We can generalize the above results further for $N \sim 2\tau_{in}$ scatterings per photon

$$f_N(\epsilon) \propto f_0 \begin{cases} (\epsilon/4\gamma_0^2\epsilon_0)\left[\ln(\epsilon/4\gamma_0^2\epsilon_0)\right]^{N-1}, & \epsilon < \epsilon_0 \\ (\epsilon/4\gamma_0^2\epsilon_0)^{-1}, & \epsilon > \epsilon_0 \end{cases} \tag{16}$$

In Figure 4, we show how the photon spectrum is affected by Comptonization with electrons as τ_{in} and number of scatterings increase. In the left panel, the IC scattered photon spectrum for fast cooled synchrotron seed photons (Equation 11, with $a = 2$ and $b = -1$) with energy $\epsilon_0 = m_e c^2$ and Maxwellian electrons with peak energy $\gamma_{e,0} = 1.1$ is shown for scattering orders $N = 0, 1, 2, 5$. As predicted by equation (16), the photon spectrum becomes gradually softer below peak energy as the scattering order increases. The photons scattering off Maxwellian electrons with fixed energy get thermalized at equilibrium to attain a high-energy exponential tail for large optical depths/scatterings. Furthermore, there is gradual flattening of the low-energy spectrum with increase in scattering order N.

In the right panel of Figure 4, we present the MCRaT simulation results for $E_{\rm inj,cr} = 2500\ m_e c^2$ and different optical depths $\tau_{in} = 1, 10, 20, 40$. The number of repeated dissipation events in the jet are $N_{rh} = 25$ with initial photon/electron/proton energy $E'_{\gamma,peak} = 2$ eV/$\gamma_{e,in} = 101$/$\gamma_{p,in} = 1.1$ and jet parameters $L = 10^{52}$ erg/s and $\Gamma = 100$. With increase in scattering order ($\propto \tau_{in}$), the high energy photon spectrum becomes steeper with a simultaneous decrease in $E_{\gamma,peak}$. These photons then populate the low energy spectrum and extend the non-thermal tail to energies much lower than $E_{\gamma,peak} \sim 0.2$ keV. The photon spectra from simulations are also considerably broader compared to the analytical results for similar values of N.

6. Summary and conclusions

We have utilized our MCRaT photospheric code to explain the distinct non-thermal behaviour of GRB prompt emission spectrum, $f_\nu \propto \nu^0 / f_\nu \propto \nu^{-1.2}$ at low/high photon energies along with observed peak energy at $E_{\gamma,peak} \sim 300$ MeV. For our simulations, we considered Comptonization of fast cooled synchrotron photons with Maxwellian electrons and for photon to particle number ratio $N_\gamma/N_e \sim 10^5$. The jet electrons are accelerated by two different mechanisms: 1. continuous energy transfer via Coulomb collisions with mono-energetic protons, 2. repeated episodic energy dissipation events that are equally spaced over time and accelerate particles back to their initial energies.

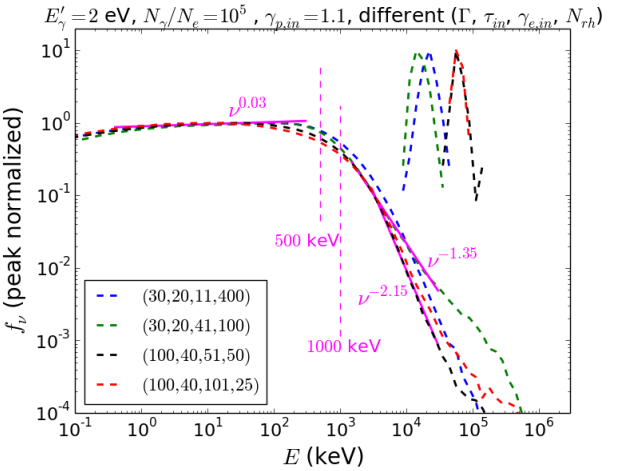


Fig. 5. **MCRaT simulation results with the best set of parameters for a jet with $N_\gamma/N_e = 10^5$.** The relativistic jet with $L = 10^{52}$ erg/s has photons with $E'_{\gamma,\rm peak} = 2$ eV and protons with $\gamma_{p,\rm in} = 1.1$. The energy injection necessary in order to produce an output photon spectrum with the observed Band-like spectral properties depends on $\tau_{\rm in}$ and Γ. Here we consider $E_{\rm inj} = 4000/2500\ m_e c^2$ for $\Gamma = 30/100$ and $\tau_{\rm in} = 20/40$, for two distinct electron energies $\gamma_{e,\rm in} = (11, 41)/(51, 101)$.

The parameters that significantly affect the spectral properties for a given N_γ/N_e are $E_{\rm inj}(\gamma_{e,\rm in}, N_{\rm rh})$, Γ and $\tau_{\rm in}$. Figure 5 shows the most probable parameter set that gives output photon spectrum with $(\alpha, \beta, E_{\gamma,\rm peak})$ very similar to the observed GRB prompt spectrum. The photons/protons in these simulations are initialized with energies $E'_{\gamma,\rm peak} = 2$ eV/$\gamma_{p,\rm in} = 1.1$ for jet parameters $L = 10^{52}$ erg/s, $N_\gamma/N_e = 10^5$ and $\Gamma \sim 30 - 100$. The particles are injected with energy $E_{\rm inj,cr} \sim 2500 - 4000\ m_e c^2$ per electron for a range of $\tau_{\rm in} \sim 20 - 40$. For smaller $\tau_{in} \sim 20$ and $\Gamma \sim 30$, $(\alpha, \beta, E_{\gamma,peak}) \sim (0, -1.4, 1$ MeV) is obtained with $E_{inj,cr} \sim 4000\ m_e c^2$ and $\gamma_{e,in} \gtrsim 40$. Although $\alpha \sim 0$ and $E_{\gamma,peak} \sim 500$ keV for $E_{inj,cr} \sim 2500\ m_e c^2$ at larger $\tau_{in} \sim 40$ and $\Gamma \sim 100$, the high energy spectrum is significantly steeper than the observed prompt spectrum with $\beta \sim -2.1$, especially for $\gamma_{e,in} \lesssim 50$.

The main results of this work can be summarised as:

(1) The electrons cool down rapidly to non-relativistic energies in the absence of jet dissipation events. This entails energy injection into the jet particles via either (continuous) Coulomb collisions or (episodic) sub-photospheric dissipation events. However, for Coulomb heating, the protons lose a considerable fraction of their energy within $\sim t_{dyn}$, for $\tau_{in} \gtrsim 10$, to attain non-relativistic energies. Therefore, continuous energy injection by protons is not sufficient to maintain electrons at $\gamma_e \sim \gamma_{e,crit}$.

(2) Energy injection can be achieved with episodic sub-photospheric dissipation events which can keep the electrons at energies $\gamma_e \gtrsim \gamma_{e,crit}$ provided that they are sufficiently energetic and frequent. For large $E_{\rm inj}$, the photon peak energy $E_{\gamma,\rm peak} \gg E_{\gamma,\rm obs} \sim 300$ keV, while for large $\tau_{\rm in}$, $E_{\gamma,\rm peak} \ll E_{\gamma,\rm obs}$ due to significant adiabatic loss. From MCRaT simulations, we quantify the $E_{inj} - \tau_{\rm in}$ correlation: injected energy $E_{inj,cr} = 4000/2500\ m_e c^2$ per electron for initial optical depth $\tau_{in} = 20/40$.

(3) In the output photon spectrum, α critically depends on $E_{\rm inj}$ whereas β and $E_{\gamma,\rm peak}$ are almost entirely determined by $\tau_{\rm in}$. With an increase in $\tau_{\rm in}$, $E_{\gamma,\rm peak}$ decreases and the high-energy photon spectrum becomes steeper. Additionally, $|\beta|$ also increases with decrease in initial electron energy $\gamma_{e,in}$ for fixed $E_{\rm inj,cr} = N_{rh,cr}(\gamma_{e,in} - 1)\ m_e c^2$. We find that $E_{\gamma,\rm peak} \sim E_{\gamma,\rm obs}$ only for smaller $\Gamma \sim 30$ - while larger $\Gamma \sim 100$ gives $E_{\gamma,peak} \sim 500$ keV at $\tau_{in} \sim 40$, the high energy photon spectrum is considerably steeper than observed.

(4) For isotropic electrons scattering isotropic photons, the scattered photon energy distribution can be analytically evaluated for lower order scatterings, given the electron and photon energy distributions. For Comptonization of synchrotron photons with Maxwellian electrons, $\alpha \sim 0$ behaviour is retained at low energies whereas $f_\nu \propto e^{-\nu}$ at high energies. Qualitatively, the low-energy non-thermal dependence is obtained from multiple scatterings and subsequent geometrical broadening of the spectrum whereas the high-energy power-law dependence is attributed to repeated episodic and continuous energy injection events.

References

1. D. Band, et al., *ApJ* **413**, 281 (1993).
2. R. D. Preece, M. S. Briggs, R. S. Mallozzi, G. N. Pendleton, W. S. Paciesas, D. L. Band, *ApJS* **126**, 19 (2000).
3. Y. Kaneko, R. D. Preece, M. S. Briggs, W. S. Paciesas, C. A. Meegan, D. L. Band, *ApJS* **166**, 298 (2006).
4. M. J. Rees, P. Meszaros, *ApJ* **430**, L93 (1994).
5. T. Piran, *Reviews of Modern Physics* **76**, 1143 (2004).
6. P. Kumar, B. Zhang, *Phys. Rep.* **561**, 1 (2015).
7. D. Giannios, *A&A* **457**, 763 (2006).
8. T. Piran, *Phys. Rep.* **314**, 575 (1999).
9. B. Zhang, et al., *ApJ* **655**, 989 (2007).

10. R. D. Preece, M. S. Briggs, R. S. Mallozzi, G. N. Pendleton, W. S. Paciesas, D. L. Band, *ApJ* **506**, L23 (1998).
11. G. Ghirlanda, A. Celotti, G. Ghisellini, *A&A* **406**, 879 (2003).
12. M. J. Rees, P. Meszaros, *ApJ* **628**, 847 (2005).
13. M. Bhattacharya, P. Kumar, *MNRAS* **491**, 4656 (2019).
14. D. Lazzati, M. C. Begelman, *ApJ* **725**, 1137 (2010).
15. M. Bhattacharya, W. Lu, P. Kumar, R. Santana, *ApJ* **852**, 1 (2018).
16. A. Chhotray, D. Lazzati, *ApJ* **802**, 132 (2015).
17. J. Granot, T. Piran, R. Sari, *ApJ* **534**, L163 (2000).
18. K. Toma, X. F. Wu, P. Meszaros, *MNRAS* **415**, 1663 (2011).
19. C. Thompson, *MNRAS* **270**, 480 (1994).
20. E. Liang, B. Zhang, F. Virgili, Z. G. Dai, *ApJ* **662**, 1111 (2007).
21. R. R. Xue, Y. Z. Fan, D. M. Wei, *A&A* **498**, 671 (2009).
22. R. Schlickeiser, Cosmic Ray Astrophysics. Springer, New York (2002).
23. D. Begue, I. A. Siutsou, G. V. Vereshchagin, *ApJ* **767**, 139 (2013).
24. G. B. Rybicki, A. P. Lightman, Radiative Processes in Astrophysics. A Wiley-Interscience publication, Wiley (1979).

Monte Carlo simulations of photospheric emission in Gamma Ray Bursts

T. M. Parsotan* and D. Lazzati
Department of Physics, Oregon State University, Corvallis, OR 97331, USA
**E-mail: parsotat@umbc.edu*

The study of Gamma Ray Bursts (GRBs) has the potential to improve our understanding of high energy astrophysical phenomena. In order to reliably use GRBs to this end, we first need to have a well-developed grasp of the mechanism that produces the radiation within GRB jets and how that relates to their structure. One model for the emission mechanism of GRBs invokes radiation produced deep in the jet which eventually escapes the jet at its photosphere. While this model has been able to explain a number of observed GRB characteristics, it is currently lacking in predictive power and in ability to fully reproduce GRB spectra. In order to address these shortcomings of the model, we have expanded the capabilities of the MCRaT code, a state of the art radiative transfer code that can now simulate optical to gamma ray radiation propagating in a hydrodynamically simulated GRB jet. Using the MCRaT code, we have constructed mock observed light curves, spectra, and polarization from optical to gamma ray energies for the simulated GRBs. Using these mock observables, we have compared our simulations of photospheric emission to observations and found much agreement between the two. Furthermore, the MCRaT calculations combined with the hydrodynamical simulations allow us to connect the mock observables to the structure of the simulated GRB jet in a way that was not previously possible. While there are a number of improvements that can be made to the analyses, the steps taken here begin to pave the way for us to fully understand the connection between the structure of a given GRB jet and the radiation that would be expected from it.

Keywords: Gamma-rays: Bursts; relativistic outflows; radiation transfer.

1. Introduction

Gamma Ray Bursts (GRBs) typically produce high energy X-ray and gamma-ray radiation that is detected within the first few tens of seconds of the event. The origin of this so called prompt emission is still not well understood even after decades of investigation. The two most common models that are used to describe the prompt emission of GRBs either employ synchrotron emission from jets with magnetic fields[1–4] or emission from the photospheres of GRB jets.[5–15]

Each model has its own advantanges and disadvantages allowing them to describe GRB observations at gamma-ray and X-ray energies in varying capacities. Models that rely on synchrotron emission are able to account for the non-thermal characteristics of GRB spectra[4,16] but are unable to fully account for various GRB observational relationships such as the Amati[17] and Yonetoku[18] relations.[1] The photospheric model on the other hand is able to reproduce the Amati and

Yonetoku relations[10,12,13,15,19,20] but has issues with reproducing typical GRB spectral parameters.[12,13] In order to fully distinguish between these models additional data at different wavelengths is needed.

There have been a number of optical prompt emission detections that can be used for the purpose of model comparison (see[21] and references therein). Additionally, there has been one detection of optical prompt polarization[22] that can also be used in comparison with optical polarization predictions obtained from synchrotron and the photospheric model calculations. The synchrotron model has traditionally been favored by these optical prompt detections[23–27] however the photospheric model has not been fully explored in its ability to reproduce these optical prompt detections.

The ability to simulate radiation signatures produced under the photospheric model has recently been improved.[11–15,21,28,29] These analyses have combined the realistic treatment of radiation with hydrodynamically simulated time-dependent GRB jets. As a result, studies of how the GRB jet structure affects the detected radiation signatures are now possible. With the ability of these state-of-the-art radiative transfer calculations to also calculate the effects of cyclo-synchrotron emission and absorption,[21,29] these types of analyses are also now possible from optical to gamma-ray energies. In this proceeding, we highlight the results of conducting radiative transfer calculations of hydrodynamically simulated GRB jets and the construction of mock observations from these calculations. We show how the mock observables can be related to the structure of the simulated GRB jets and how these analysis can help us interpret GRB observations under the photospheric model. Finally, we conclude with a summary and highlight ways that these types of global radiative transfer calculations can push the field forward.

2. Methods

2.1. *The MCRaT Code*

We have used the open-source **M**onte **C**arlo **Ra**diation **T**ransfer (MCRaT) code[30] to analyze the evolution of radiation trapped within hydrodynamical simulations of GRB jets. The MCRaT code considers photons in the astrophysical outflow that interact via Compton scattering including the effects of the full Klein-Nishina cross section with polarization. The code also considers photons that have been absorbed and emitted in the outflow due to cyclo-synchrotron (CS) emission and absorption. We summarize the algorithm here, however detailed descriptions of the code are outlined in[21] and.[29] As is outlined in Figure 1, the code loads a SRHD simulation frame and injects photons into the simulated GRB jet where the optical depth $\tau > 100$. The code will calculate the mean free paths of each photon in the simulation in order to determine which will scatter. After this calculation, MCRaT advances the simulation time to correspond to when the scattering will occur and propagates each photon by a distance corresponding to the amount of time passing in the simulation. The code then conducts the scattering considering the full Klein-Nishina

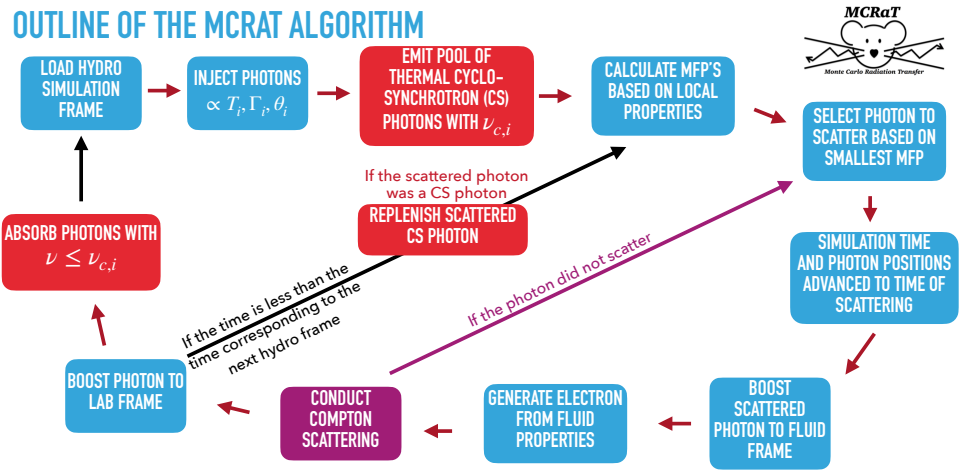


Fig. 1. The outline of the MCRaT code, see text for in depth explanation.

cross section with polarization, which means that the photons have some probability of not scattering in the electron rest frame. If the chosen photon ends up not being scattered, MCRaT chooses the next photon that would have scattered and advances the simulation time and the location of each photon. If the photon is scattered, then the new four momentum and polarization of the photon is calculated. The code continues to scatter photons until the time in MCRaT corresponds to the time in the next SRHD simulation frame. If CS emission and absorption is also taken into account in the simulation then the steps in the red boxes are taken which numerically simulate CS emission and absorption processes in MCRaT.

2.2. *Mock Observations*

With a given MCRaT simulation completed, we use the open-source ProcessMCRaT code[31] to produce mock observed light curves, spectra, and polarization measurements that can be compared to actual GRB observations on a more equal basis. This code allows photons from an MCRaT simulation to be binned in time and energy in order to produce the mock observables, similar to how actual photons measured in a given GRB event are binned. Detailed descriptions of how we construct the mock observable quantities for observers at various viewing angles with respect the jet axis, $\theta_{\rm v}$, can be found in.[12–14,21] We use the same gamma-ray and optical energy ranges as in.[29] The gamma-ray mock observables are calculated from photons with energies between 20-800 keV which corresponds to the polarimetry energy range of POLAR-2.[32] The optical mock observables are calculated from photons with energies between 1597-7820 $\mathring{A}$ ($\sim 1.5 - 7.7$ eV), which aligns with the Swift UVOT White bandpass.[33,34]

The way that we connect the mock observables to the structure of the simulated GRB jets is through equal arrival time surfaces (EATS)[7,35,36] of the photons in the GRB jet. For a given time bin, in the mock observed light curve, we can calculate surfaces that would be emitting photons towards a given observer's line of sight. These surfaces are calculated following the method of,[14] who describe the procedure for constructing these EATS.

2.3. *The Simulation Set*

The MCRaT code was used to analyze a set of two special relativistic hydrodynamical simulations (SRHD) of GRB jets. These SRHD simulations were conducted using the FLASH hydrodynamics code[37] and their properties are outlined in.[14,21,29] These two SRHD simulations are referred to as the *16TI* and *40sp_down* simulations and they represent GRB jets that have a time-constant and time-variable injection of energy, respectively.

3. Results

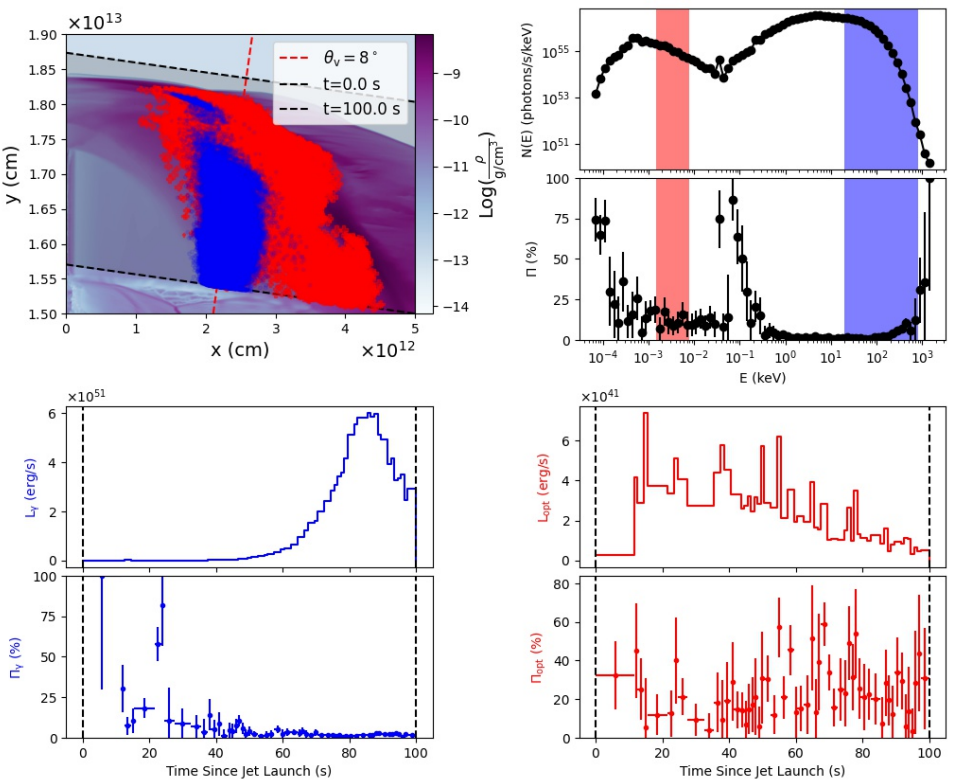


Fig. 2. The relation between the mock observed quantities and the jet structure of the *16TI* simulation for $\theta_v = 8°$. See text for full description.

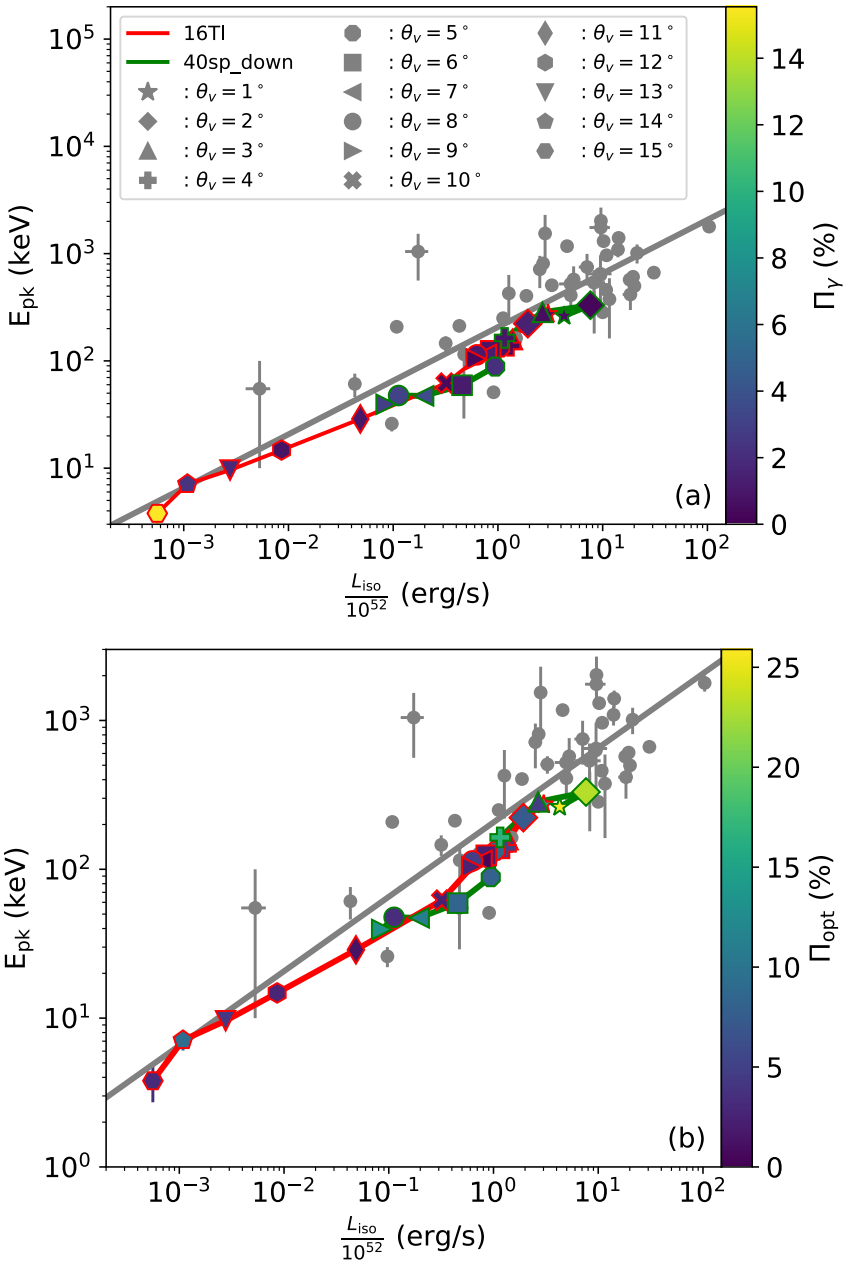


Fig. 3. The *16TI* and *40sp_down* simulations plotted alongside the Yonetoku relationship[18] in red and green lines and marker outlines, respectively. In Figure (a), the fill color of the marker denotes the mock observed time-integrated Π_γ while in Figure (b) the fill color shows the time-integrated MCRaT Π_{opt}. The different marker shapes for the MCRaT simulations show the placement of the simulations on the Yonetoku relationship as determined by observers at various θ_v. In each panel the observational relationship is shown as the grey solid line and observed data from[38] are plotted as grey markers. Figure adapted from.[29]

Our MCRaT simulations have shown that optical and gamma ray photons probe different regions of GRB outflows under the photospheric model, as is shown in Figure 2. Here, we show all the photons that have been detected by an observer located at $\theta_{\rm v} = 8^\circ$ in optical and gamma-ray energies. The top left panel shows the location of each photon in relation to the density structure of the SRHD simulated GRB jet. The density of the jet is shown as a pseudocolor plot where darker colors show denser regions and the dashed red line denoted the line of sight of the observer at $\theta_{\rm v} = 8^\circ$. The optical photons are shown in red while gamma-ray photons are depicted in blue. Furthermore, the photon markers are translucent allowing us to identify regions of the jet where the photons are densely located (due to the concentration of blue or red). Additionally, the markers are different sizes in order to show the weight of each photon in the calculation of the various mock observable quantities. The top right panel shows the time integrated spectrum of these photons with highlighted red and blue regions that highlight the energy ranges which were used to calculate the optical and gamma-ray light curves (L_γ and $L_{\rm opt}$) and polarizations (Π_γ and $\Pi_{\rm opt}$). These time-resolved mock observables are shown in the bottom 4 panels. These analysis allow us to connect the mock observables of light curves, spectra, and polarizations with the locations of where the photons are within the jet and what the jet structure is at those locations. As is shown in Figure 2, the optical photons are distrubuted widely about the observer's line of sight with their locations being focused at the dense jet-cocoon interface (JCI;[39]). The gamma-ray photons instead are mostly concentrated along the observer's line of sight with the exception of some gamma-ray photons being located near the core of the jet. Thus, optical prompt emission probes the dense JCI while the gamma-ray emission probes the portion of the jet that is directly along the observer's line of sight.

The results of our MCRaT simulations can also be used to assess how well the simulations reproduce various observational relationships, such as the Yonetoku relation,[18] and make additional predictions that can be tested against data. Figure 3 shows the comparison between the *16TI* and *40sp_down* simulations, in red and green marker edge colors and lines respectively, and the Yonetoku relation. Various observed GRBs from[38] are also shown as grey circles. We also show the time-integrated Π_γ and $\Pi_{\rm opt}$ in Figure 3(a) and (b) respectively, through the fill-color of each marker. In each panel, we see that both the *40sp_down* and *16TI* simulations are able to recover the normalization and slope of the Yonetoku relation for a variety of $\theta_{\rm v}$. With this success of the photospheric model, we can also now use the simulated polarization mock observations to make predictions. We find that the time-integrated Π_γ should increase as $\theta_{\rm v}$ increases which has been found in other studies.[40,41] Thus, bursts that are bright and energetic (in the top right of the Yonetoku relation) should have low time-integrated Π_γ. On the other hand, the time-integrated $\Pi_{\rm opt}$ should be larger for $\theta_{\rm v}$ that are close to the jet axis. As a result, we would expect that bursts located in the top right of the Yonetoku relation would have large time-integrated $\Pi_{\rm opt}$.

4. Conclusion

Using the MCRaT radiative transfer code, we have conducted post-processing analysis of the radiation expected from time-dependent special relativistic hydrodynamic simulations of GRB jets. These global radiative transfer simulations allow for the connection between observed radiation signatures and the structure of the GRB jet, which was not possible prior to these types of analysis. These analyses are commonplace in other areas of astrophysics such as galaxy formation and have proved fruitful in connecting radiation processes to the structure of galaxies and in testing observational data analysis techniques (see e.g.[42] and[43]). While these types of analysis are beginning to emerge in the study of GRBs (see references within), we need widespread adoption of these simulations within the community for any SRHD simulations that have been conducted of GRB jets. This would allow the community to build a library of GRB jet simulations and their associated radiation signatures that can be used in comparison to GRB observations, greatly enhancing our knowledge of GRB jets and the prompt emission.

References

1. B. Zhang and H. Yan, The Internal-Collision-Induced Magnetic Reconnection and Turbulence (ICMART) model of gamma-ray bursts, *The Astrophysical Journal* **726**, p. 90 (dec 2011).
2. M. J. Rees and P. Mészáros, Unsteady outflow models for cosmological gamma-ray bursts, *ApJL* **430**, L93 (August 1994).
3. F. Daigne and R. Mochkovitch, Gamma-ray bursts from internal shocks in a relativistic wind: temporal and spectral properties, *Monthly Notices of the Royal Astronomical Society* **296**, 275 (1998).
4. F. Daigne, Ž. Bošnjak and G. Dubus, Reconciling observed gamma-ray burst prompt spectra with synchrotron radiation?, *Astronomy & Astrophysics* **526**, p. A110 (2011).
5. M. J. Rees and P. Mészáros, Dissipative photosphere models of gamma-ray bursts and x-ray flashes, *The Astrophysical Journal* **628**, p. 847 (2005).
6. A. Pe'er, Temporal evolution of thermal emission from relativistically expanding plasma, *The Astrophysical Journal* **682**, p. 463 (2008).
7. A. Pe'er and F. Ryde, A theory of multicolor blackbody emission from relativistically expanding plasmas, *The Astrophysical Journal* **732**, p. 49 (2011).
8. A. Pe'er, P. Mészáros and M. J. Rees, The observable effects of a photospheric component on grb and xrf prompt emission spectrum, *The Astrophysical Journal* **642**, p. 995 (2006).
9. A. M. Beloborodov, Collisional mechanism for gamma-ray burst emission, *Monthly Notices of the Royal Astronomical Society* **407**, p. 1033 (2010).
10. D. Lazzati, B. J. Morsony, R. Margutti and M. C. Begelman, Photospheric emission as the dominant radiation mechanism in long-duration gamma-ray bursts, *The Astrophysical Journal* **765**, p. 103 (2013).
11. D. Lazzati, Monte carlo radiation transfer simulations of photospheric emission in long-duration gamma-ray bursts, *The Astrophysical Journal* **829**, p. 76 (2016).

12. T. Parsotan and D. Lazzati, A monte carlo radiation transfer study of photospheric emission in gamma-ray bursts, *The Astrophysical Journal* **853**, p. 8 (2018).
13. T. Parsotan, D. López-Cámara and D. Lazzati, Photospheric emission from variable engine gamma-ray burst simulations, *The Astrophysical Journal* **869**, p. 103 (2018).
14. T. Parsotan, D. López-Cámara and D. Lazzati, Photospheric polarization signatures from long gamma-ray burst simulations, *The Astrophysical Journal* **896**, p. 139 (2020).
15. H. Ito, J. Matsumoto, S. Nagataki, D. C. Warren, M. V. Barkov and D. Yonetoku, The photospheric origin of the yonetoku relation in gamma-ray bursts, *Nature communications* **10**, 1 (2019).
16. G. Oganesyan, L. Nava, G. Ghirlanda, A. Melandri and A. Celotti, Prompt optical emission as a signature of synchrotron radiation in gamma-ray bursts, *Astronomy & Astrophysics* **628**, p. A59 (aug 2019).
17. Amati, L., Frontera, F., Tavani, M., in 't Zand, J. J. M., Antonelli, A., ta, E. Co, Feroci, M., Guidorzi, C., e, J. Hei, etti, N. Ma, Montanari, E., tro, L. Nica, Palazzi, E., Pian, E., Piro, L. and Soffitta, P., Intrinsic spectra and energetics of bepposax gamma ray bursts with known redshifts, *A&A* **390**, 81 (2002).
18. D. Yonetoku, T. Murakami, T. Nakamura, R. Yamazaki, A. K. Inoue and K. Ioka, Gamma-ray burst formation rate inferred from the spectral peak energy-peak luminosity relation, *The Astrophysical Journal* **609**, p. 935 (2004).
19. D. Lazzati, B. J. Morsony and M. C. Begelman, Very high efficiency photospheric emission in long-duration γ-ray bursts, *The Astrophysical Journal Letters* **700**, p. L47 (2009).
20. D. López-Cámara, B. J. Morsony and D. Lazzati, Photospheric emission from long-duration gamma-ray bursts powered by variable engines, *Monthly Notices of the Royal Astronomical Society* **442**, p. 2202 (2014).
21. T. Parsotan and D. Lazzati, Photospheric prompt emission from long gamma ray burst simulations–i. optical emission, *arXiv preprint arXiv:2105.06505* (2021).
22. E. Troja, V. Lipunov, C. Mundell, N. Butler, A. Watson, S. Kobayashi, S. Cenko, F. Marshall, R. Ricci, A. Fruchter *et al.*, Significant and variable linear polarization during the prompt optical flash of grb 160625b, *Nature* **547**, 425 (2017).
23. J. L. Racusin, S. V. Karpov, M. Sokolowski, J. Granot, X. F. Wu, V. Pal'shin, S. Covino, A. J. van der Horst, S. R. Oates, P. Schady, R. J. Smith, J. Cummings, R. L. C. Starling, L. W. Piotrowski, B. Zhang, P. A. Evans, S. T. Holland, K. Malek, M. T. Page, L. Vetere, R. Margutti, C. Guidorzi, A. P. Kamble, P. A. Curran, A. Beardmore, C. Kouveliotou, L. Mankiewicz, A. Melandri, P. T. O'Brien, K. L. Page, T. Piran, N. R. Tanvir, G. Wrochna, R. L. Aptekar, S. Barthelmy, C. Bartolini, G. M. Beskin, S. Bondar, M. Bremer, S. Campana, A. Castro-Tirado, A. Cucchiara, M. Cwiok, P. D'Avanzo, V. D'Elia, M. D. Valle, A. de Ugarte Postigo, W. Dominik, A. Falcone, F. Fiore, D. B. Fox, D. D. Frederiks, A. S. Fruchter, D. Fugazza, M. A. Garrett, N. Gehrels, S. Golenetskii, A. Gomboc, J. Gorosabel, G. Greco, A. Guarnieri, S. Immler, M. Jelinek, G. Kasprowicz, V. L. Parola, A. J. Levan, V. Mangano, E. P. Mazets, E. Molinari, A. Moretti, K. Nawrocki, P. P. Oleynik, J. P. Osborne, C. Pagani, S. B. Pandey, Z. Paragi, M. Perri, A. Piccioni, E. Ramirez-Ruiz, P. W. A. Roming, I. A. Steele, R. G. Strom, V. Testa, G. Tosti, M. V. Ulanov, K. Wiersema, R. A. M. J. Wijers, J. M. Winters, A. F. Zarnecki, F. Zerbi, P. Mészáros, G. Chincarini and D. N. Burrows, Broadband observations of the naked-eye γ-ray burst grb 080319b, *Nature* **455**, 183 (sep 2008).

24. R.-F. Shen and B. Zhang, Prompt optical emission and synchrotron self-absorption constraints on emission site of GRBs, *Monthly Notices of the Royal Astronomical Society* **398**, 1936 (09 2009).
25. S. M. Tang and S. N. Zhang, Time lag between prompt optical emission and γ-rays in GRBs, *Astronomy & Astrophysics* **456**, 141 (aug 2006).
26. Y.-Z. Fan, B. Zhang and D.-M. Wei, Naked-eye optical flash from gamma-ray burst 080319b: Tracing the decaying neutrons in the outflow, *Physical Review D* **79** (jan 2009).
27. Y.-C. Zou, T. Piran and R. Sari, CLUES from the prompt emission of GRB 080319b, *The Astrophysical Journal* **692**, L92 (jan 2009).
28. H. Ito, J. Matsumoto, S. Nagataki, D. C. Warren and M. V. Barkov, Photospheric emission from collapsar jets in 3d relativistic hydrodynamics, *The Astrophysical Journal Letters* **814**, p. L29 (2015).
29. T. Parsotan and D. Lazzati, Photospheric Prompt Emission From Long Gamma Ray Burst Simulations – II. Spectropolarimetry, *arXiv e-prints* , p. arXiv:2108.11412 (August 2021).
30. T. Parsotan, lazzati-astro/MCRaT: MCRaT with Cyclo-synchrotron Emission/ Absorption (June 2021).
31. T. Parsotan, parsotat/processmcrat: Cyclo-synchrotron release (June 2021).
32. J. Hulsman, Polar-2: a large scale gamma-ray polarimeter for grbs, in *Space Telescopes and Instrumentation 2020: Ultraviolet to Gamma Ray*, 2020.
33. T. Poole, A. Breeveld, M. Page, W. Landsman, S. Holland, P. Roming, N. Kuin, P. Brown, C. Gronwall, S. Hunsberger *et al.*, Photometric calibration of the swift ultraviolet/optical telescope, *Monthly Notices of the Royal Astronomical Society* **383**, 627 (2008).
34. C. Rodrigo and E. Solano, The svo filter profile service, in *Contributions to the XIV. 0 Scientific Meeting (virtual) of the Spanish Astronomical Society*, 2020.
35. W. Deng and B. Zhang, Low energy spectral index and ep evolution of quasi-thermal photosphere emission of gamma-ray bursts, *The Astrophysical Journal* **785**, p. 112 (2014).
36. A. M. Beloborodov, Radiative transfer in ultrarelativistic outflows, *The Astrophysical Journal* **737**, p. 68 (2011).
37. B. Fryxell, K. Olson, P. Ricker, F. Timmes, M. Zingale, D. Lamb, P. MacNeice, R. Rosner, J. Truran and H. Tufo, Flash: An adaptive mesh hydrodynamics code for modeling astrophysical thermonuclear flashes, *The Astrophysical Journal Supplement Series* **131**, p. 273 (2000).
38. L. Nava, R. Salvaterra, G. Ghirlanda, G. Ghisellini, S. Campana, S. Covino, G. Cusumano, P. D'Avanzo, V. D'Elia, D. Fugazza, A. Melandri, B. Sbarufatti, S. D. Vergani and G. Tagliaferri, A complete sample of bright swift long gamma-ray bursts: testing the spectral-energy correlations, *Monthly Notices of the Royal Astronomical Society* **421**, p. 1256 (2012).
39. O. Gottlieb, E. Nakar and O. Bromberg, The structure of hydrodynamic γ-ray burst jets, *Monthly Notices of the Royal Astronomical Society* **500**, 3511 (2021).
40. H. Ito, S. Nagataki, J. Matsumoto, S.-H. Lee, A. Tolstov, J. Mao, M. Dainotti and A. Mizuta, Spectral and polarization properties of photospheric emission from stratified jets, *The Astrophysical Journal* **789**, p. 159 (2014).
41. C. Lundman, A. Pe'er and F. Ryde, Polarization properties of photospheric emission from relativistic, collimated outflows, *Monthly Notices of the Royal Astronomical Society* **440**, 3292 (2014).

42. T. Parsotan, R. Cochrane, C. Hayward, D. Angles-Alcazar, R. Feldmann, C.-A. Faucher-Giguere, S. Wellons and P. Hopkins, Realistic mock observations of the sizes and stellar mass surface densities of massive galaxies in fire-2 zoom-in simulations, *Monthly Notices of the Royal Astronomical Society* **501**, 1591 (2021).
43. R. Cochrane, C. Hayward, D. Anglés-Alcázar, J. Lotz, T. Parsotan, X. Ma, D. Kereš, R. Feldmann, C. Faucher-Giguere and P. Hopkins, Predictions for the spatial distribution of the dust continuum emission in star-forming galaxies, *Monthly Notices of the Royal Astronomical Society* **488**, 1779 (2019).

The photosphere emission spectrum of hybrid relativistic outflow for Gamma-Ray Bursts

Yan-Zhi Meng,[1,2,3] Jin-Jun Geng[1] and Xue-Feng Wu[1]

[1] *Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China*
[2] *School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China*
[3] *Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, China*
E-mail: yzmeng@nju.edu.cn (YZM), jjgeng@pmo.ac.cn (JJG) and xfwu@pmo.ac.cn (XFW)

The photospheric emission in the prompt phase is the natural prediction of the original fireball model for gamma-ray burst (GRB) due to the large optical depth ($\tau > 1$) at the base of the outflow, which is supported by the quasi-thermal components detected in several *Fermi* GRBs. However, which radiation mechanism (photosphere or synchrotron) dominates in most GRB spectra is still under hot debate. The shape of the observed photosphere spectrum from a pure hot fireball or a pure Poynting-flux-dominated outflow has been investigated before. In this work, we further study the photosphere spectrum from a hybrid outflow containing both a thermal component and a magnetic component with moderate magnetization ($\sigma_0 = L_P/L_{\rm Th} \sim 1 - 10$), by invoking the probability photosphere model. The high-energy spectrum from such a hybrid outflow is a power law rather than an exponential cutoff, which is compatible with the observed Band function in large amounts of GRBs. Also, the distribution of the low-energy indices (corresponding to the peak-flux spectra) is found to be quite consistent with the statistical result for the peak-flux spectra of GRBs best-fitted by the Band function, with similar angular profiles of structured jet in our previous works. Finally, the observed distribution of the high-energy indices can be well understood after considering the different magnetic acceleration (due to magnetic reconnection and kink instability) and the angular profiles of dimensionless entropy with the narrower core.

Keywords: Gamma-ray burst: general – radiation mechanisms: thermal – radiative transfer – scattering

1. Introduction

After decades of researches, the radiation mechanism of GRB prompt emission is still unclear.[30–33] The photospheric emission model seems to be a promising scenario.[12,19–21,24,27] The photospheric emission model naturally interprets the clustering of the peak energies[14] and the high radiation efficiency[16] observed.

Indeed, a quasi-thermal component has been found in great amounts of BATSE GRBs[25] and several *Fermi* GRBs (GRB 090902B[1]). To account for the broad spectra of other GRBs, the quasi-thermal spectrum need to be broadened. Two different mechanisms of broadening have been proposed theoretically, i.e., the subphotospheric dissipation[7,24] and the geometric broadening.[4,17,19,20,22] This geometric broadening, namely the probability photosphere model, is the result of the fact that photons can be last scattered at any place (r, Ω) inside the outflow if only the electron exists there.[3,22,23] Then, the observed photosphere spectrum is the superposition of a series of blackbodies with different temperature, thus broadened.

In studies on the probability photosphere model till now,[4,17,20] the jet is accelerated solely by the radiative pressure of the thermal photons. However, the magnetically driven acceleration[5,7] may also play an important role in the real situation, which has not been properly considered in the framework of the probability photosphere model. In this work we study the photospheric emission spectrum within the framework of the probability photosphere model for a hybrid relativistic outflow, which contains a thermal component and a magnetic component, including both the thermally and magnetically driven acceleration.

The paper is organized as follows. In Section 2, we describe the calculations of the photospheric emission spectrum for a hybrid outflow within the probability photosphere model, including the energy injections of impulsive injection and more reasonable continuous wind. The calculated spectral results and parameter dependencies are shown in Section 3. The conclusions are summarized in Section 4.

2. Probability photosphere emission from a hybrid jet

2.1. *Hybrid jet and its dynamics*

A hybrid jet is composed of a thermal component and a magnetic component, which could be described by the dimensionless entropy η and the initial magnetization parameter σ_0. The dynamics for a hybrid jet may be approximated as

$$\Gamma(r) = \begin{cases} \frac{r}{r_0}, & r_0 < r < R_{\rm ra}; \\ \Gamma_{\rm ra}\ (\frac{r}{R_{\rm ra}})^{\delta}, & R_{\rm ra} < r < R_c; \\ \Gamma_c, & r > R_c, \end{cases} \tag{1}$$

where r_0 is the radius at the jet base, $\Gamma_{\rm ra} = \max(\eta, [\eta(1+\sigma_0)]^{1/3})$ is the Lorentz factor at $R_{\rm ra}$, $\Gamma_c \simeq \eta(1+\sigma_0)$ is the coasting Lorentz factor, $R_{\rm ra}$ is the rapid acceleration radius and R_c is the coasting radius.

2.2. *Time-resolved spectra from probability photosphere emission*

2.2.1. *Impulsive injection*

For the probability photosphere model, the observed spectrum is a superposition of a series of blackbodies emitted from any place in the outflow with a certain probability, which is calculated by

$$F_\nu(\nu, t) = \frac{N_0}{4\pi d_L^2} \iint P_1(r,\mu) P_2(\nu, T) h\nu \times \delta(t - \frac{ru}{\beta c}) d\mu dr, \tag{2}$$

where $u = 1 - \beta\cos\theta$, and $N_0 = L_w/2.7k_{\rm B}T_0$ is the number of photons injected impulsively at the base of the outflow, $P_1(r,\mu)$ represents the probability density function for the final scattering to occur at the coordinates (r,θ), $\mu = \cos\theta$, $P_2(\nu, T)$ represents the probability for a photon of the observed frequency ν last-scattered at (r,θ) with the observer frame temperature of T.

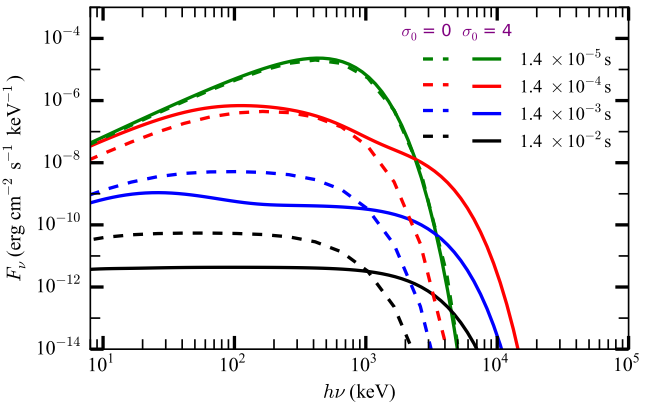


Fig. 1. The time-resolved spectra of impulsive injection for the jet with and without magnetization. The solid lines show the calculated time-resolved spectra for the hybrid jet with a hot fireball component ($\eta = 400$) and a cold Poynting-flux component (σ_0= 4). Also, a total outflow luminosity of $L_w = 10^{53}$ erg s^{-1} is assumed, base outflow radius $R_0 = 3 \times 10^8$ cm, and luminosity distance $d_{\rm L} = 4.85 \times 10^{28}$ cm ($z = 2$). Different colors represent different observational times. For comparison, the dashed lines illustrate the time-resolved spectra calculated for the jet without magnetization.[4] Obviously, at later times the time-resolved spectra extend to much higher energy and a power-law component emerges in the low-energy end.

2.2.2. *Continuous wind*

It is more realistic to consider the continuous wind from the central engine since the GRBs have relatively long duration (> 1 s). For a layer ejected from $\hat{t}$ to $\hat{t} + d\hat{t}$, the observed spectrum at the observer time t is

$$\hat{F}_\nu(\nu, t, \hat{t}) = \frac{N_0}{4\pi d_L^2} \iint P_1(r, \mu) P_2(\nu, T) h\nu \times \delta(t - \hat{t} - \frac{ru}{\beta c}) d\mu dr. \tag{3}$$

Then, integrating over all the layers, we obtain the observed time-resolved spectrum to be

$$F_\nu(\nu, t) = \int_0^t \hat{F}_\nu(\nu, t, \hat{t}) d\hat{t}. \tag{4}$$

3. Calculated results

3.1. *Impulsive injection*

The calculated time-resolved spectra are shown in Figure 1 (the solid lines). It is found that at the later times (the high-latitude emission dominates) the time-resolved spectra firstly show a power-law shape extending to a much higher energy than the early-time blackbody. Then, the power-law component vanishes gradually towards the low-energy end and the peak energy on the high-energy end remains constant.

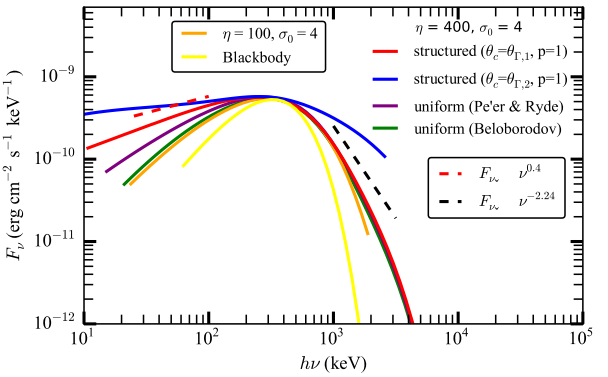


Fig. 2. The time-resolved spectra at later times ($t \geq 1$ s) of a continuous wind for the hybrid jet with magnetization σ_0= 4 and various values or angular profiles of dimensionless entropy η. Here, $\theta_{\Gamma,1} = 1/400$ and $\theta_{\Gamma,2} = 1/4000$. All the time-resolved spectra have been normalized to the same peak energy and peak flux as the case of $\eta = 400$ and σ_0= 4 (green line) to compare clearly the low-energy and high-energy indices. For comparison, the yellow line shows the spectrum of blackbody. Also, the red and black dashed lines represent respectively the average low-energy index (−0.6) and high-energy index (−3.24) fitted with the Band function for the time-resolved spectra of a large sample of single pulses.[29]

3.2. *Continuous wind*

For the continuous wind with constant wind luminosity, the spectrum at a later time ($t \geq 1$ s) which corresponds to the observed peak-flux spectrum is calculated by Equation (4). This calculated time-resolved spectrum, with the same parameters as Figure 1 ($\eta = 400$, σ_0= 4) is illustrated by the green lines (with the probability density function introduced in[3]) in Figure 2. Apparently, the spectrum on the high-energy end is a power law rather than an exponential cutoff, which is the natural result of the power-law component extending to much higher energy in Figure 2.

Furthermore, we calculate the spectrum (with the probability density function described in[23]) for the case of the dimensionless entropy η of a lateral structure, with which the observed typical low-energy spectral index $\alpha \approx -1$ can be obtained for the unmagnetized jet under the framework of the probability photosphere model.[17,20] The considered angular profile consists of an inner constant core with width θ_c and the outer power-law decreased component with power-law index p. Note that an angle-independent luminosity is considered, because the spectrum expected to be observed is formed by the photons making their last scattering at approximately $\lesssim 5/\,\Gamma_0$ (as shown in[17]) and the isotropic angular width for Lorentz factor $\theta_{c,\Gamma}$ is likely to be much smaller than the isotropic angular width for luminosity $\theta_{c,L}$.[6,13,15,26,34] The red line in Figure 2 represents the calculated spectrum for a hybrid jet with $\theta_c = \theta_{\Gamma,1} = 1/400$ and $p = 1$, corresponding to the typical value $\alpha \sim -1$ for an unmagnetized jet. While the blue line is the spectrum for hybrid jet with $\theta_c = \theta_{\Gamma,2} = 1/4000$ and $p = 1$, corresponding to the minimum value $\alpha \sim -2$ for unmagnetized jet

(see Figure 7d in[20]). To compare clearly the low-energy and high-energy indices, in Figure 2 the calculated actual time-resolved spectra have been normalized to the same peak energy and peak flux as the case of $\eta = 400$ and $\sigma_0 = 4$.

We then find that the spectrum of the hybrid jet possesses the much harder low-energy spectral index than that of the unmagnetized jet with the same angular profile, $\alpha \sim -0.6$ for $\theta_c = \theta_{\Gamma,1}$, $p = 1$ and $\alpha \sim -1$ for $\theta_c = \theta_{\Gamma,2}$, $p = 1$. The high-energy power-law index β is the same as the uniform case for $\theta_c = \theta_{\Gamma,1}$, $p = 1$ ($\beta \sim -4$, see Figure 3 and the discussion in Section 3.3), and much larger ($\beta \sim -1.6$) for $\theta_c = \theta_{\Gamma,2}$, $p = 1$. In a word, the spectrum of the hybrid jet is analogue to the empirical Band function[2] spectrum, whereas the spectrum of the unmagnetized jet corresponds to the spectrum of the empirical cutoff power-law model,[20] within the framework of the probability photosphere model.

Interestingly, in all the statistical works of a large sample of GRBs,[9–11,14,28] the average low-energy spectral index for the GRBs best-fitted by the Band function is harder ($0.1 \sim 0.3$) than that for the GRBs best-fitted by the cutoff power-law model, for both the time-integrated and the peak-flux spectra. This hardness is quite consistent with our results for the probability photosphere model discussed in the previous paragraph. Also, for the distributions of the low-energy and high-energy indices, our results (corresponding to the peak-flux spectra) are quite similar to the statistical results of the peak-flux spectra for the GRBs best-fitted by the Band function. For the low-energy spectral index, the typical value is $\alpha \sim -0.6$ and the minimum value $\alpha \sim -1$. While for the high-energy spectral index, the maximum value $\beta \sim -1.6$ in our model is close to the statistical result. The typical value $\beta \sim -4$ in our model seems to be much softer than the well-known $\beta \sim -2.5$, which we consider to arise from the assumption of the single pulse, namely without the overlap of the pulses. This consideration is proposed recently,[29] where the time-resolved spectral analysis of a large sample of single pulses has been performed and the average high-energy spectral index ($\beta \sim -3.24$) indeed is found to be much softer. Noteworthily, the average high-energy spectral index $\beta \sim -3.24$ in that work can be well reproduced with our hybrid jet model, by considering the parameter dependence on the power-law index of magnetic acceleration δ (see Figure 3 and the discussion in Section 3.3). In Figure 2, the red and black dashed lines represent the average low-energy index (-0.6) and high-energy index (-3.24) in that work, respectively.

3.3. *Parameter dependence*

In the above discussion, for the magnetically driven acceleration, we only consider the case of magnetic reconnection for the non-axisymmetric rotator.[5] Whereas, for an initially axisymmetric flow, the acceleration can happen through the kink instability.[8,18] Besides, the magnetic acceleration driven by the kink instability seems to be more rapid than the magnetic reconnection case, with a larger power-law index δ of $\sim 2/5$ or even $\sim 1/2$ (see Figure 5 in[8]). Thus, in Figure 3 we compare the

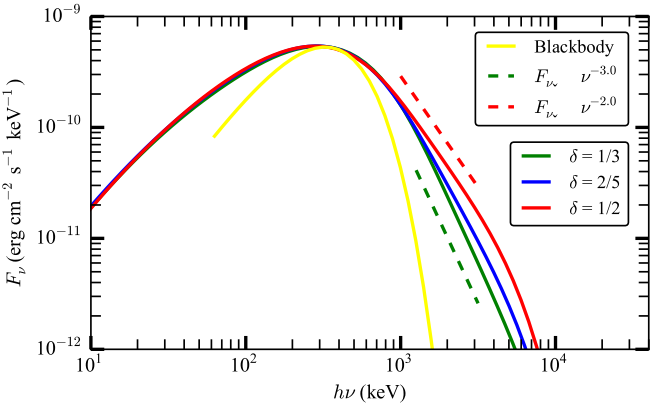


Fig. 3. The calculated time-resolved spectra at later times ($t \geq 1$ s) of continuous wind with different power-law index of magnetic acceleration δ. The green, blue and red lines are for $\delta = 1/3$, $\delta = 2/5$ and $\delta = 1/2$ respectively, along with magnetization $\sigma_0 = 9$. While the green and red dashed lines represent the high-energy spectral index $\beta = -4$ and $\beta = -3$, respectively. The yellow line is the spectrum of blackbody.

calculated spectra of our model for $\delta = 1/3$, $\delta = 2/5$ and $\delta = 1/2$, and find that their high-energy power-law indices are quite different while the low-energy indices remain the same. For $\delta = 1/3$, we have $\beta = -4$; for $\delta = 1/2$, we obtain $\beta = -3$; and β lies between -4 and -3 for $\delta = 2/5$. Surprisingly, this rough distribution of β (-4 to -3) is well consistent with the β distribution of the softer cluster for a large sample of single pulses,[29] where the β distribution seems to show the bimodal distribution with a harder typical cluster (peaks at -2.5) and a softer cluster (peaks at -3.5; see Figure 1 therein).

4. Conclusions

In this paper, by invoking the probability photosphere model we investigate the shape of the photospheric emission spectrum for the hybrid outflow, which contains a thermal component and a magnetic component with moderate magnetization ($\sigma_0 = L_P/L_{\rm Th} \sim 1 - 10$). The following conclusions are drawn.

(1) The photosphere spectrum on the high-energy end is a power law rather than an exponential cutoff.

(2) With the similar angular profiles of the dimensionless entropy η as the unmagnetized jet, considered in previous works,[17,20] the distribution of the low-energy indices (corresponding to the peak-flux spectra) for our photosphere model is quite consistent with the statistical result of the peak-flux spectra for the GRBs best-fitted by the Band function.

(3) The high-energy power-law index β for our photosphere model solely depends on the power-law index of magnetic acceleration δ. Also, the observed β distribution could be well interpreted with the photosphere model in this work.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant Nos. 11725314, 12041306, 11903019, 11833003), the Major Science and Technology Project of Qinghai Province (2019-ZJ-A10). Y.Z.M. is supported by the National Postdoctoral Program for Innovative Talents (grant No. BX20200164).

References

1. Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009, *Astrophys. J.*, 706, L138
2. Band, D., Matteson, J., Ford, L., et al. 1993, *Astrophys. J.*, 413, 281
3. Beloborodov, A. M. 2011, *Astrophys. J.*, 737, 68
4. Deng, W., & Zhang, B. 2014, *Astrophys. J.*, 785, 112
5. Drenkhahn, G. 2002, *Astron. Astrophys.*, 387, 714
6. Geng, J.-J., Zhang, B., Kölligan, A., Kuiper, R., & Huang, Y.-F. 2019, *Astrophys. J.*, 877, L40
7. Giannios, D. 2006, *Astron. Astrophys.*, 457, 763
8. Giannios, D., & Spruit, H. C. 2006, *Astron. Astrophys.*, 450, 887
9. Goldstein, A., Burgess, J. M., Preece, R. D., et al. 2012, *Astrophys. J.*, 199, 19
10. Goldstein, A., Preece, R. D., Mallozzi, R. S., et al. 2013, *Astrophys. J.*, 208, 21
11. Gruber, D., Goldstein, A., Weller von Ahlefeld, V., et al. 2014, *Astrophys. J.*, 211, 12
12. Hou, S.-J., Zhang, B.-B., Meng, Y.-Z., et al. 2018, *Astrophys. J.*, 866, 13
13. Ito, H., Just, O., Takei, Y., et al. 2021, arXiv:2105.09323
14. Kaneko, Y., Preece, R. D., Briggs, M. S., et al. 2006, *Astrophys. J.*, 166, 298
15. Lazzati, D., Morsony, B. J., & Begelman, M. C. 2007, Philosophical Transactions of the Royal Society of London Series A, 365, 1141
16. Lloyd-Ronning, N. M., & Zhang, B. 2004, *Astrophys. J.*, 613, 477
17. Lundman, C., Pe'er, A., & Ryde, F. 2013, *Mon. Not. R. Astron. Soc.*, 428, 2430
18. Lyutikov, M., & Blandford, R. 2003, arXiv:astro-ph/0312347
19. Meng, Y.-Z., Geng, J.-J., Zhang, B.-B., et al. 2018, *Astrophys. J.*, 860, 72
20. Meng, Y.-Z., Liu, L.-D., Wei, J.-J., Wu, X.-F., & Zhang, B.-B. 2019, *Astrophys. J.*, 882, 26
21. Mészáros, P., & Rees, M. J. 2000, *Astrophys. J.*, 530, 292
22. Pe'er, A. 2008, *Astrophys. J.*, 682, 463
23. Pe'er, A., & Ryde, F. 2011, *Astrophys. J.*, 732, 49
24. Rees, M. J., & Mészáros, P. 2005, *Astrophys. J.*, 628, 847
25. Ryde, F., & Pe'er, A. 2009, *Astrophys. J.*, 702, 1211
26. Tchekhovskoy, A., McKinney, J. C., & Narayan, R. 2008, *Mon. Not. R. Astron. Soc.*, 388, 551
27. Thompson, C. 1994, *Mon. Not. R. Astron. Soc.*, 270, 480
28. Yu, H.-F., Preece, R. D., Greiner, J., et al. 2016, *Astron. Astrophys.*, 588, A135
29. Yu, H.-F., Dereli-Bégué, H., & Ryde, F. 2019, *Astrophys. J.*, 886, 20
30. Zhang, B., & Yan, H. 2011, *Astrophys. J.*, 726, 90
31. Zhang, B. 2020, Nature Astronomy, 4, 210
32. Zhang, B.-B., Zhang, B., Liang, E.-W., et al. 2011, *Astrophys. J.*, 730, 141
33. Zhang, B.-B., Zhang, B., Castro-Tirado, A. J., et al. 2018a, Nature Astronomy, 2, 69
34. Zhang, W., Woosley, S. E., & MacFadyen, A. I. 2003, *Astrophys. J.*, 586, 356.

On diffusive photospheres in Gamma-Ray Bursts

G. V. Vereshchagin

ICRANet, 65122, p.le della Repubblica, 10, Pescara, Italy
ICRA and Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy
ICRANet-Minsk, National Academy of Sciences of Belarus, Nezavisimosti ave. 68, 220072 Minsk, Belarus
INAF, Viale del Parco Mellini 84, 00136 Rome, Italy
E-mail: veresh@icra.it

Photospheric emission from relativistic outflows may originate in two different regimes: photon decoupling within the outflow or radiative diffusion. I show that observed thermal component in the early afterglows of gamma-ray bursts can originate from such diffusive photospheres.

Keywords: Thermal radiation, gamma-ray bursts

1. Introduction

Gamma-ray bursts (GRBs) are strong and short irregular flashes of hard "prompt" radiation, followed by the long lasting fading out smooth "afterglow" emission spanning from sub-TeV energies to radio. Since their discovery a number of dedicated space observatories and ground based telescopes are constantly monitoring the sky daily reporting new bursts and measuring distance to their host galaxies. GRBs come in two kinds[1]: short and long, with their possible progenitors being binary neutron star mergers and collapsing massive stars reaching the endpoint of their evolution, respectively. Extremely large energies released in γ-rays ($\leq 10^{54}$ erg) as well as a short variability time (≤ 10 ms) point to ultrarelativistic outflows giving rise to the observed emission.[2]

The spectrum of the prompt emission is non-thermal, its origin can be traced to the synchrotron mechanism in relativistic shock waves, generated by the relativistic material breaking in the interstellar medium.[3] Photospheric models with possible dissipation of kinetic energy of the outflow are attractive alternative to the synchrotron models since observation of thermal radiation allows for the determination of basic hydrodynamic characteristics of the outflow from which these bursts originate.[4–6] The photons in these models are trapped and advected with the opaque outflow, until it becomes transparent. In many GRBs, a subdominant thermal component was detected during their prompt emission, while in several, such as GRB 090902B observed spectrum is almost thermal.[7,8]

Thermal components are also detected in time resolved spectra during the early afterglow in a number of GRBs.[9–14] So far several mechanisms to generate such

emission are proposed. They include a shock breakout from a progenitor star or a stellar wind[15] and a hot cocoon formed when the relativistic jet emerges from the stellar surface.[16,17] Some authors argue that shock breakouts are not energetic enough and do not last long enough to explain the observed thermal emission,[13] leaving cocoons as a favourite model. In addition, there is an alternative proposal of a cloud or a clump with small mass, accelerated by the GRB outflow.[18]

Most studies dealing with the photospheric emission, e.g.,[19–25] for a review see,[5] adopt the hydrodynamic model of a steady and infinite wind. However, finite duration of GRBs implies finite width of the wind. Winds of *finite duration* are classified as photon thin and photon thick.[4,26,27] Decoupling of photons from plasma in the latter case occurs simultaneously in the entire outflow, while in the former case photons are transported to the boundaries of the outflow by radiative diffusion, just like in nonrelativistic outflows, e.g. in supernova ejecta. Emission in this case originates not at the photospheric radius, but at smaller radii. The photon thick case, corresponding to the steady wind, appears to be justified for typical GRB parameters. Photon thin regime is overlooked in the literature, as it is assumed that at large radii the outflow is spreading[28,29] due to strong velocity gradients initially generated in the outflow, see e.g.[28,29] Such spreading outflows indeed correspond to the photon thick case.[30] However, the origin of these gradients is not clear. Even if initially present, such gradients may be erased in initial phases of expansion, e.g. by oblique shocks or other mechanisms. Hence the outflow can be photon thin. In such a case decoupling of photons from expanding plasma occurs in the diffusive regime.[27]

Radiative diffusion is known to be relevant for expanding ejecta in supernovae explosions.[31] Recently the theory of photospheric emission[27] was further extended.[32] In particular, physical parameters of such outflows were determined using the observational data.

The paper is organized as follows. The definition of relativistic photosphere is recalled in Section 2. Observational properties of diffusive photospheres are discussed in Section 3. The method for determination of physical parameters of the outflow is presented in Section 4. Physical parameters of GRB cocoons are discussed in Section 5. Case studies of GRBs with thermal emission in the early afterglow are reported in Section 6. Discussion and conclusion follow.

2. Relativistic photosphere

Consider a relativistic outflow launched at a radius R_0. The outflow is characterized by its activity time Δt, the luminosity L and mass injection rate $\dot{M}$. The associated thickness of the outflow is $l = c\Delta t$. The entropy in the region where the energy is released is parametrized by a dimensionless parameter $\eta = L/\dot{M}c^2$. Spherical symmetry is assumed, but generalization for anisotropic case with $\eta(\theta)$, where θ is the polar angle is straightforward. When $\eta \gg 1$ the bulk Lorentz factor changes

with the radial distance as

$$\Gamma \simeq \begin{cases} \dfrac{r}{R_0}, & R_0 < r < \eta R_0, \\ \\ \eta \simeq \text{const}, & r > \eta R_0, \end{cases} \tag{1}$$

During both acceleration and coasting phases the continuity equation for the laboratory number density reduces to

$$n = \begin{cases} n_0 \left(\dfrac{R_0}{r}\right)^2, & R(t) < r < R(t) + l, \\ \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $R(t)$ is the radial position of the inner boundary of the outflow.

The optical depth for a spherically symmetric outflow is[27,33]

$$\tau = \int_R^{R+\Delta R} \sigma n \left(1 - \beta \cos\theta\right) \frac{dr}{\cos\theta}, \tag{3}$$

where $R + \Delta R$ is the radial coordinate at which the photon leaves the outflow, and θ is the angle between the velocity vector of the outflow and the direction of propagation of the photon, n is the laboratory number density of electrons and positrons, which may be present due to pair production. The dominant interaction of photons in our case is Compton scattering in the non-relativistic regime, so σ is the Thomson cross section.

Electron-positron-photon plasma with baryon loading reaches thermal equilibrium before its expansion starts.[34,35] With decreasing entropy η opacity due to electrons associated with baryons increases and eventually dominates over pair opacity. For the laboratory density profile (2) one has in the radial direction

$$\tau = \begin{cases} \dfrac{1}{6}\tau_0 \left(\dfrac{R_0}{R}\right)^3, & R_0 \ll R \ll \eta R_0, \\ \\ \dfrac{1}{2\eta^2}\tau_0 \left(\dfrac{R_0}{R}\right), & \eta R_0 \ll R \ll \eta^2 R_0, \\ \\ \tau_0 \left(\dfrac{R_0}{R}\right)^2, & R \gg \eta^2 R_0, \end{cases} \tag{4}$$

where

$$\tau_0 = \frac{\sigma L}{4\pi m_p c^3 R_0 \eta} = n_0 \sigma R_0. \tag{5}$$

The first two lines correspond to a *photon thick outflow* and the third line corresponds to a *photon thin outflow*.[27]

The photospheric radius R_{ph} is defined by equating (4) to unity

$$R_{ph} = \begin{cases} R_0 \left(\frac{\tau_0}{6}\right)^{1/3}, & \tau_0 \ll \eta^3, \\ R_0 \frac{\tau_0}{2\eta^2}, & \eta^3 \ll \tau_0 \ll 4\eta^4 \frac{l}{R_0}, \\ (\tau_0 R_0 l)^{1/2}, & \tau_0 \gg 4\eta^4 \frac{l}{R_0}. \end{cases} \tag{6}$$

In eqs. (4) and (6) the regions of validity of different approximations are expressed either for radius or for parameters of the outflow. The crucial parameter which determines whether the outflow is photon thick or thin is the ratio

$$\chi = \frac{\tau_0}{4\eta^4} \frac{l}{R_0}. \tag{7}$$

The outflow is photon thin for $\chi \gg 1$ and it is photon thick otherwise.

3. Relativistic diffusive photosphere

The definition of the photosphere implies that at this position in space the outflow as a whole becomes transparent to radiation. However, emission emerges from the outflow when it is optically thick as well. Such emission is due to radiative diffusion, which transfers the energy from deeper parts of the outflow towards its surface. Naively one can think that such an effect is negligible in ultrarelativistic outflows. However, this is not the case.[4] Indeed, the comoving time, which photon takes to cross the outflow with comoving thickness $l_c = \Gamma l$ is $t_c = l_c^2/D_c$, where $D_c = c/3\sigma n_c$ is the diffusion coefficient, $n_c = n/\Gamma$ is the comoving density of the outflow. The radial coordinate of the outflow at this time is $R \simeq \Gamma c t_c$.

This diffusion radius is found in Ref. 27, and it is given by

$$R_D = \left(\tau_0 \eta^2 R_0 l^2\right)^{1/3}, \tag{8}$$

where eq. (5) has been used. It turns out to be always smaller than the photospheric radius of photon thin outflow, $R_D \ll R_{ph}$, so the radiation escapes such an outflow before it becomes transparent, just like in the supernova ejecta. In this sense the characteristic radius of the photospheric emission is not the photospheric radius found from (4), but the radius of diffusion (8). The probability distribution of last scattering of photons in diffusive photospheres is qualitatively different from the usual photospheric emission.[26] Besides, the comoving temperature of escaping radiation is different from the temperature at the photospheric radius. Applicability of the photon thin asymptotics, last line in eq. (6), can be written using eq. (8) as

$$l \ll \frac{R_D}{\sqrt{2}\eta^2}. \tag{9}$$

For larger thickness l photon thin asymptotics disappears, so in the limit $l \to \infty$ the stationary wind with photon thick asymptotics is recovered.

Adiabatic expansion implies[36] that the observed temperature of the outflow does not change while it is accelerating, and it decreases as $T_{obs} \propto R^{-2/3}$ at the coasting phase. Taking into account the finite size of emitter and the cosmological redshift z one has[37]

$$T_{obs} = \frac{\xi}{1+z} T_0 \left(\frac{\eta R_0}{R_D} \right)^{2/3}, \tag{10}$$

where ξ is a numerical factor of order unity. In estimates below $\xi = 1.48$ is assumed following Ref. 37, which is found from the Monte Carlo simulations in the infinite wind approximation, though the value of ξ in the acceleration phase and in the photon thin case could be slightly different. The temperature at the base of the outflow is

$$T_0 \simeq \left(\frac{L}{4\pi c a R_0^2} \right)^{1/4}, \tag{11}$$

where $a = 4\sigma_{SB}/c$ is the radiation constant, σ_{SB} is the Stefan-Boltzmann constant. Finally, the duration of photospheric emission for a distant observer is

$$t_a^D = (1+z) \frac{R_D}{2\eta^2 c}. \tag{12}$$

In the photon thick case the duration of thermal emission is determined by the width of the outflow l, which is unconstrained. In the photon thin case this duration is given by eq. (12) and it is a function of the diffusion radius and the Lorentz factor.

The luminosity of photospheric component scales with radius as

$$L_{ph} = L_0 \left(\frac{\eta R_0}{R} \right)^{8/3}, \tag{13}$$

and the applicability condition of the photon thin case in eq. (6) and together with the definition of diffusion radius in eq. (8) imply for the luminosity of diffusive photosphere

$$L_{thin} < L_0 \left(\frac{R_0}{\eta l} \right)^{8/3} \ll L_0. \tag{14}$$

This means that thermal emission is much weaker than the emission of the prompt radiation, if γ-rays are produced with high efficiency there. For $l \sim R_0$ this condition strongly favours small values of η and, consequently, small Lorentz factors of the outflow.

4. Determination of initial radius and bulk Lorentz factor of the outflow

Assuming that the observed thermal component in early afterglows of some GRBs is of photospheric origin, one can estimate initial radius of the outflow directly from

observations.[37] Indeed, in the ultrarelativistic regime one has

$$\mathcal{R} \equiv \left(\frac{F_{obs}^{BB}}{\sigma_{SB} T_{obs}^4} \right)^{1/2} = \zeta \frac{(1+z)^2}{d_L} \frac{R}{\Gamma}, \tag{15}$$

where R is the emission radius, d_L is the luminosity distance, ζ is a numerical factor of order unity. Following Ref. 37 $\zeta = 1.06$ is assumed for the estimates below. In Ref. 37 the emission radius R was associated with the photosphere of the *photon thick* outflows. However, this relation is valid for any ultrarelativistic emitter. Therefore, from eqs. (10), (15) and (11) the initial radius is

$$R_0 = \frac{4^{3/2} d_L}{\xi^6 \zeta^4 (1+z)^2} \mathcal{R} \left(\frac{F_{obs}^{BB}}{Y F_{obs}} \right)^{3/2}. \tag{16}$$

In the derivation of this results only two assumptions are made. First, the outflow should be coasting at ultrarelativistic speed, $\Gamma = \eta \gg 1$. Secondly, the relation $L = 4\pi d_L^2 Y F_{obs}$ is used, where Y is the fraction of the total luminosity L and the energy emitted in X and γ-rays *in the prompt phase*.

In addition to the initial radius R_0 an equation for the Lorentz factor can be obtained. Since the emitter radius for the *photon thin* outflows is the diffusion radius $R = R_D$, from eqs. (8) and (15) one obtains, see also Ref. 38

$$\frac{\eta}{l} = \frac{\zeta^{3/2} (1+z)^3}{d_L^{1/2}} \left(\frac{\sigma Y F_{obs}}{m_p c^3 \mathcal{R}^3} \right)^{1/2}. \tag{17}$$

Therefore, the Lorentz factor can be determined if l is known. In the particular case $l = R_0$ from (17) it follows the minimum Lorentz factor for which the photon thin case applies

$$\eta_{thin} = (1+z) \left(d_L \frac{Y F_{obs} \sigma}{m_p c^3 \mathcal{R}} \right)^{1/2} \frac{4^{3/2}}{\zeta^{5/2} \xi^6} \left(\frac{F_{obs}^{BB}}{Y F_{obs}} \right)^{3/2}. \tag{18}$$

The condition (9) determines the applicability limit of the photon thin case, so for $\sqrt{2}\eta^2 l = R_D$ the photon thick case is recovered[37]

$$\eta_{thick} = \left[\zeta (1+z)^2 d_L \frac{\sigma Y F_{obs}}{m_p c^3 \mathcal{R}} \right]^{1/4}. \tag{19}$$

Equations (17) and (18) allow the determination of the bulk Lorentz factor of the *photon thin* outflow, provided the measurement of the total flux F_{obs} and the parameter $\mathcal{R}$. Comparing eqs. (18) and (19) leads to the conclusion that the Lorentz factor inferred from the photon thin asymptotics is typically smaller than the one of the photon thick case; besides, it contains the inverse of the Y parameter, unlike a factor $Y^{1/4}$ in the latter case.

It is important to stress that from the theoretical point of view, given the total luminosity and the initial radius of the outflow one cannot distinguish between photon thick and photon thin cases as both sets of parameters are possible with different η and photospheric radius. These parameters are related by eq. (15), therefore

independent observational information is required in order to differentiate between the two cases.

5. GRB cocoons

Consider typical parameters relevant for GRB jets which is penetrating the progenitor.[17] In what follows introduce the notation $A_x = A/10^x$, so the luminosity L_{52} stands for 10^{52} erg/s, which is the isotropic luminosity. While the entropy of the jet can take large values, the mixing between the progenitor and the jet lowers the entropy of the cocoon, so $\eta = 10$ is chosen, as a reference value. It is also likely that the entropy is a decreasing function of the angular distance from the jet. Assume that initial radius of the wind R_0 is given by the radius of the core of the progenitor WR star $R_0 \sim 10^9$ cm, and the thickness of the wind corresponds to the size of the WR star $l_{12} = 10^{12}$ cm.[39] The crucial parameter which determines whether the outflow is photon thick or thin is the ratio

$$\chi \simeq 29 L_{52} l_{12}^{-1} \eta_1^{-5}. \tag{20}$$

For $\chi \gg 1$ the outflow is photon thin, which is the case for our fiducial parameters. Considering the extreme dependence on η, for smaller η the condition is clearly satisfied. Hence the cocoon is in the photon thin regime and therefore the radiation from the cocoons is governed by radiative diffusion. The diffusion radius is

$$R_D \simeq 4.9 \times 10^{14} L_{52}^{1/3} \eta_1^{1/3} l_{12}^{2/3} \text{ cm}, \tag{21}$$

and the arrival time corresponding to this radius is

$$t_a^D = (1+z)\frac{R_D}{2\eta^2 c} \simeq 81.7(1+z) L_{52}^{1/3} \eta_1^{-5/3} l_{12}^{2/3} \text{ s}, \tag{22}$$

which is the typical duration of thermal emission observed in early afterglows of GRBs.

The observed temperature at the diffusion radius is

$$T_{obs} = 0.12 L_{52}^{1/36} R_9^{1/6} \eta_1^{4/9} l_{12}^{-4/9} \text{keV}, \tag{23}$$

which is also a typical temperature of thermal emission in early afterglows of GRBs.[13]

Such inferred values of temperature and duration call for closer attention to the radiation properties of photon thin outflows.

6. Case studies

All GRBs reported in Ref. 13 with measured redshifts and thermal component detected in their early afterglows were considered, namely GRBs 060218, 090618, 101219B, 111123A, 111225A, 121211A, 131030A, 150727A, 151027A. Observed temperature, thermal flux and total flux were averaged for the entire duration of the thermal emission. The initial radius was found from eq. (16). Two values of the

Lorentz factor in photon thin, eq. (18), and photon thick, eq. (19), cases were determined. Then the minimum value of the Y parameter is found which allows the duration of the photospheric emission (12) to be not less than the observed duration of the thermal component. Only six cases allow both photon thick and photon thin interpretations for the photosphere; for other cases photon thin case does not apply because eq. (19) gives smaller Lorentz factor than eq. (18).

GRB 060218. This is a well studied nearby GRB[15] with record breaking duration of the thermal signal interpreted as the break out of a shock driven by a mildly relativistic shell into the dense wind surrounding the progenitor, see however Refs. 40, 41. The thermal emission in this burst with observed temperature $T_{BB} = 0.15$ keV may be also explained as a photosphere of a cocoon launched from initial radius 3.18×10^{11} cm with a mildly relativistic Lorentz factor $1.2Y^{-1} < \Gamma < 1.6Y^{1/4}$ emitting in the photon thin regime. This estimate of the Lorentz factor is in agreement with radio observation at 2 days, requiring $\Gamma \sim 2$.[42] Given that the condition $\Gamma \gg 1$ is not satisfied, results of the theory of diffusive ultrarelativistic photospheres can be applied to this case with great care. In particular, the estimated duration of the thermal signal is only $13Y$ s.

GRB 090618. This burst may represent a canonical case of photon thin outflow launched from the initial radius 10^9 cm with the Lorentz factor $3Y^{-1} < \Gamma < 40Y^{1/4}$. Assuming instantaneous energy injection with $l = R_0$ one finds $Y = 5.7$. The duration of the thermal emission with observed temperature about 1 keV is about $6Y$ s. Note that thermal emission has also been claimed in the prompt phase, with a higher temperature ranging from 54 to 12 keV.[43] Such thermal emission in the prompt phase may be interpreted as a photosphere of the photon thick outflow. Indeed, if the initially high entropy η decreases with time the outflow should experience a transition from photon thick to photon thin case.

GRB 111225A. This case is similar to GRB 060218, but with smaller initial radius of 8.3×10^9 cm and Lorentz factor in the range $1 < \Gamma < 6.0Y^{1/4}$. The duration of the thermal emission with observed temperature of 0.18 keV is about $126Y$ s, with $Y = 2.0$ for $l = R_0$. The lower bound on the Lorentz factor in unconstrained. It may correspond to a cocoon emitting in the diffusive photon thin regime.

GRB 131030A. This case is a typical long burst, similar to GRB 090618, it allows Lorentz factors in the photon thin case in the following range: $4.3Y^{-1} < \Gamma < 66Y^{1/4}$. The duration of the thermal emission with observed temperature of 1.12 keV is about $2.0Y$ s. The initial radius is 3.74×10^8 cm. For instantaneous energy injection $Y = 19$.

GRB 150727A. This case is similar to GRB 111225A with initial radius 1.1×10^9 cm and Lorentz factors in the range $1 < \Gamma < 11Y^{1/4}$, allowing for a photon thin interpretation. The duration of the thermal emission with observed temperature of 0.47 keV is about $191Y$ s, with $Y = 2.1$.

GRB 151027A. This case is similar to GRBs 090618 and 131030, however with quite large initial radius 1.5×10^{10} cm and Lorentz factors in the narrow range

$23 < \Gamma < 26 Y^{1/4}$. The duration of the thermal emission with observed temperature of 0.96 keV is about $0.47Y$ s, with $Y = 64$.

7. Discussion

The difference between the present work and the approach followed in Ref. 16 should be emphasized. The main assumption of that work is the presence of unknown dissipation mechanism, which transforms part of the kinetic energy of the outflow into radiation, postulated in Ref. 44. Such dissipation can boost the luminosity of thermal emission and it might be required to explain subdominant thermal component during the prompt emission or even the prompt emission itself, see Ref. 45. Concerning observations of this component in the early afterglow, dissipation is not required, as it is much weaker than the prompt radiation.

Similarly, there is a difference between the present work and the work in Ref. 17. There two regimes of expansion are considered: Newtonian ($v < c$) and ultrarelativistic $\eta \gg 1$. For the latter, which is of interest here, the emission was assumed to originate at the photospheric radius, given by the last line in eq. (6). As discussed in Sec. 3 above, photons in this case diffuse out much earlier, so that no photons are left in the outflow when it arrives to the photospheric radius. For this reason estimations of the luminosity and observed temperature in that paper cannot be used.

Qualitative difference in dependence of observed flux and temperature on time for photon thin outflows determine their observed properties. In particular, since the flux up to diffusion time (22) is almost constant and its luminosity is much weaker than the prompt radiation luminosity, see eq. (14), the thermal component become visible after the steep decrease of observed luminosity following the end of the prompt phase. This is indeed where such thermal component is identified in many GRBs. Its disappearance is naturally explained by diffusion of the radiation kept in the outflow. Hence it implies that no more photons are generated neither in the outflow nor in the central engine.

We show that in several GRBs, namely GRB 060218, 111225A and 150727A, the thermal component observed in the early afterglow may originate from mildly relativistic cocoons emerging from the progenitors together with the jet, due to relatively small values of inferred Lorentz factors $\Gamma < 10$.

Such emission observed in other GRBs such as 090618, 131030A and 151027A correspond to large Lorentz factors $\Gamma > 10$, indicating a jet origin of the photospheric emission. Besides, these results suggest that the progenitors of some long GRBs, in particular GRB 090618 and 131030A could be rather compact objects, with radius $l \sim 10^9$ cm.

It is important to stress that the relatively low temperature of the thermal component observed in the early afterglow with $T \sim 0.1 - 10$ keV, in contrast with typical temperatures detected during the prompt emission with $T \sim 10 - 100$ keV does not indicate small Lorentz factor of the outflow. Conversely, it may point to

photospheric origin of the thermal emission in the photon thin regime. Instead, large Lorentz factors $\Gamma \gg 1$ assumed in the model imply small mass of the emitting plasma, which is consistent with the cocoon interpretation.[17]

Notice that possible presence of thermal components both in the prompts radiation and in the early afterglow, as well as the presence of breaks in temperature dependence on time found in many cases[46–48] may correspond to the transition from photon thick to photon thin asymptotics in hydrodynamic evolution of the outflow powering GRBs.

In this work the theory of diffusive emission from relativistic photospheres is confronted with observational data on a sample of GRBs with thermal component in the early afterglows. The measurement of temperature and flux of the thermal component along with the total flux in the prompt phase are used to determine initial radii of the outflows as well as their Lorentz factors. The results indicate that in several cases (GRBs 060218, 111225A and 150727A) the inferred Lorentz factors are relatively small, $\Gamma < 10$, while in other cases (GRBs 090618, 131030A and 151027A) the inferred Lorentz factors are larger, $\Gamma > 10$. Such differences suggest two possible sources of the thermal component: mildly relativistic cocoons or highly relativistic jets. Results obtained above are valid only for those cases, where inferred Lorentz factor is relatively small, below few tens. For other cases identified in Ref. 13 inferred Lorentz factors are larger, and photon thin interpretation does not apply.

These results are the first indication that radiative diffusion may play an important role not only in nonrelativistic outflows, but also in ultrarelativistic outflows, represented by GRBs.

References

1. C. Kouveliotou, C. A. Meegan, G. J. Fishman, N. P. Bhat, M. S. Briggs, T. M. Koshut, W. S. Paciesas and G. N. Pendleton, Identification of Two Classes of Gamma-Ray Bursts, *ApJ* **413**, p. L101 (August 1993).
2. B. Zhang, *The Physics of Gamma-Ray Bursts* (Cambridge University Press, 2018).
3. P. Meszaros and M. J. Rees, Relativistic Fireballs and Their Impact on External Matter: Models for Cosmological Gamma-Ray Bursts, *ApJ* **405**, p. 278 (March 1993).
4. G. V. Vereshchagin, Physics of Nondissipative Ultrarelativistic Photospheres, *International Journal of Modern Physics D* **23**, 1430003 (Dec 2014).
5. A. Pe'er and F. Ryde, Photospheric emission in gamma-ray bursts, *International Journal of Modern Physics D* **26**, 1730018 (Jan 2017).
6. B. Zhang, Y. Wang and L. Li, Dissecting the Energy Budget of a Gamma-Ray Burst Fireball, *ApJ* **909**, p. L3 (March 2021).
7. F. Ryde, M. Axelsson, B. B. Zhang, S. McGlynn, A. Pe'er, C. Lundman, S. Larsson, M. Battelino, B. Zhang, E. Bissaldi, J. Bregeon, M. S. Briggs, J. Chiang, F. de Palma, S. Guiriec, J. Larsson, F. Longo, S. McBreen, N. Omodei, V. Petrosian, R. Preece and A. J. van der Horst, Identification and Properties of the Photospheric Emission in GRB090902B, *ApJ* **709**, L172 (February 2010).
8. F. Ryde, C. Lundman and Z. Acuner, Emission from accelerating jets in gamma-ray bursts: radiation-dominated flows with increasing mass outflow rates, *MNRAS* **472**, 1897 (December 2017).

9. K. L. Page, R. L. C. Starling, G. Fitzpatrick, S. B. Pandey, J. P. Osborne, P. Schady, S. McBreen, S. Campana, T. N. Ukwatta, C. Pagani, A. P. Beardmore and P. A. Evans, GRB 090618: detection of thermal X-ray emission from a bright gamma-ray burst, *Monthly Notices of the Royal Astronomical Society* **416**, 2078 (September 2011).
10. R. L. C. Starling, K. L. Page, A. Pe'er, A. P. Beardmore and J. P. Osborne, A search for thermal X-ray signatures in gamma-ray bursts - I. Swift bursts with optical supernovae, *Monthly Notices of the Royal Astronomical Society* **427**, 2950 (December 2012).
11. M. Sparre and R. L. C. Starling, A search for thermal X-ray signatures in gamma-ray bursts - II. The Swift sample, *Monthly Notices of the Royal Astronomical Society* **427**, 2965 (December 2012).
12. M. Friis and D. Watson, Thermal Emission in the Early X-Ray Afterglows of Gamma-Ray Bursts: Following the Prompt Phase to Late Times, *The Astrophysical Journal* **771**, p. 15 (July 2013).
13. V. Valan, J. Larsson and B. Ahlgren, Thermal components in the early X-ray afterglows of GRBs: likely cocoon emission and constraints on the progenitors, *MNRAS* **474**, 2401 (February 2018).
14. L. Izzo, A. de Ugarte Postigo, K. Maeda, C. C. Thöne, D. A. Kann, M. Della Valle, A. Sagues Carracedo, M. J. Michałowski, P. Schady, S. Schmidl, J. Selsing, R. L. C. Starling, A. Suzuki, K. Bensch, J. Bolmer, S. Campana, Z. Cano, S. Covino, J. P. U. Fynbo, D. H. Hartmann, K. E. Heintz, J. Hjorth, J. Japelj, K. Kamiński, L. Kaper, C. Kouveliotou, M. KruŻyński, T. Kwiatkowski, G. Leloudas, A. J. Levan, D. B. Malesani, T. Michałowski, S. Piranomonte, G. Pugliese, A. Rossi, R. Sánchez-Ramírez, S. Schulze, D. Steeghs, N. R. Tanvir, K. Ulaczyk, S. D. Vergani and K. Wiersema, Signatures of a jet cocoon in early spectra of a supernova associated with a γ-ray burst, *Nature* **565**, 324 (January 2019).
15. S. Campana, V. Mangano, A. J. Blustin, P. Brown, D. N. Burrows, G. Chincarini, J. R. Cummings, G. Cusumano, M. Della Valle, D. Malesani, P. Mészáros, J. A. Nousek, M. Page, T. Sakamoto, E. Waxman, B. Zhang, Z. G. Dai, N. Gehrels, S. Immler, F. E. Marshall, K. O. Mason, A. Moretti, P. T. O'Brien, J. P. Osborne, K. L. Page, P. Romano, P. W. A. Roming, G. Tagliaferri, L. R. Cominsky, P. Giommi, O. Godet, J. A. Kennea, H. Krimm, L. Angelini, S. D. Barthelmy, P. T. Boyd, D. M. Palmer, A. A. Wells and N. E. White, The association of GRB 060218 with a supernova and the evolution of the shock wave, *Nature* **442**, 1008 (August 2006).
16. A. Pe'er, P. Meszaros and M. J. Rees, Radiation from an Expanding Cocoon as an Explanation of the Steep Decay Observed in GRB Early Afterglow Light Curves, *The Astrophysical Journal* **652**, 482 (November 2006).
17. E. Nakar and T. Piran, The Observable Signatures of GRB Cocoons, *ApJ* **834**, p. 28 (January 2017).
18. R. Ruffini, G. V. Vereshchagin and Y. Wang, Thermal emission in the early afterglow of gamma-ray bursts from their interaction with supernova ejecta, *A&A* **600**, p. A131 (April 2017).
19. P. Mészáros and M. J. Rees, Steep Slopes and Preferred Breaks in Gamma-Ray Burst Spectra: The Role of Photospheres and Comptonization, *ApJ* **530**, 292 (February 2000).
20. A. Pe'er, Temporal Evolution of Thermal Emission from Relativistically Expanding Plasma, *ApJ* **682**, 463 (July 2008).
21. A. Pe'er and F. Ryde, A Theory of Multicolor Blackbody Emission from Relativistically Expanding Plasmas, *ApJ* **732**, 49 (May 2011).

22. A. M. Beloborodov, Radiative Transfer in Ultrarelativistic Outflows, *ApJ* **737**, 68 (August 2011).
23. C. Lundman, A. Pe'er and F. Ryde, A theory of photospheric emission from relativistic, collimated outflows, *MNRAS* **428**, 2430 (January 2013).
24. R. Santana, P. Crumley, R. A. Hernández and P. Kumar, Monte Carlo simulations of the photospheric process, *MNRAS* **456**, 1049 (February 2016).
25. M. Bhattacharya, W. Lu, P. Kumar and R. Santana, Monte Carlo Simulations of Photospheric Emission in Relativistic Outflows, *ApJ* **852**, p. 24 (January 2018).
26. D. Bégué, I. A. Siutsou and G. V. Vereshchagin, Monte Carlo Simulations of the Photospheric Emission in Gamma-Ray Bursts, *ApJ* **767**, p. 139 (Apr 2013).
27. R. Ruffini, I. A. Siutsou and G. V. Vereshchagin, A Theory of Photospheric Emission from Relativistic Outflows, *The Astrophysical Journal* **772**, p. 11 (July 2013).
28. T. Piran, A. Shemi and R. Narayan, Hydrodynamics of Relativistic Fireballs, *MNRAS* **263**, 861 (August 1993).
29. P. Mészáros, P. Laguna and M. J. Rees, Gasdynamics of relativistically expanding gamma-ray burst sources - Kinematics, energetics, magnetic fields, and efficiency, *ApJ* **415**, 181 (September 1993).
30. R. Ruffini, I. A. Siutsou and G. V. Vereshchagin, Spreading of ultrarelativistically expanding shell: An application to GRBs, *New Astronomy* **27**, 30 (Feb 2014).
31. D. Arnett, *Supernovae and Nucleosynthesis: An Investigation of the History of Matter from the Big Bang to the Present* (Princeton University Press, 1996).
32. G. V. Vereshchagin and I. A. Siutsou, Diffusive photospheres in gamma-ray bursts, *MNRAS* **494**, 1463 (May 2020).
33. M. A. Abramowicz, I. D. Novikov and B. Paczynski, The appearance of highly relativistic, spherically symmetric stellar winds, *ApJ* **369**, 175 (March 1991).
34. A. G. Aksenov, R. Ruffini and G. V. Vereshchagin, Thermalization of the mildly relativistic plasma, *Phys. Rev. D* **79**, p. 043008 (February 2009).
35. A. G. Aksenov, R. Ruffini and G. V. Vereshchagin, Thermalization of Electron-Positron-Photon Plasmas with an application to GRB, in *American Institute of Physics Conference Series*, American Institute of Physics Conference Series Vol. 966, January 2008.
36. R. Ruffini and G. Vereshchagin, Electron-positron plasma in GRBs and in cosmology, *Il Nuovo Cimento* **C 36**, p. 255 (May 2013).
37. A. Pe'er, F. Ryde, R. A. M. J. Wijers, P. Mészáros and M. J. Rees, A New Method of Determining the Initial Size and Lorentz Factor of Gamma-Ray Burst Fireballs Using a Thermal Emission Component, *ApJ* **664**, L1 (Jul 2007).
38. D. Bégué and S. Iyyani, Transparency Parameters from Relativistically Expanding Outflows, *ApJ* **792**, p. 42 (September 2014).
39. P. A. Crowther, Physical Properties of Wolf-Rayet Stars, *ARA&A* **45**, 177 (September 2007).
40. G. Ghisellini, G. Ghirlanda and F. Tavecchio, Did we observe the supernova shock breakout in GRB 060218?, *Monthly Notices of the Royal Astronomical Society: Letters* **382**, L77 (November 2007).
41. S. W. K. Emery, M. J. Page, A. A. Breeveld, P. J. Brown, N. P. M. Kuin, S. R. Oates and M. De Pasquale, The early optical afterglow and non-thermal components of GRB 060218, *MNRAS* **484**, 5484 (April 2019).
42. A. M. Soderberg, S. R. Kulkarni, E. Nakar, E. Berger, P. B. Cameron, D. B. Fox, D. Frail, A. Gal-Yam, R. Sari, S. B. Cenko, M. Kasliwal, R. A. Chevalier, T. Piran, P. A. Price, B. P. Schmidt, G. Pooley, D. S. Moon, B. E. Penprase, E. Ofek, A. Rau,

N. Gehrels, J. A. Nousek, D. N. Burrows, S. E. Persson and P. J. McCarthy, Relativistic ejecta from X-ray flash XRF 060218 and the rate of cosmic explosions, *Nature* **442**, 1014 (August 2006).
43. L. Izzo, R. Ruffini, A. V. Penacchioni, C. L. Bianco, L. Caito, S. K. Chakrabarti, J. A. Rueda, A. Nandi and B. Patricelli, A double component in GRB 090618: a proto-black hole and a genuinely long gamma-ray burst, *A&A* **543**, p. A10 (Jul 2012).
44. M. J. Rees and P. Mészáros, Dissipative Photosphere Models of Gamma-Ray Bursts and X-Ray Flashes, *ApJ* **628**, 847 (August 2005).
45. M. Bhattacharya and P. Kumar, Explaining GRB prompt emission with sub-photospheric dissipation and Comptonization, *MNRAS* **491**, 4656 (February 2020).
46. F. Ryde, The Cooling Behavior of Thermal Pulses in Gamma-Ray Bursts, *ApJ* **614**, 827 (October 2004).
47. F. Ryde, Is Thermal Emission in Gamma-Ray Bursts Ubiquitous?, *ApJ* **625**, L95 (June 2005).
48. F. Ryde and A. Pe'er, Quasi-blackbody Component and Radiative Efficiency of the Prompt Emission of Gamma-ray Bursts, *ApJ* **702**, 1211 (September 2009).

Summary of the parallel session GB3

G. V. Vereshchagin
ICRANet, p. della Repubblica, 10, 65100 Pescara, Italy
E-mail: veresh@icra.it

D. Bégué
Bar-llan University Ramat-Gan 5290002, Israel
E-mail: begueda@biu.ac.il

In this proceedings, we give a brief general overview of the parallel session on photospheric emission in gamma-ray bursts (GRBs), which took place during the online 16$^{\text{th}}$ Marcel Grossmann Meeting. The session covered theoretical and observational aspects of photospheric emission. In particular, spectral, temporal and polarization properties were discussed.

Keywords: Gamma-ray bursts, Photospheric emission, Thermal emission, Non-thermal emission, Jet simulations.

1. State of the art

Photospheric emission models are seen as prime contenders to explain the prompt phase of gamma-ray bursts, see *e.g.*[1] for a review. This emission is produced when the expanding jet becomes optically thin to Compton scattering,[2] which is the dominant interaction mode for photon of comoving energy $\sim$ 1keV. Photospheric emission is a naturally expected process in the widely discussed fireball model.[3–6] It can be highly efficient if the transparency takes place at small radius, that is to say close to the transition radius between accelerating and coasting phase,[3,7–9] and it naturally predicts a clustering of the spectral peak energy around 500 keV.[10]

It has long been known that photospheric models in their simplest flavors fail to explain the shallow low energy spectral slope seen in most GRBs.[11–15] This picture changed in the past 15 years when it was realized that the spectra produced at the photosphere can be highly non-thermal due to 1- energy dissipation below the photosphere either via shocks,[16–18] magnetic reconnection[19–22] or neutron decay,[23] and 2- geometrical effects due to the structure of the jets[24,25] and the photon last scattering position.[11,14,23]

The aforementioned effects are difficult to study analytically, as any analysis depends on ad hoc assumptions on the geometry of the jet, dissipation mechanisms and their localization. In recent years, 3D (GR) MHD simulations coupled to Monte-Carlo radiative transfer calculations were designed to study photospheric emission

in the context of GRBs solve those opens issues.[26–33] One of the most interesting finding is that the luminosity drop for off-axis observers is limited by the structure of the jet.

Moreover, the simulations were successful in reproducing correlations between observer position, luminosity, peak energy, and observed polarization. In particular, several important successes were obtained in explaining the Amati correlation[34] (between the isotropic total energy $E_{\rm iso}$ and the peak energy $E_{\rm p}$), the Yonetoku correlation[35] (between the isotropic luminosity $L_{\rm iso}$ and $E_{\rm p}$) and the Goletneskii relation[36] (between the luminosity and $E_{\rm p}$) by considering a distribution of off-axis observers.

Finally, the simulations permitted to obtain detailed polarization predictions. In particular, it was found from numerical simulations of photon propagation inside a jet that a high degree of time-resolved polarization requires off-axis observers, while the polarization for on-axis observers is lower.[37–39] Using a semi-analytical jet model, it was found that the polarization degree correlates to a substantial drop in luminosity.[37,40] Instead, recent hydrodynamical simulations[38,39] showed that the drop in luminosity for observer angles might be smaller than expected thanks to emission from the jet cocoon. Extending the analysis to lower frequencies it was shown that the polarization degree is larger at low energies, either thanks to the effect of synchrotron emission[41] or due to higher scattering angle towards the observer.[39]

Despite these many successes, the correlation analysis is limited to a few initial setups because of the large resources needed for each independent simulations and their post-processing analysis. In the forthcoming years, the reliability of the already obtained correlation results will be tested and refined against many more runs with different initial conditions and methods. In addition, the simulations are not yet mature enough to produce detailed spectral predictions, which could be confronted to observed data.

Concerning the spectral analysis in the context of photospheric emission, two different paths were explored in the past few years. First, detailed time-resolved spectral analysis of GBM data from many bursts were performed.[42–44] Specifically, the spectra were fitted with a cut-off power-law and the obtained low energy slope distribution were carefully investigated. It was realized that many bursts have one to several time bins with spectra harder than the limits from the synchrotron line of death. These studies concluded that photospheric emission should be the mechanism at work in those bursts, textbfsee however[45].

Second, simplified photospheric models, such as the DREAM model,[46,47] were directly fitted to observed GBM spectra[48] finding 1- good agreement for the bursts with the hardest low energy slope when considering the photospheric emission taking place at the transition between accelerating and coasting phase, and 2- good overall agreement with other bright bursts, when considering localized energy dissipation. It is expected that those studies will be extended to larger sample and refined models in the near future.

2. Talks at the session

The session was run in two blocks of three hours.

Asaf Pe'er delivered a review talk "Understanding prompt emission: where do we stand?" where he summarized basic observational information about the prompt emission and problems originating from misconceptions which are related to usage of models (including the progenitor, jet launching, dynamics and dissipation), the nature and dynamics of the outflow (e.g. baryonic vs. magnetic) and the radiation mechanisms (e.g. photospheric vs. synchrotron). He also discussed the spectral and polarization properties of non-dissipative photospheric emission, and its ability to explain observed correlation. Finally, he mentioned a new model of the prompt emission based on backscattering of the seed photons on a cold relativistic cork which appears natually in front of the jet in many progenitor models.

Ore Gottlieb presented a talk "Probing the jet launching mechanism from prompt emission of GRBs" where he described the ways to infer the information on the jet launching mechanism and its dynamics from the comparison of observations with numerical simulations of jet propagation and its the photospheric emission, focusing specifically on the role of magnetization of the jet. In particular, based on numerical hydrodynamic simulations, including self-consistent simulations within the collapsar model, he discussed and ability to produce jet tilt, intermittency, position of the magnetic dissipation region, and the role of different progenitors in jet launching and cocoon formation.

Gregory Vereshchagin reported a study "Diffusive photospheres in gamma-ray bursts" describing the results of the application of the theory of photospheric emission to the thermal radiation, detected in the early afterglows of some gamma-ray bursts. The inferred Lorentz factors of the outflow are clustered in two groups: few hundreds, indicating classical photospheric emission from ultrarelativistic outflow, and few tens, indicating diffusive regime of the photospheric emission, possibly from jet cocoons.

Tyler Parsotan delivered a talk "Monte Carlo Simulations of Photospheric Emission in Gamma Ray Bursts" in which he described the MCRaT code and its coupling to relativistic hydrodynamical simulations in order to obtain mock observations. He discussed how his code allows him to obtain predictions of the photospheric model, in particular prediction about optical prompt percursors of gamma-ray bursts.

Hirotaka Ito presented results on "Numerical simulations of photospheric emission in GRBs" focusing on the jet structure and its consequences for spectral shapes, polarization signal and correlations between observed quantities, based on hydrodynamical simulations and post processing with the radiative transport.

Yan-Zhi Meng reported the talk "Photosphere emission spectrum of hybrid relativistic outflow for gamma-ray bursts", considering both impulsive injection and continuous wind cases.

J. Michael Burgess in his talk "Spectroscopy of GRBs: Where are we now?" discussed how fitting physical models, instead of phenomenological ones, is important

in order to infer physical information from observations. He mentioned several "myths" about the synchrotron emission model such as the line-of-death and width problem and showed how proper folding through the response function of the instruments demonstrates viability of this model.

Hüsne Dereli-Bégué presented a talk "Classification of Photospheric Emission in sGRBs" where she discussed the fit of a large sample of short GRBs with a phenomenological model based on a non-dissipative photosphere model. The main conclusions are that nearly one third of the spectra are consistent with purely thermal emission, and that a large fraction of bursts may come from subphotospheric dissipation.

Björn Ahlgren discussed "Subphotospheric dissipation evaluated using joint Fermi-Swift observations" and pointed out the importance of using XRT data in addition to the BAT data from the SWIFT satellite, in order to distinguish between photospheric and synchrotron mechanisms of emission in the context of fitting of dissipative photospheric model to the data, focusing specifically on the DREAM model.

Liang Li reported on "Bayesian Time-resolved Spectroscopy of Multipulse GRBs: Variations of Emission Properties among Pulses" where he discussed construction of the catalogue of multipulse bursts, the fitting procedure and the outcomes. In particular, he discussed different trends and correlations between observed quantities such as $F - \alpha$ and $F - E_p$.

The outcome of the discussion with participation of Gregory Vereshchagin, Pe'er Asaf, J. Michael Burgess and Damien Begue is that generally speaking about gamma-ray bursts, despite strong progress both in theory and in observations there are still many unresolved issues such as the origin of the jet, the mechanism of energy dissipation, and the nature of observed variability.

The field is observationally driven, with new recent observational results such as the "anomalous" X-ray behavior of GRB 170817A and the detection of very high energy emission from several sources such as GRB 190114C. However, there are some key observational results, which still need explanation from the theory, in particular the strong spectral evolution within each pulse, the power-law decay of the peak energy, the time evolution of the temperature and so on. From the observational viewpoint, one of the most promising direction is the study of polarization of gamma-ray bursts.

3. Conclusions

To conclude, many critical progresses were made and problem addressed in the past few years. In particular, theoretical and numerical predictions start to be mature enough to allow direct comparison to observations, either via direct spectral fitting or via comparison to the global population parameters. It is expected that this trend will continue, allowing for a refinement of the photospheric emission theories, a better understanding of the data and of gamma-ray bursts.

References

1. A. Pe'er, Physics of Gamma-Ray Bursts Prompt Emission, *Advances in Astronomy* **2015**, p. 907321 (2015).
2. M. A. Abramowicz, I. D. Novikov and B. Paczynski, The appearance of highly relativistic, spherically symmetric stellar winds, *Astrophys. J.* **369**, 175 (March 1991).
3. J. Goodman, Are gamma-ray bursts optically thick?, *Astrophys. J.* **308**, L47 (September 1986).
4. B. Paczynski, Gamma-ray bursters at cosmological distances, *Astrophys. J.* **308**, L43 (September 1986).
5. M. J. Rees and P. Meszaros, Relativistic fireballs - Energy conversion and time-scales, *Mon. Not. R. Astron. Soc.* **258**, 41P (September 1992).
6. M. J. Rees and P. Meszaros, Unsteady outflow models for cosmological gamma-ray bursts, *Astrophys. J.* **430**, L93 (August 1994).
7. F. Ryde and A. Pe'er, Quasi-blackbody Component and Radiative Efficiency of the Prompt Emission of Gamma-ray Bursts, *Astrophys. J.* **702**, 1211 (September 2009).
8. D. Bégué and S. Iyyani, Transparency Parameters from Relativistically Expanding Outflows, *Astrophys. J.* **792**, p. 42 (September 2014).
9. F. Ryde, C. Lundman and Z. Acuner, Emission from accelerating jets in gamma-ray bursts: radiation-dominated flows with increasing mass outflow rates, *Mon. Not. R. Astron. Soc.* **472**, 1897 (December 2017).
10. A. Pe'er, P. Mészáros and M. J. Rees, Peak Energy Clustering and Efficiency in Compact Objects, *Astrophys. J.* **635**, 476 (December 2005).
11. A. Pe'er, Temporal Evolution of Thermal Emission from Relativistically Expanding Plasma, *Astrophys. J.* **682**, 463 (Jul 2008).
12. A. M. Beloborodov, Radiative Transfer in Ultrarelativistic Outflows, *Astrophys. J.* **737**, p. 68 (August 2011).
13. R. Ruffini, I. A. Siutsou and G. V. Vereshchagin, A Theory of Photospheric Emission from Relativistic Outflows, *Astrophys. J.* **772**, p. 11 (July 2013).
14. D. Bégué, I. A. Siutsou and G. V. Vereshchagin, Monte Carlo Simulations of the Photospheric Emission in Gamma-Ray Bursts, *Astrophys. J.* **767**, p. 139 (April 2013).
15. W. Deng and B. Zhang, Low Energy Spectral Index and E_p Evolution of Quasi-thermal Photosphere Emission of Gamma-Ray Bursts, *Astrophys. J.* **785**, p. 112 (April 2014).
16. A. Pe'er and E. Waxman, Prompt Gamma-Ray Burst Spectra: Detailed Calculations and the Effect of Pair Production, *Astrophys. J.* **613**, 448 (September 2004).
17. C. Lundman, A. Beloborodov and I. Vurm, Radiation mediated shocks in gamma-ray bursts: Pair creation, *ArXiv e-prints* (August 2017).
18. H. Ito, A. Levinson, B. E. Stern and S. Nagataki, Monte Carlo simulations of relativistic radiation-mediated shocks - I. Photon-rich regime, *Mon. Not. R. Astron. Soc.* **474**, 2828 (February 2018).
19. G. Drenkhahn and H. C. Spruit, Efficient acceleration and radiation in Poynting flux powered GRB outflows, *Astron. Astrophys.* **391**, 1141 (September 2002).
20. D. Giannios, Prompt emission spectra from the photosphere of a GRB, *Astron. Astrophys.* **457**, 763 (October 2006).
21. D. Giannios and D. A. Uzdensky, GRB and blazar jets shining through their stripes, *Mon. Not. R. Astron. Soc.* **484**, 1378 (March 2019).
22. Y.-Z. Meng, J.-J. Geng and X.-F. Wu, The Photosphere Emission Spectrum of Hybrid Relativistic Outflow for Gamma-ray Bursts, *arXiv e-prints* , p. arXiv:2107.04532 (July 2021).
23. A. M. Beloborodov, Collisional mechanism for gamma-ray burst emission, *Mon. Not. R. Astron. Soc.* **407**, 1033 (September 2010).

24. C. Lundman, A. Pe'er and F. Ryde, A theory of photospheric emission from relativistic, collimated outflows, *Mon. Not. R. Astron. Soc.* **428**, 2430 (January 2013).
25. Y.-Z. Meng, L.-D. Liu, J.-J. Wei, X.-F. Wu and B.-B. Zhang, The Time-resolved Spectra of Photospheric Emission from a Structured Jet for Gamma-Ray Bursts, *Astrophys. J.* **882**, p. 26 (September 2019).
26. D. Lazzati, Monte Carlo Radiation Transfer Simulations of Photospheric Emission in Long-duration Gamma-ray Bursts, *Astrophys. J.* **829**, p. 76 (October 2016).
27. A. Chhotray and D. Lazzati, Dynamic Monte Carlo Simulations of Radiatively Accelerated GRB Fireballs, *ArXiv e-prints* (May 2017).
28. T. Parsotan and D. Lazzati, A Monte Carlo Radiation Transfer Study of Photospheric Emission in Gamma-Ray Bursts, *Astrophys. J.* **853**, p. 8 (January 2018).
29. T. Parsotan, D. López-Cámara and D. Lazzati, Photospheric Emission from Variable Engine Gamma-Ray Burst Simulations, *Astrophys. J.* **869**, p. 103 (December 2018).
30. O. Gottlieb, A. Levinson and E. Nakar, Intermittent hydrodynamic jets in collapsars do not produce GRBs, *arXiv e-prints* , p. arXiv:2002.12384 (February 2020).
31. H. Ito, O. Just, Y. Takei and S. Nagataki, A Global Numerical Model of the Prompt Emission in Short Gamma-ray Bursts, *Astrophys. J.* **918**, p. 59 (September 2021).
32. O. Gottlieb, O. Bromberg, A. Levinson and E. Nakar, Intermittent mildly magnetized jets as the source of GRBs, *Mon. Not. R. Astron. Soc.* **504**, 3947 (July 2021).
33. O. Gottlieb, A. Lalakos, O. Bromberg, M. Liska and A. Tchekhovskoy, Black hole to breakout: 3D GRMHD simulations of collapsar jets reveal a wide range of transients, *arXiv e-prints* , p. arXiv:2109.14619 (September 2021).
34. L. Amati, F. Frontera, M. Tavani, J. J. M. in't Zand, A. Antonelli, E. Costa, M. Feroci, C. Guidorzi, J. Heise, N. Masetti, E. Montanari, L. Nicastro, E. Palazzi, E. Pian, L. Piro and P. Soffitta, Intrinsic spectra and energetics of BeppoSAX Gamma-Ray Bursts with known redshifts, *Astron. Astrophys.* **390**, 81 (July 2002).
35. D. Yonetoku, T. Murakami, T. Nakamura, R. Yamazaki, A. K. Inoue and K. Ioka, Gamma-Ray Burst Formation Rate Inferred from the Spectral Peak Energy-Peak Luminosity Relation, *Astrophys. J.* **609**, 935 (July 2004).
36. S. V. Golenetskii, E. P. Mazets, R. L. Aptekar and V. N. Ilinskii, Correlation between luminosity and temperature in γ-ray burst sources, *Nature* **306**, 451 (December 1983).
37. C. Lundman, A. Pe'er and F. Ryde, Polarization properties of photospheric emission from relativistic, collimated outflows, *Mon. Not. R. Astron. Soc.* **440**, 3292 (June 2014).
38. T. Parsotan, D. López-Cámara and D. Lazzati, Photospheric Polarization Signatures from Long Gamma-Ray Burst Simulations, *Astrophys. J.* **896**, p. 139 (June 2020).
39. T. Parsotan and D. Lazzati, Photospheric Prompt Emission From Long Gamma Ray Burst Simulations – II. Spectropolarimetry, *arXiv e-prints* , p. arXiv:2108.11412 (August 2021).
40. R. Gill, J. Granot and P. Kumar, Linear polarization in gamma-ray burst prompt emission, *Mon. Not. R. Astron. Soc.* **491**, 3343 (January 2020).
41. C. Lundman, I. Vurm and A. M. Beloborodov, Polarization of Gamma-Ray Bursts in the Dissipative Photosphere Model, *Astrophys. J.* **856**, p. 145 (April 2018).
42. H.-F. Yu, H. Dereli-Bégué and F. Ryde, Bayesian Time-resolved Spectroscopy of GRB Pulses, *Astrophys. J.* **886**, p. 20 (November 2019).
43. H. Dereli-Bégué, A. Pe'er and F. Ryde, Classification of Photospheric Emission in Short GRBs, *Astrophys. J.* **897**, p. 145 (July 2020).
44. Z. Acuner, F. Ryde, A. Pe'er, D. Mortlock and B. Ahlgren, The Fraction of Gamma-Ray Bursts with an Observed Photospheric Emission Episode, *Astrophys. J.* **893**, p. 128 (April 2020).

45. J. M. Burgess, D. Bégué, J. Greiner, D. Giannios, A. Bacelj and F. Berlato, Gamma-ray bursts as cool synchrotron sources, *Nature Astronomy* **4**, 174 (February 2020).
46. B. Ahlgren, J. Larsson, T. Nymark, F. Ryde and A. Pe'er, Confronting GRB prompt emission with a model for subphotospheric dissipation, *Mon. Not. R. Astron. Soc.* **454**, L31 (November 2015).
47. B. Ahlgren, J. Larsson, E. Ahlberg, C. Lundman, F. Ryde and A. Pe'er, Testing a model for subphotospheric dissipation in GRBs: fits to Fermi data constrain the dissipation scenario, *Mon. Not. R. Astron. Soc.* **485**, 474 (May 2019).
48. G. Vianello, R. Gill, J. Granot, N. Omodei, J. Cohen-Tanugi and F. Longo, The Bright and the Slow—GRBs 100724B and 160509A with High-energy Cutoffs at $\lesssim$100 MeV, *Astrophys. J.* **864**, p. 163 (September 2018).

Synchrotron and synchrotron self-Compton emission components in GRBs detected at very high energies

Jagdish C. Joshi* and Vikas Chand†

School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China
Key laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing 210093, China
**E-mail: jagdishcjoshi86@gmail.com*
†E-mail: vikasK2@nju.edu.cn

Soebur Razzaque

Centre for Astro-Particle Physics (CAPP) and Department of Physics, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa
E-mail: srazzaque@uj.ac.za

Gamma-Ray Bursts (GRBs) are energetic transients originating in a violent explosion of a massive star or merger of two compact objects in a binary system. These explosions create a relativistic blastwave which inevitably collides with the circumburst medium and results in external shocks. The emission produced thereby is the afterglow observed in GRBs after the initial prompt phase. The properties of the emitting region, i.e., non-thermal particle spectrum, magnetic amplification, and micro-physical parameters, etc., can be probed by monitoring and modelling the afterglow radiation. The recent detection of very high energy (VHE, > 100 GeV) gamma rays from GRBs has opened a possibility to test theoretical models such as the synchrotron self-Compton (SSC) in GRBs till late times in the afterglow phase. We find that the SSC model explains the afterglow observations of MAGIC-detected GRB 190114C and H.E.S.S.-detected GRB 180720B.

Keywords: GRBs: VHE Components, Afterglow Emission.

1. Introduction

GRBs, one of the brightest electromagnetic events in the universe, are produced in cataclysmic astrophysical events like the collapse of a massive star or mergers of compact binary stars (a neutron star – neutron star or a neutron star and a black hole system).[1,2] The initial prompt burst of the energy is primarily in gamma-rays, and its origin though still debated is linked to the energy released from the central source which is dissipated at the internal shocks or emitted at the photosphere.[3–6] Afterglow emission follows once the prompt emission phase ends in GRBs. The afterglow of GRBs is important to understand the circumburst environment, jet parameters, geometry of the jet and the central engine. Recently, a VHE component is discovered by the ground-based MAGIC telescopes[7–9] and H.E.S.S. telescopes.[10,11] This was anticipated based on the standard afterglow models and its origin was postulated via the inverse Compton scattering of the synchrotron radiation (synchrotron self-Compton).[12–22] Comparatively, other radiation channels like proton

synchrotron require acceleration of protons up to 10^{21} eV and currently are considered less favorable.[23] However, for a combination of afterglow parameters, proton synchrotron can dominate compared to electron synchrotron emission as discussed in the literature.[20,24]

Synchrotron radiation from a decelerating blastwave can explain the multiwavelength afterglow emission from radio to gamma-rays observed in GRBs. The maximum energy detected in GRB 190114C by the MAGIC telescopes and in GRB 180720B by the H.E.S.S. telescopes is higher than the maximum energy that can be produced by synchrotron mechanism, and therefore, self-Compton of these synchrotron photons in the GRB blastwave is invoked to explain these energetic photons.[21]

In this work, we discuss GRB afterglow model and provide multiwavelength modeling of the observations of GRB 180720B and GRB 190114C (Section 2). We summarize our results in Section 3.

2. GRB blast-wave expansion and afterglow radiation

We consider the synchrotron radiation from the electrons accelerated to relativistic energies in the forward shock. Further, these synchrotron photons are upscattered by the same population of relativistic electrons. We revisit the synchrotron self-Compton model by[21] and the details of the expressions used here are taken from the SSC formalism for GRBs.[25]

Table 1. Afterglow model parameters used for the interpretation of multiwavelength observation of GRB 180720B and GRB 190114C. The VHE component detection time information $t_{\rm VHE,obs}$ is taken from[10] for GRB 180720B and from[7,8] for GRB 190114C respectively.

Parameters	GRB 180720B	GRB 190114C
$t_{\rm VHE,obs}$(s)	$> 3.6 \times 10^4$	> 55
$E_{k,\rm iso}$(erg)	4.5×10^{54}	$4 \times 10^{54}(1 \times 10^{53})$
Γ	400	300(36)
$t_{\rm dec}$(s)	69	$52.6(2.3 \times 10^4)$
$t_{0,\rm SSC}$(s)	6×10^{-6}	8.5(61)
$A_\star$		0.02(0.02)
n_0(cm^{-3})	0.035	
p	2.4	2.18(2.5)
ϵ_e	0.05	0.03(0.08)
ϵ_B	1.2×10^{-5}	0.01(0.07)
ϕ	1	1 (1)

2.1. *GRB 190114C multiwavelength emission*

GRB 190114C was located at a redshift $z = 0.4245 \pm 0.005$[27] and its prompt gamma-ray emission releases isotropic energy $\sim 3 \times 10^{53}$ erg.[26] The very high energy gamma-ray detection from GRB 190114C was reported by MAGIC telescope.[26,28]

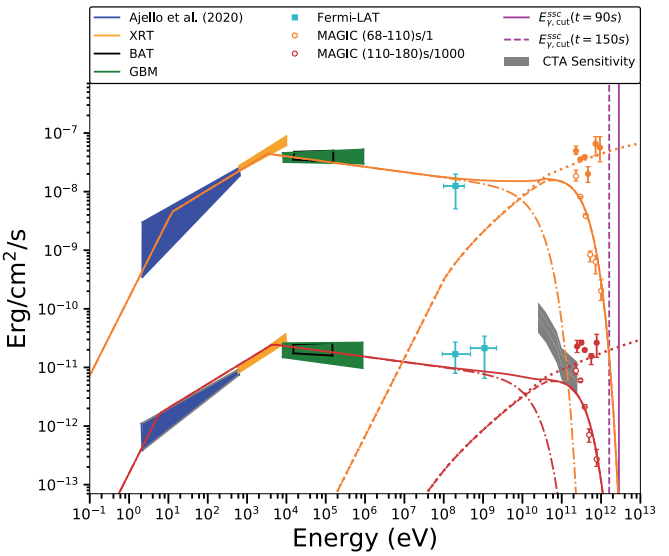


Fig. 1. SED for GRB 190114C reproduced from our earlier work,[25] with slightly different values of input model parameters. The data is extracted from.[26]

This source was observed in multiwavelengths (radio to gamma rays) by 6 satellites and 15 ground based telescopes.[26] We modelled GRB 190114C SEDs and light curves simultaneously and found that in late time the optical emission appears to be harder, which cannot be accounted for by the contribution from the host galaxy. Refreshed shocks are slow shells that will catch up with the afterglow shock region at late times.[29,30] This will enhance the energy of the forward shock medium and can be useful to explain the optical hardness seen in case of GRB 190114C. The flux below the cooling frequency due to refershed shock emission is $((1 + f^{(3+p)/4}))F_{\rm Sy}$, where $F_{\rm Sy}$ is the synchrotron flux due to the forward moving shell and $f = 5$ is the amount of higher flux compared to the optical emission of forward moving shell. The flux above the cooling frequency due to refershed shock emission is $((1 + f^{(2+p)/4}))F_{\rm Sy}$.

In our Fig. 2 we have included the contribution due to a refershed shock which can explain the late time optical flux and at the same time won't exceed the flux in X-rays. The late time onset of the afterglow constrains the initial Lorentz factor value of the slow moving shell to be ~ 36. The SSC emission is considered in the Thomson scattering regime (for details, see[25]). The Compton parameter is approximately constant and $Y \sim t_2^{-0.1}$, also shown in Figure 5.

2.2. *GRB 180720B multiwavelength emission*

GRB 180720B was located at a redshift $z = 0.653$[31] and its prompt gamma ray emission releases isotropic energy $\sim 6 \times 10^{53}$ erg.[10] In case of GRB 180720B, VHE

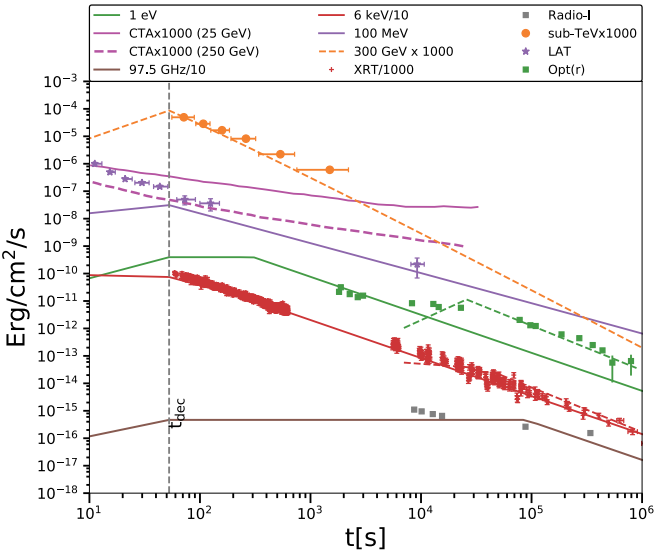


Fig. 2. Light Curves for GRB 190114C where the late time optical emission is explained via refreshed shock scenario. The data used here is extracted from[26] and Swift XRT database, https://www.swift.ac.uk/analysis/xrt/.

gamma-rays were detected by the H.E.S.S. telescopes after 10 hours.[32] Due to this late time detection we find that the emission of sub-TeV photons falls in the Klein Nishina (KN) regime, the medium is thin to pair production of VHE gamma rays, and we utilize the KN feedback effects on the cooling of electrons based on the formalism discussed by.[33] A comparison of Compton parameter is shown in Figure 5 for Thomson and KN cases. The afterglow model explains the data well as shown in Figures 3 and 4, where the SED and light curves are plotted for same set of input parameters. The model predicts higher levels of optical emission and we need to consider the absorption effects which lowers the optical flux by a factor of 0.5. The model parameters used in our work are shown in Table 1.

3. Summary and Conclusion

The VHE gamma-ray observations of two GRBs studied here are well explained using the synchrotron and SSC models. The late time optical afterglow for these two objects are in contrast to each other. In GRB 190114C the optical emission is harder and we required a refreshed shock to explain it, however for GRB 180720B the model produces higher level of optical flux by a factor of 0.5. Hence there is absorption of optical flux in the host Galaxy of GRB 180720B. The absorption of VHE photons for pair-production, in the blastwave, is negligible. The extra-galactic background light (EBL) absorption are considered for finding the maximum energy of the photons based on the model by.[34] These cut-off effects are important above

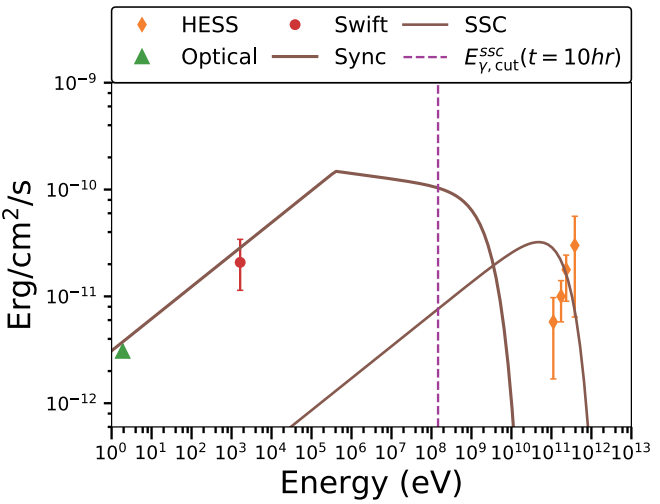


Fig. 3. SED for GRB 180720B using SSC model plotted against the HESS onservations[10] at duration 10 hr.

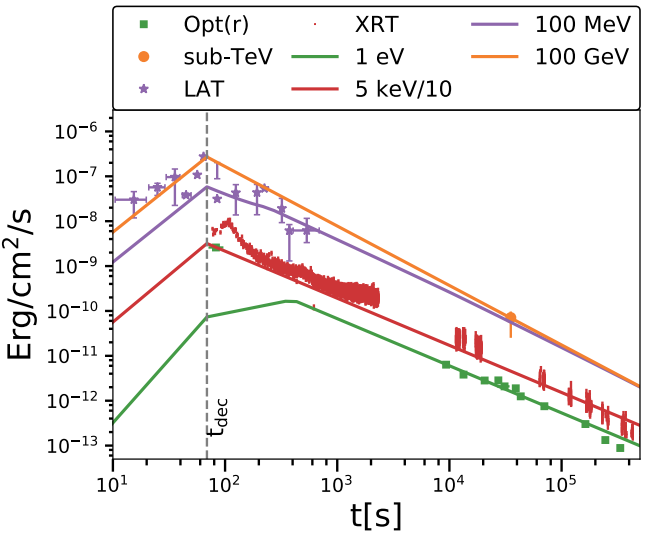


Fig. 4. Light Curves for GRB 180720B, the data used here is extracted from[10] and Swift XRT database, https://www.swift.ac.uk/analysis/xrt/.

200 GeV and 160 GeV for GRB 190114C and GRB 180720B respectively. The t^2 dependence for the blastwave flux before deceleration is used for both the ism and wind media since the flux dominates due to the area of the blastwave expansion. The bulk Lorentz factor Γ and $E_{k,\mathrm{iso}}$ and ϵ_e have approximately similar values for

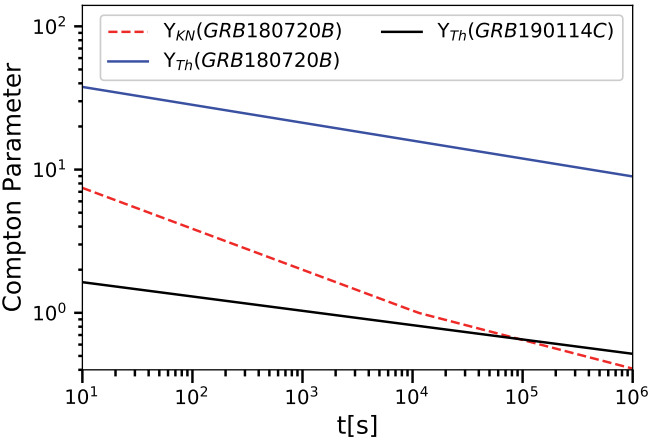


Fig. 5. Compton parameter Y estimated for two GRBs. We have modelled GRB 190114C using Thomson approximation, however for GRB 180720B we required SSC in KN regime.

both GRBs, however the fraction of energy in the magnetic field ϵ_B is very low in GRB 180720 B. This low value of ϵ_B is also reported in the SSC modelling of this source in other works.[35] Also for GRB 180720B we find the ISM medium is more suitable for modelling, which is consistent with the X-ray afterglow observations.[36]

References

1. S. E. Woosley and J. S. Bloom, The Supernova Gamma-Ray Burst Connection, *araa* **44**, 507 (September 2006).
2. D. Eichler, M. Livio, T. Piran and D. N. Schramm, Nucleosynthesis, neutrino bursts and γ-rays from coalescing neutron stars, *nature* **340**, 126 (July 1989).
3. R. Sari and T. Piran, Variability in Gamma-Ray Bursts: A Clue, *apj* **485**, 270 (August 1997).
4. S. Kobayashi, T. Piran and R. Sari, Can Internal Shocks Produce the Variability in Gamma-Ray Bursts?, *apj* **490**, p. 92 (November 1997).
5. F. Daigne and R. Mochkovitch, Gamma-ray bursts from internal shocks in a relativistic wind: temporal and spectral properties, *mnras* **296**, 275 (May 1998).
6. A. M. Beloborodov and P. Mészáros, Photospheric emission of gamma-ray bursts, *Space Science Reviews* **207**, 87 (Jul 2017).
7. MAGIC Collaboration, V. A. Acciari, S. Ansoldi *et al.*, Observation of inverse Compton emission from a long γ-ray burst, *nature* **575**, 459 (November 2019).
8. MAGIC Collaboration, V. A. Acciari, S. Ansoldi *et al.*, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *nature* **575**, 455 (November 2019).
9. V. A. Acciari, S. Ansoldi, L. A. Antonelli *et al.*, MAGIC Observations of the Nearby Short Gamma-Ray Burst GRB 160821B, *apj* **908**, p. 90 (February 2021).
10. H. Abdalla, R. Adam, F. Aharonian *et al.*, A very-high-energy component deep in the γ-ray burst afterglow, *nature* **575**, 464 (November 2019).

11. H. E. S. S. Collaboration, H. Abdalla, F. Aharonian *et al.*, Revealing x-ray and gamma ray temporal and spectral similarities in the GRB 190829A afterglow, *Science* **372**, 1081 (June 2021).
12. B. Paczynski and J. E. Rhoads, Radio Transients from Gamma-Ray Bursters, *apjl* **418**, p. L5 (November 1993).
13. P. Meszaros, M. J. Rees and H. Papathanassiou, Spectral Properties of Blast-Wave Models of Gamma-Ray Burst Sources, *apj* **432**, p. 181 (September 1994).
14. P. Mészáros and M. J. Rees, Optical and Long-Wavelength Afterglow from Gamma-Ray Bursts, *apj* **476**, 232 (February 1997).
15. M. Vietri, The Soft X-Ray Afterglow of Gamma-Ray Bursts, A Stringent Test for the Fireball Model, *apjl* **478**, L9 (March 1997).
16. A. Panaitescu and P. Mészáros, Dynamical Evolution, Light Curves, and Spectra of Spherical and Collimated Gamma-Ray Burst Remnants, *apj* **526**, 707 (December 1999).
17. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *apjl* **497**, L17 (April 1998).
18. J. Chiang and C. D. Dermer, Synchrotron and Synchrotron Self-Compton Emission and the Blast-Wave Model of Gamma-Ray Bursts, *apj* **512**, 699 (February 1999).
19. R. A. Chevalier and Z.-Y. Li, Wind Interaction Models for Gamma-Ray Burst Afterglows: The Case for Two Types of Progenitors, *apj* **536**, 195 (Jun 2000).
20. B. Zhang and P. Mészáros, High-Energy Spectral Components in Gamma-Ray Burst Afterglows, *apj* **559**, 110 (September 2001).
21. R. Sari and A. A. Esin, On the Synchrotron Self-Compton Emission from Relativistic Shocks and Its Implications for Gamma-Ray Burst Afterglows, *apj* **548**, 787 (February 2001).
22. J. Granot and R. Sari, The Shape of Spectral Breaks in Gamma-Ray Burst Afterglows, *apj* **568**, 820 (April 2002).
23. M. Vietri, GeV Photons from Ultrahigh Energy Cosmic Rays Accelerated in Gamma Ray Bursts, *prl* **78**, 4328 (June 1997).
24. S. Razzaque, A Leptonic-Hadronic Model for the Afterglow of Gamma-ray Burst 090510, *apjl* **724**, L109 (November 2010).
25. J. C. Joshi and S. Razzaque, Modelling synchrotron and synchrotron self-Compton emission of gamma-ray burst afterglows from radio to very-high energies, *mnras* **505**, 1718 (August 2021).
26. MAGIC Collaboration, V. A. Acciari, S. Ansoldi *et al.*, Observation of inverse Compton emission from a long γ-ray burst, *nature* **575**, 459 (November 2019).
27. A. J. Castro-Tirado, Y. Hu, E. Fernandez-Garcia *et al.*, GRB 190114C: refined redshift by the 10.4m GTC., *GRB Coordinates Network* **23708**, p. 1 (January 2019).
28. MAGIC Collaboration, V. A. Acciari, S. Ansoldi *et al.*, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *nature* **575**, 455 (November 2019).
29. R. Sari and P. Mészáros, Impulsive and Varying Injection in Gamma-Ray Burst Afterglows, *apjl* **535**, L33 (May 2000).
30. J. Granot, E. Nakar and T. Piran, Astrophysics: refreshed shocks from a γ-ray burst, *nature* **426**, 138 (November 2003).
31. P. M. Vreeswijk, D. A. Kann, K. E. Heintz *et al.*, GRB 180720B: VLT/X-shooter redshift., *GRB Coordinates Network* **22996**, p. 1 (January 2018).
32. H. Abdalla, R. Adam, F. Aharonian *et al.*, A very-high-energy component deep in the γ-ray burst afterglow, *nature* **575**, 464 (November 2019).
33. E. Nakar, S. Ando and R. Sari, Klein-Nishina Effects on Optically Thin Synchrotron and Synchrotron Self-Compton Spectrum, *apj* **703**, 675 (September 2009).

34. J. D. Finke, S. Razzaque and C. D. Dermer, Modeling the Extragalactic Background Light from Stars and Dust, *apj* **712**, 238 (March 2010).
35. X.-Y. Wang, R.-Y. Liu, H.-M. Zhang *et al.*, Synchrotron Self-Compton Emission from External Shocks as the Origin of the Sub-TeV Emission in GRB 180720B and GRB 190114C, *apj* **884**, p. 117 (October 2019).
36. N. Fraija, S. Dichiara, A. C. C. d. E. S. Pedreira *et al.*, Modeling the Observations of GRB 180720B: from Radio to Sub-TeV Gamma-Rays, *apj* **885**, p. 29 (November 2019).

The VERITAS gamma-ray burst follow-up program

D. Ribeiro* for the VERITAS Collaboration†

Physics Department, Columbia University,
New York, NY 10027, USA
**E-mail: dr2792@columbia.edu*

Recent detections of gamma-ray bursts (GRBs) at energies above 100 GeV demonstrate that imaging atmospheric Cherenkov telescopes (IACTs) operating in the very-high-energy range (VHE; E > 100 GeV) can provide insights into the physics of GRBs. By searching for the highest-energy photons emitted by GRBs, these telescopes can help answer questions about the particle acceleration and emission processes that occur during both the prompt and afterglow phases of GRBs. VERITAS is a very-high-energy IACT array located at the Whipple Observatory in southern Arizona, which has maintained an active GRB observing program since mid-2006. In this presentation, we will share some of the recent achievements of the VERITAS GRB follow-up program. We will discuss the development of analysis methods tailored to transient signals, and how the upper limits on the VHE emission obtained from observations of prominent bursts by VERITAS allowed us to constrain radiation mechanisms in the afterglow (e.g., for GRB 130427A) and constrain properties of the environment in which the burst took place (e.g., for GRB 150323A). Compact binary mergers that trigger short GRBs may also result in gravitational wave emission, so we will review both our follow-up program from LIGO/Virgo triggers, and also the use of archival VERITAS data to search for short GRBs based on sub-threshold events for LIGO/Virgo. Lastly, based on the properties of the VHE-detected GRBs, we will discuss recent changes to our follow-up strategy to account for the *Swift*/XRT properties for optimal VERITAS observing sensitivity.

Keywords: Gamma-ray bursts, GRBs, gravitational waves

1. Introduction

Gamma-ray bursts (GRBs) are a class of astrophysical explosions that are some of the brightest events in the universe. Emitting total energies greater than 10^{53} erg/s, these events are detected in the optical, X-ray and high-energy (HE) to very-high-energy (VHE) gamma-rays. The physical mechanisms to explain the source of the high-energy emission range from synchrotron to synchrotron-self-Compton (SSC) processes based on the environment of the GRB emission regions. Particularly important are the highest energy photons in the VHE range that can constrain the properties of the GRB emission. Ground-based gamma-ray telescopes are well suited to these observations, and the recent detections of photons above 100 GeV by H.E.S.S. and MAGIC provide insights into the physics of GRBs[1–5]

In this paper, the VERITAS GRB follow-up program is discussed. Section 2, describes the VERITAS instrument and observation capabilities. In Section 3, the

†`https://veritas.sao.arizona.edu`

GRB observations by VERITAS are summarized. In Section 4, the techniques used to improve future observations and optimize data analyses are described. Section 5 describes how the VERITAS GRB follow-up program is enhanced by the new gravitational wave trigger with the start of operation of the LIGO/Virgo gravitational wave (GW) observatory.

2. VERITAS

The Very Energetic Radiation Imaging Telescope Array System (VERITAS) is a ground-based imaging atmospheric Cherenkov telescope (IACT) array located in southern Arizona, USA. VERITAS consists of four 12-m IACTs, each equipped with a 499 photo-multiplier tube (PMT) camera and 350 hexagonal mirror facets arranged in a Davies-Cotton optical design.[6] This system enables the detection of air showers formed in the atmosphere by the cascade of particles created from the interaction of the original gamma-ray photon and molecules in the atmosphere.

VERITAS is sensitive to gamma rays of energy 100 GeV to 30 TeV. The mechanical structure of each IACT is capable of a maximum slew speed of 1 degree per second. Each camera has a field of view of 3.5 degrees. VERITAS has a 68% containment radius less than 0.1 degree at 1 TeV.[7] The recorded data is analyzed by two independent pipelines (to enable cross-checking) to reliably select gamma-ray and cosmic-ray events from the camera images, determine energy, angular direction and arrival time of the original photon, and reconstruct a skymap of the observed region.[8,9]

In the scale of short timescale events such as GRBs and other astrophysical transients, the sensitivity is statistics dependent: by one widely required standard, the detection criterion is achieved when observing a minimum number of signal counts ($N_s = 10$), the signal above the background is at least five times the systematic uncertainty in the background estimation, assumed to be 1%, and the Li & Ma statistical significance is 5σ or more.[10] These requirements weaken the achievable flux sensitivity at very short observation times ≤ 100 seconds, particularly for higher energy photons under the assumed power law spectral model, compared to longer term observation campaigns of steady sources.[11] In Figure 1, the relative short timescale sensitivities are shown for *Fermi*-LAT, VERITAS, MAGIC, H.E.S.S. and the future CTA observatories. For a Crab Nebula-like point source detection, VERITAS is capable of a 5σ detection in about 2 minutes.

3. Summary of Observations

VERITAS has followed up on more than 200 GRBs since the beginning of its GRB follow-up program in 2006. The Gamma-ray Coordinates Network (GCN) provides real-time coordinates (via *Notices*) and follow-up reports (via *Circulars*). VERITAS is subscribed to this system and has integrated an automated tracking response for incoming GCN notices. Alerts are quickly parsed and reviewed by on-site observers,

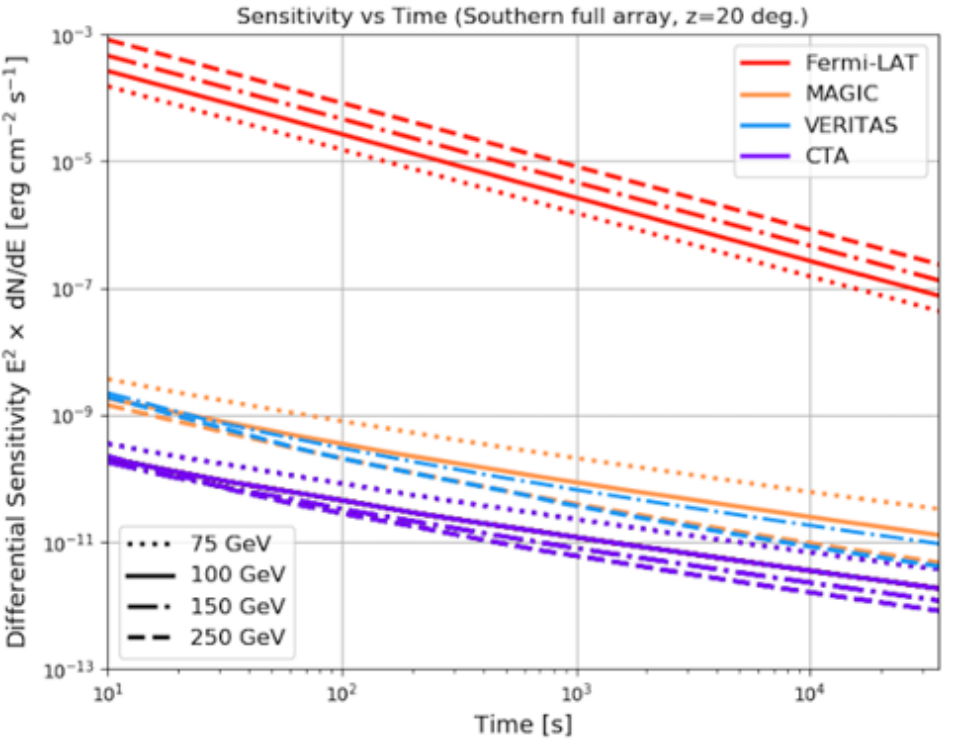


Fig. 1. The short timescale differential sensitivity of various gamma-ray observatories to Crab-like point sources (from[11]). The assumed spectral model is a power law with spectral index of -2.5. Four differential energy bins between 75 and 250 GeV are selected due to the relative overlap between the each observatory's operational energy ranges.

who assess visibility (i.e. checking for observable conditions such as pointing direction, slew time, and weather) and decide whether to follow up with a user-friendly pop-up GUI.

The distribution of the sources of GRB triggers observed is shown in Figure 2. *Swift* and *Fermi*-GBM dominate the alerts with their surveying-mode monitoring systems. Figure 3 shows the frequency of *Swift* (3a) and *Fermi*-LAT (3b) observed bursts since the beginning of the VERITAS GRB program. While *Swift* bursts are significantly more common, *Fermi*-LAT events are prioritized due to their high energy components and a few *Fermi*-LAT bursts are observed every season.

The standard protocol for observing is as follows: if the alert came from *Swift*-BAT, *Swift*-XRT or *Fermi*-LAT, observe up to trigger time plus 3 hours (for a maximum of 3 hours); if the alert came from *Fermi*-GBM only and the burst position is determined to $\leq 5^{\circ}$, observe up to the trigger time plus 1 hour (for a maximum of 1 hour). This protocol prioritizes well-localized and high-probability candidate GRBs. In some extreme cases, exceptions have been made from this protocol, such as GRB 130427A discussed in section 3.1. Figure 4 shows the distribution of delay

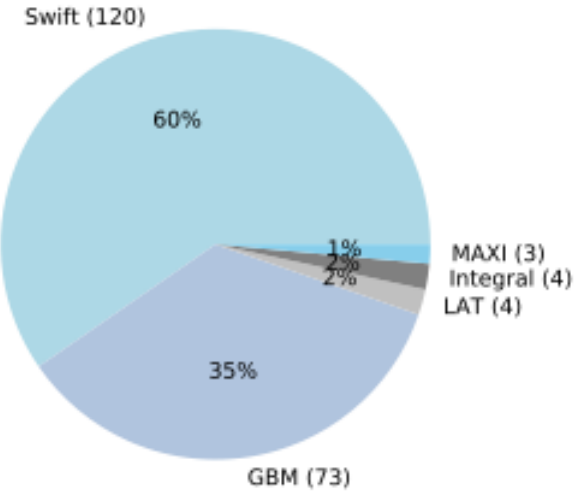
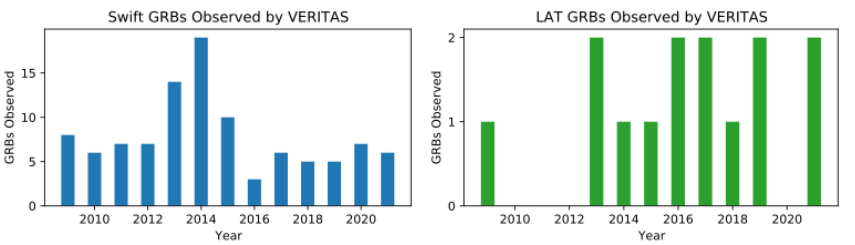


Fig. 2. Distribution of the source of trigger alerts for VERITAS followed-up GRBs.



(a) (b)

Fig. 3. Distribution of *Swift* (a) and *Fermi*-LAT (b) observed GRBs since 2009.

times: most have been observations beginning between 10^2 and 10^4 seconds after the burst detections, while a handful of exceptions were made for bright bursts with observations starting $\geq 2 \times 10^4$ seconds.

In the subsections below, three exceptional GRBs followed by VERITAS are discussed, GRB 130427A, GRB 150323A and GRB 190114C.

3.1. *GRB 130427A*

GRB 130427A was a *Fermi*-LAT-detected GRB that was bright enough for VERITAS to follow up almost 20 hours after the burst, yielding meaningful upper limits for the VHE acceleration mechanism.[12] Originally detected by *Fermi*-GBM and *Swift*-BAT, the burst was later associated with a Type Ic supernova (SN2013cq) at redshift of z=0.34.[13] The *Fermi*-LAT spectrum is consistent with a powerlaw spectral model with index of -2.2 up to 20 ks, while the later VERITAS upper

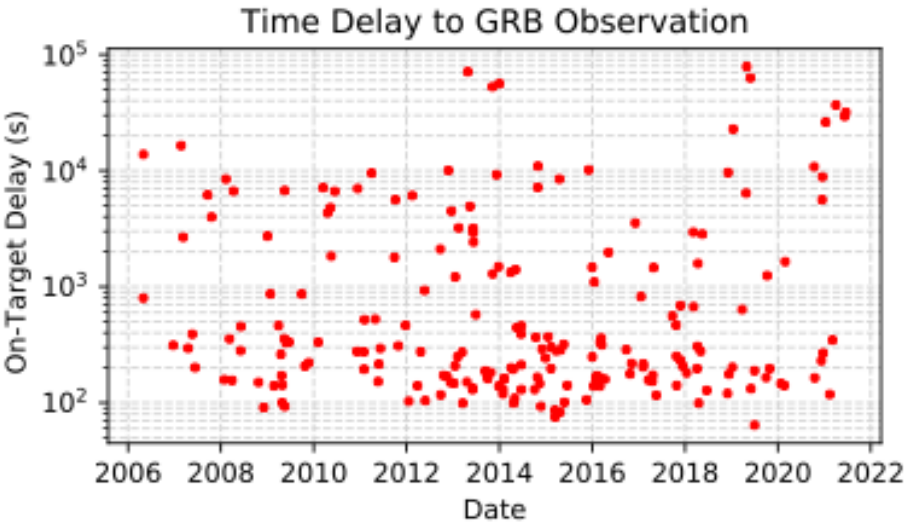


Fig. 4. Delay between the GRB trigger time and the beginning of observations by VERITAS. Most observations were followed up within 3 hours of the alert, while a few exceptionally-bright events were followed up by as much as 20 hours later. The fastest follow-up was ≈ 1 minute after the alert.

limit was calculated assuming a power law model with index -2 (see table 1). The *Fermi*-LAT light curve evolved with a temporal power-law with index of -1.4 ($dN/dt \propto t^{-1.35\pm0.08}$). GRB 130427A is one of the brightest bursts ever followed up by VERITAS.

Table 1. The VERITAS observation of GRB 130427A from (from[14]).

Day	$\mathbf{N}_{on}$	$\mathbf{N}_{off}$	α^{a}	σ^{b}	**UL** erg cm^{-2} s^{-1}
1	165	1164	0.125	1.3	9.4×10^{-12}
2	322	2120	0.143	1.1	6.6×10^{-12}
3	402	2820	0.147	-0.5	2.7×10^{-12}
Total	889	6104	0.141	0.9	3.3×10^{-12}

Note: [a] Relative ratio of exposures of On and Off observing regions.
[b] Li & Ma (1983) test statistic.[10]

Both synchrotron and SSC models proposed to explain high-energy and very-high-energy emission are constrained by the VERITAS VHE upper limit. In the synchrotron model, the VERITAS upper limit can constrain the magnetic field of an assumed interstellar medium (ISM), $B_{ISM} \leq 5\mu\text{G}\ E_{53}^{1/8}\varepsilon_{B,-4}^{1/2}n_0^{3/8}(t/t_V)^{-3/8}$ where the terms are explained in[14] (see[14,15]). In the SSC model, the VERITAS upper limit

constrains the synchrotron emission of the electron population within the emission region, limiting the Klein-Nishina energy to $E_{\rm KN} < 120$ GeV as determined by *Fermi*-LAT temporal data. However, for lower $E_{\rm KN}$, the VERITAS upper limit does not constrain the SSC model, and combined with the single zone interpretation from the early afterglow data, the synchrotron model is preferred.

3.2. *GRB 150323A*

GRB 150423A was a *Swift*-BAT-detected GRB that was observed with the standard follow-up protocol within 270 seconds of the prompt emission peak and at a relatively low redshift of $z = 0.6$. Although not necessarily noteworthy by other observatories, this burst was selected because it had one of the highest observation flux weights compared to other bursts observed by VERITAS.[16]

The derived observational weight combines the extragalactic background light (EBL) absorption factor for each burst with the time delay between burst explosion T_0 and VERITAS observation start time. The EBL absorption is defined by the burst redshift and photon energy ($e^{-\tau(z,E)}$), where the energy is chosen to be the telescope energy threshold, which is a strong function of the telescope elevation angle. The weight can be interpreted as being proportional to the approximate flux of the burst at the time of the observation, where the burst is assumed to have a temporal dependence of $1/t$.

VERITAS began the observation about 2 min after the BAT emission peak. The observation significance is 0.36σ, and with the limited EBL attenuation (only 50% at 100-200 GeV), the upper limit is $< 1\%$ of the prompt fluence. The differential upper limit (99% C.L.) at 140 GeV is 3.7×10^{-10} TeV^{-1} cm^{-2} s^{-1}, while the integral upper limit from 140 GeV to 30 TeV is 1.6×10^{-11} cm^{-2} s^{-1}. The strong upper limit on VHE emission favors an explosion into the stellar wind of a dense progenitor (Wolf-Rayet star), or a weak blast wave in a tenuous ISM.[17]

3.3. *GRB 190114c*

VERITAS also followed up the first IACT-detected GRB 190114C, observed by the MAGIC telescope[a].[1,2,18] VERITAS received the alert during daytime (UT 20:57:04) and waited for the object to rise before beginning observations several hours later. Due to moonlight, the VERITAS cameras were operating with reduced high voltage, which decreases the sensitivity to low-energy gamma-ray events. The low elevation of the burst location ($\leq 32^\circ$) during the observation further impacts the energy threshold. Additionally, clouds occluded the field of view leading to lost data through the observation. Due to these unique observation conditions, VERITAS is currently working toward a calibrated analysis pipeline custom to the challenges of this burst.

[a] `https://www.astronomerstelegram.org/?read=12390`

4. Optimizations

The observation of the prompt and afterglow emission of GRBs is a primary goal of the VERITAS follow-up program. However, due to the combination of observation conditions such as weather, moonlight, and pointing direction, with the variation of GRB flux and redshift, a significant GRB detection by VERITAS in the VHE range has remained elusive. In light of the probabilistic challenges associated with observing a bright, nearby GRB shortly after a trigger is announced, the following optimization strategies are employed: section 4.1 describes a strategy to monitor and follow-up certain bright *Swift*-XRT GRB candidates, and section 4.2 describes an offline analysis method to incorporate the time dependence of the GRB light curve into the calculation of the significance of excess photon counts.

4.1. Swift *Monitored Triggers*

Some GRBs have bright, extended afterglows in the X-ray band. These bursts may also be bright in the VHE band. Table 2 shows several VHE-detected GRBs and their inferred *Swift*-XRT X-ray flux at the time of detection (along with the corresponding delay between T_0 and the start of the observation). Using these H.E.S.S. and MAGIC detected afterglows, it is prudent to use the *Swift*-XRT inferred flux as a guide for following up on bursts that are more likely to be seen in the VHE band.

VERITAS might be missing some of the bright afterglows for GRBs that become observable at $T > T_0 + 3$ hrs, and therefore we have modified the standard GRB follow-up protocol to include the *Swift*-XRT flux (if available) to adjust the amount of time allocated for each burst. If a *Swift*-XRT flux is available, the spectrum is extrapolated to the VHE range and the light curve is extrapolated for the remainder of the observing window. GRBs with inferred fluxes $> 10^{-11}$ erg cm^{-2} s^{-1} in the X-ray band are observed for as long as they are still above that threshold. Bursts that might have been missed after passing the $T > T_0 + 3$ cutoff can now remain candidates if the X-ray band is still bright. The remaining alerts without a *Swift*-XRT counterpart follow the standard protocol. This selection threshold prioritizes bright events, increasing our probability of detecting VHE events.

4.2. *Time Dependence*

In the conventional analysis of sources observed by IACTs, background and signal regions are determined a priori so that an excess count can be calculated that characterizes the source count rate during an observation period. Following the criteria established, the statistical significance is calculated using Eq. 17 of Li & Ma.[10] However, in this formulation, both the source and background rates are assumed to be time independent. In the case of GRBs and other strongly time-dependent sources, the excess counts may be dominant in certain time bins within the observation and negligible in others, such that the excess counts summed over all time bins are not significant.

Table 2. VHE detected GRBs, as well as the bright GRB130427A, and their inferred XRT flux at the time of observation. The sustained XRT flux in these VHE detections suggests an X-ray threshold for guidance to trigger VHE follow-up observations.

GRB	**Afterglow Detected**	**Follow-up Delay**	**Inferred XRT Flux after Delay [erg cm-2 s-1]**
130427A	Detected (Fermi-LAT)	∼20 Hrs (VERITAS)	4.1×10^{-11}
180720B	Detected (H.E.S.S)	∼10 Hrs (H.E.S.S)	7.7×10^{-11}
190114C	Detected (MAGIC)	∼6 Hrs (VERITAS)	6.5×10^{-11}
190829A	Detected (H.E.S.S)	∼4.5 Hrs (H.E.S.S)	9.9×10^{-11}

To solve this problem, the time dependence of the source count rate is assumed in a parameterized form $s(t)$ and the test statistic is re-derived following the strategy outlined by Li & Ma (1983).[19] The maximum likelihood estimate is used to optimize the signal and background count rates, collected within the *On* and *Off* observation regions defined as $\langle N_{on} \rangle = (\bar{b} + \bar{s}(t))T_{on}$ and $\langle N_{off} \rangle = \bar{b}T_{off}$ for observation times T_{on} and T_{off}, respectively. Assuming a Poisson distribution of counts, the general form of the test statistic is a ratio of the likelihoods for the null and alternate hypotheses $\mathcal{L}_0$ and $\mathcal{L}$:

$$\frac{\mathcal{L}_0}{\mathcal{L}} = \frac{\bar{b}_0^{N_{on}+N_{off}}}{\left(\bar{b}+\bar{s}\right)^{N_{on}} \bar{b}^{N_{off}}}, \tag{1}$$

where $\bar{b}$ and $\bar{s}$ are the time average background and signal count rates, and $\bar{b}_0$ is the time averaged rate when all counts observed are background.

To apply the time dependence, the T_{on} observation time is split into N bins of Δt, where the probability of the number of counts for each bin is independent and $N\Delta t = T_{on}$. For a parameterized $s(t)$, at the limit of $N \to \infty$, $\Delta t \to 0$, the likelihood ratio reduces to:

$$\frac{\mathcal{L}_0}{\mathcal{L}} = \frac{b_0^{N_{on}+N_{off}}}{\left(\Pi_{t_i \in \{t_{on}\}}(b + s(t_i))\right) b^{N_{off}}} \tag{2}$$

where $s(t) \equiv \theta \times t^{-1}$. The time dependence of the signal was assumed to be 1/t since it is most similar to many GRBs light curves,[20] where θ is a normalization constant for the time-dependent count rate.

This modified likelihood ratio was validated with a Monte Carlo simulation. Using a toy-model MC of a GRB, with 5×10^5 observations starting 2 min after burst trigger, the time-dependent signal $s(t) = \theta \times t^{-1}$, and the typical IACT background rate (1 γ /min), the following simulation results provided evidence for some additional conclusions (see also Figure 2 of[19]). For a 35-min observation, the mean significance over all the simulated observations increased from 4σ to 5σ once time

dependence is included, a 25% increase in sensitivity. For a 100-min observation: the simulated observations reveal that the overall significance improves from 3.4σ to 5.2σ with time weighting for that observation time. Comparing 35-min and 100-min simulations, the time-independent method has a reduced sensitivity from 4σ to 3.4σ as expected with the increase in background counts, while the time-dependent model increases from 5σ to 5.2σ, indicating that the time dependence allows for an improved estimate of the background and signal rates for longer observation times.

The simulation showed that the significance is dependent on the length of the observation. The standard significance[10] peaks and then declines over time as the background rate accumulates counts toward the *On* region and dilutes the signal counts. However, the time-dependent likelihood ratio grows monotonically if events are weighted to account for the GRB's time dependence, up to as much as 10-25% increase from the time-independent significance (see Fig. 3 of.[19]) Some common analysis software packages within the gamma-ray astrophysics field (e.g. CTOOLs) include a temporal model in the maximum likelihood fitting in addition to the spatial and spectral models.[21]

5. Gravitational Wave Follow-Up

Compact binary coalescences (CBCs) may trigger short GRBs (sGRBs). The detection of the binary neutron star (BNS) merger GW170817 was associated with GRB 170817A, a sGRB observed by *Fermi*-GBM triggered by a kilonova.[22,23] The multi-messenger detection of this event encourages a VHE follow-up program for future GW BNS trigger candidates. Both real-time VHE follow-ups (5.1) and archival searches by VERITAS are discussed (5.2).

5.1. *Real Time VHE Follow-Up*

The LIGO BNS sensitivity indicates a horizon between 111 Mpc and 134 Mpc from the third observation run (O3) for LIGO Livingston Observatory (LLO) and LIGO Hanford Observatory (LHO), respectively.[24,25] At this distance range, VHE photons are relatively unaffected by the extragalactic background light, which could generate a favorable selection of candidate GRB triggers for VERITAS follow-up. The joint detection of both GWs and VHE emission from GRBs may be possible through the real time follow-up of GW triggers by VERITAS.

The VERITAS follow-up strategy during O3 involved calculating the visibility of the GW localization region to select an observable portion. The visibility is the amount of time available for each sky coordinate for the next 24 hours. If the moon is above the horizon, the angular distance is limited to the range between 60° and 120° while the moon illumination is low enough. Using a Greedy Traveling Salesman heuristic, the pointings that tile the observable portion are defined and sorted such that the probability covered is maximized and the slewing time is minimized. The exposure per pointing is currently set at 5 min, for a maximum of three hours total.

Pointings were limited to those visible at $> 50^\circ$ elevation from VERITAS during the night.

The validation and summary of the VERITAS GW follow-up strategy was presented in.[26] The VERITAS follow-up of GW170104 and later observations were validated with a dedicated observation of the Crab Nebula under the same conditions described above, with six 5-min pointings tiled around the Crab Nebula position (Figure 2 in[26]). Table 1 in[26] lists a selection of O3 triggers followed up by VERITAS, including the BNS event GW190425.[27]

5.2. *Archival Search*

In addition to real-time GW follow-ups, VERITAS also developed an archival search for VHE gamma-ray emission in coincident observations of sub-threshold LIGO/VIRGO GW candidates.[28] In this search, there were 7 sub-threshold candidates found in serendipitous coincidence with VERITAS archival observations (11 pointings, $\leq$ 10 hr). The candidates were selected from a published collection of sub-threshold triggers from O1.[29]

An important consideration for other observatories while evaluating this algorithm is that there was no extra cost to the IACT or a dedicated observation campaign. Although VERITAS has a small field of view (FoV) (3.5° diameter), this is large enough for coincident observations. However, the sensitivity decreases for greater distance away from the camera center. Cross-checking the list of published sub-threshold triggers for coincident VERITAS observations was completed offline. In the search, observations were considered coincident if they passed good weather selection criteria, were contained within the 90% probability region of the original LIGO/Virgo candidate map, and overlapped with the time window $-10 \leq t_0 \leq 10^4$ seconds of the coalescence time t_0.

The archival search of VERITAS observations yielded upper limits across the field of view for each observation run. For each observation, the flux upper limit was calculated at each pixel of that observation and the geometric mean was reported to characterize the observed field for comparison with the LIGO sub-threshold candidate probability region. Each of the candidate observation upper limits was then compared to a fluence model based on sGRB 090510, for a 75 Mpc distance over the [0.24 TeV, 30 TeV] energy band. This distance was chosen to match the LIGO sensitivity range for O1, while the energy range was chosen to match the VERITAS energy range.

The comparison of the GRB model to the VERITAS observations shows an estimated sensitivity to $\approx 10^{-3}$ times the fluence of the GRB. The probability that at least one truly astrophysical merger was observed by VERITAS within the observation was estimated to be 0.04%. These results indicate that archival searches of coincident VHE observations with GW candidates may constrain the fluence of the progenitor burst and provide meaningful upper limits to the broadband spectral and light curve models that characterize these events. The weak constraints

shown here indicate that stronger conclusions may be possible provided that there are enough coincident observations with sub-threshold GW candidates or more opportune events.

6. Summary and Outlook

In summary, VERITAS has an active and exciting GRB follow-up program. Since 2006, over 200 bursts have been followed-up. While no detections were made in the VHE range, meaningful upper limits were derived for a small selection of bursts. Both GRB 130427A and 150323A yielded constraints on the environment of the bursts and VHE production mechanism.

Optimizations were made to improve the likelihood of new detections. The addition of a *Swift*-XRT-guided observation strategy to optimize for bright bursts increased the amount of time dedicated to X-ray bright candidate bursts, increasing the likelihood of VHE detection. A time-dependent significance calculation improved the sensitivity to weak bursts, weighing the most likely time bins more heavily in the calculation.

The gravitational wave prompt follow-up program provides another source of triggers for sGRBs, increasing the number of GRB targets and the likelihood of detection. The GW sub-threshold archival search is a promising new method for seeking signals in archival data and setting constraints on the fluence of potential sGRB counterparts.

Acknowledgments

This research is supported by grants from the U.S. Department of Energy Office of Science, the U.S. National Science Foundation and the Smithsonian Institution, by NSERC in Canada, and by the Helmholtz Association in Germany. This research used resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science, and resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. DR is supported by NSF award 2013109. We acknowledge the excellent work of the technical support staff at the Fred Lawrence Whipple Observatory and at the collaborating institutions in the construction and operation of the instrument.

References

1. Acciari, V.A. et al, Teraelectronvolt Emission from the Gamma-Ray Burst GRB 190114c, *Nature* **575**, 455 (November 2019).
2. Derishev, E. and Piran, T., GRB afterglow parameters in the era of TeV observations: The case of GRB 190114C (June 2021).

3. Wang, X.Y., Liu, R.Y., Zhang, H.M., Xi, S.Q. and Zhang, B., Synchrotron Self-Compton Emission from External Shocks as the Origin of the Sub-TeV Emission in GRB 180720B and GRB 190114C, *The Astrophysical Journal* **884**, p. 117 (October 2019).
4. Abdalla, H. et al, A Very-High-Energy Component Deep in the Gamma-Ray Burst Afterglow, *Nature* **575**, 464 (November 2019).
5. H.E.S.S. Collaboration et al, Revealing X-Ray and Gamma Ray Temporal and Spectral Similarities in the GRB 190829a Afterglow, *Science* **372**, 1081 (June 2021).
6. Holder, J. et al, The First VERITAS Telescope, *Astroparticle Physics* **25**, 391 (July 2006).
7. Park, N., Performance of the VERITAS experiment, *Proceedings of Science* **30-July-20** (August 2015).
8. Christiansen, J., Characterization of a Maximum Likelihood Gamma-Ray Reconstruction Algorithm for VERITAS, in *Proceedings of Science*, (Sissa Medialab, Trieste, Italy, August 2017).
9. Maier, G. and Holder, J., Eventdisplay: An Analysis and Reconstruction Package for Ground-based Gamma-ray Astronomy, in *Proceedings of Science*, (Sissa Medialab Srl, August 2017).
10. Li, T.P. and Ma, Y.Q., Analysis methods for results in gamma-ray astronomy, *The Astrophysical Journal* **272**, p. 317 (September 1983).
11. Fioretti, V. et al, The Cherenkov Telescope Array Sensitivity to the Transient Sky, in *Proceedings of Science*, (Sissa Medialab, Trieste, Italy, July 2019).
12. Tam, P.H.T., Tang, Q.W., Hou, S.J., Liu, R.Y. and Wang, X.Y., Discovery of an Extra Hard Spectral Component in the High-Energy Afterglow Emission of Grb 130427a, *Astrophysical Journal Letters* **771**, p. L13 (July 2013).
13. Xu, D. et al, Discovery of the Broad-Lined Type Ic Sn 2013cq Associated with the Very Energetic Grb 130427a, *The Astrophysical Journal* **776**, p. 98 (October 2013).
14. Aliu, E. et al, Constraints on very high energy emission from GRB130427A, *Astrophysical Journal Letters* **795**, p. L3 (October 2014).
15. Kouveliotou, C. et al, Nustar Observations of GRB 130427a Establish a Single Component Synchrotron Afterglow Origin for the Late Optical to Multi-Gev Emission, *The Astrophysical Journal* **779**, p. L1 (November 2013).
16. Weiner, O.M., A New Time-Dependent Likelihood Technique for Detection of Gamma-Ray Bursts with IACT Arrays, *Proceedings of Science* **30-July-20** (September 2015).
17. Abeysekara, A.U. et al, A Strong Limit on the Very-high-energy Emission from GRB 150323A, *The Astrophysical Journal* **857**, p. 33 (April 2018).
18. Acciari, V.A. et al, Observation of Inverse Compton Emission from a Long Gamma-Ray Burst, *Nature* **575**, 459 (November 2019).
19. Weiner, O.M., Humensky, T.B., Mukherjee, R. and Santander, M., Estimating significance in observations of variable and transient gamma-ray sources, *Astroparticle Physics* **93**, 1 (July 2017).
20. Ackermann, M. et al, The First *fermi* -Lat Gamma-Ray Burst Catalog, *The Astrophysical Journal Supplement Series* **209**, p. 11 (October 2013).
21. Knödlseder, J. et al, Ctools: Cherenkov Telescope Science Analysis Software (2016).
22. Abbott, B.P. et al, Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *The Astrophysical Journal* **848**, p. L13 (October 2017).
23. Abbott, B.P. et al, Multi-messenger Observations of a Binary Neutron Star Merger, *The Astrophysical Journal* **848**, p. L12 (October 2017).

24. Abbott, R. et al, Population Properties of Compact Objects from the Second LIGO–Virgo Gravitational-Wave Transient Catalog, *The Astrophysical Journal Letters* **913**, p. L7 (May 2021).
25. Buikema, A. et al, Sensitivity and performance of the Advanced LIGO detectors in the third observing run, *Physical Review D* **102**, p. 062003 (September 2020).
26. Santander, M., Recent results from the VERITAS multimessenger program, in *36th International Cosmic Ray Conference (ICRC2019)*, September 2019.
27. Abbott, B.P. et al, GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4\ M_{\odot}$, *The Astrophysical Journal* **892**, p. L3 (March 2020).
28. Adams, C.B. et al, An Archival Search for Neutron-star Mergers in Gravitational Waves and Very-high-energy Gamma Rays, *The Astrophysical Journal* **918**, p. 66 (September 2021).
29. Magee, R. et al, Sub-threshold Binary Neutron Star Search in Advanced LIGO's First Observing Run, *The Astrophysical Journal* **878**, p. L17 (June 2019).

MAGIC view of Gamma-Ray Bursts at very high energies

A. Berti* on behalf of the MAGIC collaboration
Max-Planck-Institut für Physik,
Munich, 80805, Germany
**E-mail: aberti@mpp.mpg.de*

The detection of gamma-ray bursts (GRBs) is one of the main scientific targets pursued by the MAGIC collaboration since almost 20 years. The MAGIC telescopes were specifically designed for this purpose: the main figures of merit are the fast slewing speed (7deg/s), the low energy threshold (50 GeV at zenith) and the high sensitivity in the low energy regime. These features make MAGIC one of the most suitable instrument for the follow-up and detection of GRBs. After more than 15 years of dedicated searches, finally the first detection at teraelectronvolt energies of a GRB, namely GRB 190114C, was achieved by the MAGIC collaboration, revealing a new emission component in the afterglow phase. This discovery opened up a new era in field of GRB studies, which is now witnessing other detections, as demonstrated with the case of GRB 201216C. Furthermore, a hint of detection by MAGIC from the short and nearby GRB 160821B gives precious hints on the possible very high energy emission from this class of bursts, also in relation to searches of gravitation wave counterparts. Therefore, MAGIC is giving a crucial contribution to GRB physics, leading to a better understanding of the mechanisms underlying these peculiar objects. In this contribution I will introduce the MAGIC follow-up program, focusing on the aspects which led to the successful detection of GRBs and highlighting some key results. Finally, I will present the future challenges in these observations, discussing how MAGIC can contribute even more to the field.

Keywords: Gamma-ray bursts; MAGIC; very-high-energy

1. Introduction

The search for very high energy (VHE, $E \gtrsim 100\,\mathrm{GeV}$) emission from gamma-ray bursts (GRBs) is one of the key projects of the imaging atmospheric Cherenkov telescopes (IACTs) currently in operation:

- MAGIC (*Major Atmospheric Gamma-ray Imaging Cherenkov*), located in the Canary island of La Palma, Spain;
- H.E.S.S. (*High Energy Stereoscopic System*), operative in Namibia;
- VERITAS (*Very Energetic Radiation Imaging Telescope Array System*), located in Arizona.

In these proceedings we focus on MAGIC, a system of two Cherenkov telescopes located at about 2200 meters above sea level in the Observatorio del Roque de Los Muchachos (ORM) in La Palma, Canary Islands, Spain. MAGIC started operations in late 2003 as a single telescope (MAGIC-I) and in 2009 a second one (MAGIC-II) was added to perform stereoscopic observations. In 2012 the camera of MAGIC-I was upgraded in order to be a copy of the one installed in MAGIC-II. Both telescopes have a reflector of 17 meter diameter, corresponding to a reflective surface

of $\sim 236\,\mathrm{m}^2$. This allows MAGIC to reach an energy threshold as low as 50 GeV at trigger level when pointing close to zenith. This is an important figure of merit for the observation of GRBs, since they are sources typically located at high redshift and the VHE gamma rays are heavily absorbed due to the interaction with the low energy photons of the extragalactic background light (EBL). Other features making MAGIC a suitable IACT to perform GRB observations are the following (see e.g. ref. 1):

- the high sensitivity even at the lowest energies. MAGIC has a sensitivity of 0.7% of the Crab Nebula flux above 220 GeV in 50 h of observation;
- the large effective area, of the order of $10^5\,\mathrm{m}^2$ which allows to collect many gamma rays to perform detailed temporal and spectral studies;
- the high duty cycle, which is the result of the possibility to observe even during moon time or when pointing to high zenith angles;
- the relatively low weight thanks to the carbon fiber structure, ~ 70 t per telescope, allowing fast rotation at $\sim 7^\circ\,\mathrm{s}^{-1}$ in the so-called "fast mode";
- the automatized reaction to GRB (and other transients) alerts, reducing the possible delay introduced by a human in the loop.

Indeed, MAGIC was designed with the detection of GRBs in the VHE range as one of its primary scientific goals, therefore many hardware and software developments aimed at improving the performance of the telescopes at the lowest energies. An important step in this direction was the development of the novel Sum-Trigger system, which was used in the past to detect pulsed emission from the Crab pulsar[2] down to 25 GeV and more recently adopted to detect the Geminga pulsar in the 15-75 GeV energy range.[3]

In these proceedings we will describe the GRB follow-up program within MAGIC and the results that could be achieved in the recent years. Finally we will discuss future prospects of GRB observations with MAGIC, highlighting the scientific and technical challenges lying ahead.

2. The MAGIC GRB follow-up program

Even before the beginning of current generation IACTs operation and the launch of the *Fermi* satellite, many theoretical studies predicted that GRBs should have been emitters of high energy and very high energy gamma rays. Such studies considered different mechanisms involving both leptonic and hadronic processes. While high energy emission was detected both in the prompt and afterglow phases of GRBs with *Fermi*-LAT, IACTs could not achieve significant detection in the VHE range for many years despite a considerable amount of hours devoted to GRB observations. The search for VHE emission from GRBs is crucial to answer some important questions, such as the following (not an exhaustive list):

- do GRBs emit at VHE?
- is VHE emission relevant e.g. from the energetic point of view?

- what is the emission process (or processes)? is it a continuation of the low energy spectrum or is it a new emission process?
- can the process producing VHE gamma rays contribute also at lower energies?
- is there VHE emission both in the prompt and the afterglow phases?
- is VHE emission present in both short and long GRBs?

In order to achieve the detection of GRBs, a specific program focuses on their follow-up, including both technical and strategy aspects.

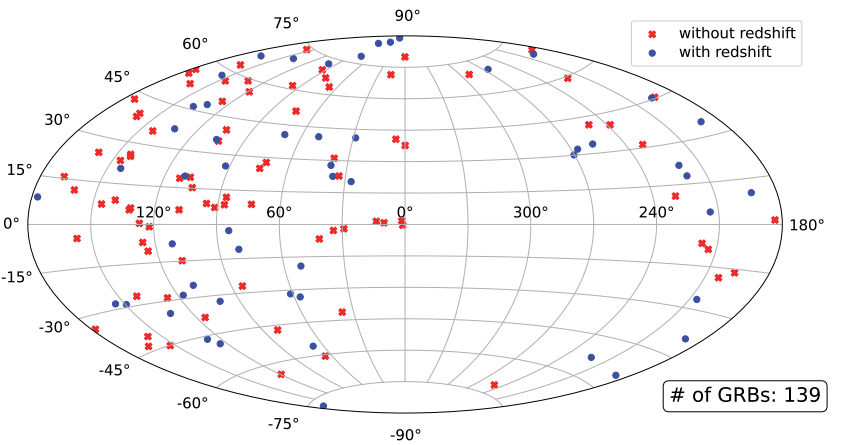


Fig. 1. Aitoff projection skymap showing the sky locations of the GRBs observed by MAGIC up to June 2021. GRBs with redshift estimation are denoted with blue filled circles, while those without redshift with red crosses.

From the technical point of view, given the small field of view of the telescopes (3.5°) and the serendipity of GRBs in both time and sky region, MAGIC relies on external alerts from other instruments. In particular, an automatic alert system (AAS) is in place since 2003 and it is used to receive GRB and other transients (high energy astrophysical neutrinos, gravitational waves, flaring *Fermi* sources, soft gamma-ray repeaters etc.) alerts, mainly from the GCN (*gamma-ray coordinates network*). After receiving a GRB alert, the AAS validates it and checks for the visibility from the MAGIC site according to some predefined criteria. If the GRB sky coordinates are immediately observable, the ongoing observation is stopped and an automatic procedure prepares the telescopes for the GRB observation. In more detail, the telescopes slew towards the GRB position in the so-called "fast mode". During the movement, other subsystems are prepared for data taking e.g. the DAQ is restarted, proper discriminator thresholds are loaded depending on the GRB position with respect to the galactic plane and to the night sky background

and mirrors are focused according to the local zenith/azimuth of the GRB. This automatic procedure was upgraded in 2013 to make it more reliable and reduce failures to a minimum. In particular, in the upgraded automatic procedure the DAQ system is not stopped but just reconfigured, and the observation is started in wobble mode, the standard pointing for MAGIC data taking.

During the years, the AAS was updated by adding new alert types and by changing the implemented follow-up strategy. For example, the AAS allows to observe GRBs even at late time with respect to the trigger time, given that *Fermi*-LAT could detect high energy emission several hours in the afterglow phase of some GRBs. A major update in the AAS was deployed in early 2021, namely the standardization to the VOEvent format.[4] Currently the AAS receives alerts from the GCN broker, but other brokers can be easily added depending on the science cases (e.g. supernovae, tidal disruption events etc.). All in all, the AAS is becoming more and more a multimessenger system, with features being added related to different follow-up strategies.

Within the GRB follow-up program, MAGIC observed 139 GRBs from 2004 up to June 2021. A skymap showing the sky positions of the observed GRBs is shown in Figure 1. Of the 139 GRBs, 53 of them have a reported redshift, of which 13 have a redshift less than 1. Most of the GRBs observed by MAGIC are long GRBs and triggered by Swift-BAT, with a minority of alerts followed-up after alerts from Fermi-GBM, Fermi-LAT and INTEGRAL (or HETE in the first years of MAGIC operation).

As far as the delay of the observation is concerned, Figure 2 shows the time delay (computed as the difference between the beginning of the observation and the trigger time of the burst) against the zenith angle at the beginning of the observation. 37 GRBs have a delay less than 100 s, 25 were followed-up within one minute from the onset. 12 GRBs were observed with a delay less than T_{90}, so that the prompt or early afterglow phases could be observed. The fastest follow-up to date was performed in the case of GRB 160821B, where the telescopes started data taking 24 s after the burst onset. Interestingly, all the GRBs detected by MAGIC or with a hint of signal were observed with a delay smaller than 100 s.

3. MAGIC results on GRB observations

3.1. *Status at MG15 meeting*

In 2018 a similar topic was presented at the 15th edition of the Marcel Grossmann meeting. At the time, MAGIC and other IACTs did not detect any GRB in the VHE range, even if a large effort was put to organize efficient GRB follow-up programs. Some interesting observations were reported by the three collaborations, namely GRB 060602B occurring in the field of view for H.E.S.S., GRB 130427A (also known as the "monster burst") by VERITAS and the short GRB 160821B by MAGIC. However no significant detection could be achieved.

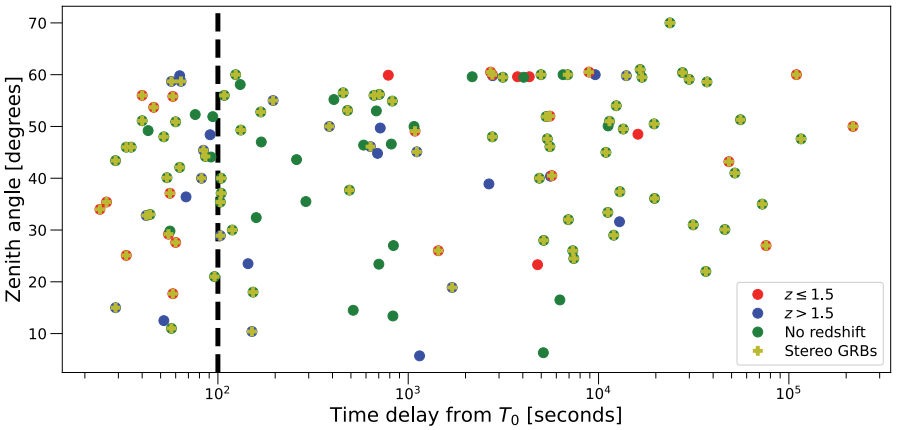


Fig. 2. Plot showing the delay of the start of the observation for GRBs observed by MAGIC up to June 2021. On the vertical axis, the zenith angle at the beginning of the observation is shown. Points with crosses denote GRBs observed in stereoscopic mode, those without were observed in mono mode.

The situation dramatically changed in 2019, which can be considered a golden year for GRB studies. Indeed, in January 2019 the MAGIC collaboration reported the historical detection of the long bright GRB 190114C in the VHE range, followed by the announcement by H.E.S.S. in May 2019 at the Cherenkov Telescope Array (CTA) Symposium in Bologna about the detection of GRB 180720B. Later in the year, at the end of August, the H.E.S.S. collaboration reported the detection of the long and very close GRB 190829A. In 2020 MAGIC reported a strong hint of signal from GRB 201015A and a detection from GRB 201216C. Therefore, the last years saw the long awaited birth of the VHE era for GRBs, finally allowing to explore the physics of these objects at such energies.

In the following section we describe the latest results by MAGIC on the search for VHE emission by MAGIC, focusing on the breakthrough detection of GRB 190114C and on the other detection of 2020.

3.2. *Recent MAGIC GRB results at MG16*

3.2.1. *GRB 190114C*

As mentioned in section 3.1, 2019 was the year starting a new era in GRB studies, which can be called the VHE era. On January 14th 2019 MAGIC detected a very significant (at the 50σ level) emission between 300 GeV and 1 TeV from the long and bright GRB 190114C,[5] initially detected by *Swift*-BAT and *Fermi*-GBM. The results on this GRB were presented in two papers, Refs. 6 and 7, one describing the MAGIC observation and detection, the second focusing on the multi-wavelength picture and on the physical interpretation of the VHE emission. In these proceedings

we will focus on the main findings from GRB 190114C observations, for more details the reader can refer to the aforementioned publications or to the contribution by D. Miceli from these proceedings series.

GRB 190114C was a bright ($E_{\rm iso} \sim 3\times10^{53}$ erg in the 1 to 10^4 keV energy range), long ($T_{90} = 362$ s as measured by *Swift*-BAT) and quite nearby ($z = 0.4245$) GRB. As shown in Figure 3, MAGIC received the alert from *Swift*-BAT 22 seconds after the onset of the GRB and started observations about 1 minute after the GRB trigger under moderate moon conditions at a relatively high zenith of 58°. From the real-time analysis the GRB was detected with a significance of 20σ in the first 20 minutes of observations above an approximate threshold of 300 GeV, later confirmed as 50σ in offline analyses, see Figure 4. The detection was reported as quickly as possible to the astrophysical community to strongly encourage the follow-up of this event at other wavelengths.

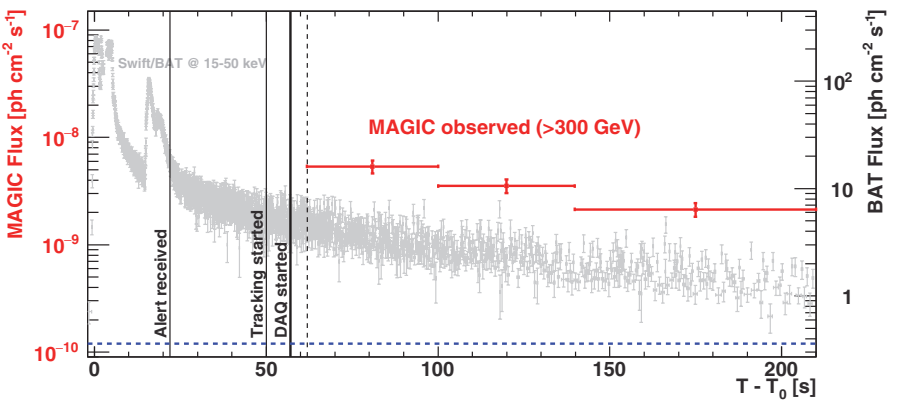


Fig. 3. Light curves for MAGIC and Swift-BAT. Vertical solid lines show the times of events related to the MAGIC automatic procedure. Extracted from Ref. 6.

One of the first questions to be answered was if the emission detected by MAGIC was related to the prompt or afterglow phase of the GRB. While the value of T_{90} may indicate that such emission could belong to the prompt, detailed spectral and temporal studies of the keV-MeV data show that at $\sim T_0 + 25$ s the properties of such low energy emission is typical of the afterglow phase. This is additionally confirmed by the similar temporal decay index between the X-ray (from *Swift*-XRT between 0.1 and 1 keV) and the VHE (between 300 GeV and 1 TeV) energy light curves.

The intrinsic spectrum of the GRB is compatible with a power law with spectral index of $\alpha_{\rm int} = -2$ between 0.2 and 1 TeV, with no indication of any break or cutoff beyond those energies at the 95% confidence level. Such flat spectrum indicated that the energy output in the VHE range is relevant, indeed turning out to be

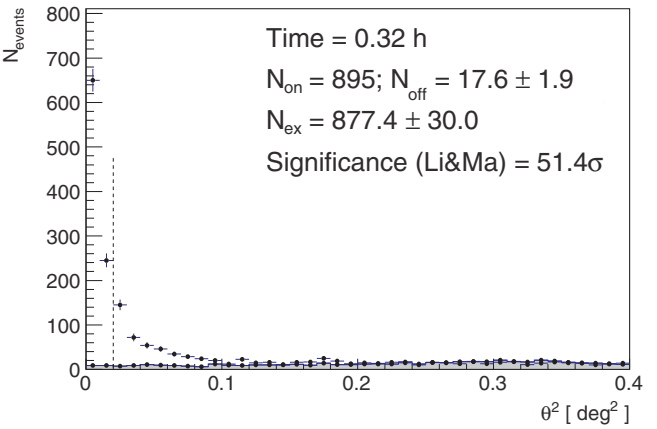


Fig. 4. Plot showing the significance of the gamma-ray signal detected by MAGIC from GRB 190114C. Extracted from Ref. 6.

comparable to the one at low energies. Given the high absorption of the VHE flux by the EBL at the redshift of the GRB, the spectrum observed by MAGIC is rather steep and best described by a power law with index $\alpha_{\rm obs} = -5.43 \pm 0.22$. This was tested against different EBL models, resulting in similar spectral indexes compatible within the statistical uncertainties.

The most important question was related to the origin of the emission detected by MAGIC. The similarity of the temporal decay in X-ray and VHE energy light curves suggests that the emission processes should be linked. While the simplest hypothesis is that the processes producing the X-ray and VHE photons are the same, namely synchrotron emission from relativistic electrons accelerated at the external shock in the afterglow of the GRB, this explanation is ruled out if one takes into account that synchrotron photons reach a maximum energy. This is the so-called burnoff limit and in the external shock of GRBs can be described by the equation $E_{\rm syn,max}(t) \approx 100(\Gamma_b(t)/1000)$GeV ($\Gamma_b(t)$ is the Lorentz factor of the GRB blastwave as a function of time). Even assuming a Lorentz factor of ~ 1000, which is not typical for GRBs and in any case it decreases with time, the maximum energy of the photons produced by synchrotron is at most around 100 GeV, even considering different density profiles of the interstellar medium density. This value is much lower than the lowest energy of the photons detected by MAGIC, therefore it is reasonable to assume that the VHE emission is due to another process. In addition, prolonging the low energy synchrotron spectrum (from GBM, XRT and LAT data) would underestimate the MAGIC flux by one order of magnitude, strengthening the conclusion that the VHE photons are indeed produced by a different mechanism. However we note that the burnoff limit interpretation assumes one emission region, while having more emission regions may allow synchrotron photons to reach higher energies.

In the assumption that the VHE emission is not due to synchrotron, the most simple alternative is the so called synchrotron-self Compton (SSC), where the synchrotron photons are upscattered by their parents electrons, thus increasing their energy. The SSC process is not new in VHE astrophysics, indeed it is commonly detected in sources like active galactic nuclei, where the spectral energy distribution shows two peaks, one at low energies due to synchrotron emission and a second one at higher energies, often at in the VHE energy range, due to the SSC process. In GRB 190114C the modeling of the GRB 190114C multi-wavelength data with a synchrotron plus SSC emission model within the external shock scenario in the afterglow shows exactly this two peaks features, therefore confirming the presence of a new emission component never observed before. Another remarkable result is the fact that the parameters describing the broadband emission of GRB 190114C have values similar to the ones found in previous studies of GRB afterglows when data only up to GeV energies were considered. This may hint to the possibility that VHE emission from SSC may be present in all GRBs and that it could be detected by IACTs if favorable conditions apply i.e. a low enough redshift and good observing conditions. This hypothesis can be confirmed only with more GRBs detected in the VHE range, as the current number of GRBs is not enough to draw firm conclusions.

3.2.2. *GRB 201015A and GRB 201216C*

The detection of GRB 190114C gave additional momentum to the searches of GRBs with MAGIC. This led to other two important observations in 2020.

The first observation concerns GRB 201015A, a burst detected by *Swift*. While the GRB showed some features resembling short GRBs, the detection of a supernova signature five days after the onset make this a long GRB.[8] Despite being at a redshift very similar to the one of GRB 190114C ($z = 0.426$), GRB 201015A has a relatively low luminosity of $E_{\rm iso} \sim 10^{50}\,{\rm erg}$, i.e. three order of magnitude lower. MAGIC started the observations after $\sim 1\,{\rm min}$ after the onset of the GRBs and it reported a hint of detection at the level of more than in 3σ from offline analyses.[9] This strong hint is very interesting if compared to the detection of the low luminosity GRB 190829A by the H.E.S.S. collaboration, located at a much lower redshift $z = 0.08$. For more details, the readers may refer to Suda et al. 2021.

In a second observation in December 2020, MAGIC detected another burst, GRB 201216C.[10] Again, this was a long GRB ($T_{90} = 48\,{\rm s}$) with similar energetics to GRB 190114C ($E_{\rm iso} \sim 5 \times 10^{53}\,{\rm erg}$), but much more distant, having a redshift of $z = 1.1$. A clear detection was achieved, and first preliminary results were presented in ICRC 2021. The most remarkable feature for this GRB is the very high redshift, enlarging the spectrum of GRB properties characterized in VHE observations.

3.2.3. *GRB 160821B*

While all GRBs currently detected at VHE belong to the long class, some progress was done also for short GRBs. Being intrinsically less energetic than long GRBs but

at relatively lower redshifts, their detection may be feasible with current IACTs. A good indication for such hypothesis is the observation of GRB 160821B by MAGIC, a short GRB detected by *Swift* located at $z = 0.16$. Additionally, short GRBs are linked with binary neutron star systems, which were proven to be gravitational waves emitters with the GW170817 event. GRB 160821B is one of the few short GRBs with a confirmed kilonova emission, pointing to a possible binary neutron star system as progenitor. As mentioned in section 2, MAGIC started data taking 24 s after the onset of GRB 160821B. The weather conditions were not favorable: moonlight was present and increasing during the observation, the zenith of the target was rather high (between 34° and 55°) and the first part of the data was affected by clouds, resulting in a reduced atmospheric transmission. Despite the unfavorable observing conditions, an evidence of a gamma-ray signal was found at a significance of $\sim 3\sigma$ above ~ 0.5 TeV.[11]

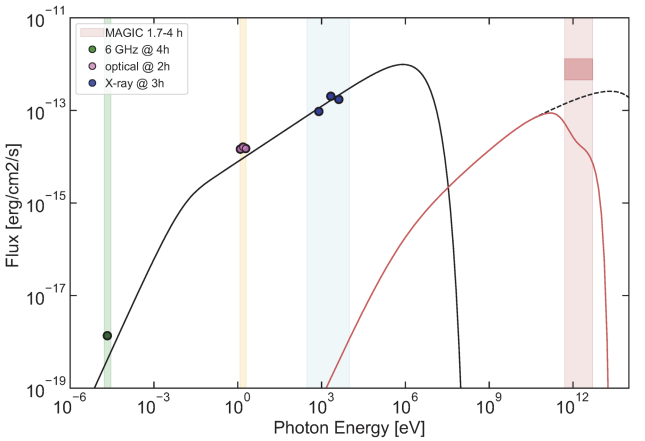


Fig. 5. Spectral energy distribution of GRB 160821B and modeling curves. The solid black line models the synchrotron radiation, while the dashed black line is the intrinsic SSC emission. The red line takes into account the absorption by the EBL. The red box only indicated the putative flux level from MAGIC. Extracted from Ref. 11.

If we assume that the gamma-ray excess is genuine emission from GRB 160821B, the flux predicted by the SSC process in the external forward shock scenario is in tension with the observed TeV flux, see Figure 5. Other scenarios were considered in order to explain the MAGIC data, from proton synchrotron emission, photohadronic cascade emission, external compton, multiple zone emission or SSC from the reverse shock. While hadronic processes are disfavored due to their inefficiency and large energy requirements, the other possibilities are not ruled out. This points to still poorly understood properties of short GRBs, which needs more observations at GeV-TeV energies in the future to fill in the picture, also in relation to gravitational wave searches.

4. Future perspectives for GRBs detection with MAGIC and conclusions

The recent detection of GRBs in the VHE regime opened a new era in this field. Such discoveries were long awaited and they are the result of a continued effort with improvements from the technical and strategy point of view. For the next years, there are other challenges that need to be faced in the search for VHE emission with MAGIC.

First of all, the number of detected GRBs is still low and they were bursts with quite different properties (e.g. distance, brightness). Therefore it is difficult to infer what could be the main factors within a GRB that favor the production of VHE gamma-rays. Despite this, a property seems to be shared among all the GRBs, namely the similar time decay between the X-ray and the VHE flux. In a recent study by Campana and collaborators instead, they discuss the possibility of a relation between a high X-ray column density and the presence of TeV emission, which seems to find confirmation considering the detected GRBs.

A second challenge for MAGIC is the detection of VHE emission in the prompt phase of a GRB. This would constitute another major breakthrough in the field, finally shedding light on which process is responsible for the initial flash of GRBs. MAGIC proved already to be able to begin GRB follow-ups very close to the onset, therefore observing during the late prompt-early afterglow phase of a dozen of GRBs. Therefore, we might expect to have a detection of VHE gamma-rays during the prompt phase to happen soon.

The third challenge was mentioned in section 3.2.3, that is the detection of short GRBs in the VHE regime. The hint of detection reported by MAGIC gives a strong indication that this goal can be achieved. In this case the number of events is quite reduced, given that the majority of GRB alerts received by MAGIC are long GRBs. However, the opportunities may increase with the beginning of the new observational run by LIGO, Virgo and Kagra expected to start in late 2022, O4. As exemplified by the GW170817/GRB170817A event, we may search for short GRBs in spatial and temporal coincidence with gravitational wave signals from the coalescence of binary neutron star or neutron star-black hole systems. As the localization of the GW events is expected to be improved, and thanks to more and more refined follow-up strategies involving both satellite-borne or ground-base telescopes, this might give MAGIC and other IACTs additional chances in detecting short GRBs.

To summarize, the recent detection of GRBs in the last years had clearly a major impact in GRB physics, and gave additional boost to searches in the VHE regime and beyond (e.g. with current and future extensive air shower arrays). The primary goal for MAGIC is to detect more GRBs, profiting of every opportunity and trying to clear the challenges mentioned above. One basic question is if SSC can model the TeV emission from GRBs in a generic way, or if other processes may instead explain it. A possible way to test such hypothesis would be to have simultaneous observation by MAGIC and Fermi-LAT partially overlapping in energy. Assuming

the two experiments give similar results in the overlapping energy region (at a few tens of GeV), a break in the MAGIC spectrum at few tens of GeV after which it rises would give a clear indication of an inverse Compton component. Such possibility is realistic given the recent result by MAGIC on the Geminga pulsar, where the SumTrigger system was used to achieve an energy threshold as low as 15 GeV so that part of the spectrum observed by MAGIC overlaps with the one by Fermi-LAT.

From the discussion in these proceedings, it is clear that the field is still young and the possibility of important discoveries is high. MAGIC proved to be up to the expectations and surely it will play a major role in the coming years as well, contributing even more to unveiling many mysteries related to GRBs.

References

1. J. Aleksić *et al.*, The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula, *Astroparticle Physics* **72**, 76 (January 2016).
2. E. Aliu *et al.*, Observation of Pulsed γ-Rays Above 25 GeV from the Crab Pulsar with MAGIC, *Science* **322**, p. 1221 (November 2008).
3. MAGIC Collaboration, Detection of the Geminga pulsar with MAGIC hints at a power-law tail emission beyond 15 GeV, *Astronomy & Astrophysics* **643**, p. L14 (November 2020).
4. R. Seaman *et al.*, Sky Event Reporting Metadata Version 2.0 IVOA Recommendation 11 July 2011 (July, 2011).
5. R. Mirzoyan *et al.*, MAGIC detects the GRB 190114C in the TeV energy domain., *GRB Coordinates Network* **23701**, p. 1 (January 2019).
6. MAGIC Collaboration *et al.*, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).
7. MAGIC Collaboration *et al.*, Observation of inverse Compton emission from a long γ-ray burst, *Nature* **575**, 459 (November 2019).
8. A. Rossi *et al.*, GRB 201015A: evidence of supernova in LBT spectra, *GRB Coordinates Network* **29306**, p. 1 (January 2021).
9. O. Blanch *et al.*, MAGIC observations of GRB 201015A: hint of very high energy gamma-ray signal, *GRB Coordinates Network* **28659**, p. 1 (October 2020).
10. O. Blanch *et al.*, GRB 201216C: MAGIC detection in very high energy gamma rays, *GRB Coordinates Network* **29075**, p. 1 (December 2020).
11. V. A. Acciari *et al.*, MAGIC Observations of the Nearby Short Gamma-Ray Burst GRB 160821B, *The Astrophysical Journal* **908**, p. 90 (February 2021).

Prospects for VHE monitoring of gamma-ray bursts with SWGO

G. La Mura,[1,*] U. Barres de Almeida,[2] R. Conceição,[1,3] A. de Angelis,[4,5,6] F. Longo,[7,8,9] M. Pimenta,[1,3] E. Prandini,[6,10] E. Ruiz-Velasco,[11] B. Tomé[1,3]
for the SWGO Collaboration

[1] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Av. Prof. Gama Pinto 2, Lisboa 1649-003, Portugal*
[*] *E-mail: glamura@lip.pt*
[2] *Brazilian Center for Physics Research (CBPF), Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro, Brazil*
[3] *Instituto Superior Técnico (IST), Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*
[4] *Dipartimento di Fisica e Astronomia - Università di Padova, Via Marzolo 8, 35131 Padova, Italy*
[5] *Dipartimento di scienze matematiche, informatiche e fisiche - Università degli Studi di Udine, Via Palladio 8, 33100 Udine, Italy*
[6] *Istituto Nazionale di Fisica Nucleare sez. Padova (INFN), Via Marzolo 8, 35131 Padova, Italy*
[7] *IFPU - Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy*
[8] *INFN, Sezione di Trieste, via A. Valerio 2, 34100 Trieste, Italy*
[9] *Dipartimento di Fisica, Università degli Studi di Trieste, via A. Valerio 2, 34100 Trieste, Italy*
[10] *INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova, Italy*
[11] *Max-Planck-Institut für Kernphysik, P.O. Box 103980, 69029 Heidelberg, Germany*

It has been established that Gamma-Ray Bursts (GRB) can produce Very High Energy radiation (VHE, $E > 100\,\mathrm{GeV}$), opening a new window for investigating particle acceleration and radiation properties in the most energetic domain. We expect that next-generation instruments in this energy regime, such as the Cherenkov Telescope Array (CTA), will mark a huge improvement in their observation. However, constraints on the target visibility and the limited duty cycle of Imaging Atmospheric Cherenkov Telescopes (IACT), affect their ability to react promptly to transient events. Here we use a grid of instrument performance estimates, based on the Extensive Air Shower (EAS) array concept proposed by the Southern Wide Field-of-view Gamma-ray Observatory (SWGO) collaboration, to evaluate SWGO's potential to detect and track VHE emission from GRBs. Observations by the *Fermi* Large Area Telescope (*Fermi*–LAT) at high energy ($E > 10\,\mathrm{GeV}$), identified some events with a distinct spectral component, which can represent a substantial fraction of the emitted energy and possibly occur very early in the process. Using models based on these properties, we estimate the possibilities that a wide field of view and large effective area ground-based monitoring facility has to probe VHE emission from GRBs. We show that the ability to monitor VHE transients with a nearly continuous scanning of the sky grants us the opportunity to simultaneously observe electromagnetic counterparts to gravitational waves and relativistic particles sources up to cosmological scales, in a way that is not accessible to IACTs.

Keywords: Instrumentation: detectors; gamma rays: general; gamma ray burst: general.

1. Introduction

The study of Gamma-Ray Bursts (GRB) represents one of the most challenging and fascinating frontiers of High Energy Astrophysics. Although they were discovered in the first years of space explorations with high energy detectors,[1] several aspects of their origin, their power, and their radiation mechanisms are still unclear. We know that GRBs are originated by ultra-relativistic jets of plasma, which are produced during the extremely fast accretion process that follows the collapse of a very massive star ($M \geq 20\,\mathrm{M}_{\odot}$) or the merger of a binary neutron star (NS), to form a *magnetar* or a black hole.[2] The process produces a flash of energetic photons, named *prompt emission*, followed by a smoothly decaying signal, called *afterglow*. Looking at the duration of the prompt phase, GRBs appear to show a bimodal distribution, where we distinguish a class of short events, with a prompt stage lasting less than 2 s, and another of long ones, whose prompt emission takes place for much longer times.[3,4] The two classes of events are consistent with two distinguished possible source types. Long GRBs have been associated with supernova explosions,[5,6] while the short ones fit better in the compact binary merger scenario,[7] as eventually demonstrated by the GRB 170817A–GW 170817 multi-messenger association that was also detected as a *kilonova*.[8–10]

In the past few years, observations performed with ground based Imaging Atmospheric Cherenkov Telescopes (IACT), such as MAGIC[11] and H.E.S.S.,[12] found evidence for Very High Energy γ-ray emission (VHE, $E \geq 100\,\mathrm{GeV}$) in the afterglow of some GRBs.[13,14] In addition, analysis of the data collected by the *Fermi* Large Area Telescope (*Fermi*-LAT)[15] demonstrated the existence of a high energy component in the spectra of some bright GRBs that extends above the 10 GeV energy range and can be detected also in association with the prompt emission.[16–18] Understanding the origin and the distribution of VHE radiation from GRBs is a key issue for their correct interpretation. It is expected that next generation γ-ray instruments, such as, for instance, the Cherenkov Telescope Array (CTA),[19] will provide exceptional new data in this field. Unfortunately, the unpredictable nature of GRBs and their fast spectral evolution makes it extremely difficult, for instruments with narrow field of view (FoV), of the order of few square degrees, to re-point and track GRB emission within very short times (below 10 s). A possible solution to this problem is using Extensive Air Shower detector arrays (EAS) to monitor large areas of the sky, with an instantaneous FoV of more than 1 sr and a nearly continuous duty cycle. Here we discuss the observational opportunities that a new instrument, based on a next-generation EAS array concept investigated by the Southern Wide-field-of-view Gamma-ray Observatory (SWGO),[20] will open in the study of GRBs.

This contribution is structured as follows: in §2 we describe the known and the expected VHE properties of GRBs; in §3 we present the potential monitoring

capabilities of SWGO; finally, in §4 we discuss our results and we draw our conclusions.

2. VHE emission from GRBs

Several arguments suggest that GRBs should be able to produce prompt VHE radiation. The most likely interpretation of GRBs invokes relativistic shocks occurring either within the magnetized jet plasma (*internal shocks*, generally associated with the prompt stage), or between the jet and the surrounding environment (*external shocks*, more often considered as the source of the afterglow emission). In both cases, the energies associated with particles and magnetic fields are high enough to produce efficient leptonic radiation mechanisms and possibly to activate hadronic processes.[21–23] The combination of temporal and spectral properties of the emitted radiation provides fundamental indications on the characteristics of the source.

In general, a large fraction of the energy that a GRB releases in the prompt phase is found in the spectral range falling between few hundreds keV to some MeV.[24] The spectrum comes in the form of a combination of power-law components, known as the *Band function*.[25] The presence of relativistic charged particles in an intense magnetic field leads to the natural expectation that powerful synchrotron and inverse Compton radiation should be emitted.[26] While the prompt emission is characterized by fast pulse-like variability, with time-scales down to few milliseconds,[27] implying extremely compact emission regions, the observation of photons up to the MeV and even the GeV energy range requires very large bulk Lorentz factors ($\Gamma >> 100$), in order to suppress pair production opacity. In these conditions, the inverse Compton scattering of synchrotron radiation may lead to the production of photons up to the TeV domain.[28] The occurrence of ultra-relativistic shocks in high density environments, on the other hand, is theoretically able to activate photo-pion interactions that can extend the spectra beyond 300 GeV[23] and lead to potential associations with neutrino events.[29]

So far, the direct detection of VHE radiation has only been possible in the afterglow of a limited number of GRBs, thanks to ground-based follow up observations.[13,30–32] However, the monitoring campaign carried out by the *Fermi*-LAT space observatory led to the identification of a high energy spectral component, coming in the form of power-law emission, that appears to be a common feature of bright GRBs and may even arise very early in the event.[33] Due to the limited collecting area and the typical short duration of GRBs, the LAT observations are not able to place strong constraints on the VHE component of GRBs, as it is suggested by the fact that *Fermi* did not detect photons above ~ 30 GeV from GRB 190114C, while the same event was firmly detected up to almost 1 TeV. As a consequence, the results of *Fermi*-LAT observations can be used as a starting point to estimate the possible extension of the GRB properties to the VHE domain and thereby evaluate their detection opportunities for other instruments.

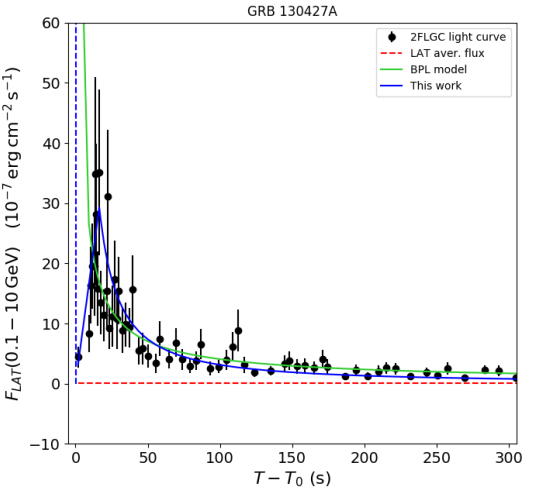


Fig. 1. Comparison between the *Fermi*-LAT light curve of GRB 130427A and the model based on Eq. (1) and Eq. (2). The vertical blue dashed line marks the start of the LAT signal temporal window, the red horizontal line is the average energy flux collected during the entire time of emission, the green continuous line is the 2FLGC power-law fit to the data, while the blue continuous line is a model using the light curve of Eq. (2).

Usually, we can express the high energy spectrum of a GRB as a function of energy in the form of:[a]

$$\frac{\mathrm{d}N(t)}{\mathrm{d}E} = N_0(t)\left(\frac{E}{E_0}\right)^{-\alpha}\exp[-\tau(E,z)] \quad [\mathrm{ph\,cm^{-2}\,s^{-1}\,GeV^{-1}}], \tag{1}$$

where $N_0(t)$ is the flux of photons per unit energy observed at time t, E_0 is the pivot energy, α is the spectral index, which is often within the range $1.5 \leq \alpha \leq 3$, with an average value close to 2, and $\tau(E,z)$ is the opacity due to pair production on Extragalactic Background Light photons (EBL), given as a function of energy and redshift. The temporal evolution of the flux is typically well represented by a power-law, or a broken power-law, that can be written as:

$$N_0(t) = \begin{cases} N_{peak}\left(\dfrac{t-T_0}{T_{peak}-T_0}\right) & \text{for} \quad T_0 \leq t \leq T_{peak} \\ N_{peak}\left(\dfrac{t}{T_{peak}-T_0}\right)^{-\gamma} & \text{for} \quad t > T_{peak}, \end{cases} \tag{2}$$

where we denoted with T_0 the trigger time, with T_{peak} the time taken to achieve peak emission, with N_{peak} the maximum flux, and with γ the temporal evolution index, which is often found to be $1 \leq \gamma \leq 2$.

[a]We assume for simplicity that the temporal evolution of the spectrum is only limited to a scaling factor, without relevant spectral changes.

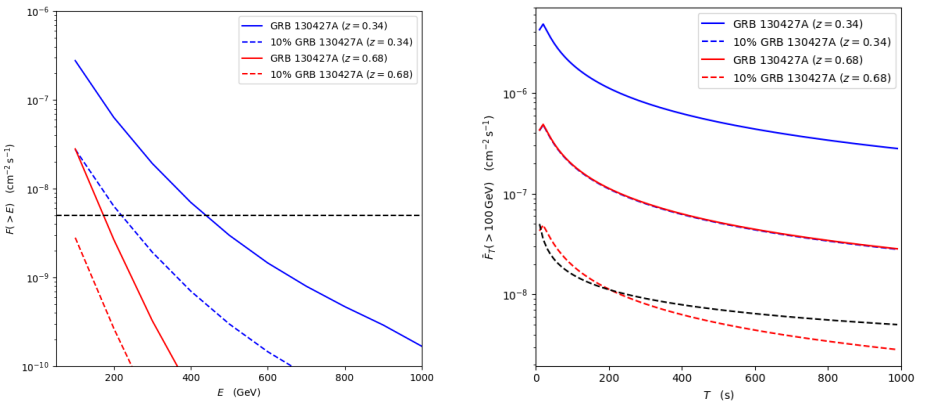


Fig. 2. **Left panel:** average photon flux expected above any given energy in 1000 s for a burst with the spectral and temporal characteristics of GRB 130427A (blue continuous line), an identical burst with 10% its strength (blue dashed line) and the same previous cases computed for twice the measured redshift (red continuous and dashed lines). The horizontal black dashed line represents a reference flux limit of $5 \cdot 10^{-9}$ ph cm^{-2} s^{-1}. **Right panel:** instantaneous integral photon fluxes, expected above 100 GeV as afunction of time, compared with the detection threshold, in the assumption that the limiting flux scales as the square root of the observation time.

Using the second catalog of *Fermi*-LAT detected GRBs (2FLGC),[33] which provides measurements of the observed photon fluxes, in the energy range between 0.1 GeV and 10 GeV, together with information on the spectral index and on the light curve shape, for a sample of GRBs observed during 10 years of regular monitoring operations, we are able to apply Eq. (1), with the inclusion of Eq. (2), to estimate the expected high energy fluxes as a function of time, as it is illustrated for instance in Fig. 1. In principle, we can extend this type of spectra to the VHE domain and, thus, obtain an estimate of the expected fluxes. In practice, this operation is not directly possible, due to the lack of a redshift measurement for the majority of the LAT detected GRBs, which implies an unknown EBL opacity in Eq. (1). Although the effects EBL are generally negligible for the observed LAT band, they become quickly very important at higher energies, with a typical γ-ray horizon set by $\tau = 1$, which, for $z \approx 1$, already occurs at $E = 100$ GeV.[34] For this reason, we combined the spectral and temporal fits, which we obtained from the LAT data, with a set of simulations, aiming at estimating the effects of EBL opacity on the VHE extension of the GRBs that resulted in the brightest LAT observed fluxes.[35]

3. GRB monitoring with SWGO

A direct consequence of Eq. (1) is that the observation of VHE radiation, particularly from fast transient sources, requires the use of large instrumented areas. This characteristic can only be offered by ground-based facilities, such as IACTs or EAS

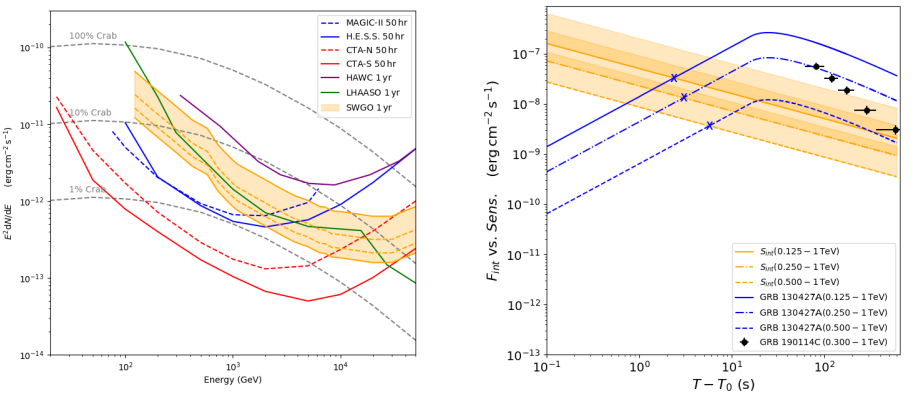


Fig. 3. **Left panel:** Spectral coverage and differential sensitivity of some currently operating and next-generation VHE observatories. The sensitivities are computed for a point-like source at a zenith distance of 20°. Different fractions of the Crab Nebula flux are shown for comparison. The shaded area is the sensitivity range expected for various configurations of SWGO.[36] **Right panel:** The expected integrated flux for a GRB 130427A-like event above 125 GeV, 250 GeV and 500 GeV, compared with the corresponding SWGO expected sensitivities. Crosses mark the estimated detection times. The VHE light curve of GRB 190114C above 300 GeV is also shown.[13]

detector arrays. The main advantage of IACT observatories is that they have excellent angular resolution and background rejection power, achieving the best possible sensitivity. On the other hand, these instruments can only operate during dark time and in clear sky conditions, implying a duty cycle of only $\sim 20\%$. In addition, the small FoV and the consequent need to respond to an external alert with target position information and subsequent re-pointing delay does not allow IACT instruments to cover the early GRB emission regime, where many critical observational constraints are expected by different theoretical models.

Thanks to their larger FoV and to the possibility to operate almost continuously, EAS detector arrays have better chances to cover a GRB and to act as an alert system themselves. The main problems of this solution are that EAS arrays need to be located at high altitudes (> 4500 m a.s.l.), in order to be reached by the particle showers initiated by γ-ray primaries with $E \leq 1$ TeV, and that they need to deal with a strong background of cosmic-ray induced air showers. The sensitivity needed to detect the expected VHE signal of GRBs can be estimated by integrating Eq. (1) in time and in energy, with the observed properties of bright LAT bursts, using known or simulated redshifts. Figure 2 shows an example of such calculation. As it is clearly illustrated by the left panel of Fig. 2, the expected signal is a function of the characteristic burst properties, such has flux, duration, spectral index and redshift, and of the energy band that is covered by the observation. In particular, the amount of flux above 100 GeV is critically affected by redshift, meaning that only bright and relatively nearby events can be detected at larger energies. Taking as reference a limiting flux of $5 \cdot 10^{-9}$ ph cm^{-2} s^{-1}, integrated above 100 GeV for an

observing time of 1000 s, which is a reasonable estimate for an instrument concept based on the SWGO reference design,[36] it is possible to reach the VHE signal expected from GRBs down to 10% the power of the brightest LAT detected event up to $z \approx 0.7$.

The SWGO Collaboration is currently investigating the configuration of a next-generation EAS array, based on Water Cherenkov Detectors (WCD), to be installed at a high altitude site in the Southern Hemisphere. By adopting a detector concept able to trigger on low energy shower particles ($\sim$ 20 MeV), with high temporal resolution ($\sim$ 2 ns), it can be shown that an array of 80 000 m^2 can improve background suppression and achieve very good monitoring sensitivities, below the TeV energy domain.[36] Figure 3 presents an illustration of this concept, compared with other currently operating and planned facilities, together with an example of its expected performance, with respect to the temporal evolution of two GRBs with a detected high energy component. If we define the integrated flux expected in the VHE domain within an observation time T as:

$$F_{int}(T) = \frac{1}{T - T_0} \int_{T_0}^{T} \mathrm{d}t \int_{E_{low}}^{1\,\mathrm{TeV}} E \frac{\mathrm{d}N(t)}{\mathrm{d}E} \mathrm{d}E, \tag{3}$$

where the spectrum is defined through Eq. (1) and Eq. (2), we can easily verify that the times required to detect the incoming fluxes are significantly below 10 s, meaning that such configurations would be particularly effective to react to the early phases of the transient. Given the promising detection chances in case of GRBs with well measured energetic components, SWGO represents an attractive solution to investigate the properties of VHE emission in the early stages of GRBs and, thus, help solving the question of whether an energetic component associated with the burst onset is a common feature of GRBs or a distinguishing property of a specific event class.

4. Conclusions

The systematic investigation of GRBs in the VHE range will cover a fundamental role in the quickly evolving field of Multi-Wavelength and Multi-Messenger Astrophysics. After the first detection of Gravitational Waves (GW), which marked a corner stone in scientific research,[37] the execution of regular monitoring campaigns increased the chances of obtaining multiple detections of fast transients. The associated observation of GW 170817 and of the short GRB 170817A, in particular, represented the first direct evidence of the compact binary merger as a viable explanation to short GRBs and it opened the way to a wealth of cosmological and high energy physics tests.[8,9] A fundamental question related to the nature of GRBs and to their distinction in classes is whether the formation of the jet and its subsequent evolution involve substantial hadronic processes, a matter that could be unambiguously solved by the association of GRBs with neutrino events or with early VHE emission.[38]

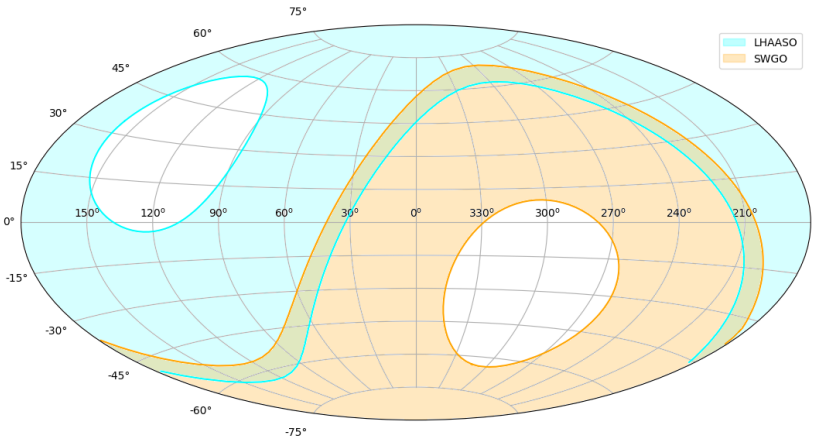


Fig. 4. Representation of the visible sky, within 30° from zenith, showing the sky regions covered by LHAASO (cyan shaded area) and by SWGO, assuming an observatory latitude close to 23° S (orange shaded area). The map is plotted in Galactic coordinates.

Evidence based on existing observations has firmly demonstrated that GRBs can produce energetic radiation and also that this spectral component may be associated with the elusive prompt emission.[17] The observation of early high energy photons, together with hints of the VHE detection of short GRBs, like in the case of GRB 160821B, challenges the external shock model predictions and suggests that additional mechanisms may be at work. Covering the VHE window, particularly in the early emission phase, will be a crucial requirement for the development of more advanced models. The ability to characterise the earliest properties of VHE emission will be fundamental both to improve the time-domain investigation of Multi-Messenger triggers, as well as to offer high quality alerts for follow-up observations that, as demonstrated by H.E.S.S. and further boosted by the upcoming investigations with CTA, can track the VHE evolution up to several hours in the afterglow.[30]

The large FoV and the nearly continuous operating time of EAS arrays make this type of instruments an ideal facility to survey the sky looking for fast transient sources. Their ability to cover large sky areas will help constraining the VHE properties of early GRB emission, with new implications on the involved radiation mechanisms. However, in order to observe events located at cosmological distances, they need to work effectively in the sub-TeV energy domain and, therefore, to be located in high altitude sites. The Large High Altitutde Air Shower Observatory (LHAASO),[39] in the Northern hemisphere, and SWGO, from the Southern hemisphere, have the potential to carry out a VHE monitoring program that will cover a large fraction of the visible sky, as illustrated in Fig. 4. If extended in the sub-TeV domain, this nearly constant scanning of the sky will help clarifying the VHE properties of GRBs by assessing the existence of spectral components that, though

predicted in well justified models, are hard to detect with instruments that need to be triggered and subsequently pointed to the target. In addition, fostering a VHE monitoring ability will offer a new window to identify sources of energetic transients, like gravitational waves and high energy neutrinos. Clearly, the identification of possible counterparts to alerts issued by continuously operating experiments, such as *IceCube* and *LIGO/VIRGO*, will undoubtedly benefit from the existence of a network of VHE monitoring programs, able to detect energetic transients and, thus, to further refine the identification of their sources and their energy production mechanisms.

Acknowledgments

The SWGO Collaboration acknowledges the support from the agencies and organizations listed here: `https://www.swgo.org/SWGOWiki/doku.php?id=acknowledgements`. This research was funded by Fundação para a Ciência e Tecnologia, under project PTDC/FIS-PAR/4300/2020 and grant DL57/2016/cP1330/cT0002. Additional funding by CNPq Productivity Research, Grant no. 311997/2019-8, by Serrapilheira Institute Grant number Serra - 1812-26906, and by FAPERJ Young Scientist Fellowship no. E-26/202.818/2019 is also acknowledged.

References

1. R. W. Klebesadel, I. B. Strong and R. A. Olson, Observations of Gamma-Ray Bursts of Cosmic Origin, *The Astrophysical Journal Letters* **182**, p. L85 (June 1973).
2. S. E. Woosley, Gamma-Ray Bursts from Stellar Mass Accretion Disks around Black Holes, *The Astrophysical Journal* **405**, p. 273 (March 1993).
3. E. P. Mazets, S. V. Golenetskii, V. N. Ilinskii, V. N. Panov, R. L. Aptekar, I. A. Gurian, M. P. Proskura, I. A. Sokolov, Z. I. Sokolova and T. V. Kharitonova, Catalog of cosmic gamma-ray bursts from the KONUS experiment data, *Astrophysics and Space Science* **80**, 3 (November 1981).
4. J. P. Norris, T. L. Cline, U. D. Desai and B. J. Teegarden, Frequency of fast, narrow γ-ray bursts, *Nature* **308**, 434 (March 1984).
5. T. J. Galama *et al.*, An unusual supernova in the error box of the γ-ray burst of 25 April 1998, *Nature* **395**, 670 (October 1998).
6. J. S. Bloom *et al.*, The unusual afterglow of the γ-ray burst of 26 March 1998 as evidence for a supernova connection, *Nature* **401**, 453 (September 1999).
7. D. Eichler, M. Livio, T. Piran and D. N. Schramm, Nucleosynthesis, neutrino bursts and γ-rays from coalescing neutron stars, *Nature* **340**, 126 (July 1989).
8. B. P. Abbott *et al.*, Multi-messenger Observations of a Binary Neutron Star Merger, *The Astrophysical Journal Letters* **848**, p. L12 (October 2017).
9. B. P. Abbott *et al.*, Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *The Astrophysical Journal Letters* **848**, p. L13 (October 2017).
10. P. K. Blanchard *et al.*, The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. VII. Properties of the Host Galaxy and Constraints on the Merger Timescale, *The Astrophysical Journal Letters* **848**, p. L22 (October 2017).

11. J. Aleksić *et al.*, The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system, *Astroparticle Physics* **72**, 61 (January 2016).
12. F. Aharonian *et al.*, Observations of the Crab nebula with HESS, *Astronomy & Astrophysics* **457**, 899 (October 2006).
13. MAGIC Collaboration, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).
14. H. Abdalla *et al.*, A very-high-energy component deep in the γ-ray burst afterglow, *Nature* **575**, 464 (November 2019).
15. W. B. Atwood *et al.*, The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission, *The Astrophysical Journal* **697**, 1071 (June 2009).
16. A. A. Abdo *et al.*, Fermi Observations of High-Energy Gamma-Ray Emission from GRB 080916C, *Science* **323**, p. 1688 (March 2009).
17. K. Asano, S. Guiriec and P. Mészáros, Hadronic Models for the Extra Spectral Component in the Short GRB 090510, *The Astrophysical Journal Letters* **705**, L191 (November 2009).
18. M. Ackermann *et al.*, Fermi Observations of GRB 090510: A Short-Hard Gamma-ray Burst with an Additional, Hard Power-law Component from 10 keV to GeV Energies, *The Astrophysical Journal* **716**, 1178 (June 2010).
19. CTA Consortium, *Science with the Cherenkov Telescope Array* (World Scientific Publishing Company, 2019).
20. U. Barres de Almeida, G. Giacinti and F. Longo, Benchmarking the Science for the Southern Wide-Field Gamma-ray Observatory (SWGO), *Proceedings of Science* **395**, p. 10 (July 2021).
21. J. I. Katz, Delayed Hard Photons from Gamma-Ray Bursts, *The Astrophysical Journal Letters* **432**, p. L27 (September 1994).
22. M. J. Rees and P. Meszaros, Unsteady Outflow Models for Cosmological Gamma-Ray Bursts, *The Astrophysical Journal Letters* **430**, p. L93 (August 1994).
23. M. Vietri, GeV Photons from Ultrahigh Energy Cosmic Rays Accelerated in Gamma Ray Bursts, *Physical Review Letters* **78**, 4328 (June 1997).
24. D. Gruber *et al.*, The Fermi GBM Gamma-Ray Burst Spectral Catalog: Four Years of Data, *The Astrophysical Journal Supplement Series* **211**, p. 12 (March 2014).
25. D. Band *et al.*, BATSE Observations of Gamma-Ray Burst Spectra. I. Spectral Diversity, *The Astrophysical Journal* **413**, p. 281 (August 1993).
26. H. Papathanassiou and P. Meszaros, Spectra of Unsteady Wind Models of Gamma-Ray Bursts, *The Astrophysical Journal Letters* **471**, p. L91 (November 1996).
27. E. Nakar and T. Piran, Temporal properties of short gamma-ray bursts, *Monthly Notices of the Royal Astronomical Society* **330**, 920 (March 2002).
28. T. Piran, The physics of gamma-ray bursts, *Reviews of Modern Physics* **76**, 1143 (October 2004).
29. M. G. Aartsen *et al.*, An All-sky Search for Three Flavors of Neutrinos from Gamma-ray Bursts with the IceCube Neutrino Observatory, *The Astrophysical Journal* **824**, p. 115 (June 2016).
30. H. Abdalla *et al.*, A very-high-energy component deep in the γ-ray burst afterglow, *Nature* **575**, 464 (November 2019).
31. O. Blanch, GRB 201216C: MAGIC detection in very high energy gamma rays, *The Astronomer's Telegram* **14275**, p. 1 (December 2020).
32. H. E. S. S. Collaboration, Revealing x-ray and gamma ray temporal and spectral similarities in the GRB 190829A afterglow, *Science* **372**, 1081 (June 2021).

33. M. Ajello *et al.*, A Decade of Gamma-Ray Bursts Observed by Fermi-LAT: The Second GRB Catalog, *The Astrophysical Journal* **878**, p. 52 (June 2019).
34. A. Saldana-Lopez, A. Domínguez, P. G. Pérez-González, J. Finke, M. Ajello, J. R. Primack, V. S. Paliya and A. Desai, An observational determination of the evolving extragalactic background light from the multiwavelength HST/CANDELS survey in the Fermi and CTA era, *Monthly Notices of the Royal Astronomical Society* **507**, 5144 (November 2021).
35. G. La Mura, U. Barres de Almeida, R. Conceição, A. De Angelis, F. Longo, M. Pimenta, E. Prandini, E. Ruiz-Velasco and B. Tomé, Gamma-ray burst detection prospects for next generation ground-based VHE facilities, *Monthly Notices of the Royal Astronomical Society* **508**, 671 (November 2021).
36. H. Schoorlemmer, R. Conceição and A. J. Smith, Simulating the performance of the Southern Wide-view Gamma-ray Observatory, *Proceedings of Science* **395**, p. 9 (July 2021).
37. B. P. Abbott *et al.*, Properties of the Binary Black Hole Merger GW150914, *Physical Review Letters* **116**, p. 241102 (June 2016).
38. M. Bustamante, J. Heinze, K. Murase and W. Winter, Multi-messenger Light Curves from Gamma-Ray Bursts in the Internal Shock Model, *The Astrophysical Journal* **837**, p. 33 (March 2017).
39. Z. Cao *et al.*, Introduction to Large High Altitude Air Shower Observatory (LHAASO), *Chinese Astronomy and Astrophysics* **43**, 457 (October 2019).

Theoretical implications on the very high energy emission from GRB 190114C

D. Miceli[1,2,*], A. Berti[3], Z. Bosnjak[4], S. Covino[5], S. Fukami[6], S. Inoue[7], F. Longo[2,8,9], R. Mirzoyan[3], E. Moretti[10], L. Nava[2,5,9], K. Noda[6], D. Paneque[3], A. Stamerra[11], Y. Suda[12], I. Vovk[6] on behalf of the MAGIC Collaboration

[1] *Laboratoire d'Annecy de Physique des Particules (LAPP), CNRS-IN2P3, 9 Chemin de Bellevue - BP 110, Annecy Cedex, France*
[*] *E-mail: davide.miceli@lapp.in2p3.fr*
[2] *INFN, Sezione di Trieste, via Valerio 2, Trieste, Italy*
[3] *Max Planck Institute for Physics, Föhringer Ring 6, Munich, Germany*
[4] *Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia*
[5] *INAF, Osservatorio Astronomico di Brera, Merate, Italy*
[6] *Institute for Cosmic Ray Research, The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Japan*
[7] *RIKEN, Hirosawa 2-1, Wako, Japan*
[8] *University of Trieste, Department of Physics, via Valerio 2, Trieste, Italy*
[9] *Institute for Fundamental Physics of the Universe (IFPU), via Beirut 12, Trieste, Italy*
[10] *Institut de Física d'Altes Energies (IFAE), Edifici Cn, Bellaterra (Barcelona), Spain*
[11] *National Institute for Astrophysics (INAF), I-00136 Rome, Italy*
[12] *Physics Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 739-8526 Hiroshima, Japan*

Since their discovery in the late 1960s Gamma-Ray Burst (GRB) emission has been deeply investigated with the help of the huge amount of data collected covering the entire electromagnetic spectrum. This large and broadband dataset was essential to constitute a general picture describing the GRB physics, revealing the most credible underlying physical processes and environmental conditions ongoing at the GRB site. A key leap in the comprehension of the GRB physics have been achieved recently, thanks to the detection of the newly energetic component in the Very High Energy (VHE, E $>$ 100 GeV) domain. The possible presence of a TeV spectral window in GRBs was predicted and theorized for several decades, but the first observational proofs of its existence were reached only in 2019 thanks to the discoveries claimed by the MAGIC and H.E.S.S. telescopes. GRB190114C was successfully detected in the TeV band by the MAGIC telescopes starting from around one minute after its trigger time and lasting for nearly 40 minutes. A successful follow-up campaign was performed and the multi-wavelength afterglow emission of the event was collected from 1 to about 2×10^{17} GHz. Such very broad dataset allows to perform unique studies on the radiation mechanisms and on the physical properties of such event. In this contribution I will describe the main results and the theoretical interpretations that have been derived from the multi-wavelength dataset of GRB190114C. In particular, the description of the TeV component detected by the MAGIC telescopes as produced via the Synchrotron Self-Compton (SSC) mechanism and its connection with the emission at lower energy bands will be presented. Such studies are a fundamental starting point for the interpretation of the current and upcoming events that will be observed in the VHE domain.

Keywords: Gamma-ray bursts; very high energy; synchrotron-self compton

1. Introduction

Gamma-Ray Bursts (GRBs) are transient phenomena produced during cataclysmic events such as the death of massive stars or the merging of binary compact objects like neutron stars and black holes.

The GRB electromagnetic radiation is characterized by initial rapid and irregular flashes bright in the keV-MeV band, the so-called *prompt* emission phase followed by a long-lasting broadband fainter emission, named *afterglow*. The duration of the prompt emission spans from milliseconds to thousands of seconds and is used to classify GRBs as *short* and *long* with a separation value of 2 s. The afterglow phase, which follows or partially overlaps with the prompt emission, is usually detectable within timescale of days, weeks or in some cases even months.

Because of its multi-wavelength nature, intense campaigns are performed to search for and follow this broadband emission from radio up to γ-rays, both with ground-based telescopes and space-born satellites. While the prompt emission mechanism is still unclear, the afterglow radiation from radio up to GeV is well described as synchrotron emission produced in the external forward shock scenario due to the interaction between the jet and the surrounding interstellar medium or stellar wind (depending on the progenitor). The possible extension of the GRB emission in the high energy (0.5 MeV $<$ E $<$ 100 GeV, HE) and to the very high energy domain (VHE, E $>$ 100 GeV) has always been one of the most debated open questions in GRB physics. The HE observations have revealed the presence of an emission component with peculiar temporal and spectral properties. The highest energy photons observed by Fermi-LAT are hardly to reconcile with the standard synchrotron afterglow scenario and hint the possibility that a new VHE emission component in GRBs is present.

A firm conclusion could be reached thanks to the observations performed by ground-based imaging Cherenkov telescopes. In particular, the MAGIC and the H.E.S.S. telescopes revealed unequivocally the presence of a VHE emission component in GRB afterglows up to TeV energies.[1] Such detections gave birth to unique studies covering several open topics of the GRB physics. In this contribution we report the analysis results and the theoretical implications derived from the TeV detection of GRB 190114C performed by the MAGIC telescopes. The interpretation of the full GRB 190114C multi-wavelength dataset with the synchrotron and Synchrotron-Self Compton (SSC) scenario and the intimate connection between the TeV emission component and the lower energy ones is presented.[2,3]

2. TeV detection and interpretation of GRB 190114C

GRB 190114C is a long GRB triggered by Swift-BAT and Fermi-GBM space instruments on 14 January 2019, 20:57:03 UT (hereafter T_0). The event was detected also by several other space instruments, namely Fermi-LAT, Swift-XRT, Swift-UVOT, AGILE, INTEGRAL/SPI-ACS, Insight/HXMT, and Konus-Wind. Triggered by space satellite alerts, the event was then followed-up and detected by

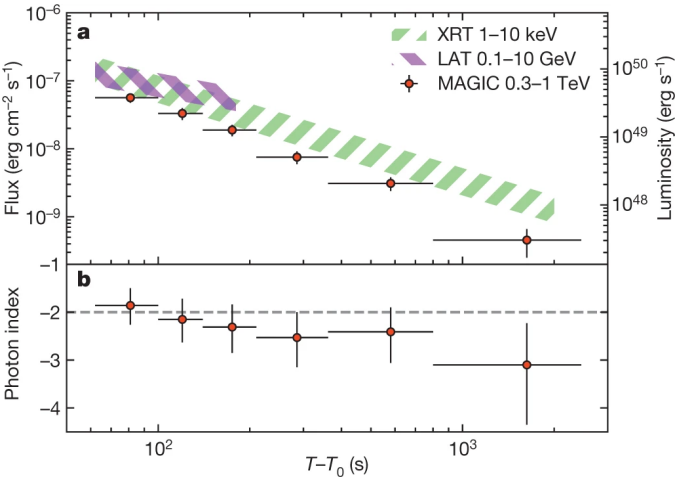


Fig. 1. **a**: MAGIC (red circles), XRT (green band) and LAT (red band) light curves. **b**: temporal evolution of the intrinsic spectral photon index.[2]

the MAGIC telescopes in the VHE band and several optical-NearInfraRed (NIR) and radio ground based telescopes. The Nordic Optical Telescope and the Gran Telescopio Canarias gave a redshift estimation of $z = 0.4245 \pm 0.0005$.

The timeline of the MAGIC observation can be described as follows:

- the Swift-BAT alert was received by the MAGIC automatic alert system at $T_0 + 22$ s. The alert was validated, the event tagged as observable and the automatic fast repositioning procedure started;
- the telescopes reached the target and started tracking at $T_0 + 50$ s;
- the data acquisition started receiving the first events at $T_0 + 57$ s and operated stably from $T_0 + 62$ s;
- observation lasted until $T_0 + 15912$ s when a zenith angle of 81.14° was reached.

The observation was performed in good weather conditions and in presence of the Moon, resulting in a night sky background approximately 6 times higher than the standard dark night conditions.

The analysis of the MAGIC dataset was performed with unprecedented accuracy performing many checks and tests to investigate the behaviour of the subsystems during the repointing and data taking. Moreover, several dedicated MC sets were produced in order to match the GRB observational conditions in the most precise way. The result of the offline analysis shows a clear detection above the 50 sigma level in the first 20 minutes of observation.

The light curve for the intrinsic flux corrected for the Extragalactic Background Light (EBL) absorption in the 0.3–1 TeV range was derived. A comparison of this light curve with the simultaneous X-ray and HE emission is shown in Figure 1.

From the evolution of the light curve and the comparison with the lower energy band emission it is possible to derive clues on the origin and the amount of power radiated in the TeV component. The TeV light curve is well described with a simple power-law temporal decay with index $\beta = 1.60 \pm 0.07$. A similar behaviour can also be seen in the X-ray and HE band. These properties, together with the absence of breaks, cutoffs or irregular variability, support the conclusion that the observed VHE emission component belong to the afterglow phase. Nevertheless, the presence of a subdominant prompt component cannot be excluded.

The comparison of the simultaneous light curves shows also that the radiated power in the TeV band is mostly comparable, within a factor of about 2, with the X-ray and HE bands. This means that a non-negligible fraction of energy is radiated in the TeV band. A rough estimate, calculated assuming that the onset of the afterglow is at $T_0 + 6$ s[4] and following the time evolution estimated by the MAGIC light curve is of around $\sim 10\%$ of the isotropic-equivalent energy of the prompt emission E_{iso}.

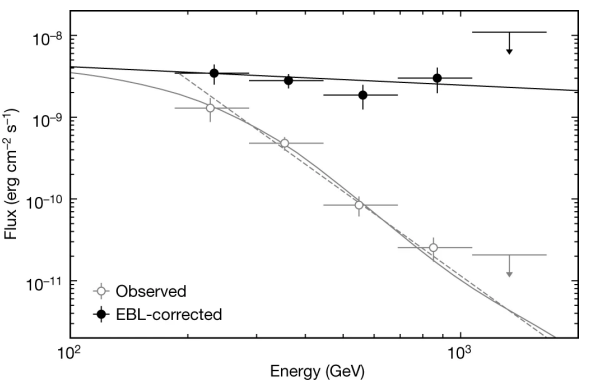


Fig. 2. GRB 190114C Observed (grey open circles) and intrinsic (black filled circles) spectral energy distribution in the time interval $T_0 + 62$ s and $T_0 + 2454$ s. The best fit power-law functions (black solid line and dashed grey line) and the EBL attenuated curve (solid grey line) are also shown.[2]

In Figure 1 also the temporal evolution of the intrinsic spectral photon index α_{int} of the TeV differential photon spectrum is shown. A constant value of $\alpha_{int} \approx -2$ is consistent with the data, considering the statistical and systematic errors. As a result, throughout the MAGIC observation, no spectral variability can be claimed. This value indicates that the radiated power is equally distributed in the 0.3-1 TeV energy range.

The observed and EBL-corrected spectral energy distributions in the 0.2-1 TeV energy range and ($T_0 + 62$ s - $T_0 + 2454$ s) time interval were also derived. Both the spectra are fitted with a simple power-law with indices $\alpha_{obs} = -5.34 \pm 0.22$ and $\alpha_{\rm int} = -2.22^{+0.23}_{-0.25}$ respectively for the observed and EBL-corrected spectrum.

This result states that the VHE component observed by MAGIC is extended up to TeV energies. Moreover, the steep observed spectrum obtained certificates the strong effect of the EBL absorption (a factor of ~ 300 at 1 TeV) in shaping the GRB spectra in the VHE domain.

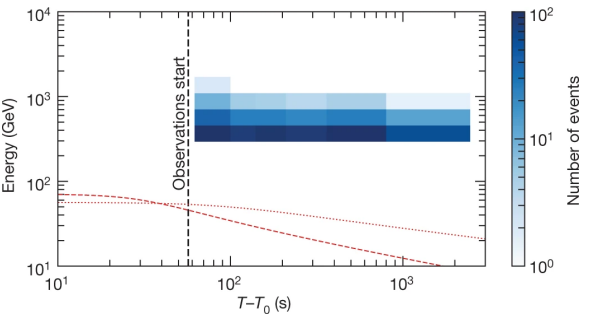


Fig. 3. Distribution of the number of GRB 190114C γ-ray events observed by MAGIC binned in time and energy and colour-coded. The synchrotron burnoff limiting curves are also displayed with two different assumptions for the external medium surrounding the GRB site: constant-density interstellar medium (dotted curve) and wind-like scenario (dashed curve).[2]

The similarities observed between the X-ray and the TeV light curves point towards the possibility that the same radiation mechanism is responsible for both components. Nevertheless, we proved that the TeV component cannot be interpreted as a simple extension in energy of the electron synchrotron emission mechanism. Indeed, in this scenario the energy of the accelerated electrons and hence of the synchrotron photons emitted is limited to a maximum value. This value can be estimated by equating the acceleration timescale and the synchrotron losses timescale, which dominates the electron energy losses. As a result, the maximum synchrotron photon energy, also called *synchrotron burnoff limit* is estimated as $E_{\mathrm{syn,max}} \sim 0.1\Gamma_{\mathrm{b}}(t)/(1+z)$ GeV where $\Gamma_{\mathrm{b}}(t)$ is the Lorentz bulk factor of the GRB jet and z is the redshift. In Figure 3 we compared the energy of the observed photons by MAGIC binned in time and energy with $E_{\mathrm{syn,max}}$ assuming two possible scenarios for the dynamical evolution of the GRB blast wave. It can be clearly seen that events detected by MAGIC are well more energetic than $E_{\mathrm{syn,max}}$ in both scenarios. This gives a strong proof that a different radiation mechanism than electron synchrotron radiation must be claimed to explain such VHE emission component in GRB 190114C.

3. Multi-wavelength afterglow modeling

A multi-wavelength interpretation is fundamental for an in-depth study of the properties of the TeV emission and its connection with the lower energy band radiation. GRB 190114C emission was detected across 17 orders of magnitude in energy and

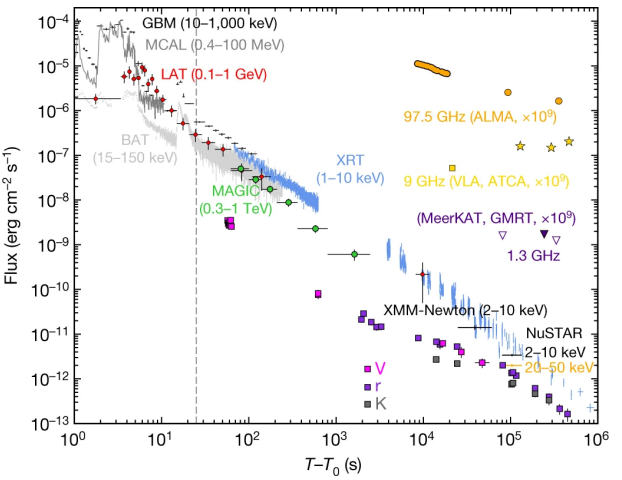


Fig. 4. Broadband light curves of GRB 190114C. Data from radio up to γ-rays are included. The vertical dashed line marks approximately the end of the prompt phase, identified as the end of the last flaring episode.[3]

by more than 20 instruments both from ground and in space (see Figure 4). In particular, the simultaneous data collected in a few time intervals in the X-ray (XRT, BAT, GBM), HE (LAT) and VHE (MAGIC) domain allow to build multi-wavelength time-binned spectra (see Figure 5). In the first time interval (68-110 s) the usual synchrotron peak flux is located in the X-ray band. Subsequently, in the HE band a decreasing flux behaviour is observed. Then, the VHE data implies the presence of a second peak in the flux and a spectral hardening for energies $E > 0.2$ TeV. This behaviour can be explained only if we assume that the TeV emission is not an extention of the X-ray synchrotron emission component but it is produced by a different radiation mechanism.

The most credible scenario able to describe consistently the multi-wavelength data is the *synchrotron and SSC external shock scenario*. In this picture, the SSC radiation mechanism produces the TeV photons while the synchrotron emission can explain the lower energy band radiation. The distribution of accelerated electrons emitting the synchrotron radiation are also responsible for the up-scattering of the same synchrotron photons through inverse Compton mechanism. As a result, a second spectral component peaking in the VHE band is generated. The afterglow emission is usually well described through external shocks. SSC radiation at TeV level can be also produced in internal shock synchrotron models for the prompt emission. However, for GRB 190114C it was estimate that it can only partially contribute ($\lesssim 20\%$) for the flux emitted at early times ($\lesssim 100$ s). As a result, the SSC radiation in the afterglow scenario is the most viable process.

The validity of this scenario was tested thanks to a numerical modeling. The full multi-wavelength dataset was modelled with a numerical code reproducing the

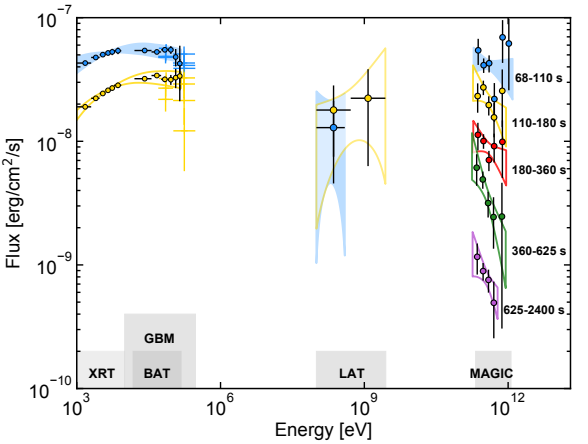


Fig. 5. Multi-band SEDs in five different time intervals. MAGIC data analysis results are corrected for the EBL attenuation. MAGIC and LAT contour regions are drawn from the 1σ error of their best-fit power-law functions. For Swift data, the regions show the 90% confidence contours for the joint fit for XRT and BAT, obtained by fitting a smoothly broken power law to the data.[3]

synchrotron plus SSC radiation emitted in the external forward shock afterglow scenario. In Figure 6 the modeling for two time intervals and for the X-ray, HE and VHE bands is shown. The numerical code solves the time-evolving electron distribution and calculates the corresponding photon emission covering the entire electromagnetic spectrum. The following radiation processes are considered: electron synchrotron emission and self-absorption, SSC emission, adiabatic losses, γ-γ absorption and emission from pairs. A set of inputs parameters containing some assumptions due to unknown properties of the acceleration process, of the shock microphysics, of the external medium and of the GRB jet and the initial conditions of the system are given in the code. The soft spectrum derived from the MAGIC dataset constrain the peak of the SSC component to be below 200 GeV. Such condition is obtained assuming that the SSC process is in Klein-Nishina regime and the γ-γ internal absorption has a non-negligible contribution in shaping the higher energy tail of the VHE spectrum.

Acceptable modeling of the multi-wavelength afterglow spectra have been found with an initial isotropic-equivalent kinetic energy of the blastwave $E_k \gtrsim 3 \times 10^{53}$ erg. The electron and magnetic field equipartition parameters are respectively $\varepsilon_e \approx 0.05-0.15$ and $\varepsilon_b \approx 0.05-1\times10^{-3}$, the density of the external medium is $n \approx 0.5-5$ cm^{-3} and the distribution of the accelerated electrons is described with a power-law of index $p \approx 2.4-2.6$. This set of values for the GRB afterglow parameters is similar to the one used for past GRB afterglow studies at lower frequencies. This is an indication that the SSC component can be a relatively common process for GRB afterglows.

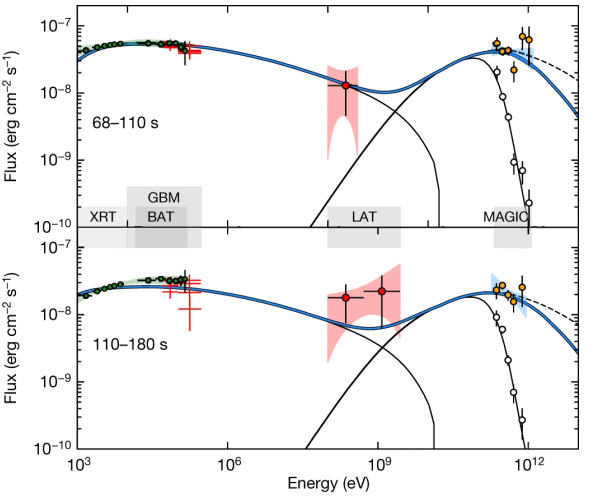


Fig. 6. Modeling of GRB 190114C spectral energy distributions in two different time intervals. Observed (thin solid line) and deabsorbed (thick blue line) spectrum are displayed. The dashed line is the SSC component neglecting the effects of internal γ-γ opacity.[3]

4. Conclusions

Observation and detection of GRB 190114C is a key milestone for the study and the interpretation of GRB physics. After decades of searches, the unambiguous discovery of a TeV component in GRBs made it possible to address for the first time ever several open questions. The newly open TeV spectral window also gives rise to new challenges for the observation and interpretation of GRBs.

A complete multiwavelength modeling of the GRB 190114C data covering the entire electromagnetic spectrum from radio up to TeV was presented. The full evolution of the afterglow emission was interpreted with the synchrotron plus SSC external forward shock scenario. The afterglow parameters used in the modeling of the X-ray, HE and VHE data have relatively common values, similar to those used in previous studies on GRBs where VHE emission was not detected. When evolving this solution to later times (see Figure 7) the optical and radio emission is overpredicted with respect to the observations. This inconsistency between the data and the modeling can be interpreted as an indication that some of the fixed parameters of the afterglow theory (e.g. the electron and magnetic field equipartition parameters) may evolve in time.

The intrinsic properties of GRB 190114C are not so exceptional when compared with the sample of the observed long GRBs. Despite being an energetic event (the isotropic-equivalent energy from the prompt is $\approx 3 \times 10^{53}$), it lies in the highest 30% distribution in energy for long GRBs.[5] This suggests that a larger sample of GRBs can have a detectable TeV component. On the other hand, favourable observational conditions and redshift play a crucial role in order to make TeV detections possible.

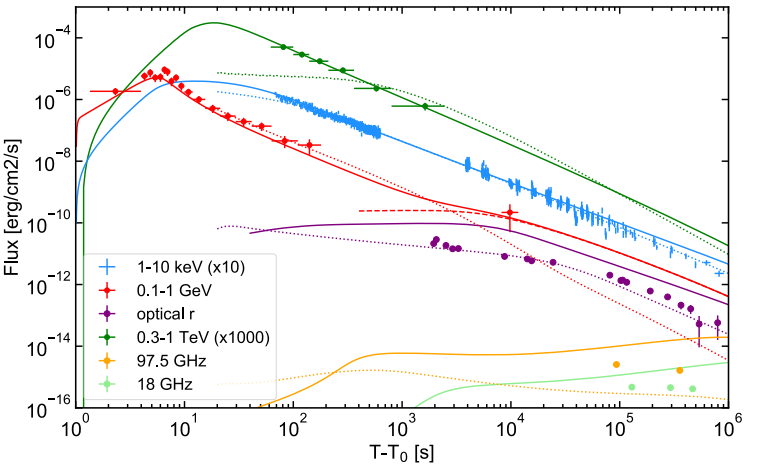


Fig. 7. Modeling of GRB 190114C afterglow light curves from radio up to TeV. Two different sets of parameters, respectively displayed in solid and dotted lines are compared to the observations. The former one is optimized for the X-ray, GeV and TeV data while the latter one well reproduces the late time optical modeling but do not explain the early time TeV emission. The dashed red line indicate the SSC contribution in the 0.1-1 GeV band.[3]

This was proved by the TeV detections of other three long GRBs: GRB 180720B,[6] GRB 190829A[7] and GRB 201216C.[8] These discoveries showed that TeV detections can be possible even at late times (hours on even a day after the trigger like in the case of GRB 180720B and GRB 190829A), or in low-luminosity events ($E_{iso} \approx 3 \times 10^{50}$ in GRB 190829A), or at higher redshift ($z = 1.1$ for GRB 201216C). The first population studies of GRBs at VHE will provide the possibility to investigate the conditions which favour a TeV detection.

The TeV photons detected from GRB 190114C were also used to give insights on some fundamental physics questions. An example is given by the study and the bounds derived on the Lorentz invariance violation effect.[9]

Furthermore, the ongoing observations of GRBs in the VHE domain will be fundamental to state the presence of a TeV emission component in short GRBs (a hint of detection, but still not conclusive, was claimed for GRB 160821B[10]) or during the prompt emission phase. Short GRBs are known to be linked with gravitational wave events. Therefore, the detection of a TeV emission component could give information on the structure of GRB jets as well as on the geometry of the merger event. The detection of TeV emission in the prompt emission phase shall provide essential information on the radiation mechanisms involved which are still unknown.

Further detections of GRBs in the VHE domain will be also useful to test the current most credible scenario for the TeV afterglow emission, namely the SSC external forward shock scenario. At the current stage, the results published for GRB 190114C,[3] GRB 180720B[6] and GRB 190829A[11] can be satisfactorily interpreted with the SSC scenario. Nevertheless, several open questions and assumptions

on the acceleration process, shock microphysics and environmental condition are still unclear.

In this evolving context the robust results obtained from the analysis and the interpretation of GRB 190114C will be a reference point for future studies of the properties of the VHE emission component in GRBs.

References

1. Zhang, B. (2019). Extreme emission seen from γ-ray bursts. Nature, 575(7783), 448-449.
2. MAGIC Collaboration, Nature, 575, 455-458 (2019)
3. MAGIC Collaboration et al., Nature, 575, 459-453 (2019)
4. Ravasio, M. E. et al., Astronomy & Astrophysics, 626, A12 (2019)
5. Nava, L. et al., Mon. Not. R. Astron. Soc. 421, 1256–1264 (2012).
6. H.E.S.S. Collaboration et al., Nature, 575, 464–467 (2019)
7. H.E.S.S. Collaboration et al., Science, 372, 1081-1085 (2021)
8. S. Fukami et al., ICRC2021 proceedings, (2021)
9. MAGIC Collaboration et al., Physical Review Letters, 125, 021301 (2020)
10. V. A. Acciari et al., Astrophysical Journal, 908, 90 (2021)
11. Salafia, O. et al, ArXiv e-prints, 2106.07169, (2021)

AGILE and GRBs: 13 years of observations

A. Ursi*, on behalf of the AGILE Team
INAF/IAPS, via del Fosso del Cavaliere 100, I-00133 Roma (RM), Italy
**E-mail: alessandro.ursi@inaf.it*

Over a time frame of 13 years, from November 2007 to November 2020, the AGILE MiniCALorimeter (MCAL; 0.4-100 MeV), detected 503 Gamma-Ray Bursts (GRBs). This sample is constituted by 44% short GRBs and 56% long GRBs, as retrieved from the study of the associated T_{50} and T_{90} burst duration distribution. For 258 GRBs, it was possible to perform a spectral analysis by adopting a single power law model; for 43 of them, also a spectral fit with a Band model with peak energy E_p above 400 keV was possible. More than 90% of these bursts were also detected by the AGILE Scientific RateMeters (RMs), providing comprehensive simultaneous observations from few tens keV to 100 MeV. The MCAL GRBs mostly consists of short-duration, spectrally hard events, due to the energy range of the detector and the adopted onboard trigger configurations, representing a burst sample which can be used to provide further insights on the high-energy component of hard-spectrum bursts.

Keywords: GRBs, high-energy, catalog, AGILE

1. Introduction

Discovered in the late 1960s,[1] Gamma-Ray Bursts (GRBs) still represent an intriguing and not completely understood phenomenon. These events consist of transients bursts of hard X- and gamma-rays, emitted by ultra-relativistic particles which have been accelerated in extragalactic central engines, releasing isotropic energies of about $E_{iso} \gtrsim 10^{51} - 10^{52}$ erg.[2] On the basis of their spectral hardness and time duration, GRBs are usually divided in short GRBs, lasting less than ~ 2 s and exhibiting a hard spectrum up to MeV energies, and long GRBs, lasting more than ~ 2 s and exhibiting a softer spectrum. The first ones are produced by the merger of binary neutron stars (BNS),[3–7] whereas the second ones are associated with Type Ic core-collapse supernovae.[8,9]

2. The AGILE satellite

The Astrorivelatore Gamma ad Immagini LEggero (AGILE) is an Italian space mission for high-energy astrophysics.[10] Its payload consists of a tungsten-silicon imaging tracker, sensitive in the 30 MeV - 50 GeV energy range, a non-imaging CsI(Tl) scintillation MiniCALorimeter (MCAL), sensitive in the 0.4-100 MeV energy range, a coded mask X-ray imager SuperAGILE (SA), sensitive in the 18-60 keV energy range, and an Anti-Coincidence (AC) system, sensitive in the 50-200 keV energy range. The suite of silicon tracker and MCAL detectors constitute the Gamma-Ray Imaging Detector (GRID).

Starting 2009, the satellite underwent a failure of the reaction wheel, losing the possibility to point sources in the sky and starting to spin around the sun-pointing axis, with an angular velocity of 1 rotation / 7 min. Spinning does not represent an issue for the MCAL detector, due to its ideal 4π sr FoV; moreover, it does not affect GRB detection, given the isotropic nature of these events. However, spinning could have an effect for what concerns spectral analysis of long GRBs, as such long-lasting events would be detected under different angles, throughout their whole duration, requiring the retrieval of several response matrices, and making the spectral fit more complex.

2.1. *MCAL and scientific ratemeters*

The AGILE Mini-CALorimeter (MCAL)[11, 12] is a non-imaging scintillation detector, sensitive between 400 keV and 100 MeV, composed of 30 CsI(Tl) scintillator bars. MCAL is a self-triggered detector, issuing a data acquisition with 2 μs time resolution whenever a count rate exceeding a given threshold is encountered. The trigger system acts on different trigger logic timescales: in particular, the so-called "hardware" logic works on short-duration timescales (0.293 ms, 1 ms, 16 ms) and adopts a static trigger with a fixed threshold count rate required to issue a data acquisition, whereas the so-called "software" logic works on longer-duration timescales (64 ms, 256 ms, 1024 ms, and 8192 ms) and adopts an adaptive trigger with a threshold count rate dependent on the background rate.

For what concerns the detection of GRBs, differently from the silicon tracker and SuperAGILE, MCAL does not require these events to lay within a given FoV, allowing a continuous monitoring of an ideal $\sim 4\pi$ sky. Moreover, MCAL focuses on the poorly investigated tens of MeV energy range, allowing to provide insights on the high-energy component of hard-spectrum bursts.

A first MCAL GRB Catalog with 84 events detected in the first two years of the AGILE mission, in the so-called "pointing mode" (2007-2009), was released in 2013[13] and a second MCAL GRB Catalog with 503 events, covering also the successive 11 years in the so-called "spinning mode" (2007-2020) was recently submitted to ApJ.[14] MCAL is part of the 3rd Inter-Planetary Network (IPN), providing GRB detections which can be used for triangulation and localization of the bursts. Starting 2017, a new real time pipeline system has been established, for the automatic offline analysis of MCAL data and the prompt delivery of GCN Notices and Circular to the scientific community, in case of a transient detection ([a]). In recent years, the onboard trigger configuration was modified to increase the MCAL detection efficiency to short-duration transients, in order to improve the detector sensitivity to possible electromagnetic counterparts of gravitational wave events, during the LIGO-Virgo observational runs O2 and O3, and in order to save onboard mass memory, to deal with severe telemetry restrictions undergone in recent years.

[a] https://gcn.gsfc.nasa.gov/agile_mcal.html

The MCAL, SA, and AC data are continuously recorded in telemetry as scientific RateMeter (RM) data, with a 0.5 s (for SA) and 1.0 s (for MCAL, and AC) time resolution, to provide a monitoring of the X- and gamma-ray background modulation throughout the satellite orbital phases. Nevertheless, the AGILE RMs allow to independently detect a large number of high-energy transients, as GRBs, Soft Gamma-ray Repeaters (SGRs), or solar flares. Despite the coarse time resolution, RM data might serve as back-up of data acquired by the onboard detectors, as well as to provide cross-check and validation of GRBs detected by MCAL.

3. Methods

A cross-search between $\sim 10^6$ AGILE MCAL triggers, issued onboard from November 2007 to November 2020, and $\sim$ 4500 IPN[15] bursts (IPN webpage:[b]), reported by the network in the same time interval, lead to the identification of 476 GRBs in the MCAL data. Due to the better efficiency of MCAL to detect short-duration transients, 109 long GRBs were only partially acquired onboard by the detector. Despite being incomplete, these events may serve as further detections which can be used to improve sky localizations via triangulation.

A parallel blind search in the MCAL data, using the offline burst search algorithm,[16] led to the detection of other 27 candidate short GRBs, not reported by other space missions. The total MCAL GRB sample is therefore constituted by 503 bursts, as reported in the second AGILE MCAL GRB Catalog,[14] recently submitted to ApJ and available on line on the ASI Space Science Data Center (SSDC) web portal at[c]. About 90% of these events have been also detected by one or more scientific RMs. A simultaneous detection of all scientific RMs would provide a broad band characterization of these events, from few tens keV to tens of MeV. In particular, the RMs detection is very suitable when dealing with partially acquired long GRBs, in order to reconstruct the whole light curve of the event and provide insights in different wavelengths.

3.1. *Time analysis*

For the 394 fully detected GRBs, it was possible to reconstruct the complete light curves of the events and to study their temporal properties. The T_{50} and T_{90} time distributions[17] of these bursts ended up with 173 short GRBs and 221 long GRBs, as shown in Fig. 1. The sample shows a large fraction (44%) of short GRBs, with respect to that retrieved by other space missions, as INTEGRAL,[18] Fermi GBM,[19] and Konus-Wind.[20] This can be ascribed to different reasons: first of all, the onboard trigger configuration optimized for the detection of short-duration transients, as already exposed in Section 2.1; the second reason is related to the relatively high-energy ($>$ 400 keV) range of MCAL, which makes the detector more sensitive

[b] http://www.ssl.berkeley.edu/ipn3/
[c] https://www.ssdc.asi.it/mcal2grbcat/

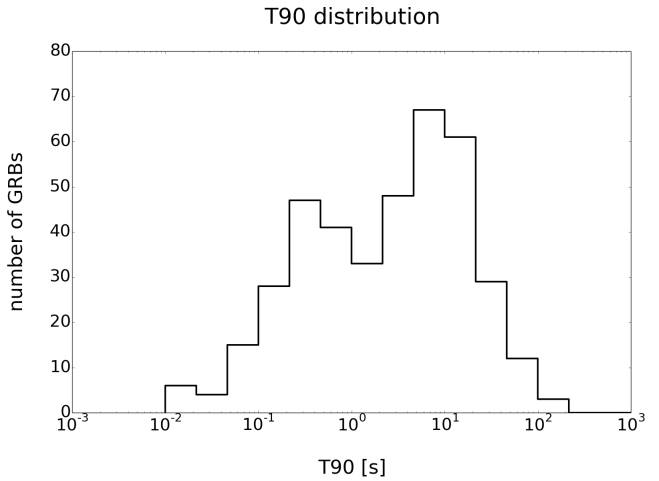


Fig. 1. T_{50} and T_{90} distributions of 394 GRBs fully detected onboard by MCAL. The typical bimodal distribution consists in 44% short GRBs and 56% long GRBs.

to hard-spectrum bursts, typically short GRBs. Moreover, the 0.4-100 MeV energy range allows to investigate only a limited spectral component of the detected bursts: a GRB will typically last longer if considering the whole spectrum including also the keV - tens of keV energy range. An example of this is shown in Fig. 2, where the same burst, GRB 200829A, is detected by the SA and the MCAL RMs, and the event clearly exhibit a longer time duration in the softer SA regime (15.3 ± 0.5 s), with respect to that observed in the harder MCAL regime (8.1 ± 0.5 s). The SA light curves have been rescaled in order to have similar background rates and to better highlight the time duration difference in the two energy ranges.

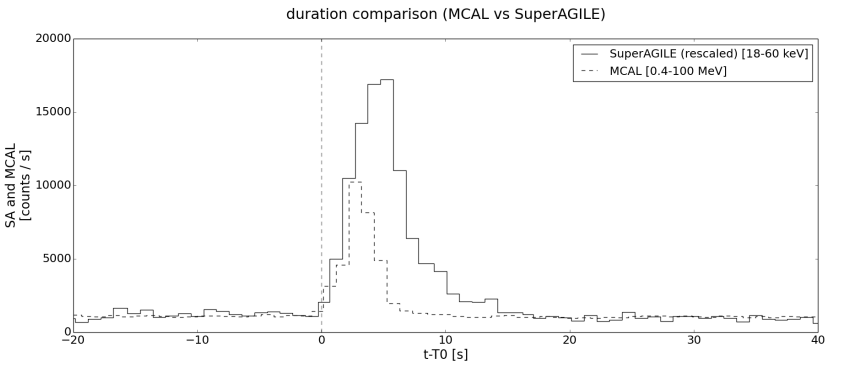


Fig. 2. GRB 200829A as detected by SuperAGILE, in the 18-60 keV energy range, and by MCAL, in the 0.4-100 MeV energy range. The event clearly lasts longer (15.3 ± 0.5 s) in the softer energy range, with respect to the MCAL high-energy range (8.1 ± 0.5 s). Data are rescaled in order to have similar background rates and highlight the time duration difference.

3.2. *Spectral analysis*

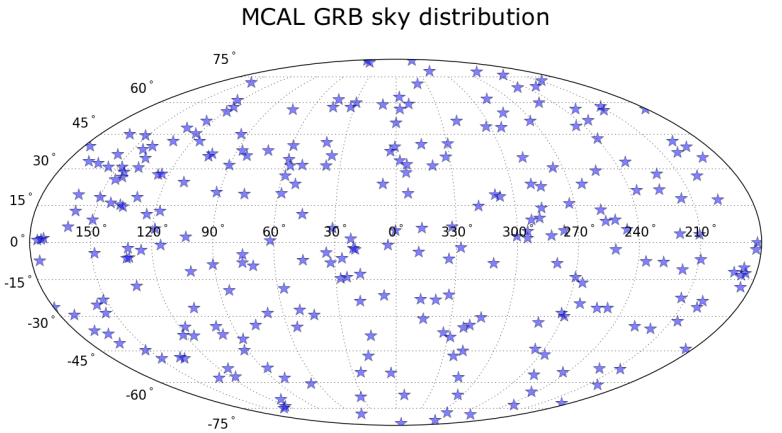


Fig. 3. Sky distribution of 276 MCAL GRBs, for which it was possible to retrieve an external localization from other space missions, or from IPN triangulations. Among them, 258 events could be fitted with a single power law model, and 43 events could be fitted also with a Band model.

In order to perform spectral analysis of the MCAL GRBs, it is necessary to retrieve the corresponding sky localizations, in order to reconstruct the appropriate response matrix of the instrument at the time of the detection. As MCAL is a non-imaging detector, localizations were retrieved from other space missions with imaging capabilities, or from IPN triangulations. In particular, it was possible to retrieve the sky localization of 276 GRBs of the MCAL sample, from detections by Fermi GBM, Fermi LAT, Swift BAT, Swift XRT, INTEGRAL ISGRI, or IPN triangulations. The spatial distibution of these events is shown in Fig. 3. Exploiting these localizations, on the basis of the available count statistics, it was possible to carry out a spectral fit for 258 GRBs. Spectral analysis was carried out using the XSpec package (version 12.9.0[21]), and adopting different fitting statistics depending on the count and background rates. Spectral fit was carried out in the fixed energy range from 400 keV to 50 MeV.

For 258 GRBs, a spectral fit with a single power law model was possible. The average photon index obtained from this sample is equal to $\langle\beta\rangle \sim -2.3$, highlighting the hardness of these bursts. The distribution of the power law photon index is shown in Fig. 4. The MCAL GRB fluxes, obtained from the spectal fitting with a power law model and estimated with a 90% confidence level, range between $3.9 \cdot 10^{-7}$ erg cm^{-2} s^{-1} and $1.0 \cdot 10^{-2}$ erg cm^{-2} s^{-1}, whereas the corresponding fluences, integrated over the related T_{90} time durations, range from $5.5 \cdot 10^{-8}$ erg cm^{-2} and $1.3 \cdot 10^{-2}$ erg cm^{-2}. The distribution of GRB fluxes is reported in Fig. 5. Moreover, for 43 particularly hard GRBs, it was possible to obtain a good spectral fit also by adopting a Band model[22] with peak energy $E_p > 400$ keV. In particular, these events

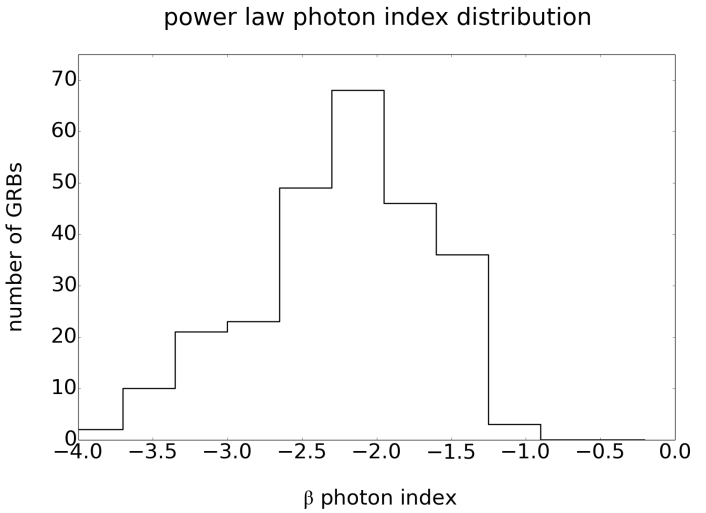


Fig. 4. Distribution of the photon index β of the single power law adopted to fit 258 MCAL GRBs. The average photon index results equal to $\langle\beta\rangle \sim -2.3$.

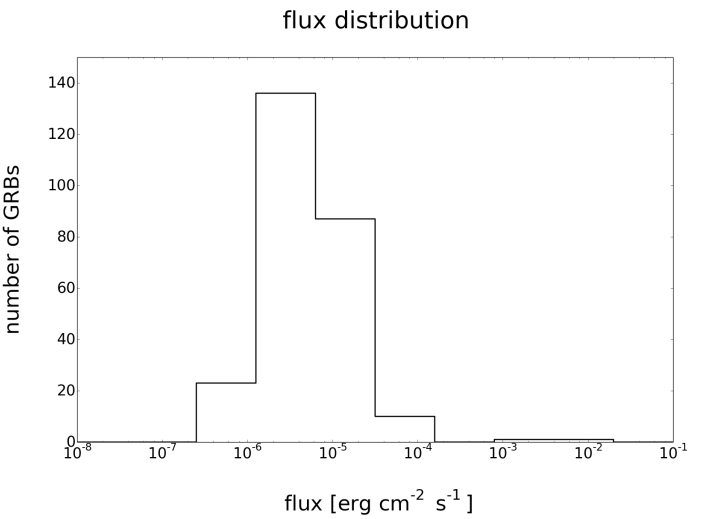


Fig. 5. Distribution of the GRB fluxes, in the 0.4-50 MeV energy range, obtained by fitting 258 burst with a single power law model.

exhibit a mean low-energy photon index $\langle\alpha\rangle \sim -0.6$, a mean high-energy photon index $\langle\beta\rangle \sim -2.5$, and a mean peak energy $\langle E_p\rangle \sim 640$ keV. The distribution of the high-energy photon index β resulting from the fit with a Band model is shown in Fig. 6.

It is interesting to notice that 92 GRBs can be fitted with power laws with photon indices $\beta > -2$, which correspond to hard-spectrum events with spectral

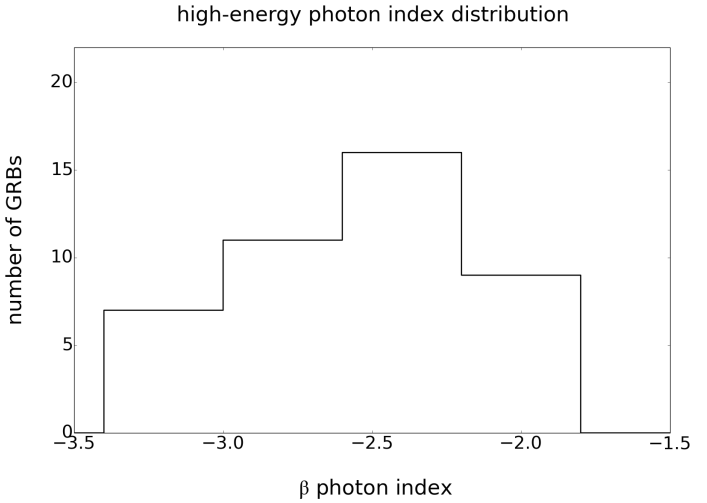


Fig. 6. Distribution of the high-energy photon index β of the Band model adopted to fit 43 MCAL GRBs. The average photon index results equal to $\langle\beta\rangle \sim -2.5$.

energy densities νF_ν with positive slopes, extending the emission up to the highest energies. These events and the 43 bursts fitted with Band models with $E_p > 400$ keV represent the hardest spectrum GRB subsample detected by the AGILE MCAL. Among them, 20 events have a simultaneous Fermi LAT detection.

3.3. *GRB population decline and Amati relation*

Adopting fluxes retrieved by spectral fits with power law model, an MCAL GRB $logN - logF$ brightness distribution can be reconstructed, as shown in Fig. 7. Such distribution provides information on the intrinsic spatial properties of GRBs and on the GRBs population decline at large distances. As MCAL is an all-sky monitor, the $logN - logF$ is not affected by on-axis area variations. Issues which can limit the monitoring of the sky, such as passages into the South Atlantic Anomaly, where the detector is switched off, or Earth occultations, which constantly hide about 35% of the available sky, can be neglected over a time frame of 13 years. The deviation at faint fluxes from a -3/2 slope, retrieved at large fluxes and typical of a Euclidean universe, makes the spatial distribution of GRBs unconsistent with a homogeneous universe.

For 8 events out of the 43 MCAL GRBs fitted with a Band model with $E_p >$ 400 keV it was possible to retrieve the corresponding redshift, provided by X-ray or optical observations of their afterglow. Using a standard cosmological model with $H_0 = 67.3$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.315$, and $\Omega_\Lambda = 0.685$, it was possible to evaluate the GRB related rest-frame parameters, such as the intrinsic peak energy $E_{p,i} = (1 + z)E_p$ of the time integrated spectrum and the isotropic equivalent

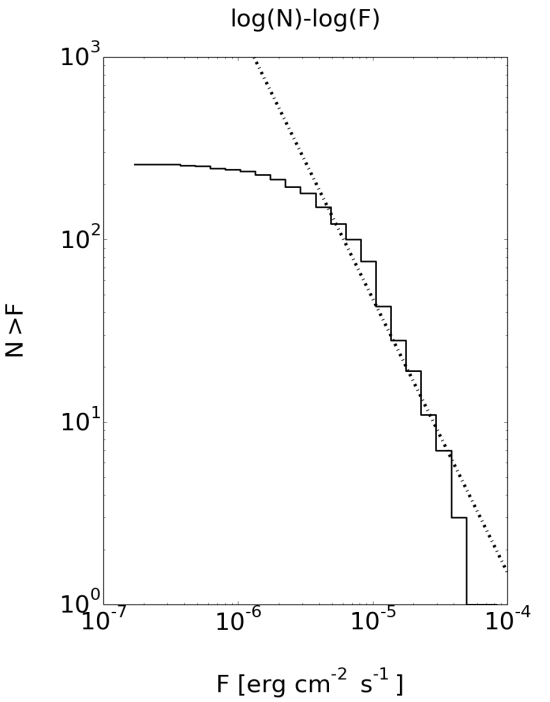


Fig. 7. GRB population decline obtained from the $logN - logF$ distribution of 258 MCAL bursts, fitted with a power law model. For faint fluxes, the brightness distribution departs from a $-3/2$ slope (dash-dotted line), typical of a Euclidean three-dimensional space, making the GRB spatial distribution not consistent with a homogenoeus universe.

energy E_{iso} in the 1-10000 keV energy range. The Amati relation[23,24] obtained from this sample results in a best-fit of the $E_{pi} = K(E_{iso}/10^{52})^m$ curve with a slope $m = 0.130 \pm 0.067$, as shown in Fig. 8. The coefficient obtained for MCAL GRBs is slightly more gentle with respect to the typical value of $m \sim 0.45$ reported by other space missions, due to the due to the non-uniform spectral hardness of the analyzed sample.

4. Conclusions

In more than 10 years activity, the AGILE MCAL detected 503 GRBs in the 0.4-100 MeV energy range. About 90% of these events were also simultaneously detected by the AGILE scientific RMs, covering a wide energy range, from 20 keV to 100 MeV. For 394 events, it was possible to reconstruct the T_{50} and T_{90} durations, ending up with 44% short GRBs and 56% long GRBs, resulting in a large fraction of short-duration bursts. 476 of these events are confirmed by simultaneous detections by other space missions, as reported by the IPN network, whereas 27 of these events are GRB candidates, so far only reported by MCAL. Using localizations provided

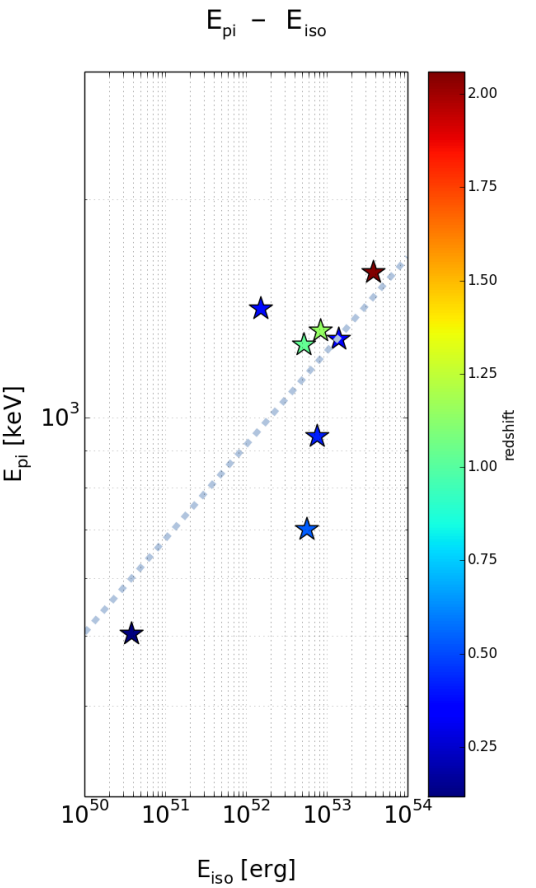


Fig. 8. Rest-frame $E_{iso} - E_{p,i}$ plane with 8 MCAL GRBs with known redshifts (colorbar), fitted with a Band model with $E_p > 400$ keV, in the 1-10000 keV energy range. The dashed line corresponds to the best-fit Amati relation, with slope $m = 0.130 \pm 0.067$.

by other missions, or by IPN triangulations, it was possible to carry out a spectral analysis of the time integrated spectrum for 258 GRBs. In particular, these events are well fitted by a single power law model, with a mean photon index $\langle\beta\rangle \sim -2.3$. For 43 events, also a Band model with peak energy $E_p > 400$ keV provided a good fit. The large fraction of bursts with durations $T_{90} < 2$ s and fittable either with power laws with photon index $\beta > -2$ or with a Band model indicates that MCAL GRBs mainly consists in short-duration, hard-spectrum bursts, which constitute an interesting and suitable sample that can be used to better investigate events with high-energy spectral components in the tens of MeV energy range.

References

1. R. W. Klebesadel, I. B. Strong and R. A. Olson, Observations of Gamma-Ray Bursts of Cosmic Origin, *Astrophys. J.l* **182**, p. L85 (June 1973).

2. N. Gehrels and P. Mészáros, Gamma-Ray Bursts, *Science* **337**, p. 932 (August 2012).
3. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2, *Physical Review Letters* **118**, p. 221101 (June 2017).
4. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence, *Astrophysical Journal Letters* **851**, p. L35 (December 2017).
5. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence, *Physical Review Letters* **119**, p. 141101 (October 2017).
6. V. Connaughton, A. Goldstein and Fermi GBM - LIGO Group, Multi-Messenger Time-Domain Astronomy with the Fermi Gamma-ray Burst Monitor, in *American Astronomical Society Meeting Abstracts #229*, American Astronomical Society Meeting Abstracts Vol. 229, January 2017.
7. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A, *Astrophysical Journal Letters* **848**, p. L14 (October 2017).
8. T. J. Galama, P. M. Vreeswijk, J. van Paradijs, C. Kouveliotou, T. Augusteijn, H. Böhnhardt, J. P. Brewer, V. Doublier, J.-F. Gonzalez, B. Leibundgut, C. Lidman, O. R. Hainaut, F. Patat, J. Heise, J. in't Zand, K. Hurley, P. J. Groot, R. G. Strom, P. A. Mazzali, K. Iwamoto, K. Nomoto, H. Umeda, T. Nakamura, T. R. Young, T. Suzuki, T. Shigeyama, T. Koshut, M. Kippen, C. Robinson, P. de Wildt, R. A. M. J. Wijers, N. Tanvir, J. Greiner, E. Pian, E. Palazzi, F. Frontera, N. Masetti, L. Nicastro, M. Feroci, E. Costa, L. Piro, B. A. Peterson, C. Tinney, B. Boyle, R. Cannon, R. Stathakis, E. Sadler, M. C. Begam and P. Ianna, An unusual supernova in the error box of the γ-ray burst of 25 April 1998, *Nature* **395**, 670 (October 1998).
9. K. Z. Stanek, T. Matheson, P. M. Garnavich, P. Martini, P. Berlind, N. Caldwell, P. Challis, W. R. Brown, R. Schild, K. Krisciunas, M. L. Calkins, J. C. Lee, N. Hathi, R. A. Jansen, R. Windhorst, L. Echevarria, D. J. Eisenstein, B. Pindor, E. W. Olszewski, P. Harding, S. T. Holland and D. Bersier, Spectroscopic Discovery of the Supernova 2003dh Associated with GRB 030329, *Astrophysical Journal Letters* **591**, L17 (July 2003).
10. M. Tavani, G. Barbiellini, A. Argan, F. Boffelli, A. Bulgarelli, P. Caraveo, P. W. Cattaneo, A. W. Chen, V. Cocco, E. Costa, F. D'Ammando, E. Del Monte, G. de Paris, G. Di Cocco, G. di Persio, I. Donnarumma, Y. Evangelista, M. Feroci, A. Ferrari, M. Fiorini, F. Fornari, F. Fuschino, T. Froysland, M. Frutti, M. Galli, F. Gianotti, A. Giuliani, C. Labanti, I. Lapshov, F. Lazzarotto, F. Liello, P. Lipari, F. Longo, E. Mattaini, M. Marisaldi, M. Mastropietro, A. Mauri, F. Mauri, S. Mereghetti, E. Morelli, A. Morselli, L. Pacciani, A. Pellizzoni, F. Perotti, G. Piano, P. Picozza, C. Pontoni, G. Porrovecchio, M. Prest, G. Pucella, M. Rapisarda, A. Rappoldi, E. Rossi, A. Rubini, P. Soffitta, A. Traci, M. Trifoglio, A. Trois, E. Vallazza, S. Vercellone, V. Vittorini, A. Zambra, D. Zanello, C. Pittori, B. Preger, P. Santolamazza, F. Verrecchia, P. Giommi, S. Colafrancesco, A. Antonelli, S. Cutini, D. Gasparrini, S. Stellato, G. Fanari, R. Primavera, F. Tamburelli, F. Viola, G. Guarrera, L. Salotti, F. D'Amico,

E. Marchetti, M. Crisconio, P. Sabatini, G. Annoni, S. Alia, A. Longoni, R. Sanquerin, M. Battilana, P. Concari, E. Dessimone, R. Grossi, A. Parise, F. Monzani, E. Artina, R. Pavesi, G. Marseguerra, L. Nicolini, L. Scandelli, L. Soli, V. Vettorello, E. Zardetto, A. Bonati, L. Maltecca, E. D'Alba, M. Patané, G. Babini, F. Onorati, L. Acquaroli, M. Angelucci, B. Morelli, C. Agostara, M. Cerone, A. Michetti, P. Tempesta, S. D'Eramo, F. Rocca, F. Giannini, G. Borghi, B. Garavelli, M. Conte, M. Balasini, I. Ferrario, M. Vanotti, E. Collavo and M. Giacomazzo, The AGILE Mission, *Astron. Astrophys.* **502**, 995 (August 2009).

11. C. Labanti *et al.*, The Mini-Calorimeter of the AGILE satellite, *proceedings of SPIE* **6266**, p. 62663Q (2006).
12. C. Labanti, M. Marisaldi, F. Fuschino, M. Galli, A. Argan, A. Bulgarelli, G. di Cocco, F. Gianotti, M. Tavani and M. Trifoglio, Design and construction of the Mini-Calorimeter of the AGILE satellite, *Nuclear Instruments and Methods in Physics Research A* **598**, 470 (January 2009).
13. M. Galli, M. Marisaldi, F. Fuschino, C. Labanti, A. Argan, G. Barbiellini, A. Bulgarelli, P. W. Cattaneo, S. Colafrancesco, E. Del Monte, M. Feroci, F. Gianotti, A. Giuliani, F. Longo, S. Mereghetti, A. Morselli, L. Pacciani, A. Pellizzoni, C. Pittori, M. Rapisarda, A. Rappoldi, M. Tavani, M. Trifoglio, A. Trois, S. Vercellone and F. Verrecchia, AGILE mini-calorimeter gamma-ray burst catalog, *Astron. Astrophys.* **553**, p. A33 (May 2013).
14. A. Ursi, M. Romani, F. Verrecchia, C. Pittori, M. Tavani, M. Marisaldi, M. Galli, C. Labanti, N. Parmiggiani, A. Bulgarelli, A. Addis, L. Baroncelli, M. Cardillo, C. Casentini, P. W. Cattaneo, A. Chen, A. Di Piano, F. Fuschino, F. Longo, F. Lucarelli, A. Morselli, G. Piano and S. Vercellone, The Second AGILE MCAL GRB Catalog: 13 Years of Observations, *Astrophysical Journal* , p. xx (2021).
15. K. Hurley, V. D. Pal'shin, R. L. Aptekar, S. V. Golenetskii, D. D. Frederiks, E. P. Mazets, D. S. Svinkin, M. S. Briggs, V. Connaughton, C. Meegan, J. Goldsten, W. Boynton, C. Fellows, K. Harshman, I. G. Mitrofanov, D. V. Golovin, A. S. Kozyrev, M. L. Litvak, A. B. Sanin, A. Rau, A. von Kienlin, X. Zhang, K. Yamaoka, Y. Fukazawa, Y. Hanabata, M. Ohno, T. Takahashi, M. Tashiro, Y. Terada, T. Murakami, K. Makishima, S. Barthelmy, T. Cline, N. Gehrels, J. Cummings, H. A. Krimm, D. M. Smith, E. D. Monte, M. Feroci and M. Marisaldi, The Interplanetary Network Supplement to the Fermi GBM Catalog of Cosmic Gamma-ray Bursts, *The Astrophysical Journal Supplement Series* **207**, p. 39 (aug 2013).
16. A. Ursi, M. Tavani, F. Verrecchia, M. Marisaldi, A. Argan, A. Trois and P. Tempesta, A New AGILE MCAL Configuration to Detect Gamma-Ray Bursts and Sub-threshold Events in the Multimessenger Era, *Astrophysical Journal* **871**, p. 27 (January 2019).
17. T. M. Koshut, W. S. Paciesas, C. Kouveliotou, J. van Paradijs, G. N. Pendleton, G. J. Fishman and C. A. Meegan, Systematic Effects on Duration Measurements of Gamma-Ray Bursts, *Astrophysical Journal* **463**, p. 570 (June 1996).
18. v. Bošnjak, D. Götz, L. Bouchet, S. Schanne and B. Cordier, The spectral catalogue of INTEGRAL gamma-ray bursts: results of the joint IBIS/SPI spectral analysis, *Astron. Astrophys.* **561**, p. A25 (2014).
19. P. N. Bhat, C. A. Meegan, A. von Kienlin, W. S. Paciesas, M. S. Briggs, J. M. Burgess, E. Burns, V. Chaplin, W. H. Cleveland, A. C. Collazzi, V. Connaughton, A. M. Diekmann, G. Fitzpatrick, M. H. Gibby, M. M. Giles, A. M. Goldstein, J. Greiner, P. A. Jenke, R. M. Kippen, C. Kouveliotou, B. Mailyan, S. McBreen, V. Pelassa, R. D. Preece, O. J. Roberts, L. S. Sparke, M. Stanbro, P. Veres, C. A. Wilson-Hodge,

S. Xiong, G. Younes, H.-F. Yu and B. Zhang, The Third Fermi GBM Gamma-ray Burst Catalog: The First Six Years, *The Astrophysical Journal Supplement Series* **223**, p. 28 (apr 2016).

20. D. S. Svinkin, D. D. Frederiks, R. L. Aptekar, S. V. Golenetskii, V. D. Pal'shin, P. P. Oleynik, A. E. Tsvetkova, M. V. Ulanov, T. L. Cline and K. Hurley, The Second Konus-Wind Catalog of Short Gamma-ray Bursts, *The Astrophysical Journal Supplement Series* **224**, p. 10 (may 2016).
21. K. A. Arnaud, XSPEC: The First Ten Years, in *Astronomical Data Analysis Software and Systems V*, eds. G. H. Jacoby and J. Barnes, Astronomical Society of the Pacific Conference Series, Vol. 101, January 1996.
22. D. Band, J. Matteson, L. Ford, B. Schaefer, D. Palmer, B. Teegarden, T. Cline, M. Briggs, W. Paciesas, G. Pendleton, G. Fishman, C. Kouveliotou, C. Meegan, R. Wilson and P. Lestrade, BATSE observations of gamma-ray burst spectra. I - Spectral diversity, *Astrophys. J.* **413**, 281 (aug 1993).
23. L. Amati *et al.*, Intrinsic spectra and energetics of BeppoSAX gamma-ray bursts with known redshifts, *Astron. Astrophys.* **390**, p. 81 (2002).
24. W. J. Azzam and M. J. Alothman, Redshift Independence of the Amati and Yonetoku Relations for Gamma-Ray Bursts, *International Journal of Astronomy and Astrophysics* **3**, 372 (January 2013).

Searching for Gamma-Ray Bursts with the High-Altitude Water Cherenkov (HAWC) observatory

K.L. Engel* for the HAWC Collaboration

Department of Physics, University of Maryland, College Park; 4150 Campus Dr, College Park, MD, USA 20742

**E-mail: klengel@umd.edu*

Satellites and imaging atmospheric Cherenkov telescopes (IACTs) have shown that gamma-ray bursts (GRBs) are capable of producing very-high-energy photons— most notably GRB 190114C, observed up to 1 TeV by the MAGIC telescopes approximately one minute after triggering the Fermi GBM and Swift BAT satellites. Particularly suited to such searches and follow-up studies is the High-Altitude Water Cherenkov (HAWC) Observatory, which monitors 1/6th of the sky at any one time, complementing the pointed observations of TeV telescopes. It covers 2/3rds of the sky every day, with near continuous uptime. The HAWC GRB program comprises two dedicated analyses: a self-triggered all-sky search and a rapid response follow-up of GRBs reported by satellites. Both methods are performed in real time at the HAWC site and additionally repeated on archival data with improved calibration and reconstruction algorithms. Recent upgrades have HAWC poised for detection of the highest-energy gamma rays associated with GRBs, which are key to developing GRB emission models as well as constraining possible beyond-the-Standard-Model physics.

Keywords: Instrumentation: detectors; gamma rays: general; gamma ray burst: general

1. Introduction

Gamma-ray bursts (GRBs) remain one of the most enigmatic astrophysical sources, with the potential to shed light on topics at the forefront of the field, including physics beyond the Standard Model. Measurements of the highest-energy gamma rays associated with GRBs are key to developing models of the relativistic jets powering emission.[1] This is because observations of a spectral cutoff at the highest photon energies can be interpreted as estimation of the bulk Lorentz factor, Γ, in the region where gamma rays are produced,[2] providing insight into the internal GRB environment as well as the expected neutrino flux from GRBs, which is sensitive to Γ (see Ref. 3). In addition, the GRB spectrum above 10 GeV, such as that searched for with the High-Altitude Water Cherenkov (HAWC) Observatory's GRB program, can be used to constrain possible Beyond-the-Standard-Model physics that contains Lorentz invariance violations.[4] Alternatively, interpreting a spectral cutoff as attenuation of GRB photons from pair-production on extragalactic background light (EBL) provides constraints on EBL density over cosmological distances.[5]

There are at least two known classes of GRBs: short GRBs have durations of less than two seconds and a peak νF_ν at Earth of ~ 1 MeV,[4] whereas long GRBs

last more than two seconds and the peak of νF_ν at Earth is $\sim$ 150–200 keV.[6] Short GRBs, with their harder spectrum, are believed to be the result of the merger of two compact objects, such as two neutron stars or a neutron star and a solar mass black hole.[7–9] Core-collapse supernova type Ic are believed to be the progenitors of long GRBs, supported by observed coincidences of the two phenomena.[10] However, because the occurrences of either class of GRB cannot be predicted, ground-based observations of such sources are limited by slewing times— especially detrimental for short GRBs.

2. The Highest-Energy Emission from GRBs

For both short and long GRBs, emission is believed to happen in highly relativistic narrow jets that point in the direction of Earth. Prompt GRB emission (that is, emission from the first seconds of the GRB) is detected in the keV–MeV band by space-borne instruments such as the *Fermi*-GBM[11] and Swift-BAT.[12] Satellites with instruments sensitive to hard gamma rays, such as the *Fermi*-LAT and imaging atmospheric Cherenkov telescopes (IACTs) such as the MAGIC telescopes and the HESS array, have shown that GRBs are capable of producing very-high-energy (VHE) photons. Prominent examples include GRB 130427A,[13] observed up to 126 GeV by the *Fermi*-LAT; GRB 180720B,[14] detected using the H.E.S.S. Array at 100 GeV $< E_\gamma <$ 1 TeV; and GRB 190114C,[15,16] observed up to 1 TeV by the MAGIC telescopes approximately one minute after triggering the *Fermi*-GBM and Swift-BAT. However, due to the limited field-of-view of IACTs, it is difficult for observations above the energy threshold for the *Fermi*-LAT to occur in the first few seconds of the GRB. By contrast, the HAWC Observatory observes 1/6th of the sky at once and has near-continuous uptime to search for GRBs, making both short GRBs and long GRBs ideal observational topics. This uninterrupted analysis of air showers would be ideal for capturing a high-energy component of the prompt emission, should one be present.

3. The High-Altitude Water Cherenkov (HAWC) Observatory

The HAWC Observatory, a successor to the Milagro Observatory,[17] is a very-high-energy ground-based air shower array located on the side of the Sierra Negra volcano in Mexico at an altitude of 4,100 m above sea level. It has a wide field-of-view of $\sim$ 2 sr at any one time (observing 2/3rds of the sky each day) and covers an energy range from $\sim$300 GeV to above 100 TeV. HAWC consists of 300 cylindrical water tanks in the main array covering a total area of 22,000 m^2. Each tank in the main array is 7.3 m in diameter and 4.5 m deep and is equipped with four (three 8″- and one 10″-diameter) upward-facing photomultiplier tubes (PMTs) anchored to the bottom of the tank.

HAWC has completed installing and integrating an additional "outrigger" array[18] composed of 345 cylindrical tanks 1.55 m in diameter and 1.65 m deep, each

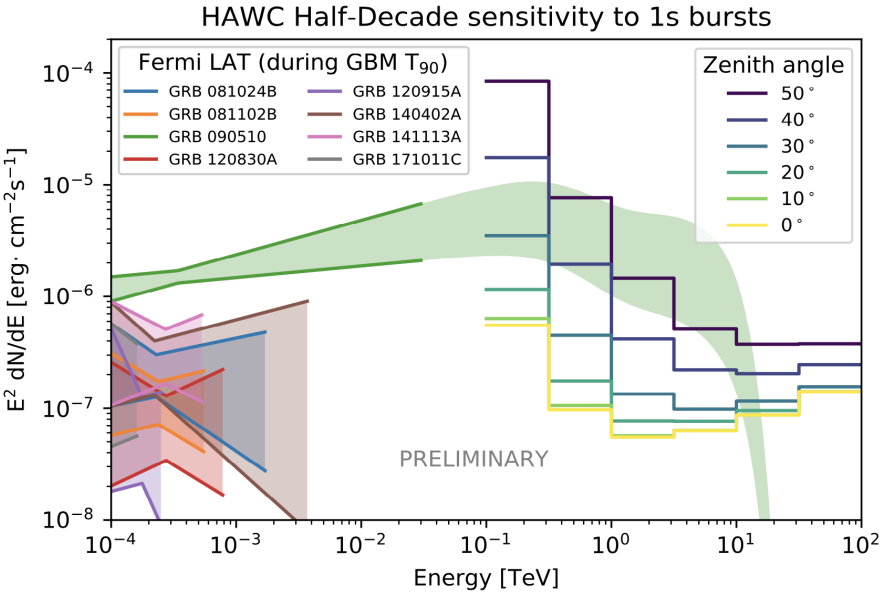
containing a single 8″ PMT. The outriggers are arranged in a concentric, circular, symmetric pattern around the main array, covering an additional instrumented area of ~ 4.5 times that of the main array. The outriggers provide enhanced sensitivity above 10 TeV by allowing the accurate determination of core position for showers off the main tank array and increase the effective area above 10 TeV by a factor of 3–4.

In both the main array and the outriggers, the PMTs detect Cherenkov light from secondary particles created in extensive air showers induced by very-high-energy gamma rays and cosmic rays incident on Earth's atmosphere. The main data acquisition system measures the arrival time of secondary particles at ground level and the amplitude of the PMT signals. This information, processed in near real time (only ~ 4 seconds delay), is used to reconstruct the arrival direction and energy of the primary particle all day, every day. The angular resolution of HAWC ranges from $\sim 0.2^\circ$ (at its highest energies) to 1.0° (at its lowest energies).[19] With these features and a duty cycle greater than 95%, HAWC is ideally suited to continuously monitor the Northern Hemisphere sky for high-energy emission from gamma-ray transients such as GRBs to find new sources as well as search archival data.

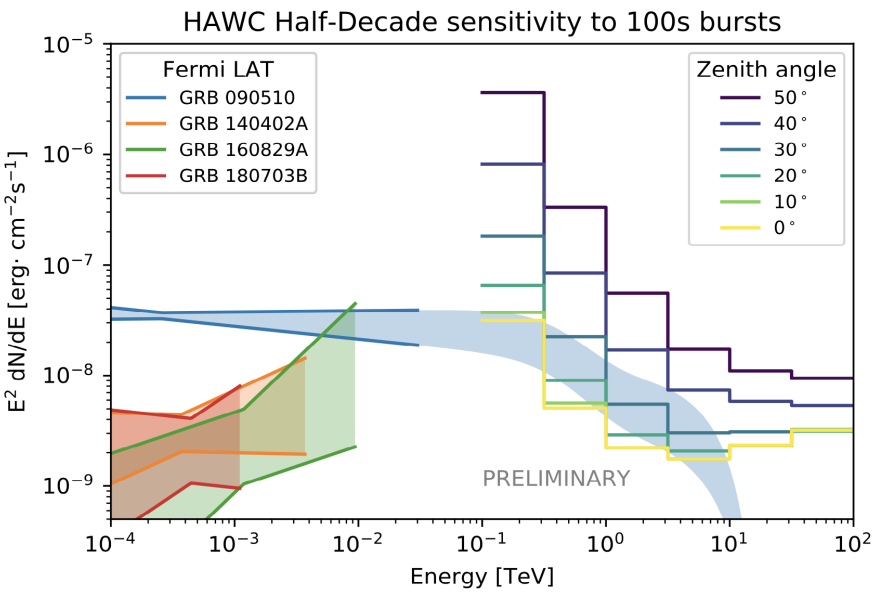
3.1. *The Gamma-Ray Detector Landscape*

HAWC's wide field-of-view and continuous operations are advantageous for detecting burst transients with emission durations shorter than the slewing times of Imaging Atmospheric Cherenkov Telescopes (IACTs), such as the prompt emission from GRBs. IACTs are also additionally limited by the need for small localization errors on triggered sources and their requirements for good weather and nighttime-only observations. While satellites are obviously very effective tools for the observation of both short and long GRBs, their limited size means satellites are not expected to provide useful statistics above tens of TeV— HAWC's effective area is > 100 times the size of *Fermi* at energies > 100 GeV. Additionally, HAWC's ability to observe GRBs before, during, and after the prompt emission phase measured by satellites is essential for determining the start time and duration of prompt emission at ~ 300 GeV— relatively unknown given the paucity of photons detected by satellites at these energies.

While HAWC is more sensitive to long GRBs, it occupies a unique position in terms of observing short GRBs due to the lack of slewing time and high energy range. This detector is therefore poised as a double-threat ad both classes of GRBs should be prime observational and analytical targets for us. The half-decade sensitivity calculations project a higher quasi-differential sensitivity than the *Fermi*-LAT[20] to both short and long bursts in the energy regime of interest herein, as shown in Figure 1.



(a) HAWC quasi-differential sensitivity to 1s bursts



(b) HAWC quasi-differential sensitivity to 100s bursts

Fig. 1. HAWC quasi-differential sensitivity to both short (a) and long (b) bursts as a function of zenith angle, defined as the mean flux in a given half-decade that would result in at least a 5σ detection half of the time. For reference we include spectra measured by the Fermi-LAT detector for short GRBs (on the order of 1s) and long GRBs (duration > 70 s), respectively. Figures by Israel Martinez-Castellanos.

4. The HAWC GRB Program

The HAWC GRB program consists of two dedicated analyses: a self-triggered, all-sky search, and rapid response follow-ups of GRBs reported by satellites. Both methods are performed in real time using the quick-look air shower reconstruction produced at the HAWC site with a latency of ∼five seconds. Additionally, each search is repeated on archival data with improved calibration and reconstruction algorithms.[1] This archival data is especially important for following up on reported emission from IACTs.

From the sensitivity of the HAWC Observatory, ~ 0.5 GRB detections per year are expected from following satellite-reported GRBs alone, assuming most GRBs do not contain an intrinsic cutoff below 200 Gev.[4] The all-sky search method yields similar expectation as the trials penalty taken when searching the full sky is roughly compensated by the ability to search data without satellite coverage.[21] Thus far, there have been no significant detections of GRB transients since the inauguration of the full HAWC detector on March 20, 2015.

4.1. *Self-Triggered, All-Sky Search*

The self-triggered all-sky method continuously searches for transient signals from GRBs above an energy threshold of 300 GeV using sliding time windows. The four time windows used are of lengths 0.2, 1, 10, and 100 seconds, typical of peak structures within GRB light curves. The all-sky search shifts each window forward in time by 10% of its width, binning air shower events during that window using a grid of $2.1^\circ \times 2.1^\circ$ square spatial bins covering all points within 50° of detector zenith,[21] as shown in Figure 2. The number of showers in each spatial bin is then compared to the expectation from well-modeled, charged cosmic-ray backgrounds, producing a p-value. Locations with an air-shower excess corresponding to a post-trials false alarm rate of one event per day are considered candidates for GRB transients and reported internally within HAWC.

This style of analysis is not limited to having to continuously search the entirety of the sky for a transient signal, but also possesses the ability to be passed an external trigger in order to perform a limited sliding time window search near the trigger time and location of known GRBs beyond just that of the satellite follow-ups described below. The analysis chain is the same whether continuous or externally triggered. This allows for the fine-tuning of the start time of very-high-energy emission when performing follow-ups. Though this analysis is performed in real-time, especially for external triggers, performing this analysis on archival data allows inspection *before* the trigger time, an ability unique to wide-field-of-view survey instruments such as HAWC, which are not limited to only looking forward in the way that IACTs are.

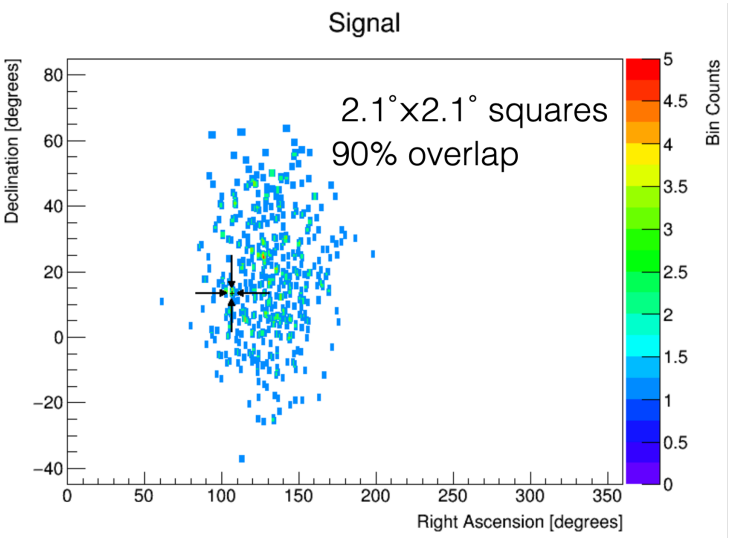


Fig. 2. The 90% overlap in spatial bins created by shifting each window by 10% of its width guarantees that the signal will be ideally oversampled such that there is a bin in which the signal is centered. Figure by Joshua Wood.

4.2. *Rapid Response Follow-Up*

The method for rapid response follow-up of reported GRBs is simpler in that it fixes the search window start time to match the external trigger time provided by a satellite (*e.g.*, the *Fermi*-GBM[11]); follow-ups of GCN Notices for triggered searches occur with $\sim$ 30 minutes latency. Additionally, the spatial portion of the search is restricted to the reported error on the GRB location. The analytical framework is the same as with the all-sky search, with air-shower events during the search window binned into spatial bins and compared against the expectation from well-modeled, charged cosmic-ray backgrounds.[22] Events can be separated based on quality and background, among other improvements. In cases when T_{90} is available with the external trigger, we search for emission occurring within T_{90} as well as $3 \times T_{90}$ and $10 \times T_{90}$ for long GRBs, and 6 and 20 seconds for short GRBs— the longer windows covering possible extended emission. Otherwise, the follow-up is performed with timescales of 1 and 20 seconds to cover typical T_{90} values as well as a 300-second window to look for extended emission. These timescales grant the ability to test for delayed-onset very-high-energy emission when performing follow-ups of these satellite-detected GRBs. The combined power of observations (or non-observations) from both satellites and ground-based observatories such as HAWC is critical for unlocking the secrets of GRBs, as shown through the agreement of the HAWC limits for both short and long GRBs with GBM observations in Figure 3.

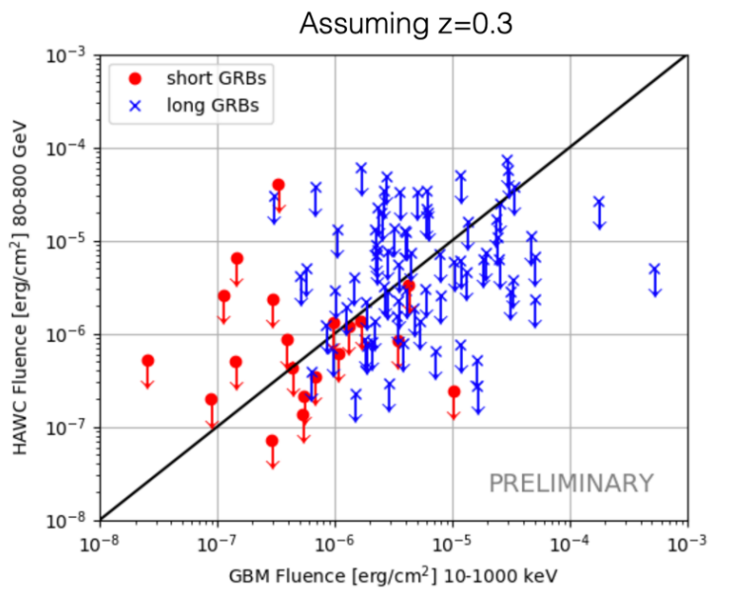


Fig. 3. Even for a non-detection, upper limits for nearby GRBs with known redshift, in combination with information from the triggering satellite(s), can place strong constraints on models.

4.3. *HAWC and the GRB Community*

Despite how ideal the HAWC Observatory is for the observation of very-high-energy emission from GRBs, a significant observation has not yet occurred. Those GRBs that have been detected to-date with a very-high-energy component—GRB 190114C,[15,16] GRB 180720B,[14] and GRB 190829A[23]—were not in the HAWC field of view. However, as shown in Figure 4, HAWC would likely have detected GRB 190114C if it had happened in its field of view.

If HAWC were to trigger on a GRB, structures are in place to alert the community and coordinate alerts between different experiments. The most prompt notification would be the use of GCNs and ATels for confirmed significant observations. HAWC has sent out 47 such circulars in the last year. Additionally, though somewhat delayed, HAWC is already using the Astrophysical Multimessenger Observatory Network (AMON)[24] for sub-threshold triggers. AMON reports "hotspot" parameters for events above the estimated cosmic-ray background level, $> 2.75\sigma$. AMON sends alerts to the GCN if an even passes the false-alarm-rate threshold of one per year, calculated using two years of scrambled datasets from both HAWC and IceCube.

5. Future Outlook

While HAWC has yet to observe a significant signal from a GRB, either self-triggered or as a follow-up, the HAWC Collaboration is very optimistic that recent

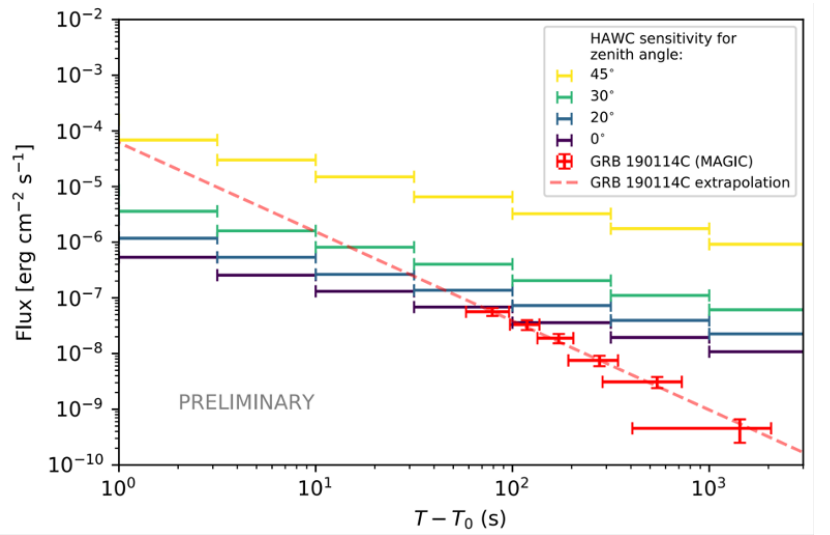


Fig. 4. HAWC's sensitivity at different zenith angles overlaid on an extrapolation of the flux detected by MAGIC from GRB 190114C. Adapted from Ref. 15.

improvements will greatly improve sensitivity to GRBs. This includes better reconstruction for low-energy events (noise reduction), increased statistics from multi-shower fits, and improved angular resolution for multiple showers on the array. Especially critical to the detection of very-high-energy emission from GRBs will be multi-bin, maximum likelihood analyses for short timescales and data reconstructed around known GRBs. The full dataset with these improvements will be ready in the coming months, with planned analyses to utilize these upgrades including both a targeted revisiting of known GRBs in reconstructed archival data, as well as updates to the self-triggered, "blind" search. With such advances at HAWC's fingertips, the HAWC Collaboration is eager for HAWC to broaden the GRB horizons and unlock their secrets through its unique position as a ground-based, wide-field-of-view observatory.

References

1. J. Wood, Results from the first one and a half years of the HAWC GRB program, *PoS* **ICRC2017**, p. 619 (2018).
2. T. Piran, Gamma-ray bursts and the fireball model, *Physics Reports* **314**, p. 575–667 (Jun 1999).
3. E. Waxman and J. Bahcall, High energy neutrinos from cosmological gamma-ray burst fireballs, *Physical Review Letters* **78**, p. 2292–2295 (Mar 1997).
4. I. Taboada and R. C. Gilmore, Prospects for the detection of grbs with hawc, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **742**, p. 276–277 (Apr 2014).

5. R. C. Gilmore, P. Madau, J. R. Primack, R. S. Somerville and F. Haardt, Gev gamma-ray attenuation and the high-redshift uv background, *Monthly Notices of the Royal Astronomical Society* **399**, p. 1694–1708 (Nov 2009).
6. S. Poolakkil *et al.*, The Fermi-GBM Gamma-Ray Burst Spectral Catalog: 10 yr of Data, *Astrophys. J.* **913**, p. 60 (2021).
7. M. Ruffert and H. Janka, Gamma-ray bursts from accreting black holes in neutron star mergers, *Astron. Astrophys.* **344**, 573 (1999).
8. S. Rosswog, E. Ramirez-Ruiz and M. B. Davies, High-resolution calculations of merging neutron stars - iii. gamma-ray bursts, *Monthly Notices of the Royal Astronomical Society* **345**, p. 1077–1090 (Nov 2003).
9. G. P. Lamb and S. Kobayashi, GRB 170817A as a jet counterpart to gravitational wave trigger GW 170817, *Mon. Not. Roy. Astron. Soc.* **478**, 733 (2018).
10. K. Z. Stanek, T. Matheson, P. M. Garnavich, P. Martini, P. Berlind, N. Caldwell, P. Challis, W. R. Brown, R. Schild, K. Krisciunas and et al., Spectroscopic discovery of the supernova 2003dh associated with grb 030329, *The Astrophysical Journal* **591**, p. L17–L20 (Jun 2003).
11. C. Meegan, G. Lichti, P. N. Bhat, E. Bissaldi, M. S. Briggs, V. Connaughton, R. Diehl, G. Fishman, J. Greiner, A. S. Hoover and et al., The fermi gamma-ray burst monitor, *The Astrophysical Journal* **702**, p. 791–804 (Aug 2009).
12. N. Gehrels, G. Chincarini, P. Giommi, K. O. Mason, J. A. Nousek, A. A. Wells, N. E. White, S. D. Barthelmy, D. N. Burrows, L. R. Cominsky, K. C. Hurley, F. E. Marshall, P. Meszaros, P. W. A. Roming, L. Angelini, L. M. Barbier, T. Belloni, P. T. Boyd, S. Campana, P. A. Caraveo, M. M. Chester, O. Citterio, T. L. Cline, M. S. Cropper, J. R. Cummings, A. J. Dean, E. D. Feigelson, E. E. Fenimore, D. A. Frail, A. S. Fruchter, G. P. Garmire, K. Gendreau, G. Ghisellini, J. Greiner, J. E. Hill, S. D. Hunsberger, H. A. Krimm, S. R. Kulkarni, P. Kumar, F. Lebrun, N. M. Lloyd-Ronning, C. B. Markwardt, B. J. Mattson, R. F. Mushotzky, J. P. Norris, B. Paczynski, D. M. Palmer, H.-S. Park, A. M. Parsons, J. Paul, M. J. Rees, C. S. Reynolds, J. E. Rhoads, T. P. Sasseen, B. E. Schaefer, A. T. Short, A. P. Smale, I. A. Smith, L. Stella, M. Still, G. Tagliaferri, T. Takahashi, M. Tashiro, L. K. Townsley, J. Tueller, M. J. L. Turner, M. Vietri, W. Voges, M. J. Ward, R. Willingale, F. M. Zerbi and W. W. Zhang, Erratum: "The Swift Gamma-ray burst mission" (ApJ, 611, 1005 [2004]), *The Astrophysical Journal* **621**, 558 (mar 2005).
13. P.-H. T. Tam, Q.-W. Tang, S.-J. Hou, R.-Y. Liu and X.-Y. Wang, Discovery of an extra hard spectral component in the high-energy afterglow emission of grb 130427a, *The Astrophysical Journal* **771**, p. L13 (Jun 2013).
14. H. Abdalla, R. Adam, F. Aharonian, F. Benkhali, E. Angüner, M. Arakawa, C. Arcaro, C. Armand, H. Ashkar, M. Backes, V. Barbosa Martins, M. Barnard, Y. Becherini, D. Berge, K. Bernlöhr, E. Bissaldi, R. Blackwell, M. Böttcher and C. Boisson, A very-high-energy component deep in the γ-ray burst afterglow, *Nature* **575**, 464 (11 2019).
15. Teraelectronvolt emission from the γ-ray burst grb 190114c, *Nature* **575**, p. 455–458 (Nov 2019).
16. P. Veres, P. N. Bhat, M. S. Briggs, W. H. Cleveland, R. Hamburg, C. M. Hui, B. Mailyan, R. D. Preece, O. J. Roberts and et al., Observation of inverse compton emission from a long γ-ray burst, *Nature* **575**, p. 459–463 (Nov 2019).
17. The Milagro Collaborations: R. Atkins et al., Milagrito, a TeV air-shower array, *NIM-A* **449**, 478 (2000).

18. V. Joshi and H. Schoorlemmer, Air shower reconstruction using HAWC and the Outrigger array, in *36th International Cosmic Ray Conference (ICRC 2019) Madison, Wisconsin, USA, July 24-August 1, 2019*, 2019.
19. A. U. Abeysekara *et al.*, Observation of the Crab Nebula with the HAWC Gamma-Ray Observatory, *Astrophys. J.* **843**, p. 39 (2017).
20. W. B. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. L. Band, G. Barbiellini and et al., The large area telescope on the fermi gamma-ray space telescope mission, *The Astrophysical Journal* **697**, p. 1071–1102 (May 2009).
21. J. R. Wood, An All-Sky Search for Bursts of Very High Energy Gamma Rays with HAWC, PhD thesis, University of Maryland, College Park2016.
22. R. Alfaro, C. Alvarez, J. D. Álvarez, R. Arceo, J. C. Arteaga-Velázquez, D. A. Rojas, H. A. A. Solares, A. S. Barber, N. Bautista-Elivar, A. Becerril and et al., Search for very-high-energy emission from gamma-ray bursts using the first 18 months of data from the hawc gamma-ray observatory, *The Astrophysical Journal* **843**, p. 88 (Jul 2017).
23. V. Chand *et al.*, Peculiar Prompt Emission and Afterglow in the H.E.S.S.-detected GRB 190829A, *Astrophys. J.* **898**, p. 42 (2020).
24. H. Ayala Solares, AMON: TeV Gamma and Neutrino Coincidence Alerts from HAWC and IceCube subthreshold data, *PoS* **ICRC2019**, p. 841 (2019).

CALET search for gamma-ray counterparts of gravitational wave events

Masaki Mori
Department of Physical Sciences, Ritsumeikan University, Kusatu, Shiga 525-8577, Japan
E-mail: morim@fc.ritsumei.ac.jp

for the CALET Collaboration*

The CALorimetric Electron Telescope (CALET) cosmic ray detector on the International Space Station (ISS) has been in operation since its launch in 2015. The main instrument, the CALET calorimeter, is monitoring the gamma ray sky from 1 GeV up to 10 TeV with a field-of-view of about 2 sr for more than five years. In this paper, we describe the analysis of gamma ray candidate events observed by the calorimeter and report on a search for gamma-ray emission from gravitational-wave event candidates announced by the LIGO/Virgo third observation run from 2019 April to 2020 March.

Keywords: Gamma rays: general – gravitational waves – methods: observational

1. Introduction

The CALorimetric Electron Telescope (CALET),[1] which was successfully launched and installed on the Japanese Experiment Module (JEM) 'Kibo'-Exposed Facility of the International Space Station (ISS) in August 2015, is a mission of the Japanese Aerospace Exploration Agency (JAXA) in collaboration with the Italian Space Agency (ASI) and NASA. The instrument has been taking science data continuously with no major interruptions since 2015 October. The main target of the CALET calorimeter is observation of high-energy cosmic rays,[2] especially electrons,[3] in the energy range from ~ 1 GeV to tens of TeV, and nuclei,[4] from ~ 10 GeV/n to ~ 100 TeV/n, but its fine detector structure allows us to observe high-energy gamma-rays[5] from ~ 1 GeV to ~ 10 TeV.

In this paper, we describe the analysis of gamma-ray candidate events observed by the CALET calorimeter and report on a search for gamma-ray emission from gravitational-wave event candidates announced by the LIGO/Virgo third observing run in the period April 2019 – March 2020. (See Ref. 6 for results during the first and second runs.) Search for low-energy emission using the CALET Gamma-ray Burst Monitor (CGBM) is reported elsewhere.[7]

2. The CALET detector

The CALET payload (Fig. 1) includes scientific and auxiliary equipment for a total mass of 613 kg, dimensions of $1.9 \times 0.8 \times 1.0$ m^3, nominal power consumption

*See the last page for the full authors list.

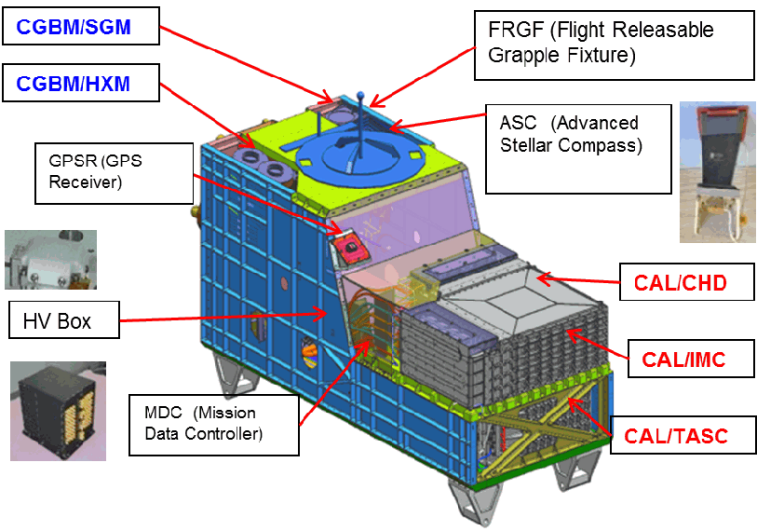


Fig. 1. The CALET payload installed at port No. 9 of the 'Kibo'-Exposed Facility of Japanese Experiment module on the ISS. The CALET calorimeter (CAL) consists of Charge Detector (CHD), Imaging Calorimeter (IMC) and Total Absorption Calorimeter (TASC). The CALET Gamma-ray Burst Monitor (CGBM) consists of two hard X-ray monitor (HXM) units and one soft gamma-ray monitor (SGM) unit.

of 507 W. The main scientific instruments are the electromagnetic calorimeter[1] and the gamma-ray burst monitor (CGBM).[8]

The CALET calorimeter is an all-calorimetric instrument designed to achieve a large proton rejection capability ($>10^5$) with a fine-grained imaging calorimeter (IMC) followed by a total absorption calorimeter (TASC). The overall thickness of CALET at normal incidence is 30 X_0 (radiation length) and ~ 1.3 proton interaction length. The charge identification of individual nuclear species is performed by a two-layered hodoscope of plastic scintillators (CHD) at the top of the apparatus, providing a measurement of the charge Z of the incident particle over a wide dynamic range ($Z = 1$ to ~ 40) with sufficient charge resolution to resolve individual elements ($\Delta Z \sim 0.1$ for light nuclei up to B, 0.3 in the Fe region) and complemented by a redundant charge determination via multiple dE/dx measurements in the IMC.

The IMC is a sampling calorimeter longitudinally segmented into 16 layers of scintillating fibers (SciFi, with 1 mm^2 squared cross-section) interspaced with thin tungsten absorbers. Alternate planes of fibers are arranged along orthogonal directions. Its surface area is 45×45 cm^2 and it can image the early shower profile in the first $3X_0$ and reconstruct the incident direction of cosmic rays with good angular resolution.

The TASC is a $27X_0$ thick homogeneous calorimeter with 12 alternate X-Y layers made of 192 lead-tungstate (PWO) logs. It measures the total energy of the incident

particle and discriminates electrons from hadrons with the help of the information from the CHD and IMC. Each log in the top layer is read by a PMT for use in the trigger system, together with CHD and IMC signals. A dual photodiode / avalanche photodiode system (APD/PD, Hamamatsu Photonics S8664-1010/S1227-33BR) is used for the read-out of the other layers. Such a readout system provides a dynamic range covering 6 orders of magnitudes and enables to measure large energy deposit expected from a proton-induced 1000 TeV shower. The TASC measures the energy of the incident particle with excellent resolution: 2% (3%) for electrons (gamma-rays) with energy $E > 10$ GeV, better than 35% for nuclei up to $\sim$ 100 TeV energy.

Charged particles and gamma rays with energy larger than 10 GeV will be triggered above a 15 MIP (minimum ionizing particle) threshold from the sum of the signals from the last two IMC SciFi belts and a 55 MIP threshold from the signal of the top layer in the TASC ("High-energy shower trigger", HE). Electrons and gamma rays in the energy range between 1 GeV and 10 GeV will be observed only for a limited exposure by reducing the IMC trigger threshold ("Low-energy shower trigger", LE, and "Low-energy gamma-ray trigger, LE-γ, respectively).

The data are transferred via two telemetry channels: one is the Ethernet channel which is operated at a medium rate of 600 kbps, and the other is a MIL-STD-1553B channel which provides low rate data at 50 kbps. In the medium rate mode, an event observed by CALET on the ISS will be transferred to the Tsukuba Space Center (TKSC) of JAXA. The low rate mode is used as a redundant channel for telemetering the CALET house keeping data plus a sample of the cosmic-ray events. The CALET data will be transmitted to TKSC via MSFC in the United State, and are sent to the Waseda CALET Operations Center (WCOC) in real time for monitoring data quality and checking for transient events.[9] The off-line data analysis is performed by the international CALET team based on the data delivered from the WCOC.

3. Analysis for gamma rays

We use two trigger modes of CAL for gamma-ray analysis: a low-energy gamma-ray (LE-γ) mode with an energy threshold $\sim$ 1 GeV used at low geomagnetic latitudes and following a CGBM burst trigger, and a high-energy (HE) mode with a threshold $\sim$ 10 GeV used in normal operation for all particles irrespective of geomagnetic latitude.[9]

The selection process of gamma-ray events used for the HE mode is essentially the same as that described in Ref. 10. For the LE-γ mode the selection and analysis are fully described in Refs. 5, 11. Here we briefly summarize the procedures.[6]

Offline trigger – In order to mitigate the inherent variability of hardware condition of flight data sample, energy deposit thresholds higher than those nominally applied by the hardware trigger are imposed both for LE-γ and HE modes.

Tracking – Event tracks are reconstructed for the HE mode using the EM track algorithm[12] developed for the electron analysis which is a powerful method for reconstructing electromagnetic showers. For the LE-γ mode we use the CC track algorithm[5] optimized for photons with energies below 10 GeV. It begins by finding clusters of hit fibers in the three bottom layers of IMC separately for the X- and Y-projections and extending the candidate tracks to upper layers of IMC. The trajectory with the highest total energy deposit is selected. The subsequent analysis applies only to contained events passing through CHD and having track lengths in TASC more than 26.4 cm for the HE mode, and well contained events whose tracks satisfy more sophisticated geometrical conditions for the LE-γ mode.[5]

Shower shape/hadronic rejection – Low energy gamma-ray events can be mimicked by albedo (i.e. upward moving) secondary charged pions from hadronic interactions in the calorimeter or the support structure. These events are vetoed by requiring to deposit more energy in the bottom layer of IMC than in the layer of pair conversion. Further rejection of events with showers not consistent with a pure electromagnetic cascade is provided by a cut on the IMC concentration, which use the lateral spread of the energy deposit distribution in the lower layers of IMC.

In order to reject hadronic events we utilize the K parameter defined as

$$K = \log_{10} F_E + R_E/2\,\mathrm{cm}$$

where F_E is the fractional energy deposit in the bottom TASC layer with respect to the total energy deposit sum in the TASC and R_E is the second moment of the lateral energy deposit distribution in the top layer of TASC. This method is developed for the derivation of the electron flux and is designed to exploit the larger spread and slower development of proton showers due to penetrating secondary pions.[13]

Charge zero – In order to select events consistent with zero primary charge, cuts are made on the energy deposits in CHD and upper IMC layers. These requirements are designed to veto charged particle events effectively. We require one of three filters utilizing CHD and upper IMC layers (see Ref. 5 for detail).

Since CALET is attached to the exposed facility of the JEM on the ISS, gamma-ray observation with CALET/CAL suffers from secondary gamma rays produced in interactions of high-energy cosmic rays with various structures of the ISS surrounding the detector. Some structures, such as the ISS truss and the JEM, are fixed to the ISS, and we can easily cut those secondary gamma rays by limiting our field-of-view (see Fig. 2 (*Left*)). However, moving structures, such as solar panels and robotic arms, produce time-varying backgrounds for gamma ray observation. In our preceding analysis, we simply rejected events coming from the field-of-view affected by moving structures.[5] We have developed moving filter algorithms to reject time-varying portions of our field-of-view by taking account of moving structures, whose operational data are supplied by JAXA, operating the 'Kibo' module, in order to maximize our exposure for cosmic gamma rays.[14]

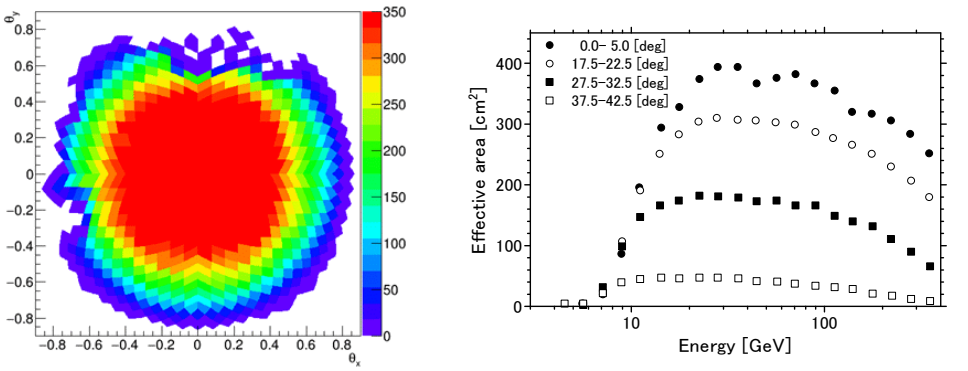


Fig. 2. (*Left*) Field-of-view of CALET/CAL, expressed in effective area (in cm^2) as a function of projected incident angles onto x- and y-plane (θ_x, θ_y). Some inefficient regions seen at edges are caused by cuts to remove secondary gamma rays from ISS structures. (*Right*) Effective area as a function of gamma-ray energy for various incidence-angle ranges.[5]

According to the detailed simulation study, we estimate the effective area and performance. The effective areas for various incident angles are shown in Fig. 2 as a function of energy.[5] The field-of-view is about 2 steradian. The angular resolution[5] and energy resolution[15] are shown in Fig. 3.

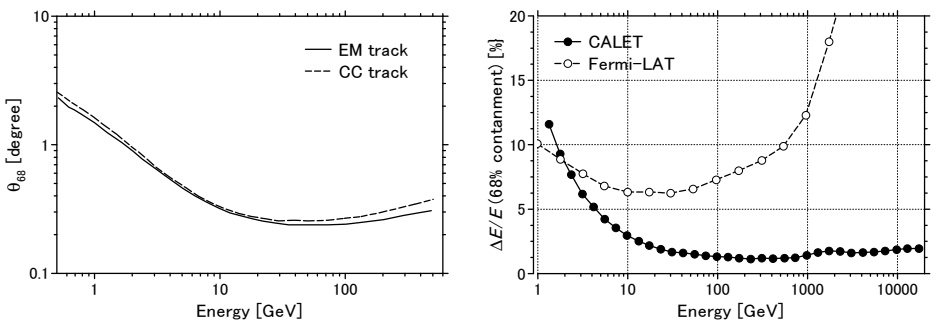


Fig. 3. (*Left*) Angular resolution of CALET for gamma rays as a function of energy, averaged over various shower starting points in IMC.[5] θ_{68} is the half angle of the cone in which 68% of gamma rays are contained. 'EM Track' is an algorithm developed for electron analysis and optimized for HE triggers, and 'CC Track' is another one developed specifically for low-energy (1–10 GeV) gamma-rays. (*Right*) Energy resolution of CALET as a function of energy,[15] compared with that of Fermi-LAT (P8R3SOURCE_V2, Total).[16]

Figure 4 shows a skymap of gamma-ray candidates (LE-γ trigger) for the data during the period 2015 November – 2018 May shown in the Galactic coordinates. Some individual sources are evident. White contours show relative exposures which are not uniform due to the inclined (51.6°) orbit of ISS.

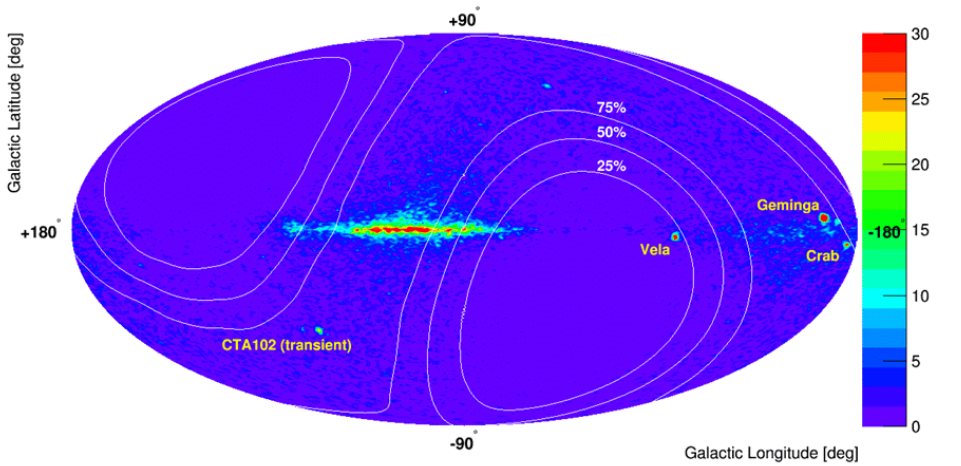


Fig. 4. Skymap of gamma-ray candidates (LE-γ trigger) shown in the Galactic coordinates for the data during the period 2015 November – 2018 May. White contours show relative exposures. Some individual sources are marked.

4. Limits on electromagnetic emission from gravitational events

We have already reported the results on the search for gamma ray emission from gravitational-wave events detected during the first and second observation run of LIGO/Virgo.[6] Also the intermediate result during the third run has been reported.[14] Here we extend the analysis for the whole third run.

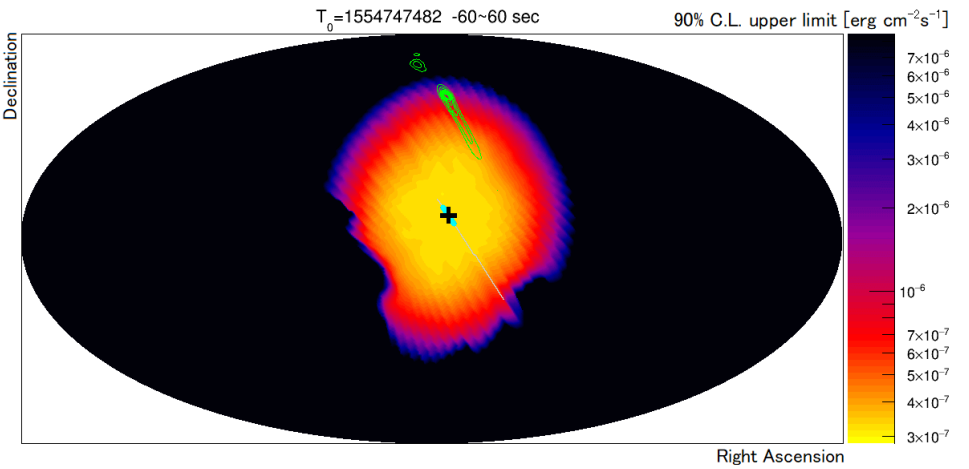


Fig. 5. 90% C.L. upper limit on S190408an energy flux in the energy region 1–10 GeV and time window $[T_0 - 60\text{s}, T_0 + 60\text{ s}]$ shown in the equatorial coordinates. The thick cyan line shows the locus of the FOV center of CAL, and the plus symbol is that at T_0. Also shown by green contours is the localization significance map of S190408an reported by LIGO/Virgo.

Since the beginning of the third run of LIGO/Virgo, we are monitoring possible gamma-ray emission based on the triggers on gravitational-event candidates supplied by the LIGO/Virgo team, in compliance with the memorandum of understanding between CALET and LIGO and Virgo. In Table 1 we summarize the CALET/CAL observations reported in GCN circulars[17] for the third run of LIGO/Virgo, from 2019 April 1 to 2020 March 27. No gamma-ray candidates have been observed in association with these triggers. Upper limits (U.L.) on the gamma-ray energy fluxes are given in unit of erg $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for the energy range 10–100 GeV except for those marked with † which are for 1–10 GeV, which corresponds to the HE and the LE-γ mode of the trigger condition of CAL around T_0. 'Coverage' is the fraction of the overlapping region of the LIGO/Virgo localization map covered by the CAL field-of-view at T_0 ('No' means there is no overlap). Also shown are the celestial coordinates of the center of CAL field-of-view at T_0. Event type in 'Comments' comes from GraceDB.[18]

An example of the energy flux limit skymaps is shown in Fig. 5 for the LIGO/Virgo trigger S190408an.[14] In this case 80% of the localization significance map of LIGO/Virgo was covered within ± 60 seconds the trigger time (T_0), but there was no gamma-ray candidate. We obtained an upper limit on gamma-ray flux of 2.3×10^{-6} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (90% C.L.) in the energy range 1–10 GeV since the CALET was in the LE-γ mode at that time..

Figure 6 shows the monthly variation of the number of LVC alerts and CALET GCN circulars. About 70% of alerts were followed and reported by CALET. Considering that the field-of-view of CAL extends to $\sim 45°$ from the zenith, the chance to have overlap with the LIGO/Virgo probable region is not small. In summary, CALET/CAL was online for 54 triggers, and for 25 triggers there were some overlapping sky area between the CALET field-of-view and the summed LIGO/Virgo probability maps. (19 in the HE mode, 6 in the LE-γ mode), but there were no gamma-ray events associated with triggers in the time window within ± 60 seconds from the trigger time (T_0) .

Although no gamma-ray event has been found for 25 triggers, upper limits on the gamma-ray flux obtained by CALET set some constraints on electromagnetic emission from gravitational-wave events as discussed in Ref. 6.

5. Summary

The CALET cosmic ray detector onboard the ISS has been monitoring cosmic gamma-rays above 1 GeV since 2015 October. In this paper we reported our procedure of gamma-ray analysis and our search for electromagnetic counterparts of gravitational-wave events upon triggers supplied by LIGO/Virgo interferometers during their third observation period, and gave upper limits on gamma-ray emission. Now, since extended operations of CALET have been approved by JAXA/NASA/ASI in March 2021 through the end of 2024, we continue observation

Table 1. Summary of CALET/CAL gamma-ray observations on gravitational-wave event candidates in the LIGO/Virgo third observing run reported in GCN circulars.[17]

GCN No.	LIGO/Virgo trigger	Trigger time T_0 (2019)	Events $T_0 \pm 60$ s	90% C.L. U.L.	Coverage	CAL α (°)	CAL δ (°)	Comments
24088	S190408an	04-08 18:18:02.288 UTC	0	2.3×10^{-6}†	80%	352.9	8.3	BBH (>99%)
24218	S190425z	04-25 08:18:05.017 UTC	0	1.0×10^{-4}	5%	131.3	-43.6	BNS (>99%)
24276	S190426c	04-26 15:21:55.337 UTC	0	2.5×10^{-5}	10%	183	-50.9	BNS (49%)
24403	S190503bf	05-03 18:54:04.294 UTC	0	4.2×10^{-5}	10%	169	-45.5	BBH (96%)
24495	S190510g	05-10 02:59:39.292 UT	0	–	No	295.7	50.8	Terrestrial (58%)
24531	S190512at	05-12 18:07:14.422 UT	0	1.9×10^{-5}	10%	214.9	37.7	BBH (99%)
24548	S190513bm	05-13 20:54:28.747 UT	0	6.0×10^{-5}†	5%	348	4.4	BBH (94%)
24593	S190517h	05-17 05:51:01.831 UT	0	–	No	126.2	-31.9	BBH (98%)
24617	S190519bj	05-19 15:35:44.398 UT	0	–	No	243.1	51.1	BBH (96%)
24648	S190521g	05-21 03:02:29.447 UT	0	6.0×10^{-6}	30%	205.7	49.2	BBH (97%)
24649	S190521r	05-21 07:43:59.463 UT	0	–	No	225.3	51.4	BBH (>99%)
24735	S190602aq	06-02 17:59:27.089 UT	0	2.9×10^{-4}	5%	127.5	45.1	BBH (99%)
24960	S190630ag	06-30 18:52:05.180 UT	0	1.2×10^{-5}	25%	84.0	31.5	BBH (94%)
24970	S190701ah	07-01 20:33:06.578 UT	0	–†	No	286.8	−1.6	BBH (93%)
25027	S190706ai	07-06 22:26:41.345 UT	0	–	No	210.4	−45.4	BBH (99%)
25033	S190707q	07-07 09:33:26.181 UT	0	2.1×10^{-6}†	20%	262.4	2.2	BBH (>99%)
25099	S190718y	07-18 14:35:12.068 UT	0	1.7×10^{-6}†	5%	195.8	−11.1	Terrestrial (98%)
25134	S190720a	07-20 00:08:36.704 UT	0	3.0×10^{-5}	25%	49.7	−32.1	BBH (99%)
25184	S190727h	07-27 06:03:33.986 UT	0	–	No	201.1	38.2	BBH (92%)
25214	S190728q	07-28 06:45:10.529 UT	0	–†	No	184.8	30.3	BBH (95%)
25390	S190814bv	08-14 21:10:39.013 UT	0	–	No	181.3	49.5	NSBH (>99%)
25536	S190828j	08-28 06:34:05.756 UT	0	–	No	13.9	12.6	BBH (>99%)
25537	S190828l	08-28 06:55:09.887 UT	0	–	No	106.9	51.0	BBH (>99%)
25647	S190901ap	09-01 23:31:01.838 UT	0	6.3×10^{-5}†	5%	353.8	16.6	BNS (86%)
25734	S190910d	09-10 01:26:19.243 UT	0	–†	No	100.8	22.9	NSBH (98%)
25735	S190910h	09-10 08:29:58.544 UT	0	9.4×10^{-6}†	10%	294.8	-5.5	BNS (61%)
25770	S190915ak	09-15 23:57:02.691 UT	0	–	No	99.7	-11.1	BBH (99%)
25830	S190923y	09-23 12:55:59.646 UT	0	1.2×10^{-5}	10%	55.3	-2.5	NSBH (68%)
25844	S190924h	09-24 02:18:46.847 UT	0	–	No	273.4	40.2	MassGap (>99%)
25891	S190930s	09-30 13:35:41.247 UT	0	3.5×10^{-5}	5%	20.7	-3.3	MassGap (95%)
25892	S190930t	09-30 14:34:07.685 UT	0	1.7×10^{-5}	5%	235.5	36.3	NSBH (74%)
26195	S191105e	11-05 14:35:21.933 UT	0	–	No	223.0	-27.4	BBH (95%)
26236	S191109d	11-09 01:07:17.221 UT	0	–	No	349.6	-16.6	BBH (>99%)
26321	S191129u	11-29 13:40:29.197 UT	0	–	No	356.9	50.7	BBH (>99%)
26358	S191204r	12-04 17:15:26.092 UT	0	–	No	269.2	34.3	BBH (>99%)
26377	S191205ah	12-05 21:52:08.569 UT	0	–	No	80.2	-32.8	NSBH (93%)
26419	S191213g	12-13 04:34:08.142 UT	0	–	No	20.4	-9.3	BNS (77%)
26465	S191215w	12-15 22:30:52.333 UT	0	–	No	222.3	40.3	BBH (>99%)
26481	S191216ap	12-16 21:33:38.473 UT	0	–†	No	186.8	13.9	BBH (99%)
26602	S191222n	12-22 03:35:37.119 UT	0	–†	No	330.3	-2.1	BBH (>99%)
26664	S200105ae	01-05 16:24:26.057 UT	0	6.5×10^{-6}	60%	50.6	-30.6	Terrestrial (97%)
26740	S200112r	01-12 15:58:38.094 UT	0	1.1×10^{-6}	5%	84.7	40.0	BBH (>99%)
26761	S200114f	01-14 02:08:18.239 UT	0	4.7×10^{-6}	80%	111.2	50.7	–
26797	S200115j	01-15 04:23:09.742 UT	0	1.7×10^{-6}	20%	84.4	45.9	MassGap (>99%)
26941	S200128d	01-28 02:20:11.903 UT	0	4.6×10^{-6}	10%	126.1	23.4	BBH (97%)
26924	S200129m	01-29 06:54:58.435 UT	0	5.7×10^{-5}	5%	288.7	-34.3	BBH (>99%)
27030	S200208q	02-08 13:01:17.991 UT	0	–	No	224.1	-41.8	BBH (>99%)
27084	S200213t	02-13 04:10:40.328 UT	0	–	No	101.4	-36.1	BBH (63%)
27149	S200219ac	02-19 09:44:15.195 UT	0	–	No	298.5	51.6	BBH (96%)
27231	S200224ca	02-24 22:22:34.406 UT	0	5.0×10^{-7}	95%	167.5	-24.8	BBH (>99%)
27232	S200225q	02-25 06:04:21.397 UT	0	–	No	157.8	-32.7	BBH (96%)
27299	S200302c	03-02 01:58:11.519 UT	0	–	No	245.4	51.6	BBH (89%)
27372	S200311bg	03-11 11:58:53.398 UT	0	–	No	191.3	51.5	BBH (>99%)
27405	S200316bj	03-16 21:57:56.157 UT	0	2.8×10^{-6}	35%	144.7	47.5	MassGap (>99%)

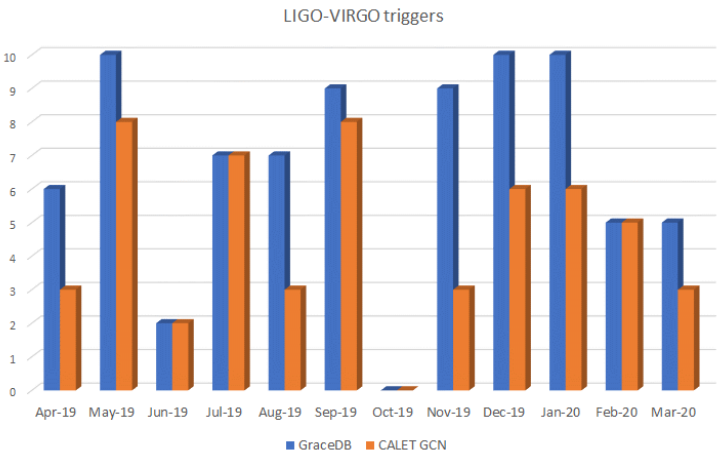


Fig. 6. Monthly variation of LVC O3 alerts and CALET GCNs.

so that we can cover the coming fourth observation run of LIGO/Virgo/KAGRA planned to start no earlier than August 2022.

Acknowledgments

We gratefully acknowledge JAXA's contributions to the development of CALET and to the operations aboard the JEM-EF on the International Space Station. We also wish to express our sincere gratitude to Agenzia Spaziale Italiana (ASI) and NASA for their support of the CALET project. This work was supported in part by JSPS Grant-in-Aid for Scientific Research (S) Number 26220708 and 19H05608, JSPS Grant-in-Aid for Scientific Research (C) Number 16K05382, and by the MEXT-Supported Program for the Strategic Research Foundation at Private Universities (2011-2015) (No. S1101021) at Waseda University. The CALET effort in Italy is supported by ASI under agreement 2013-018-R.0 and its amendments. The CALET effort in the United States is supported by NASA through Grants No. NNX16AB99G, No. NNX16AC02G, and No. NNH14ZDA001N-APRA-0075.

References

1. S. Torii for the CALET Collaboration, "The CALorimetric Electron Telescope (CALET) on the ISS: Preliminary Results from On-orbit Observations since October, 2015", *Proc. 35th ICRC* (Busan, Korea, 2017), PoS (ICRC2017) 1092; Y. Asaoka for the CALET Collaboration, "The CALorimetric Electron Telescope (CALET) on the International Space Station", *Proc. 36th ICRC* (Madison, USA, 2019), PoS (ICRC2019) 001; P. Marrocchesi for the CALET Collaboration, "New Results from the first 5 years of CALET observations on the International Space Station", *Proc. 37th ICRC* (Berlin, Germany/Online, 2021), PoS (ICRC2021) 010.

2. Pier Simone Marrocchesi for the CALET Collaboration, "CALET on the ISS: The first 5 years", in these proceedings (2021).
3. O. Adriani et al., "Extended measurement of cosmic-ray electron and positron spectrum from 11 GeV to 4.8 TeV with the calorimetric electron telescope on the International Space Station", *Phys. Rev. Lett.* **120**, 261102 (2018).
4. O. Adriani et al., "Direct Measurement of the Cosmic-Ray Proton Spectrum from 50 GeV to 10 TeV with the Calorimetric Electron Telescope on the International Space Station", *Phys. Rev. Lett.* **12**, 181102 (2019); O. Adriani et al., "Direct Measurement of the Cosmic-Ray Carbon and Oxygen Spectra from 10 GeV/n to 2.2 TeV/n with the Calorimetric Electron Telescope on the International Space Station", *Phys. Rev. Lett.*, **125**, 251102 (2020); O. Adriani et al., "Measurement of the iron spectrum in cosmic rays from 10 GeV/n to 2.0 TeV/n with the Calorimetric Electron Telescope on the International Space Station", *Phys. Rev. Lett.*, **126**, 241101 (2021).
5. N. Cannady et al., "Characteristics and Performance of the CALorimetric Electron Telescope (CALET) Calorimeter for Gamma-Ray Observations", *Astrophys. J. Suppl.* **238**, 5 (2018).
6. O. Adriani et al., "Search for GeV Gamma-Ray Counterparts of Gravitational Wave Events by *CALET*", *Astrophys. J.* **863**, 160 (2018).
7. Y. Kawakubo for the CALET Collaboration, "Gamma-ray burst observation & gravitational wave event follow-up with CALET on the International Space Station", Proc. 37th ICRC (Berlin, Germany/Online, 2021), PoS (ICRC2021) 957.
8. K. Yamaoka for the CALET Collaboration, "CALET Gamma-ray Burst Monitor (CGBM)", *Proc. 33rd ICRC* (Rio de Janeiro, Brazil, 2013), Volume 4, 2948.
9. Y. Asaoka et al., "On-orbit Operations and Offline Data Processing of CALET onboard the ISS", *Astropart. Phys.* **100**, 29 (2018).
10. M. Mori for the CALET Collaboration, "Expected Performance of CALET as a High Energy Gamma Ray Observatory", Proc. ICRC (Rio de Janeiro, Brazil, 2013), 0248, 1185.
11. N. Cannady for the CALET Collaboration, "High-Energy Gamma-ray Observations Using the CALorimetric Electron Telescope", *Proc. 35th ICRC* (Busan, Korea, 2017), PoS (ICRC2017) 720.
12. Y. Akaike for the CALET Collaboration, "CALET observational performance expected by CERN beam test", Proc. ICRC (Rio de Janeiro, Brazil, 2013), 0726, 2162.
13. O. Adriani, et al., "Energy Spectrum of Cosmic-Ray Electron and Positron from 10 GeV to 3 TeV Observed with the Calorimetric Electron Telescope on the International Space Station", *Phys. Rev. Lett.* **119**, 181101 (2017).
14. M. Mori for the CALET Collaboration, "High-Energy Gamma-ray Observations Using the CALorimetric Electron Telescope (CALET) on the ISS", *Proc. 36th ICRC* (Madison, USA, 2019), PoS (ICRC2019) 586.
15. Y. Asaoka et al., "Energy Calibration of CALET onboard the Internatinal Space Station", *Astropart. Phys.* **91**, 1 (2017).
16. "Fermi LAT Performance", `https:www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm`.
17. The Gamma-ray Coordinates Network (GCN), `https://gcn.gsfc.nasa.gov/`.
18. The Gravitational-Wave Candidate Event Database (GraceDB), `https://gracedb.ligo.org/`.

Full Authors List: The CALET Collaboration

O. Adriani[1,2], Y. Akaike[3,4], K. Asano[5], Y. Asaoka[5], E. Berti[1,2], G. Bigongiari[6,7], W. R. Binns[8], M. Bongi[1,2], P. Brogi[6,7], A. Bruno[9,10], J. H. Buckley[8], N. Cannady[11,12,13], G. Castellini[14], C. Checchia[6], M. L. Cherry[15], G. Collazuol[16,17], K. Ebisawa[18], A. W. Ficklin[15], H. Fuke[18], S. Gonzi[1,2], T. G. Guzik[15], T. Hams[11], K. Hibino[19], M. Ichimura[20], K. Ioka[21], W. Ishizaki[5], M. H. Israel[8], K. Kasahara[22], J. Kataoka[23], R. Kataoka[24], Y. Katayose[25], C. Kato[26], N. Kawanaka[27,28], Y. Kawakubo[15], K. Kobayashi[3,4], K. Kohri[29], H. S. Krawczynski[8], J. F. Krizmanic[11,12,13], P. Maestro[6,7], P. S. Marrocchesi[6,7], A. M. Messineo[30,7], J.W. Mitchell[12], S. Miyake[32], A. A. Moiseev[33,12,13], M. Mori[34], N. Mori[2], H. M. Motz[35], K. Munakata[26], S. Nakahira[18], J. Nishimura[18], G. A. de Nolfo[9], S. Okuno[19], J. F. Ormes[36], N. Ospina[16,17], S. Ozawa[37], L. Pacini[1,14,2], P. Papini[2], B. F. Rauch[8], S. B. Ricciarini[14,2], K. Sakai[11,12,13], T. Sakamoto[38], M. Sasaki[33,12,13], Y. Shimizu[19], A. Shiomi[39], P. Spillantini[1], F. Stolzi[6,7], S. Sugita[38], A. Sulaj[6,7], M. Takita[5], T. Tamura[19], T. Terasawa[40], S. Torii[3], Y. Tsunesada[41], Y. Uchihori[42], E. Vannuccini[2], J. P. Wefel[15], K. Yamaoka[43], S. Yanagita[44], A. Yoshida[38], K. Yoshida[22], and W. V. Zober[8]

[1]Department of Physics, University of Florence, Via Sansone, 1, 50019 Sesto, Fiorentino, Italy, [2]INFN Sezione di Florence, Via Sansone, 1, 50019 Sesto, Fiorentino, Italy, [3]Waseda Research Institute for Science and Engineering, Waseda University, 17 Kikuicho, Shinjuku, Tokyo 162-0044, Japan, [4]JEM Utilization Center, Human Spaceflight Technology Directorate, Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan, [5]Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwa-no-Ha, Kashiwa, Chiba 277-8582, Japan, [6]Department of Physical Sciences, Earth and Environment, University of Siena, via Roma 56, 53100 Siena, Italy, [7]INFN Sezione di Pisa, Polo Fibonacci, Largo B. Pontecorvo, 3, 56127 Pisa, Italy, [8]Department of Physics and McDonnell Center for the Space Sciences, Washington University, One Brookings Drive, St. Louis, Missouri 63130-4899, USA, [9]Heliospheric Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [10]Department of Physics, Catholic University of America, Washington, DC 20064, USA, [11]Center for Space Sciences and Technology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA, [12]Astroparticle Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [13]Center for Research and Exploration in Space Sciences and Technology, NASA/GSFC, Greenbelt, Maryland 20771, USA, [14]Institute of Applied Physics (IFAC), National Research Council (CNR), Via Madonna del Piano, 10, 50019 Sesto, Fiorentino, Italy, [15]Department of Physics and Astronomy, Louisiana State University, 202 Nicholson Hall, Baton Rouge, Louisiana 70803, USA, [16]Department of Physics and Astronomy, University of Padova, Via Marzolo, 8, 35131 Padova, Italy, [17]INFN Sezione di Padova, Via Marzolo, 8, 35131 Padova, Italy, [18]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Chuo, Sagamihara, Kanagawa 252-5210, Japan, [19]Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa, Yokohama, Kanagawa 221-8686, Japan, [20]Faculty of Science and Technology, Graduate School of Science and Technology, Hirosaki University, 3, Bunkyo, Hirosaki, Aomori 036-8561, Japan, [21]Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo, Kyoto 606-8502, Japan, [22]Department of Electronic Information Systems, Shibaura Institute of Technology, 307 Fukasaku, Minuma, Saitama 337-8570, Japan, [23]School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan, [24]National Institute of Polar Research, 10-3, Midori-cho, Tachikawa, Tokyo 190-8518, Japan, [25]Faculty of Engineering, Division of Intelligent Systems Engineering, Yokohama National University, 79-5 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan, [26]Faculty of Science, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan, [27]Hakubi Center, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan, [28]Department of Astronomy, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan, [29]Institute of Particle and Nuclear Studies, High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan, [30]University of Pisa, Polo Fibonacci, Largo B. Pontecorvo, 3, 56127 Pisa, Italy, [31]Astroparticle Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [32]Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Ibaraki College, 866 Nakane, Hitachinaka, Ibaraki 312-8508, Japan [33]Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA, [34]Department of Physical Sciences, College of Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan, [35]Faculty of Science and Engineering, Global Center for Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan, [36]Department of Physics and Astronomy, University of Denver, Physics Building, Room 211, 2112 East Wesley Avenue, Denver, Colorado 80208-6900, USA, [37]Quantum ICT Advanced Development Center, National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan, [38]College of Science and Engineering, Department of Physics and Mathematics, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo, Sagamihara, Kanagawa 252-5258, Japan, [39]College of Industrial Technology, Nihon University, 1-2-1 Izumi, Narashino, Chiba 275-8575, Japan [40]RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [41]Division of Mathematics and Physics, Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan, [42]National Institutes for Quantum and Radiation Science and Technology, 4-9-1 Anagawa, Inage, Chiba 263-8555, Japan, [43]Nagoya University, Furo, Chikusa, Nagoya 464-8601, Japan, [44]College of Science, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan

Off-axis jet scenario for early afterglow emission of low-luminosity Gamma-Ray Burst GRB 190829A

Yuri Sato[1,*], Kaori Obayashi[1], Ryo Yamazaki[1,2], Kohta Murase[3,4,5,6] and Yutaka Ohira[7]

[1] *Department of Physics and Mathematics, Aoyama Gakuin University, 5-10-1 Fuchinobe, Sagamihara 252-5258, Japan*

[2] *Institute of Laser Engineering, Osaka University, 2-6 Yamadaoka, Suita, Osaka 565-0871, Japan*

[3] *Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*

[4] *Department of Astronomy & Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*

[5] *Center for Multimessenger Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA*

[6] *Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto, Kyoto 606-8502, Japan*

[7] *Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[*] *E-mail: yuris@phys.aoyama.ac.jp*

Low-luminosity gamma-ray bursts (GRBs) have small isotropic equivalent gamma-ray energy and luminosity, compared with normal long GRBs. GRB 190829A is a member of this class. Furthermore, very-high-energy gamma-rays were detected by H.E.S.S. with $\sim 20\sigma$ significance. This event had more unusual features. First, it had much smaller isotropic equivalent gamma-ray energy than typical long gamma-ray bursts and is classified as low-luminosity GRB. Second, early X-ray and optical afterglow emission showed a rising part and simultaneously peaked at about 1400 s. We propose an off-axis jet scenario that explains these observational results. In this model, the relativistic beaming effect is responsible for the apparently small isotropic gamma-ray energy and spectral peak energy. Using a jetted afterglow model, we find that the narrow jet, which has the initial Lorentz factor of 350 and the initial jet opening half-angle of 0.015 rad, viewed off-axis can describe the observed achromatic behavior in the X-ray and optical afterglow. Another wide, baryon-loaded jet is necessary for the later-epoch X-ray and radio emissions. Derived parameters explains the very-high-energy gamma-ray flux at 20,000 s.

Keywords: Gamma-ray bursts: individual: GRB 190829A

1. Introduction

Low-luminosity gamma-ray bursts (GRBs) have small isotropic equivalent gamma-ray energy $E_{\rm iso,\gamma}$ and luminosity $L_{\rm iso,\gamma}$, compared with normal long GRBs. For most of them, supernovae are often associated. Currently, the origin is unknown. Possible scenarios are a shock breakout,[18] a low-power relativistic jet,[26] an off-axis jet,[30] and so on.

GRB 190829A is classified as a low-luminosity GRB.[4] The prompt gamma-ray emission (from $\sim$ 10 keV to MeV band) consists of two temporally separated components.[4] The burst started with less energetic emission (hereafter Episode 1

following Chand et al.[4]) with $E_{\rm iso,\gamma} = 3.2 \times 10^{49}$ erg and a peak energy (that is, the photon energy at which the νF_ν-spectrum takes a maximum) $E_p = 120$ keV. After quiescent time interval lasting about 40 s, the second brighter emission (Episode 2) with $E_{\rm iso,\gamma} = 1.9 \times 10^{50}$ erg and $E_p = 11$ keV, appeared. The observed values of $E_{\rm iso,\gamma}$ and E_p of Episode 2 are consistent with Amati relation,[1,21] while those of Episode 1 are in the region of low-luminosity GRBs. Both Episode 1 and 2 have smaller $E_{\rm iso,\gamma}$ and E_p than typical long GRBs.[14] Indeed, GRB 190829A occurred so nearby with a redshift of $z = 0.0785$ that such weak prompt emissions could be observed.

Well-sampled afterglow light curves of GRB 190829A were obtained in X-ray, optical/infrared (IR),[4] radio bands[20] and very-high-enrgy (VHE) gamma-rays.[12] VHE gamma-ray emission was detected by H.E.S.S. about 2×10^4 s after the burst trigger and its significance is $\sim 20\ \sigma$.[12] It is remarkable that early X-ray and optical/IR afterglow emission showed a rising part and simultaneously peaked at about 1.4×10^3 s. Such an "achromatic" behavior is difficult to be explained in standard afterglow model, in which the synchrotron emission has the maximum when the typical frequency ν_m crosses the observation bands.[23] In contrast, the other VHE events, GRB 190114C and 180720B, showed monotonically decaying X-ray afterglow emission.[8,29] Possible interpretations of the achromatic bump are the X-ray flare with optical counterpart,[4,32,34] the afterglow onset of baryon loaded outflow,[9] e^+e^- dust shell,[17] and the reverse shock emission.[22] Another interesting point is that late time ($t > 10^{4-5}$ s) optical/IR emissions are dominated by supernova component.[13]

In this proceeding, we consider off-axis jet scenario to explain those unusual observed properties of GRB 190829A. If the jet is viewed off-axis ($\theta_v < \theta_j$), the relativistic beaming effects cause the prompt emission to be dimmer and softer than on-axis ($\theta_v \sim 0$) viewing case.[16,27,28,30] The bulk Lorentz factor of the jet is initially so high that the afterglow emission is very dim because of the relativistic effect. As the jet decelerates, the beaming effect becomes weak, resulting in the emergence of a rising part in afterglow light curves. After the peak of the emission, the jet has smaller Lorentz factors so that the light curve only weakly depends on the viewing angle. For more details, see our paper (Sato et al.[25]).

2. Model

Following Huang et al.,[15] the dynamics of the shock is numerically computed (see Sato et al.[25] for details). The jet with isotropic-equivalent kinetic energy $E_{\rm iso,K}$, initial Lorentz factor Γ_0, and initial jet opening half-angle θ_0 decelerates via interactions with ISM and forms a thin shell. In calculating synchrotron radiation, we assume that microphysics parameters ϵ_e and ϵ_B, the energy fractions of internal energy going into radiating electrons and magnetic field, are constant. The electron energy distribution in the emitting thin shell has a power-law form with index p. In the slow cooling regime, the electron spectrum has a break at the electron cooling

Lorentz factor γ_c, where we take into account the SSC cooling in the Thomson limit as well as synchrotron energy losses.[5,24,31] Then, it has a form $N(\gamma_e) \propto {\gamma_e}^{-p}$ when $\gamma_m < \gamma_e < \gamma_c$ and $N(\gamma_e) \propto {\gamma_e}^{-p-1}$ when $\gamma_c < \gamma_e$. The flux density F_ν of the afterglow emission that arrives at the observer time T is obtained by integrating the emissivity over the equal arrival time surface (e.g., Granot, Piran, & Sari[10]). Parameters of the present model are isotropic-equivalent kinetic energy $E_{\rm iso,K}$, initial Lorentz factor Γ_0, initial jet opening half-angle θ_0, ISM density n_0, microphysics parameters ϵ_e and ϵ_B, electron power-law index p, and the viewing angle θ_v.

3. Results

In this section, we show our numerical results of synchrotron afterglow emission in the X-ray (10^{18} Hz), optical (V-band), and radio (1.3 and 15.5 GHz) bands, and compare them with observation data of GRB 190829A. The X-ray data are extracted from the *Swift* team website[a].[6,7] The optical V-band data (before the absorption correction) are obtained from Chand et al.[4] In our numerical calculation, we take the V-band extinction $A_{\rm V} = 1.5$ mag.[4] The radio data are taken from Rhodes et al.[20]

First, a single jet viewed off-axis is considered in order to discuss the observed X-ray and optical bumps around $T \sim 1.4 \times 10^2$ s. We adopt $\theta_v = 0.031$ rad, $\theta_0 = 0.015$ rad, $E_{\rm iso,K} = 4.0 \times 10^{53}$ erg, $\Gamma_0 = 350$, $n_0 = 0.01$ cm^{-3}, $\epsilon_e = 0.2$, $\epsilon_B = 5.0 \times 10^{-5}$ and $p = 2.44$. The initial opening half-angle is small, so that we refer to "narrow jet" in the following. However, the jet is still "fat" in the sense $\theta_0 > \Gamma_0^{-1}$, so that the jet dynamics is able to be discussed as in a standard manner. Our off-axis afterglow model well explains the observational results of early X-ray and optical afterglow from about 8×10^2 to 2×10^4 s. An achromatic behavior in the X-ray and optical bands is evident. The off-axis afterglow starts with a rising part because of the relativistic beaming effect.[11] As the jet decelerates, the observed flux increases. When the jet Lorentz factor becomes $\Gamma \sim (\theta_v - \theta_0)^{-1} = 65$, the afterglow light curve takes a maximum. If the adiabatic evolution ($\Gamma \propto t^{-3/8}$) is assumed, the observer time of the flux maximum is analytically given by

$$T_{pk} \sim (1+z) \left(\frac{3E_{\rm iso,K}}{4\pi n_0 m_p c^5} \right)^{\frac{1}{3}} (\theta_v - \theta_0)^{\frac{8}{3}} \quad . \tag{1}$$

For our model parameters, we get $T_{pk} \sim 2 \times 10^3$ s, which is consistent with our numerical results within a factor of two. Parameter dependence has been throughly investigated in Sato et al.[25]

Next, a two-component jet model is considered, in which another "wide jet" is introduced in addition to the narrow jet. The observed flux is simply the superposition of each jet emission components. For the wide jet, we adopt $\theta_0 = 0.1$ rad,

[a] https://www.swift.ac.uk/xrt_curves/00922968/

$E_{\rm iso,K} = 2.0 \times 10^{53}$ erg, $\Gamma_0 = 20$, $\epsilon_e = 0.4$, $\epsilon_B = 1.0 \times 10^{-5}$, and $p = 2.2$. The values of θ_v and n_0 are common for both jets. We assume that the central axes of the two jets are identical ($\theta = 0$) and that the two jets depart from the central engine ($r = 0$) at the same time. One can find in Fig. 1 that early achromatic peaks in the X-ray and optical bands are explained by the off-axis narrow jet emission (dashed lines in the right panel), and that the late X-ray and radio afterglow is interpreted with the wide jet emission (dotted lines).

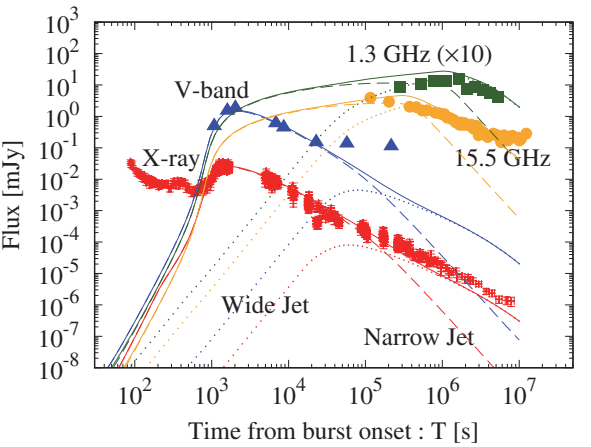


Fig. 1. Afterglow light curves calculated by our two-component jet model — solid lines are the sum of the narrow (dashed lines) and wide (dotted lines) jets in the X-ray (10^{18} Hz: red), optical (V-band: blue) and radio bands (1.3 GHz: orange, 15.5 GHz: green), which is compared with the observed data of GRB 190829A (X-ray: red points, V-band: blue triangles, 1.3 GHz: orange filled-circles, 15.5 GHz: green squares). Note that late-time ($t > 5 \times 10^4$ s) optical band is dominated by supernova component.[13]

4. Discussion

We have investigated an off-axis jet scenario in which we have invoked a two-component jet model to explain the observational results of GRB 190829A. The best-fitted model in this paper is shown by solid lines in Fig. 1. According to our model, the early X-ray and optical afterglow was off-axis emission from the narrow jet, and the late X-ray and radio afterglow came from the wide jet.

There are still some observed components that are brighter than the prediction of our jet model. They may be other components. For example, very early ($T \leq 7 \times 10^2$ s) X-ray emission should be the contribution from late prompt emission like flares. Or if the jet is structured, the early X-ray afterglow shows a plateau phase or an additional peak.[2,3,19] The observed optical flux later than $\sim 5 \times 10^4$ s is a supernova component.[13] At the late epoch ($T \sim 10^7$ s), the 15.5 GHz radio flux also exceeds our numerical result, which could be other components such as counter-jet

emission. As seen in Fig. 1, our theoretical radio fluxes in both 1.3 and 15.5 GHz with our parameters sometimes overshot the observed ones. However, the excess is only within a factor of two, and this difference may come from the uncertainty of our simple model.

The prompt emission of GRB 190829A had smaller values of the peak energy $E_{\rm p}$ and the isotropic-equivalent gamma-ray energy $E_{\rm iso,\gamma}$ than typical long GRBs. Here, it is discussed whether $E_{\rm p}$ and $E_{\rm iso,\gamma}$ from our narrow jet were typical or not if it would have been viewed on-axis ($\theta_v \approx 0$). We consider a simple emission model[16,27,28] assuming an instantaneous emission of an infinitesimally thin shell moving with the Lorentz factor Γ_0. If the narrow jet emitted Episode 1 of observed prompt emission (see § 1), that is, $E_p(\theta_v = 0.031) = 120$ keV and $E_{\rm iso,\gamma}(\theta_v = 0.031) = 3.2\times10^{49}$ erg,[4] then on-axis quantities, $E_p(0)$ and $E_{\rm iso,\gamma}(0)$, are obtained as $E_p(0) = E_p(\theta_v)/R_1 = 3.7$ MeV and $E_{\rm iso,\gamma}(0) = E_{\rm iso,\gamma}(\theta_v)/R_2 = 2.7\times10^{53}$ erg (For details, see Sato et al.[25]). These values are within the range for bursts detected so far (e.g., Zhao et al.[33]). The isotropic equivalent kinetic energy of the narrow jet just after the prompt emission is $E_{\rm iso,K} = 4.0\times10^{53}$ erg (see § 3), so that the efficiency of the prompt emission is calculated as $\eta_\gamma = E_{\rm iso,\gamma}(0)/(E_{\rm iso,\gamma}(0)+E_{\rm iso,K}) \approx 0.4$. On the other hand, if the narrow jet is responsible for Episode 2 (that is, $E_p(\theta_v = 0.031) = 11$ keV and $E_{\rm iso,\gamma}(\theta_v = 0.031) = 1.9\times10^{50}$ erg), we obtain $E_p(0) = 340$ keV and $E_{\rm iso,\gamma}(0) = 1.6\times10^{54}$ erg, which are again similar to typical long GRBs. In this case, the efficiency is $\eta_\gamma \approx 0.8$. If the narrow jet causes Episode 1, then the estimated prompt emission efficiency η_γ is almost typical, however on-axis $E_{\rm p}(0)$ is located at the highest end of the distribution for long GRBs. On the other hand, if the narrow jet produced Episode 2, then on-axis $E_{\rm p}(0)$ is smaller though η_γ is somewhat higher (but it is still comparable, and one can say that the value is reasonable considering very simple approximation of our prompt emission model). Episode 1 and 2 may be emitted from narrow and wide jets, respectively. Note that if the wide jet emits Episode 2, its efficiency is small, $\eta_\gamma \approx 5\times10^{-3}$, so that it might be natural that the narrow jet causes both Episode 1 and 2.

The VHE gamma-ray flux at 2×10^4 s is estimated, where we assume the synchrotron self-Compton emission in the Thomson limit.[24] Our narrow jet parameters determined by X-ray, optical and radio afterglows may roughly explain observed VHE gamma-ray flux at 2×10^4 s.[25] On the other hand, the VHE gamma-ray flux from wide jet has brighten until at 5×10^4 s. After this time, the flux decays.

Acknowledgments

We thank the organizers of the successful MG16 meeting. This research was partially supported by JSPS KAKENHI Grant Nos. 18H01232 (RY), 20H01901 (KM), 20H05852 (KM) and JP19H01893 (YO). R.Y. deeply appreciates Aoyama Gakuin University Research Institute for helping our research by the fund. The work of K.M. is supported by NSF Grant No. AST-1908689. Y.O. is supported by Leading Initiative for Excellent Young Researchers, MEXT, Japan.

References

1. Amati L., et al., *A&A* **390**, 81 (2002).
2. Beniamini P., Duque R., Daigne F., Mochkovitch R., *Mon. Not. R. Astron. Soc.* **492**, 2847 (2020).
3. Beniamini P., Granot J., Gill R., *Mon. Not. R. Astron. Soc.* **493**, 3521 (2020).
4. Chand V., et al., *Astrophys.* **898**, 42 (2020).
5. Dermer C. D., Chiang J., Mitman K. E., *Astrophys.* **537**, 785 (2020).
6. Evans P. A., et al., *A&A* **469**, 379 (2007).
7. Evans P. A., et al., *Mon. Not. R. Astron. Soc.* **397**, 1177 (2009).
8. Fraija N., et al., *Astrophys.* **885**, 29 (2019).
9. Fraija N., eal., 2020, arXiv, arXiv:2003.11252
10. Granot J., Piran T., Sari R., *Astrophys.* **513**, 679 (1999).
11. Granot J., Panaitescu A., Kumar P., Woosley S. E., *Astrophys. Lett.* **570**, L61 (2002).
12. H. E. S. S. Collaboration, *Science* **372**, 1081 (2021).
13. Hu Y.-D., et al., *A&A* **646**, A50 (2021).
14. Huang X.-L., Liang E.-W., Liu R.-Y., Cheng J.-G., Wang X.-Y., *Astrophys. Lett.* **903**, L26 (2021).
15. Huang Y. F., Gou L. J., Dai Z. G., Lu T., *Astrophys.* **543**, 90 (2000).
16. Ioka K., Nakamura T., *Astrophys. Lett.* **554**, L163 (2001).
17. Lu-Lu Z., et al., arXiv:2106.03466.
18. Nakar E., Sari R., *Astrophys.* **747**, 88 (2012).
19. Oganesyan G., Ascenzi S., Branchesi M., Salafia O. S., Dall'Osso S., Ghirlanda G., *Astrophys.* **893**, 88 (2020).
20. Rhodes L., et al., *Mon. Not. R. Astron. Soc.* **496**, 3326 (2020).
21. Sakamoto T., et al., *Astrophys.* **679**, 570 (2008).
22. Salafia O. S., et al., arXiv:2106.07169.
23. Sari R., Piran T., Narayan R., *Astrophys. Lett.* **497**, L17 (1998).
24. Sari R., Esin A. A., *Astrophys.* **548**, 787 (2001).
25. Sato Y., et al., *Mon. Not. R. Astron. Soc.* **504**, 5467 (2021).
26. Toma K., Ioka K., Sakamoto T., Nakamura T., *Astrophys.* **659**, 1420 (2007).
27. Yamazaki R., Ioka K., Nakamura T., *Astrophys. Lett.* **571**, L31 (2002).
28. Yamazaki R., Ioka K., Nakamura T., *Astrophys.* **593**, 941 (2003).
29. Yamazaki R., Sato Y., Sakamoto T., Serino M., *Mon. Not. R. Astron. Soc.* **494**, 5259 (2021).
30. Yamazaki R., Yonetoku D., Nakamura T., *Astrophys. Lett.* **594**, L79 (2003).
31. Zhang B., Mészáros P., *Astrophys.* **559**, 110 (2001).
32. Zhang B. T.. Murase, K., et al., 2020a, arXiv:2012.07796
33. Zhao W., et al., *Astrophys.* **900**, 112 (2020).
34. Zhao L., et al., arXiv:2012.12036.

GRB prompt phase spectra under backscattering dominated model

Mukesh Kumar Vyas* and Asaf Pe'er
Department of Physics, Bar Ilan University,,
Ramat Gan, 5290002, Israel
**E-mail: mukeshkvys@gmail.com*
asaf.peer@biu.ac.il
www.biu.ac.il

David Eichler
Ben-Gurion University,
Be'er Sheva, 84105, Israel

We propose a backscattering dominated prompt emission model for gamma ray bursts (GRB) prompt phase in which the photons generated through pair annihilation at the centre of the burst are backscattered through Compton scattering by an outflowing stellar cork. We show that the obtained spectra are capable of explaining the low and high energy slopes as well as the distribution of spectral peak energies in their observed prompt spectra.

Keywords: High energy astrophysics; Gamma-ray bursts; Relativistic jets; Theoretical models

1. Introduction

In the picture of gamma ray bursts (GRBs), collapse of a massive star gives rise to a long GRB, while a merger of two compact objects leads to a short GRB. In both the cases, strong bipolar jets are produced.[1,2] Radiation from these jets is observed if the observer is near the jet axis. The initial phase of the burst is known as prompt phase. In a typical understanding of this phase, the jet propagates inside the stellar interior and forms an envelope (stellar cork). Finally, the jet breaks out of the cork to be observed.[3,4]

An alternate view of this stage was proposed,[5–9] in which the core of the star produces electron-positron pair plasma through neutrino annihilation. This pair plasma emits radiation in form of a jet through pair annihilation and pushes the stellar material to form the stellar cork. This cork is further pushed to relativistic speeds by radiation pressure. The radiation is not able to pierce the cork and is backscattered. As the cork is moving with relativistic speeds, this radiation is beamed along the direction of the jet and is seen by the observer.

In this work, we carry out Monte Carlo simulations of the interaction of the seed photons with the cork where the backscattered photons lead to the observed spectrum. We show that the spectrum observed is capable of explaining a large range of observed spectral properties like the photon indices at low and high energies along with the spectral peak energies.

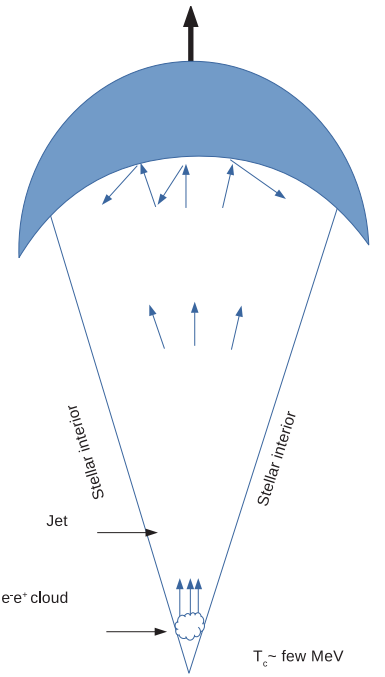


Fig. 1. Geometric representation of the system. The seed photons are generated near the centre of the bursts through pair annihilation and comptonized with the plasma present there. These photons propogate in empty jet funnel and interact with a radially outflowing stellar cork.

2. Physical picture of the system and source of the seed photons

At the time of stellar collapse, neutrinos are produced and get annihilated near the centre of the star. This annihilation produces an electron-positron pair dominated plasma. This pair plasma produces annihilation spectrum along with bremmstrahlung and is further Comptonized by the local plasma. Using Monte Carlo simulations, the resultant spectra were studied by.[10] The spectra were found to be flat with exponential decay at high energies. We obtained numerical fits to his results. For plasma density $n = 2 \times 10^{18}$ cm^{-3} the fitted spectrum is

$$F_\varepsilon = C_0 \exp\left(-\frac{C_1 \varepsilon^2}{\Theta_r^2}\right) \quad \text{KeV/s/KeV}^{-1} \tag{1}$$

Here ε photon energy normalized to electron's rest mass energy. $\Theta_r = k_B T_r / m_e c^2$ is dimensionless representation of pair temperature T_r with k_B, m_e and c are Boltzmann's constant, rest mass of the electron and light speed, respectively. $C_1 = 0.045$ and $C_0 (= 2 \times 10^{40}$ KeV s^{-1}/KeV). This fit gives reliable spectra for temperatures $\Theta_r > 0.3$. The photons with this spectral shape propagate within the jet funnel and push the stellar matter that forms a stellar cork ahead of it. The cork moves with relativistic speeds and escapes the stellar surface. The seed photons enter the

hot cork with temperature T_c (or $\Theta_c = k_B T_c / m_e c^2$). After going through Compton scattering with the electrons inside the cork, the photons gain or loose energy and get backscattered from the rear end. The geometry of this figure is explained in Figure 1. Because of the relativistic motion of the cork with Lorentz factor γ, these photons are then observed by an observer situated at angle θ_{obs} from the jet axis. The cork has curved inner surface and the photons scattered at different locations are observed at different times as well as with different energies producing a light curve as well as spectrum.

3. Solution procedure

Through a Monte Carlo code, we inject approximately 26 million seed photons with the spectral distribution according to Equation 1. These photons are launched at an adiabatically expanding and outflowing cork with temperature T_c and Lorentz factor γ with an opening angle $\theta_j = 0.1$ rad. The photons enter the cork from its rear end (Figure 1) and scattered backwards. Before escaping, the photons go though multiple scattering with energetic electrons inside the cork and thus their energies evolve. More details of the calculation procedure of the Monte Carlo code is described before.[8,11] These photons are then observed in specific angular patch $d\theta_{obs} = 0.005$ rad about the observer's location θ_{obs}. The resultant spectrum is shown in the rest frame of the GRBs with an assumption that the total energy released from the centre is equivalent to 10^{50} erg.

4. Results

In Figure 2, we plot a typical photon spectrum (photons $\mathrm{cm}^{-2}\mathrm{KeV}^{-1}$) obtained by scattering of seed photons with $\Theta_r = 3$ and cork temperature $\Theta_c = 1.4$. The cork has a bulk Lorentz factor $\gamma = 100$ and initial location at $r_i = 3 \times 10^{12}$cm from the centre of the star. The opening angle of the cork is taken to be $\theta_j = 0.1$ rad. The spectrum obtained has a low energy photon index $\alpha = -1.1$ while high energy index is $\beta = -2.75$. In Figure 3 we show the variation of spectral emissivity KeV cm^{-2}. The Spectral peak is obtained to be at $\varepsilon_{peak} = 30$ MeV. This typical spectrum obtained resembles the well known GRB prompt phase spectral shape with two power laws separated by a spectral peak.[12]

Keeping the same parameters as above, we plot the variation of ε_{peak} with observer's position θ_{obs} in Figure 4 and find that it varies from few 10 MeV to few KeV as the observing angle changes from 0 to 0.35 rad. The evolution of peak energies are followed due to relativistic kinematics. The photons are blueshifted for an on axis observer while a far off axis observer sees the photons to be having relatively less energies. This shifts the spectra to lower energies and subsequently low peak energies are observed. The obtained range of peak energies explains the typically observed range of the spectral peaks in GRB prompt phase spectra.[13]

Finally, in Figure 5 we show that the photon indices α and β are sensitive to the cork temperature Θ_c. In the range $\Theta_c = 1 - 10$, the magnitude of both indices

decreases making the spectra harder. It is understandable as the spectrum becomes harder due to high energy electrons in cork at high temperatures.

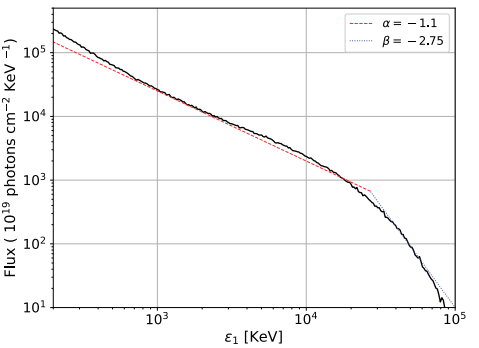


Fig. 2. Photon spectrum obtained for $\Theta_c = 1.4$, and $\theta_j = 0.1$rad. The observer is situated at $\theta_{obs} = 0.05$ rad. Here $\Theta_r = 3.0$ and $\gamma = 100$

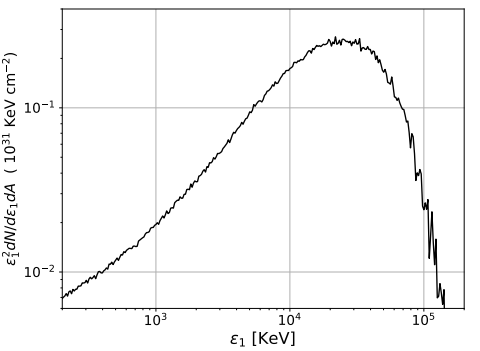


Fig. 3. Emissivity KeV cm^{-2} for the same parameters as in Figure 2

5. Conclusions

In this proceedings, we have shown that the typical spectrum of a gamma ray prompt phase is consistent with the theoretical prediction of backscattering dominated picture. We show that the spectrum has negative α and steeper β. The ranges obtained for both the indices is $\alpha \sim -1.9$ to -1.1 and $\beta \sim -3$ to -2.4. The spectral peak energies decrease as the observer is located farther from the axis or θ_{obs} is greater. In such as case, the peak energies are obtained to be in the range ε_{peak} = few 10 KeV to few 10 MeV. All the obtained ranges of these parameters, α, β and ε_{peak} are within the observed ranges of GRB prompt phase observations.[12–14]

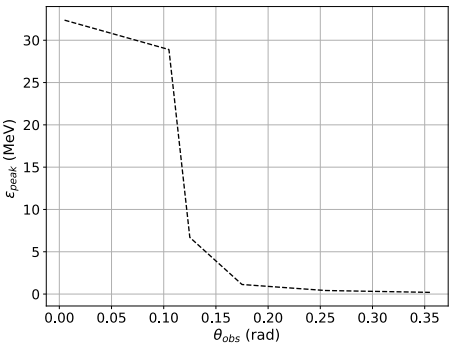


Fig. 4. Variation of the spectral peak energy ε_{peak} with observer's angle θ_{obs} for the parameters as in Figure 2.

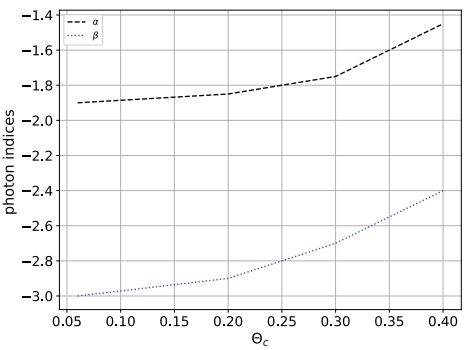


Fig. 5. Variation of photon indices α and β with the cork temperature Θ_c for $\Theta_r = 3$ and $\theta_{obs} = 0.105$ rad.

Acknowledgments

After recent demise of our collaborator late David Eichler, the completion of this work is a tribute to him.

References

1. A. Levinson and D. Eichler, Baryon purity in cosmological gamma-ray bursts as a manifestation of event horizons, *The Astrophysical Journal* **418**, p. 386 (1993).
2. A. MacFadyen and S. Woosley, Collapsars: Gamma-ray bursts and explosions in "failed supernovae", *The Astrophysical Journal* **524**, p. 262 (1999).
3. E. Ramirez-Ruiz, A. I. MacFadyen and D. Lazzati, Precursors and e± pair loading from erupting fireballs, *Monthly Notices of the Royal Astronomical Society* **331**, 197 (2002).
4. W. Zhang, S. Woosley and A. MacFadyen, Relativistic jets in collapsars, *The Astrophysical Journal* **586**, p. 356 (2003).

5. D. Eichler and H. Manis, Spectral lags explained as scattering from accelerated scatterers, *The Astrophysical Journal Letters* **689**, p. L85 (2008).
6. D. Eichler, Cloaked gamma-ray bursts, *The Astrophysical Journal Letters* **787**, p. L32 (2014).
7. D. Eichler, Short gamma-ray bursts viewed from far off-axis, *The Astrophysical Journal Letters* **869**, p. L4 (2018).
8. M. K. Vyas, A. Pe'er and D. Eichler, A backscattering-dominated prompt emission model for the prompt phase of gamma-ray bursts, *The Astrophysical Journal* **908**, p. 9 (2021).
9. M. K. Vyas, A. Pe'er and D. Eichler, Predicting spectral parameters in the backscattering dominated model for the prompt phase of GRBs, *The Astrophysical Journal Letters* **918**, p. L12 (2021).
10. A. Zdziarski, Effect of compton scattering on the pair annihilation and bremsstrahlung spectra, *Physica Scripta* **1984**, p. 124 (1984).
11. A. Pe'er, Temporal evolution of thermal emission from relativistically expanding plasma, *The Astrophysical Journal* **682**, p. 463 (2008).
12. R. D. Preece, M. S. Briggs, R. S. Mallozzi, G. N. Pendleton, W. Paciesas and D. L. Band, The batse gamma-ray burst spectral catalog. i. high time resolution spectroscopy of bright bursts using high energy resolution data, *The Astrophysical Journal Supplement Series* **126**, p. 19 (2000).
13. G. Ghirlanda, G. Ghisellini and L. Nava, Short and long gamma-ray bursts: Same emission mechanism?, *Monthly Notices of the Royal Astronomical Society: Letters* **418**, L109 (2011).
14. A. Pe'er, Physics of gamma-ray bursts prompt emission, *Advances in Astronomy* **2015** (2015).

Applying models of pulsar wind nebulae to explain X-ray plateaux following short Gamma-Ray Bursts

L. C. Strang* and A. Melatos

School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
** E-mail: lstrang@student.unimelb.edu.au*

Australian Research Council Centre of Excellence in Gravitational-wave Discovery (OzGrav), School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
** E-mail: lstrang@student.unimelb.edu.au*

Many short Gamma-Ray Bursts (sGRBs) have a prolonged plateau in the X-ray afterglow lasting up to tens of thousands of seconds. A central engine injecting energy into the remnant may fuel the plateau. A simple analytic model describing the interaction of the magnetized relativistic wind from a rapidly-rotating magnetar with the surrounding environment can reproduce X-ray plateaux and instantaneous spectra. The model is analogous to classic, well-established models of young supernova remnants and applies the underlying physics to sGRB remnants. The light curve and spectra produced by the model are compared to observations of GRB 130603B. The spectra are also used to estimate parameters of the magnetar including its poloidal field strength and angular frequency. If combined with a gravitational wave signal, this model could provide insight into multimessenger astronomy and neutron star physics.

Keywords: Gamma-ray bursts; afterglow; supernova remnants

1. Introduction

Many short gamma-ray bursts (sGRBs) display long-lived emission in the X-ray band. A subset of the X-ray afterglows are "canonical" afterglows with three components: an initial decay, a flat plateau lasting from 10 s – 10^5 s, and a final decay (typically $\sim t^{-2}$).[1,2] The luminosity and duration of the plateau are correlated, with brighter plateaux ending sooner.[3] A similar phenomenon is observed in long gamma-ray bursts (LGRBs), but the two populations are statistically distinct and likely have different progenitors for the afterglow as well as the prompt emission.[4–6]

Binary neutron star coalescence has been confirmed as a progenitor of sGRBs,[7,8] suggesting the evolution of the post-merger remnant dictates the evolution of the afterglow. X-ray plateaux have inspired several models including fireballs (with[9,10] and without[11] energy injection), fall-back accretion onto a black hole,[12] and ongoing energy injection via a central engine,[13] commonly assumed to be a millisecond magnetar.[14] In the latter scenario, the rotational energy of the magnetar is converted to X-rays via an unknown dissipative process, perhaps involving a relativistic wind.[13–19] Previous models have considered the evolution of a magnetar surrounded

by a shroud of optically-thick ejecta material[20–22] or the production of X-rays via radiative losses from interactions with the surrounding environment.[23,24]

In this work, we present results from Ref. 25 and Ref. 26 exploring a millisecond magnetar central engine through the lens of classic models of pulsar wind nebulae (PWNe), also known as plerions. The term plerion is used throughout Ref. 25 and Ref. 26. Here, we use the term PWNe throughout to emphasize the analogy to supernova remnant models such as Ref. 27 (although technically, we are discussing a magnetar wind nebula and not a pulsar wind nebula). In the PWNe model, the X-ray plateau is caused by a magnetized bubble of electrons fuelled by the relativistic wind of the magnetar. We summarize the model presented in Ref. 25 in Sec. 2 and explore associated synchrotron light curves in Sec. 3. In Sec. 4, we quote results from Ref. 26 inferring parameters of the central engine using instantaneous spectra. We conclude in Sec. 5. This paper borrows substantially from the form and content of Ref. 25 and Ref. 26; among other things, it includes Fig. 3 from Ref. 26 unchanged.

2. Pulsar Wind Model

If a neutron star survives the sGRB, it evolves under the same physics that dictates the early evolution of PWNe. Classic one-zone models for PWNe provide an analytic estimate of the synchrotron luminosity of the remnant without specifying its detailed geometry.[27,28] In Ref. 25, an analogous model is developed and applied to X-ray plateaux. In this scenario, the sGRB heralds the birth of a rapidly-rotating neutron star with a dipole, magnetar-strength external magnetic field with angular frequency $\Omega(t)$, initial angular frequency Ω_0, and polar surface magnetic field strength B_0. Simultaneously, an isotropic, relativistic blast wave (described by the Blandford-McKee self-similar solution[29]) detonates and sweeps up the surrounding interstellar medium into a spherical shell expanding at $v_{\rm s}$.

The neutron star spins down under pure magnetic-dipole braking and radiates energy at a rate

$$L_{\rm sd}(t) = L_0 \left(1 + \frac{t}{\tau}\right)^{-2} \tag{1}$$

where $L_0 \propto B_0^2 \Omega_0^4$ is the initial spin-down luminosity and $\tau \propto B_0^{-2} \Omega_0^{-2}$ is the characteristic spin-down timescale. Through analogy with classic models of PWNe, we assume the energy is extracted as a magnetized, relativistic electron-positron wind with velocity $v_{\rm w} \gg v_{\rm s}$. The ratio of the Poynting flux to kinetic-energy flux in the wind, σ, is expected to be $\gtrsim 1$ (i.e. the wind is Poynting-flux dominated) near the star and $\ll 1$ (i.e. kinetic-energy flux dominated) at distances far beyond the co-rotation radius $c/\Omega(t)$ ($\sigma \approx 10^{-3}$ for the Crab[30–32]).

As in PWNe, a reverse shock forms at radius $r_{\rm rs}$ where the ram pressure of the wind $P_{\rm ram}(r)$ balances the external static pressure $P_{\rm stat}(r)$.[28] As the shock propagates into the wind, the electrons are shocked into a power-law energy distribution $\propto E^{-a}$. Reference 25 models the electron population with a spherical, homogeneous

bubble characterized by its properties at $r_{\rm rs}$ (i.e. at $r = \dot{r}_{\rm rs}t$ for constant $\dot{r}_{\rm rs}$). Under this assumption, the fresh electrons in the energy range $[E_{-0}, E_{+0}]$ are injected into the bubble at a rate

$$\dot{N}_{\rm inj} = \frac{L_{\rm sd}(t)E^{-a}(a-2)}{(1+\sigma)\left(E_{-0}^{2-a} - E_{+0}^{2-a}\right)}. \tag{2}$$

The bubble expands at a rate $\dot{r}_{\rm rs}$, so electrons cool according to adiabatic expansion at a rate

$$-\left.\frac{{\rm d}E}{{\rm d}t}\right|_{\rm adiabatic} = \frac{E}{t}. \tag{3}$$

Given a magnetic field in the bubble $B(t)$, the electrons also cool via synchrotron radiation at a rate

$$-\left.\frac{{\rm d}E}{{\rm d}t}\right|_{\rm synch} = c_s E^2 B(t)^2 \tag{4}$$

where $c_s = \mu_0 e^4/9\pi c^3 m_e^4$. Two magnetic field configurations are presented in Ref. 25. In the first case, henceforth model A, $B(t)$ is taken to be the far-field extension of the stellar dipole field at radial distance $\dot{r}_{\rm rs}t$ (see Eq. (9) in Ref. 25). In the second case, henceforth model B, $B(t)$ is taken to be an arbitrary, constant B. Model B serves two purposes. One, it allows the PWNe model (including magnetar spin down, electron injection, and cooling mechanisms) to be assessed in general terms, even if the specific magnetic field structure in model A is a poor approximation to reality. Two, model B decouples $B(t)$ and the central engine parameters B_0 and Ω_0 (which also govern $\dot{N}_{\rm inj}$), allowing for the possibility that $B(t)$ may be affected by components of the post-merger environment other than the central engine.

The number density of electrons per unit energy $N(E,t)$ can be found via the time-dependent, inhomogeneous partial-differential equation

$$\frac{\partial N(E,t)}{\partial t} = \frac{\partial}{\partial E}\left[\left.\frac{{\rm d}E}{{\rm d}t}\right|_{\rm adiabatic} + \left.\frac{{\rm d}E}{{\rm d}t}\right|_{\rm synch}\right] + \dot{N}_{\rm inj} \tag{5}$$

where adiabatic cooling is defined as in Eq. (3), synchrotron cooling is defined in Eq. (4), and electron injection $\dot{N}_{\rm inj}$ is defined in Eq. (2). Two analytic Green's function solutions (corresponding to models A and B) are found in Ref. 25 and integrated to obtain $N(E,t)$.

3. Light Curves

In this section, we highlight results from Ref. 25 comparing the X-ray synchrotron light curves predicted by the model in Sec. 2 to those observed by the Neil Gehrels *Swift* telescope.[33–35]

The synchrotron spectrum of an electron with energy $E_{\rm c}$ may be approximated as a Dirac-delta function centred at the characteristic synchrotron frequency

$\nu_c \propto E_c^2 B(t)$. The source frame synchrotron luminosity in the range $[\nu_{\rm min}, \nu_{\rm max}]$ is

$$L_{\nu_{\rm min}-\nu_{\rm max}} = \int_{\nu_{\rm min}}^{\nu_{\rm max}} {\rm d}\nu c_s B(t)^2 E_c^2 N(E_c, t). \quad (6)$$

We define $L_{\rm X}$ as the source-frame luminosity in the band corresponding to the 0.3 keV – 10 keV band in the lab frame, i.e. the luminosity observable by *Swift*.

We use GRB 130603B as a representative example of X-ray plateaux following sGRBs because it has a known redshift[36] and has been inferred to spin down under magnetic-dipole braking.[19] We display the observed light curve in Fig. 1 and overplot with a projected light curve from the PWNe model. Black crosses are observations of GRB 130603B by *Swift* and corrected to the source frame. The blue curve is the X-ray light curve predicted by the PWNe model with $B = 10\,{\rm G}$, $B_0 = 4 \times 10^{15}\,{\rm G}$, $\Omega_0/2\pi = 140\,{\rm Hz}$, and $a = 2.8$. The parameters are chosen by hand, guided by results from Ref. 26. The observed and predicted light curve are in broad agreement.

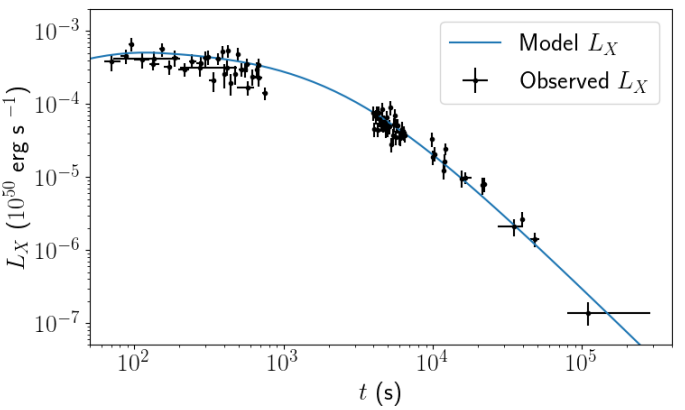


Fig. 1. X-ray luminosity in the source frame $L_{\rm X}\,(10^{50}{\rm erg\,s^{-1}})$ versus time $t\,({\rm s})$. Here "X-ray" refers to the 0.3 keV – 10 keV band observable by *Swift* corrected to the source frame. Black crosses are observations of GRB 130603B by *Swift* and corrected to the source frame. The blue curve is the X-ray light curve predicted by the PWNe model with $B = 10\,{\rm G}$, $B_0 = 4 \times 10^{15}\,{\rm G}$, $\Omega_0/2\pi = 140\,{\rm Hz}$, and $a = 2.8$. Parameters chosen by hand, guided by results from Ref. 26.

In some cases, such as GRB 090515, the plateau ends abruptly. In the context of the central engine model, this has been interpreted as the neutron star collapsing to a black hole and terminating energy injection[18] (we note, however, recent population-based arguments against this hypothesis[37]). In the PWNe model, the collapse of the neutron star halts injection only but does not remove the existing population of electrons, so the afterglow persists for some time, given approximately by the synchrotron lifetime $t_{\rm synch} \sim 2 \times 10^{-9}\,[B/(1\,{\rm G})]^{-3/2}\,[\nu_c/(1\,{\rm keV})]^{-1/2}$ s. In the X-ray band, and for the parameters in Figure 1, $t_{\rm synch}$ amounts to nanoseconds.

3.1. *Luminosity-time correlation*

The Dainotti correlation connects the plateau luminosity $L_{\rm p}$ and plateau duration $t_{\rm p}$ in GRB plateaux which states that brighter plateaux have shorter durations than dimmer plateaux.[3,38,39] This correlation has been studied extensively for LGRBs under a variety of assumptions[40,41] (see Ref. 42 for a thorough review on the correlation and its uses).

The PWNe model presented in Ref. 25 naturally reproduces the observed correlation for reasonable definitions of $L_{\rm p}$ and $t_{\rm p}$. One reasonable choice is to define the X-ray plateau duration as $t_{\rm p} = \tau$ and the corresponding luminosity as $L_{\rm p} = L_{\rm X}(t = \tau)$. The light blue region in Fig. 2 indicates the full range of ($L_{\rm p}$, $t_{\rm p}$) pairs generated by the PWNe model for $B = 5.0 \times 10^{-1}\,{\rm G}$, $E_{-0} = 2.5 \times 10^{-2}\,{\rm erg}$, $10^{12} \leq B_0/(1\,{\rm G}) \leq 10^{16}$, and $\Omega_0/2\pi \leq 10^3\,{\rm Hz}$. The black line and grey-shaded region corresponds to the best-fit correlation from Ref. 18 based on a sample of 159 sGRBs and LGRBs. The observed correlation falls entirely within the range permitted by the PWNe model.

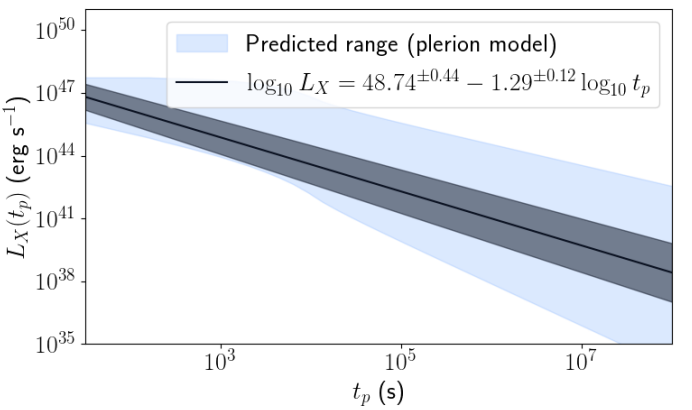


Fig. 2. X-ray plateau luminosity in the source frame $L_{\rm p} = L_{\rm X}(t_{\rm p})\,({\rm erg\,s^{-1}})$ versus X-ray plateau duration $t_{\rm p}$ (s). The black line and grey-shaded region correspond to the best-fit correlation from Ref. 18 from a sample of 159 GRBs. The light blue shaded region corresponds to the range of ($L_{\rm p}$, $t_{\rm p}$) pairs generated by the PWNe model for $B = 5.0 \times 10^{-1}\,{\rm G}$, $10^{12} \leq B_0/(1\,{\rm G}) \leq 10^{16}$, and $\Omega_0/2\pi \leq 10^3\,{\rm Hz}$. Model parameters were chosen based on results presented in Ref. 26.

4. Inferring Parameters of the Central Engine

In this section, we review results presented in Ref. 26 inferring B_0, Ω_0, $E_{\pm 0}$, a, and (for model B) B using the X-ray afterglow for six sGRBs of known redshift with an X-ray plateau. We reproduce here the necessary physics to interpret the results and leave the details of the analysis to Ref. 26.

The source-frame synchrotron spectrum at time t is

$$F_\nu(t,\nu) = N(E,t)\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\mathrm{synch}}\frac{\mathrm{d}E}{\mathrm{d}\nu}. \tag{7}$$

The synchrotron cooling rate and the characteristic synchrotron frequency depend on the magnetic field in the bubble, which is determined by B_0 and Ω_0 in model A and by B in model B. The dependence on $N(E,t)$ indicates F_ν is dependent on B, B_0, Ω_0, $E_{\pm0}$, and a, so a single point-in-time spectrum is sufficient to infer posteriors on each parameter. For a full discussion of the results and method, see Sec. 4 in Ref. 26. For both models, one has a neutron star with an approximately millisecond spin period ($\Omega_0/2\pi \lesssim 10^3\,\mathrm{Hz}$) and a magnetar-strength poloidal magnetic field ($B_0 \sim 10^{15}\,\mathrm{G}$) i.e. a millisecond magnetar. The correlations between B_0 and Ω_0 reflect their relation to the spin-down luminosity $L_0 \propto B_0^2\Omega_0^4$ and $\tau \propto B_0^{-2}\Omega_0^{-2}$. For model B, the posteriors on B cover the range $10^{-1} \lesssim B/(1\,\mathrm{G}) \lesssim 1$. This is much stronger than the magnetic field in the interstellar medium ($\sim 10^{-6}\,\mathrm{G}$) but smaller than the expected field advected outwards from the central object by the relativistic outflow (i.e. the magnetic field in model A). For both models, $E_{-0} \sim 10^{-3}\,\mathrm{erg}$, which is consistent with a population of electrons accelerated in the magnetar's magnetosphere and injected into the magnetar wind with a radiation-reaction-limited Lorentz factor.

4.1. *Spectral evolution*

In this section, we review briefly results presented in Ref. 26 using the spectrum of GRB 130603B at four separate epochs to infer B_0, Ω_0, $E_{\pm0}$, a, and B for each epoch. We label each epoch t_i with corresponding magnetic field B_i for $i \in [1,4]$ and use log uniform priors such that $10^{-6} < B_i/(1\,\mathrm{G}) < 10^6$. For each epoch with flux data F_{ν_i}, flux uncertainty σ_i, and model flux $F_\nu(t_i,\nu)$ given by Eq. (7), we define a Gaussian likelihood $P_i \propto \exp\left\{\left[F_{\nu_i} - F_\nu\left(t_i,\nu\right)\right]^2 / \left(2\sigma_i^2\right)\right\}$ such that the priors on B_0, Ω_0, $E_{\pm0}$ and a are the same at each epoch. The analysis is performed on the joint likelihood $\Pi_i P_i$.

The median values of the posterior describe a millisecond magnetar with $B_0 \approx 2\times10^{15}\,\mathrm{G}$ and $\Omega_0/2\pi \approx 600\,\mathrm{Hz}$, supplying the remnant with relativistic electrons with $a \approx 1.9$, $E_{-0} \approx 3.2\times10^{-5}\,\mathrm{erg}$, and $E_{+0} \approx 1\,\mathrm{erg}$. The median magnetic fields B_i suggest the field drops at an average rate of $0.04\,\mathrm{G\,s^{-1}}$ from $B_1 = 2\times10^2\,\mathrm{G}$ at $t_1 = 643\,\mathrm{s}$ to $B_4 = 5\times10^{-1}\,\mathrm{G}$ at $t_4 = 5735\,\mathrm{s}$. This is both slower and weaker than what is expected for the field in the termination shock of the wind in model A, which scales roughly as $B(t) \propto B_0 t^{-2}$ for $t \gtrsim \tau$, if the wind expands at a constant, relativistic speed (as in Ref. 25). The scale and temporal evolution of the B_i indicate the time-dependent magnetic field in the bubble requires revision. Several mechanisms for this are discussed in Ref. 26, including varying the expansion rate of the bubble or choosing an entirely separate magnetic field configuration.

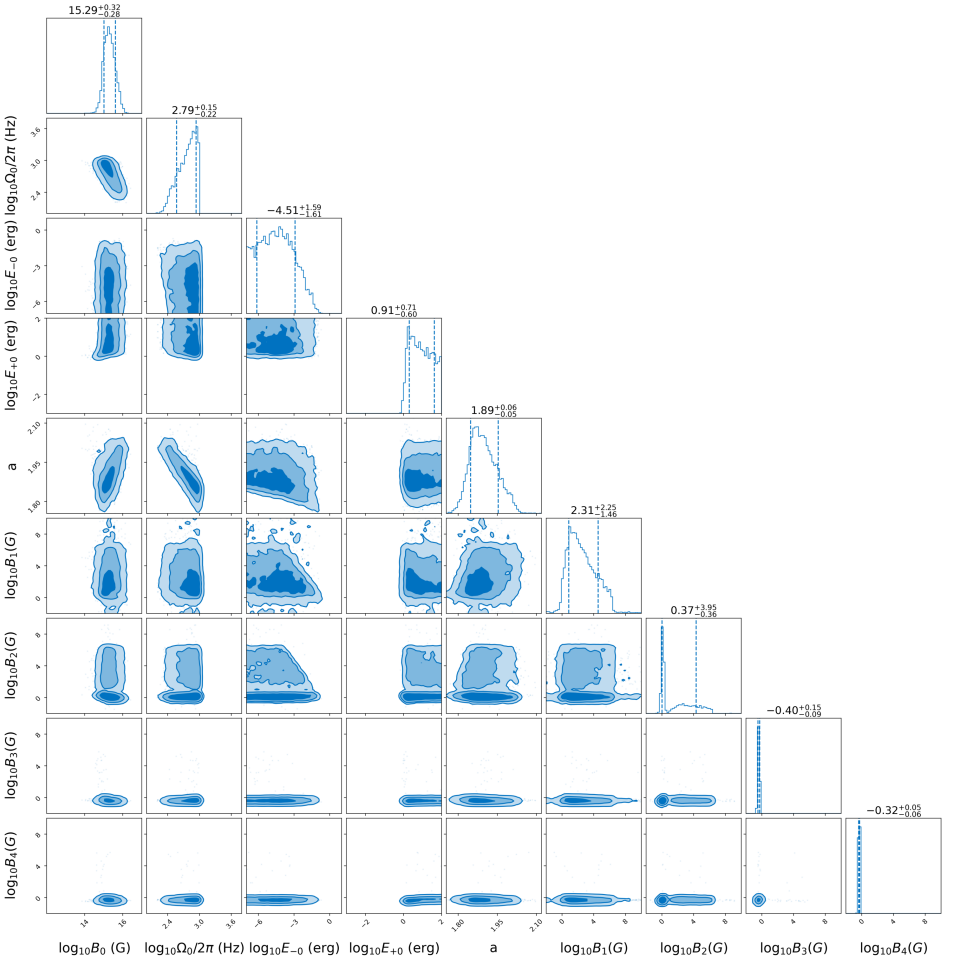


Fig. 3. Corner plot showing the posterior distribution obtained for four instantaneous spectra for GRB130603B for the parameters $\log_{10} B_0$ (G), $\log_{10} \Omega_0/2\pi$ (Hz), $\log_{10} E_{\pm 0}$ (erg), a, and $\log_{10} B_i$ $(1 \leq i \leq 4)$ (G). Figure reproduced from Ref. 26

5. Conclusion

A millisecond magnetar engine based on classic models of PWNe can explain some of the observed features of canonical sGRB afterglows. Plateau light curves are broadly consistent with the synchrotron output of an electron population $N(E, t)$ evolving under ongoing, power-law injection and adiabatic and synchrotron cooling. Such a model may be referred to as a plerion model (as in Ref. 25 and Ref. 26) or, as a synonym, a PWNe model (as we do in this work). The model is able to reproduce both the shape of the afterglow and correlations L_p–t_p. Furthermore, the instantaneous spectra can be used to infer B, B_0, Ω_0, $E_{\pm 0}$, and a within the context of the model. However, the model summarized here (originally in Ref. 25)

is highly idealized and does not include the important effects of merger ejecta on the evolution of the remnant (considered in e.g. Refs. 20–22). Also, as presented in Ref. 26, the temporal evolution of the spectrum implies a magnetic field that is inconsistent with both a dipole field (model A) and a constant field (model B). Future modeling improving on the work in Ref. 25 and integrating it into existing models for sGRB remnants should ultimately be compared to a broad sample of sGRBs.

Acknowledgments

This work makes use of data supplied by the UK Swift Science Data Centre at the University of Leicester. Parts of this research were conducted by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), through Project Number CE170100004. The work is also supported by an Australian Research Council Discovery Project grant (DP170103625).

References

1. B. Zhang, Y. Z. Fan, J. Dyks, S. Kobayashi, P. Mészáros, D. N. Burrows, J. A. Nousek and N. Gehrels, Physical Processes Shaping Gamma-Ray Burst X-Ray Afterglow Light Curves: Theoretical Implications from the Swift X-Ray Telescope Observations, *ApJ* **642**, 354 (May 2006).
2. J. A. Nousek, C. Kouveliotou, D. Grupe, K. L. Page, J. Granot, E. Ramirez-Ruiz, S. K. Patel, D. N. Burrows, V. Mangano, S. Barthelmy, A. P. Beardmore, S. Campana, M. Capalbi, G. Chincarini, G. Cusumano, A. D. Falcone, N. Gehrels, P. Giommi, M. R. Goad, O. Godet, C. P. Hurkett, J. A. Kennea, A. Moretti, P. T. O'Brien, J. P. Osborne, P. Romano, G. Tagliaferri and A. A. Wells, Evidence for a Canonical Gamma-Ray Burst Afterglow Light Curve in the Swift XRT Data, *ApJ* **642**, 389 (May 2006).
3. M. G. Dainotti, V. F. Cardone and S. Capozziello, A time-luminosity correlation for γ-ray bursts in the X-rays, *Monthly Notices of the Royal Astronomical Society* **391**, L79 (November 2008).
4. M. G. Dainotti, R. Willingale, S. Capozziello, V. Fabrizio Cardone and M. Ostrowski, Discovery of a Tight Correlation for Gamma-ray Burst Afterglows with "Canonical" Light Curves, *ApJL* **722**, L215 (October 2010).
5. A. Rowlinson, B. P. Gompertz, M. Dainotti, P. T. O'Brien, R. A. M. J. Wijers and A. J. van der Horst, Constraining properties of GRB magnetar central engines using the observed plateau luminosity and duration correlation, *Monthly Notices of the Royal Astronomical Society* **443**, 1779 (September 2014).
6. N. Rea, M. Gullón, J. A. Pons, R. Perna, M. G. Dainotti, J. A. Miralles and D. F. Torres, Constraining the grb-magnetar model by means of the galactic pulsar population, *The Astrophysical Journal* **813**, p. 92 (2015).
7. B. P. Abbott *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Physical Review Letters* **119**, p. 161101 (2017).
8. B. P. Abbott *et al.*, Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *The Astrophysical Journal Letters* **848** (2017).

9. M. J. Rees and P. Mészáros, Refreshed Shocks and Afterglow Longevity in Gamma-Ray Bursts, *ApJL* **496**, L1 (March 1998).
10. M. G. Dainotti, A. Ł. Lenart, N. Fraija, S. Nagataki, D. C. Warren, B. De Simone, G. Srinivasaragavan and A. Mata, Closure relations during the plateau emission of Swift GRBs and the fundamental plane, *Publications of the Astronomical Society of Japan* **73**, 970 (August 2021).
11. T. Piran, Gamma-ray bursts and the fireball model, *Physics Reports* **314**, 575 (June 1999).
12. P. Kumar, R. Narayan and J. L. Johnson, Properties of Gamma-Ray Burst Progenitor Stars, *Science* **321**, p. 376 (July 2008).
13. Z. G. Dai and T. Lu, Gamma-ray burst afterglows and evolution of postburst fireballs with energy injection from strongly magnetic millisecond pulsars, *Astronomy and Astrophysics* **333**, L87 (May 1998).
14. B. Zhang and P. Mészáros, Gamma-Ray Burst Afterglow with Continuous Energy Injection: Signature of a Highly Magnetized Millisecond Pulsar, *The Astrophysical Journal* **552**, L35 (May 2001).
15. B. Zhang and P. Mészáros, Gamma-Ray Burst Afterglow with Continuous Energy Injection: Signature of a Highly Magnetized Millisecond Pulsar, *ApJ Letters* **552**, L35 (May 2001).
16. Y. Fan and D. Wei, Late internal-shock model for bright x-ray flares in gamma-ray burst afterglows and grb 011121, *Monthly Notices of the Royal Astronomical Society: Letters* **364**, L42 (2005).
17. B. P. Gompertz, P. T. O'Brien, G. A. Wynn and A. Rowlinson, Can magnetar spin-down power extended emission in some short GRBs?, *Monthly Notices of the Royal Astronomical Society* **431**, 1745 (May 2013).
18. A. Rowlinson, P. T. O'Brien, B. D. Metzger, N. R. Tanvir and A. J. Levan, Signatures of magnetar central engines in short GRB light curves, *Monthly Notices of the Royal Astronomical Society* **430**, 1061 (April 2013).
19. P. D. Lasky, C. Leris, A. Rowlinson and K. Glampedakis, The braking index of millisecond magnetars, *The Astrophysical Journal Letters* **843**, p. L1 (2017).
20. B. D. Metzger and A. L. Piro, Optical and X-ray emission from stable millisecond magnetars formed from the merger of binary neutron stars, *Montly Notices of the Royal Astronomical Society* **439**, 3916 (April 2014).
21. D. M. Siegel and R. Ciolfi, Electromagnetic Emission from Long-lived Binary Neutron Star Merger Remnants. I. Formulation of the Problem, *ApJ* **819**, p. 14 (March 2016).
22. Y.-W. Yu, B. Zhang and H. Gao, Bright "Merger-nova" from the Remnant of a Neutron Star Binary Merger: A Signature of a Newly Born, Massive, Millisecond Magnetar, *ApJL* **776**, p. L40 (October 2013).
23. G. Stratta, M. G. Dainotti, S. Dall'Osso, X. Hernandez and G. De Cesare, On the Magnetar Origin of the GRBs Presenting X-Ray Afterglow Plateaus, *ApJ* **869**, p. 155 (December 2018).
24. N. Sarin, P. D. Lasky and G. Ashton, Interpreting the X-ray afterglows of gamma-ray bursts with radiative losses and millisecond magnetars, *Monthly Notices of the Royal Astronomical Society* **499**, 5986 (December 2020).
25. L. C. Strang and A. Melatos, Plerion model of the X-ray plateau in short gamma-ray bursts, *arXiv e-prints*, p. arXiv:1906.02877 (June 2019).
26. L. C. Strang, A. Melatos, N. Sarin and P. D. Lasky, Inferring properties of neutron stars born in short gamma-ray bursts with a plerion-like X-ray plateau, *Monthly Notices of the Royal Astronomical Society* **507**, 2843 (October 2021).

27. F. Pacini and M. Salvati, On the Evolution of Supernova Remnants. Evolution of the Magnetic Field, Particles, Content, and Luminosity, *The Astrophysical Journal* **186**, 249 (November 1973).
28. M. J. Rees and J. E. Gunn, The origin of the magnetic field and relativistic particles in the Crab Nebula, *Monthly Notices of the Royal Astronomical Society* **167**, 1 (April 1974).
29. R. D. Blandford and C. F. McKee, Fluid dynamics of relativistic blast waves, *Physics of Fluids* **19**, 1130 (August 1976).
30. C. F. Kennel and F. V. Coroniti, Confinement of the Crab pulsar's wind by its supernova remnant, *The Astrophysical Journal* **283**, 694 (August 1984).
31. A. Melatos and D. B. Melrose, Energy transport in a rotation-modulated pulsar wind, *Monthly Notices of the Royal Astronomical Society* **279**, 1168 (April 1996).
32. S. V. Bogovalov, D. V. Khangulyan, A. V. Koldoba, G. V. Ustyugova and F. A. Aharonian, Modelling interaction of relativistic and non-relativistic winds in binary system PSR B1259-63/SS2883 - I. Hydrodynamical limit, *Monthly Notices of the Royal Astronomical Society* **387**, 63 (June 2008).
33. N. Gehrels, G. Chincarini, P. Giommi, K. O. Mason, J. A. Nousek, A. A. Wells, N. E. White, S. D. Barthelmy, D. N. Burrows, L. R. Cominsky, K. C. Hurley, F. E. Marshall, P. Mészáros, P. W. A. Roming, L. Angelini, L. M. Barbier, T. Belloni, S. Campana, P. A. Caraveo, M. M. Chester, O. Citterio, T. L. Cline, M. S. Cropper, J. R. Cummings, A. J. Dean, E. D. Feigelson, E. E. Fenimore, D. A. Frail, A. S. Fruchter, G. P. Garmire, K. Gendreau, G. Ghisellini, J. Greiner, J. E. Hill, S. D. Hunsberger, H. A. Krimm, S. R. Kulkarni, P. Kumar, F. Lebrun, N. M. Lloyd-Ronning, C. B. Markwardt, B. J. Mattson, R. F. Mushotzky, J. P. Norris, J. Osborne, B. Paczynski, D. M. Palmer, H. S. Park, A. M. Parsons, J. Paul, M. J. Rees, C. S. Reynolds, J. E. Rhoads, T. P. Sasseen, B. E. Schaefer, A. T. Short, A. P. Smale, I. A. Smith, L. Stella, G. Tagliaferri, T. Takahashi, M. Tashiro, L. K. Townsley, J. Tueller, M. J. L. Turner, M. Vietri, W. Voges, M. J. Ward, R. Willingale, F. M. Zerbi and W. W. Zhang, The Swift Gamma-Ray Burst Mission, *The Astrophysical Journal* **611**, 1005 (August 2004).
34. P. A. Evans, A. P. Beardmore, K. L. Page, L. G. Tyler, J. P. Osborne, M. R. Goad, P. T. O'Brien, L. Vetere, J. Racusin, D. Morris, D. N. Burrows, M. Capalbi, M. Perri, N. Gehrels and P. Romano, An online repository of Swift/XRT light curves of γ-ray bursts, *Astronomy and Astrophysics* **469**, 379 (July 2007).
35. P. A. Evans, A. P. Beardmore, K. L. Page, J. P. Osborne, P. T. O'Brien, R. Willingale, R. L. C. Starling, D. N. Burrows, O. Godet, L. Vetere, J. Racusin, M. R. Goad, K. Wiersema, L. Angelini, M. Capalbi, G. Chincarini, N. Gehrels, J. A. Kennea, R. Margutti, D. C. Morris, C. J. Mountford, C. Pagani, M. Perri, P. Romano and N. Tanvir, Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs, *Monthly Notices of the Royal Astronomical Society* **397**, 1177 (August 2009).
36. A. Melandri, M. de Pasquale, S. D. Barthelmy, D. N. Burrows, M. H. Siegel and N. Gehrels, Swift Observation of GRB130603B, *GCN Report* **442** (2013).
37. P. Beniamini and W. Lu, Survival times of supramassive neutron stars resulting from binary neutron star mergers, *arXiv e-prints*, p. arXiv:2104.01181 (April 2021).
38. M. G. Dainotti, R. Willingale, S. Capozziello, V. Fabrizio Cardone and M. Ostrowski, Discovery of a Tight Correlation for Gamma-ray Burst Afterglows with "Canonical" Light Curves, *The Astrophysical Journal* **722**, L215 (October 2010).
39. J. Sultana, D. Kazanas and K. Fukumura, Luminosity Correlations for Gamma-Ray Bursts and Implications for Their Prompt and Afterglow Emission Mechanisms, *The Astrophysical Journal* **758**, p. 32 (October 2012).

40. M. G. Dainotti, S. Postnikov, X. Hernandez and M. Ostrowski, A Fundamental Plane for Long Gamma-Ray Bursts with X-Ray Plateaus, *The Astrophysical Journal* **825**, p. L20 (July 2016).
41. M. G. Dainotti, A. D. Boria and J. F. Arratia, Gamma-ray bursts with afterglow plateau phases associated with supernovae, in *Fourteenth Marcel Grossmann Meeting - MG14*, eds. M. Bianchi, R. T. Jansen and R. Ruffini, January 2018.
42. M. Dainotti, *Gamma-ray Burst Correlations; Current status and open questions* 2019.

Searching for strange quark planets

Xu Wang[1,*], Yong-Feng Huang[1,2,†] and Bing Li[3,4]

[1] *School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China*

[2] *Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, People's Republic of China*

[3] *Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, People's Republic of China*

[4] *Particle Astrophysics Division, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, People's Republic of China*

[*] *E-mail: wx_tba@163.com*

[†] *E-mail: hyf@nju.edu.cn*

Strange quark matter (SQM) may be the true ground state of matter. According to this SQM hypothesis, the observed neutron stars actually should all be strange quark stars. But distinguishing between neutron stars and strange quark stars by means of observations is extremely difficult. It is interesting to note that under the SQM hypothesis, less massive objects such as strange quark planets and strange dwarfs can also stably exist. The extremely high density and small radius of strange quark planets give us some new perspectives to identify SQM objects and to test the SQM hypothesis. First, the tidal disruption radius of strange quark planets is much smaller than normal planets, so, very close-in exoplanets can be safely identified as candidates of SQM objects. Second, gravitational waves (GW) from mergers of strange quark star-strange quark planet systems are strong enough to be detected by ground-based GW detectors. As a result, GW observation will be a powerful tool to probe SQM stars. At the same time, the tidal deformability of SQM planets can be measured to further strengthen the result.

Keywords: Strange quark stars, Exoplanets, Gravitational waves, Neutron stars

1. Introduction

The central engine of gamma-ray bursts may be compact objects such as neutron stars. However, our knowledge about matter under extreme densities is still quite poor so that the internal composition and structure of neutron stars are largely uncertain to us.[1] It has been argued that the energy per baryon of strange quark matter (SQM), which is composed of three favors of quarks (up, down and strange quarks), could be less than normal hadronic matter. As a result, SQM may be the true ground state of matter.[2,3] If this hypothesis is true, then it is even possible that all the so called "neutron stars" observed in the Universe should actually be strange quark stars.[4,5]

However, it is difficult to distinguish between strange quark stars and neutron stars through current astronomical observations. For a long time, people have tried to find the difference between neutron stars and strange quark stars in terms of mass-radius relationship, cooling rate, the minimum spin period and gravitational

wave (GW) features. But for a typical 1.4 $M_\odot$ compact object, the radius difference between a neutron star and a quark star is too small to be detected through current observational technology. Recently, Geng et al pointed out that fast radio busts may originate from the collapses of the crust of strange quark stars.[6,7] It provides an novel visual angle on these interesting objects.

It is interesting to note that, under the SQM hypothesis, quark matter is bounded by strong interaction but not gravity.[5] So, SQM can even exist stably in the form of small chunks in the universe. It implies that planets composed of strange quark matter can also exist stably.[8,9] Strange quark planets are very different from normal planets. They have a much higher mean density and a much smaller radius, which provide us with some effective new methods to test the SQM hypothesis. In our previous studies, we have suggested some new methods to identify strange quark planet.[10–15] In this article, we will summarize the results.

2. Searching for strange quark planets among close-in exoplanets

Planets cannot be too close to their host stars, otherwise they will be tidally broken up by the strong tidal force of their hosts. We can use the tidal disruption radius r_{td} to describe the shortest possible separation between a planet and its host, which can be analytically expressed as[16]

$$r_{td} \approx \left(\frac{6M}{\pi\bar{\rho}}\right)^{\frac{1}{3}}, \tag{1}$$

where M is the mass of the host star and $\bar{\rho}$ is the mean density of the planet. For the convenience of calculation, this equation can be further written as,

$$r_{td} \approx 2.37\times10^{6}\left(\frac{M}{1.4\mathrm{M}_\odot}\right)^{\frac{1}{3}}\times\left(\frac{\bar{\rho}}{4\times10^{14}\mathrm{gcm}^{-3}}\right)^{-\frac{1}{3}}\mathrm{cm}. \tag{2}$$

Suppose a planet orbits around a host star which has a typical mass of 1.4 $M_\odot$. For a strange quark planet, its mean density can be as high as 4×10^{14} g cm^{-3}. With such a high density, the r_{td} of the strange quark planet will be less than 2.37×10^{6} cm, which is only about twice the radius of a pulsar.[10,11] But for normal matter planets, the density is on the order of ~ 8 g cm^{-3}, and the corresponding r_{td} is generally larger than 8.7×10^{10} cm. Even if we take the mean density as 30 g cm^{-3}, which is already a very high value for normal matter, the derived r_{td} will still be larger than 5.6×10^{10} cm.[10,11] From these simple calculations, we argued that if the orbital radius of a planet is found to be less than 5.6×10^{10} cm, then the planet should be a strange quark planet.

An extremely close-in strange quark planet cannot be observed directly by imaging method. On the contrary, it can be relatively easily detected via pulsar timing observations. According to Kepler's law, the relationship between the orbital radius and the period can be expressed as

$$\frac{a^3}{P_{\mathrm{orb}}^2} \approx \frac{GM}{4\pi^2}, \tag{3}$$

where G is the gravity constant, a is the orbital radius and $P_{\rm orb}$ is the planet's orbital period. For planets with a smaller than $\sim 5.6 \times 10^{10}$ cm, the orbital period will be less than ~ 6100 s. Therefore, it is argued a planet with the orbital period less than 6100 s should be a strange quark planet.[10,11]

Using the above criterion, Kuerban et al. have tried to search for strange quark planets among exoplanets, especially among pulsar planets.[15] According to their results, the short period pulsar planets of PSR J0636 b, PSR J1807-2459A b and PSR 1719-14 b are good candidates of strange quark planets.

3. Searching for strange quark planets through GW observations

Since 2015, gravitational wave observations have opened a new window for astronomy[17] and are also expected to be a new tool for searching for quark stars.[18] Note that it is still quite difficult to distinguish between binary neutron star mergers and binary quark star mergers by gravitational wave observations, because these two kinds of compact stars have marginal difference in radius at the typical mass of 1.4 $M_\odot$.[19,20] However, gravitational waves may bring new opportunities in searching for strange quark planets.

A normal matter planet can not be too close to its host, otherwise it will be tidally disrupted. Consequently, the gravitational wave emissions from normal planetary systems are usually too weak to be detected. But due to the extremely high density and the very close-in orbit of strange quark planets, a strange star-strange planet system can produce very strong gravitational wave emissions, especially at the final stage of the merging process. If such a merger event occurs in the local universe, it would be detectable for the gravitational wave detectors such as Advanced LIGO and the Einstein Telescope.[12]

GW observations can also help diagnose the internal composition and internal structure of compact stars by means of tidal deformability measurements. The first tidal deformability measurement has been obtained for the binary neutron star merger event of GW170817, which gives a new constraint on the equation of state of dense matter.[21,22]

Tidal deformability is a quantity that describes the deformation of a star in a tidal field, which is defined as

$$\lambda = -\frac{Q_{ij}}{E_{ij}}, \tag{4}$$

where Q_{ij} is the induced quadrapole moment of the star and E_{ij} is the tidal field that it resides in (i.e., the gravity field produced by its companion). For convenience, this quality is often written in a dimensionless form, i.e.,

$$\Lambda = \frac{\lambda c^{10}}{G^4 m^5}. \tag{5}$$

Here, c is the speed of light in vacuum and m is the mass of the star.

Generally speaking, a larger tidal deformability means that the star will be relatively easier to be deformed in a tidal field.[22] So, tidal deformability will affect

the evolution of the gravitational wave phase, which could be perceived in GW observations.[13] Comparing the observed tidal deformability with that calculated by solving the Tolman-Oppenheimer-Volkoff equation adopting a particular equation of state (EoS), we can get useful information on the internal structure of compact stars.[23,24]

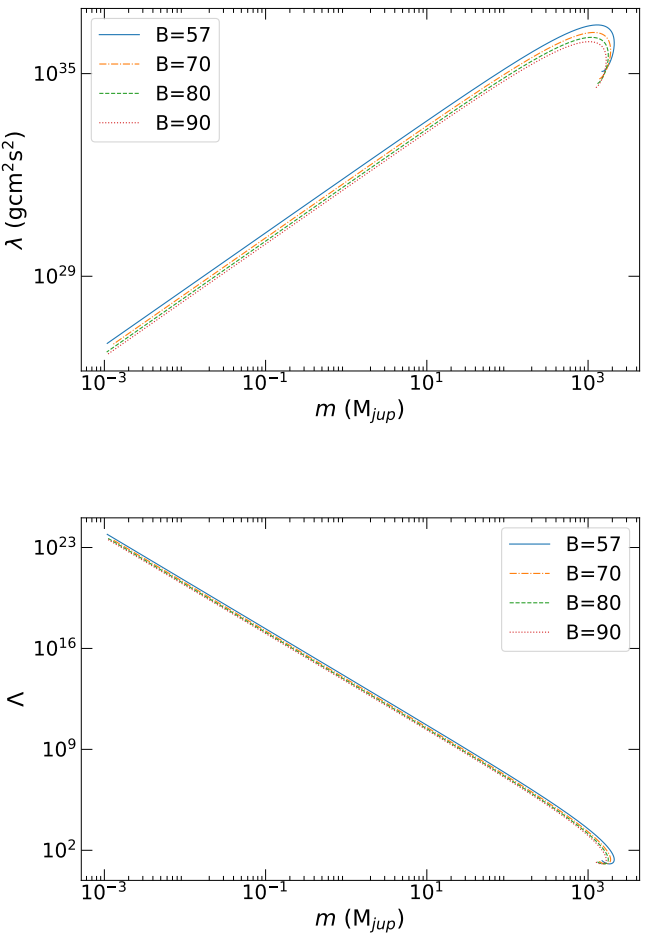


Fig. 1. Tidal deformability (the upper panel) and dimensionless tidal deformability (the lower panel) versus mass for strange quark stars and strange quark planets.[13] Different line styles represent different bag constant, which is marked in the figure in units of MeV/fm^3.

Following the method of Hinderer et al.,[23,24] we have calculated the tidal deformability of strange quark stars, paying special attention on strange dwarfs and strange planets. We engage the bag model for strange quark matter.[4] Our results are shown in Figure 1. We see that as the mass decreases, the value of the deformability also decreases, while the value of the dimensionless tidal deformability keeps

increasing. For a planet with a mass of $10^{-3}M_{\rm jup}$, its tidal deformability is up to $\sim 10^{27}$ g cm^2 s^2 and its dimensionless tidal deformability is up to $\sim 10^{23}$, which are much higher than those of normal matter planets. Therefore, the tidal deformability is a useful parameter for identifying strange quark planets.[13]

4. Summary

Discriminating between neutron stars and strange quark stars is a challenging task. In this article, we try to solve the problem from a novel point of view. We mainly concentrate on strange quark planets. It is suggested that there are basically two methods to search for strange quark planets. First, we can try to find close-in pulsar planets with the orbital period less than 6100 s. Encouragingly, at least three possible strange quark planet candidates have been found. Second, we can identify strange quark planets through gravitational wave observations. It is found that the mergers of strange star-strange planet systems can produce strong gravitational wave bursts, which can potentially be detected by the advanced LIGO and the future Einstein Telescope. The tidal deformability of strange quark dwarfs and strange quark planets is specially calculated. In short, we stress that strange quark planets could be a powerful tool for clarifying the nature of the so called "neutron stars".

Acknowledgements

This work is supported by National SKA Program of China No. 2020SKA0120300, by the National Natural Science Foundation of China (Grant Nos. 11873030, 12041306, U1938201), and by the science research grants from the China Manned Space Project with NO.CMS-CSST-2021-B11.

References

1. J. M. Lattimer and M. Prakash, The Physics of Neutron Stars, *Science* **304**, 536 (April 2004).
2. N. Itoh, Hydrostatic Equilibrium of Hypothetical Quark Stars, *Progress of Theoretical Physics* **44**, 291 (July 1970).
3. G. Baym, C. Pethick and P. Sutherland, The Ground State of Matter at High Densities: Equation of State and Stellar Models, *ApJ* **170**, p. 299 (December 1971).
4. E. Farhi and R. L. Jaffe, Strange matter, *PhRvD* **30**, 2379 (December 1984).
5. E. Witten, Cosmic separation of phases, *PhRvD* **30**, 272 (July 1984).
6. J. Geng, B. Li and Y. Huang, Repeating fast radio bursts from collapses of the crust of a strange star, *The Innovation* **2**, p. 100152 (March 2021).
7. Y. Zhang, J.-J. Geng and Y.-F. Huang, Fast Radio Bursts from the Collapse of Strange Star Crusts, *ApJ* **858**, p. 88 (May 2018).
8. N. K. Glendenning, C. Kettner and F. Weber, Possible New Class of Dense White Dwarfs, *PhRvL* **74**, 3519 (May 1995).
9. N. K. Glendenning, C. Kettner and F. Weber, From Strange Stars to Strange Dwarfs, *ApJ* **450**, p. 253 (September 1995).
10. Y. F. Huang and Y. B. Yu, Searching for Strange Quark Matter Objects in Exoplanets, *ApJ* **848**, p. 115 (October 2017).

11. A. Kuerban, J.-J. Geng, Y.-F. Huang, H.-S. Zong and H. Gong, Close-in Exoplanets as Candidates for Strange Quark Matter Objects, *ApJ* **890**, p. 41 (February 2020).
12. J. J. Geng, Y. F. Huang and T. Lu, Coalescence of Strange-quark Planets with Strange Stars: A New Kind of Source for Gravitational Wave Bursts, *ApJ* **804**, p. 21 (May 2015).
13. X. Wang, A. Kuerban, J.-J. Geng, F. Xu, X.-L. Zhang, B.-J. Zuo, W.-L. Yuan and Y.-F. Huang, Tidal Deformability of Strange Quark Planets and Strange Dwarfs, *arXiv e-prints*, p. arXiv:2105.13899 (May 2021).
14. A. Kuerban, J.-J. Geng and Y.-F. Huang, GW emission from merging strange quark star-strange quark planet systems, in *Xiamen-CUSTIPEN Workshop on the Equation of State of Dense Neutron-Rich Matter in the Era of Gravitational Wave Astronomy*, American Institute of Physics Conference Series Vol. 2127July 2019.
15. A. Kuerban, Y.-F. Huang, J.-J. Geng and H.-S. Zong, Searching for Strange Quark Matter Objects Among White Dwarfs, *arXiv e-prints*, p. arXiv:2012.05748 (December 2020).
16. J. G. Hills, Possible power source of Seyfert galaxies and QSOs, *Natur* **254**, 295 (March 1975).
17. B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari and et al., Tests of General Relativity with GW150914, *PhRvL* **116**, p. 221101 (June 2016).
18. S. Postnikov, M. Prakash and J. M. Lattimer, Tidal Love numbers of neutron and self-bound quark stars, *PhRvD* **82**, p. 024016 (July 2010).
19. A. Bauswein, R. Oechslin and H. T. Janka, Discriminating strange star mergers from neutron star mergers by gravitational-wave measurements, *PhRvD* **81**, p. 024012 (January 2010).
20. P. H. R. S. Moraes and O. D. Miranda, Probing strange stars with advanced gravitational wave detectors, *MNRAS* **445**, L11 (November 2014).
21. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *PhRvL* **119**, p. 161101 (October 2017).
22. A. Guerra Chaves and T. Hinderer, Probing the equation of state of neutron star matter with gravitational waves from binary inspirals in light of GW170817: A brief review, *Journal of Physics G Nuclear Physics* **46**, p. 123002 (December 2019).
23. T. Hinderer, Tidal Love Numbers of Neutron Stars, *ApJ* **677**, 1216 (April 2008).
24. T. Hinderer, B. D. Lackey, R. N. Lang and J. S. Read, Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral, *PhRvD* **81**, p. 123016 (June 2010).

Probe the universe by using Gamma-Ray Bursts with X-ray plateaus

Fan Xu[1,*] and Yong-Feng Huang[1,2,†]

[1] *School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China*

[2] *Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, People's Republic of China*

*E-mail: *carpedieminreality@163.com; †hyf@nju.edu.cn*

Gamma-ray Bursts (GRBs) are promising tools to probe the Universe. Especially, we can get constraints on cosmological parameters with the help of GRB luminosity relations. Recently, a tight L-T-E relation has been found. This relation involves a special subclass of GRBs, i.e., GRBs have X-ray plateaus in their X-ray afterglows. The isotropic γ-ray energy of the prompt GRB ($E_{\gamma,\rm iso}$) are linked with the duration of the plateau phase (T_a) and the X-ray luminosity at the end of the plateau phase (L_X) through this relation. Beside this relation, we find that L_X and T_a also relate to the spectral peak energy ($E_{\rm p}$) and we call this relation the L-T-$E_{\rm p}$ relation. When using these relations to constrain the cosmological parameters, we find the L-T-E relation is insensitive to the cosmological parameters and both of these relations suffer from redshift evolution. We manage to correct the redshift evolution and use the de-evolved L-T-$E_{\rm p}$ relation to constrain the cosmological parameters. We get the best-fit result as $\Omega_m = 0.389^{+0.202}_{-0.141}$ (1σ) for the flat ΛCDM model. Our result is consistent with that from other probes within 1σ confidence level. Joint constraint on cosmological parameters is presented.

Keywords: Cosmology: dark energy – gamma-ray burst: general – stars: neutron

1. Introduction

Gamma-ray Bursts (GRBs) are extremely high-energy events. The multiband observation of the GRB afterglows has helped us to better understand the true nature of GRBs. Also, the observation of the early afterglow has showed that some GRBs have a plateau phase in their X-ray afterglow. The plateau phase is often thought to be connected with extra energy injection, possibly by a millisecond magnetar as the central engine.[1–3] It is interesting to note that these compact magnetars might actually be strange quark stars.[4]

Since GRBs have a cosmological origin, we can use them as a cosmological probe. The most important application is to constrain the cosmological parameters. The method is very similar to using Supernovae Ia (SNe Ia).[5–7] We need solid luminosity relations to help us convert GRBs to standard candles.[8,9] Compared with using SNe Ia, the gamma photon from GRBs does not suffer from the extinction while they travel through the space.[10]

Several interesting relations are found for GRBs with X-ray plateau. Dainotti et al. (2008) first found a nearly anti-relation between the X-ray luminosity at the end time of the plateau (L_X) and the duration of the plateau phase at the rest frame (T_a).[11,12] This relation is also known as the Dainotti relation. The L-T-E relation

found by Xu & Huang links L_X and T_a together with the isotropic γ-ray energy of the prompt GRB ($E_{\gamma,\rm iso}$).[2] Recently, Tang et al. (2019) have revisited this relation with a sample of 174 GRBs and found $L_X \propto T_a^{-1.01} E_{\gamma,\rm iso}^{0.84}$.[13] Meanwhile, a fundamental plane relation was established by Dainotti et al. (2016) which connect L_X, T_a, and $L_{\rm peak}$.[14] Using class-specific GRB samples, these three parameters were very closely connected with each other.[15] Also, a four parameter Combo-relation was recently established which involves α, L_X, T_a, and $E_{\rm p}$, where α is the power-law timing index of the afterglow in its normal decay phase and $E_{\rm p}$ is the spectral peak energy.[16,17] Meanwhile, we have found a new relation between L_X, T_a, and $E_{\rm p}$, which we call the L-T-$E_{\rm p}$ relation.[18] These relations are extensions of the Dainotti relation and reveal the connection between the property of the X-ray plateau with the early prompt emission.

In this paper, we first check the possibility of using the L-T-E relation to constrain the cosmological parameters. We then present the new L-T-$E_{\rm p}$ relation and constraints on the cosmological parameters.

2. Difficulty of using the L-T-E relation to probe the Universe

The L-T-E relation can be written as

$$\log \frac{L_X}{10^{47}{\rm erg/s}} = a + b \log \frac{T_a}{10^3{\rm s}} + c \log \frac{E_{\gamma,\rm iso}}{10^{53}{\rm erg}}. \quad (1)$$

We use the Markov chain Monte Carlo (MCMC) algorithm to get the best-fit with the same sample provided by Tang et al. (2019)[13] and obtain $a = 1.60 \pm 0.06$, $b = -1.01 \pm 0.05$, $c = 0.85 \pm 0.04$ and $\sigma_{\rm ext} = 0.40 \pm 0.03$, i.e., $L_X \propto T_a^{-1.01\pm0.05} E_{\gamma,\rm iso}^{0.85\pm0.04}$. $\sigma_{\rm ext}$ is the the extrinsic scatter parameter arouse from some hidden variables.

When using GRBs to constrain the cosmological parameters, a solid luminosity relation that works at all redshift is needed. Hence, we divide the full sample into four redshift bins and check the evolution of the coefficients. We find that the coefficients of this relation vary a lot among different redshift bins, especially in the coefficient "c". So the L-T-E relation suffers from the redshift evolution. More importantly, because both L_X and $E_{\gamma,\rm iso}$ are related to cosmological parameters through the same dependence on the luminosity distance, so the cosmological effect is canceled out in this relation. In other words, this relation is insensitive to cosmological parameters.

3. Cosmology with the L-T-E$_{\rm p}$ relation

We have found a new relation connects L_X, T_a, and $E_{\rm p}$. We call it the L-T-$E_{\rm p}$ relation and it follows $L_X \propto T_a^{-1.08\pm0.08} E_{\rm p}^{0.76\pm0.14}$ using our sample of 121 GRBs.[18]

Likewise, we check whether this relation suffers from redshift evolution. In fact, this relation indeed evolves with redshifts. We think of two solutions to remove the evolution from this relation. The first solution is trying to find out the redshift

evolution in each parameter (i.e., L_X, T_a, and $E_{\gamma,\rm iso}$) which will be described in Section 3.1. The second one is simply adding an evolution term to account for the redshift evolution in the whole relation and the details of this solution is presented in Section 3.2.

3.1. *Redshift evolution in each parameter*

Following the Efron & Petrosian method,[19] we get the redshift evolution of L_X, T_a, and $E_{\rm p}$ separately. A simple power-law function $(1+z)^k$ is used as the evolution form of each parameter.[20] Our best-fit results are $k_{L_X} = 3.38 \pm 0.62$, $k_{T_a} = -1.54 \pm 0.30$, and $k_{E_{\rm p}} = 0.75 \pm 0.25$.[18]

After correcting for the redshift evolution of each parameter, we can get a new separate de-evolved L-T-$E_{\rm p}$ relation as

$$\log \frac{L_X/(1+z)^{k_{L_X}}}{10^{47}{\rm erg/s}} = a''' + b''' \log \frac{T_a/(1+z)^{k_{T_a}}}{10^3{\rm s}} + c''' \log \frac{E_{\rm p}/(1+z)^{k_{E_{\rm p}}}}{\rm keV}. \quad (2)$$

3.2. *Redshift evolution of the whole relation*

To get the redshift evolution of the whole relation, we simply put the possible evolution in a single power-law term of $(1+z)^{d'}$,[21] and the L-T-$E_{\rm p}$ relation can be re-written as

$$\log \frac{L_X}{10^{47}{\rm erg/s}} = a' + b' \log \frac{T_a}{10^3{\rm s}} + c' \log \frac{E_{\rm p}}{\rm keV} + d' \log(1+z). \quad (3)$$

With the MCMC method, we get the best-fit result of parameter d', which is $d' = 1.78 \pm 0.29$.[18] Then we move the evolution term to the left side of Equation 3 and get a de-evolved L-T-$E_{\rm p}$ relation

$$\log \frac{L_{X,\rm cal}}{10^{47}{\rm erg/s}} = a'' + b'' \log \frac{T_a}{10^3{\rm s}} + c'' \log \frac{E_{\rm p}}{\rm keV}, \quad (4)$$

where $L_{X,\rm cal} = L_X/(1+z)^{d'}$.

3.3. *Constraints on cosmological parameters*

With the two de-evolved relations we get above, we can now safely use them to probe the Universe. We fix $H_0 = 70.0$ km s^{-1} Mpc^{-1} for simplicity. When using GRBs to constrain the cosmological parameters, one may face the so-called "circularity problem".[22] Here we consider two methods to avoid this problem. The first method is the "simultaneously fitting method" which means to fit the relation coefficients together with the cosmological parameters.[23] The second is the "standardize with SNe Ia method" which means to use SNe Ia data set as calibrators to calculate the luminosity distances of GRBs independently from cosmological models.[8]

With the separate de-evolved L-T-$E_{\rm p}$ relation, we perform the simultaneously fitting method and get $\Omega_m = 0.76^{+0.17}_{-0.26}$ for the flat ΛCDM model.

As for the de-evolved L-T-$E_{\rm p}$ relation, we first perform the simultaneously fitting method. The best-fit result is $\Omega_m = 0.71^{+0.20}_{-0.28}$ for the flat ΛCDM model. Then we use the standardize with SNe Ia method with the "Pantheon sample" of SNe Ia.[24] We plot a Hubble diagram for GRBs, as shown in Figure 1. The low-redshift sample is calibrated by the "Pantheon sample" and shows little dispersion, while the large dispersion of the high-redshift GRB distance modulus comes from the uncertainty of the GRB relation. With the calibrated distance modulus of the high-redshift GRBs, we get the best-fit result as $\Omega_m = 0.389^{+0.202}_{-0.141}$ for the flat ΛCDM model. Also we get $\Omega_m = 0.333^{+0.188}_{-0.142}$ and $\Omega_\Lambda = 0.346^{+0.356}_{-0.249}$ for the non-flat ΛCDM model and $\Omega_m = 0.369^{+0.217}_{-0.191}$ and $w = -0.966^{+0.513}_{-0.678}$ for the flat wCDM model.

Then, we compare these results with those from other probes (i.e., SNe Ia, CMB, BAO) and find that they are consistent with each other in 1σ confidence level. Also, together with other probes, we even get a more accurate constraint on the cosmological parameters. We get $\Omega_m = 0.289 \pm 0.008$ and $\Omega_\Lambda = 0.710 \pm 0.006$ for the non-flat ΛCDM model with all the four probes. As for the flat Universe, combining the four probes, we obtain $\Omega_m = 0.295 \pm 0.006$ and $w = -1.015 \pm 0.015$ under the wCDM model and $\Omega_m = 0.291 \pm 0.005$ for the ΛCDM model. Figure 2 shows the fitting results with different probes under the flat ΛCDM model.

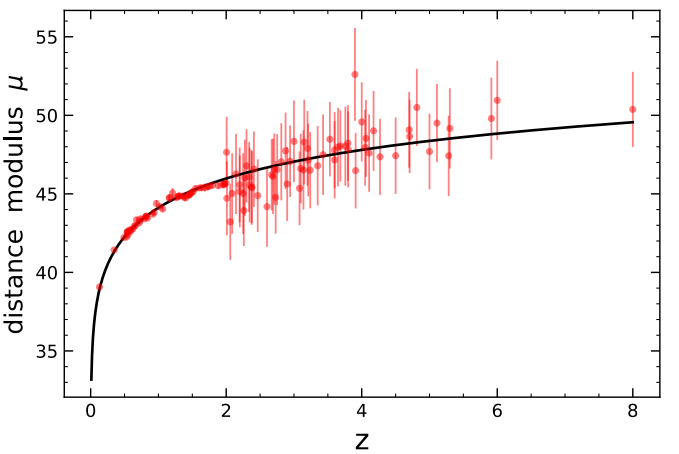


Fig. 1. Calibrated GRB Hubble diagram from the de-evolved L-T-$E_{\rm p}$ relation.[18] Red points represent our GRB sample and the black line is the theoretical distance modulus calculated with $H_0 = 70.0$ km s^{-1} Mpc^{-1} and $\Omega_m = 0.289$ under the flat ΛCDM model.

4. Summary

In our work we find that the previous L-T-E relation is not a good probe for cosmology and it is possible to use the L-T-$E_{\rm p}$ relation as a probe. We want to

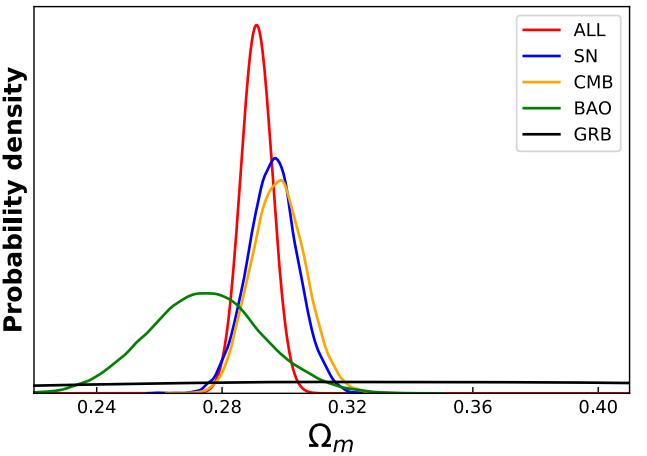


Fig. 2. Under the flat ΛCDM model, we get the probability density on Ω_m with observational data of SNe Ia (blue), CMB (orange), BAO (green), and the de-evolved L-T-$E_{\rm p}$ relation of GRBs (black). The joint constraint from all the four probes is $\Omega_m = 0.291 \pm 0.005$ (red).

stress that possible redshift evolution in GRB relations need to be removed before they can be used as cosmological probe. From our study, we can see that it is hard to get good constraints with GRBs alone, because of the large dispersion in GRB relations. However, with the improvement of GRB observations, we may tighten these relations in the future.

Acknowledgments

This work was supported by National SKA Program of China No. 2020SKA0120300, by the National Natural Science Foundation of China (Grant Nos. 11873030, 12041306, U1938201), and by the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-B11.

References

1. S. Dall'Osso, G. Stratta, D. Guetta, et al., Gamma-ray bursts afterglows with energy injection from a spinning down neutron star, *A&A* **526**, p. A121 (February 2011).
2. a M. Xu and Y. F. Huang, New three-parameter correlation for gamma-ray bursts with a plateau phase in the afterglow, *A&A* **538**, p. A134 (February 2012).
3. L. Li, X.-F. Wu, W.-H. Lei, et al., Constraining the Type of Central Engine of GRBs with Swift Data, *ApJS* **236**, p. 26 (June 2018).
4. J.-J. Geng, B. Li and Y.-F. Huang, 2021, The Innovation, in press, Repeating Fast Radio Bursts from Collapse of Strange Star Crust, *arXiv e-prints*, p. arXiv:2103.04165 (March 2021).
5. M. M. Phillips, The Absolute Magnitudes of Type IA Supernovae, *ApJL* **413**, p. L105 (August 1993).

6. A. G. Riess, A. V. Filippenko, P. Challis, et al., Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, *AJ* **116**, 1009 (September 1998).
7. S. Perlmutter, G. Aldering, G. Goldhaber, et al., Measurements of Ω and Λ from 42 High-Redshift Supernovae, *ApJ* **517**, 565 (June 1999).
8. N. Liang, W. K. Xiao, Y. Liu and S. N. Zhang, A Cosmology-Independent Calibration of Gamma-Ray Burst Luminosity Relations and the Hubble Diagram, *ApJ* **685**, 354 (September 2008).
9. L. Amati, R. D'Agostino, O. Luongo, et al., Addressing the circularity problem in the E_p-E_{iso} correlation of gamma-ray bursts, *MNRAS* **486**, L46 (June 2019).
10. F. Y. Wang, Z. G. Dai and E. W. Liang, Gamma-ray burst cosmology, *NewAR* **67**, 1 (August 2015).
11. M. G. Dainotti, V. F. Cardone and S. Capozziello, A time-luminosity correlation for γ-ray bursts in the X-rays, *MNRAS* **391**, L79 (November 2008).
12. M. Dainotti, V. Petrosian, R. Willingale, et al., Luminosity-time and luminosity-luminosity correlations for GRB prompt and afterglow plateau emissions, *MNRAS* **451**, 3898 (August 2015).
13. C.-H. Tang, Y.-F. Huang, J.-J. Geng and Z.-B. Zhang, et al., Statistical Study of Gamma-Ray Bursts with a Plateau Phase in the X-Ray Afterglow, *ApJS* **245**, p. 1 (November 2019).
14. M. G. Dainotti, S. Postnikov, X. Hernandez and M. Ostrowski, A Fundamental Plane for Long Gamma-Ray Bursts with X-Ray Plateaus, *ApJL* **825**, p. L20 (July 2016).
15. M. G. Dainotti, A. Ł. Lenart, G. Sarracino, et al., The X-Ray Fundamental Plane of the Platinum Sample, the Kilonovae, and the SNe Ib/c Associated with GRBs, *ApJ* **904**, p. 97 (December 2020).
16. L. Izzo, M. Muccino, E. Zaninoni, et al., New measurements of Ω_m from gamma-ray bursts, *A&A* **582**, p. A115 (October 2015).
17. M. Muccino, L. Izzo, O. Luongo, et al., Tracing Dark Energy History with Gamma-Ray Bursts, *ApJ* **908**, p. 181 (February 2021).
18. F. Xu, C.-H. Tang, J.-J. Geng, et al., 2021, ApJ, in press, X-ray Plateaus in Gamma-Ray Burst Afterglows and Their Application in Cosmology, *arXiv e-prints*, p. arXiv:2012.05627 (December 2020).
19. B. Efron and V. Petrosian, A Simple Test of Independence for Truncated Data with Applications to Redshift Surveys, *ApJ* **399**, p. 345 (November 1992).
20. M. G. Dainotti, V. Petrosian, J. Singal and M. Ostrowski, Determination of the Intrinsic Luminosity Time Correlation in the X-Ray Afterglows of Gamma-Ray Bursts, *ApJ* **774**, p. 157 (September 2013).
21. M. Demianski, E. Piedipalumbo, D. Sawant and L. Amati, Cosmology with gamma-ray bursts. I. The Hubble diagram through the calibrated $E_{p,I}$-E_{iso} correlation, *A&A* **598**, p. A112 (February 2017).
22. M. G. Dainotti and R. Del Vecchio, Gamma Ray Burst afterglow and prompt-afterglow relations: An overview, *NewAR* **77**, 23 (April 2017).
23. L. Amati, C. Guidorzi, F. Frontera, et al., Measuring the cosmological parameters with the $E_{p,i}$-E_{iso} correlation of gamma-ray bursts, *MNRAS* **391**, 577 (December 2008).
24. D. M. Scolnic, D. O. Jones, A. Rest, et al., The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *ApJ* **859**, p. 101 (June 2018).

A new perspective on cosmology through Supernovae Ia and Gamma Ray Bursts

B. De Simone

Department of Phyisics "E. R. Caianiello", University of Salerno, Via Giovanni Paolo II, 132, 84084, Fisciano, Salerno, Italy
E-mail: b.desimone1@studenti.unisa.it
www.df.unisa.it

V. Nielson

SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
Astronomy Department, University of Michigan, Ann Arbor, MI 48109, USA

E. Rinaldi

Physics Department, University of Michigan, Ann Arbor, MI 48109, USA
RIKEN Cluster for Pioneering Research, Theoretical Quantum Physics Laboratory, Wako, Saitama 351-0198, Japan
RIKEN iTHEMS, Wako, Saitama 351-0198, Japan
RIKEN Center for Quantum Computing, Wako, Saitama 351-0198, Japan

M. G. Dainotti*

National Astronomical Observatory of Japan, 2 Chome-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
The Graduate University for Advanced Studies, SOKENDAI, Shonankokusaimura, Hayama, Miura District, Kanagawa 240-0193, Japan
Space Science Institute, Boulder, CO, USA
E-mail: maria.dainotti@nao.ac.jp
www.guas-astronomy.jp/eng

The actual knowledge of the structure and future evolution of our universe is based on the use of cosmological models, which can be tested through the so-called 'probes', namely astrophysical phenomena, objects or structures with peculiar properties that can help to discriminate among different cosmological models. Among all the existing probes, of particular importance are the Supernovae Ia (SNe Ia) and the Gamma Ray Bursts (GRBs): the former are considered among the best standard candles so far discovered but suffer from the fact that can be observed until redshift $z = 2.26$, while the latter are promising standardizable candles which have been observed up to $z = 9.4$, surpassing even the farthest quasar known to date, which is at $z = 7.64$. The standard candles can be used to test the cosmological models and to give the expected values of cosmological parameters, in particular the Hubble constant value. The Hubble constant is affected by the so-called "Hubble constant tension", a discrepancy in more than 4 σ between its value measured with local probes and its value measured through the cosmological probes. The increase in the number of observed SNe Ia, as well as the future standardization of GRBs through their correlations, will surely be of help in alleviating the Hubble constant tension and in explaining the structure of the universe at higher redshifts. A promising class of

*Corresponding author

GRBs for future standardization is represented by the GRBs associated with Supernovae Ib/c, since these present features similar to the SNe Ia class and obey a tight correlation between their luminosity at the end of the plateau emission in X-rays and the time at the end of the plateau in the rest-frame.

Keywords: Gamma Ray Bursts, Supernovae Ia, Cosmology, Correlations, Hubble constant, Hubble tension

Introduction

Modern cosmology is based on the so-called ΛCDM model: this is the standard cosmological model based on the Cold Dark Matter (CDM) and containing the cosmological constant Λ that describes the dark energy contribution to the expansion of the cosmos. This model has been widely accepted in the scientific community and can predict the accelerated expansion of the universe, as proved by the outstanding works of Riess et al. (1998)[1] and Perlmutter et al. (1999)[2] where this effect was demonstrated through the use of SNe Ia. Despite being well received, the ΛCDM suffers from some open problems, in particular the Hubble constant (H_0) tension. This is the discrepancy, in more than 4 σ, between the value of H_0 measured with local probes (Cepheids and SNe Ia) and the value obtained through the Cosmic Microwave Background (CMB) radiation data measured by the Planck satellite. To solve this issue, many approaches have been proposed (alternative cosmological models, refined measurements of H_0, etc.), but it is necessary to use reliable probes for testing the cosmological models. From this perspective, the so-called standard candles are very important: these are astrophysical objects or events which have a fixed luminosity or that obey an intrinsic relation between the luminosity and some of the other parameters that do not depend on the luminosity themselves. To date, different objects have been standardized, in particular the SNe Ia which have a nearly uniform peak luminosity, but the problem is that these can cover only a relatively small range of redshifts: the farthest SN Ia so far observed has a redshift of $z = 2.26$.[3] The GRBs may have a key role in the future development of cosmology since they have been observed at higher redshifts than SNe Ia and quasars (currently the farthest quasar being at redshift 7.64[4]) and are able to further extend the Hubble diagram (the distance moduli versus the redshift). A particular class of GRBs, the GRBs with a plateau emission (namely, a relatively flat section of the lightcurve that follows the prompt emission and precedes the afterglow, observed not only in X-rays and optical but also in γ-rays[5]), has proven to be a promising standardizable candle through the application of tight correlations among the GRB lightcurve properties, such as the intrinsic and unbiased correlation between the luminosity at the end of the plateau emission in the X-rays, L_X, with the time at the end of the plateau emission in the rest-frame, T_X^* [6–13] (which was later confirmed also in the optical[14]) and the unbiased one between L_X and the 1s peak prompt luminosity, L_{peak}.[15,16] A combination of these two correlations led to the so-called 3D fundamental plane relation, namely a tight correspondence between L_X, T_X^*, and

L_{peak}.[17–19] For a review of these correlations, see Dainotti & Del Vecchio (2017),[20] Dainotti et al. (2018a,2018b),[21,22] and Dainotti (2019).[23] The GRB correlations are not only useful as cosmological tools, but also as discriminant among theoretical models to explain the GRBs origin, emission mechanism, and the nature of their environment.[24–29] In particular, the plateau emission and the correlations that involve its properties strongly suggest how the typical magnetic fields and spin periods of the magnetars (namely, a fast rotating neutron star) could explain the plateau emission itself (Rowlinson et al. 2014).[25] In a later paper, Rea et al. (2015)[26] show that the magnetar model can be reconciled within the GRB emission in the plateau only if supermagnetar with high magnetic field strengh are allowed. In a successive paper of Stratta et al. (2018)[30] for the first time a non-ideal modelling of spindown magnetar is fitted to the afterglow data with a statistical sample of 40 Long GRBs (LGRBs) with a well-defined plateau and 13 Short GRBs (SGRBs) including the short with extended emission. The conclusion reached in that paper is that SGRBs including the SGRBs with Extended Emission (SEE) and LGRBs can be explained within the magnetar model but with the difference that the LGRBs occupy a lower end in the magnetic field-spin period plane compared to the SGRBs which present a higher spin, P, and a higher magnetic field, B. The correlation between magnetic field and spin period follow the established physics of the spin-up line for accreting neutron star in Galactic binary systems. The B−P relation obtained perfectly matches spin-up line predictions for the magnetar model with mass accretion rates expected in the GRB prompt phase. The latter are $\sim$ 1114 orders of magnitude higher than those inferred for the Galatic accreting NSs. Thus, correlations are useful to cast more light on the physics of GRBs and this represents an effective support towards the future standardization of these transient phenomena. The current proceeding focuses on two main topics concerning cosmology with SNe Ia and GRBs: (1) the investigation of the H_0 tension through a binning approach on the Pantheon sample of SNe Ia[31] and (2) the use of the GRB fundamental planes in optical and X-ray as cosmological distance indicators and the discussion of their future use as standardizable candles.

1. On the Hubble constant tension in the SNe Ia Pantheon sample

The H_0 tension is the discrepancy in more than 4 σ between the value of H_0 estimated with local probes, such as SNe Ia and Cepheids ($H_0 = 74.03 \pm 1.42$ km $s^{-1}Mpc^{-1}$), and the value of H_0 obtained with the Planck Cosmic Microwave Background (CMB) radiation ($H_0 = 67.4 \pm 0.5$ km $s^{-1}Mpc^{-1}$). It represents one of the most important open problems in modern cosmology. To investigate it, we divided the Pantheon sample (Scolnic et al. 2018[32]), a collection of 1048 spectroscopically confirmed SNe Ia, into 3, 4, 20, and 40 bins ordered in redshift. In our analysis, we considered two cosmological models: the standard ΛCDM model and the w_0w_aCDM model, where the equation of state parameter is expressed according to the Chevallier-Polarski-Linder parametrization ($w(z) = w_0 + w_a * z/(1 + z)$).[33,34]

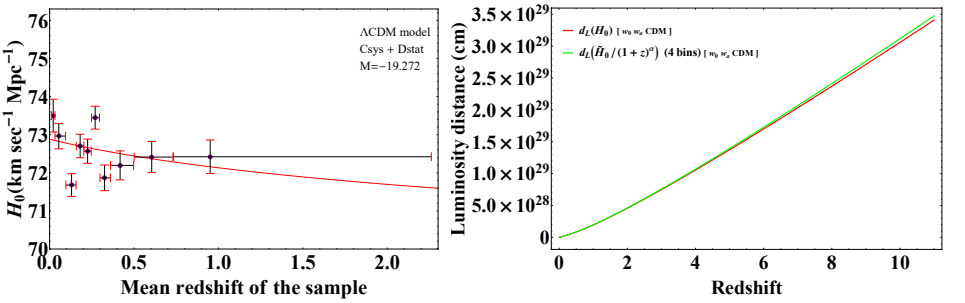


Fig. 1. **Left panel.** The fitting of the H_0 values with redshift in the case of 10 bins of SNe, assuming the ΛCDM model, using the full covariance matrix composed of systematic (C_{sys}) and statistical (D_{stat}) uncertainties,[32] and fixing $M = -19.272$ such that $H_0 = 73.5km\ s^{-1}\ Mpc^{-1}$ in the first bin. The slow decreasing trend is plain to see. **Right panel.** The comparison between the luminosity distance formula corrected with the $g(z)$ (in green) and the luminosity distance given by the standard w_0w_aCDM model (in red) as functions of the redshift.

Due to the degeneracy between the H_0 and the fiducial absolute magnitude of SNe Ia, M, we set M in such a way that locally the $H_0 = 73.5km\ s^{-1}\ Mpc^{-1}$. The fiducial values of the total matter density parameters have been assumed as $\Omega_M = 0.298$ (for the ΛCDM model) and $\Omega_M = 0.308$ (for the w_0w_aCDM model). Considering both the cosmological models, for each bin in the 3, 4, 20, and 40 divisions we performed a χ^2 minimization followed by a Monte Carlo Markov Chain simulation to obtain the best-fit H_0 values together with their 1 σ uncertainties. After obtaining these, we fitted them with the following functional form: $g(z) = H_0'/(1+z)^\alpha$, where H_0' is the value of the Hubble constant at $z = 0$ obtained with the fitting and α is the evolutionary parameter. We found that the H_0 in the Pantheon sample evolves slowly with the redshift, with an α coefficient in the order of 10^{-2} which is compatible with zero from 1.2 σ to 2.0 σ. For example, in the left panel of Figure 1 the case of fitting with 10 bins using the ΛCDM model is shown, while in the right panel a comparison between the standard luminosity distance in the w_0w_aCDM model and the corrected luminosity distance substituting H_0 with $g(z)$ is plotted. Despite the α coefficients being compatible with zero in 3 σ, the highlighted decreasing trend may affect the cosmological results. Indeed, a modification of the luminosity distance formula where H_0 is replaced with the $g(z)$ form shows how the modified luminosity distance curve departs significantly from the ΛCDM canonical one at redshift $z \sim 10$ (see the right panel of Figure 1). To check what would be the behavior of H_0 if the trend was real, we extrapolated its value from the fitting at the redshift of the most distant galaxy so far discovered ($z = 11.09$, Oesch et al. 2016[35]) and at the redshift of the last scattering surface ($z = 1100$). We obtained that the extrapolated value of H_0 at $z = 1100$ is compatible in 1 σ with the one obtained through the Planck CMB measurements. If we consider the discrepancy between the aforementioned values of H_0, namely (74.03 ± 1.42) and (67.4 ± 0.5) $kms^{-1}Mpc^{-1}$ and we compare it with the tension between the fitting values H_0'

(at $z = 0$) and $H_0(z = 1100)$, we find that our approach leads to alleviate the H_0 tension from 54% to 72% for the ΛCDM and the w_0w_aCDM models, respectively. Thus, we not only found a way to alleviate the Hubble tension, but we provided also a plausible explanation for the observed discrepancy between the values of H_0 coming from probes at different redshifts. The observed decreasing trend may be due to hidden astrophysical biases, selection effects, or even the evolution of SNe Ia lightcurve parameters, as pointed out in Nicolas et al. (2021)[36] where it was shown that the stretch parameter distribution of the Pantheon sample is affected by the drift with the redshift. If this is not the case, and the trend is real instead, the explanation for it may be found in the modified gravity theories. We propose that the $f(R)$ theories in the Jordan frame may explain the observed slow evolution of H_0. However, to prove if this trend is real or not, it is necessary to rely (a) on the future observations of SNe Ia through that will enrich the currently known samples of transient phenomena and (b) on the extension of the Hubble diagram up to redshift ranges that SNe Ia could not cover, thus calling for the use of high-z phenomena such as the GRBs.

2. Optical and X-ray GRB Fundamental Planes as Cosmological Distance Indicators

GRBs, being observed at very high-z (up to $z = 9.4$ so far[37]), have the potential to be employed as standardizable candles, extending the cosmological distance ladder beyond the redshift of SNe Ia. Therefore, we employ them to do so by first standardizing them using the 3D fundamental plane relation. 50 X-ray GRBs cut from a full sample of 222 define the platinum sample, and their corresponding fundamental plane parameter variables are determined by Monte Carlo Markov Chain sampling maintained by Gelman-Rubin statistical constraints. We use both the platinum sample and a well-defined sample of SNe Ia as a combination of probes to accurately constrain the matter content of the modern universe, Ω_M. To do so, the plane parameters are allowed to vary simultaneously with Ω_M within the chain sampling. We find our best results with the addition of a third probe, namely, Baryon Acoustic Oscillations (BAOs), given their reliability as standard rulers. The combination of these three probes allows us to constrain Ω_M to 0.306 ± 0.006, assuming a ΛCDM cosmology. Furthermore, we test, for the first time in this research field, the novel 3D optical correlation as the extension of the 3D fundamental plane in X-ray wavelengths as a cosmological tool and check its applicability compared to that of the confirmed X-ray relation. In doing so, we find that our optical GRB sample is as efficacious in the determination of Ω_M as the platinum X-ray sample. To increase the precision on the estimate of Ω_M, we consider redshift evolutionary effects to overcome common biases such as the Malmquist effect. We employ the reliable Efron & Petrosian (1992)[38] statistical method to ensure that the correlation is intrinsic to the GRB physical mechanics and not due to selection bias. It is by accounting for this that we decrease the intrinsic scatter on the X-ray plane by 44.4%, thus defining

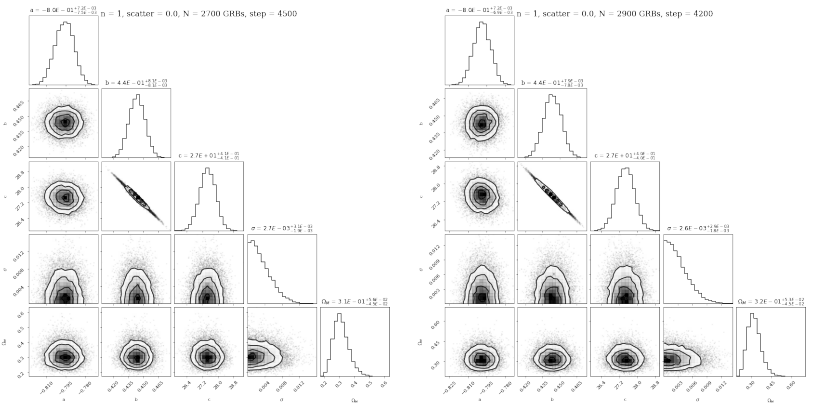


Fig. 2. **Left panel.** Contour plots with the parameters of the fundamental plane relation in X-rays (a, b, c) together with the intrinsic scatter σ and Ω_M, in the case of 2700 simulated GRBs with the evolution correction through the Efron & Petrosian method. **Right panel.** The same of the left panel, considering 2900 simulated GRBs.

the tightest 3D GRB correlation involving plateau features in the literature to date. Using this corrected sample, we confirm a value of 0.306 ± 0.006 defines Ω_M.

To both understand and predict the applicability of GRBs as standalone standard candles, we simulate additional GRBs to see how many are needed to reach uncertainties on Ω_M comparable to that of SNe Ia-derived values; we compare against Conley et al. (2011),[39] Betoule et al. (2014),[40] and Scolnic et al. (2018)[32] symmetrized error and standard deviation limits. In Figure 2, the contours for the parameters of the fundamental plane together with Ω_M after the Efron & Petrosian method application are shown in the case of 2700 (left) and 2900 (right) simulated GRBs. To do so, we first use both the full optical and X-ray GRB samples as independent bases for GRB simulation, conducted again by MCMC techniques. We find that the optical sample yields much smaller uncertainties on each simulation than the X-ray, and consequently a more constrained value of Ω_M. To increase this precision, we explore two methods of trimming our GRB sample leading to tighter planes that are in turn used as the base for simulation; an a posteriori trim using the smallest sample of GRBs for which the intrinsic scatter on the 3D plane which they defined remained near-zero, and an 'a priori' trim for which a number of possible sample sizes were tested and the one that yielded the best results after the fact is chosen. By simulating a large range of GRB sample sizes in both optical and X-ray wavelengths, we find by the construction of probability maps that the same precision as Betoule et al. (2014)[40] is achieved with only 376 simulated optical GRBs for which 47.5% of the fundamental plane variable error bars have been halved by a light curve reconstruction procedure. We find that the Conley et al. (2011)[39] limit is already achievable in most circumstances with current GRB numbers, and that the Scolnic et al. (2018)[32] limit is achieved for 1152 optical GRBs. Considering both

the detection power of future deep-space surveys THESEUS[30] and SVOM[41,42] and the nearing capability of machine learning approaches to extract unknown GRB redshifts, we conclude that GRBs will be as efficient standalone probes as SNe Ia by the year 2038. These results are interesting because, as the definition of GRBs as standard candles becomes more reliable with the introduction of the optical and X-ray fundamental planes, the addition of these astrophysical objects to SNe Ia and BAO data will soon give the most precise derivation of Ω_M ever achieved.

3. Summary and Conclusions

In this work, it was highlighted how and to what extent the GRBs may contribute to the cosmological analysis in the future. In the first part (1), the issue of the Hubble constant tension has been investigated with the Pantheon sample of SNe Ia, showing how the H_0 itself is characterized by a slowly decreasing trend with the redshift. This opens the discussion on the reason for this result: together with the possibility that a modified gravity theory could be an explanation, it is also likely that different selection biases persist in the observations of SNe Ia. This strongly suggests that the methods for biases correction, such as the Efron & Petrosian method, are needed in many astrophysical probes to achieve a reliable estimation of the cosmological parameters. Nevertheless, the limited coverage of redshift achieved by SNe Ia leads to the need for probes that can be observed at larger redshift. To this end, in the second part (2) it was shown how the 3D fundamental plane relation in the optical may help the future standardization of GRBs together with its counterpart in the X-rays. In addition, through the simulation of GRBs, it has been shown how this probe is a promising candidate to extend the Hubble diagram up to redshift greater than the ones of SNe Ia. Furthermore, a particular class of objects arouses much interest in the latest years, namely the class of LGRBs associated with Supernovae Ib/c (GRB-SNe): these manifest as supernovae appearing in the afterglow of the LGRBs and are very important since it was highlighted how the associated SNe obey a stretch-luminosity relation similar to the typical one of SNe Ia.[43] The physics behind their emission mechanism has been an object of study on several occasions[44–46] and one of the most interesting features of this class is that within the $L_X - T_X^*$ relation they show a Spearman correlation coefficient higher than the other subclasses of GRBs.[13] Thus, it is possible to use these events as a bridge between the properties of LGRBs and the ones of SNe Ib/c, giving a new perspective on the forthcoming standardization of GRB-SNe Ib/c. It is expected that the next observations of the new transients in the optical through the Subaru Telescope[47] and KISO Telescope[48] will help to investigate for selection biases and correct the current cosmological expectations on the evolution of the universe. Reliable testing of the cosmological models requires always new and more distant standardizable candles, and the GRBs have proven to be a reliable candidate for this purpose.

References

1. Observational evidence from supernovae for an accelerating universe and a cosmological constant, *The Astronomical Journal* **116**, p. 1009–1038 (September 1998).
2. S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch and T. S. C. Project, Measurements of Ω and Λ from 42 high-redshift supernovae, *The Astrophysical Journal* **517**, 565 (June 1999).
3. S. A. Rodney, A. G. Riess, D. M. Scolnic, D. O. Jones, S. Hemmati, A. Molino, C. McCully, B. Mobasher, L.-G. Strolger, O. Graur and et al., Two Type Ia Supernovae at Redshift ∼2 : Improved Classification and Redshift Determination with Medium-band Infrared Imaging, *The Astronomical Journal* **151**, p. 47 (February 2016).
4. F. Wang, J. Yang, X. Fan, J. F. Hennawi, A. J. Barth, E. Banados, F. Bian, K. Boutsia, T. Connor, F. B. Davies, R. Decarli, A.-C. Eilers, E. P. Farina, R. Green, L. Jiang, J.-T. Li, C. Mazzucchelli, R. Nanni, J.-T. Schindler, B. Venemans, F. Walter, X.-B. Wu and M. Yue, A luminous quasar at redshift 7.642, *The Astrophysical Journal Letters* **907**, p. L1 (January 2021).
5. M. G. Dainotti, N. Omodei, G. P. Srinivasaragavan, G. Vianello, R. Willingale, P. O'Brien, S. Nagataki, V. Petrosian, Z. Nuygen, X. Hernandez, M. Axelsson, E. Bissaldi and F. Longo, On the Existence of the Plateau Emission in High-energy Gamma-Ray Burst Light Curves Observed by Fermi-LAT, *The Astrophysical Journal Supplement Series* **255**, p. 13 (July 2021).
6. M. G. Dainotti, V. F. Cardone and S. Capozziello, A time–luminosity correlation for γ-ray bursts in the X-rays, *Monthly Notices of the Royal Astronomical Society: Letters* **391**, L79 (November 2008).
7. M. G. Dainotti, R. Willingale, S. Capozziello, V. F. Cardone and M. Ostrowski, Discovery of a Tight Correlation for Gamma-ray Burst Afterglows with "Canonical" Light Curves, *The Astrophysical Journal* **722**, p. L215–L219 (October 2010).
8. M. G. Dainotti, V. F. Cardone, S. Capozziello, M. Ostrowski and R. Willingale, Study of possible systematics in the L*X - Ta* correlation of Gamma Ray Bursts, *The Astrophysical Journal* **730**, p. 135 (March 2011).
9. M. G. Dainotti, V. Petrosian, J. Singal and M. Ostrowski, Determination of the intrinsic Luminosity Time Correlation in the X-ray Afterglows of GRBs, *The Astrophysical Journal* **774**, p. 157 (August 2013).
10. M. G. Dainotti, V. F. Cardone, E. Piedipalumbo and S. Capozziello, Slope evolution of GRB correlations and cosmology, *Monthly Notices of the Royal Astronomical Society* **436**, 82 (September 2013).
11. M. G. Dainotti, R. D. Vecchio, N. Shigehiro and S. Capozziello, Selection effects in Gamma Ray Bursts correlations: consequences on the ratio between GRB and star formation rates, *The Astrophysical Journal* **800**, p. 31 (February 2015).
12. R. D. Vecchio, M. G. Dainotti and M. Ostrowski, Study of GRB light curve decay indices in the afterglow phase, *The Astrophysical Journal* **828**, p. 36 (August 2016).
13. Dainotti, M. G., Nagataki, S., Maeda, K., Postnikov, S. and Pian, E., A study of gamma ray bursts with afterglow plateau phases associated with supernovae, *A&A* **600**, p. A98 (2017).
14. M. G. Dainotti, S. Livermore, D. A. Kann, L. Li, S. Oates, S. Yi, B. Zhang, B. Gendre, B. Cenko and N. Fraija, The optical luminosity–time correlation for more than 100

gamma-ray burst afterglows, *The Astrophysical Journal Letters* **905**, p. L26 (December 2020).
15. M. G. Dainotti, M. Ostrowski and R. Willingale, Towards a standard gamma-ray burst: tight correlations between the prompt and the afterglow plateau phase emission, *Monthly Notices of the Royal Astronomical Society* **418**, 2202 (December 2011).
16. M. Dainotti, V. Petrosian, R. Willingale, P. O'Brien, M. Ostrowski and S. Nagataki, Luminosity–time and luminosity–luminosity correlations for GRB prompt and afterglow plateau emissions, *Monthly Notices of the Royal Astronomical Society* **451**, 3898 (June 2015).
17. M. G. Dainotti, S. Postnikov, X. Hernandez and M. Ostrowski, A Fundamental Plane for Long Gamma-Ray Bursts with X-Ray Plateaus, *The Astrophysical Journal* **825**, p. L20 (July 2016).
18. M. G. Dainotti, X. Hernandez, S. Postnikov, S. Nagataki, P. O'brien, R. Willingale and S. Striegel, A Study of the Gamma-Ray Burst Fundamental Plane, *The Astrophysical Journal* **848**, p. 88 (October 2017).
19. M. G. Dainotti, A. Ł. Lenart, G. Sarracino, S. Nagataki, S. Capozziello and N. Fraija, The X-Ray Fundamental Plane of the Platinum Sample, the Kilonovae, and the SNe Ib/c Associated with GRBs, *The Astrophysical Journal* **904**, p. 97 (November 2020).
20. M. Dainotti and R. Del Vecchio, Gamma ray burst afterglow and prompt-afterglow relations: An overview, *New Astronomy Reviews* **77**, p. 23–61 (April 2017).
21. M. G. Dainotti, R. Del Vecchio and M. Tarnopolski, Gamma-Ray Burst Prompt Correlations, *Advances in Astronomy* **2018**, p. 1–31 (2018).
22. M. G. Dainotti and L. Amati, Gamma-ray burst prompt correlations: Selection and instrumental effects, *Publications of the Astronomical Society of the Pacific* **130**, p. 051001 (April 2018).
23. M. Dainotti, *Gamma-ray Burst Correlations*2053-2563, 2053-2563 (IOP Publishing, 2019).
24. P. Kumar, R. Narayan and J. L. Johnson, Mass fall-back and accretion in the central engine of gamma-ray bursts, *Monthly Notices of the Royal Astronomical Society* **388**, 1729 (August 2008).
25. A. Rowlinson, B. P. Gompertz, M. Dainotti, P. T. O'Brien, R. A. M. J. Wijers and A. J. van der Horst, Constraining properties of GRB magnetar central engines using the observed plateau luminosity and duration correlation, *Monthly Notices of the Royal Astronomical Society* **443**, 1779 (July 2014).
26. N. Rea, M. Gullón, J. A. Pons, R. Perna, M. G. Dainotti, J. A. Miralles and D. F. Torres, Constraining the GRB-magnetar model by means of the Galactic pulsar population, *The Astrophysical Journal* **813**, p. 92 (November 2015).
27. G. P. Srinivasaragavan, M. G. Dainotti, N. Fraija, X. Hernandez, S. Nagataki, A. Lenart, L. Bowden and R. Wagner, On the Investigation of the Closure Relations for Gamma-Ray Bursts Observed by Swift in the Post-plateau Phase and the GRB Fundamental Plane, *The Astrophysical Journal* **903**, p. 18 (October 2020).
28. M. G. Dainotti, A. . Lenart, N. Fraija, S. Nagataki, D. C. Warren, B. De Simone, G. Srinivasaragavan and A. Mata, Closure relations during the plateau emission of Swift GRBs and the fundamental plane, *Publications of the Astronomical Society of Japan* **73**, p. 970–1000 (June 2021).
29. M. G. Dainotti, V. Petrosian and L. Bowden, Cosmological evolution of the formation rate of short gamma-ray bursts with and without extended emission, *The Astrophysical Journal Letters* **914**, p. L40 (jun 2021).

30. G. Stratta, R. Ciolfi, L. Amati, E. Bozzo, G. Ghirlanda, E. Maiorano, L. Nicastro, A. Rossi, S. Vinciguerra, F. Frontera and et al., THESEUS: A key space mission concept for Multi-Messenger Astrophysics, *Advances in Space Research* **62**, p. 662–682 (August 2018).
31. M. G. Dainotti, B. D. Simone, T. Schiavone, G. Montani, E. Rinaldi and G. Lambiase, On the Hubble Constant Tension in the SNe Ia Pantheon Sample, *The Astrophysical Journal* **912**, p. 150 (May 2021).
32. D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess and et al., The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample, *The Astrophysical Journal* **859**, p. 101 (May 2018).
33. E. V. Linder, Exploring the expansion history of the universe, *Phys. Rev. Lett.* **90**, p. 091301 (March 2003).
34. M. Chevallier and D. Polarski, Accelerating universes with scaling dark matter, *International Journal of Modern Physics D* **10**, p. 213–223 (April 2001).
35. P. A. Oesch, G. Brammer, P. G. v. Dokkum, G. D. Illingworth, R. J. Bouwens, I. Labbé, M. Franx, I. Momcheva, M. L. N. Ashby, G. G. Fazio and et al., A remarkably luminous galaxy at z=11.1 measured with hubble space telescope grism spectroscopy, *The Astrophysical Journal* **819**, p. 129 (March 2016).
36. N. Nicolas, M. Rigault, Y. Copin, R. Graziani, G. Aldering, M. Briday, Y.-L. Kim, J. Nordin, S. Perlmutter and M. Smith, Redshift evolution of the underlying type ia supernova stretch distribution, *Astronomy & Astrophysics* **649**, p. A74 (May 2021).
37. A. Cucchiara, A. J. Levan, D. B. Fox, N. R. Tanvir, T. N. Ukwatta, E. Berger, T. Krühler, A. K. Yoldaş, X. F. Wu, K. Toma, J. Greiner, F. E. Olivares, A. Rowlinson, L. Amati, T. Sakamoto, K. Roth, A. Stephens, A. Fritz, J. P. U. Fynbo, J. Hjorth, D. Malesani, P. Jakobsson, K. Wiersema, P. T. O'Brien, A. M. Soderberg, R. J. Foley, A. S. Fruchter, J. Rhoads, R. E. Rutledge, B. P. Schmidt, M. A. Dopita, P. Podsiadlowski, R. Willingale, C. Wolf, S. R. Kulkarni and P. D'Avanzo, A photometric redshift of z$\sim$ 9.4 for GRB 090429B, *The Astrophysical Journal* **736**, p. 7 (June 2011).
38. B. Efron and V. Petrosian, A simple test of independence for truncated data with applications to redshift surveys, *The Astrophysical Journal* **399** (December 1992).
39. A. Conley, J. Guy, M. Sullivan, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin and et al., Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova Legacy Survey, *The Astrophysical Journal Supplement Series* **192**, p. 1 (December 2010).
40. M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann and et al., Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples, *Astronomy & Astrophysics* **568**, p. A22 (August 2014).
41. B. Cordier, J. Wei, J. L. Atteia, S. Basa, A. Claret, F. Daigne, J. Deng, Y. Dong, O. Godet, A. Goldwurm, D. Götz, X. Han, A. Klotz, C. Lachaud, J. Osborne, Y. Qiu, S. Schanne, B. Wu, J. Wang, C. Wu, L. Xin, B. Zhang and S. N. Zhang, The SVOM gamma-ray burst mission (2015).
42. B. Cordier, D. Götz and C. Motch, The SVOM mission, a pathfinder for THESEUS (2018).
43. Z. Cano, Gamma-ray burst supernovae as standardizable candles, *The Astrophysical Journal* **794**, p. 121 (Oct 2014).

44. Dainotti, M. G., Bernardini, M. G., Bianco, C. L., Caito, L., Guida, R. and Ruffini, R., Grb 060218 and grbs associated with supernovae ib/c, *A&A* **471**, L29 (2007).
45. M. Dainotti, The astrophysical trypthic: Grb, sn and urca can be extended to grb060218?, *Journal of The Korean Physical Society - J KOREAN PHYS SOC* **56** (05 2010).
46. Z. Cano, S.-Q. Wang, Z.-G. Dai and X.-F. Wu, The observer's guide to the gamma-ray burst supernova connection, *Advances in Astronomy* **2017**, p. 1–41 (2017).
47. J.-i. Watanabe, D. Kinoshita, Y. Komiyama, F. Tetsuharu, Y. Urata and F. Yoshida, Optical Follow-up of the GRB 010222 Afterglow by Subaru Telescope1, *Publications of the Astronomical Society of Japan* **53**, L27 (08 2001).
48. S. Sako, R. Ohsawa, H. Takahashi, Y. Kojima, M. Doi, N. Kobayashi, T. Aoki, N. Arima, K. Arimatsu, M. Ichiki, S. Ikeda, K. Inooka, Y. Ita, T. Kasuga, M. Kokubo, M. Konishi, H. Maehara, N. Matsunaga, K. Mitsuda, T. Miyata, Y. Mori, M. Morii, T. Morokuma, K. Motohara, Y. Nakada, S.-I. Okumura, Y. Sarugaku, M. Sato, T. Shigeyama, T. Soyano, M. Tanaka, K. Tarusawa, N. Tominaga, T. Totani, S. Urakawa, F. Usui, J. Watanabe, T. Yamashita and M. Yoshikawa, The Tomo-e Gozen wide field CMOS camera for the Kiso Schmidt telescope, in *Ground-based and Airborne Instrumentation for Astronomy VII*, eds. C. J. Evans, L. Simard and H. Takami (SPIE, 2018).

Theory of plateau phase in Gamma-Ray Bursts

Asaf Pe'er
Department of Physics, Bar Ilan University,
Ramat-Gan 52900, Israel
E-mail: asaf.peer@biu.ac.il
asafpeer2.ph.biu.ac.il/

Since its discovery in 2005, the plateau phase seen in the early x-ray afterglow of a significant fraction (10's of %) of GRBs confuse theoreticians. A close look reveals that "plateau" bursts nearly never show evidence for LAT emission, neither a strong thermal component. Using this a hint, I argue that the plateau is due to the coasting of GRB jets in a "wind"-like medius. I provide the theoretical arguments for the evolution of the lightcurve, and show how both the X-ray and optical lightcurves are naturally fitted within the framework of this model. The end of the plateau thus marks the transition between the coasting and the self-similar motion phases.

Keywords: Gamma rays: bursts; Jets: relativistic; X-rays

1. Introduction

Following the launch of the *Swift* satellite in 2004, it became obvious that the early afterglow x-ray lightcurve (minutes - hours) does not follow the previously expected self-similar decay seen at much later times. Rather, in about 60% of GRBs, a flat "plateau", lasting for $\sim 10^2 - 10^5$ s was observered in their x-ray lightcurve,[1,2] before it turns into the familiar self-similar decay, well modeled by the Blandford-McKee[3] solution.

Over the years, many attempts were made to explain this surprising result. Among the ideas discussed in the literature, are: (i) continuous energy injection that slows down the acceleration[1,4–6]; (ii) A two component jet model[7]; (iii) A forward shock emission in inhomogeneous media[8]; (iv) scattering by dust, or other modification of the ambient density by a gamma-ray trigger[9,10]; (v) A dominant reverse shock emission[11–13]; (vi) evolving microphysical parameters[9,14]; (vii) Viewing angle effect - jets viewed off-axis[8,15–17]; (viii) proto-magnetar that releases some of its energy at late times[18]; and (ix) forward shock emission, before the deceleration phase.[19] These plethora of models indicate the large uncertainty that exists in understanding this phenomenon. Furthermore, one thing in common for all these models, except for the last one, is that they all require some addition to the simple "fireball" model dynamics accompanied by synchrotron emission from shock-accelerated electrons.

The last suggestion, by Shen and Matzner[19] requires an explosion into a "wind" density profile (namely, $\rho(r) \propto r^{-2}$). It did not gain popularity, as (i) the deduced values of the Lorentz factor are lower than the 'fiducial' values, $\Gamma_i \gtrsim 100$; and (ii) the claim that this model can only account for a-chromatic afterglow, and can therefore

explain only a sub-sample of the GRB population.[20] As we show here, these two claims are incorrect.

2. Measuring GRB Lorentz factor: Is there a contradiction?

To day, there are three main methods used to estimate the Lorentz factor of the expanding GRB jets. The first relies on the opacity argument: in order to avoid copious production of pairs that would prevent observations of $>$ MeV emission one requires a high Lorentz factor. In particular, GeV emission seen by the LAT detector on board *Fermi* satellite in several bursts, require $\Gamma \gtrsim 100$.[21–24]

The second method requires identification of the onset of the self-similar expansion phase, which is accompanied by the formation and propagation of a reverse shock. For plausible parameters, emission from the reverse shock is expected mainly in the optical band, and thus an optical flash is considered as an indication for the transition to a self-similar motion. The time of this transition depends on the initial Lorentz factor, and is used to deduce its value.[25–29]

A third method of inferring the initial Lorentz factor is from quantifying the properties of the thermal emission (temperature and flux) observed during the prompt phase in a significant fraction of bright GRBs.[30–33] Indeed, within the framework of the classical GRB "fireball" model, a thermal component is unavoidable, as deep in the flow it is optically thick. The key is the adiabatic losses of the photons until they escape the plasma at the photosphere. If the photospheric radius is not much above the coasting radius, the thermal photons do not lose a large fraction of their energy, and a thermal component can be identified.[34]

In order to examine this issue, we looked at the 2^{nd} *Fermi* catalogue.[35] Out of 186 GRBs in this catalogue, only 3 show evidence for a shallow decay phase in the LAT data[36]; only one (the hard-short GRB 090510) show any evidence for a decaying plateau in the *Swift*-XRT data. However, for this specific burst, the X-ray slope during the early afterglow ($0.69^{+0.05}_{-0.06}$) can only marginally identify it as having a plateau. We therefore conclude that for bursts that show high energy emission, no evidence for a plateau phase exists.

We next looked at bursts which show evidence for a strong thermal component, and appeared in.[37–39] Non of these showed any evidence for an X-ray plateau. Finally, we looked at bursts which we were interested in analyzing. These are bursts that do show evidence for an x-ray plateau, and for which the redshift is known, and optical data is available. Non of these bursts showed any evidence for a thermal emission during their prompt phase, or for high energy ($>$ MeV) emission. We thus conclude that for bursts which show a plateau in their x-ray lightcurve, there is no indication for a high initial Lorentz factor, of $\Gamma_i \gtrsim 100$'s.

3. Dynamics and radiation

Within the context of the GRB "fireball" model, the expanding plasma jet undergoes 3 phases[40–42]: (i) an initial acceleration phase; due to light aberration, this phase

is observed to last only a fraction of a second. (ii) A coasting phase, in which the Lorentz factor of the flow is approximately constant. For "typical" GRB parameters, this phase lasts ∼few seconds. (iii) After the expanding plasma collects sufficient mass from the ambient medium (ISM), $m_{ISM} \gtrsim M_{ej}/\Gamma_i$, ($M_{ej}$ is the initially ejected mass), the flow begins its self-similar decay phase, which is well described by the Blandford-McKee[3] self-similar solution. During this phase, the Lorentz factor of the flow gradually decays. This is the phase responsible for the late time afterglow emission.

Once the flow becomes optically thin, the main radiation mechanism is synchrotron emission. Here we focus on emission from electrons accelerated to a power law distribution by the forward shock that exists in all phases, though we particularly focus on the coasting phase (ii) and late afterglow phase (iii). The accelerated electrons assume a power law distribution with power law index, p ($N_{el}(\gamma)d\gamma \propto \gamma^{-p}$). The resulting spectra is a well-known broken power law,[43,44] with 2 characteristic frequencies (omitting discussion on the self-absorption frequency, which is below the optical band is thus not relevant for the analysed signal): the peak frequency, $\nu_m^{ob.} \propto \Gamma B \gamma_{el}^2$, and the cooling frequency, $\nu_c^{ob.} \propto \Gamma^3/(B^3 r^2)$. Here, Γ is the jet (bulk) Lorentz factor, B is the magnetic field, γ_{el} is the characteristic Lorentz factor of the accelerated electrons (the minimum Lorentz factor above which a power law distribution exists), and r is the emission radius. The characteristic emission frequency of the electrons accelerated by the forward shock is ν_m, and ν_c is obtained by equating the radiative cooling time to the dynamical time. A third parameter, N_e is the total number of radiating electrons, is needed to determine the total observed flux (for bursts with known redshift).

The global spectral and temporal shapes depend on whether the cooling frequency ν_c is above or below the peak frequency, ν_m. When $\nu_c < \nu_m$ the system is in the "fast cooling" regime, where electrons lose nearly all of their energy within the dynamical time, and when $\nu_c > \nu_m$ it is in the "slow cooling". At a given observed frequency ν, there are therefore 6 possible regions, summarized as follows and marked [A] – [F]:

- "Fast cooling", $\nu_c < \nu_m$:

$$F_\nu = F_{\nu,\rm peak} \times \begin{cases} \nu_c^{-1/3}\nu^{1/3} & \nu < \nu_c < \nu_m \ [A] \\ \nu_c^{1/2}\nu^{-1/2} & \nu_c < \nu < \nu_m \ [B] \\ \nu_m^{(p-1)/2}\nu_c^{1/2}\nu^{-p/2} & \nu_c < \nu_m < \nu \ [C] \end{cases} \quad (1)$$

- "Slow cooling", $\nu_m < \nu_c$:

$$F_\nu = F_{\nu,\rm peak} \times \begin{cases} \nu_m^{-1/3}\nu^{1/3} & \nu < \nu_m < \nu_c \ [D] \\ \nu_m^{(p-1)/2}\nu^{-(p-1)/2} & \nu_m < \nu < \nu_c \ [E] \\ \nu_m^{(p-1)/2}\nu_c^{1/2}\nu^{-p/2} & \nu_m < \nu_c < \nu \ [F] \end{cases} \quad (2)$$

We adopt the standard assumption that the energy dissipated by the forward shock is used both in generating the magnetic field and in accelerating the electrons.

This assumption implies that the electrons characteristic Lorentz factor is linearly proportional to the bulk Lorentz factor, $\gamma_{el} \propto \epsilon_e \Gamma$, and similarly the energy density in the magnetic field, $B^2/8\pi = \epsilon_B u$, where u is the energy density of the material behind the shock and $\epsilon_B < 1$ is the fraction of post-shock thermal energy used in generating the magnetic field. This gives $B \propto \epsilon_B^{1/2} n^{1/2} \Gamma$.

In the literature, two scenarios are in frequent use in describing the ambient density profile. The first is a constant density interstellar medium (ISM). This scenario was thoroughly considered in the literature. It is easy to show that in this scenario, during the coasting phase, the flux continuously increases in all 6 regimes [A]–[F], as the expanding flow continuously accumulates particles from the ISM which are accelerated by the forward shock and radiate.

A second scenario (also in wide use) is that in which the GRB explosion occurs into a pre-ejected stellar wind with constant mass ejection rate and constant wind velocity. These imoly that the density drops as $\rho(r) \propto r^{-2}$. As we show below, under certain conditions, during the coasting phase the obtained lightcurve is ≈flat. We therefore focus here and below on this scenario.

Using the scaling relations derived above, it is straightforward to show that during the coasting phase where $\Gamma = cosnt$, $\nu_m \propto t^{-1}$ (following the decrease in the density with radius, which leads to a decrease in the magnetic field), and $\nu_c \propto t$ from a similar reason. During the decaying self-similar expansion, $\Gamma(r) \propto r^{-1/2}$, and a similar argument gives $\nu_m \propto t^{-3/2}$ and $\nu_c \propto t^{1/2}$. We thus find that both during the coasting and during the self-similar phases in a "wind" scenario, ν_m decreases with observed time, while ν_c increases. These behaviours are demonstrated (along side the 6 regions, [A] – [F]) in Figures 1 and 2.

The most relevant regions are regions [E] and [F]. Using the scaling relations of ν_m, ν_c and $N_e \propto r$, one obtains the scaling relations for these regions:

Table 1. Scaling relations of regions [E] and [F].

Region	Coasting phase	Self-similar decay phase
F ($\nu_c < \nu$)	$F_\nu \propto t^{(2-p)/2}\nu^{-p/2} \sim t^0$	$F_\nu \propto t^{(2-3p)/4}\nu^{-p/2} \sim t^{-1}$
E ($\nu < \nu_c$)	$F_\nu \propto t^{(1-p)/2}\nu^{-(p-1)/2} \sim t^{-0.5}$	$F_\nu \propto t^{(1-3p)/4}\nu^{-(p-1)/2} \sim t^{-1.25}$

Note: [a] Sample table footnote.

These results can now be confronted with the data. In region [F], which is always the earlier due to the rise of ν_c, one finds that at early times, the expected lightcurve is flat (if the power law index of the accelerated electrons is $p = 2$), which later decays as $F_\nu(t) \propto t^{-1}$. This is indeed what is observed in many GRBs that show a plateau. This model therefore provides an excellent explanation to the observed plateau phase. We note that as the optical band transient from region [F] to region [E] earlier than the x-ray band (see Figures on p. 3145), the optical lightcurve may very well decay during the X-ray plateau.

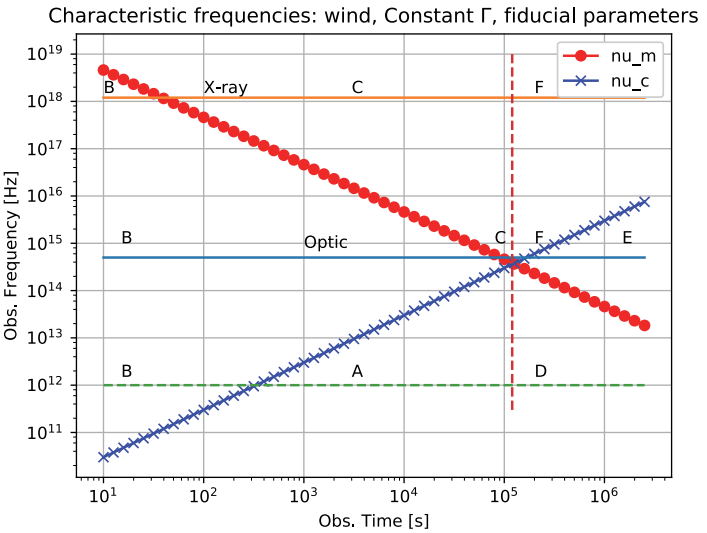


Fig. 1. Evolution of the characteristic synchrotron frequencies, ν_m and ν_c, calculated here for typical GRB parameters, during the coasting phase ($\Gamma_i = const$) for an explosion into a "wind" density environment, $\rho(r) \propto r^{-2}$. Marked are regions [A] – [F]. It is clear from the figure, that a given observed frequency change its regime in only one of two orders: Either $[B] \rightarrow [A] \rightarrow [D] \rightarrow [E]$ or $[B] \rightarrow [C] \rightarrow [F] \rightarrow [E]$.

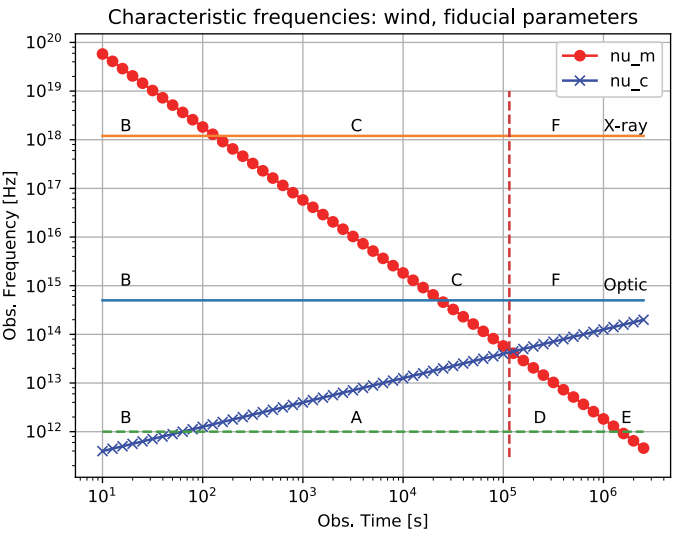


Fig. 2. Same as Figure 1, for evolution during the self-similar (decaying) expansion phase that follows the coasting phase.

4. Extracting information on the physical parameters

The excellent agreement between the theory and the data, calls for a simple interpretation according to which the transition from the plateau to the decay lightcurve marks the transition from the coasting to the self-similar expansion phases. For an explosion into a "Wind" environment, this transition time is related to the outflow parameters via

$$t^{\rm ob.}_{tran} = (1+z)\frac{9E}{32\pi A c^3 \Gamma_i^4} \tag{3}$$

where A is the proportionality constant of the wind density ($\rho(r) = Ar^{-2}$), Γ_i is the initial (coasting) Lorentz factor, E is the isotropic-equivalent energy and z is the redshift. Thus, for bursts with known redshift, the transition time provides a direct information on the value of $A\Gamma_i^4$.

Furthermore, when the lightcurve is flat, the flow is in region [F], and the parametric dependence of the x-ray flux is $\nu F_\nu(x) \propto c^3 A \Gamma_i^4 \epsilon_e d_L^{-2}$, where d_L is the luminosity distance. Thus, for bursts with known redshift (hence energy), the transition time combined with the x-ray flux provides a direct measurement on the fraction of energy given to the accelerated electrons, ϵ_e.

A decaying optical flux accompanying the flat x-ray plateau implies that the optical frequency must be in region [E]. In this case, the flux at the optical (say, R)-band is $\nu F_\nu(R) = F_{\nu,peak}\nu_m^{1/2}\nu_R^{1/2}$. For power law index $p = 2$, this gives a parametric dependence connecting $A^{7/4}\Gamma_i^3\epsilon_B^{3/4}$. This implies that for a given density (a value of A), there is a single corresponding value of the magnetization ϵ_B (as well as of the coasting Lorentz factor, Γ_i). This can be used in the opposite direction: an additional knowledge of ϵ_B, such as equipartition, or some fraction of it, can be used to determine the ambient density and the Lorentz factor.

In a recent paper [Dereli-Begue et. al., submitted] we used this method to extract information about the physical parameters of GRB outflow, as well as their initial Lorentz factor. A very interesting result, is that for the sample of GRBs with x-ray plateau, the initial Lorentz factor is typically in the range $20 \leq \Gamma_i \leq 200$, smaller than the values inferred by other methods. As explained above, there is no contradiction, and the results obtained in fact extend the known range of Lorentz factors in GRB outflows.

References

1. B. Zhang, Y. Z. Fan, J. Dyks, S. Kobayashi, P. Mészáros, D. N. Burrows, J. A. Nousek and N. Gehrels, Physical Processes Shaping Gamma-Ray Burst X-Ray Afterglow Light Curves: Theoretical Implications from the Swift X-Ray Telescope Observations, *Astrophys. J.* **642**, 354 (May 2006).
2. G. P. Srinivasaragavan, M. G. Dainotti, N. Fraija, X. Hernandez, S. Nagataki, A. Lenart, L. Bowden and R. Wagner, On the Investigation of the Closure Relations for Gamma-Ray Bursts Observed by Swift in the Post-plateau Phase and the GRB Fundamental Plane, *Astrophys. J.* **903**, p. 18 (November 2020).

3. R. D. Blandford and C. F. McKee, Fluid dynamics of relativistic blast waves, *Physics of Fluids* **19**, 1130 (August 1976).
4. J. A. Nousek, C. Kouveliotou, D. Grupe, K. L. Page, J. Granot, E. Ramirez-Ruiz, S. K. Patel, D. N. Burrows, V. Mangano, S. Barthelmy, A. P. Beardmore, S. Campana, M. Capalbi, G. Chincarini, G. Cusumano, A. D. Falcone, N. Gehrels, P. Giommi, M. R. Goad, O. Godet, C. P. Hurkett, J. A. Kennea, A. Moretti, P. T. O'Brien, J. P. Osborne, P. Romano, G. Tagliaferri and A. A. Wells, Evidence for a Canonical Gamma-Ray Burst Afterglow Light Curve in the Swift XRT Data, *Astrophys. J.* **642**, 389 (May 2006).
5. J. Granot, A. Königl and T. Piran, Implications of the early X-ray afterglow light curves of Swift gamma-ray bursts, *Mon. Not. R. Astron. Soc.* **370**, 1946 (August 2006).
6. G. Ghisellini, G. Ghirlanda, L. Nava and C. Firmani, "Late Prompt" Emission in Gamma-Ray Bursts?, *Astrophys. J.* **658**, L75 (April 2007).
7. J. L. Racusin, S. V. Karpov, M. Sokolowski, J. Granot, X. F. Wu, V. Pal'Shin, S. Covino, A. J. van der Horst, S. R. Oates, P. Schady, R. J. Smith, J. Cummings, R. L. C. Starling, L. W. Piotrowski, B. Zhang, P. A. Evans, S. T. Holland, K. Malek, M. T. Page, L. Vetere, R. Margutti, C. Guidorzi, A. P. Kamble, P. A. Curran, A. Beardmore, C. Kouveliotou, L. Mankiewicz, A. Melandri, P. T. O'Brien, K. L. Page, T. Piran, N. R. Tanvir, G. Wrochna, R. L. Aptekar, S. Barthelmy, C. Bartolini, G. M. Beskin, S. Bondar, M. Bremer, S. Campana, A. Castro-Tirado, A. Cucchiara, M. Cwiok, P. D'Avanzo, V. D'Elia, M. Della Valle, A. de Ugarte Postigo, W. Dominik, A. Falcone, F. Fiore, D. B. Fox, D. D. Frederiks, A. S. Fruchter, D. Fugazza, M. A. Garrett, N. Gehrels, S. Golenetskii, A. Gomboc, J. Gorosabel, G. Greco, A. Guarnieri, S. Immler, M. Jelinek, G. Kasprowicz, V. La Parola, A. J. Levan, V. Mangano, E. P. Mazets, E. Molinari, A. Moretti, K. Nawrocki, P. P. Oleynik, J. P. Osborne, C. Pagani, S. B. Pandey, Z. Paragi, M. Perri, A. Piccioni, E. Ramirez-Ruiz, P. W. A. Roming, I. A. Steele, R. G. Strom, V. Testa, G. Tosti, M. V. Ulanov, K. Wiersema, R. A. M. J. Wijers, J. M. Winters, A. F. Zarnecki, F. Zerbi, P. Mészáros, G. Chincarini and D. N. Burrows, Broadband observations of the naked-eye γ-ray burst GRB080319B, *Nature* **455**, 183 (September 2008).
8. K. Toma, K. Ioka, R. Yamazaki and T. Nakamura, Shallow Decay of Early X-Ray Afterglows from Inhomogeneous Gamma-Ray Burst Jets, *Astrophys. J.* **640**, L139 (April 2006).
9. K. Ioka, K. Toma, R. Yamazaki and T. Nakamura, Efficiency crisis of swift gamma-ray bursts with shallow X-ray afterglows: prior activity or time-dependent microphysics?, *Astron. Astrophys.* **458**, 7 (October 2006).
10. L. Shao and Z. G. Dai, Behavior of X-Ray Dust Scattering and Implications for X-Ray Afterglows of Gamma-Ray Bursts, *Astrophys. J.* **660**, 1319 (May 2007).
11. Z. L. Uhm and A. M. Beloborodov, On the Mechanism of Gamma-Ray Burst Afterglows, *Astrophys. J.* **665**, L93 (August 2007).
12. F. Genet, F. Daigne and R. Mochkovitch, Can the early X-ray afterglow of gamma-ray bursts be explained by a contribution from the reverse shock?, *Mon. Not. R. Astron. Soc.* **381**, 732 (October 2007).
13. R. Hascoët, F. Daigne and R. Mochkovitch, The prompt-early afterglow connection in gamma-ray bursts: implications for the early afterglow physics, *Mon. Not. R. Astron. Soc.* **442**, 20 (July 2014).
14. A. Panaitescu, P. Mészáros, N. Gehrels, D. Burrows and J. Nousek, Analysis of the X-ray emission of nine Swift afterglows, *Mon. Not. R. Astron. Soc.* **366**, 1357 (March 2006).

15. D. Eichler and J. Granot, The Case for Anisotropic Afterglow Efficiency within Gamma-Ray Burst Jets, *Astrophys. J.* **641**, L5 (April 2006).
16. G. Oganesyan, S. Ascenzi, M. Branchesi, O. S. Salafia, S. Dall'Osso and G. Ghirlanda, Structured Jets and X-Ray Plateaus in Gamma-Ray Burst Phenomena, *Astrophys. J.* **893**, p. 88 (April 2020).
17. P. Beniamini, J. Granot and R. Gill, Afterglow light curves from misaligned structured jets, *Mon. Not. R. Astron. Soc.* **493**, 3521 (April 2020).
18. B. D. Metzger, D. Giannios, T. A. Thompson, N. Bucciantini and E. Quataert, The protomagnetar model for gamma-ray bursts, *Mon. Not. R. Astron. Soc.* **413**, 2031 (May 2011).
19. R. Shen and C. D. Matzner, Coasting External Shock in Wind Medium: An Origin for the X-Ray Plateau Decay Component in Swift Gamma-Ray Burst Afterglows, *Astrophys. J.* **744**, p. 36 (January 2012).
20. P. Kumar and B. Zhang, The physics of gamma-ray bursts and relativistic jets, *Phys. Rep.* **561**, 1 (February 2015).
21. J. H. Krolik and E. A. Pier, Relativistic motion in gamma-ray bursts, *Astrophys. J.* **373**, 277 (May 1991).
22. E. E. Fenimore, R. I. Epstein and C. Ho, The escape of 100 MeV photons from cosmological gamma-ray bursts, *àaps* **97**, 59 (January 1993).
23. E. Woods and A. Loeb, Empirical Constraints on Source Properties and Host Galaxies of Cosmological Gamma-Ray Bursts, *Astrophys. J.* **453**, p. 583 (November 1995).
24. Y. Lithwick and R. Sari, Lower Limits on Lorentz Factors in Gamma-Ray Bursts, *Astrophys. J.* **555**, 540 (July 2001).
25. P. Meszaros and M. J. Rees, Optical and Long-Wavelength Afterglow from Gamma-Ray Bursts, *Astrophys. J.* **476**, 232 (February 1997).
26. S. Kobayashi, T. Piran and R. Sari, Can Internal Shocks Produce the Variability in Gamma-Ray Bursts?, *Astrophys. J.* **490**, p. 92 (November 1997).
27. R. Sari and T. Piran, GRB 990123: The Optical Flash and the Fireball Model, *Astrophys. J.* **517**, L109 (June 1999).
28. A. Panaitescu and P. Kumar, Properties of Relativistic Jets in Gamma-Ray Burst Afterglows, *Astrophys. J.* **571**, 779 (June 2002).
29. S. Kobayashi and B. Zhang, Early Optical Afterglows from Wind-Type Gamma-Ray Bursts, *Astrophys. J.* **597**, 455 (November 2003).
30. A. Pe'er, F. Ryde, R. A. M. J. Wijers, P. Mészáros and M. J. Rees, A New Method of Determining the Initial Size and Lorentz Factor of Gamma-Ray Burst Fireballs Using a Thermal Emission Component, *Astrophys. J.* **664**, L1 (July 2007).
31. A. Mizuta and S. Nagataki, Photospheric Thermal Radiation from GRB Collapsar Jets, *International Journal of Modern Physics Conference Series* **8**, 225 (2012).
32. R. Hascoët, F. Daigne and R. Mochkovitch, Prompt thermal emission in gamma-ray bursts, *Astron. Astrophys.* **551**, p. A124 (March 2013).
33. G. V. Vereshchagin and I. A. Siutsou, Diffusive photospheres in gamma-ray bursts, *Mon. Not. R. Astron. Soc.* **494**, 1463 (May 2020).
34. A. Pe'er, B.-B. Zhang, F. Ryde, S. McGlynn, B. Zhang, R. D. Preece and C. Kouveliotou, The connection between thermal and non-thermal emission in gamma-ray bursts: general considerations and GRB 090902B as a case study, *Mon. Not. R. Astron. Soc.* **420**, 468 (February 2012).
35. M. Ajello, M. Arimoto, M. Axelsson, L. Baldini, G. Barbiellini, D. Bastieri, R. Bellazzini, P. N. Bhat, E. Bissaldi, R. D. Blandford, R. Bonino, J. Bonnell, E. Bottacini, J. Bregeon, P. Bruel, R. Buehler, R. A. Cameron, R. Caputo, P. A. Caraveo, E. Cavazzuti, S. Chen, C. C. Cheung, G. Chiaro, S. Ciprini, D. Costantin, M. Crnogorcevic,

S. Cutini, M. Dainotti, F. D'Ammando, P. de la Torre Luque, F. de Palma, A. Desai, R. Desiante, N. Di Lalla, L. Di Venere, F. Fana Dirirsa, S. J. Fegan, A. Franckowiak, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Giglietto, F. Giordano, M. Giroletti, D. Green, I. A. Grenier, J. E. Grove, S. Guiriec, E. Hays, J. W. Hewitt, D. Horan, G. Jóhannesson, D. Kocevski, M. Kuss, L. Latronico, J. Li, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, S. Maldera, A. Manfreda, G. Martí-Devesa, M. N. Mazziotta, I. Mereu, M. Meyer, P. F. Michelson, N. Mirabal, W. Mitthumsiri, T. Mizuno, M. E. Monzani, E. Moretti, A. Morselli, I. V. Moskalenko, M. Negro, E. Nuss, M. Ohno, N. Omodei, M. Orienti, E. Orlando, M. Palatiello, V. S. Paliya, D. Paneque, M. Persic, M. Pesce-Rollins, V. Petrosian, F. Piron, S. Poolakkil, H. Poon, T. A. Porter, G. Principe, J. L. Racusin, S. Rainò, R. Rando, M. Razzano, S. Razzaque, A. Reimer, O. Reimer, T. Reposeur, F. Ryde, D. Serini, C. Sgrò, E. J. Siskind, E. Sonbas, G. Spandre, P. Spinelli, D. J. Suson, H. Tajima, M. Takahashi, D. Tak, J. B. Thayer, D. F. Torres, E. Troja, J. Valverde, P. Veres, G. Vianello, A. von Kienlin, K. Wood, M. Yassine, S. Zhu and S. Zimmer, A Decade of Gamma-Ray Bursts Observed by Fermi-LAT: The Second GRB Catalog, *Astrophys. J.* **878**, p. 52 (June 2019).

36. M. G. Dainotti, N. Omodei, G. P. Srinivasaragavan, G. Vianello, R. Willingale, P. O'Brien, S. Nagataki, V. Petrosian, Z. Nuygen, X. Hernandez, M. Axelsson, E. Bissaldi and F. Longo, On the Existence of the Plateau Emission in High-energy Gamma-Ray Burst Light Curves Observed by Fermi-LAT, *Astrophys. J.* **255**, p. 13 (July 2021).
37. H.-F. Yu, H. Dereli-Bégué and F. Ryde, Bayesian Time-resolved Spectroscopy of GRB Pulses, *Astrophys. J.* **886**, p. 20 (November 2019).
38. Z. Acuner, F. Ryde, A. Pe'er, D. Mortlock and B. Ahlgren, The Fraction of Gamma-Ray Bursts with an Observed Photospheric Emission Episode, *Astrophys. J.* **893**, p. 128 (April 2020).
39. H. Dereli-Bégué, A. Pe'er and F. Ryde, Classification of Photospheric Emission in Short GRBs, *Astrophys. J.* **897**, p. 145 (July 2020).
40. B. Paczynski, Gamma-ray bursters at cosmological distances, *Astrophys. J.* **308**, L43 (September 1986).
41. M. J. Rees and P. Meszaros, Relativistic fireballs - Energy conversion and time-scales, *Mon. Not. R. Astron. Soc.* **258**, 41P (September 1992).
42. M. J. Rees and P. Meszaros, Unsteady outflow models for cosmological gamma-ray bursts, *Astrophys. J.* **430**, L93 (August 1994).
43. G. B. Rybicki and A. P. Lightman, *Radiative processes in astrophysics* 1979.
44. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *Astrophys. J.* **497**, L17+ (April 1998).

Exploring the canonical behaviour of long Gamma-Ray Bursts using an intrinsic multi-wavelength afterglow correlation

S. R. Oates[a,*], J. L. Racusin[b], M. De Pasquale[c], D. Kocevski[d], M. J. Page[e], A. J. Castro-Tirado[f,g], J. Gorosabel[g,h,i,†], A. A. Breeveld[e], N. P. M. Kuin[e] and P. J. Smith[e]

[a] *School of Physics and Astronomy & Institute for Gravitational Wave Astronomy, University of Birmingham, B15 2TT, UK*

[b] *Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, Maryland 20771, USA*

[c] *Dipartimento di Scienze Matematiche e Informatiche, Scienze Fisiche e Scienze della Terra (MIFT), Universitá di Messina - Viale F. Stagno d'Alcontres, 31 - 98166 Messina, Italy*

[d] *NASA Marshall Space Flight Center, Huntsville, AL*

[e] *Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK*

[f] *Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, E-18008, Granada, Spain*

[g] *Unidad Asociada Departamento de Ingeniería de Sistemas y Automática, E.T.S. de Ingenieros Industriales, Universidad de Málaga, Spain*

[h] *Unidad Asociada Grupo Ciencias Planetarias UPV/EHU-IAA/CSIC, Departamento de Física Aplicada I, E.T.S.*

[i] *Ingeniería, Universidad del País Vasco UPV/EHU, Bilbao, Spain*

[†] *deceased*

[*] *E-mail: s.r.oates@bham.ac.uk*

In this conference proceeding we summarise our investigation of a correlation discovered between the afterglow luminosity (measured at restframe 200 s; log L_{200s}) and average afterglow decay rate (measured from restframe 200 s onwards; $\alpha_{>200s}$) of long duration Gamma-ray Burst (GRB) afterglows, found in both the optical/UV and X-ray afterglows. We examine the correlation in the X-ray light curves and find that it does not depend on the presence of specific features in the X-ray light curve. We test how the optical and X-ray parameters log $L_{O,200s}$, log $L_{X,200s}$, $\alpha_{O,>200s}$, $\alpha_{X,>200s}$ relate to each other and to parameters from the prompt emission phase. Using a Monte Carlo simulation, we explore whether these relationships are consistent with predictions of a basic standard afterglow model. We conclude that most of the correlations we observe are consistent with a common underlying physical mechanism producing GRBs and their afterglows regardless of their detailed temporal behaviour, but this basic model has difficulty explaining correlations involving $\alpha_{>200s}$. We therefore briefly discuss alternative more complex afterglow models.

Keywords: Gamma-ray bursts; correlations

1. Introduction

Gamma-ray bursts (GRBs) are intense flashes of gamma-rays that are usually accompanied by an afterglow, longer lived emission that may be detected at X-ray to radio wavelengths. Studies of single GRBs provide exceptional detail on the behaviour and physical properties of individual events. However, statistical investigations of large samples of GRBs aim to find common characteristics and correlations

that link individual events and therefore provide insight into the mechanisms common to GRBs. Statistical investigations performed so far have found a number of trends and correlations within and linking the prompt gamma-ray emission and the afterglow emission e.g., Ref. 1–10.

In this conference proceeding, we focus on the discovery of a correlation, found in a sample of optical/UV afterglow light curves,[11] between the logarithmic brightness ($\log L_{O,200s}$; measured at restframe 200 s and at a restframe wavelength 1600 Å), and their average decay rate ($\alpha_{>200s}$; measured from restframe 200 s onwards with a single power-law). This correlation has also been found at X-ray wavelengths.[12] To gain insight into the origin of the luminosity-decay correlation, we investigated the X-ray sample and how it relates to other GRB properties.[12,13] Ref. 12 discovered the $\log L_{200s} - \alpha_{>200s}$ in the X-ray afterglows observed by the *Swift* X-ray Telescope (XRT[14]). Ref. 13 compared the parameters of the optical/UV $\log L_{200s} - \alpha_{>200s}$ correlation with the equivalent values from the X-ray and also explored their relationship to properties of the prompt emission, namely the isotropic energy E_{iso} and the peak energy E_{peak}. In the following, we will provide a summary of these papers.

2. Sample Selection

The X-ray light curves were retrieved from the University of Leicester *Swift*-XRT Team GRB repository.[15,16] Ref. 12 selected X-ray afterglows of *Swift*/BAT detected GRBs, that were observed between December 2004 and March 2014, that had measured redshifts, had at least 3 light curve bins and started within a factor of restframe 200 s, t_{200s}. The final sample includes 246 GRBs; 237 long and 9 short. The count rate light curves were converted to flux density at 1 keV using the spectral index from the automated fits to the photon counting mode data, and then to intrinsic luminosity. All light curve fitting is performed in the count rate domain.

The Ref. 13 sample, used to compare the properties of the X-ray, optical afterglows together with their prompt emission parameters, consists of 48 long GRBs that overlap the Ref. 11 and Ref. 12 samples. The optical/UV luminosity light curves were produced at a common wavelength of 1600 Å.[11]

To measure luminosity at 200 s, $\log L_{200s}$, for the optical/UV light curves we interpolated between 100 and 2000 s and for the X-ray we measured the luminosity at 200 s from the best-fit light curve model.[17] To obtain the average decay rate $\alpha_{>200s}$, a single power-law was fit to each optical/UV and X-ray light curve from 200 s onwards.

For some X-ray afterglows, an initial steep decay, associated with the tail of the prompt emission,[18] can contaminate the initial part of the X-ray light curve. Of the 246 X-ray light curves, the steep decay segment is found to contaminate 23 X-ray light curves at restframe 200 s. We identify a light curve segment to have a prompt origin if there is a steep to shallow transition with $\Delta\alpha > 1.0$. In these situations the average decay index is measured with a simple power-law fit to data beyond restframe 200 s and after the steep to shallow transition. In order to get a better

estimate of the afterglow luminosity at restframe 200 s, we extrapolate back to restframe 200 s the first segment of the best-fit light curve that is not contaminated by the prompt emission (see also Ref. 12).

To compare the afterglow properties with the prompt emission properties we determined the isotropic energy E_{iso} and restframe peak energy, E_{peak} from the γ-ray emission, following Ref. 17. Of the 48 GRBs that overlap the optical and X-ray samples,[11,12] we determined E_{peak} for 44 GRBs and E_{iso} for 47 GRBs.

We determine the strength of the correlation using the IDL tool *r_correlate*, which measures the Spearman rank coefficient (R_{sp}), and its corresponding null hypothesis probability (p). We also use the partial Spearman rank correlation to test the dependence of each correlation on redshift.

We perform a linear regression analysis using the IDL routines *fitexy* and *sixlin*: *fitexy* is used when both parameters have errors, *sixlin* is used when we do not know the errors on one or both parameters. Since there are only a handful of GRBs with errors on the E_{iso} and E_{peak} parameters, we choose to discard errors in both parameters and use *sixlin* when determining the strength and significance of each correlation with one of these parameters involved.

3. Results

3.1. $\log L_{X,200s} - \alpha_{X,>200s}$ *correlation*

Within the sample of 246 X-ray afterglow light curves we see evidence for a correlation between $\log L_{X,200s}$ and $\alpha_{X,>200s}$, similar to that in the optical/UV.[11] The X-ray afterglow sample consists of long and short GRBs and the light curves are often more complex than the optical/UV, consisting of features such as plateaus and flares that may add scatter or influence the correlation. We test these effects by dividing the sample by specific characteristics and reproducing the same analyses. We report all of these tests and the final correlation in Table 1 and Fig. 1.

The first test is to determine if the $\log L_{X,200s} - \alpha_{X,>200s}$ correlation is observed for both short and long GRBs. Separating GRBs in to long and short classes, we find that long GRBs are significantly correlated, but no significant correlation is found for the short GRBs. This suggests that their is some intrinsic difference in the afterglow properties of short and long GRBs, be it their environment or jet dynamics. For all further tests of the $\log L_{X,200s} - \alpha_{X,>200s}$ correlation, we exclude short GRBs.

X-ray flares have been shown to have an internal rather than external shock origin and may be a potential source of contamination in measuring the $\log L_{X,200s}$ and $\alpha_{X,>200s}$ parameters. We first separate those afterglows with X-ray flares (without removing flaring intervals) and those without flares, and find that the two samples show very similar correlation strengths and slopes, but with slightly more scatter in the sample with flares. We do not exclude GRBs with flares, instead we exclude the flaring intervals from the light curves and refit them to obtain $\log L_{X,200s}$ and

$\alpha_{\rm X,avg,>200s}$. Re-running the Spearman rank test, on the entire sample we find a tighter correlation. Henceforth, we therefore, continue our investigation using the flare-removed average decay fits.

The average decay rate may also be influenced by the plateau phase observed in some X-ray afterglows (e.g Ref. 17). For example, a light curve with an extremely long plateau (e.g. GRB 060729; Ref. 19) may have a shallower average decay, whereas GRBs without plateaus would be steeper. In addition Refs. 20, 21 have shown that there is a relationship between the time and flux of the end of the X-ray plateau, which could be another manifestation of the log $\rm L_{200s} - \alpha_{>200s}$ correlation. We thus divide the X-ray sample in to those with and without plateaus and to determine if the correlation is still present in both sub-samples. The results in Table 1 indicate that the luminosity-average decay correlation is present and similarly significant in both sub-samples. This suggests that the plateau feature is not responsible for producing the correlation and is not solely responsible for regulating the average afterglow decay.

Table 1. Linear regression and correlation statistics for each test of the X-ray light curve sample. The partial Spearman rank coefficient tests the dependence of the given set of parameters on redshift. For the regular or partial Spearman rank coefficient, the corresponding null hypothesis is given to its right. This table is adapted from Table 2 in Ref. 12.

Sample	Parameters		Spearman	Null	Partial	Null	Best fit linear regression		Number
	x-axis	y-axis	Rank	Hypothesis	Spearman Rank	Hypothesis	Slope	Constant	in Sample
Short	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	-0.07	> 0.10	0.11	> 0.10	$0.16^{+0.10}_{-0.39}$	$-3.40^{+11.16}_{-2.77}$	9
Long	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.59	$\ll 10^{-6}$	0.59	$\ll 10^{-6}$	$0.27^{+0.04}_{-0.04}$	$-6.99^{+1.23}_{-1.10}$	237
Flares	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.58	$\ll 10^{-6}$	0.56	$\ll 10^{-6}$	$0.30^{+0.07}_{-0.06}$	$-7.91^{+1.84}_{-2.17}$	134
No Flares	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.59	$\ll 10^{-6}$	0.64	$\ll 10^{-6}$	$0.28^{+0.03}_{-0.04}$	$-7.27^{+1.28}_{-1.01}$	103
Plateau	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.58	$\ll 10^{-6}$	0.55	$\ll 10^{-6}$	$0.26^{+0.05}_{-0.06}$	$-6.81^{+1.84}_{-1.43}$	156
No Plateau	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.57	$\ll 10^{-6}$	0.61	$\ll 10^{-6}$	$0.26^{+0.06}_{-0.05}$	$-6.82^{+1.60}_{-1.83}$	81
Final	log $\rm L_{X,200s}$	$\alpha_{X,>200s}$	0.59	$\ll 10^{-6}$	0.59	$\ll 10^{-6}$	$0.27^{+0.04}_{-0.04}$	$-6.99^{+1.23}_{-1.11}$	237

3.2. *Prompt emission and afterglow parameter comparison*

We now compare the parameters of the optical/UV and X-ray log $\rm L_{200s} - \alpha_{>200s}$ correlations using the 48 GRBs that overlap both samples. The results are given in Table 2. Using the same GRBs for the optical and X-ray log $\rm L_{200s} - \alpha_{>200s}$ correlation we find the slopes of the linear regressions are consistent at 1σ. When swapping the X-ray and optical/UV luminosity and decay parameters, i.e log $\rm L_{O,200s}$ versus $\alpha_{\rm X,>200s}$ and log $\rm L_{X,200s}$ versus $\alpha_{\rm O,>200s}$ we find similar strength relationships. Strong correlations are also observed when correlating log $\rm L_{O,200s}$ vs log $\rm L_{X,200s}$ and $\alpha_{\rm O>200s}$ vs $\alpha_{\rm X,>200s}$.

In Table 2 we also provide the results of the comparison of the parameters of the optical/UV and X-ray luminosity-decay correlations with the prompt emission parameters: log $\rm E_{iso}$ and $\rm E_{peak}$. Comparison of the optical/UV and X-ray luminosity with log $\rm E_{iso}$ indicates strong correlations and the slope of the linear regressions are consistent to within 1σ. We also provide the results of the comparison of log $\rm E_{iso}$

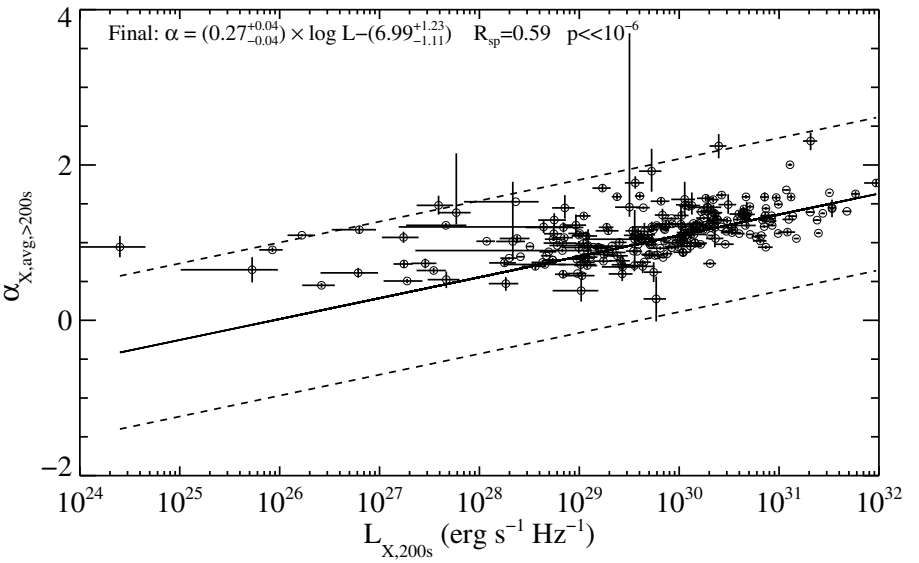


Fig. 1. Final average decay - luminosity correlation using the sample that includes flare correction and those only of long duration, with corrections and sub-sample optimisation described in §3. The solid line indicates the best fit regression, and the dashed lines indicates the 2σ deviation. This figure is reproduced from Fig. 9 of Ref. 12.

Table 2. For each pair of parameters examined, this table contains: the Spearman rank correlation coefficient with its associated null hypothesis; the coefficient of the partial Spearman rank with its associated null hypothesis; the slope and constant values provided by the best fit linear regression. We also provide the 1σ error of the Spearman rank coefficient. Table is a reproduction of Table 2 from Ref. 13.

Parameters		Spearman Rank	Null	Partial	Null	—Best fit linear regression—	
x-axis	y-axis	Coefficient	Hypothesis	Spearman Rank	Hypothesis	Slope	Constant
log $L_{O,200s}$	log $L_{X,200s}$	0.81 (0.05)	5.26×10^{-12}	0.70	2.85×10^{-8}	0.91 ± 0.22	1.04 ± 6.94
$\alpha_{O,>200s}$	$\alpha_{X,>200s}$	0.77 (0.07)	1.10×10^{-10}	0.75	1.27×10^{-9}	0.97 ± 0.10	0.25 ± 0.09
log $L_{O,200s}$	$\alpha_{O,>200s}$	0.58 (0.11)	1.90×10^{-5}	0.50	2.85×10^{-4}	0.28 ± 0.04	-7.72 ± 1.31
log $L_{X,200s}$	$\alpha_{X,>200s}$	0.69 (0.09)	8.03×10^{-8}	0.63	1.58×10^{-6}	0.26 ± 0.05	-6.71 ± 1.39
log $L_{O,200s}$	$\alpha_{X,>200s}$	0.60 (0.12)	6.87×10^{-6}	0.52	1.53×10^{-4}	0.29 ± 0.03	-8.13 ± 1.08
log $L_{X,200s}$	$\alpha_{O,>200s}$	0.65 (0.10)	5.58×10^{-7}	0.60	7.58×10^{-6}	0.32 ± 0.06	-8.70 ± 1.68
log E_{iso}	$\alpha_{O,>200s}$	0.54 (0.12)	9.05×10^{-5}	0.44	1.96×10^{-3}	0.21 ± 0.05	-10.22 ± 2.57
log E_{iso}	$\alpha_{X,>200s}$	0.57 (0.11)	3.12×10^{-5}	0.47	8.70×10^{-4}	0.21 ± 0.04	-9.60 ± 2.16
log E_{iso}	log $L_{O,200s}$	0.76 (0.06)	4.51×10^{-10}	0.66	4.59×10^{-7}	1.09 ± 0.13	-25.27 ± 6.92
log E_{iso}	log $L_{X,200s}$	0.83 (0.05)	5.04×10^{-13}	0.76	4.78×10^{-10}	1.10 ± 0.15	-27.81 ± 7.89
log E_{peak}	$\alpha_{O,>200s}$	0.45 (0.13)	2.05×10^{-3}	0.38	1.20×10^{-2}	0.48 ± 0.17	-0.22 ± 0.41
log E_{peak}	$\alpha_{X,>200s}$	0.48 (0.13)	9.22×10^{-4}	0.40	7.52×10^{-3}	0.48 ± 0.15	0.03 ± 0.36
log E_{peak}	log $L_{O\times200s}$	0.66 (0.11)	1.16×10^{-6}	0.58	3.51×10^{-5}	2.97 ± 0.76	24.53 ± 1.95
log E_{peak}	log $L_{X,200s}$	0.75 (0.10)	4.74×10^{-9}	0.70	1.38×10^{-7}	2.97 ± 0.67	22.50 ± 1.73

with $\alpha_{O,>200s}$ and $\alpha_{X,>200s}$. The Spearman rank coefficients are smaller than those found between luminosity and log E_{iso}, but do still indicate a correlation. Within 1σ the slopes of the linear regression for both the optical/UV and X-ray $\alpha_{>200s}$ versus log E_{iso} are consistent with each other. Similar results can are also found for

the four optical/UV and X-ray parameters versus E_{peak}, see Table 2. However, the relationships involving E_{peak} are weaker in comparison to the relationships observed with log E_{iso}; consistent with that found by Ref. 6.

4. Discussion

The log $L_{200s} - \alpha_{>200s}$ correlation, observed in the optical/UV and X-ray light curves, suggests that the brightest afterglows decay more quickly than the fainter afterglows. This points towards a common underlying mechanism producing the afterglow emission in the X-ray and optical/UV afterglows. We can therefore generally exclude models that invoke different emission mechanisms that would result in the log $L_{200s} - \alpha_{>200s}$ correlation being observed in only one frequency band.

Pre-*Swift* observations of late time X-ray afterglows also seemed to suggest the brightest X-ray afterglows decay more quickly than fainter afterglows,[22–24] but a larger sample including some of the first *Swift* X-ray light curves[25] was not able to support previous claims (see also Ref. 12). In this analysis, the correlation between luminosity and temporal behaviour is investigated at a much earlier time, when there is greater spread in the luminosity distribution, and the average decay index is determined using almost the entire observed afterglow.

We also have shown that the X-ray and optical/UV log L_{200s} are correlated with log E_{iso} and E_{peak}. This is consistent with previous studies (e.g. Refs. 23, 26–28), in particular Refs. 6 and 29, who performed a similar study using early X-ray luminosity, 5−10 minutes after trigger. We have also shown that the optical/UV and X-ray $\alpha_{>200s}$ are correlated with log E_{iso} and E_{peak}. Altogether, these correlations indicate that the GRBs with the brightest, fastest, decaying afterglows also have the largest observed prompt emission energies and typically larger peak spectral energy.

We now investigate if these observations are consistent with the predictions of a basic standard afterglow model; an isotropic outflow with no reverse shock or energy injection. The standard afterglow model predicts different relationships between L, α and other parameters depending on the location and ordering of the synchrotron spectral frequencies relative to the observing bands. Therefore to obtain the expected relationships between various parameters for a sample of GRBs we performed a Monte Carlo simulation. Using 10^4 trials, we simulated the optical/UV (at 1600 Å) and X-ray (at 1 keV) flux densities for 48 GRBs using equation 8 of Ref. 30 and equations 4, 5 and 6 given in Ref. 31 for $F_{\nu,max}$, ν_m and ν_c. In this simulation we assume that all GRBs are produced in a constant density medium. To compute $F_{\nu,max}$, ν_m and ν_c we needed to determine values for the microphysical parameters. These were selected at random from log-normal distributions which had 3σ intervals ranging between: 0.01-0.3 for the fraction of energy given to the electrons, ϵ_e; $5\times10^{-4}-0.5$ for the fraction of energy given to the magnetic field, ϵ_B, and $10^{-3}-10^{3}\mathrm{cm}^{-2}$ for the density of the external medium. For the electron energy index p, we centred the distribution at 2.4, as determined by Ref. 32, however, we set

the 1σ width to be 0.2 rather than 0.59. Since the closure relations fail for p values < 2, we re-selected p when $p < 2$ was selected. The value of p along with the position of ν_c relative to the observed band and redshift (selected from a uniform distribution with the range 0.5 - 4.5, a similar range as the observed sample) dictate the values of α, β and the k-correction (as given in Ref. 33).

For the 48 GRBs in each trial, we selected a prompt emission energy from a log-normal distribution with a 3σ range $10^{51} - 10^{54}$ erg; a range similar to that of the GRBs in this sample. We picked a random value between 10% and 99% for the efficiency, which we used to convert the prompt emission energy into kinetic energy. Once all the microphysical parameters, redshift and kinetic energy had been selected, we were then able to determine the position of ν_c and thus knew where it was in relation to ν_O and ν_X. With this information, we then calculated the value of the optical/UV and X-ray fluxes and converted these to luminosity. As a byproduct of calculating the optical/UV and X-ray luminosities, we also have simulated distributions for E_{iso} and α. Therefore we also produce predictions for comparisons that involve these parameters. For the parameters of 48 GRBs in each trial, we performed linear regression using the IDL routine *sixlin*, and we also calculated the Spearman rank coefficient. The results of the simulation can be found in Table 3.

Table 3. The Spearman rank coefficient and linear regression parameters as predicted by the synchrotron model for a sample of 48 GRBs. These values were computed with a Monte Carlo simulation with 10^4 trials. Table is a reproduction of Table 1 from Ref. 13.

Parameters		Simulated Spearman	—Best fit linear regression for simulation—	
x-axis	y-axis	Rank Coefficient	Slope	Constant
$\log L_{O,200s}$	$\log L_{X,200s}$	0.92 ± 0.0	0.82 ± 0.0	3.76 ± 1.25
$\alpha_{O,>200s}$	$\alpha_{X,>200s}$	0.74 ± 0.0	1.10 ± 0.1	0.04 ± 0.17
$\log L_{O,200s}$	$\alpha_{O,>200s}$	0.30 ± 0.14	0.04 ± 0.02	-0.31 ± 0.65
$\log L_{X,200s}$	$\alpha_{X,>200s}$	0.20 ± 0.14	0.04 ± 0.03	0.10 ± 0.78
$\log E_{iso}$	$\alpha_{O,>200s}$	0.06 ± 0.15	0.03 ± 0.06	-0.32 ± 2.91
$\log E_{iso}$	$\alpha_{X,>200s}$	0.09 ± 0.15	0.04 ± 0.06	-0.76 ± 3.13
$\log E_{iso}$	$\log L_{O,200s}$	0.51 ± 0.11	4.43 ± 1.03	-200.76 ± 54.10
$\log E_{iso}$	$\log L_{X,200s}$	0.54 ± 0.11	3.28 ± 0.71	-142.22 ± 37.33

In the basic standard afterglow model, the optical/UV and X-ray emission is produced by the same mechanism in an isotropic outflow, we would therefore expect to see relationships between $\log L_{O,200s}$ & $\log L_{X,200s}$ and $\alpha_{O,>200s}$ versus $\alpha_{X,>200s}$. Our observed relationships between these parameters can therefore be explained easily by the standard afterglow model and are fully consistent with the simulations. A relationship between $\log E_{iso}$ and $\log L_{200s}$ is also expected in the standard afterglow model, but the comparison of our observed relationship to the simulations suggests that the observed linear regression slope is less steep than predicted by the simulation. Furthermore, the relationships we observe, between $\log L_{200s}$ and $\alpha_{>200s}$,

and log E_{iso} and $\alpha_{>200s}$, are not expected in the standard afterglow model and are not predicted by the simulations.

Since the standard afterglow does not succeed in fully predicting all of our observed correlations, it is likely that a more complex outflow model is required. This conclusion is similar to that drawn during the separate investigation of the optical/UV log $L_{200s} - \alpha_{>200s}$ decay correlation.[11]

4.1. *Alternative Models*

There are three main possibilities that could make the outflow complex enough to be able to reproduce the observed correlations. The first is that perhaps there is some mechanism or parameter that controls the amount of energy given to and distributed during the prompt and afterglow phases and that also regulates the afterglow decay rate. This should occur in such a way that for events with the largest gamma-ray isotropic energy, the energy given to the afterglow is released quickly, resulting in an initially bright afterglow which decays rapidly. Conversely, if the gamma-ray isotropic energy is smaller, then the afterglow energy is released slowly over a longer period, the afterglow will be less bright initially and decay at a slower rate.

The second possibility is that the correlations could be a geometric effect, perhaps the result of the observer's viewing angle. Jets viewed away from the jet-axis may have fainter afterglows that decay less quickly in comparison to afterglows

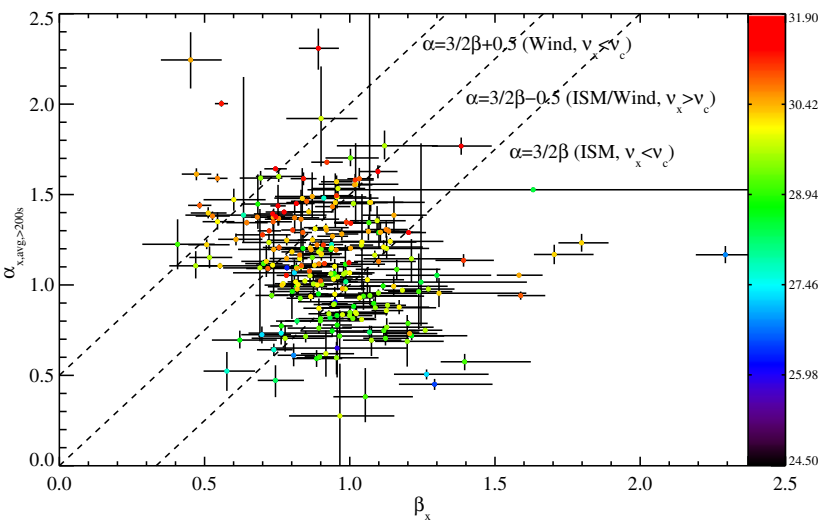


Fig. 2. The average temporal decay ($\alpha_{X,>200s}$) and average spectral energy index (β_X) are compared with log $L_{X,200s}$ (colour scale), demonstrating consistency and trends with the closure relations (dashed lines). The high luminosity (redder) points are roughly consistent with wind-like environments. This figure is reproduced from Fig. 15 of Ref. 12.

viewed closer to the centre of the jet (see Fig 3. of Ref. 4). Similarly, this will also affect the observed prompt emission, with jets viewed off-axis appearing to have lower isotropic energy and lower peak spectral energy.[34]

The third possibility could be related to the circumburst environment. The closure relations relate α and the spectral index β through different relationships depending on the ordering of the synchrotron spectral parameters and the density profile or the external medium. If the correlation is affected by the circumburst environment, we expect to see GRBs with the highest luminosities favouring a particular environment. No apparent correlation is observed in the optical/UV.[11] However in the X-ray (see Figure 2), the highest luminosity GRBs tend toward the lines demarcating the r^{-2} wind environment. The ambiguity in the $\nu_x > \nu_c$ cases prohibits us from making a strong statement on the role of circumburst environment, but it may be another possible contribution in that the initially brightest GRB afterglows may be more likely to live in wind-like environments (see also Ref. 35).

5. Conclusions

This proceeding has summarised the work presented in Refs. 12 and 13. We have shown that the correlation between luminosity (measured at restframe 200 s; log L_{200s}) and average decay rate (measured from 200 s; $\alpha_{>200s}$) is observed in the X-ray light curve sample as well as the optical/UV.[11] When we rerun the correlations with the GRBs that overlap the optical/UV and X-ray samples we find the luminosity-decay correlations are consistent. This suggests a single underlying mechanism producing the correlations in both bands and it is not dependent on their detailed temporal behaviour. We also show significant correlations between the logarithmic optical/UV and X-ray luminosity (log $L_{O,200s}$, log $L_{X,200s}$) and the optical/UV and X-ray decay indices ($\alpha_{O,>200s}$ and $\alpha_{X,>200s}$) and all four of these parameters are found to be correlated with the prompt emission parameters: isotropic energy (E_{iso}) and restframe peak spectral energy (E_{peak}). Together these correlations imply that the GRBs with the brightest afterglows in the X-ray and optical/UV bands, decay the fastest and they also have the largest observed prompt emission energies and typically larger peak spectral energy. This suggests that what happens during the prompt phase has direct implications on the afterglow.

We used a Monte Carlo simulation to examine whether the standard afterglow model is able to explain the observed correlations. Overall, observed correlations between the luminosities in both the X-ray and optical/UV bands and between the luminosities and the isotropic energy are consistent with the predictions of the simulation. However, observed relationships involving the average decay indices with either luminosity at 200 s or the isotropic γ-ray energy are not consistent with the simulation. We therefore suggest that a more complex afterglow or outflow model is required to produce all the observed correlations. This may be due to either a viewing angle effect or by some mechanism or physical property controlling the

energy release within the outflow. The environment in which the GRB exploded may also contribute to the observed correlation.

Acknowledgements

This research has made use of data obtained from the High Energy Astrophysics Science Archive Research Center (HEASARC) and the UK Swift Science Data Centre provided by NASA's Goddard Space Flight Center and the University of Leicester, UK, respectively. MJP, AAB, NPMK and PJS acknowledge the support of the UK Space Agency.

References

1. L. Amati, F. Frontera, M. Tavani, J. J. M. in't Zand, A. Antonelli, E. Costa, M. Feroci, C. Guidorzi, J. Heise, N. Masetti, E. Montanari, L. Nicastro, E. Palazzi, E. Pian, L. Piro and P. Soffitta, Intrinsic spectra and energetics of BeppoSAX Gamma-Ray Bursts with known redshifts, *Astron. Astrophys.* **390**, 81 (July 2002).
2. G. Ghirlanda, G. Ghisellini and D. Lazzati, The Collimation-corrected Gamma-Ray Burst Energies Correlate with the Peak Energy of Their νF_ν Spectrum, *Astrophys. J.* **616**, 331 (November 2004).
3. M. G. Dainotti, V. F. Cardone and S. Capozziello, A time-luminosity correlation for γ-ray bursts in the X-rays, *Mon. Not. R. Astr. Soc.* **391**, L79 (November 2008).
4. A. Panaitescu and W. T. Vestrand, Taxonomy of gamma-ray burst optical light curves: Identification of a salient class of early afterglows, *Mon. Not. R. Astr. Soc.* **387**, 497 (June 2008).
5. M. G. Bernardini, R. Margutti, J. Mao, E. Zaninoni and G. Chincarini, The X-ray light curve of gamma-ray bursts: Clues to the central engine, *Astron. Astrophys.* **539**, p. A3 (March 2012).
6. P. D'Avanzo, R. Salvaterra, B. Sbarufatti, L. Nava, A. Melandri, M. G. Bernardini, S. Campana, S. Covino, D. Fugazza, G. Ghirlanda, G. Ghisellini, V. La Parola, M. Perri, S. D. Vergani and G. Tagliaferri, A complete sample of bright Swift Gamma-ray bursts: X-ray afterglow luminosity and its correlation with the prompt emission, *Mon. Not. R. Astr. Soc.* **425**, 506 (September 2012).
7. L. Li, E.-W. Liang, Q.-W. Tang, J.-M. Chen, S.-Q. Xi, H.-J. Lü, H. Gao, B. Zhang, J. Zhang, S.-X. Yi, R.-J. Lu, L.-Z. Lü and J.-Y. Wei, A Comprehensive Study of Gamma-Ray Burst Optical Emission. I. Flares and Early Shallow-decay Component, *Astrophys. J.* **758**, p. 27 (October 2012).
8. E.-W. Liang, L. Li, H. Gao, B. Zhang, Y.-F. Liang, X.-F. Wu, S.-X. Yi, Z.-G. Dai, Q.-W. Tang, J.-M. Chen, H.-J. Lü, J. Zhang, R.-J. Lu, L.-Z. Lü and J.-Y. Wei, A Comprehensive Study of Gamma-Ray Burst Optical Emission. II. Afterglow Onset and Late Re-brightening Components, *Astrophys. J.* **774**, p. 13 (September 2013).
9. E. Zaninoni, M. G. Bernardini, R. Margutti, S. Oates and G. Chincarini, Gamma-ray burst optical light-curve zoo: Comparison with X-ray observations, *Astron. Astrophys.* **557**, p. A12 (September 2013).
10. A. Panaitescu, W. T. Vestrand and P. Woźniak, Peaks of optical and X-ray afterglow light curves, *Mon. Not. R. Astr. Soc.* **433**, 759 (July 2013).
11. S. R. Oates, M. J. Page, M. De Pasquale, P. Schady, A. A. Breeveld, S. T. Holland, N. P. M. Kuin and F. E. Marshall, A correlation between the intrinsic brightness and

average decay rate of Swift/UVOT gamma-ray burst optical/ultraviolet light curves, *Mon. Not. R. Astr. Soc.* **426**, L86 (October 2012).
12. J. L. Racusin, S. R. Oates, M. de Pasquale and D. Kocevski, A Correlation between the Intrinsic Brightness and Average Decay Rate of Gamma-Ray Burst X-Ray Afterglow Light Curves, *Astrophys. J.* **826**, p. 45 (July 2016).
13. S. R. Oates, J. L. Racusin, M. De Pasquale, M. J. Page, A. J. Castro-Tirado, J. Gorosabel, P. J. Smith, A. A. Breeveld and N. P. M. Kuin, Exploring the canonical behaviour of long gamma-ray bursts using an intrinsic multiwavelength afterglow correlation, *Mon. Not. R. Astr. Soc.* **453**, 4121 (November 2015).
14. D. N. Burrows, J. E. Hill, J. A. Nousek, J. A. Kennea, A. Wells, J. P. Osborne, A. F. Abbey, A. Beardmore, K. Mukerjee, A. D. T. Short, G. Chincarini, S. Campana, O. Citterio, A. Moretti, C. Pagani, G. Tagliaferri, P. Giommi, M. Capalbi, F. Tamburelli, L. Angelini, G. Cusumano, H. W. Bräuninger, W. Burkert and G. D. Hartner, The Swift X-Ray Telescope, *Space Science Reviews* **120**, 165 (October 2005).
15. P. A. Evans, A. P. Beardmore, K. L. Page, L. G. Tyler, J. P. Osborne, M. R. Goad, P. T. O'Brien, L. Vetere, J. Racusin, D. Morris, D. N. Burrows, M. Capalbi, M. Perri, N. Gehrels and P. Romano, An online repository of Swift/XRT light curves of γ-ray bursts, *Astron. Astrophys.* **469**, 379 (July 2007).
16. P. A. Evans, A. P. Beardmore, K. L. Page, J. P. Osborne, P. T. O'Brien, R. Willingale, R. L. C. Starling, D. N. Burrows, O. Godet, L. Vetere, J. Racusin, M. R. Goad, K. Wiersema, L. Angelini, M. Capalbi, G. Chincarini, N. Gehrels, J. A. Kennea, R. Margutti, D. C. Morris, C. J. Mountford, C. Pagani, M. Perri, P. Romano and N. Tanvir, Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs, *Mon. Not. R. Astr. Soc.* **397**, 1177 (August 2009).
17. J. L. Racusin, E. W. Liang, D. N. Burrows, A. Falcone, T. Sakamoto, B. B. Zhang, B. Zhang, P. Evans and J. Osborne, Jet Breaks and Energetics of Swift Gamma-Ray Burst X-Ray Afterglows, *Astrophys. J.* **698**, 43 (June 2009).
18. B. Zhang, Y. Z. Fan, J. Dyks, S. Kobayashi, P. Mészáros, D. N. Burrows, J. A. Nousek and N. Gehrels, Physical Processes Shaping Gamma-Ray Burst X-Ray Afterglow Light Curves: Theoretical Implications from the Swift X-Ray Telescope Observations, *Astrophys. J.* **642**, 354 (May 2006).
19. D. Grupe, C. Gronwall, X. Wang, P. W. A. Roming, J. Cummings, B. Zhang, P. Mészáros, M. D. Trigo, P. T. O'Brien, K. L. Page, A. Beardmore, O. Godet, D. E. vanden Berk, P. J. Brown, S. Koch, D. Morris, M. Stroh, D. N. Burrows, J. A. Nousek, M. McMath Chester, S. Immler, V. Mangano, P. Romano, G. Chincarini, J. Osborne, T. Sakamoto and N. Gehrels, Swift and XMM-Newton Observations of the Extraordinary Gamma-Ray Burst 060729: More than 125 Days of X-Ray Afterglow, *Astrophys. J.* **662**, 443 (June 2007).
20. M. G. Dainotti, R. Willingale, S. Capozziello, V. Fabrizio Cardone and M. Ostrowski, Discovery of a Tight Correlation for Gamma-ray Burst Afterglows with "Canonical" Light Curves, *Astrophys. J. Letters* **722**, L215 (October 2010).
21. M. G. Dainotti, V. Petrosian, J. Singal and M. Ostrowski, Determination of the Intrinsic Luminosity Time Correlation in the X-Ray Afterglows of Gamma-Ray Bursts, *Astrophys. J.* **774**, p. 157 (September 2013).
22. M. Boër and B. Gendre, Evidences for two Gamma-Ray Burst afterglow emission regimes, *Astron. Astrophys.* **361**, L21 (September 2000).
23. C. Kouveliotou, S. E. Woosley, S. K. Patel, A. Levan, R. Blandford, E. Ramirez-Ruiz, R. A. M. J. Wijers, M. C. Weisskopf, A. Tennant, E. Pian and P. Giommi, Chandra Observations of the X-Ray Environs of SN 1998bw/GRB 980425, *Astrophys. J.* **608**, 872 (June 2004).

24. B. Gendre and M. Boër, Decay properties of the X-ray afterglows of gamma-ray bursts, *Astron. Astrophys.* **430**, 465 (February 2005).
25. B. Gendre, A. Galli and M. Boër, X-Ray Afterglow Light Curves: Toward A Standard Candle?, *Astrophys. J.* **683**, 620 (August 2008).
26. M. De Pasquale, L. Piro, B. Gendre, L. Amati, L. A. Antonelli, E. Costa, M. Feroci, F. Frontera, L. Nicastro, P. Soffitta and J. in't Zand, The BeppoSAX catalog of GRB X-ray afterglow observations, *Astron. Astrophys.* **455**, 813 (September 2006).
27. M. Nysewander, A. S. Fruchter and A. Pe'er, A Comparison of the Afterglows of Short- and Long-duration Gamma-ray Bursts, *Astrophys. J.* **701**, 824 (August 2009).
28. D. A. Kann, S. Klose, B. Zhang, D. Malesani, E. Nakar, A. Pozanenko, A. C. Wilson, N. R. Butler, P. Jakobsson, S. Schulze, M. Andreev, L. A. Antonelli, I. F. Bikmaev, V. Biryukov, M. Böttcher, R. A. Burenin, J. M. Castro Cerón, A. J. Castro-Tirado, G. Chincarini, B. E. Cobb, S. Covino, P. D'Avanzo, V. D'Elia, M. Della Valle, A. de Ugarte Postigo, Y. Efimov, P. Ferrero, D. Fugazza, J. P. U. Fynbo, M. Gålfalk, F. Grundahl, J. Gorosabel, S. Gupta, S. Guziy, B. Hafizov, J. Hjorth, K. Holhjem, M. Ibrahimov, M. Im, G. L. Israel, M. Jeĺinek, B. L. Jensen, R. Karimov, I. M. Khamitov, Ü. Kiziloğlu, E. Klunko, P. Kubánek, A. S. Kutyrev, P. Laursen, A. J. Levan, F. Mannucci, C. M. Martin, A. Mescheryakov, N. Mirabal, J. P. Norris, J.-E. Ovaldsen, D. Paraficz, E. Pavlenko, S. Piranomonte, A. Rossi, V. Rumyantsev, R. Salinas, A. Sergeev, D. Sharapov, J. Sollerman, B. Stecklum, L. Stella, G. Tagliaferri, N. R. Tanvir, J. Telting, V. Testa, A. C. Updike, A. Volnova, D. Watson, K. Wiersema and D. Xu, The Afterglows of Swift-era Gamma-ray Bursts. I. Comparing pre-Swift and Swift-era Long/Soft (Type II) GRB Optical Afterglows, *Astrophys. J.* **720**, 1513 (September 2010).
29. R. Margutti, E. Zaninoni, M. G. Bernardini, G. Chincarini, F. Pasotti, C. Guidorzi, L. Angelini, D. N. Burrows, M. Capalbi, P. A. Evans, N. Gehrels, J. Kennea, V. Mangano, A. Moretti, J. Nousek, J. P. Osborne, K. L. Page, M. Perri, J. Racusin, P. Romano, B. Sbarufatti, S. Stafford and M. Stamatikos, The prompt-afterglow connection in gamma-ray bursts: a comprehensive statistical analysis of Swift X-ray light curves, *Mon. Not. R. Astr. Soc.* **428**, 729 (January 2013).
30. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *Astrophys. J. Letters* **497**, L17 (April 1998).
31. B. Zhang, E. Liang, K. L. Page, D. Grupe, B.-B. Zhang, S. D. Barthelmy, D. N. Burrows, S. Campana, G. Chincarini, N. Gehrels, S. Kobayashi, P. Mészáros, A. Moretti, J. A. Nousek, P. T. O'Brien, J. P. Osborne, P. W. A. Roming, T. Sakamoto, P. Schady and R. Willingale, GRB Radiative Efficiencies Derived from the Swift Data: GRBs versus XRFs, Long versus Short, *Astrophys. J.* **655**, 989 (February 2007).
32. P. A. Curran, R. L. C. Starling, A. J. van der Horst and R. A. M. J. Wijers, Testing the blast wave model with Swift GRBs, *Mon. Not. R. Astr. Soc.* **395**, 580 (May 2009).
33. E. Berger, S. R. Kulkarni and D. A. Frail, A Standard Kinetic Energy Reservoir in Gamma-Ray Burst Afterglows, *Astrophys. J.* **590**, 379 (June 2003).
34. E. Ramirez-Ruiz, J. Granot, C. Kouveliotou, S. E. Woosley, S. K. Patel and P. A. Mazzali, An Off-Axis Model of GRB 031203, *Astrophys. J. Letters* **625**, L91 (June 2005).
35. M. De Pasquale, S. Schulze, D. A. Kann, S. Oates and B. Zhang, Physical properties of rapidly decaying Afterglows, in *EAS Publications Series*, eds. A. J. Castro-Tirado, J. Gorosabel and I. H. Park, EAS Publications Series, Vol. 61, July 2013.

Kilonova emission observed so far: A comparison with AT2017gfo

A. Rossi

INAF - Osservatorio di Astrofisica e Scienza dello Spazio, via Piero Gobetti 93/3, 40129 Bologna, Italy

E-mail: andrea.rossi@inaf.it

AT2017gfo is the first kilonova (KN) that could be extensively monitored in time both photometrically and spectroscopically. Moreover, it is the first optical counterpart of a gravitational wave source and it is associated with the short gamma-ray burst GRB 170817A. Here I present our search for the fingerprints of AT2017gfo-like kilonova emissions in the optical/NIR light curves of 39 short GRBs with known redshift. Afterwards, I show how, for the first time, our results allow us to study separately the range of luminosity of the blue and red components of AT2017gfo-like kilonovae in short GRBs. With these results at hand, I show up to which redshift a KN can be followed up by some of the current and future observatories.

Keywords: Kilonovae; Gamma-ray bursts; Neutron stars

1. Introduction

For a long time only indirect evidence associated short Gamma-ray bursts (GRBs) to the merging of compact objects, but the simultaneous detection of the gravitational wave (GW) source GW170817[1] by aLIGO/AdVirgo[2,3] and the short Gamma-ray burst GRB 170817A[4–6] has provided the first direct evidence that at least a fraction of short GRBs is associated with the merging of two neutron stars (NSs). At the same time, the discovery of AT2017gfo, the optical counterpart of GW170817,[7] allowed us, for the first time, to follow-up spectroscopically the elusive astrophysical phenomena known as "kilonova",[8,9] i.e. a thermal emission powered by the radioactive decay of elements formed via r-process nucleosynthesis in the ejecta of the NS-NS merger e.g.,[10,11]. In particular, the observations are consistent with a kilonova characterised by a blue, rapidly decaying, component and a red, more slowly evolving one. Therefore, one can distinguish between the *blue* and *red* components, depending on whether the rest-frame effective wavelength is brighter in the optical or in the NIR, respectively.

In the following, we aim to further investigate the possible range of kilonova luminosity and identify possible new kilonova candidates, comparing the optical/NIR light curves of all short GRBs (sGRBs) with known redshift up to June 2019 with those of AT2017gfo.

2. Kilonovae known to date

Although AT2017gfo is the most outstanding and clear demonstration of the existence of kilonovae, the majority of the known kilonovae are associated to sGRBs, although often the association is not very strong. Among those, the most clear evidence was found in the optical counterparts of 6 sGRBs: 050709A,[12] 060614A,[13,14] 080503,[15,16] 130603B,[17,18] 150101B,[19] 160821B,[20–23] and 070809.[24] In addition to those, there are the magnetar-powered kilonovae identified by Gao *et al.*[16] (GRBs 050724, 070714B). Unfortunately, in the case of GRB 080503 the redshift could not be measured either via spectroscopy of the afterglow or of the host galaxy; in the case of GRB 070809 the redshift is not well defined. Recently, a new KN candidate has been found associated to GRB 200222A, although this claim is not secure.[25,26] Therefore, we define a golden sample that includes all 7 GRBs with kilonova candidates claimed in the literature that have accurate redshift.

Note that the evidence of kilonova in GRBs 050724 070714B has been found by modelling the X-rays and optical light curves with magnetar-powered KN model. These magnetar KN (MKN) are expected to be more luminous, have longer duration but also bluer than the the classical KN. It should also be noted that in all these possible kilonova identifications (but GRB 150101B), the emission was preceded by a bright GRB afterglow indicating an on-axis configuration, thus suggesting that the kilonova emission may exceed the afterglow luminosity even for on-axis GRBs.

3. Optical/NIR light curves of AT2017gfo and of short GRB afterglows

The sample of short GRBs consists of 39 objects within the redshift range $0.1 \leq z \leq 2.2$.[27] As described in,[27] to have a coverage from UV to NIR bands between $\sim$1.5 and $\sim$10.5 days after the GW trigger we built 3 early SEDs using photometric observations from the literature[7,28–34] and ten spectra of AT2017gfo[33,35] Data have been corrected for foreground Galactic extinction. In order to compare AT2017gfo with sGRB optical counterparts, we first built a set of rest-frame light curves of AT2017gfo for all the filters in which there are sGRB observations. Afterwards, we compared, in each filter, the optical and NIR luminosities of AT2017gfo with those of the sGRB counterparts in our selected sample.[27] We have distinguished between the *blue* and *red* components, depending on whether the rest-frame effective wavelength is below or above 9000 Å, respectively.

4. Possible new kilonova candidates

A kilonova is expected to show a shallow evolution close to its maximum brightness. Therefore, it can be distinguished from the standard afterglow behaviour, which has a constant power-law decay according to the standard fireball model.[36] In this context we considered a decay to be anomalously shallow when its decaying index is $\alpha < 0.75$; see also.[27] The blue component of GRB 090515 has a shallow decay

in the first 24 hours and is just $\sim$ 2 times fainter than AT2017gfo at the same time. This is also true in the case of GRB 080905A, and thus may be compatible with a kilonova with a luminosity similar to the blue counterpart of GRB 090515. However, the error of the decay index is too large ($\alpha = 0.4 \pm 1.3$) and thus we do not consider this sGRB. We also identify possible shallow decay in the cases of 150423A and 150424A which however can be interpreted otherwise. In particular, the shallow decay of GRB 150423A is too early to be compared with AT2017gfo. A more detailed analysis, which consists in modelling together the optical and X-ray light curves, is needed to further investigate the presence of kilonova emission in all cases.

5. Constraints to the kilonovae luminosity

Considering the blue component spectral band, we observe a wide range in luminosity. In particular, we can exclude the presence of an AT2017gfo-like kilonova in two cases (GRBs 050509B and 061201). Few cases with an anomalous shallow decay, namely GRBs 050724, 060614, 070714B, and 150424A, are more than 10 times brighter than AT2017gfo. Note that in the case of GRBs 150423A and 150424A the optical and X-ray light curves are very similar to those of GRB 050724 and GRB 060614, which optical re-brightening is interpreted as magnetar-powered kilonova emission by.[16] In this case, the additional source of energy would explain why they are all much brighter than AT2017gfo.

On the contrary, all the kilonovae detected in the NIR (GRBs 050709, 130603B, 160821B), are no more than 3 times brighter or fainter than AT2017gfo after 2 days. Intriguingly, there are no NIR upper limits below AT2017gfo luminosity after ~ 50 hours. In other words, in all cases when observations comparable to AT2017gfo NIR emission exist the kilonova counterpart has been detected. This suggests that all red kilonovae detected so far have similar luminosity, although we are aware that the numbers are not high enough for meaningful statistics.

6. Future perspectives

In Figure 1 I show how the peak luminosity of the blue KN changes with redshift, assuming the luminosity constraints found by[27] and summarized above. It shows that the search and follow-up of the electromagnetic (EM) counterparts will be possible during O4 and O5, like what has been done in the case GW 170817, provided hat the whole sky-map can be covered down to r$\sim$ 21 mag. Facilities like, VST, will be advantaged, but should concentrate on the smaller skymaps of tens of sq. degrees, or the regions of higher probability, to maximize the return and their gain in depth.

The Binary neutron star horizon expected for the 3rd generation of GW interferometers will extend above z=1, with the bulk of events concentrated at $z < 0.5$. Although the increment in distance translates in a huge increment in the number of GW events, including BNS and their EM counterparts, the localization will not

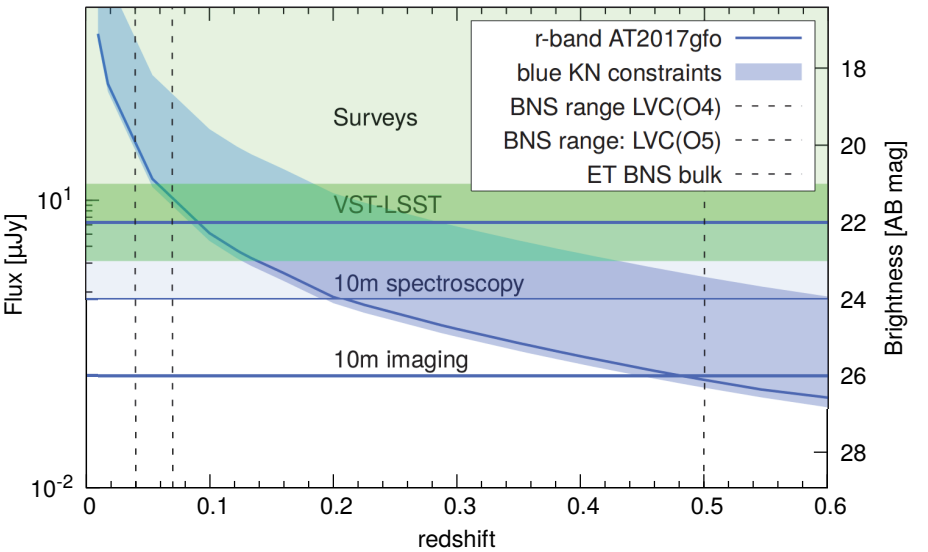


Fig. 1. Peak brightness of AT2017gfo in the r-band at different redshifts, within the constraints we derived for an AT2017gfo-like KN from the golden sample. The vertical line is the aLIGO/AVirgo detection limits for a BNS event during O4, O5 and ET. The last one is limited at $z \lesssim 0.5$ at which ET will be able to detect most of the GW signals from merging BNS. The horizontal lines are different detection limits for different class of telescopes with an exposure time of 10 min.

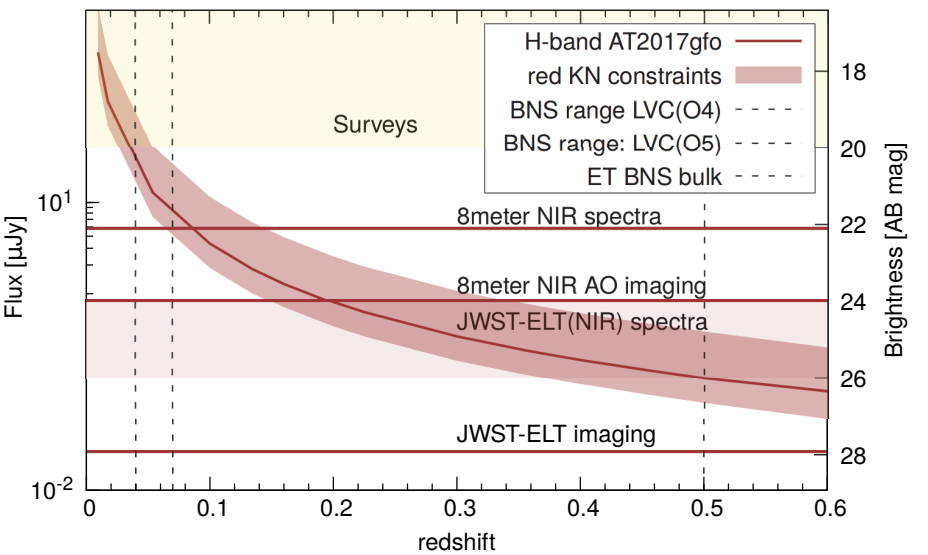


Fig. 2. The same as shown in Figure 1 but in the H-band.

improve much in respect to the present situation during LVK runs: 50% of all events will have localized with error of more than 1000 sq degrees, and only below redshift 0.2 at least 50% of events will be localized below 100 sq degrees. In this situation, what are the chances to find, follow-up and characterize the EM counterparts?

A shown in Figure 1, below redshift 0.2 the search of the EM counterpart can still be succesful thanks to the most advanced surveys if they capabilities and strategy will be adjusted to reach r 21 mags and beyond, perhaps sacrificing sky covergae to optimize depths, when this is possible. In particular, the VST telescope, or even better the Vera Rubin telescope, will be able to localize most events, or at list the brightest events. The search becomes difficult at higher redshifts, with only Vera Rubin able to successfully find the brightest events. On the contrary, once localized, the follow-up and characterization via spectroscopy will not be a problem, as the class of the 10m telescope has enough sensitivity and is well equipped for the follow-up.

More critical is the situation in the NIR, where no many facilities may be able to reach the necessary depth to successfully search and localize the EM counterpart. Still, the characterization will be guaranteed by the class of 8-10m and extremely large telescopes, together with the Nancy Roman space observatory. This situation is even more dramatic dramatic if we think that the blue kilonova is expected only if there is no prompt collapse in a black hole, and therefore the characteristic feature in common to both cases i.e., the red kilonova is the only that is always expected. One solution may be offered by a space observatory capable to detect the gama-ray signal and slew in the direction of the burst to localize its afterglow, like the current *Swift*, the future *SVOM* or the proposed *THESEUS* and *GAMOW* missions. In this case the bright afterglow can be localized up to high redshift. The only limitation is that only on-axis (or close) events will be followed up, being the emission collimated. However, it also means that all the sGRBs at $z < 0.5$ and many at $z < 1$ will be associated to a gravitational wave signal, allowing to measure both distance and redshift, constrain both masses and inclinations with implications n the cosmological studies and equation of state of the progenitor NS.

References

1. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Physical Review Letters* **119**, p. 161101 (October 2017).
2. LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso and et al., Advanced LIGO, *Classical and Quantum Gravity* **32**, p. 074001 (April 2015).
3. F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin and et al., Advanced Virgo: A second-generation interferometric gravitational wave detector, *Classical and Quantum Gravity* **32**, p. 024001 (January 2015).

4. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A, *Astroph. J.* **848**, p. L14 (October 2017).
5. V. Savchenko, C. Ferrigno, E. Kuulkers, A. Bazzano, E. Bozzo, S. Brandt, J. Chenevez, T. J.-L. Courvoisier, R. Diehl, A. Domingo, L. Hanlon, E. Jourdain, A. von Kienlin, P. Laurent, F. Lebrun, A. Lutovinov, A. Martin-Carrillo, S. Mereghetti, L. Natalucci, J. Rodi, J.-P. Roques, R. Sunyaev and P. Ubertini, INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817, *Astroph. J.* **848**, p. L15 (October 2017).
6. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya and et al., Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *Astroph. J.* **848**, p. L13 (October 2017).
7. D. A. Coulter, R. J. Foley, C. D. Kilpatrick, M. R. Drout, A. L. Piro, B. J. Shappee, M. R. Siebert, J. D. Simon, N. Ulloa, D. Kasen, B. F. Madore, A. Murguia-Berthier, Y.-C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest and C. Rojas-Bravo, Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source, *Science* **358**, 1556 (December 2017).
8. L.-X. Li and B. Paczyński, Transient Events from Neutron Star Mergers, *Astroph. J.* **507**, L59 (November 1998).
9. B. D. Metzger, G. Martínez-Pinedo, S. Darbha, E. Quataert, A. Arcones, D. Kasen, R. Thomas, P. Nugent, I. V. Panov and N. T. Zinner, Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei, *Mon. Not. R. Astron. Soc.* **406**, 2650 (Aug 2010).
10. D. Kasen, B. Metzger, J. Barnes, E. Quataert and E. Ramirez-Ruiz, Origin of the heavy elements in binary neutron-star mergers from a gravitational-wave event, *Nature* **551**, 80 (November 2017).
11. B. D. Metzger, T. A. Thompson and E. Quataert, A Magnetar Origin for the Kilonova Ejecta in GW170817, *Astroph. J.* **856**, p. 101 (April 2018).
12. Z.-P. Jin, K. Hotokezaka, X. Li, M. Tanaka, P. D'Avanzo, Y.-Z. Fan, S. Covino, D.-M. Wei and T. Piran, The Macronova in GRB 050709 and the GRB-macronova connection, *Nature Communications* **7**, p. 12898 (September 2016).
13. B. Yang, Z.-P. Jin, X. Li, S. Covino, X.-Z. Zheng, K. Hotokezaka, Y.-Z. Fan, T. Piran and D.-M. Wei, A possible macronova in the late afterglow of the long-short burst GRB 060614, *Nature Communications* **6**, p. 7323 (June 2015).
14. Z.-P. Jin, X. Li, Z. Cano, S. Covino, Y.-Z. Fan and D.-M. Wei, The Light Curve of the Macronova Associated with the Long-Short Burst GRB 060614, *Astroph. J.* **811**, p. L22 (October 2015).
15. D. A. Perley, B. D. Metzger, J. Granot, N. R. Butler, T. Sakamoto, E. Ramirez-Ruiz, A. J. Levan, J. S. Bloom, A. A. Miller, A. Bunker, H.-W. Chen, A. V. Filippenko, N. Gehrels, K. Glazebrook, P. B. Hall, K. C. Hurley, D. Kocevski, W. Li, S. Lopez, J. Norris, A. L. Piro, D. Poznanski, J. X. Prochaska, E. Quataert and N. Tanvir, GRB 080503: Implications of a Naked Short Gamma-Ray Burst Dominated by Extended Emission, *Astroph. J.* **696**, 1871 (May 2009).
16. H. Gao, B. Zhang, H.-J. Lü and Y. Li, Searching for Magnetar-powered Merger-novae from Short GRBS, *Astroph. J.* **837**, p. 50 (March 2017).

17. N. R. Tanvir, A. J. Levan, A. S. Fruchter, J. Hjorth, R. A. Hounsell, K. Wiersema and R. L. Tunnicliffe, A 'kilonova' associated with the short-duration γ-ray burst GRB 130603B, *Nature* **500**, 547 (August 2013).
18. E. Berger, W. Fong and R. Chornock, An r-process Kilonova Associated with the Short-hard GRB 130603B, *Astroph. J.* **774**, p. L23 (Sep 2013).
19. E. Troja, L. Piro, G. Ryan, H. van Eerten, R. Ricci, M. H. Wieringa, S. Lotti, T. Sakamoto and S. B. Cenko, The outflow structure of GW170817 from late-time broad-band observations, *Mon. Not. R. Astron. Soc.* **478**, L18 (July 2018).
20. Z.-P. Jin, X. Li, H. Wang, Y.-Z. Wang, H.-N. He, Q. Yuan, F.-W. Zhang, Y.-C. Zou, Y.-Z. Fan and D.-M. Wei, Short GRBs: Opening Angles, Local Neutron Star Merger Rate, and Off-axis Events for GRB/GW Association, *Astroph. J.* **857**, p. 128 (April 2018).
21. M. M. Kasliwal, O. Korobkin, R. M. Lau, R. Wollaeger and C. L. Fryer, Infrared Emission from Kilonovae: The Case of the Nearby Short Hard Burst GRB 160821B, *Astroph. J.* **843**, p. L34 (July 2017).
22. E. Troja, A. J. Castro-Tirado, J. Becerra González, Y. Hu, G. S. Ryan, S. B. Cenko, R. Ricci, G. Novara, R. Sánchez-Rámirez, J. A. Acosta-Pulido, K. D. Ackley, M. D. Caballero García, S. S. Eikenberry, S. Guziy, S. Jeong, A. Y. Lien, I. Márquez, S. B. Pand ey, I. H. Park, T. Sakamoto, J. C. Tello, I. V. Sokolov, V. V. Sokolov, A. Tiengo, A. F. Valeev, B. B. Zhang and S. Veilleux, The afterglow and kilonova of the short GRB 160821B, *Mon. Not. R. Astron. Soc.* **489**, 2104 (Oct 2019).
23. G. P. Lamb, N. R. Tanvir, A. J. Levan, A. de Ugarte Postigo, K. Kawaguchi, A. Corsi, P. A. Evans, B. Gompertz, D. B. Malesani, K. L. Page, K. Wiersema, S. Rosswog, M. Shibata, M. Tanaka, A. J. van der Horst, Z. Cano, J. P. U. Fynbo, A. S. Fruchter, J. Greiner, K. E. Heintz, A. Higgins, J. Hjorth, L. Izzo, P. Jakobsson, D. A. Kann, P. T. O'Brien, D. A. Perley, E. Pian, G. Pugliese, R. L. C. Starling, C. C. Thöne, D. Watson, R. A. M. J. Wijers and D. Xu, Short GRB 160821B: A Reverse Shock, a Refreshed Shock, and a Well-sampled Kilonova, *Astroph. J.* **883**, p. 48 (Sep 2019).
24. Z.-P. Jin, S. Covino, N.-H. Liao, X. Li, P. D'Avanzo, Y.-Z. Fan and D.-M. Wei, A kilonova associated with GRB 070809, *Nature Astronomy* , p. 461 (Oct 2019).
25. W. Fong, T. Laskar, J. Rastinejad, A. R. Escorial, G. Schroeder, J. Barnes, C. D. Kilpatrick, K. Paterson, E. Berger, B. D. Metzger, Y. Dong, A. E. Nugent, R. Strausbaugh, P. K. Blanchard, A. Goyal, A. Cucchiara, G. Terreran, K. D. Alexander, T. Eftekhari, C. Fryer, B. Margalit, R. Margutti and M. Nicholl, The Broadband Counterpart of the Short GRB 200522A at z = 0.5536: A Luminous Kilonova or a Collimated Outflow with a Reverse Shock?, *Astroph. J.* **906**, p. 127 (January 2021).
26. B. O'Connor, E. Troja, S. Dichiara, E. A. Chase, G. Ryan, S. B. Cenko, C. L. Fryer, R. Ricci, F. Marshall, C. Kouveliotou, R. T. Wollaeger, C. J. Fontes, O. Korobkin, P. Gatkine, A. Kutyrev, S. Veilleux, N. Kawai and T. Sakamoto, A tale of two mergers: constraints on kilonova detection in two short GRBs at z $\sim$ 0.5, *Mon. Not. R. Astron. Soc.* **502**, 1279 (March 2021).
27. A. Rossi, G. Stratta, E. Maiorano, D. Spighi, N. Masetti, E. Palazzi, A. Gardini, A. Melandri, L. Nicastro, E. Pian, M. Branchesi, M. Dadina, V. Testa, S. Brocato, S. Benetti, R. Ciolfi, S. Covino, V. D'Elia, A. Grado, L. Izzo, A. Perego, S. Piranomonte, R. Salvaterra, J. Selsing, L. Tomasella, S. Yang, D. Vergani, L. Amati and J. B. Stephen, A comparison between short GRB afterglows and AT2017gfo: shedding light on kilonovae properties, *arXiv e-prints* , p. arXiv:1901.05792 (Jan 2019).
28. N. R. Tanvir, A. J. Levan, C. González-Fernández, O. Korobkin, I. Mandel, S. Rosswog, J. Hjorth, P. D'Avanzo, A. S. Fruchter, C. L. Fryer, T. Kangas, B. Milvang-Jensen, S. Rosetti, D. Steeghs, R. T. Wollaeger, Z. Cano, C. M. Copperwheat,

S. Covino, V. D'Elia, A. de Ugarte Postigo, P. A. Evans, W. P. Even, S. Fairhurst, R. Figuera Jaimes, C. J. Fontes, Y. I. Fujii, J. P. U. Fynbo, B. P. Gompertz, J. Greiner, G. Hodosan, M. J. Irwin, P. Jakobsson, U. G. Jørgensen, D. A. Kann, J. D. Lyman, D. Malesani, R. G. McMahon, A. Melandri, P. T. O'Brien, J. P. Osborne, E. Palazzi, D. A. Perley, E. Pian, S. Piranomonte, M. Rabus, E. Rol, A. Rowlinson, S. Schulze, P. Sutton, C. C. Thöne, K. Ulaczyk, D. Watson, K. Wiersema and R. A. M. J. Wijers, The Emergence of a Lanthanide-rich Kilonova Following the Merger of Two Neutron Stars, *Astroph. J.* **848**, p. L27 (October 2017).

29. M. R. Drout, A. L. Piro, B. J. Shappee, C. D. Kilpatrick, J. D. Simon, C. Contreras, D. A. Coulter, R. J. Foley, M. R. Siebert, N. Morrell, K. Boutsia, F. Di Mille, T. W.-S. Holoien, D. Kasen, J. A. Kollmeier, B. F. Madore, A. J. Monson, A. Murguia-Berthier, Y.-C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest, C. Adams, K. Alatalo, E. Bañados, J. Baughman, T. C. Beers, R. A. Bernstein, T. Bitsakis, A. Campillay, T. T. Hansen, C. R. Higgs, A. P. Ji, G. Maravelias, J. L. Marshall, C. M. Bidin, J. L. Prieto, K. C. Rasmussen, C. Rojas-Bravo, A. L. Strom, N. Ulloa, J. Vargas-González, Z. Wan and D. D. Whitten, Light curves of the neutron star merger GW170817/SSS17a: Implications for r-process nucleosynthesis, *Science* **358**, 1570 (December 2017).
30. P. A. Evans, S. B. Cenko, J. A. Kennea, S. W. K. Emery, N. P. M. Kuin, O. Korobkin, R. T. Wollaeger, C. L. Fryer, K. K. Madsen, F. A. Harrison, Y. Xu, E. Nakar, K. Hotokezaka, A. Lien, S. Campana, S. R. Oates, E. Troja, A. A. Breeveld, F. E. Marshall, S. D. Barthelmy, A. P. Beardmore, D. N. Burrows, G. Cusumano, A. D'Aì, P. D'Avanzo, V. D'Elia, M. de Pasquale, W. P. Even, C. J. Fontes, K. Forster, J. Garcia, P. Giommi, B. Grefenstette, C. Gronwall, D. H. Hartmann, M. Heida, A. L. Hungerford, M. M. Kasliwal, H. A. Krimm, A. J. Levan, D. Malesani, A. Melandri, H. Miyasaka, J. A. Nousek, P. T. O'Brien, J. P. Osborne, C. Pagani, K. L. Page, D. M. Palmer, M. Perri, S. Pike, J. L. Racusin, S. Rosswog, M. H. Siegel, T. Sakamoto, B. Sbarufatti, G. Tagliaferri, N. R. Tanvir and A. Tohuvavohu, Swift and NuSTAR observations of GW170817: Detection of a blue kilonova, *Science* **358**, 1565 (December 2017).
31. S. Covino, K. Wiersema, Y. Z. Fan, K. Toma, A. B. Higgins, A. Melandri, P. D'Avanzo, C. G. Mundell, E. Palazzi, N. R. Tanvir, M. G. Bernardini, M. Branchesi, E. Brocato, S. Campana, S. di Serego Alighieri, D. Götz, J. P. U. Fynbo, W. Gao, A. Gomboc, B. Gompertz, J. Greiner, J. Hjorth, Z. P. Jin, L. Kaper, S. Klose, S. Kobayashi, D. Kopac, C. Kouveliotou, A. J. Levan, J. Mao, D. Malesani, E. Pian, A. Rossi, R. Salvaterra, R. L. C. Starling, I. Steele, G. Tagliaferri, E. Troja, A. J. van der Horst and R. A. M. J. Wijers, The unpolarized macronova associated with the gravitational wave event GW 170817, *Nature Astronomy* **1**, 791 (November 2017).
32. E. Troja, L. Piro, H. van Eerten, R. T. Wollaeger, M. Im, O. D. Fox, N. R. Butler, S. B. Cenko, T. Sakamoto, C. L. Fryer, R. Ricci, A. Lien, R. E. Ryan, O. Korobkin, S.-K. Lee, J. M. Burgess, W. H. Lee, A. M. Watson, C. Choi, S. Covino, P. D'Avanzo, C. J. Fontes, J. B. González, H. G. Khandrika, J. Kim, S.-L. Kim, C.-U. Lee, H. M. Lee, A. Kutyrev, G. Lim, R. Sánchez-Ramírez, S. Veilleux, M. H. Wieringa and Y. Yoon, The X-ray counterpart to the gravitational-wave event GW170817, *Nature* **551**, 71 (November 2017).
33. E. Pian, P. D'Avanzo, S. Benetti, M. Branchesi, E. Brocato, S. Campana, E. Cappellaro, S. Covino, V. D'Elia, J. P. U. Fynbo, F. Getman, G. Ghirlanda, G. Ghisellini, A. Grado, G. Greco, J. Hjorth, C. Kouveliotou, A. Levan, L. Limatola, D. Malesani, P. A. Mazzali, A. Melandri, P. Møller, L. Nicastro, E. Palazzi, S. Piranomonte, A. Rossi, O. S. Salafia, J. Selsing, G. Stratta, M. Tanaka, N. R. Tanvir, L. Tomasella,

D. Watson, S. Yang, L. Amati, L. A. Antonelli, S. Ascenzi, M. G. Bernardini, M. Boër, F. Bufano, A. Bulgarelli, M. Capaccioli, P. Casella, A. J. Castro-Tirado, E. Chassande-Mottin, R. Ciolfi, C. M. Copperwheat, M. Dadina, G. De Cesare, A. di Paola, Y. Z. Fan, B. Gendre, G. Giuffrida, A. Giunta, L. K. Hunt, G. L. Israel, Z.-P. Jin, M. M. Kasliwal, S. Klose, M. Lisi, F. Longo, E. Maiorano, M. Mapelli, N. Masetti, L. Nava, B. Patricelli, D. Perley, A. Pescalli, T. Piran, A. Possenti, L. Pulone, M. Razzano, R. Salvaterra, P. Schipani, M. Spera, A. Stamerra, L. Stella, G. Tagliaferri, V. Testa, E. Troja, M. Turatto, S. D. Vergani and D. Vergani, Spectroscopic identification of r-process nucleosynthesis in a double neutron-star merger, *Nature* **551**, 67 (November 2017).

34. P. S. Cowperthwaite, E. Berger, V. A. Villar, B. D. Metzger, M. Nicholl, R. Chornock, P. K. Blanchard, W. Fong, R. Margutti, M. Soares-Santos, K. D. Alexander, S. Allam, J. Annis, D. Brout, D. A. Brown, R. E. Butler, H.-Y. Chen, H. T. Diehl, Z. Doctor, M. R. Drout, T. Eftekhari, B. Farr, D. A. Finley, R. J. Foley, J. A. Frieman, C. L. Fryer, J. García-Bellido, M. S. S. Gill, J. Guillochon, K. Herner, D. E. Holz, D. Kasen, R. Kessler, J. Marriner, T. Matheson, E. H. Neilsen, Jr., E. Quataert, A. Palmese, A. Rest, M. Sako, D. M. Scolnic, N. Smith, D. L. Tucker, P. K. G. Williams, E. Balbinot, J. L. Carlin, E. R. Cook, F. Durret, T. S. Li, P. A. A. Lopes, A. C. C. Lourenço, J. L. Marshall, G. E. Medina, J. Muir, R. R. Muñoz, M. Sauseda, D. J. Schlegel, L. F. Secco, A. K. Vivas, W. Wester, A. Zenteno, Y. Zhang, T. M. C. Abbott, M. Banerji, K. Bechtol, A. Benoit-Lévy, E. Bertin, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, D. L. DePoy, S. Desai, J. P. Dietrich, A. Drlica-Wagner, T. F. Eifler, A. E. Evrard, E. Fernandez, B. Flaugher, P. Fosalba, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. A. Goldstein, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, B. Jain, D. J. James, T. Jeltema, M. W. G. Johnson, M. D. Johnson, S. Kent, E. Krause, R. Kron, K. Kuehn, N. Nuropatkin, O. Lahav, M. Lima, H. Lin, M. A. G. Maia, M. March, P. Martini, R. G. McMahon, F. Menanteau, C. J. Miller, R. Miquel, J. J. Mohr, E. Neilsen, R. C. Nichol, R. L. C. Ogando, A. A. Plazas, N. Roe, A. K. Romer, A. Roodman, E. S. Rykoff, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, I. Sevilla-Noarbe, M. Smith, R. C. Smith, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, R. C. Thomas, M. A. Troxel, V. Vikram, A. R. Walker, R. H. Wechsler, J. Weller, B. Yanny and J. Zuntz, The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. II. UV, Optical, and Near-infrared Light Curves and Comparison to Kilonova Models, *Astroph. J.* **848**, p. L17 (October 2017).

35. S. J. Smartt, T.-W. Chen, A. Jerkstrand, M. Coughlin, E. Kankare, S. A. Sim, M. Fraser, C. Inserra, K. Maguire, K. C. Chambers, M. E. Huber, T. Krühler, G. Leloudas, M. Magee, L. J. Shingles, K. W. Smith, D. R. Young, J. Tonry, R. Kotak, A. Gal-Yam, J. D. Lyman, D. S. Homan, C. Agliozzo, J. P. Anderson, C. R. Angus, C. Ashall, C. Barbarino, F. E. Bauer, M. Berton, M. T. Botticella, M. Bulla, J. Bulger, G. Cannizzaro, Z. Cano, R. Cartier, A. Cikota, P. Clark, A. De Cia, M. Della Valle, L. Denneau, M. Dennefeld, L. Dessart, G. Dimitriadis, N. Elias-Rosa, R. E. Firth, H. Flewelling, A. Flörs, A. Franckowiak, C. Frohmaier, L. Galbany, S. González-Gaitán, J. Greiner, M. Gromadzki, A. N. Guelbenzu, C. P. Gutiérrez, A. Hamanowicz, L. Hanlon, J. Harmanen, K. E. Heintz, A. Heinze, M.-S. Hernandez, S. T. Hodgkin, I. M. Hook, L. Izzo, P. A. James, P. G. Jonker, W. E. Kerzendorf, S. Klose, Z. Kostrzewa-Rutkowska, M. Kowalski, M. Kromer, H. Kuncarayakti, A. Lawrence, T. B. Lowe, E. A. Magnier, I. Manulis, A. Martin-Carrillo, S. Mattila, O. McBrien, A. Müller, J. Nordin, D. O'Neill, F. Onori, J. T. Palmerio, A. Pastorello,

F. Patat, G. Pignata, P. Podsiadlowski, M. L. Pumo, S. J. Prentice, A. Rau, A. Razza, A. Rest, T. Reynolds, R. Roy, A. J. Ruiter, K. A. Rybicki, L. Salmon, P. Schady, A. S. B. Schultz, T. Schweyer, I. R. Seitenzahl, M. Smith, J. Sollerman, B. Stalder, C. W. Stubbs, M. Sullivan, H. Szegedi, F. Taddia, S. Taubenberger, G. Terreran, B. van Soelen, J. Vos, R. J. Wainscoat, N. A. Walton, C. Waters, H. Weiland, M. Willman, P. Wiseman, D. E. Wright, Ł. Wyrzykowski and O. Yaron, A kilonova as the electromagnetic counterpart to a gravitational-wave source, *Nature* **551**, 75 (November 2017).

36. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *Astroph. J.* **497**, p. L17 (April 1998).

Neutrinos and gamma-ray production from proton-proton interactions in binary-driven hypernovae framework

S. Campion

Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy
E-mail: stefano.campion@roma1.infn.it
www.phys.uniroma1.it/fisica

ICRA, Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy

J. D. Melon Fuksman

ICRA, Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy
E-mail: david.melon@icranet.org

J. A. Rueda Hernandez

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122 , Italy,
E-mail: jorge.rueda@icra.it

ICRA, Dipartimento di Fisica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, Roma, 00185, Italy

ICRANet-Ferrara, Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, Ferrara, 44122, Italy

Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, Ferrara, 44122, Italy

INAF, Istituto de Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, Rome, 00133, Italy

In this paper we make a prediction on the intensity and energies of neutrinos and photons coming out from the decay chain of the π-meson and μ-lepton, produced by proton-proton inelastic interaction in energetic $\left(E_{\rm iso} \gtrsim 10^{52}\ {\rm erg}\right)$ long gamma-ray bursts (Long-GRBs), within the type I binary-driven hypernova (BdHN) framework.

Keywords: Long-GRBs BdHN; *pp* interaction; GeV-TeV ν/γ

1. Introduction

In the last years multi-messenger astronomy became a fundamental technique to get complementary informations from a single cosmic source. The aim of this technique is to connect different messengers, from the same source, emerging from the same process or from a different one an which are connected spatially and temporally. The advent of new facilities able to detect high-energy astrophysical sources, as gamma-ray bursts (GRBs) or active galactic nuclei (AGN), have been/is important for the development of this field of research.

In this paper we study the neutrino and photons emission emerging from the proton–proton inelastic interaction, for the specific model of binary driven hypernova (BdHN) of long GRB. The BdHN I progenitor is a binary system composed of a carbon-oxygen star (CO_{core}) and a neutron star (NS) companion, where the CO_{core} explosion as supernova (SN) triggers an accretion process onto the NS that, for orbital periods of a few minutes, leads to the formation a black hole (BH). Close to the *cavity* formed around the BH—due to the accretion and the collapse processes—an electron–positron plasma (created in the BH formation), during its isotropic and self-accelerating expansion, it engulfs different amounts of the ejecta baryons along different directions, leading to a direction-dependent proton Lorentz factor. The protons engulfed in the ejecta along the high-density direction ($\sim 10^{23}$ particle/cm^3) reach energies in the range $1.24 \lesssim E_p \lesssim 6.14$ GeV and interact with the ejecta unshocked protons. The protons engulfed along the low density direction, instead, reach energies ~ 1 TeV and interact with the low-density (~ 1 particle/cm^3) protons of the interstellar medium (ISM). The *pp* interactions, for both the density regions, give rise to γ/ν with energies $E_{\gamma,\nu} \leq 5$ GeV and $10 \leq E_{\gamma,\nu} \leq 10^3$ GeV, respectively, for which we calculate the spectra and luminosity.

1.1. *BdHN I: from MeV to GeV and TeV neutrinos*

There are two principal classes of GRBs (see Ref. 1 for the subdivision of long and short GRBs in seven different subclasses). The "short bursts" originate in the mergers of binaries composed of a neutron star (NS) accompanied either by another NS, or a white dwarf (WD) or a black hole (BH), or mergers of binary WDs. The "long bursts" split in four different subclasses, all of them originating from binaries which have been called *binary-driven hypernovae* (BdHNe) of type I, II, III and IV.[2] In this paper, we focus only on the BdHN of type I.

The progenitors of a BdHN system could be zero-age main-sequence (ZAMS) stars ($\gtrsim 10 - 12\ M_\odot$).[3,4] The core–collapse of one of the primary and the consequent formation of a NS at it's center, leads to a system composed by this new born NS and the secondary original star. The system then evolves through mass-transfer episodes and possibly multiple common-envelope epochs.[5] These binary interactions lead to the ejection of the outermost and lighter (H/He) outermost layers of the secondary star. At this stage, the binary is composed of a CO_{core} (or a helium star) and a NS companion.[3,6,7] The different types of BdHN are distinguishable by their final state. As regards the BdHN I, the evolution of the system proceed through the

following steps. The CO_{core} explodes as SN and forms a new NS (hereafter νNS) at its center. At the same time, the ejected material triggers a hypercritical accretion process onto the NS companion. For compact binaries (orbital periods $P \sim 5$ min), the accretion onto the NS brings to the formation of a BH by gravitational collapse (see Refs. 4, 7). This final system is called BdHN I[a], and it explain energetic long bursts with isotropic energy in gamma-rays $E_{\rm iso} \gtrsim 10^{52}$ erg and peak energy $E_{p,i} \gtrsim 200$ keV. The observed occurrence rate of BdHN I ~ 1 Gpc^{-3} y^{-1}.[9]

1.1.1. *Baryonic content available for proton-proton interactions in a BdHN I*

Let's start to analyze now the structure of the baryonic matter present in the BdHN I. The three-dimensional simulations of this system,[2,7,10] pointed out that the SN ejecta, although starts expanding in a spherically symmetric way, becomes highly asymmetric by the accretion process onto the NS[7,10] and the BH formation.[11] Due to this morphology and distribution of baryonic matter around the binary component, the $e^{\pm}$–plasma created in the process of BH formation (see Refs. 12–15 for details on the creation and dynamics of the plasma), which expands isotropically from the newborn BH site, experiences a different dynamics along different directions.[16]

These studies have been specialized in the case where the $e^{\pm}$–plasma engulfs a limited amount of baryons, characterized by a *baryon load* parameter $B \lesssim 10^{-2}$. The *baryon load* is defined as $B \equiv M_b c^2/E_{e^+e^-}$, namely the ratio between the baryon rest-mass energy respect to the $e^{\pm}$ energy. Such low values of B allow the plasma to reach transparency with high Lorentz factor $\Gamma \sim 1/B \gtrsim 10^2$, needed to explain the gamma-ray prompt emission of the GRB. We denote with γ the Lorentz factor of a single particle, and with Γ the one of bulk motion.

In the direction pointing from the CO_{core} to the accreting NS, outwards and lying on the orbital plane, the NS and the BH formation cave a region characterized by very poor baryon pollution, a *cavity* (see Refs. 7, 10, 11). In the other directions along the orbital plane, the $e^{\pm}$–plasma penetrates inside the SN ejecta at $\sim 10^8$–10^{10} cm, and evolves engulfing much larger amounts of baryons, finally reaching transparency at 10^{12} cm with $\Gamma \lesssim 4$.

In order to count the baryon mass engulfed by the $e^{\pm}$–plasma, the baryon load parameter is defined as $B \equiv M_b c^2/E_{e^+e^-}$, namely the ratio between the baryon rest-mass energy respect to the e^+e^- energy. For low baryon load values ($10^{-2} - 10^{-3}$), the transparency of the $e^{\pm}$–plasma occurs for high Lorentz factor $\Gamma \sim 1/B$. For high baryon load, as we will show, the transparency is reach with $\Gamma \sim 4$.

We are ready now to set up the properties of the incident and target protons. Therefore, in BdHN I we have, at least, two types of proton–proton interactions:

(I) Interaction of the protons with low $\Gamma \sim 7$ within the self-accelerated e^+e^-p–plasma that penetrates the SN ejecta, with the unshocked protons ahead the plasma expansion front, at rest inside the ejecta.

[a] We refer the reader to Refs. 2, 8 for details on the other BdHN systems, BdHN II–IV.

(II) Interaction of the protons with $\Gamma \sim 10^2$–10^3 engulfed in the self-accelerated e^+e^-p–plasma in the direction of least baryon density around the newborn BH, with the protons at rest of the interstellar medium (ISM), at a distance $\sim 10^{16}$ cm from the system (see e.g. Ref. 17 for details).

In Sec. 2, we consider the interaction in the high density region (interaction (I) above) and describe the processes connected with the interaction along this direction. In Sec. 3, we study the interaction occurring along the low density region (interaction (II) above). Since the energy range of the interacting and produced particles are different in the two regions, also the models and the approaches used to study the interaction and get the final results differs. Finally, in Sec. 4 we summarize the main results of this work.

2. *pp* interactions inside the high-density ejecta

In this section, we analyze the *pp* interaction occurring when the $e^+e^-\gamma$–plasma starts to expands along the high-density direction and to engulf the protons present in the SN ejecta, forming a $e^+e^-\gamma p$–plasma. Consequently these protons are accelerated and interact inelastically with the target protons ahead of them, still at rest. If the the kinematic energy threshold for the production of the resonance Δ^{++} is reached, the products of the scattering are: photons (from the decay of the π^0), two ν_μ and one ν_e—considering also their antiparticle—(from the decay of $\pi^\pm$ and, consequently, of the secondary created $\mu^\pm$).

In order to study this evolution of the system, we have performed relativistic hydrodynamic (RHD) simulations—performed with a one-dimensional implementation of the RHD module of the `PLUTO` code[18]—of the $e^\pm$–plasma dynamics inside the SN ejecta (see Ref. 16,19 for additional details), described by a system of partial differential equations in only radius and time coordinates. The simulation starts at the moment of BH formation, with initial conditions taken from the final configuration of the numerical simulations in Ref. 7:

(i) The SN remnant is obtained from the explosion of the $\mathrm{CO_{core}}$ of mass $M_{\mathrm{CO}} = 11.15\ M_\odot$ evolved from a ZAMS star of $30\,M_\odot$. M_{CO} is composed by $2\ M_\odot$ for the mass of the νNS (collapsed iron core) and $9.15\,M_\odot$ conform the total ejecta mass (envelope mass).
(ii) The orbital period is $P \approx 5$ min, i.e. a binary separation $a \approx 1.5 \times 10^{10}$ cm.
(iii) The ejecta have negligible pressure, is considered to be in homologous expansion ($v(r) \propto r$) and at rest as seen from the expanding plasma.
(iv) The baryon load of the $e^\pm$–plasma is not isotropic since the baryon density is different along different directions. The density profile of the ejecta, at the BH formation time, decays with distance as a power-law, i.e. $\rho \propto (R_0 - r)^\alpha$ (see Ref. 7).
(v) The total isotropic energy of the $e^\pm$–plasma is set to $E_{e^+e^-} = 3.16 \times 10^{53}$ erg. This implies, according to the above ejecta properties, a baryon load parameter of $B = 51.75$.

The evolution from these initial conditions leads to the formation of a shock and to its subsequent expansion until reaching the outermost part of the SN.

We consider a protons energy distribution in the laboratory frame as a monochromatic function $J_p(E_p) \propto \delta(E_p - E_p^0)$ since, in the comoving frame, it can be represented by Maxwell–Boltzmann distribution. As one can guess from the energy conservation, the protons energy decrease very fast. Then we concentrate on the first stages of the expansion $\left(10^8 \lesssim r \lesssim 10^{10}\ \mathrm{cm}\right)$.

From the RHD simulations described above, we have extracted the quantities necessary for our calculations of the γ/ν spectra, namely the protons Lorentz factor at the shock front, γ_p, their energy, their density, n_{sh}, and the density of the unshocked protons, n_t.

In order to get these quantities, we have built a process that use the pressure inside the plasma, the protons Lorentz factor at the front and the protons mean–free–path λ_p to recognize the plasma front position and the possible interaction protons (above and below the front position).

From this procedure we got that the protons energy varies between the range $1.24 \leq E_p \leq 6.14$ GeV, which is high enough to produce secondary particles. The proton energy threshold to produce pions in the final state is, for the interaction $pp \to pn\pi^+$, $E_{p,\mathrm{Th}} = 1228$ MeV and, for $pp \to pp\pi^0$, $E_{p,\mathrm{Th}} = 1217$ MeV.

In the region of the expansion of the shock, at every radius, the average number density of the target protons $\langle n_t \rangle$ in the remnant varies in the range $5 \times 10^{23}\ \mathrm{cm}^{-3} \lesssim n_t \lesssim 8 \times 10^{23}\ \mathrm{cm}^{-3}$ (the higher value occurs in the beginning of the simulation, namely at the beginning of the expansion). The protons number density at the front of the expanding shell, $\langle n_{\mathrm{sh}} \rangle$, does not vary much either; it is in the range $(0.5\text{–}9) \times 10^{25}\ \mathrm{cm}^{-3}$ (the maximum value occurs in the region close to the initial radius of the expansion).

2.1. *Particles spectra*

We turn now to the describe how we have calculated the spectrum of the particles emerging form the π/μ decay[b]. For our low energy protons, the differential pp cross-section, $d\sigma(E_\pi, E_p)/dE_\pi$, is given in Ref. 20. We notice that the cross-section is different for the three types of pions $\pi^+,\ \pi^-,\ \pi^0$. The pions production rate can be computed as

$$Q_\pi(E_\pi) = cn_t \int_{E_\pi}^{E_p^{\max}} J_p(E_p) \frac{d\sigma(E_\pi, E_p)}{dE_\pi} dE_p, \tag{1}$$

with $J_p(E_p)$ the monochromatic protons energy distribution, n_t the number density of the target protons.

[b]See Ref. 19 for more details on the calculations and for the plot of the particles spectra.

We denote with $\Phi_a = dN_a/dE_a$ the spectrum of the particle a, with N_a the particle number density per unit time[c].

The spectrum of photons emerging from π^0 decay is given by

$$\Phi_{\pi^0\to\gamma\gamma}\left(E_\gamma\right) = 2\int_{E^{\min}(E_\gamma)}^{E_\pi^{\max}} \frac{Q_\pi(E_\pi)}{\sqrt{E_\pi^2 - m_\pi^2 c^4}} dE_\pi, \tag{2}$$

where $E^{\min}\left(E_\gamma\right) = E_\gamma + m_{\pi^0}^2 c^4/(4E_\gamma)$ can be derived by the kinematics of the process and the factor 2 takes into account the two produced photons.

The spectrum of $\nu_{\mu^{(1)}}$ is similarly given by

$$\Phi_{\pi\to\mu\nu_\mu}\left(E_{\nu_\mu}\right) = \frac{1}{\lambda}\int_{E^{\min}\left(E_{\nu_\mu}\right)}^{E_\pi^{\max}} \frac{Q_\pi(E_\pi)\theta\left(\lambda - \frac{E_{\nu_\mu}}{E_\pi}\right)}{\sqrt{E_\pi^2 - m_\pi^2 c^4}} dE_\pi, \tag{3}$$

where $E^{\min}\left(E_{\nu_\mu}\right) = E_{\nu_\mu}/\lambda + (\lambda/4)\left(m_\pi^2 c^4/E_{\nu_\mu}\right)$; $\lambda = 1 - r_\pi$, with $r_\pi = (m_\mu/m_\pi)^2$, is the maximum energy fraction that the neutrino emerging from the direct decay can take from the pion. The value $E_\pi^{\max}$ is calculated from the kinematics and results to be:

$$E_\pi^{\max}\left(E_p\right) = \gamma_{\rm CM}\left(E_p\right)\left[E_\pi^{\rm CM}\left(E_p\right) + \beta_{\rm CM}\left(E_p\right) p_f^{\rm CM}\left(E_p\right)\right], \tag{4}$$

where CM indicate that the Lorentz factor, the momentum and the velocity are calculated in centre-of-mass frame.

Instead, the spectrum of the $\nu_{\mu^{(2)}}$ are given by:

$$\Phi_{\pi\to\mu\to\nu_\mu/\nu_e}\left(E_\nu\right) = \int_{E^{\min}(E_\nu)}^{E_\pi^{\max}} \frac{Q_\pi(E_\pi)}{\sqrt{E_\pi^2 - m_\pi^2 c^4}}\, g\left(\frac{E_\nu}{E_\pi}\right) dE_\pi, \tag{5}$$

where the functions $g(z)$ are derived in Ref. 21 and represent the ν spectra after the decay chain $\pi \to \mu \to \nu$.

The function $g(z)$ can be decomposed as the sum of an unpolarized spectrum, $g^0(z)$, plus a polarized one, $g^{\rm pol}(z)$ (see Appendix in Ref. 21). We have used the formula for the relativistic limits (given in that paper), $\beta_\pi \to 1$, which is well satisfied since the pions Lorentz factor assumes values $4.5 \leq \gamma_\pi \leq 34.5$. We have calculated the spectrum for both unpolarized and polarized case (see Ref. 19).

Altogether, the spectra of the different particles varies between values $10^{23} \leq E^2 dN_a/dE \leq 10^{31}$ erg/cm^3/s at energies $0.3 \leq E_a \leq 5$ GeV.

At this point, we can calculate the particle emissivity at every radius, ϵ_a^i, integrating over the energy of the particle a (from 0.3 to $E_\pi^{\max}$) its spectrum times E_a and, then, integrate ϵ_a^i over the volume V_i to get the total luminosity L_a. The interacting volume at the radius i is calculated as $V_i = 4\,\pi\, r_i^2\, \lambda_i$, where λ_i is the mean-free path of the protons of energy E_p^i in the shell front. Finally, the total

[c]Henceforth we denote as $\nu_{\mu^{(1)}}$ the muonic neutrino/antineutrino from the direct pion decay, $\pi \to \mu\nu_\mu$, and $\nu_{\mu^{(2)}}$ the neutrino/antineutrino from the consequent muon decay, $\mu \to e\nu_\mu\nu_e$.

energy emitted for each particle a, is given integrating L_a over the entire emission time, namely (Δt_i, is the time the shell spends to cover the distance between r_{i-1} and r_i with the velocity β_i, $\Delta t_i \approx \Delta r_i/(c\,\beta_i)$):

$$\mathcal{E}_a \approx \sum_{i=2}^{n} L_a^i \times \Delta t_i = \sum_{i=2}^{n} L_a^i \times \frac{\Delta r_i}{c\beta_i}, \tag{6}$$

The total energy emitted in photons is $\sim 4 \times 10^{52}$ erg, while the different types of ν it varies between 0.5 and 3.5×10^{51} erg. Summing all the energy emitted for all the particle, we obtain a total energy release of 9.11×10^{51} erg, that is $\approx 2.9\%$ of the energy of initial energy of the $\gamma e^{\pm}$–plasma.

3. TeV protons interacting with the ISM

In this section, we now consider the interaction of incident protons engulfed by the $\gamma e^{\pm}$–plasma in the direction of the circumburst medium (CBM) with low baryon load—we assume here $B = 10^{-3}$—with target protons of the ISM. The interaction with the ISM—located at a distance between $10^{16} \leq r \leq 10^{17}$ cm from the BH site—occurs in a spherical shell. Thus in this case, the number density of the target is $n_{\rm ISM} \sim 1$ cm^{-3} and the transparency is reached far away from the BH site, with ultra-relativistic Lorentz factor of up to $\gamma_p = \Gamma \sim 10^3$. Therefore, we adopt here that the incident protons have energies ~ 1 TeV.

The number of engulfed protons, N_p, is given by the ration between the isotropic energy of the $e^{\pm}$ times the baryon load over the protons mass. For $E_{e^+e^-} \sim 10^{53}$ erg and $B = 10^{-3}$, $N_p = 6.65 \times 10^{52}$.

Since the energy range differs from the previous case, for the computation of the particles spectrum here we have used the results obtained in Ref. 22, where a parameterization of the pp cross-section for high-energy protons and an analytic formula for the spectrum of the emerging particles is derived.

Following Ref. 22 and denoting $J_p(E_p)$ the protons energy distribution—in units cm^{-3} TeV^{-1}—in this case the production rate for the secondary particle a, $\Phi_a(E_a) \equiv dN_a/dE_a$ (for $E_p \geq 0.1$ TeV) is given by

$$\Phi_a(E_a) = c\,n_p \int_{E_a}^{\infty} \sigma_{\rm inel}^{pp}(E_p) J_p(E_p) F_a\,(x, E_p)\,\frac{dE_p}{E_p}, \tag{7}$$

with n_p the target protons density (we assume it is 1 particles/cm^3), $\sigma_{\rm inel}^{pp}(E_p)$ is the inelastic pp cross-section, $x \equiv E_a/E_p$ and F_a is the specific spectrum for the particle a derived in Ref. 22.[d]

[d]We notice that in Ref. 22 they do not distinguish between π^+ and π^-, as well as between ν and $\bar{\nu}$. So, differently from the low density/energy case, here we do not distinguish between ν and $\bar{\nu}$ too.

3.0.1. *Particles spectra*

Here the protons energy distribution is $J_p\left(E_p\right) = A\delta\left(E_p - E_p^0\right)$, with $E_p^0 = 1$ TeV and $A = N_p/V$, with N_p the number of interacting protons over the volume $V = 4\pi\left(r_2^3 - r_1^3\right)/3$, with $r_1 = 10^{16}$ cm and $r_2 = 10^{17}$ cm.

The total emitted energy for each particle a here is calculated as

$$\mathcal{E}_{tot,\ a} = \Delta t \sum_i \epsilon_a^i \times \Delta V_i = \ \epsilon_a\ \Delta t\ \Delta V, \tag{8}$$

where $\Delta V_i = 4/3\,\pi\left(r_i^3 - r_{i-1}^3\right)$, with $r_i = r_{i-1} + \delta s$—δs is the space covered by the protons with their velocities in one second.[e]

The luminosity is given by $L_a = \epsilon_a \times \Delta V_{\text{last shell}}$, where: ϵ_a the emissivity of the particle a calculated by Eq. (7), $\Delta V_{\text{last shell}} = 4/3\,\pi\left(r_2^3 - r_*^3\right)$ and $r_* = 10^{17} - \delta s$ cm and $r_2 = 10^{17}$ cm. $r_2 - r_*$ represents the last ISM shell where the accelerated protons interact—the protons spends $\Delta t = \Delta r/c\,\beta = \ 3 \times 10^6$ seconds to cross the entire ISM region and, then, it is useless to talk about total luminosity.

Then, the emissivity of all the secondary emerging particles from the *pp* interaction are given by substituting the $F_a\left(x, E_p\right)$ functions derived in Ref. 22 inside Eq. (7). We got the following results for: the total energy $\mathcal{E}_a$ emitted through all emitting region (calculated via Eq. (8)), the luminosity of the last emitting shell L_a and the particle energy for which we have the maximum value of the spectrum $E^2 dN_\gamma/dE$, $E_a^{\max}$:

(1) for photons: $\mathcal{E}_\gamma = 5.41 \times 10^{43}$ erg, $L_\gamma = 1.01 \times 10^{43}$ erg s^{-1} and $E_\gamma^{\max} = 91.62$ GeV;
(2) for muonic neutrino from direct pion decay $\left(\nu_{\mu^{(1)}}\right)$: $\mathcal{E}_{\nu_{\mu^{(1)}}} = 1.60 \times 10^{43}$ erg, $L_{\nu_{\mu^{(1)}}} = 3.01 \times 10^{42}$ erg s^{-1} and $E_{\nu_{\mu^{(1)}}}^{\max} = 44.72$ GeV;
(3) for muonic neutrino from muon decay $\left(\nu_{\mu^{(2)}}\right)$: $\mathcal{E}_{\nu_{\mu^{(2)}}} = 1.98 \times 10^{43}$ erg, $L_{\nu_{\mu^{(2)}}} = 3.71 \times 10^{42}$ erg s^{-1} and $E_{\nu_{\mu^{(2)}}}^{\max} = 63.9$ GeV.

From these results we deduce that the principal differences between the interactions considered in Sec. 2 and Sec. 3 resides in the density of the target particles in both interaction regions. Another, but less efficient, factor that has influence on the obtained results is the difference in the interacting protons energy, which is reflected onto the *pp* cross-section.

4. Summary, discussion and conclusions

In this work we have computed the photon and neutrino production via proton-proton interactions occurring within the BdHNe I scenario for energetic long GRBs, which we have recalled in Sec. 1. From the dynamics of the BdHN I formation process leads to a highly asymmetric SN ejecta around the newborn BH site. Therefore,

[e]Note that ϵ_a^i does not depend on the radius, $\epsilon_a^i = \epsilon_a$, and then $\Delta V = 4/3\,\pi\left(r_2^3 - r_1^3\right)$.

the $e^{\pm}$–plasma created in the BH formation process, during its isotropic expansion and self-acceleration, engulfs different amounts of matter of the surrounding SN ejecta, depending on the direction. Consequently, either the observation of the emitted particles are direction dependent.

In Sec. 2 we have studied the *pp* interaction along the high-density direction inside the SN ejecta. The secondary particles production occurs in the first ~ 1.5 s of the plasma expansion inside the ejecta. The $e^{\pm}$–plasma starts to expand inside the SN ejecta, engulfing protons and accelerating them to energies of a few GeV. For these low energetic protons, we adopted the cross-section given in Ref. 20. At later times, the energy of the protons is below the threshold for the production of secondary particles by *pp* inelastic scattering. The obtained results tell us that the neutrinos and photons have energies $E_{\nu_{\mu^{(1)}}} \leq 2$ GeV and E_{γ}, $E_{\nu_{\mu^{(2)}}}$, $E_{\nu_e} < 5$ GeV, with associated total energy (integrated over the whole emitting region) of $\sim 10^{50}$–10^{51} erg.

In the second part, Sec. 3, we have studied the expansion of the plasma along the direction of low baryon load $\left(B = 10^{-3}\right)$ (see, e.g., Refs. 1,23). Here, the expansion of the $\gamma e^{\pm}$–plasma along the low-density directions allows a self-acceleration of it that brings the engulfed protons to energies of up to $E_p \sim 1$ TeV ($\gamma_p \sim 10^3$). For these higher energetic protons we have followed the approach presented in Ref. 22, where the *pp* cross-section and the spectra of the emerging particles are parametrized. In this case, the secondary particles have energies $1 \leq E_a \leq 10^3$ GeV and the total emitted energy is $\sim 10^{43}$ erg.

We found also that the spectrum of the emerging particles follow approximately a cutoff power-law function (CPL) $f(E) \propto E^{\alpha} \exp E/E_c$, with spectral index $1 \leq \alpha < 3$, depending on the considered particle and the interaction region. The power-law term usually derives from the spectral index of the primary interacting protons (see, for example, Ref. 24), but since we considered a spherical expansion of the $\gamma e^{\pm} B$–plasma, and a monochromatic protons energy distribution (at each radius of the expansion), we deduce that the power-law term is intrinsic in the considered process. Instead, the exponential decay is explained by the kinematic of the process. Indeed a fraction of the parent proton energy is taken by the pion; for the direct pion decay, the $\nu_{\mu^{(1)}}$ can take, at maximum, a fraction λ of the pion energy, while the muon can take, at maximum, the entire pion energy.

Now we try to give some numbers concerning the ν detection with Earth Detectors. In general, since GRB occur at cosmological distances and this makes neutrino detection very challenging because of the very low neutrino flux arriving to the Earth.

We can obtain order-of-magnitude estimate of the probability of detection of these neutrinos considering three detectors: SuperKamiokande (SK), Hyper-Kamiokande (HK) and IceCube (IC), calculating the detection horizon.[f] The

[f]For this purpose, we use the best experimental conditions, i.e. the peak neutrino luminosity and the corresponding neutrino energy.

number of neutrinos per-unit-time and per-unit-area that arrive to the detector is

$$\frac{d^2N_\nu}{dSdt} = \frac{\mathcal{E}_\nu}{4\pi D^2 E_\nu^*}, \tag{9}$$

where $\mathcal{E}_\nu$ is the total energy emitted during 1 s in neutrinos of energy E_ν, D is the luminosity distance to the source, $E_\nu^* = E_\nu/(1+z)$ is the redshifted neutrino energy, and z the source cosmological redshift.

The number of detectable neutrino, $N_\nu^{\rm det}$, is calculated as the number of neutrino per unit of time and area that arrives to the detector, given by Eq. (9), times the cross-section for the neutrino-nucleon interaction, $\sigma_{\nu N}$ (see, e.g. Ref. 25, 26), times the number of probable interacting baryon in the detector $N_b^{\rm det} = N_A \times M_{\rm int}$—where N_A is the Avogadro number and $M_{\rm int}$ the interacting mass inside the detector—times the integration time of the detector $T_{\rm int}$, that we set to 1 second (in agreement with the time-interval of the emission inside the ejecta):

$$N_\nu^{\rm det} = \frac{d^2N_\nu}{dSdt} \times T_{\rm int} \times \sigma_{\nu N} \times N_b^{\rm det}, \tag{10}$$

For the three considered detectors is $N_b^{\rm det} = N_A \times (22.5\,{\rm kton}) = 1.35 \times 10^{34}$ baryons (for SK); $N_b^{\rm det} = (560\ {\rm kton}) \times N_A = 3.37 \times 10^{35}$ baryons (for HK); for the *Deep Core Detector* of IC, the effective volume is a function of the neutrino energy: $N_b^{\rm det} = (10\ {\rm Mton}) \times N_A = 6.022. \times 10^{36}$ baryons; $N_b^{\rm det} = (20\ {\rm Mton}) \times N_A = 1.2044 \times 10^{37}$ baryons; $N_b^{\rm det} = (30\ {\rm Mton}) \times N_A = 1.8066 \times 10^{37}$ baryons.

Now, using the relations and estimates above and considering the Hubble-Lemaître law, $c\,z = H_0 D_h$ (with $H_0 = 72$ km s^{-1} Mpc^{-1}), we can obtain the neutrino-detection horizon, D_h, i.e. the luminosity distance to the source for which we have $N_\nu^{\rm det} = 1$:

$$D_h = \frac{K H_0}{2c} + \frac{1}{2}\sqrt{\frac{K^2 H_0^2}{c^2} + 4K}, \tag{11}$$

where $K = \mathcal{E}_\nu \sigma_{\nu N} T_{\rm int} N_b^{\rm det}/(12\pi E_\nu)$.

For the low energy neutrinos ($E_\nu \leq 2$ GeV, at the production site, i.e. in the source frame) there are additional considerations that decrease the probability to detect these neutrino, namely: 1) the fact that the energy need to be divided for the factor $1+z$ and that for low energetic ν the cross-section for $\nu - N$ interaction is very low—$\sim 10^{-39}$ cm^2 (see, for example, Ref. 25, 26); 2) for these low energies, there is much background noise by atmospheric neutrinos, created in the atmospheric showers by cosmic rays and solar neutrinos.

The high energetic neutrinos ($E_\nu \leq 10^3$ GeV) from the low-density region could be, in principle, more easily detected than the previous case since at these energies there is no background noise from atmospheric and solar neutrinos and the cross-section assumes higher values $\sigma_{\nu N} \sim 10^{-37}$ cm^2.

The calculations of the horizon distance through Eq. (11) give the following results that

(I) for the ν from the high density region, $0.2 \leq D_h \leq 11.12$ Mpc;
(II) for the ν from the low density region, $22 \leq D_h \leq 954$ pc.

In both cases, we obtained the higher values with the Deep Core of the IC detector, while the lower ones are for the SK detector. This analysis suggests that only the IceCube detector might detect the neutrinos here analyzed, i.e. neutrinos produced by *pp* interactions in the context of BdHNe I, especially the one coming from the high-density region, even if all of them have the right energy range of sensibility to detect our neutrinos.

References

1. R. Ruffini, J. Rueda, M. Muccino, Y. Aimuratov, L. Becerra, C. Bianco, M. Kovacevic, R. Moradi, F. Oliveira, G. Pisani *et al.*, On the classification of GRBs and their occurrence rates, *The Astrophysical Journal* **832**, p. 136 (2016).
2. Y. Wang, J. A. Rueda, R. Ruffini, L. Becerra, C. Bianco, L. Becerra, L. Li and M. Karlica, Two Predictions of Supernova: GRB 130427A/SN 2013cq and GRB 180728A/SN 2018fip, *ApJ* **874**, p. 39 (March 2019).
3. J. A. Rueda and R. Ruffini, On the Induced Gravitational Collapse of a Neutron Star to a Black Hole by a Type Ib/c Supernova, *ApJ* **758**, p. L7 (October 2012).
4. L. Becerra, F. Cipolletta, C. L. Fryer, J. A. Rueda and R. Ruffini, Angular Momentum Role in the Hypercritical Accretion of Binary-driven Hypernovae, *ApJ* **812**, p. 100 (October 2015).
5. T. M. Tauris, N. Langer and P. Podsiadlowski, Ultra-stripped supernovae: Progenitors and fate, *MNRAS* **451**, 2123 (August 2015).
6. C. L. Fryer, J. A. Rueda and R. Ruffini, Hypercritical Accretion, Induced Gravitational Collapse, and Binary-Driven Hypernovae, *ApJ* **793**, p. L36 (October 2014).
7. L. Becerra, C. L. Bianco, C. L. Fryer, J. A. Rueda and R. Ruffini, On the Induced Gravitational Collapse Scenario of Gamma-ray Bursts Associated with Supernovae, *ApJ* **833**, p. 107 (December 2016).
8. J. A. Rueda, R. Ruffini and Y. Wang, Induced Gravitational Collapse, Binary-Driven Hypernovae, Long Gramma-ray Bursts and Their Connection with Short Gamma-ray Bursts, *Universe* **5**, p. 110 (May 2019).
9. R. Ruffini, J. A. Rueda, M. Muccino, Y. Aimuratov, L. M. Becerra, C. L. Bianco, M. Kovacevic, R. Moradi, F. G. Oliveira, G. B. Pisani and Y. Wang, On the Classification of GRBs and Their Occurrence Rates, *ApJ* **832**, p. 136 (December 2016).
10. L. Becerra, C. L. Ellinger, C. L. Fryer, J. A. Rueda and R. Ruffini, SPH Simulations of the Induced Gravitational Collapse Scenario of Long Gamma-Ray Bursts Associated with Supernovae, *ApJ* **871**, p. 14 (Jan 2019).
11. R. Ruffini, J. D. Melon Fuksman and G. V. Vereshchagin, On the Role of a Cavity in the Hypernova Ejecta of GRB 190114C, *ApJ* **883**, p. 191 (Oct 2019).
12. G. Preparata, R. Ruffini and S.-S. Xue, The dyadosphere of black holes and gamma-ray bursts, *A&A* **338**, L87 (October 1998).
13. R. Ruffini, J. D. Salmonson, J. R. Wilson and S.-S. Xue, On evolution of the pair-electromagnetic pulse of a charged black hole, *Astronomy and Astrophysics Supplement* **138**, 511 (September 1999).

14. R. Ruffini, J. D. Salmonson, J. R. Wilson and S. S. Xue, On the pair-electromagnetic pulse from an electromagnetic black hole surrounded by a baryonic remnant, *A&A* **359**, 855 (July 2000).
15. C. L. Bianco, R. Ruffini and S. S. Xue, The elementary spike produced by a pure e^+e^- pair-electromagnetic pulse from a Black Hole: The PEM Pulse, *A&A* **368**, 377 (March 2001).
16. R. Ruffini, Y. Wang, Y. Aimuratov, U. Barres de Almeida, L. Becerra, C. L. Bianco, Y. C. Chen, M. Karlica, M. Kovacevic, L. Li, J. D. Melon Fuksman, R. Moradi, M. Muccino, A. V. Penacchioni, G. B. Pisani, D. Primorac, J. A. Rueda, S. Shakeri, G. V. Vereshchagin and S.-S. Xue, Early X-Ray Flares in GRBs, *ApJ* **852**, p. 53 (January 2018).
17. L. Izzo, R. Ruffini, A. V. Penacchioni, C. L. Bianco, L. Caito, S. K. Chakrabarti, J. A. Rueda, A. Nandi and B. Patricelli, A double component in GRB 090618: A proto-black hole and a genuinely long gamma-ray burst, *A&A* **543**, p. A10 (July 2012).
18. A. Mignone, C. Zanni, P. Tzeferacos, B. Van Straalen, P. Colella and G. Bodo, The pluto code for adaptive mesh computations in astrophysical fluid dynamics, *The Astrophysical Journal Supplement Series* **198**, p. 7 (2011).
19. S. Campion, J. D. M. Fuksman and J. A. R. Hernandez, Neutrino production from proton-proton interactions in binary-driven hypernovae, *arXiv preprint arXiv:1910.10439* (2019).
20. S. R. Blattnig, S. R. Swaminathan, A. T. Kruger, M. Ngom and J. W. Norbury, Parametrizations of inclusive cross sections for pion production in proton-proton collisions, *Phys. Rev. D* **62**, p. 094030 (Oct 2000).
21. P. Lipari, Lepton spectra in the earth's atmosphere, *Astroparticle Physics* **1**, 195 (1993).
22. S. R. Kelner, F. A. Aharonian and V. V. Bugayov, Energy spectra of gamma rays, electrons, and neutrinos produced at proton-proton interactions in the very high energy regime, *Phys. Rev. D* **74**, p. 034018 (Aug 2006).
23. R. Ruffini, Y. Wang, Y. Aimuratov, U. B. de Almeida, L. Becerra, C. L. Bianco, Y. C. Chen, M. Karlica, M. Kovacevic, L. Li, J. D. M. Fuksman, R. Moradi, M. Muccino, A. V. Penacchioni, G. B. Pisani, D. Primorac, J. A. Rueda, S. Shakeri, G. V. Vereshchagin and S.-S. Xue, Early x-ray flares in GRBs, *The Astrophysical Journal* **852**, p. 53 (2018).
24. T. K. Gaisser, R. Engel and E. Resconi, *Cosmic rays and particle physics* (Cambridge University Press, 2016).
25. P. Zyla *et al.*, Review of Particle Physics, *PTEP* **2020**, p. 083C01 (2020).
26. Y. Hayato, A neutrino interaction simulation program library neut, *Acta Phys. Polon.* **40**, 2477 (2009).

General relativistic turbulence in spherically symmetric core-collapse supernovae simulations

L. Boccioli* and G. J. Mathews

Center for Astrophysics, Department of Physics, University of Notre Dame, 225 Nieuwland Science Hall, Notre Dame, IN 46556, USA
**E-mail: lbocciol@nd.edu*

E. P. O'Connor

The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden

It is generally believed that General Relativity (GR) is of secondary importance in the explosion of core-collapse supernovae (CCSN). However, as 3D simulations are becoming more and more detailed, GR effects can be strong enough to change the hydrodynamics of the supernova and affect the explosion. Since a 3D simulation in full GR is computationally extremely challenging, it is valuable to modify simulations in a spherically symmetric spacetime to incorporate 3D effects. This permits exploration of the parameter dependence of CCSN with a minimum of computational resources. In this proceedings contribution we report on the formulation and implementation of general relativistic neutrino-driven turbulent convection in the spherically symmetric code `GR1D`. This is based upon *STIR*, a recently proposed Newtonian model based on mixing length theory. When the parameters of this model are calibrated to 3D simulations, we find that our GR formulation significantly alters the correspondence between progenitor mass and explosion vs. black-hole formation. We therefore believe that, going forward, simulating CCSNe in full GR is of primary importance.

Keywords: Supernovae - Simulations - Mixing Length Theory - General Relativistic Hydrodynamics

1. Introduction

This proceedings contribution is based on work published in Ref. [1]. Core-collapse supernovae (CCSNe) have been at the core of cutting-edge computational research for more than 50 years. Despite that, the details of the mechanisms driving the explosion still remain unknown, even though significant progress has been made since the first attempts at explaining CCSNe.[2–4]

Historically, one-dimensional (1D) spherically symmetric simulations were able to assess the crucial role of neutrinos in aiding the expansion of the shock through the so-called delayed neutrino-heating mechanism.[5,6] Two-dimensional (2D) simulations[7,8] and three-dimensional (3D) simulations,[9] have been only recently accessible thanks to the fast technological improvements of the last three decades.

Spherically symmetric simulations, however, have unfortunately, not led to self-consistent explosions of Fe-core CCSNe (which are the most common) since they involve all stars with masses > 11 $M_\odot$. On the other hand, there have been several simulations in 2D and 3D that have led to successful explosions.[10–16] However, 2D simulations have been recently shown[17] to favor explosions by an artificial enhancement of neutrino-heating behind the shock via an inverse turbulence cascade which is not present in 3D. Therefore, only 3D simulations can provide the final explanation as to what causes the explosion. However, despite the technological improvements of the last few decades, 3D simulations continue to present a difficult computational challenge, even for modern supercomputers.

In comparison, modern 1D simulations are significantly faster to run and are also more consistent across different codes.[18] In other words, when the initial conditions are the same, different groups obtain similar results. This guarantees a somewhat solid foundation, making 1D simulations an ideal tool to study how different input physics can affect the explosion of supernovae. To do that, first one needs to artificially trigger an explosion. In a recent paper Couch et al.[19] (hereafter CWO20) developed *STIR* (Simulated Turbulence In Reduced-dimensionality), a parametric model based upon Mixing Length Theory (MLT) that incorporates the effects of 3D turbulence in spherically symmetric simulations.

The simulations from CWO20 use Newtonian hydrodynamics and only partially include general relativistic effects through a General Relativistic Effective Potential (GREP) from Ref. [20], which is a common practice in the supernova community. However, we know that General Relativity (GR) plays an important role in the explosion of supernovae.[10,21] Hence, simulations in full GR are desirable, and in this proceedings contribution we summarize the extension of the *STIR* model to a general relativistic treatment.[1] Throughout the manuscript, we adopt natural units, i.e. $G = c = M_\odot = 1$.

2. Methods

2.1. *The STIR model of Couch et al. (2020)*

A detailed description of STIR can be found in Refs. 22, 23 and CWO20. There, it is shown that the effects of turbulence can be treated as a perturbation on the background fluid. After a Reynolds decomposition of the compressible Euler equations, and several other approximations valid in typical supernova thermodynamic environments, one arrives to the following equation describing the evolution of the turbulent kinetic energy:

$$\begin{aligned} \frac{\partial \rho v_{\rm turb}^2}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}[r^2(\rho v_{\rm turb}^2 v_r - \rho D_K \nabla v_{\rm turb}^2)] \\ = -\rho v_{\rm turb}^2 \frac{\partial v_r}{\partial r} + \rho v_{\rm turb}\omega_{\rm BV}^2 \Lambda_{\rm mix} - \rho \frac{v_{\rm turb}^3}{\Lambda_{\rm mix}} \quad , \end{aligned} \tag{1}$$

where

$$\Lambda_{\rm mix} = \alpha_{\rm MLT} \frac{P}{\rho g}, \tag{2}$$

$$\omega_{\rm BV}^2 = g_{\rm eff} \left(\frac{1}{\rho} \frac{\partial \rho}{\partial r} - \frac{1}{\rho c_s^2} \frac{\partial P}{\partial r} \right). \tag{3}$$

In the above equations, ρ is the mass density, v_r is the radial velocity, $\Lambda_{\rm mix}$ is the mixing length, $\omega_{\rm BV}$ is the Brunt-Väisälä frequency, c_s is the sound speed, D_K is a diffusion coefficient due to turbulence and $g_{\rm eff}$ is the magnitude of the local effective acceleration. For a fluid in hydrostatic equilibrium, $g_{\rm eff}$ simply reduces to the local gravitational acceleration g. More generally, however, in the rest frame one should take the acceleration of the fluid into account. Therefore, the total acceleration $g_{\rm eff}$ can be expressed as:

$$g_{\rm eff} = g - v_r \frac{\partial v_r}{\partial r} \ , \tag{4}$$

as described in CWO20.

The mixing length $\Lambda_{\rm mix}$ is the average distance that a convective element will travel before being mixed with (and increasing the internal energy of) the surrounding material. The Brunt-Väisälä frequency $\omega_{\rm BV}$ is the rate at which the convective elements are rising. As one can notice from Eq. (3), $\omega_{\rm BV}^2$ can be either positive or negative: when $\omega_{\rm BV}^2 > 0$ the fluid is convectively unstable, i.e. convection is generated; when $\omega_{\rm BV}^2 < 0$ the fluid is convectively stable, i.e. convection is damped. Ultimately, the main parameter of the model is $\alpha_{\rm MLT}$, which scales the mixing length to the pressure scale height in Eq. (2). Typically $\alpha_{\rm MLT} \sim O(1)$.

The diffusion coefficient D_K is defined as:

$$D_K = \alpha_K v_{\rm turb} \Lambda_{\rm mix} \ . \tag{5}$$

Similar terms appear in the internal energy, electron fraction, and neutrino energy density evolution equations (for the complete set of hydrodynamic equations used in the model, (see Eqs. 25-29 and 33 in CWO20). Therefore, strictly speaking, *STIR* has 4 additional free parameters: α_K, α_e, $\alpha_{\rm Ye}$, α_ν. However, the convective motions are not very sensitive to the value of these parameters, so we set them to 1/6 for simplicity, consistent with the choices of Ref. [24] and CWO20.

In the next section we describe a general relativistic version of the model described above.

2.2. *STIR in General relativity*

The first attempts to create a general relativistic model for convection date back to Ref. [25]. We follow the same approach, but using a slightly different formalism.

All the simulations described here[1] were run with the open-source, spherically symmetric, general relativistic code `GR1D`.[26] The Boltzmann equation for neutrino

transport is solved using an M1-scheme, with opacity tables generated using the open-source code `NuLib`.[27]

The metric evolved in `GR1D` is Schwarzchild-like:

$$\begin{aligned} ds^2 &= g_{\mu\nu}x^\mu x^\nu \\ &= -\alpha(r,t)^2dt^2 + X(r,t)^2dr^2 + r^2d\Omega^2, \end{aligned} \tag{6}$$

where α and X can be expressed as functions of a metric potential ϕ (which reduces to the Newtonian potential in the Newtonian limit) and the enclosed gravitational mass $M_{\rm grav}$:

$$\begin{aligned} \alpha(r,t) &= \exp[\phi(r,t)], \\ X(r,t) &= \left(1 - \frac{2M_{\rm grav}(r,t)}{r}\right)^{-1/2}. \end{aligned} \tag{7}$$

For the present work, we first note that turbulence is mostly relevant far from the proto-neutron star (PNS) where GR effects can be treated as a perturbation. Therefore, one can simply make a few changes to the terms in Eq. (1) without having to re-derive the entire Reynold's decomposition. The expression for $\omega^2_{\rm BV}$, however, must be carefully re-derived. Far from the PNS, we invoke the following:

(1) replace the conserved variable ρ with its GR counterpart, i.e. $D = WX\rho$, where $W = (1-v^2)^{-1/2}$ and $v = Xv_r$;
(2) multiply the RHS of Eq. (1) by αX.
(3) multiply the spatial flux in Eq. (1) by α/X (see Ref. [26] for more details on the derivation of the hydrodynamic equations in `GR1D`).

The expression of $\omega^2_{\rm BV}$ can be derived using conservation of momentum for a convective eddy in a background fluid in hydrostatic equilibrium. The case of a fluid with non zero acceleration can be derived with ad-hoc corrections for general relativistic effects. The derivation can be found in Ref. [1], and it leads to:

$$\omega^2_{\rm BV} = \frac{\alpha^2}{\rho h X^2}\left(\frac{{\rm d}\phi}{{\rm d}r} - v\frac{\partial v}{\partial r}\right)\left(\frac{\partial \rho(1+\epsilon)}{\partial r} - \frac{1}{c_s^2}\frac{\partial P}{\partial r}\right), \tag{8}$$

where $v = Xv_r$.

The main difference between Eqs. (3) and (8) is the inclusion of $\partial\rho\epsilon/\partial r$ in the latter. In the gain region the internal energy decreases with radius, i.e. $\partial\rho\epsilon/\partial r < 0$. This decreases the magnitude of $\omega_{\rm BV}$ and therefore the amount of turbulence that is generated. We will come back to this in Section 3.2

3. Results: Comparison with 3D simulations

3.1. *Results using an effective potential*

Inspired by the work of CWO20, we compare our GREP model to the `mesa20_LR_v` 3D simulation,[28] by using the same setup chosen by CWO20. That is, we simulate

the collapse of a 20 $M_\odot$ progenitor from,[29] adopting the SFHo EOS[30] and assuming Nuclear Statistical Equilibrium (NSE) everywhere. The algorithm used in `FLASH` to solve the neutrino radiation transport closely resembles the one used in `GR1D`.[27] Additionally, the set of `NuLib` opacities we adopted is the same used by CWO20. Finally, we use 12 neutrino energy groups geometrically spaced up to ∼250 MeV.

The upper panels of Figure 1 (modified from Ref. [1]) show the shock radius versus time and the turbulent velocity profile at ∼135 ms post-bounce for our GREP model (to be compared to Figures 1 and 2 of CWO20). The main difference between our results using `GR1D` and the ones from CWO20 using `FLLASH` is that `GR1D` consistently gives larger values for the turbulent velocity at a given $\alpha_{\rm MLT}$. This then translates into larger shock radii at a given time. Except for these small differences, the agreement between the two models is very good, and both yield explosions for $\alpha_{\rm MLT} \gtrsim 1.2$.

When it comes to the comparison with the 3D results, however, our MLT-like model does not captures some features that are present in the 3D case. Specifically, the profile at ∼135 ms post-bounce of the convective speed in 3D has a longer tail at 50-80 km. This has already been noticed by CWO20, and it is due to angular variations present in the 3D model, rather than a deeper convection extending in the region below the gain layer. In our model, convection shuts off at 80 km, as one would expect, since that is approximately the location of the gain layer. A more interesting difference is the lack of PNS convection at 25 km, not captured by our MLT-like model. A possible explanation for this is that STIR is not taking lepton number-driven convection into account, which is not easily tractable with MLT models, and therefore a more careful treatment of this type of convection might ease the discrepancy with the 3D results. We are currently working on adjusting some of the parameters of STIR deep inside the PNS to match the 3D result, and therefore analyze the impact of PNS convection on the explosion, but this goes beyond the scope of this conference proceedings.

3.2. *Results using GR*

One can compare the results obtained using the simple GREP approach with results in full GR. We show the results using full GR in the bottom panels of Figure 1, while the upper panels refer to the runs using our GREP model. The most important difference to point out is the range of $\alpha_{\rm MLT}$ used in the GR and GREP simulations. To produce shock radii and turbulent velocities that are similar to the GREP results, the value of $\alpha_{\rm MLT}$ that needs to be used in full GR is ∼ 20% larger. The reason behind this increase in $\alpha_{\rm MLT}$ lies in the expression of $\omega_{\rm BV}$. As pointed out in Section 2.2, including the internal energy gradient into eq. (8) is the main difference between the GREP and GR models.

In the gain region, where turbulent convection is most relevant, the gradient of the internal energy is negative. This decreases the magnitude of $\omega^2_{\rm BV}$, making the fluid more stable against convection. Including $\rho\epsilon$ in the definition of $\omega_{\rm BV}$ is hence

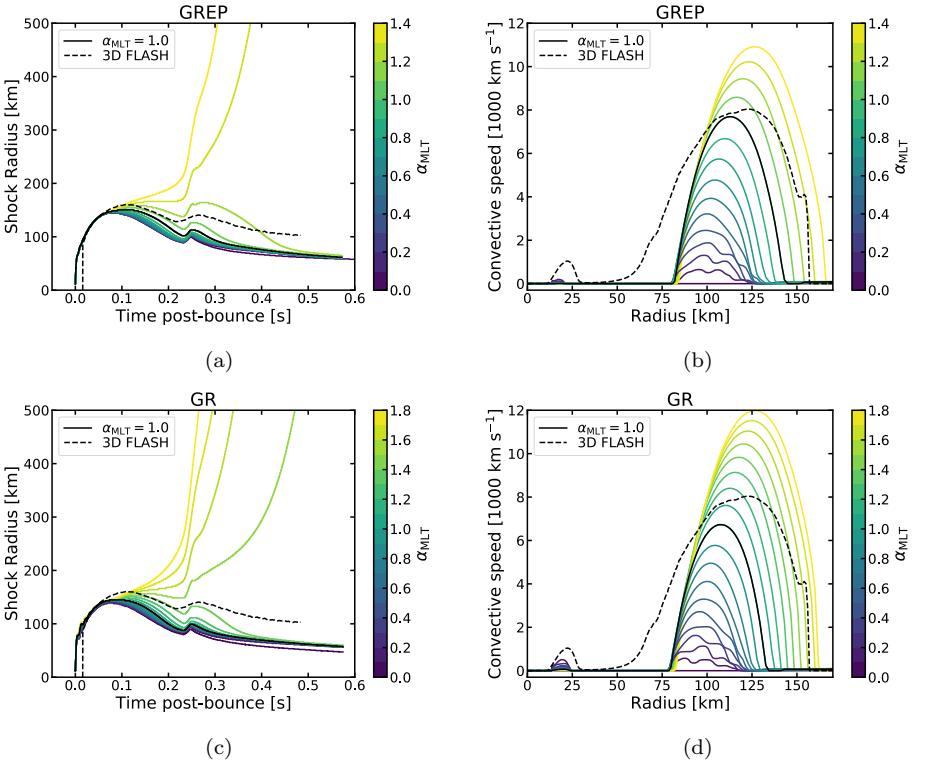


Fig. 1. The plots on the upper row were generated using our GREP model, while the plots on the bottom row were generated using full GR. Panels (a,c) show the time evolution of the shock radius for different values of the parameter $\alpha_{\rm MLT}$, and can be compared to Figure 2 from CWO20. Panels (b,d) show a snapshot at ~ 135 ms post bounce of $v_{\rm turb}$, and can be compared to Figure 1 from CWO20. The dashed lines represent the 3D simulation from Ref. [28].

needed to realistically characterize turbulent convection. If one takes the form of $\omega_{\rm BV}$ from Eq. (8) and implements it in the GREP model, the value of $\alpha_{\rm MLT}$ needed to achieve an explosion increases, becoming comparable to the one used in the GR model.

3.3. *Progenitor Study*

In the previous Sections we summarized the validation of our turbulent convection model[1] by comparing it to the 3D results of Ref. [31]. In this section, we summarize the use[1] of our calibrated models to simulate the collapse and subsequent shock revival of 20 progenitors from Ref. [32]. We use three different values of $\alpha_{\rm MLT}$, for which the fraction of successful explosions is roughly between 25% and 80%.

Our GREP model generates results[1] that are compatible with the ones obtained by CWO20 shifted by $\Delta\alpha_{\rm MLT} \simeq 0.05$. If one compares the left panel of Figure 2 with Figure 6 from CWO20, it is clear that our model tends to yield explosions for

slightly smaller values of $\alpha_{\rm MLT}$. This shift mainly depends on two differences between our model and the one from CWO20: (i) we used a finer resolution in space and energy; (ii) the numerical algorithms used to solve the hydrodynamic equations and the neutrino transport are different.

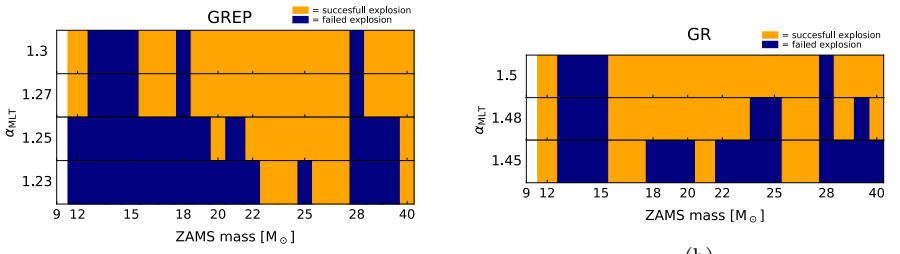


Fig. 2. Explosion pattern (modified from Ref. [1]) of CCSN for the GREP (left panel) and GR (right panel) models as a function of the Zero Age Main Sequence mass. Orange bands represent successful explosions (i.e. the shock has reached 500 km), while dark blue bands represent failed explosions.

Even more importantly, Figure 2 shows that a General Relativistic treatment of turbulent convection does not simply reproduce the results obtained by the GREP model. By looking at Figure 1 one might conclude that, since the value of $\alpha_{\rm MLT}$ required to generate an explosion in GR is larger, using GR with larger values of $\alpha_{\rm MLT}$ would produce the same patterns shown in the left panel of Fig. 2. However, that is not the case, and GR modifies the explosion pattern of CCSNe. To even more accurately characterize the differences between the patterns of explodability in the GR and GREP models, a systematic study with hundreds of progenitors and more values of $\alpha_{\rm MLT}$ would be desired. That, however, is beyond the scope of this conference proceedings.

We can conclude from Figure 2 that general relativity changes which progenitors are more likely to explode. By focusing on the patterns associated with $\alpha_{\rm MLT} = 1.27$ and $\alpha_{\rm MLT} = 1.48$ one can see that, using the GREP model, the 24, 25 and 30 $M_\odot$ progenitors explode, whereas the 18 $M_\odot$ doesn't. In the GR model it is the exact opposite. It should be pointed out that the pattern of explodability generated by the GR model with $\alpha_{\rm MLT} = 1.48$ is intermediate between the results of CWO20 and Ref. [32]: (i) like the former (but unlike the latter), it shows failed explosions for low mass progenitors with M = 13-15 $M_\odot$; (ii) like the latter (and unlike the former) it shows that higher mass progenitors with M = 24-25 $M_\odot$ result in failed explosions.

It is worth mentioning that the explodability pattern obtained using the GR model with $\alpha_{\rm MLT} = 1.48$ cannot be reproduced by the GREP model. One can see that, from the left panel of Figure 2, for $\alpha_{\rm MLT} = 1.23$, the 25 $M_\odot$ progenitor fails, and all progenitors below 22 $M_\odot$ fail as well. This shows that the GREP model

cannot produce successful explosions for progenitors with masses 15-18 $M_\odot$ and at the same time failed explosions of the 24-25 $M_\odot$ progenitors, like we see in the right panel of Figure 2. Notably, the GR model with $\alpha_{\rm MLT} = 1.5$ reproduces the same pattern of explodability found in the GREP model with $\alpha_{\rm MLT} = 1.27$ (with the only exception of the 18 $M_\odot$ progenitor). This tells us that: (i) the threshold between failed and successful explosions is a steep function of $\alpha_{\rm MLT}$; and (ii) GR can reproduce the GREP results for large values of $\alpha_{\rm MLT}$ and large explosion fractions.

Overall, our results[1] have shown that including general relativistic effects can modify how turbulence behaves in a one-dimensional, MLT-like model. It is hard to predict if this effect will translate to multi-dimensional simulations, given the differences between 1D and multi-D. Nonetheless, these results suggest that GR (as opposed to GREP) can have a significant impact on the explosion of CCSNe. A detailed comparison between full-GR and Newtonian simulations, performed with 2D and 3D codes across multiple progenitors, will clarify whether this effect translates to higher dimensions, where turbulent convection is generated self-consistently.

4. Conclusions

In this proceedings contribution we have summarized our development[1] of extended *STIR*, the MLT-like model of CWO20, to a full general relativistic formalism. Our implementation of STIR in `GR1D` can reproduce the results of CWO20 when using the same GREP model that they developed. The GR version of STIR needs larger values of $\alpha_{\rm MLT}$ to achieve shock dynamics that are similar to the ones obtained with a GREP model.[1] The reason behind this is that, as can be seen from Eq. (8), one needs to include the gradient of the internal energy gradient in the expression for $\omega_{\rm BV}$. This reduces the magnitude of $\omega^2_{\rm BV}$ in the gain region, which inhibits the generation of turbulence. The net result is that larger values of $\Lambda_{\rm mix}$ (and therefore of $\alpha_{\rm MLT}$) are needed to develop a convective mixing that is as strong as the one obtained without the inclusion of GR.

After comparing our model to the 3D results of Ref. [28], we simulated[1] the collapse and subsequent shock expansion of 20 different progenitors[32] for different values of $\alpha_{\rm MLT}$, for both the GREP and GR models. Our main finding[1] is that GR changes the pattern of explodability of CCSNe. Specifically, the 24 $M_\odot$ and the 25 $M_\odot$ progenitors need comparatively much larger values of $\alpha_{\rm MLT}$ to explode with the GR model. This produces an explodability as a function of progenitor mass that is intermediate between the results of CWO20 and Ref. [32]. However, the GR model also shows, for values of $\alpha_{\rm MLT}$ that yield large explosion fractions (i.e. $\alpha_{\rm MLT} = 1.5$), an explodability that is compatible with the results obtained using the GREP model.

Acknowledgments

The authors would like to thank Sean Couch, Andre da Silva Schneider and Mackenzie Warren for fruitful discussions. Work at the University of Notre Dame supported

by the U.S. Department of Energy under Nuclear Theory Grant DE-FG02-95-ER40934. EOC would like to acknowledge Vetenskapsrådet (the Swedish Research Council) for supporting this work under award numbers 2018-04575 and 2020-00452.

References

1. L. Boccioli, G. J. Mathews and E. P. O'Connor, General Relativistic Neutrino-driven Turbulence in One-dimensional Core-collapse Supernovae, *ApJ* **912**, p. 29 (May 2021).
2. S. A. Colgate and R. H. White, The Hydrodynamic Behavior of Supernovae Explosions, *ApJ* **143**, p. 626 (March 1966).
3. W. D. Arnett, Gravitational collapse and weak interactions, *Canadian Journal of Physics* **44**, 2553 (January 1966).
4. G. Sonneborn, B. Altner and R. P. Kirshner, The Progenitor of SN 1987A: Spatially Resolved Ultraviolet Spectroscopy of the Supernova Field, *ApJ* **323**, p. L35 (December 1987).
5. H. A. Bethe and J. R. Wilson, Revival of a stalled supernova shock by neutrino heating, *ApJ* **295**, 14 (August 1985).
6. S. W. Bruenn, Stellar core collapse - Numerical model and infall epoch, *ApJS* **58**, 771 (August 1985).
7. D. S. Miller, J. R. Wilson and R. W. Mayle, Convection above the Neutrinosphere in Type II Supernovae, *ApJ* **415**, p. 278 (September 1993).
8. M. Herant, W. Benz, W. R. Hix, C. L. Fryer and S. A. Colgate, Inside the Supernova: A Powerful Convective Engine, *ApJ* **435**, p. 339 (November 1994).
9. C. L. Fryer and M. S. Warren, Modeling Core-Collapse Supernovae in Three Dimensions, *ApJL* **574**, L65 (July 2002).
10. B. Müller, H.-T. Janka and A. Marek, A New Multi-dimensional General Relativistic Neutrino Hydrodynamics Code for Core-collapse Supernovae. II. Relativistic Explosion Models of Core-collapse Supernovae, *ApJ* **756**, p. 84 (September 2012).
11. E. J. Lentz, S. W. Bruenn, W. R. Hix, A. Mezzacappa, O. E. B. Messer, E. Endeve, J. M. Blondin, J. A. Harris, P. Marronetti and K. N. Yakunin, Three-dimensional Core-collapse Supernova Simulated Using a 15 $M_{\odot}$ Progenitor, *ApJL* **807**, p. L31 (July 2015).
12. H.-T. Janka, T. Melson and A. Summa, Physics of Core-Collapse Supernovae in Three Dimensions: A Sneak Preview, *Annual Review of Nuclear and Particle Science* **66**, 341 (October 2016).
13. S. W. Bruenn, E. J. Lentz, W. R. Hix, A. Mezzacappa, J. A. Harris, O. E. B. Messer, E. Endeve, J. M. Blondin, M. A. Chertkow, E. J. Lingerfelt, P. Marronetti and K. N. Yakunin, The Development of Explosions in Axisymmetric Ab Initio Core-collapse Supernova Simulations of 12-25 M Stars, *ApJ* **818**, p. 123 (February 2016).
14. E. P. O'Connor and S. M. Couch, Two-dimensional Core-collapse Supernova Explosions Aided by General Relativity with Multidimensional Neutrino Transport, *ApJ* **854**, p. 63 (February 2018).
15. B. Müller, T. M. Tauris, A. Heger, P. Banerjee, Y.-Z. Qian, J. Powell, C. Chan, D. W. Gay and N. Langer, Three-dimensional simulations of neutrino-driven core-collapse supernovae from low-mass single and binary star progenitors, *MNRAS* **484**, 3307 (April 2019).
16. A. Burrows, D. Radice, D. Vartanyan, H. Nagakura, M. A. Skinner and J. C. Dolence, The overarching framework of core-collapse supernova explosions as revealed by 3D FORNAX simulations, *MNRAS* **491**, 2715 (January 2020).

17. S. M. Couch and C. D. Ott, The Role of Turbulence in Neutrino-driven Core-collapse Supernova Explosions, *ApJ* **799**, p. 5 (January 2015).
18. E. O'Connor, R. Bollig, A. Burrows, S. Couch, T. Fischer, H.-T. Janka, K. Kotake, E. J. Lentz, M. Liebendörfer, O. E. B. Messer, A. Mezzacappa, T. Takiwaki and D. Vartanyan, Global comparison of core-collapse supernova simulations in spherical symmetry, *Journal of Physics G: Nuclear and Particle Physics* **45**, p. 104001 (sep 2018).
19. S. M. Couch, M. L. Warren and E. P. O'Connor, Simulating turbulence-aided neutrino-driven core-collapse supernova explosions in one dimension, *The Astrophysical Journal* **890**, p. 127 (feb 2020).
20. A. Marek, H. Dimmelmeier, H. T. Janka, E. Müller and R. Buras, Exploring the relativistic regime with Newtonian hydrodynamics: An improved effective gravitational potential for supernova simulations, *A&A* **445**, 273 (January 2006).
21. J. R. Wilson and G. J. Mathews, *Relativistic Numerical Hydrodynamics* (Cambridge University Press, 2003).
22. Q. A. Mabanta and J. W. Murphy, How Turbulence Enables Core-collapse Supernova Explosions, *ApJ* **856**, p. 22 (March 2018).
23. Q. A. Mabanta, J. W. Murphy and J. C. Dolence, Convection-aided Explosions in One-dimensional Core-collapse Supernova Simulations. I. Technique and Validation, *ApJ* **887**, p. 43 (December 2019).
24. B. Müller, M. Viallet, A. Heger and H.-T. Janka, The Last Minutes of Oxygen Shell Burning in a Massive Star, *ApJ* **833**, p. 124 (December 2016).
25. K. S. Thorne, Validity in General Relativity of the Schwarzschild Criterion for Convection, *ApJ* **144**, p. 201 (April 1966).
26. E. O'Connor and C. D. Ott, A new open-source code for spherically symmetric stellar collapse to neutron stars and black holes, *Classical and Quantum Gravity* **27**, p. 114103 (June 2010).
27. E. O'Connor, An Open-source Neutrino Radiation Hydrodynamics Code for Core-collapse Supernovae, *ApJS* **219**, p. 24 (August 2015).
28. E. P. O'Connor and S. M. Couch, Exploring Fundamentally Three-dimensional Phenomena in High-fidelity Simulations of Core-collapse Supernovae, *ApJ* **865**, p. 81 (October 2018).
29. R. Farmer, C. E. Fields, I. Petermann, L. Dessart, M. Cantiello, B. Paxton and F. X. Timmes, On Variations of Pre-Supernova Model Properties, *The Astrophysical Journal Supplement Series* **227**, p. 22 (dec 2016).
30. A. W. Steiner, M. Hempel and T. Fischer, Core-collapse Supernova Equations of State Based on Neutron Star Observations, *ApJ* **774**, p. 17 (September 2013).
31. S. M. Couch and E. P. O'Connor, High-resolution Three-dimensional Simulations of Core-collapse Supernovae in Multiple Progenitors, *ApJ* **785**, p. 123 (April 2014).
32. T. Sukhbold, T. Ertl, S. E. Woosley, J. M. Brown and H.-T. Janka, Core-Collapse Supernovae from 9 to 120 Solar Masses Based on Neutrino-Powered Explosions, *The Astrophysical Journal* **821**, p. 38 (April 2016).

Mountain formation by repeated, inhomogeneous crustal failure in a neutron star

A. D. Kerin[1] and A. Melatos[1,2]

[1] *School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2] *Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), University of Melbourne, Parkville, VIC 3010, Australia*

As a neutron star spins down mechanical strain accumulates in the solid crust up to and beyond the point of failure. To model the repeated macroscopic failure of the crust an idealised cellular automaton is developed, with nearest-neighbour interactions representing strain dissipation and redistribution, and including thermal losses. The probability distribution functions (PDFs) of the size and waiting times of failure events in the automaton are presented. The final failure event of a star's life occurs when the star spins down to $\approx (5 \pm 3)\%$ of its birth frequency with implications for transient events e.g. rotational glitches. In addition the automaton is able to predict the star's mass-quadrupole moment and gravitational wavestrain at all points in the star's life, with relevance to future gravitational wave surveys such as those carried out with the Laser Interferometer Gravitational Wave Observatory (LIGO).

Keywords: Asteroseismology – gravitational waves – stars: evolution – stars: neutron – stars: rotation

1. Introduction

The mechanical failure of the crust of a rotating neutron star, and the associated creation of mountains, has been suggested as a source of continuous gravitational waves,[1–3] and transient events such as rotational glitches[4,5] and fast radio bursts.[6] No such gravitational waves have been observed despite multiple recent searches,[7–13] but theoretical arguments exist that the first detection is close.[2] It should be noted that other reasonable explanations for transients have been suggested, e.g. superfluid vortices in the context of glitches.[14,15]

The macroscopic, tectonic process by which the crust fails and mountains form is unknown.[16,17] By analogy with the Earth, one expects crustal failure to involve large-scale, inhomogeneous elements such as fault lines, plates, subduction zones, stress zones, and so on, even though the tensile properties of a neutron star's crust are likely to be different to those of terrestrial rocks (and indeed are not understood completely at present). Modelling has primarily focused on the microphysics of crustal failure.[18–22] In particular Ref. 19 found local cracking does not occur due to the extreme pressure and the crust material fails globally. There have been some investigations into the evolution of the whole star as it evolves through a quasistatic sequence of elastic equilibria.[23,24,31] However, these do not consider the impact of the stick-slip dynamics of repeated local failures involving inhomogenous tectonic features, e.g. tectonic plates, which lead to history dependent outcomes.

Here we present a model of local mountain formation caused by mechanical failure of the crust on macroscopic scales driven by spin down. The model describes the tectonics of crustal failure through a cellular automaton, similar in spirit to the block-and-spring automata used to describe the far-from-equilibrium, stick-slip dynamics of terrestrial earthquakes.[25] Section 2 deals with how the crust deforms due to spin down, the consequent strain build up, and the effects of repeated local failures. Section 3 briefly presents some of the results of this model, in particular the probability distribution functions (PDFs) of the heat dissipated in failure events and the time between events. Section 4 presents the gravitational wavestrain over time as a function of the automaton parameters. For simplicity we do not consider a magnetic field, which is unlikely to be important in ordinary neutron stars (although this may not be true in magnetars).

This paper summarises the key aspects of the model and its output as developed in Ref. 26. This summary borrows heavily from the structure and content of Ref. 26. Among other things, Figs. 1 and 2 are reproduced from the latter reference.

2. Tectonics of Mountain Formation

2.1. *Global deformation due to spin down*

A neutron star is composed of a superfluid core and a solid crystalline crust.[21,22,27–30] Other components, e.g. transitional layers containing nuclear pasta at the crust-core interface,[16] are neglected here for simplicity.

As a neutron star spins down the balance of centrifugal and gravitational forces changes and so does the equilibrium figure of the crust. We use the method of Ref. 31 to model the secular deformation due to spin down, which in turn allows for the calculation of the mechanical strain at every point in the crust. In Ref. 27 Eqs. (3)-(8) define a vector which describes the change in the star's equilibrium figure due to spin down from which the mechanical strain can be calculated, see also Eqs. (1)-(7) in Ref. 26.

2.2. *Local microscopic failure*

We propose that the strain built up over the course of spin down causes the crust to fail locally when strained beyond breaking, and that some fraction of the energy released from the strained crystal lattice of the crust causes the centre of mass of the failed section to move radially outwards, creating a mountain.

Estimates of the breaking strain of neutron star crust material vary many orders of magnitude, e.g. Ref. 32 found a breaking strain of $\sim 10^{-5}$ and the more recent results of Ref. 19 instead found a breaking strain of ~ 0.1. We follow the results of Ref. 19 and consider a breaking strain in the range 0.075 to 0.11. Additionally Ref. 19 found that the material fails in a global manner, i.e. local cracks do not form, but the simulation volume is $\sim 10^{11}$ fm^3 and the macroscopic crust likely behaves differently (see Sec. 2.3 in this paper and Sec. 2.3 in Ref. 26).

When the crust fails it deforms plastically. When a material is deformed plastically a fraction of the plastic work done, β, is dissipated as heat. The value of β depends on a variety of factors such as the material, current strain, strain rate, and deformation history.[33–37] As per Ref. 35 terrestrial metals such as copper and steel have $0.75 \lesssim \beta \lesssim 0.95$, however such metals are unlikely to be good analogues for neutron star crust material. It is extremely difficult to calculate a realistic value of β and we do not attempt to do so. Instead we conservatively set $\beta = 0.9$, taking the position that the significant majority of the plastic work done is dissipated as heat.

2.3. *Repeated global failure: An automaton*

Ref. 19 found, using a direct simulation of nucleonic crystal (neutron star crust material), that the material fails in a global sense. Due to the extreme pressure local failure, i.e. cracking, is not observed. However the simulation volume is $\sim 10^{11}$ fm^3 and the macroscopic crust likely behaves differently, as noted in Sec. 2.2. In reality there are likely to be macroscopic features like seismic faults. To model these stick-slip and stress-relax dynamics[25] we create an automaton.

We divide the crust into N discrete cells; cells are indexed by i, with $0 \leq i < N$, and have position (r_i, θ_i, ϕ_i). The half-annular shape of the cells (dictated by assumed symmetry) and their coordinates are discussed in detail in Sec. 2.3 and Appendix A of Ref. 26. Each cell is assigned a breaking strain, σ_i, uniformly randomly between 0.075 and 0.11, informed by the stress-strain curve calculated in Ref. 19. Strain is greatest at the base of the crust, so the elastic potential energy density is greatest there too. As such the strain of each cell is evaluated at the base of the crust.

Over the course of spin down the strain of each cell, γ_i, increases. We define failure as when a cell has strain equal to or greater than its breaking strain, $\gamma_i \geq \sigma_i$. A failed volume of material is less able to support mechanical loads and so shifts the load to the neighbouring volume.[38] To model this we implement a nearest-neighbours interaction. Upon failure adjacent cells receive a fraction of the strain "lost" by the failed cell, with

$$\gamma_{i\pm1} \mapsto \gamma_{i\pm1} + \frac{1}{3}D(1-A)\gamma_i, \tag{1}$$

$$\gamma_{N-1-i} \mapsto \gamma_{N-1-i} + \frac{1}{3}D(1-A)\gamma_i, \tag{2}$$

where $0 \leq D \leq 1$ and $0 \leq A \leq 1$ are constants of the automaton, and cells indexed by $i = 0$ and $i = N-1$ are adjacent.[a] $(1-A)\gamma_i$ is the amount of strain "lost" by cell i upon failure. Some portion of the released strain, $D(1-A)\gamma_i$, is redistributed amongst cell i's nearest neighbours as per Eqs. (1) and (2). The remaining portion

[a]The cells closest to the North pole, $i = 0, N-1$, and South pole, $i = \lfloor N/2 \rfloor$ for N odd, or $i = N/2, N/2-1$ for N even, are adjacent to two cells not three. As such during strain redistribution Eq. (2) does not apply and Eq. (1) should be evaluated with a factor of 1/2 rather than 1/3.

of strain, $(1-D)(1-A)\gamma_i$, is converted into plastic work deforming the crust, i.e. moving cell i. After the strain of the failed cell's nearest neighbours is increased the failed cell's own strain is updated accordingly,

$$\gamma_i \mapsto A\gamma_i. \tag{3}$$

The quantity A is the fraction of strain a failed cell retains upon failure, and D is the fraction of strain released from the failed cell that is redistributed to the nearest neighbours. We call A and D the "dissipation" and "redistribution" parameters respectively.

In the failure process described above the crust is deformed plastically. The failed cell i is moved purely radially (θ_i and ϕ_i are unchanged) a distance Δr_i to model the plastic deformation,

$$r_i \mapsto r_i + \Delta r_i. \tag{4}$$

Δr_i is calculated by equating the energy associated with plastic deformation with the work done moving the cell against the gravitational-centrifugal potential a distance Δr_i. The potential is given by[31]

$$\Phi(\vec{r}) = -\pi G\rho_{\text{core}}\left[2R^2 - \frac{2r^2}{3} - \frac{4e^2r^2}{15}P_2(\cos\theta)\right], \tag{5}$$

where G is the gravitational constant, ρ_{core} is the density of the fluid core, and e is the eccentricity of the ellipse defined by the meridional cross-section of the star.

In general the mechanical potential energy U stored in a body of volume V under a strain γ is given by the strain-energy formula

$$\frac{\mu\gamma^2}{2} = \frac{U}{V}, \tag{6}$$

where μ is the shear modulus of the material. The mechanical potential energy density of a given cell i is then $\mu\gamma_i^2/2$. As described above if a given cell i with strain γ_i fails then the amount of strain that is converted into plastic work is $(1-A)(1-D)\gamma_i$. We calculate Δr_i from

$$\frac{\mu(1-\beta)}{2}[(1-A)(1-D)\gamma_i]^2 = \Delta r_i\rho_{\text{crust}}\left.\frac{d\Phi(\vec{r})}{dr}\right|_{\vec{r}=\vec{r}_i}, \tag{7}$$

where ρ_{crust} is the density of the crust, and β is the coefficient of thermal dissipation. As per Sec. 2.2 we take the conservative position of $\beta = 0.9$, i.e. we assume the significant majority of the plastic work done on the crust during failure is dissipated as heat.

The state of each cell in the automaton is described by five numbers $[r_i(t_n), \theta_i(t_n), \phi_i(t_n), \gamma_i(t_n), \sigma_i(t_n)]$, which are updated at each time-step. Here the i index refers to the cell index and t_n corresponds to the time-step. The initial shape of the star is a Maclaurin spheroid with eccentricity $e = 0.1$, a representative value.

The specific steps of the automaton are given:

(1) Choose the values of A and D.
(2) Initialise $[r_i(t_0), \theta_i(t_0), \phi_i(t_0)]$.
(3) Randomly assign each cell a breaking strain in the range $0.075 \leq \sigma_i \leq 0.11$, and set $\gamma_i = 0$.
(4) Using the method of Ref. 31 calculate the deformation vectors and increase the strain of each cell for a small spin down of $\delta\Omega$ and the fiducial values below in Sec. 2.4.
(5) If a cell i has $\gamma_i \geq \sigma_i$ then it undergoes failure.
    (a) Redistribute the strain of failed cell i to its nearest neighbours as per Eqs. (1) and (2) synchronously, i.e. all cells with $\gamma_i \geq \sigma_i$ are identified and fail simultaneously before Eqs. (1) and (2) are applied.
    (b) Move cell i radially as per Eqs. (4) and (7).
    (c) Decrease the strain in cell i as per Eq. (3).
    (d) Assign to cell i a new breaking strain, uniformly randomly in the range $0.075 \leq \sigma_i \leq 0.11$.
    (e) Repeat the steps (5)(a) to (5)(e) until $\gamma_i < \sigma_i$ for all $0 \leq i < N$.
(6) Decrease Ω by $\delta\Omega$.
(7) Repeat from step (4) until one has $\Omega \leq \delta\Omega$.

2.4. *Fiducial parameters*

The fiducial values used for this automaton are: $\mu = 2.4 \times 10^{29}$ Jm^{-3}, total radius $R = 10.5$ km, crust-core radius $R' = 9.5$ km, $\rho_{\rm core} = 6.38 \times 10^{17}$ kgm^{-3}, $\rho_{\rm crust} = 10^{17}$ kgm^{-3}, total stellar mass of 1.4 solar masses, $\Omega(0)/2\pi = 800$ Hz, and $\dot{\Omega}(0)/2\pi = -1 \times 10^{-8}$ Hzs^{-1}. For $A = D = 0.5$ and the above fiducial values the failure of a cell on the equator with $\sigma_i = 0.1$ leads to $\Delta r_i \approx 0.04$ mm. For the purpose of calculating the gravitational wavestrain we consider a star that is a distance $d = 1$ kpc away from the Earth.

In this automaton we set $N = 400$. The typical angular velocity step per automaton time-step is $\delta\Omega = (1.25 \times 10^{-4})\Omega(0)$. To convert $\Omega(t)$ into time t in what follows, we use $\Omega(t) = \Omega(0)(1 + t/\tau)^{-1/2}$ where $\tau = -\Omega(0)/2\dot{\Omega}(0) = 4 \times 10^{10}$ s, is the electromagnetic spin-down timescale.

3. Event Statistics

The automaton allows for the location of every part of the crust to be known at every point in time over the course of the star's life. Additionally the time, location, and energy of failure events are tracked. Numerous observables are studied as functions of t in Ref. 26. In Sec. 3.1 we focus on the PDF of event sizes. The PDF of the time between one event and the next one (the waiting time) and the age of the star at which tectonic activity ceases are presented in Sec. 3.2, and the correlation between event size and waiting time is presented in Sec. 3.3.

3.1. *Size PDF*

Mechanical failure of the crust has been linked to multiple types of transient events such as rotational glitches,[4,5] soft gamma-ray repeaters[22,39,40] and fast radio bursts.[6] Testable models of the transients' statistics, such as the automaton presented here, are needed to differentiate between crustal failure and other viable explanations of the transients e.g. superfluid vortices in the context of rotational glitches.[14,41–44]

In the upper panel of Fig. 1 we present the PDFs for the heat dissipated in an event, i.e. its size, given as ΔE. The left-hand side of Eq. (7) gives the amount of energy done as plastic work on the crust not lost as heat. Thus multiplying by $\beta/(1-\beta)$ gives the energy that is dissipated as heat in a failure event. The qualitative form of the PDF is similar across parameter space, a tight peak with a tail in increasing ΔE. Close to $A = D = 0$ events are largest with $\langle \Delta E \rangle \approx 5.4 \times 10^{36}$ J and the tail is most strongly suppressed; events with $\Delta E > \langle \Delta E \rangle$ are relatively less likely than in other regions of parameter space. Close to $A = D = 1$ events are smallest with $\langle \Delta E \rangle \approx 1.9 \times 10^{33}$ J and the tail is much more pronounced; events with $\Delta E > \langle \Delta E \rangle$ are relatively more likely than in other regions of parameter space. The decrease in $\langle \Delta E \rangle$ with increasing A and D follows from Eq. (7). The change in the tail comes from larger A and D more easily facilitating avalanches. More frequent avalanches means larger events are more frequent. For larger A cells remain close to critical after failure, and are thus more prone to fail when a neighbour does. For larger D the nearest-neighbours interaction is simply stronger.

3.2. *Waiting-time PDF*

We define the waiting time of an event, Δt, as the time between that event and the next one. In the lower panel of Fig. 1 we present the PDFs of waiting time for a variety of values of A and D. There is a steep rise for small Δt with an approximately exponential tail for high Δt, the probability density peaks at $\Delta t \approx 0.1\tau$. Recall τ is the spin-down timescale as defined in Sec. 2.4. The waiting-time PDF is approximately exponential, with the peak at $\Delta t \approx 0.1\tau$ an artifact of time discretisation. The frequency decrement, $\delta\Omega$, is constant. Correspondingly the time-steps are small early in the star's life and large later on. As such short waiting-time events are restricted to early in the star's life whereas long waiting-time events can happen throughout. In contrast to the PDF of event sizes the PDF of waiting times is unaffected by A or D. This is because the driver of events is the rate of strain build up, which is governed by spin down and unaffected by A or D.

In order for failure to occur at all we find that the star must be born with an initial rotational frequency $\gtrsim 750$ Hz, consistent with the results of Ref. 27. However once failure does occur, it continues until the star is near totally spun down, $\approx 5 \pm 3\%$ of the birth frequency, i.e. $t/\tau \approx 100$ to 2500. This is because the only way for strain to exit the system is through failure events, it otherwise continually builds up.

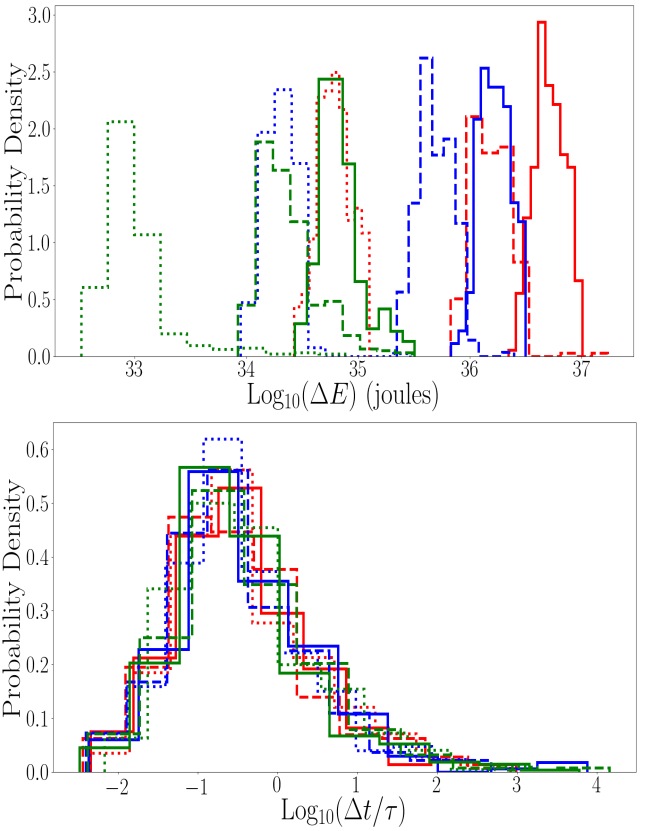


Fig. 1. Size PDFs, measured in joules, for a range of A and D (top panel). Waiting-time PDFs, measured in units of spin-down timescale, τ, for a range of A and D (bottom panel). In both panels red corresponds to $A = 0.1$, blue to $A = 0.5$ and green to $A = 0.9$. The solid lines correspond to $D = 0.1$, dashed to $D = 0.5$ and dotted to $D = 0.9$. Each PDF is constructed from all the events from five simulations. This figure is copied from Ref. 26.

3.3. *Size-waiting-time correlations*

Also of interest are the correlations between the size of an event and the waiting time (recall the waiting time is the time until the next event). We find that the Spearman rank coefficient is ≈ 0.3 uniformly across parameter space. On average larger events affect more of the crust and so, post-failure, cause a larger fraction of the crust to be further from failure, meaning that the average time until the next event is longer. However larger events do not necessarily affect more of the crust, due to the random breaking strains of cells, causing the correlation to be weak. Additionally it is possible for a large event to simply not involve the next-closest-to-failure cell and thus do nothing to delay the next event, further weakening the correlation.

4. Gravitational Radiation

Continuous gravitational waves are the subject of many recent, albeit unsuccessful, searches.[7–13] As such we are motivated to calculate the signal expected from possible sources. Here we consider an isolated spinning-down neutron star with mountains forming due to mechanical failure of the crust. Other potential sources of continuous waves include X-ray binaries[45] and accreting neutron stars.[46] In this model the location of every part of the crust, and the rotational frequency, is known for all of the star's life, so the gravitational wavestrain can be calculated at all times.

In Fig. 2 we plot the wavestrain over time from a neutron star. The two dynamical factors that determine the wavestrain are the changing failure-induced mass-quadrupole moment and the decaying rotational frequency, with $h_0 \propto Q\Omega(t)^2/d$, where h_0 is the gravitational wavestrain, Q is the mass-quadrupole moment, and d is the distance from Earth. At early times spin down is rapid, causing strain to accumulate quickly and many failures to occur, leading to a rapid increase in Q, thus causing the sharp initial rise in Fig. 2. As the star spins down $\dot{\Omega}(t)$ decreases, this slows the rate of mountain creation and causes Q to increase more slowly. At $t/\tau \approx 5$ to 10 the effect of increasing Q balances with the effect of decreasing $\Omega(t)$, causing the wavestrain reach its maximum value. After the peak in wavestrain the decreasing frequency leads to $h_0 \propto (1 + t/\tau)^{-1}$ for $t \gtrsim 10\tau$.

Similar to the size of events the wavestrain is largest near $A = D = 0$ peaking at $h_0 \approx 5 \times 10^{-28}$ and smallest near $A = D = 1$ peaking at $h_0 \approx 1.6 \times 10^{-31}$. Larger events create larger mountains and so lead to a greater waverstrain. Wavestrain reaches a maximum at $t/\tau \approx 5$ to 10, corresponding to $\Omega(t)/2\pi \approx 250$ to 320 Hz. There is no trend with A or D in when the wavestrain peaks. The timing of the peak

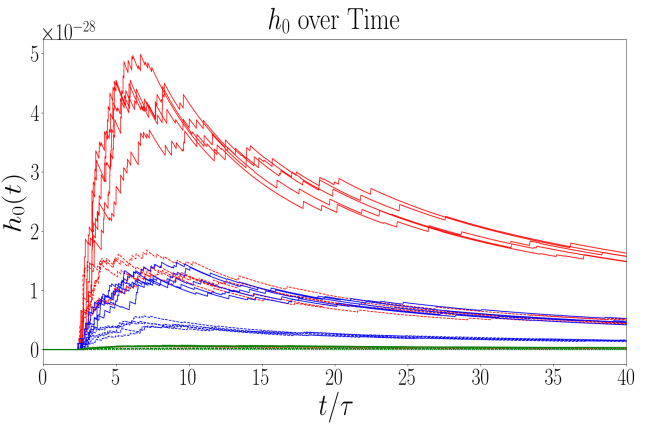


Fig. 2. Gravitational wavestrain, $h_0(t)$, as a function of time, t, in units of the spin-down timescale τ. Each line is the wavestrain of an individual run, of which there are five per value of (A, D). Red corresponds to $A = 0.1$, blue to $A = 0.5$ and green to $A = 0.9$. The solid lines correspond to $D = 0.1$, dashed to $D = 0.5$, dotted to $D = 0.9$. This figure is copied from Ref. 26.

is determined by the rate of events and spin down, neither of which is impacted by A or D.

It should be noted that spin down induced-failure is not the only potential source of mountains on neutron stars. Mountains may form on accreting stars[46] and r-mode oscillations may also lead to a detectable current-quadrupole gravitational wave signal.[47]

5. Conclusion

Electromagnetic spin down drives centrifugal crustal deformation and repeated crustal failure in rotation-powered pulsars. Crustal failure may be responsible for impulsive phenomena like rotational glitches.[4,5] It may also create mountains which lead to the continuous emission of gravitational radiation.[1–3] Molecular dynamics simulations reveal that the crust fails homogeneously, without cracking, under the high pressures on neutron stars.[19] However the simulations are conducted on picometre scales and cannot resolve inhomogeneities on macroscopic length scales (e.g. $\gtrsim 1$ m) such as tectonic plates, faults, and so on, which are seeded by dislocations and other impurities and grow under repeated failure. In this paper, we develop a phenomenological model in the form of a cellular automaton to capture - in a highly idealised form - some of the far-from-equilibrium physics of repeated failure.

We present an idealised cellular automaton to model the evolution of the crust as it both spins down secularly and deforms plastically due to mechanical failure. The secular deformation and strain build up are modelled using the method of Ref. 31. We make use of a nearest-neighbours interaction to model the redistribution and dissipation of strain and the plastic deformation following crustal failure, including the effects of thermal dissipation.

In this model the time, location and size of every failure event are known, as is position of every part of the crust at every point in time. This allows for a wide variety of observables to be predicted, e.g. total tectonic activity, rate of tectonic activity, and number of events. These observables, among others, are investigated in greater detail in Ref. 26. We find the ΔE PDF is qualitatively similar for all values of A and D, namely a tight peak with a tail in increasing ΔE. Events are largest near $A = D = 0$ with $\langle \Delta E \rangle \approx 5.4 \times 10^{36}$ J and smallest near $A = D = 1$ with $\langle \Delta E \rangle \approx 1.9 \times 10^{33}$ J. The gravitational wavestrain follows a similar pattern due to larger events creating larger mountains. The maximum emitted wavestrain in the star's life is greatest near $A = D = 0$ with $h_0 \approx 5 \times 10^{-28}$ and smallest near $A = D = 1$ with $h_0 \approx 1.6 \times 10^{-31}$. In contrast we find that the waiting-time PDF is insensitive to A and D.

The key predictions of this model are (i) in order for failure to occur at all the star must be born spinning $\gtrsim 750$ Hz, consistent with the results of Ref. 27, (ii) once the first failure has occurred failures will continue until $\Omega(t) \approx (0.05 \pm 0.03)\Omega(0)$, and (iii) there is a weak positive correlation (Spearman rank coefficient ≈ 0.3) between the size of an event and the time until the next one.

We plan to improve this calculation in future. Refs. 23 and 24 extend the formalism of Ref. 31. These extensions account for the stratification of the star i.e. they allow for the consideration of a star of multiple variously dense layers, rather than requiring a two component crust-core formulation, additionally they model perturbations from chemical equilibrium.

References

1. K. Riles, Gravitational waves: Sources, detectors and searches, *Progress in Particle and Nuclear Physics* **68**, 1 (2013).
2. G. Woan, M. Pitkin, B. Haskell, D. Jones and P. Lasky, Evidence for a minimum ellipticity in millisecond pulsars, *The Astrophysical Journal Letters* **863**, p. L40 (2018).
3. B. Reed, A. Delbel and C. Horowitz, Modeling the galactic neutron star population for use in continuous gravitational wave searches, *arXiv preprint arxiv:2104.007711* (2021).
4. M. Ruderman, Crust-breaking by neutron superfluids and the vela pulsar glitches, *The Astrophysical Journal* **203**, 213 (1976).
5. J. Middleditch, F. E. Marshall, Q. D. Wang, E. V. Gotthelf and W. Zhang, Predicting the starquakes in psr j0537–6910, *The Astrophysical Journal* **652**, p. 1531 (2006).
6. A. G. Suvorov and K. D. Kokkotas, Young magnetars with fracturing crusts as fast radio burst repeaters, *Monthly Notices of the Royal Astronomical Society* **488**, 5887 (2019).
7. B. Abbott, B. P, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt *et al.*, Narrow-band search for gravitational waves from known pulsars using the second ligo observing run, *Physical Review D* **99**, p. 122002 (2019).
8. B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Directional limits on persistent gravitational waves using data from advanced ligo's first two observing runs, *Physical Review D* **100**, p. 062001 (2019).
9. B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Search for gravitational waves from scorpius x-1 in the second advanced ligo observing run with an improved hidden markov model, *Physical Review D* **100**, p. 122002 (2019).
10. B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Searches for continuous gravitational waves from 15 supernova remnants and fomalhaut b with advanced ligo, *The Astrophysical Journal* **875**, p. 122 (2019).
11. B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt *et al.*, Searches for gravitational waves from known pulsars at two harmonics in 2015–2017 ligo data, *The Astrophysical Journal* **879**, p. 10 (2019).
12. R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Gravitational-wave constraints on the equatorial ellipticity of millisecond pulsars, *The Astrophysical journal letters* **902**, p. L21 (2020).
13. M. A. Papa, J. Ming, E. V. Gotthelf, B. Allen, R. Prix, V. Dergachev, H.-B. Eggenstein, A. Singh and S. J. Zhu, Search for continuous gravitational waves from the

central compact objects in supernova remnants cassiopeia a, vela jr., and g347. 3–0.5, *The Astrophysical Journal* **897**, p. 22 (2020).
14. P. Anderson and N. Itoh, Pulsar glitches and restlessness as a hard superfluidity phenomenon, *Nature* **256**, 25 (1975).
15. L. Warszawski and A. Melatos, Gross–pitaevskii model of pulsar glitches, *Monthly Notices of the Royal Astronomical Society* **415**, 1611 (2011).
16. M. Caplan and C. Horowitz, Colloquium: Astromaterial science and nuclear pasta, *Reviews of Modern Physics* **89**, p. 041002 (2017).
17. F. Gittins, N. Andresson and D. Jones, Modelling neutron stars, *arXiv preprint arXiv:2009.12794* (2021).
18. C. Horowitz and J. Hughto, Molecular dynamics simulation of shear moduli for coulomb crystals, *arXiv preprint arXiv:0812.2650* (2008).
19. C. Horowitz and K. Kadau, Breaking strain of neutron star crust and gravitational waves, *Physical Review Letters* **102**, p. 191102 (2009).
20. A. Chugunov and C. Horowitz, Breaking stress of neutron star crust, *Monthly Notices of the Royal Astronomical Society: Letters* **407**, L54 (2010).
21. C. Horowitz, J. Hughto, A. Schneider and D. Berry, Neutron star crust and molecular dynamics simulation, *arXiv preprint arXiv:1109.5095* (2011).
22. D. Baiko and A. Chugunov, Breaking properties of neutron star crust, *Monthly Notices of the Royal Astronomical Society* **480**, 5511 (2018).
23. E. Giliberti, M. Antonelli, G. Cambiotti and P. Pizzochero, Incompressible analytical models for spinning-down pulsars, *Publications of the Astronomical Society of Australia* **36** (2019).
24. E. Giliberti, G. Cambiotti, M. Antonelli and P. Pizzochero, Modelling strains and stresses in continuously stratified rotating neutron stars, *Monthly Notices of the Royal Astronomical Society* **491**, 1064 (2020).
25. R. Burridge and L. Knopoff, Model and theoretical seismicity, *Bulletin of the seismological society of america* **57**, 341 (1967).
26. A. Kerin and A. Melatos, Mountain formation by repeared, inhomogeneous crustal failure in a neutron star, *Submitted to the Monthly Notices of the Royal Astronomical Society* (2021).
27. F. Fattoyev, C. Horowitz and H. Lu, Crust breaking and the limiting rotational frequency of neutron stars, *arXiv preprint arXiv:1804.04952* (2018).
28. A. I. Chugunov, Neutron star crust in voigt approximation: general symmetry of the stress–strain tensor and an universal estimate for the effective shear modulus, *Monthly Notices of the Royal Astronomical Society: Letters* **500**, L17 (2020).
29. A. Kozhberov and D. Yakovlev, Deformed crystals and torsional oscillations of neutron star crust, *Monthly Notices of the Royal Astronomical Society* **498**, 5149 (2020).
30. A. A. Kozhberov, Breaking stress of coulomb crystals in the neutron star crust, *arXiv preprint arXiv:2011.04397* (2020).
31. L. M. Franco, B. Link and R. I. Epstein, Quaking neutron stars, *The Astrophysical Journal* **543**, p. 987 (2000).
32. R. Smoluchowski, Frequency of pulsar starquakes, *Physical Review Letters* **24**, p. 923 (1970).
33. D. Rittel, On the conversion of plastic work to heat during high strain rate deformation of glassy polymers, *Mechanics of Materials* **31**, 131 (1999).
34. P. Rosakis, A. Rosakis, G. Ravichandran and J. Hodowany, A thermodynamic internal variable model for the partition of plastic work into heat and stored energy in metals, *Journal of the Mechanics and Physics of Solids* **48**, 581 (2000).

35. D. Macdougall, Determination of the plastic work converted to heat using radiometry, *Experimental mechanics* **40**, 298 (2000).
36. J. Hodowany, G. Ravichandran, A. Rosakis and P. Rosakis, Partition of plastic work into heat and stored energy in metals, *Experimental mechanics* **40**, 113 (2000).
37. G. Ravichandran, A. J. Rosakis, J. Hodowany and P. Rosakis, On the conversion of plastic work into heat during high-strain-rate deformation, in *AIP conference proceedings*, (1)2002.
38. R. V. R. Hertzberg and J. Hertberg, *Deformation and Fracture mechanics of engineering materials*, 5 edn. (John Wiley and Sons Inc, 2013).
39. D. Kaplan, S. Kulkarni, M. Van Kerkwijk, R. Rothschild, R. Lingenfelter, D. Marsden, R. Danner and T. Murakami, Hubble space telescope observations of sgr 0526–66: New constraints on accretion and magnetar models, *The Astrophysical Journal* **556**, p. 399 (2001).
40. K. Hurley, S. Boggs, D. Smith, R. Duncan, R. Lin, A. Zoglauer, S. Krucker, G. Hurford, H. Hudson, C. Wigger *et al.*, An exceptionally bright flare from sgr 1806–20 and the origins of short-duration γ-ray bursts, *Nature* **434**, 1098 (2005).
41. L. Warszawski and A. Melatos, Knock-on processes in superfluid vortex avalanches and pulsar glitch statistics, *Monthly Notices of the Royal Astronomical Society* **428**, 1911 (2013).
42. W. Fulgenzi, A. Melatos and B. Hughes, Radio pulsar glitches as a state-dependent poisson process, *Monthly Notices of the Royal Astronomical Society* **470**, 4307 (2017).
43. A. Melatos, G. Howitt and W. Fulgenzi, Size-waiting-time correlations in pulsar glitches, *The Astrophysical Journal* **863**, p. 196 (2018).
44. J. B. Carlin and A. Melatos, Autocorrelations in pulsar glitch waiting times and sizes, *Monthly Notices of the Royal Astronomical Society* **488**, 4890 (2019).
45. G. Ushomirsky, L. Bildsten and C. Cutler, Gravitational waves from low-mass x-ray binaries: A status report, in *AIP Conference Proceedings*, (1)2000.
46. G. Ushomirsky, C. Cutler and L. Bildsten, Deformations of accreting neutron star crusts and gravitational wave emission, *Monthly Notices of the Royal Astronomical Society* **319**, 902 (2000).
47. N. Andersson and K. D. Kokkotas, The r-mode instability in rotating neutron stars, *International Journal of Modern Physics D* **10**, 381 (2001).

Gravitational waves from neutrino mass generating phase transitions

Nobuchika Okada
Department of Physics and Astronomy, University of Alabama, Tuscaloosa, Alabama 35487, USA

Osamu Seto*
Institute for the Advancement of Higher Education, Hokkaido University, Sapporo 060-0817, Japan, and
Department of Physics, Hokkaido University, Sapporo 060-0810, Japan
**E-mail: seto@particle.sci.hokudai.ac.jp*

Some particle physics models with an additional $U(1)$ gauge interaction are interesting because those address the origin of neutrino masses. We show that, in a wide class of models, such an extra $U(1)$ gauge symmetry breaking in the early universe can be first-order phase transition and hence generate a detectable amplitude of stochastic gravitational wave radiation in future experiments. We also discuss the parameter dependence and a possible UV completion.

Keywords: Gravitational waves; Beyond the standard model.

1. Introduction

The detection of gravitational waves (GWs) provides information on the evolution of the Universe at an early stage. Cosmological GWs could originate from, for instance, quantum fluctuations during inflationary expansion[1] and phase transitions.[2,3] The first-order phase transition in the early Universe is one of the promising sources of GWs stochastic background.[4–6] After a first-order phase transition, the dynamics of bubble collision[7–11] and subsequent turbulence of the plasma[12–16] and sonic waves generate GWs.[17–19] These might be within the reach of future space interferometer experiments such as the Big Bang Observer (BBO)[20] and DECi-hertz Interferometer Observatory (DECIGO)[21]; or even ground-based detectors such as Advanced LIGO (aLIGO),[22] and Einstein Telescope (ET),[23] if the transition temperature is below $\mathcal{O}(10^7)$ GeV.[24–26]

One of the important open questions in particle physics is the origin of nonvanishing neutrino masses established by various neutrino oscillation phenomena. The most attractive idea for explaining the tiny neutrino masses is the so-called seesaw mechanism with heavy Majorana right-handed (RH) neutrinos N_R.[27–30] From analogy to the fact that masses of gauge bosons and fermions in the standard model (SM) of particle physics are generated by the breakdown of the electroweak SM gauge symmetry, it is natural to suppose that the masses of heavy RH neutrinos are also generated by developing the vacuum expectation value (VEV) of a Higgs field which breaks a certain symmetry at a high energy scale.

Our Universe has undergone several phase transitions since its birth. The well-known examples are the quark-hadron, aka QCD, transition and the electroweak phase transition. Other promising transitions would be associated with the generation of RH neutrino Majorana masses, and the breakdown of the Great Unified Theory (GUT) gauge symmetry.

The spectrum of stochastic GWs produced by the first-order electroweak phase transition with an extended Higgs sector has been investigated due to interest in the electroweak baryogenesis.[4–6] Among GWs generated by an extra $U(1)$ symmetry breaking,[31–41] one of the simplest and the most promising symmetry is the gauged $U(1)_{B-L}$ (baryon number minus lepton number) symmetry,[42–45] since the SM also poses the $B-L$ symmetry as an accidental global symmetry. In the standard $U(1)_{B-L}$ charge assignment, three RH neutrinos must be introduced to satisfy the conditions for cancellation of the gauge and gravity anomaly. The $U(1)_{B-L}$ may come from a larger GUT symmetry.

In the case that the $U(1)_{B-L}$ symmetry breaking takes place at an energy scale higher than the TeV scale, it is very difficult for any collider experiments to address the mechanism of the symmetry breaking and the RH neutrino mass generation. On the other hand, GWs from the first-order phase transition associated with the spontaneous $U(1)_{B-L}$ gauge symmetry breaking can probe the mechanism. This is what we point out here. For GWs generated by a TeV scale $U(1)_{B-L}$ phase transition, see literature.[32,33,38]

First, we investigate GWs from the first-order phase transition associated with the spontaneous $U(1)_{B-L}$ gauge symmetry breaking within the minimal Higgs model[39] and a non-minimal Higgs model.[35] We show that the first-order phase transition of the Higgs $B-L$ potential can generate a sufficiently large GWs amplitude to be detected in future experiments.

Next, we consider an ultraviolet (UV) completion of such an extra $U(1)$ extended SM. A primary candidate for the completion is an $SO(10)$ GUT model, in which the extra $U(1)$ gauge group along with the SM gauge group is embedded, and all the SM fermions and RH neutrinos in each generation are also unified into a single **16** representation of $SO(10)$ (see, for example, a review[46] and references therein). Among several possible paths of symmetry breaking from the $SO(10)$ to the SM gauge group, we consider the following two step breaking. The first breaking $SO(10)$ into $SU(5)\times U(1)$ takes place at a very high scale $M_{\mathrm{SO}(10)}$. As the second step, the $SU(5)$ breaks to the SM gauge group at a scale $M_{\mathrm{SU}(5)} \simeq 10^{16}$ GeV. We show that the parameter set compatible with $SO(10)$ unification can produce a detectable GWs spectrum.[41]

This paper is organized as follows: In the next section, we describe the minimal and non-miminal $U(1)_{B-L}$ model and then derive the resultant GWs spectrum by estimating the latent heat and the transition timescale of the phase transition. In Sect. 3, after we describe the outline for the $SO(10)$ unification of the $U(1)$ extended SM based on the gauge group of $SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_X$.

through an intermediate path with the $SU(5) \times U(1)_X$ unification, we compute the GWs spectrum generated by the first-order phase transition of $U(1)_X$ and present the resultant GWs spectrum for various sets of the model parameters. The last section is devoted to our summary.

2. Gravitational waves from $U(1)_{B-L}$ phase transition

After we briefly summarize the properties of GWs produced by a first-order phase transition in the early Universe, we show the amplitude and the spectrum of GWs from the first-order phase transition associated with the spontaneous $U(1)_{B-L}$ gauge symmetry breaking within the minimal Higgs model[39] and a non-minimal Higgs model.[35]

2.1. *Scalar potential parameters related to the GW spectrum*

We consider the following two types of tree-level scalar potential. One is the potential for the minimal model:

$$V_0(\Phi_2) = -M_{\Phi_2}^2 \Phi_2 \Phi_2^\dagger + \frac{1}{2}\lambda_2 (\Phi_2 \Phi_2^\dagger)^2, \tag{1}$$

and the other is a non-minimal form with two Higgs fields:

$$\begin{aligned} V_0(\Phi_1, \Phi_2) = {} & \frac{1}{2}\lambda_1 (\Phi_1 \Phi_1^\dagger)^2 + \frac{1}{2}\lambda_2 (\Phi_2 \Phi_2^\dagger)^2 + \lambda_3 \Phi_1 \Phi_1^\dagger (\Phi_2 \Phi_2^\dagger) \\ & + M_{\Phi_1}^2 \Phi_1 \Phi_1^\dagger - M_{\Phi_2}^2 \Phi_2 \Phi_2^\dagger - A(\Phi_1 \Phi_1 \Phi_2^\dagger + \Phi_1^\dagger \Phi_1^\dagger \Phi_2). \end{aligned} \tag{2}$$

The subscript i of Φ_i stands for the $B - L$ charge of each Φ_i field. Here, M_{Φ_i}, λ_i, and A are real positive parameters. We omit the SM Higgs field part and its interaction terms because of those little importance. The Yukawa interactions of N_R are given by

$$\mathcal{L}_{Yukawa} \supset -\sum_{i=1}^{3}\sum_{j=1}^{3} Y_D^{ij} \overline{l_L^i} H N_R^j - \frac{1}{2}\sum_{k=1}^{3} Y_{N^k} \Phi_2 \overline{N_R^{k\,C}} N_R^k + \text{H.c.}, \tag{3}$$

where the first term is the neutrino Dirac Yukawa coupling, and the second is the Majorana Yukawa couplings. Once the Higgs field Φ_2 develops a nonzero VEV, the $U(1)$ gauge symmetry is broken and the Majorana mass terms of the RH neutrinos and the mass of the extra gauge boson are generated. After the electroweak symmetry breaking, tiny neutrino masses are generated through the seesaw mechanism. The experimental bounds on the extra gauge boson are derived as the LEP constraint $m_{Z'}/g_{B-L} \gtrsim 6$ TeV[47,48] and the LHC Run 2 constraints $m_{Z'} \gtrsim 3.9$ TeV[49–52] for $g_{B-L} \simeq 0.7$.

In order to realize a first-order phase transition, the magnitude of the cubic term in the scalar potential

$$V(\varphi) \sim m^2 \varphi^2 - E\varphi^3 + \lambda\varphi^4, \tag{4}$$

is crucial, where φ is defined as $\Phi = \varphi/\sqrt{2}$. There are two possibilities of the origin of the cubic term: One case is that the effective cubic term is induced as loop thermal corrections, and the other is that the term exists in the tree level potential. The cases (1) and (2) above are representative samples of each case. The former corresponds to the case of thermal correction and the later does to the tree level model respectively.

In calculation of the phase transition, we include one-loop corrections to the scalar potential for zero temperature $\Delta V_{1-\text{loop}}(\varphi)$ and that for a finite temperature T given by

$$\Delta V_T(\varphi) = \sum_s g_s \frac{T^4}{2\pi^2} J_B(m_s^2(\varphi)/T^2) - \sum_f g_f \frac{T^4}{2\pi^2} J_F(m_f^2(\varphi)/T^2) + \sum_v g_v \frac{T^4}{2\pi^2} J_B(m_v^2(\varphi)/T^2), \tag{5}$$

where g_i, with $i = s$ (scalars), f (fermions) and v (vectors) denotes the number of internal degrees of freedom, $m_i(\varphi)$ are φ dependent masses for i-th particles. $J_{B(F)}$ is an auxiliary function in thermal corrections.[53,54] As is well known, $J_B(x^2)$ is expressed as

$$J_B(x^2) \simeq -\frac{\pi^4}{45} + \frac{\pi^2}{12}x^2 - \frac{\pi}{6}(x^2)^{3/2} + ..., \tag{6}$$

at the high temperature expansion. The third term induces the effective cubic term that could make the transition of the first order even if the tree level potential does not contain a cubic term as in the potential (1).

For our numerical calculations, we have implemented our models into the public code `CosmoTransitions`,[55] where both zero- and finite-temperature one-loop effective potentials with the corrections for resummation,

$$V_{\text{eff}}(\varphi, T) = V_0(\varphi) + \Delta V_{1-\text{loop}}(\varphi) + \Delta V_T(\varphi, T), \tag{7}$$

have been calculated in the $\overline{\text{MS}}$ renormalization scheme. Here, as a caveat, we note that there is a long-standing open problem of gauge dependence on the use of the effective Higgs potential. Our results are also subjects of this issue[56,57] and should be regarded as a reference value.

By using `CosmoTransitions`, we find the bounce solution of φ and the bounce action for the transition, the bubble nucleation temperature $T_\star$,[58] and the latent heat energy density ϵ in the radiation dominated Universe. With the bounce solution and action, we obtain the dimensionless transition timescale

$$\frac{\beta}{H_\star} \simeq T\frac{dS}{dT}\bigg|_{T_\star} = T\frac{d(S_E^3/T)}{dT}\bigg|_{T_\star}, \tag{8}$$

with

$$\beta \equiv -\frac{dS}{dt}\bigg|_{t_\star}, \tag{9}$$

from the bounce action[10] in the four-dimensional Minkowski space S and in the three-dimensional Euclidean space S_E^3. Here, $t_\star$ is the bubble nucleation time.

2.2. *The GW spectrum*

There are three main processes and mechanisms of GWs production: bubble collisions, turbulence and sound waves after bubble collisions. The resulting spectrum of the generated background GWs is given by

$$\Omega_{GW}(f) = \Omega_{GW}^{\rm coll}(f) + \Omega_{GW}^{\rm sw}(f) + \Omega_{GW}^{\rm turb}(f), \tag{10}$$

in terms of the density parameter Ω_{GW}. The three terms on the right side indicate the GW produced by bubble collisions, sound waves, and turbulence, respectively. We adopt the fitting formula of spectrum commonly adopted in literature[4–6] as well as our papers.[35,39,41]

We only show predicted GWs spectra for our benchmark points of the minimal Higgs model (1) with expected sensitivities of various future experiments,[59] while one may find the similar spectrum of GWs in non-minimal model.[35] One benchmark point $(g_{B-L}, v_2, \lambda_2) = (0.44, 4\,\text{TeV}, 1.5\times10^{-4})$ is shown by the green curve, while the other $(g_{B-L}, v_2, \lambda_2) = (0.46, 3.8\times10^3\,\text{TeV}, 4.0\times10^{-4})$ is shown by the red curve in Fig. 1. The green curve point is similar to the lowest energy case satisfying the LHC limit. An approximate relationship can be seen that $f_{\rm peak}$ is proportional to the VEV of Φ_2, v_2. Next, we discuss the resulting GWs spectrum gauge coupling dependency. The GWs spectrum for various values of the $B-L$ gauge coupling constant for the fixed value of $v_2 = 10$ TeV and $\lambda_2 = 0.002$ is shown in Fig. 2. We have found a mild dependence for the frequency but the amplitude is quite sensitive. The largest amplitude is obtained for $0.35 \lesssim g_{B-L} \lesssim 0.4$. Here, we note that the GWs amplitude in Figs. 1 and 2 have been re-evaluated with taking the suppression effect due to the short-lasting sonic wave[60–64] into account and smaller than previous estimations.[39]

3. UV completion by SO(10)

3.1. *From $U(1)_{B-L}$ to $U(1)_X$*

We consider the UV completion of the $U(1)$ extended SM by $SO(10)$ GUT via the intermediate step of $SU(5)\times U(1)_X$ unification. To realize this partial unification, we generalize $U(1)_{B-L}$ of the minimal $B-L$ model to $U(1)_X$, under which the charge of an SM field is defined as a linear combination of its hyper-charge and $B-L$ charge, $q_X = Yx + Q_{B-L}$, with x being a real constant.[65,66] The particle content of this model is listed in Table 1. Except for the introduction of the new parameter x, the model properties are quite similar to those of the minimal $B-L$ model, which is realized as the special case of $x=0$. Another limit $|x| \gg 1$ is called a "hyper-charge oriented" case.[67–69]

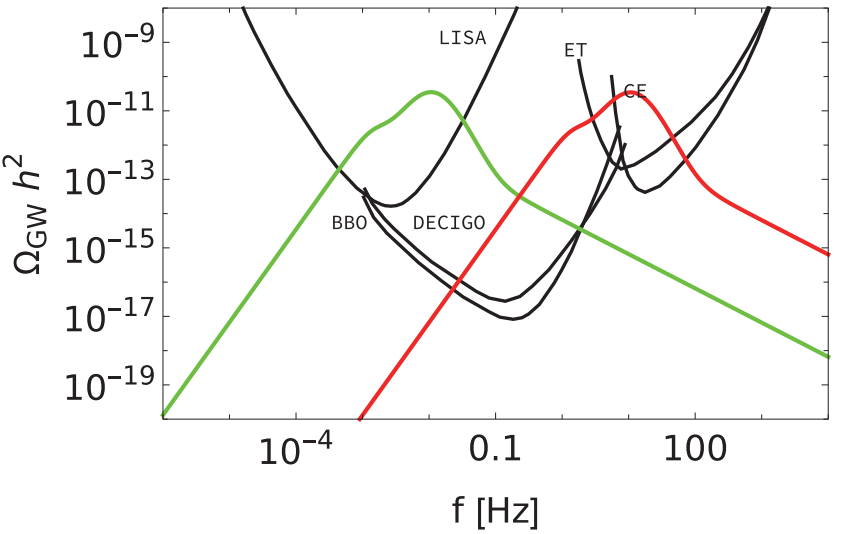


Fig. 1. The predicted GWs spectra for two benchmark points are shown. The expected sensitivities of each indicated experiments[59] are shown by black solid curves.

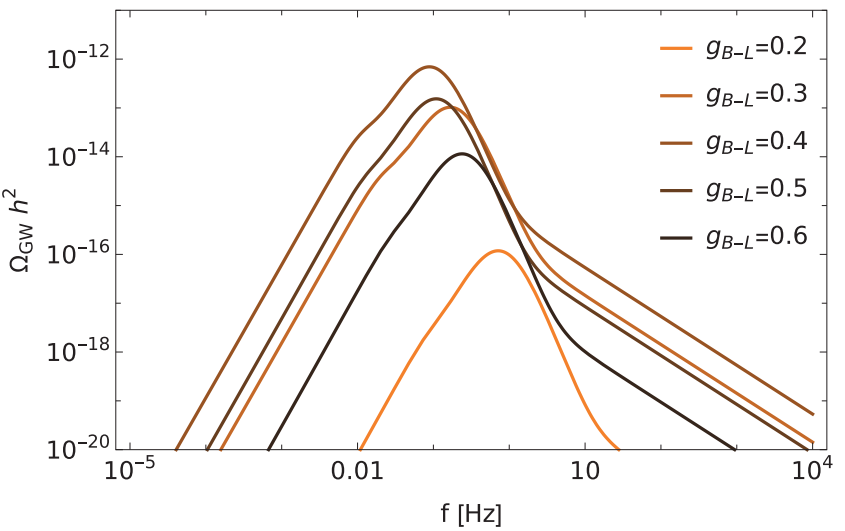


Fig. 2. The predicted GWs spectrum for various values of g_{B-L} for $v_2 = 10$ TeV and $\lambda_2 = 0.002$.

3.2. $SU(5) \times U(1)_X$ *unification*

3.2.1. $SU(5) \times U(1)_X$ *embedding*

We now consider the embedding,

$$SU(5) \times U(1)_X \supset SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_X. \tag{11}$$

Table 1. The particle content of the minimal $U(1)_X$ model. In addition to the SM particle content ($i = 1, 2, 3$), three RH neutrinos N_R^i ($i = 1, 2, 3$) and one $U(1)_X$ Higgs field Φ_2 are introduced.

	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	$U(1)_X$
q_L^i	**3**	**2**	$\frac{1}{6}$	$\frac{1}{6}x + \frac{1}{3}$
u_R^i	**3**	**1**	$\frac{2}{3}$	$\frac{2}{3}x + \frac{1}{3}$
d_R^i	**3**	**1**	$-\frac{1}{3}$	$-\frac{1}{3}x + \frac{1}{3}$
l_L^i	**1**	**2**	$-\frac{1}{2}$	$-\frac{1}{2}x - 1$
e_R^i	**1**	**1**	-1	$-x - 1$
N_R^i	**1**	**1**	0	-1
H	**1**	**2**	$-\frac{1}{2}$	$-\frac{1}{2}x$
Φ_2	**1**	**1**	0	-2

As in the standard $SU(5)$ GUT,[70] the charge conjugation of RH down quarks and left-handed leptons are embedded in $\mathbf{5}^*$ representation of $SU(5)$, while left-handed quarks, the charge conjugation of RH up quarks, and the charge conjugation of RH charged leptons are embedded in **10** representation:

$$\mathbf{5}^* \supset d_R^{iC} \oplus \ell_L^i, \qquad \mathbf{10} \supset q_L^i \oplus u_R^{iC} \oplus \ell_R^{iC}. \tag{12}$$

This is possible for $x = -4/5$ and hence the $SU(5)$ unification leads to a quantization of $U(1)_X$ charge.[71]

3.2.2. *Gauge coupling unification to $SU(5)$*

A simple setup to achieve the unification of the three SM gauge couplings is to introduce two pairs of vector-like quarks ($Q + \bar{Q}$ and $D + \bar{D}$) with TeV scale masses, M_Q and M_D, respectively. Their representations are listed in Table 2. In the presence of the exotic quarks, the SM gauge couplings are successfully unified at $M_{\mathrm{SU(5)}} \simeq 10^{16}$ GeV.[72–78] This unification scale corresponds to the proton lifetime of $\tau_p \simeq 10^{38}$ yr, which is much longer than the current experimental lower limit of $\tau(p \to \pi^0 e^+) \simeq 10^{34}$ yr reported by the Super-Kamiokande collaboration.[79]

Table 2. Representations of the vector-like quarks.

	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	$U(1)_X$
Q	**3**	**2**	$\frac{1}{6}$	$\frac{1}{5}$
$\bar{Q}$	$\mathbf{3}^*$	**2**	$-\frac{1}{6}$	$-\frac{1}{5}$
D	**3**	**1**	$-\frac{1}{3}$	$\frac{3}{5}$
$\bar{D}$	$\mathbf{3}^*$	**1**	$\frac{1}{3}$	$-\frac{3}{5}$

We show the gauge coupling unification by solving the renormalization group (RG) equations of the SM gauge couplings including the contributions of exotic quarks Q and D also into the beta function coefficients. Our results are shown in Fig. 3. We set the vector-like quark masses to be $M_Q = M_D = 1.5$ TeV, which

satisfy the latest LHC bounds.[80,81] We find that three SM gauge couplings are successfully unified at $M_{\rm SU(5)} \simeq 2.24 \times 10^{16}$ GeV. We will discuss the RG evolution for g_χ in the next subsection.

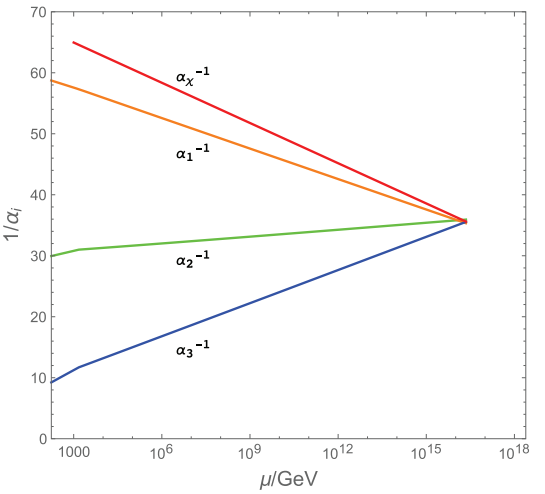


Fig. 3. The RG evolution of the gauge couplings of the $U(1)$ extended SM with the vector-like quarks. Three SM gauge couplings are unified at $M_{\rm SU(5)} \simeq 2.24 \times 10^{16}$ GeV. The results for the case of $M_{\rm SU(5)} = M_{\rm SO(10)}$.

3.3. *SO*(10) *unification*

The $SU(5) \times U(1)_X$ can be embedded into $SO(10)$. The decomposition of several $SO(10)$ multiplets to $SU(5) \times U(1)_X$ is given by[82]:

$$
\begin{aligned}
SO(10) &\supset SU(5) \times U(1)_X \\
\mathbf{10} &= \mathbf{5}(-2/5) + \mathbf{5}^*(2/5), \\
\mathbf{16} &= \mathbf{1}(1) + \mathbf{5}^*(-3/5) + \mathbf{10}(1/5), \\
\mathbf{45} &= \mathbf{1}(0) + \mathbf{10}(-4/5) + \mathbf{10}^*(4/5) + \mathbf{24}(0), \\
\mathbf{126} &= \mathbf{1}(2) + \mathbf{5}^*(2/5) + \mathbf{10}(6/5) + \mathbf{15}^*(-6/5) + \mathbf{45}(-2/5) + \mathbf{50}^*(2/5).
\end{aligned}
$$

The SM fermions and RH neutrinos are embedded in **16** representation. The SM Higgs doublet (H) is embedded in **10** representation,[a] while the $U(1)_X$ Higgs field (Φ_2) is in **126** representation. Exotic quarks $D + \bar{D}$ in $(\mathbf{5}, 3/5) + (\mathbf{5}^*, -3/5)$ are embedded in $\mathbf{16}^* + \mathbf{16}$ multiplets in the $SO(10)$ GUT, while $Q + \bar{Q}$ in $(\mathbf{10}, 1/5) + (\mathbf{10}^*, -1/5)$ are embedded in $\mathbf{16} + \mathbf{16}^*$ multiplets. Similarly to the embedding of

[a]To be precise, for deriving realistic SM fermion mass matrices, the SM Higgs doublet is identified with a linear combination of $SU(2)_L$ doublets in **10** and **126** representations. See Eq. (15) for the Yukawa coupling in the $SO(10)$ GUT.

$U(1)_Y$ into $SU(5)$, the $SO(10)$ GUT normalization of $U(1)_X$ is given by

$$g_X = \sqrt{\frac{5}{8}} g_\chi, \qquad Q_\chi = \sqrt{\frac{5}{8}} Q_X. \tag{13}$$

For simplicity, we assume the $SO(10)$ symmetry breaking to the $U(1)_X$ extended SM by non-zero vacuum expectation values (VEVs) of $\langle \mathbf{1}(0) \rangle$ and $\langle \mathbf{24}(0) \rangle$ in a **45**-representation Higgs field:

$$SO(10) \underset{\langle \mathbf{1}(0) \rangle}{\longrightarrow} SU(5) \times U(1)_X \underset{\langle \mathbf{24}(0) \rangle}{\longrightarrow} SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_X. \tag{14}$$

The final $U(1)_X$ breaking can be realized by a non-zero VEV of $\Phi_2^\dagger = \mathbf{1}(2) \subset \mathbf{126}$ Higgs field. The Yukawa interactions for the SM fermions are given by

$$\mathcal{L}_{\text{Yukawa}} \supset Y_{10} \mathbf{16}_f \mathbf{16}_f \mathbf{10}_H + Y_{126} \mathbf{16}_f \mathbf{16}_f \mathbf{126}_H^\dagger, \tag{15}$$

where $\mathbf{16}_f$ is a fermion multiplet (the generation index is suppressed), and $\mathbf{10}_H$ and $\mathbf{126}_H$ are Higgs fields. Referring the above decomposition, one can see that the VEV of $\mathbf{1}(2) \subset \mathbf{126}$ Higgs breaks the $U(1)_X$ symmetry and generates Majorana masses of RH neutrinos in $\mathbf{16}_f$ through the Yukawa coupling Y_{126} in Eq. (15).

3.4. *Gauge coupling unification to SO(10)*

After the successful unification of the SM gauge groups to $SU(5)$ at $M_{\text{SU(5)}}$ we consider the unification of $SU(5) \times U(1)_\chi \to SO(10)$ at $M_{\text{SO(10)}}$. The simplest case is $M_{\text{SU(5)}} = M_{\text{SO(10)}}$, then the running coupling $g_\chi(\mu)$ is determined so as to satisfy the unification condition $g_\chi(M_{\text{SU(5)}}) = g_i(M_{\text{SU(5)}})$ $(i = 1, 2, 3)$. The result is shown in the left-handed panel of Fig. 3. For a case of $M_{\text{SU(5)}} < M_{\text{SO(10)}}$, we need to solve the RG evolution of g_5 and g_χ in the energy range of $M_{\text{SU(5)}} < \mu < M_{\text{SO(10)}}$. We omit discussion of those cases here, however, we have shown that the unification of $SU(5) \times U(1)_\chi$ into $SO(10)$ is possible.

3.5. *The GW spectrum*

First, we show the dependence of the GWs spectrum on the energy scale of symmetry breaking, or equivalently, the VEV (v_2) scale. In Fig. 4 we show the GWs spectrum for different symmetry breaking scales for a fixed value of $\lambda_2 = 6 \times 10^{-4}$. As expected, the peak frequency becomes higher, as the symmetry breaking occurs at higher energies. The black solid curves denote the expected sensitivities of each indicated experiment: LISA, DECIGO, BBO, ET, and Cosmic Explore (CE). The expected sensitivity curves for each experiment were based on one of the latest results.[59]

Next, we study how the $U(1)_X$ Higgs quartic coupling and Yukawa coupling affect the resultant GWs spectrum. As a Yukawa coupling increases, the peak amplitude decreases with the peak frequency increasing. This is because fermion loops generate only thermal mass term and not an effective trilinear term in the thermal

potential (5), which weakens the first order phase transition. Assuming the hierarchy among the Yukawa couplings as $Y_{N_3} \equiv Y_N \gg Y_{N_2}, Y_{N_1}$, for simplicity, we show in Fig. 5 the GWs spectrum for various values of Y_N and λ_2 for $g_\chi = 0.463$ and $v_2 = 1$ PeV.

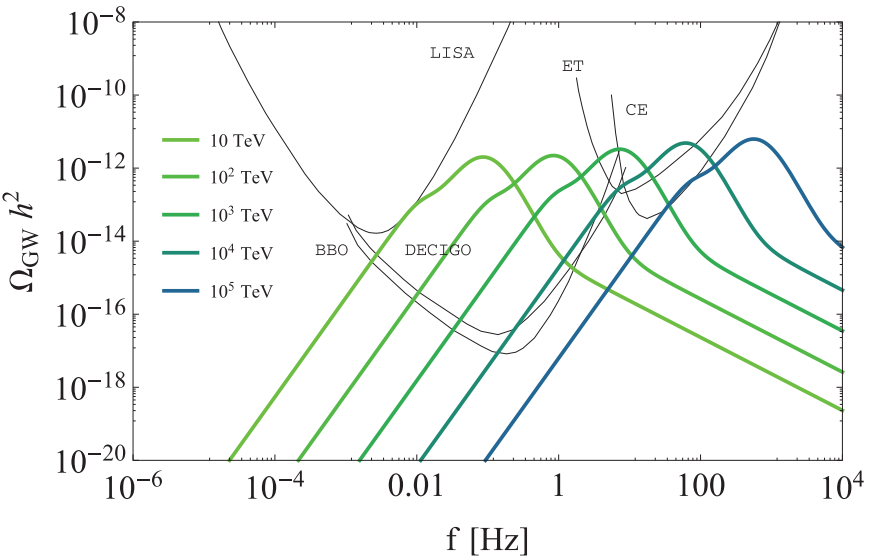


Fig. 4. The predicted GWs spectrum for various symmetry breaking scales for $\lambda_2 = 6 \times 10^{-4}$. The difference of the symmetry breaking scale is indicated by colors as shown in the legends. The expected sensitivities of each indicated experiments[59] are shown by black solid curves.

4. Summary

In this paper, we have considered the $U(1)_X$ extended SM and studied the spectrum of stochastic GWs generated by the first-order phase transition associated with the extra $U(1)_X$ symmetry breaking in the early Universe. This breaking is responsible for the generation of Majorana masses of RH neutrinos and may be surveyed by the GWs spectrum.

We also have shown a UV completion of the $U(1)_X$ extended SM by an $SO(10)$ GUT. We have found that the first-order phase transition triggered by this extra $U(1)$ symmetry breaking can be strong enough to generate GWs with a detectable size of amplitude if the $U(1)_X$ Higgs quartic coupling is small enough and the symmetry breaking scale (the bubble nucleation temperature $T_\star$) is smaller than about 10^5 (10^4) TeV.

We have also clarified the dependence of the resultant GWs spectrum on the RH neutrino Majorana Yukawa couplings, in other words, the mass scale of RH neutrinos. It can be seen as the reduction of the amplitude and the shift of peak

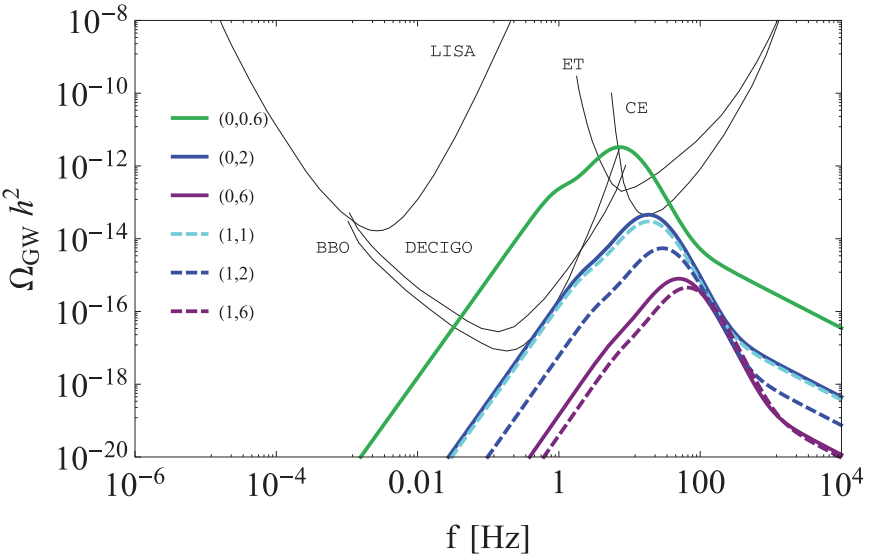


Fig. 5. The predicted GWs spectrum for various values of Y_N and λ_2 for $g_\chi = 0.463$ and $v_2 = 1$ PeV. Parameters in the legend denote $(Y_N, \lambda_2 \times 10^3)$.

frequency of GWs background, which is very similar to the effect of the variation of the $U(1)_X$ Higgs quartic coupling nevertheless.

Acknowledgments

We thank Taiki Hasegawa and Hikaru Uchida for collaboration which results were in part based on. This work is supported in part by the U.S. DOE Grant No. DE-SC0012447 (N.O.), the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants No. 19K03860, No. 19H05091, No. 19K03865, and No. 21H00060. (O.S.).

References

1. A. A. Starobinsky, JETP Lett. **30**, 682 (1979) [Pisma Zh. Eksp. Teor. Fiz. **30**, 719 (1979)].
2. E. Witten, Phys. Rev. D **30**, 272 (1984).
3. C. J. Hogan, Mon. Not. Roy. Astron. Soc. **218**, 629 (1986).
4. C. Caprini and D. G. Figueroa, Class. Quant. Grav. **35**, 163001 (2018).
5. A. Mazumdar and G. White, Rept. Prog. Phys. **82**, no. 7, 076901 (2019).
6. C. Caprini, M. Chala, G. C. Dorsch, M. Hindmarsh, S. J. Huber, T. Konstandin, J. Kozaczuk, G. Nardini, J. M. No, K. Rummukainen, P. Schwaller, G. Servant, A. Tranberg and D. J. Weir, JCAP **03**, 024 (2020).
7. M. S. Turner and F. Wilczek, Phys. Rev. Lett. **65**, 3080 (1990).
8. A. Kosowsky, M. S. Turner and R. Watkins, Phys. Rev. D **45**, 4514 (1992).
9. A. Kosowsky, M. S. Turner and R. Watkins, Phys. Rev. Lett. **69**, 2026 (1992).

10. M. S. Turner, E. J. Weinberg and L. M. Widrow, Phys. Rev. D **46**, 2384 (1992).
11. A. Kosowsky and M. S. Turner, Phys. Rev. D **47**, 4372 (1993).
12. M. Kamionkowski, A. Kosowsky and M. S. Turner, Phys. Rev. D **49**, 2837 (1994).
13. A. Kosowsky, A. Mack and T. Kahniashvili, Phys. Rev. D **66**, 024030 (2002).
14. A. D. Dolgov, D. Grasso and A. Nicolis, Phys. Rev. D **66**, 103505 (2002).
15. G. Gogoberidze, T. Kahniashvili and A. Kosowsky, Phys. Rev. D **76**, 083002 (2007).
16. C. Caprini, R. Durrer and G. Servant, JCAP **0912**, 024 (2009).
17. M. Hindmarsh, S. J. Huber, K. Rummukainen and D. J. Weir, Phys. Rev. Lett. **112**, 041301 (2014).
18. M. Hindmarsh, S. J. Huber, K. Rummukainen and D. J. Weir, Phys. Rev. D **92**, 123009 (2015).
19. M. Hindmarsh, Phys. Rev. Lett. **120**, 071301 (2018).
20. G. M. Harry, P. Fritschel, D. A. Shaddock, W. Folkner and E. S. Phinney, Class. Quant. Grav. **23**, 4887 (2006); Erratum: [Class. Quant. Grav. **23**, 7361 (2006)].
21. N. Seto, S. Kawamura and T. Nakamura, Phys. Rev. Lett. **87**, 221103 (2001).
22. G. M. Harry [LIGO Scientific Collaboration], Class. Quant. Grav. **27**, 084006 (2010).
23. M. Punturo *et al.*, Class. Quant. Grav. **27**, 194002 (2010).
24. C. Grojean and G. Servant, Phys. Rev. D **75**, 043507 (2007).
25. P. S. B. Dev and A. Mazumdar, Phys. Rev. D **93**, 104001 (2016).
26. C. Balazs, A. Fowlie, A. Mazumdar and G. White, Phys. Rev. D **95**, 043505 (2017).
27. P. Minkowski, Phys. Lett. **67B**, 421 (1977).
28. T. Yanagida, Conf. Proc. C **7902131**, 95 (1979).
29. M. Gell-Mann, P. Ramond, and R. Slansky, Conf. Proc. C **790927**, 315 (1979).
30. R. N. Mohapatra and G. Senjanovic, Phys. Rev. Lett. **44**, 912 (1980).
31. R. Jinno, K. Nakayama and M. Takimoto, Phys. Rev. D **93**, 045024 (2016).
32. R. Jinno and M. Takimoto, Phys. Rev. D **95**, 015020 (2017).
33. W. Chao, W. F. Cui, H. K. Guo and J. Shu, Chin. Phys. C **44**, no. 12, 123102 (2020).
34. K. Hashino, M. Kakizaki, S. Kanemura, P. Ko and T. Matsui, JHEP **1806**, 088 (2018).
35. N. Okada and O. Seto, Phys. Rev. D **98**, 063532 (2018).
36. K. Hashino, R. Jinno, M. Kakizaki, S. Kanemura, T. Takahashi and M. Takimoto, Phys. Rev. D **99**, no. 7, 075011 (2019).
37. V. Brdar, A. J. Helmboldt and J. Kubo, JCAP **1902**, 021 (2019).
38. C. Marzo, L. Marzola and V. Vaskonen, Eur. Phys. J. C **79**, no. 7, 601 (2019).
39. T. Hasegawa, N. Okada and O. Seto, Phys. Rev. D **99**, no. 9, 095039 (2019).
40. N. Haba and T. Yamada, Phys. Rev. D **101**, no. 7, 075027 (2020).
41. N. Okada, O. Seto and H. Uchida, PTEP **2021**, no. 3, 033B01 (2021).
42. J. C. Pati and A. Salam, Phys. Rev. D **8**, 1240-1251 (1973).
43. A. Davidson, Phys. Rev. D **20**, 776 (1979).
44. R. N. Mohapatra and R. E. Marshak, Phys. Rev. Lett. **44**, 1316 (1980) [Erratum-ibid. **44**, 1643 (1980)].
45. R. E. Marshak and R. N. Mohapatra, Phys. Lett. B **91**, 222 (1980).
46. T. Fukuyama, Int. J. Mod. Phys. A **28**, 1330008 (2013).
47. M. Carena, A. Daleo, B. A. Dobrescu and T. M. P. Tait, Phys. Rev. D **70**, 093009 (2004).
48. J. Heeck, Phys. Lett. B **739**, 256 (2014).
49. N. Okada and S. Okada, Phys. Rev. D **93**, 075003 (2016).
50. N. Okada and S. Okada, Phys. Rev. D **95**, 035025 (2017).
51. N. Okada and O. Seto, Mod. Phys. Lett. A **33**, 1850157 (2018).
52. S. Okada, Adv. High Energy Phys. **2018**, 5340935 (2018).
53. M. L. Bellac, *Thermal Field Theory*, Cambridge University Press (1996).

54. J. I. Kapusta and C. Gale, *Finite-Temperature Field Theory: Principles and Applications*, Cambridge University Press (2011).
55. C. L. Wainwright, Comput. Phys. Commun. **183**, 2006 (2012).
56. C. L. Wainwright, S. Profumo and M. J. Ramsey-Musolf, Phys. Rev. D **86**, 083537 (2012).
57. C. W. Chiang and E. Senaha, Phys. Lett. B **774**, 489 (2017).
58. S. J. Huber and T. Konstandin, JCAP **0809**, 022 (2008).
59. K. Schmitz, JHEP **01**, 097 (2021).
60. J. Ellis, M. Lewicki and J. M. No, JCAP **1904**, 003 (2019).
61. J. Ellis, M. Lewicki, J. M. No and V. Vaskonen, JCAP **1906**, 024 (2019).
62. D. Cutting, M. Hindmarsh and D. J. Weir, Phys. Rev. Lett. **125**, no. 2, 021302 (2020).
63. M. Hindmarsh and M. Hijazi, JCAP **1912**, no. 12, 062 (2019).
64. J. Ellis, M. Lewicki and J. M. No, JCAP **07**, 050 (2020)
65. T. Appelquist, B. A. Dobrescu and A. R. Hopper, Phys. Rev. D **68** 035012 (2003).
66. S. Oda, N. Okada and D. s. Takahashi, Phys. Rev. D **92**, no. 1, 015026 (2015).
67. N. Okada, S. Okada and D. Raut, Phys. Rev. D **95** no. 5, 055030 (2017).
68. S. Oda, N. Okada, D. Raut and D.-s. Takahashi, Phys. Rev. D **97** no. 5, 055001 (2018).
69. N. Okada, S. Okada and Q. Shafi, Phys. Lett. B **810**, 135845 (2020).
70. H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
71. N. Okada, S. Okada and D. Raut, Phys. Lett. B **780**, 422 (2018).
72. U. Amaldi, W. de Boer, P. H. Frampton, H. Furstenau and J. T. Liu, Phys. Lett. B **281**, 374 (1992).
73. J. L. Chkareuli, I. G. Gogoladze and A. B. Kobakhidze, Phys. Lett. B **340**, 63 (1994).
74. J. L. Chkareuli, I. G. Gogoladze and A. B. Kobakhidze, Phys. Lett. B **376**, 111 (1996).
75. D. Choudhury, T. M. P. Tait and C. E. M. Wagner, Phys. Rev. D **65**, 053002 (2002).
76. D. E. Morrissey and C. E. M. Wagner, Phys. Rev. D **69**, 053001 (2004).
77. I. Gogoladze, B. He and Q. Shafi, Phys. Lett. B **690**, 495 (2010).
78. H. Y. Chen, I. Gogoladze, S. Hu, T. Li and L. Wu, Eur. Phys. J. C **78**, no. 1, 26 (2018).
79. K. Abe *et al.* [Super-Kamiokande Collaboration], Phys. Rev. D **95**, no. 1, 012004 (2017).
80. A. M. Sirunyan *et al.* [CMS Collaboration], JHEP **1808**, 177 (2018).
81. A. M. Sirunyan *et al.* [CMS Collaboration], Phys. Rev. D **100**, no. 7, 072001 (2019).
82. R. Slansky, Phys. Rept. **79**, 1 (1981).

Efficiency of registration of chirp bursts and signals of collapsing stars by the Euro-Asian network of GW interferometers

V. N. Rudenko[1,2,3], S. L. Andrusenko[3], D. P. Krichevskiy[1,3], G. D. Manucharyan[1,3]

[1] *Sternberg Astronomical Institute, Lomonosov State University, Universitetskii prospect 13, Moscow, 119234, Russia*

[2] *Faculty of Physics, Lomonosov Moscow State University, Leninskie Gory, Moscow 119991 Russia*

[3] *Bauman Moscow State Technical University, Physics Department, 2-nd Baumanskaya street 5, building 1, Moscow, 105005, Russia*

In this report, we evaluate the performance of the proposed Euro-Asian network (EAN) of gravitational wave (GW) interferometers, which is planned taking into account the location of one of the detectors at the RAS Novosibirsk Scientific Center. EAN is formed by four detectors: VIRGO, KAGRA, LIGO India and Novosibirsk. The efficiency of this configuration is calculated based on typical numerical criteria for wide area networks.[1] One of the key criteria is the accuracy of reconstructing the parameters of GW bursts, which links the calculation of this criterion to a specific class of astrophysical sources. We present results for calculations performed for the chirp signals from the of relativistic binaries inspiral and for signals from rotating collapsing stars. Maximizing the integral criterion we find the optimal orientation of the Novosibirsk detector, which is specified by the angle between the south direction and the bisector of the Michelson arms of the GW interferometer.

Keywords: Gravitational waves; Network of detectors; Binary merger; gravitational collapse

1. Introduction

In September 2015, the first direct registration of a gravitational-wave burst from the merger of a relativistic binary, whose components were evaluated as black holes (BH), took place. The detection of this event was carried out using LIGO detectors.[2] After there were several other similar registrations. A qualitative step was the registration by three detectors (including a similar interferometer VIRGO in Europe) GW170814 burst from the merger of BH binary ($M = 30M_{\odot}$) from the distance of 540 Mps,[3] which allowed to reduce the localization zone of the source on the celestial sphere by an order of magnitude, up to $\sim 60\ deg^2$. A gravitational wave (GW) signal from neutron stars (NS) merger was registered, coinciding with GRB170817A gamma burst (with 1.7 s delay).[4] One of the most recent significant steps is the registration of gravitational waves from the neutron star – black hole binaries coalescence.[5] All these facts allow to claim confidently real occurrence of a new gravitational-wave channel of astrophysical information and heuristic value of multi-messenger astronomy, i.e. strategy of parallel observation of transients on detectors of different physical nature. However, there is still no observation of gravitational waves from supernova as well as significant coincidence between

LIGO/VIGRO data and neutrino detectors like NOvA[6] and IceCube[7] experiments. Gravitational waves from core-collapse supernovae can serve as an important source of information about the processes occurring during a given event.[8]

In this context, we continue the discussion of a European-Asian network (EAN)[9] which consists of four antennas in the northern hemisphere: VIRGO in Italy, KAGRA in Japan, LIGO-India in India and the planned new additional antenna in Novosibirsk. In order to assess the scientific feasibility and efficiency of such a network the calculation of its main characteristics was performed in the approach developed in Refs. 1, 10.

Here we consider two sources of gravitational wave signals: inspiral of a relativistic binary and collapsing stars at the protoneutron star stage which is subjected to rotational instabilities. For both sources we define the optimal orientation angle of possible detector in Novosibirsk by maximizing the integral effiency criteria.

Table 1 shows the coordinates of the detectors in question. The detector orientation angle γ is defined as the angle between the southward direction at the detector location and the bisector of the angle formed by its arms, measured counterclockwise.

Table 1. Detector data; all angles given in degrees.

Detector	Latitude l	Longitude L	Orientation γ
VIRGO	43.6	-10.5	206.5
KAGRA	36.4	-137.3	163.3
LIGO India	19.6	-77.0	254.0
Novosibirsk	55.0	-82.9	to be defined

It's worth noting that in addition to the planning Novosibirsk interferometer, today a project of search for neutrino and gravitational correlations using the OGRAN gravitational detector and the BUST neutrino telescope is being developed in Russia.[11]

2. Criteria of a network

To estimate efficiency of a network of ground based detectors it's necessary to construct power patterns of individual components and the whole network. Here we review the basic information necessary for construction of the pattern and calculation of the criteria.

In the long wavelength approximation (the GW wavelength is much larger than the interferometer arm length L) the detector response can be evaluated as

$$h(t) = \frac{\delta L}{L} = F_+(\theta, \varphi, \psi)h_+(t) + F_\times(\theta, \varphi, \psi)h_\times(t), \tag{1}$$

where $F_+(\theta, \varphi, \psi)$, $F_\times(\theta, \varphi, \psi)$ are the antenna pattern functions for the two polarizations, which are functions of the polar angle θ and the azimuth angle φ of the

spherical coordinate system (XY is the detector plane) and the polarization angle of the GW ψ.

Antenna pattern functions have the following form (in the coordinate frame, which basis vectors, coincide with the direction of the detector arms):

$$F_+ = \frac{1}{2}(1+\cos^2\theta)\cos 2\varphi\cos 2\psi - \cos\theta\sin 2\varphi\sin 2\psi, \tag{2}$$

$$F_\times = \frac{1}{2}(1+\cos^2\theta)\cos 2\varphi\sin 2\psi + \cos\theta\sin 2\varphi\cos 2\psi. \tag{3}$$

In Ref. 12 it's shown that for a network of N detectors network antenna power pattern P^N:

$$P^N = \sum_{k=1}^{4}(F_{+,k}^2 + F_{\times,k}^2). \tag{4}$$

In order to choose the optimal detector angle in Novosibirsk, we use 3 independent criteria presented in Refs. 1, 10. These three criteria form an integral criterion, which is to be maximized by changing the orientation of the Novosibirsk detector, to find the most effective angle.

2.1. *Polarization criterion I*

Criterion I characterizes ability of the network to assess the polarization of the received GW. Following Ref. 1 we define + and × integral functions for a network of four detectors:

$$F^N = \frac{1}{2}\sqrt{F_1^2 + F_2^2 + F_3^2 + F_4^2}, \tag{5}$$

where N - stands for a network function and $F_1 ... F_4$ all either correspond to the + or × polarization. Obviously F^N depend on the polarization angle of ψ.

Calculation of I is carried out in dominant polarization frame (DPF).[13] In DPF for each point on the celestial sphere $(\alpha;\delta)$ (in equatorial coordinate system point is defined by right ascension $\alpha \in [-\pi;\pi]$ and declination $\delta \in [-\frac{\pi}{2};\frac{\pi}{2}]$) a polarization angle that maximizes the network factor F_+^N and minimizes $F_\times^N$ is chosen. Consequently, for this direction $(\alpha;\delta)$ the condition $F_+^N \geq F_\times^N$ is valid. The condition of approximate equality of factors $F_\times^N$ and F_+^N has to be kept, i.e. $\frac{F_\times^N}{F_+^N} \approx 1$. This means that the gravity detector network will be sensitive to both gravity wave polarizations. It follows that a minimum difference of $\left|F_\times^N - F_+^N\right|$ should be sought for all $(\alpha;\delta)$. This leads to the quantitative formulation of the polarization criterion I[1]:

$$I = \left(\frac{1}{4\pi}\oiint \left|F_+^N(\alpha;\delta) - F_\times^N(\alpha;\delta)\right|^2 d\Omega\right)^{-1/2}, \tag{6}$$

where averaging of $\left|F_\times^N - F_+^N\right|$ over celestial sphere takes place ($d\Omega$ - solid angle). The bigger is I the smaller is the averaged difference $\left|F_\times^N - F_+^N\right|$.

2.2. *Localization criterion D*

Criterion D characterizes the ability of a network to define angular position of a source. In astrometry the problem of a source localization on celestial sphere of a radiation source is solved by a method of triangulation. Triangulation is based on the difference in time between the registration of signals by network detectors. The further apart the detectors are, the greater the time delay is. To maximize the source location accuracy on the celestial sphere, the telescopes should be placed as far apart from each other as possible. According to Ref.,[1] for a network of four detectors D is calculated as the area of the triangle formed by the three detectors in the network, which has the largest area among all possible combinations. If O - the center of the Earth, A, B, C - points where the detectors are located, the area of the corresponding triangle:

$$S_{ABC} = 1/2|[\overrightarrow{AC}, \overrightarrow{AB}]| = 1/2|[\overrightarrow{OC} - \overrightarrow{OA}; \overrightarrow{OB} - \overrightarrow{OA}]|. \tag{7}$$

2.3. *Parameters reconstruction criterion R*

Criterion R characterizes the possibility of reconstruction the parameters of the signal of a known analytical form. According to the Maximum likelihood estimation in the additive Gaussian noise background model, the parameters of the received signal are evaluated by the Rao-Cramer bound. The best possible estimates are obtained using the Fisher information matrix $\Gamma_{\alpha\beta}$[14] in accordance with the formula

$$\Gamma_{\alpha\beta} = Re\left\{4\int_{f_{min}}^{f_{max}} \frac{\overline{\partial_\alpha \tilde{h}(f)}\partial_\beta \tilde{h}(f)}{S_n(f)} df\right\}, \tag{8}$$

where $\tilde{h}(f)$ is the Fourier image of response of the detector, the line above the Fourier image of response represents the complex conjugate, and $S_n(f)$ is the spectral noise density of a single detector. In this paper we assume for simplicity that all detectors have the same noise properties presented in.[15]

Rao-Cramer bound determines the best possible accuracy of parameter P estimation[14]:

$$\delta P^2 = \left(\Gamma_N^{-1}\right)_{PP}, \tag{9}$$

where $\Gamma_N = \sum_{i=1}^{N} \Gamma_i$, i.e. the Fisher information matrix for detector network, is the sum of the corresponding detector matrices constituting the network. The inverse value of the celestial-averaged relative error is a numerical expression of criterion R:

$$R = \left(\frac{1}{4\pi}\oiint \left(\frac{\delta P}{P}\right)^2 d\Omega\right)^{-1/2} = \left\langle \frac{\delta P}{P} \right\rangle^{-1}. \tag{10}$$

Maximization of criterion R leads to the minimum relative error averaged over the celestial sphere in the estimation of the parameter.

2.4. *Integral criterion C*

Integral criterion C is used to compare different configurations of a network:

$$C = \sqrt{\left(\frac{I}{I_{max}}\right)^2 + \left(\frac{D}{D_{max}}\right)^2 + \left(\frac{R}{R_{max}}\right)^2}. \tag{11}$$

Criteria I, D and R together define a three-dimensional space that can be used to define a point which describes a particular configuration. Maximization of C by orientation angle of the detector in Novosibirsk gives the optimal orientation angle and leads to values $I(\gamma_{Nsk}^{max}) = I_{max}, D(\gamma_{Nsk}^{max}) = D_{max}$ and $R(\gamma_{Nsk}^{max}) = R_{max}$ (from the definition it's clear $C_{max} = 1$).

3. Sources

3.1. *Binary inspiral*

As a first typical source of gravitational radiation we consider inspiral of a relativistic binary. For our purposes a rather simplified Newtonian form of gravitational chirp signal, which does not take into account post-Newtonian corrections is sufficient[16]:

$$h(t) = \frac{4}{d_L}\sqrt{F_+^2 + F_\times^2}\frac{G^{5/3}}{c^4}\mathcal{M}^{5/3}(\pi f)^{2/3} cos\left(\Phi + \Psi\right), \tag{12}$$

$$\mathcal{M} = \mu^{3/5} M^{2/5}, \tag{13}$$

$$\Phi = -2\left(\frac{c^3(T-t)}{5GM}\right)^{5/8}, \tag{14}$$

$$f = \frac{1}{\pi M}\left[\frac{5}{256}\frac{c^5 M}{G(T-t)}\right]^{3/8}, \tag{15}$$

where $\mathcal{M}$ is the chirp mass of a system, d_L is the distance to the source, Φ is the phase, f is the frequency, Ψ is the initial phase, M and μ is the total and reduced masses of the binary system, T is the moment of coalescence.

3.2. *Collapsing star*

As a second source of gravitational radiation we consider a core-collapse supernovae. During the core-collapse there exist many mechanisms of gravitational waves radiation on different stages of the process.[17] As Fisher matrix approach (8) requires analytical form of the signal we have considered gravitational waves from long-lived rotational instabilities of a proto-neutron star. If the key result of this instabilities is bar deformation that the radiation can be simulated by radiation from a rotating cylinder (axis of rotation is a bisector of the cylinder axis) with a Gaussian exponent which is introduced phenomenologically to take into account finiteness of

the signal:

$$h(t) = \sqrt{F_+^2 + F_\times^2}\frac{GM\omega^2 L^2}{3c^4 r}(1-3\varepsilon^2)e^{-\frac{4(t-t_0)^2}{t_0^2}}cos(2\omega t+\phi), \tag{16}$$

$$\phi = \pi + \tan^{-1}\frac{F_+}{F_\times}, \tag{17}$$

where G is Newtonian constant of gravitation, c – speed of light, M – mass of the source, L – length of the cylinder, R – radius of the cylinder, ω – source angular frequency, t_0 – characteristic signal length and parameter $\varepsilon = \frac{R}{L}$ characterizes the degree of deformation and is used for calculation of criterion R (10). Gravitational radiation carries away energy from the system so ω decreases over time. However for typical values of source parameters[17] $\omega = 2\pi \times 700$ rad/s, $L = 20$ km, $R = 5$ km, $M = M_\odot$ and $t_0 = 1$ s calculation of gravitational waves luminosity via Einstein's formula[18]:

$$L_{GW} = \frac{G}{5c^5}\left\langle \sum_{j,k=1}^{3} \dddot{Q}_{jk}\left(t-\frac{r}{c}\right)\dddot{Q}_{jk}\left(t-\frac{r}{c}\right)\right\rangle, \tag{18}$$

where Q_{jk} – reduced quadrupole moment and dot denotes time derivative, leads to $\frac{\Delta\omega}{\omega} \sim 10^{-7}$, so we consider ω to be constant.

4. Numerical results and conclusions

To define the optimal orientation angle of possible Novosibirsk detector for the to sources in question we maximize the integral criterion C (11). We have chosen the following typical values for the first source parameters: a binary neutron star with masses of $1.4M_\odot$, without spins, located at a distance of 1 Gpc from the Earth and with an orbital plane perpendicular to the line-of-sight. The results of numerical integration are presented in Figure 1. Criterion D in our approximation does not depend on the orientation angle of the detector in Novosibirsk. The most sensitive to changes in orientation angle is criterion I. The maximum value of integral criterion C is achieved at $\gamma_{Nsk}^{max} = 13^\circ$.

In the case of gravitational collapse the strain $h(t)$ is expected to be several orders of magnitude less than for binary coalescence ($h \sim 10^{-22} - 10^{-20}$ at 10 kpc) so we can't assume isotropic distribution of sources over the celestial sphere. Instead we integrate over the Milky Way disk (neglecting the fact that the sensitivity may be sufficient for detection of the signal from several nearby galaxies). Due to rotation of the Earth antenna pattern functions F_+ and $F_\times$ depend on time (and consequently $P = P(t)$, $\phi = \phi(t)$) thus we average functions of time over 24 hours which also makes analytical calculation of Fourier image $\tilde{h}(f)$ in (8) possible. The parameters of a source are those ones from Section 3.2. The calculation results are shown in Figure 2. Again I is again the most sensitive criterion to changes of orientation angle. The maximum value of integral criterion C is achieved at $\gamma_{Nsk}^{max} = 40^\circ$.

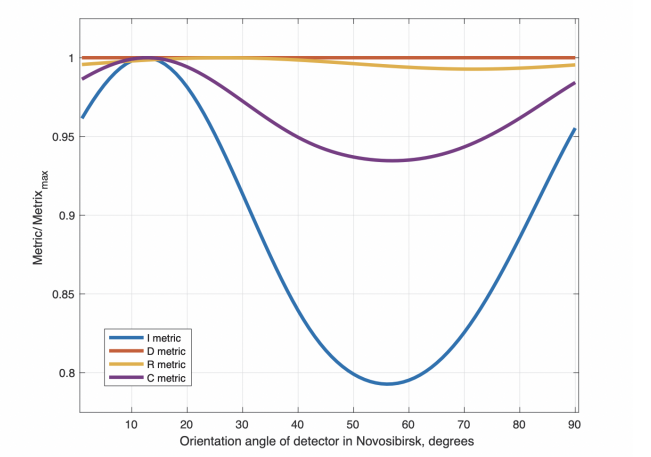


Fig. 1. Dependence of all criteria on orientation angle of detector in Novosibirsk for binary inspiral.

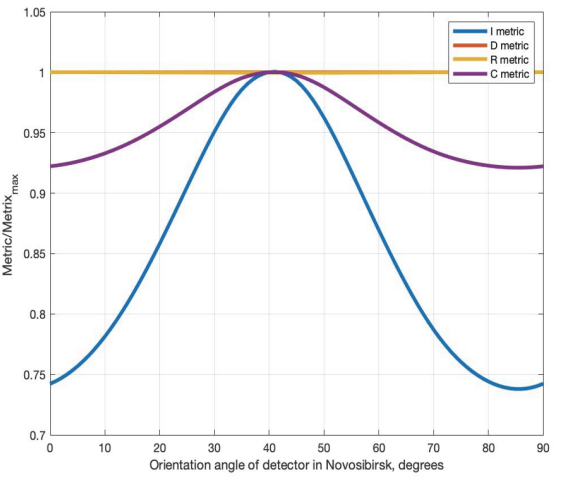


Fig. 2. Dependence of all criteria on orientation angle of detector in Novosibirsk for gravitational collapse.

The final choice of the orientation depends on the most topical problems at the time of construction of the detector in Novosibirsk. To conclude we notice that the choice of the source is a limitation of this work, because more significant physics is encrypted in a more complex structure of the signal from the core collapse, but such signals are model-dependent and do not have an analytical form (e.g.[19]), which does not allow using them within the framework of this approach.

The obtained result shows that in one specific network it is impossible to indicate the orientation of the interferometer in Novosibirsk, which would be optimal both

for GW signals from the coalescence of relativistic binaries and for signals from collapses. This is despite the fact that the collapse model with the stage of rotating cylindrical bar instability is closest to the picture of a merging bynaries at the inspiral stage. At the moment, it is possible to recommend the choice of an averaged orientation with an appropriate estimate of the loss in the accuracy of estimating the parameters of the received signals.

Acknowledgments

The authors are grateful to Prof. A.N.Morozov and Prof. V.O Gladyshev from the Fundamental Physics Faculty of Bauman Moscow Technical University for stimulation of this analysis. This work was supported by the grant RFBR 19-29-11010.

References

1. P. Raffai *et al.*, Optimal networks of future gravitational-wave telescopes, *Class. Quantum Grav.* **30**, p. 155004 (2013).
2. B. P. Abbott *et al.*, Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, p. 061102 (Feb 2016).
3. B. P. Abbott *et al.*, Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence, *Phys. Rev. Lett.* **119**, p. 141101 (Oct 2017).
4. B. P. Abbott *et al.*, Gw170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (Oct 2017).
5. R. Abbott *et al.*, Observation of gravitational waves from two neutron star–black hole coalescences, *ApJL* **915**, p. L5 (jun 2021).
6. M. A. Acero *et al.*, Search for multimessenger signals in nova coincident with ligo/virgo detections, *Phys. Rev. D* **101**, p. 112006 (Jun 2020).
7. D. Veske *et al.*, Multi-messenger searches via IceCube's high-energy neutrinos and gravitational-wave detections of LIGO/Virgo, *PoS* **ICRC2021**, p. 950 (2021).
8. C. L. Fryer and K. C. B. New, Gravitational Waves from Gravitational Collapse, *Living Rev. Relativ.* **14**, p. 1 (2011).
9. V. Rudenko, S. Andrusenko, D. Krichevskiy and G. Manucharyan, Gw interferometer euro-asian network: Detection characteristics for signals of known shape, *Universe* **6**, p. 140 (2020).
10. Y.-M. Hu *et al.*, Global optimization for future gravitational wave detector sites, *Class. Quantum Grav.* **32**, p. 105010 (apr 2015).
11. V. N. Rudenko, Y. M. Gavrilyuk, A. V. Gusev, D. P. Krichevskiy, S. I. Oreshkin, S. M. Popov and I. S. Yudin, Gravitational wave detector ogran as multi-messenger project of ras-msu, *International Journal of Modern Physics A* **35**, p. 2040007 (2020).
12. L. S. Finn, Aperture synthesis for gravitational-wave data analysis: Deterministic sources, *Physical Review D* **63**, p. 102001 (2001).
13. S. Klimenko *et al.*, Localization of gravitational wave sources with networks of advanced detectors, *Phys. Rev. D* **83**, p. 102001 (May 2011).
14. B. Kocsis, Z. Haiman, K. Menou and Z. Frei, Premerger localization of gravitational-wave standard sirens with lisa: Harmonic mode decomposition, *Phys. Rev. D* **76**, p. 022003 (Jul 2007).
15. LIGO, (2009), LIGO Document T0900288-v3.
16. L. S. Finn and D. F. Chernoff, Observing binary inspiral in gravitational radiation: One interferometer, *Phys. Rev. D* **47**, 2198 (Mar 1993).

17. S. E. Gossan *et al.*, Observing gravitational waves from core-collapse supernovae in the advanced detector era, *Phys. Rev. D* **93**, p. 042002 (Feb 2016).
18. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields: Volume 2* (Butterworth-Heinemann, 1980).
19. H. Dimmelmeier *et al.*, Gravitational wave burst signal from core collapse of rotating stars, *Phys. Rev. D* **78**, p. 064056 (Sep 2008).

Joint analysis method on gravitational waves and low-energy neutrinos to detect core-collapse supernovae

O. Halim
Istituto Nazionale di Fisica Nucleare (INFN) sez. di Trieste, Italy
E-mail: odysse.halim@ts.infn.it

C. Casentini
Istituto Nazionale di Astrofisica - Istituto di Astrofisica e Planetologia Spaziali (INAF -IAPS), Rome, Italy

M. Drago
Università di Roma La Sapienza, I-00185 Roma, Italy,
INFN, Sezione di Roma, I-00185 Roma, Italy

V. Fafone
University of Rome Tor Vergata, Rome, Italy,
INFN sez. di Roma Tor Vergata, Rome, Italy

K. Scholberg
Department of Physics, Duke University, Durham, NC, USA

C. F. Vigorito
University of Turin, Italy,
INFN sez. di Torino, Italy

G. Pagliaroli
Gran Sasso Science Institute (GSSI), L'Aquila, Italy,
INFN sez. di LNGS, Assergi, Italy

Core-collapse supernovae produce copious low-energy neutrinos and are also predicted to radiate gravitational waves. These two messengers can give us information regarding the explosion mechanism. The gravitational wave detection from these events are still elusive even with the already advanced detectors. Here we give a concise and timely introduction to a new method that combines triggers from GW and neutrino observatories; more details shall be given in a forthcoming paper.[1]

Keywords: Multimessenger, supernova, core-collapse, low-energy neutrino, gravitational wave.

1. Introduction

Core-collapse supernovae (CCSNe) are expected to produce multimessenger signals such as neutrinos, gravitational waves (GWs), as well as multi-wavelength electromagnetic waves.[2,3] Low-energy neutrinos (LENs) are expected to be produced by

these events, with the average energy $\sim$ 10 MeV and these LENs are from the majority of the total energy budget ($\sim 10^{53}$ erg) of the CCSNe.

LENs from a CCSN have been successfully detected with the observation of SN1987A in Large Magellanic Cloud by Kamiokande-II,[4] IMB,[5] and Baksan.[6] Today, there are several LEN detectors waiting for these astrophysical events such as Super-Kamiokande[7] (Super-K), LVD,[8] KamLAND,[9] and IceCube[10] with the horizon up to the edge of our galaxy and beyond. These detectors are in fact in a multi-detector collaboration for this effort to produce low-latency alerts under SuperNova Early Warning System (SNEWS).[11,12]

Moreover, the era of multimessenger astronomy involving GWs has just begun with the detection of the binary neutron star merger.[13] Recently, GW search has been in the phase of O3 data taking, which is done by LIGO[14] detectors (4-km arm in Hanford and Livingston USA), Virgo[15] detector (3-km arm in Cascina Italy), and KAGRA[16] (3-km arm in Gifu Prefecture, Japan in the latest period of O3). In total, the GW science data taking has been done for 3 observing runs (O1, O2, and O3) from 2015 to 2020, and currently the detectors undergo improvements in order to be more sensitive for O4 (expected in August 2022). CCSNe are expected to produce GW signals with broad physical processes.[17–20] Detecting GWs from these sources will enable us to study the physical processes. As we will see later on, the detection capability of GWs could be improved by multimessenger search involving LENs.

Here, we provide a timely description of a strategy to combine GWs and LENs for a multimessenger search following the chart in Fig. 1. More details discussion will be given in a forthcoming paper,[1] which has just been accepted by JCAP. This method is based on our previous works.[21,22] We construct a time-coincident strategy and test it with the simulated signals for both GW and LEN data.

In our strategy, we use the `coherent WaveBurst` (cWB) pipeline[23–26] for GW data analysis from simulations. This is a model-agnostic algorithm pipeline that is used to search GWs from CCSNe.[27] Moreover, we also simulate LEN signals arriving in several neutrino detectors and analyse them[a]. At this stage, we use a new approach (introduced in[22]) for neutrino analysis in order to exploit the temporal behaviour of CCSNe. In the next step, we perform a temporal-coincidence analysis between the two messengers. Different messenger data are analysed separately and then combined together for possible GW-LEN signals. This could be interesting for online networks such as SNEWS or offline search for sub-threshold signals.

This paper is organised as follows. In Sec. 2, we discuss the emission models for each messenger. Then, in Sec. 3, the data and analysis by our strategy will be presented. In the end, we show the implementation of our strategy in Sec. 4.

[a]Note that we employ no detailed neutrino-detector simulation.

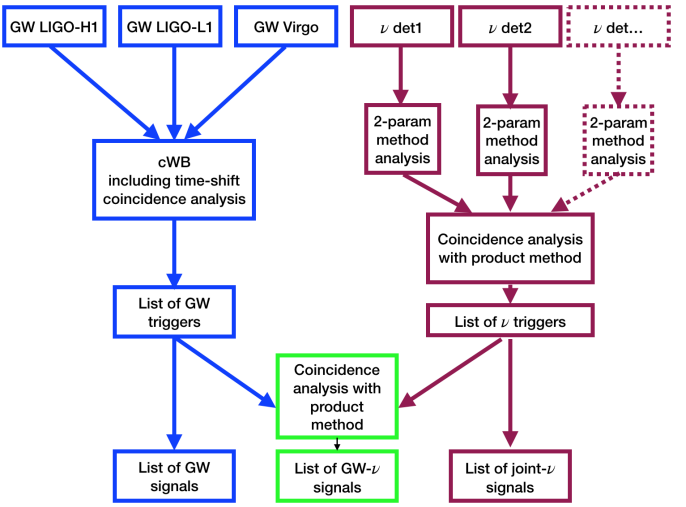


Fig. 1. The schematic view of the multimessenger GW-LEN strategy.

2. Messengers from Core-Collapse Supernovae

A CCSN is expected to produce a $O(10\,\mathrm{ms} - 1\,\mathrm{s})$ GW burst and an impulsive $O(10 - 30\,\mathrm{s})$ emission of $O(10\,\mathrm{MeV})$ LENs[28b], which highly depend on the CCSN initial conditions: progenitor mass, rotation, etc. Thus, it would be ideal to perform a simultaneous GW-LEN analysis from *the same numerical simulation*. Unfortunately, no simultaneous GW-LEN numerical simulations have provided successful CCSN explosion, or if there are, the simulations can only provide both signals for the first half of a second, until the explosion. Meanwhile, the LEN emission duration is expected to last about 10 seconds. Therefore, we combine the currently available GW signals and LEN signals from different simulations with similar progenitor masses, and handle them as if they are from the simultaneous simulation.

2.1. *Gravitational Wave Emission*

We provide the detail of our GW waveforms in Tab. 1. We use the GW signals from the 3D neutrino-radiation hydrodynamics simulations of Radice *et al.*[30] (named as "Rad") with the various zero age main sequence masses in order to have the successful explosions from the low-mass and the failed explosions from the high-mass. Besides, we also take into account models with rapid rotation and high magnetic field which produce much stronger GWs. In this case, we take GW waveforms from two simulations: the Dimmelmeier[31] ("Dim") and the Scheidegger[32] ("Sch"). The stellar progenitors of Dim and Sch have strong rotation and magnetic field. These constraints make the models less favourable to happen than the neutrino-radiation

[b]These signals could also be emitted by "failed" SNe.[29]

mechanism.[27,28,33] Nevertheless, we would not rule out any models since we have not yet detected any CCSN GWs. Thus, we try to cover as broad as possible the uncertainty band from theoretical predictions: the lower limit is from the Rad model, and the upper limit is from the Dim and the Sch models.

Table 1. Waveforms from the considered CCSN simulations. In the columns: emission type and reference, waveform identifier, waveform abbreviation for this manuscript, progenitor mass, angle-averaged root-sum-squared strain $h_{\rm rss}$, frequency when the GW energy spectrum peaks, and emitted GW energy.

Waveform Family	Waveform Identifier	Abbr.	Mass $M_\odot$	$h_{\rm rss}$ @10 kpc $\left[10^{-22}\,\frac{1}{\sqrt{\rm Hz}}\right]$	$f_{\rm peak}$ [Hz]	$E_{\rm GW}$ $[10^{-9}\,M_\odot c^2]$
Radice[30]	s25	Rad25	25	0.141	1132	28
3D simulation;	s13	Rad13	13	0.061	1364	5.9
h_+ & $h_\times$; (Rad)	s9	Rad9	9	0.031	460	0.16
Dimmelmeier[31]	dim1-s15A2O05ls	Dim1	15	1.052	770	7.685
2D simulation;	dim2-s15A2O09ls	Dim2	15	1.803	754	27.880
h_+ only; (Dim)	dim3-s15A3O15ls	Dim3	15	2.690	237	1.380
Scheidegger[32]	sch1-R1E1CA$_L$	Sch1	15	0.129	1155	0.104
3D simulation;	sch2-R3E1AC$_L$	Sch2	15	5.144	466	214
h_+ & $h_\times$; (Sch)	sch3-R4E1FC$_L$	Sch3	15	5.796	698	342

2.2. *Low-energy Neutrino Emission*

Similar to the previous emission, we consider two models for the LEN emissions. First, we take the signals from the numerical simulations of Hüdepohl obtained for a progenitor of $11.2M_\odot$; without the collective oscillations,[34] with the time-dependent neutrino luminosities, and average energies. The simulation provides the first 7.5 seconds of the neutrino emission with the analytical extension in order to reach 10 seconds of the signal. The average LEN energies from before collapse up to the simulated 0.5 s after bounce are (c.f. Table 3.4 of Ref.[34]): $\langle E_{\nu_e}\rangle = 13$ MeV, $\langle E_{\bar\nu_e}\rangle = 15$ MeV and $\langle E_{\nu_x}\rangle = 14.6$ MeV.

Second, we also use a parametric model for neutrino emission from Pagliaroli *et al.*,[35] focusing on the best-fit emission from SN1987A data and the model provides the average energies of $\langle E_{\nu_e}\rangle = 9$ MeV, $\langle E_{\bar\nu_e}\rangle = 12$ MeV and $\langle E_{\nu_x}\rangle = 16$ MeV. This signal has a temporal structure:

$$F(t,\tau_1,\tau_2) = (1-e^{-t/\tau_1})e^{-t/\tau_2}, \tag{1}$$

where τ_1 and τ_2 are parameters governing the emission, representing the rise and decay timescales of the neutrino signal. The best-fit values on these parameters are ~ 0.1 s and ~ 1 s.[36]

Here, we consider only the main interaction channel for water and scintillator detectors: the inverse beta decay (IBD) $\bar\nu_e + p \to n + e^+$. We also consider standard MSW neutrino oscillations on the flux $\Phi_{\bar\nu_e}$ at the detectors. The flux is an admixture

of the unoscillated flavor fluxes at the source, i.e. $\Phi_{\bar{\nu}_e} = P \cdot \Phi_{\bar{\nu}_e} + (1-P)\Phi_{\bar{\nu}_x}$. Here, P describes the survival probability for the $\bar{\nu}_e$. The value of $P \simeq 0$ is for the Inverted Hierarchy (IH), while $P \simeq 0.7$ is for the Normal Hierarchy (NH). The expected number of IBD events can be seen in Tab. 2 with the reference distance of 10 kpc.

Table 2. Average number of IBD events for a CCSN exploding at 10 kpc for the considered models detected by Super-K,[7] LVD,[8] and KamLAND,[9] with the assumed energy thresholds ($E_{\rm thr}$).

Model (identifier)	Progenitor Mass	Super-K ($E_{\rm thr} = 6.5$ MeV)	LVD ($E_{\rm thr} = 7$ MeV)	KamLAND ($E_{\rm thr} = 1$ MeV)
Pagliaroli[35] (SN1987A)	$25\,M_\odot$	4120	224	255
Hüdepohl[34] (Hud)	$11.2\,M_\odot$	2620	142	154

3. Data and Analysis Strategy

Here, we will discuss the data and analysis for GWs, for LENs, and for a combined multimessenger search. We consider a conservative threshold on global false alarm rate (FAR) of 1/1000 years.

3.1. *Gravitational Wave Analysis*

We employ the cWB[37,38c] algorithm for the GW data analysis, a widely used software in the LIGO-Virgo-KAGRA collaborations; which is also used for the study of CCSN targeted-search.[27,39] The cWB software does not need any waveform templates; it combines coherently the excess energy of the data from the involved GW detectors. A maximum likelihood analysis is used to search for the GW candidates and their parameters. The significance is estimated comparing foreground the detection statistics ρ (of the candidates) with a background distribution from the time-shift procedure. We simulate Gaussian noise with a spectral sensitivity based on the expected Advanced LIGO[40] and Advanced Virgo detectors.[41,42] We simulate ~ 16 days of data and ~ 20 years of background livetime.

Waveforms from the models in Sec. 2.1 have been simulated for several distances: 5, 15, 20, 50, 60, 700 kpc. In terms of sky direction, for the shorter distances (5, 15, 20) we use the Galactic model given in,[43] whereas for the longer (50 and 60 kpc) distances we take fixed directions, either towards the Large or Small Magellanic Clouds, and additionally for the 700 kpc we use the Andromeda direction. We focus

[c]cWB home page, `https://gwburst.gitlab.io/`;
public repositories, `https://gitlab.com/gwburst/public`
documentation, `https://gwburst.gitlab.io/documentation/latest/html/index.html`.

mainly on the 5-60 kpc distances for the multimessenger analysis, while the distance between 60 and 700 kpc are used for the efficiency curve[d]. We inject 1 GW signal per 100 seconds, wide enough for two consecutive waveforms. In terms of sky positions, we apply $\sim$ 2500 different realizations over all the sky direction for each distance and model.

We choose a threshold of 864 per day for GW candidates to be used for our multimessenger analysis, to satisfy the requirement of the combined FAR of 1/1000 years. The efficiency curves for each distance can be seen in Fig. 2 as the ratio of the number of recovered injections ($\mathrm{FAR_{GW}} < 864/\mathrm{day}$) to the $\sim$ 2500 total injections.

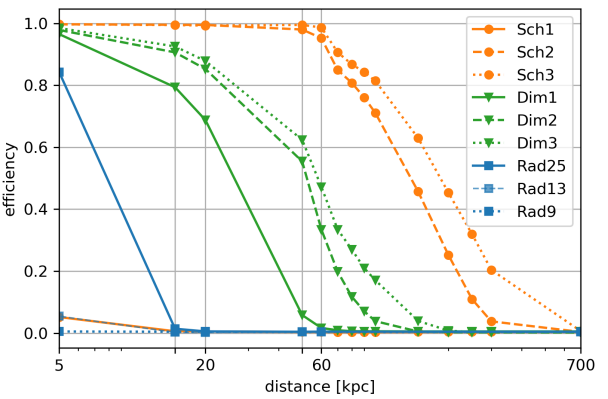


Fig. 2. Efficiency curve of GW search from Advanced LIGO and Advanced Virgo detectors for the various models in Tab. 1, given a FAR threshold of 864/day. Sch1-3 are for the Scheidegger model with different frequency peaks, Dim1-3 are for the Dimmelmeier model with different h_{rss} values, and Rad1-3 are for the Radice model with different progenitor masses

3.2. *Neutrino Analysis: Expanding the Neutrino Detection Horizon*

The standard LEN analysis for CCSNe[44–46] employs a binning in a time series data set with a sliding time window of $w = 20$ seconds. The group of events in a bin is called a *cluster* and the number of events in a cluster is called multiplicity m. We can assume the multiplicity distribution of background events follows a Poisson distribution. For each i-th cluster, we calculate the *imitation frequency* (f^{im}), which is correlated with the significance, defined as,

$$f_i^{\mathrm{im}}(m_i) = N \times \sum_{k=m_i}^{\infty} P(k), \tag{2}$$

[d]For the simulations between 60 and 700 kpc, we use the Andromeda position, though no known astronomical objects are expected around that area.

where $P(k)$ is the Poisson term: the probability in which a cluster of multiplicity k is simply due to the background,

$$P(k) = \frac{(f_{\rm bkg} w)^k e^{-f_{\rm bkg} w}}{k!}, \tag{3}$$

with $N = 8640$ is the total number of bins in one day, calculated from the 10-s overlapping window between two consecutive bins, to tackle the boundary problems. This f^{im} is basically called FAR in the GW community.

From our previous work,[47] we can exploit the temporal behavior of LEN signals with an additional cluster parameter called $\xi_i \equiv \frac{m_i}{\Delta t_i}$, with Δt_i as the i-th cluster duration: the time difference between the last and the first event. Clearly, Δt_i has a maximum value of the bin width itself, which is 20 seconds. A cluster is considered when $m_i \geq 2$, thus $\xi_i \geq 0.1$.

Previous work[47] uses ξ as a cut value. In this work we develop further in order improve the estimation of the imitation frequency. Particularly, it is formulated a new modified **2-parameter** (m and ξ) imitation frequency, called $F^{\rm im}$:

$$F_i^{\rm im}(m_i, \xi_i) = N \times \sum_{k=m_i}^{\infty} P(k, \xi_i), \tag{4}$$

where the term $P(k, \xi_i)$ is the joint probability of a cluster with multiplicity k **and** ξ_i substituting $P(k)$ previously (Eq. 3). Finally, it is convenient to write (see c.f. App. A of[1] and c.f. Sec. 7.1. of[21] for more detail),

$$F_i^{\rm im}(m_i, \xi_i) = N \times \sum_{k=m_i}^{\infty} P(k) \int_{\xi=\xi_i}^{\infty} {\rm PDF}(\xi \geq \xi_i | k) d\xi. \tag{5}$$

Fig. 3 shows the ξ probability density function (PDF) for the Super-K detector (as an example). The black solid line represents the PDF of ξ values due to the background, i.e. PDF($\xi|k$).

We simulate ~ 10 years of background data for each neutrino detector with the assumption that the background frequencies $f_{\rm bkg}$: 0.012 Hz for Super-K,[46] 0.015 Hz for KamLAND,[48] and 0.028 Hz for LVD.[44] Moreover, the CCSN simulated signals (see Sec. 2) are injected into the background data for each detector model (the starting times are simultaneous with the GW injections, with reasonable time delay due to detectors' positions) and for each source distance. These clusters are extracted via the Monte Carlo method and the injection rate is 1 per day.

The clusters are considered signal candidates if their standard $f^{\rm im} \leq 1/{\rm day}$, which has been set in order to reach the global FAR of 1/1000 years. The efficiency for each distance D can be defined as,

$$\eta(D) = \frac{N_{\rm r,s}(D)}{N_{\rm inj,s}(D)}, \tag{6}$$

where $N_{\rm r,s}$ is the number of candidates and $N_{\rm inj,s}$ is the total number of injections. With the recipes up to now, the efficiency curves can be constructed (see Fig. 4).

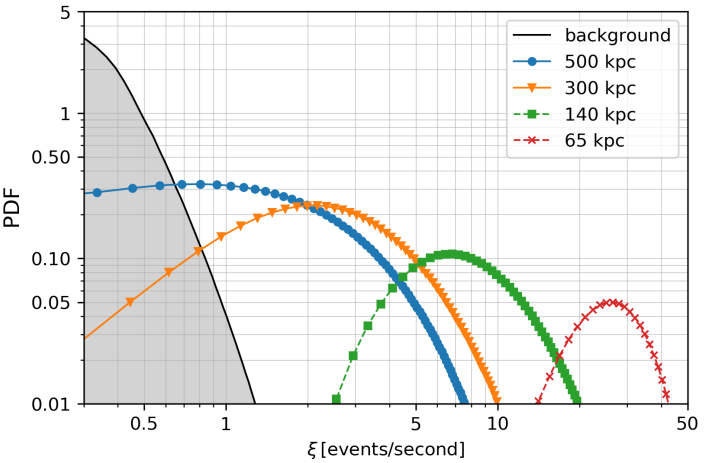


Fig. 3. PDF of ξ values for background (black line) and signal clusters (colored-lines) for different distances in the case of the Super-K detector. Data are from.[47]

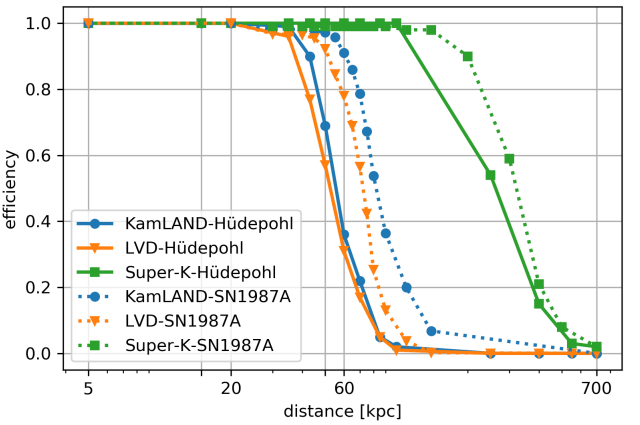


Fig. 4. The efficiency curves of LEN detectors for the Hud (continuous lines) and SN1987A (dashed lines). The imitation frequency threshold is 1/day.

3.3. *Multimessenger Analysis*

The ultimate goal of our work is to construct a multimessenger analysis, combining both LEN and GW data sets (green boxes in Fig. 1). We simply perform a temporal-coincidence analysis between the GW and LEN lists and the statistical significance can be estimated by the combination of its FAR. The joint coincidences are defined as "CCSN candidates".

The $\mathrm{FAR_{GW}}$ is acquired via the time-shifting method (Sec. 3.1) while, the FAR_ν from LEN is obtained by the product method based on SNEWS[11] i.e.,

$$\mathrm{FAR}_\nu = \mathrm{Nd} \times w_\nu^{\mathrm{Nd}-1} \prod_{i=1}^{\mathrm{Nd}} F_i^{\mathrm{im}}, \tag{7}$$

where Nd is the number of LEN detectors, w_ν is the time window of coincidence analysis, and F_i^{im} is the 2-parameter imitation frequency of the clusters.

All in all, the multimessenger $\mathrm{FAR_{glob}}$ from "CCSN candidates" can be written[e],

$$\mathrm{FAR_{glob}} = \mathrm{Net} \times w_c^{\mathrm{Net}-1} \prod_{X=1}^{\mathrm{Net}} \mathrm{FAR}_X, \tag{8}$$

where Net is the number of sub-networks, w_c is the temporal coincidence window between GW and LEN, and FAR_X is the *false-alarm-rate* ($X = \{\nu, \mathrm{GW}\}$). Thus, the *false-alarm-probability* taking into account Poisson statistics can be written,

$$\mathrm{FAP} = 1 - e^{-\mathrm{FAR} \times \mathrm{livetime}}, \tag{9}$$

where "livetime" is the common observing time among the network.

Eq. 8 and 9 are used to compare the performance of our 2-parameter method (Eq. 5) with the standard 1-parameter method (Eq. 2). This performance is discussed as efficiency values, analogous with Eq. 6. For the multimessenger step, we define a "detection" in a network when $\mathrm{FAP} \geq 5\sigma$[f].

4. Results

Here, we discuss the results of our method. First, we mention the single-detector neutrino analysis performance. Next, we show the sub-network of neutrino detectors. Finally, we provide the global network of GW-LEN multimessenger analysis.

4.1. *Improvement of the Single-Detector LEN Analysis*

We work our method on the simulated single-detector data for KamLAND, LVD, and Super-K. To show the improvement, we discuss in the following, as a leading example, the case of the KamLAND detector. We consider a CCSN at 60 kpc with the LEN profile following SN1987A model (the first row of Tab. 2). We perform a 10 years of KamLAND simulated background data and we inject randomly the simulated LENs with the rate of 1/day, thus, 3650 total injections.

All clusters, either due to background or injections, are plotted in Fig. 5 in a ξ vs multiplicity plane. Blue crosses represents the clusters due to injections while the yellow inverted-triangles are due to background. Because of the statistical Poisson fluctuation in the Monte Carlo simulation method, the injection clusters may have

[e] Thorough discussion on the choice of coincidence analysis can be seen in.[21]
[f] $5\sigma \approx 5.7 \times 10^{-7}$

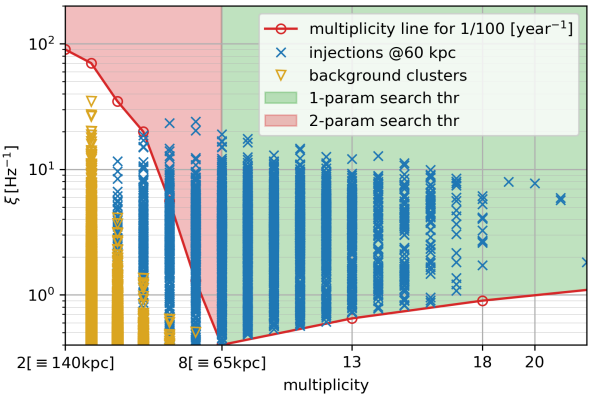


Fig. 5. The ξ-multiplicity map for KamLAND with the simulated background (yellow-triangle) and injection (blue-cross) clusters generated with the SN1987A model at 60 kpc.

various multiplicity values. Besides, background events are also fluctuating in this 20-sec window.

We estimate the associated imitation frequency (or FAR_ν[g]) for each recovered cluster. This imitation frequency in 1-parameter analysis is a function of only the cluster multiplicity via Poisson statistics. To satisfy the SNEWS requirement of $\mathrm{FAR}_\nu \leq 1/100$ year, the multiplicity threshold is 8. This means clusters should lie on the green area of Fig. 5 to satisfy the standard 1-parameter requirement. This limit on the multiplicity could be understood as a KamLAND horizon for CCSN search of $\simeq$ 65 kpc with the model based on SN1987A, in fact it is the average multiplicity for this CCSN configuration: $\langle m_i \rangle = 8$.

On the other hand, the imitation frequency of our 2-parameter method requires each cluster to satisfy the requirement of both multiplicity and the ξ value following Eq. 4. That equation gives us the red line in Fig. 5 for $\mathrm{FAR}_\nu = 1/100$ years[h] required for the 2-parameter method. All clusters above the red line pass the 2-parameter method requirement. This also relax the requirement allowing multiplicity lower than 8 as long as a specific ξ value is satisfied. Thus, the red area in Fig. 5 represents the improvement area of our 2-parameter method comparing with the 1-parameter. In addition, it is clear from the figure that all simulated background clusters (yellow triangles) are well below the red threshold line.

Quantitatively, Fig. 5 shows an increase of the detection efficiency with details given in the first row of Tab. 3. The KamLAND efficiency at 60 kpc is improved from 73% by the standard 1-parameter method to 83% by our 2-parameter method. Moreover, we mention also that there are 75198 noise triggers in this data set and none of them have FAR lower than 1/100 years for both methods. The efficiency

[g]As previously stated in Sec. 3.2, the imitation frequency is actually the FAR. So, FAR_ν is imitation frequency for KamLAND data, either f^{im} or F^{im} depending on the context.
[h]The threshold corresponding to the current SNEWS requirement.

increase could also mean we can expand the detection horizon of the detector since the expected multiplicity is proportional to the inverse of squared-distance[i]. The results for KamLAND are representative for all the scenarios we investigated. Similar improvements can be seen for other scenarios (see c.f. App. C of[1]).

4.2. *The LEN Detector Sub-Network*

We apply our method for the sub-network of LEN detectors: KamLAND and LVD, since their efficiency curves are compatible; see Fig. 4. The threshold for LEN detector network is 5σ in FAP_ν (Eq. 9). We compare the efficiencies in the second and third row of Tab. 3.

Table 3. Efficiency (η) comparison between 1-parameter and 2-parameter method for analysis of KamLAND (with the SN1987A model) and KamLAND-LVD (with the Hud model) and for $\mathrm{FAP}_\nu > 5\sigma$.

Analysis	Model	Distance [kpc] $[> 5\sigma]$	$\eta_{1\mathrm{param}}$ $[> 5\sigma]$	$\eta_{2\mathrm{param}}$
KamLAND	SN1987A	60	**2665/3654=72.9%**	**3026/3654=82.8%**
KamLAND-LVD	Hud	50	**47/108=43.5%**	**59/108=54.6%**
KamLAND-LVD	Hud	60	**19/107=17.8%**	**28/107=26.2%**

At a distance of 50 kpc with a $\mathrm{FAR}_\nu \leq 1/100$ years, we can detect 12% and 26% CCSNe for LVD and KamLAND, respectively. Meanwhile, if the detectors are in a network looking for coincidences within w_ν, it is possible to recover $\sim 43\%$ with 5σ threshold. When we start using 2-parameter method taking into account the ξ distribution for each detector, this efficiency grows to $\sim 55\%$.

Analogously, for CCSNe at 60 kpc, the signals with $\mathrm{FAR}_\nu \leq 1/100$ years can be detected only 3% and 7% for LVD and KamLAND. Whereas, as a network this efficiency increases to 18% and to 26% for 1-parameter and 2-parameter method, respectively for 5σ threshold.

In order to quantify the result, we will mention the results for the SN1987A model with a CCSN at 60 kpc recovered by the LVD-KamLAND network. The improvement in efficiency is from 85% to 93%. These results can be compared with first row of Tab. 3 for the single detector analysis.

4.3. *The Multimessenger Network*

In the multimessenger network, the temporal coincidence window is $w_c = 10$ seconds between GW and LEN lists. We use the 5σ threshold to claim a multimessenger

[i]The number of events is basically a flux.

"detection" of injected signals. We focus on emphasizing the improvement of combining GWs and LENs, which means when the detection efficiency of both LEN and GW detectors is less than 100%. In other words, it is to combine detectors with comparable detection efficiencies. The horizon of the GW network is highly influenced by the assumed GW emission model, see Fig. 2. In this case, the Dim2 model (see Tab. 1) has the GW detection horizon comparable with the LVD and KamLAND detectors, which is more or less the Large Magellanic Cloud distance. Thus, we perform the method for the global network of LIGO-Virgo, LVD, and KamLAND. Results for other GW models can be seen in c.f. App. B of.[1]

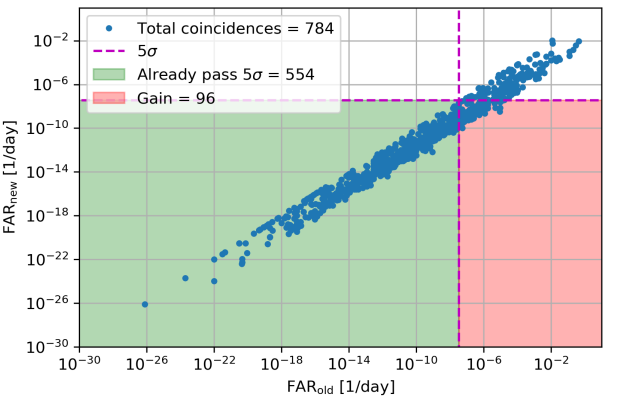


Fig. 6. The joint FAR of GW-LEN candidates obtained with the 2-parameter method ($\mathrm{FAR_{new}}$ y-axis) vs the 1-parameter ($\mathrm{FAR_{old}}$ x-axis) with the **KamLAND** (SN1987A-model) and HLV (Dim2-model) at 60 kpc.

For the coincidence analysis of KamLAND detector with the HLV (Hanford, Livingston, Virgo) GW network, we can see the FAR comparison in Fig. 6 for the case of CCSNe at 60 kpc and with the LENs from SN1987A model and the GW from Dim2 model. The magenta dashed-line is 5σ significance threshold. Either $\mathrm{FAR_{old}}$ or $\mathrm{FAR_{new}}$ values lower than this line mean they pass 5σ. Thus, all clusters (blue dots) lie on the green area are recovered by the 1-parameter method. While, the red area is the improvement: the clusters recovered by the 2-parameter but not by the 1-parameter. The first row of Tab. 4 gives us the efficiency of the 1-parameter and the 2-parameter method. This yields an additional $\sim 12\%$ of signals from the 2-parameter that are discarded by the 1-parameter.

The cWB performs and analyses 2346 GW injections, and out of that, 784 of them have $\mathrm{FAR_{GW}} < 864/\mathrm{day}$. These candidates' significances are far too low to be even taken as *sub-threshold* triggers. Then, temporal coincidence analysis is performed with the input of these GW triggers and the list of LEN clusters. There are 554 coincident candidates that pass 5σ ($\sim 71\%$ of the GW triggers) with the

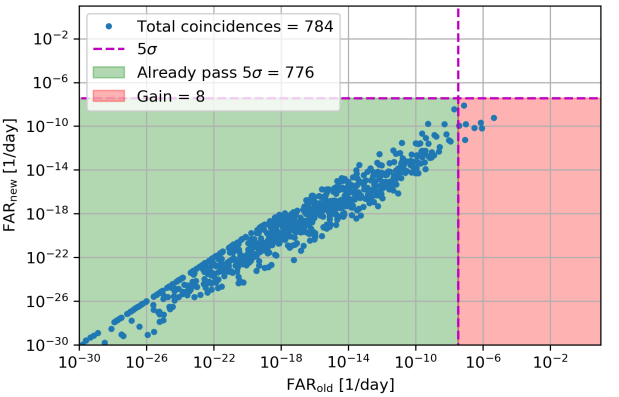


Fig. 7. The joint FAR of GW-LEN candidates obtained with the 2-parameter method ($\mathrm{FAR_{new}}$ y-axis) vs the 1-parameter ($\mathrm{FAR_{old}}$ x-axis) with the **KamLAND-LVD** (SN1987A-model) and HLV (Dim2-model) at 60 kpc.

Table 4. Efficiency comparison of the 1-parameter ($\eta_{1\mathrm{param}}$) and the 2-parameter ($\eta_{2\mathrm{param}}$) method for Fig. 6 (upper row) and 7 (lower row). The first column indicates the specific network of detectors and models. The second column shows the GW results with the threshold on the $\mathrm{FAR_{GW}} < 864$/day). The third and last columns are the efficiency with $> 5\sigma$ significance.

Network & Type of Injections	Recovered $\mathrm{FAR_{GW}} < 864$/d	$\eta_{1\mathrm{param}}$ [$> 5\sigma$]	$\eta_{2\mathrm{param}}$ [$> 5\sigma$]
HLV-KAM (Dim2-SN1987A)	784/2346= 33.4%	554/784= **70.7%**	650/784= **82.9%**
HLV-KAM-LVD (Dim2-SN1987A)	784/2346= 33.4%	776/784= **99.0%**	784/784= **100%**

standard 1-parameter method. However, if we employ the new 2-parameter method, we can get additional 110 recovered signals. This means the efficiency increases to $\sim 83\%$ (the first row of Tab. 4).

We also study further this method involving the LVD detector to be in the triple-detector configuration. We show the result in Fig. 7 and quantitative summary in the second row of Tab. 4. The improvement of our method seems less evident. This is clear due to the fact that the efficiency cannot go beyond the maximum value of 33.4%, where all the GW triggers from cWB are coincident with the list of LEN clusters with $> 5\sigma$ significance.

We then provide the result by using the Hud model in exchange of the SN1987A model. The FARs can be seen in Fig. 8 and the efficiency values in Tab. 5. This 2-parameter method gives us $\sim 7\%$ more recovered signals. We found that on average, the FARs the injections are $O(10^3)$ smaller with the 2-parameter than the 1-parameter for both emission models.

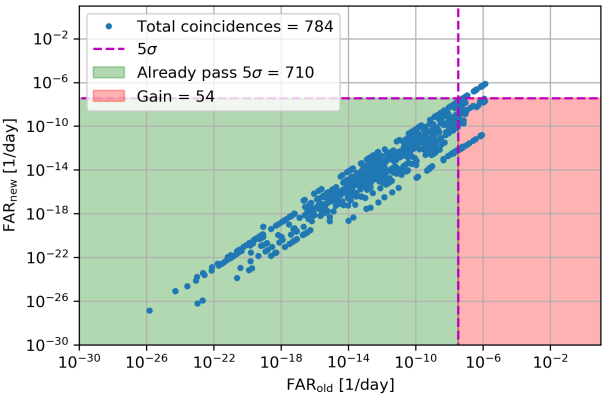


Fig. 8. The joint FAR of GW-LEN candidates obtained with the 2-parameter method ($\mathrm{FAR}_{\mathrm{new}}$ y-axis) vs the 1-parameter ($\mathrm{FAR}_{\mathrm{old}}$ x-axis) with the KamLAND-LVD (**Hud-model**) and HLV (Dim2-model) at 60 kpc.

Table 5. Efficiency comparison of the 1-parameter ($\eta_{1\mathrm{param}}$) and the 2-parameter ($\eta_{2\mathrm{param}}$) method for Fig. 8. The columns are analogous to Table 4.

Network & Type of Injections	Recovered $\mathrm{FAR}_{\mathrm{GW}} < 864$/d	$\eta_{1\mathrm{param}}$ [$> 5\sigma$]	$\eta_{2\mathrm{param}}$ [$> 5\sigma$]
HLV-KAM-LVD (Dim2-Hud)	784/2346= 33.4%	710/784= **90.6%**	764/784= **97.5%**

All in all, let us summarize the results. Fig. 2 for Dim2 model shows that the GW network HLV, by applying a threshold $\mathrm{FAR}_{\mathrm{GW}} \leq 864$/day, recovers about $\sim 33\%$ of the injected signals at 60 kpc. Notice that these GW triggers are far from statistically significant and clearly 0% passes a 5σ threshold. With this global network analysis and with our 2-parameter method, the multimessenger detection efficiency grows to $\sim 33\%$. If taking into account the weaker emission like Hud model, the detection efficiency arrives at the value of $33.4\% \cdot 97.5\% = 32.6\%$.

5. Conclusion

This paper is a timely description of the multimessenger strategy to combine GWs and LENs to search for CCSNe. More detail results will be discussed in a forthcoming paper[1] that has been accepted by JCAP.

Various models of GW and LEN emissions have been considered and used for a set of simultaneous injections into the background from the considered detectors. The data are then analysed with our proposed 2-parameter method comparing with the standard 1-parameter method. The performance is exercised at the level of efficiency estimation between these two methods for various analysis steps.

In general a multimessenger analysis can give a better efficiency to low-significant signals.

We highlight the improvement of the neutrino analysis with the ξ parameter in terms of FAR and FAP. This method could be of interest of the online alert system like SNEWS2.0.

All in all, the multimessenger campaign between GWs and LENs with the proposed method can increase the overall efficiency. Due to the higher sensitivity in LEN detectors, we can also do a targeted search in the (hopefully near) future to search for GWs from CCSNe.

Acknowledgement

The authors gratefully acknowledge the support of the NSF for the provision of computational resources. The work of GP is partially supported by the research grant number 2017W4HA7S "NAT-NET: Neutrino and Astroparticle Theory Network" under the program PRIN 2017 funded by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR).

References

1. O. Halim, C. Casentini, M. Drago, V. Fafone, K. Scholberg, C. F. Vigorito and G. Pagliaroli, Multimessenger analysis strategy for core-collapse supernova search: Gravitational waves and low-energy neutrinos (2021), arXiv:2107.02050.
2. G. Pagliaroli, F. Vissani, E. Coccia and W. Fulgione, Neutrinos from supernovae as a trigger for gravitational wave search, *Phys. Rev. Lett.* **103**, p. 031102 (Jul 2009).
3. I. Leonor, L. Cadonati, E. Coccia, S. D'Antonio, A. D. Credico, V. Fafone, R. Frey, W. Fulgione, E. Katsavounidis, C. D. Ott, G. Pagliaroli, K. Scholberg, E. Thrane and F. Vissani, Searching for prompt signatures of nearby core-collapse supernovae by a joint analysis of neutrino and gravitational wave data, *Classical and Quantum Gravity* **27**, p. 084019 (apr 2010).
4. K. Hirata *et al.*, Observation of a neutrino burst from the supernova sn1987a, *Phys. Rev. Lett.* **58**, 1490 (Apr 1987).
5. R. M. Bionta *et al.*, Observation of a neutrino burst in coincidence with supernova 1987a in the large magellanic cloud, *Phys. Rev. Lett.* **58**, p. 1494 (1987).
6. E. Alexeyev *et al.*, Detection of the neutrino signal from sn 1987a in the lmc using the inr baksan underground scintillation telescope, *Physics Letters B* **205**, 209 (1988).
7. Y. Fukuda *et al.*, The super-kamiokande detector, *Nucl. Instrum. Meth.* **A501**, 418 (2003).
8. M. Aglietta *et al.*, The most powerful scintillator supernovae detector: Lvd, *Il Nuovo Cimento A Series 11* **105**, 1793 (1992), cited By 103.
9. F. Suekane *et al.*, An overview of the kamland 1-kiloton liquid scintillator2004. arXiv:physics/0404071.
10. M. Aartsen *et al.*, The icecube neutrino observatory: Instrumentation and online systems, *Journal of Instrumentation* **12**, P03012 (mar 2017).
11. P. Antonioli *et al.*, SNEWS: The supernova early warning system, *New Journal of Physics* **6**, 114 (sep 2004).
12. S. A. Kharusi *et al.*, SNEWS 2.0: a next-generation supernova early warning system for multi-messenger astronomy, *New Journal of Physics* **23**, p. 031201 (mar 2021).

13. B. P. Abbott *et al.*, Multi-messenger observations of a binary neutron star merger*, *The Astrophysical Journal* **848**, p. L12 (oct 2017).
14. J. Aasi *et al.*, Advanced LIGO, *Class. Quant. Grav.* **32**, p. 074001 (2015).
15. F. Acernese *et al.*, Advanced Virgo: a second-generation interferometric gravitational wave detector, *Class. Quant. Grav.* **32**, p. 024001 (2015).
16. T. Akutsu *et al.*, Overview of KAGRA: Detector design and construction history (2020), arXiv:2005.05574.
17. C. D. Ott, The gravitational-wave signature of core-collapse supernovae, *Classical and Quantum Gravity* **26**, p. 063001 (feb 2009).
18. E. Abdikamalov, G. Pagliaroli and D. Radice, Gravitational waves from core-collapse supernovae (2020), arXiv:2010.04356.
19. J. Powell and B. Müller, Three-dimensional core-collapse supernova simulations of massive and rotating progenitors, *Mon. Not. Roy. Astron. Soc.* **494**, 4665 (2020).
20. M. Szczepanczyk, J. Antelis, M. Benjamin, M. Cavaglia, D. Gondek-Rosinska, T. Hansen, S. Klimenko, M. Morales, C. Moreno, S. Mukherjee, G. Nurbek, J. Powell, N. Singh, S. Sitmukhambetov, P. Szewczyk, J. Westhouse, O. Valdez, G. Vedovato, Y. Zheng and M. Zanolin, Detecting and reconstructing gravitational waves from the next galactic core-collapse supernova in the advanced detector era (2021), arXiv:2104.06462.
21. O. Halim, Searching for core-collapse supernovae in the multimessenger era: Low energy neutrinos and gravitational waves, PhD thesis, Gran Sasso Science Institute (GSSI)2020.
22. O. Halim *et al.*, Expanding core-collapse supernova search horizon of neutrino detectors, *Journal of Physics: Conference Series* **1468**, p. 012154 (feb 2020).
23. S. Klimenko and G. Mitselmakher, A wavelet method for detection of gravitational wave bursts, *Classical and Quantum Gravity* **21**, S1819 (sep 2004).
24. S. Klimenko *et al.*, A coherent method for detection of gravitational wave bursts, *Classical and Quantum Gravity* **25**, p. 114029 (may 2008).
25. M. Drago, Search for transient gravitational wave signals with unknown waveform in the ligo virgo network of interferometric detectors using a fully coherent algorithm, PhD thesis, Università degli Studi di Padova2010.
26. V. Necula, S. Klimenko and G. Mitselmakher, Transient analysis with fast wilson-daubechies time-frequency transform (jun 2012).
27. B. P. Abbott *et al.*, Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced ligo and advanced virgo, *Phys. Rev. D* **101**, p. 084002 (Apr 2020).
28. H.-T. Janka, Explosion mechanisms of core-collapse supernovae, *Annual Review of Nuclear and Particle Science* **62**, 407 (2012).
29. E. O'Connor and C. D. Ott, Black Hole Formation in Failing Core-Collapse Supernovae, *Astrophys. J.* **730**, p. 70 (2011).
30. D. Radice, V. Morozova, A. Burrows, D. Vartanyan and H. Nagakura, Characterizing the Gravitational Wave Signal from Core-Collapse Supernovae, *Astrophys. J. Lett.* **876**, p. L9 (2019).
31. H. Dimmelmeier *et al.*, Gravitational wave burst signal from core collapse of rotating stars, *Phys. Rev. D* **78**, p. 064056 (Sep 2008).
32. S. Scheidegger, R. Kaeppeli, S. C. Whitehouse, T. Fischer and M. Liebendoerfer, The Influence of Model Parameters on the Prediction of Gravitational wave Signals from Stellar Core Collapse, *Astron. Astrophys.* **514**, p. A51 (2010).
33. S. Woosley and A. Heger, The Progenitor stars of gamma-ray bursts, *Astrophys. J.* **637**, 914 (2006).

34. L. Hüdepohl, Neutrinos from the formation, cooling and black hole collapse of neutron stars, PhD thesis, Technische Universität München2014.
35. G. Pagliaroli *et al.*, Improved analysis of sn1987a antineutrino events, *Astroparticle Physics* **31**, 163 (2009).
36. F. Vissani, G. Pagliaroli and M. L. Costantini, A parameterized model for supernova electron antineutrino emission and its applications, *J. Phys. Conf. Ser.* **309**, p. 012025 (2011).
37. S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva and G. Mitselmakher, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, *Phys. Rev. D* **93**, p. 042004 (2016).
38. M. Drago, V. Gayathri, S. Klimenko, C. Lazzaro, E. Milotti, G. Mitselmakher, V. Necula, B. O'Brian, G. A. Prodi, F. Salemi, M. Szczepanczyk, S. Tiwari, V. Tiwari, G. Vedovato and I. Yakushin, Coherent waveburst, a pipeline for unmodeled gravitational-wave data analysis (2021), arXiv:2006.12604.
39. B. P. Abbott *et al.*, First targeted search for gravitational-wave bursts from core-collapse supernovae in data of first-generation laser interferometer detectors, *Phys. Rev. D* **94**, p. 102001 (Nov 2016).
40. B. P. Abbott *et al.*, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Rel.* **23**, p. 3 (2020).
41. J. Aasi *et al.*, Advanced LIGO, *Class. Quant. Grav.* **32**, p. 074001 (2015).
42. F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, *Class. Quant. Grav.* **32**, p. 024001 (2015).
43. M. J. Szczepańczyk, Multimessenger astronomy with gravitational waves from core-collapse supernovae, PhD thesis, Embry-Riddle Aeronautical University2018.
44. N. Agafonova *et al.*, Implication for the core-collapse supernova rate from 21 years of data of the large volume detector, *Astrophys. J.* **802**, p. 47 (2015).
45. M. Ikeda *et al.*, Search for supernova neutrino bursts at super-kamiokande, *Astrophys. J.* **669**, 519 (2007).
46. K. Abe *et al.*, Real-time supernova neutrino burst monitor at super-kamiokande, *Astropart. Phys.* **81**, 39 (2016).
47. C. Casentini *et al.*, Pinpointing astrophysical bursts of low-energy neutrinos embedded into the noise, *JCAP* **1808**, p. 010 (2018).
48. K. Eguchi *et al.*, First results from kamland: Evidence for reactor antineutrino disappearance, *Phys. Rev. Lett.* **90**, p. 021802 (Jan 2003).

A cryogenic and superconducting inertial sensor for the Lunar Gravitational–Wave Antenna, the Einstein Telescope and Selene-physics

F. Badaracco*, J. V. van Heijningen and E. C. Ferreira

Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain, Louvain-La-Neuve, B-1348, Belgium
**E-mail: francesca.badaracco@uclouvain.be*
http://www.uclouvain.be/

A. Perali

School of Pharmacy, Physics Unit, SuperNano Laboratory, University of Camerino, 62032 - Camerino (MC), Italy
http://www.supermaterials.org

INAF - Sezione di Camerino, Via Gentile III da Varano, 27, 62032 - Camerino (MC) - Italy

The Lunar Gravitational–Wave Antenna is a proposed low–frequency gravitational–wave detector on the Moon surface. It will be composed of an array of high-end cryogenic superconducting inertial sensors (CSISs). A cryogenic environment will be used in combination with superconducting materials to open up pathways to low–loss actuators and sensor mechanics. CSIS revolutionizes the (cryogenic) inertial sensor field with a modelled displacement sensitivity at 0.5 Hz of 3 orders of magnitude better than the current state–of–the–art. It will allow the Lunar Gravitational–Wave Antenna to be sensitive below 1 Hz, down to 1 mHz and it will also be employed in the forthcoming Einstein Telescope—a third-generation gravitational–wave detector which will make use of cryogenic technologies and that will have an enhanced sensitivity below 10 Hz. Moreover, CSIS seismic data could also be employed to obtain new insights about the Moon interior and what we can call the Selene-physics.

Keywords: Inertial sensor, Superconducting, Cryogenic, Seismic sensor, Lunar science, Lunar gravitational-wave antenna, Einstein Telescope

1. Introduction

The development of the Cryogenic Superconducting Inertial Sensor (CSIS) originates in the Einstein Telescope (ET) framework. ET will need a highly precise sensor to monitoring motion effects caused by the low–vibration cooling applied to its penultimate suspension stage, which indeed will operate at cryogenic temperatures and in vacuum.[1] Because of its extreme sensitivity and capability of working at cryogenic temperatures, CSIS will also be deployed on the Moon, exploiting it as a detector to reveal gravitational waves (GWs). The Lunar Gravitational–Wave Antenna (LGWA)[2] aims to exploit Moon's response to passing GWs and its resulting

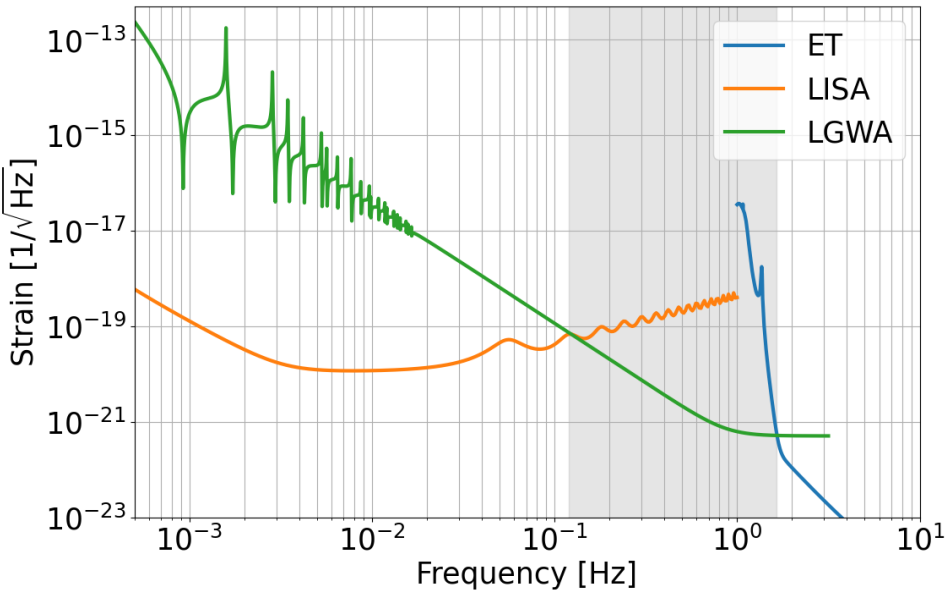


Fig. 1. Comparison in strain sensitivity between three future GW detectors. The notch–peak structures in the LGWA curve around 1–10 mHz result from the response of the Moon to passing GWs. Combining the performance of CSIS with the expected surface motion due to the lunar response yields the LGWA strain sensitivity. The grey rectangle highlights the gap that LGWA would bridge. See Ref.[2] for more info about the LGWA sensitivity curve.

surface motion. It will employ an array of 4 CSISs which can be regarded as the readout of a detector constituted by the Moon. The seismic data that LGWA would record will be used to study the Moon interior, its seismicity and its formation. This field of study is addressed as Selene-physics (the geophysics of the Moon).

LGWA will bridge the sensitivity gap between ET and the Laser Interferometer Space Antenna (LISA)[3] (see Fig. 1) between 0.12 and 1.5 Hz.

Improving detector sensitivity in the low–frequency band (below 10 Hz) entails a huge technical effort: Earth–bound GW detectors, like ET, are limited by the Newtonian noise,[4] while LGWA poses some difficulties deriving from deploying, assembling and powering the necessary instruments on the Moon.

The three detectors of which the sensitivity curves are shown in Fig. 1 are in development to extend our ability to detect GWs at lower frequencies. Signals of binary systems of (super–)massive black holes are found at these frequencies. Additionally, it will allow us to observe binary neutron stars already hours before the final merging.[5,6] Release of early warnings for the Electromagnetic (EM) follow–up will then be possible. The longer the observational time, the better the parameter estimation will be, also allowing to perform the sky–localization solely with the aid of LGWA (by exploiting the rotation of the Moon around the Earth). Polarization measurements and general relativity tests will also be possible.[2] Finally, it is

impossible to know, but exciting to imagine what kinds of unknown can be detected once we open up that low–frequency band.

2. A technical design for CSIS

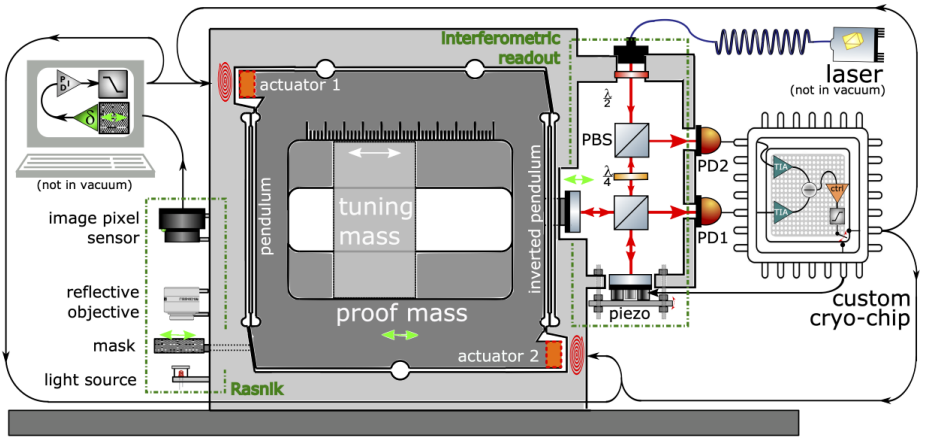


Fig. 2. Design of the CSIS sensor. We can identify four main parts: two readouts (highlighted by green dotted rectangles), spiral actuators (orange and red) and the proof mass (at the center). In Sec. 2 more details about the actuators and the proof mass are provided, while the two readouts are described in Secs. 3 and 4.

The design of CSIS is presented in Fig. 2. We can identify its four main parts: the Rasnik readout, the interferometric readouts, the two actuators on either side of the proof mass, and the niobium monolithic sensor mechanics. The latter will be the core of our sensor: a 1 kg mass suspended in a Watt's linkage configuration[7] (a combination of a pendulum and an inverted one) fabricated by Electrical Discharge Machining (EDM). The monolithic design allows to avoid having separate mechanical parts which would cause thermal dissipation.[8] Moreover, in Fig. 2 we see a tuning mass that will serve to adjust the resonance frequency of the sensor.

At cryogenic temperatures we can dramatically reduce the thermal noise, especially by employing materials that are superconductive ($T_c = 9.2$ K). At ~ 5 K the mechanical Q–factor of a niobium CSIS will be of the order of 10^4 (see Ref.[9]), which is a significant improvement with respect to the room temperature version. Concerning LGWA, permanent shadow zones (of the order of several km) on the Moon are a convenient way to provide low and stable temperatures (<40 K).[10] This will make less challenging reaching the temperatures where the niobium behaves as a superconductor.

The actuators of Fig. 2 will constitute of a thin layer of niobium deposited on a ceramic substrate and then attached to the sensor frame. By applying a current

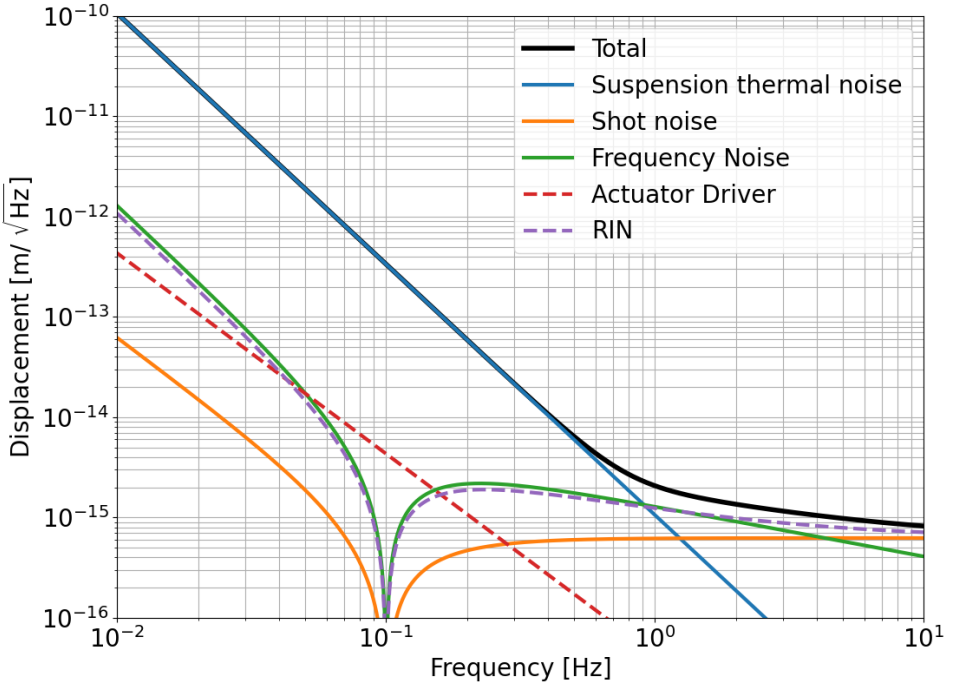


Fig. 3. Sensitivity and relative noises of niobium CSIS. The thermal noise was calculated considering a Q–factor of 10^4 and a temperature of 5 K. The other noises where calculated based on Ref.[9], where RIN is referring to the Relative Intensity Noise of the laser source of the interferometric readout and the actuator driver is calculated following Eq. 1.

to the spiral coil, a magnetic field will be created. Thanks to the Meissner effect, it will be expelled from the niobium of the proof mass, thus exerting a force on it.

To calculate CSIS sensitivity of Fig. 3, we need to take into account also the noise produced by the actuator driver:

$$n_{\mathrm{AD}} = \frac{\beta \mathrm{V_{DAC}}}{\mathrm{R_s m}\omega^2}, \tag{1}$$

where m is the mass, $\mathrm{R_s}$ is the sampling resistance (which determines the current into the actuator), $\mathrm{V_{DAC}}$ is the DAC voltage noise and β is the coil response value based on the latest simulations.

Another important advantage of employing a superconducting material is that it will avoid the thermal dissipation due to the eddy currents. This was a significant problem for the room temperature sensor version, where voice coils were employed to actuate the proof mass. In this way, the mechanical Q–factor was found to be highly limited by viscous damping associated with eddy currents induced on the moving metal surfaces by the voice coil stray field. The two readouts in Fig. 2 will be instead described in the next two sections.

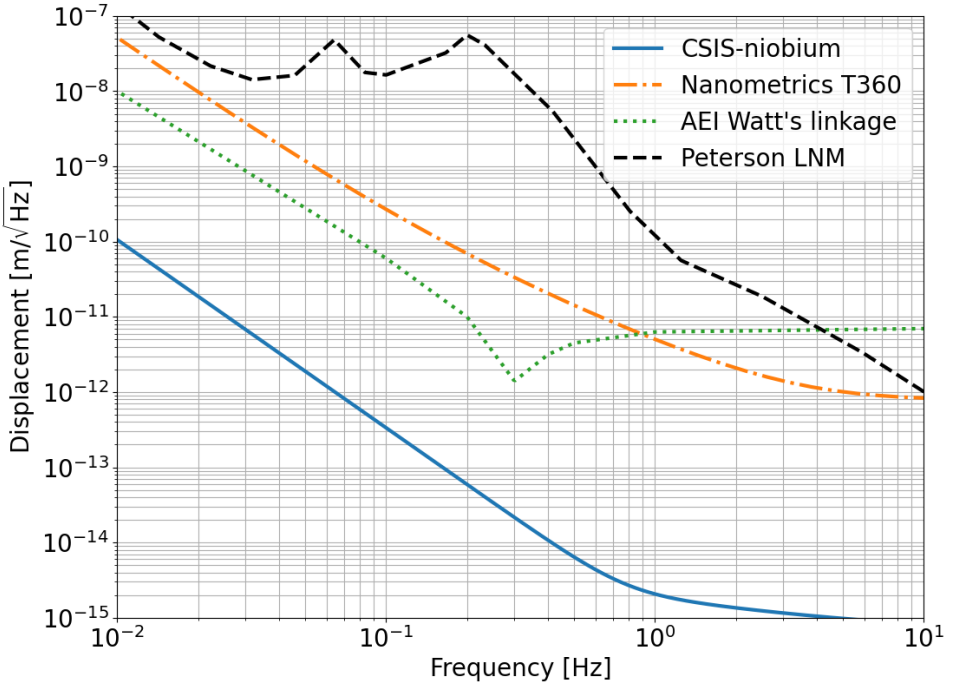


Fig. 4. Sensitivity of CSIS made of niobium compared to that of the most sensitive seismic sensors available. The Peterson's Low Noise Model (LNM) is depicted for comparison.

Thanks to its monolithic design, the employment of a material with a high Q-factor and thanks to two different readouts, CSIS will be able to reach sensitivities of 2 fm/$\sqrt{\text{Hz}}$ at 1 Hz (see Fig. 3). This is a factor 3000 better than the most sensitive seismic sensors (see Fig. 4, where the Peterson's Low Noise Model (LNM)[11] is depicted for comparison).

3. Interferometric readout

The interferometric readout that will be implemented in CSIS is based on that of Ref.[12] and it can be seen on the right side of Fig. 2 and in Fig. 6. It can be described as a simple Michelson interferometer with one more beamsplitter that allows to read the light coming back towards the laser. This second beamsplitter is needed to remove the common noise (such as intensity fluctuations in the laser power) by taking the differential signal between the two outputs. In this way, we can reach a shot noise limited sensitivity. The subtraction will be performed by a chip designed to work at cryogenic temperatures using CMOS technology.

The differential signal (see Fig. 5) will be fed through a control filter to the actuators which will dynamically lock the proof mass. This results in highly reduced relative motion between the proof mass and its frame and increased readout

dynamic range. Moreover, keeping the differential signal as close as possible to a fixed working point ensures optimal subtraction of the common noises. Indeed, in the room temperature version, it was found that the the desired shot-noise-limit performance was not yet achieved.[13] The reason was that there was still some residual motion of the proof mass, thus resulting in a sub-optimal subtraction of the two photodiodes signals. To solve this, we can use a second control loop to reduce as much as possible the residual motion of the proof mass: the Rasnik readout.

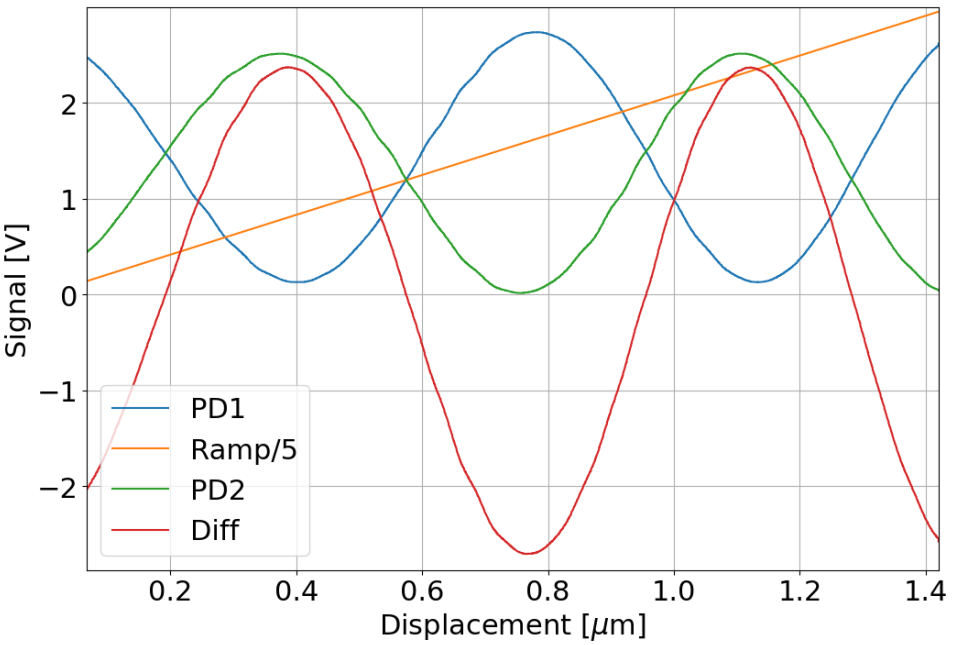


Fig. 5. Interference fringes obtained with a 1538 nm laser and moving one of the mirrors by injecting a ramp signal in its piezo actuator (PK44M3B8P2-Thorlabs). Using the known wavelength we find a piezo conversion factor of 96.1 nm V^{-1}. The minima and maxima are not exactly separated by the same distance probably due to non-linearities of the piezo actuator. The working points of the control loop are at 0 V of the differential signal.

Polarizing optics divert all the available power onto the photodiodes, avoiding dumping part of it. This is a important feature when working at cryogenic temperatures and it also means less shot noise for given injected power. Indeed, in Tab. 1 we can compare the powers received by the photodiodes and the overall shot noise when polarizing and non-polarizing optics are used. The total shot noise expected in the interferometer output depends on the amount of power falling on the photodiodes as $\mathrm{SN_i} \propto 1/\sqrt{P_\mathrm{i}}$. The total shot noise is then $\mathrm{SN_{tot}} = \sqrt{\mathrm{SN}^2_{\mathrm{PD}_1} + \mathrm{SN}^2_{\mathrm{PD}_2}}$.

We thus get a factor $\sqrt{3}$ of reduction in the shot noise when polarizing optics are employed with respect to the non-polarizing optics case.

Table 1. Comparison of interferometric readout outputs and shot noise without and with polarizing optics for a given input power P_0.

	power on PD1	power on PD2	SN_{tot}
previous configuration[12]	$P_0/4$	$P_0/8$	$\propto \sqrt{12/P_0}$
with polarizing optics	$P_0/2$	$P_0/2$	$\propto \sqrt{4/P_0}$

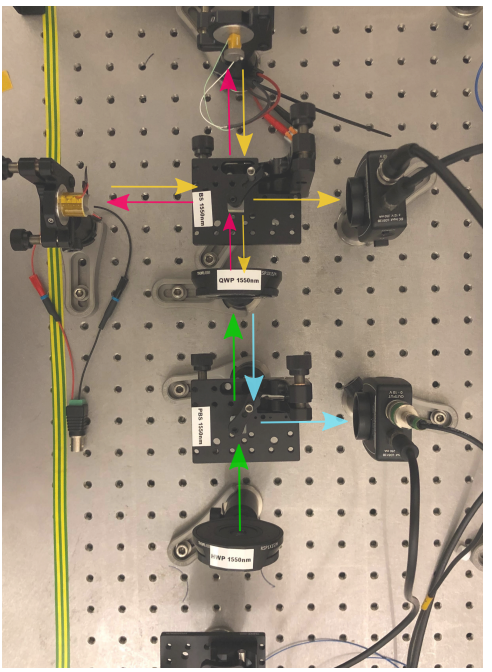


Fig. 6. Readout interferometer set up (laser on the bottom and photodiodes on the right). The arrows represent the light polarization: the green arrow right after the half wave plate (AHWP05M-1600-Thorlabs) is the linearly polarized light along the transmission axis of the polarizing beamsplitter (PBS104-Thorlabs), which, after passing through a quarter wave plate (AQWP05M-1600-Thorlabs), becomes circularly polarized (pink arrow). Reflection on the mirrors then reverses the handedness (yellow arrow). The second passage through the quarter wave plate with reversed handedness produces a linear polarization rotated by 90° degrees, which allows the light to be reflected by the polarizing beamsplitter instead of being transmitted.

4. Rasnik readout

The control loop driven by the Rasnik readout will have a larger dynamic range, capable of damping the proof mass resonance and thus reducing the proof mass residual motion. A Rasnik uses a back-lit chessboard mask which has its image projected via an objective on a pixel camera. The mask moves with the proof mass to which it is attached. The image that is recorded by the camera is then fed to a processing unit which performs a 2-dimensional Fourier transform. This transform will show a peak because of the many periodically spaced black-white transitions in the chessboard image. This peak shifts according to the motion of the mask and

a peak fitting algorithm returns then the Rasnik displacement measurement. The obtained sensitivity is 7 pm/$\sqrt{\text{Hz}}$ and a sub-pm/$\sqrt{\text{Hz}}$ sensitivity is expected given the implementation of certain improvements.[14] The beauty of this system is that its dynamic range is only dependent on the chessboard mask dimensions, which can be designed arbitrarily large.

5. Final remarks

The development of highly sensitive inertial sensors in the GW community opens the path to new GW detector concepts like LGWA, but also to the possibility of studying the Moon in more detail, giving rise to Selene-physics.

CSIS is the product of R&D in the ET community and it is the perfect candidate to be employed as readout in LGWA. It will reach a sensitivity of a few fm/$\sqrt{\text{Hz}}$ from 0.5 Hz onwards and, therefore, it will become the most sensitive inertial sensor capable of working at cryogenic temperatures in the low–frequency band.

References

1. ET Steering Committee Editorial Team, ET design report update 2020, *ET-0007A-20* (2020).
2. J. Harms *et al.*, Lunar Gravitational-wave Antenna, *The Astrophysical Journal* **910**, p. 1 (Mar 2021).
3. K. Danzmann and the LISA study team, LISA: laser interferometer space antenna for gravitational wave measurements, *Classical and Quantum Gravity* **13**, A247 (November 1996).
4. J. Harms, Terrestrial gravity fluctuations, *Living Reviews in Relativity* **22** (October 2019).
5. M. L. Chan, C. Messenger, I. S. Heng and M. Hendry, Binary neutron star mergers and third generation detectors: Localization and early warning, *Physical Review D* **97** (June 2018).
6. S. Grimm and J. Harms, Multiband gravitational-wave parameter estimation: A study of future detectors, *Physical Review D* **102** (July 2020).
7. J. Liu, L. Ju and D. G. Blair, Vibration isolation performance of an ultra-low frequency folded pendulum resonator, *Physics Letters A* **228**, 243 (April 1997).
8. A. Bertolini, R. DeSalvo, F. Fidecaro, M. Francesconi, S. Marka, V. Sannibale, D. Simonetti, A. Takamori and H. Tariq, Mechanical design of a single-axis monolithic accelerometer for advanced seismic attenuation systems, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **556**, 616 (January 2006).
9. J. V. van Heijningen, A fifty-fold improvement of thermal noise limited inertial sensitivity by operating at cryogenic temperatures, *Journal of Instrumentation* **15**, P06034 (June 2020).
10. D. A. Paige, M. A. Siegler, J. A. Zhang, P. O. Hayne, E. J. Foote, K. A. Bennett, A. R. Vasavada, B. T. Greenhagen, J. T. Schofield, D. J. McCleese, M. C. Foote, E. DeJong, B. G. Bills, W. Hartford, B. C. Murray, C. C. Allen, K. Snook, L. A. Soderblom, S. Calcutt, F. W. Taylor, N. E. Bowles, J. L. Bandfield, R. Elphic, R. Ghent, T. D. Glotch, M. B. Wyatt and P. G. Lucey, Diviner lunar radiometer observations of cold traps in the moon's south polar region, *Science* **330**, 479 (October 2010).

11. J. Peterson, Observation and modeling of seismic background noise, *Open-file report* **93-322** (1993).
12. M. B. Gray, D. E. McClelland, M. Barton and S. Kawamura, A simple high-sensitivity interferometric position sensor for test mass control on an advanced ligo interferometer, *Optical and Quantum Electronics* **31**, 571 (1999).
13. J. V. van Heijningen, Turn up the bass! Low-frequency performance improvement of seismic attenuation systems and vibration sensors for next generation gravitational wave detectors, *PhD thesis, Vrije Universiteit* (2018).
14. H. van der Graaf *et al.*, The ultimate performance of the rasnik 3-point alignment system, *arXiv:2104.03601v4* (2021).

Space gravitational wave antenna DECIGO and B-DECIGO

S. Kawamura[1,2,*] and the DECIGO working group

[1] *Physics department, Nagoya University, Nagoya, Aichi, 464-8602, Japan*
[2] *The Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Nagoya, Aichi, 464-8602, Japan*
[*] *E-mail: kawamura@u.phys.nagoya-u.ac.jp*

Deci-hertz Interferometer Gravitational Wave Observatory (DECIGO) is the future mission that aims to detect gravitational waves between 0.1 Hz and 10 Hz. DECIGO has four clusters, and one cluster of DECIGO consists of three differential Fabry-Perot interferometers with three drag-free spacecraft. Among various science targets of DECIGO, the detection of primordial gravitational waves is crucial. We are now updating the DECIGO design to ensure the detection of the primordial gravitational waves. We aim to launch B-DECIGO first at the earliest in 2032 as a pathfinder mission of DECIGO. B-DECIGO will not only establish the necessary technologies for DECIGO, but also accomplish a variety of important sciences.

Keywords: DECIGO; B-DECIGO; interferometer; primordial gravitational wave

1. Introduction

Since the first detection of gravitational waves by LIGO and Virgo[1] in 2015, gravitational waves from various sources such as black hole binary, neutron star binary,[2] intermediate-mass black hole binary,[3] and black hole/neutron star binary coalescences[4] have been detected by LIGO and Virgo. Especially the neutron star binary coalescence was accompanied by short gamma-ray burst[5] and electromagnetic-wave observations of the afterglow.[6]

In the future, gravitational wave astronomy will be further developed by the third-generation ground-based detectors, such as Einstein Telescope (ET)[7] and Cosmic Explorer (CE),[8] as well as space detectors, such as LISA.[9] Ground-based detectors aim to detect gravitational waves above 10 Hz, while LISA aims below 0.1 Hz. Detection of gravitational waves from compact star binaries between the two frequency bands is useful because it can play a role of follow-up of the gravitational waves already detected by LISA, as well as a role of a predictor of the gravitational waves to be detected by the ground-based detectors. Furthermore, this mid-frequency band has the potential of establishing a deep and effective window for gravitational waves coming from various sources, not only because expected gravitational-wave signals are larger at lower frequencies, but also the confusion limiting noise caused by irresolvable gravitational wave signals coming from white dwarf binaries.

One of the future missions, aiming at detecting gravitational waves in this mid-frequency (between 0.1 Hz and 10 Hz), is Deci-hertz Interferometer Gravitational Wave Observatory (DECIGO).[10,11] In this paper, pre-conceptual design and aimed science of DECIGO are explained in sections 2 and 3. Then the necessary update of the DECIGO design is described in section 4. In section 5, B-DECIGO, which is a pathfinder mission of DECIGO with solid scientific targets, is explained. Finally, in section 6, conclusions are summarized.

2. Design of DECIGO

The concept of DECIGO was born in 2001. At that time, the main science target was the direct measurement of the acceleration of the Universe. Since then, we have elaborated the design of DECIGO by implementing Fabry-Perot arm cavities, and finally, we established a pre-conceptual design of DECIGO.

As shown in Fig. 1, one cluster of DECIGO consists of three differential Fabry-Perot interferometers with three drag-free spacecraft. Each spacecraft has two floating mirrors, which are shared by two interferometers. The relative position of the mirror with respect to the spacecraft is measured by local position sensors attached to the spacecraft, and the position signals are fed back to the thrusters attached to the spacecraft to control the position of the spacecraft. The distance between the two spacecraft is 1,000 km. Laser light illuminating the interferometers has a wavelength of 515 nm and a power of 10 W. The mirrors have a radius of 0.5 m and a mass of 100 kg. The Fabry-Perot cavities have a finesse of 10. There are four clusters altogether placed separately, with two clusters at the same place in the heliocentric Earth-trail orbit, as shown in Fig. 2. The separation of the spacecraft is useful to increase the angular resolution of the sources, while placing the two clusters at the same place is useful to take correlation measurements to detect stochastic gravitational-wave background, such as primordial gravitational waves.

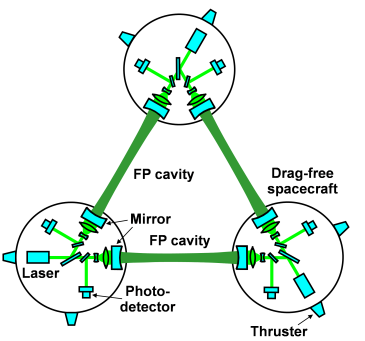


Fig. 1. Pre-conceptual design of DECIGO.

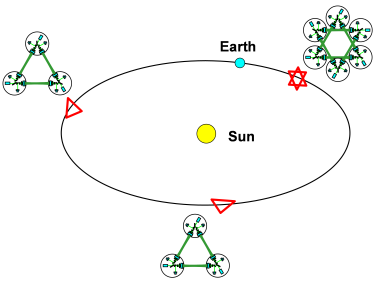


Fig. 2. Orbit of DECIGO.

3. Science of DECIGO

The target sensitivity of DECIGO and expected gravitational-wave signals from various sources are shown in Fig. 3. The target sensitivity of one cluster of DECIGO is $4 \times 10^{-24}\,\mathrm{Hz}^{-1/2}$ around 1 Hz. The sensitivity is limited by radiation pressure noise below 0.1 Hz and shot noise above 10 Hz. The shot-noise-limited sensitivity decreases above the cavity pole frequency, about 10 Hz, because the gravitational wave signals are canceled by the long storage time of the light in the arm cavities.

One cluster of DECIGO can detect gravitational waves coming from a large number of black hole binaries, for example, 1,000-solar-mass black holes at a redshift of $z = 10$ with a high signal-to-noise ratio. This could reveal the formation mechanism of intermediate-mass black holes. DECIGO can also detect gravitational waves from a large number of neutron star binaries, for example, at a redshift of $z = 1$ with a good signal-to-noise ratio. This can provide a prediction of the exact time and location of neutron star binary coalescences for ground-based gravitational-wave detectors as well as both terrestrial and space electromagnetic-wave observatories.[12]

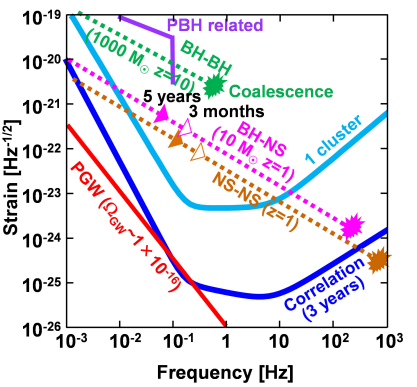


Fig. 3. Target sensitivity of DECIGO and expected gravitational wave signals.

This could also lead to the direct measurement of the acceleration of the expansion of the Universe, because it could cause a phase delay of gravitational wave signals.[10,13,14] This could provide reliable information about dark energy. DECIGO can also detect gravitational waves from, for example, about ten solar-mass black hole and neutron star binaries at a redshift of $z = 1$. This could provide a factor of 10^4 more precise test for the general theory of relativity.[15] In addition, DECIGO could detect gravitational waves coming from density fluctuations at the beginning of the Universe, which could have produced primordial black holes as candidates of dark matters.[16] DECIGO also has a variety of other science targets.[17–20]

The target sensitivity of DECIGO can be improved to $7 \times 10^{-26}\,\mathrm{Hz}^{-1/2}$ around 1 Hz by taking a correlation of two clusters st the same place for three years. This could enable detection of the primordial gravitational waves and revelation of the true inflation mechanism.[21] Moreover, DECIGO could observe a parity violation between the two circular polarizations,[22] separate the tensor, scalar, and vector modes of gravitational waves,[23] and reveal the thermal history after the inflation.[24,25] The design sensitivity of DECIGO was established more than ten years ago in such a way that DECIGO could detect the primordial gravitational waves corresponding to normalized energy of $\Omega_{\mathrm{GW}} = 2 \times 10^{-15}$, which was the upper limit of the primordial gravitational waves at that time. However, analysis based on the observations by the Planck satellite lowered the upper limit of Ω_{GW} to 1×10^{-16}. As a result, the DECIGO sensitivity became no longer good enough to detect the primordial gravitational waves.

4. Update of DECIGO design and target sensitivity

We are currently updating the design parameters of DECIGO, so that the new target sensitivity, determined by the updated design parameters, could ensure detection of the primordial gravitational waves with a good margin. We take the following three steps for the update. (1) For a given radius of the mirror, the cavity length, reflectivity of the mirrors, and laser power are optimized for the best signal-to-noise ratio, considering only the quantum noise as the noise source of DECIGO. (2) The same optimization is performed, considering other noise sources, such as the thermal noise of the mirrors. (3) The design and the corresponding sensitivity are updated, considering practical issues, such as the constraints of the mirror size.

We have completed the first step so far. For a given radius of the mirrors, we used a set of the three free design parameters (the cavity length, reflectivity of the mirrors, and laser power) to calculate the signal-to-noise ratio for the primordial gravitational waves. In the process of the calculation, optical loss due to the diffraction of the light, determined by the radius of the mirrors and the cavity length, was properly considered.[26] The signal-to-noise ratio was calculated between 0.1 Hz and 1 Hz, assuming that the confusion limiting noise caused by gravitational waves coming from extra-galactic white dwarf binaries does not exist above 0.1 Hz. The laser power of 100 W at the maximum was assumed to be available, although the default

design of the laser power for DECIGO is 10 W. The result of the optimization is shown in Fig. 4.[27] It indicates that by increasing the radius of the mirror from 0.5 m (default value) to 1 m, the cavity length can be extended from 1,000 km (default value) to about 5,000 km. This results in the improvement of the signal-to-noise ratio from 1.6 to about 100 with also the optimized reflectivity of the mirrors and laser power. It should be noted that in this optimization, the mass of the mirror is maintained to be constant, 100 kg (default value), even with the increased radius of the mirrors by shortening the thickness of the mirrors. We are now doing the same kind of optimization considering other noise sources, such as the thermal noise of the mirrors.

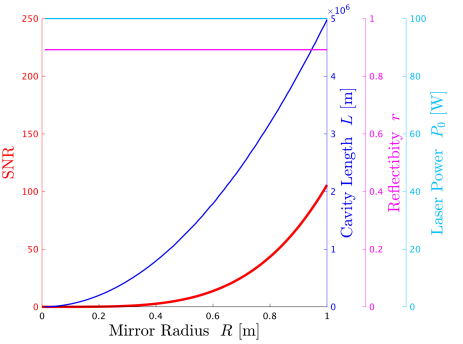


Fig. 4. Optimization of the cavity length, reflectivity of the mirrors, and laser power for a given radius of the mirrors in terms of the best signal-to-noise ratio.[27]

We are also investigating the possibility of implementing the quantum locking technique with an optical spring to DECIGO. We found that the quantum locking with an optical spring could improve the signal-to-noise ratio of the quantum noise to the primordial gravitational waves from 1.6 to 214 without changing the mirror size.[28,29]

5. B-DECIGO

We aim to launch B-DECIGO[30] at the earliest in 2032 before DECIGO. B-DECIGO can play a role in demonstrating the required technologies for DECIGO as well as a role of a real gravitational wave detector. B-DECIGO consists of three differential Fabry-Perot interferometers with three drag-free spacecraft. The distance between two spacecraft is 100 km, which is one-tenth of that of DECIGO. The laser wavelength is the same as that of DECIGO, and the laser power is 1 W (one-tenth of that of DECIGO). The mirror radius is 0.15 m (one-third of that of DECIGO), and the mass is 30 kg (one-third of that of DECIGO). The finesse of the arm cavities is 100 (ten times larger than that of DECIGO). B-DECIGO has one cluster, and the orbit is still under consideration.

The target sensitivity of B-DECIGO is shown in Fig. 5. The target sensitivity of B-DECIGO is lower than that of DECIGO approximately by a factor of ten. Since it has only one cluster, the correlation between the two clusters is not available. Although we cannot expect to detect the primordial gravitational waves, B-DECIGO can still bring us a variety of science.[31] It is expected to predict the timing of the neutron star binary coalescences together with the location of the source 100 times a year. It could reveal the origin of about 30 solar-mass black holes. It could estimate the parameter of compact star binaries much better than before. B-DECIGO could also demonstrate the method of removing the foreground gravitational wave signals coming from neutron star binaries, which is necessary for DECIGO. It could also demonstrate the technologies required to realize DECIGO, such as the Fabry-Perot cavities in space and drag-free system.

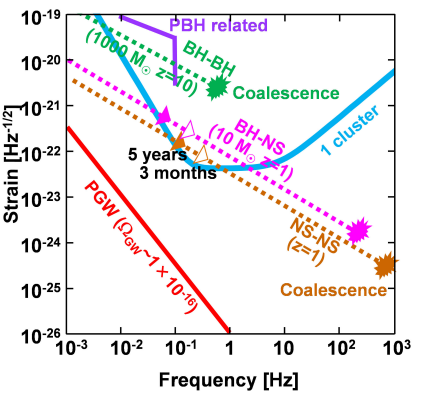


Fig. 5. Target sensitivity of B-DECIGO and expected gravitational wave signals.

6. Conclusions

DECIGO will accomplish a variety of outstanding sciences, including direct observation of the primordial gravitational waves. We are now investigating the improvement of target sensitivity of DECIGO to ensure the detection of the primordial gravitational waves. B-DECIGO will not only accomplish a variety of important sciences but also establish the necessary technologies for DECIGO.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, grant number JP19H01924.

References

1. B. P. Abbott, et al., *Phys. Rev. Lett.* **116** (2016) 061102.
2. B. P. Abbott, et al., *Phys. Rev. Lett.* **119** (2017) 161101.
3. R. Abbott, et al., *Phys. Rev. Lett.* **125** (2020) 101102.
4. R. Abbott, et al., *Astrophys. J. Lett.* **915:L5** (2021).
5. A. Goldstein, et al., *ApJ* **848** (2017) L14.
6. D. A. Coulter, et al., *Science* **358** (2017) 1556.
7. M. Punturo, et al., *Class. Quantum Grav.* **27** (2010) 194002.
8. B. P. Abbott, et al., *Class. Quantum Grav.* **34** (2017) 044001.
9. K. Danzmann, et al., *Nature Physics* **11** (2015) 613.
10. N. Seto, S. Kawamura, and T. Nakamura, *Phys. Rev. Lett.* **87** (2001) 221103.
11. S. Kawamura, et al., *Int. J. Mod. Phys. D* **28** (2019) 1845001.
12. R. Takahashi and T. Nakamura, *Astrophys. J.* **596** (2003) L231.
13. A. Nishizawa, et al., *Phys. Rev. D* **85** (2012) 044047.
14. A. Nishizawa, A. Taruya, and S. Saito, *Phys. Rev. D* **83** (2011) 084045.
15. K. Yagi and T. Tanaka, *Prog. Theor. Exp. Phys.* **123** (2010) 1069.
16. R. Saito and J. Yokoyama, *Phys. Rev. Lett.* **102** (2009) 161101.
17. N. Seto, *Astrophys. J.* **677** (2008) L55.
18. M. Kakizaki, S. Kanemura, and T. Matsui, *Phys. Rev. D* **92** (2015) 115007.
19. S. Hou, et al., *Phys. Rev. D* **103** (2021) 044005.
20. A. Piórkowska-Kurpas, et al., *ApJ* **908** (2021) 196.
21. S. Kuroyanagi, T. Chiba, and N. Sugiyama, *Phys. Rev. D* **79** (2009) 103501.
22. N. Seto, *Phys. Rev. D* **75** (2007) 061302(R).
23. A. Nishizawa, A. Taruya, and S. Kawamura, *Phys. Rev. D* **81** (2010) 104043.
24. S. Kuroyanagi, K. Nakayama, and J. Yokoyama, *Prog. Theor. Exp. Phys.* (2015) 013E02.
25. K. Nakayama, et al., *Phys. Rev. D* **77**, 124001.
26. S. Iwaguchi, et al., *Galaxies* **9** (2021) 9010009.
27. T. Ishikawa, et al., *Galaxies* **9** (2021) 9010014.
28. R. Yamada, et al., *Phys. Lett. A* **384** (2020) 126626.
29. R. Yamada, et al., *Phys. Lett. A* **402** (2021) 127365.
30. T. Nakamura, et al., *Prog. Theor. Exp. Phys.* (2016) 093E01.
31. S. Isoyama, H. Nakano, and T. Nakamura, *Prog. Theor. Exp. Phys.* (2018) 073E01.

Summary of the parallel session GW2

Dongfeng Gao*, Wei-Tou Ni†, Jin Wang‡, Ming-Sheng Zhan§, and Lin Zhou¶

Wuhan Institute of Physics and Mathematics, APM, Chinese Academy of Sciences, Wuhan 430071, China

*dfgao@wipm.ac.cn

†*wei-tou.ni@wipm.ac.cn*

‡*wangjin@wipm.ac.cn*

§*mszhan@wipm.ac.cn*

¶*lzhou@wipm.ac.cn*

This article summarizes the talks in the session GW2 of *the Sixteenth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories*, 5-10 July, 2021, on *Mid-frequency (0.1-10 Hz) gravitational waves: Sources and detection methods* with a review on strain power spectral density amplitude of various mid-frequency gravitational wave projects/concepts and with extended summaries on the progress of ZAIGA project and on the conceptual study of AMIGO.

Keywords: Mid-frequency gravitational waves; Earth-based gravitational wave detectors; Moon-based gravitational wave detectors; Space-borne gravitational wave detectors

1. Introduction

The mid-frequency GW (Gravitational Wave) band (0.1-10 Hz) between the LIGO-Virgo-KAGRA detection band and LISA-TAIJI/TIANQIN detection band is rich in GW sources. In addition to the intermediate BH (Black Hole) Binary coalescence, it can also come from the inspiral phase of stellar-mass coalescence and from compact binaries falling into intermediate BHs. Detecting mid-frequency GWs enables us to study the compact object population, to test general relativity and beyond-the Standard-Model theories, to explore the stochastic GW background and so on. In addition to DECIGO and BBO, the detection proposals under study include AEDGE, AIGSO, AION, AMIGO, DO, ELGAR, GLOC, INO, LGWA, MAGIS, MIGA, SOGRO, TIAGO, TOBA, ZAIGA, etc. Great advances have accumulated since MG15.

In the GW2 parallel session on 7 July 2021, there were ten talks.[1–10] First, an outlook of the mid-frequency GW detection and AMIGO were reviewed[1] followed by three talks on the mid-frequency astrophysical/cosmological sources and their propagation through the inhomogeneous universe for the GW observations from the Earth, the Moon and the Space.[2–4] The next two talks were on the Lunar Gravitational-Wave Antenna (LGWA) and its sensor development.[5,6] The last four talks were on Earth-based detectors TOBA,[7] ZAIGA[8] and AION,[9] and space-borne detectors AEDGE,[9] DECIGO[10] and B-DECIGO.[10]

Although there have been activities across all GW frequency spectrum,[11] current activities on GW detection are largely in the high-frequency band (10 Hz-10 kHz) of LIGO-Virgo-KAGRA and low-frequency band (0.1 mHz-100 mHz) of LISA-TAIJI/TIANQIN together with PTAs (300 pHz-100 nHz) and CMB polarization observations (1 aHz-10 fHz). Nevertheless, activities in the mid-frequency GW band have increased steadily since last decade. In the 2030s, ET and Cosmic Explorer will be deployed in the high-frequency band; LISA and TAIJI/TIANQIN will be deployed in the mHz low-frequency band. To complete the map, some mid-frequency GW detectors are desired to be deployed also. Table 1 lists the strain power spectral density (psd) amplitude at 0.1 Hz, 1 Hz and 10 Hz for various mid-frequency GW detectors together with the strain psd amplitude for ET at 1 Hz & 10 Hz and Cosmic Explorer at 10 Hz, and the strain psd amplitude for LISA at 0.1 Hz and for & TAIJI & TIANQIN at 0.1 Hz, 1 Hz and 10 Hz.

Harms *et al.*[31] have discussed ground-based GW detector concepts to extend the present ground-based interferometers' detection spectral range to middle-frequency band 0.1–10 Hz. They find that although there are no fundamental limits to the detector sensitivity in this band, the technical limits for the Newtonian-noise cancellation from infrasound and seismic surface fields are challenging to overcome.[31,32] Newtonian noise needs to be measured and subtracted. ET and Cosmic Explorer need to deal with this issue to certain extent and their sensitivity extends only down to 1 Hz and 5 Hz amplitudes respectively. ET is listed at 10 Hz and 1 Hz and Cosmic Explorer is listed at 10 Hz in Table 1 for comparison. On the lower frequency side, LISA, TAIJI and TIANQIN are listed for comparison. For space missions, local gravity noise (Newtonian noise needs to be compensated, and/or measured and subtracted. Due to the success of LISA Pathfinder mission,[33,34] this issue is largely solved.

The goal on the strain psd noise of ET at 1 Hz is $\sim 1 \times 10^{-21}\mathrm{Hz}^{-1/2}$, while that of LISA at 0.1 Hz is $4 \times 10^{-20}\mathrm{Hz}^{-1/2}$. Any new mid-frequency proposals aiming at a starting time after ET and LISA turn-on can keep these sensitivities in mind for their designs.

In Section 2, we discuss and review sources and scientific goals for mid-frequency GW detection. In Section 3, we discuss Earth-based detectors and summarize the TOBA, AION and ZAIGA talks. In Section 4, we discuss Moon-based detectors and summarize the LGWA and GLOC talks. In Section 5, we discuss space-borne GW detectors and summarize the AEDGE, AMIGO, DECIGO and B-DECIGO talks.

2. Sources and Scientific Goals

The general scientific goals of mid-frequency GW detectors are:

(i) to bridge the spectral gap between high-frequency and first-generation low-frequency GW sensitivities for detecting intermediate mass BH coalescence;

(ii) to detect inspiral phase and predict time of stellar-mass binary black hole coalescence together with neutron star coalescence or neutron star-black hole

Table 1. Strain power spectral density (psd) amplitude at 0.1 Hz, 1 Hz and 10 Hz for various mid-frequency GW detectors. In the first column, [E] means Earth-based, [M] means Moon-based, and [S] means space-borne. All units are in $\mathrm{Hz}^{-1/2}$.

GW detector or detector concept	Method of GW detection	Strain psd amplitude at 0.1 Hz	Strain psd amplitude at 1 Hz	Strain psd amplitude at 10 Hz
ET[12][E]	Micheson Interferometry with Fabry-Perot Cavities	n.a.	1×10^{-21}	1×10^{-24}
CE[13][E]	Micheson Interferometry with Fabry-Perot Cavities	n.a.	n.a.	6×10^{-25}
LISA[14][S]	TDI interferometry	4×10^{-20}	4×10^{-19}	n.a.
TAIJI[15][S]	TDI interferometry	3.3×10^{-20}	3.3×10^{-19}	3.3×10^{-18}
TIANQIN[16][S]	TDI interferometry	1.5×10^{-20}	3.5×10^{-20}	4×10^{-19}
aSOGRO[17][E]	SQUID capacitance-bridge transducer	5×10^{-20}	3×10^{-21}	1×10^{-21}
TOBA[18][E]	Torsion Bar	1×10^{-19}	8×10^{-20}	5×10^{-20}
ELGAR/MIGA[19][E]	Laser-Linked Atom Interferometry	3×10^{-19}	5×10^{-22}	5×10^{-22}
MAGIS-1 km[20][E]	Laser-Linked Atom Interferometry	n.a.	1×10^{-21}	n.a.
AION-1 km[21][E]	Laser-Linked Atom Interferometry	n.a.	1×10^{-21}	n.a.
ZAIGA[22][E]	Laser-Linked Atom Interferometry	1.4×10^{-18}	2.4×10^{-21}	1.2×10^{-21}
LGWA[23][M]	Inertial sensors with station location measurements	1×10^{-20}	1×10^{-20}	n.a.
GLOC[24][M]	Micheson Interferometry with with Fabry-Perot Cavities	n.a.	1×10^{-23}	1×10^{-24}
AEDGE[25][S]	Laser-Linked Atom Interferometry	3.5×10^{-23}	3×10^{-23}	n.a.
AIGSO[26][S]	Atom Interferometry	7×10^{-19}	1×10^{-20}	3×10^{-22}
INO-d[27][S]	Clocks with laser link	$1 - 2 \times 10^{-18}$	$1 - 2 \times 10^{-18}$	$1 - 2 \times 10^{-18}$
b-AMIGO[28][S]	Michelson X0/X TDI	5.5×10^{-21}	3×10^{-21}	6×10^{-21}
AMIGO[28][S]	Michelson X0/X TDI	4.3×10^{-21}	1×10^{-21}	2.3×10^{-21}
e-AMIGO[28][S]	Michelson X0/X TDI	4.3×10^{-21}	5×10^{-22}	8.5×10^{-22}
TIANGO[29][S]	TDI Interferometry	1×10^{-21}	4×10^{-22}	4×10^{-22}
DO-Conservative[30][S]	TDI Interferometry	1.5×10^{-22}	1.3×10^{-22}	3×10^{-22}
DO-Optimal[30][S]	TDI Interferometry	4×10^{-23}	2×10^{-23}	4×10^{-23}
B-DECIGO[10][S]	Fabry-Perot arm cavities	1.5×10^{-22}	5×10^{-23}	8×10^{-23}
DECIGO 1 cluster[10][S]	Fabry-Perot arm cavities	1.5×10^{-23}	5×10^{-24}	8×10^{-23}
DECIGO 3-year correlation[10][S]	Fabry-Perot arm cavities	3×10^{-25}	7×10^{-26}	3×10^{-25}

coalescence for ground interferometers, e.g., the inspiral GWs from sources like GW150914;[35]

(iii) to detect compact binary inspirals for studying stellar evolution and galactic population;

(iv) to detect middle frequency GWs from compact binaries falling into intermediate mass BHs for multi-band observation with ground-based GW detectors;[36]

(v) to study the compact object population and to estimate/explore the astrophysical stochastic background from stellar object inspirals;

(vi) to determine Supermassive Black Hole properties using GW radiation from surrounding stellar-mass BH binaries;[3,37]

(vii) to test general relativity and beyond-the-standard-model theories.

In subsection 2.1, the goal (vi) will be discussed. General treatments on astrophysical GW sources and associated goals for detection in the middle-frequency GW bands can be found in Refs. [35, 36, 38, 39]. In subsection 2.2, Generation, propagation and detection of GWs in inhomogeneous universe will be discussed. In section 3-5, specific scientific goals for individual detector will be stated.

2.1. *Determination of Supermassive Black Hole properties using GW radiation from surrounding stellar-mass BH binaries*[3,37]

With space-borne gravitational-wave observatories, one may use stellar-mass BH binary as a signal carrier to probe modulations induced by a central supermassive BH (SMBH) to further place constraints on the SMBH's properties. As an example, Yu and Chen[3,36] showed that the de Sitter precession of the inner stellar-mass binary's orbital angular momentum around the angular momentum of the outer orbit would be detectable when the precession period is comparable to the duration of observation, typically a few years. The precession could be combined with the Doppler shift arising from the outer orbital motion to determine the mass of the SMBH and the outer orbital separation with percent-level accuracy. Joint detection with Earth-based detector would extend the detection threshold to a precession period to $\sim$100 yr.

2.2. *Generation, propagation and detection of GWs in inhomogeneous universe*

In the talk, Fier *et al.*[2,40] presented their recent studies on GWs produced by remote compact astrophysical sources. To describe such GWs properly, Fier *et al.* introduced three scales, the typical wavelength of GWs, the scale of the cosmological perturbations, and the size of the observable universe. For GWs to be detected by the current and foreseeable detectors, they can be well approximated as high-frequency GWs. The spatial, traceless, and Lorentz gauge conditions can be imposed simultaneously to simplify the field equations, even when the background is not vacuum, as long as the high-frequency GW approximation is valid. Fier *et al.* developed the geometrical optics approximation to such GWs, and calculated the gravitational integrated Sachs-Wolfe effects due to the presence of the cosmological scalar and tensor perturbations to read out explicitly the dependence of the amplitude, phase and luminosity distance of the GWs.

3. Earth-Based Detectors – TOBA, AION and ZAIGA

Earth is noisy for strain measurement. From 10 Hz to 10 kHz, vibration noise needs to be suppressed. This issue is basically solved by active and passive vibration isolation after decades of research and development. In the mid-frequency band of GW detection on Earth, the Newtonian noise (the gravity gradient noise) cannot be suppressed. It needs to be measured and subtracted. TOBA uses torsion bars, and AION and ZAIGA use atom interferometry to measure the Newtonian noise and subtract it.

3.1. *TOBA*

TOBA (TOrsion-Bar Antenna) is a mid.-frequency gravitational-wave antenna. It is formed by two bar-shape test masses, each suspended as a torsion pendulum. Tidal effect originated by incoming gravitational wave will be detected as differential angular motion of these two bars. The fundamental sensitivity is $10^{-19}\mathrm{Hz}^{-1/2}$ at 0.1 Hz frequency band, assuming 10-m scale cryogenic detector at 4 K with the rotations of the bars measured with Fabry–Perot cavities at the end of the bars.[7,18] Though this sensitivity is not comparable with space antennae, it is sufficient to observe some intermediate-mass black-hole inspirals in our universe. Also, operation on ground is advantageous in development time and cost. The Phase-III prototype is to complete the demonstration of noise reduction reaching a sensitivity of about $10^{-15}\mathrm{Hz}^{-1/2}$ at 0.1 Hz. Measuring the terrestrial gravity fluctuation with a sensitivity of about or below $10^{-15}\mathrm{Hz}^{-1/2}$ at 0.1 Hz is useful for two geophysical purpose – earthquake early warning and Newtonian noise reduction for GW detection.

3.2. *AION*

AION (Atom Interferometer Observatory and Network) is a proposed experimental programme using cold strontium atoms to search for ultra-light dark matter, to explore gravitational waves in the mid-frequency, and to probe other frontiers in fundamental physics. AION has 3 stages of development, AION-10 m, AION-100 m and AION-1 km in parallel to MAGIS-10 m, MAGIS-100 m and MAGIS-1 km. AION would share many technical features with the MAGIS experimental programme, and synergies would flow from operating AION in a network with MAGIS, as well as with other atom interferometer experiments such as MIGA, ZAIGA and ELGAR.[9,21]

3.3. *ZAIGA*

ZAIGA (the Zhaoshan long-baseline Atom Interferometer Gravitation Antenna) is a proposed underground long-baseline atom interferometer (AI) facility (see Fig. 1), aiming for experimental research on gravitation and related problems.[22] It is located in the mountain named Zhaoshan which is about 80 km southeast to Wuhan city.

ZAIGA will consist of gravitational wave detection (ZAIGA-GW), dark matter detection (ZAIGA-DM), high-precision test of the equivalence principle (ZAIGA-EP), clock-based gravitational red-shift measurement (ZAIGA-CE-R), rotation measurement and gravitomagnetic effect (ZAIGA-RM), and geological and geophysical measurement (ZAIGA-GG). The first stage of the project (2021-2025) will include building a vertical 300-meter long tunnel to interrogate three 10-meter atom interferometers and atomic clocks (aiming for ZAIGA-EP, ZAIGA-DM and ZAIGA-CE-R), and building a horizontal 1.5-km long tunnel (aiming for ZAIGA-RM and ZAIGA-GG). Depending on the performance in the first stage, the second stage of the project will construct an array of 1-km arm-length horizontal triangular tunnels and be dedicated to ZAIGA-GW.

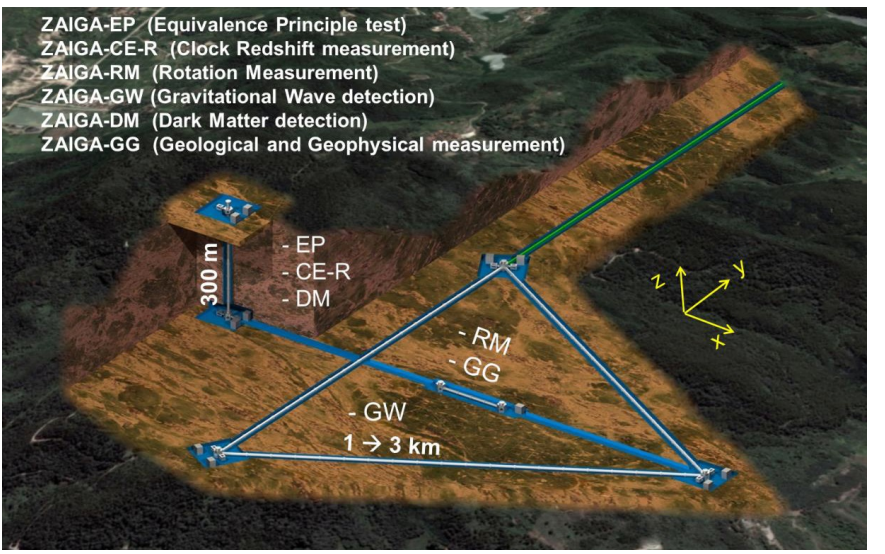


Fig. 1. Schematic diagram of ZAIGA design.

The site exploration started immediately after the ZAIGA plan was put forward. Relevant work for site exploration keeps on going now. The following progress on the technical and theoretical aspects of ZAIGA has also been made.

In the past three years, with our 10-m dual-species atom interferometer, we carried out a joint mass-energy test of the equivalence principle (EP), by using rubidium atoms with specified mass and internal energy.[41] By extending the four-wave double-diffraction Raman transition method (4WDR) to ^{85}Rb and ^{87}Rb atoms with different angular momenta, we can measure their differential gravitational acceleration. The Eötvös parameter, η, of the four paired combinations ($^{85}Rb|F = 2\rangle - ^{87}Rb|F = 1\rangle$, $^{85}Rb|F = 2\rangle - ^{87}Rb|F = 2\rangle$, $^{85}Rb|F = 3\rangle - ^{87}Rb|F = 1\rangle$ and $^{85}Rb|F = 3\rangle - ^{87}Rb|F = 2\rangle$) were measured to be $\eta_1 = (1.5 \pm 3.2) \times 10^{-10}$,

$\eta_2 = (-0.6 \pm 3.7) \times 10^{-10}$, $\eta_3 = (-2.5 \pm 4.1) \times 10^{-10}$ and $\eta_4 = (-2.7 \pm 3.6) \times 10^{-10}$, respectively. The violation parameters of mass and internal energy are constrained to $\eta_0 = (-0.8 \pm 1.4) \times 10^{-10}$ and $\eta_E = (0.0 \pm 0.4) \times 10^{-10}$. This work opens a door for joint tests of two attributes beyond the traditional pure mass or energy tests of EP with quantum systems. It also provides us a firm technical foundation for our proposed ZAIGA-EP experiment. For the rotation measurement, Yao *et al.* in our group has improved the precision of the measurement by one-order of magnitude recently and keep on advancing the precision and accuracy.[42]

Recently, a theoretical study on the possibility of detecting the ultralight scalar dark matter (DM) with ZAIGA was completed.[43] Starting with a popular scalar DM model, where the DM field is assumed to linearly couple to the standard model fields through five coupling parameters, the DM candidate can be detected by determining whether the five coupling parameters are zero or not. The solution to the DM field contains a background oscillation term and a local exponential fluctuation term. The DM signals in ZAIGA has been calculated. For a pair of AIs vertically separated by 300 meters, the DM-induced differential phase consists of three contributions, coming from the DM-induced changes in atomic internal energy levels, atomic masses and the gravitational acceleration. For a pair of AIs horizontally separated by several kilometers, the signal comes only from the DM-induced changes in atomic internal energy levels. The constraints on five DM coupling parameters for ZAIGA-DM have been estimated. It turns out that the proposed constraints would complement the MICROSCOPE space experiment[44] and would be several orders of magnitude better in higher-mass parametric regions. It also clearly shows the advantages and disadvantages of the two configurations of ZAIGA-DM. For the vertical configuration, the advantage is that all the five DM coupling parameters can be constrained, and the disadvantage is that the arm-length is not easy to extend. For the horizontal configuration, the advantage is that the arm-length is relatively easy to extend, and the disadvantage is that only two DM parameters can be constrained. For other type of DM candidates, the relevant studies are in progress.

4. Moon-Based Detectors – LGWA and GLOC

Moon is much quieter seismologically and does not have an atmosphere. This makes Moon itself a possible good detector by monitoring its surface motion excited by GWs. This also makes an interferometer GW detector based on the Moon possible.

4.1. *Lunar Gravitational-Wave Antenna*[5,23]

One of the first concepts proposed for the detection of gravitational waves by Joseph Weber is to monitor the quadrupolar vibrations of elastic bodies excited by GWs. At laboratory scale, these experiments became known as resonant-bar detectors, which form an important part of the history of GW detection. Due to the dimensions of

these bars, the targeted signal frequencies were in the kHz range. It was also Weber who suggested to monitor vibrations of Earth and Moon to search for gravitational waves in the mHz band. His Lunar Surface Gravimeter was deployed on the Moon in 1972 by the Apollo 17 crew. A design error made it impossible to carry out the intended search for GWs, but the idea remains intriguing. Harms *et al.*[5,23] have proposed a new concept, the Lunar Gravitational-Wave Antenna (LGWA), based on Weber's idea. LGWA would have a rich GW and multi-messenger science case with galactic binaries and more massive black-hole binaries. It would also serve as a high-precision geophysical station shedding light on the interior structure of the Moon, the mechanisms of moonquakes, and the Moon's formation history. The key component is a next-generation, high-sensitivity seismometer to be deployed on the Moon. For its most sensitive realization, LGWA would have to be deployed in a permanent shadow near the south or north pole of the Moon to benefit from the natural cryogenic environment. This would improve the sensitivity of the seismometer and also provide a lower-noise environment due to the absence of thermally induced seismic events that were observed by the Apollo seismometers. Powering of the seismic stations and data transfer pose additional challenges for such a deployment.

4.1.1. *Inertial sensors for the LGWA*[6]

The core of LGWA will be composed of an array of high-end seismic sensors: CSIS (Cryogenic Superconducting Inertial Sensor) on the Moon's surface. A cryogenic environment will be used in combination with superconducting materials to open up pathways to low-loss actuators and sensor mechanics.

CSIS revolutionizes the (cryogenic) inertial sensor field by obtaining a displacement sensitivity at 0.5 Hz of 3 orders of magnitude better than current state-of-art. It will allow LGWA to be sensitive below 1 Hz, down to 1 mHz. It will also be employed in the forthcoming Einstein Telescope (ET) - a third-generation gravitational-wave detector that will make use of cryogenic technologies and will have an enhanced sensitivity below 10 Hz. Moreover, CSIS seismic data could also be employed to get new insights about the Moon's interior... and the selenphysics (the Moon's geophysics). A technical design for CSIS with both interferometric readout and Rasnik readout is presented at the meeting.

4.2. *GLOC (Gravitational-Wave Lunar Observatory for Cosmology)*[4,24]

Taking advantage of the vacuum environment and the low seismic activities of the Moon, GLOC proposes to build a full laser-interferometric triangle-shape GW detector with 40 km arm length to explore fundamental cosmology and multi-band, multi-messenger astrophysics from the Moon. The speaker focused on goals that are unique to GLOC, and compare the detection landscape of the elusive intermediate-mass black holes between GLOC and other prominent space-based missions, 3G detectors and deci-Hertz concepts.

5. Space-Borne GW Detectors

The space environments are much quieter in terms of the gravity gradient noise consideration as compared to the Earth and Moon environments. It is so quiet that the major gravity gradient noise comes from local environments inside the spacecraft. In space, the GW strains propagating to the detector are measured above the geodetic motion or the desired controlled motion/position of the spacecraft. For spacecraft to follow geodesics, one needs to perform drag-free control. For spacecraft to follow desired motion/position over the geodetic motion, one needs to actuate acceleration on the geodetic motion in addition to basic drag-free control. LISA Pathfinder launched on December 3, 2015 has completely met the LISA drag-free noise requirement and has successfully demonstrated the first generation drag-free technology requirement for space detection of GWs.[33,34]

Many space GW detection proposals need to use constant/equal arm configurations. They are AEDGE,[9,25] AIGSO,[26,45] DECIGO/B-DECIGO,[10] etc. AIGSO has 10 km arm length, the shortest arm length among these mission proposals. The calculated actuation acceleration needed to maintain such orbits for AIGSO is around 10 pm s^{-2}.[45,46]

In the mission of LISA Pathfinder, different levels of force and torque authority were implemented, from the nominal configuration with x-force authority (on the sensitive line-of-sight axis) of 1100 pm s^{-2} to the URLA configuration levels, with x-force authority of 26 pm s^{-2}.[47] The published LPF differential acceleration noise floor is established by measurements in this configuration. Specifically, LISA Pathfinder demonstrated that when a constant out of the loop force with amplitude of 11.2 pN was applied to the sensitive axis of TM1 (Test Mass 1) for reducing the gravitational imbalance between the TMs, this force does not introduce significant noise or calibration errors.[47] Basically, the accelerometer part of the constant-arm technology is already demonstrated by LISA Pathfinder for AIGSO.

B-DECIGO has a nominal arm length of 100 km, DECIGO 1,000 km, and AMIGO 10,000 km. The actuation accelerations needed are respectively 10, 100, and 1000 times more than AIGSO. While the actuation accelerations needed for constant arm implementation of b-DECIGO and DECIGO is still basically in the LISA Pathfinder nominal configuration range, the actuation accelerations for constant arm AMIGO is one order larger. On what noise level could the actuation accelerations be done is an issue that needs to be studied and demonstrated carefully for AMIGO. A suggestion is to use an additional test mass (i.e. a pair) to alternate with the original one.[46,48]

GW space missions do not usually require large communication capacity. The deployment to 2-4 degrees behind or ahead of the Earth orbit is no more complicated as to L1 or L2 Sun-Earth Lagrange points. We hereby provide *an example for taking less than a week to reach the (pre)-science orbits.* A last-stage launch from 300 km LEO (Low Earth Orbit) to an eccentric Hohmann orbit with apogee 262931 km (The period of this Hohmann transfer orbit is about 6 days.) (Fig. 2). It takes 3 days

(half an elliptic orbit) for this transfer from perigee to apogee (Fig. 2. (b)). This apogee can be designed to be the closest encounter point with respect to Earth of the center of mass of the 3 S/C traced back in time geodetically of the 2-degree-behind-the-Earth formation. From here, a Δv_1's of about 1.6 km/s are needed to allow the 3 S/C to enter their respective (pre-) science orbits (Fig. 2. (c)) when the calibration, commissioning and various science operations can be started. For example, it is the AMIGO-S-2-4deg orbit for AMIGO and the time epoch to reach the 2-degree-behind-the-Earth point is 2030-Jan-1st 12:00:00.[48] (AMIGO-Earth-like solar orbits with the AMIGO formation varying between 2 and 4 degrees behind the Earth orbit starting at epoch JD2462503.0 (2030-Jan-1st 12:00:00) in J2000 mean-equatorial solar-system-barycentric coordinate system.) There are similar orbit configurations for other time epochs and for ahead of the Earth situations. For ahead of the Earth situations, the apogees of the Hohmann orbits need to be on the other side of the Earth.

In the following, we summarize AEDGE,[9] AMIGO,[1] and DECIGO and B-DECIGO[10] in our session.

5.1. *AEDGE*[9,25]

Atomic Experiment for Dark Matter and Gravity Exploration (AEDGE) is a concept for a space experiment using cold atoms to search for ultra-light dark matter, and to detect gravitational waves in the mid-frequency band. It will also complement other planned searches for dark matter, and exploit synergies with other gravitational wave detectors. Examples are given in the talk of how its gravitational-wave measurements could explore the assembly of super-massive black holes, first-order phase transitions in the early universe and cosmic strings.[9,25] AEDGE will be based upon technologies now being developed for terrestrial experiments using cold atoms, and will benefit from the space experience obtained with, e.g., LISA and cold atom experiments in microgravity.

5.2. *AMIGO*

In 2017, a middle-frequency GW mission AMIGO (Astrodynamical Middle-frequency Interferometric GW Observatory) with arm length 10000 km was proposed to have significant sensitivity in this frequency band 0.1 Hz-10 Hz to bridge the gap between the ground detection sensitivity and the LISA sensitivity and yet to be a first-generation candidate for space GW missions.[51] The basic mission concept is in Ref. [51]. Stellar-size black hole inspirals[35] such as the ones like GW150914 will be sure sources of AMIGO. The AMIGO observation of the inspiral phase of these sources will enable us to predict the coalescence time precisely for the ground multi-messenger observations.

In [48], first studies on possible schemes of implementation of AMIGO were presented. Both the solar orbit and earth orbit options together with deployment

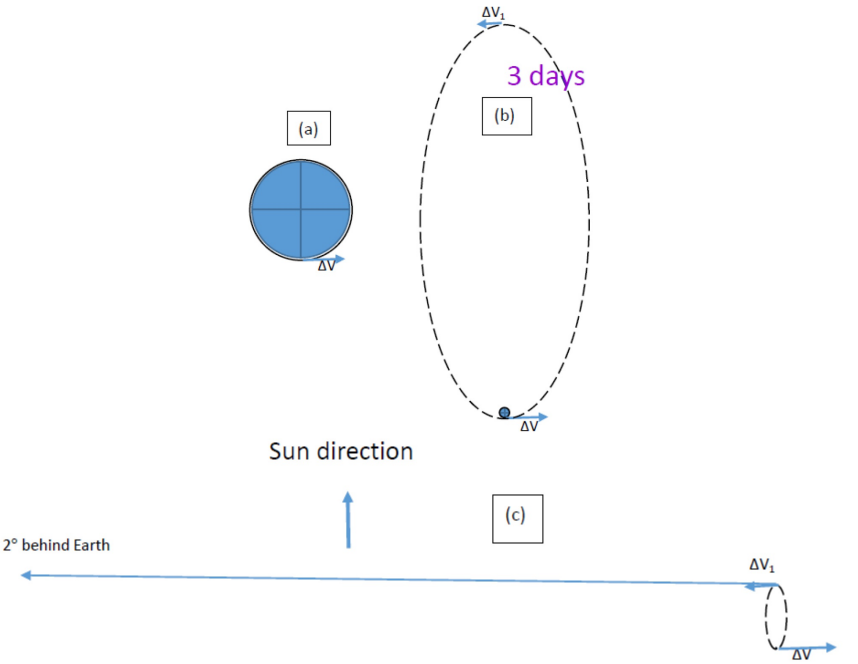


Fig. 2. An option for taking less than a week from launch to reaching the (pre-)science orbit for AMIGO-S-2-4 deg. (a) Launch to 300 km LEO (Low Earth Orbit) parking orbit. (b) A last-stage launch from 300 km LEO to an eccentric Hohmann orbit with apogee 262931 km (The period of this Hohmann transfer orbit is about 6 days.); It takes 3 days (half an elliptic orbit) for this transfer from perigee to apogee. (c) This apogee is designed to be the closest encounter point with respect to Earth of the center of mass of the 3 S/C traced back in time geodetically of the 2-degree-behind-the-Earth AMIGO formation]. From here, Δv_1's of about 1.6 km/s are needed to allow the 3 S/C to enter their respective (pre-)science orbits reaching 2-degree-behind-the-Earth at 2030-Jan-1st 12:00:00.[28,49,50]

strategy were discussed. The first-generation TDIs (Time Delay Interferometry's) for all orbit options studied were calculated and found that all the heliocentric orbit options satisfy the frequency-noise suppression requirement, but the geocentric orbit options do not satisfy the requirement. From this study, the heliocentric option is preferred. The issue on feasibility of constant equal-arm implementation was studied. For the solar-orbit options, the acceleration to maintain a constant equal-arm formation can be designed to be less than 15 nm/s^2 with the thruster requirement in the 15 μN range.[46,48] AMIGO would be a good place to test and implement the constant equal-arm option. Fuel requirement, thruster noise requirement and test mass acceleration actuation requirement were considered.

In [28], the core noise requirements on position noises and acceleration noises were updated. From these design white position noises and acceleration noises, the GW sensitivities (Fig. 3) were obtained for the first-generation Michelson X TDI configuration of b-AMIGO (baseline AMIGO), AMIGO, and e-AMIGO (enhanced

AMIGO). In Fig. 3, the sensitivities for the corresponding AMIGO-5's (50,000 km nominal arm length) are also shown. Steps to implement the AMIGO mission concept were indicated in view of the current technology development.

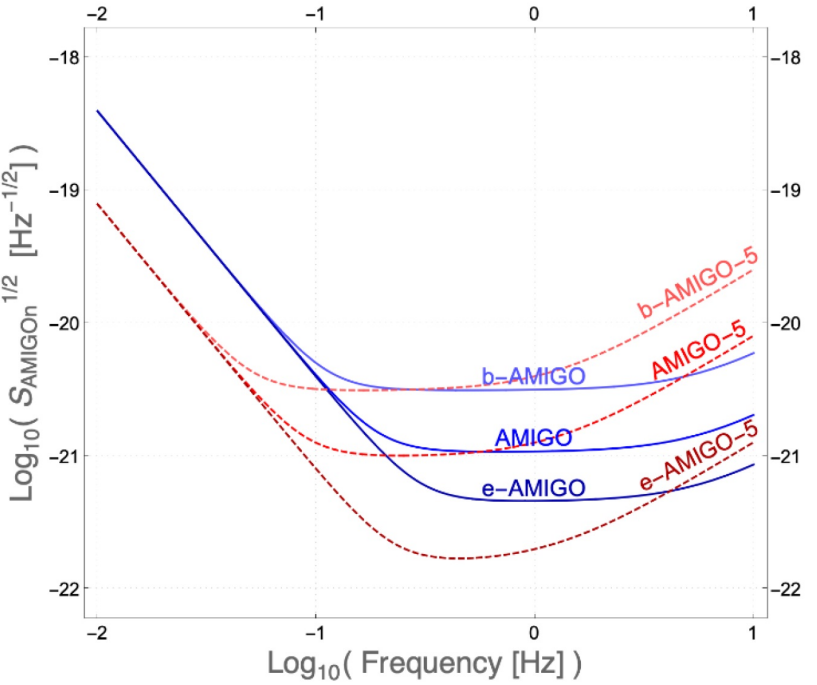


Fig. 3. Strain power spectral density (psd) amplitude vs. frequency for various AMIGO-5 proposals as compared to AMIGO proposals. The solid lines are for AMIGO's (10,000 km nominal arm length); the dashed lines are for AMIGO-5's (50,000 km nominal arm length).[28]

5.3. *DECIGO and B-DECIGO*[10]

DECIGO was conceived in 2001 with the main scientific target to directly measure the acceleration of the Universe. The design has been elaborated with laser Fabry-Perot arm cavities. DECIGO consists of four clusters with each cluster having three differential Fabry-Perot interferometers with three drag-free spacecraft forming an equilateral triangle of nominal arm cavity length 1000 km. Lasers illuminating the cavity have power 10 W and wavelength 515 nm. The current design has main mirrors with diameter 1 m and mass 100 kg forming cavities of finesse 10. The four clusters also form a nearly equilateral triangle in the heliocentric orbits with two clusters in Earth-trailing orbits. The other two clusters are separated from each other and from the near-Earth clusters by 120 degrees in the heliocentric orbits. The DECIGO team recognized that "... analysis based on the observations

by the Planck satellite lowered the upper limit of GW to 1×10^{-16} As a result, the DECIGO sensitivity became no longer good enough to detect the primordial gravitational waves". Nevertheless, with target sensitivity the best among all the present mid-frequency concepts (Table 1), it may still bring us some clues on the very early universe, e.g. revealing the thermal history after the inflation[52,53] etc.

The design of B-DECIGO is down-scaled from DECIGO, consisting of one cluster with three spacecraft separated from each other by 100 km. The arm cavity mirrors have diameter of 0.3 m with mass of 30 kg, and cavity finesse of 100. The laser wavelength is 515 nm and laser power 1 W. With target strain psd amplitude about one order higher than DECIGO, it serves as a pathfinder for DECIGO, and with the noise sensitivity well enough, it is also a good probe to the many mid-frequency GW sources discussed in section 2. DECIGO team aim to launch B-DECIGO at the earliest in 2032.

6. Discussion and Outlook

Middle frequency band (0.1-10 Hz) is important for GW detection to bridge the gap between high-frequency GW detection and low-frequency GW detection for various scientific goals. We have seen various proposals. It is hopeful to have some mid-frequency GW detectors in the 2030s to do multiband astronomy and multi-messenger astronomy.

Acknowledgements

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under grant No. XDB21010100, and by the National Key Research and Development Program of China under Grant No. 2016YFA0302002.

References

1. W.-T. Ni, Outlook of the Mid-frequency GW Detection and AMIGO, GW2 Parallel session talk of the MG16 Meeting (2021).
2. J. Fier, X. Fang, B. Li, S. Mukohyama, A. Wang, and T. Zhu, Generation, propagation and detection of GWs in inhomogeneous universe, GW2 Parallel session talk of the MG16 Meeting (2021).
3. H. Yu and Y. Chen, Direct Determination of Supermassive Black Hole Properties with Gravitational-Wave Radiation from Surrounding Stellar-Mass Black Hole Binaries, GW2 Parallel session talk of the MG16 Meeting (2021).
4. K. Jani, Fundamental Cosmology & Multi-band, Multi-messenger Astrophysics from the Moon, GW2 Parallel session talk of the MG16 Meeting (2021).
5. J. Harms, Lunar Gravitational-Wave Antenna, GW2 Parallel session talk of the MG16 Meeting (2021).
6. F. Badaracco, J. van Heijningen, E. Camilo Ferreira, and A. Perali, A cryogenic & superconducting inertial sensor for the Lunar Gravitational Wave Antenna and for selenphysics, GW2 Parallel session talk of the MG16 Meeting (2021), Proceedings of MG16, accepted.

7. M. Ando, S. Takano, C. P. Ooi, Y. Oshima, and Y. Michimura, TOBA: A Ground-Based Mid.-Frequency Gravitational-Wave Antenna, GW2 Parallel session talk of the MG16 Meeting (2021).
8. D. Gao on behalf of the ZAIGA collaboration, Recent progress on ZAIGA, GW2 Parallel session talk of the MG16 Meeting (2021).
9. M. Lewicki, AION and AEDGE: Gravitational physics with atom interferometry, GW2 Parallel session talk of the MG16 Meeting (2021).
10. S. Kawamura and DECIGO working group, Space gravitational wave antenna DECIGO and B-DECIGO, GW2 Parallel session talk of the MG16 Meeting (2021), Proceedings of MG16, accepted.
11. K. Kuroda, W.-T. Ni and W.-P. Pan, Gravitational waves: Classification, methods of detection, sensitivities, and sources, Int. J. Mod. Phys. D 24, 1530031 (2015); also in One Hundred Years of General Relativity: From Genesis and Empirical Foundations to Gravitational Waves, Cosmology and Quantum Gravity, Chap. 10, ed. W.-T. Ni (World Scientific, Singapore, 2017).
12. ET sensitivities, http://www.et-gw.eu/index.php/etsensitivities.
13. M. Evans *et al.*, A Horizon Study for Cosmic Explorer: Science, Observatories, and Community, CE-P2100003-v7, arXiv:2109.09882 [astro-ph.IM].
14. LISA mission sensitivities, https://www.elisascience.org/articles/lisa-mission/first-gravitational-wave-observatory-space/sensitivity.
15. W.-H. Ruan, Z.-K. Guo, R.-G. Cai, and Y.-Z. Zhang, Taiji program: Gravitational-wave sources, *Int. J. Mod. Phys. A* **35**, 2050075 (2020); arXiv:1807.09495.
16. J. Luo, L.-S. Chen, H.-Z. Duan *et al.*, TianQin: A space-borne gravitational wave detector, *Class. Quantum Grav.* **33**, 035010 (2016).
17. H. J. Paik, M. V. Moody and R. S. Norton, SOGRO – Terrrestrial full-tensor detector for mid-frequency gravitational waves, *Int. J. Mod. Phys. D* **29**, 1940001 (2020).
18. T. Shimoda, S.Takano, C. P. Ooi, N. Aritomi, A. Shoda, Y. Michimura, and M. Ando, Torsion-bar antenna: A ground-based mid-frequency and low-frequency gravitational wave detector, *Int. J. Mod. Phys. D* **29**, 1940003 (2020), arXiv:1812.01835.
19. B. Canuel *et al.*, ELGAR – a European Laboratory for Gravitation and Atom-interferometric Research, *Class. Quantum Grav.* **37**, 225017 (2020).
    B. Canuel *et al.*, Exploring gravity with the MIGA large scale atom interferometer, *Sci. Rep.* **8**, 14064 (2018).
20. MAGIS-100 collaboration, Matter-wave Atomic Gradiometer Interferometric Sensor (MAGIS-100), *Quantum Sci. Technol.* **6**, 044003 (2021), arXiv:2104.02835.
21. L. Badurina, E. Bentine, D. Blas, K. Bongs, D. Bortoletto, AION: An Atom Interferometer Observatory and Network, *JCAP* **05**, 011 (2020), arXiv:1911.11755v1 [astro-ph.CO].
22. M.-S. Zhan, J. Wang, W.-T. Ni *et al.*, ZAIGA: Zhaoshan Long-baseline Atom Interferometer Gravitation Antenna, *Int. J. Mod. Phys. D* **29**, 1940005 (2020), arXiv:1903.09288 [physics.atom-ph].
23. J. Harms *et al.*, Lunar Gravitational-wave Antenna, *Astrophys. J.*, **910**, 1 (2021).
24. K. Jani and A. Loeb, Gravitational-Wave Lunar Observatory for Cosmology, arXiv:2007.08550.
25. Y. A. El-Neaj *et al.*, AEDGE: Atomic Experiment for Dark Matter and Gravity Exploration in Space, *EPJ Quantum Technology* **7**, 6 (2020), arXiv:1908.00802.
26. D. Gao, J. Wang, and M. Zhan, Atomic Interferometric Gravitational-Wave Space Observatory (AIGSO), *Commun. Theor. Phys.* **69**, 37 (2018).
27. T. Ebisuzaki, H. Katori, J. Makino, A. Noda, H. Shinkai, T. Tamagawa, INO: Interplanetary network of optical lattice clocks, *Int. J. Mod. Phys. D* **29**, 1940002 (2020), arXiv:1809.10317.

28. W.-T. Ni, Core noise requirements, and GW sensitivities of AMIGO, arXiv:2106.12432 [gr-qc; astro-ph.IM].
29. K. A. Kuns, H. Yu, Y. Chen, and R. X. Adhikari, Astrophysics and cosmology with a deci-hertz gravitational-wave detector: TianGO, *Phys. Rev. D* **102**, 043001 (2020), arXiv:1908.06004 [gr-qc, astro-ph.CO, astro-ph.HE].
30. Manuel Arca Sedda *et al.*, The missing link in gravitational-wave astronomy, *Experimental Astronomy* **51**, 1427 (2021).
31. J. Harms, B. J. J. Slagmolen, R. X. Adhikari *et al.*, Low-frequency terrestrial gravitational-wave detectors, *Phys. Rev. D* **88**, 122003 (2013).
32. J. Harms, Terrestrial gravity fluctuations, *Living Rev. Relativity* **18**, 3 (2015) (DOI 10.1007/lrr-2015-3).
33. M. Armano, H. Audley, G. Auger *et al.*, Sub-Femto-g Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results, *Phys. Rev. Lett.* **116**, 231101 (2016).
34. M. Armano, H. Audley and J. Baird *et al.*, Beyond the Required LISA Free-Fall Performance: New LISA Pathfinder Results down to 20 μHz, *Phys. Rev. Lett.* **120**, 061101 (2018).
35. A. Sesana, Prospects for multiband gravitational-wave astronomy after GW150914, *Phys. Rev. Lett.* **116**, 231102 (2016).
36. X. Chen and W.-B Han, Extreme-mass-ratio inspirals produced by tidal capture of binary black holes, *Communications Physics* **1**, 53 (2018) and references therein.
37. H. Yu and Y. Chen, Direct Determination of Supermassive Black Hole Properties with Gravitational-Wave Radiation from Surrounding Stellar-Mass Black Hole Binaries, *Phys. Rev. Lett.* **126**, 021101 (2021).
38. I. Mandel, A. Sesana, and A. Vecchio, The astrophysical science case for a decihertz gravitational-wave detector, *Class. and Quantum Grav.* **35**, 054004 (2018), and references therein.
39. Manuel Arca Sedda *et al.*, The Missing Link in Gravitational-Wave Astronomy: Discoveries waiting in the decihertz range, *Class. Quantum Grav.* **37**, 215011 (2020), arXiv:1908.11375.
40. J. Fier, X. Fang, B. Li, S. Mukohyama, A. Wang, and T. Zhu, Gravitational wave cosmology I: High frequency approximation, *Phys. Rev. D* **103**, 123021 (2021).
41. L. Zhou *et al.*, Joint mass-and-energy test of the equivalence principle at the 10^{-10} level using atoms with specified mass and internal energy, *Phys. Rev. A* **104**, 022822 (2021).
42. Z.-W. Yao *et al.*, Self-alignment of a large-area dual-atom-interferometer gyroscope using parameter-decoupled phase-seeking calibrations, *Phys. Rev. A* **103**, 023319 (2021).
43. W. Zhao, X. Mei, D. Gao, J. Wang, and M.-S. Zhan, Ultralight scalar dark matter detection with ZAIGA, arXiv: 2110.11564.
44. P. Touboul, G. Métris, M. Rodrigues, Y. Andréc and A. Robert, The MICROSCOPE space mission to test the Equivalence Principle, *Comptes Rendus Physique*, **21**, 139 (2020), and references therein.
45. G. Wang, D. Gao, W.-T. Ni, J. Wang and M.-S. Zhan, Orbit design for space atom-interferometer AIGSO, *Int. J. Mod. Phys. D* **29**, 1940004 (2020).
46. G. Wang, W.-T. Ni, and A.-M. Wu, Orbit design and thruster requirement for various constant-arm space mission concepts for gravitational-wave observation, *Int. J. Mod. Phys. D* **29**, 1940006 (2020), arXiv:1908.05444 [gr-qc].
47. M. Armano, H. Audley, J. Baird *et al.*, LISA Pathfinder Performance Confirmed in an Open-Loop Configuration: Results from the Free Fall Actuation Mode, *Phys. Rev. Lett.* **123**, 111101 (2019).

48. W.-T. Ni, G. Wang, and A.-M. Wu, Astrodynamical middle-frequency interferometric gravitational wave observatory AMIGO: Mission concept and orbit design, *Int. J. Mod. Phys. D* **29**, (2020) 1940007, arXiv:1909.04995 [gr-qc].
49. A.-M. Wu, W.-T. Ni and G. Wang, Deployment Simulation for LISA Gravitational Wave Mission, IAC-17-A2.1.4, 68th International Astronautical Congress, 25-29 September 2017, Adelaide, Australia (2017).
50. A.-M. Wu *et al.*, Deployment of AMIGO, paper in preparation.
51. W.-T. Ni, Gravitational Wave (GW) Classification, Space GW Detection Sensitivities and AMIGO (Astrodynamical Middle-frequency Interferometric GW Observatory), Proceedings of Joint Meeting of 13th International Conference on Gravitation, Astrophysics and Cosmology, and 15th Italian-Korean Symposium on Relativistic Astrophysics, Ewha Womans University, Seoul, Korea, July 3-7, 2017, *EPJ Web of Conferences* **168**, 01004 (2018); arXiv:1709.05659 [gr-qc].
52. S. Kuroyanagi, K. Nakayama, and J. Yokoyama, Prospects of determination of reheating temperature after inflation by DECIGO, *Prog. Theor. Exp. Phys.*, 013E02 (2015).
53. K. Nakayama *et al.*, Space-based gravitational-wave detectors can determine the thermal history of the early Universe, *Phys. Rev. D* **77**, 124001 (2008).

Salient features of the optimised PyCBC IMBH search

Koustav Chandra and Archana Pai
Department of Physics, Indian Institute of Technology Bombay, Powai, Mumbai 400 076, India

V. Villa-Ortega and T. Dent
IGFAE, Campus Sur, Universidade de Santiago de Compostela, 15782 Spain

C. McIsaac, I. W. Harry and G. S. Cabourn Davies
University of Portsmouth, Portsmouth, PO1 3FX, United Kingdom

K. Soni
Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India

Matched-filter based PyCBC searches have successfully detected $\sim \mathcal{O}(50)$ compact binary merger signals in the LIGO-Virgo data. While most PyCBC searches have been designed to detect stellar-mass binaries, we present here a new search that is optimised to detect short-duration gravitational wave (GW) signals emitted by intermediate-mass black hole mergers. When applied to the first half of the third observation run data, the optimised search re-identified the intermediate mass black hole (IMBH) binary event, GW190521, with a false alarm rate of 1 in 727 yrs, significantly lower than the previous PyCBC search result of 1 in 0.94 yr. Analysis of simulated signals from IMBH binaries with generically spinning components shows an increase in sensitivity by a factor of 1.2 to 3 over previous PyCBC searches.

Keywords: Gravitational Waves, Matched-Filtering, Intermediate-Mass Black Holes

1. Introduction

Coalescing compact binaries are amongst the primary sources of gravitational wave (GW) with frequencies accessible to current second-generation ground-based GW interferometers. Detection of $\sim \mathcal{O}(50)$ stellar-mass binary mergers (the majority being the binary black holes) has already ushered us into the age of observational strong-field gravity. They have so far given us unique insights into the black hole population that we hardly knew existed, with one of them being the first confirmed detection of a massive black hole system GW190521 with a remnant IMBH.[1,2]

GW190521 was observed in the first half of the third observing run of the Advanced LIGO and Advanced Virgo detectors. The compact binary merger event lasted for ~ 0.1 s in the detector band and had barely any observable pre-merger phase of the signal, clearly reflecting the binary's high total mass. The detailed Bayesian parameter estimation showed the merger event to be consistent with the merger of two black holes in a mildly precessing orbit, with component masses of $85^{+21}_{-14}\,M_\odot$ and $66^{+17}_{-18}\,M_\odot$ and a remnant black hole of $142^{+28}_{-16}\,M_\odot$ falling in the mass

range of intermediate-mass black holes.[1] The merger event producing this remnant was detected by a search sensitive to generic transient at high confidence with a false alarm rate of $< 1/4900$ years.[3]

Traditionally, the unmodeled GW transient search for short-duration GW signals is performed using the coherent waveburst (cWB) search[4] which is built on the rationale that a GW signal from a compact binary coalescence (CBC), chirp or not, is expected to be a transient waveform and it will produce some localised excess of energy if projected in the time-frequency plane. Therefore, identifying such an excess coherently across two or more detectors should give a strong indication of a GW event. Together with this, model-based matched-filter searches, such as PyCBC[5–9] and GstLAL,[10] targeting generic binaries have also been used for the IMBH binary searches in the LIGO-Virgo data (ADD O1-O2 and O3 paper link).

While unmodeled cWB searches are tuned to look for short-duration signals,[11] the standard matched-filter-based searches were not. The lack of *optimisation* towards IMBH binary signals made them suffer from the "look-elsewhere" effect. Further, improvements in the sensitivity of the detectors at low frequencies[12–14] before the beginning of the third observing run of Advanced LIGO and Advanced Virgo (O3) motivated us to adapt our matched-filter algorithms to the target search space. Here, we will summarise the salient features of one of the optimised templated searches, namely PyCBC-IMBH that was developed and deployed for the IMBH binary search. This optimised search was used together with other search pipelines to look for IMBH binaries in the O3 run, and the reader can find the results of it in.[15] For complete information on the algorithm itself, we direct the reader to the optimised PyCBC IMBH search paper.[16]

2. The Optimised PyCBC-IMBH Search

Below we summarise the salient features of the optimised PyCBC-IMBH search.[16]

Data cleaning: The gravitational wave strain data is often plagued with loud background noise, which can significantly hamper the search sensitivity.[17] Thus, we begin our analysis by windowing out very high amplitude excursions ($> 50\sigma$ deviation from Gaussian noise) in the whitened data.[6] Since empirically, these loud noise transients tend to be correlated with neighbouring quieter noise transients, we also *discard* any LIGO triggers which are either 1s before or 2.5s after the centre of the gating window.[16]

Template bank: After this data cleaning process, we use matched-filtering to detect the target signals hidden in the detector noise.[5,18,19] We assume that our target signals are well-modelled and deterministic, and to a good approximation, we can consider them to be buried in wide-sense stationary Gaussian noise. However, the binary's parameters are not known a priori, forcing us to use a discrete bank of filter waveforms, aka templates. Our bank of templates is designed to intercept the dominant quadrupolar harmonics of a non-precessing binary.[20] Specifically,

we use the reduced-order representation of the spinning effective one-body model, SEOBNRv4,[21,22] for our templates, and we *constrain* ourselves to binaries with redshifted (detector frame) total mass between $100 - 600 M_{\odot}$.[16] Additionally, our templates have a minimum component mass of $40 M_{\odot}$ and a mass ratio between 1/1 and 1/10. We do not use relatively lower mass ratios in the templates as binaries with low mass ratios emit gravitational wave signals with significantly higher subdominant harmonics that are not modelled by our waveform approximant. We also don't include any templates with a duration less than 0.07 s., measured from a fixed starting frequency of 15 Hz, to reduce the false alarms arising from short-duration noisy transients. Since information on the spins of IMBH binaries is lacking, we allow our templates to have dimensionless spin components along the orbital angular momentum to be between ± 0.998. The reader can find additional details about the template bank in.[16]

Although our search targets non-precessing binaries, we expect our search to retain its sensitivity towards moderately precessing binaries, mainly because our target systems are relatively short-lived within the detector bandwidth.[23,24] Furthermore, we also expect our search to be equally sensitive to binaries with low eccentricity as the emission of gravitational waves will circularise the binary system before the merger, and because of the shortness of our templates and the detector bandwidth, we will only be able to intercept the last few cycles of IMBH binary's inspiral.[25]

Time-phase consistency between detectors: Single-detector triggers are identified based on peaks in the matched filter signal-to-noise ratio (SNR) time series. Unfortunately, our detector data is polluted with several non-stationary instrumental artefacts that are not loud enough to be gated out during data pre-processing. Technically, this makes our search a detection problem in non-Gaussian noise. Having a network of simultaneously operating similar sensitivity detectors help a lot as we can demand that the time difference between triggers at different detectors is consistent with an astrophysical signal, i.e. within the light-travel time up to expected uncertainty. Besides, if the trigger is due to a GW signal, then it must share consistent phase and amplitude characteristics. Hence the *coincidence test* significantly reduces the number of noisy artefacts that is picked up by the search.

Signal consistency tests: But there can be chance coincidences between noise triggers. So, before the coincidence test is performed, there is a need to differentiate a signal from a glitch at the single detector level. We do this in multiple stages. We begin by only considering the loudest trigger within a pre-determined window of time to reduce false alarms, and we subject these triggers to two different signal-consistency tests: the χ^2_r-test[26] and the $\chi^2_{r,sg}$-test.[27] The former checks if the morphology of the triggers matches with that of the best-matched template, while the latter checks for any excess power beyond the maximum frequency of the template. The output of these two tests is used to amend the trigger SNR to reduce the apparent loudness of a noise trigger as the target signal will return values closer

to unity. We also account for any short-term non-stationarity of the detector noise that can lead to wrongful estimation of the trigger SNR.[28–30]

Given our target space, our search is susceptible to short but loud noise triggers motivating us to use stricter discriminators. We chose to discard any trigger with $\chi^2_r > 10$ or with short-term power spectral density (PSD) measure of over ten times the expectation from stationary noise.[16]

Event significance and ranking: We subject the surviving single-detector triggers to the coincidence test and assign each *multi-detector candidate* a rank based on their response to the coincidence test.[8,9] Our candidate events can be coincidentally observed either in one of the possible two detector combinations or in all three detectors. This is because we analyse all possible two and three detector combinations. Finally, a statistical significance is assigned to each of these candidates by comparing its rank against simulated background noise that is generated by finding fictitious coincidences between detector data.[6] We achieve this by shifting the pivot detector(s) data with respect to a fixed reference detector by a time that is considerably larger than the light-travel time between the detectors. By repeating this procedure, we produce more than 10^4 years worth of noise background.

In the next section, we provide details of how the PyCBC algorithm benefits from the optimisation made.

3. Summary of Sensitivity Comparison

To quantify the benefit of our restricted analysis, we compared the sensitivity of our optimised search against existing PyCBC-based searches with overlapping parameter space.[16] We compared the PyCBC-IMBH search against the PyCBC-broad and the PyCBC-BBH search. The former looks for binaries consisting of either neutron stars or black holes or both, while the latter is a "focused" search for binary black hole (BBH) covering a restricted range of masses. We resort to an injection campaign for the comparison. This involved adding simulated signals to real GW data from the first half of the third observing run (O3a) run as part of the analysis without any physical actuation and then estimate the sensitivity of each of the searches in terms of sensitive volume time $\langle VT \rangle$.[31,32] Our simulated signals mimicked those from generically spinning BBH with detector frame total mass between $100 - 600 M_\odot$ and having a mass ratio between $1/1 - 1/10$. The signals themselves were either the dominant quadrupolar mode or the full symphony emitted by such binaries. We found that the PyCBC-IMBH search is $\sim 1.2 - 3$ times more sensitive than the PyCBC-broad search and is $\sim 1.1 - 12.6$ times more sensitive than the PyCBC-BBH search at a false alarm rate threshold of 1 in 100 years.

We also compare our search against two other searches, namely cWB and GstLAL-IMBH, that were used by the LIGO Scientific, Virgo & Kagra collaboration (LVK) to look for IMBH binaries in O3 run.[15] For this purpose, we utilise the results of the injection campaign that were used by the LVK to place upper limits on the merger rate density for a discrete set of IMBH binaries. The dataset is publicly

Table 1. Summary of the single pipeline sensitive volume-time $\langle VT \rangle$. Here, M_T is the total source frame mass of the binary and $q = m_2/m_1 \leq 1$ denotes the mass ratio of the binary. $\chi_{\rm eff}$ is the mass-weighted combination of the spins parallel to the orbital angular momentum $\vec{L}$ while $\chi_{\rm p}$ captures the average amount of precession exhibited by a generically precessing system over many cycles defined at reference frequency of 11 Hz during the inspiral.

M_T ($M_\odot$)	q	$\chi_{\rm eff}$	$\chi_{\rm p}$	$\langle VT \rangle_{\rm PyCBC}$[Gpc^{3}yr]	$\langle VT \rangle_{\rm cWB}$[Gpc^{3}yr]	$\langle VT \rangle_{\rm GstLAL}$ [Gpc^{3}yr]
120	1/2	0.00	0.00	11.51	8.93	8.18
120	1/4	0.00	0.00	4.64	3.54	3.04
120	1/5	0.00	0.00	3.11	2.42	2.06
120	1/7	0.00	0.00	1.67	1.24	1.05
120	1/10	0.00	0.00	0.82	0.60	0.51
150	1/2	0.00	0.00	11.95	9.96	7.40
200	1	0.00	0.00	14.84	12.76	10.27
200	1/2	0.00	0.00	10.35	9.47	6.56
200	1/4	0.00	0.00	3.90	3.69	1.93
200	1/7	0.00	0.00	1.34	1.29	0.53
220	1/10	0.00	0.00	0.57	0.60	0.22
250	1/4	0.00	0.00	2.99	2.83	1.80
300	1/2	0.00	0.00	6.27	5.86	5.38
350	1/6	0.00	0.00	0.72	0.75	0.44
400	1	0.00	0.00	4.82	4.55	4.29
400	1/2	0.00	0.00	3.08	3.33	2.68
400	1/3	0.00	0.00	1.68	1.94	1.29
400	1/4	0.00	0.00	1.11	1.22	0.70
400	1/7	0.00	0.00	0.44	0.47	0.20
440	1/10	0.00	0.00	0.20	0.20	0.09
500	2/3	0.00	0.00	1.87	1.97	1.45
600	1	0.00	0.00	0.90	0.79	0.50
600	1/2	0.00	0.00	0.65	0.81	0.34
800	1	0.00	0.00	0.15	0.14	0.11
200	1	0.80	0.00	38.54	29.60	30.99
400	1	0.80	0.00	18.08	15.88	16.07
600	1	0.80	0.00	4.97	5.35	4.03
800	1	0.80	0.00	0.89	0.88	0.37
200	1	-0.80	0.00	10.07	9.28	6.39
400	1	-0.80	0.00	1.90	1.86	1.66
600	1	-0.80	0.00	0.22	0.22	0.11
800	1	-0.80	0.00	0.04	0.05	0.03
200	1	0.51	0.42	26.38	20.88	19.42
200	1/2	0.14	0.42	13.97	12.36	9.01
200	1/4	0.26	0.42	7.67	7.65	4.13
200	1/7	0.32	0.42	3.35	3.59	1.44
400	1	0.51	0.42	10.53	9.38	9.19
400	1/2	0.14	0.42	4.96	5.42	4.24
400	1/4	0.26	0.42	2.58	3.63	1.88
400	1/7	0.32	0.42	1.11	1.79	0.65
600	1	0.51	0.42	2.44	2.49	1.72
600	1/2	0.14	0.42	1.12	1.45	0.58
800	1	0.51	0.42	0.10	0.18	0.01

available at `https://dcc.ligo.org/LIGO-P2100179/public`, and it contains details about the parameters of each of the injections and also their associated p-value as reported by the contributing searches. The signals themselves were numerically simulated and are designed to imitate GWs from a representative population of IMBH binaries with generic spins. Additional details about the injections can be found in.[15]

To estimate $\langle VT \rangle$ using the publicly available dataset, we start off by calculating the number of signals that are recovered by each of three contributing searches. We do this by considering only those injections whose p-value is less than the minimum p-value of the loudest noise trigger, namely that of 200214_224526 in the.[15] This trigger has a minimum p-value of 9.2×10^{-2} and was only reported by the weakly model-dependent cWB search. The $\langle VT \rangle$ is then given by

$$\langle VT \rangle = \frac{N_{\rm rec}}{N_{\rm total}} \langle VT \rangle_{\rm total} \tag{1}$$

where the first term denotes the fraction of injections that are recovered below the p-value threshold while the second term denotes the total spacetime volume covered by a given injection set.

We tabulate the results of this analysis for each of the three pipelines for each simulated IMBH binary in Table 1. Amplitude and phase errors arising from detector calibration have not been included in the analysis, but a statistical uncertainty of 4%-7% is expected. We find that the PyCBC-IMBH search has higher sensitivity as compared to the GstLAL-IMBH for all of the mass bins and is comparably sensitive (if not better for most of the symmetric and/or relatively low total mass bins) to cWB search.

The latter can be understood as follows. As discussed in Sec. 2, the PyCBC-IMBH search uses only the dominant multipole templates in matched filtering. Hence, the sensitivity drops with a decrease in mass ratio as compared to the cWB search. This is because a decrement in mass ratio corresponds to an increase in the contribution coming from the sub-dominant harmonics. The cWB search being a weakly modelled search, is oblivious to this effect and hence has higher sensitivity than template-based searches. Furthermore, if one compares the sensitivity of a given search for a given total mass but decreasing mass ratio, one will observe that the search performance degrades owing to an overall decrease in the intrinsic luminosity of the system involved.

Concerning the total mass of a binary, the PyCBC search tends to have higher sensitivity for relatively lower total mass as signals from such systems tend to have a longer inspiral within the detector bandwidth, thus allowing for a larger SNR. At higher total mass, the cWB search has higher sensitivity because of its efficient signal-noise discriminators, which allow it to distinguish the signal from noisy transients better, thus allowing for a relatively lower p-value and hence greater sensitivity. This also explains why the PyCBC search has a poorer performance for a system with a given mass distribution but having a relatively anti-aligned spin.

GW signals from a significantly precessing binary can have a notably suppressed inspiral right before the merger, causing a characteristic sine-Gaussian morphology. The search summarised here uses non-precessing quadrupolar templates meaning the SNR recovered would be lower and the χ^2_r higher owing to mismatch with templates.

4. Conclusion

Here, we summarise the salient features of the optimisation made to the existing PyCBC search to target short-duration GWs from massive BBHs. This optimisation enabled us to improve the search performance and re-detect the event GW190521 with improved statistical significance. We also compare the sensitivity of PyCBC-IMBH with other searches used for the detection of IMBH binaries using the public data.

Detection of IMBH and its subsequent studies has the potential to provide valuable insight into the BBH population. Hence searches like this are going to be increasingly important. While we are primarily concerned here with detection of IMBH binaries, this type of search can be used to detect highly red-shifted BBHs which we expect to see in the next few years with the advent of third-generation GW observatories like the Einstein Telescope and Cosmic Explorer. Thus, the future of this field is quite exciting!

Acknowledgments

KC is extremely grateful to the organisers of the conference for hosting such a wonderful meeting in challenging circumstances. We thank Gayathri V. and Keith Riles for their comments on an earlier draft of this proceeding. The authors are grateful for the computational resources and data provided by the LIGO Laboratory which is funded by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. The open data is available in the Gravitational Wave Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. The authors also acknowledge the use of the IUCAA LDG cluster, Sarathi, for computational/numerical work. KC acknowledges the MHRD, Government of India, for the fellowship support. AP's research is supported by SERB-Power-fellowship grant SPF/2021/000036, DST, India. VVO and TD acknowledge financial support from Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019-2022), by European Union ERDF, and by the "María de Maeztu" Units of Excellence program MDM-2016-0692 and the Spanish Research State Agency. GSCD and IWH acknowledge the STFC for funding through grant ST/T000333/1. CM was supported by the STFC through the DISCnet Centre for Doctoral Training. KS acknowledges the Inter-University Centre of Astronomy and Astrophysics (IUCAA), India, for the fellowship support.

We want to thank all of the essential workers who put their health at risk during this ongoing COVID-19 pandemic. Without their support, we would not have completed this work. We offer condolences to people who have lost their family members during this pandemic.

References

1. R. Abbott *et al.*, GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$, *Phys. Rev. Lett.* **125**, p. 101102 (2020).
2. R. Abbott *et al.*, Properties and Astrophysical Implications of the 150 $M_{\odot}$ Binary Black Hole Merger GW190521, *Astrophys. J. Lett.* **900**, p. L13 (2020).
3. M. Szczepańczyk *et al.*, Observing an intermediate-mass black hole GW190521 with minimal assumptions, *Phys. Rev. D* **103**, p. 082002 (2021).
4. S. Klimenko *et al.*, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, *Phys. Rev. D* **93**, p. 042004 (2016).
5. B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown and J. D. E. Creighton, FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* **85**, p. 122006 (2012).
6. S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, *Class. Quant. Grav.* **33**, p. 215004 (2016).
7. T. Dal Canton and I. W. Harry, Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run, *arXiv e-prints* (5 2017).
8. A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst and D. A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search, *Astrophys. J.* **849**, p. 118 (2017).
9. G. S. Davies, T. Dent, M. Tápai, I. Harry, C. McIsaac and A. H. Nitz, Extending the PyCBC search for gravitational waves from compact binary mergers to a global network, *Phys. Rev. D* **102**, p. 022004 (2020).
10. C. Messick *et al.*, Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data, *Phys. Rev. D* **95**, p. 042001 (2017).
11. B. P. Abbott *et al.*, Search for intermediate mass black hole binaries in the first and second observing runs of the Advanced LIGO and Virgo network, *Phys. Rev. D* **100**, p. 064064 (2019).
12. D. Davis *et al.*, LIGO detector characterization in the second and third observing runs, *Class. Quant. Grav.* **38**, p. 135014 (2021).
13. R. P. Fisher, G. Hemming, M.-A. Bizouard, D. A. Brown, P. F. Couvares, F. Robinet and D. Verkindt, Dqsegdb: A time-interval database for storing gravitational wave observatory metadata, *SoftwareX* **14**, p. 100677 (2021).
14. B. P. Abbott *et al.*, A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, *Class. Quant. Grav.* **37**, p. 055002 (2020).
15. The LIGO Scientific Collaboration, the Virgo Collaboration and the KAGRA Collaboration, Search for intermediate mass black hole binaries in the third observing run of Advanced LIGO and Advanced Virgo, *arXiv e-prints*, p. arXiv:2105.15120 (May 2021).
16. K. Chandra, V. Villa-Ortega, T. Dent, C. McIsaac, A. Pai, I. W. Harry, G. S. C. Davies and K. Soni, An optimized PyCBC search for gravitational waves from intermediate-mass black hole mergers, *Phys. Rev. D* **104**, p. 042004 (2021).
17. M. Cabero *et al.*, Blip glitches in Advanced LIGO data, *Class. Quantum Grav.* **36**, p. 155010 (2019).

18. B. Sathyaprakash and S. Dhurandhar, Choice of filters for the detection of gravitational waves from coalescing binaries, *Phys. Rev. D* **44**, 3819 (1991).
19. B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, *Phys. Rev. D* **53**, 6749 (1996).
20. P. Ajith *et al.*, Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins, *Phys. Rev. Lett.* **106**, p. 241101 (2011).
21. A. Bohé, S. Marsat and L. Blanchet, Next-to-next-to-leading order spin–orbit effects in the gravitational wave flux and orbital phasing of compact binaries, *Class. Quant. Grav.* **30**, p. 135009 (2013).
22. M. Pürrer, Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins, *Phys. Rev. D* **93**, p. 064041 (2016).
23. I. Harry, S. Privitera, A. Bohé and A. Buonanno, Searching for Gravitational Waves from Compact Binaries with Precessing Spins, *Phys. Rev. D* **94**, p. 024012 (2016).
24. K. Chandra, V. Gayathri, J. C. Bustillo and A. Pai, Numerical relativity injection analysis of signals from generically spinning intermediate mass black hole binaries in Advanced LIGO data, *Phys. Rev. D* **102**, p. 044035 (2020).
25. A. Ramos-Buades, S. Tiwari, M. Haney and S. Husa, Impact of eccentricity on the gravitational wave searches for binary black holes: High mass case, *Phys. Rev. D* **102**, p. 043005 (2020).
26. B. Allen, χ^2 time-frequency discriminator for gravitational wave detection, *Phys. Rev. D* **71**, p. 062001 (2005).
27. A. H. Nitz, Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers, *Class. Quant. Grav.* **35**, p. 035016 (2018).
28. T. Venumadhav, B. Zackay, J. Roulet, L. Dai and M. Zaldarriaga, New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo, *Phys. Rev. D* **101**, p. 083030 (2020).
29. T. Venumadhav, B. Zackay, J. Roulet, L. Dai and M. Zaldarriaga, New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of Advanced LIGO, *Phys. Rev. D* **100**, p. 023011 (2019).
30. S. Mozzon, L. Nuttall, A. Lundgren, T. Dent, S. Kumar and A. Nitz, Dynamic Normalization for Compact Binary Coalescence Searches in Non-Stationary Noise, *Class. Quant. Grav.* **37**, p. 215014 (2020).
31. R. Biswas, P. R. Brady, J. D. E. Creighton and S. Fairhurst, The Loudest event statistic: General formulation, properties and applications, *Class. Quant. Grav.* **26**, p. 175009 (2009), [Erratum: Class. Quant. Grav. 30, 079502 (2013)].
32. B. P. Abbott *et al.*, Upper Limits on the Rates of Binary Neutron Star and Neutron Star–black Hole Mergers From Advanced Ligo's First Observing run, *Astrophys. J. Lett.* **832**, p. L21 (2016).

Matter shells modifying gravitational wave signals

Monos Naidoo*, Nigel T. Bishop and Petrus J. van der Walt

Department of Mathematics, Rhodes University, Grahamstown, 6140, South Africa
**E-mail: monos.naidoo@gmail.com*
www.ru.ac.za

As detections of mergers of compact bodies begin to flow in, and as we enter an era of precision GW measurements, our understanding of compact bodies, their physics and that of the surrounding astrophysical environment, will continue to grow and at times even be challenged. The need to revise the mass bounds of compact bodies such as BHs and NSs and the possibility of the existence of GW echoes are just some of consequences of the first few years of GW detection. In previous work, using linearised perturbation theory, we made the novel finding that a dust shell will cause a GW to be modified both in magnitude and phase, but without any energy being transferred to or from the dust. We extend our analysis to matter shells surrounding compact body mergers and to intervening matter in cosmology. Instead of only monochromatic GW sources, as we used in our initial investigation, we also consider burst-like GW sources. The thin density shell approach is modified to include thick shells by considering concentric thin shells and integrating. Solutions are then found for these burst-like GW sources using Fourier transforms. In the context of cosmology, apart from the gravitational redshift, the effects are too small to be measurable. We show that GW echoes that are claimed to be present in the LIGO data of certain events, could not have been caused by a matter shell. We do find, however, that matter shells surrounding BBH mergers, BNS mergers, and CCSNe could make modifications of order a few percent to a GW signal. These modifications are expected to be measurable in GW data with current detectors if the event is close enough and at a detectable frequency; or in future detectors with increased frequency range and amplitude sensitivity.

Keywords: Gravitational waves; Gravitational wave echoes; Bondi-Sachs; Matter shell; Linearized perturbation theory.

1. Introduction

Claims have been made that echoes have been detected in the LIGO data of the binary black hole (BBH) merger GW150914,[1,2] and also from the binary neutron star (BNS) merger GW170917.[3] It has been postulated that various exotic compact objects (ECOs) may cause echoes in a gravitational wave (GW) signal and that these may be the cause of the echoes claimed to be observed. These claims have been contested,[4,5] sparking a debate and responses in defense of the claims[6,7] with further substantiations.[3,8] A number of ECO scenarios have been investigated. At the quantum level, Hawking's information paradox suggests Planck-scale modifications of black hole horizons (firewalls[9]) and other modifications to black hole structure (fuzzballs[10]). Dark matter particles have been suggested surrounding star-like objects.[11] Other postulates include stars with interiors consisting of self-repulsive, de Sitter spacetime, surrounded by a shell of ordinary matter (gravastars[12]). Then

there are Boson stars, which are macroscopic objects made up of scalar fields.[13] All these objects are compact bodies mimicking black holes, but without a horizon. One consequence of these horizonless structures is that ingoing gravitational waves produced in a merger may reflect multiple times off effective radial potential barriers. The gravitational waves may be, in effect, trapped between effective radial potential barriers causing the waves to be 'bounced' off these barriers several times with wave packets leaking out to infinity at regular times. These gravitational wave signals, 'trailing' the main (outward bound) signal are referred to as gravitational wave echoes.[14–16]

Whilst much of the discussion around gravitational wave echoes has been within the context of 'new physics'[17–23] (beginning with[14]), the possibility that the echoes may be a consequence of the astrophysical environment should also be investigated. One such scenario is the effect of a matter shell surrounding the gravitational wave source. This case was considered using numerical methods in,[24] and was also mentioned in our previous work on the effects of matter on GW propagation.[25]

Echoes from the astrophysical environment were considered in,[26] by placing a massive shell that is not infinitely thin. This was a generalisation of the approach of[27] which was based on an infinitely thin shell. As we have shown in our paper[25] the case for a thin shell can be generalised to that of a thick shell by considering several concentric thin shells and integrating[26] showed that the deviation from Schwarzschild ringdown in their astrophysical estimations were relatively small except for a large mass which indicated that for the majority of astrophysical scenarios the effect would be relatively small. However they did note that considerations of dark matter around black holes (or compact body mergers) would leave some parametric freedom for echoes as well,[24] studied both the combination of contributions of modifications of the Schwarzschild geometry near the surface, and a nonthin shell of matter surrounding the compact body/merger. They found that a massive shell at a distance could be distinguished from the purely Schwarzschild evolution of perturbations. However, for the situation of new physics near the surface of a compact object, (a wormhole in their case), the strong echoes of the surface dominate the echoes of the distant shell. Furthermore, they found that it would take an extraordinarily large mass, located sufficiently close to the wormhole, to lead to discernable changes in the main echoes of the surface and that these changes would be relatively small. The interaction of GWs with matter has also been studied in cosmology,[28,29] with the objective of using GW observations to constrain the properties of dark matter.

In previous work,[25] we showed that a thin spherical dust shell surrounding a gravitational wave source, causes the gravitational wave to be modified both in magnitude and phase, but without any energy being transferred to or from the dust. That work suggests the possibilty of GW echoes. In Sec. 2 we describe the problem considered in this scenario, the assumptions made and the key results.

The solution of[25] is for a monchromatic GW source. A general waveform may be decomposed into a sum of Fourier components, and the technical details are given

in Sec. 3. The decomposition is implemented within a Matlab script using the Fast Fourier Transform, and validation results are reported in this section.

Secs. 4 and 5 are about the astrophysics of the various scenarios considered. In[30] we investigated whether a matter shell could explain the echoes that may exist in the LIGO data of GW150914 and GW170817. It was found that the shell would need to have such a large mass that it would constitute a black hole and this is discussed in Sec. 4. Our results show that if GW echoes are confirmed in another GW event, and with a relative magnitude and delay time similar to that of GW150914 and GW170817, then a matter shell would not be a viable explanation, so strengthening the case that an ECO would have been observed.

A key factor in determining the echo properties of a matter shell is the echo delay time, which in the cases above was of order 1s. If the echo delay time is much smaller, order 1ms or smaller, then shell properties that are physically acceptable could lead to measurable effects; however, the short delay time would mean that the effect would not appear as an echo in the usual sense, but rather as a modification to the original signal. Examples of such signal modifications are given in Sec. 5.1 for a matter shell around an event like GW150914; in Sec. 5.2 for black hole quasinormal mode signal following a binary neutron star merger; and in Sec. 5.3 for the case of core collapse supernovae (CCSNe).

Our conclusions are discussed in Sec. 6.

2. Effect of a shell of matter on a GW

In[25] we considered the scenario of a thin shell of matter surrounding a gravitational wave source such as a compact binary merger, as shown schematically in Fig. 1. The spacetime around the GW source in[25] would be otherwise empty except for the surrounding shell of matter. Confining the investigation to a thin shell does not preclude the case of a thick shell. Results can easily be applied to a series of concentric thin shells and then integrated to give the effect for a thick matter shell. The EOS is taken, as a start, to be that of dust. The results show that the shell modifies the outgoing GWs in both phase and magnitude without contradicting previous results about energy transfer. The problem is set up within the Bondi-Sachs formalism for the Einstein equations with coordinates based on outgoing null hypersurfaces.[31,32] The null hypersurfaces, are labelled by the coordinate $x^0 = u$, the angular coordinates by x^A $(A = 2, 3)$ and the surface area radial coordinate by $x^1 = r$. The angular coordinates (e.g. spherical polars (θ, ϕ)) label the null ray generators of a hypersurface $u =$ constant. The Bondi-Sachs metric then describes a general spacetime, which may be written as

$$ds^2 = -\left(e^{2\beta}\left(1+\frac{W}{r}\right) - r^2 h_{AB}U^A U^B\right)du^2 - 2e^{2\beta}dudr \\ - 2r^2 h_{AB}U^B dudx^A + r^2 h_{AB}dx^A dx^B\,, \tag{1}$$

where, $h^{AB}h_{BC} = \delta^A_C$. The condition that r is a surface area coordinate implies $\det(h_{AB}) = \det(q_{AB})$, where, q_{AB} is a unit sphere metric (e.g. $d\theta^2 + \sin^2\theta d\phi^2$).

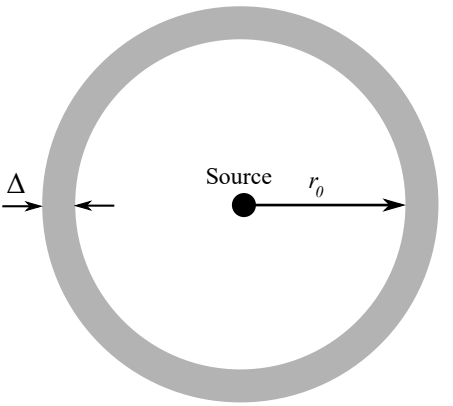


Fig. 1. The problem of a GW source which is surrounded by a spherical shell of mass M_S located between $r = r_0$ and $r = r_0 + \Delta$, where r is the distance from the source.

The GW strain far from the source is written $\mathcal{H}_{M0} = r(h_+ + ih_\times)$, where $h_+, ih_\times$ are the usual polarization modes in the TT gauge. Now suppose that, in the absence of the matter shell,

$$\mathcal{H}_{M0} = \Re(H_{M0}\exp(2i\pi f u))\,{}_2Z_{2,2}\,, \tag{2}$$

where f is the frequency (assumed to be monochromatic) of the GWs; H_{M0} is a constant determined by the physics of the GW source; and ${}_2Z_{2,2}$ is a spin-weighted spherical harmonic related to the usual ${}_sY_{\ell,m}$ as specified in.[33,34] Then, as found in,[25] the introduction of a spherical shell around the GW source of mass M_S, radius r_0 and thickness Δ modifies the wave strain to:

$$\mathcal{H} = \Re\left(H_{M0}\left(1 + \frac{2M_S}{r_0} + \frac{iM_S}{\pi r_0^2 f} + \frac{iM_S e^{-4i\pi r_0 f}}{4\pi r_0^2 f} + \mathcal{O}\left(\frac{M_S\Delta}{r_0^2}, \frac{M_S}{r_0^3 f^2}\right)\right)\exp(2i\pi f u)\right)\,{}_2Z_{2,2}\,. \tag{3}$$

Each of the terms containing M_S in Eq. (3) represents a correction to the wave strain in the absence of the shell.[25] The first correction, $2M_S/r_0$, is part of the gravitational red-shift effect, the main consequence of which is a reduction in the frequency; this effect is well-known, and henceforth we will assume that GW waveforms to be considered have allowed for this effect. The second term, $iM_S/(\pi r_0^2 f)$, is out of phase with the leading terms $1 + 2M_S/r_0$ and hence represents a phase shift of the GW. This term, to $\mathcal{O}(M_S)$, does not change the magnitude of $\mathcal{H}$ and hence has no effect on the energy of the GW. The presence of $e^{-4\pi i r_0 f}$ in the third term

describes a change in the magnitude of $\mathcal{H}$, as verified in.[25] In this context, the modified signal would then be interpreted as an echo of the main signal. The echo varies from the main signal in both magnitude and phase, with the magnitude of the echo described by

$$R = \frac{M_S}{4\pi r_0^2 f} \tag{4}$$

relative to the original signal.

3. Solution for a burst-like GW source using Fourier transforms

The time-delay of the echo is $2r_0$, but the echo's magnitude depends on the wave frequency f. The GW sources reported to date are not monochromatic but are burst-like. Such a source may be decomposed into its Fourier components and Eq. (3) applied to each component, and the echo signal obtained by summing over the transformed components. Because the magnitude of the transformation is frequency-dependent, the echo signal will have a form more complicated than simply a time-delay and magnitude change to the original signal. This effect is now analyzed.

We replace $\Re(H_{M0}\exp(2i\pi f u))$ in Eq. (2) by $h(u)$ defined in the interval $u_0 \leq u \leq u_{N-1}$; and then construct a discrete representation of $h(u)$, $h_k = h(u_k)$ ($k = 0, \cdots, N-1$), with the u_k on a regular grid, i.e. $u_{k+1} - u_k = \delta$ for $k = 0, \cdots, N-2$. Note that if the highest frequency that needs to be resolved is f_m, then N should be chosen so that $(u_{N-1} - u_0)/(N-1) < 1/(2f_m)$, i.e. to satisfy the Nyquist condition. The discrete Fourier transform[35] of $\{h_k\}$ is

$$H_n = \sum_{k=0}^{N-1} h_k \exp\left(\frac{2\pi i k n}{N}\right) \quad \text{with inverse} \quad h_k = \frac{1}{N}\sum_{n=0}^{N-1} H_n \exp\left(\frac{-2\pi i k n}{N}\right). \tag{5}$$

Then defining $H_{2,n}$ and $H_{3,n}$ to be coefficients in the transform domain of the second (phase-shift) and third (echo) terms of Eq. (3), we have

$$\begin{aligned}
&H_{2,n} = \frac{iM_S H_n}{\pi r_0^2 f_n}\,, \quad H_{3,n} = \frac{iM_S H_n \exp(-4\pi i r_0 f_n)}{4\pi r_0^2 f_n}\,, \quad n = 1, \cdots, \frac{N}{2}\,,\\
&H_{2,n} = H^*_{2,N-n}\,, \quad H_{3,n} = H^*_{3,N-n}\,, \quad n = \frac{N}{2}+1, \cdots, N-1\,,\\
&H_{2,0} = H_{3,0} = 0\,,
\end{aligned} \tag{6}$$

where * denotes the complex conjugate, and where we have used the condition that, in the time domain, all quantities are real. It is being assumed that N is even, and normally $N = 2^m$ (with m an integer) for convenience when using the fast Fourier transform; further

$$f_n = \frac{n(N-1)}{N(u_{N-1} - u_0)}\,. \tag{7}$$

Then $h_{2,k}, h_{3,k}$ are found on applying the inverse discrete Fourier transformation.

A Matlab script that implements the calculation of the previous paragraph is available as Supplementary Material in.[30] The script was checked by applying it to a monochromatic signal $h(u) = \Re(-i\exp(2\pi i f u))$. The errors $e_{2,k}$ in $h_{2,k}$ and $e_{3,k}$ in $h_{3,k}$ are

$$\begin{aligned} e_{2,k} &= \left| h_{2,k} - \Re\left(iM_S/(\pi r_0^2 f)\exp(2i\pi f u)\right)\right| \\ e_{3,k} &= \left| h_{3,k} - \Re\left(iM_S/(4\pi r_0^2 f) e^{-4\pi i r_0 f}\exp(2i\pi f u)\right)\right| . \end{aligned} \tag{8}$$

For the case $u_0 = 0$ms, $u_{N-1} = 100$ms, $f = 1$kHz, $r_0 = 1.25$ms ($\approx$km), and $M_S = 0.25$ms ($\approx M_\odot$), we found:

N	$\|\|e_{2,k}\|\|$	$\|\|e_{3,k}\|\|$
2^9	4.0×10^{-4}	1.6×10^{-3}
2^{15}	1.6×10^{-6}	2.7×10^{-5}
2^{21}	2.4×10^{-8}	4.3×10^{-7}

(9)

where $||e_k||$ is defined to be

$$||e_k|| = \sqrt{\frac{\sum_{k=0}^{N-1} e_k^2}{N}} . \tag{10}$$

Thus the errors are tending to zero, and N can be chosen so as to attain a desired accuracy. Note that an error of order machine precision is achieved for special values of the frequency $f = k(N-1)/N/(u_{N-1} - u_0)$ with k an integer; in this case we would have that the cyclic assumption of the discrete Fourier transform would be satisfied, i.e. $u_i = u_{i+N}$.

4. Could matter shells explain the GW echo claims?

In,[1] the first of the tentative search for echoes, the authors find evidence for the existence of echoes in the first detection event[36] **GW150914**. They find further comparable evidence for echoes from the events GW151012[37] (then referred to as LVT151012) and GW151226.[38] The references report a number of echo events for GW150914, with the first occurring at about 0.3s after merger; therefore, if caused by a matter shell, the radius would be about 45,000km. The magnitude of the echo was about 0.0992 times that of the original signal. Using 132Hz for the frequency, which is its value when the amplitude was at its maximum at the end of the merger phase, and applying Eq. (4) gives $M_S \simeq 740{,}000 M_\odot$. Such a mass within a radius of 45,000km would constitute a black hole, so the scenario of an echo caused by a shell can be discounted for GW150914.

Extending their investigations to the first BNS detection **GW170817**,[39] the authors of[1] find evidence again of the existence of echoes in the postmerger event.[3] The echo was reported to occur at frequency $f_{echo} \simeq 72$ Hz, approximately 1.0 sec after the BNS merger event. The inspiral signal is at 72Hz about 4.0s before merger, so if the reported echo is caused by a matter shell it must have a radius of about

2.5s$\simeq$ 750,000km. Assuming that the magnitude of the echo signal must be at least 0.01$\times$ that of the original signal and applying Eq. (4), it follows that for the echo to be caused by a shell it would have a mass of approximately $10^7 M_\odot$. Now, a mass of $10^7 M_\odot$ inside a radius of 750,000km would constitute a black hole, so the scenario of an echo caused by a shell can be discounted for GW170817.

The above two examples illustrate the general difficulty of producing a GW echo by means of a matter shell. An echo, in the usual sense, is a repeat of the original signal after a short time delay, which in practice must be at least hundreds of ms, corresponding to a shell radius of at least $\sim$ 20,000km. Now, Eq. (4) shows that the shell mass $M_S = 4\pi R r_0^2 f$ which, for expected values of the frequency f, will be large – either implying a black hole and thus not feasible, or requiring an unexpected astrohysical scenario. A much smaller shell radius would avoid these difficulties, but the effect of the shell would be seen as a modification of the original signal, rather than as an echo. Some possible scenarios are presented in the next section.

5. GW events and matter shells

5.1. *GW150914 in the presence of a matter shell*

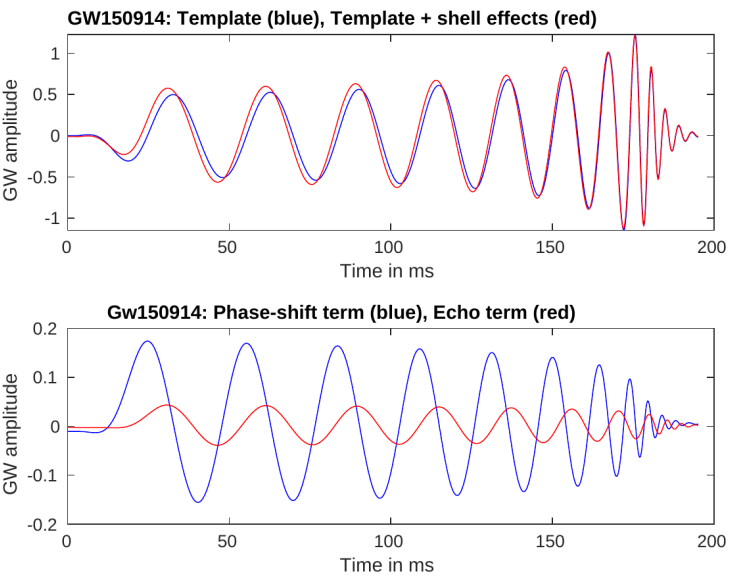


Fig. 2. The effect of a matter shell of radius 3ms (about 900km) and mass 0.3ms (about $60 M_\odot$) on the signal of GW150914. The top panel shows the original signal in blue, and the original signal plus modifications due to the shell in red. The lower panel shows the modifications due to the phase-shift term in blue, and due to the echo term in red.

Although a matter shell could not explain the echoes that might exist in GW150914 data, we now investigate how a hypothetical matter shell might modify the signal from a binary black hole merger. We consider the example case of a shell at radius 3ms (about 900km) and of mass 0.3ms (about $60M_\odot$) and the signal of GW150914[40–42] (Of course, the astrophysical evidence does not suggest the existence of such a shell). The results are shown in Fig. 2. The top panel shows that there is a small but noticeable modification to the template signal, particularly at early times; this is because the frequency is lower at early times and so the modification effects are larger. The bottom panel shows the contributions of the phase-shift term (h_2, blue) and the echo term (h_3, red); it is noticeable that, unlike the template signal, these terms decrease in magnitude as the frequency increases with time.

The accuracy of the results presented in Fig. 2 is limited since the formalism used in[25] assumed a weak field GW source, which is not the case for two black holes at merger. In particular, GWs reflected by the shell would be partially absorbed by the black holes, so reducing the magnitude of the echo contribution to the GW signal.

5.2. *Binary Neutron Star (BNS) mergers*

BNS GW events that have been observed include GW170817[43] and GW190425.[44] Of these, GW170817 was at a higher signal to noise ratio and the event was observed post-merger in the electromagnetic spectrum,[45] indicating that the post-merger object contained a large amount of free matter; we will therefore focus on this event.

The relevant source parameters reported for the event are[46,47]: total mass $M_1 + M_2 = 2.74M_\odot$, and radii $R_1 = 10.8$km, $R_2 = 10.7$km. The reported GW signal increased in amplitude and frequency until 500Hz, at which stage the signal finished; i.e. the signal was observed during the inspiral phase, but ended as the two objects started to merge. The merger probably produced a central remnant of a neutron star or a black hole, in which case GWs in the lowest quasinormal mode would have been produced; however, the frequency of these GWs would be about 1.5 to 3kHz for a neutron star, or about 6kHz for a $2M_\odot$ black hole, which is outside the sensitivity band of the LIGO detectors.

In order to estimate the possible effect of matter on GWs emitted from a central remnant, we consider the model of a spherical shell of mass $M_S = 0.7M_\odot$ and radius $r_0 = 25$km around a GW source at either 6kHz or 2kHz. We find that the phase-shift term $2iM_S/(r_0^2\nu)$ evaluates to

$$\frac{iM_S}{r_0^2\pi f} = 0.0263i \;\text{ or }\; 0.0788i\,, \tag{11}$$

for 6kHz or 2kHz, respectively. The echo effect would be 1/4 of the above values, and the delay would be 0.1667ms, which is the same as the wave period (at 6kHz),

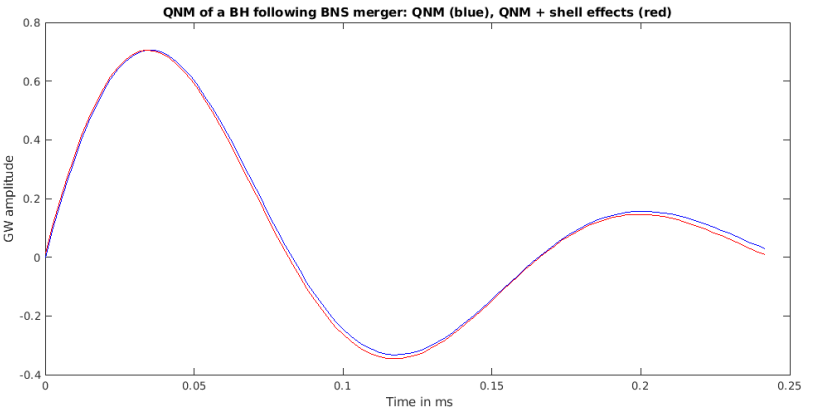


Fig. 3. The effect of a matter shell of radius 25km and mass $0.7M_{\odot}$ on a quasinormal mode (QNM) signal of a $2M_{\odot}$ remnant of a binary neutron star merger. The original signal is in blue, and the original signal plus modifications due to the shell is in red.

or less than a wave period (at 2kHz). Thus, in the future if GW detectors in the kHz band are operational and if a BNS event occurs at high signal to noise ratio, then shell effects would affect the GW signal in a measurable way, although the small delay time means that it would be difficult to disentangle the echo and phase-shift effects from the main signal. The modifications to the quasinormal mode signal are illustrated in Fig. 3. However, there are a number of caveats that should be noted:

- The model in[25] assumed that the shell is static, but the aftermath of a BNS meger will be highly dynamical.
- The hypothesis of a shell forming is not supported by a detailed numerical simulation; indeed, since the system started as an inspiral, the matter outside the remnant should have a ring-like structure. Thus, the matter effects could be much smaller than shown in Eq. (11) for an observer on the axis of rotation of the system.
- The comment at the end of Sec. 5.1 about absorption of GWs by a black hole applies here.

So the quantitative values in Eq. (11) should be interpreted as indicative of the order of magnitude of the interaction of GWs with matter, rather than as precise estimates. It should also be noted that if the numerical modeling includes all the matter, and if the simulation run period includes the quasinormal mode ringdown, then shell effects would already be included in the simulation.

5.3. *Core collapse supernovae*

We next turn our attention to an anticipated candidate of detectable GW waves: core collapse supernovae (CCSNe). Whilst BBH and BNS mergers are currently

the only GW events picked up by LIGO and VIRGO, supernovae are expected to produce, under certain conditions, GW waves detectable by the current generation of interferometers or those on the horizon. For now, all detection of supernovae have been confined to electromagnetic detection. Photons originate at the outer edge of a star and hence provide only limited information on the interior regions. The detection of GWs which are the result of the aspherical motion of the inner regions will provide a wealth of information on these regions and the mechanism leading to the supernova explosion, where all the four fundamental forces of nature are involved.

Whilst the central engines and inner regions of CCSNe have yet to be fully understood, there exist several studies of their progenitors and the subsequent evolution and detection.[48–51] For stars of mass larger than $8M_\odot$, evolution normally proceeds through several stages of core burning and then to core collapse once nuclear fusion halts when there are no further burning processes to balance the gravitational attraction. Typically, these cores are iron cores, with the critical mass signalling the onset of core collapse ranging from $1.3M_\odot$ to $1.7M_\odot$. The core breaks into two during the collapse, with the inner core of $0.4M_\odot$ to $0.6M_\odot$ in sonic contact and collapsing homologously and the outer core collapsing supersonically. The inner core reaches supranuclear densities of $\sim 2\times 10^{14}\mathrm{g/cm}^3$ where the nuclear matter stiffens, resulting in a bounce of the inner core. The resulting shock wave is launched into the collapsing outer core. However, the shock loses energy to dissociation of iron nuclei, stalling at ∼150 km within ∼10 ms after formation. Many computationally demanding simulations exist[52–54] for generation of GWs from CCSNe.

5.3.1. *Model for the GW modifications due to matter around the inner core*

Ref.[54] presents GW waveforms from simulations of CCSNe for various zero age main sequence (ZAMS) masses in the range $9M_\odot, \cdots, 60M_\odot$. The GW signal starts with an initial burst of duration about 50ms and frequency about 100Hz, followed by a quiescent period. We model this part of the waveform as

$$\begin{aligned} &h_+ + ih_\times \sim \sin(0.2\pi u)\sin(0.02\pi u)\,{}_2Z_{2,2}\,, \quad 0\le u\le 50\\ &h_+ + ih_\times = 0\,, \quad 50\le u\le 100\,. \end{aligned} \tag{12}$$

GWs are generated by aspherical motions in the inner core, commencing just after the bounce. The inner core is surrounded by the outer core, treated as a thick matter shell, and we now model its modifications to the GW signal. The shell has an inner radius r_{in}, an outer radius r_{out}, and density at $r = r_{in}$ of ρ_0 with density fall-off $\rho \propto r^{-1-a}$ with $1/2 \le a \le 2$. The effect of the whole shell is obtained by decomposing it into thin shells and then integrating. The result for the echo term is not a simple analytic expression and will be evaluated numerically. However, the phase shift term does give a simple analytic result

$$\int_{r_{in}}^{r_{out}} \frac{iM_s}{r^2\pi f}dr = i\int_{r_{in}}^{r_{out}} \frac{4\pi r^2\rho_0(r_{in})^{1+a}}{r^{1+a}r^2\pi f}dr = \frac{4i\rho_0 r_{in}}{af}\left(1-\left(\frac{r_{in}}{r_{out}}\right)^a\right), \tag{13}$$

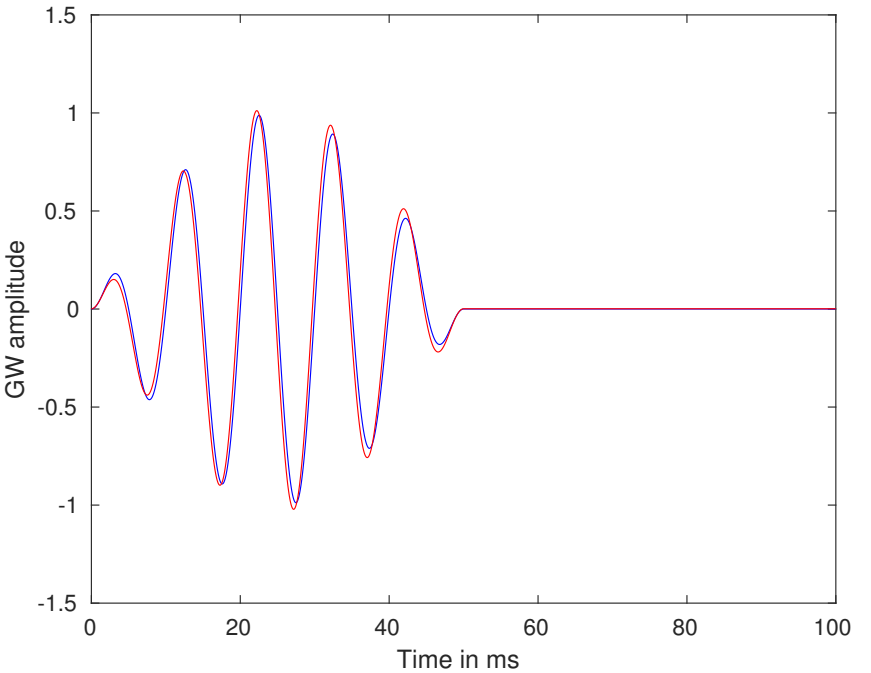


Fig. 4. The effect on the initial burst in a CCSNe GW signal of a matter shell extending from a radius 30km to the shock boundary at 150km. The original signal is in blue, and the signal plus modifications is shown in red.

where f is the GW frequency. Fig. 4 shows the original signal given by Eq. (12) in blue, and the original signal plus shell modifications in red, for the case $r_{in} = 0.1$ms (≈ 30km), $r_{out} = 0.5$ms (≈ 150km), $a = 1$, and $\rho_0 = 0.05/\text{ms}^2$ ($\approx 0.75{\times}10^{12}\text{g/cm}^3$). These values model: the inner core as a proto-neutron star (PNS) of radius 30km, and whose oscillations generate the GWs; the shock boundary as having a radius of 150km; and the density at the inner radius ($\approx 0.75 \times 10^{12}\text{g/cm}^3$) at a couple of orders below the supranuclear density. For this model, Eq. (13) evaluates to $0.16i$, and the total mass of the shell to $0.503M_\odot$.

There is some uncertainty in the parameter values that should be used in modeling the matter shell around the inner core. The simple form of Eq. (13) shows how varying the parameters would change the magnitude of the shell effect.

6. Conclusion

There are astrophysical scenarios which can be regarded as comprising a shell of matter around a GW source, and this paper has investigated in what way the GW signal would be affected. The investigation started with GW events for which echoes have been claimed to exist in the LIGO data, and it was found that such echoes

could not be caused by a matter shell. Thus, an unambiguous observation of GW echoes in the future would favour the existence of ECOs.

We investigated the effect of matter shells in three specific example cases. The first was a binary black hole merger analogous to GW150914, surrounded by a hypothetical matter shell at radius 900km and mass $60M_{\odot}$. Astrophysically, such a shell is highly unlikely to exist, but this case is useful as it well illustrates some of the features of the shell effects on the waveform. The next case considered was the quasinormal mode signal from the remnant of a binary neutron star merger like GW170917. In this case, it is known that there is a substantial amount of matter around the remnant, although the extent to which the shell model is appropriate is unclear. The final case considered was that of a core collapse supernova. Although GWs from such events have not been observed, they are regarded as potential sources; and here it is clear that the proto neutron star, in which the GWs are generated, is surrounded by shells of matter. Of the three cases, the core collapse supernova is that which yielded the largest shell modifications to the GWs, and for which the predictions are most reliable.

The effects of matter shells are small but measurable if the signal to noise ratio is sufficiently high. As GW observations become more accurate, through both hardware developments and, as time passes, the increasing chance of observing nearby events, these effects will need to be taken into account.

Acknowledgments

This work was supported by the National Research Foundation, South Africa, under grant numbers 118519 and 114815.

References

1. J. Abedi, H. Dykaar and N. Afshordi, Echoes from the Abyss: Tentative evidence for Planck-scale structure at black hole horizons, *Phys. Rev. D* **96**, p. 082004 (2017).
2. R. S. Conklin, B. Holdom and J. Ren, Gravitational wave echoes through new windows, *Phys. Rev. D* **98**, p. 044021 (Aug 2018).
3. J. Abedi and N. Afshordi, Echoes from the Abyss: A highly spinning black hole remnant for the binary neutron star merger GW170817, *JCAP* **11**, p. 010 (2019).
4. G. Ashton, O. Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. D. Meadors, A. B. Nielsen, A. Nitz and J. Westerweck, Comments on: "Echoes from the abyss: Evidence for Planck-scale structure at black hole horizons" (12 2016).
5. J. Westerweck, A. Nielsen, O. Fischer-Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. Meadors and A. H. Nitz, Low significance of evidence for black hole echoes in gravitational wave data, *Phys. Rev. D* **97**, p. 124037 (2018).
6. J. Abedi, H. Dykaar and N. Afshordi, Echoes from the Abyss: The Holiday Edition!, *arXiv e-prints*, p. arXiv:1701.03485 (January 2017).
7. J. Abedi, H. Dykaar and N. Afshordi, Comment on: "Low significance of evidence for black hole echoes in gravitational wave data", *arXiv e-prints*, p. arXiv:1803.08565 (March 2018).

8. J. Abedi and N. Afshordi, Echoes from the Abyss: A Status Update, *arXiv e-prints*, p. arXiv:2001.00821 (January 2020).
9. A. Almheiri, D. Marolf, J. Polchinski and J. Sully, Black Holes: Complementarity or Firewalls?, *JHEP* **02**, p. 062 (2013).
10. O. Lunin and S. D. Mathur, AdS/CFT duality and the black hole information paradox, *Nucl. Phys. B* **623**, 342 (2002).
11. G. F. Giudice, M. McCullough and A. Urbano, Hunting for Dark Particles with Gravitational Waves, *JCAP* **10**, p. 001 (2016).
12. P. O. Mazur and E. Mottola, Gravitational vacuum condensate stars, *Proc. Nat. Acad. Sci.* **101**, 9545 (2004).
13. S. L. Liebling and C. Palenzuela, Dynamical Boson Stars, *Living Rev. Rel.* **20**, p. 5 (2017).
14. V. Cardoso, E. Franzin and P. Pani, Is the gravitational-wave ringdown a probe of the event horizon?, *Phys. Rev. Lett.* **116**, p. 171101 (2016), [Erratum: Phys. Rev. Lett. 117, 089902 (2016)].
15. V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela and P. Pani, Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale, *Phys. Rev. D* **94**, p. 084031 (2016).
16. V. Cardoso and P. Pani, Tests for the existence of black holes through gravitational wave echoes, *Nature Astron.* **1**, 586 (2017).
17. H. Nakano, N. Sago, H. Tagoshi and T. Tanaka, Black hole ringdown echoes and howls, *Progress of Theoretical and Experimental Physics* **2017** (Jul 2017).
18. A. Testa and P. Pani, Analytical template for gravitational-wave echoes: Signal characterization and prospects of detection with current and future interferometers, *Phys. Rev. D* **98**, p. 044018 (2018).
19. Y.-T. Wang, Z.-P. Li, J. Zhang, S.-Y. Zhou and Y.-S. Piao, Are gravitational wave ringdown echoes always equal-interval?, *Eur. Phys. J. C* **78**, p. 482 (2018).
20. P. Bueno, P. A. Cano, F. Goelen, T. Hertog and B. Vercnocke, Echoes of Kerr-like wormholes, *Phys. Rev. D* **97**, p. 024040 (2018).
21. A. Maselli, S. H. Völkel and K. D. Kokkotas, Parameter estimation of gravitational wave echoes from exotic compact objects, *Phys. Rev. D* **96**, p. 064045 (2017).
22. C. Barceló, R. Carballo-Rubio and L. J. Garay, Gravitational wave echoes from macroscopic quantum gravity effects, *JHEP* **05**, p. 054 (2017).
23. R. Carballo-Rubio, F. Di Filippo, S. Liberati and M. Visser, Phenomenological aspects of black holes beyond general relativity, *Phys. Rev. D* **98**, p. 124009 (2018).
24. R. A. Konoplya, Z. Stuchlík and A. Zhidenko, Echoes of compact objects: New physics near the surface and matter at a distance, *Phys. Rev. D* **99**, p. 024007 (Jan 2019).
25. N. T. Bishop, P. J. van der Walt and M. Naidoo, Effect of a low density dust shell on the propagation of gravitational waves, *Gen. Rel. Grav.* **52**, p. 92 (2020).
26. E. Barausse, V. Cardoso and P. Pani, Can environmental effects spoil precision gravitational-wave astrophysics?, *Phys. Rev. D* **89**, p. 104059 (2014).
27. P. Leung, Y. Liu, W. Suen, C. Tam and K. Young, Quasinormal modes of dirty black holes, *Phys. Rev. Lett.* **78**, 2894 (1997).
28. G. Goswami, G. K. Chakravarty, S. Mohanty and A. R. Prasanna, Constraints on cosmological viscosity and self-interacting dark matter from gravitational wave observations, *Phys. Rev. D* **95**, p. 103509 (May 2017).
29. G. Baym, S. P. Patil and C. J. Pethick, Damping of gravitational waves by matter, *Phys. Rev. D* **96**, p. 084033 (Oct 2017).

30. M. Naidoo, N. T. Bishop and P. J. van der Walt, Modifications to the signal from a gravitational wave event due to a surrounding shell of matter, *arXiv e-prints*, p. arXiv:2102.00060 (January 2021).
31. H. Bondi, M. G. J. van der Burg and A. W. K. Metzner, Gravitational waves in general relativity VII. Waves from axi-symmetric isolated systems, *Proc. R. Soc. London* **A269**, 21 (1962).
32. R. Sachs, Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time, *Proc. Roy. Soc. London* **A270**, 103 (1962).
33. N. T. Bishop and L. Rezzolla, Extraction of gravitational waves in numerical relativity, *Living Rev. Relativ.* **19**, p. 1 (Oct 2016).
34. N. T. Bishop, Linearized solutions of the Einstein equations within a Bondi-Sachs framework, and implications for boundary conditions in numerical simulations, *Class. Quantum Grav.* **22**, 2393 (2005).
35. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes*, 2nd edn. (Cambridge University Press, New York, 1992).
36. B. Abbott *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**, p. 061102 (2016).
37. A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown and B. Krishnan, 1-OGC: The first open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO data, *Astrophys. J.* **872**, p. 195 (2019).
38. B. P. Abbott *et al.*, GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence, *Phys. Rev. Lett.* **116**, p. 241103 (2016).
39. B. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya and et al., Gw170817: Observation of gravitational waves from a binary neutron star inspiral, *Physical Review Letters* **119** (Oct 2017).
40. B. P. Abbott *et al.*, A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, *Class. Quant. Grav.* **37**, p. 055002 (2020).
41. G. W. O. S. Center, GW150914 Template Data `https://www.gw-openscience.org/GW150914data`.
42. G. W. O. S. Center, Binary Black Hole Signals in LIGO Open Data `https://losc.ligo.org/s/events/GW150914/LOSC_Event_tutorial_GW150914.html`.
43. B. Abbott *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (2017).
44. B. Abbott *et al.*, GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$, *Astrophys. J. Lett.* **892**, p. L3 (2020).
45. B. P. Abbott et. al., Multi-messenger Observations of a Binary Neutron Star Merger, *Astrophys. J. Lett.* **848**, p. L12 (2017).
46. B. Abbott *et al.*, Observing gravitational-wave transient GW150914 with minimal assumptions, *Phys. Rev. D* **93**, p. 122004 (2016), [Addendum: Phys.Rev.D 94, 069903 (2016)].
47. B. Abbott *et al.*, GW170817: Measurements of neutron star radii and equation of state, *Phys. Rev. Lett.* **121**, p. 161101 (2018).
48. B. Müller, Hydrodynamics of core-collapse supernovae and their progenitors, *Astrophysics* **6**, p. 3 (2020).
49. E. Abdikamalov, G. Pagliaroli and D. Radice, Gravitational Waves from Core-Collapse Supernovae (10 2020).
50. S. E. Woosley, A. Heger and T. A. Weaver, The evolution and explosion of massive stars, *Rev. Mod. Phys.* **74**, p. 1015 (November 2002).

51. S. Woosley and A. Heger, Nucleosynthesis and Remnants in Massive Stars of Solar Metallicity, *Phys. Rept.* **442**, 269 (2007).
52. H. Andresen, B. Müller, E. Müller and H.-T. Janka, Gravitational Wave Signals from 3D Neutrino Hydrodynamics Simulations of Core-Collapse Supernovae, *Mon. Not. Roy. Astron. Soc.* **468**, 2032 (2017).
53. H. Andresen, E. Müller, H. Janka, A. Summa, K. Gill and M. Zanolin, Gravitational waves from 3D core-collapse supernova models: The impact of moderate progenitor rotation, *Mon. Not. Roy. Astron. Soc.* **486**, 2238 (2019).
54. D. Radice, V. Morozova, A. Burrows, D. Vartanyan and H. Nagakura, Characterizing the Gravitational Wave Signal from Core-Collapse Supernovae, *Astrophys. J. Lett.* **876**, p. L9 (2019).

Odd-dimensional gravitational waves from a binary system on a three-brane

D. V. Gal'tsov[a,*] and M. Khlopunov[a,b,†]

[a] *Faculty of Physics, Lomonosov Moscow State University, Moscow, 119899, Russia*

[b] *Institute of Theoretical and Mathematical Physics, Lomonosov Moscow State University, Moscow 119991, Russia*

E-mail: * *galtsov@phys.msu.ru,* † *khlopunov.mi14@physics.msu.ru*

We consider gravitational radiation in the presence of non-compact extra dimensions. If their number is odd, all spacetime becomes odd-dimensional and formation of gravitational radiation becomes non-trivial because of violation of the Huygens principle. Gravitational waves travel with the speed of light, while the full retarded gravitational field of a localized source propagates with all velocities lower or equal to the speed of light, so special care is needed to extract radiation. Here we consider a simplified model consisting of two point masses moving on a three-brane embedded in five-dimensional bulk. Particles are assumed to interact through a massless scalar field living on the same brane, while gravitational radiation is emitted into the full five-dimensional space. We use the Rohrlich-Teitelboim approach to radiation, extracting the radiative component of the retarded gravitational field via splitting of the energy-momentum tensor. The source term consists of the local contribution from the particles and the non-local contribution from the scalar field stresses. The latter is computed using the DIRE approach to the post-Newtonian expansions. In the non-relativistic limit, we find an analog of the quadrupole formula containing the integral over the full history of motion, preceding the retarded moment of time. We compute gravitational radiation and study the orbit evolution of the non-relativistic circular binary system on the brane.

Keywords: Extra dimensions, Huygens principle, gravitational radiation, brane

1. Introduction

Gravitational radiation in the higher-dimensional spacetimes has become an intriguing problem in the last twenty years due to development of the gravity theories with large extra dimensions.[1–4] Recent successes of the gravitational-wave astronomy give us the new tool for exploring extra dimensions.[5–10] Some constraints on their size, number and geometry have already been extracted.[11–16] Another new tool relevant to extra dimensions is shadow of a black hole.[17–19]

With regard to the theoretical study of multidimensional radiation, then in the majority of literary sources only radiation in even space-time dimensions is studied,[20–24] while the odd dimensions were discussed mainly in the more academic context of the problem of radiation reaction.[25–28] Huygens principle violation in odd dimensions[29–31] leads to significant difference of field propagation in even and odd dimensions. In both cases, signal from the instantaneous flash reaches the observation point in time required to propagate with the speed of light. However,

in odd dimensions, after the arrival of the primary signal an endless tail is observed which is not the case in the even dimensions. Mathematically, this is related to the localisation of the odd-dimensional retarded Green's functions not only on the light cone, but also inside it:

$$G^{2\nu+1}_{\rm ret}(x) = \frac{(-1)^\nu}{(2\pi)^\nu}\frac{d^{\nu-1}}{(rdr)^{\nu-1}}\frac{\theta(t)\,\theta(t^2-r^2)}{\sqrt{t^2-r^2}},\ \nu\in\mathbb{N}, \tag{1}$$

where $t = x^0$, and $r = |\mathbf{x}|$. The odd-dimensional gravitational radiation of the non-relativistic binary systems has been studied[32] using the effective field theory formalism,[33,34] which is insensitive to the dimensionality of spacetime. But since it is based on the Fourier decompositions, the obtained analog of the quadrupole formula does not contain any information about the gravitational field in the wave zone and does not reveal the role of the tail in the formation of the gravitational-wave signal. So here we consider five-dimensional radiation essentially using the space-time picture.

2. The setup

We consider a simplified model of a binary system whose motion is confined inside the four-dimensional subspace (brane). Particles interact with each other only through the massless scalar field localised on the same brane, while gravitational radiation is five-dimensional. Our localisation mechanism is purely kinematical: if the initial conditions are restricted to the brane, and the mediator scalar field is four-dimensional, the system remains in the brane forever. Such a system admits the stable elliptical orbits, and the interaction field is free from tails, providing us with a simple setup to study the features of the odd-dimensional gravitational radiation related to the Huygens principle violation.

Gravitational radiation of such a system can be described by the linearised theory on the Minkowski background without use of the quadratic part of the Ricci tensor corresponding to the gravitational stresses.[35] However, one still needs to take into account the scalar field stresses corresponding to the energy of the particles interaction:

$$\Box\bar{h}_{MN}(x) = -2\kappa_5\left[T^{\rm P}_{MN}(x)+T^{\rm F}_{MN}(x)\right],\quad \partial^M\bar{h}_{MN}=0, \tag{2}$$

$$T^{\rm P}_{MN}(x) = \delta^\mu_M\delta^\nu_N\sum_{a=1}^{2}m_a\int d\tau_a\,\dot{z}_{a\mu}\dot{z}_{a\nu}\,\delta^{(4)}(x-z_a)\delta(x^4), \tag{3}$$

$$T^{\rm F}_{MN}(x) = \frac{1}{4\pi}\delta^\mu_M\delta^\nu_N\left(\partial_\mu\varphi\partial_\nu\varphi-\frac{1}{2}\eta_{\mu\nu}\partial^\alpha\varphi\partial_\alpha\varphi\right)\delta(x^4), \tag{4}$$

where $M,N=\overline{0,4}$ and $\mu,\nu=\overline{0,3}$, and $\eta_{MN}=\mathrm{diag}(-1,1,1,1,1)$. Here, $z^\mu_a(\tau_a)$ are the world lines of the particles and $\dot{z}^\mu_a = dz^\mu_a/d\tau_a$, m_a are their masses, and κ_5 is the five-dimensional gravitational constant. We assume that the brane lies in the $x^4=0$ hypersurface.

The effective energy-momentum tensor of the gravitational field, by analogy with the four-dimensional theory,[36] is given by

$$t^{\rm rad}_{MN}(x) = \frac{1}{4\kappa_5}\left\langle \partial_M \bar{h}^{\rm TT}_{ij} \partial_N \bar{h}^{\rm TT}_{ij} \right\rangle, \quad i,j = \overline{1,4}, \tag{5}$$

where a periodic motion of the source is assumed, and bracket $\langle \ldots \rangle$ denotes averaging over the period. In Eq. (5), we use the metric deviations in the transverse-traceless gauge

$$h^{\rm TT}_{0M} = 0, \quad h^{\rm TT}_{ii} = 0, \quad \partial^j h^{\rm TT}_{ij} = 0, \tag{6}$$

in which gravitational waves have five independent polarisations. Namely, polarisations of the plane wave propagating along the x^3-coordinate are presented by the matrix

$$\bar{h}^{\rm TT}_{ij} = \begin{pmatrix} h_+ - \frac{1}{2}h_\circ & h_\times & 0 & h_{14} \\ h_\times & -h_+ - \frac{1}{2}h_\circ & 0 & h_{24} \\ 0 & 0 & 0 & 0 \\ h_{14} & h_{24} & 0 & h_\circ \end{pmatrix}, \tag{7}$$

where the "cross" and "plus" polarisations are the same as in the four-dimensional theory[35,36]

$$h_+ = \frac{1}{2}\left(h_{11} - h_{22}\right), \quad h_\times = h_{12}, \tag{8}$$

and $h_\circ$ is the "breathing" mode[7]

$$h_\circ = \frac{2}{3}h_{44} - \frac{1}{3}\left(h_{11} + h_{22}\right). \tag{9}$$

The brane-living observer would detect only three of them – the standard "plus" and "cross", and the "breathing" polarisation, corresponding to the uniform shrinking and stretching of the probe masses circle lying in the plane orthogonal to the wave propagation direction.[7] We find that the breathing polarisation is non-vanishing even when both the source and an observer live on the brane.

3. Point particles contribution

To calculate the point particles contribution into the gravitational radiation we use the Rohrlich-Teitelboim definition of radiation[37–39] (see also[40–43]) based on the Lorentz-invariant decomposition of the on-shell energy-momentum tensor of the retarded field in the far zone. The Lorentz-invariant distance from the particle trajectory $z^M(\tau)$,

$$\hat{\rho} \equiv \hat{v}^M \hat{X}_M, \quad \hat{\rho} \xrightarrow{r \gg |\mathbf{z}|} r \tag{10}$$

is equivalent to the spatial distance from the system to the observation point. Here, the following notations are used:

$$\hat{X}^M = x^M - z^M(\hat{\tau}) \equiv \hat{\rho}\hat{c}^M, \quad \hat{c}^M\hat{v}_M = 1, \quad \hat{c}^2 = 0, \tag{11}$$

$$\hat{v}^M = v^M(\hat{\tau}), \quad v^M = \frac{dz^M}{d\tau}, \tag{12}$$

and $\hat{\tau}$ is the retarded proper time defined as

$$\left(x^M - z^M(\hat{\tau})\right)^2 = \hat{X}^2 = 0, \; x^0 \geq z^0(\hat{\tau}). \tag{13}$$

Then, in D dimensions, the on-shell energy-momentum tensor can be expanded in the inverse powers of $\hat{\rho}$ as follows

$$T^{MN} = T^{MN}_{\text{Coul}} + T^{MN}_{\text{mix}} + T^{MN}_{\text{rad}} \tag{14}$$

$$T^{MN}_{\text{Coul}} \sim \frac{A^{MN}}{\hat{\rho}^{2D-4}}, \quad T^{MN}_{\text{mix}} \sim \frac{B^{MN}}{\hat{\rho}^{2D-5}} + \ldots + \frac{C^{MN}}{\hat{\rho}^{D-1}}, \quad T^{MN}_{\text{rad}} \sim \frac{D^{MN}}{\hat{\rho}^{D-2}}. \tag{15}$$

Here, the most short-range part T^{MN}_{Coul} is the energy-momentum tensor of the Coulomb part of the field; T^{MN}_{mix} is the mixed part, which is absent in $D = 3$ and consists of more that one term in $D > 4$. The remaining long-range part of the energy-momentum tensor T^{MN}_{rad} has the properties allowing one to associate it with the energy-momentum flux carried by the emitted part of the field. Indeed, it is separately conserved $\partial_M T^{MN}_{\text{rad}} = 0$; it is proportional to the direct product of two null vectors $T^{MN}_{\text{rad}} \sim \hat{c}^M\hat{c}^N$, corresponding to the propagation of the associated energy-momentum flux exactly with the speed of light $\hat{c}_M T^{MN}_{\text{rad}} = 0$. Finally, it falls down as $T^{MN}_{\text{rad}} \sim 1/r^{D-2}$ and gives positive definite energy-momentum flux through the distant sphere of the area $\sim r^{D-2}$.

These properties hold both in even and odd dimensions. However, due to the Huygens principle violation, in odd dimensions the emitted part of the energy-momentum tensor T^{MN}_{rad} depends on the entire history of the source's motion preceding the retarded time $\hat{\tau}$, while in even ones it is completely determined by the state of the source at this moment. Also, due to the form of the energy-momentum tensor (5), which is the bilinear expression of the field derivatives, one can define its emitted part, by analogy with that of the energy-momentum tensor

$$(\partial_M \bar{h}^{\text{TT}}_{ij})^{\text{rad}} \sim 1/\hat{\rho}^{D/2-1}. \tag{16}$$

Therefore, the radiated energy flux through the distant $(D-2)$-dimensional sphere of radius r is given as

$$W_D = \int T^{0i}_{\text{rad}}\, n^i\, r^{D-2}\, d\Omega_{D-2}, \tag{17}$$

where $d\Omega_{D-2}$ is the angular element, and $n^i = x^i/r$ is a unit spacelike vector in the observation direction.

Considering the gravitational radiation from the non-relativistic binary system we demonstrate that to find the non-relativistic approximation of the emitted part

of the gravitational field produced by point particles one have to introduce, besides the small particles velocities $|\mathbf{v}| \ll 1$, another small parameter

$$\mathbf{s}(\tau) = \frac{\mathbf{z}(\hat{\tau}) - \mathbf{z}(\tau)}{\hat{\tau} - \tau}, \quad \lim_{\tau \to \hat{\tau}} \mathbf{s}(\tau) = \hat{\mathbf{v}}, \quad |\mathbf{s}| \sim |\mathbf{v}|, \tag{18}$$

which is of order of particles velocities for any moment of proper time. Then, we find the point particles contribution into the gravitational radiation from the binary system as

$$(\partial_M \bar{h}^{\rm P}_{ij})^{\rm rad} = -\frac{\mu\kappa_5 \bar{c}_M}{2^{3/2}\pi^2 r^{3/2}} \int_{-\infty}^{\bar{\tau}} d\tau \frac{\dot{a}_i v_j + 2a_i a_j + v_i \dot{a}_j}{(\bar{\tau} - \tau)^{1/2}}, \quad \bar{c}_M = [-1, \mathbf{n}], \tag{19}$$

where $z^i = z_2^i - z_1^i$ is the relative coordinate of the system, $\mu = m_1 m_2/(m_1 + m_2)$ is its reduced mass, and $\bar{\tau} = t - r$ is the retarded proper time calculated up to the leading order. Note that the emitted part of the gravitational field (19) is proportional to the null vector $\bar{c}_M$ leading to the corresponding energy-momentum tensor (5) being proportional to the direct product of two null vectors, in accordance with the Rohrlich-Teitelboim approach, and depends on the history of the system's motion preceding the retarded time $\bar{\tau}$.

4. Scalar field contribution

The scalar field stresses contribution into the gravitational radiation is calculated by use of the DIRE approach to the post-Newtonian expansions.[44] It is based on the splitting of the spacetime into the near zone, whose size is of order of the characteristic wavelength of gravitational radiation $\mathcal{R} = \mathcal{S}/|\mathbf{v}| \sim \lambda_{\rm GW}$ ($\mathcal{S}$ is the characteristic size of the binary system), and the radiation zone being exterior to the near zone.

In our calculations we assume that the observation point is in the radiation zone, the retardation of the scalar field inside the near zone is negligible, and that the energy-momentum density of the scalar field is non-vanishing only in the near zone being large enough due to the slow motion of the particles. Then, we arrive at the scalar field contribution into the gravitational radiation in the form

$$(\partial_M \bar{h}^{\rm F}_{ij})^{\rm rad} = -\frac{\mu\kappa_5 \bar{c}_M}{2^{5/2}\pi^2 r^{3/2}} \int_{-\infty}^{\bar{\tau}} d\tau \, \frac{\ddot{a}_i z_j + 2\dot{a}_i v_j + 2a_i a_j + 2v_i \dot{a}_j + z_i \ddot{a}_j}{(\bar{\tau} - \tau)^{1/2}}. \tag{20}$$

As the point particles contribution (19), this expression is proportional to the null vector $\bar{c}_M$ and depends on the history of the particles motion.

Combining the point-like (19) and the non-local scalar field (20) contributions, we find the emitted part of the total gravitational field of the binary system:

$$(\partial_M \bar{h}^{\rm TT}_{ij})^{\rm rad} = -\frac{\kappa_5 \bar{c}_M}{2^{5/2}\pi^2 r^{3/2}} \int_{-\infty}^{\bar{\tau}} d\tau \, \frac{\overset{\cdots\cdot}{Q}{}^{\rm TT}_{ij}}{(\bar{\tau} - \tau)^{1/2}}, \quad Q_{ij} = \mu \left[z_i z_j - \frac{1}{4} \delta_{ij} z_k z_k \right]. \tag{21}$$

As in the four-dimensional theory,[36] it is determined by the transverse-traceless part of the quadrupole moment of the binary system but now depends on the history of motion.

Therefore, using the Eqs. (5) and (17) we obtain the five-dimensional analog of the quadrupole formula for the angular distribution of the gravitational radiation power

$$\frac{dW_5}{d\Omega_3} = \frac{\kappa_5}{128\pi^4}\left\langle \mathcal{A}_{ij}^{\mathrm{TT}}\mathcal{A}_{ij}^{\mathrm{TT}}\right\rangle, \quad \mathcal{A}_{ij}^{\mathrm{TT}} = \int_{-\infty}^{\bar{\tau}} d\tau\, \frac{\overset{\cdots\cdot}{Q}{}_{ij}^{\mathrm{TT}}}{(\bar{\tau}-\tau)^{1/2}}. \tag{22}$$

depending on the entire history of the binary system's motion preceding the retarded time $\bar{\tau}$. Recently, the analogous formula was obtained by use of the spectral decompositions of the retarded Green's functions,[45] which are insensitive to the dimensionality of the spacetime, and can be considered as a confirmation of our result.

5. Binary system on a circular orbit

As the simplest application of the obtained five-dimensional quadrupole formula, we consider the gravitational radiation from the non-relativistic binary system on the circular orbit. Integrating the angular distribution (22), we find the total gravitational radiation power of the system on the circular orbit

$$W_5^{\mathrm{circ}} = \frac{5}{9\pi}\kappa_5\mu^2 R_{\mathrm{s}}^4\omega_{\mathrm{s}}^7, \tag{23}$$

where R_{s} is the orbital radius of the system, and ω_{s} is its frequency of the orbital motion. Also, using the energy conservation law

$$\frac{dE_{\mathrm{tot}}}{dt} = -W_5^{\mathrm{circ}}, \tag{24}$$

where E_{tot} is the total mechanical energy of the point particles, we analyse the quasi-circular orbit shrinking of the binary system arriving at the evolution laws for the orbital frequency and radius

$$\omega_{\mathrm{s}}(t) = \left(\frac{9\pi}{55\kappa_5}\right)^{3/11}(\sqrt{\mu}g_1g_2)^{-2/11}\,(t_{\mathrm{coal}}-t)^{-3/11}, \tag{25}$$

$$R_{\mathrm{s}}(t) = R_0\left(\frac{t_{\mathrm{coal}}-t}{t_{\mathrm{coal}}-t_0}\right)^{2/11}, \tag{26}$$

where t_{coal} is the moment of time corresponding to the coalescence of the binary system $R_{\mathrm{s}} \to 0$, $\omega_{\mathrm{s}} \to \infty$, and $R_{\mathrm{s}}(t_0) = R_0$. Note that the obtained Eqs. (25) and (26) differ significantly from their four-dimensional analogs

$$\omega_{\mathrm{s}}^{\mathrm{4D}}(t) \sim (t_{\mathrm{coal}}-t)^{-3/8}, \quad R_{\mathrm{s}}^{\mathrm{4D}}(t) \sim (t_{\mathrm{coal}}-t)^{1/4}, \tag{27}$$

what could be expected, given the five-dimensional nature of gravity and infinite size of extra dimension in our model.

Acknowledgements

The work of M. Kh. was supported by the "BASIS" Foundation Grant No. 20-2-10-8-1. The work of D.G. was supported by the Russian Foundation for Basic Research on the project 20-52-18012, and the Scientific and Educational School of Moscow State University "Fundamental and Applied Space Research".

References

1. N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, The Hierarchy problem and new dimensions at a millimeter, *Phys. Lett. B* **429**, 263 (1998).
2. L. Randall and R. Sundrum, A Large mass hierarchy from a small extra dimension, *Phys. Rev. Lett.* **83**, 3370 (1999).
3. L. Randall and R. Sundrum, An Alternative to compactification, *Phys. Rev. Lett.* **83**, 4690 (1999).
4. G. R. Dvali, G. Gabadadze and M. Porrati, 4D gravity on a brane in 5D Minkowski space, *Phys. Lett. B* **485**, 208 (2000).
5. A. O. Barvinsky and S. N. Solodukhin, Echoing the extra dimension, *Nucl. Phys. B* **675**, 159 (2003).
6. C. Deffayet and K. Menou, Probing Gravity with Spacetime Sirens, *Astrophys. J. Lett.* **668**, L143 (2007).
7. D. Andriot and G. Lucena Gómez, Signatures of extra dimensions in gravitational waves, *JCAP* **06**, 048 (2017) [erratum: JCAP **05**, E01 (2019)].
8. H. Yu, Z. C. Lin and Y. X. Liu, Gravitational waves and extra dimensions: a short review, *Commun. Theor. Phys.* **71**, no. 8, 991 (2019).
9. O. K. Kwon, S. Lee and D. D. Tolla, Gravitational Waves as a Probe of the Extra Dimension, *Phys. Rev. D* **100**, 084050 (2019).
10. M. Corman, C. Escamilla-Rivera and M. A. Hendry, Constraining extra dimensions on cosmological scales with LISA future gravitational wave siren data, *JCAP* **02**, 005 (2021).
11. L. Visinelli, N. Bolis and S. Vagnozzi, Brane-world extra dimensions in light of GW170817, *Phys. Rev. D* **97**, no. 6, 064039 (2018).
12. K. Pardo, M. Fishbach, D. E. Holz and D. N. Spergel, Limits on the number of spacetime dimensions from GW170817, *JCAP* **07**, 048 (2018).
13. K. Chakravarti, S. Chakraborty, K. S. Phukon, S. Bose and S. SenGupta, Constraining extra-spatial dimensions with observations of GW170817, *Class. Quant. Grav.* **37**, no. 10, 105004 (2020).
14. S. Chakraborty, K. Chakravarti, S. Bose and S. SenGupta, Signatures of extra dimensions in gravitational waves from black hole quasinormal modes, *Phys. Rev. D* **97**, no. 10, 104053 (2018).
15. K. Chakravarti, S. Chakraborty, S. Bose and S. SenGupta, Tidal Love numbers of black holes and neutron stars in the presence of higher dimensions: Implications of GW170817, *Phys. Rev. D* **99**, no. 2, 024036 (2019).
16. A. K. Mishra, A. Ghosh and S. Chakraborty, Constraining extra dimensions using observations of black hole quasi-normal modes, [arXiv:2106.05558 [gr-qc]].
17. S. Vagnozzi and L. Visinelli, Hunting for extra dimensions in the shadow of M87*, *Phys. Rev. D* **100**, no. 2, 024020 (2019).
18. I. Banerjee, S. Chakraborty and S. SenGupta, Silhouette of M87*: A New Window to Peek into the World of Hidden Dimensions, *Phys. Rev. D* **101**, no. 4, 041301 (2020).
19. J. C. S. Neves, Constraining the tidal charge of brane black holes using their shadows, *Eur. Phys. J. C* **80**, no. 8, 717 (2020).

20. B. P. Kosyakov, Exact solutions of classical electrodynamics and the Yang-Mills-Wong theory in even-dimensional space-time, *Theor. Math. Phys.* **119**, 493 (1999).
21. V. Cardoso, O. J. C. Dias and J. P. S. Lemos, Gravitational radiation in D-dimensional space-times, *Phys. Rev. D* **67**, 064026 (2003).
22. V. Cardoso, M. Cavaglia and J. Q. Guo, Gravitational Larmor formula in higher dimensions, *Phys. Rev. D* **75**, 084020 (2007).
23. A. Mironov and A. Morozov, Radiation beyond four space-time dimensions, *Theor. Math. Phys.* **156**, 1209 (2008).
24. B. P. Kosyakov, Electromagnetic radiation in even-dimensional spacetimes, *Int. J. Mod. Phys. A* **23**, 4695 (2008).
25. D. V. Galtsov, Radiation reaction in various dimensions, *Phys. Rev. D* **66**, 025016 (2002).
26. P. O. Kazinski, S. L. Lyakhovich and A. A. Sharapov, Radiation reaction and renormalization in classical electrodynamics of point particle in any dimension, *Phys. Rev. D* **66**, 025017 (2002).
27. Y. Yaremko, Self-force in 2+1 electrodynamics, *J. Phys. A* **40**, 13161 (2007).
28. E. Shuryak, H. U. Yee and I. Zahed, Self-force and synchrotron radiation in odd space-time dimensions, *Phys. Rev. D* **85**, 104007 (2012).
29. J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Dover Publications, New York, 2014).
30. R. Courant and D. Hilbert, *Methods of Mathematical Physics: Partial Differential Equations*, Wiley Classics Library (Wiley, New York, 2008).
31. D. Ivanenko and A. Sokolov, *Sov. Phys. Dokl.* **36**, 37 (1940); *Classical Field Theory* (State Publishing House of Technical and Theoretical Literature, Moscow, 1951) (in Russian) [*Klassische Feldtheorie* (Akademie-Verlag, Berlin, 1953)].
32. V. Cardoso, O. J. C. Dias and P. Figueras, Gravitational radiation in d>4 from effective field theory, *Phys. Rev. D* **78**, 105010 (2008).
33. W. D. Goldberger and I. Z. Rothstein, An Effective field theory of gravity for extended objects, *Phys. Rev. D* **73**, 104029 (2006).
34. W. D. Goldberger, Les Houches lectures on effective field theories and gravitational radiation, [arXiv:hep-ph/0701129 [hep-ph]].
35. S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972).
36. M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, (OUP, Oxford, 2007).
37. F. Rohrlich, The definition of electromagnetic radiation, *Nuovo Cim* **21**, 811 (1961).
38. F. Rohrlich, *Classical Charged Particles* (World Scientific, Singapore, 2007).
39. C. Teitelboim, Splitting of the maxwell tensor - radiation reaction without advanced fields, *Phys. Rev. D* **1**, 1572 (1970) [erratum: *Phys. Rev. D* **2**, 1763 (1970)].
40. D. V. Gal'tsov and P. Spirin, Radiation reaction reexamined: Bound momentum and Schott term, *Grav. Cosmol.* **12**, 1 (2006).
41. B. P. Kosyakov, Radiation in electrodynamics and in Yang-Mills theory, *Sov. Phys. Usp.* **35**, 135 (1992).
42. P. A. Spirin, Massless field emission in the space-time of extra dimensions, *Grav. Cosmol.* **15**, 82 (2009).
43. D. V. Gal'tsov and M. Khlopunov, Synchrotron radiation in odd dimensions, *Phys. Rev. D* **101**, no. 8, 084054 (2020).
44. M. E. Pati and C. M. Will, PostNewtonian gravitational radiation and equations of motion via direct integration of the relaxed Einstein equations. 1. Foundations, *Phys. Rev. D* **62**, 124015 (2000).
45. Y. Z. Chu, Electromagnetic and Gravitational Radiation in All Dimensions: A Classical Field Theory Treatment, [arXiv:2107.14744 [gr-qc]].

Developments in numerical relativity and gravitational wave observations

Nigel T. Bishop
Department of Mathematics (Pure and Applied), Rhodes University, Makhanda 6139, South Africa
E-mail: n.bishop@ru.ac.za
www.ru.ac.za

This paper provides a summary of the fourteen talks that were presented in the session GW4 on various aspects of numerical relativity and computation concerning gravitational waves.

Keywords: Numerical relativity, gravitational waves

1. Introduction

A total of fourteen talks were presented over two days, covering a number of different topics within numerical relativity and computation concerning gravitational waves (GWs). Since the authors of each talk are publishing their work in these Proceedings or elsewhere, the purpose of this article is not to present any of the technical details of the talks. Rather, this session summary is aimed at presenting an overview of, and background information about, the various topics covered. As such, it is intended to be accessible to those working on other aspects of general realativty and astrophysics.

2. Inclusion of null infinity

Numerical relativity simulations of events such as compact object mergers, are normally carried out in a finite region, and the GW emission is estimated from the geometry near the outer boundary. However, because of the nonlinear nature of the Einstein equations, GWs are unambiguously defined only at future null infinity. Now the Einstein equations have the property that they can be compactified in a radial null direction and remain regular, so there are numerical approaches that compute GWs at future null infinity. The method commonly utilized is known as characteristic extraction, in which geometric data near the outer boundary of a merger simulation is used to provide boundary data to a characteristic code, i.e. one in which the spacetime is foliated by outgoing null cones. An alternative approach uses hyperboloidal slicing, in which the spacetime slices are spacelike in the central region, and become asymptotically null in the far region; however, this approach is less developed and has not yet produced gravitational waveforms.

Giannakopoulos *et al.*[1,2] presented results concerning the well-posedness of the Einstein equations in characteristic coordinates. In partial differential equation (PDE) theory, well-posedness in the standard L_2 norm is characterized by a system that is strongly (or symmetric) hyperbolic. However, it was shown that the characteristic Einstein equations form a PDE system that is only weakly hyperbolic. Further, it was demonstrated, using numerical examples, that weak hyperbolicity can have the effect of reducing the order of convergence of a code from that expected from considerations of the truncation error.

Gautam *et al.*[3] discussed a dual foliation formulation on hyperboloidal slices. There has been much effort over many years to implement a successful hyperboloidal scheme, but there remain problems of instability and of the need to regularize quantities that can become unbounded at null infinity. The essential idea presented is that these problems can be resolved by making a nonlinear change of variables, and the approach has been tested and shown to work in the special case of spherical symmetry. The approach is currently being extended to the general case, and if successful, this would be a significant achievement.

3. Surrogate models

Surrogate models have become an essential tool in searches in detector data for GWs from binary black hole mergers. Numerical relativity provides the most accurate waveforms, but they are computationally expensive and can take over a month for a single waveform. Since the merger phase space has, in principle, up to 7 dimensions, it is not feasible to populate it with numerical relativity waveforms. A surrogate waveform can be regarded as a mapping from (a restricted part of) the merger phase space to give a waveform. The quality of the model depends on how the mapping function is parametrized, and then fixing the parameter values so as to be consistent with known numerical relativity waveforms.

Islam *et al.*[4] reported a surrogate model for waveforms produced when the orbit is eccentric rather than circular. Eccentricity has long been neglected in waveform calculations, because it is well known that the reaction to GW emission damps out eccentricity at a rate that is much faster than that of the inspiral. However, there are possible astrophysical scenarios which could lead to the orbit being eccentric near merger, and the identification of such an event would be significant in terms of the understanding of formation channels of binary black hole systems. The model reported is restricted, at this stage, to non-spinning equal-mass black hole binary systems.

4. Effect of matter on GW propagation

As GWs pass through matter, some energy will be transferred to the matter causing the GW to be attenuated, but in astrophysical scenarios this effect is so small that it has been completely negelcted. However, recent work has shown that matter

can affect GW propagation in other ways, i.e. as a phase shift or as an echo. The possibility of a GW echo is interesting because there are (controversial) reports that echoes have been observed in GW data, caused by the remnant being some form of exotic compact object (ECO).

Naidoo *et al.*[5] applied a model of a GW source inside a shell of matter to GW events for which there are claims of echo observations. It was shown that any echo could not have been produced by the effect of a matter shell, since if it were the shell would have to be so massive as to constitute a black hole. Thus, confirmation of an echo in GW data would provide evidence in favour of the ECO hypothesis. However, it was also shown that there are astrophysical events, such as a binary neutron star merger and especially a core collapse supernova, for which the echo and phase-shift effects would alter the GW signal sufficiently for the difference to be measurable.

5. Hadron-quark phase transition in a binary neutron star merger

During the merger of two neutron stars, the matter becomes denser than nuclear matter and with a temperature of order $10^{12}\,^{\circ}$K, and it is feasible that a phase transition from hadronic matter to quark matter may occur. Such a phase transition would change the effective equation of state, and therefore would affect the motion of the matter and thus the emitted GW signal.

Hanauske *et al.*[6] considers whether a hadron-quark phase transition could lead to an observational signature of the event in the GW signal.[7–10] Computational models for a number of possible post-merger scenarios are constructed, and results obtained for the GW signal in runs to about 20ms. The results show that an increase in the frequency of the post-merger GW signal would indicate the occurrence of a phase transition.

6. Alternative theories of gravity

General relativity is consistent with every observation and experiment to date. However, there are reasons for investigating alternative theories. General relativity is not consistent with quantum mechanics, so it needs to be modified in a way that would affect its properties at a very small scale. On the other hand, at the very large scale of cosmology, the issues of the unknown nature of dark matter and dark energy may be resolved by using an alternative theory. Clearly, any alternative theory would need to be consistent with observations including those by LIGO/Virgo, and that motivates the need to calculate GW emission from compact object coalescence in the theory.

Lim *et al.*[11] reported the construction of a numerical code for an effective field theory with a Lagrangian that is that of general relativity plus terms that are quadratic in the Ricci tensor. The resulting field equations involve up to 4^{th} order derivatives. The code uses the harmonic gauge, and is restricted to the case

of spherical symmetry. It has been tested with initital data that is Minkowski or Schwarzschild plus noise, and has been shown to be stable and convergent.

Khlopunov *et al.*[12] calculates the GW emission due to a system of two point masses on a 3-brane embedded in a 5-dimensional spacetime, using analytical methods. In the non-relativistic limit, a generalization of the quadrupole formula is obtained.

7. GW memory

The memory effect in GW theory is well-known. Consider a system of free particles, each particle being at rest relative to the others, in a region of spacetime that can be regarded as initially Minkowskian. Supose that the particles are now perturbed by GWs, and that afterwards the spacetime region is again Minkowskian. Then the particles will be at rest relative to each other, but in general the positions of the particles relative to each other will have changed.

Grant *et al.*[13,14] generalizes the memory concept to the case that the particles have initial relative velocity and acceleration (due to forces other than gravity), and obtains formulas in terms of the GWs and non-radiative quantities for the memory-like changes to these quantities. In the case of zero relative velocity and acceleration, the formulas reduce to what is already known.

8. Searches for intermediate mass black hole (IMBH) mergers in GW data

IMBHs are usually defined as being in the mass range $10^2 M_\odot \leq M \leq 10^5 M_\odot$. They are less massive than the supermassive black holes at galactic centres, and too massive to have been formed by stellar collapse, because the pair instability gap means that stellar collapse cannot produce a black hole in the mass range $50 M_\odot \lessapprox M \lessapprox 130 M_\odot$. The event GW190521 is estimated to have been caused by the merger of black holes with approximate masses $85 M_\odot$ and $66 M_\odot$ yielding a remnant with mass about $142 M_\odot$, providing conclusive evidence for the existence of IMBHs. This observation has motivated searches that focus on identifying IMBH events in LIGO/Virgo data.

PyCBC is a free and open software package used to search for GW events in detector data. Chandra *et al.*[15] reported a PyCBC-based search that is optimized to identify IMBH events. The search re-identified GW190521 with a much better false alarm rate than that of the original search.

Gayathri *et al.*[16] presented, on behalf of the LIGO, Virgo and KAGRA collaborations, the results of their searches for IMBH events in the observing period O3.

9. Searches for a GW signal of a spinning neutron star

An object spinning with frequency f, unless it is exactly axisymmetric, emits GWs at frequency $2f$. Pulsars are spinning neutron stars that emit regular radio pulses,

and thus have rotational frequencies that are accurately determined. However, searches for GW signals from a number of pulsars have not yielded any positive result, and have only been able to place upper bounds on the magnitude of such signals. It is well-known that glitches occur in the radio pulses, indicating that some form of deformation is occurring in the neutron star. Thus it can be expected that there are periods of time during which the neutron star would be emitting GWs, but when such periods occur and their duration is not known.

Fesik *et al.*[17,18] presented a method for localising a long-duration signal in GW data, that is, to estimate the start time and duration of the signal. The method applies when the signal duration is much longer than the wave period, so more than a few hours. This knowledge can then be used to improve the signal-to-noise ratio of the recovered signal.

10. Machine learning

In recent years, the use of machine learning has become more and more commonplace throughout science, and this certainly includes applications to GW calculations and GW data analysis. The key advantages of machine learning are that: (a) in many circumstances it has proven to be remarkably accurate; and (b) there are now a number of software packages available, so from a user perspective, implementation can be straightforward. A common application of machine learning is the construction of a function that maps an input vector $\mathbf{x}$ to an output vector $\mathbf{y}$, i.e $\mathbf{y} = \mathcal{F}(\mathbf{x})$. We need to know the expected output $\mathbf{z}_i$ in a number of cases, and then construct $\mathcal{F}$ by minimizing the error $||\mathbf{z}_i - \mathbf{y}_i||$; this process is called *training* the network, and the quality of the network is described by the magnitude of the error, particularly for data points that are not used in the training.

Mishra[19] applied machine learning to search for a coherent WaveBurst in GW detector data. The network is trained on calculated waveforms of binary black hole mergers; the complete waveform is not used, but rather certain statistics representing the waveform are calculated, and these are used for the training. The resulting network is reported to improve the detection of binary black hole merger events by between 15% and 25%.

Field *et al.*[20,21] use machine learning to determine orbital dynamics of binary black hole systems from GW measurements. The system is written in the form of a universal differential equation, $\dot{\mathbf{x}} = \mathcal{F}(\mathbf{x})$, where $\mathbf{x}$ is a vector with components representing various aspects of the system dynamics. Once a choice of $\mathcal{F}$ is made, $\mathbf{x}$ is determined and then the quadrupole formula is applied to find the emitted GWs. Given observed (or numerically calculated) GW data from a binary black hole system, the next step is network training, i.e. to find $\mathcal{F}$ by minimizing the mismatch between this data and that associated with the choice made for $\mathcal{F}$. It was found that this model makes quite accurate waveform predictions outside the time interval over which the model is trained.

References

1. Thanasis Giannakopoulos, Nigel T. Bishop, David Hilditch, Denis Pollney and Miguel Zilhao, *Hyperbolicity of General Relativity in null foliations*, Book of Abstracts, p. 325 (2021).
2. Thanasis Giannakopoulos, David Hilditch and Miguel Zilhao, *Hyperbolicity of General Relativity in Bondi-like gauges*, Phys. Rev. D **102**, 064035 (2020).
3. Shalabh Gautam, Alex Vañó-Viñuales, Edgar Gasperin, David Hilditch and Sukanta Bose, *Einstein Field Equations in Spherical Symmetry on Hyperboloidal Slices: Dual Foliation Formulation*, Book of Abstracts, p. 324 (2021).
4. Tousif Islam, Vijay Varma, Jackie Lodman, Scott Field, Gaurav Khanna, Mark Scheel, Harald Pfeiffer, David Gerosa and Lawrence Kidder, *Eccentric binary black hole surrogate models for the gravitational waveform andremnant properties: comparable mass, nonspinning case*, Book of Abstracts, p. 328 (2021).
5. M Naidoo, Nigel T. Bishop and Petrus van der Walt, *Matter shells modifying gravitational wave signals*, Book of Abstracts, p. 328, and in these Proceedings (2021).
6. Matthias Hanauske, Gloria Montana, Lukas Weih and Luciano Rezzolla, *Gravitational-wave signatures of the hadron-quark phase transition in binary compact star mergers*, Book of Abstracts, p. 326 (2021).
7. M. Hanauske and L. R. Weih, *Neutron star collisions and gravitational waves*, Astronomische Nachrichten **342**, 788 (June 2021).
8. L. R. Weih, M. Hanauske and L. Rezzolla, *Postmerger Gravitational-Wave Signatures of Phase Transitions in Binary Mergers*, Physical review letters **124**, p. 171103 (May 2020).
9. E. R. Most, L. J. Papenfort, V. Dexheimer, M. Hanauske, S. Schramm, H. Stöcker and L. Rezzolla, *Signatures of Quark-Hadron Phase Transitions in General-Relativistic Neutron-Star Mergers*, Physical review letters **122**, p. 061101 (February 2019).
10. G. Montaña, L. Tolós, M. Hanauske and L. Rezzolla, *Constraining twin stars with GW170817*, Physical Review D **99**, p. 103009 (May 2019).
11. Hyun Lim and Aaron Held, *Dynamics of Quadratic Gravity in Spherical Symmetry*, Book of Abstracts, p. 323 (2021).
12. Mikhail Khlopunov and Dmitry Galtsov, *Odd-dimensional gravitational waves from the binary system on three-brane*, Book of Abstracts, p. 329, and in these Proceedings (2021).
13. Alexander Grant and David Nichols, *Memory-like effects due to relative velocity and acceleration*, Book of Abstracts, p. 327 (2021).
14. Alexander Grant and David Nichols, *Persistent gravitational wave observables: Curve deviation in asymptotically flat spacetimes*, `https://arxiv.org/abs/2109.03832` (2021).
15. Koustav Chandra, Veronica Villa-Ortega, Thomas Dent, Connor McIsaac, Archana Pai, Ian Harry, Gareth Davies and Kanchan Soni, *An optimised PyCBC search for gravitational waves from intermediate-mass black hole mergers*, Book of Abstracts, p. 324, and in these Proceedings (2021).
16. V. Gayathri, The LIGO Scientific Collaboration, The Virgo Collaboration and The KAGRA Collaboration, *Intermediate mass black hole search in LIGO-Virgo's third observing period*, Book of Abstracts, p. 325 (2021).
17. Liudmila Fesik and Maria Alessandra Papa, *Localisation of a long-duration gravitational wave signal in time*, Book of Abstracts, p. 327 (2021).
18. Liudmila Fesik and Maria Alessandra Papa, *The detectability of long-duration gravitational wave signals*, Book of Abstracts, p. 327 (2021).

19. Tanmaya Mishra, *Optimization of model independent gravitational wave search using machine learning*, Book of Abstracts, p. 326 (2021).
20. Scott Field, Brendan Keith and Akshay Khadse, *Learning orbital dynamics of binary black hole systems from gravitational wave measurements*, Book of Abstracts, p. 323 (2021).
21. Brendan Keith, Akshay Khadse and Scott E. Field, *Learning orbital dynamics of binary black hole systems from gravitational wave measurements*, Phys. Rev. Research **3**, 043101 (2021).

Insights into the Galactic Center environment from VHE gamma-ray observations with ground-based facilities

C. Fruck*

Physics Department, Technical University Munich,
James-Franck-Str. 1, 85748, Garching, Germany
E-mail: cfruck@ph.tum.de
www.ph.tum.de

The dynamic center of our galaxy is known to host a source of TeV gamma rays since the very beginning of the 21st century and a link to the supermassive black hole at the Galactic Center has been speculated on ever since. But not only the point-like source, spatially coincident with SgrA*, can be observed from the ground using the Imaging Air Cherenkov Telescope technique, but also diffuse emission from the vicinity, spanning more than one degree along the Galactic plane and emitting a remarkably hard energy spectrum, reaching energies well beyond 10 TeV.

Recent observations by the H.E.S.S., MAGIC and VERITAS facilities have enabled detailed studies of the dynamics of high-energy particles in Galactic Center region that indicate a link between the diffuse component and central point-like gamma-ray source. These studies suggest the presence of a powerful cosmic-ray accelerator in close proximity to SgrA*. This could potentially even be one of the long-sought-after Galactic PeVatrons, needed in order to explain the cosmic-ray spectrum up to the the feature called 'knee' at around 10^{15} eV.

Keywords: Galactic Center, SMBH, VHE gamma-ray astronomy

1. Introduction

The Milky Way Galaxy falls into the category of bared spiral galaxies, with a moderate star forming rate within the spiral arms and a significantly lower star forming rate inside the bar and bulge, towards its dynamic center. The bar and bulge span 1.5 and 4.4 kpc in size, respectively, while the the galactic disk extends over more than 30 kpc. A much smaller region at the center of the Galactic bulge contains a much higher density of young stars when compared to the surrounding regions. Such a high star-formation rate can only be maintained by a high density of molecular gas in this region, which is referred to as the Central Molecular Zone (CMZ) of the Milky Way Galaxy. The high star formation rate inside the CMZ is also the reason for the high density of high energy astrophysical phenomena being observed in this area of the sky. Those include pulsar wind nebulae (PWN), supernova remnants (SNR)[1] and presumably also millisecond pulsars (MSP).[2] All those types of

*Now at: Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

objects can potentially accelerate charged particles to TeV energies and beyond, which can in turn produce very high energy (VHE) gamma rays through secondary processes like Bremsstrahlung or inverse Compton scattering in case of electrons or deep inelastic scattering in case of protons.

Another potential source of VHE gamma rays is the strong accumulation of dark matter (DM) that is expected around the Galactic Center (GC).[3] Some models predict DM particles to interact or decay with gamma-rays in the final state. The high expected DM density at the GC, paired with its relative proximity makes it one of the prime observational targets for indirect DM searches.[4] Strong foregrounds from astrophysical sources on the other hand complicate the search.

At the very center of the Milky Way Galaxy's rotation, spanning less than a 0.1 pc in size, resides the so-called S-star cluster with numerous O and B-type stars, whose rapid movement can be traced over a time scale of years with high resolution observations in the infrared.[5] The velocity of these stars and the tight elliptical orbits around a common dynamical center allow to quantify the central mass concentration, $\approx 4.4 \cdot 10^6\,\mathrm{M}_\odot$, and to exclude nearly all other explanations, except for the presence of a super-massive black hole (SMBH) at the dynamic center of our galaxy.[5] This mass concentration is spatially consistent with the point-like radio source SgrA*,[6] thus probably caused by non-thermal emission from the vicinity of the black hole (BH). The gravitational radius of such a BH would be

$$r_\mathrm{G} = \frac{2\,G\,M}{c^2} = 1.3 \cdot 10^{12}\mathrm{cm} \approx 10\,\mu\,\mathrm{arcsec}\,, \tag{1}$$

which is within reach for the most ambitious very-long-baseline interferometry (VLBI) campaigns, such as the *Event Horizon Telescope* (EHT) that utilizes a global network of radio telescopes.[7]

SgrA* is also visible in X-ray,[8] adding further evidence for non-thermal processes being at work and the acceleration of charged particles to high energies. In radio, infrared, as well as in X-rays, SgrA* shows minute to day-scale variability, changing its brightness over several orders of magnitude.[9,10] These observations indicate that the site of emission of non-thermal radiation must be comparable to the size of the BH itself and thus likely located in close proximity to the event horizon. Also from X-ray observations, echos on the scale of tens of parsecs have been identified within the surrounding structures of dense gas. These are believed to stem from past very strong X-ray flares caused by SgrA*, now visible as echos that can be seen traveling through the surrounding environment over a time scale of several years.[11] Another hint for an episode in the more distant past of strongly increased intensity of acceleration of high energy charged particles at the Galactic Center are the giant bubble-like structures found in X-rays[12] and gamma rays,[13] extending far outside the galactic disk in both directions. These structures can not be directly linked to SgrA* with certainty, but could potentially be the result of increased accretion activity onto the GC black hole.

2. VHE Gamma-Ray Observations from the Ground

The very dense and active astrophysical environment within the CMZ and especially the link between the SMBH, the closest to Earth by a large margin) and non-thermal processes visible in radio and X-ray, make the GC a very interesting target for observations in gamma rays. The *Fermi* large area telescope (LAT) in low earth orbit since 2008 is the most sensitive all-sky detector for gamma-rays in the energy range between 100 MeV and about 1 TeV.[14] While the GC hosts strong sources found also in the the *Fermi* source catalog, including a point source coincident with SgrA*,[15] the spatial resolution in the energy range around 10 GeV below is rather poor, while the effective collection area needed to cope with the low flux above that energy, is way large than the only about 1 m^2 offered by *Fermi*.

This is where ground-based facilities come into play, that in fact predate *Fermi* by more than a decade. The most successful class of ground-based instruments for observations in the very high energy (VHE) range between about 50 GeV and 100 TeV are so-called imaging air Cherenkov telescopes (IACTs). IACTs detect short ($\approx$ 3 ns long) and faint (hundred to a couple of ten thousand photons collected by the telescope) flashes of Cherenkov light produced in extensive air showers produced by gamma-rays interacting with the atmosphere generating charged leptons. These Cherenkov light flashes, mostly generated in the lower stratosphere and high troposphere, illuminate an area of several ten thousand m^2 on the ground at sea level. The light is collected by optical telescopes with simple geometry but large mirror area and fast as well as sensitive cameras, capable of time resolution on the order of nanoseconds. The reconstruction of the original direction of the VHE photons is strongly helped by employing stereoscopy and observing the same event with two or more telescopes placed within the light pool. The technique was pioneered by the Whipple[16] and HEGRA[17] instruments, but became a more established branch of astronomy with the next generation of telescopes. The three most important of those installations are the H.E.S.S.,[18] MAGIC[19] and VERITAS[20] experiments, located in Namibia, La Palma and the US, respectively.

The first detection of the GC in VHE gamma rays was already made with Whipple in 2004,[21] quickly followed up by detections with H.E.S.S. (2004)[22] and MAGIC (2006).[23] Later deep observations by H.E.S.S. have reveled further details of the GC VHE emission. In 2006, after subtracting the point sources at SgrA* and G0.9+0.1, the presence of diffuse emission throughout the CMZ has been reveled, spanning about 200 pc in width along the Galactic plane. The emission has been attributed to high energy protons interacting with the dense molecular clouds in the region.[24] The position estimate of the Galactic Center point source could be greatly refined, far below the gamma-ray PSF of the instrument. This lead to the exclusion of the SNR SgrA East as other potential source apart from SgrA*.[25] The spectral shape of the GC point source has been measured with high precision, yielding an index of around 2.2 and an exponential cut off at $\approx$ 10 TeV.[26]

3. Recent Observational Results

The recent increased interest that the GC received in VHE gamma rays has been sparked by the discovery of a gas cloud in 2011, on a highly eccentric orbit towards SgrA* in infrared, that led to speculations about a potential increase in accretion rate and possibly also flaring activity of the non-thermal emission of the object.[27] The pericenter passage was predicted to happen in 2013/14 with possible effects on the accretion rate around the time or later.[28] These revelations lead to a multi-year observation campaign by MAGIC, starting in 2012 and leading to a publication of the results covering the first four years of monitoring the gamma-ray source.[29] No variability of the GC point source has been detected until 2015, neither at energies above 1 TeV, nor at higher energies above 10 TeV, where some hadronic emission models predict a much faster reaction of the gamma-ray flux to newly injected particles compared to lower energy gamma rays.[30]

In 2016 the H.E.S.S. collaboration probed the extended emission from the Galactic Ridge in great detail, reveling that in contrast to the spectrum of the point-source, the gamma-ray spectrum emitted by this region reaches more than 30 TeV in energy without a hint for a cut-off.[31] Assuming proton-proton interactions as source for the emission, proton energies would thus not cut off before reaching around 1 PeV in energy. Another important finding from those observations is that the flux density of protons required to explain the observed gamma-ray emission profile, under the assumption of a density profile of molecular gas from CO emissions, best fits a $1/r$ shape, where r is the distance from the location of SgrA*.[31] Such an observation is compatible with cosmic rays (CRs) being accelerated in the vicinity of SgrA* and undergoing diffusion processes on their path away from the source. Combining both of those observations provides evidence that the GC might indeed host a Galactic PeVatron, potentially even linked to the SMBH SgrA*.

These findings were largely confirmed in 2020 by a new analysis of a large set of MAGIC data collected around the GC and CMZ. A new data analysis tool (SkyPrism[32]) for extended sources that allows fitting of 2D spatial models to gamma-ray sky maps binned in energy, similar to what is done by the *Fermi* tools,[33] was utilized for this analysis. This way individual energy spectra for 3 point-like (SgrA*, the PWN G0.9+01 and a recently discovered source inside the Galactic Radio Arc) and one extended source (the diffuse Galactic Ridge emission) could be extracted.[34] The energy spectrum of the diffuse component, in contrast to earlier findings by H.E.S.S., in this case shows a 2σ hint for a cut-off at around 20 TeV. It should however be noted that the regions probed and the methods used differ for both results. While the H.E.S.S. analysis extracted events from a cut anulus with inner radius of 0.1 and outer radius of 0.4 degrees around SgrA* and excluding the Arc source ("pacman" region),[31] the MAGIC analysis used a CS radio map that extends about 1.4 degrees and 0.8 along the Galactic plane as spatial template for this emission component.[34] Earlier in 2018 in second H.E.S.S publication concerning

the gamma-ray emission from the CMZ, a similar extraction region has been used for the spectrum of the diffuse component, showing a spectral shape compatible to the one found with MAGIC, exhibiting a similar downward trend above 10 TeV but not further commenting on it. Also the roughly $1/r$ spatial profile of the underlying proton flux density has been confirmed by MAGIC,[34] tending to somewhat higher values of α (≈ 1.2), especially when assuming an underlying constant CR-sea component in addition to the $1/r^{\alpha}$ profile.

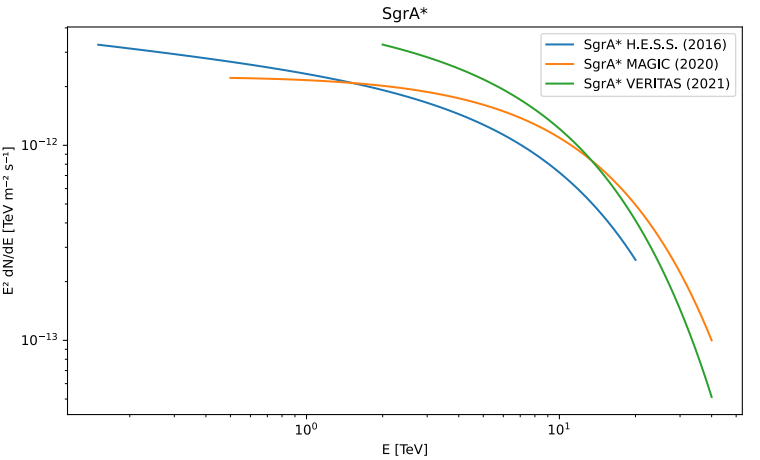


Fig. 1. Best fit SEDs of the point source coincident with SgrA* measured with H.E.S.S.,[31] MAGIC[34] and VERITAS.[35]

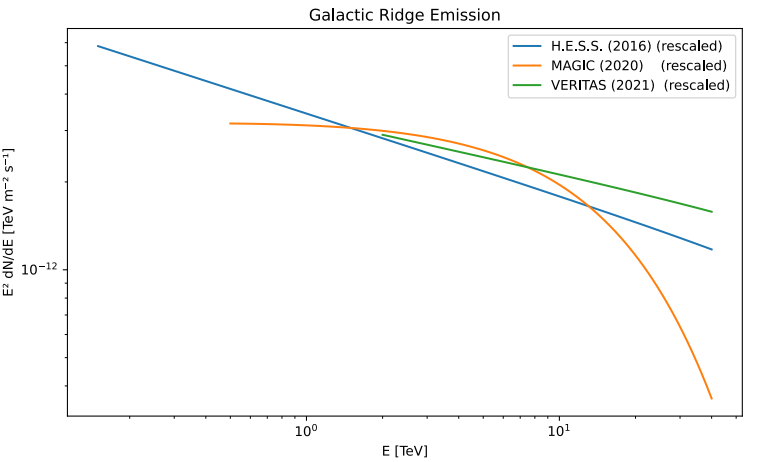


Fig. 2. Best fit SEDs of the Galactic Ridge emission measured with H.E.S.S.,[1] MAGIC[34] and VERITAS.[35]

The latest results of the current VERITAS GC observation campaign have been published in 2021, showing a nearly 10 y light curve of SgrA*, without a hint for variability, neither at energies above 2 TeV, nor above 5 TeV.[35] This study also presented an energy spectrum of the diffuse emission from the CMZ, extracted from Circular regions along the Galactic plane, but avoiding close proximity to SgrA*. Despite the complementary nature compared to the "pacman" region probed by H.E.S.S. (2016)[31] and the overlap with H.E.S.S. (2018)[1] and MAGIC (2020),[34] here no hint for a cut-off has been detected, with the spectrum continuing with an index around -2.2 until 30 TeV.

The best-fit spectral energy distributions (SEDs) from recent observations of the Galactic Center region[1,31,34,35] are illustrated in Figures 1 (spectrum of SgrA*) and 2 (Galactic Ridge). The SEDs measured for the Galactic Ridge have been slightly re-scaled in flux for better comparison, as not all correspond to the exact same extraction region. The Galactic Ridge spectra measured with H.E.S.S. and VERITAS have been fitted with a power law, while the MAGIC spectrum corresponds to a power law with exponential cut-off.

4. Conclusions and Outlook

The VERITAS result[35] highlights once more that the debate about the nature of the GC gamma-ray source and and CR accelerator is still open and further observations are necessary in order to gain even better insight to the underlying physics. It is unclear weather the current generation of Cherenkov telescopes are sufficient for the task at hand, or if new discoveries will mainly rely on the next generation of IACTs, the telescopes of the Cherenkov Telescope Array (CTA) currently under construction.[36] Also instruments utilizing the water Cherenkov technique like the recently commissioned HAWC[37] and LHAASO[38] could play an important role in the future. While on one hand lacking the angular resolution, those instruments do not suffer from the limited field of view of IACTs and can accumulate vast amounts of data with relatively low irreducible background. This makes them perfectly suited for putting the GC and CMZ into context with the larger scale emission from the rest of the Galactic plane. Being located on the northern hemisphere, HAWC and LHAASO themselves are not ideally suited for the task, but the upcoming Southern Wide-Field Gamma-Ray Observatory (SWGO) might well be.[39]

References

1. H. E. S. S. Collaboration *et al.*, Characterising the VHE diffuse emission in the central 200 parsecs of our Galaxy with H.E.S.S., *Astronomy and Astrophysics* **612**, p. A9 (April 2018).
2. R. Bartels, S. Krishnamurthy and C. Weniger, Strong Support for the Millisecond Pulsar Origin of the Galactic Center GeV Excess, *Physical Review Letters* **116**, p. 051102 (February 2016).
3. O. Y. Gnedin and J. R. Primack, Dark Matter Profile in the Galactic Center, *Physical Review Letters* **93**, p. 061302 (August 2004).

4. L. Bergström, P. Ullio and J. H. Buckley, Observability of gamma rays from dark matter neutralino annihilations in the Milky Way halo, *Astroparticle Physics* **9**, 137 (August 1998).
5. R. Genzel, F. Eisenhauer and S. Gillessen, The Galactic Center massive black hole and nuclear star cluster, *Reviews of Modern Physics* **82**, 3121 (December 2010).
6. K. Y. Lo, D. C. Backer, R. D. Ekers, K. I. Kellermann, M. Reid and J. M. Moran, On the size of the galactic centre compact radio source: diameter <20 AU, *Nature* **315**, 124 (May 1985).
7. D. Castelvecchi, How to hunt for a black hole with a telescope the size of Earth, *Nature* **543**, 478 (March 2017).
8. F. K. Baganoff *et al.*, Rapid X-ray flaring from the direction of the supermassive black hole at the Galactic Centre, *Nature* **413**, 45 (September 2001).
9. D. Haggard, M. Nynka, B. Mon, N. d. l. C. Hernandez, M. Nowak, C. Heinke, J. Neilsen, J. Dexter, P. C. Fragile, F. Baganoff, G. C. Bower, L. R. Corrales, F. C. Zelati, N. Degenaar, S. Markoff, M. R. Morris, G. Ponti, N. Rea, J. Wilms and F. Yusef-Zadeh, Chandra Spectral and Timing Analysis of Sgr A\ast's Brightest X-Ray Flares, *The Astrophysical Journal* **886**, p. 96 (November 2019).
10. T. Do *et al.*, Unprecedented Near-infrared Brightness and Variability of Sgr A\ast, *The Astrophysical Journal* **882**, p. L27 (September 2019).
11. M. Clavel, R. Terrier, A. Goldwurm, M. R. Morris, G. Ponti, S. Soldi and G. Trap, Echoes of multiple outbursts of Sagittarius A* revealed by Chandra, *Astronomy & Astrophysics* **558**, p. A32 (October 2013).
12. P. Predehl *et al.*, Detection of large-scale X-ray bubbles in the Milky Way halo, *Nature* **588**, 227 (December 2020).
13. M. Su, T. R. Slatyer and D. P. Finkbeiner, Giant Gamma-Ray Bubbles from Fermi-Lat: Active Galactic Nucleus Activity or Bipolar Galactic Wind?, *The Astrophysical Journal* **724**, 1044 (November 2010).
14. W. B. Atwood *et al.*, The large area telescope on the fermi gamma-ray space telescope mission, *The Astrophysical Journal* **697**, 1071 (May 2009).
15. S. Abdollahi *et al.*, Fermi Large Area Telescope Fourth Source Catalog, *The Astrophysical Journal Supplement Series* **247**, p. 33 (March 2020).
16. J. e. a. Kildea, The Whipple Observatory 10m gamma-ray telescope, 1997–2006, *Astroparticle Physics* **28**, 182 (October 2007).
17. H. Collaboration *et al.*, First Results on the Performance of the HEGRA IACT Array, *arXiv:astro-ph/9704098* (April 1997).
18. W. Benbow, The Status and Performance of H.E.S.S., *AIP Conference Proceedings* **745**, 611 (February 2005).
19. J. Aleksić *et al.*, The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system, *Astroparticle Physics* **72**, 61 (January 2016).
20. N. Park, Performance of the VERITAS experiment, in *Proceedings of The 34th International Cosmic Ray Conference — PoS(ICRC2015)*, (SISSA Medialab, August 2016).
21. K. Kosack *et al.*, TeV Gamma-Ray Observations of the Galactic Center, *The Astrophysical Journal* **608**, p. L97 (May 2004).
22. F. Aharonian *et al.*, Very high energy gamma rays from the direction of Sagittarius A*, *Astronomy and Astrophysics* **425**, L13 (October 2004).
23. J. Albert *et al.*, Observation of Gamma Rays from the Galactic Center with the MAGIC Telescope, *The Astrophysical Journal* **638**, p. L101 (January 2006).

24. F. Aharonian *et al.*, Discovery of very-high-energy gamma-rays from the Galactic Centre ridge, *Nature* **439**, 695 (February 2006).
25. HESS Collaboration *et al.*, Localizing the VHE gamma-ray source at the Galactic Centre, *Monthly Notices of the Royal Astronomical Society* **402**, 1877 (March 2010).
26. F. Aharonian *et al.*, Spectrum and variability of the Galactic center VHE gamma-ray source HESS J1745–290, *Astronomy & Astrophysics* **503**, 817 (September 2009).
27. S. Gillessen *et al.*, A gas cloud on its way towards the supermassive black hole at the Galactic Centre, *Nature* **481**, 51 (January 2012).
28. S. Gillessen *et al.*, New Observations of the Gas Cloud G2 in the Galactic Center, *The Astrophysical Journal* **763**, p. 78 (January 2013).
29. M. L. Ahnen *et al.*, Observations of Sagittarius A* during the pericenter passage of the G2 object with MAGIC, *Astronomy & Astrophysics* **601**, p. A33 (May 2017).
30. D. R. Ballantyne, M. Schumann and B. Ford, Modelling the time-dependence of the TeV gamma-ray source at the Galactic Centre, *Monthly Notices of the Royal Astronomical Society* **410**, 1521 (January 2011).
31. H. E. S. S. collaboration *et al.*, Acceleration of petaelectronvolt protons in the Galactic Centre, *Nature* **531**, 476 (March 2016).
32. I. Vovk, M. Strzys and C. Fruck, Spatial likelihood analysis for MAGIC telescope data - From instrument response modelling to spectral extraction, *Astronomy & Astrophysics* **619**, p. A7 (November 2018).
33. Fermi Science Support Development Team, Fermitools: Fermi Science Tools, *Astrophysics Source Code Library* , p. ascl:1905.011 (May 2019).
34. V. A. Acciari *et al.*, MAGIC observations of the diffuse gamma-ray emission in the vicinity of the Galactic center, *Astronomy & Astrophysics* **642**, p. A190 (October 2020).
35. C. B. Adams *et al.*, VERITAS Observations of the Galactic Center Region at Multi-TeV Gamma-Ray Energies, *The Astrophysical Journal* **913**, p. 115 (June 2021).
36. J. Knödlseder, The Cherenkov Telescope Array, *arXiv:2004.09213 [astro-ph]* (April 2020).
37. A. U. Abeysekara *et al.*, The 2HWC HAWC Observatory Gamma-Ray Catalog, *The Astrophysical Journal* **843**, p. 40 (June 2017).
38. X. Bai *et al.*, The Large High Altitude Air Shower Observatory (LHAASO) Science White Paper, *arXiv:1905.02773 [astro-ph]* (May 2019).
39. P. Abreu *et al.*, The Southern Wide-Field Gamma-Ray Observatory (SWGO): A Next-Generation Ground-Based Survey Instrument for VHE Gamma-Ray Astronomy, *arXiv:1907.07737 [astro-ph]* (July 2019).

The TAIGA experiment

M. Tluczykont
Institut für Experimentalphysik, Universität Hamburg, Hamburg, D-22 761 Germany
martin.tluczykont@physik.uni-hamburg.de

I.I. Astapov
National Research Nuclear University MEPhI, Moscow, 115409 Russia

A.K. Awad
Institut für Experimentalphysik, Universität Hamburg, Hamburg, D-22 761 Germany

P.A. Bezyazeekov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

M. Blank
Institut für Experimentalphysik, Universität Hamburg, Hamburg, D-22 761 Germany

E.A. Bonvech
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.N. Borodin
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

A.V. Bulan
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

M. Brückner
Deutsches Elektronen-Synchrotron DESY, Zeuthen, 15738 Germany

N.M. Budnev
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A. Chiavassa
Physics Department of the University of Torino and Istituto Nazionale di Fisica Nucleare, Torino, 10125 Italy

D.V. Chernov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.N. Dyachok
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.R. Gafarov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.Yu. Garmash
Novosibirsk State University, Novosibirsk, 630090 Russia

Budker Institute of Nuclear Physics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

V.M. Grebenyuk
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Dubna University, Dubna, Moscow oblast, 141980 Russia

O.A. Gress
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

E. Gress
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

T.I. Gress
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

O.G. Grishin
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.A. Grinyuk
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

D. Horns
Institut für Experimentalphysik, Universität Hamburg, Hamburg, D-22 761 Germany

A.L. Ivanova
Novosibirsk State University, Novosibirsk, 630090 Russia

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

N.N. Kalmykov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.V. Kindin
National Research Nuclear University MEPhI, Moscow, 115409 Russia

S.N. Kiryuhin
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

R.P. Kokoulin
National Research Nuclear University MEPhI, Moscow, 115409 Russia

K.G. Kompaniets
National Research Nuclear University MEPhI, Moscow, 115409 Russia

E.E. Korosteleva
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.A. Kozhin
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

E.A. Kravchenko
Novosibirsk State University, Novosibirsk, 630090 Russia

Budker Institute of Nuclear Physics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

A.P. Kryukov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

L.A. Kuzmichev
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.A. Lagutin
Altai State University, Barnaul, Altai krai, 656049 Russia

M. Lavrova
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

E. Lemeshev
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

B.K. Lubsandorzhiev
Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

N.B. Lubsandorzhiev
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.D. Lukanov
Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

D.S. Lukyantsev
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

R.R. Mirgazov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

R. Mirzoyan
Max Planck Institute for Physics, Munich, 80805 Germany

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

R.D. Monkhoev
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

E.A. Osipova
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.L. Pakhorukov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A. Pan
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

L.V. Pankov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.D. Panov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.A. Petrukhin
National Research Nuclear University MEPhI, Moscow, 115409 Russia

D.A. Podgrudkov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.A. Poleschuk
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

E.G. Popova
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A. Porelli
Deutsches Elektronen-Synchrotron DESY, Zeuthen, 15738 Germany

E.B. Postnikov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.V. Prosin

Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.S. Ptuskin

Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, Russian Academy of Sciences, Troitsk, Moscow, 142190 Russia

A.A. Pushnin

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

R.I. Raikin

Altai State University, Barnaul, Altai krai, 656049 Russia

A.Y. Razumov

Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

G.I. Rubtsov

Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

E.V. Ryabov

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

Y.I. Sagan

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Dubna University, Dubna, Moscow oblast, 141980 Russia

V.S. Samoliga

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

I. Satyshev

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Yu.A. Semeney

Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.A. Silaev

Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.A. Silaev(junior)

Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.Yu. Sidorenkov

Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

A.V. Skurikhin
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

A.V. Sokolov
Novosibirsk State University, Novosibirsk, 630090 Russia

Budker Institute of Nuclear Physics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

Y. Suvorkin
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

L.G. Sveshnikova
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

V.A. Tabolenko
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

A.B. Tanaev
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

B.A. Tarashansky
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

M. Ternovoy
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

L.G. Tkachev
Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Dubna University, Dubna, Moscow oblast, 141980 Russia

N. Ushakov
Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

A. Vaidyanathan
Novosibirsk State University, Novosibirsk, 630090 Russia

P.A. Volchugov
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

N.V. Volkov
Altai State University, Barnaul, Altai krai, 656049 Russia

D. Voronin
Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

R. Wischnewski
Deutsches Elektronen-Synchrotron DESY, Zeuthen, 15738 Germany

I.I. Yashin
National Research Nuclear University MEPhI, Moscow, 115409 Russia

A.V. Zagorodnikov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

D.P. Zhurov
Research Institute of Applied Physics, Irkutsk State University, Irkutsk, 664003 Russia

Irkutsk National Research Technical University, Irkutsk, Russia

The Tunka Advanced Instrument for Gamma-ray and cosmic ray Astrophysics (TAIGA) is a hybrid experiment for the measurement of Extensive Air Showers (EAS) with good spectral resolution in the TeV to PeV energy range. In this domain, the long-sought Pevatrons can be detected. Currently the hybrid TAIGA detector combines two wide angle shower front Cherenkov light sampling timing arrays (HiSCORE and Tunka-133), two 4m class, 10° aperture Imaging Air Cherenkov Telescopes (IACTs) and 240 m^2 surface and underground charged particle detector stations. Our goal is to introduce a new hybrid reconstruction technique, combining the good angular and shower core resolution of HiSCORE with the gamma-hadron separation power of imaging telescopes. This approach allows to maximize the effective area and simultaneously to reach a good gamma-hadron separation at low energies (few TeV). At higher energies, muon detectors are planned to enhance gamma-hadron separation. During the commissioning phase of the first and second IACT, several sources were observed. First detections of known sources with the first telescope show the functionality of the TAIGA IACTs. Here, the status of the TAIGA experiment will be presented, along with first results from the current configuration.

Keywords: Gamma-rays: experiments; Observations: Crab Nebula.

1. Introduction

TAIGA stands for Tunka Advanced Instrument for Gamma-ray and cosmic ray Astrophysics, and is a hybrid instrument designed to access the gamma-ray energy range from approx. 1 TeV to several 100s of TeV. The experiment is located at an altitude of 675m above sea level on the site of the Tunka-133[1] cosmic ray array (51°48'35"N, 103°04'02"E). Observations in this energy range are motivated by several questions of Astroparticle and particle physics. The presence of cosmic ray (CR) accelerators within our Galaxy can be inferred from observations of diffuse and extended gamma-ray emission in the High-Energy (HE)[2] and Very-High-Energy (VHE) range[3–5] e.g. These Gamma-ray emissions result from the decay of neutral

pions, which are produced in interactions of the accelerated cosmic rays with the ambient interstellar medium, i.e. molecular gas clouds, which effectively act as CR tracers.[6] The gamma-ray energy range from 10 TeV to few 100 TeV is crucial for the search for the Galactic Pevatrons, which accelerate cosmic ray particles up to knee-energies (approx. 3 PeV proton energy). Due to the typical inelasticity of hadronic interactions, the cutoff energy of Pevatrons in the gamma-ray regime is located at energies of several 100 TeV. Hard gamma-ray spectra in this energy range are much less ambiguously attributable to the hadronic rather than leptonic emission mechanism, because the cross-section for inverse Compton scattering drops with increasing energy (Klein-Nishina regime). Recent neutrino observations[7] and the detection of gamma-rays at ultra high energies by LHAASO[8] emphasize the potential of the search for the Galactic Pevatrons. The detection of Pevatrons, and their spectroscopic, and morphological resolution in their cutoff energy regime are the main physics goals of TAIGA.

Beyond 10^{17} eV, the origin of CR is assumed to be extra-Galactic (EG). Due to the development of intergalactic pair cascades, the accelerators of EG cosmic rays (EGCR) might exhibit a halo-like gamma-ray emission. Furthermore, one can expect an enhancement of EGCR from the local galactic supercluster.[9] The intergalactic pair cascades from that direction could create an anisotropic gamma-ray emission, observable with a wide field of view.

In the energy range of 100 TeV to several PeV, absorption due to pair creation becomes relevant also for Galactic sources. The attenuation of Galactic gamma-ray signals via production of e^+e^--pairs reaches a maximum around 100 TeV from the Galactic interstellar radiation field (IRF) and at 2 PeV from the cosmic microwave background (CMB).[10] This opens up the possibility to measure the density of the IRF, provided the distance of the emitting object is known. At PeV energies, the relevant field is the CMB, which is universal and very well known. Therefore, a possibility might exist to measure distances from a CMB absorption feature, i.e. cutoff in energy spectra of the most energetic Galactic gamma-ray sources. Another variable comes into play when considering the possibility of photon/axion conversion[11] or photon/hidden-photon oscillations,[12] because axions and hidden-photons could propagate without interaction, reconvert to photons, thus reducing the attenuation effect by pair production. Another fundamental effect on the attenuation by pair production might be the modification of the e^+e^- pair production threshold due to Lorentz Invariance violation. In a more direct way, other, non-thermal, dark matter candidates (e.g. wimpzillas) with masses ranging from 10^{12} GeV to 10^{16} GeV can be searched for in the energy range considered.

Furthermore, the chemical composition of Cosmic rays can be measured with TAIGA in the transition range from a Galactic to extragalactic origin ($\approx 10^{15}$ eV to 10^{17} eV) and in the energy range above, where the composition of cosmic rays is still poorly understood.[13] Furthermore, TAIGA will allow measurements of cosmic ray anisotropies beyond 100 TeV primary energy. Finally, observations of EAS can

also address fundamental questions of particle physics such as measurements of the proton-proton cross-section, or the search for quark-gluon plasma.

2. The TAIGA Experiment

In the TAIGA experiment, a hybrid approach using different detection techniques is used. While both Cherenkov photon and particle identification are used, the term hybrid also, and foremostly, refers to the hybrid Cherenkov technique. The angle-integrating air shower timing array TAIGA-HiSCORE is combined with the imaging air Cherenkov Telescopes TAIGA-IACT. The timing array is a cost effective method for the instrumentation of very large detector areas, and it provides good angular, core position, and energy resolution in the energy range above few 10s of TeV. The IACTs provide the air shower image shape for good gamma-hadron separation. When used in stand alone mode (i.e. no stereo), one IACT can detect air showers at a distance of 300 m - and more - from the shower core. Both detector components together provide the most relevant key parameters: direction and core from HiSCORE-TAIGA and image shape (width, length) from IACT-TAIGA. Placed at a distance of about 600 m from each other, 4 IACTs can cover a HiSCORE array of more than 1 km^2 area. It has been shown that using such a hybrid approach with only few IACTs per km^2, combined with a HiSCORE array, achieves a competitive gamma-hadron separation in the energy range above 10 TeV.[14] At energies higher than 100 TeV, the measurement of the muon component of the EAS using particle detectors on the surface and underground will enhance the gamma-hadron separation further.

Currently, TAIGA consists of a total of 120 HiSCORE stations deployed on an area of 1 km^2, and two 4 m class imaging air Cherenkov telescopes (IACTs) with a 10° aperture. A third IACT is planned for 2022. The distance between the IACTs was chosen as 300 m in the prototype phase of TAIGA. In future, distances of up to 600 m are planned. Additionally, scintillator stations for charged particle detection with a total surface of 240 m^2 are deployed above and below ground for the measurement of the muon component of EAS. The layout of the TAIGA-HiSCORE and TAIGA-IACT components is shown in Figure 1. The current TAIGA pilot array is used for a proof of principle of the detector concept. In future, an optimized array layout will take advantage of the hybrid method, that allows larger separations between the individual IACTs.

2.1. *TAIGA-HiSCORE*

HiSCORE is an array of angle-integrating air Cherenkov detector stations, distributed over an area of approximately 1 km^2. As shown in Figure 1, the stations are arranged in rows offset to each other with inter-station distances of 75 m to 150 m. Each station consists of 4 8 inch (and 10 inch) photomultiplier tubes (PMTs, ElectronTubes and Hamamatsu) with a segmented Winston cone (Alanod 4300UP foil), resulting in a light sensitive area of 0.5 m^2 per station. The viewing cone of

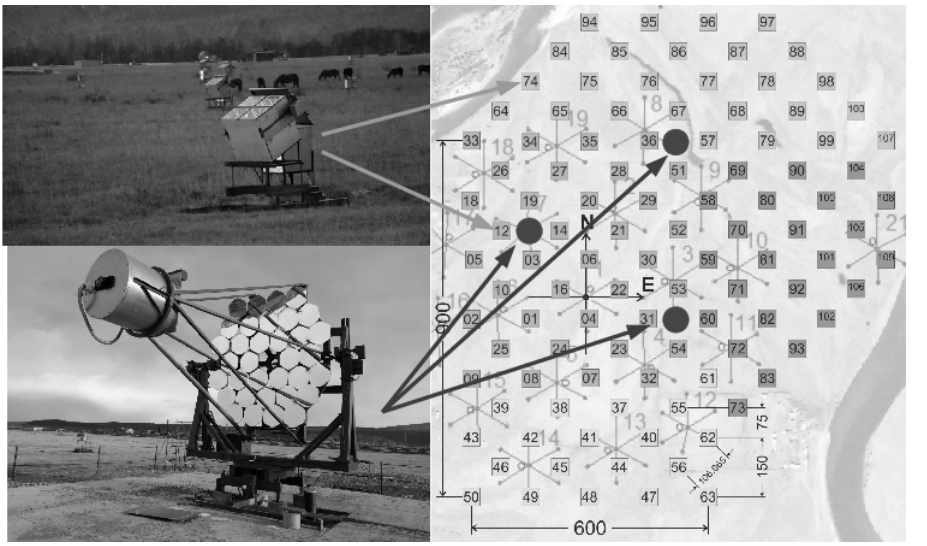


Fig. 1. Layout of the TAIGA-HiSCORE array. Only the HiSCORE and IACT components are shown. The first two IACTs were deployed from 2016 to 2019. The third IACT will be deployed until 2022. The HiSCORE array currently consists of 89 stations and will be extended to 120 stations by the end of 2021.

one station has a diameter of 60°, resulting in an effective field of view of 0.6 sr. An analog summator sums up the signals of all four channels. The sum is used for triggering, and the individual channels are read out with a DRS 4-based 9-channel readout board at a sampling rate of 2 GHz. Four anode signals and four dynode signals (low gain channel) are read out. Channel number 9 is used for time synchronization, based on the sampling of a 100 MHz clock signal. This custom time-synchronization was cross checked with the WhiteRabbit system. It could be verified that the time synchronization works on a level of 0.2 ns relative timing accuracy.[15,16] Each detector station is connected to the central DAQ via optical fibre. The signal traces are processed to extract the amplitude and the timing (half-rise-time, full width half maximum). Based on these parameters, the air shower arrival direction, core impact position, and energy are reconstructed using the methods developed for Tunka-133 and HiSCORE.[1,16,17] Monte Carlo simulations were used to evaluate the detector performance, resulting in resolutions for core position (O(15 m)), energy (20%), and direction (0.1°). Moreover, the serendipituous discovery of a pulsed signal from the CATS lidar onboard the international space station allowed to verify the absolute pointing of the experiment.

2.2. *TAIGA-IACT*

Currently, two IACTs are in operation on the TAIGA site. A third IACT is under construction. The TAIGA IACTs are based on a Davies-Cotton design, with

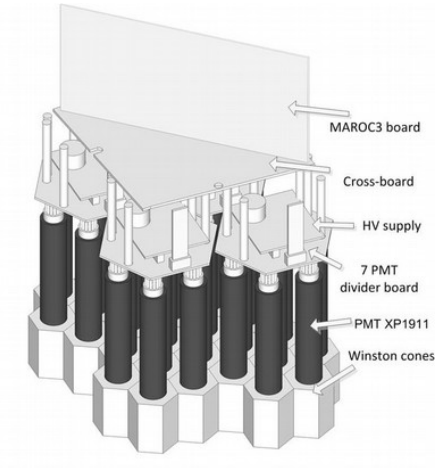


Fig. 2. Layout of a camera cluster of 28 PMTs. The high voltage is supplied to subgroups of 7 pixels. Each cluster is equipped with a MAROC-3 ASIC board.

a tesselated mirror dish of 34 facets with 60 cm diameter each, resulting in a mirror area of about 10 m^2. Each axis of the alt-az dish mount mechanics is driven by a stepper motor equipped with shaft-encoders for pointing positioning. Additionally, a sky-CCD system is used for pointing calibration. A verification of the absolute pointing accuracy of 0.02°, was carried out by the measurement of the anode current in the central camera PMT-pixel in a drift-scan accross a known bright star position.[18] The PMT camera of IACT-1 consists of 560 XP1911 PMTs (diameter 19mm). The second IACT is equipped with 595 PMTs. Both cameras have a field of view of 9.6° and are placed at the focal point at 4.75 m distance from the respective dishes. For enhanced light collection, and noise suppression from stray light, a Winston Cone plate is attached to the PMT camera plane. The PMTs are organized in clusters of 28 pixels (see Figure 2). All channels are read out with a 64 channel ASIC MAROC-3 board. The board provides a fast shaper which is used for triggering and a slow shaper which is used for read out. The camera electronics for the first two TAIGA-IACT cameras are described elsewhere.[19,20]

Air shower event images, such as shown in Figure 3, are reconstructed using the pixel amplitude distributions in the camera, either based on classical moment analysis,[21] or more advanced machine learning algorithms.[22]

2.3. *TAIGA Muon array*

The TAIGA-Muon array[23] consists of scintillation counters on a total area of currently 240 m^2. The muon detector stations are distributed above and below ground. In a future stage, a 5 times larger instrumented area is planned. Each TAIGA muon detector consists of four triangular sectors of scintillator material. The sectors are connected with wavelength shifting bars that guide the light to small PMTs (FEU-85). Measurements of cosmic rays have yielded an average cosmic ray muon amplitude of 31p.e.±20%. The TAIGA-Muon array develops its power

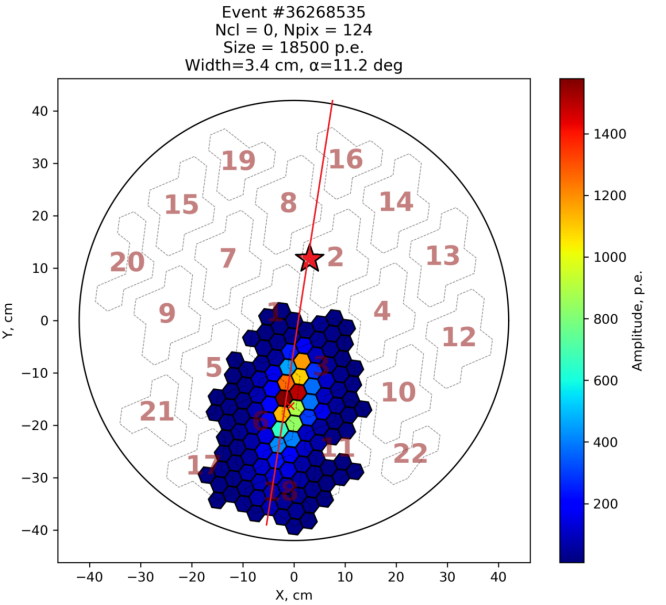


Fig. 3. Real data event for IACT1. The numbered shapes correspond to the above described clusters. The star shows the direction of the air shower event as reconstructed by the HiSCORE array. The total amplitude summed over pixels (size) is 18500 photoelectrons (p.e.). The alpha angle of the major axis is 11.2 degrees, and the measured image with (second image moment) is 0.41 degrees.

at higher energies, starting at about 100 TeV, where the measurement of the muon component of EAS helps in discriminating hadrons from gamma-rays, as well as identifying the nature of the hadronic cosmic rays.

2.4. *Hybrid technique*

A large instrumented detector area is key for accessing very high energies from few TeV to several 100s of TeV. Using only IACTs and the well-established stereoscopic technique requires a large number of channels per instrumented km^2, because firstly, each IACT has a large number of camera channels, and secondly, the IACTs cannot be placed too far away from each other, in order to detect the same EAS in at least two IACTs for stereoscopic reconstruction. As opposed to that, the instrumentation of large areas with a timing array such as HiSCORE is more cost effective, requiring considerably lower number of channels per km^2. However, while gamma-hadron separation is possible with HiSCORE alone to a certain degree (and improving with increasing energy[17]), the separation power stays below the quality reached with IACTs. Therefore, TAIGA implements a novel hybrid technique, taking advantage of the strengths of both Cherenkov techniques. The effective area is maximized using the timing array principle, and the gamma-hadron separation is optimized using

the IACTs. More specifically, the strong points of HiSCORE are the directional and core position resolution. The strength of the IACTs is the image shape. The IACTs can be placed so far apart from each other, that each IACT is operated in monoscopic mode, meaning that the Cherenkov lightpool of one air shower event only hits one IACT, as shown in Figure 4. Without requiring a stereoscopic trigger, the effective area of each IACT can be fully exploited and only 4 IACTs are needed in order to cover an area of the order of $1\,\mathrm{km}^2$. The loss of reconstruction quality when using the IACTs in monoscopic mode as compared to stereoscopic mode can be recuperated using the information of the HiSCORE array on angular and core position, as illustrated in Figure 5. The major axis of the IACT image in that Figure is pointing towards the core position, which cannot be reconstructed with the IACT alone, but is provided by the HiSCORE array. For comparison, a classical method for hadron suppression in stereoscopic systems is to use the mean scaled width parameter, for which the core position is reconstructed stereoscopically and the image widths of all IACTs are scaled according to MC-expectation. A similar parameter in TAIGA is the hybrid scaled width, which is the image width from one IACT, scaled with the MC-expectation for the given image size and using the core position and direction as reconstructed by HiSCORE. It was shown that the hybrid scaled width achieves competitive rejection power.[14] As opposed to stereoscopic systems, this method does not suffer from deteriorating core impact resolution at large distances (over the full HiSCORE array, the core resolution is the same) and also does not require closely spaced IACTs. The sensitivity of the air Cherenkov components of the TAIGA experiment was estimated using MC-simulations and is shown in Figure 6.

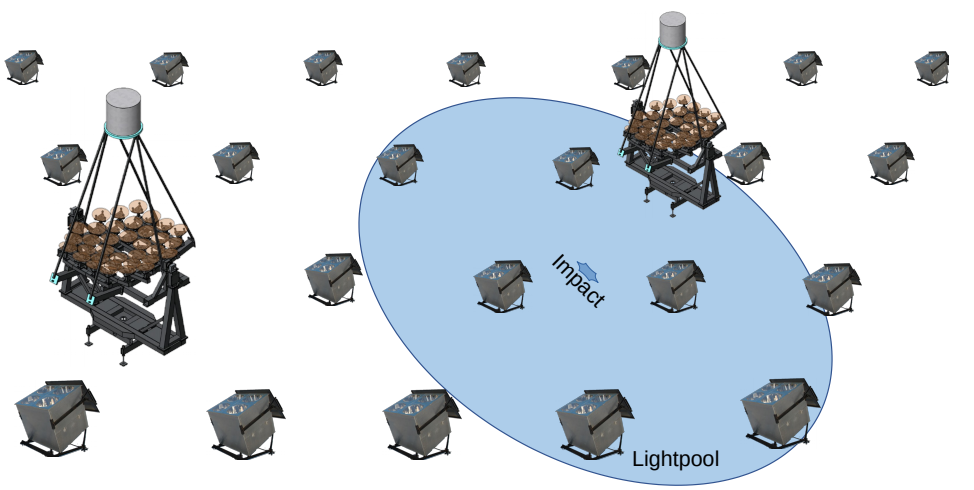


Fig. 4. Schematical drawing of the principle of hybrid operation.

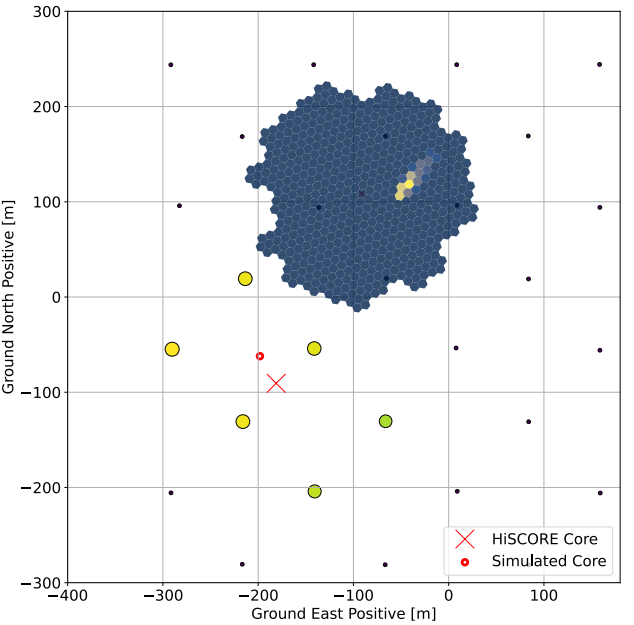


Fig. 5. Simulated hybrid event: as explained in Figure 4, the EAS impact position is reconstructed using the HiSCORE array. The major IACT image axis projected to the ground reference frame (not drawn) is pointing towards the position of the reconstructed core impact position.

3. First results

3.1. *TAIGA-HiSCORE*

The angle integrating HiSCORE array was installed in several stages since 2016. The performance of the array was estimated using from MC-simulations and real data. The relative angular and core position resolutions using a subdivision of the array into subarrays (chessboard method) could be reproduced in simulations, verifying the expected performance of TAIGA-HiSCORE.[24] The relative station timing is calibrated using stationary light sources and air shower data. A fast moving source in the field of view of the array turned out to be the CATS Lidar onboard the international space station. This light source could be used to verify time synchronization and absolute array pointing.[15,16] As compared to the Tunka-133 array, the measurement of the cosmic ray energy spectrum could be extended to lower energies with TAIGA-HiSCORE, as indicated in Figure 7.

3.2. *TAIGA-IACT*

The MC detector simulation for the TAIGA-IACTs is done using two simulation chains: a custom MC simulation chain, and an adaptation of the *sim_telarray* package.[25] The image pixel photoelectron scale is calibrated using an LED illuminating

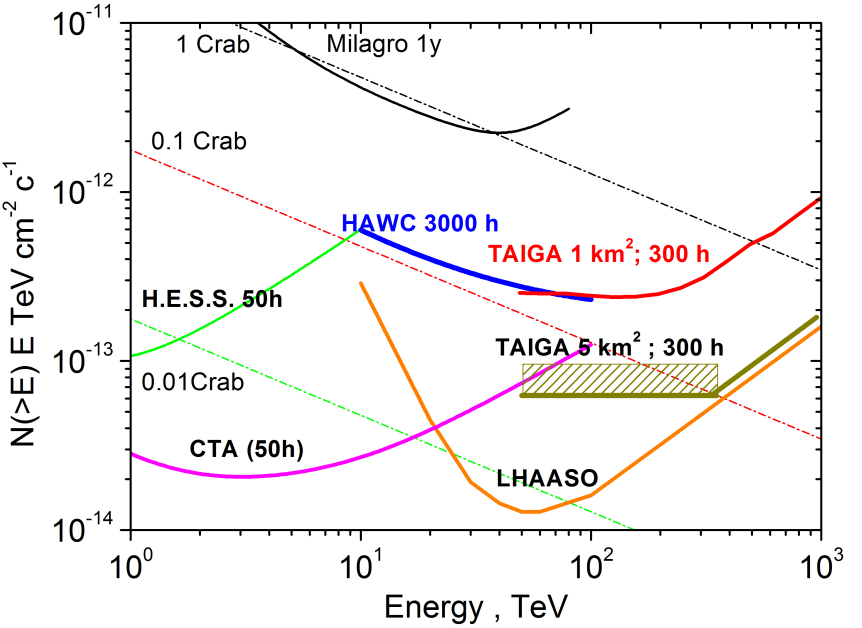


Fig. 6. Sensitivity of the TAIGA experiment for 300 hours of observations in hybrid mode (without muon detectors). A comparison to sensitivities of other experiments is not straight forward, since also the angular and spectroscopic resolutions play an important role for measurements of the morphology and spectra of astrophysical objects. TAIGA has a relative energy resolution of better than 20% and an angular resolution of the order of 0.1°.

the whole PMT camera. Raw data are processed using two independent reconstruction chains to obtain air shower images (amplitude distributions of all pixels for each event). Four different high-level analyses, based on the robust Hillas-parameters, were applied to these image data. Figure 8 shows a comparison of the amplitude distributions of the second-hottest pixel in real data and simulation. A comparison of simulated and real image width is shown in Figure 9. The simulations describe real data reasonably well.

After commissioning of the first TAIGA-IACT, our primary goal was to demonstrate that the IACT is operational. For this purpose, observations of the standard candle, the Crab Nebula, were carried out. After weather quality selection, a total data set of 40.5 h is obtained for analysis. Observations were taken in the wobble mode (Wobble offset 1.2deg). The position of the Crab Nebula inside the IACT camera is taken as the on-source region and the background was estimated from off-source regions at identical distances to the camera center, from the same dataset. The distribution of the alpha angle between the major image axis and the position of the Crab Nebula inside the camera is shown in Figure 10. Selecting events with an image size of at least 120 p.e., a clear excess can be observed. With a zenith angle of 31deg, and taking into accout the altitude of the observation level, the efficiency

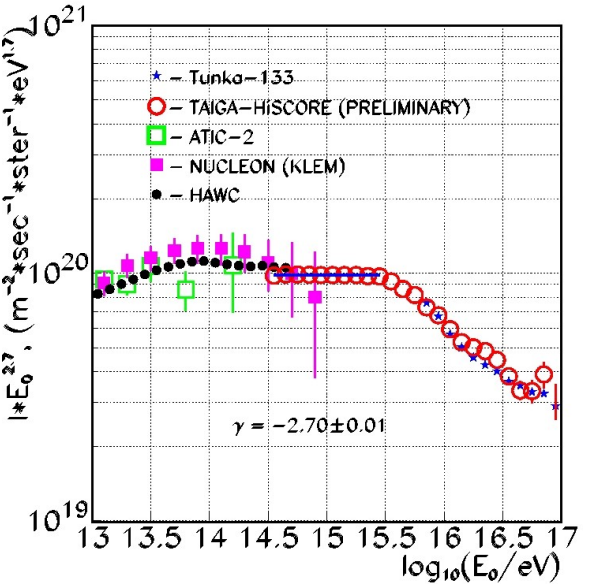


Fig. 7. The all-particle cosmic ray spectrum measured with TAIGA-HiSCORE, compared to different measurements by other experiments.

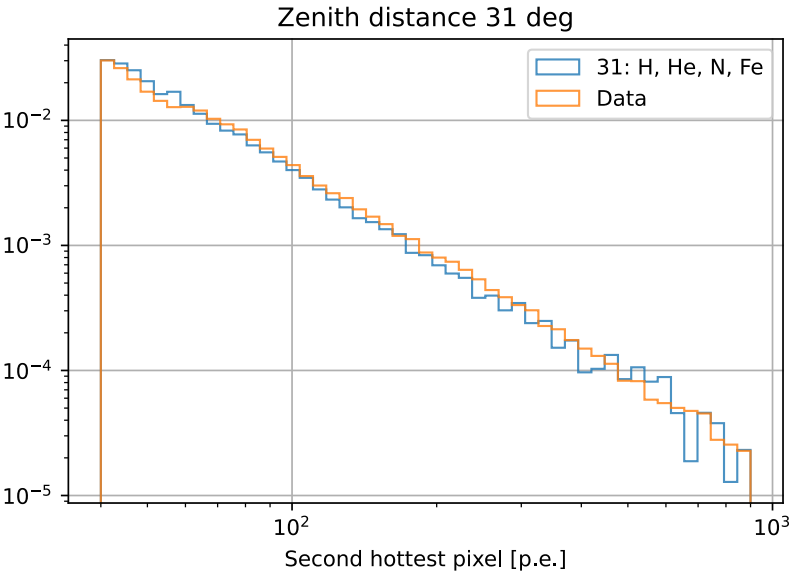


Fig. 8. Distribution of the amplitude of the second-hottest pixel in IACT-1. Data are shown in light grey, simulations are shown in dark grey.

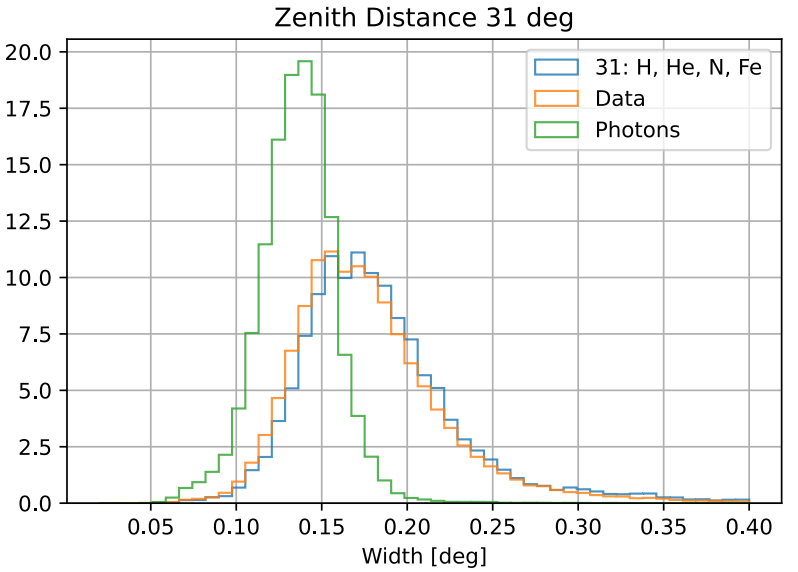


Fig. 9. Distribution of the image width for MC and real data.

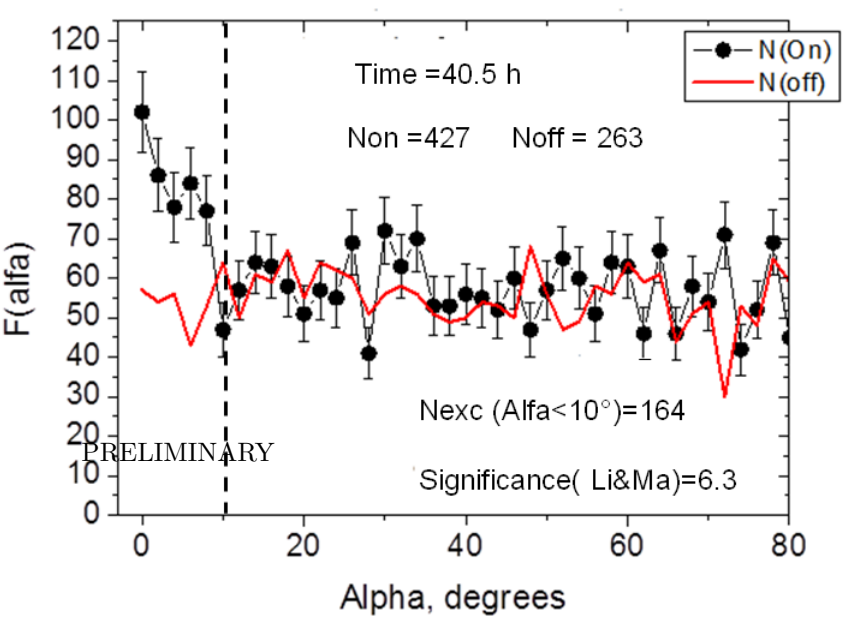


Fig. 10. Distribution of the alpha angle for all runs of a dataset of 42 hours from the Crab Nebula. These preliminary results are currently being finalized to be included in an upcoming journal publication.

of the optical system, and the cuts applied, the energy threshold of this data set is 4.5 TeV . The significance[26] of the observed excess is $6.3\,\sigma$ (using a cut on alpha smaller than 10deg).

4. Conclusion

The novel hybrid technique of the TAIGA pilot array combines the angle integrating air Cherenkov timing techinque with 4 m class imaging air Cherenkov telescopes, taking advantage of the strengths of both techniques, i.e. large instrumented areas with the HiSCORE array, and gamma-hadron separation from image shape using the IACTs. Additionally, a scintillator detector for particle detection above and below ground will allow the measurement of the muon component, for improved gamma-hadron separation at higher energies. The expected sensitivity of the future $5\,\mathrm{km}^2$ stage is in the region of $10^{-13}\mathrm{photons\,cm^{-2}s^{-1}}$. First stages of the 3 TAIGA detector components are in operation. Monte Carlo simulations were compared to real data, showing a good understanding of the detecots. Observations of the Crab Nebula for 40.5 h above 4.5 TeV resulted in an excess of $6.3\,\sigma$.

Acknowledgments

The work was performed at the UNU "Astrophysical Complex of MSU-ISU" (agreement 13.UNU.21.0007). The work is supported by Russian Foundation for Basic Research (grants 19-52-44002, 19-32-60003), the Russian Science Foundation (grant 19-72-20067 (Section D,E), the Russian Federation Ministry of Science and High Education (projects FZZE-2020-0017, FZZE-2020-0024), by the Deutsche Forschungsgemeinschaft (DFG, TL 51/6-1) and by the Helmholtz Association (HRJRG-303) and by European Union's Horizon 2020 programme (No. 653477). D.H., M.T, A.K.A. and M.B. acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC 2121 "Quantum Universe" — 390833306.

References

1. S. F. Berezhnev, D. Besson, N. M. Budnev, A. Chiavassa, O. A. Chvalaev, O. A. Gress, A. N. Dyachok, S. N. Epimakhov, A. Haungs, N. I. Karpov, N. N. Kalmykov, E. N. Konstantinov, A. V. Korobchenko, E. E. Korosteleva, V. A. Kozhin, L. A. Kuzmichev, T. Collab.) *et al.*, *Nucl. Instrum. Methods* **692**, p. 98 (2012).
2. S. D. Hunter, D. L. Bertsch, J. R. Catelli, T. M. Dame, S. W. Digel, B. L. Dingus, J. A. Esposito, C. E. Fichtel, R. C. Hartman, G. Kanbach, D. A. Kniffen, Y. C. Lin, H. A. Mayer-Hasselwander, P. F. Michelson, C. von Montigny, R. Mukherjee, P. L. Nolan, E. Schneid, P. Sreekumar, P. Thaddeus and D. J. Thompson, *The Astrophysical Journal* **481**, 205 (May 1997).
3. A. A. Abdo, B. Allen, D. Berley, E. Blaufuss, S. Casanova, C. Chen, D. G. Coyne, R. S. Delay, B. L. Dingus, R. W. Ellsworth and et al., *The Astrophysical Journal* **658**, p. L33–L36 (Feb 2007).

4. F. Aharonian, A. G. Akhperjanian, A. R. Bazer-Bachi, B. Behera, M. Beilicke, W. Benbow, D. Berge, K. Bernlöhr, C. Boisson, O. Bolz and et al., *Astronomy & Astrophysics* **481**, p. 401–410 (Jan 2008).
5. A. Abramowski, F. Aharonian, F. A. Benkhali *et al.*, *Nature* **531**, p. 476–479 (Mar 2016).
6. S. Gabici and F. A. Aharonian, *The Astrophysical Journal* **665**, p. L131–L134 (Aug 2007).
7. T. Glauch, in *Proceedings of the 3rd International Symposium on Cosmic Rays and Astrophysics (ISCRA-2021), 8–10 June 2021.*
8. Z. Cao, F. Aharonian, Q. An *et al.*, *Nature* **594**, p. 33–36 (2021).
9. T. Kneiske and D. Horns, in *38th COSPAR Scientific Assembly*,
10. I. V. Moskalenko, T. A. Porter and A. W. Strong, *Astr. Phys. J. L.* **640**, L155 (April 2006).
11. F. D. Steffen, *European Physical Journal C* **59**, 557 (January 2009).
12. H.-S. Zechlin, D. Horns and J. Redondo, in *American Institute of Physics Conference Series (2008)*, eds. F. A. Aharonian, W. Hofmann and F. Rieger.
13. L. Anchordoqui, M. T. Dova, A. Mariazzi, T. McCauley, T. Paul, S. Reucroft and J. Swain, *Ann. Phys. (N.Y.)* **314**, p. 145 (2004), hep-ph/0407020.
14. M. Kunnas, PhD thesis, University of Hamburg (2017), https://ediss.sub.uni-hamburg.de/handle/ediss/7582.
15. O. Gress, I. Astapov, N. Budnev, P. Bezyazeekov, A. Bogdanov, V. Boreyko, M. Brückner, A. Chiavassa, O. Chvalaev, A. Dyachok, T. Gress, S. Epimakhov, E. Fedoseev, A. Gafarov, N. Gorbunov, V. Grebenyuk, (TAIGA Collab.) *et al.*, *Nucl. Instrum. Methods A* **845**, p. 367 (2017).
16. A. Porelli, D. Bogorodskii, M. Brückner, N. Budnev, O. Chvalaev, A. Dyachok, S. Epimakhov, T. Eremin, O. Gress, T. Gress, D. Horns, A. Ivanova, S. Kiruhin, E. Konstantinov, E. Korosteleva, M. Kunnas *et al.*, *J. Phys.: Conf. Ser.* **632**, p. 012041 (2015).
17. D. Hampf, M. Tluczykont and D. Horns, *Nucl. Instrum. Methods A* **712**, p. 137 (2013).
18. D. Zhurov, in *Proceedings of the 36th International Cosmic Ray Conference (ICRC2019).*
19. N. Lubsandorzhiev, I. Astapov, P. Bezyazeekov, V. Boreyko, A. Borodin, M. Brueckner, N. Budnev, A. Chiavassa, A. Dyachok, O. Fedorov, A. Gafarov, A. Garmash, N. Gorbunov, V. Grebenyuk, O. Gress, T. Gress, (TAIGA collab.) *et al.*, in *Proceedings of the 35th International Cosmic Ray Conference (ICRC2017).*
20. N. Lubsandorzhiev, in *Proceedings of the 36th International Cosmic Ray Conference (ICRC2019).*
21. A. M. Hillas, in *Proceedings of the 19th International Cosmic Ray Conference (1985).*
22. R. Parsons and S. Ohm, *Eur. Phys. J. C* **80**, p. 363 (2020).
23. I. I. Yashin, I. I. Astapov, N. S. Barbashina, A. G. Bogdanov, V. Boreyko, N. M. Budnev, M. Büker, M. Bruckner, A. Chiavassa, O. B. Chvalaev, A. V. Gafarov, N. Gorbunov, V. Grebenyuk, O. A. Gress, A. Grinyuk, O. G. Grishin, (TAIGA collab.) *et al.*, *J. Phys.: Conf. Ser.* **675**, p. 032037 (2016).
24. M. Tluczykont, O. Gress, E. Korosteleva, L. Kuzmichev, A. Pakhorukov, A. Porelli, V. Prosin, L. Sveshnikova, R. Wischnewski, I. Astapov, P. Bezyazeekov, V. Boreyko, A. Borodin, M. Brueckner, N. Budnev, A. Chiavassa, (TAIGA Collab.) *et al.*, in *Proceedings of the 35th International Cosmic Ray Conference (ICRC2017).*
25. K. Bernlöhr, *Astropart. Phys.* **30**, p. 149 (2008).
26. T. Li and Y. Ma, *Astrophys. J.* **272**, p. 317 (1983).

Science perspectives of the Southern Wide-field Gamma-ray Observatory (SWGO)

K.L. Engel* for the SWGO Collaboration

Department of Physics, University of Maryland, College Park; 4150 Campus Dr, College Park, MD, USA 20742

**E-mail: klengel@umd.edu*

We outline the science case for SWGO, the Southern Wide-field Gamma-ray Observatory, an air-shower array planned for construction at a high-altitude site in the Southern Hemisphere. This next-generation, wide-field-of-view, gamma-ray survey instrument will be sensitive to gamma rays from $\sim$ 100 GeV to beyond 100 TeV energies and, due to its location and features, will be complementary to other gamma-ray observatories such as HAWC, LHAASO, and CTA. SWGO's scientific endeavors will cover a broad range of exciting topics, such as monitoring the transient sky at very high energies, unveiling Galactic and extragalactic particle accelerators, probing particle physics beyond the Standard Model, and the characterization of the cosmic ray flux. With its novel design, SWGO stands poised to provide the world's best sensitivity across many facets of our exciting field.

Keywords: Instrumentation: detectors; gamma rays: general

1. Introduction

With an abundance of "paradigm-shifting" discoveries in the field of high-energy particle astrophysics in the last decade, it is critical for the continued growth of these topics and movement into potential new discovery space that there be facilities in both hemispheres capable of continuously surveying the gamma-ray sky in the time and space domains. Ideally, such an instrument would have a very wide field of view, making it capable of providing transient alerts to the astrophysics community in addition to having an abundance of archival data for performing follow-ups from other notifications. It is also important that this new facility be ground-based, collecting information from extensive air showers, as necessitated by the declining flux emitted by cosmic sources at very-high energies. We believe that the Southern Wide-field Gamma-ray Observatory (SWGO), a water-Cherenkov facility to be located in the Southern Hemisphere, fills this pivotal role within the astrophysics and high-energy particle physics communities.

The complementarity between IACTs and a wide-field-of-view ground-based particle array such as that outlined herein is summarized by Table 1, from the Science Case White Paper for a wide-field-of-view gamma-ray observatory in the Southern Hemisphere.[1]

This work is intended to highlight endeavors presented in Ref. 1, as well as additional efforts since the time of that document's publication, and is the collective work of the whole of the SWGO Collaboration.

Table 1. Comparison of typical performance of current and planned IACT arrays and ground particle arrays for gamma-ray astronomy. See Ref. 1 for details.

	IACT Arrays	Ground-particle Arrays
Field-of-view	3°–10°	90°
Duty cycle	10%–30%	>95%
Energy range	30 GeV – >100 TeV	∼500 GeV – >100 TeV
Angular resolution	0.05°–0.02°	0.4°–0.1°
Energy resolution	∼7%	60%–20%
Background rejection	>95%	90%–99.8%

2. The Southern Wide-field Gamma-ray Observatory (SWGO) Concept

SWGO is a proposed future water Cherenkov experiment in the Southern Hemisphere for the ground-based detection of astrophysical gamma rays. The core unit will use the water-Cherenkov technique to observe intensive air showers from primary particles incident on the Earth's atmosphere and would be the most sensitive gamma-ray observatory in the Southern Hemisphere above 10 TeV, with expected sensitivity up to PeV energies. With an array of such detections, similar to the operational mode of the HAWC Observatory in the Northern Hemisphere,[2] the energy and direction of the incident gamma ray that produced the shower can be determined.

The exact design of SWGO is potentially dependent on site choice, with water access, construction costs, and infrastructure feasibility all playing deciding roles in addition to compatibility with science-driven main design goals. The aim of the design such a next-generation observatory, no matter the specifics, is to achieve roughly an order of magnitude higher sensitivity over the current-generation of instruments like HAWC.[1]

Being the most sensitive observatory in the Southern Hemisphere above ~ 20 TeV (as seen in Figure 1) would make SWGO an ideal complement not only to Northern Hemisphere observatories, but also to CTA South. It's complementarity with Northern Hemisphere observatories is well visualized by comparing SWGO's projected field of view with that of the HAWC Observatory, shown in Figure 2. By placing SWGO in the Southern Hemisphere, we would have a prime view to monitor sources and source regions previously inaccessible or mostly in accessible to HAWC, such as the Galactic Center and the Magellanic Clouds, as well as more ideal viewing of source regions such as the Fermi Bubbles.

3. The Core Science Case

The following list is just an overview of the science we intend to do with SWGO once built, with promising sensitivity studies currently being conducted on all fronts.

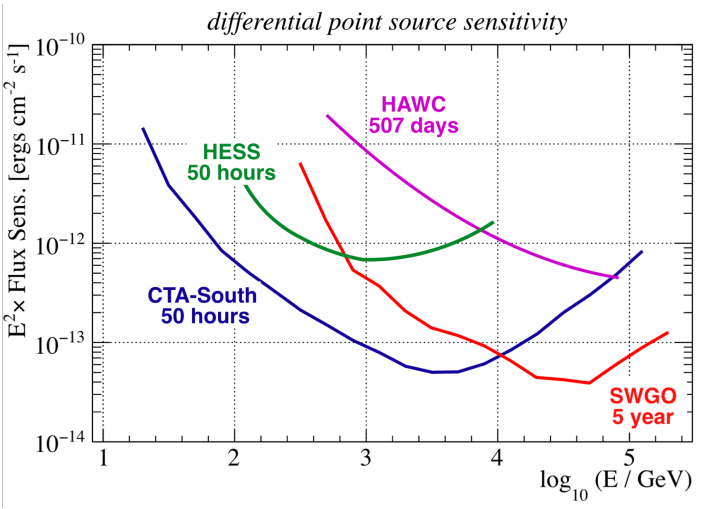


Fig. 1. The five-year differential point-source sensitivity of SWGO, anticipated to bbe the best in the Southern sky above tens of TeV. Figure from Ref. 3.

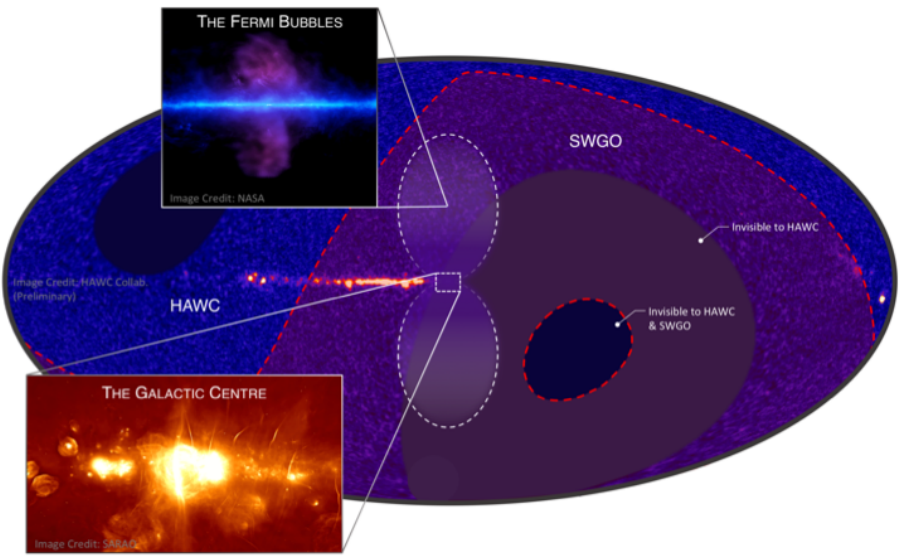


Fig. 2. Sky coverage of SWGO in Galactic coordinates (red dashed line) overlaid on the HAWC significance map containing over 50 sources. Figure from Ref. 3.

The core science cases of SWGO, as enumerated in Ref. 1, are:

- Detection of short-timescale phenomena[4]— low-energy threshold for detection of short-timescale (<1 hr) transient events down to 100 GeV

- Search for PeVatrons[5]— Improved sensitivity up to a few 100s of TeV to search for PeV Galactic particle accelerators
- Pulsar Wind Nebulae (PWNe) and Gamma-ray halos[6]— Unique potential for accessing the high-energy end of the Galactic population
- Dark matter and diffuse emission[7,8]— Unique access to the Galactic Center and Halo at the high-energy end of the spectrum
- Cosmic rays[9]— Unique complement to LHAASO for anisotropy studies, with the capability to reach low angular scale. Additionally, good muon tagging implies good mass resolution for composition studies up to the *knee*

3.1. *The High-Energy Transient Sky*

There are four key targets of interest in the high-energy transient sky—gravitational waves (GWs), fast radio bursts (FRBs), high-energy neutrinos, and gamma-ray bursts (GRBs)—to which the applied properties of SWGO will be a unique advantage for observing.

3.1.1. *Gravitational Waves*

GRB 170817A and the subsequent kilonova emission associated with binary-neutron-star (BNS) merger GW 170817 launched the field of GW multimessenger astronomy. SWGO's wide field of view and large, unbiased duty cycle would allow the recording of real-time, high-energy gamma-ray data for all GW events falling into its field of view. This would provide coverage of high-uncertainty regions, locate very-high-energy counterparts, and search for currently unmodeled burst-like GW signals. SWGO could even provide real-time notifications to other Southern Hemisphere telescopes and observatories, *e.g.,* CTA, so that they might perform detailed follow-up observations.

3.1.2. *Fast Radio Bursts*

Monitoring approaches by IACTs of FRBs are less than ideal as the duty cycle of FRBs is poorly constrained. A wide-field-of-view survey instrument like SWGO is therefore ideal for Galactic FRB follow-ups to provide constraints on gamma-ray emission in coincidence with bursts detected in the radio domain, as well as monitoring for repeaters. Simulation studies have also predicted that 60–70% of FRBs are located in the Southern Hemisphere, so SWGO's location will be incredibly advantageous for the study of these objects.

3.1.3. *High-Energy Neutrinos*

As evidence for the interaction of high-energy hadronic particles, high-energy neutrinos are crucial in the search for sources of high-energy cosmic-ray accelerators,

but a multimessenger approach is necessary to determine the source of the astrophysical flux of high-energy neutrinos. This multimessenger approach, towards which SWGO's large, unbiased duty cycle will be an invaluable tool, greatly aids neutrino studies as high-energy gamma-ray observations allow precise localization of the emission region, allowing for not only the identification of their astrophysical source, but also information about overall source energetics and energy- and time-dependent power output.

3.1.4. *Gamma-Ray Bursts*

Although electromagnetic emission from GRBs is widely observed down to the radio, the VHE domain remains largely uncharted. The transient and unpredictable nature of GRBs combined with intrinsic alert delays by X-ray satellites, makes it difficult for IACTs to observe GRBs rapidly enough to catch the prompt-to-early afterglow phases in which very-high-energy emission is either rarely, or as yet not observed. However, SWGO's wide field of view, large, unbiased duty cycle, and sensitivity $\lesssim 1$ TeV would allow for the monitoring of all nearby GRBs inside its field of view, increasing the number of observations without the need for external triggers, in addition to the benefits of its archival data analysis (see Figure 3). Additionally, follow-ups in SWGO's energy range would potentially allow for the discrimination between proposed emission scenarios— a longstanding topic of great interest within the field.

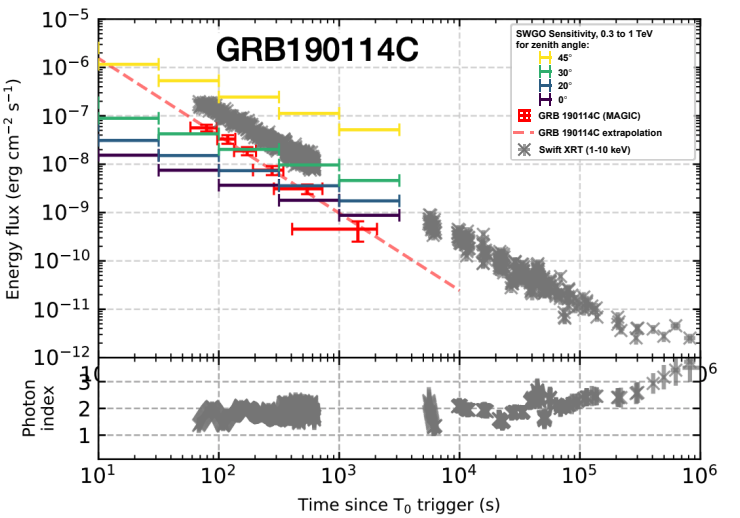


Fig. 3. SWGO's sensitivity to a GRB like GRB 190114C, including results from the MAGIC telescopes for comparison. Figure adapted from Ref. 10.

3.2. *Galactic Particle Acceleration*

The study of very-high-energy gamma rays is one of the most powerful methods to search the Galactic plane for astrophysical particle accelerators, including PeVatrons— sources able to accelerate hadrons up to the "knee" of the cosmic-ray spectrum. With its large energy range, excellent sensitivity above 10 TeV, and ideal design features, SWGO will be able to isolate Galactic PeVatrons from the background of softer gamma-ray sources. Its design means that SWGO is well equipped to handle the source confusion, diffuse gamma-ray backgrounds, and spatially extended emission regions that typically complicate these kinds of analyses, making its sensitivity to astrophysical particle accelerators in the local Galactic neighborhood one of the greatest strengths of the proposed observatory. From this, we are enthusiastic that SWGO will be the optimal instrument to discover sources such as the microquasar SS433.[11] Such assets will also allow SWGO to identify hard-spectrum sources for long-term observation with CTA, as well as complement the coverage provided of the Northern Hemisphere by HAWC and LHAASO.

3.3. *Pulsar Wind Nebulae and Gamma-Ray Halos*

First observed by HAWC,[12] the all-sky observations of which SWGO is capable will allow the testing of large gamma-ray halos as potential sources of the observed positron excess. Studying gamma rays from such potential positron sources provides crucial input to the positron excess puzzle. SWGO will survey nearby pulsars (age $< 10^6$ yr; see Figure 4) in the Southern Hemisphere, where its wide field of view is ideal for measuring gamma-ray halos around the co-located pulsar wind nebulae (PWNe), including from millisecond pulsars not (yet) discovered with current instruments. The large extents of these gamma-ray halos allow us to study the propagation of particles within them in unprecedented detail, unveiling the properties of the accelerators and the medium around them.

3.4. *The Extragalactic Very-High-Energy Sky*

Several models currently exist to explain the spectra of Active Galactic Nuclei (AGN) from radio to very-high-energy gamma rays, with additional data needed to unravel the mechanisms at work in these extreme environments. The very-high-energy extragalactic sky is dominated by Blazars, radio-loud AGN with jets oriented near our line of sight, whose second power-spectrum peak in the gamma rays has a much debated origin. SWGO is expected to be able to detect very-high-energy blazars at both short and long time scales, as well as perform searches in an unbiased way over the full Southern sky due to its near-continuous duty cycle. Such observations will provide information on the intergalactic magnetic field, radiative processes and acceleration mechanisms, periodicity, and beyond-the-Standard-Model physics. SWGO's characteristics are also ideal for the crucial evaluation of the significance of neutrino-blazar flare correlations, such as implied by observations of TXS 0506+056.[13]

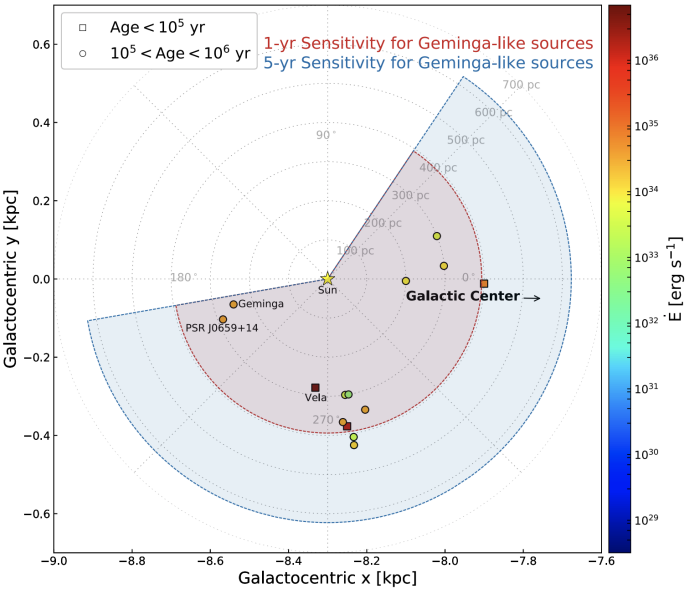


Fig. 4. Pulsars, age $< 10^6$ yr., within 500 pc, showing the one-year (red) and five-year (blue) sensitivity of SWGO to Geminga-like sources at these locations. Figure from Ref. 1.

3.5. *Physics Beyond the Standard Model*

Returning to the sky that we would expect to see with SWGO in Figure 5, one can see that SWGO would have access to several prominent dark matter targets that cannot be seen, or cannot be seen well, by HAWC in the Northern Hemisphere, such as the Galactic Center, the Magellanic Clouds, and a population of dwarf galaxies recently discovered with optical surveys like DES.

3.5.1. *WIMP Searches with Cosmic Gamma Rays*

Dark matter annihilation is theorized to produce Standard Model particles through weak interactions which, in turn, produce gamma rays, mainly via pion decay but also through inverse Compton scattering of photons off produced electron-positron pairs. Weakly Interacting Massive Particles, or WIMPs, are promising dark matter candidates with $E_{CM} \sim 100$ GeV. Their masses can range from $\sim$5–100 TeV, but the heaviest particles, $m_\chi > 10$ TeV, that can only be searched for with an instrument with a high sensitivity, are best probed with an indirect experiment like SWGO. Heavy dark matter (100 TeV–100 PeV) is also theorized beyond the WIMP model, where SWGO's wide field of view and sensitivity to the highest energies will make it a crucial player in searching for indirect signals from dark matter.

The Galactic Center is the closest dense region of dark matter and, as seen in Figure 5, SWGO's location in the Southern Hemisphere will allow this to be a prime

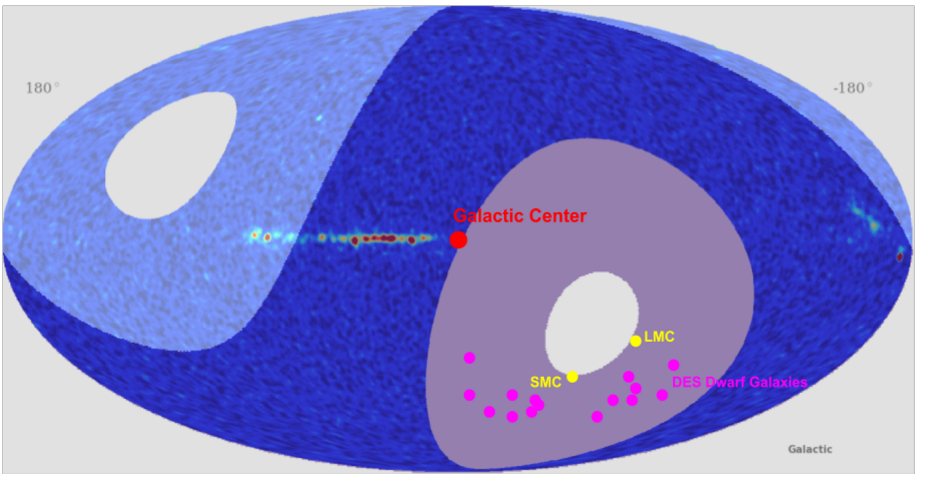


Fig. 5. HAWC skymap with SWGO view (shaded) overlaid. Figure from Andrea Albert.

target for observations. This is also an opportunity to highlight the complementarity of SWGO with the *Fermi*-LAT and CTA. Combined, these experiments will explore thermal WIMPs from 5 GeV–100 TeV.[14] There is an overlap in mass between experiments, which will allow for multiple potential detections, but it is SWGO that extends the reach of such studies up to 100 TeV— one of many exciting reasons to build SWGO.

3.5.2. *Axion-Like Particles*

As SWGO will have its best sensitivity at the highest energies, we can also look for Axion-like particles (ALPs), a generalization of the axion and a well-motivated dark matter candidate. Gamma rays from AGN can convert to ALPs in magnetic fields in the environment through which they propagate. These particles can travel unattenuated through the extragalactic background light (EBL) and convert back into gamma rays in the Milky Way's magnetic field. This conversion results in a high-energy tail in the observed energy spectrum—as seen in Figure 6—to which SWGO would be sensitive.

3.5.3. *Dark Matter Searches in Dwarf Galaxies*

We can also look for dark matter annihilation gamma rays from dwarf galaxies, many of which have recently been discovered by optical surveys in the Southern Hemisphere, making them prime targets for SWGO. These dwarf spheroidals (dSphs) are dark-matter-rich Milky Way satellites that are nearby and essentially background-free. With its wide field of view and large duty cycle, SWGO will observe dozens of dSphs every day, giving us the ideal survey of the sources needed

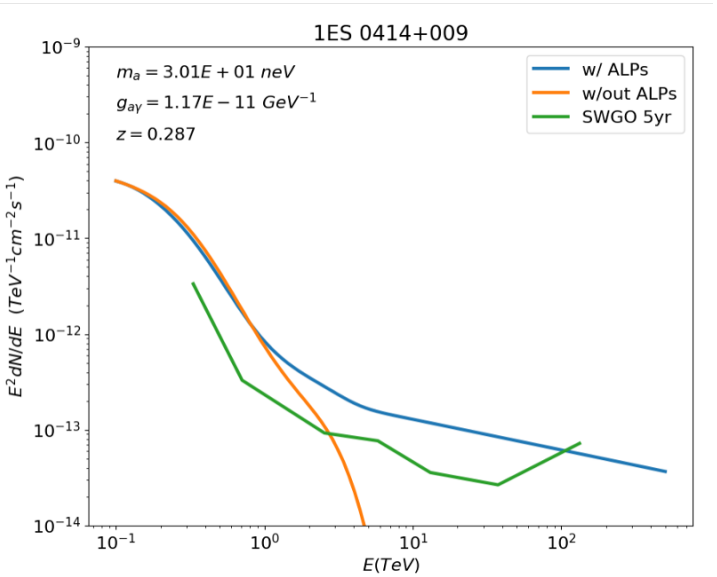


Fig. 6. ALP conversions could produce sharp spectral features in the observed spectrum of extragalactic sources which would require good energy resolution and large event statistics to resolve well. Figure from Ref. 1.

to place strong constraints on dark matter annihilation and decay. Unlike similar experiments, SWGO will have archival data at all locations within its field of view, which means if a new dwarf is discovered, there is already data on it to be analyzed with no time delay to collect it, such as might be experienced with IACTs.

3.5.4. *Lorentz Invariance Violation*

If Lorentz Invariance is violated, then gamma rays above an energy threshold ($E^{(n)}_{LIV}$) rapidly decay into e^+e^- pairs, so seeing a photon at the highest energies will set a limit on the energy scale of Lorentz Invariance Violation (LIV)— The current highest gamma-ray energy observed is 1,140 TeV seen by LHAASO.[15] SWGO's sensitivity up to several PeV would allow gamma rays to be probed for this signature of LIV at the highest energies ever observed. The detection of multiple PeV gamma rays will provide the world-leading constraints on LIV.

3.5.5. *Primordial Black Holes*

Primordial Black Holes (PBHs) are black holes created in the early Universe with masses ranging from the Planck mass to supermassive black holes. PBHs in some mass ranges are candidates to constitute some of the observed dark matter in the Galaxy. PBHs evaporate through Hawking Radiation, with this evaporation increasing substantially towards the end of their lives, producing a burst of very-high-energy gamma rays. As these are transient sources with a runaway emission,

SWGO's wide field of view and large duty cycle make it an optimal instrument with which to observe PBHs of initial mass $m_{PBH} \sim 5 \times 10^{14}$ g. The expected limits are 30 times better than the data-based limits currently placed by HAWC,[16] as shown in Figure 7.

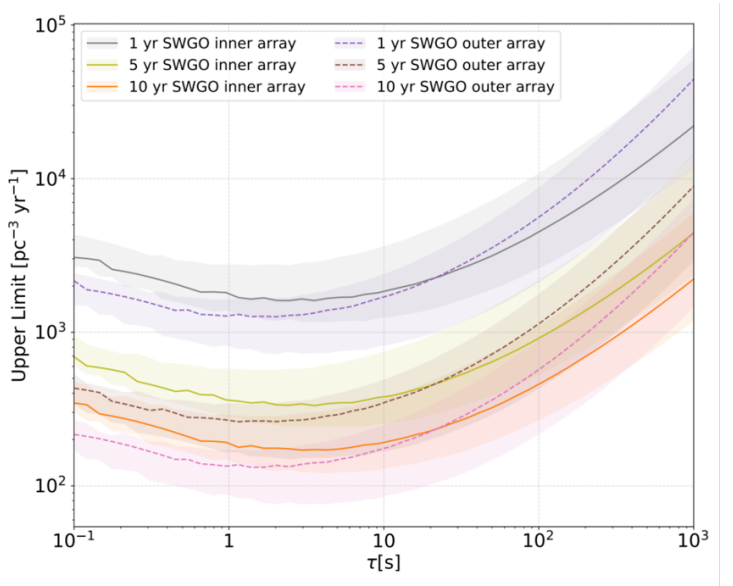


Fig. 7. The expected limits on the local burst rate density of PBHs. Figure from Ref. 17.

3.6. *Cosmic Rays*

The main feature in the energy spectrum of Galactic cosmic rays is the so-called "knee", which is characterized by a steepening of the spectral index from ~ -2.7 to ~ -3.1 at about 3 PeV. Understanding the origin of the "knee" is the key for a comprehensive theory of the origin of cosmic rays up to the highest observed energies. SWGO is uniquely suited to detect sources of cosmic rays with energies in excess of 1 PeV—expected to produce very-high-energy gamma rays above 100 TeV[18–21]—and measure any cutoffs in their spectra that indicate the maximum energy to which they are able to accelerate particles. With its wide aperture and its almost 100% duty cycle, SWGO will be able to gather significantly larger electron event statistics than IACTs, which may help in extending the electron spectrum further than the 20 TeV already measured, potentially revealing more details of the local source distribution. Additionally, the large sky coverage of SWGO would enable the search for anisotropy in the electron arrival direction at different angular scales, which would provide additional very relevant information for the understanding of Galactic particle acceleration and cosmic-ray transport.

4. Looking Forward

As a ground-based cosmic-gamma-ray observatory sensitive up to PeV energies, SWGO would be the most sensitive gamma-ray observatory in the Southern Hemisphere above 10 TeV, with complementarities among both Northern Hemisphere observatories, as well as other Southern Hemisphere observatories. With its wide field of view, near 100% duty cycle, ad high-energy reach, SWGO will provide the world's best sensitivity to emission from high-energy transients, propagation of particles within gamma-ray halos, PeVatrons in the galactic plane, several dark matter candidates, and Lorentz Invariance Violation.

References

1. A. Albert *et al.*, Science Case for a Wide Field-of-View Very-High-Energy Gamma-Ray Observatory in the Southern Hemisphere (2 2019).
2. A. U. Abeysekara *et al.*, Measurement of the Crab Nebula at the Highest Energies with HAWC, *Astrophys. J.* **881**, p. 134 (2019).
3. P. Abreu *et al.*, The Southern Wide-Field Gamma-Ray Observatory (SWGO): A Next-Generation Ground-Based Survey Instrument for VHE Gamma-Ray Astronomy (7 2019).
4. F. Schüssler and K. Satalecka, All-Sky time domain astrophysics with Very High Energy Gamma rays, *BAAS* **51**, p. 357 (May 2019).
5. P. Cristofari, A. Albert, A. Carraminana, S. Casanova, B. Dingus, M. Duvernois, N. Fraija, H. Fleischhack, J. Goodman, T. Greenshaw, J. Harding, A. Haungs, B. Hona, P. Huentemeyer and V. Joshi, Where are the pevatrons? (05 2019).
6. H. Fleischhack, A. Albert, C. Alvarez, R. Arceo, H. A. A. Solares, J. F. Beacom, R. Bird, C. A. Brisbois, K. S. Caballero-Mora, A. Carraminana, S. Casanova, P. Cristofari, P. Coppi, B. L. Dingus, M. A. DuVernois, K. L. Engel, J. A. Goodman, T. Greenshaw, J. P. Harding, B. Hona, P. H. Huentemeyer, H. Li, T. Linden, K. Malone, J. Martinez-Castro, M. A. Mostafa, M. U. Nisa, C. Riviere, F. S. Greus, A. Sandoval, A. J. Smith, W. Springer, T. Sudoh, K. Tollefson, A. Zepeda and H. Zhou, Pulsars in a Bubble? Following Electron Diffusion in the Galaxy with TeV Gamma Rays, *BAAS* **51**, p. 311 (May 2019).
7. A. Viana, Searching for TeV Dark Matter in the Milky Way Galactic Center, *BAAS* **51**, p. 308 (May 2019).
8. A. Albert, K. S. Caballero, P. M. Chadwick, B. L. Dingus, K. L. Engel, J. A. Goodman, T. Greenshaw, J. P. Harding, P. Huentemeyer, J. S. Lapington, J. Lundeen, J. Martinez-Castro, M. U. Nisa, H. Schoorlemmer, K. Tollefson, A. Viana and A. Zepeda, Searching for Sources of TeV Particle Dark Matter in the Southern Hemisphere, *BAAS* **51**, p. 202 (May 2019).
9. Fraija, A. Albert, J. C. Arteaga-Velazquez, H. A. Ayala Solares, K. S. Caballero-Mora, P. Cristofari, G. Di Sciascio, J. C. Diaz-Velez, M. A. DuVernois, K. L. Engel, A. Galvan-Gamez, J. A. Garcia-Gonzalez, T. Greenshaw, J. P. Harding, A. Haungs, O. Martinez-Bravo, J. Martinez, M. A. Mostafa, I. Torres, A. Zepeda and H. Zhou, Cosmic rays in the TeV to PeV primary energy range, *Astro2020: Decadal Survey on Astronomy and Astrophysics* **2020**, p. 459 (May 2019).
10. V. A. Acciari *et al.*, Teraelectronvolt emission from the γ-ray burst GRB 190114C, *Nature* **575**, 455 (2019).

11. A. U. Abeysekara *et al.*, Very high energy particle acceleration powered by the jets of the microquasar SS 433, *Nature* **562**, 82 (2018), [Erratum: Nature 564, E38 (2018)].
12. A. U. Abeysekara *et al.*, Extended gamma-ray sources around pulsars constrain the origin of the positron flux at Earth, *Science* **358**, 911 (2017).
13. D. Van Eijk, Tau Neutrinos in IceCube, KM3NeT and the Pierre Auger Observatory, *SciPost Phys. Proc.* **1**, p. 030 (2019).
14. A. Viana, H. Schoorlemmer, A. Albert, V. de Souza, J. P. Harding and J. Hinton, Searching for Dark Matter in the Galactic Halo with a Wide Field of View TeV Gamma-ray Observatory in the Southern Hemisphere, *JCAP* **12**, p. 061 (2019).
15. Z. Cao *et al.*, Exploring Lorentz Invariance Violation from Ultra-high-energy Gamma Rays Observed by LHAASO (6 2021).
16. A. Albert *et al.*, Constraining the Local Burst Rate Density of Primordial Black Holes with HAWC, *JCAP* **04**, p. 026 (2020).
17. R. López-Coto, M. Doro, A. de Angelis, M. Mariotti and J. P. Harding, Prospects for the Observation of Primordial Black Hole evaporation with the Southern Wide Field of View Gamma-ray Observatory (3 2021).
18. N. Fraija, E. Aguilar-Ruiz, A. Galván-Gámez, A. Marinelli and J. A. de Diego, Study of the PeV neutrino, γ-rays, and UHECRs around the lobes of Centaurus A, *MNRAS* **481**, 4461 (December 2018).
19. P. Cristofari, A. Albert, A. Carramiñana, S. Casanova, B. L. Dingus, M. A. DuVernois, N. Fraija, H. Fleischhack, J. A. Goodman and T. Greenshaw, Where are the pevatrons?, in *BAAS*, May 2019.
20. N. Fraija, M. Araya, A. Galván-Gámez and J. A. de Diego, Analysis of Fermi-LAT observations, UHECRs and neutrinos from the radio galaxy Centaurus B, *JCAP* **2019**, p. 023 (August 2019).
21. J. Holder, E. Amato, R. Bandiera, R. Bird, A. Bulgarelli, V. V. Dwarkadas, N. Giglietto, P. Goldoni, J. Hinton and B. Hnatyk, Understanding the Origin and Impact of Relativistic Cosmic Particles with Very-High-Energy Gamma-rays, in *BAAS*, May 2019.

Laue lenses: Focusing optics for hard X/soft Gamma-ray astronomy

L. Ferro*, M. Moita, P. Rosati, R. Lolli, C. Guidorzi and F. Frontera

Department of Physics and Earth Science, University of Ferrara, Via Giuseppe Saragat 1, Ferrara (FE), 44122, Italy
** E-mail: frrlsi@unife.it*
www.fst.unife.it

E. Virgilli, E. Caroli, N. Auricchio, J. B. Stephen, C. Labanti, F. Fuschino and R. Campana

OAS of Bologna, INAF, Via Piero Gobetti 93/3, Bologna (BO), 40129, Italy
www.oas.inaf.it

C. Ferrari

Institute of Materials for Electronics and Magnetism (IMEM), CNR, Parco Area delle Scienze 37/A, Parma (PR), 43124, Italy
www.imem.cnr.it

S. Squerzanti

INFN of Ferrara, Via Giuseppe Saragat 1, Ferrara (FE), 44122, Italy
www.fe.infn.it

M. Pucci

National Institute of Optics (Istituto Nazionale di Ottica, CNR-INO), CNR, Largo Enrico Fermi 6, Firenze (FI), 50125, Italy
www.ino.cnr.it

S. del Sordo and C. Gargano

IASF of Palermo, INAF, Via Ugo La Malfa 153, Palermo (PA), 90146, Italy
www.ifc.inaf.it

Hard X-/soft Gamma-ray astronomy is a key field for the study of important astrophysical phenomena such as the electromagnetic counterparts of gravitational waves, gamma-ray bursts, black holes physics and many more. However, the spatial localization, imaging capabilities and sensitivity of the measurements are strongly limited for the energy range >70 keV due to the lack of focusing instruments operating in this energy band. A new generation of instruments suitable to focus hard X-/ soft Gamma-rays is necessary to shed light on the nature of astrophysical phenomena which are still unclear due to the limitations of current direct-viewing telescopes. Laue lenses can be the answer to those needs. A Laue lens is an optical device consisting of a large number of properly oriented crystals which are capable, through Laue diffraction, of concentrating the radiation into the common Laue lens focus. In contrast with the grazing incidence telescopes commonly used for softer X-rays, the transmission configuration of the Laue lenses

allows us to obtain a significant sensitive area even at energies of hundreds of keV. At the University of Ferrara we are actively working on the modelization and construction of a broad-band Laue lens. In this work we will present the main concepts behind Laue lenses and the latest technological developments of the TRILL (Technological Readiness Increase for Laue Lenses) project, devoted to the advancement of the technological readiness of Laue lenses by developing the first prototype of a lens sector made of cylindrical bent crystals of Germanium.

Keywords: Hard X/soft Gamma-ray astronomy; Laue Lenses; Focusing Telescopes.

1. Laue lenses: A new way to look at the X-ray/Gamma sky

Laue lenses are innovative X and Gamma-ray optics based on the Bragg's law of diffraction in crystals[1]:

$$2d_{hkl} \sin \theta_B = n \frac{hc}{E} \tag{1}$$

Where θ_B is the Bragg's angle, which is the angle between the diffraction planes of the crystal and the diffracted beam, d_{hkl} is the inter-planar spacing of diffraction planes of the crystal with Miller indexes (hkl), E is the energy of the diffracted photon, n is the diffraction order and hc corresponds to 12.4 keV · Å.

The crystals can be arranged in such a way that the X-ray beam can be reflected by a thin, superficial layer of the crystal (Bragg geometry, or Reflection geometry) or it can cross the whole crystal and be transmitted over it (Laue geometry, or Transmission geometry).

Optics based on Bragg's law of diffraction in Laue configuration can be of great interest in the astrophysical context because (1) Bragg's diffraction is effective up to energies of few MeV and (2) the transmission configuration can be exploited to increase the effective area of the optics.

A Laue lens, indeed, is exactly based on this concept. A Laue lens can be visualized as a spherical cap of radius R covered by crystal tiles oriented in such a way that the the radiation coming from the sky is transmitted through the crystals and sent to the focal point. The focal distance of the lens f is equal to half the curvature radius of the spherical cap[2] (Fig. 1).

There are different configurations for the crystals on the cap which can be chosen to optimize the effective area in different ways; the simplest one, which is the configuration we are using for our prototypes, is to place the crystals in concentric rings at a distance r from the axis of the lens. From the Bragg's equation, we can see that the centroid of the energy spectrum of the photons diffracted by each ring can be expressed as[2]:

$$E = \frac{hc}{2d_{hkl}} \sin \left[\frac{1}{2} \arctan \left(\frac{f}{r} \right) \right] \sim \frac{hc}{d_{hkl}} \frac{f}{r} \tag{2}$$

Where the approximation holds for small Bragg's angle, which is the case of hard X and Gamma rays. Higher energy photons are diffracted from crystals in the inner region of the lens, while lower energy photons interact with the outermost regions

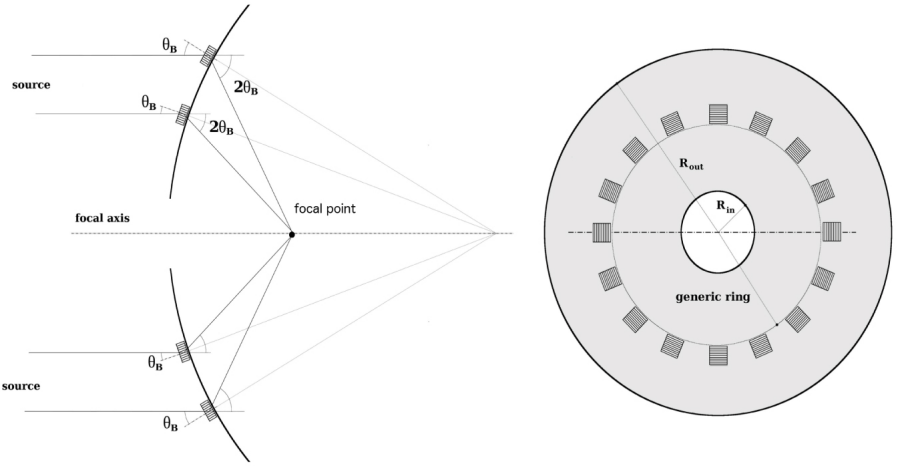


Fig. 1. Side and top view of a Laue lens. The crystals are positioned on a spherical support in such a way that the radiation coming from the sky, parallel to the optical axis of the lens, interacts with the crystals and is focused. The diffraction planes of the crystals are oriented in such a way that the angle between them and the incoming X-ray beam is equal to the Bragg's angle θ_B, so the angle between the diffracted beam and the incoming beam is $2\theta_B$. The focal lens is equal to half the curvature radius of the spherical cap.

of the optics (Fig. 2). With Laue lenses, we expect to be able to build a lightweight (<150 kg), stable optics which will enable real concentration of high energy radiation up to 600 keV at a focal distance of 20 m. Thanks to the unprecedented imaging capabilities of Laue lenses, we will be able to obtain a Point Spread Fuction (PSF) with a width of ∼30 arcsec in the energy band 50 keV - 600 keV. Comparing this with one of today's most advanced high energy instruments, the IBIS aboard INTEGRAL, whose PSF is of about 12 arcmin we can immediately understand the quantum leap in imaging capabilities that would be brought by an instrument such as a Laue lens, which will also allow us to make spectroscopy and even X-ray Compton polarimetry with an unprecented quality.

2. Bent perfect crystals

Several types of crystals can be used to build a Laue lens and the imaging capabilities and effective area of the optics are strictly related to the physical and chemical nature of the chosen crystals. Two are the families of crystals that we considered for our applications: perfect crystals and mosaic crystals.

A perfect crystal is a crystal in which the lattice plane are all oriented in the same direction. The rocking curve of the crystal has a Gaussian shape and it centered on the position satisfying Bragg's equation. The Full Width at Half Maximum (FWHM) of the rocking curve is called Darwin width.[1] A mosaic crystal, instead, can be seen as made up by microscopic perfect crystals, called *crystallites*, whose

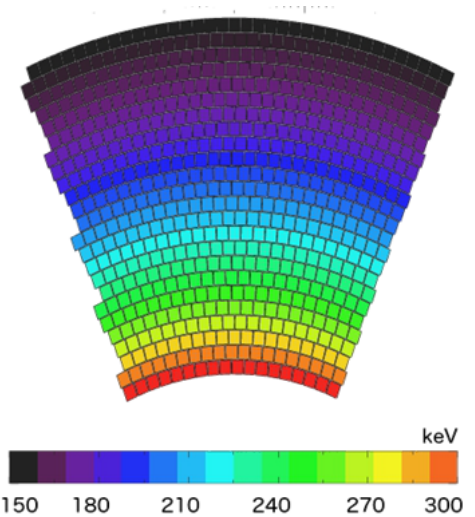


Fig. 2. In a spherical Laue lens, all the crystals at the same radial distance from the center of the lens are sensitive to the same energy range, so by nesting multiple rings it is possible to build a broad-band Laue lens. The highest energies are focused by the innermost crystals, while the lower energies are focused by the outermost crystals.

diffraction planes are misaligned around an average direction following a Gaussian distribution. The FWHM of the distribution of the cristallites is called *mosaicity* of the crystal.[1] The presence of a distribution of cristallites increases the passband of the crystals, however it also generates an effect called *mosaic defocusing*, which enlarges the PSF.

With the past HAXTEL project,[3] flat mosaic crystals of copper crystals were successfully used to build a prototype of a short-focal Laue lens. However they show some evident limits:

- Flat crystals don't have true concentration capability: they can redirect X-ray beams, however the size of the image produced by a flat crystal is comparable to the cross section of the crystal itself. This can be a problem especially in the case of long-focal lenses, for which very small crystals would be required to obtain an adequately small image.
- Flat crystals, expecially of the perfect type, show a very narrow passband.[1] Even though they can work perfectly for a narrow band Laue lens, the reduced pass-band is a problem for a broadband Laue lens, since it can be responsible of a very uneven variation of the effective area of the instrument.
- The diffraction efficiency for flat crystals cannot be higher than 50%.[1] This comes from the fact that the photon traversing the crystal have the same probability of undergoing an even or odd number of diffraction processes, limiting the maximum efficiency theoretically achievable.

The use of bent crystals allow us to overcome those limitations. A bent crystal is a crystal in which the average direction of the diffraction planes varies inside the volume of the crystals according to a definite curvature radius. If a crystal is cilindrically bent, which is the easiest type of curvature configuration attainable, then it acts as a concentrator along the bent direction, focusing the radiation coming from an astrophysical source to a focal distance equal to half the curvature radius of the crystal itself.[4] This allows us both to shrink the cross section of the image in a region smaller than size of the crystal tile, and to increase the energy passband of each crystal (Fig. 3).

This holds true both for perfect and mosaic crystals, however bent perfect crystals have some interesting properties which can make them particularly interesting for a Laue lens.

Bending a perfect crystal to a chosen curvature radius induce an internal secondary curvature on specific diffraction planes in a way related to the primary curvature radius of the crystal.[5] In this way, the average direction of the diffraction planes varies continuously inside the crystal, slightly increasing their angular spread, whose average value is called *quasi-mosaicity* of the crystal.[6]

According to the dynamical theory of diffraction in bent perfect crystals, the bending of a perfect crystal can also increase the reflectivity of the bent planes above the limit of 50%.[5] This can be explained intuitively in the following way: since the diffraction planes are curved, the probability that the beam crossing the crystal

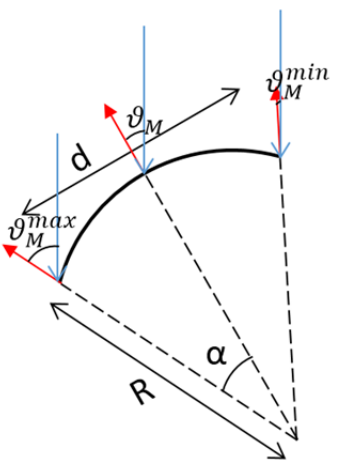


Fig. 3. Explanation of the concentration effect given by cylindrical bent crystal: due to the fact that the crystal is bent, the average direction of the diffraction planes (red arrows) varies continuously inside the crystals, which means that the Bragg's angle of the incoming radiation (blue arrows) changes accordingly. The result is that the X-ray beam is concentrated and that the crystal's energy pass-band is enlarged of a quantity which depends on the length of the focusing direction d of the crystal, its curvature radius R and its average diffracted energy.

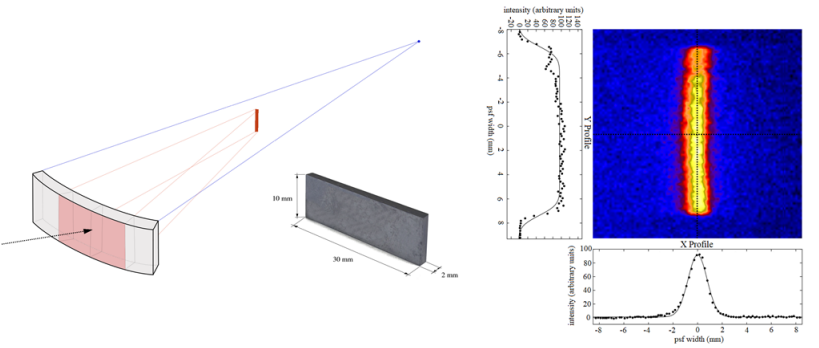


Fig. 4. Right: Schematics of the concentration effect of one cylindrical bent, 30×10 mm^2 crystals such as the one used in our tests. The radiation (in red) is concentrated along the focusing direction and the image is produced at a distance equal to half the curvature radius of the crystal. Left: Experimental image of a concentrated 150 keV X-ray beam obtained at the LARIX Facility. The area of the crystal is of 30×10 cm^2, while the area of the image is of 0.1×10 cm^2, which means that the beam was concentrated in an area 30 times smaller than the size of the crystal itself.

undergoes further diffractions gets reduced in a significant way. Experimental results suggest that an increase of the reflection efficiency can be effectively achieved.[7]

A crystal can be curved in an elastic way, by applying mechanical clamps to its surface, or in a plastic way, by subjecting it to a permanent deformation.[8] A mechanical bending of the crystal can bring some excessive complications for an astrophysical Laue lens, since adding further mechanical parts would increase the weight of the optics, the complexity of the design and the opacity to radiation, so crystals permanently bent are required for this type of applications.

Different methods to obtain a self-standing, cylindrical curvature have been studied and the surface lapping technique, developed at the IMEM/CNR in Parma, is one of the methods that reached a good technological maturity[9].[10] This technique consists in inducing a controlled damage by lapping one of the surfaces of the crystals, which generates an internal strain able to bend the crystal in self-standing way. The parameters of the lapping process, such as the grain of the machine, the sanding time and the final thickness of the crystals after the lapping procedure, can be set to bend the crystals to the desired curvature radius. The process can be paired with an etching process which can be used to reduce an excess of the curvature radius. By combining both methods, the curvature radius of the crystal can be fine-tuned in a very precise way. The focusing capabilities and the pass-band increase of crystals bent by surface lapping have been experimentally confirmed at our facility.[11]

For our prototype of Laue lens, we aim to use bent perfect crystals of Silicon and Germanium. Past tests done with mosaic GaAs(220) crystals curved to a radius of 40 m showed that bent crystals are able to focus high energy X-rays (Fig. 4). Now that the focusing capabilities of bent crystals has been widely assessed, we are moving to perfect Si(111) and Ge(111). According to the dynamical theory of

diffraction, the (111) planes will show the properties of increased efficiency described by Malgrange's theory of bent perfect crystals.

3. The TRILL project

The TRILL project (Technological Readiness Level Increase for Laue Lenses) is a project funded by the Italian Space Agency (ASI) for the increase of the Technological Readiness Level (TRL) of technologies with a potential use in future space missions. The aim of the TRILL project is to increase the TRL of a Laue lens in all its aspects, from the production of the crystals, to their integration on the lens itself, to the performance testing using an adequate spectroscopic-imager detectors used as focal plane detectors.

To build our first prototype of Laue lens, we decided for a modular approach: the full Laue lens in divided in spherical sectors called *petals* and each petal is then divided in smaller *modules*. Every module can host some tenth of crystals (Fig. 5). Aim of the TRILL project is to build 4 modules of a Laue lens and join them together in a first prototype of a subsector of a petal.

The TRILL project is structured on the following main tasks:

(1) To define a reliable and repeatable way to bend the crystals with no deterioration of their properties. To avoid distortion on the PSF of the single crystals, it is crucial that the curvature radius of the bent Si and Ge crystals is as uniform as possible and within an accuracy of ± 2 m from the desired curvature radius of 40 m.
(2) To find the best materials and bonding method to build a module of Laue lens. To reach the desired PSF size of 30 arcsec, we need that the crystal tiles are

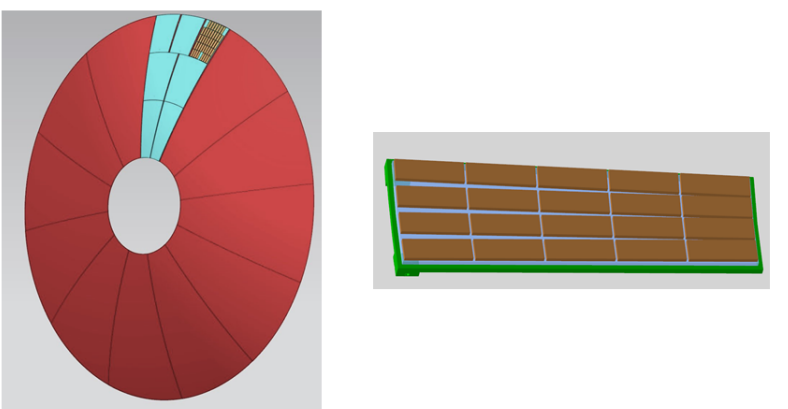


Fig. 5. Left: Schematic CAD model of a full Laue lens. The lens is divided in spherical sectors, called petals (red), and each petal is divided in a series of modules (cyan). The crystals are fixed on each module. Right: CAD model of a module of Laue lens, with 20 crystals on top.

set with an accuracy better than 10 arcsec, which requires us to find a very fine way to position the crystals.

(3) To find a way to join the modules together in one single piece. The modules will be connected in a way that will grant the possibility to adjust the position of each piece in an active way, to allow us to correct the position of the modules whenever required.

In this section, we illustrate the status of advancement of each of those 3 tasks.

The manufacturing of the crystals is done at the CNR/IMEM in Parma. Here, the crystals of Silicon and Germanium that we want to use for our prototypes are cut from wafers to obtain $30 \times 10 \times 2$ mm^3 tiles. Each tile is then bent via surface lapping to the desired curvature radius of 40 m. The procedure in the past was first tested on mosaic GaAs(220) crystals with a mosaicity of 13 arcsec. From those previous tests, we discovered that there is range of values for the crystal curvature radius for which the distortion of the image induced by a distorted curvature radius is completely masked by the mosaic defocusing effect. This allows us to reduce the constraints on the accuracy of the curvature radius of the crystals according to the value of the angular dispersion of the diffractive planes.

In the case of perfect Ge(220) crystals, the lapping procedure generates a mosaic structure on the diffraction planes of the crystals, which acquire a mosaicity of $\sim$5 arcsec. Given this value of mosaicity, if the curvature radius of the crystals is within $\pm$2 m from the nominal value of 40 m, the induced distortion on the PSF is negligible.

Last tests on Ge(220) gave very promising results: the values of the curvature radius of this first sample of 38 Ge(220) crystals were in the range [37.9 $\pm$ 0.9;

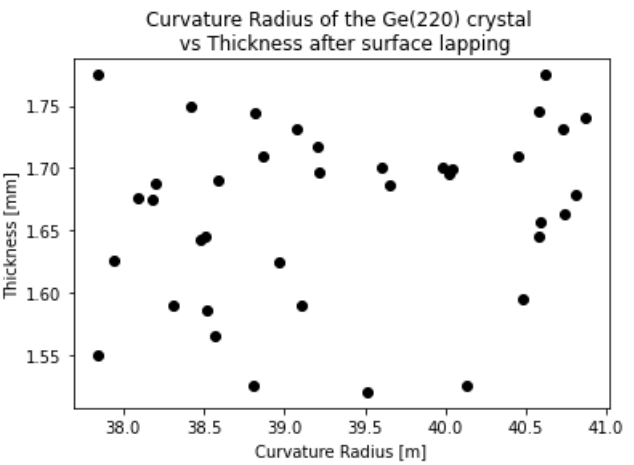


Fig. 6. Curvature radius of the sample of 38 bent crystals of Ge(220) vs their thickness after the surface lapping procedure. All the curvature radius are in the limit $\pm$2 m from the nominal curvature radius of 40 m.

40.9 ± 0.7] m, which means that every crystal is suitable for our applications. The average value of the curvature radii distribution is of 39.3 ± 0.2 m, with a standard deviation of 1 m (Fig. 6).

The lapping procedure, however, removes part of the material and the thickness of the crystals gets significantly decreased. The final thickness of the crystals after the lapping procedure is in the interval [1.520 ± 0.005; 1.775 ± 0.005] mm, with an average value of (1.665 ± 0.002) mm (Fig. 6).

The module assembly is the second, crucial task of the TRILL project. The crystals will be bonded on an adequate substrate. Ray-tracing simulations of a full Laue lens made by bent perfect crystals of Ge(111) allowed us to establish that it is necessary a positioning accuracy of the crystals'diffracting planes <10 arcsec to obtain the desired PSF accuracy of 30 arcsec for the full lens. Given that the focal length of the lens is of 20 m, this required accuracy translates in a positional accuracy of the crystals of the order of few microns, which means that high precision bonding techniques are required to integrate the crystal on the support of the prototype.

Currently each Laue lens module consists of a glass trapezoid substrate on which the crystals are glued with a low-shrinkage UV-curable glue. Each substrate is made of clear quartz, chosen due to its transparency and its low CTE. The substrate is shaped as an isosceles trapezium with bases 68 mm and 56 mm long, while the height is 183 mm. Given that the cross section of the crystals is 30×10 mm^2 and that the spacing between the crystals glued on the substrate will be of about 1 mm, each glass support can hold about 30 crystals. The surfaces of the substrate was bent to a curvature radius of 40 mm, the same of the crystals, at the National Institute of Optics of the National Research Council, in Florence. The substrate is hold in vertical position by an INVAR steel frame.

The crystals are positioned on the glass substrate using a motorized hexapod with six degrees of freedom: three translations and three rotations. The hexapod has a accuracy of the order of 1 micron on the translational movements and of 2×10^{-5} radians on the rotations (Fig. 7, left).

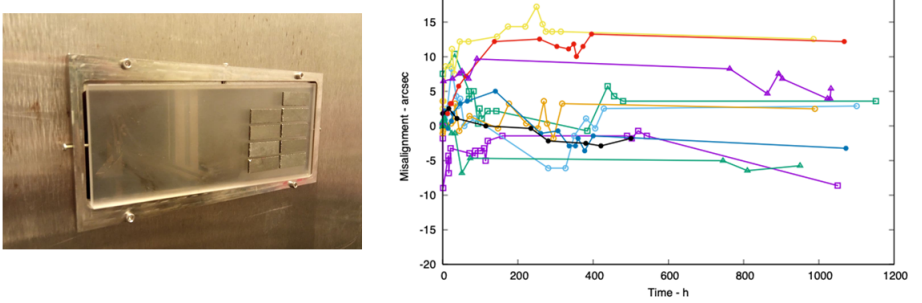


Fig. 7. Left: Laue lens module with nine crystals glued on top. The steel frame supporting the glass substrate is visible. Right: Misalignment of the crystals vs time from bonding.

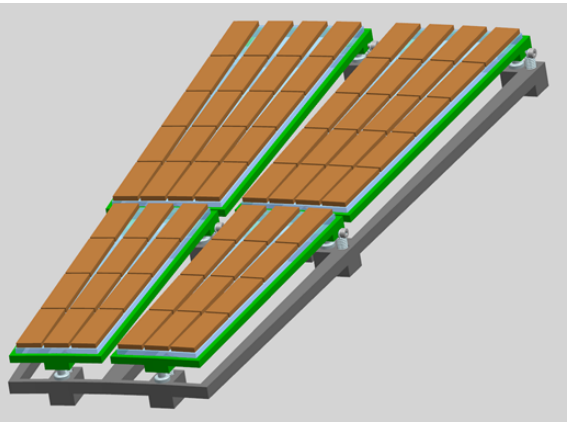


Fig. 8. CAD model of the four Laue lens module assembled. The module will be connected to a steel frame by means of micrometric screws.

The glue currently used to bond the crystals to the glass substrate is the *DYMAX OP-61-LS*, an UV curable, single component optical adhesive. This glue was choosen for its very low linear shrinkage of 0.03% of its length and for its high viscosity, which makes it suitable for a vertical bonding.

The crystals are glued with the following procedure: (1) the crystal is positioned on the hexapod and oriented to the proper Bragg's angle, then (2) a small drop of glue is deposited on the substrate with a glue dispenser, (3) the crystal is then put in contact with the glue and the position is fine tuned to correct eventual slight change in the position of the crystal, (4) the glue is cured with an UV lamp and finally (5) the crystal is released from the hexapod.

The position of the crystals is then tested immediately after the bonding and also when some time has passed. We found that the misalignment of the crystals after 50 days from bonding is within the interval +15/-10 arcsec (Fig. 7, right). These first results are quite good, since we are starting to get close to our target of <10 arcsec misalignment. However at the moment this technique shows some problems of stability in time and repeatibility. In particular, we are studying a better way to measure the quantity of glue deposited on the substrate and to control the illumination uniformity of the UV light used for the curing process.

The last task, i.e. the study of a way to assemble the modules together, is currently in study phase. At the moment, we plan to join the modules together by building a metal support frame in which the different modules will be fixed with high precision micrometric screws (Fig. 8). The screws will be set in position by using miniaturized piezoelectric motors, which will allow us to actively change the orientation of the module in any moment and to reduce as much as possible the effect that external solicitation can have on the final alignment of the prototype and, latter, of the full Laue lens itself.

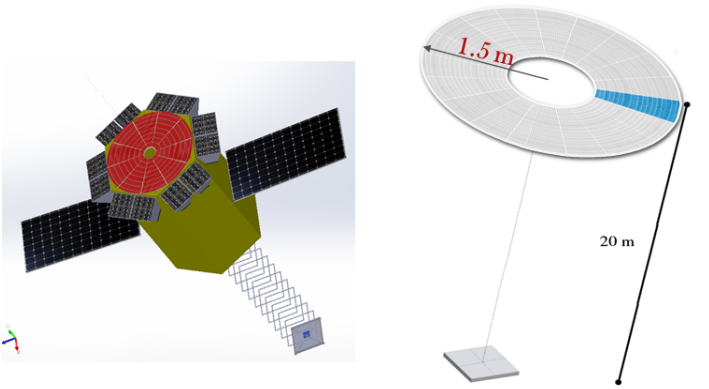


Fig. 9. Left: Representation of ASTENA's in-flight configuration. The Laue lens is shown on the top of spacecraft, in red. Right: Schematic representation of NFT. NFT will be made by a Laue lens optics with 3 m of diameter and 20 m of focal length. The focal plane detector will be a CZT spectral-imager detector.

The next step of the TRILL project will be to start the production of the bent Si(111) and Ge(111) crystals, to further optimize the bonding process and build the 4 modules and, finally, to assemble the modules together.

4. The NFT aboard ASTENA

Our studies on Laue lenses, including the TRILL project, are oriented to develope the technology for building the Narrow Field Telescope (NFT) on-board ASTENA, the Advanced Surveyor for Transient Events and Nuclear Astrophysics, a concept mission that we proposed for the ESA Call "Voyage 2050"[12].[13] The Narrow Field Telescope will be a revolutionary hard X/soft Gamma-ray focusing telescope working in the energy band 50-700 keV based on the technology of Laue lenses (Fig. 9). The NFT will be made by a Laue lens optics of 3 m of diameter with 20 m focal length. The lens will be made by about 19500 crystal tiles of perfect Si(111) and Ge(111) of size $30\times10\times2$ mm^3, bent to a curvature radius of 40 m. In the context of the TRILL project, we will then test the capabilities of the crystals that will be used for the lens and define the construction method for the whole lens.

The assumed focal plane detector will be a pixelated CdZnTe spectral-imager detector, with a size of $8\times8\times8$ cm^3, a pitch of 300 μm and an efficiency >80% in the whole energy band of NFT.

The combination between the characteristics of the optics and the detector will grant NFT of an unprecedented angular resolution in the sub-MeV energy range of 30 arcsec and a point source localization accuracy <10 arcsec, with a Field of View of 4 arcmin. NFT will bring a leap in sensitivity of two order of magnitude with respect to the best current instruments working in the same energy bands, opening a new range of possibilities for high energy astronomical observations.

Acknowledgements

This work is partly supported by the AHEAD-2020 Project grant agreement 871158 of the European Union's Horizon 2020 Programme and by the ASI-INAF agreement no. 2017-14-H.O "Studies for future scientific missions".

References

1. W. H. Zachariasen, *Theory of X-ray Diffraction in Crystals* (Wiley, 1945).
2. F. Frontera and P. V. Ballmoos, Laue gamma-ray lenses for space astrophysics:status and prospects (2011).
3. F. Frontera, G. Loffredo, A. Pisa, F. Nobili, V. Carassiti, F. Evangelisti, L. Landi, S. Squerzanti, E. Caroli, J. B. Stephen, K. H. Andersen, P. Courtois, N. Auricchio, L. Milani and B. Negri, Focusing of gamma-rays with Laue lenses: first results, in *Space Telescopes and Instrumentation 2008: Ultraviolet to Gamma Ray*, eds. M. J. L. Turner and K. A. Flanagan (SPIE, 2008).
4. E. Virgilli, F. Frontera, V. Valsan, V. Liccardo, V. Carassiti, S. Squerzanti, M. Statera, M. Parise, S. Chiozzi, F. Evangelisti, E. Caroli, J. Stephen, N. Auricchio, S. Silvestri, A. Basili, F. Cassese, L. Recanatesi, V. Guidi, V. Bellucci, R. Camattari, C. Ferrari, A. Zappettini, E. Buffagni, E. Bonnini, M. Pecora, S. Mottini and B. Negri, The laue project and its main results (2014).
5. A. Authier and C. Malgrange, Diffraction Physics, *Acta Crystallographica Section A* **54**, 806 (Nov 1998).
6. R. Camattari, V. Guidi, V. Bellucci and A. Mazzolari, The 'quasi-mosaic' effect in crystals and its applications in modern physics, *Journal of Applied Crystallography* **48** (08 2015).
7. V. Bellucci, V. Guidi, R. Camattari and I. Neri, Calculation of diffraction efficiency for curved crystals with arbitrary curvature radius, *Journal of Applied Crystallography* **46** (04 2013).
8. R. K. Smither, K. A. Saleem, D. E. Roa, M. A. Beno, P. von Ballmoos and G. K. Skinner, High diffraction efficiency, broadband, diffraction crystals for use in crystal diffraction lenses, *Experimental Astronomy* **20**, 201 (December 2005).
9. C. Ferrari, E. Buffagni, E. Bonnini and A. Zappettini, X-ray diffraction efficiency of bent GaAs mosaic crystals for the LAUE project, in *Optics for EUV, X-Ray, and Gamma-Ray Astronomy VI*, eds. S. L. O'Dell and G. Pareschi, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 88612013.
10. E. Buffagni, E. Bonnini, C. Ferrari, E. Virgilli and F. Frontera, X-ray characterization of curved crystals for hard x-ray astronomy, in *EUV and X-ray Optics: Synergy between Laboratory and Space IV*, eds. R. Hudec and L. Pina, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 95102015.
11. E. Virgilli, F. Frontera, P. Rosati, E. Bonnini, E. Buffagni, C. Ferrari, J. B. Stephen, E. Caroli, N. Auricchio, A. Basili and S. Silvestri, Focusing effect of bent GaAs crystals for γ-ray Laue lenses: Monte Carlo and experimental results, *Experimental Astronomy* **41**, 307 (February 2016).
12. F. Frontera, E. Virgilli, C. Guidorzi, P. Rosati, R. Diehl, T. Siegert, C. Fryer, L. Amati, N. Auricchio, R. Campana, E. Caroli, F. Fuschino, C. Labanti, M. Orlandini, E. Pian, J. B. Stephen, S. Del Sordo, C. Budtz-Jorgensen, I. Kuvvetli, S. Brandt, R. M. C. da Silva, P. Laurent, E. Bozzo, P. Mazzali and M. D. Valle, Understanding the origin of the positron annihilation line and the physics of supernova explosions, *Experimental Astronomy* **51**, 1175 (Jun 2021).

13. C. Guidorzi, F. Frontera, G. Ghirlanda, G. Stratta, C. G. Mundell, E. Virgilli, P. Rosati, E. Caroli, L. Amati, E. Pian, S. Kobayashi, G. Ghisellini, C. Fryer, M. D. Valle, R. Margutti, M. Marongiu, R. Martone, R. Campana, F. Fuschino, C. Labanti, M. Orlandini, J. B. Stephen, S. Brandt, R. C. d. Silva, P. Laurent, R. Mochkovitch, E. Bozzo, R. Ciolfi, L. Burderi and T. Di Salvo, A deep study of the high–energy transient sky, *Experimental Astronomy* **51**, 1203 (Jun 2021).

ASTENA: A mission concept for a deep study of the transient gamma-ray sky and for nuclear astrophysics

E. Virgilli[1,2,*], F. Frontera[1,3], P. Rosati[1,3,4], C. Guidorzi[1,3,4], L. Ferro[3], M. Moita[3], M. Orlandini[1], F. Fuschino[1,2], R. Campana[1,2], C. Labanti[1], E. Marchesini[1], E. Caroli[1], N. Auricchio[1], J. B. Stephen[1], C. Ferrari[5], S. Squerzanti[4], S. Del Sordo[6], C. Gargano[6] and M. Pucci[7]
on behalf of the ASTENA Collaboration

[1] *INAF - OAS di Bologna, Via Piero Gobetti 93/3, I-40129 Bologna, Italy*
[2] *INFN Sezione di Bologna Viale C. Berti Pichat, 6/2 – 40127 Bologna, Italy*
[3] *Dept. of Physics and Earth Science, Univ. of Ferrara, Via Saragat 1, I-44122, Ferrara, Italy*
[4] *Istituto Nazionale di Fisica Nucleare, INFN-Sezione di Ferrara, 44122 Ferrara, Italy*
[5] *CNR-IMEM Institute, Parco Area delle Scienze 37/A, 43124 Parma, Italy*
[6] *INAF/IASF-Palermo, Via Ugo La Malfa 153, 90146 Palermo, Italy*
[7] *CNR - Istituto Nazionale di Ottica - Largo Fermi 6, 50125 Firenze, Italy*
[*] *E-mail: enrico.virgilli@inaf.it*

Gamma-ray astronomy is a branch whose potential has not yet been fully exploited. The observations of elemental and isotopic abundances in supernova (SN) explosions are key probes not only of the stellar structure and evolution but also for understanding the physics that makes Type-Ia SNe as standard candles for the study of the Universe expansion properties. In spite of its crucial role, nuclear astrophysics remains a poorly explored field mainly for the typical emission lines intensity which are vanishing small and requires very high sensitivities of the telescopes. Furthermore, in spite that the Galactic bulge-dominated intensity of positron annihilation line at 511 keV has been measured, its origin is still a mystery due to the poor angular resolution and insufficient sensitivity of the commonly employed instrumentation in the sub-MeV energy domain. To answer these scientific issues a jump in sensitivity and angular resolution with respect to the present instrumentation is required. Conceived within the EU project AHEAD, a new high energy mission, capable of tackling the previously mentioned topics, has been proposed. This concept of mission named ASTENA (Advanced Surveyor of Transient Events and Nuclear Astrophysics), includes two instruments: a Wide Field Monitor with Imaging and Spectroscopic (WFM-IS, 2 keV - 20 MeV) capabilities and a Narrow Field Telescope (NFT, 50 - 700 keV). Thanks to the combination of angular resolution, sensitivity and large FoV, ASTENA will be a breakthrough in the hard X and soft gamma–ray energy band, also enabling polarimetry in this energy band. In this talk the science goals of the mission are discussed, the payload configuration is described and expected performances in observing key targets are shown.

Keywords: Soft gamma-rays; hard x-ray telescope; Focusing gamma rays; Bragg diffraction; bent crystals.

1. Introduction

Within the H2020 European program AHEAD,[1,2] (integrated Activities for High Energy Astrophysics Domain) devoted to the assessment of future gamma-ray experiments, a Scientific Advisory Group recommended the prime scientific questions that might be addressed by a future space mission operating in the gamma-ray energy domain. The high-priorities themes resulted to be the nuclear astrophysics and

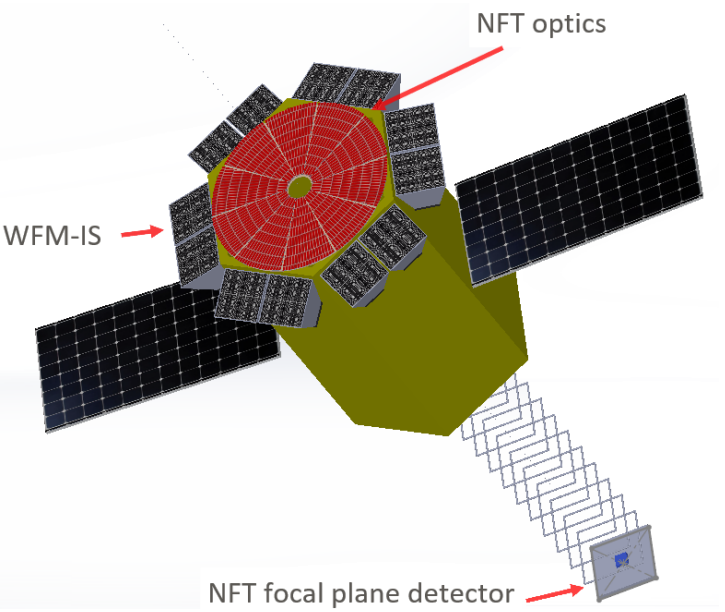


Fig. 1. Drawing of the ASTENA mission. The NFT (in red) is a 20 m focal length Laue lens made with bent Germanium and Silicon crystals. At the focal distance is positioned a focal plane detector (blue) which is a solid state device made with 4 layers of Cadmium Zinc Telluride. The focal distance is achieved through a deployable mast which at launch is fully enclosed in the spacecraft and in operative condition extends from the bottom of the spacecraft for 15 m, which is a reasonable distance for the stability of an extendable structure. The WFM-IS composed of 12 Position Sensitive Detectors (PSDs) distributed around the NFT and oriented 15 degrees two by two outwards with respect to the Laue lens axis in order to extend the FoV of the overall instrument.

the study of the transient sky. According to those themes, the ASTENA (Advanced Surveyor for Transient Events and Nuclear Astrophysics) has been designed. The ASTENA concept mission, which is shown in Fig. 1, is a broad energy pass-band experiment composed by two complementary instruments. The first is a Wide Field Monitor with Imaging and Spectrometric capabilities (WFM-IS) with an outstanding broad energy pass-band from 2 keV to 20 MeV. It consists on 12 coded mask cameras deployed in a circular pattern around the hexagonal spacecraft and oriented at 15 degrees with respect to the spacecraft axis. The overall Field of View (FoV) of the instrument is $\sim$2 steradians. The second instrument is a narrow FoV (few arcminutes) telescope based on a Laue lens with a geometric area of $\sim$7 m^2 and 20 m focal length, capable to focus photons in the broad energy pass-band 50 - 700 keV on a solid state detector. The Narrow Field Telescope (NFT) represents a technological breakthrough as its optics, based on the diffraction from bent crystals, provides an unprecedented sensitivity with respect to any other mission flown and operative in the same energy pass-band. At launch, we expect to keep the WFM-IS cameras and the focal plane detector stored within the cylindrical spacecraft

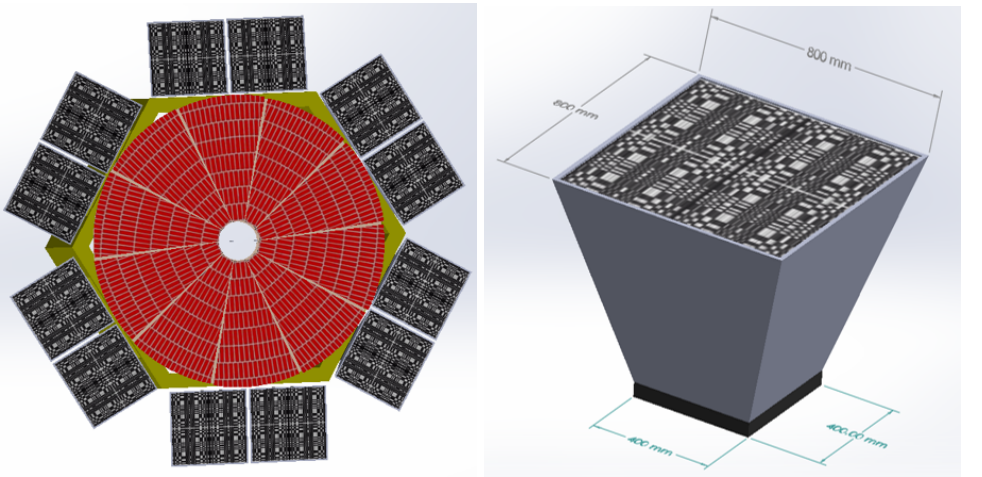


Fig. 2. Left: Top view of the instruments on board ASTENA. The 12 PSD of the WFM/IS are coupled 2 by two and are positioned around the hexagonal structure of the spacecraft. Each PSD is surmounted by a coded mask which ensures the imaging capability up to 150 keV. The red part is the broad energy pass-band Laue with 15 m focal length. Right: Detail of one PSD in which is visible the coded mask and the collimator surrounding the instrument. At the bottom (dark profile) is visible the detection plane made with scintillator bars which are coupled with SDDs (see text for the explanation of the instrument working principle).

whose diameter is 1.5 m and the length is 5 m. In the operational configuration an extendable boom brings the focal plane detector 15 m apart from the bottom of the spacecraft while a mechanism discloses and tilts the coded mask cameras at their nominal position and angle. In the following currently on-going project called AHEAD2020 the mission concept is being refined and optimized. The mission has been proposed in the ESA call "Voyage 2050"[3] as a future medium class mission for hard X-/soft gamma-ray astrophysics.[4,5] The final recommendation of the Senior Committee has confirmed the key importance of the high energy observations from space with high sensitivity and capable of enabling spectro-polarimetry based on new technologies, particularly in synergy with gravitational wave astronomy for resolving some of the fundamental questions still unanswered in astrophysics related to the nucleosynthesis in explosive events or to the accretion mechanism on compact sources.

2. The ASTENA configuration

The ASTENA mission concept builds on the complementarity between the two instruments on board. The WFM-IS is composed by 12 coded mask cameras equipped with Position Sensitive Detectors (visible in Fig.2 - left) (PSDs) distributed two by two around the hexagonal spacecraft envelope. Each pair of PSDs in the same hexagon side is co-aligned and directed towards the same direction which is radially tilted outwards of 15 degrees with respect to the satellite axis. Each PSD has a size

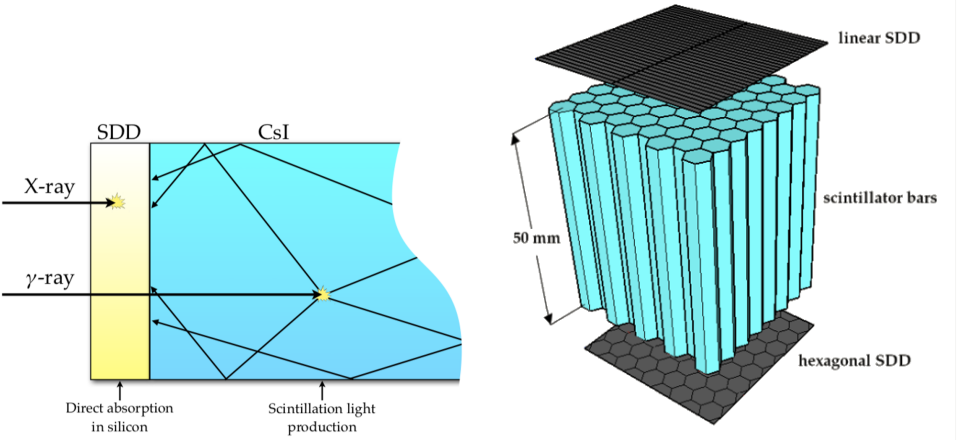


Fig. 3. Left: working principle of the WFM-IS. An hexagonal scintillator bar (cyan) is coupled at both ends (only one end is visible in the picture) with an SDD (yellow). The low energy radiation (< 30 keV) is directly absorbed in the SDD while higher energy photons pass through the SDD and interact in the scintillator bar. The scintillation light, which is produced at some distance from the surface of the scintillator bar and is reflected by the lateral surfaces, is collected by the two SDDs. The comparison between the two signals allows for the determination of the depth of interaction. Right: disposition of the scintillator bars (cyan) and of the top and bottom SDDs (grey). The top SDD is equipped with an array of linear anodes therefore the charge collected provides information only in one direction. In the bottom SDD the anodic structure has a hexagonal shape in order to fully exploit this geometry which minimizes the directional bias for polarimetric measurements, if compared with a detector with square (cubic) pixelization (voxelization).

of 43×42 cm^2 and is made of ~ 6500 scintillator bars with hexagonal cross section (5 mm between flat sides) and 50 mm long. Both ends of the scintillator bars are optically coupled with two Silicon Drift Detectors (SDDs) 400 μm thick. The instrument is based on the same detection principle of the X and Gamma-ray Imager and Spectrometer (XGIS) on board the THESEUS mission[6] which was proposed as 5^{th} medium-class mission for the ESA Cosmic Vision Programme (M5).

The detection principle (see Fig. 3) is based on a different interaction of the radiation with the system, depending on the energy of the interacting photons. Low energy photons (< 30 keV) are directly absorbed in the SDD while photons with higher energy pass through the SDD and interact in the scintillator bar. The scintillation light, which is produced at some distance from the ends of the scintillator bars, is reflected by the lateral surfaces which are properly polished and wrapped in order to reflect as much as possible the light produced, and it is collected by the two SDDs optically coupled at both ends of the bar. For high energy photons (>30 keV) the detector is position sensitive with a 3-D resolution given that the position of interaction of the gamma-ray photon along the bar can be reconstructed by comparing the signal collected from the two opposed SDDs. The top SDDs (those facing the sky) have the dual purpose of directly detecting the low energy photons (below few tens of keV) and of collecting the scintillation light emitted

by the scintillator bar when photons with higher energy pass through them. They have a linear anodic structure therefore are sensitive only to one direction (X or Y) with spatial resolution of 1.25 mm. Given that each pair of PSDs placed at the same side of the spacecraft have perpendicular anodes, they behave like a 2-D detector. Instead, the bottom SDD has hexagonal shape and it is used to get the 3-D position sensitivity. In this way we can minimize the background through the Compton kinematics reconstruction of the trajectory of the photons and exploit the polarimetric capability of the instrument. Different scintillator materials as Cesium Iodide (CsI(Tl)), Gadolinium Aluminium Gallium Garnet (GAGG), Lutetium Yttrium Orthosilicate (LYSO(Ce)) are under investigation in order to find the optimal properties to fit the instrument requirements. Each PSD is surmounted by a double scale[7] square coded mask with side of 80 cm at a distance of 70 cm from the detector plane (see Fig. 2 - right). The double scale enables imaging with Point Source Location Accuracy (PSLA) of 1 arcmin for a 7σ signal for photons with energy $<$ 30 keV and PSLA of 5 arcmin for photons in the energy range 30 - 150 keV. In Fig. 4 (left) the PSLA as a function of the significance of the observation for different configurations of the WFM-IS is reported, compared with required PSLA of 1 arcminute. In Fig. 4 (right) the integrated sensitivity of the overall WFM-IS is reported as a function of the integration time, divided in three relevant pass-bands according to the different detection principle employed (2 - 10 keV: direct detection in the SDDs, 30 - 150 keV: interaction with the scintillator bars, 150 keV - 10 MeV: uncollimated interaction in the scintillator bar). Above 150 keV the coded mask is transparent to radiation and an effective imaging capability can be enabled through two features: 1. by exploiting the Compton kinematics for the photon direction reconstruction and 2. by taking advantage of the different measured photon intensity from the 6 blocks which are - in general - differently oriented with respect to the direction of the observed event, except for perfectly on-axis sources. With these configuration the instrument can provide a FoV of $\sim$2 sr with an angular resolution of a few arcmin and an unprecedented energy pass-band of 2 keV - 20 MeV which has never been achieved so far with a single device.

The second instrument is a 20 m long focal length Laue lens based on bent crystals made with Germanium and Silicon. The crystals are distributed in 43 concentric rings and, according to the spherical geometry and to the Bragg law, the lower energies are reflected from outer rings while inner rings are responsible for focusing the higher energies. The size of the crystals have been chosen to be 30×10 mm^2, the longer dimension being the focusing direction. In the other direction, no concentration is expected due to the cylindrical curvature of the tiles. The cross section of the tiles has been chosen in order to tile the overall geometric area with a moderate number of crystals ($\sim$19500) by minimizing the amount of uncovered area (the optics filling factor results above 85%). Depending on the energy to be diffracted the crystal thickness is optimized in order to maximize the diffraction efficiency. For both materials the 111 planes are assumed in order to exploit the

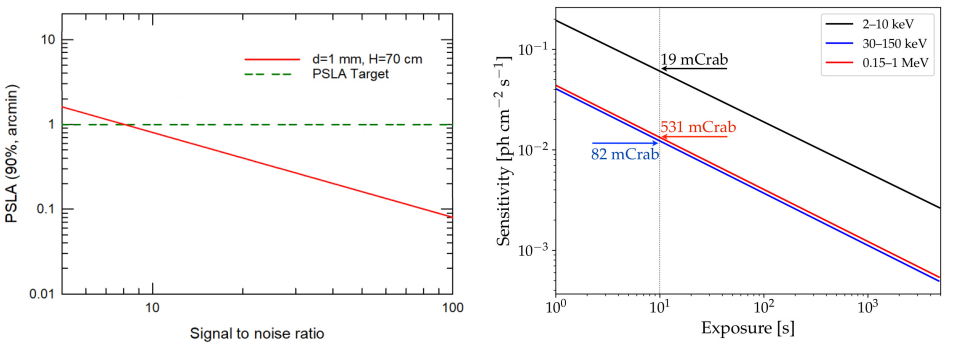


Fig. 4. Left: Point Source Location Accuracy (PSLA) for the WFM-IS as a function of the signal to noise ratio. The parameter d is the distance between the anodes of the top SDDs. The requird PSLA is also shown (green dashed line) Right: Integrated continuum sensitivity expressed in photons/cm^2/s as a function of the exposure time for the three significant energy pass-bands acquired for direct absorption in the SDDs (black curve, 2 - 10 keV), through the scintillation bars (30 - 150 keV, blue curve) and as a Compton detector with no coded mask (150 keV - 1 MeV, red curve).

so called secondary curvature of the diffraction planes induced in crystals for some crystallographic orientation (including the 111) allowing to achieve a throughput which overcomes the limitation of the 50% of the incident beam, which is the limit of mosaic and flat perfect crystals.

Orders of diffraction higher than the fundamental are being considered in the simulations in order to maximize the effective area, particularly in the energy range 500 - 520 keV (at the expenses of the effective area at lower energies) in order to increase the sensitivity at the energy of interest for the detection of the weak e^+/e^- annihilation line from the Galactic center (see Sect. 3 for further details). At 20 m from the optics is placed the focal plane detector which is kept at the correct position with a deployable mast. The detector is made with 4 layers of Cadmium Zinc Telluride (CZT) with cross section 80 $\times$ 80 mm^2 each with a thickness of 20 mm. The detector has a 3-D spatial resolution of about 300 μm in all directions. This is achievable with a proper disposition of the anodes on the top and bottom of each CZT layer. Thanks to the overall detector thickness its detection efficiency is greater than 80% at 600 keV with energy resolution of 1% at 511 keV.

The NFT continuum sensitivity which has been derived at Low Earth Orbit (LEO) by comparison with the background measured by the Spectrometer on-board INTEGRAL SPI[8] is reported in Fig. 6 (top). The curve has been estimated for an integration time of 10^5 s and at 3σ confidence level. For comparison, in the plot has been reported a number of past and present experiments exploiting the direct view principle (equipped with coded masks or collimators) or using focusing telescopes. The energy band in which it is estimated is $\Delta E = E/2$, except in the band from 50 keV to 62.5 keV, in which the energy band linearly increases from $\Delta E = E/4$ to $\Delta E = E/2$, due to the absence of crystals diffracting energies

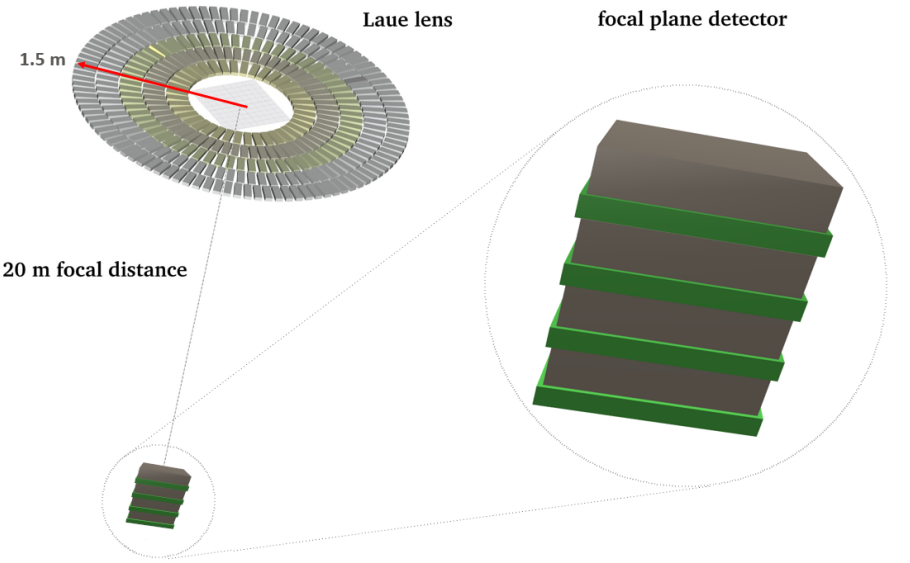


Fig. 5. Left: The optics of the NFT composed by 43 rings of crystal tiles made with Silicon and Germanium. The optics has a radius of 1.5 m for a total geometric area of about 7 m^2. At 20 m from the Laue lens is placed the focal plane detector. Right: a Geant4 drawing of the main components of the focal plane detector which is made of 4 layers of a solid state Cadmium Zinc Telluride (CZT, dark grey) with 3-D spatial resolution, spectral and polarimetric capabilities.

below 50 keV. Uncertainties in the realization of the optics have been also taken into account. The curvature radius of the crystals have been considered within 5% of the nominal radius and the mounting accuracy of about 10 arcseconds with respect to the nominal positioning. The presence of such deviations from the nominal lens realization requirements have effects on the size of the PSF and, ultimately, on the focusing power of the Laue lens. This unprecedented sensitivity is achieved thanks to the use of bent crystals with optimized thickness. Figure 6 (bottom) shows the sensitivity at measuring the flux of emission lines, at 3σ confidence level and for 10^5 s integration time. An intrinsic FWHM of 2 keV has been considered for the lines (e.g. see [9] for the 158 keV line from SN2014J). The narrow line sensitivity is about 1 order of magnitude better than SPI on board INTEGRAL at 511 keV. These values of sensitivities are mainly due to the use of bent crystals, to the transmission geometry and to the focal length that allows for an unprecedented large collecting area.

Laue optics have the great advantage of drastically reducing the instrumental background as the photons are mainly collected into an area of a few mm^2. The drawback is that, as far as the source moves out of the focal axis, the diffracted image spreads over a larger area generating ring-shaped images for coma aberrations. Through ray-tracing and Monte Carlo simulations it has been estimated that,

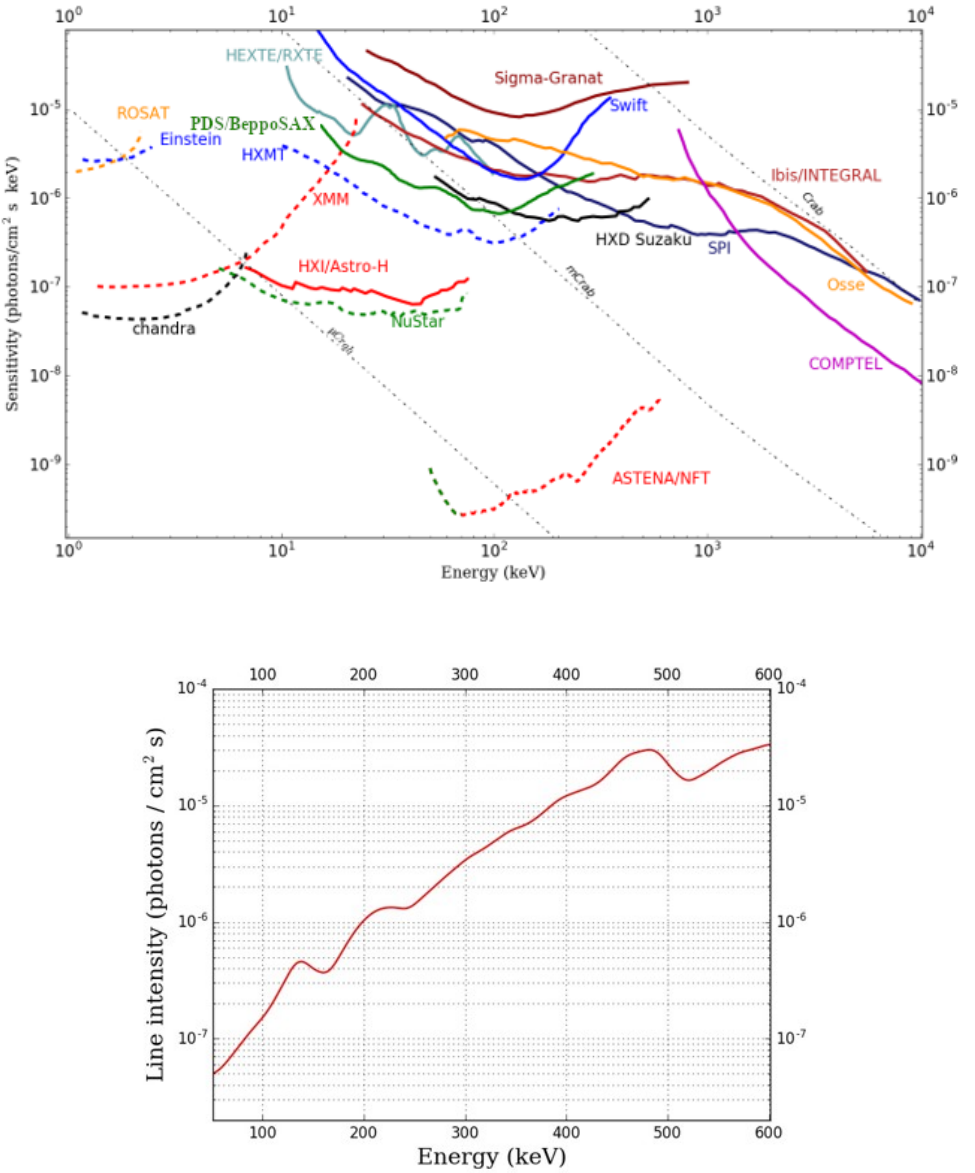


Fig. 6. Top: NFT continuum sensitivity at 3σ confidence level and in 10^5 s of observation time, with $\Delta E = E/2$, except in the band 50 to 62.5 keV, where ΔE linearly increases from $\Delta E = E/4$ to $\Delta E = E/2$, compared with the sensitivity at the same significance and with the same integration time of other missions or experiments (dashed lines represent other focusing experiment, continuum curves correspond to direct view instruments). Bottom: Expected line sensitivity for the NFT, calculated for an integration time of 10^5 s, at 3σ confidence level.

with the defined optics dimension and crystal size, photons from off-axis sources (>4 arcmin) are scattered into a broad detection area.[10] This reduces the benefits of the focusing process. Therefore the FOV of the NFT has been fitted with the PSLA of the WFM-IS in order to exploit the sensitivity of the former for nearly on-axis sources with the excellent broad-band survey capabilities of the latter.

3. Key science with ASTENA

For some relevant scientific key subjects that can be tackled with the instruments on board ASTENA see [10, 11]. Here we summarize these subjects. As mentioned in Sect. 1 the main objectives of the ASTENA mission concept are mainly two. First, ASTENA intends to investigate the intriguing key questions related to the gamma-ray lines in astrophysics. Furthermore, the high sensitivity of the on board instruments joined with the large pass-band of the WFM-IS and their polarimetric capabilities would make ASTENA an ideal tool for shading light in the transient sky, providing also an important contributions to the multi-messenger synergy.

3.1. *The transient sky observed with ASTENA*

The discovery of the Gravitational Wave (GW) event GW170817[12] and its electromagnetic counterpart GRB170817A[13–16] marked the birth of the multi-messenger study of the transient sky. In the near future, this plural effort will solve fundamental astrophysics and cosmological questions. Furthermore, it will boost the discovery of sources of known and unknown classes of transients. Short GRBs represent the main class of transient which has been already confirmed to be associated with powerful GW events.

Polarization level in the range 30 - 80% of the prompt emission of the GRBs has been claimed for a few dozens events (a summary of the properties of the events can be found in [17, 18]) but, due to the limited statistical significance of the results, their confirmation is not definitive. These uncertainties are mainly due to the lack of sensitivity of the instruments used as gamma-ray polarimeters. A significant detection could probe the geometry of the magnetic field and its intensity which are precious information for shedding light into the jet composition and dissipation mechanism. Thanks to its polarimetric capabilities, along with a large detection area and broad pass-band, the WFM-IS would be the ideal instrument to measure the degree of polarization of the prompt emission, to perform a detailed time-resolved study and to evaluate the dependence of the polarization degree with energy.

Instruments capable of measuring the electromagnetic counterpart of GW events have a crucial importance for different reasons. Firstly, they can independently confirm the astrophysical nature of a GW trigger, particularly for the faintest events. Furthermore, since present interferometers have large uncertainties in the sky localization, a high sensitive wide FoV instrument like the WFM-IS could provide the localization of the order of 1 arcmin with the possibility of performing followup

observation with a soft gamma-ray telescope with an unprecedented sensitivity and angular resolution as the NFT.

It must be mentioned that with the launch of the Laser Interferometer Space Antenna (LISA) a plethora of GW events from stellar mass BH binaries will be detected in the mHz regime with a sky localization of the order of 1 deg^2. Such detections will work as alert for the observation of the same events at higher frequencies weeks/months later (with an accuracy of the order of tens of seconds) from ground interferometers. Thanks to the large FoV and to the PSLA of the WFM-IS will be possible to point in advance the instrument for the detection of the prompt hard X-ray counterpart of the event as well as the detection of the delayed hard X-ray emission through the NFT.

After the discovery of the GRB afterglow, till now this emission has been observed almost in all energy bands from IR to the soft X-ray regime up to ~ 10 keV. Afterglow detection in the 0.1–10 GeV, in the sub-TeV (100–440 GeV) and in the TeV energy regions have been reported.[19,20] One of the most important open issues that have still to be settled is the afterglow emission in the sub-MeV/MeV region which is almost completely undetected except for some events detected above few tens of keV[21–23] This is mainly due to the low flux emitted in this energy band, joined with the lack of sensitivity for the present instrumentation. From the prompt emission detected with the WFM-IS and though a fast repointing it would be possible, thanks to the sensitivity of the NFT, to measure the hard x-ray afterglow as well as its polarization level.

3.2. *Nuclear astrophysics with ASTENA*

In spite of its importance for understanding the inner regions of the astrophysical sources emitting hard X and mainly gamma-rays, nuclear astrophysics is still a theoretical field and almost experimentally unexplored, mainly due to observational limits. The reason is that present instrumentation in the hard-X and gamma-ray regime has absent or crude imaging capability and low sensitivity, compared with instruments in other wavelengths. One of the most relevant open issues in astrophysics is the origin of the 511 keV positron annihilation line observed from the Galactic Center (GC). Discovered in the seventies[24] neither the origin of the gamma emission, nor the responsible for the positron production have been found. The 511 keV annihilation line still represents a puzzle mainly due to the limitation of both sensitivity and angular resolution of current instruments operative in the 0.5 MeV region. To date, the best mapping of the line was obtained through SPI on-board INTEGRAL.[25] A variety of potential astrophysical sources responsible of the 511 keV signal and/or of the production of the positrons have been proposed, including type Ia supernovae,[26] GRBs,[27] microquasars,[28] low-mass X-ray binaries,[29] and neutron star mergers.[30] The high number of X–ray sources in the GC suggests the possibility that the 511 keV line is due to the emission of a number of discrete - unresolved - sources. The detection of a transient 0.5 MeV emission from V404 Cygni[31] is in

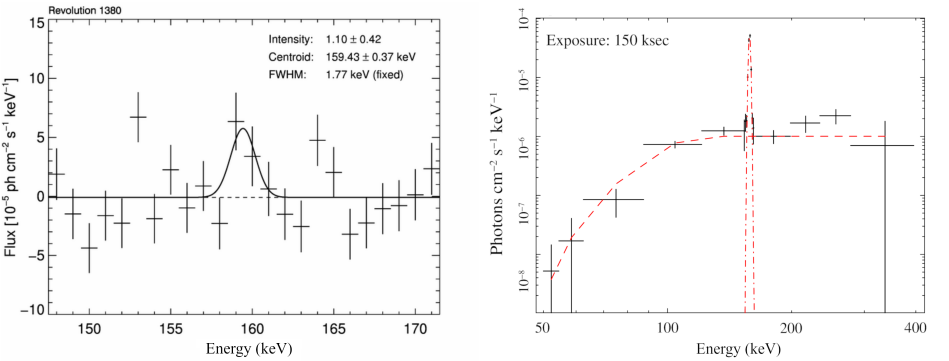


Fig. 7. Left: the 158 keV line due to the ^{56}Ni decay as observed from SN2014J as with the SPI on board INTEGRAL 3 weeks after the explosion. Reprinted from [9]. Right: the same line as observed with the NFT onboard ASTENA with 150 ks integration time. The simulated data (black points) consists of a continuum modeled with an absorbed power-law (red dashed line) with spectral index $\alpha = 0$ normalization 10^{-6} ph/cm^2/s @ 1 keV plus a Gaussian (red dot-dashed line) centered at 158 keV with σ=0.75 keV and normalization 1.1×10^{-4} ph/cm^2/s.

favour of this hypothesis. It is also believed that the 511 keV observed map would only represent the annihilation sites and not the positron sources. The propagation of the positrons away from the source could be the cause of a general broadening which originates the diffuse 511 keV emission. No point sources of annihilation radiation have yet been detected in the GC. Nevertheless, the angular resolution of INTEGRAL/SPI does not provide any definitive information about structure in the emission, even with several years of integrated data.[32] Under these considerations it is clear that observations of the GC with much higher angular resolution than that achievable with SPI aboard INTEGRAL together with high sensitive instruments are required in order to distinguish and localize - if any - discrete source of annihilation radiation.

Another topic under study that is worth to be mentioned is related to the nucleosynthesis of heavy elements in Type-Ia supernovae (SNe-Ia). Gamma-rays escaping the ejecta of SN-Ia can be used as tools for studying both the structure of the exploding star and the characteristics of the explosion. One of the key points is to estimate the amount of ^{56}Ni which is probably the most important physical parameter underlying the observed correlation of SNe-Ia luminosities with their light curves. The only direct way to estimate this amount is through the decay chain $^{56}\text{Ni} \rightarrow {}^{56}\text{Co} \rightarrow {}^{56}\text{Fe}$. The decay produces gamma-rays that ultimately power the optical light curve of the SN-Ia. Most models predict that the ejected material is not transparent to gamma-rays at least for tens of days since the explosion. As the ejecta proceeds, it becomes gradually more transparent to gamma-rays. Contrarily to the expectations, only two weeks after SN2014J through SPI/INTEGRAL observations, the characteristic gamma-ray lines 158 keV and 812 keV from the ^{56}Ni decay have been detected [9] and later confirmed at 5σ confidence level also with

ISGRI/INTEGRAL.[33] The early observations of gamma-rays from ^{56}Ni support the hypotesis of the presence of this element at the star surface. The intensity of the gamma-ray spectra mainly depends on the mass and distribution of the nickel mass. High sensitivity measurements focused at particular narrow lines would shed light on the SNe-Ia explosion mechanism and, ultimately to the correct distribution and mass of ^{56}Ni.

As an example of performances with ASTENA, in Fig. 7 we show the 158 keV line detected with SPI/INTEGRAL from SN2014J[9] for an exposure time of 150 ks. The estimated line intensity of $(1.1\pm 0.4) \times 10^{-4}$ ph/cm^2/s corresponds to a detection confidence level of 2.5σ. For comparison, with the same integration time, a spectrum consisting of a Gaussian profile (centroid energy 158 keV, $\sigma = 0.75$ keV) superposed to a continuum described by a power-law with spectra slope α=0 and normalization 10^{-6} ph/cm^2/s @ 1 keV has been simulated in order to estimate how it could be observed with the NFT aboard ASTENA. The same significance achieved by SPI in 150 ks is obtained in $\sim$1 ks.

4. Conclusions

Future missions for hard X-ray astrophysics require higher sensitivity than current instrumentation and capability of observing the sky in a broad energy pass-band. Such features will be crucial for solving some of the most important open issues. In addition, hard-X and soft gamma-ray polarimetry, which are still in their infancy, will play a crucial role in combination with spectroscopy, timing and imaging, providing information about the degree and the direction of polarization of the incident radiation. Particularly in hard X-rays, whose emission comes directly from the central engine of the sources, polarimetry is a powerful tool for investigating the magnetic field and the distribution of matter around astrophysical objects. Moreover, today hard X-/gamma-ray astrophysics suffers from the absence of devices that allow the radiation to be focused, the only method through which it is possible to increase the signal to noise of the observations and to suppress the instrumental background.

The ASTENA mission concept includes two complementary instruments through which it is possible to carry out both polarimetry and imaging in the hard X-ray regime. The WFM-IS is an array of 12 monitors that can effectively work in an unprecedented energy pass-band for a single instrument (2 keV - 20 MeV) thanks to the coupling of SDDs a and scintillator bars. Thanks to the adoption of a double scale coded mask, the instrument enables localization accuracy of about 1 arcmin and 5 arcmin in the energy pass-bands $<$ 30 keV and 30 - 150 keV, respectively. With this location accuracy, the WFM-IS is in complete synergy with the second instrument on board, the NFT: for the first time a focusing optics based on a Laue lens will be operative in a broad pass-band and with a FoV of a 4-5 arcminutes and an angular resolution of a few tens of arcseconds. Thanks to its focusing capability it provides an outstanding sensitivity for detecting source polarization[34] and both for

the continuum and for nuclear lines. The unprecedented broad pass band and the large FoV of the WFM-IS will allow to make surveys and to detect faint transient sources, including GRBs. The NFT, with its deep sensitivity for nearly on-axis sources, will be the ideal tool for followup observations of detected transient events.

In this work we have presented the ASTENA mission concept and details on both instruments aboard. We have described some key science issues that can be tackled with ASTENA as a stand-alone experiment and in synergy with other experiments, including its contribution to the multi-messenger astrophysics. We expect that a mission like ASTENA will be a breakthrough in general astrophysics and in particular for answering the questions that are still central in high energy astrophysics.

Acknowledgements

This work has been partially supported with the financial contribution from the ASI-INAF agreement n. 2017-14-H.0 "Studi per future missioni scientifiche", the ASI-INAF agreement n. 2018-10-H.1-2020 "HERMES Technologic Pathfinder", and the AHEAD EU Horizon 2020 project (Integrated Activities in the High Energy Astrophysics Domain), grant agreement n. 871158.

References

1. L. Piro, L. Natalucci and Ahead Consortium, AHEAD: Integrated Activities in the High Energy Astrophysics Domain, in *Exploring the Hot and Energetic Universe: The first scientific conference dedicated to the Athena X-ray observatory*.
2. L. Piro, The Activities for the High-Energy Astrophysics Domain (AHEAD), in *Whereabouts and Physics of the Roaming Baryons in the Universe*.
3. F. Favata, G. Hasinger, L. J. Tacconi, C. S. Arridge and K. S. O'Flaherty, Introducing the Voyage 2050 White Papers, contributions from the science community to ESA's long-term plan for the Scientific Programme, *Experimental Astronomy* **51**, 551 (June 2021).
4. F. Frontera, E. Virgilli, C. Guidorzi, P. Rosati, R. Diehl, T. Siegert, C. Fryer, L. Amati, N. Auricchio, R. Campana, E. Caroli, F. Fuschino, C. Labanti, M. Orlandini, E. Pian, J. B. Stephen, S. Del Sordo, C. Budtz-Jorgensen, I. Kuvvetli, S. Brandt, R. M. Curado da Silva, P. Laurent, E. Bozzo, P. Mazzali and M. Della Valle, Understanding the origin of the positron annihilation line and the physics of the supernova explosions, *White paper in response to Voyage 2050 long term planning of ESA science programme* **1**, p. 20 (Oct 2019).
5. C. Guidorzi, F. Frontera, G. Ghirlanda, G. stratta, C. Mundell, E. Virgilli, P. Rosati, E. Caroli, L. Amati, E. Pian, S. Kobayashi, G. Ghisellini, C. Fryer, M. Della Valle, E. Margutti, M. Marongiu, R. Martone, R. Campana, F. Fuschino, C. Labanti, M. Orlandini, J. B. Stephen, S. Brandt, R. M. Curado da Silva, L. P., M. R., E. Bozzo, R. Ciolfi, L. Burderi and T. Di Salvo, Deeper and broder: future observations in the X-/gamma-ray band of known and unknown, *White paper in response to Voyage 2050 long term planning pf ESA science programme* **1**, p. 20 (Oct 2019).
6. L. Amati, P. O'Brien, D. Götz, E. Bozzo, C. Tenzer *et al.*, The THESEUS space mission concept: science case, design and expected performances, *Advances in Space Research* **62**, 191 (July 2018).

7. G. K. Skinner and J. E. Grindlay, Coded masks with two spatial scales, *Astronomy and Astrophysics* **276**, p. 673 (September 1993).
8. G. Vedrenne, J. P. Roques, V. Schönfelder, P. Mandrou, G. G. Lichti, A. von Kienlin, B. Cordier, S. Schanne, J. Knödlseder, G. Skinner, P. Jean, F. Sanchez, P. Caraveo, B. Teegarden, P. von Ballmoos, L. Bouchet, P. Paul, J. Matteson, S. Boggs, C. Wunderer, P. Leleux, G. Weidenspointner, P. Durouchoux, R. Diehl, A. Strong, M. Cassé, M. A. Clair and Y. André, SPI: The spectrometer aboard INTEGRAL, *Astronomy and Astrophysics* **411**, L63 (November 2003).
9. R. Diehl, T. Siegert, W. Hillebrandt, S. A. Grebenev, J. Greiner, M. Krause, M. Kromer, K. Maeda, F. Röpke and S. Taubenberger, Early ^{56}Ni decay gamma rays from SN2014J suggest an unusual explosion, *Science* **345**, 1162 (September 2014).
10. F. Frontera, E. Virgilli, C. Guidorzi, P. Rosati, R. Diehl, T. Siegert, C. Fryer, L. Amati, N. Auricchio, R. Campana, E. Caroli, F. Fuschino, C. Labanti, M. Orlandini, E. Pian, J. B. Stephen, S. Del Sordo, C. Budtz-Jorgensen, I. Kuvvetli, S. Brandt, R. M. C. da Silva, P. Laurent, E. Bozzo, P. Mazzali and M. D. Valle, Understanding the origin of the positron annihilation line and the physics of supernova explosions, *Experimental Astronomy* **51**, 1175 (June 2021).
11. C. Guidorzi, F. Frontera, G. Ghirlanda, G. Stratta, C. G. Mundell, E. Virgilli, P. Rosati, E. Caroli, L. Amati, E. Pian, S. Kobayashi, G. Ghisellini, C. Fryer, M. D. Valle, R. Margutti, M. Marongiu, R. Martone, R. Campana, F. Fuschino, C. Labanti, M. Orlandini, J. B. Stephen, S. Brandt, R. C. d. Silva, P. Laurent, R. Mochkovitch, E. Bozzo, R. Ciolfi, L. Burderi and T. Di Salvo, A deep study of the high-energy transient sky, *Experimental Astronomy* **51**, 1203 (June 2021).
12. B. P. Abbott, LIGO Scientific Collaboration and Virgo Collaboration, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *prl* **119**, p. 161101 (October 2017).
13. A. von Kienlin, C. Meegan and A. Goldstein, GRB 170817A: Fermi GBM detection., *GRB Coordinates Network* **21520**, p. 1 (January 2017).
14. V. Connaughton, GBM-LIGO Group, L. Blackburn, M. S. Briggs, J. Broida, E. Burns, J. Camp, T. Dal Canton, N. Christensen, A. Goldstein, R. Hamburg, C. M. Hui, P. Jenke, D. Kocevski, N. Leroy, T. Littenberg, J. McEnery, R. Preece, J. Racusin, P. Shawhan, K. Siellez, L. Singer, J. Veitch, P. Veres and C. Wilson-Hodge, LIGO/Virgo G298048: Fermi GBM trigger 170817.529 and LIGO single IFO trigger, *GRB Coordinates Network* **21506**, p. 1 (August 2017).
15. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A, *ApJ Letter* **848**, p. L14 (October 2017).
16. V. Savchenko, C. Ferrigno, E. Kuulkers, A. Bazzano, E. Bozzo, S. Brandt, J. Chenevez, T. J. L. Courvoisier, R. Diehl, A. Domingo, L. Hanlon, E. Jourdain, A. von Kienlin, P. Laurent, F. Lebrun, A. Lutovinov, A. Martin-Carrillo, S. Mereghetti, L. Natalucci, J. Rodi, J. P. Roques, R. Sunyaev and P. Ubertini, INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817, *ApJ Letter* **848**, p. L15 (October 2017).
17. M. L. McConnell, High energy polarimetry of prompt GRB emission, *New Astronomy Review* **76**, 1 (February 2017).

18. R. Gill, M. Kole and J. Granot, GRB Polarization: A Unique Probe of GRB Physics, *Galaxies* **9**, p. 82 (October 2021).
19. H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E. O. Angüner, M. Arakawa, C. Arcaro, C. Armand, H. Ashkar, M. Backes, V. Barbosa Martins, M. Barnard, Y. Becherini, D. Berge, K. Bernlöhr, E. Bissaldi, R. Blackwell, M. Böttcher, C. Boisson, J. Bolmont, S. Bonnefoy, J. Bregeon, M. Breuhaus, F. Brun, P. Brun, M. Bryan, M. Büchele, T. Bulik, T. Bylund, M. Capasso, S. Caroff, A. Carosi, S. Casanova, M. Cerruti, T. Chand, S. Chandra, A. Chen, S. Colafrancesco, M. Curyło, I. D. Davids, C. Deil, J. Devin, P. deWilt, L. Dirson, A. Djannati-Ataï, A. Dmytriiev, A. Donath, V. Doroshenko, J. Dyks, K. Egberts, G. Emery, J. P. Ernenwein, S. Eschbach, K. Feijen, S. Fegan, A. Fiasson, G. Fontaine, S. Funk, M. Füßling, S. Gabici, Y. A. Gallant, F. Gaté, G. Giavitto, L. Giunti, D. Glawion, J. F. Glicenstein, D. Gottschall, M. H. Grondin, J. Hahn, M. Haupt, G. Heinzelmann, G. Henri, G. Hermann, J. A. Hinton, W. Hofmann, C. Hoischen, T. L. Holch, M. Holler, D. Horns, D. Huber, H. Iwasaki, M. Jamrozy, D. Jankowsky, F. Jankowsky, A. Jardin-Blicq, I. Jung-Richardt, M. A. Kastendieck, K. Katarzyński, M. Katsuragawa, U. Katz, D. Khangulyan, B. Khélifi, J. King, S. Klepser, W. Kluźniak, N. Komin, K. Kosack, D. Kostunin, M. Kreter, G. Lamanna, A. Lemière, M. Lemoine-Goumard, J. P. Lenain, E. Leser, C. Levy, T. Lohse, I. Lypova, J. Mackey, J. Majumdar, D. Malyshev, V. Marandon, A. Marcowith, A. Mares, C. Mariaud, G. Martí-Devesa, R. Marx, G. Maurin, P. J. Meintjes, A. M. W. Mitchell, R. Moderski, M. Mohamed, L. Mohrmann, C. Moore, E. Moulin, J. Muller, T. Murach, S. Nakashima, M. de Naurois, H. Ndiyavala, F. Niederwanger, J. Niemiec, L. Oakes, P. O'Brien, H. Odaka, S. Ohm, E. de Ona Wilhelmi, M. Ostrowski, I. Oya, M. Panter, R. D. Parsons, C. Perennes, P. O. Petrucci, B. Peyaud, Q. Piel, S. Pita, V. Poireau, A. Priyana Noel, D. A. Prokhorov, H. Prokoph, G. Pühlhofer, M. Punch, A. Quirrenbach, S. Raab, R. Rauth, A. Reimer, O. Reimer, Q. Remy, M. Renaud, F. Rieger, L. Rinchiuso, C. Romoli, G. Rowell, B. Rudak, E. Ruiz-Velasco, V. Sahakian, S. Sailer, S. Saito, D. A. Sanchez, A. Santangelo, M. Sasaki, R. Schlickeiser, F. Schüssler, A. Schulz, H. M. Schutte, U. Schwanke, S. Schwemmer, M. Seglar-Arroyo, M. Senniappan, A. S. Seyffert, N. Shafi, K. Shiningayamwe, R. Simoni, A. Sinha, H. Sol, A. Specovius, M. Spir-Jacob, Ł. Stawarz, R. Steenkamp, C. Stegmann, C. Steppa, T. Takahashi, T. Tavernier, A. M. Taylor, R. Terrier, D. Tiziani, M. Tluczykont, C. Trichard, M. Tsirou, N. Tsuji, R. Tuffs, Y. Uchiyama, D. J. van der Walt, C. van Eldik, C. van Rensburg, B. van Soelen, G. Vasileiadis, J. Veh, C. Venter, P. Vincent, J. Vink, H. J. Völk, T. Vuillaume, Z. Wadiasingh, S. J. Wagner, R. White, A. Wierzcholska, R. Yang, H. Yoneda, M. Zacharias, R. Zanin, A. A. Zdziarski, A. Zech, A. Ziegler, J. Zorn, N. Żywucka, F. de Palma, M. Axelsson and O. J. Roberts, A very-high-energy component deep in the gamma-ray burst afterglow, *Nature* **575**, 464 (November 2019).
20. MAGIC Collaboration, V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González, W. Bednarek, L. Bellizzi, E. Bernardini, A. Berti, J. Besenrieder, W. Bhattacharyya, C. Bigongiari, A. Biland, O. Blanch, G. Bonnoli, Ž. Bošnjak, G. Busetto, A. Carosi, R. Carosi, G. Ceribella, Y. Chai, A. Chilingaryan, S. Cikota, S. M. Colak, U. Colin, E. Colombo, J. L. Contreras, J. Cortina, S. Covino, G. D'Amico, V. D'Elia, P. da Vela, F. Dazzi, A. de Angelis, B. de Lotto, M. Delfino, J. Delgado, D. Depaoli, F. di Pierro, L. di Venere, E. Do Souto Espiñeira, D. Dominis Prester, A. Donini, D. Dorner, M. Doro, D. Elsaesser, V. Fallah Ramazani, A. Fattorini, A. Fernández-Barral, G. Ferrara, D. Fidalgo, L. Foffano, M. V. Fonseca, L. Font, C. Fruck, S. Fukami, S. Gallozzi, R. J. García López, M. Garczarczyk, S. Gasparyan, M. Gaug, N. Giglietto,

F. Giordano, N. Godinović, D. Green, D. Guberman, D. Hadasch, A. Hahn, J. Herrera, J. Hoang, D. Hrupec, M. Hütten, T. Inada, S. Inoue, K. Ishio, Y. Iwamura, L. Jouvin, D. Kerszberg, H. Kubo, J. Kushida, A. Lamastra, D. Lelas, F. Leone, E. Lindfors, S. Lombardi, F. Longo, M. López, R. López-Coto, A. López-Oramas, S. Loporchio, B. Machado de Oliveira Fraga, C. Maggio, P. Majumdar, M. Makariev, M. Mallamaci, G. Maneva, M. Manganaro, K. Mannheim, L. Maraschi, M. Mariotti, M. Martínez, S. Masuda, D. Mazin, S. Mićanović, D. Miceli, M. Minev, J. M. Miranda, R. Mirzoyan, E. Molina, A. Moralejo, D. Morcuende, V. Moreno, E. Moretti, P. Munar-Adrover, V. Neustroev, C. Nigro, K. Nilsson, D. Ninci, K. Nishijima, K. Noda, L. Nogués, M. Nöthe, S. Nozaki, S. Paiano, J. Palacio, M. Palatiello, D. Paneque, R. Paoletti, J. M. Paredes, P. Peñil, M. Peresano, M. Persic, P. G. Prada Moroni, E. Prandini, I. Puljak, W. Rhode, M. Ribó, J. Rico, C. Righi, A. Rugliancich, L. Saha, N. Sahakyan, T. Saito, S. Sakurai, K. Satalecka, K. Schmidt, T. Schweizer, J. Sitarek, I. Šnidarić, D. Sobczynska, A. Somero, A. Stamerra, D. Strom, M. Strzys, Y. Suda, T. Surić, M. Takahashi, F. Tavecchio, P. Temnikov, T. Terzić, M. Teshima, N. Torres-Albà, L. Tosti, S. Tsujimoto, V. Vagelli, J. van Scherpenberg, G. Vanzo, M. Vazquez Acosta, C. F. Vigorito, V. Vitale, I. Vovk, M. Will, D. Zarić and L. Nava, Teraelectronvolt emission from the gamma-ray burst GRB 190114C, *Nature* **575**, 455 (November 2019).

21. E. Maiorano, N. Masetti, E. Palazzi, F. Frontera, P. Grandi and et al., The puzzling case of GRB 990123: multiwavelength afterglow study, *Astronomy and Astrophysics* **438**, 821 (August 2005).
22. C. Kouveliotou, J. Granot, J. L. Racusin, E. Bellm, G. Vianello, S. Oates, C. L. Fryer, S. E. Boggs, F. E. Christensen, W. W. Craig, C. D. Dermer, N. Gehrels, C. J. Hailey, F. A. Harrison, A. Melandri, J. E. McEnery, C. G. Mundell, D. K. Stern, G. Tagliaferri and W. W. Zhang, NuSTAR Observations of GRB 130427A Establish a Single Component Synchrotron Afterglow Origin for the Late Optical to Multi-GeV Emission, *ApJ Letter* **779**, p. L1 (December 2013).
23. A. Martin-Carrillo, L. Hanlon, M. Topinka and et al., GRB 120711A: An intense INTEGRAL burst with long-lasting soft γ-ray emission and a powerful optical flash, *Astronomy and Astrophysics* **567**, p. A84 (Jul 2014).
24. I. Johnson, W. N. and R. C. Haymes, Detection of a Gamma-Ray Spectral Line from the Galactic-Center Region, *Astrophysical Journal* **184**, 103 (August 1973).
25. J. Knödlseder and GRI Consortium, GRI: The Gamma-Ray Imager Mission, in *The Obscured Universe. Proceedings of the VI INTEGRAL Workshop (2007)*,
26. E. Kalemci, S. E. Boggs, P. A. Milne and S. P. Reynolds, Searching for Annihilation Radiation from SN 1006 with SPI on INTEGRAL, *ApJ Letter* **640**, L55 (March 2006).
27. G. Bertone, A. Kusenko, S. Palomares-Ruiz, S. Pascoli and D. Semikoz, Gamma-ray bursts and the origin of galactic positrons, *Physics Letters B* **636**, 20 (April 2006).
28. N. Guessoum, P. Jean and N. Prantzos, Microquasars as sources of positron annihilation radiation, *Astronomy and Astrophysics* **457**, 753 (October 2006).
29. N. Prantzos, Astrophysical gamma-ray lines: A probe of stellar nucleosynthesis and star formation, in *5th INTEGRAL Workshop on the INTEGRAL Universe (2004)*,
30. R. Bartels, F. Calore, E. Storm and C. Weniger, Galactic binaries can explain the Fermi Galactic centre excess and 511 keV emission, *MNRAS* **480**, 3826 (November 2018).
31. T. Siegert, R. Diehl, J. Greiner, M. G. H. Krause, A. M. Beloborodov, M. C. Bel, F. Guglielmetti, J. Rodriguez, A. W. Strong and X. Zhang, Positron annihilation signatures associated with the outburst of the microquasar V404 Cygni, *Nature* **531**, 341 (March 2016).

32. L. Bouchet, J. P. Roques and E. Jourdain, On the Morphology of the Electron-Positron Annihilation Emission as Seen by Spi/integral, *Astrophysical Journal* **720**, 1772 (Sep 2010).
33. J. Isern, P. Jean, E. Bravo, J. Knödlseder and et al., Gamma-ray emission from SN2014J near maximum optical light, *Astrophysical Journal* **588**, p. A67 (Apr 2016).
34. M. Moita, M. Ferro, U. Frontera, O. Caroli, R. Virgilli, P. Stephen, F. Curado da Silva, P. Maia and A. Del Sordo, "Polarimetric prospects of a new hard X-/soft gamma-ray space mission for next decades", these proceedings.

Polarimetric prospects of a new hard X-soft gamma-ray space mission for next decades

M. Moita*, L. Ferro and F. Frontera
Dept. of Physics and Earth Science, University of Ferrara, Ferrara, Italy
**E-mail: mmoita@fe.infn.it*

E. Caroli, E. Virgilli and J. B. Stephen
OAS of Bologna INAF Bologna, Bologna, Italy

R. M. Curado da Silva and M. Maia
LIP-Laboratório de Instrumentação e Física Experimental de Partículas, Coimbra, Portugal

S. del Sordo
IASF-Palermo INAF Palermo, Palermo, Italy

The measurement of the polarization of the high-energy photons from cosmic sources has now become a key observational parameter for understanding the emission mechanisms and the geometry of the active regions involved. Therefore, a mandatory requirement for new instrumentation in this energy regime will provide high sensitivity for polarimetric measurements associated with spectroscopy and imaging.

In this perspective, the Advanced Surveyor of Transient Events and Nuclear Astrophysics (ASTENA) mission, which includes two main instruments: the Wide field monitor (WFM-IS), with a large effective area and a wide energy passband (2 keV – 20 MeV); and the Narrow Field Telescope (NFT), with a broad energy passband (50–600 keV) with focusing capabilities based on the use of an advanced Laue lens; will both provide high sensitivity for polarimetric measurements. Furthermore, both instruments will include spectometers with a good 3D spatial resolution allowing to perform 3D Compton polarimetry, increasing the possibilities to optimize the event selection. Herein, we report on the results of a Monte Carlo study devoted to optimize the configuration of both instruments, in particular, the modulation factor (Q), the events detection efficiency (Eff) and the Minimum Detectable Polarization (MDP).

Keywords: Polarimetry, Gamma-ray detectors, Space instrumentation, Compton scattering.

1. Introduction

Polarimetry of high-energy sources is a key observational parameter for a better understanding of the emission mechanisms and geometry of a wide number of cosmic objects and events such as GRBs, pulsars, binary black holes and active galactic nuclei, whose emissions are expected to be polarized. The field of hard

X-ray polarimetry has been quite active in the last two decades: missions such as INTEGRAL, AstroSat and POLAR have provided some important scientific inputs in understanding emission mechanism and geometry in X-ray pulsars,[1–3] magnetic field structure in Pulsar Wind Nebula (PWN),[1,4] disk-jet interplay in black hole X-ray binaries (XRBs)[5,6] and hard X-ray emission mechanism in GRB prompt emission.[7–9] Though these results are marred with large uncertainties and a firm conclusion is not possible, most of these findings are extremely interesting as they pose new challenges to the existing theories. One interesting and common feature in all these measurements is a systematic increase in polarization within the hard X-ray band which makes this energy regime extremely promising for polarimetry experiments.[10]

As a part of the European project AHEAD (integrated Activities in the High Energy Astrophysics Domain), a new concept of high energy mission named ASTENA (Advanced Surveyor of Transient Events and Nuclear Astrophysics) has been proposed and accepted for feasibility and performance study.[11] In the current AHEAD 2020 project, which is a continuation of AHEAD, we deepen the feasibility study of ASTENA and increase our knowledge on the potential performances of the instruments on board concerning spectral, imaging and, particularly in this context, polarization capabilities. In this work we present the results of a study that aims to reliably and realistically evaluate the performance in terms of polarimetry that can be achieved with the ASTENA and to optimize the design of the detection units to achieve this aim taking into account the requirements.

2. ASTENA

The ASTENA configuration is described in,[12] and it is shown in Fig. 1. Its main properties are summarized in Table 1. The instrumentation on board consists of a Wide Field Monitor–Imaging Spectrometer (WFM-IS) with a 2 keV–20 MeV passband, and a Narrow Field Telescope (NFT) with a 50–600 keV passband.

Table 1. Summary of the main properties of the WFM-IS and of the NFT on board ASTENA.

	WFM-IS	NFT
Energy pass-band	2 keV – 20 MeV	50 – 600 keV
Total useful area[a]	$\sim$ 5800 cm^2 ($<$ 30 keV) $\sim$ 6700 cm^2 (30–150 keV) $\sim$ 13800 cm^2 (>200 keV)	7 m^2 (projected)
Field of View	2 sr	4 arcmin
Angular resolution	6 arcmin	$\sim$ 30 arcsec HPD
Point source localization accuracy	1 arcmin	$<$ 10 arcsec
3σ, **Continuum Sensitivity** (ph/cm^2 s keV)	1×10^{-4} (10 s, 30–150 keV)	1.2×10^{-9} (10^5 s, 225–375 keV)
3σ, 10^5 s **Line sensitivity** (ph/cm^2 s)	-	5×10^{-7} (178 keV)

Note: [a]Total geometric area through the mask or collimator.

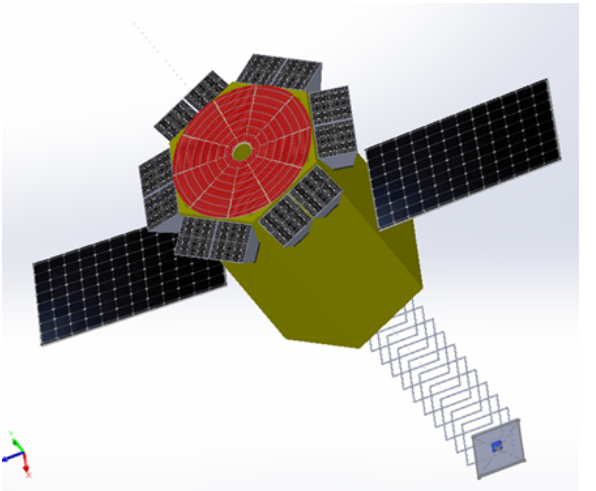


Fig. 1. The ASTENA satellite in inflight configuration. The WFM/S modules surround the NFT Laue lens (front), while its focal plane is positioned 20 m apart by a booming structure (back).

The NFT is a Laue lens focusing telescope with a 3 m diameter and 20 m focal length, composed by a large number of crystal tiles in transmission configuration, that are disposed in such a way that they concentrate the incident radiation onto a common focal spot.[13] The lens proposed for the NFT is made of ∼19500 bent crystal tiles of Si(111) and Ge(111), with 40 m curvature radius (within 5% uncertainty). The crystal tiles have a 30×10 mm^2 cross section, and an optimized thickness vs. energy for maximizing the diffraction efficiency, with the condition of a maximum thickness of 5 mm. The crystals are disposed on 43 rings, with an outermost ring radius of 149 cm and an innermost of 18 cm, ensuring an energy passband of 50 keV (outermost rings) to 600 keV (innermost rings). Current Monte Carlo simulations, ray-trace analyses and laboratory experiments evaluated that an angular resolution of 0.5 arcmin can be achieved, with a useful FOV of approximately 4 arcmin.[14]

Concerning the NFT focal plane detector, the adopted configuration consists of 4 layers of drift strip detectors of CdZnTe,[15] each layer with a cross section a 80×80 mm^2 and a thickness of 20 mm. This configuration will ensure a detection efficiency higher than 80% in the entire energy band of the NFT, a spectroscopy response of 1% @ 511 keV and fine spatial resolution (0.2–0.3 mm) in three dimensions,[16] allowing the correct sampling of the Point Spread Function (PSF) of the lens, that is of the order of a few mm^2. Thanks to the unprecedented sensitivity that can be achieved with the Laue lens focusing technique and the high segmentation level the focal plane detector, NFT is expected to provide fine polarization measurements.

The WFM-IS consists of an array of 12 units, two units on each side of the hexagon surrounding the NFT, accommodated as shown in Fig. 1. All the units are

offset by 15 deg with respect to the axis of the NFT. Each WFM-IS unit consists of a 40×40 array of detection elements, each made of a scintillator bar (CsI(TI) or similar) 5 cm long, with an hexagonal cross section of about 75 mm^2, viewed by two Silicon Drift Detectors (SDD) 0.4 mm thick, one on the top (side of entrance of the celestial photons) and other on the bottom. The functioning principle of this detector is similar to that adopted for the X-Gamma-ray Imaging Spectrometer (XGIS) aboard the THESEUS mission concept.[17] It has the great advantage of a very broad passband (2 keV–20 MeV), a 3-D position sensitivity to energy losses in the scintillator bars and a very low intrinsic background. The hexagonal cross section of the bars, which allows us to obtain a low geometric systematic effect in the polarization measurements, associated with the 3D position sensitivity of the detector makes the WFM-IS suitable for fine polarization measurements.[11]

3. Compton Polarimetry with ASTENA

In the hard X- and soft γ-rays regime the measurement of the polarization status of a photon beam rely on the Compton scattering mechanism. The detectors on-board ASTENA provide the segmentation needed to offer an efficient method to measure the linear polarisation of incoming photons using Compton scattering because, with a coincidence event logic, each element/pixel can act at the same time both as a scattering element and as a detection unit allowing to record the pairs of events that provide polarimetric information: one is the Compton scattering of the incoming photon, the other is the interaction by Compton or photoelectric effect of the secondary photon.[10]

The cross-section of a hard X-ray photon to interact with a free electron is given by the Compton scattering cross-section described by the Klein–Nishina equation[18]:

$$\frac{d\sigma}{d\Omega} = \frac{r_0^2 \varepsilon^2}{2} \left(\frac{1}{\varepsilon} + \varepsilon - 2 \sin^2 \theta \cos^2 \varphi \right) \tag{1}$$

where r_0^2 is the classic electron radius, $\varepsilon = E'/E$ the ratio between the energy of the scattered photon and the energy of the incoming photon, θ is the scattering angle of the scattered photon measured from the direction of the incident photon, and φ is the azimuth scattering angle with respect to the electric vector of the incident photon. On a segmented detector the distribution of the azimuthal angle measured on a large sample of data has a sinusoidal behavior with period π

$$N(\phi) = a_0(1 + a_1 \cos(2(\phi - a_2 + \pi/2))) \tag{2}$$

where a_0 is the normalization constant, $Q = a_1$ the modulation factor, and $P = a_2$ the polarization angle which are obtained from a fit to the data. This distribution is the called "modulation curve", which is built dividing the scattering map on a fixed number of angular sectors, covering 360°, and integrating the counts inside each

angular sector. The aperture of the angular sectors should be chosen taking into account the discretization level of the detector (i.e., the pixel dimensions) and the counts statistics. Because of the segmented nature of the detector, the integration of count inside a defined angular sector should take into account effects due to pixels that lie across their borders. Methods to account for these pixels are, for example, that of distributing uniformly the pixel counts proportionally to the area of the pixel intercepted by two adjacent angular sectors. For the analyses performed the number of bins of the azimuthal distribution was optimized for each instrument since it affects the shape of the modulation curve.

In the characterization of a polarimeter it is important to describe its response not only to polarized, but also to unpolarized signals. For a perfectly uniform and symmetric instrument the modulation curve from a 0% polarized signal should be totally flat. However, the detector geometry introduces an asymmetry that is important to correct the signal from this component. We correct the measured modulation curve from the instrument response to a unpolarized signal by applying the following expression for each angle bin:

$$N_{corr}(\phi) = \frac{N_p(\phi)}{N_{np}(\phi)} N_{np}^{max} \tag{3}$$

where $N_{corr}(\phi)$ is the corrected azimuthal distribution, $N_p(\phi)$ is the polarized azimuthal distribution, $N_{np}(\phi)$ the non polarized modulation azimuthal distribution, and N_{np}^{max} the maximum of the non polarized azimuthal distribution. In practice, the systematic uncertainty in the polarization measurement will be estimated from the results of an extensive on-ground calibration campaign of the instruments, using both polarized and unpolarized monochromatic beams of gamma rays.

The minimum detectable polarization (MDP) of an instrument for a given celestial source type is an essential parameter to estimate its polarimetric potential. For a space polarimeter in a background dominated environment, the MDP with a 99% confidence level, is given by[19]:

$$MDP = \frac{4.29}{A.\epsilon.S_F.Q_{100}} \sqrt{\frac{A.\epsilon.S_f + B}{T}} \tag{4}$$

where, A is the effective area, ϵ is the events' efficiency, S_F is the source flux over the selected energy band (photons cm^{-2} s^{-1}), B is the background noise count rate (counts/s), T is the observation time (s) and Q_{100} is the polarimetric modulation factor for a 100% polarized source.

High polarimetric performance for hard X/soft γ-rays can be achieved fulfilling the following requirements: high scattering efficiency, fine spatial resolution, and fine spectroscopy. These characteristics allow a large flexibility in scattered events selection, and therefore several event filter can be implemented to optimize the modulation factor and the event efficiency, and consequently improve the MDP.

4. Polarimetric Analyses for a Monochromatic Source

4.1. *NFT*

In order to fulfil the requirements on the ASTENA NFT, and following the knowhow acquired by both experiments on CdTe/CZT detectors used as scattering polarimeter and development on 3D CZT spectrometers,[20] a simple model of the NFT focal plane has been implemented using MEGALib tools.[21] Using the MEGAlib/Geomega package, we implement a numerical model made of a stack of 4 thick layers of CZT separated by a layer (10 mm thick) of a PCB type mixture passive material, representative of the read-out front-end electronics board (Fig. 2). Each CZT layer is 20 mm thick and 80 mm × 80 mm wide. The CZT layers were divided in voxels with dimensions 0.25 x 0.25 x 0.25 mm^3 each. The spectroscopic response of the detector is simply modeled by a relationship derived from measurements on CZT drift strips spectrometers.[22] In all the performed simulation runs, we set the low energy threshold to 5 keV, that is a reasonable value at room temperature as demonstrated by the same experimental results.

The Laue lens configuration will impinge an almost monochromatic beam for each crystal ring on the detector, each one having a small inclination angle with respect to the axis of the lens, ranging from 0.4° (at the energy of 600 keV, corresponding to the inner radius) to 4.5° (at the energy of 50 keV, corresponding to the external radius). This will produce rings of photons with similar energy hitting the detector inside a PSF. We know from previous experiments that an off-axis beam introduce geometric effects on the detector that affect the modulation factor.[23] However, since the Laue lens presents a 180° symmetry, for each beam there will be a symmetric beam with opposite off-axis angle. This configuration will correct any possible systematic error due to the off-axis beam and the results will be similar as we impinge an on-axis beam on the detector. In fact, a systematic study was performed in order to validate this assumption. We evaluated the modulation factor by varying the number of pair of symmetric off-axis beams from 1 to 128, obtaining a distribution of values for the modulation factor with a standard deviation of 0.02 from the central value which is comparable with the error on the values of the modulation factor. To simulate a full Laue lens would require ∼20000 beams, each with its off-axis angle, however, by taking into account the aforementioned results, we decided to use instead a Gaussian cone shape beam, 20 m from the detector, which produce a Gaussian shape PSF with a FWHM of 0.15 cm on the top of the detector. This will simplify the simulation code and speed up the simulations without compromising the result.

To ensure the quality of the selected events we perform a check by Compton kinematics – by evaluating the energy of the two first interactions and the corresponding scattering angle we ensure that it corresponds to the expected values and the event can be selected as a "good event". This also can be done for multiple interactions, by using the energy of the first interaction, the sum of the following interactions

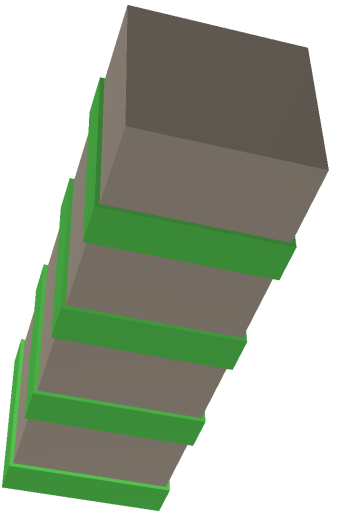


Fig. 2. Monte Carlo model of the the NFT focal plane. The gray volumes represent the segmented CZT spectrometer, while the green ones represent the readout electronics boards.

and the scattering angle from the first and second interaction. This method ensures that all the events selected are true Compton events and have polarimetric information on it. This method also removes the double events that occur in neighbour pixels. These type of events can introduce anomalies because the scattering angle is difficult to know. Since we consider the interaction in the center of pixel these types of events will be considered as a $\sim 90^{\circ}$ scattering event (high polarimetric information). However, if the interactions occur near the edges of the pixels the scattering angle can be $\sim 0^{\circ}$ (low polarimetric information).

Due to the focusing properties of the NFT we know that the first interaction is always inside the PSF. These allows a further correction since we only select the events which complies with this condition. This will reduce the background or any spurious events that occur in the detector.

After the selections, the scattering maps and modulation curves are drawn. To correct the systematic effects introduced by the detector geometry we also simulate an unpolarized source in the same conditions as the polarized source and we corrected the distributions with equation 3. As an example Fig. 3 presents the scattering maps obtained for a 300 keV unpolarized beam and for a 100% polarized beam with a polarization vector direction of 60°. By diving the scattering maps into 24 bins we achieved the azimuthal distributions presented in Fig. 3, below each scattering map. The corrected modulation curve is presented in Fig. 4. The corrected curve is then fitted with equation 2 to get the modulation factor and the polarization angle.

In Fig. 5 it is presented the modulation factor as function of the energy for the NFT focal plane for monochromatic beams ranging from 100 to 600 keV. These

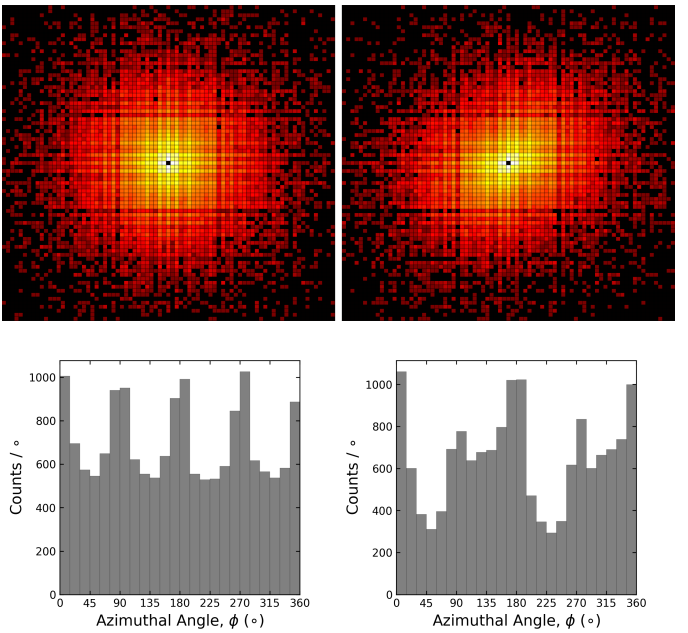


Fig. 3. Scattering maps and azimuthal distributions obtained for an unpolarized (left) and 100% polarized (right) beam. The beam energy is 300 keV and the polarization vector direction is 60° for the 100% polarized case.

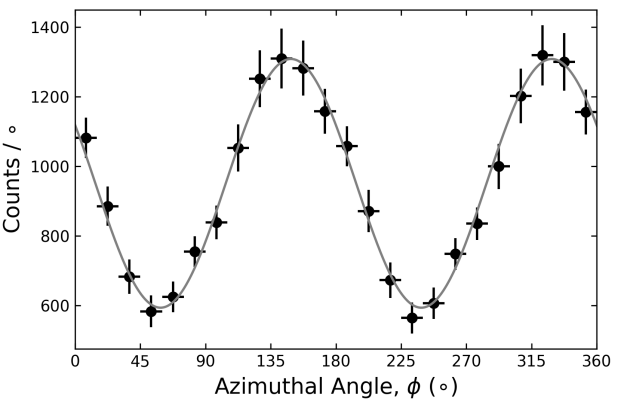


Fig. 4. Corrected modulation curve obtained when we apply the correction in Eq. (3) on the histograms of Fig. 3.

results showed that the NFT Focal Plane Detector (FPD) has potential to perform polarimetric analysis in the whole energy range presenting modulation factors >0.2. As expected, the best modulation factor is obtained at energies ∼300 keV.

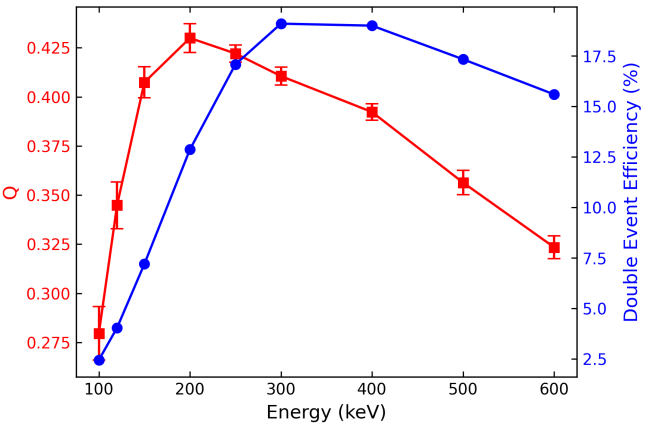


Fig. 5. NFT modulation factor and event efficiency as function of the energy.

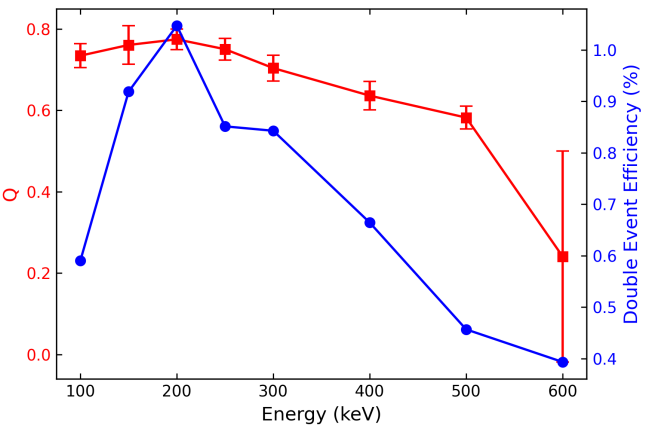


Fig. 6. NFT modulation factor and event efficiency as function of the energy for events with scattering angles of $\sim 90°$.

The percentage of selected events shows its maximum at $\sim$200 keV where the Compton cross section prevails over the other interactions in the CZT.

Due to the highly segmented detector we can also select events that maximize the modulation factor like events with scattering angles $\sim 90°$ (along the detector axis). By making a selection in the Z coordinates we ensure that all the events have $\sim 90°$. In Fig. 6 it is shown the modulation factor as a function of the energy taking account this selection. As can be seen, there is a big increase of the modulation factor when compared with the results without selection. The modulation factor reaches values near $\sim$0.8 for energies $\sim$200 keV. However the efficiency is very affected and reduced to values below $\sim$1%.

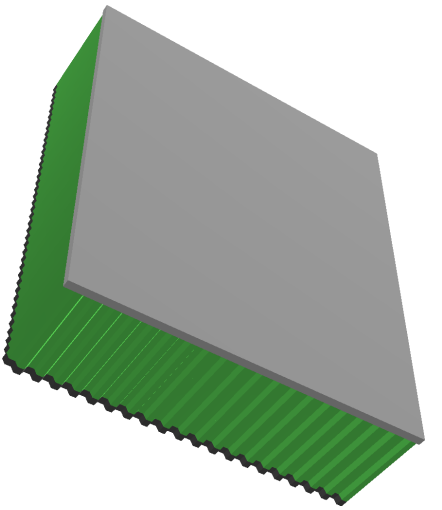


Fig. 7. A geometry sketch of the the WFM-IS unit. The green volume represent the CsI scintillators, while the gray ones represent the Si SDDs, light gray on top and dark gray on bottom.

4.2. *WFM-IS*

The Monte Carlo model of one WFM-IS unit is presented in Fig. 7. It consists of an hexagonal 40x40 array of CsI scintillators (for a total of 205 bars), with a distance between contiguous centres across hexagon flat sides of 5 mm, and a length of 50 mm. To mimic the depth resolution of the detector we segmented the bars in 10 segments of 5 mm each. On the top the detector has a linear SDD, 0.4 mm thick, and, on bottom an hexagonal single anode SDD, 0.4 mm thick, with a distance between hexagon flat sides of 5 mm. To construct the full WFM-IS we copied 12 units and placed along an hexagonal configuration, 2 units in each hexagon side, with a distance from the center (lens axis) of 1.5 m.

Since the WFM-IS is a wide field instrument, the beam used for the simulations was a far point source. For the WFM-IS we do not know were the beam undergoes the first interaction, all the scintillating bars will be scattering and absorbing elements, so we selected the events that comply the Compton kinematics and we corrected the signal data with the unpolarized data. In Fig. 8 the scattering maps are shown for a 500 keV unpolarized and 100% polarized source.

A systematic study was performed using monochromatic sources from 125 keV to 5 MeV. Fig. 9 shows the modulation factor and efficiency obtained for one WFM-IS unit and the full WFM-IS. As can be seen, for energies below 1 MeV the modulation factor shows values >0.2, however for energies above 1 MeV the modulation factor drops drastically becoming impossible to get polarization results. The reason for this is that for energies > 1 MeV both the theoretic value of the modulation factor and the efficiency of the Compton process drop significantly.[10] When comparing

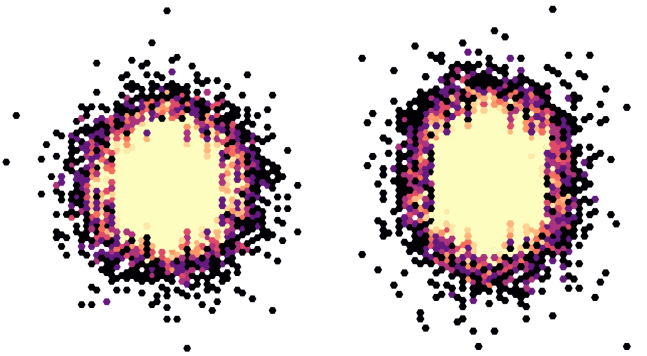


Fig. 8. Scattering maps obtained for a 500 keV unpolarized (left) and 100% polarized (right) beam with the WFM-IS unit. The polarization vector direction is 0° for the 100% polarized case.

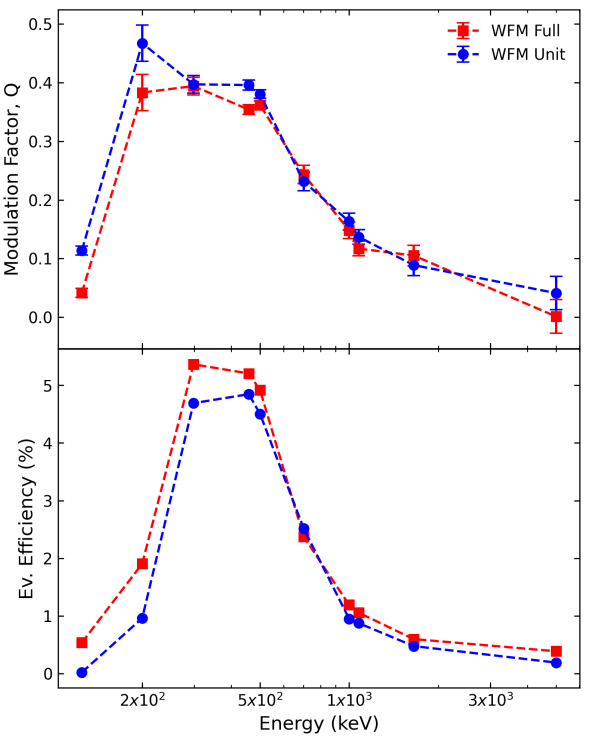


Fig. 9. Modulation factor (top) and event efficiency (bottom) as function of the energy for one WFM-IS unit and for the full WFM-IS.

the results for one WFM-IS unit and for the full WFM-IS the results of the modulation factor are similar, however the event efficiency increases when using the full WFM-IS.

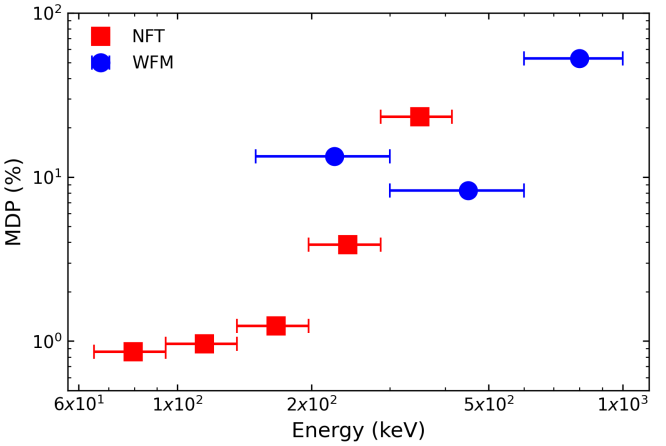


Fig. 10. Minimum detectable polarization (MDP) versus energy for a 100 mCrab source when observing with NFT and WFM-IS for 100 ks.

5. Polarimetric Sensitivity of ASTENA

In order to estimate the potential MDP of the ASTENA mission in a LEO (low Earth orbit) when observing a Crab type source, we simulated a source with a power law spectrum (normalization at 1 keV: 10.74 photons keV^{-1} s^{-2} cm^{-2}; photon index: 2.17) for each instrument individually and obtained the modulation factor and event efficiency. The simulated energy range was in accordance with the instrument energy passband: for the NFT 65 – 600 keV; and for the WFM-IS 150 keV – 1 MeV. As before, we simulated an on-axis Gaussian cone beam to mimic the 1.5 mm diameter of the Lens PSF for the NFT and a on-axis far source for the WFM-IS.

The background components were modeled in detail also using the MEGAlib environment tools for both NFT and WFM-IS using the works.[24–28] Background simulations were performed for both instruments individually and, to determine the real background for polarimetric analyses, we also performed the same selection process as the data. First we checked the events by Compton kinematics and, for the NFT data, we check if the first interaction is inside the PSF.

The effective area of the NFT was computed analytically as the product of the geometrical area of the instrument and the reflection efficiency of the lens[14] times the effective area for the WFM-IS was estimated using MEGAlib simulations.

The calculated MDP as a function of the energy when observing a 100 mCrab source for 100 ks is presented in Fig. 10. As can be seen, for energies < 200 keV the NFT presents MDP values $\sim$1%. However, above that energy it starts to substantially degrade because of the decrease of effective area of the Laue lens. For energies above 300 keV the WFM-IS presents a better MDP than the NFT meaning that the instruments can complement each other in energy bands.

6. Conclusions

In the next missions for medium energy astrophysics, polarimetry will be a fundamental requirement for the solution of important scientific problems related to the nature and geometry of the high energy cosmic sources emission. The ASTENA mission, under study as part of the European project AHEAD 2020, thanks to the high efficient, fine spectral resolution and high segmentation of the detectors onboard will be able to make polarimetry a standard way of observation by definitively opening this observation window also in the field of hard X- and soft γ-rays.

In this work we presented the results of an ongoing study which aims at a reliable and realistic evaluation of the performance in terms of polarimetry achievable with ASTENA, and obtain information for optimizing the design of the detection systems. The results show that the NFT will be able to achieve unprecedented polarametric capabilities for energies between 60 - 300 keV, while the WFM-IS will be capable to provide polarimetric measurements up to 1 MeV. ASTENA polarimetric performances show that such instruments hold the potential to address gamma-ray polarimetry with high sensitivity and contribute to expanding this new astrophysics knowledge window.

Acknowledgements

This work has been partially supported with the financial contribution from the ASI-INAF agreement n. 2017-14-H.0 “Studi per future missioni scientifiche”, the ASI-INAF agreement n. 2018-10-H.1-2020 “HERMES Technologic Pathfinder”, and the AHEAD EU Horizon 2020 project (Integrated Activities in the High Energy Astrophysics Domain), grant agreement n. 871158.

References

1. S. V. Vadawale, Chattopadhyay *et al.*, Phase-resolved X-ray polarimetry of the Crab pulsar with the AstroSat CZT Imager, *Nature Astronomy* **2**, p. 50–55 (2018).
2. M. Forot, P. Laurent, I. A. Grenier, C. Gouiffès and F. Lebrun, Polarization of the crab pulsar and nebula as observed by the INTEGRAL/IBIS telescope, *The Astrophysical Journal* **688**, L29 (nov 2008).
3. E. Jourdain and J.-P. Roques, 2003–2018 monitoring of the crab nebula polarization in hard x-rays with INTEGRAL SPI, *The Astrophysical Journal* **882**, p. 129 (sep 2019).
4. A. J. Dean, D. J. Clark, J. B. Stephen, V. A. McBride, L. Bassani, A. Bazzano, A. J. Bird, A. B. Hill, S. E. Shaw and P. Ubertini, Polarized gamma-ray emission from the crab, *Science* **321**, 1183 (2008).
5. P. Laurent, J. Rodriguez, J. Wilms, M. Cadolle Bel, K. Pottschmidt and V. Grinberg, Polarized gamma-ray emission from the galactic black hole cygnus x-1, *Science* **332**, p. 438–439 (Mar 2011).
6. E. Jourdain, J. P. Roques, M. Chauvin and D. J. Clark, Separation of two contributions to the high energy emission of cygnus x-1: Polarization measurements with INTEGRAL SPI, *The Astrophysical Journal* **761**, p. 27 (nov 2012).

7. S.-N. Zhang, M. Kole *et al.*, Detailed polarization measurements of the prompt emission of five gamma-ray bursts, *Nature Astronomy* **3**, 258 (Mar 2019).
8. T. Chattopadhyay, S. V. Vadawale, E. Aarthy, N. P. S. Mithun, V. Chand, A. Ratheesh, R. Basak, A. R. Rao, V. Bhalerao, S. Mate, A. B., V. Sharma and D. Bhattacharya, Prompt emission polarimetry of gamma-ray bursts with the AstroSat CZT imager, *The Astrophysical Journal* **884**, p. 123 (Oct 2019).
9. D. Götz, P. Laurent, F. Lebrun, F. Daigne and Ž. Bošnjak, Variable polarization measured in the prompt emission of grb 041219a using ibis on board INTEGRAL, *The Astrophysical Journal* **695**, L208 (Apr 2009).
10. T. Chattopadhyay, Hard x-ray polarimetry – an overview of the method, science drivers and recent findings (2021).
11. F. Frontera, "understanding the origin of the positron annihilation line and the physics of the supernova explosions".
12. F. Frontera, E. Virgilli, C. Guidorzi, P. Rosati, R. Diehl, T. Siegert, C. Fryer, L. Amati, N. Auricchio, R. Campana, E. Caroli, F. Fuschino, C. Labanti, M. Orlandini, E. Pian, J. B. Stephen, S. Del Sordo, C. Budtz-Jorgensen, I. Kuvvetli, S. Brandt, R. M. C. da Silva, P. Laurent, E. Bozzo, P. Mazzali and M. D. Valle, Understanding the origin of the positron annihilation line and the physics of supernova explosions, *Experimental Astronomy* **51**, 1175 (June 2021).
13. F. Frontera and P. von Ballmoos, Laue Gamma-Ray Lenses for Space Astrophysics: Status and Prospects, *X-Ray Optics and Instrumentation, 2010. Special Issue on X-Ray Focusing: Techniques and Applications, id.215375* **2010** (2010).
14. E. Virgilli, V. Valsan, F. Frontera, E. Caroli, V. Liccardo and J. B. Stephen, Expected performances of a Laue lens made with bent crystals, *Journal of Astronomical Telescopes, Instruments, and Systems* **3**, p. 044001 (October 2017).
15. I. Kuvvetli, C. Budtz-Jørgensen, E. Caroli and N. Auricchio, CZT drift strip detectors for high energy astrophysics, *Nuclear Instruments and Methods in Physics Research A* **624**, 486 (December 2010).
16. I. Kuvvetli, C. Budtz-Jørgensen, A. Zappettini, N. Zambelli, G. Benassi, E. Kalemci, E. Caroli, J. B. Stephen and N. Auricchio, A 3D CZT high resolution detector for x- and gamma-ray astronomy, in *High Energy, Optical, and Infrared Detectors for Astronomy VI*, eds. A. D. Holland and J. Beletic (SPIE, 2014).
17. R. Campana, F. Fuschino, C. Labanti, L. Amati, S. Mereghetti, M. Fiorini, F. Frontera, G. Baldazzi, P. Bellutti and et al., The X-Gamma Imaging Spectrometer (XGIS) onboard THESEUS, *Mem. Soc. Astron. Italiana* **89**, p. 137 (2018).
18. O. Klein and Y. Nishina, Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac, *Zeitschrift für Physik* **52**, 853 (1929).
19. M. C. Weisskopf, E. H. Silver, H. L. Kestenbaum, K. S. Long and R. Novick, A precision measurement of the X-ray polarization of the Crab Nebula without pulsar contamination., *The Astrophysical Journal* **220**, L117 (March 1978).
20. L. Abbene *et al.*, Recent advances in the development of high-resolution 3d cadmium–zinc–telluride drift strip detectors, *Journal of Synchrotron Radiation* **27**, 1564 (2020).
21. A. Zoglauer, Megalib: Medium energy gamma-ray astronomy library.
22. I. Kuvvetli, Development of cdznte detector systems for space applications, PhD thesis, Technical University of Denmark 2003.
23. R. M. C. da Silva *et al.*, Polarimetry study with a CdZnTe focal plane detector, *IEEE Transactions on Nuclear Science* **58**, 2118 (2011).

24. T. Mizuno, T. Kamae, G. Godfrey, T. Handa, D. J. Thompson, D. Lauben, Y. Fukazawa and M. Ozaki, Cosmic-ray background flux model based on a gamma-ray large area space telescope balloon flight engineering model, *The Astrophysical Journal* **614**, p. 1113–1123 (Oct 2004).
25. M. Kole, M. Pearce and M. Muñoz Salinas, A model of the cosmic ray induced atmospheric neutron environment, *Astroparticle Physics* **62**, 230 (2015).
26. E. Churazov, S. Sazonov, R. Sunyaev and M. Revnivtsev, Earth X-ray albedo for cosmic X-ray background radiation in the 1–1000 keV band, *Monthly Notices of the Royal Astronomical Society* **385**, 719 (02 2008).
27. S. Sazonov, E. Churazov, R. Sunyaev and M. Revnivtsev, Hard X-ray emission of the Earth's atmosphere: Monte Carlo simulations, *Monthly Notices of the Royal Astronomical Society* **377**, 1726 (05 2007).
28. M. Türler, M. Chernyakova, T. J.-L. Courvoisier, P. Lubiński, A. Neronov, N. Produit and R. Walter, Integralhard x-ray spectra of the cosmic x-ray background and galactic ridge emission, *Astronomy and Astrophysics* **512**, p. A49 (Mar 2010).

Prospect for WHIM detection in the cosmic web by *SRG/eROSITA*[1]

H. Tanimura* and N. Aghanim

Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, Bâtiment 121, 91405 Orsay, France

**E-mail: hideki.tanimura@ias.u-psud.fr*

Hydrodynamical simulations predict that the cosmic web contains the majority of the missing baryons in the form of plasma, called the warm-hot intergalactic medium (WHIM). However, its direct measurement through X-ray emissions has been prevented for decades due to the weak signal and the complex morphology of cosmic filaments.

We report the first statistical detection of X-ray emission from cosmic web filaments with the *ROSAT* data. We identified more than 15,000 large-scale filaments, spanning 30-100 Mpc length, in the SDSS survey and statistically detected X-ray emissions from the WHIM at 4.2 sigma confidence level using the *ROSAT* maps. Given this detection, we can expect a much more significant detection from *SRG/eROSITA* and indeed predicted the detectability of the WHIM. The prediction shows that stacking ∼2000 filaments only would lead to a 5σ detection with an average gas temperature of the WHIM as low as ∼0.3 keV.

Keywords: Clusters; intracluster medium; large-scale structure of Universe; cosmic background radiation

1. Missing baryon problem

Recent CMB observations indicate that baryons comprise 4.9% of the total energy density of the Universe,[2] and it is in agreement with the detected baryons at early times ($z > 2$), for example, through measurements of quasar absorption lines. However, at late times ($z < 2$), ∼30% of baryons have not been confirmed,[3] called the "Missing Baryon Problem." Hydrodynamical simulations predict that most of the missing baryons are expected to be in the form of hot plasma with the temperature range of 10^5–10^7 K and density of about 10–100 times the average cosmic value, referred to as the warm-hot intergalactic medium (WHIM), and to be located in filaments between galaxy clusters.[4] Therefore, extensive searches have been performed for the WHIM with far-UV, X-ray, and the thermal Sunyaev-Zel'dovich (tSZ) effect. However, the measurement is difficult because the signal is relatively weak, and the morphology of the source is complex.

2. WHIM detection with *ROSAT* data

To probe the WHIM in filaments, we used the filament catalog identified in the SDSS spectroscopic galaxies[5] and the *ROSAT* X-ray count-rate maps at six energy bands in 0.1 – 2.0 keV[6] [a]. We stacked the *ROSAT* maps at the positions of ∼15,000

[a] http://www.jb.man.ac.uk/research/cosmos/rosat/

cosmic filaments (30 to 100 Mpc at $0.2 < z < 0.6$) and detected the X-ray emissions for the first time at a significance of 4.2σ in the energy band of 0.56–1.21 keV (left panel in Fig. 1). While no significant detection was found at the other five energy bands, this result is predicted by hydrodynamic simulations,[7,8] which suggest that the relative contribution of the WHIM to the total CXB peaks at about 0.4–0.9 keV.

We performed an X-ray spectral analysis with the APEC model for the stacked signals at six energy bands and estimated the average gas density and temperature at the cores (< 2 Mpc) of the filaments (right panel in Fig. 1). The average central gas overdensity was found to be $\delta = 30 \pm 15$, assuming a cylindrical filament with a density distribution following a β-model with β=2/3. We compared this measurement with the other statistical tSZ measurement[9] and found that these results are consistent. The average gas temperature was also found to be $0.9^{+1.0}_{-0.6}$ keV, and it is a bit larger than the tSZ result of ~ 0.1 keV. This difference may imply that the gas temperature is higher at the core than at the outskirt, considering that the X-ray gas temperature was estimated within $\sim$2 Mpc from the filament spines, whereas the tSZ temperature was the average within $\sim$5 Mpc.

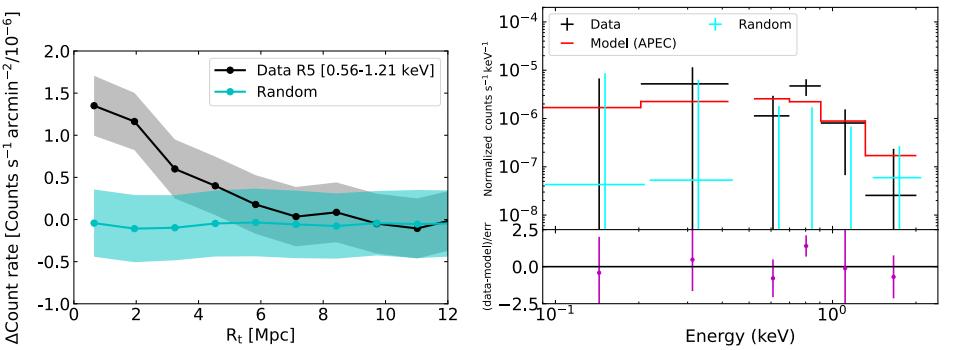


Fig. 1. *Left*: Average radial X-ray profile of 15,165 filaments in the energy band of 0.56–1.21 keV in black with the 1σ uncertainties in gray. Average radial X-ray profile from 1,000 random samplings in cyan. For each random sampling, 15,165 filaments are displaced at random positions on the *ROSAT* map. *Right*: Excesses of X-ray emission at the cores of 15,165 filaments in the six *ROSAT* energy bands in black. Amplitudes of X-ray emissions from 1,000 random samplings in cyan. Fit to the data with the APEC model in red. Ratio of data and model at each energy band is shown at the bottom in magenta.

3. Prospect of WHIM detection with *SRG/eROSITA*

A definitive statement on the gas temperature trend in filaments requires higher quality data, such as from the ongoing *SRG/eROSITA* all-sky survey. Therefore, we simulated the *SRG/eROSITA* observation of the X-ray emission from cosmic filaments based on the gas density and temperature estimated with the *ROSAT*

data and evaluated its detectability. The result shows that we can expect a $\sim 5\sigma$ detection by stacking only 100 filaments (left panel in Fig. 2).

While the detection becomes more challenging when the gas temperature is lower, we found that the stacking of ∼2,000 filaments would lead to a 5σ detection of X-ray emission, even with an average gas temperature as low as ∼0.3 keV (right panel in Fig. 2).

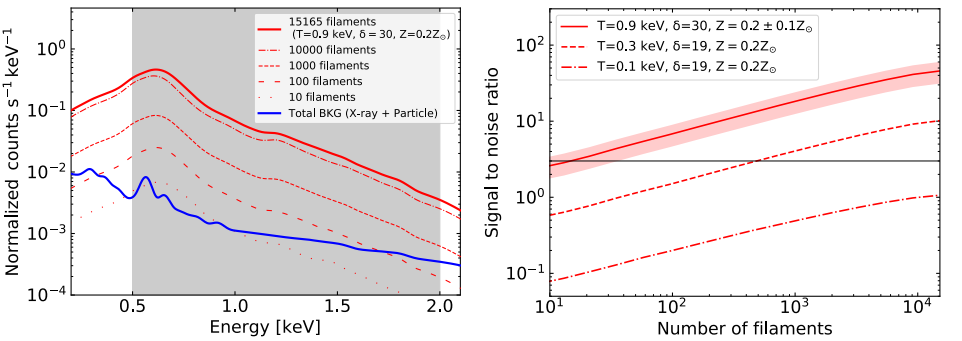


Fig. 2. *Left*: Simulated *SRG/eROSITA*'s energy spectra from the gas at the cores of the stacked 10, 100, 1,000, 10,000, and 15,165 filaments in red. The stacked spectra are calculated with overdensity = 30, temperature = 0.9 keV, and metallicity = $0.2Z_\odot$. *SRG/eROSITA*'s Background energy spectrum in blue. *Right*: Signal-to-noise ratios as a function of the number of stacked filaments are shown in red for "overdensity = 30, temperature = 0.9 keV and metallicity = $0.2Z_\odot$ at 0.5–2.0 keV", "overdensity = 19, temperature = 0.3 keV and metallicity = $0.2Z_\odot$ at 0.3–0.8 keV", and "overdensity = 19, temperature = 0.1 keV and metallicity = $0.2Z_\odot$ at 0.3–0.5 keV". Uncertainties due to the metallicity of $(0.2\pm0.1)Z_\odot$ are shown in light red area. Horizontal black line indicates a 3σ detection line.

Acknowledgement

This research has been supported by the funding for the ByoPiC project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement ERC-2015-AdG 695561.

References

1. H. Tanimura *et al.*, *A&A* **643**, p. L2 (November 2020).
2. Planck Collaboration, *A&A* **641**, p. A6 (September 2020).
3. J. M. Shull, B. D. Smith and C. W. Danforth, *ApJ* **759**, p. 23 (November 2012).
4. R. Cen and J. P. Ostriker, *ApJ* **650**, 560 (October 2006).
5. N. Malavasi *et al.*, *A&A* **642**, p. A19 (October 2020).
6. S. L. Snowden *et al.*, *ApJ* **485**, 125 (August 1997).
7. L. A. Phillips *et al.*, *ApJ* **554**, L9 (June 2001).
8. E. Ursino and M. Galeazzi, *ApJ* **652**, 1085 (December 2006).
9. H. Tanimura *et al.*, *A&A* **637**, p. A41 (May 2020).

The role of eXTP in the multi-messenger astronomy era

G. Stratta

INAF, Istituto di Astrofisica e Planetologia Spaziali, Rome, Italy
E-mail: giulia.stratta@inaf.it

Gor Oganesyan

Gran Sasso Science Institute, Viale F. Crispi 7, I-67100, L'Aquila (AQ), Italy

INFN - Laboratori Nazionali del Gran Sasso, I-67100, L'Aquila (AQ), Italy

The first detection of gravitational waves on 2015 with the Advanced LIGO and Advanced Virgo interferometers has opened a new observational window in the Universe. The last decade has also welcomed decisive discoveries in neutrino astronomy. Expected advances of gravitational wave and neutrino detectors by the end of the 2020s will mark the start of a golden era of multi-messenger astrophysics. The most promising multi-messenger sources in the high-energy sky, e.g. GRBs, AGNs, magnetars, are among the main targets for the enhanced X-ray Timing and Polarimetry (eXTP). In this proceeding, we describe the possible role of eXTP in the context of multi-messenger astronomy and in particular on the synergies with gravitational wave interferometers at the sensitivity expected by the end of the twenties.

Keywords: X-ray sources; gravitational waves; neutrinos; multi-messenger astrophysics.

1. Introduction

In the quickly evolving framework of multi-messenger astronomy, two breakthrough results have been achieved in the last few years that magnificently show the power of combining different "messengers" from the same source as well as the crucial role of high-energy observatories.

On 17 August 2017 a binary neutron star merger (GW170817) has been detected with the network of Advanced LIGO (aLIGO,[1]) and Advanced Virgo (AdV,[2]) gravitational wave interferometers. The source was found to be spatially and temporally coincident with a short Gamma Ray Burst GRB170817.[3,4] This association marked the first direct evidence that at least a fraction of short GRB progenitors are NS-NS mergers. Short GRB progenitor nature so far was based on indirect evidence collected in the last 15 years that goes from the lack of any core collapse supernova associated with a short GRB to the explosion site within the host galaxy, typically far from star forming regions (see[5] for a review). Both the burst and afterglow properties of GRB 170817 were found to be consistent with a significant inclination angle of the binary system. In particular, the afterglow properties provided

the first empirical evidence of a narrowly collimated jet and allowed to define its energy angular distribution.[6,7] Another emission component distinguished from the afterglow was also detected as a fading opt-NIR source (AT2017gfo)[8] and identified with the theoretically predicted thermal emission (kilonova) powered by the intense radioactive decay from instable heavy nuclei freshly-formed during the neutron star merger.[8,9] Past kilonova observations where claimed in few short GRBs and their identification relied on few data points (e.g.[10]). The exceptional high-quality temporal and spectral data of AT2017gfo allowed us to study in great details the kilonova properties for the first time.[11,12]

The other breakthrough discovery was achieved on 22 September 2017 when a high-energy neutrino event was detected with IceCube[13] and found to be spatially consistent with a gamma-ray emitting blazar at z=0.34 (TXS0506+056). The association has a confidence level lower than the case of GW/GRB170817 (i.e. $\sim 3\sigma$) but is an important finding that provide crucial hints on the origin of the identified diffuse flux of astrophysical very-high-energy neutrinos (10 TeV-10 PeV) detected by IceCube[14] and the possible role of blazars within most extreme particle accelerators in the Universe.

In the next few years, gravitational wave and neutrino detectors will undergo significant upgrades reaching unprecedented sensitivities. A new gravitational wave detector located in India (LIGO-India[a]) will join the current network of second generation (2G) interferometers formed by the two aLIGO in the USA, AdV in Italy and KAGRA in Japan.[15,16] Both aLIGO and AdV will reach a sensitivity larger than the nominal one with the upgraded configurations "A+" and "Virgo+"[17] during the fifth observation run (the start of which is currently scheduled for mid 2025). The completed network will be able to detect neutron star binary mergers (the most promising source for an associated electromagnetic counterpart) up to 330 Mpc, with an expected rate of few tens of detections per year (updated to the third observation run results).[17] By the end of the twenties, major upgrades are under consideration for aLIGO in the "Voyager" configuration[18] that will test some of the key technologies need for the third generation (3G) interferometers as the Cosmic Explorer (CE,[18,19]) and Einstein Telescope (ET,[20–22]) expected to be operational by mid-thirties. The "Voyager" configuration will possibly place the GW detection rate mid-way between the 2G and 3G detectors by the end of the twenties. The 3G detectors will have about one order of magnitude better sensitivity than current ones, with an expected rate of compact binary coalescences (CBCs) per year of the order of $O(10^5)$, up to distances that will cover the peak of star formation epoch and far beyond.[22]

Gravitational wave source sky localisation from the interferometer network is based on the triangulation method. By the end of the twenties, the completed GW detector network will have improved sky localisation capabilities, with average

[a]https://dcc.ligo.org/LIGO-M1100296/public

uncertainty regions of the order of few tens of square degrees for CBCs.[17] At distances larger than $\sim$ 100 Mpc, this scenario challenges the identification of the electromagnetic counterparts at optical wavelengths due to the large number of transient sources expected in such large areas and the incompleteness of galaxy catalogs above 100 Mpc. On the contrary, the high-energy transient sky is much less populated and gamma-ray or X-ray counterparts detected with all sky monitors might represent the most suitable way to pinpoint the GW source in the sky.

The GW source localisation uncertainties strongly depends on the presence of a network. This is still uncertain for the 3G interferometers during the thirties. A single GW detector such as ET, thanks to its sensitivity at low frequency, is able to localize about 100 binary neutron star merger per years within a few tens of square degrees up to a redshift of 0.3. These sources are detected before the mergers (some of them also few hours before the merger). For a network of detectors including ET and CE, the number of well localized events increases of an order of magnitude. However, going to large distances the sky-localization becomes of hundreds of square degrees for a large majority of events (e.g.,[22,23]). Early warnings from GW detectors allows ground and space-based telescopes to slew onto the identified sky region before the advent of the expected electromagnetic transient emission. Early warning alerts of the order of few minutes for the most nearby sources (i.e. NS-NS mergers at few tens of Mpc) are planned also for the 2G next observation runs[b].

For neutrino detectors, the Km3NeT[24] in the Mediterrean Sea will provide IceCube comparable sensitivites in the Northern Emisphere on the second half of the twenties and it will work in synergies with the Gigaton Volume Baikal late neutrino detector. In the Southern Emisphere, IceCube-gen2 will gradually increase the current IceCube sensitivity up to one order of magnitude,[25] boosting the current astrophysical neutrino detection rate to values that will significantly increase the chances to find the electromagnetic counterparts. The neutrino sky localization accuracy depends on neutrino flavour, where long tracks topology (for ν_μ) provide angular resolution down to 0.1-0.2 deg and a 2π sr sky coverage, while cascade topology (for ν_e and most ν_τ) angular resolution are typically 3-5 deg for the whole 4π sky, thus still requiring an electromagnetic counterpart and detectors with sky localization capabilities better than few arcmin to pinpoint the host galaxy and/or identify the source.

The launch of the enhanced X-ray Timing and Polarimetry mission eXTP,[26] currently scheduled by 2027[c], coincides with the middle/final stages of the above described major upgrades of neutrino and gravitational wave detectors. The recent breakthrough results described above, point to the gamma-ray bursts and active galactic nuclei among the most promising classes of multi-messenger sources that will be observed in the next years. Further promising candidates are represented by

[b] $https://emfollow.docs.ligo.org/userguide/early_warning.html$
[c] https://www.isdc.unige.ch/extp/

magnetars and core collapsing supernovae. All these sources are among the main targets for eXTP and a number of white papers describe all the science we will be able to do with eXTP.[27–30] The onboard wide field monitor (WFM, 2-50 keV, 3.7 sr field of view), with few arcmin-scale sky localization capabilities, and two sets of focusing telescope arrays with spectroscopic and polarimetric capabilities (SFA and PFA), will allow eXTP to play a crucial role in finding and characterizing the electromagnetic counterparts of the gravitational wave and neutrino sources by the end of the twenties.

In the following sections we focus on the capabilities of eXTP to observe two specific classes of multi-messenger sources: the compact binary coalescences (CBC) formed by two neutron stars (NS-NS) or a neutron star and a stellar-mass black hole (NS-BH), and core collapsing massive stars (ccSNe). For NS-NS mergers and ccSNe, a metastable or stable neutron star remnant may form after the main event. Both NS-N/NS-BH and ccSNe, as well as the possible NS remnant, have a number of theoretically expected emission components in the electromagnetic spectrum and in particular at high energies.

2. NS-NS/NS-BH mergers

Several electromagnetic components are expected in the soft/hard X-ray bands after the merger phase of coalescing NS-NS/NS-BH systems. A short GRB, possibly accompanied by a soft extended emission (see below), is expected soon after[d] the coalescence if the observer line of sight lies along the jet axis or within a certain angle range that depends on the combination of the source distance, the jet energetic angular distribution and the detector sensitivity. The collimated nature of gamma-ray emission implies that only a fraction of GW-detected NS-NS/NS-BH can be observed as short GRBs. To make a rough estimate, by assuming that all NS-NS/NS-BH produce a GRB, the number of GW+GRB detection is $f_b = (1-cos\theta_{jet})$ times lower, where f_b can be as low as 0.001 for $\theta_{jet} \sim 2-3$ deg. Interestingly, recent results on the jet structure based on the observation of GRB170817,[31] shows that at large viewing angles gamma-ray emission get fainter but also softer, thus allowing sensitive wide field monitors to cover energies down to a few keV, as eXTP/WFM, best suited for detecting the electromagnetic counterpart of most nearby NS-NS/NS-BH events that are more likely off-axis. Simultaneous detection of GRB+GW can provide a wealth of information on jet launching mechanisms exploiting for instance the temporal delay between the GRB trigger and the GW detection. Also a non-detection of a GRB counterpart can provide useful insights on jet formation efficiencies in NS-NS/NS-BH.

The impact of the jet with the circumburst environment gives rise to an afterglow emission that, depending on the viewing angle from the jet axis, it can show an initial steep decay followed by a plateau and then a typical $t^{-\alpha}$ decay (with $\alpha \sim 1-2$), or an

[d]For GW170817, the short burst was observed 1.7 seconds after the merger epoch.[4]

initial slow rising behaviour followed by a late time power law decay in case of large viewing angles. At very late time, e.g. months up to years after the NS-NS/NS-BH merger, the impact of the nearly isotropic ejecta of neutron-rich matter released during the merger, may eventually produce a mildy-relativistic blast wave from which synchrotron radiation could be observed in X-rays (e.g.,[32] see also §2.1.3). For NS-NS mergers, if a NS remnant is formed, nearly isotropic X-ray emission produced from the pulsar nebula might escape if successfully ionizes the surrounding ejecta of matter released during the merger. Such non-thermal X-ray emission is expected on timescales of the order of days for certain values of the pulsar magnetic field and ejecta mass and with luminosities up to $\sim 5 \times 10^{44}$ erg s^{-1}.[33,34]

We will see in the next sections how eXTP will play an important role in detecting these high-energy components, significantly contributing to multi-messenger observational campaigns by the end of the twenties.

2.1. *eXTP and GRB 170817-like sources*

In this section we present preliminary results from data simulations of X-ray emission from a GRB170817-like source observed with eXTP. For these simulations we used calibration files released by March 2021 and the data analysis software package XSPECv12.10.[35]

2.1.1. *Prompt emission*

GRB 170817 is so far the only short GRB that has been associated with a NS-NS merger system detected through GW emission (see §1). Despite a high-energy signal rather typical for short GRBs, with duration of $\sim$ 2s and peak photon flux of 3.7 ± 0.9 ph cm^{-2} s^{-1} in the 10-1000 keV band,[36] given the very small distance of the source ($\sim$ 40 Mpc,[3]), the isotropic equivalent luminosity $L_{iso} = 1.6 \times 10^{47}$ erg s^{-1}[37] is several orders of magnitude lower than the average value for short GRBs ($\sim 10^{52}$ erg s^{-1}, see e.g.[38]). According to Goldstein et al.,[36] the burst spectrum is best fitted by a power-law model with exponential cut-off $f(E) = A(E/E_0)^{\alpha} e^{-E(2+\alpha)/E_p}$ in the first 0.6 s, with $\alpha = -0.62 \pm 0.40$, peak energy[e] $E_p = 185 \pm 62$ keV, 10-1000 keV flux $(3.1 \pm 0.7) \times 10^{-7}$ erg cm^{-2} s^{-1} and fluence $(1.8 \pm 0.4) \times 10^{-7}$ erg cm^{-2}. During the tail of the burst, the spectrum is best fitted by a black-body with KT=(10.3 ± 1.5) keV, flux $(0.5 \pm 0.1) \times 10^{-7}$ erg cm^{-2} s^{-1} and fluence $(0.61 \pm 0.12) \times 10^{-7}$ erg cm^{-2} in the 10-1000 keV. By using these best-fit models, and introducing photoelectric absorption with equivalent hydrogen column density[f] $N_H = 7.5 \times 10^{20}$ cm^{-2},[36] we perform simulations of eXTP/WFM observations of the main burst in the 2-50 keV energy band and find that a GRB 170817-like burst at the same distance could be

[e]The peak energy is the cut-off energy multiplied by $(\alpha + 2)$

[f]based on the host galaxy NGC4993 dust extinction $A_V = 0.338$

confidently detected ($\sim 7\sigma$) during the first 0.6 s, if spectral parameters were at the 1σ soft side of the error margins, confirming past results.[27] We estimate that a GRB170817-like source could be detectable with eXTP/WFM under this spectral assumption up to about 70 Mpc.

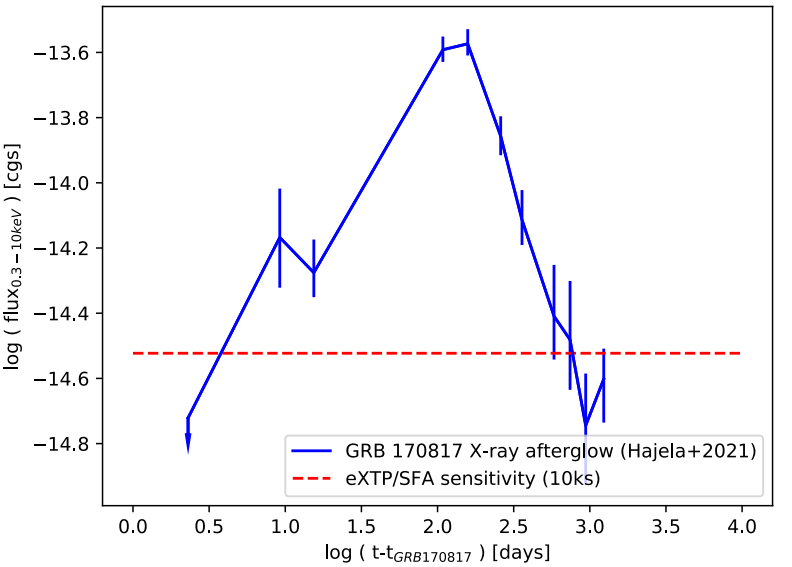


Fig. 1. X-ray (0.3-10 keV) flux of the short GRB 170817 afterglow that was associated with a NS-NS merger progenitor. Data are from Hajela et al. 2021 and references therein. The eXTP/SFA sensitivity will allow us to monitor a similar source from the very first phases up to the late ones, providing accurate sky localization. Horizontal line indicates the eXTP/SFA sensitivity for 10 ks of exposure from Zhang et al. 2016 and[26] computed in the 0.3-10 keV by assuming an absorbed power law model with phton index $\Gamma = 1.7$ and $N_H = 5 \times 10^{20}$ cm^{-2}.

2.1.2. *Afterglow emission*

The X-ray afterglow emission of GRB 170817 was detected with Chandra only 9 days after the merger as a faint source spatially coincident with the optical counterpart AT2017gfo and with flux $\sim 5 \times 10^{-15}$ erg cm^{-2} s^{-1} (0.3-10 keV,[39]). GRB170817 X-ray afterglow showed a slow rising behaviour that reached a peak of the emission after 100 days, with X-ray peak flux of a few 10^{-14} erg cm^{-2} s^{-1} (see Fig. 1). The overall properties of GRB170817 are in good agreement with the predictions for a GRB observed with a large inclination angle ($\sim 20-40$ deg) with respect to the jet axis. Given the eXTP/SFA sensitivities by assuming 10 ks of exposure ($\sim 2\times10^{-15}$ in the 2-10 keV energy band and $\sim 5\times10^{-16}$ in the 0.5-2 keV energy band, see Zhang et al. 2016), an X-ray afterglow of a GRB 170817-like source could be observed at almost all epochs, if accurate source sky localization is provided. The light curve

long monitoring provided by eXTP/SFA is crucial to disentangle among different jet structure modellings that in turn encode the physics of jet formation and launching mechanisms in GRBs.

2.1.3. *Very late time X-ray emission*

Currently, the X-ray emission of GRB 170817 is still detectable and shows an intriguing possible rebrightening phase 4 years after the merger epoch.[40] Among the possible interpretations, this rebrightening could be the evidence of the theoretically predicted non-thermal emission produced by the impact of the kilonova ejecta into the circumburst medium,[32,41] though other interpretations are under investigations.[42] Measured spectrum during this phase assuming a power-law model, has photon index $\Gamma = 1.6$ and an integrated flux of 2.5×10^{-14} erg cm^{-2} s^{-1} in the 0.3-10 keV energy range.[32] We find that by assuming the spectral parameters provided by,[32] such late time X-ray emission could be detected with eXTP/SFA at more than 5σ level with 20 ks of exposure.

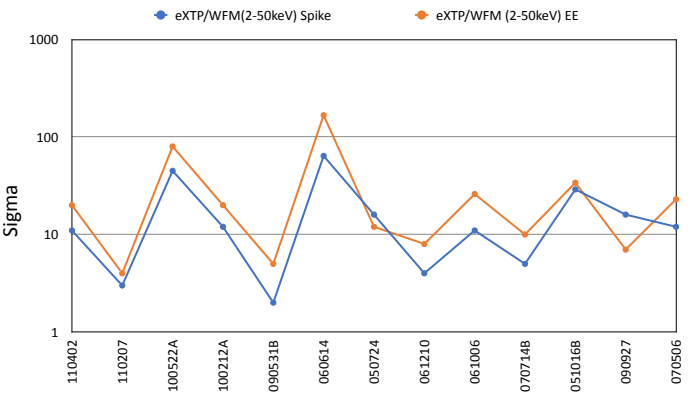


Fig. 2. Detection significance of a sample of short GRB spikes (blue) with Extended Emission (EE, orange). The x-axis shows the short GRB names, where the last 8 has measured redshift and cover a range from z=0.125 for GRB060614 to z=2.31 for GRB070506.

2.2. *Short GRB with Extended Emission*

An interesting perspective that potentially can increase the simultaneous GRB+GW detection is represented by the subclass of short GRBs that, following the main spike with hard spectral properties, show a prolongued softer and fainter Extended Emission (EE) that lasts after few tens of seconds.[43–46] The fraction of short GRBs with EE is still very uncertain and from high-energy detector catalogues range from $\sim 2 - 20\%$. However, recent studies over a large sample of short GRBs detected

with the coded mask telescope onboard the Neil Gehrels Swift Observatory (Swift, hereafter,[47]) and for which the X-ray telescope (XRT) observations could start during the EE, showed that providing a spectral coverage down to sof X-rays (i.e. $\sim$ few keV), the fraction of short GRBs with EE can be as high as 50% or more.[48] Among the possible hypothesis on the origin of the extended emission there is the presence of a magnetar formed soon after the merger of two neutron stars.[49,50] The important consequence in this case is that the extended emission component is expected to have a much lower degree of collimation with respect to the main spike that characterizes the short GRB.[33,50] If so, the EE detection rate can be in principle higher than short GRBs, making wide field monitors sensitive to the X-ray band as eXTP/WFM best suited to detect the electromagnetic counterparts of a potentially large fraction of NS-NS mergers.

We simulate 5 short GRB with EE with measured spectral parameters from Fermi/GBM and Swift/BAT joint analysis plus 8 short GRB with EE observed with Swift/BAT with known redshift. We take best fit parameters from Kaneko et al.[44] For these simulations we used calibration files released by March 2021 and the data analysis software package XSPECv12.10.[35] We find that EE component is always detected with eXTP/WFM and in most cases with better significance than the hard spike (see Figure 2).

2.3. *Polarisation of GRB jets*

2.3.1. *The prompt emission*

Despite the fact that GRBs have been observed more than forty years, the basic properties of the GRB jets are not yet constrained. We still can not discriminate baryonic jets[51,52] from ones dominated by the Poynting flux[53,54] by only analysing the spectral and temporal properties of short lasting prompt emission pulses ($\sim$0.1 s) at the keV-MeV energy range (see[55] and[56] for a review). We also do not have a clear understanding of the dissipation processes that lead to the GRB production. Our knowledge is limited by the lack of an interpretation of the GRB spectra which are typically modelled by the two broken power laws smoothly connected at $\sim$ 200 keV for LGRBs and at higher energies for SGRBs.[57] The observed spectra are too narrow to fit the fast cooling synchrotron radiation model and they are too broad compared with the simple thermal emission.[57,58] The absence of a single and straightforward emission mechanism to explain the typical GRB spectra has left us with big number of diverse physical models that correspond to very different GRB jet models. We can roughly classify the GRB production models by their dissipation side. The "standard" model assumes that the prompt emission is produced above the jet photosphere by the synchrotron emission of the non-thermal electrons that are produced by the internal shocks.[59,60] The "expected" parameters of the internal shocks model return fast cooling regime of the synchrotron radiation to dominate in the prompt phase, however it contradicts the observed values of the photon indices

at the low energies.[58] Other class of models assume that the dissipation of the jet takes the place below the jet photosphere via the radiation dominated shocks.[61,62] The complex nature of the formation of the final spectrum via many Compton scatterings allows to produce spectra much broader than a simple thermal ones, thus close to the observed GRB spectra. However, it remains quite challenging to explain the observed variability, the power law structure of the high energy tail of the GRB spectra and the long lasting X-ray tails of the GRB pulses in these models.[63] There are numerous effects that can modify the spectral expectations from internal shocks model or any other optically thin synchrotron sources (e.g. non-thermal particles from magnetic re-connection), including the anisotropy of electron's pitch angles or presence of an external source of photons (see[64] for a brief review). While all these models could explain some of the observed GRB spectra, or at least partially (at the low or the high energy part), their prediction on the level of polarisation and the evolution of the polarisation angle is quite different, also depending on the structure of the local magnetic field (see[65] for a review). Therefore, the polarisation measure of the prompt phase is crucial for discriminating among GRB jet models including the acceleration processes (the magnetic field orientation).

So far the attempts to measure the GRB polarisation over short time-scales in the soft γ-rays returned inconclusive results due to contradictions between instruments and methods of data reduction performed by the Gamma-Ray Burst Polarimeter (GAP) (Yonetoku et al. 2011), POLAR,[66] and the Astrosat CZTI[67] instruments. The future γ-ray polarimeters, such as POLAR-2, LEAP and SPHiNX, have the capabilities to improve the polarisation measure accuracy of the prompt emission phase. However, measuring the polarization in the X-rays is equally interesting and has its own advantages. While the prompt emission pulses are very short in the soft γ-rays, we do observe regularly their tails (X-ray steep decay phase) for much longer time (100-1000 s) in 0.3-10 keV thanks to Swift/XRT. The X-ray steep decay is quite bright and starts at $\sim 10^{-7}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ at the peak of the prompt pulse and drops in two orders of magnitude in 100-1000 s. This phase is modelled as remnant of the prompt emission, i.e. the emission we observe when the GRB pulse is diminished at the jet co-moving frame.[68] Recently, it was demonstrated that X-ray steep decay is the lower energy prompt emission originated by the adiabatically cooling jet after its heat source is switched off.[69] Measuring X-ray polarization in the steep decay phase by the Polarimetry Focusing array (PFA) onboard eXTP would be a direct probe of the prompt emission polarisation (see[70]). The sensitivity of PFA is enough to constrain the linear polarisation degree for exposures of 100-1000 s given the extraordinary brightness of the X-ray steep decay phase. It requires a precise sky localisation of each detected GRB by the wide field γ or X-ray instruments, comparable with the extremely narrow FoV of PFA. The localisation level of arcminutes are expected to be provided by the eXTP/WFM and also by the next generation nano-satellite detectors such as HERMES.[71] Therefore, the prompt communication between eXTP instruments and advanced networks of γ-ray

telescopes could provide a unique tool to study early X-ray polarisation. The presence of precursors and the quiescent periods in the temporal structure of LGRBs could be useful to send early trigger to detect bright X-ray counterparts of GRBs by PFA.

While *Swift*/XRT has been able to detect the steep decay phase of LGRBs, it is still technically challenging to promptly slew in the case of SGRBs. Given the spectral differences between long and short GRBs, one would expect physical differences between their jets and the dissipation processes. Therefore, the early detection of the X-ray emission from SGRBs and the measurement of its polarisation is of a great interest. The early warning and localisation of GWs prior to the BNS merger by the next generation GW detectors, such as the Einstein Telescope,[72] could be a unique trigger for the detection of the early X-ray polarisation of the the tail of the prompt emission from SGRBs.

2.3.2. *The afterglow emission*

Although we have very little knowledge on the physics of the prompt emission production, GRB afterglow emission (lasting for days) has a quite convincing interpretation. It is widely accepted that the dissipation of the GRB jet in the circumburst medium via shocks produces afterglow emission.[73–75] There have been several indications of detection of the linear polarisation in the late afterglow emission at the level of few percent[76] and a detection of circular polarisation in the afterglow of GRB 121024A.[77] The linear polarisation of level of few percent has been found at very early times,[78] when the emission is dominated by the reverse shock propagating into the GRB jet ejecta. Thus these measurements result to be essential for probing the shock physics by constraining the orientation of the magnetic field relative to the bulk motion, as well as for understanding the nature of the GRB jet by itself (reverse shock). Furthermore, nearby GRB sources can be detected off-axis, as the multi-messenger event of GW 170817/GRB 170817A. An upper limit of 12 percent in the linear polarisation at 2.8 GHz was obtained 244 days after the binary neutron star merger.[79] This upper limit allowed us to establish that the magnetic field has a finite parallel component, calling for turbulence amplified magnetic field in the external shock.[80]

X-ray domain in the physics of the GRB afterglow has a special importance for the following reasons. While the radio, optical and Very High Energy emissions in the afterglow phase are typically consistent with the standard external shocks model, *Swift* has discovered, that roughly half of the X-ray afterglow emission deviated from the standard forward shock emission, showing a long lasting (up to $\sim 10^4$ s) plateau phase, which is characterised by nearly constant flux. The X-ray plateau emission is can be produced by the additional energy input from the GRB central engine (e.g. newborn magnetar)[68] or it can be caused by the viewing angle effects[81] or by the delayed prompt emission from the jet wings.[82,83] In either case, the measure of the polarisation would be essential for constraining the physics

of this mysterious GRB phase, simply because the polarisation would probe the structure of the magnetic field of the emission surface. From the observational perspective, the plateau phase has some advantages; it appears later compared to the steep decay phase giving more time to slew on, lasts longer ($\sim 10^4$ s), it is relatively bright ($\sim 10^{-10}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ comparable with the PFA sensitivity) and with absence of significant spectral evolution, which could remove polarisation. In the era of the Einstein Telescope, it would be also possible to constrain the BNS postmerger signal for relatively nearby sources. Combining the information from the BNS postmerger signals together with the X-ray polarisation from the plateau phase will be extremely powerful to shed light on our understanding of the origin of the central engine powering SGRBs.

Another intriguing possibility to probe the prompt emission is to measure the polarisation of the late-time X-ray flares. *Swift* has discovered numerous X-ray flares superimposed to the afterglow emission.[84] These flares appear up to 10^5 s after the GRB detection and their variability suggests their internal nature, i.e. they could be hardly produced by the external shocks. Therefore, they are often interpreted as late-time prompt emission. Their late-time appearance and relative brightness increases the chances for PFA to measure their linear polarisation,[85] thus providing complementary observations to the the next generation γ-ray polarimeters. From one side, the X-ray flares polarisation measurements would probe the prompt emission zone, and on the other side, the comparison between the polarisation measurements at early times in γ-rays and at late times in X-rays would finally help us to establish the nature of the late time X-ray activity.

3. Core collapse supernovae and magnetars

Core collapse supernovae (ccSNe) have been claimed as another potential class of high-frequency gravitational wave sources,[86–88] as well as neutrino sources,[89] that current and/or future ground-based GW detectors and neutrino detectors may reveal in the next years. The ccSNe energy output transferred with GWs however is still very uncertain depending on the rather unnown explosion mechanism. Recent modelling contraints on possible GW energy values, obtained from a search of GW transients associated with optically detected ccSNe within ~ 20 Mpc during the first two observational runs of aLIGO and AdV (between 2015 and 2017), suggest that ccSNe GW detection may happen in the next years for sources up to few tens of Mpc in a few, very optimistic scenarios.[90] The GW signal detection from ccSNe, possibly combined with neutrinos detection, represents a unique opportunity to probe the inner dynamics inaccessible to electromagnetic observations. The electromagnetic counterpart of ccSNe plays a key role in allowing to confirm or discard the likely very faint GW signal by providing a precise sky localization and an estimate of the epoch of the explosion.

At high-energies, the best GW-detected ccSNe electromagnetic counterparts would be the typical long GRBs due to their brightness and the expected temporal

proximity to the epoch of the GW burst originating from the core collapse. The rate of typical long GRBs, however, is expected to be extremely low due to the small fraction of ccSNe known to be associated with such energetic events (only about few %, e.g.[91]), and to the combination of the GRB collimated nature and the short distances that GW detectors may reach for this type of burst signals. However, a subclass of anomalously faint and soft long GRBs, named Low Luminosity GRBs (LLGRBs), have been suggested to represent a distincted class of long GRBs that populate the local Universe at a much higher rate (up to 10^3 times the typical long GRBs). The redshift distribution of LLGRBs detected so far is well below 1, where the most nearby was detected at z=0.0085 (37 Mpc).[92] These properties make LLGRBs an interesting class of sources that may enhance the chances of simultaneous detection with GWs with high-energy wide field monitors. LLGRBs are also softer than typical long GRBs and this highlight the role of wide field monitors with peak sensitivity between few to few tens of keV as eXTP/WFM. We simulated how eXTP/WFM would detect GRB 060218, a typical LLGRB observed with Swift/BAT. This was a close GRB at 139 Mpc showing an unusual long and soft burst, with duration 2740s.[93] The burst was best fitted by a simple power law spectrum in the 15-150 keV with photon index $\Gamma = 2.5 \pm 0.1$, fluence of 5.8×10^{-6} erg cm^{-2} and peak flux 2×10^{-8} erg cm^{-2} s^{-1}. By simulating with XSPECv12.0 the eXTP/WFM observations using the most recent calibration files we find that with 1ks of exposure this burst could have been clearly detected with significance of $\sim 40\ \sigma$.

Another electromagnetic counterpart at high-energy is represented by the X-ray emission from the ccSNe shock breakout (SBO). The advantage of such component with respect to LLGRBs is that it is thought to be not collimated. However, the SBO phenomenology lack of statistics so far, with only one event observed (XRO 080109/SN2008D,[94]). The 0.3-10 keV spectrum of XRO080109 was consistent with a power law with photon index $\Gamma = 2.3 \pm 0.3$, equivalent hydrogen column density $N_H = 6.8^{+1.8}_{-1.5} \times 10^{21}$ cm^{-2} and peak flux of 7×10^{-10} erg s^{-1} cm^{-2}.[94] Such low fluxes could not had been detected by the majority of current wide field monitors. Indeed, its discovery happened serendipitously by falling within the Swift/XRT field of view while performing a follow-up campaign of another source in the very same galaxy. By assuming the best fit spectral model of XRO080109, very preliminary simulations with eXTP/WFM with 400 s of exposure show that this source would not be detected, confining the SBO detection with eXTP/WFM within our Galaxy.

4. Conclusions

The eXTP expected launch date and lifetime are timely for a full synergy with upgraded second generation GW detectors and with next generation neutrino detectors. Among the most promising multimessenger sources there are the compact binary mergers from which the high-energy electromagnetic counterpart is expected in the form of short GRBs and their afterglows and possibly additional X-ray emission,

yet to be confirmed, as from the impact of the kilonova with the circumburst medium. We have shown how eXTP/WFM can observe short GRBs, including the "Extended emission" observed in a fraction of short GRBs. The eXTP/SFA could have monitored almost all the evolution of the associated X-ray afterglow of GRB 170817 up to very late times and the eXTP/WFM detect the soft component of the prompt emission up to 70 Mpc. Both neutrino and GW are expected from cc-SNe, a subclass of which is the progenitor of long GRBs and in particular of the much more numerous low luminosity GRBs that are populating the nearby Universe. eXTP/WFM can play a crucial role in detecting the associated GRB and providing accurate burst timing and sky localization that will help in refining GW/neutrino event detection confidence and parameter estimation. In particular, the eXTP/WFM sky localization will allow us to trigger follow-up campaigns by next generation facilities as ELT, SKA, CTA, etc. to characterize GW sources with unprecedented sensitivities. eXTP/SPA is capable to detect the GRB prompt and afterglow emission polarisation to disentangle among different jet launching mechanisms as well as to shed light on the nature of the late time activity observed in the X-ray afterglow of several events.

References

1. J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari et al., Advanced LIGO, *Class. Quantum Grav.* **32**, p. 074001 (April 2015).
2. F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin et al., Advanced Virgo: A second-generation interferometric gravitational wave detector, *Class. Quantum Grav.* **32**, p. 024001 (January 2015).
3. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X., V. B. Adya et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Letter* **119**, p. 161101 (October 2017).
4. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari et al., Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *ApJL* **848**, p. L13 (October 2017).
5. E. Berger, Short-Duration Gamma-Ray Bursts, *Ann. Rev. Astron. Astrophys.* **52**, 43 (August 2014).
6. G. Ghirlanda, O. S. Salafia, Z. Paragi, M. Giroletti, J. Yang, B. Marcote, J. Blanchard, I. Agudo, T. An, M. G. Bernardini, R. Beswick, M. Branchesi, S. Campana, C. Casadio, E. Chassand e-Mottin, M. Colpi, S. Covino, P. D'Avanzo, V. D'Elia, S. Frey, M. Gawronski, G. Ghisellini, L. I. Gurvits, P. G. Jonker, H. J. van Langevelde, A. Melandri, J. Moldon, L. Nava, A. Perego, M. A. Perez-Torres, C. Reynolds, R. Salvaterra, G. Tagliaferri, T. Venturi, S. D. Vergani and M. Zhang, Compact radio emission indicates a structured jet was produced by a binary neutron star merger, *Science* **363**, 968 (Mar 2019).
7. K. P. Mooley, A. T. Deller, O. Gottlieb, E. Nakar, G. Hallinan, S. Bourke, D. A. Frail, A. Horesh, A. Corsi and K. Hotokezaka, Superluminal motion of a relativistic jet in the neutron-star merger GW170817, *Nature* **561**, 355 (September 2018).

8. B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari et al., Multi-messenger Observations of a Binary Neutron Star Merger, *ApJL* **848**, p. L12 (October 2017).
9. D. A. Coulter, R. J. Foley, C. D. Kilpatrick, M. R. Drout, A. L. Piro, B. J. Shappee, M. R. Siebert, J. D. Simon, N. Ulloa, D. Kasen, B. F. Madore, A. Murguia-Berthier, Y. C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest and C. Rojas-Bravo, Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source, *Science* **358**, 1556 (Dec 2017).
10. N. R. Tanvir, A. J. Levan, A. S. Fruchter, J. Hjorth, R. A. Hounsell, K. Wiersema and R. L. Tunnicliffe, A 'kilonova' associated with the short-duration γ-ray burst GRB 130603B, *Nature* **500**, 547 (August 2013).
11. E. Pian, P. D'Avanzo, S. Benetti, M. Branchesi, E. Brocato, S. Campana, E. Cappellaro, S. Covino, V. D'Elia, J. P. U. Fynbo, F. Getman, G. Ghirlanda, G. Ghisellini, A. Grado, G. Greco, J. Hjorth, C. Kouveliotou, A. Levan, L. Limatola, D. Malesani, P. A. Mazzali, A. Melandri, P. Møller, L. Nicastro, E. Palazzi, S. Piranomonte, A. Rossi, O. S. Salafia, J. Selsing, G. Stratta, M. Tanaka, N. R. Tanvir, L. Tomasella, D. Watson, S. Yang, L. Amati, L. A. Antonelli, S. Ascenzi, M. G. Bernardini, M. Boër, F. Bufano, A. Bulgarelli, M. Capaccioli, P. Casella, A. J. Castro-Tirado, E. Chassande-Mottin, R. Ciolfi, C. M. Copperwheat, M. Dadina, G. De Cesare, A. di Paola, Y. Z. Fan, B. Gendre, G. Giuffrida, A. Giunta, L. K. Hunt, G. L. Israel, Z. P. Jin, M. M. Kasliwal, S. Klose, M. Lisi, F. Longo, E. Maiorano, M. Mapelli, N. Masetti, L. Nava, B. Patricelli, D. Perley, A. Pescalli, T. Piran, A. Possenti, L. Pulone, M. Razzano, R. Salvaterra, P. Schipani, M. Spera, A. Stamerra, L. Stella, G. Tagliaferri, V. Testa, E. Troja, M. Turatto, S. D. Vergani and D. Vergani, Spectroscopic identification of r-process nucleosynthesis in a double neutron-star merger, *Nature* **551**, 67 (Nov 2017).
12. S. J. Smartt, T. W. Chen, A. Jerkstrand, M. Coughlin, E. Kankare, S. A. Sim, M. Fraser, C. Inserra, K. Maguire, K. C. Chambers, M. E. Huber, T. Krühler, G. Leloudas, M. Magee, L. J. Shingles, K. W. Smith, D. R. Young, J. Tonry, R. Kotak, A. Gal-Yam, J. D. Lyman, D. S. Homan, C. Agliozzo, J. P. Anderson, C. R. Angus, C. Ashall, C. Barbarino, F. E. Bauer, M. Berton, M. T. Botticella, M. Bulla, J. Bulger, G. Cannizzaro, Z. Cano, R. Cartier, A. Cikota, P. Clark, A. De Cia, M. Della Valle, L. Denneau, M. Dennefeld, L. Dessart, G. Dimitriadis, N. Elias-Rosa, R. E. Firth, H. Flewelling, A. Flörs, A. Franckowiak, C. Frohmaier, L. Galbany, S. González-Gaitán, J. Greiner, M. Gromadzki, A. N. Guelbenzu, C. P. Gutiérrez, A. Hamanowicz, L. Hanlon, J. Harmanen, K. E. Heintz, A. Heinze, M. S. Hernandez, S. T. Hodgkin, I. M. Hook, L. Izzo, P. A. James, P. G. Jonker, W. E. Kerzendorf, S. Klose, Z. Kostrzewa-Rutkowska, M. Kowalski, M. Kromer, H. Kuncarayakti, A. Lawrence, T. B. Lowe, E. A. Magnier, I. Manulis, A. Martin-Carrillo, S. Mattila, O. McBrien, A. Müller, J. Nordin, D. O'Neill, F. Onori, J. T. Palmerio, A. Pastorello, F. Patat, G. Pignata, P. Podsiadlowski, M. L. Pumo, S. J. Prentice, A. Rau, A. Razza, A. Rest, T. Reynolds, R. Roy, A. J. Ruiter, K. A. Rybicki, L. Salmon, P. Schady, A. S. B. Schultz, T. Schweyer, I. R. Seitenzahl, M. Smith, J. Sollerman, B. Stalder, C. W. Stubbs, M. Sullivan, H. Szegedi, F. Taddia, S. Taubenberger, G. Terreran, B. van Soelen, J. Vos, R. J. Wainscoat, N. A. Walton, C. Waters, H. Weiland, M. Willman, P. Wiseman, D. E. Wright, Ł. Wyrzykowski and O. Yaron, A kilonova as the electromagnetic counterpart to a gravitational-wave source, *Nature* **551**, 75 (Nov 2017).
13. IceCube Collaboration, Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A, *Science* **361**, p. eaat1378 (July 2018).

14. IceCube Collaboration, Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector, *Science* **342**, p. 1242856 (November 2013).
15. K. Somiya, Detector configuration of KAGRA-the Japanese cryogenic gravitational-wave detector, *Classical and Quantum Gravity* **29**, p. 124007 (June 2012).
16. Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi and H. Yamamoto, Interferometer design of the kagra gravitational wave detector, *Phys. Rev. D* **88**, p. 043007 (Aug 2013).
17. B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, L. S. C. Kagra Collaboration and VIRGO Collaboration, Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Reviews in Relativity* **23**, p. 3 (September 2020).
18. D. Reitze *et al.*, Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO, *Bull. Am. Astron. Soc.* **51**, p. 035 (2019).
19. B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, C. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, N. Aggarwal, O. D. Aguiar, A. Ain, P. Ajith, B. Allen, P. A. Altin, S. B. Anderson, W. G. Anderson, K. Arai, M. C. Araya, C. C. Arceneaux, J. S. Areeda, K. G. Arun, G. Ashton, M. Ast, S. M. Aston, P. Aufmuth, C. Aulbert, S. Babak, P. T. Baker, S. W. Ballmer, J. C. Barayoga, S. E. Barclay, B. C. Barish, D. Barker, B. Barr, L. Barsotti, J. Bartlett, I. Bartos, R. Bassiri, J. C. Batch, C. Baune, A. S. Bell, B. K. Berger, G. Bergmann, C. P. L. Berry, J. Betzwieser, S. Bhagwat, R. Bhandare, I. A. Bilenko, G. Billingsley, J. Birch, R. Birney, S. Biscans, A. Bisht, C. Biwer, J. K. Blackburn, C. D. Blair, D. G. Blair, R. M. Blair, O. Bock, C. Bogan, A. Bohe, C. Bond, R. Bork, S. Bose, P. R. Brady, V. B. Braginsky, J. E. Brau, M. Brinkmann, P. Brockill, J. E. Broida, A. F. Brooks, D. A. Brown, D. D. Brown, N. M. Brown, S. Brunett, C. C. Buchanan, A. Buikema, A. Buonanno, R. L. Byer, M. Cabero, L. Cadonati, C. Cahillane, J. C. Bustillo, T. Callister, J. B. Camp, K. C. Cannon, J. Cao, C. D. Capano, S. Caride, S. Caudill, M. Cavaglià, C. B. Cepeda, S. J. Chamberlin, M. Chan, S. Chao, P. Charlton, B. D. Cheeseboro, H. Y. Chen, Y. Chen, C. Cheng, H. S. Cho, M. Cho, J. H. Chow, N. Christensen, Q. Chu, S. Chung, G. Ciani, F. Clara, J. A. Clark, C. G. Collette, L. Cominsky, M. Constancio, D. Cook, T. R. Corbitt, N. Cornish, A. Corsi, C. A. Costa, M. W. Coughlin, S. B. Coughlin, S. T. Countryman, P. Couvares, E. E. Cowan, D. M. Coward, M. J. Cowart, D. C. Coyne, R. Coyne, K. Craig, J. D. E. Creighton, J. Cripe, S. G. Crowder, A. Cumming, L. Cunningham, T. D. Canton, S. L. Danilishin, K. Danzmann, N. S. Darman, A. Dasgupta, C. F. D. S. Costa, I. Dave, G. S. Davies, E. J. Daw, S. De, D. DeBra, W. D. Pozzo, T. Denker, T. Dent, V. Dergachev, R. T. DeRosa, R. DeSalvo, R. C. Devine, S. Dhurandhar, M. C. Díaz, I. D. Palma, F. Donovan, K. L. Dooley, S. Doravari, R. Douglas, T. P. Downes, M. Drago, R. W. P. Drever, J. C. Driggers, S. E. Dwyer, T. B. Edo, M. C. Edwards, A. Effler, H.-B. Eggenstein, P. Ehrens, J. Eichholz, S. S. Eikenberry, W. Engels, R. C. Essick, T. Etzel, M. Evans, T. M. Evans, R. Everett, M. Factourovich, H. Fair, S. Fairhurst, X. Fan, Q. Fang, B. Farr, W. M. Farr, M. Favata, M. Fays, H. Fehrmann, M. M. Fejer, E. Fenyvesi, E. C. Ferreira, R. P. Fisher, M. Fletcher, Z. Frei, A. Freise, R. Frey, P. Fritschel, V. V. Frolov, P. Fulda, M. Fyffe, H. A. G. Gabbard, J. R. Gair, S. G. Gaonkar, G. Gaur, N. Gehrels, P. Geng, J. George, L. Gergely, A. Ghosh, A. Ghosh, J. A. Giaime, K. D. Giardina, K. Gill, A. Glaefke, E. Goetz, R. Goetz, L. Gondan, G. González, A. Gopakumar, N. A. Gordon, M. L. Gorodetsky, S. E. Gossan, C. Graef, P. B. Graff, A. Grant, S. Gras, C. Gray, A. C. Green, H. Grote, S. Grunewald, X. Guo, A. Gupta, M. K. Gupta, K. E. Gushwa, E. K. Gustafson, R. Gustafson, J. J. Hacker, B. R. Hall, E. D. Hall, G. Hammond,

M. Haney, M. M. Hanke, J. Hanks, C. Hanna, M. D. Hannam, J. Hanson, T. Hardwick, G. M. Harry, I. W. Harry, M. J. Hart, M. T. Hartman, C.-J. Haster, K. Haughian, M. C. Heintze, M. Hendry, I. S. Heng, J. Hennig, J. Henry, A. W. Heptonstall, M. Heurs, S. Hild, D. Hoak, K. Holt, D. E. Holz, P. Hopkins, J. Hough, E. A. Houston, E. J. Howell, Y. M. Hu, S. Huang, E. A. Huerta, B. Hughey, S. Husa, S. H. Huttner, T. Huynh-Dinh, N. Indik, D. R. Ingram, R. Inta, H. N. Isa, M. Isi, T. Isogai, B. R. Iyer, K. Izumi, H. Jang, K. Jani, S. Jawahar, L. Jian, F. Jiménez-Forteza, W. W. Johnson, D. I. Jones, R. Jones, L. Ju, K. Haris, C. V. Kalaghatgi, V. Kalogera, S. Kandhasamy, G. Kang, J. B. Kanner, S. J. Kapadia, S. Karki, K. S. Karvinen, M. Kasprzack, E. Katsavounidis, W. Katzman, S. Kaufer, T. Kaur, K. Kawabe, M. S. Kehl, D. Keitel, D. B. Kelley, W. Kells, R. Kennedy, J. S. Key, F. Y. Khalili, S. Khan, Z. Khan, E. A. Khazanov, N. Kijbunchoo, C.-W. Kim, C. Kim, J. Kim, K. Kim, N. Kim, W. Kim, Y.-M. Kim, S. J. Kimbrell, E. J. King, P. J. King, J. S. Kissel, B. Klein, L. Kleybolte, S. Klimenko, S. M. Koehlenbeck, V. Kondrashov, A. Kontos, M. Korobko, W. Z. Korth, D. B. Kozak, V. Kringel, C. Krueger, G. Kuehn, P. Kumar, R. Kumar, L. Kuo, B. D. Lackey, M. Landry, J. Lange, B. Lantz, P. D. Lasky, M. Laxen, A. Lazzarini, S. Leavey, E. O. Lebigot, C. H. Lee, H. K. Lee, H. M. Lee, K. Lee, A. Lenon, J. R. Leong, Y. Levin, J. B. Lewis, T. G. F. Li, A. Libson, T. B. Littenberg, N. A. Lockerbie, A. L. Lombardi, L. T. London, J. E. Lord, M. Lormand, J. D. Lough, H. Lück, A. P. Lundgren, R. Lynch, Y. Ma, B. Machenschalk, M. MacInnis, D. M. Macleod, F. Magaña-Sandoval, L. M. Zertuche, R. M. Magee, V. Mandic, V. Mangano, G. L. Mansell, M. Manske, S. Márka, Z. Márka, A. S. Markosyan, E. Maros, I. W. Martin, D. V. Martynov, K. Mason, T. J. Massinger, M. Masso-Reid, F. Matichard, L. Matone, N. Mavalvala, N. Mazumder, R. McCarthy, D. E. McClelland, S. McCormick, S. C. McGuire, G. McIntyre, J. McIver, D. J. McManus, T. McRae, S. T. McWilliams, D. Meacher, G. D. Meadors, A. Melatos, G. Mendell, R. A. Mercer, E. L. Merilh, S. Meshkov, C. Messenger, C. Messick, P. M. Meyers, H. Miao, H. Middleton, E. E. Mikhailov, A. L. Miller, A. Miller, B. B. Miller, J. Miller, M. Millhouse, J. Ming, S. Mirshekari, C. Mishra, S. Mitra, V. P. Mitrofanov, G. Mitselmakher, R. Mittleman, S. R. P. Mohapatra, B. C. Moore, C. J. Moore, D. Moraru, G. Moreno, S. R. Morriss, K. Mossavi, C. M. Mow-Lowry, G. Mueller, A. W. Muir, A. Mukherjee, D. Mukherjee, S. Mukherjee, N. Mukund, A. Mullavey, J. Munch, D. J. Murphy, P. G. Murray, A. Mytidis, R. K. Nayak, K. Nedkova, T. J. N. Nelson, A. Neunzert, G. Newton, T. T. Nguyen, A. B. Nielsen, A. Nitz, D. Nolting, M. E. N. Normandin, L. K. Nuttall, J. Oberling, E. Ochsner, J. O'Dell, E. Oelker, G. H. Ogin, J. J. Oh, S. H. Oh, F. Ohme, M. Oliver, P. Oppermann, R. J. Oram, B. O'Reilly, R. O'Shaughnessy, D. J. Ottaway, H. Overmier, B. J. Owen, A. Pai, S. A. Pai, J. R. Palamos, O. Palashov, A. Pal-Singh, H. Pan, C. Pankow, F. Pannarale, B. C. Pant, M. A. Papa, H. R. Paris, W. Parker, D. Pascucci, Z. Patrick, B. L. Pearlstone, M. Pedraza, L. Pekowsky, A. Pele, S. Penn, A. Perreca, L. M. Perri, M. Phelps, V. Pierro, I. M. Pinto, M. Pitkin, M. Poe, A. Post, J. Powell, J. Prasad, V. Predoi, T. Prestegard, L. R. Price, M. Prijatelj, M. Principe, S. Privitera, L. Prokhorov, O. Puncken, M. Pürrer, H. Qi, J. Qin, S. Qiu, V. Quetschke, E. A. Quintero, R. Quitzow-James, F. J. Raab, D. S. Rabeling, H. Radkins, P. Raffai, S. Raja, C. Rajan, M. Rakhmanov, V. Raymond, J. Read, C. M. Reed, S. Reid, D. H. Reitze, H. Rew, S. D. Reyes, K. Riles, M. Rizzo, N. A. Robertson, R. Robie, J. G. Rollins, V. J. Roma, G. Romanov, J. H. Romie, S. Rowan, A. Rüdiger, K. Ryan, S. Sachdev, T. Sadecki, L. Sadeghian, M. Sakellariadou, M. Saleem, F. Salemi, A. Samajdar, L. Sammut, E. J. Sanchez, V. Sandberg, B. Sandeen, J. R. Sanders, B. S. Sathyaprakash, P. R. Saulson, O. E. S. Sauter, R. L. Savage, A. Sawadsky, P. Schale, R. Schilling, J. Schmidt, P. Schmidt,

R. Schnabel, R. M. S. Schofield, A. Schönbeck, E. Schreiber, D. Schuette, B. F. Schutz, J. Scott, S. M. Scott, D. Sellers, A. S. Sengupta, A. Sergeev, D. A. Shaddock, T. Shaffer, M. S. Shahriar, M. Shaltev, B. Shapiro, P. Shawhan, A. Sheperd, D. H. Shoemaker, D. M. Shoemaker, K. Siellez, X. Siemens, D. Sigg, A. D. Silva, A. Singer, L. P. Singer, A. Singh, R. Singh, A. M. Sintes, B. J. J. Slagmolen, J. R. Smith, N. D. Smith, R. J. E. Smith, E. J. Son, B. Sorazu, T. Souradeep, A. K. Srivastava, A. Staley, M. Steinke, J. Steinlechner, S. Steinlechner, D. Steinmeyer, B. C. Stephens, R. Stone, K. A. Strain, N. A. Strauss, S. Strigin, R. Sturani, A. L. Stuver, T. Z. Summerscales, L. Sun, S. Sunil, P. J. Sutton, M. J. Szczepańczyk, D. Talukder, D. B. Tanner, M. Tápai, S. P. Tarabrin, A. Taracchini, R. Taylor, T. Theeg, M. P. Thirugnanasambandam, E. G. Thomas, M. Thomas, P. Thomas, K. A. Thorne, E. Thrane, V. Tiwari, K. V. Tokmakov, K. Toland, C. Tomlinson, Z. Tornasi, C. V. Torres, C. I. Torrie, D. Töyrä, G. Traylor, D. Trifirò, M. Tse, D. Tuyenbayev, D. Ugolini, C. S. Unnikrishnan, A. L. Urban, S. A. Usman, H. Vahlbruch, G. Vajente, G. Valdes, D. C. Vander-Hyde, A. A. van Veggel, S. Vass, R. Vaulin, A. Vecchio, J. Veitch, P. J. Veitch, K. Venkateswara, S. Vinciguerra, D. J. Vine, S. Vitale, T. Vo, C. Vorvick, D. V. Voss, W. D. Vousden, S. P. Vyatchanin, A. R. Wade, L. E. Wade, M. Wade, M. Walker, L. Wallace, S. Walsh, H. Wang, M. Wang, X. Wang, Y. Wang, R. L. Ward, J. Warner, B. Weaver, M. Weinert, A. J. Weinstein, R. Weiss, L. Wen, P. Weßels, T. Westphal, K. Wette, J. T. Whelan, B. F. Whiting, R. D. Williams, A. R. Williamson, J. L. Willis, B. Willke, M. H. Wimmer, W. Winkler, C. C. Wipf, H. Wittel, G. Woan, J. Woehler, J. Worden, J. L. Wright, D. S. Wu, G. Wu, J. Yablon, W. Yam, H. Yamamoto, C. C. Yancey, H. Yu, M. Zanolin, M. Zevin, L. Zhang, M. Zhang, Y. Zhang, C. Zhao, M. Zhou, Z. Zhou, X. J. Zhu, M. E. Zucker, S. E. Zuraw, J. Zweizig and J. H. and, Exploring the sensitivity of next generation gravitational wave detectors, *Classical and Quantum Gravity* **34**, p. 044001 (Jan 2017).

20. M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge, S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik, E. Calloni, G. Cella, E. Chassande Mottin, S. Chelkowski, A. Chincarini, J. Clark, E. Coccia, C. Colacino, J. Colas, A. Cumming, L. Cunningham, E. Cuoco, S. Danilishin, K. Danzmann, G. De Luca, R. De Salvo, T. Dent, R. De Rosa, L. Di Fiore, A. Di Virgilio, M. Doets, V. Fafone, P. Falferi, R. Flaminio, J. Franc, F. Frasconi, A. Freise, P. Fulda, J. Gair, G. Gemme, A. Gennai, A. Giazotto, K. Glampedakis, M. Granata, H. Grote, G. Guidi, G. Hammond, M. Hannam, J. Harms, D. Heinert, M. Hendry, I. Heng, E. Hennes, S. Hild, J. Hough, S. Husa, S. Huttner, G. Jones, F. Khalili, K. Kokeyama, K. Kokkotas, B. Krishnan, M. Lorenzini, H. Lück, E. Majorana, I. Mandel, V. Mandic, I. Martin, C. Michel, Y. Minenkov, N. Morgado, S. Mosca, B. Mours, H. Müller Ebhardt, P. Murray, R. Nawrodt, J. Nelson, R. Oshaughnessy, C. D. Ott, C. Palomba, A. Paoli, G. Parguez, A. Pasqualetti, R. Passaquieti, D. Passuello, L. Pinard, R. Poggiani, P. Popolizio, M. Prato, P. Puppo, D. Rabeling, P. Rapagnani, J. Read, T. Regimbau, H. Rehbein, S. Reid, L. Rezzolla, F. Ricci, F. Richard, A. Rocchi, S. Rowan, A. Rüdiger, B. Sassolas, B. Sathyaprakash, R. Schnabel, C. Schwarz, P. Seidel, A. Sintes, K. Somiya, F. Speirits, K. Strain, S. Strigin, P. Sutton, S. Tarabrin, A. Thüring, J. van den Brand, C. van Leewen, M. van Veggel, C. van den Broeck, A. Vecchio, J. Veitch, F. Vetrano, A. Vicere, S. Vyatchanin, B. Willke, G. Woan, P. Wolfango and K. Yamamoto, The Einstein Telescope: a third-generation gravitational wave observatory, *Class. Quantum Grav.* **27**, p. 194002 (October 2010).

21. S. Hild, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia et al., Sensitivity studies for third-generation gravitational wave observatories, *Classical and Quantum Gravity* **28**, p. 094013 (May 2011).
22. M. Maggiore, C. Van Den Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone and M. Sakellariadou, Science case for the Einstein telescope, *J. Cosmol. Astropart. Phys.* **2020**, p. 050 (March 2020).
23. M. L. Chan, C. Messenger, I. S. Heng and M. Hendry, Binary neutron star mergers and third generation detectors: Localization and early warning, *Phys. Rev. D* **97**, p. 123014 (June 2018).
24. S. Adrián-Martínez, M. Ageron, F. Aharonian, S. Aiello, A. Albert, F. Ameli, E. Anassontzis, M. Andre, G. Androulakis, M. Anghinolfi and et al., Letter of intent for KM3NeT 2.0, *J. Phys. G Nucl. Phys.* **43**, p. 084001 (August 2016).
25. IceCube-Gen2 Collaboration, IceCube-Gen2: A Vision for the Future of Neutrino Astronomy in Antarctica, *arXiv e-prints*, p. arXiv:1412.5106 (December 2014).
26. S. Zhang, A. Santangelo, M. Feroci, Y. Xu, F. Lu, Y. Chen, H. Feng, S. Zhang, S. Brandt, M. Hernanz, L. Baldini, E. Bozzo, R. Campana, A. De Rosa, Y. Dong, Y. Evangelista, V. Karas, N. Meidinger, A. Meuris, K. Nandra, T. Pan, G. Pareschi, P. Orleanski, Q. Huang, S. Schanne, G. Sironi, D. Spiga, J. Svoboda, G. Tagliaferri, C. Tenzer, A. Vacchi, S. Zane, D. Walton, Z. Wang, B. Winter, X. Wu, J. J. M. in 't Zand, M. Ahangarianabhari, G. Ambrosi, F. Ambrosino, M. Barbera, S. Basso, J. Bayer, R. Bellazzini, P. Bellutti, B. Bertucci, G. Bertuccio, G. Borghi, X. Cao, F. Cadoux, R. Campana, F. Ceraudo, T. Chen, Y. Chen, J. Chevenez, M. Civitani, W. Cui, W. Cui, T. Dauser, E. Del Monte, S. Di Cosimo, S. Diebold, V. Doroshenko, M. Dovciak, Y. Du, L. Ducci, Q. Fan, Y. Favre, F. Fuschino, J. L. Gálvez, M. Gao, M. Ge, O. Gevin, M. Grassi, Q. Gu, Y. Gu, D. Han, B. Hong, W. Hu, L. Ji, S. Jia, W. Jiang, T. Kennedy, I. Kreykenbohm, I. Kuvvetli, C. Labanti, L. Latronico, G. Li, M. Li, X. Li, W. Li, Z. Li, O. Limousin, H. Liu, X. Liu, B. Lu, T. Luo, D. Macera, P. Malcovati, A. Martindale, M. Michalska, B. Meng, M. Minuti, A. Morbidini, F. Muleri, S. Paltani, E. Perinati, A. Picciotto, C. Piemonte, J. Qu, A. Rachevski, I. Rashevskaya, J. Rodriguez, T. Schanz, Z. Shen, L. Sheng, J. Song, L. Song, C. Sgro, L. Sun, Y. Tan, P. Uttley, B. Wang, D. Wang, G. Wang, J. Wang, L. Wang, Y. Wang, A. L. Watts, X. Wen, J. Wilms, S. Xiong, J. Yang, S. Yang, Y. Yang, N. Yu, W. Zhang, G. Zampa, N. Zampa, A. A. Zdziarski, A. Zhang, C. Zhang, F. Zhang, L. Zhang, T. Zhang, Y. Zhang, X. Zhang, Z. Zhang, B. Zhao, S. Zheng, Y. Zhou, N. Zorzi and J. F. Zwart, The enhanced X-ray Timing and Polarimetry mission—eXTP, *Science China Physics, Mechanics, and Astronomy* **62**, p. 29502 (February 2019).
27. J. J. M. in 't Zand, E. Bozzo, J. Qu, X.-D. Li, L. Amati, Y. Chen, I. Donnarumma, V. Doroshenko, S. A. Drake, M. Hernanz, P. A. Jenke, T. J. Maccarone, S. Mahmoodifar, D. de Martino, A. De Rosa, E. M. Rossi, A. Rowlinson, G. Sala, G. Stratta, T. M. Tauris, J. Wilms, X. Wu, P. Zhou, I. Agudo, D. Altamirano, J.-L. Atteia, N. A. Andersson, M. C. Baglio, D. R. Ballantyne, A. Baykal, E. Behar, T. Belloni, S. Bhattacharyya, S. Bianchi, A. Bilous, P. Blay, J. Braga, S. Brandt, E. F. Brown, N. Bucciantini, L. Burderi, E. M. Cackett, R. Campana, S. Campana, P. Casella, Y. Cavecchi, F. Chambers, L. Chen, Y.-P. Chen, J. Chenevez, M. Chernyakova, C. Jin, R. Ciolfi, E. Costantini, A. Cumming, A. D'Aì, Z.-G. Dai, F. D'Ammando, M. De Pasquale, N. Degenaar, M. Del Santo, V. D'Elia, T. Di Salvo, G. Doyle, M. Falanga, X. Fan, R. D. Ferdman, M. Feroci, F. Fraschetti, D. K. Galloway, A. F. Gambino, P. Gandhi, M. Ge, B. Gendre, R. Gill, D. Götz, C. Gouiffès, P. Grandi, J. Granot, M. Güdel, A. Heger,

C. O. Heinke, J. Homan, R. Iaria, K. Iwasawa, L. Izzo, L. Ji, P. G. Jonker, J. José, J. S. Kaastra, E. Kalemci, O. Kargaltsev, N. Kawai, L. Keek, S. Komossa, I. Kreykenbohm, L. Kuiper, D. Kunneriath, G. Li, E.-W. Liang, M. Linares, F. Longo, F. Lu, A. A. Lutovinov, D. Malyshev, J. Malzac, A. Manousakis, I. McHardy, M. Mehdipour, Y. Men, M. Méndez, R. P. Mignani, R. Mikusincova, M. C. Miller, G. Miniutti, C. Motch, J. Nättilä, E. Nardini, T. Neubert, P. T. O'Brien, M. Orlandini, J. P. Osborne, L. Pacciani, S. Paltani, M. Paolillo, I. E. Papadakis, B. Paul, A. Pellizzoni, U. Peretz, M. A. Pérez Torres, E. Perinati, C. Prescod-Weinstein, P. Reig, A. Riggio, J. Rodriguez, P. Rodríguez-Gil, P. Romano, A. Różańska, T. Sakamoto, T. Salmi, R. Salvaterra, A. Sanna, A. Santangelo, T. Savolainen, S. Schanne, H. Schatz, L. Shao, A. Shearer, S. N. Shore, B. W. Stappers, T. E. Strohmayer, V. F. Suleimanov, J. Svoboda, F. K. Thielemann, F. Tombesi, D. F. Torres, E. Torresi, S. Turriziani, A. Vacchi, S. Vercellone, J. Vink, J.-M. Wang, J. Wang, A. L. Watts, S. Weng, N. N. Weinberg, P. J. Wheatley, R. Wijnands, T. E. Woods, S. E. Woosley, S. Xiong, Y. Xu, Z. Yan, G. Younes, W. Yu, F. Yuan, L. Zampieri, S. Zane, A. A. Zdziarski, S.-N. Zhang, S. Zhang, S. Zhang, X. Zhang and M. Zingale, Observatory science with eXTP, *Science China Physics, Mechanics, and Astronomy* **62**, p. 29506 (February 2019).

28. A. Santangelo, S. Zane, H. Feng, R. Xu, V. Doroshenko, E. Bozzo, I. Caiazzo, F. Coti Zelati, P. Esposito, D. González-Caniulef, J. Heyl, D. Huppenkothen, G. Israel, Z. Li, L. Lin, R. Mignani, N. Rea, M. Orlandini, R. Taverna, H. Tong, R. Turolla, C. Baglio, F. Bernardini, N. Bucciantini, M. Feroci, F. Fürst, E. Göğüş, C. Güngör, L. Ji, F. Lu, A. Manousakis, S. Mereghetti, R. Mikusincova, B. Paul, C. Prescod-Weinstein, G. Younes, A. Tiengo, Y. Xu, A. Watts, S. Zhang and S.-N. Zhan, Physics and astrophysics of strong magnetic field systems with eXTP, *Science China Physics, Mechanics, and Astronomy* **62**, p. 29505 (February 2019).
29. A. De Rosa, P. Uttley, L. Gou, Y. Liu, C. Bambi, D. Barret, T. Belloni, E. Berti, S. Bianchi, I. Caiazzo, P. Casella, M. Feroci, V. Ferrari, L. Gualtieri, J. Heyl, A. Ingram, V. Karas, F. Lu, B. Luo, G. Matt, S. Motta, J. Neilsen, P. Pani, A. Santangelo, X. Shu, J. Wang, J.-M. Wang, Y. Xue, Y. Xu, W. Yuan, Y. Yuan, S.-N. Zhang, S. Zhang, I. Agudo, L. Amati, N. Andersson, C. Baglio, P. Bakala, A. Baykal, S. Bhattacharyya, I. Bombaci, N. Bucciantini, F. Capitanio, R. Ciolfi, W. K. Cui, F. D'Ammando, T. Dauser, M. Del Santo, B. De Marco, T. Di Salvo, C. Done, M. Dovčiak, A. C. Fabian, M. Falanga, A. F. Gambino, B. Gendre, V. Grinberg, A. Heger, J. Homan, R. Iaria, J. Jiang, C. Jin, E. Koerding, M. Linares, Z. Liu, T. J. Maccarone, J. Malzac, A. Manousakis, F. Marin, A. Marinucci, M. Mehdipour, M. Méndez, S. Migliari, C. Miller, G. Miniutti, E. Nardini, P. T. O'Brien, J. P. Osborne, P. O. Petrucci, A. Possenti, A. Riggio, J. Rodriguez, A. Sanna, L. Shao, M. Sobolewska, E. Sramkova, A. L. Stevens, H. Stiele, G. Stratta, Z. Stuchlik, J. Svoboda, F. Tamburini, T. M. Tauris, F. Tombesi, G. Torok, M. Urbanec, F. Vincent, Q. Wu, F. Yuan, J. J. M. in 't Zand, A. A. Zdziarski and X. Zhou, Accretion in strong field gravity with eXTP, *Science China Physics, Mechanics, and Astronomy* **62**, p. 29504 (February 2019).
30. A. L. Watts, W. Yu, J. Poutanen, S. Zhang, S. Bhattacharyya, S. Bogdanov, L. Ji, A. Patruno, T. E. Riley, P. Bakala, A. Baykal, F. Bernardini, I. Bombaci, E. Brown, Y. Cavecchi, D. Chakrabarty, J. Chenevez, N. Degenaar, M. Del Santo, T. Di Salvo, V. Doroshenko, M. Falanga, R. D. Ferdman, M. Feroci, A. F. Gambino, M. Ge, S. K. Greif, S. Guillot, C. Gungor, D. H. Hartmann, K. Hebeler, A. Heger, J. Homan, R. Iaria, J. i. Zand, O. Kargaltsev, A. Kurkela, X. Lai, A. Li, X. Li, Z. Li, M. Linares, F. Lu, S. Mahmoodifar, M. Méndez, M. Coleman Miller, S. Morsink, J. Nättilä, A. Possenti, C. Prescod-Weinstein, J. Qu, A. Riggio, T. Salmi, A. Sanna, A. Santangelo,

H. Schatz, A. Schwenk, L. Song, E. Šrámková, B. Stappers, H. Stiele, T. Strohmayer, I. Tews, L. Tolos, G. Török, D. Tsang, M. Urbanec, A. Vacchi, R. Xu, Y. Xu, S. Zane, G. Zhang, S. Zhang, W. Zhang, S. Zheng and X. Zhou, Dense matter with eXTP, *Science China Physics, Mechanics, and Astronomy* **62**, p. 29503 (February 2019).

31. O. S. Salafia, G. Ghirlanda, S. Ascenzi and G. Ghisellini, On-axis view of GRB 170817A, *Astron. Astrophys.* **628**, p. A18 (Aug 2019).
32. A. Hajela, R. Margutti, J. S. Bright, K. D. Alexander, B. D. Metzger, V. Nedora, A. Kathirgamaraju, B. Margalit, D. Radice, E. Berger, A. MacFadyen, D. Giannios, R. Chornock, I. Heywood, L. Sironi, O. Gottlieb, D. Coppejans, T. Laskar, Y. Cendes, R. Barniol Duran, T. Eftekhari, W. Fong, A. McDowell, M. Nicholl, X. Xie, J. Zrake, S. Bernuzzi, F. S. Broekgaarden, C. D. Kilpatrick, G. Terreran, V. A. Villar, P. K. Blanchard, S. Gomez, G. Hosseinzadeh, D. J. Matthews and J. C. Rastinejad, The emergence of a new source of X-rays from the binary neutron star merger GW170817, *arXiv e-prints*, p. arXiv:2104.02070 (April 2021).
33. B. D. Metzger and A. L. Piro, Optical and X-ray emission from stable millisecond magnetars formed from the merger of binary neutron stars, *Mon. Not. R. Astron. Soc.* **439**, 3916 (April 2014).
34. R. Ciolfi, W. Kastaun, J. V. Kalinani and B. Giacomazzo, First 100 ms of a long-lived magnetized neutron star formed in a binary neutron star merger, *Phys. Rev. D* **100**, p. 023005 (Jul 2019).
35. K. Arnaud, B. Dorman and C. Gordon, XSPEC: An X-ray spectral fitting package (October 1999).
36. A. Goldstein, P. Veres, E. Burns, M. S. Briggs, R. Hamburg, D. Kocevski, C. A. Wilson-Hodge, R. D. Preece, S. Poolakkil, O. J. Roberts, C. M. Hui, V. Connaughton, J. Racusin, A. von Kienlin, T. Dal Canton, N. Christensen, T. Littenberg, K. Siellez, L. Blackburn, J. Broida, E. Bissaldi, W. H. Cleveland, M. H. Gibby, M. M. Giles, R. M. Kippen, S. McBreen, J. McEnery, C. A. Meegan, W. S. Paciesas and M. Stanbro, An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A, *Astrophys. J. Lett.* **848**, p. L14 (October 2017).
37. B. B. Zhang, B. Zhang, H. Sun, W. H. Lei, H. Gao, Y. Li, L. Shao, Y. Zhao, Y. D. Hu, H. J. Lü, X. F. Wu, X. L. Fan, G. Wang, A. J. Castro-Tirado, S. Zhang, B. Y. Yu, Y. Y. Cao and E. W. Liang, A peculiar low-luminosity short gamma-ray burst from a double neutron star merger progenitor, *Nature Communications* **9**, p. 447 (January 2018).
38. G. Ghirlanda, O. S. Salafia, A. Pescalli, G. Ghisellini, R. Salvaterra, E. Chassande-Mottin, M. Colpi, F. Nappo, P. D'Avanzo, A. Melandri, M. G. Bernardini, M. Branchesi, S. Campana, R. Ciolfi, S. Covino, D. Götz, S. D. Vergani, M. Zennaro and G. Tagliaferri, Short gamma-ray bursts at the dawn of the gravitational wave era, *Astron. Astrophys.* **594**, p. A84 (October 2016).
39. E. Troja, L. Piro, H. van Eerten, R. T. Wollaeger, M. Im, O. D. Fox, N. R. Butler, S. B. Cenko, T. Sakamoto, C. L. Fryer, R. Ricci, A. Lien, R. E. Ryan, O. Korobkin, S.-K. Lee, J. M. Burgess, W. H. Lee, A. M. Watson, C. Choi, S. Covino, P. D'Avanzo, C. J. Fontes, J. B. González, H. G. Khandrika, J. Kim, S.-L. Kim, C.-U. Lee, H. M. Lee, A. Kutyrev, G. Lim, R. Sánchez-Ramírez, S. Veilleux, M. H. Wieringa and Y. Yoon, The X-ray counterpart to the gravitational-wave event GW170817, *Nature* **551**, 71 (November 2017).
40. A. Hajela, R. Margutti, J. Bright, K. D. Alexander, W. Fong, E. Berger, R. Chornock, D. L. Coppejans, P. Blanchard, V. A. Villar, T. Eftekhari, A. Kathirgamaraju, D. Giannios, T. Laskar, J. Zrake, A. MacFadyen and K. Paterson, Continued Chandra observations of GW170817 at 3.3 years since merger, *GRB Coordinates Network* **29375**, p. 1 (January 2021).

41. V. Nedora, D. Radice, S. Bernuzzi, A. Perego, B. Daszuta, A. Endrizzi, A. Prakash and F. Schianchi, Dynamical ejecta synchrotron emission as a possible contributor to the changing behaviour of GRB170817A afterglow, *MNRAS* **506**, 5908 (October 2021).
42. E. Troja, B. O'Connor, G. Ryan, L. Piro, R. Ricci, B. Zhang, T. Piran, G. Bruni, S. B. Cenko and H. van Eerten, Accurate flux calibration of GW170817: is the X-ray counterpart on the rise?, *arXiv e-prints*, p. arXiv:2104.13378 (April 2021).
43. Z. F. Bostancı, Y. Kaneko and E. Göğüş, Gamma-ray bursts with extended emission observed with BATSE, *MNRAS* **428**, 1623 (January 2013).
44. Y. Kaneko, Z. F. Bostancı, E. Göğüş and L. Lin, Short gamma-ray bursts with extended emission observed with Swift/BAT and Fermi/GBM, *MNRAS* **452**, 824 (September 2015).
45. J. P. Norris and J. T. Bonnell, Short Gamma-Ray Bursts with Extended Emission, *ApJ* **643**, 266 (May 2006).
46. A. Lien, T. Sakamoto, S. D. Barthelmy, W. H. Baumgartner, J. K. Cannizzo, K. Chen, N. R. Collins, J. R. Cummings, N. Gehrels, H. A. Krimm, C. B. Markwardt, D. M. Palmer, M. Stamatikos, E. Troja and T. N. Ukwatta, The Third Swift Burst Alert Telescope Gamma-Ray Burst Catalog, *ApJ* **829**, p. 7 (September 2016).
47. P. T. O'Brien, R. Willingale, J. Osborne, M. R. Goad, K. L. Page, S. Vaughan, E. Rol, A. Beardmore, O. Godet, C. P. Hurkett, A. Wells, B. Zhang, S. Kobayashi, D. N. Burrows, J. A. Nousek, J. A. Kennea, A. Falcone, D. Grupe, N. Gehrels, S. Barthelmy, J. Cannizzo, J. Cummings, J. E. Hill, H. Krimm, G. Chincarini, G. Tagliaferri, S. Campana, A. Moretti, P. Giommi, M. Perri, V. Mangano and V. LaParola, The Early X-Ray Emission from GRBs, *ApJ* **647**, 1213 (August 2006).
48. S. Kisaka, K. Ioka and T. Sakamoto, Bimodal Long-lasting Components in Short Gamma-Ray Bursts: Promising Electromagnetic Counterparts to Neutron Star Binary Mergers, *ApJ* **846**, p. 142 (September 2017).
49. B. D. Metzger, D. Giannios, T. A. Thompson, N. Bucciantini and E. Quataert, The protomagnetar model for gamma-ray bursts, *MNRAS* **413**, 2031 (May 2011).
50. N. Bucciantini, J. Arons and E. Amato, Modelling spectral evolution of pulsar wind nebulae inside supernova remnants, *MNRAS* **410**, 381 (January 2011).
51. B. Paczynski, Gamma-ray bursters at cosmological distances, *ApJl* **308**, L43 (September 1986).
52. J. Goodman, Are gamma-ray bursts optically thick?, *ApJl* **308**, p. L47 (September 1986).
53. V. V. Usov, Millisecond pulsars with extremely strong magnetic fields as a cosmological source of γ-ray bursts, *Nature* **357**, 472 (June 1992).
54. C. Thompson, A model of gamma-ray bursts., *MNRAS* **270**, 480 (October 1994).
55. T. Piran, The physics of gamma-ray bursts, *Rev. Mod. Phys.* **76**, 1143 (October 2004).
56. P. Kumar and B. Zhang, The physics of gamma-ray bursts and relativistic jets, *Phys. Rep.* **561**, 1 (February 2015).
57. R. D. Preece, M. S. Briggs, R. S. Mallozzi, G. N. Pendleton, W. S. Paciesas and D. L. Band, The Synchrotron Shock Model Confronts a "Line of Death" in the BATSE Gamma-Ray Burst Data, *ApJl* **506**, L23 (October 1998).
58. G. Ghisellini, A. Celotti and D. Lazzati, Constraints on the emission mechanisms of gamma-ray bursts, *MNRAS* **313**, L1 (March 2000).
59. R. Narayan, B. Paczynski and T. Piran, Gamma-ray bursts as the death throes of massive binary stars, *ApJl* **395**, L83 (August 1992).
60. M. J. Rees and P. Meszaros, Unsteady outflow models for cosmological gamma-ray bursts, *ApJl* **430**, L93 (August 1994).

61. D. Eichler and A. Levinson, A Compact Fireball Model of Gamma-Ray Bursts, *ApJ* **529**, 146 (January 2000).
62. A. Pe'er, P. Mészáros and M. J. Rees, The Observable Effects of a Photospheric Component on GRB and XRF Prompt Emission Spectrum, *ApJ* **642**, 995 (May 2006).
63. R. Barniol Duran and P. Kumar, Adiabatic expansion, early X-ray data and the central engine in GRBs, *MNRAS* **395**, 955 (May 2009).
64. S. Ascenzi, G. Oganesyan, M. Branchesi and R. Ciolfi, Electromagnetic counterparts of compact binary mergers, *J. Plasma Phys.* **87**, p. 845870102 (February 2021).
65. R. Gill, M. Kole and J. Granot, Grb polarization: A unique probe of GRB physics (2021).
66. N. Produit, T. W. Bao, T. Batsch, T. Bernasconi, I. Britvich, F. Cadoux, I. Cernuda, J. Y. Chai, Y. W. Dong, N. Gauvin, W. Hajdas, M. Kole, M. N. Kong, R. Kramert, L. Li, J. T. Liu, X. Liu, R. Marcinkowski, S. Orsi, M. Pohl, D. Rapin, D. Rybka, A. Rutczynska, H. L. Shi, P. Socha, J. C. Sun, L. M. Song, J. Szabelski, I. Traseira, H. L. Xiao, R. J. Wang, X. Wen, B. B. Wu, L. Zhang, L. Y. Zhang, S. N. Zhang, Y. J. Zhang and A. Zwolinska, Design and construction of the POLAR detector, *Nuclear Instruments and Methods in Physics Research A* **877**, 259 (January 2018).
67. S. Vadawale, Prospects of hard X-ray polarimetry with pixillated CZT detectors, in *Astronomical Society of India Conference Series, year = 2013, series = Astronomical Society of India Conference Series*,
68. B. Zhang, Y. Z. Fan, J. Dyks, S. Kobayashi, P. Meszaros, D. N. Burrows, J. A. Nousek and N. Gehrels, Physical processes shaping gamma-ray burst x-ray afterglow light curves: Theoretical implications from the SwiftX-ray telescope observations, *The Astrophysical Journal* **642**, 354 (May 2006).
69. S. Ronchini, G. Oganesyan, M. Branchesi, S. Ascenzi, M. G. Bernardini, F. Brighenti, S. Dall'Osso, P. D'Avanzo, G. Ghirlanda, G. Ghisellini, M. E. Ravasio and O. S. Salafia, Spectral index-flux relation for investigating the origins of steep decay in γ-ray bursts, *Nature Communications* **12**, p. 4040 (January 2021).
70. R. Gill and J. Granot, Temporal evolution of prompt GRB polarization, *Monthly Notices of the Royal Astronomical Society* **504**, 1939 (04 2021).
71. F. Fuschino, R. Campana, C. Labanti, Y. Evangelista, M. Feroci, L. Burderi, F. Fiore, F. Ambrosino, G. Baldazzi, P. Bellutti, R. Bertacin, G. Bertuccio, G. Borghi, D. Cirrincione, D. Cauz, F. Ficorella, M. Fiorini, M. Gandola, M. Grassi, A. Guzman, G. L. Rosa, M. Lavagna, P. Lunghi, P. Malcovati, G. Morgante, B. Negri, G. Pauletta, R. Piazzolla, A. Picciotto, S. Pirrotta, S. Pliego-Caballero, S. Puccetti, A. Rachevski, I. Rashevskaya, L. Rignanese, M. Salatti, A. Santangelo, S. Silvestrini, G. Sottile, C. Tenzer, A. Vacchi, G. Zampa, N. Zampa and N. Zorzi, HERMES: An ultra-wide band X and gamma-ray transient monitor on board a nano-satellite constellation, *Nuclear Instruments and Methods in Physics Research A* **936**, 199 (August 2019).
72. M. Maggiore, C. Van Den Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone and M. Sakellariadou, Science case for the Einstein telescope, *jcap* **2020**, p. 050 (March 2020).
73. B. Paczynski and J. E. Rhoads, Radio Transients from Gamma-Ray Bursters, *ApJl* **418**, p. L5 (November 1993).
74. P. Mészáros and M. J. Rees, Optical and Long-Wavelength Afterglow from Gamma-Ray Bursts, *Astrophys. J.* **476**, 232 (February 1997).
75. R. Sari, T. Piran and R. Narayan, Spectra and Light Curves of Gamma-Ray Burst Afterglows, *ApJl* **497**, L17 (April 1998).

76. S. Covino and D. Gotz, Polarization of prompt and afterglow emission of Gamma-Ray Bursts, *Astronomical and Astrophysical Transactions* **29**, 205 (January 2016).
77. K. Wiersema, S. Covino, K. Toma, A. J. van der Horst, K. Varela, M. Min, J. Greiner, R. L. C. Starling, N. R. Tanvir, R. A. M. J. Wijers, S. Campana, P. A. Curran, Y. Fan, J. P. U. Fynbo, J. Gorosabel, A. Gomboc, D. Götz, J. Hjorth, Z. P. Jin, S. Kobayashi, C. Kouveliotou, C. Mundell, P. T. O'Brien, E. Pian, A. Rowlinson, D. M. Russell, R. Salvaterra, S. di Serego Alighieri, G. Tagliaferri, S. D. Vergani, J. Elliott, C. Fariña, O. E. Hartoog, R. Karjalainen, S. Klose, F. Knust, A. J. Levan, P. Schady, V. Sudilovsky and R. Willingale, Circular polarization in the optical afterglow of GRB 121024A, *Nature* **509**, 201 (May 2014).
78. C. G. Mundell, D. Kopač, D. M. Arnold, I. A. Steele, A. Gomboc, S. Kobayashi, R. M. Harrison, R. J. Smith, C. Guidorzi, F. J. Virgili, A. Melandri and J. Japelj, Highly polarized light from stable ordered magnetic fields in GRB 120308A, *Nature* **504**, 119 (December 2013).
79. A. Corsi, G. W. Hallinan, D. Lazzati, K. P. Mooley, E. J. Murphy, D. A. Frail, D. Carbone, D. L. Kaplan, T. Murphy, S. R. Kulkarni and K. Hotokezaka, An Upper Limit on the Linear Polarization Fraction of the GW170817 Radio Continuum, *ApJl* **861**, p. L10 (July 2018).
80. R. Gill and J. Granot, Constraining the magnetic field structure in collisionless relativistic shocks with a radio afterglow polarization upper limit in GW 170817, *MNRAS* **491**, 5815 (February 2020).
81. P. Beniamini, R. Duque, F. Daigne and R. Mochkovitch, X-ray plateaus in gamma-ray bursts' light curves from jets viewed slightly off-axis, *MNRAS* **492**, 2847 (February 2020).
82. G. Oganesyan, S. Ascenzi, M. Branchesi, O. S. Salafia, S. Dall'Osso and G. Ghirlanda, Structured Jets and X-Ray Plateaus in Gamma-Ray Burst Phenomena, *ApJ* **893**, p. 88 (April 2020).
83. A. Panaitescu, X-Ray Afterglows from the Gamma-Ray Burst "Large-angle" Emission, *ApJ* **895**, p. 39 (May 2020).
84. D. N. Burrows, P. Romano, A. Falcone, S. Kobayashi, B. Zhang, A. Moretti, P. T. O'Brien, M. R. Goad, S. Campana, K. L. Page, L. Angelini, S. Barthelmy, A. P. Beardmore, M. Capalbi, G. Chincarini, J. Cummings, G. Cusumano, D. Fox, P. Giommi, J. E. Hill, J. A. Kennea, H. Krimm, V. Mangano, F. Marshall, P. Mészáros, D. C. Morris, J. A. Nousek, J. P. Osborne, C. Pagani, M. Perri, G. Tagliaferri, A. A. Wells, S. Woosley and N. Gehrels, Bright X-ray Flares in Gamma-Ray Burst Afterglows, *Science* **309**, 1833 (September 2005).
85. Y. Z. Fan, B. Zhang and D. Proga, Linearly Polarized X-Ray Flares following Short Gamma-Ray Bursts, *ApJl* **635**, L129 (December 2005).
86. J. Logue, C. D. Ott, I. S. Heng, P. Kalmus and J. H. C. Scargill, Inferring core-collapse supernova physics with gravitational waves, *Phys. Rev. D* **86**, p. 044023 (August 2012).
87. N. Andersson, J. Baker, K. Belczynski, S. Bernuzzi, E. Berti, L. Cadonati, P. Cerdá-Durán, J. Clark, M. Favata, L. S. Finn, C. Fryer, B. Giacomazzo, J. A. González, M. Hendry, I. S. Heng, S. Hild, N. Johnson-McDaniel, P. Kalmus, S. Klimenko, S. Kobayashi, K. Kokkotas, P. Laguna, L. Lehner, J. Levin, S. Liebling, A. MacFadyen, I. Mandel, S. Marka, Z. Marka, D. Neilsen, P. O'Brien, R. Perna, J. Read, C. Reisswig, C. Rodriguez, M. Ruffert, E. Schnetter, A. Searle, P. Shawhan, D. Shoemaker, A. Soderberg, U. Sperhake, P. Sutton, N. Tanvir, M. Was and S. Whitcomb, The transient gravitational-wave sky, *Classical and Quantum Gravity* **30**, p. 193002 (October 2013).

88. J. Powell and B. Müller, Gravitational wave emission from 3D explosion models of core-collapse supernovae with low and normal explosion energies, *MNRAS* **487**, 1178 (July 2019).
89. D. Guetta, R. Rahin, I. Bartos and M. Della Valle, Constraining the fraction of core-collapse supernovae harbouring choked jets with high-energy neutrinos, *MNRAS* **492**, 843 (February 2020).
90. B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos and et al., Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced LIGO and advanced Virgo, *Phys. Rev. D* **101**, p. 084002 (April 2020).
91. M. Della Valle, Supernovae and gamma-ray bursts connection, *Astronomical and Astrophysical Transactions* **29**, 99 (January 2016).
92. V. D'Elia, S. Campana, A. D'Aì, M. De Pasquale, S. W. K. Emery, D. D. Frederiks, A. Lien, A. Melandri, K. L. Page, R. L. C. Starling, D. N. Burrows, A. A. Breeveld, S. R. Oates, P. T. O'Brien, J. P. Osborne, M. H. Siegel, G. Tagliaferri, P. J. Brown, S. B. Cenko, D. S. Svinkin, A. Tohuvavohu and A. E. Tsvetkova, GRB 171205A/SN 2017iuk: A local low-luminosity gamma-ray burst, *Astronomy and Astrophysics* **619**, p. A66 (November 2018).
93. S. Campana, V. Mangano, A. J. Blustin, P. Brown, D. N. Burrows, G. Chincarini, J. R. Cummings, G. Cusumano, M. Della Valle, D. Malesani, P. Mészáros, J. A. Nousek, M. Page, T. Sakamoto, E. Waxman, B. Zhang, Z. G. Dai, N. Gehrels, S. Immler, F. E. Marshall, K. O. Mason, A. Moretti, P. T. O'Brien, J. P. Osborne, K. L. Page, P. Romano, P. W. A. Roming, G. Tagliaferri, L. R. Cominsky, P. Giommi, O. Godet, J. A. Kennea, H. Krimm, L. Angelini, S. D. Barthelmy, P. T. Boyd, D. M. Palmer, A. A. Wells and N. E. White, The association of GRB 060218 with a supernova and the evolution of the shock wave, *Nature* **442**, 1008 (August 2006).
94. A. M. Soderberg, E. Berger, K. L. Page, P. Schady, J. Parrent, D. Pooley, X. Y. Wang, E. O. Ofek, A. Cucchiara, A. Rau, E. Waxman, J. D. Simon, D. C. J. Bock, P. A. Milne, M. J. Page, J. C. Barentine, S. D. Barthelmy, A. P. Beardmore, M. F. Bietenholz, P. Brown, A. Burrows, D. N. Burrows, G. Byrngelson, S. B. Cenko, P. Chandra, J. R. Cummings, D. B. Fox, A. Gal-Yam, N. Gehrels, S. Immler, M. Kasliwal, A. K. H. Kong, H. A. Krimm, S. R. Kulkarni, T. J. Maccarone, P. Mészáros, E. Nakar, P. T. O'Brien, R. A. Overzier, M. de Pasquale, J. Racusin, N. Rea and D. G. York, An extremely luminous X-ray outburst at the birth of a supernova, *Nature* **453**, 469 (May 2008).

CALET on the ISS: The first 5 years

Pier Simone Marrocchesi* for the CALET Collaboration†

Dept. of Physical Sciences, Earth and Environment, Via Roma 56, I-53100 Siena, Italy
and INFN Sezione di Pisa, Largo Bruno Pontecorvo 3, I-56127 Pisa, Italy
**E-mail: piersimone.marrocchesi@pi.infn.it*

Abstract—The CALorimetric Electron Telescope CALET is collecting science data on the International Space Station since October 2015 with excellent and continuous performance. Energy is measured with a deep homogeneous calorimeter (1.2 nuclear interaction lengths, 27 radiation lengths) preceded by an imaging pre-shower (3 radiation lengths, 1mm granularity) providing tracking and 10^{-5} electron/proton discrimination. Two independent sub-systems identify the charge Z of the incident particle from proton to iron and above (Z<40). CALET measures the cosmic-ray electron+positron flux up to 20 TeV, gamma rays up to 10 TeV, and nuclei up to 1 PeV. In this paper, we report the on-orbit performance of the instrument and summarize the main results obtained during the first 5 years of operation, including the electron+positron energy spectrum and the individual spectra of protons, heavier nuclei and iron. Solar modulation and gamma-ray observations are also concisely reported, as well as transient phenomena and the search for gravitational wave counterparts.

Keywords: Cosmic rays; high energy astrophysics; space-borne experiments.

1. INTRODUCTION

CALET is a cosmic-ray experiment designed for long-term observations of charged and neutral cosmic radiation on the ISS. The instrument is managed by an international collaboration led by the Japanese Space Agency (JAXA) with the participation of the Italian Space Agency (ASI) and NASA. It was launched on August 19, 2015 with the Japanese carrier H-IIB, delivered to the ISS by the HTV-5 Transfer Vehicle, and installed on the Japanese Experiment Module Exposure Facility (JEM-EF). The science program of CALET addresses several outstanding questions of high-energy astroparticle physics including the origin of cosmic rays (CR), the possible presence of nearby astrophysical CR sources, the acceleration and propagation of primary and secondary elements in the galaxy, and the nature of dark matter. The design of CALET is optimised for high precision measurements of the electron+positron spectrum with an accurate scan of the energy interval already covered by previous experiments and its extension to the region above 1 TeV. Given the high energy resolution of CALET for electrons, a detailed study of the spectral shape might reveal the presence of nearby sources of acceleration as well as possible indirect signatures of dark matter.[1,2]

With its individual element resolution in the charge identification of cosmic rays, CALET is also carrying out direct measurements of the spectra and relative

†See the last page for the full authors list.

abundances of light and heavy cosmic nuclei[3,4] from proton to iron, in the energy interval from $\sim$50 GeV (10 GeV/n) for the lighter (heavier) nuclei to several hundred TeV. The abundances of the rare CR trans-iron elements up to Z$\sim$40 are studied with a dedicated program of long term observations.[5]

2. THE CALET INSTRUMENT

The CALET main telescope is an all-calorimetric instrument comprised of three sub-detectors. The CHarge Detector (CHD) is positioned at the top of the apparatus and consists of a two layered hodoscope of plastic scintillators paddles (14 paddles per layer). It performs the charge identification of individual nuclear species, providing a measurement of the charge Z of the incident particle over a wide dynamic range (from Z = 1 up to Z = 40).[6] The IMaging Calorimeter (IMC) is a fine grained sampling calorimeter, segmented longitudinally into 16 layers of scintillating fibers (with 1 mm^2 square cross-section), read out individually and arranged in pairs along orthogonal directions. Each pair is interleaved with thin tungsten absorbers (for a total thickness of 3 X_0). It reconstructs the early shower profile and the impinging particle trajectory with good angular resolution, while providing also a redundant charge measurement.[1] The third detector is the Total AbSorption Calorimeter (TASC), an homogeneous calorimeter with 12 layers of lead-tungstate (PWO) logs arranged in pairs along orthogonal directions. With its 27 X_0 thickness and shower imaging capability, it measures electrons and gamma-rays with an excellent energy resolution, providing high discrimination against hadronic cascades. The total thickness of the main telescope is equivalent to 30 X_0 and 1.3 proton interaction lengths (λ_I), the geometrical factor is 0.12 m^2 sr. A more detailed description of the instrument can be found in Ref. 7 and in the Supplemental Material (SM) of Ref. 9.

3. FLIGHT OPERATIONS AND CALIBRATIONS

The commissioning of CALET aboard the ISS was successfully completed at the beginning of October 2015. Since then, the instrument has been taking science data continuously with no significant interruptions.[2] The on-orbit operations are controlled via the JAXA Ground Support Equipment (JAXA-GSE) in Tsukuba by the Waseda CALET Operations Center (WCOC) at Waseda University, Tokyo.

As of April 30, 2021 a total observation time of more than 2027 days was integrated with a live time fraction $\sim$85% of the total time, and $\sim$2.7 billion events collected above 1 GeV. The exposure with the high-energy (HE) trigger mode, designed to maximize the collection power for electrons above 10 GeV, and other high-energy shower events was $\sim$ 178 m^2 sr day.

Energy calibrations of each channel of CHD, IMC, and TASC is performed with penetrating proton and He particles selected in-flight by a dedicated trigger mode. Raw signals are corrected for light output non-uniformity, gain differences among the channels, position and temperature dependence, as well as temporal

gain variations.[7,8] Correlations among the four gain ranges for each TASC channel are calibrated with flight data, and responses from consecutive ranges are linked together to provide a seamless transition. In this way, a dynamic range spanning more than six orders of magnitude is achieved, allowing observations from one minimum ionizing particle to 1 PeV showers.

4. COSMIC-RAY DIRECT MEASUREMENTS WITH CALET ON THE ISS

4.1. *The Electron Spectrum*

The CALET collaboration reported their first measurement of the inclusive electron+positron spectrum in the energy range from 10 GeV to 3 TeV (Ref. 9) within a fiducial subset of the acceptance. Soon after, the DArk Matter Particle Explorer (DAMPE) collaboration published their all-electron spectrum in the energy interval from 25 GeV to 4.6 TeV (Ref. 10).

The latter publication was followed by a number of papers speculating about the origin of a possible peak-like structure near 1.4 TeV in DAMPE data. An updated version of the CALET all-electron spectrum was published, covering the energy

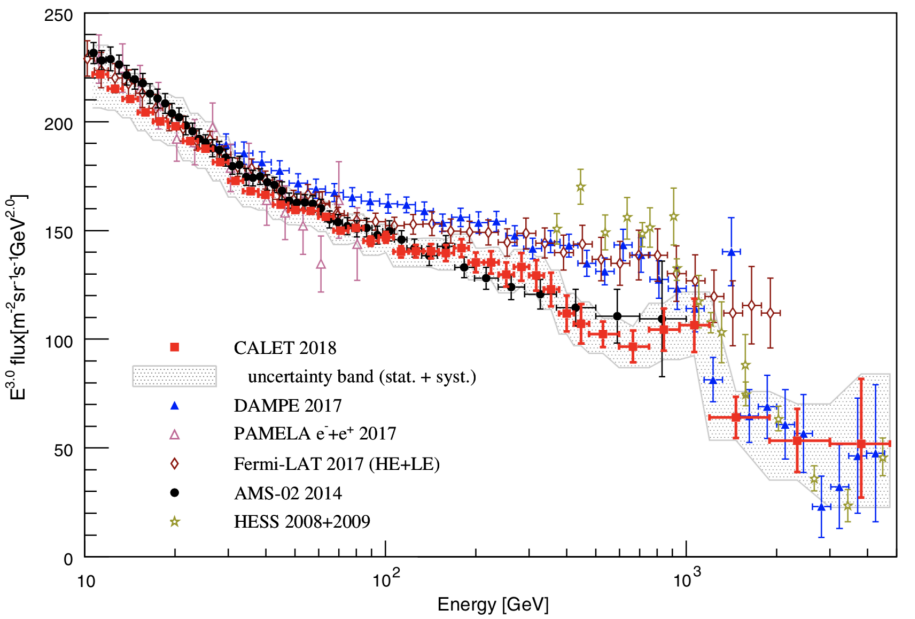


Fig. 1. Direct measurements of the electron + positron flux by space-borne experiments including,[10,11 12–14] and from ground-based experiments.[15,16] The CALET 2018 data[11] are shown as red filled circles in the energy interval 11 GeV to 4.8 TeV. The width of each bin is shown as a horizontal bar, statistical errors as vertical bars. The gray band indicates the quadratic sum of statistical and systematic errors (not including the uncertainty on the energy scale).

range from 11 GeV to 4.8 TeV (Ref. 11) with 780 days of flight data and the full geometrical acceptance. It reported a new analysis with doubled statistics at $E>$ 475 GeV and included one additional energy bin between 3 and 4.8 TeV (Figure 1). The width of each bin is shown as a horizontal bar, the statistical errors as vertical bars, while the gray band is representative of the quadratic sum of statistic and systematic errors. A comprehensive study of the systematic uncertainties was performed as described in Refs. 9, 11 and Supplemental Material therein. A constant electron identification efficiency of 70% was achieved above 30 GeV, with a proton contamination level of 2–5% below 1 TeV and ∼10–20% above.

Taking the currently available experimental data at face-value, we notice that the all-electron spectrum data seem to fork into two groups of measurements: AMS-02 + CALET and Fermi/LAT + DAMPE, with good consistency within each group, but with only marginal overlap between the two, possibly indicating the presence of unknown systematic errors. The CALET spectrum is consistent with AMS-02 below ∼1TeV where both experiments have a good electron identification capability albeit using different detection techniques. CALET observation of a flux suppression above ∼1TeV is consistent with DAMPE within errors. No peak-like structure was found at 1.4 TeV in CALET data, irrespective of the energy binning. After re-binning with the same set of energy bins as DAMPE, an inconsistency between the two measurements emerged[11] with a 4 σ significance, the latter including the systematic errors quoted by both experiments.

New results on the analysis of electrons based on the first five years of CALET observations will be presented at the upcoming ICRC2021 conference.

4.2. *The Proton Spectrum*

Cosmic-ray energies from the GeV scale to the multi-TeV region have been explored – in separate subranges – by magnetic spectrometers (e.g., BESS-TeV, PAMELA, and AMS-02), calorimeters (e.g., ATIC, CREAM, NUCLEON, and DAMPE) and Cherenkov / Transition Radiation instruments (e.g., TRACER). In the intermediate energy region from 200 GeV to 800 GeV a deviation from a single power-law (SPL) was observed in both proton and helium spectra by CREAM,[17–19] PAMELA[20,21] and confirmed with high statistics measurements by AMS-02,[22] CALET,[23] and DAMPE.[24]

The first proton paper published by CALET[23] reported a proton flux measurement where, for the first time, a single space-borne instrument was able to cover the whole interval of proton energies from 50 GeV to 10 TeV thanks to its large dynamic range. The proton flux was extracted from the data collected from October 13, 2015 to August 31, 2018 (1054 days) on the ISS using only 40% of the total acceptance. A detailed study of the systematic uncertainties was reported in the same paper and in the Supplementary Material therein.[23] This is of particular relevance because CR flux measurements are well known to be affected by relatively large systematic errors, often specific of each instrument. CALET proton data (Fig. 2)

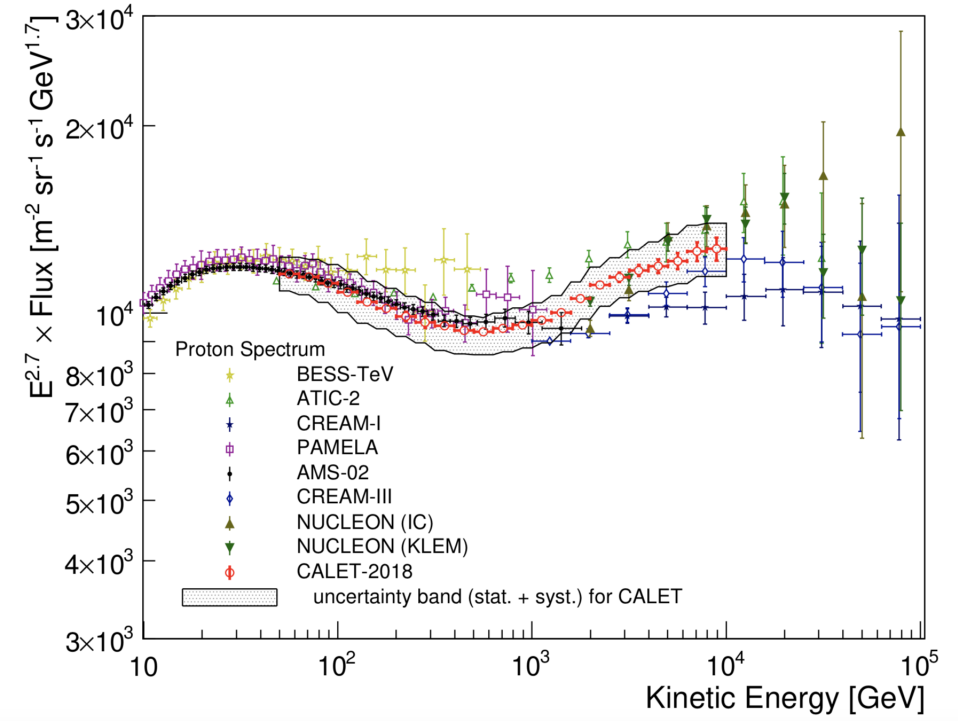


Fig. 2. Cosmic-ray proton spectrum measured by CALET from 50 GeV to 10 TeV published in Ref. 23. The gray band indicates the quadratic sum of statistical and systematic errors.

are consistent with AMS-02 but extend to higher energies by nearly one order of magnitude, showing a very smooth transition of the spectral index from -2.81 ± 0.03 (in the region 50–500 GeV) to -2.56 ± 0.04 (in 1–10 TeV), thereby confirming the existence of a spectral hardening and providing evidence of a deviation from a single power law by more than 3σ.

An update of CALET proton analysis, based on 5 years of data on the ISS, will be presented at the upcoming ICRC2021 conference together with preliminary results on the helium flux.

4.3. *The Spectra of Heavier Nuclei*

The observation of a spectral hardening in proton and helium, as well as in carbon and oxygen spectra,[18,20,22,25,26] have opened a new and unexpected scenario in CR phenomenology. In particular, the high statistics observations by AMS-02, up to a maximum detectable rigidity (MDR) of a few TV, clearly show that primary elements have a very similar rigidity dependence above ∼60 GV and that secondary elements (like Li, Be and B) also show a flux hardening, though with subtle differences that might be attributed to propagation effects (secondaries propagate first

as primaries and then as secondaries). Therefore, it is very important to extend the presently available measurements to the multi-TeV region and investigate the energy dependence of the spectral index for individual nuclear species with high accuracy. CALET is carrying out extensive measurements of the energy spectra, relative abundances and secondary-to-primary ratios of cosmic-ray nuclei.

Preliminary CALET results on the B/C ratio and on the spectra of heavier nuclei from neon to iron (Fig. 3) were previously reported (see for instance Refs. 4, 27, 28). In the following we will focus on the CALET published spectra of C, O and Fe.

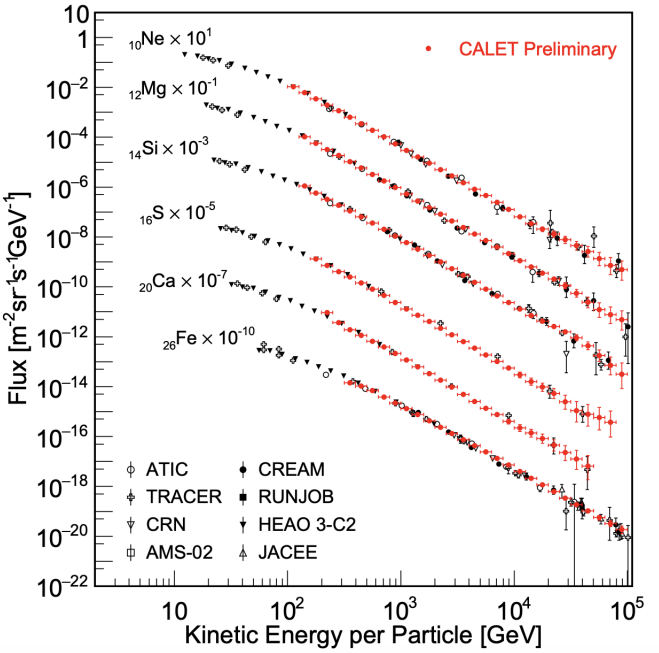


Fig. 3. Preliminary results of energy spectra of heavy primary components of Ne, Mg, Si, S, Ca and Fe as a function of energy par particle compared with previous observations. The Error bars of CALET data[28] represent the statistical uncertainty only.

4.4. *Carbon and Oxygen Spectra*

The energy spectra of carbon and oxygen and their flux ratio were measured by CALET in the energy range from 10 GeV/n to 2.2 TeV/n and published in Ref. 29. CALET observations (Fig. 4) allow to exclude a single power law spectrum for C and O at the level of more than 3σ. A spectral index increase (spectrum flattening) $\Delta\gamma = 0.166 \pm 0.042$ (carbon) and $\Delta\gamma = 0.158 \pm 0.053$ (oxygen) were measured above 200 GeV/n, respectively. The fluxes of C and O were found to share the same energy dependence with a constant C/O flux ratio 0.911 ± 0.006 above 25 GeV/n.

While the above results are consistent with the ones reported by AMS-02 for the same elements, the absolute normalization of CALET data is significantly lower than AMS-02, but in agreement with previous experiments (including PAMELA for carbon). For more details please refer to Ref. 29 and the Supplementary Material therein.

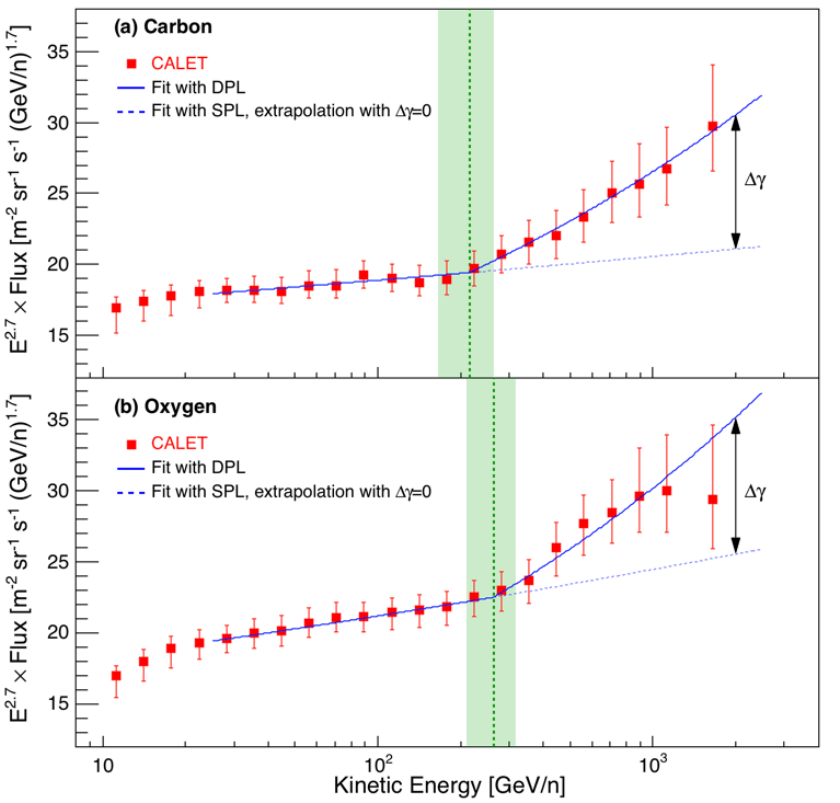


Fig. 4. Fit of the CALET (a) C and (b) O energy spectra[29] with a Double Power Law (DPL) function (blue line) in the energy range [25, 2000] GeV/n. The flux is multiplied by $E^{2.7}$ where E is the kinetic energy per nucleon. Error bars are the sum in quadrature of statistical and systematic uncertainties. The dashed blue lines represent the extrapolation of an Single Power Law function fitted to the data in the energy range [25, 200] GeV/n. $\Delta\gamma$ is the spectral index change above the transition energy E_0 (vertical green dashed line). The green band shows the uncertainty error on E_0 from the DPL fit.

4.5. *The Iron Spectrum*

In a recent paper,[30] the CALET collaboration reported their first measurement of the energy spectrum of cosmic-ray iron from 10 GeV/n to 2.0 TeV/n.

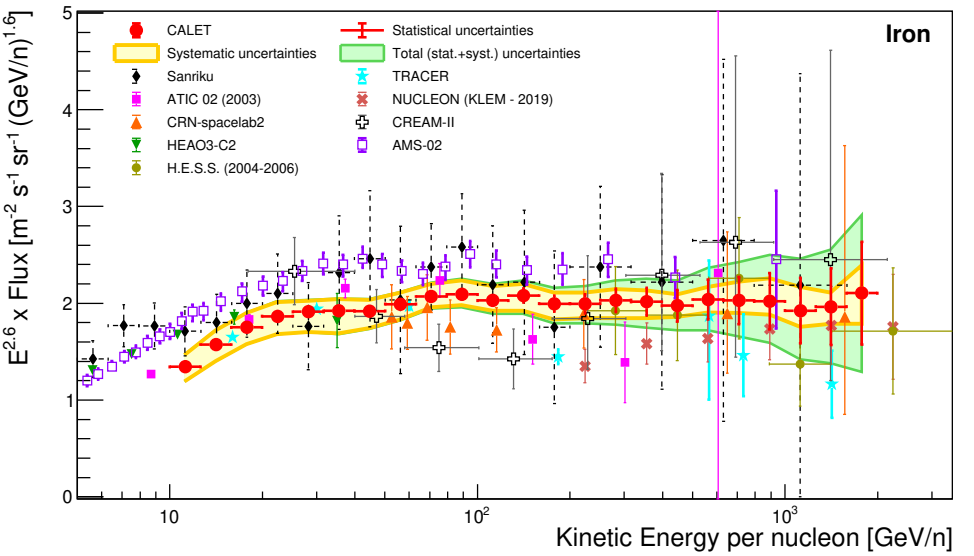


Fig. 5. CALET iron flux[30] as a function of kinetic energy per nucleon in GeV (with multiplicative factor $E^{2.6}$). The error bars of the CALET data (red filled circles) represent the statistical uncertainty only. The yellow band indicates the quadrature sum of systematic errors, while the green band indicates the quadrature sum of statistical and systematic errors. Also plotted are the data points from other direct measurements.[31–39]

The analysis is based on 4.4 years of observations and the measurement achieves a significantly better precision than most of the existing measurements of the same element. The CALET iron differential spectrum in kinetic energy per nucleon is shown in Fig. 5, where uncertainties including statistical and systematic errors are bounded within a green band. The spectrum is compared with the results from space-based (HEAO3-C2,[31] CRN,[32] AMS 02,[33] NUCLEON[34]) and balloon-borne experiments (ATIC-02,[35] TRACER,[36] CREAM-II,[37] Sanriku[38]), as well as ground-based observations (H.E.S.S.[39]). The CALET spectrum is consistent with ATIC 02 and TRACER at low energy and with CRN and HESS at high energy. CALET and NUCLEON iron spectra have similar shapes, while they differ in the absolute normalization of the flux. The latter turns out to be higher for CALET than for CRN by ∼10% on average, while it is lower by 14% with respect to Sanriku. CALET and AMS-02 iron spectra have a very similar shape (Fig. S12 of the Supplemental Material of Ref. 30), but differ in the absolute normalization of the flux by ∼ 20%.

Taking into account the average size of the large systematic errors reported in the literature, CALET data turn out to be consistent with previous measurements within the uncertainty error band, both in spectral shape and normalization. Below 50 GeV/n the spectral shape is found to be similar to the one observed for primaries

lighter than iron. Above the same energy, CALET observations are consistent with the hypothesis of an SPL spectrum up to 2 TeV/n, i.e., the flattening observed above a few hundred GeV/nucleon in the p-C-O spectra does not appear to be present in the iron spectrum in the sub-TeV region. Beyond this limit, the uncertainties given by the available statistics and large systematics do not allow yet to draw a significant conclusion on a possible deviation from a single power law. An SPL fit in this region yields a spectral index value $\gamma = -2.60 \pm 0.03$.

4.6. *The Observation of Gamma-Rays*

CALET can identify gamma-rays and measure their energies from ∼1 GeV to the TeV region. Both CHD and the first IMC layers are used in the offline analysis as anti-coincidence against incoming charged particles, taking advantage of the high granularity of the IMC.

Gamma-ray candidates are also required to deposit more energy in the bottom IMC layers than in the upper ones where pair conversion takes place. In addition to the HE trigger, CALET implements a LE-γ trigger extending the sensitivity to gamma rays with primary energies down to ∼1 GeV. This dedicated trigger is activated only at low geomagnetic latitudes (to avoid an increase of the dead-time) and it is also enabled whenever a gamma-ray burst is triggered by the Calet Gamma-Ray Burst Monitor (CGBM).[40] The first two years of data allowed a complete characterization of the performance of CALET as a gamma-ray instrument, the optimization of the event selection criteria, the determination of the effective area, Point Spread Function (PSF) and absolute pointing accuracy. Measured signals from gamma-ray bright point sources and diffuse galactic emission were found to be in agreement with simulated results and expectations from Fermi-LAT data.[41] The spectra from sources like Crab, Geminga, and Vela pulsars were measured by CALET and tested for consistency with parameterised LAT spectra. These results confirmed the sensitivity of the calorimeter in observing bright, persistent sources.[42] The gamma-ray sky observed by CALET using the LE-γ trigger is shown in Fig. 6.

CALET can also detect gamma-ray transients by means of the CGBM operating in the energy range of 7 keV–20 MeV. As of April 2021, 246 GRBs have been detected, 12% of which were classified as short, with an average rate of ∼44.6 /year.

A search for electromagnetic counterparts of gravitational waves (GW) triggered by LIGO/Virgo was performed with a combined analysis of the CGBM and the calorimeter. Candidate signals compatible with gamma-ray emission were searched for in time intervals of tens of seconds centered on the reported trigger times of GW151226, GW170104, GW170608, GW170814, and GW170817 events. No signal was detected for all GW events; upper limits on gamma-ray emission were set for GW151226 (CAL + CGBM) and GW170104 (CAL), while GW170608, GW170814, GW170817 turned out to be outside the CALET field-of-view.[43,44]

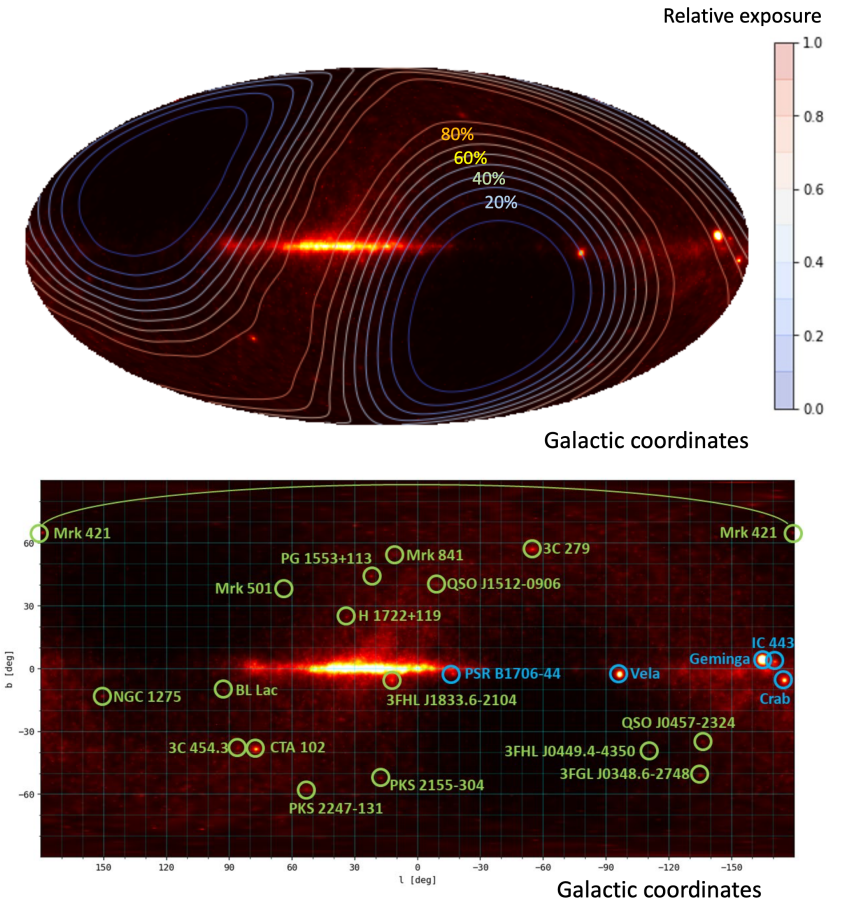


Fig. 6. Top: CALET gamma-ray sky map (observation period: 2015.11.01-2020.07.31) with the LE-γ trigger (E >1 GeV), shown in a Mollweide projection of galactic coordinates. White contours show the relative level of exposure compared to the maximum on the sky; bottom: point sources observed during the same period (with >1 GeV), including the Crab, Geminga, and Vela pulsars.

5. SUMMARY AND PERSPECTIVES

CALET was successfully launched on Aug. 19, 2015. The instrument performance has been very stable during all the scientific observation period from Oct. 13, 2015. CALET measurements of the electron spectrum were published in two papers,[9,11] the latter with improved statistics and extended energy range from 11 GeV to 4.8 TeV. The extension to five years of CALET on-orbit operations provided an increase of the available statistics in the electron observations by a factor $\sim$3 thereby contributing to a better understanding of the detector and of the systematic errors. A search for possible spectral footprints of nearby electron sources in the region above $\sim$1 TeV is in progress.

The wide dynamic range and excellent charge identification capability allow CALET to measure nuclei in cosmic rays from proton to iron and above, with an energy reach approaching the PeV scale. The proton spectrum was published up to 10 TeV (Ref. 23); C and O spectra to 2.2 TeV/n (Ref. 29), and Fe to 2.0 TeV/n (Ref. 30). The spectral index dependence on energy confirmed a spectral hardening for p, C, O with a smooth onset at a few hundred GeV. Measurements of the energy spectra and composition of all primary and secondary nuclei (and of their ratios) are ongoing. The relative abundance of the ultra heavy nuclei up to $Z = 40$ has also been preliminarily analyzed.[5]

The performance of the gamma-ray measurements has confirmed CALET's capability to observe the diffuse component and bright point-sources in the gamma-ray sky from ∼1 GeV to 100 GeV and above (Fig. 6). The continuous GeV gamma-ray sky observation with CALET complements the coverage by other missions and may help to identify unexplored high-energy emissions from future transient events. The latter phenomena are studied with the CGBM.

Follow-up observations were carried out in the X-ray and gamma-ray band of GW events during LIGO/Virgo observation campaigns.[43,44]

Solar modulation is constantly monitored and studied. Since the start of observations in 2015/10, a steady increase in the 1-10 GeV all-electron flux has been observed to present. In the past two years, the flux has reached the maximum flux observed with PAMELA during the previous solar minimum period.[45] Solar energetic particles (SEP) are also studied at high geomagnetic latitudes.

High statistics detection of MeV electrons originating from the radiation belt allows the study of relativistic electron precipitation.[46] This is one of the topics of Space Weather studies[47] which were added as additional observational targets for CALET after the start of on-orbit operations.

Important updates of CALET electron and proton analyses, as well as preliminary results on He, B, B/C analyses, will be presented at the upcoming ICRC2021 conference (highlights in Ref. 48).

The so far excellent performance of the instrument and the outstanding quality of the data suggest that a long-term strategy of CALET observations will contribute to a deeper understanding of cosmic-ray phenomena. CALET operations on the ISS have been recently approved for an extension to the end of 2024 (at least).

ACKNOWLEDGMENTS

We gratefully acknowledge JAXA's contributions to the development of CALET and to the operations aboard the JEM-EF on the International Space Station. We also wish to express our sincere gratitude to Agenzia Spaziale Italiana (ASI) and NASA for their support of the CALET project. This work was supported in part by JSPS Grant-in-Aid for Scientific Research (S) Number 26220708 and 19H05608, JSPS Grant-in-Aid for Scientific Research (B) Number 17H02901, JSPS Grant-in-Aid for Research Activity Start-up No.20K22352 and by the MEXT-Supported

Program for the Strategic Research Foundation at Private Universities (2011–2015) (No.S1101021) at Waseda University. The CALET effort in Italy is supported by ASI under agreement 2013-018-R.0 and its amendments. The CALET effort in the United States is supported by NASA through Grants No. NNX16AB99G, No. NNX16AC02G, and No. NNH14ZDA001N-APRA-0075.

References

1. S. Torii (for the CALET Collab.), PoS (ICRC2019) 142 (2019).
2. Y. Asaoka (for the CALET Collab.), PoS (ICRC2019) 001 (2019).
3. S. Torii and P. S. Marrocchesi (for the CALET Collab.), Adv. Sp. Res. **64**(12), 2531 (2019).
4. Y. Akaike (for the CALET Collab.), PoS (ICRC2019) 034 (2019).
5. B. F. Rauch (for the CALET Collab.), PoS (ICRC2019) 130 (2019).
6. P. S. Marrocchesi, O. Adriani, Y. Akaike, M. G. Bagliesi, A. Basti, G. Bigongiari, S. Bonechi, M. Bongi, M. Y. Kim, T. Lomtadze, P. Maestro, T. Niita, S. Ozawa, Y. Shimizu, and S. Torii, Nucl. Instrum. Methods A **659**, 477 (2011).
7. Y. Asaoka, Y. Akaike, Y. Komiya, R. Miyata, S. Torii, O. Adriani, K. Asano, M. G. Bagliesi, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, et al. (CALET Collab.), Astropart. Phys. **91**, 1 (2017).
8. Y. Asaoka, S. Ozawa, S. Torii, O. Adriani, Y. Akaike, K. Asano, M. G. Bagliesi, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, et al. (CALET Collab.), Astropart. Phys. **100**, 29 (2018).
9. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, G. Collazuol, et al. (CALET Collab.), Phys. Rev. Lett. **119**, 181101 (2017).
10. G. Ambrosi, Q. An, R. Asfandiyarov, P. Azzarello, P. Bernardini, B. Bertucci, M. S. Cai, J. Chang, D. Y. Chen, H. F. Chen, J. L. Chen, W. Chen, M. Y. Cui, T. S. Cui, A. D'Amone, A. De Benedittis, et al. (DAMPE Collab.), Nature **552**, 63 (2017).
11. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, E. Berti, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, et al. (CALET Collab.), Phys. Rev. Lett. **120**, 261102 (2018).
12. O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, M. Bongi, V. Bonvicini, S. Bottai, A. Bruno, F. Cafagna, D. Campana, P. Carlson, M. Casolino, G. Castellini, C. De Santis, et al. (PAMELA Collab.), Riv. Nuovo Cimento **40**, 473 (2017).
13. S. Abdollahi et al. (The Fermi-LAT Collab.), Phys. Rev. D **95**, 082007 (2017).
14. M. Aguilar, D. Aisa, A. Alvino, G. Ambrosi, K. Andeen, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, A. Bartoloni, L. Basara, M. Battarbee, R. Battiston, et al. (AMS Collab.), Phys. Rev. Lett. **113**, 221102 (2014).
15. F. Aharonian, A. G. Akhperjanian, U. Barres de Almeida, A. R. Bazer-Bachi, Y. Becherini, B. Behera, W. Benbow, K. Bernlöhr, C. Boisson, A. Bochow, V. Borrel, I. Braun, E. Brion, J. Brucker, P. Brun, R. Bühler, et al. (H.E.S.S. Collab.), Phys. Rev. Lett. **101**, 261104, (2008).
16. F. Aharonian, A. G. Akhperjanian, G. Anton, U. Barres de Almeida, A. R. Bazer-Bachi, Y. Becherini, B. Behera, K. Bernlöhr, A. Bochow, C. Boisson, J. Bolmont, V. Borrel, J. Brucker, F. Brun, P. Brun, R. Bühler, et al. (H.E.S.S. Collab.), Astron. Astrophys. **508**, 561 (2009).

17. Y. S. Yoon, T. Anderson, A. Barrau, N. B. Conklin, S. Coutu, L. Derome, J. H. Han, J. A. Jeon, K. C. Kim, M. H. Kim, H. Y. Lee, J. Lee, M. H. Lee, S. E. Lee, J. T. Link, A. Menchaca-Rocha, et al., Astrophys. J. **839**, 5 (2017).
18. H. S. Ahn, P. Allison, M. G. Bagliesi, J. J. Beatty, G. Bigongiari, J. T. Childers, N. B. Conklin, S. Coutu, M. A. DuVernois, O. Ganel, J. H. Han, J. A. Jeon, K. C. Kim, M. H. Lee, L. Lutz, P. Maestro, et al., Astrophys. J. Lett. **714**, L89 (2010).
19. Y. S. Yoon, H. S. Ahn, P. S. Allison, M. G. Bagliesi, J. J. Beatty, G. Bigongiari, P. J. Boyle, J. T. Childers, N. B. Conklin, S. Coutu, M. A. DuVernois, O. Ganel, J. H. Han, J. A. Jeon, K. C. Kim, M. H. Lee, et al., Astrophys. J. **728**, 122 (2011).
20. O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, L. Bonechi, M. Bongi, V. Bonvicini, S. Borisov, S. Bottai, A. Bruno, F. Cafagna, D. Campana, R. Carbone, P. Carlson, et al. (PAMELA Collab.), Science **332**, 69 (2011).
21. O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, M. Bongi, V. Bonvicini, S. Borisov, S. Bottai, A. Bruno, F. Cafagna, D. Campana, R. Carbone, P. Carlson, M. Casolino, et al. (PAMELA Collab.), Astrophys. J. **765**, 91 (2013).
22. M. Aguilar, D. Aisa, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, A. Bartoloni, L. Basara, M. Battarbee, et al. (AMS Collab.), Phys. Rev. Lett. **114**, 171103 (2015).
23. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, E. Berti, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, A. Bruno, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, et al. (CALET Collab.), Phys. Rev. Lett. **122**, 181102 (2019).
24. Q. An, R. Asfandiyarov, P. Azzarello, P. Bernardini, X. J. Bi, M. S. Cai, J. Chang, D. Y. Chen, H. F. Chen, J. L. Chen, W. Chen, M. Y. Cui, T. S. Cui, H. T. Dai, A. D'Amone, A. De Benedittis, et al. (DAMPE Collab.), Sci. Adv. **5** eaax3793 (2019).
25. M. Aguilar, D. Aisa, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, A. Bartoloni, L. Basara, M. Battarbee, et al. (AMS Collab.), Phys. Rev. Lett. **115**, 211101 (2015).
26. M. Aguilar, L. Ali Cavasonza, B. Alpat, G. Ambrosi, L. Arruda, N. Attig, S. Aupetit, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, A. Bartoloni, L. Basara, S. Başeğmez-du Pree, M. Battarbee, et al. (AMS Collab.), Phys. Rev. Lett. **119**, 251101 (2017).
27. P. Maestro, P. S Marrocchesi, G. Bigongiari, and P. Brogi, Adv. Space Res. 64, 2538 (2019).
28. Y. Akaike (for the CALET Collab.), J. Phys.: Conf. Ser. **1181**, 012042 (2019).
29. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, E. Berti, G. Bigongiari, W. R. Binns, M. Bongi, P. Brogi, A. Bruno, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, et al. (CALET Collab.), Phys. Rev. Lett. **125**, 251102 (2020).
30. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, E. Berti, G. Bigongiari, W. R. Binns, M. Bongi, P. Brogi, A. Bruno, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, G. Collazuol, et al. (CALET Collab.), Phys. Rev. Lett. **126**, 241101 (2021).
31. J. J. Engelmann, P. Ferrando, A. Soutoul, P. Goret, E. Juliusson, L. Koch-Miramond, N. Lund, P. Masse, B. Peters, N. Petrou, and I. L. Rasmussen, Astron. Astrophys. **233**, 96 (1990).
32. D. Müller, S. P. Swordy, P. Meyer, J. L'Heureux, and J. M. Grunsfeld, Astrophys. J. 374, **356** (1991).
33. M. Aguilar, L. Ali Cavasonza, M. S. Allen, B. Alpat, G. Ambrosi, L. Arruda, N. Attig, F. Barao, L. Barrin, A. Bartoloni, S. Başeğmez-du Pree, R. Battiston, M. Behlmann,

B. Beischer, J. Berdugo, B. Bertucci, et al. (AMS Collab.), Phys. Rev. Lett. **126**, 041104 (2021).
34. V. Grebenyuk, D. Karmanov, I. Kovalev, I. Kudryashov, A. Kurganov, A. Panov, D. Podorozhny, A. Tkachenko, L. Tkachev, A. Turundaevskiy, O. Vasiliev, A. Voronin (NUCLEON Collaboration), Adv. Space Res. **64**, 2546 (2019).
35. A. D. Panov, J. H. Adams, H. S. Ahn, G. L. Bashinzhagyan, J. W. Watts, J. P. Wefel, J. Wu, O. Ganel, T. G. Guzik, V. I. Zatsepin, I. Isbert, K. C. Kim, M. Christl, E. N. Kouznetsov, M. I. Panasyuk, E. S. Seo, et al. (ATIC Collaboration), Bull. Russ. Acad. Sci. **73**, 564 (2009).
36. M. Ave, P. J. Boyle, F. Gahbauer, C. Höppner, J. R. Hörandel, M. Ichimura, D. Müller, and A. Romero-Wolf (TRACER Collab.), Astrophys. J. **678**, 262 (2008).
37. H. S. Ahn, P. Allison, M. G. Bagliesi, L. Barbier, J. J. Beatty, G. Bigongiari, T. J. Brandt, J. T. Childers, N. B. Conklin, S. Coutu, M. A. DuVernois, O. Ganel, J. H. Han, J. A. Jeon, K. C. Kim, M. H. Lee, et al. (CREAM Collab.), Astrophys. J. **707**, 593 (2009).
38. M. Ichimura, M. Kogawa, S. Kuramata, H. Mito, T. Murabayashi, H. Nanjo, T. Nakamura, K. Ohba, T. Ohuchi, T. Ozawa, Y. Yamada, H. Matsutani, Z. Watanabe, E. Kamioka, K. Kirii, M. Kitazawa, et al., Phys. Rev. D **48**, 1949 (1993).
39. F. Aharonian, A. G. Akhperjanian, A. R. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge, K. Bernlöhr, C. Boisson, O. Bolz, V. Borrel, I. Braun, E. Brion, A. M. Brown, R. Bühler, I. Büsching, S. Carrigan, et al. (H.E.S.S. Collab.), Phys. Rev. D **75**, 042004 (2007).
40. K. Yamaoka, A. Yoshida, T. Sakamoto, I. Takahashi, T. Hara, T. Yamamoto, Y. Kawakubo, R. Inoue, S. Terazawa, R. Fujioka, K. Senuma, S. Nakahira, H. Tomida, S. Ueno, S. Torii, M. L. Cherry, S. Ricciarini (CALET Collab.), in *Proceedings of the 7th Huntsville Gamma-Ray Burst Symposium, GRB 2013, eConf C1304143, 41 (2013)*.
41. N. Cannady, Y. Asaoka, F. Satoh, M. Tanaka, S. Torii, M. L. Cherry, M. Mori, O. Adriani, Y. Akaike, K. Asano, M. G. Bagliesi, E. Berti, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, et al. (CALET Collab.), Astrophys. J. Suppl. S. **238**, 5 (2018).
42. M. Mori et al. (for the CALET Collab.), PoS (ICRC2019) 586 (2019).
43. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, E. Berti, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, et al. (CALET Collab.), Astrophys. J. Lett. **863**, 160 (2018).
44. O. Adriani, Y. Akaike, K. Asano, Y. Asaoka, M. G. Bagliesi, G. Bigongiari, W. R. Binns, S. Bonechi, M. Bongi, P. Brogi, J. H. Buckley, N. Cannady, G. Castellini, C. Checchia, M. L. Cherry, G. Collazuol, et al. (CALET Collab.), Astrophys. J. Lett. **829**, L20 (2016).
45. S. Miyake et al. (for the CALET Collab.), PoS (ICRC2019) 1126 (2019).
46. R. Kataoka, Y. Asaoka, S. Torii, T. Terasawa, S. Ozawa, T. Tamura, Y. Shimizu, Y. Akaike, M. Mori (CALET Collab.), Geophys. Res. Lett. **43**, 4119 (2016).
47. A. Bruno et al. (for the CALET Collab.), PoS (ICRC2019) 1063 (2019).
48. P. S. Marrocchesi et al. (for the CALET Collab.), PoS (ICRC2021) 010 (2021).

Full Authors List: CALET Collaboration

O. Adriani[1,2], Y. Akaike[3,4], K. Asano[5], Y. Asaoka[5], E. Berti[1,2], G. Bigongiari[6,7], W. R. Binns[8], M. Bongi[1,2], P. Brogi[6,7], A. Bruno[9,10], J. H. Buckley[8], N. Cannady[11,12,13], G. Castellini[14], C. Checchia[6], M. L. Cherry[15], G. Collazuol[16,17], K. Ebisawa[18], A. W. Ficklin[15], H. Fuke[18], S. Gonzi[1,2], T. G. Guzik[15], T. Hams[11], K. Hibino[19], M. Ichimura[20], K. Ioka[21], W. Ishizaki[5], M. H. Israel[8], K. Kasahara[22], J. Kataoka[23], R. Kataoka[24], Y. Katayose[25], C. Kato[26], N. Kawanaka[27,28], Y. Kawakubo[15], K. Kobayashi[3,4], K. Kohri[29], H. S. Krawczynski[8], J. F. Krizmanic[11,12,13], P. Maestro[6,7], P. S. Marrocchesi[6,7], A. M. Messineo[30,7], J.W. Mitchell[12], S. Miyake[32], A. A. Moiseev[33,12,13], M. Mori[34], N. Mori[2], H. M. Motz[35], K. Munakata[26], S. Nakahira[18], J. Nishimura[18], G. A. de Nolfo[9], S. Okuno[19], J. F. Ormes[36], N. Ospina[16,17], S. Ozawa[37], L. Pacini[1,14,2], P. Papini[2], B. F. Rauch[8], S. B. Ricciarini[14,2], K. Sakai[11,12,13], T. Sakamoto[38], M. Sasaki[33,12,13], Y. Shimizu[19], A. Shiomi[39], P. Spillantini[1], F. Stolzi[6,7], S. Sugita[38], A. Sulaj[6,7], M. Takita[5], T. Tamura[19], T. Terasawa[40], S. Torii[3], Y. Tsunesada[41], Y. Uchihori[42], E. Vannuccini[2], J. P. Wefel[15], K. Yamaoka[43], S. Yanagita[44], A. Yoshida[38], K. Yoshida[22], and W. V. Zober[8]

[1]Department of Physics, University of Florence, Via Sansone, 1, 50019 Sesto, Fiorentino, Italy, [2]INFN Sezione di Florence, Via Sansone, 1, 50019 Sesto, Fiorentino, Italy, [3]Waseda Research Institute for Science and Engineering, Waseda University, 17 Kikuicho, Shinjuku, Tokyo 162-0044, Japan, [4]JEM Utilization Center, Human Spaceflight Technology Directorate, Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan, [5]Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwa-no-Ha, Kashiwa, Chiba 277-8582, Japan, [6]Department of Physical Sciences, Earth and Environment, University of Siena, via Roma 56, 53100 Siena, Italy, [7]INFN Sezione di Pisa, Polo Fibonacci, Largo B. Pontecorvo, 3, 56127 Pisa, Italy, [8]Department of Physics and McDonnell Center for the Space Sciences, Washington University, One Brookings Drive, St. Louis, Missouri 63130-4899, USA, [9]Heliospheric Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [10]Department of Physics, Catholic University of America, Washington, DC 20064, USA, [11]Center for Space Sciences and Technology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA, [12]Astroparticle Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [13]Center for Research and Exploration in Space Sciences and Technology, NASA/GSFC, Greenbelt, Maryland 20771, USA, [14]Institute of Applied Physics (IFAC), National Research Council (CNR), Via Madonna del Piano, 10, 50019 Sesto, Fiorentino, Italy, [15]Department of Physics and Astronomy, Louisiana State University, 202 Nicholson Hall, Baton Rouge, Louisiana 70803, USA, [16]Department of Physics and Astronomy, University of Padova, Via Marzolo, 8, 35131 Padova, Italy, [17]INFN Sezione di Padova, Via Marzolo, 8, 35131 Padova, Italy, [18]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Chuo, Sagamihara, Kanagawa 252-5210, Japan, [19]Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa, Yokohama, Kanagawa 221-8686, Japan, [20]Faculty of Science and Technology, Graduate School of Science and Technology,, Hirosaki University, 3, Bunkyo, Hirosaki, Aomori 036-8561, Japan, [21]Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa Oiwakecho, Sakyo, Kyoto 606-8502, Japan, [22]Department of Electronic Information Systems, Shibaura Institute of Technology, 307 Fukasaku, Minuma, Saitama 337-8570, Japan, [23]School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan, [24]National Institute of Polar Research, 10-3, Midori-cho, Tachikawa, Tokyo 190-8518, Japan, [25]Faculty of Engineering, Division of Intelligent Systems Engineering, Yokohama National University, 79-5 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan, [26]Faculty of Science, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan, [27]Hakubi Center, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan, [28]Department of Astronomy, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan, [29]Institute of Particle and Nuclear Studies, High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan, [30]University of Pisa, Polo Fibonacci, Largo B. Pontecorvo, 3, 56127 Pisa, Italy, [31]Astroparticle Physics Laboratory, NASA/GSFC, Greenbelt, Maryland 20771, USA, [32]Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Ibaraki College, 866 Nakane, Hitachinaka, Ibaraki 312-8508, Japan [33]Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA, [34]Department of Physical Sciences, College of Science and Engineering, Ritsumeikan University, Shiga 525-8577, Japan, [35]Faculty of Science and Engineering, Global Center for Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan, [36]Department of Physics and Astronomy, University of Denver, Physics Building, Room 211, 2112 East Wesley Avenue, Denver, Colorado 80208-6900, USA, [37]Quantum ICT Advanced Development Center, National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan, [38]College of Science and Engineering, Department of Physics and Mathematics, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo, Sagamihara, Kanagawa 252-5258, Japan, [39]College of Industrial Technology, Nihon University, 1-2-1 Izumi, Narashino, Chiba 275-8575, Japan [40]RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [41]Division of Mathematics and Physics, Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan, [42]National Institutes for Quantum and Radiation Science and Technology, 4-9-1 Anagawa, Inage, Chiba 263-8555, Japan, [43]Nagoya University, Furo, Chikusa, Nagoya 464-8601, Japan, [44]College of Science, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan

The fluxes of charged cosmic rays as measured by the DAMPE satellite

Paolo Bernardini (on behalf of the DAMPE Collaboration)
Dipartimento di Matematica e Fisica, Università del Salento
and Istituto Nazionale di Fisica Nucleare
Lecce, 73100, Italy
E-mail: paolo.bernardini@le.infn.it

DAMPE (DArk Matter Particle Explorer) is a satellite-born experiment promoted by the Chinese Academy of Sciences, with the collaboration of Italian and Swiss agencies. Since December 2015, DAMPE flies at the altitude of 500 km and collects data smoothly. The detector is made of four sub-detectors: top layers of plastic scintillators, a silicon-tungsten tracker, a BGO calorimeter (32 radiation lengths), and a bottom boron-doped scintillator to detect delayed neutrons. The main goal of the experiment is the search for indirect signals of Dark Matter in the electron and photon spectra with energies up to 10 TeV. Furthermore DAMPE studies cosmic charged and gamma radiation. Moreover, the calorimeter depth and the large acceptance allow to measure cosmic ray fluxes in the range from 20 GeV up to hundreds of TeV with unprecedented precision. An overview of the latest results about the charged cosmic rays will be presented.

Keywords: Cosmic rays; Satellite; DAMPE.

1. Introduction

The DArk Matter Particle Explorer (DAMPE) is a satellite-born experiment and it is funded by the strategic space projects of the Chinese Academy of Sciences with the active contribution of Italian and Swiss universities and institutions.[1,2] It is devoted to search for indirect dark matter signatures, to measure the fluxes of charged cosmic rays and to study the high energy gamma signal of astrophysical origin. In this paper the dark matter search and the gamma astronomy will be neglected to focus the reader's attention on the recent results about charged cosmic rays (electrons, protons and heavier nuclei).

2. The detector

The detector has been designed not only to reach the research goals. The final setup is an optimal compromise between scientific needs and reduction of power consumption and weight. In Fig. 1 the four sub-detectors of the DAMPE experiment are shown:

- PSD, in place of Plastic Scintillator strip Detector, used to measure the charge (Z) of incoming particles thanks to the Bethe-Bloch formula. The anti-coincidence with the PSD signal is also exploited to select gamma events.

- STK, in place of Silicon-Tungsten tracKer-converter, to reconstruct the trajectory of particles and to confirm the charge measurement. The tungsten is devoted to convert photons in electron-positron pairs.
- BGO, in place of bismuth germanate calorimeter (14 layers) of about 32 radiation lengths, that is ~ 1.6 nuclear interaction lengths, enough to get high resolution in energy measurement. The BGO bars allow to reconstruct the shower image in two views, then it is possible to distinguish between hadronic and electromagnetic ones.
- NUD, in place of NeUtron Detector, to further increase the hadronic shower rejection power.

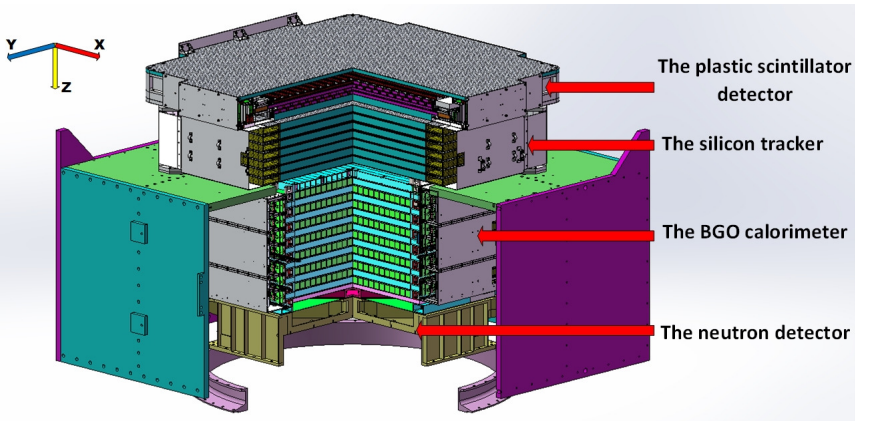


Fig. 1. The DAMPE detector with its 4 sub-detectors.

DAMPE performances were verified by a series of beam tests at CERN.[1,3,4] The wide energy range of electron beam and the high purity of proton beam allowed to check energy resolution, linearity (Fig. 2, left) and e/p separation (see Fig. 7 in the following). Also beams of argon and lead fragments were used to check the PSD capability to measure Z for heavy ions (Fig. 2, right for the argon beam). Furthermore sea-level muon test have been performed during different stages of the DAMPE assembly. Then the energy response to Minimum Ionizing Particles (MIPs), the efficiency and the detector alignment have been checked with a large sample of atmospheric muon tracks.

3. In-flight DAMPE performances

The satellite has been successfully launched on December 17, 2015. After few days all detectors were powered on and the data acquisition began very soon. We would like to stress that the detector works properly and flies at the altitude of 500 km on a Sun-synchronous orbit since almost 6 years, in spite of its foreseen 3-years life.

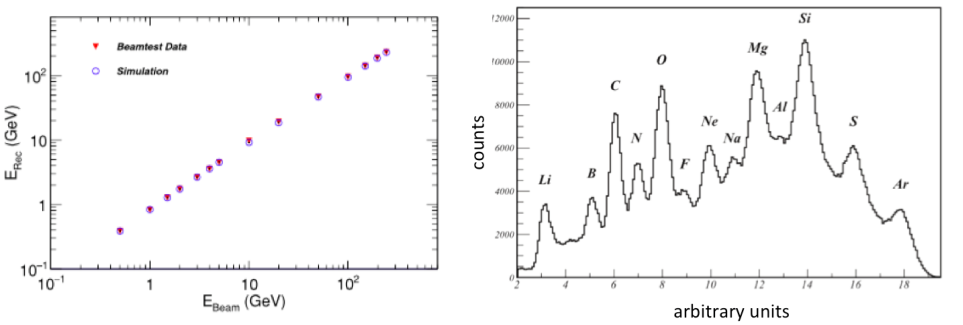


Fig. 2. Beam test - Left panel: Reconstructed energy versus electron beam energy (simulated and beam-test data are shown). Right panel: PSD charge spectrum of fragments of an argon beam (helium peak has been removed).

On orbit the detector is calibrated by using the standard scientific data and dedicated calibration data. The data acquisition system runs continuously in observation mode for most of the time and switches to calibration mode for 40 seconds twice per orbit. All the calibration results (pedestal values, dynode ratios, energy gains, trigger thresholds and so on) are consistent with the ground ones and very stable after temperature correction. The STK alignment allows a significant improvement of the tracking resolution.[5] In Fig. 3 (left panel) the PSD signal due to MIP protons is plotted as a function of time in a period of about 4 months. A similar plot is shown in the right panel of Fig. 3 for BGO signal due to no-showering helium nuclei. The high stability of the detectors is confirmed by these and many other checks.

4. Cosmic Rays (CR)

After more than 100 years since the CR discovery many questions are yet without answer. The spectra before the knee do not fit perfectly with a simple power-law and the canonical model of CR acceleration in Super Nova shock does not explain the observed structures. Precise measurements for different chemical components are essential for testing the CR models (sources, production, acceleration, diffusion in the galaxy, presence of different populations, and so on).

DAMPE can contribute to CR studies measuring the spectrum structures for different nuclei in the range from tens of GeV up to hundreds of TeV. The possibility to reach such high energies with on-orbit detectors put a bridge between space and ground experiments and pave the way to a combined analysis.

The measurement of CR spectra is based on the estimate of charge and energy of incoming particles. By means of the STK tracking (supported also by the information on the fired BGO bars) the path-length in PSD is reconstructed. Then the PSD signal allows to measure the charge (square root of the energy release in PSD is proportional to path-length and Z). In Fig. 4 the elemental CR spectrum is

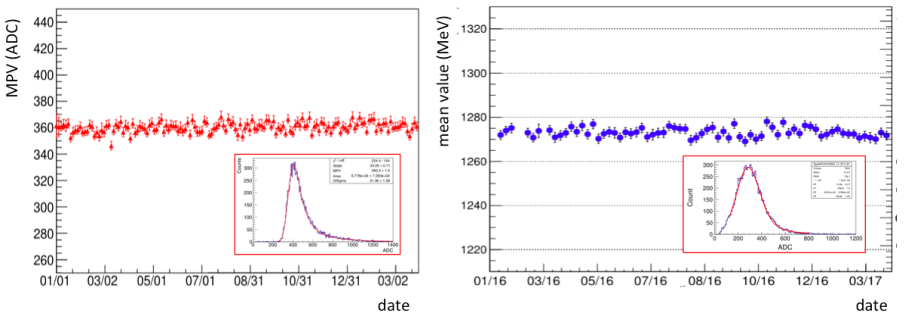


Fig. 3. Left panel: the deposited energy in PSD for minimum ionizing protons is fitted with the convolution of a Landau function with a Gaussian one (insert). The Most Probable Value (MPV) is shown as a function of time. Right panel: similar plot for minimum ionizing helium nuclei in BGO (the mean value of the deposited energy is displayed versus time instead of MPV).

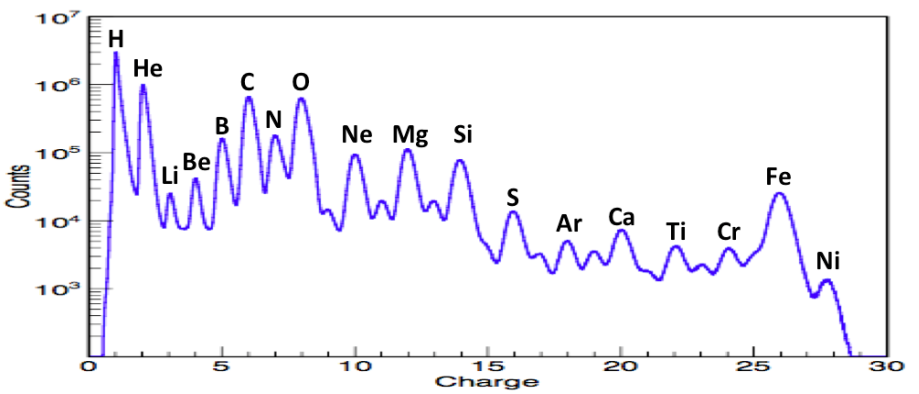


Fig. 4. Charge spectrum reconstructed on the basis of the deposited energy in the PSD. Almost all Z-peaks are visible up to nickel.

shown from hydrogen up to iron and beyond. By means of the template-fit procedure the different peaks are separated and the mutual contamination is estimated (an example is shown in Fig. 5 where the data-Montecarlo agreement is excellent).

Lastly the energy is estimated by the signal in the BGO calorimeter. Its unprecedented thickness allows to push the direct measurement of CR spectra to highest energies. The imaging allows also to distinguish the features of hadronic showers from electromagnetic ones.

The DAMPE spectra for protons[6] and helium nuclei[7] are presented in Fig. 6. The analyses for heavier nuclei are in progress. Both the spectra in Fig. 6 show a similar shape inconsistent with a single power-law. The fit of these spectra improves significantly using smoothly-broken power-laws in order to reproduce a hardening followed by a softening. The observation of the hardening confirms previous measurements, instead the softening is observed for the first time by DAMPE with such

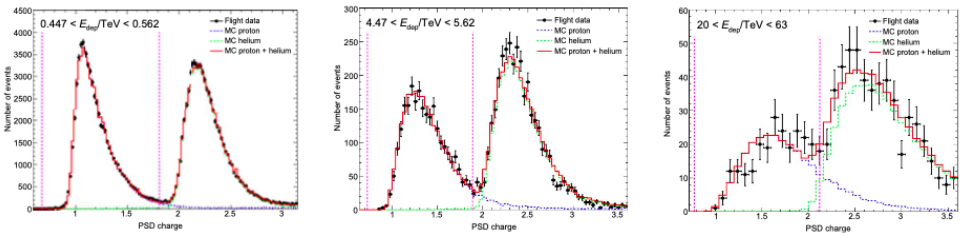


Fig. 5. Combined PSD signal for protons and helium nuclei, for different ranges of BGO energy (447-562 GeV, 4.47-5.62 TeV, 20-63 TeV). The colour codes to distinguish on-orbit and simulated data are shown in the plot. The cut used to select proton candidates is represented by the vertical dotted lines.

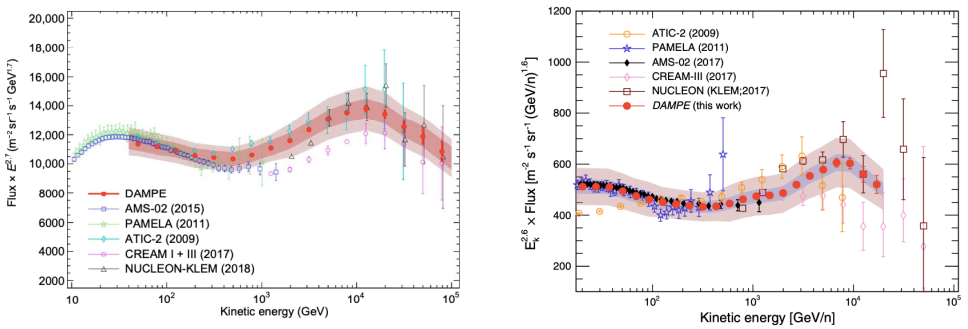


Fig. 6. Spectra for cosmic protons (left panel) and helium nuclei (right panel). For protons (helium) the flux is multiplied for $E^{2.7}$ ($E^{2.6}$). Two changes of spectral index are visible in both the spectra (hardening followed by softening).

significance. The break-energy for the softening suggests a dependence on particle charge, although a dependence on mass can not be ruled out yet.

5. Cosmic electrons and positrons (CRE)

Some measurements[8–14] reported an unexpected structure of CRE flux. Furthermore evidence for a spectral break in the TeV range of the CRE flux has been provided by indirect measurements[15,16] affected by significant uncertainties.

In two years of data taking DAMPE measured CRE spectrum in the range from 25 GeV to 4.6 TeV[17] with very low background. Indeed the imaging properties of the BGO calorimeter are exploited to recognise the showers initiated by CRE. The PID algorithm is based on the spread of the BGO hits (*spread*) and on the fraction ($F_{\rm last}$) of energy on the lower layer with respect to the sum of energies released on all the BGO layers. The quantities *spread* and $F_{\rm last}$ are combined in the parameter

$$\zeta = F_{\rm last}\ \frac{(spread/\mathrm{mm})^4}{8\times 10^6}. \qquad (1)$$

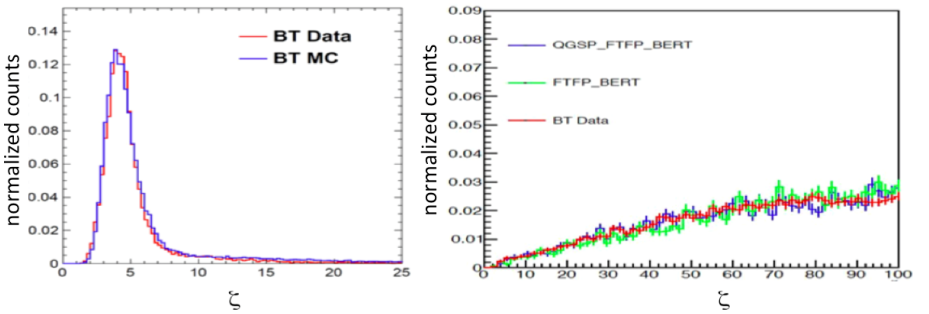


Fig. 7. Beam test (BT) - The ζ-distributions for electrons (left panel) and for protons (right panel). In the proton plot the real data are compared with simulations based on different models of the hadronic interactions.

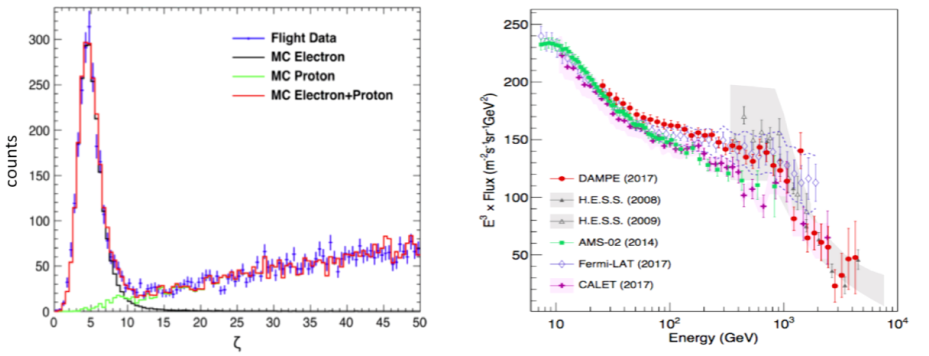


Fig. 8. Left: Distribution of the ζ parameter (the cut $\zeta < 8.5$ is applied to select the CRE sample). Right: electron and positron spectrum measured by different experiments (DAMPE: red dots, HESS systematic errors: grey band).

The efficiency of the ζ-parameter in order to distinguish electrons and protons has been verified in the beam test (Fig. 7). Then the CRE sample is selected by requiring $\zeta < 8.5$ (see Fig. 8, left panel). The CRE flux measured by DAMPE as a function of energy is shown in right panel of Fig. 8 with measurements of others experiments. Also in this case a smoothly broken power-law is preferred to fit the largest part of the spectrum.

6. Conclusions

The DAMPE detector works extremely properly since its launch more than 5 years ago and it is expected to operate stably in the next few years. Many analyses based on DAMPE data are going on successfully. The measurements of CRE, proton and helium spectra are particularly remarkable. Indeed the direct detection of spectral breaks confirms previous direct and indirect measurements, clarifies the shape of

the spectra at highest energies. These results put severe constraints about models of acceleration and diffusion of galactic cosmic rays. Many new results are expected from the analysis of the complete sample of DAMPE data.

References

1. Chang J., Chinese J. Space Science **34**, 550 (2014).
2. Chang J. *et al.*, Astropart. Physics **95**, 6 (2017).
3. Azzarello P. *et al.*, Nuclear Instruments and Methods A **831**, 378 (2016).
4. Zhang Z. *et al.*, Nuclear Instruments and Methods A **836**, 98 (2016).
5. Tykhonov A. *et al.*, Nuclear Instruments and Methods A **893**, 43 (2018).
6. Ambrosi G. *et al.* (DAMPE Collaboration), *Science Advances 5* (2019) eaax3793.
7. Alemanno F. *et al.* (DAMPE Collaboration), *Physics Review Letters 126* (2021) 201102.
8. Chang J. *et al.* (ATIC Collaboration), Nature **456**, 362 (2008).
9. Abdo A.A. *et al.* (Fermi-LAT Collaboration), Phys. Rev. Lett. **102**, 181101 (2009).
10. Abdollahi S. *et al.* (Fermi-LAT Collaboration), Phys. Rev. D **95**, 082007 (2017).
11. Aguilar M. *et al.* (AMS-02 Collaboration), Phys. Rev. Lett. **113**, 221102 (2014).
12. Adriani O. *et al.* (PAMELA collaboration), Phys. Rev. Lett. **106**, 201101 (2011).
13. Adriani O. *et al.* (CALET Collaboration), Phys. Rev. Lett. **119**, 181101 (2017).
14. Adriani O. *et al.* (CALET Collaboration), Phys. Rev. Lett. **120**, 261102 (2018).
15. Aharonian F. *et al.* (H.E.S.S. Collaboration), Phys. Rev. Lett. **101**, 261104 (2008).
16. Aharonian F. *et al.* (H.E.S.S. Collaboration), Astron. Astrophys. **508**, 561 (2009).
17. Ambrosi G. *et al.* (DAMPE Collaboration), Nature **552**, 63 (2017).

Recent results from the Pierre Auger Observatory

E. Roulet for the Pierre Auger Collaboration
Centro Atómico Bariloche
Av. Bustillo 9500, Bariloche, 8400, Argentina
E-mail: roulet@cab.cnea.gov.ar
Full author list: https://www.auger.org/archive/authors_2021_07.html

The Pierre Auger Observatory has by now achieved an exposure of order 10^5 km^2 sr yr, exploring about 85% of the sky. In this talk, I will review some of the recent results, including the detailed measurements of the features in the cosmic ray spectrum, the study of the anisotropies in the cosmic ray arrival directions both at large and intermediate angular scales, the inferred mass composition, and multimessenger searches.

Keywords: Cosmic rays

The Pierre Auger Observatory has been operational since January 2004 and has by now reached an accumulated exposure of order 10^5 km^2 yr sr (see Fig. 1). This enormous increase in the number of detected cosmic rays (CRs) with ultrahigh energies, by more than an order of magnitude with respect to previous observations, has allowed measuring the CR spectrum with unprecedented detail,[1] firmly establishing the suppression at the highest energies, above 47 EeV, as well as the various features of the spectrum, including the discovery of a softening at 14 EeV, the observation of the ankle at 5 EeV, the second-knee at 0.16 EeV, and the low-energy ankle at 28 PeV (see Fig. 2).

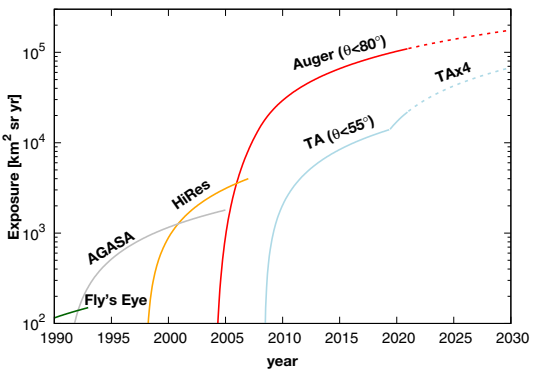


Fig. 1. Exposure vs. time of ultrahigh-energy air shower arrays. The previous experiments (Fly's Eye, HiRes and AGASA) as well as the Auger and Telescope Array exposures used in anisotropy searches at the highest energies are shown, together with the extrapolations to the next years involving the enlarged TAx4 upgrade.

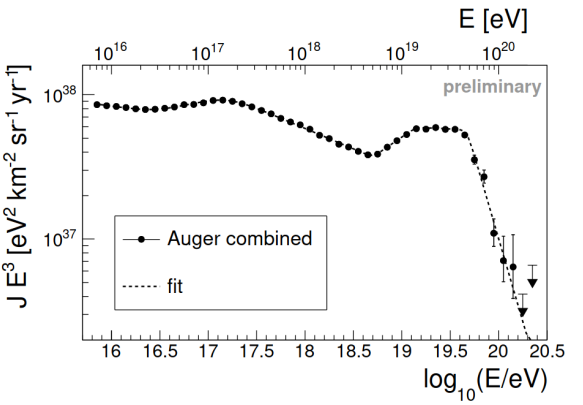


Fig. 2. Combined spectrum from measurements obtained with different techniques together with a broken power-law fit (see ref.1 for details).

By observing events with zenith angles up to 80°, i.e., covering 85% of the sky, it also became possible to determine with high significance an anisotropy in the CR arrival direction distribution at energies above 8 EeV, which has a characteristic dipolar pattern with an amplitude of about 7%.[2] The direction of this anisotropy, with the reconstructed dipole pointing about 115° away from the direction towards the Galactic center, provides a clear indication supporting the extragalactic origin of the CRs at these energies (see left panel of Fig. 3). Considering several energy bins above 8 EeV, the dipole amplitude is found to actually increase with energy, and the direction of the fitted dipole in all bins lies not far from the direction of the outer spiral arm of the Galaxy.[3] At energies below 8 EeV, no significant anisotropies were observed, so that relevant upper bounds on the equatorial dipole amplitudes were set down to energies of 0.03 EeV, and interestingly the right ascension phase of the flux modulations below about 1 EeV is not far from the right ascension of the Galactic center.[4] Hints of anisotropies at intermediate angular scales, of order 20°, have been obtained at energies above about 40 EeV, with the main excess lying around the direction towards the nearby radiogalaxy Centaurus A (right panel of Fig. 3). This radiogalaxy then provides a possible candidate source of ultrahigh-energy CRs, although other potential CR sources are also located in the same region of the sky, such as the starburst galaxies NGC4945 and M83. A likelihood ratio test comparing the CR arrival directions above different threshold energies with a catalog of nearby starburst galaxies, flux-weighted and smeared, reveals a maximum significance above a threshold of 38 EeV for an angular scale of a top-hat window of 25°.[5]

The hybrid nature of the Auger Observatory, with both surface detectors (SD) and fluorescence detectors (FD), allows it to perform sensitive measurements of the mass composition of the primary CRs. The SD consists of instrumented water tanks that measure the Cherenkov light emitted by relativistic secondary charged particles

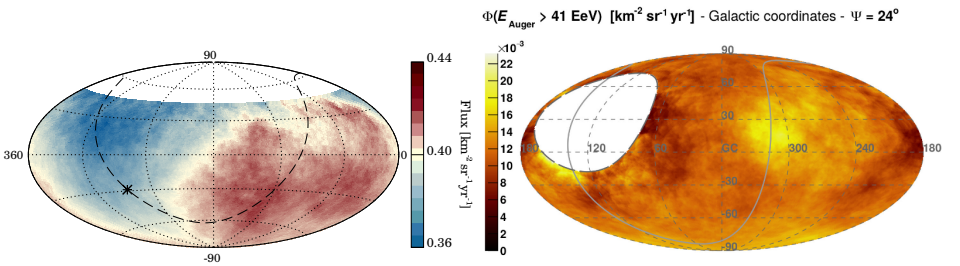


Fig. 3. CR flux map above 8 EeV averaged on windows of 45° radius (left panel, in Equatorial coordinates) and above 41 EeV averaged of 24° radius windows (right panel, Galactic coordinates).

crossing them and is hence sensitive to both the muonic and the electromagnetic components of the showers reaching ground level. The FD telescopes measure instead the longitudinal development in the atmosphere of the CR induced showers by observing, with a $\sim 15\%$ duty cycle, the light emitted by the N_2 air molecules that get excited by the passage of the electromagnetic component of the shower. The FD allows for an almost calorimetric measurement of the primary energy, which is also used to calibrate the SD energy assignment. Moreover, the atmospheric depth at which the showers reach their maximum development, $X_{\rm max}$, provides crucial information about the mass composition of the CRs, because lighter primaries give rise to more penetrating showers than heavier ones since these last can be considered as a superposition of many lower energy showers induced by the individual nucleons. The measurements of the distribution of $X_{\rm max}$ values allowed us to establish[6] that at energies of a few EeV the average CR masses are light, being dominated by H and He primaries, and they become increasingly heavier as the energy is further increased, lying in the ballpark of the expectations from primary N nuclei at energies of 30 EeV and possibly even heavier at higher energies (the actual masses inferred depend, however, on the hadronic interaction model adopted in the analysis). The narrow spread in the fluctuations on the values of $X_{\rm max}$ measured above the ankle energy also implies that there is little admixture between elements with disparate masses at these energies. These measurements are depicted in Fig. 4.

A combined fit to the spectrum and anisotropy measurements within an astrophysical scenario with equal luminosity sources having power-law spectra with rigidity dependent cutoffs has been performed in ref. 7.

Besides the measurement of $X_{\rm max}$ with FD, also the SD has been used to obtain composition information,[8] exploiting the rise-times of the measured signals in the triggered detectors. Given that the shower muons travel straight from their production points high in the atmosphere, while the electromagnetic component of photons, electrons, and positrons have a more diffusive propagation in the atmosphere, the muonic signal is expected to be more concentrated in the early part of the shower front while the electromagnetic one will have a larger spread. Hence, shorter rise times are expected in the signals from showers with a more abundant

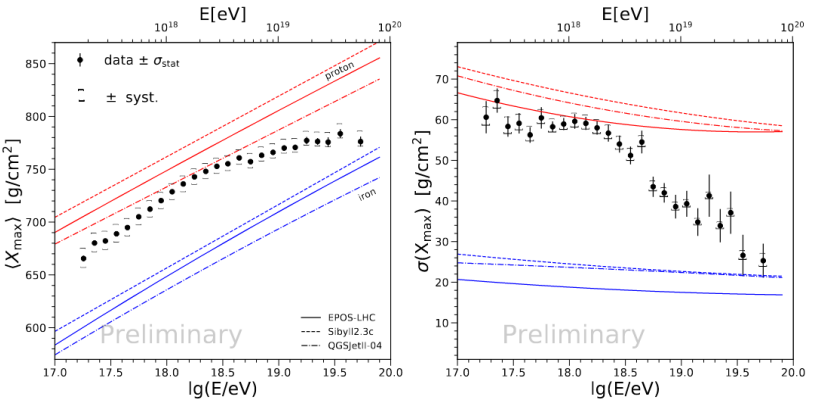


Fig. 4. Mean $X_{\max}$ and its dispersion as a function of energy, as measured with the Auger fluorescence detectors.

muonic component, as is the case for the showers induced by heavier nuclei. These measurements can then be correlated with the FD measurements of the depth of shower maxima. Given the larger statistics obtained with SD, which has a 100% duty cycle, it has become possible to obtain more detailed information about the composition at the highest energies, actually splitting the highest energy FD bin in four SD bins. In this way the trend towards a heavier composition for increasing masses is confirmed with SD up to the highest energies.

Another search performed has been to look for a component of photon-induced showers, which would be very poor in muons, and this allowed to set stringent bounds on the primary photon fluxes above EeV energies.[9] The upper bounds strongly disfavor exotic scenarios for the production of UHECRs, such as those involving superheavy particle decays (see left panel in Fig. 5).

Also ultrahigh-energy neutrinos have been searched through the study of inclined showers.[10] Very inclined showers produced by ordinary nuclear-induced cascades, which initiate high in the atmosphere, would have their electromagnetic component totally attenuated by the time they reach ground level. On the other hand, neutrino-induced showers can be initiated at any depth in the atmosphere, and those originating close to the detector will have a significant electromagnetic component, looking then 'younger' than the hadronic ones, what could be exploited to single them out from the background. Actually, the most sensitive neutrino search is that looking for showers produced by Earth-skimming tau neutrinos interacting in the crust of the Earth and producing tau leptons that manage to exit from the soil before decaying to produce the atmospheric shower that may be observed. The most sensitive energy range for these searches is around the EeV, which just happens to be the range where the fluxes of cosmogenic neutrinos are expected to peak. In particular, cosmogenic neutrinos could arise from the photo-pion production taking place when CR protons with energies exceeding the GZK threshold of about 40 EeV

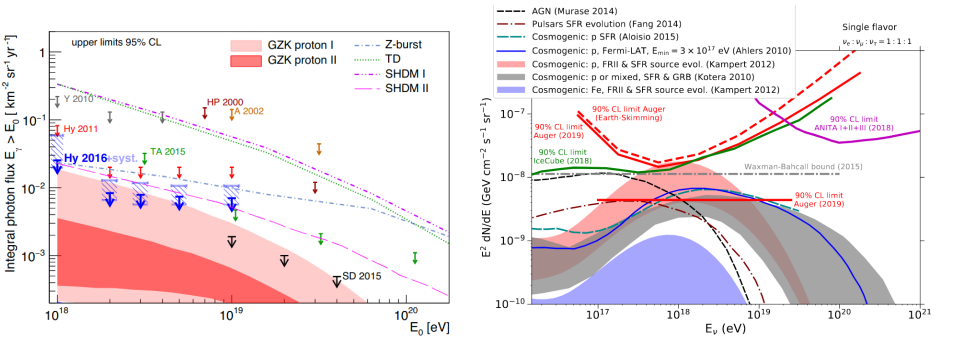


Fig. 5. Bounds on photon fluxes (left) and neutrino fluxes (right), compared to some model expectations.

interact with CMB photons, and the neutrinos produced in the charged pion decays have typical energies of about $E_p/20$. No neutrino candidates have yet been observed in these searches, allowing us to set bounds that start to constrain some of the most optimistic scenarios for the production of cosmogenic neutrinos, which are those with protons dominating the CR fluxes at the highest energies. However, more realistic scenarios with a predominantly heavier composition at the highest energies predict fluxes below the present bounds (see right panel of Fig. 5).

Searches of neutrinos in association with the first neutron star binary merger detected by the LIGO-Virgo Collaboration also led to negative results,[11] in spite of the lucky coincidence that the event just happened to be a few degrees below the horizon, a situation in which the sensitivity to Earth-skimming tau neutrinos is largely enhanced. It is interesting to note that the model predictions for the case in which the jet of this closeby merger had been aligned with the line of sight were not very far from the Auger sensitivity, but the lack of an observed signal is not unexpected, given the slight misalignment of the jet in this particular merger. Also searches in the direction towards the blazar TXS 0506+056, from which a $\sim$ 300 TeV neutrino was observed by IceCube, revealed no neutrino candidate, allowing to constrain the neutrino flux from this object at the highest energies.[12]

The Pierre Auger Observatory is at present undergoing an upgrade, consisting in adding 4 m^2 scintillators on top of each SD and using faster electronics, which in combination with the water-Cherenkov detectors should allow for a cleaner separation between the electromagnetic and muonic components of the showers. This separation should make it possible to obtain composition information on an event by event basis with the SD detector, allowing, for instance, to improve the sensitivity of the anisotropy studies by restricting them to just the lighter CR component (which is less deflected by the Galactic and extragalactic magnetic fields). Also radio antennas are being added to each SD detector to observe the radio emission from the electromagnetic component of the showers, which should help in particular to study the composition of the inclined showers. This new phase of the Observatory

should allow for improved studies of all the topics just discussed, as well as to better handle the still remaining open issues.

References

1. Aab et al. (Pierre Auger Collaboration), Phys. Rev. Lett. 125 (2020) 121106; V. Novotný for the Pierre Auger Collaboration, PoS(ICRC2021)324
2. A. Aab et al. (Pierre Auger Collaboration), Science 357 (2017) 1266
3. A. Aab et al. (Pierre Auger Collaboration), Astrophys. J. 868 (2018) 4
4. A. Aab et al. (Pierre Auger Collaboration), Astrophys. J. 891 (2020) 142
5. A. Aab et al. (Pierre Auger Collaboration), Astrophys. J. Lett. 853 (2018) L29; J. Biteau for the Pierre Auger Collaboration, PoS(ICRC2021)307
6. A. Aab et al. (Pierre Auger Collaboration), Phys. Rev. D 90 (2014) 122005; A. Yushkov for the Pierre Auger Collaboration, PoS(ICRC2019)483
7. A. Aab et al. (Pierre Auger Collaboration), JCAP 04 (2017) 038
8. A. Aab et al. (Pierre Auger Collaboration), Phys. Rev. D 96 (2017) 122003; C. Todero Peixoto for the Pierre Auger Collaboration, PoS(ICRC2019)440
9. A. Aab et al. (Pierre Auger Collaboration), JCAP 04 (2017) 009
10. A. Aab et al. (Pierre Auger Collaboration), JCAP 10 (2019) 022
11. ANTARES Collaboration, IceCube Collaboration, The Pierre Auger Collaboration, and LIGO Scientific Collaboration and Virgo Collaboration, Astrophys. J. Lett. 850 (2017) L35
12. A. Aab et al. (Pierre Auger Collaboration), Astrophys. J. 902 (2020)105

The HERD space mission

F.C.T. Barbato on behalf of the HERD Collaboration

Gran Sasso Science Institute, L'Aquila, Italy
INFN Laboratori Nazionali del Gran Sasso, Assergi, L'Aquila, Italy
E-mail: felicia.barbato@gssi.it

The High Energy cosmic-Radiation Detection (HERD) space mission is now being designed, as a result of an international collaboration among several chinese and european institutions, to make cosmic ray (CR) direct measurements at the highest possible energies with current technologies.

HERD primary scientific goals include precise measurements of the energy spectra of CR individual species up to few PeV, reaching the knee of the all-particle spectrum, and study electrons and photon of spectra from GeV up to tens of TeV, also contributing to multimessenger observations together with other satellites and ground-based experiments.

In order to reach these goals HERD is configured to accept incident particles from its top and the four lateral sides. The baseline design includes covering the top and fours sides with: the Silicon Charge Detector (SCD), for incident particle trajectory and charge measurement, the Plastic Scintil- lator Detector (PSD), for photon tagging and precise charge measurement, and a scintillating Fiber Tracker (FIT). The core of the facility is made by a LYSO crystal calorimeter (CALO) that with its 3 interaction lengths and 55 radiation lengths will allow the measurement of incident gamma-rays, electrons and cosmic ray nuclei with unprecedented resolution and 3D reconstruction. In addition, on one side a Transition Radiation detector (TRD) will be installed for on-orbit calibration of the CALO.

Keywords: Cosmic rays; Chinese Space Station; space missions; SiPM; LYSO.

1. Introduction

Cosmic rays are messengers with which we use to explore the Universe and its laws. Despite being discovered more than one century ago, our knowledge about them is still limited. In the recent years, some progresses have been made in our knowledge and understanding of them thanks to several experiments. Nevertheless, many fundamental questions on this topic are still open: where are they produced? how are they accelerated to such energies? how do they propagate in the Universe?

A big issue concerns the so-called *knee* region of the primary cosmic ray spectrum. Here (at $\sim$ PeV energies), the spectrum steepens. This behaviour was firstly observed in 1958, nevertheless the cause remains still unknown. Because of the rapidly falling intensity of the cosmic ray flux with energy, in order to explore this region we need experiments with a large acceptance and operating over years. In past this role was played by ground based experiments. Nevertheless, this kind of experiments are not able to perform mass composition studies. For this reason, we need new spaceborne observatories performing studies aiming at finding an answer to this question.

The High Energy cosmic Radiation Detection (HERD) facility is an international space mission that will start operation around 2027 to make cosmic ray direct measurements at the highest possible energies with advanced technologies. It will be installed on the Chinese Space Station (CSS) where it will observe the high energy sky for more than 10 years. The HERD project involves researchers from Chinese institutions, Italy, Switzerland and Spain. The unprecedented acceptance and the 3D segmented calorimeter allow HERD to represent the future of the direct cosmic ray detection and a missing observational link among space-based and ground-based experiments.

2. Scientific objectives

HERD primary scientific goals include:

- extending the measurement of cosmic ray flux up to a few PeV testing the theory of the knee structure as due to galactic source acceleration limit;
- extending the measurement of $e^+ + e^-$ flux up to several tens of TeV testing the hypothesis of the expected cutoff at high energy distinguishing between DM or astrophysical origin of positron excess and looking for possible other structures of the spectra and eventual anisotropies;
- large acceptance, high sensitivity to γ-rays up to several hundreds of GeV for a long term gamma ray sky also searching for γ-lines associated to DM annihilation, while accomplishing a γ-ray sky survey up to very high energy.

Cosmic ray origins Since their discovery in 1910, our knowledge about cosmic ray physics experienced several progresses. Today, we are in another golden era for cosmic rays, due mainly to their connection with physics at the highest energy, indirect search for particle dark matter and site and mechanism of cosmic accelerators. Information about the latter topic can be obtained from the energy spectrum of cosmic rays. It is expected to follow a power-law for energies below the "knee" (at 3-4 PeV) because of the canonical shock acceleration of particles, see FIG. 1.

Nevertheless, several experiments observed changes in the power-law spectral indices for protons, helium, and heavy nuclei. Current data suggest a hardening of the spectra above about 0.2 TeV/nucleon. The hardening at energies above 200-400 GeV is well established since the first observation by PAMELA,[2] while the softening at energies above 10 TeV is observed by different experiments with the first strong evidence in DAMPE data.[3] The deviations from single power law of the spectra motivate extensive investigations for deeper understanding of the acceleration and propagation mechanisms or of new possible cosmic ray sources. Several theoretical solutions have been proposed, exploiting different hypotheses on source properties/populations, acceleration/propagation mechanisms and particle physics issues at high energies, in order to give a satisfactory explanation of this behaviour.[4]

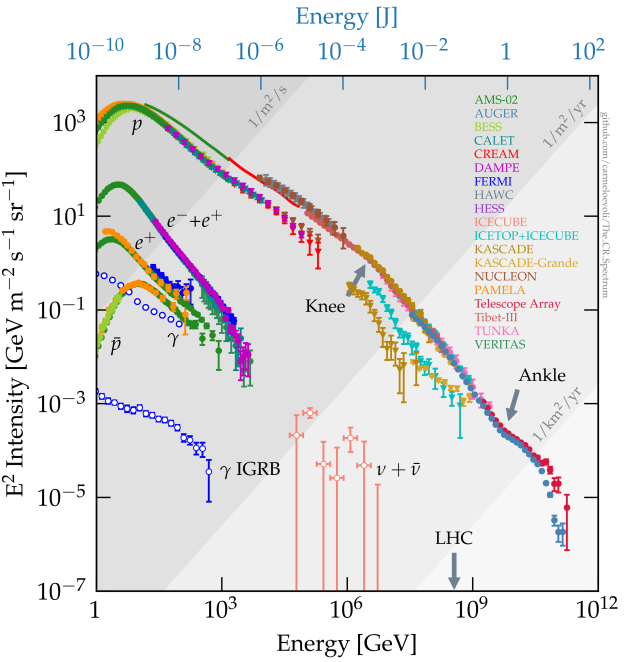


Fig. 1. Up-to-date summary of measurements of CR energy spectra.[1] For energies below ~ 100 TeV, the spectra of different species are shown, and for higher energies the all-particle spectra are plotted. References of data: CREAM; ATIC; AMS-02; PAMELA; Fermi-LAT; HESS; IceCube; Akeno; Tibet; AGASA; HiRes; Auger.

So one of the HERD objectives will be the extension of the measurement of p and He flux up to a few PeV. The HERD data can clearly reveal the knee structure of both protons and Helium nuclei and can critically address the possible Z-dependence, A-dependence, or constant knee of different compositions, which are very important to understand the physical nature of the knee of CRs, see FIG. 2. Measurements of secondary over primary ratio, such as B/C, are important to probe the propagation of CRs. HERD will extend the precise measurements of B/C up to a few TeV/nucleon, thus accurately determine the propagation behaviour of CRs.

Finally, taking into account that the Iron nuclei are the end-products of stellar nucleosynthesis, they are thus very good tracer of the acceleration sites of CRs. HERD is expected to strongly improve such measurements.

Measurement of cosmic electron+positron spectrum and Dark Matter signal search In 2009, PAMELA experiment reported a positron excess[5] in the CR flux at energies 10-100 GeV, later it was very precisely measured by AMS-02 in an extended energy range up to 700 GeV.[6] Such a signal is generally expected from dark matter annihilation. However, the hard positron spectrum and large amplitude are difficult to achieve in most conventional WIMP models. A variety of theoretical

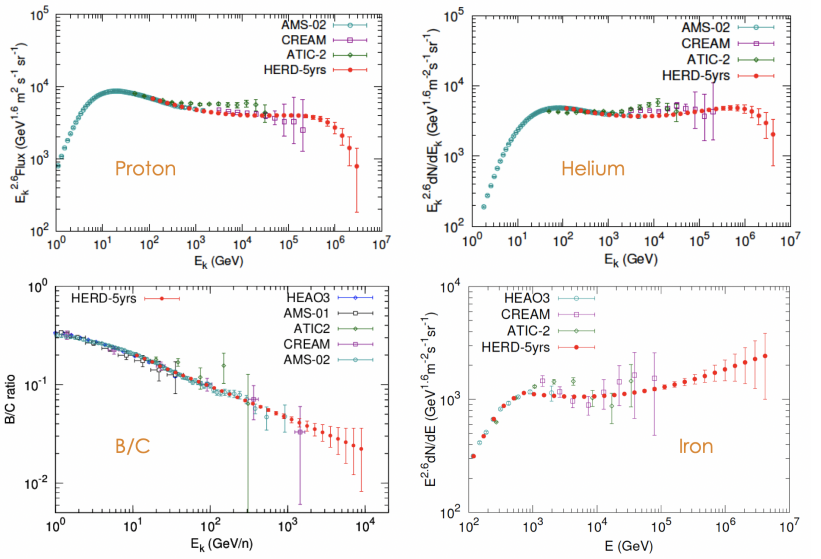


Fig. 2. Expected p, He, B/C and Fe energy spectra measured with 5yr exposure of HERD, compared with that observed by other experiments.

studies focusing on the DM interpretation of this excess have been performed in the literature, however the origin of these high energy positrons remains debatable.

Evidently, it is of a great importance to extend this measurement with greater precision at higher energy scales in order to distinguish the origin of the high energy electrons/positrons, either from DM annihilation or from astrophysical sources.

In this sense, progress has been made by the satellite experiment DAMPE (launched in 2015) which in 2017 reported its first result of CR electrons/positrons, showing a cutoff at ~ 0.9 TeV.[7] It has long been believed that the shape of the spectral cutoff is different for the astrophysical source (softer) and DM annihilation (harder).

Compared with DAMPE, the HERD experiment will have a larger acceptance and can extend the measurement of the electron-positron spectrum to higher energies with optimal precision, see FIG. 3, thus providing valuable insight on high energy astrophysics and DM studies.

A high energy gamma ray observatory HERD will be the most sensitive sky surveyer with very wide field of view for gamma rays between tens of GeV and several TeV, thus extending the Fermi-LAT observations at higher energies (E>300 GeV) and additionally complementing CTA, the most sensitive high energy γ-ray telescope with a narrow field of view.

The observation of the γ-ray sky is among the main scientific objectives of the HERD collaboration being relevant in understanding the nature of DM while also studying cosmic ray acceleration and propagation mechanisms in the Galaxy.[8]

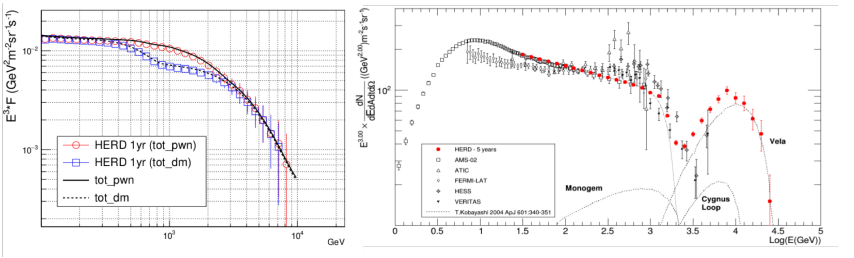


Fig. 3. Expected $e^+ + e^-$ flux in 1 year with PWN or DM sources on the left. Expected e++e- flux in 5 years on the right.

Information about cosmic ray spectra and intensities in distant locations, indeed, could be extracted by measuring the diffuse γ-ray emission, that traces energetic particle interactions, primarily protons and electrons, in the Interstellar Medium (ISM).[9] Moreover, the understanding of Galactic diffuse emission is also crucial to dark matter studies, since γ-rays can be produced in cosmic ray interactions mainly via decay of neutral pions and possibly from the annihilation of dark matter weakly interacting massive particles (WIMPs). The Galactic Center (GC) is expected to be the brightest source of DM annihilations in the γ-ray sky by several orders of magnitude. Although due to several astrophysical processes in the crowded GC region make it extremely difficult to disentangle the DM signal from conventional emissions, the DM-induced γ-ray emission is expected to be so large there that its search is of crucial relevance.[10] In the analyses of Fermi data of the Galactic center, the diffuse γ-ray backgrounds and discrete sources, as we model them today, seem to account for the majority of the detected γ-ray emission. Nevertheless, residual emission not accounted for by the above mentioned models of standard sources could still be present,[11][12] and be interpreted as product of DM annihilation.[13] The spectrum and spatial distribution of the GeV excess is consistent with what can be expected from Weakly Interacting Massive Particle (WIMP) with mass of a few tens of GeV of annihilating to standard model particles in the inner Galactic halo. With significantly improved sensitivity, HERD could measure with high resolution these spectra. Furthermore, the higher effective area especially at TeV energies may help in extending the energy range in which we detect the Galactic diffuse emission. This would provide a unique bridge for ground-based experiments sensitive in the TeV energy range.

3. The HERD payload

The HERD payload will be installed onboard of the CSS around 2027 and will be operating for 5 to 10 years, see TAB 1. Its heart, the calorimeter (CALO), will be surrounded on 5 sides by the other subdetectors i.e. the Fiber Tracker (FIT), the Plastic Scintillator Detector (PSD) and the Silicon Charge Detector (SCD), while a Transition Radiation Detector (TRD) will be placed only on one lateral side,

see FIG. 4. This configuration will guarantee a wide field of view and more than one order of magnitude increase in acceptance with respect to current experiments, see TAB 2.

Table 1. Mission requirements of the HERD payload

Mission requirements	Value
Life Time	>10 years
Orbit	Circular LEO
Altitude	340-450 km
Inclination	42°
FOV	± 70 °
Mass	4t

Table 2. Main requirements of the HERD payload

Item	Value
Energy range (e/γ)	10 GeV - 100 TeV; >0.1 GeV(γ)
Energy range (nucleus)	30 GeV - 3 PeV
Angular resolution (e/γ)	0.1 deg.@10 GeV
Charge measurement (nucleus)	0.05 - 0.15 c.u.
Energy resolution (e)	1% @ 200 GeV
Energy resolution (p)	20% @ 100 GeV - PeV
e/p separation	$\sim 10^{-6}$
Geometric factor (e)	>3 m^2Sr @ 200 GeV
Geometric factor (p)	>2 m^2Sr @ 100 GeV

The Silicon Charge Detector The silicon charge detector is the outermost shell of HERD. It will measure the incident particle trajectory and charge of cosmic rays up to Z=26. Its position is chosen to ensure a reliable charge measurement and to avoid fragmentation phenomena of high energy particles. The current design follows the example of the DAMPE Silicon Tracker (STK). Each detection unit, i.e. each side, contains 8 layers of 300 μm microstrip silicon detectors mounted with alternating orthogonal direction strip directions onto low-density aluminum honeycombs, see FIG. 5. It is highly segmented to minimize the backscattered secondary particles coming from the CALO. The global active area is about 60 m^2.

We simulated several samples of proton, helium, carbon, oxygen, silicon, and iron nuclei with energies from 10 GeV/n to 1 TeV/n and with isotropic incidence to study the performances of the SCD. We found that the combined charge resolution of 8 layers is below 0.3 c.u. for all samples. Test beams are ongoing at the CERN SPS and PS ion beam lines to directly measure charge resolution of SCD prototypes.

The Plastic Scintillator Detector The PSD is the second detector encountered by an incoming particle. It works as an anticoincidence detector (i.e. it is able to discriminate incident γ-rays from charged particles) and provides an additional charge measurement of incoming cosmic-ray nuclei in a range up to Z $\leq$26.[14]

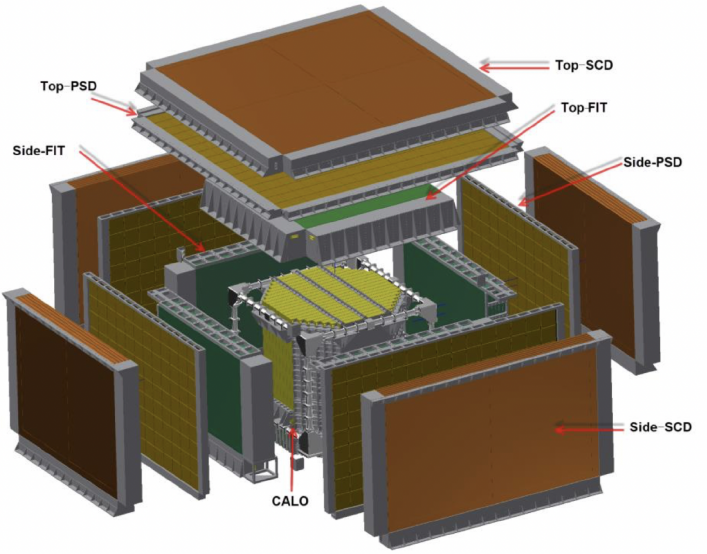


Fig. 4. Exploded model of HERD. In orange the silicon charge detector (SCD), in dark yellow the plastic scintillator detector (PSD), in green the fiber tracker (FIT) and in light yellow the calorimeter (CALO).

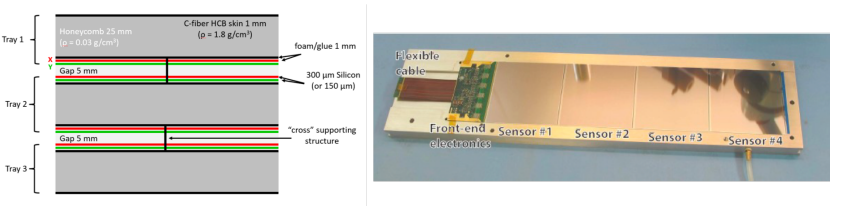


Fig. 5. Scheme of the SCD on the left. A prototype of SCD on the right.

Currently there are two designs for the PSD: one made by plastic scintillator bars the other made by plastic scintillator tiles. In both cases the single detection unit (i.e. bar, tile) is readout by silicon photomultipliers (SiPMs). The use of SiPMs will ensure, in addition to exceptional detection performances, low-weight, negligible power consumption, compactness and non-sensitivity to magnetic fields.

The bar option forsees two orthogonal layers of trapezoidal bars, see FIG. 6

The tile option forsees two staggered layers of square tiles, see FIG. 7.

Both configurations show advantages and disadvantages, mainly related to the optimal number of readout channels versus back–splash (or back–scattering) effects.

Currently both prototypes are undergoing beam tests at CERN with the aim to define the best scintillator type and size as well as SiPM quantity and model.

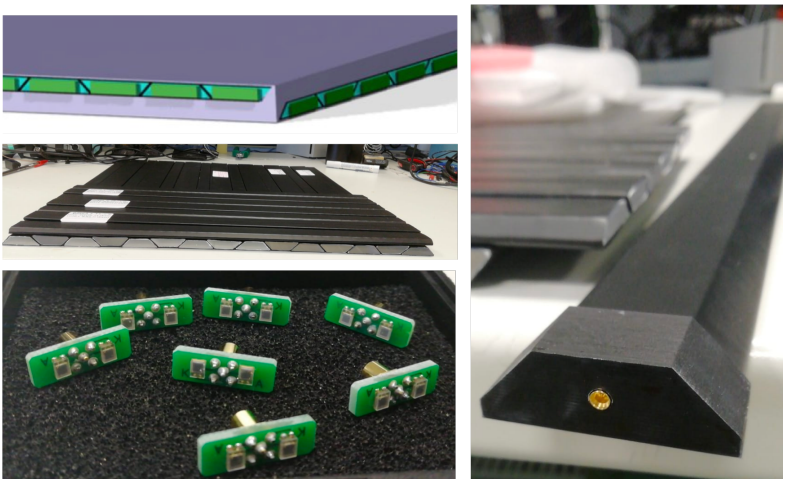


Fig. 6. Scheme of the PSD made with bars on the upper left. The PSD bar prototype undergoing the beam test at CERN on central left. A detail of the SiPMs used for the readout on the bottom left. A detail of the complete detection unit (the bar) undergoing the beam test at CERN on the right.

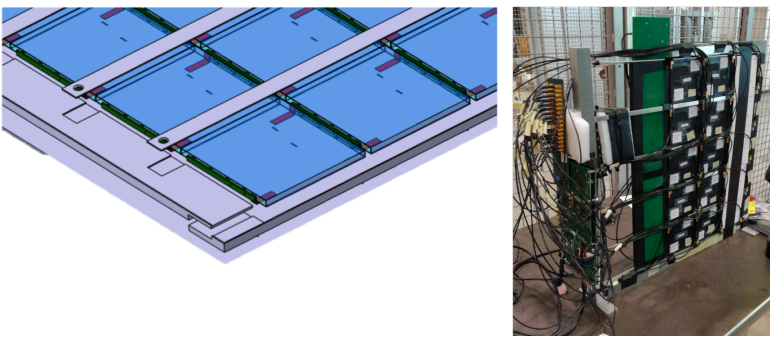


Fig. 7. Detail of the scheme of the PSD made with tiles on the left. The PSD tiles prototype undergoing the beam test at CERN on the right.

Fiber Tracker The FIT is the third detector hit by an impinging particle. It has several aims:

- reconstructing the trajectories and the charge absolute value (Z) of charged cosmic rays;
- favouring the conversion of low energy gamma rays and reconstructing the tracks of the generated electrons and positrons;
- acting as a redundance of the SCD and offering the possibility of cross-calibration among subdetectors.

In the current design,[15] each side of the FIT is made by 7 tracking planes, for 7 independent measurements of the position of a traversing charged particle, see FIG. 8. Each tracking plane consists of two layers measuring the two orthogonal spatial coordinates.

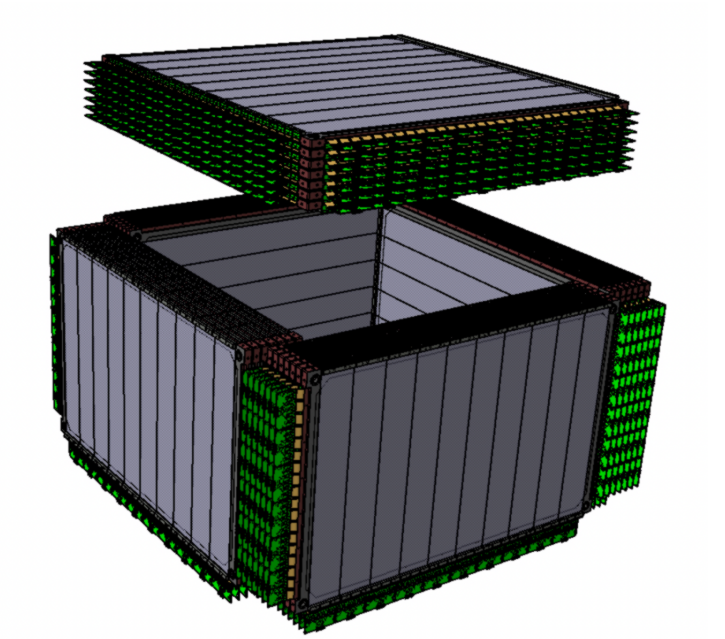


Fig. 8. Scheme of the FIT detector.

The tracking planes of the top sector are made of 10 FIT modules on both the x layers and y layers, while the tracking planes of the side sectors are made of 6 FIT modules on the x layers and 10 FIT modules on the y layers. A module includes one scintillating fiber mat and three SiPM-arrays to readout the scintillation light induced by the particles hitting the mat.

A FIT module prototype made of a 77 cm long fiber mat and one SiPM array was exposed to a 400 GeV/c primary proton beam and a fragmentation ion beam created by a 150 GeV/c lead beam hitting a beryllium target at CERN. A spatial resolution of 45 m and a mean hit efficiency of 99.6% were measured for protons. The response to nuclei from helium up to beryllium was studied and a charge resolution better than 15% was found.

The calorimeter The heart of the HERD instrument is the CALO: it is a homogeneous, isotropic, 3D segmented calorimeter accepting particles coming from each surface.[16] The performance of a detector based on these ideas was accurately studied by the CaloCube *R&D* project. By exploiting those results, the baseline design

of the CALO consists of about 7500 LYSO cubes with edge length of 3 cm, corresponding to about 2.6 radiation lengths (X_0) and 1.4 Moliere radius. The CALO external envelope is similar to an octagonal prism and the crystals are arranged on vertical layers with different dimension, see FIG. 9.

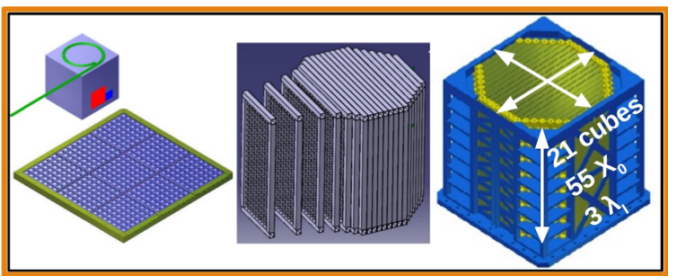


Fig. 9. Scheme of the CALO detector.

The total depth of the CALO for vertical particles is about 55 X_0 (radiation length) and $3\lambda_I$ (interaction length).

The scintillation light of each crystal is read-out by two independent systems: the first one consists of WaveLength Shifting fibers (WLS) coupled to image Intensified scientific CMOS (IsCMOS) cameras, the second one is made of photo-diodes (PD) connected to custom front-end electronics chips named HIDRA. This design, here named "double read-out system", achieves the capability of cross-calibrating the scintillation light measurement. Recent articles by calorimetric CR experiments speculated about the energy scale calibration as a possible source of unknown uncertainty, particularly for the electron flux measurement. Both the "double read-out" system and the TRD detector will strongly increase the understanding of HERD calorimeter energy scale with respect to previous CR detectors. Furthermore, both WLS and PD systems provide independent fast trigger information which will be employed to improve the HERD trigger capabilities.

The Transition Radiation Detector The TRD, installed on a lateral face of the detector, is needed to calibrate the response of the calorimeter to high energy hadronic showers. Ground calibration of space calorimeters, indeed, can only go to 400 GeV at maximum by using CERN SPS beam. For TeV protons, only few particles have showers all contained in the calorimeter and only visible deposited energy is recorded by detectors. Energy of primary proton is then derived by data extrapolation of low energy protons, which is not reliable.

A Transition Radiation Detector (TRD) could be used to calibrate the incident particle with energy in a range in which transition radiation is generated but not saturated. The energy of TR is proportional to the Lorentz factor γ of the incident

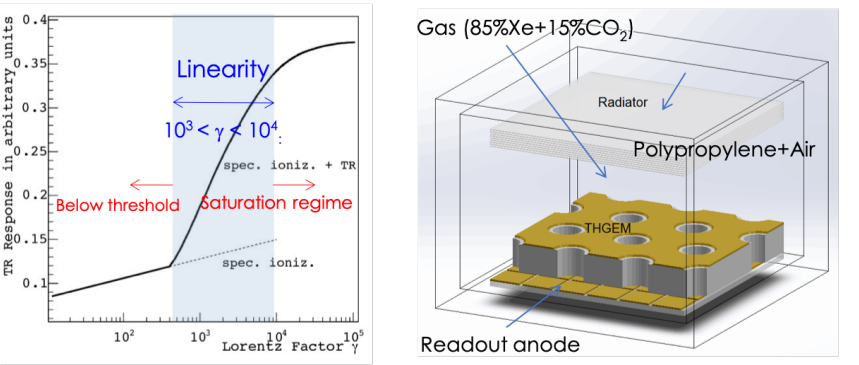


Fig. 10. Working principle of the TRD detector on the left. Scheme of the TRD detector on the right.

charged particle. The radiant photons will be generated at $\gamma{\sim}10^3$ and saturated at $\gamma{\sim}10^4$, the corresponding energy region of proton is about 1-10 TeV, see FIG. 10.

By detecting the X rays generated by TR, the absolute energy of the incident charged particle in the TeV region can be obtained by measuring the Lorentz factor.

The large effective area of TRD is achieved by a modular and highly redundant design. The TRD is arranged in 3 layers each consisting of 9 detection modules. Each detection module consists of a TR radiator, an X-ray gaseous detector with an effective area of $20{\times}20$ cm^2 and read-out electronics, supported by grid-like frames.

4. Summary

The HERD detector will be installed onboard CSS on 2027, with a lifetime of 5–10 years. The main scientific objectives regarding this initiative reside in the fields of CR physics, gamma-rays and (indirect) DM searches.

HERD will be capable of providing high-precision measurements including single-element spectral indexes along with spectral hardenings/softenings of nuclei and it will investigate the "knee" in hadronic CRs, for the first time via direct observations.

Moreover it will provide precise measurements of secondary-to-primary ratios (i.e., B/C) in an extended energy range,giving information on CR propagation in the Galaxy.

A great effort will also be invested in probing fine structures in the leptonic component $(e^+ + e^-)$, especially at the highest-achievable energies with great precision.

HERD will also be involved in high energy gamma-ray observations throughout a broad energy range due to its wide FOV, as well as assisting in observations of ground-based experiments with a narrow FOV.

Acknowledgements

The author would like to thank all members of the HERD Collaboration for supporting this work with valuable comments and fruitful discussions.

References

1. C. Evoli, *The Cosmic-Ray Energy Spectrum* (Zenodo, 2020).
2. O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, L. Bonechi, M. Bongi, V. Bonvicini, S. Borisov, S. Bottai, A. Bruno, F. Cafagna, D. Campana, R. Carbone, P. Carlson, M. Casolino, G. Castellini, L. Consiglio, M. P. D. Pascale, C. D. Santis, N. D. Simone, V. D. Felice, A. M. Galper, W. Gillard, L. Grishantseva, G. Jerse, A. V. Karelin, S. V. Koldashov, S. Y. Krutkov, A. N. Kvashnin, A. Leonov, V. Malakhov, V. Malvezzi, L. Marcelli, A. G. Mayorov, W. Menn, V. V. Mikhailov, E. Mocchiutti, A. Monaco, N. Mori, N. Nikonov, G. Osteria, F. Palma, P. Papini, M. Pearce, P. Picozza, C. Pizzolotto, M. Ricci, S. B. Ricciarini, L. Rossetto, R. Sarkar, M. Simon, R. Sparvoli, P. Spillantini, Y. I. Stozhkov, A. Vacchi, E. Vannuccini, G. Vasilyev, S. A. Voronov, Y. T. Yurkin, J. Wu, G. Zampa, N. Zampa and V. G. Zverev, Pamela measurements of cosmic-ray proton and helium spectra, *Science* **332**, 69 (2011).
3. F. Alemanno, Q. An, P. Azzarello, F. C. T. Barbato, P. Bernardini, X. J. Bi, M. S. Cai, E. Catanzani, J. Chang, D. Y. Chen, J. L. Chen, Z. F. Chen, M. Y. Cui, T. S. Cui, Y. X. Cui, H. T. Dai, A. D'Amone, A. De Benedittis, I. De Mitri, F. de Palma, M. Deliyergiyev, M. Di Santo, T. K. Dong, Z. X. Dong, G. Donvito, D. Droz, J. L. Duan, K. K. Duan, D. D'Urso, R. R. Fan, Y. Z. Fan, K. Fang, F. Fang, C. Q. Feng, L. Feng, P. Fusco, M. Gao, F. Gargano, K. Gong, Y. Z. Gong, D. Y. Guo, J. H. Guo, X. L. Guo, S. X. Han, Y. M. Hu, G. S. Huang, X. Y. Huang, Y. Y. Huang, M. Ionica, W. Jiang, J. Kong, A. Kotenko, D. Kyratzis, S. J. Lei, S. Li, W. L. Li, X. Li, X. Q. Li, Y. M. Liang, C. M. Liu, H. Liu, J. Liu, S. B. Liu, W. Q. Liu, Y. Liu, F. Loparco, C. N. Luo, M. Ma, P. X. Ma, T. Ma, X. Y. Ma, G. Marsella, M. N. Mazziotta, D. Mo, X. Y. Niu, X. Pan, A. Parenti, W. X. Peng, X. Y. Peng, C. Perrina, R. Qiao, J. N. Rao, A. Ruina, M. M. Salinas, G. Z. Shang, W. H. Shen, Z. Q. Shen, Z. T. Shen, L. Silveri, J. X. Song, M. Stolpovskiy, H. Su, M. Su, Z. Y. Sun, A. Surdo, X. J. Teng, A. Tykhonov, H. Wang, J. Z. Wang, L. G. Wang, S. Wang, X. L. Wang, Y. Wang, Y. F. Wang, Y. Z. Wang, Z. M. Wang, D. M. Wei, J. J. Wei, Y. F. Wei, S. C. Wen, D. Wu, J. Wu, L. B. Wu, S. S. Wu, X. Wu, Z. Q. Xia, H. T. Xu, Z. H. Xu, Z. L. Xu, Z. Z. Xu, G. F. Xue, H. B. Yang, P. Yang, Y. Q. Yang, H. J. Yao, Y. H. Yu, G. W. Yuan, Q. Yuan, C. Yue, J. J. Zang, F. Zhang, S. X. Zhang, W. Z. Zhang, Y. Zhang, Y. J. Zhang, Y. L. Zhang, Y. P. Zhang, Y. Q. Zhang, Z. Zhang, Z. Y. Zhang, C. Zhao, H. Y. Zhao, X. F. Zhao, C. Y. Zhou and Y. Zhu, Measurement of the cosmic ray helium energy spectrum from 70 gev to 80 tev with the dampe space mission, *Phys. Rev. Lett.* **126**, p. 201102 (May 2021).
4. J. Blumer, R. Engel and J. R. Horandel, Cosmic Rays from the Knee to the Highest Energies, *Prog. Part. Nucl. Phys.* **63**, 293 (2009).
5. I. Cholis, D. P. Finkbeiner, L. Goodenough and N. Weiner, The PAMELA positron excess from annihilations into a light boson, **2009**, 007 (Dec 2009).
6. L. Feng, R.-Z. Yang, H.-N. He, T.-K. Dong, Y.-Z. Fan and J. Chang, AMS-02 positron excess: new bounds on dark matter models and hint for primary electron spectrum hardening, *Phys. Lett. B* **728**, 250 (2014).

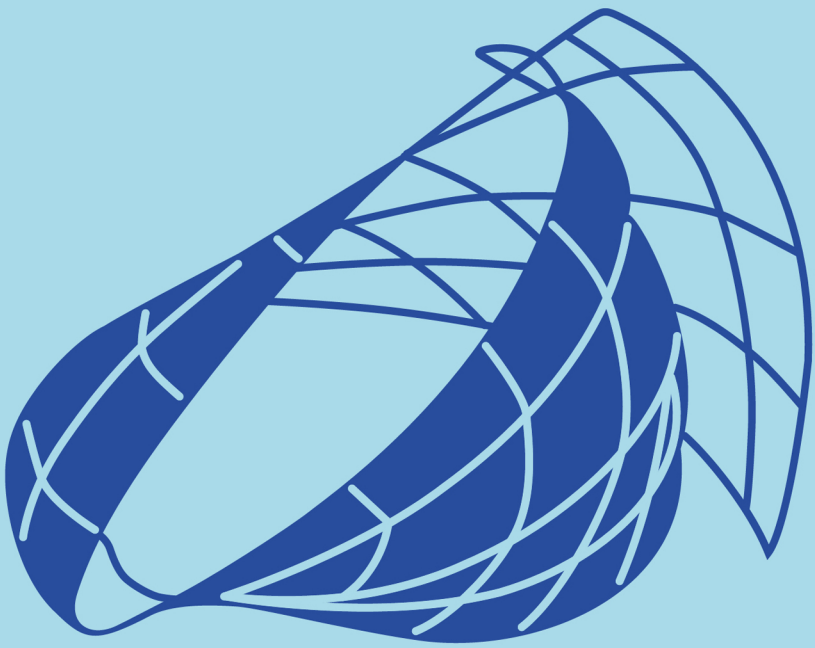
7. G. Ambrosi *et al.*, Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons, *Nature* **552**, 63 (2017).
8. R. C. Haight, M. B. Chadwick, T. Kawano and P. Talou, Front matter for volume 769, *AIP Conference Proceedings* **769**, p. frontmatter (2005).
9. S. D. Hunter, D. L. Bertsch, J. R. Catelli, T. M. Dame, S. W. Digel, B. L. Dingus, J. A. Esposito, C. E. Fichtel, R. C. Hartman, G. Kanbach, D. A. Kniffen, Y. C. Lin, H. A. Mayer-Hasselwander, P. F. Michelson, C. von Montigny, R. Mukherjee, P. L. Nolan, E. Schneid, P. Sreekumar, P. Thaddeus and D. J. Thompson, Egret observations of the diffuse gamma-ray emission from the galactic plane, *apj* **481**, 205 (May 1997).
10. G. A. Gómez-Vargas, M. A. Sánchez-Conde, J.-H. Huh, M. Peiró, F. Prada, A. Morselli, A. Klypin, D. G. Cerdeño, Y. Mambrini and C. Muñoz, Constraints on WIMP annihilation for contracted dark matter in the inner Galaxy with the Fermi-LAT, *Journal of Cosmology and Astroparticle Physics* **2013**, p. 029 (2013).
11. V. Vitale, A. Morselli *et al.*, Indirect search for dark matter from the center of the milky way with the Fermi-Large Area Telescope, *arXiv:0912.3828* (2009).
12. B. Zhou, Y.-F. Liang, X. Huang, X. Li, Y.-Z. Fan, L. Feng and J. Chang, GeV excess in the Milky Way: The role of diffuse galactic gamma-ray emission templates, *Physical Review D* **91**, p. 123010 (2015).
13. A. Achterberg, S. Amoroso, S. Caron, L. Hendriks, R. R. De Austri and C. Weniger, A description of the Galactic Center excess in the Minimal Supersymmetric Standard Model, *Journal of Cosmology and Astroparticle Physics* **2015**, p. 006 (2015).
14. D. Kyratzis, F. Alemanno, C. Altomare, F. Barbato, P. Bernardini, P. Cattaneo, I. D. Mitri, F. de Palma, L. D. Venere, M. D. Santo, P. Fusco, F. Gargano, F. Loparco, S. Loporchio, G. Marsella, M. Mazziotta, F. Pantaleo, A. Parenti, R. Pillera, A. Rappoldi, G. Raselli, M. Rossella, D. Serini, L. Silveri, A. Surdo and L. Wu, The plastic scintillator detector of the herd space mission, *Proceeding of Science* **ICRC 2021**.
15. C. P. et al., Fit: the scintillating fiber tracker of the herd space mission, *Proceeding of Science* **ICRC 2021**.
16. L. P. et al., Design and expected performances of the large acceptance calorimeter for the herd space mission, *Proceeding of Science* **ICRC 2021**.

This page intentionally left blank

PART D

PROCEEDINGS OF THE SIXTEENTH

MARCEL GROSSMANN MEETING ON GENERAL RELATIVITY



Editors

Remo Ruffini

Gregory Vereshchagin

World Scientific

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

This page intentionally left blank

PART D

THE SIXTEENTH
MARCEL GROSSMANN MEETING

On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories

Proceedings of the MG16 Meeting
on General Relativity
Online, 5–10 July 2021

Editors

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Gregory Vereshchagin
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

Series Editor

Remo Ruffini
University of Rome "La Sapienza", Rome, Italy
International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Names: Marcel Grossmann Meeting on General Relativity (16th : 2021 : Online) | Ruffini, Remo, editor. | Vereshchagin, Gregory, editor.

Title: The sixteenth Marcel Grossmann meeting on recent developments in theoretical and experimental general relativity, astrophysics, and relativistic field theories : proceedings of the MG16 meeting on general relativity online, 5–10 July 2021 / editors: Remo Ruffini, University of Rome "La Sapienza", Rome, Italy, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy, Gregory Vereshchagin, International Center for Relativistic Astrophysics Network (ICRANet), Pescara, Italy.

Other titles: Proceedings of the MG16 meeting on general relativity online, 5–10 July 2021

Description: New Jersey : World Scientific, [2023] | Includes bibliographical references.

Identifiers: LCCN 2022047088 | ISBN 9789811269769 (set ; hardcover) | ISBN 9789811266584 (v. 1) | ISBN 9789811266591 (v. 2) | ISBN 9789811266607 (v. 3) | ISBN 9789811266614 (v. 4) | ISBN 9789811269776 (set ; ebook)

Subjects: LCSH: General relativity (Physics)--Congresses. | Gravitation--Congresses. | Quantum gravity--Congresses. | Cosmology--Congresses. | Astrophysics--Congresses.

Classification: LCC QC173.6 .M37 2021 | DDC 523.01--dc23/eng/20220929

LC record available at https://lccn.loc.gov/2022047088

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2023 by Editors

All rights reserved.

This is an Open Access volume published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 (CC BY-NC) License. Further distribution of this work is permitted, provided the original work is properly cited.

For any available supplementary material, please visit
https://www.worldscientific.com/worldscibooks/10.1142/13149#t=suppl

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore

THE MARCEL GROSSMANN MEETINGS

Series Editor: REMO RUFFINI

Publications in the Series of Proceedings

Proceedings of the Sixteenth Marcel Grossmann Meeting on General Relativity
(Virtual Meeting, 2021)
Edited by G. Vereshchagin, R. Ruffini
World Scientific, 2022

Proceedings of the Fifteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2018)
Edited by E.S. Battistelli, R.T. Jantzen, R. Ruffini
World Scientific, 2022

Proceedings of the Fourteenth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2015)
Edited by M. Bianchi, R.T. Jantzen, R. Ruffini
World Scientific, 2017

Proceedings of the Thirteenth Marcel Grossmann Meeting on General Relativity
(Stockholm, Sweden, 2012)
Edited by K. Rosquist, R.T. Jantzen, R. Ruffini
World Scientific, 2015

Proceedings of the Twelfth Marcel Grossmann Meeting on General Relativity
(Paris, France, 2009)
Edited by T. Damour, R.T. Jantzen, R. Ruffini
World Scientific, 2012

Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity
(Berlin, Germany, 2006)
Edited by H. Kleinert, R.T. Jantzen, R. Ruffini
World Scientific, 2007

Proceedings of the Tenth Marcel Grossmann Meeting on General Relativity
(Rio de Janiero, Brazil, 2003)
Edited by M. Novello, S. Perez-Bergliaffa, R. Ruffini
World Scientific, 2005

Proceedings of the Ninth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 2000)
Edited by V.G. Gurzadyan, R.T. Jantzen, R. Ruffini
World Scientific, 2002

Proceedings of the Eighth Marcel Grossmann Meeting on General Relativity
(Jerusalem, Israel, 1997)
Edited by T. Piran
World Scientific, 1998

Proceedings of the Seventh Marcel Grossmann Meeting on General Relativity
(Stanford, USA, 1994)
Edited by R.T. Jantzen and G.M. Keiser
World Scientific, 1996

Proceedings of the Sixth Marcel Grossmann Meeting on General Relativity
(Kyoto, Japan, 1991)
Edited by H. Sato and T. Nakamura
World Scientific, 1992

Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity
(Perth, Australia, 1988)
Edited by D.G. Blair and M.J. Buckingham
World Scientific, 1989

Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity
(Rome, Italy, 1985)
Edited by R. Ruffini
World Scientific, 1986

Proceedings of the Third Marcel Grossmann Meeting on General Relativity
(Shanghai, People's Republic of China, 1982)
Edited by Hu Ning
Science Press – Beijing and North-Holland Publishing Company, 1983

Proceedings of the Second Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1979)
Edited by R. Ruffini
North-Holland Publishing Company, 1982

Proceedings of the First Marcel Grossmann Meeting on General Relativity
(Trieste, Italy, 1975)
Edited by R. Ruffini
North-Holland Publishing Company, 1977

SPONSORS

International Center for Relativistic Astrophysics Network (ICRANet)
International Center for Relativistic Astrophysics (ICRA)

FREEDOM OF MOVEMENT FOR SCIENTISTS

The Marcel Grossmann Meetings were founded with the premise that scientists of all nations have a right to meet to exchange knowledge independent of national borders.

ACKNOWLEDGEMENTS

We acknowledge the outstanding job done before, during and after the meeting by the ICRANet/ICRA administrative and secretarial staff: Cristina Adamo, Silvia Latorre, Elisabetta Natale, and Cinzia di Niccolo. Finally this meeting and its proceedings could not have functioned without the dedicated IT support of the ICRANet system manager Gabriele Brandolini, with some temporary assistance from Domenico La Selva and Damiano Verzulli. We would like to thank Linda Kwan from World Scientific for valuable assistance during preparation of these proceedings.

ORGANIZING BODIES OF THE SIXTEENTH MARCEL GROSSMANN MEETING

INTERNATIONAL ORGANIZING COMMITTEE

Blair David, Choquet Bruhat Yvonne, Damour Thibault, De Bernardis Paolo, Everitt C. W. Francis, Fryer Chris, Haensch Theodor, Henneaux Marc, Jones Christine, Kerr Roy, Kleinert Hagen, Kunz Jutta, Laemmerzahl Claus, Longair Malcolm, Mirabel Felix, Mirzoyan Razmik, Piran Tsvi, Rueda Jorge, Ruffini Remo (chair), Sasaki Misao, Sato Humitaka, Sunyaev Rashid, 't Hooft Gerard, Weinberg Steven, Yau Shing-Tung, Zhang Bing

LOCAL ORGANIZING COMMITTEE

Adamo Cristina, Bianco Carlo Luciano, Brandolini Gabriele A., di Niccolo Cinzia, Latorre Silvia, La Selva Domenico, Li Liang, Loppini Alessandro, Natale Elisabetta, Verzulli Damiano, Vereshchagin Gregory (chair), Wang Yu

INTERNATIONAL COORDINATING COMMITTEE

ALBANIA: Hafizi M. - ARGENTINA: Arguelles C., Scoccola C., Reula O., Romero G.E. - ARMENIA: Sahakyan N. - AUSTRALIA: Blair D., Ju L., Lun A., Manchester D., Melatos A., Quinn P., Scott S.M., Steele J.D. - AUSTRIA: Aichelburg P.C., Schindler S. - BELARUS: Kilin S., Prakapenia M., Siutsou I. - BELGIUM: Henneaux M. - BOLIVIA: Aguirre C.B. - BOSNIA: Pasic V. - BRAZIL: Barres de Almeida U., Coelho Goulart J., Dalmolin F.T., de Lima Rafael C.R., Guzzo M., Maia C., Malheiro M., Romero Filho C.A., Shellard R.C., Zen Vasconcellos C. - BULGARIA: Yazadjiev S. - CANADA: Singh D., Smolin L., Turok N. - CHILE: Bauer F., Bunster W.C., Giacomini A. - CHINA (MAINLAND): Cai R., Cai Y., Cao Z., Chang J., Chen J., Chen X., Dai Z., Feng L.-L., Han W., Jing Y., Li T.-P., Lin W., Lou Y.-Q., Luo J., Mei J., Tam T., Wang A., Wang Y., Wu X.-P., Wu Y.-L., Yuan Y.-F., Zhang B.-B., Zhang S.-N., Zhao G. - CHINA (TAIWAN): Chen Chiang-Mei, Chen Pisin, Lee Da-Shin, Lee Wo-Lung, Ni Wei-Tou - COLOMBIA: Bargueño de Retes P., Gonzalez G., Higuera Garzon M.A., Núñez L., Romano A.E., Valenzuela Toledo C.A., Zuluaga J.I. - CROATIA: Dominis Prester D., Karlica M., Milekovic M., Smolcic V., Smolic I., Suric T. - CUBA: Perez Martinez A., Pérez Rojas H. - CZECH REPUBLIC: Bicak J., Stuchlik Z. - DENMARK: Naselsky P. - ECUADOR: Contreras E. - EGYPT: Tawfik A.N., Wanas M.I. - ESTONIA: Einasto J., Saar E. - FINLAND: Poutanen J., Volovik G. - FRANCE: Brillet A., Buchert T., Chardonnet P., Coullet P., de Freitas Pacheco J.A., Deruelle N., Iliopoulos J., Lamanna G., Mignard F. - GEORGIA: Lavrelashvili George, Machabeli Giorgi - GERMANY: Biermann P., Blumlein J., Di Piazza A., Fritzsch H., Genzel R., Gilmozzi R., Hasinger G., Hehl F., Keitel C., Kiefer C., Mirzoyan R.,

Neugebauer G., Nicolai H., Renn J., Ringwald A., Ruediger A. - GREECE: Batakis N.A., Cotsakis S., Vagenas E.C. - HUNGARY: Fodor G., Levai P. - ICELAND: Bjornsson G., Jakobsson P. - INDIA: Chakrabarti S.K., Iyer B., Padmanabhan T., Souradeep T. - IRAN: Baghram S., Bavarsad E., Eslam Panah B., Firouzjahi H., Haghighat M., Mansouri R., Mashhoon B., Shakeri S., Sobouti Y., Taghi Mirtorabi M. - IRELAND: O'Murchada N. - ISRAEL: Milgrom M., Nakar E., Pe'er A., Piran T. - ITALY: Belinski V., Bianchi M., Bianco C.L., Cherubini C., Della Valle M., Falciano S., Filippi S., Haardt F., Menotti P., Merafina M., Pani P., Ricci F., Treves A., Vereshchagin G.V., Vitale S., Xue S.- S. - JAPAN: Fujimoto M.-K., Makishima K., Nakamura T., Sato K., Shibata M. - KAZAKHSTAN: Abishev M., Aimuratov Y., Boshkayev K., Mychelkin E.G., Spitaleri C. - KOREA (PYEONGYANG): Kim J.S. - KOREA (SEOUL): Kim S.P., Kim S.-W., Lee H.K., Lee H.-W., van Putten M. - KYRGYZSTAN: Gurovich V.Ts. - LIBYA: Gadri M. - MEXICO: Breton N., Cervantes-Cota J.L., Fraija Cabrera N.I., García-Diaz A.A., Macías Alvarez A., Mielke Eckehard W., Quevedo H., Rodriguez L.F. - NETHERLANDS: Slagter R. - NEW ZEALAND: Visser M., Wiltshire D. - NORWAY: Elgaroy O., Fonseca Mota D., Knutsen H. - PAKISTAN: Qadir A., Qamar S. - PERU: Vargas T. - POLAND: Belczynski K., Demianski M., Lewandowski Jerzy, Nurowski P., Sokolowski L. - PORTUGAL: Costa M., Da Silva A., Lemos J.P.S., Lobo F., Moniz P., Silva L.O. - ROMANIA: Visinescu M. - RUSSIA: Aksenov A., Arkhangelskaja I., Bisnovatyi-Kogan G., Blinnikov S., Chechetikin V.M., Cherepaschuk A.M., Khriplovich I., Lipunov V.M., Lukash V.N., Novikov I., Rudenko V.N., Starobinsky A.A. - SERBIA: Djordjevic G., Jovanovic P., Knezevic Z., Pankov-Hzvojevic M., Popovic L., Prodanovic T., Sijacki D., Simic S. - SLOVAKIA: Balek V. - SLOVENIA: Cadez A., Gomboc A., Zavrtanik D. - SOUTH AFRICA: Larena J., Maharaj S. - SPAIN: Elizalde E., Ibanez J., Perez M.J., Verdaguer E. - SWEDEN: Abramowicz M.A., Marklund M., Ryde F. - SWITZERLAND: Durrer R., Jetzer P. - TURKEY: Aliev A., Gurses M. - UKRAINE: Novosyadlyj B., Zaslavskii O., Zhuk A. - UNITED ARAB EMIRATES: Fernini I. - UNITED KINGDOM: Cruise A.M., Frenk Carlos S., Green M., Mavromatos N., Perry M., Willingale R. - USA: Abel T., Ashtekar A., Bardeen J., Carlstrom J., Cornish N., Dermer C., Fan X., Flanagan E., Fraschetti F., Fryer C., Incera V., Jantzen R.T. (chairperson), Kolb R., Laguna P., Lousto C., Madau Piero, Mathews Grant, Matzner Richard, Melia Fulvio, Mester John, Michelson Peter, Nordtvedt Kenneth, Parker Leonard, Pretorius F., Pullin J., Shapiro I., Shapiro S., Shoemaker D., Smoot G., Stiavelli M., Teukolsky S., van Nieuwenhuizen P., Zhang B. - UZBEKISTAN: Ahmedov B., Zalaletdinov R.M. - VATICAN CITY: Gionti G. - VENEZUELA: Fuenmayor E. - VIETNAM: Long H.N.

MARCEL GROSSMANN AWARDS

Sixteenth Marcel Grossmann Meeting

Institutional Awards

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

Individual Awards

DEMETRIOS CHRISTODOULOU
"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

GERARD 't HOOFT
"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

TSVI PIRAN
"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".

STEVEN WEINBERG
"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Each recipient is presented with a silver casting of the TEST sculpture by the artist A. Pierelli. The original casting was presented to His Holiness Pope John Paul II on the first occasion of the Marcel Grossmann Awards.

15th Marcel Grossmann Meeting
July 2018, Rome, Italy

Institutional Awards

PLANCK SCIENTIFIC COLLABORATION (ESA)
"for obtaining important constraints on the models of inflationary stage of the Universe and level of primordial non-Gaussianity; measuring with unprecedented sensitivity gravitational lensing of Cosmic Microwave Background fluctuations by large-scale structure of the Universe and corresponding B-polarization of CMB, the imprint on the CMB of hot gas in galaxy clusters; getting unique information about the time of reionization of our Universe and distribution and properties of the dust and magnetic fields in our Galaxy"

- presented to Jean-Loup Puget, the Principal Investigator of the High Frequency Instrument (HFI)

HANSEN EXPERIMENTAL PHYSICS LABORATORY AT STANFORD UNIVERSITY
"to HEPL for having developed interdepartmental activities at Stanford University at the frontier of fundamental physics, astrophysics and technology"

- presented to Research Professor Leo Hollberg, HEPL Assistant Director

Individual Awards

LYMAN PAGE
"for his collaboration with David Wilkinson in realizing the NASA Explorer WMAP mission and as founding director of the Atacama Cosmology Telescope"

RASHID ALIEVICH SUNYAEV
"for the development of theoretical tools in the scrutinising, through the CMB, of the first observable electromagnetic appearance of our Universe"

SHING-TUNG YAU
"for the proof of the positivity of total mass in the theory of general relativity and perfecting as well the concept of quasi-local mass, for his proof of the Calabi conjecture, for his continuous inspiring role in the study of black holes physics"

14th Marcel Grossmann Meeting
July 2015, Rome, Italy

Institutional Award

EUROPEAN SPACE AGENCY (ESA)
"for the tremendous success of its scientific space missions in astronomy, astrophysics, cosmology and fundamental physics which have revolutionized our knowledge of the Universe and hugely benefited science and mankind"

- presented to its Director General Johann-Dietrich Woerner

Individual Awards

KEN'ICHI NOMOTO
"for heralding the role of binary systems in the evolution of massive stars"

MARTIN REES
"for fostering Research in black holes, gravitational waves and cosmology"

YAKOV G. SINAI
"for applying the mathematics of chaotic systems to physics and cosmology"

SACHIKO TSURUTA
"for pioneering the physics of hot neutron stars and their cooling"

FRANK C.N. YANG
"for deepening Einstein's geometrical approach to physics in the best tradition of Paul Dirac and Hermann Weyl"

T.D. LEE (award received by Yu-Qing Lou on behalf of Prof. T.D. Lee)
"for his work on white dwarfs motivating Enrico Fermi's return to astrophysics and guiding the basic understanding of neutron star matter and fields"

13th Marcel Grossmann Meeting
July 2012, Stockholm, Sweden

Institutional Award

ALBANOVA
for its innovative status as a joint institute established by Stockholm University and the Royal Institute of Technology and for fostering contributions to cosmology and astrophysics in the profound scientific tradition established by Oskar Klein.

- presented to the Rector of Stockholm University, Prof. Kåre Bremer.

Individual Awards

DAVID ARNETT
for exploring the nuclear physics and yet unsolved problems of the endpoint of thermonuclear evolution of stars, leading to new avenues of research in physics and astrophysics.

VLADIMIR BELINSKI and I.M. KHALATNIKOV
for the discovery of a general solution of the Einstein equations with a cosmological singularity of an oscillatory chaotic character known as the BKL singularity.

FILIPPO FRONTERA
for guiding the Gamma-ray Burst Monitor Project on board the BeppoSAX satellite, which led to the discovery of GRB X-ray afterglows, and to their optical identification.

12th Marcel Grossmann Meeting
July 2009, Paris, France

Institutional Award

INSTITUT DES HAUTES ÉSTUDES SCIENTIFIQUE (IHÉS)
for its outstanding contributions to mathematics and theoretical physics, and notably for having renewed basic geometrical concepts, and having developed new mathematical and physical aspects of spacetime.

- presented to Prof. Jean-Pierre Bourguignon

Individual Awards

JAAN EINASTO
for pioneering contributions in the discovery of dark matter and cosmic web and fostering research in the historical Tartu Observatory.

CHRISTINE JONES
for her fundamental contributions to the X-ray studies of galaxies and clusters tracing their formation and evolution and for her role in collaborations using clusters to study dark matter and in analyzing the effects of outbursts from supermassive black holes on the intracluster gas.

MICHAEL KRAMER
for his fundamental contributions to pulsar astrophysics, and notably for having first confirmed the existence of spin-orbit precession in binary pulsars.

11th Marcel Grossmann Meeting
July 2006, Berlin, Germany

Institutional Award

FREIE UNIVERSITÄT BERLIN
for the successful endeavor of re-establishing — in the spirit of the Humboldt tradition — freedom of thinking and teaching within a democratic society in a rapidly evolving cosmos

- presented to Dr. Dieter Lenzen, President of FUB

Individual Awards

ROY KERR
for his fundamental contribution to Einstein's theory of general relativity: "The gravitational field of a spinning mass as an example of algebraically special metrics."

GEORGE COYNE
for his committed support for the international development of relativistic astrophysics and for his dedication to fostering an enlightened relationship between science and religion.

JOACHIM TRUMPER
for his outstanding scientific contributions to the physics of compact astrophysical objects and for leading the highly successful ROSAT mission which discovered more than 200,000 galactic and extragalactic X-ray sources: a major step in the observational capabilities of X-ray astronomy and in the knowledge of our universe.

10th Marcel Grossmann Meeting
July 2003, Rio de Janeiro, Brazil

Institutional Award

CBPF (Brazilian Center for Research in Physics)
for its role as a teaching and research institution and as a place originating fundamental physics ideas in the exploration of the universe.

- presented to its founders Cesar Lattes, Josè Leite Lopez and Jayme Tiomno

Individual Awards

YVONNE CHOQUET-BRUHAT AND JAMES W. YORK, JR.
for separate as well as joint work in establishing the mathematical framework for proving the existence and uniqueness of solutions to Einstein's gravitational field equations.

YUVAL NE'EMAN
for his contributions to science, epistimology, mathematics and physics from subnuclear to space sciences.

9th Marcel Grossmann Meeting
July 2000, Rome, Italy

Institutional Award

SOLVAY INSTITUTES
for identifying and recording in discussions by the protagonists the crucial developments of physics and astrophysics in the twentieth century.

- presented to Jacques Solvay

Individual Awards

CECILLE AND BRYCE DEWITT
for promoting General Relativity and Mathematics research and inventing the "summer school" concept.

RICCARDO GIACCONI
for opening, five successive times, new highways for exploring the Universe.

ROGER PENROSE
for extending the mathematical and geometrical foundations of General Relativity.

8th Marcel Grossmann Meeting
June 1997, Jerusalem

Institutional Award

HEBREW UNIVERSITY
for its role as a cradle of Science and Humanities and for hosting the manuscripts of Albert Einstein.

- presented to M. Magidor, President of the Hebrew University of Jerusalem

Individual Awards

TULLIO REGGE
for his contributions to the interface between mathematics and physics leading to new fields of research of paramount importance in relativistic astrophysics and particle physics.

FRANCIS EVERITT
for leading the development of extremely precise space experiments utilizing superconducting technology to test General Relativity and the Equivalence Principle.

7th Marcel Grossmann Meeting
June 1994, Stanford, USA

Institutional Award

SPACE TELESCOPE SCIENCE INSTITUTE
for its critical role in the direction and operation of the Hubble Space Telescope, a truly unique international laboratory for the investigation and testing of general relativity in the context of modern astrophysics and cosmology.

- presented to Peter Stockman

Individual Awards

SUBRAHMANYAN CHANDRASEKHAR
for his contributions to the analysis of gravitational phenomena from Newton to Einstein and especially for leading the way to relativistic astrophysics with the concept of critical mass for gravitational collapse.

JIM WILSON
for having built on his experience in nuclear physics, thermonuclear reactions, and extensive numerical simulation to create a new testing ground for the novel concepts of relativistic astrophysics.

6th Marcel Grossmann Meeting
June 1991, Kyoto, Japan

Institutional Award

RITP
for keeping alive first in Hiroshima and them in Kyoto research in relativity, cosmology, and relativistic field theory and the development of a school of international acclaim.

- presented to Professor K. Tomita

Individual Awards

MINORU ODA
for participating in the pioneering work of the early sixties in X-ray astronomy and for his subsequent molding of an agile and diversified Japanese scientific space program investigating the deepest aspects of relativistic astrophysics.

STEPHEN HAWKING
for his contributions to the understanding of spacetime singularities and of the large scale structure of the Universe and of its quantum origins.

5th Marcel Grossmann Meeting
August 1988, Perth, Australia

Institutional Award

THE UNIVERSITY OF WESTERN AUSTRALIA
for its contributions to relativistic astrophysics.

- presented to the Vice Chancellor, Professor Robert Smith

Individual Awards

SATIO HAYAKAWA
for his contributions to research in gamma, X-ray and infrared radiation as well as cosmic rays.

JOHN ARCHIBALD WHEELER
for his contributions to geometrodynamics and Einstein's visions.

4th Marcel Grossmann Meeting
July 1985, Rome, Italy

Institutional Award

THE VATICAN OBSERVATORY
for its contributions to the origin and development of astrophysics.

- presented to His Holiness Pope John Paul II

Individual Awards

WILLIAM FAIRBANK
for his work in gravitation and low temperature physics.

ABDUS SALAM
for his work in unifying fundamental interactions.

Institutional Awards for the Spektrum-Roentgen-Gamma (SRG) mission

"for the creation of the world's best X-ray map of the entire sky, for the discovery of millions of previously unknown accreting supermassive black holes at cosmological redshifts, for the detection of X-rays from tens of thousands of galaxy clusters, filled mainly with dark matter, and for permitting the detailed investigation of the growth of the large-scale structure of the universe during the era of dark energy dominance".

S.A. LAVOCHKIN ASSOCIATION
- presented to its Designer General **Alexander Shirshakov**

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
- presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
- presented to Professor **Rashid Sunyaev,** Principal Investigator of SRG Observatory in Russia

On Tuesday June 29, 2021, the following 31 astro-ph appeared:

1. https://arxiv.org/abs/2106.14517
2. https://arxiv.org/abs/2106.14518
3. https://arxiv.org/abs/2106.14519
4. https://arxiv.org/abs/2106.14520
5. https://arxiv.org/abs/2106.14521
6. https://arxiv.org/abs/2106.14522
7. https://arxiv.org/abs/2106.14523
8. https://arxiv.org/abs/2106.14524
9. https://arxiv.org/abs/2106.14525
10. https://arxiv.org/abs/2106.14526
11. https://arxiv.org/abs/2106.14527
12. https://arxiv.org/abs/2106.14528
13. https://arxiv.org/abs/2106.14529
14. https://arxiv.org/abs/2106.14530
15. https://arxiv.org/abs/2106.14531
16. https://arxiv.org/abs/2106.14532
17. https://arxiv.org/abs/2106.14533
18. https://arxiv.org/abs/2106.14534
19. https://arxiv.org/abs/2106.14535
20. https://arxiv.org/abs/2106.14536
21. https://arxiv.org/abs/2106.14537
22. https://arxiv.org/abs/2106.14541
23. https://arxiv.org/abs/2106.14542
24. https://arxiv.org/abs/2106.14543
25. https://arxiv.org/abs/2106.14544
26. https://arxiv.org/abs/2106.14545
27. https://arxiv.org/abs/2106.14546
28. https://arxiv.org/abs/2106.14547
29. https://arxiv.org/abs/2106.14548
30. https://arxiv.org/abs/2106.14549
31. https://arxiv.org/abs/2106.14550

S.A. LAVOCHKIN ASSOCIATION
presented to its Designer General **Alexander Shirshakov**

Dr Alexander Shirshakov

S.A. Lavochkin Association created the Navigator space platform carrying German eRosita and Russian ART-XC X-Ray Telescopes, organized the launch of SRG Orbital X-Ray Observatory to the second Lagrangian point of the Sun-Earth system at a distance of 1.5 million km from the Earth and managed the observatory flight and the daily reception of its scientific data on Earth for 23.5 months.

Dr Alexander Shirshakov, Designer General of the S.A. Lavochkin Association, is specialized in design, manufacture, testing, launch and control of S/C for scientific purposes. Among those S/C launched, there are the «Radiostron» Astrophysical Observatory (2011) and the «Spektr-RG» space observatory (2019), while the planned S/C launches are «Luna-25» and «Exomars».

Dr Shirshakov started his career in 1973, working as an engineer of the State Unitary Enterprise «NPO named by S.A. Lavochkin» in Khimki (Russian Federation). Starting from 1989 he has played multiple roles within the Lavochkin Association, been appointed head of the group, head of the sector, head of department, deputy head of the complex, head of the branch, director of the center, deputy head of the Design Bureau, deputy General Designer and deputy General Director.

Dr Shirshakov is an editorial board Member of the reviewed edition of «Vestnik of Lavochkin Association». Since 2017, he is also member of the General Designer council. He has been awarded Honored Mechanical engineer of the Russian Federation as well as Agency-level award of the Russian Federal Space Agency.

MAX PLANCK INSTITUTE FOR EXTRATERRESTRIAL PHYSICS (MPE)
presented to Professor **Peter Predehl**, Principal Investigator of eROSITA

Professor Peter Predehl

eROSITA is the soft X-ray telescope on-board the Russian-German Spektr-RG mission which was successfully launched from Baikonur on July 13, 2019 and placed in a halo orbit around the L2 point. 30 years after ROSAT, eROSITA performs an all-sky survey with an unprecedented sensitivity, spectral and angular resolution. Clusters of galaxies are the largest collapsed objects in the Universe. Their formation and evolution is dominated by gravity, i.e. Dark Matter, while their large scale distribution and number density depends on the geometry of the Universe, i.e. Dark Energy. X-ray observations of clusters of galaxies provide information on the rate of expansion of the Universe, the fraction of mass in visible matter, and the amplitude of primordial fluctuations which are the origin of clusters of galaxies and the whole structure of the universe. eROSITA has been designed to detect at least 100.000 clusters of galaxies and to detect systematically more than 3 million obscured accreting Black Holes. eROSITA will also allow to study the physics of galactic X-ray source

populations, like pre-main sequence stars, supernova remnants and X-ray binaries. The eROSITA telescope consists of seven identical Wolter-1 mirror modules. A novel detector system has been developed by MPE on the basis of the successful XMM-Newton pn-CCD technology. MPE is the scientific lead institute of eROSITA, responsible for the development of the instrument, the operation, the analysis software and data archive. Peter Predehl led this development as Principal Investigator of eROSITA and German lead scientist of the SRG mission for more than 15 years until the completion of the first of eight surveys in 2020. At this time eROSITA has already discovered more than 1 million X-ray sources, more than all X-ray observatories of the last 50 years together. This demonstrates that the design goals of the mission will easily be fulfilled.

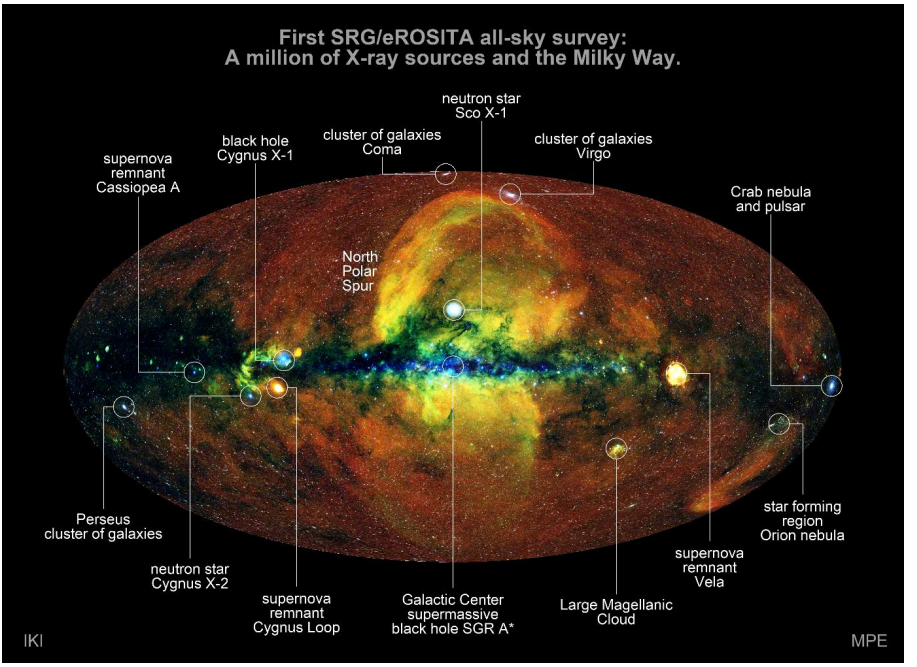
SPACE RESEARCH INSTITUTE (IKI) OF THE RUSSIAN ACADEMY OF SCIENCES
presented to Professor **Rashid Sunyaev**

Professor Rashid Sunyaev

Space Research Institute (IKI) of the Russian Academy of Sciences was responsible for developing the overall concept and scientific program of the SRG Orbital observatory and played a leading role in developing the ART-XC telescope and the entire SRG observatory as part of the Russian space science program carried out by Roskosmos Corporation in the interests of the Russian Academy of Sciences.

During the flight to the L2 point of the Sun-Earth system, SRG with German (eRosita) and Russian (ART-XC named after Mikhail Pavlinsky) X-ray Telescopes aboard performed calibrations and long duration Performance Verification observations of a dozen of targets and deep fields. Starting in the middle of December 2019, the SRG scanned the whole sky three times. During these scans, SRG discovered two million point X-ray sources: mainly quasars, stars with hot and bright coronae, and more than 30 thousand clusters of galaxies. There is a competition and synergy in the search for clusters of galaxies between SRG and the ground-based Atacama Cosmology and South Pole Telescopes, which are searching for clusters of galaxies in microwave spectral band using Sunyaev-Zeldovich effect. SRG provided the X-Ray map of the whole sky in hard and soft bands, the last is now the best among existing. The huge samples of the X-ray selected quasars at the redshifts up to z = 6.2 and clusters of galaxies will be used for well-known cosmological tests and detailed study of the growth of the large scale structure of the Universe during and after reionization. SRG/eRosita is discovering every day several extragalactic objects which increased or decreased their brightness more than 10 times during half of the year after the previous scan of the same one-degree wide strip on the sky. A significant part of these objects has observational properties similar to the Events of Tidal Disruption of a star orbiting in the vicinity of the supermassive black hole. ART-XC discovered a lot of bright galactic and extragalactic transients.

Rashid Sunyaev is the Principal Investigator of SRG mission in Russia, director-emeritus of the Max-Planck Institute for Astrophysics and Maureen and John Hendricks distinguished visiting professor of the Institute for Advanced Study, Princeton.



Individual Awards

Professor **DEMETRIOS CHRISTODOULOU**

"For his many lasting contributions to the foundation of mathematical physics including the dynamics of relativistic gravitational fields. Notably for: contributing in 1971, at the age of 19, to derive with Remo Ruffini the mass-energy formula of black holes as a function of their angular momentum, charge and irreducible mass. Christodoulou turned then to the study of partial differential equations and mathematical physics, to which he remained dedicated for the rest of his career. Highlights in this area include the theoretical discovery of the nonlinear memory effect of gravitational waves (Phys. Rev. Letters 1991), the monograph (1993) in collaboration with Sergiu Klainerman on the global nonlinear stability of the Minkowski spacetime, the monograph (2009) on the formation of black holes in pure general relativity by imploding gravitational waves, and the monographs (2007 and 2019) on the formation and further development of shocks in fluids."

Professor Demetrios Christodoulou

It was back in 1967 that Achille Papapetrou mentioned the case of the 16-year-old Demetrios Christodoulou to John Archibad Wheeler. Wheeler interviewed Demetrios in Paris and brought him immediately to Princeton where he was registered as an undergraduate at the university. After one year he entered the graduate school and started collaborating with me. At the time I was working with Wheeler on the effective potential approach to geodesics co-rotating and counter-rotating (see e.g. reference in The Classical Theory of Fields (Landau and Lifshitz, 1980) in the Kerr metric (later renamed as ISCO; see e.g. (Gravitation Misner, Thorne, Wheeler. 1973). In parallel, Frank Zerilli was working on the gravitational radiation emitted by the fall of a test particle in a Schwarzschild black hole (Zerilli 1970). From these limited conceptual arena Charles Misner and later Kip Thorne launched a program for the detection of gravitational waves on the Earth; see e.g. Misner 1974, Abbott et al. 2016, Abbott et al. 2017. See however Davis et al. 1972, Rodriguez et al. 2018 and J.A. Rueda et al. 2018.

A new approach started with the arrival of Demetrios: he was just creating mathematics following his needs. We identified the reversible and irreversible transformations of a Kerr black hole. Wheeler advanced a thermodynamic analogy. I addressed the need of identifying the concept of irreducible mass (from the Italian "irriducibile"), and was Demetrios's contribution to integrate, overnight, the differential equation for infinitesimal reversible transformations which led to the finite mass-energy formula of a Kerr black hole. That evening, while walking back home through IAS woods, I expressed to Wheeler the great relevance of the newly found formula by Demetrios and proposed to let Demetrios be the single author of this article, admiring his great mathematical talent. Wheeler agreed. The Editor of PRL objected since in that two pages article the Fig. 2 by Wheeler and myself was still unpublished. Actually that Fig. 2 followed a discussion I previously had with Penrose in Florence (Penrose 1961) which allowed us to present there, for the first time, a "Penrose Process". Some difficulties in achieving this process were obvious from the example in Fig. 2, which Roger later recognized himself (Penrose & Floyd 1971). The Editor finally agreed on our written request and the paper appeared on September 17, 1970 (Christodoulou, 1970). On January 1971 appeared my article with Johnny introducing the Black Hole (Ruffini & Wheeler, 1971), with the new physics we were developing in Princeton, including the concept of the "ergosphere". On march 1 1971 we submitted the mass formula of the Kerr Newmann metric, including the relation between the surface area of the horizon and the irreducible (Christodoulou & Ruffini, 1971). On March 11, 1971 the same results were independently confirmed by Steven Hawking, extending further the applicability of our equation (Hawking 1971).

The thesis was successfully discussed by a committee including Eugene Wigner (see Fig. 1), one of the closest collaborators of Albert Einstein and David Wilkinson (see Fig. 2), the head of the NASA WMAP mission, and Johnny and myself as supervisors. The new message was clear: Black Holes, far from being a sink of energy, were energy sources emitting "in principle" 50% of their mass energy, being extractable (Christodoulou & Ruffini, 1971).

Fig. 1 and Fig. 2: Demetrios during his thesis presentation with Eugene Wigner (Fig. 1) and David Wilkinson (Fig.2). Johnny and I were supervisors, ready to intervene in case of need, but no need of intervention was necessary! Wigner elaborated the aphorism of Niels Bohr "Interesting = wrong" in the most definite "very interesting if true = totally wrong".

Demetrios turned soon to the study of partial differential equations and mathematical physics, to which he dedicated for the rest of his career and results were published in four monographs: (Christodoulou and Klainerman 1994, Christodoulou 2007, Christodoulou 2009, Christodoulou 2019). In 1968, Johnny proposed to Demetrios the collapse of a "geon" composed of massless scalar field as a second topic for his thesis. It took almost forty years for him to solve this problem, extended by Demetrios to the focusing of gravitational waves leading to black hole formation (Christodoulou 2009).

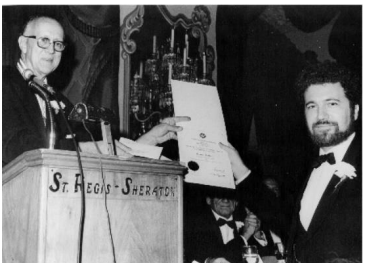


Fig. 3: Prof. Remo Ruffini receiving the Cressy Morrison Award of the New York Academy of Sciences, 1972 for the discovery of the first Black Hole in our galaxy Cygnus X1.

A "long march" started on 12 December 1970 with the launch of the Uhuru satellite by Riccardo Giacconi. Early in 1971 an almost daily conversation with him and Herb Gursky at the Smithsonian Astrophysical Observatory, leading to the discovery of binary X-ray sources. This was soon followedby the announcement of Cygnus X1 identified as the first black hole in our galaxy (Ruffini 1973); see e.g. Gursky & Ruffini 1975, which contained as well the first publicannouncement of the Discovery of Gamma Ray burst, as well as Giacconi & Ruffini 1980, 2009; see Figs. 3 and 4).

Fig. 4: In the second row, from left to right, there are, among others: E. T. Newman, S. Chandrasekhar (Nobel 1983), R. Giacconi (Nobel 2002), R. Ruffini, A. Treves, A. Hewish (Nobel 1974), D. Arnett, J.H. Taylor (Nobel 1993), J. Wilson, R. Penrose (Nobel 2020), as well as J. Bahcall, T. Damour, T. Piran et al.

Today, after fifty years, this "long march" has reached a definite result: through the grandest observational multi-wavelength effort in the history of mankind, from space, ground and underground observatories, we are finally finding evidence that black holes are "alive" and their "extractable energy" in our mass formula (Christodoulou & Ruffini, 1971), is the energy source of the most energetic cosmological sources: gamma ray bursts (GRBs),the active galactic nuclei (AGNs) as well as the ultra-high energy cosmic rays (UHECRs) (Ruffini et al. 2021 and references therein). Their "inner engine", has three independent components: 1) a Kerr black hole which is neither in a stationary state nor in vacuum, 2) a background magnetic field aligned with the black hole rotation axis, and 3) an extremely diluted fully ionized plasma (Moradi et al. 2021).There is no role in this inner engine for ISCO. Indeed a new electro dynamical field equations describe the synchrotron radiation emitted close to the black hole horizon, they point to a discrete and repetitive emission of "blackholic quanta" in the MeV and in the GeV. The magnitudes and the emission time scales of these quanta, for M87 and GRB 130427A, are expressed as a function of the above three parameters (Rueda & Ruffini, 2021). A long lasting GeV emission with a luminosity decreasing as a temporal power law, allows for the first time in GRBs, the determination of the black hole mass and spin as well as their time evolution perfectly fulfilling our mass energy formula (Christodoulou & Ruffini, 1971): a long lasting emission process profoundly different from the traditional process of continued gravitational contraction.

Remo Ruffini

Professor **GERARD 't HOOFT**

"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon".

Professor Gerard 't Hooft

Prof. Gerard 't Hooft has been a full Professor at the Utrecht University (the Netherlands), since 1977. Nowadays, he is an Emeritus Professor at that University. During his career, he has paid extended scientific visits to CERN (Geneva), Harvard, Stanford, Princeton and Duke University, NC. In 1999, together with M. Veltman, he received the Nobel Prize in Physics, awarded by The Royal Swedish Academy of Sciences, *"For elucidating the quantum structure of electroweak interactions in physics"*.

Prof. 't Hooft's main subjects of research includes:

– Gauge Theories for the sub-atomic particles and forces, various aspects and ingredients of what is now called "The Standard Model of the sub-atomic particles: renormalizability, topological features such as magnetic monopoles and instantons, 1/N expansions.

– Theories for the quantization of the gravitational force and black holes: producing models for the quantum properties of a black hole, as derived from Standard Model and General Relativity alone; its topological features such as antipodal identification.

– Fundamental theories underlying quantum mechanics, in particular returning determinism and reality to the dynamics of the tiniest material entities in his universe.

Prof. 't Hooft has been awarded the Wolf Prize of the State of Israel (1982), the Pius XI Medal (Vatican City, 1983), the Lorentz Medal (KNAW Amsterdam, 1986) as well as the Spinoza Premium (Netherlands Organization for Scientific Research NWO, 1995).

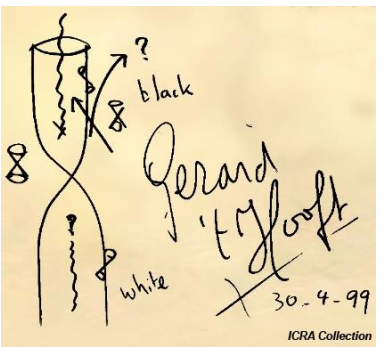


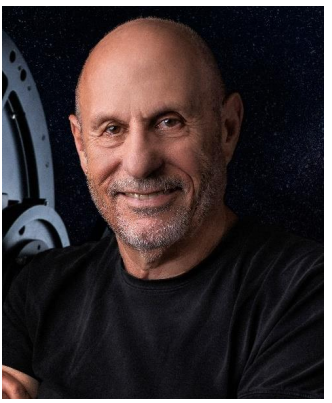
Fig. 2: The signature of Gerard 't Hooft on the wall of ICRA Room 301 (April 4, 1999).

A special event took place at ICRA on April 30, 1999. Prof. Ruffini invited Gerard 't Hooft to Rome to discuss a boundary condition for a quantum field on the black hole horizon, a topic Prof. Ruffini discussed in a previous article "Black-hole evaporation in the Klein-Sauter-Heisenberg-Euler formalism" with Thibault Damour (Phys. Rev. D 14, 332, 1976), but which needed to be examined in more detail. Prof. Ruffini planned to direct Gerard's attention to some specific aspects of this problem. Because we have traditionally been very attentive in spending ICRA travel funds, ICRA offered Gerard to come to Rome on a reduced fare weekend ticket arriving Friday and departing Monday. He had a great relaxing weekend together with Prof. Ruffini following his seminar, which among other things allowed Gerard to sign the wall in our ICRA Room (see Fig. 2), and during this splendid Rome spring weekend he also was able to find a missing factor of 2 in a formula in Prof. Ruffini's 1971 paper with Demetri Christodoulou on the black hole mass formula. The following October, Gerard received the Nobel prize, which meant that we could no longer get away with bringing him to Rome on a cheap ticket! Ever since Gerard has been in our MG IOC helping us with the preparation of the meetings. We are very happy to announce this MG16 Award to Gerard 't Hooft with the motivating phrase *"for his persistent devotion to the study of the quantum field theory boundary conditions at the black hole horizon"*.

Remo Ruffini

Professor **TSVI PIRAN**

"for extending relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory".



Professor Tsvi Piran

Tsvi Piran is the emeritus Schwartzmann professor at the Hebrew University of Jerusalem. He obtained his PhD in Physics, in 1976 from the Hebrew University working on the collisional Penrose process. Piran returned to the Hebrew University at 1981after being a post doc at Oxford and Texas and a long-term member at the IAS at Princeton. In 1982 he initiated and directed the first ever summer school on Gravitational Waves that took place at Les Houches. Piran was a visiting professor at Harvard, Columbia and New York and a Moore scholar at Caltech.

Piran's research deals with numerous aspects of relativistic astrophysics, ranging from the foundation of numerical relativity to modeling of observer relativistic phenomena and analytic work on the fate of gravitational collapse. Piran's research work focuses mostly on black holes and in particular on gamma-ray bursts. He was among the first to point out their cosmological origin and their association with merging neutron stars and heavy r-process nucleo synthesis. Piran's achievements were recognized in the 2019 EMET prize for Physics.

Professor **STEVEN WEINBERG**

"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory".

Professor Steven Weinberg. Photo courtesy of Matt Valentine.

Steven Weinberg is a member of the Physics and Astronomy Departments at The University of Texas at Austin. His research has covered a broad range of topics in quantum field theory, elementary particle physics and cosmology. He has been honored with numerous awards, including the Nobel Prize in Physics, the National Medal of Science, the Heinemann Prize in Mathematical Physics and in 2020, the Breakthrough Prize. He is a member of the US National Academy of Sciences, Britain's Royal Society, and other academies in the USA and abroad. The American Philosophical Society awarded him the Benjamin Franklin Medal, with a citation that said he is "considered by many to be the preeminent theoretical physicist alive in the world today." His books for physicists include *Gravitation and Cosmology*, the three-volume work *The Quantum Theory of Fields*, *Cosmology* and published in April of 2021, *Foundations of Modern Physics*. Educated at Cornell, Copenhagen, and Princeton, he also holds honorary degrees from sixteen other universities. He taught at Columbia, Berkeley, M.I.T., and Harvard, where he was Higgins Professor of Physics, before coming to Texas in 1982.

Fig. 1: Chuo Pei Yuan and Cheng Ning Yang at MG2 in Trieste, Italy (1979).

The Sixteenth Marcel Grossmann Meeting (MG16) is a very special one in many respects: it will take place during a pandemic and in spite of the many difficulties, we have decided not to postpone it but to organize it as a virtual meeting. As described on the MG series webpage, these meetings started in 1975 with the first meeting at the International Centre for Theoretical Physics (ICTP) in Trieste (Italy) that I organized with Nobel Prize winner Abdus Salam. A second meeting followed in 1979, with a significantly larger participation including Nobel Laurate Cheng Ning Yang and a Chinese delegation led by Chuo Pei Yuan (see Fig. 1), including Fang Li-Zhi who had accompanied me during my entire first visit to China in 1979. The first truly international MG meeting followed in 1982 in Shanghai (China):this represented an especially important step forward both for the meeting and for China. A multi-millennia *"motto"* in China, which was then proclaimed on banners everywhere, read *"Friends from all over the world are welcomed"*.

We were soon at an impasse over the participation of scientists from Israel, since no diplomatic relations existed between China and Israel at that time and the Israeli scientists were not to be allowed to attend the meeting. A long negotiation began. The boundary conditions were clearly set by Steven Weinberg, a member of the present MG16 IOC: no MG meetings on Einstein's theory of general relativity could occur without the participation of Israeli scientists. The intervention of Yuval Ne'emann, also a member of the MG IOC then as well as the Minister of Science of Israel (see Fig.2), proposed a compromise that would admit at least one Israeli scientist. I went to Beijing alone, meeting every morning for a week with 12 Chinese representatives led by Chuo Pei Yuan going over all possible options. I stayed in an isolated villa not far from Tiananmen Square, accompanied by the 3 volumes of Matteo Ricci (RI MA TO) to keep me company. No solution was in sight the entire week. At the last moment, just before my departure, an agreement was finally reached allowing two Israeli scientists into China. The historic compromise would admit Gerard Tauber and Tsvi Piran into China using a special ICRA travel document I had proposed for them to be able to participate in the meeting, accepted by the Chinese Ambassador in Rome. This modified the thousand-year Chinese *"motto"* to read *"Scientists from all over the world are welcomed"*. The event was extremely beneficial for China and signaled the truly international nature of the MG meetings.

Fig. 2: From right to left: Chaim Weizmann, President of Israel; Yuval Ne'emann, Minister of Science of Israel; R. Ruffini.

I kept on meeting Tauber in the years which followed (see Fig. 3). Soon after, Yuval Ne'emann visited China. The development of bilateral relations, including military cooperation and economical tights, grow exponentially until the establishment of normal diplomatic relations between Israel and China in 1992.

Fig. 3: From right to left: Arrigo Finzi, Remo Ruffini, Gerard Tauber and Konrad Bleuler.

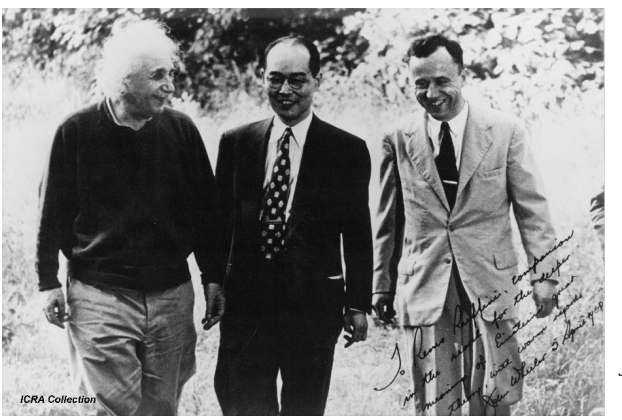


Fig. 4: Albert Einstein, Hideki Yukawa and John. A. Wheeler with a hand-written dedication to Remo Ruffini "To Remo Ruffini, companion in the search for the deeper meaning of Einstein great theory. With warm regards, John Wheeler 5 April 1968".

Given their key role played in the foundations of the MG meetings, I am very happy to propose on behalf of the MG16 IOC, two special Marcel Grossmann Individual Awards: one to Steven Weinberg for *"for unwavering support for the MG meetings since their inception, a true companion in the search for the deeper meaning of Einstein's great theory"* and another one to Tsvi Piran, *"for extending Relativistic astrophysics across international frontiers, a true companion in the search for the deeper meaning of Einstein's great theory"*, in the words of John A. Wheeler's photo dedication to me (see Fig. 4).

Remo Ruffini

PREFACE

Since 1975, the Marcel Grossmann Meetings on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories have been organized in order to provide opportunities for discussing recent advances in gravitation, general relativity and relativistic field theories, emphasizing mathematical foundations, physical predictions and experimental tests. The objective of these meetings is to elicit exchange among scientists that may deepen our understanding of spacetime structures as well as to review the status of ongoing experiments aimed at testing Einstein's theory of gravitation either from the ground or from space. Previous meetings have been held in Trieste (MG1: 1975) and (MG2: 1979), Shanghai (MG3: 1982), Rome (MG4: 1985, MG9: 2000), Perth (MG5: 1988), Kyoto (MG6: 1991), Stanford (MG7: 1994), Jerusalem (MG8: 1997), Rio (MG10: 2003), Berlin (MG11: 2006), Paris (MG12: 2009), Stockholm (MG13: 2012), MG14 in 2015 and MG15 in 2018 both in Rome.

Due to the COVID-19 pandemic spreading in the last two years the decision was taken to organize the Sixteenth Marcel Grossmann meeting for the first time in history entirely online. Despite numerous challenges, related to the organization of large worldwide event, MG16 showed the strongest ever interest from the scientific community with a record-breaking number of almost 1200 registered participants and of more than 1000 speakers.

The traditional six-day schedule has been modified to account for different time zones of the speakers and each day the program of the meeting was divided in three blocks with the reference to the Central European Summer Time. The first block was starting at 06:30 in the morning, allowing comfortable time for speakers from Asia and Oceania. The second block was held in the daytime in Europe and Africa. The third block was starting in the afternoon and ending at 19:30 allowing accommodation of the speakers from the Americas. Each day the blocks of plenary sessions were interchanging with the blocks of about 30 parallel sessions each, making this one of the most intense MG meetings ever. All this was possible thanks to recent developments in communication technologies. The Indico open-source software was selected as a web platform for this meeting, while Zoom platform was adopted for the video-conferencing. The meeting was streamed on ICRANet YouTube channel.

The meeting started on Monday July 5 with the Award ceremony. The individual awards went to Demetrios Christodoulou, Tsvi Piran, Gerard 't Hooft and Steven Weinberg, while the Institutional Awards went to the S.A. Lavochkin Association, to the Max Planck Institute for Extraterrestrial Physics – MPE and to the Space Research Institute IKI of the Russian Academy of Sciences. Overall there were 54 plenary talks, 4 public lectures and 5 roundtables and about 90 parallel sessions. The plenary session "Events in Relativistics Astrophysics" on Monday have seen the contributions from Rashid Sunyaev, Michael Kramer, James Miller-Jones, Felix Mirabel. The public lectures were delivered by Razmik Mirzoyan, Asghar Qadir

and Mohammad Bagheri. Plenary talks on Tuesday session "Black holes and the Quantum" by Juan Maldacena, Ahmed Almheiri, Gerard 't Hooft, Mihalis Dafermos, Sergiu Klainerman, Abhay Ashtekar and Frank Wilczek were bracketed by two roundtables on "New results from SRG/eRosita" with the participation of Andrea Merloni, Prof. Rashid Sunyaev, Alexander Lutovinov, Chandreyee Maitra, Esra Bulbul and "Solar neutrinos and Borexino" with the participation of Gianpaolo Bellini and Wick Haxton. Plenary talks on Wednesday in the session "Lambda CDM tensions" by George Efstathiou, Scolnic Daniel, Marc Kamionkowski, Wendy Freedman, Priya Natarajan and Licia Verde were followed by the roundtable "Precision cosmology" with the participation of Licia Verde, Marc Kamionkowski, Piero Rosati, and the public lecture by Francis Halzen. Two blocks of Thursday plenary sessions "Black holes in GRBs" and "Precision tests" included the talks by Roy Kerr, Yuan Ha, Lorenzo Amati, Elena Pian, Carlos Raúl Argüelles, Di Li, Jianglai Liu, Claus Lämmerzahl, Gerhard Heinzel and Ignazio Ciufolini and were followed by the roundtable "GRB 170817A and GRB 190829A" with the participation of Eleonora Troja, Liang Li, Rahim Moradi, Jorge Armando Rueda Hernandez. Two plenary blocks on Friday "Massive stars" and "Physics behind stellar collapse" included the talks by Selma de Mink, Norbert Langer, Jiri Bicak and Tomáš Ledvinka, Ivan De Mitri, Rahim Moradi and Giancarlo Cella. Finally, two plenary blocks on Saturday "Current and future missions" have seen the talks by Shuang-Nan Zhang, Weimin Yuan, Makoto Tashiro, Ruoyu Liu, Jean-Luc Atteia, Jim Hinton and Nicholas White and were followed by the roundtable "What is in our Galactic center" with the participation of Reinhard Genzel, Carlos Raúl Argüelles, Andreas Krut, Jorge Armando Rueda Hernandez, Eduar Becerra Vergara. The program of the meeting can be found at the official website http://www.icra.it/mg/mg16 and at ICRANet Indico website https://indico.icranet.org/event/1/.

These proceedings include about 400 papers containing the results presented at the Sixteenth Marcel Grossmann meeting. The plenary papers from the meeting have been published in International Journal of Modern Physics D as they were submitted. The table of contents includes also the links to YouTube videos with talks given at the meeting and cover plenary talks, public lectures, roundtables and all parallel sessions. The general link to the videos from MG16 is: https://www.youtube.com/watch?v=QFe1lsSid-o&list=PLr5RLbSWSonsaOnZukBDs0qsNIWM8AvRF.

As the editors we would like to express our gratitude to all the chairpersons of the parallel sessions at MG16, who peer-reviewed the papers submitted for these proceedings, as well as to the ICRANet secretariat office and in particular to Cinzia di Niccolo, Elisabetta Natale and Yasmina Di Domizio, as well as to ICRANet system manager Gabriele Brandolini for their help in preparation of this publication.

Remo Ruffini and Gregory Vereshchagin
November 2021

This page intentionally left blank

Contents

Publications in this Series v

Sponsors and Acknowledgements vii

Organizing Committees ix

Marcel Grossmann Awards xi

Preface xxxvi

PART A
PLENARY SESSIONS

New results from testing relativistic gravity with radio pulsars
Michael Kramer 3

Dragging of inertial frames by matter and waves
Jiří Bičák and Tomáš Ledvinka 22

Probes of the progenitors, engines and physics behind stellar collapse
Chris L. Fryer 39

The observation of high-energy neutrinos from the cosmos: Lessons learned for multimessenger astronomy
Francis Halzen 59

The first results of PandaX-4T
Jianglai Liu on behalf of the PandaX Collaboration 85

XRISM: X-ray imaging and spectroscopy mission
Makoto S. Tashiro and the XRISM team 95

The SVOM mission
J.-L. Atteia, B. Cordier and J. Wei on behalf of the SVOM Collaboration 104

Quantum field theory with boundary conditions at the horizons
Gerard 't Hooft 133

The development of general relativity and the cosmological constant
Asghar Qadir 143

The irreducible mass of Christodoulou-Ruffini-Hawking mass formula
Yuan K. Ha 154

Reshaping our understanding on structure formation with the quantum nature of the dark matter
C. R. Argüelles, E. A. Becerra-Vergara, A. Krut et al. 164

First results of LHAASO
Ruo-Yu Liu for the LHAASO Collaboration 180

On the MG16 awards 2021
Remo Ruffini 196

The white dwarf binary merger model of GRB 170817A
J. A. Rueda, R. Ruffini, Liang Li et al. 217

PARALLEL SESSIONS

Accretion

• MHD Processes Near Compact Objects
Chairperson: Sergey Moiseenko
YouTube link: https://youtu.be/2WMTg06ZmV8

A semi-implicit multidimensional unstructured gas dynamical solver for astrophysical applications
Ilya A. Kondratyev and Sergey G. Moiseenko 242

Magnetized neutron stars propagating through a non-uniform ISM
O. D. Toropina, M. M. Romanova and R. V. E. Lovelace 255

Calculation of the kinetic coefficients of arbitrary degenerate electrons in magnetized dense matter
M. V. Glushikhina and G. S. Bisnovatyi-Kogan 264

Modeling of magnetic fields of accretion discs, using no-z- and RZ-approximations
M. V. Pashentseva and E. A. Mikhailov 272

• Accretion Discs and Jets
Chairpersons: Audrey Trova and Shokoufe Faraji
YouTube link: https://youtu.be/29Wj9RCVEKw

Limiting effects in tori clusters
D. Pugliese and Z. Stuchlik 280

Hydrodynamical transport of angular momentum in accretion disks in the presence of nonlinear perturbations due to noise
Subham Ghosh and Banibrata Mukhopadhyay 295

Properties of accretion disc models in the background quadrupole
Shokoufe Faraji 307

Magnetized tori around a uniformly accelerating black hole
Shokoufe Faraji and Audrey Trova 317

Multifrequency behaviour of high mass X-ray binaries (Time lag between optical and X-ray outbursts)
Franco Giovannelli 321

Active Galactic Nuclei

• The Black Hole in M87
Chairpersons: Brian Punsly and Jorge Rueda
YouTube link: https://youtu.be/l1lTgksyJag

Rotation of the crescent image of M87* and polarization of its ESE hotspot
Krzysztof Nalewajko 339

Magnetic reconnection in jet-accretion disk systems
Chandra B. Singh, Elisabete M. de Gouveia Dal Pino, Luis H. S. Kadowaki et al. 344

• Machine Learning in Astronomy: AGN, Transient Events, Cosmology and Others
Chairpersons: Rahim Moradi and Yu Wang
YouTube link: https://youtu.be/ErqrmMZQsBk

Exact fractal model of the universe and possible machine learning methods for the verification of the fractality
A. A. Kirillov, E. P. Savelova and P. O. Vladykina 352

Estimating the photometric redshifts of galaxies and QSOs using regression techniques in machine learning
A. Momtaz, M. H. Salimi and S. Shakeri 368

Deep learning in quasar physics
F. Rastegar Nia, M. T. Mirtorabi, R. Moradi, Y. Wang and A. Vafaei Sadr 382

Cosmological density field emulation and gravitational wave inference based on dimensionality reduction and supervised machine learning
Miguel Conceição, António da Silva and Alberto Krone-Martins 391

Unsupervised photometric detection of galaxy cluster candidates in large surveys
Ana Carvalho, Alberto Krone-Martins and Antonio da Silva 409

• Multiwavelength and Multi-Messenger Observations of Active Galactic Nuclei
Chairpersons: Paolo Giommi and Narek Sahakyan

YouTube link: https://youtu.be/-Hyu2NQsExg

Time-dependent lepto-hadronic modeling of the emission processes in blazar jets
S. Gasparyan, D. Bégué and N. Sahakyan 429

Multiwavelength study of high-redshift blazars
G. Harutyunyan and D. Israyelyan 445

Alternative Theories

• Extended Theories of Gravity and Quantum Cosmology
Chairpersons: Yi-Fu Cai and Wentao Luo

YouTube links: https://youtu.be/ADRr9DfV5zM
https://youtu.be/eOzpiC1cFkU
https://youtu.be/kYzJds_JIp8

Quantum gravity phenomenology from thermodynamics of spacetime
A. Alonso-Serrano and M. Liška 462

Gauge theory of gravity based on the correspondence between the 1^{st} and the 2^{nd} order formalisms
David Benisty 479

$U(1)$ local strings in hybrid metric-Palatini gravity
Tiberiu Harko, Francisco S. N. Lobo and Hilberto M. R. da Silva 485

Inflationary supersymmetric FRLW quantum cosmology
N. E. Martínez-Pérez, C. Ramirez and V. M. Vázquez-Báez 499

Effective $f(R)$ actions for modified loop quantum cosmologies
Ana Rita Ribeiro, Daniele Vernieri and Francisco S. N. Lobo 517

Probing multiverse using gravitational wave observations
Moe Kukihara and Kazuhiro Hayama 531

Operator ordering ambiguity in observables of quantum cosmology
Harkirat Singh Sahota and Kinjalk Lochan 538

Decoupled quark stars in self-interacting Brans-Dicke gravity
M. Sharif and Amal Majid 548

Big-bounce in projectively invariant Nieh-Yan models: The Bianchi I case
Flavio Bombacigno, Simon Boudet, Gonzalo J. Olmo and Giovanni Montani 561

Late time cosmology with derivatives of matter Lagrangian
Shahab Shahidi 576

On the semiclassical and quantum picture of the Bianchi I polymer dynamics
E. Giovannetti, G. Montani and S. Schiattarella 588

Quantum corrections to the Bianchi II transition under local rotational invariance
Sara F. Uria, David Brizuela and Ana Alonso-Serrano 597

• **Mathematical Problems of Relativistic Physics: Classical and Quantum**
Chairpersons: A. Shadi Tahvildar-Zadeh and Michael Kiessling

YouTube links: https://youtu.be/9Dr3M9Kb2jo
https://youtu.be/gMOykpapJ5A

The hypercomplex medium as storage of physical equations
Alexander P. Yefremov 605

The Maxwell-Bopp-Landé-Thomas-Podolsky-Einstein system for a static point source
Érik Amorim 619

On recent developments in the theory of relativistic dissipative fluids
V. Hoang 627

Adiabatic solutions in general relativity as null geodesics on the space of boundary diffeomorphisms
Emine Şeyma Kutluk 635

The point spectrum of the Dirac Hamiltonian on the zero-gravity Kerr-Newman spacetime
M. Kiessling, E. Ling and A. S. Tahvildar-Zadeh 648

Causal fermion systems: Classical gravity and beyond
Felix Finster 661

Newman-Penrose-Debye formalism for fields of various spins in pp-wave backgrounds
Aleksandr Kulitskii and Elena Yu Melkumova 679

Gravitational geometric phase
Banibrata Mukhopadhyay, Tanuman Ghosh and Soumya Kanti Ganguly 689

Retarded potentials and radiation in odd dimensions
D. V. Gal'tsov and M. Khlopunov 699

Wave propagation in the anti-deSitter optical metric
D. García-Peláez and C. S. López-Monsalvo 713

New approaches to constrained dynamics and Hamilton-Jacobi procedures in general relativity
D. Salisbury, J. Renn and K. Sundermeyer 719

Orientability of space from electromagnetic quantum fluctuations
N. A. Lemos and M. J. Rebouças 725

Essential self-adjointness of Dirac operators under the influence of general-relativistic gravity
Michael K.-H. Kiessling, A. Shadi Tahvildar-Zadeh and Ebru Toprak 736

• **Wormholes, Energy Conditions and Time Machines**
Chairpersons: Francisco Lobo and Diego Rubiera-Garcia

YouTube links: https://youtu.be/tu_3Wqcd9Ys
https://youtu.be/NqN1c-2fv8Y

Relic magnetic wormholes as possible source of toroidal magnetic fields in galaxies
A. A. Kirillov and E. P. Savelova 743

Wormhole geometries induced by action-dependent Lagrangian theories
Ismael Ayuso, Francisco S. N. Lobo and José P. Mimoso 756

Gravitational lensing by wormholes in binary systems
S. Pietroni 774

Hyper-fast positive energy warp drives
E. W. Lentz 779

From black-bounce to traversable wormhole, and beyond
Alex Simpson 787

Tractor beams, pressor beams, and stressor beams within the context of general relativity
Matt Visser, Jessica Santiago and Sebastian Schuster 808

A singularity theorem for evaporating black holes
E.-A. Kontou, B. Freivogel and D. Krommydas 822

Circularly symmetric thin-shell wormholes in F(R) gravity with (2+1)-dimensions
Cecilia Bejarano, Ernesto F. Eiroa and Griselda Figueroa-Aguirre 831

Warp drive dynamic solutions considering different fluid sources
Osvaldo L. Santos-Pereira, Everton M. C. Abreu and Marcelo B. Ribeiro 840

Symmetries and geometry of spacetime: Towards a new paradigm
Francisco Cabral, Francisco S. N. Lobo and Diego Rubiera-Garcia 856

• Theories of gravity: Alternatives to the cosmological and particle standard models
Chairpersons: Stefano Bellucci and Orlando Luongo

YouTube links: https://youtu.be/-aQcrYoQfBM
https://youtu.be/WsCwPb5OQhY
https://youtu.be/JMnrUfBgqVU
https://youtu.be/pPhZY-bbsew

Thermodynamics of scalar-tensor gravity: A new approach
Valerio Faraoni 876

Two body dynamics in a quadratic modification of general relativity
Soham Bhattacharyya 883

Alternatives to Λ: Torsion, generalized couplings, and scale invariance
C. J. A. P. Martins, C. M. J. Marques, C. B. D. Fernandes et al. 907

Model-independent test of scalar-tensor gravity theory by reconstructing scalar mode of GW170817
Yuya Gushima and Kazuhiro Hayama 921

Cosmology in the novel scalar-tensor representation of $f(R,T)$ gravity
Tiago B. Gonçalves, João Luís Rosa and Francisco S. N. Lobo 932

On the interaction between electromagnetic, gravitational, and plasma related perturbations on LRS class II spacetimes
P. Semrén 943

Condition for expansion-collapse duality between Einstein and Jordan frames
Dipayan Mukherjee, H. K. Jassal and Kinjalk Lochan 958

The model of dark energy based on the quantum-mechanical uncertainty relation
Yu. V. Dumin 967

• Conformal Dilaton Gravity and Related Issues
Chairperson: Reinoud Jan Slagter

YouTube link: https://youtu.be/A3Ygi3YBs5A

Conformal dilaton gravity, antipodal mapping and black hole physics on a warped spacetime
R. J. Slagter 978

From neutrino masses to the full size of the universe — Some intriguing aspects of the tetron model
B. Lampe 999

Summary parallel session AT5
Reinoud Jan Slagter 1014

• Horava-Lifshitz Gravity
Chairperson: Anzhong Wang

YouTube links: https://youtu.be/vPT1dH1zITE
https://youtu.be/5z7zhpiDOpw

Boundary conditions for the Klein-Gordon field on Lifshitz spacetimes
Lissa de Souza Campos 1017

Dynamical system analysis of Bianchi-I spacetimes in $f(R)$ gravity
Saikat Chakraborty, Kazuharu Bamba and Alberto Saa 1026

Cosmological implications in modified Hořava-Lifshitz gravity
Abdul Jawad, Kazuharu Bamba and Farwa Khurshid 1038

Finite action principle and wormholes
Jan Chojnacki and Jan Kwapisz 1046

Strange quark stars in Hořava gravity
Grigoris Panotopoulos 1054

Shadows of Kerr-like black holes in $4D$ Einstein–Gauss–Bonnet gravity and constraints from EHT observations
Sushant G. Ghosh and Rahul Kumar Walia 1069

Wormhole interaction in 2d Horăva-Lifshitz quantum gravity
Jan Ambjørn, Yuki Hiraga, Yoshiyasu Ito and Yuki Sato 1084

Nature of singularities in vector-tensor theories of gravity
V. H. Satheeshkumar 1095

Hořava-Lifshitz and Einstein-Æther gravity in the light of Event Horizon Telescope observations of M87*
Emmanuel N. Saridakis 1104

Hořava-Lifshitz gravity in $(3 + 1)$ dimensions coupled with anisotropic matter and possible constraints from GRB 170817A
Tao Zhang and Fu-Wen Shu 1112

Summary of the parallel session AT6
Anzhong Wang 1119

• Ghost-Free Models of Modified Gravity: Massive Gravity, Horndeski and DHOST Theories, Other Related Models; Their Properties and Solutions
Chairpersons: Dmitry Gal'tsov and Michael Volkov

YouTube links: https://youtu.be/l8KHUPnT2D8
https://youtu.be/48OHIKpgNqs

Non-local R^2-like inflation, gravitational waves and Non-Gaussianities
K. Sravan Kumar 1124

Palatini kinetic scalar-tensor theory: Analytical and numerical solutions
D. V. Gal'tsov and D. S. Bushuev 1136

PART B

Black Holes: Theory and Observations/Experiments

• Theoretical and Observational Studies of Astrophysical Black Holes
Chairperson: Alexander Zakharov

YouTube links: https://youtu.be/fiv_MH-N2kw
https://youtu.be/GYoOb17GvE8

Reconstruction of a star motion in the vicinity of black hole from the redshift of the electromagnetic spectrum
S. O. Komarov and A. K. Gorbatsievich 1151

Shadows of hairy Kerr black holes and constraints from M87*
Sushant G. Ghosh and Misba Afrin 1167

Displacement memory and BMS symmetries
Shailesh Kumar 1179

Physical black holes in semiclassical gravity
Sebastian Murk and Daniel R. Terno 1196

• Black Hole Thermodynamics
Chairperson: Hernando Quevedo

YouTube links: https://youtu.be/XmZDf5mrXQk
https://youtu.be/amseL2qykfk

Information recovery from evaporating rotating charged black holes
Zhi-Wei Wang, Samuel L. Braunstein and Saurya Das 1212

Black hole thermodynamics from entanglement mechanics
S. Mahesh Chandran and S. Shankaranarayanan 1223

Thermodynamics of charged black hole
M. Sharif and Amjad Khan 1238

Rindler trajectories and Rindler horizons in the Schwarzschild spacetime
Kajol Paithankar and Sanved Kolekar 1250

Linear growth of the two-point function for the Unruh state in $1+1$ dimensional black holes
Paul R. Anderson, Zachary P. Scofield and Jennie Traschen 1255

Stress-energy tensor for a quantized scalar field in a four-dimensional black hole that forms from the collapse of a null shell
Shohreh Gholizadeh Siahmazg, Paul R. Anderson, Raymond D. Clark and Alessandro Fabbri 1265

Microscopic model building for black hole membranes from constraints of symmetry
Swastik Bhattacharya and S. Shankaranarayanan 1275

Einstein-Maxwell-Dilaton-Axion mass formulas for black holes with struts and strings
Dmitri Gal'tsov, Gérard Clément and Igor Bogush 1291

• Black Holes in Alternative Theories of Gravity
Chairpersons: Jutta Kunz and Kamal Hajian

YouTube links: https://youtu.be/FRNGJKhiw7c
https://youtu.be/Tjfmuut1Eo0

Holography for rotating black holes in $f(T)$ gravity
Masoud Ghezelbash 1308

Infinitely degenerate exact Ricci-flat solutions in $f(R)$ gravity
Semin Xavier, Jose Mathew and S. Shankaranarayanan 1319

Universe in a black hole with spin and torsion
Nikodem Popławski 1327

Asymptotically flat black hole solution in modified gravity
Surajit Kalita and Banibrata Mukhopadhyay 1337

Shadow of a charged black hole surrounded by an anisotropic matter field
Javier Badía and Ernesto F. Eiroa 1343

Constraining modified gravity theories with physical black holes
Sebastian Murk 1351

Penrose suggestion as to pre-Planck-era black holes showing up in present universe data sets discussed, with a possible candidate as to GW radiation which may provide initial CMBR data
A. W. Beckwith 1359

Summary of the parallel session BH3
Kamal Hajian and Jutta Kunz 1372

Binaries

• **Explosive Events Associated with Compact-Object Binary Mergers**
Chairpersons: Chris Belczynski and Jorge Rueda

YouTube links: https://youtu.be/Dwq1ZU3gKrg
https://youtu.be/nw02ylI6R2M

Uncertainties in kilonova modeling
C. L. Fryer, C. J. Fontes, O. Korobkin et al. 1391

• **Post-Newtonian and Post-Minkowskian Corrections for Binary Gravitating Systems**
Chairperson: Johannes Bluemlein

YouTube link: https://youtu.be/wfiLG5r08yE

Tutti-Frutti method: Recent developments in the PN/PM/SF treatment of the gravitational two-body problem
Donato Bini and Andrea Geralico 1405

• **Multichannel Studies of Nonstationary Relativistic Stars**
Chairperson: Vladimir Lipunov

YouTube link: https://youtu.be/usn2PlU_qFA

GRB observations on cubesate satellites in the Universat–SOCRAT project
Sergey I. Svertilov, Michail I. Panasyuk, Vitaly V. Bogomolov et al. 1412

Multiwavelength observations of GRB160625B by MASTER, Lomonosov, Konus-Wind and three stage collapse
V. M. Lipunov, V. A. Sadovnichy, M. I. Panasyuk et al. 1429

The role of the magnetic fields in GRB outflows
N. Jordana-Mitjans, C. G. Mundell, S. Kobayashi et al. 1449

MASTER optical observations of the blazar TXS0506+056 during the IC170922A
V. M. Lipunov, K. Zhirkov, V. G. Kornilov et al. 1467

Boson Stars

• Scalar Fields in Cosmology
Chairpersons: Carlos Herdeiro and Alfredo Macias

YouTube links: https://youtu.be/SyLoguueGKk
https://youtu.be/f9nuo8Jvw-w

A short review on nonlinear perturbation theory of structure formation for modified gravity
Jorge L. Cervantes-Cota and Alejandro Aviles 1474

Testing modified gravity theories with marked statistics
Alejandro Aviles 1494

Dark matter as condensed phase of generic bosons
Elías Castellanos and Jorge Mastache 1513

Cosmic Microwave Background

• Cosmic Backgrounds from Radio to Far-IR
Chairperson: Carlo Burigana

YouTube link: https://youtu.be/4e3Cj5wahck

New Planck tSZ map and its cosmological analysis
H. Tanimura, M. Douspis and N. Aghanim 1527

The CMB dipole: Eppur si muove
R. M. Sullivan and D. Scott 1532

High angular resolution Sunyaev Zel'dovich observations: The case of MISTRAL
E. S. Battistelli, E. Barbavara, P. de Bernardis et al. 1542

Cosmological and astrophysical results exploiting magnification bias with high-z sub-millimetre galaxies
L. Bonavera, M. M. Cueli and J. Gonzalez-Nuevo 1557

The impact of the Lorentz symmetry violation on the CMB polarization
Seddigheh Tizchang, Rohoollah Mohammadi and She-Sheng Xue 1571

Cosmic backgrounds from the radio to the far-infrared: Recent results and perspectives from cosmological and astrophysical surveys
Carlo Burigana, Elia Sefano Battistelli, Laura Bonavera et al. 1579

• **New Horizons in Cosmology with CMB Spectral Distortions**
Chairpersons: Jens Chluba and Andrea Ravenni

YouTube links: https://youtu.be/uBYLO4Smw3o
https://youtu.be/5oAPfzAe35k

CMB μT cross-correlations as a probe of PBH scenarios
Ogan Özsoy and Gianmassimo Tasinato 1609

Theoretical and numerical aspects of CMB spectral distortions from non-thermal energy injections
Sandeep Kumar Acharya, Jens Chluba and Abir Sarkar 1628

BISOU: A balloon project to measure the CMB spectral distortions
B. Maffei, M. H. Abitbol, N. Aghanim et al. 1633

Cosmic microwave background spectral distortions constraints on decaying dark matter particles and axion-like particles using *COBE/FIRAS* and *EDGES*
Boris Bolliet 1645

The COSmic Monopole Observer (COSMO)
S. Masi, E. Battistelli, P. de Bernardis et al. 1654

• **Status of the $\mathbf{H_0}$ and $\boldsymbol{\sigma_8}$ Tensions: Theoretical Models and Model-Independent Constraints**
Chairpersons: Joan Solà Peracaula and Adrià Gómez-Valent

YouTube links: https://youtu.be/VNWZ1Bzjus4
https://youtu.be/zEBOCwvetKE
https://youtu.be/nzgC7qV9H_Y

Measuring the Hubble constant H_0 from gravitational lensing
Liliya L. R. Williams 1672

Extra components consistency in the Hubble tension and BBN
Osamu Seto and Yo Toda 1686

Gravitational anomalies, axions and a string-inspired running vacuum model in Cosmology
Nick E. Mavromatos 1693

Early and not so early dark energy. What do cosmological observations tell us about them?
Adrià Gómez-Valent, Ziyang Zheng, Luca Amendola, Valeria Pettorino and Christof Wetterich 1713

Renormalized $\rho_{\rm vac}$ without m^4 terms
Cristian Moreno-Pulido and Joan Solà Peracaula 1733

BD-ΛCDM and running vacuum models: Theoretical background and current observational status
Javier de Cruz Pérez, Joan Solà Peracaula, Adrià Gómez-Valent and Cristian Moreno-Pulido 1752

Cosmological tensions: Hints for a new concordance model?
E. Di Valentino 1770

Solving both H_0 and σ_8 tensions in $f(T)$ gravity
Emmanuel N. Saridakis 1783

Precision Cosmology and Hubble tension in the era of LSS surveys
G. Fanizza 1792

• **Effects of Primordial Perturbations Enhancement: From Black Holes Formation to CMB Anomalies**
Chairpersons: Antonio Enea Romano and Krzysztof Turzynski
YouTube link: https://youtu.be/frjjONXbd1M

Primordial black holes arise when the inflaton falls
Keisuke Inomata 1803

Effects of the modification of gravity on the production of primordial black holes
Sergio Andrés Vallejo Peña 1809

Cosmic Strings

• **Cosmic Strings**
Chairpersons: Reinoud Jan Slagter and Batool Imtiaz
YouTube link: https://youtu.be/ZpaU82ZUHzM

$U(1)$ local strings in generalized hybrid metric-Palatini gravity
Hilberto M. R. da Silva, Tiberiu Harko, Francisco S. N. Lobo and João Luís Rosa 1820

New evidence of the azimuthal alignment of quasars spin vector in Large Quasar Groups and cosmic strings
R. J. Slagter 1835

Summary parallel session cosmic strings I
Reinoud Jan Slagter and Batool Imtiaz 1848

• **From Cosmic Strings to Superstrings**
Chairpersons: Carlos Martins and Ivan Rybak
YouTube link: https://youtu.be/LJFtV_4aSAg

Scaling solutions of wiggly cosmic strings
A. R. R. Almeida and C. J. A. P. Martins 1851

High resolution calibration of string network evolution
J. R. C. C. C. Correia and C. J. A. P. Martins 1871

Radiation from Global Cosmic Strings using adaptive mesh refinement
Amelia Drew and E. P. S. Shellard 1891

Analysing the scaling density of axion strings
A. Lopez-Eiguren 1898

Electroweak axion string and superconductivity
Yu Hamada, Yoshihiko Abe and Koichi Yoshioka 1912

Dark Energy and Large Scale Structure

• **Dark Energy and the Accelerating Universe**
Chairpersons: Alexei Starobinky and David Polarski
YouTube links: https://youtu.be/JZEPRS_rqbE
https://youtu.be/3aG_AT4UzWE
https://youtu.be/ZlTWquG-pFk

Hints for the $H_0 - r_d$ tension in uncorrelated Baryon Acoustic Oscillations dataset
Denitsa Staicova 1923

Observational constraints on nonlinear matter extensions of general relativity
E.-A. Kolonia and C. J. A. P. Martins 1935

Constraining the dark energy-dark matter interaction model using low-redshift observations
Archana Sangwan, Joseph P. J. and S. Shankaranarayanan 1948

On the evolution of inhomogeneous perturbations in the ΛCDM model and $f(R)$ modified gravity theories
T. Schiavone and G. Montani 1961

Soft dark energy and soft dark matter
Emmanuel N. Saridakis 1970

A simple parametrisation for coupled dark energy
Vitor da Fonseca, Nelson J. Nunes and Tiago Barreiro 1979

• **Cosmography with Gravitational Lensing**
Chairpersons: Claudio Grillo and Mimoza Hafizi
YouTube link: https://youtu.be/FXplIAvUDBM

A tale of two double quasars: Hubble constant tension or biases?
L. J. Goicoechea and V. N. Shalyapin 1990

Dark Matter

• **Interacting Dark Matter**
Chairpersons: Nikolaos Mavromatos
YouTube links: https://youtu.be/zOshKJlwD-Y
https://youtu.be/JR44dh2GYik

Dark energy and dark matter unification from dynamical space time: BBN constraints
D. Benisty 2005

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2013

LHC experiments for long-lived particles of the dark sector
Vasiliki A. Mitsou 2029

Constraining the interactions in the dark sector with cosmological data
Adrià Gómez-Valent, Valeria Pettorino and Luca Amendola 2050

Running vacuum interacting with dark matter or with running gravitational coupling. Phenomenological implications
Joan Solà Peracaula 2069

Dark matter properties from the Fornax globular cluster timing: Dynamical friction and cored profiles
D. Blas 2089

Growth of linear perturbations in a universe with superfluid dark matter
S. Banerjee, S. Bera and D. F. Mota 2101

Interacting dark sector in the late Universe: Mapping fields and fluids, and observational signatures
Joseph P. J. and S. Shankaranarayanan 2119

The role of self interactions in the cosmological evolution of warm dark matter
R. Yunis, C. R. Argüelles, D. López Nacir et al. 2127

Interaction energy between a charged medium and its electromagnetic field as a dark matter candidate
Mayeul Arminjon 2139

• Dark Matter Searches with Liquid Xenon and Argon Detectors and Self Gravitating Systems and Dark Matter
Chairpersons: Marco Merafina and Soroush Shakeri and She-Sheng Xue

YouTube link: https://youtu.be/H9oGYnGq9pI

The maximum mass of dilute axion stars
Pierre-Henri Chavanis 2149

A dark matter solution for the XENON1T electron excess and the galactic center 511 keV line
Yasaman Farzan 2174

Preliminary results of rich galaxy clusters' spatial distribution analysis on CfA2 Redshift Survey data: Compact objects or dark matter presence at redshift less 0.032
I. V. Arkhangelskaja, A. M. Galper, L. N. Khanh and D. N. Dorosheva 2189

• Dark Matter: Beyond ΛCDM
Chairpersons: Carlos Argüelles and Andreas Krut

YouTube links: https://youtu.be/hdKeo5L7pYE
https://youtu.be/i0IPHXzmV-s

Probing the nature of dark matter with Milky Way subhaloes
M. R. Lovell 2202

Addressing classical cosmological back-reaction with multiple scales
Yonadav Barry Ginat 2217

Imaging formation process for DM profiles
Omar de J. Cabrera-Rosas and Tonatiuh Matos 2222

The self-gravitating Fermi gas in Newtonian gravity and general relativity
Pierre-Henri Chavanis 2230

• Dark Matter and Rare Processes
Chairpersons: Rita Bernabei and Zurab Berezhiani

YouTube links: https://youtu.be/wheVrbETP_0
https://youtu.be/dVdZ4TIDxt0

The dark matter: DAMA/LIBRA and its perspectives
R. Bernabei, P. Belli, V. Caracciolo et al. 2252

Dark matter directionality approach
R. Bernabei, P. Belli, V. Caracciolo et al. 2272

Collapse models under test by high sensitivity γ-ray and X-ray measurements
C. Curceanu, Kristian Piscicchia, Massimiliano Bazzi et al. 2288

Leptophilic dark matter at linear colliders
P. S. Bhupal Dev 2296

DM6 session: Dark matter and rare processes
R. Bernabei and Z. Berezhiani 2316

• **The Nature of Galactic Halos**
Chairpersons: Francesco De Paolis and Asghar Qadir
YouTube link: https://youtu.be/qIwVnNxi0n0

Primordial black holes as dark matter candidates in the Galactic halo
Lindita Hamolli, Mimoza Hafizi, Francesco De Paolis and Achille A. Nucita 2319

Giant cosmic ray halos around M31 and the Milky Way
S. Recchia, S. Gabici, F. A. Aharonian and V. Niro 2335

A nearly complete census of intergalactic gas using the kinematic Sunyaev-Zel'dovich effect
Chaves-Montero, Jonás 2345

Searching for Intermediate Mass Black Holes in the Milky Way's galactic halo
A. Franco, A. A. Nucita, F. De Paolis, F. Strafella and M. Maiorano . . . 2352

Virial clouds evolution from the last scattering up to the formation of first stars
Noraiz Tahir, Asghar Qadir, Muhammad Sakhi and Francesco De Paolis 2360

Testing Weyl-modified gravity on M31 and Milky Way
Muhammad Bilal and Asghar Qadir 2365

PART C

Education

• **Teaching Einsteinian Physics to School Students**
Chairpersons: David Blair and Matteo Luca Ruggiero
YouTube links: https://youtu.be/W-WV6J8kprg
https://youtu.be/UHYpwKQ09SU

Teaching relativity: A paradigm change
F. Herrmann and M. Pohlig 2371

Teaching relativity: Computer-aided modeling
F. Herrmann and M. Pohlig 2381

Solstices and Equinoxes in 1703 at the meridian line of St. Maria degli Angeli in Rome, and the stellar aberration of Sirius
Costantino Sigismondi and Silvia Pietroni 2388

Positional astrometry at arcsecond accuracy using historical instruments, with light equipment
Costantino Sigismondi and Lorenzo Ricciardi 2398

Daily, seasonal, and equinoctial solar paths on a school soccer field
Costantino Sigismondi 2411

Teaching relativity at the AstroCamp
C. J. A. P. Martins 2415

Sungrazing comets as General Relativistic gravitational probes
Silvia Pietroni and Costantino Sigismondi 2424

The three Summer solstice's markers of 1721 unveiled in the Basilica of Santa Maria degli Angeli in Rome
Costantino Sigismondi 2428

Einstein-First: Bringing children our best understanding of reality
A. Popkova, K. Adams, S. Boublil et al. 2438

Exact Solutions

• **Exact Solutions in Four and Higher Dimensions**
Chairpersons: David Blair and Matteo Luca Ruggiero

YouTube link: https://youtu.be/GHZUS5-4gVQ

Kundt spacetimes in the Einstein–Gauss–Bonnet theory
R. Švarc, J. Podolský and O. Hruška 2453

Exact decoupled solutions in curvature-matter coupled gravity
M. Sharif and Fizza Furqan 2464

Tolman-Oppenheimer-Volkov conditions beyond spherical symmetry
José P. Mimoso, Alan Maciel and Morgan Le Delliou 2479

A spherically symmetric stiff fluid spacetime in light of cosmic structure formation
Daniele Gregoris 2497

• **Exact Solutions (Including Higher Dimensions)**
Chairperson: Susan Scott

YouTube link: https://youtu.be/-hJjJvmmOew

Three-parameter solution for the null-surface formulation in 2+1 dimensions
Tina A. Harriott and J. G. Williams 2510

New exact stationary cylindrical anisotropic fluid solution of GR
M.-N. Célérier 2522

Early Universe

• **Quantum Fields**
Chairperson: Andrei Lebed

YouTube links: https://youtu.be/BLaTp0r0TkQ
https://youtu.be/ZifgaSDy5Vc

Hydrodynamic representation and energy balance for the Dirac and Weyl fermions in curved space-times
Tonatiuh Matos, Omar Gallegos and Pierre-Henri Chavanis 2533

Breakdown of the Equivalence Principle for a composite quantum body
A. G. Lebed 2551

Extended DeWitt-Schwinger subtraction scheme, heavy fields and decoupling
Antonio Ferreiro and Jose Navarro-Salas 2557

Renormalization and decoupling for the Yukawa model in curved spacetime
Sergi Nadal-Gisbert, Antonio Ferreiro and José Navarro-Salas 2562

Trace anomaly and evaporation of spherical black holes
P. Meda 2573

On decay of the false Unruh vacuum
A. Shkerin 2587

Behaviour of noise kernel in de Sitter and FRW space-times
Ankit Dhanuka and Kinjalk Lochan 2600

Breaking Buchdahl: Ultracompact stars in semiclassical gravity
Julio Arrechea, Carlos Barceló, Raúl Carballo-Rubio and Luis J. Garay 2608

Einstein anomaly with tensors of odd order in six dimensional curved space
Kohei Yamamoto and Satoshi Yajima 2619

Quantum memory and BMS symmetries
Sanved Kolekar and Jorma Louko 2623

• Topological Methods, Global Existence Problems, and Spacetime Singularities
Chairperson: Spiros Cotsakis

YouTube link: https://youtu.be/H8Itnc_C10c

Gravitational singularities, scattering maps for bouncing, and structure-preserving algorithms
Philippe G. LeFloch 2630

Brane-world asymptotics in a nonlinear fluid bulk
I. Antoniadis, S. Cotsakis and Ifigeneia Klaoudatou 2645

Primordial synchronization of Mixmaster spatial points
Spiros Cotsakis 2657

• The Early Universe
Chairperson: Stefano Ansoldi

YouTube links: https://youtu.be/m80hHWgOlFs
https://youtu.be/6tyaZ8MMVtw

Quintessential inflation from Lorentzian slow roll
David Benisty 2663

Condensed light, quantum black holes and L-CDM cosmology: Experimentally suggested and tested unified approach to dark matter, dark energy, cosmogenesis and two-stage inflation
Victor Borsevici 2672

Helical magnetic fields lead to baryogenesis
Ashu Kushwaha and S. Shankaranarayanan 2692

Polymer Quantization of the Isotropic Universe: Comparison with the Bounce of Loop Quantum Cosmology
G. Barca, E. Giovannetti, F. Mandini and G. Montani 2700

General relativistic evolution equations for density perturbations in open, flat and closed FLRW universes and the problem of structure formation
Pieter G. Miedema 2708

Constraining beyond ΛCDM models with 21cm intensity mapping forecast observations combined with latest CMB data
M. Berti 2726

Entropy and irreversible processes in gravity and cosmology
Llorenç Espinosa-Portalés and Juan García-Bellido 2737

Fundamental Interactions and Stellar Evolution

• Why and How the Sun and the Stars Shine: The Borexino Experiment
Chairpersons: Giampaolo Bellini, Dmitry Naumov, Gioacchino Ranucci, Gemma Testera

YouTube links: https://youtu.be/hh5wDnM8miU
https://youtu.be/TG-HgBf7W4s
https://youtu.be/HLiis2LFeEs

Experimental detection of the CNO cycle
B. Caccianiga, N. Rossi, G. Testera et al. 2753

Borexino detector performances
A. Caminata, M. Agostini, K. Altenmüller et al. 2765

Study of antineutrinos from the Earth and the Cosmos with the Borexino detector
Sandra Zavatarelli, M. Agostini, K. Altenmuller et al. 2774

Unveiling the engine of the Sun: Measurements of the *pp*-chain solar neutrinos with Borexino
D. Guffanti, A. C. Re and O. Smirnov 2785

Electron neutrino survival probability in the energy range 200 keV–15 MeV
Marco Pallavicini on behalf of the Borexino Collaboration 2804

The relevance of pp-chain and CNO-cycle neutrino measurements for solar physics
F. L. Villante and A. M. Serenelli 2815

Role of the CNO cycles in stars
A. Ianni 2835

Geoneutrino observation
Tadao Mitsui for the KamLAND Collaboration 2840

Synthesis of the session: Why and how the Sun and the stars shine
Gianpaolo Bellini 2845

• **Rotation in Stellar Evolution**
Chairperson: Georges Meynet

YouTube link: https://youtu.be/9LNoiwa0nv8

The internal rotation of low-mass stars from solar and stellar seismology
G. Buldgen and P. Eggenberger 2848

The rotation of supermassive stars
L. Haemmerlé 2865

Fast Transients

• **What Can We Learn from a Growing Sample of Fast Radio Bursts?**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/yo4n1SgfUrQ
https://youtu.be/NwGooPauhjU

Cosmology with high-redshift FRBs
A. Fialkov 2880

• **Non Standard Cosmological Probes**
Chairpersons: Duncan Lorimer, Victoria Kaspi and Bing Zhang

YouTube links: https://youtu.be/EEFUgiFeMck
https://youtu.be/Ryb15AINfMs

Closing the cosmological loop with the redshift drift
C. J. A. P. Martins, C. S. Alves, J. Esteves, A. Lapel and B. G. Pereira 2890

Gamma-Ray Bursts as potential cosmological probes
L. Izzo 2906

Surface brightness fluctuations: The method and future applications
Michele Cantiello 2915

Preliminary results of analysis of Ia supernovae redshift distributions on data of the Asiago Supernova and Open Supernova Catalogues
I. V. Arkhangelskaja 2930

• **Photospheric Emission in GRBs**
Chairpersons: Gregory Vereshchagin and Damien Bégué

YouTube links: https://youtu.be/ZifnzoUXIFc
https://youtu.be/n7vWAVhFXiU

Understanding prompt emission: Where do we stand?
Asaf Pe'er 2946

On explaining prompt emission from GRB central engines with photospheric emission model
M. Bhattacharya and P. Kumar 2957

Monte Carlo simulations of photospheric emission in Gamma Ray Bursts
T. M. Parsotan and D. Lazzati 2972

The photosphere emission spectrum of hybrid relativistic outflow for Gamma-Ray Bursts
Yan-Zhi Meng, Jin-Jun Geng and Xue-Feng Wu 2982

On diffusive photospheres in Gamma-Ray Bursts
G. V. Vereshchagin 2989

Summary of the parallel session GB3
G. V. Vereshchagin and D. Bégué 3002

• High and Very High Energy Emission from Gamma-Ray Bursts
Chairpersons: Francesco Longo and Fabian Schüssler

YouTube links: https://youtu.be/ZUkLyyYowaM
https://youtu.be/zwN1mNVeUzA

Synchrotron and synchrotron self-Compton emission components in GRBs detected at very high energies
Jagdish C. Joshi, Vikas Chand and Soebur Razzaque 3009

The VERITAS gamma-ray burst follow-up program
D. Ribeiro for the VERITAS Collaboration 3017

MAGIC view of Gamma-Ray Bursts at very high energies
A. Berti on behalf of the MAGIC Collaboration 3030

Prospects for VHE monitoring of gamma-ray bursts with SWGO
G. La Mura, U. Barres de Almeida, R. Conceição et al. 3041

Theoretical implications on the very high energy emission from GRB 190114C
D. Miceli, A. Berti, Z. Bosnjak et al. on behalf of the MAGIC Collaboration 3052

AGILE and GRBs: 13 years of observations
A. Ursi on behalf of the AGILE Team 3062

Searching for Gamma-Ray Bursts with the High-Altitude Water Cherenkov (HAWC) observatory
K. L. Engel for the HAWC Collaboration 3074

• **Electromagnetic Counterparts of Compact Binary Mergers**
Chairpersons: Jonathan Granot and Paz Beniamini

YouTube link: https://youtu.be/AS7bwaT48us

CALET search for gamma-ray counterparts of gravitational wave events
Masaki Mori for the CALET Collaboration 3084

• **Unusual and New Types of Gamma-Ray Bursts**
Chairperson: Binbin Zhang

YouTube link: https://youtu.be/VrF0iU8Q6us

Off-axis jet scenario for early afterglow emission of low-luminosity Gamma-Ray Burst GRB 190829A
Yuri Sato, Kaori Obayashi, Ryo Yamazaki, Kohta Murase and Yutaka Ohira 3095

• **Gamma-Ray Burst Correlations: Observational Challenges and Theoretical Interpretation**
Chairpersons: Maria Giovanna Dainotti and Liang Li

YouTube links: https://youtu.be/gcEy2h6y1jg
https://youtu.be/m2pCT6RAmI0
https://youtu.be/GTCvcVhBjN4

GRB prompt phase spectra under backscattering dominated model
Mukesh Kumar Vyas, Asaf Pe'er and David Eichler 3101

Applying models of pulsar wind nebulae to explain X-ray plateaux following short Gamma-Ray Bursts
L. C. Strang and A. Melatos 3107

Searching for strange quark planets
Xu Wang, Yong-Feng Huang and Bing Li 3118

Probe the universe by using Gamma-Ray Bursts with X-ray plateaus
Fan Xu and Yong-Feng Huang 3124

A new perspective on cosmology through Supernovae Ia and Gamma Ray Bursts
B. De Simone, V. Nielson, E. Rinaldi and M. G. Dainotti 3130

Theory of plateau phase in Gamma-Ray Bursts
Asaf Pe'er 3141

Exploring the canonical behaviour of long Gamma-Ray Bursts using an intrinsic multi-wavelength afterglow correlation
S. R. Oates, J. L. Racusin, M. De Pasquale et al. 3150

• **GRB 170817A and Binary Models**
Chairpersons: Marica Branchesi and Giulia Stratta
YouTube link: https://youtu.be/FRBvkaLX5WU

Kilonova emission observed so far: A comparison with AT2017gfo
A. Rossi 3162

• **Binary-Driven Hypernovae of Type 1, 2 and 3**
Chairpersons: Carlo Luciano Bianco, Christian Cherubini and Simonetta Filippi
YouTube link: https://youtu.be/8IFLfget3C0

Neutrinos and gamma-ray production from proton-proton interactions in binary-driven hypernovae framework
S. Campion, J. D. Melon Fuksman and J. A. Rueda Hernandez 3172

General relativistic turbulence in spherically symmetric core-collapse supernovae simulations
L. Boccioli, G. J. Mathews and E. P. O'Connor 3184

Gravitational Waves

• **Sources of Gravitational Waves**
Chairperson: Andrew Melatos
YouTube links: https://youtu.be/Lr8b9nKsFm4
https://youtu.be/xuqJOXPDyVY
https://youtu.be/qzOrgojHsMg
https://youtu.be/KCJoWMd71pE

Mountain formation by repeated, inhomogeneous crustal failure in a neutron star
A. D. Kerin and A. Melatos 3194

Gravitational waves from neutrino mass generating phase transitions
Nobuchika Okada and Osamu Seto 3206

Efficiency of registration of chirp bursts and signals of collapsing stars by the Euro-Asian network of GW interferometers
V. N. Rudenko, S. L. Andrusenko, D. P. Krichevskiy and G. D. Manucharyan 3219

Joint analysis method on gravitational waves and low-energy neutrinos to detect core-collapse supernovae
O. Halim, C. Casentini, M. Drago et al. 3228

• **Mid-frequency Gravitational Waves (0.1–10 Hz): Sources and Detection Methods**
Chairperson: Wei-Tou Ni
YouTube link: https://youtu.be/sJ6A7a73Vxw

A cryogenic and superconducting inertial sensor for the Lunar Gravitational–Wave Antenna, the Einstein Telescope and Selene-physics
F. Badaracco, J. V. van Heijningen, E. C. Ferreira and A. Perali 3245

Space gravitational wave antenna DECIGO and B-DECIGO
S. Kawamura and the DECIGO working group 3254

Summary of the parallel session GW2
Dongfeng Gao, Wei-Tou Ni, Jin Wang et al. 3261

• **Numerical Relativity and Gravitational Wave Observations**
Chairperson: Nigel Bishop

YouTube links: https://youtu.be/eZPytAU4Zmk
https://youtu.be/th-3KqkxDnU

Salient features of the optimised PyCBC IMBH search
Koustav Chandra, Archana Pai, V. Villa-Ortega et al. 3277

Matter shells modifying gravitational wave signals
Monos Naidoo, Nigel T. Bishop and Petrus J. van der Walt 3286

Odd-dimensional gravitational waves from a binary system on a three-brane
D. V. Gal'tsov and M. Khlopunov 3301

Developments in numerical relativity and gravitational wave observations
Nigel T. Bishop 3309

High Energy

• **Very High Energy Gamma Rays**
Chairpersons: Razmik Mirzoyan and Alessandro De Angelis

YouTube link: https://youtu.be/jHM1RH20ZyM

Insights into the Galactic Center environment from VHE gamma-ray observations with ground-based facilities
C. Fruck 3316

The TAIGA experiment
M. Tluczykont, I. I. Astapov, A. K. Awad et al. 3324

Science perspectives of the Southern Wide-field Gamma-ray Observatory (SWGO)
K. L. Engel for the SWGO Collaboration 3343

• **Future Missions for High-Energy Astrophysics**
Chairpersons: Filippo Frontera and Shaolin Xiong

YouTube link: https://youtu.be/JgDsZX6RUkU

Laue lenses: Focusing optics for hard X/soft Gamma-ray astronomy
L. Ferro, M. Moita, P. Rosati et al. 3355

ASTENA: A mission concept for a deep study of the transient gamma-ray sky and for nuclear astrophysics
E. Virgilli, F. Frontera, P. Rosati et al. on behalf of the ASTENA Collaboration 3368

Polarimetric prospects of a new hard X-soft gamma-ray space mission for next decades
M. Moita, L. Ferro, F. Frontera et al. 3385

• **The SRG Mission: First Results from eROSITA and ART-XC**
Chairperson: Andrea Merloni
YouTube link: https://youtu.be/l0t1B716UcM

Prospect for WHIM detection in the cosmic web by *SRG/eROSITA*
H. Tanimura and N. Aghanim 3400

• **eXTP – Enhanced X-Ray Timing and Polarimetry Mission**
Chairpersons: Marco Feroci and Fangjun Lu
YouTube link: https://youtu.be/3d08eKmsImI

The role of eXTP in the multi-messenger astronomy era
G. Stratta and Gor Oganesyan 3403

• **Observations of HE and UHE Cosmic Rays**
Chairpersons: Ivan De Mitri and Fabio Gargano
YouTube link: https://youtu.be/l9DZyBMWO5c

CALET on the ISS: The first 5 years
Pier Simone Marrocchesi for the CALET Collaboration 3427

The fluxes of charged cosmic rays as measured by the DAMPE satellite
Paolo Bernardini on behalf of the DAMPE Collaboration 3442

Recent results from the Pierre Auger Observatory
E. Roulet for the Pierre Auger Collaboration 3449

The HERD space mission
F.C.T. Barbato on behalf of the HERD Collaboration 3455

PART D

History of Relativity

• **The "Fall and Rise" of Betelgeuse**
Chairperson: Costantino Sigismondi
YouTube link: https://youtu.be/VmbrE2gYmOM

The observation of the stars in daytime and near the horizon
Costantino Sigismondi and Paolo Ochner 3471

Fall and Rise of Betelgeuse: The summary of HR1 session
Costantino Sigismondi and Paolo Ochner 3475

Photometry of Betelgeuse at daylight
Otmar Nickel 3479

Evidence for dynamical changes in Betelgeuse using multi-wavelength data
Sneha Kachhara, Sandip V. George, Ranjeev Misra and G. Ambika 3485

The curious case of Betelgeuse
Jacco Th. van Loon 3494

Variable stars observed from city sites: The 2500 AAVSO-SGQ database
Costantino Sigismondi and Paolo Ochner 3501

Betelgeuse: An introductory course to observational astronomy
Costantino Sigismondi and Tiziana Pompa 3507

The meridian line of the Vatican obelisk to study the stellar aberration
Costantino Sigismondi and Lorenzo Ricciardi 3513

Betelgeuse, Sirius and the stars in the roman *Settecento*
Costantino Sigismondi 3519

• **History of Relativity, Gravitation and Cosmology**
Chairperson: Luis Crispino

YouTube link: https://youtu.be/RNbPUSp95PQ

On Einstein's last bid to keep a stationary cosmology
Salvador Galindo-Uribarri and Jorge L. Cervantes-Cota 3536

Jayme Tiomno: Relativity, gravity, cosmology, and the Marcel Grossmann Meetings
William D. Brewer 3547

A look inside Feynman's route to gravitation
M. Di Mauro, S. Esposito and A. Naddeo 3563

Towards detecting gravitational waves: A contribution by Richard Feynman
M. Di Mauro, S. Esposito and A. Naddeo 3576

Stellar gravitational collapse, singularity formation and theory breakdown
Kiril Maltsev 3596

The Hamilton-Jacobi analysis by Peter Bergmann and Arthur Komar of classical general relativity
D. Salisbury 3626

• Time and Philosophy in Physics
Chairperson: Shokoufe Faraji

YouTube link: https://youtu.be/986v-V5JJEk

The passage of time and top-down causation
Barbara Drossel 3631

Explaining time's passage
Jonathan J. Dickau 3646

A glimpse to Feynman's contributions to the debate on the foundations of quantum mechanics
M. Di Mauro, S. Esposito and A. Naddeo 3657

Summary of the parallel session HR3
Shokoufe Faraji 3671

Neutron Stars

• Dense Matter in Compact Stars
Chairpersons: Alessandro Drago and Jorge Rueda

YouTube links: https://youtu.be/U5Yr0oDVhqY
https://youtu.be/MvMDFh-1_bc

Massive compact stars in the two-families scenario
P. Char, A. Drago and G. Pagliara 3677

Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond
J. Soldateschi, N. Bucciantini and L. Del Zanna 3684

Screening and elastic properties of the NS crust in the OCP approximation
D. Barba González, C. Albertus Torres and M. A. Pérez-García 3703

Tidal deformability as a probe of dark matter in neutron stars
D. Rafiei Karkevandi, S. Shakeri, V. Sagun and O. Ivanytskyi 3713

Binary neutron star mergers with quark matter equations of state
Atul Kedia, Hee Il Kim, Grant Mathews and In-Saeng Suh 3732

Probing dense matter physics with transiently-accreting neutron stars: The case of source MXB 1659-29
Melissa Mendes, Andrew Cumming, Charles Gale and Farrukh J. Fattoyev 3736

• **Compact Stars as Laboratories for Testing Strong Gravity**
Chairpersons: Aurora Perez Martinez and César Augusto Zen Vasconcellos

YouTube link: https://youtu.be/C1KGecdNEfs

Vacuum properties and astrophysical implications
A. Pérez Martínez, M. Pérez-Garcia, E. Rodríguez Querts and A. Romero Jorge 3756

Testing extended theories of gravity with GRBs
L. Mastrototaro 3762

• **Pulsar Power in Physics and Astrophysics and Pulsars and Pulsar Systems at High Energies**
Chairpersons: Andrea Possenti and Pak-Hin Tam

YouTube links: https://youtu.be/gAG29DZwUbM
https://youtu.be/ytdUBFrViHI

News and views regarding PSR J1757–1854, a highly-relativistic binary pulsar
A. D. Cameron, M. Bailes, V. Balakrishnan et al. 3774

On the origin of the unique isolated X-Ray pulsar 1E 161348-5055 with 6.7 hr. spin period
V. Yu. Kim 3785

Advantages of including globular cluster millisecond pulsars in Pulsar Timing Arrays
M. Maiorano, F. De Paolis, A. A. Nucita and A. Franco 3791

Searching for pulsars in globular clusters with the MeerKAT Radio Telescope
F. Abbate on behalf of the MeerTIME/TRAPUM Collaboration 3799

Precision Tests

• **Gravitational Lensing and Shadows**
Chairpersons: Perlick Volker and Oleg Tsupko

YouTube links: https://youtu.be/6DXXWpMJ3IQ
https://youtu.be/F-K1gXn71_Y
https://youtu.be/B-V8r2HMztw

Gravitational lensing by rotating Simpson–Visser black holes
Sushant G. Ghosh and Shafqat Ul Islam 3812

Killing tensors in foliated spacetimes and photon surfaces
Igor Bogush, Kirill Kobialko and Dmitri Gal'tsov 3827

Decoding black hole metrics from the interferometric pattern of relativistic images
V. Bozza 3839

Symplectic evolution of an observed light bundle
N. Uzun 3844

Shadow of black holes with a plasma environment in 4D Einstein-Gauss-Bonnet gravity
Javier Badía and Ernesto F. Eiroa 3856

Aspects of neutrino mass hierarchy in gravitational lensing
Himanshu Swami 3865

Photon regions in stationary axisymmetric spacetimes and umbilic conditions
K. V. Kobialko and D. V. Gal'tsov 3874

Gravitational lensing by charged accelerating black holes
Torben C. Frost 3885

• Experimental Gravitation
Chairpersons: Angela di Virgilio and Claus Lammerzahl

YouTube link: https://youtu.be/mXcxztQ0nyk
https://youtu.be/tE3gIUBviTc

A manmade experiment aimed to clarify the gravity law in the Solar system
Alexander P. Yefremov and Alexandra A. Vorobyeva 3905

Gravitomagnetic field generation using high permittivity materials in superconducting magnetic energy storage devices
G. V. Stephenson 3910

Large ring laser gyroscopes: Geometry stabilization and laser control
U. Giacomelli, N. Beverini, G. Carelli et al. 3920

Dark gravitomagnetism with LISA and gravitational waves space detectors
A. Tartaglia, M. Bassan, G. Pucacco, V. Ferroni and D. Vetrugno 3929

Light rays in the Solar system experiments: Phases and displacements
Pravin Kumar Dahal and Daniel R. Terno 3942

The Ginger project – preliminary results
C. Altucci, F. Bajardi, A. Basti et al. 3956

• **Variation of the Fundamental Constants, Tests of the Fundamental Symmetries and Probes of the Dark Sector**
Chairpersons: Angela Victor Flambaum and Yevgeny Stadnik

YouTube links: https://youtu.be/NAZA-0tWHak
https://youtu.be/juhDGBJ12Lg
https://youtu.be/zxGgDP2sn60

Varying fundamental constants and dark energy in the ESPRESSO era
C. J. A. P. Martins 3963

• **Dragging is Never Draggy: MAss and CHarge Flows in GR**
Chairperson: Oldrich Semerak

YouTube link: https://youtu.be/ho31IgLNxu8

Testing the general relativistic nature of the Milky Way rotation curve with Gaia DR2
Mariateresa Crosta 3970

Spinning cylinders in general relativity: A canonical form for the Lewis metrics of the Weyl class
L. Filipe O. Costa, José Natário and N. O. Santos 3982

Magnetized black holes: The role of rotation, boost, and accretion in twisting the field lines and accelerating particles
Ondřej Kopáček and Vladimír Karas 3999

Spinning particle: Is Newton-Wigner the only way?
V. Witzany 4010

Gravitomagnetic resonance and gravitational waves
Matteo Luca Ruggiero and Antonello Ortolan 4019

Quantum Gravity

• **Loop Quantum Gravity**
Chairpersons: Marcin Kisielowski and Jerzy Lewandowski

YouTube links: https://youtu.be/VaLPseYWh9E
https://youtu.be/WFIgMqrQ07U

Studying the EPRL spinfoam self-energy
Pietropaolo Frisoni 4026

A spin foam framework for the black-to-white hole transition
Farshid Soltani 4045

Holographic properties of the bulk-to-boundary transmission of information in regions of quantum space
Eugenia Colafranceschi 4062

• **Quantum Gravity Phenomenology**
Chairpersons: Giovanni Amelino-Camelia and Jerzy Kowalski-Glikman

YouTube links: https://youtu.be/vYnVb2zNl0o
https://youtu.be/icK8z80Hm-Y
https://youtu.be/svXLa0yhyYY

Minimal length discretization and properties of modified metric tensor and geodesics
Abdel Nasser Tawfik, Fady T. Farouk, F. Salah Tarabia and Muhammad Maher 4074

The structure of the multiverse from the entanglement entropy
Samuel Barroso Bellido 4082

Effective field theory from relativistic Generalized Uncertainty Principle
Vasil N. Todorinov, Saurya Das and Pasquale Bosso 4088

Stelle gravity as the limit of quantum gravity with maximal momentum
V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das 4107

Baryon asymmetry and minimum length
Saurya Das, Mitja Fridman, Gaetano Lambiase and Elias C. Vagenas . . . 4114

On quantum gravity and quantum gravity phenomenology
Douglas Edmonds, Djordje Minic and Tatsu Takeuchi 4126

WKB approach to the gravity-matter dynamics: A cosmological implementation
G. Maniccia and G. Montani 4146

Natural evidence for fuzzy sphere noncommutative geometry: Super-Chandrasekhar white dwarfs
Surajit Kalita, T. R. Govindarajan and Banibrata Mukhopadhyay 4159

A model of polymer gravitational waves: Theory and some possible observational consequences
Angel Garcia-Chung, James B. Mertens, Saeed Rastgoo, Yaser Tavakoli and Paulo Vargas Moniz 4166

• Loop Quantum Gravity: Cosmology and Black Holes
Chairpersons: Jorge Pullin and Parampreet Singh

YouTube links: https://youtu.be/CAuAK31Ukho
https://youtu.be/JlYnEzRiRR0

Primordial power spectrum from a matter-ekpyrotic scenario in loop quantum cosmology
Bao-Fei Li, Sahil Saini and Parampreet Singh 4178

The primordial power spectra in modified loop quantum cosmology
Bao-Fei Li, Javier Olmedo, Parampreet Singh and Anzhong Wang 4188

Primordial perturbations in kinetically dominated regimes of classical and quantum cosmology
B. Elizaga Navascués, R. Jiménez-Llamas and G. A. Mena Marugán 4193

Revisiting the Hamiltonian formalism of the Ashtekar–Olmedo–Singh black hole model
Alejandro García-Quismondo and Guillermo A. Mena Marugán 4211

A comparison of different choices of clocks in a reduced phase space quantization in loop quantum cosmology with an inflationary potential using effective techniques
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4228

Initial conditions in LQC/mLQCs
Bao-Fei Li, Parampreet Singh and Anzhong Wang 4234

Holonomy corrections in effective midisuperspace models
A. Alonso-Bardaji and D. Brizuela 4239

Infrared signatures of quantum bounce in collapsing geometry
Harkirat Singh Sahota and Kinjalk Lochan 4247

Effective black hole interior and the Raychadhuri equation
Keagan Blanchette, Saurya Das, Samantha Hergott and Saeed Rastgoo . . . 4256

Effect of loop quantization prescriptions on the physics of non-singular gravitational collapse
Kristina Giesel, Bao-Fei Li and Parampreet Singh 4267

Summary of the parallel session QG3
Jorge Pullin and Parampreet Singh 4272

Strong Field

• Strong Electromagnetic and Gravitational Field Physics: From Laboratories to Early Universe

Chairpersons: Sang Pyo Kim and She-Sheng Xue

YouTube links: https://youtu.be/kexqTayqFiU
https://youtu.be/laymGp6x9Hg

Reliable equations of state of viscous strong and electroweak matter
A. Tawfik 4277

Neutral fermion pair production by Sauter-like magnetic step
T. C. Adorno, Zi-Wang He, S. P. Gavrilov and D. M. Gitman 4290

On the magnetic field screening in strong crossed electromagnetic field
S. Campion, J. A. Rueda Hernandez, S.-S. Xue and R. Ruffini 4306

Particle creation by strong fields and quantum anomalies
José Navarro-Salas 4317

Constraints on the non-minimal coupling of electromagnetic fields from astrophysical observations
Susmita Jana and S. Shankaranarayanan 4326

New partial resummation of the QED effective action
Silvia Pla and José Navarro-Salas 4337

Can a detector detect soft photons
Sanved Kolekar and Jorma Louko 4347

Breaking of the adiabatic invariance in the production of particles by strong fields
P. Beltrán-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla 4352

Dynamics of relativistic electrons in non-uniform magnetic fields and its applications in quantum computing and astrophysics
Srishty Aggarwal and Banibrata Mukhopadhyay 4362

Validity of the semiclassical approximation in 1+1 electrodynamics: Numerical solutions to the linear response equation
Ian M. Newsome 4374

On Kerr black hole perfect MHD processes in Doran coordinates
C. Cherubini, S. Filippi, A. Loppini et al. 4387

Tadpole contribution to magnetic photon-graviton conversion
N. Ahmadiniaz, F. Bastianelli, F. Karbstein and C. Schubert 4393

• **The Effects of (Non)Linear Electrodynamics on the Properties of Astrophysical/Gravitational Compact Objects**
Chairperson: Seyed Hossein Hendi

YouTube link: https://youtu.be/SAIAXtQhavE

Correspondence of gamma radiation coming from GRBs and magnetars based on the effects of nonlinear vacuum electrodynamics
Tursynbek Yernazarov, Medeu Abishev and Yerlan Aimuratov 4401

Absorption of massless scalar waves by electrically charged regular black holes
Marco A. A. de Paula, Luiz C. S. Leite and Luís C. B. Crispino 4410

White Dwarfs

• **White Dwarf Explosions**
Chairpersons: Robert Fisher and María Pilar Ruiz Lapuente

YouTube links: https://youtu.be/ndaW6u2xuOo
https://youtu.be/IwjYaQaqJeI

Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass
Friedrich K. Röpke, Florian Lach, Sabrina Gronow et al. 4420

Type Ia supernovae and their explosive nucleosynthesis: Constraints on progenitors
Shing-Chi Leung and Ken'ichi Nomoto 4427

Charged polarized white dwarfs with finite temperature as a possible source of type Ia supernovae
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4447

• **White Dwarfs, Magnetic Compact Stars and Nuclear Astrophysics**
Chairpersons: Manuel Malheiro and Jaziel Goulart Coelho

YouTube link: https://youtu.be/onicFElJQnA

CTCV J2056-3014 and other fast-spinning white dwarfs
C. V. Rodrigues, R. Lopes de Oliveira, A. Bruch et al. 4453

Gravitational waves from fast-spinning white dwarfs
M. F. Sousa, J. C. N. de Araujo and J. G. Coelho 4461

Highly magnetized white dwarfs: Implications and current status
B. Mukhopadhyay, M. Bhattacharya, A. J. Hackett et al. 4475

Electron captures and stability of white dwarfs
N. Chamel, L. Perot, A. F. Fantina et al. 4488

Massive hot white dwarfs: Consequences of finite temperature in the structure and on the onset of instabilities
Sílvia P. Nunes, José D. V. Arbañil and Manuel Malheiro 4508

A study of the infrared emission of SGR/AXPs in a disk scenario and its implications for their origin
Sarah Villanova Borges 4514

Particle acceleration and high energy emission in the white dwarf binaries AE Aquarii and AR Scorpii
P. J. Meintjes, S. T. Madzime, Q. Kaplan et al. 4522

Study the effects of anisotropy on the highly magnetized white dwarfs
Debabrata Deb, Banibrata Mukhopadhyay and Fridolin Weber 4532

List of Participants 4545

PART D

This page intentionally left blank

The observation of the stars in daytime and near the horizon

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
www.icra.it/solar

Paolo Ochner
Padova University and Asiago Astrophysical Observatory, Italy

Traditionally when a star was near the solar conjunction, its magnitude was not measured from groundbased observations. In this way the lightcurves of variable stars show yearly gaps of about two or three months each year. Some observations made in twilight and in daytime in Rome of Betelgeuse, Antares and Venus are described.

Keywords: Atmospheric seeing, Airmass extinction, Daytime astronomical observations of stars.

1. Introduction: Betelgeuse and its 2020 secondary minimum

On 31 march 2020 Betelgeuse's luminosity was again in its rising phase, after the deep minimum Worldwide discussed. The observations of this phase continued in the evening until April 2020, on a dark sky. By the end of April the star was entering in twilight visibility for Rome, and its luminosity measurement with Argelander's method become complicate. The sky background rised, and the Purkinje effect affected the perception of the stellar luminosity by naked eye. Moreover the airmass' extinction near the horizon is affected by the humidity, which is a variable effect with the season and with the meteorological local conditions. Air turbulence also complicates the signal to noise ratio of the star. The cases of Antares at the meridian at 21° of altitude at sunset and of Venus at the meridian in afternoon, both observed in Vatican, are also discussed.

2. Sky background, airmass and comparison stars

A bright star is visible through a telescope during the day, the conditions required are: clear sky and exact pointing. For Betelgeuse in twilight, near the Western horizon in end of May in Rome, a dried tree in the foreground was used to aim at the star's position with a monocular or a small Newtonian telescope of 76 mm mirror.

For measuring the stellar magnitude the usual comparisons stars were not visible: Aldebaran, red as Betelgeuse, was already set and Procyon, white and normally

brighter than Betelgeuse, was not visible neither close to a landmark. Only Mercury, reddish, was bright and low enough to be seen and compared. γ Geminorum, white, also was detectable close to Betelgeuse and to another landmark.

These conditions allowed to verify that a "regular" $1/\sin$ cosecant law for the airmass extinction does not apply for the evening horizon in may in Rome.

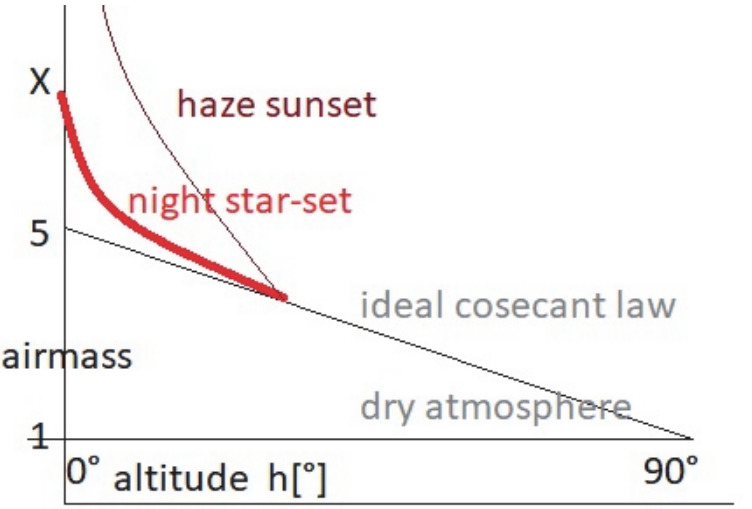


Fig. 1. Airmass extinction in dry atmosphere, night star set and haze sunset (as in end of August, in Rome). The deviation from the cosecant law are relevant below 15°.

Other experiences with sea sunsets showed strong departures from the cosecant law near the horizon in presence of humidity: the solar light can be extinct by an equivalent up to 30 airmasses of haze, or only by 5 airmasses in drier days. In the former situation the lower limb of the Sun disappears before reaching the horizon, while in the latter the Sun remains bright until the last glimpse of light.

2.1. *Atmospheric seeing*

Air turbulence is stronger in daytime, and large bubbles of hot air from locations close to the observer defocus the star. This fact complicates the visibility of the star on a bright sky background. Antares at the meridian, at 21° of altitude at sunset on 21 August 2021 was visible less than 50% of the duration of the observation, continuously disappearing and reappearing without having changed the focus of the telescope 10×42; only Venus, under clear sky, appeared steadily visible in daytime, through a monocular 7x18 on October 5, 2021. In other days of October 2021 the defocus affected also the observation of Venus.

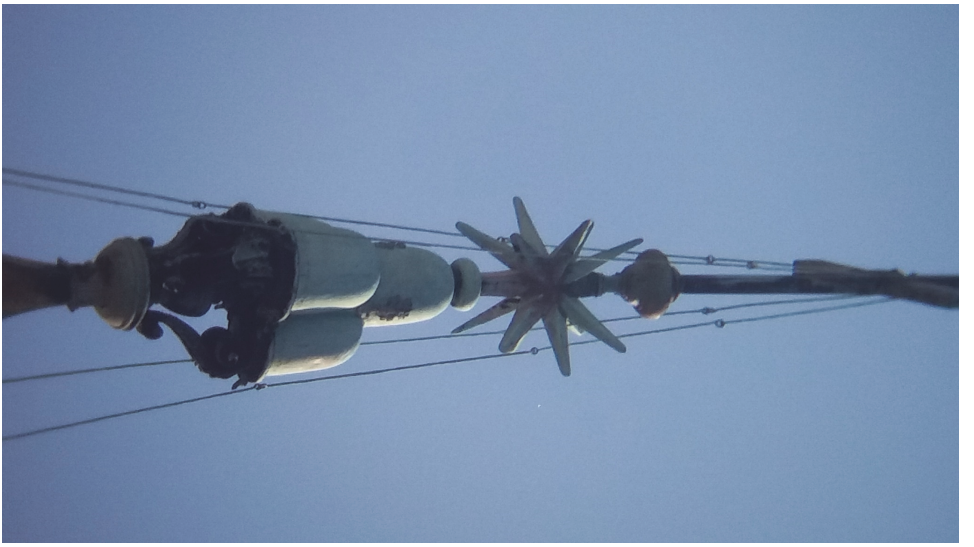


Fig. 2. A photo of Venus in daytime on October 5, 2021 at 16:10 local time, with the top of st. Peter's Obelisk in Vatican. In this photo Venus is below the center of the tridimensional star, as a bright point on the blue sky.

3. The early morning reappearances of Betelgeuse in September

My observations of Betelgeuse made in September from 2012 to 2021, and in general the first observations after the solar conjunction, showed a systematic increase of the perceived luminosity. Aldebaran, Procyon and Rigel are always used as comparison stars, and the airmass correction was carefully considered. The detector is my eye, and this systematic luminosity increasing, can be due to a physiological effect like the Purkinje's one, even if I applied the recommendation of not staring at the star. *Exempli gratia* on September 15, 2021 a careful observation, with the star "bracketed" with Aldebaran and Procyon and Rigel, yielded magnitude 0.27, 0.25 magnitudes brighter than V-band observations in AAVSO database for the same days. Moreover this luminosity is brighter than average maxima of Betelgeuse, which was nearly at the meridian on a starting twilight phase. Other September's luminous cases have to be attributed to the heavy haze at the Eastern horizon: either facing the Adriatic sea in Lanciano (265 m above sea level), either from the Asiago highland (1000 m). The clearness of the morning sky, and the low humidity dispersed in the atmosphere can be the responsible of such peculiar situation.

3.1. *Recovering airmass on the AAVSO-SGQ observations' catalogue*

The airmass of Betelgeuse during all these particular observations in September ranges from 2 to 3, and can be calculated on AAVSO-SGQ catalogue by selecting

SGQ as observer, in searching Alf Ori observations (and all the other variable stars observed).

The location of the observation is always reported in the notes, and the time is required by the application form. So the data for recomputing the airmass are made available.

4. Conclusions

The Argelander's method consists to bracket the luminosity of the star in study among two comparison stars. For Betelgeuse it is necessary to apply always the airmass corrections, because the angular distance of the comparison stars, usually Aldebaran and Procyon, produce artificial difference in luminosity.

The airmass correction can depart from the cosecant law near the horizon due to the haze (fig. 1).

Defocus effects of the air turbulence complicate this measurement. From Rome the magnitude of Betelgeuse has been measured as late as 31 May, in 2021. From the southern hemisphere in Australia Rod Stubbings could observe until the same date in 2021 and until 10 of June in 2020.

From Brasil, Florianopolis, Alexandre Amorim could observe from 8 July in the morning twilight.

The observations in V-band of Otmar Nickel were performed in daytime, during the solar conjunction, allowing a ungapped monitoring of the luminosity of this star.

References

1. C. Sigismondi, *Gerbertus* **13**, 65 (2020).
2. C. Sigismondi, ATel 13601 (2020), `https://www.astronomerstelegram.org/?read=13601`.
3. C. Sigismondi, Egress of Venus from St. Peter's obelisk (October 27, 2021), `https://youtu.be/ZWTg8xSYiuU`.
4. O. Nickel, *Photometry of Betelgeuse at daylight* MGM XVI Proceedings (2021).

Fall and Rise of Betelgeuse: The summary of HR1 session

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
www.icra.it/solar

Paolo Ochner
Padova University and Asiago Astrophysical Observatory, Italy

The occurrence of the deep minimum of Betelgeuse in 2020 has been object of a Worldwide debate, mainly on the public media. The session dedicated to Betelgeuse in the XVI Marcel Grossmann Meeting focused the many aspects of this debate, offering a good state of art of this subject, after and during an unprecedented observational effort from ground, airborne and from the space.

Keywords: Betelgeuse; Red giant; Variable stars.

1. Betelgeuse among the red giant variable stars

The special session HR1 *History of Relativity I* of the XVI Marcel Grossmann Meeting was dedicated to the Great Dimming of Betelgeuse, occurred from December 2019 to February 2020. This meeting gathered prominent scholars in this research field, and relevant representatives of the amateur astronomers who participated to the observational campaign of Betelgeuse (Wolfgang Vollmann from Vienna and Rod Stubbings from Australia).

The impact of Betelgeuse's dimming in the media, last year, is an index of how such an astronomical phenomenon can drive interest of the great public, and for how long, and why. One of the most recurrent issues was the supernova ending of its stellar history (because it was believed or announced to be imminent, as a consequence of the three months dimming... and this common mistake in the press, even specialized, was the reason why Relativistic Astrophysicists decided to clarify the nature of a potential gravitational collapse). Pre-supernova stages of supergiants are very poorly known, even for SN 1987A (Massimo Turatto, 2020 ICRANet Betelgeuse meeting), but a core-collapse event lasts only an hour, as recalled in the plenary session of this meeting (Chris Fryer).

In any case Betelgeuse ranks among the ten brightest stars in the sky, and it is variable, with amplitudes normally around 0.5 magnitudes from minimum to maximum, which become 1.0 magnitudes for the Great Dimming of 2020.

2. Observing Betelgeuse, the culture of mankind, and the didactic

Betelgeuse is part of the night sky culture of mankind, probably documented since Lascaux caves in the Pyrenees (Cro-Magnon civilization, 40000 years ago) and present in Greek mythology. Its dimming in 2020 has excited again the human fantasy, as well as an unprecedented observational campaign in all part of the spectrum.

Fig. 1. The Lascaux painting in the cave. The Taurus, at top center, is probably an Urus (Bos Primigenius). From this general view the detail of astronomical interest is shown below.

Fig. 2. The detail where Orion would appear. The Pleiades are also identified, Orion's belt is four point aligned in the previous global image. Andrea Dupree drawn our attention to this document.

This variability can be followed potentially by all, by applying some basic airmass correction to the reference stars and Betelgeuse itself (Costantino Sigismondi and Tiziana Pompa) even in twilight. The natural continuation of this approach, which is also didactic, is the daylight observations of Betelgeuse from ground, by Otmar Nickel, used to integrate the V and visual light curve of Betelgeuse obtained with the dark night sky. The search of peculiarities of "the painful shoulder"[4] of Orion is a natural way to approach this phenomenon, while the modern data on Betelgeuse magnitudes is less than two centuries (Sir John Herschel in 1837-41). How different is Betelgeuse from Antares and other red supergiant stars? In the timescales of stellar evolution how peculiar was this dimming? This provocation came from Jacco van Loon, the chairman of GAPS 2021, a 254-participants online meeting held in June 2021 on Red Supergiants.

3. Date of this meeting among the light curve of Betelgeuse

The date of this meeting occurred after a whole pulsational period (425 days) from the Great Minimum dated around 11 February 2020. This was the occasion to compare the Great Minimum with the following one, in order to verify the uniqueness of the 2020 event, or, conversely, its statistical fluctuation's nature. The light curves presented at this parallel session were all updated to the day before, thanks to the daytime and spacebased observations.

4. Time series analysis issues in the lightcurve of Betelgeuse

The time-series analysis of the light curve of Betelgeuse, to identify possible signatures of changes in the nature of the physical pulsations, has been the subject treated by Sandip V. George. The issue of new oscillation regime with period shorter than the main one of 425 days, seemed to appear from the preliminary light curves presented by Edward Guinan and Andrea K. Dupree: they exploited satellite and groundbased daytime measurements to follow the luminosity of the star during the months of solar conjunction.

5. ID card of Betelgeuse

The stellar parameters of Betelgeuse have been investigated in depth by Meridith Joyce by comparing observational data with synthetic data obtained with the best program of stellar evolution simulation, MESA. Mass, radius and distance have been reconsidered with this technique, while the measurements with astrometric satellites (from Hipparcos to GAIA) are affected by the asymmetry of the stellar surface of Betelgeuse. The opportunity to consider a recent merging history with a smaller companion was part of the presentation of Alexey Bobrick, concerning the dynamical history of Betelgeuse. The opportunity to observe Betelgeuse in the mid-infrared part of the spectrum from the stratosphere was presented by Graham

Harper, who showed that the variation of temperature on the photosphere of Betelgeuse, might be responsible for the dimming, more than "dust eclipses" produced by material ejections along the line of sight. Higher spatial resolution images obtained with the Hubble Space Telescope are confirming this issue, and the great hot and cold spots on the surface of the star are reflecting the convective cells on this supergiant, as the presentations of Edward Guinan and Andrea Dupree have shown from various perspectives.

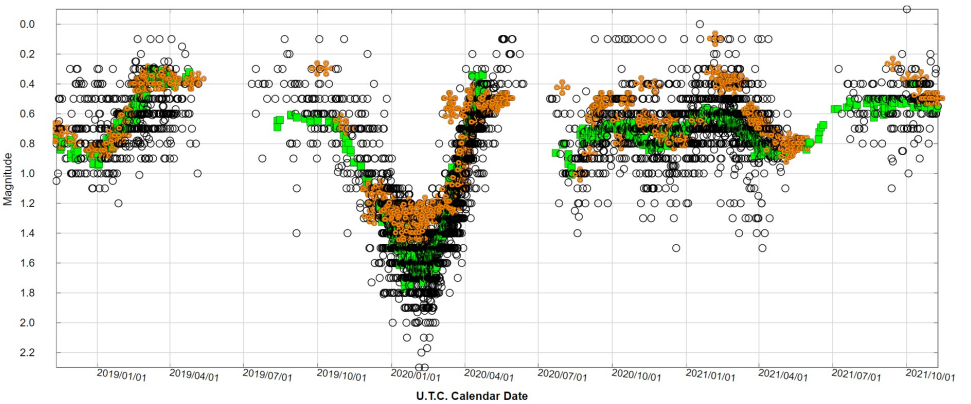


Fig. 3. The AAVSO light curve of Betelgeuse in the last three years (November 2018-2021). Evidenced in orange cross are 310 visual observations of the author. Note the green data in V-band during the Summer of 2021, by Otmar Nickel: they firstly cover completely the yearly gap during the solar conjunction. The visual data in the Summer 2020 are of Rod Stubbings from near Melbourne, Alexandre Amorim from Florianopolis and myself from Rome till end of May. The other V-band green data: 458 are of Wolfgang Vollmann who follows Betelgesue since 1976.

6. Perspectives

This session, held on July 8th, 2021, can be a reference for future studies on Betelgeuse and Red Giants in general, as well as an ideal introductory event to modern stellar variability, from naked eye observations to stellar interferometry techniques to airborne and space telescopes and satellite observations. The goal of having a good state-of-art meeting about Betelgeuse, its astrophysics and its observations, has been obtained.

References

1. Betelgeuse dimming: the state of the star International workshop (2020), `http://icranet.org/index.php?option=com_content&task=view&id=1281`.
2. GAPS 2021, IAU meeting 2021 `https://gaps2021.wixsite.com/conference`.
3. HR1 session of Marcell Grossmann Meeting 2021, `https://youtu.be/VmbrE2gYmOM`.
4. S. Wilk, JAAVSO **17**, 43, 1999 `https://app.aavso.org/media/jaavso/1743.pdf`.

Photometry of Betelgeuse at daylight

Otmar Nickel
zum Schollberg 11, 55129 Mainz, Germany
E-mail: otmar.nickel@web.de

In a backyard observatory in Mainz, Germany, several stars have been observed at daylight using a 250mm Newton telescope and a CCD camera. The sky background was measured on several days; values of 1.8 - 4.7 mag/arcsec2 were found at angles of 10°-100° distance to the Sun. Photometry of Betelgeuse was performed as a first attempt in 2020 with stacked images of Betelgeuse; it was improved in 2021 by using a neutral density filter (1 percent transmission) and measuring calibration and extinction coefficients on 4-8 bright reference stars. Photometry of Betelgeuse resulted in calculated errors of less than 0.05 mag from February to July 2021. It could be shown, that daylight observations of Betelgeuse can fill the observational gap with reliable magnitude data, when Betelgeuse is near to the Sun.

Keywords: Betelgeuse, photometry

1. Overview

Fig. 1 shows a light curve of Betelgeuse in the V-band, taken by many amateur astronomers with digital cameras (Kafka S, 2021). Unfortunately there are gaps in this curve, when Betelgeuse is near to the Sun, that means at a distance of less than about 30°. To fill these gaps there are 3 ways:
- Observations in twilight close to horizon (preferable in the southern hemisphere)
- Observations with space probes distant from earth (e.g. STEREO-A)
- Observations at daylight
In this paper I will report about daylight observations of Betelgeuse, which I have done this and the last year.

2. Equipment

The telescope of my observatory is a Newtonian telescope with 250 mm aperture on an equatorial mount. The mount is computer controlled with encoders in both axes. Images are acquired by a cooled CCD camera ATIK 460exm with a V filter and a neutral density filter (1 percent transmission), to reduce the light intensity. The camera is capable of exposures down to 1 ms. The front of the tube is prolonged with a cylindrical shade to reduce stray light from the Sun.

3. Sky brightness

The sky brightness can be measured in magnitudes per square arcsec; fig. 2 shows own measurements from 17 cloudless days (July to September 2020 and February to

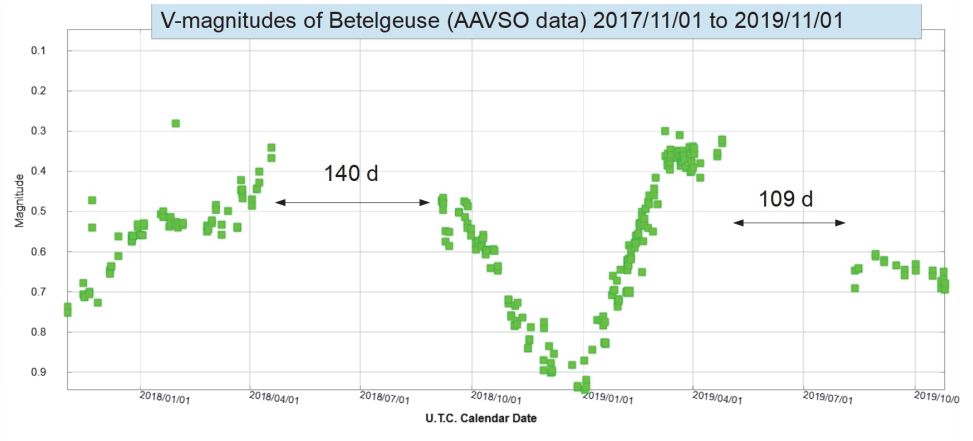


Fig. 1. Light curve of Betelgeuse 2017/11 to 2019/11 (AAVSO data)

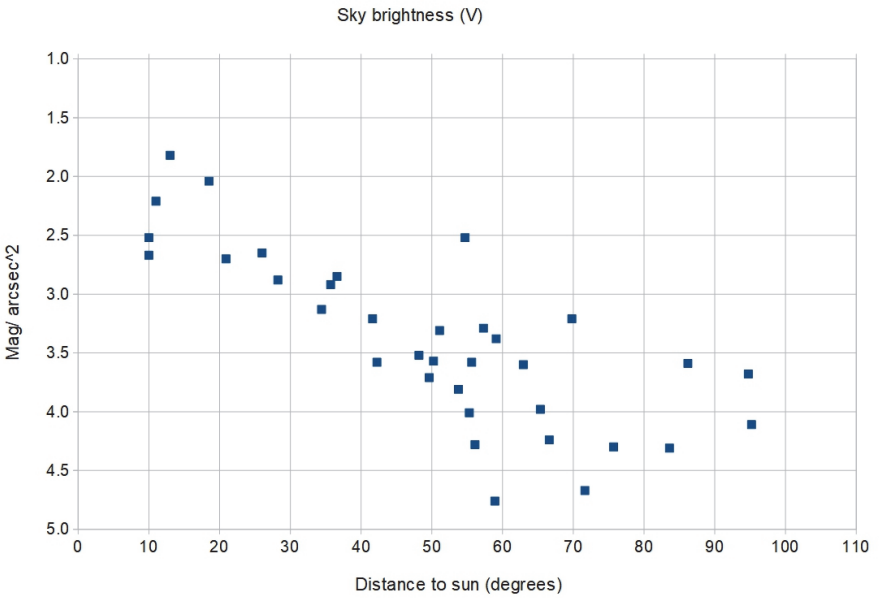


Fig. 2. Measured sky brightness versus solar angular distance

April 2021) at my observatory as graph. The sky brightness depends mainly on the angular distance to the Sun (which is shown here on the horizontal axis). Between 60 and 100 degrees the sky brightness seems to be stable, but here the scattering of data is very high, because other parameters (e.g. airmass and humidity) have more

influence. It can be shown, that stars can be imaged with reasonable signal-to-noise ratio, if their brightness in magnitudes is at least in the order of the sky brightness in mag per square arcsec. Therefore stars up to 2 magnitudes could be observed near to the Sun.

4. Methods

To get images of a star at daytime with a reasonable signal-to-noise ratio, one has to record a large number of images with short exposure time and to add them together to a "stacked" image, as seen in fig. 3.

To evaluate the brightness of a star, first an instrumental magnitude has to be processed, which is proportional to the logarithm of pixel counts corrected by the sky background. The same procedure is done for one or more stars, whose magnitude is known to get a standardized magnitude by comparison.

I did first measurements of Betelgeuse (in 2020) with only 1 comparison star, namely Aldebaran (alpha Tau). To get the correct magnitude, it is necessary, to measure also the extinction of all stars used for the comparison. Therefore the method was improved in 2021 by using an ensemble of 4-8 reference stars with brightness of up to 2 magnitudes and measuring the extinction constant.

The standard magnitude V_{var} can be calculated by the following equation (Da Costa GS, 1992):

$$V_{var} = m_{inst} + T_v(B-V)_{var} + m_0 - k_v X_{var} \tag{1}$$

where

m_{inst} = instrumental magnitude of variable
T_v = color transformation constant for V-Band
$(B-V)_{var}$ = color index of variable
m_0 = zero point (depending on sensitivity of the system)
k_v = extinction constant
X_{var} = airmass of the variable star

m_0 and k_v were determined for each measurement from the color corrected instrumental magnitudes of the ensemble of reference stars by linear regression.

5. Results

Fig. 4 shows the 2020 light curve of Betelgeuse with the results of my first attempts for daylight measurements (seen as crosses). The 3 red points are measurements of the STEREO-A space probe (Dupree A, et al, 2020). The errors of the daylight measurements were in the order of 0.07 mag.

Fig. 5 shows the 2021 light curve, again with my daylight measurements, seen as crosses. The gap of values could be relatively well filled with data, the errors were less than 0.05 mag, even at the closest distance of 16° to the Sun. There is a light

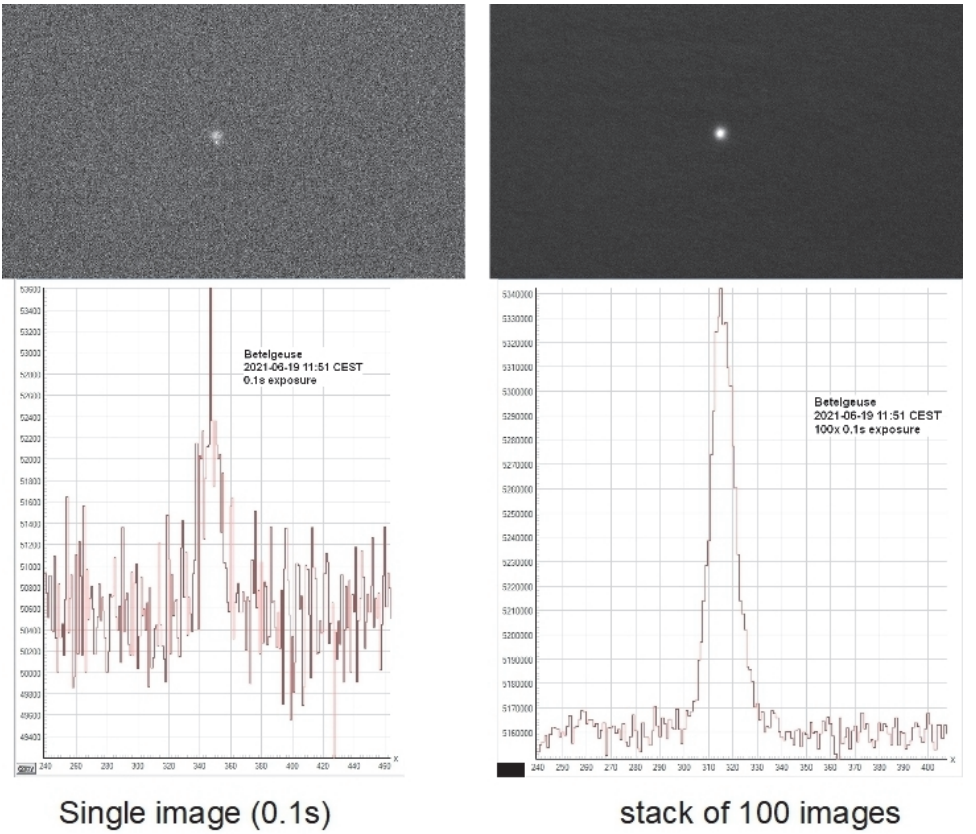


Fig. 3. Images and profiles of Betelgeuse near to sun, left: single image, right: stack of 100 images

systematic error of 0.02 magnitudes; the reason for that may be a slight error of the color transformation, which can be significant for Betelgeuse with its highly red color index.

The method was tested also on a star which comes even closer to the Sun than Betelgeuse: The V magnitude of Alpha Ari was measured on 6 days at a distance to the Sun of 35° down to 10°. The mean value of all measurements was 2.0 ±0.035, which is very close to the Hipparcos magnitude (2.01 mag).

6. Conclusions

- Measurements of the V magnitude of Betelgeuse are possible with astronomical cameras at daylight, even at closest distance to the Sun
- The method gives reliable results with an error of less than 0.05 mag (clear sky provided)

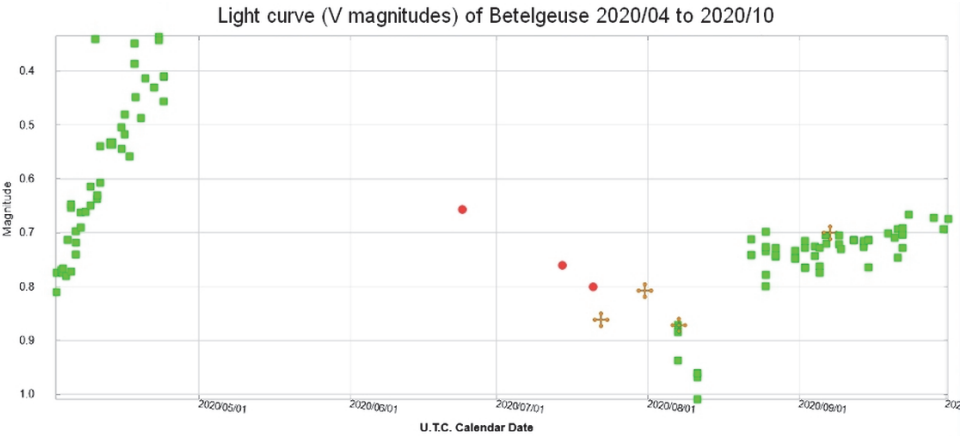


Fig. 4. Light curve of Betelgeuse 2020; green points: V mag data (AAVSO data base), crosses: daylight data, red dots: STEREO-A data

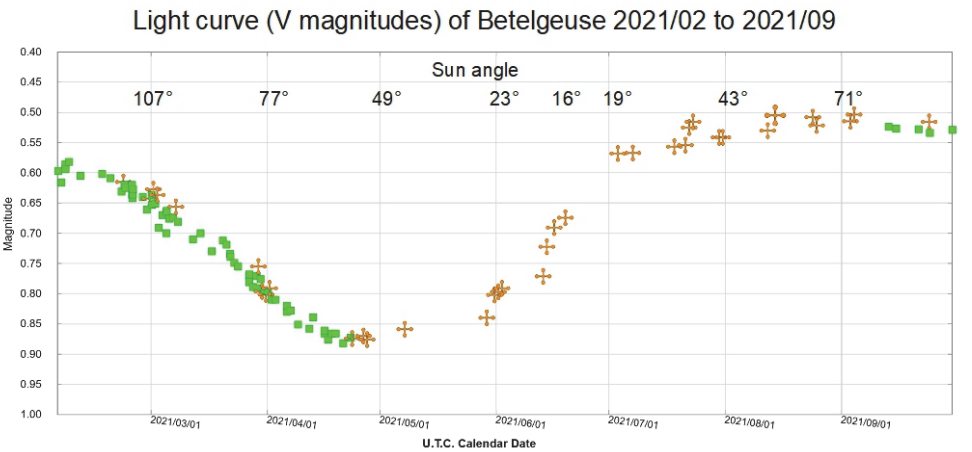


Fig. 5. Light curve of Betelgeuse 2021 (Kafka S, 2021); green points: V mag data (AAVSO data base), crosses: daylight data

- The solar conjunction gap of the light curve can be filled in the future, if the method is used also by other observers with suitable equipment.

Acknowledgements

The author would like to thank Prof. Sigismondi for the invitation to the MG16 meeting and for his help in preparing this paper.

References

1. Kafka, S., *Observations from the AAVSO International Database* (2021) `https://www.aavso.org`
2. Da Costa, GS, *Basic photometry techniques* in: Astronomical CCD Observing and Reduction Techniques **23**, 90 (1992)
3. Dupree, A and Guinan, E and Thompson, WT, *Photometry of Betelgeuse with the STEREO mission while in the glare of the Sun from Earth* The Astronomer's Telegram **13901** (2020)

Evidence for dynamical changes in Betelgeuse using multi-wavelength data

Sneha Kachhara

Indian Institute of Science Education and Research (IISER) Tirupati, Tirupati, India
E-mail: snehakachhara@students.iisertirupati.ac.in

Sandip V. George

Department of Computer Science, University College London, London WC1E 6BT, United Kingdom
Interdisciplinary Center for Psychopathology and Emotion Regulation, University Medical Center Groningen
University of Groningen, The Netherlands
E-mail: sandip.george@ucl.ac.uk

Ranjeev Misra

Inter University Center for Astronomy and Astrophysics (IUCAA) Pune, Pune, India
E-mail: rmisra@iucaa.ac.in

G. Ambika

Indian Institute of Science Education and Research (IISER) Tirupati, Tirupati, India
E-mail: g.ambika@iisertirupati.ac.in

The reasons behind the Great Dimming and subsequent rising in the brightness of Betelgeuse between October 2019 and March 2020 still continues to baffle astronomers. It has been shown by George et al. (2020) that critical slowing down preceded the dimming event. This suggested that the dimming was as a result of the change in the nature of the nonlinear dynamics of the star. In this work we present additional evidence for dynamical changes in Betelgeuse prior to the Great Dimming event, using nonlinear time series analysis. We study the relations between the different bands in the photometry data collected from the Wing photometery (IR/near-IR) and Wasatonic observatory (V-band). We also analyse how the early warning signals studied previously changed during and after the Great Dimming.

Keywords: Critical slowing down, alpha orionis, recurrence quantification analysis, wing photometry.

1. Overview

The unprecedented dimming and subsequent brightening events in Betelgeuse between October 2019 and March 2020 have since been studied in extensive detail.[1–3] Three main hypotheses have been suggested as reasons for the dimming, namely changes in pulsation dynamics, star spots and a dust cloud. Combinations of these

hypotheses have also been suggested as reasons for the dimming.[4,5] We will explore each of these explanations below, before explaining how the nonlinear dynamics of the light curve can contribute towards understanding the dimming.

The most prominent reason suggested for the Great Dimming observed in Betelgeuse has been the formation of a dust cloud. This has been put forward as an explanation by multiple authors. Measurements of effective temperature have suggested that no significant drop was observed during the dimming.[6] In addition, high angular resolution images showed a significant dimming in the southern hemisphere of Betelgeuse during the dimming event.[7] Observations from multiple wavelengths suggest that an outflow from the star itself, enhanced by pulsations, condensed around the southern hemisphere of the star.[4,5] It has also been suggested that the dimming could be explained by an ejection from the star that cools below 3000 K to form molecules, before condensing to dust.[8]

The dust hypothesis has been disputed as the sole reason for the dimming by a number of authors. Radiative transfer models using data from sub-millimeter wavelengths suggested that the dimming was a result of changes in the photosphere as opposed to external dust.[9] Further, observations during the dimming suggested that there was little change in the IR flux.[10] Moreover, Wing three-filter TiO and near-IR photometry found that the effective temperature during the dimming could be much lower than previously calculated, suggesting that the dimming observed in the visible spectrum could be due to photospheric motions.[11] Many of these authors proposed the presence of spots on the stellar surface as an alternative reason for the dimming of Betelgeuse.[9,12]

Finally, it has been suggested that the observed dimming was as a result of changes in pulsational dynamics.[11,13,14] Multiple authors have pointed out that the Great Dimming coincided with the minima of both the 430 days and 5.8 year periods.[11,15,16] However, the observed decrease in temperature in the star could not be explained by pulsational dynamics alone.[12]

Nonlinear time series analysis on the light curve of Betelgeuse prior to the major dimming event suggested that a critical slowing down was observed in the light curve well before the actual dimming occurred. When a nonlinear dynamical system undergoes a change of state, it often does so suddenly and with no change in the mean response. However, prior to many such changes in state (called a critical transition), multiple time series quantifiers called early warning signals (EWS) are shown to increase.[17] The presence of critical slowing down in Betelgeuse indicated that the star was about to undergo a change of state in its dynamical properties. It was suggested that the dimming phenomenon was the change of state itself or occurred as a result of it.[14] This work suggested that the reasons behind the dimming could be explained by studying the light curve before the Great Dimming occurred.

In this paper we explore the dynamical properties of Betelgeuse in greater detail, by studying how they changed before, during and after the Great Dimming. We do this by analysing Wing three-filter TiO and V band photometry measured at

the Wasatonic Observatory[11,18] and V/Vis data from the American Association of Variable Star Observers (AAVSO).[19] We first gain further evidence for a dynamical change in Betelgeuse by studying the multiwavelength Wing data. We then study changes in the linear correlations between the wavelengths and study how they changed over the period leading up to the dimming. Finally we study how the quantifiers studied by George et al.[14] changed in the AAVSO light curve, in the period leading up to, during and after the Great Dimming.

2. Analysis

2.1. *Wing Three-filter and V-band Photometry*

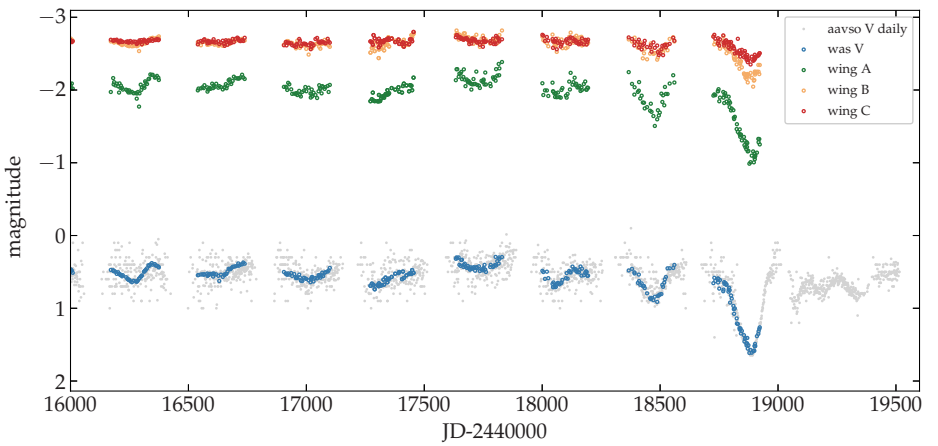


Fig. 1. Variation of intensity in Betelgeuse, as captured in Vis/V band AAVSO data (grey), V-band photometry from Wasatonic laboratory (blue), and Infrared Wing three-filter (A: green, B: orange, C: red) data.

Wing photometry consists of an eight band measurement of electromagnetic radiation from an astronomical source, initially developed for studying M supergiant stars.[20] Later, a simplified three-filter system was proposed that measures infrared magnitude, color and spectral type for a star.[21] In this study, photometric data (IR/near-IR) from three Wing filters,[21] and a wideband V filter collected by the Wasatonic observatory, are analyzed. The V-band filter is centered around 5550 Å. The Wing A filter is dominated by the TiO 7190 Å. The Wing B and C filters are continuum bands centered around 7500 Å and 10240 Å respectively. More information about the data may be found in Harper et al. (2020).[11]

We first calculate the cross-correlations (see figure 2 (a)), using Person Correlation Coefficient among data from different bands in a sliding window fashion with a window size = 300 points. The V-band and A band are already known to exhibit

strong correlation,[11] but we find that this correlation decreases prior to dimming. It is interesting that this trend starts much before the actual dimming episode, hinting towards a prolonged process. Moreover, we find this trend to be simultaneous in all pairs of data.

We then employ nonlinear time series analysis techniques to analyze the multivariate data. Building on dynamical systems theory, our techniques assume the existence of a state space representation for the data. Since dynamical features of the source are intimately related to the topological properties in state space, its quantification can yield valuable information.

One of these helpful properties is recurrence: the tendency of a system to visit the neighborhoods of previously occupied states. Recurrence is a prominent feature of deterministic dynamics and can be easily captured in terms of Euclidean distances among points on the attractor (trajectories) in state space.[22] When presented in matrix form, it is known as the Recurrence matrix (R). A 2-dimensional visual representation of R is known as the recurrence plot (RP). Statistical measures on RP, collectively known as the Recurrence Quantification Analysis (RQA), reveal subtle characteristics of the dynamics at play. Changes in some of these RQA measures may serve as EWS for a system close to a dynamical transition.[23] For the multivariate data, we make use of Recurrence Rate (RR) to study dynamical features of the Betelgeuse light curve.

We analyze recurrence properties of the data in the same sliding window fashion, keeping distance threshold constant at 0.1 for calculation of RP. We do this separately for each band, and present the results for RR (fraction of points in the recurrence matrix) in figure 2 (b). Again, we find an increasing trend in all wavelengths. Interestingly, the trend is most consistent for IR bands. Further, it corresponds well to the correlation indicated in figure 2 (a). This indicates that if there is a dynamical process at play, it may not be a peripheral phenomenon. Similar variation across bands hints at the possibility of change at the stellar scale, probably in the pulsation dynamics.

These results agree well with EWS observed in the AAVSO data,[19] reported previously.[14] However, the length of available data limits an extensive analysis, at this point in time.

2.2. *Beyond the Great Dimming*

In this section we present results on how the nonlinear parameters of the light curve changed after the Great Dimming, using V band data from the AAVSO.[19] In George et al. (2020),[14] we reported signatures of dynamical transitions in this data in the form of critical slowing down. We reported an increase in multiple EWS including the autocorrelation at lag 1 (ACF(1)), variance and recurrence based measures in the data leading up to the dimming. The increase in these quantifiers indicated

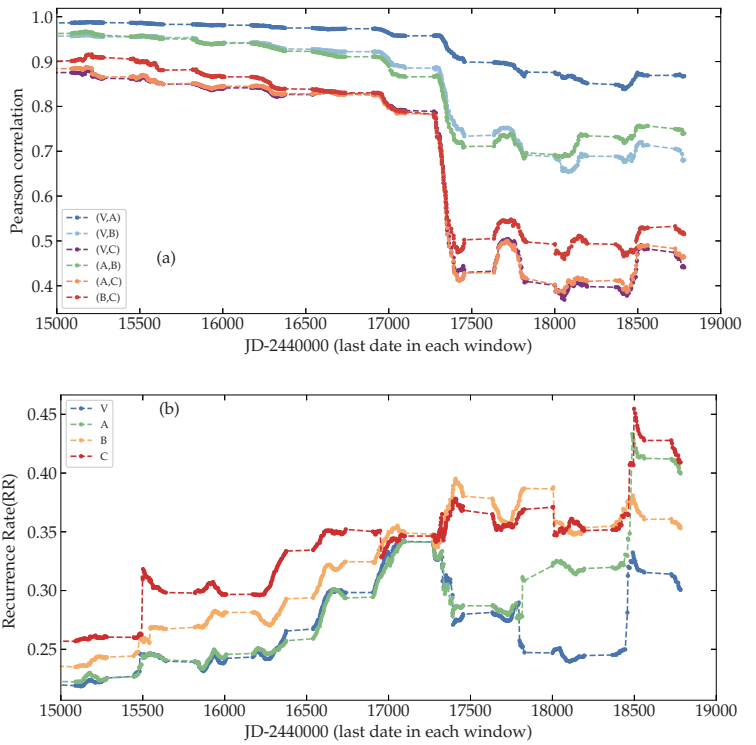


Fig. 2. (a) Cross-correlation between pairs of data from different bands calculated in a sliding window fashion upto (but not including) dimming. (b) Recurrence Rate (RR) obtained with fixed recurrence threshold = 0.1, for each band. Window size = 300.

that a critical transition was imminent. It was assumed that the Great Dimming occurred as a result of this critical transition. In the present paper we go beyond the Great Dimming and analyse how the EWS changed since.

2.2.1. *Autocorrelation and Variance*

To do this we initially bin the data from 1990 onwards into 10 day bins, and calculate the ACF(1) and variance in windows of size 300. This is presented in figure 3.1. Both the ACF(1) and Variance show an increase leading up to the dimming, a sharp rise when the Great Dimming occurred, and a slow reduction after the dimming ended. In figure 3.2 we select an area around the dimming, bin the data into 5 day bins and choose a window size of 40 points. The data again shows the same trends, with a rise leading up to the dimming, a sharp rise during the Great Dimming and a fall since. With the smaller window and bin size, the error on the quantifiers increases (changing approximately as $\frac{1}{\sqrt{n}}$).

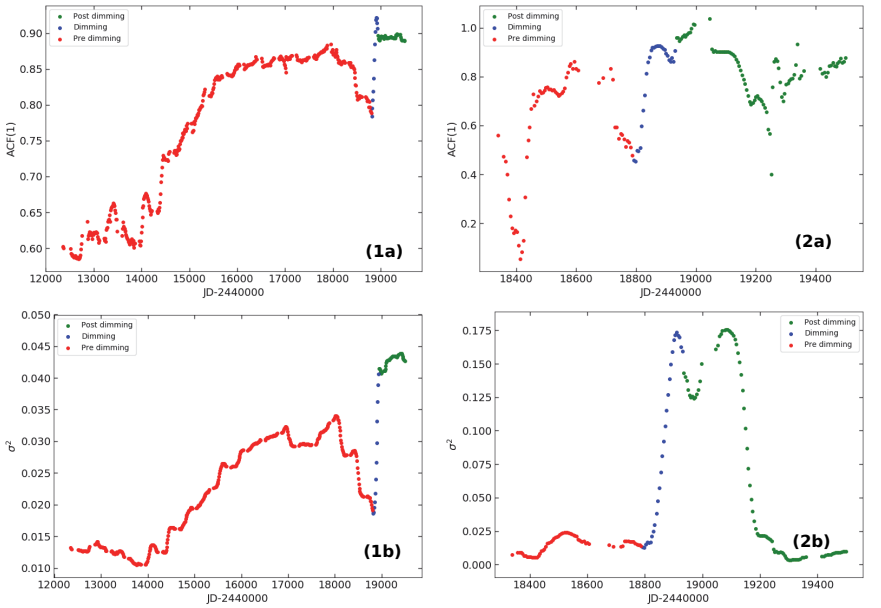


Fig. 3. Variation in the (a) Autocorrelation at lag 1 and (b) Variance calculated from the light curve of Betelgeuse upto the dimming (red), during the dimming (blue) and post dimming (green). The left panels (figures 1a and 1b) show the variation from 1990 onwards using a window size of 300 points on data binned every 10 days. The right panels zoom into a region around the dimming, with data binned every 5 days with a window size of 40.

2.2.2. *Recurrence Measures*

We then analyze AAVSO data in terms of recurrence measures as well. For this, we fix the recurrence rate (RR) as 0.1, and calculate Determinism (DET) and Laminarity (LAM) in the same sliding window fashion. We find a consistent rise in these measures prior to dimming, followed by a sudden dip and subsequent rise (see figure 4 (a) and (b)). We also present typical recurrence plots for two windows near dimming episode (one prior and one including dimming) in figure 4 (c) and (d) respectively as a visual aid. Clearly, the sudden drop in brightness affects the structure of the RP, and the recurrence measures reflect it. However, it is interesting to see the rise prior to dimming, possibly caused by dynamical factors (details discussed in George et al.[14]). The trend in post dimming behaviour seems in agreement which could be verified with future observations in the coming years.

3. Discussion

We presented the nonlinear analysis on two datasets measuring the brightness of Betelgeuse before, during and after the Great Dimming of 2019-20. We conducted a multivariate time series analysis of the data from Wing-IR photometry and the

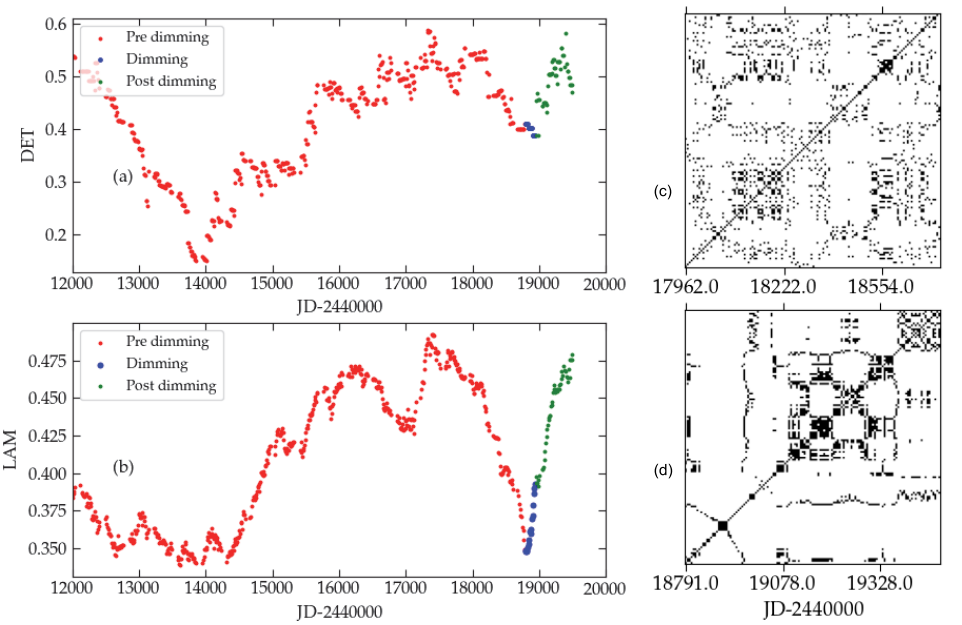


Fig. 4. Recurrence measures (a) Determinism (DET) and (b) Laminarity (LAM) calculated from the light curve of Betelgeuse upto the dimming (red), during (blue), and post dimming (green). The left panels (figures a and b) show the variation from 1990 onwards using a window size of 300 points on data binned every 10 days. The right panels (figures c and d) show typical recurrence plots for prior and during dimming respectively. The recurrence rate was fixed at 0.1.

V band photometry from Wasatonic laboratory. We show that prior to the dimming, the cross correlations and individual recurrence measures vary considerably. Our results suggest that a change in the dynamics of Betelgeuse commenced in advance of the Great Dimming of 2019-20. We also analyse early warning signals in AAVSO light curve for the period including and after the Great Dimming. While not conclusive, we see evidence that the values of the early warning signals have reduced since the Great Dimming.

Our results make a strong case that the dynamics of Betelgeuse underwent significant changes leading up to the Great Dimming. Variable stars are known to undergo changes in their nonlinear properties as they evolve.[24–26] In this study we observe that the recurrence properties, cross correlations and early warning signals change prior to the Great Dimming. In addition, Betelgeuse is known to exhibit multiple frequencies in its light curve, including a short period, thought to be driven by the κ mechanism, a long period thought to be convection driven[27] and a recently detected overtone of the short period.[11,28] It is possible that the pulsational dynamics of Betelgeuse, including the birth of new periods may have preceded the Great Dimming, and these need to be examined in detail.

Taken together with the study on the light curve of Betelgeuse by George et al. (2020),[14] our results seem to indicate that the Great Dimming was not an isolated incident, and to understand the reasons for the dimming, we must consider changes in the dynamics before it.

Acknowledgments

We acknowledge Dr. Graham Harper for useful discussions, and Richard Wasatonic and Prof. Edward Guinan for the multi wavelength data used in this study. We acknowledge with thanks the variable star observations from the AAVSO International Database contributed by observers worldwide and used in this research. SVG acknowledges financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovative programme (ERC-CoG-2015; No 681466 awarded to M. Wichers). SK acknowledges financial support from the Council of Scientific and Industrial Research (CSIR), India.

References

1. E. F. Guinan, R. J. Wasatonic and T. J. Calderwood, The fainting of the nearby red supergiant betelgeuse, *The Astronomer's Telegram* **13341**, p. 1 (2019).
2. E. F. Guinan, R. J. Wasatonic and T. J. Calderwood, Updates on the "fainting" of betelgeuse, *The Astronomer's Telegram* **13365**, p. 1 (2019).
3. C. Sigismondi, Rapid rising of betelgeuse's luminosity, *The Astronomer's Telegram* **13601**, p. 1 (2020).
4. A. K. Dupree, K. G. Strassmeier, L. D. Matthews, H. Uitenbroek, T. Calderwood, T. Granzer, E. F. Guinan, R. Leike, M. Montargès, A. M. Richards *et al.*, Spatially resolved ultraviolet spectroscopy of the great dimming of betelgeuse, *The Astrophysical Journal* **899**, p. 68 (2020).
5. Kravchenko, K., Jorissen, A., Van Eck, S., Merle, T., Chiavassa, A., Paladini, C., Freytag, B., Plez, B., Montargès, M. and Van Winckel, H., Atmosphere of betelgeuse before and during the great dimming event revealed by tomography, *A&A* **650**, p. L17 (2021).
6. E. M. Levesque and P. Massey, Betelgeuse just is not that cool: Effective temperature alone cannot explain the recent dimming of betelgeuse, *The Astrophysical Journal Letters* **891**, p. L37 (2020).
7. M. Montargès, E. Cannon, E. Lagadec, A. de Koter, P. Kervella, J. Sanchez-Bermudez, C. Paladini, F. Cantalloube, L. Decin, P. Scicluna *et al.*, A dusty veil shading betelgeuse during its great dimming, *Nature* **594**, 365 (2021).
8. B. Davies and B. Plez, The impact of winds on the spectral appearance of red supergiants, *Monthly Notices of the Royal Astronomical Society* (2021).
9. T. E. Dharmawardena, S. Mairs, P. Scicluna, G. Bell, I. McDonald, K. Menten, A. Weiss and A. Zijlstra, Betelgeuse fainter in the submillimeter too: An analysis of jcmt and apex monitoring during the recent optical minimum, *The Astrophysical Journal Letters* **897**, p. L9 (2020).
10. R. Gehrz, J. Marchetti, S. McMillan, T. Procter, A. Zarling, J. Bartlett and N. Smith, Betelgeuse remains steadfast in the infrared, *The Astronomer's Telegram* **13518**, p. 1 (2020).

11. G. M. Harper, E. F. Guinan, R. Wasatonic and N. Ryde, The photospheric temperatures of betelgeuse during the great dimming of 2019/2020: No new dust required, *The Astrophysical Journal* **905**, p. 34 (2020).
12. S. Alexeeva, G. Zhao, D.-Y. Gao, J. Du, A. Li, K. Li and S. Hu, Spectroscopic evidence for a large spot on the dimming betelgeuse, *Nature communications* **12**, 1 (2021).
13. E. F. Guinan and R. J. Wasatonic, The continued unprecedented fading of betelgeuse, *The Astronomer's Telegram* **13410**, p. 1 (2020).
14. S. V. George, S. Kachhara, R. Misra and G. Ambika, Early warning signals indicate a critical transition in betelgeuse, *Astronomy & Astrophysics* **640**, p. L21 (2020).
15. J. R. Percy, What's up with betelgeuse?, *Journal of the Royal Astronomical Society of Canada* **114** (2020).
16. C. Sigismondi, Betelgeuse 2020 dimming: getting the minimum, *Gerbertvs, International Academic Publication on History of Medieval Science* **13**, 39 (2020).
17. M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk and G. Sugihara, Early-warning signals for critical transitions, *Nature* **461**, 53 (2009).
18. G. M. Harper, E. Chambers, W. D. Vacca, H. Wiesemeyer, D. Fadda, C. deWitt, R. Wasatonic, A. M. S. Richards, N. Ryde, C. Fischer, M. J. Richter, E. F. Guinan, R. Minchin, U. U. Graf and S. Colditz, Sofia upgreat/fifi-ls emission line observations of betelgeuse during the great dimming of 2019/2020, *The Astronomical Journal* **Accepted** (2021).
19. S. Kafka, Observations from the AAVSO International Database (2021).
20. N. White and R. Wing, Photoelectric two-dimensional spectral classification of m supergiants, *The Astrophysical Journal* **222**, 209 (1978).
21. R. F. Wing, Three-color narrow-band photoelectric photometry of red variables, *Journal of the American Association of Variable Star Observers (JAAVSO)* **21**, 42 (1992).
22. N. Marwan, M. C. Romano, M. Thiel and J. Kurths, Recurrence plots for the analysis of complex systems, *Physics reports* **438**, 237 (2007).
23. N. Marwan, S. Schinkel and J. Kurths, Recurrence plots 25 years later—gaining confidence in dynamical transitions, *EPL (Europhysics Letters)* **101**, p. 20007 (2013).
24. J. R. Buchler and G. Kovacs, Period doubling bifurcations and chaos in w virginis models, *The Astrophysical Journal* **320**, L57 (1987).
25. R. Szabó, Z. Kolláth, L. Molnár, K. Kolenberg, D. Kurtz, S. Bryson, J. Benkő, J. Christensen-Dalsgaard, H. Kjeldsen, W. Borucki *et al.*, Does kepler unveil the mystery of the blazhko effect? first detection of period doubling in kepler blazhko rr lyrae stars, *Monthly Notices of the Royal Astronomical Society* **409**, 1244 (2010).
26. S. V. George, R. Misra and G. Ambika, Fractal measures and nonlinear dynamics of overcontact binaries, *Communications in Nonlinear Science and Numerical Simulation* **80**, p. 104988 (2020).
27. R. B. Stothers, Giant convection cell turnover as an explanation of the long secondary periods in semiregular red variable stars, *The Astrophysical Journal* **725**, p. 1170 (2010).
28. M. Joyce, S.-C. Leung, L. Molnár, M. Ireland, C. Kobayashi and K. Nomoto, Standing on the shoulders of giants: new mass and distance estimates for betelgeuse through combined evolutionary, asteroseismic, and hydrodynamic simulations with mesa, *The Astrophysical Journal* **902**, p. 63 (2020).

The curious case of Betelgeuse

Jacco Th. van Loon

Lennard-Jones Laboratories,
Keele University, ST5 5BG, UK
E-mail: j.t.van.loon@keele.ac.uk
www.astro.keele.ac.uk/∼jacco

Betelgeuse is the nearest red supergiant, one of the brightest stars in our sky, and statistically speaking it would be expected to be "typical". Yet it exhibits many features that seem "curious", to say the least. For instance it has a high proper motion. It rotates fast. It has little dust. It dimmed unexpectedly. Is any of these, and other, phenomena atypical, and taken together does it make Betelgeuse atypical? This is important to know, because we need to know whether Betelgeuse might be a prototype of red supergiants in general, or certain subclasses of red supergiants, since we can study it in such great detail. It is also important to know as it may be a link to understanding other, apparently atypical cases such as supernova 1987A, and maybe even such exotica as Thorne-Żytkov objects. Studying this question in itself helps us understand how we deal with rarity and coincidence in understanding the Universe we live in.

Keywords: Red supergiants – stellar variability – dust formation – supernova progenitors – Philosophy of Science – Betelgeuse.

1. Prologue

Not far in the sky from the most famous supernova remnant, the Crab Nebula, the most famous red supergiant, Betelgeuse, has been a familiar sight for peoples in either hemisphere. Distinctively orange, and one of the brightest lights on the nightly celestial firmament, its prominent position in the corner of arguably the most iconic constellation, Orion, Betelgeuse commands our attention and this in itself makes it special. This perspective aims to challenge the uniqueness and surprising behaviour of many a favourite star, and what it tells us about us. It could be seen as a more philosophical sequel to the Paris 2012 workshop on Betelgeuse.[1]

2. Is Betelgeuse typical?

It's got to be, right? Surely, as the nearest example of a red supergiant, Betelgeuse must be ordinary? This is because the chances are favourable for a draw of one from among many to be drawn from the most common.

While this may often be true, and hence unremarkable and thus unnoticed, it is not always the case, as is expected when considering many draws from samples that do include rare examples. (The basis for many false associations, let alone causal relationships.) For instance, the first supernova witnessed by naked eye for three centuries, and the best studied, supernova 1987A wasn't! It had been expected to

have been a star like Betelgeuse to have exploded, and yet it was all but. (Though they may have much more in common than at first sight.[1])

We are not placed in a typical red supergiant environment. While within a Bubble created by multiple supernovæ, we are not currently situated in a spiral arm. So perhaps Betelgeuse is typical for this particular environment, but not necessarily for red supergiants that are largely found in the denser, more actively star-forming regions of a spiral galaxy such as the spectacular H II regions in the prominent spiral arms several times more distant, or the even more distant massive clusters near the Galactic Centre.

But all things considered, Betelgeuse does not stand out in either luminosity, temperature, mass-loss rate or dust content – in fact it looks like a fairly commonly identified SN II-P progenitor![2] (But this may partly be a selection bias of the progenitor identification strategies.)

3. Is Betelgeuse atypical?

Just like most people look similar – four limbs, a head with a pair of eyes and ears and a nose upfront – when inspected more closely individual traits set us apart. This doesn't make each specimen a species, and in the end, we have much more in common than differentiates us, both in terms of the way we look and the way we behave.

Among Betelgeuse's perplexities count its location at some distance from sites of recent star formation, its large space motion, and a picture-perfect bow shock.[3] Upon close inspection, it spins at a baffling rate[4,5] and also is super-nitrogenous.[6]

And then came the 'Great Dimming'.[7] (Which may have nothing or all to do with the above.) A drop in brightness by half relegated Betelgeuse to the realm of much more inconspicuous stars, but sent alarm bells ringing with fears – or excitement – for its imminent luminous demise.

But none have been scrutinised as much as Betelgeuse. Brief late phases of underrepresented stars at birth, red supergiants themselves are rare, so few are seen up as close as Betelgeuse. It is not how well we know Betelgeuse, but how poorly we know the other red supergiants that defines its curiousness.

4. Our one-sided view of Betelgeuse

Because of the attention it has recently received, we shall first unpick how rare is Betelgeuse's Great Dimming, before we return to its full portfolio of properties.

By our very own nature as observers and thinking minds, and our position in the Universe, we suffer from anthropocentric bias: just because *we* didn't see it before, or elsewhere, doesn't make it rare per se. The 20$^{\text{th}}$-century artist Pablo Picasso exposed this imperfect picture by recovering what is hidden from view in a form referred to as "cubism". But how do we do that in reality? Do we know it? Do we infer it? Or do we ignore it? If you look out into a field and see what looks like a

black lamb, what do you conclude? That:

- lambs are black?
- lambs exist that are black?
- lambs exist of which at least one side is black?
- apparitions exist that look like lambs of which at least one side is black?

Despite well-known examples of non-isotropy in astronomy (cf. exoplanet transits, γ-ray bursts, pulsars) it is too often overlooked.

Betelgeuse forms dust clouds episodically, as convection cells cause dark spots[8] that induce cooler circumstellar conditions where dust can form,[9] likely in an already cooler parcel of lifted gas.[10] In the case of Betelgeuse this is exacerbated because Betelgeuse is on the cusp between a warm chromosphere and dusty wind.[11] Counting the number of dust clouds and considering their projected distances from Betelgeuse, a dust cloud forms about once every five years!

Imagine Betelgeuse were cubic and we only faced one side at a time (compare a standard die), then we'd only see one in every six events randomly happening on one of its sides. (In reality the polar and equatorial regions may display different behaviour especially in the case of a rapidly spinning Betelgeuse.) This means we *may* have had a Great Dimming *on our side* once every few decades (cf. the visual lightcurve[12]), but we *certainly* missed Great Dimmings happening *on another side*!

It is tempting to expect a dust cloud to form at every minimum in the $\sim$ 2000 day cycle. But that depends on whether this periodicity represents radial pulsation (in which case it is true) or convectional modulation – in which case the five-yearly cloud would always be in front, though other clouds would be forming more frequently throughout, above cool convection cells not seen by us, hence being in tension with observational evidence. It is more likely that a dust cloud forms at a pulsational minimum but only above a cool convection cell (the latter causing the $\sim$ 400-day cycle, but again that's our biased perspective, there could be a cool convection cell *somewhere* on the surface at all times). In fact, forming at some height above the surface, such dust cloud is less likely to be seen directly in front of the star than the convection cell is – at one stellar radius above the surface this chance is already diminished by a factor four. (Note that even the cloud purported to have caused the Great Dimming did not cover the entire face of Betelgeuse.[7]) The Great Dimming might have been our treat, but bread and butter for Betelgeuse.

Previous dimmings have been disputed, but it is not outrageous if of a few expected *chance* occurrences (as opposed to *predicted* events) just one has materialised. A century is not a very long time span in the life of a star, even that of a red supergiant (one pro mille!), and Betelgeuse may have exhibited many more dimmings, and may exhibit many more to come. Our Sun has arguably been more dramatic, as we know it so well: the sunspot cycle is a fairly regular rythm of activity, but we are still astounded by the Maunder Minimum of the late-17$^{\text{th}}$ and early-18$^{\text{th}}$ centuries – what if the telescope had only been invented a century later? (We would never have known.)

5. How common are anomalies?

If Betelgeuse ought to be common for being the nearest red supergiant, then so should be Antares for being only the second-nearest red supergiant. But as we argued, this is not a given. It would be much less likely, though, if both Betelgeuse *and* Antares were special. This then begs the question: is Betelgeuse different from Antares?

Antares has similar luminosity, temperature, convection,[13] mass loss, (low) dust content and lightcurve (with similar dimmings). Like Betelgeuse, it produces discrete dust clouds.[14,15] But Antares is a *slow* rotator and is accompanied by a hot star (on a sufficiently wide orbit not to be directly affected by it). How different does that make them?

Imagine stars are characterised by five features, that each have a 1:10 chance of being anomalous (imagine a ten-sided die). Then 1:2 stars are expected to be anomalous. Every other star. So a star is just as likely to be common as it is to be anomalous, despite the majority of stars to be common in any given feature. It should not come as a surprise if both Betelgeuse and Antares had some properties that are rare among the general population of red supergiants. This is what makes individual people distinctive.

Imagine another scenario, where a penguin among other penguins in a colony is distinguished by:

- the colour of its coat: blue (as opposed to brown).
- the texture of its coat: smooth (as opposed to fluffy).
- the colour of its cheeks: yellow.
- the colour of the underbill: orange.

Does that make this individual penguin anomalous in four ways? Most definitely not! It is anomalous in just one way: being an adult (surrounded by infants)!

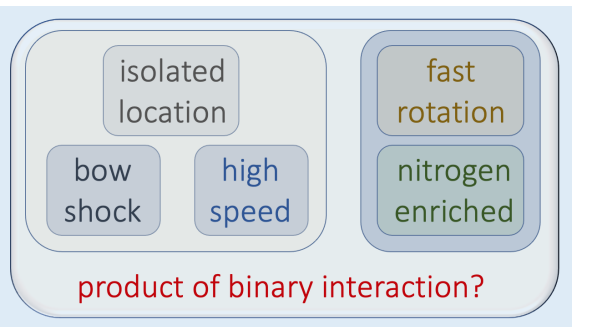


Fig. 1. Betelgeuse has various characteristics which have been treated as peculiarities. Irrespective of whether these are indeed oddities at all, they can be reduced to fewer characteristics – possibly just one – for which there may well be a totally reasonable – if not banal – explanation.

Returning to the curious case of Betelgeuse, its isolated location, high speed and bow shock are all related, and so may be its fast rotation and nitrogen enrichment – both can be reconciled into one: binary interaction (Fig. 1). And a common one at that: 10–40 per cent of massive stars will be affected by a companion.[16–18]

Then which is more special: Betelgeuse or Antares? Neither – both cases are expected equally: the result of binary interaction (Betelgeuse) and multiple but unaffected (Antares). The nearby Universe thus seems to be living up to expectations remarkably well!

6. How to find out how common is Betelgeuse?

Moving onwards, if we want to find out what red supergiants typically look like, how they generally evolve, and how and why some may deviate from the norm, we will not succeed in this by studying Betelgeuse harder and longer, but by statistical studies of samples of other red supergiants, such as:

- lightcurve studies based on sparse and/or time-limited data but for large samples in nearby galaxies ('Great Dimmings' may be seen in other galaxies, too[19]).
- nitrogen abundance measurements and their link with rotation.
- radial velocity monitoring and spectral energy distributions, to determine binary fractions.
- three-dimensional space motion studies in conjunction with rejuvenation scenarios, to determine past binary interactions.
- interferometric size measurements in relation to the location on the Hertzsprung–Russell Diagram, and mass (loss) measurements from seismology.

Fig. 2. What draws your eye? The white canvas that occupies most of what you're looking at? Or the black dot offset from the centre, which only covers a tiny fraction of your view? We are drawn to the exceptional, not the norm.

7. Epilogue

The take-away message from this discussion is that when making inferences we need to account for biases. We tend to look for the distinctive, not the common. For instance, when we look at a white panel with a black dot (Fig. 2) our attention is drawn to the dot even though it occupies much less space than the rest of the canvass. Just as we occupy ourselves with the few per cent of baryonic Universe, preferentially in its condensed forms, leaving the vastness of space largely neglected. If we saw less of the canvass, we might have missed the black dot altogether, but if we saw more we might find out black dots to be rather common!

Acknowledgements

I'd like to thank Costantino Sigismondi for organising the Betelgeuse session at the 16^{th} Marcel Grossmann meeting and for inviting me to speak.

References

1. J. Th. van Loon, Betelgeuse and the Red Supergiants, in Betelgeuse Workshop 2012, eds. P. Kervella, T. Le Bertre and G. Perrin, EAS Publications Series, Vol. 60, pp. 307–316 (2013).
2. S. J. Smartt, Observational Constraints on the Progenitors of Core-Collapse Supernovæ: The Case for Missing High-Mass Stars, *Publications of the Astronomical Society of Australia* **32**, 16 (2015).
3. L. Decin, N. L. J. Cox, P. Royer, *et al.*, The enigmatic nature of the circumstellar envelope and bow shock surrounding Betelgeuse as revealed by Herschel. I. Evidence of clumps, multiple arcs, and a linear bar-like structure, *Astronomy and Astrophysics* **548**, A113 (2012).
4. H. Uitenbroek, A. K. Dupree and R. L. Gilliland, Spatially Resolved Hubble Space Telescope Spectra of the Chromosphere of α Orionis, *The Astronomical Journal* **116**, 2501 (1998).
5. P. Kervella, L. Decin, A. M. S. Richards, G. M. Harper, I. McDonald, E. O'Gorman, M. Montargès, W. Homan and K. Ohnaka, The close circumstellar environment of Betelgeuse. V. Rotation velocity and molecular envelope properties from ALMA, *Astronomy and Astrophysics* **609**, A67 (2018).
6. D. L. Lambert, J. A. Brown, K. H. Hinkle and H. R. Johnson, Carbon, nitrogen and oxygen abundances in Betelgeuse, *The Astrophysical Journal* **284**, 223 (1984).
7. M. Montargès, E. Cannon, E. Lagadec, *et al.*, A dusty veil shading Betelgeuse during its Great Dimming, *Nature* **594**, 365 (2021).
8. X. Haubois, G. Perrin, S. Lacour, *et al.*, Imaging the spotty surface of Betelgeuse in the H band, *Astronomy and Astrophysics* **508**, 923 (2009).
9. P. Kervella, G. Perrin, A. Chiavassa, S. T. Ridgway, J. Cami, X. Haubois and T. Verhoelst, The close circumstellar environment of Betelgeuse. II. Diffraction-limited spectro-imaging from 7.76 to 19.50 μm with VLT/VISIR, *Astronomy and Astrophysics* **531**, A117 (2011).

10. P. Kervella, T. Verhoelst, S. T. Ridgway, G. Perrin, S. Lacour, J. Cami and X. Haubois, The close circumstellar environment of Betelgeuse. Adaptive optics spectro-imaging in the near-IR with VLT/NACO, *Astronomy and Astrophysics* **504**, A115 (2009).
11. J. Th. van Loon, M.-R. L. Cioni, A. A. Zijlstra and C. Loup, An empirical formula for the mass-loss rates of of dust-enshrouded red supergiants and oxygen-rich Asymptotic Giant Branch stars, *Astronomy and Astrophysics* **438**, 273 (2005).
12. M. Joyce, S.-C. Leung, M. László, M. Ireland, C. Kobayashi and K. Nomoto, Standing on the Shoulders of Giants: New Mass and Distance Estimates for Betelgeuse through Combined Evolutionary, Asteroseismic, and Hydrodynamic Simulations with MESA, *The Astrophysical Journal* **902**, 63 (2020).
13. K. Ohnaka, G. Weigelt and K.-H. Hofmann, Vigorous atmospheric motion in the red supergiant star Antares, *Nature* **548**, 310 (2017).
14. K. Ohnaka, Imaging the outward motion of clumpy dust clouds around the red supergiant Antares with VLT/VISIR, *Astronomy and Astrophysics 568*, A17 (2014).
15. E. Cannon, M. Montargès, A. de Koter, L. Decin, M. Min, E. Lagadec, P. Kervella, J. O. Sundqvist and H. Sana, The inner circumstellar dust of the red supergiant Antares as seen with VLT/SPHERE/ZIMPOL, *Monthly Notices of the Royal Astronomical Society* **502**, 369 (2021).
16. H. Sana, S. E. de Mink, A. de Koter, N. Langer, C. J. Evans, M. Gieles, E. Gosset, R. G. Izzard, J.-B. Le Bouquin and F. R. N. Schneider, Binary Interaction Dominates the Evolution of Massive Stars, *Science* **337**, 444 (2012).
17. S. E. de Mink, H. Sana, N. Langer, R. G. Izzard and F. R. N. Schneider, The Incidence of Stellar Mergers and Mass Gainers among Massive Stars, *The Astrophysical Journal* **782**, 7 (2014).
18. J. J. Eldridge, E. R. Stanway, L. Xiao, L. A. S. McClelland, G. Taylor, M. Ng, S. M. L. Greis and J. C. Bray, Binary Population and Spectral Synthesis Version 2.1: Construction, Observational Verification, and New Results, *Publications of the Astronomical Society of Australia* **34**, 58 (2017).
19. J. E. Jencson, D. J. Sand, J. E. Andrews, N. Smith, J. Pearson, J. Strader, S. Valenti, E. R. Beasor and B. Rothberg, An Exceptional Dimming Event for a Massive, Cool Supergiant in M51, *The Astrophysical Journal* **submitted** (arXiv:2110.11376).

Variable stars observed from city sites: The 2500 AAVSO-SGQ database

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

Paolo Ochner
Padova University and Asiago Astrophysical Observatory, Italy

The observations of Betelgeuse since Christmas 2011 were started either to study the star and to contribute to observational astronomy from urban contexts, with high level of light pollution, using the naked eye. Since then 811 observations to November 9, 2021 have been made and communicated to the AAVSO international database. Antares and also other interesting long-period variable stars like VY CMa, RS Oph (recurrent nova), V766 Cen and several novae and SN2014J are part of the 2.5 K database of visual observations, obtained with the Argelander method applied to naked eye or telescopic observations. It is an invitation for the young generations, to contribute to these studies with careful and feasible observations, which require a reduction for airmass extinction of the variable star and of the comparison stars, when they are not on the same narrow field of view. The possibility to deduce the airmass by reading the notes (containing the location of the observation) of all the AAVSO-SGQ catalogue is finally presented as a point of force of my observations.

Keywords: Mira, Betelgeuse; Eta Carinae; VY Canis Majoris; RS Ophiuchi, Antares; V766 Cen, Novae, SN2014J

1. Theory and observations in history of astronomy

I observe variable stars since 1997, I have started with Mira. The "candidacy" of Mira Ceti as Betlehem Star because it is close to the position of the Jupiter-Saturn triple conjunction of 6-7 b.C. was conceived by me at the Pontifical University of Lateran and the study was continued in Yale (2001-2004) with Dorrit Hoffleit. Maxima correlation function ruled out Mira for having two consecutive bright maxima, but this property was confirmed in the "oldest" Myra-type: R Leonis, R Hydrae and χ Cygni, from their almost four-century-long lightcurves. To a bright maximum normally a dim one follows, as the correlation function of their consecutive maxima shows. These studies were supported by historical and personal visual observations. Since 2011 I observe two first magnitude variable stars, with airmass correction to reach 0.01 magnitudes accuracy with naked eye.

AAVSO observer Sebastian Otero first claimed this accuracy, as for his first data on Nova Centauri 2013, observed also by me from Porto Alegre and Rio de Janeiro.

Betelgeuse's 811 observations in 10 years, include the deep minimum of 2020.

My SGQ contributions to AAVSO-database, paralleled with Betelgeuse's V-band measurements, help to define the "personal equations" present in all $25M$ visual observations before CCD era (1911-on) going back to 1893 for Betelgeuse.

2. Light curves completeness

The temporal extension of lightcurves is crucial to understand the stellar behaviour. With respect to Betelgeuse, more complicate cases are Antares and Aldebaran, less variable and with distant comparison stars. The Southern hypergiants VY CMa, low in the roman horizon, V766 Cen and Eta Carinae from South America in 1999, 2003 and 2013-2014 have been also monitored with binoculars or small telescopes. Delta Scorpii (Be binary star, whose system was at periastron in 2011) has also been studied from Rio de Janeiro, since 2011. Concluding: my 2.5K direct visual observations in nearly 25 years are data and experience useful to understand and simulate the accuracy of the World largest visual observations database for variable stars, the AAVSO one.

3. Betelgeuse among the long-period variable star

Betelgeuse is indeed the brightest variable star, but not the bigger one. Roughly the pulsational periods are proportional to the mass of the star through the relation that we found in the harmonic oscillators between the elastic force and the frequency and the mass: $K \times x = F = m\omega^2 \times x$.

Then $\omega = \sqrt{K/m}$. This physical assumption is very simple, and basic; it not fulfills all observational details nowadays available, but is a good starting point to interpret the difference of 9 years of observations of Betelgeuse with the ones of Antares.

The oscillation of Betelgeuse is evident and its main period is visible, while Antares shows either an oscillation and a rising trend.

Eta Carinae in 15 years (from 1999 to 2014) also showed a rising trend, while V766 Cen and VY CMa have been measured only for a short period, below one month, and other observers or future observations will allow to detect a trend.

4. Betelgeuse and Antares: Airmass correction and trends

From 2011 to 2021 Antares in the AAVSO database has been observed 619 times by 13 persons (only 3 with more than 10 observations). In the same time interval Betelgeuse has been observed nearly 14000 times by 360 observers. Not necessarily more observations are better, because of the "personal equations", or the systematic effect of each observer can experience for the Purkinje effect (physiological and slightly different from one individual to another), or the airmass extinction not applied when the comparison stars are far apart from the star.

The data of all the observers fill completely a full magnitude range for these reasons.

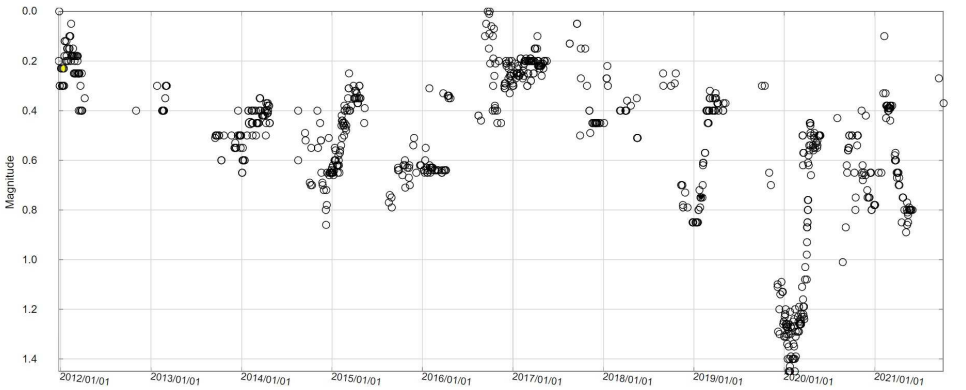


Fig. 1. The observations of Betelgeuse from 2011 to 2021 by Costantino Sigismondi, SGQ AAVSO observer.

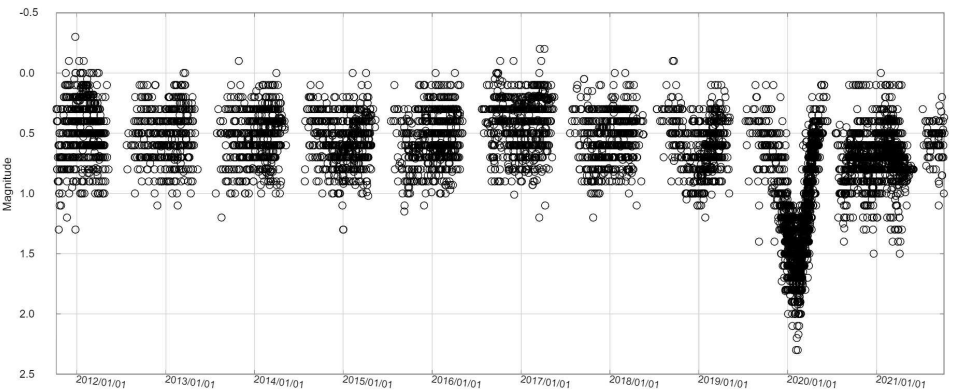


Fig. 2. All visual observations of Betelgeuse from 2011 to 2021 by the AAVSO observers: the oscillations are less evident, only the one during the deep minimum is clear. This was because Betelgeuse was at the meridian during this phase, around midnight, and its altitude was nearly constant night after night. The magnitude of the minimum from all data has a wide uncertainty. Only few visual data, still the majority for Betelgeuse, are corrected for the airmass: this is the cause of such uncertainties. The yearly gap due to solar conjunction is the most clear feature.

In the case of Betelgeuse I suppose that the airmass extinction is not applied by the majority of the observers, so I select only the obervers that I know personally as applying that correction.

Antares is a more Southern star (less observers in the Southern hemisphere), and even more separated by the comparison stars, then its measurements are more complicate. The amplitude of its variability is smaller than Betelgeuse's. These difficulties may explain the rather low number of observations available: 4% with respect to Betelgeuse. The oscillation I have detected are not clearly distinct from a yearly period, which seems related to the altitude above the horizon of the star

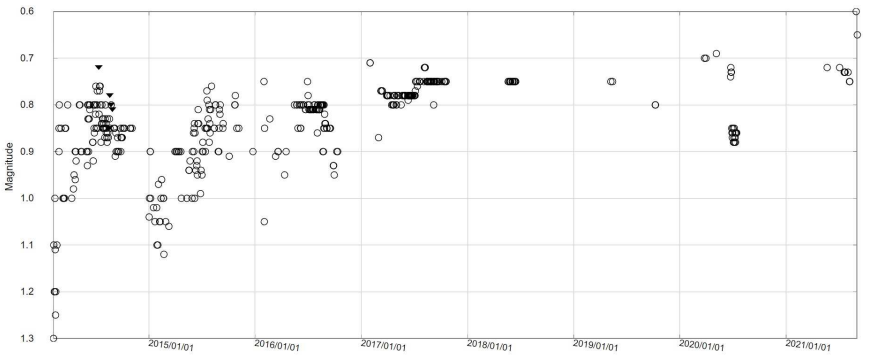


Fig. 3. The observations of Antares from 2014 to 2021 by Costantino Sigismondi, SGQ AAVSO observer.

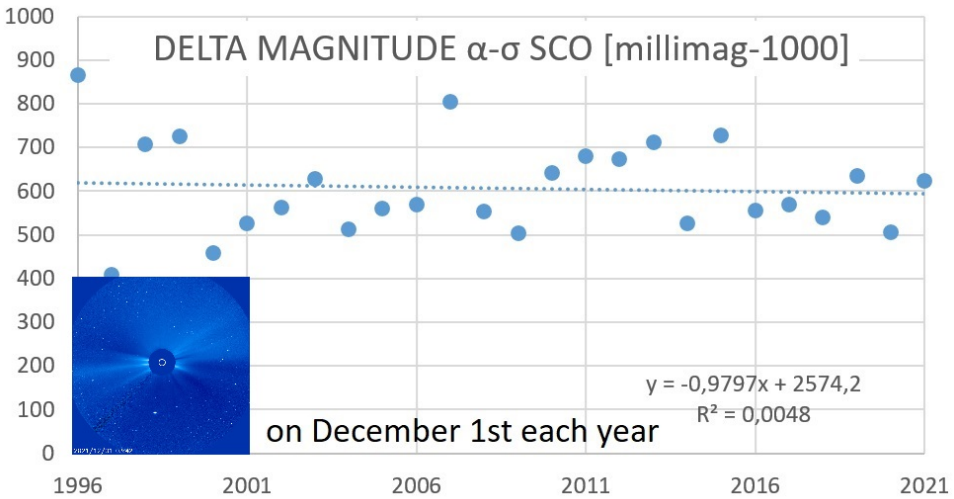


Fig. 4. The observations of Antares from 1996 to 2021 obtained by LASCO C3 Coronograph with the differential photometry made with σ Scorpii. The icon shows the image of Dec 1st, 2021.

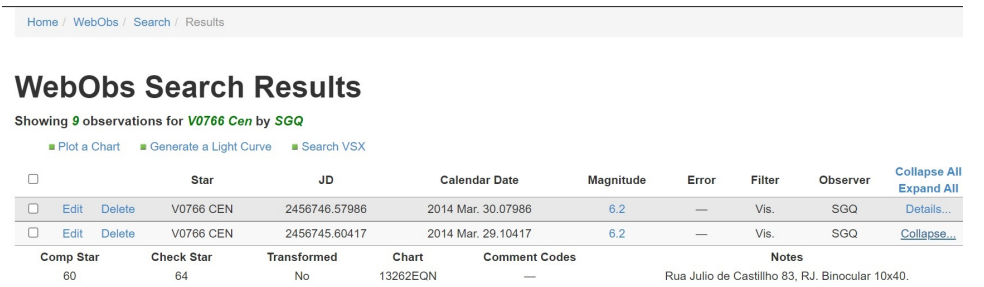
observed at a given evening hour, and not to the star itself. The trend of the increase of the luminosity is more evident from these data.

As Betelgeuse, Antares presents rather rapid changes of luminosity during its period of visibility. No trend over the 2.5 decades can be guessed by either the visual data and the sampling each 1st of December, made with the SOHO LASCO C3 Coronograph, which includes for some days a year Antares in its field of view. That Coronograph could show stars of 5th magnitude under low activity of the solar corona. Antares is nearly saturated with this instrument, because it shows a lateral photons' spilling.

5. Recovering airmass on the AAVSO-SGQ observations' catalogue

The airmass of all variable stars can be calculated on AAVSO-SGQ catalogue by selecting SGQ as observer, in searching Alf Ori observations (and all the other variable stars observed).

The location of the observation is always reported in the notes, and the time is required by the application form. So the data for recomputing the airmass are made available.


Home / WebObs / Search / Results

WebObs Search Results

Showing *9* observations for *V0766 Cen* by *SGQ*

Plot a Chart | Generate a Light Curve | Search VSX

			Star	JD	Calendar Date	Magnitude	Error	Filter	Observer	Collapse All / Expand All
☐	Edit	Delete	V0766 CEN	2456746.57986	2014 Mar. 30.07986	6.2	—	Vis.	SGQ	Details...
☐	Edit	Delete	V0766 CEN	2456745.60417	2014 Mar. 29.10417	6.2	—	Vis.	SGQ	Collapse...

Comp Star	Check Star	Transformed	Chart	Comment Codes	Notes
60	64	No	13262EQN	—	Rua Julio de Castillho 83, RJ. Binocular 10x40.

Fig. 5. Example on V766 Centauri of the information contained in the notes: RJ is the city, Rio de Janeiro, and the address allows to have from google maps the geographical coordinates of the observation. From the time the altitude h° in the sky of the star is obtained, yielding the airmass with the cosecant law $1/\sin(h°)$

6. Conclusions and Perspectives: The quality of SGQ observations

The luminosity of Betelgeuse allowed to study that star in each condition from urban sites, namely for my observations Paris, Rome, Rio de Janeiro, Warsaw, without very dark skies neither without the possibility to be well screened from the light directed toward my eyes. The oscillations at the main period of 1.2 years are clearly detected.

More complicate has been the study of **Antares** which is farther from its comparison stars than Betelgeuse, and which is closer to the horizon for all the European cities. The oscillation of Antares is ranging from 1.16 to 0.88 according to the General Catalgue of Variable Stars 5.1, but my observations of the Summer 2021 found it around the magnitude 0.70 ± 0.05, at its maximum since 7 years. For Antares the comparison with Saturn was useful in the last years, and the planet has also a "light-curve" due to the geometrical change of the rings' plane. The small oscillations of Antares are not clearly detected for the aforementioned difficulties: at the horizon there are more obstacles and the atmospheric extinction is not linear with the cosecant of the altitude $1/sin(h°)$. In the General Catalogue of Variable Stars 5.1 the is no value for the oscillating period of Antares. It is classified as LC variable, and according to wikipedia it means: *A slow irregular variable (ascribed the GCVS types L, LB and LC) is a variable star that exhibit no or very poorly defined*

periodicity in their slowly changing light emissions. These stars have often been little-studied, and once more is learnt about them, they are reclassified into other categories such as semiregular variables. A long period study on the variability of Antares has conducted by me following a 2014 study on Delta Scorpii, with the SOHO C3 coronograph. The next datum from SOHO C3 will be available 15 days after the closure of these proceedings, on December 1st 2021. This measure will be interesting since in the summer 2021 Antares has shown the brightest luminosity of the last 7 years. The two datasets of **Eta Carinae** in 1999 and in 2013-4 allow to see its slow rising. The other hypergiant **V766 Cen** and **VY CMa** have been observed for a short time slot, and will be compared with future observations. For them the differential magnitude has been estimated with stars in the same or near the telescopic field of view, being them not visible to the naked eye. Actually Eta Carinae could be visible with the naked eye, but not in the city lights. To observe it from Rio de Janeiro (Copacabana seaside with a strong illumination in 2003 and Observatorio Nacional in 2013) I had to use a small telescope. The recurrent nova **RS Oph** (2021, 24 observations along the month 8 August-12 September down to magnitude 9 with a 76 mm telescope, fig. 5), other Novae in Sagittarius from 2015 to 2018 and the SN 2014J are part of this personal database of 2564 high quality visual observations to date (9 November 2021), available to the science community.

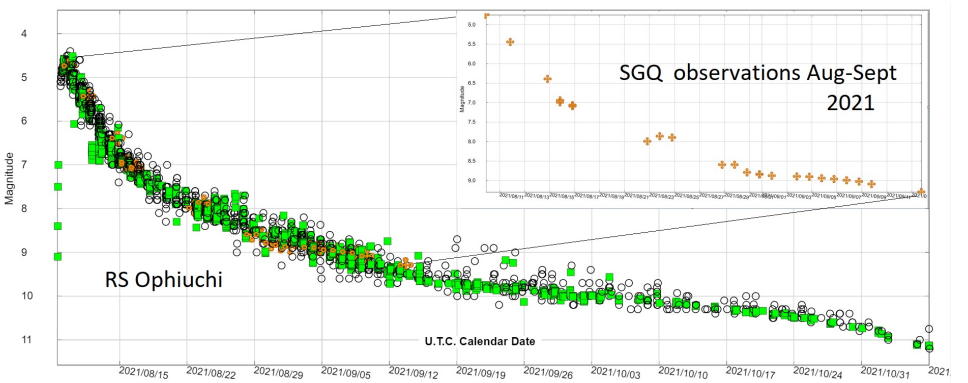


Fig. 6. The AAVSO light curve of RS Oph with the SGQ contributions evidenced in orange. They are compared with the V-band observations, and they are in very good agreement.

References

1. Antares in the General Catalogue of Variable Stars (2021), `http://www.sai.msu.su/gcvs/cgi-bin/search.cgi?search=Alf+Sco`.
2. C. Sigismondi, AAVSO-SGQ database (insert SGQ in the observers' code) `https://app.aavso.org/webobs/search/`.
3. Curt Renz, Saturn Stellar Magnitude `https://www.curtrenz.com/saturn.html`.
4. C. Sigismondi, et al. (2014) *Photometry of Delta Scorpii from 1996 to 2013 using SOHO LASCO C3 coronograph* - `https://arxiv.org/abs/1410.8492`.

Betelgeuse: An introductory course to observational astronomy

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

Tiziana Pompa
Liceo Galilei, Pescara, Italy

Betelgeuse is the brightest variable star in the sky, with a pulsational main period of 14 months and typical magnitude variation of 0.5 magnitudes. In 2020 its deep dimming was of one whole magnitude, particularly evident to each attentive observer. The technique of airmass correction to detect the stellar signal from the atmospheric extinction treats the naked eye observations as the digital astronomical observations. Harlow Shapley said in 1930s "theories come and go, while a good observation stands forever" and he spoke about naked eye observations, nowadays they are less considered in science mainly because of philosophical reasons concerning the subjectivity of knowledge and the deceit of the senses, while the knowledge of the eye physiology and of data reduction could recover the correct relationship between the human being and the Universe.

Keywords: Betelgeuse; Red giant; Variable stars, Atmospheric Extinction, Quadrant.

1. General Overview on observing Betelgeuse

According to Dorrit Hoffleit (1907-2007) director of the Maria Mitchell Observatory and astronomer at Yale University, Mira is the educational star. But it requires the use of a telescope at its minimum, and a binocular for its maximum phase, to spot it in city lights. Betelgeuse is a semiregular variable star of first magnitude, in the most famous constellation in the sky, visible from both hemispheres because it is on the celestial equator. It is visible (to the naked eye) for nine months a year. Its magnitude estimate, according to the Argelander method, requires other first magnitude stars, separated by several degrees, at rather different airmasses.

Procyon, Aldebaran, Pollux, Castor as well as γ Geminorum, Regulus, Denebola, Mars and even Mercury have been used as comparison stars, and their airmass corrections included.

Since January 2019 we steered a project in schools to observe and measure the light curve of Betelgeuse. We observed Betelgeuse and its comparison stars with a special colored quadrant where the angular height in sight corresponds to the airmass exctinction with respect to the zenit. The schools were the Lyceum Morgagni and Ferraris in Rome and Ostia and the Lyceum Galilei in Pescara.

☐	Edit	Delete	ALF ORI	2458869.58333	2020 Jan. 21.08333	1.40	—	Vis.	SGQ	Details...
☐	Edit	Delete	alf Ori	2458869.30556	2020 Jan. 20.80556	1.45	—	Vis.	SGQ	Details...
☐	Edit	Delete	alf Ori	2458866.35764	2020 Jan. 17.85764	1.29	—	Vis.	SGQ	Collapse...
Comp Star	**Check Star**	**Transformed**	**Chart**	**Comment Codes**	**Notes**					
Bet Gem mv=1.16	Alf Gem mv=1.58	No	Tirion2000.0/Simbad	—	Pescara, front sea 7th floor, Betelgeuse dimming meeting. Naked eye with airmass correction.					
☐	Edit	Delete	ALF ORI	2458866.27431	2020 Jan. 17.77431	1.35	—	Vis.	SGQ	Collapse...
Comp Star	**Check Star**	**Transformed**	**Chart**	**Comment Codes**	**Notes**					
et Gem mv=1.16	Alf Gem mv=1.58	No	Tirion2000.0/Simbad	UL	Pescara, ICRANet, Betelgeuse dimming meeting. Naked eye with airmass correction.					
☐	Edit	Delete	alf Ori	2458866.22014	2020 Jan. 17.72014	1.45	—	Vis.	SGQ	Collapse...
Comp Star	**Check Star**	**Transformed**	**Chart**	**Comment Codes**	**Notes**					
et Gem mv=1.16	Alf Gem mv=1.58	No	Tirion2000.0/Simbad	UL	Pescara, ICRANet, Betelgeuse dimming meeting. Naked eye with airmass correction.					
☐	Edit	Delete	ALF ORI	2458865.35764	2020 Jan. 16.85764	1.30	—	Vis.	SGQ	Details...
☐	Edit	Delete	ALF ORI	2458864.42500	2020 Jan. 15.92500	1.26	—	Vis.	SGQ	Details...
☐	Edit	Delete	ALF ORI	2458863.39583	2020 Jan. 14.89583	1.28	—	Vis.	SGQ	Details...
☐	Edit	Delete	ALF ORI	2458862.26042	2020 Jan. 13.76042	1.40	—	Vis.	SGQ	Details...

Fig. 1. The observations made during the meeting of 2020 during the Great Dimming.

The great dimming of Betelgeuse was then observed and measured to the nearest 0.01 magnitude at naked eye (fig. 1).

These observations have been sent to AAVSO, as contributes of students to citizen science. Similar operations have been realized with the students of the Astrophysics Laboratory, Sapienza University of Rome. Summer vactions' homework are "offered" by Antares, a first magnitude stars greatly separated by its more suitable comparison stars. Recently Saturn worked as reference, but a planet is not a constant standard candle. The introduction to photometry, colorimetry and spectrometry with these naked-eye cases is complete and historically meaningful.

2. Colorimetry: From the spectrum to the Newton's disk

The color of the stars is something lost in city environment, because lost is the visibility of the sky in general, owing to the artificial light directed toward our eyes and toward the sky. Betelgeuse is clearly a red star, and it can be noticed by comparing it with Procyon or Rigel, normally used as comparison stars for magnitude's estimates, or Sirius, that are rather close to Betelgeuse.

The Newton's disk for solar spectrum has been replicated for Betelgeuse with red-yellowish tonalities.

The connection between the color of the star and the its temperature can be done by noticing the percentage of color of a given wavelength, as represented in the Newton's disk. Basically the Wien's law is represented with the apple pie diagram: the largest sector corresponds to the wavelength of the maximum intensity in the spectrum.

3. Betelgeuse the brightest long-period variable star

Among the variable stars Betelgeuse is indeed the brightest, and the most simple one to be detected thanks to its location in Orion, near the celestial equator. This

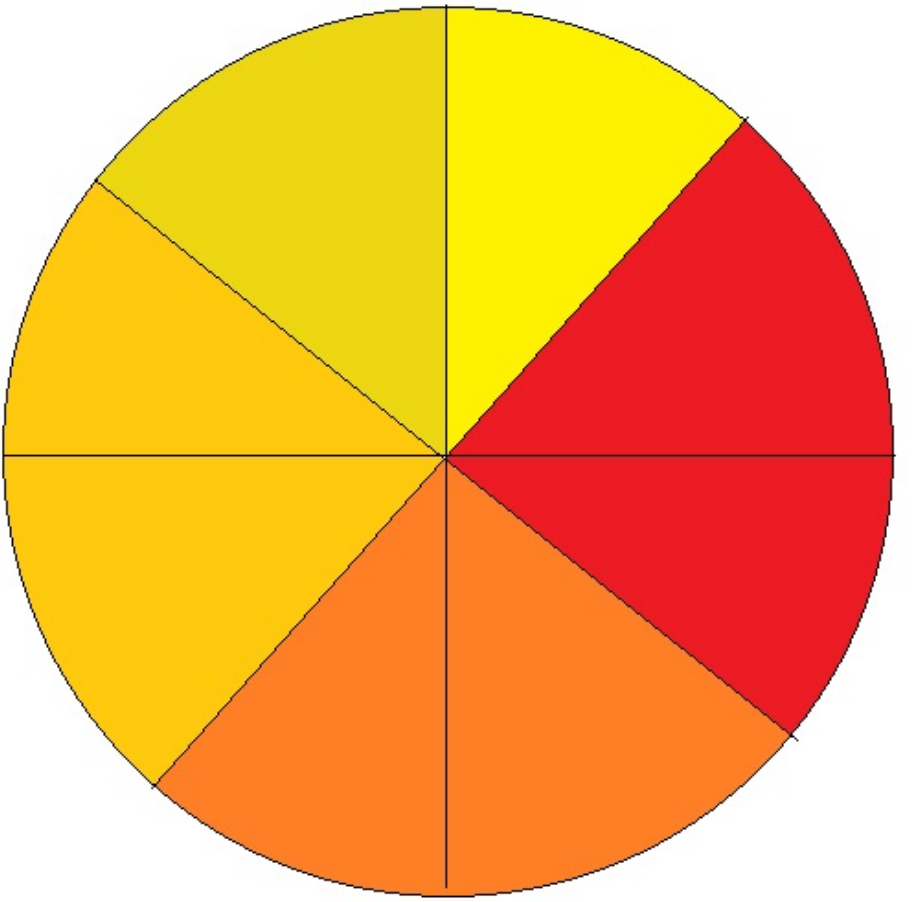


Fig. 2. Colorimetry: the Newton's disk of Betelgeuse as realized by 14 years old students in 2020.

position makes Betelgeuse popular either from the Northern and the Southern hemispheres. The period of main oscillation is 1 year and two months, and the secondary modulating period is almost six years. In a reasonable space of time achievable by a school class program, the star can be followed from a maximum through the following minimum to the next one maximum, i.e. a complete oscillation. This allows the students to familiarize with the oscillating nature of such phenomenon, and primarily with the fact that among the "fixed" stars something changes and within a factor of two in intensity (2.512 times for a whole magnitude gap from minimum to maximum) even if the star is a supergiant as Betelgeuse is.

4. The quadrant for the airmass correction

Hanging a weight on the top right corner of the quadrant in figure 3, the thread shows the magnitude's correction and the corresponding angular altitude of the

star. The brighter stars used as comparison stars with the Argelander's method are reported near the bottom left corner: Aldebaran, Procyon, Rigel, Spica, Regulus and Deneb. There are not Pollux and Castor of Gemini, neither Bellatrix in Orion, stars dimmer than 1.2 magnitude, which in January 2019, when the quadrant was designed, were out of the last four-decades range of variation of Betelgeuse.

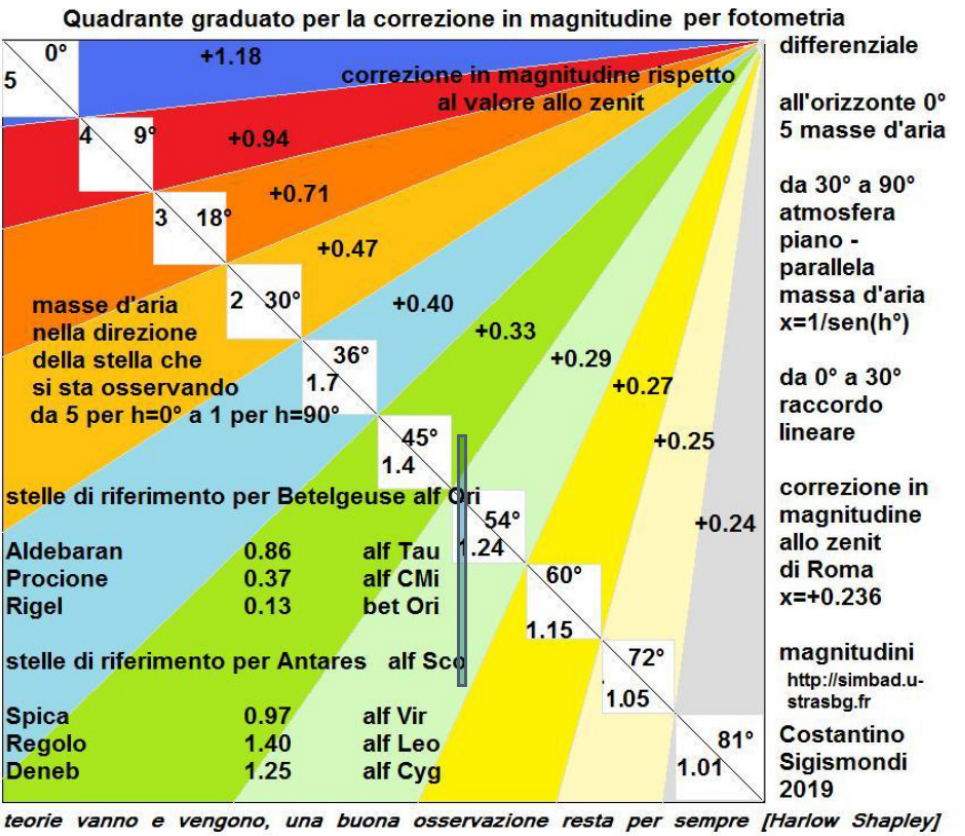


Fig. 3. The quadrant prepared for the observations of Betelgeuse for the students of the Morgagni Lyceum in Rome and the Galilei Lyceum in Pescara.

5. Data reduction

It is not obvious for a young student that an astronomical observation, or an experiment in general, require a data treatment, to be presented or to be analyzed to obtain science results. In particular an observation with naked eyes seems something really far from the data analysis concept. While a light-curve shows the intrinsic "behaviour" of a star, and it is exactly what we want to observe and measure, a single observation made with the naked eye seems a static datum. The difference in altitude of the stars determines a different atmospheric extinction acting on them,

Δmag da aggiungere	Altezza della stella di riferimento di uguale magnitudine										Atmosfera piano parallela con 0.23 magnitudini per massa d'aria valido al livello del mare														
Altezza stella variabile	15	16	17	18	19	20	21	22	23	24	25	30	35	40	45	50	55	60	65	70	75	80	85	90	
15	0,00	-0,05	-0,10	-0,14	-0,18	-0,22	-0,25	-0,27	-0,30	-0,32	-0,34	-0,43	-0,49	-0,53	-0,56	-0,59	-0,61	-0,62	-0,63	-0,64	-0,65	-0,66	-0,66	-0,66	
16	0,05	0,00	-0,05	-0,09	-0,13	-0,16	-0,19	-0,22	-0,25	-0,27	-0,29	-0,37	-0,43	-0,48	-0,51	-0,53	-0,55	-0,57	-0,58	-0,59	-0,60	-0,60	-0,60	-0,60	
17	0,10	0,05	0,00	-0,04	-0,08	-0,11	-0,14	-0,17	-0,20	-0,22	-0,24	-0,33	-0,39	-0,43	-0,46	-0,49	-0,51	-0,52	-0,53	-0,54	-0,55	-0,55	-0,56	-0,56	
18	0,14	0,09	0,04	0,00	-0,04	-0,07	-0,10	-0,13	-0,16	-0,18	-0,20	-0,28	-0,34	-0,39	-0,42	-0,44	-0,46	-0,48	-0,49	-0,50	-0,51	-0,51	-0,51	-0,51	
19	0,18	0,13	0,08	0,04	0,00	-0,03	-0,06	-0,09	-0,12	-0,14	-0,16	-0,25	-0,31	-0,35	-0,38	-0,41	-0,43	-0,44	-0,45	-0,46	-0,47	-0,47	-0,48	-0,48	
20	0,22	0,16	0,11	0,07	0,03	0,00	-0,03	-0,06	-0,08	-0,11	-0,13	-0,21	-0,27	-0,31	-0,35	-0,37	-0,39	-0,41	-0,42	-0,43	-0,43	-0,44	-0,44	-0,44	
21	0,25	0,19	0,14	0,10	0,06	0,03	0,00	-0,03	-0,05	-0,08	-0,10	-0,18	-0,24	-0,28	-0,32	-0,34	-0,36	-0,38	-0,39	-0,40	-0,40	-0,41	-0,41	-0,41	
22	0,27	0,22	0,17	0,13	0,09	0,06	0,03	0,00	-0,03	-0,05	-0,07	-0,15	-0,21	-0,26	-0,29	-0,31	-0,33	-0,35	-0,36	-0,37	-0,38	-0,38	-0,38	-0,38	
23	0,30	0,25	0,20	0,16	0,12	0,08	0,05	0,03	0,00	-0,02	-0,04	-0,13	-0,19	-0,23	-0,26	-0,29	-0,31	-0,32	-0,33	-0,34	-0,35	-0,36	-0,36	-0,36	
24	0,32	0,27	0,22	0,18	0,14	0,11	0,08	0,05	0,02	0,00	-0,02	-0,11	-0,16	-0,21	-0,24	-0,27	-0,28	-0,30	-0,31	-0,32	-0,33	-0,33	-0,33	-0,34	
25	0,34	0,29	0,24	0,20	0,16	0,13	0,10	0,07	0,04	0,02	0,00	-0,08	-0,14	-0,19	-0,22	-0,24	-0,26	-0,28	-0,29	-0,30	-0,31	-0,31	-0,31	-0,31	
30	0,43	0,37	0,33	0,28	0,25	0,21	0,18	0,15	0,13	0,11	0,08	0,00	-0,06	-0,10	-0,13	-0,16	-0,18	-0,19	-0,21	-0,22	-0,22	-0,23	-0,23	-0,23	
35	0,49	0,43	0,39	0,34	0,31	0,27	0,24	0,21	0,19	0,16	0,14	0,06	0,00	-0,04	-0,08	-0,10	-0,12	-0,14	-0,15	-0,16	-0,16	-0,17	-0,17	-0,17	
40	0,53	0,48	0,43	0,39	0,35	0,31	0,28	0,26	0,23	0,21	0,19	0,10	0,04	0,00	-0,03	-0,06	-0,08	-0,09	-0,10	-0,11	-0,12	-0,12	-0,13	-0,13	
45	0,56	0,51	0,46	0,42	0,38	0,35	0,32	0,29	0,26	0,24	0,22	0,13	0,08	0,03	0,00	-0,03	-0,04	-0,06	-0,07	-0,08	-0,09	-0,09	-0,09	-0,10	
50	0,59	0,53	0,49	0,44	0,41	0,37	0,34	0,31	0,29	0,27	0,24	0,16	0,10	0,06	0,03	0,00	-0,02	-0,03	-0,05	-0,06	-0,06	-0,07	-0,07	-0,07	
55	0,61	0,55	0,51	0,46	0,43	0,39	0,36	0,33	0,31	0,28	0,26	0,18	0,12	0,08	0,04	0,02	0,00	-0,02	-0,03	-0,04	-0,04	-0,05	-0,05	-0,05	
60	0,62	0,57	0,52	0,48	0,44	0,41	0,38	0,35	0,32	0,30	0,28	0,19	0,14	0,09	0,06	0,03	0,02	0,00	-0,01	-0,02	-0,03	-0,03	-0,03	-0,04	
65	0,63	0,58	0,53	0,49	0,45	0,42	0,39	0,36	0,33	0,31	0,29	0,21	0,15	0,10	0,07	0,05	0,03	0,01	0,00	-0,01	-0,02	-0,02	-0,02	-0,02	
70	0,64	0,59	0,54	0,50	0,46	0,43	0,40	0,37	0,34	0,32	0,30	0,22	0,16	0,11	0,08	0,06	0,04	0,02	0,01	0,00	-0,01	-0,01	-0,01	-0,01	
75	0,65	0,60	0,55	0,51	0,47	0,43	0,40	0,38	0,35	0,33	0,31	0,22	0,16	0,12	0,09	0,06	0,04	0,03	0,02	0,01	0,00	0,00	-0,01	-0,01	
80	0,66	0,60	0,55	0,51	0,47	0,44	0,41	0,38	0,36	0,33	0,31	0,23	0,17	0,12	0,09	0,07	0,05	0,03	0,02	0,01	0,00	0,00	0,00	0,00	
85	0,66	0,60	0,56	0,51	0,48	0,44	0,41	0,38	0,36	0,33	0,31	0,23	0,17	0,13	0,09	0,07	0,05	0,03	0,02	0,01	0,01	0,00	0,00	0,00	
90	0,66	0,60	0,56	0,51	0,48	0,44	0,41	0,38	0,36	0,34	0,31	0,23	0,17	0,13	0,10	0,07	0,05	0,04	0,02	0,01	0,01	0,00	0,00	0,00	

Fig. 4. The airmass table for Rome with $\Delta Mag = 0.236/sin(h)$.

Fig. 5. The quadrants realized for the special event of the Betelgeuse Deep Minimum Meeting of January, 17th 2020.

and this computation has to be done for each of the observed stars: Betelgeuse and the two bracketing comparison stars (Aldebaran and Procyon, usually). This operation generate a system of two first order equations in the incognita *magnitude* of the variable star. The solution of such system is a range of magnitudes for the variable star, hopefully as narrow as 0.05 or 0.02 magnitudes. The data reduction can be made also by using Stellarium 0.20.2, ad higher versions, where an extinction's model is implemented. With respect to Betelgeuse Stellarium airmass fails below 10° and our quadrant is more reliable.

6. Conclusions and Perspectives

Betelgeuse is a good "student star" or the "educational star" following Dorrit Hoffleit. She was applying such definition for Mira Ceti (from 2nd magnitude to 9th) but the students were College students with an astronomical telescope available. Betelgeuse requires attention indeed, but it is visible always to the naked eye. It is visible also from the light-polluted cities, and the observations can be done from their home, as homework, rather easily from September-November before sunrise to April-May after sunset, and nearly all night long in between. The possibility to participate to AAVSO observational campaigns, with a real citizen science contribution can stimulate the most interested students to pursue the way of a scientific career in astronomy.

References

1. Betelgeuse dimming: the state of the star International workshop (2020), http://icranet.org/index.php?option=com_content&task=view&id=1281.
2. http://www.aavso.org.
3. C. Sigismondi, Quadrante per Magnitudini (2019) https://docs.google.com/document/d/1UDJIOmdYZnBEAK4pDzdCWjEoGVBqjjbp1gL9HvrmbiY/edit?usp=sharing.
4. http://www.stellarium.org.

The meridian line of the Vatican obelisk to study the stellar aberration

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
www.icra.it/solar

Lorenzo Ricciardi
University of Rome 3,
Rome, Italy

The first evidence of the stellar aberration come out from telescopic meridian observations, including daytime ones. The distance in ecliptic longitude between the star and the Sun has to change from 180° to 90° to have the maximum effect in the delay of the transit, with respect to a constant sidereal day of 23h 56m 04.09s. We present the measurements of meridian transits of Antares and Ras Alhague during the Summer 2021, at the meridian line of St. Peter's square, whose gnomon is the Vatican obelisk. The possibility to observe such transits in daylight has been also verified with Antares (21 August, 2021) and Venus (October 2021). The geometry of the instrument and its configuration is ideal to introduce students to the observations of this special relativistic effect, from an experimental point of view. The transits observed from the same position, the Scorpio zodiacal disk, from October 16 to November 11, 2021 permitted to scan the axis of the obelisk and the one of the uppermost cross, finding an offaxis of (16±3) mm West for the upper pyramid and a further deviation of (5 ± 3) mm West of the cross.

Keywords: Vatican obelisk; Meridian line; Stellar Aberration; Ecliptic longitude; Sidereal day.

1. The Vatican obelisk and the meridian line of St. Peter's square

The Egyptian obelisk was carried in Rome by Emperor Caligola and set on the Vatican Circus. It was the only one which survived along the middle ages, considered the witness of the martyrdom of the apostle St. Peter. Pope Sixtus V and the architect Domenico Fontana moved it in front of the Basilica of St. Peter in 1586 with the special design to emulate and surpass the Augustus' obelisk mentioned by Plinius the Elder in his Naturalis Historia. The Dominican astronomer Egnazio Danti who also worked at the Gregorian Reformation of the Calendar (1582) drafted the project of a meridian line, which would have been the largest in the World, his untimely death on 22nd October 1586, 40 days after the translation of the obelisk, delayed this work of 230 years. Only in 1817 Cardinal Peter Maccarani, with the astronomer rev. Filippo Luigi Gigli, realized the line in granite, 7cm wide and 70 meter's long from Cancer to Capricorn. The characteristics of such line are measured and reported in this paper, they have been used as zero calibration of the instruments in the measurements of the Earth's rotation period.

2. The stellar aberration

Published in 1727 as a discovery of a new motion of the fixed stars, by the Royal Astronomer James Bradley, the stellar aberration moves the stars around an ellipse. The ecliptic longitude of the Sun determines the longitude's aberration amount for a given star: it is maximum ± 20.62 arcsec if the star is on the ecliptic plane, and with the Sun in conjunction or opposition, it is zero in both quadratures. Out of the ecliptic plane a latitude component appears, which equals the longitudinal one at the ecliptic pole, where γ Draconis, the star observed by James Bradley, it is located. We observed Antares between 130° to 100° from the Sun and Ras Alhague in evening, twilight and sunset conditions; the aberration is proportional to the difference of the cosine of such angles or 47% of the 20.62 arcsec, i.e. 9 arcsec. A difference in 9 arcsec is corresponding to 0.6 s in transit time, with respect to a constant sidereal day. In 30 days, as the duration of this series of observations, it corresponds to a drift in the duration of the sidereal day of +0.023 s/day.

The aberration value of 9 arcsec corresponds to twice the parallax maximum estimated error of 4.5 arcsec (corresponding to 2 mm on the meridian line at 92 m from the gnomon on the Capricorn's disk).

2.1. *Reducing the parallax error*

A mirror located in the center of the meridian line, always in the same point, is aimed with the monocular at its center. The top of the obelisk is at 92 meters of distance, and an error of pointing of ±2 mm is possible, this yield a maximum angular uncertainty of ±4.5 arcsec.

3. The observations at the meridian line of obelisk's occultations: A differential measure

The inclination coefficient 235.872 s = 3 m 55.872 s is the difference with 24 hours of the measured average sidereal day. The event measured is the disappearance of Antares behind the obelisk, called *first contact* behind the obelisk.

The figure 1 represents the linear fit of the UTC transit times observed with a monocular as ingress and egress time behind the obelisk, as seen from the same point of the end of the meridian line next to the Capricorn sign. The average sidereal day is 23h 56m 04.128s ±0.02, which is +0.038 ±0.02 s larger than the standard value for it. The special relativistic effect +0.0207 s is within a sigma from our result.

These measurements are made all from the same place of the meridian line, near the black star of the Capricorn's disk for Antares and Venus (in afternoon meridian transit during its Eastern solar elongation) and near the black star of Virgo's disk for Ras Alhague, so they are differential. Systematic deviations along the meridian line do not affect them.

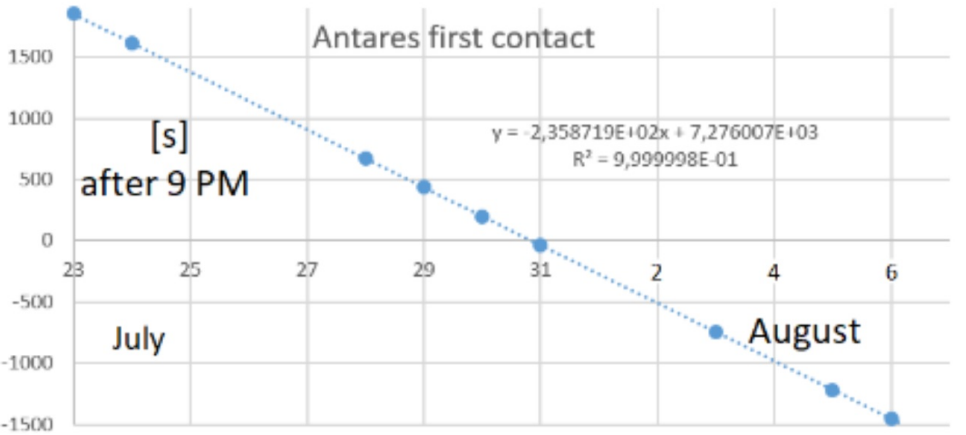


Fig. 1. The average sidereal day from 23 July (Sun $\lambda = 121°$) to 21 August 2021 (Sun $\lambda = 149°$). Antares has $\lambda = 250°$. In this timespan the ecliptic elongation of Antares from the Sun ranges from $\lambda = 129°$ to $\lambda = 101°$. The difference of the cosines is 0.2, multiplied by 20.6 arcsec, the aberration constant, yields 9 arcsec, which in time is 0.6 s. In the 29 days of this plot the theoretical value is is 0.0207 s/day, and the measured one is 0.038±0.02 s/day.

4. Local deviations from the North of the meridian line

A general astrometric recognition of the meridian line has been done on the zodiacal disks, obtaining the following deviations with respect to the axis of the obelisk.

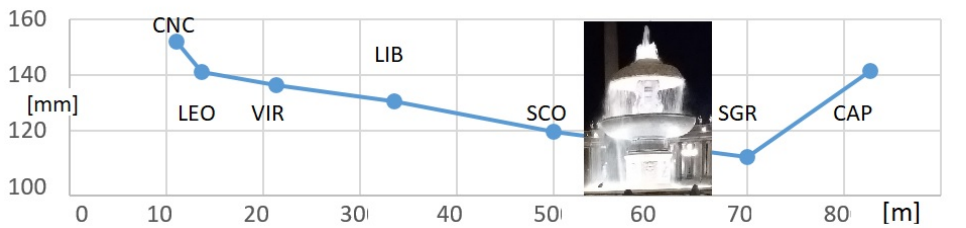


Fig. 2. Deviations in mm from the obelisk's axis along the line.

The cross on top of the obelisk is not aligned with its axis, being slightly west. The cross has been probably used for aligning the Capricorn with the Summer signs. They are at the same westward distance from the obelisk axis, being aligned with the celestial North within 2 mm along 71 meters, which corresponds to 6 arcseconds: the top of the technology of 1817.

5. Orthogonality of the cross over the obelisk

After calibrating the meridian line with respect to the North, it is possible to calibrate the gnomon (the obelisk with the cross) with respect to the vertical of its axis.

Fig. 3. The Southern face of the obelisk in full Sun: on the top the dark Southern arm if the cross is slightly on the left from the obelisk's axis.

The technique consists in measuring the difference between the observed transit and the calculated one, and transforming it into linear distances with the local velocity of the shadow. A series of observations made from October 16 to November 11, 2021 has been made to verify the orthogonality of the cross above the obelisk, and its symmetry axis. The Sun, as seen from the star of the Scorpio's marble zodiacal disk, started to pass behind the top of the cross, and it scanned all the features (cross, star, mounts, pyramid and the uppermost sections of the monolith). The instant in which the Sun passed through their symmetry axis was compared with the transit time calculated with the ephemerides. After many verifications in St. Maria degli Angeli meridian line to the nearest arcsecond, the ephemerides of Stellarium 0.20.2 have been used for St. Peter's square meridian line. The final result of this study has an uncertainty of $\pm$ 3 mm, due to the location of the telescope or of the camera, in case of direct vision, on the marble disk.

With respect to the symmetry axis of the obelisk:

- the axis of the obelisk from the point is defined (10 and 11 November)
- the star is 16.5 mm West of the axis (20 October)
- the middle part of the cross is further 10 mm West (18 October)
- the uppermost part of the cross is further 5 mm West (16 October)

Being the errorbar $\pm$ 3mm the cross can be vertical as well as slightly inclined. The four tyrants have the function of stabilization during windy days. On 14 october, a windy day, the oscillations of the cross have been detected. The cross

Fig. 4. The Sun at noon on November 10, 2021, as seen from the Scorpio's black star.

is 90 mm wide and the oscillation were 5 mm - 10 mm. All the structure on top of the obelisk is 5 meter high, then the oscillations are small. The orthogonality of the whole structure verified from 64 m, has an angular accuracy of $\pm\ 10''$.

6. Discussion and conclusions

The measurement of the length of the sidereal day with a fixed star is not a simple experiment. It is necessary to have a fixed instrument and a very low parallax. The Vatican obelisk was the fixed reference frame used for this experiment, and the parallax was reduced by observing from the center of the granite meridian line of St. Peter's square. A mirror was used to aim at the same point of the line, to reduce the parallax down to 4 arcsec. The final result over 31 days on Antares from 21 July to 21 August 2021 allowed to detect the contribution of Special Relativity as stellar light aberration within less than a sigma (0.02 s) of the predicted theoretical value of the average sidereal day: 23h 56m 04.138 s $\pm$0.02 s, 0.0273 s larger than the aberrated value.

An extremely good result obtained with such historical instrument, the obelisk-meridian line of St. Peter's in the Vatican. A monocular, a mirror and a UTC synchronized watch and an audio-recorder has been a light equipment to perform such an experiment. The didactic value of this experience is paramount.

The deviation of the line from North and of the obelisk with the uppermost cross with the vertical are measured with a few millimeters accuracy, by comparing the transit's timings with the ephemerides.

References

1. J. Bradley, *Phil. Trans.* **35**, 637-661 (1727-1728).
2. C. Sigismondi, Venus transit of October 27, 2021 at the vatican obelisk. `https://drive.google.com/drive/u/2/folders/1ugl5TkhzQnE9qJcfJgx2BkiNnUIfKRAX`.

Betelgeuse, Sirius and the stars in the roman *Settecento*

Costantino Sigismondi
ICRA, Sapienza University of Rome,
Rome, Italy
E-mail: sigismondi@icra.it
http://www.icra.it/solar

Betelgeuse was reported as a magnitude "1-" in the Almagest of Ptolemy (150 AD), while it is of first magnitude star in the Uranometria of Bayer (1603). The latter with the stellar catalogue of Philippe de la Hire were used to include in 1701 the stars in the Great Meridian Line of St. Maria degli Angeli by Francesco Bianchini (1662-1729). For Sirius and Arcturus a series of star were positioned on the floor of the Basilica in correpondence with their positions up to two hours before their meridian transit, with 15 minutes of interval in the case of Sirius. Other stars in the floor represent the motion of the Sun on 20 August 1702 and in the Equinoxes, so that since two hours before the meridian transit it was possible to see how the Sun approached the Celestial Equator. Their astrometrical calibration with the Sun along the year is here published for the first time: for some of the stars the timing accuracy of their location is within a few seconds, as well as their distance from the pinhole, while in other cases it may reflect the "quantization" due to the terracotta bricks on the former floor of the Basilica. Because of the precession and proper motion the position of Sirius on the line changed from 1701 to 2021, and it has taken into account in this analysis. The black marble stars of the Vatican obelisk meridian line (1817), are calibrated with respect to the 3D star on top of the obelisk, by projecting the Sun behind it through a 7×18 prismatic monocular.

Keywords: Betelgeuse; Stellar Variability, History of Astronomy, Meridian Line, Santa Maria degli Angeli, Vatican obelisk, Rome, Francesco Bianchini (1662-1729).

1. Positional astronomy in 1700 in Rome

During the *Settecento* in Rome the research in astronomy was dedicated to Celestial Mechanics,[1] as in the other Eurpean Countries. In the various observatories present in Rome (the Jesuits at the Collegio Romano, the Dominicans at Minerva, Bianchini in St. Maria degli Angeli) the measure of their geographical coordinates with lunar eclipses was performed and compared with the other European cities. A lunar occultation was firstly observed by Giovanni Battista Audiffredi at Minerva, the one of Spica, with evidence of multiple nature of this stellar system. This observation was realized in view of a better knowledge of the lunar orbit. Francesco Bianchini (1669-1729) in St. Maria degli Angeli observed also three lunar eclipses in 1703 in order to verify the lunar theory in the Gregorian Calendar, and we used them to recover the UTC of the solar meridian transits. The coordinates of the stars used to measure the solar longitude were taken from the Catalogue of Philippe de la Hire (1700), of the Observatory of Paris. Bianchini placed 22 stars on the marble of the meridian line st. Maria degli Angeli, and Bellatrix, the western shoulder of Orion, or γ Orionis and Betelgeuse, the oriental one, are reported with their magnitude in

figure 4. Two of the stars have been found in the wrong place: both of the Northern hemisphere, placed in the Southern hemisphere. The reason was a misreading of the coordinates in the book of Philippe de la Hire, like if that work was hurried, without any time for revision. Other stars have been placed with an extreme care and precision, like for Sirius.

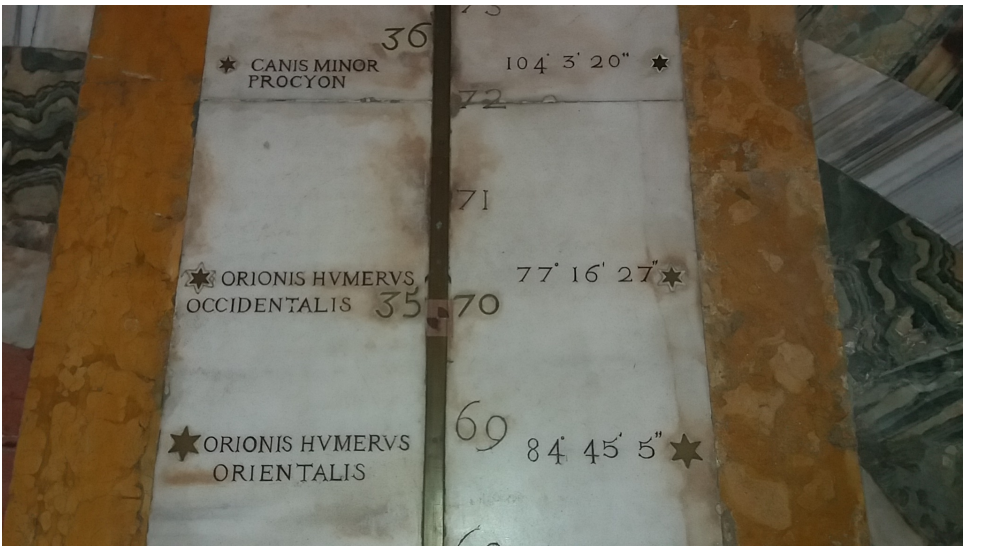


Fig. 1. Procyon, Bellatrix (occidentalis) and Betelgeuse (orientalis) on the meridian line of St. Maria degli Angeli, with their magnitude. Bellatrix is smaller than Betelgeuse. 35° is the zenithal distance of Bellatrix, $70 = 100 \times tan35°$.

2. Magnitudes and stellar positions in St. Maria degli Angeli

Bianchini wanted to represent the zodiacal signs with the corresponding constellations of the Atlas of Johannes Bayer, the Uranometria (1603). Francesco Tedeschi realized the marble *tarsie* and Bianchini provided the magnitude's scale according the Catalogue. Also on the meridian line the magnitudes were attributed according to that Catalogue, not being stellar variability still recognized as a general property of the stars.

The position of Sirius on the meridian line is the one of 1700 "precessioned" to 1701. The precession was the only one motion of the celestial sphere to be known at that time. The proper motion was discovered by Edmund Halley on Arcturus' coordinates few years later in 1718. The stellar aberration, which is annual, was discovered by James Bradley in 1727 on γ Draconis.

If we compute the position of Sirius with the coordinates of today, applying the precession down to 1701 we do not found its position on the meridian line, because of its proper motion in declination of 1.211 arcsec/year Southwards, which

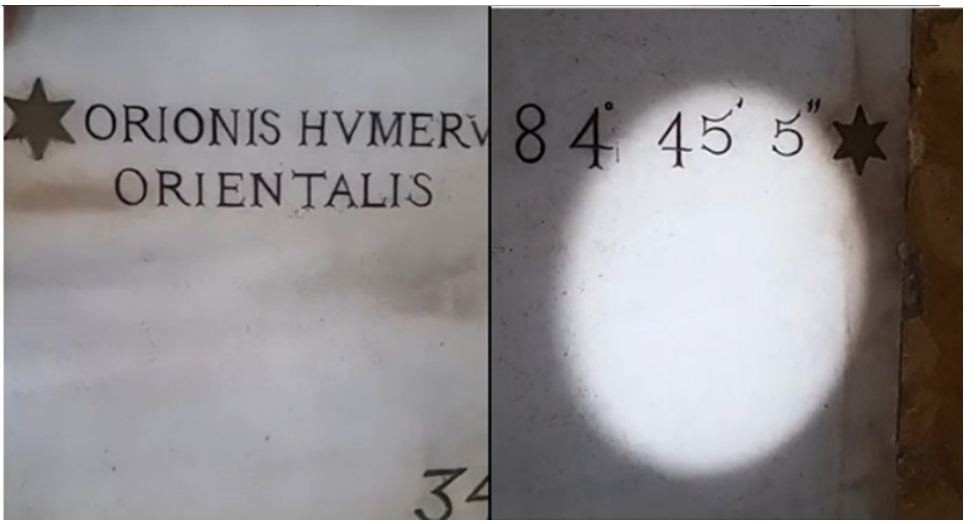


Fig. 2. The Sun passing over the position of Betelgeuse in the spring of 2018. The ecliptic longitude $84°45'5''$ of 1701 is reported on the meridian line, in correspondence of its declination $34°$ is the zenithal angle of the Northern limb of the Sun at the moment of the photo.

Fig. 3. The star of the celestial North pole in the Basilica of St. Maria degli Angeli.

in 320 years totalized 387 arcsec, i.e. 0.68 centesimal parts or 137 mm towards the Capricorn sign.

The positions of the stars of series of Sirius, on the hyperbola of Sirius' path on the floor of the main hall, the original roman hall of 30 meters of heigth, have been placed to mark each 15 minutes the approach of Sirius to the meridian line.

Fig. 4. The ellipse of 2100 orbit of the Polaris, around the North Celestial Pole.

Fig. 5. The constellation of Leo at the meridian line of St. Maria degli Angeli, represented with its stars and corresponding magnitudes, according to the Uranometria (1603).

Excepted the star of -2 hours, which is absent without signs of effraction, all other stars are placed within 10 seconds of accuracy, from the actual time of meridian transit. This verification has been realized on February 5 and November 8 2021 by observing the Sun transiting on the same path. Knowing the instant declination of the Sun, also the distance from the pinhole's vertical has been verified.

As example I report the observations of the star 1 hour 45 minutes, 45 and 30 minutes before the meridian transit on November 8th 2021.

- first contact 10:08:12 – 11:08:10 – 11:22:34.5
- last contact 10:09:42 – 11:09:40 – 11:24:02
- star inside for 01:30 – 01:30 – 01:27.5

The conjunction was respectively at 10:08:57, 11:08:55 and 11:23:18.5 and the solar meridian transit occurred at 11:54:04, then with an advance respectively of 1h45m07s 45m09s and 30m40s. The time difference was -7, -9, and -40 seconds. The 30 minutes-before star shows a wrong position in timing (azimuth) but a correct position in declination (distance from the pinhole's projection) since the duration of the timing is nearly constant (90 s) or compatible with the motion in declination of the Sun during 1h 15 min of 0.86 arcminutes.

Moreover the solar transit lasted 2 m 21 s, and the transit over the star 1 m 30 s, this meant that the center of the star was separated by the solar one by 33 s (see fig.1).

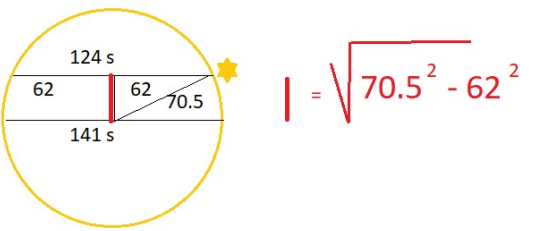


Fig. 6. The theorem of Pitagora can be applied on the solar ellipse, because the properties of its chords are preserved in the projection on the floor of the Basilica. The data are of the meridian transit of 7 Novembre 2021: 141 s is the transit's duration and 124 s is the time of Sirius star inside the solar moving ellipse on the floor.

By using the same technique to the ten stars of Sirius all over the floor of the Basilica we found the stars with azimut (timing) errors of 7 to 40 seconds, or 5 cm to 16 cm, while the declination (distance from the pinhole's vertical point) is correct, since the duration of the contacts of the solar limb with the stars is constant.

The azimuthal dimension of 14 cm, is compatible with the presence of terracotta bricks on the ancient floor of St. Maria degli Angeli, the one realized by Michelangelo.

3. Equinoctial line and Papal path of 20 August 1702

It seems like the stars, originally, were all placed at the center of the closest terracotta bricks, placed with the long side parallel to the main hall of the Basilica. Another way of describing that fact is that 11 bricks were prepared with a star in

Fig. 7. The Sun on the meridian line at Sirius' declination on November 7th 2021. The star is included in the solar image. It transit lasted 141 s and the "time of flight" of Sirius was 84 s.

Fig. 8. The solar image "climbs" the column of Vanvitelli on November 7th 2021.

the middle, and later they were used to substitute the selected bricks by the actual path of Sirius. The stars of the Celestial Equator show an alignment not exactly perpendicular to the meridian line, by two arcminutes, or 11 mm over 20 meters, the timings are accurate. It may reflect the use of a "paleo" meridian line which

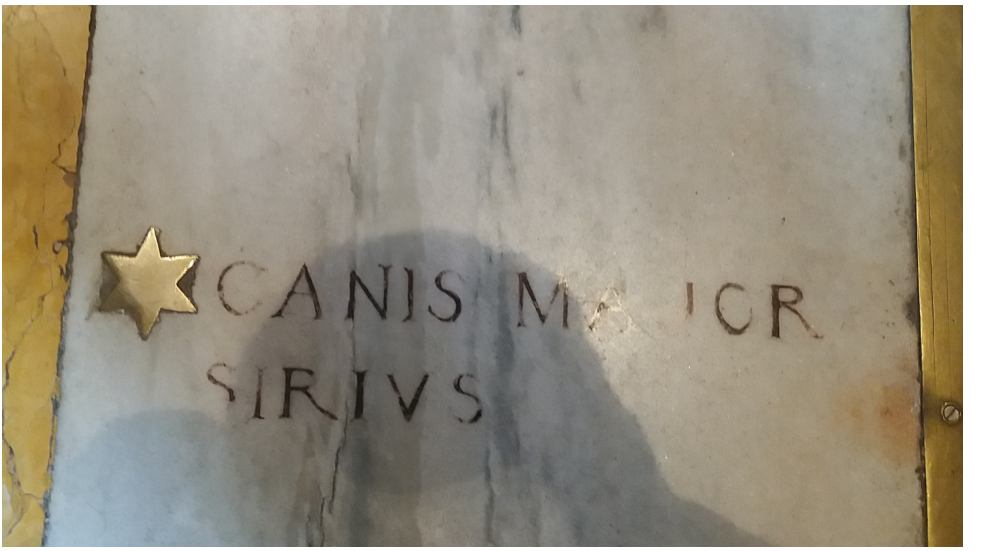


Fig. 9. The star of Sirius on the marble *Imetto* of the meridian line.

Fig. 10. The Sun four minutes before the meridian transit of 8 November: its position with respect to Sirius' star is nearly tangent. With respect to 7 Nov 2021 the Sun is 16.5′ more South, about 50% of its diameter. This image, with the one of figure 6, account for this fact precisely, being a differential measure. The Sun moves perpendicularly to the brass line.

was better aligned with the North than the present one, modified after the intervent of Giacomo Filippo Maraldi, an astronomer of Cassini's observatory visiting Rome when Bianchini was tracing the Great Gnomon. He had his own observing place in the Tower of Biscia in Palazzo Venezia, from which he transmitted the visual

signal to S. Maria degli Angeli. The boreal line, 6 meters long and a few months "older" than the Great Line, was used by Bianchini only to measure the latitude of the pinhole, and later abandoned, but it was better aligned to the North than the Great one. This would explain the 2 arcminutes of shift of the Equator, materialized by the alignment of four stars placed each 30 minutes before the meridian transit.

For the stars of the papal path the accuracy on their location is slightly better than of Sirius, but always a few cm of difference with respect to the theoretical positions are found.

Fig. 11. One of the stars of the solar phat of August 20, 1702, when the pope Clement XI visited the meridian line. Photo of August 19, 2021.

4. Marble disks in St. Peter's square in 1817

Their location on the meridian line is connected with the top features of the obelisk. The marble is really of fine quality, offering a very good reflection to the projected image through a telescope. Also the chromatic aberration is well visible.

Fig. 12. The star of Scorpio on the meridian line of St. Peter's square. The line was realized in 1817, the director of the works and committent was Cardinal Pietro Maccarani, head of the Fabric of St. Peter. The art of marble *tarsie* was still at its apogee. Photo of 20 October 2021.

5. Conclusions and Perspectives

The stars of St. Maria degli Angeli described in this text are important to understand the technology used to place them, owing their astronomical significance. The verification of their positions requires still more observations, because on 5 February and 8 November 2021 not all stars were measured. Better data are available for 19 August 2021 for the stars of the papal visit. Has Bianchini placed the stars, and

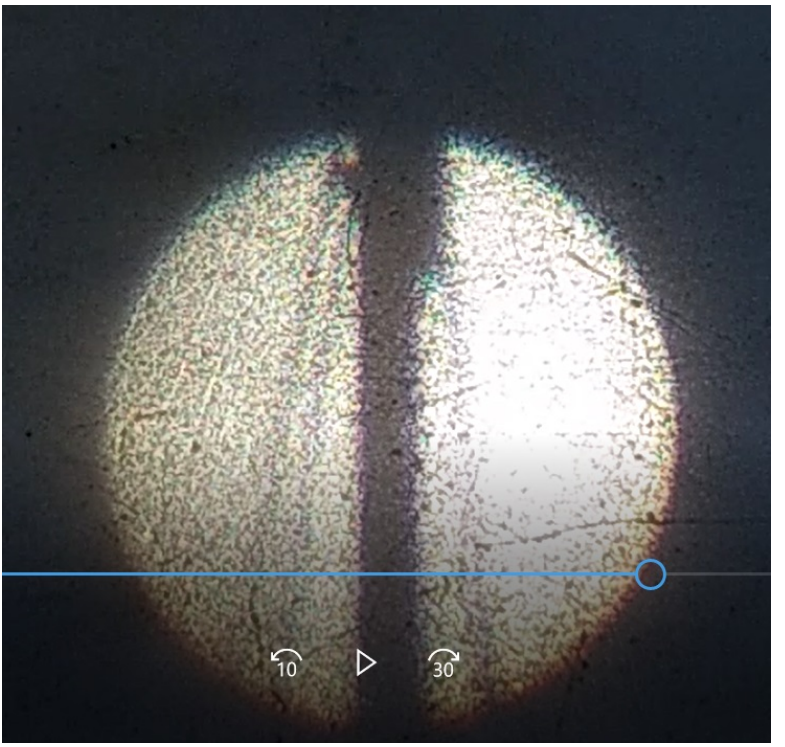


Fig. 13. The Sun projected on the Scorpio marble on October 18, 2021, at the meridian transit: the central part of the uppermost cross is visible.

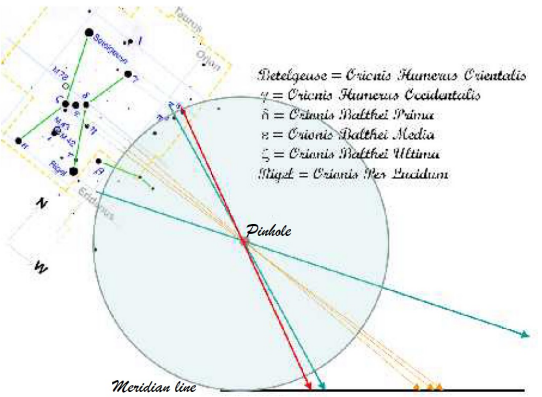


Fig. 14. The projection of the stars of Orion along the meridian line through the pinhole.

Fig. 15. A star on the floor of St. Peter's Basilica.

Vanvitelli left them in the same place? Has Bianchini placed the stars of Sirius in the center of the bricks and then changed only the relevant ones? Or was continuous the position of the stars on the floor of the Basilica since the inauguration of the meridian line in 1702?

In the Vatican the stars on the meridian line of St. Peter's square and the stars on the obelisk belong to different centuries: nineteenth and sixteenth one. Pope Sixtus V placed some obelisks in the squares of Rome (Laterano and Piazza del Popolo and St. Peter's square) as well as the one of St. Maria Maggiore. On the top of these obelisks a three-dimensional star is always present. It is like the crossing point of all the direction, a geodetic point near each relevant point of the Urbs. Sixtus V was the ruler of the urbanistic of Rome, and its plan is effective still nowadays. I presented here the observations in St. Peter's square made with a small telescope, like the one that Bianchini used in the Basilica of St. Maria degli Angeli to observe the stars transiting through the window over the pinhole. The sidereal period of the Earth's rotation was recovered to the nearest 1/100 s, by using the vatican obelisk

Fig. 16. Before and after the solar meridian transit, as seen from Scorpio on November 10, 2021.

Fig. 17. The ghost image of the Sun behind the three-dimensional star superimposed to the obelisk's profile, and its nearly 100 m-long shadow at 10:40 AM on November 18, 2021. Both the Sun and its halo are overexposed, and only the ghost image holds the solar position's information.

Fig. 18. The Sun projected near the Scorpio's star (left Oct 20, 2021, -20 s transit's advance) and Sagittarius' star (right: Nov 18, 2021, -27.8 s). The connection between the star and the solar occultation by the three-dimensional star above the obelisk is not given by the dates reported on the zodiacal disk, but there is a 3-5 days anticipation, due also to the concave shape of the square.

Fig. 19. The dimension of the solar image projected on November 18, 2021 compared with 0.50 euro. The mini-telescope has 7 x magnification, with respect to a pinhole image. This image was projected at 300 mm of distance from the telescope's objective. The pinholhe image would have sized 3 mm (always 1 to 100 the ratio diameter:focal length of the solar image). Then the image is 21 mm of transverse diameter. The meridional diameter is rather 2 times the transverse one.

as sighting reference. This technique has been adapted also to hazy days, when the image cannot be projected to the marble: a direct vision of the obelisk is realized as in fig. 16. The calibration of the line with respect to the celestial North, and of the obelisk and the uppermost cross with respect to the vertical axis of symmetry of the obelisk is another fine technique, to inspect the details of the wonderful manufacture of about two millenia B. C. the obelisk from Heliopolis, dedicated to the Sun healer from blindness. Art, religion and science are profoundly entangled in this approach to the stars in the roman *Settecento*.

Fig. 20. The comparison between the direct image (overexposed halo) and the ghost image, on November 17, 2021 behind the Northern fountain's water. The top of the obelisk, as seen from the meridian black star is visible, also in the left-down pane, where the Sun's halo saturates most of the image, and some water's droplets appear black. The instant of the meridian transit was chosen for this screenshot, as the green ghost image shows, being divided into two equal parts by the top cross. The meridian transit occurred at 11:54:41, 28 s before the ephemerides' one at 11:55:09. The use of the small telescope allowed to reach always a timing accuracy better than one second. The offaxis angle of the telescope was evaluated by the motion of the projected image in term of the solar angular diameter of $32.4'$ entering the Sagittarius. The lens is placed on the meridian line axis within 0.5 mm. The transit lasted 2 min 16 s, $\varnothing_\odot = 32'\ 05'', 16''$ less for Nov 18.

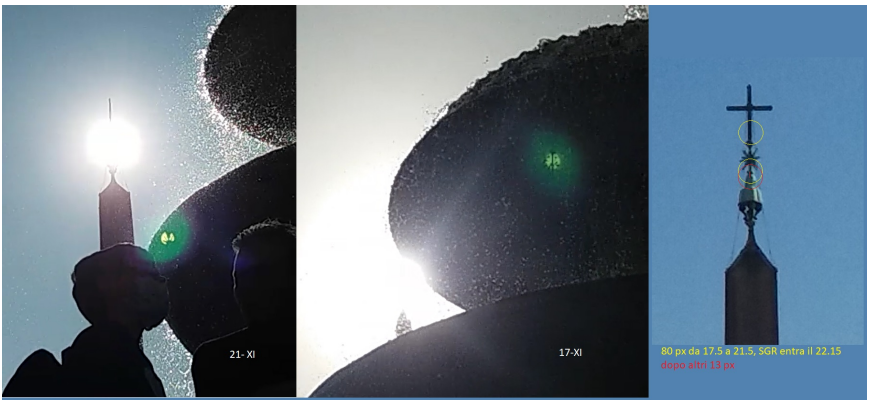


Fig. 21. The Sun filmed from the Sagittarius' black star on November 17 (center) and 21st (left). The position of the Sun is reported on the cross on 17 November and below the 3-D star on 21st. The red circle is where the ingress in Sagittarius occurred on 22.15 November 2021.

Fig. 22. The Sun at 12:07:40 of 4 December 2021, 8 minutes after the meridian transit. The yellow empty disk represents the position of the Sun at the meridian, and the red one the position at the Winter solstice of December 21, 2021. The photo is taken from the zodiacal marble disk of the Capricorn, on the black star. The alignment between the Sun at the solstice and the 3-dimensional star is confirmed as in the other austral zodiacal signs of Scorpio and Sagittarius.

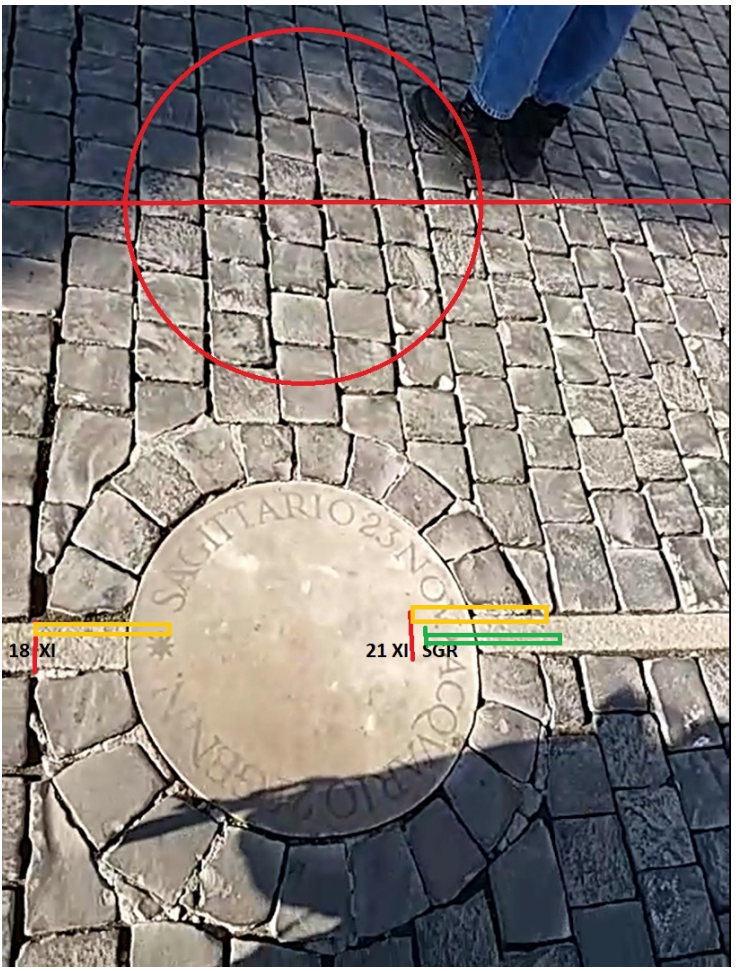


Fig. 23. The principle of the shadow's motion perpendicular to the meridian line around the meridian transit has been used to draw this image: the red line and circle are used to predict the position of the Sun 25 minutes later at the meridian transit: the Sun disappears at the ground level at the two extremes of the meridian diameter. Then the Sagittarius' marble disk has been located in the sole point were the Sun could illuminated all the disk's meridian diameter at the ingress of Sagittarius. The photo is of Sunday 21 November 2021, 16 hours before the ingress of the Sun into Sagittarius, and 30 minutes before the papal Angelus.

References

1. C. Sigismondi, *Meridiani e Longitudini a Roma* Sapienza University of Rome, 2006 and Biblioteca Casanatense 2005. `https://rosa.uniroma1.it/rosa03/semestrale_di_geografia/article/view/15345`, `https://casanatense.beniculturali.it/?p=75`.
2. C. Sigismondi, *Lo Gnomone Clementino* Gerbertus **7** 1, 80, 2014.
3. C. Sigismondi, *Lo Gnomone Clementino: Astronomia Meridiana in Chiesa dal '700 ad oggi* `https://arxiv.org/ftp/arxiv/papers/1106/1106.2498.pdf` 2016.
4. M. Catamo and C. Lucarini, *Il Cielo in Basilica* Arpa-Agami, Roma, 2012.

5. C. Sigismondi, *Sirius' path in the Basilica*: 5 feb https://drive.google.com/drive/u/2/folders/1vRCHMTOueVmARxydUVbmabzSQEjAzB_v 8 nov https://drive.google.com/drive/u/2/folders/1z9uP-3qODZAbKVOwKORwlFhRPxedHqw4.
6. C. Sigismondi, Meridian transits in St. Peter's square in November 2021 from Sagittarius' disk 17 nov https://drive.google.com/drive/u/2/folders/1qS8Xrg0dz1Mt91oEdHLdMhnRPZ7il11K 18 nov https://drive.google.com/drive/u/2/folders/1S5Ovd6cQTfcVzmYkUVCvlz8g7T2So9Pw 20 nov https://drive.google.com/drive/u/2/folders/1N5XLETPkApu7SE9e2ngdS-_KyJninrbw 21 nov https://drive.google.com/drive/u/2/folders/1g3N5pBNTVwDL26ei_zTql1hklcq_rnyc.
7. Papa Francesco, *La mia idea di arte* 37-42 Mondadori, Milano 2016.

On Einstein's last bid to keep a stationary cosmology

Salvador Galindo-Uribarri* and Jorge L. Cervantes-Cota†

Departamento de Física, Instituto Nacional de Investigaciones Nucleares, Carretera México Toluca Km. 36.5, Ocoyoacac, C.P. 52750, Edo. Mex., México

**E-mail: salvador.galindo@inin.gob.mx*

†E-mail: jorge.cervantes@inin.gob.mx

It is commonly known that the steady-state model of the universe was proposed and championed in a series of influential papers around mid-twenty century by Fred Hoyle, Hermann Bondi, and Thomas Gold. In contrast it is little known that, many years before, Albert Einstein briefly explored the same idea; that is of a "dynamic steady state" universe. In 1931 during his first visit to Caltech, Einstein tried to develop a model where the universe expanded and where matter was supposed to be continuously created. This latter process was proposed by him to keep the matter density of the universe constant. However, Einstein shortly abandoned the idea. The whole event has already been described and analyzed by C. O'Raifeartaigh and B. McCann in 2014. It is the purpose of this brief note to point out what might have prompted Einstein to consider a continuous creation of matter and the prevailing circumstances at that time that drove Einstein's intent.

Keywords: History of Physics; Cosmology; Einstein.

1. Introduction

The first relativistic model of the cosmos is due to Einstein. His model entailed the idea of a universe both isotropic and homogeneous on the largest scales (idea known as "Cosmological Principle"). At that time, it was the accepted view that the universe was stationary. This was not unreasonable since relative velocities of stars are small. To achieve a stationary universe, Einstein added in 1917 to his original field equations, the so-called cosmological constant "λ" to counterbalance the effects of gravity and attain a static universe[1]:

$$G_{\mu\nu} - \lambda\, g_{\mu\nu} = -\kappa \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right) . \tag{1}$$

Einstein showed that the value of the constant he introduced was proportional to the mean mass density ρ of the universe and inversely proportional to the square of R, its radius of curvature,

$$\lambda = \frac{\kappa\rho}{2} = \frac{1}{R^2} . \tag{2}$$

The introduction of λ in the original field equations was accepted by the early few practitioners of General Relativity as convenient to keep the current standard view at the time, of a static universe.

However, Einstein was uneasy with his forced introduction of the λ constant. Several of his papers show his uneasiness. For instance, in a 1919 publication Einstein stated:

"...the general theory of relativity requires that the universe be spatially finite. [This] requires the introduction of a new universal constant λ, standing in a fixed relation to the total mass of the universe (or respectively, to the equilibrium density of matter). This is gravely detrimental to the formal beauty of the theory".[2]

It was not until 1931 that Einstein at long last dropped off λ from his field equations. During those days, Einstein was visiting professor in Pasadena California attending an invitation made by Robert Andrews Millikan to spend a short season at Caltech (from late December 1930 to early March 1931). Soon after his arrival to California he discussed Hubble's redshift measurements and its implications with Mount Wilson Observatory astronomers.

Today there is still a justified widespread view that Einstein discarded the cosmological constant immediately after he was satisfied of the validity of Hubble's evidence for a non-static universe. This rendered λ in his field equations, redundant. In April of that same year, Einstein submitted for publication, a paper to the Sitzungsberichte der Preussischen Akademie der Wissenschaften where he rejected his cosmological term as superfluous and no longer justified. In his own words, "theoretically unsatisfactory".[3]

But the view that Einstein dropped off λ just before his meetings with the Mount Wilson astronomers (in early Jan 1931) is not strictly exact. In the interval between his arrival to California and his 1931 Sitzungsberichte paper (April 1931), he made a last effort to model a "dynamic steady state" universe, keeping λ in his field equations. This was serendipitously discovered in 2013 by Cormac O'Raifeartaigh in an unpublished Einstein's manuscript kept in the Albert Einstein Archives (AEA) maintained by the Hebrew University of Jerusalem.[4] The first translation into English of the manuscript, contents and its analysis, has been already covered in 2014 by his discoverer Cormac O'Raifeartaigh and his colleague Bruce McCann.[5] In addition, the manuscript contents were also commented in a note added in proof by Harry Nussbaumer and later also reviewed by him.[6,7]

2. A significant finding

The discovery by O'Raifeartaigh of the unpublished manuscript showing a model still using the lambda constant, came as a great surprise to everyone. From the moment Einstein arrived at Caltech it seemed that he had already accepted a non-static universe. Various reports in the local press affirm this. And consequently, people supposed that Einstein had already scrapped the constant from his field equations as unnecessary.

As a celebrity, during his visit to California, Einstein's activities and sayings were reported daily by the press. In Jan 3 the New York Times (NYT) reported that during an interview given the previous day Einstein stated: "New observations

by Hubble and Humason (astronomers at Mount Wilson) concerning the red shift of light in distant nebulae make the presumptions near that the general structure of the universe is not static".[8] The next following month the NYT (Feb 5) another front-page story, informed that Einstein delivered a lecture the previous day (Feb 4) where he no longer held to the model of "Stable universe".[9]

On January 15, Einstein attended a welcome dinner in his honor at "The Athenaeum", a club house for the California Institutes Associates (a group of promoters of southern California scientific and scholarly research). The dinner was attended by around 200 guests among them a selected group of astronomers that had collaborated with their own research in testing relativity.[10] The final dinner speech was delivered by one of them, Walter S Adams, director of Mount Wilson. Adams had verified Einstein's prediction of gravitational redshift.[11] During his speech Adams highlighted the problem of the "nature and structure of the universe" and he announced that:

"Professor Einstein is now inclined to consider the most promising line of attack on the problem to be based on theories of a non-static universe, the general equations of which have been developed so ably by Dr. Richard Chase Tolman, of the California institute of technology".[12]

Despite the multiple examples that can be cited, affirming Einstein's alleged intention to follow a non-static approach, the 2014 discovery by O'Raifeartaigh of the manuscript, revealed that Einstein temporarily followed a different scheme in his unpublished document.

3. The manuscript

The unpublished manuscript is entitled "Zum kosmologischen Problem", that is located in the Albert Einstein Archives (AEA) (draft, 1931. Doc [2-112]). It is a signed, four-page handwritten manuscript by Einstein on American paper. Assigned by AEA to January or February 1931.

The cosmological model depicted in the manuscript was not previously detected given that the first words of its title are identical to those of Einstein's 1931 Sitzungsberichte paper (April 1931). For this reason, it was assumed to be a draft of the latter publication.

The manuscript has already been extensively analyzed by O'Raifeartaigh and colleagues[13] and in the here cited papers by Nussbaumer, so we shall limit ourselves to giving a succinct description of its contents based on those publications.

4. The model in the manuscript

In the manuscript Einstein explores a solution to his field equations retaining λ that could be compatible with Hubble's observations; to be precise an expanding universe in which the density of matter does not change over time. Einstein starts his analysis by choosing the metric of flat space expanding exponentially (De Sitter

metric):

$$ds^2 = -e^{\alpha t}(dx_1^2 + dx_2^2 + dx_3^2) + c^2 dt^2. \tag{3}$$

So, the distance between two points increases over time as $e^{\alpha t/2}$, and he remarks: "one can thus account for Hubbel's [sic] Doppler effect by giving the masses (thought of as uniformly distributed) constant co-ordinates over time".

In his calculations he finds that α^2 (representing the expansion of the universe) is related to its overall density ρ as,

$$\alpha^2 = \frac{\kappa}{3}\rho. \tag{4}$$

That is, ρ determines the expansion. Since the redshift measurements at the time suggested that the universe expansion is constant, so he concluded that the density must be constant as well.

In the final part of his manuscript, Einstein proposes that the density of matter remains constant by supposing a continuous formation of matter in empty space. He then observes that: "For the density to remain constant, new particles of matter must be continually formed within that volume from space."

At the end of the manuscript, he associates the cosmological constant with an energy of space: "by setting the λ-term, space itself is not empty of energy". Consequently, the continuous creation of matter becomes associated to the λ cosmological constant.

Einstein's "dynamic steady state" universe simultaneously incorporated the observed expansion of the universe together with a new paradigm, namely of an expanding universe that maintained its isotropy and spatial homogeneity the same as it always has and always will (later dubbed the perfect cosmological principle) by means of the continuous creation of matter.

The "dynamic steady state" model was so compelling to set aside. However, Einstein didn't submit his manuscript for publication. O'Raifeartaigh and colleagues detected the possible reason why Einstein gave up publishing it.[13] He must have noticed on revision, that his model contained a flaw. Once the error is corrected the model leads to the trivial solution $\rho = 0$, that is an empty universe of matter. Therefore, he promptly abandoned this attempt.

We must ask now what it was that attracted Einstein to consider the possibility of a universe where there is a continuous creation of matter. To find a possible answer we shall consider the reception at that time, of an expanding universe as opposed to prevailing static views.

5. Universe's age and Earth's age tension

Towards the beginning of the 1930s the main objection to the interpretation of the Hubble constant as the indicator of the universe expansion was that it's measured value implied a younger universe than that of Earth's accepted age.

Before 1930's, scientific estimates of the age of the Earth dated from mid-19th century, from Helmholtz-Kelvin gravitational contraction ages,[14] passing through those resting on geology (i.e., the amount of salt in the oceans, sediments) and - after a fierce debate between geologist and physicists[15] - finally deferred to an age based on early-20th century radioactive dating.[16] This produced by the 1930's, an estimate of the universe age of around 4 billion years.[17]

In 1931, at the time Einstein was visiting Caltech, the matter of determining the age of the earth was not entirely resolved to the extent that the National Research Council decided to appoint a committee to investigate to settle the question.[18]

On the other hand, the cosmological age of the universe based on the universe expansion, was estimated by taking the reciprocal of the value of the Hubble constant (Hubble's time). Hubble first evaluation of his eponymous constant was of around 500 kilometers per second per megaparsec. This implied a universe's age of about two billion years, which was in a tense contradiction with the estimated age of the Earth, as just mentioned, of about four billion years. As consequence, such mismatch created room for doubt, Einstein himself included. Some critics questioned that the observed nebulae redshifts, were in fact a manifestation of the Doppler's effect. Such was the panorama that reigned in the early 30s.

This incongruity raised two possible explanations: on the one hand Hubble constant value was wrong, or on the other hand, Doppler's shift needed a novel interpretation, as it is said today "perhaps new physics".

As it is well known, the Hubble constant estimation involved measurement of distance and velocity of a variable star (a Cepheid) belonging to a galaxy in question. This involved monitoring the apparent brightness of the Cepheid variable to obtain its period. Then, using Shapley's Cepheid calibration, its absolute brightness was established and consequently the distance to the galaxy where the Cepheid was located. At that time there was little doubt on the correctness of Shapley's calibration. Also, measurement of recession velocity of galaxies even in those days, was straightforward. Spectral recordings had already achieved good accuracy. So, Hubble's constant value was in little doubt. Eventually, some astronomers pointed out that Cepheid's calibration could be slightly inaccurate as the apparent star's luminosity could be diminished by interstellar media and that had to be accounted for. It must be remembered that It was not until mid-1940's that Walter Baade identified two Cepheid types and made a major correction to Shapley's calibration and thus to Hubble's value.[19] Now we know that the problem was on the way off value of Hubble's constant at that time.

On the second possibility, that the incongruity between the ages of the earth and the universe had its origin in an erroneous interpretation of the observed redshift, Fritz Zwicky gave in 1929 an alternative explanation. Zwicky, a resident scholar at Caltech, suggested the concept of "tired light".[20] This was a hypothetical redshift mechanism where photons lost energy over time through interactions with other particles in their trajectories through a static universe. So, the more distant objects

would appear redder than more nearby ones. It is pertinent to mention that the term "tired light" was not used by Zwicky but later coined to refer to this concept by Caltech cosmologist Richard Tolman in the early 1930s in relation to the so called Tolman Surface Brightness Test.

6. Earlier dynamic stationary state cosmologies

Zwicky was not alone. There was also an academic minority that put forward similar ideas to Zwicky's explanation even before his 1929 suggestion. All these "tired light" propositions had in common the assumption that photons heading earthward, somehow interchanged with the intergalactic medium, part of their energy, thus red shifting their frequencies.

"Tired light" was championed by the few enthusiasts of steady state cosmologies. Its appeal as we shall see resides on the fact that it circumvents the standard interpretation of the redshift as an indicator of galaxy recession and thus safeguards their view of a static universe.

The Nobel laureate Walther Nernst and William Duncan Macmillan, a well-known Chicago professor, were leading advocates of a steady-state universe. What these had in common is that they presuppose the universe as eternal, self-preserving structure in which matter and radiation are constantly been transformed into one another to keep the universe in a stationary balanced state thus avoiding a heat-dead of the universe. A cosmic destiny they both abhorred.

MacMillan's supposes in his own version of "tired light": "That there is a leakage of energy from the photon [...] due perhaps to an inherent instability in the photon, or, possibly, [due] to collisions with other photons". He concludes therefore: "it is evident that the frequency [of the photon] declines with the energy, and the lines of the spectrum are shifted toward the red".[21]

In addition, he considers that the "evaporated" energy from the photon continues to exist as abundant radiant energy of "very low frequency". A kind of primitive version of the modern CMB radiation (but the latter being Black Body radiation). Then he puts forward the possibility that perhaps the "evaporated" energy "disappears into the fine structure of space and reappears eventually in the structure of the atom". In other words, he proposes a mechanism of matter creation.

Another notable hypothesis was that of the Nobel laureate, Walther Nernst, who suggested that some photons are partially absorbed by the luminiferous æther (which he accepts to exist).[22] In his 1928 essay "Phyco-Chemical Considerations in Astrophysics", Nernst outlines the notion of matter creation in his stationary universe:

"I may therefore hold fast to the hypothesis uttered by me that, just as the principle of the stationary condition of the cosmos demands that the radiation of the stars be absorbed by the luminiferous æther, so also finally the same thing happens with mass, and that, conversely, strongly active elements are continually

being formed from the æther, though naturally not in amounts demonstrable to us,...".[23]

It is worth noticing that both, MacMillan and Nernst, claim energy transfer occurs by an unknown process yet to be discovered.

The "tired light" proposal did not vanish into oblivion as alternative explanation to redshift been caused by recessional motion. As late as 1935, Edwin Hubble himself and Richard Tolman (both at Caltech during Einstein's several yearly visits) investigated the possibility that "tired light" might be an alternative interpretation of redshift.

They did that comparing the surface brightness of galaxies as a function of their redshift (applying the so called "Tolman Surface Brightness Test"[24]). According to Tolman Test, the relationship between surface brightness of a galaxy and its redshift differs in the case of a static universe from that of an expanding one. In a coauthored publication they explored this possibility.[25] In their joint paper and to simplify their analysis, they employed Einstein's static model of the universe: ... "combined with the assumption that the photons emitted by a nebula [a galaxy] lose energy on their journey to the observer by some unknown effect, which is linear with distance, and which leads to a decrease in frequency, without appreciable transverse deflection". In short, they employed Einstein's model plus "tired light". As result of their analysis, they conceded: "Until further evidence is available, both the present writers wish to express an open mind with respect to the ultimately most satisfactory explanation of the nebular red-shift" ..."They both incline to the opinion, however, that if the red-shift is not due to recessional motion, its explanation will probably involve some quite new physical principles".

7. On Einstein's arrival at Caltech

Up until now, we have seen that in the late 1920s and early 1930s, Hubble's law was commonly interpreted as a demonstration that the universe was truly expanding. However, there was still the problem of reconciling the age of the universe with that of the earth. This inconsistency, in addition to raising a reasonable doubt about the interpretation of the redshift, gave rise to the credibility of the theories of a stationary universe. For some, the appeal of a universe without evolving through time resided in that such universe has no beginning and no end, so it converts the "age of the universe" as a senseless question issue, consequently eliminating the earth-universe ages paradox.

Attempting to develop a stationary cosmological model was indeed an attractive motivation for Einstein, since it would avoid his well-known aversion to a universe that has a beginning and reinforce his well-known paradigm of a static universe. But adopting such an idea required reliable and robust observational evidence of the continuous creation of matter.

As we will see below, fresh "evidence" (that turned out to be illusory) of the continuous creation of matter emerged just as Einstein arrived at Caltech. This came

from Millikan (Einstein host at Caltech) so it must have seemed reliable to Einstein. We must remember that Millikan received the Nobel prize partly for experimentally verifying Einstein's photoelectric equation.

8. Birth cries of atoms

On the 1st of February 1931, during Einstein's stay at Caltech, a paper by Millikan and his collaborator H. Cameron appeared in Physical Review. Its title was "A more accurate and more extended cosmic-ray ionization-depth curve, and the present evidence for atom-building".[26] This paper contained what it would seem to be the experimental evidence Einstein needed, that is, observational evidence of continuous matter creation. In our opinion, it was this publication what encouraged Einstein to use the continual creation of matter hypothesis in his unpublished manuscript. So, Einstein made his last bid to keep a stationary cosmology.

Millikan's 1931 paper is the apogee of a series of publications on cosmic rays where he explores their composition and origins. But, before we comment on this publication it is convenient to recall very briefly Millikan's research on cosmic rays. During the early 20's he and his collaborators made a series of observations on board of balloons, high altitude peaks, and different geographical locations, at various latitudes and in the depths of lakes proving in his 1926 publication that cosmic rays were of extraterrestrial origin.[27] Millikan believed that his measurements proved that the primary cosmic rays were energetic photons. He also stated that cosmic rays, "...must arise from nuclear changes of some sort..." but far more energetic than any radioactive change thus far on record".

In the same 1926 paper he suggested that cosmic rays probably came from among the following three nuclear processes: a) The capture of an electron by the nucleus of a light atom, b) the formation of helium out of hydrogen, or c) some new type of "nuclear change", such as the condensation of radiation into atoms.

Regarding third suggestion he made above, we must emphasize that Millikan here raises the creation of matter from the "condensation of radiation into atoms." this idea was not his, as he himself stated in a note written by him for the journal Science in 1930.[28] In the note he recalls having in 1915 discussions on the "running down of the universe" (i.e., when the universe has reached a state of maximum entropy) at the University of Chicago with his then colleague William Duncan Macmillan (of whom we have already commented) where the issue of atom building was discussed. Millikan recalls,

"In our conversations at Chicago W. D. Macmillan constantly held out for the view that a still further step forward should be taken and that the idea of "running down of the universe" should be given up by the assumption that atom building went on in space by the condensation of radiation into atoms. He discussed his idea in detail with me in the year 1915, and in July, 1918 he published it in full".[29]

In 1928 Millikan and Cameron found that incoming cosmic rays could be grouped into three independent energy bands centered at 26, 110, and 120 (MeV).[30] They

argued that these bands are produced by the release of photons when eventually a "sudden union" (i.e., fusion) between atoms occur. To support their assertion, they observed that the center of the bands agreed respectively with 26 MeV (which is just about the mass defect of helium), 110 MeV (which is close to the mass defect of oxygen and nitrogen) and 220 MeV (to that of silicon.) So, their inference was that the three photon bands reaching the earth must be generated by the "sudden union" of: 4 hydrogen atoms fused to form helium, 14 to form nitrogen, 16 to form oxygen and 28 to form silicon". Secondary electrons, they claimed, were produced in the atmosphere by Compton scattering of gamma rays. Millikan believed that space was filled with a tenuous gas of electrons and protons (the latter he called "positive electrons"). So, to get around the "running down of the universe" he assumed that "These building stones [protons and electrons] are continuously being replenish throughout the heavens by the condensation with the aid of some as yet wholly unknown mechanism of radiant heat into positive and negative electrons".

Millikan supposed discovery of the continuous formation of matter went beyond the scientific sphere provoking the attention of among the scientific and the lay publics. In a statement made by Millikan to the press, cosmic rays were "birth cries of atoms, a Millikan phrase that achieved a good deal of currency".[31]

Regarding Millikan's 1931 publication the one which appeared on print just at the time Einstein was hosted by Millikan, this was basically a continuation of his 1928 paper including revised energy calculations that further bolstered his tenacious ideas. A short time later Arthur Holly Compton showed that not all cosmic rays were photons, but at least a large part of them consists of charged particles. This led to a debate between the two Nobel laureates. The debate went on for some time with Compton at the winning side but that's another story.

9. Final comments

Einstein before long abandoned the idea of keeping the λ constant in his field equations. In his next publication, the 1931 Sitzungsberichte paper, he does not make use of the constant anymore.[3] At the end of the paper Einstein adds some remarks about the age of the universe problem, which was quite severe without the use of the λ constant. Today we know that the problem was on the way off value of Hubble's constant at that time. Einstein signals two possible errors that may be occurring in the initial approach to the problem. First, he insinuates that the matter distribution might be inhomogeneous and that a homogeneity approximation may be illusional. Then he adds that in astronomy one should be cautious with large extrapolations in time. His first comment is surprising as it seems to indicate some reservations on the cosmological principle.

The next following winter (1931-1932), Einstein was back in Caltech for a second stay where he met De Sitter. Together they formulated a model today known as the Einstein-de Sitter universe, assuming a flat space with no cosmological constant and with an expansion velocity asymptotically approaching zero in the infinite future.[32]

This became the standard model up to the mid-1990s. This was the Einstein's last intent to produce a cosmological model.

Acknowledgments

JLCC acknowledges support by CONACyT project 283151.

References

1. A. Einstein (1917). Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. Sitz. König. Preuss. Akad. 142-152. Or 'Cosmological considerations in the general theory of relativity' CPAE 6 (Doc. 43).
2. A. Einstein (1919). Spielen Gravitationsfelder im Aufbau der materiellen Elementarteilchen eine wesentliche Rolle? Sitz. König. Preuss. Akad. 349-356. Or 'Do gravitation fields play an essential part in the structure of the elementary particles of matter?' CPAE 7 (doc.17).
3. A. Einstein (1931). Sitzungsber. Preuss. Akad. Wiss. 235-37.
4. C. O'Raifeartaigh (2014). Einstein's steady-state cosmology. Physics World, 27(09), 30.
5. C. O'Raifeartaigh and B. McCann, (2014). Einstein's cosmic model of 1931 revisited: An analysis and translation of a forgotten model of the universe. Eur. Phys. J. H 39 (2014) 63-85, e-Print: 1312.2192 [physics.hist-ph].
6. H. Nussbaumer (2014). Einstein's conversion from his static to an expanding universe. EPJ-H, Eur. Phys. J. H. Vol. 39, 37-62, DOI: 10.1140/epjh/e2013-40037-6. http://arxiv.org/abs/1311.2763.
7. H. Nussbaumer (2018). Einstein's aborted attempt at a dynamic steady-state universe. Acta Historica Astronomiae, 64, 463-477.
8. New York Times, p. 1 (Jan. 3, 1931).
9. New York Times, p. 1 and 17 (Feb. 5, 1931).
10. A. C. Balch, Professor Einstein at the California Institute of Technology. Science, 73, 375-379 (1931).
11. W. S. Adams (1925). The relativity displacement of the spectral lines in the companion of Sirius. Proceedings of the National Academy of Sciences of the United States of America, 11(7), 382.
12. R. A. Millikan and W. S. Adams, The reason and the results of Dr. Einstein's visit to the California Institute of Technology. Science 73380-381, (1931).
13. C. O'Raifeartaigh, B. McCann, W. Nahm, and S. Mitton (2014). Einstein's exploration of a steady-state model of the universe. The European Physical Journal H, 39(3), 353-367.
14. J. D. Burchfield (1990). Lord Kelvin and the Age of the Earth. University of Chicago Press.
15. H. Hellman and H. Hellman (1998). Great feuds in science: Ten of the liveliest disputes ever. New York: Wiley.
16. C. Lewis (2002). The Dating Game: One man's search for the Age of the Earth. Cambridge University Press.
17. L. Badash (1968). Rutherford, Boltwood, and the age of the earth: The origin of radioactive dating techniques.
18. The Age of the Earth. Bulletin of the pp. 487. Washington, D.C. (1931). Geological Magazine, 68(11), 523-525. doi:10.1017/S0016756800098459 academy of science.

19. W. Baade (1944). The resolution of Messier 32, NGC 205, and the central region of the Andromeda nebula. ApJ 100 137-146.
20. F. Zwicky (1929). "On the Redshift of Spectral Lines Through Interstellar Space". Proceedings of the National Academy of Sciences. 15 (10): 773-779.
21. W. D. MacMillan (1932). Velocities of the spiral nebulae. Nature, 129(3246), 93-93.
22. W. Nernst (1928). Physico-chemical considerations in astrophysics. Journal of the Franklin Institute, 206(2), 135-142.
23. W. Nerst (1928) ibid ref.[22], p 141.
24. R. C. Tolman (1930). On the estimation of distances in a curved universe with a non-static line element. Proceedings of the National Academy of Sciences of the United States of America, 16(7), 511.
25. E. Hubble and R. C. Tolman (1935). Two methods of investigating the nature of the nebular redshift. The Astrophysical Journal, 82, 302.
26. R. A. Millikan and G. H. Cameron (1931). A more accurate and more extended cosmic-ray ionization-depth curve, and the present evidence for atom-building. Physical Review, 37(3), 235.
27. R. A. Millikan and G. H. Cameron (1926). High frequency rays of cosmic origin III. Measurements in snow-fed lakes at high altitudes. Physical Review, 28(5), 851.
28. R. A. Millikan (1930). Remarks on the history of cosmic radiation. Science, 71(1851), 640-641.
29. W. D. MacMillan (1918). On stellar evolution. The astrophysical journal, 48, 35.
30. R. A. Millikan and G. H. Cameron (1928). The origin of the cosmic rays. Physical review, 32(4), 533.
31. D. J. Kevles (1979). Robert A. Millikan. Scientific American, 240(1), 118-125. p 122.
32. A. Einstein and W. de Sitter (1932). On the relation between the expansion and the mean density of the Universe. PNAS, 18, 213-214.

Jayme Tiomno: Relativity, gravity, cosmology, and the Marcel Grossmann Meetings

William D. Brewer

Fachbereich Physik, Freie Universität Berlin,
14195 Berlin, F.R. Germany
E-mail: william.brewer@fu-berlin.de
www.users.physik.fu-berlin.de/∼wbrewer/index.html

Jayme Tiomno belonged to the 'founder's generation' of physicists in Brazil. He began working in relativity theory early in his career, at a time when it was not at all 'fashionable', through the influence of his early mentor, Mario Schenberg in São Paulo. When he went to graduate school in Princeton, in February 1948, his advisor there, John Wheeler, gave him a project in General Relativity, even though this was more than 4 years before Wheeler's 'turn' from nuclear and particle physics to field theory and gravitation.

Tiomno and Wheeler however soon discovered their mutual interest in meson decays, and Tiomno's Masters and PhD theses were on topics from particle physics, which remained his major field of interest for the following 20 years, during which he collaborated with Abdus Salam, among others. Only when he returned to Princeton in 1971, a refugee from the oppressive dictatorship in Brazil, did he again begin working in gravitation and field theory, having missed the 'golden age' initiated in part by Wheeler's group.

At the IAS, Tiomno experienced a renaissance of his interest in field theory, working with Remo Ruffini and others. He continued this work in the 1980's after he was able to return to the CBPF in Rio de Janeiro (which he had helped to found). His participation in the Marcel Grossmann Meetings was limited but significant.

Keywords: Physics Brazil; History of Physics; Marcel Grossmann Meetings; General Relativity; Gravitation

1. Introduction

Jayme Tiomno was born in Rio de Janeiro on April 16, 1920, the son of Russian-Jewish immigrants who had arrived in Brazil about 10 years earlier. There was nothing in his background that might suggest that he would become an eminent theoretical physicist. But his parents both valued education and saw to it that he and his siblings (an older brother, a twin sister, and two younger sisters) were all able to attend college. In the year of his birth, on Sept. 7, the Brazilian national holiday, the *Universidade do Rio de Janeiro* (URJ) was founded as the first real university in Brazil (as a collection of faculties and institutes combined from existing institutions).

After his childhood and youth spent in small cities in Minas Gerais, Tiomno's family returned in 1934 to Rio, where Jayme completed his *curso complementar* in medicine and then began studying medicine in 1938 at the *Universidade do Brazil* (UB – as the URJ was now called). In early 1939, Jayme was enrolled by his brother

for Natural History at the *Universidade do Distrito Federal* (UDF), a newer research university founded in the capital in 1935 by Anisio Teixeira, an eminent educator who was at the time Director of Education of the Federal District.

Fig. 1. (a) Jayme Tiomno, 1941. (b) Tiomno, Wheeler, Ruffini, 2001 (c) Mario Schenberg, 1937.

Tiomno qualified for the physics course, but was unable to study at the UDF because it was closed by the Vargas regime. Its physics curriculum was incorporated into the *Faculdade Nacional de Filosofia* (FNFi) and moved to the UB. Tiomno later had to decide between medicine and physics, and he chose the latter, graduating in 1941. Figure 1 shows Tiomno at age 21, when he graduated from the FNFi, and also 60 years later, with his Princeton mentor John Wheeler and his later collaborator Remo Ruffini, at Wheeler's 90th birthday in 2001; and Tiomno's first graduate mentor, Mario Schenberg, about the time Tiomno began his university career.

Tiomno entered military service, and was able to do some research with his mentor, the experimentalist Joaquim da Costa Ribeiro, and complete his *licenciatura* (teaching certificate) during the War. However, his real interest lay in theory, inspired by his Italian professor of theoretical physics, Luigi Sobrero (a former assistant of Levi-Civitá), and after the War, he obtained a fellowship to do graduate research at the Universidade de São Paulo (USP), in the department founded by Gleb Wataghin, under Wataghin's star pupil, the theoretician Mario Schenberg. The photo in Fig. 2 shows Tiomno's professors Sobrero and Costa Ribeiro as well as the teaching assistants in physics at the FNFi in May, 1942. Many of them later became noted scientists.

Tiomno spent most of 1946 at USP, studying hard to catch up in modern physics, to which he had hardly been exposed at the FNFi. After his fellowship ended, he returned to Rio in early 1947, but was then offered an assistantship with Mario Schenberg at USP, where he went in mid-year of 1947. He worked with Schenberg on formulating gravitation in Minkowski space, his introduction to relativity. The project was completed but never published, partly because relativity, and especially non-Einsteinian interpretations of gravitation, were hard to publish in 1948, and

Fig. 2. The teaching staff in physics at the FNFi, 1942. From left: Paulo Alcântara Gomes, Elisa Frota-Pessôa, Jayme Tiomno, Joaquim da Costa Ribeiro, Luigi Sobrero, Leopoldo Nachbin, José Leite Lopes, and Mauricio Matos Peixoto.

partly due to Schenberg's political difficulties, which forced him to go into exile in Europe in the following years.

Tiomno, in the meantime, on the recommendations of Schenberg and José Leite Lopes, was granted a fellowship for graduate work at Princeton by the U.S. State Department. He had to leave Brazil abruptly in early February 1948 to meet the starting deadline for his fellowship. In Princeton, he joined the group of John Archibald Wheeler, already a well-known theoretician, working in nuclear and particle physics. Nevertheless, Wheeler started Tiomno on a project involving point particles in GR (and eventually their gravitational-radiative damping). Tiomno was frustrated with the project and received little help from Wheeler (and from Einstein). Wheeler had already suggested this topic to Leopold Infeld, who in 1949 published the result with his assistant Alfred Schild.[1] By then, Tiomno and Wheeler had turned their interest to the pion-muon decay chain, and Tiomno spent the next 20 years working mainly in particle physics, apart from two collaborative projects with David Bohm – one a conformally-invariant formulation of the Dirac theory, which Tiomno only completed 12 years later; the other the inclusion of spin into Bohm's deterministic version of quantum mechanics (1955).

Fig. 3. John Wheeler and Jayme Tiomno in Princeton, Summer 1948.

When Wheeler (Fig. 3) left on sabbatical in June, 1949, Tiomno, with his fresh MSc degree, collaborated with Chen Ning Yang and David Bohm on weak-interaction physics and on relativistic quantum mechanics. He then began his PhD thesis work with Eugene Wigner as mentor. He finished his thesis in September 1950 and returned to São Paulo on October 10.

While still at Princeton, Tiomno also met Richard P. Feynman and Cécile Morette, and invited both to visit the newly-founded *Centro Brasileiro de Pesquisas Físicas* (CBPF) in Rio, which he had helped to plan and implement. He later also encouraged Bohm to apply (successfully) for the vacant chair of Advanced Physics in São Paulo (vacated by Gleb Wataghin when he returned to Italy in 1949). The young David Bohm is shown in Fig. 4, about the time that he completed his undergraduate studies.

Fig. 4. David Bohm, around 1940.

Figure 5 is a photo of a meeting at Princeton in early 1949, where several young Brazilian physicists conferred about the newly-founded CBPF in Rio, and met the Japanese theoretician Hideki Yukawa, on sabbatical in Princeton at the time. More information on Tiomno's life and career can be found in our recent biography.[2]

2. The development of gravitation, General Relativity and cosmology research after 1950

Modern gravitational physics dates from the early 20th century. Lorentz, Poincaré, Fitzgerald and others had speculated about applying the principle of relativity to electromagnetism (Maxwell theory). Albert Einstein realized the significance of those speculations and formulated his Special Relativity in 1905. Ten years later, with the help of Marcel Grossmann, he completed his General Theory of Relativity (GR), the modern theory of gravitation. After the early period from 1915–1925, when solutions to Einstein's field equations were found (e.g. by K. Schwarzschild, 1916: gravitational singularity; A. Einstein, 1917: closed, static universe; W. de Sitter, 1917: zero-density universe; A. Friedmann, G. Lemâitre, 1924/27: expanding

Fig. 5. A group of young Brazilians, together with Hideki Yukawa (behind, center), in Princeton, early 1949. César Lattes (behind, left), José Leite Lopes (below, center) and Jayme Tiomno (below, right) were the chief founders of the CBPF (Brazilian Center for Physics Research, Rio de Janeiro), which began operations about this time. At the upper right is Walter Schützer, and at lower left is Hervasio de Carvalho.

universe), Einstein began working on a unified theory of gravity and electromagnetism, and active work on GR decreased after about 1928. Some of the early contributors to GR theory are shown in Fig. 6.

In the period 1935–1955, Einstein published with N. Rosen, L. Infeld, B. Hoffmann, P. Bergmann, W. Pauli, V. Bargmann, and E.G. Strauss on some details of GR theory, and some other authors – notably J.R. Oppenheimer et al. (1939: collapse of stellar remnants) and K. Gödel (1949: stationary rotating homogeneous universe with timelike loops) also made important contributions. But only in the early 1950's did the 'great revival' of GR and its applications to relativistic astrophysics and cosmology begin, stimulated initially by J.A. Wheeler and his school (Princeton, University of Maryland, Caltech), and independently by Bryce DeWitt and Cécile Morette DeWitt (University of North Carolina, University of Texas), and in England by D.W. Sciama, C. Isham, R. Penrose, and S. Hawking (in London and Cambridge), and in the USSR by Y. Zel'dovitch, I. Novikov, and R. Sunyaev.

Fig. 6. (a) Karl Schwarzschild, ca. 1910. (b) Willem de Sitter. (c) Georges Lemâitre.

Tiomno's interest in field theory and GR was rekindled in 1966/67, when he spent a year as Associate of the ICTP (International Centre for Theoretical Physics) in Trieste (founded in 1964 by Abdus Salam, with Tiomno's active participation). While there, he reconnected with two Argentine physicists whom he had known earlier, in Rio and in London: Juan José Giambiagi and Carlos Guido Bollini. Their joint papers, especially 'A Linear Theory of Gravitation' (1970), mark Tiomno's revived interest in gravitation theory. Giambiagi, Tiomno and Bollini are shown in Fig. 7 while at the ICTP in 1966/67, and also shown are Jayme Tiomno with his wife Elisa Frota-Pessôa in Trieste at the same time. That interest was intensified during his stay in Princeton in 1971/72, after his blacklisting in Brazil by the military dictatorship. There, he interacted with Wheeler's now flourishing school. In the years 1972–75, he published 13 articles on gravitation and field theory. In addition to his two more senior collaborators, Remo Ruffini and Leonard Parker, his coauthors included the younger physicists C.V. Vishveshwara, Marc Davis, Jeffrey M. Cohen, Frank J. Zerilli, Robert M. Wald, and Reinhard A. Breuer. Many of those publications dealt with radiation – both electromagnetic and gravitational – from particles orbiting or falling into black holes; but they also included speculations on pulsars and the balance of forces on particles near black holes, as well as the possibility and properties of gravitational 'synchrotron radiation'. Figure 8 shows Tiomno and Elisa in Princeton in 1971, and also Tiomno's farewell letter to Wheeler upon leaving Princeton in June, 1972.

The text of Tiomno's letter to Wheeler when he left Princeton reads: *"We are leaving tomorrow, looking forward to a vague possibility of remaining in Rio. We were much pleased with our stay in Princeton. I have profited much from it and started a new phase of my career. Even the disagreeable dispute with a colleague has had a positive result to stress my friendship with you and to make me recover from the difficulty of writing papers. Here are some of them, which would not exist*

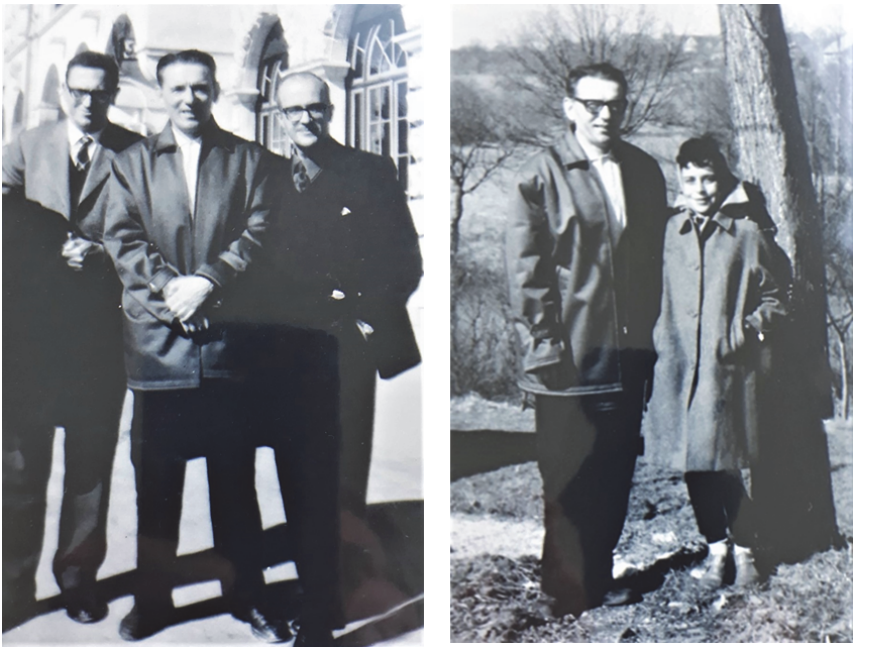
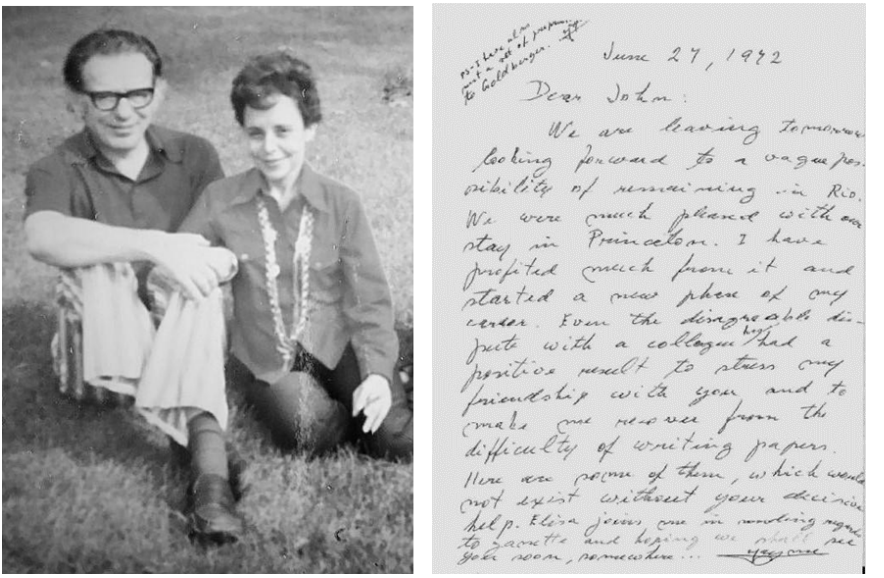


Fig. 7. (a) J.J. Giambiagi, J. Tiomno, C.G. Bollini, ICTP 1966. (b) Jayme Tiomno and Elisa Frota-Pessôa, ICTP 1966.

without your decisive help. Elisa joins me in sending regards to Janette and hoping we shall see you soon, somewhere . . . Jayme. PS. I have also sent a set of preprints to Goldberger."

We can only speculate on the precise details of the 'dispute' mentioned; it was probably a conflict over the publication of an article written jointly by Breuer, Ruffini, Tiomno and Vishveshwara (which appeared later in Phys. Rev. D. See Ref. 1 for more details).

After returning to Brazil in mid-1972, Jayme and Elisa were confronted with their continued blacklisting by the military dictatorship, which considered them to be 'subversive' (a charge which verges on the ridiculous, in retrospect). They could not even work at the private foundation CBPF, which they had helped to found and establish. Finally (after a Papal dispensation), they were given the opportunity to work at the Catholic University (PUC) in Rio de Janeiro, albeit with limited privileges. They remained there until after the Amnesty in 1979, and then still refused to beg for reinstatement to their university positions. Jayme finally rejoined the 'reformed' CBPF at the end of 1980; it was now an independent laboratory of the national science research funding agency CNPq. There, he became an honored and very active member of the faculty, retiring as 'Researcher emeritus' in 1992. For more information on the development of relativity and gravitation research after their 'rennaissance' in the 1950's, see the references.[3–5]



June 27, 1972

Dear John:

We are leaving tomorrow looking forward to a vague possibility of remaining in Rio. We were much pleased with our stay in Princeton. I have profited much from it and started a new phase of my career. Even the disagreeable dispute with a colleague had a positive result to strain my friendship with you and to make me recover from the difficulty of writing papers. Here are some of them, which would not exist without your decisive help. Elisa joins me in sending regards to Jeanette and hoping we shall see you soon, somewhere ... Yours

Fig. 8. (a) Tiomno and Elisa, Princeton 1971. (b) Tiomno's 'farewell letter' to Wheeler, June 1972.

3. Tiomno's Scientific Work

Like Wheeler's scientific career, Tiomno's lifework can be divided into three distinct periods, although not so sharply:

– The first was his early work on relativity and particle physics, which extended from 1947 through about 1967. Highlights of this period were his work on the Weak Interaction (initially called the 'Universal Fermi Interaction', a term coined by Tiomno and C.N. Yang in 1949); his work on S-matrix theory and causality (with Walter Schützer; see Fig. 5); his prediction of the K* meson (with N. Zagury and A.L.L. Videira); and work on hyperons and global symmetry (with Abdus Salam).
– The second period began in the late 1960's and extended to around 1981. It was characterized by a shift in interest toward field theory, gravitation and cosmology, and was stimulated by collaborations with Giambiagi and Bollini as well as with his Princeton colleagues in 1971/72.
– The third period might be called his "eclectic phase", after he had returned to the CBPF in 1980, when he was involved in numerous projects relating to particle physics, Special Relativity, General Relativity, field theory, gravitation and cosmology, ending with his interest in relativistic rotating systems in the mid-1990's.

In the following, we give some selected examples of Tiomno's publications from the 1970's to the mid-1980's, as an illustration of the variety and the relevance of his work in his second and third periods.

3.1. *Some Selected Publications, 1970-1985*

(1) 'A Linear Theory of Gravitation', in: Lettere al Nuovo Cimento **3**, 65 (1970) (with C.G. Bollini and J.J. Giambiagi).
(2) 'Equivalence of Lorentz Transformations and Foldy-Wouthuysen Transformation for Free Spinor Fields', in: Physica **53**, 581–601 (1971).
(3) 'Electromagnetic Field of a Particle Moving in a Spherically-Symmetric Black-Hole Background', in: Lettere al Nuovo Cimento **3**, 211 (1972) (with R. Ruffini and C.V. Vishveshwara).
(4) 'Can Synchrotron Gravitational Radiation Exist?', in: Physical Review Letters **28**, 1352 (1972) (with Marc Davis, R. Ruffini and F. Zerilli).
(5) 'Pulses of Gravitational Radiation of a Particle Falling Radially into a Schwarzschild Black Hole', in: the Physical Review **D5**, 2932 (1972) (with Marc Davis and R. Ruffini).
(6) 'Maxwell Equations in a Spherically Symmetric Black-Hole Background', in: Lettere al Nuovo Cimento **5**, 851–855 (1972).
(7) 'Pair-Producing Electric Fields and Pulsars', in: the Astrophysical Journal **178**, 809 (1972) (with L. Parker).
(8) 'Polarization of Gravitational Synchrotron Radiation', in: Lettere al Nuovo Cimento **4**, 857 (1972) (with R.A. Breuer and C.V. Vishveshwara).
(9) 'Balancing of Electromagnetic and Gravitational Forces and Torque Between Spinning Particles at Rest', in: the Physical Review **D7**, 356 (1973).
(10) 'Gyromagnetic Ratio of a Massive Body', in: the Physical Review **D7**, 998 (1973) (with J.M. Cohen and R.M. Wald).
(11) 'Electromagnetic Field of Rotating Charged Bodies', in: the Physical Review **D7**, 992 (1973).
(12) 'Charged point particles with magnetic moments in General Relativity', in: Revista Brasileira de Física **8**, 350 (1978) (with Ricardo M. Amorim).
(13) 'On the relation between fields and potentials in non-abelian gauge fields', in: Revista Brasileira de Física **9**, 1 (1979) (with J.J. Giambiagi and C.G. Bollini).
(14) 'Singular potentials and analytic regularizations in classical Yang-Mills Equations', in: Journal of Math. Phys. **20**, 1967 (1979) (with C.G. Bollini and J.J. Giambiagi).
(15) 'Gauge field copies', in: Phys. Letters **83B**, 185 (1979) (with J.J. Giambiagi and C.G. Bollini).
(16) 'Wilson Loops and Related Strings for the instanton and its variational derivatives', in: Il Nuovo Cimento **59**, 412 (1980) (with C.G. Bollini and J.J. Giambiagi).
(17) 'Wilson Loops in Kerr Gravitation', in: Letters to Il Nuovo Cimento **81**, 13 (1981) (with C.G. Bollini and J.J. Giambiagi).
(18) 'Geodesic Motion and Confinement in Gödel's Universe', in: the Physical Review **D27**, 779 (1983) (with M. Novello and I.D. Soares).

(19) 'Gödel-type metric in Einstein-Cartan Spaces', in: Contributed Papers to the 10th International Conference on General Relativity and Gravitation, p. 507 (1983) (with A. Teixeira and J. Duarte).
(20) 'Homogeneity of Riemannian Spacetimes of Gödel type', in: the Physical Review **D28**, 1251 (1983) (with M. Rebouças).
(21) 'Pseudoscalar Mesons and Scalar diquarks—decay constants', in: Il Nuovo Cimento **81A**, 485 (1984) (with I. Bediaga, E. Predazzi, A.F.S. Santoro and M.H.G. Souza).
(22) 'Gravitational coupling of scalar and fermionic fields to matter vorticity: Microscopic asymmetries', in: Revista Brasileira de Física, Suppl. **14**, 372 (1984) (with I.D. Soares).
(23) 'Lifetimes in a quark-diquark system', in: Lettere al Nuovo Cimento **42**, 92–96 (1985) (with A.F.S. Santoro, I. Bediaga, M.G.H. Souza and E. Predazzi).
(24) 'On Experiments to detect possible failures of Relativity Theory', in: Foundations of Physics **15**, No. 9, pp. 945–961 (1985) (with W. Rodrigues).
(25) 'Gluon and qq mixing: [1440] system', in: Zeitschrift f. Physik **C30**, 493 (1985) (with A.C.B. Antunes, F. Caruso and E. Predazzi).
(26) 'Experiments to Detect Possible Weak Violations of Special Relativity', in: Physical Review Letters **55**, 143 (1985) (with A.K.A. Maciel).

Of special note are the publications that Tiomno himself listed as particularly important when he was nominated in 1995 for the Physics Prize of the Third-World Academy of Sciences (TWAS - an organization founded by Abdus Salam to complement the work of the ICTP). These include Items 1, 4, 9, 14, 16, 19, and 26 in the above list – and they illustrate how his interest in gravitation and field theory developed over the years.

4. ICRA and the MG Meetings. MG-X

The first international conference series to be initiated for the newly-revived field of General Relativity and Gravitation carried exactly that name – GR-1 was held at Chapel Hill, NC/USA in January 1957. It was organized by Bryce and Cécile (Morette) DeWitt, who had obtained positions at the University of North Carolina. In fact, it had a predecessor, a conference organized in Bern, Switzerland and held in July 1955 to celebrate the 50th anniversary of the initial publication of (Special) Relativity theory. It is often called 'GR-0', while later conferences in the series are listed as 'GR-1' etc. GR-2 took place in a former palace in Royaumont, near Paris, a location that was suggested by Cécile Morette DeWitt. Tiomno attended it, since he was in Europe at the end of his sabbatical year in London, 1958/59. This conference series continues today, now at three-year intervals, and GR-23 is scheduled for July 2022 in Beijing. Figure 9 shows Tiomno and Elisa in Paris in June 1959, at the end of their London stay, when he attended the Royaumont conference (GR-2). Tiomno also later published a paper in the proceedings of GR-10 (Item 19 in the above list).

Fig. 9. Jayme Tiomno and Elisa Frota-Pessôa in Paris, June 1959 (at the west end of the *Pont d'Iena*).

Another conference series was started in Dallas, TX/USA in 1963. It is called the 'Texas Symposia on Relativistic Astrophysics', and its initial motivation was to consider the newly-discovered Quasars in the light of relativity theory. This was the first time that the term 'Relativistic Astrophysics' was used publicly. An amusing account of the genesis of these conferences, which continue today at (usually) two-year intervals (the 31st is planned for December 2021 in Prague), was given by Engelbert L. Schucking, one of the original organizers from 1963, in *Physics Today* (August 1989, pp. 46-52).

In 1975, two of Tiomno's former collaborators, Abdus Salam and Remo Ruffini, founded the conference series 'Marcel Grossmann Meetings (on Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories)', which should be familiar to the audience of this talk. Ruffini was at the time still officially at the Institute for Advanced Study in Princeton, but was on leave in 1975 as a visitor to the University of Western Australia (Nedlands)

and the University of Kyoto in Japan. The following year, he accepted a faculty position at the University of Catania, Sicily. Abdus Salam had been Head of the Department of Theoretical Physics at Imperial College London since 1957 (where Tiomno was one of the first visiting scientists, in the academic year 1958/59), and he became Director of the ICTP Trieste after its founding in 1964. The Centre offered a good platform for hosting the MG conferences, and the first two were held there in 1975 and 1979. MG-III, held in Shanghai in 1982, established the three-year interval still continued today. It was the result of the friendship between Remo Ruffini, who became Professor of Physics at his *alma mater*, the University of Rome 'La Sapienza' in 1978, and a Chinese visitor, Fang Li Zhi, whom he hosted in the late 1970's (cf. MG-XVI Awards Ceremony, Part 2). Later conferences in the series have been held in many cities on most of the continents.

Given his association with their founders and his own work in field theory, gravitation and cosmology, it was natural that Jayme Tiomno would have an interest in those meetings. But Tiomno, in spite of his close connection with the two founders of the MG conference series, did not attend the earlier MG meetings, both in Trieste; they took place during the 'leaden years' in Brazil, when he may not have been free to travel, and he had also not done any new, relevant research since leaving Princeton in 1972. By the time of MG-III, held in Shanghai/China in 1982, Remo Ruffini invited Mario Novello, from Tiomno's DRP Department at the CBPF, to participate in the meeting. Tiomno himself did not however attend the MG-III meeting; it was far from Brazil and perhaps difficult for him to obtain a visa.

Fig. 10. (a) Tiomno, Elisa and Mario Novello, 1991. (b) Tiomno with his last collaborator, Ivano Soares, in 1997.

In 1985, Ruffini and Abdus Salam, together with several other prominent scientists, founded the ICRA organization (International Center for Relativistic Astrophysics), based in Rome and Pescara, Italy (ICRANet). By 1985, Tiomno's time

had come, and he gave an invited paper on the LET–SRT controversy at MG-IV, held that year for the first time in Rome. Remo Ruffini and Abdus Salam were no doubt pleased to give him the opportunity to experience an MG conference, and Novello's enthusiasm also played an important role in convincing him to attend. After the meeting in Rome, Tiomno resumed his contacts with his former collaborators, and in 1994, the year of MG-VII, held in Stanford/CA, USA, he was a member of the 'Committee of the Americas', part of the international coordinating committee for the conference; but he again did not attend the meeting itself—by then, as *researcher emeritus*, he was reducing his overseas travel and participation in conferences, and encouraging some of his former students and collaborators to actively take over those functions, including Mario Novello and Marcelo Rebouças, among others.

Prior to the MG-X Meeting planned for 2003, the Brazilian physicists offered to host the conference at the CBPF, and Mario Novello served as chairman of its local organizing committee. At that meeting, the Institutional Award, presented at each conference to an outstanding institution in research and teaching, was given to the CBPF 'for its role as a teaching and research institution and as a place originating fundamental physics ideas in the exploration of the universe', and presented to its founders, César Lattes, José Leite Lopes, and Jayme Tiomno, who were all present.

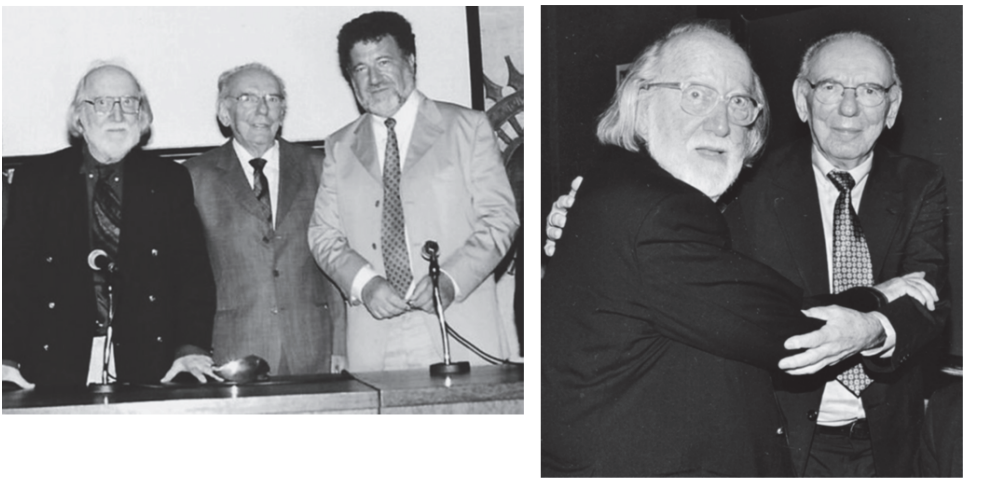


Fig. 11. (a) José Leite Lopes, Tiomno, and Remo Ruffini at the prize ceremony, MG-X, 2003. (b) Leite Lopes and Tiomno at MG-X in Rio, 2003.

That same year, ICRANet, the international networking arm of ICRA, was officially chartered. Ruffini needed at least three states to create it, and at the time, he had just two: the Vatican and Armenia. Brazil's membership in ICRANet was

also prepared at the 2003 Meeting, and Mario Novello later became its representative for Brazil (a post now occupied by Ulisses Barres de Almeida of the CBPF). Brazil formally agreed to join in 2005, and its membership was finalized in 2011.

Thus, although Jayme Tiomno himself was not strongly engaged in these meetings, his influence continues to be felt in them through the *Centro* that he helped to found, and through his former students and collaborators, who are actively involved in them, both scientifically and organizationally.

5. Conclusions

The three major strands of fundamental physics in the 20th century have wound their way through the decades, producing many offshoots and sometimes seeming to fuse into a single strand. They each experienced periods of hectic activity and rapid progress, and longer periods of relative neglect and apparent stagnation.

Particle physics reached its preliminary climax in the late 1970's with the formulation of the Standard Model of Particle Physics (combining the electroweak unified theory with quantum chromodynamics to describe three of the four fundamental interactions (weak, strong, electromagnetic); however leaving gravity out of the picture). In the intervening 40 years, the SMPP has been completed and perfected, but although no one really believes it to be the 'last word' on the microscopic world, the predicted 'new physics' beyond the SMPP has failed to appear, both experimentally and theoretically.

Gravitational physics has also produced the Standard Model of Cosmology (based on GR), and it has been enriched by many new discoveries and technical advances. But the SMC leaves many open questions and unsolved problems: dark matter, dark energy, inflation; and its fundamental constant, the Hubble constant, is currently the subject of conflicting measured values.

Fundamental quantum mechanics has given rise to a number of important offshoots: quantum optics, quantum information, quantum computing; but the interpretation of quantum mechanics is still elusive, as is its combination with gravity – quantum gravity.

Jayme Tiomno contributed significantly to all three of those major strands. But also, and for him more importantly, he contributed to the establishment and improvement of physics education and research in his native country, Brazil. This was his most serious goal during his long lifetime, and he made major contributions to its present success. The lesson to be learned from his lifelong efforts is that serious and honest work often leads to unexpected and lasting successes, even though it faces drastic obstacles and may appear hopeless at the time.

Acknowledgments

The author thanks Prof. Alfredo Tiomno Tolmasquim for much effort in the preparation of our joint book, reference,[2] upon which much of this article is based. Readers should consult it for many more references to original sources. Thanks are due also to Prof. Luis Carlos Bassalo Crispino, who was chairman of the session of MG-XVI where this paper was presented.

References

1. Leopold infeld and Alfred Schild, 'On the Motion of Test Particles in General Relativity', Rev. Mod. Phys. **21**, 408 (1949).
2. '*Jayme Tiomno: A Life for Science, A Life for Brazil*', William Dean Brewer and Alfredo Tiomno Tolmasquim, Springer/Nature, Cham/CH, June 2020; ISBN 978-3-030-41010-0.
3. Alexander Blum, Roberto Lalli, and Jürgen Renn, *Isis* **106**, No. 3, 598-620 (2015).
4. Alexander Blum and Dieter Brill, *ArXiv*:1905.05988v1 (2019).
5. '*Building the General Relativity and Gravitation Community During the Cold War*', Roberto Lalli, Springer, 2017; ISBN 973-3-319-54654-4.

A look inside Feynman's route to gravitation

M. Di Mauro

Dipartimento di Matematica, Università di Salerno, Via Giovanni Paolo II, Fisciano (SA), 84084, Italy
E-mail: madimauro@unisa.it

S. Esposito* and A. Naddeo†

INFN, Sezione di Napoli, C. U. Monte S. Angelo, Via Cinthia, Napoli, 80125, Italy
**E-mail: sesposit@na.infn.it*
†E-mail: anaddeo@na.infn.it

In this contribution we report about Feynman's approach to gravitation, starting from the records of his interventions at the Chapel Hill Conference of 1957. As well known, Feynman was concerned about the relation of gravitation with the rest of physics. Probably for this reason, he promoted an unusual, field theoretical approach to general relativity, in which, after the recognition that the interaction must be mediated by the quanta of a massless spin-2 field, Einstein's field equations should follow from the general properties of Lorentz-invariant quantum field theory, plus self-consistency requirements. Quantum corrections would then be included by considering loop diagrams. These ideas were further developed by Feynman in his famous lectures on gravitation, delivered at Caltech in 1962-63, and in a handful of published papers, where he also introduced some field theoretical tools which were soon recognized to be of general interest, such as ghosts and the tree theorem. Some original pieces of Feynman's work on gravity are also present in a set of unpublished lectures delivered at Hughes Aircraft Company in 1966-67 and devoted primarily to astrophysics and cosmology. Some comments on the relation between Feynman's approach to gravity and his ideas on the quantum foundations of the fundamental interactions are included.

Keywords: Gravitation; Quantum field theory; Loop diagrams.

1. Introduction: A timeline

Among the many scientific interests that Feynman had in the 1950s and in the 1960s, a prominent place was taken by the understanding of the relation of gravitation to the rest of physics,[a] and in particular the assessment of its consistency with the uncertainty principle.[b] Feynman likely began to seriously think about gravity in

[a] "Next we shall discuss the possible relation of gravitation to other forces. There is no explanation of gravitation in terms of other forces at the present time. It is not an aspect of electricity or anything like that, so we have no explanation. However, gravitation and other forces are very similar, and it is interesting to note analogies." (Ref. 1, Vol. 1, Sec. 7-7); "My subject is the quantum theory of gravitation. My interest in it is primarily in the relation of one part of nature to another." (Ref. 2, p. 697).

[b] "...it would be important to see whether Newton's law modified to Einstein's law can be further modified to be consistent with the uncertainty principle. This last modification has not yet been completed." (Ref. 1, Vol. 1, Sec. 7-8).

the early 1950s, just after completing his work on quantum electrodynamics. This was attested by Murray Gell-Mann[3] in a paper written for a Physics Today special issue devoted to Feynman's legacy,[c] as well as by Bryce S. DeWitt,[4] in a letter to Agnew Bahnson written in November 1955.[5] Feynman's efforts began in a period in which general relativity, after a stagnation which lasted about thirty years, was gradually emerging as a mainstream research area, giving rise to a process known as the *renaissance* of general relativity.[6]

A crucial event to consider, in order to reconstruct Feynman's approach to gravitation, is the 1957 Chapel Hill conference[4] on *The Role of Gravitation in Physics*, which was also pivotal to the renaissance of general relativity. At that conference, the gravitational physics community delineated the tracks along which subsequent work would develop in the subsequent decades. The main threads were the following[7]: classical gravity, quantum gravity, and the classical and quantum theory of measurement (as a bridge between the previous two topics). In that conference Feynman's work on gravity, of which nothing had been published yet, was presented for the first time, and put on paper in the records.[4] In fact these written records testify that by the time of the conference he had already deeply thought, and performed computations, about each of the above listed three topics, focusing in particular on classical gravitational waves, on the arguments in favor of quantum gravity from fundamental quantum mechanics, and finally on quantum gravity itself. Thus, in this contribution, we take as starting point for our reconstruction Feynman's interventions at Chapel Hill and follow the development of his work in the subsequent years, until the late sixties, where he apparently lost interest in the subject.

The ultimate goal of Feynman's work was the development of a quantum field theory of gravitation, which led him to face deep conceptual as well as mathematical issues, such as divergent integrals and the lack of unitarity beyond the tree level approximation. This task accompanied Feynman for some years, as stated in a letter[8] he wrote to Viktor F. Weisskopf in 1961 ("As you know, I am studying the problem of quantization of Einstein's General Relativity. I am still working out the details of handling divergent integrals which arise in problems in which some virtual momentum must be integrated over"), as well as reported by William R. Frazer[9] in a short summary of the talks given at the La Jolla conference, later in the same year.[d] A first comprehensive account of Feynman's results on quantum gravity can

[c]Indeed here Gell-Mann remembered his visit to Caltech during the Christmas vacation of 1954-55, the discussions with Feynman about quantum gravity, and the fact that Feynman "had made considerable progress".

[d]The International Conference on the Theory of Weak and Strong Interactions was held in June 14-16, 1961, at the University of California, San Diego, in La Jolla. Here, Geoffrey F. Chew gave his celebrated talk on the S-matrix,[10] while an afternoon session was devoted to the theory of gravitation, where Feynman reported on his work on the renormalization of the gravitational field, and recognized non-unitarity as the main difficulty, which was shared also by Yang-Mills theory.

be found in the talk he gave in 1962 at the Warsaw conference,[e] whose written version was later published as a regular paper in Acta Physica Polonica.[2] Further details were given by Feynman much later, in a couple of papers[11,12] published in 1972 in the Festschrift for John A. Wheeler's 60th birthday,[13] which were written in a period in which he was already deeply absorbed in the study of partons and strong interactions.

Among the main sources which contribute to offer a clear account of Feynman's work on gravity issues, it is mandatory to include the famous Caltech Lectures on Gravitation,[14] delivered in 1962-63 and aimed to advanced graduate students and postdocs. Finally, there is some unpublished material, included in two sets of lectures, given in the 1960s at the Hughes Aircraft Company, which only recently have been made available on the web.[f] In particular, the 1966-67 set of lectures,[16] which were devoted to astronomy, astrophysics and cosmology, contains an introductory treatment of general relativity, with an emphasis on applications to the main subjects. While sharing many similarities with the above quoted Caltech lectures,[14] the Hughes treatment offers to the attentive reader several original points. In those same years Feynman had succeeded in finding a new derivation of Maxwell's equations,[17,18] and a generalization of this approach to gravity is suggested (but not pursued) in several places in the Hughes lectures on astrophysics, as well as in the set of lectures given in the following year and devoted to electromagnetism.[19]

After outlining the main steps and sources which helped us to reconstruct the full development of Feynman's work on gravitation,[g] in this contribution we focus on two key issues: the formulation of quantum gravity as a quantum field theory of a massless spin-2 field, the graviton, in whole analogy with quantum electrodynamics, which is the content of Section 2, and the unitarity and renormalization issues arising beyond the tree level approximation, presented in Section 3. Our concluding remarks close the paper.

2. Gravity as a quantum field theory

Feynman's strategy in approaching gravity was firstly outlined at Chapel Hill conference in a series of critical comments (Ref. 4, pp. 272-276), in which a non-geometric and field theoretical line of attack is put forward. His starting point was an hypothetical, counterfactual situation, in which scientists would discover the principles of Lorentzian quantum field theories before general relativity.[h] The main concern in

[e]The International Conference on General Relativity and Gravitation was held in Jabłonna and Warsaw in July 25-31, 1962, with Leopold Infeld as the chairman of the local organizing committee. The discussion focused on three main topics: the general properties of gravitational radiation, the quantization of gravity and the exact solutions of the Einstein field equations.

[f]See Ref. 15 for a brief account of Feynman's involvement in teaching at the Hughes Aircraft Company.

[g]See Ref. 20 for a comprehensive account of this work.

[h]In fact, in Ref. 4, p. 273, Feynman said: "Instead of trying to explain the rest of physics in terms of gravity I propose to reverse the problem by changing history. Suppose Einstein never existed [...]".

such a situation would then be to include a new force, the gravitational one, in the framework of quantum field theory. This approach would be later completed in the first part of the Caltech lectures.[14] Feynman's reasoning was the following: on the basis of the general principles of quantum field theory and of experimental results it is possible to conclude that gravity, as any other force, has to be mediated by exchanges of a virtual particle, which in this case is a massless neutral spin-2 quantum, the graviton. Thus, by constructing a Lorentz invariant quantum field theory of the graviton[i] and by imposing certain consistency requirements, full general relativity should be recovered. Clearly, by following the same procedure for the spin-1 case Maxwell's equations are obtained (in this case it is much simpler, being the theory linear). Such an approach testifies his ideas about fundamental interactions as manifestations of underlying quantum theories,[18] which were expressed by him several times, for example in the Hughes Lectures[19]:

> I shall call conservative forces, those forces which can be deduced from quantum mechanics in the classical limit. As you know, Q.M. is the underpinning of Nature (Ref. 19, p. 35).

Let us describe the steps in more detail. First of all, one has to establish the spin of the mediating quantum. Both in the Caltech[14] and in the 1966-67 Hughes lectures,[16] the choice of a spin-2 mediator was justified by the observation that energy, which is the source of the gravitational force, grows with the velocity. The same observation ruled out the possibility of a spin-0 field, because the associated charge would decrease with the velocity. This result can be traced back to an old argument by Einstein (never published but recalled in Ref. 26, pp. 285-290), according to which the vertical acceleration of a body would depend on its horizontal velocity, and in particular would be zero for light, making light deflection impossible. Once the spin of the graviton is established, one can easily construct the linearized theory of the associated field, which is a massless spin-2 field. Nonlinearity then comes into play because the graviton has to couple with anything carrying energy-momentum, then also with itself, and this coupling must be universal. The resulting nonlinear theory is general relativity. General covariance and the geometric interpretation of general relativity are finally recovered as a byproduct of the gauge invariance of the theory[j]. For Feynman, this was only half of the whole story. In fact, by pushing calculations beyond tree level, quantum gravity effects would be taken into account. This was Feynman's ultimate goal, i.e. obtaining a quantum theory of gravity, which in this approach amounts to the quantization of another field.

---

[i]The linear theory for a massless spin-2 field and its massive counterpart was completely worked out by Markus Fierz and Wolfgang E. Pauli in Ref. 21, on the basis of a previous work by Paul A. M. Dirac,[22] while iterative arguments similar to Feynman's ones were later put forward by Suraj N. Gupta[23] and Robert H. Kraichnan[24] in an attempt to generate infinite nonlinear terms both in the Lagrangian and in the stress-energy tensor. See for details Ref. 25.

[j]See Ref. 14, p. 113.

Let us describe how Feynman sketched the above procedure at the Chapel Hill conference (Ref. 4, pp. 272-276). In whole analogy with electrodynamics, he wrote down the following action:

$$\int \left(\frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A^\nu}{\partial x^\mu}\right)^2 \mathrm{d}^4x + \int A_\mu j^\mu \mathrm{d}^4x + \frac{m}{2}\int \dot{z}_\mu^2 \mathrm{d}s + \frac{1}{2}\int T_{\mu\nu}h^{\mu\nu}\mathrm{d}^4x + \int (\text{second power of first derivatives of } h), \qquad (1)$$

where $h_{\mu\nu}$ is the new field associated with the graviton, i.e. a symmetric second-order tensor field, which satisfies second order linear equations of the kind:

$$h^{\mu\nu,\sigma}{}_{,\sigma} - 2\overline{h}^{\mu}{}_{\sigma}{}^{,\nu\sigma} = \overline{T}^{\mu\nu}, \qquad (2)$$

where the *bar* operation is defined on a generic second rank tensor $X_{\mu\nu}$ as:

$$\overline{X}_{\mu\nu} = \frac{1}{2}\left(X_{\mu\nu} + X_{\nu\mu}\right) - \frac{1}{2}\eta_{\mu\nu}X^\sigma{}_\sigma. \qquad (3)$$

The equations of motion for particles also follow from the above action:

$$g_{\mu\nu}\ddot{z}^\nu = -\left[\rho\sigma,\mu\right]\dot{z}^\rho\dot{z}^\sigma, \qquad (4)$$

where $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $\eta_{\mu\nu}$ is the Minkowski metric, and $[\rho\sigma,\mu]$ are the Christoffel symbols of the first kind.

However, when the field $h_{\mu\nu}$ is coupled to the matter according to Eq. (4), the corresponding stress energy tensor $T_{\mu\nu}$ does *not* obey to a conservation law, leading to a consistency problem. This happens at variance with electromagnetism, where Maxwell equations guarantee conservation of the current j^μ. Instead here $T_{\mu\nu}$ does not take into account the effect of gravity on itself, which requires nonlinearity.[k] Thus a suitable $T_{\mu\nu}$ has to be found in order to satisfy the condition $\partial_\nu T^{\mu\nu} = 0$. The solution to this consistency issue[l] can be obtained by adding to the action a nonlinear third order term in $h_{\mu\nu}$, which leads to the following equation for $T_{\mu\nu}$:

$$g_{\mu\lambda}T^{\mu\nu}{}_{,\nu} = -\left[\rho\nu,\lambda\right]T^{\rho\nu}. \qquad (5)$$

One can then go on to higher order approximations, until the procedure converges. But finding the general solution of Eq. (5) is a really difficult task in the absence of a standard procedure. Feynman's idea was to look for an expression for the $T_{\mu\nu}$, and hence for the action, that is invariant under the following infinitesimal transformation of the whole tensor field $g_{\mu\nu}$:

$$g'_{\mu\nu} = g_{\mu\nu} + g_{\mu\lambda}\frac{\partial A^\lambda}{\partial x^\nu} + g_{\nu\lambda}\frac{\partial A^\lambda}{\partial x^\mu} + A^\lambda\frac{\partial g_{\mu\nu}}{\partial x^\lambda}, \qquad (6)$$

where the 4-vector A^λ is the generator. This is a geometric transformation in a Riemannian manifold with metric, hence one can say that geometry gives the metric

[k] Feynman noticed that nonlinearity was necessary in order to explain the precession of the perihelion of Mercury (as discussed in Ref. 14, p. 75).

[l] See for details Refs. 4, p. 274 and 14, pp. 78-79.

$g_{\mu\nu}$. As such, in Feynman's approach geometry comes into play at the end and not at the beginning. By working out calculations, the full nonlinear Einstein gravitational field equations are obtained.

Lectures 3-6 of Feynman's graduate course on gravitation (Ref. 14) contain all the details of the above procedure. Interestingly, a proof is also given of the ability of this field theory based formalism to reproduce key physical effects of curved spacetime geometry. For instance, in Lecture 5 (Ref. 14, pp. 66-69) it is shown that, in the action of a scalar field, the time dilation $t \to t' = t\sqrt{1+\epsilon}$ exactly reproduces the effect of a constant weak gravitational field described by the tensor $g_{44} = 1+\epsilon$, $g_{ii} = -1$, $i = 1, 2, 3$. Incidentally such an effect plays a pivotal role in producing the right result for the precession of Mercury's perihelion. Before moving to applications, Feynman devoted some lectures to the discussion of the usual geometric approach to gravity and of its link with the field theory based approach (Ref. 14, Lectures 7-10):

> Let us try to discuss what it is that we are learning in finding out that these various approaches give the same results (Ref. 14, p. 112).

Despite advocating one approach over another, Feynman was in fact intrigued by the double nature of gravity, which has both a geometrical interpretation and a field interpretation, and in Section 8.4 of Ref. 14 he recognized how an explanation may be provided by gauge invariance. Indeed a viable procedure may be established in order to obtain the invariance of the equations of physics under space dependent variable displacements. This amounts to looking for a more general Lagrangian, to be obtained by adding to the old one new terms, involving a gravity field. The net result is a new picture of gravity, as the field corresponding to a gauge invariance with respect to displacement transformations.

Summing up, Feynman succeeded in obtaining the full nonlinear Einstein equations by means of a consistency argument applied to a Lorentz invariant quantum field theory of the graviton. According to John Preskill and Kip S. Thorne,[25] it is likely that he was completely unaware of the earlier work by Gupta[23] and Kraichnan,[24] which as mentioned had been developed along a similar line of attack, but was still incomplete. So he would have developed his approach independently, besides getting more complete results. A Lorentz-invariant field theoretical approach to gravity would have been later pursued also by Steven Weinberg,[27,28] albeit quite different from Feynman's one,[m] as well as by Stanley Deser.[29–32] Indeed Deser's approach, while being similar to Feynman's one, was more elegant and general and led to completion the whole program started with Kraichnan and Gupta. Finally, a rigorous and general analysis of the relation between spin-2 theories and general covariance was carried out by Robert M. Wald.[33]

[m] Weinberg's approach relied on the analyticity properties of graviton-graviton scattering amplitudes, and it was quite more general, since Weinberg actually proved that a quantum theory of a massless spin-2 field can only be consistent if this field universally couples to the energy-momentum of matter, hence the equivalence principle has to be obeyed.

3. Fighting with loops: The renormalization of gravity

After obtaining the Einstein-Hilbert action and the full nonlinear Einstein gravitational field equations, Feynman's efforts were mainly directed towards discussing quantum field theory issues beyond the tree level approximation, i.e. loop diagrams, unitarity and renormalization.

To the best of our knowledge, these issues were publicly addressed for the first time by Feynman in 1961 at the already mentioned La Jolla conference,[9] where nonlinearity was recognized as the very source of difficulty within both gravitation and Yang-Mills theories. Indeed the sources of the gravitational field are energy and momentum, and the gravitational field carries energy and momentum itself. In the same way, the source of a Yang-Mills field is the isotopic spin current, and the Yang-Mills field carries isotopic spin itself. This means that both the gravitational field and the Yang-Mills field are self-coupled, resulting in nonlinear field theories. A difficulty that nonlinearity brings about is the fact that loop diagrams seem to clash with unitarity.

As recalled in the introduction, Feynman's results on quantum gravity can be found in a report of the talk given at the 1962 Warsaw conference, later published in Acta Physica Polonica,[2n] with many details discussed much later in the two 1972 Wheeler Festschrift papers.[11,12]

Also in this case Feynman followed his original strategy, leaving aside quantization of space-time geometry, constructing a quantum field theory for the graviton, and working out results at different perturbative orders. Since the goal was the quantum theory, the Einstein equations and the corresponding Lagrangian were assumed as a starting point, rather than derived from scratch. The theory was coupled to a scalar field, and perturbative calculations up to the next to leading order were pursued. This implied the inclusion of loop diagrams, which make quantum corrections to enter the game. In Feynman's own words:

> I started with the Lagrangian of Einstein for the interacting field of gravity and I had to make some definition for the matter since I'm dealing with real bodies and make up my mind what the matter was made of; and then later I would check whether the results that I have depend on the specific choice or they are more powerful (Ref. 2, p. 698).

The metric was split in the following way:

$$g_{\mu\nu} = \delta_{\mu\nu} + \kappa h_{\mu\nu}; \tag{7}$$

where the Minkowski metric is here denoted by $\delta_{\mu\nu}$ and κ is a dimensionful coupling constant. Substituting (7) and expanding, the Lagrangian for gravity coupled with

[n] As pointed out by Trautman in some recently published memories (34, p. 406), the text of Feynman's plenary lecture became available too late to be included in the proceedings, therefore it was published only in 1963 as a regular paper.

a scalar field can be cast in the form:

$$L = \int \left(h_{\mu\nu,\sigma}\overline{h}_{\mu\nu,\sigma} - 2\overline{h}_{\mu\sigma,\sigma}\overline{h}_{\mu\sigma,\sigma}\right) \mathrm{d}\tau + \frac{1}{2}\int \left(\phi_{,\mu}^2 - m^2\phi^2\right) \mathrm{d}\tau \tag{8}$$
$$+ \kappa \int \left(\overline{h}_{\mu\nu}\phi_{,\mu}\phi_{,\nu} - m^2\frac{1}{2}h_{\sigma\sigma}\phi^2\right) \mathrm{d}\tau + \kappa \int (hhh)\, \mathrm{d}\tau + \kappa^2 \int (hh\phi\phi)\, \mathrm{d}\tau + ...$$

where the bar operation has been defined in (3) and a schematic notation has been adopted for the highly complex higher order terms. The first two terms are simply the free Lagrangians of the gravitational field and of matter, respectively. Before considering radiative corrections the classical solution of the problem was worked out, which involved the variation of Eq. (8) with respect to h and, then, to ϕ, giving rise to the following equations of motion with a source term:

$$h_{\mu\nu,\sigma\sigma} - \overline{h}_{\sigma\nu,\sigma\mu} - \overline{h}_{\sigma\mu,\sigma\nu} = \overline{S}_{\mu\nu}\,(h,\phi)\,, \tag{9}$$
$$\phi_{,\sigma\sigma} - m^2\phi = \chi\,(\phi,h)\,. \tag{10}$$

A close inspection revealed that Eq. (9) was singular, so that Feynman was forced to resort to the invariance of the Lagrangian under the transformation:

$$h'_{\mu\nu} = h_{\mu\nu} + 2\xi_{\mu,\nu} + 2h_{\mu\sigma}\xi_{\sigma,\nu} + \xi_\sigma h_{\mu\nu,\sigma}, \tag{11}$$

ξ_μ being arbitrary. This meant that the source $S_{\mu\nu}$ had to be divergenceless in order to make Eq. (9) consistent. Finally, by making the gauge choice $\overline{h}_{\mu\sigma,\sigma} = 0$, the law of the gravitational interaction of two systems by means of the exchange of a virtual graviton was obtained. Feynman then went on computing other processes, such as an interaction vertex coupling two particles and a graviton and the gravitational analog of gravitational Compton effect (i.e. with a graviton replacing the photon).

After these preliminary calculations, Feynman went to the next to leading order approximation, thus encountering diagrams with closed loops:

> However the next step is to take situations in which we have what we call closed loops, or rings, or circuits, in which not all momenta of the problem are defined (Ref. 2, pp. 703-704).

He realized that working out closed loop diagrams required the solution of a number of conceptual issues, and he succeeded in showing that any diagram with closed loops can be expressed in terms of sums of on shell tree diagrams, which is the content of his celebrated tree theorem (which was treated in detail in Ref. 11). Further details on the statement of the tree theorem and, in particular, on the nature of the proof for the one-loop case were given by Feynman in the discussion section (Ref. 2, pp. 714-717), while answering some related questions by DeWitt. But the main problem to face in carrying out one-loop calculations was the lack of unitarity, due to the presence of contributions arising from the unphysical longitudinal polarization states of the graviton, which did not cancel as they should. Following a suggestion

by Gell-Mann,[o] Feynman considered the simpler Yang-Mills case (his results in this case were summarized in Ref. 12) and found the same pathological behavior:

> But this disease which I discovered here is a disease which exist in other theories. So at least there is one good thing: gravity isn't alone in this difficulty. This observation that Yang-Mills was also in trouble was of very great advantage to me. [...] the Yang-Mills theory is enormously easier to compute with than the gravity theory, and therefore I continued most of my investigations on the Yang-Mills theory, with the idea, if I ever cure that one, I'll turn around and cure the other (Ref. 2, p. 707).

The solution to this issue was obtained by expressing each loop diagram as a sum of trees and then computing the trees. This worked even if the process of opening a loop by cutting a graviton line implies the replacement of a virtual graviton with a real transverse one. Finally, in order to guarantee gauge invariance the sum of the whole set of tree diagrams corresponding to a given process has to be taken.

The same results, according to Feynman, could be obtained by direct integration of the closed loop. In the last case a mass-like term has to be added to the Lagrangian to avoid singularity but at the price of breaking gauge invariance. At the same time a contribution has to be subtracted, which is obtained by making a ghost particle (with spin-1 and Fermi statistics) to go around the loop and artificially coupled to the external field. In this way both unitarity and gauge invariance would be restored. This procedure was worked out also for Yang-Mills theory, but in that case the ghost particle has spin-0.[12]

Once successfully solved the one-loop case, Feynman's efforts pointed toward a further generalization of the above procedure to two or more loops:

> Now, the next question is, what happens when there are two or more loops? Since I only got this completely straightened out a week before I came here, I haven't had time to investigate the case of 2 or more loops to my own satisfaction. The preliminary investigations that I have made do not indicate that it's going to be possible so easily gather the things into the right barrels. It's surprising, I can't understand it; when you gather the trees into processes, there seems to be some loose trees, extra trees (Ref. 2, p. 710).

But, in Feynman's words, preliminary attempts seem to suggest that novel difficulties enter the game when dealing with two or more loops, as also mentioned in the last of his published Lectures on Gravitation (Ref. 14, Lecture 16, pp. 211-212). Here, once again, he recognized in the lack of unitarity of some sums of diagrams the main source of the observed pathological behavior and pointed out that a similar

[o] "I suggested that he try the analogous problem in Yang-Mills theory, a much simpler nonlinear gauge theory than Einsteinian gravitation." (Ref. 3, p. 53).

feature was shared also by Yang-Mills theory. Finally, while hinting at the problem of finding ghost rules for high order diagrams, he argued in favor of the non-renormalizability of gravity as a consequence of these difficulties:

> I do not know whether it will be possible to develop a cure for treating the multi-ring diagrams. I suspect not – in other words, I suspect that the theory is not renormalizable. Whether it is a truly significant objection to a theory, to say that it is not renormalizable, I don't know (Ref. 14, Lecture 16, pp. 211-212).

It is not clear whether Feynman was suggesting a link between non-unitarity and non-renormalizability issues. But in any case Feynman's results played a prominent role in the development of gauge theory and quantum gravity. Feynman's rules for ghosts were later generalized to all orders by DeWitt,[35–37] while Ludvig D. Faddeev and Viktor N. Popov[38] derived them in a much simpler way, by means of functional integral quantization, setting the standard for all subsequent work in the field. In particular, DeWitt proved that Yang-Mills theory and quantum gravity are in fact unitary at two[35] and arbitrarily many loops.[36,37] However, while Yang-Mills theory was later shown to be renormalizable (cf. Refs. 39-42), gravity presented divergences which could not be renormalized (cf. Refs. 43-46), confirming Feynman's suspect. It should be mentioned that, in subsequent years, modified theories of gravity, characterized by an action quadratic in the curvature, have been put forward. Unlike ordinary general relativity, these theories are renormalizable but not unitary.[47]

It is worth mentioning that, unlike most of his contemporaries, Feynman did not think about non-renormalizability as a signature of inconsistency of a theory, as also recalled by Gell-Mann,[p] and claimed by Feynman himself in one of his last interviews, given in January 1988:

> The fact that the theory has infinities never bothered me quite so much as it bother others, because I always thought that it just meant that we've gone too far: that when we go to very short distances the world is very different; geometry, or whatever it is, is different, and it's all very subtle (48, p. 507).

In fact, within the modern view on quantum field theory, which was developed in 1970s, non-renormalizability is considered only a signature of the fact that the theory loses its validity at energies higher that a certain scale. Nevertheless, one has an *effective field theory*, which works and can be useful to make predictions under that scale (interesting historical discussions can be found in Refs. 49 and 50). This is true also for gravity.[51] But, as remembered by John P. Preskill in a recent talk

[p] "He was always very suspicious of unrenormalizability as a criterion for rejecting theories" (Ref. 3, p. 53).

(Ref. 52, slide 37), although he anticipated this view, apparently Feynman was not really at ease with it:

> I spoke to Feynman a number of times about renormalization theory during the mid-80s (I arrived at Caltech in 1981 and he died in 1988). I was surprised on a few occasions how the effective field theory viewpoint did not come naturally to him". [...] Feynman briefly discusses in his lectures on gravitation (1962) why there are no higher derivative terms in the Einstein action, saying this is the "simplest" theory, not mentioning that higher derivative terms would be suppressed by more powers of the Planck length.

As a further remark, let us notice that in the last years Feynman's tree theorem has spurred a renewed interest in researchers working in the context of advanced perturbative calculations and generalized unitarity (cf. Refs. 53-57).

4. Concluding remarks

In this paper we focused on Feynman's contributions to the research in quantum gravity, starting from his interventions at Chapel Hill conference in 1957 and ending with the Wheeler festschrift papers. His approach was field theoretical rather than geometric, reflecting his strong belief in the unity of Nature, which is quantum at the deepest level. Quantization of gravity, according to him, simply had to be considered as the quantization of another field, the spin-2 graviton field. In this way full general relativity would be recovered at leading order, while the inclusion of loop diagrams brought into the picture a bunch of new difficulties. In this respect, Feynman's struggle against loops, while succeeding at one-loop order, failed with two- and higher-loop diagrams. Nevertheless, his results triggered further efforts and some tools he developed, such as the tree theorem, have recently become of widespread use among people working on scattering amplitudes.

References

1. R. P. Feynman, R. B. Leighton and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1963).
2. R. P. Feynman, Quantum theory of gravitation, *Acta Phys. Polon.* **24**, 697 (1963).
3. M. Gell-Mann, Dick Feynman - The guy in the office down the hall, *Phys. Today* **42**, 50 (1989).
4. C. DeWitt-Morette and D. Rickles, *The Role of Gravitation in Physics*, Report from the 1957 Chapel Hill Conference (Edition Open Access, 2011).
5. Letter from Bryce DeWitt to Agnew Bahnson, dated 15 November 1955, as cited in,[4] p. 25.
6. A. S. Blum, R. Lalli and J. Renn, The renaissance of General Relativity: How and why it happened, *Annalen der Physik* **528**, 344 (2016).
7. P. G. Bergmann, Summary of the Chapel Hill Conference, *Rev. Mod. Phys.* **29**, 352 (1957).
8. R. P. Feynman, unpublished letter to Victor F. Weisskopf, February 1961; in *Richard P. Feynman Papers*, California Institute of Technology Archives, Box 3, Folder 8.

9. W. R. Frazer, Theory of Weak and Strong Interactions, *Physics Today* **14**(12), 80 (1961).
10. G. F. Chew, *S-matrix theory of strong interactions: A lecture note and reprint volume*, Frontiers in physics series (W. A. Benjamin, New York, 1961).
11. R. P. Feynman, Closed loop and tree diagram, in Ref.[13].
12. R. P. Feynman, Problems in quantizing the gravitational field and the massless Yang-Mills field, in Ref.[13].
13. J. R. Klauder (ed.), *Magic Without Magic - John Archibald Wheeler. A Collection Of Essays In Honor Of His 60th Birthday* (Freeman, San Francisco, 1972).
14. R. P. Feynman, F. B. Morinigo, W. G. Wagner and B. Hatfield, *Feynman Lectures on Gravitation* (Addison-Wesley, Reading, MA, 1995).
15. M. Di Mauro, S. Esposito and A. Naddeo, When Physics Meets Biology: A Less Known Feynman, *Transversal Int. J. Hist. Sci.* **4**, 163 (2018).
16. R.P. Feynman, *Astronomy, Astrophysics, and Cosmology*, lectures at Hughes Aircraft Company, Notes taken and transcribed by John T. Neer, 1966-67.
17. R. De Luca, M. Di Mauro, S. Esposito and A. Naddeo, Feynman's different approach to electromagnetism, *Eur. J. Phys.* **40**, 065205 (2019).
18. M. Di Mauro, R. De Luca, S. Esposito and A. Naddeo, Some insight into Feynman's approach to electromagnetism, *Eur. J. Phys.* **42**, 025206 (2021).
19. R.P. Feynman, *Electrostatics, Electrodynamics, Matter-Waves Interacting, Relativity*, lectures at Hughes Aircraft Company, Notes taken and transcribed by John T. Neer, 1967-68.
20. M. Di Mauro, S. Esposito and A. Naddeo, A road map for Feynman's adventures in the land of gravitation, *Eur. Phys. H* **46**, 22 (2021).
21. M. Fierz and W. Pauli, On relativistic wave equations for particles of arbitrary spin in an electromagnetic field, *Proc. Roy. Soc. Lond. A* **173**, 211 (1939).
22. P. A. M. Dirac, Relativistic wave equations, *Proc. Roy. Soc. Lond. A* **155**, 447 (1936).
23. S. N. Gupta, Gravitation and Electromagnetism, *Phys. Rev.* **96**, 1683 (1954).
24. R. H. Kraichnan, Special-Relativistic Derivation of Generally Covariant Gravitation Theory, Phys. Rev. **98**, 1118 (1955).
25. J. Preskill and K. S. Thorne, Foreword to Ref.[14].
26. A. Einstein, Notes on the Origin of the General Theory of Relativity, in *Ideas and Opinions*, translated by Sonja Bargmann, (Crown, New York, 1954).
27. S. Weinberg, Derivation of gauge invariance and the equivalence principle from Lorentz invariance of the S- matrix, *Phys. Lett.* **9**, 357 (1964).
28. S. Weinberg, Photons and gravitons in perturbation theory: Derivation of Maxwell's and Einstein's equations, *Phys. Rev.* **138**, B988 (1965), B988.
29. S. Deser, Self-interaction and gauge invariance, *Gen. Rel. Grav.* **1**, 9 (1970).
30. D. G. Boulware and S. Deser, Classical General Relativity Derived from Quantum Gravity, *Annals Phys.* **89**, 193 (1975).
31. S. Deser, Gravity From Self-interaction in a Curved Background, *Class. Quant. Grav.* **4**, L99 (1987).
32. S. Deser, Gravity from self-interaction redux, *Gen. Rel. Grav.* **42**, 641 (2010).
33. R. M. Wald, Spin-2 Fields and General Covariance, *Phys. Rev. D* **33**, 3613 (1986).
34. A. Trautman and D. Salisbury, Memories of my early career in relativity physics, *Eur. Phys. J. H* **44**, 391 (2019).
35. B. S. DeWitt, Theory of radiative corrections for non-abelian gauge fields, *Phys. Rev. Lett.* **12**, 742 (1964).
36. B. S. DeWitt, Quantum Theory of Gravity. 2. The Manifestly Covariant Theory, *Phys. Rev.* **162**, 1195 (1967).

37. B. S. DeWitt, Quantum Theory of Gravity. 3. Applications of the Covariant Theory, *Phys. Rev.* **162**, 1239 (1967).
38. L. D. Faddeev and V. N. Popov, Feynman Diagrams for the Yang-Mills Field, *Phys. Lett. B* **25**, 29 (1967).
39. G. 't Hooft, Renormalization of Massless Yang-Mills Fields, *Nucl. Phys. B* **33**, 173 (1971).
40. G. 't Hooft, Renormalizable Lagrangians for Massive Yang-Mills Fields, *Nucl. Phys. B* **35**, 167 (1971).
41. G. 't Hooft and M. J. G. Veltman, Regularization and Renormalization of Gauge Fields, *Nucl. Phys. B* **44**, 189 (1972).
42. G. 't Hooft and M. J. G. Veltman, Combinatorics of gauge fields, *Nucl. Phys. B* **50**, 318 (1972).
43. G. 't Hooft and M. J. G. Veltman, One loop divergencies in the theory of gravitation, *Ann. Inst. H. Poincare Phys. Theor. A* **20**, 69 (1974).
44. M. H. Goroff and A. Sagnotti, Quantum gravity at two loops, *Phys. Lett. B* **160**, 81 (1985).
45. M. H. Goroff and A. Sagnotti, The Ultraviolet Behavior of Einstein Gravity, *Nucl. Phys. B* **266**, 709 (1986).
46. A. E. M. van de Ven, Two loop quantum gravity, *Nucl. Phys. B* **378**, 309 (1992).
47. K. S. Stelle, Renormalization of Higher Derivative Quantum Gravity, *Phys. Rev. D* **16**, 953 (1977).
48. J. Mehra, *The Beat of a different drum: The Life and science of Richard Feynman* (Oxford University Press, Oxford, 1994).
49. S. Weinberg, On the Development of Effective Field Theory, *Eur. Phys. J. H* **46**, 6 (2021).
50. S. Weinberg, Effective field theory, past and future, *Int. J. Mod. Phys. A* **31**, 1630007 (2016).
51. C. P. Burgess, Quantum gravity in everyday life: General relativity as an effective field theory, *Living Rev. Rel.* **7**, 5 (2004), and references therein.
52. J. Preskill, *Feynman after 40*, Talk given at APS April Meeting, 16 April 2018, `http://theory.caltech.edu/~preskill/talks/APS-April-2018-Feynman-4-3.pdf`.
53. R. Britto, Loop Amplitudes in Gauge Theories: Modern Analytic Approaches, *J. Phys. A* **44**, 454006 (2011).
54. I. Bierenbaum, S. Catani, P. Draggiotis and G. Rodrigo, Feynman's Tree Theorem and Loop-Tree Dualities, *PoS* **LC2010**, 034 (2010).
55. S. Caron-Huot, Loops and trees, *JHEP* **05**, 080 (2011).
56. M. Maniatis, Application of the Feynman-tree theorem together with BCFW recursion relations, *Int. J. Mod. Phys. A* **33**, 1850042 (2018).
57. M. Maniatis, Application of Britto-Cachazo-Feng-Witten recursion relations and the Feynman-tree theorem to the four-gluon amplitude with all plus helicities, *Phys. Rev. D* **100**, 096022 (2019).

Towards detecting gravitational waves: A contribution by Richard Feynman

M. Di Mauro

Dipartimento di Matematica, Università di Salerno, Fisciano, 84084, Italy
E-mail: madimauro@unisa.it

S. Esposito* and A. Naddeo†

INFN, Sezione di Napoli, C. U. Monte S. Angelo, Via Cinthia, Napoli, 80125, Italy
**E-mail: sesposit@na.infn.it*
†*E-mail: anaddeo@na.infn.it*

An account of Richard Feynman's work on gravitational waves is given. Feynman's involvement with this subject can be traced back to 1957, when he attended the famous Chapel Hill conference on the Role of Gravitation in Physics. At that conference, he presented in particular the celebrated sticky bead argument, which was devised to intuitively argue that gravitational waves must carry energy, if they exist at all. While giving a simple argument in favor of the existence of gravitational waves, Feynman's thought experiment paved the way for their detection and stimulated subsequent efforts in building a practical detecting device. Feynman's contributions were systematically developed in a letter to Victor Weisskopf, completed in February 1961, as well as in his Caltech Lectures on Gravitation, delivered in 1962-63. There, a detailed calculation of the power radiated as gravitational radiation was performed, using both classical and quantum field theoretical tools, leading to a derivation of the quadrupole formula and its application to gravitational radiation by a binary star system. A comparison between the attitudes of Feynman and of the general relativity community to the problems of gravitational wave physics is drawn as well.

Keywords: History of physics; History of relativity; Gravitational waves.

1. Introduction

The contributions of Richard P. Feynman to physics notoriously touched all four fundamental interactions of Nature. His first breakthrough occurred in the theory of electromagnetism, with his version of covariant quantum electrodynamics that earned him the 1965 Nobel prize,[1] which was developed by him between the late 1940s and the early 1950s. A few years later, Feynman's interest had focused on another fundamental interaction, namely gravity. Feynman first mentioned this new interest of his to Murray Gell-Mann, who much later recalled:

> We first discussed it when I visited Caltech during the Christmas vacation of 1954-55 [...] I found that he had made considerable progress.[2]

In view of this statement,[a] it is conceivable that Feynman's interests were directed towards gravity some time before 1954, probably shortly after completing his work on quantum electrodynamics. While working also on other subjects, most notably condensed matter physics and the theory of weak interactions,[5] Feynman nurtured his interest in gravity at least until the second half of the 1960s. In these fifteen years, he gave several contributions to the both classical and quantum gravity, some of which would later be recognized as pivotal for the field. The first public contribution to gravity was given by Feynman at the Chapel Hill conference in 1957,[4] where he took active part to many of the numerous discussion sessions. After that, records of his later work are documented by several published and unpublished sources. For detailed historical accounts of Feynman's contributions to gravitational physics, see Refs. 6 and 8.

In this paper, we focus on Feynman's work on gravitational waves. As well-known, at Chapel Hill Feynman proposed a simple and intuitive thought experiment in favor of the existence of gravitational waves. This "sticky bead" argument presumably contributed to inspiring the pioneering experimental efforts by Joseph Weber. In fact, Feynman's qualitative argument was supported and extended by detailed calculations, several results of which were presented at Chapel Hill as well. While the original calculations are presumably lost, Feynman included a detailed description of an updated version of them in a letter he wrote to Viktor F. Weisskopf in early 1961.[9] These updated computations were also included in the famous graduate lectures on gravitation he delivered at Caltech in the academic year 1962-63.[7] They show that Feynman had addressed most of the contemporary theoretical problems of the subject, showing that gravitational waves carry energy, and that they can be detected by suitable instruments; moreover, he computed in detail the energy radiated by binary self-gravitating systems. Feynman's comments clearly show that he regarded his results as compelling (and probably somewhat straightforward as well, since he did not publish them) solutions to the above mentioned issues. Indeed, in his subsequent work on gravity, Feynman did not deal with gravitational waves any more, focusing instead on quantum gravity.

The paper is organized as follows. In Sect. 2 we briefly sketch the history of gravitational waves prior to Feynman's engagement with them, highlighting the theoretical problems which were the concern of scientists in the 1950s and 1960s. In Sect. 3 we briefly introduce the historical context in which the Chapel Hill conference took place and we summarize Feynman's contributions to it. In Sect. 4 we examine the letter to Weisskopf, and the parts of the Lectures on Gravitation which discuss gravitational waves. We close this section with a summary of Feynman's solutions to the theoretical issues in gravitational wave research. In Sect. 5 we give some comments on Feynman's work on gravitational waves, in particular in connection with his general views on gravity and on fundamental interactions in general.

[a] This was confirmed by Bryce S. DeWitt, who declared: "Even by 1955 Feynman claimed to have spent a great deal of effort on the problem of gravitation".[3]

In Sect. 6 we outline some contemporary and later developments, and in particular we describe the solutions to the theoretical issues that are currently accepted by the scientific community. Sect. 7 is devoted to conclusions.

2. Brief history of gravitational waves up to the 1950s

The main open problem concerning gravitational waves, when Feynman tackled them, was their very existence as physical entities.[b] In fact, general relativity enjoys general covariance, which means that coordinate systems have no intrinsic meaning. While very elegant and powerful, this symmetry also poses great difficulties, which have kept haunting general relativity practitioners since the earliest days of the theory. Namely, it is not easy to distinguish between real physical phenomena and mere artifacts due to a bad coordinate choice, hence invariant characterizations are needed. One of the earliest examples is the Schwarzschild singularity,[11] which occurs in the spherically symmetric solution found by Karl Schwarzschild. In fact, the metric describing this solution in spherical coordinates displays a singularity on a surface located at a given value of the radial coordinate. This singularity is now well-known to be apparent, and it is removed by choosing a different coordinate system,[12] but it took more than forty years to really understand what this meant. Two other, strictly linked examples, are given by the definition of energy in general relativity, and by gravitational waves themselves. When Albert Einstein first showed in 1916[13] that the gravitational field equations he had written down the year before[14] predict in the weak field limit the existence of wave-like excitations, and when in 1918[15] he corrected a major mistake in his 1916 paper, finding his celebrated quadrupole formula describing the leading-order energy emission through gravitational radiation, he chose a particular coordinate system, in which his equations looked formally analogous to the equations of electromagnetism. Einstein immediately recognized that some of the wave solutions he had found were spurious, and could be eliminated by a coordinate change (they were flat space in curved coordinates), while the remaining ones were not spurious, and carried energy according to his formula.

2.1. *The question of the existence of gravitational waves*

A weak point in Einstein's derivation was spotted and corrected by Sir Arthur S. Eddington,[16] namely, the fact that Einstein's choice of coordinates explicitly contained the speed of light, hence the propagation speed of gravitational waves may have been inadvertently put in by hand by the German physicist. Eddington improved and extended Einstein's work, giving a more rigorous proof of the fact that some of Einstein's solution were not coordinate artifacts, they carried energy and they propagated with the speed of light. Finally, Eddington rederived the

[b]For details on the history of gravitational waves, with a focus on the theoretical side, we refer to Ref. 10.

quadrupole formula, correcting a minor error by Einstein (an overall factor of 2), and applied it to the computation of energy radiated away by a rigid rotator. However, in complaining about the lack of an invariant way of distinguishing physical from spurious waves, Eddington referred to the latter in a very suggestive way, stating that, being coordinate artifacts, they propagate with the "speed of thought" (Ref. 17, p. 130). This generated a misunderstanding according to which Eddington attributed this characteristic to gravitational waves in general, thus questioning their very existence. Indeed, doubts concerning the existence of gravitational waves did not stop bothering the scientists working in general relativity for a long time. Another problem was posed by the fact that, unlike electromagnetism, gravitation is described by a highly nonlinear theory, while gravitational waves had been found only in the linearized approximation. The possibility thus existed that gravitational waves were in fact only artifacts of the linearized theory, meaning that there is no solution to the full gravitational field equations describing gravitational fields propagating as waves. This point of view was shared for some time by Einstein himself, after in 1937 he had tried with Nathan Rosen to find plane wave solutions to the full gravitational field equations.[18] Indeed, the two physicists found singularities in their solutions, and interpreted them as obstructions to the existence of propagating solutions to the equations.[c] After a well-known incident with a referee,[20] Einstein finally realized that the singularities he had found were coordinate artifacts, and that in fact he and Rosen had found an exact solution to his field equation, which enjoyed cylindrical symmetry. This was one of the first solutions to the full theory describing gravitational waves.[d] However, the question of the physical effects ascribable to these waves was still unclear, since it seemed that they did not carry energy, as still argued by Rosen himself in 1955 (Ref. 22, pp. 171-174), and anyway the Einstein-Rosen solution was non-physical, since its source was not localized. What was needed was an exact solution describing waves radiating from a localized center, which was still unavailable at the time. In any case, the fact that even Einstein himself had doubted that gravitational waves could exist certainly contributed to making them a neglected field of study for twenty more years.

A very important breakthrough occurred in 1956, when Felix A. E. Pirani managed to give an invariant characterization of gravitational radiation, in terms of spacetime curvature rather than energy,[23,24] giving in fact birth to the modern popular description of gravitational waves as propagating ripples of curvature. This work was performed in the context of Pirani's investigations of the physical meaning of the Riemann curvature tensor, based on the geodesic deviation equation[25] and on the classification of curvature tensors in terms of their principal null directions.[26] In particular, using the geodesic deviation equation, Pirani showed that the passage of a gravitational wave should modify the proper distance between test particles,

[c] Einstein mentioned this in a letter to Max Born (Ref. 19, letter 71).
[d] Actually, Einstein and Rosen had rediscovered a solution first found by Guido Beck in 1925.[21]

as an effect of the variation of the curvature. Thus, a measurable effect could be ascribed at least in principle to gravitational waves.

2.2. *The question of energy loss*

The issue of existence was not the only one which was at stake in gravitational wave physics in the 1950s. Even taking existence for granted, indeed, there was a problem concerning gravitational radiation by self-gravitating systems, such as binary stars. The first applications by Einstein and Eddington involved rigid bodies such as rotators, which can be considered to be kept together by the electromagnetic interaction, while a bound state kept together by gravity itself is beyond the reach of the linearized approximation, which was the only context in which gravitational energy radiation had been studied. In fact, many physicists even doubted that gravitational waves were radiated at all by such a system. And even if radiation did occur, it was unclear whether the Einstein quadrupole formula, which had been obtained in the linearized theory, correctly described the resulting energy loss. The issue had been addressed for the first time in 1941 by Lev D. Landau and Evgenij M. Lifshitz, in the first edition of their textbook,[27] where the binary system was treated by using the post-Newtonian approximation, and energy loss was computed by applying the quadrupole formula. However, there was no consensus about the correctness of their calculation, and the question was still open in the 1950s.

2.3. *Summary of the open problems*

From what we said above, it emerges that the situation concerning gravitational wave research was far from clear in the 1950s, with many basic theoretical problems and no hope of experimental guidance. We may summarize the main issues at stake as follows:

(1) The actual existence of gravitational waves as physical entities, their speed of propagation and their energy conveying;
(2) The existence of gravitational waves as solutions to the full nonlinear gravitational field equations;
(3) The radiation of gravitational waves by self-gravitating systems and the correctness of the quadrupole formula in describing the energy loss.

3. The Chapel Hill conference

As we saw, before the 1950s research on gravitational waves was carried out by only a handful of pioneers. This lack of interest was in fact not only the result of the doubts expressed by Einstein and also attributed to Eddington. The whole field of general relativity remained dormant for thirty years, from the mid-1920s to the mid-1950s, after an initial burst of activity. During this period, later baptized "low-watermark" phase,[28–30] general relativity was considered by most physicists to be

a rather esoteric and heavily mathematical subject, with tenuous relations with the physical world.

Around the mid-1950's, things began to change, and gravity slowly regained a prominent place in physics. Pivotal to this change of attitude, which was dubbed the "renaissance" of general relativity,[31–33] was the organization of a series of international conferences entirely devoted to the subject. The renaissance of general relativity was characterized by the fact that physicists began studying the theory in a less speculative fashion than before, by focusing on its physical implications, and by developing dedicated, i.e. intrinsically general relativistic methods. That was also a time where researchers from other fields, most notably particle and field theory, began to be interested on gravitation and to apply their methods to it. One of them was of course Feynman, who started working on gravity just in this period, and it was in one of these conferences, namely the one (organized by DeWitt and his wife Cécile DeWitt-Morette) at the University of North Carolina at Chapel Hill in January 1957,[4] that he first announced his results in the field, as we anticipated in the introduction. In his many contributions to the discussions at Chapel Hill, he clearly delineated his attitude towards gravity, and presented a host of results he had already obtained. Besides the arguments on favor of the existence of gravitational waves, which we discuss in the following subsections, Feynman sketched a proposal for a field-theoretical approach to classical gravity. His view, shared by other particle physicists, was indeed that gravity had to be regarded as a fundamental interaction like all the others, and hence it had to be described by a field theory, which would also encode quantum corrections to the classical theory. Indeed, within a few years, he managed to derive the field equations of general relativity (as summarized in Ref. 7), and he attacked the problem of quantum gravity, from this perspective.[6,8] Finally, the Chapel Hill conference is notably one of the few occasions where Feynman expressed his views about foundational quantum issues, which were inspired by the problem of the quantization of gravity.[4,6,8,34]

3.1. *The sticky bead argument*

Feynman presented his famous sticky bead argument in a session devoted to discussing the necessity of gravity quantization (in which he also proposed a thought experiment in order to argue in favor of that). In fact, Feynman had arrived a day later at the conference[e], skipping the session devoted to gravitational waves. However, further discussion of gravitational waves in sessions devoted to quantum gravity was not off-topic, since the existence of these waves is of course very important for the quantization of the theory. Thus, in the middle of a discussion about whether the gravitational field had to be quantized even in the absence of gravitational waves, Feynman proposed a simple physical argument in favor of their existence (Ref. 4, p. 260 and pp. 279-281). He argued that if gravitational waves

[e] An amusing and well-known recollection of his arrival in North Carolina can be found in Ref. 35.

exist, they must carry energy and this, together with the existence of these waves as solutions of the linearized theory, was enough for him to be confident in their existence. He expressed this intuitive feeling with the words: "My instincts are that if you can feel it, you can make it" (Ref. 4, p. 260). Feynman's reasoning was detailed in an expanded version of his remarks, included in the records of the conference[4]:

> I think it is easy to see that if gravitational waves can be created they can carry energy and do work. Suppose we have a transverse-transverse wave generated by impinging on two masses close together. Let one mass A carry a stick which runs past touching the other B. I think I can show that the second in accelerating up and down will rub the stick, and therefore by friction make heat. I use coordinates physically natural to A, that is so at A there is flat space and no field [...]. Then Pirani at an earlier section gave an equation for the motion of a nearby particle. (Ref. 4, p. 279).

The result by Pirani is of course that obtained in Refs. 23 and 24 (which indeed had been presented earlier at the conference, cf. Ref. 4, p. 141), stating that the displacement η of the mass B (measured from the origin A of the coordinate system) in the field of a gravitational wave satisfies the following differential equation:[f]

$$\ddot{\eta}^a + R^a_{0b0}\eta^b = 0, \qquad (a, b = 1, 2, 3) \tag{1}$$

R being the Riemann curvature tensor at the point A. Feynman continued:

> the curvature tensor [...] does not vanish for the transverse-transverse gravity wave but oscillates as the wave goes by. So, η on the RHS is sensibly constant, so the equation says the particle vibrates up and down a little (with amplitude proportional to how far it is from A on the average, and to the wave amplitude.) Hence it rubs the stick, and generates heat. (Ref. 4, p. 279)

The upshot is that the stick absorbs energy from the gravitational wave, which means that the latter carries energy and is able to do work.

3.2. *Feynman's theoretical detector*

In connection with the device he considered, Feynman went on considering a possible gravitational wave detector in which there are four, rather than two test masses, which oscillate as shown in Figure 1.

This thought device was used to compute how much energy gravitational waves would carry. The device was described by Feynman as follows:

> If I use 4 weights in a cross, the motions at a given phase are as in the figure (Fig. 1). Thus a quadrupole moment is generated by the wave. Now the

[f] This follows from the geodesic deviation equation (Ref. 4, p. 141; Refs. 23 and 24).

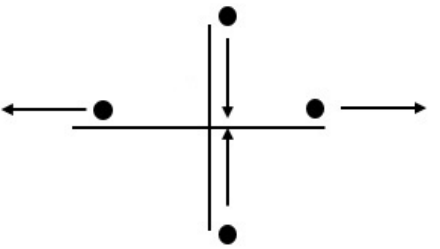


Fig. 1. Scheme of Feynman's gravitational wave detector.

> question is whether such a wave can be generated in the first place. First since it is a solution of the equations (approx.) it can probably be made. Second, when I tried to analyze from the field equations just what happens if we drive 4 masses in a quadrupole motion of masses like the figure above would do - even including the stress-energy tensor of the machinery which drives the weights, it was very hard to see how one could avoid having a quadrupole source and generate waves. Third my instinct is that a device which could draw energy out of a wave acting on it, must if driven in the corresponding motion be able to create waves of the same kind. [...] If a wave impinges on our "absorber" and generates energy - another "absorber" placed in the wave behind the first must absorb less because of the presence of the first, (otherwise by using enough absorbers we could draw unlimited energy from the waves). That is, if energy is absorbed the wave must get weaker. How is this accomplished? Ordinarily through interference. To absorb, the absorber parts must move, and *in moving generate a wave* which interferes with the original wave in the so-called forward scattering direction, thus reducing the intensity for a subsequent absorber. In view therefore of the detailed analysis showing that gravity waves can generate heat (and therefore carry energy proportional to R^2 with a coefficient which can be determined from the forward scattering argument). I conclude also that these waves can be generated and are in every respect real. (Ref. 4, p. 280, emphasis in the original)

In this quote, Feynman's point of view about the existence of gravitational waves is evident: since they exist as solutions of the field equations (albeit in the linearized approximation), they carry energy, and they can be generated (in fact, the detector generates secondary waves), then they must be a real phenomenon.

3.3. *Binary systems*

Within the discussion (Ref. 4, p. 260), Feynman also quoted a result concerning the calculation of the energy radiated by a self-gravitating two-body system in a circular orbit. This result shows that he had addressed also this issue, performing

detailed calculations. The quoted result is:

$$\frac{\text{Energy radiated in one revolution}}{\text{Kinetic energy content}} = \frac{16\pi}{15}\frac{\sqrt{mM}}{m+M}\left(\frac{u}{c}\right)^5, \tag{2}$$

where m and M are the masses of the two bodies, and u is their relative velocity. In the case of the Earth-Sun system, this formula describes as expected a tiny effect, leading to a huge order of magnitude (about 10^{26} years) for the lifetime of the motion of the Earth around the Sun. Despite the smallness of the effect, this is a definite prediction. Feynman stated to have performed a full analysis of this problem some time before (Ref. 4, p. 280), but he only quoted the result, without presenting his computations. Fortunately, Feynman reported a version of them four years later, in the letter written to Victor F. Weisskopf,[9] to which we now turn.

4. The letter to Weisskopf and the sixteenth lecture on gravitation

As anticipated, a complete and systematic description of Feynman's computations on gravitational waves can be found in a letter he wrote to Weisskopf in February 1961,[9] to answer a question that the Austrian physicist had asked him in the fall of 1960, concerning the reality of gravitational radiation. The letter was so detailed that it was subsequently included in the material distributed to the students attending the Caltech lectures in 1962-63. Indeed, the sixteenth lecture of that course (the last one in the published version[7]) dealt with gravitational waves, and reproduced part of the computations showed in the letter itself, with additional interesting comments.

The computations that Feynman reported in the letter and in the lectures were surely related to those he had performed before the Chapel Hill conference, but included several improvements that Feynman had conceived meanwhile. Indeed, at the end of the letter, he recalled the conference, claiming that

> It was this entire argument used in reverse that I made at a conference in North Carolina several years ago to convince people that gravity waves must carry energy. [...] Only as I was writing this letter to you did I find this simpler argument from the Lagrangian (Ref. 9, p. 14).

What Feynman meant by "argument used in reverse" and by "simpler argument from the Lagrangian" will be clear by the end of this section.

The most striking fact about the computations is that they are performed adopting a quantum point of view, treating general relativity as a quantum field theory, and then considering the classical limit:

> My view is quantum mechanical, but the classical limits are easily derived. (Ref. 9, p. 6)

Only later the same computation is repeated from the usual, classical point of view, in the framework of linearized gravity. As we saw, in fact, in that period Feynman

was struggling with the quantization of gravity, building on the field theoretical viewpoint that he had advocated at Chapel Hill. In the beginning of 1961, he had recognized that loop diagrams in the theory present major difficulties, but he was still far from a solution (later in the same year he would report on his progress at the La Jolla conference[36]). However, he could safely study classical gravity by applying his field theoretical methods to tree diagrams, which are the only ones relevant to the classical limit (while loop diagrams describe quantum corrections). This is what he did in the letter,[g] claiming that:

> I am studying the problem of quantization of Einstein's General Relativity. I am still working out the details of handling the divergent integrals which arise in problems in which some virtual momentum must be integrated over. But for cases of radiation, without the radiation corrections, there is no difficulty. (Ref. 9, p. 1)

The "radiation corrections" Feynman was referring to are of course what are usually called radiative corrections, i.e. higher order corrections to the processes, which include the troublesome loop diagrams. In fact, in the Lectures Feynman adopted this point of view in studying many other problems in classical gravity:

> All other problems in the theory of gravitation[h] we shall attack first by the quantum theory; in order to obtain the classical consequences on macroscopic objects we shall take the classical limits of our quantum answers. (Ref. 7, p. 163)

The field theoretical methods that Feynman used are intrinsically perturbative. Feynman explicitly discussed the validity of this approximation, and of the linearized approximation he used in the classical computation, stating that this is a natural and obvious choice in view of the extreme weakness of the gravitational interaction. He stated in fact that:

> I won't be concerned with fields of arbitrary strength [...]. I am surprised that I find people objecting to an expansion of this kind, because it would be hard to find a situation where numerically a perturbation series is more justified! (Ref. 9, p. 1)

which is confirmed in the Lectures:

> There needs to be no apology for the use of perturbations, since gravity is far weaker than other fields for which perturbation theory seems to make extremely accurate predictions. (Ref. 7, p. 207)

[g] A conceptually analogous calculation was performed in 1936 by Matvei Bronstein,[37] who derived the quadrupole formula with the correct factor of 2 by computing the emission coefficients of transverse gravitons.

[h] Here Feynman was referring to all problems, apart from that of a spherical mass distribution and cosmology.

From these quotations, it is evident that Feynman considered the contemporary debate around the existence of gravitational waves in the full nonlinear theory of general relativity, as opposed to the linearized theory, to be pointless, since the perturbative/linearized theory would furnish an excellent approximation in most cases of interest.[i]

4.1. *The quantum computation*

To perform his computation, Feynman started from linearized gravity, which from his point of view was the linear theory of a massless charged spin-2 quantum field coupled to matter. For simplicity, he considered scalar (spin-0) matter, and he computed the probabilities for single graviton emission in several processes involving matter particles, in the low energy limit. The result is that the probability of emission of a graviton with frequency ω in the solid angle $\mathrm{d}\Omega$ is (Ref. 9, p. 7; Ref. 7, p. 214):

$$P = a^2 \frac{\mathrm{d}\Omega}{4\pi} \frac{\mathrm{d}\omega}{\omega} \frac{\lambda^2}{4\pi^2}, \tag{3}$$

where a is a kinematical factor and λ is the gravitational coupling constant (which is proportional to the square root of Newton's constant G).

Feynman knew that the probabilities of radiation of massless quanta become universal in the low energy limit, i.e. the probability of emission of a soft quantum in an interaction is independent of the spin of the interacting quanta (hence the choice of scalar matter implies no loss of generality), and most importantly of the kind of interaction which is involved. This happens because the low-energy limit is equivalent to the long-wavelength limit, in which the physics is insensitive to short-distance details. In particular, soft graviton emission occurs, with the same probability, also when matter quanta interact gravitationally. Since classical radiation is related to soft quantum emission, the implication is that masses accelerating under the reciprocal gravitational interactions must emit gravitational radiation as any other accelerating mass:

> Although this was worked out for spin zero particles, the result, as $\omega \to 0$, is the same for all, spin 1/2, photons, etc. It is correct for collisions irrespective of the kind of force. For example, Schiff asked me if two masses under their mutual gravitation can radiate. The answer is certainly yes. (Ref. 9, p. 7)

According to Feynman, these results proved that self-gravitating system must radiate gravitational waves. This view was confirmed in the Lectures:

[i]He also acknowledged the fact that the problem was at the moment of purely academic interest, stating that "the most "practical" thing to do is to go to zero order and forget the whole problem altogether" (Ref. 9, p. 1).

> As far as the radiation is concerned, the exact nature of the over-all scattering process is not important. I emphasize the last point because there are always some theorists who go about mumbling some mystical reasons to claim that the radiation would not occur if the scattering is gravitational - there is no basis for these claims; as far as we are concerned, radiation of gravity waves is as real as can be; the sun-earth rotation must be a source of gravitational waves. (Ref. 7, p. 215).

Thus the issue of gravitational radiation by self-gravitating systems was considered solved by Feynman. We emphasize that his quantum point of view was pivotal to that conclusion, since his reasoning relied on a well-known result of quantum field theory, namely the universality of soft graviton emission. Once again, Feynman showed some annoyance towards the physicists who continued to have doubts, while the solution was so clear to him.

4.2. *The classical computation and the quadrupole formula*

In the second part of the letter, Feynman switched to a classical computation, which was considered to be more appropriate for macroscopic objects, such as celestial bodies:

> for the motion of big objects such as planets and stars it may be more consistent to work in the classical limit. (Ref. 7, p. 215)

The classical computation was developed in close analogy with electrodynamics, taking into account that in the case of gravitation the source is a tensor $S_{\mu\nu}$ rather than a vector J_μ. The differential equation which has to be solved to find classical gravitational waves is (in the notation of Ref. 7):

$$\Box^2 \overline{h}_{\mu\nu} = \lambda S_{\mu\nu}, \tag{4}$$

where $h_{\mu\nu}$ is the field describing deviations of the metric from the flat Minkowski metric and the bar operation is defined as:

$$\overline{h}_{\mu\nu} = \frac{1}{2}\left(h_{\mu\nu} + h_{\nu\mu}\right) - \frac{1}{2}\eta_{\mu\nu} h^{\sigma}{}_{\sigma}. \tag{5}$$

$\Box^2$ is the usual, flat-space, d'Alembertian operator. As well-known, the above equation corresponds to the linearized Einstein equations in the de Donder gauge $\partial^\mu \overline{h}_{\mu\nu} = 0$. When all quantities are periodic with a given frequency ω, the solution at a point 1 located at a distance from the source much greater than the linear dimensions of the source itself (which is the region where $S_{\mu\nu}$ is expected to be large):

$$\overline{h}_{\mu\nu}(\vec{r}_1) = -\frac{\lambda}{4\pi r_1} e^{i\omega r_1} \int d^3\vec{r}_2 S_{\mu\nu}(\vec{r}_2) e^{-i\vec{K}\cdot\vec{r}_2}, \tag{6}$$

with $|\vec{r}_1| \gg |\vec{r}_2|$ as stated. This approximation holds for nearly all cases of astronomical interests, where wavelengths are much longer than the system's dimensions,

such as for instance binary stars and the Earth-Sun system. Finally, Feynman assumed that the motions of matter in the source were slow, so that he could resort to the post-Newtonian approximation. This is where the computations in the last section of Lecture 16 of the Caltech Lectures (Ref. 7, pp. 218-220) stopped, while in the letter Feynman went on computing the power radiated by the above waves in the quadrupole approximation which, for a periodic motion of frequency ω, read (Ref. 9, p. 11):

$$\frac{\text{mean energy radiated}}{\text{sec.}} = \frac{1}{5} G\omega^6 \sum_{ij} \left| Q'_{ij} \right|^2 , \tag{7}$$

where $Q'_{ij} = Q_{ij} - \frac{1}{3}\delta_{ij} Q_{kk}$, $Q_{ij} = \sum_a m_a R^a_i R^a_j$ is the quadrupole mass moment, and the root mean square average is taken. The result (7) was then shown to agree with the quantum mechanical probability (3) computed in the first part of the letter (after suitable averaging). To strengthen this result, Feynman had considered earlier (Ref. 9, p. 10) an analogous comparison for the case of electromagnetism, finding again perfect correspondence.

Feynman then applied his formulas to the case of a circularly rotating double star (Ref. 9, p. 12), promptly recovering the result (2). In sum, Feynman's classical computation was analogous to that performed earlier by Landau and Lifshitz. However, his treatment was simpler, since he managed by using a very ingenious method for computing the energy density, which essentially exploits the fact that for classical transverse plane waves the linearized gravity Lagrangian reduces to that of the harmonic oscillator. This spared him formal complications such as the cumbersome energy-momentum pseudo-tensor, which was used by Landau and Lifschitz. This is in fact the "simpler argument from the Lagrangian" which he referred to in the end of the letter.

The above treatment was then supplemented with a thorough analysis of the effects of a gravitational wave impinging on a device made of two test particles placed on a rod with friction, in this way providing a more complete description of the working principle of the gravitational wave detector previously introduced at Chapel Hill. A second absorber, made of four moving particles in a quadrupole configuration, as in Figure 1, was also considered, and shown to be able to absorb energy from a gravitational wave acting on it. The oscillating device was shown to re-radiate waves with the same energy content. Apparently, at Chapel Hill, Feynman had derived the expression for the radiated energy (which was then applied to the binary system case) from the detailed study of the quadrupolar detector, either because at that time he did not know how to extract the definition of energy from the gravitational Lagrangian, or because the existing methods were considered to be too cumbersome by him. In the letter, instead, Feynman found a simple way of taking the opposite route (from this we can understand why Feynman claimed that the argument at Chapel Hill had been "used in reverse"), since the quadrupole formula was derived in general from the linearized Einstein equations and then applied to the detector case.

Thus, Feynman had shown that gravitational waves carry energy, and he had computed how much, also in the case of a binary star, satisfactorily addressing both the first and the third of the above issues. At the end of Lecture 16 of Ref. 7, Feynman commented again on the energy content of gravitational waves:

> What is the power radiated by such a wave? There are a great many people who worry needlessly at this question, because of a perennial prejudice that gravitation is somehow mysterious and different - they feel that it might be that gravity waves carry no energy at all. We can definitely show that they can indeed heat up a wall, so there is no question as to their energy content. The situation is exactly analogous to electrodynamics (Ref. 7, p. 219).

and in the letter, while recalling the Chapel Hill conference he commented:

> I was surprised to find a whole day at the conference devoted to this question, and that "experts" were confused. That is what comes from looking for conserved energy tensors, etc. instead of asking "can the waves do work?" (Ref. 9, p. 14).

These two quotes again display Feynman's annoyance towards the relativity community. The fact that relativists kept arguing about these problems without reaching an agreement, and that they privileged sophisticated formal tools to plain physical thinking, definitely disturbed Feynman, who repeatedly expressed his annoyance, and may have contributed to his loss of interest in the subject.

4.3. *Feynman's answer to the three issues*

The detailed study of Feynman's work on gravitational waves reveals what he thought the solutions to the issues listed in Sect. 2 should be, at least from a theoretical point of view, pending the ultimate experimental verification. In summary, these were:

(1) Feynman's thought experiments and calculations clearly show that gravitational waves exist and that they carry energy, while the fact that they propagate with the speed of light is a consequence of the fact that they emerged from massless quanta, i.e. the gravitons;
(2) Since gravity is so weak, it makes perfect sense to study gravitational waves using the linearized approximation;
(3) Gravitationally-bound system emit gravitational radiation as any other bound system, and the energy loss is described by the quadrupole formula in the approximations valid in most situations of interest.

5. Comments on Feynman's attitude towards gravity and gravitational waves

As emphasized, Feynman's approach to gravitational waves, and to gravitation in general, was much closer in spirit to particle physics than to general relativity. In fact, Feynman preferred a field theoretical approach over the geometrical one which had become the standard one among relativists. In his opinion, indeed, the latter approach divided gravity from other fundamental interactions, preventing the investigation of the relation of gravitation with the rest of physics, and also posing great obstacles towards its quantization.

A second motivation which underlies Feynman's choice of using quantum field theoretical methods for investigating classical gravitational radiation and other problems in classical gravity, was his belief that quantum mechanics underlies Nature at the most fundamental level, with the forces we observe at the macroscopic level emerging from quantum interactions in the classical limit. This is most clearly expressed by him in some unpublished lectures:

> I shall call conservative forces, those forces which can be deduced from quantum mechanics in the classical limit. As you know, Q.M. is the underpinning of Nature... (Ref. 38, p. 35)

Feynman recognized that the main problem with gravity was the fact that general relativity had received only a handful of experimental confirmations, while the subjects of gravitational waves and quantum gravity were confined to pure theory. These subjects were being investigated by the relativity community by resorting to sophisticated mathematics, in the hope that rigor could be a substitute for empirical evidence. Feynman had a different point of view. To him, the best way to proceed was to try to extract physical predictions from the theory *as if* these were easily accessible experimentally. This viewpoint was strongly advocated by him at Chapel Hill in some critical comments:

> There exists, however, one serious difficulty, and that is the lack of experiments. [...] so we have to take a viewpoint of how to deal with problems where no experiments are available. There are two choices. The first choice is that of mathematical rigor. [...] The attempt at mathematical rigorous solutions without guiding experiments is exactly the reason the subject is difficult, not the equations. The second choice of action is to "play games" by intuition and drive on. Take the case of gravitational radiation. [...] I think the best viewpoint is to pretend that there are experiments and calculate. In this field since we are not pushed by experiments we must be pulled by imagination. (Ref. 4, pp. 271-2)

and shortly after

> The real challenge is not to find an elegant formalism, but to solve a series of problems whose results could be checked. This is a different point of view. Don't be so rigorous or you will not succeed. (Ref. 4, p. 272)

This is in fact the viewpoint that Feynman adopted in his investigations of gravitational waves, but also in his work on quantum gravity.[8]

6. Contemporary and subsequent developments

When the Chapel Hill conference took place, the time was ripe for the existence of gravitational waves to be accepted by most physicists. Thus, the work of Pirani, and Feynman's sticky bead argument, managed to convince people that gravitational waves belong to the realm of physical phenomena, and suggested ways to investigate their physical effects. In fact, an argument similar to Feynman's was conceived independently by a leading relativist, Hermann Bondi, who was also present at Chapel Hill. This is referred to by several of Bondi's remarks at the conference, such as the following one, made after Pirani's talk (Ref. 4, p. 142): "Can one construct in this way an absorber for gravitational energy by inserting a $d\eta/d\tau$ term, to learn what part of the Riemann tensor would be the energy producing one, because it is that part that we want to isolate to study gravitational waves?". The absorber of gravitational energy is just a more formal description of the "stickiness" of Feynman's sticky-bead. Shortly after the conference, Bondi published his variant of the sticky-bead device,[39] although unlike Feynman he did not succeed in relating the intensity of the gravitational wave to the amount of energy carried by it. Inspired by these discussions, another participant to the conference, Joseph Weber, soon undertook his lifetime effort to experimentally detect gravitational waves[40]; Weber built his famous resonant bar detector in 1966[41] and even announced the first experimental detection of gravitational waves coming from the center of the Milky Way in 1969.[42] While this finding was later disproved, Weber's works triggered all the subsequent research on the detection of gravitational waves, which culminated in the monumental detections by the LIGO-Virgo collaboration in 2016.[43] Among all the physical results that this brought about, the confirmation came that gravitational waves move with the speed of light.[44]

Thus, it may be said that the first issue, that of the existence and physical effects of gravitational waves, was solved after Chapel Hill, with agreement among the scientific community. The same cannot be said of the other two issues. The relativity community reached an agreement on the status of gravitational waves in the full nonlinear theory only some years later, thanks to the rigorous work performed in that framework by Bondi, Pirani, Ivor Robinson and Andrzej Trautman, among others. In particular, Robinson and Trautman found exact solutions to the equations of

general relativity describing waves radiated from bounded sources.[45] Furthermore, in the sixties, Bondi, Rainer W. Sachs, Ezra T. Newman and Roger Penrose,[46–50] gave a satisfactory definition of the energy radiated at infinity as gravitational radiation by an isolated gravitating system in an asymptotically flat spacetime. Thus, the modern theoretical characterization of gravitational waves had emerged by the late 1960s.

The last issue took still longer to be settled, since it crucially involved the problem of motion in general relativity and the issue of radiation reaction, whose foundations were still quite shaky.[51] The doubts concerning whether binary systems radiated gravitational energy were fought by crucial empirical input, which came in the mid-1970s with the discovery of the first binary neutron star.[52,53j] This did not immediately settle the quadrupole formula controversy yet,[54] but it prompted many efforts towards its resolution. The confirmation of came when Thibault Damour[55] showed that there was quantitative agreement between the emission rate computed from the quadrupole formula and observations. In fact, it is only in the final stages of the merging of compact objects that higher (mass and current) momenta need to be included (see e.g. Ref. 56 and references therein).

7. Conclusions

In the above pages, we have described the work performed by Richard Feynman in the physics of gravitational waves. We have seen how his attitude towards physics, and towards fundamental interactions, shaped this work, and how he managed to give sound and substantially correct answers to the main theoretical problems of the time. Feynman's physical intuition and sound calculations convinced him early on of the existence of gravitational waves and gravitational radiation. His work contributed to the settlement of the issue, and played a role in triggering the first experimental efforts towards their detection. Feynman did not share the theoretical concerns of most relativists, who continued to argue about other issues, until rigorous solutions and/or experimental input were available, and considered those arguments pointless, harshly criticizing them. This may have been contributed to his loss of interest in the subject, starting from the late 1960s, when he began to be involved to the only fundamental interaction to whose understanding he had not contributed yet, that is, the strong interaction.[5] But that is another story.

Acknowledgments

The authors would like to thank the staff of the Caltech archives for providing them a copy of Feynman's letter to Weisskopf.[9]

[j]In fact, so far, all direct and indirect evidence for gravitational waves has come from binary systems, which hence are the only confirmed sources.

References

1. R. P. Feynman, The development of the space-time view of quantum electrodynamics, *Science* **153**, no. 3737, 699 (1966).
2. M. Gell-Mann, Dick Feynman-The Guy in the Office Down the Hall, *Phys. Today* **42**, 50 (1989).
3. Letter from Bryce DeWitt to Agnew Bahnson, dated 15 November 1955, cited in Ref. 4, p. 25.
4. C. DeWitt-Morette and D. Rickles, *The Role of Gravitation in Physics*, Report from the 1957 Chapel Hill Conference (Edition Open Access, Berlin, 2011).
5. J. Mehra, *The Beat of a different drum: The Life and science of Richard Feynman*, (Oxford University Press, Oxford, 1994).
6. J. Preskill and K. S. Thorne, Foreword to Ref. 7.
7. R. P. Feynman, F. B. Morinigo, W. G. Wagner and B. Hatfield, *Feynman Lectures on Gravitation*, (Addison-Wesley, Reading, MA, 1995).
8. M. Di Mauro, S. Esposito and A. Naddeo, A road map for Feynman's adventures in the land of gravitation, *Eur. Phys. J. H* **46**, no.1, 22 (2021), [arXiv:2102.11220 [physics.hist-ph]].
9. R. P. Feynman, unpublished letter to Victor F. Weisskopf, February 1961; in Richard P. Feynman Papers, California Institute of Technology Archives, Box 3, Folder 8.
10. D. Kennefick, *Traveling at the speed of thought: Einstein and the quest for gravitational waves* (Princeton University Press, Princeton, 2007).
11. K. Schwarzschild, Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie (On the gravitational field of a mass point according to Einstein's theory), *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, **1916**, 189 (1916), [arXiv:physics/9905030 [physics]].
12. M. D. Kruskal, Maximal extension of Schwarzschild metric, *Phys. Rev.* **119** (1960), 1743-1745.
13. A. Einstein, Näherungsweise Integration der Feldgleichungen der Gravitation (Approximate Integration of the Field Equations of Gravitation), *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916**, 688 (1916).
14. A. Einstein, Die Feldgleichungen der Gravitation (The Field Equations of Gravitation), *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915**, 844 (1915).
15. A. Einstein, Über Gravitationswellen (On Gravitational Waves), *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1918**, 154 (1918).
16. A. S. Eddington, The propagation of gravitational waves, *Proc. Roy. Soc. Lond.* A **102**, 268 (1922).
17. A. S. Eddington, *The Mathematical Theory of Relativity*, 2nd edn. (Cambridge University Press, Cambridge, 1960).
18. A. Einstein and N. Rosen, On Gravitational waves, *J. Frank. Inst.* **223**, 43 (1937).
19. A. Einstein and M. Born, *The Born-Einstein Letters: Friendship, Politics, and Physics in Uncertain Times* (MacMillan, New York, 2005).
20. D. Kennefick, Einstein versus the *Physical Review*, *Physics Today* **58**, 9, 43 (2005).
21. G. Beck, Zur Theorie binärer Gravitationsfelder (The Theory of binary gravitational fields), *Z. Phys.* **33**, 713 (1925).
22. A. Mercier and M. Kervaire (eds.), *Fünfzig Jahre Relativitätstheorie* (Birkhäuser, Basel, 1956).
23. F. A. E. Pirani, On the Physical significance of the Riemann tensor, *Acta Phys. Pol.* **15**, 389 (1956), reprinted in *Gen. Rel. Grav.* **41**, 1215 (2009).
24. F. A. E. Pirani, Invariant formulation of gravitational radiation theory, *Phys. Rev.* **105**, 1089 (1957).

25. J. L. Synge, On the deviation of geodesics and null-geodesics, particularly in relation to the properties of spaces of constant curvature and indefinite line-element, *Annals Math.* **35**, 705 (1934).
26. A. Z. Petrov, Klassifikacya prostranstv opredelyayushchikh polya tyagoteniya, *Uch. Zapiski Kazan. Gos. Univ.* **114 (8)**, 55 (1954) English translation: Classification of spaces defined by gravitational fields, *Gen. Rel. Grav.* **32 (8)**, 1665 (2000).
27. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields,* first English edition, Addison-Wesley, Cambridge, 1951.
28. J. Eisenstaedt, La relativité générale à l'étiage: 1925-1955, *Arch. Hist. Exact Sci.* **35**, 115 (1986).
29. J. Eisenstaedt, Trajectoires et impasses de la solution de Schwarzschild, *Arch. Hist. Exact Sci.* **37**, 275 (1987).
30. J. Eisenstaedt, The low water mark of general relativity, 1925-1955, in D. Howard, J. Stachel (eds.), *Einstein and the History of General Relativity*, (Birkhäuser, Basel, 1989), pp. 277-292.
31. C. M. Will, *Was Einstein Right? Putting General Relativity to the Test*, (Oxford University Press, Oxford, 1986).
32. C. M. Will, *The renaissance of general relativity*, in P. C. W. Davies (ed.), *The New Physics*, (Cambridge University Press, Cambridge, 1989).
33. A. S. Blum, R. Lalli and J. Renn, *The Renaissance of General Relativity in Context*, *Einstein Studies* **16** (Birkhäuser, Cham, 2020).
34. H. D. Zeh, Feynman's interpretation of quantum theory, *Eur. Phys. J. H* **36**, 63 (2011).
35. R. P. Feynman, *Surely You're Joking, Mr. Feynman! - Adventures of a curious character* (W.W. Norton & Co., New York, 1985).
36. W. R. Frazer, Theory of Weak and Strong Interactions, *Phys. Tod.* **14**(12), 80 (1961).
37. M. P. Bronstein, Quantentheories schwacher Gravitationsfeldern (Quantum theory of weak gravitational fields), *Physikalische Zeitschrift der Sowietunion* **9**, 140 (1936), English translation by M. A. Kurkov, edited by S. Deser, *Gen. Relativ. Grav.* **44**, 267 (2012).
38. R. P. Feynman, *Electrostatics, Electrodynamics, Matter-Waves Interacting, Relativity*, lectures at the Hughes Aircraft Company, Notes taken and transcribed by J. T. Neer, 1967-68.
39. H. Bondi, Plane gravitational waves in general relativity, Nature **179**, 1072 (1957).
40. J. Weber, Detection and Generation of Gravitational Waves, *Phys. Rev.* **117**, 306 (1960).
41. J. Weber, Observation of the Thermal Fluctuations of a Gravitational-Wave Detector, *Phys. Rev. Lett.* **17**, 1228 (1966).
42. J. Weber, Evidence for discovery of gravitational radiation, *Phys. Rev. Lett.* **22**, 1320 (1969).
43. B. P. Abbott et al. [LIGO Scientific and Virgo], Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**, 061102 (2016) [arXiv:1602.03837 [gr-qc]].
44. B. P. Abbott et al. [LIGO Scientific, Virgo, Fermi-GBM and INTEGRAL], Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *Astrophys. J. Lett.* **848**, L13 (2017), [arXiv:1710.05834 [astro-ph.HE]].
45. I. Robinson and A. Trautman, Spherical Gravitational Waves, *Phys. Rev. Lett.* **4**, 431 (1960).
46. H. Bondi, M. G. J. van der Burg and A. W. K. Metzner, Gravitational waves in general relativity. 7. Waves from axisymmetric isolated systems, *Proc. Roy. Soc. Lond. A* **269**, 21 (1962).

47. R. K. Sachs, Gravitational waves in general relativity. 8. Waves in asymptotically flat space-times, *Proc. Roy. Soc. Lond. A* **270**, 103 (1962).
48. R. K. Sachs, Gravitational radiation, In *Relativity, Groups and Topology*, C. DeWitt and B. S. DeWitt eds. (Gordon and Breach, New York, 1964) pp. 521-562.
49. E. Newman and R. Penrose, An Approach to gravitational radiation by a method of spin coefficients, *J. Math. Phys.* **3**, 566 (1962).
50. R. Penrose, Zero rest mass fields including gravitation: Asymptotic behavior, *Proc. Roy. Soc. Lond. A* **284**, 159 (1965).
51. J. Ehlers, A. Rosenblum, J. N. Goldberg and P. Havas, Comments on Gravitational Radiation Damping and Energy Loss in Binary Systems, *Astrophys. J. Lett.* **208**, L77 (1976).
52. R. A. Hulse and J. H. Taylor, Discovery of a pulsar in a binary system, *Astrophys. J. Lett.* **195**, L51 (1975).
53. J. Taylor, L. Fowler, P. McCulloch, Measurements of general relativistic effects in the binary pulsar PSR1913 + 16. *Nature* **277**, 437 (1979).
54. D. Kennefick, The binary pulsar and the quadrupole formula controversy, *Eur. Phys. J. H* **42** (2017) no.2, 293-310.
55. T. Damour, Gravitational radiation reaction in the binary pulsar and the quadrupole formula controversy, *Phys. Rev. Lett.* **51**, 1019 (1983).
56. L. Blanchet, Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries, *Living Rev. Rel.* **17**, 2 (2014), [arXiv:1310.1528 [gr-qc]].

Stellar gravitational collapse, singularity formation and theory breakdown

Kiril Maltsev

Heidelberg Institute for Theoretical Studies,
Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany
E-mail: kiril-maltsev@h-its.org
www.h-its.org

Department of Physics and Astronomy, University of Heidelberg,
Im Neuenheimer Feld 226, 69120 Heidelberg, Germany
www.uni-heidelberg.de

We review three definitions (missing point(s) unsteadiness, infinite quadratic curvature invariant, and geodesic incompleteness) of what a gravitational singularity is, and argue that prediction of a gravitational singularity is problematic for General Relativity (GR), indicating breakdown of Lorentzian geometry, only insofar as it concerns the infinite curvature singularity characterization. In contrast, the geodesic incompleteness characterization is GR's innovating hallmark, which is not meaningfully available in Newtonian gravity formulations (locally infinite density field, and locally infinite gravitational force) of what a gravitational singularity is. It is the continuous, non-quantized, nature of Lorentzian geometry which admits gravitational contraction be continued indefinitely. The Oppenheimer-Snyder 1939 analytical solution derives formation of a locally infinite curvature singularity *and* of incomplete geodesics, while Penrose's 1965 theorem concerns formation of incomplete (null) geodesics only. We critically examine the main physical arguments against gravitational singularity formation in stellar collapse, with scope restriction to decades spanning in between Schwarzschild's 1916 solution and Penrose's 1965 singularity theorem. As the most robust curvature singularity formation counter-argument, we assess Markov's derivation of an upper bound on the quadratic curvature invariant $R_{\mu\nu\lambda\delta}R^{\mu\nu\lambda\delta} \leq \frac{1}{\ell_P^4}$ from a ratio of natural constants $\hbar$, c and G, in connection with Wheeler's grounding of the premise that the Planck scale ℓ_P is ultimate.

Keywords: Geodesic incompleteness; Infinite curvature; Schwarzschild incompressible fluid; Einstein-Rosen bridge; ISCO; Chandrasekhar mass; TOV mass; Core collapse; Compact stars; Oppenheimer-Snyder collapse; Singularity theorems; Curvature bound; Theory breakdown.

1. Characterization of a gravitational singularity

Over the period of over hundred years, the idea that gravitational singularity formation, predicted by General Relativity, is physical - representing a genuine feature of reality - was iteratively regarded with skepticism. The scientific community's doubts at times were based on mere conceptual reservation and reluctance to accept a novel paradigm, while at other times had justification by strong physical argument. In Sec. (1), we will outline how (not) to conceive what a gravitational singularity in General Relativity is, and compare two contrasting interpretational

views: realism and anti-realism about singularity formation. We will provide a simple field-theoretical argument in favour of the interpretation of the prediction of curvature divergence as an indication of theory breakdown in Sec. (1.3). The goal of the subsequent Sections (2) and (3) is to explore some of the main arguments that have historically (with a scope restriction to 1916-1965) been advocated as preclusions to gravitational singularity formation in stellar collapse, and to assess their relevance for the present.

1.1. *Coordinate vs. geometric singularity*

A first important propaedeutic worth a recall is the demarcation between a coordinate singularity and a geometric singularity. Let $(\mathcal{M}, g_{\mu\nu})$ be an ordered pair, with $\mathcal{M}$ a manifold, associated with a spacetime, and $g_{\mu\nu}$ a Lorentzian metric with signature $(-,+,+,+)$, defined on the manifold $\mathcal{M}$ and associated with geometric properties of the spacetime. A coordinate singularity indicates the limit of applicability of a coordinate chart to represent a metric tensor $g_{\mu\nu}$. The line element ($ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu$) of the single-parameter (M, associated with mass) Schwarzschild metric in the Schwarzschild spherical coordinate chart $(x^0, x^1, x^2, x^3) = (ct, r, \theta, \phi)$ reads

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 g_\Omega. \tag{1}$$

As $r \to 2GM$, the radial component $g_{rr} \to \infty$ while the temporal component $g_{tt} \to 0$. Yet, the Gullstrand-Painlevé, the Kruskal-Szekeres and the Eddington-Finkelstein coordinates are examples of coordinate systems for which the Schwarzschild metric, when transformed onto another coordinate chart ($ds'^2 = g'_{\mu\nu}\, dx'^\mu\, dx'^\nu$), is regular at the Schwarzschild boundary (which, in Schwarzschild coordinates, locates at the $r = 2GM$ two-sphere) of the set of nested round spheres. To illustrate, the line element in Eddington-Finkelstein coordinates (v, r, θ, ϕ) reads

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dv^2 \pm 2\, dv\, dr + r^2 g_\Omega \tag{2}$$

where a tortoise coordinate $r^* = r + 2GM\, ln|\frac{r}{2GM} - 1|$ is introduced, satisfying $\frac{dr^*}{dr} = \left(1 - \frac{2GM}{r}\right)^{-1}$, to define the time-advanced ingoing (for $v = t + r^*$) and time-retarded outgoing (for $v = t - r^*$) coordinates, respectively. Representation of the Schwarzschild metric in Eddington-Finkelstein coordinates shows curves to be extendible beyond the Schwarzschild boundary. According to the Principle of General Covariance, coordinate systems play no role in the formulation of fundamental laws of physics; they only have a descriptive-representational role. No observer, and therefore no coordinate system is privileged over another, yet the capability of physical representation by one coordinate chart may break down while others do not: an illucidating GR-independent example is the failure of representation of the North (at $\theta = 0$) and of the South Pole (at $\theta = \pi$) by spherical coordinates, while their representation is regular in cartesian coordinates. In GR, changing a

coordinate chart implies a re-definition of the coordinate axes relative to a proper inertial frame. While the introduction of a new coordinate chart allows to transform away the apparent singularity at the Schwarzschild boundary $r = 2GM$, the central metric field irregularity at $r = 0$ persists in the maximally extended Schwarzschild spacetime. The Kretschmann scalar curvature invariant is a coordinate-independent measure that allows to identify a curvature singularity in the geometry of the metric field. It is defined as full contraction of the Riemann tensor with itself, summing over all of its components:

$$K = R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}. \tag{3}$$

Its evaluation on the Schwarzschild metric yields $K = \frac{48G^2M^2}{r^6}$, which is finite at the Schwarzschild boundary $r = 2GM$, but divergent at $r = 0$.

1.2. *Missing point(s) unsteadiness, infinite quadratic curvature invariant, or geodesic incompleteness?*

Calculation of the Kretschmann scalar is one method to tell apart a genuine singularity in the geometry of the metric field from a mere coordinate singularity pertinent to a coordinate representation of the metric. There has, however, been debate over what the appropriate definition of a gravitational singularity is, which appropriately characterizes it physically. In particular, as summarized and cross-examined in Ref. 1, three different main conceptions (all of which have sub-classes) have been proposed:

(1) missing point(s) unsteadiness,
(2) infinite curvature, and
(3) path inextendibility.

Missing point(s) unsteadiness. The conception that a gravitational singularity is a missing point, or a set of missing points, on the spacetime manifold, suggests that a spacetime is singular if it contains a 'hole'.[2] Then, the metric field $g_{\mu\nu}$ is discontinuous at a point $p \notin \mathcal{M}$ or a set of points $\{p_i\} \notin \mathcal{M}$, which have been removed from the $\mathcal{M}$ spacetime manifold, akin to a tear in a piece of cloth. Conceptually, a missing point on the spacetime manifold suggests the absence of spacetime itself at that point. The motivation for this singularity characterization stems from the fact that derivative operators, such as the Riemann tensor, are locally ill-defined at points at which the metric field $g_{\mu\nu}$ is discontinuous - at points that lie outside of its domain. The demand that the metric is continuous and differentiable everywhere formalizes into the requirement of non-degeneracy: $det(g_{\mu\nu}) \neq 0$. The condition is fulfilled if all the components of both the covariant ($g_{\mu\nu}$) and the contravariant ($g^{\mu\nu}$) metric tensor are regular. Then, a (removable) gravitational singularity violates the steadiness condition for any choice of coordinate chart: the missing point, or points, on the spacetime manifold give rise to the impossibility of its smooth

local extension, and thereby to unsteadiness in the gravitational field. This historically original notion of a gravitational singularity has been deployed in the works of Schwarzschild, Einstein, Hilbert, and contemporaries.

The missing point(s) singularity definition suggests that the singular spacetime manifold could be 'patched' by the appendage of the boundary that embeds the singularity. The main burden to accepting this notion of singularity is that the boundary construction in the neighbourhood of the singularity turns into the location problem[1]: there is no Lorentz-covariant location of the missing point(s) on the spacetime manifold, which would be coordinate- and therefore observer-independent. Assuming that there are no more additional dimensions to the spacetime, there is no way of imagining it being embedded in a larger topological space, with respect to which one could identify the locations of the subspace point(s) to be missing out.

Infinite quadratic curvature invariant. There are many sub-classes to definitions of a gravitational singularity in terms of an infinite curvature pathology. They differ in choice of the curvature tensor and of the divergence criterion. One of the curvature singularity classes is identified by the Kretschmann scalar, a quadratic curvature invariant defined with the Riemann tensor. It is used as an indicator of the presence of a local curvature singularity, when $K \to \infty$. A local observer is then subject to tidal forces that arise from the differences in "strength" of the gravitational field between two neighbouring points in space-time, and the curvature singularity manifests in its growth without bound. However, a well-behaved quadratic curvature invariant does not imply that all other local curvature measures are regular: there are examples of space-times where the quadratic Riemann curvature invariants are vanishing but the Ricci curvature tensor $R_{\mu\nu}$ diverges.[3] Again others, such as the conformal Misner singularities, where curvature scalars are well-behaved but the Riemann tensor evaluated with some local tetrad is divergent. Yet the advantage of curvature tensor *products* over curvature tensors is the former's coordinate-independence.

A problem with a local curvature singularity characterization, even if defined by virtue of curvature tensor products, is that it is incomplete: there are examples of well-defined space-times in which the state of motion of the observer is decisive in determining the physical response, as the putative singularity is traversed: whether or not an object is rotating about its axis, or subject to proper acceleration in the direction of motion, determines whether the object gets torn into pieces by infinite tidal forces, or whether it survives the cross-over.[4] In the latter example, the curvature pathology arises for a non-rotating observer resting on a geodesic, and an improved local gravitational singularity definition needs be reformulated accordingly. However, even this extension of the singularity definition is insufficiently precise, given a number of counter-examples: a point in spacetime may be well-defined for an observer travelling on a complete geodesic and experiencing well-behaved tidal forces all the way along, while this same point exhibits a curvature

pathology for an observer travelling on an incomplete geodesic.[3] Finally, therefore, the most general gravitational singularity definition that aims to invoke a locally infinite curvature invariant needs to confess that the curvature pathology arises locally along a curve that is incomplete. This raises the question whether infinite curvature is even a necessary feature of a gravitational singularity, rather than an inessential singularity property, since the formation of an incomplete path may entail but does not require a curvature pathology. The relation between curvature divergence and geodesic incompleteness remains a delicate problem that still seeks its resolution in Lorentzian differential geometry, but for which there seem to be two options:

(1) a gravitational singularity is defined by geodesic incompleness (or, more generally, path inextendibility) alone. It does not imply divergence of any local curvature measure, which is an additional feature, realized in some singular spacetimes but not in others.
(2) a gravitational singularity is defined by geodesic incompleteness concurring with a locally divergent quadratic curvature invariant; curvature divergence endows geodesic incompleteness with a physical (and local) characterization.

Important conclusions can be drawn from comparison of the curvature singularity characterization in Einstein's gravity theory to the Newtonian gravity theory counterpart. The formulation of the Raychaudhuri-Komar singularity theorem[a] in the framework of Newton-Cartan gravity,[8] a geometrized tensor reformulation of Newtonian gravity theory suggests that

- in Newtonian gravity[b], curvature divergence (as measured by the geometrized Poisson equation $R_{\mu\nu}\xi^\mu\xi^\nu = 4\pi\rho$ where ξ^μ is an arbitrary time-like vector) arises as consequence of a locally (infinite density) pathology in the stress-energy field, which is source of the gravitational field, while

[a]The Raychaudhuri-Komar singularity theorem[5,6] predicts formation of an infinite matter density singularity from generic conditions. It applies to a perfect fluid (described by the stress-energy tensor in the form of eq. (11)), whose timelike ($u^\mu u_\mu = -1$) velocity field u^μ is geodesic and non-rotational. If the derivative of the divergence $\theta \equiv \nabla_\mu u^\mu$ along the geodesic congruence is positive (negative) at an instant of time and the convergence condition $R_{\rho\nu}u^\rho u^\nu \geq 0$ holds, then the energy density ρ of the fluid diverges in the infinite past (future) along every integral curve of u^μ. The above mentioned modern formulation of the theorem is taken from Ref. 7.

[b]The key difference is the following: while in Newtonian gravity a gravitational force field singularity may arise only as consequence of a pathological infinite energy density matter state, Einstein's gravity theory allows for formation of gravitational singularities in a vacuum ($T_{\mu\nu} = 0$), analytically as solutions to the vacuum Einstein Field Equations. A generic example is the collision of two plane gravitational waves ($R_{\mu\nu} = 0$) in an otherwise flat background. Tipler's general plane-wave singularity theorem[9] proves the collision spacetime be future null geodesically incomplete. In addition, in the collision region of the 'sandwiched' GWs, under certain conditions (which depend on polarization of the colliding GWs) a curvature singularity may form at the focused point, or a coordinate singularity instead, along with a Killing-Cauchy horizon.[10–12]

- geodesic incompleteness is a pathology in the gravitational field itself, independent of pathology or non-pathology of a matter state. This singularity characterization is not meaningfully available in Newtonian gravity theory.[c]

This distinction reminds of the interest to enquire not only about how a singularity is defined, but also how it is generically formed under physical conditions.

Geodesic incompleteness. The definition of a gravitational singularity in terms of geodesic incompleteness is the least problematic and most widely accepted singularity characterization. In contrast to a complete geodesic that is continued indefinitely, an incomplete path has a finite affine length parameter (which is proper time τ, for a time-like vector field). One of the ways to formalize geodesic completeness reads as follows:

> *A spacetime is singular if it contains an incomplete causal geodesic* $\gamma : [0, a) \to \mathcal{M}$ *such that there is no extension* $\gamma' : \mathcal{M} \to \mathcal{M}'$ *for which* $\gamma' \circ \gamma$ *is extendible.*[13]

The demand for unique geodesics that do not cross implies that the causal geodesic is not well-defined beyond the affine length a, at which the affine parameter converges. A relativistic spacetime $(\mathcal{M}, g_{\mu\nu})$ is timelike ($\gamma^\mu\gamma_\mu = -1$), null ($\gamma^\mu\gamma_\mu = 0$), or spacelike ($\gamma^\mu\gamma_\mu = +1$) geodesically incomplete if it entails an incomplete timelike, null, or spacelike geodesic, respectively. The example of the Taub-NUT space-time shows that a space-time can be time-like complete but null-incomplete, and vice versa.[14] Geodesic incompleteness implies that the worldline of a freely falling observer terminates, with no possibility of further extension.[d] A background presupposition of path inextendibility is that the spacetime in the theory is *maximally extended* and that it does not form a subset of a larger dimensional, more extensive geometry into which the manifold (e.g., a four-dimensional spacetime) is embedded.

1.3. *Theory breakdown?*

Singularity realism vs. anti-realism. The above mentioned three gravitational singularity definitions in General Relativity are conceptually quite distinct pathologies in the geometry of the gravitational field, and do not imply one another: missing point(s) unsteadiness singularity suggests the (local) absence of spacetime itself,

[c] See Ref. 8 for an opposing view. According to our assessment, the author confounds the geodesic incompleteness characterization of a gravitational singularity in Newton-Cartan gravity with that of the missing point(s) unsteadiness.

[d] In Refs. 1, 15, 16 a more inclusive *b-incompleteness* criterion is advocated as gravitational singularity definition. According to b-incompleteness, the inextendible path can but does not need be a geodesic: A maximal spacetime is singular if and only if it contains an inextendible path of a generalized affine length. Since every incomplete geodesic is b-incomplete, the geodesic incompleteness characterization is sufficient for the argument of this paper.

while an infinite quadratic curvature invariant suggests that tidal forces become infinite at a point in space-time, which itself may be well-defined on the manifold. Geodesic incompleteness suggests the finitude of worldlines in a spacetime, which nevertheless may be well-defined and endowed with well-behaved finite quadratic curvature invariants at any point on the manifold.

The predictions of gravitational singularities by General Relativity can, for each of the singularity classes separately, be physically interpreted in one of the two ways: either as

(1) a theory deficiency, or as
(2) opening up a profoundly novel viewpoint of properties of space-time geometry.

The stance of singularity realism is supported by arguments such as the empirical success of General Relativity as the predictive theory, and the reasoning that the prediction of -say- geodesic incompleteness does not prima facie logically imply any contradiction or internal inconsistency within the gravitation theory. The realist stance is also inclined to accept a gravitational collapse singularity as part of the definition of a Black Hole.

In contrast, the stance of singularity anti-realism is supported by lessons learnt from the pre-history of radical theory changes in physics, which teach us to only partially confess to the belief in some statements made by the current best physical theory, but not in others. Of particular concern are the prior experiences with singularities, in ray optics, in phase transition hydrodynamics, in Newtonian N-body gravitational dynamics, in Maxwellian electromagnetism and in other subfields of physics[e], all of which nail down the appearance of a *divergent quantity* as an artifact of the mathematical model that is deployed. Anti-realism about gravitational singularities in GR is supported by the mathematical ill-definition and non-differentiability of geometrical quantities, as far as the infinite curvature characterization is concerned. The relevance of the argument is demonstrated below upon the example of the electromagnetic interaction potential between two fermionic charges.

Theory change: The case of Electromagnetism. In classical electromagnetism, the Coulomb potential between two point masses carrying electromagnetic charges q_1 and q_2 reads

$$V(r) = -\frac{1}{4\pi\epsilon}\frac{q_1 q_2}{r}, \tag{4}$$

where r is the relative distance between the point masses. For vanishing distance $r \to 0$, the interaction potential $V(r)$ diverges. The physics of the electromagnetic interaction is, however, inappropriately described by the classical Maxwellian field theory at small relative distances at scale of order of magnitude of the de Broglie wavelengths of the charged particles. Here, effects of their quantum nature become non-negligible, and Quantum Electrodynamics takes over as descriptive theory

[e]See Ref. 19 for a brief discussion of some of the above mentioned examples.

instead: the basic interaction potential between two fermion charges Q_1 and Q_2 in Minkowski space-time $\eta_{\mu\nu}$ and the relative wave-vector $\mathbf{k}$ turns into the integral

$$V(r) = -\frac{-g^2\hbar c Q_1 Q_2}{4\pi}\frac{1}{(2\pi)^3}\int d^3k\; e^{\frac{i\mathbf{k}\cdot\mathbf{r}}{\hbar}}\frac{4\pi\eta_{\mu\nu}}{k^2}. \tag{5}$$

The upshot from this comparison is that the prediction of a singularity qua divergency in the interaction potential marks the limits of an empirically adequate applicability of a theory: it arises as an artifact of the theory description at coarse-grained scale. Similarly, General Relativity is a classical field theory that describes the phenomenology of gravitation at a macroscopic scale at which principles of classical physics apply. However, gravitational collapse - which is central subject of this work- involves processes in the strong gravity regime at microscopic scales $\delta r \to 0$, at which GR is not tested as gravitation theory, and hence lacks the basis for trust in its predictions.

2. Singularities in the interior Schwarzschild metric (1916–1939)

2.1. *Schwarzschild's incompressible fluid model, and Einstein's singularity resolution attempts*

Infinite pressure requirement. In his first 1916 paper,[20] Schwarzschild derives an analytical solution to the gravitational field *exterior* to a gravitating point mass, completing earlier work of approximated solutions found by Einstein. Eq. (1) is a simplified writing of the metric due to Droste[21] who found the same solution independently, that renounces from redundant variable transformations, and where the radial coordinate r is defined as a radial measure away[f] from the point mass M. In contrast, in his second 1916 publication[22] he calculates[g] the gravitational field *inside and outside* of, not a point mass, but of a static non-rotating spherically symmetric incompressible fluid. The externally measured radius of the fluid sphere is given by $P_a = \sqrt{\frac{3}{\varkappa\rho_0}}\sin\chi_a$, and internal properties of the fluid are supposed to give rise to an inner pressure source $P(r)$. In order to calculate the pressure distribution, Schwarzschild proceeds by imposing a Hydrostatic Equilibrium (HE) condition and the vanishing pressure boundary condition $P(P_a) = 0$ onto the fluid. From there, he finds an analytic expression to the pressure gradient $P(\chi) = \rho_0\left(\frac{2\cos\chi_a}{3\cos\chi_a - \cos\chi} - 1\right)$ be proportional to the density and to $v = \frac{2}{3\cos\chi_a - \cos\chi}$ which Schwarzschild refers

[f] Note that a covariant measure of spatial distance between two nested spheres at r_1 and r_2 is obtained from the proper distance $\varrho = \int_{r_1}^{r_2} dr\,\sqrt{g_{rr}}$.

[g] The line element of the interior Schwarzschild solution in the original coordinates used by Schwarzschild takes the non-trivial form:

$$ds^2 = \left(\frac{3cos\chi_a - cos\chi}{2}\right)^2 dt^2 - \frac{3}{\varkappa\rho_0}(\chi^2 + sin^2 d\theta^2 + sin^2\chi\, sin^2\theta\, d\phi^2). \tag{6}$$

Schwarzschild derives a metric with two independent parameters (χ_a and ρ_0) associated with internally measured radius and density of the incompressible fluid, respectively. Here, $\varkappa = 8\pi k^2$ where k is the Gaussian gravitation constant.

to as the "*naturally measured light velocity*".[22] After a critical matter concentration threshold[h] is surpassed, v, and thereby also the pressure, diverge at the center ($\chi = 0$). There, the gravitational field becomes infinite, and introduces a discontinuity into metric $g_{\mu\nu}$. From his calculation follows the requirement of an infinite pressure at the center of the sphere in order to maintain its HE, which he evaluates as physically unacceptable, and as a breakdown of the incompressible constant density fluid model. He summarizes his finding as follows:[i]

> "*In the center of the sphere ($\chi = 0$) the speed of light and pressure become infinite, as soon as* $\cos\chi_a = 1/3$*, and as the fall velocity has become equal to* $\sqrt{8/9}$ *of the (naturally measured) speed of light. Therewith a limit of concentration is given, beyond which the sphere of incompressible fluid may not exist. Applying our equations to* $\cos\chi_a < 1/3$*, one would obtain unsteadiness already above the center of the sphere [...]. Over these solutions, which indeed are physically meaningless, since they result in infinite pressure in the center, one can pass over to the limiting case of a mass concentrated at a point [...]. For an externally measuring observer follows (...) that the sphere of given gravitational mass* $\alpha/2k^2$ *may have no smaller externally measured radius than* $P_a = \alpha$*.*"[22]

The Newtonian central singularity counterpart, and pragmatism on the "impassible barrier". In Droste's coordinates, Schwarzschild's analytical 1916 solution entailed the prediction of two irregularities: the central singularity at the center of the incompressible fluid sphere ($r = 0$), and at the Schwarzschild coordinate singularity at $r_S = 2GM$ - above the center of the sphere but beneath the internal radius of the incompressible fluid. Both irregularities were perceived as troublesome by contemporaries, however, the Schwarzschild singularity has historically been assessed as the major problem. Coordinate transformations of the type of eq. (2) to eliminate the unsteadiness at the Schwarzschild surface have been under development, yet their significance was not fully clarified until the work of Lemaitre (1932); but various other considerations of how to handle with the prediction of irregularities in the metric field were brought forward instead. The central $r = 0$ singularity has likely not been considered as a threat to General Relativity for the reason that it already occurred in the gravitational potential $\phi(r) = -\frac{-GMm}{r}$ between two gravitating point masses M and m in Newtonian gravity. The Schwarzschild singularity, however, was novel and predicted to occur in the deep interior of stars that soon were modelled by Schwarzschild's ideal fluid ansatz: for instance, the Schwarzschild radius of the Sun amounts to roughly

[h] The critical matter concentration threshold is obtained from applying the condition $\cos\chi_a = \frac{1}{3}$ to the externally measured radius $P_a = \sqrt{\frac{3}{\varkappa\rho_0}}\sin\chi_a$, from which follows $P_a = \alpha$ for a point mass and $P_a = \frac{9}{8}\alpha$ for an incompressible fluid sphere. $\alpha/2k^2$ is the gravitational mass.

[i] Translation from the German original into English by the author K. M.

3 km, far below the Sun's radius of about 0.7 million km. Eddington called the Schwarzschild surface the "*magic circle*", and conjectured that no particles inside a star may penetrate this "*impassible*"[2] barrier. Reason is that, as Droste remarked in Ref. 21, the penetration would take an infinitely long time ($g_{tt} \to 0$), or as Einstein put it, "*a clock kept at this place would go at the rate zero*"[23] but nothing could develop without time, according to this line of reasoning.

Observation of stars was at that time period based solely on their surface properties[j], and traced on the Hertzsprung-Russell diagram. Since the prediction of the Schwarzschild singularity thereby escaped any empirical test, the prevailing logical positivist attitude to science, which assumed that any empirically unverifiable statement is meaningless, is likely to have fostered a less concerned view of the Schwarzschild singularity: no stars had been known to have a physical radius smaller than $2GM$, thus the Schwarzschild singularity appeared tolerable so long as it did not produce any falsifiable predictions. Nevertheless, during the Paris conference (1922), Hadamard contemplated the possibility of the Schwarzschild surface surpassing the physical radius of a star. Einstein conceived this scenario as a threat to the theory of General Relativity, and aimed to disprove the possibility of the "Hadamard catastrophe" on physical grounds.

The innermost stable circular orbit. Einstein was guided by an understanding of Mach's principle as implying that the metric field is determined by its matter-energy content, for "*inertia originates in a kind of interaction between bodies*".[24] Guided by Mach's principle, he interpreted a singularity in the gravitational field as an indication of a pathological matter state that gives rise to it. Hence, he aimed to prove that while stellar matter is source of the gravitational field, it is impossible to accumulate matter neither at their center of mass at $r = 0$, nor at the Schwarzschild surface $r = r_S$, where the singularities were supposed to arise. Then -according to his line of reasoning- irregularities in the metric field are unphysical, so long as they are not occupied by matter, which originates the properties of the gravitational field. In his 1939 paper, Einstein's goal was to demonstrate that this holds true in an allegedly more realistic model of stellar matter than Schwarzschild's perfect fluid ansatz. He starts the paper with an expansion of the argument brought forward by Schwarzschild himself, pointing at the inadequacy of incompressible fluid assumption:

> "*The concept of an incompressible liquid is not compatible with relativity theory as elastic waves would have to travel with infinite velocity. It would be necessary, therefore, to introduce a compressible liquid whose equation of state excludes the possibility of sound signals with a speed in excess of*

[j] Asteroseismology, the study of stellar interior structure through their global oscillation modes, which are excited in the interior and induce structural imprints onto photometric observables, had not yet been developed at that time.

> *the velocity of light. But the treatment of any such problem would be quite involved; besides, the choice of such an equation of state would be arbitrary within wide limits, and one could not be sure that thereby no assumptions have been made which contain physical impossibilities.*"[23]

Instead, Einstein proposes another model: a spherically symmetric self-gravitating many-body system composed of non-interacting point particles of equal unit mass. For this system, Einstein calculates the interior gravitational field and equations of motion, and finds the innermost stable circular orbit (ISCO) to lie above the Schwarzschild radius in GR, in contrast to the treatment of the same many-body system in Newtonian gravity, which does not yield any lower bound on the radius of the ISCO. Orbits of a radius tighter than the Schwarzschild radius, as he finds, require orbital velocities that exceed the speed of light, which is prohibited by the principles of relativity theory. From there, he concludes that the irregularities in the metric field at $r = r_S$ and at $r = 0$ are physically irrelevant, because these regions do not get populated by particles in the self-gravitating many-body system.

Einstein's calculations were correct and insightful in their own right, however he draw false conclusions from them. With his background premise that all test particle orbits always remain stable, he excluded possibility of unstable orbits, such as an inspiral. A more cautious conclusion would have been that there are no *stable* particle orbits within a certain distance away from the Schwarzschild radius. A spiral-like particle trajectory, of orbital velocity smaller than the speed of light yet beneath the innermost stable orbital radius, is not impeded from arising by Einstein's calculation: Einstein's calculation shows that an orbit radius lower than r_S cannot be circularly stable.

The Einstein-Rosen bridge. Einstein was even ready to sacrifice the original version of the gravitational field equations for the sake of doing away gravitational singularities in the Schwarzschild metric. At the same time, he was opportunist[25] about turning the prediction of field singuilarities into a solution to the problem of particle representation and of particle motion in General Relativity. With regard to the latter problem, he saw the logical independence of the geodesic law from the field equations as a weakness of GR. One of his approaches to a remedy was to derive the geodesic law from application of the covariant conservation law $\nabla_\nu T^{\mu\nu} = 0$ to the Einstein Field Equations (EFE) at the limiting case of matter-free space ($T^{\mu\nu} = 0$). From there would follow that, "*[i]f one assumes that matter is arranged in narrow 'world tubes', one obtains from this by an elementary consideration the theorem that the axes of those 'world tubes' are geodesic lines (in the absence of electromagnetic fields). This means: the law of motion is a consequence of the field law.*".[26] With regard to the former problem, Einstein assessed the representation of matter by the stress-energy tensor as phenomenological approximation, as a crude substitute of known (quantum) properties of matter. Instead, singularities in the

gravitational field, seemingly problematic at first, opened up a novel approach to resolve another problem. His proposal was the representation of elementary particles by pole singularities in an empty gravitational field:

> "*All attempts to represent matter by an energy-momentum tensor are unsatisfactory and we wish to free our theory from any particular choice of such a tensor. Therefore we shall deal here only with gravitational equations in empty space, and matter will be represented by singularities of the field.*"[27]

In this approach unifying particle representation and motion, particles in GR are represented by localized pole singularities in the metric fields $g_{\mu\nu}$, which outside the singularities are regular elsewhere and satisfy the source-free ($T_{\mu\nu} = 0$) EFE, while particle geodesics are represented as displacements of pole singularities in and by the gravitational field obeying the source-free EFE. This approach was a stepstone toward a unified field theory Einstein was questing for, yet Einstein presumed that the total field (obtained from what he refers to as the "*completed field equations*"), unifying the gravitational field and the matter field, would eliminate singularities:

> "*[...] one will have to say that the field of a material particle may the less be viewed as a pure gravitational field the closer one comes to the position of the particle. If one had the field equation of the total field, one would be compelled to demand that the particle themselves would everywhere be describable as singularity free solutions of the completed field-equations.*"[28]

The prediction of singularities in solutions to the EFE fostered his view that GR is a preliminary, not the final field theory. The belief in the singularity representation of particles was further shaken Silberstein's argument[29] that, due to the existence of a *static* axisymmetric solution to the field equations representing two pole singularities, either the field equations are wrong, or the problem of motion cannot be treated with it, or particles cannot be represented as singularities. Reason is that the two particles so represented are, in Silberstein's construction, immobile and not accelerated in each other's gravitational field. Influenced by Silberstein's argument, in Einstein's 1935 paper with his student Rosen,[30] the authors present a formalism that casts the representation of a particle from the point mass singularity representation at the origin of the Schwarzschild metric onto a non-singular particle representation obtained from a topological bridge construction connecting two asymptotically flat regions of spacetime. To get there, Einstein and Rosen proceed by two essential steps. First, they modify the source-free EFE by transforming the Ricci tensor (which equals $R_{\mu\nu} = R^{\lambda}_{\mu\nu\lambda} = g^{\lambda\rho}R_{\rho\mu\nu\lambda} = 0$ in the source-free field) to the tensor density

$$R_{\mu\nu} \rightarrow R^{*}_{\mu\nu} = g^{2}R_{\mu\nu} \tag{7}$$

where g is the determinant of the metric field $g_{\mu\nu}$. The effect it had in their study of the flat Minkowski metric in Rindler coordinates is that the irregularity condition $g = 0$, since the new field equations have no denominators in g and do not result in 'divide by zero' divergences. Hence, Einstein & Rosen's move was to make the unsteadiness singularity condition $det(g_{\mu\nu}) = 0$ itself irrelevant, by the mutilation of the source-free field equations. Second, Einstein & Rosen perform a coordinate transformation, $u^2 = r - 2GM$, which shifts the Schwarzschild singularity down to the origin of the coordinate system, thereby transforming away the entire central region $r \in [0, 2GM[$ and, therewith, the central singularity. They interpret the line element ds^2 resulting from the spherically symmetric solution to the mutilated field equations as describing a spacetime with two congruent sheets (at $u < 0$, and at $u > 0$) cross-connected by a "bridge" at $u = 0$ (hence $r = 2GM$). Despite that g vanishes at the shifted $u = 0$ origin, the unsteadiness singularity condition is invalidated by the mutilated EFE for an empty gravitational field:

> "*The whole field consists of two equal halves, separated by the surface of symmetry* $x_1 = 0$*, such that for the corresponding points* (x_1, x_2, x_3, x_4) *and* $(-x_1, x_2, x_3, x_4)$ *the* g_{ik} *are equal. As a result we find that, although we are permitting the determinant* g *to take the value* 0 *(for* $x_1 = 0$*), no change of sign of* g *and in general no change in the 'inertial index' of the quadratic form occurs.*"[30]

Electrically neutral[k] particles are henceforth described by topological ripples in the singularity-free solution of the new source-free field equations, for which the problem of particle representation and motion coincides. While the topological bridge construction extends the spacetime region from $u > 0$ onto another $u < 0$ (and vice versa), it has later shown to be non-traversible.[31] What remains problematic in the Einstein-Rosen argument is the assessment of whether a spacetime is singular or not in coordinates in which the Schwarzschild metric is not maximally extended. The maximally extended Schwarzschild metric is given by the Kruskal-Szekeres spacetime,[32] which analogously allows for a topological bridge construction connecting two spherically symmetric halves. However the crucial difference to the Einstein-Rosen calculation is that, in the Kruskal-Szekeres spacetime, the symmetric halves are glued together at the $r = 0$ line. In consequence, gravitational singularity formation is not avoided, since the Kretschmann scalar diverges at $r = 0$ in the maximally extended Kruskal-Szekeres spacetime.

[k]The authors move on to consider the case of a charged particle. To account for the electromagnetic charge, they solve the mutilated field equations with a $T^{\mu\nu}$ source term that represents the Maxwell electromagnetic field. Here they introduce a second modification to the EFE, consisting in the change of sign (from positive to negative) of the stress-energy tensor $T^{\mu\nu}$ on the right-hand side of the EFE for the sake of again obtaining the topological bridge as solution to the metric field from the unified treatment of gravitation and electromagnetism. The change of sign, on the other hand, forces Einstein & Rosen to concede the possibility of negative stress-energy.[25]

2.2. *Chandrasekhar's mass limit vs. Eddington's equilibrium hypothesis*

In 1926 Eddington published seminal work[33] on the theory of internal structure, energy generation, radiation transfer and evolution of stars, which remained the primary source of reference for generations of astrophysicists (although the process of nuclear fusion as the dominant supply source of stellar inner energy was not clarified until work of Bethe & Bloch (1939)). Henceforth, the scientific community was equipped with more realistic models of stellar matter, which allowed to address the *astrophysical* viability of the prediction of gravitational singularities in stars modelled by the Schwarzschild ideal hydrodynamics ansatz. The 'standard model' of stellar structure assumed contributions to the Equation of State (EOS) $P = P(\rho, T)$ from classical ideal gas pressure $P \propto \rho T$ and from the Stefan-Boltzmann law of black body radiation pressure $P \propto T^4$ as the main pressure sources that counterbalance gravitational contraction of the star and maintain HE of the perfect fluid in its interior. This 'standard model' was an adequate approximation for most types of observed stars, but not for the faint White Dwarf stars. Based on comparison of observed stellar bolometric luminosity and total mass tuples, Milne (1930) suggested that in order for a steady equilibrium state to always hold, the classical ideal gas law must break down beyond a critical core luminosity L_c, at which he argued that the matter configuration in the White Dwarf star must condense to a quantum degenerate gas.[34]

Breakdown of the (polytropic) Equation of State. Chandrasekhar's achievement consisted in the calculation of an upper mass limit[35,36] for mechanical stability of an electron degenerate White Dwarf star, beyond which the Pauli pressure of electrons is not sufficent to counterbalance collapse of the star under the weight of the nuclei it is composed of. In his calculation, Chandrasekhar models the electrons by a quantum ideal Fermi gas, which respect the Pauli Exclusion Principle (a maximum of two electrons, with anti-aligned spins in the 'up' and 'down' states, are allowed to occupy the same classical phase space cell) and whose energy levels are occupied up to the Fermi energy maximum according to Fermi-Dirac quantum statistics. The Pauli degeneracy pressure relates to electron density ρ_e by the polytropic EOS law $P(r) = K\rho(r)^\gamma$, which is independent of temperature, and where the adiabatic index equals $\gamma = 5/3$ (non-relativistic degeneracy) or $\gamma = 4/3$ (special relativistic degeneracy). The electron density itself is dependent on nuclear composition and density of the White Dwarf, and the parameter K takes this into account. Chandrasekhar works with a simple White Dwarf model composed solely of protons and electrons. In a Newtonian gravity treatment, Chandrasekhar proceeds by imposing the HE condition

$$\frac{dP(r)}{dr} = -\frac{Gm(r)}{r^2}\rho(r) \tag{8}$$

where $m(r)$ is the shell mass up to radius r, obtained from the volume integral over the total density profile $\rho(r)$. He further assumes that the density distribution peaks in the center ($r = 0$) of the White Dwarf. From there he follows that there is an upper limit to total mass M, for any given density gradient $\rho(r)$, at which electron degeneracy pressure is still sufficient to maintain HE. As the star contracts under its own gravity, electrons need to speed up in their motion in order to retain the necessary phase space volume that is required to accomodate all of the electron states in accord with the Pauli principle. He finds that, as *all* the electrons inside the White Dwarf attain relativistic degeneracy, the polytropic index (and thereby also the pressure) decreases, and at a critical Chandrasekhar mass (which, in general, depends on nuclear composition of the star, and for which the solar mass unit $\odot$ is commonly used) the internal radius of the White Dwarf tends to zero:

> "*We are bound to assume therefore that a state must come beyond which the equation of state (...) is not valid, for otherwise we are led to the physically inconceivable result that for* $M = 0.92 \bigodot \beta^{-\frac{3}{2}}, r_1 = 0$, *and* $\rho = \infty$."[37]

Chandrasekhar's fluid ansatz was different from Schwarzschild's in that Schwarzschild calculated the resulting pressure required to sustain HE for any incompressible fluid of given density and radius, while Chandrasekhar instead assumed a relativistic electron degeneracy pressure source to calculate the maximum mass that still maintains the fluid's HE, as function of a specified matter density gradient that peaks in the center. However both the authors arrive at the compatibility of their solution equations with the formation of a singular density ($\rho \to \infty$) point mass, and both suggest to dismiss this conclusion under sacrifice of the scope of validity of their respective solution equations: Schwarzschild's turn was to suggest that the assumption of the incompressible fluid breaks down, while Chandrasekhar's turn instead was to argue that it is not the general framework of the perfect fluid ansatz itself but the polytropic EOS employed therein what breaks down[1]. The Chandrasekhar mass limit improved an earlier upper mass limit estimate

[1]Apart from dismissing the scope of validity of the polytropic EOS, another turn is to reject the presumption that the HE condition continues to be valid as the Chandrasekhar mass is approached. Neither Schwarzschild nor Chandrasekhar question the appropriateness of the HE condition in the aforementioned works *before* the critical mass concentration limit is reached. To anticipate, as is known today from a more realistic treatment of matter in electron degenerate stellar cores, nuclear and particle interactions may trigger hydrodynamic instabilities and the energy release that power explosions as the Chandrasekhar mass is approached. Thermonuclear reactions inside a White Dwarf, composed of -for example- a degenerate Carbon-Oxygen core that is surrounded by an outer shell of mainly Helium ions, under certain yet not well understood conditions[38] (one possible progenitor mechanism is the accretion of hot matter onto the White Dwarf from a companion star) ignite a deflagration or detonation before the Chandrasekhar mass limit is attained, which unbinds (thermonuclear runaway energy release from Carbon fusion is one of the possible explosion mechanisms) the star in a Type 1a Supernova, and leaves no remnant behind. In more massive stars, an (effective) Chandrasekhar mass limit indeed sets the condition for onset of stellar core collapse (CC), onward of which the HE assumption becomes an inappropriate constraint on the physics: violent Neutron Star formation is one of the outcomes of CC, after -according to current

by Stoner (1930) who assumed a constant density fluid. With the theoretical model of gravitational implosion ($r_1 \to 0$) of a sufficiently massive White Dwarf star, he demonstrated that (Newtonian) gravitational singularity formation is not merely a mathematical curiosity but a possible outcome of stellar evolution that connects to stellar progenitor models.

The final equilibrium conjecture. Driven by views of a philosophy of science which exclude the possibility of gravitational implosion, Eddington vehemently dismissed Chandrasekhar's mathematically rigorous calculation of an upper mass limit for mechanical stability of a White Dwarf star. During the meeting of the astronomical society, he expressed concerns about the relativistic degeneracy formula deployed by Chandrasekhar, suggesting that quantum degeneracy is lifted when relativistic effects become important:

> "*The formula is based on a combination of relativity mechanics and non-relativity quantum theory, and I do not regard the offspring of such a union as born in lawful wedlock (...) There is no such thing as relativistic degeneracy!*"[39]

Instead, he conjectured that when the nuclear energy supply is exhausted, the star reaches a final stable equilibrium state at which radiation pressure compensates further gravitational contraction:

> "*The star has to go on radiating and radiating and contracting and contracting until, I suppose, it gets down to a few km radius, when gravity becomes strong enough to hold in the radiation, and the star can at last find peace.*"[39]

His argument was heuristic and not grounded upon an elaborated theoretical final equilibrium model, but it was influential enough to prevent Chandrasekhar's work from receiving a wider attention.

2.3. *Tolman-Oppenheimer-Volkoff mass limit vs. unexplored Equation of State effects*

(More) compact stars. The discovery of the neutron[40] in 1932 opened up new directions in nuclear stellar astrophysics. But even prior to its discovery, Landau

'standard' CC Supernova models- a neutrino-driven convection-enhanced explosion has ejected a significant fraction of mass from the collapsing core. The explosive mass ejection allows the stellar post-collapse remnant find a new HE state at ultra-dense matter conditions, which -in the simplest treatment- is again modelled by a polytropic Equation of State: HE is now sustained by neutron degeneracy pressure of a quantum ideal gas composed of neutrons.

presented a model[m] of a more compact stellar core, taking form of a "*gigantic nucleus*",[41] composed of a stable matter configuration at which the protons and electrons still co-exist at densities even greater than the limiting $\rho_c \propto 10^{12} \frac{g}{cm^3}$ assumed in the original Chandrasekhar mass White Dwarf. While Landau did not explicitly invoke a neutron composition in his stellar core model, he anticipated the idea that stable matter configurations at even higher densities than the White Dwarf do exist, and still are adequately approximated by a polytropic EOS, taking degeneracy pressure into account. Few years later, in 1934 the astronomers Baade and Zwicky suggested the explosive transformation of an ordinary star into a Neutron Star as an explanation of observations of supernova transients.

> "*If super-novae are giant analogues to ordinary novae, we may expect that ionized gas shells are expelled from them at great speeds. If this assumption is correct, part of the cosmic rays should consist of protons and heavier ions. [...] We have tentatively suggested that the super-nova process represents the transition of an ordinary star into a neutron star. If neutrons are produced on the surface of an ordinary star they will 'rain' down towards the center if we assume that the light pressure on neutrons is practically zero. This view explains the speed of the star's transformation into a neutron star.*"[42]

The details of the nuclear physics involved therein, in particular the transition of the protons, bounded together by the strong nuclear force, into neutrons by virtue of undergoing Electron Capture, were developed in subsequent years by Hund (1936) and by Gamow (1937).[43] Because of the half-integer spin of the neutrons, Pauli pressure of the neutron ideal Fermi gas, rather than of the electron gas, is responsible for maintaining stability of the compact star, withstanding further gravitational contraction. Oppenheimer & Serber (1938) re-estimated the upper mass limit for a stable degenerate neutron core modelled by a polytropic EOS, but corrected the result of Landau, which posed too tight constraints on nuclear binding energy.[44]

Unexplored EOS effects. A treatment of degenerate neutron cores in General Relativity followed soon hereafter: it was based on Tolman's solution ansatz to the interior Schwarzschild field of a spherically symmetric static matter distribution[45,46]:

$$ds^2 = e^v dt^2 - \frac{1}{1 - \frac{2Gm(r)}{r}} dr^2 + r^2 g_\Omega \tag{10}$$

[m] He arrives at the important formula

$$M_0 = \frac{3.1}{m^2} \left(\frac{\hbar c}{G} \right)^{3/2} \tag{9}$$

where M_0 is the limiting mass that is sufficient to maintain the (Newtonian) HE, and the parameter m is the mass of a nuclear ion per one electron. For a degenerate White Dwarf core, $m \simeq 2m_p$ where m_p is the proton rest mass, and from there follows the well-known $M_0 \simeq 1.44 M_\odot$ Chandrasekhar limit.

where $m = m(r) = \int_0^r dr' 4\pi r'^2 \rho(r')$ is the volume integral over energy density up to the radial coordinate r, and where the coefficient $v = v(r)$ obtained from the fluid's EOS. Oppenheimer & Volkoff (1939) adopted this ansatz for their calculation of the gravitational field inside neutron cores,[47] which they assumed to have density $\rho(r)$, isotropic pressure $P(r)$, be at rest ($u^\mu = 0$) and under the absence of mechanical stresses. Under these assumptions, the stress-energy tensor of the perfect fluid

$$T^{\mu\nu} = \left(\rho(r) + \frac{P(r)}{c^2} u^\mu u^\nu - P(r) g^{\mu\nu} \right) \tag{11}$$

simplifies, with $T^1_1 = T^2_2 = T^3_3 = -P$ and $T^4_4 = \rho$ remaining as the only non-vanishing diagonal components. Plugging them and Tolman's metric ansatz into the EFE $R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$, they derive a general relativistic HR condition, which -in a modern reformulation of the solution equations- reads

$$\frac{dP(r)}{dr} = -\frac{Gm(r)}{r^2} \rho(r) \left(1 + \frac{P(r)}{\rho(r) c^2} \right) \left(1 + \frac{4\pi r^3 P(r)}{m(r) c^2} \right) \left(1 - \frac{2Gm(r)}{rc^2} \right)^{-1}. \tag{12}$$

In the general relativistic HE, both the stress-energy field *and* the gravitational field are stationary. The solution to the TOV equation (12) requires the supply of an EOS $P = P(\rho, T)$, which reduces to $P = P(\rho)$ if temperature-independent (as in the polytropic case), and which coincides with Schwarzschild's 1916 solution to the interior field if the constant density assumption $\rho(r) = \rho_0$ is made. In their calculation of the upper mass limit to general relativistic stability, Oppenheimer & Volkoff model the neutron core by a variant of a more realistic polytropic EOS than Landau's, which however only takes into account Fermi degeneracy pressure of the neutron gas, and neglects contributions to the EOS from thermal energy and internal particle interaction forces. Under these assumptions, they derive a general relativistic upper mass limit $M_{TOV} \leq \frac{3}{4} M_\odot$. At the same time, they comment on solutions to the field equations for greater core masses:

> "*For masses greater than $\frac{3}{4} M_\odot$ there are no static equilibrium solutions. (...) There would then seem to be only two answers possible to the question of the 'final' behaviour of very massive stars: either the equation of state we have used so far fails to describe the behaviour of highly condensed matter that the conclusions reached above are qualitatively misleading, or the star will contract indefinitely (...) although more and more slowly, never reaching true equilibrium.*"[47]

The Tolman-Oppenheimer-Volkoff mass limit is seen as a first general relativistic (albeit rather coarse, given a simplistic treatment of matter in the neutron core) demarcation between a Neutron Star vs. Black Hole final fate of core collapse. As an alternative to complete gravitational collapse, the authors trace out a similar argument to Chandrasekhar's, namely that the EOS used in their calculation might

become inadequate at more extreme matter conditions[n] and that yet unknown pressure sources (which not necessarily need to arise from fermionic degeneracy pressure) might be responsible for restoring HE of the star. In that case, the TOV condition eq. (12), and resulting mass limit, needs be re-evaluated with a yet unexplored EOS, in place of the polytropic ansatz which accounts for fermionic degeneracy pressure.

3. Oppenheimer-Snyder collapse (1939–1965)

3.1. *Indefinite gravitational contraction*

After Oppenheimer and Volkoff established the upper mass limit for stability of a neutron core in 1939, Oppenheimer and his student Snyder that same year sought for solutions to the EFE for the case that the TOV mass limit is exceeded by polytrope fluid models of mass $M > M_{TOV} = 0.75\,M_{\odot}$, beyond neutron star densities ($\rho_c \propto 10^{15} \frac{g}{cm^3}$). Their result is that they do not find any stable matter states that allow for *stationary* solutions to the field equations for $g_{\mu\nu}$ that would respect the TOV equilibrium condition. Instead, under the idealizing assumption of vanishing pressure $P(r) = 0$, they find an analytical *evolved* solution to the gravitational field, describing gravitational collapse of freely falling matter henceforth simply modelled as dust. They reformulate Tolman's ansatz to the interior Schwarzschild metric (eq. (10)) in a coordinate system that is co-moving ($u^{\mu} = 0$) with the collapsing matter. As result, the stress-energy tensor again takes the simple form of eq. (11), but -given the zero pressure assumption- the only non-vanishing component is the density $T^4_4 = \rho$. The gravitational contraction of the star is continued indefinitely: as the continued spherically symmetric gravitational contraction shrinks the sphere of dust down to asymptotically zero, the central density diverges, local measures of

[n]When taking a leap further ahead into the future, the reliance of compact star models on developments in nuclear and particle physics becomes evident: after the postulation of the subnuclear quark particles[48,49] as constituents of hadronic matter and therefore of nucleons in 1964, Ivanenko & Kurdgelaidze in 1965 constructed a simple Quark Star model.[50] With the discovery of the theoretical description of the constituents of matter at a higher energy scale, a novel model for HE, sustained by degeneracy pressure of the constituent fermions, becomes conceivable. The Quark Star model assumes that the quarks constituting the neutron, bound together by the strong interaction at coupling strength $\alpha_s(k^2)$ that logarithmically decreases with scale, at high enough temperatures undergo a deconfinement phase transition to asymptotic freedom. The deconfinement of quarks justifies application of, once again, the quantum ideal Fermi gas model. The free quarks collectively reach a stable subnuclear HE configuration sustained by degeneracy pressure of the fermionic quark gas. Whether or not Quark Stars exist in nature is an unresolved scientific enquiry, especially in light of the fact that according to more recent, state-of-art models of Neutron Stars, the EOS of their (possibly deconfined and quark degenerate) cores is not well understood.[51]

The lower the mass, the lower does the quark degeneracy pressure need to be in order to maintain HE. Putatively, Quark Stars are therefore physically reasonable mimickers of very *low* mass Black Hole, but not of intermediate mass, massive, and supermassive Black Hole candidates. It is unlikely that the degeneracy pressure from any of the established fermions known in today's Standard Model of elementary particle physics is sufficient for the corresponding exotic compact star to act as supermassive Black Hole mimicker. However, certain models of compact Boson Stars open up new directions for physical speculation.

curvature diverge, and incomplete geodesics form.[o] The admission of the formation of a point mass singularity by the solution equations has been found before in the perfect fluid models of Schwarzschild, Chandrasekhar, Landau and Oppenheimer & Volkoff. What is remarkable[p] about the Oppenheimer & Snyder solution is that, along with the curvature singularity (which is, by virtue of the EFE, consequence of a locally divergent stress-energy field), their calculations also predict geodesic incompleteness.

Non-vanishing pressure and mass loss effects. The Oppenheimer-Snyder type dust collapse model fulfills the conditions for the Raychaudhuri-Komar singularity theorem to apply, which implies formation of an infinite density point mass provided that the Weak Energy Condition (see Sec. (3.3)) holds: the Oppenheimer-Snyder collapse assumes a perfect fluid, which is irrotational, and geodesic (the absence of proper acceleration is consequence of the $P = 0$ assumption). From an infinite density point mass, the EFE imply a *curvature* singularity. Though what is more, the authors affirm their belief in the validity of the solution equations even for the case of non-zero pressure $P \neq 0$:

> "*We believe that the general features of the solution obtained in this way give a valid indication even for the case that the pressure is not zero, provided that the mass is great enough to cause collapse (...) Of course, actual stars would collapse more slowly than the example which we studied analytically because of the effect of the pressure of the matter, of radiation, and of rotation.*"[52]

At the same time, they enumerate a number of physical mechanisms that could prevent a star from surpassing the TOV mass limit, after all of its thermonuclear internal energy resources have been exhausted:

(1) the fission of the star into fragments by its rotation,

[o] Setting the proper time τ of a local observer co-moving with the collapsing matter of the outermost dust shell in relation with the coordinate time measured by a distant observer t outside the collapsing star, and taking the asymptotic limit $t \to \infty$, they find that proper time of a time-like geodesic converges to a finite $\tau \to \tau_0$. As the collapsing star shrinks beneath its Schwarzschild radius, time-like geodesics become space-like. The authors calculate the asymptotic space-like configuration of the gravitational field inside the collapsing star: the temporal component $g_{tt} \to 0$ vanishes while the radial component g^{rr} reaches a finite configuration in the interior $r < r_S$ except at $r = 0$ where it becomes infinite. In addition, a coordinate singularity develops at $r = r_S$.

[p] The authors also anticipate the idea that the two-sphere $r = r_S$ constitutes an event horizon, preventing an escape from the collapsing star and isolating it from any causal interaction with the exterior:

> "*After this time τ_0 an observer comoving with the matter would not be able to send a light signal from the star; the cone within which a signal can escape has closed entirely.*"[52]

(2) photon radiation pressure blowing off outer mass layers,
(3) significant enough mass radiation.

From the perspective of modern astrophysical stellar evolution models of the late pre-collapse evolutionary stages, the above mentioned effects are unlikely to be strong enough to be a general argument preventing implosion of a physically realistic massive core.[q]

3.2. *The Landau-Wheeler arguments, and Boson Stars*

Landau pointed out in his 1932 and 1938 papers that a compression of fermions, over course of stellar gravitational collapse, to a point mass, with all fermions occupying the same phase space cell, leads to violation of fermionic quantum statistics,[41,56] which prohibits occupation of the same quantum state by more than one fermion. The formation of a point mass singularity is further complicated by other quantum principles. In particular, the Heisenberg uncertainty principle, according to which position and momentum of a particle cannot both be precisely constraint at the same time. Energy compactness considerations, advocated by Wheeler, according to which a quantum particle of given rest-mass energy is not allowed to localize at scales smaller than its Compton wavelength in order to be well-defined. Wheeler confronted Oppenheimer at the Solvay Conference (1958) to argue on heuristic grounds that the collapsing (fermionic) nuclear matter must dissolve into electromagnetic, gravitational and/or neutrino radiation.[57] In a way, his opposition to Oppenheimer's result was akin to Eddington's to Chandrasekhar's work, in that Wheeler took up Eddington's idea that radiation becomes held in the star, which though still maintains its final equilibrium. But Wheeler went further with an elaborated model of a *Geon* as a HE candidate. In 1955, a decade before the stationary

[q](1): According to more recent computer simulations of rotating core collapse that deploy a polytrope EOS,[53] the infalling matter is indeed fragmented into an inner core contracting subsonically and an outer core contracting supersonically. The centrifugal force has the effect of deforming the inner core, and reducing the available gravitational energy to 'push' the outer core into an explosion. Rotation therefore has impact on the fate of the collapsing core regarding whether it explodes or not, but is not an implosion preventing mechanism that generalizes. (2): The high energy photon radiation in the hot stellar core leads to photo-disintegration of heavy nuclear ions (mostly iron group elements), which a more realistic massive core is composed of. Therefore, the photo-disintegration facilitates rather than prevents gravitational collapse, for it occurs at the price of taking away internal energy from the star: thermal radiation pressure contributes to the total pressure that maintains HE, and photo-disintegration has the overall effect of reducing it.[54] (3): Stellar winds in late phase massive star evolution do significantly contribute to mass loss, however the mass radiation mostly takes away mass from the envelope rather than from the core of the star, which -due to its greater compactness- is stronger bound gravitationally. Mass ejection from the core is strong enough only in the case of a CC Supernova explosion, signalling the violent birth of a proto-Neutron Star. However, a successful launch of an explosion is only one of the possible outcomes of stellar CC, along with a Failed Supernova, a Weak Supernova that precedes Black Hole formation by fallback of ejected matter onto the proto-Neuton Star, and the direct quiescent Black Hole formation.[55]

Newman-Kerr electrovacuum solution was found, Wheeler presented a solution to the EFE describing a spherically symmetric confined classical radiation field. His idea was that radiation is trapped in form of standing waves in closed circular toroid tracks of radius of the Geon, and collectively arranges to form stable macroscopic objects.[58] These Geon configurations were supposed to consist of either of trapped electromagnetic waves or of trapped gravitational waves, held together by their own self-gravity in a stable equilibrium state. Since their reduced wavelength is negligibly small compared to the radius, the Geon decay rate is negligible over long time scales. By proposing a non-singular bosonic HE alternative to gravitational implosion, Wheeler further elaborated not only Eddington's but also Landau's argument. Though the transition of fermionic mass-energy to bosonic states in a collapsing star still remained a conjecture, without a specified generic transition mechanism. Alas, his model -as well as variants thereof- turned out to be unstable: a small equilibrium perturbation to the confined radiation field leads to onset of its implosion.[r]

Following a different approach in line with Wheeler's *Geometrodynamics* research program, Wheeler together with his student Fuller returned to the idea of a non-singular Einstein-Rosen bridge connecting two symmetrical regions of an empty spacetime, and studied Wormhole solutions in the maximally extended Kruskal-Szekeres metric, with an introduced minimum proper circumference at the connecting "throat" in order to obtain a singularity-free metric.[31] But it was found that these suffer from qualitatively the same problem as Geon models, namely that Wormholes in 4-dimensional, general relativistic spacetime are unstable against equilibrium perturbations in a dynamic problem (which results in shrinkage of circumference of the throat down to zero, and divergence of curvature invariants after finite proper time), unless even higher-dimensional space-time or quantum theoretical considerations are taken into account that stabilize the topological bridge.

[r]However, when a complex scalar field φ is considered instead of the electromagnetic field, the equation of motion for the complex scalar field (the Klein-Gordon equation) admits stable geon-like soliton solutions, stabilizing as a Boson Star. Their gravitational field is sourced by the self-gravity of the bosonic waves, which are dispersed in accordan with the Klein-Gordon equation, satisfying the laws of Bose-Einstein quantum statistics: bosonic particles of mass μ confined within an object of size $2R$ possess a proper velocity $v \propto (2R\mu)^{-1}$. It is the Heisenberg uncertainty pressure of the bosons which provides the pressure (along with a self-interaction term, which adds either a positive or a negative contribution to pressure, depending on the nature of the self-scattering potential $V(\varphi)$) to counterbalance contraction and maintain HE. However, given a limitation by the finite internal radius, and thereby the proper velocity of the Bosons, there is a maximum mass to stability of a Boson Star, $M_{BS} \propto \frac{m_P^2}{\mu}$ (whose precise value is determined from specific boson star models), beyond which the Heisenberg uncertainty pressure is insufficient to withstand gravitational collapse (m_P is the Planck mass).[59] It therefore seems not implausible that Boson Stars, even if they exist, are only an intermediate stable configuration state, before gravitational collapse of the compact Boson Star occurs if it has become sufficiently massive.

3.3. *The Khalatnikov-Lifshitz asymmetry argument, and Penrose's 1965 theorem*

The Oppenheimer & Snyder paper convinced Landau about the reality of complete gravitational collapse of matter, but not of the irregularities in the metric field by the formation of singularities (referring to all three notions defined in Sec. 1). As one of the leading figures of the Soviet school assessed the prediction of gravitational singularities in collapsing stars as well as in Early Universe Cosmology to be the major unsolved problem of theoretical physics, his judgement would shape the thinking of a generation of Soviet theorists. The Landau school accepted the reality of gravitational collapse sooner than the Wheeler school, and would endow it with the name "Frozen Star", grounded upon effects at the Schwarzschild boundary, apparent to an external observer, rather than interior singularity hallmarks.[s] Doubting singularity formation as physically viable, Khalatnikov and Lifshitz in the early 60s advocated that in any physically realistic stellar gravitational collapse scenario, the condition of spherical symmetry, inherent in the Schwarzschild metric, would be violated.[60] In realistic asymmetric collapse, they argued that nonlinearity effects would dominate the interaction between gravitational field and matter, resulting in chaotic dynamics and impeding formation of the central singularity. Instead, they interpreted the central singularity as an artifact prediction that arises from the imposed spherical symmetry condition in the Oppenheimer-Snyder collapse.

Penrose's work on the singularity theorem[61] historically was motivated as a counter to the Khalatnikov-Lifshitz argument. The theorem states that, provided that a set of conditions are fulfilled, the production of geodesic incompleteness singularities is generic to General Relativity theory, independent of spherical symmetry assumptions:

> *The question has been raised as to whether this singularity is, in fact, simply a property of the high symmetry assumed. The matter collapses radially inwards to the single point at the center, so that a resulting space-time catastrophe there is perhaps not surprising. Could not the presence of perturbations which destroy the spherical symmetry alter the situation drastically? [...] It will be shown that, after a certain critical condition has been unfilled, deviations from spherical symmetry cannot prevent space-time singularities from arising.*[61]

[s] Characteristics of the Frozen Star definition are

- the infinite gravitational redshift of light signals, down to zero, emitted from the Schwarzschild boundary, and
- the illusory coordinate singularity effects that, as measured in coordinate time of a distant external observer, it takes infinitely long for the outermost coontracting stellar surface to collapse under the star's Schwarzschild boundary; as well as the infinite gravitational time delay of a test object asymptotically approaching the Schwarzschild boundary from outside but never reaching it.

Penrose starts off with three premises (geodesic completeness, the weak energy condition, and the causality condition) and shows that they lead to a contradiction, provided that a closed future-directed trapping surface, which is not necessarily spherically symmetric, has formed. The resolution of the contradiction requires that one of the premises is refuted. If one holds onto the latter two conditions, then the geodesic completeness assumption needs be rejected. Put differently, the conditions that need be fulfilled in order for the singularity theorem to apply are

- *Weak Energy Condition*: $R_{\mu\nu}\dot{\gamma}^{\mu}\dot{\gamma}^{\nu} \geq 0$, strictly speaking a geometrical condition that demands the convergence of null-like and time-like vector fields γ.[t]
- *Causality Condition*: the spacetime has an initial Cauchy hypersurface, and by global hyperbolicity is time-orientable (no closed null-like or time-like curves).
- *Boundary Condition*: formation of a closed future-trapped surface.

Assuming that the above conditions are satisfied throughout the entire course of stellar gravitational collapse, Penrose's theorem guarantees formation of incomplete null geodesics[u]. For the case of a collapsing star modelled by the Schwarzschild ansatz, the outermost trapped surface forms at $r_b = r_S$ where r_b is the outer boundary of the spherically symmetric collapsing star. Under which generic conditions a trapping surface forms, and whether it is reasonable to assume that the weak energy and causality conditions apply throughout the entire course of gravitational core collapse, are separate research questions that require further investigation, going beyond the scope of this work.

The geodesic incompleteness singularity theorem is independent of the EFE. Therefore, it applies to any theory of gravity, e.g. $f(R)$ theories, formulated using the framework of Lorentzian geometry, that satisfies the stated causality, energy and boundary conditions. The Penrose theorem does not state 'where' in space-time the incomplete geodesics locally form. Moreover, the theorem makes no statement about local curvature divergence. The lack of a physical characterization of geodesic incompleteness formation is by some (e.g. Landau & Lifschitz in Ref. 62) seen as weakness of the theorem - the theorem does not exemplify the reasons for which the geodesic incompleteness occurs: whether or not the path inextendibility is due to

[t] In GR, the Ricci tensor $R_{\mu\nu}$ is related to $T_{\mu\nu}$ by the EFE, which gives the geometrical condition a conceptual interpretation in terms of energy related physical quantities.

[u] A key tool used in the formulation and proof of the singularity theorems is the focusing effect of the Raychaudhuri equation. The equation describes the divergence behaviour of a congruence family of geodesics. The divergence of a congruence is defined as the derivative of the logarithm of the determinant of the congruence volume. Raychaudhuri's equation then describes the evolution of geodesic congruences. The closed trapping surface condition implies that light rays emitted in opposite directions orthogonal to a closed space-like 2-sphere converge. Null geodesics include the entire boundary of the proper future: when they converge from opposite sides, there is no more future boundary. The congruence volume of parallel null geodesics reaches zero within finite affine length. Consequence is the generic prediction that the relativistic spacetime $(\mathcal{M}, g)$ is not future null geodesically complete.

curvature divergence, is left open. The latter fact can, however, instead be seen as a strength of the theorem, in that it allows to treat the geodesic incompleteness and infinite curvature singularity notions apart from one another: the Penrose theorem does not need to relate geodesic incompleteness to any pathological mass-energy or curvature state in order to apply.

3.4. *The Planck-Wheeler length, and Markov's curvature bound*

Curvature fluctuations at Planck scale. Together with his students Harrison and Wakano, Wheeler in 1958 mapped out, in a research study that combines theoretical and numerical work, a catalogue of the endpoints of stellar thermonuclear evolution for a wide range of stellar masses.[57] The catalogue played a pivotal role in his later conviction that along with the White Dwarf and Neutron Star compact remnant endpoints of stellar evolution, the complete gravitational collapse as predicted by the Oppenheimer & Snyder model, although it makes crude approximative assumptions, is a possible outcome of evolution of massive stars. Along with exclusion of alternatives to complete gravitational collapse, the decisive role[64] for Wheeler's conviction was played by computer simulations of CC, in particular those of Colgate & White (1966),[63] which adopted more realistic EOS models of stellar matter, and thereby undermined Wheeler's earlier objection that the Oppenheimer-Snyder collapse model is physically unreasonable. Wheeler's updated stance on the issue of the final fate was that mass-energy disappears, as it gets torn into pieces by the divergent tidal forces at the center of mass. The complete gravitational collapse results in a gravitating vacuum ($T_{\mu\nu} = 0$), which he later called "*mass without mass*", or "*disembodied mass*".[65] However, with regard to the prediction of curvature singularity formation as part of the final fate, Wheeler expressed mistrust in GR as the appropriate theory of gravity at smallest scales. His conviction was that the Planck length $\ell_P = \sqrt{\frac{\hbar G}{c^3}}$ is the minimum scale at which space is resolved in an ultimate theory of gravity, and at which he expected space to exhibit "*quantum foam*"[66] of random curvature fluctuations, including microscopic wormholes. Whether or not a gravitational singularity forms over course of unhalted implosion would then be assessed by a successor theory, after the "*fiery marriage*" of quantum theory and gravitation. The important point he raised is that -while he accepted reality of gravitational collapse- the curvature singularity might be an artifact prediction that arises neither because a yet unexplored pressure source has not been taken into account, nor because of idealization, but because the theory of gravity itself breaks down.

Maximal energy density. Aiming at the same goal (the theoretical preclusion of curvature singularity formation) but following a different line of argument, Markov in 1947 first stated the conditions under which there arises an upper limiting bound to curvature,[67] an idea he would explore and elaborate further over the subsequent

decades (see e.g. Refs. 68, 69) His proposal was that a limiting quadratic curvature invariant

$$R_{\mu\nu\lambda\delta}R^{\mu\nu\lambda\delta} \leq \frac{1}{\ell_P^4} \tag{13}$$

is consequence of a maximal density $\rho_q = \frac{c^5}{G^2\hbar} \simeq 10^{94}\frac{g}{cm^3}$ of mass-energy that is permitted to occur in nature. The limiting energy density, a construction from the ratio of the three natural constants c, $\hbar$ and G of special relativity (SR), quantum theory (QT) and gravitation, remains a conjecture. Though he gives a motivation for it, from consideration of the underlying principles of the theories that provide the limiting constraints: from SR, that the propagation velocity of any signal cannot exceed c (a principle held so sacred by Einstein); and from QT, that the action cannot be less than $\hbar$ (a condition which is violated in classical physics). At the same time, Markov highlights that it is the nonlinearity in the differential gravitational field equations coupling stress-energy to measures of curvature, which needs to supply the energy cut-off to avoid UV divergences and the renormalization problem of quantum field theory.[v]

4. Conclusions

The decades spanning in between Schwarzschild's analytical 1916 solution to the EFE and Penrose's 1965 singularity theorem produced a rich pool of thought about the nature of gravitational singularities, and about the viability of their physical realization in stellar gravitational collapse. The examination of the arguments in favour or against realism on singularity formation requires to take into account considerations from differential geometry, stellar astrophysics, and history of science. We have investigated a long list of singularity formation counter-arguments and, while telling the three gravitational singularity notions apart, came to the following conclusions:

- Geodesic incompleteness is an innovative (as it is not available in Newtonian gravity), the most general and a non-local singularity characterization, which is compatible with but does not require divergence in any curvature quantities, and which does not prima facie contradict principles neither of general relativity nor of quantum theory. Prediction of geodesic incompleteness by Penrose's theorem therefore does not indicate breakdown of GR.

[v]As the maximal density limit is reached, he argues that a phase transition of matter-energy into a de Sitter vacuum takes place. The latter conjecture has been be elaborated by Gliner and Sakharov in 1966, with applications to singularity resolution in Cosmology: the negative pressure $P = -\rho < 0$ obeyed by the false vacuum $T_{\mu\nu} = \Lambda g_{\mu\nu}$ introduces a repulsive term into the field equations and violates, by virtue of the EFE with Λ included on the right-hand side, the weak energy condition. With the unclarified nature of Λ, the de Sitter phase transition is a non-singular albeit speculative outcome of gravitational collapse, whose investigation goes beyond the scope of this work.

- An infinite quadratic curvature invariant is not a trustworthy prediction of GR, because it is a local singularity characterization evaluating properties of the gravitational field down at microscopic scales below the Planck length, while GR is a phenomenological theory of gravity. A microscopic theory of gravity theory, perhaps a unified field theory Einstein and Wheeler were questing for, would reconcile the physics of the gravitational field itself with principles of quantum theory. In addition, there are several sub-classes of infinite curvature singularities, which are problematic in their own right.
- Curvature singularity avoidance is achieved in general not by introduction of a yet unexplored speculative EOS which either appropriately describes mass-energy at extreme conditions and restores HE in a collapsing star, or implies a phase transition that introduces a repulsive term by virtue of the EFE (even if there are such remedies, GR would still allow for analytic solutions to EFE with singular curvature, which remains problematic), but by modification of the gravitation theory itself. Markov's bound on the quadratic curvature invariant is consequence of a limiting mass-energy density conjecture, from which it is derived. However, theoretical work has shown that -under certain conditions- curvature singularities form in GR even from a vacuum spacetime, as in the case of colliding gravitational waves, independent of matter. We therefore propose to retain Markov's conclusion (eq. 13) but to reverse the logic: a limiting density of both stress-energy and of gravitational energy is consequence of a bound on the quadratic curvature invariant. The curvature bound needs be introduced from first principle quantization considerations, which GR -as a macroscopic theory- does not itself supply from within the gravity theory.
- Curvature divergence signals the breakdown of Lorentzian geometry, independent of the EFE: the continuous nature of Lorentzian geometry permits gravitational contraction in Oppenheimer-Snyder collapse be continued indefinitely, without that a limiting scale is ever reached. GR does not cure a problem it itself inherits from Newtonian gravity: since in either theory measures of distance are non-quantized, a curvature pathology in GR is permitted to occur in GR for the same reason that allows for infinite force in Newtonian gravity. If quantum theory is not well-defined at scales smaller than the Planck length, it is not evident why GR should be. In addition, it is the lowest scale at which Lorentz symmetry applies. ℓ_P therefore is a natural unit measure for quantization of the quadratic curvature invariant, which can only be achieved by a quantizing tessellation of Lorentzian geometry in units of Planck length and Planck time. Assuming that physics at scales below the Planck scale is meaningless, the curvature bound ($R_{\mu\nu\lambda\delta}R^{\mu\nu\lambda\delta} = \frac{1}{\ell_P^4}$ limit) impedes formation of any energy density singularity in any inertial frame, because it makes space contraction to a point an impossibility.

Acknowledgments

The author is grateful to Tushar Menon, Erik Curiel, Fritz Roepke, Fabian Schneider, Dennis Lehmkuhl, Roger Penrose, Martin Lesourd and Juliusz Doboszewski for discussions on related topics, and to Siobhan Roberts for the motivation to write. The work of K. M. is supported by the Klaus Tschira Foundation.

References

1. E. Curiel, The Analysis of Singular Spacetimes, *Philosophy of Science* **66**, S119 (sep 1999).
2. J. Earman and J. Eisenstaedt, Einstein and Singularities, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **30**, 185 (jun 1999).
3. S. W. Hawking, G. F. R. Ellis and R. K. Sachs, The Large Scale Structure of Space-Time, **27**, 91 (apr 1974).
4. G. F. R. Ellis and B. G. Schmidt, Singular space-times, **8**, 915 (nov 1977).
5. A. Raychaudhuri, Relativistic Cosmology. I, **98**, 1123 (may 1955).
6. A. Komar, Necessity of Singularities in the Solution of the Field Equations of General Relativity, **104**, 544 (oct 1956).
7. J. M. M. Senovilla and D. Garfinkle, The 1965 Penrose singularity theorem, *Classical and Quantum Gravity* **32**, p. 124008 (jun 2015).
8. J. O. Weatherall, What Is a Singularity in Geometrized Newtonian Gravitation?, *Philosophy of Science* **81**, 1077 (dec 2014).
9. F. J. Tipler, Singularities from colliding plane gravitational waves, *Phys. Rev. D* **22**, 2929 (dec 1980).
10. U. Yurtsever, Structure of the singularities produced by colliding plane waves, *Physical Review D* **38**, 1706 (sep 1988).
11. T. Wang, J. Fier, B. Li, G. Lü, Z. Wang, Y. Wu and A. Wang, Singularities of plane gravitational waves in Einstein's general relativity, *General Relativity and Gravitation* **52** (feb 2020).
12. K. Q. Abbasi, I. Hussain and A. Qadir, Probing Szekeres' colliding sandwich gravitational waves, *The European Physical Journal Plus* **136** (may 2021).
13. C. J. Clarke, *The analysis of space-time singularities*, no. 1 (Cambridge University Press, 1993).
14. C. W. Misner, *Taub-Nut Space as a Counterexample to almost anything*, in *Relativity Theory and Astrophysics. Vol.1: Relativity and Cosmology*, ed. J. Ehlers 1967, p. 160.
15. B. Schmidt, A new definition of singular points in general relativity, *General relativity and gravitation* **1**, 269 (1971).
16. J. Earman, *Bangs, crunches, whimpers, and shrieks: Singularities and acausalities in relativistic spacetimes* (Oxford University Press, 1995).
17. R. Geroch, What is a singularity in general relativity?, *Annals of Physics* **48**, 526 (1968).
18. J. K. Beem, Some examples of incomplete space-times, *General Relativity and Gravitation* **7**, 501 (jun 1976).
19. I. Stewart, Physical infinity, in *Infinity: A Very Short Introduction*, (Oxford University Press, mar 2017) pp. 91–102.

20. K. Schwarzschild, Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften Berlin*, 189 (1916).
21. J. Droste, The field of a single centre in Einstein's theory of gravitation, and the motion of a particle in that field, *Ned. Acad. Wet., SA* **19**, p. 197 (1917).
22. K. Schwarzschild, Über das Gravitationsfeld einer Kugel aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie, *Sitzungsberichte der königlich preußischen Akademie der Wissenschaften zu Berlin*, 424 (1916).
23. A. Einstein, On a Stationary System With Spherical Symmetry Consisting of Many Gravitating Masses, *The Annals of Mathematics* **40**, p. 922 (oct 1939).
24. A. Einstein, Letter to Ernst Mach. Zurich, 25 June, reported at pag. 544 of. CW Misner, KS Thorne, and JA Wheeler, Gravitation (1973).
25. J. Eisenstaedt, Dark Bodies and Black Holes, Magic Circles and Montgolfiers: Light and Gravitation from Newton to Einstein, *Science in Context* **6**, 83 (1993).
26. A. Einstein and J. Grommer, Allgemeine Relativitätstheorie und Bewegungsgesetz, *Albert Einstein: Akademie-Vorträge: Sitzungsberichte der Preußischen Akademie der Wissenschaften 1914–1932*, 2 (1927).
27. A. Einstein and L. Infeld, On the motion of particles in general relativity theory, *Canadian Journal of Mathematics* **1**, 209 (1949).
28. M. Black, A. Einstein and P. A. Schilpp, Autobiographical Notes, *Journal of Symbolic Logic* **15**, p. 157 (1949).
29. L. Silberstein, Two-centers solution of the gravitational field equations, and the need for a reformed theory of matter, *Physical Review* **49**, p. 268 (1936).
30. A. Einstein and N. Rosen, The Particle Problem in the General Theory of Relativity, *Physical Review* **48**, 73 (jul 1935).
31. R. W. Fuller and J. A. Wheeler, Causality and Multiply Connected Space-Time, *Phys. Rev.* **128**, 919 (oct 1962).
32. M. D. Kruskal, Maximal Extension of Schwarzschild Metric, *Phys. Rev.* **119**, 1743 (sep 1960).
33. A. S. Eddington, *The Internal Constitution of the Stars*, 1926).
34. C. R. Almeida, Stellar equilibrium vs. gravitational collapse, *The European Physical Journal H* **45**, 25 (feb 2020).
35. S. Chandrasekhar, The Maximum Mass of Ideal White Dwarfs, *The Astrophysical Journal* **74**, p. 81 (jul 1931).
36. S. Chandrasekhar, XLVIII.The density of white dwarf stars, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **11**, 592 (feb 1931).
37. S. Chandrasekhar and E. A. Milne, The Highly Collapsed Configurations of a Stellar Mass, *Monthly Notices of the Royal Astronomical Society* **91**, 456 (mar 1931).
38. W. Hillebrandt, M. Kromer, F. K. Röpke and A. J. Ruiter, Towards an understanding of Type Ia supernovae from a synthesis of theory and observations, *Frontiers of Physics* **8**, 116 (april 2013).
39. January 11 Meeting of the Royal Astronomical Society, *The Observatory* **58**, 33 (1935).
40. J. Chadwick, Possible Existence of a Neutron, *Nature* **129**, 312 (feb 1932).
41. L. D. Landau, On the theory of stars, *Phys. Z. Sowjetunion* **1**, p. 152 (1932).
42. W. Baade and F. Zwicky, Remarks on Super-Novae and Cosmic Rays, *Physical Review* **46**, 76 (jul 1934).
43. D. G. Yakovlev, P. Haensel, G. Baym and C. Pethick, Lev Landau and the concept of neutron stars, *Physics-Uspekhi* **56**, p. 289 (2013).
44. J. R. Oppenheimer and R. Serber, On the Stability of Stellar Neutron Cores, *Physical Review* **54**, 540 (oct 1938).

45. R. C. Tolman, Effect of Inhomogeneity on Cosmological Models, *Proceedings of the National Academy of Sciences* **20**, 169 (mar 1934).
46. R. C. Tolman, Static Solutions of Einstein's Field Equations for Spheres of Fluid, *Physical Review* **55**, 364 (feb 1939).
47. J. R. Oppenheimer and G. M. Volkoff, On Massive Neutron Cores, *Physical Review* **55**, 374 (feb 1939).
48. G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, in *Developments in the quark theory of hadrons. Vol. 1. 1964 - 1978*, eds. D. B. Lichtenberg and S. P. Rosen 2 1964, pp. 22–101.
49. M. Gell-Mann, A schematic model of baryons and mesons, *Physics Letters* **8**, 214 (feb 1964).
50. D. D. Ivanenko and D. F. Kurdgelaidze, Hypothesis concerning quark stars, *Astrophysics* **1**, 251 (1967).
51. J. M. Lattimer and M. Prakash, Neutron Star Structure and the Equation of State, *The Astrophysical Journal* **550**, 426 (mar 2001).
52. J. R. Oppenheimer and H. Snyder, On Continued Gravitational Contraction, *Physical Review* **56**, 455 (sep 1939).
53. S. Yamada and K. Sato, Numerical study of rotating core collapse in supernova explosions, *The Astrophysical Journal* **434**, p. 268 (oct 1994).
54. R. Kippenhahn, A. Weigert and A. Weiss, *Stellar Structure and Evolution* (Springer Berlin Heidelberg, 2012).
55. A. Heger, C. L. Fryer, S. E. Woosley, N. Langer and D. H. Hartmann, How Massive Single Stars End Their Life, **591**, 288 (jul 2003).
56. L. Landau, Origin of Stellar Energy, *Nature* **141**, 333 (feb 1938).
57. B. K. Harrison, M. Wakano and J. A. Wheeler, Matter-energy at high density: End point of thermonuclear evolution, *La structure et évolution de l'univers*, 124 (1958).
58. J. A. Wheeler, Geons, *Physical Review* **97**, 511 (jan 1955).
59. L. Visinelli, Boson Stars and Oscillatons: A Review (September 2021).
60. E. M. Lifshitz and I. M. Khalatnikov, Investigations in relativistic cosmology, *Advances in Physics* **12**, 185 (april 1963).
61. R. Penrose, Gravitational Collapse and Space-Time Singularities, *Physical Review Letters* **14**, 57 (jan 1965).
62. L. D. Landau, and E. Lifshitz, *The Classical Theory of Fields: Volume 2* (Butterworth-Heinemann, 1975).
63. S. A. Colgate and R. H. White, The Hydrodynamic Behavior of Supernovae Explosions, *The Astrophysical Journal* **143**, p. 626 (mar 1966).
64. S. Furlan, John Wheeler Between Cold Matter and Frozen Stars: The Road Towards Black Holes (2021).
65. J. A. Wheeler, The lesson of the black hole, *Proceedings of the American Philosophical Society* **125**, 25 (1981).
66. J. A. Wheeler, On the nature of quantum geometrodynamics, **2**, 604 (dec 1957).
67. M. Markov, O Predelnom Lambda-Protsesse, *Zhurnal eksperimentalnoi i teoreticheskoi fiziki* **17**, 848 (1947).
68. M. Markov, Supplement of the Progress of Theoretical Physics, Commemoration Issue for 30th Anniversary of Meson Theory by Dr, *H. Yukawa*, p. 85 (1965).
69. M. Markov, Limiting density of matter as a universal law of nature, *ZhETF Pisma Redaktsiiu* **36**, 214 (1982).

The Hamilton-Jacobi analysis by Peter Bergmann and Arthur Komar of classical general relativity

D. Salisbury
Physics Department, Austin College,
Sherman, Texas 75090, USA
E-mail: dsalisbury@austincollege.edu
www.austincollege.edu

Peter Bergmann initiated in 1966 an application of Hamilton-Jacobi techniques to general relativity. Little had been done by this time on extending this analysis to gauge theories. He proved that when, as in the case of Einstein's theory, the phase space generator of evolution consisted of a linear combination of constraints, the Hamilton principal function must be independent of spacetime coordinates. Also the Hamilton Jacobi equations that determined this functional of the 3-metric retained their form under phase space functionals that were invariant under the action of the spacetime diffeomorphism group. Komar followed up beginning in 1967 with a series of papers in which he proved that a complete solution of the Hamilton Jacobi equations was determined by a commuting set of diffeomorphism invariants. These invariants thereby labeled equivalence classes of solutions of Einstein's equations under the action of the full four-dimensional diffeomorphism group. Furthermore, this set satisfied canonical commutation relations with another invariant set. The hope and expectation was that these invariants could be promoted to quantum operators in a quantum theory of gravity. This framework will be contrasted with J. A. Wheeler's geometrodynamical program in which the only underlying covariance group is spatial diffeomorphisms. The full spacetime diffeomorphism symmetry is replaced by the notion of 'multi-fingered' time. A related dispute concerning the 'sandwich conjecture' will be discussed, relevant to the functional integral approach to quantum gravity. Two three geometries cannot determine a corresponding four geometry if they lie in distinct four dimensional diffeomorphism equivalence classes.

Keywords: Hamilton-Jacobi equations; geometrodynamics; quantum gravity.

1. Introduction

The following is a brief historical overview of work on a Hamilton-Jacobi approach to general relativity that was undertaken by Peter Bergmann and Arthur Komar in the 1960's and 1970's. A more detailed initial version, discussing both the relation of their research to previous and concurrent approaches, and to later progress, appears in Ref. 1. A further revision and expansion is in progress. The emphasis throughout their investigations was on the full four-dimensional diffeomorphism covariance of Einstein's theory. This lead to a divergence with the geometrodynamical approach of John Wheeler and company where only the spatial covariance is fully respected.

2. Bergmann's initial Hamilton-Jacobi analysis of general relativity

It is not widely recognized that it was Peter Bergmann who pointed out to Peres prior to the publication of his groundbreaking paper in Ref. 2 that his S appearing

as an argument in the four general relativistic constraint equations should be interpreted "as the Hamilton-Jacobi functional for the gravitational field." Of course the following are now identified as the Wheeler-DeWitt equations,

$$\mathcal{H}_\mu \left(g_{ab}, \frac{\delta S}{\delta g_{cd}} \right) = 0, \tag{1}$$

where the $\mathcal{H}_\mu \left(g_{ab}, p^{cd} \right)$ are the secondary constraints in general relativity. Bergmann proved in Ref. 3 that in a theory in which the Hamiltonian is constrained to vanish S could not depend explicitly on the time. The argument applied equally well to spatial dependence, as noted first in Ref. 4. Thus $S = S[g_{ab}(\vec{x})]$. Bergmann also showed that the Hamilton-Jacobi equations were form invariant under canonical transformations generated by diffeomorphism invariants. The fact that the numerical value of S is altered under the action of $\mathcal{H}_0$ presented a puzzle. The question arose whether this could this be inconsistent with the accepted notion of 'frozen time'.

3. Komar's isolation of solution equivalence classes

Komar observed in Ref. 4 that although there were only four Hamilton-Jacobi equations the principal function S delivered $6 \times \infty^3$ expressions for the momenta,

$$p^{ab}(\vec{x}) = \frac{\delta S}{\delta g_{ab}(\vec{x})}, \tag{2}$$

and therefore the $p^{ab}(\vec{x})$ are not uniquely determined. Two additional constraints needed to be imposed, with $A = 1, 2$,

$$\alpha^0_A \left[g_{ab}(\vec{x}), \frac{\delta S}{\delta g_{cd}(\vec{x})} \right] - \alpha_A(\vec{x}) = 0. \tag{3}$$

From the fact that

$$\frac{\delta \alpha^0_A}{\delta g_{ab}} + \frac{\delta \alpha^0_A}{\delta p^{cd}} \frac{\delta^2 S}{\delta g_{cd} \delta g_{ab}} = 0, \tag{4}$$

and similarly for the $\mathcal{H}_\mu$ it follows that

$$\left\{ H_\mu, \alpha^0_A \right\} = \frac{\delta^2 S}{\delta g_{ab} \delta g_{cd}} \left(- \frac{\delta H_\mu}{\delta p^{ab}} \frac{\delta \alpha^0_A}{\delta p^{cd}} + \frac{\delta H_\mu}{\delta p^{cd}} \frac{\delta \alpha^0_A}{\delta p^{ab}} \right) = 0. \tag{5}$$

In other words, the α^0_A must be diffeomorphism invariants (and they must also commute with each other.)

The constant values of $\alpha^0_A \left[g_{ab}(\vec{x}), p^{cd}(\vec{x}) \right]$ identify equivalence classes under the action of the spacetime diffeomorphism group. In Ref. 5 he showed that there existed invariant functionals β^A_0 that were canonically conjugate to the α^0_A. However, as formulated at this stage by Komar, one cannot yet obtain solutions of Einstein's equations by setting $\beta^A(\vec{x}) = \frac{\delta S}{\delta \alpha_A(\vec{x})}$. One still requires a temporal coordinate - like the 'intrinsic' q^0 that appears in the free particle action.

In Ref. 5 he showed that there existed invariant functionals β^A_0 that were canonically conjugate to the α^0_A. However, as formulated at this stage by Komar, one

cannot yet obtain solutions of Einstein's equations by setting $\beta^A(\vec{x}) = \frac{\delta S}{\delta \alpha_A(\vec{x})}$. One still requires a temporal coordinate - like the 'intrinsic' q^0 that appears in the free relativistic particle action, with increment given by

$$dS_p = p_\mu dq^\mu, \tag{6}$$

with constraint $p^2 + m^2 = 0$. This can be compared to vacuum general relativity where the non-vanishing contribution to the increment in the action takes the form

$$dS_{gr} = \int d^3x \, p_{ab} dg^{ab}, \tag{7}$$

with constraints $\mathcal{H}_\mu = 0$.

In the particle case one can choose the 'intrinsic time' $t = q^0$ as the evolution parameter and also solve for p_0 resulting in

$$dS_p = -\left(\vec{p}^2 + m^2\right)^{1/2} dt + p_a dq^a. \tag{8}$$

This yields the complete Hamilton principal function

$$S_p(q^a, t; \alpha^b) = -\left(\vec{\alpha}^2 + m^2\right)^{1/2} t + \alpha_a q^a. \tag{9}$$

The analogue of the gravitational α_0^A in this case is p^a. The analogue of the canonical conjugate β_B^0 would be the reparameterization constant $q^a - p^a q^0/p^0$. The general solution is obtained from

$$\beta^a = \frac{\partial S_p}{\partial \alpha^a}$$

Bergmann and Komar, Ref. 6, had explicitly recognized this type of emergence of intrinsic time evolution. Earlier, Komar in Ref. 7 had proposed that intrinsic curvature-based coordinates could be constructed using Weyl curvature scalars. He and Bergmann in Ref. 8 proved that these scalars depended only on g_{ab} and p^{cd}. The question naturally arise as to why Bergmann and Komar did not proceed with the use of intrinsic coordinates in their Hamilton-Jacobi treatment. A Bergmann quote in Ref. 9 from 1971 is revealing: "Although intrinsic coordinates lead, in principle, to a complete set of observables in general relativity, their defects, of which the most glaring is their deviation from Lorentz coordinates, render this procedure illusory. It appears preferable to retain coordinates that are approximately, or asymptotically Lorentzian and hence not to destroy one's intuition." Thus in spite of prolonged occupation with a Hamiltonian formulation of the the underlying general covariance of general relativity, Bergmann still seemed to have ceded undue importance to the more familiar Poincaré symmetry of conventional field theory.

However, as shown in Ref. 10 it is in principle possible to carry out a canonical change of variable in the non-vanishing increment dS_{GR} to intrinsic spacetime coordinates $x^\mu = X^\mu(g_{ab}, p^{cd})$, analogous to the parameter choice $t = q^0$ in the free relativistic particle model. This is a corrected version of Refs. 11 and 12. Current

work with Kurt Sundermeyer and Jürgen Renn is in preparation. One makes a canonical change of variables such that

$$\begin{aligned} dS_{gr} &= \int d^3x\, p^{ab} dg_{ab} \\ &= \int d^3x \left(\pi_\mu dX^\mu + p^A dg_A + \frac{\delta G}{\delta g_{ab}} dg_{ab} + \frac{\delta G}{\delta g_A} dg_A + \frac{\delta G}{\delta X^\mu} dX^\mu \right). \end{aligned} \tag{10}$$

One must find a generator $G[g_{ab}, X_A, g_B]$ such that $p_{ab} = \frac{\delta G}{\delta g_{ab}}$. Then

$$dS'_{gr} := d(S_{gr} - G) = \int d^3x \left(\pi_\mu dX^\mu + p^A dg_A \right). \tag{11}$$

Next choose the X^μ as intrinsic coordinates, i.e., set $x^\mu = X^\mu$. Finally, one eliminates the canonical conjugates to X^μ, π_ν, by solving the constraints. Then we have the resulting intrinsic Hamilton-Jacobi equation

$$\pi_0 \left[g_A, \frac{\delta S'_{gr}}{\delta g_B}, x^\mu \right] + \frac{\partial S'_{gr}}{\partial x^0} = 0. \tag{12}$$

From the complete solutions $S'_{gr}\,[g_A, x^\mu; \alpha_B]$ one can obtain the full set of physically distinct solutions of Einstein's equations from

$$\beta^A = \frac{\delta S'_{gr}}{\delta \alpha_A}. \tag{13}$$

4. Contrast with geometrodynamics

The contrast of this program with Wheeler's geometrodynamics cannot be overstated. The multifingered time approach assumed that the full four-dimensional diffeomorphism symmetry had been lost. States should be labeled by the $2 \times \infty^3$ diffeomorphism invariants $\alpha_A(\vec{x})$, and not by three-geometries.

References

1. D. Salisbury, Observables and Hamilton-Jacobi approaches to general relativity. I. Early history, *arXiv:2106.10774[gr-qc]* (2021).
2. A. Peres, On Cauchy's problem in general relativity - II., *Il Nuovo Cimento* **26**, 53 (1962).
3. P. G. Bergmann, Hamilton-Jacobi and Schrödinger theory in theories with first-class Hamiltonian constraints, *Physical Review* **144**, 1078 (1966).
4. A. B. Komar, Hamilton-Jacobi quantization of general relativity, *Physical Review* **153**, 1385 (1967).
5. A. B. Komar, Hamilton-Jacobi version of general relativity, *Physical Review* **170**, 1195 (1968).
6. P. G. Bergmann and A. B. Komar, Status report on the quantization of the gravitational field, in *Recent Developments in General Relativity*, ed. L. Infeld (Pergamon Press, 1962) pp. 31–46.
7. A. B. Komar, Construction of a complete set of independent observables in the general theory of relativity, *Physical Review* **111**, 1182 (1958).

8. P. G. Bergmann and A. B. Komar, Poisson brackets between locally defined observables in general relativity, *Physical Review Letters* **4**, 432 (1960).
9. P. G. Bergmann, Hamilton-Jacobi theory with mixed constraints, *Annals of the New York Academy of Sciences* **172**, 571 (1971).
10. D. Salisbury, J. Renn and K. Sundermeyer, Restoration of four-dimensional diffeomorphism covariance in canonical general relativity: An intrinsic Hamilton-Jacobi approach.
11. D. Salisbury, J. Renn and K. Sundermeyer, Restoration of four-dimensional diffeomorphism covariance in canonical general relativity: An intrinsic Hamilton-Jacobi approach, *arXiv:1508.01277v6* (2021).
12. D. Salisbury, J. Renn and K. Sundermeyer, Erratum. Restoration of four-dimensional diffeomorphism covariance in canonical general relativity: An intrinsic Hamilton Jacobi approach, *International Journal of Modern Physics A* **33**, p. 1892001 (2018).

The passage of time and top-down causation

Barbara Drossel

Institute of Condensed Matter Physics, Technical University of Darmstadt, Hochschulstr. 6, 64289 Darmstadt, Germany
E-mail: drossel@fkp.tu-darmstadt.de

It is often claimed that the fundamental laws of physics are deterministic and time-symmetric and that therefore our experience of the passage of time is an illusion. This talk will critically discuss these claims and show that they are based on the misconception that the laws of physics are an exact and complete description of nature. I will argue that all supposedly fundamental deterministic and time-symmetric laws have their limitations and are supplemented by stochastic and irreversible elements. In fact, a deterministic description of a system is valid only as long as interactions with the rest of the world can be ignored. The most famous example is the quantum measurement process that occurs when a quantum system interacts with a macroscopic environment such as a measurement apparatus. This environment determines in a top-down way the possible outcomes of the measure- ment and their probabilities. I will argue that more generally the possible events that can occur in a system and their probabilities are the result of top-down influences from the wider context. In this way the microscopic level of a system is causally open to influences from the macroscopic environment. In conclusion, indeterminism and irreversibility are the result of a system being embedded in a wider context.

Keywords: Indeterminism, irreversibility, contextual emergence, top-down causation, nonreductionism

1. Introduction

The success of physics at explaining, calculating and controling processes in nature has led to the widespread belief that the equations of physics describe accurately how the state of a system changes with time under the influence of the various physical forces. Even more, many physicists think that the properties of the fundamental equations of physics are also properties of nature. Since these equations are deterministic, it is concluded that nature also is deterministic, with the time development being completely determined by the initial state and the fundamental laws. Furthermore, these laws are invariant under changing the direction of time. Based on this, many physicists think that the perceived direction of time is merely a consequence of very special initial conditions of the universe, and that all processes in nature could in principle also run backwards in time. Einstein is the most prominent scientist who considered the irreversible passage of time as an illusion since the present contains already the future, as it is fixed by the deterministic laws of physics and the initial conditions.

The implications of this understanding of physics are immense. It follows in particular that our experience of being agents that can act in the "now" and thus bring about future developments is also an illusion. And if everything is fully determined by the laws of physics, this contradicts our experience that our decisions are driven by non-physical causes such as values and logic and purpose.

While many philosophers of science buy the bold claims made by physicists, others are more skeptical. Indeed, there are many reasons to be skeptical. Physics as a science is always an approximate and incomplete description of nature and not a direct image of what goes on in nature. And many physicists not working with supposedly 'fundamental' physics such as quantum physics, but in condensed matter physics or soft matter physics, are aware of these limitations of the basic physical theories at describing nature. Instead of deriving the properties of their systems starting from the many-particle Schrödinger equation (which would be the fundamental equation for an object that consists of a huge number of atoms), they write down simpler models and effective theories that capture more directly the phenomenon to be described. These theories are often even in logical contradiction with the supposed fundamental theory, as explained in the following quote by Nobel laureate Anthony Leggett[1]:

> No significant advance in the theory of matter in bulk has ever come about through derivation from microscopic principles. [...] I would confidently argue further that it is in principle and forever impossible to carry out such a derivation. [...] The so-called derivations of the results of solid state physics from microscopic principles alone are almost all bogus, if 'derivation' is meant to have anything like its usual sense. Consider as elementary a principle as Ohm's law. As far as I know, no-one has ever come even remotely within reach of deriving Ohm's law from microscopic principles without a whole host of auxiliary assumptions ('physical approximations'), which one almost certainly would not have thought of making unless one knew in advance the result one wanted to get, (and some of which may be regarded as essentially begging the question). This situation is fairly typical: once you have reason to believe that a certain kind of model or theory will actually work at the macroscopic or intermediate level, then it is sometimes possible to show that you can 'derive' it from microscopic theory, in the sense that you may be able to find the auxiliary assumptions or approximations you have to make to lead to the result you want. But you can practically never justify these auxiliary assumptions, and the whole process is highly dangerous anyway: very often you find that what you thought you had 'proved' comes unstuck experimentally (for instance, you 'prove' Ohm's law quite generally only to discover that superconductors don't obey it) and when you go back to your proof you discover as often as not that you had implicitly slipped in an assumption that begs the whole question. [...] I claim then that the important advances in macroscopic physics come

> essentially in the construction of models at an intermediate or macroscopic level, and that these are logically (and psychologically) independent of microscopic physics. Examples of the kind of models I have in mind which may be familiar to some readers include the Debye model of a crystalline solid, the idea of a quasiparticle, the Ising or Heisenberg picture of a magnetic material, the two-fluid model of liquid helium, London's approach to superconductivity .... In some cases these models may be plausibly represented as 'based on' microscopic physics, in the sense that they can be described as making assumptions about microscopic entities (e.g. 'the atoms are arranged in a regular lattice'), but in other cases (such as the two-fluid model) they are independent even in this sense. What all have in common is that they can serve as some kind of concrete picture, or metaphor, which can guide our thinking about the subject in question. And they guide it in their own right, and not because they are a sort of crude shorthand for some underlying mathematics derived from 'basic principles.'

The goal of this paper is to show that the claim that nature is at its most fundamental level reversible and deterministic is wrong. First, I want to explain that the time-reversible and deterministic nature of the supposedly fundamental physical equations are in fact idealizations that hold only under very specific conditions. In particular, the studied objects must be carefully isolated from interacting with the rest of the world. By taking a closer look at the areas of physics that are supposedly based on these fundamental equations, I will point out instances where indeterministic and irreversible additional elements are added to the theory when needed. This shows that the programme of deriving everything *only* from the fundamental equations is never realized in practice. I will furthermore argue that indeterminism and irreversibility are closely related, as the first implies the second.

The central part of my paper will focus on the influence of the environment and on top-down effects. By looking at what happens in a quantum measurement, we will see that the environment determines the possible events and their probabilities. This will lead us to a more general discussion of how stochasticity and irreversibility are the result of top-down effects from the context. A consequence of this is a contextually emergent view of nature, where not everything is controlled bottom-up by microphysics but there are also top-down influences from the surrounding context - even from without physics. This prepares the conclusion that while physics underlies everything that happens in nature, it does not determine everything.

2. Irreversibility and indeterminism in the 'fundamental' theories of physics

The following considerations show that no theory that was once or is now thought to be fundamental can do without adding irreversible elements.

2.1. *Classical mechanics*

Classical mechanics is based on Newton's law

$$m\frac{d^2\vec{x}}{dt^2} = \vec{F}(\vec{x})\,. \tag{1}$$

This law is deterministic as the position $\vec{x}$ and velocity $\vec{v} = d\vec{x}/dt$ at a given moment in time determine the future time evolution of $\vec{x}$ and $\vec{v}$. The law is invariant under time reversal as a change of the sign of t does not change the law. Consequently, when a time evolution of $\vec{v}$ and $\vec{x}$ has occurred from time $t = 0$ to $t = t_f$, the reverse time evolution takes place when starting at time t_f and inverting the velocity $\vec{v}$. When applying Newton's law to many particles that interact via mutual forces, the same conclusions are obtained. Due to its impressive success at unifying Kepler's laws of planetary motion and Galileis laws for falling bodies, Newton's theory was considered for a long time an exact and comprehensive description of the physical world, but there have always been cautious voices pointing out its limitations, see.[2,3] Only with the advent of the theory of relativity and quantum mechanics did it become clear to everybody that this view is wrong. Sometimes I wonder why physicists still make this mistake to believe that their most recent theories are exact and universal....

Back to classical mechanics: Even before the mentioned developments in the 20th century, it was clear that classical mechanics cannot do without irreversible elements: When comparing the equations to reality, one needs to include friction. For instance, a pendulum that swings freely under the influence of gravity will eventually come to rest as friction reduces the swinging amplitude with time. When including friction, Newton's law becomes

$$m\frac{d^2\vec{x}}{dt^2} = \vec{F}(\vec{x}) - \eta\frac{d\vec{x}}{dt} \tag{2}$$

with a friction term proportional to the velocity, the strength of which is determined by the friction coefficient η. Even the apparently so perpetual motion of the planets on their orbits is slowed down by friction due to tidal forces.

Classical mechanics also knows stochasticity. As explained so well by Gisin,[4,5] determinism is valid only if position and velocity have infinite precision. But physics is always limited to a finite number of bits when measuring and calculating the motion of objects. For chaotic systems, the time evolution of which depends extremely sensitively on the initial values, this means in practice that prediction of the future time evolution becomes impossible beyond a certain time horizon. There are good reasons to conclude that this indeterminism is not just a matter of our limited abilities, but an inherent feature of nature, see also.[5–7]

If we accept that chaotic systems are inherently stochastic, the irreversible nature of friction is coupled to stochasticity on the atomic level of description: Viewed microscopically, the slowing down of the pendulum is due to collisions of the pendulum with the randomly moving molecules in the air. However, the motion of the atoms is not deterministic since it is chaotic when described by Newton's laws.

2.2. *Classical electrodynamics*

Classical electrodynamics is governed by Maxwell's four equations,

$$\nabla \cdot \vec{E} = 4\pi\varrho \tag{3}$$

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \tag{4}$$

$$\nabla \cdot \vec{B} = 0 \tag{5}$$

$$\nabla \times \vec{B} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t} \tag{6}$$

These equations are again deterministic as an initial configuration of the electric and magnetic fields $\vec{E}$ and $\vec{B}$ and of the charge and current density distributions ϱ and $\vec{j}$ determine the future evolution of the fields, and using Newton's equation for the motion of the charges under the influence of the fields they also determine the time evolution of the charge and current densities.

The equations are also time reversible: When reversing the direction of time and that of currents and magnetic fields, Maxwell's equations remain unchanged. So for each time evolution there is an equally realistic reversed time evolution (as nature knows of no preferred orientation for currents or magnetic fields).

However, there is an important subfield of electrodynamics where the time reversal invariance is broken: This is the emission of radiation from accelerated charges. Using the electrostatic potential ϕ and the magnetic vector potential $\vec{A}$ instead of $\vec{E}$ and $\vec{B}$, the emitted radiation takes the form

$$\phi(\vec{r},t) = \int d^3r' \frac{\varrho\left(\vec{r}\,', t - \frac{|\vec{r}-\vec{r}\,'|}{c}\right)}{|\vec{r}-\vec{r}\,'|}\,, \tag{7}$$

$$\vec{A}(\vec{r},t) = \frac{1}{c}\int d^3r' \frac{\vec{j}\left(\vec{r}\,', t - \frac{|\vec{r}-\vec{r}\,'|}{c}\right)}{|\vec{r}-\vec{r}\,'|}\,. \tag{8}$$

This means that the potentials (and consequently the fields) at a position $\vec{x}$ at time t are determined by the motion of the charges at earlier times $t - \frac{|\vec{r}-\vec{r}\,'|}{c}$, which is simply the time electromagnetic radiation takes to propagate with the velocity of light from the point of origination $\vec{r}\,'$ to the point of measurement $\vec{r}$. All electromagnetic fields thus have localized sources in the temporal past.[8]

When explaining the reasons why emission of radiation is described by the retarded solutions, one has to resort to microscopic stochasticity: Radiation emitted by localized sources is absorbed by walls etc. The reverse process would be that different walls conspire to emit radiation that converges from all directions to a localised sink where it is completely absorbed and turned into motion of electrical charges. However, this 'conspiracy' is argued to be impossible as the atoms of the walls perform random thermal motion and therefore emit incoherent thermal radiation and not radiation that is correlated over large spatial scales. The thermal motion of the wall atoms is considered to be stochastic.

There is a second way how stochasticity enters electrodynamics: When trying to describe what happens at the atomic scale when light is emitted, one needs to resort to a quantum version of electrodynamics, which is quantum electrodynamics and describes the emission of light in terms of the stochastic emission of energy quanta (each containing an energy $\hbar\omega$ with ω being the frequency) from the source. And this brings us to the next section, which deals with quantum mechanics.

2.3. *Quantum mechanics*

The basic equation of nonrelativistic quantum mechanics for one particle is the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{x},t) = \hat{H}\Psi(\vec{x}) \tag{9}$$

with the Hamilton operator

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\vec{x})\,. \tag{10}$$

Here, the first term is the kinetic energy, and the second the external potential. Given the initial state $\psi(\vec{x},0)$, the state at other times t are given by the relation

$$\Psi(\vec{x},t) = e^{-i\hat{H}t/\hbar}\,\Psi(\vec{x},0) \tag{11}$$

if the potential $V(\vec{x})$ is not explicitly time dependent.

This equation is deterministic. This means that the initial state, combined with the Hamiltonian determines the future time evolution. It is also invariant under time reversal: The time-reversed Schrödinger equation is solved by the complex conjugate wave function Ψ^*, and this does not affect the observables as they are calculated from expressions that contain products of Ψ and Ψ^*. All this holds also for the many-perticle version of the Schrödinger equation.

However, quantum mechanics is incomplete without a rule for how to calculate the outcome of a measurement. And this rule says that out of all possible measurement outcomes of an observable, one of them (let us call the corresponding state ϕ_n) is chosen stochastically with a probability that is given by $|\langle\phi_n|\Psi\rangle|^2$. This process is irreversible, as the reverse process (that the measurement apparatus or photographic plate returns to the pre-measurement state and emits the particle that has been absorbed during measurement) is not observed in nature. In this way, quantum mechanics includes from the onset irreversibility and stochasticity. There are a number of interpretations of quantum mechanics that attempt to explain away this dichotomy between the Schrödinger equation and the measurement process by describing also all the atoms of the measurment device by quantum mechanics, usually invoking decoherence in one way or another[9]; however, to many people these interpretations remain unsatisfactory as they cannot really explain why in a single run of an experiment only one of the possible outcomes is observed.[10,11] In one way

or another, stochasticity kreeps into any interpretation. It does so even in Bohmian mechanics, where all stochasticity of the future time evolution is contained in the random features of the initial state.

2.4. *Quantum field theory*

The theory that is often considered the most fundamental one is quantum field theory, which is a relativistic theory for many particles and has several building blocks that take into account the different types of interactions and particles. It is a combination of the Glashow-Salam-Weinberg model for the electroweak interaction and of quantum chromodynamics for the strong interaction. Just as for nonrelativistic quantum mechanics, this theory includes two parts. Unitary time evolution according to a Hamilton operator is only applied between preparation and measurement of particles. The measurement outcome (e.g. of a scattering experiment or particle collision experiment) is again described by a probability that is calculated in a similar way as above. This means that everything written above for nonrelativistic quantum mechanics applies also to quantum field theory.

2.5. *Thermodynamics*

Thermodynamics is considered fundamental only by some scientists. But these scientists argue that thermodynamics might be even more fundamental than the other fields of physics. The most famous quotation in this direction is by Sir Arthur Eddington (From his book *The Nature of the Physical World*[12]):

> The law that entropy always increases holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations - then so much the worse for Maxwell's equations. If it is found to be contradicted by observation - well, these experimentalists do bungle things sometimes. But if your theory is found to be against the Second Law of Thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation.

Einstein had a similary high opinion of thermodynamics:

> A law is more impressive the greater the simplicity of its premises, the more different are the kinds of things it relates, and the more extended its range of applicability. [...] It is the only physical theory of universal content, which I am convinced, that within the framework of applicability of its basic concepts will never be overthrown.

Now, the second law of thermodynamics is about the irreversibility of nature: All processes run in the direction in which the entropy of the universe increases. The relations of thermodynamics can be obtained also from a microscopic description,

which is that of statistical physics. But statistical physics is not deterministic, as it is based on probabilities for the different possible microscopic states a system can take. This means that we have again a link between irreversibility and stochasticity in this field of physics.

3. A reversible, deterministic theory cannot give irreversibility

It is often argued that a 'fundamental' microscropic theory that is deterministic and reversible, such as classical mechanics or quantum mechanics, can give rise to irreversibility. Even textbooks on statistical mechanics often make this claim and use arguments based on coarse-graining. In this section, I will demonstrate that the 'derivations' of irreversibility from a microscopic reversible and deterministic theory all employ additional assumptions that are not part of the theory. They all invoque stochasticity or randomness in one way or another and are therefore not truly deterministic. As soon as stochasticity is employed, irreversibility follows naturally. The first subsection will argue that stochasticity leads to irreversibilty, and the second subsection will show how stochasticity is smuggled into the derivations of the second laws of thermodynamics from classical mechanics.

3.1. *Stochasticity gives irreversibility*

In the previous section, we have seen that irreversible processes are always coupled to stochasticity when considered on the atomic level, for instance when friction is described in terms of atomic collisions or when the approach to thermodynamic equilibrium is described by stochastic transitions between different microscopic states. More general considerations suggest that irreversibility and stochasticity are indeed intimately connected: Progress of time is perceived via changes. These changes occur in the form of events. Whenever one of a set of possible events occurs, the open future becomes the definite past. Thus, events are both irreversible and stochastic. There are several theories that are based on such events constituting the fabric of spacetime itself on the Planck scale.[13,14] But we need not know the most fundamental level in order to see that any stochastic event is also irreversible. This is because once the event has occurred the state prior to this event cannot be retrodicted. For a quantum measurement, e.g., the outcome does not allow to reconstruct the incoming wave function. A stochastic event itself is a step by which an indeterminate future gives way to a definite outcome. Against this, it is sometimes argued that probabilities can also be applied backwards for retrodiction. Yes, but this is qualitatively different: probabilities used in retrodiction are either Bayesian probabilities that are based on incomplete knowledge of the past, or they are frequencies in an ensemble of events. In principle, the past that precedes a given event is fixed in either case. In contrast, the future is not fixed if stochasticity is real and not merely apparent.

3.2. *The hidden assumption in so-called 'derivations' of the Second Law from classical mechanics*

In the following, I will use classical mechanics to demonstrate that additional assumptions are smuggled in when irreversibility is 'derived' from a reversible deterministic theory. A similar type of argument can be made for the supposed derivations of the Second law from a many-particle Schrodinger equation.[6] In both cases, the additional arguments are randomness of initial conditions and of statistical independence of degrees of freedom. The main ideas behind the 'derivations' of the Second Law from classical mechanics are as follows:

(i) A gas is modelled as a conservative mechanical system of $N \simeq 10^{23}$ small balls with hard-core repulsion, enclosed in a container of volume V with perfectly reflecting walls. Since energy does not change during time evolution, the trajectory of this mechanical system in $6N$-dimensional phase space stays within the $6N-1$-dimensional energy shell in which the initial state is placed. (ii) It is generally assumed, even though this is not proven, that the dynamics of this system is ergodic. This means that a "typical" trajectory approches each point of the energy shell with arbitrary precision ϵ if enough time has passed. Or, equivalently, an initially small and compact phase-space volume of dimension $6N$ (representing an ensemble of very similar initial conditions) will become stretched and folded to the extent that eventually some part of it will be in every small volume of size ϵ^{6N-1} in the energy shell. (iii) The vast majority of cells of size ϵ^{6N-1} correspond to macrostates with maximum entropy, which have (among other properties) an even distribution of density over the entire volume. (iv) Therefore, if starting in one of the few cells that correspond to a low-entropy initial state, after sufficiently long time the state of the system will 'almost certainly' be in a cell of maximum entropy.

Of course, each of these steps could be discussed in depth, in particular the starting assumption of infinite-precision phase space points and infinite-precision trajectories. I will focus on the last step and accept all preceding ones for the sake of brevity. In this last step, additional assumptions creep in: The last step is based on the assumption that the trajectory taken by the system is a 'typical' trajectory, or, equivalently, that among all the possible initial states that lie within the initial energy-shell volume element, the true initial state is a randomly chosen one. This is an assumption of randomness of the initial state. The laws of classical mechanics would not be violated if all particles of the gas would gather in irregular time intervals in some corner of the container. Among all possible initial states that agree with each other within a desired degree of precision, there would be particular initial states that show such an atypical behavior. We therefore have to introduce the additional assumption that the 'true' initial state is not such an 'atypical' one that leads to unexpected low-entropy future states. This is somewhoat similar to what one does in electrodynamics when ruling out spatially localized sources of radiation that lie in the future.

Instead of postulating random initial conditions, one could equivalently argue that the initial state is not fixed with infinite precision but only with finite precision, and that randomness comes in as the time evolution of the system proceeds and more and more bits that were not fixed initially become relevant for fixing the ongoing time evolution. In my view there are many reasons to prefer this latter perspective, and it has been defended most forcefully by Nicolas Gisin.[5]

But there is more to stochasticity than that it is always added to 'deterministic' theories when real-world problems are addressed. Stochasticity depends on context. The next section focuses on this issue.

4. Stochasticity is context dependent

We have seen above that whenever the environment of a system is included irreversibility comes in: In mechanical systems, the environment causes friction. In electrodynamics, considering the emission of radiation means that there is an environment into which this radation can be emitted. In quantum mechanics, the interaction of a quantum particle with the rest of the world, which is here represented by a measurment device, causes the irreversible measurement event. Thermodynamics is a very interesting field of science from this point of view as it is a prime example of the environment determining what happens in a system: The environment imposes the temperature and the chemical potential. The environment performs changes on the system (such as pushing a piston or connecting a hot and a cold object) the effect of which is calculated in thermodynamics problems.

It is usually not acknowledged that stochasticity also requires a context. Karl Popper emphasized it in his paper "The propensity interpretation of probability",[15] where he argues that probabilities are defined only when the conditions under which the experiment is to be repeated are specified. Probabilities are thus defined relative to a setup in which the event of interest occurs. Popper justifies the propensity view by the findings of quantum mechanics. Indeed, the "fundamental" stochastic event is a quantum event. Its standard example is a measurement. Let us take for illustrative purposes the Stern-Gerlach experiment. It is the simplest quantum measurement as the observable to be measured is a spin-1/2, with only two measurement results, which we denote as $|+\rangle$ and $|-\rangle$. The measurement device measures the spin with respect to direction chosen by the experimenter. Let this direction be the positive z direction. If the incoming particles have not been polarized such that they point in the z direction, both measurement outcomes $|+\rangle$ and $|-\rangle$ are possible, and their probabilities depend on the preparation procedure. Now, the experimenter could have chosen to measure the spin with respect to another direction, for instance the x direction. In this case the possible measurement results would be again $|+\rangle$ and $|-\rangle$, but now these two results signify an orientation parallel or antiparallel to the direction of the x axis and no longer to the z axis. In general, the probabilities for the two measurement results are also different compared to a measurement in z direction.

This example demonstrates that the possible events and their probabilities are not an intrinsic property of the quantum particle but the combined effect of the way how the quantum particle has been prepared and of the measurement device. Had the experimenter chosen to measure position or energy instead of spin, the possible results and their probabilities would again have been different. Had the experimenter decided not to perform any measurement, the quantum particle would have remained in the state generated by the preparation procedure.

Now, quantum measurements are just one class of quantum mechanical events by which a quantum particle undergoes an irreversible, stochastic transition and exchanges energy with a macroscopic, finite-temperature object (which is a classical object). There need not be an experimenter, the particle could interact with the rest of the world in a different way. Examples are nuclear fusion reactions in the core of the sun or the emission of photons into space from the warm surface of the earth. Again, the fact that these events happen and their probabilities depend on the context.

So far, not many models for quantum measurement take this context explicity into account. Many authors, however, speak in a more general sense of quantum contextuality.[16,17] The Copenhagen interpretation of the measuremeent process makes an explicit distinction between classical objects (the measurment setup) and quantum objects, but does not detail how this difference arises. More explicit considerations of how the context is involved in a quantum measurement are advanced by Ellis[18] and more recently by Drossel and Ellis.[11]

This insight that there is a top-down effect from the environment or the larger spatial scale to the quantum particle leads us to the next section where we introduce the concepts of contextual emergence and top-down causation.

5. Contextual emergence and top-down causation

Physics is viewed by many people as a reductionistic science where the properties of an object are derived from its parts and their interactions. Consequently, those theories that deal with the most microscopic objects (i.e., particle physics and quantum field theory) are considered to be the most fundamental ones and to describe, at least in principle, everything that occurs on larger scales. But this is a one-sided view that ignores the various ways in which the context or the larger structure influences what happens at a smaller scale. To understand this better, it is useful to think of nature as a hierarchy of objects, with the objects on the lower level being the parts of the objects on the next hierarchical level. For instance, one can build the hierarchy Elementary particle - Atom - Crystal - Earth - Solar system - Galaxy - Universe; or, choosing a hierarchy that involves us humans: Elementary particle - Atom - Molecule - Cell - Organ - Individual - Society.

Now, describing an object in terms of its parts and their interaction is practised successfully by all scientists. For instance, describing the conducting properties of a metal, physicists build a model of electrons performing collisions with the lattice

defects and the lattice vibrations (the so-called phonons) of the metal. Or, when exploring the cycle of growth and division of a biological cell, scientists describe this process by a network of molecular interactions and reactions. This success of the reductionist procedure often veils the view for the top-down influences that are equally important: The metal block provides the environment for lattice vibrations to exist, and the temperature of the metal block determines how many phonons of which frequency are present. The lattice structure of the metal determines what the conduction bands of the metal look like and thus strongly affects the way how electrons can propagate in the material. The wider context, such as the connection to a power supply is the prerequisite for conduction happening at all. Similarly, the growth and division of cells is regulated by mechanical and chemical cues from the embedding tissue to which the cell belongs. Furthermore, the cell depends on a continuous supply of nutrients and building material from its environment. For any other system of the hierarchies mentioned above, we can similarly list a variety of top-down influences of the whole on its parts. This topic of top-down causation is discussed extensively by George Ellis in his book "How can physics underlie the mind".[19]

All this means that reduction, successful as it is, is only part of the story. The complement of reduction is emergence: while reduction considers how the whole can be explained in terms of its parts, the concept of emergence considers the qualitatively new properties of the whole, which are not properties of the parts. One distinguishes between different concepts of emergence. Proponents of weak emergence hold that the emergent properties, even though they appear surprising and qualitatively different from the properties of the parts, can ultimately be fully accounted for by the properties of the parts. For a physicist this means that the theory describing the parts and their interactions contains at least implicitly all the phenomena observed in the whole. It is closely associated with a reductionist view. In contrast, proponents of strong emergence hold that the emergent properties are explained only partially by the parts and their interactions and cannot be fully reduced to them. The notion that captures best the type of strong emergence that occurs in physics is that of contextual emergence. It emphasizes that the existence of the parts and their properties depend on the context, the boundary conditions, etc. Until know, strong (or contextual) emergentists are a minority among physicists. The reductionist paradigm prevails despite the many good reasons to question it.

With this understanding of contextual emergence and top-down causation, we will gain a deeper understanding of how indeterminism enables our experience of the passage of time and of being agents.

6. Indeterminism and top-down causation

It is often argued that indeterminism is of no help at justifying free will and agency. The reason given is that stochasticity is completely random in the sense that random events are not influenced by anything else and can therefore not be part of

a conscious act or decision - which is not random but follows logical reasoning or is based on values. However, our discussion of the quantum measurement process above has shown that stochastic events are not completely random. On the contrary, the possibility of these events and the nature of these events is set by the context. This is not appreciated very often. Karl Popper made the interesting suggestion that chance at the lower level is necessary for top-down causation from the higher level,[20] and he has been criticized for it.[21] But I think that he is right, and so do others.[7] Only when the entities at the lower level are not fully controlled by the microscopic laws can they respond to the higher level.

In biological systems, the influence of the environment on the types of stochastic events and their effects is evident at all levels: Let us take gene expression. The process by which transcription factors bind to the promoter region of a gene and enable transcription has many random elements, as diffusion and binding of transcription factors is subject to thermal noise. However, the context determines which transcription factors are activated at all. For instance, when I experience joy, this triggers the discharge of endorphins, which in turn stimulates the production of further endorphines via gene expression. On a larger timescale, evolution has shaped the transcription factors that we possess and the affinities to the different genes, resulting in the binding and dissociation rates that characterize the stochastic dynamics in our cells.

As another example, consider the brain. The neurons in our brain fire "randomly". At the same time, this electric activity reflects the brain activity, for instance while I formulate the text of this paper and type it. There is a wider context that influences the stochastic events in the brain: The past experiences have shaped the connection patterns and the strengths of the synapses of our neural network, and this in turn affects how activity propagates through the network and which neurons or groups of neurons are triggered by which others. Although there is randomness in the individual firing events of neurons, and on a more microscopic level in the opening and closing of ion channels, the frequency of firing and the neurons involved and the temporal sequence of which regions become active and which output neurons pass the signal on (for instance to my fingers while I am typing) is not random at all. Somehow it even reflects my creative activity while I develop my arguments for this text.

The constructive role of randomness in biological systems is discussed in more depth in several review articles.[22–24]

7. Conclusions

The view of physics that I have defended in this article is very different from the reductionist view that is propagated in many textbooks. The main problem with that view is that theories are applied far beyond their range of applicability. Any physical theory is an approximation and not an exact description of the physical processes of interest. And by no means it is a complete description of everything that

is going on in the physical world. I have argued that even though the supposedly fundamental theories are deterministic and time-reversible, these theories are in practice supplemented by irreversible and stochastic features. There is no reason to assume that irreversibility and stochasticity are only apparent. Such a claim is based on ideology and not on evidence. I have argued that irreversibility and stochasticity enter when the influence of the environment on a system is taken into account. This focused the attention on the often-neglected topics of emergence and top-down causation. It is the context that sets the stage for the possible events that can happen in the system, and the context also influences the probabilities of these events.

It follows that our experience of time flowing and us being agents that can influence what happens is not contradicted by physics. Physics enables everything and underlies everything, but it does not determine everything. On the contrary, physics appears to be sufficiently complex to allow even life and consciousness to function on a material platform.[19] This bears some resemblence to Turing machines in computational science, which are complex enough to allow for any possible type of calculation.

Top-down influences come in different shapes. Top-down causation from the wider material context is easier to understand but is only part of the story. In particular the two biological examples of Section 6 illustrate that there is also an influence from the nonmaterial world of ideas, rationality, goals and desires on the activity of our genes and neurons. It may well be that physics will never find an access to describing these influences. The goal of my paper is far more modest, namely to make the case that physics cannot rule out that these influences are present. Physics is not causally closed and does not encompass everything that happens in our world.

To conclude, we should not let naive interpretations of the laws of physics constrain our view of reality. On the contrary, we should trust our most immediate experiences that are the basis for us being capable of thinking and acting, and based on these we should critically examine those simplistic metaphysical claims.

References

1. A. J. Leggett, On the nature of research in condensed-state physics, *Foundations of Physics* **22**, 221 (1992).
2. M. van Strien, Was physics ever deterministic? the historical basis of determinism and the image of classical physics, *The European Physical Journal H* **46**, 1 (2021).
3. F. Del Santo, Indeterminism, causality and information: Has physics ever been deterministic?, in *Undecidability, Uncomputability, and Unpredictability*, (Springer, 2021) pp. 63–79.
4. N. Gisin, Indeterminism in physics, classical chaos and bohmian mechanics: Are real numbers really real?, *Erkenntnis*, 1 (2019).
5. N. Gisin, Real numbers are the hidden variables of classical mechanics, *Quantum Studies: Mathematics and Foundations* **7**, 197 (2020).

6. B. Drossel, On the relation between the second law of thermodynamics and classical and quantum mechanics, in *Why More Is Different*, (Springer, 2015) pp. 41–54.
7. F. Del Santo and N. Gisin, Physics without determinism: Alternative interpretations of classical physics, *Physical Review A* **100**, p. 062107 (2019).
8. S. Weinstein, Electromagnetism and time-asymmetry, *Modern Physics Letters A* **26**, 815 (2011).
9. W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, *Reviews of modern physics* **75**, p. 715 (2003).
10. M. Schlosshauer, Decoherence, the measurement problem, and interpretations of quantum mechanics, *Reviews of Modern Physics* **76**, p. 1267 (2005).
11. B. Drossel and G. Ellis, Contextual wavefunction collapse: An integrated theory of quantum measurement, *New Journal of Physics* **20**, p. 113025 (2018).
12. A. Eddington, *The nature of the physical world: The Gifford Lectures 1927* (BoD–Books on Demand, 2019).
13. M. Cortês and L. Smolin, Quantum energetic causal sets, *Physical Review D* **90**, p. 044035 (2014).
14. R. D. Sorkin, Spacetime and causal sets, *Relativity and gravitation: Classical and quantum*, 150 (1990).
15. K. R. Popper, The propensity interpretation of probability, *The British journal for the philosophy of science* **10**, 25 (1959).
16. P. Grangier and A. Auffèves, What is quantum in quantum randomness?, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **376**, p. 20170322 (2018).
17. G. Jaeger, Quantum contextuality and indeterminacy, *Entropy* **22**, p. 867 (2020).
18. G. F. Ellis, On the limits of quantum theory: Contextuality and the quantum–classical cut, *Annals of Physics* **327**, 1890 (2012).
19. G. Ellis, *How can physics underlie the mind?: Top-down causation in the human context* (Springer Heidelberg, 2016).
20. K. R. Popper and J. C. Eccles, *The Self and Its Brain: An Argument for Interactionism* (Springer Heidelberg Berlin, 1977).
21. T. O'Connor and H. Y. Wong, Emergent properties, in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Metaphysics Research Lab, Stanford University, 2015) Summer 2015 edn.
22. R. Noble and D. Noble, Harnessing stochasticity: How do organisms make choices?, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, p. 106309 (2018).
23. K. Wiesenfeld and F. Moss, Stochastic resonance and the benefits of noise: from ice ages to crayfish and squids, *Nature* **373**, 33 (1995).
24. L. S. Tsimring, Noise in biology, *Reports on Progress in Physics* **77**, p. 026601 (2014).

Explaining time's passage

Jonathan J. Dickau

Independent Researcher, ISGRG member
Poughkeepsie, New York 12603, USA
E-mail: jonathan@jonathandickau.com
www.jonathandickau.com

The way we experience time is in the accumulation of experiences and events that happen in the moment, and then are behind us. Since the time of Anaximander at least; philosophers have tried to explain both the nature of time and its origin or basis. In modern times; scientists are the ones exploring the domain of time, so now *they* attempt to explain the nature and basis of time – with varying degrees of success. This is complicated because explanations from Classical Physics or Relativity are different from, and incompatible with, answers from Quantum Mechanics, so we hope Quantum Gravity theories will help resolve this. Recent advances in Mathematics hold promise for a unified basis explaining both the thermodynamic and quantum-mechanical time arrows in a way that consistently informs our Philosophy. However; we may need to explore beyond the island of familiar Maths, to reconcile the divergent pictures of how and why time passes.

Keywords: Origin of time; passage of time; basis of time; Philosophy of time; Physics of time.

1. Introduction

While we imagine time is a single entity; the reality of time is more complex and more interesting. As far back as Anaximander[1]; philosophers have tried to make sense of that complexity, and his analogy of stepping into a moving stream actually comes very close to the reality we observe, since things keep moving while events are happening. So the simplest realities of life are governed by a rule that assures most decisions will be complex instead of simple. Coupling Plato's observation "time is the image of eternity[2]" with the insight that space is a projection of infinity can help us make sense of his deeper message. But then; we see glimpses into quantum gravity from the ancient Greek philosophers. Rovelli's book "Reality is Not What it Seems[3]" wonderfully recounts insights from those individuals that extend to topics including quantum gravity and the nature of time. While I agree with his view that we find the seeds of quantum gravity and the birth of atomism among ancient philosophers; we differ somewhat on whether time is fundamental or primordial on that basis. The resolution involves both cosmology and quantum gravity. So this paper explores how recent developments in Mathematics might have a deeper relevance than was previously appreciated to advance our understanding of time. I will discuss the examples cited above in greater detail and offer unique insights from my own research.[4,5]

Already in 1922, concluding a debate with Bergson; Einstein said "There is no such thing as the time of the philosopher.[6]" So for some it is a dead issue. But the debate goes on in the Physics community and among modern philosophers, over the consequences of how physical theory affects our understanding of time. This debate centers on the sharply differing views of time in Relativity and Quantum Mechanics. And there is strong evidence both views must be true. So if the traditional outlook is that time is a single entity for all; that view is broken. But this emphasizes the need to create new ways to conceptualize what we have learned in Physics and incorporate it into our Philosophy.[7] And yet; Anaximander's ancient view of time as a stream which continues to flow even as you are stepping in it remains enlightening in modern times. So there is hope for reconciling Physics with Philosophy, and knowledge that by developing quantum gravity theories; we gain a better philosophical understanding of time. I think the reverse is also true; it is easier to choose or develop better approaches to quantum gravity, by understanding the Philosophy of time better. Prevailing QG theories share many features and predictions with considerable common ground to explore.[8] So I examine ideas of Connes,[9] Longo,[10] Gisin,[11] as well as Drossel and Ellis,[12] on the emergence of time from quantum gravity, with some unique insights on wavefunction collapse from Dieter Zeh,[13] and a discussion of the possibility of global asymmetry as it applies to whether time is primordial or fundamental.[14]

2. What Is There to Explain?

If we had a better understanding of the extreme micro- and macro-scale in Physics; time would not be an enigma at all. But from my view; people are unaware they live in the 'Goldilocks zone' on an island of comfortable norms. The realm of familiar Maths is perfectly sufficient for exploring anything anywhere on the island; so who could want more? Theoretical Physics seeks the origins of events of our current surface-layer reality in deeper layers or causal elements further back in time. But if we cut ourselves off from the possibility those origins are higher-dimensional; we omit some probable realities that solve standing problems in Physics. This is because evolutive properties arise unavoidably when the underlying or governing algebra is non-commutative or non-associative. Time's passage may thus be caught up in how higher-order spaces are reduced in dimension through cosmological transitions to obtain a cosmos with familiar properties, such as the one we inhabit. Evolution arises automatically if the dimensions are high enough, as any system's degrees of freedom approach infinity; so it becomes almost certain that a space or its underlying algebra will evolve, or is self-evolving, once certain initial conditions are met. This involves a unit of minimum uncertainty or action, and at least a modicum of energy; but physicists observe that these conditions are handily met at the Planck scale. One might question creating something from nothing, but we are uncertain absolute nothing ever existed. So there is little argument, about reasons for time to move

inexorably forward, if we assume the universe started from variations or variability and moved toward conditions.

And yet; people ask "what can cause numbers to change on their own or physical parameters change of their own accord?" It seems more reasonable to accept that virtual particles can pop in and out of existence in pairs, rather than that objects and spaces are evolving processes. But to explain time; we may need to accept that everything evolves on its own, unless it is otherwise constrained. To understand why time passes; I suggest we start with the assumption that objects and spaces are not necessarily static or fixed, and could be inherently time-evolving. This is precisely what Alain Connes proposed, when in 2000 he emphatically wrote "Noncommutative measure spaces evolve with time![15]" This result is based largely on the findings of Tomita[16] and Takesaki[17] studying modular Hilbert algebras, but Connes has generalized that work to create a new kind of differential geometry. This methodology is more nearly universal than people are aware of, however. Even if some of his ideas are not verified by current evidence; his line of reasoning is not falsified thereby, because intrinsic time evolution undeniably arises in higher-d Maths. This idea or the broader context it suggests provides a framework for a more realistic yet more flexible view of time. In this view time is, at its root, inherently or automatically progressing; originating from the dynamism of intrinsic evolution of forms and spaces in higher dimensions because the non-commutativity and/or non-associativity of their underlying algebra induces algebraic and geometric directionality and/or sequentiality.

This approach offers a unified basis for the quantum-mechanical and thermodynamic arrows of time that resolves some paradoxes or quandaries. If we see time in the tendency for objects and spaces to evolve automatically, once the degrees of freedom are high enough; we have a way to unite the Math and the Physics of time that resolves the paradox of differing time arrows from various models. This is perhaps a more major advancement than has been appreciated to this point. It restores seeing time as an accumulation of experiences or events that occur and are then behind us in a view of Quantum Mechanics where Heisenberg's uncertainty or 'unsharp measure' is due to spacetime being non-commuting. All interactions or observations accumulate because quantum-mechanical reality is like Anaximander's stream. But if we assume the evolution of the early cosmos was higher-dimensional too; we can extend intrinsic time to canonical time evolution in the initial origin at the Planck time through the inflationary period at least until baryogenesis, and likely continuing to play a part until decoupling or recombination occurred, or at a 5-d $\rightarrow$ 4-d boundary[18,19] that is seen as a black hole $\rightarrow$ white hole transition[20,21] in some theories. Time evolves from higher-d spaces through cosmological transitions, in these models. The Aikyon theory of Singh[22,23] employs Connes' intrinsic time and the octonions to explain both cosmic origins and particle physics without such transitions, but affirms other aspects of my present work. So we now examine a proposed basis for these varied ideas.

3. Explaining What We Observe

Great successes in Physics assume what we observe is built and can be explained from the bottom up, from the smallest components or the simplest applicable formulas, but there is another way to explain what we see, where nature employs top-down methods taking advantage of higher-order and higher-dimensional mathematical forms, as well. If we allow a broader spectrum of possibilities at the cosmos' origin, by assuming that both bottom-up and top-down strategies are at work at the same time; a picture emerges where dimensionality has an upper and lower limit at the outset of geometrogenesis, which later converges toward a single value.[24] This lets emergent forms be shaped by higher-d figures like the Monster Group and E_8, during the earliest phases of cosmological evolution, even while they are being built from lower-d components. From this radically top-down outlook, even when coupled with causal structure theories of quantum gravity; the conventional view using bottom-up reasoning is inside-out, and we need a view where we see the universe from the outside-in, to fully understand it. The disparity is partly due to the way most people have learned Maths, where the simplest elements are taught first and advanced ideas are derived from the simplicities. However; we know there are patterns that emerge in complex structures from which simple rules can be derived.

This dichotomy is called additive vs. formant synthesis. In the one case; you add bits of clay to make a sculpture or individual harmonic tones to create the desired waveform, while in the other you chip bits away from the outside or remove some overtones from a more complex waveform, to obtain the desired shape. Here I propose that nature has utilized and always uses both methods to shape the cosmos. It builds larger and more complex forms using smaller pieces, but it also pares physical reality down from a palette of possible forms, and borrows structure from more complex forms to create simpler ones. In this way; nature can employ the full range of Mathematics and exploit the true organizing power of what resides in higher dimensions. But the outside-in view is difficult to obtain on the island of common Maths where mathematical structures are built from the bottom-up, rather than inherited from orderly patterns in higher-level structure. To gain *that* view we need to look beyond the familiar island where features like commutativity and associativity can be taken for granted, and explore the unfamiliar expanse of non-commutative and non-associative algebras and geometry. Luckily mathematicians have already charted a lot of the content that exists in higher-d spaces, and we are learning how some of it works to shape the laws of Physics. A clear example is found in the normed division algebras, and in the reduction of the octonions to the quaternions, the complex numbers, and the reals.

$$\mathbb{O} \supset \mathbb{H} \supset \mathbb{C} \supset \mathbb{R} \tag{1}$$

The octonions are the most general number type, 8-dimensional with one real and seven imaginary parts but are non-associative; the quaternions are 4-d with three imaginary parts but are non-commutative; the complex numbers have one real

and one imaginary part, but are well-behaved; and only the reals encode a constant value. Imaginary numbers encode variation or the freedom to vary by a specified amount in a certain direction or orientation. Though they are the granddaddy of familiar types, the octonions are thought to be weird and difficult.[25] Viewing the imaginary dimensions as rotations; the octonions reduce to the quaternions if 4 of their 7 axes are fixed, that reduce to the complex numbers if 2 of the remaining 3 are fixed, and if the last rotation is halted, only the real-numbered value remains. The octonion algebra itself is a reduction because 480 possible multiplication tables before choosing a starting place and direction reduce to 16, 8 left- and 8 right-handed algebras.[26] And once any table is chosen, we must employ only that one table thereafter. Each step in any octonion calculation thus proceeds from optiony toward specificity.

$$Smooth \supseteq Top \supseteq Meas \tag{2}$$

$$Gas \supseteq Liquid \supseteq Solid \tag{3}$$

In the relations above; we see this pattern of moving from variability or variations to definite conditions as a more general bridge between the fields of Math and Physics that needs to be further explored. Moving from left to right; we see that acquiring a surface or boundary makes a space or object topological, or indicates a phase change from a gas to a liquid state. Then having a fixed metric makes objects or spaces measurable, or indicates the transition or phase change to the solid state. This reflects the sensibility of the earlier expression where applying successive constraints to the octonions eventually gets us to the real numbers. We observe that evolutive properties arising from directionality in higher-d Maths project onto or influence what comes after, such that things tend to evolve from highly variable states toward specific conditions and discrete possibilities. One might question the applicability of this line of reasoning to real-world Physics or everyday life. It certainly has a place for answering questions in early universe Cosmology, in Quantum Gravity, and the underpinnings of Quantum Mechanics. But the actual footprint of non-commutative and non-associative algebra and geometry is scarcely known. That is why it is important for today's Physics researchers to look into evolutive properties that arise in these Maths, as a possible source or root cause for the evolution of time.

The idea that spaces with certain properties have a built-in time evolution is neither well-known nor well-understood, except among a handful of ardent researchers in this arena. But when Tomita[27] opened that door, by discovering intrinsic evolution in modular Hilbert algebras; Connes and others were quick to apply that notion to a broader class of objects and spaces. To walk through that door, we must go beyond the island of familiar Maths and embrace the idea that non-commutativity and non-associativity are more of a blessing than a curse to Physics,[28] because they assure directed evolution. But we must first accept that if reality is quantum-mechanical, non-commutative ordering factors automatically arise

due to Heisenberg uncertainty making the space we live in non-commutative. Notably; this issue also arises in Relativity[29] as factor-ordering problems sometimes, when applying the equivalence principle. However, it happens often. Cycles of action must be undertaken in a specific sequence both when doing octonion algebra and with common tasks like painting or baking,[30] where a process of ratcheting accumulation happens in both cases. So the demands of additional rules of order and sequence one must learn to use non-commutative and non-associative algebras are not unnatural or onerous, and in fact represent the same laws of directional evolution all natural processes must follow.

$$\delta: \rightarrow \mathrm{Aut}(M) \tag{4}$$

Connes[31] uses the above expression, where $\mathrm{Out}(M) = \mathrm{Aut}(M)/\mathrm{Int}(M)$ the quotient group of automorphisms of M by its normal subgroup of inner automorphisms, to show how modular or Tomita flow gives rise to intrinsic time evolution. He emphatically asserts that non-commutative spaces evolve with time and this is greatly expanded in recent work by Longo,[32] showing why non-commutative spaces must evolve, and explaining the basis for time's emergence in detail. Non-commutativity is enough to assure time will evolve, and we undeniably live in a space that is non-commutative. The background state of physical reality is not what Classical Physics would imagine a vacuum to be, a static space absent of contents. We can point to quantum uncertainty or virtual particles to explain this, but time's directionality can also be seen to result from residing in a higher-d spacetime instead of 3-d space plus time. Longo extends Connes' arguments to a more robust connection between quantum mechanics and thermodynamics using the language of operator algebras, and shows how time emerges as a result. But he sees this as resulting from wavefunction collapse, and I do not see emergent time as dependent upon that feature. Here I remind the reader of an outside-in view offered by Dieter Zeh[33] suggesting that the global wavefunction is more real or fundamental than particles or quantum transitions, and persists while we measure. But recent work by Peter Morgan,[34] using the Koopman-von Neumann formalism, suggests we can choose between collapse and no-collapse models depending on the context, so long as we talk about subsequent measurements rather than states.

Things go from possibility toward actuality. However; this dynamism does not depend on properties arising in higher-d spaces and hyper-complex algebras or from infinities, as in the above examples, so higher-d attributes are not essential for top-down mechanisms to work. Gisin[35] points out that even real numbers, which are commonly viewed as fixed quantities, are only fixed by a process of determination. From an intuitionist view; the exact or precise value of a quantity is not known before it is determined in a process. So if there is a specific real number with a large number of digits; we may not find out what extended values are for a long time, and information on what is beyond a given point is no better than random data. This is like a stochastic background in the realm of the indeterminate, which influences our view of time. Gisin speaks to this issue[36] and also has novel ideas

about wavefunction collapse.[37] This work informs the Contextual Collapse model of Drossel and Ellis,[38] a top-down approach which treats the quantum measurement problem in detail, where measurement is by nature a multi-stage process. Their work addresses shortcomings in decoherence theory and other models, which they handle individually, concluding that a robust description includes both Quantum and Classical elements because any physical measurement unavoidably includes a mix of both, depending on the context. They further extend the conversation to how quantum measurements are a special case, projections of the wavefunction; of a larger class they call events which are ubiquitous in nature.

4. What Kind of Philosophy Is Best for Physics?

Is there a single best Philosophy of Physics? While some would argue that learning Philosophy is a side trip for those learning Physics; that in itself is a Philosophy about how Science is done. So it is unavoidable that the two are interwoven. This is aptly illustrated by Rovelli[39] in his paper arguing that Philosophy and Physics need each other, contrasting the schools of Isocrates and Plato. While Isocrates taught that practical skills and their application should be cultivated; Plato advocated the quest for knowledge of how or why things work, and how they came to be as they are. While Plato was among the first to be called a philosopher; we put both in that category in modern times, and their teachings are seen as two competing schools of Philosophy. Aristotle argued that general theory supports practice, and one can argue that both approaches are needed to make progress in some areas. But the debate about which philosophy is best goes on today, writ large on the face of politics. This influences funding for all areas of Science, depending on the philosophy of the prevailing political party. This split is seen to be connected to our perception of time, in that liberals have a more forward-facing view, being focused on creating a better future, while conservatives tend to focus on preserving the legacy and traditions of our past, which is more rearward-facing. Here again; people fiercely debate which philosophy is correct, or benefits us most, rather than seeing that both viewpoints are needed.

If asked what kind of Philosophy is best for Physics; I would advise people to seek an inclusive or encompassing view of reality. We need to be explorers, unafraid to look beyond the boundaries, if we want to make important discoveries or developments. And we must be wary of information silos that present a self-consistent but wrong-headed view of the world. The incorrect usage of the word recombination persists in astrophysics, for example, to describe what happens at the horizon of last scattering during decoupling; even though early-universe cosmology tells us nuclei and electrons never were combined into atoms before then. This illustrates how skilled scholars work on an island of familiar and comfortable norms, with fierce pressure to remain within the bounds of their area of specialization. The self-consistent view within the silo is often reinforced by walls that are polished

mirror smooth. An outsider and generalist like myself gets a different view from others when attending quantum gravity lectures, for example, because the common ground stands out as starkly as the differences between approaches, and I have no need for a clear winner. I think all of the current lines of research inform us in useful ways, and are not mutually-exclusive. But we should not stop exploring or seeking other options to explore, just because we have some viable explanations for gravity on the table.

So the ideal Philosophy for Physics always leaves the door open to learning more about what we think we know, discovering something new or unexpected, and so on. We are unavoidably inside a construction that is the cosmos, and further limited because we can see only what is within the Hubble radius. We are on the inside, looking out at the larger cosmos. This is why I advocate seeking the outside-in view of Math and Physics, as a radical extension of the top-down view, and as a possible cure for the view that making things from the bottom-up cannot only create all form but aptly explains all we see. A basic fact about explorers is needing somewhere to explore. Perhaps we need to get off-planet to obtain the perspective we need for the outside-in view beyond our limitations. Or maybe we only need to seek that perspective from higher ground philosophically. To see a circle in its entirety; one must be off the page, in the 3rd dimension. Likewise; to draw that circle, one must be in a dimension one higher than the surface on which you are drawing. If this as a general pattern; there is always a need to discern what the structure we are living in would look like from one dimension higher. But the block universe of General Relativity, of spacetime as seen from the outside, is a severe limitation. So we need to frame other constructs that give us better scaffolding for our ideas of the view from out there. And that is the main thrust of this paper; to inspire others who might help to create that structure.

When entertaining questions about Philosophy and Physics; one cannot forget the way Mathematics always makes its way onto the scene. Plato brought to light the notion of a mathematical ideal or archetype which is projected onto physical reality. We now have a much more sophisticated notion of how that works, but the idea persists to this day. I think we need to be hyper-dimensional Platonists, with a guide to the Atlas of Lie Groups[40] in every College library where they teach higher-level Math and Physics subjects. In this way; we will speed up the process of discovery fueled by what we know. But we need to introduce some ideas at an earlier age, so tomorrow's young people will have the benefit of what we as adults already know. This is what David Blair[41] seeks to do with the "Einstein First" program, which teaches advanced concepts based on our latest knowledge, and then works that back into the structure of what students have already learned. This allows young people to benefit from the understanding of their elders, in a context that is not intimidating for teachers with a limited knowledge of Relativity. We should attempt to do the same with other areas of Mathematics and Physics so the next generation will have the benefit of what we have learned. This might

not be possible if we teach things solely from the bottom-up view. What if we had young people building Zome models[42] of E_8 instead? At least we need to give kids a chance to see what these figures look like, at Science Centers and Math Museums.

The nexus of Philosophy and Physics must be strong, for either topic to provide useful information. This means it is incumbent on the organizers of events and publications to provide a context for insightful contributions that forge a stronger connection between what appear to be separated views, or to push the envelope of developments on the outskirts. The 16th Marcel Grossmann conference was exemplary on both of these fronts, in its spirit of cooperation, openness, and inclusivity. The organizers deliberately included minority views and off-limits topics in both the choice of plenary speakers and in the various breakout sessions. So people were encouraged to look outside of their silos somewhat, and to explore what is beyond the island of comfortable mainstream assumptions. During MG16; a Philosophy of Physics approaching the ideal was observed. But in the world at large; there is a lot of progress to be made forging connections to support a higher ideal. What we see now is a world that largely ignores the advice of both philosophers and physicists, in favor of following popular icons. There is much to do, if we want the ideas that enlighten physicists to be meaningful for everyone, but things being discovered in Physics have a meaning that unavoidably affects all equally. There is no point to arguing Physics is devoid of meaning. Therefore our Philosophy must include the notion that Physics is relevant.

We owe much of our current understanding to the philosophers of ancient Greece, not least the notion of atomism with indivisibly small bits or increments in the cosmos or the way it is constructed, as was well-explained in Rovelli's book.[43] Leucippus likely brought the seed of this idea to Democritus from Anaximander's school. But we can see that atomism applied to Plato's comment that time is the projected image of eternity, shows up as every atom in the stream of Anaximander having a duration or persistence in time. This is observed as a half-life, when examining the same phenomenon in nuclear or sub-atomic Physics, due in part to the relativistic effects of time dilation. But here I note that parcels of space needed to have persistence in time already for those particles to exist. That is why it is important to continue our exploration of quantum gravity theories and try to understand the cosmological context for these theories in a broader way. So physicists should see it as a responsibility to build bridges between the philosophically disconnected islands of thought, by applying a Philosophy of Physics that supports our common endeavor to learn the secrets of the cosmos. Encouraging students as well as researchers to explore beyond what is known is important, and willingness to question our own boundaries is an important first step. We must be ready to leave the world of comfortable norms behind, when examining cosmic origins, to fully understand how things got to be the way they are now.

References

1. Anaximander, (c. 611-546 BC).
2. Plato, *Timaeus* 37d (c. 360 BC).
3. C. Rovelli, *Reality is not What it Seems — the Journey to Quantum Gravity* (Allen Lane, 2016).
4. J. J. Dickau, *Gravitation by Condensation*, in *Fundamental Physics and Physics Education Research*, eds. B. G. Sidharth, J. C. Murillo, M. Michelini, C. Perea (Springer, Cham., 2020) https://doi.org/10.1007/978-3-030-52923-9_7.
5. J. J. Dickau, The Intrinsic Time of Alain Connes and Evolutive Properties in higher-d Algebras, *Prespacetime Jour.*, **12**, 3, pp. 283–299 (2020).
6. A. Einstein, quote from a debate with Henri Bergson (1922).
7. C. Rovelli, Physics Needs Philosophy. Philosophy Needs Physics, *Found. Phys.*, **48**, 5, pp. 481–491 (2018).
8. J. J. Dickau, Confluence of Theories, *Prespacetime Jour.*, **11**, 4, pp. 346–351 (2020).
9. A. Connes, A View of Mathematics, unpublished essay, pg. 16, which now part of the *UNESCO Encyclopedia of Life Support Services* (2004).
10. R. Longo, The Emergence of Time, *Expositiones Mathematicae*, **38**, 2, pp. 240–258, (2020) https://doi.org/10.1016/j.exmath.2020.01.005, arXiv:1910.13926.
11. N. Gisin, Mathematical languages shape our understanding of time in physics, *Nature Physics*, **16**, pp. 114–116, (2020) https://doi.org/10.1038/s41567-019-0748-5, arXiv:2002.01653.
12. B. Drossel and G. F. R Ellis, Contextual Wavefunction Collapse: An integrated theory of quantum measurement, *New Jour. Phys.*, **20**, 113025, arXiv:1807.08171 (2018).
13. H. D. Zeh, There are no quantum jumps, nor are there particles!, *Phys. Lett. A*, **172**, 4, pp. 189–192 (1993).
14. D. Schleicher and D. Dudko, Core Entropy of Quadratic Polynomials, *Arnold Math Jour.*, **6**, pp. 333–385, arXiv:1412.8760 (2020).
15. A. Connes, *Noncommutative Geometry Year 2000*, in *Visions in Mathematics: GAFA 2000 Special volume, Part II*, pp. 481–559, Birkhäuser Basel, arXiv:math/0011193 (2000).
16. M. Tomita, Quasi-standard von Neumann algebras, unpublished preprint (1967).
17. M. Takesaki, Tomita's Theory of Modular Hilbert Algebras and its Applications, *Lect. Notes Math*, Vol. 128 Springer-Verlag, Berlin-New York (1970).
18. G. Dvali, G. Gabadadze, and Porrati, 4D Gravity on a Brane in 5D Minkowski Space, *Phys. Lett. B*, **485**, 1–3 pp. 208–214, arXiv:hep-th/0005016 (2000).
19. C. de Rham, et al. Cascading Gravity: Extending the Dvali-Gabadadze-Porrati Model to Higher Dimension, *Phys. Rev. Lett.*, **100**, 251603 (2008).
20. R. Pourhasan, N. Afshordi, and R. Mann, Out of the White Hole: a holographic origin for the Big Bang, *Jour. of Cosmo. and Astroparticle Phys.*, **2014**, arXiv:1309.1487; (2014).
21. N. Poplawski, Radial motion into an Einstein–Rosen bridge, *Phys. Lett. B*, **687**, 2–3, pp. 110–113, arXiv:0902.1994 (2010).
22. T. P. Singh, Octonions, trace dynamics and non-commutative geometry: A case for unification in spontaneous quantum gravity, *Z. Naturforschung A*, **75**, 12, pp. 1051–1062, arXiv:2006.16274; (2020).
23. T. P. Singh, Trace dynamics and division algebras: towards quantum gravity and unification, *Z. Naturforschung A*, **76**, 2, 131–162, arXiv:2009.05574 (2021).
24. J. J. Dickau, Bimetric Convergence, *Prespacetime Jour.*, **11**, 4, pp. 294–304 (2020).
25. J. C. Baez, The octonions, *Bull. Am. Math Soc.*, **39**, pp. 145–205, arXiv:math/0105155 (2002).

26. R. Lockyer, 16 Octonion Algebras – 30 Aliases – 480 Representations, preprint, (2010) can be found at: `http://www.octospace.com/files/16_Octonion_Algebras_-_480_Representations.pdf`.
27. M. Tomita, ibid Ref. 16 (1967).
28. P. C. Kainen, *An octonion model for physics*; in *Proceedings of 4th Intl. Conf. on Emergence, Coherence, Hierarchy, and Organization (ECHO IV)*, Odense, Denmark (2000) found at: `http://faculty.georgetown.edu/kainen/octophys.pdf` and also at: `https://www.researchgate.net/publication/2472211_An_Octonion_Model_for_Physics`.
29. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Factor-ordering problems in the Equivalence Principle*, chap. 16, sec. 3, pp. 388–392 in: *Gravitation*, W.H. Freeman and Co., New York (1973).
30. J. J. Dickau, Painting, Baking and Non-Associative Algebra, *Prespacetime Jour.*, **11**, 3, pp. 265–268 (2020).
31. A. Connes, ibid Ref. 15 (2000).
32. R. Longo, ibid Ref. 10 (2020).
33. H. D. Zeh, ibid Ref. 13 (1993).
34. P. W. Morgan, The collapse of a quantum state as a signal analytic sleight of hand, preprint, arXiv:2101.10931 (2021).
35. N. Gisin, ibid Ref. 11 (2020).
36. N. Gisin, *Time Really Passes, Science Can't Deny That*, in *Time in Physics* eds. R. Renner and S. Stupar, Springer, pp. 1–15, arXiv:1602.01497 (2017).
37. N. Gisin, *Collapse. What Else?* in *Collapse of the Wavefunction* Ed. Shan Gao, Cambridge Uni. Press, pp. 207–224, arXiv:1701.08300 (2018).
38. B. Drossel and G. F. R Ellis, ibid Ref. 12 (2018).
39. C. Rovelli, ibid Ref. 3 (2018).
40. *Atlas of Lie Groups and Representations*, (2007) with much explanatory material found at: `http://www.liegroups.org/`.
41. D. Blair, *The Einstein-First Project*, (2021) much information found at: `https://www.einsteinianphysics.com/`.
42. D. A. Richter, *Gosset's Figure in 8 Dimensions, A Zome Model*, found at: `http://homepages.wmich.edu/~drichter/gossetzome.htm` which is constructed from Zometools, found at: `https://www.zometool.com/`.
43. C. Rovelli, ibid Ref. 7 (2016).

A glimpse to Feynman's contributions to the debate on the foundations of quantum mechanics

M. Di Mauro

Dipartimento di Matematica, Università di Salerno, Via Giovanni Paolo II, Fisciano (SA), 84084, Italy
E-mail: madimauro@unisa.it

S. Esposito* and A. Naddeo†

INFN, Sezione di Napoli, C. U. Monte S. Angelo, Via Cinthia, Napoli, 80125, Italy
**E-mail: sesposit@na.infn.it*
†*E-mail: anaddeo@na.infn.it*

The broad debate on foundational issues in quantum mechanics, which took place at the famous 1957 Chapel Hill conference on *The Role of Gravitation in Physics*, is here critically analyzed with an emphasis on Richard Feynman's contributions. One of the most debated questions at Chapel Hill was whether the gravitational field had to be quantized and its possible role in wave function collapse. Feynman's arguments in favor of the quantization of the gravitational field, based essentially on a series of gedanken experiments, are here discussed. Then the related problem of the wave function collapse, for which Feynman hints to decoherence as a possible solution, is discussed. Finally, another topic is analyzed, concerning the role of the observer in a closed Universe. In this respect, Feynman's many-worlds characterization of Everett's approach at Chapel Hill is discussed, together with later contributions of his, including a kind of Schrödinger's cat paradox, which are scattered throughout the 1962-63 Lectures on Gravitation. Philosophical implications of Feynman's ideas in relation to foundational issues are also discussed.

Keywords: Gedanken experiment; Wave function collapse; Many-worlds.

1. Introduction: The Chapel Hill conference

Richard Feynman's most famous contribution in quantum theory is undoubtedly his celebrated path integral approach.[1] His contributions to the debate on interpretational issues are instead much less well known.[2,3] The first such contributions took place in the wide discussions which characterized the 1957 Chapel Hill conference,[4] whose title was: *The Role of Gravitation in Physics*. The Chapel Hill conference, which played a key role in triggering the so called *Renaissance* of general relativity[5] was of capital importance in establishing the future research lines in the field. Broadly, the main tracks were[6]: classical gravity, quantum gravity, and the classical and quantum theory of measurement (as a link between the previous two topics). In particular, foundational quantum issues were widely discussed by researchers attending the conference. The main motivation for addressing them came from the

fundamental question of whether the gravitational field had to be quantized, like other fundamental fields, or not. As its title declares, the ambitious goal of the conference was the merging of general relativity with the rest of physics and, especially, with the world of elementary particle physics, which was ruled by quantum mechanics, hence it was (and it still is, of course) very important to understand physics in regimes where both gravity and quantum mechanics become important. This was clearly stated by Peter G. Bergmann in the opening session of the second half of the conference, which was in fact focused on problems within the quantum domain (Ref. 4, p. 165):

> Physical nature is an organic whole, and various parts of physical theory must not be expected to endure in "peaceful coexistence." An attempt should be made to force separate branches of theory together to see if they can be made to merge, and if they cannot be united, to try to understand why they clash. Furthermore, a study should be made of the extent to which arguments based on the uncertainty principle force one to the conclusion that the gravitational field must be subject to quantum laws: (a) Can quantized elementary particles serve as sources for a classical field? (b) If the metric is unquantized, would this not in principle allow a precise determination of both the positions and velocities of the Schwarzschild singularities of these particles?

The answer to these questions was expected to lead to a novel perspective on the ordinary notions of space and time, by introducing uncertainty relations for quantities such as distances and volumes. Physicists hoped that, as a result, the divergences which plagued quantum field theory would get suppressed by the gravitational field. Along with the formal discussion on the main approaches to the quantization of the gravitational field (namely the canonical, the functional integral and the covariant perturbative approaches), which would have been developed in the following decades, a lot of time was devoted to conceptual issues, especially in discussion sessions. A broad debate arose on the problem of quantum measurement, the main conceptual question being:

> What are the limitations imposed by the quantum theory on the measurements of space-time distances and curvature? (Ref. 4, p. 167)

or, equivalently

> What are the quantum limitations imposed on the measurement of the gravitational mass of a material body, and, in particular, can the principle of equivalence be extended to elementary particles? (Ref. 4, p. 167)

As the editors of the written records of the conference[4] emphasized, an answer to the above questions could not be simply found in dimensional arguments, since the

Planck mass does not set a lower limit to the mass of a particle whose gravitational field can be measured. Indeed, a simple argument shows that the gravitational field of *any* mass can in principle be measured, thanks to the "long tail" of the Newtonian force law (Ref. 4, pp. 167-8).

The great part of the discussions of interest to us here developed in Section VIII of the conference (Ref. 4, pp. 243-60), where the focus was mainly on the contradictions eventually arising in the logical structure of quantum theory if gravity quantization is not assumed, but in fact many even more foundamental issues for quantum physics in general were touched as well. Feynman greatly contributed to these discussions, proposing several gedanken experiments in order to argue in favor of the necessity of gravitational quantization, and hinting to decoherence as a viable solution to the problem of wave function collapse. His contributions in fact triggered a wide debate on the quantum measurement problem and on the existence and meaning of macroscopic quantum superpositions. Further, in the subsequent session of the conference, which was the closing one (Ref. 4, pp. 263-278), he gave a "many-worlds" characterization of Hugh Everett III's relative state interpretation of quantum mechanics,[7] which had just been described for the first time by Everett's advisor, John A. Wheeler. It is interesting to notice that, in fact, the Chapel Hill conference was one of the few places where Feynman was directly involved in discussions about the foundations and interpretation of quantum mechanics.[2]

The Chapel Hill discussions played a pivotal role in triggering subsequent research on foundational issues, mainly on the quantum measurement problem and on the various possible interpretations of quantum mechanics, with an emphasis on the consequences of Everett's one. In this respect it is worth mentioning further contributions by Feynman, who despite not being often involved with foundational quantum issues, nonetheless kept thinking deeply on them also in the following years, especially when working on gravity, as hinted for example in a letter written in 1961 to Viktor Weisskopf[8]:

> How can we experimentally verify that these waves are quantized? Maybe they are not. Maybe gravity is a way that quantum mechanics fails at large distances.

as well as in several suggestions and considerations scattered throughout his 1962-63 graduate lectures on gravitation.[9]

In this paper, we historically and critically analyze Feynman's contributions to the debate about the foundations of quantum mechanics and gravity, with an emphasis on the Chapel Hill discussions, which constitute the main source about his thoughts on the matter. In particular, Section 2 deals with the problem of the quantization of the gravitational field while in Sections 3 and 4 the focus is on wave function collapse following a measurement and on Everett's relative state interpretation of quantum mechanics. Finally our concluding remarks are summarized in Section 5.

2. Quantization of the gravitational field: Gedanken experiments

In this Section we report on Feynman's arguments in favor of the quantization of the gravitational field. As a matter of fact, Feynman believed that nature cannot be half classical and half quantum and shared Bergmann's general ideas, as recalled by himself in some concluding remarks made at the end of the conference:

> The questions raised in the last three days have to do with the relation of gravity to the rest of physics. We have gravity - electrodynamics - quantum theory - nuclear physics - strange particles. The problem of physics is to put them all together. The original problem after the discovery of gravity was to put gravity and electrodynamics together since that was essentially all that was known. Therefore, we had the unified field theories. After quantum theory one tries to quantize gravity (Ref. 4, p. 272).

In particular, for Feynman gravity, like the other fundamental interaction that we experience at the macroscopic, classical level, i.e. electromagnetism, has a quantum foundation (see e.g. Refs. 10–12). Coherently with this, he developed an approach to quantum gravity (whose first hints were given at Chapel Hill, cf. Ref. 4, pp. 271-276) which was characterized by being fully quantum from the beginning.[9,13]

Feynman's arguments in favor of the quantization of gravity were totally different from the usual dimensional arguments, which rely on the assumption that gravity has to dominate over all other interactions at the Planck scale. According to him quantum effects involving gravity should be subject to probing without invoking such high scales. He provided further evidence to support this claim by putting forward thought experiments designed to show that, if quantum mechanics is required to hold for objects massive enough to produce a detectable gravitational field (the opposite would require a modification of quantum mechanics), then the gravitational field has to be quantized, in order to avoid contradictions. This view supported the hope that quantization of the metric would help taming the divergences present in quantum field theory and, in this way, it was relevant also for the theory of elementary particles. Let us thus retrace the discussion in more detail.

Section VIII began with Thomas Gold stating that, in the absence of phenomena not explainable by classical gravity, the only way to argue in favor of quantization of gravity is the presence of logical contradictions, of which he was not convinced (Ref. 4, pp. 243-244). Bryce S. DeWitt pointed out that difficulties may arise in the presence of quantized matter fields depending on the choice of the quantum expectation value of the stress-energy tensor as the source of the gravitational field (Ref. 4, p. 244). Indeed a measurement may change this expectation value, which implies a change of the gravitational field itself. The classical theory of gravitation is valid because fluctuations are negligible at the scale where gravitational effects become sizeable. At this point Feynman put forward his first thought experiment, which was a variant of the two-slit diffraction experiment, in which a mass indicator has been put behind the two-slit wall. Within a space-time region whose linear

dimensions are of order L in space and L/c in time, the uncertainty on the gravitational potential (divided by c^2 so that it is dimensionless and thus homogeneous to a metric) is in general $\Delta g = \sqrt{\frac{hG}{c^3L^2}} = \frac{L_P}{L}$, where $L_P = \sqrt{\frac{hG}{c^3}}$ is the Planck length.[a] The order of magnitude of the potential generated by a mass M within the spatial part of the considered region is $g = \frac{MG}{Lc^2}$, which implies $\Delta g = \frac{\Delta MG}{Lc^2}$. A comparison with the previous general expression gives a mass uncertainty $\Delta M = \frac{c^2 L_P}{G} \approx 10^{-5}$ grams, if the time of observations is less than L/c. On the other side, by allowing an infinite time, M would be obtained with infinite accuracy. This option, according to Feynman, could not take place for a mass M fed into the two-slit apparatus, so that the apparatus will not be able to uncover the difficulty unless M is at least of order 10^{-5} grams. Thus he concluded that

> Either gravity must be quantized because a logical difficulty would arise if one did the experiment with a mass of order 10^{-4} grams, or else [...] quantum mechanics fails with masses as big as 10^{-5} grams (Ref. 4, p. 245).

Subsequent discussions moved on more formal matters such as the meaning of the equivalence principle in quantum gravity, with Feynman contrasting Helmut Salecker's suggestion of a possible violation of the equivalence principle in the quantum realm. Then a further remark by Salecker himself brought the participants' attention back to the topical issue, the necessity of gravitational quantization. In particular, as explained by an Editor's note (see Ref. 4, p. 249), Salecker hinted to the possibility to build up an action-at-a-distance theory of gravitation in whole analogy to the electromagnetic case, with charged quantized particles acting as a source of a unquantized Coulomb field. At this stage Frederik J. Belinfante proposed the quantization of both the static part and the transverse part of the gravitational field (the last one describing gravitational radiation) as a way to circumvent difficulties related to the choice of the expectation value of the stress energy tensor as the source of the gravitational field. As noticed by Zeh,[2] Belinfante's proposal reflects his ideas on the ontology of quantum mechanics, which point toward an epistemic rather than ontic interpretation of the wave function:

> "There are two quantities which are involved in the description of any quantized physical system. One of them gives information about the general dynamical behavior of the system, and is represented by a certain operator (or operators). The other gives information about our knowledge of the system; it is the state vector [...] the state vector can undergo a sudden change if one makes an experiment on the system. The laws of nature therefore unfold continuously only as long as the observer does not bring extra knowledge of his own into the picture." (Ref. 4, p. 250).

[a] This follows from an argument given by Wheeler in Ref. 4, pp. 179-180, involving a path integral for the gravitational field.

Belinfante's description used the Heisenberg picture, at variance with Feynman who reasoned in terms of wave functions as dynamical objects. Indeed, according to Belinfante "the wave function [...] must change for reasons beyond the system's physical dynamics. He does not refer to ensembles of wave functions or a density matrix in order to represent incomplete knowledge" (Ref. 2, p. 65).

Feynman promptly replied with his second gedanken experiment, a Stern-Gerlach experiment with a gravitational apparatus. More in detail, he considered a spin-1/2 particle going through a Stern-Gerlach apparatus and then crossing the first or the second of two counters (denoted as 1 and 2, respectively), each one connected by means of a rod to an indicator. He took the indicator as a little ball with a diameter of 1 cm, going up or down depending on the position of the object, at counter 1 or 2, respectively. Quantum mechanics, as underlined by Feynman, provides in principle an amplitude for the ball up and an amplitude for the ball down[b]. However, the ball is chosen to be macroscopic, so as to be able to produce a detectable gravitational field, which in turn can move a probe ball. In other words a channel between the object and the observer is established via the gravitational field. This ideal experiment led Feynman to infer the following conclusion about gravity quantization:

> Therefore, there must be an amplitude for the gravitational field, provided that the amplification necessary to reach a mass which can produce a gravitational field big enough to serve as a link in the chain does not destroy the possibility of keeping quantum mechanics all the way. There is a bare possibility (which I shouldn't mention!) that quantum mechanics fails and becomes classical again when the amplification gets far enough, because of some minimum amplification which you can get across such a chain. But aside from that possibility, if you believe in quantum mechanics up to any level then you have to believe in gravitational quantization in order to describe this experiment (Ref. 4, p. 251).

A subsequent answer to a question by Hermann Bondi[c] further highlighted, according to Zeh (Ref. 2 p. 67), Feynman's position against the epistemic interpretation of the wave function:

> I don't really have to measure whether the particle is here or there. I can do something else: I can put an inverse Stern-Gerlach experiment on and bring the beams back together again. And if I do it with great precision, then I arrive at a situation which is not derivable simply from the information that there is a 50 percent probability of being here and a 50 percent probability of

[b] Zeh[2] points out how Feynman's description of the measurement process is very close to the standard measurement and registration device proposal by John von Neumann.[14]

[c] "What is the difference between this and people playing dice, so that the ball goes one way or the other according to whether they throw a six or not?" (Ref. 4, p. 252).

> being there. In other words, the situation at this stage is not 50-50 that the die is up or down, but *there is an amplitude* that it is up and an amplitude that it is down – a complex amplitude – and as long as it is still possible to put those amplitudes together for interference you have to keep quantum mechanics in the picture. (Ref. 4, p. 252, our emphasis)

3. Wave function collapse

The aim of this Section is to highlight Feynman's ideas about the quantum measurement problem. Indeed, from the last sentence of the last quote, one can infer a clear reference to the problem of wave function collapse, as well as a hint to decoherence as a possible solution. A further suggestion can be found in a subsequent remark:

> Well, it's a question of what goes on at the level where the ball flips one way or the other. In the amplifying apparatus there's already an uncertainty - loss of electrons in the amplifier, noise, etc. - so that by this stage the information is completely determined. Then it's a die argument. You might argue this way: Somewhere in your apparatus this idea of amplitude has been lost. You don't need it any more, so you drop it. The wave packet would be reduced (or something). Even though you don't know where it's reduced, it's reduced. And then you can't do an experiment which distinguishes *interfering* alternatives from just plain odds (like with dice). (Ref. 4, p. 252, our emphasis)

According to Feynman, wave packet reduction occurs somewhere in his experimental apparatus thanks to the amplifying mechanism, so that a huge amount of amplification (via the macroscopic gravitational field of the ball) may be effective in changing amplitudes to probabilities. Then he wondered about the possibility to devise an experiment able in principle to avoid the wave packet reduction due to the amplification process. Subsequent criticism by Leon Rosenfeld and Bondi further stimulated Feynman's intuition. So he was led to envisage a sort of quantum interference in his experiment, driven by the gravitational interaction between macroscopic balls, which could be described by means of a quantum field with suitable amplitudes taking a value or another value, or to propagate here and there. But, as Bondi suggested, any irreversible element should be removed, such as for instance the possibility of gravitational links to radiate. Probably, this is another point in which a further hint to the role of decoherence in destroying quantum interference can be recognized, even if in Bondi's and Feynman's words a reference is made only to classical irreversibility.[2] In fact the meaning of decoherence, its origin and its role in smearing out phase relations as well as in triggering the transition to classicality, were still unclear in the late 1950s[d]. The discussion went on with Feynman arguing that quantum interference might eventually take place with a mass of macroscopic

[d]For a historical and research account on decoherence, see Ref. 15 and references therein.

size, say about 10^{-5} gram or 1 gram, and hinting to the possible role of gravity in destroying quantum superpositions. In his words:

> There would be a *new principle*! It would be *fundamental*! The principle would be: – *roughly*: *Any piece of equipment able to amplify by such and such a factor* (10^{-5} grams or whatever it is) *necessarily must be of such a nature that it is irreversible.* It might be true! But at least it would be fundamental because it would be a new principle. There are two possibilities. Either this principle – this missing principle – is right, *or* you can amplify to any level and still maintain interference, in which case it's absolutely imperative that the gravitational field be quantized... *I believe*! *or* there's another possibility which I have't thought of (Ref. 4, pp. 254-255, emphasis in original).

The same ideas would have been pursued later by Feynman in his Lectures on Gravitation,[9] where he considered "philosophical problems in quantizing macroscopic objects" and hinted to the possibility of a failure of quantum mechanics induced by gravity:

> I would like to suggest that it is possible that quantum mechanics fails at large distances and for large objects. Now, mind you, I do not say that I think that quantum mechanics *does* fail at large distances, I only say that it is not inconsistent with what we do know. If this failure of quantum mechanics is connected with gravity, we might speculatively expect this to happen for masses such that $\frac{GM^2}{\hbar c} = 1$, of M near 10^{-5} grams, which corresponds to some 10^{18} particles (Ref. 9, pp. 12-13).

Within the same set of lectures (Ref. 9, p. 14), the possibility is recognized that amplitudes may reduce to probabilities for a sufficiently complex object, thanks to a smearing effect on the evolution of the phases of all parts of the object. Such a smearing effect could have a gravitational origin. A similar idea is expressed in the end of the letter to Weisskopf,[8] as already mentioned in the Introduction. This shows how Feynman was open-minded with respect to all possibilities, despite his strong belief in the quantum nature of reality.

The debate on gravity quantization here highlighted and, in particular, Feynman's deep insights, triggered subsequent research on the possibility of a gravity-induced collapse of wave function, as a viable solution to the measurement problem in quantum mechanics.[16–25] The idea is attractive since gravity is ubiquitous in nature, and gravitational effects depend on the size of objects, so it has been greatly developed in the subsequent years up to the present day. Along this line of thinking, Roger Penrose suggested that a conflict emerges when a balanced superposition of two separate wave packets representing two different position of a massive object is considered.[22,23] In his words (see Ref. 26, p. 475): "My own point of view is that as soon as a *significant* amount of space-time curvature is introduced, the rules of quantum linear superposition must fail". Clearly the conflict is the result of putting

together the general covariance of general relativity and the quantum mechanical superposition principle. By assuming in each space-time the validity of the notions of stationarity and energy, and by taking the difference between the identified time-translation operators as a measure of the ill-definiteness of the final superposition's energy, the decay time for the balanced superposition of two mass distributions can be estimated:

$$t_D = \frac{\hbar}{\Delta E_{grav}}. \tag{1}$$

Here ΔE_{grav} is the gravitational self-energy of the difference between the mass distributions of each of the two locations of the object. This means that massive superpositions would immediately undergo collapse. The same idea led Penrose to introduce the so called Schrödinger-Newton equation,[24] which governs a peculiar non-linear evolution of the center-of-mass wave function. Its soliton-like dynamics is the result of a competition between a partial shrinking due to the non-linear term and a partial spreading due to the usual dynamical term.

More recently, other collapse models(called *dynamical* collapse models) have been put forward,[27–30] where the collapse of the wave function is induced by a different mechanism, i.e. the interaction with a random source such as an external noise source.

From the experimental side, a lot of proposals have been made as well, mainly aimed to explore a parameter range where both quantum mechanics and gravity are significant. For instance it has been possible to create a quantum superposition state with complex organic molecules with masses of the order $m = 10^{-22}$ kg.[31,32] Recent experimental proposals involve matter-wave interferometers,[33–35] quantum optomechanics[36–39] and magnetomechanics.[40] Today a major challenge is, on one hand, to design viable experiments at the interface between the quantum and classical worlds, while on the other hand to reveal and discriminate gravity-induced decoherence from environmental decoherence in such experiments. However, despite such a huge theoretical and experimental effort, a definite answer to the quantum measurement problem as well as to the problem of the emergence of classical from the quantum world is still lacking.

4. The role of the observer in a closed Universe

In this Section we deal with the issue of the role of the observer in a closed Universe, as emerged from discussions carried out in the closing session of Chapel Hill conference. Here Wheeler gave the first public presentation of Everett's relative-state interpretation of quantum mechanics,[7,41] as follows:

> General relativity, however, includes the space as an integral part of the physics and it is impossible to get outside of space to observe the physics. Another important thought is that the concept of eigenstates of the total

> energy is meaningless for a closed Universe. However, there exists the proposal that there is one "universal wave function". This function has already been discussed by Everett, and it might be easier to look for this "universal wave function" than to look for all the propagators (Ref. 4, p. 270).

Feynman promptly replied by characterizing of Everett's approach as "many-worlds":

> The concept of a "universal wave function" has serious conceptual difficulties. This is so since this function must contain amplitudes for all possible worlds depending on all quantum-mechanical possibilities in the past and thus one is forced to believe in the equal reality of an infinity of possible worlds (Ref. 4, p. 270).

The same idea will be presented some years later in the Lectures on Gravitation (Ref. 9, pp. 13-14), where Feynman also discussed the role of the observer in quantum mechanics. The Schrödinger's cat paradox allowed him to illustrate the difference between the results of a measurement carried out by an external as well as an internal observer. While the external observer describes his results by an amplitude, with the system collapsing into a well-defined final state after the measurement, according to the internal observer the results of the same measurement are given by a probability. Clearly the absence of an external observer leads to a paradox, which is much more effective when considering the whole Universe as described by a complete wave function. This Universe wave function is governed by a Schrödinger equation and implies the presence of an infinite number of amplitudes, which bifurcate from each atomic event. In other words, according to Feynman, the Universe is constantly spitting into an infinite number of branches, as a result of the interactions between its components. Here the key observation is that interactions play the role of measurements. As a consequence, an inside observer knows which branch the world has taken, so that he can follow the track of his past. Feynman's conclusion brings into play a conceptual problem:

> Now, the philosophical question before us is, when we make an observation of our track in the past, does the result of our observation become real in the same sense that the final state would be defined if an outside observer were to make the observation (Ref. 9, p. 14)?

A further discussion of the meaning of the wave function of the Universe is carried out later in the Lectures on Gravitation (Ref. 9, pp. 21-22). Here Feynman restated his "many-worlds" characterization of Everett's approach with reference to a "cat paradox on a large scale", from which our world eventually could be obtained by a "reduction of the wave packet". Concerning this reduction, he wondered about the relation between Everett's approach and collapse mechanisms of whatever origin. Perhaps it may be relevant to quote a comment by John P. Preskill, within a talk

about the Feynman legacy, given at the APS April Meeting (cf. Ref. 42, slide 29):

> When pressed, Feynman would support the Everett viewpoint, that all phenomena (including measurement) are encompassed by unitary evolution alone. According to Gell-Mann, both he and Feynman already held this view by the early 1960s, without being aware of Everett's work[e]. However, in 1981 Feynman says of the many-worlds picture: "It's possible, but I am not very happy with it".

Historically, Everett's proposal[7] was the first attempt to go beyond the Copenhagen interpretation in order to apply quantum mechanics to the Universe as a whole. To this end, the well known separation of the world into "observer" and "observed" has to be superseded by promoting the observer to be part of the system, while the usual quantum rules are still effective in measuring, recording or doing whatever operation. Quantum fluctuations of space-time in the very early Universe have also to be properly taken into account. On the other hand, this proposal lacks an adequate description of the origin of the quasi-classical realm as well as a clear explanation of the meaning of the branching of the wave function.

Everett's work has been further developed by many authors[43–48] and its more recent generalization is known as *decoherent histories approach to quantum mechanics of closed systems*. A characterizing feature of this formulation is that neither observers nor their measurements play a prominent role. Furthermore the so called *retrodiction*, namely the ability to construct a history of the evolution of the Universe towards its actual state by using today's data and an initial quantum state, is allowed. The process of prediction requires to select out decoherent sets of histories of the Universe as a closed system, while decoherence in this context plays the same role of a measurement within the Copenhagen interpretation. Decoherence is a much more observer-independent concept and gives a clear meaning to Everett's branches, the main issue being to identify mechanisms responsible for it. In particular, it has been shown that decoherence is frequent in the Universe for histories of variables such as the center of mass positions of massive bodies.[49]

5. Concluding remarks

In this paper we have retraced Feynman's thoughts on foundational issues in quantum mechanics. Such ideas were rarely expressed by him, who mainly linked his name to his masterful application of thus theory, but they clearly emerged whenever he focused on the problem the interface of quantum mechanics and gravity, with the ensuing great conceptual questions. These problems, such as that of the collapse of the wave function, of macroscopic superpositions and of observers which are parts of the observed system itself, are of course not limited to the gravitational

[e]We do not agree that Feynman was not aware of Everett's work in the 1960s, since he commented on it at Chapel Hill in 1957.

realm, but lie at the very heart of quantum physics. Despite much progress having been achieved since Feynman's times, part of which had been anticipated by him and other participants to the Chapel Hill conference, much work still needs to be done, before his most famous remark[50] "I think I can say that nobody understands Quantum Mechanics" can be considered to be no longer true.

References

1. R. P. Feynman, Space-time approach to non-relativistic quantum mechanics, *Rev. Mod. Phys.* **20**, 367 (1948).
2. H. D. Zeh, Feynman's interpretation of quantum theory, *Eur. Phys. J. H* **36**, 63 (2011).
3. M. Di Mauro, S. Esposito and A. Naddeo, A road map for Feynman's adventures in the land of gravitation, *Eur. Phys. H* **46**, 22 (2021).
4. C. DeWitt-Morette and D. Rickles, *The Role of Gravitation in Physics*, Report from the 1957 Chapel Hill Conference (Edition Open Access, 2011).
5. A. S. Blum, R. Lalli and J. Renn, The renaissance of General Relativity: How and why it happened, *Annalen der Physik* **528**, 344 (2016).
6. P. G. Bergmann, Summary of the Chapel Hill Conference, *Rev. Mod. Phys.* **29**, 352 (1957).
7. H. Everett III, "Relative state" formulation of quantum mechanics, *Rev. Mod. Phys.* **29**, 454 (1957).
8. R. P. Feynman, unpublished letter to Victor F. Weisskopf, February 1961; in *Richard P. Feynman Papers*, California Institute of Technology Archives, Box 3, Folder 8.
9. R. P. Feynman, F. B. Morinigo, W. G. Wagner and B. Hatfield, *Feynman Lectures on Gravitation* (Addison-Wesley, Reading, MA, 1995).
10. R. De Luca, M. Di Mauro, S. Esposito and A. Naddeo, Feynman's different approach to electromagnetism, *Eur. J. Phys.* **40**, 065205 (2019).
11. M. Di Mauro, R. De Luca, S. Esposito and A. Naddeo, Some insight into Feynman's approach to electromagnetism, *Eur. J. Phys.* **42**, 025206 (2021).
12. R. P. Feynman, *Electrostatics, Electrodynamics, Matter-Waves Interacting, Relativity*, lectures at Hughes Aircraft Company, Notes taken and transcribed by John T. Neer, 1967-68.
13. R. P. Feynman, Quantum theory of gravitation, *Acta Phys. Polon.* **24**, 697 (1963).
14. J. von Neumann, *Mathematische Grundlagen der Quantenmechanik (Mathematical foundations of quantum mechanics)* (Springer, Berlin, 1932), Chap. 6.
15. W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, *Rev. Mod. Phys.* **75**, 715 (2003).
16. F. Károlyházy, Gravitation and quantum mechanics of macroscopic objects, *Il Nuovo Cimento A* **42**, 390 (1966).
17. A. Frenkel, Spontaneous localizations of the wave function and classical behavior, *Found. Phys.* **20**, 159 (1990).
18. A. Frenkel, A Tentative Expression of the Károlyházy Uncertainty of the Space-Time Structure Through Vacuum Spreads in Quantum Gravity, *Found. Phys.* **32**, 751 (2002).
19. L. Diosi, Gravitation and quantum mechanical localization of macroobjects, *Phys. Lett. A* **105**, 199 (1984).
20. L. Diosi, Models for universal reduction of macroscopic quantum fluctuations, *Phys. Rev. A* **40**, 1165 (1989).

21. L. Diosi, Notes on certain Newton gravity mechanisms of wavefunction localization and decoherence, *J. Phys. A: Math. Gen.* **40**, 2989 (2007).
22. R. Penrose, *Gravity and state-vector reduction*, in *Quantum Concepts in Space and Time*, eds. R. Penrose and C. J. Isham, (Clarendon Press, Oxford, 1986), p. 129.
23. R. Penrose, On gravity's role in quantum state reduction, *Gen. Rel. Grav.* **28**, 581 (1996).
24. R. Penrose, Quantum computation, entanglement and state reduction, *Phil. Trans. R. Soc. Lond. A* **356**, 1927 (1998).
25. R. Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe* (Jonathan Cape, London, 2004).
26. R. Penrose, *The emperor's new mind: Concerning computers, minds, and the laws of physics* (Oxford University Press, Oxford, 1989).
27. G. C. Ghirardi, R. Grassi and A. Rimini, Continuous spontaneous reduction model involving gravity, *Phys. Rev. A* **42**, 1057 (1990).
28. G. C. Ghirardi, A. Rimini and T. Weber, Unified Dynamics for Microscopic and Macroscopic Systems, *Phys. Rev. D* **34**, 470 (1986).
29. P. Pearle, Ways to describe dynamical state-vector reduction, *Phys. Rev. A* **48**, 913 (1993).
30. A. Bassi and G. C. Ghirardi, Dynamical reduction models, *Phys. Rept.* **379**, 257 (2003).
31. S. Gerlich, S. Eibenberger, M. Tomandl, S. Nimmrichter, K. Hornberger, P. J. Fagan, J. Tuxen, M. Mayor and M. Arndt, Quantum interference of large organic molecules, *Nat. Commun.* **2**, 263 (2011).
32. S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor and J. Tuxen, Matter-wave interference of particles selected from a molecular library with masses exceeding 10 000 amu, *Phys. Chem. Chem. Phys.* **15**, 14696 (2013).
33. I. Pikovski, M. Zych, F. Costa and C. Brukner, Universal decoherence due to gravitational time dilation, *Nat. Phys.* **11**, 668 (2015).
34. M. Zych, F. Costa, I. Pikovski and C. Brukner, Quantum interferometric visibility as a witness of general relativistic proper time, *Nat. Commun.* **2**, 505 (2011).
35. Y. Margalit, Z. Zhou, S. Machluf, D. Rohrlich, Y. Japha and R. Folman, A self-interfering clock as a "which path" witness, *Science* **349**, 1205 (2015).
36. S. Bose, K. Jacobs and P. L. Knight, Scheme to probe the decoherence of a macroscopic object, *Phys. Rev. A* **59**, 3204 (1999).
37. W. Marshall, C. Simon, R. Penrose and D. Bouwmeester, Towards quantum superpositions of a mirror, *Phys. Rev. Lett.* **91**, 130401 (2003).
38. F. Maimone, G. Scelza, A. Naddeo and V. Pelino, Quantum superpositions of a mirror for experimental tests for nonunitary Newtonian gravity, *Phys. Rev. A* **83**, 062124 (2011).
39. O. Romero-Isart, A. C. Pflanzer, F. Blaser, R. Kaltenbaek, N. Kiesel, M. Aspelmeyer and J. I. Cirac, Large Quantum Superpositions and Interference of Massive Nanometer-Sized Objects, *Phys. Rev. Lett.* **107**, 020405 (2011).
40. H. Pino, J. Prat-Camps, K. Sinha, B. P. Venkatesh and O. Romero-Isart, On-chip quantum interference of a superconducting microsphere, *Quantum Sci. Technol.* **3**, 25001 (2018).
41. S. Osnaghi, F. Freitas and O. Freire Jr., The origin of the Everettian heresy, *Stud. Hist. Phil. Mod. Phys.* **40**, 97 (2009).
42. J. Preskill, *Feynman after 40*, Talk given at APS April Meeting, 16 April 2018, `http://theory.caltech.edu/~preskill/talks/APS-April-2018-Feynman-4-3.pdf`.

43. M. Gell-Mann and J. B. Hartle, Classical equations for quantum systems, *Phys. Rev. D* **47**, 3345 (1993).
44. R. B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, Cambridge, 2002).
45. R. Omnes, *Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, 1994)
46. M. Gell-Mann and J. B. Hartle, *Quantum Mechanics in the Light of Quantum Cosmology*, in *Complexity, Entropy, and the Physics of Information*, ed. W. H. Zurek, SFI Studies in the Sciences of Complexity, Vol. VIII, (Addison Wesley, Reading, MA, 1990).
47. P. C. Hohenberg, An Introduction to consistent quantum theory, *Rev. Mod. Phys.* **82**, 2835 (2010).
48. J. B. Hartle, *The Quantum Mechanics of Closed Systems*, in *Directions in General Relativity*, eds. B. L. Hu, M. P. Ryan and C. V. Vishveshwara, Volume 1, (Cambridge University Press, Cambridge, 1993).
49. E. Joos and H. D. Zeh, The emergence of classical properties through interaction with the environment, *Zeit. Phys. B* **59**, 223 (1985).
50. R. P. Feynman, *The Character of Physical Law* (MIT Press, Boston, MA, 1967).

Summary of the parallel session HR3

Shokoufe Faraji

University of Bremen, Center of Applied Space Technology and Microgravity (ZARM), 28359 Germany

This is the summary of the parallel session entitled "Time and Philosophy in Physics", chaired by Shokoufe Faraji in the sixteenth Marcel Grossmann Meeting. This parallel session aimed to discuss open issues related to Time and fundamental laws from different perspectives in a complementary point of view.

Keywords: Time in Physics, Time in Philosophy; passage of Time; fundamental laws

1. Introduction

It is astonishing that we can understand the Universe in the way we study it. However, concurrently celebrating the achievements of science, we should respect its limits and not claim more than it can actually achieve or attain, where by pushing its limits, complex philosophical issues may arise. However, the concern is about when theories are applied beyond their range of applicability. Since the visible Universe may not give us enough information to characterize the laws of physics completely, thus we need a more comprehensive framework that not only incorporates science but can also go beyond the limits.

Therefore, for example the questions on the domain of validity of the fundamental laws of physics and the nature of existence of infinities are sound viable. However, maybe above all, the nature of Time and the order present in nature is the greatest secret in the Universe.[1–5]

2. The passage of Time and top-down causation

This is reported by Prof. Barbara Drossel.

There is a common underlying agreement among many physicists to consider the properties of the fundamental equations of physics are also properties of nature. Therefore, since these equations are deterministic, it is concluded that nature is also deterministic. Besides, the Time evolution is completely determined by the initial state and the fundamental laws. Furthermore, these laws are invariant under changing the direction of Time. However, any physical theory is an approximation and not an exact description of the physical processes of interest. And by no means it is a complete description of everything that is going on in the physical world. Even a deterministic and Time-reversible fundamental theory is in practice supplemented by irreversible and stochastic features. In conclusion, the irreversibility and stochasticity enter when the influence of the environment on a system is taken into

account, which is related to emergence and top-down causation and microsystems are causally open to influences from the macroscopic environment.

3. Temporal asymmetry and causality

This is reported by Prof. Wayne Myrvold.

The temporal asymmetry between past and future permeates virtually every aspect of the world of our experience. As far as we know, it has no counterpart in the laws of fundamental physics. One option is to trace this asymmetry to a fact about the early state of the universe, either taken as a contingent, unexplained fact, or as a consequence of some physical principle. However, there is an alternate route to explaining temporally asymmetric phenomena, according to how work on the process of equilibration is done. This does not involve any hypothesis about the past, rather it is founded on conditions that we can apply at any Time, closely related to the concept of causality.[6] In fact, temporal asymmetry enters our explanations because the very notion of what it means to explain something is temporal asymmetry. To sum up, the conceptions of casual order and temporal order are not independent and in fact they are two sides of the same coin.

Therefore, the absence of temporally asymmetry in the fundamental laws is no obstacle to explaining that asymmetry. What is invoked is a temporal asymmetry in the very notion of explanation.

4. Temporal ordering in Philosophy and Physics

This is reported by Prof. Norman Sieroka.

Time and temporal order are one of the oldest and most eminent concepts in both philosophy and physics. One of the main aims of philosophy is to coordinate the human experience in structural analysis and comparison. For this, one needs to acknowledge the differences in types of Time and Time scales. Then we realize many issues about Time are issues about the ordering of events both within and across different types and scales.[7]

From the philosophical point of view, mainly there are two different types of temporal orderings. The first one, the presentism or possibilism perspective; events being characterized by the present, past or future, and try to focus on mental states and experience; however, this view may have trouble in physics if simultaneity is not absolute. The second view is Eternalism; the ordering relation is only about events being earlier or later and focuses on physical states, and there is no vital distinction between past, present and future. However, it seems the second one to be of importance to physics.

To sum up, the relation between these two orderings and their inner dynamics can be interesting in the context of the interaction between philosophy and physics. Besides, investigation of the notion of an *event* seems fundamentally essential in both areas.

5. Time in Quantum Gravity: A false problem

This is reported by Prof. Carlo Rovelli.

There are several problems of Time in Physics that are related to each other. Indeed, always a lot of confusion comes from mixing problems that are distinct one from another. In particular, an extensive discussion of Time in Quantum Gravity was caused in the 60th by the Wheeler-de Witt equation, which at first sight suggests Time is frozen and there is no evolution. However, dealing with Time in General Relativity is a slightly different than in non-relativistic Physics.[8] The difficulty for developing Time in General Relativity is to struggle with the meaning of the coordinates. In fact, coordinates in non-relativistic field Theory are measurable by rods and clocks; in contrast to General Relativity, there is nothing physical attached to coordinates. In fact, all possible versions of Time in General Relativity is a quantity that is locally dependent on the Gravitational field. This means the entire formulation of dynamics as evolution in Time does not fit within the scope of General Relativity. In fact, instead of conceptualizing the evolution of physical variables evolving with respect to a variable, we are forced to consider the evolution of a variable with respect to one another. So by this perception, the relativistic dynamics is not frozen in Quantum Gravity and simply described in a different language. Thus, we can predict the relation among a number of variables, and there is no mystery about *emergence of Time* in Quantum Gravity. In fact, many tentative quantum gravity theories, from the Wheeler-de Witt equation to Loop Quantum Gravity, do not specify a Time variable, and yet they are predictive. Of course, this view has substantial consequences, for example, there are some indications that the Time might be discrete.

6. The issue of Time in Fundamental Physics

This is reported by Prof. Abhay Ashtekar.

The Notion of Time in fundamental physics has undergone radical revisions over centuries dramatically.[9] However, new observational windows continually open up in the process. In particular, issues at the forefront of cosmology and quantum gravity pose new conceptual and technical challenges. The same is true in particle Physics, GPS systems or Quantum Information Theory.

A pleasing feature is coming together from physical and operational ideas, fundamental concepts from General Relativity, the associated Hamiltonian constraint, and a post-Pauli understanding of how one should think of Time in Quantum Mechanics. In fact, in Quantum Mechanics, the focus is on the extended system by considering the clock. In Quantum Gravity, there is no such thing as Time in Quantum mechanics, and the extended system kinematics is already contained the dynamical variables of General Relativity.

However, it does not mean *everything is understood and well-controlled.* Many refinements, extensions and technical issues are still open; however, there is no conceptual roadblock to this understanding.

7. Time in General Relativity and Quantum Mechanics

This is reported by Prof. Claus Laemmerzahl.

The nature of Time is characteristic of the underlying theory. Within the formalism of General Relativity, it is possible to define or characterize a distinguished clock and Time operationally, and this is the so-called standard clock providing proper Time. In addition, within Quantum Mechanics, it is possible to define a clock and the corresponding Time. This clock is an atomic clock, and it provides Time in the unit of the second. It can be proven with high precision that both clocks are compatible though they are based on entirely different notions.[10] Based on such standard clocks, we define Special Relativity effects such as the gravitational redshift, gravitational Time delay and gravitomagnetic clock effect. This compatibility breaks down in strong gravitational fields and also in generalized theories of gravity.

It would be of some interest to study the relativistic two-body problem and another Time scale from chirp. In addition, studying a bound system in an external gravitational field like the Earth-Moon system in the field of the Sun, seems necessary to our understanding.

8. Time's passage

This is reported by Janathan Dickau.

The way we experience Time is in the accumulation of experiences and events that happen in the moment. On the one hand, philosophers have tried to explain both the nature of Time and its origin. On the other hand, the meaning of Time in different areas of physics like in Classical and Quantum Mechanics, and Relativity are different. So we hope for example Quantum Gravity theories will help resolve this issue. Recent advances in Mathematics also hold promise for a unified basis explaining both the thermodynamic and quantum-mechanical Time arrows in a consistent way. It is important to continue our exploration of quantum gravity theories and try to understand the cosmological context for these theories in a broader way. So physicists should see it as a responsibility to build bridges between the philosophically disconnected islands of thought, by applying a Philosophy of Physics that supports our common endeavor to learn the secrets of the cosmos.

9. A glimpse of Feynman's contributions to the debate on the Foundations of Quantum Mechanics

This is reported by Dr. Adele Naddeo.

One of the most debated questions in the famous 1957 Chapel Hill conference on "The Role of Gravitation in Physics" was whether or not the gravitational field has to be quantized.[11] Feynman proposed several gedanken experiments in order to argue in favor of the necessity of gravitational quantization, and hinting to decoherence as a viable solution to the problem of wave function collapse.[12] Feynman believed that nature cannot be half classical and half quantum; therefore, we had the unified field theories. After quantum theory one tries to quantize gravity!

Another issue was related to the limitations that quantum quantum theory may impose on the measurements of space-time distances and curvature. However, since the Planck mass does not set a lower limit to the mass of a particle whose gravitational field can be measured, so the gravitational field of any mass can in principle be measured.

Feynman also discussed the role of the observer in quantum mechanics. The Schrödinger's cat paradox allowed him to illustrate the difference between the results of a measurement carried out by an external as well as an internal observer.[13] While the external observer describes his results by an amplitude, according to the internal observer the results of the same measurement are given by a probability. In other words, according to Feynman, the Universe is constantly spitting into an infinite number of branches, as a result of the interactions between its components. As a consequence, an inside observer knows which branch the world has taken, so that he can follow the track of his past. However, in 1981 Feynman says of the many-worlds picture: *It's possible, but I am not very happy with it.*

10. Conclusion

Physics enables everything and underlies everything, but it does not mean it determines everything.[5] Perhaps the world is much more complicated than can be explained only by science. Thus, we should trust our most immediate experiences, new conceptual and technical challenges. In fact, they are the basis and guides for being capable of thinking and acting. Based on these, we could critically examine and take care of metaphysical claims and our ultimate questions like: does the world covered by science describe the whole reality? What about underlying assumptions and axioms in cosmology? What about all the philosophical issues related to "the interpretation of quantum mechanics," "measurement Problem," and the ontological character of quantum states? Why does the Universe have such a unique structure that life can exist?

References

1. R. Penrose, *The road to reality: A complete guide to the laws of the universe* (Vintage, London, 2005).
2. G. F. Ellis, Physics in the real universe: Time and spacetime, *General relativity and gravitation* **38**, 1797 (2006).
3. G. Auletta, G. F. Ellis and L. Jaeger, Top-down causation by information control: From a philosophical problem to a scientific research programme, *Journal of The Royal Society Interface* **5**, 1159 (2008).
4. J. Birriel, Fashion, faith, and fantasy in the new physics of the universe fashion, faith, and fantasy in the new physics of the universe by roger penrose, published by princeton university press (sept. 27, 2016), 520 pages. isbn-10:0691119791, *The Physics Teacher* **56**, 482 (10 2018).
5. G. Ellis, *How can physics underlie the mind?: Top-down causation in the human context* (Springer Heidelberg, 2016).

6. W. Myrvold and J. Christian, *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* 01 2008.
7. N. Sieroka, Philosophie der zeit, *Nova acta Leopoldina: Abhandlungen der Kaiserlich Leopoldinisch-Carolinisch Deutschen Akademie der Naturforscher* **425**, 23 (12 2020).
8. C. Rovelli, Time in quantum gravity: An hypothesis, *Phys. Rev. D* **43**, 442 (Jan 1991).
9. A. Ashtekar, Time in Fundamental Physics, *arXiv e-prints*, p. arXiv:1312.6322 (December 2013).
10. V. Perlick, Characterization of standard clocks by means of light rays and freely falling particles, *Gen. Relativ. Gravitation* (1987).
11. C. M. Dewitt and D. Rickles, The role of gravitation in physics : report from the 1957 chapel hill conference (2011).
12. M. Di Mauro, S. Esposito and A. Naddeo, A road map for Feynman's adventures in the land of gravitation, *European Physical Journal H* **46**, p. 22 (December 2021).
13. H. D. Zeh, Feynman's interpretation of quantum theory, *European Physical Journal H* **36**, 63 (July 2011).

Massive compact stars in the two-families scenario

P. Char*, A. Drago† and G. Pagliara†

*Space Sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, Bât. B5a, 4000 Liège, Belgium

†Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy and INFN Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy

We review the two-families scenario of compact stars. Under the hypothesis that strange quark matter is the ground state of dense matter one can infer the existence of two separated branches of stars: hadronic stars and quark stars. While the first branch is populated by very compact and light stars, the second branch can be populated by very massive stellar objects. A possible confirm to this scenario derives from the gravitational waves detections GW170817 and GW190814, the first one suggesting the existence of stars with radii as small as 11km for $M \sim 1.4 M_\odot$ and the second one suggesting the existence of a compact star with a mass above $2.5 M_\odot$.

Keywords: Compact stars, Binary mergers

1. Introduction

The opening of the Gravitational Waves (GWs) astronomy has already led to some important and new measurements of the structure of neutron stars (NSs) and constraints on the population and mass distribution of neutron stars binaries[1] and neutron star - black hole (BH) binaries.[2] There are two merger events on which we want to focus here: GW170817[3] and GW190814.[4] The first one has being unambiguously associated to a NS-NS merger and it is the first example of multimessenger astronomy since, in this case, also the electromagnetic counterparts GRB170817A and AT2017gfo have been detected. With a total mass of $2.74 M_\odot$, this event is associated to the merger of two NSs with a mass around $1.37 M_\odot$ and thus far below the two solar mass limit observed through pulsar timing measurements, see e.g.[5]. Interestingly by combining the constraints on the tidal deformability obtained from the GW signal and the properties of the electromagnetic signals, a precious constraint on the structure of NSs has been found: the radius of the $1.4 M_\odot$ mass NS $R_{1.4} = 11.0^{+0.9}_{-0.6}$km at 90% credible interval.[6] Such a small radius suggests a quite soft equation of state for dense nucleonic matter and one can predict a maximum mass $m_{\rm max}$ value of $\sim 2.3 M_\odot$.

On the other hand, GW190814 has been produced by the merger of a $23 M_\odot$ BH and a compact object with mass in the range $2.5 - 2.67 M_\odot$. The nature of the low mass companion is unclear, it could be the lightest BH or the heaviest NS ever discovered. By assuming that it is indeed a NS, one would get a very strong

constraint on the equation of state: only very stiff equations of state could lead to such massive compact stars, see e.g.[7]. Thus, a tension emerges between the constraints obtained from GW170817 and GW190814.

The two-families scenario[8–14] proposes a solution to this tension based on the hypothesis of coexistence of hadronic stars (HS) and quark stars (QS) the first being very compact and the second being very massive. In the following we will review this scenario outlining the differences with respect to the standard one family scenario and other astrophysical scenarios with two separated branches of compact stars.

2. Modelling the equation of state and structure of compact stars

The equation of state of dense matter is rather uncertain and indeed NSs are considered in nuclear and particle physics as natural laboratories for matter at extremely high densities, order of ten times the nuclear saturation density n_s. There are however a few theoretical arguments that suggest, qualitatively, what could happen if matter is compressed (in NSs by gravity) above n_s. First, the increase of the baryon chemical potential must lead, at some unknown density, to the appearance of new baryons, namely hyperons or delta resonances. The appearance of new degrees of freedom is necessary accompanied by a softening of the equation of state. The value of the threshold density for the emergence of these degrees of freedom and how much the equation of state is softened due to their appearance depend on the details of the theoretical modelling of the equation of state. With the present knowledge one cannot exclude that NSs in which also hyperons and delta resonance (called hadronic stars, HSs) form could be as massive as $2M_\odot$, see for instance[15] where also kaon condensates are considered in the computation of the equation of state. If the maximum mass of NSs $M_{max} \sim 2.6M_\odot$, it is difficult to explain such a massive object in terms of a HS (at the moment there are no calculations showing that HS could be so massive). Thus, the threshold for the formation of hyperons and delta would be above the central density n_c reached in NSs. For $m_{\rm max} = 2.6M_\odot$ it has been found in[16] that $n_c \leq 4.5n_s$, one would thus conclude that no new degrees of freedom are produced up to $4.5n_s$. Such a possibility could be realised if an extremely strong hyperon-hyperon repulsion is assumed as e.g. in the calculations of.[17] Here we will adopt the most common result that hyperons and/or delta resonance appear in dense matter already at about twice n_s.[18] The resulting softening, if no ad hoc repulsion term between hyperons is added (see e.g.[19]), is such that M_{max} for HSs could be as low as $1.6M_\odot$. Moreover radii as small as 10km are obtained for the most massive configurations. Thus very compact objects are interpreted as being HSs, we will discuss one example of mass-radius relation within this scenario later on.

How to explain then the existence of massive NSs, with masses possibly as high as $2.6M_\odot$? Since in the two-families scenario massive stars are QSs, the equation of state of quark matter must be rather stiff. Let us consider the simple constant speed of sound (CSS) equation of state for which the pressure p and the energy

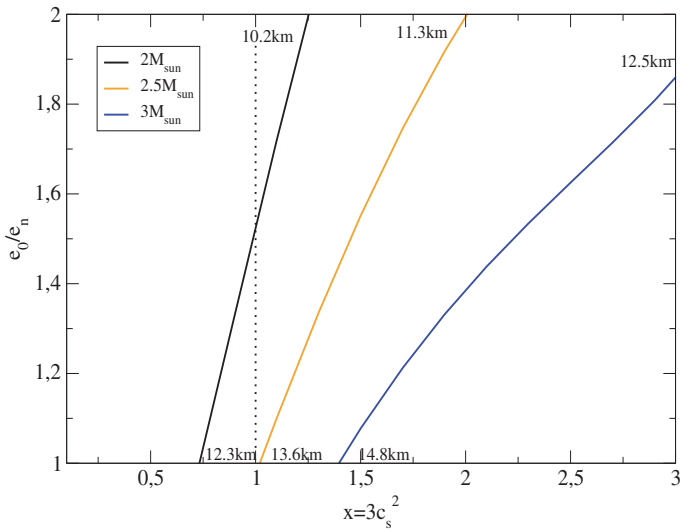


Fig. 1. Lines of constant $m_{\max}$ in the $e_0 - x$ plane. The vertical dashed line indicates the conformal limit, $c_s^2 = 1/3$.

density e are related by the relation: $p = c_s^2(e - e_0)$ where e_0 is the energy density at zero pressure and c_s is the (constant) speed of sound. The stiffness is regulated by e_0 and c_s: the larger the value of c_s and/or the smaller the value of e_0, the stiffer the equation of state. Indeed, by solving the TOV equations by adopting such an equation of state one can show, through numerical calculations, that $m_{\max}$ as function of e_0 and c_s can be described by the following parametrization: $m_{max}(x, e_0) = (e_s/e_0)^{1/2}(0.11 + 3.27x - 1.04x^2 + 0.13x^3)$ where $x = c_s^2/(1/3)$ and e_s is the energy density of nuclear matter $e_s \sim 150$ MeV/fm^3, see.[20] For c_s the upper bound is 1, namely the causal limit whereas the lower bound for e_0 can be taken at $e_0 = e_s$, namely by assuming that strange quark matter has an energy density at $p = 0$ not smaller than the one of nuclear matter.

In Fig. 1 we display the constant maximum mass curves in the x - e_0 plane. As one can notice, if one sets $c_s = 1/3$, namely to its asymptotic value as predicted by perturbative QCD (the so called conformal limit), one can reach values for $m_{\max}$ as high as $\sim 2.5 M_\odot$ if e_0 is set to e_s. Thus for $c_s = 1/3$ one could touch the lower limit for the low mass companion of GW190814. Small deviations from the conformal limit would allow to easily reach its upper limit, namely $2.67 M_\odot$. It is interesting to also discuss the radii of massive compact stars. For a fixed value of maximum mass, the radius of $m_{\max}$ decreases for increasing values of c_s (and e_0). For instance for $m_{\max} = 2M_\odot$, $r_{\max}$ ranges from 12.3km (for $x = 0.7$) to 10.2 (for $x = 1.2$) km. Interestingly, the analysis of the NICER data on PSR J0740+6620, provides $R = 12.35 \pm 0.75$ km for $M = 2.08 \pm 0.07 M_\odot$.[21] In our scenario, and under the hypothesis that $m_{\max}$ is actually $\sim 2M_\odot$, one would infer that c_s needs to be close to the conformal limit in order to explain the radius of PSR J0740+6620.

This is at odd with the so-called twin-stars scenario where a strong first order phase transition is assumed to occur from nucleonic matter to quark matter. The necessary softening associated with the phase transition must be compensated by large values of c_s ($c_s \sim 1$) in order to explain the existence of massive stars, see e.g.[22].

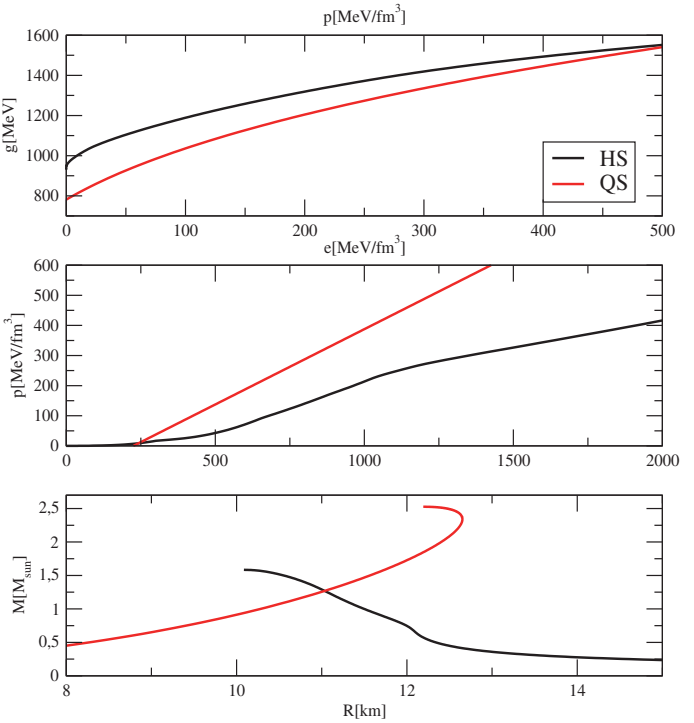


Fig. 2. Upper panel: Gibbs potential g as a function of the pressure for hadronic and quark matter. The phase with the lowest g is thermodynamically preferred. Middle panel: relation between pressure and energy density for hadronic and quark matter. Lower panel: mass-radius relations for HSs and QSs.

An important point when discussing the possibility of separated branches of compact stars is the thermodynamical consistency of the equations of state. In the two-families scenario, the branch of QSs is populated via the conversion of a HS. The conversion itself, to start, needs a specific condition on the thermodynamical variable of the two phases, namely at the same pressure (and same temperature) the new phase must have a smaller Gibbs potential according to the second principle of thermodynamics. To compute the Gibbs potential $g = (e + p)/n$ we need to first compute the baryon density in the CSS model as a function of the pressure. By using the relation $p = n^2 \frac{\partial(e/n)}{\partial n}$ one can show that:

$$n = n_0 \left(\frac{p(1 + c_s^2) + c_s^2 e_0}{c_s^2 e_0} \right)^{\frac{1}{1+c_s^2}} \tag{1}$$

where n_0 is the baryon density at zero pressure. In Fig. 2, upper panel we show the comparison between the Gibbs potential of a hadronic equation of state, SFHo-HD taken from[13] and a CCS equation of state of quark matter with $c_s^2 = 0.5$, $e_0 = 1.5e_s$ and $n_0 = 1.8n_s$. The Gibbs potential of the quark phase is below the one of the hadronic phase at least up to very large values of the pressures (~ 500 MeV/fm^3) that are not reached within the HSs branch (for which the maximum central pressure is ~ 330 MeV/fm^3 for this specific example). This means that quark matter can be nucleated from hadronic matter, the only necessary condition being the amount of strangeness in the hadronic phase. In,[13] we have estimated that the critical mass of HSs allowing for nucleation varies in the range $1.5 - 1.6M_\odot$. In the middle panel of Fig. 2, we display the relation between pressure and energy density for the two equations of state. One can notice that the quark matter equation of state is stiffer in the $p - e$ plane and thus leads to larger maximum masses. Although the quark equation of state is stiffer than the hadronic one, is anyway preferred (once the nucleation and the conversion process start) due to its smaller Gibbs potential. Finally in the lower panel, we display the mass radius curve of HSs and QSs. In this specific example, the quark equation of state leads to a maximum mass of $2.5M_\odot$ and this could potentially explain the nature of the low mass companion of GW190814 as being a QS. At the same time, the HSs branch is populated by very compact stars, as small as 10km. The event GW170817 is interpreted in our scenario as a merger of a HS and a QS. The computed average tidal deformability, $\tilde{\Lambda}$, is in agreement with the constraints obtained from the GW signal, see.[13]

3. Conclusions

We have reviewed the two-families scenario proposed in.[8] In this scenario two brances of compact stars exist, HSs which could be very compact and QSs which could be very massive. This scenario can therefore explain why certain sets of data (both on GW and EM signals) suggest small radii ($R < 11$km) and other sets of data suggest large masses, as large as $2.5M_\odot$. The coexistence of these branches is possible because the quark matter equation of state can be rather stiff (in the $p - e$ plane) while being thermodynamically preferred with respect to the hadronic matter equation of state (in the $g - p$ plane). Interestingly, the two branches separate for values of mass of $\sim 1.6M_\odot$ at variance with other scenarios with separated branches, as e.g. the twin stars scenario, in which the new branch opens up at $\sim 2M_\odot$,[22]

In the two-families scenario the specific process of conversion of a HS into a QS has been extensively studied, see[23–25]: basically it is a combustion process which could be very fast (with time scales of the order of ms) due to hydrodynamical instabilities. The new stellar object has a larger radius with respect to the progenitor HS as studied in detail in Ref.[26]. In the twin-stars scenario instead, the formation of a hybrid star is triggered by a mechanical instability with respect to the gravitational collapse. It has to be shown that from a $2M_\odot$ star (where the twin branch starts)

one can form a much lighter ($M \sim 1.6 M_\odot$) and compact ($R \sim 10$ km) star (see[22]) preventing the collapse to a BH.

Finally, another difference between the two-families scenario and the twin stars scenario concerns the value of the speed of sound of quark matter: it can stay below or quite close to the conformal limit ($c_s^2 = 1/3$) in the former while it must be set to the causal limit ($c_s^2 = 1$) in the latter. In this case one should find out a physical mechanism that allows to match the pQCD results at asymptotic densities according to which c_s^2 must tend again to 1/3.

References

1. B. P. Abbott *et al.*, GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_\odot$, *Astrophys. J. Lett.* **892**, p. L3 (2020).
2. R. Abbott, Observation of Gravitational Waves from Two Neutron Star-Black Hole Coalescences, *ApJL*, **915**, p. L5 (2021).
3. B. Abbott *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (2017).
4. R. Abbott *et al.*, GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, *Astrophys. J. Lett.* **896**, p. L44 (2020).
5. P. Demorest, T. Pennucci, S. Ransom, M. Roberts and J. Hessels, Shapiro Delay Measurement of A Two Solar Mass Neutron Star, *Nature* **467**, 1081 (2010).
6. C. D. Capano, I. Tews, S. M. Brown, B. Margalit, S. De, S. Kumar, D. A. Brown, B. Krishnan and S. Reddy, GW170817: Stringent constraints on neutron-star radii from multimessenger observations and nuclear theory (2019).
7. F. Fattoyev, C. Horowitz, J. Piekarewicz and B. Reed, GW190814: Impact of a 2.6 solar mass neutron star on nucleonic equations of state (7 2020).
8. A. Drago, A. Lavagno and G. Pagliara, Can very compact and very massive neutron stars both exist?, *Phys. Rev.* **D89**, p. 043014 (2014).
9. A. Drago, A. Lavagno, G. Pagliara and D. Pigato, The scenario of two families of compact stars, *Eur. Phys. J.* **A52**, p. 40 (2016).
10. A. Drago and G. Pagliara, Merger of two neutron stars: predictions from the two-families scenario, *Astrophys. J. Lett.* **852**, p. L32 (2018).
11. G. Wiktorowicz, A. Drago, G. Pagliara and S. Popov, Strange quark stars in binaries: formation rates, mergers and explosive phenomena, *Astrophys. J.* **846**, p. 163 (2017).
12. G. F. Burgio, A. Drago, G. Pagliara, H. J. Schulze and J. B. Wei, Are Small Radii of Compact Stars Ruled out by GW170817/AT2017gfo?, *Astrophys. J.* **860**, p. 139 (2018).
13. R. De Pietri, A. Drago, A. Feo, G. Pagliara, M. Pasquali, S. Traversi and G. Wiktorowicz, Merger of compact stars in the two-families scenario, *Astrophys.J.* **881**, p. 122 (2019).
14. I. Bombaci, A. Drago, D. Logoteta, G. Pagliara and I. Vidaña, Was GW190814 a Black Hole–Strange Quark Star System?, *Phys. Rev. Lett.* **126**, p. 162702 (2021).
15. V. B. Thapa, M. Sinha, J. J. Li and A. Sedrakian, Massive Δ-resonance admixed hypernuclear stars with antikaon condensations, *Phys. Rev. D* **103**, p. 063004 (2021).
16. C. Drischler, S. Han, J. M. Lattimer, M. Prakash, S. Reddy and T. Zhao, Limiting masses and radii of neutron stars and their implications, *Phys. Rev. C* **103**, p. 045808 (2021).

17. D. Lonardoni, A. Lovato, S. Gandolfi and F. Pederiva, Hyperon Puzzle: Hints from Quantum Monte Carlo Calculations, *Phys. Rev. Lett.* **114**, p. 092301 (2015).
18. A. Drago, A. Lavagno, G. Pagliara and D. Pigato, Early appearance of delta isobars in neutron stars, *Phys. Rev.* **C90**, p. 065809 (2014).
19. S. Weissenborn, D. Chatterjee and J. Schaffner-Bielich, Hyperons and massive neutron stars: vector repulsion and SU(3) symmetry, *Phys. Rev.* **C85**, p. 065802 (2012), [Erratum: Phys. Rev. C90, no. 1, 019904 (2014)].
20. S. Traversi, P. Char, G. Pagliara and A. Drago, Speed of sound in dense matter and two families of compact stars (2 2021).
21. M. C. Miller *et al.*, The Radius of PSR J0740+6620 from NICER and XMM-Newton Data, *Astrophys. J. Lett.* **918**, p. L28 (2021).
22. J.-E. Christian and J. Schaffner-Bielich, Confirming the existence of twin stars in a NICER way (9 2021).
23. A. Drago, A. Lavagno and I. Parenti, Burning of an hadronic star into a quark or a hybrid star, *Astrophys. J.* **659**, 1519 (2007).
24. G. Pagliara, M. Herzog and F. K. Röpke, Combustion of a neutron star into a strange quark star: The neutrino signal, *Phys. Rev. D* **87**, p. 103007 (2013).
25. A. Drago and G. Pagliara, Combustion of a hadronic star into a quark star: The turbulent and the diffusive regimes, *Phys. Rev. C* **92**, p. 045801 (2015).
26. A. Drago and G. Pagliara, Why can hadronic stars convert into strange quark stars with larger radii, *Phys. Rev. D* **102**, p. 063003 (2020).

Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond

J. Soldateschi*, N. Bucciantini and L. Del Zanna

Dipartimento di Fisica e Astronomia, Università degli Studi di Firenze, and INFN - Sezione di Firenze, Via G. Sansone 1,
I-50019 Sesto F. no (Firenze), Italy
**E-mail: jacopo.soldateschi@unifi.it*

INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5,
I-50125 Firenze, Italy

Neutron stars harbour extremely powerful magnetic fields, leading to their shape being deformed. Their magnetic deformation depends both on the geometry - and strength - of their internal magnetic field and on their composition, encoded by the equation of state. However, both the details of the internal magnetic structure and the equation of state of the innermost part of neutron stars are mostly unkown. We performed a study of numerical models of magnetised, static, axisymmetric neutron stars in general relativity and in one of its most promising extensions, scalar-tensor theories. We did so by using several realistic equations of state currently allowed by observational and nuclear physics constraints, considering also those for strange quark stars. We show that it is possible to find simple relations among the magnetic deformation of a neutron star, its Komar mass, and its circumferential radius in the case of purely poloidal and purely toroidal magnetic configurations satisfying the equilibrium criterion in the Bernoulli formalism. These relations are quasi-universal, in the sense that they mostly do not depend on the equation of state. Our results, being formulated in terms of potentially observable quantities, could help to understand the magnetic properties of neutron stars interiors and the detectability of continuous gravitational waves by isolated neutron stars, independently of their equation of state. In the case of scalar-tensor theories, these relations depend also on the scalar charge of the neutron stars, thus potentially providing a new way to set constraints on the theory of gravitation.

Keywords: Neutron stars; magnetic deformation; scalar-tensor theories.

1. Introduction

The most compact material objects in the known Universe are neutron stars (NSs), which are also known to harbour extremely powerful magnetic fields, especially in the sub-class known as magnetars[1–4]: the surface magnetic field of NSs has been found to be in the range 10^{8-12}G for radio and γ-ray pulsars,[5–7] while estimates in magnetars reach 10^{15}G.[8,9] While the surface and magnetospheric magnetic fields of NSs can be probed through many different methods,[10,11] their internal magnetic field remains mostly unknown, both in its strength and geometry. Predictions expect that values up to 10^{16}G and 10^{17-18}G could be reached inside magnetars and newly-born proto-NSs, respectively.[12–14] On the other hand, while it is known that neither purely poloidal nor purely toroidal magnetic configurations are stable[15,16]

and mixed configurations are more favoured,[17–19] the extact magnetic field geometry in the interior of NSs is less clear. Fortunately, these magnetic fields present potentially observable effects to the structure of NSs: they can modify the torsional oscillations of NSs,[20,21] their cooling properties[22,23] and their deformation.[24,25]

In this scenario, another unknown regarding the internal structure of NSs enters the game: the equation of state (EoS) of NSs remains mostly unconstrained. Recent observational results rejected the validity of many EoS: the observation of very massive NSs [e.g. the most massive NS known to date, with a mass potentially reaching $\sim 2.28 M_\odot$[26]]; the limits on the stiffness of NSs[27,28] obtained by the first observation of gravitational waves (GWs) emitted by a binary NS merger[29]; the results of the NICER telescope[30,31] on the possible NSs radii. However, this uncertainity is further enhanced by the effect that the strong magnetic field buried inside NSs exherts on their composition, for example determining the presence of exotic particles or the existence of a superconducting phase.[32–36] For these reasons, undertanding and constraining the interplay between the magnetic field of NSs and their EoS is fundamental in order to improve our knowledge of these objects.

Other than observations in the electromagnetic domain, GWs can help to probe the inner structure of NSs. Of particular importance are continuous GWs (CGWs), which are emitted, for example, by a time-varying deformation. Such scenario can happen e.g. in the case of magnetically-deformed NSs whose magnetic axis is not aligned with their rotation axis.[25,37,38] Since magnetic fields found in regular pulsars are too low to cause a significant deviation from spherical symmetry,[24] newly-born proto-NSs and millisecond magnetars are the most promising sources of CGWs,[39] along with millisecond pulsars which possess a superconducting core.[38]

While the problem of the inner composition and magnetic field of NSs is an interesting problem on its own, it is deeply intertwined with the long-standing quest for the definitive theory of gravity. Even if general relativity (GR) remains the best theory to model the gravitational interaction, it has long been known that our understanding of gravity through GR presents some issues on the galactic and cosmological scales.[40] While one possibility is to introduce a dark sector,[41,42] another path is to introduce modification to GR, leading to alternative theories of gravity.[43] Among the possible alternatives to GR, scalar-tensor theories (STTs) - in which gravity is mediated also by a scalar field non-minimally coupled to the spacetime metric - have gathered a strong interest because of their simplicity, because they derive from some suggested theories of quantum gravity,[44] in general they satisfy the weak equivalence principle, and they are free of the issues affecting other alternatives to GR.[45–47] Of great importance in the development of STTs was the discovery of a non-perturbative strong field effect, exhibited in some of these theories, named 'spontaneous scalarisation',[48] which leads to strong deviations from GR in the vicinity of compact material objects (i.e. NSs), while it remains unconstrained in the weak-field limit. The presence of a scalar field leads to an enrichment of the physics of NSs, for example by causing the emission of additional modes of

GWs[49,50] or by modifing the mass-radius relation of NSs, the binary NS merger dynamics,[51] the frequency of NS normal modes,[52] their tidal and rotational deformation,[53,54] their magnetic deformation[55–57] and the light propagation properties in their vicinity.[58] While massless STTs have been mostly ruled out by observations,[59–61] STTs with a massive scalar field are still viable.[62–64] As anticipated, part of the phenomenology of STTs is degenerate with the EoS of NSs, for example concerning the mass-radius relation or their deformabilities. For such reason, disentangling their roles is important in order to have a more informative interpretation of observations.

In the following we assume a signature $\{-,+,+,+\}$ for the spacetime metric and use Greek letters μ, ν, λ, ... (running from 0 to 3) for 4D spacetime tensor components, while Latin letters i, j, k, ... (running from 1 to 3) are employed for 3D spatial tensor components. Moreover, we use the dimensionless units where $c = G = \mathrm{M}_\odot = 1$, and we absorb the $\sqrt{4\pi}$ factors in the definition of the electromagnetic quantities. All quantities calculated in the Einstein frame (E-frame) are denoted with a bar $(\bar{\cdot})$ while all quantities calculated in the Jordan frame (J-frame) are denoted with a tilde $(\tilde{\cdot})$.

2. Neutron stars in general relativity

The spacetime metric in the case of static, axisymmetric configurations can be well approximated[65,66] using the conformally flat condition (CFC).[67,68] Then, for spherical-like coordinates $x^\mu = [t, r, \theta, \phi]$, the line element is

$$g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = -\alpha^2\mathrm{d}t^2 + \psi^4\left[\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right] , \tag{1}$$

where $g_{\mu\nu}$ is the spacetime metric, with determinant g, $\alpha(r,\theta)$ is the lapse function and $\psi(r,\theta)$ is the conformal factor. The energy-momentum tensor for a magnetised ideal fluid is[69,70]

$$T^{\mu\nu} = (\rho + \varepsilon + p)\, u^\mu u^\nu + p g^{\mu\nu} + F^\mu_\lambda F^{\nu\lambda} - \frac{1}{4}F^{\lambda\kappa}F_{\lambda\kappa}g^{\mu\nu} , \tag{2}$$

where ρ is the rest mass density, ε is the internal energy density, p is the pressure, u^μ is the four-velocity and $F^{\mu\nu}$ is the Faraday tensor. The 3+1 formalism[71,72] can be used to recast the equations in a more computationally-friendly way; under these assumptions, the Einstein equations for the metric become two Poisson-like equations, one for ψ and one for $\alpha\psi$.[19]

As for the magnetohydrodynamics (MHD) quantities, using the pseudo-enthalpy h, related to the pressure and density by $\mathrm{d}\ln h = \mathrm{d}p/(\rho + \varepsilon + p)$, Euler's equation becomes the 'generalised Bernoulli integral',[19]

$$\ln\left(\frac{h}{h_\mathrm{c}}\right) + \ln\left(\frac{\alpha}{\alpha_\mathrm{c}}\right) - \mathcal{M} = 0 , \tag{3}$$

where $\mathcal{M}$ is the magnetisation function and h_c and α_c are the values of h and α at the centre of the star, respectively (having assumed $\mathcal{M}_\mathrm{c} = 0$). The magnetisation

function in the case of a purely poloidal magnetic field is assumed to be $\mathcal{M} = k_{\rm pol} A_\phi$, where $k_{\rm pol}$ is the poloidal magnetisation constant and A_ϕ is the ϕ-component of the vector potential, computed by solving the relativistic Grad-Shafranov equation. For a purely toroidal magnetic field, we instead use $\mathcal{M} = -k_{\rm tor}^2(\rho+\varepsilon+p)\mathcal{R}^2$, where $k_{\rm tor}$ is the toroidal magnetisation constant and $\mathcal{R}^2 = \alpha^2\psi^4 r^2 \sin^2\theta$. In the first case, the poloidal components of the magnetic field are found through the Grad-Shafranov equation[66,73] for A_ϕ; in the second case, the toroidal component is proportional to $(\rho+\varepsilon+p)\mathcal{R}^2/\alpha$. Finally, the EoS closes the system of equations. We describe the EoS we used in Sect. 5. For a more detailed description of the equilibrium formalism the reader is referred to Refs. 55–57.

3. Scalar-tensor theories in a nutshell

The action of massless STTs in the J-frame, according to the 'Bergmann-Wagoner formulation',[74–76] is

$$\tilde{S} = \frac{1}{16\pi}\int \mathrm{d}^4x\sqrt{-\tilde{g}}\left[\varphi\tilde{R} - \frac{\omega(\varphi)}{\varphi}\tilde{\nabla}_\mu\varphi\tilde{\nabla}^\mu\varphi\right] + \tilde{S}_{\rm p}\left[\tilde{\Psi},\tilde{g}_{\mu\nu}\right] , \tag{4}$$

where $\tilde{g}$ is the determinant of the spacetime metric $\tilde{g}_{\mu\nu}$, $\tilde{\nabla}_\mu$ its associated covariant derivative, $\tilde{R}$ its Ricci scalar, $\omega(\varphi)$ is the coupling function of the scalar field φ, and $\tilde{S}_{\rm p}$ is the action of the physical fields $\tilde{\Psi}$. In the E-frame, the action is obtained by making the conformal transformation $\bar{g}_{\mu\nu} = \mathcal{A}^{-2}(\chi)\tilde{g}_{\mu\nu}$, where $\mathcal{A}^{-2}(\chi) = \varphi(\chi)$ and χ is a redefinition of the scalar field in the E-frame, related to φ by $\mathrm{d}\chi/\mathrm{d}\ln\varphi = \{[\omega(\varphi)+3]/4\}^{1/2}$. In the E-frame, the scalar field is minimally coupled to the metric: Einstein's field equations retain their usual form in the E-frame, but the energy-momentum tensor is now the sum of the one describing the fluid and electromagnetic fields and of the scalar field one. On the other hand, the scalar field is minimally coupled to the physical fields in the J-frame: the MHD equations in the J-frame have the same expression as in GR. In addition to the metric and MHD equations, in STTs we have an additional equation to solve for the scalar field. In the E-frame it is

$$\Delta\chi = -4\pi\bar{\psi}^4\alpha_{\rm s}(\chi)\mathcal{A}^4\tilde{T} - \partial\ln\left(\bar{\alpha}\bar{\psi}^2\right)\partial\chi , \tag{5}$$

where $\Delta = f^{ij}\hat{\nabla}_i\hat{\nabla}_j$ and $\hat{\nabla}_i$ are, respectively, the 3D Laplacian and nabla operator of the flat space metric f_{ij}, $\partial f\partial g = \partial_r f\partial_r g + (\partial_\theta f\partial_\theta g)/r^2$, $\alpha_{\rm s}(\chi) = \mathrm{d}\ln\mathcal{A}/\mathrm{d}\chi$ and $\tilde{T} = 3\tilde{p} - \tilde{\varepsilon} - \tilde{\rho}$ is the trace of the J-frame energy momentum tensor of the fluid and electromagnetic fields. We used an exponential coupling function $\mathcal{A}(\chi) = \exp\{\alpha_0\chi + \beta_0\chi^2/2\}$.[48] The α_0 parameter regulates the effects of the scalar field in the weak-field limit, while the β_0 parameter controls spontaneous scalarisation. We chose $\alpha_0 = -2\times10^{-4}$ and $\beta_0 \in [-6,-4.5]$. The most recent observational constraints to date require that, for massless scalar fields, $|\alpha_0| \lesssim 1.3\times10^{-3}$ and $|\beta_0| \gtrsim 4.3$[61]; in the case of massive scalar fields, lower values of β_0 are allowed,[77] as long as the screening radius is smaller than the separation of the binary NSs whose

observation confirmed the absence of dipolar GWs.[78,79] We emphasise that results found in a massless STT regarding the internal structure of a NS are also valid for STTs containing a screening effect as long as the screening radius is larger than the NS radius; as such, our results regarding the magnetic deformation of NSs are valid also in the case of a massive scalar field, with a mass such that its screening radius is larger than the NS radius but lower than the binary separation. For a more detailed description of the formulation of STTs within the 3+1 formalism the reader is referred to Refs. 55–57.

4. The magnetic deformation of polytropic neutron stars

We now focus on the case of static NSs in the weak magnetic field regime, where the magnetic deformation of the star is well approximated by a perturbative approach; this was shown to happen for $B_{\rm max} \lesssim 10^{17}$G,[80,81] where $B_{\rm max} = \max[\sqrt{B_i B^i}]$ and B_i are the components of the magnetic field. We focus only on the range of masses corresponding to stable configurations. All the results shown are computed using the `XNS` code,[19,55–57,66,80,82] which solves the coupled equations for the metric, scalar field, and MHD structure of a NS under the assumptions of stationarity and axisymmetry, adopting conformal flatness and maximal slicing.

In Newtonian gravity[83,84] and in GR,[16,66] in the limit of weak magnetic fields, the quadrupole deformation e of a magnetised NS can be expressed as a linear function of $B^2_{\rm max}$.[66] Equivalently, instead of using $B^2_{\rm max}$ one can parametrise e also in terms of $\mathcal{H}/W$, where $\mathcal{H}$ is the magnetic energy of the NS, defined in the J-frame as

$$\tilde{\mathcal{H}} = \pi \int \mathcal{A}^3 \tilde{B}_i \tilde{B}^i \sqrt{\bar{\gamma}} {\rm d}r {\rm d}\theta \, , \tag{6}$$

W being its binding energy.

In STTs, we found that $\bar{e}$ still has a linear trend with $\tilde{B}^2_{\rm max}$ (or $\tilde{\mathcal{H}}/\bar{W}$), but with coefficients that depend more strongly on the baryonic mass M_0 (which is the same in the J-frame and in the E-frame) than in GR; the strength of this dependence is related to the value of β_0. In the limit $\tilde{B}_{\rm max} \to 0$, keeping fixed M_0 and β_0:

$$|\bar{e}| = c_{\rm B} \tilde{B}^2_{\rm max} + \mathcal{O}\left(\tilde{B}^4_{\rm max}\right), \quad |\bar{e}| = c_{\rm H} \frac{\tilde{\mathcal{H}}}{\bar{W}} + \mathcal{O}\left(\frac{\tilde{\mathcal{H}}^2}{\bar{W}^2}\right), \tag{7}$$

where $c_{\rm B} = c_{\rm B}(M_0, \beta_0)$ and $c_{\rm H} = c_{\rm H}(M_0, \beta_0)$ are the 'distortion coefficients', and $\tilde{B}_{\rm max}$ is normalised to 10^{18}G.

The coefficients $c_{\rm B}$ and $c_{\rm H}$ as functions of M_0, for various β_0, are shown in Fig. 1, in the case of NSs endowed with a purely toroidal or a purely poloidal magnetic field and described by the POL2 EoS $\tilde{p} = K_{\rm a} \tilde{\rho}^{\gamma_{\rm a}}$, with $\gamma_{\rm a} = 2$ and $K_{\rm a} = 110$ in dimensionless units. The black line represents GR, which corresponds to $\alpha_0 = \beta_0 = 0$. The other lines stand for $\beta_0 = \{-4.5, -4.75, -5.0, -5.25, -5.5, -5.75, -6.0\}$ and $\alpha_0 = -2 \times 10^{-4}$. The effect of scalarisation is visible in the distinctive change in the slope as the scalarised sequences depart from the GR one. The more negative

β_0 is, the more enhanced are the modifications with respect to GR. At a fixed M_0, scalarised NSs have lower distortion coefficient and quadrupole deformation than the GR models with the same mass for most of the scalarisation range. We found that, in the the ranges $1.2 \leq M_0/\mathrm{M}_\odot \lesssim 2.4$ and $-6 \leq \beta_0 \leq -4.5$, the distortion coefficients are well approximated by a combination of power laws of the baryonic mass M_0, the J-frame circumferential radius $\tilde{R}_\mathrm{c}$ and the E-frame scalar charge $\bar{Q}_\mathrm{s}$.[55] In particular

$$c_\mathrm{B} \approx c_1 M_{0;1.6}^{\alpha} R_{10}^{\beta} \left[1 - c_2 Q_1^{\gamma} M_{1.6}^{\delta} R_{10}^{\rho}\right] \ , \tag{8}$$

where $M_{0;1.6}$ is M_0 in units of $1.6\mathrm{M}_\odot$, R_{10} is $\tilde{R}_\mathrm{c}$ in units of 10km, Q_1 is $\bar{Q}_\mathrm{s}$ in units of $1\mathrm{M}_\odot$, $[c_1, \alpha, \beta, c_2, \gamma, \delta, \rho] = [0.16, -2.22, 4.86, 0.87, 1.32, -1.27, -2.21]$ in the toroidal case and $[c_1, \alpha, \beta, c_2, \gamma, \delta, \rho] = [0.077, -1.99, 5.80, 1.38, 1.22, -0.86, -3.49]$ in the poloidal case. The coefficient c_H in the toroidal cases is similar to that in the poloidal case; in fact, it is possible to find a functional form over the entire mass range that holds for both magnetic configuartions with an accuracy of few percents:

$$c_\mathrm{H} \approx 0.5 + \mathcal{F}(M_0)\mathcal{T}(M_0, \bar{Q}_\mathrm{s}, \tilde{R}_\mathrm{c}) \times \begin{cases} 0.65 \text{ for toroidal} \ , \\ 1.02 \text{ for poloidal} \ , \end{cases} \tag{9}$$

where $\mathcal{F}(M_0)$ does not depend explicitly on the scalar field and described the role of the equation of state, $\mathcal{T}(M_0, \bar{Q}_\mathrm{s}, \tilde{R}_\mathrm{c})$ represents the correction due to the presence of scalarisation, and the last numerical factor differentiates the oblate from the prolate geometry, which depends on the magnetic field configuration. We found that:

$$\mathcal{F}(M_0) \quad = 4.98 - 1.95 M_{0;1.6} \ , \tag{10}$$

$$\mathcal{T}(M_0, \bar{Q}_\mathrm{s}, \tilde{R}_\mathrm{c}) = 1 - \frac{1.90}{R_{10}^{2.45}} \left(\frac{Q_1}{M_{0;1.6}}\right)^{1.3} \ . \tag{11}$$

5. Selection of equations of state

The results shown in Sect. 4 are found using a simple polytropic law which, while allowing us to compare our results to previous studies,[16,19] is not allowed by observations. For this reason, we further computed the distortion coefficients for a selection of EoS allowed by observational and nuclear physics constraints.[57] In particular, we selected 13 different EoS that span a diverse range of calculation methods and particle contents. All EoS we used, except the polytropic one, are chosen to satisfy the lates constraints: they reach a maximum mass of at least $\sim 2.05\mathrm{M}_\odot$,[85] satisfy various nuclear physics constraints,[86] are not too stiff,[87] and give the NS a radius between $\sim$10km and $\sim$14km for $1.4\mathrm{M}_\odot$ mass models[88,89]. We chose 6 EoS which contain only $npe\mu$ particles (APR,SLY9,BL2,DDME2,NL3$\omega\rho$,SFH), 2 EoS containing also hyperons (DDME2-Y,NL3$\omega\rho$-Y), 2 EoS containing an *uds* quark matter domain treated with the Nambu-Jona-Lasinio model (BH8, BF9), 2 EoS containing an *uds* quark matter domain treated with the MIT bag model or perturbative QCD (SQM1, SQM2), as well as the POL2 EoS.

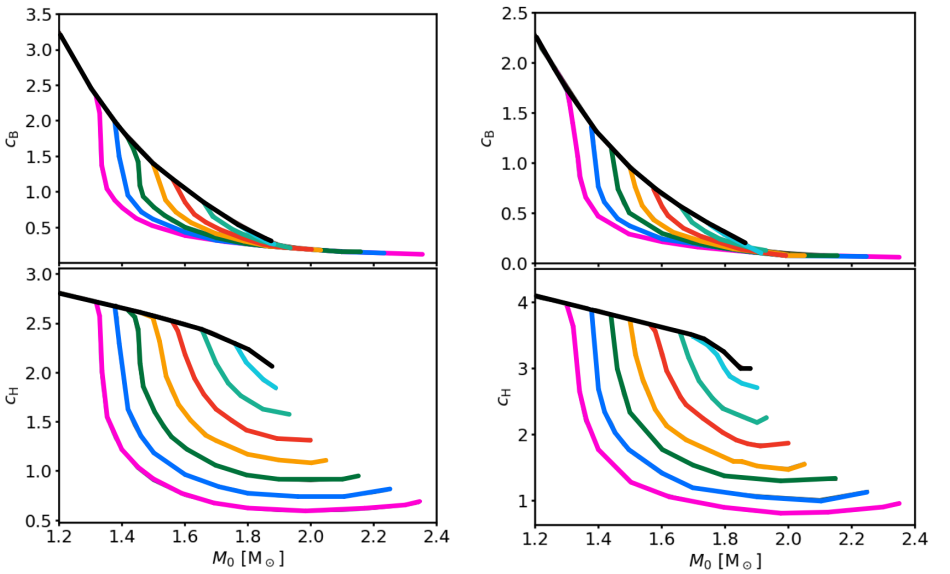


Fig. 1. Distortion coefficients c_B (top panels) and c_H (bottom panels) as functions of the baryonic mass M_0 of models with a purely toroidal magnetic field (left panels) and with a purely poloidal magnetic field (right panels), for various value of β_0: from $\beta_0 = -6$ (magenta curve) to $\beta_0 = -4.5$ (light blue curve) increasing by 0.25 with every line. The black curve corresponds to GR. Adapted from Soldateschi et al., A&A, 645, A39 (2021).

In Fig. 2 we plot the Komar mass $\bar{M}_k$ against the circumferential radius $\tilde{R}_c$ for models of un-magnetised, static NSs computed with the described EoS. The left panel refers to GR, while the right panel to STT with $\beta_0 = -6$. From the left panel of Fig. 2 we see that the NS radii have values ranging from ∼10km to ∼14km for most EoS, excluding the less compact SQM2 and especially the POL2 EoS. The maximum masses are around ∼2-2.2M$_\odot$ for most EoS, except for the NL3$\omega\rho$ and the POL2 EoS.

6. Quasi-universal relations for the magnetic deformation of neutron stars

We computed the distortion coefficients in Eq. 7 for the EoS described in Sect. 5, as well as another distortion coefficient:

$$|\bar{e}| = c_s \tilde{B}_s^2 + \mathcal{O}\left(\tilde{B}_s^4\right) , \tag{12}$$

where $\tilde{B}_s$ is the magnetic field calculated at the pole of the NS, at the surface, normalised to 10^{18}G. The coefficients c_B and c_H contain quantities that are not directly accessible by observations, because they require to know the details of the internal structure and magnetic field geometry of NSs. On the other hand, c_s may prove to be more useful to compare our results to observations. This last coefficient is

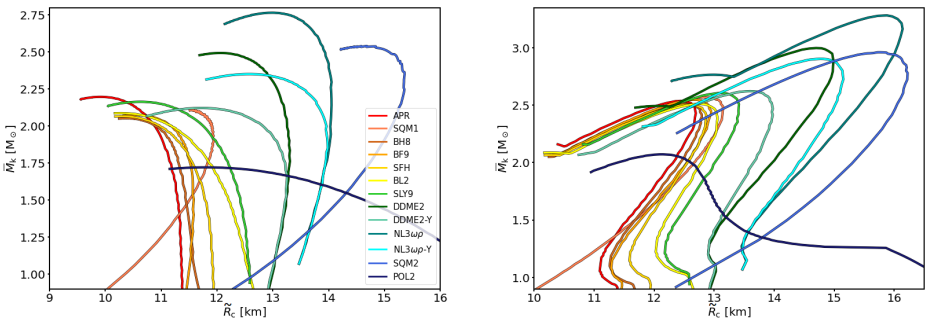


Fig. 2. Komar mass $\bar{M}_{\rm k}$ against circumferential radius $\tilde{R}_{\rm c}$ for un-magnetised, static models of NSs computed with the EoS described in Sect. 5 in GR (left plot) and in STT with $\beta_0 = -6$ (right plot). The EoS are color-coded, and ordered in the legend, according to the compactness $C = M_{\rm k}/R_{\rm c}$ calculated at $M_{\rm k} = 1.4{\rm M}_\odot$ in GR: red for the highest compactness and blue for the lowest compactness. Adapted from Soldateschi et al., ArXiv e-prints (2021), arXiv:2106.00603v2 [astro-ph.HE].

defined only for configurations endowed with a poloidal magnetic field (the toroidal one being hidden under the surface).

We found that $c_{\rm B}, c_{\rm H}$ and $c_{\rm s}$ are similar for all standard EoS (i.e. all EoS except SQM1, SQM2 and POL2) as functions of the Komar mass.[57] For this reason, we chose to consider the dependence also on another potentially observable quantity, namely the circumferential radius $\tilde{R}_{\rm c}$, to understand whether the spread among the various EoS could be further reduced. We used the 'principal component analysis' (PCA) technique to find the best-fit relation between $c_{\rm B,H,s}$, $\bar{M}_{\rm k}$ and $\tilde{R}_{\rm c}$. We found the following formulas approximating $c_{\rm B,H}$ in GR to a satisfying level of accuracy for all standard EoS (thus the name 'quasi-universal relations'):

$$c_{\rm B}^{\rm PCA} = \begin{cases} 0.13^{+0.03}_{-0.02} R_{10}^{5.45} M_{1.6}^{-2.41} \text{ for poloidal,} \\ \\ 0.25^{+0.03}_{-0.03} R_{10}^{5.03} M_{1.6}^{-2.07} \text{ for toroidal,} \end{cases} \tag{13}$$

$$c_{\rm H}^{\rm PCA} = \begin{cases} 5.77^{+0.04}_{-0.06} - 0.77 R_{10} - 4.14 M_{1.6} - 0.27 M_{1.6}^2 + \\ \quad +0.07 R_{10}^2 + 2.28 M_{1.6} R_{10} \text{ for poloidal,} \\ \\ 7.02^{+0.05}_{-0.07} - 5.22 R_{10} - 2.76 M_{1.6} - 0.12 M_{1.6}^2 + \\ \quad +1.92 R_{10}^2 + 1.51 M_{1.6} R_{10} \text{ for toroidal,} \end{cases} \tag{14}$$

where $M_{1.6} = \bar{M}_{\rm k}/1.6{\rm M}_\odot$. Moreover, we found that, in GR, $c_{\rm s}$ is well-approximated by the following relation:

$$c_{\rm s}^{\rm PCA} = 2.97^{+0.12}_{-0.23} R_{10}^{4.61} M_{1.6}^{-2.80} \, . \tag{15}$$

The coefficient $c_{\rm s}$ computed using Eq. 15 is plotted against the value computed with formula Eq. 12 in Fig. 3 (top left plot), together with the error of the approximation (bottom left plot). The dashed black line is a reference $c_{\rm s} = c_{\rm s}^{\rm PCA}$ bisecting line, which represents a perfect approximation. The superscript and subscript in the first coefficient of Eq. 15 are the values defining the purple and magenta lines bounding the magenta shaded area in Fig. 3, top left plot. The dark blue line, bounding the shaded blue area, marks the 90th percentile of the errors (the bounds containing 90% of the results). The corresponding values of these errors are showed with lines of the same colour in the bottom left plot. The approximation $c_{\rm s}^{\rm PCA}$ holds with a very small error, under a few percents. Moreover, we found that the quasi-universal relations for $c_{\rm B}$ hold with a slightly larger error, around $\sim 10\%$. The approximation for $c_{\rm H}$ is even more accurate, with an error that is mostly under $\sim 1\%$. Moreover, we found that performing $c_{\rm H}^{\rm PCA} \to 5/3 c_{\rm H}^{\rm PCA} - 0.9$ allows one to use the coefficients found in the toroidal case also in the poloidal configurations with a $\sim 2\%$ error.

Applying the quasi-universal relations derived in the case of standard EoS, Eqs. 13-14-15, to the polytropic models computed with the POL2 EoS, leads to larger errors, especially in the case of $c_{\rm B}$. We refer the reader to Ref. 57 for more details.

In the case of STTs, we find quasi-universal relations for $\Delta c_{\rm B} = |c_{\rm B} - c_{\rm B}^{\rm GR}|$, $\Delta c_{\rm H} = |c_{\rm H} - c_{\rm H}^{\rm GR}|$ and $\Delta c_{\rm s} = |c_{\rm s} - c_{\rm s}^{\rm GR}|$, where $c_{\rm B}^{\rm GR}, c_{\rm H}^{\rm GR}$ and $c_{\rm s}^{\rm GR}$ are the relations found in the GR case: Eqs. 13-14-15 respectively. In this case, we also allowed the dependence of the distortion coefficients on the scalar charge $\bar{Q}_{\rm s}$. We found the following quasi-universal relations:

$$\Delta c_{\rm B}^{\rm PCA} = \begin{cases} 0.03^{+0.05}_{-0.03} R_{10}^{8.23} M_{1.6}^{-5.08} Q_1^{2.60} \text{ for poloidal,} \\ \\ 0.06^{+0.09}_{-0.05} R_{10}^{5.96} M_{1.6}^{-3.52} Q_1^{1.95} \text{ for toroidal,} \end{cases} \tag{16}$$

$$\Delta c_{\rm H}^{\rm PCA} = \begin{cases} 1.96^{+0.17}_{-0.18} R_{10}^{0.72} M_{1.6}^{-1.96} Q_1^{1.54} \text{ for poloidal,} \\ \\ 1.49^{+0.26}_{-0.17} R_{10}^{0.75} M_{1.6}^{-1.81} Q_1^{1.55} \text{ for toroidal,} \end{cases} \tag{17}$$

$$\Delta c_{\rm s}^{\rm PCA} = 0.92^{+0.20}_{-0.27} R_{10}^{4.77} M_{1.6}^{-4.50} Q_1^{1.71} \ . \tag{18}$$

The quasi-universal relation in Eq. 18 is plotted against the corresponding value $\Delta c_{\rm s}$, computed through formula Eq. 12, in Fig. 3 (top right plot). The bottom right plot displays the relative error of the quasi-universal relation. The dashed line is a reference $\Delta c_{\rm s} = \Delta c_{\rm s}^{\rm PCA}$ bisecting line, which would be a perfect approximation. We can see that the approximation for $c_{\rm s}$ holds with an error that is around 10%. Moreover, we found that the quasi-universal relations for $\Delta c_{\rm B}$ hold with an error of

$\sim 50\%$. The approximation for $\Delta c_{\rm H}$ is more accurate, with a relative error of just a few percents. Similarly to what we found in the GR case, performing $c_{\rm H}^{\rm PCA} \to 3/2 c_{\rm H}^{\rm PCA} - 0.2$ allows us to use the coefficients found in the toroidal case in the poloidal geometry with a $\sim 10\%$ error.

Using Eq. 16 for models computed using the POL2 EoS, errors in approximating $\Delta c_{\rm B}$ remain of the same magnitude in the poloidal case, while they increase by $\sim 20\%$ in the toroidal case. Instead, Eq. 17 holds also for the POL2 EoS, at the expense of an error reaching a few tens of percents for the approximation of $\Delta c_{\rm H}$, as in the case of approximating $\Delta c_{\rm s}$ using Eq. 18 for polytropic models computed using the POL2 EoS.

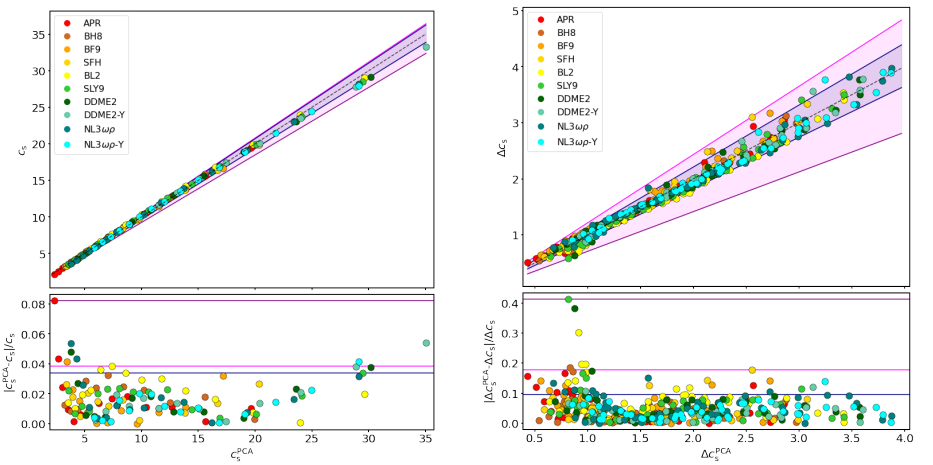


Fig. 3. Top left plot: distortion coefficient $c_{\rm s}$, calculated according to Eq. 12 in GR, versus its approximation $c_{\rm s}^{\rm PCA}$ calculated with the quasi-universal relation in Eq. 15. Top right plot: Difference $\Delta c_{\rm s}$ between the distortion coefficient $c_{\rm s}$, calculated according to Eq. 12 in STT with $\beta_0 \in \{-6, -5.75, -5.5, -5\}$, and the GR quasi-universal relation in Eq. 15. This is plotted versus its approximation $\Delta c_{\rm s}^{\rm PCA}$, calculated with the quasi-universal relation in Eq. 18. The corresponding relative deviations from the PCA are given in the bottom plots. The dashed lines are $c_{\rm B,H} = c_{\rm B,H}^{\rm PCA}$ and $\Delta c_{\rm s} = \Delta c_{\rm s}^{\rm PCA}$, respectively. The magenta shaded areas comprise all data points and the purple and magenta lines represent the upper and lower bounds of Eq. 15 and Eq. 18, respectively; the dark blue lines bounding the shaded blue area mark the 90th percentile error region. The EoS are color-coded, and ordered in the legend, according to the compactness $C = M_{\rm k}/R_{\rm c}$ calculated at $M_{\rm k} = 1.4 {\rm M}_\odot$ in GR: red for the highest compactness and blue for the lowest compactness. Adapted from Soldateschi et al., ArXiv e-prints (2021), arXiv:2106.00603v2 [astro-ph.HE].

As in the case of the POL2 EoS, we incur in larger errors if we apply the quasi-universal relations we found to the case of the SQM1 and SQM2 EoS. In GR, $c_{\rm B}^{\rm PCA}$ is a factor $\sim 0.4 - 0.8$ lower than $c_{\rm B}$. The maximum error for approximating $c_{\rm H}$ with Eq. 14 increases to $\sim 8 - 12\%$ in the case of SQM1, and $\sim 5\% (\sim 40\%)$ for purely poloidal (toroidal) magnetic fields in the case of SQM2. Moreover, $c_{\rm s}^{\rm PCA}$ is at

most a factor ~ 1.4 higher than $c_{\rm s}$. In STT, $\Delta c_{\rm B}^{\rm PCA}$ in the poloidal case is around a factor ~ 2 lower (higher) than $\Delta c_{\rm B}$ for the SQM1 (SQM2) EoS; in the toroidal case, it is a factor ~ 2 lower for both SQM1 and SQM2. The maximum error for approximating $\Delta c_{\rm H}$ using Eq. 17 increases to $\sim 30\%(\sim 50\%)$ for purely poloidal (toroidal) magnetic fields. Finally, $\Delta c_{\rm s}^{\rm PCA}$ is around a factor ~ 1.7 lower than $\Delta c_{\rm s}$.

A time-varying quadrupolar deformation leads to the emission of GWs. While in GR these are only of tensor nature, in the case of STTs a scalar channel is also present, which can contain any multipolar component. We only focus here on quadrupolar modes of GWs, both of tensor and scalar nature. As we have shown, NSs in GR and in STTs posses quite different quadrupolar deformations; for this reason, we compare the amount of GWs emitted by NSs in these two modes. To this end, we introduce the following quantity:

$$\mathcal{S} = \left| \frac{q_{\rm s}}{q_{\rm g}} \right|, \tag{19}$$

where

$$q_{\rm s} = 2\pi \int \alpha_{\rm s} \mathcal{A}^4 \tilde{T} \left(3\sin^2\theta - 2\right) r^4 \sin\theta {\rm d}r{\rm d}\theta, \tag{20}$$

$$q_{\rm g} = \int \left[\pi \mathcal{A}^4 (\tilde{\varepsilon} + \tilde{\rho}) - \frac{1}{8} \partial\chi \partial\chi \right] r^4 \sin\theta \left(3\sin^2\theta - 2\right) {\rm d}r{\rm d}\theta. \tag{21}$$

The quantity $q_{\rm s}$ is related to the source of scalar waves. The mass quadrupole $q_{\rm g}$ is $\bar{I}_{\rm zz} - \bar{I}_{\rm xx} = \bar{e}\bar{I}_{\rm zz}$ and is the source of tensor waves. Thus, the ratio $\mathcal{S}$ computed for a given NS model measures the relative amount of quadrupolar GWs emitted in the scalar and tensor channels. If $\mathcal{S} < 1$, the majority of GWs will be emitted in the tensor channel; if $\mathcal{S} > 1$, in the scalar channel. We found that the following quasi-universal relations hold for $\mathcal{S}$:

$$\mathcal{S}^{\rm PCA} = \begin{cases} 1.98^{+0.18}_{-0.05} R_{10}^{-0.71} M_{1.6}^{-0.54} Q_1^{1.22} \text{ for poloidal,} \\ \\ 1.99^{+0.18}_{-0.07} R_{10}^{-0.74} M_{1.6}^{-0.60} Q_1^{1.23} \text{ for toroidal.} \end{cases} \tag{22}$$

In Fig. 4 we plot the values of $\mathcal{S}$, computed through Eq. 19 for $\beta_0 \in \{-6, -5.75, -5.5, -5\}$, against their PCA approximation, computed through Eq. 22. The approximation is very accurate in both magnetic geometries, with an error that is mostly concentrated under $\sim 4\%$. Moreover, we see that the coefficients in Eq. 22 are almost identical in the poloidal and the toroidal cases, thus the two approximations are practically equivalent. It is possible that this similarity means that there exist a relation between the sources of scalar and tensor waves, $q_{\rm s}$ and $q_{\rm g}$, that does not depend on either the magnetic field geometry or the EoS.

The quasi-universal relations we found are independent on the EoS of standard NSs. For this reason they may be useful in disentangling the effects of the EoS and

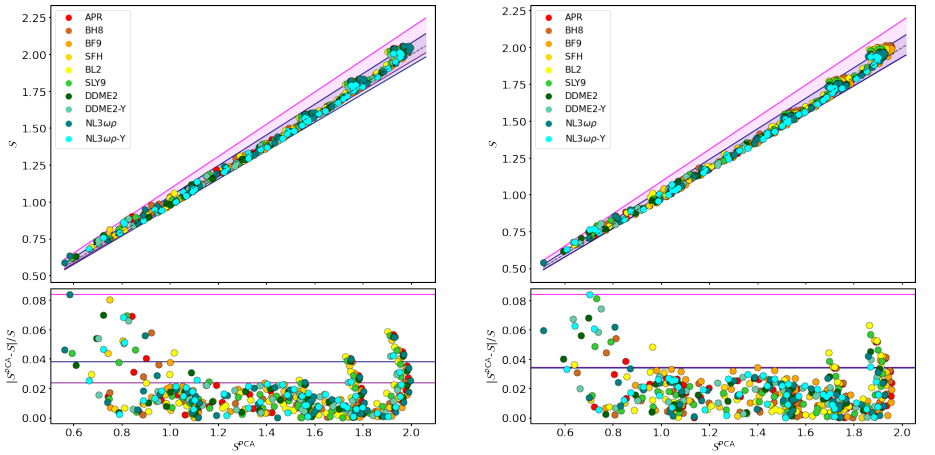


Fig. 4. Ratio $\mathcal{S}$ between scalar and tensor quadrupolar GW losses, calculated according to Eq. 19 in STT with $\beta_0 \in \{-6, -5.75, -5.5, -5\}$. This is plotted versus its approximation $\mathcal{S}^{\rm PCA}$, calculated with the quasi-universal relations in Eq. 22 (top plot in each panel). The corresponding relative devitations from the PCA are given in the bottom plot in each panel. The left panel refers to a purely poloidal magnetic field; the right panel refers to a purely toroidal magnetic field. The dashed line is $\mathcal{S} = \mathcal{S}^{\rm PCA}$. The magenta shaded area comprises all data points and the purple and magenta lines represent the upper and lower bounds of Eq. 22; the dark blue lines bounding the shaded blue area mark the 90th percentile error region. The EoS are color-coded, and ordered in the legend, according to the compactness $C = M_{\rm k}/R_{\rm c}$ calculated at $M_{\rm k} = 1.4{\rm M}_\odot$ in GR: red for the highest compactness and blue for the lowest compactness. Adapted from Soldateschi et al., ArXiv e-prints (2021), arXiv:2106.00603v2 [astro-ph.HE].

the magnetic field structure of NSs, leaving their internal magnetic structure as the only major unknown in GR. In this sense, the most promising relation is that of $c_{\rm s}$. On the one hand, $c_{\rm s}$ can be computed from its definition Eq. 12. In this case, one needs to measure both the magnetic field strength at the surface of the NS, $B_{\rm s}$, and its quadrupolar deformation e, which, in turn, can be estimated from the strain of CGWs emitted by the NS: $h_0 \propto eI$, where I is the moment of inertia of the NS along its rotation axis, which can be computed through an EoS-independent relation by knowing the NS mass and its radius.[90] On the other hand, our quasi-universal relation Eq. 15 allows one to estimate $c_{\rm s}^{\rm PCA}$ by knowing just the NS mass and radius. Comparing these values of $c_{\rm s}$ and $c_{\rm s}^{\rm PCA}$ can give an insight into the internal magnetic geometry of the emitting NS: the quasi-universal relation Eq. 15 was found in the case of a purely poloidal field, that is an extreme configuration which exherts the maximum possible magnetic deformation on the NS. For example, a lower value of $c_{\rm s}$ inferred by observations may imply the existence of a toroidal component of the NS internal magnetic field, which counteracts the effect of the poloidal component and results in a lower magnetic deformation. In the opposite case, another source of deformation may be present other than the magnetic field.

The quasi-universal relations Eqs. 13-14 may be more useful to constrain $B_{\rm max}$ or $\mathcal{H}/W$ themselves, being these quantities generally not observable. In particular,

$B_{\rm max} \approx (e/c_{\rm B})^{1/2}$ and $\mathcal{H}/W \approx e/c_{\rm H}$. As in the case of $c_{\rm s}$, we expect $c_{\rm B,H} < c_{\rm B,H}^{\rm PCA}$. If the NS mass and radius are known, one can esitimate a lower bound for both $B_{\rm max}$ and $\mathcal{H}/W$ by using the quasi-universal relations found in the purely poloidal and purely toroidal case. Moreover, the quasi-universal relations for $c_{\rm B}$, $c_{\rm H}$ and $c_{\rm s}$ can be useful to easily compute the distortion coefficients from the mass and radius of a model, without having to fully simulate the NS model.

In the case of STTs, some of the effects of the EoS are degenerate with the presence of a scalar charge inside the NS. For this reason, quasi-universal relations Eqs. 16-17-18 may help to understand whether a distortion coefficient inferred through relations Eqs. 8-12, is compatible with the observed NS possessing a scalar charge, independently from its EoS. Relation Eq. 22 does not involve knowing the strength of the magnetic field, thus it is probably more promising. In particular, following our previous argument, it may be that $\mathcal{S} \approx \mathcal{S}^{\rm PCA}$ independently of the magnetic configuration. Then, the observation of CGWs coming from a given NS of known mass, radius, distance d and spin period P translates into a lower bound for a function of the scalar charge. On the other hand, the non-observation of CGWs from a given NS can be translated into an upper bound.

Finally, we can use the quasi-universal relations we found to assess the detectability of known NSs, in GR. Since the strain of CGWs emitted by a deformed NS, rotating with frequency $f_{\rm rot}$ and at a distance d, is $h_0 \propto eIf_{\rm rot}^2/d$, by using the quasi-universal relation for $c_{\rm s}$ Eq. 15 we can compute h_0 for the pulsars of the ATNF[91] catalogue. In this case, $f_{\rm rot}$ and d are taken from the data in ATNF catalogue, while $M_{\rm k}, B_{\rm s}$ are generated from the expected distributions.[92,93] The radius $R_{\rm c}$ is computed by assuming the two most diverse standard EoS among the ones we considered, APR and NL3$\omega\rho$, and using the corresponding mass-radius diagram. Our results are shown in Fig. 5, where they are compared to the sensitivity of the advanced LIGO (aLIGO) detector in the design configuration[a] and of the Einstein Telescope (ET) detector in the D configuration[b]: the blue (red) solid line represents the sensitivity of aLIGO (ET), while the blue (red) dot-dashed and dashed lines show the minimum strain detectable by aLIGO (ET) assuming a 1 month or 2 years observing time, respectively. The x-axis represents the frequency f of the emitted CGWs. In the general case, CGWs are emitted at two frequencies, $f = f_{\rm rot}$ and $f = 2f_{\rm rot}$. In this case, we consider only the $f = 2f_{\rm rot}$ wave. As we can see from Fig. 5, many millisecond pulsars contained in the ATNF catalogue have a chance to emit CGWs which could be potentially observed with the future ET detector, especially considering a 1 month or 2 year observing campaign. As for aLIGO, in the case of 1 month or 2 years observing time, a few tens of the ATNF millisecond pulsars could potentially be detected through CGWs. We note that the difference in radii given by the two different EoS does not significantly alter the strain of the

[a]The aLIGO design densitivity curves can be found at `https://dcc.ligo.org/LIGO-T1800044/public`.
[b]The ET sensitivity curves can be found at `http://www.et-gw.eu/index.php/etsensitivities`.

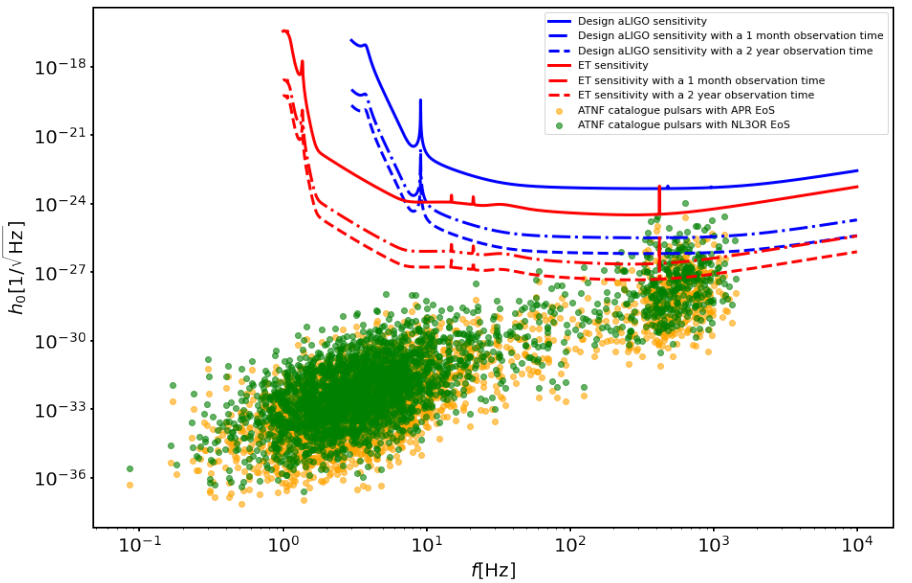


Fig. 5. CGWs strain h_0 computed for pulsars contained in the ATNF catalogue through the use of quasi-universal relation Eq. 15 and by assuming the APR EoS (orange points) or the NL3$\omega\rho$ EoS (green points). The x-axis represents the frequency f of the emitted CGWs. The blue (red) solid line represents the sensitivity of the aLIGO (ET) detector, while the blue (red) dot-dashed and dashed lines show the minimum strain detectable by aLIGO (ET) assuming a 1 month or 2 years observing time, respectively.

emitted CGWs. However, pulsars with smaller rotation period are much less likely to be observed, even with a 3rd generation detector like ET.

References

1. R. C. Duncan and C. Thompson, Formation of Very Strongly Magnetized Neutron Stars: Implications for Gamma-Ray Bursts, *ApJ* **392**, p. L9 (June 1992).
2. C. Thompson and R. C. Duncan, Neutron Star Dynamos and the Origins of Pulsar Magnetism, *ApJ* **408**, p. 194 (May 1993).
3. C. Thompson and R. C. Duncan, The soft gamma repeaters as very strongly magnetized neutron stars - I. Radiative mechanism for outbursts, *MNRAS* **275**, 255 (July 1995).
4. C. Thompson and R. C. Duncan, The Soft Gamma Repeaters as Very Strongly Magnetized Neutron Stars. II. Quiescent Neutrino, X-Ray, and Alfven Wave Emission, *ApJ* **473**, p. 322 (December 1996).
5. E. Asseo and D. Khechinashvili, The role of multipolar magnetic fields in pulsar magnetospheres, *MNRAS* **334**, 743 (08 2002).
6. H. C. Spruit, The source of magnetic fields in (neutron-) stars, in *Cosmic Magnetic Fields: From Planets, to Stars and Galaxies*, eds. K. G. Strassmeier, A. G. Kosovichev and J. E. Beckman, IAU Symposium, Vol. 259, April 2009.

7. L. Ferrario, A. Melatos and J. Zrake, Magnetic Field Generation in Stars, *Space Sci. Rev.* **191**, p. 77 (October 2015).
8. S. A. Olausen and V. M. Kaspi, The McGill Magnetar Catalog, *ApJS* **212**, p. 6 (May 2014).
9. S. B. Popov, Origins of magnetars in binary systems, *A&AT* **29**, 183 (January 2016).
10. N. Rea, P. Esposito, R. Turolla, G. L. Israel, S. Zane, L. Stella, S. Mereghetti, A. Tiengo, D. Götz, E. Göğüş and C. Kouveliotou, A Low-Magnetic-Field Soft Gamma Repeater, *Science* **330**, p. 944 (November 2010).
11. R. Staubert, J. Trümper, E. Kendziorra, D. Klochkov, K. Postnov, P. Kretschmar, K. Pottschmidt, F. Haberl, R. E. Rothschild, A. Santangelo, J. Wilms, I. Kreykenbohm and F. Fürst, Cyclotron lines in highly magnetized neutron stars, *A&A* **622**, p. A61 (February 2019).
12. L. Del Zanna and N. Bucciantini, Covariant and 3 + 1 equations for dynamo-chiral general relativistic magnetohydrodynamics, *MNRAS* **479**, 657 (06 2018).
13. R. Ciolfi, W. Kastaun, J. V. Kalinani and B. Giacomazzo, First 100 ms of a long-lived magnetized neutron star formed in a binary neutron star merger, *Phys. Rev. D* **100**, p. 023005 (July 2019).
14. K. Franceschetti and L. Del Zanna, General Relativistic Mean-Field Dynamo Model for Proto-Neutron Stars, *Universe* **6**, p. 83 (June 2020).
15. K. H. Prendergast, The Equilibrium of a Self-Gravitating Incompressible Fluid Sphere with a Magnetic Field. I, *ApJ* **123**, p. 498 (May 1956).
16. J. Frieben and L. Rezzolla, Equilibrium models of relativistic stars with a toroidal magnetic field, *MNRAS* **427**, p. 3406 (December 2012).
17. R. Ciolfi and L. Rezzolla, Twisted-torus configurations with large toroidal magnetic fields in relativistic stars, *MNRAS* **435**, L43 (August 2013).
18. K. Uryū, E. Gourgoulhon, C. M. Markakis, K. Fujisawa, A. Tsokaros and Y. Eriguchi, Equilibrium solutions of relativistic rotating stars with mixed poloidal and toroidal magnetic fields, *Phys. Rev. D* **90**, p. 101501 (November 2014).
19. A. G. Pili, N. Bucciantini and L. Del Zanna, Axisymmetric equilibrium models for magnetized neutron stars in general relativity under the conformally flat condition, *MNRAS* **439**, 3541 (2014).
20. L. Samuelsson and N. Andersson, Neutron star asteroseismology. Axial crust oscillations in the Cowling approximation, *MNRAS* **374**, 256 (January 2007).
21. H. Sotani, Torsional oscillations of neutron stars with highly tangled magnetic fields, *Phys. Rev. D* **92**, p. 104024 (November 2015).
22. D. Page, J. M. Lattimer, M. Prakash and A. W. Steiner, Minimal Cooling of Neutron Stars: A New Paradigm, *ApJS* **155**, 623 (December 2004).
23. D. N. Aguilera, J. A. Pons and J. A. Miralles, The Impact of Magnetic Field on the Thermal Evolution of Neutron Stars, *ApJ* **673**, p. L167 (February 2008).
24. B. Haskell, L. Samuelsson, K. Glampedakis and N. Andersson, Modelling magnetically deformed neutron stars, *MNRAS* **385**, 531 (March 2008).
25. R. O. Gomes, H. Pais, V. Dexheimer, C. Providência and S. Schramm, Limiting magnetic field for minimal deformation of a magnetized neutron star, *A&A* **627**, p. A61 (July 2019).
26. D. Kandel and R. W. Romani, Atmospheric circulation on black widow companions, *ApJ* **892**, p. 101 (April 2020).
27. T. L. S. Collaboration and the Virgo Collaboration, Gw170817: Measurements of neutron star radii and equation of state, *Phys. Rev. Lett.* **121**, p. 161101 (October 2018).
28. A. Bauswein, Equation of state constraints from multi-messenger observations of neutron star mergers, *Ann. Physics* **411**, p. 167958 (December 2019).

29. T. L. S. Collaboration and V. Collaboration, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (October 2017).
30. P. T. H. Pang, I. Tews, M. W. Coughlin, M. Bulla, C. Van Den Broeck and T. Dietrich, Nuclear-Physics Multi-Messenger Astrophysics Constraints on the Neutron-Star Equation of State: Adding NICER's PSR J0740+6620 Measurement, *arXiv e-prints*, p. arXiv:2105.08688 (May 2021).
31. N.-B. Zhang and B.-A. Li, Impacts of NICER's Radius Measurement of PSR J0740+6620 on Nuclear Symmetry Energy at Suprasaturation Densities, *arXiv e-prints*, p. arXiv:2105.11031 (May 2021).
32. M. Ruderman, Spin-driven changes in neutron star magnetic fields, *J. Astrophys. Astron.* **16**, 207 (June 1995).
33. J. L. Zdunik and P. Haensel, Maximum mass of neutron stars and strange neutron-star cores, *A&A* **551**, p. A61 (March 2013).
34. P. Costa, M. Ferreira, H. Hansen, D. P. Menezes and C. Providência, Phase transition and critical end point driven by an external magnetic field in asymmetric quark matter, *Phys. Rev. D* **89**, p. 056013 (March 2014).
35. B.-J. Cai, F. J. Fattoyev, B.-A. Li and W. G. Newton, Critical density and impact of Δ (1232) resonance formation in neutron stars, *Phys. Rev. C* **92**, p. 015802 (July 2015).
36. A. Drago, A. Lavagno, G. Pagliara and D. Pigato, The scenario of two families of compact stars, *Eur. Phys. J. A* **52**, p. 40 (February 2016).
37. M. Bocquet, S. Bonazzola, E. Gourgoulhon and J. Novak, Rotating neutron star models with a magnetic field, *A&A* **301**, p. 757 (September 1995).
38. C. Cutler, Gravitational waves from neutron stars with large toroidal b fields, *Phys. Rev. D* **66**, p. 084025 (October 2002).
39. S. Dall'Osso and L. Stella, Millisecond Magnetars, *arXiv e-prints*, p. arXiv:2103.10878 (March 2021).
40. E. Papantonopoulos, *Modifications of Einstein's Theory of Gravity at Large Distances*, Lecture Notes in Physics (Springer International Publishing, 2015).
41. V. Trimble, Existence and Nature of Dark Matter in the Universe, *ARA&A* **25**, 425 (September 1987).
42. P. J. E. Peebles and B. Ratra, The cosmological constant and dark energy, *Rev. Mod. Phys.* **75**, 559 (April 2003).
43. S. Capozziello and M. de Laurentis, Extended Theories of Gravity, *Phys. Rep.* **509**, 167 (2011).
44. T. Damour, F. Piazza and G. Veneziano, Runaway Dilaton and Equivalence Principle Violations, *Phys. Rev. Lett.* **89**, p. 081601 (August 2002).
45. A. DeFelice, M. Hindmarsh and M. Trodden, Ghosts, instabilities, and superluminal propagation in modified gravity models, *J. Cosmology Astropart. Phys.* **2006**, p. 005 (August 2006).
46. A. De Felice and T. Tanaka, Inevitable Ghost and the Degrees of Freedom in f(R,G) Gravity, *Prog. Theor. Phys.* **124**, 503 (September 2010).
47. O. Bertolami and J. Páramos, Viability of nonminimally coupled f (R) gravity, *Gen. Relativ. Gravit.* **48**, p. 34 (March 2016).
48. T. Damour and G. Esposito-Farèse, Nonperturbative strong-field effects in tensor-scalar theories of gravitation, *Phys. Rev. Lett.* **70**, 2220 (April 1993).
49. D. M. Eardley, D. L. Lee and A. P. Lightman, Gravitational-wave observations as a tool for testing relativistic gravity, *Phys. Rev. D* **8**, 3308 (November 1973).

50. P. T. H. Pang, R. K. L. Lo, I. C. F. Wong, T. G. F. Li and C. Van Den Broeck, Generic searches for alternative gravitational wave polarizations with networks of interferometric detectors, *Phys. Rev. D* **101**, p. 104055 (May 2020).
51. M. Shibata, K. Taniguchi, H. Okawa and A. Buonanno, Coalescence of binary neutron stars in a scalar-tensor theory of gravity, *Phys. Rev. D* **89**, p. 084005 (April 2014).
52. H. Sotani and K. D. Kokkotas, Stellar oscillations in scalar-tensor theory of gravity, *Phys. Rev. D* **71**, p. 124038 (June 2005).
53. P. Pani and E. Berti, Slowly rotating neutron stars in scalar-tensor theories, *Phys. Rev. D* **90**, p. 024025 (Jul 2014).
54. D. D. Doneva, S. S. Yazadjiev, N. Stergioulas and K. D. Kokkotas, Rapidly rotating neutron stars in scalar-tensor theories of gravity, *Phys. Rev. D* **88**, p. 084060 (October 2013).
55. J. Soldateschi, N. Bucciantini and L. Del Zanna, Axisymmetric equilibrium models for magnetised neutron stars in scalar-tensor theories, *A&A* **640**, p. A44 (August 2020).
56. J. Soldateschi, N. Bucciantini and L. Del Zanna, Magnetic deformation of neutron stars in scalar-tensor theories, *A&A* **645**, p. A39 (2021).
57. J. Soldateschi, N. Bucciantini and L. Del Zanna, Quasi-universality of the magnetic deformation of neutron stars in general relativity and beyond, *arXiv e-prints*, p. arXiv:2106.00603 (June 2021).
58. N. Bucciantini and J. Soldateschi, Iron line from neutron star accretion discs in scalar tensor theories, *MNRAS* **495**, L56 (April 2020).
59. C. M. Will, The Confrontation between General Relativity and Experiment, *Living Rev. Relativ.* **17** (2014).
60. L. Shao, N. Sennett, A. Buonanno, M. Kramer and N. Wex, Constraining nonperturbative strong-field effects in scalar-tensor gravity by combining pulsar timing and laser-interferometer gravitational-wave detectors, *Phys. Rev. X* **7**, 041025 (October 2017).
61. G. Voisin, I. Cognard, P. C. C. Freire, N. Wex, L. Guillemot, G. Desvignes, M. Kramer and G. Theureau, An improved test of the strong equivalence principle with the pulsar in a triple star system, *A&A* **638**, p. A24 (June 2020).
62. F. M. Ramazanoğlu and F. Pretorius, Spontaneous Scalarization with Massive Fields, *Phys. Rev. D* **93**, p. 064005 (March 2016).
63. S. S. Yazadjiev, D. D. Doneva and D. Popchev, Slowly rotating neutron stars in scalar-tensor theories with a massive scalar field, *Phys. Rev. D* **93**, p. 084038 (April 2016).
64. R. Rosca-Mead, C. J. Moore, U. Sperhake, M. Agathos and D. Gerosa, Structure of neutron stars in massive scalar-tensor gravity, *Symmetry* **12**, p. 1384 (2020).
65. A. Oron, Relativistic magnetized star with poloidal and toroidal fields, *Phys. Rev. D* **66**, p. 023006 (July 2002).
66. A. G. Pili, N. Bucciantini and L. Del Zanna, General relativistic models for rotating magnetized neutron stars in conformally flat space-time, *MNRAS* **470**, 2469 (2017).
67. J. R. Wilson and G. J. Mathews, *Relativistic Numerical Hydrodynamics*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2003).
68. J. A. Isenberg, Waveless approximation theories of gravity, *Int. J. Mod. Phys. D* **17**, 265 (February 2008).
69. N. Bucciantini and L. Del Zanna, A fully covariant mean-field dynamo closure for numerical 3 + 1 resistive GRMHD, *MNRAS* **428**, 71 (January 2013).
70. N. Tomei, L. Del Zanna, M. Bugli and N. Bucciantini, General relativistic magnetohydrodynamic dynamo in thick accretion discs: fully non-linear simulations, *MNRAS* **491**, 2346 (January 2020).

71. M. Alcubierre, *Introduction to 3+1 Numerical Relativity*, International Series of Monographs on Physics (OUP Oxford, 2008).
72. É. Gourgoulhon, *3+1 Formalism in General Relativity: Bases of Numerical Relativity*, Lecture Notes in Physics (Springer Berlin Heidelberg, 2012).
73. L. Del Zanna and C. Chiuderi, Exact solutions for symmetric magnetohydrodynamic equilibria with mass flow, *A&A* **310**, 341 (June 1996).
74. P. G. Bergmann, Comments on the scalar-tensor theory, *Int. J. Theor. Phys.* **1**, 25 (May 1968).
75. R. V. Wagoner, Scalar-Tensor Theory and Gravitational Waves, *Phys. Rev. D* **1**, 3209 (June 1970).
76. D. I. Santiago and A. S. Silbergleit, On the energy-momentum tensor of the scalar field in scalar-tensor theories of gravity, *Gen. Relativ. Gravit.* **32**, p. 565–582 (April 2000).
77. D. D. Doneva and S. S. Yazadjiev, Rapidly rotating neutron stars with a massive scalar field—structure and universal relations, *J. Cosmology Astropart. Phys.* **2016**, 019 (nov 2016).
78. X. Zhang, T. Liu and W. Zhao, Gravitational radiation from compact binary systems in screened modified gravity, *Phys. Rev. D* **95**, p. 104027 (May 2017), arXiv: 1702.08752.
79. X. Zhang, R. Niu and W. Zhao, Constraining the scalar-tensor gravity theories with and without screening mechanisms by combined observations, *Phys. Rev. D* **100**, p. 024038 (July 2019).
80. A. G. Pili, N. Bucciantini and L. Del Zanna, General relativistic neutron stars with twisted magnetosphere, *MNRAS* **447**, 2821 (March 2015).
81. N. Bucciantini, A. G. Pili and L. D. Zanna, The role of currents distribution in general relativistic equilibria of magnetized neutron stars, *MNRAS* **447**, 1 (2015).
82. N. Bucciantini and L. Del Zanna, General relativistic magnetohydrodynamics in axisymmetric dynamical spacetimes: the X-ECHO code, *A&A* **528**, A101 (2011).
83. D. G. Wentzel, Hydromagnetic Equilibria., *ApJS* **5**, p. 187 (December 1960).
84. J. P. Ostriker and J. E. Gunn, On the Nature of Pulsars. I. Theory, *ApJ* **157**, p. 1395 (September 1969).
85. E. Fonseca, H. T. Cromartie, T. T. Pennucci, P. S. Ray, A. Y. Kirichenko, S. M. Ransom, P. B. Demorest, I. H. Stairs, Z. Arzoumanian, L. Guillemot, A. Parthasarathy, M. Kerr, I. Cognard, P. T. Baker, H. Blumer, P. R. Brook, M. DeCesar, T. Dolch, F. A. Dong, E. C. Ferrara, W. Fiore, N. Garver-Daniels, D. C. Good, R. Jennings, M. L. Jones, V. M. Kaspi, M. T. Lam, D. R. Lorimer, J. Luo, A. McEwen, J. W. McKee, M. A. McLaughlin, N. McMann, B. W. Meyers, A. Naidu, C. Ng, D. J. Nice, N. Pol, H. A. Radovan, B. Shapiro-Albert, C. M. Tan, S. P. Tendulkar, J. K. Swiggum, H. M. Wahl and W. W. Zhu, Refined Mass and Geometric Measurements of the High-mass PSR J0740+6620, *ApJ* **915**, p. L12 (July 2021).
86. M. Fortin, C. Providencia, A. R. Raduta, F. Gulminelli, J. L. Zdunik, P. Haensel and M. Bejger, Neutron star radii and crusts: uncertainties and unified equations of state, *Phys. Rev. C* **94** (2016).
87. A. Guerra Chaves and T. Hinderer, Probing the equation of state of neutron star matter with gravitational waves from binary inspirals in light of GW170817: A brief review, *J. Phys. G: Nucl. Part. Phys.* **46**, p. 123002 (December 2019).
88. A. Bauswein, O. Just, H.-T. Janka and N. Stergioulas, Neutron-star radius constraints from gw170817 and future detections, *ApJ* **850**, p. L34 (Dec 2017).

89. T. E. Riley, A. L. Watts, P. S. Ray, S. Bogdanov, S. Guillot, S. M. Morsink, A. V. Bilous, Z. Arzoumanian, D. Choudhury, J. S. Deneva, K. C. Gendreau, A. K. Harding, W. C. G. Ho, J. M. Lattimer, M. Loewenstein, R. M. Ludlam, C. B. Markwardt, T. Okajima, C. Prescod-Weinstein, R. A. Remillard, M. T. Wolff, E. Fonseca, H. T. Cromartie, M. Kerr, T. T. Pennucci, A. Parthasarathy, S. Ransom, I. Stairs, L. Guillemot and I. Cognard, A NICER View of the Massive Pulsar PSR J0740+6620 Informed by Radio Timing and XMM-Newton Spectroscopy, *arXiv e-prints*, p. arXiv:2105.06980 (May 2021).
90. C. Breu and L. Rezzolla, Maximum mass, moment of inertia and compactness of relativistic stars, *MNRAS* **459**, 646 (June 2016).
91. R. N. Manchester, G. B. Hobbs, A. Teoh and M. Hobbs, The Australia Telescope National Facility Pulsar Catalogue, *AJ* **129**, 1993 (April 2005).
92. J. Antoniadis, T. M. Tauris, F. Ozel, E. Barr, D. J. Champion and P. C. C. Freire, The millisecond pulsar mass distribution: Evidence for bimodality and constraints on the maximum neutron star mass, *arXiv e-prints*, p. arXiv:1605.01665 (May 2016).
93. C.-A. Faucher-Giguère and V. M. Kaspi, Birth and Evolution of Isolated Radio Pulsars, *ApJ* **643**, 332 (May 2006).

Screening and elastic properties of the NS crust in the OCP approximation

D. Barba González*, C. Albertus Torres† and M.A. Pérez-García‡

Department of Fundamental Physics, University of Salamanca, Plaza de la Merced s/n E-37008, Salamanca, Spain

** david.barbag@usal.es, † albertus@usal.es, ‡ mperezga@usal.es*

We study the stress tensor of the outer crust of Neutron Stars (NSs) under the approximation of the one component plasma (OCP) with screening. In our approach a system of identical ions is placed in an electron degenerate Fermi sea. By using Molecular Dynamics simulations at finite temperature we obtain the stress tensor components and associated pressure including Ewald sums for the electron screened ion potential. Our results show that a careful characterization of the crystallized phase at low temperature is necessary in order to obtain the true ground state. These ion phases are of interest for modelling the effects of large tidal forces and the gravitational wave signal linked to the violent events in binary NS mergers and possible crust crunching in NS continuous emission.

Keywords: Neutron star, dense matter, crust

1. Introduction

The extreme conditions of matter inside Neutron Stars (NSs) are hard to grasp based on our knowledge of regular low density condensed matter on Earth. Nevertheless it is crucial for our understanding of these objects to theoretically model their interior. High density matter present in their core is thought to be a homogeneous system of relativistic particles, mostly neutron rich. Typical NSs are believed to be formed out of nucleons and possibly heavier particles and a lepton sector providing electrical charge neutrality. Located in their outer shells, the crust is far from being a trivial phase.[1] It is composed of non-homogeneous matter, typically a set of irregular shapes, pasta phases,[2] that at lower densities end up forming a regular ion lattice. Besides, in order to maintain electrical charge neutrality there must be a degenerate electron Fermi gas surrounding these structures. Some of the neutrons leaking from these irregular shapes can pair into superfluid phases in the lower density part of the NS crust.[3]

Experimental constraints on the nature of NSs have been difficult to obtain, only in the last decades satellite missions or big water tank detectors have been able to deepen inside the electromagnetic or neutrino probes of dynamics of their interior in different stages of their birth and evolution. The dawn of Multi-Messenger Physics is thus a promising tool to help understand different aspects of NSs and has proven possible to observationally test some of the theoretical assumptions about the physical conditions of not only the crust, but also the core and the

global Equation of State (EoS) of nuclear matter. In particular, another type of messengers, gravitational waves, show that the elastic properties of this material play a significant role in the strength and shape of measured signals involving these objects.

In this contribution we are interested in evaluating the impact of screening on forces, potential energies and stress tensor, as they are key to understand some mechanical properties of the very outer crust, which typically extend within hundreds of meters in the most external layers of the object.[4] Matter in this region is formed by a mixture of ions and electrons are stripped off them forming a degenerate Fermi sea. Typical ion mean densities[5] belong to the range between around 10^{-5} to 10^{-4} fm^{-3}.

As mentioned, in this region one expects to find a state of matter called Coulomb liquid, formed by different nuclear species, embedded in a degenerate electron gas. This gas makes sure that the system is kept charge neutral, and also screens the electromagnetic interactions between the ions so that this interaction is no longer pure Coulomb-wise but accounts for screening due to the electron gas. Thus the potential created by a positive charge Z_i (in units of electron charge e) at position $\vec{r_i}$, that describes this interaction at r is the so-called *Debye potential*, and has the form of a two-body Coulomb interaction with an exponential tail

$$\phi_i\left(\vec{r}\right) = \frac{Z_i}{\left|\vec{r} - \vec{r_i}\right|} e^{-\frac{\left|\vec{r} - \vec{r_i}\right|}{\lambda}}, \tag{1}$$

where λ is the Thomas-Fermi screening length

$$\lambda = \frac{\pi^{\frac{1}{2}}}{\left(4\alpha k_F\right)^{\frac{1}{2}} \left(k_F^2 + m_e^2\right)^{\frac{1}{4}}} \sim \frac{1}{2k_F}\sqrt{\frac{\pi}{\alpha}}. \tag{2}$$

k_F and m_e are the electron Fermi momentum and mass, respectively. α is the fine structure constant.

This potential ϕ_i is such that the interaction described tends to a pure Coulomb interaction when the screening length is much larger than the typical interparticle distance. In terms of the ion species that will be considered to describe the material in the outer NS crust, this will arise after the material in the hot protoNS cools down and its composition finally stabilizes and freezes to a ground state reached at $T \sim 10^5$ K. It is expected that the realistic composition in this layer is that of a multi-component plasma, comprised of different ion species. However, we would like to approach this complex problem by investigating only one, the most frequent nucleus, at a given density and temperature. This is known in the literature as the one-component plasma (OCP).[6]

Our approach is thus to assume that the material under inspection is comprised of a single ion species with fixed electrical charge and atomic mass number (Z, A)

embedded in the degenerate electron gas of number density equal to that of protons $n_e = n_p$, and whose interaction is described by the Debye potential. Note that in our calculation we will be using Molecular Dynamics and solving the equations of motion of ion degrees of freedom described in the quantum perspective as Gaussian charge density distributions. Similarly to the construction for the Coulomb interaction (which has an infinite range) we will consider the efficient Ewald summation technique[7] for the Debye potential.

2. Molecular Dynamics and the Ewald summation technique

Molecular Dynamics (MD) is a many-body technique to simulate the microscopic behaviour of matter in which the equations of motion of the individual degrees of freedom (ions in our case) are integrated while interacting via a prescribed potential. Typically for the integration of the equations of motion an algorithm is used. We use a Velocity Verlet one in our work. The final goal of this is obtaining the time evolution of the positions and velocities of the ions from which derived properties can be calculated once the system is thermalized. We use the NVT ensamble where ions (Fe or more general species with different Z and A) are the degrees of freedom for which we integrate the individual equations of motion, using the aforementioned Debye potential, as we are in the OCP approximation. The number of ions, size of the simulation box $L = V^{1/3}$, and the kinetic energy (or temperature T) remain fixed during the evolution starting from a random distribution. Also, the lepton fraction $Y_e = \frac{N_e}{N_p+N_n}$ obtained from number of protons (N_p), neutrons (N_n) and electrons ($N_e = N_p$) is fixed, as well as the OCP parameters (Z_i, A_i), for which we use fixed $Z_i = Z$, $A_i = A$. The crystallization parameter in the Coulomb case is defined as $\Gamma^{OCP} = \frac{Z^2}{k_B T}\left(\frac{4\pi}{3} n_I\right)^{1/3}$ and for $\Gamma^{OCP} > 175$ an ordered state is thought to arise.

The dynamics of the N_I ions are obtained by solving Newton's law $\vec{F_i} = m_i \vec{a_i}$, $i = 1, N_I$ assuming that the resulting potential is created by adding that of each ion. Note that the charge of each ion, ρ_i is spread into a Gaussian distribution, a more realistic case than the frequently assumed point-like charges, populating the layers in outer NS crusts. More in detail we have

$$\rho_{i,a}(r) = Z_i \left(\frac{a}{\pi}\right)^{\frac{3}{2}} e^{-ar^2}, \tag{3}$$

where a is related to the Gaussian width, σ_a, as $a = \frac{1}{2\sigma_a^2}$ and r is the distance coordinate taken from the center of the Gaussian source.

The Debye interaction potential produced by such a Gaussian source with charge and inverse width Z_i, a at $\vec{r_i}$ is then obtained by solving the Poisson equation

$$-\frac{1}{r}\frac{d^2}{dr^2}(r\phi(r)) + \frac{\phi(r)}{\lambda^2} = 4\pi \left[Z \left(\frac{a}{\pi}\right)^{3/2} e^{-ar^2} \right], \tag{4}$$

so that the solution is

$$\phi_{Zi,a}\left(\vec{r}\right) = \frac{Z_i}{2|\vec{r}-\vec{r_i}|} e^{\frac{1}{4a\lambda^2}} \left[e^{-\frac{|\vec{r}-\vec{r_i}|}{\lambda}} \mathrm{erfc}\left(\frac{1}{2\sqrt{a}\lambda} - \sqrt{a}|\vec{r}-\vec{r_i}| \right) \right.$$
$$\left. - e^{\frac{|\vec{r}-\vec{r_i}|}{\lambda}} \mathrm{erfc}\left(\frac{1}{2\sqrt{a}\lambda} + \sqrt{a}|\vec{r}-\vec{r_i}| \right) \right], \tag{5}$$

with erfc the complementary error function.

Even if the interaction between ions in our simulations is thought to be screened, it can reach longer than the size of the cubic simulation box $L = \left(\frac{N_I}{n_I}\right)^{1/3}$ where N_I and n_I are the number of ions and ion number density used. A proper treatment of the long-range part of the interaction is required. Usually, this is partly dealt with via periodic boundary conditions (PBC), thinking of the system as a infinite lattice of simulation boxes in which the ions interact with the closest neighbors of the different images of a selected box. By using the Ewald summation technique one takes advantage of including all the long-range interactions between ions irrespectively of the values of the screening length and the size of the simulation box. Some previous works have been developed for pure Coulomb potentials,[8] but to our knowledge this has not been done for electron screened versions. In this contribution we have developed the relevant expressions for the MD calculation of the microscopic state of matter.

In order to calculate the energies and accelerations in the Ewald summation scheme, we must first introduce *spurious* Gaussian charges into the system to screen the real ones. These are centered at the same positions as the real charges but have opposite sign, $-Z_i$, and display an inverse squared width α_{Ewald}

$$\rho_{-Z_i,\alpha_{\mathrm{Ewald}}} = -Z_i \left(\frac{\alpha_{\mathrm{Ewald}}}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_{\mathrm{Ewald}} r^2}. \tag{6}$$

These are the so-called *screening charges*. Consequently as a way of maintaing the electrical charge neutrality of the system, we are forced to introduce compensating charges that are identical in shape to the screening ones except for the charge, which in this case is positive $\rho_{Z_i,\alpha_{\mathrm{Ewald}}} = Z_i \left(\frac{\alpha_{\mathrm{Ewald}}}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_{\mathrm{Ewald}} r^2}$.

It is key to note here that the parameter α_{Ewald} is an arbitrary square inverse length that we introduced for convenience to screen the original interaction. The physical observables that one can obtain from the simulation thus must not depend on this parameter.

The total interaction energy U_{tot} is then divided into three parts. Firstly, a short-range part, $U_{\mathrm{short-range}}$, which is the interaction between the real charges and the sum of the real plus the screening charges. The total interaction potential, $\phi_{\mathrm{tot},i}$ concerning this short-range part of the interaction is produced by this sum of real plus screening charges.

$$\phi_{\mathrm{tot,i}}\left(|\vec{r}-\overrightarrow{r_i}|\right)=\frac{Z_i}{2\,|\vec{r}-\overrightarrow{r_i}|}\left\{e^{\frac{-|\vec{r}-r_i|}{\lambda}}\left[e^{\frac{1}{4a\lambda^2}}\,\mathrm{erfc}\left(\frac{1}{2\sqrt{a}\lambda}-\sqrt{a}\,|\vec{r}-\overrightarrow{r_i}|\right)\right.\right.$$
$$\left.-e^{\frac{1}{4\alpha_{\mathrm{Ewald}}\,\lambda^2}}\,\mathrm{erfc}\left(\frac{1}{2\sqrt{\alpha_{\mathrm{Ewald}}}\,\lambda}-\sqrt{\alpha_{\mathrm{Ewald}}}\,|\vec{r}-\overrightarrow{r_i}|\right)\right]$$
$$-e^{\frac{|\vec{r}-r_i|}{\lambda}}\left[e^{\frac{1}{4a\lambda^2}}\,\mathrm{erfc}\left(\frac{1}{2\sqrt{a}\lambda}+\sqrt{a}\,|\vec{r}-\vec{r}_i|\right)\right.$$
$$\left.\left.-e^{\frac{1}{4\alpha_{\mathrm{Ewald}}\lambda^2}}\,\mathrm{erfc}\left(\frac{1}{2\sqrt{\alpha_{\mathrm{Ewald}}}\lambda}+\sqrt{\alpha_{\mathrm{Ewald}}}\,|\vec{r}-\vec{r}_i|\right)\right]\right\}. \tag{7}$$

When using the Coulomb interaction one can verify that expressions have a closed, analytical expression, but for the Debye case, instead, one has to perform numerical integrations each time step of the simulation in order to obtain the forces (as gradient of this potential). The final expression for this short-range part of the interaction energy is given by

$$U_{\mathrm{short-range}}=\frac{1}{2}\sum_{i=1}^{N_I}\sum_{j\neq i=1}^{N_I}2Z_j\left(\frac{a}{\pi}\right)^{\frac{1}{2}}\frac{e^{-ar_{ij}^2}}{r_{ij}}\int_0^\infty r'\phi_{\mathrm{tot},i}\left(r'\right)e^{-ar'^2}\sinh\left(2ar_{ij}r'\right)dr'$$
$$-\frac{2\pi}{V}\sum_{=1i}^{N_I}\sum_{j=1}^{N_I}Z_j\int_0^\infty r'^2\phi_{\mathrm{tot,i}}\left(r'\right)dr', \tag{8}$$

where $r_{ij}=|\vec{r_i}-\vec{r_j}|$ is the distance between the N_I ions labeled i and j.

The usefulness of the separation into short and long range contributions stands clear here as the long-range energy term has a dependence in the interparticle distance that tails off quickly, so that it converges very rapidly in real space. The upper limit of the numerical integration is set, practically, at the finite value $r'=\sqrt{2}L$.

The second energy term, U_{l}, of the scheme is produced by the interaction between real charges and the compensating charges in all the simulation boxes that exist in the infinite lattice that was created using PBCs. This is done by transforming Poisson's equation from the r-coordinate to the Fourier k-space. The derivation's endpoint is a summation over reciprocal lattice vectors, the interaction going to zero when the module of these vectors is high enough so that the sum converges in the Fourier space,

$$U_{\mathrm{l}}=\frac{1}{2}\sum_{i,j=1}^{N_I}\sum_{\vec{k}\neq 0}\frac{4\pi Z_iZ_j}{V\left(k^2+\frac{1}{\lambda^2}\right)}e^{\frac{-k^2}{4}\left(\frac{1}{a}+\frac{1}{\alpha_{\mathrm{Ewald}}}\right)}e^{i\vec{k}(\vec{r_i}-\vec{r_j})}, \tag{9}$$

where $\vec{k}=\frac{2\pi}{L}\left(n_{k,x},n_{k,y},n_{k,z}\right)$ being $n_{k,x},n_{k,y},n_{k,z}\in\mathbb{Z}$ integers in the Cartesian XYZ directions. Finally, it remains to substract the interaction between the real particle and its own compensating charge in the same simulation box, as it is included spuriously in the long-range part. It also includes integrals that are to be

numerically solved. This energy from self-interaction is given by

$$U_{self} = 2\pi \left(\frac{a}{\pi}\right)^{\frac{3}{2}} \sum_{i=1}^{N_I} Z_i \int_0^{\infty} r'^2 \phi_{Z_i,\alpha_{\text{Ewald}}}(r')\, e^{-ar'^2} dr', \tag{10}$$

where $\phi_{Z_i,\alpha_{\text{Ewald}}}$ is the potential produced by the compensating charges alone, from equation 5. The global expression for the interaction energy is $U_{tot} = U_{\text{short-range}} + U_{\text{l}} - U_{self}$.

3. Stress tensor in the OCP approximation

In our study we focus on the OCP stress tensor. We can write the total stress tensor as follows[7]

$$\Pi_{\alpha\beta}^{\text{tot}} = \frac{1}{L^3} \sum_{i=1}^{N_I} m \dot{\vec{r}}_{i\alpha} \dot{\vec{r}}_{j\beta} + \Pi_{\alpha\beta} \tag{11}$$

where m is the mass of ions. This second-order rank tensor can be put into the form $\Pi_{\alpha\beta} = \Pi_{\alpha\beta}^{\text{real}} + \Pi_{\alpha\beta}^{\text{recip}}$, arising from the interparticle forces. As the interactions in real space may be written in a pairwise fashion, the configurational part of the pressure tensor, $\Pi_{\alpha\beta}^{\text{real}}$, can be evaluated directly from the real forces by employing the virial expression

$$\Pi_{\alpha\beta}^{\text{real}} = \frac{1}{2V} \sum_{i,j\neq i=1}^{N_I} \left(F_{ij,\alpha}^{\text{real}}\, r_{ij,\beta} + F_{ij,\beta}^{\text{real}} r_{ij,\alpha}\right), \tag{12}$$

where α, β are tensorial indexes running from 0 to 3 (time and XYZ spatial coordinates). We use Minkowski metric $g_{\alpha,\beta} = diag(1,-1,-1,-1)$.

The real space forces arise from the short-range part of the potential and take the form

$$\begin{aligned} \vec{F}_{\text{ij, short-range}} &= 2\left(\frac{a}{\pi}\right)^{\frac{1}{2}} \frac{Z_j e^{-ar_{ij}^2}}{r_{ij}^2} \left(\frac{\vec{r}_{ij}}{r_{ij}}\right) \\ &\times \left\{ \int_0^{\infty} r' \phi_{\text{tot,i}}(r')\, e^{-ar'^2} \left[\left(1+2ar_{ij}^2\right)\sinh\left(2ar_{ij}r'\right) - 2ar_{ij}r'\cosh\left(2ar_{ij}r'\right)\right] dr' \right\}, \end{aligned} \tag{13}$$

while the reciprocal part is obtained as

$$\begin{aligned} \Pi_{\alpha\beta}^{\text{recip}} &= \frac{1}{2} \sum_{\vec{k}\neq 0}^{\infty} \frac{4\pi}{V^2\left(k^2 + \frac{1}{\lambda^2}\right)} e^{\frac{-k^2}{4}\left(\frac{1}{a}+\frac{1}{\alpha_{Ewald}}\right)} e^{i\vec{k}\cdot(\vec{r}_i - \vec{r}_j)} \\ &\times \left\{ \delta_{\alpha\beta} - 2\left[\frac{1}{4}\left(\frac{1}{a}+\frac{1}{\alpha}\right) + \frac{1}{\left(k^2+\frac{1}{\lambda^2}\right)}\right] k_\alpha k_\beta \right\}. \end{aligned} \tag{14}$$

Note that when solving the equations of motions for the ions the forces involved are not only those of real space but also the reciprocal space contribution from U_l. Their form reads

$$\vec{F}_{\mathrm{i},l} = \frac{1}{2}\sum_{\vec{k}\neq 0} \frac{8\pi Z_i}{V\left(k^2+\frac{1}{\lambda^2}\right)} e^{-\frac{-k^2}{4}\left(\frac{1}{\alpha}+\frac{1}{a_{\mathrm{Ewald}}}\right)} \times \left\{ \sin\left(\vec{k}\cdot\vec{r}_i\right)\left[\sum_j Z_j \cos\left(\vec{k}\cdot\vec{r}_j\right)\right] - \cos\left(\vec{k}\cdot\vec{r}_i\right)\left[\sum_j Z_j \sin\left(\vec{k}\cdot\vec{r}_j\right)\right]\right\}\vec{k}. \tag{15}$$

4. Results of the simulated OCP crystal

In this section we show the results obtained from our OCP MD simulations. In our NVT ensamble we use fixed N_I ions in a box of length $L = \left(\frac{N_I}{n_I}\right)^{1/3}$ for a given outer crust ion number density n_I at temperature T. Typically we set $T = 1$ MeV. In Fig. 1 we show two snapshots (planes XZ, XY) of particle positions for a run with $n_I = 0.017\mathrm{fm}^{-3}$, and $N_I = 50$ particles. We set $a = 2\ fm^{-2}$ and $a_{Ewald} = 10\ fm^2$. Initial random positions are shown in red while those after thermalization ($t = 10^4$ fm/c) are depicted in blue. As shown an ordered crystal forms after this time evolution.

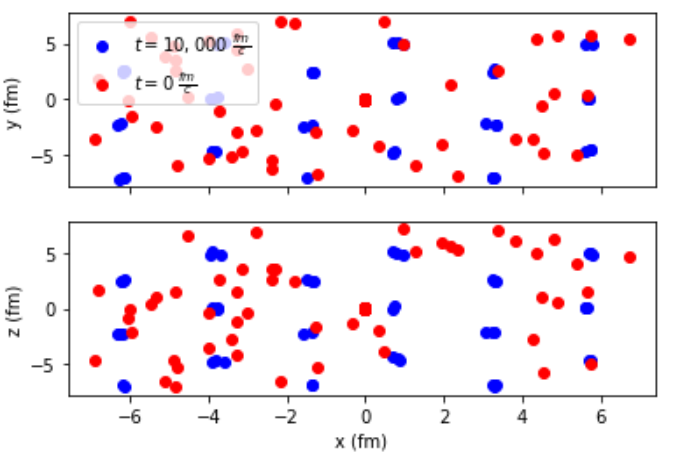


Fig. 1. Two dimensional snapshots of particle positions for a run with $n_I = 0.017\,\mathrm{fm}^{-3}$, and $N_I = 50$ particles with $A = 56$, $Z = 26$. Initial positions are shown in red while those after thermalization ($t = 10^4$ fm/c) are depicted in blue. As shown an ordered crystal forms after this time evolution.

In Fig. 2 we show the pair correlation function $g(r)$ versus distance in units of simulation box length L with the same conditions as those depicted in system from Fig. 1. At small distances the repulsive interaction leads the distribution, however

at larger distances the crystallized configurations arise as signaled by the different peaks. Normalization of $g(r)$ is such that $\int_0^\infty n_I g(r) 4\pi r^2 dr = N_I - 1$.

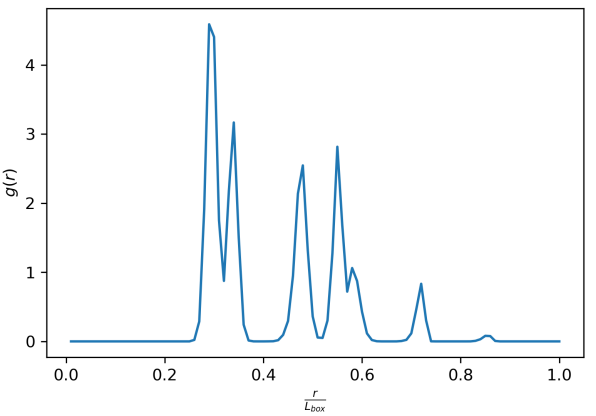


Fig. 2. Pair correlation function for $N_I = 50$ and same conditions as those in Fig. 1. Short range repulsion is shown along with the different peaks signaling the appearance of an ordered state with different neighbouring distances.

In Fig. 3 we show the energy per nucleon for a system of $N_I = 10$ ions with $A = 56$ at density $n_I = 0.0016$ fm^{-3}. Using the Ewald sum (we set $a = 2\ fm^{-2}$ and $a_{Ewald} = 10\ fm^{-2}$) we explore the convergence and actual value of energy per nucleon as the electron screening length changes. We use $\lambda = 10, 50, 100$ fm and we also test the Coulomb case where $\lambda \to \infty$. We show that the screened interaction in the infinite range converges to existing calculations using the raw Coulomb case.[8] In the depicted case the change can be nearly 50%, making the system less bound.

In Fig. 4 we show the pressure as a function of the mass density along with the non-diagonal stress tensor components. We use $a = 2\ fm^{-2}$ and $a_{Ewald} = 10\ fm^{-2}$ for a system of $N_I = 25$ ions with $Z = 26$, $A = 56$ at different densities. Pressure is obtained for the crystallized sytem as $P = \frac{1}{3} Tr \Pi_{\alpha\beta}$, but an additional electron pressure is present from the Fermi gas such that $P_e = \frac{\pi^3 \hbar^2}{15m} \left(\frac{3n_e}{\pi}\right)^{\frac{5}{3}}$ and $n_e = n_p$ in the charge neutral system. It is important to note that the true ground state of the system must provide a zero pressure crystal (as it does not exert external force per unit area) and this is fitted by considering a variable box length in our simulation. The off-diagonal components are zero as the crystal is not distorted from equilibrium (no shear).

We conclude that a careful evaluation of screening effects in the evaluation of the energies as well as forces is key to obtain the true binding of ion matter in the outer crust of NSs. We find that when the Ewald sum is implemented screened matter is less bound than the regular Coulomb lattice and this has an impact on forces

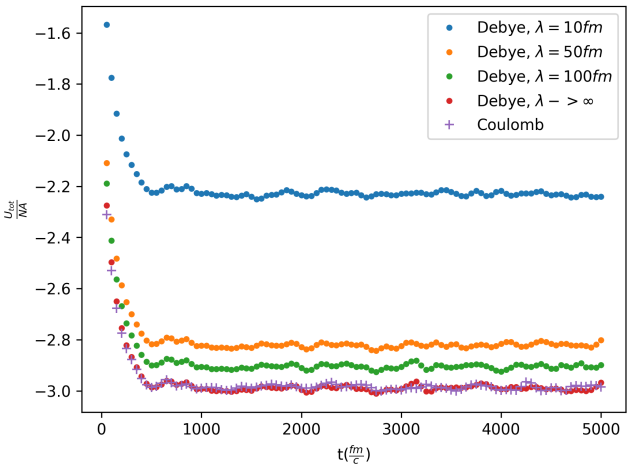


Fig. 3. Energy per nucleon as a function of simulation time for ions with $Z = 26$, $A = 56$, T=1 MeV and $n_I = 0.0016$ fm^{-3}. Colors represent simulations for different values of electron screening length λ, and for the Coulomb case. Ewald sums using Debye expressions in the limit $\lambda \to \infty$ provide results that converge to existing Coulomb calculations.[8]

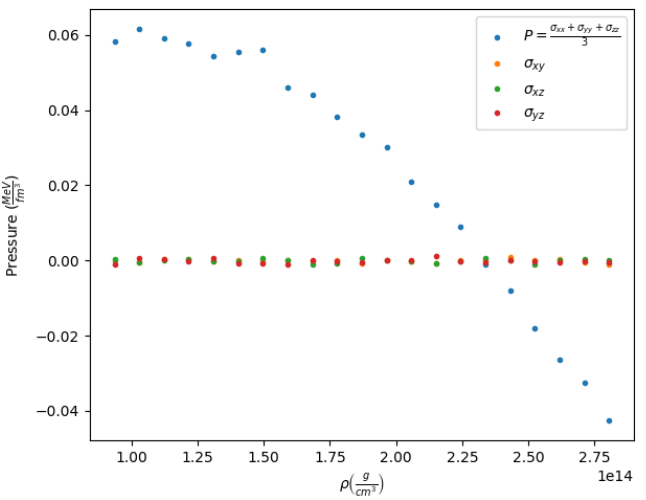


Fig. 4. Ion pressure and off-diagonal elements of the stress tensor as a function of mass density for ions with $Z = 26$, $A = 56$, T=1 MeV. It is seen that the total ion pressure is zero for a given value of the density. This is relevant to the definition of the true ground state of the system.

and thus on the mechanical properties derived from the stress tensor. In a future contribution we will detail how sizable this correction is with a complete analysis from density and temperature variation.

Acknowledgment

We acknowledge financial support from Junta de Castilla y León through the grant SA096P20, Agencia Estatal de Investigación through the grant PID2019-107778GB-100, Spanish Consolider MultiDark FPA2017-90566-REDC and PHAROS COST Actions MP1304 and CA16214. We thank A. Aguado for useful comments.

References

1. M. E. Caplan and C. J. Horowitz, Reviews of Modern Physics, 89, (2017).
2. Horowitz, C. J., M. A. Pérez-García, and J. Piekarewicz, *Phys. Rev. C* 69, 045804 (2004).
3. B. Haskell, A. Sedrakian, chapter in The Physics and Astrophysics of Neutron Stars, Astrophysics and Space Science Library, Springer, 457 (2018).
4. N. Chamel and P. Haensel, *Living Rev. Relativity*, **11**, (2008), 10.
5. J. M. Pearson, N. Chamel, A. Y. Potekhin, A. F. Fantina, C. Ducoin, A. K. Dutta and S. Goriely, *MNRAS* 481, 2994–3026 (2018).
6. A. F. Fantina et al., *Astronomy and Astrophysics* 633, A149 (2020).
7. A. Aguado, and P. A. Madden, *J. Chem. Phys.* 119, 7471 (2003).
8. G. Watanabe et al., *Phys. Rev. C* **68** 035806 (2003).

Tidal deformability as a probe of dark matter in neutron stars

D. Rafiei Karkevandi

ICRANet-Isfahan, Isfahan University of Technology, Isfahan, 84156-83111, Iran
E-mail: davood.rafiei64@gmail.com

S. Shakeri

Department of Physics, Isfahan University of Technology, ICRANet-Isfahan, Isfahan University of Technology, Isfahan, 84156-83111, Iran
E-mail: s.shakeri@iut.ac.ir

V. Sagun

CFisUC, Department of Physics, University of Coimbra, Coimbra, 3004-516, Portugal
E-mail: violetta.sagun@uc.pt

O. Ivanytskyi

Institute of Theoretical Physics, University of Wroclaw, Wroclaw, 50-204, Poland
E-mail: oleksii.ivanytskyi@uwr.edu.pl

The concept of boson stars (BSs) was first introduced by Kaup and Ruffini and Bonazzola in the 1960s. Following this idea, we investigate an effect of self-interacting asymmetric bosonic dark matter (DM) according to Colpi et al. model for BSs (1986) on different observable properties of neutron stars (NSs). In this paper, the bosonic DM and baryonic matter (BM) are mixed together and interact only through gravitational force. The presence of DM as a core of a compact star or as an extended halo around it is examined by applying different boson masses and DM fractions for a fixed coupling constant. The impact of DM core/halo formations on a DM admixed NS properties is probed through the maximum mass and tidal deformability of NSs. Thanks to the recent detection of Gravitational-Waves (GWs) and the latest X-ray observations, the DM admixed NS's features are compared to LIGO/Virgo and NICER results.

Keywords: Bosonic Dark matter, Complex scalar field, Neutron Star, Gravitational-Wave

1. Introduction

The evidence for the existence of dark matter (DM) which constitutes up to 85% of the matter in the universe, is implied from various astrophysical and cosmological observations. However, despite the enormous experimental efforts in the past decades, the nature of these particles remains elusive. In addition to various terrestrial experiments, compact astrophysical objects such as neutron stars (NSs) can be served as valuable natural detectors to constrain the properties of DM. The presence

of DM in NS interior depending on the various hypothesis of introducing them and their features, could have signficant effects on the properties of NSs.[1–8]

There are different scenarios assuming existence of DM in the NS interiors, which are mostly based on the DM accumulation during different stages of stellar evolution. The main of these stages are : a) progenitor, b) main sequence star, c) supernova explosion with formation of a proto-NS, and d) equilibrated NS.[4,9–13] As an alternative way, DM can be produced during the supernova explosion or NS merger leading to presence of DM in the NSs.[2,3] High level of DM fraction inside NS is reachable through other mechanisms such as (i) dark compact objects or DM clumps as the accretion center of baryonic matter (BM),[3,14,15] (ii) DM captured by NS in a binary system including Dark star and Dark star–NS merger,[7,15,16] (iii) NS can pass through a region in the Galaxy with extremely high DM density leading to accumulation of vast amount of DM.[16–19]

Two main and qualitatively different pictures leading to potentially observable impact of DM on the NS properties are a) Self-annihilating DM affecting luminosity, effective temperature and cooling process of NSs[20–25] and b) Asymmetric DM (ADM) with negligible annihilation rate caused by particle-antiparticle asymmetry in the dark sector.[26–29] We consider the second possibility allowing stable and massive DM particles to reside in a core of the NS. It was pointed out that the presence of DM particles in stellar cores can significantly decrease the mass of a compact object.[18,19,30,31] However, it was shown that light DM particles form an extended halo around the NS and can increase its gravitational mass.[2,4] It is worth mentioning that both of the aforementioned cases for combination of DM and BM within NS are known as DM admixed NS.

Regarding the ADM model, generally two methods have been utilized so far to extract the properties of the DM admixed NS from the Tolman-Oppenheimer-Volkof (TOV) equations.[32,33] 1) Single fluid formalism, for which an Equation of State (EoS) is considered for the whole star by inserting DM-BM interactions.[1,7,34–37] 2) Two-fluid formalism, for which DM and BM interact only through gravitational force, and two individual EoSs have to be considered for the DM and BM fluids.[15,16,38–41]

In this research, we apply a two-fluid formalism for the DM admixed NS. Bosonic DM is described by a complex scalar field with repulsive self-interaction. Historically, this model has been applied to describe hypothetical self-gravitating objects composed of bosons, so called boson stars. The idea of BS was first proposed by Kaup[42] and Ruffini-Bonazzola[43] for non-interacting bosons. The Heisenberg uncertainty principle was the only source of pressure of the BS matter resisting gravitational contraction. This leads to much lower maximum mass of BS compared to the Chandrasekhar mass. The pressure of the BS matter was significantly increased by introducing repulsive self-interaction proposed by Colpi et al.[44] Within this approach, stellar mass objects are supported by the DM particle mass about hundreds MeV and dimensionless coupling constant is of order of unity (see a comprehensive

review on BSs in[45–47]). Another component, i.e. BM, is modeled by the induced surface tension (IST) EoS. It was successfully applied to describe the nuclear matter, heavy-ion collision data and dense matter existing inside NS.[48–50]

With this set up we study effects of self-repulsive bosonic ADM on the mass-radius (M-R) profile and tidal deformability parameter[2,3,13,36,37,51–54] inferred from the GW signals related to post-merger stages of NSs.[55–59] Such a combined analysis based on the recent LIGO/Virgo results[60,60,61,61] opens a new possibility to study the internal structure of compact objects which may contain DM.

To be specific, in this work we consider a model of Colpi et al.[44] with sub-GeV DM particles of mass $m_\chi \sim \mathcal{O}(100$ MeV) and self-coupling constant $\lambda = \pi$. We analyze two key observational constraints of NSs, i.e. maximum mass and tidal deformability. The first of them is based on the NICER observation of the heaviest known pulsar PSR J0740+6620 with mass $2.072^{+0.067}_{-0.066} M_\odot$[62] and corresponds to requiring the maximal stellar mass to be at least $M_{max} = 2M_\odot$. The merger event GW170817[63] leads to the second constraint on the dimensionless tidal deformability $\Lambda \leq 580$ for $M = 1.4M_\odot$.[64] Using these constraints we probe DM admixed NSs at various masses and fractions of DM.

The rest of the paper is organized as follows. In Sec. 2 the DM and BM EoSs are described. In Sec. 3 we explain the two-fluid TOV formalism, DM halo and DM core formations and their impacts on mass-radius profile of DM admixed NSs. Sec. 4 is devoted to probing the effect of DM halo/core configurations on the tidal deformability. Our conclusions will be presented in Sec.5. We use units in which $\hbar = c = G = 1$.

2. Bosonic DM and BM models

2.1. *Dark Boson Star*

In the following we apply a model of complex scalar field as the bosonic DM with repulsive self-interaction potential, $V(\phi) = \frac{\lambda}{4}|\phi|^4$, minimally coupled to gravity and described by the action

$$S = \int d^4x\sqrt{-g}\left[\frac{M_{Pl}^2}{2}R - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi^* - \frac{1}{2}m_\chi^2|\phi|^2 - \frac{1}{4}\lambda|\phi|^4\right], \tag{1}$$

where m_χ is the boson mass, λ stands for the dimensionless coupling constant and M_{Pl} corresponds to the Planck mass.[44,65] In this setup a coherent scalar field is governed by both Klein-Gordon and Einstein equations which can potentially form Bose-Einstein Condensate (BEC) if the temperature is sufficiently low.[66,67] It has been assumed a spherically symmetric configuration for the scalar field $\phi(r,t) = \Phi(r)e^{i\omega t}$ and a static metric to rewrite Klein-Gordon-Einstein (K.G.E) equations to a set of ordinary differential equations.[44] This leads to the following EoS describing

a self-interacting and self-gravitating bosonic system so-called BS

$$P = \frac{m_\chi^4}{9\lambda}\left(\sqrt{1+\frac{3\lambda}{m_\chi^4}\rho}-1\right)^2 . \tag{2}$$

We recently presented an alternative derivation of this EoS in locally flat space-time by using the mean-field approximation (see appendix of[68]). This equation is valid in the parameter region for λ as

$$\lambda \gg 4\pi(m_\chi/M_{Pl})^2 = 8.43\times10^{-36}\left(\frac{m_\chi}{100\,\mathrm{MeV}}\right)^2 \tag{3}$$

In this limit which is called strong coupling regime, the system can be approximated as a prefect fluid and the anisotropy of pressure will be ignored.[45,69,70] Stellar mass BSs can be formed for $\lambda \sim \mathcal{O}(1)$ and $m_\chi \sim \mathcal{O}(100\ \mathrm{MeV})$,[45,71] in this section, we focus on this range of model parameters.

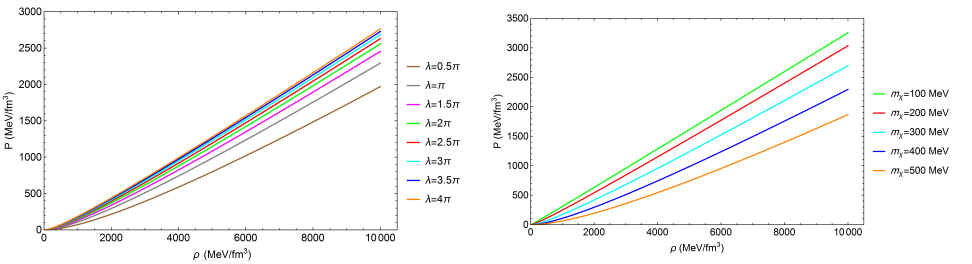


Fig. 1. Pressure as a function of density for bosonic matter obtained for $\lambda = \pi$ and various DM masses as labeled (right); for $m_\chi = 400$ MeV and different values of coupling constant (left).

Fig. 1 shows pressure of bosonic matter as a function of density for different masses and coupling constants as labeled. As it is shown in Fig. 1, (right panel) the pressure decreases with increasing of boson masses and (left panel) the pressure rises with the enhancement of the coupling constant or equivalently increasing the repulsive force between bosons.

In different density limits one can approximate Eq. (2) in a typical polytrope form $P = K\rho^\gamma$, where polytropic index at low density $\gamma \simeq 2$ and it smoothly reaches to $\gamma \simeq 1$ at high density. At low density regime, the bosonic DM EoS, Eq. (2), is reduced to

$$P \approx \frac{\lambda}{4m_\chi^4}\rho^2 . \tag{4}$$

However, for high density regime or correspondingly for very light bosons or high coupling constant, Eq. (2) reaches to radiation EoS with $P \approx \rho/3$. Similar equation to Eq. (4) has been obtained so far for a dilute self-interacting boson gas in a self-gravitating system (BS) known as Gross-Pitaevskii-Poisson (G.P.P) equation.[70,72,73] In fact, the G.P.P equation describes the BEC phase in a dilute

gas where only two-body mean field interaction is considered near zero temperature.[74–76]

In Fig. 2, we present the M-R diagrams of BSs obtained by solving TOV equation for Eq. (2). As it is indicated in the right panel, by decreasing the boson mass, the maximum gravitational mass of BSs increases and even goes above $2M_{\odot}$[62,77,78] and the corresponding radius goes well above typical NS radius. In the left panel, it is shown that higher self-coupling constant at fixed mass $m_{\chi} = 400$ MeV leads to higher maximum masses of BSs. Both the decreasing of boson mass and increasing of the coupling constant cause an enhancement in pressure of the system and consequently the rise of the maximum mass.

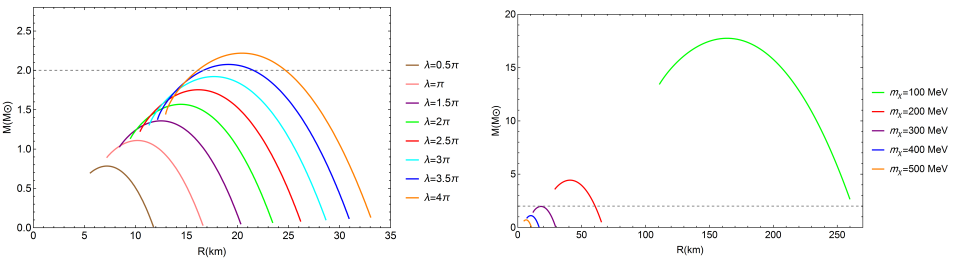


Fig. 2. M-R profile for BSs based on Eq. (2), (right panel) for the fixed value of coupling constant $\lambda = \pi$ and different boson masses. (Left panel) Calculations are made for fixed boson mass $m_{\chi} = 400$ MeV and different values of coupling constant as labeled at the figure. The gray dashed line shows the $2M_{\odot}$ limit.

Moreover, the variation of compactness $\mathcal{C} = M/R$ with respect to mass of BSs for different values of m_{χ} and λ is presented in Fig. 3. It shows the same dimensionless maximum compactness $\mathcal{C}_{(max)} \simeq 0.16$ for all cases. We see that the maximum compactness of a BS based on Eq. (2) is independent of free parameters of the model, namely m_{χ} and λ and for all the parameter space is well below the black hole formation limit $\mathcal{C} = 0.5$.[46,79]

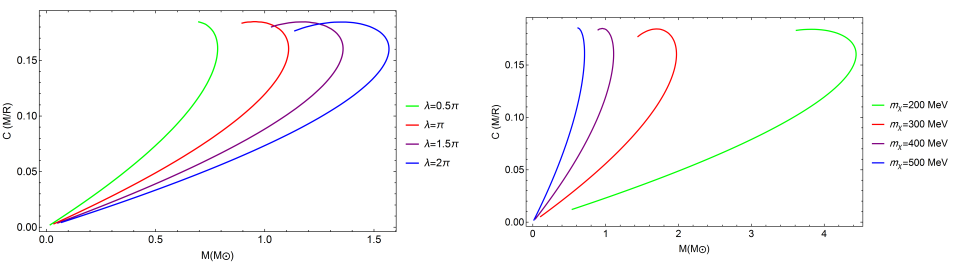


Fig. 3. Compactness of BSs as a function of their mass obtained for a fixed coupling constant $\lambda = \pi$ and different values of boson mass (right); fixed particle's mass $m_{\chi} = 400$ MeV and various λ values (left).

2.2. *Neutron Star*

For the baryon component (NS matter), we use the unified EoS with induced surface tension (IST) where both the short-range repulsion and long-range attraction between baryons have been taken into account.[49,50] The IST EoS reproduces the nuclear matter properties,[80] fulfills the proton flow constraint,[81] provides a high-quality description of hadron multiplicities created during the nuclear-nuclear collision experiments[82] as well as the matter inside compact stars.[48–50] The EoS is in a very good agreement with latest NS observations providing the maximum mass $M_{max} = 2.08 M_\odot$ and radius of the $1.4 M_\odot$ star equals to $R_{1.4} = 11.37$ km.[83] In our work the crust part of the NS's EoS is described via the polytropic EoS with $\gamma = 4/3$.[4] In Fig. 4, the change of pressure for a same density regime is plotted for BM and DM EoSs (left panel) and we show the M-R profiles of the NS and BS based on our considered EoSs (right panel).

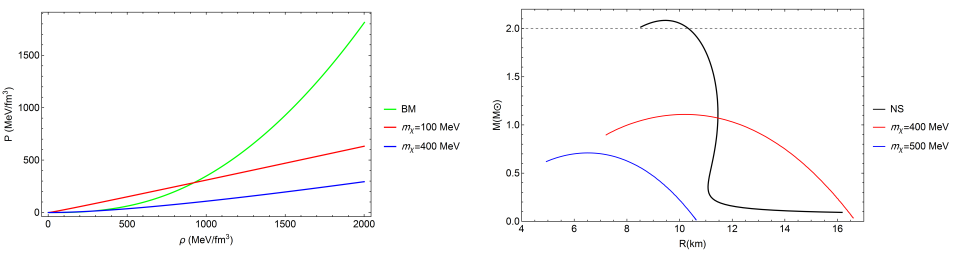


Fig. 4. Comparing BM and DM EoSs, left panel shows pressure vs. energy density for two different values of boson mass and $\lambda = \pi$. Right panel indicates M-R profiles for the NSs and BSs, considering $m_\chi = 400, 500$ MeV and the same coupling constant.

3. Two-fluid TOV equations and maximum mass

In order to study compact objects formed by the admixture of BM and DM that interact only through gravity, we use two-fluid TOV formalism[15,16] shown by Eqs. (5-6). Here $p = p_B + p_D$ and $M = M_T = \int_0^r 4\pi r^2 \epsilon_B(r) dr + \int_0^r 4\pi r^2 \epsilon_D(r) dr$. It can be seen that the total pressure and mass of the object have two contributions from BM and DM fluids shown by B and D indices. In order to solve two-fluid TOV equations

$$\frac{dp_{\rm B}}{dr} = -\left(p_{\rm B} + \epsilon_{\rm B}\right) \frac{M + 4\pi r^3 p}{r(r - 2M)}\,, \tag{5}$$

$$\frac{dp_{\rm D}}{dr} = -\left(p_{\rm D} + \epsilon_{\rm D}\right) \frac{M + 4\pi r^3 p}{r(r - 2M)}, \tag{6}$$

two central conditions related to both of the fluids have to be considered. By fixing two central pressures (p_B and p_D) together with the initial conditions at the center

of the star ($M_B(r \simeq 0) = M_D(r \simeq 0) \simeq 0$) the Eqs. (5-6) are numerically integrated up to the radius at which the pressure of one of the components vanishes. In principle this radius can be realized as DM radius R_D or BM radius R_B. In the former case the DM distributed only inside the core while BM extends to larger radius ($R_B > R_D$), then we set $p_D(r > R_D) = 0$ and continue the numerical integration to reach the visible radius of the star where $p_B(R_B) = 0$. When we have a BM core, DM can exist as an extended halo around the core with $R_D > R_B$, where $p_B(r > R_B) = 0$. It should be mentioned that for both DM core and DM halo cases, the core of the object is a mixture of DM and BM. Based on our extensive analysis there is another possibility of DM admixed NSs' configurations for which $R_B \approx R_D$ and DM distributed within the entire NS (see Fig. 5).

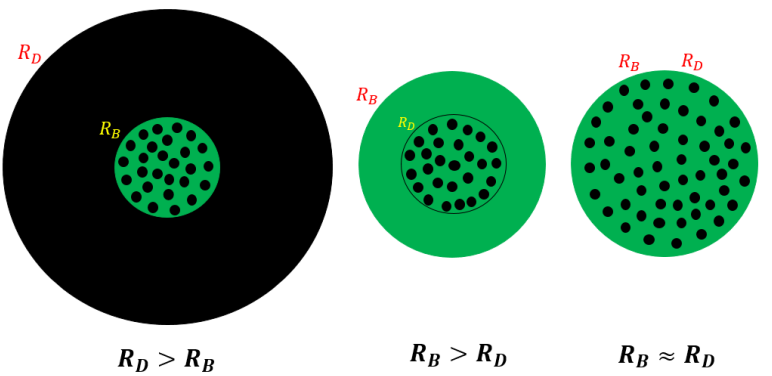


Fig. 5. Three possible configurations of a DM admixed NS, (left) DM halo, (middle) DM core and (right) DM is distributed in a whole NS. Note that for the DM core and halo cases the core of the object is a mixture of BM and DM. Green and black colors denote BM and DM, respectively.

For all of the possible DM admixed NSs' structures, the total gravitational mass of the mixed object is

$$M_T = M_B(R_B) + M_D(R_D). \tag{7}$$

However, the observable radius of the star is still defined by R_B, this is due to the visibility of R_B compare to R_D and technical difficulties of indirect detection of dark radius R_D. Furthermore, the DM fraction that determines the amount of DM in a DM admixed NS is defined as

$$F_\chi = \frac{M_D(R_D)}{M_T}. \tag{8}$$

Hereafter an effect of DM on NS properties is studied for m_χ of about hundreds MeV and and fixed coupling constant $\lambda = \pi$. Fig. 6 shows energy density profiles for a DM admixed NS where the BM (dashed red curves) and DM (solid red curves)

components are plotted separately. The energy density profiles for pure BM and DM stars are presented by solid black and green curves, respectively. Here we consider $\lambda = \pi$ and $F_\chi = 20\%$, while the central values of pressure for BM and DM components are chosen in such a way that a desired DM fraction F_χ has been obtained.

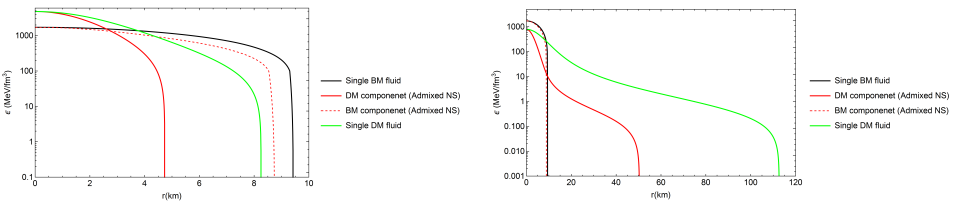


Fig. 6. Energy density profiles for pure BM and DM stars (black and green curves) shown together with the slitted DM and BM components of a DM admixed NS (solid and dashed red curves). Left panel corresponds to a DM core formation, while the right one to a DM halo, for $m_\chi = 400$ MeV and $m_\chi = 100$ MeV, respectively. For both of the cases, coupling constant is fixed at $\lambda = \pi$ and $F_\chi = 20\%$.

On the left panel which is obtained for $m_\chi = 400$ MeV, a DM core with $R_D \approx 5$ km is embedded in a BM structure with a larger radius. On the right panel, we fixed DM mass to $m_\chi = 100$ MeV which leads to the formation of a DM halo around the BM fluid with much larger radius. Interestingly, we see that for both DM core and DM halo formations, a reduction occurs in the energy density and the radius of DM and BM fluids in the mixed object compare to pure BM/DM star. This effect is much larger for the DM component and shows that the properties of the single DM fluid have significant effects in the admixed NS and in fact underly their features.

By comparing the left and right panels of Fig. 6, we see a transition from DM core to DM halo by changing m_χ from 400 MeV to 100 MeV. Therefore, by a thorough analysis of an effect of model parameters, as a general behaviour, we can conclude that light DM particles with $m_\chi < 200$ MeV, for low DM fractions, tend to form a halo around a NS, while heavier ones would mainly create a DM core inside a compact star (for more detailed analysis see[68]).

The M-R profiles for DM admixed NSs are shown in Fig. 7 in which $M = M_B + M_D$. Here R is the outermost radius of the star which is determined by R_B for the DM core and R_D for the DM halo. The solid black curve shows the M-R relation for the BM fluid (without DM), the gray dashed line indicates the $2M_\odot$ constraint on maximum mass of NSs and the shaded regions colored in magenta and cyan denote the causality and GR limits, respectively.

As it is shown in Fig. 7, two different boson masses, 100 MeV and 400 MeV lead to a DM halo and a DM core formations, respectively. In the left panel ($m_\chi = 400$ MeV), it is indicated that DM core formation causes a decrease of the maximum

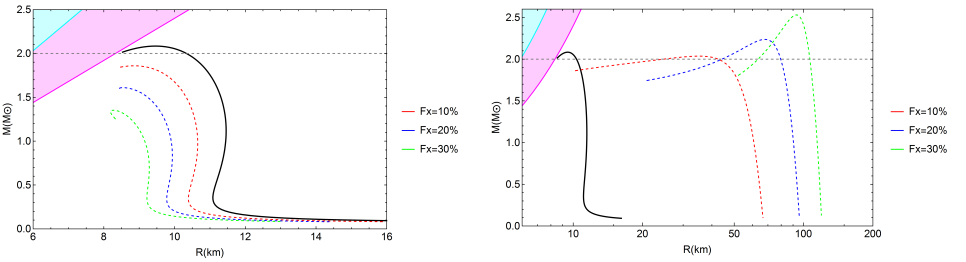


Fig. 7. Mass-Radius profiles for DM admixed NSs for $m_\chi = 400$ MeV (left) which corresponds to a DM core formation and $m_\chi = 100$ MeV (right) that represents an extended DM halo formation around a NS. Coupling constant is fixed to $\lambda = \pi$ and different F_χ are considered as labeled.

mass well below the $2M_\odot$ constraint,[62,77,78] and also a reduction of the corresponding radius. However, for $m_\chi = 100$ MeV (see the right panel in Fig. 7) for which a DM halo is formed around a NS, both maximum mass and radius are increased. Regarding the radius of the object for DM halo formation, it is increased significantly since is determined by R_D. It is seen that higher DM fractions enhance both of the above behaviours for DM core/halo structures. Note that for the cases in which a DM halo is formed ($R_D > R_B$), the visible radius of the star remains to be R_B.

To summarize, we see here that an effect of bosonic DM with repulsive self-interaction onto NSs is in agreement with previous studies which considered different DM models. Thus, an existence of a DM core decreases the maximum star's mass and the corresponding radius, while the formation of a DM halo increases these quantities.[2–4,30] In this regard, the most massive NS observed by NICER,[62] PSR J0740+6620 ($M_{max} \simeq 2M_\odot$) is compatible with the DM admixed NS scenario.

4. Tidal deformability of a DM admixed NS

GW signal from NS-NS mergers introduce tidal deformability as a new observable quantity to probe the internal structure of NSs and constrain their macroscopic features.[84–87] In this section, we analyse an impact of self-interacting bosonic DM on the tidal deformability of a DM admixed NS.

The idea of tidal deformability was first proposed by Tanja Hinderer in 2008[88,89] which comes from the fact that in a binary system of NSs both of the objects are deformed owing to the imposed tidal forces.[90–92] The tidal deformability expresses the ability of the gravitational field to change the quadrupole structure of a NS which alter the rotational phase of the binary system. Therefore, the GW signal is influenced during the inspiral phase due to the deformation effects of NSs when the binary orbital radius becomes comparable to the radius of NS. In fact, taking tidal deformability into account produces a phase shift in GW signal and accelerates the inspiral which leads to an earlier merging.[37,85,93]

The induced quadrupole moment Q_{ij} of a NS due to the external tidal field of its companion $\mathcal{E}_{ij}$ can be parameterized as follows[88,90]

$$Q_{ij} = \lambda_t \mathcal{E}_{ij} \,, \tag{9}$$

where λ_t is the tidal deformability parameter and can be defined based on k_2, the tidal love number, which is calculated from the system of equations including the TOV one. As is evident, k_2 and the tidal deformability strongly depend on the star's EoS.[88–90]

$$\lambda_t = \frac{2}{3} k_2 R^5 \tag{10}$$

Unlike λ_t which has dimension, dimensionless tidal deformability Λ can be defined as,

$$\Lambda = \frac{\lambda_t}{M^5} = \frac{2}{3} k_2 \left(\frac{R}{M}\right)^5 . \tag{11}$$

where R and M are the radius and mass of the compact star. It should be mentioned that R in a DM admixed NS is the outermost radius of the object which for a DM halo $R = R_D$ and for a DM core $R = R_B$. As an observational constraint on the tidal deformability, we take $\Lambda_{1.4} = 190^{+390}_{-120}$ reported by[64] for $M = 1.4 M\odot$ in the case of GW170817.

In the following, we investigate the effect of self-interacting bosonic DM, as a DM core or a DM halo, on the tidal deformability Λ of a mixed object at various m_χ and F_χ. In Figs. 8 and 9 the variation of Λ is shown in terms of total mass and radius of DM admixed NSs. In these figures the gray horizontal dashed lines indicate the LIGO/Virgo upper bound $\Lambda_{1.4} = 580$,[64] the gray solid vertical lines show $M_T = 1.4 M_\odot$ and the colored dashed and solid vertical lines stand for $R_{1.4}$ radius for the corresponding model parameters. The tidal deformability calculated for the pure baryonic EoS is denoted by the solid black curve and its $\Lambda_{1.4}$ value is about 285 which is well below the LIGO/Virgo constraint.

As a general behaviour in these plots, it can be seen that tidal deformability is a decreasing function of total mass and rises by increasing the radius which is related to the definition of this parameter as a function of R/M through Eq. (11). It follows from the M-R profile of a combined system NS+DM, that approaching the maximum mass of the equilibrium sequence decreases the stellar radius and, consequently, R/M. In other words the lowest value of Λ corresponds to the maximum mass and minimum radius of the DM admixed NS.

The effect of variation of Λ caused by changing the DM mass at fixed coupling constant $\lambda = \pi$ and DM fraction $F_\chi = 10\%$ is shown on Fig. 8. It is seen that for low DM masses $m_\chi = 100, 120, 150$ MeV, leading to formation of the DM halo, $\Lambda_{1.4}$ is higher than in the cases of higher m_χ and purely baryonic NS. Indeed, the corresponding $R_{1.4}$ significantly exceeds the values obtained for purely baryonic NS. For $m_\chi = 300, 400, 500$ MeV, however, the situation is different. Formation of the

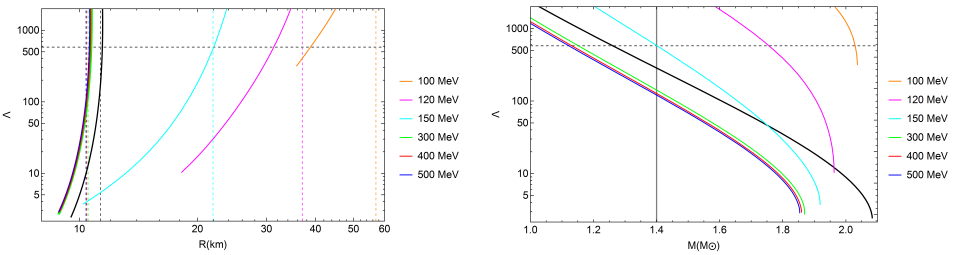


Fig. 8. Tidal deformability (Λ) in terms of total mass (right) and outermost radius (left) for stable sequences of DM admixed NSs. Various boson masses are considered, $m_\chi = 100, 120, 150$ MeV correspond to a DM halo formation while for $m_\chi = 300, 400, 500$ MeV a DM core is formed inside NS. Coupling constant and DM fraction are fixed at π and 10%, respectively.

DM core reduces the corresponding tidal polarizability, for which the $\Lambda - R$ curves are very similar to purely baryonic case. Regarding Fig. 8, we can conclude that DM halo yields large $\Lambda_{1.4}$, which even can exceed the observational constraint, while the DM core lowers Λ making it consistent with $\Lambda_{1.4} \leq 580$. This is related to the effect, which was mentioned in the previous section. Namely, DM halo increases the mass and radius of DM admixed NSs while DM core decreases these quantities. It is worth mentioning that GW observations during the inspiral phase of NS-NS coalescence correspond to lower frequencies detectable by Ad. LIGO. At this regime typical interstellar separation is $r < 150$ km. In order to prevent the technical difficulties caused by the overlap of DM halos we restrict their radii as $R_D \leq 75$ km.[2,3]

To give more insight, Fig. 9 shows modification of tidal polarizability due to variation of the DM fraction from 5% to 15% calculated at fixed $\lambda = \pi$ and $m_\chi = 100, 400$ MeV, corresponding to DM halo and DM core, respectively. As it is seen, higher F_χ increases $\Lambda_{1.4}$ and $R_{1.4}$ at $m_\chi = 100$ MeV and decreases theses parameters at $m_\chi = 400$ MeV. Remarkably, for $m_\chi = 100$ MeV (solid lines) tidal polarizability of the $M = 1.4 M\odot$ star exceeds 580 even at $F_\chi = 5\%$ since in this case Λ is more sensitive to DM fraction than at $m_\chi = 400$ MeV. Despite at this later case depicted by the dashed lines $\Lambda_{1.4}$ is in agreement with the upper observational constraint, the reduction of tidal deformability and $R_{1.4}$ should be consistent with the lower observational limits $\Lambda_{1.4} \gtrsim 70$ and $R_{1.4} \gtrsim 11$ km.[64,94–97]

As the final remark of this section, in Fig. 10, we show the effect of increasing and decreasing tidal deformability in a NS with DM halo/core and purely baryonic one. It was explained in the beginning of this section that tidal deformability parameter shows how much the compact object is deformed due to the gravitational potential of its companion. Thus in this illustration, we see that DM admixed NS with a DM halo can be more deformed since it has higher values of Λ compared to purely baryonic NS and the DM admixed one with the DM core. In addition, considering tidal love number k_2 we note that mixed compact objects with stiffer EoS are more deformable due to the DM halo compared to the ones with softer EoS producing the DM core.

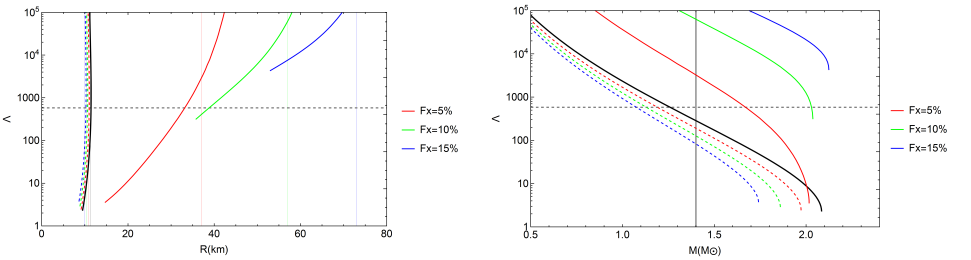


Fig. 9. Tidal deformability (Λ) in terms of total mass (right) and outermost radius (left) for stable sequences of DM admixed NSs. Two boson masses are considered, $m_\chi = 100$ MeV (solid lines) and $m_\chi = 400$ MeV (dashed lines) which correspond to DM halo and DM core formations, respectively. Various DM fractions are considered as labeled.

In summary, we note that in full agreement with the previous studies[2,3] DM halo increases tidal deformability, while DM core decreases it. Meanwhile, upper constraint on tidal deformability, $\Lambda_{1.4} = 580$, related to GW170817 event,[64] has been considered.

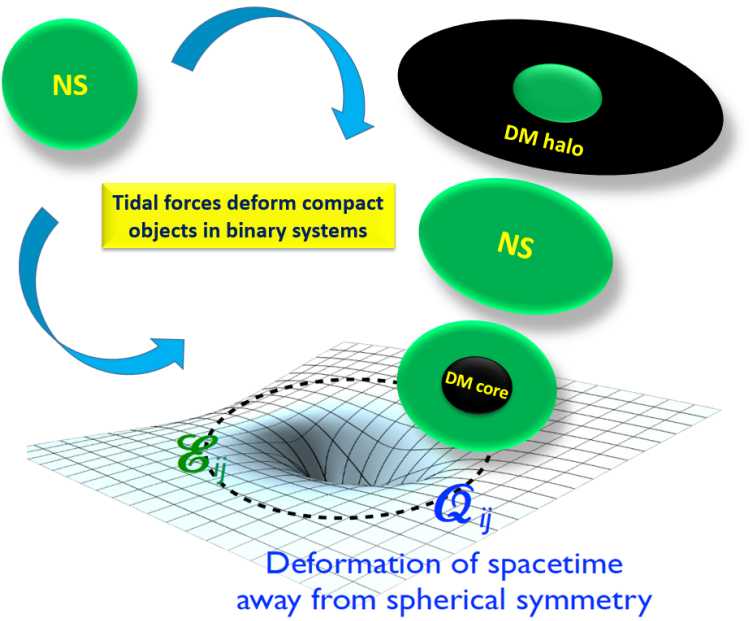


Fig. 10. The effect of increasing and decreasing of tidal deformability on DM admixed NS's deformation is compared with the pure BM object (NS). It is shown that higher value of Λ indicates more deformation in the compact object, therefore a DM halo deforms more in comparison with a pure NS and DM admixed NS with a DM core.

5. Conclusion and Outlook

Treating DM as a self-repulsing complex scalar field, various properties of single fluid BSs and two-fluid NSs has been studied within the TOV formalism. It is shown that for $\lambda = \pi$, light DM particles ($m_\chi \lesssim 200$ MeV) form BSs with much larger maximum mass and radius compared with typical NSs, while heavy DM particles lead to formation of BSs with much smaller radius and mass. Furthermore, we showed that for low DM fractions ($F_\chi < 20\%$), light bosons create a halo around the DM admixed NSs, while heavier DM particles form a DM core inside the BM component.

The effect of bosonic DM as a halo/core has been examined by considering the maximum mass, radius and tidal deformability of a DM admixed NS. We have indicated that DM halo formation causes an increase in the aforementioned observable quantities while a DM core reduces all of them. Considering various m_χ and F_χ, the maximum mass and tidal deformability of the mixed object has been compared to the latest upper observational bounds, $M_{max} = 2M\odot$ and $\Lambda_{1.4} \leq 580$ inferred from NICER (PSR J0740+6620) and LIGO/Virgo (GW170817) detections.

Regarding the impact of DM halo and DM core formations on NS's observable parameters and applying the observational limits for NSs' features, one could constrain the parameter space of DM model such as mass and coupling constant and also the amount of DM inside the compact object. In this regard, an extensive investigation has been done recently in[68] by the same authors of the present paper in which a constraint has been imposed on F_χ for sub-GeV DM particles by taking M_{max} and $\Lambda_{1.4}$ bounds. Moreover, as DM core decreases the visible radius of the DM admixed NS (R_B) and DM halo increases the invisible dark radius of the object (R_D), radius constraint for typical NSs ($M \approx 1.4M\odot$) and most massive ones ($M \approx 2M\odot$)[62,96,97] could be utilized to impose more stringent limits on DM parameter space and its fraction. In addition, any unusual observational results of NSs' properties could be explained by the DM admixed NS model. For instance, there are many effort among the community to explain the nature of the secondary compact object in the GW190814[98] event with the mass about $2.6M\odot$ being higher than the maximum NS one. There are some works explaining the mass of this strange object by the DM core or halo formation within the DM admixed NS scenario.[52,99,100] Regarding our model, as an example, $m_\chi = 50$ MeV, $\lambda = \pi$ and $F_\chi \approx 20\%$ lead to formation of a DM admixed NS with $M_T = 2.6M\odot$ and detectable radius about 10 km. As a final remark, upcoming modern facilities such as X-ray (NICER,[101] ATHENA,[102] eXTP[103] and STROBE-X[104]) and radio (MeerKAT,[105] ngVLA[106] and SKA[107]) telescopes, as well as GW (LIGO/Virgo/KAGRA[108] and Einstein[109,110]) detectors, shown in Fig. 11, would provide vast numbers of promising results for NSs' features bringing us to a golden age of NS investigations and consequently could help us to shed light on the nature of DM and its possible existence in compact objects.

Fig. 11. Applying various innovative telescopes covering all kinds of observations from GW and X-ray to radio waves provide a unique opportunity for compact objects' research which may solve the puzzle of DM.

Acknowledgments

V.S. acknowledges the support from the Fundação para a Ciência e Tecnologia (FCT) within the projects UID/04564/2021. The work of O.I. was supported by the Polish National Science Center under the grant No. 2019/33/B/ST9/03059.

References

1. G. Panotopoulos and I. Lopes, Dark matter effect on realistic equation of state in neutron stars, *Physical Review D* **96** (Oct 2017).
2. A. E. Nelson, S. Reddy and D. Zhou, Dark halos around neutron stars and gravitational waves, *Journal of Cosmology and Astroparticle Physics* **2019**, p. 012–012 (Jul 2019).
3. J. Ellis, G. Hütsi, K. Kannike, L. Marzola, M. Raidal and V. Vaskonen, Dark Matter Effects On Neutron Star Properties, *Phys. Rev. D* **97**, p. 123007 (2018).
4. O. Ivanytskyi, V. Sagun and I. Lopes, Neutron stars: New constraints on asymmetric dark matter, *Phys. Rev. D* **102**, p. 063028 (Sep 2020).
5. Z. Rezaei, Study of Dark-Matter Admixed Neutron Stars using the Equation of State from the Rotational Curves of Galaxies, *Astrophys. J.* **835**, p. 33 (2017).
6. A. de Lavallaz and M. Fairbairn, Neutron Stars as Dark Matter Probes, *Phys. Rev. D* **81**, p. 123521 (2010).
7. M. I. Gresham and K. M. Zurek, Asymmetric Dark Stars and Neutron Star Stability, *Phys. Rev. D* **99**, p. 083008 (2019).

8. Y. Dengler, J. Schaffner-Bielich and L. Tolos, The Second Love Number of Dark Compact Planets and Neutron Stars with Dark Matter, *arXiv e-prints*, p. arXiv:2111.06197 (November 2021).
9. J. F. Navarro, C. S. Frenk and S. D. M. White, The Structure of cold dark matter halos, *Astrophys. J.* **462**, 563 (1996).
10. R. Ruffini, C. R. Argüelles and J. A. Rueda, On the core-halo distribution of dark matter in galaxies, *Monthly Notices of the Royal Astronomical Society* **451**, p. 622–628 (May 2015).
11. C. Argüelles, A. Krut, J. Rueda and R. Ruffini, Novel constraints on fermionic dark matter from galactic observables i: The milky way, *Physics of the Dark Universe* **21**, p. 82–89 (Sep 2018).
12. A. Del Popolo, M. Le Delliou and M. Deliyergiyev, Neutron Stars and Dark Matter, *Universe* **6**, p. 222 (November 2020).
13. R. Ciancarella, F. Pannarale, A. Addazi and A. Marciano, Constraining mirror dark matter inside neutron stars, *Phys. Dark Univ.* **32**, p. 100796 (2021).
14. I. Goldman, R. Mohapatra, S. Nussinov, D. Rosenbaum and V. Teplitz, Possible implications of asymmetric fermionic dark matter for neutron stars, *Physics Letters B* **725**, p. 200–207 (Oct 2013).
15. P. Ciarcelluti and F. Sandin, Have neutron stars a dark matter core?, *Physics Letters B* **695**, p. 19–21 (Jan 2011).
16. F. Sandin and P. Ciarcelluti, Effects of mirror dark matter on neutron stars, *Astroparticle Physics* **32**, p. 278–284 (Dec 2009).
17. A. Del Popolo, M. Deliyergiyev and M. Le Delliou, Solution to the hyperon puzzle using dark matter, *Phys. Dark Univ.* **30**, p. 100622 (2020).
18. M. Deliyergiyev, A. Del Popolo, L. Tolos, M. Le Delliou, X. Lee and F. Burgio, Dark compact objects: An extensive overview, *Physical Review D* **99** (Mar 2019).
19. A. Li, F. Huang and R.-X. Xu, Too massive neutron stars: The role of dark matter?, *Astropart. Phys.* **37**, 70 (2012).
20. C. Kouvaris, WIMP Annihilation and Cooling of Neutron Stars, *Phys. Rev. D* **77**, p. 023006 (2008).
21. S. A. Bhat and A. Paul, Cooling of Dark-Matter Admixed Neutron Stars with density-dependent Equation of State, *Eur. Phys. J. C* **80**, p. 544 (2020).
22. J. Fuller and C. Ott, Dark Matter-induced Collapse of Neutron Stars: A Possible Link Between Fast Radio Bursts and the Missing Pulsar Problem, *Mon. Not. Roy. Astron. Soc.* **450**, L71 (2015).
23. J. F. Acevedo, J. Bramante and A. Goodman, Nuclear fusion inside dark matter, *Phys. Rev. D* **103**, p. 123022 (2021).
24. A. Sedrakian, Axion cooling of neutron stars, *Phys. Rev. D* **93**, p. 065044 (2016).
25. A. Sedrakian, Axion cooling of neutron stars. II. Beyond hadronic axions, *Phys. Rev. D* **99**, p. 043011 (2019).
26. D. E. Kaplan, M. A. Luty and K. M. Zurek, Asymmetric Dark Matter, *Phys. Rev. D* **79**, p. 115016 (2009).
27. J. Shelton and K. M. Zurek, Darkogenesis: A baryon asymmetry from the dark matter sector, *Phys. Rev. D* **82**, p. 123512 (2010).
28. K. Petraki and R. R. Volkas, Review of asymmetric dark matter, *Int. J. Mod. Phys. A* **28**, p. 1330028 (2013).
29. C. Kouvaris and N. G. Nielsen, Asymmetric Dark Matter Stars, *Phys. Rev. D* **92**, p. 063526 (2015).
30. S. Leung, M. Chu and L. Lin, Dark-matter admixed neutron stars, *Phys. Rev. D* **84**, p. 107301 (2011).

31. Q.-F. Xiang, W.-Z. Jiang, D.-R. Zhang and R.-Y. Yang, Effects of fermionic dark matter on properties of neutron stars, *Phys. Rev. C* **89**, p. 025803 (Feb 2014).
32. R. C. Tolman, Static solutions of Einstein's field equations for spheres of fluid, *Phys. Rev.* **55**, 364 (1939).
33. J. R. Oppenheimer and G. M. Volkoff, On Massive neutron cores, *Phys. Rev.* **55**, 374 (1939).
34. A. Das, T. Malik and A. C. Nayak, Confronting nuclear equation of state in the presence of dark matter using GW170817 observation in relativistic mean field theory approach, *Phys. Rev. D* **99**, p. 043016 (2019).
35. H. C. Das, A. Kumar, B. Kumar, S. Kumar Biswal, T. Nakatsukasa, A. Li and S. K. Patra, Effects of dark matter on the nuclear and neutron star matter, *Mon. Not. Roy. Astron. Soc.* **495**, 4893 (2020).
36. D. Sen and A. Guha, Implications of Feebly Interacting Dark Sector on Neutron Star Properties and Constraints from GW170817, *Mon. Not. Roy. Astron. Soc.* **504**, p. 3 (2021).
37. H. C. Das, A. Kumar and S. K. Patra, Effects of dark matter on the inspiral properties of the binary neutron star (4 2021).
38. S. Mukhopadhyay, D. Atta, K. Imam, D. Basu and C. Samanta, Compact bifluid hybrid stars: Hadronic Matter mixed with self-interacting fermionic Asymmetric Dark Matter, *Eur. Phys. J. C* **77**, p. 440 (2017), [Erratum: Eur. Phys. J. C 77, 553 (2017)].
39. X. Li, F. Wang and K. Cheng, Gravitational effects of condensate dark matter on compact stellar objects, *Journal of Cosmology and Astroparticle Physics* **2012**, p. 031–031 (Oct 2012).
40. X. Li, T. Harko and K. Cheng, Condensate dark matter stars, *JCAP* **06**, p. 001 (2012).
41. L. Tolos and J. Schaffner-Bielich, Dark compact planets, *Physical Review D* **92** (Dec 2015).
42. D. J. Kaup, Klein-Gordon Geon, *Phys. Rev.* **172**, 1331 (1968).
43. R. Ruffini and S. Bonazzola, Systems of selfgravitating particles in general relativity and the concept of an equation of state, *Phys. Rev.* **187**, 1767 (1969).
44. M. Colpi, S. Shapiro and I. Wasserman, Boson Stars: Gravitational Equilibria of Selfinteracting Scalar Fields, *Phys. Rev. Lett.* **57**, 2485 (1986).
45. F. E. Schunck and E. W. Mielke, General relativistic boson stars, *Class. Quant. Grav.* **20**, R301 (2003).
46. S. L. Liebling and C. Palenzuela, Dynamical Boson Stars, *Living Rev. Rel.* **20**, p. 5 (2017).
47. L. Visinelli, Boson Stars and Oscillatons: A Review (9 2021).
48. K. A. Bugaev, V. V. Sagun, A. I. Ivanytskyi, I. P. Yakimenko, E. G. Nikonov, A. V. Taranenko and G. M. Zinovjev, Going beyond the second virial coefficient in the hadron resonance gas model, *Nucl. Phys. A* **970**, 133 (February 2018).
49. V. V. Sagun, I. Lopes and A. I. Ivanytskyi, The induced surface tension contribution for the equation of state of neutron stars, *Astrophys. J.* **871**, p. 157 (2019).
50. V. Sagun, I. Lopes and A. Ivanytskyi, Neutron stars meet constraints from high and low energy nuclear physics, *Nucl. Phys. A* **982**, 883 (2019).
51. A. Quddus, G. Panotopoulos, B. Kumar, S. Ahmad and S. K. Patra, GW170817 constraints on the properties of a neutron star in the presence of WIMP dark matter, *J. Phys. G* **47**, p. 095202 (2020).
52. K. Zhang and F.-L. Lin, Constraint on hybrid stars with gravitational wave events, *Universe* **6**, p. 231 (2020).

53. A. Le Tiec and M. Casals, Spinning Black Holes Fall in Love, *Phys. Rev. Lett.* **126**, p. 131102 (2021).
54. A. Das, T. Malik and A. C. Nayak, Confronting nuclear equation of state in the presence of dark matter using gw170817 observation in relativistic mean field theory approach, *Physical Review D* **99** (Feb 2019).
55. J. Ellis, A. Hektor, G. Hütsi, K. Kannike, L. Marzola, M. Raidal and V. Vaskonen, Search for Dark Matter Effects on Gravitational Signals from Neutron Star Mergers, *Phys. Lett. B* **781**, 607 (2018).
56. M. Bezares, D. Viganò and C. Palenzuela, Gravitational wave signatures of dark matter cores in binary neutron star mergers by using numerical simulations, *Phys. Rev. D* **100**, p. 044049 (2019).
57. M. Bezares and C. Palenzuela, Gravitational Waves from Dark Boson Star binary mergers, *Class. Quant. Grav.* **35**, p. 234002 (2018).
58. C. Horowitz and S. Reddy, Gravitational Waves from Compact Dark Objects in Neutron Stars, *Phys. Rev. Lett.* **122**, p. 071102 (2019).
59. A. Bauswein, G. Guo, J.-H. Lien, Y.-H. Lin and M.-R. Wu, Compact Dark Objects in Neutron Star Mergers (12 2020).
60. B. P. Abbott *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (2017).
61. B. P. Abbott *et al.*, GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$, *Astrophys. J. Lett.* **892**, p. L3 (2020).
62. T. E. Riley *et al.*, A NICER View of the Massive Pulsar PSR J0740+6620 Informed by Radio Timing and XMM-Newton Spectroscopy, *Astrophys. J. Lett.* **918**, p. L27 (2021).
63. B. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya and et al., Gw170817: Observation of gravitational waves from a binary neutron star inspiral, *Physical Review Letters* **119** (Oct 2017).
64. B. P. Abbott *et al.*, GW170817: Measurements of neutron star radii and equation of state, *Phys. Rev. Lett.* **121**, p. 161101 (2018).
65. A. Maselli, P. Pnigouras, N. G. Nielsen, C. Kouvaris and K. D. Kokkotas, Dark stars: Gravitational and electromagnetic observables, *Phys. Rev. D* **96**, p. 023005 (2017).
66. A. Arbey, J. Lesgourgues and P. Salati, Galactic halos of fluid dark matter, *Phys. Rev. D* **68**, p. 023511 (Jul 2003).
67. A. Suárez, V. H. Robles and T. Matos, A Review on the Scalar Field/Bose-Einstein Condensate Dark Matter Model, *Astrophys. Space Sci. Proc.* **38**, 107 (2014).
68. D. Rafiei Karkevandi, S. Shakeri, V. Sagun and O. Ivanytskyi, Bosonic Dark Matter in Neutron Stars and its Effect on Gravitational Wave Signal, *arXiv e-prints*, p. arXiv:2109.03801 (September 2021).
69. E. W. Mielke and F. E. Schunck, Boson stars: Alternatives to primordial black holes?, *Nucl. Phys. B* **564**, 185 (2000).
70. P.-H. Chavanis and T. Harko, Bose-Einstein Condensate general relativistic stars, *Phys. Rev. D* **86**, p. 064011 (2012).
71. C. Pacilio, M. Vaglio, A. Maselli and P. Pani, Gravitational-wave detectors as particle-physics laboratories: Constraining scalar interactions with a coherent inspiral model of boson-star binaries, *Phys. Rev. D* **102**, p. 083002 (2020).
72. X. Li, T. Harko and K. Cheng, Condensate dark matter stars, *JCAP* **06**, p. 001 (2012).
73. P.-H. Chavanis, Mass-radius relation of self-gravitating Bose-Einstein condensates with a central black hole, *Eur. Phys. J. Plus* **134**, p. 352 (2019).

74. F. Dalfovo, S. Giorgini, L. P. Pitaevskii and S. Stringari, Theory of bose-einstein condensation in trapped gases, *Reviews of Modern Physics* **71**, p. 463–512 (Apr 1999).
75. J. Rogel-Salazar, The gross–pitaevskii equation and bose–einstein condensates, *European Journal of Physics* **34**, p. 247–257 (Jan 2013).
76. C. J. Pethick and H. Smith, *Bose-Einstein Condensation in Dilute Gases*, 2 edn. (Cambridge University Press, 2008).
77. J. Antoniadis *et al.*, A Massive Pulsar in a Compact Relativistic Binary, *Science* **340**, p. 6131 (2013).
78. H. T. Cromartie *et al.*, Relativistic Shapiro delay measurements of an extremely massive millisecond pulsar, *Nature Astron.* **4**, 72 (2019).
79. P. Amaro-Seoane, J. Barranco, A. Bernal and L. Rezzolla, Constraining scalar fields with stellar kinematics and collisional dark matter, *JCAP* **11**, p. 002 (2010).
80. V. V. Sagun, A. I. Ivanytskyi, K. A. Bugaev and I. N. Mishustin, The statistical multifragmentation model for liquid-gas phase transition with a compressible nuclear liquid, *Nucl. Phys. A* **924**, 24 (2014).
81. A. I. Ivanytskyi, K. A. Bugaev, V. V. Sagun, L. V. Bravina and E. E. Zabrodin, Influence of flow constraints on the properties of the critical endpoint of symmetric nuclear matter, *Phys. Rev. C* **97**, p. 064905 (2018).
82. V. V. Sagun, K. A. Bugaev, A. I. Ivanytskyi, I. P. Yakimenko, E. G. Nikonov, A. V. Taranenko, C. Greiner, D. B. Blaschke and G. M. Zinovjev, Hadron resonance gas model with induced surface tension, *European Physical Journal A* **54**, p. 100 (June 2018).
83. V. Sagun, G. Panotopoulos and I. Lopes, Asteroseismology: radial oscillations of neutron stars with realistic equation of state, *Phys. Rev. D* **101**, p. 063025 (2020).
84. C. Raithel, F. Özel and D. Psaltis, Tidal deformability from GW170817 as a direct probe of the neutron star radius, *Astrophys. J. Lett.* **857**, p. L23 (2018).
85. K. Chatziioannou, Neutron star tidal deformability and equation of state constraints, *Gen. Rel. Grav.* **52**, p. 109 (2020).
86. E. R. Most, L. R. Weih, L. Rezzolla and J. Schaffner-Bielich, New constraints on radii and tidal deformabilities of neutron stars from GW170817, *Phys. Rev. Lett.* **120**, p. 261103 (2018).
87. S. De, D. Finstad, J. M. Lattimer, D. A. Brown, E. Berger and C. M. Biwer, Tidal Deformabilities and Radii of Neutron Stars from the Observation of GW170817, *Phys. Rev. Lett.* **121**, p. 091102 (2018), [Erratum: Phys. Rev. Lett. 121, 259902 (2018)].
88. T. Hinderer, Tidal Love numbers of neutron stars, *Astrophys. J.* **677**, 1216 (2008).
89. T. Hinderer, B. D. Lackey, R. N. Lang and J. S. Read, Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral, *Phys. Rev. D* **81**, p. 123016 (2010).
90. S. Postnikov, M. Prakash and J. M. Lattimer, Tidal Love Numbers of Neutron and Self-Bound Quark Stars, *Phys. Rev. D* **82**, p. 024016 (2010).
91. T. Zhao and J. M. Lattimer, Tidal Deformabilities and Neutron Star Mergers, *Phys. Rev. D* **98**, p. 063020 (2018).
92. S. Han and A. W. Steiner, Tidal deformability with sharp phase transitions in (binary) neutron stars, *Phys. Rev. D* **99**, p. 083014 (2019).
93. B. D. Lackey and L. Wade, Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars, *Phys. Rev. D* **91**, p. 043002 (2015).
94. M. C. Miller *et al.*, PSR J0030+0451 Mass and Radius from *NICER* Data and Implications for the Properties of Neutron Star Matter, *Astrophys. J. Lett.* **887**, p. L24 (2019).

95. T. E. Riley *et al.*, A *NICER* View of PSR J0030+0451: Millisecond Pulsar Parameter Estimation, *Astrophys. J. Lett.* **887**, p. L21 (2019).
96. G. Raaijmakers, S. K. Greif, K. Hebeler, T. Hinderer, S. Nissanke, A. Schwenk, T. E. Riley, A. L. Watts, J. M. Lattimer and W. C. G. Ho, Constraints on the dense matter equation of state and neutron star properties from NICER's mass-radius estimate of PSR J0740+6620 and multimessenger observations (5 2021).
97. M. C. Miller *et al.*, The Radius of PSR J0740+6620 from NICER and XMM-Newton Data, *Astrophys. J. Lett.* **918**, p. L28 (2021).
98. R. Abbott *et al.*, GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, *Astrophys. J. Lett.* **896**, p. L44 (2020).
99. H. C. Das, A. Kumar and S. K. Patra, Dark matter admixed neutron star as a possible compact component in the GW190814 merger event, *Phys. Rev. D* **104**, p. 063028 (2021).
100. B. K. K. Lee, M.-c. Chu and L.-M. Lin, Can the GW190814 secondary component be a bosonic dark matter admixed compact star? (10 2021).
101. A. L. Watts, Constraining the neutron star equation of state using Pulse Profile Modeling, *AIP Conf. Proc.* **2127**, p. 020008 (2019).
102. R. Cassano *et al.*, SKA-Athena Synergy White Paper (7 2018).
103. J. J. M. in 't Zand *et al.*, Observatory science with eXTP, *Sci. China Phys. Mech. Astron.* **62**, p. 029506 (2019).
104. P. S. Ray *et al.*, STROBE-X: X-ray Timing and Spectroscopy on Dynamical Timescales from Microseconds to Years (3 2019).
105. M. Bailes *et al.*, MeerTime - the MeerKAT Key Science Program on Pulsar Timing, *PoS* **MeerKAT2016**, p. 011 (2018).
106. R. J. Selina, The next generation very large array: A technical overview `https://arxiv.org/ftp/arxiv/papers/1806/1806.08405.pdf` (2018).
107. A. Weltman *et al.*, Fundamental physics with the Square Kilometre Array, *Publ. Astron. Soc. Austral.* **37**, p. e002 (2020).
108. R. Abbott *et al.*, Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences, *Astrophys. J. Lett.* **915**, p. L5 (2021).
109. M. Punturo *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, *Class. Quant. Grav.* **27**, p. 194002 (2010).
110. M. Maggiore *et al.*, Science Case for the Einstein Telescope, *JCAP* **03**, p. 050 (2020).

Binary neutron star mergers with quark matter equations of state

Atul Kedia[1,*], Hee Il Kim[2], Grant Mathews[1] and In-Saeng Suh[1,3]

[1] *Center for Astrophysics, Department of Physics, University of Notre Dame, Indiana, USA*
[2] *Center for Quantum Spacetime, Sogang University, Seoul 04107, Korea*
[3] *Center for Research Computing, University of Notre Dame, Indiana, USA*
[*] *E-mail: akedia@nd.edu, atulkedia93@gmail.com*

With recent observations of gravitational wave signals from binary neutron star mergers (BNSM) by LIGO-Virgo-KAGRA (LVK) Collaboration and NICER, the nuclear equation of state (EoS) is becoming increasingly testable by analysis with numerical simulations. Numerous simulations currently exist exploring the EoS at different density regimes for the constituent neutron stars. In this paper we summarize the GR three-dimensional hydrodynamics based simulations of BNSMs for EoSs with a specific emphasis on quark matter EoS at the highest densities.

Keywords: Neutron star binaries; Gravitational waves; Nuclear equation of state.

1. Introduction

Neutron stars are an ideal laboratory in which to probe the properties of matter at very high density. The microphysics of nuclear interactions in a neutron star reflects in its large structural features like its mass and radius. This results in modulation of the evolution of their binaries. This amplification of subatomic physics makes probing the physics at scales 10^{-14} m plausible at neutron star size 10^4 m scales. In particular, neutron star binary systems provide a means to analyze the pressure of nuclear matter at all domains of nuclear densities. Indeed, the first detection of gravitational waves from the binary neutron star merger GW170817 by the LIGO-Virgo Collaboration and the pulsar PSR J0740 by NICER has provided fundamental new insights into the nature of dense neutron-star matter.[1,2] The detected gravitational wave signal depends upon the tidal distortion of the neutron stars as they approach merger. For example, in the LIGO analysis[1,3] the tidal polarizability[4,5] was deduced from post-Newtonian dynamics implying that the radius of the stars of 1.4 $M_\odot$ is in the range 10.5 km $\leq R \leq$ 13.3 km. This has placed tight constrains on the equation of state (EoS) for nuclear matter as the stars approach merger.

Here we summarize the work done in the regime of quantum chromodynamics (QCD) formed during the merged neutron star binaries. The detection of the gravitational radiation during the postmerger could be used as a sensitive probe of both the order of the quark-hadron phase transition and the properties of matter in the non-perturbative regime of QCD.

The prospect of the postmerger evolution being used to explore EoS issues has been proposed for some time.[6] Indeed, there have been several investigations into the effects of the formation of quark matter in the BNSM.[7–10] For the most part these

studies have considered effects of a first-order phase transition and the formation of a mixed quark-hadron phase. In this case the first-order transition can soften the EoS and hasten the formation of the black hole.

The neutron star-ringdown occurs in a frequency range 2-5 kHz and thus the strain strength for a binary at 50 Mpc is not easily accessible to aLIGO/aVirgo/KAGRA. However, the third generation GW observatories, the Einstein Telescope and the Cosmic Explorer, will have enhanced sensitivities in this frequency range and will be susceptible to observing postmerger evolution of a BNSM. Further, there is a suggestion in the literature (see Refs. 11, 12) of post merger energy output in gravitational radiation from the GW170817 event that appears to be an extended ringdown (see however Ref. [13].) In Ref. [12] it was hypothesized that such extended emission might result from spin down of a magnetar.

2. Equation of state

To describe the evolution of matter completely, the hydrodynamics equations require an additional constraint that relates the various state variables of the matter, i.e. pressure, density, electron fraction, chemical potentials, etc. in a neutron star.[14,15] Constraints on the equation of state (EoS) have been placed by aLIGO based upon the tidal polarizability deduced from the chirp associated with event GW170817.[1,3] Any realistic description of the equation of state (EoS) of matter formed in the merger of neutron stars must also include the consequences of a transition between hadronic matter and quark matter. As the merged system collapses to a black hole it unavoidably encounters all dense phases of matter, particularly the transition to quark matter.

It is worth noting that as the baryon density and chemical potential increase the QCD strong coupling constant α_s approaches unity and a nonperturbative approach to QCD is imperative. In particular, There is rich physics in this region of the quark-matter phase diagram including the generation of constituent quark masses, due to chiral symmetry breaking,[16] and quark pairing leading to color superconductivity.[17] The evolution of these effects until the asymptotic regime must be described during the collapse.

The transition from hadronic matter to quark matter is not yet fully understood.[18] However, people have proposed many interesting models to describe the transition from hadronic matter to quark matter at densities between 2-5 ρ_{nuc}.[19,20] Considering strongly interacting quark states for the crossover density region, people could have made the EoSs satisfying 2 times solar mass observational bound.[21] Also, many models having the first order phase transition features have been proposed to describe the transition.

3. Power spectral density of the gravitational waves

In addition to f_{max} which could be obtained from the instantaneous change of the phase of h, the power spectral density (PSD) of the strain, h, poses to reveal further

features about the star's EoS. Refs. [22, 10] shows the presence of high frequency spectral features in the f_1, f_2 and f_3, as defined in Ref. [22] modes for binaries with different EoS description. The positioning and strength of the frequencies depends on the nature of the EoS. One of the important aspects about the power spectra of the mergers is that the strength of the PSD at 2 kHz to 5 kHz frequencies is weak enough that the current LVK detectors are unable to discern them from the noise. It is also pointed out that the third generation gravitational wave interferometers, the Einstein Telescope and the Cosmic Explorer, would have higher sensitivities at these frequencies making them more likely to resolve the f-modes. Further, Ref. [22] notes the occurrence of 'quasiuniversal' relations relating f-modes with compactness and the maximum frequency of chirp with tidal deformability of a single stable NS for the given EoS. These relations can be explained partially by the stiffness of the equation of state, softer EoSs leading to higher f-modes. Recent studies of the BNSM with the phase transtion have examined whether the features of the phase transition could be imprinted in the f-mode frequencies, for example, the shift of the peak frequencies.[8,23] We are currently investigating the spectral features of the BNSM with a specific crossover EoS and have found some interesting postmerger behaviour showing elongated duration owning to the stiffened EoS.[18,19,24]

4. Conclusion

We have briefly summarized some recent studies on the BNSM with quark matter. An observation of the postmerger and the collapse in a BNSM could be used to determine the nature of the phase transition and the physics at the crossover densities. The next generation GW detectors will probe and unveil the physics of this dense matter physics with their highly improved power of detectabilities at higher GW frequencies.

Finally, noting a caveat that there have been no studies fully considering such as MHD and neutrino transport for the simulations, however, we expect that essential features of the non-perturbative characters of the QCD would be revealed and confirmed without taking these considerations into account.

References

1. B.P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. 119, 161101 (2017).
2. M. C. Miller et al 2021, *The Astrophysical Journal Letters*, 918 L28
3. B. P. Abbott, et al. "Tests of general relativity with GW170817." Phys. Rev. Lett. 123.1 (2019): 011102.
4. T. Damour, M. Soffel, and C. Xu, Phys. Rev. D 45, 1017 (1992).
5. E. E. Flanagan and T. Hinderer, Phys. Rev. D 77, 021502 (2008).
6. A. Bauswein, H.-T. Janka, K. Hebeler, and A. Schwenk, Phys. Rev. D 86, 063001 (2012).

7. E. R. Most, L. Jens Papenfort, V. Dexheimer, M. Hanauske, H. Stoecker, and L. Rezzolla, On the de-confinement phase transition in neutron-star mergers, Eur. Phys. J. A 56, 59 (2020)
8. L. R. Weih, M. Hanauske, and L. Rezzolla, Postmerger Gravitational-Wave Signatures of Phase Transitions in Binary Mergers, Phys. Rev. Lett. 124, 171103 (2020).
9. S. L. Liebling, C. Palenzuela, and L. Lehner, Effects of High Density Phase Transitions on Neutron Star Dynamics, Class. Quant. Grav. 38, 115007 (2021)
10. A. Prakash, D. Radice, D. Logoteta, et al., arXiv:2106.07885 (2021)
11. H. M. P van Putten and D. V. Massimo, MNRAS, 482, L46 (2019).
12. H. M. P van Putten, D. V. Massimo, and A. Levinson, Astrophys. J. Lett. 876, L2 (2019).
13. M. Oliver, D. Keitel, A. Miller, et al. MNRAS, 485, 843 (2019).
14. D. Radice, A. Perego, F. Zappa, and S. Bernuzzi Ap. J. Lett. 852.2 (2018): L29.
15. J. Olson, *et al.*, arXiv:1612.08992 (2016).
16. Hatsuda, T., and Kunihiro, T. 1994, PhR, 247, 221
17. M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, 2008, RvMP, 80, 1455.
18. Baym, G., Hatsuda, T., Kojo, T., et al. 2018, RPPh, 81, 056902
19. G. Baym, S. Furusawa, T. Hatsuda, T. Kojo, H. Togashi, ApJ, 885, 42 (2019)
20. E. Annala, E., T. Gorda, A. Kurkela, et al., Nature Physics, (2020), 16(9), 907-910.
21. P. B. Demorest, T. Pennucci, S. M. Ransom, et al. , Nature (London) 467, 1081 (2010).
22. K. Takami, L. Rezzolla, and L. Baiotti, Physical Review D 91.6 (2015): 064001
23. S. Blacker, NUF Bastian, A. Bauswein, et al., Physical Review D 102.12 (2020): 123023.
24. A. Kedia, H.I. Kim, I.S. Suh, G. Mathews, (2021) in preparation.

Probing dense matter physics with transiently-accreting neutron stars: The case of source MXB 1659-29

Melissa Mendes*, Andrew Cumming and Charles Gale

Department of Physics, McGill University, 3600 rue University, Montreal, QC, H3A 2T8, Canada
** E-mail: melissa.mendessilva@mail.mcgill.ca*

Farrukh J. Fattoyev

Department of Physics, Manhattan College, Riverdale, NY 10471, USA

Recent observational data on transiently-accreting neutron stars has unequivocally shown fast-cooling sources, such as in the case of neutron star MXB 1659-29. Previous calculations have estimated its total neutrino luminosity and heat capacity, as well as suggested that direct Urca reactions take place in 1% of the volume of the core. In this paper, we reproduce the inferred luminosity of this source with detailed models of equations of state (EOS) and nuclear pairing gaps. We show that three superfluidity gap models are inconsistent with data for all EOS and another three are disfavoured because of fine tuning arguments. We also calculate the total heat capacity for all constructed stars and show that independent observations of mass and luminosity could set constraints on the core superfluidity of a source as well as the density slope of the symmetry energy, (L). This is an important step towards defining a universal equation of state for neutron stars and therefore, towards a better understanding of the phase diagram of asymmetric matter at high densities.

Keywords: Symmetry energy; neutron star composition; direct Urca; transiently-accreting neutron stars

1. Introduction

The composition of the core of neutron stars is still unknown. Their extreme densities make it impossible for similar conditions to be reproduced experimentally in laboratories, and their strongly-coupled hadronic nature represents a serious challenge for first-principles calculations. Therefore, there is a lot of uncertainty when it comes to the particle content and matter organization in neutron star cores. Even in the simplest model of neutron stars made only of neutrons, protons, electrons and muons, there is no agreement on the star's proton fraction – which determines electron and muon fractions, by the charge neutrality requirement – or how to describe proton superconductivity and neutron superfluidity, both expected to happen in matter at high densities.[1] The proton fraction is important because it determines whether neutrino cooling occurs via modified Urca or direct Urca reactions, which differ by orders of magnitude in emissivity.

Although BCS theory with polarization corrections [2,3] can describe superfluidity and superconductivity in a low density medium, at high densities, in-medium effects

prevent an accurate description of neutron and proton pairing. This uncertainty forbids a precise determination of the critical temperatures as a function of density, that is, the gap amplitude and width of the pairing gap in neutron star cores. Nonetheless, a wide range of theoretical calculations of superfluidity at high densities exist, which have been summarized in the functional analytic form in Ref. 4. Its parameters mimic the effect of in-medium interactions, and thus define a range of possible models for nucleon pairing at high densities, each one with their specific amplitudes and widths.

These different gap models will predict different stellar cooling rates, as shown in Ref. 4, because both superfluidity and superconductivity suppress neutrino production through direct or modified Urca reactions.[5] Furthermore, Cooper pair formation produces pair-breaking-formation (PBF) cooling reactions, which enhance the star's cooling when its temperature is close to the superfluid critical temperature.[5] The suppression or enhancement of neutrino emission due to superfluidity depends on density, and therefore location within the neutron star core. By comparing these predictions with the star's observed luminosity, one can exclude or favor specific gap models.

Transiently-accreting neutron stars are ideal systems to perform these observations. These stars periodically accrete matter from a companion star, in cycles of accretion outbursts followed by periods of quiescence, in which accretion halts.[6,7] By observing X-ray emissions after the accretion period, the surface temperature of the star can be found. Combining that information with the mass accretion rate, one can obtain a consistent estimate of the star's neutrino luminosity and determine whether fast cooling reactions (such as direct Urca) or slow cooling reactions (such as modified Urca) are taking place in the neutron star's core.[7]

The source MXB 1659-29 is particularly interesting because it has shown more than one accretion-quiescence cycle, which allows for repeated measurements of accretion rates during outburst, luminosity and temperature during quiescence. Analysis of the energy balance between accretion-driven crustal heating in outburst and neutrino cooling during quiescence leads to the conclusion that this source is undergoing fast cooling processes in its core.[7,8] Furthermore, it has been suggested in Ref. 8 that measuring the decrease of external temperature of the neutron star MXB 1659-29 over an interval of ten years, combined with the observation of its neutrino luminosity, could set constraints on its total heat capacity, and therefore on its composition. This analysis would represent another avenue for finding, for example, the relative contribution of leptons to the total heat capacity of a specific source based on direct observations, which could set further constraints on its superfluidity.

Other transiently accreting sources such as KS 1731-260 have inferred neutrino cooling luminosities inconsistent with fast cooling, implying that they have slow cooling processes operating in their cores.[9] To explain this data, one should consider a combination of equation of state (EOS) describing a star's particle content, and a collection of nuclear pairing gap models that accommodate both fast and slow

cooling processes. In this paper, we create detailed realistic scenarios for MXB 1659-29, to determine how the observed neutrino cooling rate of that source constrains the input physics. We use gap models described in Ref. 4, and a family of relativistic mean-field (RMF) EOS[10] based on FSUGold2, first described in Ref. 11. Their particularities are discussed in section 2. In section 3, we investigate the agreement of the calculated luminosities with data and we verify the suggestion made in Ref. 8 that direct Urca processes occur in approximately 1% of the volume of the star's core. Finally, in section 4, we discuss which gap models accurately describe this scenario and compute predictions for the neutron star heat capacity in each case.

2. Formalism

To determine the particle composition of the core of neutron stars, we use a family of equations of state (EOS) that are developed from the RMF model FSUGold2.[11] The original FSUGold2 model was created to reproduce the ground-state properties of finite nuclei such as binding energy and charge radii, their monopole response, and the maximum observed neutron star mass. In particular, the original model predicts both a stiff symmetry energy and a soft equation of state for symmetric nuclear matter, that is, matter with equal number of protons and neutrons. The symmetry energy is an essential ingredient of the EOS that strongly impacts the structure, dynamics, and composition of neutron stars and has received considerable attention over the last decade .[12,13] Customarily, one expands the total energy per nucleon $E(\rho,\alpha)$ at zero temperature—where $\rho=\rho_{\rm n}+\rho_{\rm p}$ is the total baryon density and $\alpha=(\rho_{\rm n}-\rho_{\rm p})/\rho$ is the neutron-proton asymmetry parameter—around the energy of isospin symmetric matter with $\alpha=0$,

$$E(\rho,\alpha)=E_{\rm SNM}(\rho)+E_{\rm sym}(\rho)\cdot\alpha^2+\mathcal{O}\left(\alpha^4\right) \tag{1}$$

where $E_{\rm SNM}(\rho)\equiv E(\rho,0)$ is the energy per nucleon in symmetric nuclear matter (SNM) and $E_{\rm sym}(\rho)$ the symmetry energy, which represents a correction of second order in parameter α, to the symmetric limit. No odd powers of α appear in the expansion above because the nuclear force is assumed to be isospin symmetric. To characterize the behavior of both $E_{\rm SNM}(\rho)$ and $E_{\rm sym}(\rho)$ near the nuclear saturation density $\rho_{\rm sat}\approx 0.15$ fm^{-3}, one can further expand these quantities in a Taylor series,[14]

$$E_{\rm SNM}(\rho)=B+\frac{1}{2}Kx^2+\cdots \tag{2}$$

$$E_{\rm sym}(\rho)=J+Lx+\frac{1}{2}K_{\rm sym}x^2+\cdots \tag{3}$$

where $x=(\rho-\rho_{\rm sat})/3\rho_{\rm sat}$ is a dimensionless parameter that quantifies the deviations of the density from its value at saturation, B is the energy per nucleon and K is the incompressibility coefficient of SNM. Similarly, J and $K_{\rm sym}$ represent the symmetry energy and the incompressibility coefficient of symmetry energy at saturation density and serve as corrections to the binding energy and incompressibility.

Unlike in SNM, whose pressure vanishes at saturation density, the slope of the symmetry energy L, and consequently, pressure of pure neutron matter, do not vanish at $\rho_{\rm sat}$.

Since L is poorly constrained experimentally, we generated a family of the FSUGold2 models that are identical in their predictions for the EOS of SNM but vary in their predictions for the symmetry energy. It is well known that, at a sub-saturation density of $\rho \approx 0.1$ fm^{-3}, which represents an average value between the central and surface densities, the symmetry energy is well constrained by the binding energy of heavy nuclei with a significant neutron excess.[15–17] By fixing this value of the symmetry energy, we obtain a family of models with differing (J, L) values that are identical in their predictions for the ground state properties of finite nuclei but predict a range of the neutron skin thicknesses and neutron star radii consistent with the current experimental and observational data.[18–22]

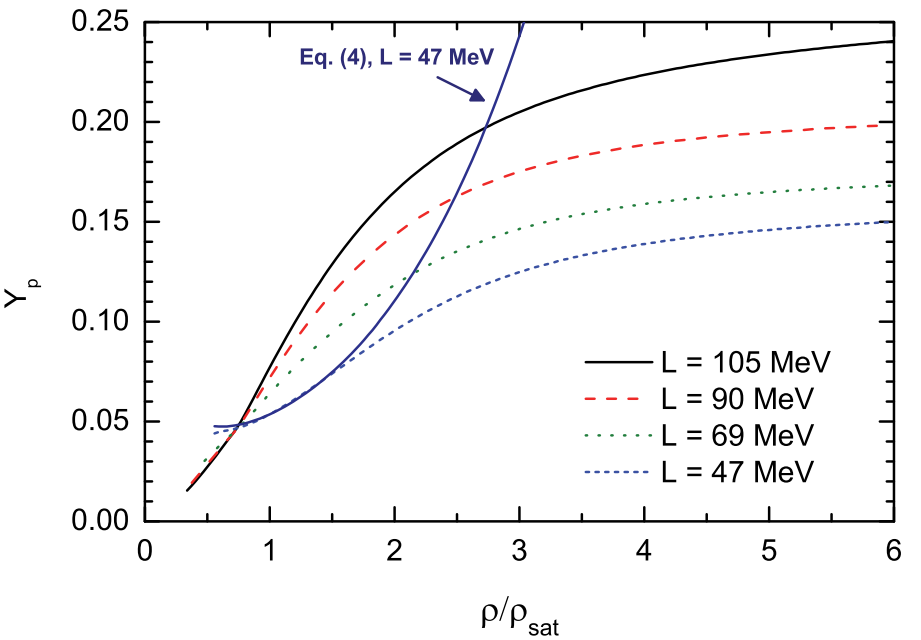


Fig. 1. Proton fraction $Y_{\rm p}$ versus baryon density ρ normalized by saturation density for EOS in the FSUGold2 family. Different curves correspond to EOS with different L values, in MeV. The blue solid curve shows the analytic approximation (Eq. 4).

Cold-catalyzed matter in neutron stars, ie. matter in the ground state with lowest energy per nucleon, is in a state of chemical equilibrium and is assumed to be electrically neutral. Therefore one must impose a charge neutrality and chemical equilibrium condition to obtain an EOS of neutron-star matter. We consider a

minimal model in which neutrons, protons, electrons and muons are present inside neutron stars. Increasing the parameter L leads to a larger symmetry energy at supersaturation densities. This in turn makes it difficult for protons to convert into neutrons and thus increases the proton fraction $Y_{\rm p} = \rho_{\rm p}/\rho$ in the innermost region of the star, as shown in Fig. 1. For matter composed of neutrons, protons and electrons, Ref. 23 shows this proportionality explicitly by writing the beta equilibrium condition as a function of the proton fraction, such that for low proton fraction,

$$Y_{\rm p} \simeq \frac{64}{3\pi^2\rho_{\rm sat}\,(3x+1)}\left(\frac{J+Lx+\frac{1}{2}K_{\rm sym}x^2}{\hbar c}\right)^3 . \tag{4}$$

As displayed in Fig. 1, this approximation is only valid up to about $1.5\rho_{\rm sat}$, where proton fractions are small. Here we used the predicted bulk parameters of symmetry energy at saturation density $\rho_{\rm sat} = 0.15$ fm^{-3}, $J = 30.6$ MeV, $L = 47.0$ MeV and $K_{\rm sym} = 54.0$ MeV to generate the blue solid curve in Fig. 1.

The outer crust is described by the EOS from Ref. 24, whereas the EOS for the inner crust is described in Ref. 25. For the core we use the FSUGold2 family of EOS with different values of the slope of the symmetry energy L. We assume a non-rotating spherically-symmetric neutron star, and solve the Tolman-Oppenheimer-Volkoff (TOV) equations

$$\frac{{\rm d}P}{{\rm d}r} = -\frac{\mathcal{E}(r)}{c^2}\frac{Gm(r)}{r^2}\left[1+\frac{P(r)}{\mathcal{E}(r)}\right]\left[1+\frac{4\pi r^3P(r)}{m(r)c^2}\right]\left[1-\frac{2Gm(r)}{c^2r}\right]^{-1} , \tag{5}$$

$$\frac{{\rm d}m}{{\rm d}r} = 4\pi r^2\frac{\mathcal{E}(r)}{c^2} , \tag{6}$$

$$\frac{{\rm d}\phi}{{\rm d}r} = -\frac{1}{\mathcal{E}(r)+P(r)}\frac{{\rm d}P}{{\rm d}r} \tag{7}$$

where $m(r)$ is the mass within radius r, $P(r)$ is the pressure, $\mathcal{E}(r)$ is the energy density and $\phi(r)$ is the gravitational potential such that at the surface of the star, $r = R$ and $m = M$, the pressure vanishes, $P(R) = 0$ and $\phi(R) = \frac{1}{2}\ln(1-2GM/c^2R)$.

Since our goal is to reproduce the inferred neutrino luminosity of MXB 1659-29, we consider the fast cooling process of direct Urca only. If there are no muons participating, direct Urca cooling takes place through the reactions

$$n \to p + e^- + \bar{\nu}_e, \quad p + e^- \to n + \nu_e. \tag{8}$$

This process conserves momentum only if

$$k_{Fn} \le k_{Fp} + k_{Fe}, \tag{9}$$

which implies that for direct Urca reactions the proton fraction must exceed a threshold value

$$Y_p \ge Y_{\rm p\,dUrca} = \frac{\left[\left(3\pi^2\hbar^3\rho Y_n\right)^{1/3} - \left(3\pi^2\hbar^3\rho_e\right)^{1/3}\right]^3}{3\pi^2\hbar^3\rho}, \tag{10}$$

as explained in Ref. 5. Here, $k_{\rm Fx}$ are the Fermi momenta, a function of ρ_x, the number density for each species and the particle fraction Y_x. If there are no muons in the star at all, Eq. (10) can be simplified because of the star's charge neutrality condition $\rho_p = \rho_e$, hence $Y_{\rm p\,dUrca} = 1/9 \approx 0.11$. However, for the EOS studied here, $Y_{\rm p\,dUrca} \neq 1/9$ because even when muons are not participating in direct Urca reactions, they are present, such that the charge neutrality condition becomes $\rho_p = \rho_e + \rho_\mu$, thus, for our set of EOS, we have $0.131 \leq Y_{\rm p\,dUrca} \leq 0.138$.

The direct Urca neutrino luminosity is given by

$$L_{\nu_{\rm dUrca}} = \int_0^{R_{\rm core}} \frac{4\pi r^2 \epsilon_0^{\rm dUrca\ total}}{\left(1-(2Gm(r)/c^2 r)\right)^{1/2}} dr, \tag{11}$$

where the integral is over the neutron star core, and the local neutrino emissivity is[5]

$$\epsilon_0^{\rm dUrca\ total} = \epsilon_0^{\rm dUrca\ e^-} + \epsilon_0^{\rm dUrca\ \mu^-} \tag{12}$$

$$\epsilon_0^{\rm dUrca\ e^-} = \frac{457\pi}{10080} G_{\rm F}^2 \cos^2\theta_{\rm C} \left(1+3g_{\rm A}^2\right) \frac{m_n^* m_{\rm p}^* m_e^*}{h^{10}c^3} \left(k_{\rm B}T\right)^6 \Theta_{\rm npe}, \tag{13}$$

where we used the weak coupling constant $G_{\rm F} = 1.43610^{-62}{\rm Jm}^3$, the Cabibbo angle $\theta_{\rm C} = 0.227$, the axial vector coupling constant $g_{\rm A} = -1.2601\left(1 - \frac{\rho}{4.15(\rho_0+\rho)}\right)$. $\rm M_x^*$ represents the effective mass of species x and $\Theta_{\rm npe}$ is a step function specifying the densities where direct Urca reactions can happen, respecting momentum conservation. Note that, in general, muons can participate in these reactions at high densities, so that

$$n \to p + \mu^- + \bar{\nu}_{\mu^-}, \quad p + \mu^- \to n + \nu_{\mu^-} \tag{14}$$

Their emissivities are

$$\epsilon_0^{\rm dUrca\ \mu^-} = \epsilon_0^{\rm dUrca\ e^-}. \tag{15}$$

Including superfluidity and superconductivity in the neutron star core model changes the neutrino luminosity calculations. The formation of Cooper pairs reduces the number of neutrons and/or protons available for participating in direct Urca reactions, therefore the rate $\epsilon^{\rm dUrca}$ is exponentially reduced, as described in Ref. 5, according to

$$\epsilon^{\rm dUrca} = \epsilon_0^{\rm dUrca} R \tag{16}$$

$$R = \exp\left(-v^*_{\rm triplet/singlet}\right) \tag{17}$$

$$\text{for } v^*_{\rm triplet} = 2.376\left(\rm T_c/T\right), \text{where } \rm T_c = 0.1187\Delta(k_{Fx}) \tag{18}$$

$$\text{and } v^*_{\rm singlet} = 1.764\left(\rm T_c/T\right), \text{where } \rm T_c = 0.5669\Delta(k_{Fx}). \tag{19}$$

Here T_c is the critical temperature, calculated according to each gap model parametrization, and T is the local temperature of the core, $T = \tilde{T}\exp(-\phi(r))$, where $\tilde{T}$ is the temperature of the isothermal core as seen at infinity. The heat

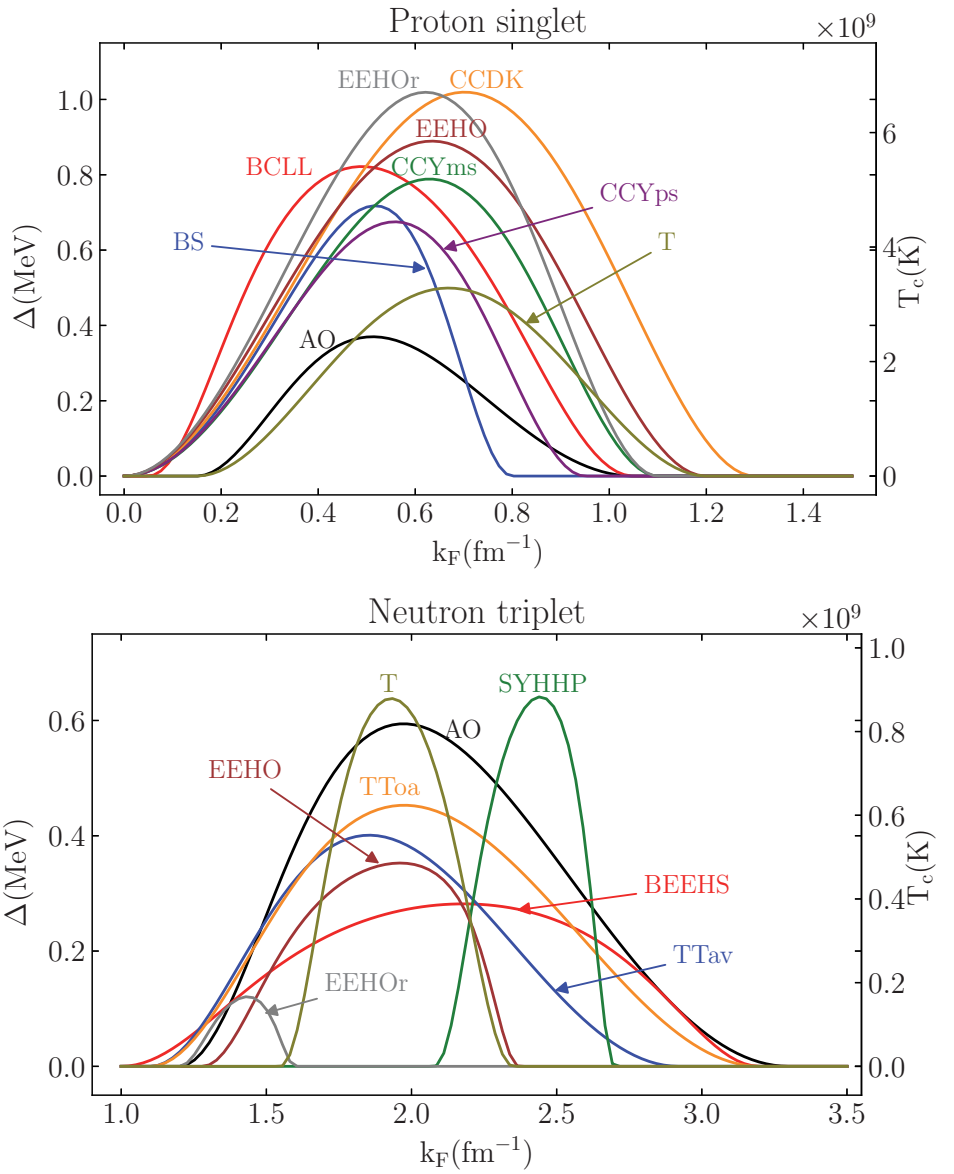


Fig. 2. Gap model parametrizations Δ (MeV) as a function of Fermi momenta k_F (fm^{-1}) for all models studied here (see Ref. 4 and references therein). (a) Proton singlet superconductivity (b) Neutron triplet superfluidity.

capacity of neutrons and protons is similarly reduced when they are superfluid or superconducting, $C_{p,n}^{\text{superfluid}} = C_{p,n}\ R$.

Equations (17)–(19) are simplifications of the results of full calculations, that find the proper quantum state rearrangements in the energy levels of the matter once

neutron and proton pairing are taken into account.[5] We acknowledge the difference in the reduction rates between the full calculation and the approximations (Eq. (17)–(19)), especially at $T \ll T_c$ and when superconductivity and superfluidity are simultaneously present in the star's core. This difference becomes large only well-below the critical temperature T_c, where the neutrinos are already suppressed, thus our results are not expected to change qualitatively. Results of full calculations will be explored in a forthcoming paper.

To characterize the different gap models, we work with the polynomial parametrizations for the superfluid gap[4]

$$\Delta\left(k_{\mathrm{Fx}}\right) = \Delta_0 \frac{\left(k_{\mathrm{Fx}} - k_0\right)^2}{\left(k_{\mathrm{Fx}} - k_0\right)^2 + k_1} \frac{\left(k_{\mathrm{Fx}} - k_2\right)^2}{\left(k_{\mathrm{Fx}} - k_2\right)^2 + k_3}, \tag{20}$$

where Δ_0, k_0, k_1, k_2 and k_3 are free parameters fitted to adjust the amplitudes and widths of each gap model. The resulting superfluid gaps are shown in Fig. 2 and their parameters can be found in Ref. 4.

3. Results

3.1. *Direct Urca cooling and the dependence on gap model*

We first calculated neutron star models without the effects of pairing, to determine when direct Urca reactions are kinematically allowed, as a function of the neutron star mass M and slope of the symmetry energy L. The direct Urca threshold for all EOS is shown in Fig. 3. All stars with masses above the $\mathrm{M_{dUrca}}$ curve can have part or the entirety of their core emitting neutrinos through direct Urca reactions. For low L EOS, only stars with large masses ($M \geq 1.8 M_\odot$) have a proton fraction large enough to allow direct Urca. On the right hand side of the plot, high L EOS, which have a larger proton fraction, can accommodate direct Urca reactions even for very low mass neutron stars ($M \leq 1.0 M_\odot$). Note that we include neutron stars with $M < 1.0 M_\odot$ to show the full parameter space, although it is not expected that such low mass neutron stars are formed in reality.

The range of parameter space in which direct Urca cooling can occur may change once we include superfluidity, as a consequence of the exponential suppression of direct Urca rates in the presence of Cooper pairs. Depending on the range of density over which superfluidity occurs, there can be different effects on the direct Urca parameter space. For example, shown in Fig. 4, the gap model NT_{T} predicts that for $L \geq 60\,\mathrm{MeV}$, part of the core of all neutron stars with masses $M \leq (1.7\text{–}1.8)\, M_\odot$ will be superfluid, thus reducing their total neutrino luminosity. Therefore, owing to the combination of the direct Urca threshold and the suppression by superfluidity, in this case only high mass stars $M \geq (1.7\text{–}1.8)\, M_\odot$ can significantly cool through direct Urca reactions, even for large values of L. In all figures where superfluidity gap models are taken into consideration, we also plot their closing curves, which are obtained by finding the densities for which the predicted gap model $T_c(\rho)$ matches

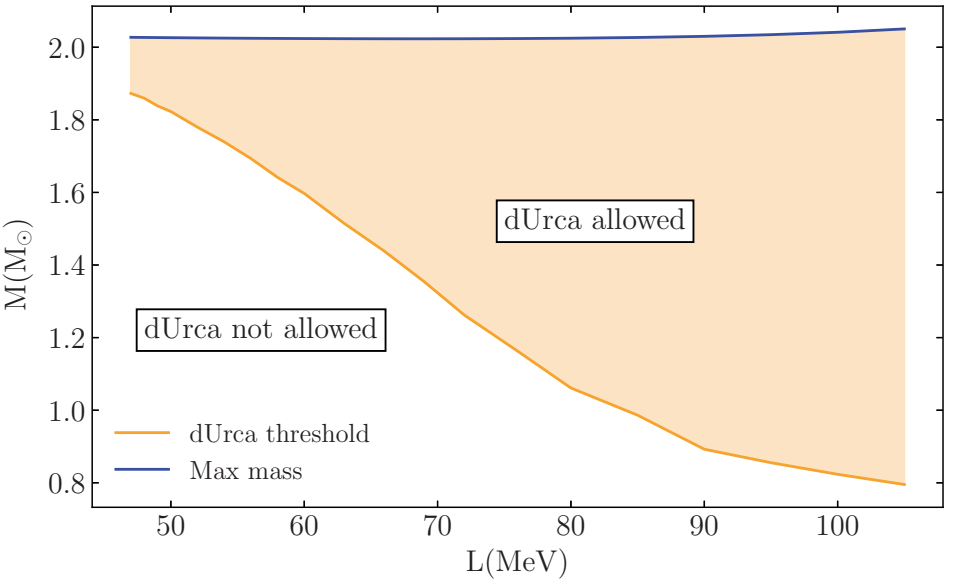


Fig. 3. Direct Urca threshold calculation (orange curve) for all EOS without nuclear pairing. The shaded area corresponds to the phase space where direct Urca reactions are allowed, the maximum mass reached for each EOS is also shown (blue curve).

the star's core temperature T, for $\tilde{T} = 2.5 \times 10^7$K, the inferred core temperature for MXB 1659-29.[8]

Another interesting situation is when there is a late opening of the superfluid gap, as in the case of $NT_{\rm SYHHP}$, Fig. 5. In this situation, for large L ($L \geq 70\,$MeV) and low mass stars ($M \leq 1.3M_\odot - 1.4M_\odot$), direct Urca reactions are allowed to happen because there is no superfluidity in the core yet. Cooper pair formation happens for larger mass stars or lower L, when direct Urca is suppressed, until there is no more superfluidity, that is, the gap closes, at higher masses. Therefore, direct Urca reactions are allowed again, close to maximum mass stars. In this situation, there are two regions of the phase diagram where direct Urca processes are significant: large L and low mass stars or very high mass stars ($M \approx 2.0M_\odot$) for all L.

There are also cases in which the superfluid part of the core is in a region where direct Urca reactions do not happen. Then, superfluidity does not change the calculations of neutrino luminosity shown in Fig. 3. This is the case for gap models $\rm NT_{EEHOr}$, $\rm PS_{BS}$ and $\rm PS_{CCYps}$. The last possible situation is represented by gap models $\rm NT_{AO}$, $\rm NT_{BEEHS}$ and $\rm NT_{TToa}$, which predict that the whole core of the neutron star is superfluid, severely suppressing direct Urca for all EOS and all masses.

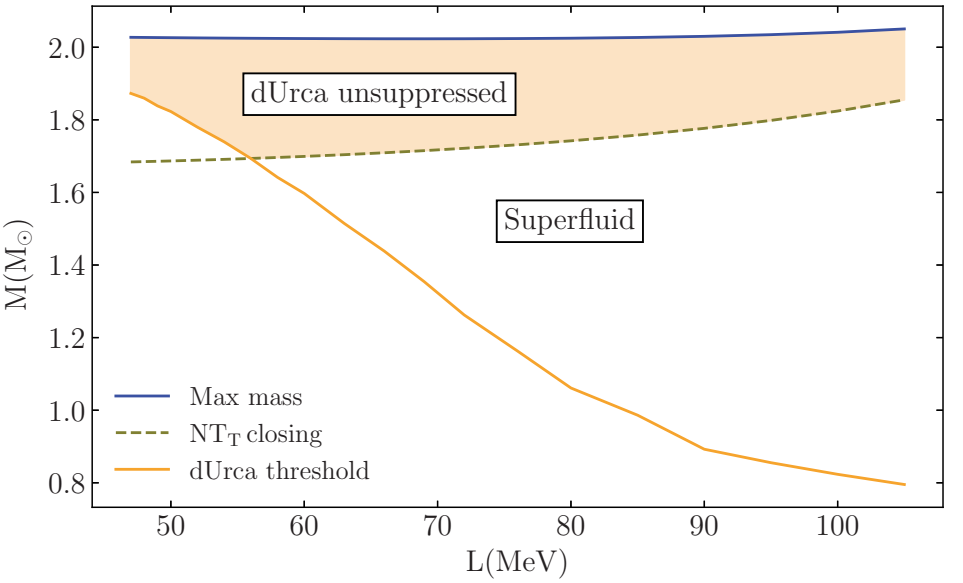


Fig. 4. Direct Urca threshold calculation (orange curve) and closing curve of superfluid gap model NT_T (dashed olive). The area between the orange and dashed olive lines corresponds to exponential suppression of direct Urca reactions. Above the dashed line, there is no suppression.

3.2. *Application to MXB 1659-29*

We now discuss the predictions of our model for the accreting neutron star MXB 1659-29. As described in Ref. 8, knowing the total mass accreted onto the neutron star during an outburst, one can estimate the total energy deposited in the crust. Assuming this energy is almost completely conducted into the core, and, after the core reaches thermal equilibrium, radiated away by neutrinos, we get an estimate of the star's neutrino luminosity. For the source MXB 1659-29, this luminosity is estimated to be $L_\nu \approx 3 \times 10^{27}$ J/s $= 3 \times 10^{34}$ erg/s.

This estimate was shown[8] to be insensitive (less than a factor of two) to variations in accretion rate during outburst and the outburst recurrence time, neutron star mass or EOS, for a core temperature of $\tilde{T} = 2.5 \times 10^7$ K and a light element envelope composition. As detailed in Ref. 9, the envelope composition is a crucial parameter in modelling the relationship between the internal temperature of the star and its surface temperature. For this source, a heavy element envelope would correspond to a core temperature of $\tilde{T} = 5 \times 10^7$ K, which increases the star's predicted total neutrino luminosity by two orders of magnitude, however, this scenario is inconsistent with the measured accretion rates. Ref. 8 also found that the envelope's impurity parameter and distance uncertainties can change the inferred luminosity above by an additional factor of $\lesssim 2$. For this study, we will take the value $L_\nu = 3 \times 10^{34}$ erg/s as the correct value of neutrino luminosity.

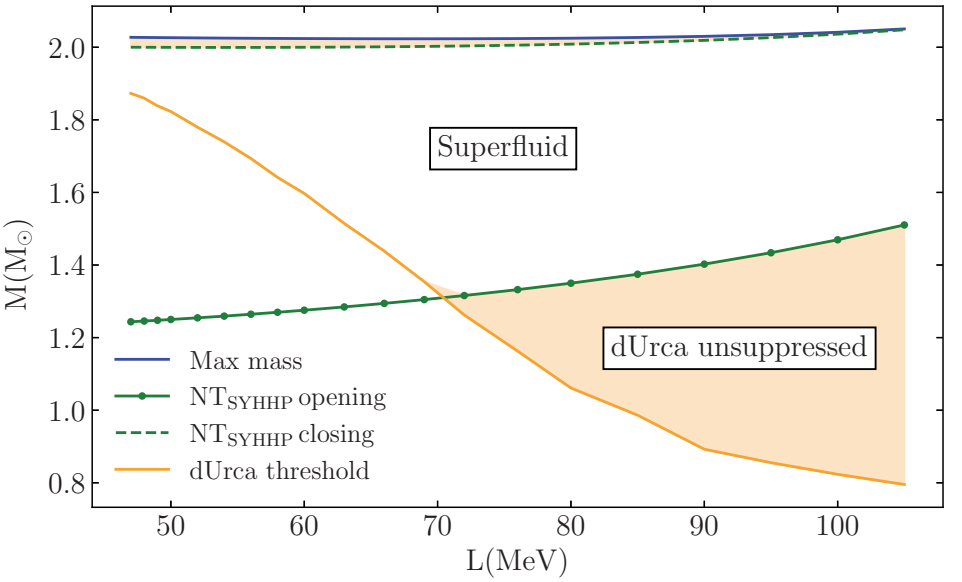


Fig. 5. The direct Urca threshold (orange curve), and the opening (dotted green) and closing (dashed green) curves of superfluid gap $\mathrm{NT_{SYHHP}}$. The area bounded by the orange, the dotted green and the dashed green lines corresponds to exponential suppression of direct Urca reactions. In all highlighted areas, there is no suppression.

To attempt to match the inferred neutrino luminosity of the source MXB 1659-29, we calculate the masses of stars with that total luminosity, for all combinations of EOS and gap models. The results are shown in Fig. 6. Panel (a) displays only the no superfluidity case, whereas panel (b) contains the predictions for neutron triplet superfluidity and panel (c), for proton singlet superconductivity. In all cases, we observe that only a small fraction of the volume of the core, from 0.10% to 1.07% depending on the EOS, is involved in unsuppressed direct Urca cooling. These volume fractions are shown in the bottom panel of panel (a).

We also calculated the total neutron star heat capacity, for all combinations of EOS and gap models. The results are shown in Fig. 7. Note that larger mass stars have larger heat capacities, which is a trend one observes in the figure above and in Fig. 8. Specifically, for proton superconductivity, high L EOS have lower mass stars achieving the observed luminosity of $L_\nu = 3 \times 10^{34}$ erg/s, thus Fig. 7a shows a general trend of lower heat capacities for higher Ls. The opposite trend happens for neutron superfluidity, where most heat capacity curves increase with increasing L, reflecting that higher mass stars are produced in that region. The C-shaped curves on the right bottom of Fig. 8b are a consequence of this phenomenon. Note that different L EOS can generate stars with the same mass and luminosity, for a given gap model. The fact that we are not comparing same size stars on Fig. 7 explains

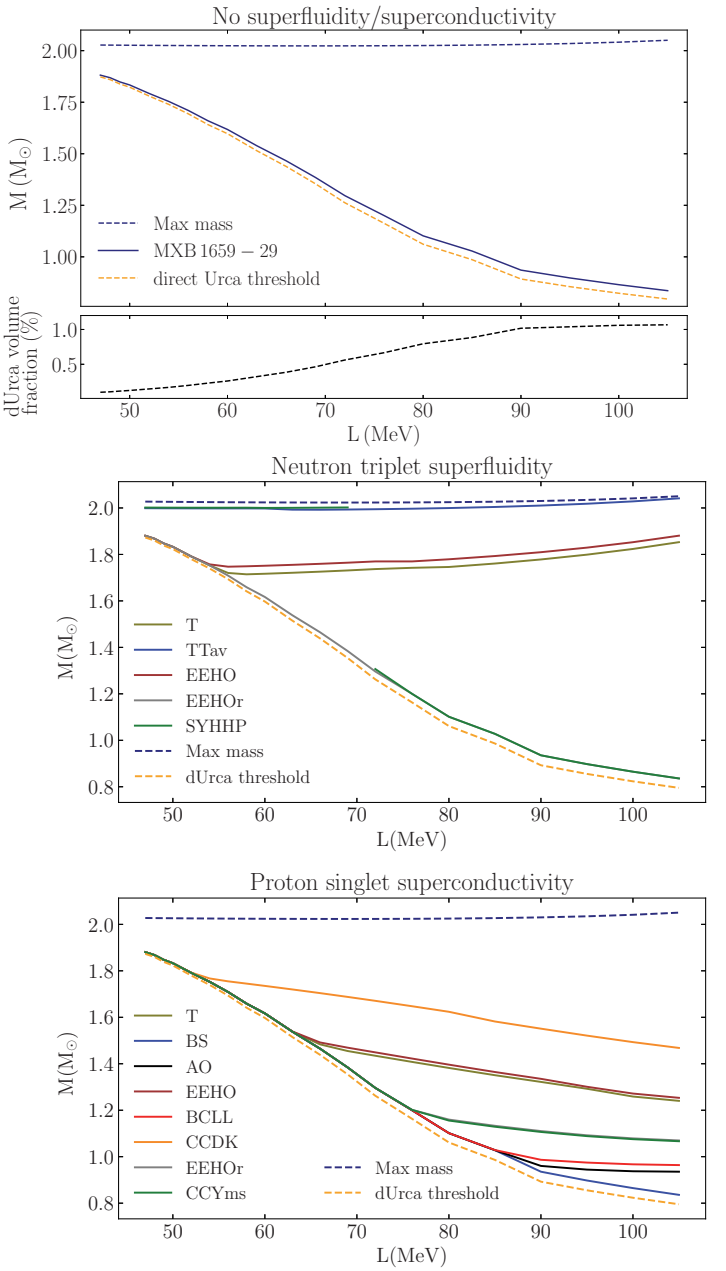
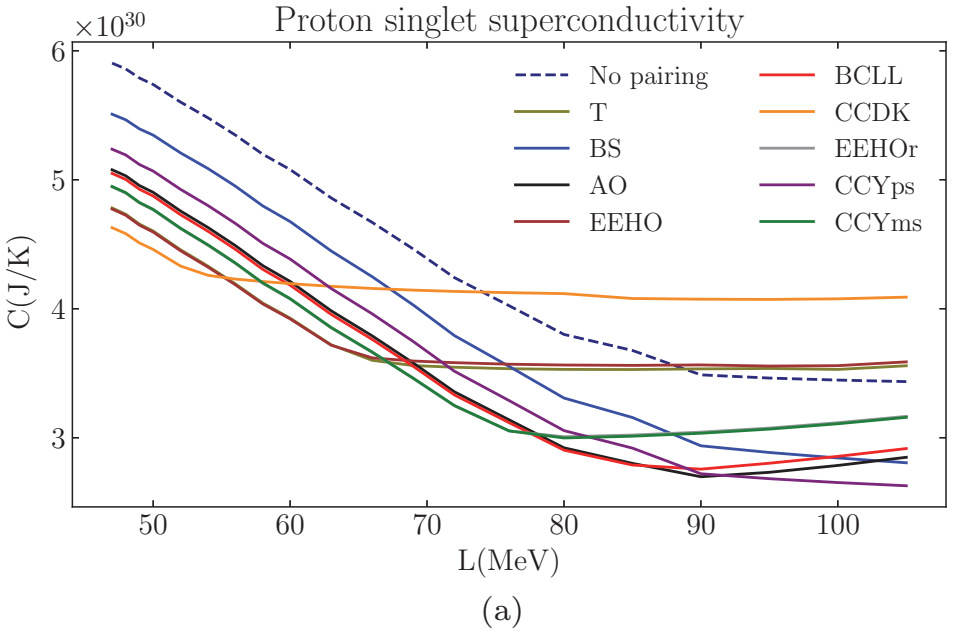


Fig. 6. Masses of stars with luminosity $L_\nu = 3{\cdot}10^{27}$ J/s, corresponding to MXB 1659-29 source, for all EOS. (a) No superfluidity/superconductivity (dark blue curve). The bottom panel displays the percentage of the core volume involved in unsupressed direct Urca reactions. (b) For all gap models of neutron triplet superfluidity. The curve displayed in panel a is under the gap model $\mathrm{NT_{EEHOOr}}$ curve. (c) For all gap models of proton singlet superconductivity. Gap models $\mathrm{PS_{CCYps}}$ and $\mathrm{PS_{BS}}$ predict the same curve.



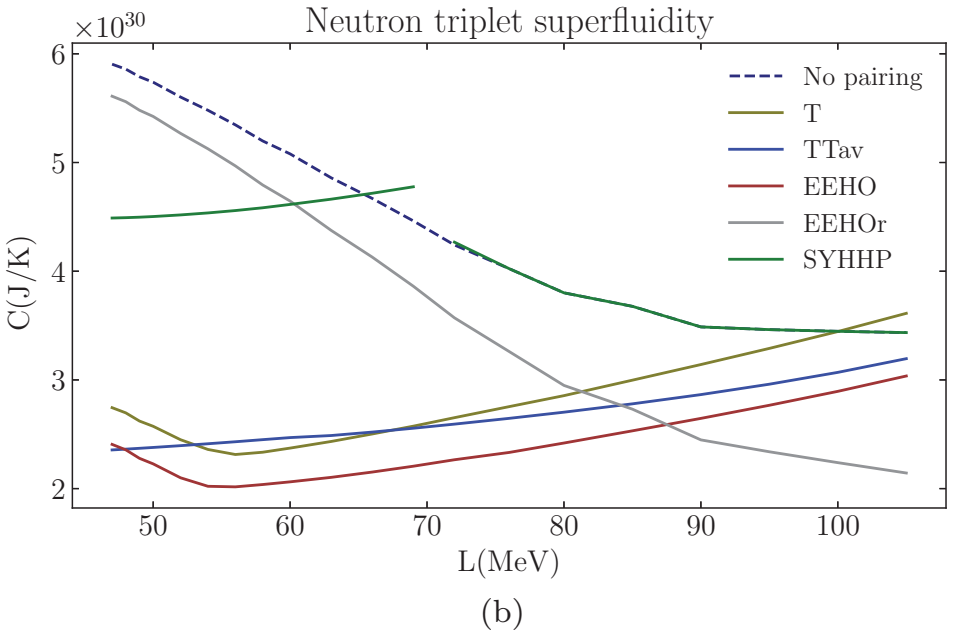
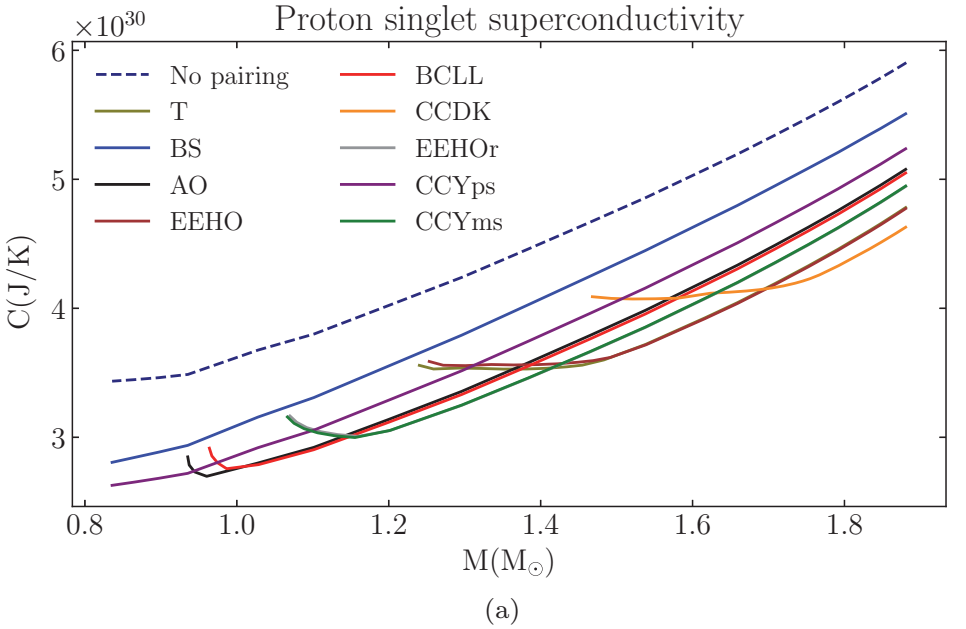


Fig. 7. Total heat capacity C (J/K) versus L (MeV) for stars with luminosity $L_\nu = 3 \cdot 10^{27}$ J/s for all EOS studied here. (a) Proton singlet superconductivity (b) Neutron triplet superfluidity.



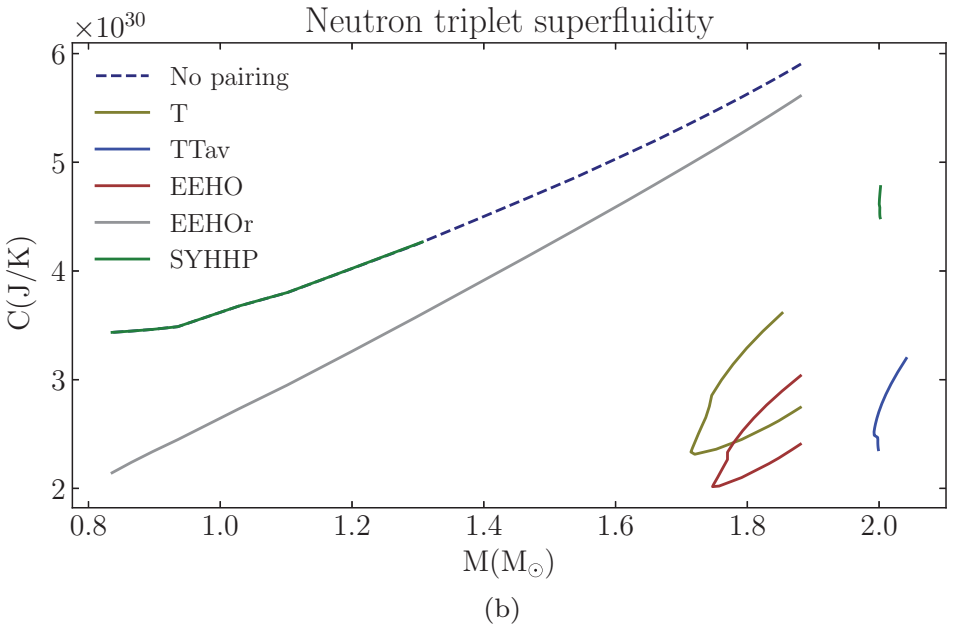


Fig. 8. Total heat capacity C (J/K) versus total mass M ($M_\odot$) for stars with luminosity $L_\nu = 3 \cdot 10^{27}$J/s for all EOS studied here. (a) Proton singlet superconductivity (b) Neutron triplet superfluidity.

why some heat capacity curves of superfluid stars overcome the heat capacity curve of stars without nuclear pairing. This situation does not happen in Fig. 8.

4. Discussion

Comparing the predictions of each gap model and EOS combination with the inferred neutrino luminosity of the accreting neutron star MXB 1659-29 (Fig. 6), we find that three models are disfavoured and can be excluded. $\mathrm{NT_{AO}}$, $\mathrm{NT_{BEEHS}}$ and $\mathrm{NT_{TToa}}$ predict that the whole core of the star will be superfluid at the core temperature of MXB 1659-29, hence direct Urca reactions are completely suppressed and we are not able to reproduce the inferred neutrino luminosity of MXB 1659-29. Other fast cooling processes would be equally suppressed by superfluidity, such that these models describe slow cooling stars for any EOS, thus, are inconsistent with the data. These gap models have similar opening and closing densities, but different amplitudes of the critical temperature T_c, suggesting that their location and width are the determinant factors in the luminosity prediction. This result is a consequence of the low temperature of MXB 1659-29's core in comparison with the models' T_c (Fig. 2). Therefore, we expect that any gap models with similar opening and closing densities will be unable to reproduce the data.

The other gap models can reproduce the observed neutrino luminosity. However, gap models $\mathrm{NT_{EEHOr}}$, $\mathrm{PS_{BS}}$ and $\mathrm{PS_{CCYps}}$ are disfavoured, particularly for high L EOS. These models close before the onset of direct Urca for all EOS, hence their predictions for neutrino luminosity are the same as in the case without superfluidity. Their calculated star masses are very close to the direct Urca threshold masses, and for high L ($L \geq 85$ MeV), they predict that all neutron stars will be fast cooling, which is in disagreement with observations. For low L, their predictions agree with data if less than 0.5% of the core volume of the star is involved in unsuppressed direct Urca reactions, which suggests fine-tuning because it requires that the neutron star mass happens to lie within a narrow range of masses near the dURCA threshold.

Of the remaining gap models, we highlight NT_{SYHHP}, which predicts that only very high mass stars have unsuppressed direct Urca for low and intermediate L ($L \leq 70\,\mathrm{MeV}$), whereas for high L, both low mass stars and very high mass stars cool through this fast process. This superfluid model is the only one predicting direct Urca cooling in low mass stars ($M \leq 1.3\,M_\odot$), thus, a future observation consistent with this situation would favour this model and $L \geq 70\,\mathrm{MeV}$. At the same time, it is able to accommodate slow cooling for stars with masses $2.0\,M_\odot \geq M \geq 1.3\,M_\odot$, in agreement with luminosity data from other transiently-accreting and isolated sources, for example, the ones studied in Refs. 26, 27.

Proton superconductivity models $\mathrm{PS_{AO}}$, $\mathrm{PS_{BCLL}}$, $\mathrm{PS_{EEHOr}}$ and $\mathrm{PS_{CCYms}}$ also predict fast cooling for low mass stars with $L \geq 70\,\mathrm{MeV}$, but, differently from the superfluid model NT_{SYHHP}, by themselves they do not accomodate slow cooling for stars with masses $M \geq 1.3\,M_\odot$, thus are inconsistent with data from other sources. However, a more realistic description of a neutron star would include a

combination of neutron superfluidity and proton superconductivity, which can significantly change the results shown here. Therefore, the proton superconductivity models above can not be excluded under this argument.

On Reference 9, it was discussed that the heat capacity and luminosity ratio can be written as

$$\frac{C_{38}}{L_{\nu,35}} = \left(\frac{\Delta\tilde{T}/\tilde{T}}{0.3\%}\right)^{-1} \left(\frac{t_q}{10\mathrm{yr}}\right) \tilde{T}_8^{-1} \tag{21}$$

where $\Delta\tilde{T}$ corresponds to the difference in temperature measurements after a quiescence time interval t_q, $C_{38} = C/10^{38}$ and $L_{\nu,35} = L_\nu/10^{35}$. Therefore, knowing $\Delta\tilde{T}/\tilde{T}$ and the inferred luminosity of a source, one can find its total heat capacity, which is intrinsically dependent on the star's particle composition. In Figure 9, we show our calculations for the heat capacities and luminosities of stars with different masses and gap models. For better visualization, only neutron triplet results for two EOS are shown. We display $\Delta T_{\mathrm{eff}}^{\infty}/T_{\mathrm{eff}}^{\infty}$, which is related to $\Delta\tilde{T}/\tilde{T}$ through $\Delta\tilde{T}/\tilde{T} = 1.8\,\Delta T_{\mathrm{eff}}/T_{\mathrm{eff}}$.

The difference between heat capacities for each triplet gap model is a factor of 2 at most, thus, one needs a few percent precision in observations of temperature variation to discriminate between particle composition scenarios.[8] Note that the intersection between $\Delta\tilde{T}/\tilde{T}$ curves and the vertical line marking the luminosity of source MXB 1659-29 is only relevant for $\Delta\tilde{T}/\tilde{T} < 2\%$. Bigger temperature variations for that luminosity would imply that the star is completely superfluid, such that the heat capacity is dominated by leptonic contributions. That is an unrealistic scenario, as discussed before.

Knowing L with precision can eliminate or favour certain gap models. However, the current predictions from gravitational waves[21,22] and nuclear experiments, such as PREX-II[10] are in tension, thus, unable to set reliable constraints on L. For this reason, we chose to investigate the whole phase space of L for the EOS family studied here. At the moment, our results are unable to set constraints for the EOS, however, with independent measurements of luminosity and mass of a neutron star source, this goal could be achieved. In this scenario, it will be particularly relevant to include more fast cooling processes in this framework, beyond direct Urca, as well as other equations of state, potentially with exotic components like hyperons, pions or free quarks.

We highlight that, for all cases studied here, the fraction of the core volume undergoing unsuppressed direct Urca is around 1%, which suggests that neutron stars with even higher neutrino luminosities and lower core temperatures should exist. Candidates for such cold stars are the sources SAX J1808.4-3658 and 1H 1905+000.[6] Further research is necessary to confirm whether the cooling of these sources is consistent with direct Urca reactions. Alternatively, one can speculate whether source MXB 1659-29 is actually a neutron star with a quark core or if it cools through exotic processes, such as neutrino emission from pion or kaon condensates. In some

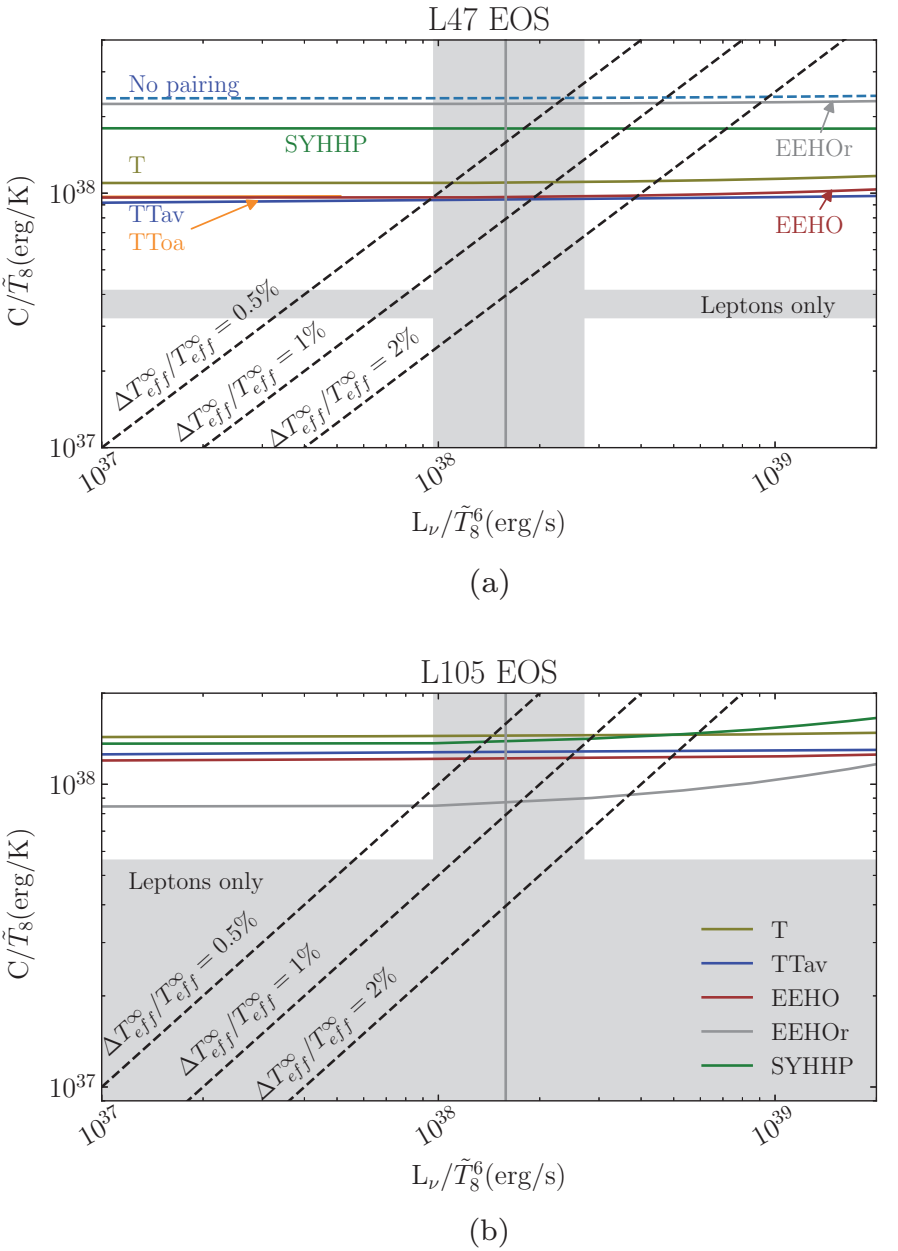


Fig. 9. Total heat capacity $C/\tilde{T}_8$ (erg/K) versus neutrino luminosity $L_\nu/\tilde{T}_8^6$ (erg/s) for neutron triplet gap models studied here. Only stars able to cool through direct URCA are displayed. $C_{leptons}$ is the lower limit of only electrons and muons contributing, mimicking a completely superfluid star. The vertical line and grey region around it correspond to the inferred luminosity of $L_\nu = 3 \times 10^{34}$ erg/s and uncertainty for source MXB 1659-29, more details on Ref. 8. The diagonal lines $\Delta T^{\infty}_{eff}/T^{\infty}_{eff}$ indicate temperature variation of the star's effective temperature after a period of ten years in quiescence. (a) *L*47 EOS (b) *L*105 EOS. The "no pairing" curve is under the *SYHHP* curve.

of these cases, their emissivities could be lower by up to 3 orders of magnitude, as shown in Ref. 5, resulting in larger fractions of the core volume cooling through direct Urca emission.

5. Summary

In this paper, we studied a family of hadronic EOS with different values of the slope of the symmetry energy, L, combined with several models of proton superconductivity and neutron superfluidity in the neutron star core. Comparing the direct Urca neutrino luminosity calculations with data from transiently-accreting source MXB 1659-29, we can rule out gap models that predict superfluidity in the whole star's core, $\mathrm{NT_{AO}}$, $\mathrm{NT_{BEEHS}}$ and $\mathrm{NT_{TToa}}$. We predict that pairing models with similar opening and closing densities can also be excluded. Gap models $\mathrm{NT_{EEHOr}}$, $\mathrm{PS_{BS}}$ and $\mathrm{PS_{CCYps}}$ are able to describe the observed luminosity but are disfavoured, because, for high L, they predict all neutron stars will be fast cooling and for low L, the masses need to be close to the direct Urca threshold. In a forthcoming paper, we will investigate neutron superfluidity and proton superconductivity simultaneous presence in the star's core.

We also calculated the total heat capacities for all stars which match the inferred neutrino luminosity of MXB 1659-29. Their values are within a factor of 2, hence one needs to distinguish between 1% and 2% of a variation of temperature over a decade[8] to differentiate between the particle composition scenarios. If independent observations of mass and luminosity can be made, they could be used to determine an accurate gap model description of the star's nuclear pairing, as well as the L parameter of its EOS. Furthermore, we were able to construct several detailed realistic scenarios where MXB 1659-29's estimated luminosity was obtained, consistent with direct Urca cooling at around 1% core volume, as previously estimated. This low percentage implies that stars with even larger neutrino luminosities and colder cores should exist. Alternatively, other fast cooling processes from exotic components with a lower emissivity could be operating over a larger fraction of the core.

Acknowledgments

We thank Sangyong Jeon for useful discussions and comments. This work was funded in part by the Natural Sciences and Engineering Research Council of Canada. M. M. acknowledges support by the Schlumberger Foundation Faculty for the Future Fellowship program. A. C. is a member of the Centre de Recherche en Astrophysique du Québec (CRAQ). M. M. and A. C. are members of the McGill Space Institute (MSI).

References

1. D. Page, J. M. Lattimer, M. Prakash and A. W. Steiner, *Stellar superfluids*, in *Novel Superfluids*, eds. K. H. Bennemann and J. B. Ketterson, International series of monographs on physics, Vol. 2 (Oxford, United Kingdom : Oxford University Press, 2013), ch. 21.
2. J. Bardeen, L. N. Cooper and J. R. Schrieffer, Theory of Superconductivity, *Phys. Rev.* **108**, 1175 (1957).
3. A. Gezerlis and J. Carlson, Strongly paired fermions: Cold atoms and neutron matter, *Phys. Rev. C* **77**, p. 032801 (2008).
4. W. C. G. Ho, K. G. Elshamouty, C. O. Heinke and A. Y. Potekhin, Tests of the nuclear equation of state and superfluid and superconducting gaps using the Cassiopeia A neutron star, *Phys. Rev. C: Nucl. Phys.* **91**, p. 015806 (2015).
5. D. G. Yakovlev, A. D. Kaminker, O. Y. Gnedin and P. Haensel, Neutrino emission from neutron stars, *Phys. Rep.* **354**, 1 (2001).
6. R. Wijnands, N. Degenaar and D. Page, Cooling of Accretion-Heated Neutron Stars, *J. Astrophys. Astron.* **38**, p. 49 (2017).
7. J. M. Lattimer, A Rapidly Cooling Neutron Star, *Physics Online Journal* **11**, p. 42 (2018).
8. E. F. Brown, A. Cumming, F. J. Fattoyev, C. J. Horowitz, D. Page and S. Reddy, Rapid Neutrino Cooling in the Neutron Star MXB 1659-29, *Phys. Rev. Lett.* **120**, p. 182701 (2018).
9. A. Cumming, E. F. Brown, F. J. Fattoyev, C. J. Horowitz, D. Page and S. Reddy, Lower limit on the heat capacity of the neutron star core, *Phys. Rev. C* **95**, p. 025806 (2017).
10. B. T. Reed, F. J. Fattoyev, C. J. Horowitz and J. Piekarewicz, Implications of PREX-II on the equation of state of neutron-rich matter, *Phys. Rev. Lett.* **126**, p. 172503 (2021).
11. W.-C. Chen and J. Piekarewicz, Building relativistic mean field models for finite nuclei and neutron stars, *Phys. Rev. C* **90**, p. 044305 (2014).
12. C. J. Horowitz, E. F. Brown, Y. Kim, W. G. Lynch, R. Michaels, A. Ono, J. Piekarewicz, M. B. Tsang and H. H. Wolter, A way forward in the study of the symmetry energy: experiment, theory, and observation, *J. Phys. G* **41**, p. 093001 (2014).
13. B.-A. Li, A. Ramos, G. Verde and I. Vidana, Topical issue on nuclear symmetry energy, *The European Physical Journal A* **50**, p. 1 (2014).
14. J. Piekarewicz and M. Centelles, Incompressibility of neutron-rich matter, *Phys. Rev. C* **79**, p. 054311 (2009).
15. C. J. Horowitz and J. Piekarewicz, Neutron star structure and the neutron radius of Pb-208, *Phys. Rev. Lett.* **86**, p. 5647 (2001).
16. R. J. Furnstahl, Neutron radii in mean field models, *Nucl. Phys. A* **706**, 85 (2002).
17. Z. Zhang and L.-W. Chen, Constraining the symmetry energy at subsaturation densities using isotope binding energy difference and neutron skin thickness, *Phys. Lett. B* **726**, 234 (2013).
18. D. Adhikari *et al.*, An Accurate Determination of the Neutron Skin Thickness of ^{208}Pb through Parity-Violation in Electron Scattering, *Phys. Rev. Lett.* **126**, p. 172502 (2021).
19. T. E. Riley *et al.*, A NICER View of PSR J0030+0451: Millisecond Pulsar Parameter Estimation, *Astrophys. J. Lett.* **887**, p. L21 (2019).

20. M. C. Miller *et al.*, PSR J0030+0451 Mass and Radius from NICER Data and Implications for the Properties of Neutron Star Matter, *Astrophys. J. Lett.* **887**, p. L24 (2019).
21. B. P. Abbott *et al.*, GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* **119**, p. 161101 (2017).
22. B. P. Abbott *et al.*, GW170817: Measurements of neutron star radii and equation of state, *Phys. Rev. Lett.* **121**, p. 161101 (2018).
23. J. M. Lattimer and A. W. Steiner, Constraints on the symmetry energy using the mass-radius relation of neutron stars, *Eur. Phys. J. A* **50**, p. 40 (2014).
24. G. Baym, C. Pethick and P. Sutherland, The Ground State of Matter at High Densities: Equation of State and Stellar Models, *Astrophys. J.* **170**, 299 (1971).
25. J. Negele and D. Vautherin, Neutron star matter at sub-nuclear densities, *Nucl. Phys. A* **207**, 298 (1973).
26. S. Beloin, S. Han, A. W. Steiner and K. Odbadrakh, Simultaneous fitting of neutron star structure and cooling data, *Phys. Rev. C* **100**, p. 055801 (2019).
27. D. Page, J. M. Lattimer, M. Prakash and A. W. Steiner, Neutrino emission from cooper pairs and minimal cooling of neutron stars, *Astrophys. J.* **707**, 1131 (2009).

Vacuum properties and astrophysical implications

A. Pérez Martínez[1,*], M. Pérez-Garcia[1,†], E. Rodríguez Querts[2,‡] and A. Romero Jorge[2,§]

[1] *Departamento de Física Fundamental, Universidad de Salamanca, Plaza de la Merced s/n 37008, Spain*

[2] *Departamento de Física Teórica, Instituto de Cibernética Matemática y Física (ICIMAF), Calle E esq 15 No. 309, Vedado, La Habana 10400, Cuba*

[*] *aurorapm1961@usal.es,* [†] *maperezga@usal.es,* [‡] *elizabeth@icimaf.cu,* [§] *adrian@icimaf.cu*

QED's predictions that photons propagating in a magnetized vacuum should feel the vacuum birefringence are still standing. Magnetars have strong magnetic fields and may give us signals of this effect through the delay of photons travelling from this source to detectors and the polarization position or by the angle and degree of polarization of the radiation emitted. Starting from non linear electrodynamics, we analyze and discuss for weak and strong field approximations the theoretical predictions for both using a toy model of rotating neutron stars with dipolar magnetic field shape and photon trajectories that lie radial.

Keywords: Magnetic field, Neutron stars, Polarization, Magnetized vacuum.

1. Introduction

Quantum Electrodynamics is a very well-established and successful theory with great numbers of experimental test. The theory conceives that fluctuations from the virtual electron-positron pairs, the vacuum, give rise to very interesting phenomena, becoming a magnetized vacuum in non-linear interaction theory. One of these properties is birenfrigence, which means that electromagnetic waves propagating perpendicular to a constant electric or magnetic field suffer changes in their speed of propagation. This velocity is a function of the external field. As incredible as it seems, this phenomenon has not been detected experimentally yet. It is not clear the reason behind that. It could be due to the smallness of the effect of the magnetic fields generated in the lab, but it is not ruled out that it represents a manifestation of new physics.

One of the most relevant experiment designs for getting birefringence is the Polarization of Vacuum with Laser (PVLAS) experiment.[1] A magnetic field of 5.5×10^4 G was used in this experiment. The PVLAS-team blamed axions[a] as responsible for the discrepancy between the established theoretical QED birefrigence and the null experimental signal.[1]

[*] Departamento de Física Teórica, Instituto de Cibernética Matemática y Física (ICIMAF), Calle E esq 15 No. 309, Vedado, La Habana 10400, Cuba.

[a] hypothetical elementary particles with low mass, candidates for dark matter, and in this context, photons could decay in axions and a new physics may appear.

However Astrophysics could give information about birefringence and other exotic phenomena of the QED-vacuum. The Universe is our main lab, while experiments on Earth are being improved.Very exotic objects like neutron stars have huge magnetic fields that could give us signals of birefrigence of the vacuum but also the inestability of the vacuum and pair production. Neutron stars have magnetic fields of the order of Schwinger's critical field $B_c = 4.41 \times 10^{13}$G,[2] and beyond for magnetars.[3] The birefringence signal from these objects could be measured through the time delay that should undergo the emitted polarized radiation of NS to be detected.

Our work is focused on obtaining a theoretical estimate of the angle of polarization as a consequence of vacuum birefringence, considering that the pulsar magnetosphere is a magnetized vacuum. In that case, we analyze the propagation of photons in two limits, weak and strong magnetic fields, and consider the dipolar shape of the magnetic field.

We started using effective Euler Heisenberg Lagrangian[4] to obtain the modes of polarization of photons traveling perpendicular to the magnetic field ($k \perp B$). This calculation is equivalent to the previous one obtained by solving the photon dispersion equation, considering the radiative corrections given by the magnetized photon self-energy.[5]

2. Euler Heisenberg non-linear electrodynamics

The non linear Euler Heisenberg Lagrangian for the weak $B \ll Bc$,$E \ll Ec$ and strong limits $B \gg Bc$, $E \gg Ec$ is presented in this section. Despite the fact that this formulation does not cover all microscopic photon-photon interactions, it is very useful because it describes vacuum as a non-linear optical medium allowing us to investigate specific phenomena using traditional techniques and interpretation of non-linear optics in media.

For the approximation $B \ll B_c$ the EH lagrangian[4] has the form

$$\mathcal{L}_{EH}^{W} = \mathcal{L}_0 + \xi(\mathcal{L}_0^2 + \frac{7\epsilon_0^2 c^2}{4}(\boldsymbol{E}.\boldsymbol{B})^2), \;\; \mathcal{L}_0 = \frac{\epsilon_0}{2}(E^2 - c^2 B^2), \tag{1}$$

$\mathcal{L}_0$ is the linear Lagrangian density and ϵ_0 and c are the dielectric constant and the speed of light in the vacuum respectively. QED vacuum corrections emerge through the constant $\xi = \frac{8\alpha^2\hbar^3}{45 m_e^4 c^5}$ which encloses the interaction of virtual pairs with magnetic field and depends on charge and mass of electrons. The vacuum: electron-positron virtual pairs have an indirect photon interaction responsible of non linearity of lagrangian Eq. (1). In the limit $B \gg B_c$ we have for the EH Lagrangian the following expression[4,6]

$$\mathcal{L}_{EH}^{S} = -\frac{B^2}{2} + \frac{m^4}{24\pi^2}\frac{B}{B_c}\left(ln\left(\frac{B}{2B_c}\right) + \frac{6}{\pi^2}\xi^{i\prime}(2)\right). \tag{2}$$

where $\frac{6}{\pi^2}\xi^{i\prime}(2) = -0.5699610$ with $\zeta(x)$ the Riemann zeta-function. This limit is important in astrophysics because compact objects, such as neutron stars (magnetars),

have a surface magnetic field of order $10^{13}-10^{15}$ G[7] which is higher than the critical magnetic field of B_c.

We start off from Lagrangian in two limits: weak and strong magnetic field to study the photon propagation perpendicular to a constant magnetic field B_e (in the direction x_3) in a vacuum. The photon can be described as electromagnetic wave with electric and magnetic field E_w and B_w which are orthogonal each other. Therefore, the total electric and magnetic fields are $E_t = E_w$, $B_t = B_w + B_e$.[8] The modified Maxwell equation solutions will be determined by whether the external magnetic field B_e is aligned with or orthogonal to the wave magnetic field B_w ($B_e \perp B_w$, $B_e \parallel B_w$), resulting in two physical polarization modes.[9,10] According to the case, we label along the paper $i = \perp, \parallel$ perpendicular and parallel respectively. Assuming E_w and B_w as plane wave $\sim exp(i(kz - \omega t)$ to solve modified Maxwell equations, doing some algebra and retaining only first term corrections, we may get the dispersion equations.[11] For the weak and strong field approximation reads as

$$\omega_{\parallel}^{W} = \mid \boldsymbol{k} \mid \sqrt{(1 - \frac{\xi}{4}(\boldsymbol{B_e} \times \hat{\boldsymbol{k}})^2)}, \quad \omega_{\perp}^{W} = \mid \boldsymbol{k} \mid \sqrt{(1 - \frac{7}{4}\xi(\boldsymbol{B_e} \times \hat{\boldsymbol{k}})^2)}, \tag{3}$$

$$\omega_{\parallel}^{S} = \mid \boldsymbol{k} \mid \sqrt{(1 - \frac{\alpha}{3\pi}(\hat{\boldsymbol{b}} \times \hat{\boldsymbol{k}}))}, \quad \omega_{\perp}^{S} = \mid \boldsymbol{k} \mid \sqrt{(1 - \frac{\alpha}{3\pi}\frac{B_e}{B_c}(\hat{\boldsymbol{b}} \times \hat{\boldsymbol{k}}))}. \tag{4}$$

For photon propagation perpendicular to the external magnetic field $\theta = \pi/2$. For weak field, both modes depend on the square of the external magnetic field. However, the strong magnetic field limit (SFL) has different behavior. In that case, we obtain where only the perpendicular mode depends on the magnetic field but linearly, while the mode parallel is independent of the magnetic field.

We define $\Delta n^{S,W} \equiv n_{\perp} - n_{\parallel}$ as the difference between the refraction indices of the two polarization modes,

$$\Delta n^{S} = \frac{\alpha}{4\pi}\frac{2}{15}\left(\frac{B_e}{B_c}\right)^2, \qquad \Delta n^{W} = \frac{\alpha}{4\pi}\frac{2}{3}\left(\frac{B_e}{B_c} - 1\right). \tag{5}$$

Let us note that in a magnetized vacuum, not only do photons acquire different velocities according to their polarization, but also the polarization changes once a photon has travelled a certain distance. Initial photon linear polarization becomes elliptical and the angular rotation of the polarization plane of the electromagnetic wave travelling a path length L is $\phi = \Delta n^{S,W} \omega L$.

3. Polarization radiation of neutron stars

Our focus in this section is to revisit the polarization that would occur when photons propagate in the magnetosphere of a neutron star, considering only the effects of a magnetized vacuum. As we mentioned before, NS are very exotic astronomical objects, characterized by huge densities and magnetic fields. The surface magnetic field could be around 10^{12-15} G, but its geometry is complicated, and it is impossible to consider it constant and uniform throughout the star.

Therefore, to do our study we have to make three important approximations: photons propagate in a magnetosphere considered as a magnetized vacuum, the magnetic field shape is dipolar and photon trajectories are radial. These assumptions allow us to calculate the polarization degree (PD) of radiation using the evolution equation, $\frac{\partial \mathbf{s}}{\partial \hat{r}} = \hat{\mathbf{\Omega}} \times \mathbf{s}$, $|\hat{\mathbf{\Omega}}| = \frac{\omega}{c}\Delta n^{W,S}$, where $\hat{r}$ is the direction of propagation of the photons, $\mathbf{s} = \{S_1/S_0, S_2/S_0, S_3/S_0\}$ is the normalized Stokes vector and $\hat{\mathbf{\Omega}}$ is related to the refraction index (defined in Eq (5)). We estimate the polarization degree PD of photons consider the adiabatic approximation $\hat{\mathbf{\Omega}}\left(\frac{1}{|\hat{\mathbf{\Omega}}|}\frac{\partial|\hat{\mathbf{\Omega}}|}{\partial\hat{r}}\right)^{-1} \gtrsim 1/2$, obtained in.[13] This condition allows us to calculate the polarization-limiting radius r_L at which the degree of polarization begins to be constant and the polarization becomes elliptical. We consider a dipolar shape configuration for the magnetic field as $B_e(r) = B_0\left(\frac{r_0}{r}\right)^3$, with B_0 and r_0 are the surface magnetic field and radius of the neutron star respectively. Because, regardless of the value of the surface magnetic field B_0, the important values of magnetic field strength are those where decoupling of polarization modes occurs, whether far or close to the neutron star, our calculations will be performed for the entire range of magnetic fields using the weak and strong field approximations. If the decoupling modes occur far enough from the star's surface, $r_L \gg r_0$, as the magnetic field goes as r^{-3}, the magnetic field will be significantly smaller. Then, $\Omega \sim \Delta n^W \sim B_e^2$. When decoupling arises for $r_0 \leq r_L < 2r_0$ the strong field approximation, $B_0/B_c \gtrsim 1$, is suitable, and $\Omega \sim \Delta n^S \sim B_e$. We assume rotating NS for both magnetic field regimes. Then, the angle ϕ is given by the relation $r_L - r_0 = \phi\frac{P}{2\pi}$ where P is the period of the pulsar.[12] The computed polarization angle is given by the equations

$$\phi(\omega, B, P) = \begin{cases} \frac{2\pi}{cP}\left(\left[\frac{\alpha}{90\pi}\frac{\omega}{c}\left(\frac{B_0}{B_c}\right)^2 r_0^6\right]^{1/5} - r_0\right), & B < B_c \text{ (weak)} \\ \frac{2\pi}{cP}\left(\left(\frac{\alpha}{9\pi}\frac{\omega}{c}\frac{B_0}{B_c}r_0^3\right)^{1/2} - r_0\right), & B > B_c \text{ (strong)} \end{cases} \tag{6}$$

and polarization degree is $\Pi(\omega, B, P) = 4\pi/cP\left((3/r_L)^2 + (4\pi/cP)^2\right)^{-1/2}$. For higher frequencies, the radiation from neutron stars' magnetosphere is substantially polarized (fig. 1). Furthermore, neutron stars or magnetars with a higher magnetic field, release strongly polarized radiation. The polarization degree (PD) of radiation of a pulsar is inversely related to its period; millisecond pulsars have more polarization than second period pulsars. The behavior of angle and polarization degree with the magnetic field is depicted in fig. 1 for weak and strong field approximations. For magnetic fields greater than $2B_c$, strong field approximation contributes to an increase the polarization degree and polarization angle compared to weak field approximation.

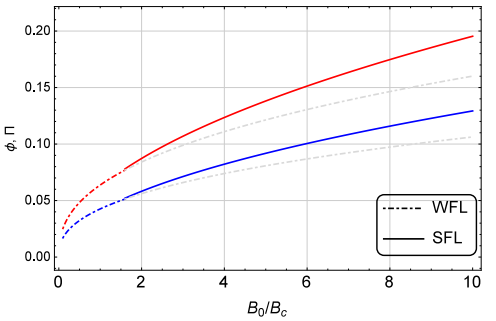


Fig. 1. Comparison of the polarization degree (Blue line) and polarization angle (Red line) for weak field limit (dotdashed line) and strong field limit (solid line) as function of the surface magnetic field of the pulsar. We take typical neutron star parameters, a radius $r_0 = 10$ km, a period of $P = 1$ s and radiation of $\omega = 2$ MeV of energy.

Our results are in agreement with previous obtained by Heyl et al[13] when weak field approximation is taken into account. Furthermore, studies on $RXJ1856.5-3764$ for polarization of visible and X radiation revealed that it is essential to include the influence of magnetized vacuum to explain the observational data of polarization of this source.[14]

For photon energy of $\omega = 1MeV$ and typical values of the parameters $B_0 = 1.5 \times 10^{13}G$, $M = 1.4M_{\odot}$, radius $r_0 = 12km$, and $P = 7.06s$, the angle and degree of polarization for $RXJ1856.5-3764$ owing to simply the effect of magnetized vacuum give $\phi = 6.4 \times 10^{-3}$ and $\Pi = 0.42\%$.[15,16] Obviously, when geometric factors are taken into account, this value rises.[14]

4. Conclusions

From non-linear electrodynamics we have obtained the refraction index and different phase velocity of photon propagating in strong magnetic field according to the polarization modes. We have also obtained the polarization angle that the radiation acquires as a consequence of crossing a magnetized vacuum. The calculations have been done in weak and strong field approximation for a constant magnetic field.

In addition, we investigated the influence of a magnetized vacuum on the polarization of radiation emitted by rotating NS. To do so, we considered the magnetosphere to be a magnetized vacuum with a dipolar magnetic field and photons traveling along radial trajectories.

Our findings are consistent with previous results based on weak field approximation.[13] More energetic photons have a greater polarization degree than the lower ones. Besides, the polarization degree is strongly dependent on the period of rotation of neutron stars since this can raise or decrease the polarization limiting radii, causing the couple or decouple of the propagation modes.

The authors of the work[14] claim that there is no way of explaining the polarization data from visible and X rays of the pulsar $RXJ1856.5-3764$ without

considering the effect of magnetized vacuum. As a consequence, it might be a test of vacuum birefringence, albeit we expect it will be tested in terrestrial laboratories.

However, not all regarding vacuum birefringence has been said, and the door is opened for new physics. This work was undertaken as a starting point for developing more realistic models of polarization radiation emission of NS which will allow us to identify in robust way those linked to vacuum birefringence and may validate the QED prediction.

Acknowledgments

A.P.G and A.P.M acknowledge the support of the Agencia Estatal de Investigación through the grant PID2019-107778GB-100 from Junta de Castilla y Leon, Spanish Consolider MultiDark FPA2017-90566-REDC and PHAROS COST Actions MP1304 and CA16214. The work of E.R.Q. and A.W.R.J. were supported by the project of No. NA211LH500-002 from AENTA-CITMA, Cuba.

References

1. A. Ejlli, F. Della Valle, U. Gastaldi, G. Messineo, R. Pengo, G. Ruoso and G. Zavattini, *Phys. Rept.*, **871**, 1-74 (2020).
2. R. Ciolfi, *Astron. Nachr*, **335**, 624-629 (2014).
3. C. Kouveliotou, S. Dieters, T. Strohmayer et al., *Nature*, **393**, 235-237 (1998).
4. W. Heisenberg and H. Euler, *Z. Phys.*, **98**, nos. 11-12, 714-732 (1936).
5. A. W. Romero Jorge, E. Rodríguez Querts, A. Pérez Martínez et al. *Astron. Nachr.*, **340**, 852-586 (2020).
6. V. I. Ritus, *arXiv: High Energy Physics - Theory* (1998).
7. R. C. Duncan, *arXiv: Astrophysics*, **526**, 830-841 (2000).
8. H. Pérez Rojas, *J. Exp. Theor. Phys.*, **76**, 1 (1979).
9. H. Pérez Rojas and A. E. Shabad, *Ann. Phys.*, **121**, 432-455 (1979).
10. H. Pérez Rojas and E. Rodríguez Querts. *Phys. Rev. D*, **79**, 093002 (2009).
11. H. P. Rojas and E. R. Querts, *Jour. of Mod. Phys. D*, **19**, No. 810, 1599-1608 (2010).
12. J. S. Heyl and N. J. Shaviv, *Mon. Not. R. Astron. Soc.*, **311**, 555–564 (2000).
13. Jeremy S. Heyl, Nir J. Shaviv, Don Lloyd, *Monthly Notices of the Royal Astronomical Society*, Volume **342**, 1, 134–144 (2003).
14. R. P. Mignani, V. Testa, D. G. Caniulef, R. Taverna, R. Turolla, S. Zane and K. Wu, *Mon. Not. Roy. Astron. Soc.*, **465**, no. 1, 492-500 (2017).
15. A. M. Pires, F. Haberl, V. E. Zavlin, C. Motch, S. Zane, M. M. Hohle, *Astronomy and Astrophysics*, **563**, A50 (2014).
16. S. B. Popov, R. Taverna and R. Turolla, *Mon. Not. Roy. Astron. Soc.*, **464**, no. 4, 4390-4398 (2017).

Testing extended theories of gravity with GRBs

L. Mastrototaro

Dipartimento di Fisica "E.R Caianiello", Università degli Studi di Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy
Istituto Nazionale di Fisica Nucleare - Gruppo Collegato di Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy
E-mail: lmastrototaro@unisa.it

We present our studies on the neutrino pairs annihilation into electron-positron pairs ($\nu\bar{\nu} \to e^- e^+$) near the surface of a neutron star in the framework of extended theories of gravity. The latter modifies the maximum energy deposition rate near to the photonsphere and it might be several orders of magnitude greater than that computed in the framework of General Relativity. These results provide a rising in the Gamma-Ray Bursts energy emitted from a close binary neutron star system and might be a fingerprint of modified theories of gravity, changing our view of astrophysical phenomena.

Keywords: Extended theories of gravity, GRB, Neutrino energy deposition, Black Holes, Neutron Stars

1. Introduction

General Relativity (GR) is without any doubt the best theory of gravitational interaction. Its predictions have been tested with high accuracy on scales of the solar system (for example, the precession of the Mercury perihelion, the photons deviation and the gravitational leasing effect), on astrophysical scales (the gravitational waves), and cosmological scales (the cosmic microwave background radiation (CMBR) and the formation of primordial light elements (Big Bang Nucleosynthesis)). Despite these results, there are still open questions that make GR incomplete. The latter arise at short distances and small time scales (black hole and cosmological singularities, respectively), or at large distance scales, the rotational curve of the galaxies and the observed accelerated phase of the present Universe, for which any predictability is lost.

To solve these issues, deviations from the GR (hence from the Hilbert-Einstein action on which GR is based) are needed, and new ingredients, such as dark matter and dark energy, are required for fitting the present picture of our Universe.[1–6] Indeed, in the last years, several *alternative* or *modified* theories of gravity have been proposed, which try to answer all the opened questions of GR and the Cosmological Standard Model. To give an example, higher-order curvature invariants than the simple Ricci scalar R, allow getting inflationary behaviour, removing the primordial singularity, as well as explaining the flatness and horizon problems[7,8] (for further applications, see Refs.[9–37]). On the other hand, one can tend to preserve the GR and the Hilbert-Einstein action, adding only some two unidentified components: dark matter and dark energy. These two different approaches try to solve the same

problems but at the moment there is not a final solution for that (modified gravity theories do not manage to solve all the problems while dark matters elements are still missing). One of the consequences of these approaches is that the metric tensor $g_{\mu\nu}$ describing the gravitational field generated by a massive source gets modified with respect to GR metric, in particular, the Schwarzschild or Kerr metric. The latter are recovered in the limit in which the parameters characterizing some specific theory of gravity beyond GR are set to zero.

In this proceeding, we highlight the differences between GR and extended theories of gravity arising from the mechanism of generation of Gamma-Ray Bursts (GRBs). We focus, in particular, on GRBs powered by neutrino annihilation processes $\nu\bar{\nu} \to e^+e^-$.[a]

The neutrinos annihilation process is relevant in many astrophysical frameworks: in the stellar envelope, as well as on the delay shock mechanism into the Type II Supernova (at late time, from the hot proto-neutron star, the energy is deposited into the supernova envelope via neutrino pair annihilation and neutrino-lepton scattering). These processes augment the neutrino heating of the envelope generating a successful supernova explosion.[44] Moreover, the neutrino annihilation process has been also proposed as the mechanism that power GRB in a binary neutron star system, which is the topic of this work. Simulations and analytical estimations performed within GR show that the mechanism is not sufficient to generate the required energy for explaining short GRB. Such a conclusion changes, as we will show, if the gravitational background is described by modified theories.

2. Neutrino energy deposition formulation

In this section we recall the main features to treat the energy deposition in curved spacetimes.[45,46] Previous calculations of the $\nu\bar{\nu} \to e^-e^+$ reaction in the vicinity of a neutron star have been first based on Newtonian gravity,[47,48] then the effect of gravity has been incorporated for static stars,[44,49] and then extended to rotating stars.[45,50] We consider the spacetime around a black hole described by the following diagonal metric

$$g_{\mu\nu} = \mathrm{diag}\left(-f(r), h(r), r^2, r^2\sin^2\theta\right) , \tag{1}$$

The energy deposition per unit time and per volume is given by (considering $c = \hbar = 1$)[51]

$$\begin{aligned}\dot{q}(r) = &\iint f_\nu(\mathbf{p}_\nu, r) f_{\overline{\nu}}(\mathbf{p}_{\overline{\nu}}, r) \\ &\times \left[\sigma|\mathbf{v}_\nu - \mathbf{v}_{\overline{\nu}}|\varepsilon_\nu\varepsilon_{\overline{\nu}}\right]\frac{\varepsilon_\nu + \varepsilon_{\overline{\nu}}}{\varepsilon_\nu\varepsilon_{\overline{\nu}}} d^3\mathbf{p}_\nu d^3\mathbf{p}_{\overline{\nu}} ,\end{aligned} \tag{2}$$

[a] Indeed the state of the art of NS-NS merger simulations is special relativistic, axisymmetric hydrodynamic simulations for the final black hole-torus systems[38–43] from which emerge that neutrino-pair annihilation in ordinary GR models seems to be not efficient enough to power GRBs and the Blandford-Znajek process is currently considered a more promising mechanism for launching. In extended theories of gravity, the situation can be different as we will show in this proceeding.

where $f_{\nu,\overline{\nu}}$ are the neutrino number densities in phase space, $\mathbf{v}_\nu$ the neutrino velocity, and σ is the rest frame cross section. Since the term $\sigma|\mathbf{v}_\nu - \mathbf{v}_{\overline{\nu}}|\varepsilon_\nu\varepsilon_{\overline{\nu}}$ is Lorentz invariant, it can be calculated in the center-of-mass frame, and turns out to be

$$\sigma|\mathbf{v}_\nu - \mathbf{v}_{\overline{\nu}}|\varepsilon_\nu\varepsilon_{\overline{\nu}} = \frac{DG_F^2}{3\pi}(\varepsilon_\nu\varepsilon_{\overline{\nu}} - \mathbf{p}_\nu \cdot \mathbf{p}_{\overline{\nu}}c^2)^2\,, \tag{3}$$

with $G_F = 5.29 \times 10^{-44}$ cm^2 MeV^{-2} the Fermi constant,

$$D = 1 \pm 4\sin^2\theta_W + 8\sin^4\theta_W\,, \tag{4}$$

$\sin^2\theta_W = 0.23$ the Weinberg angle, and the plus sign for electron neutrinos and antineutrinos while the minus sign for muon and tau type. $T(r)$ is the temperature measured by the local observer and $\Theta(r)$ is the angular integration factor. At these energies, the mass of the electrons can be neglected and it is possible to obtain that the general expression of the rate per unit time and unit volume of the $\nu\bar{\nu} \to e^+e^-$ process[44]

$$\dot{q} = \frac{7DG_F^2\pi^3\xi(5)}{2}[kT(r)]^9\Theta(r)\,. \tag{5}$$

The evaluation of $T(r)$ and $\Theta(r)$ account for the gravitational redshift and path bending. To write the latter in terms of observed luminosity L_∞ one has to has to consider that temperature, like energy, varies linearly with red-shift and following the procedure of Ref.[44], one finds that:

$$\Theta(r) = \frac{2\pi^3}{3}(1-x)^4(x^2+4x+5)\,, \tag{6}$$

$$T(r) = \frac{\sqrt{f(R)}}{\sqrt{f(r)}}T(R)\,, \tag{7}$$

$$L_\infty = f(R)L(R)\,, \tag{8}$$

$$L(R) = L_\nu + L_{\overline{\nu}} = \frac{7}{4}\,a\pi R^2T^4(R)\,. \tag{9}$$

Here R is the neutrinosphere radius (the spherical surface where the stellar material is transparent to neutrinos and from which neutrinos are emitted freely), $L(R)$ is the neutrino luminosity, a the radiation constant, $x = \sin^2\theta_r$, θ_r is the angle between the trajectory and the tangent velocity in terms of local radial and longitudinal velocities[45] for which one can obtain that[46]

$$\cos\theta_r = \frac{R}{r}\sqrt{\frac{f(r)}{f(R)}}. \tag{10}$$

This relation comes from the fact that the impact parameter b is constant on all the trajectory and is related to $\cos\theta_r$ by the relation

$$b = \left(\frac{f(r)}{r\cos\theta_r}\right)^{-1}\,. \tag{11}$$

Moreover, photosphere radius R_{ph} exists below which a massless particle can not be emitted tangent to the stellar surface. The present discussion is therefore restricted to $R > R_{ph}$. The neutrino emission properties hence mainly depend on the geometry of spacetime. From the equation of the velocities

$$\dot{r}^2 = \left(E\dot{t} - L\dot{\phi}\right) f(r)\,, \quad \dot{\phi} = \frac{L}{r^2}, \quad \dot{t} = -\frac{E}{f(r)},$$

where E and L are the energy and angular momentum at the infinity, one gets the effective potential V_{eff}, such that the photonsphere radius follows from the condition $\frac{\partial V_{\rm eff}}{\partial r} = 0$. The circular orbit is derived by imposing $\dot{r}^2 = 0$. We note en passant that these results reduce to ones derived in the case of the Schwarzschild geometry, $R_{ph} = 3M$, as calculated in.[44]

The integration of $\dot{q}$ from R to infinity gives the total amount of local energy deposited by the neutrino annihilation process (for a single neutrino flavour) for time units

$$\dot{Q} = 4\pi \int_R^\infty dr \frac{r^2}{\sqrt{f(r)}}\,\dot{q}\,. \tag{12}$$

According to,[44] the total energy deposition from the neutrinosphere radius R_6 to infinity (for a symmetric spherical star that emits neutrinos from a spherical neutrinosphere) is given by

$$\dot{Q}_{51} = 1.09 \times 10^{-5} \mathcal{F}\left(\frac{M}{R}\right) D L_{51}^{9/4} R_6^{-3/2}\,. \tag{13}$$

Here $\dot{Q}_{51}$ is expressed in units of 10^{51} erg s^{-1}, L_{51} is neutrino luminosity in units of 10^{51} erg s^{-1}, $D = 1 \pm 4\sin^2\theta_W + 8\sin^4\theta_W$, $\sin^2\theta_W = 0.23$ and the plus sign is for electron neutrinos and antineutrinos while the minus sign is for muon and tau type, R_6 is the radius in units of 10 km and, for a generic diagonal metric of the form $g_{\mu\nu} = \left(g_{00}, -1/g_{00}, -r^2, -r^2\sin^2\theta\right)$, the function $\mathcal{F}\left(\frac{M}{R}\right)$ is given by

$$\mathcal{F}\left(\frac{M}{R}\right) = 3g_{00}(R)^{9/4} \int_1^{\rm R_{ch}} (x-1)^4(x^2+4x+5)\frac{y^2 g_{11}(yR)^{1/2} dy}{g_{00}(yR)^{9/2}}\,. \tag{14}$$

In the Newtonian limit one gets $\mathcal{F}(0) = 1$ so that it is convenient, for our analysis, to consider the ratio $\dot{Q}_{\rm GR}/\dot{Q}_{\rm Newt} = \mathcal{F}(M/R)$. Usually, almost all energy is carried out by electron neutrino thus we can approximate Eq. (13) considering $D = 1.23$.

Equations (13) and (14) allow obtaining the deposited energy of neutrinos and the energy that can be emitted to powering GRB. In the next Section, we will study the ratio (14) for astrophysical objects in various modified gravity theories.

3. Neutrino deposition in modified gravity

In what follows, we present three relevant cases that explain how relevant could be the modification to GR in this context. First of all, we take into consideration the

Einstein dilaton Gauss-Bonnet gravity. The action is[52]

$$S = \frac{1}{8\pi}\int d^4x\sqrt{-g}\left(\frac{R}{2} - \frac{1}{2}\partial_\mu\psi\partial^\mu\psi + \alpha\psi L_{GB}\right) \ , \tag{15}$$

where ψ is a scalar field and L_{GB} is the Gauss-Bonnet invariant: $L_{GB} = R^2 - 4R^{\alpha\beta}R_{\alpha\beta} + R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}$. The solution considered is the Sotiriou-Zhau solution(solution in perturbation theory[53]), which lead to results in Fig. 1[46]:

$$ds^2 = -f(r)dt^2 + h(r)dr^2 + r^2d\Omega^2 \ , \tag{16}$$

with

$$f(r) = \left(1 - \frac{2m}{r}\right)\left(1 + \sum_n A_n\bar{a}^n\right) \ ;$$
$$h(r) = \left(1 - \frac{2m}{r}\right)^{-1}\left(1 + \sum_n B_n\bar{a}^n\right) \ .$$

where, to the second order:

$$A_1 = B_1 = 0 \ ;$$
$$A_2 = -\frac{49}{40m^3r} - \frac{49}{20m^2r^2} - \frac{137}{30mr^3} - \frac{7}{15r^4} + \frac{52m}{15r^5} + \frac{40m^2}{3r^6} \ ;$$
$$B_2 = \frac{49}{40m^3r} + \frac{29}{20m^2r^2} + \frac{19}{10mr^3} - \frac{203}{15r^4} - \frac{436m}{15r^5} - \frac{184m^2}{3r^6} \ .$$

The maximum value taken for α, considering the perturbative regime of the solution, shown an increase of the 50% for the maximum amount of energy deposition respect to GR.

The second case that we want to analyze is the Brans Dicke theory. It represents a generalization of general relativity, where gravitational effects are in part due to geometry, in part due to a scalar field. The action is[54]

$$S = \int d^4x\sqrt{-g}\left[\psi R + \frac{16\pi}{c^4}L - \omega(\psi)\right] \ , \tag{17}$$

where L is the Lagrangian density of all the matter, including all non-gravitational field, ψ is a scalar field and ω is its Lagrangian density.

With this Lagrangian, expressing the line element in the isotropic form, we obtain the solution[54]

$$ds^2 = -e^{2\alpha}dt^2 + e^{2\beta}\left[dr^2 + r^2d\Omega^2\right] \ , \tag{18}$$

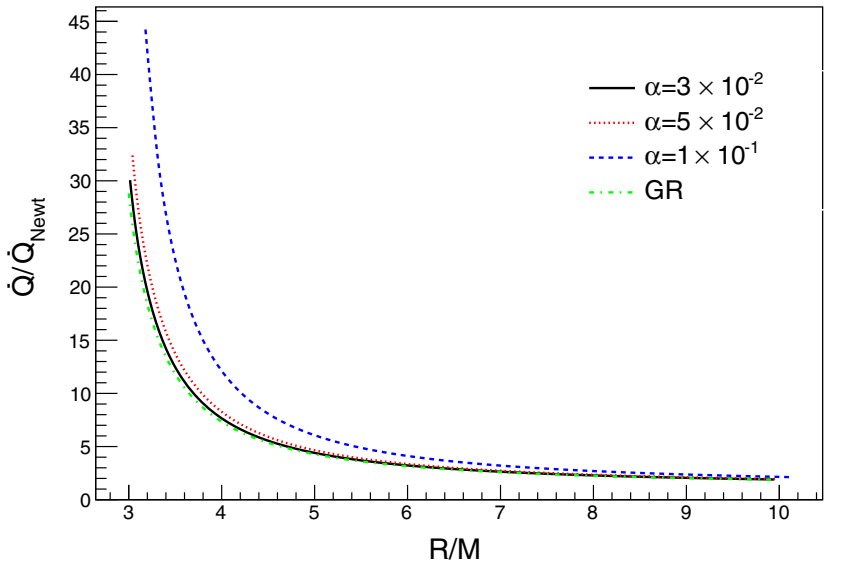


Fig. 1. Ratio of energy deposition $\dot{Q}$ for the the Sotiriou-Zhau metric to total Newtonian energy deposition $\dot{Q}_{\rm Newt}$ for different values of α. The green curve shows the GR energy deposition for comparison.

where

$$\lambda = \sqrt{(C+1)^2 - C(1-\frac{C\omega}{2})}$$

$$e^{2\alpha} = e^{2\alpha_0}\left[\frac{1-\frac{B}{r}}{1+\frac{B}{r}}\right]^{\frac{2}{\lambda}} ,$$

$$e^{2\beta} = e^{2\beta_0}\left(1+\frac{B}{r}\right)^4\left[\frac{1-\frac{B}{r}}{1+\frac{B}{r}}\right]^{\frac{2(\lambda-C+1)}{\lambda}} ,$$

$$\psi = \psi_0\left[\frac{1-\frac{B}{r}}{1+\frac{B}{r}}\right]^{-\frac{C}{\lambda}} ,$$

with ω positive constant and

$$\alpha_0 = \beta_0 = 0 \ ,$$
$$\psi_0 = \frac{4+2\omega}{3+2\omega} \ ,$$
$$C \sim -\frac{1}{2+\omega} \ ,$$
$$B \sim \frac{M}{2\sqrt{\psi_0}} \ .$$

Using this metric, we obtain the shape for energy deposition in Fig. 2. Even with this model we have an enhancement of about 50% respect to the maximum value of $\dot{Q}/\dot{Q}_{\rm Newt}$ 30 in GR.

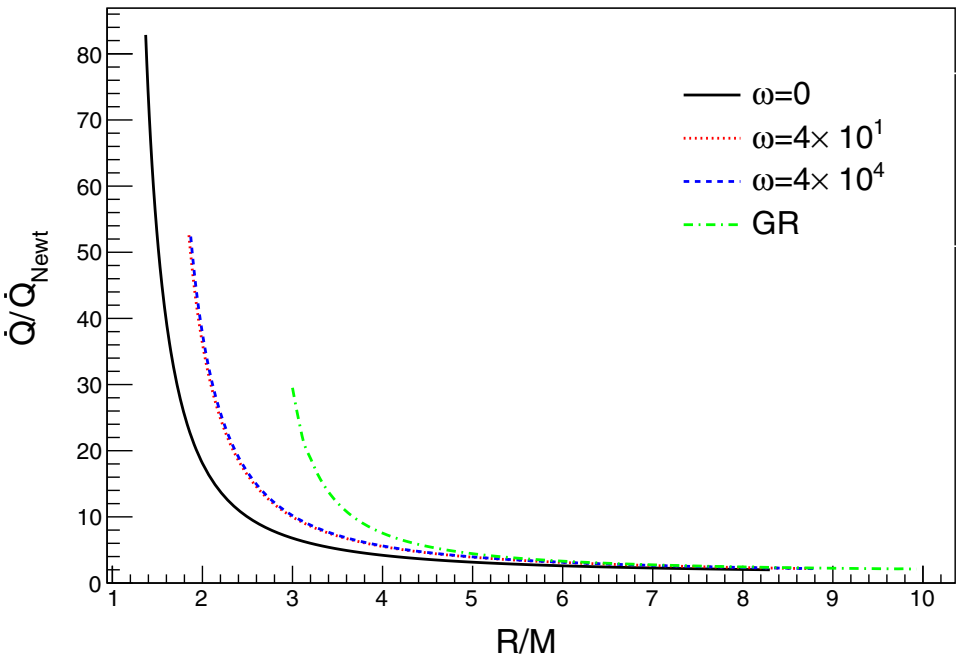


Fig. 2. Ratio of energy deposition $\dot{Q}$ for metric in Eq. (18) to total Newtonian energy deposition $\dot{Q}_{\rm Newt}$ for different value of ω. The green curve shows the GR energy deposition for comparison.

Finally, we discuss the case of a BH surrounded by quintessence field. In this case

$$g_{00}(r) = 1 - \frac{2M}{r} - \frac{c}{r^{3\omega_q - 1}} \ , \tag{19}$$

where c is a positive constant and $-1 < \omega_q < -1/3$. The quintessence parameter is constrained by the fact that increasing c, the model passes from describing a black hole with an event horizon to representing a naked singularity.

We chose to show only the case with $\omega = -0.4$, for which we obtain results in Fig. 3 and an enhancement of a factor 25 with respect to GR.

3.1. *GRB enhancement*

The above results show that modified gravity provides an enhancement of the neutrino annihilation process as compared to GR. Such an enhancement is relevant

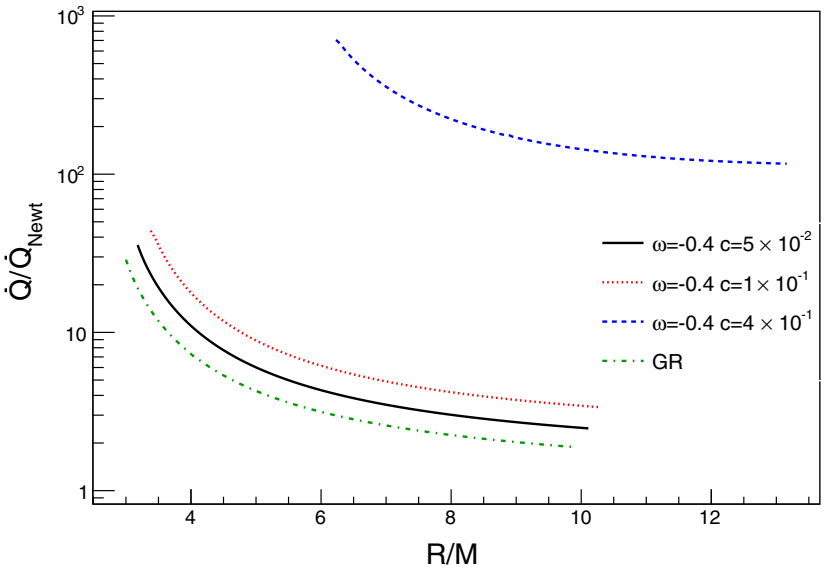


Fig. 3. Ratio of total energy deposition $\dot{Q}$ for $\omega = -0.4$ to total Newtonian energy deposition $\dot{Q}_{\rm Newt}$ for three values of the parameter c. The green curve shows the GR energy deposition for comparison.

for powering GRBs for the model given by a closed neutron stars binary merging system. Neutrino emission happens in the last phase of the merging and the final configuration is a black hole (or neutron star) with an accretion disk. With the developed formalism, we can not describe the disk emission (we are considering a spherical system), so we restrict ourselves to the central BH. Taking into account total energy emitted into neutrinos from the central BH of $\mathcal{O}(10^{52})$ erg, a radius of $R = 20$ km, the maximum possible total energy released in GRB is

$$Q_{\rm GR} \sim 2.5 \times 10^{49} \text{ erg}\,, \tag{20}$$

which is too small to explain the short GRB from neutron star merging. Instead, considering the maximum enhancement that modified gravity induces shown in Figs. 1, 2 and 3, we have that the total possible energy released in GRB can exceed the maximum energy of $\mathcal{O}(10^{52})$ erg (we have to remind that we are considering only the neutrino emission from the central BH and therefore the true emitted energy is larger considering the whole BH+disk configuration). Moreover, considering that the deposited energy is converted very efficiently to the relativistic jet energy, we infer a constraint on the quintessence model. It is possible to obtain that the maximum allowed value of $\mathcal{F}(R_{ph})$ is $\mathcal{O}(10^4)$. The contour plots given in Figs. 4 and 5 show the value of $\mathcal{F}(R_{\rm ph})$ for allowed value of ω_q and c. Therefore, one can infer, for $\mathcal{F}(R_{\rm ph}) \sim \mathcal{O}(10^4\text{-}10^5)$, the values of the parameter c that are not allowed in the considered scenario.

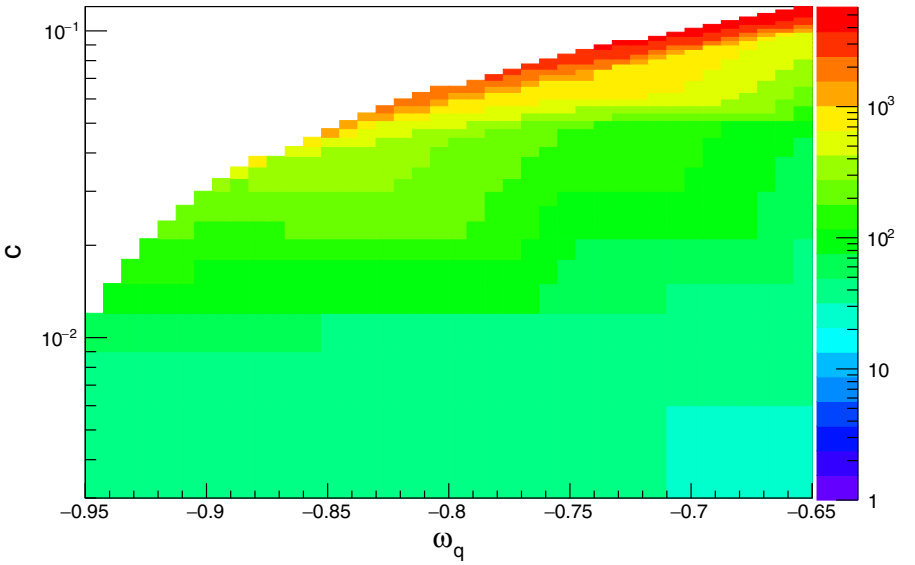


Fig. 4. Contour plot for $\omega_q \in]-1, -0.65[$. On the y axis are reported the excluded values of c due to the creation of a naked singularity (white part) and, on the right, the values of $\mathcal{F}(R_{\rm ph})$. It can be also seen the values of the parameter c, for which $\mathcal{F}(R_{\rm ph}) \sim \mathcal{O}(10^4\text{-}10^5)$, that are excluded by the energy deposition bounds.

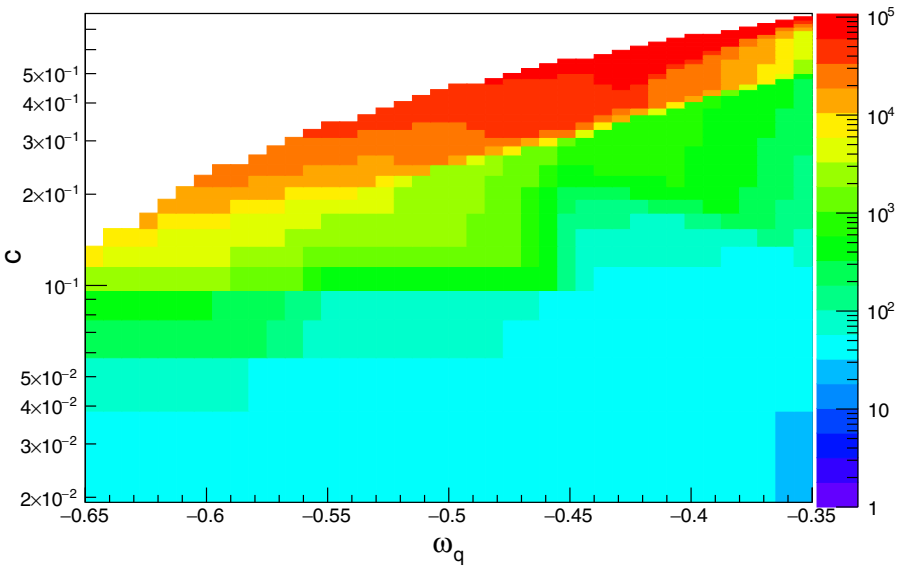


Fig. 5. Contour plot for $\omega_q \in]-0.65, -0.35[$. On the y axis are reported the excluded values of c due to the creation of a naked singularity (white part) and, on the right, the values of $\mathcal{F}(R_{\rm ph})$. It can be also seen the values of the parameter c, for which $\mathcal{F}(R_{\rm ph}) \sim \mathcal{O}(10^4\text{-}10^5)$, that are excluded by the energy deposition bounds.

4. Conclusion

In conclusion, we have analyzed the neutrino pair annihilation process $\nu\bar{\nu} \to e^+e^-$ near a BH in some modified gravitational theories. We have shown that, owing to a shift of the photosphere radius, there is an enhancement of the deposited energy rate ratio with respect to GR. Such an enhancement could be a relevant mechanism for the generation of GRBs in close neutron star binary merging, for which neutrino pairs annihilation has been proposed as a possible source. Moreover, the released energy may be larger than the Short GRB maximum energy observed of $\mathcal{O}(10^{52})$ erg as it happens in the quintessence model (cfr. Ref.[55]). In that case, one can constrain the value of ω and c such that the released energy is inferior to $\mathcal{O}(10^{52})$ erg. The $\nu\bar{\nu} \to e^+e^-$ processes are of great importance in astrophysics, as well as in modified gravity, since they lead to considerable differences with respect to GR and its phenomenology. Therefore, the results presented in this proceeding could provide a new astrophysical framework to search for a signature of gravitational theories beyond GR.

References

1. A. G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116**, 1009 (1998).
2. S. Perlmutter *et al.*, Measurements of Ω and Λ from 42 high redshift supernovae, *Astrophys. J.* **517**, 565 (1999).
3. S. Cole *et al.*, The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications, *Mon. Not. Roy. Astron. Soc.* **362**, 505 (2005).
4. G. Hinshaw *et al.*, Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: temperature analysis, *Astrophys. J. Suppl.* **170**, p. 288 (2007).
5. S. M. Carroll, The Cosmological constant, *Living Rev. Rel.* **4**, p. 1 (2001).
6. V. Sahni and A. A. Starobinsky, The Case for a positive cosmological Lambda term, *Int. J. Mod. Phys. D* **9**, 373 (2000).
7. A. A. Starobinsky, A New Type of Isotropic Cosmological Models Without Singularity, *Adv. Ser. Astrophys. Cosmol.* **3**, 130 (1987).
8. A. A. Starobinsky, The Perturbation Spectrum Evolving from a Nonsingular Initially De-Sitter Cosmology and the Microwave Background Anisotropy, *Sov. Astron. Lett.* **9**, p. 302 (1983).
9. H. Oyaizu, M. Lima and W. Hu, Nonlinear evolution of $f(R)$ cosmologies. II. Power spectrum, *Phys. Rev. D* **78**, p. 123524 (Dec 2008).
10. L. Pogosian and A. Silvestri, The pattern of growth in viable f(R) cosmologies, *Phys. Rev. D* **77**, p. 023503 (2008), [Erratum: Phys. Rev. D 81, 049901 (2010)].
11. I. Sawicki and W. Hu, Stability of cosmological solutions in $f(R)$ models of gravity, *Phys. Rev. D* **75**, p. 127502 (Jun 2007).
12. B. Li and J. D. Barrow, Cosmology of $f(R)$ gravity in the metric variational approach, *Phys. Rev. D* **75**, p. 084010 (Apr 2007).
13. T. Clifton, Higher powers in gravitation, *Phys. Rev. D* **78**, p. 083501 (Oct 2008).
14. T. Clifton and J. D. Barrow, The power of general relativity, *Phys. Rev. D* **72**, p. 103005 (Nov 2005).
15. S. Capozziello and G. Lambiase, Higher order corrections to the effective gravitational action from Noether symmetry approach, *Gen. Rel. Grav.* **32**, 295 (2000).

16. S. Capozziello and G. Lambiase, Nonminimal derivative coupling and the recovering of cosmological constant, *Gen. Rel. Grav.* **31**, 1005 (1999).
17. S. Capozziello, G. Lambiase and H. Schmidt, Nonminimal derivative couplings and inflation in generalized theories of gravity, *Annalen Phys.* **9**, 39 (2000).
18. S. Nojiri and S. D. Odintsov, Inhomogeneous equation of state of the universe: Phantom era, future singularity, and crossing the phantom barrier, *Phys. Rev. D* **72**, p. 023003 (Jul 2005).
19. S. Capozziello, V. F. Cardone, E. Elizalde, S. Nojiri and S. D. Odintsov, Observational constraints on dark energy with generalized equations of state, *Phys. Rev. D* **73**, p. 043512 (Feb 2006).
20. I. Brevik, E. Elizalde, S. Nojiri and S. D. Odintsov, Viscous little rip cosmology, *Phys. Rev. D* **84**, p. 103508 (Nov 2011).
21. S. Nojiri and S. D. Odintsov, The New form of the equation of state for dark energy fluid and accelerating universe, *Phys. Lett. B* **639**, 144 (2006).
22. S. Nojiri and S. D. Odintsov, Non-singular modified gravity unifying inflation with late-time acceleration and universality of viscous ratio bound in F(R) theory, *Prog. Theor. Phys. Suppl.* **190**, 155 (2011).
23. G. Lambiase, Thermal leptogenesis in $f(R)$ cosmology, *Phys. Rev. D* **90**, p. 064050 (Sep 2014).
24. G. Lambiase, S. Mohanty and A. R. Prasanna, Neutrino coupling to cosmological background: A review on gravitational Baryo/Leptogenesis, *Int. J. Mod. Phys. D* **22**, p. 1330030 (2013).
25. G. Lambiase and G. Scarpetta, Baryogenesis in $f(R)$ theories of gravity, *Phys. Rev. D* **74**, p. 087504 (Oct 2006).
26. B. Jain, V. Vikram and J. Sakstein, Astrophysical Tests of Modified Gravity: Constraints from Distance Indicators in the Nearby Universe, *Astrophys. J.* **779**, p. 39 (2013).
27. L. Lombriser, A. Slosar, U. Seljak and W. Hu, Constraints on f(R) gravity from probing the large-scale structure, *Phys. Rev. D* **85**, p. 124038 (2012).
28. S. Ferraro, F. Schmidt and W. Hu, Cluster abundance in $f(R)$ gravity models, *Phys. Rev. D* **83**, p. 063503 (Mar 2011).
29. L. Lombriser, F. Schmidt, T. Baldauf, R. Mandelbaum, U. Seljak and R. E. Smith, Cluster density profiles as a test of modified gravity, *Phys. Rev. D* **85**, p. 102001 (May 2012).
30. F. Schmidt, A. Vikhlinin and W. Hu, Cluster constraints on $f(R)$ gravity, *Phys. Rev. D* **80**, p. 083505 (Oct 2009).
31. H. Motohashi, A. A. Starobinsky and J. Yokoyama, Cosmology Based on f(R) Gravity Admits 1 eV Sterile Neutrinos, *Phys. Rev. Lett.* **110**, p. 121302 (2013).
32. L. Buoninfante, G. Lambiase and L. Petruzziello, Quantum interference in external gravitational fields beyond General Relativity (4 2021).
33. A. Capolupo, G. Lambiase, A. Stabile and A. Stabile, Virial theorem in scalar tensor fourth order gravity, *Eur. Phys. J. C* **81**, p. 650 (2021).
34. V. A. S. V. Bittencourt, M. Blasone, F. Illuminati, G. Lambiase, G. G. Luciano and L. Petruzziello, Quantum nonlocality in extended theories of gravity, *Phys. Rev. D* **103**, p. 044051 (2021).
35. G. Lambiase, M. Sakellariadou and A. Stabile, Constraints on extended gravity models through gravitational wave emission, *JCAP* **03**, p. 014 (2021).
36. N. Bernal, A. Ghoshal, F. Hajkarim and G. Lambiase, Primordial Gravitational Wave Signals in Modified Cosmologies, *JCAP* **11**, p. 051 (2020).

37. G. Tino, L. Cacciapuoti, S. Capozziello, G. Lambiase and F. Sorrentino, Precision Gravity Tests and the Einstein Equivalence Principle, *Prog. Part. Nucl. Phys.* **112**, p. 103772 (2020).
38. M. Ruffert and H. T. Janka, Gamma-ray bursts from accreting black holes in neutron star mergers, *Astron. Astrophys.* **344**, 573 (April 1999).
39. R. Popham, S. E. Woosley and C. Fryer, Hyperaccreting black holes and gamma-ray bursts, *The Astrophysical Journal* **518**, 356 (Jun 1999).
40. T. D. Matteo, R. Perna and R. Narayan, Neutrino trapping and accretion models for gamma-ray bursts, *The Astrophysical Journal* **579**, 706 (Nov 2002).
41. S. Fujibayashi, Y. Sekiguchi, K. Kiuchi and M. Shibata, Properties of neutrino-driven ejecta from the remnant of a binary neutron star merger: Pure radiation hydrodynamics case, *The Astrophysical Journal* **846**, p. 114 (Sep 2017).
42. O. Just, M. Obergaulinger, H. T. Janka, A. Bauswein and N. Schwarz, Neutron-star merger ejecta as obstacles to neutrino-powered jets of gamma-ray bursts, *Astrophys. J. Lett.* **816**, p. L30 (2016).
43. F. Foucart, M. D. Duez, L. E. Kidder, R. Nguyen, H. P. Pfeiffer and M. A. Scheel, Evaluating radiation transport errors in merger simulations using a monte carlo algorithm, *Phys. Rev. D* **98**, p. 063007 (Sep 2018).
44. J. D. Salmonson and J. R. Wilson, General relativistic augmentation of neutrino pair annihilation energy deposition near neutron stars, *Astrophys. J.* **517**, 859 (1999).
45. A. Prasanna and S. Goswami, Energy deposition due to neutrino pair annihilation near rotating neutron stars, *Phys. Lett. B* **526**, 27 (2002).
46. G. Lambiase and L. Mastrototaro, Effects of modified theories of gravity on neutrino pair annihilation energy deposition near neutron stars, *ApJ.* **904**, 1 (2020).
47. J. Cooperstein, L. van den Horn and E. A. Baron, Neutrino flows in collapsing stars: a two-fluid model, *ApJ* **309**, p. 653 (1986).
48. J. Cooperstein, L. J. van den Horn and E. A. Baron, Neutrino Pair Energy Deposition in Supernovae, *ApJ* **321**, p. L129 (1987).
49. J. D. Salmonson and J. R. Wilson, Neutrino annihilation between binary neutron stars, *Astrophys. J.* **561**, 950 (2001).
50. R. Mallick, A. Bhattacharyya, S. K. Ghosh and S. Raha, General Relativistic effect on the energy deposition rate for neutrino pair annihilation above the equatorial plane along the symmetry axis near a rotating neutron star, *Int. J. Mod. Phys. E* **22**, p. 1350008 (2013).
51. J. Goodman, A. Dar and S. Nussinov, Neutrino Annihilation in type II Supernovae, *Astrophys. J. Lett.* **314**, L7 (1987).
52. S. Mukherjee and S. Chakraborty, Horndeski theories confront the Gravity Probe B experiment, *Phys. Rev. D* **97**, p. 124007 (2018).
53. T. P. Sotiriou and S.-Y. Zhou, Black hole hair in generalized scalar-tensor gravity: An explicit example, *Phys. Rev. D* **90**, p. 124063 (2014).
54. C. Brans and R. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev.* **124**, 925 (1961).
55. G. Lambiase and L. Mastrototaro, Neutrino pair annihilation ($\nu\bar{\nu} \to e^- e^+$) in the presence of quintessence surrounding a black hole (12 2020).

News and views regarding PSR J1757–1854, a highly-relativistic binary pulsar

A. D. Cameron* and M. Bailes

ARC Centre of Excellence for Gravitational Wave Discovery (OzGrav)
Centre for Astrophysics and Supercomputing, Swinburne University of Technology,
PO Box 218, Hawthorn VIC 3122, Australia
**E-mail: andrewcameron@swin.edu.au*

V. Balakrishnan, D. J. Champion, P. C. C. Freire, M. Kramer and N. Wex

Max-Planck-Institut für Radioastronomie (MPIfR)
Auf dem Hügel 69, D-53121 Bonn, Germany

S. Johnston

Australia Telescope National Facility, CSIRO Space and Astronomy
PO Box 76, Epping NSW 1710, Australia

A. G. Lyne and B. W. Stappers

Jodrell Bank Center for Astrophysics, University of Manchester
Alan Turing Building, Oxford Road, Manchester M13 9PL, UK

M. A. McLaughlin, N. Pol and H. Wahl

Department of Physics and Astronomy, West Virginia University
PO Box 6315, Morgantown, WV 26506, USA

C. Ng

Dunlap Institute for Astronomy & Astrophysics, University of Toronto
50 St. George Street, Toronto, ON M5S 3H4, Canada

A. Possenti

Istituto Nazionale di Astrofisica (INAF), Osservatorio Astronomico di Cagliari
Via della Scienza 5, 09047 Selargius (CA), Italy

A. Ridolfi

Istituto Nazionale di Astrofisica (INAF), Osservatorio Astronomico di Cagliari
Via della Scienza 5, 09047 Selargius (CA), Italy

Max-Planck-Institut für Radioastronomie (MPIfR)
Auf dem Hügel 69, D-53121 Bonn, Germany

We provide an update on the ongoing monitoring and study of the highly-relativistic double neutron star binary system PSR J1757–1854, a 21.5-ms pulsar in a highly eccentric, 4.4-hour orbit. The extreme nature of this pulsar's orbit allows it to probe a parameter space largely unexplored by other relativistic binary pulsars. For example, it displays

one of the highest gravitational wave (GW) luminosities of any known binary pulsar, as well as the highest rate of orbital decay due to GW damping. PSR J1757–1854 is also notable in that it is an excellent candidate for exploring new tests of General Relativity and other gravitational theories, with possible measurements of both Lense-Thirring precession and relativistic orbital deformation (through the post-Keplerian parameter δ_θ) anticipated within the next 3–5 years.

Here we present a summary of the latest interim results from the ongoing monitoring of this pulsar as part of an international, multi-telescope campaign. This includes an update of the pulsar's long-term timing and post-Keplerian parameters, new constraints on the pulsar's proper motion and corresponding Shklovskii kinematic correction, and new limits on the pulsar's geodetic precession as determined by monitoring for secular changes in the pulse profile. We also highlight prospects for future work, including an updated timeline on new relativistic tests following the introduction of MeerKAT observations.

Keywords: Gravitation; binaries: close; pulsars: individual: PSR J1757–1854

1. Introduction

For many decades, relativistic binary pulsars have been one of the key tools for studying different theories of gravity in the strong field regime, the current paradigm being Einstein's theory of General Relativity (GR). Discovered in 2016 as part of the HTRU-S Galactic Plane pulsar survey,[1,2] the 21.5-ms pulsar PSR J1757–1854 ranks as one of the most extreme examples of this class. Its high eccentricity ($e \simeq 0.606$) and compact orbit (measured via the projected semi-major axis, $x \simeq 2.24\,\mathrm{lt\text{-}s}$) around a neutron star companion combine some of the best properties of other notable relativistic binary pulsars, including the high eccentricity of PSR B1913+16[3] (the 7.75-hr orbital period *'Hulse-Taylor pulsar'*, with $e \simeq 0.617$) and the compact orbit of PSR J0737–3039A/B[4,5] (the 2.45-hr orbital period *'Double Pulsar'*, with $x \simeq 1.42\,\mathrm{lt\text{-}s}$). The extreme nature of PSR J1757–1854 is further demonstrated by multiple records set by the pulsar, including the highest acceleration ($\sim 700\,\mathrm{m\,s^{-2}}$) and highest relative velocity ($\sim 1000\,\mathrm{km\,s^{-1}}$) seen in a binary pulsar, as well as the highest rate of orbital decay due to gravitational wave damping and one of the shortest predicted merger times of any Galactic double neutron star system ($\sim 76\,\mathrm{Myr}$). Together, these properties make PSR J1757–1854 a highly promising candidate for providing new insights into theories of gravity.

Here, we describe some of the ongoing aspects of the work involving this pulsar, in advance of a more detailed peer-reviewed publication anticipated by mid-2022.

2. Current timing and post-Keplerian tests

Since its discovery in 2016, PSR J1757–1854 has been observed by multiple radio telescopes as part of an international collaboration, including the Parkes[a] (Australia), Effelsberg (Germany), Lovell (UK), Green Bank (GBT; USA) and MeerKAT (South Africa) telescopes. This data was analysed using a standard pulsar timing

[a] Also known by its indigenous Wiradjuri name *'Murriyang'*.

technique,[6] wherein summed, average pulse profiles are cross-correlated against high signal-to-noise (S/N) template profiles to produce a dataset of high-precision pulse 'times of arrival' (TOAs). Each TOA represents the mean arrival time of an integrated pulse profile from the pulsar. These are then iteratively fit using a *timing ephemeris* in order to develop a high-precision model of the pulsar's behaviour.

Currently, the GBT accounts for the overwhelming majority of PSR J1757–1854's TOAs, having produced over 22,000 TOAs since it began observing in mid-2016. Data from the GBT is recorded in a full-Stokes coherent search mode at both L and S-bands (800 MHz bandwidth centered at 1500 and 2000 MHz respectively). We currently record one 4.4-hr orbit at each frequency every two months.

Table 1. Latest timing parameters for PSR J1757–1854, based on data from the GBT and Parkes between MJD 57405–59363 and employing the DDH[7] binary model. Values in parentheses give the 1-σ uncertainties on the final digit.

Astrometric & spin parameters	
Right ascension, α (J2000)	17:57:03.78412(2)
Declination, δ (J2000)	−18:54:03.359(4)
Spin period, P (ms)	21.4972318900292(6)
Spin period derivative, $\dot{P}$	$2.627335(9) \times 10^{-18}$
Dispersion measure (pc cm^{-3})	378.203(2)
Proper motion in RA, μ_α (mas yr^{-1})	−4.36(12)
Proper motion in DEC, μ_δ (mas yr^{-1})	−0.8(14)
Period and position epoch (MJD)	57701
Orbital parameters	
Orbital period, P_b (d)	0.183537835854(7)
Eccentricity, e	0.6058171(3)
Projected semimajor axis, x (lt-s)	2.2378060(15)
Epoch of periastron, T_0 (MJD)	57700.92599421(3)
Longitude of periastron, ω (°)	279.34090(9)
Post-Keplerian parameters	
Rate of periastron advance, $\dot{\omega}$ (° yr^{-1})	10.36498(3)
Einstein delay, γ (ms)	3.5891(16)
Orbital period derivative, $\dot{P}_\mathrm{b}$	$-5.294(6)\times10^{-12}$
Orthometric amplitude, h_3 (μs)	5.08(18)
Orthometric ratio, ς	0.899(10)
Mass measurements (based on $\dot{\omega}$ and γ)	
Pulsar mass, m_p ($\mathrm{M}_\odot$)	1.3406(5)
Companion mass, m_c ($\mathrm{M}_\odot$)	1.3922(5)
Inclination angle, i (°)	85.0(2)

The most recent timing ephemeris for PSR J1757–1854 is given in Table 1. This includes five theory-independent post-Keplerian (PK) relativistic parameters, from which we can derive three independent tests of gravity. A mass-mass diagram showing the mass constraints imposed by each PK parameter under GR is shown

in Figure 1. By fixing a solution for the masses of the pulsar and the companion neutron star using the intersection of $\dot{\omega}$ and γ (the most precisely measured PK parameters), the further intersection of each remaining PK parameter provides one additional test of consistency. Both the orthometric amplitude and ratio parameters of the Shapiro delay[7] (h_3 and ς) are consistent to within 3.7% and 1.9% of their GR-predicted values respectively, and are approximately consistent with the $\dot{\omega} - \gamma$ intersection. However, although the observed value of the orbital period derivative $\dot{P}_\mathrm{b}$ is also consistent with GR to within at least 1%, its location on the mass-mass diagram is inconsistent with the constraints set by $\dot{\omega}$ and γ under the assumption of GR being correct. This is further discussed in Section 2.1.

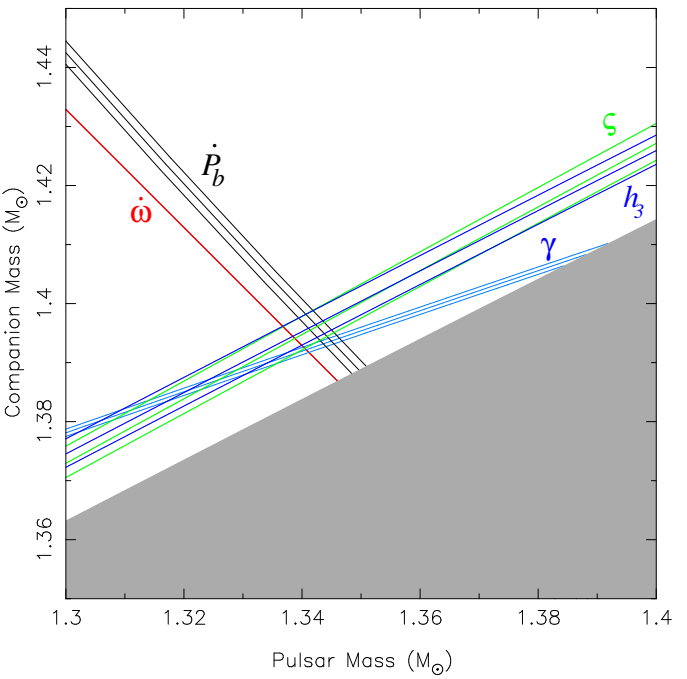


Fig. 1. Mass-mass diagram for PSR J1757–1854. Each coloured triplet of lines shows the constraints (with 1-σ uncertainty) placed by each of the measured PK parameters in Table 1 on the mass of the pulsar and the companion neutron star, in this case according to GR. The measured uncertainty of $\dot{\omega}$ is so small that it cannot be seen at this scale. The grey region in the bottom right is excluded due to orbital geometry.

2.1. *Proper motion and the radiative test of gravity*

A recent development in the timing of PSR J1757–1854 has been the first constrained measurement of the pulsar's proper motion, which is also listed in Table 1. The component in right ascension (μ_α) is detectable at approximately 36 σ, while the component in declination (μ_δ) remains poorly constrained, although its contribution

to the total proper motion remains small given its value relative to μ_α. Together, these values indicate a total proper motion of

$$\mu_{\rm T} = \sqrt{\mu_\alpha^2 + \mu_\delta^2} = 4.4\,(3)\ {\rm mas\,yr^{-1}}. \tag{1}$$

This allows us to attempt to quantify the contributions contaminating the observed value of the orbital period derivative, $\dot{P}_{\rm b,obs}$. We define the excess contribution as

$$\dot{P}_{\rm b,exs} = \dot{P}_{\rm b,obs} - \left(\dot{P}_{\rm b,GR} + \dot{P}_{\rm b,Gal} + \dot{P}_{\rm b,Shk}\right), \tag{2}$$

where $\dot{P}_{\rm b,GR}$ is the expected intrinsic contribution from GR (here calculated using the mass values derived from the observed $\dot{\omega}$ and γ as given in Table 1); $\dot{P}_{\rm b,Gal}$ is the contribution from the acceleration of the pulsar within the Galactic potential[8] (dependent on position, distance and the chosen model of the Galactic potential); and $\dot{P}_{\rm b,Shk}$ is the contribution from the Shklovskii effect[9] (dependent on distance and proper motion).

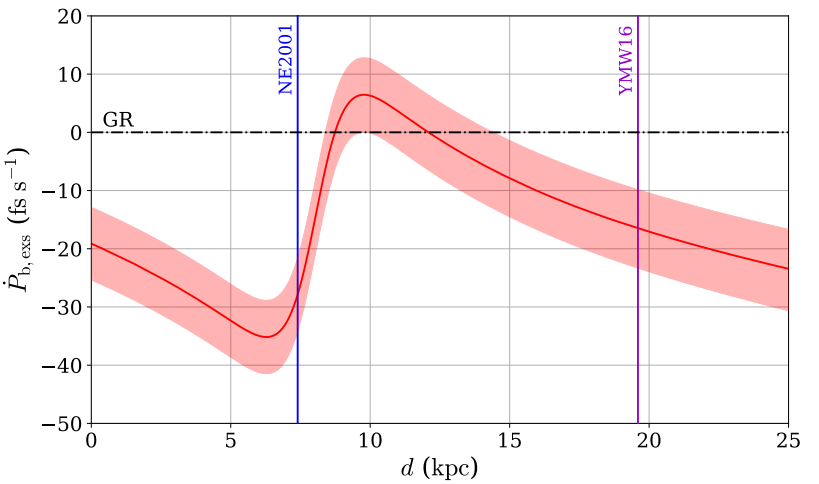


Fig. 2. The excess contribution to the orbital period derivative $\dot{P}_{\rm b,exs}$ as calculated by Equation 2, plotted (red) as a function of distance. The shaded region shows the 1-σ uncertainty. $\dot{P}_{\rm b,Gal}$ was calculated using the McMillan Galactic potential.[10] The vertical blue (left) and magenta (right) lines indicate the NE2001[11] and YWM16[12] DM-distance estimates respectively. The horizontal dashed black line shows the GR expectation of $\dot{P}_{\rm b,exs} = 0$.

For the radiative $\dot{P}_{\rm b}$ test of GR to be considered 'passed', $\dot{P}_{\rm b,exs}$ must be consistent with zero, given that the other contributions have been accurately accounted for. However, although we have now quantified the pulsar's proper motion, we still lack an accurate estimate of the pulsar's distance. This is demonstrated in Figure 2, where we show how $\dot{P}_{\rm b,exs}$ depends on the distance to PSR J1757–1854.

Assuming GR, the distance to PSR J1757–1854 should fall between approximately 8.5–14.5 kpc. Therefore, available distance estimates based on the pulsar's dispersion measure (DM), including both the NE2001[11] (7.4 kpc) and the YMW16[12] (19.6 kpc) electron density models, are currently excluded. This is also shown in Figure 2. However, we note that these DM-distance estimates typically come with large uncertainties,[13] making a definitive determination difficult. Additionally, the McMillan Galactic potential model[10] we have used in our analysis has known shortcomings at distances of approximately 7–9 kpc in the central regions of the Galaxy near PSR J1757–1854's position, further complicating this assessment.

Without an accurate distance to PSR J1757–1854, the utility of the radiative test of gravity is unfortunately limited. As seen in Figure 2, the unknown distance introduces an approximate uncertainty to $\dot{P}_{\rm b,exs}$ of at least $\pm 20 \times 10^{-15}$, more than three times the current uncertainty of $\dot{P}_{\rm b,obs} = 6 \times 10^{-15}$, such that the precision of the radiative test of GR is limited to about 0.4 %. We are currently exploring ways of further refining the estimated distance to PSR J1757–1854 (e.g. through VLBI astrometry). However, by assuming GR, we will still be able to inform our understanding of the pulsar's evolution. For example, the currently allowed distance range indicates a transverse velocity based on $\mu_{\rm T}$ between 180–300 $\mathrm{km\,s^{-1}}$, well in agreement with simulations of the neutron star kick associated with the second progenitor supernova.[2]

3. Attempts to detect the presence of geodetic precession

Our current understanding of the binary evolution of PSR J1757–1854 suggests the strong likelihood of a misalignment between the pulsar's spin vector and the orbital angular momentum vector.[2] As a result, we anticipate the presence of geodetic precession at a rate of $\Omega_{\rm geod} \simeq 3.07\,^\circ\,\mathrm{deg\,yr^{-1}}$, which if detected may serve as another test of GR. We have therefore analysed the current GBT dataset for evidence of changes in the pulse profile, which would indicate a changing line-of-sight through the emission cone of the pulsar as expected from geodetic precession.[14]

Our method adapts an approach previously applied[15] to PSR J1022+1001. Each GBT observation (at both L and S-bands) spanning greater than 80% the orbital period is summed in time, frequency and polarisation before being fit by a template composed of three component Gaussian curves. We then normally re-sample each profile bin using the off-pulse root-mean-squared noise value to generate at least 1,000 randomly 're-noised' profiles, to which the template is re-fit. From each new fitted template, the positions and amplitudes of the major and minor peaks are then measured, as well as the intercept positions at 10% and 50% of the profile's maximum amplitude. The mean and standard deviation of each parameter's statistical distribution across all fitted templates are then taken as the nominal value and its uncertainty.

This approach has for the first time confirmed the presence of statistically significant profile change. An example of this is given in Figure 3, which shows that

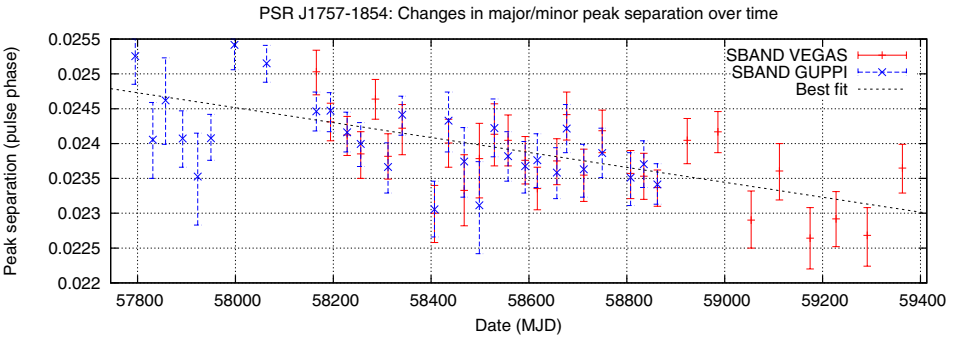


Fig. 3. Change in phase separation of the major and minor peaks of PSR J1757–1854's pulse profile, as measured by the GBT at S-band using the GUPPI and VEGAS backends. The line of best fit corresponds to a decreasing separation of $0.14\,(2)\,^{\circ}\,\mathrm{yr}^{-1}$.

the separation between the two peaks of PSR J1757–1854's profile (termed the 'major' and 'minor' peaks) is decreasing (i.e. moving closer together) by an average rate of approximately $0.14\,(2)\,^{\circ}\,\mathrm{yr}^{-1}$. The width of the profile at both 10% and 50% also appears to be shrinking at a commensurate rate, with all trends most pronounced at S-Band due to the decreased influence of scattering (with respect to lower-frequency L-band observations). The precise physical implications of these changes are to be explored in a future publication, but their detection provides fundamental evidence for the presence of geodetic precession, as well as evidence for the suspected spin/orbit misalignment present in the binary system.

4. Searches for the companion neutron star

If PSR J1757–1854's neutron star companion was confirmed to be an active radio pulsar, it would immensely enhance the scientific utility of this pulsar and its binary system. This is most obviously demonstrated by the prior example of the Double Pulsar, where the presence of the second pulsar PSR J0737–3039B allowed additional tests of gravity beyond those normally available, including the detection of geodetic precession[16] and an independent determination of the mass ratio.[17]

We have therefore conducted an ongoing search for evidence of pulsations from the neutron star companion to PSR J1757–1854. This search has focused on observations from the GBT, which are recorded in a coherently-dedispersed search mode so as to allow for optimal search sensitivity across all pulse periods. Given our precise knowledge of PSR J1757–1854's orbit as detailed in Table 1, each observation is re-sampled using the inferred orbit of the companion via the custom software package PYSOLATOR.[18] This allows the companion to appear stationary such that any pulsations would appear as a single narrow peak in a resulting Fourier spectrum. To account for the small degree of uncertainty in the companion orbit, different

re-sampled trials are produced using values of the mass ratio $q = m_{\rm p}/m_{\rm c}$ between $0.9554 < q < 0.9661$ in increments of 0.00005. The resulting Fourier power spectra from each observation are then stacked and summed, boosting the total S/N of any pulsations.

To date, this search has unfortunately returned negative results, although the complete GBT dataset has yet to be analysed. In total, approximately 389.7 hr of GBT observations are currently on file, although there is significant overlap in this dataset between the VEGAS and GUPPI backends during intervals when both were recorded simultaneously. Ongoing searches for the companion neutron star will continue, as the companion may become detectable in the event that it precesses into view in the near future, given the expected precession rate of $\Omega_{\rm geod} \simeq 2.97\,^\circ\,{\rm deg\,yr^{-1}}$.

5. Future observations and anticipated tests of gravity

5.1. *Addition of MeerKAT*

Although observations with South Africa's MeerKAT telescope have been ongoing since 2019 as part of the MeerTIME program,[19] all these observations took place with MeerKAT's L-band receiver, which has a large 856 MHz bandwidth centered at 1283.5 MHz. This frequency band is located at lower frequencies than those of the GBT L-band receiver and therefore is much more susceptible to pulse broadening via scattering than the GBT L-band.[20] As a consequence, even though MeerKAT can achieve about twice the S/N of the GBT for this target, the available timing precision from both telescopes is roughly equivalent.

These effects are expected to be mitigated as part of the ongoing rollout of a fleet of MeerKAT S-band receivers (with a bandwidth of approximately 1750 MHz centred at 2625 MHz), designed by the MPIfR. The decreased influence of scattering at these higher frequencies combined with the wider bandwidth and the pulsar's relatively flat spectral index suggest an improvement in TOA precision to within 10 μs or better (compared with a typical TOA precision of 25–30 μs currently achievable using similar TOA integration times with the current GBT and MeerKAT receivers). This will provide a significant advantage in the ongoing timing of the pulsar.

5.2. *Future tests of gravity*

PSR J1757–1854 is expected to provide new insight into two rarely-explored tests of gravity within the next few years. The first of these is the *relativistic orbital deformation*, characterised by the PK parameter δ_θ, which describes the deviation of the orbit from a pure Keplerian ellipse due to the effects of relativity. The measurability of δ_θ scales with both eccentricity and $\dot{\omega}$, positioning PSR J1757–1854 as an ideal system for providing constraint in the short-term. Meanwhile, the Hulse-Taylor pulsar[21] (lower $\dot{\omega}$) and the Double Pulsar (lower eccentricity) have to date only weakly constrained δ_θ despite having been studied for far longer.

The second anticipated test involves the detection of Lense-Thirring precession, which to date has not been detected in a double neutron star binary system. We intend to detect this effect via the measurement of a change in the semi-major axis,[22] $\dot{x}_{\rm LT}$. Although in general $\dot{x}_{\rm LT}$ becomes negligible for small misalignment angles between the pulsar spin and orbital angular momentum vectors, our recent detection of geodetic precession confirms that a significant misalignment is likely. The Lense-Thirring effect is also proportional to the neutron star moment of inertia; should the Lense-Thirring effect be detected here, it would lead to a rare measurement of this quantity and thereby provide insight into the neutron star equation of state.

The timescale over which each of these parameters is anticipated to be constrained was determined by a simulation of the anticipated timing campaigns. A simulation was produced by extrapolating the current GBT timing campaign until the end of 2024, and including an equivalent MeerKAT S-band campaign comprising one full orbit recorded every two months, starting in 2022. Timing ephemerides were fit to the data at set intervals, while averaging the results of 30 realisations of the dataset. Based on these simulations, we anticipate the achievement of a 3-σ constraint of δ_θ between approximately 2024–2025, and of $\dot{x}_{\rm LT}$ between 2025–2026.

Acknowledgments

The Parkes radio telescope is part of the Australia Telescope National Facility which is funded by the Australian Government for operation as a National Facility managed by CSIRO. We acknowledge the Wiradjuri people as the traditional owners of the Observatory site. The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. Pulsar research at the Jodrell Bank Centre for Astrophysics and the observations using the Lovell Telescope are supported by a consolidated grant from the STFC in the UK. This work is also based on observations with the 100-m telescope of the Max-Planck-Institut für Radioastronomie at Effelsberg. AC and MB acknowledge OzGrav ARC grant CE170100004. AP and AR acknowledge that (part of) this work has been funded using resources from the research grant "iPeska" (P.I. Andrea Possenti) funded under the INAF national call Prin-SKA/CTA approved with the Presidential Decree 70/2016. AP and AR also acknowledge the support from the Ministero degli Affari Esteri e della Cooperazione Internazionale - Direzione Generale per la Promozione del Sistema Paese - Progetto di Grande Rilevanza ZA18GR02. AR acknowledges continuing valuable support from the Max-Planck Society. MAM and HW are members of the NANOGrav Physics Frontiers Center and supported by NSF awards 1430284 and 2020265.

References

1. M. J. Keith, A. Jameson, W. van Straten, M. Bailes, S. Johnston, M. Kramer, A. Possenti, S. D. Bates, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, N. D'Amico, L. Levin, P. L. McMahon, S. Milia and B. W. Stappers, The High Time Resolution Universe Pulsar Survey - I. System configuration and initial discoveries, *Monthly Notices of the Royal Astronomical Society* **409**, 619 (December 2010).
2. A. D. Cameron, D. J. Champion, M. Kramer, M. Bailes, E. D. Barr, C. G. Bassa, S. Bhandari, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, R. P. Eatough, C. M. L. Flynn, P. C. C. Freire, A. Jameson, S. Johnston, R. Karuppusamy, M. J. Keith, L. Levin, D. R. Lorimer, A. G. Lyne, M. A. McLaughlin, C. Ng, E. Petroff, A. Possenti, A. Ridolfi, B. W. Stappers, W. van Straten, T. M. Tauris, C. Tiburzi and N. Wex, The High Time Resolution Universe Pulsar Survey - XIII. PSR J1757-1854, the most accelerated binary pulsar, *Monthly Notices of the Royal Astronomical Society* **475**, L57 (March 2018).
3. R. A. Hulse and J. H. Taylor, Discovery of a pulsar in a binary system, *Astrophysical Journal* **195**, L51 (January 1975).
4. M. Burgay, N. D'Amico, A. Possenti, R. N. Manchester, A. G. Lyne, B. C. Joshi, M. A. McLaughlin, M. Kramer, J. M. Sarkissian, F. Camilo, V. Kalogera, C. Kim and D. R. Lorimer, An increased estimate of the merger rate of double neutron stars from observations of a highly relativistic system, *Nature* **426**, 531 (December 2003).
5. A. G. Lyne, M. Burgay, M. Kramer, A. Possenti, R. N. Manchester, F. Camilo, M. A. McLaughlin, D. R. Lorimer, N. D'Amico, B. C. Joshi, J. Reynolds and P. C. C. Freire, A Double-Pulsar System: A Rare Laboratory for Relativistic Gravity and Plasma Physics, *Science* **303**, 1153 (February 2004).
6. D. R. Lorimer and M. Kramer, *Handbook of Pulsar Astronomy* (Cambridge Univ. Press, Cambridge, 2005).
7. P. C. C. Freire and N. Wex, The orthometric parametrization of the Shapiro delay and an improved test of general relativity with binary pulsars, *Monthly Notices of the Royal Astronomical Society* **409**, 199 (November 2010).
8. T. Damour and J. H. Taylor, On the Orbital Period Change of the Binary Pulsar PSR 1913+16, *Astrophysical Journal* **366**, p. 501 (January 1991).
9. I. S. Shklovskii, Possible Causes of the Secular Increase in Pulsar Periods, *Soviet Astronomy* **13**, p. 562 (feb 1970).
10. P. J. McMillan, The mass distribution and gravitational potential of the Milky Way, *Monthly Notices of the Royal Astronomical Society* **465**, 76 (February 2017).
11. J. M. Cordes and T. J. W. Lazio, NE2001.I. A New Model for the Galactic Distribution of Free Electrons and its Fluctuations, *arXiv e-prints* (July 2002).
12. J. M. Yao, R. N. Manchester and N. Wang, A New Electron-density Model for Estimation of Pulsar and FRB Distances, *Astrophysical Journal* **835**, p. 29 (January 2017).
13. A. D. Cameron, D. J. Champion, M. Bailes, V. Balakrishnan, E. D. Barr, C. G. Bassa, S. Bates, S. Bhandari, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, C. M. L. Flynn, A. Jameson, S. Johnston, M. J. Keith, M. Kramer, L. Levin, A. G. Lyne, C. Ng, E. Petroff, A. Possenti, D. A. Smith, B. W. Stappers, W. van Straten, C. Tiburzi and J. Wu, The High Time Resolution Universe Pulsar Survey - XVI. Discovery and timing of 40 pulsars from the southern Galactic plane, *Monthly Notices of the Royal Astronomical Society* **493**, 1063 (March 2020).
14. M. Bailes, Geodetic precession in binary pulsars, *Astronomy and Astrophysics* **202**, 109 (August 1988).

15. P. V. Padmanabh, E. D. Barr, D. J. Champion, R. Karuppusamy, M. Kramer, A. Jessner and P. Lazarus, Revisiting profile instability of PSR J1022+1001, *Monthly Notices of the Royal Astronomical Society* **500**, 1178 (January 2021).
16. R. P. Breton, V. M. Kaspi, M. Kramer, M. A. McLaughlin, M. Lyutikov, S. M. Ransom, I. H. Stairs, R. D. Ferdman, F. Camilo and A. Possenti, Relativistic Spin Precession in the Double Pulsar, *Science* **321**, p. 104 (July 2008).
17. M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire and F. Camilo, Tests of General Relativity from Timing the Double Pulsar, *Science* **314**, 97 (October 2006).
18. A. Ridolfi, PYSOLATOR: Remove orbital modulation from a binary pulsar and/or its companion (March 2020).
19. M. Bailes, A. Jameson, F. Abbate, E. D. Barr, N. D. R. Bhat, L. Bondonneau, M. Burgay, S. J. Buchner, F. Camilo, D. J. Champion, I. Cognard, P. B. Demorest, P. C. C. Freire, T. Gautam, M. Geyer, J. M. Griessmeier, L. Guillemot, H. Hu, F. Jankowski, S. Johnston, A. Karastergiou, R. Karuppusamy, D. Kaur, M. J. Keith, M. Kramer, J. van Leeuwen, M. E. Lower, Y. Maan, M. A. McLaughlin, B. W. Meyers, S. Osłowski, L. S. Oswald, A. Parthasarathy, T. Pennucci, B. Posselt, A. Possenti, S. M. Ransom, D. J. Reardon, A. Ridolfi, C. T. G. Schollar, M. Serylak, G. Shaifullah, M. Shamohammadi, R. M. Shannon, C. Sobey, X. Song, R. Spiewak, I. H. Stairs, B. W. Stappers, W. van Straten, A. Szary, G. Theureau, V. Venkatraman Krishnan, P. Weltevrede, N. Wex, T. D. Abbott, G. B. Adams, J. P. Burger, R. R. G. Gamatham, M. Gouws, D. M. Horn, B. Hugo, A. F. Joubert, J. R. Manley, K. McAlpine, S. S. Passmoor, A. Peens-Hough, Z. R. Ramudzuli, A. Rust, S. Salie, L. C. Schwardt, R. Siebrits, G. Van Tonder, V. Van Tonder and M. G. Welz, The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime, *Publications of the Astronomical Society of Australia* **37**, p. e028 (July 2020).
20. M. Kramer, I. H. Stairs, V. Venkatraman Krishnan, P. C. C. Freire, F. Abbate, M. Bailes, M. Burgay, S. Buchner, D. J. Champion, I. Cognard, T. Gautam, M. Geyer, L. Guillemot, H. Hu, G. Janssen, M. E. Lower, A. Parthasarathy, A. Possenti, S. Ransom, D. J. Reardon, A. Ridolfi, M. Serylak, R. M. Shannon, R. Spiewak, G. Theureau, W. van Straten, N. Wex, L. S. Oswald, B. Posselt, C. Sobey, E. D. Barr, F. Camilo, B. Hugo, A. Jameson, S. Johnston, A. Karastergiou, M. Keith and S. Osłowski, The relativistic binary programme on MeerKAT: science objectives and first results, *Monthly Notices of the Royal Astronomical Society* **504**, 2094 (June 2021).
21. J. M. Weisberg and Y. Huang, Relativistic Measurements from Timing the Binary Pulsar PSR B1913+16, *Astrophysical Journal* **829**, p. 55 (September 2016).
22. T. Damour and J. H. Taylor, Strong-field tests of relativistic gravity and binary pulsars, *Physical Review D (Particles, Fields, Gravitation, and Cosmology)* **45**, 1840 (March 1992).

On the origin of the unique isolated X-Ray pulsar 1E 161348-5055 with 6.7 hr. spin period

V. Yu. Kim

Fesenkov Astrophysical Institute
Almaty, 050020, Observatroriya 23, Kazakhstan
E-mail: kim@aphi.kz
www.aphi.kz

A scenario for the formation of an isolated X-ray pulsar 1E161348-5055 with an anomalously long period of 6.7 hours is proposed. It is shown that this pulsar can be a descendant of a massive X-ray binary system, which disintegrated about 2000 years ago after a supernova explosion caused by the core collapse of a massive component. X-ray radiation of this object in the present epoch is generated as a result of accretion of matter onto (about 10 million years old) neutron star from the residual non-Keplerian accretion disk. The pulsar's nebula RCW 103 is a supernova remnant formed by the explosion of its massive companion in the final evolutionary phase of a massive binary system.

Keywords: X-Ray pulsar, isolated neutron star, accretion, stellar evolution

Introduction

X-ray pulsar 1E 161348-5055 (hereinafter 1E 1613) was discovered in 1979 by means of the Einstein cosmic observatory, as a point X-ray source in the supernova remnant RCW 103.[1] The distance to RCW 103 was approximately estimated as $d \simeq 3.2 \pm 0.1$ kpc from the analysis of absorption of radio emission by neutral hydrogen in the 21 cm line.[2] The estimated age of RCW 103 from the expansion rate and size of the nebula is about $\tau \simeq 2000 \pm 1000$ yr.[2] The nebula is an atypical remnant of a SN II type. It has an almost rounded shape with a fibrous structure, a low expansion rate (~ 1100 km/s) and it has a relatively small spatial size 7.7 pc for its age.[3] Nebulae with such parameters account for less than 20 % of the currently known SN II supernova remnants, the explosion of which apparently occurred in a high-density gaseous medium.[4] After the discovery of this pulsar, it was suggested that this object is an isolated cooling neutron star in the peculiar supernova remnant RCW 103.[5] However, in the period 1997-1999, during the study of 1E 1613 by the ASCA space observatory, significant variations in the source brightness were detected and the thermal component emitted by a hot spot $kT \sim 0.6 - 0.8$ keV with a radius of $a_{\rm p} \sim 600$ m,[6,7] was found. It is not typical for a cooling neutron star, but typical for magnetized neutron stars accreting matter on their surface. This served as the basis for the hypothesis that 1E 1613 is a neutron star accreting matter in the area of the magnetic poles. The average X-ray luminosity of the source is $L_{\rm x} \sim 10^{34}$ erg/s.[6,7] The report about the possible X-Ray variability of 1E 1613 with a period of ~ 6 hours, discovered from the results of observations

of this source by the "Chandra" observatory, stimulated the search for an optical companion of the neutron star. Observations made for this purpose with the ESO VLT telescope in the near infrared, however, allowed to establish only the upper limit of the luminosity of a hypothetical companion $L_{\rm ir} \sim 10^{31}$ erg/s, which could be a star of a later spectral class than M4.[8] This result became a strong argument in favor of the fact that 1E 1613 is not part of a close binary system and it is an isolated neutron star, whose radiation is caused either by accretion of matter from the residual disk or by the rapid dissipation of a superstrong magnetic field.[9]

In 2006, by means of the XMM-Newton space telescope, it was possible to establish the spin period of 1E 1613 $P_{\rm s} \simeq 6.67 \pm 0.003$ hr.[10] None of the known pulsars had such a long spin period. Almost immediately after the discovery of the period, several hypotheses were put forward about the nature of object 1E 1613. One of the first idea was the assumption that this object is a unique magnetar accreting matter from the surrounding residual (fall-back) disk, with a magnetic field $> 10^{15}$ G.[9,10] It was also hypothesized that 1E 1613 is a magnetar in a low-mass close binary system.[5] Subsequently, it was suggested that 1E 1613 is a millisecond pulsar in a close binary system, where the observed period (6.7 hours) was assumed as an orbital period (the so-called period mimicry).[11]

In 2011, from the analysis of the pulsation profiles obtained from space observatories: "Swift", "Chandra", "XMM-Newton", the upper limit of the rotation deceleration was calculated $\dot{P}_{\rm s} \leq 1.6 \times 10^{-9}$ s/s or $\dot{\nu}_{\rm s} \leq -2.8 \times 10^{18}$ Hz/s (here $\nu_{\rm s} = 1/P_{\rm s}$ - spin frequency).[12] Analyzing their result, the authors of this discovery noted that the moment of force applied to a neutron star in the present epoch is significantly less than the value expected in all previously proposed models of this source. The high stability of pulsations apparently indicates that the spin period of the star at the present epoch is close to the equilibrium period of its spin rotation, the value of which weakly depends on changes in the accretion rate in the residual disk. It was shown in the article[13] that such a situation is realized in the scenario of Magnetic Levitation accretion, in which a neutron star with a magnetic field 10^{12} G accrets onto its surface matter from the residual non-Keplerian magnetic disk.

1. Formation of an isolated pulsar during the decay of a binary system

We consider a situation in which 1E 1613 is a neutron star, born in a close massive binary system during the first supernova explosion. We assume that this did not lead to the disintegration of the system and the neutron star formed a close pair with its massive companion. The initial rotation period of the neutron star was fractions of a second and increased as the star successively passed the ejector and propeller states. Upon completion of this phase, the star passed into an accretor state, in which it remained until the second supernova explosion, caused by the collapse of the core of its massive companion.This event most likely led to the disintegration of the system,[14] and the old neutron star passed into the state of an isolated pulsar,

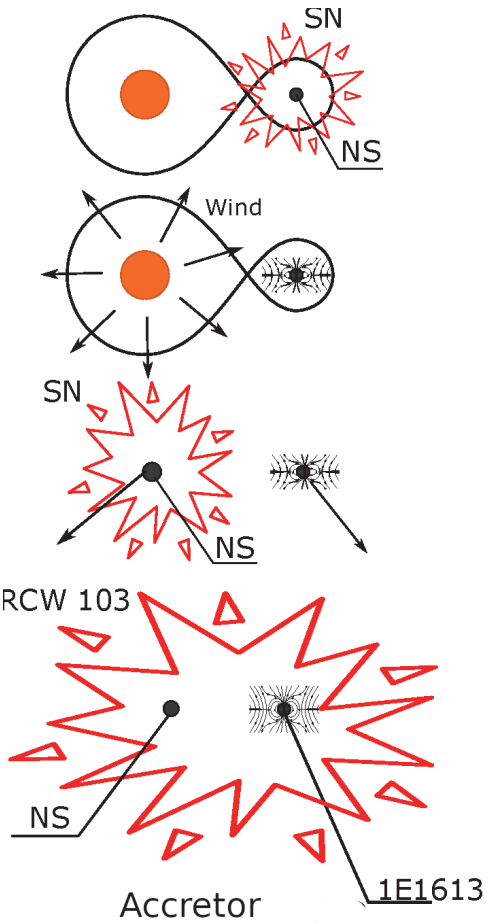


Fig. 1. Possible way of formation of isolated X-ray pulsar as a descendant of high-mass X-ray binary system

immersed in the remnant of its companion supernova, observed in the present epoch as the RCW 103 nebula.

2. Magnetorotational evolution

The scenario of the evolution of a high-mass X-ray binary system (HMXB) described above has been sufficiently studied by now and is perceived by most authors as canonical (see the article [Postnov et al., 2014][15] and the literature cited there). The lifetime of such a system is estimated by the evolution time of its optical component on the main sequence with mass $M_{\rm opt}$ corresponds to[16]:

$$t_{\rm ms} \simeq 6 \times 10^6\, yr \times \left(\frac{M_{\rm opt}}{M_\odot}\right)^{-5/2} \qquad (1)$$

Numerical modeling of the magnetorotational evolution of a neutron star in the framework of such a scenario with allowance for the dissipation of its magnetic field was carried out by Urpin et al.[17] According to the results of their calculations, the magnetic field of the star during $t_{\rm ms}$ decreases by an order of magnitude due to diffusion and accretion screening.[18] In particular, with an initial value of $\sim 10^{13}$ G, the magnetic field on the surface of a star that has evolved as part of HMXB system is $\sim 10^{12}$ G. Such a neutron star spends most of its evolution in the ejector state (see formula 17 in the paper[19]):

$$t_{\rm ej} \simeq 3 \times 10^6\, yr\ \times \mu_{30.5}^{-1}\, I_{45}\, \dot{M}_{14}^{1/2}\, \upsilon_7^{-1/2} \tag{2}$$

Here: $\mu_{30.5} = \mu_{\rm ns}/3 \times 10^{30}$ G cm^3 - normalized magnetic dipole moment of a neutron star, $I_{45} = I_{\rm ns}/10^{45}$ g cm^2 - its moment of inertia, $\dot{M}_{14} = \dot{M}/10^{14}$ g/s - normalized accretion rate, $\upsilon_7 = \upsilon_{\rm rel}/10^7$ cm/s - related velocity of a neutron star into surrounded medium. During this stage, its magnetic field decreases by an average of 3 times (the rapid cooling of the star during this phase significantly reduces the rate of diffusion of the magnetic field in its crust[17]).

As the rotation period of the star reaches a critical value (see formula 16 in the paper[19]), it passes into the propeller state.

$$P_{\rm ej} \sim 1.2\,{\rm s}\ \times \mu_{30.5}^{1/2}\, \dot{M}_{14}^{-1/4}\, \upsilon_7^{-1/4} \tag{3}$$

At this stage, the spin-down torque applied to the star by the penetrating under its Bondi radius gas is determined by the expression $K_{\rm sd}^{\rm pr} \sim \mu_{\rm ns}^2/r_{\rm m}^3$. The duration of the propeller stage substantially depends on the geometry and physical parameters of the accretion flow.

$$\tau_{\rm pr} = \frac{\pi\, I\, r_{\rm m}^3}{\mu_{\rm ns}\, P_{\rm ej}} \tag{4}$$

This parameter reaches its minimum possible value in the case of the scenario of Magnetic Levitation accretion:[13]

$$\tau_{\rm pr}^{\rm (ml)} \simeq 6 \times 10^3\, yr\ \times \alpha_{0.1}^{6/13}\, \mu_{30}^{-29/26}\, I_{45}\, m^{3/13}\, T_6^{-6/13}\, \dot{M}_{14}^{-35/52}\, \upsilon_7^{1/4} \tag{5}$$

In the paper[13] we showed that 1E 1613 could be in the accretor stage, the material for the accretion is supplied by the surrounding residual Magnetic Levitating disk or ML-disk formed from the supernova remnant RCW 103. The magnetic field strength 1E 1613 in this model (Magnetic Levitation accretion) corresponded to 10^{12} G, which corresponds to the canonical value of the magnetic field of neutron stars. For detailed explanation, see articles [Ikhsanov et al., 2013, 2015].[13,20]

3. Conclusions

The main conclusion of this study is to explain the origin of an isolated X-ray pulsar with a period of 6.7 hours within the canonical scenario of the evolution of a massive X-ray binary system without invoking the hypothesis of superstrong

magnetic fields on the surface of a neutron star. Our proposed model is consistent with modern concepts of a relatively small value of the initial rotation period of neutron stars and does not require invoking the additional assumption of fall-back accretion onto a neutron star after its birth.

The age of a neutron star 1E 161348-5055, manifesting itself in the present epoch as an isolated X-ray pulsar, is comparable to the lifetime of a massive X-ray binary system, estimated by Eq. (1) and amounts to several million years. The lifetime of this star in the state of an isolated X-ray pulsar, however, is much shorter. It corresponds to the time elapsed since the collapse of the massive X-ray binary, which is comparable to the age of the RCW 103 nebula and is about 2000 years (see also the article[20]).

The residual accretion disk surrounding the magnetosphere of a neutron star in the present epoch could have been formed during its evolution as part of a massive X-ray binary system or after the disintegration of the system during the capture of matter by an old neutron star from an envelope ejected by its massive component during a supernova explosion.

Isolated X-ray pulsars with long periods, within the framework of our model, are descendants of the widest pairs in which a massive component at the last stage of its evolution does not fill its Roche lobe and the capture of matter from the stellar wind remains the dominant mechanism of mass exchange between the components of the system at all stages of its evolution. The lifetime of such a pulsar depends on the mass of the residual disk and the physical parameters of the gas, which determine the rate of diffusion of the magnetic field in it. The main candidates for descendants of close binary systems are isolated X-ray pulsars located in the supernova remnants formed during the explosion of massive companions.

Acknowledgments

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP09258811).

References

1. I. Tuohy and G. Garmire, Discovery of a compact X-ray source at the center of the SNR RCW 103, *Astrophys. J.* **239**, L107 (1980).
2. J. L. Caswell, J. D. Murray, R. S. Roger, D. J. Cole and D. J. Cooke, Neutral hydrogen absorption measurements yielding kinematic distances for 42 continuum sources in the galactic plane, *Astron. and Astrophys.* **45**, 239 (1975).
3. L. Carter, J. Dickel and D. Bomans, Expansion of the Supernova Remnant RCW 103, *Publ. Astron. Soc. Pacific* **109**, 990 (1997).
4. D. Marsden, R. E. Lingenfelter, R. E. Rothschild, and J. C. Higdon, Nature versus Nurture: The Origin of Soft Gamma-Ray Repeaters and Anomalous X-ray Pulsars, *Astrophys. J.* **550**, 397 (2001).
5. F. Pizzolato, M. Colpi, A. De Luca, S. Mereghetti, A. Tiengo, 1E 161348-5055 in the Supernova Remnant RCW 103: A Magnetar in a Young Low-Mass Binary System?, *Astrophys. J.* **681**, 530 (2008).

6. E. V. Gotthelf, R. Petre and U. Hwang, The Nature of the Radio-quiet Compact X-ray Source in Supernova Remnant RCW 103, *Astrophys. J.* **487**, L175 (1997).
7. E. V. Gotthelf, R. Petre and G. Vasisht, X-ray Variability from the Compact Source in the Supernova Remnant RCW 103, *Astrophys. J.* **514**, L107 (1999).
8. G. G. Pavlov, D. Sanwal and M.A. Teter, Young Neutron Stars and Their Environments, in *IAU Symp. 218*, p. 239, eds. F. Camilo and B. M. Gaensler (San Francisco: ASP, 2004).
9. X.-D. Li, The Nature of the Compact X-ray Source in Supernova Remnant RCW 103, *Astrophys. J.* **666**, L81 (2007).
10. A. de Luca, P. A. Caraveo, S. Mereghetti, A. Tiengo and G. F. Bignami, A Long-Period, Violently Variable X-ray Source in a Young Supernova Remnant, *Science* **313**, 814 (2006).
11. H. Bhadkamkar and P. Ghosh, Young pre-low-mass X-ray binaries in the propeller phase. Nature of the 6.7-h periodic X-ray source 1E 161348-5055 in RCW 103, *Astron. and Astrophys.* **1297**, 506 (2009).
12. P. Esposito, R. Turollia, A. De Luca, G. L. Israel, A. Possenti and D. N. Burrows, Swift monitoring of the central X-ray source in RCW 103, *Monthly. Not. Roy. Astron. Soc.* **418**, 170 (2011).
13. N. R. Ikhsanov, V. Yu. Kim, N. G. Beskrovnaya and L. A. Pustil'nik, A new look at the origin of the 6.67 hr period X-ray pulsar 1E 161348-5055, *Astrophys. and Space Sci.* **346**, 105 (2013).
14. S. B. Popov and M. E. Prokhorov, Progenitors with enhanced rotation and the origin of magnetars, *Monthly. Not. Roy. Astron. Soc.* **367**, 732 (2006).
15. K. A. Postnov and L. R. Yungelson, The Evolution of Compact Binary Star Systems, *Living Rev. in Relativity* **17**, 3 (2014).
16. D. Bhattacharya and E. P. J. van den Heuvel, Formation and evolution of binary and millisecond radio pulsars, *Phys. Rep.* **203**, 1 (1991).
17. V. Urpin, D. Konenkov and U. Geppert, Evolution of neutron stars in high-mass X-ray binaries, *Monthly. Not. Roy. Astron. Soc.* **299**, 73 (1998).
18. G. S. Bisnovatyi-Kogan and B. V. Komberg, Pulsars and close binary systems, *Soviet Astronomy* **51**, 373 (1974).
19. N. R. Ikhsanov, Signs of magnetic accretion in the young Be/X-ray pulsar SXP 1062, *Monthly. Not. Roy. Astron. Soc.* **434**, L39 (2012).
20. N. R. Ikhsanov, V. Yu. Kim and N. G. Beskrovnaya, A scenario of the formation of isolated X-ray pulsars with anomalously long period, *Astronomy Reports* **59**, 25 (2015).

Advantages of including globular cluster millisecond pulsars in Pulsar Timing Arrays

M. Maiorano*, F. De Paolis, A. A. Nucita and A. Franco
Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Via Arnesano, I-73100 Lecce, Italy, INFN, Sezione di Lecce, Via Arnesano, I-73100 Lecce, Italy
* *E-mail: michele.maiorano@le.infn.com*

Even though Pulsar Timing Arrays already have the potential to detect the gravitational wave background by finding a quadrupole correlation in the timing residuals, this goal has not yet been achieved. Motivated by some theoretical arguments, we analyzed some advantages of including the millisecond pulsars within globular clusters, especially those in their cores, in current and future Pulsar Timing Array projects for detecting the gravitational waves emitted by an ensemble of supermassive black holes.

Keywords: Black Holes; Gravitational Waves; Pulsars.

1. Introduction

Within the context of the Albert Einstein theory of General Relativity,[1] gravitational waves (GWs) are space perturbations propagating at the speed of light in the vacuum. Due to the smallness of their amplitude, which in most cases is less than 10^{-16}[a], their detection is a challenging problem and, in fact, it took a hundred years since they were predicted for the first time. Nowadays, ground-based gravitational interferometers, such as those of the LIGO (Laser Interferometer Gravitational-Wave Observatory), KAGRA (Kamioka Gravitational Wave Detector) and VIRGO collaboration, are currently the only instruments capable of efficiently detecting GWs. Their sensitivity allows to detect high-frequency (i.e. in the range 10–10^3 Hz) GWs emitted by compact object systems, such as black hole binaries,[2] neutron star binaries[3] and neutron star-black hole binaries,[4] in the final phase of the merger. The detection of low-frequency (i.e. in the frequency range $\simeq 10^{-5}$–1 Hz) GWs is more challenging because it requires space-based gravitational interferometers such as LISA (Laser Interferometer Space Antenna), which is planned to be launched for 2034 and consists of a constellation of three satellites, separated by about 1.5×10^6 km, in a triangular configuration, orbiting with Earth around the Sun.[5,6] However, ultra-low frequency (i.e. in the frequency range $\simeq 10^{-10}$–10^{-6} Hz) GWs, that may be generated by sources relevant for astrophysics and cosmology, such as supermassive black hole binaries (SMBHBs)[7] and cosmic strings,[8] and form the gravitational

[a]This is a dimensionless quantity describing how a spatial dimension is stretched or squeezed by the passage of a GW.

wave background (GWB), are out of the range of any kind of gravitational interferometers and can only be detected by Pulsar Timing Arrays (PTAs)[b].

PTAs constantly monitor the pulsed radio emission from the most stable isolated and binary millisecond pulsars across the sky to detect variation between the observed pulse time of arrival (ToA) and the one expected from the timing model. Such variation, referred to as timing residuals, may be GW-induced. Indeed, GWs can modify the distance between the Earth and the MSP, causing advances or delays in the pulse ToAs.[10,11] As GWs, many other effects may induce timing residuals, so they must be distinguished from GWs. At present, the only way of doing this is by verifying that the cross-correlations between the timing residuals of each MSP pair follow the Hellings and Downs function, which is due to the quadrupole nature of GWs.[12]

The main PTA collaborations are the European Pulsar Timing Array (EPTA), the Indian Pulsar Timing Array (InPTA), the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), and the Parkes Pulsar Timing Array (PPTA), and all of them are part of the International Pulsar Timing Array (IPTA). These PTAs have timed about forty MSPs for more than ten years with an accuracy that should be sufficient, in principle, to detect gravitational waves. Some clues of the presence of a common red process compatible with a GWB was found, but there is still not a strong enough evidence for or against it.[13–15] The reason behind that is yet unclear. Most likely, this could be due to the small number of MSPs available for PTAs, so continuing to add suitable MSPs should allow claiming detection, or to an insufficient observation time spawn, so it is essential to continue collecting data for many more years. Another option is that the standard GW search technique needs to be complemented by independent methods.

2. Back to the Theory

The GW-induced timing residuals are described, in a compact form, by[c]

$$r(t,\Omega^i) = \mathrm{Re}\left\{ \frac{h}{j\omega} \sum_{A=+,\times} \frac{1}{2} \frac{p^i p^j e^A_{ij}}{(1+\Omega_i p^i)} \left[e^{j(\omega\tilde{t}-\alpha^A)} - e^{j(\omega\tilde{t}-\omega\tau(1-\Omega_i p^i)-\alpha^A)} \right]_0^t \right\} \quad (1)$$

where h, ω, A, e^A_{ij}, and α^A indicate the GW amplitude, angular frequency, polarization state, polarization tensor, and the initial phase, respectively, $\tilde{t}$ is a mute variable, τ is the time distance between the Solar System Barycenter (SSB), Ω_i and p^i indicate the direction versors of the GW source and the MSP, respectively. The two terms in the square brackets are known as Earth[d] term and Pulsar term, and

[b]For a quick graphic overview of the different GW sources and GW detectors, see Ref. 9.
[c]For a complete theoretical description, see Ref. 16.
[d]The name "Earth" is a legacy of early work on PTAs.[17] In fact, since the Earth is not an inertial reference frame, the observer is assumed in SSB, whose position is known with great precision.

describe the metric in the proximity of the two objects[e]. Eq. (1) implies that, for an arbitrary MSP set, the Earth term is the same, while the Pulsar term changes. Since the latter term gives rise to an uncorrelated contribution in timing residuals, it is neglected in the standard GWB search[f]. This assumption is not entirely true if the MSP set is confined in a very small region, as in the case of globular clusters (GCs). Therefore, the GW-induced timing residuals of MSPs within GCs, especially those in their cores, should be strongly correlated, so considering these MSPs may offer a new instrument for GW detection.

3. Simulated Environment

The main target of the GW search with PTA is the GWBs produced by a hypothetical SMBHB population. Such a detection would give crucial information on the history of the Universe and, since that has not yet been achieved, it is worth considering improving PTAs by including MSPs within GCs.

We simulated the GW-induced timing residuals on GC core MSPs due to an SMBHB population. We arbitrarily assumed that the SMBHB chirp mass, distance, and GW emission frequency follow a log-normal distribution[g] and are in the ranges 10^8–10^9 $M_\odot$, 10^8–10^9 pc, and 10^{-9}–10^{-8} Hz, respectively, and that the SMBHB right ascension and declination follow a uniform distribution (see Fig. 1).

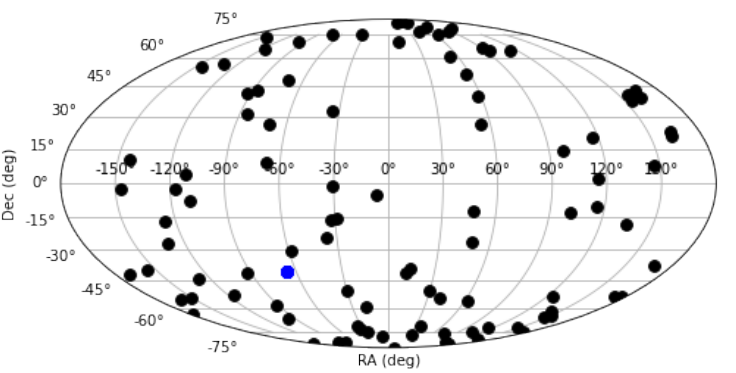


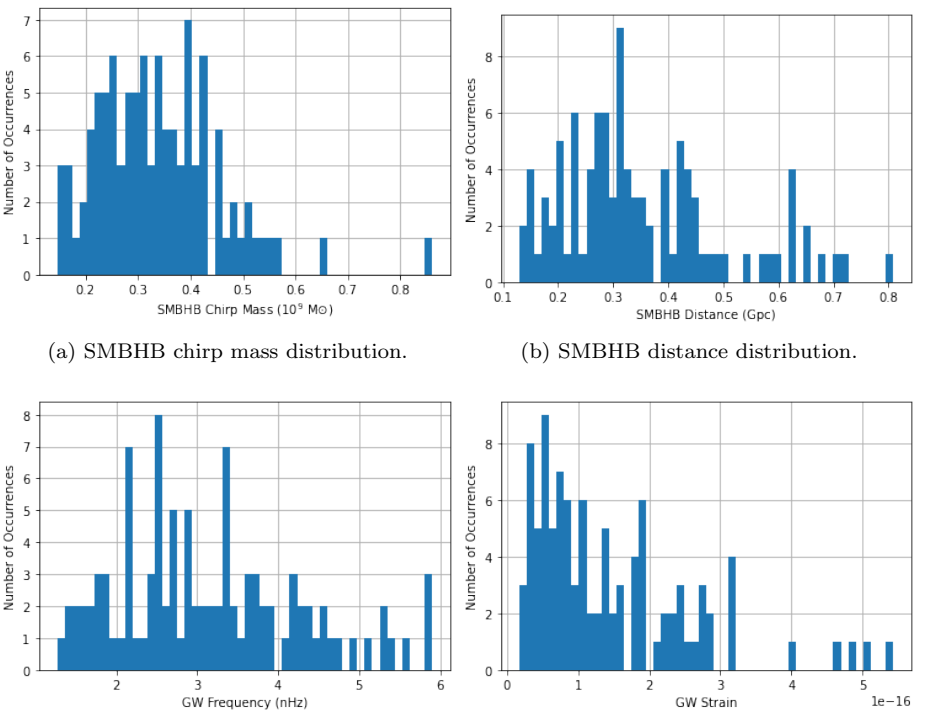
Fig. 1. Simulation of the SMBHB population. The position of 100 simulated SMBHBs, randomly scattered all over the sky is shown. The black dots indicate the SMBHB coordinates (RA and Dec), while the blue dot indicates the simulated GC coordinates.

[e]In this paper geometrical units c=G=1 have been adopted.
[f]However, it is important to remark that the pulsar term may play an important role in the detection of GW emitted by single sources.
[g]The properties of the SMBHB population are still a matter of debate. However, this paper aims to explain how PTAs can be improved by including MSPs within GCs, so we took the results in Refs. 18–20 just as a guide for building SMBHB distributions compatible with the GW-induced timing residuals observable by PTAs.

The SMBHB strain is, in good approximation, given by:

$$h = \sqrt{\frac{32}{5}} \frac{\mathcal{M}^{5/3}}{D} \left(\frac{2\pi f}{1+z} \right)^{2/3} \tag{2}$$

where $\mathcal{M}$, D, z, and f indicate the SMBHB chirp mass, distance, cosmological redshift, and GW emission frequency, respectively. Eq. (2) allows to determine the GW strain distribution for our SMBHB population (see Fig. 2).



(a) SMBHB chirp mass distribution.

(b) SMBHB distance distribution.

(c) GW frequency distribution.

(d) GW strain distribution.

Fig. 2. SMBHB parameter simulation. Panel (a) shows the SMBHB chirp mass log-normal distribution, expressed in units of $10^9\ M_\odot$. Panel (b) shows the SMBHB distance log-normal distribution, expressed in Gpc. Panel (c) shows the GW frequency log-normal distribution, expressed in nHZ. Panel (d) shows the GW strain distribution, expressed in units of 10^{-16}. On the vertical axis of each panel is plotted the number of occurrences.

We also assumed that the GC core MSP distance[h], right ascension, and declination with respect to the GC center follow a uniform distribution[i]. To get a realistic

[h]In this paper, MSP distance refers to the distance with respect to the diametrical plane passing through the GC center and orthogonal to the line-of-sight.

[i]In accordance with what is described by the GC core mass-density distribution.[21]

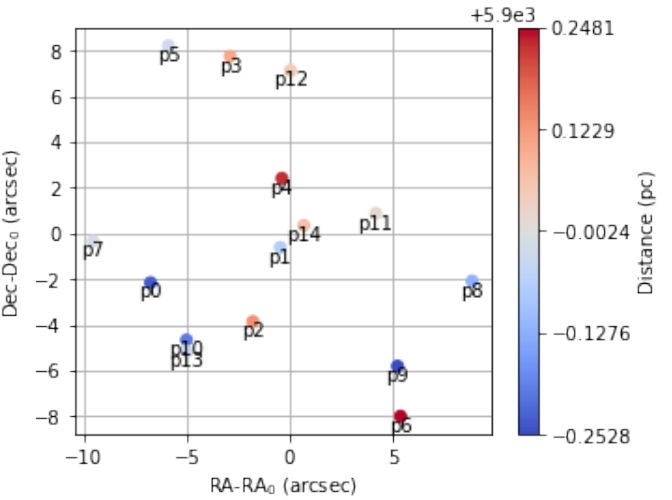


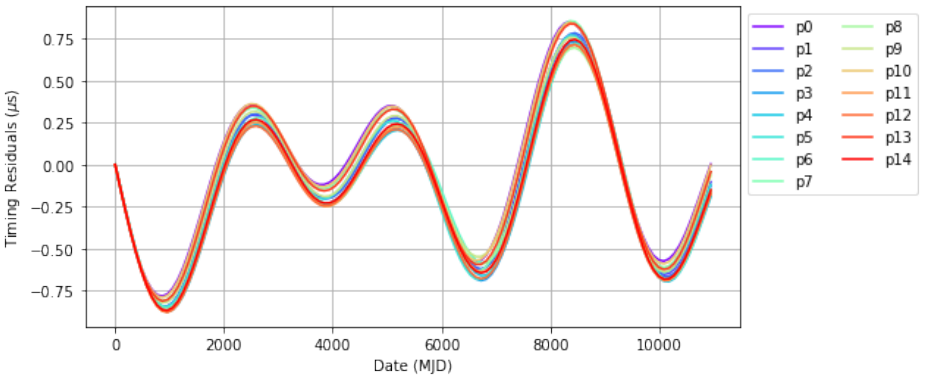
Fig. 3. Simulation of the MSP set. The position of 15 simulated GC core MSPs, randomly scattered within a GC core, whose distance and angular radius are 5.9 kpc and 0.16 arcmin, respectively, is shown. The dots indicate the MSP coordinates (RA and Dec), with respect to the GC center (RA_0 and Dec_0), expressed in arcsec, while the dot colors indicate, as described by the color bar, the MSP distance with respect to the GC core center, expressed in pc (the MSP distance with respect to the SSB can be obtained by summing the values expressed by the labels of the color bar with the quantity on its top).

result, we took the GC core distance, angular radius, and MSP number of the Terzan 5 GC since it is one of the most populated by MSPs[22–24] (see Fig. 3).

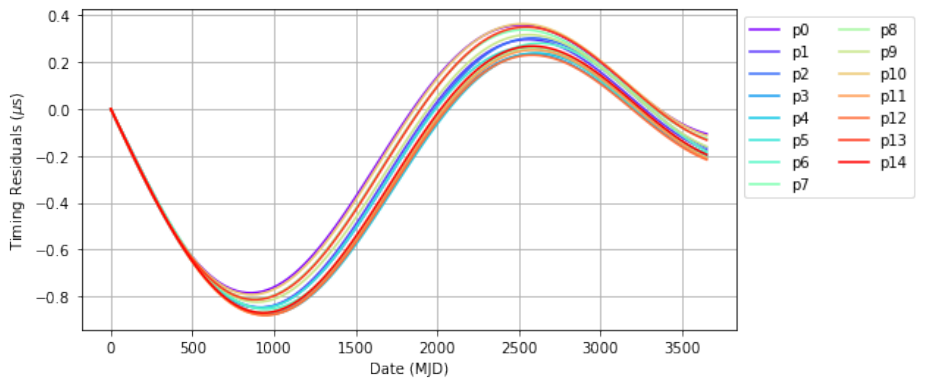
We finally calculated the simulated GW-induced timing residuals corresponding to each GC core MSP due to the overall contribution of the GW emission from each SMBHB (see Fig. 4).

4. Results

The simulation results (see Fig. 4) confirm what has been deduced in Sec. 2. In the PTA frequency band, the simulated GW-induced timing residual corresponding to each GC core MSP due to the overall contribution of the GW emission from each SMBHB appear to be strongly correlated. This property can be very useful in the GW search. Observing such correlation for the GC core MSPs would lead to a strong evidence for the GWB, especially if supported by stronger hints of a common red process compatible with a GWB. GC core MSPs can also be helpful for determining the quadrupolar nature of the cross-correlations between the timing residuals of each MSP pair. Indeed, the Hellings and Downs function has a global maximum when the angular distance between an MSP pair is null. Of course, it is unlikely to have an aligned MSP pair by chance, but this frequently happens in GCs due to physical reasons. Therefore, GCs offer the unique possibility of obtaining cross-correlation



(a) Simulated 30 years GW-induced timing residuals.



(b) Simulated 10 years GW-induced timing residuals.

Fig. 4. GW-induced timing residuals. Panel (a) shows the simulated GW-induced timing residuals corresponding to each GC core MSP due to the overall contribution of the GW emission from each SMBHB, plotted over a time interval of 30 years. Panel (b) shows the GW-induced timing residuals, plotted over a time interval of only 10 years. In both panels, the colored curves indicate the simulated GW-induced timing residuals, and have a different color for each GC core MSP, as described by the legend. On the horizontal axis of both panels the date measured from the arbitrary beginning of the simulated ToA observation, expressed in MJD, is plotted. On the vertical axis of both panels the timing residuals, expressed in μs, are plotted.

data for low angular separation MSPs, which can be very important because they might challenge or confirm the prediction of the Hellings & Downs function.

5. Conclusion

The proximity between the MSPs lying in a given GC, especially those in the GC core, can be exploited to get unique information about GWs and, for this reason, it is worth considering improving PTAs by including them.

It is important to note that, in the presented analysis, it has been considered an ideal situation to highlight some of the advantages of including MSPs within

GCs in PTAs. On the other hand, this class of MSPs is currently not timed with a precision adequate to detect GW-induced timing residual.[25,26]

Nevertheless, we expect that to change soon, thanks to new detectors, such as the Five hundred meter Aperture Spherical Telescope (FAST),[27] the MeerKAT radio telescope,[28] and the Square Kilometre Array (SKA).[29] SKA, in particular, will play a crucial role in ultra-low frequency GW detection. With its unprecedented sensitivity, it will be possible to time significantly better the GC MSPs, opening the possibility of adopting the strategy proposed in this paper, which is a summary of Ref. 30.

Acknowledgments

We warmly acknowledge Andrea Possenti, of the Istituto Nazionale di Astrofisica (INAF), for many useful discussions. We also acknowledge the support by the Theoretical Astroparticle Physics (TAsP) and Euclid projects of the Istituto Nazionale di Fisica Nucleare (INFN).

References

1. A. Einstein, Die grundlage der allgemeinen relativitätstheorie, *Ann. Phys.* **354**, 769 (1916).
2. B. P. Abbott *et al.*, Observation of gravitational waves from a binary black hole merger, *PRL* **116**, 061102 (2016).
3. B. P. Abbott *et al.*, Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB170817A, *ApJ* **848**, L13 (2017).
4. R. Abbott *et al.*, Observation of gravitational waves from two neutron star-black hole coalescences, *ApJ* **915**, L5 (2018).
5. P. Amaro-Seoane *et al.*, Laser interferometer space antenna, *eprint arXiv:1702.00786* (2017).
6. Lisa `https://https://sci.esa.int/web/lisa`.
7. M. Rajagopal and R. W. Romani, Ultra-low-frequency gravitational radiation from massive black hole binaries, *ApJ* **446**, 543 (1995).
8. T. Damour and A. Vilenkin, Gravitational wave bursts from cusps and kinks on cosmic strings, *PRD* **64**, 064008 (2001).
9. Gravitational-wave sensitivity curves `https://cplberry.com/2015/01/10/1408-0740/`.
10. M. V. Sazhin, Opportunities for detecting ultralong gravitational waves, *Astronomicheskii Zhurnal* **55**, 65 (1978).
11. S. Detweiler, Pulsar timing measurements and the search for gravitational waves, *ApJ* **234**, 1100 (1979).
12. R. W. Hellings and G. S. Downs, Upper limits on the isotropic gravitational radiation background from pulsar timing analysis, *ApJ* **265**, L39 (1983).
13. Z. Arzoumanian *et al.*, The NANOgrav 12.5 yr data set: Search for an isotropic stochastic gravitational-wave background, *ApJ* **905**, L34 (2020).
14. S. Blasi, V. Brdar and K. Schmitz, Has NANOGrav found first evidence for cosmic strings?, *PhRvL* **126**, 041305 (2021).

15. B. Goncharov *et al.*, On the evidence for a common-spectrum process in the search for the nanohertz gravitational-wave background with the parkes pulsar timing array, *ApJ* **917**, L19 (2021).
16. M. Maggiore, *Gravitational Waves. Vol. 2: Astrophysics and Cosmology* (Oxford University Press, 2008).
17. A. N. Lommen, Pulsar timing arrays: the promise of gravitational wave detection, *RPPh* **78**, 124901 (2015).
18. M. Tucci and M. Volonteri, Constraining supermassive black hole evolution through the continuity equation, *A&A* **600**, A64 (2017).
19. M. Celoria, R. Oliveri, A. Sesana and M. Mapelli, Lecture notes on black hole binary astrophysics, *eprint arXiv:1807.11489* (2018).
20. N. Sanchis-Gual, V. Quilis and J. A. Font, Estimate of the gravitational-wave background from the observed cosmological distribution of quasars, *PRD* **104**, 024027 (2021).
21. J. Binney and S. Tremaine, *Galactic dynamics* (Princeton University Press, 1987).
22. M. Cadelano *et al.*, Discovery of three new millisecond pulsars in terzan 5, *ApJ* **855**, 125 (2018).
23. The ATNF pulsar database `https://www.atnf.csiro.au/research/pulsar/psrcat/`.
24. Pulsars in globular clusters `https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html`.
25. N. D'Amico *et al.*, Timing and searching millisecond pulsars in globular clusters ATNF Proposal, (2009).
26. P. C. C. Freire, Long-term observations of the pulsars in 47 Tucanae - II. Proper motions, accelerations and jerks, *MNRAS* **471**, 857 (2017).
27. R. Nan, The five-hundred aperture spherical radio telescope (FAST) project, *IJMPD* **20**, 989 (2011).
28. M. Bailes *et al.*, The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime, *PASA* **37**, e028 (2020).
29. A. Weltman *et al.*, Fundamental physics with the square kilometre array, *PASA* **37**, e002 (2020).
30. M. Maiorano, F. De Paolis and A. A. Nucita, Including millisecond pulsars inside the core of globular clusters in pulsar timing arrays, *in press on EPJP (https://doi.org/10.1140/epjp/s13360-021-02098-0), eprint arXiv:2110.11046* (2021).

Searching for pulsars in globular clusters with the MeerKAT Radio Telescope

F. Abbate on behalf of the MeerTIME/TRAPUM Collaboration
Max Planck Institut für Radioastronomie,
Auf dem Hügel 69, D-53121, Bonn, Germany
E-mail: abbate@mpifr-bonn.mpg.de

Globular clusters are known to host an unusually large population of millisecond pulsar when compared to the Galactic disk. This is thanks to the high rate of dynamical encounters occurring in the clusters that can create the conditions to efficiently recycle neutron stars into millisecond pulsars. The result is a rich population of pulsars with properties and companions difficult or impossible to replicate in the Galactic disk. For these reasons, globular clusters have been and still are a prime target of searches for new and exciting pulsars. Because of their large distances, the limiting factor inhibiting these discoveries is the telescope sensitivity. The MeerKAT radio telescope, a 64-dish interferometer in South Africa, guarantees unrivalled sensitivity for globular clusters in the southern sky. Observations of well-studied globular clusters with MeerKAT have already returned more than 35 new pulsars with many more expected. These exciting discoveries will help us to understand more about the neutron star equation of state, stellar evolution, accretion physics and to hunt for intermediate mass black holes. In this talk I will present the prospects and current discoveries of the globular cluster working group in the MeerTIME and TRAPUM programmes.

Keywords: Pulsars; globular dusters.

1. Pulsars in globular clusters

Globular clusters contain a very rich population of pulsars with 232 known in 36 cluster[a] and this population is quite peculiar when compared with the one in the Galactic disk. The abundance of pulsars per unit of stellar mass in globular clusters is ~ 100 times higher than that of the disk and the vast majority of the pulsars are millisecond pulsars (MSP) with rotation periods shorter than a few tens of ms.

The difference between the population of pulsars in globular clusters and in the Galactic disk is caused by the starkly different environments. Globular clusters have no sign of recent star formation and have very high stellar densities in their cores that can reach up to $\sim 10^{4-5}$ $M_\odot\,pc^{-3}$ compared to $\sim 1-10$ $M_\odot\,pc^{-3}$ in the Solar neighbourhood. This high stellar density favours dynamical encounters between stars that can lead to exchange encounters where neutron stars become part of new binary systems. The exchanges of companions increase the chance of the neutron star going through an accretion phase boosting the number of low mass

[a]An up-to-date count can be found in the webpage `https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html`

X-ray binaries (LMXB) that evole into MSPs. Once a neutron star has been recycled into an MSP, it is likely to remain active for a time even longer than the Hubble time due to the low magnetic field.

Due to the peculiar environment within globular clusters, the possibility of going through multiple stages of recycling and the lack of recent star formation, the majority of pulsars have very fast rotational periods usually smaller than 20 ms. The fastest pulsar known, PSR J1748−2446ad,[1] with a rotational period of 1.395 ms resides in the globular cluster Terzan 5. Not all of the pulsars in globular clusters are, however, very fast. There is a small population of pulsars where the accretion process was interrupted before completion,[2] possibly by external encounters, with period between 20-100 ms called 'mildly recycled'. Furthermore, there is, in some specific clusters, a small number of slow pulsars with periods over 100 ms with the record of 1004.04 ms being held by B1718−19A in NGC 6342.[3] The presence of these apparently 'young' pulsars is in contradiction with the known formation scenarios and the lack of new star formation. Apart from dynamic disruption of LMXB[2] other possible explanations have been put forward like accretion-induced collapse, direct collisions with a main sequence star,[4] or electron capture supernova of an OMgNe white dwarf.[5]

The MSPs populations of globular clusters is also quite different from the one in the Galactic disk. The percentage of isolated to binary MSPs can vary a lot between clusters with some like NGC 6624 and NGC 6517 having mostly isolated pulsars and others like M62 and M5 having mostly binary pulsars. This is linked to the dynamical parameters of the clusters.[2]

The multiple dynamical interactions create a population of binary MSPs that are in some cases very different from the Galactic disk population. Eclipsing systems with non-degenerate companions like 'black widows' and 'redbacks' are common.[3,6,7] Binary systems can acquire high eccentricities and massive companions through successive interactions or through exchange encounters like M15C,[8] NGC 1851A,[9] NGC 6544B,[10] and NGC6652A[11] . A very peculiar triple system with a white dwarf and a Jupiter-mass companion has also been observed in a globular cluster.[12,13]

1.1. *Science with pulsars in globular clusters*

The scientific achievements made possible by this unique population of pulsars together with the possibility of observing a large number of pulsars in the same telescope beam have made globular clusters a high priority target in the past decades.

The internal composition of neutron stars is determined by the equation-of-state that can be probed by measuring their masses. This can be achieved through pulsar timing by measuring the so-called post-Keplerian parameters, which quantify relativistic orbital effects.[14] One of these parameters is the relativistic precession of the periastron that is relatively easily accessible in highly eccentric systems, and can be used to place constraints on the total mass of the systeem assuming

general relativity. Thanks to this method, some neutron stars with the potential of breaking the current mass limit have been identified.[15,16] The importance of determining neutron star masses is not limited to the value of the maximum mass, also the minimum mass and the distribution of masses can be used to test various formation and recycling scenarios.[17]

Another avenue of research that can be pursued in globular clusters regards the maximum rotation speed. Currently the record of 1.395 ms is held by J1748-2446ad in Terzan 5[1] but even faster pulsars might be present in clusters and set more constraints on the equation of state.

Studies of eclipsing pulsars like 'black widows' and 'redbacks' are essential to study the possible evolution scenarios of MSPs and to constrain the properties of the companions.[18,19]

If we change focus from specific pulsars to the entire population of pulsars in a single cluster, the science perspective broadens to include the properties of the globular clusters themselves. The gravitational effects of the globular clusters on the pulsars are accessible through the derivatives of the rotational period.[20] With a significant number of pulsars, this can give us access to the structural parameters of the clusters and explore the possibility of the presence of an intermediate mass black hole in the center.[21–23]

Using the dispersion measures (DM) and estimating the 3-d position of the pulsars inside the cluster, it becomes possible to look for diffuse gas. This method has led to the first detection of ionized gas in a globular cluster.[23,24] If we add the information coming from the rotation measure, it becomes possible to probe also the magnetic fields. This led to the possible discovery of a magnetized outflow from the Galaxy in the direction of the globular cluster 47 Tucanae.[25]

The numerous scientific possibilities suggest that globular clusters pulsars should still be considered extremely profitable targets for fututre research. Simulations show that a significant population of potentially interesting pulsars is still undiscovered in the clusters.[26,27] These could include MSP-black hole systems, which would be invaluable for tests of gravitational theories.[28] However, despite the efforts, the number of new discoveries in the last ten years has been very low. The ones that were discovered were found in archival data using new search techniques.[29–31] This suggests that we had reached the sensitivity limit of the current observing facilities. To uncover the numerous and potentially interesting pulsars new and more sensitive facilities are needed. One such facility is FAST[32] in China that became operative in 2016 and has already discovered 32 pulsars in globular clusters.[33–35] Another new facility that recently came online and has the potential to revolutionize the field is MeerKAT[36] in South Africa.

2. Advantages of the MeerKAT radio telescope

The MeerKAT radio telescope is an interferometer composed of 64 antennas each with a diameter of 13.5 m. The maximum baseline within antennas is 8 km.

Thanks to its location in South Africa at a latitude of $-30°$, it can observe the entire Southern sky. When observing with all antennas it can reach telescope gains four times higher than Parkes, the only other radio telescope capable of observing pulsars in cm-wavelengths below a declination of $-45°$. The gain is also 1.4 times higher than the Green Bank Telescope giving MeerKAT a significant advantage when looking at the central regions of the Galaxy that are observable at high elevations for long hours.

MeerKAT is currently able to observe in UHF-band (544-1088 MHz), L-band (856-1712 MHz) showing remarkably RFI-clear bands. In near future also the S-band (1750-3500 MHz) receivers will be available for observations. A description of the capabilities and potentialities of MeerKAT as a pulsar observatory can be found in Bailes et al. 2020.[37]

Of the MeerKAT Large Science Projects (LSPs) that are currently using MeerKAT for observations, two include pulsars as the main target: MeerTIME[38] and TRAPUM.[39]

2.1. *MeerTIME*

MeerTIME[b] is an LSP based on timing already known pulsars with the high sensitivity of MeerKAT that started collecting data in early 2019. It has four main scientific goals: timing relativistic binary pulsars in order to measure the masses of the systems and test theories of gravity[40]; the Thousand Pulsars Array that monitors a wide variety of pulsars to study their geometries, magnetospheres and interstellar effects[41]; a pulsar timing array that monitors high precision pulsars to contribute to the search for nanohertz gravitational waves [Spiewak et al. submitted]; time pulsars in globular clusters.[42]

The project plans to observe globular cluster pulsars in order to study the eclipses, measure the masses of the pulsars, and study the properties of the globular clusters themselves.

MeerTIME can observe up to four tied array beams using the Pulsar Timing User Supplied Equipment (PTUSE) machines as the main data acquisition system. These beams can either be in "timing" mode where a pulsar is folded in real time using an ephemeris or in "search" mode where the data is saved in a "filterbank" style file and can be folded later or searched for new pulsars or single pulses from previously known ones. For both modes the observations are done with very high time resolution (9 μs), full polarimetric information, and real-time coherent dedispersion.

Observing globular clusters has the advantage that a large number of pulsars fall within the same beam of the telescope and we are therefore capable of observing them simultaneously if the data is recorded in "search" mode. However, the beam of MeerKAT using the full array with a baseline of 8 km would be too small to

[b]http://www.meertime.org

observe all of the pulsars in some globular clusters so in most cases only the central 44 antennas with a maximum baseline of 1 km are used.

2.2. *TRAPUM*

The TRAnsients and PUlsars with MeerKAT (TRAPUM[c]) LSP focuses on searching for new pulsars. It searches for pulsars in unidentified Fermi sources, supernova remnants, pulsar wind nebulae, in the Galactic plane, in nearby galaxies and in globular clusters. The globular cluster working group is focused on searching both in clusters with known pulsars (which allows de-dispersion at the known DM of the cluster) and in clusters with no known pulsars. The project started taking data in April 2020.

TRAPUM uses Filterbanking Beam-former User Supplied Equipment (FBFUSE) and the Accelerated Pulsar Search User Supplied Equipment (APSUSE), a 60 node computing cluster, to synthesize up to ~ 400 beams and record incoherently de-dispersed search-mode data for each one. This allows us to cover a large portion of the sky (up to $\sim 1-4$ arcmin in radius depending on the configuration) while keeping the full sensitivity of MeerKAT. In this way it is possible to cover the entire cluster with one pointing and search for pulsars in the center and in the outskirts of the cluster simultaneously. However, due to the very high data rate, it is not possible to retain full temporal and spectral resolution nor full polarimetric information.

Because of the complementarity of science goals, the globular cluster working groups of MeerTIME and TRAPUM have decided to collaborate sharing observing time and resources. This collaboration has already resulted in two publications.[42,43]

3. New discoveries

The joint collaboration has discovered 36 new pulsars in globular clusters[d]. The discoveries are summarized in Table 1 and 2 and can be split chronologically into two groups: the ones made with the PTUSE machine before TRAPUM became operative and the ones made possible by the APSUSE machine of TRAPUM.

3.1. *MeerTIME discoveries*

The first batch of discoveries made with observations taken by MeerTIME is already published.[42] This publication contains the description and the properties of 8 new millisecond pulsars found in six different globular clusters. Three of them are isolated pulsars while the rest are found in binary systems. The profiles of these pulsars are shown in Figure 1.

[c]http://www.trapum.org

[d]the full up-to-date list of discoveries can be found at http://www.trapum.org/discoveries.html

Table 1. Properties of the pulsars discovered by MeerTIME/TRAPUM using the PTUSE machine

Pulsar name	Period (ms)	DM (pc cm^{-3})	notes
Ter 5 an	4.8023	237.7	Binary
M62G	4.6081	113.6	Binary
NGC 6522D	5.5369	192.7	Isolated
NGC 6624G	6.0912	86.20	Binary
47 Tuc ac	2.7456	24.46	Ecl. Binary
47 Tuc ad	3.7460	24.41	Ecl. Binary
NGC 6752F	8.4854	33.20	Isolated
NGC 6624H	5.1301	86.85	Isolated
NGC 6440G	5.2157	219.7	Isolated
NGC 6440H	2.8486	222.6	Binary
M28M	9.5689	119.23	Binary

Table 2. Properties of the pulsars discovered by MeerTIME/TRAPUM using the APSUSE machine

Pulsar name	Period (ms)	DM (pc cm^{-3})	notes
NGC 6752G	4.7902	33.27	Isolated
M28N	3.3429	119.33	Binary
NGC 6342B	2.5683	71.44	Isolated
NGC 1851B	2.8162	52.07	Isolated
NGC 1851C	5.5648	52.05	Isolated
NGC 1851D	4.5543	52.17	Binary
NGC 1851E	5.5952	51.95	Binary
NGC 1851F	4.3294	51.63	Binary
NGC 1851G	3.8028	51.01	Binary
NGC 1851H	5.5061	52.26	Binary
NGC 6752H	2.0155	33.25	Isolated
NGC 6752I	2.6582	33.34	Isolated
NGC 6624I	4.3195	87.35	Isolated
NGC 6624J	20.8995	86.85	Isolated
NGC 6441E	251.1587	221.16	Isolated
NGC 6441F	6.0006	228.22	Binary
NGC 6441G	5.3522	229.20	Isolated
NGC 1851I	32.6538	52.42	Binary
NGC 1851J	6.6329	52.06	Isolated
NGC 1851K	4.6920	51.93	Isolated
NGC 1851L	2.9586	51.23	Binary
NGC 1851M	4.7977	51.66	Isolated
NGC 1851N	5.5679	51.11	Isolated
J1823-3022	2497.72	96.9	Isolated
NGC 6624K	2.7686	87.46	Isolated

These discoveries include two new eclipsing millisecond pulsars 47 Tuc ac and 47 Tuc ad. Fitting for the orbits of these pulsars suggest that 47 Tuc ac is a ‘black widow’ pulsar while 47 Tuc ad is a ‘red-back’. Both of these discoveries were then

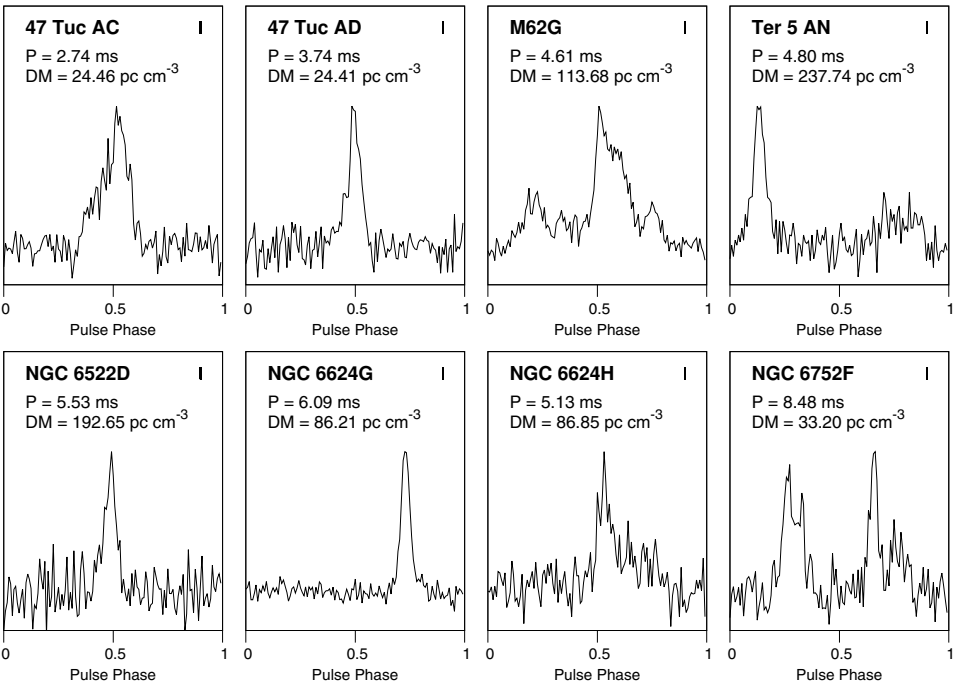


Fig. 1. Profiles of the eight new pulsar discoveries published in Ridolfi et al. 2021.[42]

re-detected in previous Parkes observations but due to scintillation effects and eclipses the detections were too few to allow for deriving phase-connected timing solutions.

Possibly the most interesting pulsar in this set of discoveries is NGC 6624G. This pulsar is found in a binary with orbital period of 1.54 days and eccentricity of 0.38. Because of the very high eccentricity it was possible to use only 11 months of data to measure the rate of advance of periastron $\dot{\omega} = 0.217 \pm 0.004$ deg yr^{-1}. Assuming that this advance is caused entirely by relativistic effects and using formulas derived in General Relativity (GR) it is possible to derive the total mass of the system, $M_{tot} = 2.65 \pm 0.07 \mathrm{M}_{\odot}$. This parameter alone is not enough to measure precisely the mass of the pulsar itself, see Figure 2, but suggests that the neutron star is particularly heavy. Measurements of further post-keplerian parameters will be needed to accurately determine the mass. The companion mass is also particularly heavy suggesting that it was not the original companion that recycled the pulsar but a new companion gained in an exchange encounter. This is the type of secondary exchange products that we have been looking for.

Another newly discovered pulsar that has the potential of being particularly massive is Ter 5 an. This binary has an eccentricity of 0.0066, two orders of magnitude higher than the average millisecond pulsar binaries found in the Galactic field. It is likely caused by distant stellar encounters in the dense cluster environment. Also

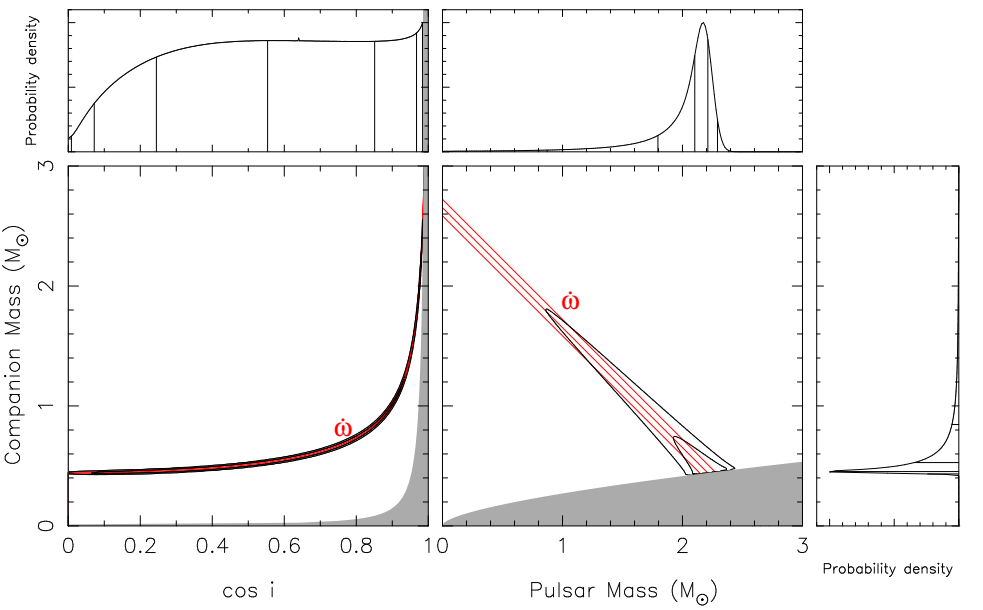


Fig. 2. Mass-inclination and mass-mass diagrams for NGC 6624G. The main square panels depict the $\cos i$-M_c and M_c-M_p planes. The grey areas in the former are excluded by the requirement that the mass of the pulsar is positive, the grey areas in the latter are excluded by the mass function and the requirement that $\sin i \leq 1$. The red lines depict the masses consistent with the measurement of $\dot{\omega}$ and its ± 1-σ uncertainties, under the assumption that this effect is dominated by the GR contribution and that GR is the correct theory of gravity. The contours include 68.3 and 95.4% of a 2-D probability distribution function (pdf) derived from the χ^2 of `TEMPO` fits that assumed all GR effects to be according to the masses and orbital inclination at each point. The side panels show the probability density functions for the $\cos i$ (top left), M_p (top right) and M_c (right) derived by marginalizing the aforementioned 2-D pdfs. we obtain a median for the pulsar mass of 2.1 $M_\odot$, but with a tail of probability that extends to lower masses: there is a 31% probability of $M_p < 2$ $M_\odot$ and a 3.8% probability of $M_p < 1.4$ $M_\odot$. Thus, either the pulsar is very massive, or it has a massive companion; the system was likely formed in an exchange encounter. Credits: Ridolfi et al. 2021.[42]

here, owing to the recovery of the system in data from the Green Banks Telescope and the resulting long baseline, it was possible to measure an advance of the periastron with a rate of $\dot{\omega} = 0.009 \pm 0.001$ deg yr^{-1}. With the assumption that it fully caused by GR effects, the total mass becomes $M_{tot} = 2.95 \pm 0.52 M_\odot$. The mass-inclination and mass-mass diagrams are shown in Figure 3. The uncertainties in this case are higher allowing for a greater range in pulsar mass.

The other pulsars discovered by MeerTIME are in the process of being published. Of these, one is particularly interesting: NGC 6440H. It is a binary pulsar with a very light companion. Assuming a pulsar mass of 1.4 $M_\odot$, the companion has a minimum mass of 0.0062 $M_\odot$. This very low mass of the companion suggests that it could be brown dwarf or a planet. If the planetary nature of the companion is confirmed, this would be only the second pulsar with such a companion.[12,13]

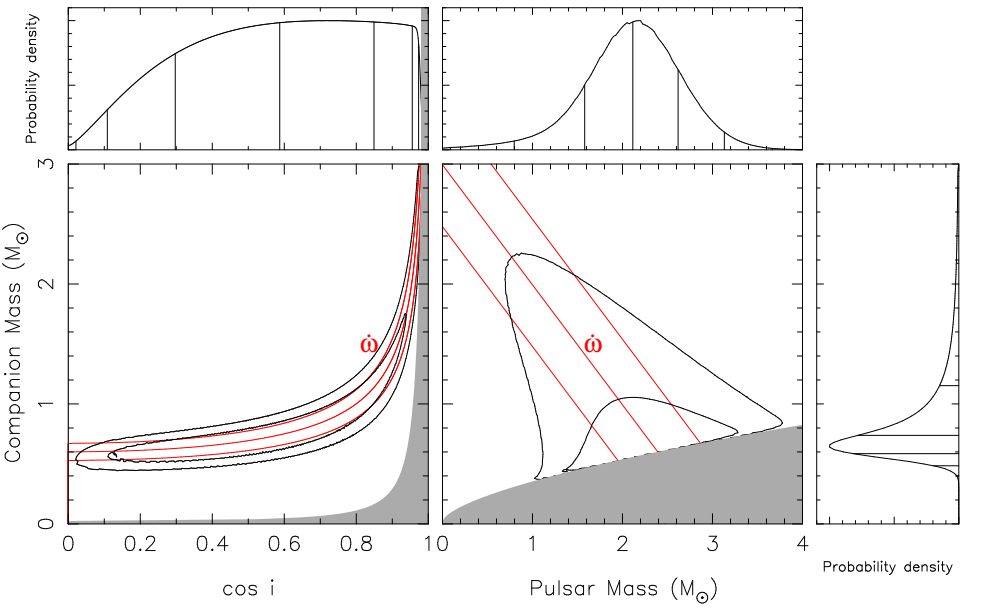


Fig. 3. Same mass-inclination and mass-mass diagrams as in Figure 2 but for Ter 5 an. The estimated median pulsar mass is 2.13 $M_\odot$ and median companion mass is 0.75 $M_\odot$. Credits: Ridolfi et al. 2021.[42]

3.2. *TRAPUM discoveries*

Only a small fraction of the observations of globular clusters with the APSUSE machine with TRAPUM have currently been processed. The large volumes of data produced in these observations require a large processing power to be properly analyzed. Despite this, the new pulsar discoveries are already more than double than those made using the PTUSE machines and only the central 44 antennas. The list is shown in Table 2.

More than half of these discoveries come from just one globular cluster NGC 1851. This is a core-collapsed cluster that was known to contain only one pulsar NGC 1851A. TRAPUM observations in just the central beam have confirmed the presence of 13 new pulsars. This large population makes NGC 1851 the third globular cluster for number of known pulsars together with M28. Once phase connected ephemerides are determined for all these pulsars it will be possible to probe the potential well and test the presence of gas and magnetic fields within the cluster.

Two of these new discoveries, NGC 1851D and NGC 1851E, are binary systems with very high eccentricities. NGC 1851E has an orbital period of around 7 days and an eccentricity of $e = 0.71$. This high value means that in the near future it will be possible to measure the rate of precession of the periastron to high accuracy and therefore the total mass of the system. Early fits to the orbit suggest that, assuming a pulsar mass of 1.4 $M_\odot$, the companion has a minimum mass of 1.5 $M_\odot$.

This suggests that there is a high probability of this being another double neutron star system.

The vast majority of the discoveries are fast spinning MSPs but there are some exceptions: NGC 6624J and NGC 1851I have periods respectively of 20.89 and 32.65 ms, that suggest that they are mildly recycled pulsars. Another exceptions is NGC 6441E, a slow pulsar with a period of 251.15 ms. Particularly interesting is also J1823-3022 with a period of 2497.72 ms. This potentially record breaking pulsar has been found in the outskirts of the globular cluster NGC 6624 and has a DM that is higher than the rest of the pulsars in the cluster. Because of these reasons it is not clear if it is really a member of the cluster or if it is a Galactic disk pulsar positioned along the same line of sight. All of these slow pulsars are in core-collapsed clusters validating the correlation between spin periods and dynamical state of the cluster suggested previously.[2]

These early results from a fraction of the observations suggest that many new discoveries are possible by analyzing the data that has already been captured.

4. Future plans

Early observations have showed that the future of the two projects MeerTIME and TRAPUM can prove revolutionary for the science of globular cluster pulsars.

MeerTIME will focus on timing the already discovered pulsars in order to measure the masses of the most interesting binaries and characterise the eclipses of the ecplising 'black-widows' and 'redbacks'. Furthermore, by improving the timing models of all pulsars it will be possible the study the properties of the clusters like the gas content, magnetic field and the presence of intermediate mass black holes.

TRAPUM, on the other hand, will continue to search for new pulsars broadening the scope in order to include clusters with no previously known pulsars. The searches will involve all of the available bands, going from the UHF band, where the pulsars are brighter but the interstellar medium is more noisy to the L band and the S band where the effects of scattering and the interstellar medium are minimized.

In closing, the last few years have shown an almost unprecedented growth in the number of pulsars in globular clusters going from 150 in 2018 to over 230 in 2021. The largest contributor is the MeerKAT telescope with 36 but a very close second is FAST with 32 new discoveries. These numbers are likely to increase in the following years and with them also the possibility of finding even more exotic systems like a double millisecond pulsar binary or a pulsar-black hole binary.

References

1. J. W. T. Hessels, S. M. Ransom, I. H. Stairs, P. C. C. Freire, V. M. Kaspi and F. Camilo, A Radio Pulsar Spinning at 716 Hz, *Science* **311**, 1901 (March 2006).
2. F. Verbunt and P. C. C. Freire, On the disruption of pulsar and X-ray binaries in globular clusters, *Astronomy and Astrophysics* **561**, p. A11 (January 2014).

3. A. G. Lyne, J. D. Biggs, P. A. Harrison and M. Bailes, A long-period globular-cluster pulsar in an eclipsing binary system, *Nature* **361**, 47 (January 1993).
4. A. G. Lyne, R. N. Manchester and N. D'Amico, PSR B1745-20 and Young Pulsars in Globular Clusters, *Astrophysical Journal Letters* **460**, p. L41 (March 1996).
5. J. Boyles, D. R. Lorimer, P. J. Turk, R. Mnatsakanov, R. S. Lynch, S. M. Ransom, P. C. Freire and K. Belczynski, Young Radio Pulsars in Galactic Globular Clusters, *The Astrophysical Journal* **742**, p. 51 (November 2011).
6. F. Camilo, D. R. Lorimer, P. Freire, A. G. Lyne and R. N. Manchester, Observations of 20 Millisecond Pulsars in 47 Tucanae at 20 Centimeters, *The Astrophysical Journal* **535**, 975 (June 2000).
7. S. M. Ransom, J. W. T. Hessels, I. H. Stairs, P. C. C. Freire, F. Camilo, V. M. Kaspi and D. L. Kaplan, Twenty-One Millisecond Pulsars in Terzan 5 Using the Green Bank Telescope, *Science* **307**, 892 (February 2005).
8. S. B. Anderson, P. W. Gorham, S. R. Kulkarni, T. A. Prince and A. Wolszczan, Discovery of two radio pulsars in the globular cluster M15, *Nature* **346**, 42 (July 1990).
9. P. C. Freire, Y. Gupta, S. M. Ransom and C. H. Ishwara-Chandra, Giant Metrewave Radio Telescope Discovery of a Millisecond Pulsar in a Very Eccentric Binary System, *Astrophysical Journal Letters* **606**, L53 (May 2004).
10. R. S. Lynch, P. C. C. Freire, S. M. Ransom and B. A. Jacoby, The Timing of Nine Globular Cluster Pulsars, *The Astrophysical Journal* **745**, p. 109 (February 2012).
11. M. E. DeCesar, S. M. Ransom, D. L. Kaplan, P. S. Ray and A. M. Geller, A Highly Eccentric 3.9 Millisecond Binary Pulsar in the Globular Cluster NGC 6652, *The Astrophysical Journal Letters* **807**, p. L23 (July 2015).
12. S. E. Thorsett, Z. Arzoumanian, F. Camilo and A. G. Lyne, The Triple Pulsar System PSR B1620-26 in M4, *The Astrophysical Journal* **523**, 763 (October 1999).
13. S. Sigurdsson, H. B. Richer, B. M. Hansen, I. H. Stairs and S. E. Thorsett, A Young White Dwarf Companion to Pulsar B1620-26: Evidence for Early Planet Formation, *Science* **301**, 193 (July 2003).
14. J. Antoniadis, P. C. C. Freire, N. Wex, T. M. Tauris, R. S. Lynch, M. H. van Kerkwijk, M. Kramer, C. Bassa, V. S. Dhillon, T. Driebe, J. W. T. Hessels, V. M. Kaspi, V. I. Kondratiev, N. Langer, T. R. Marsh, M. A. McLaughlin, T. T. Pennucci, S. M. Ransom, I. H. Stairs, J. van Leeuwen, J. P. W. Verbiest and D. G. Whelan, A Massive Pulsar in a Compact Relativistic Binary, *Science* **340**, p. 448 (April 2013).
15. P. C. C. Freire, S. M. Ransom, S. Bégin, I. H. Stairs, J. W. T. Hessels, L. H. Frey and F. Camilo, Eight New Millisecond Pulsars in NGC 6440 and NGC 6441, *The Astrophysical Journal* **675**, 670 (March 2008).
16. P. C. C. Freire, A. Wolszczan, M. van den Berg and J. W. T. Hessels, A Massive Neutron Star in the Globular Cluster M5, *The Astrophysical Journal* **679**, 1433 (June 2008).
17. A. Ridolfi, P. C. C. Freire, Y. Gupta and S. M. Ransom, Upgraded Giant Metrewave Radio Telescope timing of NGC 1851A: A possible millisecond pulsar - neutron star system, *Monthly Notices of the Royal Astronomical Society* **490**, 3860 (December 2019).
18. M. H. van Kerkwijk, V. M. Kaspi, A. R. Klemola, S. R. Kulkarni, A. G. Lyne and D. Van Buren, Optical Observations of the Binary Pulsar System PSR B1718-19: Implications for Tidal Circularization, *The Astrophysical Journal* **529**, 428 (January 2000).
19. S. Bogdanov, J. E. Grindlay and M. van den Berg, An X-Ray Variable Millisecond Pulsar in the Globular Cluster 47 Tucanae: Closing the Link to Low-Mass X-Ray Binaries, *The Astrophysical Journal* **630**, 1029 (September 2005).

20. E. S. Phinney, Pulsars as Probes of Globular Cluster Dynamics (January 1993).
21. B. J. Prager, S. M. Ransom, P. C. C. Freire, J. W. T. Hessels, I. H. Stairs, P. Arras and M. Cadelano, Using Long-term Millisecond Pulsar Timing to Obtain Physical Characteristics of the Bulge Globular Cluster Terzan 5, *The Astrophysical Journal* **845**, p. 148 (August 2017).
22. B. B. P. Perera, B. W. Stappers, A. G. Lyne, C. G. Bassa, I. Cognard, L. Guillemot, M. Kramer, G. Theureau and G. Desvignes, Evidence for an intermediate-mass black hole in the globular cluster NGC 6624, *Monthly Notices of the Royal Astronomical Society* **468**, 2114 (June 2017).
23. F. Abbate, A. Possenti, A. Ridolfi, P. C. C. Freire, F. Camilo, R. N. Manchester and N. D'Amico, Internal gas models and central black hole in 47 Tucanae using millisecond pulsars, *Monthly Notices of the Royal Astronomical Society* **481**, 627 (November 2018).
24. P. C. Freire, M. Kramer, A. G. Lyne, F. Camilo, R. N. Manchester and N. D'Amico, Detection of Ionized Gas in the Globular Cluster 47 Tucanae, *The Astrophysical Journal* **557**, L105 (August 2001).
25. F. Abbate, A. Possenti, C. Tiburzi, E. Barr, W. van Straten, A. Ridolfi and P. Freire, Constraints on the magnetic field in the Galactic halo from globular cluster pulsars, *Nature Astronomy* **4**, 704 (March 2020).
26. M. Bagchi, D. R. Lorimer and J. Chennamangalam, Luminosities of recycled radio pulsars in globular clusters, *Monthly Notices of the Royal Astronomical Society* **418**, 477 (November 2011).
27. P. J. Turk and D. R. Lorimer, An empirical Bayesian analysis applied to the globular cluster pulsar population, *Monthly Notices of the Royal Astronomical Society* **436**, 3720 (December 2013).
28. K. Liu, R. P. Eatough, N. Wex and M. Kramer, Pulsar-black hole binaries: Prospects for new gravity tests with future radio telescopes, *Monthly Notices of the Royal Astronomical Society* **445**, 3115 (December 2014).
29. Z. Pan, G. Hobbs, D. Li, A. Ridolfi, P. Wang and P. Freire, Discovery of two new pulsars in 47 Tucanae (NGC 104), *Monthly Notices of the Royal Astronomical Society* **459**, L26 (June 2016).
30. M. Cadelano, S. M. Ransom, P. C. C. Freire, F. R. Ferraro, J. W. T. Hessels, B. Lanzoni, C. Pallanca and I. H. Stairs, Discovery of Three New Millisecond Pulsars in Terzan 5, *The Astrophysical Journal* **855**, p. 125 (March 2018).
31. B. C. Andersen and S. M. Ransom, A Fourier Domain "Jerk" Search for Binary Pulsars, *The Astrophysical Journal Letters* **863**, p. L13 (August 2018).
32. R. Nan, D. Li, C. Jin, Q. Wang, L. Zhu, W. Zhu, H. Zhang, Y. Yue and L. Qian, The Five-Hundred Aperture Spherical Radio Telescope (fast) Project, *International Journal of Modern Physics D* **20**, 989 (January 2011).
33. Z. Pan, S. M. Ransom, D. R. Lorimer, W. C. Fiore, L. Qian, L. Wang, B. W. Stappers, G. Hobbs, W. Zhu, Y. Yue, P. Wang, J. Lu, K. Liu, B. Peng, L. Zhang and D. Li, The FAST Discovery of an Eclipsing Binary Millisecond Pulsar in the Globular Cluster M92 (NGC 6341), *The Astrophysical Journal Letters* **892**, p. L6 (March 2020).
34. L. Wang, B. Peng, B. W. Stappers, K. Liu, M. J. Keith, A. G. Lyne, J. Lu, Y.-Z. Yu, F. Kou, J. Yan, P. Jiang, C. Jin, D. Li, Q. Li, L. Qian, Q. Wang, Y. Yue, H. Zhang, S. Zhang, Y. Zhu and FAST Collaboration, Discovery and Timing of Pulsars in the Globular Cluster M13 with FAST, *The Astrophysical Journal* **892**, p. 43 (March 2020).
35. Z. Pan, L. Qian, X. Ma, K. Liu, L. Wang, J. Luo, Z. Yan, S. Ransom, D. Lorimer, D. Li and P. Jiang, FAST Globular Cluster Pulsar Survey: Twenty-four Pulsars Discovered in 15 Globular Clusters, *The Astrophysical Journal Letters* **915**, p. L28 (July 2021).

36. R. S. Booth and J. L. Jonas, An Overview of the MeerKAT Project, *African Skies* **16**, p. 101 (March 2012).
37. M. Bailes, A. Jameson, F. Abbate, E. D. Barr, N. D. R. Bhat, L. Bondonneau, M. Burgay, S. J. Buchner, F. Camilo, D. J. Champion, I. Cognard, P. B. Demorest, P. C. C. Freire, T. Gautam, M. Geyer, J. M. Griessmeier, L. Guillemot, H. Hu, F. Jankowski, S. Johnston, A. Karastergiou, R. Karuppusamy, D. Kaur, M. J. Keith, M. Kramer, J. van Leeuwen, M. E. Lower, Y. Maan, M. A. McLaughlin, B. W. Meyers, S. Osłowski, L. S. Oswald, A. Parthasarathy, T. Pennucci, B. Posselt, A. Possenti, S. M. Ransom, D. J. Reardon, A. Ridolfi, C. T. G. Schollar, M. Serylak, G. Shaifullah, M. Shamohammadi, R. M. Shannon, C. Sobey, X. Song, R. Spiewak, I. H. Stairs, B. W. Stappers, W. van Straten, A. Szary, G. Theureau, V. Venkatraman Krishnan, P. Weltevrede, N. Wex, T. D. Abbott, G. B. Adams, J. P. Burger, R. R. G. Gamatham, M. Gouws, D. M. Horn, B. Hugo, A. F. Joubert, J. R. Manley, K. McAlpine, S. S. Passmoor, A. Peens-Hough, Z. R. Ramudzuli, A. Rust, S. Salie, L. C. Schwardt, R. Siebrits, G. Van Tonder, V. Van Tonder and M. G. Welz, The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime, *Publications of the Astronomical Society of Australia* **37**, p. e028 (July 2020).
38. M. Bailes, E. Barr, N. D. R. Bhat, J. Brink, S. Buchner, M. Burgay, F. Camilo, D. Champion, J. Hessels, A. Jameson, S. Johnston, A. Karastergiou, R. Karuppusamy, V. Kaspi, M. Keith, M. Kramer, M. McLaughlin, K. Moodley, S. Oslowski, A. Possenti, S. Ransom, F. Rasio, J. Sievers, M. Serylak, B. Stappers, I. Stairs, G. Theureau, W. van Straten, P. Weltevrede and N. Wex, MeerTime - the MeerKAT Key Science Program on Pulsar Timing (January 2016).
39. B. Stappers and M. Kramer, An Update on TRAPUM (January 2016).
40. M. Kramer, I. H. Stairs, V. Venkatraman Krishnan, P. C. C. Freire, F. Abbate, M. Bailes, M. Burgay, S. Buchner, D. J. Champion, I. Cognard, T. Gautam, M. Geyer, L. Guillemot, H. Hu, G. Janssen, M. E. Lower, A. Parthasarathy, A. Possenti, S. Ransom, D. J. Reardon, A. Ridolfi, M. Serylak, R. M. Shannon, R. Spiewak, G. Theureau, W. van Straten, N. Wex, L. S. Oswald, B. Posselt, C. Sobey, E. D. Barr, F. Camilo, B. Hugo, A. Jameson, S. Johnston, A. Karastergiou, M. Keith and S. Osłowski, The relativistic binary programme on MeerKAT: science objectives and first results, *Monthly Notices of the Royal Astronomical Society* **504**, 2094 (June 2021).
41. S. Johnston, A. Karastergiou, M. J. Keith, X. Song, P. Weltevrede, F. Abbate, M. Bailes, S. Buchner, F. Camilo, M. Geyer, B. Hugo, A. Jameson, M. Kramer, A. Parthasarathy, D. J. Reardon, A. Ridolfi, M. Serylak, R. M. Shannon, R. Spiewak, W. van Straten, V. Venkatraman Krishnan, F. Jankowski, B. W. Meyers, L. Oswald, B. Posselt, C. Sobey, A. Szary and J. van Leeuwen, The Thousand-Pulsar-Array programme on MeerKAT - I. Science objectives and first results, *Monthly Notices of the Royal Astronomical Society* **493**, 3608 (April 2020).
42. A. Ridolfi, T. Gautam, P. C. C. Freire, S. M. Ransom, S. J. Buchner, A. Possenti, V. Venkatraman Krishnan, M. Bailes, M. Kramer, B. W. Stappers, F. Abbate, E. D. Barr, M. Burgay, F. Camilo, A. Corongiu, A. Jameson, P. V. Padmanabh, L. Vleeschower, D. J. Champion, W. Chen, M. Geyer, A. Karastergiou, R. Karuppusamy, A. Parthasarathy, D. J. Reardon, M. Serylak, R. M. Shannon and R. Spiewak, Eight new millisecond pulsars from the first MeerKAT globular cluster census, *Monthly Notices of the Royal Astronomical Society* **504**, 1407 (June 2021).
43. F. Abbate, M. Bailes, S. J. Buchner, F. Camilo, P. C. C. Freire, M. Geyer, A. Jameson, M. Kramer, A. Possenti, A. Ridolfi, M. Serylak, R. Spiewak, B. W. Stappers and V. Venkatraman Krishnan, Giant pulses from J1823-3021A observed with the MeerKAT telescope, *Monthly Notices of the Royal Astronomical Society* **498**, 875 (October 2020).

Gravitational lensing by rotating Simpson–Visser black holes

Sushant G. Ghosh

Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi 110025, India; Astrophysics and Cosmology Research Unit, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag 54001, Durban 4000, South Africa
E-mail: sgghosh@gmail.com, sghosh2@jmi.ac.in

Shafqat Ul Islam

Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi 110025, India
E-mail: shafphy@gmail.com

We investigate gravitational lensing by rotating Simpson-Visser black hole in the strong field limit and find that deflection angle α_D, photon sphere radius x_m and ratio of the flux of the first image to all other images r_{mag} decreases. In contrast, angular position θ_1, angular separation s increases more rapidly with l and their behaviour is similar to that of the Kerr black hole. By considering the supermassive black holes NGC 4649, NGC 1332, Sgr A* and M87* as rotating Simpson-Visser black holes, we found that the latter can be quantitatively distinguished from the Kerr black hole when probed by gravitational lensing effects. The deviation of the lensing observables $\Delta\theta_1$ and Δs by considering the Sgr A* and M87* as rotating Simpson-Visser black holes from Kerr black hole for $0 < l/2M < 0.6$ ($a/2M = 0.45$) are in the range $0.0422 - 0.11658$ μas and $0.031709 - 0.08758$ μas, which are too small to be distinguished by the current Event Horizon Telescope observations, and one has to wait for future observations by new generation EHT that can pin down the exact constraint.

Keywords: Lensing, Simpson-Visser black hole

1. Introduction

The singularity theorems of Hawking and Penrose[1] (see also Ref.[2]) established that the end state of gravitational collapse of a sufficiently massive star ($\sim 3.5M_\odot$) is a gravitational singularity i.e., spacetime ceases to exist and hence marking a breakdown of the laws of physics. Apart from the quantum gravity which is expected to resolve the singularity issue,[3] significant attention has also shifted to regular models which can allow one to understand the black hole interior. Sakharov[4] and Gliner[5] pioneered the idea of regular models that suggests a de Sitter core and the cosmological vacuum obeys the equation of state. Motivated by Sakharov[4] and Gliner,[5] Bardeen[6] proposed the first regular black hole with horizons.[7,8] The Bardeen black hole spacetime asymptotically reduces to Schwarzschild black hole[7] but acts like the de Sitter spacetime near the origin thus eventually settling with a regular center. Later, it was shown in Ref.[9] that Bardeen's model is an exact solution of general relativity coupled to nonlinear electrodynamics thereby an alteration of the

Reissner−Nordström black hole solution.[10] Bardeen's proposal with non-linear electrodynamics as the source has been extended to regular black holes,[11–13] and also on regular rotating black[8,14,15] holes.

Lately, Simpson-Visser propose an interesting spherically symmetric regular black hole spacetime[16–18] where a parameter l is introduced to cause a repulsive force to avoid singularity and is responsible for the regularisation of the metric at $r = 0$. The Simpson-Visser metric which interpolates between the Schwarzschild black hole and the Morris–Thorne traversable wormhole is regular everywhere as the scalar.[16] The Simpson-Visser spacetime received significant attention with a discussion on the energy conditions, causal structure, and innermost stable circular orbit (ISCO),[16] regularity, quasi-local mass, energy conditions and the causal structure of black bounce models,[18] gravitational lensing[19] and shadow formation.[22,23]

Graviatational lensing(GL), a vital and prior prediction of general relativity(GR) primarily in the strong field regime,[24] is considered an important test of GR. Since then, GL proved to be an essential tool for providing insights on the spacetime.[25] In the strong field regime, photon encircles the black hole more than once resulting in a bending angle greater than 2π radians and results in shadow,[26] photon rings[27] and relativistic images.[28–32] GL for various spacetimes has been discussed in detail.[29–39] Virbhadra and Ellis developed a numerical method[33] to study the large deflection of light rays in strong field. Later, Bozza analysed analytically the strong field gravitational lensing for general spherically symmetric and static spacetime.[29–32] Tsukamoto[37] worked on similar lines to study the lensing by spherically symmetric and static spacetime. GL began to be crucial in 1990's which triggered quantitative studies by Kerr black holes[29,40,41] and has since then recieved significant attention[42–49] due to the tremendous advancement of observational facilities.

This paper investigates the strong gravitational lensing by rotating Simpson-Visser black holes[20] and assess the phenomenological differences with the Kerr black holes; in particular, we discuss the effect of deformation parameter l on gravitational lensing observables and time delay between the relativistic images. Further, considering the supermassive black holes as the lens, we obtain the positions, separation, magnification, and time delay in the formation of relativistic images.

We adopt natural units ($8\pi G = c = 1$)

2. Rotating Simpson-Visser black holes

The rotating Simpson-Visser black hole metric apart from mass M and angular momentum a, has an additional parameter l, which in Boyer-Lindquist coordinates reads as[20]

$$ds^2 = -\left(1 - \frac{\sqrt{x^2 + l^2}}{\Sigma}\right) dt^2 + \frac{\Sigma}{\Delta} dx^2 + \Sigma d\theta^2 - \frac{2a \sin^2\theta \sqrt{x^2 + l^2}}{\Sigma} dt d\phi + \frac{\mathbb{A} \sin^2\theta}{\Sigma} d\phi^2, \tag{1}$$

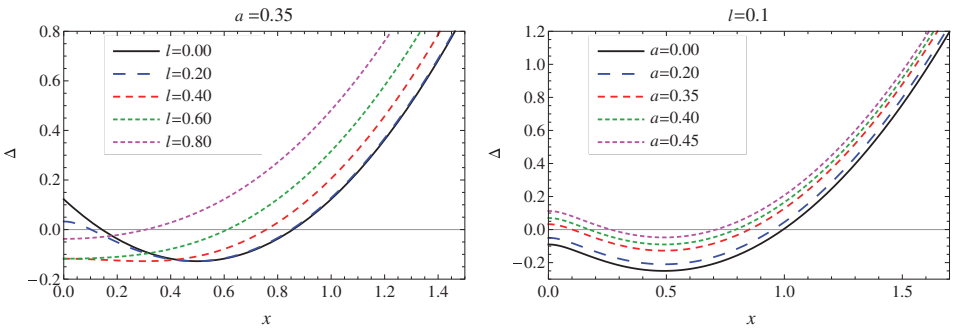


Fig. 1. The horizons (zeroes of Δ) of rotating Simpson-Visser black holes. The case $l = 0$ (Kerr black hole) is shown for comparison. For $l > l_c$ (e.g., for $a = 0.35$, $l_c \approx 0.1429$ and for $a = 0.45$, $l_c \approx 0.282$), the black holes admit only event horizon.

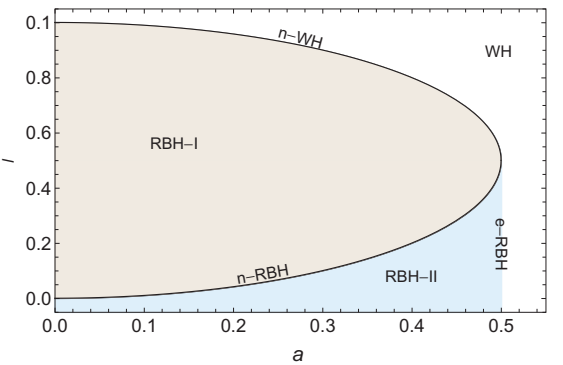


Fig. 2. Parameter space for the rotating Simpson-Visser black hole and the corresponding spacetime structure. For $a = 0.3$ the rotating Simpson-Visser black hole admits both Cauchy and event horizon for $0.0 < l_{cb} < 0.1$ and only event horizon for $0.1 < l_c < 0.9$. For $a = 0.45$, both Cauchy and event horizon exist for $0.0 < l_{cb} < 0.282055$ and only event horizon is present for $0.282055 < l_c < 0.717945$. We measure the quantities a and l in units of Schwarzschild radius $2M$.

where $\Sigma = x^2 + l^2 + a^2\cos^2\theta$, $\Delta = x^2 + l^2 + a^2 - 2\sqrt{x^2 + l^2}$ and $\mathbb{A} = (x^2 + l^2 + a^2)^2 - \Delta a^2 \sin^2\theta$. We measure the quantities r, a, t, and l in units of Schwarzschild radius $2M$[31] and use x instead of radius r.

Moreover, the metric (1) is composed of identical portions around $x = 0$ which is a regular surface of finite size for $l \neq 0$, and the observer can easily cross.[20] Depending on the value of the parameter l, the metric (1) is either a regular black hole or a traversable wormhole.[20]

The points of coordinate singularities, which exist at $\Delta = 0$, are also the horizons of the black hole metric (1) and are located at $x_\pm$ (cf. Fig. 1). The horizons of the metric (1) exist when $|a| < 0.5$ and $l < 0.5 + \sqrt{(0.5)^2 - a^2}$ (cf. Fig. 1). Consequently, following cases are possible for the metric (1)

- For $l < 0.5 - \sqrt{(0.5)^2 - a^2}$, we have two horizons $x_\pm$, corresponding to the radii of outer (event) (x_+) and inner (Cauchy) (x_-) horizons (This corresponds to RBH-II, in Fig. 2).
- For $0.5 - \sqrt{(0.5)^2 - a^2} < l < 0.5 + \sqrt{(0.5)^2 - a^2}$, only event horizon exists. (This corresponds to RBH-I, in Fig. 2).
- For $|a| = 0.5$ and $l < 0.5$, Cauchy and event horizon correspond to extremal black hole (e-RBH) with degenerate horizons, while for $|a| = 0.5$ and $l > 0.5$, there are no horizons and so no black hole.
- The case $|a| > 0.5$ with no horizons, the metric (1) describes a traversable wormhole with a time-like throat, which becomes null for $|a| = 0.5$.[20]

We restrict the values of a and l in the range $|a| < 0.5$ and $0 < l < 0.5 + \sqrt{(0.5)^2 - a^2}$ for at least one horizon to exist and to qualify the spacetime as a black hole. We shall also consider the values $x > 0$ as the solution (1) is symmetric around $x = 0$.

3. Gravitational lensing

The rotating Simpson-Visser black hole in the equatorial plane takes the form

$$\mathrm{d}s^2 = -A(x)dt^2 + B(x)dx^2 + C(x)d\phi^2 - D(x)dt\,d\phi, \tag{2}$$

where

$$A(x) = 1 - \frac{1}{\sqrt{\Sigma}}, \quad B(x) = \frac{\Sigma}{\Delta}, \quad C(x) = \frac{\mathbb{A}}{\Sigma}, \quad D(x) = \frac{2a}{\sqrt{\Sigma}},$$

$\Sigma = x^2 + l^2$ and $\mathbb{A} = (x^2 + l^2 + a^2)^2 - \Delta a^2$. Further, we also assume the source, which is located behind the lens, to be almost aligned along the optical axis such that the images are highly magnified and very prominent.

The projection of four-momentum p^μ along the two commuting Killing vectors $\xi^\mu_{(t)}$ and $\xi^\mu_{(\phi)}$ results in two linearly independent conserved quantities; energy E and axial angular momentum component L.

$$p_t = g_{tt}p^t + g_{t\phi}p^\phi = -E, \qquad p_\phi = g_{t\phi}p^t + g_{\phi\phi}p^\phi = L. \tag{3}$$

Solving the two coupled equations results in the following equations of motion for t and ϕ coordinates[56]

$$\Sigma\frac{dt}{d\tau} = a(L - aE\sin^2\theta) + \frac{x^2 + l^2 + a^2}{\Delta}[E(x^2 + l^2 + a^2) - La] \tag{4}$$

$$\Sigma\frac{d\phi}{d\tau} = \frac{L}{\sin^2\theta} - aE + \frac{a}{\Delta}[E(x^2 + l^2 + a^2) - La], \tag{5}$$

We follow the Hamilton-Jacobi equation and the integral approach pioneered by Carter[57] to obtain the equations of motion in the first-order differential form for

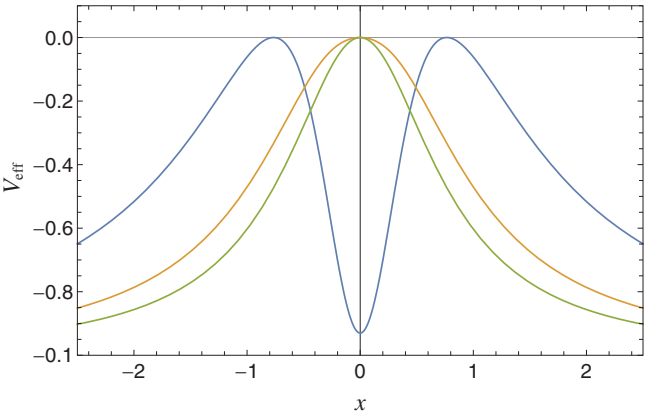


Fig. 3. The effective potential as a function of radial coordinate x. (a) The blue line represents the effective potential for rotating Simpson-Visser black hole black hole with $a = 0.334$ and $l = 0.698$ (b) the orange line represents the wormhole throat with $a = 0.5$ and $l = 0.7$ (c) the green line represents the traversable wormhole with $a = 0.884$ and $l = 0.746$.

the rotating Simpson-Visser black hole spacetime (1), where the equations descend from

$$\frac{\partial S}{\partial \tau} = -\frac{1}{2} g^{\mu\nu} \frac{\partial S}{\partial x^{\mu}} \frac{\partial S}{\partial x^{\nu}}, \tag{6}$$

where S is the action and τ is the affine parameter along the geodesic. Using the conserved quantities p_t and p_ϕ, the Jacobi action S admits a separable solution in the first integral form as follows[57]

$$S = -Et + L\phi + S_x(x) + S_\theta(\theta). \tag{7}$$

Here, $S_x(x)$ and $S_\theta(\theta)$, respectively, are functions only of the x and θ coordinates. On using the four-momentum relations

$$\frac{dS_x}{dx} = p_x = g_{xx} p^x = \frac{\Sigma}{\Delta} \frac{dx}{d\tau}, \tag{8}$$

$$\frac{dS_\theta}{d\theta} = p_\theta = g_{\theta\theta} p^\theta = \Sigma \frac{d\theta}{d\tau}, \tag{9}$$

in Eq. (6), the following system of first-order ordinary differential equations are obtained

$$\Sigma \frac{dx}{d\tau} = \pm \sqrt{\mathcal{R}(x)}\ , \tag{10}$$

$$\Sigma \frac{d\theta}{d\tau} = \pm \sqrt{\Theta(\theta)}\ , \tag{11}$$

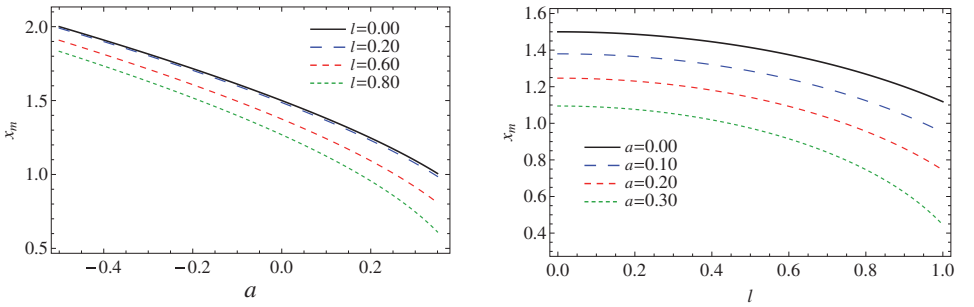


Fig. 4. Plot showing radii of unstable circular photon orbit for rotating Simpson-Visser black hole (dashed curve) in comparison to Kerr black hole (solid curve in the left plot).

where the effective potentials $\mathcal{R}(x)$ and $\Theta(\theta)$ for radial and polar motion are given by

$$\mathcal{R} = [E(x^2 + l^2 + a^2) - La]^2 - \Delta[(L - aE)^2 + \mathcal{K}], \tag{12}$$

$$\Theta = \mathcal{K} - \cos^2\theta\left[a^2E^2 + \frac{L^2}{\sin^2\theta}\right]. \tag{13}$$

$\mathcal{K}$ is a separability constant related to the Carter constant that derives its existence from a Killing tensor. The motion of the photon is symmetric around the equatorial plane in the region $\mathcal{R} \geq 0$ and $\Theta(\theta) \geq 0$. The bound orbits can be identified by the solutions of[56, 58]

$$\mathcal{R}|_{(x=x_m)} = \left.\frac{\partial\mathcal{R}}{\partial x}\right|_{(x=x_m)} = 0, \text{ and } \left.\frac{\partial^2\mathcal{R}}{\partial x^2}\right|_{(x=x_m)} > 0, \tag{14}$$

where x_m is the unstable photon orbit radius. Eq.(10) can be rewritten as

$$\dot{x}^2 + V_{\text{eff}} = 0 \tag{15}$$

where V_{eff} is the effective potential. For simplicity we take $E = 1$ and plot the effective potential in Fig. 3.

The deflection angle of a photon moving on the equatorial plane using the geodesics Eqs. (5) and (10), for rotating Simpson-Visser black hole is given by[32]

$$\alpha_D(x_0) = 2\int_{x_0}^{\infty} \frac{\sqrt{A_0 B}\,(2AL + D)}{\sqrt{4AC + D^2}\sqrt{A_0 C - AC_0 + L\,(AD_0 - A_0 D)}}dx - \pi, \tag{16}$$

where x_0 is the distance of closest approach of the light to the lens, L is the projection of angular momentum and A_0 is defined by $A(x_0)$. Equation (14) gives the radii of photon sphere which can be rewritten as[32]

$$A(x)C'(x) - A'(x)C(x) + L(A'(x)D(x) - A(x)D'(x)) = 0. \tag{17}$$

The radii of unstable photon sphere is depicted in Fig. 4 and is smaller when compared with the Kerr black hole.[32] Further, the deflection angle when expanded in

the SDL takes the form

$$\alpha_D(\theta) = -\bar{a}\log\left(\frac{\theta D_{OL}}{u_m} - 1\right) + \bar{b} + \mathcal{O}\left(u - u_m\right), \tag{18}$$

where u is the impact parameter and u_m is the value of impact parameter at $x_0 = x_m$ and is independent of the parameter l (cf. Table 1). The deflection angle in Eq.(18) is depicted in Fig. 6 wherein it is evident that deflection angle diverges at u_m.

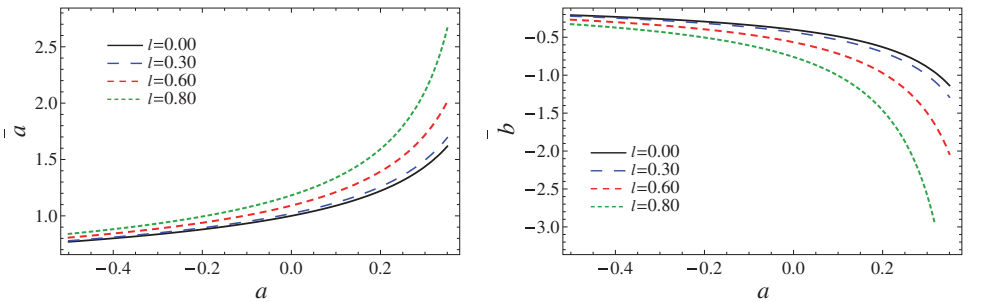


Fig. 5. Plot showing the behaviour of strong lensing coefficients $\bar{a}$ and $\bar{b}$. The solid black curve corresponds to Kerr black hole.

Table 1. Estimates for the strong lensing coefficients $\bar{a}$, $\bar{b}$ and the impact parameter u_m/R_s. We measure the quantities a and l in units of Schwarzschild radius $2M$.

		Lensing Coefficients		
a	l	$\bar{a}$	$\bar{b}$	u_m/R_s
	0.00	1.00000	-0.400230	2.59808
	0.25	1.01419	-0.423912	2.59808
	0.50	1.06066	-0.506783	2.59808
0.00	0.75	1.15470	-0.698106	2.59808
	0.95	1.29219	-1.03047	2.59808
	0.00	1.43692	-0.891314	1.91924
	0.25	1.47594	-0.965136	1.91924
	0.50	1.61535	-1.256170	1.91924
0.30	0.75	1.97301	-2.180740	1.91924
	0.90	2.52527	-4.037170	1.91924
	0.00	2.58352	-2.76735	1.42221
	0.30	2.79948	-3.26402	1.42221
	0.40	3.01084	-3.79273	1.42221
0.45	0.50	3.3693	-4.7799	1.42221
	0.60	4.05135	-6.94521	1.42221

The behaviour of coefficients $\bar{a}$ and $\bar{b}$ is similar to the Kerr black hole ($l = 0$) and spherical Simpson-Visser ($a = 0$) (cf. Fig. 5). The deflection angle for rotating

Simpson-Visser black hole in SDL (cf. Fig. 6) like Kerr black hole, is more than 2π radians and multiple images of the source are formed.

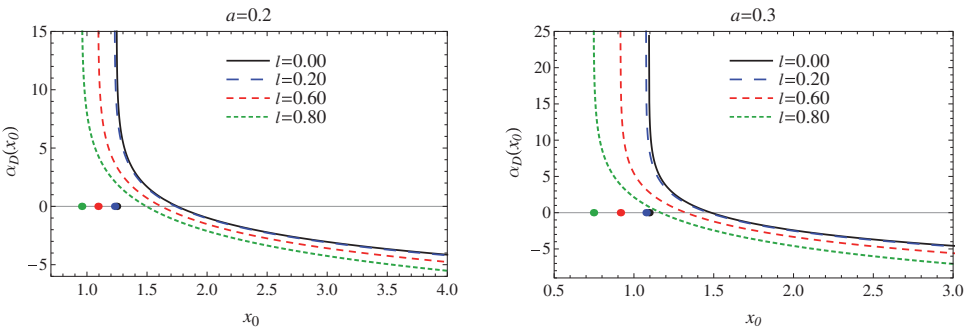


Fig. 6. Plot showing the variation of light deflection angle as a function of minimum distance x_0 for different values of a and l.

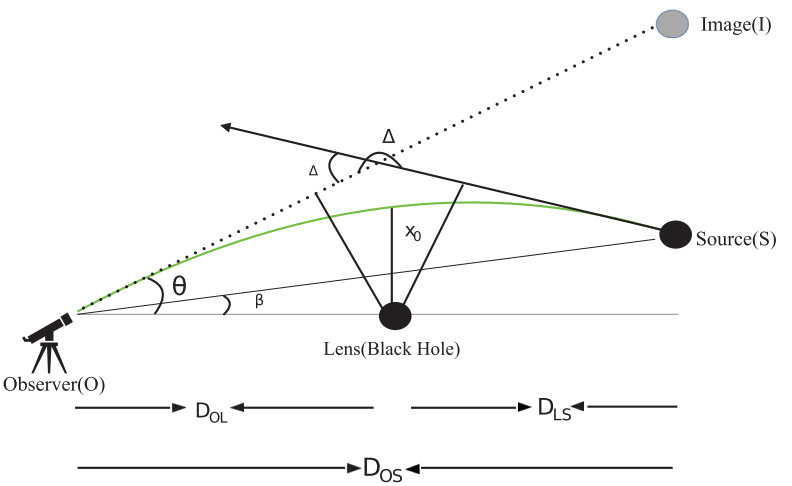


Fig. 7. Schematic diagram depicting the lensing configuration. α_D represent the deflection angle, where as θ and β are the angular positions of the image and source, respectively.

3.1. *Observables*

Next, we consider the lens equation to define a geometrical relationship between the observer, the source and the lens (cf. Fig. 7) as given in,[31,54]

$$\beta = \theta - \frac{D_{LS}}{D_{OL} + D_{LS}} \Delta\alpha_n, \tag{19}$$

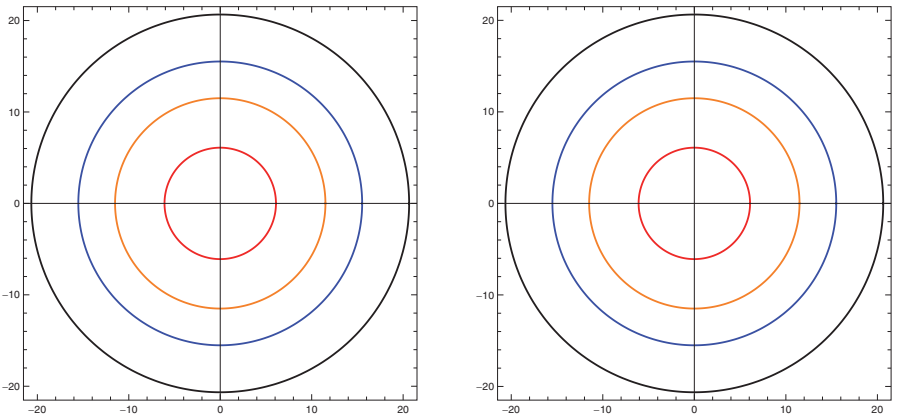


Fig. 8. The polar plot of θ_n^E (for $n = 1$): The angular radius of the outermost Einstein ring. Here, we consider Kerr black hole ($a = 0.25$ and $l = 0$) (left) and rotating Simpson-Visser balck hole ($a = 0.25$ and $l = 0.7$) (right) as supermassive black holes Sgr A* (black), M87* (blue), NGC 4649 (orange) and NGC 1332 (red).

where $\Delta\alpha_n$ is the extra deflection angle with $n \in N$ and $0 < \Delta\alpha_n \ll 1$. It has been acessed in[54] that lens equation discussed in[55] is bit more accurate than Eq.(19) but the latter has an advantage in analysing the observational effects more precisely. We shall estimate the observables for the strong gravitational lensing by rotating Simpson-Visser black hole as in.[31,32,42,52] Using the lens Eq. (19) and Eq. (18), the angular separation between the lens and the $n-$th image is given by[32]

$$\theta_n = \frac{u_m}{D_{OL}}(1 + e_n) + \frac{D_{OL} + D_{LS}}{D_{LS}} \frac{u_m e_n}{\bar{a} D_{OL}} (\beta - {\theta_n}^0), \tag{20}$$

where

$$e_n = \exp\left(\frac{\bar{b}}{\bar{a}} - \frac{2n\pi}{\bar{a}}\right). \tag{21}$$

Here the first term corresponds to value of θ when photon travels $2n\pi$ radians around the black hole and second term is the correction part to $2n\pi$.[31] For $\beta = 0$, a ring-shaped image is produced, known as Einstein rings.[25] Solving Eq. (20) for $\beta = 0$ and $D_{OL} = D_{LS}$, the angular radius of the Einstein rings can be obtained as[31,51]

$$\theta_n^E = \frac{u_m}{D_{OL}} (1 + e_n), \tag{22}$$

where $n = 1$ is the angular radius of the outermost Einstein ring (cf. Fig. 8).[42,52]

The brightness or magnification of the image is preserved in the deflection of the light but the appearance of the solid angle changes, thus magnifying the brightness

of the images. The magnification of n-loop image is given by[31,32]

$$\mu_n = \left|\frac{\beta}{\theta}\frac{\partial\beta}{\partial\theta}\right|^{-1}_{\theta_n^0} = \frac{1}{\beta}\left[\frac{u_m}{D_{OL}}(1+e_n)\left(\frac{D_{OL}+D_{LS}}{D_{LS}}\frac{u_m e_n}{D_{OL}\bar{a}}\right)\right]. \tag{23}$$

Table 2. Estimates for the observable θ_∞ and time delay $\Delta T_{2,1}$ for supermassive black holes at the center of nearby galaxies. These observables are independent of l.

	Sgr A*		M87*		NGC 4649		NGC 1332	
a	$\theta_\infty(\mu$as)	$\Delta T_{2,1}$(min)	$\theta_\infty(\mu$as)	$\Delta T_{2,1}$(hrs)	$\theta_\infty(\mu$as)	$\Delta T_{2,1}$(hrs)	$\theta_\infty(\mu$as)	$\Delta T_{2,1}$(hrs)
0.00	26.3299	11.4968	19.782	289.647	14.6615	210.328	7.76719	113.185
0.30	19.4504	8.4928	14.6133	213.968	10.8307	155.374	5.73777	83.612
0.45	14.4132	6.29295	10.8289	158.543	8.02585	115.127	4.2518	61.9538

The Eq. (23) diverges at $\beta \to 0$, suggesting that the perfect alignment maximises the possibility of the detection of the images. If the first image can be distinguished from the rest packed ones, we can have three characteristic observables[31] as

$$\theta_\infty = \frac{u_m}{D_{OL}}, \tag{24}$$

$$s = \theta_1 - \theta_\infty \approx \theta_\infty \exp\left(\frac{\bar{b}}{\bar{a}} - \frac{2\pi}{\bar{a}}\right), \tag{25}$$

$$r_{\rm mag} = \frac{\mu_1}{\sum_{n=2}^{\infty}\mu_n} \approx \frac{5\pi}{\bar{a}\,\log(10)}, \tag{26}$$

where, θ_∞ and θ_1 are the angular positions of the innermost image and outermost image, respectively and s is the angular separation between them. $r_{\rm mag}$ gives the ratio of the flux of the first image and the all other images. By observationally measuring these observables and inverting the Eqs. (24-26), the lensing coefficients $\bar{a}$, $\bar{b}$, the minimum impact parameter u_m can be obtained.[53]

The time delay, which is another important observable, is defined as the time difference between the formation of relativistic images considering the source with luminosity variations. If the time signal between the first and second image can be distinguished, the time delay $\Delta T_{2,1}$, when images are on the same side of the lens, can be approximated as[50]

$$\begin{aligned}\Delta T_{2,1} &\approx \frac{2\pi}{\bar{a}\sqrt{c_{2m}}}\frac{2x^2\sqrt{B(x)A(x_0)}[C(x)-LD(x)]}{x_0\sqrt{C(x)(D(x)^2+4A(x)C(x))}}\left(1-\frac{1}{\sqrt{A(x_0)}f(z,x_0)}\right)\\ &= 2\pi u_m,\end{aligned} \tag{27}$$

To measure the time delay, it is necessary that the source should be pulsating and the pulses show up in the images with a temporal phase thus making it possible to measure the time difference.

Table 3. Estimates for the lensing observables for supermassive black holes at the center of nearby galaxies for different values of a and l.

		Sgr A*		M87*		NGC 4649		NGC 1332		
a	l	$\theta_1(\mu as)$	s (μas)	$\theta_1(\mu as)$	s (μas)	$\theta_1(\mu as)$	s (μas)	$\theta_1(\mu as)$	s (μas)	r_{mag}
	0.00	26.3628	0.0329517	19.8068	0.0247571	14.6799	0.0183488	7.77691	0.00972061	6.82188
	0.25	26.3652	0.0353455	19.8086	0.0265556	14.6812	0.0196818	7.77762	0.0104268	6.72647
	0.50	26.3735	0.0436767	19.8148	0.032815	14.6858	0.0243209	7.78008	0.0128844	6.43173
0.00	0.75	26.3922	0.0623325	19.8289	0.0468314	14.6962	0.0347092	7.78558	0.0183878	5.90792
	0.95	26.4216	0.091705	19.8509	0.0688993	14.7126	0.051065	7.79425	0.0270526	5.27931
	0.0	19.5823	0.131981	14.7125	0.0991593	10.9042	0.0734923	5.7767	0.0389338	4.74756
	0.25	19.5936	0.143258	14.721	0.107632	10.9105	0.0797717	5.78003	0.0422604	4.62205
	0.50	19.6331	0.18278	14.7507	0.137326	10.9325	0.101779	5.79169	0.0539194	4.22317
0.30	0.75	19.717	0.266607	14.8137	0.200306	10.9792	0.148458	5.81642	0.078648	3.45759
	0.90	19.777	0.326616	14.8587	0.245392	11.0126	0.181873	5.83412	0.0963504	2.70145
	0.0	14.8471	0.433859	11.1548	0.325965	8.26744	0.24159	4.37982	0.127986	2.64054
	0.30	14.8893	0.476064	11.1865	0.357674	8.29094	0.265092	4.39227	0.140437	2.43684
	0.40	14.9206	0.50742	11.21009	0.38123	8.3084	0.28255	4.40152	0.149689	2.26578
0.45	0.50	14.9537	0.540453	11.2349	0.40605	8.32679	0.30094	4.41126	0.15943	2.02472
	0.60	14.9636	0.550440	11.2424	0.41355	8.33235	0.306507	4.41421	0.162377	1.15820

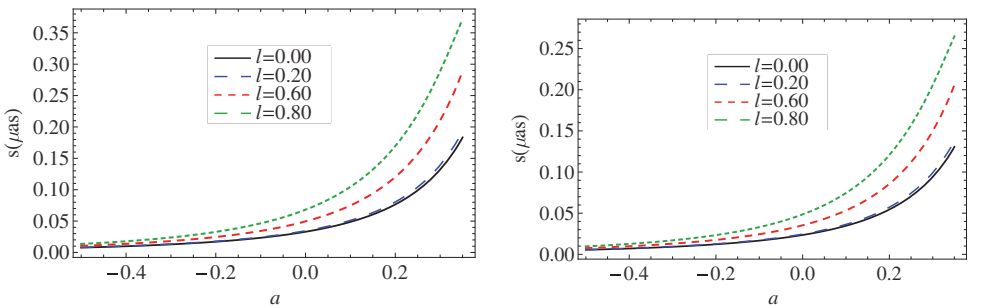


Fig. 9. Plot showing strong lensing observables s for the Sgr A*(left panel) and M87*(right panel) black holes.

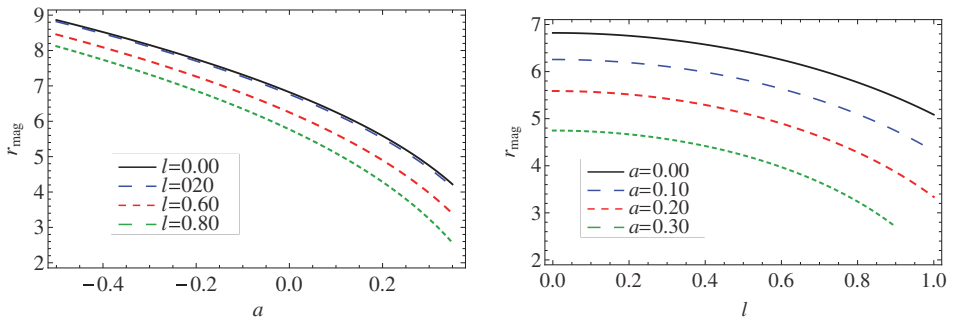


Fig. 10. Plot showing strong lensing observables r_{mag} for different parameters. r_{mag} is independent of the black hole mass or its distance from the observer.

4. Applications to supermassive black holes

Here, we model the supermassive black holes Sgr A*, M87*, NGC4649 and NGC1332 as the rotating Simpson-Visser black holes to estimate and compare the observables in SDL with the GR counterparts, and use mass and distance as in the Ref.[42]. The observables θ_∞ and time delay $\Delta T_{2,1}$ are independent of the parameter l and are shown in Table 2. We observe that θ_1 for Sgr A* and M87* is in the range of 14.84 μas $< \theta_1 <$ 26.36 μas and 11.15 μas $< \theta_1 <$ 19.80 μas, respectively. The observables s and r_{mag} for rotating Simpson-Visser black holes are depicted in Fig. 9 and 10, whereas their values, along with the position of the first image θ_1 are tabulated in Table 3. Further, the angular separation s between the first and inner most image due to the rotating Simpson-Visser black hole for Sgr A* and M87* are in the range of 32.95 $n$as $< s <$ 433.8 $n$as and 24.75 $n$as $< s <$ 413.55 $n$as, respectively (cf. Table 3). Although the rotating Simpson-Visser black holes deviate from the Kerr black hole but the deviation are beyond the threshold of the current EHT observation (cf. Table 4) and should wait for the next generation event horizon telescope (ngEHT). From the brightness difference, we find that Kerr black holes produces less intense images. The angular position and angular separation of the relativistic images are larger for the Sgr A* black hole than the M87*, NGC4649 and NGC1332 black holes. The time delay of the first image from that of the second image $\Delta T_{2,1}$, for Sgr A*, M87*, NGC4649 and NGC1332 respectively, can reach $\sim$ 11.49 min, $\sim$ 289.6 hrs, $\sim$ 210.3 hrs and $\sim$ 113.2 hrs (cf. Table 2).

Table 4. Deviation of the lensing observables of rotating Simpson Visser black holes from Kerr black hole for supermassive black holes at the center of nearby galaxies for $a = 0.3$ and $a = 0.45$. For $a = 0.3$ the rotating Simpson-Visser black hole admits both Cauchy and event horizon for $0.0 < l_{cb} < 0.1$ and only event horizon for $0.1 < l_c < 0.9$. For $a = 0.45$, both Cauchy and event horizon exist for $0.0 < l_{cb} < 0.282055$ and only event horizon is present for $0.282055 < l_c < 0.717945$. Here $\Delta(X) = X_{RSV} - X_{Kerr}$.

		Sgr A*		M87*		NGC 4649		NGC 1332		
a	l	$\Delta\theta_1(\mu$as)	Δs (μas)	$\Delta\theta_1(\mu$as)	Δs (μas)	$\Delta\theta_1(\mu$as)	Δs (μas)	$\Delta\theta_1(\mu$as)	Δs (μas)	Δr_{mag}
	0.25	0.0112	0.0112769	0.0080	0.00803363	0.00627	0.00627941	0.0149	0.0149856	-0.125517
	0.50	0.0508	0.0507993	0.0362	0.0361893	0.02828	0.028287	0.0149	0.0149856	-0.524392
	0.75	0.1346	0.134626	0.0959	0.0959075	0.0749	0.0397141	0.0397	0.078648	-1.28997
0.30	0.90	0.1946	0.194635	0.1386	0.138658	0.108381	0.0574166	0.0574	0.0963504	-2.04612
	0.30	0.0422	0.042205	0.0317	0.0317098	0.0235	0.0235018	0.0124	0.0124505	-0.2037
	0.40	0.0735	0.073560	0.0552	0.0552727	0.0409	0.0409655	0.02170	0.0217022	-0.3747
	0.50	0.1065	0.10659	0.08008	0.0800856	0.05935	0.0593557	0.03144	0.0314447	-0.615823
0.45	0.60	0.1165	0.116582	0.08758	0.0875896	0.06491	0.0649173	0.03439	0.0343911	-0.956684

5. Conclusions

We have investigated the gravitational lensing of light in strong field limit due to rotating Simpson-Visser black holes which has an additional deviation parameter l besides mass M and spin a. We find that $\bar{a}$ increases whereas $\bar{b}$ and deflection angle α_D decrease with increasing l. On the other hand, u_m depends only on the spin of

the black hole and is independent of the black hole parameter l. The observables θ_∞ and $\Delta T_{2,1}$ do not depend of l. Further, the relative magnification of the first and innermost images is independent of the black hole mass or its distance from the observer. Interestingly, it turns out that the separation between the two relativistic images for Simpson-Visser black holes is greater than that for the Kerr black holes.

Although, the gravitational lensing analysis (cf. Figs. 6, 9 and 10) and horizon structure (cf. Figs. 1 and 2), constrain the rotating Simpson-Visser black holes parameter l but the deviation produced in lensing observables by taking the rotating Simpson-Visser black holes as the supermassive black holes, namely, Sgr A*, M87*, NGC 4649 and NGC 1332, from those for Kerr black hole for $0 < l < 0.6$ ($a = 0.45$) are $\mathcal{O}(\mu\text{as})$ (cf. Table 4). Our results are the generalization of previous discussions, on the Kerr black holes, in a more general setting in the limit $l \to 0$, reduced exactly to *vis-à-vis* the Kerr black hole results.

Acknowledgments

S.G.G. and S.U.I. would like to thank SERB-DST for the ASEAN project IMRC/AISTDF/CRD/2018/000042.

References

1. S.W. Hawking and R. Penrose, Proc. R. Soc. A **314**, 529 (1970).
2. S.W. Hawking and G.F.R. Ellis, The Large Scale Structure of Space and Time (Cambridge University Press, Cambridge, 1973).
3. J.A. Wheeler, Relativity, Groups, and Topology, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).
4. A.D. Sakharov, Sov. Phys. JETP, **22**, 241 (1966).
5. E.B. Gliner, Sov. Phys. JETP, **22**, 378 (1966).
6. J. Bardeen, in *Proceedings of GR5* (Tiflis, U.S.S.R., 1968).
7. S. Ansoldi, [arXiv:0802.0330 [gr-qc]].
8. S.G. Ghosh, Eur. Phys. J. C **75**, 532 (2015).
9. E. A. Beato and A. Garcia, Phys. Lett. B **464**, 25 (1999); E.A. Beato and A. Garcia, Gen. Relativ. Gravit. **31**, 629 (1999).
10. H. Reissner, Ann. Phys. **355**, 9 (1916).
11. E. A. Beato and Alberto Garcia Phys. Rev. Lett. **80**, 5056 (1998); S.A. Hayward, Phys. Rev. Lett. **96**, 031103 (2006); I. Dymnikova, Gen Relativity Gravity **24**, 235 (1992); Z. Y. Fan and X. Wang, Phys. Rev. D **94**, 124027 (2016); O.B. Zaslavskii, Phys. Rev. D **80**, 064034 (2009).
12. J.P.S. Lemos and V.T. Zanchin, Phys. Rev. D **83**, 124005 (2011); J. C. S. Neves and A. Saa, Phys. Lett. B **734**, 44 (2014); L. Balart and E. C. Vagenas, Phys. Rev. D **90**, 124045 (2014); S. G. Ghosh and S. D. Maharaj, Phys. Dark Univ. **31**, 100793 (2021); M. S. Ali and S. G. Ghosh, Phys. Rev. D **98**, 084025 (2018); A. Kumar, D. Veer Singh and S. G. Ghosh, Eur. Phys. J. C **79**, 275 (2019); S. G. Ghosh, Class. Quant. Grav. **35**, 085008 (2018).
13. A. Cisterna, G. Giribet, J. Oliva and K. Pallikaris, Phys. Rev. D **101**, 124041 (2020).
14. A. Kumar, S. G. Ghosh and S. D. Maharaj, Phys. Dark Univ. **30**, 100634 (2020); R. Kumar, A. Kumar and S. G. Ghosh, Astrophys. J. **896**, 89 (2020); R. Kumar and

S. G. Ghosh, Class. Quant. Grav. **38**, 8 (2021); R. Kumar, S. G. Ghosh and A. Wang, Phys. Rev. D **100**, 124024 (2019); R. Kumar and S. G. Ghosh, Astrophys. J. **892**, 78 (2020); A. Abdujabbarov, M. Amir, B. Ahmedov and S. G. Ghosh, Phys. Rev. D **93**, 104004 (2016); M. Amir and S. G. Ghosh, Phys. Rev. D **94**, 024054 (2016); S. G. Ghosh and M. Amir, Eur. Phys. J. C **75**, 553 (2015); M. Amir and S. G. Ghosh, JHEP **07**, 015 (2015); S. G. Ghosh and S. D. Maharaj, Eur. Phys. J. C **75**, 7 (2015); C. Bambi and L. Modesto, Phys. Lett. B **721**, 329 (2013).

15. A. Eichhorn and A. Held, [arXiv:2103.07473 [gr-qc]]. A. Eichhorn and A. Held, [arXiv:2103.13163 [gr-qc]].
16. A. Simpson and M. Visser, JCAP **02**, 042 (2019).
17. A. Simpson, P. Martin-Moruno and M. Visser, Class. Quant. Grav. **36**, 145007 (2019).
18. F. S. N. Lobo, M. E. Rodrigues, M. V. d. S. Silva, A. Simpson and M. Visser, [arXiv:2009.12057 [gr-qc]].
19. N. Tsukamoto, Phys. Rev. D **103**, 024033 (2021).
20. J. Mazza, E. Franzin and S. Liberati, [arXiv:2102.01105 [gr-qc]].
21. K. A. Bronnikov, R. A. Konoplya and T. D. Pappas, [arXiv:2102.10679 [gr-qc]].
22. H. C. D. L. Junior, L. C. B. Crispino, P. V. P. Cunha and C. A. R. Herdeiro, [arXiv:2102.07034 [gr-qc]].
23. R. Shaikh, K. Pal, K. Pal and T. Sarkar, [arXiv:2102.04299 [gr-qc]].
24. L. C. B. Crispino and D. Kennefick, Nature Phys. **15**, 416 (2019); L. C. B. Crispino and S. Paolantonio, Nature Astron. **4**, 6 (2020).
25. A. Einstein, Science **84**, 506 (1936); S. Liebes, Phys. Rev. **133**, B835 (1964); Y. Mellier, Ann. Rev. Astron. Astrophys. **37**, 127 (1999); M. Bartelmann and P. Schneider, Phys. Rept. **340**, 291(2001); F. Schmidt, Phys. Rev. D **78**, 043002 (2008); Guzik, Jacek and Jain, Bhuvnesh and Takada, Masahiro, Phys. Rev. D. **81**, 023503 (2010).
26. J. L. Synge,Mon. Not. R. Astron. Soc. **131**, 463 (1966).
27. S. E. Gralla, D. E. Holz and R. M. Wald, Phys. Rev. D **100**, 024018 (2019); M. D. Johnson, A. Lupsasca, A. Strominger, G. N. Wong, S. Hadar, D. Kapec, R. Narayan, A. Chael, C. F. Gammie and P. Galison, *et al.* Sci. Adv. **6**, 1310 (2020); S. E. Gralla and A. Lupsasca, Phys. Rev. D **102**, 124003 (2020).
28. C. Darwin, Proc. R. Soc. A **249**, 180 (1959); P. V. P. Cunha and C. A. R. Herdeiro, Gen. Rel. Grav. **50**, 42 (2018).
29. V. Bozza, Gen. Rel. Grav. **42**, 2269 (2010).
30. V. Bozza, S. Capozziello, G. Iovane and G. Scarpetta, Gen. Rel. Grav. **33**, 1535 (2001).
31. V. Bozza, Phys. Rev. D **66**, 103001 (2002).
32. V. Bozza, Phys. Rev. D **67**, 103006 (2003).
33. K. S. Virbhadra and G. F. R. Ellis, Phys. Rev. D **62**, 084003 (2000).
34. S. Frittelli, T. P. Kling and E. T. Newman, Phys. Rev. D **61**, 064021 (2000).
35. E. F. Eiroa, G. E. Romero and D. F. Torres, Phys. Rev. D **66**, 024010 (2002)
36. S. V. Iyer and A. O. Petters, Gen. Rel. Grav. **39**, 1563 (2007).
37. N. Tsukamoto, Phys. Rev. D **95**, 064035 (2017); N. Tsukamoto and Y. Gong, Phys. Rev. D **95**, 064034 (2017).
38. K. S. Virbhadra and C. R. Keeton, Phys. Rev. D **77**, 124014 (2008).
39. R. Shaikh, P. Banerjee, S. Paul and T. Sarkar, JCAP **07**, 028 (2019).
40. Rauch K P and Blandford R D, Astrophys. J. **46** 421 (1994).
41. V. Bozza, Phys. Rev. D **78**, 063014 (2008).
42. S. G. Ghosh, R. Kumar and S. U. Islam, JCAP **03**, 056 (2021).
43. S. W. Wei, Y. X. Liu, C. E. Fu and K. Yang, JCAP **1210**, 053 (2012).
44. K. Beckwith and C. Done, Mon. Not. Roy. Astron. Soc. **359**, 1217 (2005).
45. Y. W. Hsiao, D. S. Lee and C. Y. Lin, Phys. Rev. D **101**, 064070 (2020).

46. D. Kapec and A. Lupsasca, Class. Quant. Grav. **37**, 015006 (2020).
47. S. E. Gralla and A. Lupsasca, Phys. Rev. D **101**, 044031 (2020).
48. O. James, E. von Tunzelmann, P. Franklin and K. S. Thorne, Class. Quant. Grav. **32**, 065001 (2015).
49. P. Cunha, V.P., C. A. R. Herdeiro and E. Radu, Universe **5**, 220 (2019).
50. V. Bozza and L. Mancini, Gen. Rel. Grav. **36**, 435 (2004).
51. Thomas Muller, Phys. Rev. D **77**, 124042 (2008).
52. S. U. Islam and S. G. Ghosh, [arXiv:2102.08289 [gr-qc]].
53. S. U. Islam, R. Kumar and S. G. Ghosh, JCAP **09**, 030 (2020); R. Kumar, S. U. Islam and S. G. Ghosh, Eur. Phys. J. C **80**, 1128 (2020).
54. V. Bozza, Phys. Rev. D **78**, 103005 (2008).
55. H. C. Ohanian, Am. J. Phys. 55 (May, 1987).
56. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, New York, 1992).
57. B. Carter, Phys. Rev. D **174**, 1559 (1968).
58. V. P. Frolov and A. Zelnikov, *Introduction to Black Hole Physics* (Oxford University Press, New York, 2011).

Killing tensors in foliated spacetimes and photon surfaces

Igor Bogush*, Kirill Kobialko† and Dmitri Gal'tsov‡
Faculty of Physics, Moscow State University, 119899, Moscow, Russia
**E-mail: igbogush@gmail.com*
†*E-mail: kobyalkokv@yandex.ru*
‡*E-mail: galtsov@phys.msu.ru*

We discuss a recently proposed geometric method[1] for constructing a nontrivial Killing tensor of rank two in a foliated spacetime of codimension one that lifts trivial Killing tensors from slices to the entire manifold. The existence of nontrivial Killing tensor is closely related to generalized photon surfaces. The method is illustrated on some known cases and used to construct the hitherto unknown Killing tensor for the Nutty dyon in dilaton-axion gravity.

Keywords: Killing tensors, photon surfaces, gravitational lensing, black hole shadows.

1. Introduction

Killing tensors of the second rank express hidden symmetries of spacetime[2,3] providing integrals of motion for geodesics and wave operators in field theories. Some constructive procedures to find them were suggested for spaces with a warped/twisted product structure,[4] spaces admitting a hypersurface orthogonal Killing vector field,[5,6] or special conformal Killing vector fields.[7] Recently, some deformed Kerr metrics attracted attention when trying to find new physics in ultracompact astrophysical objects. Classes of such metrics were listed that admit Killing tensors,[8,9] but for others the separation of variables in geodesic equations is not guaranteed.

The new procedure[1] suggested for foliated spacetime of codimension one is based on lifting the trivial Killing tensors in slices with an arbitrary second fundamental form.[1] This remove some particular assumptions about slices, so, hopefully, we can move further in the study of separability. The method is closely related to formalism of fundamental photon surfaces introduced in.[10] This generalizes one previous observation[11] on the relationship between spacetime separability and spherical photon orbits. It is also worth noting that our method does not require sovling diffenerial equations at all.

As illustrations, we apply this new technique to some conventional metrics of Petrov type D demonstrating that this technique allows to obtain known Killing tensors[12,13] purely algebraically, without solving any differential equations. Then we successfully apply new technique to find Killing tensor for Gal'tsov-Kechkin solution[14,15] in dilaton-axion gravity which belongs to the general Petrov type I. The new method reveals the nature of Killing tensors as arising from isometries of low-dimensional slices of a smooth foliation.

In Sec. 2 we briefly describe the equations for the Killing vectors and Killing tensors of rank two. In Sec. 3 we consider spacetimes with foliation of codimension one and present the equations governing the interplay between symmetries in the bulk and in the slices. Eventually we describe the Killing generating technique. In Sec. 4 we reveal the connection between the Killing tensors and the fundamental photon surfaces. Sec. 5 provides examples with axial symmetry in different models.

2. Conventions

Let M be a Lorentzian manifold of dimension m with scalar product $\langle\,\cdot\,,\,\cdot\,\rangle$ and Levi-Civita connection ∇[a].

Definition 2.1. A vector field $\mathcal{K}: M \to TM$ is called a Killing vector field if[16]

$$\underset{\mathcal{X}\leftrightarrow\mathcal{Y}}{\mathrm{Sym}}\,\{\langle\nabla_{\mathcal{X}}\mathcal{K},\mathcal{Y}\rangle\} = 0, \quad \forall\mathcal{X},\mathcal{Y}\in TM. \tag{1}$$

Definition 2.2. A linear self-adjoint mapping $K(\,\cdot\,): TM \to TM$ is called a Killing mapping if

$$\underset{\mathcal{X}\leftrightarrow\mathcal{Y}\leftrightarrow\mathcal{Z}}{\mathrm{Sym}}\,\{\langle\nabla_{\mathcal{X}}K(\mathcal{Z}),\mathcal{Y}\rangle\} = 0, \quad \forall\mathcal{X},\mathcal{Y},\mathcal{Z}\in TM, \tag{2}$$

where the linear mapping $\nabla_{\mathcal{X}}K(\,\cdot\,): TM \to TM$ is defined as follows

$$\nabla_{\mathcal{X}}K(\mathcal{Y}) \equiv \nabla_{\mathcal{X}}(K(\mathcal{Y})) - K(\nabla_{\mathcal{X}}\mathcal{Y}), \quad \forall\mathcal{X},\mathcal{Y}\in TM. \tag{3}$$

One can introduce a Killing tensor as a symmetric form $K(\mathcal{X},\mathcal{Y}) = \langle K(\mathcal{X}),\mathcal{Y}\rangle$, which is associated with the conservation law quadratic in momenta. Let $\mathcal{K}_\alpha$ be a set of n Killing vector fields. Then, one can define the following trivial Killing mapping

$$K(\mathcal{X}) = \alpha\mathcal{X} + \sum_{\alpha,\beta=1}^{n} \gamma^{\alpha\beta}\,\langle\mathcal{X},\mathcal{K}_\alpha\rangle\,\mathcal{K}_\beta, \quad \gamma^{\alpha\beta} = \gamma^{\beta\alpha}, \tag{4}$$

where α and $\gamma^{\alpha\beta}$ is the set of $n(n+1)/2+1$ independent constants in M. Note that the trivial Killing mapping does not give new conservation laws. However, one can show the existence of manifolds with nontrivial Killing tensors, which are not associated with the manifold isometries directly.

3. Generation of a non-trivial Killing tensor

Consider a timelike/spacelike *foliation* of the manifold M by a smooth family of hypersurfaces S_Ω parameterized by $\Omega\in\mathbb{R}$ (slices) with the lapse function φ, and

[a]Here we also use the notation $\underset{\mathcal{X}\leftrightarrow\mathcal{Y}}{\mathrm{Sym}}\,\{B(\mathcal{X},\mathcal{Y})\} \equiv B(\mathcal{X},\mathcal{Y}) + B(\mathcal{Y},\mathcal{X})$.

vector field ξ normal to slices ($\langle\xi,\xi\rangle \equiv \epsilon = \pm 1$). Then, the second fundamental form ${}^{\Omega}\sigma(\,\cdot\,,\,\cdot\,) : TS\times TS \to \mathbb{R}$ and the mean curvature of slices S_Ω are defined as follows

$$ {}^{\Omega}\sigma(X,Y) \equiv \epsilon\,\langle \nabla_X Y, \xi\rangle\,, \quad \forall X,Y \in TS, \qquad H \equiv \mathrm{Tr}({}^{\Omega}\sigma)/(m-1). \tag{5} $$

In particular, one can decompose the Killing vector field into the sum $\mathcal{K} = \mathcal{K}_\Omega + k_N\xi$, with the normal $k_N\xi$ and tangent $\mathcal{K}_\Omega \in TS_\Omega$ components. In the general case, the projection $\mathcal{K}_\Omega$ is not a Killing vector in the slices of foliation.[1,10] An exception is the case of totally umbilic or totally geodesic slice, where the projection of any Killing field is a conformal or ordinary Killing vector field respectively. Such slices arise if the field generating the foliation is a (conformal) Killing field and/or the spacetime has the structure of a warped/twisted product.[16] Therefore, the generation of the Killing vectors in M from Killing vectors $\mathcal{K}_\Omega$ with a nontrivial normal component k_N is possible in the case of the totally geodesic slices only. As we will see further, the case of Killing tensors is more intricate.

Similarly, the Killing mapping can be split to normal and tangent components $K(\,\cdot\,) = K_\Omega^{(\,\cdot\,)} + k_N^{(\,\cdot\,)}\xi$, where $k_N^{(\,\cdot\,)}\xi$ is a normal component and $K_\Omega^{(\,\cdot\,)} \in TS_\Omega$ is a tangent component. In the case of totally geodesic slices, one can lift the Killing tensor from the slice to the whole manifold and obtain a nontrivial normal component $k_N^{(\,\cdot\,)}$. This particular case of totally geodesic slices was considered, for example, in the Ref.[5]. Moreover, if we consider the conformal Killing tensors, a similar technique can be applied in the warped spacetimes,[4] where the foliation slices are totally umbilic.[16] In this paper, we consider the Killing tensor lift technique for arbitrary slices (not totally geodesic submanifolds). In this case, the second fundamental form ${}^{\Omega}\sigma$ is not trivial, and Killing equations imply $k_N^X = 0$, $K_\Omega^\xi = 0$. Then, the family of Killing mappings $K_\Omega : TS_\Omega \to TS_\Omega$ can be lifted from the slices to the Killing mapping $K(\,\cdot\,) = K_\Omega^{(\,\cdot\,)} + k_N^{(\,\cdot\,)}\xi$ in the manifold M with nontrivial normal components, if the following equations hold[1,10]

$$ k_N^X = 0, \quad K_\Omega^\xi = 0, \quad \xi(k_N^\xi) = 0, \quad X(k_N^\xi) = 2k_N^\xi \nabla_X \ln\varphi - 2\nabla_{K_\Omega^X}\ln\varphi, \tag{6a} $$

$$ \underset{X\leftrightarrow Y}{\mathrm{Sym}} \left\{ \frac{1}{2}\cdot\left\langle \nabla_\xi K_\Omega^X, Y\right\rangle + \epsilon\cdot {}^{\Omega}\sigma(X,K_\Omega^Y) - \epsilon k_N^\xi \cdot {}^{\Omega}\sigma(X,Y)\right\} = 0. \tag{6b} $$

Suppose that the manifold M has a collection $n \le m-2$ of linearly independent Killing vector fields $\mathcal{K}_\alpha$ tangent to the slices S_Ω of the foliation $\mathcal{F}_\Omega$. Then, such vectors $\mathcal{K}_\alpha$ are also Killing vectors in the slices S_Ω, and a trivial Killing mapping of the form (4) is always defined. Substituting this mapping into equations (6a), (6b) we obtain

$$ X(k_N^\xi) = 2(k_N^\xi - \alpha)\nabla_X \ln\varphi, \quad \xi(k_N^\xi) = 0, \quad X(\alpha) = 0, \quad X(\gamma^{\alpha\beta}) = 0, \tag{7a} $$

$$ 2\epsilon(k_N^\xi - \alpha)\cdot {}^{\Omega}\sigma(X,Y) = \xi(\alpha)\,\langle X,Y\rangle + \sum_{\alpha,\beta=1}^{n} \xi(\gamma^{\alpha\beta})\,\langle X,\mathcal{K}_\alpha\rangle\,\langle \mathcal{K}_\beta, Y\rangle\,, \tag{7b} $$

for any $X, Y \in TS_\Omega$. There is always a trivial solution for these equations

$$k_N^\xi = \alpha, \quad \xi(\alpha) = 0, \quad \xi(\gamma^{\alpha\beta}) = 0, \tag{8}$$

corresponding to the trivial Killing tensor in M. However, in some cases it can also have nontrivial solutions, which corresponds to the nontrivial Killing tensor and new conservation laws. Let us additionally assume that the Gramian matrix $\mathcal{G}_{\alpha\beta} = \langle \mathcal{K}_\alpha, \mathcal{K}_\beta \rangle$ is not degenerate ($\mathcal{G} \equiv \det(\mathcal{G}_{\alpha\beta}) \neq 0$). Then, we can introduce a basis $\{\mathcal{K}_\alpha, e_a\}$ in S_Ω in such a way that $e_a \in \{\mathcal{K}_\alpha\}^\perp$ with $a = 1, \ldots, m-n-1$. A non-trivial Killing tensor can be generated using the technique from the following theorem[1]:

Theorem 3.1. *Let the manifold M contains a collection of $n \leq m-2$ Killing vector fields $\mathcal{K}_\alpha$ with a non-degenerate Gramian $\mathcal{G}_{\alpha\beta} = \langle \mathcal{K}_\alpha, \mathcal{K}_\beta \rangle$, tangent to the foliation slices S_Ω (partially umbilic if $n < m-2$) with the second fundamental form*[b]

$${}^\Omega\sigma = \begin{pmatrix} -\frac{1}{2}\epsilon \cdot \xi\mathcal{G}_{\alpha\beta} & 0 \\ 0 & h^\Omega \cdot \langle e_a, e_b \rangle \end{pmatrix} \tag{9}$$

Then, there is a nontrivial Killing tensor on manifold M, if the following steps can be successfully completed:

***Step one**: Check compatibility and integrability conditions (10), (11)*

$$X(h^\Omega \cdot \varphi^3) = 0, \tag{10}$$

$$X\left(\mathcal{G}^{\alpha\beta} - \frac{\epsilon}{2h^\Omega} \cdot \xi\mathcal{G}^{\alpha\beta}\right) = 0. \tag{11}$$

***Step two**: Obtain α from (12) and check the condition (13)*

$$\xi \ln \xi(\alpha) = \xi \ln h^\Omega - 2\epsilon h^\Omega, \tag{12}$$

$$X(\alpha) = 0. \tag{13}$$

***Step three**: Define $\gamma^{\alpha\beta}$ from (14) using the conditions $\xi\nu^{\alpha\beta} = 0$, $X\gamma^{\alpha\beta} = 0$.*

$$\gamma^{\alpha\beta} = \epsilon\frac{\xi(\alpha)}{2h^\Omega} \cdot \mathcal{G}^{\alpha\beta} - \nu^{\alpha\beta}, \tag{14}$$

***Step four**: Using the functions found in the previous steps, construct a Killing map and the corresponding Killing tensor:*

$$K(\mathcal{X}) = \alpha\mathcal{X} + \sum_{\alpha,\beta=1}^{n} \gamma^{\alpha\beta} \langle \mathcal{X}, \mathcal{K}_\alpha \rangle \mathcal{K}_\beta + \frac{\xi(\alpha)}{2h^\Omega} \langle \mathcal{X}, \xi \rangle \xi. \tag{15}$$

[b]The form of the left upper block is a consequence of the tangent Killing vectors, and this does not impose a new condition. The non-diagonal zero elements is a new condition, which is satisfied in many applications. The right lower block is a condition of partially umbilical surfaces, which imposes constraints if $\dim\{e_a\} > 1$.

4. Connection with photon submanifolds

Consider the case of a manifold with two Killing vectors spanning a timelike surface ($\mathcal{G} < 0$). Let us define a Killing vector field ρ^α with index $\alpha = 1, 2$ numbering Killing vectors of the basis $\{\mathcal{K}_\alpha\}$, which is supposed to have constant components $\rho^\alpha = (\rho, 1)$. The quantity ρ can be called the generalized impact parameter (see[10] for details). However, one can choose arbitrary parametrization of ρ^α up to the norm.

Proposition 4.1. *The fundamental photon surface is a partially umbilic surface with a second fundamental form of the form (9), with the following connection between h^Ω and $\mathcal{G}_{\alpha\beta}$[10]*

$$\rho^\alpha \mathcal{M}_{\alpha\beta}\rho^\beta = 0, \qquad \mathcal{M}_{\alpha\beta} \equiv \frac{1}{2h^\Omega}\cdot\xi\mathcal{G}_{\alpha\beta} - \mathcal{G}_{\alpha\beta} - \frac{1}{2h^\Omega}\cdot\xi\ln\mathcal{G}\cdot\mathcal{G}_{\alpha\beta}. \tag{16}$$

If the surface under consideration is totally umbilic $\mathcal{M}_{\alpha\beta} = 0$, it is obviously a fundamental photon surfaces for any ρ. Since totally umbilic surfaces usually exist in spherically symmetric solutions (both static and non-static) or non-rotating solutions with NUT-charge,[17] and they have been considered in detail in a number of works,[18,19] we will focus on the case $\mathcal{M}_{\alpha\beta} \neq 0$.

Consider the foliation generating a nontrivial Killing tensor in accordance with Theorem 3.1, and ask the question whether its slice S_Ω is a fundamental photon surface. First of all, we need to solve the quadratic equation (16) for ρ and check the condition $\rho^\alpha\mathcal{G}_{\alpha\beta}\rho^\beta \geq 0$. It has nontrivial solution if the eigenvalues of the matrix $\mathcal{M}_{\alpha\beta}$ have different signs, that is $\mathcal{M} \equiv \det(\mathcal{M}_{\alpha\beta}) < 0$. Then the solution for ρ reads as

$$\rho = \frac{-\mathcal{M}_{12} \pm \sqrt{-\mathcal{M}}}{\mathcal{M}_{11}}. \tag{17}$$

Condition $\rho^\alpha\mathcal{G}_{\alpha\beta}\rho^\beta \geq 0$ is satisfied if the following inequality holds

$$\pm 2(\mathcal{G}_{12}\mathcal{M}_{11} - \mathcal{G}_{11}\mathcal{M}_{12})\sqrt{-\mathcal{M}} - 2\mathcal{G}_{11}\cdot\mathcal{M} + \mathcal{M}_{11}\cdot\mathcal{G}\cdot\mathrm{Tr}(\mathcal{M}) \geq 0, \tag{18}$$

where $\mathrm{Tr}(\mathcal{M}) = \mathcal{M}_{\alpha\beta}\mathcal{G}^{\alpha\beta} = -2-(2h^\Omega)^{-1}\cdot\xi\ln\mathcal{G}$. Equation (18) defines the so-called photon region,[20,21] which arises as a flow of fundamental photon surfaces.[10] However, it has not been proven that the expression (17) for ρ is constant in every slice. But the integrability condition (11) guarantees that it is true in fact.[1] Therefore, we have the following theorem.

Theorem 4.1. *Let S_Ω be a non totally umbilic foliation slice with compact spatial section satisfying all conditions of the theorem 3.1 for $dim\{\mathcal{K}_\alpha\} = 2$. Then maximal subdomain $U_{PS} \subseteq S_\Omega$ such that the inequality (18) holds for all $p \in U_{PS}$ is a fundamental photon surface*[c].

[c] In the case of not compact spatial section, the slice is not a fundamental photon surface by definition.[10] However, the theorem can be generalized for such not compact surfaces.

In particular, the region $U_{PR} \subseteq M$, such that the inequality (18) holds for any point $p \in U_{PR}$, is a photon region. This theorem generalizes the connection between the existence of Killing tensors of this type and photon surfaces or spherical null geodesics, which was noted in Refs.[11,19,22]. Unfortunately, in the opposite direction, the theorem is not fair, since the existence of fundamental photon surfaces does not guarantee the existence of the Killing tensor. As a counterexample, one can suggest Zipoy-Voorhees metric[23] where the fundamental photon surfaces exists[24] but there is no nontrivial Killing tensor.[25] Nevertheless, the existence of fundamental photon surfaces can serve as a sign that the Killing tensor can be presented in the corresponding metric, and it is advisable to check the conditions of consistency and integrability.

5. Axially symmetric spacetimes

Consider a Lorentzian manifold M with the metric tensor

$$ds^2 = -f(dt - \omega d\phi)^2 + \lambda dr^2 + \beta d\theta^2 + \gamma d\phi^2. \tag{19}$$

where all metric components depend on r and θ only and the foliation with timelike slices $r = \Omega$. Generally, this metric possesses two Killing vectors $\mathcal{K}_1 = \partial_t$, $\mathcal{K}_2 = \partial_\varphi$. One can find that the second fundamental form of these slices has the form (9) and other quantities are

$$\xi = \lambda^{-1/2}\partial_r, \quad h^\Omega = -\frac{1}{2}\lambda^{-1/2} \cdot \partial_r \ln\beta, \quad \varphi = \lambda^{1/2}, \quad \mathcal{G}^{\alpha\beta} = \frac{1}{\gamma}\begin{pmatrix} \omega^2 - \gamma f^{-1} & \omega \\ \omega & 1 \end{pmatrix}. \tag{20}$$

In this case, the number of Killing vector is one less than the slices dimension, so the boundary $n \geq m - 2$ saturates and the partially umbilic condition just imposes a relation on h^Ω. The compatibility and integrability conditions (10), (11) take the form

$$\partial_\theta(\lambda \cdot \partial_r \ln\beta) = 0, \qquad \partial_\theta\left(\mathcal{G}^{\alpha\beta} + \frac{1}{\partial_r \ln\beta}\partial_r\mathcal{G}^{\alpha\beta}\right) = 0. \tag{21}$$

The Eq. (12) can be solved as follows

$$\alpha = A_\theta \cdot \beta + B_\theta, \tag{22}$$

where the arbitrary functions A_θ, B_θ depend on θ only, obeying the condition $\partial_\theta\alpha = 0$. As the result, we have one more necessary condition for the case in this section: the function β must be of the form

$$\beta(r,\theta) = \beta_1(\theta)\beta_2(r) + \beta_3(\theta), \tag{23}$$

where $\beta_{1,2,3}$ are some functions of the corresponding variables. In particular, the normal component is $k_N^\xi = B_\theta$. Next, we can define the matrix γ:

$$\gamma^{\alpha\beta} = -\beta A_\theta \cdot \mathcal{G}^{\alpha\beta} - \nu^{\alpha\beta}. \tag{24}$$

The integrability condition guarantees that $\gamma^{\alpha\beta}$ always satisfies the equations (7) for some $\nu^{\alpha\beta}$ depending only on θ. On the other hand, we have to find a $\nu^{\alpha\beta}$ that makes the equation $\partial_r \gamma^{\alpha\beta} = 0$ true. Therefore, we can omit the θ-dependent part in $\gamma^{\alpha\beta}$ to some constant matrix instead of looking for $\nu^{\alpha\beta}$. Combining everything together, we get the final Killing tensor in the holonomic basis

$$K^{\mu\nu} = \alpha g^{\mu\nu} + \sum_{\alpha,\beta=t,\phi} \gamma^{\alpha\beta} \mathcal{K}^{\mu}_{\alpha} \mathcal{K}^{\nu}_{\beta} - \beta A_\theta \lambda^{-1} \delta^{\mu}_{r} \delta^{\nu}_{r}. \tag{25}$$

The compatibility and integrability conditions, as well as the condition on the function β, are invariant under the multiplicative transformations of the form

$$\lambda \to \lambda' = u(r)\lambda, \quad \beta \to \beta' = v(\theta)\beta. \tag{26}$$

If β possesses the aforementioned form (23), one can simplify the integrability condition by the substitution $\mathcal{G}^{\alpha\beta} = \tilde{\mathcal{G}}^{\alpha\beta} \cdot \beta_1/\beta$. Then, the integrability condition is $\partial_\theta \partial_r \tilde{\mathcal{G}}^{\alpha\beta} = 0$, which is solved by $\tilde{\mathcal{G}}^{\alpha\beta} = \tilde{\mathcal{G}}^{\alpha\beta}_r(r) + \tilde{\mathcal{G}}^{\alpha\beta}_\theta(\theta)$. This generalizes the result of Ref.[26], where the similar condition was obtained from the separability of the Hamilton-Jacobi equation. In our case, we have also included the $\beta_1(\theta)$ term. Furthermore, the compatibility condition and the function form (23) leads to the form of $\lambda = \lambda_r(r)\beta/\beta_1$, where λ_r is an arbitrary function of r.

5.1. *Kerr metric*

As a simple illustration, consider Kerr solution in the Boyer-Lindquist coordinates:

$$ds^2 = -\frac{\Delta - a^2 \sin^2\theta}{\Sigma}(dt - \omega d\phi)^2 + \Sigma\left(\frac{dr^2}{\Delta} + d\theta^2 + \frac{\Delta \sin^2\theta}{\Delta - a^2\sin^2\theta} d\phi^2\right), \tag{27}$$

$$\Sigma = r^2 + a^2\cos^2\theta, \qquad \omega = \frac{-2Mar\sin^2\theta}{\Delta - a^2\sin^2\theta}, \tag{28a}$$

$$\Delta = r(r-2M) + a^2. \tag{28b}$$

In the Kerr metric, $\beta = r^2 + a^2\cos^2\theta$, $\lambda = \beta/\Delta$, satisfy the compatibility condition. One can explicitly verify that $\mathcal{G}^{\alpha\beta}$ satisfies the integrability equation. In this case $\alpha = r^2$, $A_\theta = 1$ and $k^{\xi}_{N} = B_\theta = -a^2\cos^2\theta$ (here we have fixed the multiplicative integration constant, which appears due to the linearity of Killing equations). The part of $\gamma^{\alpha\beta}$ independent on θ reads

$$\gamma^{\alpha\beta} = \Delta^{-1} \begin{pmatrix} (a^2+r^2)^2 & a(a^2+r^2) \\ a(a^2+r^2) & a^2 \end{pmatrix}. \tag{29}$$

Finally, we get α and $\gamma^{\alpha\beta}$, which correspond to the well-known nontrivial Killing tensor for Kerr solution

$$\begin{aligned} K^{\mu\nu} &= r^2 g^{\mu\nu} + \Delta^{-1} S^\mu S^\nu - \Delta \delta^{\mu}_{r} \delta^{\nu}_{r}, \\ S^\mu &= s\delta^{\mu}_{t} + a\delta^{\mu}_{\varphi}, \qquad s = r^2 + a^2. \end{aligned} \tag{30}$$

5.2. *Plebanski-Demianski solution*

This is a less trivial example of the type D solution of Einstein-Maxwell equations with the cosmological constant, which contains also an acceleration parameter. The metric line element ds^2 is more conveniently presented in the conformally related frame using the Boyer-Lindquist coordinates:

$$\Omega^2 ds^2 = \Sigma\left(\frac{dr^2}{\Delta_r}+\frac{d\theta^2}{\Delta_\theta}\right)+\frac{1}{\Sigma}\left((\Sigma+a\chi)^2\Delta_\theta\sin^2\theta-\Delta_r\chi^2\right)d\phi^2 \tag{31}$$

$$+\frac{2}{\Sigma}\left(\Delta_r\chi-a(\Sigma+a\chi)\Delta_\theta\sin^2\theta\right)dtd\phi-\frac{1}{\Sigma}\left(\Delta_r-a^2\Delta_\theta\sin^2\theta\right)dt^2, \tag{32}$$

where we have defined the functions

$$\Delta_\theta = 1-a_1\cos\theta-a_2\cos^2\theta, \quad \Delta_r = b_0+b_1r+b_2r^2+b_3r^3+b_4r^4, \tag{33}$$

$$\Omega = 1-\lambda(N+a\cos\theta)r, \quad \Sigma = r^2+(N+a\cos\theta)^2, \quad \chi = a\sin^2\theta-2N(\cos\theta+C), \tag{34}$$

with the following constant coefficients in Δ_θ and Δ_r:

$$a_1 = 2aM\lambda-4aN\left(\lambda^2(k+\beta)+\frac{\Lambda}{3}\right), \quad a_2 = -a^2\left(\lambda^2(k+\beta)+\frac{\Lambda}{3}\right), \quad b_0 = k+\beta, \tag{35}$$

$$b_1 = -2M, \quad b_2 = \frac{k}{a^2-N^2}+4MN\lambda-(a^2+3N^2)\left(\lambda^2(k+\beta)+\frac{\Lambda}{3}\right), \tag{36}$$

$$b_3 = -2\lambda\left(\frac{kN}{a^2-N^2}-(a^2-N^2)\left(M\lambda-N\left(\lambda^2(k+\beta)+\frac{\Lambda}{3}\right)\right)\right), \tag{37}$$

$$b_4 = -\left(\lambda^2k+\frac{\Lambda}{3}\right), \tag{38}$$

$$k = \frac{1+2MN\lambda-3N^2\left(\lambda^2\beta+\frac{\Lambda}{3}\right)}{1+3\lambda^2N^2(a^2-N^2)}(a^2-N^2), \quad \lambda = \frac{\alpha}{\omega}, \quad \omega = \sqrt{a^2+N^2}\,. \tag{39}$$

Generally, the coordinates t and r range over the whole $\mathbb{R}$, while θ and ϕ are the standard coordinates on the unit two-sphere. Seven independent parameters $M, N, a, \alpha, \beta, \Lambda, C$ can be physically interpreted as mass, NUT parameter (magnetic mass), rotation parameter, acceleration parameter, $\beta = e^2+g^2$ comprises the electric e and magnetic g charges, Λ is the cosmological constant, and the constant C defines the location of the Misner string.

The first step is to check the compatibility and integrability conditions (10), (11). The first one holds if $\alpha \cdot a = 0$, i.e. either the acceleration α or the rotation a are zero. Indeed, as shown in Ref.[12], the general PD solution with acceleration possesses a conformal Killing tensor, but not the usual one. In the case $a = 0$, the second condition does not hold. So, further we will consider the solution with zero acceleration $\alpha = 0$, which corresponds to the dyonic Kerr-Newman-NUT-AdS solution.

In the **second step** we pick up the r-dependent part from $\beta = \Sigma/\Delta_\theta$ for α

$$\beta = \frac{\Sigma}{\Delta_\theta} \quad \Rightarrow \quad \alpha = r^2, \quad A_\theta = \Delta_\theta. \tag{40}$$

Similarly, as the **third step**, the r-dependent part for $\gamma^{\alpha\beta}$ is defined as

$$\gamma^{\alpha\beta} = \Delta_r^{-1} \begin{pmatrix} s^2 & as \\ as & a^2 \end{pmatrix}, \quad s = \Sigma + a\chi = r^2 + a^2 - 2aCN + N^2. \tag{41}$$

In the last **fourth step**, we obtain the nontrivial Killing tensor for the Kerr-Newman-NUT-AdS metric:

$$K^{\mu\nu} = r^2 g^{\mu\nu} + \Delta_r^{-1} S^\mu S^\nu - \Delta_r \delta^\mu_r \delta^\nu_r, \quad S^\mu = s\delta^\mu_t + a\delta^\mu_\varphi. \tag{42}$$

6. Gal'tsov-Kechkin (GK) solution

In 1994 one of the present authors in collaboration with O. Kechkin derived the general stationary charged black hole solution within the Einstein-Maxwell- dilaton-axion (EMDA) gravity, which is the $\mathcal{N} = 4, D = 4$ supergravity consistently truncated to the theory with one vector field.[14] This solution was seven-parametric, containing mass M, electric and magnetic charges Q, P, rotation parameter a, NUT N and asymptotic values of the dilaton and axion (irrelevant for the metric) as independent parameters. Less general (without NUT) solution was derived by A. Sen[27] in the context of the dimensionally reduced effective action of the heterotic string, and now it is commonly referred as Kerr-Sen metric. Non-rotating solutions with NUT were independently obtained by Kallosh et al.[28] and Johnson and Myers.[29] Now the Kerr-Sen metric is often considered as a deformed Kerr solution in modeling deviations from the standard picture of black holes.

The GK solution[14] was shown[15] to belong to type Petrov I, contrary to Kerr-Newman-NUT solution in the Einstein-Maxwell gravity. The same was shown for Kerr-Sen solution without NUT.[30] Though the Kerr-Sen metric is not type D, the Hamilton-Jacoby equation was shown to be separable for it,[31,32] though the Killing tensor was not found explicitly (for further discussion see[9]). But for the solution with NUT, it was claimed that no second rank Killing tensor exists.[33] So here we use the new technique to resolve this controversy.

The line element of the GK solution can be written in the form (27) where the functions Δ, ω, Σ are redefined as follows

$$\begin{aligned}
&\Delta = (r - r_-)(r - 2M) + a^2 - (N - N_-)^2, \\
&\Sigma = r(r - r_-) + (a\cos\theta + N)^2 - N_-^2, \\
&\omega = \frac{-2w}{\Delta - a^2 \sin^2\theta}, \quad w = N\Delta\cos\theta + a\sin^2\theta\,(M(r - r_-) + N(N - N_-))\,.
\end{aligned}$$

The full solution also contains Maxwell, dilaton ϕ and axion κ fields (whose form can be found in[14]), and represents a family with seven parameters: mass M,

NUT charge N, rotation parameter a, electric and magnetic charges Q and P, and asymptotic complex axidilaton charge $z_\infty = \kappa_\infty + i\mathrm{e}^{-2\phi_\infty}$, irrelevant for the metric. The following abbreviations are used:

$$r_- = \frac{M|\mathcal{Q}|^2}{|\mathcal{M}|^2}, \quad N_- = \frac{N|\mathcal{Q}|^2}{2|\mathcal{M}|^2}, \quad \mathcal{M} = M + iN, \quad \mathcal{Q} = Q - iP, \quad \mathcal{D} = -\frac{\mathcal{Q}^{*2}}{2\mathcal{M}}. \tag{43}$$

The metric is presented in the Kerr-like form, but the metric functions are essentially different. The solution reduces to Kerr-NUT for $Q = P = 0$.

Now we apply our procedure to construct the Killing tensor. It can be easily verified that the consistency (10) and the integrability (11) conditions are satisfied. The uplift turns out to be as simple as in the vacuum Kerr example, leading to the result

$$\begin{aligned} &K^{\mu\nu} = r(r - r_-)g^{\mu\nu} + \Delta^{-1}S^\mu S^\nu - \Delta\delta^\mu_r\delta^\nu_r, \\ &S^\mu = s\delta^\mu_t + a\delta^\mu_\varphi, \qquad s = r(r - r_-) + a^2 + N^2 - N_-^2. \end{aligned} \tag{44}$$

This expression is new and is applicable both to the Kerr-Sen solution $N = N_- = 0$, and for the full GK solution.

7. Conclusions

In this article, we reviewed new geometric method of generating Killing tensor in spacetimes with codimension one foliation.[1] Using our general lift equations one can try to rise a trivial Killing tensor defined in the slices into a nontrivial Killing tensor in the bulk. For this, the foliation must satisfy some consistency and integrability conditions, which we have presented explicitly. Furthermore, we have completely solved the lifting equations in terms of the functions α, $\gamma^{\alpha\beta}$ and formulated the theorem (3.1) for generation of the non-trivial Killing tensor.

Finding a foliation compatible with integrability and consistency conditions can be a challenging task. The existence of such a foliation means that the slices represent fundamental photon surfaces provided the corresponding inequalities hold. This generalizes the result of Ref.[19] to the case of general stationary spaces. Conversely, the existence of fundamental photon surfaces, though does not guarantee the existence of the Killing tensor, but may serve as an indication that such tensor may exist. Therefore, it is recommended to check the consistency and integrability conditions for fundamental photon surfaces, if these are known. This property makes the search for fundamental photon surfaces important also for studying the integrability of geodesic motion. It is tempting to conjecture that the existence of fundamental photon surfaces implies the existence of Killing tensor if the slice is equipotential or spherical.[34]

Using this technique, we were able to derive Killing tensor for EMDA dyon with NUT (GK solution[14]), which is also valid for Kerr-Sen solution as a particular case.

Acknowledgements

The work was supported by the Russian Foundation for Basic Research on the project 20-52-18012Bulg-a, and the Scientific and Educational School of Moscow State University "Fundamental and Applied Space Research". I.B. is also grateful to the Foundation for the Advancement of Theoretical Physics and Mathematics "BASIS" for support.

References

1. K. Kobialko, I. Bogush and D. Gal'tsov, Killing tensors and photon surfaces in foliated spacetimes, *Phys. Rev. D* **104**, p. 044009 (2021).
2. B. Carter, Killing Tensor Quantum Numbers and Conserved Currents in Curved Space, *Phys. Rev. D* **16**, p. 3395 (1977).
3. V. Frolov, P. Krtous and D. Kubiznak, Black holes, hidden symmetries, and complete integrability, *Living Rev. Rel.* **20**, p. 6 (2017).
4. P. Krtous, D. Kubiznak and I. Kolar, Killing-Yano forms and Killing tensors on a warped space, *Phys. Rev. D* **93**, p. 024057 (2016).
5. D. Garfinkle and E. N. Glass, Killing Tensors and Symmetries, *Class. Quant. Grav.* **27**, p. 095004 (2010).
6. D. Garfinkle and E. N. Glass, Killing-Yano tensors in spaces admitting a hypersurface orthogonal Killing vector, *J. Math. Phys.* **54**, p. 032501 (2013).
7. R. Rani, S. B. Edgar and A. Barnes, Killing tensors and conformal Killing tensors from conformal Killing vectors, *Class. Quant. Grav.* **20**, p. 1929 (2003).
8. Z. Carson and K. Yagi, Asymptotically flat, parameterized black hole metric preserving Kerr symmetries, *Phys. Rev. D* **101**, p. 084030 (2020).
9. G. O. Papadopoulos and K. D. Kokkotas, On Kerr black hole deformations admitting a Carter constant and an invariant criterion for the separability of the wave equation, *Gen. Rel. Grav.* **53**, p. 21 (2021).
10. K. V. Kobialko and D. V. Gal'tsov, Photon regions and umbilic conditions in stationary axisymmetric spacetimes: Photon Regions, *Eur. Phys. J. C* **80**, p. 527 (2020).
11. G. Pappas and K. Glampedakis, On the connection of spacetime separability and spherical photon orbits, *arXiv*, p. 1806.04091 (2018).
12. D. Kubiznak and P. Krtous, On conformal Killing-Yano tensors for Plebanski-Demianski family of solutions, *Phys. Rev. D* **76**, p. 084036 (2007).
13. M. Vasudevan, Integrability of some charged rotating supergravity black hole solutions in four and five dimensions, *Phys. Lett. B* **624**, p. 287 (2005).
14. D. V. Galtsov and O. V. Kechkin, Ehlers-Harrison type transformations in dilaton - axion gravity, *Phys. Rev. D* **50**, p. 7394 (1994).
15. A. Garcia, D. Galtsov and O. Kechkin, Class of stationary axisymmetric solutions of the Einstein-Maxwell dilaton - axion field equations, *Phys. Rev. Lett.* **74**, p. 1276 (1995). doi:10.1103/PhysRevLett.74.1276
16. B. Y. Chen, Pseudo-Riemannian Geometry, δ-Invariants and Applications, *WorldScientific, Hackensack*, (2011).
17. D. V. Gal'tsov and K. V. Kobialko, Completing characterization of photon orbits in Kerr and Kerr-Newman metrics, *Phys. Rev. D* **99**, p. 084043 (2019)
18. C. M. Claudel, K. S. Virbhadra and G. F. R. Ellis, The Geometry of photon surfaces, *J. Math. Phys.* **42**, p. 818-838 (2001).
19. Y. Koga, T. Igata and K. Nakashi, Photon surfaces in less symmetric spacetimes, *Phys. Rev. D* **103**, p. 044003 (2021).

20. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon Regions and Shadows of Kerr-Newman-NUT Black Holes with a Cosmological Constant, *Phys. Rev. D* **89**, p. 124004 (2014).
21. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon Regions and Shadows of Accelerated Black Holes, *Int. J. Mod. Phys. D* **24**, p. 1542024 (2015).
22. K. Glampedakis and G. Pappas, Modification of photon trapping orbits as a diagnostic of non-Kerr spacetimes, *Phys. Rev. D* **99**, p. 124041 (2019).
23. H. Kodama and W. Hikida, Global structure of the Zipoy-Voorhees-Weyl spacetime and the delta=2 Tomimatsu-Sato spacetime, *Class. Quant. Grav.* **20**, p. 5121 (2003).
24. D. V. Gal'tsov and K. V. Kobialko, Photon trapping in static axially symmetric spacetime, *Phys. Rev. D* **100**, p. 104005 (2019)
25. G. Lukes-Gerakopoulos, The non-integrability of the Zipoy-Voorhees metric, *Phys. Rev. D* **86**, p. 044013 (2012).
26. T. Johannsen, Regular Black Hole Metric with Three Constants of Motion, *Phys. Rev. D* **88**, p. 044002 (2013).
27. A. Sen, Rotating charged black hole solution in heterotic string theory, *Phys. Rev. Lett.* **69**, p. 1006 (1992).
28. R. Kallosh, D. Kastor, T. Ortin and T. Torma, Supersymmetry and stationary solutions in dilaton axion gravity, *Phys. Rev. D* **50**, p. 6374 (1994).
29. C. V. Johnson and R. C. Myers, Stringy twists of the Taub - NUT metric, *arXiv*, p. 9409177 (1994).
30. A. Burinskii, Some properties of the Kerr solution to low-energy string theory, *Phys. Rev. D* **52**, p. 5826 (1995).
31. R. A. Konoplya, Z. Stuchlík and A. Zhidenko, Axisymmetric black holes allowing for separation of variables in the Klein-Gordon and Hamilton-Jacobi equations, *Phys. Rev. D* **97**, p. 084044 (2018).
32. X. G. Lan and J. Pu, Observing the contour profile of a Kerr–Sen black hole, *Mod. Phys. Lett. A* **33**, p. 1850099 (2018).
33. H. M. Siahaan, Kerr-Sen-Taub-NUT spacetime and circular geodesics, *Eur. Phys. J. C* **80**, p. 1000 (2020).
34. C. Cederbaum and G. J. Galloway, Photon surfaces with equipotential time-slices, *J. Math. Phys.* **62**, p. 032504 (2021).

Decoding black hole metrics from the interferometric pattern of relativistic images

V. Bozza
Dipartimento di Fisica "E.R. Caianiello", Università di Salerno, Fisciano, 84084, Italy
E-mail: valboz@sa.infn.it
www.unisa.it

The image of a source around a black hole is replicated an infinite number of times by photons performing an arbitrary number of turns around the black hole. Such relativistic images generate a characteristic staircase structure in the complex visibility measured in interferometric observations. For a quasi-static source, we present analytical formulae describing the visibility function in the strong deflection limit of gravitational lensing. These formulae are then used to extract the properties of the metric from features in the visibility.

Keywords: Black hole, gravitational lensing, interferometry

1. Introduction

The sensational observation of the image of the supermassive black hole at the center of M87* by the EHT collaboration has opened the era of the investigation of black holes at high resolution.[1] Black holes typically appear as a dark spot surrounded by a light ring created by the radiation emitted by the accreting matter revolving around the black holes thanks to the strong gravitational fields.[2,3] This is possible because an unstable circular orbit for photons exists at a specific radius, depending on the black hole parameters and the inclination of the photon trajectory with respect to the equatorial plane. For this reason, any compact sources around the black hole generate two infinite sequences of images asymptotically approaching the light ring.[4–7]

Higher order images of the accretion disk should generate concentric light rings asymptotically approaching the shadow border.[8] Such structure would be detectable in very long baseline interferometry, since the complex visibility should have a staircase structure, with each step generated by an image at specific order.[9]

We propose to describe the complex visibility generated by a compact source using the tools of gravitational lensing in the strong deflection limit. With an analytical description at hand, it is then possible to track the dependence of the observables on the black hole metric and disentangle the metric parameters from the source characteristics.[10]

2. Relativistic images of a compact source

Consider a source at distance r_S from a black hole at angle ϕ_S from the optical axis (i.e. the line joining the observer to the black hole). The direct image will be formed by weakly deflected light rays and will be the largest one, thus dominating the lower angular frequencies in the visibility. Relativistic images are formed by light rays performing n turns around the black hole before reaching the observer.[4] The angular separation of these images from the black hole in the observer sky can be described in the strong deflection limit by a simple exponential formula[11]

$$\theta_{n,\pm} = \theta_m(1 + \epsilon_{n,\pm}), \tag{1}$$

where θ_m is the angular radius of the shadow,

$$\epsilon_{n,\pm} = \eta_O \eta_S e^{\frac{b \pm \phi_S - 2n\pi}{a}}, \tag{2}$$

$$\eta_O = 1 - \frac{r_m}{r_O} \qquad \eta_S = 1 - \frac{r_m}{r_S}, \tag{3}$$

r_O is the observer's distance and r_m is the radius of the unstable circular orbit for photons.

The quantities r_m, u_m, a and b are all functions of the specific black hole metric and can be calculated by standard algorithms.[11] Therefore, in principle, the observation of higher order images can be used to constrain the metric of the black hole. The tangential extension of higher order images $\Delta\vartheta_I$ is the same as the tangential extension of the source $\Delta\vartheta_S$ in the hypothesis of spherical symmetry. The radial extension is obtained by differentiation of Eq. (1)

$$\Delta\theta_{n,\pm} = \Delta\theta_0 \; \epsilon_{n,\pm} \tag{4}$$

$$\Delta\theta_0 = \frac{\theta_m}{a}\Delta\phi_S + \theta_m \frac{r_m}{r_S^2}\left(\frac{1}{\eta_S} + \frac{g_1(\eta_S)}{a}\right)\Delta r_S. \tag{5}$$

Higher order n are exponentially thinner with n. The coefficient $\Delta\theta_0$ does not depend on n, instead.

3. Complex visibility of a compact source

Interferometric observations measure the amplitude of the complex visibility, which is calculated by a Fourier transform of the image pattern.[9] If we assume a static source with a Gaussian profile, the images will be aligned along one axis in the observer sky (say the x-axis). The complex visibility expressed in the angular frequencies (u, v) conjugate to the angular coordinates (x, y) will be

$$V(u,v) = N(v) \sum_{p=\pm 1} \sum_{n=1}^{+\infty} \epsilon_{n,p} \; e^{-2\pi^2 \Delta\theta_{n,p}^2 u^2} e^{-2\pi\imath p\theta_{n,p} u}, \tag{6}$$

$$N(v) = 2\pi I_0 \theta_m \Delta\theta_0 \Delta\vartheta_S e^{-2\pi^2 \theta_m^2 \Delta\vartheta_S^2 v^2}. \tag{7}$$

The shape of the visibility in the v direction $N(v)$ does not depend on n and is the same for all contributions. In the u direction, each image contributes with a Gaussian function whose size scales with $1/\Delta\theta_n$ and thus increases exponentially with n. As the contribution of order n decays, the order $n+1$ dominates, generating the staircase structure already suggested in Ref. 9. Fig. 1 shows this structure for various source positions ϕ_S.

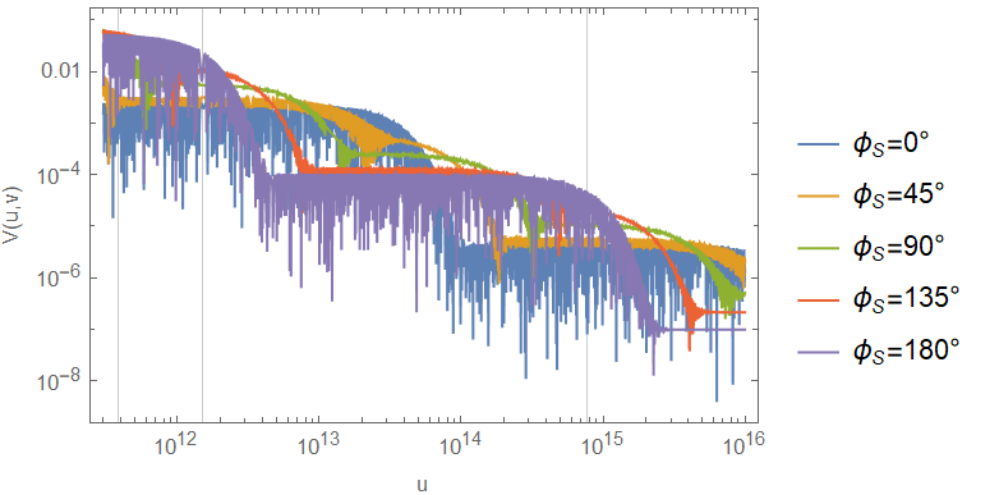


Fig. 1. Visibility function in Schwarzschild space-time for various source azimutal angles. The three vertical lines correspond to Moon baseline, Lagrangian point L2 baseline and Jupiter semi-major axis respectively at the wavelength $\lambda = 1$mm.

The direct image only contributes to low angular frequencies and rapidly decays, leaving space to higher order images at high angular frequencies. Such contributions are analytically expressed by Eq. (6), which can be used to relate the main characteristics such as amplitude, size and periodicity of each contribution to the characteristics of the black hole metric.[10]

Simple algebra allows to invert such expressions so as to write down the gravitational lensing parameters as functions of the observables.[10] The coefficient a is given by the ratio of the amplitude $h_{n,+}$ of contributions by images of consecutive orders

$$a = \frac{2\pi}{\ln\left(h_{n,+}/h_{n+1,+}\right)}. \tag{8}$$

The source angular position is obtained by comparing the amplitude of contributions by positive and negative parity images

$$\phi_S = \frac{a}{2}\ln\left(h_{n,+}/h_{n,-}\right). \tag{9}$$

The angular radius of the shadow and the coefficient b are derived from the frequencies $\nu_{n,\pm}$ of the periodicities in the visibility at scales when two contributions interfere

$$\theta_m = \frac{1}{4}\left\{e^{\frac{2\pi}{a}}\left[\nu_{n,-} + e^{-\frac{2\phi_S}{a}}\left(\nu_{n,-} - \nu_{n,+}\right)\right] - \nu_{n,+}\right\}\left(\coth\frac{\pi}{a} - 1\right), \tag{10}$$

$$\eta_O \eta_S e^{\frac{b-2n\pi}{a}} = \frac{2(\nu_{n,+} - \nu_{n,-})}{\nu_{n,-}\left(e^{\frac{\phi_S}{a}} + e^{-\frac{\phi_S}{a}}\right) - \nu_{n,+}\left(e^{\frac{\phi_S - 2\pi}{a}} + e^{-\frac{\phi_S}{a}}\right)}. \tag{11}$$

These analytical formulae allow to constrain the coefficients of gravitational lensing in the strong deflection limit, which are functions of the black hole metric. It could be possible, therefore, to distinguish between different metrics through the observation of the visibility of a compact source.[10]

4. Discussion and conclusions

As pointed before, the baselines needed for the resolution of higher order images are of the order of millions km.[9] The great scientific expectations on such observations are pushing forward projects for very long baseline interferometry in space, which could be available in the next decades.[12,13]

The physical setup analyzed in this study is very basic: a single static compact source at some distance from a spherically symmetric black hole. In this framework, the complex visibility generated by the infinite sequence of higher order images is simple enough to be described in a fully analytical form. The properties of the metric can be extracted in a clean and source-independent way, something that is quite difficult for non-compact sources or with the whole accretion flow.[14]

However, in order to apply this idealized case to real astrophysical situations, we need a powerful source dominating over the accretion flow with its higher order images. In the mm band this may happen with a very powerful flare in the accretion flow due to some local change in the magnetic fields.[15] In the Near Infrared, we may consider bright stars orbiting the black hole as compact sources.[16] Secondly, in order to have all images aligned on the same axis, we need perfect spherical symmetry and a quasi-static source. In fact, higher order images are formed by photons taking longer paths as they revolve around the black hole. Therefore, such images would follow an orbiting source with some delay.[17] This limits the applicability of this framework to sources far enough to neglect their orbital motion.

In spite of these limitations, this fully analytical calculation stands as a reference limiting case for more complicated calculations and opens the investigation to more complete physical setups.

References

1. Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, W. Alef *et al.*, First M87 Event Horizon Telescope Results. i. The Shadow of the Supermassive Black Hole, *ApJ* **875**, p. L1 (April 2019).

2. J. M. Bardeen, Timelike and null geodesics in the Kerr metric., in *Black Holes (Les Astres Occlus)*, January 1973.
3. V. Perlick and O. Y. Tsupko, Calculating black hole shadows: review of analytical studies, *arXiv e-prints*, p. arXiv:2105.07101 (May 2021).
4. C. G. Darwin, The gravity field of a particle, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **249**, 180 (1959).
5. H. C. Ohanian, The black hole as a gravitational "lens", *American Journal of Physics* **55**, 428 (May 1987).
6. K. S. Virbhadra and G. F. R. Ellis, Schwarzschild black hole lensing, *Phys. Rev. D* **62**, p. 084003 (October 2000).
7. V. Bozza, Gravitational lensing in the strong field limit, *Phys. Rev. D* **66**, p. 103001 (November 2002).
8. J. P. Luminet, Image of a spherical black hole with thin accretion disk., *A&A* **75**, 228 (May 1979).
9. M. D. Johnson, A. Lupsasca, A. Strominger, G. N. Wong *et al.*, Universal interferometric signatures of a black hole's photon ring, *Science Advances* **6**, p. eaaz1310 (March 2020).
10. F. Aratore and V. Bozza, Decoding a black hole metric from the interferometric pattern of relativistic images, *arXiv e-prints*, p. arXiv:2107.05723 (July 2021).
11. V. Bozza and G. Scarpetta, Strong deflection limit of black hole gravitational lensing with arbitrary source distances, *Physical Review D* **76**, p. 083008 (2007).
12. D. Pesce, K. Haworth, G. J. Melnick *et al.*, Extremely long baseline interferometry with Origins Space Telescope, in *Bulletin of the American Astronomical Society*, September 2019.
13. A. S. Andrianov, A. M. Baryshev, H. Falcke *et al.*, Simulations of M87 and Sgr A* imaging with the Millimetron Space Observatory on near-Earth orbits, *MNRAS* **500**, 4866 (January 2021).
14. A. Chael, M. D. Johnson and A. Lupsasca, Observing the Inner Shadow of a Black Hole: A Direct View of the Event Horizon, *arXiv e-prints*, p. arXiv:2106.00683 (June 2021).
15. Event Horizon Telescope Collaboration, K. Akiyama, J. C. Algaba, A. Alberdi *et al.*, First M87 Event Horizon Telescope Results. VIII. Magnetic Field Structure near The Event Horizon, *ApJ* **910**, p. L13 (March 2021).
16. Gravity Collaboration, R. Abuter, A. Amorim, M. Bauböck, J. P. Berger *et al.*, Detection of faint stars near Sagittarius A* with GRAVITY, *A&A* **645**, p. A127 (January 2021).
17. V. Bozza and L. Mancini, Time Delay in Black Hole Gravitational Lensing as a Distance Estimator, *General Relativity and Gravitation* **36**, 435 (February 2004).

Symplectic evolution of an observed light bundle

N. Uzun

Univ Lyon, Ens de Lyon, Univ Lyon1, CNRS, Centre de Recherche Astrophysique de Lyon Lyon, UMR5574, F69007, France
E-mail: nezihe.uzun@ens-lyon.fr
http://www.univ-lyon1.fr

Each and every observational information we obtain from the sky regarding the brightnesses, distances or image distortions resides on the deviation of a null geodesic bundle. In this talk, we present the symplectic evolution of this bundle on a reduced phase space. The resulting formalism is analogous to the one in paraxial Newtonian optics. It allows one to identify any spacetime as an optical device and distinguish its thin lens, pure magnifier and rotator components. We will discuss the fact that the distance reciprocity in relativity results from the symplectic evolution of this null bundle. Other potential applications like wavization and its importance for both electromagnetic and gravitational waves will also be summarized.

Keywords: General relativity, geodesic deviation, optics, symplectic

1. Introduction

When we make an observation on the sky, we deduce the brightness, distance and image distortion information about astrophysical objects via a bundle of null rays. This geodesic bundle connects the emitter and the observer that are located at two distant regions of a spacetime. Nevertheless, they receive reciprocal information about each other.

For example, an observer located at point O on Earth, obtains the distance to a galaxy cluster located at point G, by making use of the observed solid angle at point O, and the estimated cross–sectional area on the sky at point G. This gives the angular diameter distance, D_A. Such a distance estimation is mostly relevant for standard rulers in cosmology as, for example, baryon acoustic oscillations. Likewise, an observer at point G can estimate the distance to a point O in a similar fashion. According to an observer on Earth, on the other hand, this information is contained in the luminosity distance, D_L. Namely, by using the observed flux

$$F_{\text{observed}} = \frac{\text{Source luminosity}}{\text{Surface area}} = \frac{L_{\text{source}}}{4\pi D_U^2}, \tag{1}$$

one can deduce an *uncorrected* luminosity distance, D_U. When the relativistic corrections are included, one obtains a corrected luminosity distance, D_L, via

$$D_L = (1+z)^{-1} D_U, \tag{2}$$

where z is the redshift factor. The relationship between the luminosity distance and the angular diameter distance is given by Etherington's distance reciprocity theorem[1]

$$D_L = (1+z)D_A. \tag{3}$$

The distance reciprocity, eq (3), was originally derived for light propagation within a single spacetime. This means that the initial point of the propagation is an observation point at which the geodesic deviation vector becomes zero. Eventually, it can be shown that the distance reciprocity follows from the symmetries of the Riemann curvature tensor.[1,2]

Previously, we showed that[3] one can extend the proof for more generic scenarios. Namely, for a ray bundle propagating in multiple geometries, one needs to propagate a bundle with arbitrary initial conditions. Such a scenario was used, for example, in Fleury *et al.*[4,5] In those investigations, there exists a 4×4 Wronskian matrix which takes the initial ray bundle variables to the final ones. A similar construction can be found[6–9] where the Wronskian in question is 8–dimensional and it includes more information regarding the light bundle evolution. We demonstrated[3] that the Wronskian matrix of Fleury *et al.* is indeed symplectic and the most general form of the distance reciprocity follows from the symplectic symmetries of this transformation matrix. This was achieved by following a Hamiltonian formalism defined on a 4–dimensional phase space. Note that our approach is analogous to the reduced phase space optics formalism developed for classical paraxial rays.[10,11]

In classical optics, one starts with Fermat's action in order to obtain a Hamiltonian formalism in which rays are the geodesic solutions of the corresponding optical metric of Fermat. Likewise, we implemented the usage of geodesic actions applied to a null bundle. The outcome is an effective geodesic deviation action which was originally derived up to higher orders in Vines' work.[12]

The aim of this talk is to give a brief summary of our previous formalism,[3] along with discussing the importance of implementing such new techniques to gravitational optics. We summarize the aforementioned symplectic phase space approach in Section (2). Some perspectives on the symplectic ray bundle evolution is given in Section (3) and we conclude with Section (4).

2. Symplectic evolution of light bundles on phase space

2.1. *Vines' bi–local geodesic deviation action*

It is known that connecting two geodesics is, in general, non–local. Therefore, if one wants to obtain a geodesic deviation action, one needs to consider a bi–local formalism. Vines[12] achieves this for generic curves by following the definition of Synge's world function[13] $\sigma(x,y)$. This is a bi–local object which depends on two spacetime points x and y. Those two points are connected by a unique geodesic, Σ,

such that

$$\sigma(x,y) = \frac{1}{2}\begin{cases} (\text{proper distance})^2, & \Sigma : \text{spacelike} \\ 0, & \Sigma : \text{null} \\ -(\text{proper time})^2, & \Sigma : \text{timelike.} \end{cases}$$

Now consider a bundle whose outer most null geodesic, $\Upsilon(v)$, has a tangent vector $\vec{k}'$ (See Fig. 1.). The central null geodesic, $\chi(v)$, with a tangent vector $\vec{k}$, can also be parameterized with the same affine parameter v. This is called isosynchronous parameterization. Then those tangent vectors satisfy $\nabla_{\vec{k}'}\vec{k}' = 0$ and $\nabla_{\vec{k}}\vec{k} = 0$ where ∇ denotes the covariant derivative operator defined at a specified point. Moreover, physical sizes of the objects on the sky are estimated by the proper sizes. Therefore, a spacelike world function, being the measure of proper distance between two spacetime points, is the most relevant tool for our construction.

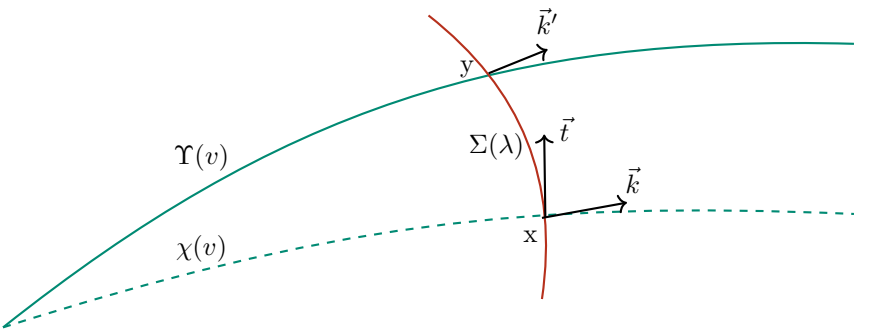


Fig. 1. The central null geodesic $\chi(v)$ is plotted with dashed lines. The red curve represents $\Sigma(\lambda)$ given by Synge's world function that is spacelike. The outermost null geodesic $\Upsilon(v)$ can be uniquely obtained through $\chi(v)$ and $\Sigma(\lambda)$.

Now consider a spacelike geodesic, $\Sigma(\lambda)$, that connects $\Upsilon(v)$ to $\chi(v)$. One can define an exponential map on $\Sigma(\lambda)$ via the derivatives of the world function $\sigma(x,y)$. This results in a bi–local object $\vec{\eta}$, whose variation with respect to the affine parameter that parameterizes the null bundle, is given by[3,12]

$$\dot{\eta}^\alpha = \frac{\mathbb{D}\eta^\alpha}{dv} = \left(k^\beta \nabla_\beta + k'^\mu \nabla_\mu\right)\eta^\alpha. \tag{4}$$

Note that for an arbitrarily large separation one needs to consider the covariant derivatives of $\vec{\eta}$ at two distinct points along the null bundle as in eq. (4). In that case, one can write the action functional of $\Upsilon(v)$, i.e., $S_\Upsilon = \int \frac{1}{2}k'^2 dv$, by making use of the action functional, $S_\chi = \int \frac{1}{2}k^2 dv$, of the central null geodesic, $\chi(v)$, and some additional terms,[3,12] i.e.,

$$S_\Upsilon = S_\chi + \int \frac{1}{2}\left[2\vec{k}\cdot\dot{\vec{\eta}} + \dot{\vec{\eta}}\cdot\dot{\vec{\eta}} - R_{\vec{\eta}\vec{k}\vec{\eta}\vec{k}} + O(\vec{\eta},\dot{\vec{\eta}})^3\right] dv. \tag{5}$$

Here, $R^{\alpha}{}_{\gamma\beta\delta}$ is the Riemann curvature tensor. In order to obtain the second integral, that appears on the r.h.s of eq. (5), one solves eq. (4) for $\vec{k}'$ and substitutes it in $S_{\Upsilon} = \int \frac{1}{2} k'^2 dv$. However, this is not a straightforward calculation as it requires the bi–local object $\dot{\eta}^{\alpha}$ to be Taylor expanded in the coincidence limit. We advice the reader to see the original derivation of Vines[12] for generic scenarios or our previous paper[3] for further details in this context.

On the other hand, for the observations on the sky, what is relevant is a thin bundle of null geodesics, rather than the ones with large separations. Thus, once the local limit of eq. (5) is taken for an infinitesimally thin bundle, one can consider $\vec{\eta}$ as a local geodesic deviation vector. Then, one obtains an effective action

$$S = \int \left(\frac{1}{2} \dot{\vec{\eta}} \cdot \dot{\vec{\eta}} + \frac{1}{2} R_{\vec{\eta}\vec{k}\vec{k}\vec{\eta}} \right) dv, \tag{6}$$

whose extremization results in a first order geodesic deviation equation

$$\ddot{\eta}^{\alpha} = R^{\alpha}{}_{\vec{k}\vec{k}\vec{\eta}}\,. \tag{7}$$

Note that as from eq. (6), and onwards, the overdot represents a local covariant derivative with respect to the central ray $\vec{k}$.

In the next section, we will introduce a local tetrad and an associated screen basis around an observation point. This will allow us to rewrite the action functional (6) in terms of the observable variables in the following sections.

2.2. *Screen basis, cross–sections and solid angles*

Consider an observer with 4–velocity $\vec{u}$. Assume that (s)he observes a thin null bundle with a central geodesic whose tangent vector is given by $k^{\alpha} = \omega\,(u^{\alpha} + r^{\alpha})$. Here, the observed frequency of light is given by $\omega = -\vec{k} \cdot \vec{u}$ and $\vec{r}$ is a spacelike vector and thus $\vec{u} \cdot \vec{r} = 0$. We will also consider a 2–dimensional spacelike screen represented by the Sachs basis,[14,15] s^{α}_{a}, with $\{a, b\} = \{1, 2\}$. Then,

$$\vec{s}_a \cdot \vec{s}_b = \delta_{ab}, \qquad \vec{u} \cdot \vec{s}_a = 0, \qquad \vec{r} \cdot \vec{s}_a = 0, \qquad \nabla_{\vec{k}} \vec{s}_a = 0, \tag{8}$$

are satisfied. This orthogonal screen is the one on which the observables are projected.

For an observational bundle, $\vec{k} \cdot \vec{\eta} = 0$ holds initially as the null bundle converges into a vertex point where the observer is located. Note that this condition is preserved throughout the evolution of the bundle as (i) $\nabla_{\vec{k}} \vec{\eta} = \nabla_{\vec{\eta}} \vec{k}$ holds due to the propagation vector, $\vec{k}$, and its Jacobi field, $\vec{\eta}$, being Lie dragged along each other,[16] (ii) the propagation vector satisfies the null condition, (iii) the first order geodesic deviation equation, (7), is satisfied. Therefore, one can decompose the deviation vector as[3,15]

$$\vec{\eta} = \eta^{k} \vec{k} + \eta^{1} \vec{s}_1 + \eta^{2} \vec{s}_2. \tag{9}$$

Note that we will denote components of $\vec{\eta}$ residing on the orthogonal screen as $\boldsymbol{\eta} := \eta^{1} \vec{s}_1 + \eta^{2} \vec{s}_2$.

Those screen–projected components are useful in identifying observationally relevant objects. For example, the cross–sectional area, $d\mathcal{A}$, of an extended object and an observed solid angle, $d\Theta$, are respectively given by[2]

$$d\mathcal{A} := \left|\eta^1 \wedge \eta^2\right|, \qquad d\Theta := \left|\frac{d\eta^1}{d\ell} \wedge \frac{d\eta^2}{d\ell}\right|. \tag{10}$$

where $|d\ell| = \omega dv$ is the infinitesimal proper length written in terms of the affine parameter, v, of the null bundle and the observed frequency of light, ω. The symbol $\wedge$ represents the exterior product.

In the next section, we will introduce a Hamiltonian formalism associated with the action functional, (6). This Hamiltonian will be defined on a phase space whose coordinates are constructed via the screen–projected deviation vector, $\boldsymbol{\eta}$, and its derivatives, $\dot{\boldsymbol{\eta}}$.

2.3. *A reduced phase space and its Hamiltonian*

In general relativity, the standard way of constructing an optical phase space resides on a $3+1$ decomposition of a spacetime. Specifically, one chooses the phase space coordinates as the ones induced on a 3–dimensional spacelike hypersurface and the phase space momenta are chosen as the induced 3–momenta of a photon. Alternatively, we show that[3] a 4–dimensional phase space can be effectively constructed for a null bundle via treating the screen–projected deviation vector, as phase space Darboux coordinates, q^a, and its derivatives along the bundle as phase space momenta, p_b. Then a phase space vector can be defined as

$$\boldsymbol{\varkappa} = \begin{bmatrix} q^a \\ p_b \end{bmatrix} = \begin{bmatrix} \eta^1 \\ \eta^2 \\ \dot{\eta}_1 \\ \dot{\eta}_2 \end{bmatrix}. \tag{11}$$

Note that, the overdot in eq. (11) now denotes a standard total derivative with respect to the evolution parameter v as we are considering only the dyad components of $\vec{\eta}$.

Let us reconsider the effective geodesic deviation action functional given in eq. (6). By (i) treating the integrant of the action functional as a Lagrangian, (ii) using the symmetries of the Riemann tensor and (iii) rewriting the Lagrangian in terms of the phase space coordinates, one can obtain a reduced Lagrangian. A passage to the Hamiltonian formalism gives us a reduced Hamiltonian

$$H = \frac{1}{2}\delta^{ab}\dot{\eta}_a\dot{\eta}_b - \frac{1}{2}\mathcal{R}_{ab}\,\eta^a\eta^b, \tag{12}$$

where $\mathcal{R}_{ab} := R_{a\vec{k}\vec{k}b}$ is known as the *optical tidal matrix* in cosmology.[15,17]

Then, the corresponding Hamiltonian equations can be written as a matrix equation

$$\dot{\boldsymbol{\varkappa}} = \mathbf{L_H}\,\boldsymbol{\varkappa}, \qquad \text{with} \qquad \mathbf{L_H} = \left[\begin{array}{c|c} \mathbf{0_2} & \delta^{ab} \\ \hline \mathcal{R}_{ab} & \mathbf{0_2} \end{array}\right], \tag{13}$$

where $\mathbf{L_H}$ is the Hamiltonian matrix representation of the Lie operator defined through the reduced Hamiltonian, H, given in eq (12) and $\mathbf{0_2}$ is a 2×2 zero matrix.

Let us now investigate the solutions of the Hamiltonian equations, (13), in the next section.

2.4. *Symplectic evolution and its symmetries*

The Hamiltonian equations, (13), presented in the previous section are linear equations as $\mathbf{L_H} = \mathbf{L_H}(v)$. Then, its solutions are given by a linear transformation of an input phase vector, $\boldsymbol{\varkappa}_0$, defined at v_0, which gives an output phase space vector, $\boldsymbol{\varkappa}$, at some value of v, i.e.,

$$\boldsymbol{\varkappa} = \mathbf{T}\,(v, v_0)\,\boldsymbol{\varkappa}_0. \tag{14}$$

Substituting eq. (14) back into eq. (13) shows that the ray bundle transfer matrix, $\mathbf{T}$, satisfies a similar equation as the phase space vector $\boldsymbol{\varkappa}$, i.e., $\dot{\mathbf{T}} = \mathbf{L_H}\,\mathbf{T}$. Then, its solution is given by taking an ordered exponential (OE) map,

$$\mathbf{T}\,(v, v_0) = \mathrm{OE}\left[\int_{v_0}^{v} \mathbf{L_H} dv\right] \mathbf{T}\,(v_0, v_0)\,, \qquad \text{with} \qquad \mathbf{T}\,(v_0, v_0) = \mathbf{I_4}, \tag{15}$$

and $\mathbf{I_4}$ is a 4×4 identity matrix. In that case, $\mathbf{T}$, is a symplectic matrix which satisfies

$$\mathbf{T}^{\intercal}\,\boldsymbol{\Omega}\,\mathbf{T} = \boldsymbol{\Omega}, \qquad \det \mathbf{T} = 1, \tag{16}$$

as exponential map of Hamiltonian matrices are symplectic matrices. Here, $^{\intercal}$ denotes the transpose operator and $\boldsymbol{\Omega}$ is the well–known *fundamental symplectic matrix* whose components are given by

$$\Omega^{ij} = \left[\begin{array}{c|c} \mathbf{0_2} & \mathbf{I_2} \\ \hline -\mathbf{I_2} & \mathbf{0_2} \end{array}\right]. \tag{17}$$

This allows one to treat the ray bundle evolution in general relativity analogous to a dynamical problem in classical mechanics. A similar approach has indeed been developed in paraxial Newtonian optics much earlier (Cf. references in[10,11]). In that case, this approach is also known as the *ABCD–matrix method* as it is very common to represent a symplectic matrix, $\mathbf{T}$, in a block form

$$\mathbf{T} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]. \tag{18}$$

In our case, the submatrices in eq. (18) are all 2×2 matrices which satisfy certain symmetry conditions

$$\begin{gathered} \mathbf{A}\mathbf{B}^{\intercal},\ \mathbf{A}^{\intercal}\mathbf{C},\ \mathbf{B}^{\intercal}\mathbf{D} \text{ and } \mathbf{C}\mathbf{D}^{\intercal} \text{ are symmetric,} \\ \mathbf{A}\mathbf{D}^{\intercal} - \mathbf{B}\mathbf{C}^{\intercal} = \mathbf{I_2}, \end{gathered} \tag{19}$$

due to $\mathbf{T}$ being a symplectic matrix, i.e., eq. (16) is satisfied. Then, the Hamiltonian equations, (13), become equivalent to,

$$\begin{aligned} \dot{\mathbf{A}} &= \mathbf{C}, & \mathbf{A}\,(v_0, v_0) &= \mathbf{I_2}, \\ \dot{\mathbf{B}} &= \mathbf{D}, & \mathbf{B}\,(v_0, v_0) &= \mathbf{0_2}, \\ \dot{\mathbf{C}} &= \mathcal{R}\mathbf{A}, & \mathbf{C}\,(v_0, v_0) &= \mathbf{0_2}, \\ \dot{\mathbf{D}} &= \mathcal{R}\mathbf{B}, & \mathbf{D}\,(v_0, v_0) &= \mathbf{I_2}. \end{aligned} \tag{20}$$

Therefore, once the solutions of the equation set above is found, one can find the explicit form of the transfer matrix, $\mathbf{T}$, and the geodesic deviation variables of the null bundle in question. This allows one to estimate certain observables as we discuss in the next section.

3. Some perspectives on the symplectic ray bundle evolution

3.1. *Distance reciprocity*

The relationship of the ray bundle transfer matrix, $\mathbf{T}$, to the observables is not immediately obvious. In order to estimate the distance to an object located at G, for example, one starts with propagating the bundle with the initial phase space vector $\boldsymbol{\varkappa}_o = (\mathbf{0}, \dot{\eta}_o)^\intercal$ at the observation point O (See Fig. 2.). Then, one can calculate the cross–sectional area, $d\mathcal{A}$, at G by (i) making use of the definitions of $d\mathcal{A}$ and the solid angle, $d\Theta$, given in eq. (10), (ii) inputting the information of the measured $d\Theta$ at point O in the initial phase space vector, $\boldsymbol{\varkappa}_o = (\mathbf{0}, \dot{\eta}_o)^\intercal$, and (iii) considering the ray bundle transformation matrix $\mathbf{T}$ that transfers the initial phase space vector to a final one as in eq. (14). In that case, the estimated distance in question is the angular diameter distance given by $D_A = (d\mathcal{A}_G/d\Theta_O)^{1/2}$. Likewise, a similar distance estimation can be done by an observer located at point G. Then, the estimated distance is the luminosity distance given by $D_L = (d\mathcal{A}_O/d\Theta_G)^{1/2}$.

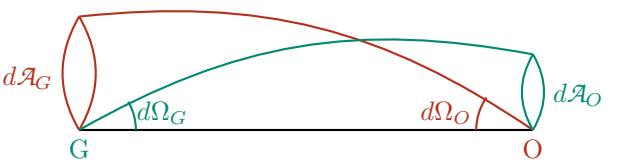


Fig. 2. Observer at O measures a solid angle $d\Omega_O$ and estimates the cross-sectional area of an astrophysical object at G as $d\mathcal{A}_G$. Same thing is applicable for an observer at G as well. Reciprocity of the estimated distances is written as $d\mathcal{A}_O/d\Omega_G = (1+z)^2\, d\mathcal{A}_G/d\Omega_O$.

As the solution of a phase space vector is given by eq. (14) in which $\mathbf{T}$ is represented in a block form (18), one obtains the angular diameter and luminosity distances respectively by making use of only the submatrix $\mathbf{B}$, i.e.,

$$D_A = \omega_O \det\left[\mathbf{B}\,(v_G, v_O)\right]^{1/2}, \qquad D_L = \omega_G \det\left[\mathbf{B}\,(v_O, v_G)\right]^{1/2}. \tag{21}$$

Here, ω_O and ω_G are the frequencies measured at respective points through which a redshift parameter is defined as $1+z=\omega_G/\omega_O$. Note that the frequency terms appear in the process of transforming the proper length, as it appears in the definition of solid angle in eq.(10), to the affine parameter, v, as we discussed in Section (2.2). The abbreviation "det" refers to the determinant of a given matrix.

In order for Etherington's distance reciprocity, eq. (3), to hold one needs to show that $\det\left[\mathbf{B}\left(v_G,v_O\right)\right]=\det\left[\mathbf{B}\left(v_O,v_G\right)\right]$ holds. Indeed, this was previously proven by making use of the symmetries of the Riemann tensor.[15] Note that such a case is applicable for light propagation within a single geometry. For scenarios that include light propagation in multiple spacetimes, one needs to propagate the bundle not only from the initial vertex point but also from arbitrary initial phase space points to other arbitrary ones. In that case, the information about the light bundle transformation is not only contained in the submatrix $\mathbf{B}$ but in all of the submatrices of $\mathbf{T}$. For example, this is applicable for light propagation in Swiss–cheese type models.[4,5]

Previously, we showed that distance reciprocity follows from the symplectic conditions, eq. (19), of the ray bundle transfer matrix for generic scenarios.[3] To be more specific, consider a ray bundle matrix, $\mathbf{T}(v_f,v_i)$, which takes an initial phase space vector, $\boldsymbol{\varkappa}(v_i)$, to a final one, $\boldsymbol{\varkappa}(v_f)$. Such a linear transformation should be equal to its inverse, once the initial and final points are traversed, i.e., $\mathbf{T}(v_i,v_f)=\mathbf{T}^{-1}(v_f,v_i)$. Moreoever, symplectic matrices form a group, thus inverse of a symplectic matrix is also symplectic. This requires $\mathbf{T}^{-1}(v_f,v_i)=\boldsymbol{\Omega}^{-1}\,\mathbf{T}^{\intercal}(v_f,v_i)\boldsymbol{\Omega}$. Then, it is easy to show that the distance reciprocity indeed follows from $\mathbf{T}(v_i,v_f)=\boldsymbol{\Omega}^{-1}\,\mathbf{T}^{\intercal}(v_f,v_i)\boldsymbol{\Omega}$. See our previous work[3] for further details in which we show this result more explicitly.

3.2. *Decomposition of an optical propagation*

One of the earliest predictions of Einstein's general relativity was on the light propagation within our Solar System. Back then, the idea of a spacetime structure was so hard to grasp that the astronomers conceptualized the effect of a massive object on the trajectory of light as "bending". Similarly, the name "gravitational lensing" remained within the community after Eddington performed his famous Solar eclipse experiment. Even more than 100 years after this observation, the researchers in the field use the term gravitational lensing in a loose sense.

On the other hand, in Newtonian optics a lens is defined more rigorously. For the paraxial case, for example, it is identified as a linear transformation that creates a shearing affect on the phase space defining the ray propagation. Though, lenses are not the only type of optical devices and there is a common practice in Newtonian optics in order to identify the components of light propagation. This is achieved by an Iwasawa decomposition[18] where a symplectic ray transformation matrix is decomposed into its submatrices belonging to a nilpotent subgroup, an abelian subgroup and a maximally compact subgroup for 1–dimensional systems.

This allows one to identify the thin lens, the pure magnifier and the fractional Fourier transformer components of an optical propagation problem uniquely. In higher dimensions, such an identification is known as Iwasawa factorization whose components can be obtained via the submatrices of the transformation matrix **T** (Cf.[10] for further details).

Previously, we proposed that any spacetime can be treated as an optical device within our formalism[3] which allows one to identify the lenses, the magnifiers and the rotators of an optical propagation in general relativity in a rigorous manner. In that case, spacetimes can be compared and contrasted in terms of their significance in different types of optical transformations, by making use of the Iwasawa factorization of their symplectic ray bundle transfer matrices.

3.3. *From rays back to waves*

Finding the exact solutions of the Maxwell's equation on a curved background is not an easy task in general relativity. There exists a similar problem in Newtonian optics for light propagation in media with an arbitrary refractive index. Within the paraxial regime of Newtonian optics, there exist certain "wavization" techniques that are adopted from quantum mechanics in order to find approximate solutions of the Maxwell equations.

The analogy between classical optics and quantum mechanics follows from the similarities between the Schrödinger equation of a particle wavefunction and the Maxwell equation of a complex amplitude of a classical wave in the scalar theory.[19,20] The $\hbar \to 0$ limit of quantum mechanics is said to give classical mechanics, and $\lambda \to 0$ limit of wave optics gives ray optics. Moreover, both the classical mechanics and the paraxial ray optics are constructed on symplectic phase spaces. Thus, in the first order limit, a wave picture can be attained back by making use of some phase space techniques that were initially introduced within quantum mechanics. In that case, phase space distribution functions, like Wigner function (or other analogous definitions), and metaplectic operators that are responsible for the spreading of a wave, play the central role in the wavization process.[11]

We proposed[3] that such methods can be adopted from Newtonian optics as we also reduced the problem of light bundle propagation into a symplectic evolution problem in phase space. Note that this is immediately relevant for gravitational lensing studies because when the wavelength of light is not much smaller than the gravitational lensing object, the wave effects on lensing become important. It is known that the wave effects amplify the estimated intensity of a wave in general.[21,22] Also, an earlier result showed that the wave and ray optics do not agree in terms of the position of the image formation for a specific geometry.[23] However, there is not much known in terms of the image formation in wave optics of light.[24] There are limited number of studies that investigate the wave optics and those usually only focus on the lensing problem on static background.[25–27] We propose that symplectic methods can be quite useful in this respect.

Moreover, it is known that gravitational waves also have null propagation vectors similar to light rays. Therefore, geometric optics is relevant for the investigation of gravitational waves at the linear order.[28,29] Also, the interaction of gravitational waves with matter fields is known to be weak. Thus, their coherence is expected to be preserved even over cosmological distances.[30] Then, the interplay between the ray and wave behaviours can be used to study gravitational waves within our formalism. Also note that wave effects on the lensing of the gravitational waves have been investigated in some studies only lately.[24,31–37] However, all of those studies focus on wave effects on specific geometries.

In addition, recently, new ideas are introduced to the literature regarding the wave effects. For example, the spin Hall effect which is normally studied in condensed matter field, is known to bring corrections to the geometrical optics limit in the case of light rays. Oancea *et al.*[38] found an analogous effect for the case of gravitational waves. Similarly, the Aharonov-Bohm effect is known from quantum electromagnetism. Baraldo *et al.*[30] found that there exists an analogous effect for the gravitational waves. Namely, a shift is predicted on the interference pattern of the gravitational waves due to the angular momentum of a lensing object.

Therefore, we propose that once a wavization procedure is applied to our symplectic null bundle propagation, evolution of gravitational waves and their lensing can be studied within more generic scenarios as well.

4. Conclusion

In this talk, we summarized our work[3] which reduces an observational light bundle evolution in general relativity analogous to a classical optical problem in phase space. This requires simultaneous implementation of a null geodesic action to the outermost null curve and the central null curve of an observational light bundle. One then obtains an effective geodesic deviation action which can be written in terms of the screen–projected variables of the null geodesic deviation. This reduces the problem by one order and a 4–dimensional reduced phase space can be obtained by treating the screen–projections of the geodesic deviation vectors as canonical coordinates and their derivatives along the central null ray as canonical momenta. The Hamiltonian equations of this first order problem can be written as a matrix equation whose solutions are obtained via linear symplectic transformations. We shortly discussed (i) a generic proof of the distance reciprocity in general relativity by making use of the symmetries of the symplectic ray bundle transfer matrix, (ii) decomposition of the light bundle propagation into thin lens, pure magnifier and fractional Fourier transformation portions such that any spacetime can be rigorously identified as an optical device, (iii) the relevance of our method to the wave optics for electromagnetic and gravitational waves which have been drawing increasing interest in the literature lately.

Acknowledgements

The author thanks Thomas Buchert for his comments. This work is a part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ERC advanced Grant 740021ARTHUS, PI: Thomas Buchert).

References

1. I. M. H. Etherington, On the definition of distance in general relativity, *Philos. Mag.* **15**, p. 761 (1933).
2. G. F. R. Ellis, Relativistic cosmology, *Gen. Rel. Grav.* **41**, p. 581 (2009).
3. N. Uzun, Reduced phase space optics for general relativity: Symplectic ray bundle transfer, *Class. Quant. Grav.* **37**, p. 045002 (2020).
4. P. Fleury, H. Dupuy and J. P. Uzan, Interpretation of the Hubble diagram in a nonhomogeneous universe, *Phys. Rev. D* **87**, p. 123526 (2013).
5. P. Fleury, Swiss-cheese models and the Dyer-Roeder approximation, *J. Cosmol. Astropart. Phys.* **06**, p. 054 (2014).
6. M. Grasso, M. Korzyński and J. Serbenta, Geometric optics in general relativity using bilocal operators, *Phys. Rev. D* **99**, p. 064038 (Mar 2019).
7. M. Korzyński and E. Villa, Geometric optics in relativistic cosmology: New formulation and a new observable, *Phys. Rev. D* **101**, p. 063506 (Mar 2020).
8. M. Korzyński, J. Miśkiewicz and J. Serbenta, Weighing the spacetime along the line of sight using times of arrival of electromagnetic signals, *Phys. Rev. D* **104**, p. 024026 (Jul 2021).
9. M. Grasso, E. Villa, M. Korzyński and S. Matarrese, Isolating nonlinearities of light propagation in inhomogeneous cosmologies, *Phys. Rev. D* **104**, p. 043508 (Aug 2021).
10. K. B. Wolf, *Geometric Optics on Phase Space*, Theoretical and Mathematical Physics, Theoretical and Mathematical Physics (Springer, Berlin Heidelberg, 2004).
11. A. Torre, *Linear Ray and Wave Optics in Phase Space: Bridging Ray and Wave Optics via the Wigner Phase-Space Picture* (Elsevier Science, 2005).
12. J. Vines, Geodesic deviation at higher orders via covariant bitensors, *Gen. Rel. Grav.* **47**, p. 59 (2015).
13. J. L. Synge, *Relativity: the general theory*, Series in Physics, Series in Physics (North Holland, 1960).
14. R. Sachs, Gravitational waves in general relativity. VI. The outgoing radiation condition, *Proc. R. Soc. London* **264**, p. 309 (1961).
15. V. Perlick, Gravitational Lensing from a Spacetime Perspective, *Living Rev. Rel.* **7**, p. 9 (2004).
16. R. M. Wald, *General Relativity* (Chicago Univ. Pr., Chicago, USA, 1984).
17. S. Seitz, P. Schneider and J. Ehlers, Light propagation in arbitrary spacetimes and the gravitational lens approximation, *Class. Quantum Grav.* **11**, p. 2345 (1994).
18. K. Iwasawa, On Some Types of Topological Groups, *Ann. Math.* **50**, p. 507 (1949).
19. D. Dragoman, *Phase space correspondence between classical optics and quantum mechanics*, Progress in Optics Vol. 43 (Elsevier, 2002), p. 433.
20. D. Dragoman, Applications of the Wigner Distribution Function in Signal Processing, *EURASIP J. Appl. Signal Process.* **2005**, p. 1520 (2005).
21. T. T. Nakamura and S. Deguchi, Wave Optics in Gravitational Lensing, *Prog. Theor. Phys. Supplement* **133**, 137 (01 1999).

22. N. Matsunaga and K. Yamamoto, The finite source size effect and the wave optics in gravitational lensing, *J. Cosmol. Astropart. Phys.* **01**, p. 023 (2006).
23. E. Herlt and H. Stephani, Wave Optics of the Spherical Gravitational Lens. II. Diffraction of a Plane Electromagnetic Wave by a Black Hole, *Int. J. Theor. Phys.* **17**, 189 (March 1978).
24. Y. Nambu, Wave optics and image formation in gravitational lensing, *J. Phys.: Conference Series* **410**, p. 012036 (Feb 2013).
25. Y. Nambu and S. Noda, Wave optics in black hole spacetimes: The Schwarzschild case, *Class. Quant. Grav.* **33**, p. 075011 (Apr 2016).
26. S. G. Turyshev and V. T. Toth, Diffraction of electromagnetic waves in the gravitational field of the sun, *Phys. Rev. D* **96**, p. 024008 (Jul 2017).
27. S. G. Turyshev and V. T. Toth, Wave-optical treatment of the shadow cast by a large sphere, *Phys. Rev. A* **97**, p. 033810 (Mar 2018).
28. K. S. Thorne, The theory of gravitational radiation — An introductory review, in *Gravitational Radiation*, January 1983.
29. S. R. Dolan, Geometrical optics for scalar, electromagnetic and gravitational waves on curved spacetime, *Int. J. Mod. Phys. D* **27**, p. 1843010 (2018).
30. C. Baraldo, A. Hosoya and T. T. Nakamura, Gravitationally induced interference of gravitational waves by a rotating massive object, *Phys. Rev. D* **59**, p. 083001 (Mar 1999).
31. R. Takahashi and T. Nakamura, Wave effects in gravitational lensing of gravitational waves from chirping binaries, *Astrophys. J.* **595**, 1039 (2003).
32. A. J. Moylan, D. E. McClelland, S. M. Scott, A. C. Searle and G. V. Bicknell, Numerical wave optics and the lensing of gravitational waves by globular clusters, in *11th Marcel Grossmann Meeting on General Relativity*, 10 2007.
33. C.-M. Yoo, T. Harada and N. Tsukamoto, Wave Effect in Gravitational Lensing by the Ellis Wormhole, *Phys. Rev. D* **87**, p. 084045 (2013).
34. R. Takahashi, Arrival time differences between gravitational waves and electromagnetic signals due to gravitational lensing, *Astrophys. J.* **835**, p. 103 (2017).
35. L. Dai and T. Venumadhav, On the waveforms of gravitationally lensed gravitational waves, *arXiv: 1702.04724* (2 2017).
36. L. Dai, S.-S. Li, B. Zackay, S. Mao and Y. Lu, Detecting lensing-induced diffraction in astrophysical gravitational waves, *Phys. Rev. D* **98**, p. 104029 (Nov 2018).
37. A. K. Meena and J. S. Bagla, Gravitational lensing of gravitational waves: wave nature and prospects for detection, *Mon. Not. Roy. Astro. Soc.* **492**, 1127 (2020).
38. M. A. Oancea, J. Joudioux, I. Y. Dodin, D. E. Ruiz, C. F. Paganini and L. Andersson, Gravitational spin hall effect of light, *Phys. Rev. D* **102**, p. 024075 (Jul 2020).

Shadow of black holes with a plasma environment in 4D Einstein-Gauss-Bonnet gravity

Javier Badía[1,2,*] and Ernesto F. Eiroa[1,†]

[1] *Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), Casilla de Correo 67, Sucursal 28, 1428, Buenos Aires, Argentina*
[2] *Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria Pab. I, 1428, Buenos Aires, Argentina*
[*] *E-mail: jbadia@iafe.uba.ar*
[†] *E-mail: eiroa@iafe.uba.ar*

We study the shadow cast by rotating black holes surrounded by plasma in the context of the 4D Einstein-Gauss-Bonnet theory of gravity. The metric for these black holes results from applying the Newman-Janis algorithm to a spherically symmetric solution. We obtain the contour of the shadow for a plasma frequency model that allows a separable Hamilton-Jacobi equation. We introduce three observables in order to characterize the position, size, and shape of the shadow.

Keywords: Black hole shadow; Modified gravity; Plasma.

1. Introduction

It is well known that Einstein-Gauss-Bonnet theory in four spacetime dimensions is purely topological and thus equivalent to general relativity. A novel gravity theory was recently proposed[1] by rescaling the coupling constant and taking the limit $D \to 4$ in the D-dimensional Einstein-Gauss-Bonnet theory, with the intention of overcoming this standard result. But this approach, based on some particular solutions, lacks of a complete set of field equations and does not have an intrinsically four-dimensional description in terms of a covariantly conserved rank-2 tensor in four dimensions.[2] A theory with a Gauss-Bonnet term in four dimensions that provide a solution to these problems was then presented,[3,4] which propagates a scalar field in addition to the metric tensor and the full action belongs to the Horndeski class of scalar-tensor theories of gravity. The action, obtained by a regularization procedure in a way that is free from divergences, produces well behaved second-order field equations. This theory, usually dubbed regularized or scalar-tensor 4D Einstein-Gauss-Bonnet gravity (4DEGB), includes a nonvanishing contribution coming from the Gauss-Bonnet term.[3,4]

The number of theoretical studies on black hole shadows[5,6] has sharply increased since the Event Horizon Telescope (EHT)[7,8] collaboration obtained the first reconstructed image of the supermassive black hole M87*, located at the center of the giant elliptical galaxy M87. Many articles have been published on this topic in recent years, here we can mention only a few of them,[9–12] in which more references can be found. The presence of plasma surrounding a black hole affects the characteristics of the shadow,[13–15] because photons with different frequencies follow distinct

trajectories, resulting in chromatic effects. In this work, we consider the shadow due to rotating black holes[16,17] in the scalar-tensor 4DEGB gravity, within a plasma environment. We define three observables to describe how the presence of plasma affects the size, the deformation and the position of the shadow. We analyze in detail the case of a Shapiro plasma, with the shadow seen by an equatorial observer. We use units such that $G = c = \hbar = 1$.

2. Black holes in 4D Einstein-Gauss-Bonnet gravity

The scalar-tensor 4DEGB action reads[3,4]

$$S = \int_{\mathcal{M}} d^4x\sqrt{-g}\Big[R + \gamma\Big(4G^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - \phi\mathcal{G} + 4\Box\phi(\nabla\phi)^2 + 2(\nabla\phi)^4\Big)\Big] + S_m. \quad (1)$$

The theory is free of divergences and belongs to the Horndeski class, with functions $G_2 = 8\gamma X^2$, $G_3 = 8\gamma X$, $G_4 = 1 + 4\gamma X$, and $G_5 = 4\gamma \ln X$, where $X = -(1/2)\nabla_\mu\phi\nabla^\mu\phi$. The field equations are obtained[3,4] by varying this action with respect to the metric

$$G_{\mu\nu} = \gamma\mathcal{H}_{\mu\nu} + T_{\mu\nu}, \quad (2)$$

where $\mathcal{H}$ is a complicated second order function in the derivatives of the scalar field ϕ; there is another equation that is found[3,4] by varying with respect to the scalar field ϕ. The expressions of these equations are not relevant for our purposes.

The 4DEGB field equations in vacuum admit the spherically symmetric and asymptotically flat solution[1,4] with the metric given by

$$ds^2 = -\left(1 - \frac{2m(r)}{r}\right)dt^2 + \left(1 - \frac{2m(r)}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2), \quad (3)$$

where

$$m(r) = \frac{r^3}{64\pi\gamma}\left(\sqrt{1 + \frac{128\pi\gamma M}{r^3}} - 1\right). \quad (4)$$

The parameter γ of the theory has units of mass squared. We only consider $\gamma > 0$, since the square root becomes imaginary for a finite value of r if $\gamma < 0$. The limit $\gamma \to 0$ corresponds to the Schwarzschild geometry. The black hole mass is M, since we have $m(r) \to M$ as $r \to \infty$.

A rotating solution was subsequently found by applying a modified version of the Newman-Janis algorithm,[16,17] having in the Boyer-Lindquist coordinates the form

$$ds^2 = -\frac{\rho^2\Delta}{\Sigma}dt^2 + \frac{\Sigma\sin^2\theta}{\rho^2}\left[d\varphi - \frac{2am(r)r}{\Sigma}dt\right]^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2, \quad (5)$$

where $a = J/M$ is the rotation parameter and

$$\begin{aligned}\rho^2 &= r^2 + a^2\cos^2\theta,\\ \Delta &= r^2 - 2m(r)r + a^2,\\ \Sigma &= (r^2 + a^2)^2 - a^2\Delta\sin^2\theta.\end{aligned} \quad (6)$$

The limit $\gamma \to 0$ recovers the Kerr black hole. The radius r_h of the event horizon corresponds to the largest real and positive solution of the equation $\Delta = 0$. It can be seen numerically that the condition $\gamma/M^2 < 0.00129$ is necessary for the existence of the event horizon, otherwise there is a naked singularity.

3. Photon geodesics in a plasma environment

The motion of photons in a pressureless, nonmagnetized plasma is governed by the Hamiltonian[18]

$$\mathcal{H} = \frac{1}{2}\left(g^{\mu\nu}(x)p_\mu p_\nu + \omega_p(x)^2\right), \tag{7}$$

with $g^{\mu\nu}$ the inverse of the metric tensor, x^μ the spacetime coordinates, p^μ the conjugate momenta, and ω_p the plasma electron frequency

$$\omega_p^2 = \frac{4\pi e^2}{m_e} N, \tag{8}$$

where e and m_e are the electron charge and mass respectively, and N the electron number density. We assume that the plasma frequency is stationary and axisymmetric, so it does not depend on the Boyer-Lindquist coordinates t and φ. As it happens for the Kerr geometry,[13] the inequality $\omega_0^2 \geq -g_{tt}\omega_p^2$ is a necessary and sufficient condition for a light ray with frequency ω_0 (measured by an observer at infinity) to exist at a given point of the spacetime.

3.1. *Hamilton-Jacobi equation*

We can write down the Hamilton-Jacobi (HJ) equation for photons, which reads

$$\mathcal{H}\left(x, \frac{\partial S}{\partial x}\right) = 0 \tag{9}$$

and we introduce the ansatz

$$S = -\omega_0 t + p_\varphi \varphi + S_r(r) + S_\theta(\theta). \tag{10}$$

By substituting it in the HJ equation, we arrive at

$$\begin{aligned}\Delta(S_r')^2 - \frac{1}{\Delta}\left[(r^2+a^2)^2\omega_0^2 + 4am(r)r\omega_0 p_\varphi + a^2 p_\varphi^2\right] \\ +(S_\theta')^2 + a^2\omega_0^2\sin^2\theta + \frac{p_\varphi^2}{\sin^2\theta} + \rho^2\omega_p^2 = 0,\end{aligned} \tag{11}$$

where the prime denotes the derivative with respect to the corresponding coordinate. The HJ equation is separable if and only if the plasma frequency takes the form[13]

$$\omega_p^2 = \frac{f_r(r) + f_\theta(\theta)}{\rho^2}, \tag{12}$$

with $f_r(r)$ and $f_\theta(\theta)$ arbitrary functions of the coordinates r and θ, respectively. The spacetime is axisymmetric, stationary, and asymptotically flat, therefore the

quantities $p_t = -\omega_0$ and p_φ are conserved along the geodesics of photons; $E = \omega_0$ is the photon energy and p_φ is the z component of the angular momentum. Since ω_p only depends on the coordinates r and θ, the quantities E and p_φ are still conserved in the presence of plasma. A third constant of motion for photons is $\mathcal{H} = 0$. By substituting ω_p into the Hamilton-Jacobi equation, we can separate it as

$$(S'_\theta)^2 + \left(a\omega_0 \sin\theta - \frac{p_\varphi}{\sin\theta}\right)^2 + f_\theta = -\Delta(S'_r)^2 + \frac{1}{\Delta}\left[\omega_0(r^2+a^2) - ap_\varphi\right]^2 - f_r. \quad (13)$$

Since the left hand side is a function only of θ and the right hand side a function only of r, they should both be equal to a constant $\mathcal{K}$. For convenience, we use instead the Carter constant,[19] defined by $\mathcal{Q} = \mathcal{K} - (p_\varphi - a\omega_0)^2$, as the fourth constant of motion. From the Hamilton equations, using that $\dot{x}^\mu = p^\mu = g^{\mu\nu}p_\nu$ (the dot denotes the derivative with respect to the curve parameter λ) and $p_\nu = \partial S/\partial x^\nu$, the equations of motion read

$$\rho^2 \dot{t} = \frac{r^2+a^2}{\Delta}P(r) - a(a\omega_0 \sin^2\theta - p_\varphi), \quad (14)$$

$$\rho^2 \dot{\varphi} = \frac{a}{\Delta}P(r) - a\omega_0 + \frac{p_\varphi}{\sin^2\theta}, \quad (15)$$

$$\rho^2 \dot{r} = \pm\sqrt{R(r)}, \quad (16)$$

$$\rho^2 \dot{\theta} = \pm\sqrt{\Theta(\theta)}, \quad (17)$$

where

$$R(r) = P(r)^2 - \Delta[\mathcal{Q} + (p_\varphi - a\omega_0)^2 + f_r], \quad (18)$$

$$\Theta(\theta) = \mathcal{Q} + \cos^2\theta\left(a^2\omega_0^2 - \frac{p_\varphi^2}{\sin^2\theta}\right) - f_\theta, \quad (19)$$

$$P(r) = \omega_0(r^2+a^2) - ap_\varphi. \quad (20)$$

These first-order equations determine the movement of the photons in the presence of plasma.

3.2. *Spherical photon orbits*

The orbits of photons with constant r should fulfill the two conditions given by $R(r) = R'(r) = 0$; such solutions are unstable, satisfying $R''(r) > 0$. From the equation $R(r) = 0$ we find that

$$\mathcal{Q} + (p_\varphi - a\omega_0)^2 + f_r = \frac{(\omega_0(r^2+a^2) - ap_\varphi)^2}{\Delta}. \quad (21)$$

Then, substituting into $R'(r) = 0$, we obtain

$$R' = 4\omega_0 r(\omega_0(r^2+a^2) - ap_\varphi) - \frac{\Delta'}{\Delta}(\omega_0(r^2+a^2) - ap_\varphi)^2 - \Delta f'_r = 0, \quad (22)$$

which is a quadratic equation for p_φ, with the solution

$$p_\varphi = \frac{\omega_0}{a}\left[r^2 + a^2 - \frac{2r\Delta}{\Delta'}\left(1 \pm \sqrt{1 - \frac{\Delta' f'_r}{4\omega_0^2 r^2}}\right)\right]. \quad (23)$$

Then, after some algebra, we can obtain $\mathcal{Q}$ in terms of r, which reads

$$\mathcal{Q} = -\frac{\omega_0^2 r^4}{a^2} + \frac{4\omega_0^2 r^2 \Delta}{a^2 \Delta'}\left[r - \frac{2}{\Delta'}\left(\Delta - a^2\right)\right]\left(1 \pm \sqrt{1 - \frac{\Delta' f_r'}{4\omega_0^2 r^2}}\right) + \frac{\Delta f_r'}{\Delta' a^2}\left(\Delta - a^2\right) - f_r. \tag{24}$$

We have found the critical values of p_φ and $\mathcal{Q}$ associated with the spherical orbits of photons. Any trajectory should have $\Theta \geq 0$, then

$$\mathcal{Q} + \cos^2\theta\left(a^2\omega_0^2 - \frac{p_\varphi^2}{\sin^2\theta}\right) - f_\theta(\theta) \geq 0. \tag{25}$$

This equation defines the photon region: spherical orbits exist at values of r and θ for which this inequality is satisfied.

3.3. *Shadow*

For a distant observer, the shadow is determined by the set of photon directions that never reach infinity and instead cross the event horizon. The boundary consists of those rays that asymptotically approach the spherical photon orbits of the spacetime, so they have the same conserved quantities. In order to relate the directions in the sky with the constants of motion, we take an observer at rest in the asymptotically flat region (large r_o) with an inclinaton angle θ_o from the spin axis of the black hole, and we construct the orthonormal tetrad as usual[9]:

$$\mathbf{e}_{\hat{t}} = \partial_t, \qquad \mathbf{e}_{\hat{r}} = \partial_r, \qquad \mathbf{e}_{\hat{\theta}} = \frac{1}{r_\mathrm{o}}\partial_\theta, \qquad \mathbf{e}_{\hat{\varphi}} = \frac{1}{r_\mathrm{o}\sin\theta_\mathrm{o}}\partial_\varphi, \tag{26}$$

so that the components of the four-momentum in this frame read

$$p^{\hat{t}} = \omega_0, \qquad p^{\hat{r}} = p^r, \qquad p^{\hat{\theta}} = r_\mathrm{o} p^\theta, \qquad p^{\hat{\varphi}} = r_\mathrm{o}\sin\theta_\mathrm{o} p^\varphi = \frac{p_\varphi}{r_\mathrm{o}\sin\theta_\mathrm{o}}. \tag{27}$$

For the plasma model, we assume that

$$\lim_{r\to\infty} \omega_p(r,\theta) = 0, \tag{28}$$

which is equivalent to

$$\lim_{r\to\infty} \frac{f_r(r)}{r^2} = 0, \tag{29}$$

so photons propagate in vacuum far away from the black hole. We adopt the celestial coordinates for an observer at infinity[5]

$$\alpha = -r_\mathrm{o}\frac{p^{\hat{\varphi}}}{p^{\hat{t}}}\bigg|_{r_\mathrm{o}\to\infty}, \tag{30}$$

$$\beta = -r_\mathrm{o}\frac{p^{\hat{\theta}}}{p^{\hat{t}}}\bigg|_{r_\mathrm{o}\to\infty}; \tag{31}$$

replacing p^θ in terms of the conserved quantities, they read

$$\alpha = -\frac{p_\varphi}{\omega_0 \sin\theta_\mathrm{o}}, \tag{32}$$

$$\beta = \pm\frac{1}{\omega_0}\sqrt{\mathcal{Q} + \cos^2\theta_\mathrm{o}\left(a^2\omega_0^2 - \frac{p_\varphi^2}{\sin^2\theta_\mathrm{o}}\right) - f_\theta(\theta_\mathrm{o})}. \tag{33}$$

The contour of the black hole shadow is described by the parametric curve $(\alpha(r), \beta(r))$.

3.4. *Observables*

We define three observables[10]: the area of the shadow, its oblateness, and the horizontal displacement of its centroid, as follows:

- The area is calculated from
$$A = 2\int \beta\, d\alpha = 2\int_{r_+}^{r_-} \beta(r)|\alpha'(r)|\, dr, \tag{34}$$
with the factor 2 arising from the up-down symmetry of the shadow and taking the plus sign in Eq. (33).
- The oblateness is related to the deformation of the shadow as compared to a circle
$$D = \frac{\Delta\alpha}{\Delta\beta}, \tag{35}$$
where $\Delta\alpha$ and $\Delta\beta$ are the horizontal and vertical extent of the shadow.
- The horizontal coordinate of the centroid is given by
$$\alpha_c = \frac{2}{A}\int \alpha\beta\, d\alpha = \frac{2}{A}\int_{r_+}^{r_-} \alpha(r)\beta(r)|\alpha'(r)|\, dr, \tag{36}$$
with the same factor 2 as in the definition of the area.

These observables allow a characterization of the shadow by only three numbers.

4. Example: Shapiro plasma

We consider, in order to provide an example, the well known model of plasma consisting of dust that is at rest at infinity, introduced by Shapiro.[20] In the Kerr spacetime, the mass density and the squared plasma frequency go as $r^{-3/2}$, and they are independent of θ to a very good approximation. However, for this plasma distribution the equations cannot be brought into a separable form; therefore, we take the frequency to have an additional θ dependency[13] by adopting $f_r(r) = \omega_c^2\sqrt{M^3 r}$ and $f_\theta(\theta) = 0$, resulting in a plasma electron frequency

$$\omega_p^2 = \omega_c^2 \frac{\sqrt{M^3 r}}{r^2 + a^2\cos^2\theta}, \tag{37}$$

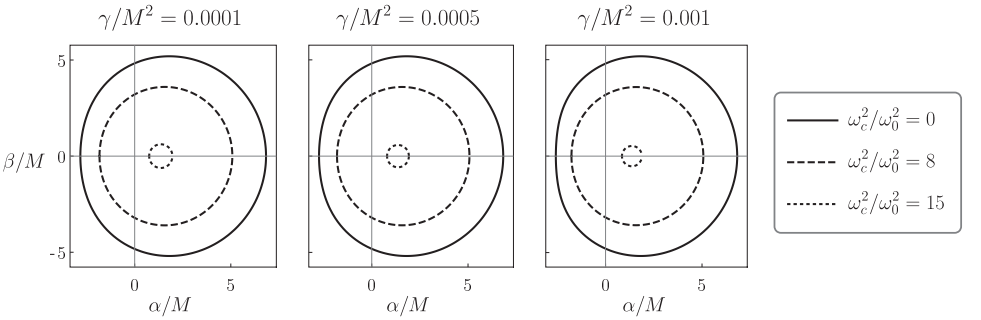


Fig. 1. Shadow of a $4D$ Einstein-Gauss-Bonnet black hole with spin $a/M = 0.9$ surrounded by a Shapiro-type plasma distribution with $f_r(r) = \omega_c^2\sqrt{M^3 r}$, for an equatorial observer.

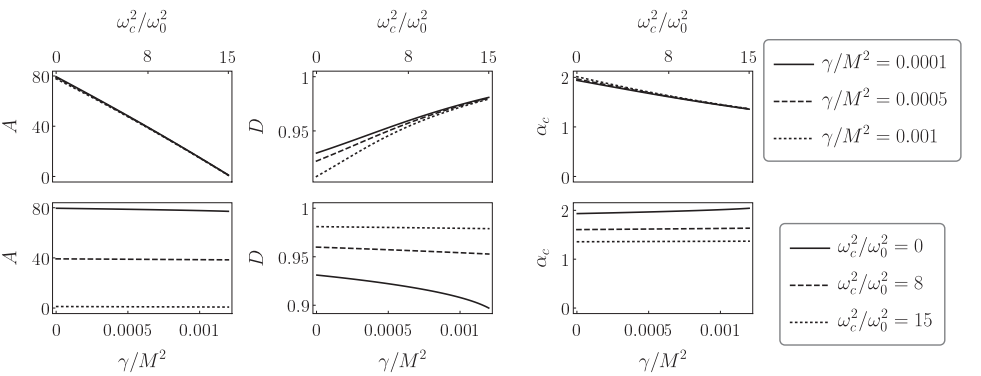


Fig. 2. The area (A), the oblateness (D), and the centroid (α_c) of the shadow. Top: the observables as functions of the frequency ratio ω_c/ω_0, for three values of the EGB parameter γM^2. Bottom: the observables as functions of γM^2, for three values of the frequency ratio ω_c/ω_0.

where ω_c is a constant that characterizes the fluid. At large distances from the black hole, the metric approaches the Kerr one, i.e. $m(r) \to M$, so we expect that the Shapiro model is still valid; but at small distances the metric may be quite different, so we assume that the plasma density is not significantly affected. The trajectories of photons with frequency ω_0 depend only on the ratio ω_c/ω_0. We present in Figs. 1 and 2 some plots of black hole shadows and the corresponding observables, for an equatorial observer, $a/M = 0.9$, and suitably chosen values of the other parameters. For a given value of the plasma frequency ω_c, we can see that:

- For fixed parameter γ/M^2, the shadow reduces in size, is less deformed, and has a smaller centroid displacement, as the photon frequency ω_0 decreases. In particular, the shadow disappears entirely below a certain photon frequency, due to the appearance of the forbidden region. This forbidden region starts as two caps around the poles and grows towards the equatorial plane as the frequency decreases, leading to a reduction of the shadow size.

- For a fixed photon frequency ω_0, the shadow slightly reduces in size, is less deformed, and the centroid displacement slightly grows, as γ/M^2 increases. The effect of increasing the parameter γ/M^2 is more prominent for high frequency photons, particularly in the case of the oblateness.

The existence of plasma surrounding the black hole results in a smaller and less deformed shadow than in vacuum, which corresponds to $\omega_c = 0$.

5. Final remarks

We have studied how the presence of plasma modifies the shadow corresponding to rotating black holes in the scalar-tensor 4-dimensional Einstein-Gauss-Bonnet theory with respect to the vacuum case. These black holes were obtained in previous works from spherically symmetric solutions by applying the Newman-Janis algorithm. We have neglected the gravitational influence of the plasma itself and any processes of scattering, emission or absorption. We have shown that the Hamilton-Jacobi equation for photons is separable if the plasma frequency obeys the usual separability condition introduced for the Kerr spacetime. Light follows timelike curves in the presence of plasma, resulting in a modification of the photon regions and frequency-dependent forbidden regions, where photons cannot travel. We have obtained the expressions for the celestial coordinates of the shadow contour as viewed by a far away observer, which reduce to the already known for vacuum when the plasma frequency goes to zero or the photon frequency to infinity. In order to characterize the size, the shape, and the position, we have presented three observables: the area, the oblateness, and the centroid of the shadow. With the intention to provide a concrete example, we have taken a variation of the Shapiro plasma distribution, which models dust at rest at infinity surrounding a black hole. We have considered an equatorial observer, for simplicity and also because the effects on the shadow in this case are more prominent. For a fixed the plasma frequency $\omega_c \neq 0$ and given values of the mass M and the parameter γ, the shadow becomes smaller and less deformed as the photon frequency ω_0 decreases, with the appearance of a forbidden region around the black hole as a distinctive feature. This forbidden region begins as two caps around the poles and grows towards the equatorial plane for decreasing ω_0, resulting in a sharp reduction of the shadow size and in its eventual disappearance. The presence of plasma always leads to a smaller and less deformed shadow than in the case of vacuum ($\omega_c = 0$). The work presented here is a particular case of the shadow cast by a class of rotating black holes surrounded by plasma, which results from applying the Newman-Janis algorithm to spherically symmetric spacetimes[15] determined by a mass function $m(r)$ satisfying that $m(r) \to M$ when $r \to \infty$.

Acknowledgments

This work has been supported by CONICET and Universidad de Buenos Aires.

References

1. D. Glavan and C. Lin, Einstein-Gauss-Bonnet Gravity in Four-Dimensional Spacetime, *Phys. Rev. Lett.* **124**, 081301 (2020).
2. M. Gurses, T. C. Sisman, and B. Tekin, Is there a novel Einstein–Gauss–Bonnet theory in four dimensions?, *Eur. Phys. J. C* **80**, 647 (2020).
3. P. G. S. Fernandes, P. Carrilho, T. Clifton, and D. J. Mulryne, Derivation of regularized field equations for the Einstein-Gauss-Bonnet theory in four dimensions, *Phys. Rev. D* **102**, 024025 (2020).
4. R. A. Hennigar, D .Kubizňák, R. B. Mann, and C. Pollack, On taking the $D \to 4$ limit of Gauss-Bonnet gravity: Theory and solutions, *JHEP* **07** (2020) 027.
5. J. M. Bardeen, in *Black Holes*, Eds. C. De Witt, B. S. De Witt, p. 215-239 (1973).
6. H. Falcke, F. Melia, and E. Agol, Viewing the Shadow of the Black Hole at the Galactic Center, *Astrophys. J.* **528**, L13 (2000).
7. K. Akiyama *et al.* (Event Horizon Telescope), First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, L1 (2019).
8. K. Akiyama *et al.* (Event Horizon Telescope), First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring, *Astrophys. J. Lett.* **875**, L5 (2019).
9. N. Tsukamoto, Black hole shadow in an asymptotically flat, stationary, and axisymmetric spacetime: The Kerr-Newman and rotating regular black holes, *Phys. Rev. D* **97**, 064021 (2018).
10. J. Badía and E. F. Eiroa, Influence of an anisotropic matter field on the shadow of a rotating black hole, *Phys. Rev. D* **102**, 024066 (2020).
11. H. C. D. Lima Junior, L. C. B. Crispino, P. V. P. Cunha, and C. A. R. Herdeiro, Spinning black holes with a separable Hamilton–Jacobi equation from a modified Newman–Janis algorithm, *Eur. Phys. J. C* **80**, 1036 (2020).
12. V. Perlick and O. Y. Tsupko, Calculating black hole shadows: review of analytical studies, arXiv:2105.07101 [gr-qc].
13. V. Perlick and O. Y. Tsupko, Light propagation in a plasma on Kerr spacetime: Separation of the Hamilton-Jacobi equation and calculation of the shadow, *Phys. Rev. D* **95**, 104003 (2017).
14. H. Yan, Influence of a plasma on the observational signature of a high-spin Kerr black hole, *Phys. Rev. D* **99**, 084050 (2019).
15. J. Badía and E. F. Eiroa, Shadow of axisymmetric, stationary and asymptotically flat black holes in the presence of plasma, *Phys. Rev. D* **104**, 084055 (2021).
16. R. Kumar and S. G. Ghosh, Rotating black holes in 4D Einstein-Gauss-Bonnet gravity and its shadow, *JCAP* **07** (2020) 053.
17. S. Wei and Y. Liu, Testing the nature of Gauss–Bonnet gravity by four-dimensional rotating black hole shadow, *Eur. Phys. J. Plus* **136**, 436 (2021).
18. J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).
19. B. Carter, Global Structure of the Kerr Family of Gravitational Fields, *Phys. Rev.* **174**, 1559 (1968).
20. S. Shapiro, Accretion onto black holes: The emergent radiation spectrum. III. Rotating (Kerr) black holes, *Astrophys. J.* **189**, 343 (1974).

Aspects of neutrino mass hierarchy in gravitational lensing

Himanshu Swami

Department of Physical Sciences,
Indian Institute of Science Education & Research (IISER) Mohali,
SAS Nagar, Manauli-140306, Punjab, India
E-mail: himanshuswami@iisermohali.ac.in

We study neutrino flavour oscillations in curved spacetime especially emphasising neutrino lensing in Schwarzschild spacetime, and highlight the role that neutrinos lensing induced by a gravitational source can play in inferring neutrino absolute mass and mass hierarchy. Further, the wave packet approach in neutrino lensing and its modifications compared with the plane wave approach is analysed. Finally, we discuss the decoherence effect in the wave packet approach and its possible use in inferring absolute neutrino mass and neutrino mixing parameters.

Keywords: Neutrino mass hierarchy; gravitational lensing.

1. Neutrino oscillations in flat and curved spacetime

It is now clear from experiments that neutrinos have masses, and that they mix and neutrino flavour eigenstates $|\nu_\alpha\rangle$ are now related to their mass eigenstates $|\nu_i\rangle$ by 3×3 unitary matrix[1–3] U

$$|\nu_\alpha\rangle = \sum_i U^*_{\alpha i} |\nu_i\rangle, \tag{1}$$

where $\alpha = e, \mu, \tau$ and $i = 1, 2, 3$. A neutrino flavour state propagation from source S to detector D, located at $\mathbf{x}_S$ and $\mathbf{x}_D$ respectively, is described by

$$|\nu_\alpha\left(t_D, \mathbf{x}_D\right)\rangle = N \sum_i U^*_{\alpha i} \psi_i\left(x_D, x_S\right) |\nu_i\left(t_S, \mathbf{x}_S\right)\rangle\,, \tag{2}$$

where N is a normalisation factor. $\psi_i\left(x_D, x_S\right)$ is the wavefunction, evolved between the time of production (t_S) and detection (t_D), for i^{th} massive neutrino . Assuming neutrino wave-function as a plane wave, the flavour transition probability from $\nu_\alpha \to \nu_\beta$ at the detection point is given by

$$\mathcal{P}_{\alpha\beta} \equiv \left|\langle\nu_\beta|\nu_\alpha(t_D, \mathbf{x}_D)\rangle\right|^2 = \sum_{i,j} U_{\beta i} U^*_{\beta j} U_{\alpha j} U^*_{\alpha i}\, \exp(-i(\Phi_i - \Phi_j))\,. \tag{3}$$

Now, neutrino mass eigenstates develop different phases Φ_i, which gives rise to neutrino oscillation phenomena.[4] Massive neutrino phase in flat spacetime is given by

$$\Phi_i = E_i(t_D - t_S) - \mathbf{p}_i \cdot (\mathbf{x}_D - \mathbf{x}_S)\,. \tag{4}$$

Assuming that all the mass eigenstates in a flavour eigenstate initially produced have equal momentum or energy.[4,5] This assumption together with $(t_D - t_S) \simeq |\mathbf{x}_D - \mathbf{x}_S|$

for relativistic neutrinos ($E_i \gg m_i$) gives

$$\Delta\Phi_{ij} \equiv \Phi_i - \Phi_j \simeq \frac{\Delta m_{ij}^2}{2E_0} \left|\mathbf{x}_D - \mathbf{x}_S\right|, \tag{5}$$

where $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$. E_0 is the average energy of the relativistic neutrinos produced at S. Substitution of Eq. (5) in Eq. (3) for two flavours of neutrinos lead to the following oscillation formula:

$$\mathcal{P}_{e\mu} = \sin^2 2\alpha \, \sin^2 \left(\frac{\Delta m_{12}^2 \, L}{4E_0}\right) \tag{6}$$

where $L = |\mathbf{x}_D - \mathbf{x}_S|$ and the angle α parameterize elements of matrix U. We see a constant universal shift in the squared masses, i.e. $m_i^2 \to m_i^2 + C$, leaves the expression of $\mathcal{P}_{\alpha\beta}$ unchanged, therefore, $\mathcal{P}_{\alpha\beta}$ depends on the difference of squared masses and not on the absolute masses. In a curved spacetime, the phase in Eq. (4) can be replaced by its covariant form[6]

$$\Phi_i = \int_S^D p_\mu^{(i)} \, dx^\mu \,, \tag{7}$$

where $p_\mu^{(i)}$ is the canonical conjugate momentum to the coordinate x^μ for the i^{th} neutrino mass eigenstate. As in a general curved spacetime, more than one different geodesics starting from the same spacetime point can intersect later, therefore, accounting for all neutrino paths (m) a neutrino flavour state is now given by

$$\left|\nu_\alpha \left(t_D, \mathbf{x}_D\right)\right\rangle = \sum_i U_{\alpha i}^* \sum_m N \psi_i^m \left(x_D, x_S\right) \left|\nu_i \left(t_S, \mathbf{x}_S\right)\right\rangle . \tag{8}$$

Now, normalisation (N), in general, can depend on the path m.

2. Neutrino lensing in Schwarzschild spacetime

We now examine neutrino oscillations in Schwarzschild spacetime specifically emphasizing on the neutrino lensing. The line element in Schwarzschild metric is given by

$$ds^2 = B\left(r\right) dt^2 - \frac{1}{B\left(r\right)} dr^2 - r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right), \tag{9}$$

where $B\left(r\right) = 1 - \frac{R_s}{r}$, and R_s is Schwarzschild radius. With no lose of generality, we will take $\theta = \pi/2$. Further, it is assumed that the neutrinos are produced as a plane wave.

In the weak gravity limit, there are two paths denoted by impact parameters b_1 and b_2 that neutrino may take to reach at the detector location O. Neutrino flavor transtion probability from α flavour (produced at the source) to β flavour (detected at the detector location) is, then, given by[7]

$$\begin{aligned}\mathcal{P}_{\alpha\beta}^{\text{lens}} = |N|^2 \Bigg(& 2 \sum_i |U_{\beta i}|^2 |U_{\alpha i}|^2 (1 + \cos(\Delta b^2 B_{ii})) + \sum_{i,j\neq i} U_{\beta i} U_{\beta j}^* U_{\alpha j} U_{\alpha i}^* \\ & \times \big(\exp(-i\Delta m_{ij}^2 A_{11}) + \exp(-i\Delta m_{ij}^2 A_{22}) + 2\cos\left(\Delta b^2 B_{ij}\right) \\ & \times \exp\left(-i\Delta m_{ij}^2 A_{12}\right) \big) \Bigg), \end{aligned} \tag{10}$$

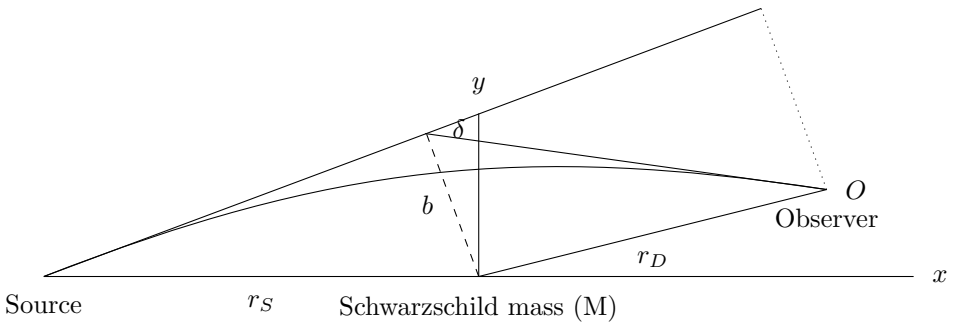


Fig. 1. Diagrammatic representation of weak lensing of neutrinos. Neutrinos are created at the source (S) and get lensed by a Schwarzschild mass (M) and detected at location (O).

with

$$|N|^2 = \left(2 + 2\sum_i |U_{\alpha i}|^2 \cos(\Delta b^2 B_{ii})\right)^{-1} \tag{11}$$

where $\Delta b^2_{12} \equiv \Delta b^2$. It is noted that many similar versions of Eq. (10) are available in literature,[9–11] however, with some subtle differences. Though, the role of normalisation (Eq. (11)) has been neglected in literature[9,11] which is crucial in understanding the neutrino oscillation interference effects.

The change in oscillation phase calculated along neutrino massive trajectory for path (m) is given by

$$\Phi_j^{massive} \simeq \frac{m_j^2}{E_0}(r_S + r_D)\left(1 - \frac{b_m^2}{2r_S r_D} + \frac{2GM}{r_S + r_D}\right), \tag{12}$$

where b_m is the impact parameter and related with deflection angle (δ) as $\delta = -2R_s/b + O\left(R_s m_j^2/bE_0^2\right)$.[8] However, the phase change if calculated along null trajectory for path (m) gives[9]

$$\Phi_j^{null} \simeq \frac{m_j^2}{2E_0}(r_S + r_D)\left(1 - \frac{b_m^2}{2r_S r_D} + \frac{2GM}{r_S + r_D}\right), \tag{13}$$

which is half of $\Phi_j^{massive}$. Despite of this difference in oscillation phase, the bulk of literature employs null ray approximations to calculate the change in the neutrino phase. So, to keep in tune with the bulk of the existing literature on neutrino oscillation, we will employ null ray approximation to study weak lensing. Nevertheless, the results of lensing remain qualitatively consistent if one uses massive trajectories instead of null trajectories. The other quantities in Eq. (10) are related to the phase difference $\Delta\Phi_{ij}^{mn}$ as[7]

$$\Delta\Phi_{ij}^{mn} = \Phi_i^m - \Phi_j^n = \Delta m_{ij}^2\, A_{mn} + \Delta b_{mn}^2\, B_{ij}\,, \tag{14}$$

where

$$A_{mn} = \frac{(r_S + r_D)}{2E_0}\left(1 + \frac{2GM}{r_S + r_D} - \frac{\sum b_{mn}^2}{4r_S r_D}\right),$$

$$B_{ij} = -\frac{\sum m_{ij}^2}{8E_0}\left(\frac{1}{r_S} + \frac{1}{r_D}\right). \tag{15}$$

Here, $\sum b_{mn}^2 = b_m^2 + b_n^2$ and $\sum m_{ij}^2 = m_i^2 + m_j^2$. We note that unless mixing angle $\alpha = \pi/4$ (in the two flavour case) (1) normalization Eq. (11) is not invariant under $m_i \leftrightarrow m_j$ due to $|U_{\alpha i}|^2$ and (2) the neutrino phase Eq. (13) is not invariant under a universal shift $m_i^2 + C$ due to B_{ij} term, unless $\Delta b_{mn}^2 \equiv b_m^2 - b_n^2$ vanishes. Hence, transition probability will be sensitive to both aforementioned points. Fig. 2 shows flavour transition probability for the lensed neutrino, which are lensed by the Sun and detected on the Earth. The Earth trajectory is taken to be circular in the plots. From Fig. 2 one can see the difference in transition probability as the value of mass m_1 changes. Also, the change in mass ordering gives quite a significant difference in transition probabilities and oscillation frequency. We observe that neutrino oscillation frequency increases as we increase the value of m_1. Therefore, one can utilise neutrino lensing to infer neutrino mass. One such way to extract information about absolute mass is to find the flipping point (ϕ_D^e), i.e., where all flavour changing probabilities become the same. The flipping angle in the two flavour case is given by[7]

$$\sum m^2 = \frac{16n\pi E_0 r_S r_D}{Z\Delta b^2} + \frac{16E_0 r_S r_D}{Z\Delta b^2}\cot^{-1}\left(\tan\left(\frac{\Delta m^2 Z\Delta b^2}{16E_0 r_S r_D}\right)\right.$$
$$\left.\times\, G\left(r_S, r_D, \phi_D^e, \alpha, \Delta m^2, R_s\right)\right), \tag{16}$$

$$G_{\pm}\left(r_S, r_D, \phi_D^e, \alpha, \Delta m^2, R_s\right)$$
$$\equiv \frac{2\cos 2\alpha\cos^2\alpha - \sin^2 2\alpha\cos\zeta \pm \sqrt{\sin^4 2\alpha\cos^2\zeta + \sin^2 2\alpha\cos^2 2\alpha}}{2\cos 2\alpha\cos^2\alpha + \sin^2 2\alpha\cos\zeta \mp \sqrt{\sin^4 2\alpha\cos^2\zeta + \sin^2 2\alpha\cos^2 2\alpha}}, \tag{17}$$

where $\zeta \equiv \Delta m^2\left(r_S + r_D + R_s - (r_S + r_D)\sum b^2/4r_S r_D\right)/2E_0$, $Z = r_S + r_D$. In the flipping equation (16), all the measurable quantities are on the right-hand side of the equation, and the term related to absolute mass is on the left-hand side.

3. Neutrino decoherence in gravitational lensing

In a realistic setup neutrinos are produced and detected as a wave packet. So writing detector and neutrino source states in momentum space by using non coordinate tetrad basis,[12] we arrive at the amplitude of flavour transition probability in the

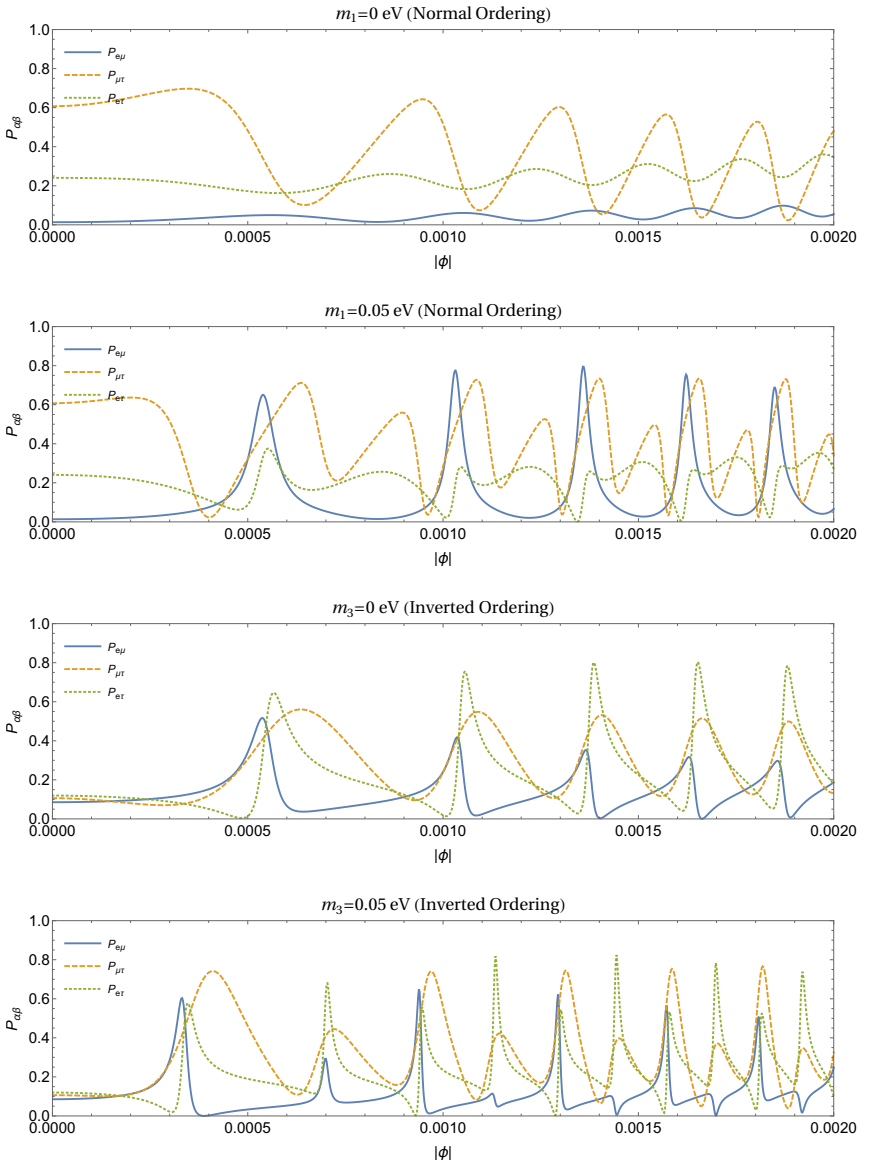


Fig. 2. Probability of $\nu_\alpha \to \nu_\beta$ conversion, for different α and β, as function of azimuthal angle ϕ in three flavour case. We take $R_s = 3$ km, $r_D = 10^8$ km, $r_S = 10^5 r_D$ and $E_0 = 10$ MeV. Neutrino mass squared differences (Δm^2_{ij}), mixing angle and the Dirac CP phase values are taken from (NuFIT 4.1 (2019)) global fit.[3]

wave packet picture and is given by[4, 13, 14]

$$\mathcal{A}^m_{\alpha\beta} \equiv \langle \nu_\beta(\vec{x}_D)|\nu_\alpha(t,\vec{x})\rangle = \sum_i U_{\beta i} U^*_{\alpha i} \int \frac{d^3p}{(2\pi)^3} f^{D*}_{i,\vec{x}}\left(\vec{p},\vec{p}^{\,D}_i\right) f^S_{i,\vec{x}}\left(\vec{p},\vec{p}^{\,S}_i\right) e^{-i\Phi^m_i} \,. \tag{18}$$

The probability of transition $\nu_\alpha \to \nu_\beta$, then, can be computed from the amplitude using[7,14]

$$P_{\alpha\beta} = \frac{\left|\sum_m \mathcal{A}^m_{\alpha\beta}\right|^2}{\sum_\beta \left|\sum_m \mathcal{A}^m_{\alpha\beta}\right|^2} \,. \tag{19}$$

Here, we assume the detection process is time-independent[4] to arrive at the Eq. (18). Further, taking $f^{S,D}_{i,\vec{x}}\left(\vec{p},\vec{p}^{\,S,D}_i\right)$ as Guassian which has a sharp peak around $\vec{p}^{\,S}_i$. We, then, substitute Φ^m_i with its series expansion at $\vec{p}=\vec{p}^{\,S}_i$ in Eq. (18)

$$\Phi^m_i(\vec{p}) = \Phi^m_i(\vec{p}^{\,S}_i) \,+\, \vec{X}^m_i \cdot (\vec{p}-\vec{p}^{\,S}_i) \,+\, \mathcal{O}(p^2)\,, \tag{20}$$

where $\vec{X}^m_i = \partial_{\vec{p}}\Phi^m_i(\vec{p}=\vec{p}^{\,S}_i)$. Furthermore, for simplification, we take $\vec{p}^{\,D,S}_i = \vec{p}^{\,D,S}$ and $\sigma_{i\,D,S} = \sigma_{D,S}$, to arrive at following transition amplitude[14] $\mathcal{A}^m_{\alpha\beta}$ and transition probability[14] $P_{\alpha\beta}$

$$\mathcal{A}^m_{\alpha\beta} = 2\sqrt{2}\left(\frac{\sigma_D\sigma_S}{\sigma_D^2+\sigma_S^2}\right)^{3/2} \sum_i U_{\beta i}U^*_{\alpha i}\, e^{-i\left(\Phi^m_i - \frac{\sigma_S^2}{\sigma_D^2+\sigma_S^2}\vec{X}^m_i\cdot\left(\vec{p}^{\,D}-\vec{p}^{\,S}\right)\right)} \times e^{-\frac{\sigma_D^2\sigma_S^2|\vec{X}^m_i|^2+\left(\vec{p}^{\,D}-\vec{p}^{\,S}\right)^2}{2\left(\sigma_D^2+\sigma_S^2\right)}}\,, \tag{21}$$

$$P_{\alpha\beta} = \frac{\sum\limits_{i,j} U^*_{\beta i}U_{\alpha i}U_{\beta j}U^*_{\alpha j}\sum\limits_{m,n} e^{-i\Phi^{mn}_{ij}}\, e^{-D^{mn}_{ij}}}{\sum\limits_i U_{\alpha i}U^*_{\alpha i}\sum\limits_{m,n} e^{-i\Phi^{mn}_{ii}}\, e^{-D^{mn}_{ii}}}\,, \tag{22}$$

where

$$\Phi^{mn}_{ij} \equiv \left(\Phi^m_i - \Phi^n_j\right) - \frac{\bar{\sigma}^2}{\sigma_D^2}\left(\vec{p}^{\,D}-\vec{p}^{\,S}\right)\cdot\left(\vec{X}^m_i - \vec{X}^n_j\right)\,, \tag{23}$$

$$D^{mn}_{ij} = \mathbf{X}^{mn}_{ij} - \mathbf{X}^{\hat{m}\hat{n}}_{\hat{i}\hat{i}}\,, \tag{24}$$

$$\mathbf{X}^{mn}_{ij} \equiv \frac{1}{2}\bar{\sigma}^2\left(|\vec{X}^m_i|^2 + |\vec{X}^n_j|^2\right)\,, \tag{25}$$

where $\bar{\sigma}^2 = \sigma_D^2\sigma_S^2/(\sigma_D^2+\sigma_S^2)$. From Eq. (22), we define D^{mn}_{ij} to be decoherence factor. In D^{mn}_{ij} (Eq. (24)), the term $\mathbf{X}^{\hat{m}\hat{n}}_{\hat{i}\hat{i}}$ is choosen be the the smallest among $\mathbf{X}^{mn}_{ij}$, such that D^{mn}_{ij} are, by construction, greater than or equal to zero. We note that as D^{mn}_{ij} get large, probabilities approach to values which only depend on neutrino mixing matrix elements, i.e., $P_{\alpha\beta} \to U^*_{\beta 1}U_{\beta 1} = |U_{\beta 1}|^2$. We, also, note that the oscillation phase obtained in Eq. (23) is modified from the one obtained assuming neutrinos as plane waves. Now, oscillation phase also depends on the source and the detector wave profile. The difference between plane wave phase and modified phase become negligible if the neutrino wave packets at production and detection follow $\vec{p}^{\,D} \simeq \vec{p}^{\,S}$ condition. Now considering the Schwarzschild neutrino lensing, taking $b_2 < b_1$ and $m_1 < m_2 < m_3$, one finds the term D^{mn}_{ij}, which approaches last to zero,

is given by D_{12}^{11}. Hence, the coordinate distance at which the transition probability gets suppressed by at least a factor of e^{-1} is given by following condition[14]

$$(r_D + r_S)\left(1 - \frac{b_1^2}{2r_S r_D} + \frac{R_S}{r_D + r_S}\right) \geq 2\sqrt{\frac{2}{B(r_S)}}\frac{E_0^2}{\bar{\sigma}\sqrt{m_2^4 - m_1^4}}\,. \tag{26}$$

Eq. (26) can readily be used to estimate decoherence length, i.e., the distance to which the coherent oscillations last. For the radial outward propagation in the flat and Schwarzschild spacetime, respectively, with the same dilution in probability e^{-1} and written in physical length, is given by

$$L_p^F = r_D - r_S \geq 2\sqrt{2}\frac{E_0^2}{\bar{\sigma}\sqrt{m_2^4 - m_1^4}}, \tag{27}$$

and

$$L_p^S - R_s \ln\frac{r_D}{r_S} \geq 2\sqrt{\frac{2}{B(r_S)}}\frac{E_0^2}{\bar{\sigma}\sqrt{m_2^4 - m_1^4}}\,. \tag{28}$$

From Eq. (27) and Eq. (28) we observe for the same physical distance $L_p^F = L_p^S$ between the source and the detector, neutrino decoher more in the flat spacetime, whereas due to R_s term neutrino decoher less in the Schwarzschild spacetime. This is true even for the non radial and the lensing case. Also, from Eq. (26), Eq. (27) and Eq. (28), we see that rate of dilution depend not only on Δm^2, but also on $\sum m^2$ term. Hence, length at which oscillations cease to exist due to suppression of all terms in transition probability can be used to infer absolute neutrino masses. To make above point clear we take two neutrino flavour case and plot change in transition probability taking $y_D = 0$ and varying only x_D, also now we have $x_D = r_D$. Fig. 4 shows maximum and minimum probability as a function of r_D near dechorence region, we see for $m_1 = 0$ eV decoherence occurs, i.e. $P_{e\mu} \to \sin^2\alpha = 1/2$, later compares to the $m_1 = 0.1$ eV case. We also note from Fig. 3, that decoherence is consistent with $D_{12}^{11} \geq 2$. Right panel of Fig. 3 compute correlations between $\bar{\sigma}$ and r_D for $D_{12}^{11} = 1$. The estimation is done for two different values of the lightest neutrino mass m_1 keeping $m_2^2 - m_1^2 = 10^{-3}eV^2$ fixed.

4. Summary

By studying neutrino lensing in Schwarzschild background, we explicated that neutrino lensing is sensitive to neutrino mass ordering and absolute neutrino masses, contrary to the vacuum neutrino flavour oscillations in flat spacetime. In curved spacetime, neutrinos can take more than one path to arrive at the detector; this gives rise to path difference in the neutrino phase, giving a novel contribution in oscillation probabilities that depends on absolute neutrinos mass. We presented a method for inferring individual neutrino mass from flipping angles/points, i.e. locations in the space where all the neutrino flavours have equal probabilities of being detected by the detector. Using the Sun-Earth model, we explicitly demonstrate

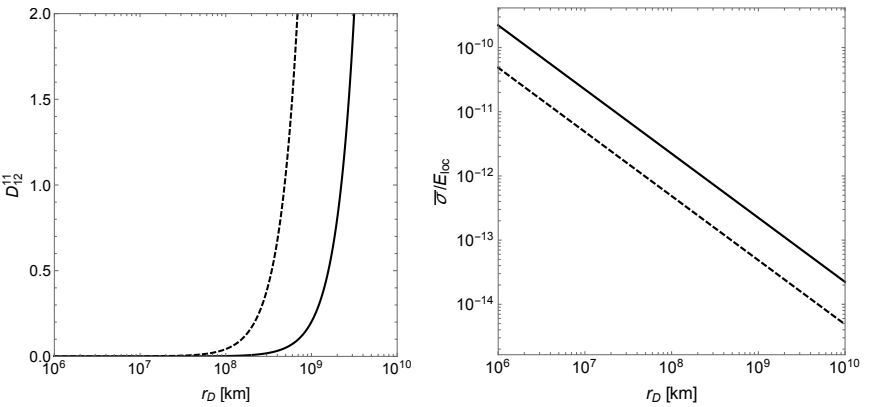


Fig. 3. Left panel: the damping factor D_{12}^{11} as function of r_D for $\bar{\sigma}/E_{\rm loc} = 10^{-13}$. Right panel: contours corresponding to $D_{12}^{11} = 1$. In both the panels, the solid (dashed) line corresponds to $m_1 = 0$ ($m_1 = 0.1$) eV. The other parameters are $r_S = 10^5 r_D$, $R_S = 3$ km, $m_2^2 - m_1^2 = 10^{-3}\,\text{eV}^2$ and $E_{\rm loc} = 10$ MeV.

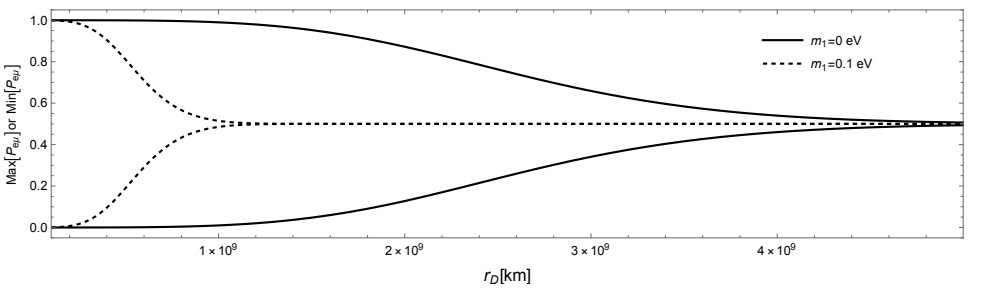


Fig. 4. Maximum and minimum transition probability envelop as a function of r_D for two flavour case. The solid (dashed) line corresponds to $m_1 = 0$ ($m_1 = 0.1$) eV. The mixing angle is $\alpha = \pi/4$ and all the other parameters are as given in the caption of Fig. 3.

oscillation sensitivity on individual neutrino masses in neutrino lensing. Neutrino oscillation frequency increases as we increase the value of m_1. Further, using the Gaussian wavepacket profile for the neutrino and the detector, we find wave packet profile modifies the neutrino phase. Now, the oscillation phase depends on the source and detector wave packet profile $\sigma_{D,S}$ and $\vec{p}^{D,S}$, and reduces to the plane wave neutrino phase when mean momentum $\vec{p}^{D,S}$ of the source and detector are same. The wave packet approach also introduces a new length scale, i.e. decoherence length, after which neutrino oscillations cease to exist. The decoherence length depends not only on Δm^2 but also on $\sum m^2$. Also, in a weak gravity limit, neutrino travels more in the Schwarzschild spacetime as compare to the flat spacetime before decoupling due to decoherence. Further, we highlight that neutrino flavour transition probability settles to the fundamental values of neutrino mixing matrix elements after decoherence. Therefore, the sensitivity of flavour oscillations probabilities and neu-

trino decoherence on $\sum m^2$ can reveal absolute neutrino mass and their ordering. Hence, neutrino lensing can be one of the viable methods for inferring individual neutrino mass. Conversely, one can also use decoherence length and neutrino mass to infer values of neutrino mixing parameters.

References

1. F. Capozzi, G. L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, "Status of three- neutrino oscillation parameters, circa 2013," *Phys. Rev.* **D89**, 093018 (2014), https://arxiv.org/abs/1312.2878.
2. P. F. de Salas, D. V. Forero, C. A. Ternes, M. Tortola, and J. W. F. Valle, "Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity," *Phys. Lett.* **B782**, 633-640 (2018), https://arxiv.org/abs/1708.01186.
3. Ivan Esteban, M. C. Gonzalez-Garcia, Alvaro Hernandez-Cabezudo, Michele Maltoni, and Thomas Schwetz, "Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ_{23}, δ_{CP}, and the mass ordering," *JHEP* **01**, 106 (2019), https://arxiv.org/abs/1811.05487.
4. Evgeny Kh. Akhmedov and Alexei Yu. Smirnov, "Paradoxes of neutrino oscillations," *Phys. Atom. Nucl.* **72**, 1363-1381 (2009), https://arxiv.org/abs/0905.1903.
5. E. Kh. Akhmedov and A. Yu. Smirnov, "Neutrino oscillations: Entanglement, energy-momentum conservation and QFT," *Found. Phys.* **49**, 1279-1306 (2011), https://arxiv.org/abs/1008.2077.
6. Christian Y. Cardall and George M. Fuller, "Neutrino oscillations in curved space-time: An Heuristic treatment," *Phys. Rev.* **D55**, 7960-7966 (1997), https://arxiv.org/abs/hep-ph/9610494.
7. Himanshu Swami, Kinjalk Lochan, and Ketan M. Patel, "Signature of neutrino mass hierarchy in gravitational lensing," *Phys. Rev.* **D102**, 024043 (2020), https://arxiv.org/abs/2002.00977.
8. Thanu Padmanabhan, "Gravitation: Foundations and Frontiers," *Cambridge University Press*, ISBN:978-0-521-88223-1 (2010).
9. N. Fornengo, C. Giunti, C. W. Kim, and J. Song, "Gravitational effects on the neutrino oscillation," *Phys. Rev.* **D56**, 1895-1902 (1997), https://arxiv.org/abs/hep-ph/9611231.
10. Roland M. Crocker, Carlo Giunti, and Daniel J. Mortlock, "Neutrino interferometry in curved spacetime," *Phys. Rev.* **D69**, 063008 (2004), https://arxiv.org/abs/hep-ph/0308168.
11. Jean Alexandre and Katy Clough, "Black hole interference patterns in flavor oscillations," *Phys. Rev.* **D98**, 043004 (2018), https://arxiv.org/abs/1805.01874.
12. E. Mitsou and J. Yoo, "Tetrad Formalism for Exact Cosmological Observables," SpringerBriefs in Physics (Springer, 2020), https://arxiv.org/abs/1908.10757.
13. E. K. Akhmedov and J. Kopp, "Neutrino oscillations: Quantum mechanics vs. quantum field theory," *JHEP* **04**, 008 (2010), [Erratum: JHEP 10, 052 (2013)], https://arxiv.org/abs/1001.4815.
14. Himanshu Swami, Kinjalk Lochan, and Ketan M. Patel, "Aspects of gravitational decoherence in neutrino lensing," https://arxiv.org/abs/2106.07671.

Photon regions in stationary axisymmetric spacetimes and umbilic conditions

K. V. Kobialko* and D. V. Gal'tsov†

Faculty of Physics, Moscow State University, 119899, Moscow, Russia

E-mail: kobyalkokv@yandex.ru

†*E-mail: galtsov@phys.msu.ru*

We present the fundamentals of the recently proposed geometric description[1] of *photon regions* in terms of foliation into *fundamental photon hypersurfaces*, which satisfies the umbilic condition for the subbundle of the tangent bundle defined by the generalized impact parameter.

Keywords: Photon regions, photon surfaces, gravitational lensing, black hole shadows.

1. Introduction

Formation of shadows and relativistic images of stationary black holes is closely related to *photon regions*,[2,3] which are defined as compact domains where photons can travel endlessly without escaping to infinity or disappearing at the event horizon. Indeed, the boundary of the gravitational shadow corresponds to the set of light rays that inspiral asymptotically onto the part of the spherical surfaces in photon regions on which closed spherical photon orbits are located.[4–6]

Spherical surfaces in the photon region are just as important for determining the shadow of a stationary black hole as the *photon surfaces*[7,8] in the static case[a]. Recall that an important property of the photon surfaces is established by the theorem asserting that these are timelike *totally umbilic* hypersurfaces S in spacetime. This means that their second fundamental form II[11] is proportional to the induced metric:

$$\mathrm{II}(X,Y) = \mathrm{H}\,\langle X,Y\rangle\,, \quad \forall X,Y \in TS. \tag{1}$$

This property can serve as a constructive definition for analyzing photon surfaces instead of solving geodesic equations. It is especially useful in the cases when the geodesic equations are non-separable, and their analytic solution can not be found.[12–17]

However, in rotating spacetime such as Kerr, spherical surfaces in the photon region do not fully satisfy the umbilic condition and have a boundary. Such surfaces usually form a family, parameterized by the value of the azimuthal impact parameter $\rho = L/E$, where L, E are the integrals of motion corresponding to the timelike and azimuthal Killing vector fields.[18,19] To describe these surfaces and the photon

[a]For recent review of strong *gravitational lensing* and shadows see.[6,9,10]

region geometrically, we introduce the concept of *partially umbilic* submanifolds that weaken the condition (1). Namely, it is possible to impose the condition (1) not on *all* vectors from the tangent space TS, but only on some subset of TS specified by the azimuthal impact parameter. In addition, we must specify the boundary conditions for the submanifolds so that the photon does not escape through them. Together, this leads to the definition of *fundamental photon submanifolds*[1,20] - as generalization of *fundamental photon orbits*.[17] The slices consisting of the fundamental photon surfaces form generalized photon regions.

In this article, we give a concise presentation of the main concepts of the geometric approach to fundamental photon submanifolds and regions. Section 2 describes the partition of the tangent space of the manifold into sectors specified by the azimuthal impact parameter of the geodesics $\rho = L/E$. Then, in the section 3, we introduce the notion of partially umbilic submanifolds on the ρ-constrained sector of the tangent space and define the *fundamental photon submanifolds*. The Section 4 contains the geometric definition of the photon region.

2. Geodesic classes

Let M be a m dimensional Lorentzian manifold[11] with scalar product $\langle \cdot , \cdot \rangle$, Levi-Civita connection ∇, a tangent bundle TM and supposed to possess two commuting Killing vector fields $\mathcal{K}_\alpha$ ($\alpha = t, \varphi$) defining a stationary axisymmetric spacetime such that $\mathcal{G} = \det(\mathcal{G}_{\alpha\beta}) < 0$, where $\mathcal{G}_{\alpha\beta} = \langle \mathcal{K}_\alpha, \mathcal{K}_\beta \rangle$.

Let us define a Killing vector field $\hat{\rho} \in \{\mathcal{K}_\alpha\}$[20] with index numbering Killing vectors fields of the frame $\{\mathcal{K}_\alpha\}$

$$\hat{\rho} = \rho^\alpha \mathcal{K}_\alpha, \quad \rho^\alpha = (\rho, 1), \tag{2}$$

which is determined by arbitrary constant parameter ρ. In addition, we will introduce a vector field $\hat{\tau}$ in $\{\mathcal{K}_\alpha\}$ orthogonal to $\hat{\rho}$:

$$\tau^\alpha = \mathcal{G}^{\alpha\lambda}\epsilon_{\lambda\beta}\rho^\beta, \qquad \langle \hat{\tau}, \hat{\tau} \rangle = -\langle \hat{\rho}, \hat{\rho} \rangle, \qquad \langle \hat{\tau}, \hat{\rho} \rangle = 0, \tag{3}$$

where $\epsilon_{\lambda\beta}$ is the two-dimensional Levi-Civita tensor.

Then there is frame $\{\hat{\tau}, e_a\} \cong \hat{\rho}^\perp$, such that $\{e_a\}$ ($a = 1, m-2$) is a frame in euclidean orthogonal complement $\{\mathcal{K}_\alpha\}^\perp$[b].

Let γ be some geodesic on M, and $\dot{\gamma}$ denotes the tangent vector field to γ. Then there is the following general relationship between geodesics γ and orthogonal complement $\hat{\rho}^\perp$.

Proposition 2.1. *At every point $p \in M$ there is a one-to-one correspondence*[c] *between geodesics γ (with nonzero energy $E \equiv -\langle \mathcal{K}_t, \dot{\gamma} \rangle \neq 0$) with impact parameter $\rho = -\langle \mathcal{K}_\varphi, \dot{\gamma} \rangle / \langle \mathcal{K}_t, \dot{\gamma} \rangle$ and tangent vector fields $\mathcal{X} \in \hat{\rho}^\perp$ with $\langle \mathcal{K}_t, \mathcal{X} \rangle \neq 0$.*

[b] Orthogonal complement $^\perp$ defined in.[11] If $\langle \hat{\rho}, \hat{\rho} \rangle = 0$, then $\hat{\tau}$ and $\hat{\rho}$ are simply proportional and the orthogonal complement $\hat{\rho}^\perp$ will contain only one null vector $\hat{\rho}$ and spacelike e_a.

[c] Accurate to the geodesics reparametrization.

Proof. If geodesic γ has ρ as an impact parameter, then $\rho = -\langle \mathcal{K}_\varphi, \dot\gamma\rangle / \langle \mathcal{K}_t, \dot\gamma\rangle$ and multiplying by $\langle \mathcal{K}_t, \dot\gamma\rangle \neq 0$ we get $0 = \langle \rho\mathcal{K}_t + \mathcal{K}_\varphi, \dot\gamma\rangle = \langle \hat\rho, \dot\gamma\rangle$ and therefore $\dot\gamma \in \hat\rho^\perp$. If $\langle \hat\rho, \mathcal{X}\rangle = 0$ and $\langle \mathcal{K}_t, \mathcal{X}\rangle \neq 0$, then at any point $p \in M$ the any vector $\mathcal{X}|_p$ is a tangent vector to some geodesic γ, which always exists and unique at least in some vicinity of $p \in M$ as solution of ODE with initial conditions $\gamma(0) = p$ and $\dot\gamma(0) = \mathcal{X}|_p$. The geodesic γ has ρ as an impact parameter insofar as $0 = \langle \hat\rho, \dot\gamma(0)\rangle = \langle \rho\mathcal{K}_t + \mathcal{K}_\varphi, \dot\gamma(0)\rangle$ and we get $\rho = -\langle \mathcal{K}_\varphi, \dot\gamma(0)\rangle / \langle \mathcal{K}_t, \dot\gamma(0)\rangle$. □

It is clear that in the general case the Killing vector field $\hat\rho$ can be timelike on the some part of the manifold M. In this case, its orthogonal complement $\hat\rho^\perp$ will not have the Lorentzian signature everywhere, and therefore not all manifolds will be available for null geodesics with a given impact parameter ρ. So, our goal is to find the suitable region $\mathcal{C} \subset M$ in the original manifold M such that $\langle \hat\rho, \hat\rho\rangle|_{\mathcal{C}} \geq 0$.

Proposition 2.2. *If $\nabla_{\hat\rho}\hat\rho \neq 0$ for all null $\hat\rho$, the smooth function $\langle \hat\rho, \hat\rho\rangle$ defines m dimensional manifold with boundary*

$$\mathcal{C} \subset M : \langle \hat\rho, \hat\rho\rangle|_{\mathcal{C}} \geq 0, \tag{4}$$

with interior $\mathcal{O} : \langle \hat\rho, \hat\rho\rangle|_{\mathcal{O}} > 0$ and $m-1$ dimensional boundary $\partial\mathcal{C} : \langle \hat\rho, \hat\rho\rangle|_{\partial\mathcal{C}} = 0$ with outward normal

$$\mathcal{N} = \nabla_{\hat\rho}\hat\rho, \quad \langle \mathcal{N}, \hat\rho\rangle = 0. \tag{5}$$

Proof. Let $\mathcal{X} \in TM$ be an arbitrary vector field in M, then using Killing equation we get

$$\langle \nabla \langle \hat\rho, \hat\rho\rangle, \mathcal{X}\rangle \equiv \nabla_{\mathcal{X}} \langle \hat\rho, \hat\rho\rangle = 2\langle \nabla_X \hat\rho, \hat\rho\rangle = -2\langle \nabla_{\hat\rho}\hat\rho, \mathcal{X}\rangle, \tag{6}$$

$$\Updownarrow$$

$$\nabla \langle \hat\rho, \hat\rho\rangle = -2\nabla_{\hat\rho}\hat\rho. \tag{7}$$

Thus $\nabla \langle \hat\rho, \hat\rho\rangle|_{\partial\mathcal{C}} \neq 0$ and the boundary $\partial\mathcal{C}$ is a hypersurface in M with the normal field $\mathcal{N}$ proportional to $\nabla_{\hat\rho}\hat\rho$. Choosing an outward normal, i.e. directed so that the function $\langle \hat\rho, \hat\rho\rangle$ decreases along the $\mathcal{N}$ we get first condition in (5). Applying Killing equation again we get second condition in (5) since

$$\langle \mathcal{N}, \hat\rho\rangle = \langle \nabla_{\hat\rho}\hat\rho, \hat\rho\rangle = -\langle \nabla_{\hat\rho}\hat\rho, \hat\rho\rangle = 0. \tag{8}$$

□

Definition 2.1. A connected manifold $\mathcal{C}$ will be called causal region. The manifold $\mathcal{O}$ will be called accessible region.[1]

From the point of view of geodesics, and, in particular, the fundamental photon orbits, the region $\mathcal{C}$ represents an accessible region for the null geodesics motion in some effective potential.[17,21] Physical meaning of the causal region $\mathcal{C}$ is that any point can be theoretically observable for any observer in the same region. This causal region may contain spatial infinity (if any) and then will be observable for

an asymptotic observer. In some cases, several causal regions may exist, while null geodesics with a given ρ cannot connect one to another. The boundary $\partial\mathcal{C}$ of the causal region is defined as the branch of the solution of the equation $\langle\hat{\rho},\hat{\rho}\rangle = 0$ and is the set of turning points of null geodesics.

3. Photon submanifold

Let M and S be Lorentzian manifolds, of dimension m and n respectively, and $S \to M$ an isometric embedding[11] defining S as a submanifold[d] in M. We adopt here the following convention for the second fundamental form II of the submanifold[11]:

$$\nabla_Y X = \mathcal{D}_Y X + \mathrm{II}(X,Y), \quad X,Y \in TS, \tag{9}$$

where $\mathcal{D}_X Y \in TS$, $\mathrm{II}(X,Y) \in TS^{\perp}$ and ∇ and $\mathcal{D}$ are the Levi-Civita connections on M and S respectively.

Definition 3.1. We will call a submanifold S invariant, if the Killing vector fields $\mathcal{K}_\alpha$ and $[\mathcal{K}_\alpha,\mathcal{K}_\beta]$ in M are tangent vector fields to S.

For invariant submanifolds the Killing vectors of M will be also the Killing vectors on the submanifold S[20] and well defined restrictions $\hat{\rho}^{\perp}$ and $\mathcal{C}$ on TS and S respectively. We now define a weakened version of the standard umbilic condition (1) requiring it to be satisfied only for some subbundle $V \subset TS$ in the tangent bundle TS.

Definition 3.2. A submanifold S will be called totally V umbilic if[1]

$$\mathrm{II}(X,Y) = \mathrm{H}|_V \langle X,Y\rangle, \quad \forall X,Y \in V. \tag{10}$$

In particular, every totally umbilic submanifold is trivially totally V umbilic for any V. We also note that in the general case $\mathrm{H}|_V$ appearing in this formula is only part of the mean curvature[11] i.e the trace of II on the subbundle V. For invariant totally V umbilic submanifolds, an important theorem on the behavior of null geodesics holds, generalizing the classical result.[7,11]

Proposition 3.1. *Any null geodesic γ with impact parameter ρ in an invariant Lorentzian submanifold $S \subset \mathcal{O}$ is a null geodesic in M if and only if S is totally $\hat{\rho}^{\perp}$ umbilic submanifold.*

Proof. Let S be a totally $\hat{\rho}^{\perp}$ umbilic invariant Lorentzian submanifold and γ be an arbitrary affinely parameterized null geodesic with impact parameter ρ in S i.e. $\mathcal{D}_{\dot\gamma}\dot\gamma = 0$ and $\dot\gamma \in \hat{\rho}^{\perp} \subset TS$. Then by the Gauss decomposition (9)

$$\nabla_{\dot\gamma}\dot\gamma = \mathcal{D}_{\dot\gamma}\dot\gamma + \mathrm{II}(\dot\gamma,\dot\gamma) = \mathrm{H}|_{\hat{\rho}^{\perp}} \langle\dot\gamma,\dot\gamma\rangle = 0, \tag{11}$$

consequently γ is a null geodesic in M.

[d] A hypersurface if $n = m-1$

Conversely, let every null geodesic γ with impact parameter ρ in S be a null geodesic in M, then from the Gauss decomposition (9)

$$\mathrm{II}(\dot\gamma,\dot\gamma)=0, \tag{12}$$

for any null $\dot\gamma\in\hat\rho^\perp\subset TS$. Since we limited ourselves to the accessible region $\mathcal{O}$ we can choose an orthonormal Lorentzian frame $\{\hat\tau/||\hat\tau||,e_a\}\cong\hat\rho^\perp\subset TS$ $(a=1,n-2)$. Then the equality (12) for null vectors $\dot\gamma=\hat\tau/||\hat\tau||\pm e_a$ in the new frame takes the form

$$\mathrm{II}(\hat\tau/||\hat\tau||,\hat\tau/||\hat\tau||)+\mathrm{II}(e_a,e_a)=0,\quad \mathrm{II}(\hat\tau/||\hat\tau||,e_a)=0. \tag{13}$$

And for null vectors $\dot\gamma=\hat\tau/||\hat\tau||\pm(e_a\pm e_b)/\sqrt{2}$

$$\mathrm{II}(e_a,e_b)=0. \tag{14}$$

□

Physical meaning of the Proposition 3.1 is that the null geodesics with a given impact parameter ρ initially touching the spatial section of the invariant totally $\hat\rho^\perp$ umbilic submanifold remain on it for an arbitrarily long time, unless of course they leave it across the boundary. This is a well-known property of a photon sphere and its generalization - a photon surface.[7] Thus, we obtain a generalization of the classical definition of the photon surfaces to the case of a class of geodesics with a fixed impact parameter.

It is useful to obtain an equation for the second fundamental form of the totally $\hat\rho^\perp$ umbilic submanifold in the original basis $\{\mathcal{K}_\alpha,e_a\}$.

Proposition 3.2. *For the invariant $\hat\rho^\perp$ umbilic submanifold, the second fundamental form* II *in the basis $\{\mathcal{K}_\alpha,e_a\}$ has the form*

$$\mathrm{II}=\begin{pmatrix} -\frac{1}{2}\nabla^\perp\mathcal{G}_{\alpha\beta} & -\sum_{A=1}^{m-n}\epsilon_A\left\langle\nabla_{\xi_A}\mathcal{K}_\alpha,e_b\right\rangle\xi_A \\ -\sum_{A=1}^{m-n}\epsilon_A\left\langle\nabla_{\xi_A}\mathcal{K}_\beta,e_a\right\rangle\xi_A & \mathrm{H}|_{\hat\rho^\perp}\left\langle e_a,e_b\right\rangle \end{pmatrix}, \tag{15}$$

where $\mathrm{H}|_{\hat\rho^\perp}$ and $\mathcal{G}_{\alpha\beta}$ satisfy the master equation[20]

$$\rho^\alpha\mathcal{M}_{\alpha\beta}\rho^\beta=0,\quad \mathcal{M}_{\alpha\beta}=\frac{1}{2}\nabla^\perp\left(\mathcal{G}^{-1}\mathcal{G}_{\alpha\beta}\right)-\mathrm{H}|_{\hat\rho^\perp}\left(\mathcal{G}^{-1}\mathcal{G}_{\alpha\beta}\right), \tag{16}$$

$$\mathcal{G}^{\alpha\lambda}\epsilon_{\lambda\beta}\rho^\beta\mathrm{II}(\mathcal{K}_\alpha,e_a)=0, \tag{17}$$

and the derivative along the unit normals $\xi_A\in TS^\perp$ $(A=1,m-n)$ of the submanifold S is defined as

$$\nabla^\perp(\,\cdot\,)=\sum_{A=1}^{m-n}\epsilon_A\nabla_{\xi_A}(\,\cdot\,)\xi_A,\quad \epsilon_A\equiv\left\langle\xi_A,\xi_A\right\rangle. \tag{18}$$

Proof. From the Killing equation for $X\in\{e_a\}$ we find

$$\mathrm{II}(X,\mathcal{K}_\alpha)=\pi^\perp(\nabla_X\mathcal{K}_\alpha)=\sum_{A=1}^{m-n}\epsilon_A\left\langle\nabla_X\mathcal{K}_\alpha,\xi_A\right\rangle\xi_A=-\sum_{A=1}^{m-n}\epsilon_A\left\langle\nabla_{\xi_A}\mathcal{K}_\alpha,X\right\rangle\xi_A. \tag{19}$$

Using the Killing equation again we get

$$\langle \nabla_{\mathcal{K}_\alpha} \mathcal{K}_\beta, \xi_A \rangle = - \langle \nabla_{\xi_A} \mathcal{K}_\beta, \mathcal{K}_\alpha \rangle = -\nabla_{\xi_A} \langle \mathcal{K}_\alpha, \mathcal{K}_\beta \rangle - \langle \xi_A, \nabla_{\mathcal{K}_\beta} \mathcal{K}_a \rangle . \tag{20}$$

Then, using the involutivity condition $[\mathcal{K}_\alpha, \mathcal{K}_\beta] \in TS$ we finally find

$$\mathrm{II}(\mathcal{K}_\alpha, \mathcal{K}_\beta) = \sum_{A=1}^{m-n} \epsilon_A \langle \nabla_{\mathcal{K}_\alpha} \mathcal{K}_\beta, \xi_A \rangle \xi_A = -\frac{1}{2} \sum_{A=1}^{m-n} \epsilon_A \nabla_{\xi_A} \langle \mathcal{K}_\alpha, \mathcal{K}_\beta \rangle \xi_A. \tag{21}$$

and

$$\mathrm{II}(\hat{\tau}, \hat{\tau}) = -\frac{1}{2} \tau^\alpha \tau^\beta \nabla^\perp \mathcal{G}_{\alpha\beta} = -\rho^\alpha \rho^\beta \left(\frac{1}{2} \nabla^\perp \mathcal{G}_{\alpha\beta} + \mathcal{G}^{\lambda\gamma} \epsilon_{\lambda\alpha} \nabla^\perp \epsilon_{\gamma\beta} \right) \tag{22}$$

$$= \frac{1}{2} \rho^\alpha \rho^\beta \left(-\nabla^\perp \mathcal{G}_{\alpha\beta} + \mathcal{G}_{\alpha\beta} \nabla^\perp \ln \mathcal{G} \right). \tag{23}$$

Substituting this expression into the equation (10), we get (16). □

Alternatively, the master equation can be rewritten as expression for the mean curvature

$$\mathrm{H}|_{\hat{\rho}^\perp} = \frac{1}{2} \nabla^\perp \ln \left(\mathcal{G}^{-1} \langle \hat{\rho}, \hat{\rho} \rangle \right). \tag{24}$$

If the submanifold under consideration is totally umbilic we'll get $\mathcal{M}_{\alpha\beta} = 0$.

The notion of an invariant $\hat{\rho}^\perp$ umbilic submanifold is however too general (as is the notion of an umbilic surface by itself[22]) and is not yet defined at the boundary of the causal region $\partial\mathcal{C}$. Generally speaking, these submanifolds are geodesically not complete (in the sense that null geodesics can leave them across the boundary) or have a non-compact spatial section (geodesics can go into the asymptotic region). Moreover, for each ρ there can be an infinite number of them, just as there are an infinite number of umbilic surfaces, but only one photon sphere in the static Schwarzschild[23] solution. Therefore, it is necessary to introduce a more specific definition of fundamental photon submanifolds.[1,20]

Definition 3.3. A fundamental photon submanifold $S \subset \mathcal{C}$ is an invariant Lorentzian submanifold with compact spatial section such that:

(a) The boundary ∂S (if any) lie in $\partial\mathcal{C}$.

(b) The second fundamental form II has the form (15) and satisfies the master equation/inequality

$$\mathcal{M}(\hat{\rho}, \hat{\rho}) = 0, \quad \langle \hat{\rho}, \hat{\rho} \rangle \geq 0. \tag{25}$$

In the case $n = m - 1$, the fundamental photon submanifold is a timelike fundamental photon hypersurface (FPH).

Proposition 3.3. *Every null geodesic γ with an impact parameter ρ at least once touching an arbitrary fundamental photon submanifold S lies in it completely i.e. $\gamma \subset S$.*

Proof. Obviously condition (b) in Definition 3.3, by virtue of Proposition 3.1, prevents null geodesics from leaving the fundamental photon submanifold at all interior points $(S/\partial S) \cap \mathcal{O}$.

Conditions (a-b) for boundary points ∂S and interior points $S \cap \partial\mathcal{C}$ prevents the possibility of null geodesics to leave fundamental photon submanifolds through them. Indeed, $\partial\mathcal{C}$ is the set of turning points for null geodesics in M that cannot move in the direction of the normal $\mathcal{N}|_{\partial S}$, while condition $\mathrm{II}(\hat{\rho},\hat{\rho})|_{\partial S} = 0$ following from (25) not only prevents geodesics from moving in the normal directions ξ_A but also ensures that the normal $\mathcal{N}_{\partial S}$ to ∂S in S coincides with the normal $\mathcal{N}|_{\partial S}$ to the $\partial\mathcal{C}$ in M.

$$\mathcal{N}|_{\partial S} = \nabla_{\hat{\rho}}\hat{\rho}|_{\partial S} = \{\mathcal{D}_{\hat{\rho}}\hat{\rho} + \mathrm{II}(\hat{\rho},\hat{\rho})\}\,|_{\partial S} = \mathcal{D}_{\hat{\rho}}\hat{\rho}|_{\partial S} = \mathcal{N}_{\partial S}. \tag{26}$$

□

From this statement, it is clear that the so-defined fundamental photon submanifolds have trapping properties even at the boundary and contain

(a) non-periodic trapped photon orbits,

(b) periodic fundamental photon orbits.[17]

In stationary and axisymmetric geometry, there exists a vector field

$$\hat{\omega} = \omega^{\alpha}\mathcal{K}_{\alpha}, \quad \omega^{\alpha} = (1,\omega), \tag{27}$$

orthogonal to the all spatial slices Σ of manifold M[e]. In such slices, one can define fundamental photon submanifolds in terms of *principal curvatures*.

Proposition 3.4. *The principal curvatures λ of the spatial section S' of the invariant totally $\hat{\rho}^{\perp}$ umbilic submanifold S satisfy the master equation*

$$\lambda_a - \lambda_{\varphi} = \nabla^{\perp}\ln(||\hat{\rho}||/||\hat{\omega}||), \tag{28}$$

with

$$\lambda_{\varphi} = -\frac{1}{2}\nabla^{\perp}\ln\mathcal{G}_{\varphi\varphi}, \quad \lambda_a = \mathrm{H}|_{\hat{\rho}^{\perp}}. \tag{29}$$

Proof. From the general theory for the second fundamental form II' of the spatial section $S' \subset \Sigma$ of the invariant totally $\hat{\rho}^{\perp}$ umbilic submanifold S we find that

$$\mathrm{II}'(X,Y) = \mathrm{II}(X,Y) = \mathrm{H}|_{\hat{\rho}^{\perp}}\langle X,Y\rangle, \quad \mathrm{II}'(\mathcal{K}_{\varphi},\mathcal{K}_{\varphi}) = \mathrm{II}(\mathcal{K}_{\varphi},\mathcal{K}_{\varphi}) = -\frac{1}{2}\nabla^{\perp}\mathcal{G}_{\varphi\varphi},$$

$$\mathrm{II}'(\mathcal{K}_{\varphi},X) = \mathrm{II}(\mathcal{K}_{\varphi},X) = -\sum_{A=1}^{m-n}\epsilon_A\langle\nabla_{\xi_A}\mathcal{K}_{\varphi},X\rangle\,\xi_A, \quad X \in \{e_a\}.$$

In what follows, we will assume that $\mathrm{II}(\mathcal{K}_{\varphi},X) = 0$.[1] Due to the shape of the second fundamental form of the spatial slices Σ and the expression for mixed components, we find $\mathrm{II}(\hat{\omega},X) = 0$, and therefore from $\mathrm{II}(\mathcal{K}_{\varphi},X) = 0$ follows $\mathrm{II}(\mathcal{K}_t,X) = 0$ and

[e]The norm $||\hat{\omega}||$ of vector field $\hat{\omega}$ is called the lapse function.

the second umbilic condition (17) is fulfilled identically. Further, from the master equation (24) and expressions for the determinant $\mathcal{G}$

$$\begin{aligned}\mathcal{G} &= \langle \mathcal{K}_t, \mathcal{K}_t \rangle \langle \mathcal{K}_\varphi, \mathcal{K}_\varphi \rangle - \langle \mathcal{K}_t, \mathcal{K}_\varphi \rangle^2 \\ &= (\langle \hat{\omega}, \hat{\omega} \rangle - 2\omega \langle \mathcal{K}_t, \mathcal{K}_\varphi \rangle - \omega^2 \langle \mathcal{K}_\varphi, \mathcal{K}_\varphi \rangle) \langle \mathcal{K}_\varphi, \mathcal{K}_\varphi \rangle - \omega^2 \langle \mathcal{K}_\varphi, \mathcal{K}_\varphi \rangle^2 \\ &= \langle \hat{\omega}, \hat{\omega} \rangle \langle \mathcal{K}_\varphi, \mathcal{K}_\varphi \rangle . \end{aligned} \tag{30}$$

we get (28). □

4. Photon region

We now define the concept of a fundamental photon region and a fundamental photon function – a generalization of the classical three-dimensional photon region in the Kerr metric.[2,3,24]

Definition 4.1. The fundamental photon function PF will be called the mapping[1]

$$PF : \rho \to \bigcup S \tag{31}$$

which associates with each ρ one or the union of several fundamental photon submanifolds with the same ρ.

The function PF can be continuous, defining some connected smooth submanifold in the extended manifold $\{M, \rho\}$. At the same time, several continuous functions PF can exist in which different FPHs correspond to one ρ. In particular, for a given ρ, photon and antiphoton FPHs ((un)stable photon surface[25]) can occur simultaneously, indicating the instability of the solution.[8]

Definition 4.2. The fundamental photon region is the complete image of the function PF

$$PR = \bigcup_{\rho} PF. \tag{32}$$

A fundamental photon region is a standard region in the space M in which there are fundamental photon orbits and, in particular, the classical photon region in the Kerr metric. However, the mapping PF can several times cover the image of PR or part of it when the parameter ρ is continuously changed. For example, in the case of a static space, PR is covered at least two times, i.e. PF is a two-sheeted function.

The photon region can also be described using an algebraic equation.[20] To do this, rewrite the equations and the inequality (25) as

$$\mathcal{M}_{tt}\rho^2 + 2\mathcal{M}_{t\varphi}\rho + \mathcal{M}_{\varphi\varphi} = 0, \tag{33}$$

$$\mathcal{G}_{tt}\rho^2 + 2\mathcal{G}_{t\varphi}\rho + \mathcal{G}_{\varphi\varphi} \geq 0. \tag{34}$$

Then, excluding ρ, we find

$$2(-\mathcal{M}_{t\varphi} \pm \sqrt{\mathcal{M}_{t\varphi}^2 - \mathcal{M}_{tt}\mathcal{M}_{\varphi\varphi}})(\mathcal{G}_{t\varphi} - \mathcal{G}_{tt}\mathcal{M}_{t\varphi}/\mathcal{M}_{tt})/\mathcal{M}_{tt} + (\mathcal{G}_{\varphi\varphi} - \mathcal{G}_{tt}\mathcal{M}_{\varphi\varphi}/\mathcal{M}_{tt}) \geq 0, \tag{35}$$

or

$$\pm 2(\mathcal{G}_{t\varphi}\mathcal{M}_{tt} - \mathcal{G}_{tt}\mathcal{M}_{t\varphi})\sqrt{-\mathcal{M}} - 2\mathcal{G}_{tt} \cdot \mathcal{M} + \mathcal{M}_{tt} \cdot \mathcal{G} \cdot \mathrm{Tr}(\mathcal{M}) > 0. \tag{36}$$

These inequality describe a generalized photon region.

5. Conclusion

In this article, we briefly presented a purely geometric approach to defining characteristic surfaces and regions filled with closed photon orbits, based on some generalization of umbilic hypersurfaces. The main new concept is a *partially* umbilic surface, which has umbilic properties with respect to a correctly defined subbundle of the tangent bundle. This approach does not address the integration of geodesic equations, and thus is applicable to spacestimes with a non-integrable geodesic structure.

We tried to give a more clear and concise idea of the main geometric notions presented in,[1] supplementing them with a number of new useful expressions and relations, which, in particular, turned out to be useful for analyzing their connection with Killing tensor fields.[20] We hope that this formalism will pave the way for obtaining new topological constraints, Penrose-type inequalities (and other estimates),[26–28] uniqueness theorems,[23,29–33] similar to ones known for photon spheres and transversally trapping surfaces.[34,35]

The work is supported by the Russian Foundation for Basic Research on the project 20-52-18012Bulg-a, and the Scientific and Educational School of Moscow State University “Fundamental and Applied Space Research”.

References

1. K. V. Kobialko and D. V. Gal'tsov, Photon regions and umbilic conditions in stationary axisymmetric spacetimes: Photon Regions, *Eur. Phys. J. C* **80**, p. 527 (2020).
2. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon Regions and Shadows of Kerr-Newman-NUT Black Holes with a Cosmological Constant, *Phys. Rev. D* **89**, p. 124004 (2014).
3. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon Regions and Shadows of Accelerated Black Holes, *Int. J. Mod. Phys. D* **24**, p. 1542024 (2015).
4. D. C. Wilkins, Bound Geodesics in the Kerr Metric, *Phys. Rev. D* **5**, p. 814 (1972).
5. E. Teo, Spherical orbits around a Kerr black hole, *Gen. Rel. Grav.* **53**, p. 10 (1972).
6. V. I. Dokuchaev and N. O. Nazarova, Silhouettes of invisible black holes, *Usp. Fiz. Nauk* **190**, p. 627-647 (2020).
7. C. M. Claudel, K. S. Virbhadra and G. F. R. Ellis, The Geometry of photon surfaces, *J. Math. Phys.* **42**, p. 818-838 (2001).

8. G. W. Gibbons and C. M. Warnick, Aspherical Photon and Anti-Photon Surfaces, *Phys. Lett. B* **763**, p. 169-173 (2016).
9. V. Perlick and O. Y. Tsupko, Calculating black hole shadows: Review of analytical studies, *arXiv*, p. 2105.07101 (2021).
10. P. V. P. Cunha and C. A. R. Herdeiro, Shadows and strong gravitational lensing: A brief review, *Gen. Rel. Grav.* **50**, p. 42 (2018).
11. B. Y. Chen, Pseudo-Riemannian Geometry, δ-Invariants and Applications, *World Scientific, Hackensack* (2011).
12. N. J. Cornish and G. W. Gibbons, The Tale of two centers, *Class. Quant. Grav.* **14**, p. 1865 (1997).
13. P. V. P. Cunha, J. Grover, C. Herdeiro, E. Radu, H. Runarsson and A. Wittig, Chaotic lensing around boson stars and Kerr black holes with scalar hair, *Phys. Rev. D* **94**, p. 104023 (2016).
14. O. Semerak and P. Sukova, Free motion around black holes with discs or rings: Between integrability and chaos - I, *Mon. Not. Roy. Astron. Soc.* **404**, p. 545 (2010).
15. J. Shipley and S. R. Dolan, Binary black hole shadows, chaotic scattering and the Cantor set, *Class. Quant. Grav.* **33**, p. 175001 (2016).
16. P. V. P. Cunha, C. A. R. Herdeiro and M. J. Rodriguez, Does the black hole shadow probe the event horizon geometry?, *Phys. Rev. D* **97**, p. 084020 (2018).
17. P. V. P. Cunha, C. A. R. Herdeiro and E. Radu, Fundamental photon orbits: Black hole shadows and spacetime instabilities, *Phys. Rev. D* **96**, p. 024039 (2017).
18. D. V. Gal'tsov and K. V. Kobialko, Completing characterization of photon orbits in Kerr and Kerr-Newman metrics, *Phys. Rev. D* **99**, p. 084043 (2019).
19. D. V. Gal'tsov and K. V. Kobialko, Photon trapping in static axially symmetric spacetime, *Phys. Rev. D* **100**, p. 104005 (2019).
20. K. Kobialko, I. Bogush and D. Gal'tsov, Killing tensors and photon surfaces in foliated spacetimes, *Phys. Rev. D* **104**, p. 044009 (2021).
21. G. Lukes-Gerakopoulos, The non-integrability of the Zipoy-Voorhees metric, *Phys. Rev. D* **86**, p. 044013 (2012).
22. L. M. Cao and Y. Song, Quasi-local photon surfaces in general spherically symmetric spacetimes, *arXiv*, p. 1910.13758 (2019).
23. C. Cederbaum and G. J. Galloway, Photon surfaces with equipotential time-slices, *J.Math. Phys.* **62**, p. 032504 (2021).
24. J. Grover and A. Wittig, Black Hole Shadows and Invariant Phase Space Structures, *Phys. Rev. D* **96**, p. 024045 (2017).
25. Y. Koga and T. Harada, Stability of null orbits on photon spheres and photon surfaces, *Phys. Rev. D* **100**, p. 064040 (2019).
26. T. Shiromizu, Y. Tomikawa, K. Izumi and H. Yoshino, Area bound for a surface in a strong gravity region, *PTEP* **2017**, p. 033E01 (2017).
27. X. H. Feng and H. Lu, On the size of rotating black holes, *Eur. Phys. J. C* **80**, p. 551 (2020).
28. R. Q. Yang and H. Lu, Universal bounds on the size of a black hole, *Eur. Phys. J. C* **80**, p. 949 (2020).
29. C. Cederbaum and G. J. Galloway, Uniqueness of photon spheres in electro-vacuum spacetimes, *Class. Quant. Grav.* **33**, p. 075006 (2016).
30. S. S. Yazadjiev, Uniqueness of the static spacetimes with a photon sphere in Einstein-scalar field theory, *Phys. Rev. D* **91**, p. 123013 (2015).
31. S. Yazadjiev and B. Lazov, Uniqueness of the static Einstein–Maxwell spacetimes with a photon sphere, *Class. Quant. Grav.* **32**, p. 165021 (2015).

32. S. Yazadjiev and B. Lazov, "Classification of the static and asymptotically flat Einstein-Maxwell-dilaton spacetimes with a photon sphere," *Phys. Rev. D* **93**, p. 083002 (2016).
33. M. Rogatko, Uniqueness of photon sphere for Einstein-Maxwell-dilaton black holes with arbitrary coupling constant, *Phys. Rev. D* **93**, p. 064003 (2016).
34. H. Yoshino, K. Izumi, T. Shiromizu and Y. Tomikawa, Transversely trapping surfaces: Dynamical version, *PTEP* **2020**, p. 023E02 (2020).
35. H. Yoshino, K. Izumi, T. Shiromizu and Y. Tomikawa, Formation of dynamically transversely trapping surfaces and the stretched hoop conjecture, *PTEP* **2020**, p. 053E01 (2020).

Gravitational lensing by charged accelerating black holes

Torben C. Frost

ZARM, University of Bremen, 28359 Bremen, Germany and Institute for Theoretical Physics, Leibniz University Hannover, 30167 Hannover, Germany
E-mail: torben.frost@zarm.uni-bremen.de

Current astrophysical observations show that on large scale the Universe is electrically neutral. However, locally this may be quite different. Black holes enveloped by a plasma in the presence of a strong magnetic field may have acquired a significant electric charge. We can also expect that some of these charged black holes are moving. Consequently to describe them we need spacetime metrics describing moving black holes. In general relativity such a solution is given by the charged C-de Sitter-metric. In this article we will assume that it can be used to describe moving charged black holes. We will investigate how to observe the electric charge using gravitational lensing. First we will use elliptic integrals and functions to solve the geodesic equations. Then we will derive lens equation, travel time and redshift. We will discuss the impact of the electric charge on these observables and potential limitations for its observation.

Keywords: Black Holes; Gravitational Lensing

1. Introduction

X-ray and gravitational wave observations indicate that stellar mass black holes are widely distributed in our galaxy.[1,2] In addition, the motion of gas and stars[3,4] around galactic centres as well as the recent observation of the shadow of the compact object in the centre of the galaxy M87[5] indicate that in the centre of most galaxies we can find at least one supermassive black hole (SMBH). While the resolution of these observations are not yet precise enough to exclude all alternative theories of gravity or different types of compact objects all observational features indicate that these compact objects can be described either by the Schwarzschild metric or the Kerr metric.

Concurrent astrophysical observations show that our Universe is electrically neutral. This indicates that all astrophysical objects, among them black holes, do not carry a significant electric charge. However, if a black hole is surrounded by a plasma with sufficiently large magnetic fields, the black hole may acquire a significant electric charge.[6] In general relativity charged black holes that are at rest and do not carry a spin are described by the Reissner-Nordström metric, or, if we want to take the cosmological expansion into account, by the Reissner-Nordström-de Sitter metric. In the case that we consider real astrophysical environments it is very unlikely that black holes are at rest. Instead it is more realistic to assume that they constantly undergo accelerated motion. Consequently we need spacetimes describing accelerating black holes to describe them accurately. In the framework of general relativity a family of such solutions is described by the charged C-de Sitter metric.[7]

The metric is an exact solution to Einstein's electrovacuum field equations with cosmological constant. It generalises the Reissner-Nordström-de Sitter metric by introducing the acceleration parameter α in addition to the mass parameter m, the electric charge e and the cosmological constant Λ. The charged C-de Sitter metric is axisymmetric and static and describes a charged accelerating black hole with cosmological constant. The acceleration of the black hole results from conical singularities on the axes, which are commonly interpreted as a string pulling the black hole and a strut pushing the black hole. These conical singularities make the spacetime on a first view appear unphysical. However, it may still serve as an approximation for black holes undergoing accelerated motion. Gravitational lensing in the C-metric only attracted attention relatively recently. Grenzebach, Perlick and Lämmerzahl[8,9] investigated the shadow of accelerating black holes in the whole Plebański-Demiański family with acceleration.[10] Sharif and Iftikhar[11] calculated the deflection angle for light rays in the equatorial plane for accelerating Kerr-NUT black holes. The method they used explicitly assumes that the light rays are located in the equatorial plane of the spacetime. However, Alrais Alawadi, Batic and Nowakowski[12] and subsequently Frost and Perlick[13] demonstrated that the C-metric possesses a photon cone $\vartheta_{\mathrm{ph}} \neq \pi/2$ and therefore the method of Sharif and Iftikhar cannot be directly applied. Alrais Alawadi, Batic and Nowakowski[12] calculated the deflection angle on the photon cone. Only a short time later Frost and Perlick[13] investigated gravitational lensing in the C-metric. In their work Frost and Perlick[13] used the canonical form of the elliptic integrals and Jacobi's elliptic functions to solve the geodesic equations. Then, following the apporach of Grenzebach, Perlick and Lämmerzahl,[8] they fixed an observer in the region of outer communication and introduced an orthonormal tetrad to relate latitude and longitude on the observer's celestial sphere to the constants of motion of the light rays. Using the conventions in Bohn et al.[14] they formulated a lens equation, they derived the redshift and the travel time of individual light rays on the celestial sphere of the observer. In this article we will extend their approach to the charged C-de Sitter metric. For this purpose the remainder of this article is structured as follows. As not all readers may be familiar with the charged C-de Sitter metric we will provide a short overview of its physical properties in Section 2. In Section 3 we will demonstrate how to solve the equations of motion. In Section 4 we will first introduce the orthonormal tetrad and show how to parameterise the light rays using the angles on the observer's celestial sphere. Then we will introduce the lens map, formulate the lens equation, and calculate the redshift and the travel time of light rays. In Section 5 we will shortly summarise our results and conclusions. Throughout the whole article we will use geometric units with $c = G = 1$. The metric signature is chosen as $(-,+,+,+)$.

2. The Charged C-de Sitter-Metric

The charged C-de Sitter metric is a solution to Einstein's electrovacuum field equations with cosmological constant. It belongs to the Plebański-Demiański family of

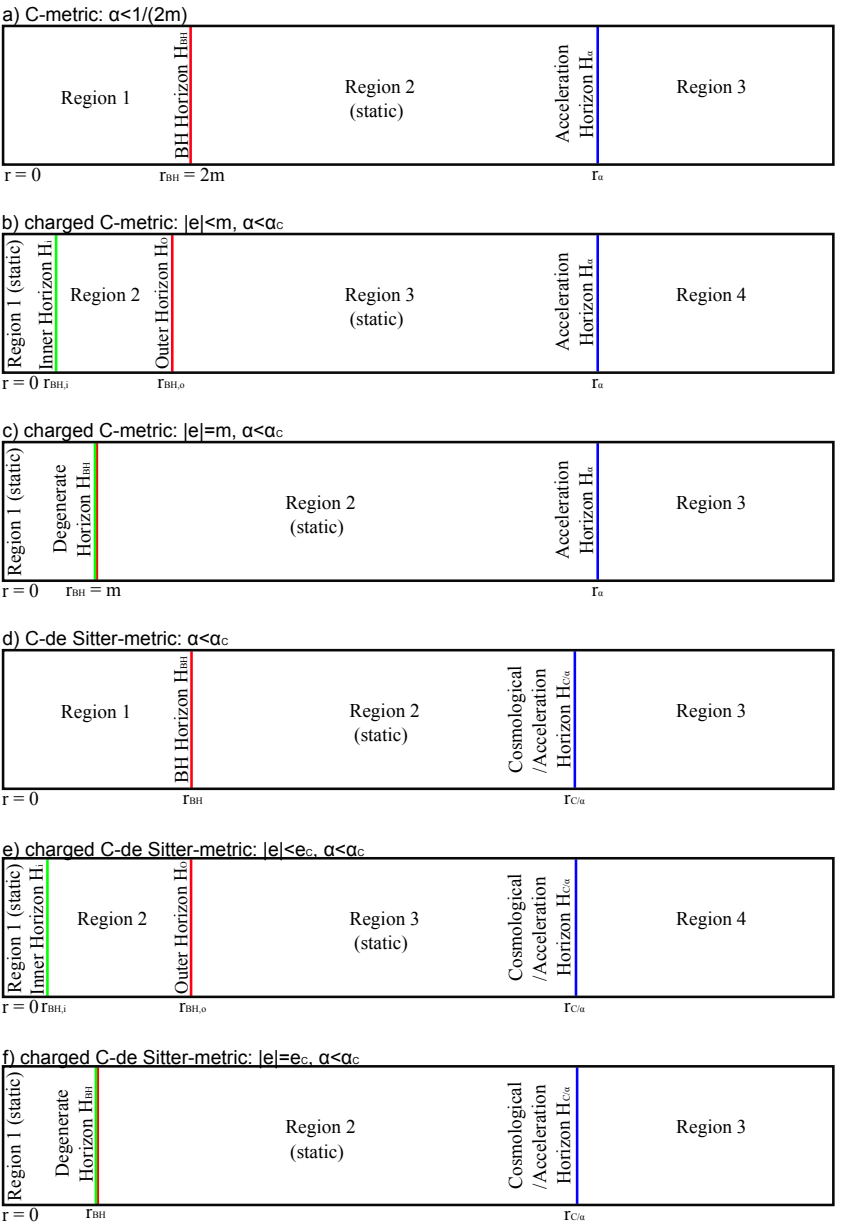


Fig. 1. Position of the curvature singularity at $r = 0$ and the coordinate singularities in a) the C-metric, the charged C-metric with b) $|e| < m$ and c) $|e| = m$, d) the C-de Sitter metric and the charged C-de Sitter metric with e) $|e| < e_{\rm C}$ and f) $|e| = e_{\rm C}$. Note that the angular coordinates are suppressed and other singularities are not shown.

spacetimes of Petrov type D[10] and is axisymmetric and static. The spacetime is characterised by four parameters: the mass parameter m, the electric charge e, the cosmological constant Λ, and the acceleration parameter α. Its line element reads[7]

$$g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = \frac{1}{\Omega(r,\vartheta)^2}\left(-Q(r)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{Q(r)} + \frac{r^2\mathrm{d}\vartheta^2}{P(\vartheta)} + r^2\sin^2\vartheta P(\vartheta)\mathrm{d}\varphi^2\right), \quad (1)$$

where

$$Q(r) = (1-\alpha^2r^2)\left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) - \frac{\Lambda}{3}r^2, \quad (2)$$

$$P(\vartheta) = 1 - 2\alpha m\cos\vartheta + \alpha^2e^2\cos^2\vartheta, \quad (3)$$

$$\Omega(r,\vartheta) = 1 - \alpha r\cos\vartheta. \quad (4)$$

When we set $\alpha = 0$ the metric reduces to the Reissner-Nordström-de Sitter family of spacetimes including the Schwarzschild metric ($e = 0$, $\Lambda = 0$), the Reissner-Nordström metric ($\Lambda = 0$), and the Schwarzschild-de Sitter metric ($e = 0$). For $e = 0$ and $\Lambda = 0$ the metric reduces to the regular C-metric. For $e = 0$ it reduces to the C-de Sitter metric and for $\Lambda = 0$ it reduces to the charged C-metric. In accordance with observations we choose $0 < m$. Theoretically the electric charge e can take any real value but because the line element (1) only contains its square we restrict it such that we have $0 \leq e \leq e_\mathrm{C}$. Because we live in an expanding Universe we will assume that $0 \leq \Lambda < \Lambda_\mathrm{C}$. α can be restricted from the symmetry of the spacetime. When we replace α by $-\alpha$ and substitute $\vartheta \to \pi - \vartheta$ we see that (1) is invariant and consequently we can restrict the acceleration parameter to $0 < \alpha < \alpha_\mathrm{C}$. Here, e_C, Λ_C and α_C are critical parameters that have to be chosen according to the desired interpretation of the spacetime. In the charged C-de Sitter metric the time coordinate t can assume any real value. In addition we choose the angular coordinates ϑ and φ such that they represent the angular coordinates on the two sphere S^2. The range of the r coordinate is determined by the singularity structure of the spacetime, in particular the singularities arising from $Q(r)$ and $\Omega(r,\vartheta)$. We will discuss these singularities and their implications for the r coordinate in the following. The spacetime admits several singularities. We start by discussing the singularities arising from $Q(r)$. In total $Q(r)$ can lead to up to five singularities in the metric. The equation $Q(r) = 0$ can lead to up to four singularities. In addition the metric has another singularity at $r = 0$. The singularity at $r = 0$ is a curvature singularity. For $e \neq 0$ the curvature singularity is timelike. Because lightlike and timelike geodesics cannot cross the curvature singularity particle motion is limited to $0 < r$. In addition the conformal factor becomes singular when $\Omega(r,\vartheta) = 0$. This singularity corresponds to conformal infinity and limits the radius coordinate r. The singularity starts at $r = 1/\alpha$ on the axis $\vartheta = 0$ and extends to $r = \infty$ for $\pi/2 \leq \vartheta$. The other singularities are only coordinate singularities r_H that can be removed using appropriate coordinate transformations (for the C-metric the procedure is

demonstrated in Ref. 15). In this article we restrict to black hole spacetimes and consequently all coordinate singularities with $0 < r_{\rm H}$ are horizons. In general the horizon struture of each metric depends on the parameters of the spacetime. For each spacetime the critical parameters $e_{\rm C}$, $\Lambda_{\rm C}$ and $\alpha_{\rm C}$ have to be chosen such that all horizons are real and that they lead to physically meaningful black hole spacetimes. The horizon structure of the whole family of charged C-de Sitter metrics is shown in Fig. 1. Panel a) shows the horizon struture of the C-metric. The C-metric has two horizons, a black hole horizon at $r_{\rm BH} = 2m$ and an acceleration horizon at $r_{\rm BH} < r_\alpha = 1/\alpha$. Between the horizons the spacetime is static. In the other two regions the spacetime is non-static. Panels b) and c) show the horizon struture for the charged C-metric. The main difference to the C-metric is that the spacetime now contains an inner and an outer black hole horizon $r_{\rm BH,i} \leq r_{\rm BH,o} < r_\alpha$. When $e = m$ both horizons coincide and form a degenerate horizon at $r_{\rm BH} = m$. The regions $0 < r < r_{\rm BH,i}$ and $r_{\rm BH,o} < r < r_\alpha$ are static. The other two regions are non-static. When we include the cosmological constant Λ (panels d) to f)) the radius coordinates of the horizons change, however, the horizon structure remains unchanged. For a more detailed discussion of the horizon structure in the C-metric we refer the interested reader to Ref. 13. In our case we always choose $e_{\rm C}$ such that $P(\vartheta)$ does not lead to singularities in the metric. Finally, the spacetime has two singularities on the axes at $\sin\vartheta = 0$. As demonstrated in Refs. 7 and 15 these singularities are conical singularites. They are associated with a deficit angle ($\vartheta = 0$) and a surplus angle ($\vartheta = \pi$). The charged C-de Sitter metric is usually interpreted such that it describes a charged accelerating black hole with cosmological constant. In its full analytic extension the spacetime describes two causally separated charged black holes accalerating away from each other. The conical singularity at $\vartheta = 0$ is commonly interpreted as a string pulling the black hole, while the conical singularity at $\vartheta = \pi$ is interpreted as a strut pushing the black hole. As demonstrated in Refs. 7 and 15 one of the conical singularities can always be removed, however, as in our previous work[13] here we will retain both singularities to show the effects of both, the string and the strut on geodesic motion.

3. Equations of Motion

In the charged C-de Sitter metric the equations of motion for light rays are fully separable. In Mino parameterisation[16] they read[8,13]

$$\frac{\mathrm{d}t}{\mathrm{d}\lambda} = \frac{r^2 E}{Q(r)}, \tag{5}$$

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = E^2 r^4 - r^2 Q(r) K, \tag{6}$$

$$\left(\frac{\mathrm{d}\vartheta}{\mathrm{d}\lambda}\right)^2 = P(\vartheta) K - \frac{L_z^2}{\sin^2\vartheta}, \tag{7}$$

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{L_z}{\sin^2\vartheta P(\vartheta)}, \tag{8}$$

where the Mino parameter λ is related to the affine parameter s by

$$\frac{\mathrm{d}\lambda}{\mathrm{d}s} = \frac{\Omega(r,\vartheta)^2}{r^2}. \tag{9}$$

Here, the three constants of motion E, L_z and K are the energy, the angular momentum about the z axis and the Carter constant of the light ray, respectively. In this article we choose the energy E such that $\mathrm{d}t/\mathrm{d}\lambda > 0$. In the following we will briefly discuss the equations of motion, the turning points of the r and the ϑ motion and identify the locations of the photon sphere and the photon cone. Then we will turn to solving the equations of motion. We will derive the solutions to the equations of motion for arbitrary initial conditions $(x_i^\mu) = (x^\mu(\lambda_i)) = (t_i, r_i, \vartheta_i, \varphi_i)$ closely following the procedures described in Refs. 13 and 17. Because the main focus of this article will be to apply our results to gravitational lensing we will limit our discussion to lightlike geodesics in the region of outer communication between photon sphere and the cosmological/acceleration horizon.

3.1. *The r Motion*

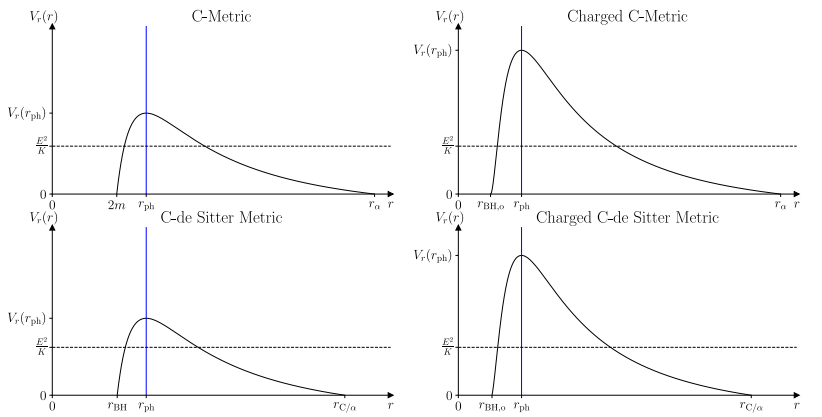


Fig. 2. Potential $V_r(r)$ of the r motion in the C-metric (top left), the charged C-metric (top right), the C-de Sitter metric (bottom left) and the charged C-de Sitter metric (bottom right) for $\Lambda = 1/(200m^2)$, $e = m$, and $\alpha = 1/(10m)$. The axes have the same scale in all four plots.

We begin with discussing the r motion. Following Ref. 13 we first rewrite (6) in terms of the potential $V_r(r)$

$$\frac{1}{r^4 K}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + V_r(r) = \frac{E^2}{K}, \tag{10}$$

where

$$V_r(r) = \left(\frac{1}{r^2} - \alpha^2\right)\left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) - \frac{\Lambda}{3}. \tag{11}$$

Fig. 2 shows the potentials for the C-metric (top left), the charged C-metric (top right), the C-de Sitter metric (bottom left) and the charged C-de Sitter metric (bottom right). For all four metrics the potential has a similar structure. The most striking difference occurs close to the (outer) black hole horizon. For $e \neq 0$ the maximum of $V_r(r)$ is much higher than in the uncharged case. In addition for $\Lambda > 0$ $V_r(r)$ approaches zero at a smaller radius coordinate $r_{\mathrm{C}/\alpha} < r_\alpha$. The maximum of the potential at $E^2/K = V_r(r_{\mathrm{ph}})$ marks the radius coordinate of an unstable photon sphere. We can calculate the radius coordinate of the photon sphere by calculating $\mathrm{d}r/\mathrm{d}\lambda = \mathrm{d}^2r/\mathrm{d}\lambda^2 = 0$. Combining both conditions leads to the determining equation

$$r^3 + \frac{1 - \alpha^2 e^2}{m\alpha^2} r^2 - \frac{3}{\alpha^2} r + \frac{2e^2}{m\alpha^2} = 0 \tag{12}$$

for the radius of the photon sphere r_{ph}. It is remarkable that although the spherical symmetry is broken by the acceleration parameter the cosmological constant does not have any effect on the radius coordinate of the photon sphere. The roots of this equation can be analytically obtained using Cardano's method. As long as $e^2 < 1/\alpha^2$ we get three real solutions. Only the largest root lies in the region of outer communication $r_{\mathrm{BH,o}} < r < r_{\mathrm{C}/\alpha}$ and thus marks the position of the photon sphere. When $e = 0$ the polynomial reduces to second order and r_{ph} is given by[12,13]

$$r_{\mathrm{ph}} = \frac{6m}{1 + \sqrt{1 + 12\alpha^2 m^2}}. \tag{13}$$

We can read from Fig. 2 that at r_{ph} $V_r(r_{\mathrm{ph}})$ has always a maximum and thus orbits of geodesics on the photon sphere are unstable. Here, unstable means that if the orbit of a lightlike geodesic on the photon sphere was radially infinitessimally perturbed the light ray would either fall into the black hole or escape to (conformal) infinity. Using the potential we can distinguish three different types of lightlike motion. For $E^2/K > V_r(r_{\mathrm{ph}})$ lightlike geodesics have no turning points. Here we have to distinguish between the cases $K = 0$ (we will see in Section 3.3 that these geodesics are principal null geodesics) and $K > 0$. For $E^2/K > V_r(r_{\mathrm{ph}})$ and $K = 0$ (6) was already solved in Ref. 13. In this case the solution reads:

$$r(\lambda) = \frac{r_i}{1 - i_{r_i} r_i E\,(\lambda - \lambda_i)}, \tag{14}$$

where $i_{r_i} = \operatorname{sgn}\left(\mathrm{d}r/\mathrm{d}\lambda|_{r=r_i}\right)$. In the case $E^2/K > V_r(r_{\mathrm{ph}})$ and $K > 0$ (6) has two real and two complex conjugated roots. We label them such that $r_1 > r_2$ and $r_3 = \bar{r}_4 = R_3 + iR_4$. To solve (6) we now substitute[17,18]

$$r = \frac{r_1\bar{R} - r_2 R + (r_1\bar{R} + r_2 R)\cos\chi_r}{\bar{R} - R + (\bar{R} + R)\cos\chi_r}, \tag{15}$$

where

$$R = \sqrt{(R_3 - r_1)^2 + R_4^2} \quad \text{and} \quad \bar{R} = \sqrt{(R_3 - r_2)^2 + R_4^2}. \tag{16}$$

Then we follow the steps described in Appendix B of Ref. 13 and obtain the solution in terms of Jacobi's elliptic cn function. It reads

$$r(\lambda) = \frac{r_1\bar{R} - r_2 R + (r_1\bar{R} + r_2 R)\mathrm{cn}\left(a_r\left(\lambda - \lambda_i\right) + \lambda_{r_i,k_1}, k_1\right)}{\bar{R} - R + (\bar{R} + R)\mathrm{cn}\left(a_r\left(\lambda - \lambda_i\right) + \lambda_{r_i,k_1}, k_1\right)}, \tag{17}$$

where

$$a_r = i_{r_i}\sqrt{\left(E^2 + \left(\alpha^2 + \frac{\Lambda}{3}\right)K\right)R\bar{R}}, \qquad \lambda_{r_i,k_1} = F_L(\chi_{r_i}, k_1), \tag{18}$$

$$\chi_{r_i} = \arccos\left(\frac{(r_i - r_2)R - (r_i - r_1)\bar{R}}{(r_i - r_2)R + (r_i - r_1)\bar{R}}\right), \; k_1 = \frac{(R + \bar{R})^2 - (r_1 - r_2)^2}{4R\bar{R}}.$$

The condition $E^2/K = V_r(r_{\mathrm{ph}})$ characterises lightlike geodesics asymptotically coming from or going to the photon sphere. In this case (6) has four real roots, two of which are equal. We label them such that $r_1 = r_2 = r_{\mathrm{ph}} > r_3 > r_4$. In this case we substitute

$$r = r_3 + \frac{3a_{3,r}}{12y - a_{2,r}}, \tag{19}$$

where

$$a_{2,r} = 6\left(E^2 + \left(\alpha^2 + \frac{\Lambda}{3}\right)K\right)r_3^2 - 6m\alpha^2 K r_3 - (1 - \alpha^2 e^2)K, \tag{20}$$

$$a_{3,r} = 4\left(E^2 + \left(\alpha^2 + \frac{\Lambda}{3}\right)K\right)r_3^3 - 6m\alpha^2 K r_3^2 - 2(1 - \alpha^2 e^2)K r_3 + 2mK. \tag{21}$$

Then we follow the steps described in Section 3.2.3 in Ref. 13 and obtain as solution for $r(\lambda)$:

$$r(\lambda) = r_3 - \frac{(r_{\mathrm{ph}} - r_3)(r_3 - r_4)}{r_{\mathrm{ph}} - r_3 - (r_{\mathrm{ph}} - r_4)\tanh^2\left(b_r + i_{r_i}\sqrt{\frac{a_{3,r}(r_{\mathrm{ph}} - r_4)}{4(r_{\mathrm{ph}} - r_3)(r_3 - r_4)}}(\lambda_i - \lambda)\right)}, \tag{22}$$

where

$$b_r = \mathrm{artanh}\left(\sqrt{\frac{(r_i - r_4)(r_{\mathrm{ph}} - r_3)}{(r_i - r_3)(r_{\mathrm{ph}} - r_4)}}\right). \tag{23}$$

All remaining lightlike geodesics have $E^2/K < V_r(r_{\mathrm{ph}})$. These geodesics have four real roots and can pass through a turning point. Outside the photon sphere this turning point is always a minimum. For solving (6) we now label the roots such that $r_1 = r_{\mathrm{min}} > r_2 > r_3 > r_4$. Then we substitute[17,18]

$$r = r_2 + \frac{(r_1 - r_2)(r_2 - r_4)}{r_2 - r_4 - (r_1 - r_4)\sin^2\chi_r}. \tag{24}$$

Again we apply the steps described in Appendix B of Ref. 13. This time we obtain the solution to (6) in terms of Jacobi's elliptic sn function

$$r(\lambda) = r_2 + \frac{(r_1 - r_2)(r_2 - r_4)}{r_2 - r_4 - (r_1 - r_4)\mathrm{sn}^2\left(c_r\left(\lambda - \lambda_i\right) + \lambda_{r_i,k_2}, k_2\right)}, \tag{25}$$

where

$$c_r = \frac{i_{r_i}}{2}\sqrt{\left(E^2 + \left(\alpha^2 + \frac{\Lambda}{3}\right)K\right)(r_1 - r_3)(r_2 - r_4)}, \quad \lambda_{r_i,k_2} = F_L(\chi_{r_i}, k_2), \tag{26}$$

$$\chi_{r_i} = \arcsin\left(\sqrt{\frac{(r_i - r_1)(r_2 - r_4)}{(r_i - r_2)(r_1 - r_4)}}\right), k_2 = \frac{(r_2 - r_3)(r_1 - r_4)}{(r_1 - r_3)(r_2 - r_4)}.$$

3.2. *The t Coordinate*

Now we calculate the time coordinate t. The right-hand side of (5) depends only on $r(\lambda)$. Therefore, we first rewrite (5) as a differential of r. For this purpose we divide (5) by the root of (6). Then we integrate over r starting at $t(\lambda_i) = t(r_i) = t_i$. The resulting integral reads

$$t(\lambda) = t_i + \int_{r_i\ldots}^{\ldots r(\lambda)} \frac{E r'^2 \mathrm{d}r'}{Q(r')\sqrt{E^2 r'^4 - r'^2 Q(r')K}}. \tag{27}$$

Here, we have to choose the sign of the root in accordance with the r motion and the dots in the limits of the integration indicate that we have to split the integral at the turning points. The procedure to integrate (27) is straight forward. We have to distinguish the same types of motion as for $r(\lambda)$ in Section 3.1. First we perform a partial fraction decomposition of $r'^2/Q(r')$. Then we follow the procedure described in Refs. 13 and 17 to express (27) in terms of elementary functions and elliptic integrals. Note that some of the elliptic integrals cannot immediately be expressed in form of elementary functions and elliptic integrals of first, second and third kind. However, we can easily rewrite them using the procedures described in Appendix B in Ref. 17 and Appendix A in Ref. 13.

3.3. *The ϑ Motion*

Again we start by rewriting (7) in terms of the potential $V_\vartheta(\vartheta)$

$$\frac{\sin^2\vartheta}{K}\left(\frac{\mathrm{d}\vartheta}{\mathrm{d}\lambda}\right)^2 + V_\vartheta(\vartheta) = -\frac{L_z^2}{K}, \tag{28}$$

where

$$V_\vartheta(\vartheta) = -\sin^2\vartheta(1 - 2\alpha m\cos\vartheta + \alpha^2 e^2\cos^2\vartheta). \tag{29}$$

Fig. 3 shows the potentials $V_\vartheta(\vartheta)$ of the ϑ motion for the C-(de Sitter) metric (left) and the charged C-(de Sitter) metric (right). Although in the right plot we set $e = m$ the differences are invisibly small. The potential has always a minimum at a

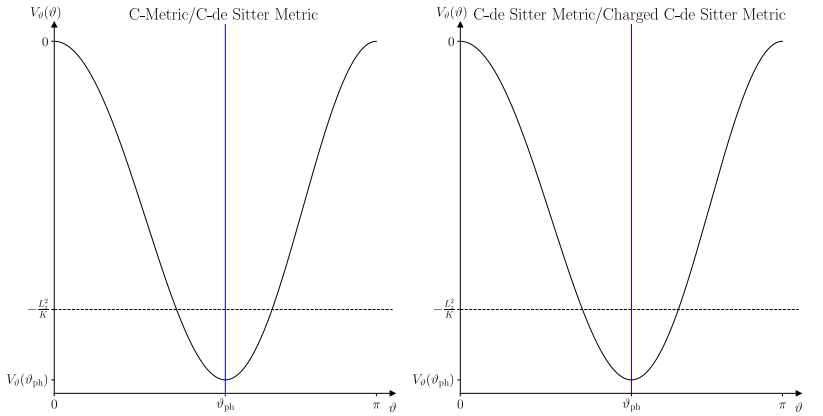


Fig. 3. Potential $V_\vartheta(\vartheta)$ of the ϑ motion in the C-metric/C-de Sitter metric (left) and the charged C-metric/C-de Sitter metric (right) for $e = m$ and $\alpha = 1/(10m)$. The axes have the same scale in both plots.

single value ϑ. At this minimum we have $\mathrm{d}\vartheta/\mathrm{d}\lambda = \mathrm{d}^2\vartheta/\mathrm{d}\lambda^2 = 0$. Combining both conditions leads to the determining equation

$$\cos^3\vartheta - \frac{3m}{2\alpha e^2}\cos^2\vartheta + \frac{1-\alpha^2e^2}{2\alpha^2e^2}\cos\vartheta + \frac{m}{2\alpha e^2} = 0. \tag{30}$$

We determine the roots of this equation using Cardano's method. Within the permissible range $0 \leq \vartheta \leq \pi$ the equation has exactly one real root $\vartheta_{\rm ph}$. This is the photon cone of the charged C-(de Sitter) metric. Note that we cannot determine the limits $e \to 0$ or $\alpha \to 0$ from the calculated $\vartheta_{\rm ph}$ directly. When we approach these limits we see that for $e \to 0$ $\vartheta_{\rm ph}$ reduces to[12, 13]

$$\vartheta_{\rm ph} = \arccos\left(\frac{-2\alpha m}{1+\sqrt{1+12\alpha^2m^2}}\right) \tag{31}$$

and for $\alpha \to 0$ it reduces to $\vartheta_{\rm ph} = \pi/2$. The photon cone is unstable. Here unstable means that when light rays on the photon cone are infinitessimally perturbed in ϑ direction they will begin to oscillate between a minimum $\vartheta_{\rm min}$ and a maximum $\vartheta_{\rm max}$. Based on the potential $V_\vartheta(\vartheta)$ we can distinguish two different types of motion. From (7) we can immediately read that we have $L_z = 0$ when $K = 0$ and the right-hand side of (7) vanishes. Because $L_z = 0$ implies $\mathrm{d}\varphi/\mathrm{d}\lambda = 0$ these are the already in Section 3.1 mentioned principal null geodesics. Also for lightlike geodesics on the photon cone ($-L_z^2/K = V_\vartheta(\vartheta_{\rm ph})$) the right-hand side of (7) vanishes because we have two coinciding roots at $x = \cos\vartheta_{\rm ph}$. Thus for principal null geodesics and lightlike geodesics on the photon cone the solution to (7) reads $\vartheta(\lambda) = \vartheta_i$. For all other lightlike geodesics we have $-L_z^2/K \neq V_\vartheta(\vartheta_{\rm ph})$. These geodesics oscillate between a minimum $\vartheta_{\rm min}$ and a maximum $\vartheta_{\rm max}$ and each geodesic can potentially have arbitrary many turning points. In this case we first rewrite (7) in terms of

$x = \cos\vartheta$. It is easy to see that in the new parameterisation (7) reduces to a polynomial of third order in x when $e = 0$. This case was already treated in Ref. 13 and thus we do not reproduce it here. For $e \neq 0$ we always have two turning points, however, depending on the choice of L_z and K the right-hand side of (7) has two different root structures. In the first case two of the roots are real. We label them such that $x_2 = \cos\vartheta_{\max} < x_1 = \cos\vartheta_{\min}$. The other two roots are complex conjugated and we label them as $x_3 = \bar{x}_4 = X_3 + iX_4$. As first step to solve (7) we now substitute[18]

$$x = \frac{x_1\bar{X} + x_2 X + (x_1\bar{X} - x_2 X)\cos\chi_\vartheta}{\bar{X} + X + (\bar{X} - X)\cos\chi_\vartheta}, \tag{32}$$

where X and $\bar{X}$ are given by (16). In the second step we again follow the steps described in Appendix B of Ref. 13. We obtain $\vartheta(\lambda)$ in terms of Jacobi's elliptic cn function. It reads

$$\vartheta(\lambda) = \arccos\left(\frac{x_1\bar{X} + x_2 X + (x_1\bar{X} - x_2 X)\mathrm{cn}\left(a_\vartheta\left(\lambda - \lambda_i\right) + \lambda_{\vartheta_i,k_3}, k_3\right)}{\bar{X} + X + (\bar{X} - X)\mathrm{cn}\left(a_\vartheta\left(\lambda - \lambda_i\right) + \lambda_{\vartheta_i,k_3}, k_3\right)}\right), \tag{33}$$

where $i_{\vartheta_i} = \mathrm{sgn}\left(\mathrm{d}\vartheta/\mathrm{d}\lambda|_{\vartheta=\vartheta_i}\right)$ and

$$a_\vartheta = i_{\vartheta_i}\sqrt{\alpha^2 e^2 K X\bar{X}}, \qquad \lambda_{\vartheta_i,k_3} = F_L(\chi_{\vartheta_i}, k_3), \tag{34}$$

$$\chi_{\vartheta_i} = \arccos\left(\frac{(\cos\vartheta_i - x_1)\bar{X} + (\cos\vartheta_i - x_2)X}{(x_1 - \cos\vartheta_i)\bar{X} + (\cos\vartheta_i - x_2)X}\right),\ k_3 = \frac{(x_1 - x_2)^2 - (X - \bar{X})^2}{4X\bar{X}}.$$

In the second case all four roots are real and we label them such that $x_4 = \cos\vartheta_{\max} < x_3 = \cos\vartheta_{\min} < x_2 < x_1$. In this case we substitute[18]

$$x = x_1 - \frac{(x_1 - x_3)(x_1 - x_4)}{x_1 - x_3 + (x_3 - x_4)\sin^2\chi_\vartheta}. \tag{35}$$

We again follow the steps described in Appendix B of Ref. 13. This time we obtain $\vartheta(\lambda)$ in terms of Jacobi's elliptic sn function

$$\vartheta(\lambda) = \arccos\left(x_1 - \frac{(x_1 - x_3)(x_1 - x_4)}{x_1 - x_3 + (x_3 - x_4)\mathrm{sn}^2\left(b_\vartheta\left(\lambda_i - \lambda\right) + \lambda_{\vartheta_i,k_4}, k_4\right)}\right), \tag{36}$$

where

$$b_\vartheta = \frac{i_{\vartheta_i}}{2}\sqrt{\alpha^2 e^2 K(x_1 - x_3)(x_2 - x_4)}, \qquad \lambda_{\vartheta_i,k_4} = F_L(\chi_{\vartheta_i}, k_4) \tag{37}$$

$$\chi_{\vartheta_i} = \arcsin\left(\sqrt{\frac{(\cos\vartheta_i - x_4)(x_1 - x_3)}{(x_1 - \cos\vartheta_i)(x_3 - x_4)}}\right),\ k_4 = \frac{(x_1 - x_2)(x_3 - x_4)}{(x_1 - x_3)(x_2 - x_4)}.$$

3.4. *The φ Motion*

The φ motion is governed by (8). We immediately read that for $L_z = 0$ the right-hand side vanishes. For $K = 0$ these are the principal null geodesics. In this case we obtain the solution to (8) as $\varphi(\lambda) = \varphi_i$. If $K \neq 0$ these are geodesics crossing the axes. What happens to these geodesics now depends on whether we assume that

the string and the strut are opaque or transparent. In the former case the lightlike geodesics terminate at the string and we observe it as a black line blocking out all light emitted by sources located behind it. In the latter case the lightlike geodesics pass through the string. To see what happens when a lightlike geodesic passes through the axes we consider a series of lightlike geodesics continually approaching the axes for $L_z > 0$ and $L_z < 0$.[13] We observe that for $L_z > 0$ and $L_z < 0$ the φ coordinate has different limits. Hence geodesics crossing the axes have two different continuations and the φ coordinate is not continuous.

Next we turn to lightlike geodesics with a double root at $x = \cos\vartheta_{\mathrm{ph}}$ ($-L_z^2/K = V_\vartheta(\vartheta_{\mathrm{ph}})$). These are lightlike geodesics on the photon cone with $\vartheta(\lambda) = \vartheta_{\mathrm{ph}}$. In this case the right-hand side of (8) is constant and the solution $\varphi(\lambda)$ reads

$$\varphi(\lambda) = \varphi_i + \frac{L_z(\lambda - \lambda_i)}{\sin^2\vartheta_{\mathrm{ph}} P(\vartheta_{\mathrm{ph}})}. \tag{38}$$

Last we turn to lightlike geodesics with $-L_z^2/K \neq V_\vartheta(\vartheta_{\mathrm{ph}})$. In this case we want to express $\varphi(\lambda)$ in terms of elementary functions and elliptic integrals. For this purpose we first substitute $x = \cos\vartheta$ in (7) and (8). Then we divide (8) by the root of (7) and integrate. Now $\varphi(\lambda)$ reads

$$\varphi(\lambda) = \varphi_i + \int_{\cos\vartheta_i \ldots}^{\ldots \cos\vartheta(\lambda)} \frac{L_z \mathrm{d}x'}{(1-x'^2)P(x')\sqrt{(1-x'^2)P(x')K - L_z^2}}. \tag{39}$$

Here, we have to choose the sign of the root such that it corresponds to the sign of the $\cos\vartheta$ motion and the dots in the limits indicate that we have to split the integral at the turning points. Again we rewrite the elliptic integral using elementary functions and the canonical forms of the elliptic integrals. For this purpose we now perform a partial fracion decomposition of $(1-x^2)^{-1}P(x)^{-1}$. Then we use (41) in Ref. 13 and (32) and (35) to rewrite (39) in terms of elementary functions and elliptic integrals of first, second and third kind. Note that we again encounter elliptic integrals which do not immediately have a canonical form. Again we rewrite them using the procedures described in Appendix B in Ref. 17 and Appendix A in Ref. 13 (although here we have $n^2/(n^2-1) < 1$ and not $1 < n^2/(n^2-1)$ the basic integration procedure is the same).

4. Gravitational Lensing

4.1. *Celestial Sphere*

In astronomy the position of a light source on the sky is identified using latitude and longitude coordinates. For this purpose astronomers fix the main target of their observation at the centre of their image and measure the position of all other objects relative to the centre. In our discussion of gravitational lensing in the charged C-de Sitter metric we follow this approach. For this purpose we first fix an observer at coordinates $(x^\mu_O) = (t_O, r_O, \vartheta_O, \varphi_O)$ in the region of outer communication between photon sphere and cosmological/acceleration horizon. Then we choose the black hole

as main target of our observation and introduce an orthonormal tetrad following Refs. 8 and 13:

$$e_0 = \frac{\Omega(r,\vartheta)}{\sqrt{Q(r)}}\partial_t\Big|_{(x^\mu_O)}, \quad e_1 = \frac{\Omega(r,\vartheta)\sqrt{P(\vartheta)}}{r}\partial_\vartheta\Big|_{(x^\mu_O)}, \tag{40}$$

$$e_2 = -\frac{\Omega(r,\vartheta)}{r\sin\vartheta\sqrt{P(\vartheta)}}\partial_\varphi\Big|_{(x^\mu_O)}, \quad e_3 = -\Omega(r,\vartheta)\sqrt{Q(r)}\partial_r\Big|_{(x^\mu_O)}.$$

We will call the angles on the celestial sphere surrounding the observer Σ (latitude) and Ψ (longitude). The angle Σ is measured from the axis along e_3 connecting lens and observer and the angle Ψ is measured in the direction of e_2 from the axis along the direction of e_1.

Before we can investigate gravitational lensing in the charged C-de Sitter metric we first have to derive the relations between the constants of motion E, L_z and K and the angles on the observer's celestial sphere Σ and Ψ. For this purpose we consider a lightlike geodesic ending at the position of the observer. Now we first write down the tangent vector to this geodesic. It reads

$$\frac{\mathrm{d}\eta}{\mathrm{d}\lambda} = \frac{\mathrm{d}t}{\mathrm{d}\lambda}\partial_t + \frac{\mathrm{d}r}{\mathrm{d}\lambda}\partial_r + \frac{\mathrm{d}\vartheta}{\mathrm{d}\lambda}\partial_\vartheta + \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda}\partial_\varphi. \tag{41}$$

At the position of the observer we can also express the tangent vector of the geodesic by the angles on the observer's celestial sphere and the tetrad vectors e_0, e_1, e_2 and e_3

$$\frac{\mathrm{d}\eta}{\mathrm{d}\lambda} = \sigma\left(-e_0 + \sin\Sigma\cos\Psi e_1 + \sin\Sigma\sin\Psi e_2 + \cos\Sigma e_3\right), \tag{42}$$

where σ is a normalisation constant

$$\sigma = g\left(\frac{\mathrm{d}\eta}{\mathrm{d}\lambda}, e_0\right). \tag{43}$$

In our convention we have $E > 0$ and thus σ has to be negative. Because the Mino parameter λ is defined up to an affine transformation without loss of generality we can choose $\sigma = -r_O^2/\Omega(r_O,\vartheta_O)^2$.[13] Now we insert σ in (42) and (43). Then a comparison of the coefficients of (41) and (42) leads to[13]

$$E = \frac{\sqrt{Q(r_O)}}{\Omega(r_O,\vartheta_O)},\ L_z = \frac{r_O\sqrt{P(\vartheta_O)}\sin\vartheta_O\sin\Sigma\sin\Psi}{\Omega(r_O,\vartheta_O)},\ K = \frac{r_O^2\sin^2\Sigma}{\Omega(r_O,\vartheta_O)^2}. \tag{44}$$

4.2. *Angular Radius of the Photon Sphere*

The shadow of a black hole is a higly idealised concept. It is constructed as follows. First we place an observer in the region of outer communication between photon sphere and cosmological/acceleration horizon. Then we distribute light sources everywhere except between observer and the black hole. Therefore the former region is associated with brightness on the observer's sky while the latter is associated with

darkness. Now we shoot back light rays exactly on the boundary between these areas of brightness and darkness. These are geodesics asymptotically going to the photon sphere. They have the same constants of motion as light rays on the photon sphere and thus a double root at r_{ph}. We now use this fact to calculate the angular radius of the shadow. For this purpose we first insert (44) in (6). Then we employ that $\mathrm{d}r/\mathrm{d}\lambda|_{r=r_{\mathrm{ph}}} = 0$. Now we solve for Σ and get as angular radius of the shadow[13]

$$\Sigma_{\mathrm{ph}} = \arcsin\left(\frac{r_{\mathrm{ph}}}{r_O}\sqrt{\frac{Q(r_O)}{Q(r_{\mathrm{ph}})}}\right). \tag{45}$$

4.3. *Lens Map*

The lens map or lens equation maps images on the observer's sky back to the source surface. This article is an extension of the work presented in Ref. 13 and therefore we will closely follow their definition of the lens map. To set up the lens map we first distribute light sources on a sphere with radius $r_O < r_L$ in the region of outer communication. Then we shoot light rays backwards in time from the position of the observer (x^μ_O). Some of the light rays, but not all, will intersect with the sphere of light sources. These geodesics form a map from the celestial sphere of the observer to the sphere of light sources

$$(\Sigma, \Psi) \to (\vartheta_L(\Sigma, \Psi), \varphi_L(\Sigma, \Psi)). \tag{46}$$

This is our lens equation. Now we have to calculate $\vartheta_L(\Sigma, \Psi)$ and $\varphi_L(\Sigma, \Psi)$ to obtain the lens map for the charged C-de Sitter metric. For this purpose we set $(x^\mu_i) = (x^\mu_O)$ and insert (44) into the solutions of the equations of motion derived in Section 3. Due to the symmetry of the charged C-de Sitter metric we can choose t_O and φ_O arbitrarily. Similarly λ is only defined up to an affine transformation. Therefore, to ease all following calculations we choose them such that $\lambda_O = 0$, $t_O = 0$ and $\varphi_O = 0$.

As first step towards calculating $\vartheta_L(\Sigma, \Psi)$ and $\varphi_L(\Sigma, \Psi)$ we calculate the Mino parameter $\lambda_L < \lambda_O$

$$\lambda_L = \int_{r_O \cdots}^{\cdots r_L} \frac{\Omega(r_O, \vartheta_O)\mathrm{d}r'}{\sqrt{Q(r_O)r'^4 - r_O^2 \sin^2 \Sigma r'^2 Q(r')}}. \tag{47}$$

Again for lightlike geodesics passing through a turning point we have to split the integral and choose the sign of the root in the denominator in agreement with the r motion. Now we insert λ_L in the appropriate solution for $\vartheta(\lambda)$ in Section 3.3 and obtain $\vartheta_L(\Sigma, \Psi)$. Next we count the number of turning points n of the ϑ motion. For this purpose we calculate the Mino parameter up to the first turning point of the ϑ motion λ_0. In the next step we calculate the difference of the Mino parameter between two subsequent turning points of the ϑ motion $\Delta\lambda$. Now we count how many turning points occur while $\lambda_L < \lambda_n = \lambda_0 + n\Delta\lambda$. Finally we calculate $\varphi_L(\Sigma, \Psi)$ from (39) as described in Section 3.4.

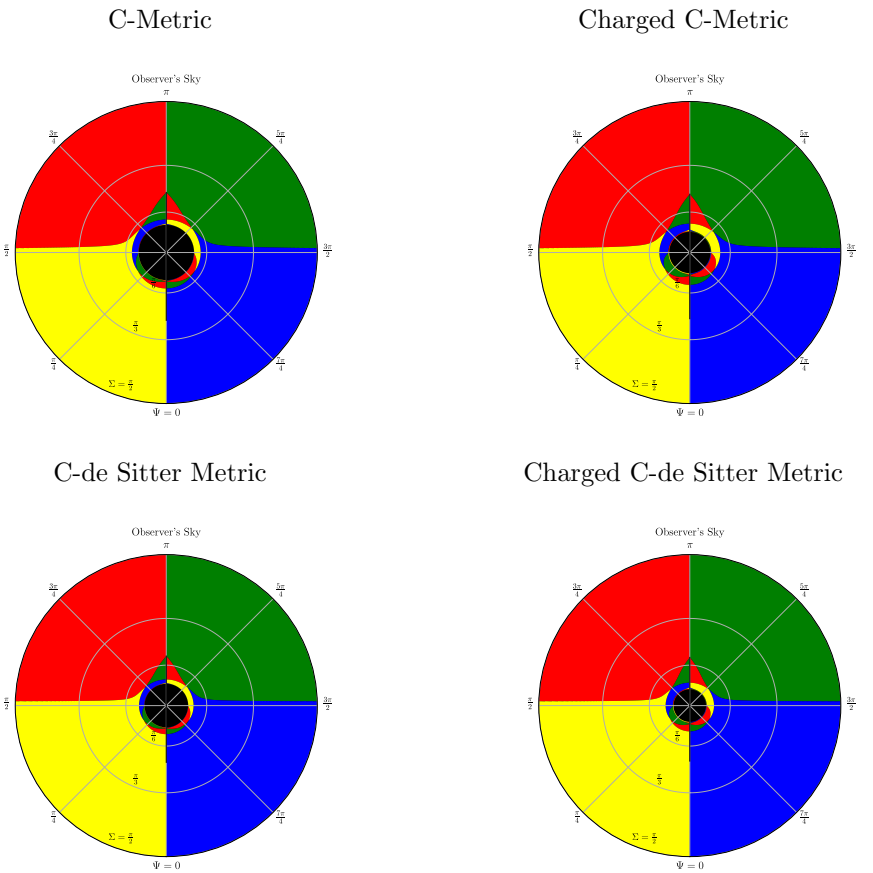


Fig. 4. Lens equation for the C-metric[13] (top left), the charged C-metric (top right), the C-de Sitter metric (bottom left) and the charged C-de Sitter metric (bottom right) for $\Lambda = 1/(200m^2)$, $e = m$ and $\alpha = 1/(10m)$. The observer is located at $r_O = 8m$, $\vartheta_O = \pi/2$ and the sphere of light sources is located at $r_L = 9m$. The colour convention follows Refs. 13 and 14 and is as follows: $0 \leq \vartheta_L \leq \frac{\pi}{2}$: red/green; $\frac{\pi}{2} < \vartheta_L \leq \pi$: yellow/blue $0 \leq \varphi_L < \pi$: green/blue; $\pi \leq \varphi_L < 2\pi$: red/yellow. The black lines at $\Psi = 0$ and $\Psi = \pi$ mark light rays that cross the axes at least once.

Fig. 4 shows plots of the lens equation in stereographic projection for the C-metric (top left), the charged C-metric (top right), the C-de Sitter metric (bottom left) and the charged C-de Sitter metric (bottom right) for $\Lambda = 1/(200m^2)$, $e = m$ and $\alpha = 1/(10m)$. The observer is located at $r_O = 8m$ and $\vartheta_O = \pi/2$. The sphere of light sources is located at $r_L = 9m$. The black lines at $\Psi = 0$ and $\Psi = \pi$ mark lightlike geodesics that cross the string or the strut at least once. The black circle in the centre of each image is the shadow of the black hole. The images clearly show that the angular radius Σ_{ph} of the shadow decreases when $0 < \Lambda$ and $0 < e$. All four images basically show the same features up to a scaling. The symmetry with respect to $\Psi = \pi/2$ and $\Psi = 3\pi/2$ is clearly broken. Thus in all four cases

light rays ending parallel to the surface $\vartheta = \pi/2$ generally come from light sources not located at $\vartheta = \pi/2$. However, the shape of the features shown in the images is clearly symmetric with respect to the line marked by $\Psi = 0$ and $\Psi = \pi$. On the right side of the line of symmetry at the outer boundary of each image we first have a region coloured in blue and green. In this region we find images where the covered angle $0 < \Delta\varphi < \pi$. These are images of first order. Adjacent to these images, closer to the shadow we find a region coloured in yellow and red. In this region we find images for which the light ray covered the angle $\pi < \Delta\varphi < 2\pi$. These are images of second order. If we go closer to the shadow we also find images of third and fourth order. The boundaries between the images of different orders mark the positions of the critical curves. On the left side of the line of symmetry we find the same features, however, the ordering of the colours is reversed. In addition we observe that at $\Psi = \pi$ images of second order already occur further away from the shadow than for $\Psi = 0$. This implies that light rays passing close to the string cover the same angle $\Delta\varphi$ faster than light rays passing close to the strut.

4.4. *Redshift*

The redshift z of a light ray relates its energy at the time of emission by a light source to its energy at the position at which it is detected by an observer. It is directly accessible to observation via the frequency shift of atomic or molecular absorption lines. We will now construct redshift maps. For this purpose we will use the same settings and results we obtained from constructing the lens map in Section 4.3. In this setting observer and light source are static and the corresponding general redshift formula can be found in Ref. 19, pp. 45. After inserting the metric coefficients g_{tt} of the charged C-de Sitter metric it reads[13]

$$z = \sqrt{\frac{g_{tt}|_{x_O}}{g_{tt}|_{x_L}}} - 1 = \sqrt{\frac{Q(r_O)}{Q(r_L)}}\frac{\Omega(r_L, \vartheta_L(\Sigma, \Psi))}{\Omega(r_O, \vartheta_O)} - 1. \tag{48}$$

Fig. 5 shows plots of the redshift maps for the same observer-source geometry as for the lens maps in Fig. 4 In all four plots the outer region is dominated by redshifts. The redshift has two maxima around $\Psi = \pi$ and at $\Psi = 0$ (close to the shadow). In addition in all images we observe a region of blueshifts centered around $\Psi = 0$. Another crescent shaped region of blueshifts can be found at $\Psi = \pi$ close to the shadow. What are now the effects of the cosmological constant Λ and the electric charge e? Comparing the images with $\Lambda = 0$ in the upper row with the images with $\Lambda = 1/(200m^2)$ in the lower row shows that when we turn on the cosmological constant the redshift range shifts from $-1 < z < 2$ to $-0.8 < z < 6$, respectively. The two blueshift areas move to lower latitudes closer to the centre of the shadow. In addition the fraction of the images covered by blueshifts decreases. The effects of the electric charge e are less strongly pronounced. When we compare the images on the left ($e = 0$) to the images on the right ($e = m$) we observe that in the top row the areas with blueshift move to slightly lower latitudes while they still seem

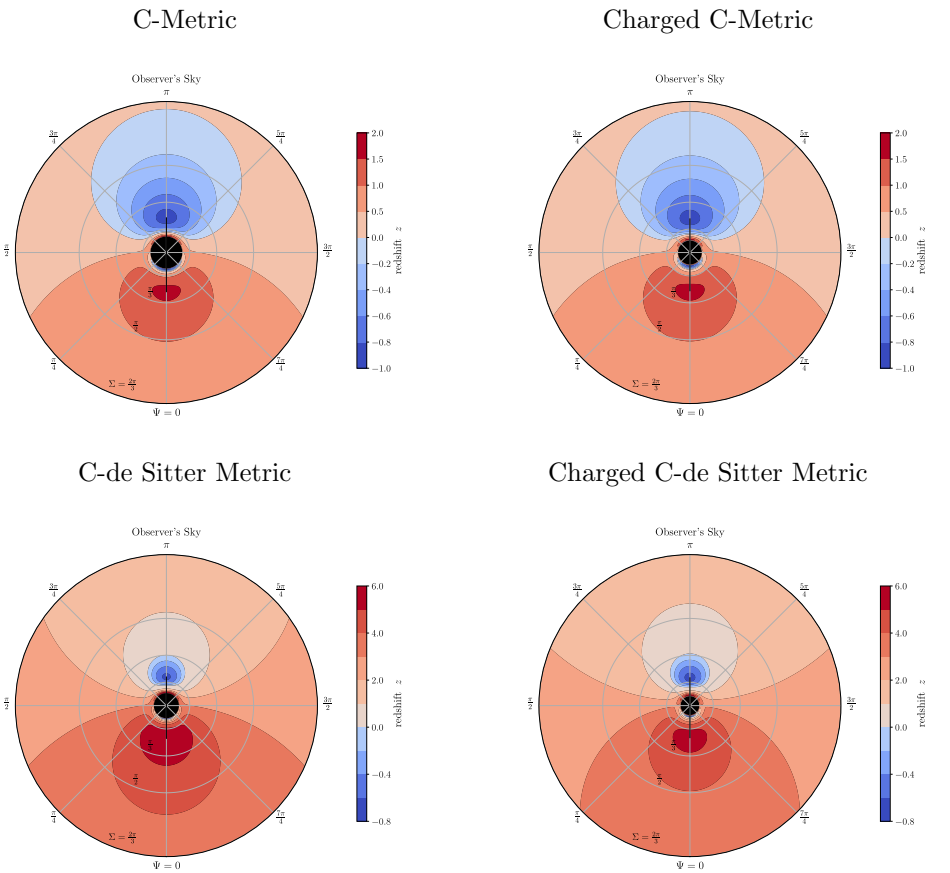


Fig. 5. Redshift maps for the C-metric[13] (top left), the charged C-metric (top right), the C-de Sitter metric (bottom left) and the charged C-de Sitter metric (bottom right) for $\Lambda = 1/(200m^2)$, $e = m$ and $\alpha = 1/(10m)$. The observer is located at $r_O = 8m$, $\vartheta_O = \pi/2$ and the sphere of light sources is located at $r_L = 9m$. The black lines at $\Psi = 0$ and $\Psi = \pi$ mark light rays that cross the axes at least once.

to cover roughly the same fraction of the image. However, comparing both images in the lower row indicates that in the presence of a cosmological constant turning on the electric charge leads to an increase of the fraction of the image covered by blueshifts. In addition the areas of blueshifts shift to lower latitudes but appear to be located at angular distances further away from the shadow.

4.5. *Travel Time*

The travel time $T = t_O - t_L$ measures in terms of the time coordinate t the elapsed time between the emission of a light ray by a source at a time t_L and the detecion of the same light ray by an observer at the time t_O. We obtain the travel time integral

C-Metric/Charged C-Metric C-de Sitter Metric/Charged C-de Sitter Metric

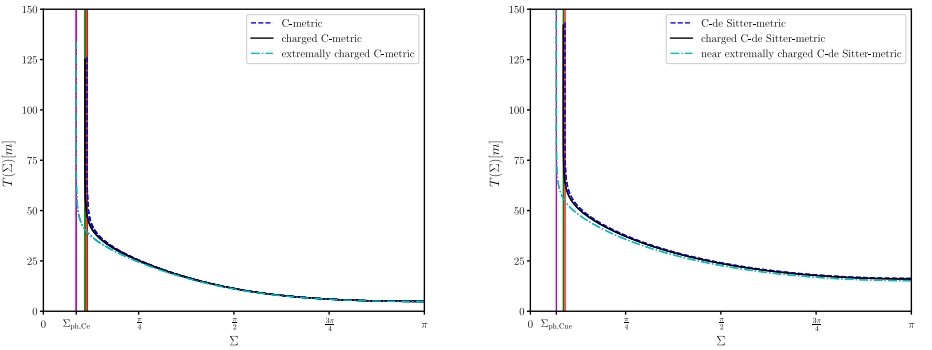


Fig. 6. Travel time $T(\Sigma)$ for the C-metric and the charged C-metric (left), and the C-de Sitter metric and the charged C-de Sitter metric (right) for $\Lambda = 1/(200m^2)$, $e = m$ in the extremal and near extremal case, otherwise $e = m/2$ and $\alpha = 1/(10m)$. The observer is located at $r_O = 8m$ and the sphere of light sources is located at $r_L = 9m$. $\Sigma_{\mathrm{ph,Ce}}$ and $\Sigma_{\mathrm{ph,Cne}}$ mark the position of the angular radius of the shadow on the observer's celestial sphere for the extremally charged C-metric and the near extremally charged C-de Sitter metric, respectively.

by inserting (44) and $t_O = 0$ in (27). We evaluate the travel time as described in Section 3.2. Fig. 6 shows the travel time for the (charged) C-metric (left) and the (charged) C-de Sitter metric (right) for $\Lambda = 1/(200m^2)$, $e = m/2$ in the regular case, $e = m$ in the (near) extremal case and $\alpha = 1/(10m)$. The observer is located at $r_O = 8m$ and the light source is located at $r_L = 9m$. The left plot shows that for $\Sigma > \pi/2$ in the C-metric and the charged C-metric the travel time is roughly the same. When we turn on the cosmological constant this drastically changes. In comparison to the (charged) C-metric the travel time gets significantly longer. However, turning on the electric charge in presence of a positive cosmological constant leads to a decrease of the travel time. In addition both plots indicate that for $e = m/2$ the travel time close to Σ_{ph} is slightly shorter than for $e = 0$ while it is significantly longer for $e = m$.

5. Summary and Implications for Observations

In this article we extended the work presented for the C-metric in Ref. 13 to the charged C-de Sitter metrics. In the first part we discussed and solved the equations of motion using elementary as well as elliptic functions and ellitpic integrals. In the second part we used the derived analytical solutions to investigate gravitational lensing in the charged C-de Sitter metrics.

How can we now use these results in combination with observational data to distinguish between the different black hole spacetimes and in particular to measure the electric charge of a black hole?

In the charged C-de Sitter metrics the shape of the shadow is always circular. Its angular radius decreases with increasing Λ, e and α. However, because the distance between the observer on Earth and the black hole lens is not a priori known it alone cannot be used to determine the nature of the black hole. The length of the travel time in particular close to the shadow is more characteristic for each spacetime. Unfortunately in real astrophysical settings (multiple imaging systems) we can only measure travel time differences leading to similar ambiguities as for the shadow. However, the charged C-de Sitter spacetimes also admit two very characteristic lensing features that distinguish them from their non-accelerating counterparts ($\alpha = 0$). The first characteristic is the breaking of symmetry with respect to the equatorial plane on the celestial sphere of the observer in the lens maps. Here, the most salient feature was the observation that images of second order occur at larger (lower) angular distance from the shadow close to the string (strut). The second characteristic is that the redshift z is a function of the coordinates on the observer's celestial sphere. The breaking of symmetry can be tested by observing multiple images from the same light source gravitationally lensed by a black hole. In such a system we can measure the position of the images on the celestial sphere of the observer relative to the lens and then compare it with theoretical predictions. Similarly although our construction of the redshift map is highly idealised we may be able to find astrophysical systems similar to this configuration. In such a system, we can measure the redshift of known emission lines and construct a partial redshift map to determine if it is a function of the coordinates on the observer's celestial sphere. Observing the symmetry breaking and showing that the measured redshift is a function on the observer's celestial sphere tells us with high certainty that the observed black hole is accelerating and can be described by one of the charged C-de Sitter metrics. However, due to the ambiguity introduced by the a-priori unknown distance between observer and black hole lens we cannot use these measurements alone to determine Λ, e and α. Therefore, to correctly identify the nature of the spacetime describing the black hole we observe, and having the chance to accurately measure the mass parameter m, the cosmological constant Λ, the electric charge e and the acceleration parameter α we need to combine high accuracy measurements of the angular diameter of the shadow, the redshift function on the observer's celestial sphere and of the position of and the travel time differences between two or more multiple images from the same source. Unfortunately even with very high accuracy observations for most black holes the electric charge e is likely to be far to low to be measured and one may only be able to estimate an upper limit.

Acknowledgments

I would like to thank Volker Perlick for the helpful discussions. I acknowledge financial support from QuantumFrontiers. I also acknowledge support from Deutsche Forschungsgemeinschaft within the Research Training Group 1620 Models of Gravity.

References

1. J. M. Corral-Santana, J. Casares, T. Muñoz-Darias, F. E. Bauer, I. G. Martínez-Pais and D. M. Russell, BlackCAT: A catalogue of stellar-mass black holes in X-ray transients, *Astron. Astrophys.* **587**, p. A61 (2016).
2. R. Abbott *et al.*, Population Properties of Compact Objects from the Second LIGO–Virgo Gravitational-Wave Transient Catalog, *Astrophys. J. Lett.* **913**, p. L7 (2021).
3. K. Gültekin, E. M. Cackett, J. M. Miller, T. Di Matteo, S. Markoff and D. O. Richstone, A CHANDRA Survey of Supermassive Black Holes with Dynamical Mass Measurements, *Astrophys. J.* **749**, p. 129 (2012).
4. GRAVITY Collaboration, Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole, *Astron. Astrophys.* **615**, p. L15 (2018).
5. The Event Horizon Telescope Collaboration, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *Astrophys. J. Lett.* **875**, p. L1 (2019).
6. E. Castellanos, J. C. Degollado, C. Lämmerzahl, A. Macías and V. Perlick, Bose-Einstein condensates in charged black–hole spacetimes, *JCAP* **2018**, p. 043 (2018).
7. J. B. Griffiths and J. Podolský, *Exact Space-Times in Einstein's General Relativity* (Cambridge University Press, Cambridge, 2009).
8. A. Grenzebach, V. Perlick and C. Lämmerzahl, Photon regions and shadows of accelerated black holes, *Int. J. of Mod. Phys. D* **24**, p. 1542024 (2015).
9. A. Grenzebach, *The Shadow of Black Holes*, Springer Briefs in Physics, Springer Briefs in Physics (Springer, Heidelberg, 2016).
10. J. F. Plebanski and M. Demianski, Rotating, Charged, and Uniformly Accelerating Mass in General Relativity, *Ann. Phys. (N. Y.)* **98**, 98 (1976).
11. M. Sharif and S. Iftikhar, Equatorial gravitational lensing by accelerating and rotating black hole with NUT parameter, *Astrophys. Space Sci.* **361**, p. 36 (2016).
12. M. Alrais Alawadi, D. Batic and M. Nowakowski, Light bending in a two black hole metric, *Class. Quantum Grav.* **38**, p. 045003 (2021).
13. T. C. Frost and V. Perlick, Lightlike geodesics and gravitational lensing in the spacetime of an accelerating black hole, *Class. Quantum Grav.* **38**, p. 085016 (2021).
14. A. Bohn, W. Throwe, F. Hébert, K. Henriksson, D. Bunandar, M. A. Scheel and N. W. Taylor, What does a binary black hole merger look like?, *Class. Quantum Grav.* **32**, p. 065002 (2015).
15. J. B. Griffiths, P. Krtouš and J. Podolský, Interpreting the C-metric, *Class. Quantum Grav.* **23**, 6745 (2006).
16. Y. Mino, Perturbative approach to an orbital evolution around a supermassive black hole, *Phys. Rev. D* **67**, p. 084027 (2003).
17. S. E. Gralla and A. Lupsasca, Null geodesics of the Kerr exterior, *Phys. Rev. D* **101**, p. 044032 (2020).
18. H. Hancock, *Elliptic Integrals*, Mathematical Monographs, 1st edn. (New York John Wiley & Sons, 1917).
19. N. Straumann, *General Relativity*, Graduate Texts in Physics, Graduate Texts in Physics, 2 edn. (Springer, Heidelberg, 2013).

A manmade experiment aimed to clarify the gravity law in the Solar system

Alexander P. Yefremov* and Alexandra A. Vorobyeva†
Institute of Gravitation and Cosmology, RUDN University,
Miklukho-Maklay str. 6, Moscow, Russia, 117198
* *E-mail: efremov-ap@rudn.ru*
† *E-mail: vorobyeva-aa@rudn.ru*
www.rudn.ru

Ultra-sensitivity of a planet's gravity assist to changes of the test-body impact parameter prompts a space experiment testing the nature of the gravitational field in the Solar system. The Sun, Earth and Venus serve as the space lab with a primitive space probe (space ball) as a test body moving on a ballistic trajectory from the Earth to Venus (rendering GA) and backwards to the Earth's orbit. We explain why in Newton and Einstein gravity, the probe's final positions (reached at the same time) may differ greatly; an Earth's observer can measure the gap.

Keywords: Gravitation; General Relativity; Gravity assist.

Introduction

Recent experiments supporting validity of GR, in our opinion, have at least two weak features. First, we are, in fact, dealing with isolated observations of very rare natural phenomena, mostly it is motion and merging of some massive objects. Second, these objects are very far from the observer. If binary pulsars are observed at distances of "only" 10 thousand light-years, then the sources of the gravitational radiation is determined at about 1.3 billion light-years. So, in fact we examine the past.

This seems to violate the basic principles of an experiment: possibility to repeated it and to deal with physical laws "valid here and now". But it is not the case yet, therefore, the claims about the absolute acceptability of Einstein's theory of gravity are hardly convincing enough.

Well, we understand that a repeatable experiment testing very small GR effects in the near space (in the Solar system) is a very challenging task.

In this presentation we will try to demonstrate that such an experiment (and a series of experiments) aimed to specify the law of the Sun's gravity with good accuracy is quite possible.

The Gravity Assist Maneuver

The main idea of the experiment is to use the so-called gravity assist maneuver of a space probe near a close planet.[1–3]

Reading papers on the space missions (where the gravity assist has become a habitual instrument) we paid attention at a great precision of the impact parameter value (or value of the close parameter — pericenter, the shortest distance between the planet rendering the assistance and a spacecraft). On the Solar system scale of hundred million kilometers, the impact parameter is normally precalculated with the accuracy of one kilometer.

This precision should reflect a great sensibility of results of the gravity assist maneuver to the pericenter value; small deflection must entail great trouble for the mission. However, we could find no sensible information about the gravity-assist sensitivity function. So, we had to determine it ourselves.

The Standard Flight

At first, we designed a scheme of a "standard flight" in Newton's gravity of the Sun. It represents a space probe flight in a planetary system comprising a central star and planets imitating the Sun, Earth, and Venus. For simplicity, it is assumed that the planets move in circular orbits, however, the other physical parameters of the system are very close to the observation data.

The pattern flight looks as follows. A space probe is launched from the Earth (or from the Earth's orbit — since we ignore here non-essential Earth's gravity). From the launch point at initial zero-time moment and azimuth (polar angle) also zero,

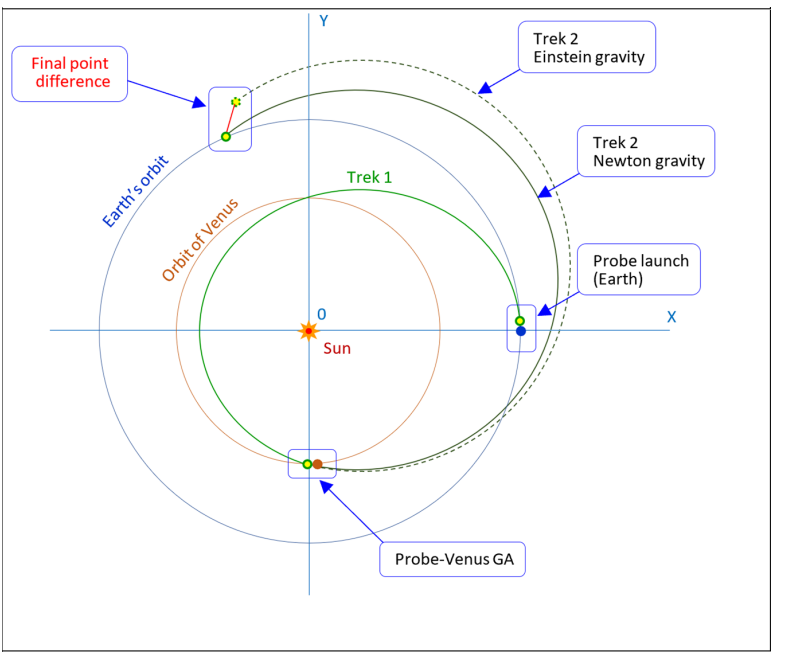


Fig. 1. The standard flight and the Gravity assist effects for different gravity laws.

the probe falls freely towards Venus by a Kepler elliptic leg (trek 1 in Fig. 1), and after a sufficiently prolonged flight it arrives at the Venus gravity-assist point. The initial conditions are chosen so, that the probes' perihelion has azimuth 180 deg, while the gravity assist point has azimuth of 270 deg, and strictly precomputed pericenter value. The Venus gravity assist accelerates the probe so that after flight by another elliptic leg (trek 2 in Fig. 1) it returns to the Earth's orbit, and at the assigned time sends a signal received by the Earth's observer who thus determines the probe's final position.

The Gravity-Assist Sensitivity as a Function of Pericenter Value

Slightly changing the initial conditions, we manage to make the chosen pericenter shorter by 1 km and let it move freely to the final position the same time. Then the calculated distance between the probe's final positions (with the pericenter difference of 1 km) gives us the gravity- assist sensitivity function for the chosen pericenter value.

Subsequently changing the initial conditions, we found the sensitivity values analytically for 20 pericenter lengths (from 18 000 km to 8 000 km — that is from 12 000 km to 2 000 km altitude) and obtained the following result. The sensitivity turns out quite a non-linear function.

At a greater pericenter value (18 000 km) the sensitivity is about 20 000 (this means that one-kilometer difference (i.e. 17 999 km) throws the probe aside its assigned final point at 20 000 km. In other words, the sensitivity is about 104. At smaller pericenter value, about 10 000 km (it is 4 000 km altitude) the sensitivity grows ten times as 105. That is at the final point the probe is thrown aside at 100 000 km.

Thus, the gravity assist maneuver may work as a powerful amplifier of small pericenter changes caused by some physical agents on its first elliptic leg (Trek 1).

We should say that we cross-checked the analytical evaluation of the sensitivity function values by a pure numerical construction of respective probe trajectories with the help of iterative program (realized within Python 3 system), choosing the time interval of 1 second, so that the probe's trajectories are built on 35–50 million points.

The results of analytical calculation and numerical construction of the trajectories turned out quite close.

The plots of GA sensitivity function are represented in Figs. 2–3.

The first physical factor to distort Kepler's Trek 1 is Einstein's general relativistic gravity; so we made preliminary computations, and again by two methods: analytically and numerically (and again the results are very close).

The GR gravity brings three reasons making he pericenter shorter. First, the GR-ellipse slightly shrinks, second, it is subject to a slow precession, and third, on a shorter trajectory the probe arrives at the gravity-assist point a earlier so that Venus is met a bit closer.

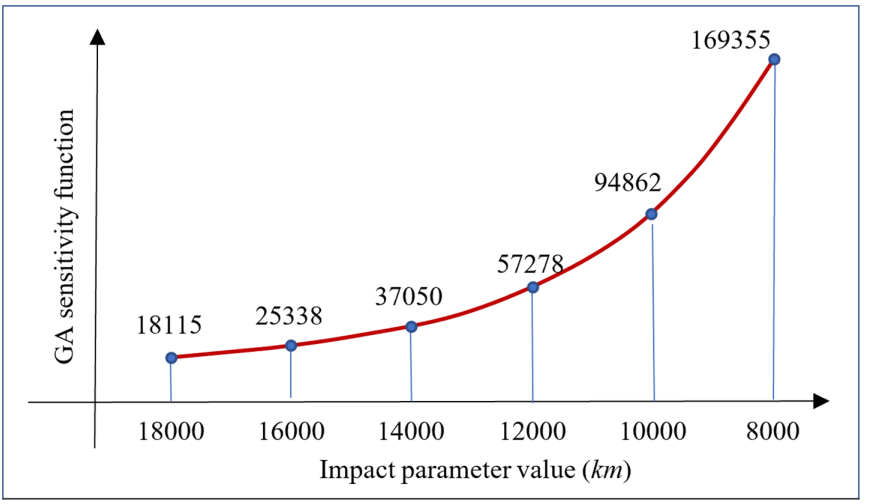


Fig. 2. Math calculation of GA sensitivity as function of IP value.

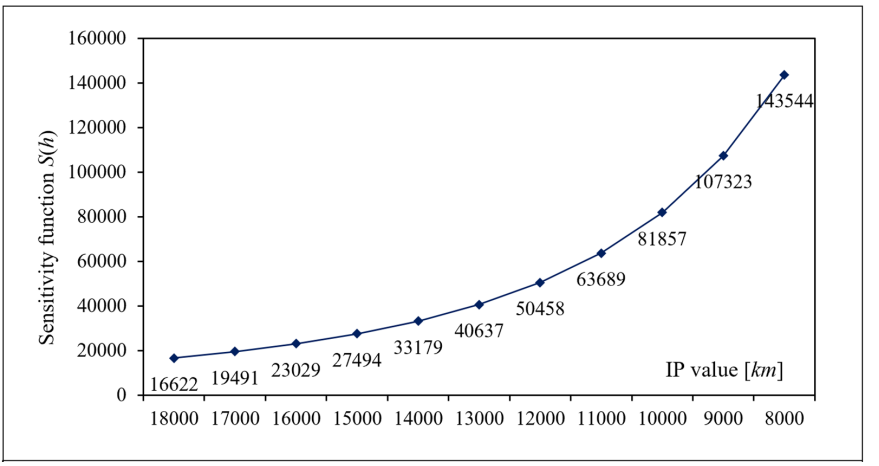


Fig. 3. GA sensitivity function $S(h) = dl/dh$ where dl is FP difference [km] for $dh = 1$ km.

Then under identical initial conditions the probe's pericenter becomes about 40 km shorter. Therefore, if the pericenter is 18 000 km, then at the final point the observer must detect about 800 000 km distance from assigned position. This difference will be much more (millions of km) for shorter pericenter values. These distances are big enough; their experimental detection by the observer should indicate in favor of Einstein's gravity.

Conclusion

The experiment seems not too hard to perform and it does not require excessive costs. Existing computational means allow distinguishing pure gravity effects among other physical agents, influence of other planets, Sun's oblateness, solar wind, etc. The space probe need not be a sophisticated station packed with expensive tools; a primitive space-ball able to send receivable signals is sufficient. We think that many cosmic agencies of different states and private companies can afford this job. We hope, someone will be the first.

References

1. A. P. Yefremov, Sensitivity of the Gravity Assist to Variations of the Impact Parameter, *Grav. Cosmol.* **26**, 118 (2020).
2. A. P. Yefremov and A. A. Vorobyeva, A planet's gravity assist as a powerful amplifier of small physical effects in the Solar system, *Acta Astronautica* **180**, 205, 2021.
3. A. P. Yefremov and A. A. Vorobyeva, A "space-ball" experiment to specify the nature of gravity in the solar system. [Submitted to: *Acta Astronautica* (2021)]

Gravitomagnetic field generation using high permittivity materials in superconducting magnetic energy storage devices

G. V. Stephenson
Seculine Consulting
Cupertino, CA 95014, USA
E-mail: seculine@gmail.com

A method is described for creating a measurable unbalanced gravitational acceleration using a gravitomagnetic field surrounding a superconducting toroid as described by Forward.[1] An experimental superconducting magnetic energy storage toroid configuration of wound superconducting nanowire is proposed to create a measurable acceleration field along the axis of symmetry, providing experimental confirmation of the additive nature of a Lense–Thirring derived gravitomagnetic field. In the present paper gravitational coupling enhancement of this effect is explored using high-permittivity material, as predicted by Sarfatti[2] and his modification to Einstein's general relativity field equations for gravitational coupling in matter.

Keywords: Gravitational; Gravitomagnetic; Lense–Thirring; Superconducting magnetic energy storage; SMES; Nanorods; Nanowires; Super dielectric materials; SDM; Super capacitors; High permittivity; Gravitomagnetic permeability; Gravitational coupling.

1. Introduction

When Forward[1] proposed a gravitomagnetic toroid to produce unbalanced gravitational force in 1962, a prototype was quite impractical. Recent advances in high-temperature superconducting (HTSC) nanorod wires (nanowires)[3,4] has enabled a new class of superconducting magnetic energy storage (SMES) devices operating at current densities sufficient to develop measurable gravitomagnetic fields.

In the present paper, an experimental SMES toroid configuration is proposed that uses a super dielectric material (SDM; e.g., high-permittivity super capacitor) to substantially improve gravitational coupling of mass flow to the curvature of spacetime as predicted by Sarfatti.[2] Depending on the nature of gravitational coupling, it is predicted that a set of standard accelerometers could measure acceleration fields along the axis of symmetry of the toroidal coil, thus providing experimental confirmation of the additive nature of the gravitomagnetic fields, as well as the production of a linear component of the overall acceleration field (Fig. 1).

In the instantiation of Forward's gravitational generation coil described in this paper, superconducting electron flow provides the change in mass current in the toroid. A high-permittivity material such as a SDM[5] is added to the space in the center of the toroid to improve gravitational coupling. Alternatively or additionally high-permittivity insulators may be used as an exterior jacket to the conducting portion of the nanowire to further boost gravitational coupling; gravitomagnetic effects are largest adjacent to the mass flow.

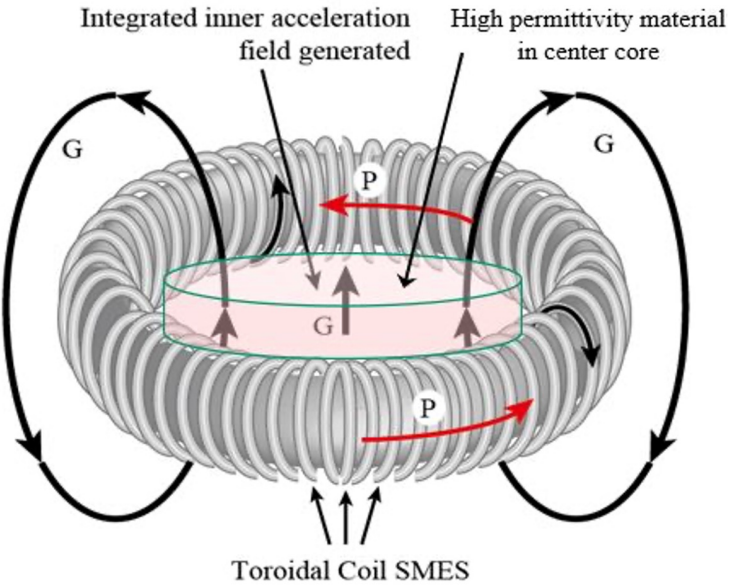


Figure 1. Gravitational force generation coil from Forward[1] with an inspiraling mass current, with a vector potential P, creating gravitomagnetic field G, with high-index material in the center to improve coupling.

2. Gravitomagnetic Force Equation for Toroid Mass Flow

As first developed in Forward,[1] the linear force G_f developed by gravitomagnetic force in the mass flow toroid of Fig. 1 is given by Eq. (1)

$$G_\mathrm{f} = \left(\frac{\eta}{4\pi}\right)\left(\frac{N\dot{T}r^2}{R^2}\right), \tag{1}$$

where η is gravitomagnetic permeability, $\eta = \eta_\mathrm{o}\eta_\mathrm{r}$,[3] N is the number of turns in the toroid coil winding, $\dot{T}$ represents the mass flow, r is the cross-sectional radius of the toroid, and R is the overall radius of the toroid.

Mass flow can be generated by an electrical current by virtue of the electrons' mass in motion. Single-electron mass flow (Fig. 2) is given by mass momentum:

$$T_\mathrm{e} = p_\mathrm{e} = (\Omega \times r)m_\mathrm{e}, \tag{2}$$

where Ω is the angular rate, angular velocity is $v = \Omega \times r$ in the classical case.[6]

Change in mass flow for the single-electron flow shown in Fig. 2 is

$$\dot{T}_\mathrm{e} = \dot{p}_\mathrm{e} = a \cdot m_\mathrm{e} = (\Omega \times v)m_\mathrm{e}. \tag{3}$$

This is equivalent to centripetal force,

$$\dot{T}_\mathrm{e} = F_\mathrm{e} = \frac{m_\mathrm{e}v^2}{r} = m_\mathrm{e}a_\mathrm{e} = m_\mathrm{e}\left(\omega^2 r\right), \tag{4}$$

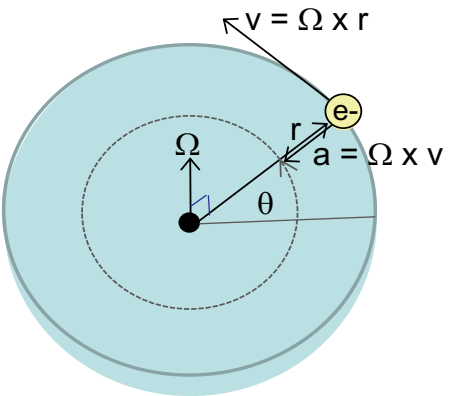


Figure 2. Electron orbit around one loop of SMES toroid.

where ω is the angular rate:

$$\omega = \frac{2\pi}{t_p} = \frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{5}$$

3. Current in Idealized SMES

We now estimate the possible inspiraling current flux enabled by the emerging SMES technology as it relates to the core geometry constraints described in Fig. 3, a toroid with torus geometry. We start with the assumptions needed to calculate the number of turns, N.

From Eq. (1), the torus assumptions made in Fig. 3 can be factored into Eq. (6):

$$G_{\mathrm{f}} = \left(\frac{\eta_{\mathrm{o}}\eta_{\mathrm{r}}}{4\pi}\right) S \left(\frac{N\dot{T}r^2}{R^2}\right), \tag{6}$$

where

G_{f} = gravitomagnetic force
η_{o} = absolute gravitomagnetic permeability
η_{r} = relative gravitomagnetic permeability
N = number of turns in the torus coil
S = Sarfatti scaling factor
$\dot{T}$ = change in mass or mass flow
r = cross-section radius of torus
R = centerline radius of torus

Note that Eq. (6) differs from Forward's original formulation in that it includes the scaling factor S which accounts for the scaling of gravitational coupling due to high-index matter or metamaterials, which was first expressed by Sarfatti as

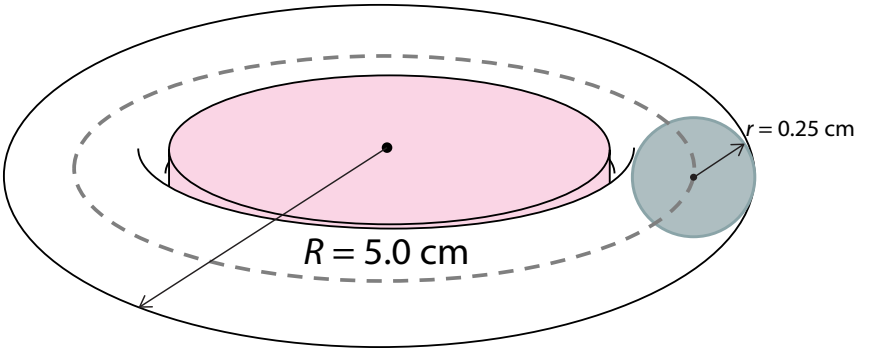


Figure 3. Toroid with a torus-shaped core geometry, including high-permittivity scaling matter in center.

an additional zero-rank tensor in Einstein's field equations[2,7] based on combining Lentz's[8] work with that of Medina and Stephany:[9]

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \left(\frac{8\pi G}{c^4}\right) ST_{\mu\nu}, \tag{7}$$

where

$R_{\mu\nu}$ = Ricci curvature tensor,
$Rg_{\mu\nu}$ = Ricci scalar curvature,
G = gravitational constant
c = speed of light
S = Sarfatti scaling tensor
$T_{\mu\nu}$ = Stress Energy Tensor

For the purposes of describing an idealized case with a realistic geometry, we describe a device bounded by a 10 cm toroid (outer mold line diameter), shown in Fig. 3 and Fig. 4, with a 0.5 cm cross-sectional diameter. Furthermore, 16 sectors are defined as shown in Fig. 4.

We further add additional assumptions regarding conductor wrapping around the toroid to determine constraints on the number of conductive loops that can be accommodated using the described technology.[3] As shown in Fig. 5 via cross section, we assume here a conductor winding depth of 0.1 cm.

The inner edge of a single segment of the cross section, C_s, is 1/16 of the toroid's centerline circumference:

$$C_s = \frac{2\pi r_i}{16} = 1.77\,\text{cm} \tag{8}$$

With a depth D, the minimum inner loop cross-sectional area can be described as

$$A_{sec} = D \cdot C_s = (0.1\,\text{cm})(1.77\,\text{cm}) = 0.177\,\text{cm}^2. \tag{9}$$

This area is shown packed with conductors in Fig. 6 in depth and along the sector circumference.

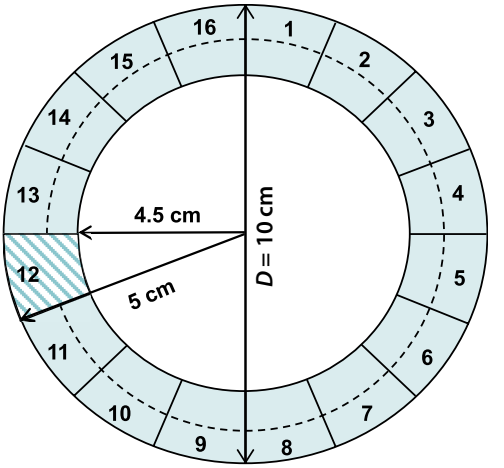


Figure 4. Torus section definition.

Assuming each nanowire conductor has a diameter d_c of 100 µm, the cross-sectional area of each conductor will be given by

$$A_c = \pi r_c^2 = 7.854 \times 10^{-9}\,\mathrm{m}^2. \tag{10}$$

For packing nanowire conductors in a cross-sectional area described in Fig. 6, assume as a worst case a rectangular area described by the shortest edges such that

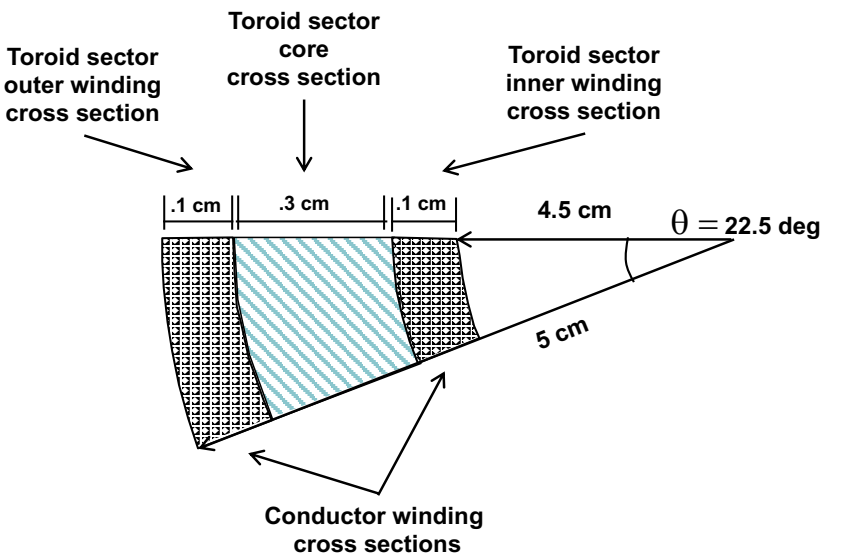


Figure 5. Conductor cross section in each sector.

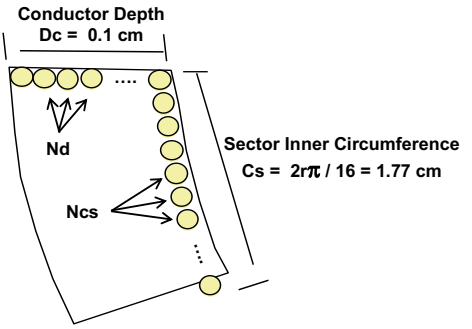


Figure 6. Conductor packing in each sector.

a number of conductors in depth, N_{d}, may be packed in one dimension, with the number of conductors, N_{cs}, packed in the other dimension. These packing counts may be calculated in Eq. (11) and Eq. (12) as follows:

$$N_{\mathrm{d}} = D_{\mathrm{c}}/d_{\mathrm{c}} = \frac{0.1\,\mathrm{cm}}{100\,\mu\mathrm{m}} = 10 \tag{11}$$

$$N_{\mathrm{cs}} = C_{\mathrm{s}}/d_{\mathrm{c}} = \frac{1.77\,\mathrm{cm}}{100\,\mu\mathrm{m}} = 177 \tag{12}$$

The total number of windings by sector will therefore be the product $N_{\mathrm{d}} \times N_{\mathrm{cs}}$:

$$N_{\mathrm{sec}} = N_{\mathrm{d}} \cdot N_{\mathrm{cs}} = (10)(177) = 1770. \tag{13}$$

With 16 sectors, the total number of windings for the entire toroid will be

$$N = 16N_{\mathrm{sec}} = 28{,}320. \tag{14}$$

What is $\dot{T}$ with the forgoing assumptions? In this idealized case electrons circulate about a coil of circumference c_{r}, or slightly larger:

$$c_{\mathrm{r}} = 2\pi r = 1.57\,\mathrm{cm}. \tag{15}$$

3.1. *Electron motion*

Assume further a supply voltage of 16 kV, resulting in 16KeV of kinetic energy for each electron, which corresponds with the upper limit of a nonrelativistic case, where $v = 0.25\mathrm{c}$, so that $\gamma = 1.06 \approx 1.0$.

Then from Eq. (4) for nonrelativistic circular motion, the vector change in DC current flow is

$$\dot{T}_{\mathrm{e}} = \frac{m_{\mathrm{e}} v^2}{r}, \tag{16}$$

which, for a single electron, has the following values

$$m_{\mathrm{e}} = \text{mass of the electron} = 9.11 \times 10^{-31}\,\mathrm{kg}$$
$$v = \text{velocity of the electron} = 0.25\mathrm{c} = 0.75 \times 10^{8}\,\frac{\mathrm{m}}{\mathrm{s}}$$
$$r = 0.25\,\mathrm{cm}$$

for the assumed geometry

The angular acceleration of the electron is

$$a_{\mathrm{e}} = \frac{v^2}{r} = 2.25 \times 10^{18}\,\frac{\mathrm{m}}{\mathrm{s}^2} \tag{17}$$

The change in mass flow represents centripetal acceleration in the case of circular motion:

$$\dot{T}_{\mathrm{e}} = m_{\mathrm{e}} \cdot a_{\mathrm{e}} = 20.48 \times 10^{-13}\,\mathrm{N}. \tag{18}$$

Eq. (18) corresponds to the change in mass flow for one electron in one loop of coil. Total mass flow change is therefore the mass flow change per electron times the number of electrons:

$$\dot{T} = \dot{T}_{\mathrm{e}} \cdot N_{\mathrm{e}}. \tag{19}$$

3.2. *Electron current*

What is the number of electrons, N_{e}, in motion in one loop (part of the mass flow) at a given time for an assumed velocity of $v = .25\mathrm{c}$? N_{e} in one loop can be described by the current I times the period of a single loop circulation Δt:

$$N_{\mathrm{e}} = I \cdot \Delta t, \tag{20}$$

where the period of an orbit can be described by

$$\Delta t = \frac{c_{\mathrm{r}}}{v} = \frac{2\pi r}{v} = 0.2094\,\mathrm{ns}. \tag{21}$$

What is the possible current inside the idealized device for the case where the entire winding is in series? We assume the max current stays below the critical current density of $250\,\frac{\mathrm{MA}}{\mathrm{m}^2}$.[3,4]

Current is limited by the maximum permissible current density and the cross section of the conductor:

$$I = J \cdot A_{\mathrm{c}}, \tag{22}$$

where J is material dependent. For the nanowire assumed in Ref. 3, $J = 250\,\frac{\mathrm{MA}}{\mathrm{m}^2}$ and the cross-sectional area, $A_{\mathrm{c}} = 7.854 \times 10^{-9}\,\mathrm{m}^2$, as given in Eq. (10). Therefore, the maximum current for this conductor diameter is $I = 1.96\,\mathrm{A}$.

Expanding on Eq. (22), the number of electrons N_{e} in circulation in one loop may be calculated by noting that there are 6.2415×10^{18} electrons per Coulomb:

$$N_{\mathrm{e}} = \left(\frac{\text{electron}}{\text{Coulomb}}\right) I \left(\frac{\text{Coulomb}}{\text{second}}\right) \cdot \Delta t = 5.12 \times 10^{9}\,\text{electron} \tag{23}$$

4. Forces in Idealized SMES

Expressing Eq. (19) as force per electron times the number of electrons in motion in one loop:

$$\dot{T} = \dot{T}_{\mathrm{e}} \left(\frac{\mathrm{Newton}}{\mathrm{electron}} \right) \cdot N_{\mathrm{e}} = 5.25\,\mathrm{mN} \tag{24}$$

Thus, each loop experiences about 5.25 mN of integrated centripetal force ($\dot{T}$) due to the electrons in circulation within. We now describe the scale factor to couple this force to the gravitomagnetic effect. Revisiting Eq. (6), which describes the overall linear force developed at the center of the toroidal coil, total gravitomagnetically developed force will be

$$G_{\mathrm{f}} = (\eta_{\mathrm{o}}\eta_{\mathrm{r}})\,(S) \left(\frac{N\dot{T}r^2}{4\pi R^2} \right) = (\eta_{\mathrm{o}}\eta_{\mathrm{r}})\,(S)(0.0295\,\mathrm{N}), \tag{25}$$

where known variables have been grouped on the right and unknown variables have been collected on the left. What are the correct values for η_{o} and η_{r}?

4.1. *Predicted force using gravitational potential scaling*

If η_{o} goes as $G/2\mathrm{c}$ as does gravitomagnetic potential (Ref. 10, Eq. (1.5)), then

$$\eta_{\mathrm{o}} = -\frac{G}{2\mathrm{c}} = 1.11 \times 10^{-19}. \tag{26}$$

This leads to $G_{\mathrm{f}} = (\eta_{\mathrm{r}})3.275 \times 10^{-21}$. Values of η_{r} are experimentally unknown at this time. However, if values of η_{r} track values of ϵr^2, then values as high as $\eta_{\mathrm{r}} = 10^{16}$ may be possible, yielding $G_{\mathrm{f}} = 0.03275$ mN. For a test mass of 1 kg this is equivalent to an easily measurable 3.35 μG. This is for a scaling factor of $G/2\mathrm{c}$. If the scaling factor is G/c^2, then the effect drops to an unmeasurably low 1.12×10^{-14} G, even for $\eta_{\mathrm{r}} = 10^{16}$.

4.2. *Predicted force using gravitational field scaling*

If the scaling factor is G/c^3 or G/c^4, then the effect is negligible for $\eta_{\mathrm{r}} = 10^{16}$, so a much higher-index material will need to be found to improve coupling, or other means must be found to improve either the S tensor or the gravitomagnetic permeability η_{r}.

Consider the case where η_{o} goes as $8\pi\mathrm{G}/\mathrm{c}^4$ as does the coupling constant in Einstein's Field Equations:

$$\eta_{\mathrm{o}} = \frac{8\pi G}{\mathrm{c}^4} = 2.07 \times 10^{-43}, \tag{27}$$

which gives the total gravitational force developed as

$$G_{\mathrm{f}} = (\eta_{\mathrm{o}}\eta_{\mathrm{r}})\,(S)(0.0295\,\mathrm{N}) = (\eta_{\mathrm{r}})\,(S)0.61 \times 10^{-44}\,\mathrm{N}. \tag{28}$$

A value of $\eta_r = 10^{16}$ gives $G_f = (S)4.43 \times 10^{-23}\,\mathrm{N}$, which requires a scalar value of S of at least an additional 10^{16} to obtain a measurable value for G_f. For an isotropic material, S can be written as[7]

$$S = \frac{1}{2}\left(\epsilon^2 + \frac{1}{\mu^2}\right) = \frac{1}{2}\frac{n^4+1}{\mu^2}. \tag{29}$$

Therefore, either a permittivity of $\epsilon = 10^8$ or an index of $n = 10^4$ with $\mu = 1$ would be sufficient to provide the necessary 16 orders of magnitude in coupling improvement. However, care should be taken that effects captured by the S factor are not duplicative of those accounted for in the value of η_r.

Super dielectrics can provide up to $\epsilon = 10^9$,[5] which, if we simultaneously had $\eta_r = 10^{16}$ via some unrelated phenomena, would improve the overall generated gravitational force to a marginally measurable level:

$$G_f = (\eta_r)\,(S) \times 0.61 \times 10^{-44} = 0.061\,\mathrm{nN}. \tag{30}$$

Alternatively metamaterials could also be investigated to establish whether non-isotropic materials may create a more advantageous gravitational coupling.

4.3. *Lenz's law implications*

Is there an equivalent of Lenz's Law for gravitomagnetics? Consider a hypothetical case where a Forward toroid capable of supporting sufficient current and coupling to develop 1 G of acceleration. As a thought experiment, imagine disconnecting the power supply and shorting the toroidal coil to itself, such that it is one continuous closed loop of nanowire, and equip a flying craft with such a coil.

At the surface of the Earth such a properly oriented toroidal coil would be immersed in a 1 G gravitational field. Should not such a field generate a counter-current in a toroidal coil capable of supporting a current of that level, essentially providing a self-powered counter field? If so, a craft so equipped would be able to hover using a counter-field equal but opposite to the Earth's gravitational field without need of additional power sources.

Similarly, additional closed-loop toroidal coils, properly oriented, could also be used to "current charge," collecting energy from a static gravitational field, potentially providing power that could be directed to other propulsive or nonpropulsive functions of such a hypothetical craft.

5. Conclusions

An argument is made for using high-permittivity materials in the core or donut hole or nanowire insulators of an SMES toroid to improve gravitomagnetic coupling for the creation of an acceleration field, possibly of measurable amplitude. Improved coupling will be beneficial for shrinking device scale and complexity and overcoming the very weak coupling between mass flow and gravitomagnetic spacetime curvature.

Acronyms

HTSC – high temperature superconductor
SDM – super dielectric material
SMES – superconducting magnetic energy storage

Acknowledgments

The author wishes to acknowledge Jack Sarfatti for helpful correspondence on his insights into altering Einstein's field equations in the presence of matter and metamaterials. The financial support of Seculine Consulting is gratefully acknowledged.

References

1. R. Forward, Guidelines to antigravity, *Proc. Gravit. Res. Found.* (New Boston, 1962).
2. J. Sarfatti, Explaining US Navy close encounters with tic tac UAV metric engineering, *Proc. Estes Park Adv. Prop. Workshop* (Space Studies Institute, North Hollywood, 2020).
3. G. Stephenson, W. Rieken, and A. Bhargava, Extended cases of laboratory generated gravitomagnetic field measurement devices, *J. High Energy Phys. Gravit. Cosmol.* **5**, 375–394 (2019), https://doi.org/10.4236/jhepgc.2019.52021.
4. W. Rieken, A. Bhargava, R. Horie, J. Akimitsu, and H. Daimon, $YBa_2Cu_3O_x$ superconducting nanorods, *Jpn. J. Appl. Phys.* **57**, Aritcle 023101 (2018).
5. S. Fromille and J. Phillips, Super dielectric materials, *Materials*, **7** (12), 8197–8212 (2014), https://doi.org/10.3390/ma7128197.
6. J. M. Knudsen and P. G. Hjorth, *Elements of Newtonian Mechanics: Including Nonlinear Dynamics* 3rd edn. (Springer, New York, 2000).
7. G. Stephenson, 2000–2020 summary of gravitational work, *APEC Proc.*, https://doi.org/10.13140/RG.2.2.23735.34720.
8. E. W. Lentz, *Breaking the Warp Barrier: Hyper-Fast Solitons in Einstein–Maxwell-Plasma Theory* (2020), https://arxiv.org/abs/2006.07125.
9. R. Medina and J. Stephany, The Energy–Momentum Tensor of Electromagnetic Fields in Matter (2017), https://arxiv.org/abs/1703.02109v1.
10. B. Mashhoon, *Gravitoelectromagnetism: A Brief Review* (2008), https://arxiv.org/abs/gr-qc/0311030.

Large ring laser gyroscopes: Geometry stabilization and laser control

U. Giacomelli
GSSI
L'Aquila, Italy
E-mail: umberto.giacomelli@gssi.it
www.gssi.it

N. Beverini, G. Carelli, D. Ciampin, A. Di Virgilio, F. Fuso, E. Maccioni and P. Marsili
Dipartimento di Fisica
Università di Pisa
Pisa, Italy

A. Di Virgilio and A. Simonelli
INFN-Sez Pisa
Pisa, Italy

Ring laser Gyroscopes (RLG) are very versatile devices that find application in many fields as navigation, seismology and geophysics. Moreover, thanks to their sensitivity and accuracy, in the last years they have been used in fundamental physics research field.

GINGER (Gyroscopes IN GEneral Relativity) research group aims to exploit a large RLG to test general relativity theory. Our research team has two working RLG, both with a square shape, one installed in Pisa and named GP2. (1.6 m side), and the other installed in the INFN underground laboratory of Gran Sasso near L'Aquila named GINGERINO (3.6 m side). The final goal of GINGER is to measure the Earth rotation rate with enough precision to take into consideration general relativity predicted corrections.

To reach this target, one of the requirements is the stability of the laser and the optical cavity of the RLG. We will show the last developed techniques aimed to satisfy this stability requirement. Working on GP2 we have tested two different techniques to control the ring shape. One is based on the stabilization of the two Fabry-Pèrot resonators formed along the square diagonals by the opposite mirrors of the RLG. The other, consist in controlling the ring perimeter by monitoring its free spectral range through a beet-note between one of the counterpropagating beams and a frequency stabilized laser source. We will show the characteristics, the potentialities and the tests of these two methods.

Keywords: Large Frame Ring Laser, Sagnac Effect, frequency control, Laser stabilization

1. Introduction

Large frame ring laser gyroscopes (RLG), which exploit Sagnac effect, are the most sensitive devices for detecting absolute angular motions in a huge range of frequency, extending from kHz down to DC.[1] Thanks to their sensitivity and accuracy, in the last years they have been used in the fundamental physics research field.

In this framework, the project GINGER (Gyroscopes IN GEneral Relativity) of the Istituto Italiano di Fisica Nucleare (INFN) was developed in 2010.[2,3] The goal of the project is to measure the Lense-Thirring and De Sitter effects using an on

Earth device via a very precise and accurate measurement of the Earth rotation rate. Binding rigidly a RLG to the ground it is possible to exploit the Sagnac effect to measure the Earth rotation rate. Due to this effect, two counter-propagating beams that circulate in a rotating close path have an optical frequency difference proportional to the rotation rate. This frequency difference, known as the Sagnac frequency (f_S), is linked to the reference frame rotation rate via the geometrical characteristics of the ring as descibed in Eq. (1)

$$f_S = \frac{4A}{\lambda P}\vec{\Omega} \cdot \hat{n} \tag{1}$$

where A and P are respectively the area and the perimeter of the optical path, λ is laser wavelength without the path rotation, $\vec{\Omega}$ is the rotation rate of the reference frame and $\hat{n}$ is the normal area versor of the optical path.

It is possible to demonstrate that Eq. (1) is still valid in General Relativity (GR) with only one difference: $\vec{\Omega}$.[4] In fact, in the relativistic formulation, $\vec{\Omega}$ includes all the gravitational field effects, which is exactly what we want to measure.

$$\vec{\Omega} = \vec{\Omega}_{\oplus} + \vec{\Omega}_G + \vec{\Omega}_{LT} \tag{2}$$

$\vec{\Omega}_{\oplus} = 7.29 \times 10^{-5}$ rad/s is the Earth instantaneous angular rotation velocity, $\vec{\Omega}_G$ is the geodetic (or De Sitter effect), $\vec{\Omega}_{LT}$ is the Lense-Thirring effect.

These terms are estimated to be, respectively, $6.98 \cdot 10^{-10}\Omega_{\oplus}$ the geodetic part and $2.31 \cdot 10^{-10}\Omega_{\oplus}$ the Lense-Thirring one.[5,6] This estimation gives the sensitivity level required to be able to measure the general relativity correction upon the Earth rotation rate.

GINGER research group aims to reach this level of sensitivity with an array of large frame RLGs. In order to reconstruct the Earth rotation rate vector with enough precision to evaluate the De Sitter and Lense-Thirring contribution. To obtain this goal the also the ratio $4A/\lambda P$ of Eq. (1) must be known and stable at a level of 10^{-10}. This ratio is known as *scale factor* (for simplicity in the following we will refer to this quantity as scale factor or k). In the following, we will describe two different techniques to stabilize this value.

2. GP2 and the geometry control

Our research group has two RLGs, the first one, named GP2[7–9] installed in the INFN (Istituto Nazionale di Fisica Nucleare) laboratory in Pisa, and the second one, GINGERINO[10–14] installed in the underground laboratory of LNGS (Laboratori Nazionale del Gran Sasso) near L'Aquila.

This second one is our prototype of high precision large frame RLG. It is a square ring of 3.6m side located inside a quiet environment like those of the underground laboratories of LNGS inside the Gran Sasso mount. It has been built to test the quality of this kind of site for the installation of a device for GR measurements.

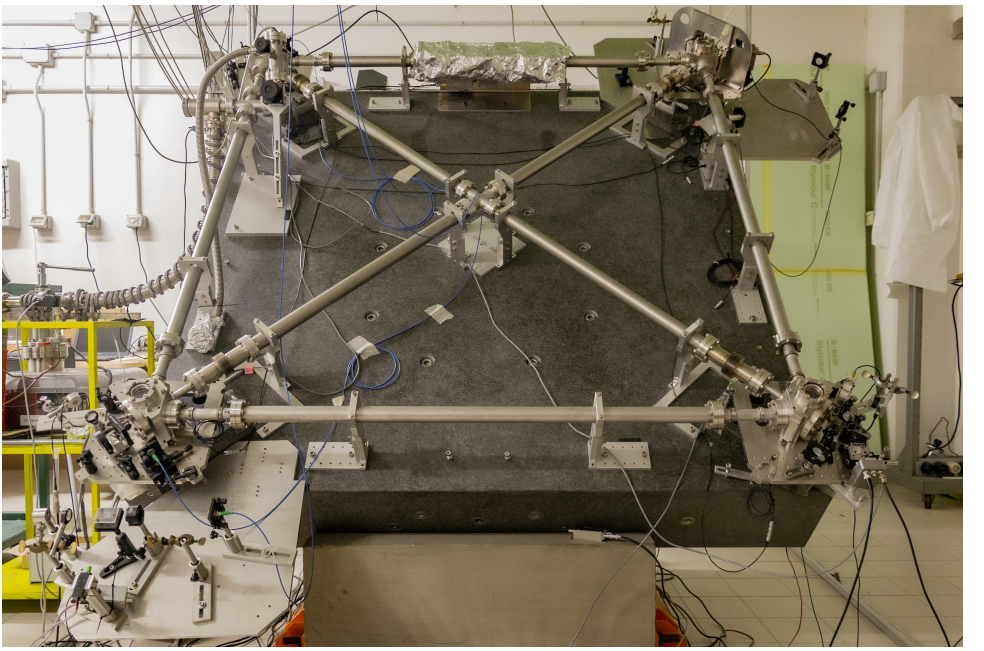


Fig. 1. Pictures of GP2.

Moreover, it has also shown that can be a suitable instrument also for geophysics ad geodesy.[15–17]

In this work we will focus on the first one GP2, that is our *test* ring laser (see Fig. 1). It is a square ring of 1.6m side, tilted of $\sim 45^{o}$ in order to maximize the scalar product in Eq. (1). In addition to the square cavity, it has other two resonance cavities: the diagonals, and 4PZT actuators (one for each corner) able to move the whole corner along the diagonal direction. The structure of both our RLGs is etherolithic. This means that active geometry stabilization is essential to fix the scale factor value during the measurement period.

We have developed two different methods to control the ring geometry, the first is based on controlling the length of square diagonals and the second is based on stabilizing its perimeter length.

2.1. *Diagonal Control*

First of all, we have designed a mathematical model to parameterize all possible deformation that can affect the ring shape.[8] Basing this model on the diagonals deformation we have identified six fundamental deformations. Tn the hypothesis of small perturbation of a perfect square we can reconstruct any kind of shape perturbation using these six fundamental deformations. Using this parametrization

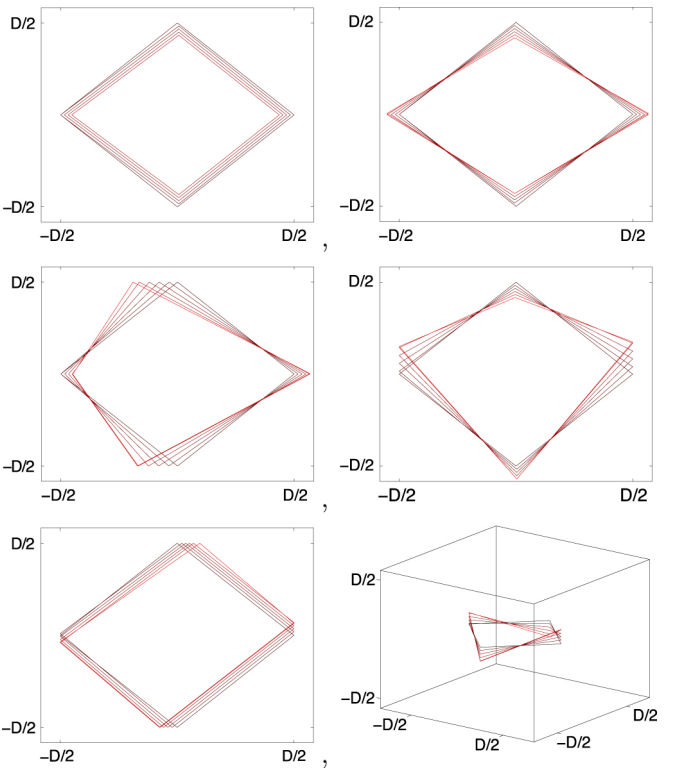


Fig. 2. Graphical representation of the optical paths six fundamental deformation of ring shape. First row: diagonal common mode stretching E_1 (left); differential mode stretching E_2 (right). Second row: shear planar deformations E_3 (left), E_4 (right). Third row: diagonal tilt E_5 (left); out of plane tilt E_6 (right). From Fig. 2 of [8]

we have recalculate the scale factor obtaining:

$$k = \frac{L}{\lambda}\left[1 - \frac{\tau_1}{\sqrt{2}L} + \frac{\tau_1^2}{2L^2} - \frac{\tau_2^2}{4L^2} - \frac{2L+\sqrt{2}r}{4L\left(r-\sqrt{2}L\right)^2}\left(\tau_3^2+\tau_4^2\right) - \frac{L+2\sqrt{2}r}{L\left(4r-\sqrt{2}L\right)^2}\tau_6^2\right] \tag{3}$$

were each τ_α represent the coefficients that quantify the magnitude of each deformation, L the is the unperturbed ring side length and r is the curvature radius of each mirror. The assumption of of small perturbation means that $10^{-2} < \tau_\alpha/L < 10^{-1}$. In view of this, from Eq. (3) it is clear that an accuracy of 1 part in 10^{10} on k can be reached only if the relative amplitude of E_1 deformation is 10^{-10} or less. This means that stabilizing only the diagonals length is enough to reach our purpose.

We have realized an opto-electronic system able to transfer the frequency stability of an He-Ne iodine stabilized laser to the length stability of a Fabry-Pérot cavity (the diagonals). Firstly we have tested this system on two probe cavities on optical bench[7] and then we have implemented it on GP2.[9]

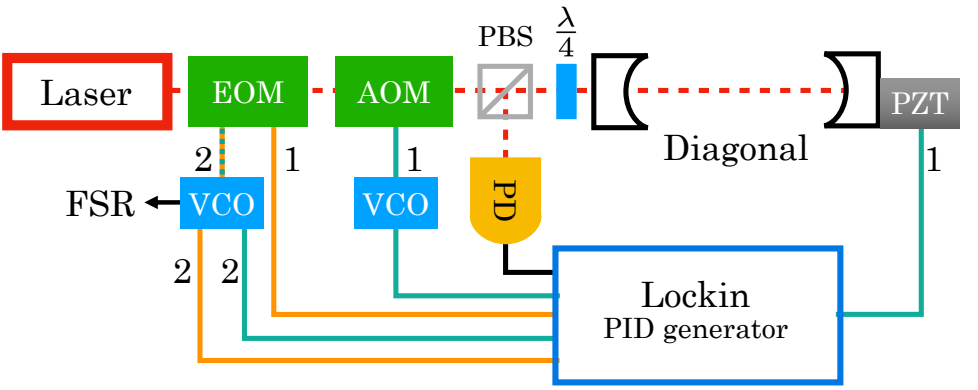


Fig. 3. Control loop scheme. Orange lines are the modulation signals, green lines are the corrections generated by the feedback system and black line is the modulated signal. Lines labeled with 1 are the first control loop and those ones with 2 are the second.

Fig. 3 shows a schematization of a control loop for a single diagonal. The control system is divided into two loops (see lines labeled with 1 and 2 in Fig. 3). The first is the one that fixes the length of the diagonal via two different actuators the piezoelectric (PZT) and the Acousto-Optic Modulator (AOM). The second one is used to measure the cavity Free Spectral Range (FSR) in order to test the control system efficiency.

The results of the measurement with this control technique are summarized in Fig. 4. The plots show the comparison between the measured Earth rotation rate estimated by the reconstructed Sagnac frequency and the Earth rotation rate

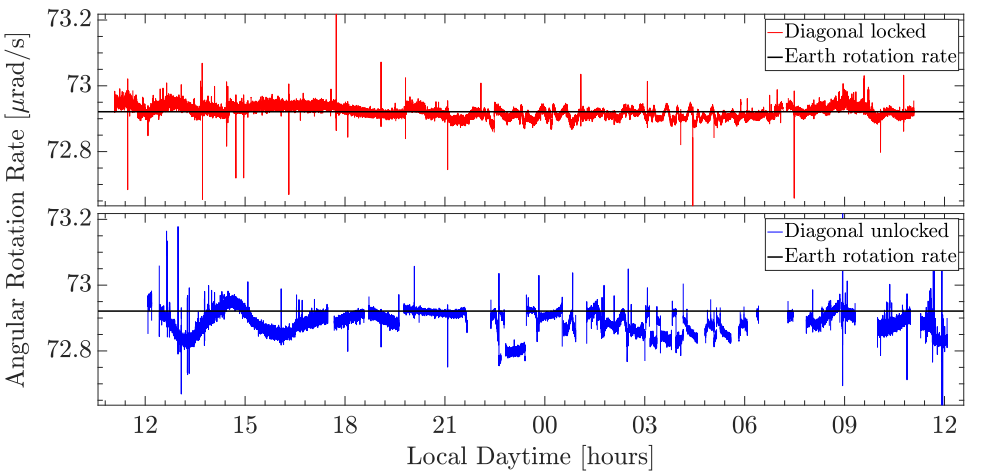


Fig. 4. Reconstructed value of Earth rotation rate from Sagnac frequency in μrad/s compared with the known value while the diagonals are locked (TOP) and free-running (BOTTOM).

nominal value. Both quantities are expressed in μrad/s. On top, the comparison when the diagonals are locked is shown, while on the bottom when the diagonals are unlocked.

A clear improvement due to the presence of the cavity stabilization system is the increase of the duty-cycle of the measurements. In our case, the duty-cycle is defined as the ratio between the period during which it is possible to reconstruct the Sagnac frequency, divided by the entire acquisition time. It is straightforward to see the effect of the stabilization on the so-defined duty-cycle. The "holes" present in the bottom plot are generated by the period of split mode of the RLG. In this condition, we have a duty-cycle of about 78%. This value increases up to 99% when we active the lock as the top plot shows. In addition, the presence of the control loop increases the effectiveness of the frequency reconstruction algorithm reducing the spread of the data with respect to the Earth rotation rate known value.

A more detailed discussion of this technique can be found in [9, 18].

2.2. *Perimeter Control*

Our studies demonstrate that the diagonal control technique is the main way to obtain the necessary stability for GR measurement. Nevertheless, we explored other geometry control methods. One of them consists of keeping fixed the ring perimeter.

There are two possible ways to measure the ring perimeter, the first consists of generating a beat-note between one of the two counter-propagating beams and a frequency stabilized laser. The second is the self-beat-note technique. If the laser inside the ring has enough power to excite the fundamental and the first longitudinal mode, using a fast photodiode we can acquire the beat-note of these two modes that

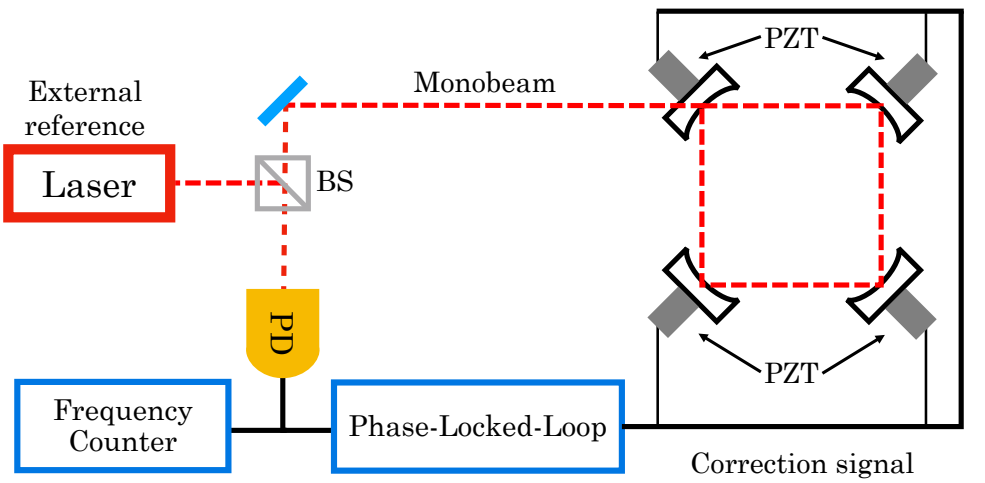


Fig. 5. Scheme of perimeter lock loop via comparison with an external laser source. The one that exploit the self-beat-note is analogous but without the Beam Splitter (BS) and the external laser.

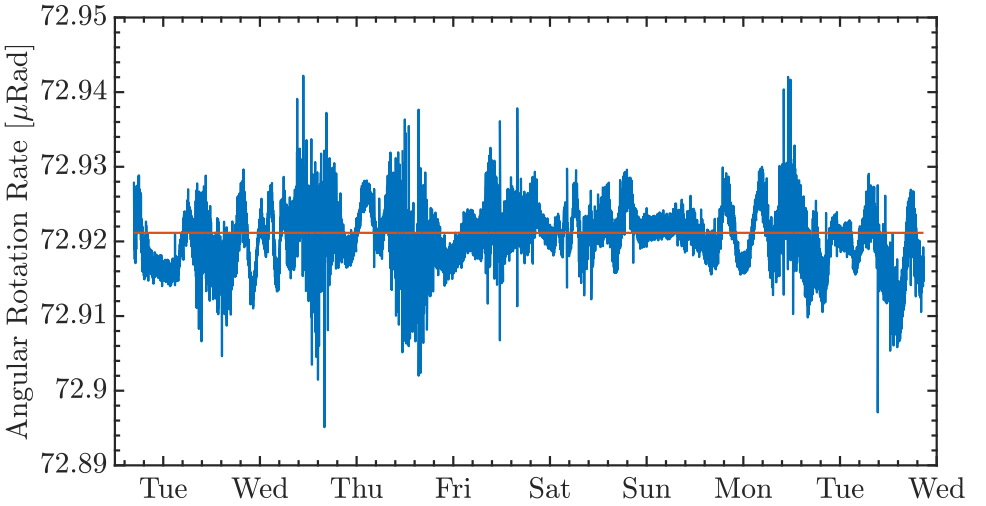


Fig. 6. Reconstructed value of Earth rotation rate from Sagnac frequency in μrad/s compared with the known value while the perimeter is locked.

is exactly the Free Spectral Range (FSR) of the cavity. This method has an intrinsic lower level of stability than the first one. So we chose the external reference laser method.

Fig. 5 shows a schematization of the perimeter control loop. The beat-note between monobeam and reference laser is acquired from a fast photodiode and sent to a frequency counter, used as a monitor, and to a Phased-Locked-Loop (PLL), used to generate the correction signal.

As in the case of diagonal control, we tested the perimeter lock loop acquiring the Sagnac frequency. Fig. 6 shows the result of this test. Also in this case we have both the benefits present in the diagonals control method. The spread of the data around the nominal value of Earth rotation rate is about 0.3‰ and the duty-cycle close to 99%.

3. Conclusions

In this work, we have shown two different methods to stabilize the geometry of a large frame RLG. The diagonal control method is the one suitable for GR application. Our mathematical model demonstrates that we can fix the scale factor value simply by controlling the diagonal length. Also, preliminary tests on GP2 showed important enhanced in the quality of the Sagnac frequency data in view of an application on a new generation device dedicated to measuring De Sitter and Lense-Thirring contribution to the Earth rotation rate.

Despite it is not the chosen one for GR measurement, also the perimeter control method has some interesting features. It is easier to realize, especially if we use the

self-beat-note technique. Moreover, as in the case of diagonal control, it increases the duty-cycle of an RLG at a level neat to 100% and it improves the Sagnac frequency data quality. Thanks to these it can be useful for the geological and geophysical application of an RLG. In this field, the requested stability is lower than the one of GR.

References

1. K. U. Schreiber and J.-P. R. Wells, Invited Review Article: Large ring lasers for rotation sensing, *Review of Scientific Instruments* **84**, p. 041101 (4 2013).
2. A. Di Virgilio, K. U. Schreiber, A. Gebauer, J.-P. R. We, A. Tartaglia, J. Belfi, N. Beverini and A. Ortolan, A laser gyroscope system to detect the gravito-magnetic effect on Earth, *International Journal of Modern Physics D* **19**, 2331 (12 2010).
3. A. Di Virgilio, J. Belfi, W.-T. Ni, N. Beverini, G. Carelli, E. Maccioni and A. Porzio, GINGER: A feasibility study, *The European Physical Journal Plus* **132**, p. 157 (4 2017).
4. F. Bosi, G. Cella, A. Di Virgilio, A. Ortolan, A. Porzio, S. Solimeno, M. Cerdonio, J. P. Zendri, M. Allegrini, J. Belfi, N. Beverini, B. Bouhadef, G. Carelli, I. Ferrante, E. Maccioni, R. Passaquieti, F. Stefani, M. L. Ruggiero, A. Tartaglia, K. U. Schreiber, A. Gebauer and J.-P. R. J.-P. R. Wells, Measuring gravitomagnetic effects by a multi-ring-laser gyroscope, *Physical Review D* **84**, p. 122002 (12 2011).
5. A. Tartaglia, A. Di Virgilio, J. Belfi, N. Beverini and M. L. Ruggiero, Testing general relativity by means of ring lasers: Ring lasers and relativity, *European Physical Journal Plus* **132**, 1 (2 2017).
6. A. Di Virgilio, J. Belfi, F. Bosi, F. Morsani, G. Terreni, N. Beverini, G. Carelli, U. Giacomelli, E. Maccioni, A. Ortolan, A. Porzio, C. Altucci, R. Velotta, A. Donazzan, G. Naletto, D. Cuccato, A. Beghi, M. Pelizzo, M. Ruggiero, A. Tartaglia, G. De Luca and G. Saccorotti, The GINGER Project, in *Nuclear and Particle Physics Proceedings*, 10 2017.
7. J. Belfi, N. Beverini, D. Cuccato, A. Di Virgilio, E. Maccioni, A. Ortolan, R. Santagata, A. D. Virgilio, E. Maccioni, A. Ortolan and R. Santagata, Interferometric length metrology for the dimensional control of ultra-stable ring laser gyroscopes, *Classical and Quantum Gravity* **31**, p. 225003 (11 2014).
8. R. Santagata, A. Beghi, J. Belfi, N. Beverini, D. Cuccato, A. Di Virgilio, A. Ortolan, A. Porzio and S. Solimeno, Optimization of the geometrical stability in square ring laser gyroscopes, *Classical and Quantum Gravity* **32**, p. 055013 (3 2015).
9. N. Beverini, G. Carelli, A. Di Virgilio, U. Giacomelli, E. Maccioni, F. Stefani and J. Belfi, Length measurement and stabilization of the diagonals of a square area laser gyroscope, *Classical and Quantum Gravity* **37**, p. 065025 (3 2020).
10. J. Belfi, N. Beverini, F. Bosi, G. Carelli, D. Cuccato, G. De Luca, A. Di Virgilio, A. Gebauer, E. Maccioni, A. Ortolan, A. Porzio, R. Santagata, A. Simonelli and G. Terreni, First Results of GINGERino, a deep underground ringlaser, *Review of Scientific Instruments* **88**, p. 034502 (1 2016).
11. J. Belfi, N. Beverini, F. Bosi, G. Carelli, D. Cuccato, G. De Luca, A. Di Virgilio, A. Gebauer, E. Maccioni, A. Ortolan, A. Porzio, G. Saccorotti, A. Simonelli and G. Terreni, Deep underground rotation measurements: GINGERino ring laser gyroscope in Gran Sasso, *Review of Scientific Instruments* **88**, p. 034502 (3 2017).
12. J. Belfi, N. Beverini, G. Carelli, A. D. Virgilio, U. Giacomelli, E. Maccioni, A. Simonelli, F. Stefani, G. Terreni, A. Di Virgilio, U. Giacomelli, E. Maccioni,

A. Simonelli, F. Stefani and G. Terreni, Analysis of 90 day operation of the GINGERINO gyroscope, *Appl. Opt.* **57**, 5844 (10 2018).
13. A. D. V. Di Virgilio, A. Basti, N. Beverini, F. Bosi, G. Carelli, D. Ciampini, F. Fuso, U. Giacomelli, E. Maccioni, P. Marsili, A. Ortolan, A. Porzio, A. Simonelli and G. Terreni, Underground Sagnac gyroscope with sub-prad/s rotation rate sensitivity: Toward general relativity tests on Earth, *Physical Review Research* **2**, p. 032069 (9 2020).
14. S. Capozziello, C. Altucci, F. Bajardi, A. Basti, N. Beverini, G. Carelli, D. Ciampini, A. D. V. Di Virgilio, F. Fuso, U. Giacomelli, E. Maccioni, P. Marsili, A. Ortolan, A. Porzio, A. Simonelli, G. Terreni and R. Velotta, Constraining theories of gravity by GINGER experiment, *Eur. Phys. J. Plus* **136**, p. 394 (2021).
15. A. Simonelli, J. Belfi, N. Beverini, A. D. V. Di Virgilio, U. Giacomelli, G. De Luca and H. Igel, Love waves trains observed after the MW 8.1 Tehuantepec earthquake by an underground ring laser gyroscope, in *AGU Fall Meeting, S33G-2954* (New Orleans, 2017).
16. A. Simonelli, H. Igel, J. Wassermann, J. Belfi, A. Di Virgilio, N. Beverini, G. De Luca and G. Saccorotti, Rotational motions from the 2016, Central Italy seismic sequence, as observed by an underground ring laser gyroscope, *Geophysical Journal International* **214**, 705 (7 2018).
17. A. Simonelli, G. De Luca, U. Giacomelli, G. Terreni and A. Di Virgilio, Observation by Means of An Underground Ring Laser Gyroscope of Love Waves Generated in the Mediterranean Sea: Source Direction and Comparison with Models, *Seismological Research Letters* (3 2020).
18. U. Giacomelli, Geometry control and stabilization of a large ring laser gyroscope, PhD thesis, Università di Pisa.

Dark gravitomagnetism with LISA and gravitational waves space detectors

A. Tartaglia

OATo, INAF and DISAT, Politecnico
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
E-mail: angelo.tartaglia@inaf.it
www.polito.it

M. Bassan and G. Pucacco

Dip. Fisica, Tor Vergata University and INFN/Roma2, Via della Ricerca Scientifica 1
00133 Rome, Italy
E-mail: bassan@roma2.infn.it

V. Ferroni and D. Vetrugno

Dip. Fisica, Trento University and TIFPA, Via Sommarive 14
38123 Povo TN, Italy
E-mail: valerio.ferroni@unitn.it

We present here the proposal to use the LISA interferometer for detecting the gravitomagnetic field due to the rotation of the Milky Way, including the contribution given by the dark matter halo. The galactic signal would be superposed to the gravitomagnetic field of the Sun. The technique to be used is based on the asymmetric propagation of light along the closed contour of the space interferometer (Sagnac-like approach). Both principle and practical aspects of the proposed experiment are discussed. The strategy for disentangling the sought for signal from the kinematic terms due to proper rotation and orbital motion is based on the time modulation of the time of flight asymmetry. Such modulation will be originated by the annual oscillation of the plane of the interferometer with respect to the galactic plane. Also the effect of the gravitomagnetic field on the polarization of the electromagnetic signals is presented as an in principle detectable phenomenon.

Keywords: Gravito-magnetism; Galactic halo; Sagnac effect; Space interferometers.

1. Introduction

Among the possible space-times consistent with the Einstein equations and with various matter distributions, a special interest deserve those endowed with a chiral symmetry about the time axis of a given observer. The chiral symmetry may sometime appear as a simple local coordinates artifact, in which case the symmetry disappears with an appropriate choice of the reference frame. Clearly, the most interesting situations are when the symmetry is indeed in the curvature of space-time, which happens when the source of gravity is a spinning mass distribution. In terms of the metric tensor, the presence of the symmetry we are interested in is manifested by non-null time-space off-diagonal terms. It is of course always possible to choose

a local reference frame where all off-diagonal terms are brought to zero, but this is in general not possible *globally*, when the source of gravity is spinning. In weak field conditions (the most common situation in the universe) the symmetry we are considering appears in the form of a *gravito-magnetic* field. The name derives from the fact that in weak field approximation the Einstein equations assume a form quite similar to Maxwell's equations of electromagnetism (e.m.), so that the interaction between moving masses (just as moving charges in e.m.) may be described in terms of forces due to two vector fields: the *gravito-electric* (analogous to the electric field, but always attractive), and the *gravito-magnetic* field (analogous to the magnetic field of classical e.m.). This formalism is well known and we refer to the vast literature and to basic texts of general relativity (GR): see for example Ref. [1].

It is, however, worth remarking that the gravito-magnetic force is usually much weaker than the gravito-electric one, making thus hard to identify measurable effects. So far, physically relevant phenomena associated with the aforementioned symmetry are expected, upstream of any approximation, in strong field conditions such as near Kerr black holes or, when the gravito-electromagnetic approximation is applied, in the behaviour of peculiar systems where gravity is still strong enough to allow the gravito-magnetic component to emerge in an observable way. The latter is the case of the double pulsar, whose internal dynamics, readable from the signals received by our radiotelescopes, indirectly hints to the gravito-magnetic (GM) interactions between the spins and the orbital motions of the two stars in the pair.[2]

In the Solar System, the effects due to the angular momenta of the Sun and of the planets are extremely weak. Until now the only measured effects are those of the terrestrial gravito-magnetic field. They have been observed monitoring the motion of the moon along its orbit by Laser ranging[3] and doing the same with the LAGEOS and LAGEOS2 satellites:[4] what was found was the Lense-Thirring drag[5] on their motion. Direct evidence has been obtained by the dedicated Gravity Probe B experiment which measured the induced precession on gyroscopes carried on a circumterrestrial satellite in polar orbit.[6] Laser ranging is also the main tool used to analyze the orbit of the dedicated LARES mission, which is producing the best accuracy so far for the GM field of the Earth.[7] On the surface of our planet it is expected to measure the GM field using ring lasers.[8–10]

A way to compensate for the weakness of the sought effects is to use very large sensors and measuring devices, that could not be hosted on a single spacecraft or in a laboratory on Earth. This is one of the reasons pushing toward peculiar configurations of a plurality of measuring devices in space, either around the Earth[11] or at the scale of the inner solar system.[12]

An extremely interesting opportunity is the LISA interferometer,[13] designed and under development for detecting gravitational waves. What we propose here is to use LISA, which, as we shall see, is already fit for the purpose, also to reveal the GM field in which the device is immersed. The interesting sources that could be studied

are both the solar angular momentum and the angular momentum of the Milky Way (or, to say better, the part of it which is relevant for the solar system). A further reason of interest for the galactic angular momentum is that it reasonably depends both on the visible and on the dark component of our galaxy. So far, dark matter (DM), which is an important ingredient of the cosmic cocktail, is postulated only on the basis of its direct gravitational (gravito-electric) effect. However, if DM exists and we trust GR, we may expect also phenomena related to the proper rotation of DM distributions. It is currently accepted that most galaxies (including ours) are immersed in huge DM halos; if so, it is also reasonable to expect such halos to rotate together with the visible part of the galaxy: in fact, if dark and visible interact gravitationally, any little localized inhomogeneity on one side or the other acts as a hook that drags or brakes the other component so that, after some remote transitory, both end up revolving together. The consequence is that DM should represent a relevant contribution to the angular momentum of the whole, then to the GM component of the gravitational interaction.

In what follows we outline how LISA could be used to measure both the angular momentum of the Sun and of the Milky Way. Reliable estimates of the angular momentum of the visible Milky Way exist: by difference, we could thus measure the galactic Dark content.

2. Internal space-time of the Milky Way

Globally and in the first instance we can attribute an axial symmetry to the distribution of matter of the Milky Way and imagine also that it is in uniform rotation condition (stationary space-time). If so the typical line element may be written as follows:

$$ds^2 = U(r,z)c^2dt^2 - 2N(r,z)rd\phi cdt - W(r,z)r^2d\phi^2 - Q(r,z)(dr^2 + dz^2) \qquad (1)$$

Cylindrical space coordinates have been used. An additional symmetry we assume is reflection symmetry about the galactic plane. U, N, W, Q are dimensionless functions of r and z only. They are everywhere regular (except possibly in the origin). The only constraint expressing the reflection symmetry is

$$\frac{\partial U}{\partial z}\Big|_{z=0} = \frac{\partial N}{\partial z}\Big|_{z=0} = \frac{\partial W}{\partial z}\Big|_{z=0} = \frac{\partial Q}{\partial z}\Big|_{z=0} = 0 \qquad (2)$$

Our interest is on the view point of an observer at rest with the Sun, so for the moment it is convenient to assume that the r axis is corotating with the Milky Way. This assumption does not spoil the symmetry we have assumed, but implies that the U, N, W, Q functions implicitly include also the effects of the choice of the observer (which in general treatments is assumed to be at rest with the center of the mass distribution and located at infinity, whereas here it is co-moving with the Sun).

We cannot a priori say more about the shape of the functions because we are considering an observer inside the mass distribution (visible or dark it may be) and

we do not know exact solutions of Einstein's equations within a rotating distribution of matter. Nonetheless, if we succeed in detecting specific effects of the rotation of the Milky Way, we may also deduce relevant information on the way DM is distributed. This is because the non trivial parts of the elements of the metric tensor contain the mass density of the Milky Way and the rotation curve of the stars in it: since we know how the visible mass is distributed and the rotation curve of the galaxy (which we have assumed to be the same for DM also), from the evaluation of the total GM field we may deduce relevant information on the DM density distribution, which contributes to the effect.

2.1. *Galactic gravito-magnetism*

Line element (1) is quite general, but we may equally well adopt the language and the images of the weak field, since that is indeed the case. We may also focus our attention on the situation on and near the galactic plane considering that the Sun is not far from it. This said, we read the ratio N/U as the only non-zero component of a three-vector, which is the analog of the vector potential $\bar{A}$ of classical electromagnetism.[14]

Under the above conditions, whenever the metric tensor is independent of time the equation of geodesic motion can be written in a form analogous to Lorentz's equation of e.m.

$$\frac{d^2\bar{x}}{dt^2} = -c^2\left(\bar{\nabla}U - 2\frac{\bar{v}}{c}\times\bar{\nabla}\times\bar{A}_g\right) \tag{3}$$

$\bar{v}$ and $\bar{x}$ are the three-dimensional velocity and position of the test mass with respect to the observer. The functions U and $\bar{A}_g$ are given in terms of the elements of the metric tensor $g_{\mu\nu}$ by:

$$U = g_{00};\quad A_{(g)i} = \frac{g_{0i}}{g_{00}} \tag{4}$$

In our case it is

$$\bar{A}_g = \left(0, \frac{N}{U}, 0\right)\quad \text{or}\quad \bar{A}_g = \frac{N}{U}\hat{\phi} \tag{5}$$

Here $\hat{\phi}$ is the unit transverse vector in the direction of rotations about the symmetry axis. Our space basis for covariant vectors is $(dr, rd\phi, dz)$ so that all components of $\bar{A}_g$ are dimensionless and the field lines of the vector are circles centered on the space symmetry axis, contained in planes perpendicular to that axis. The next step, formally recognizing the weak field approximation, is to calculate a *gravito-magnetic field* $\bar{B}_g$ from $\bar{A}_g$ just in the same way we would do in classical electromagnetism:

$$\bar{B}_g = \bar{\nabla}\times\bar{A}_g = \frac{N}{U}\left(\frac{\partial_z U}{U} - \frac{\partial_z N}{N}\right)\hat{r} + \frac{N}{U}\left(\frac{\partial_r N}{N} - \frac{\partial_r U}{U}\right)\hat{z} \tag{6}$$

∂_r and ∂_z are a shorthand notation for $\partial/\partial r$ and $\partial/\partial z$; each component of $\bar{B}_g$ has the dimension of the inverse of a length.

If we limit ourselves to the symmetry plane (the galactic plane) the r component of the field vanishes and the vector turns out to be perpendicular to the plane, then parallel to the axis of the Milky Way.

3. The propagation of light

In the peculiar space-time we are considering, the propagation of light displays an interesting behaviour. Whenever some physical device (mirrors, waveguides or else) constrains light to move along a closed path, the time of flight (ToF) it takes to make a full turn is different depending on whether the travel is right- or respectively left-handed. The sense of the circulation is with respect to the field lines of $\bar{B}_g$.

Actually this asymmetry is not peculiar of light only. It holds true for any signal or voyager, provided its velocity is locally the same in both directions at any position along the path.[15] The effect we are recalling is sometimes improperly called 'Sagnac effect', though the latter is an effect of special instead of general relativity and concerns the non inertial rotational motion of the observer rather than the curvature of space-time.

When considering the situation from a four-dimensional viewpoint, light leaves the source at a given event of the worldline of the observer, then comes back to the observer at a different point along its worldline: the path is of course open. When starting in the opposite direction, the light ray will meet again the observer at a different event of its worldline. Locally measuring the proper time (τ) difference between the arrivals (the length of the intercepted interval of the observer's worldline) we deduce an information concerning the proper angular momentum of the source of gravity (possibly combined with rotational motion of the observer, if it is present).

The difference in proper time can be expressed in terms of the components of the metric tensor. Using the standard notation of GR we have:[16]

$$\Delta\tau = -\frac{2}{c}\sqrt{g_{00}} \oint \frac{g_{0i}}{g_{00}} dx^i \qquad i = 1, 2, 3 \tag{7}$$

The square root of g_{00} is evaluated at the position of the observer and accounts for the gravitational field there; it does not coincide with the g_{00} under the integral sign, since the latter expresses the gravitational field along the integration path. Here x^i is a generic space coordinate and the integration is performed along the space closed contour traveled by the light beam.

Restricting the description to the space-time with the symmetries described above, and using the typical notation of gravito-electromagnetism, the time delay becomes:

$$|\Delta\tau| = \frac{2}{c}\sqrt{U} \oint A_{gi} dx^i = \frac{2}{c}\sqrt{U} \int \bar{B}_g \cdot \hat{u}_n dS \tag{8}$$

The last term on the right is obtained applying the Stokes theorem of usual three-dimensional geometry and converts the line integral into a flux; dS is the surface element of the area contoured by the closed path in space and $\hat{u}_n$ is the unit three-vector perpendicular to the given surface element.

3.1. *Galactic gravito-magnetism in the Solar System*

Our purpose is to consider an experiment carried out at the scale of the inner solar system and indeed at the Earth's orbit. The galactic GM field $\bar{B}_g$ does in fact depend on the distance from the center of the Milky Way. The Sun is approximately $R_S \cong 2.35 \times 10^{20}$m (28,000 light years) away from the center of the Milky Way. If we think of an instrument revolving around the Sun in correspondence of the orbit of our planet, its distance from the center of the Milky Way changes at most by plus or minus 1 astronomical unit (AU), i.e. $\Delta R_S \sim 3 \times 10^{11}$ m. The relative fluctuation of the distance, under these conditions, would then be $\Delta R_S/R_S \sim 10^{-9}$. Even though we have no explicit expression for the functions appearing in line element (1), just looking at what happens in electromagnetic analogies, we may reasonably expect $\Delta B_g/B_g$ to be of the same order of magnitude as $\Delta R_S/R_S$. If so, if we additionally assume such a relative change to be negligible for our purposes, it follows that in practice $\bar{B}_g$ of the Milky Way is a constant for our experiment. Under this assumption eq. (8) simplifies to:

$$|\Delta\tau| \cong \frac{2}{c}\sqrt{U}B_g S \cos\gamma \tag{9}$$

S is now the total area enclosed in the path of light, assuming that the contour is contained in a plane; γ is the angle between the perpendicular to that plane and the direction of the gravito-magnetic field of the Milky Way at the Sun. As far as the Sun may be considered to lay in the galactic plane, the direction of $\bar{B}_g$ coincides with that of the axis of the Milky Way.

If we are able to measure $|\Delta\tau|$, we may then deduce from eq. (9) information on U and in particular on N, since U, being related to the local ordinary gravitational potential, can also be obtained by other means.

4. Space interferometers and LISA

It follows from previous considerations and from eq. (9) that a good strategy to detect the presence of a gravito-magnetic field is to resort to the measurement of time asymmetries in the propagation of light; eqs. (8) and (9) show that we need a vast area S in order to enhance the sensitivity of the experiment. This is why appropriate configurations of receivers and transponders in space are particularly interesting. For instance, it has been proposed[11] to use constellations of satellites around the Earth (such as the Galileos), but even more appropriate and promising may be the opportunity to use LISA, an orbiting interferometer designed to detect gravitational waves around the mHz frequency band.

LISA will monitor the distance between pairs of free falling test masses (TM) at a distance $L = 2.5$ million km with a continuous interferometric laser ranging scheme to detect strain distance variation due to gravitational waves. The TM pairs will be lodged in three spacecraft arranged to form a big equilateral triangular constellation. So far we considered the observer as being co-moving with the Sun, but actually the plane of the LISA constellation will form an angle $\beta = 30^o$ with the ecliptic plane and that plane is at an angle $\alpha \simeq 60^o$ with respect to the galactic plane; the whole constellation will rotate around the Sun at a distance of 1 AU, lagging by 20^o the orbit of the Earth; it will also rotate on its plane, around the center of mass, with the same 1 year period. The geometry of such configuration is shown in Fig. 1.

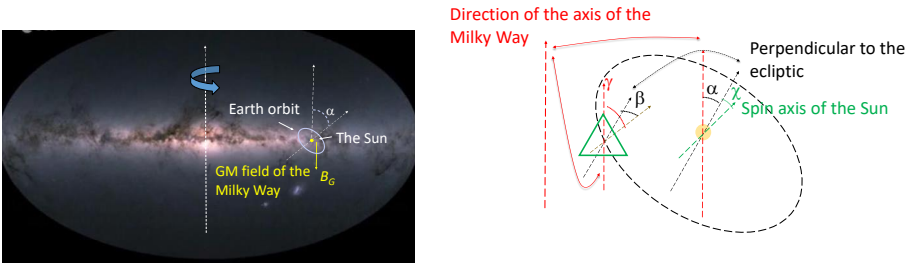


Fig. 1. The LISA interferometer in the Milky Way. a) (left) Freely falling constellation in the terrestrial orbit. b) (right) Orientation of the triangular constellation with respect to the galaxy and to the ecliptic.

Keplerian dynamics causes, in the yearly rotation period, changes of up to 10^4 km in armlength and 10 m/s in spacecraft velocity: this phenomenon, called *flexing* of the arms, forbids equal path interferometry. However, the use of post-processing techniques (Time Delay Interferometry: TDI)[17,18] allows to synthesise equal arm interferometry to better than 1 ns. Among TDI combinations there are several that mimic the output of a Sagnac interferometer,[19,20] thus suppressing the GW signals. By interfering electromagnetic signals traveling in opposite directions along the LISA triangle, we could measure the ToF difference at one of the corners: Eq. (8) will then allow us to deduce the intensity of the gravito-magnetic flux through the interferometer.

In the reference frame comoving with the spacecraft, $\bar{B}_g$ will contain various contributions including those due to the proper and orbital motion of the device (Sagnac effect). The components of interest would be the galactic contribution (both from visible and dark matter) and the gravito-magnetic field originating from the angular momentum of the Sun. In principle also the Earth and other planets would contribute, but it is easily verified that those components would indeed be much weaker than the galactic and solar ones.

4.1. *Gravito-magnetism from the Sun*

The GM field of the Sun, $\bar{B}_{g\odot}$, has a simple form, as long as our star is treated as a compact spherical spinning source. The configuration of the field is dipolar and in the solar equatorial plane, where $\bar{B}_{g\odot}$ is perpendicular to the plane, it is (net of kinematic contributions, i.e. for a non-rotating observer in a Sun-centered reference frame)[8]

$$\bar{B}_{g\odot} = \frac{2G}{c^3 r^3}\bar{J}_{\odot} = 2.8 \cdot 10^{-28}\mathrm{m}^{-1} \tag{10}$$

G is Newton's constant and $\bar{J}_{\odot}$ is the angular momentum of the Sun; r is the observer's distance from the center of the star.

In our approximation, we can take $U = 1 - 2\frac{GM_{\odot}}{c^2 r} \simeq 1$, and the ToF asymmetry, for the solar effect alone, would then be:

$$|\Delta\tau|_{\odot} \simeq \frac{2}{c}\int |\bar{B}_{g\odot}| \cos\eta dS \simeq \frac{4G}{c^4}\int \frac{|\bar{J}_{\odot}|}{r^3}\cos\eta dS \tag{11}$$

The flux is calculated over the LISA triangle: actually, the area of the constellation is tilted with respect to the ecliptic, and this has two consequences: 1) part of the area is slightly above and part slightly below the ecliptic plane so that there $\bar{B}_{g\odot}$ has also a small radial component; and 2) the three spacecrafts have different instantaneous distances from the Sun ($\delta r/r \lesssim 0.6\%$) and therefore experience a different strength of the dipolar field. In the following, we shall neglect, for sake of simplicity, both corrections. η is the angle between the normal to the plane of the triangle and the axis of the Sun. Since the latter is in turn inclined with respect to the north of the ecliptic[21] by the angle $\chi \simeq 7^{\circ}$, η changes periodically during the year oscillating between $\eta_{min} = \beta - \chi$ and $\eta_{max} = \beta + \chi$; β and χ are shown in Fig. 1b.

5. Proposed measurements

The experiment we would like to consider uses LISA as the triangular closed path along which e.m. signals are sent in opposite directions, in order to measure the ToF difference. Letting, for the moment, all practical problems aside, we note a multiplicity of contributions simultaneously acting upon the interferometer. GR is a non-linear theory so that, in principle, it is not an easy task to disentangle the various terms. In particular, beside the solar and galactic contribution, we have the kinematic terms related to the rotational movements of the configuration, i.e. the proper rotation of the triangle, its orbital motion around the Sun and the rotation of the Sun about the axis of the Milky Way. These kinematic terms may legitimately be considered as manifestations of the Sagnac effect; in a treatment that wanted to be "exact" all terms (both kinematic signals and "physical" i.e. GM terms)

combine non-linearly with each other, so that it is extremely difficult to distinguish them. Fortunately, the fact that all effects are small or very small (the GM terms are expected to be much weaker than the kinematic ones) helps us, in the sense that in an approximate treatment we are allowed to truncate the expansion at the lowest order. Indeed the lowest approximation is the linear one where a simple superposition principle is applied.

Even accepting the linear approximation, the problem remains of separating the different addends of the sum. This separation would be possible if we knew at least the kinematic terms with an accuracy better than the size of the unknown "physical" contributions. This is a difficult task, because the gravitomagnetic effect is expected to be many orders of magnitude smaller than the Sagnac effect; we need some sort of signature that highlights the terms of interest and the "marking" we may exploit is the time dependence of the signals. Looking at Fig. 1b we see that our triangle oscillates yearly with respect to the galactic plane between an angle $\gamma_{max} \simeq \alpha + \beta$ and an angle $\gamma_{min} \simeq \alpha - \beta$, where $\alpha \simeq 60^o$ is the angle between the ecliptic and galactic planes. The oscillation is not simply sinusoidal and we use $\simeq$ instead of just $=$ because the axes of the Milky Way, of the ecliptic and of the LISA triangle never turn out to be exactly co-planar. Furthermore, when looking at the solar contribution, we find the already mentioned oscillation with respect to the solar spin axis.

Thus, both GM contributions exhibit yearly periodicity but with a different phase, whereas the Sagnac effect is nearly constant with small harmonic components (including the yearly one) due to flexing, which can be, in principle, calculated and subtracted.

As for the detectable signal, we may estimate its minimal amplitude for the Sun GM from eq. (11), with a LISA constellation area, neglecting the above mentioned oscillations, $S = (\sqrt{3}/4)L^2 \simeq 2.7 \times 10^{18}\mathrm{m}^2$. The target displacement sensitivity for LISA, is $c|\Delta\tau_{min}| \sim 10^{-11}$ m. With this sensitivity both the solar GM and the galactic one should be detectable by LISA (see Fig. 2), although with marginal signal to noise ratio.

Eq. 9 shows that the minimum detectable GM field would then be

$$B_{(min)} \sim 1.8 \times 10^{-30}\mathrm{m}^{-1} \tag{12}$$

In order to see what this number means, consider that the terrestrial GM field on the surface of the Earth (at the equator) is $\sim 10^{-22}\mathrm{m}^{-1}$. Proper Sagnac terms (which depend on the angular velocities around the Sun, around the constellation center and around the galactic center) are many orders of magnitude larger than the value in eq.(12), whereas the solar contribution is only a factor 100 above the minimum detectable signal. On the basis of estimates of the angular momentum of the visible mass, the galactic GM field is expected to be just above the edge of detectability.

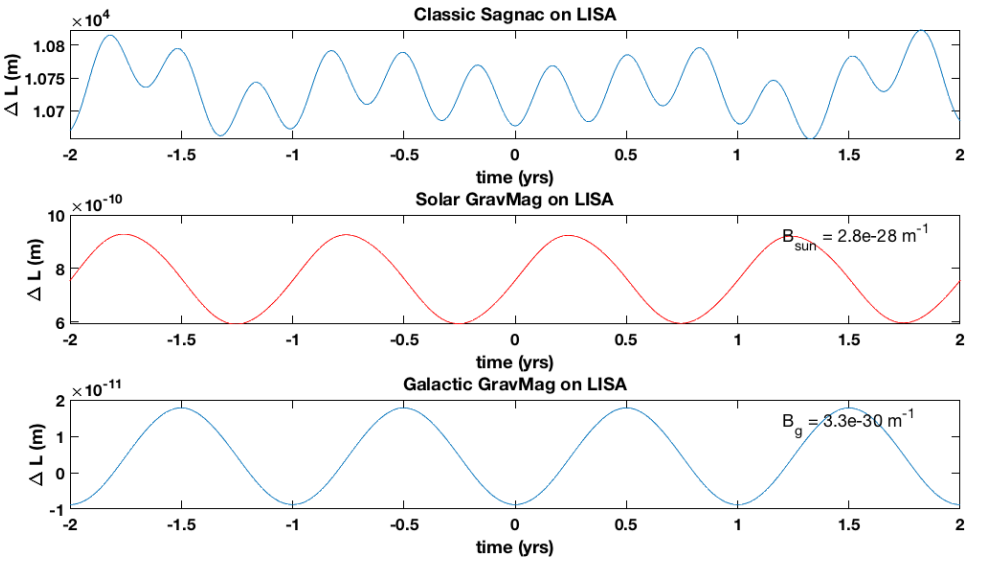


Fig. 2. Time dependence of the expected signals (expressed as lengths) on a noiseless LISA constellation:
TOP: Sagnac effect, due to the triangle rotations both around the Sun and around its center. The signal has an almost constant value, with small ($\sim 1\%$) oscillations due to the flexing of the arms and ensuing change of the area.
MIDDLE: GM effect from the Sun; the yearly modulation is due to the inclination of $J_\odot$ with respect to the ecliptic.
BOTTOM: signal from the galactic Gravito Magnetism; this measurement would be performed *within* the source mass current, and no certainties can be offered about its strength. We hypothized, for this plot, a field $B_{MW} = 3 \cdot 10^{-30} m^{-1}$, neglecting the contribution of the external part of the Milky Way (that would actually act to reduce this strength).

6. GM effect on light polarization

The only GM effect we have considered so far is the asymmetry in the ToF of light, but there are other consequences to be considered. The analogy with classical e.m. suggests that we may also expect the analog of the Faraday effect. In other words, if a linearly polarized light beam propagates along the field lines of a GM field the direction of the polarization vector should rotate around the direction of the propagation.[23]

Using the formalism of GR, the electromagnetic field is described (in four dimensions) by the so called Faraday anti-symmetric tensor $F_{\mu\nu}$ and, regardless of any approximation, when there is propagation producing a four-dimensional area $\delta S^{0i} = (\delta\tau \delta x^i) = (\delta x^i)^2/c$ (movement along the direction of x^i at the speed of light) the corresponding change in $F_{\mu\nu}$ is

$$\delta F^{\mu\nu} = (R^\mu_{\epsilon 0 i} F^{\epsilon\nu} + R^\nu_{\epsilon 0 i} F^{\mu\epsilon}) \delta S^{0i} \tag{13}$$

$R^\mu_{\nu\lambda\rho}$ are the components of the Riemann tensor which account for the curvature of space-time.

Whenever the propagation takes place over a distance l in a direction non-perpendicular to the GM field, the change in the Faraday tensor corresponds to a rotation of the polarization vector of the wave around the propagation line by an angle ψ:[23]

$$\psi \simeq -\frac{\sqrt{g_{00}}}{2}\bar{B}_g \cdot \bar{l} = \pm\frac{l}{2}\frac{N}{\sqrt{U}}\left(\frac{\partial_r N}{N} - \frac{\partial_r U}{U}\right) \tag{14}$$

The last expression in Eq. (14) refers to the case of the Milky Way and uses the notation introduced in Eq. (6) applied on the galactic plane and for light travelling in a direction perpendicular to that plane.

The solar GM will produce, in any LISA arm, a rotation of the polarization plane of $\sim 3 \cdot 10^{-19}$ prad. The two laser beams propagating in each arm in opposite directions would undergo opposite rotations. In principle the effect would cumulate with the laser beam going around the constellation if the transmission of the light from satellite to satellite preserves the polarization. LISA will use linearly polarized light for the strain measurements. The unwanted polarisation component present in the received beam will be removed, when entering the measurement chain, by a PBS (polarising beam splitter); this fraction, containing the GM signal, could be in principle detected by an additional, dedicated photo diode. In fact, in the LISA-Pathfinder (LPF) mission it has been observed that, because of non perfect performance of the PBS, part of this light can still contaminate the strain measurement, yielding additional noise in the form of read-out noise and spurious laser pressure. Thus the GM rotation effect in LISA would possibly result in a decreased sensitivity of the observatory. That signal could be reconstructed in post-processing, as it was done in LPF, using a well shaped model of the measurement chain.[22]

7. Conclusion

We have discussed the relevant sources that could contribute to the gravito-magnetic field in the solar system far from the planets and within 1 AU from the Sun. In particular we have considered the relevance of the field in producing an asymmetry in the propagation of electromagnetic signals along closed space paths. We have then discussed the possibility to use the big triangle formed by the LISA satellite constellation applying an approach similar to that of the Sagnac effect.

A quantitative evaluation of the effects shows that the sensitivity of such an arrangement would be sufficient to measure the solar GM field. Such measurement would give an important information on the internal structure of the Sun, from which the angular momentum of our star depends. But LISA, used *à la Sagnac*, would probably be able to reveal the weak GM field of the Milky Way as well. This is increasingly interesting and important since the galactic gravitomagnetism depends of course on the distribution of visible matter in the Milky Way, but also,in a non marginal way, on the presence of dark matter, that, on the whole, is expected to have a mass much greater than that visible.

The GM field also causes a rotation of the polarization of the e.m. signals that will travel along the arms of the interferometer. The effect is tiny, and its detection is extremely challenging, but should in principle be considered, in view of future technological advances. This would yield additional and complementary information on the intensity of the field.

LISA will send its "raw" data to Earth, to be off-line processed and syntethized in interferometric signals: in view of this strategy, pursuing the search for GM signatures on the data, a challenging, though not impossible task, should require no further hardware development, and no additional task for the mission.

For these reasons and for the importance of the objectives pursued it would be worthwhile to carry out the experiment.

References

1. Misner C. W., Thorne K. S. and Wheeler J. A., *Gravitation* (Freeman, New York, 1973).
2. Breton R. P. et al., *Relativistic Spin Precession in the Double Pulsar* - Science **321**, 104 (2008).
3. Nordtvedt K., *Lunar Laser Ranging - a comprehensive probe of post-Newtonian gravity* in Proceedings of Villa Mondragone International School of Gravitation and Cosmology, 2002; arXiv, gr-qc/0301024 (2003).
4. Ciufolini I. and Pavlis E. C., *A confirmation of the general relativistic prediction of the Lense-Thirring effect* - Nature, **431** (7011) 958 (2004).
5. Lense J. and Thirring H., Phys. Z. **19**, 156 (1918); the English translation is available in B. Mashhoon, F.W. Hehl, and D. S. Theiss, Gen. Relativ. Gravit. **16**, 711 (1984).
6. Everitt, C. W. F., et al., *Gravity probe B: Final results of a space experiment to test general relativity*, Phys. Rev. Lett. **10**, 221101 (2011).
7. Lucchesi, D. M. et al., *A 1% Measurement of the Gravitomagnetic Field of the Earth with Laser-Tracked Satellites Universe* **6**, 139 (2020).
8. Bosi F. et al., *Measuring gravitomagnetic effects by a multi-ring-laser gyroscope*, Phys. Rev. D **84**, 122002-1–122002-23 (2011).
9. Di Virgilio, A. et al., *A ring lasers array for fundamental physics*, C. R. Phys. **15**, 866–874 (2014).
10. Tartaglia, A., et al., *Testing general relativity by means of ring lasers*, Eur. Phys. J. Plus **132**, 73 (2017).
11. Ruggiero M. L. and Tartaglia A., *Test of gravitomagnetism with satellites around the Earth*, European Physical Journal Plus, **134**, 205 (2019).
12. Tartaglia A. et al., *Detecting the gravito-magnetic field of the dark halo of the Milky Way - the LaDaHaD mission concept*, Experimental Astronomy, https://doi.org/10.1007/s10686-021-09700-4 (2021).
13. Wanner G. et al., *Space-based gravitational wave detection and how LISA Pathfinder successfully paved the way*, Nature Physics 15 200-202, 2019.
14. Tartaglia A., *Rotation Effects and The Gravito-Magnetic Approach*, in *General Relativity and Gravitational Physics: 16th Sigrav Conference on General Relativity and Gravitational Physics*; G. Esposito, G. Lambiase, G. Marmo, G. Scarpetta and G. Vilasi Eds., (AIP Conference Proceedings **751**, 136-145, 2005); https://doi.org/10.1063/1.1891538.

15. Ruggiero M. L. and Tartaglia A., *A note on the Sagnac effect for matter beams*, Eur. Phys. J. Plus **130**, 90 (2015).
16. Kajari E., Buser M., Feiler C. and Schleich W. P., *Rotation in relativity and the propagation of light*, Riv. Nuovo Cimento Soc. Ital. Fis. **32**, 339 (2009).
17. Tinto M. and J. W. Armstrong, *Cancellation of laser noise in an unequal-arm interferometer detector of gravitational radiation* Phys. Rev. D 59, 102003 (1999).
18. Cornish N. Neil J. Cornish and Hellings R. W. *The effects of orbital motion on LISA time delay interferometry*, Class. Quantum Grav. 20 4851 (2003).
19. Shaddock D., *Operating LISA as a Sagnac interferometer*, Phys. Rev. D 69, 022001 (2004).
20. Muratore M., Vetrugno D. and Vitale S., *Revisitation of time delay interferometry combinations that suppress laser noise in LISA*, Class. Quantum Grav. 37, 185019 (2020).
21. Pijpers, F. P., *Helioseismic determination of the solar gravitational quadrupole moment* Monthly Notices of the Royal Astronomical Society **297**, L76-L80 (1998).
22. Kaune B., *In-orbit Stability Analysis of the LISA Pathfinder Optical Metrology: Photoreceivers and Polarisation*, Doctoral Thesis, Der Fakultät für Mathematik und Physik der Gottfried Wilhelm Leibniz Universität Hannover, 2021.
23. Ruggiero M. L. and Tartaglia A., *Gravitational Faraday rotation in binary pulsar systems*, Monthly Notices of the Royal Astronomical Society, **374**, 847-851 (2007).

Light rays in the Solar system experiments: Phases and displacements

Pravin Kumar Dahal* and Daniel R. Terno

Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia

**E-mail: pravin-kumar.dahal@hdr.mq.edu.au*

Geometric optics approximation is sufficient to describe the effects in the near-Earth environment. In this framework Faraday rotation is purely a reference frame (gauge) effect. However, it cannot be simply dismissed. Establishing local reference frame with respect to some distant stars leads to the Faraday phase error between the ground station and the spacecraft of the order of 10^{-10} in the leading post-Newtonian expansion of the Earth's gravitational field. While the Wigner phase of special relativity is of the order 10^{-4}–10^{-5}. Both types of errors can be simultaneously mitigated by simple encoding procedures. We also present briefly the covariant formulation of geometric optic correction up to the subleading order approximation, which is necessary for the propagation of electromagnetic/gravitational waves of large but finite frequencies. We use this formalism to obtain a closed form of the polarization dependent correction of the light ray trajectory in the leading order in a weak spherically symmetric gravitational field.

Keywords: Geometric optics; Wigner phase; Gravitational Faraday rotation; Post-Newtonian expansion; Gravitational spin Hall effect.

1. Introduction

Space deployment of quantum technology[1–3] brings it into a weakly relativistic regime. As an unintended but fortunate side effect, low-Earth orbit (LEO) quantum communication satellites provide new opportunities to test fundamental physics. Once the tiny putative physical effects fall within the sensitivity range of these devices, they may impose constraints on practical quantum communications, timekeeping, or remote sensing tasks.[4–6] A more futuristic technology, such as the proposed solar gravity telescope[7–9] actually needs the general-relativistic effects for its operation.

For flying qubits that are implemented as polarization states of photons[10] the dominant source of relativistic errors in this setting is the Wigner rotation (or phase), an effect special relativity (SR).[4,6] Gravitational polarization rotation, also known as the gravitational Faraday effect[11,12] occurs in a variety of astrophysical systems, such as accretion disks around astrophysical black hole candidates[13] or gravitational lensing.[14] This effect was the subject of a large number of theoretical investigations, primarily within the geometric optics approximation,[11,12,14–17] and also from the perspective of quantum communications.[18] Interpretation of these results was until recently sometimes contradictory, as it is important to carefully analyze the relation between the reference frames of the emitter and the detector. Moreover, even if at the leading order the gravitationally induced polarization

rotation in the near-Earth environment is pure gauge effect, it cannot be simply dismissed. We provide a simple estimation of this emitter- and observer-dependent phase and give its explicit form in several settings.

Already in the leading post-eikonal order trajectories of light beams are affected by polarization.[19–21] The optical gravitational spin Hall effect has recently received a comprehensive treatment in Refs. 22, 23. Using this formalism we obtain a closed form of the leading correction in case of a weak spherically symmetric gravitational field. As expected, on the scale of the Solar system — be it the near-Earth environment or a focal plane of the solar gravity telescope at 600a.u.— the effect is negligible. However, it is interesting conceptually and its scaling indicates that it may play a much more important role in the strong gravity regions.[23]

The rest of this article is organized as follows. In Sec. 2 we review the SR effects. Polarization rotation in general stationary spacetimes is described in Sec. 3, where we also evaluate the effects in communication with Earth-orbiting satellites. In Sec. 4 we briefly summarize the main techniques of Refs. 22, 23 and then obtain the polarization-dependent changes in the light ray trajectories in the solar system experiments.

We work with $c = \hbar = G = 1$. The constants G and c are restored in a small number of expressions where their presence is helpful. The spacetime metric $g_{\mu\nu}$ has a signature $-+++$. The four-vectors are distinguished by the sans font, k, $k^\mu = (\mathsf{k})^\mu$. The three-dimensional spatial metric is denoted as γ_{mn}, and three-dimensional vectors are set in boldface, $\mathbf{k}$, or are referred to by their explicit coordinate form, k^m. The inner product in the metric γ is denoted as $\mathbf{k}\cdot\mathbf{f}$, and the unit vectors in this metric are distinguished by the caret, $\hat{\mathbf{k}}$, $\hat{\mathbf{k}}\cdot\hat{\mathbf{k}} \equiv 1$. Post-Newtonian calculations employ a fiducial Euclidean space. Euclidean vectors are distinguished by arrows, $\vec{k}$. Components of the two types of vectors may coincide, $(\mathbf{v})^m = (\vec{b})^m$, but $\vec{b}\cdot\vec{k} = \sum_{k=1}^{3} v^m k^m$. Accordingly, the coordinate distance is the Euclidean length of the radius vector, $r \equiv \sqrt{\vec{x}\cdot\vec{x}}$.

2. Wigner rotation and special relativistic effects

Consider one round of communications between the ground station (GS) and a low energy orbit spacecraft (SC). The problem is most conveniently analyzed in the geocentric system, the origin of which coincides with the centre of the Earth at the moment of emission. If we direct the z-axis of the system along the Earth's angular momentum, then the velocity $\mathbf{v}$ of the GS lies in the xy-plane. Whereas, the velocity $\mathbf{u}$ of the SC and the initial propagation direction $\hat{\mathbf{k}}$ are arbitrary when expressed in the global frame. This setting is depicted in Fig. 1.

Quantum states of a photon with a definite four-momentum $\mathsf{k} = (|\mathbf{k}|, \mathbf{k})$ can be represented either as Hilbert space vectors or as complex polarization vectors in the usual three-dimensional space,

$$|\Psi_{\mathsf{k}}\rangle = f_+|\mathsf{k},+\rangle + f_-|\mathsf{k},-\rangle \quad \Leftrightarrow \quad \hat{\mathbf{f}}_{\mathbf{k}} = f_+\hat{\mathbf{b}}^+_{\hat{\mathbf{k}}} + f_-\hat{\mathbf{b}}^-_{\hat{\mathbf{k}}}, \tag{1}$$

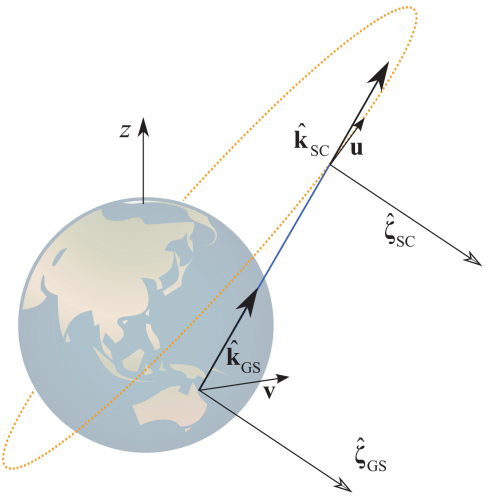


Fig. 1. Scheme of the communication round between the ground station (GS) and the spacecraft (SC). The light ray is highlighted in blue. All the vectors are indicated in the Earth-centered inertial frame. In the flat spacetime approximation the unit tangent vectors to the ray $\hat{\mathbf{k}}_{GS}$ and $\hat{\mathbf{k}}_{SC}$ at the GS and the SC, as well as the direction to an infinitely distant guide star, $\hat{\boldsymbol{\zeta}}_{GS}$ and $\hat{\boldsymbol{\zeta}}_{SC}$, respectively coincide. Velocity of the GS at the emission of the signal is $\mathbf{v}$ and velocity of the SC at the moment of detection is $\mathbf{u}$.

with the transversal vectors $\hat{\mathbf{b}}^{\pm}_{\hat{\mathbf{k}}}$, $\mathbf{k}\cdot\hat{\mathbf{b}}^{\pm}_{\hat{\mathbf{k}}}=0$. The correspondence is rooted in the relationship between finite-dimensional and unitary representations of the Poincaré group.[24,25]

Unitary operators U that describe the state transformation between the Lorentz frames are obtained via the induced representation of the Poincaré group. Basis states that correspond to an arbitrary momentum $(\mathsf{k})^{\mu}\equiv k^{\mu}$ are defined with the help of standard Lorentz transformation $L(k^{\mu})$, that takes the four momentum from the standard value k^{μ}_{S} to k^{μ}. For massless particles, $k^{\mu}_{\mathrm{S}}=(1,0,0,1)$ and

$$L(k^{\mu})=R(\hat{\mathbf{k}})B_z(\xi_{|\mathbf{k}|}), \tag{2}$$

where $R(\hat{\mathbf{k}})=R_z(\phi)R_y(\theta)$ rotates the z-axis into the direction $\hat{\mathbf{k}}$ by performing rotations around the y- and z-axis by angles θ and ϕ, respectively. These rotations follow the boost $B_z(\xi_{|\mathbf{k}|})$ along the z-axis, that brings the magnitude of momentum to $|\mathbf{k}|$.

The states of an arbitrary momentum are defined via

$$|\mathsf{k},\pm\rangle:=U\big(L(\mathsf{k})\big)|\mathsf{k}_{\mathrm{S}},\pm\rangle, \tag{3}$$

while the standard right- and left-circular polarization vectors are defined as

$$\hat{\mathbf{b}}^{\pm}_{\hat{\mathbf{k}}}:=R(\hat{\mathbf{k}})(\hat{\mathbf{x}}\pm i\hat{\mathbf{y}})/\sqrt{2}, \tag{4}$$

where the linear polarization vectors are $\hat{\mathbf{b}}_1:=R(\hat{\mathbf{k}})\hat{\mathbf{x}}$ and $\hat{\mathbf{b}}_2:=R(\hat{\mathbf{k}})\hat{\mathbf{y}}$, respectively.

Alternatively, these vectors can be obtained as

$$\hat{\mathbf{b}}_2 = \frac{\hat{\mathbf{z}} \times \hat{\mathbf{k}}}{|\hat{\mathbf{z}} \times \hat{\mathbf{k}}|}, \qquad \hat{\mathbf{b}}_1 = \hat{\mathbf{b}}_2 \times \hat{\mathbf{k}}. \tag{5}$$

The explicit form of the polarization four-vector f_k, $\mathsf{f}_\mathsf{k} \cdot \mathsf{k} = 0$ depends on the gauge.[4, 18, 26]

Under arbitrary Lorentz transformation, states transform (apart from the normalization factor) via

$$U(\Lambda)|\mathsf{k}, \pm\rangle = U(\Lambda L(\mathsf{k}))|\mathsf{k}_\mathrm{S}, \pm\rangle = U(L(\Lambda\mathsf{k}))U(W)|\mathsf{k}_s, \pm\rangle, \tag{6}$$

where the transformation $W = L^{-1}(\Lambda\mathsf{k})\Lambda L(\mathsf{k})$ to the subgroup (Wigner's little group) group that leaves k_s invariant,

$$U(W)|\mathsf{k}_\mathrm{S}, \sigma\rangle = \sum_{\pm'} D(W)_{\sigma\sigma'}|\mathsf{k}_\mathrm{S}, \sigma'\rangle. \tag{7}$$

The matrices $D(W)_{\sigma\sigma'}$ form the representation of the little group. For massless particles an element of the little group W can be decomposed as

$$W = SR_z(\varpi), \tag{8}$$

where S is a translation in the (xy)-plane and $R_z(\varpi)$ a rotation. As translations do not contribute to the physical degrees of freedom of photons, the state $|\mathsf{k}, \pm\rangle$ transforms as

$$U(\Lambda)|\mathsf{k}, \pm\rangle = e^{\pm i\varpi}|\Lambda\mathsf{k}, \pm\rangle. \tag{9}$$

There are no generic explicit expressions for ϖ. Their evaluation is not considerably simpler if $\Lambda = \mathcal{R}$, where $\mathcal{R}$ is a rotation (as there is no risk of confusion we use the same designation for the four-dimensional matrices of spatial dimensions and for their 3×3 blocks). However, as the transformation law of $\hat{\mathbf{f}}$ can be obtained from the three-dimensional form of the Lorentz transformations of the transversal electromagnetic wave, in this case[4, 27]

$$U(\mathcal{R})|\Psi_\mathsf{k}\rangle \Leftrightarrow \mathcal{R}\hat{\mathbf{f}}_\mathbf{k}. \tag{10}$$

Moreover, an arbitrary rotation around the direction $\hat{\mathbf{b}}_2$, $R_{\hat{\mathbf{b}}_2}(\alpha)$, does not introduce a phase ϖ.[16] This provides the motivation for introduction of the so-called Newton gauge that we review below.

In communications with the Earth-orbiting satellites settings of Fig. 1 the Wigner phase is the dominant relativistic effect.[28] While the Wigner phase is of the order 10^{-4}–10^{-5}, we will see below that establishing the local reference frame with respect to some distant stars leads to the Faraday phase error of the order of 10^{-10}.

3. Faraday rotation and general relativistic effects

Here we describe the effects of gravity on polarization in the geometric optics approximation, wave it can be considered a vector that is simply affected by the geodesic motion of null particles to which it is attached. We present the gravitational Faraday effect in a way that clearly separates the gauge-independent part from the effects of the reference frame choices[16] and is convenient for the near-Earth calculations that use the post-Newtonian approximation. Then we demonstrate that at the leading order the polarization rotation is a purely gauge effect and evaluate it for a practically useful choice of the reference frames.

The equation of geometric optics are obtained by performing the short wave expansion of the wave equation in the Lorentz gauge,[29,30]

$$A^{\mu} = a^{\mu} e^{i\psi}, \tag{11}$$

where a^{μ} is the slowly varying complex amplitude and ψ is the rapidly varying real phase. In later calculations, we use the wave vector $k_{\mu} := \psi_{,\mu}$, the squared amplitude $a = (a^{\mu *} a_{\mu})^{1/2}$ and the polarization vector $f^{\mu} = a^{\mu}/a$ is transverse to the trajectory, $f^{\mu} k_{\mu} = 0$.

The eikonal equation

$$g^{\mu\nu} \frac{\partial \psi}{\partial x^{\mu}} \frac{\partial \psi}{\partial x^{\nu}} = 0, \tag{12}$$

is the leading term in the expansion of the wave equation[29,30] is the Hamilton-Jacobi equation for a free massless particle on a given background spacetime. It allows description of light propagation in terms of fictitious massless particles.

The wave vector k_{μ}, which is normal to the hypersurface of constant phase is null $k^{\mu} k_{\mu} = 0$, and is geodesic

$$k^{\mu} \nabla_{\mu} k^{\nu} = 0, \tag{13}$$

as it is the gradient of a scalar function. Similarly, the polarization vector f^{μ} is parallel propagated along it

$$k^{\mu} \nabla_{\mu} f^{\nu} = 0. \tag{14}$$

A convenient three dimensional representation of the evolution of polarization vectors is possible in stationary spacetimes. Static observers follow the congruence of timelike Killing geodesics that define projection from spacetime manifold $\mathcal{M}$ onto the three dimensional space Σ_3, $\Pi : \mathcal{M} \to \Sigma_3$. The metric $g_{\mu\nu}$ on $\mathcal{M}$ can be written in terms of three dimensional scalar h, a vector $\mathbf{g}$ with component g_m and a metric γ_{mn} on Σ_3 as

$$ds^2 = -h \left(dx^0 - g_m dx^m \right)^2 + dl^2, \tag{15}$$

where $h = -g_{00}$, $g_m = -g_{0m}/g_{00}$, and the three dimensional distance $dl^2 = \gamma_{mn} dx^m dx^n$, where

$$\gamma_{mn} = g_{mn} - \frac{g_{0m} g_{0n}}{g_{00}}. \tag{16}$$

Using the relationships between the three and four dimensional covariant derivatives, the propagation equations in a stationary spacetimes result in the following three dimensional equations

$$\frac{D\hat{\mathbf{k}}}{Ds} = \mathbf{\Omega} \times \hat{\mathbf{k}}, \qquad \frac{D\hat{\mathbf{f}}}{Ds} = \mathbf{\Omega} \times \hat{\mathbf{f}}, \tag{17}$$

where D/Ds is the covariant derivative in three dimensional space with metric γ_{mn}. Thus, both the polarization and propagation vectors are rigidly rotated with an angular velocity

$$\mathbf{\Omega} = 2\boldsymbol{\omega} - (\boldsymbol{\omega}.\hat{\mathbf{k}})\hat{\mathbf{k}} - \mathbf{E}_g \times \hat{\mathbf{k}}, \tag{18}$$

where $\boldsymbol{\omega}$ and $\mathbf{E}_g$ could be interpreted as the gravitoelectric and gravitomagnetic field respectively

$$\boldsymbol{\omega} = -\frac{1}{2} k_0 \nabla \times \mathbf{g}, \qquad \mathbf{E}_g = -\frac{\nabla h}{2h}. \tag{19}$$

In flat spacetimes, polarization basis is uniquely fixed by the Wigner little group construction. However, in general curved background, Wigner construction must be performed at every point. This is because, in the absence of a global reference direction, the standard polarization triad $(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \hat{\mathbf{k}})$ is different at every locations. Given such choice, the net polarization can be found by starting with the initial polarization $f_{\text{fin}}(x_{\text{in}})$, parallel propagating it according to the rule Eq. (17) and then using the decomposition

$$f_{\text{fin}}(x_{\text{fin}}) = \cos\chi \hat{\mathbf{b}}_1 + \sin\chi \hat{\mathbf{b}}_2, \tag{20}$$

to read off an angle. For Schwarzschild spacetime, $\boldsymbol{\omega} = 0$, and thus differentiation of this equation gives

$$\frac{d\chi}{ds} = \frac{1}{\hat{\mathbf{f}} \cdot \hat{\mathbf{b}}_1} \hat{\mathbf{f}} \cdot \frac{D\hat{\mathbf{b}}_2}{Ds}. \tag{21}$$

This equation implies that polarization rotation is a pure gauge effect. The phase remains zero if the standard directions are set with the help of the local free fall acceleration $\mathbf{w}$ of a stationary observer. At each point in the spacetime we choose the direction of the standard reference momentum, or equivalently the z-axis of our standard polarization triad, to be $\hat{\boldsymbol{\zeta}} := \mathbf{w}$. For a photon with momentum $\mathbf{k}$ we choose the linear polarization vector $\hat{\mathbf{b}}_2$ to point in the direction $\hat{\boldsymbol{\zeta}} \times \hat{\mathbf{k}}$, and finally we choose $\hat{\mathbf{b}}_1 := \hat{\mathbf{b}}_2 \times \hat{\mathbf{k}}$ such that it completes the orthonormal triad $(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \hat{\mathbf{k}})$. This construction is known as the Newton gauge.[16] With this convention $\mathbf{\Omega} = -\mathbf{E}_g \times \mathbf{k} \equiv \Omega \hat{\mathbf{b}}_2$ and thus $\chi \equiv 0$ along the trajectory. However, such choice of standard polarization direction is practically unfeasible. We will see below the consequences of setting the z-axis with the help of a guide star.

Electromagnetic radiation and massless particles are not affected by Newtonian gravity. The post-Newtonian expansion[31] is conveniently organized in powers of $\epsilon^2 \sim GM/c^2\ell \sim v^2/c^2$, where $-GM/\ell$ is the (maximal) typical potential and v is

a typical velocity of massive particles. The parameter ϵ helps to keep track of the orders and is set to unity at the end of the calculations. The leading post-Newtonian contributions are of order ϵ^2; to take gravitomagnetic effects into account, we need contributions up to ϵ^3.

The post-Newtonian expansion of the metric near a single slowly rotating quasi-rigid gravitating body, up to ϵ^3, assuming that the underlying theory of gravity is general relativity is

$$ds^2 = -V^2(r)c^2dt^2 + \vec{R}\cdot d\vec{x}cdt + W^2(r)d\vec{x}\cdot d\vec{x}, \tag{22}$$

where

$$V(r) = 1 - \epsilon^2\frac{U}{c^2}, \qquad W(r) = 1 + \epsilon^2\frac{U}{c^2}. \tag{23}$$

The Newtonian gravitational potential $-U = -GMQ(r,\theta)/r \simeq -GM/r$ depends on the mass M and the higher order multipoles. The frame-dragging term is

$$\vec{R} = -\epsilon^3\frac{4G}{c^3}\frac{\vec{J}\times\vec{x}}{r^3}, \tag{24}$$

where $\vec{J}$ is the angular momentum of the rotating body. Hence, we see that the gauge invariant polarization rotation is absent in the leading order post-Newtonian expansion and the Faraday phase at order ϵ^2 is a reference frame effect. We again revert to the units $G = c = 1$. To obtain leading order contributions to the phase and polarization, photon trajectories only need to be expanded up to ϵ^2

$$\vec{k} = \vec{n} - \epsilon^2\frac{2M}{r(t)}\vec{n} - \epsilon^2\frac{2M\vec{d}}{d^2}\left(\frac{\vec{x}(t)\cdot\vec{n}}{r(t)} - \frac{\vec{x}_0\cdot\vec{n}}{r_0}\right), \tag{25}$$

where $\vec{x}(t) = \vec{x}_0 + (t-t_0)\vec{n} + \mathcal{O}(\epsilon^2)$ and $\vec{d} = \vec{x}_0 - (\vec{x}_0\cdot\vec{n})\vec{n}$ is the vector joining the centre of the earth and the point of closest approach of the unperturbed ray. Here, $\vec{n}$ is the initial propagation direction, $\vec{n}\cdot\vec{n} = 1$.

The leading order post-Newtonian metric is spherically symmetric. Note that the initial polarization that is perpendicular to the propagation plane remains perpendicular to it. So, we select the reference frame differently from that we have done in special relativistic calculations. Here we focus only on gravitational effects and treat them separately from the effects of rotation and relative motion. We take $z = 0$, the plane where the ray from GS to SC lies, set their velocities zero and consider the polarization vector

$$\vec{f} = (0,0,1) = \text{const}. \tag{26}$$

The gauge dependent Faraday phase results from the change in the definitions of standard polarization directions along the trajectory. The reference directions $\hat{\boldsymbol{\zeta}}_i$ (unit vector pointing to the distant star), $i =$ GS, SC are obtained from the tangents to the rays from the fixed guide star that arrive to the GS and SC respectively.

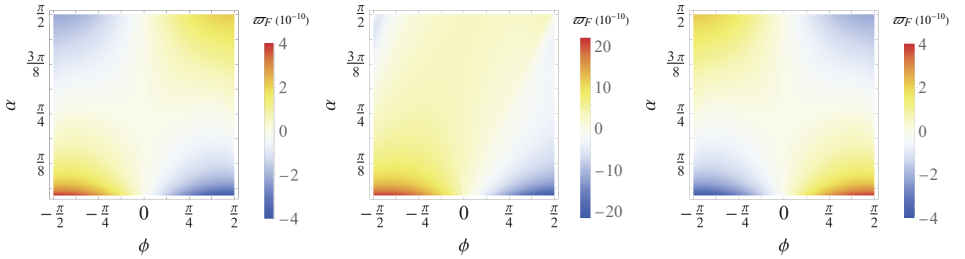


Fig. 2. Faraday phase $\varpi_{\rm F}$ depends on the choice of the reference direction $\vec{\zeta}(\alpha,\beta)$ and the initial propagation direction $\vec{n}(\phi)$. The angles are defined in Eqs. (31) and (32). *(left)* $\beta = 0$. *(center)* $\beta = 7\pi/16$. *(right)* $\beta = \pi$.

We assume the reference star to be infinitely far. Approximating the differences in reference directions as arising solely form the gravitational field of the earth,

$$\vec{\zeta}_i = \vec{l}\left(1 - \epsilon^2 \frac{2M}{r_i}\right) - \epsilon^2 \frac{2M\vec{d}_i}{d_i^2}\left(\frac{\vec{x}_i\cdot\vec{l}}{r_i} - 1\right), \tag{27}$$

where $\vec{l}\cdot\vec{l} = 1$; $-\vec{l}$ is the flat spacetime direction from the infinitely distant star to the observers. Standard polarization vectors at GS and SC could be defined as

$$\hat{\mathbf{b}}_{2i} = \frac{\vec{\zeta}_i \times \vec{k}}{|\vec{\zeta}_i \times \vec{k}|}, \qquad \hat{\mathbf{b}}_{1i} = \hat{\mathbf{b}}_{2i} \times \hat{\mathbf{k}}. \tag{28}$$

So, from Eq. (21), we get

$$\Delta\chi \cos\chi = \hat{\mathbf{f}}_{SC}\cdot\hat{\mathbf{b}}_{2SC} - \hat{\mathbf{f}}_{GS}\cdot\hat{\mathbf{b}}_{2GS} \tag{29}$$

or,

$$\Delta\chi = \frac{1}{\hat{\mathbf{f}}_{GS}\cdot\hat{\mathbf{b}}_{1GS}}(\hat{\mathbf{f}}_{SC}\cdot\hat{\mathbf{b}}_{2SC} - \hat{\mathbf{f}}_{GS}\cdot\hat{\mathbf{b}}_{2GS}). \tag{30}$$

If we choose the x-axis to pass through GS, then

$$\vec{n} = (\cos\phi, \sin\phi, 0), \qquad -\frac{1}{2}\pi < \phi < \frac{1}{2}\pi, \tag{31}$$

$$\vec{l} = (\cos\alpha, \sin\alpha\sin\beta, \sin\alpha\cos\beta), \qquad 0 \le \alpha \le \frac{1}{2}\pi, \tag{32}$$

where the altitude of the guided star is $\pi/2 - \alpha$. When the reference direction $\vec{\zeta}$ and the propagation direction $\vec{k}$ are collinear, then the standard polarization directions are undefined. If $\vec{\zeta}$ lies in the plane determined by GS, SC and centre of earth, then the Faraday phase is zero. Moreover, the post-Newtonian phase fails in the limit of $\alpha = 0$ or $\beta = \pi/2, 3\pi/2$. The plot of the Faraday phase for different reference and propagation direction is shown in Fig. 2.

4. Polarization-dependent trajectory

Taking into account the post-eikonal terms modifies the description of light propagation along the null rays that are geodesic trajectories of massless particles. These particles are still null, but their motion is no more geodesic. It is most conveniently described by using a specially designed null tetrad.[17,32] It is introduced as follows.

Trajectory $X(s)$ of a null particle with the tangent vector K can be written as

$$\frac{dX^\mu}{ds} = K^\mu, \qquad K^\mu K_\mu = 0. \tag{33}$$

This trajectory is typically is not a geodesic,[22,23] and the acceleration vector w is defined by

$$w^\mu := K^\nu K^\mu_{;\nu}, \qquad K^\mu w_\mu = 0. \tag{34}$$

The second tetrad vector n, $\mathsf{n}^2 = 0$ is chosen to satisfy

$$\mathsf{K} \cdot \mathsf{n} = -1. \tag{35}$$

A pair of complex conjugated null vectors, m and $\bar{\mathsf{m}} = \mathsf{m}^*$, are built from two spacelike vectors and satisfy

$$\mathsf{m} \cdot \bar{\mathsf{m}} = 1. \tag{36}$$

All other inner products vanish and the metric can be expressed as

$$g_{\mu\nu} = -K_{(\mu} n_{\nu)} + m_{(\mu} \bar{m}_{\nu)}. \tag{37}$$

It is possible to reparameterize the null curves and rescale the vectors of the null tetrad while preserving orthonormality relations,[17,32]

$$\mathsf{K} \to A\mathsf{K}, \qquad \mathsf{n} \to A^{-1}\mathsf{n}, \tag{38}$$

$$\mathsf{K} \to \mathsf{K}, \quad \mathsf{m} \to \mathsf{m} + a\mathsf{K}, \quad \bar{\mathsf{m}} \to \bar{\mathsf{m}} + a^*\mathsf{K}, \quad \mathsf{n} \to \mathsf{n} + a^*\mathsf{m} + a\bar{\mathsf{m}} + aa^*\mathsf{K}, \tag{39}$$

$$\mathsf{m} \to e^{i\phi}\mathsf{m}, \qquad \mathsf{m} \to e^{-i\phi}\bar{\mathsf{m}}. \tag{40}$$

Using this freedom it is possible to choose the tetrad in such a way[23] that the vector n is parallel propagated along the ray,

$$\nabla_{\mathsf{K}} \mathsf{n} = 0, \tag{41}$$

while

$$\nabla_{\mathsf{K}} \mathsf{m} = -\kappa \mathsf{n}, \qquad \nabla_{\mathsf{K}} \bar{\mathsf{m}} = -\kappa^* \mathsf{n}, \tag{42}$$

where

$$\kappa = -\mathsf{w} \cdot \mathsf{m} = -m^\mu K^\nu K_{\mu;\nu}. \tag{43}$$

With this choice of the null tetrad the polarization-dependent correction to the trajectory that takes into account the frequency ω is finite, and given by[23]

$$\frac{D^2 X^\mu}{Ds^2} = K^\nu K^\mu_{;\nu} = -\frac{i\sigma}{\omega} R^\mu{}_{\nu\alpha\beta} K^\nu m^\alpha \bar{m}^\beta =: F^\mu, \tag{44}$$

where $R^{\mu}{}_{\nu\alpha\beta}$ is the Riemann tensor. At this stage no assumptions about the spacetime were made. Effects of this term in the solar system are adequately described by the weak field expansion of the Schwarzschild metric. We will shortly see that in the leading order the polarization-dependent acceleration is of the order of ϵ^2. We also treat Eq. (44) as a correction to the null geodesic. Thus the uperturbed motion can be confined to the equatorial plane.

The trajectory is decomposed as

$$X^{\mu}(s) = x^{\mu}(s) + \alpha x^{\mu}_{\sigma}(s), \tag{45}$$

where $x^{\mu}(s)$ is a null geodesic (that we assume can be affinely parameterized by s), $x^{\mu}_{\sigma}(s)$ is the polarization-dependent correction, and the formal arbitrary infinitesimal parameter α is introduced to ease the perturbative manipulations. The tangent vector

$$K^{\mu} = k^{\mu} + \alpha\Lambda^{\mu}, \tag{46}$$

is null so,

$$k^{\mu}\Lambda_{\mu} = 0. \tag{47}$$

For the weak gravity regime it is convenient use the usual post-Newtonian expansion (the standard choice of the coordinates (t, x, y, z), but without switching to the coordinate time t as an evolution parameter). On the dimensional grounds we expect as the famous light deflection[29,31] is the ϵ^2 effect ($\Delta\phi = 2r_g/\ell$, where ℓ is the impact parameter and $r_g = 2M$ is the gravitational radius), the leading polarization-dependent effect is of the order $\alpha = \epsilon^2\lambda/\ell$, where λ is the characteristic wave length. Taking this into account and writing M explicitly, we note that at the leading order

$$k^{\mu} = k^{\mu}_{(0)} + Mk^{\mu}_{(1)}, \tag{48}$$

$$\Lambda^{\mu} = Mq^{\mu}, \tag{49}$$

where the components $k^{\mu}_{(0)}$ are constant. Taking into account that $q^{\mu}_{;\nu} = q^{\mu}_{,\nu} + \mathcal{O}(M)$, we have

$$K^{\nu}K^{\mu}_{;\nu} = Mk^{\nu}_{(0)}q^{\mu}_{,\nu} + \mathcal{O}(M^2). \tag{50}$$

It is convenient to describe the unperturbed null geodesic using the full Schwarzschild solution. The symmetry allows to restrict the motion to the equatorial plane. The tangent can be taken as

$$k^{\mu} = \left(\frac{1}{f(\bar{r})}, \pm\sqrt{1 - \frac{\ell^2}{\bar{r}^2}f(\bar{r})}, 0, \frac{\ell}{\bar{r}^2}\right), \tag{51}$$

where $f(\bar{r}) = 1 - 2M/r$. Here $\bar{r}$ is the Schwarzschild radial coordinate (i.e. the circumferential radius). In the weak field regime it is related to the isotropic radial coordinate $r = \sqrt{\vec{x}\cdot\vec{x}}$ as $\bar{r} \approx r + M$.

We used the freedom of rescaling the null vectors to set the energy of a fictitious photon to $E = 1$, so the reduced angular momentum $\ell := L/E$ equals to the impact

parameter.[17] The sign of k^r depends on whether the photon approaches the origin (i.e. the Sun), or moves away from it. In the equatorial plane we have $\mathsf{k}\nabla_{\mathsf{k}}\mathsf{k} = 0$.

The second vector of the null tetrad

$$n^\mu = \left(\frac{\bar{r}^2}{2\big(\bar{r}^2 - \ell^2 f(\bar{r})\big)}, \mp \frac{\bar{r} f(\bar{r})}{2\sqrt{\bar{r}^2 - \ell^2 f(\bar{r})}}, 0, \frac{\ell f(\bar{r})}{2\big(\bar{r}^2 - \ell^2 f(\bar{r})\big)} \right), \tag{52}$$

is chosen so that $\mathsf{k}\cdot\mathsf{n} = -1$, and the complex-conjugate pair of null vectors is given by

$$m^\mu = \frac{1}{\sqrt{2}} \left(\frac{i\ell}{\sqrt{\bar{r}^2 - \ell^2 f(\bar{r})}}, 0, \frac{1}{r}, \frac{i}{\sqrt{\bar{r}^2 - \ell^2 f(\bar{r})}} \right), \qquad \bar{m}^\mu = m^{\mu *}. \tag{53}$$

These three vectors are to be adjusted to satisfy the conditions required for the validity of Eq. (44). This adjustment is quite straightforward, as in our setting

$$\nabla_{\mathsf{k}}\mathsf{n} = i\alpha(\mathsf{m} - \bar{\mathsf{m}}), \tag{54}$$

and

$$\nabla_{\mathsf{k}}\mathsf{m} = i\alpha\mathsf{k}, \qquad \nabla_{\mathsf{k}}\bar{\mathsf{m}} = -i\alpha\mathsf{k}, \tag{55}$$

where

$$\alpha = \pm \frac{\ell(r - 3M)}{\sqrt{2} r (\bar{r}^2 - \ell^2 f(\bar{r}))}, \tag{56}$$

where the sign of α coincides with that of k^r.

In the leading order approximation the trajectories are geodesic, and the parallel transport of k results in $\kappa = 0$. In turn it enables the parallel transport of the tetrad if the function a satisfies

$$a_{,r} = -i\alpha/k^r = \frac{i\ell(\bar{r} - 3M)}{\sqrt{2}\big(\bar{r}^2 - \ell^2 f(\bar{r})\big)^{3/2}}, \tag{57}$$

where the sign does not change on transition from decreasing to increasing distance.

Null geodesics in the Schwarzschild spacetime are classified[17] according to the behaviour of the three roots of the equation $C(\bar{r}) = r^3 - \ell^2(\bar{r} - 2M) = 0$. Only one root is relevant for the trajectories we study,

$$\bar{r}_0 = \ell - M + \mathcal{O}(M^2). \tag{58}$$

Up to the terms of the third order in M the roots are $2M$ and $r \pm \bar{r}_0$. Using this we find

$$\int a_{,r} dr = \frac{i\ell(3rM - \bar{r}_0^2)}{\sqrt{2(\bar{r}^2 - \bar{r}_0^2)}\bar{r}_0^2} + \mathcal{O}(M^3), \tag{59}$$

that is sufficient to obtain the desired $a(\bar{r})$ at any point of the trajectory apart from $r = \bar{r}_0$.

The same expression is valid for both increasing and decreasing r. For $r > \ell \sim \bar{r}_0 \gg M$ we can approximate it as

$$a = -\frac{i\ell}{\sqrt{2}\sqrt{\bar{r}^2 - \ell^2}} - \frac{3iM}{\sqrt{2}\ell}\left(1 - \frac{\bar{r}^3}{(\bar{r}^2 - \ell^2)^{3/2}}\right) + \mathcal{O}(M^2). \tag{60}$$

It is convenient to express the acceleration F^μ in terms of the original tetrad vectors as

$$F^\mu = -\frac{i\sigma}{\omega} R^\mu{}_{\alpha\beta\gamma}(k^\alpha m^\beta \bar{m}^\gamma + a k^\alpha k^\beta \bar{m}^\gamma + a^* k^\alpha m^\beta k^\gamma). \tag{61}$$

Only one component is non-zero at the leading order:

$$F^\theta = -\frac{\sigma}{\omega}\frac{3\ell M}{\bar{r}^4}\left(\frac{1}{\sqrt{\bar{r}^2 - \ell^2 f(\bar{r})}} + \frac{\sqrt{2}a(\bar{r})\ell}{\bar{r}^2}\right). \tag{62}$$

Using again the decomposition of $C(\bar{r})$ it becomes

$$F^\theta \approx \frac{\sigma}{\omega}\frac{3\ell M}{\bar{r}^4\sqrt{\bar{r}^2 - \bar{r}_0^2}}\left(-1 + \frac{\ell^2}{\bar{r}^2}\right) = -\frac{\sigma}{\omega}\frac{3\ell M}{\bar{r}^6}\frac{\bar{r}^2 - \ell^2}{\sqrt{\bar{r}^2 - \bar{r}_0^2}}, \tag{63}$$

We evaluate the correction $x_\sigma(s)$ using the post-Newtonian formalism that was outlined above. It is important to note that to use only the leading order ϵ^2 while using the order of α polarization correction to describe a particular trajectory is unjustified, as the next post-Newtonian correction for the geometric optics term is still higher. Using the setting that is described below $\epsilon^2 = r_g/\ell = 4.2 \times 10^{-6}$, while $\lambda/\ell = 1.4 \times 10^{-9}$ for $\lambda = 1$m. However, as the exact general relativistic calculation within the geometric optics approximation leads to polarization-independent results, calculation of the polarization-dependent shift can be done using only the ϵ^2 expansion.

To obtain an estimate of the effect we neglect the difference between $\bar{r}_0$ and ℓ, essentially regularizing the expressions near the closest approach radius. The difference between the isotropic radial coordinate that is used in the post-Newtonian analysis and the Schwarzschild radial coordinate is of the order of $2M \ll \ell$ and affects only the higher-order terms. The unperturbed motion is confined to the z-plane. We can assume that $\vec{k} = (-1, 0, 0)$, that corresponds to $\ell > 0$. In this approximation

$$F^z = +\frac{3\sigma\ell M}{\omega r^5}\sqrt{r^2 - \ell^2} + \mathcal{O}(M^2), \tag{64}$$

while all other components are zero at the leading order. Then the explicit equation for the leading polarization-dependent correction is

$$q^z_{,x} = -\frac{3\sigma\ell}{\omega r^5}\sqrt{r^2 - \ell^2}. \tag{65}$$

We are interested in a light ray that comes from afar ("minus infinity"), grazes the sun and is detected at some finite (but large) distance $r_2 \gg \ell$, the initial off-plane displacement and velocity are zero. Similarly to the calculations of corrections

for the classical tests of general relativity[29,31] we use the unperturbed (i.e. flat spacetime trajectory with $r_0 = \ell$

$$y = \ell, \qquad x = -t. \tag{66}$$

Then the correction $\delta z = Mq^z$ is obtained by an elementary integration. For the far field ($r \gg \ell$) the trajectory approaches a straight line with

$$\theta_\infty = \frac{2\sigma}{\omega}\frac{M}{\ell^2}, \tag{67}$$

giving at the distance r from the sun the transversal shift

$$\Delta z = \frac{4M}{\omega\ell^2} r. \tag{68}$$

This shift coincides with the result obtained using the polarization-dependent part of the angle of deflection $\Delta\phi$ between in-and outgoing polarized rays that was reported in Ref. 20. On the other hand, this approximation is too crude to identify a convergence of the initially divergent trajectories that was reported in Ref. 22), and a more delicate approximations will be investigated in the follow-up work.

5. Summary

Describing propagation of electromagnetic waves in vacuum in terms of rays that follow null geodesics is a very good approximation in the high-frequency regime. Within this approximation the polarization rotation in the Schwarzschild metric, and as a result in the leading post-Newtonian approximation, is a purely gauge effect. This phase will be present as a consequence of practical methods of setting up reference frames in the Earth-to-spacecraft communications. However, for the near-Earth experiments these effects are typically about 10^{-5} weaker than the SR effects. Observable gravity-induced deviations from the geometric optics approximation are expected only in ultra-strong gravitational fields.

References

1. S.-K. Liao, *et al.*, *Nature* **549**, 43 (2017).
2. P. Kómár *et al.*, *Nature Phys.* **10**, 582 (2014); H. Dai *et al.*, *Nature. Phys.* **16**, 848 (2020).
3. D. K. L. Oi, *et al.*, *EPJ Quantum Tech.* **4**, 6 (2017); L. Mazzarella, *et al.*, *Cryptography* **4**, 7 (2020).
4. A. Peres and D. R. Terno, *Rev. Mod. Phys.* **76**, 93 (2004).
5. R. B. Mann and T. C. Ralph, *Relativisitc Quantum Information*, *Class. Quant. Grav.* **29** (22) (2012).
6. D. Rideout *et al.*, *Class. Quant. Grav.* **29**, 224011 (2012).
7. Landis, G. A., arXiv:1604.06351 (2016).
8. S. G. Turyshev, and V. T. Toth, *Phys. Rev. D* **96**, 024008 (2017).
9. V. T. Toth and S. G. Turyshev, *Phys. Rev. D* **103**, 124038 (2021).
10. F. Flamini, N. Spagnolo, and F. Sciarrino, *Rep. Prog. Phys.* **82**, 016001 (2019).

11. G. V. Skrotskiĭ, *Sov. Phys. Dokl.* **2**, 226 (1957); J. Plebanski, *Phys. Rev.* **118**, 1396 (1960)
12. M. Nouri-Zonoz, *Phys. Rev. D* **60**, 024013 (1999); M. Sereno, *Phys. Rev. D* **69**, 087501 (2004); S. M. Kopeikin and B. Mashhoon, *Phys. Rev. D* **65**, 064025 (2002).
13. R. F. Stark and P. A. Connors, *Nature* **266**, 429 (1977)
14. P. P. Kronberg, C. C. Dyer, E.M. Burbidge, and V. T. *Astrophys. J.* **613**, 672 (2004).
15. F. Fayos and J. Llosa, *Gen. Rel. Grav.* **14**, 865 (1982).
16. A. Brodutch, T. F. Demarie, and D. R. Terno, *Phys. Rev. D* **84**, 104043 (2011); A. Brodutch and D. R. Terno, *Phy. Rev. D* **84**, 121501((R) (2011).
17. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, Oxford, England, 1992).
18. P. M. Alsing and G. J. Stephenson, Jr, arXiv:0902.1399; M. Rivera-Tapia, A. Delgao, G. Rubilar, *Class. Quant. Grav.* **37**, 195001 (2020).
19. B. Mashhoon, *Phy. Rev. D* **11**, 2679 (1974).
20. P. Gosselin, A. Bérard, and H. Mohrbach, *Phy. Rev. D* **75**, 084035 (2007).
21. V. P. Frolov and A. A. Shoom, *Phy. Rev. D* **84**, 044026 (2011).
22. M. A. Oancea, J. Joudioux, I. Y. Dodin, D. E. Ruiz, C. F. Paganini, and L. Andersson, *Phys. Rev. D* **102**, 024075 (2020).
23. V. Frolov, *Phys. Rev. D* **102**, 084013 (2020).
24. E. Wigner, *Ann. Math.* **40**, 149 (1939). https://doi.org/10.2307/1968551.
25. W.-K. Tung, *Group Theory in Physics* (World Scientific, Singapore, 1985).
26. D. Han, Y. S. Kim, and D. Son, *Phys. Rev. D* **31**, 328 (1985). https://doi.org/10.1103/PhysRevD.31.328.
27. N. H. Lindner, A. Peres, and D. R. Terno, *J. Phys. A* **36**, L449 (2003).
28. P. K. Dahal and D. R. Terno, *Phys. Rev. A (in press)*, arXiv:2106.13426 (2021).
29. C.W. Misner, K. S. Thorn, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
30. A. I. Harte, *Gen. Relativ. Gravit.* **51**, 14 (2019).
31. C. M. Will, *Theory and Experiment in Gravitational Physics*, 2nd edition (Cambridge University Press, 2018).
32. H. Stephani, D. Kramer, M. A. H. MacCallum, C. Hoenselaers, and E. Herlt, *Exact Solutions of Einstein's Field Equations*, 2nd ed. (Cambridge University Press, Cambridge, 2003).

The Ginger project – preliminary results

C. Altucci[1,2], F. Bajardi[1,2], A. Basti[3,4], N. Beverini[3,4], S. Capozziello[1,2], G. Carelli[3,4,*], D. Ciampini[3,4], G. De Luca[5], R. Devoti[5], G. Di Stefano[5], A.D.V. Di Virgilio[4], F. Fuso[3,4], U. Giacomelli[3,4], A. Govoni[5], E. Maccioni[3,4], P. Marsili[3,4], A. Ortolan[6], A. Porzio[2,7], A. Simonelli[4], G. Terreni[4] and R. Velotta[1,2]

[1] *Università di Napoli Dipartimento di Fisica*
[2] *Instituto Nazionale di Fisica Nucleare – Sezione di Napoli*
[3] *Università di Pisa Dipartimento di Fisica "E. Fermi",*
[4] *Instituto Nazionale di Fisica Nucleare – Sezione di Pisa*
[5] *Instituto Nazionale di Geofisica e Vulcanolgia*
[6] *Instituto Nazionale di Fisica Nucleare – Laboratori Nazionali di Legnaro*
[7] *Consiglio Nazionale delle Ricerche – SPIN*
[*] *E-mail: giorgio.carelli@unipi.it*
https://web.infn.it/GINGER/index.php/it/home

GINGER (Gyroscopes IN General Relativity) project aims to directly measure the Lense–Thirring (LT) and de Sitter (dS) effects on the Earth and it is based on an array of underground Ring Laser Giroscopes (RLG), the most sensitive inertial sensors to measure the rotation rate of the Earth. Since LT and dS act on a ring laser as angular rotation vectors summed to the earth rotation rate, by using at least two gyroscopes it is possible to retrieve the General Relativity (GR) contribution to the rotation rate. The kinematic component is independently measured by the International Earth Rotation and Reference Systems Service (IERS) with very high accuracy. Ginger could also make it possible to discriminate among different theories, minimizing modeling. The measurement of the Earth angular rotation rate, in order to be fruitful for a fundamental physics test, has to have sensitivity of 1 part in 10^9 or better. The most recent analysis of our prototype GINGERINO data indicate a sensitivity better than 1 part 10^{12}, i.e. 0.1% of the LT term.

Keywords: Ring laser gyroscope, Lense–Thirring, gravito-magnetic effect, length of the day

1. Introduction

The debate on gravitational theories, to extend or modify General Relativity is very active for the issues related to ultra-violet and infra-red behavior of Einstein's theory. The first is connected to the Quantum Gravity problem, the other to Dark Matter and Dark Energy governing large scale structure and cosmology. Up to now, no final theory, capable of explaining gravitational interaction at any scale, has been formulated, and it is important to say that for the discrimination between different theories sensitivity breakthrough is certainly required. Earth based experiments can be in principle advantageous allowing a thorough analysis of the systematic, little modelling of external perturbation and the feasibility of upgrades. The GINGER project, an array of Sagnac gyroscopes, can be used for Earth based gravity measurements, in particular to constrain parameters of gravity theories, like scalar-tensor or

Horava-Lifshitz gravity, by considering their post-Newtonian limits matched with experimental. This effect has been measured by the Gravity Probe B space experiments (10% level)[1] and by Lageos and Lares, reaching the 4% level.[2,3] Gravity Probe B has finished its operational life, while Lares is still providing data, and Lares II will be launched soon. This test is based on reconstructing the trajectories of geodetic satellites, i.e. mechanical simple objects to minimise the effect of the different drags affecting the trajectories, and combining laser ranging data coming from different stations. The Lense-Thirring test using the reconstruction of the nodes of the trajectories, requires the independent map of the Earth gravity fields, independently measured, and the final result at the end is limited by the accuracy of the accuracy of the map. It is necessary to accurately model the zonal tides, since their influence on the nodes of the geodetic satellites is larger than the Lense Thirring itself. The test provided by GINGER, which being Earth based provides the measurement at fixed latitude, is very different. The Lense-Thirring, relying on the Sagnac effect, does not require to combine data coming from different places or to use the accurate map of the Earth gravitational field. Recently, it has been pointed out by Jay Tasson of Carlton that RLG can effectively contribute to the Lorentz Violation quest. In this respect Jay has shown that a level of sensitivity of 1 part in 10^9 for the Earth rotation rate would provide interesting measurements of two Lorentz-violating terms in the framework of the Standard-Model Extension. In one case, sensitivities that are competitive with recent laboratory and perhaps solar system tests would result.[4]

The Earth rotation is a tool to investigate fundamental physics, since it is related to GR terms, as de Sitter and Lense–Thirring, and can provide unique data to investigate Lorentz violation. Its usefulness for fundamental physic is connected to sensitivity, which is quite often express in relative to the average Earth rotation rate; the boundary to be meaningful to fundamental physics is to reach 1 part in 10^9 and long term continuous operation. Present high sensitivity ring laser gyroscope have already fulfilled those requirements.

The Sagnac effect states that two light beams counter propagating inside a closed path complete the path at different times, t_c and t_{cc} for clockwise and anti-clockwise beam respectively, if the closed path rotates; in general the Sagnac effect, $\Delta t = t_c - t_{cc}$ is proportional to the non reciprocity between the two paths. Based on the Sagnac effect, several devices have been developed mainly for inertial navigation. The closed path of the devices with higher sensitivity is defined by an optical ring cavity, and two different technology are feasible: passive and active. The Sagnac gyroscope is identified by its area vector, and measures the scalar product between the local angular rotation and its area vector, in return the gravitomagnetic, and geodetic components act as angular rotations summed to the cinematic ones. The passive gyroscopes requires an external laser source. The active device, called Ring Laser Gyro (RLG), has an active medium inside the cavity, to feed the two counter propagating beams. It is a rather convenient solution, since external laser source

with high spectral purity is not required. At present the RLGs, with perimeter from 10 up to 36 m, have the sensitivity record for angular rotation rate measurement.

RLGs are a very promising tool for GR measurements given their features:

- since the effect of the Earth is a function of the latitude, they works at a fixed latitude, can be oriented at will and an independent gravity map of the Earth is not needed;
- they are not averaged, the synchronization of different clocks is not necessary.

Moreover the working prototypes demonstrated unattended continuous operation for months, typically sub-prad/s sensitivity in 1 second of measurement, very large bandwidth, fast response, in principle as fast as milli-seconds and very large dynamic range. High sensitivity RLG are attached to the Earth crust, so are able to measure the Earth rotation rate Ω_{geo}; at present several RLGs are operative: in Germany, G[5] of the geodetic observatory of Wettzell and ROMY in the geophysical observatory of Bavaria,[6] GINGERINO in the underground Gran Sasso laboratory, in Italy,[7] ER1 of the University of Canterbury,[8] in New Zealand, and HUST-1[9] passive gyroscope at HUST, part of the TianQin project, at Wuhan in China.

2. Our text bench: The Earth rotation

The GINGER project is devoted to fundamental physics and it is based on an array of RLG, a first prototype, GINGERINO was built inside the underground Gran Sasso laboratory to validate the site for GINGER. The Earth angular rotation is the natural "test bench" of our prototype. The device is taking data with high duty cycle since several years. Being an inertial sensor, it is sensitive to Ω_{geo} and its variations as polar motion, annual and Chandler wobbles, tides and crust deformations; quantities constantly measured with very high accuracy by the international system IERS. All this signal can be used to investigate the characteristics and the sensitivity of GINGERINO. Its sensitivity has been carefully investigated with standard statistical means, using 103 days of continuous operation and the available geodesic measurements of the Earth angular rotation rate. Sensitivity of 0.1 frad/s appears, with 600 s bandwidth at frequency of 40 days, indicating the feasibility of 1 part 10^{13} of the Earth rotation rate, and accordingly for GINGER the gravitomagnetism test on Earth at the 0.1% level or even better. Fig. 2 shows a view of the experimental apparatus. The general relativity test will require high accuracy, not only sensitivity. It is certainly challenging, but it is an apparatus which will be able to provide unique information for fundamental physics, and to other research fields, as geodesy and geophysics in general.

RLGs have the drawback to deal with the non linear dynamic of the laser, a problem that has highly limited their diffusion. Our analysis of the data from

GINGERINO shows that the large disturbances induced by the dinamic of the laser can be subtracted by a suitable tecnique. We are confident that the technique is finally mature to propose GINGER, a project based in 10 years experience and which contains many experimental details aiming at improving the long term stability, reducing instrumental disturbances.

The true Sagnac signal ω_s is recovered taking into account the laser dynamic, and assuming that it is $\omega_s = \omega_{geo} + \omega_{local}$, where ω_{geo} indicates the scalar product of the total variations of the Earth with the RLG area vector, ω_{local} the signals of local origin, related to temperature fluctuations and local tilts. Comparing the recorded data with the IERS measurements, the different terms are recovered using linear regression and standard statistical means based on minimum square, providing clear indication about the sensitivity of the apparatus, for example see Fig. 1.[10]

GINGERINO operates far from external disturbances and protected from large thermal excursions in the deep underground environment of Gran Sasso. Its data are compared with the evaluated global signal provided by IERS. he analysis takes into account and eliminates the nonlinear laser disturbances and recovers the global IERS signal with all tiny features.[11,12] At the same time disturbances of an instrumental and local origin are estimated using the environmental signals as temperature and local tilts. The obtained residuals, i.e. the unmodeled part of the Sagnac signal indicate a rotational sensitivity below the frad/s level. Disturbances are mainly of an instrumental origin, suggesting the need for improvements in the mechanical design of the RLG structure. By injecting probe signals in the GINGERINO data, we conclude that the sensitivity is 0.1 frad/s with 600 s integration time, more than a factor one hundred below the expected noise for this class of instruments.[13,14] In the near future it will be necessary to investigate this aspect from a theoretical side, and to develop a detailed Monte Carlo study in order to carefully investigate the analysis procedure.

3. Next step

We are ready to build a large-frame 3-dimensional array of ring laser gyroscopes to be mounted inside the under- ground Gran Sasso Laboratory (LNGS). The Ginger final aim is to measure the Earth angular velocity in the laboratory co-rotating frame and to compare it with the Earth angular velocity as observed in the Cosmic inertial frame. GR foresees that, in low field approximation, the De Sitter or geo-electric effect and Lense–Thirring or geo-magnetic effect have to be taken into account. For this purpose it is necessary an instrumental accuracy of the order of 10^{-14} rad s^{-1}. Such an accuracy offers also the possibility of measuring fundamental geodetic parameters related to the fluctuations of the Earth rotational velocity and of the Earth axis orientation, the so-called 'length of the day' and 'polar motion'. In parallel to the experimental studies onto our running prototypes, we have investigated different array configurations. The goal was finding the most simple and

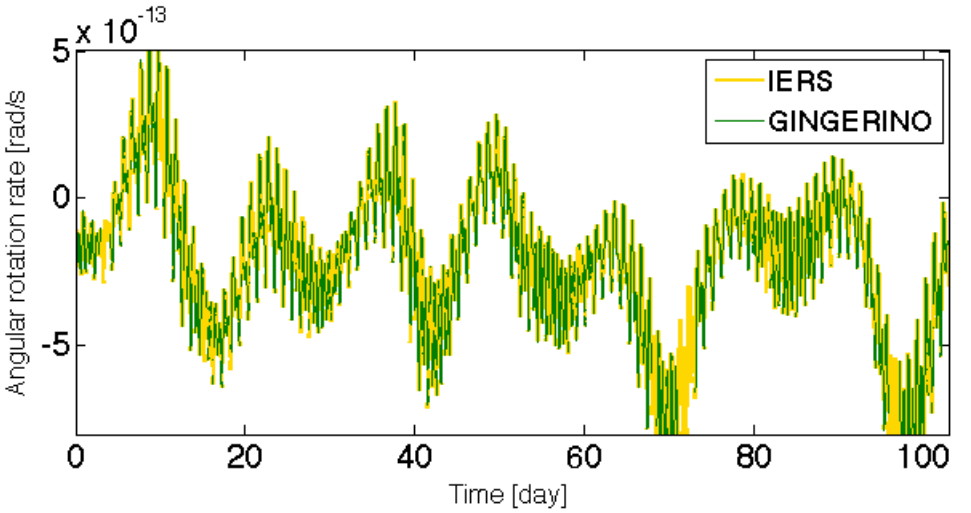


Fig. 1. Agreement of the variations of the LoD Earth rotation rate variations with higher frequency effects due to deformations, evaluated with the linear regression and the data provided by the international system IERS.

cost-effective setup, a sort of minimal instrument that enables the measurement of GR effects with the required reliability. We are fully aware that for this kind of very high sensitivity measurements, redundancy, i.e., the possibility of correlating measurements coming from different instruments and possibly different locations, is a key feature for reducing the influence of systematic as well as stochastic noise sources. In a "future scenario" a second apparatus, not necessarily co-located with the first one, is an option for further reducing the measurement uncertainty and, possibly, accessing further very tiny gravitational effects. As already stated, the final aim of GINGER is to reconstruct, with the required precision, the Earth rotation vector. In principle, measuring a 3D vector by scalar quantities, like the Sagnac frequencies, would require three independent measurements that fixes a local 3D frame. From this basic concept, we arrived, in 2011, at the octahedral configuration mentioned above. To reduce the number of required rings one should have an "external" information or measurement on the vector, or should be able to set two rings so that the plane defined by their oriented area vectors contains the Earth rotation one.[15] So doing this, the pair of rings define a plane where the rotation vector locally lies, and the problem dimensionality reduces from 3 to 2. We note that this plane, in the case of the Earth rotation, is nothing but the meridian one. Thus, in principle, a pair of rings would play the game and, if the conditions on their positioning are fulfilled with a given precision, the full vector can be reconstructed. Then, the problem movesto a different perspective, how to properly align the two rings so that the plane defined by their area vectors contains the Earth angular velocity vector.

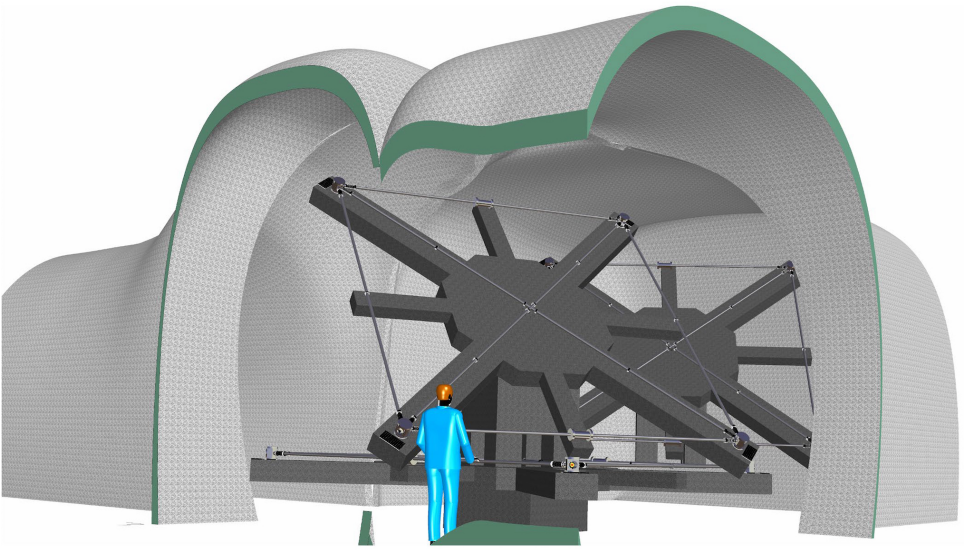


Fig. 2. Pictorial view of GINGER (Gyroscopes IN GEneral Relativity), based on an array of RLG, underground located in order to reduce the environmental disturbances and provide suitable very low frequency measurements

We have proven that the best configuration for the two rings is to orient one along the Earth rotation axis. With this choice, the Sagnac frequency of the ring is at its maximum value (a saddle point). As a consequence, orientation errors will affect only quadratically the measured signal and, very important fact, the RLG at maximum provides the modulus of the angular rotation rate. The relative orientation of the second ring is not so critical, although a horizontal ring can be preferable, owing to the possibility of locally defining its orientation by the local vertical. This is, however, a critical issue because its orientation with respect to the angular rotation axis must be known with a very high precision. The comparison with the RLG at the maximum value of the Sagnac frequency provides with adequate sensitivity the absolute orientation with respect to the rotational axis. The addition of a third RLG completes the measurement in all three degrees of freedom and provides redundancy, which is certainly welcome taking into account the very small signals we are looking for.

References

1. C. W. F. Everitt, D. B. DeBra, B. W. Parkinson, J. P. Turneaure, J. W. Conklin, M. I. Heifetz, G. M. Keiser, A. S. Silbergleit, T. Holmes, J. Kolodziejczak, M. Al-Meshari, J. C. Mester, B. Muhlfelder, V. G. Solomonik, K. Stahl, P. W. Worden, W. Bencze, S. Buchman, B. Clarke, A. Al-Jadaan, H. Al-Jibreen, J. Li, J. A. Lipa, J. M. Lockhart, B. Al-Suwaidan, M. Taber and S. Wang, Gravity probe b: Final results of a space experiment to test general relativity, *Phys. Rev. Lett.* **106**, p. 221101 (May 2011).

2. I. Ciufolini and E. C. Pavlis, A confirmation of the general relativistic prediction of the Lense–Thirring effect, *Nature* **431**, p. 958 (Oct 2004).
3. I. Ciufolini, A. Paolozzi, E. C. Pavlis, G. Sindoni, J. Ries, R. Matzner, R. Koenig, C. Paris, V. Gurzadyan and R. Penrose, An improved test of the general relativistic effect of frame-dragging using the lares and lageos satellites, *The European Physical Journal C* **79**, p. 872 (2019).
4. S. Moseley, N. Scaramuzza, J. D. Tasson and M. L. Trostel, Lorentz violation and sagnac gyroscopes, *Phys. Rev. D* **100**, p. 064031 (Sep 2019).
5. K. U. Schreiber and J.-P. R. Wells, Invited review article: Large ring lasers for rotation sensing, *Review of Scientific Instruments* **84**, p. 041101 (2013).
6. H. Igel, K. U. Schreiber, A. Gebauer, F. Bernauer, S. Egdorf, A. Simonelli, C.-J. Liny, J. Wassermann, S. Donner, C. Hadziioannou, S. Yuan, A. Brotzer, J. Kodet, T. Tanimoto, U. Hugentobler and J.-P. R. Wells7, ROMY: A Multi-Component Ring Laser for Geodesy and Geophysics, *Geophysical Journal International* **225** (2021), ggaa614.
7. A. Di Virgilio, J. Belfi, W.-T. Ni, N. Beverini, G. Carelli, E. Maccioni and A. Porzio, GINGER: A feasibility study, *The European Physical Journal Plus* **132**, p. 157 (Apr 2017).
8. D. Zou, R. J. Thirkettle, A. Gebauer, G. K. MacDonald, K. U. Schreiber and J.-P. R. Wells, Gyroscopic performance and some seismic measurements made with a 10 meter perimeter ring laser gyro housed in the ernest rutherford building, *Appl. Opt.* **60**, 1737 (Feb 2021).
9. K. Liu, F. Zhang, Z. Li, X. Feng, K. Li, Y. Du, K. Schreiber, Z. Lu and J. Zhang, Noise analysis of a passive resonant laser gyroscope., *Sensors* **20** (2020).
10. A. D. V. Di Virgilio, A. Basti, N. Beverini, F. Bosi, G. Carelli, D. Ciampini, F. Fuso, U. Giacomelli, E. Maccioni, P. Marsili, A. Ortolan, A. Porzio, A. Simonelli and G. Terreni, Underground sagnac gyroscope with sub-prad/s rotation rate sensitivity: Toward general relativity tests on earth, *Phys. Rev. Research* **2**, p. 032069 (Sep 2020).
11. A. D. V. Di Virgilio, N. Beverini, G. Carelli, D. Ciampini, F. Fuso and E. Maccioni, Analysis of ring laser gyroscopes including laser dynamics, *The European Physical Journal C* **79**, p. 573 (Jul 2019).
12. A. D. V. Di Virgilio, N. Beverini, G. Carelli, D. Ciampini, F. Fuso, U. Giacomelli and E. Maccioni, Identification and correction of sagnac frequency variations, *The European Physical Journal C* **80**, p. 163 (2020).
13. A. D. Di Virgilio, C. Altucci, F. Bajardi, A. Basti, N. Beverini, S. Capozziello, G. Carelli, D. Ciampini, F. Fuso, U. Giacomelli, E. Maccioni, P. Marsili, A. Ortolan, A. Porzio, A. Simonelli, G. Terreni and R. Velotta, Sensitivity limit investigation of a sagnac gyroscope through linear regression analysis, *The European Physical Journal C* **81**, p. 400 (2021).
14. A. Basti, N. Beverini, F. Bosi, G. Carelli, D. Ciampini, A. D. V. Di Virgilio, F. Fuso, U. Giacomelli, E. Maccioni, P. Marsili, G. Passeggio, A. Porzio, A. Simonelli and G. Terreni, Effects of temperature variations in high-sensitivity sagnac gyroscope, *The European Physical Journal Plus* **136**, p. 537 (2021).
15. A. Tartaglia, A. Di Virgilio, J. Belfi, N. Beverini and M. L. Ruggiero, Testing general relativity by means of ring lasers, *The European Physical Journal Plus* **132**, p. 73 (Feb 2017).

Varying fundamental constants and dark energy in the ESPRESSO era

C. J. A. P. Martins

Centro de Astrofísica da Universidade do Porto, and
Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto,
Rua das Estrelas, 4150-762 Porto, Portugal
E-mail: Carlos.Martins@astro.up.pt

The observational evidence for the recent acceleration of the universe shows that canonical theories of cosmology and particle physics are incomplete and that new physics is out there, waiting to be discovered. A compelling task for astrophysical facilities is to search for, identify and ultimately characterize this new physics. I present very recent developments in tests of the stability of nature's fundamental constants, as well as their impact on physics paradigms beyond the standard model. Specifically I discuss new observational constraints at low redshifts and at the BBN epoch, and highlight their different implications for canonical quintessence-type models and for non-canonical string-theory inspired models. Finally I also present new forecasts, based on realistic simulated data, of the gains in sensitivity for these constraints expected from ELT-HIRES, on its own and in combination with Euclid.

Keywords: Cosmology; Fundamental constants; Dark energy.

1. Introduction

The observational evidence for the acceleration of the universe shows that our canonical theories of cosmology and particle physics are at least incomplete, and possibly incorrect. Is dark energy a cosmological constant (i.e. vacuum energy)? If the answer is yes, it is ten to some large power times smaller than our Quantum Field Theory based expectations. If the answer is no, then the Einstein Equivalence Principle must be violated. Either way, new physics is out there, waiting to be discovered; we must search for, identify and characterize this new physics.

The CosmoESPRESSO team uses the universe as a laboratory to address, with precision spectroscopy and other observational, computational and theoretical tools, 6 grand-challenge questions:

(1) Are the laws of physics universal?
(2) Is gravity just geometry (i.e., a fictitious force)?
(3) What makes the universe accelerate?
(4) Can we find fossil relics of the early universe?
(5) How do we optimize next-generation facilities?
(6) How do we prepare next-generation (astro) physicists?

In what follows I will highlight recent contributions of the CosmoESPRESSO team to this fundamental quest, pertaining to fundamental constants.

2. Scalars, because they are there

We know since 2012 (thanks to the LHC) that fundamental scalar fields are among Nature's building blocks. Even before this discovery they were already widely used in cosmology, e.g., to describe inflation, cosmic defects, dynamical dark energy, and dynamical fundamental couplings. We also expect that cosmological scalar fields will naturally couple to the rest of the model, leading to long-range forces and "varying constants".[1–3] Of particular interest in what follows are electromagnetic sector couplings, which yield spacetime variations of the fine-structure constant, with multiple testable fingerprints. A recent theoretical and observational overview of the field can be found in Ref. 4.

One of the most actively pursued test of the stability of fundamental couplings consists of high resolution astrophysical spectroscopy measurements of the fine-structure constant α, the proton to electron mass ratio μ, the proton gyromagnetic ratio g_p, or combinations thereof. A recent joint likelihood analysis of all currently available data[5] leads to a very mild (one to two standard deviations of statistical significance) preference for astrophysical variations (as compared to the local laboratory values), at the parts per million level of relative variation, All this data spans the redshift range $0.2 < z < 4.2$. However, it is also known that, at least in the case of α measurements which make up most of the full dataset, there are systematics at the level of at least parts per million. A new generation of more robust measurements is therefore necessary.

3. Aiming higher

In the most natural and physically realistic models for time (redshift) variations of fundamental couplings, such variations are monotonic. Moreover, one expects that the putative scalar field driving this variation will be significantly damped at the onset of the dark energy domination phase of the universe (i.e., the recent acceleration phase). Both of these are theoretical motivations for testing the stability of fundamental constants such as α, though direct measurements, at higher redshifts—in other words, deep into the matter era, where any relative variations are expected to be larger.

An example of work towards this goal relies on using observations of the redshift $z = 7.085$ quasar J1120+0641[6] to constrain variations of α over the redshift range $z = 5.51$ to $z = 7.06$. These led to the four highest redshift direct measurements of α (the previous highest redshift direct measurement was at $z = 4.18$), with the latter corresponding to a look-back time 12.96 Gyr (for the current best-fit cosmological parameters in the standard ΛCDM model). A total of about 30 hours of data from the X-SHOOTER spectrograph on the European Southern Observatory's Very Large Telescope (VLT) was used, which also makes this the first direct measurement of α in the infrared. Finally, the analysis relied on a new AI-based method, aiming to remove some possible dependency of the final result on human-made choices.

The weighted mean strength of the electromagnetic force over this redshift range in this location in the universe, reported in Ref. 7, is measured to be

$$\frac{\Delta\alpha}{\alpha} = \frac{\alpha(z) - \alpha_0}{\alpha_0} = (-2.18 \pm 7.27) \times 10^{-5}\,, \tag{1}$$

where α_0 denotes the local laboratory value, i.e. we find no evidence for a time variation. The sensitivity of the combined measurement is only at the tens of parts per million level, as compared to the parts per million (nominal) sensitivity of optical measurements at lower redshifts, but this first result should primarily be seen as a proof of concept analysis, demonstrating that such high redshift measurements can indeed be done.

As for measurements in the optical, the arrival of the ESPRESSO spectrograph,[8] operating at the combined Coudé focus of the VLT, enables new and more stringent tests. Preliminary analyses already demonstrate that the dominant sources of systematics limiting previous high resolution spectrographs are not present,[9] and the first results of its measurements of the fine-structure constant should be reported soon.

Broadly speaking, the direct impact of ESPRESSO on cosmology and fundamental physics will come from at least five different types of observations:

(1) Direct measurements of the fine-structure constant α, at redshifts between $z \sim 1$ and $z \sim 4$, using various ions
(2) Direct measurements of the proton to electron mass ratio μ, at redshifts between $z \sim 2$ and $z \sim 4$, using molecular Hydrogen and Carbon Monoxide
(3) Direct measurements of the cosmic microwave background temperature at redshifts, at redshifts between $z \sim 2$ and $z \sim 4$, using Carbon Monoxide and neutral Carbon
(4) New measurements of the primordial Deuterium abundance (of relevance for BBN, as discussed in what follows)
(5) Additional probes, including deep spectra, lensed QSOs and precursor redshift drift measurements.

In addition to their direct fundamental relevance, these will also impact the quest to characterize dark energy properties.[10]

4. BBN with GUTs

Big Bang Nucleosynthesis is one of the cornerstones of the Hot Big Bang model, but its success has been limited by the long-standing Lithium problem;[11] more recently, a possible Deuterium discrepancy has also been suggested.[12]

If a fiducial theoretical model is chosen and the relevant sensitivity coefficients are known, BBN can be studied perturbatively. This approach is Well-known for relevant cosmological parameters, such as the neutron lifetime, number of neutrinos and baryon-to-photon ratio.[12] More recently, this has been self-consistently

extended for a broad class of Grand Unified Theories,[13–15] where all couplings are allowed to vary.

The Lithium problem can be expressed as a statistical preference for a 70% depletion of its theoretically predicted primordial (cosmological) abundance, with respect to the astrophysically measured one,[15] with otherwise standard physics. When the analysis is repeated for a broader class of GUT models, one finds[15] a mild (ca. two to three standard deviations) statistical preference for larger values of the fine-structure constant at the BBN epoch, at the parts per million level of relative variation, while the preferred Lithium depletion drops to about 65%. These results are qualitatively consistent across various models, although quantitatively the best-fit values of the relevant parameters do have a mild model dependence, further discussed in Ref. 15.

This means that the preference for a $\Delta\alpha/\alpha > 0$ is not due to the Lithium problem, but to the aforementioned Deuterium discrepancy. A few ppm variation of α solves D discrepancy, given their positive correlation. Such a variation is consistent with all other currently available cosmological, astrophysical and local constraints. This helps with the Lithium problem, reducing the astrophysical depletion required, but only by a moderate amount. Thus the most likely explanation for the Lithium problem is an astrophysical one, and Ref. 15 further shows that the amount of depletion needed to solve the Lithium problem can be accounted for by transport processes of chemical elements in stars—specifically, the combination of atomic diffusion, rotation and penetrative convection. On the other hand, the Helium4 abundance has relatively little statistical weight in the above analysis, given the very tight observational constraints on the primordial Deuterium abundance.

These results show that BBN is a very sensitive probe of new physics: it is quite remarkable that one can constrain the strength of the electromagnetic interaction, to parts per million level, when the universe was seconds to minutes old. Going forward, improving observed abundances of Deuterium and Helium4 by a factor of 2 to 3 will lead to stringent tests of this GUT class of models, and in particular will confirm or rule out the current preference for a $\Delta\alpha/\alpha > 0$. A cosmological measurement of the Helium3 abundance (which is currently not possible) would also provide a key consistency test of the underlying physics. These will be crucial tasks for the next generation of high-resolutions ultra-stable astrophysical spectrographs.

5. Dynamical dark energy

Realistic models for a varying fine-structure constant usually rely on a fundamental or effective scalar field, and can be phenomenologically divided into two broad classes, dubbed Class I and Class II, for which astrophysical tests of the stability of α will play different roles as probes of the underlying cosmological model.[4]

Class I contains the models where the same degree of freedom yields dynamical dark energy and the varying α. In this case the cosmological evolution of α is

parametrically determined, meaning that its redshift dependence can be expressed as a function of cosmological parameters (including the matter density and those describing the dark energy equation of state) together with one or more parameter describing the coupling of the scalar field to the electromagnetic sector. In these models, constraints on α constrain the dark energy equation of state,[16–18] and in particular may enable an observational discriminating test between freezing and thawing dark energy models.[19] The simplest example of this class are quintessence models, provided one does not forget the coupling to the electromagnetic sector.

A recent example are the Euclid forecast constraints on dark energy coupled to electromagnetism, with astrophysical and laboratory data.[20] These show that α measurements improve the Euclid dark energy Figure of merit by between 8 and 26 percent, depending on the correct fiducial model (with larger improvements occurring in the null case, where the fiducial model is ΛCDM). Inter alia, these forecasts confirm the expectation[21] that increasing redshift lever arm of the measurements, which the α data enables, is crucial.

Class II contains the models where the dynamical degree of freedom responsible for the varying α has a negligible effect on cosmological dynamics, and therefore has little or no contribution to the acceleration of the universe. Such models are identifiable through consistency tests comparing cosmological and astrophysical (or local) data, and in this case α measurements still constrain model parameters. Typically they do not directly constrain cosmological parameters, but they may still indirectly help constrain them by breaking degeneracies in the overall parameter space (which will include cosmological and particle physics parameters). The simplest example of this class are the Bekenstein type models.[22,23] In some cases, such as the Dirac-Born-Infeld type models, astrophysical measurements of α may be the only possible observational probe that, given already available constraints, can distinguish these models from ΛCDM.[24]

Importantly, in all these models the scalar field inevitably couples to nucleons, leading to Weak Equivalence Principle violations.[2,3] Therefore measurements of α constrain the Eotvos parameter η. The current bound, including the available α measurements together with the MICROSCOPE bound,[25] has been reported in Ref. 26, and is

$$\eta < 4 \times 10^{-15} \,. \tag{2}$$

This is three times stronger than the bound from MICROSCOPE alone, and 30 time tighter than the best ground-based direct bounds, with the caveat that the constraint includes a mild model dependence (ca. factor of 2). This will be further improved by ESPRESSO, and later on by ELT-HIRES. ESPRESSO observations can probably reach a sensitivity of 2×10^{-16} (about 5 times better than the final MICROSCOPE results, due to appear imminently) while the ELT-HIRES expected sensitivity is at the few times 10^{-18} level, similar to that of proposed STEP.

6. So what's your point?

The acceleration of the universe shows that canonical theories of cosmology and particle physics are incomplete, if not incorrect. Precision astrophysical spectroscopy provides a direct and competitive probe of the (still unknown) new physics that must be out there. This already provides highly competitive constraints, and will take an increasingly stronger role in the coming years.

So far one can say that nothing is varying at the few parts per million level of relative variation. This is already a very tight bound. It is orders of magnitude stronger than the few percent level of the constraints on the deviation of the dark energy equation of state from a cosmological constant, and also one order of magnitude stronger than the Cassini bound on the Eddington parameter. Tests of the stability of α also lead to the best available Weak Equivalence Principle constraint, a situation that is expected to remain even after MICROSCOPE's final results.

The ESPRESSO spectrograph is here, and new and more robust measurements are coming soon. Together with the final MICROSCOPE results, stringent new tests will become possible. In the longer term, the ELT will be the flagship tool in a new generation of precision consistency tests of fundamental physics, leading to competitive 'guaranteed science' implications for dark energy as well as unique synergies with other facilities, including Euclid and the SKA.

Acknowledgments

This work was financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), and by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the framework of the project POCI-01-0145-FEDER-028987 and PTDC/FIS-AST/28987/2017.

References

1. R. H. Dicke, Experimental relativity, in *Relativité, Groupes et Topologie: Proceedings, École d'été de Physique Théorique, Session XIII, Les Houches, France, Jul 1 - Aug 24, 1963*, 1964.
2. S. M. Carroll, Quintessence and the rest of the world, *PRL* **81**, 3067 (1998).
3. T. Damour and J. F. Donoghue, Phenomenology of the Equivalence Principle with Light Scalars, *Class. Quant. Grav.* **27**, p. 202001 (2010).
4. C. J. A. P. Martins, The status of varying constants: A review of the physics, searches and implications, *Rep. Prog. Phys.* **80**, p. 126902 (2017).
5. C. J. A. P. Martins and M. Vila Miñana, Consistency of local and astrophysical tests of the stability of fundamental constants, *Phys. Dark Univ.* **25**, p. 100301 (2019).
6. D. J. Mortlock *et al.*, A luminous quasar at a redshift of z = 7.085, *Nature* **474**, p. 616 (2011).
7. M. R. Wilczynska *et al.*, Four direct measurements of the fine-structure constant 13 billion years ago, *Sci. Adv.* **6**, p. eaay9672 (2020).

8. F. Pepe *et al.*, ESPRESSO at VLT - On-sky performance and first results, *Astron. Astrophys.* **645**, p. A96 (2021).
9. T. M. Schmidt *et al.*, Fundamental physics with ESPRESSO: Towards an accurate wavelength calibration for a precision test of the fine-structure constant, *Astron. Astrophys.* **646**, p. A144 (2021).
10. A. Leite, C. Martins, P. Molaro, D. Corre and S. Cristiani, Dark energy constraints from ESPRESSO tests of the stability of fundamental couplings, *PRD* **94**, p. 123512 (2016).
11. B. D. Fields, The primordial lithium problem, *Ann. Rev. Nucl. Part. Sci.* **61**, 47 (2011).
12. C. Pitrou, A. Coc, J.-P. Uzan and E. Vangioni, A new tension in the cosmological model from primordial deuterium?, *Mon. Not. Roy. Astron. Soc.* **502**, p. 2474 (2021).
13. M. T. Clara and C. J. A. P. Martins, Primordial nucleosynthesis with varying fundamental constants: Improved constraints and a possible solution to the Lithium problem, *Astron. Astrophys.* **633**, p. L11 (2020).
14. C. J. A. P. Martins, Primordial nucleosynthesis with varying fundamental constants: Degeneracies with cosmological parameters, *Astron. Astrophys.* **646**, p. A47 (2021).
15. M. Deal and C. J. A. P. Martins, Primordial nucleosynthesis with varying fundamental constants: Solutions to the Lithium problem and the Deuterium discrepancy, *Astron. Astrophys.* **653**, p. A48 (2021).
16. C. J. A. P. Martins and A. M. M. Pinho, Fine-structure constant constraints on dark energy, *Phys. Rev. D* **91**, p. 103501 (2015).
17. C. J. A. P. Martins, A. M. M. Pinho, R. F. C. Alves, M. Pino, C. I. S. A. Rocha and M. von Wietersheim, Dark energy and Equivalence Principle constraints from astrophysical tests of the stability of the fine-structure constant, *JCAP* **08**, p. 047 (2015).
18. C. J. A. P. Martins, A. M. M. Pinho, P. Carreira, A. Gusart, J. López and C. I. S. A. Rocha, Fine-structure constant constraints on dark energy: II. Extending the parameter space, *Phys. Rev. D* **93**, p. 023506 (2016).
19. J. M. A. V. Boas, D. M. N. Magano, C. J. A. P. Martins, A. Barbecho and C. Serrano, Distinguishing freezing and thawing dark energy models through measurements of the fine-structure constant, *Astron. Astrophys.* **635**, p. A80 (2020).
20. M. Martinelli *et al.*, Euclid: constraining dark energy coupled to electromagnetism using astrophysical and laboratory data (5 2021).
21. E. Calabrese, M. Martinelli, S. Pandolfi, V. F. Cardone, C. J. A. P. Martins, S. Spiro and P. E. Vielzeuf, Dark Energy coupling with electromagnetism as seen from future low-medium redshift probes, *Phys. Rev. D* **89**, p. 083509 (2014).
22. A. C. O. Leite and C. J. A. P. Martins, Current and future constraints on Bekenstein-type models for varying couplings, *Phys. Rev. D* **94**, p. 023503 (2016).
23. C. S. Alves, A. C. O. Leite, C. J. A. P. Martins, T. A. Silva, S. A. Berge and B. S. A. Silva, Current and future constraints on extended Bekenstein-type models for a varying fine-structure constant, *Phys. Rev. D* **97**, p. 023522 (2018).
24. V. C. Tavares and C. J. A. P. Martins, Varying alpha generalized Dirac-Born-Infeld models, *Phys. Rev. D* **103**, p. 023525 (2021).
25. P. Touboul *et al.*, Space test of the Equivalence Principle: first results of the MICROSCOPE mission, *Class. Quant. Grav.* **36**, p. 225006 (2019).
26. C. J. A. P. Martins and M. Prat Colomer, Fine-structure constant constraints on late-time dark energy transitions, *Phys. Lett. B* **791**, 230 (2019).

Testing the general relativistic nature of the Milky Way rotation curve with Gaia DR2

Mariateresa Crosta
Italian National Institute for Astrophysics, OATo, Turin, Italy
E-mail: mariateresa.crosta@inaf.it

The ESA mission Gaia directly measures the kinematics of the stellar component of the Galaxy with the goal to create the largest, most precise three-dimensional map of the Milky Way. The very core of the Gaia data analysis and processing involves General Relativity to guarantee accurate scientific products. In parallel, any Galactic model should be developed consistently with the relativistic-compliant kinematics delivered by Gaia. In this respect, this contribution presents the first test for a relativistic Galactic rotation curve with the Gaia second release products. Both a general relativistic model and a classical analogue were fit to the best-ever kinematics, derived exclusively from Gaia data, of a carefully selected homogenous sample of disk stars tracing the axisymmetric part of the Galactic potential. The relativistic rotation curve results statistically indistinguishable from its state-of-the-art dark-matter-based analogue. This supports the ansatz that the background geometry could drive the stellar velocities in the plane of our Galaxy far away from its center and mimic dark matter. Furthermore, one of Einstein's equations provides the necessary baryonic matter density to close the observed gap with respect to the expected Newtonian velocities without the need of extra mass.

Keywords: Gravitation; Galaxy: kinematics and dynamics; General Relativity; Relativistic Astrometry; Gravitational Astrometry; Local Cosmology; Dark Matter

1. General Relativity in the Gaia context

Gaia is the ESA cornerstone astrometric mission[1] - a wide European effort involving almost 450 scientists - launched in December 2013. Gaia implements Relativistic Astrometry for the data analysis and processing that, as part of fundamental physics, can be named Gravitational Astrometry.

Astrometry is the oldest branch of astronomy, it provides positions and motions of stars since Babilonian times. Astrometrists have compared tiny movements in the sky with increasing accuracy through the ages. In practice they measured smaller and smaller angles. The same measurement principle applies for the Gaia mission, but in order to achieve high precision, nowadays we need to operate from space. Gaia provides positions (direction and distance) along with velocities of over 1 billion stars in our Galaxy with an accuracy of up to 10 millionths-of arcsecond, i.e. 1-microarsecond (μas), that corresponds to the angular size of the star on the top of Mole as seen from Jupiter. At the end-of-mission the astrometric accuracies will be better than 5-10 μas for brighter stars and 130-600μas for faint targets. Nevertheless, smaller angles mean larger distances. In astrometry the location of an

object is considered reliable if the relative parallax error is less 10%. This implies that with 1 microsecond of accuracy we obtain reliable distances up to the Galactic scale.[2] Then, the final goal is unraveling the structure, the formation, and the evolution of the Milky Way (MW). As matter a fact, since Gaia is continuously scanning the sky, the main goal is to produce a map of about 1-2 billion celestial objects in 3 dimensions. Each celestial object will be observed on average 70 times (a true Galilean method applied to the sky!). Thanks to the combinations on the focal plane of astrometry (positions, proper motions parallaxes), photometry (spectral classification, photometric distances, brightness, temperature, mass, age, chemical composition) and spectroscopy (radial velocity, chemical abundances), the Gaia survey is detailed star by star along with the knowledge of their intrinsic properties.

According to different parallaxes accuracies, the Gaia scientific outputs span from the Solar System scale to the MW one and comprise, among many others, tests on General Relativity (GR), ten thousands exoplanet detections within 200 pc, accurate calibration of the HR diagram, accurate distance scale, cosmological tests such as local Λ-CDM predictions.[2] However, to guarantee all its scientific outputs Gaia requires an appropriate relativistic modelling of the observable and the observer. Gaia is operating at L2 point of the Earth-Sun system and the few-microarcsecond level of its measurements requires to consider the overlapping non-stationary weak gravitational fields due to the Solar System masses, in particular to account for the off-diagonal terms of the IAU metric. Then, Relativistic Astrometry implies a fully general-relativistic analysis of the light trajectory, from the observational data back to the positions of light-emitting stars.[3] Since Gaia is implementing absolute Relativistic Astrometry for the first time, inside the Gaia consortium for the Gaia data processing and analysis (DPAC) there exist two GR models, which set up two independent astrometric solutions for the celestial sphere. Among the six DPAC centers across Europe, the only one specialized in the treatment and validation of the satellite astrometric data is run by ALTEC (located at Turin) under the scientific supervision of the astromeric group INAF-OATo for ASI. The DPCT hosts the systems of the Astrometric Verification Unit (AVU), which is in charge, for DPAC, of the verification, through the Global Sphere Reconstruction (GSR), of the absolute astrometry achieved through the baseline astrometric model. The adoption of these two models applied to the same set of data, besides to be an independent verification tool to guarantee the Gaia scientific outputs (for example to control systematics), are in itself a GR test.[4] Discrepancies that cannot be fixed by the models deserve to be investigated as possible GR shortcomings.

2. Gravitational Astrometry at Milky Way Scale

For the Gaia-like observer the weak gravitational regime turns out to be "strong" when one has to perform high accurate measurements. The position and velocity data, comprising the outputs of the Gaia mission, are fully GR compliant. Given a

relativistic approach for the data analysis and processing, any subsequent exploitations should be consistent with the precepts of the theory underlying the astrometric model. Therefore, a fully relativistic model for the MW structure should be applied as well.

A GR structure of MW can assure a coherent laboratory for Local Cosmology, in particular for testing how well distances and kinematics at the scale of the MW disk compare with the Λ-CDM model predictions based on the Concordance Cosmological Model.[2]

Although the Λ-CDM considers homogeneous and isotropic solutions of the Einstein Equations (EE), in the most advanced simulations Λ-CDM cosmology assumes an average FRLW evolution while growth in structure is treated by Newtonian N-body simulations. Ray-tracing as true observables from Gaia can provide constraints on the parameters included in the simulations and the use of Gaia data must be parallel with the utilisation of the most advanced cosmological simulations with baryonic matter (gas and stars). In such a context, flat galactic rotation curves, a longest outstanding problem for astronomy, provide the main observational support to the hypothesis of a surrounding dark matter. The simplest explanation is that galaxies contain far more mass than can be accounted by the bright stellar objects residing in the galactic disk. Adding a surrounding "dark matter" halo allows a good fit to data.

On the other hand, stellar kinematics, as tracer of gravitational potential, is the most reliable observable for gauging different matter contributions. Rotation curves (RC) are distinctive features of spiral galaxies like our Milky Way, a sort of a kinematical/dynamical signature, like the HR diagram for the astrophysical content. And since Gaia by routinely scanning individual sources throughout the whole sky directly measures the (relativistic) kinematics of the stellar component, the rotation curve of the MW lends itself as a first test for a GR Galaxy with Gaia data. This seems to contradict the common assumption that at MW scale the relativistic effects are usually negligible and locally Newton approximation for the gravitational potential is retained valid at each point. In fact, the dynamics of galaxies is usually treated by considering the Newtonian limit of EE, in practice it boils down to solving Poisson's equation in order to derive the velocities tracing the observed RC. However, if one supposes a MW background geometry due a flat Minkowskian metric plus small perturbations as done for the Gaia relativistic models, the typical velocity of stars in our Galaxy is $v_{Gal} = 250$ km/s, thus the lowest order of accuracy of the perturbation terms to be retained for an hypothetical pN Galaxy metric is $(v_{Gal}/c)^2 \approx 100$mas and $(v_{Gal}/c)^3 \approx 120\mu$as. Since the individual DR2 astrometric error is $\leq 100\mu$as throughout most of its magnitude range,[16] "weak" relativistic effect could be relevant for a Gaia-like observer. This indicates that is worth investigating the small curvature limit in GR at MW scale, bearing in mind that it may not coincide with the Newtonian regime. To answer one needs to compare classical and GR models for the MW that derive from different field equations.

3. Classical and General Relativistic Models for the Milky Way

The classical model (MWC) for the MW structure takes into account several substructures, comprising a bulge, a thin and thick disc, and an halo. For the bulge one can adopt a Plummer density profile[5]

$$\rho_b = \frac{3b_b^2 M_b}{4\pi(r^2+b_b^2)^{5/2}}, \tag{1}$$

where, in cylindrical coordinates, the bulge spherical radius is $r=\sqrt{R^2+z^2}$, with $b_b = 0.3$ kpc the Plummer radius and M_b is the total bulge mass.

As for the thin and thick MW disks, one can consider a double-component stellar disk modelled as two Miyamoto-Nagai potentials. This function is also approximated with a double exponential disk as in McMillan[6] and Korol et al.[7] The most general description[5,8,9] is expressed in the form

$$\rho_d(R,z) = \frac{M_d b_d^2}{4\pi} \frac{\left[a_d R^2 + (a_d + 3\sqrt{z^2+b_d^2})(a_d+\sqrt{z^2+b_d^2})^2\right]}{\left[R^2+(a_d+\sqrt{z^2+b_d^2})^2\right]^{5/2}(z^2+b_d^2)^{3/2}}, \tag{2}$$

where M_d is the total (thin or thick) disk mass, a and b are the scale–length and scale–height, $b_{td} = 0.25$ kpc and $b_{Td} = 0.8$ kpc the thin and thick disk scale–heights, respectively.

Finally, for the MWC, a standard Navarro-Frank-White (NFW) model describes the DM halo (McMillan[6], Bovy[8], Navarro et al.[10])

$$\rho_h(r) = \rho_0^{halo} \frac{1}{(r/A_h)(1+r/A_h)^2}, \tag{3}$$

where ρ_0^{halo} is the DM halo density scale and A_h its (spherical) scale radius.

Each component contributes to define the potential in the Poisson's equation

$$\nabla^2\Phi = 4\pi G\rho_{tot} = 4\pi G(\rho_b + \rho_{td} + \rho_{Td} + \rho_h) \tag{4}$$

from which the velocity profile is derived

$$V^2 = R\,(d\Phi_{tot}/dR)\,. \tag{5}$$

On the other hand, Einstein's equation is in general very difficult to solve analytically, making it even the more difficult to detail a metric for the whole Galaxy as a multi-structured object. As first attempt, we adopted in Crosta et al.[11] an axisymmetric and stationary Galactic metric-disk and a simple relativistic model suitable to represent the Galactic disk as dust in equilibrium at a large distance from the centre.[12]

As well known, in a stationary and axisymmetric space-time there exist two commuting killing vector fields, k^α (time-like) and m^α (always zero on the axis of symmetry) and a coordinate system $\{t,\phi,r,z\}$ adapted to the symmetries,[13,14] whose line element for a rotating perfect fluid takes the form:

$$ds^2 = -e^{2U}(dt+Ad\phi)^2 + e^{2U}\left(e^{2\gamma}(dr^2+dz^2) + Wd\phi^2\right), \tag{6}$$

where $e^{2U}, e^{2\gamma}$ are conformal factors and U, A, W depend only on coordinates $\{r, z\}$. The time coordinate t (time-like far enough from the metric source) spans in the range $[-\infty, +\infty]$ and ϕ is the azimuthal angular coordinate in the range $[0, 2\pi]$.[13] For the general dust solution[14] we have[a]:

$$-e^{2U} = (k|k),\ -Ae^{2U} = (k|m),\ e^{-2U}W^2 - A^2e^{2U} = (m|m). \tag{7}$$

In addition

$$m^\alpha = \partial^\alpha_\phi,\ k^\alpha = \partial^\alpha_t,\ \partial_t g_{ij} = \partial_\phi g_{ij} = 0,\ g_{\phi a} = g_{ta} = 0, \tag{8}$$

where $a = r, z$. Because of the two dimensional Laplace equation one can choose $W = r^2$. For rigidly rotating dust (i.e. shearfree and expansionfree) $U = 0$ and there exist a time-like Killing vector (linear combination of k^α and m^α with constant coefficient) parallel to the four-velocity of the fluid u^α, i.e. the co-rotating one chosen by Balasin and Grumiller,[12] proportional to ∂^α_t.[14] Then, by setting $e^{2\gamma} \equiv e^\nu$, $N = -A$ and $e^{2U} = 1$, eq. (6) becomes the line element adopted by Balasin and Grumiller[12]:

$$ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta = -dt^2 + 2Nd\phi dt + (r^2 - N^2)d\phi^2 + e^\nu(dr^2 + dz^2). \tag{9}$$

Such a model may serve as a crude model for galaxies with the stars considered as dust grains. A rigidly rotating disk of dust is the universal limit of rigidly rotating perfect fluid configurations where the pressure-to-energy-density ratio vanishes. The dust assumptions implies that the masses inside a large portion of the Galaxy interact only gravitationally and reside far from the central bulge region. As a matter of fact, stellar encounters become effective below the parsec scale, on the other hand the Galaxy could be considered globally isolated around 25 kpc where flaring effects emerge, indicating the onset of external gravitational perturbations.

Given a stationary and axisymmetric space-time, the unit tangent vector field of a general spatially circular orbit can be expressed as

$$u^\alpha = \Gamma\left(k^\alpha + \beta m^\alpha\right), \tag{10}$$

where β is the constant angular velocity (with respect to infinity) and Γ the normalization factor. Equation (10) represents a class of observers that includes static ones ($\beta = 0$), and can be parametrized either by β or equivalently by the linear velocity, say ζ, with respect to the ZAMOs (Z^α), the locally non-rotating observer with zero angular momentum as:

$$u^\alpha = \gamma\left(e^\alpha_{\hat{0}} + \zeta^{\hat{\phi}} e^\alpha_{\hat{\phi}}\right), \tag{11}$$

where $\gamma = -(u|Z)$ is the Lorentz factor, $e^\alpha_{\hat{0}}$ is the unit normal to the t=constant hypersurfaces, and $e^\alpha_{\hat{\phi}}$ the ϕ unit direction of the orthonormal frame adapted to the ZAMOs. ZAMO frames, indeed, have a non-zero angular velocity with respect to flat infinity and move on worldlines orthogonal to the hypersurfaces t=constant.

[a]Symbol (|) stands for the scalar product relative to the chosen metric.

The line element (9) can be rewritten in terms of the lapse $M = r/\sqrt{(r^2 - N^2)}$ and the shift factor $M^\phi = N/(r^2 - N^2)$ as follows

$$ds^2 = -M^2 dt^2 + (r^2 - N^2)\left(d\phi + M^\phi dt\right)^2 + e^\nu (dr^2 + dz^2), \tag{12}$$

where

$$Z^\alpha = (1/M)(\partial_t - M^\phi \partial_\phi). \tag{13}$$

Then,[15]

$$\zeta^{\hat{\phi}} = \frac{\sqrt{g_{\phi\phi}}}{M}(\beta + M^\phi). \tag{14}$$

In order to trace the relativistic velocity profile from Gaia data, one has to consider that the local (static) barycentric observer as defined in the Barycentric Celestial Reference System (BCRS)[3] reduces to be $\propto \partial_t^\alpha$ far away from the Solar System.

In such a context the spatial velocity of the co-rotating dust of the BG model ($\beta = 0$) turns out to be proportional to the off-diagonal term of the chosen metric, indicated with N, which could imply gravitational dragging, namely

$$|V(r,z)| \equiv \zeta^{\hat{\phi}} = \frac{N(r,z)}{r} \propto g_{0\phi}. \tag{15}$$

The metric function N has been determined by Balasin and Grummiller by solving the set of the Einstein field equations and results as

$$N(r,z) = V_0(R_{out} - r_{in}) + \frac{V_0}{2}\sum_{\pm}\left(\sqrt{(z \pm r_{in})^2 + r^2} - \sqrt{(z \pm R_{out})^2 + r^2}\right), \tag{16}$$

which is limited by the constraint $|z| < r_{in}$ and depends only on parameters V_0, R_{out}, r_{in}, corresponding, respectively, to the flat regime velocity, the extension of the MW disk and the bulge radius.

Finally, the velocity profiles on the equatorial plane is

$$V^{BG}(R) = \frac{V_0}{R}\left(R_{out} - r_{in} + \sqrt{r_{in}^2 + R^2} - \sqrt{R_{out}^2 + R^2}\right). \tag{17}$$

4. Testing the Geometry-Driven Rotation Curve of the Milky Way with Gaia DR2

To study if the flatness of the MW RC is geometry driven, we selected stars from the Gaia DR2 archive[16] according to the following strict criteria: i) a complete Gaia DR2 astrometric dataset; ii) parallaxes good to 20%; iii) Gaia-measured velocity along the line of sight, namely radial velocity, with better than 20% uncertainties; iv) materialization of the sample by means of a cross-matched entry in the 2MASS catalog.[17] These steps allow to reconstruct a proper 6D phase-space location occupied by each individual star as derived by the same observer.

The requirements above left us with a very homogenous sample of 5277 early type stars and 325 classical type I Cepheids, where 99.4% of the sample span in a range

of 11 kpc and is below 1 kpc from the galactic plane (a characteristic scale height for the validity of the BG-RC model). To date it represents the best angular-momentum sustained stellar population of the Milky Way that better traces its observed RC. The quantities extracted from the Gaia DR2 archive are transformed from their natural ICRS reference frame to its galactocentric cylindrical counterpart.[18]

Since the parameter space is too large, a simple nonlinear fit could not be adeguate, the Markov Chain Monte Carlo method to determine the unknown parameters and their uncertainties was used.[11] There are several literature works providing the parameter values for each component. For the fit the more recent ones were adopted. In the following tables the best fit estimates as the median of the posteriors and their 1σ level credible interval are reported.

Table 1. Parameters of BG's model.[11] θ, σ_θ^- and σ_θ^+ are the mean and the 1σ credible interval limits from the posteriors of the parameters.

BG model	θ	σ_θ^-	σ_θ^+
r_{in} [kpc]	0.39	-0.25	+0.36
R_{out}[kpc]	47.87	-14.80	+23.96
V_0 [km/s]	263.10	-16.44	+25.93
e^{ν_0}	0.083	-0.014	+0.014

Table 2. Parameters of the MWC model.[11] θ, σ_θ^- and σ_θ^+ are the mean and the 1σ credible interval limits from the posteriors of the parameters.

MWC model	θ	σ_θ^-	σ_θ^+
$M_b[10^{10}\mathrm{M}_\odot]$	1.0	-0.4	+0.4
$M_{td}[10^{10}\mathrm{M}_\odot]$	3.9	-0.4	+0.4
$M_{Td}[10^{10}\mathrm{M}_\odot]$	4.0	-0.5	+0.5
a_{td}[kpc]	5.2	-0.5	+0.5
a_{Td}[kpc]	2.7	-0.4	+0.4
$\rho_0^{halo}[\mathrm{M}_\odot\mathrm{pc}^{-3}]$	0.009	-0.003	+0.004
A_h [kpc]	17	-3	+4

In figure 1, the red and blue curves show the best fit to the BG and MWC models, respectively. For the likelihood analysis reported in Crosta et al.[11] the two models appear almost identically consistent with the data, indicating that weak fields in GR can drive the MW rotation curve as the classical dark-matter-based analogue.

The fitted value r_{in}= 0.39 kpc, besides being quite close to the value adopted for the Plummer radius (b_b = 0.3 kpc), provides an independent measurements of the radial size of the MW bulge directly from the velocity data. The least constrained parameter of the BG model is the radial extension of the MW disk, as expected due to the limited radial coverage of the Gaia-only velocity data used.

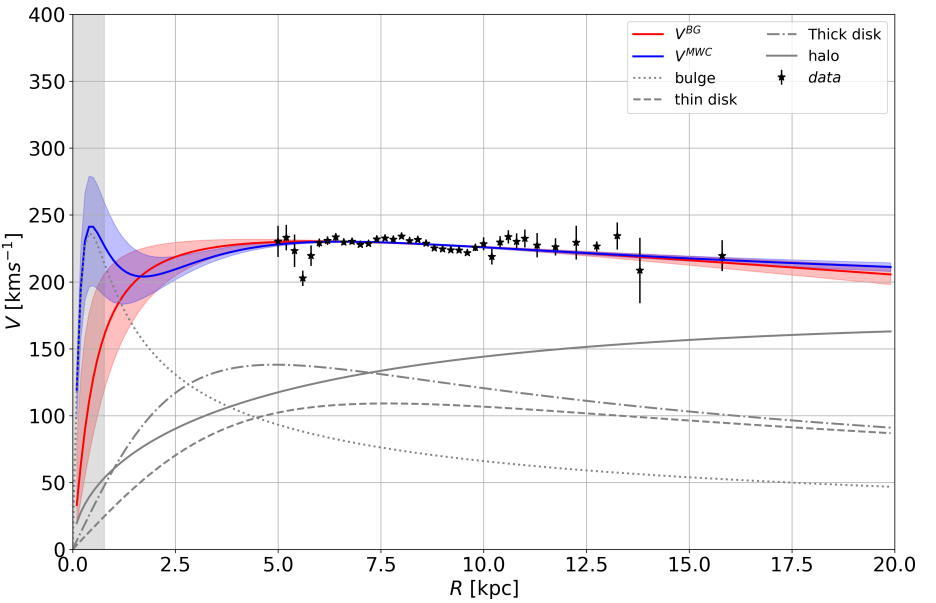


Fig. 1. From M. Crosta, M. Giammaria, M.G. Lattanzi and E. Poggio.[11] The azimuthal velocity profile of the MW as derived from the sample of disk tracers selected from the Gaia DR2. The black starred symbols represent the median values within a radial bin of about 0.2 kpc. The corresponding error bars are computed via bootstrapping. The other grey curves represent the kinematical substructures that contribute to the MWC model, in particular the grey solid line shows the contribution of the NFW halo. The coloured areas represent the reliability intervals of the fitted curves; note that for $R < 5$ kpc both the classical and the relativistic curves are very uncertain because of the lack of data in that region.The grey vertical band represents twice the value of r_{in} estimated with the BG model.

Differently form Balasin and Grummiller, in Crosta et al.[11] also the metric factor $e^{-\nu}$ was estimated via the Einstein equation for the density

$$\rho(R,z) = e^{-\nu(R,z)} \frac{1}{8\pi R^2} \left[(\partial_R N(R,z))^2 + (\partial_z N(R,z))^2 \right]. \tag{18}$$

Figure 2 two shows that the two models provide a similar baryonic mass distribution in the Galactic disk over the radial range explored by the Gaia data.

From the fit of the data the local relativistic baryonic matter density results $\rho(R = R_\odot, z = 0) \equiv \rho_\odot = 0.083 \pm 0.006\, \mathrm{M}_\odot \mathrm{pc}^{-3}$ that is in agreement with independent current estimates.[19–21] Note that, differently form the classical approach, the local mass density was used as the observed datum at Sun position in the likelihood function.

Finally, in order to disentangle the gravitational dragging component, which has not newtonian counterpart, we compared: i) the MWC baryonic-only contribution with the effective Newtonian profile V_{eN}^{BG}, which was calculated by implementing the BG density profile as described in Binney and Tremaine[22] and Almeida et al.[23];

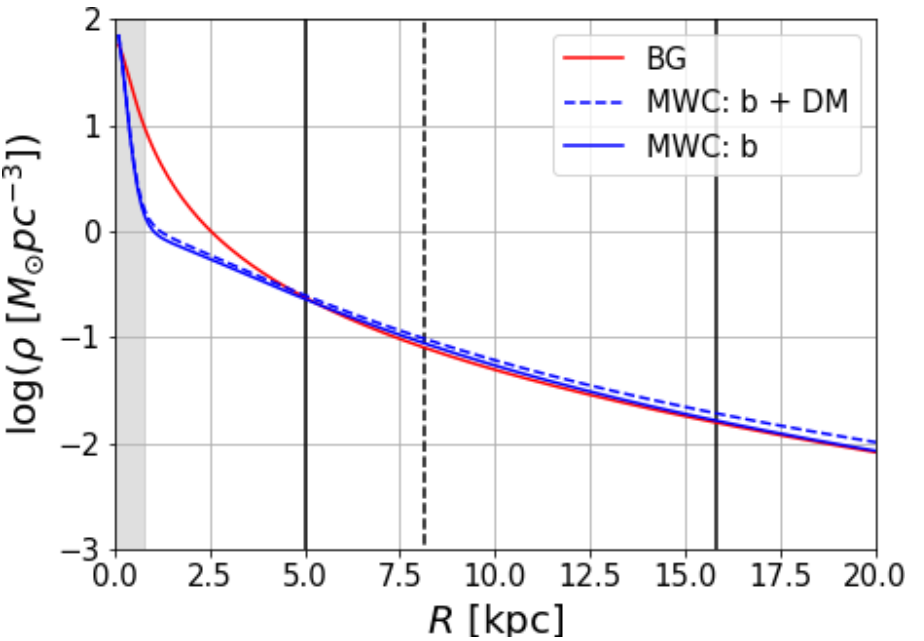


Fig. 2. From M. Crosta, M. Giammaria, M.G. Lattanzi and E. Poggio.[11] The density profile of the MW at $z = 0$ derived from 100 random draws from the posterior distribution of the fit. The red solid line is the BG model, while the blue dashed line represents the total matter contribution for the MWC model (i.e. the sum of the bulge and the two disks as the baryonic counterpart, plus the dark matter halo). The blue solid line shows the contribution of the sole baryonic matter. The vertical black solid lines limit the range of our data, while the vertical black dashed line indicates the Sun position in the Galaxy. Finally, the grey vertical band represents twice the value of r_{in} estimated with the BG model.

ii) the MWC dark matter-only contribution with the "dragging curve" traced by subtracting the effective component V_{eN}^{BG} to the total V^{BG}. More precisely, this last step implied a minimization procedure to get the effective BG disk half- thickness $|z|_{eff}$, that yields $|z|_{eff} = 0.215$ kpc (very close to the value of adopted for the classical thin disc), namely

$$(V_{drag}^{BG}(R_i; |z|_{eff}|) = \sqrt{(V^{BG}(R))^2 - (V_{eN}^{BG}(R; |z|_{eff}))^2}. \tag{19}$$

The results are illustrated in figure 3. Over the range covered by the Gaia data, the relativistic "dragging" component overlaps the one due to the halo in the MWC models. For R $\leq$ 5 kpc (the region we cannot constrain with the Gaia data at the moment), the two effective Newtonian velocities differ sharply. This could indicate the breaking point for the direct applicability of the BG model to the Milky Way, as it calls for a more suitable relativistic description of its central regions.

Despite the above limitations, the results points clearly to the possibility that a gravitational "dragging-like" effect due to the geometry background could sustain a flat RC.

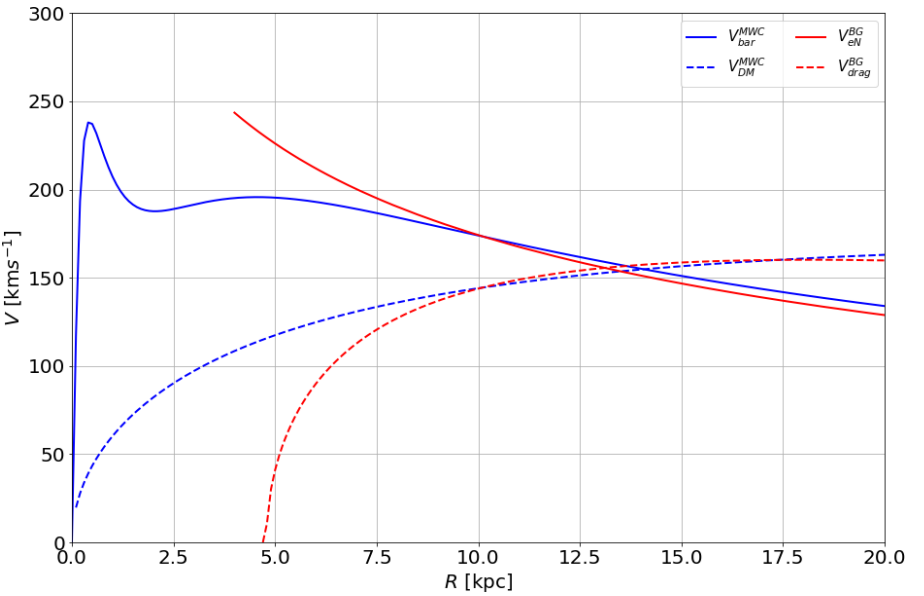


Fig. 3. From M. Crosta, M. Giammaria, M.G. Lattanzi and E. Poggio.[11] Red and blue colors refer to the BG and MWC model, respectively. Solid lines represent the relativistic effective Newtonian rotation curve, V^{BG}_{eN}, and its analogue for the MWC model, as contributed by the total of baryonic mass. The dashed lines show the MWC halo component alone, and the gravitational dragging contribution to V^{BG}, V^{BG}_{drag}, obtained by subtracting V^{BG}_{eN} from V^{BG} itself.

5. Concluding Remarks and Future Perspectives

By setting a coherent GR framework, one can effectively establish to what extent the MW structure is dictated by the standard theory of gravity. The paper Crosta et al.[11] explored the weak relativistic regime of Einstein's equation for the galactic dynamics and gives a possible interpretation of the MW RC as traced by the Gaia data by considering the background curvature, in some respects how space tells mass to move. If we look at DM proprieties, it does not absorb or emit light but it exerts and responds only to the gravity force; it enters the calculation as extra mass (halo) required to justify the flat galactic rotational curves. A gravitational dragging "DM-like" effect driving the MW velocity rotation curve could imply that geometry - unseen but perceived as manifestation of gravity according to Einstein's equation - is responsible of the flatness at large Galactic radii. In spite of some inadequacies, the simplified BG model adopted was proven also to be quite useful to estimate the (external) radial size of the Galactic bulge and the disc thickness at radial distances R $\geq$4 kpc. Obviously, more data and much improved mathematical models are necessary to confirm such a scenario.

For the observational side the sample size will increases from current 6000 to more than 100 thousands upper main sequence disc stars, with the addition of early-type B stars, in perspective of the full Gaia Data Release 3 expected in 2022.[24] Also

data from spectroscopic surveys (e.g. SDSS, APOGEE, LAMOST, RAVE, GES - Gaia ESO Survey, GALAH) will be merged.

For the theoretical side, the findings described in the previous section push on the fully use of Einstein's theory, point the lack of mathematical solutions for GR Galactic models and state the need to develop more complex background "geometries" to account for the MW multi-structures in concomitance of the incoming and increasingly accurate Gaia data releases along with other Galactic observations targeting the Galactic center. One attempt could be to study, e.g., the class of Lewis and Papapetrou metrics to encompass all the different MW structures and consider different conformal factors to be fitted with the Gaia data (as was done for the density in the BG model). Once a MW geometry is provided, one can extend it to other similar galaxies and include a fully GR kinematics.

With more physically appropriate metrics, along with adequate solution, the Galaxy can play a reference role for other galaxies, much like the Sun for stellar models.

Acknowledgements

This contribution has made use of data products from: the ESA Gaia mission (gea.esac.esa.int/archive/), funded by national institutions participating in the Gaia Multilateral Agreement; and the Two Micron All Sky Survey (2MASS,www.ipac.caltech.edu/2mass).

References

1. Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., Babusiaux, C. , Bailer-Jones, C. A. L., Bastian, U., Biermann, M., Evans, D. W. and et al., The Gaia mission, *A&A*, **595**, pp. A1 (2016).
2. M. G. Lattanzi, Astrometric cosmology, in *Memorie della Societá Astronomica Italiana*, **83**, p.1033 (2012), https://ui.adsabs.harvard.edu/abs/2012MmSAI..83.1033L.
3. M. Crosta, A. Geralico, M. G. Lattanzi, A. Vecchiato, General relativistic observable for gravitational astrometry in the context of the Gaia mission and beyond, *Phys. Rev. D*, **96**,104030 (2017).
4. M. Crosta, Astrometry in the 21st century. From Hipparchus to Einstein, *Nuovo Cimento Rivista Serie*, **42**, 10 (2019).
5. E. Pouliasis, P. Di Matteo, & M. Haywood, A Milky Way with a massive, centrally concentrated thick disc: New Galactic mass models for orbit computations, *A&A*, **598**, A66 (2017).
6. P. J. McMillan, The mass distribution and gravitational potential of the Milky Way, *MNRAS*, **465**, 76-94 (2017).
7. V. Korol, E. M. Rossi, and E. Barausse, A multimessenger study of the Milky Way's stellar disc and bulge with LISA, Gaia, and LSST, *MNRAS*, **483**, 5518-5533 (2019).
8. J. Bovy, Galpy: A python Library for Galactic Dynamics, *ApJ*, **216**, 29 (2015).
9. D. A. Barros, J. R. D. Lepine, & W. S. Dias, Models for the 3D axisymmetric gravitational potential of the Milky Way galaxy - A detailed modelling of the Galactic disk, *A&A*, **593**, A108 (2016).

10. J. F. Navarro, C. S. Frenk, and S. D. M. White, The Structure of Cold Dark Matter Halos, *ApJ*, **462**, 563 (1996).
11. M. Crosta, M. Giammaria, M. G. Lattanzi, and E. Poggio, On testing CDM and geometry-driven Milky Way rotation curve models with Gaia DR2, *MNRAS*, **496**, Issue 2 (2020).
12. H. Balasin, & D. Grumiller, Non-Newtonian behavior in weak field general relativity for extended rotating sources, *Int. Journal of Mod. Phys. D*, **17**, 475 (2008).
13. F. de Felice & J. S. Clarke, Relativity on curved manifolds (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1990).
14. H. Stephani, D. Kramer, M. Maccallum, C. Hoenselaers, and E. Herlt, Exact Solutions of Einstein's Field Equations. Second edition (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2009).
15. F. de Felice & D. Bini, Classical Measurements in Curved Space-Times (Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2010)
16. Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones et al., Gaia Data Release 2. Summary of the contents and survey properties, *A&A*, **616**, pp. A1(2018).
17. M. F. Skrutskie et al., The two micron all sky survey (2MASS), *AJ*, **131**, 1163 (2006).
18. F. Mignard et al., Gaia Data Release 2. The celestial reference frame (Gaia-CRF2), *A&A*, **616**, id.A14 (2018).
19. S. Garbari, C. Liu, J. I. Read, G. Lake G., A new determination of the local dark matter density from the kinematics of K dwarfs, *MNRAS*, **425**, 1445 (2012).
20. O. Bienaymé et al., Weighing the local dark matter with RAVE red clump stars, *A&A*, **571**, A92 (2014).
21. C. F. McKee, A. Parravano, D. J. Hollenbach, Stars, Gas, and Dark Matter in the Solar Neighborhood, *ApJ*, **814**, 13 (2015).
22. J. Binney & S. Tremaine, Galactic Dynamics, 2nd ed. (Princeton, NJ: Princeton Univ. Press, 2008).
23. A. O. F. de Almeida, O. F. Piattella, D. C. Rodrigues, A method for evaluating models that use galaxy rotation curves to derive the density profiles, *MNRAS*, **462**, Issue 3 (2016).
24. Details for the Data Release Scenario: http://www.cosmos.esa.int/web/gaia/release.

Spinning cylinders in general relativity: A canonical form for the Lewis metrics of the Weyl class

L. Filipe O. Costa* and José Natário†

GAMGSD, Departamento de Matemática, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, 1049-001, Portugal

** E-mail: lfilipecosta@tecnico.ulisboa.pt*

† E-mail: jnatar@math.ist.utl.pt

N. O. Santos

Sorbonne Université, UPMC Université Paris 06, LERMA, UMRS8112 CNRS, Observatoire de Paris-Meudon, 5, Place Jules Janssen, F-92195 Meudon Cedex, Paris, F-92195, France

E-mail: Nilton.Santos@obspm.fr

In the main article [CQG **38** (2021) 055003], a new "canonical" form for the Lewis metrics of the Weyl class has been obtained, depending only on three parameters — Komar mass and angular momentum per unit length, plus the angle deficit — corresponding to a coordinate system fixed to the "distant stars" and an everywhere timelike Killing vector field. Such form evinces the local but non-global static character of the spacetime, and striking parallelisms with the electromagnetic analogue. We discuss here its generality, main physical features and important limits (the Levi-Civita static cylinder, and spinning cosmic strings). We contrast it on geometric and physical grounds with the Kerr spacetime — as an example of a metric which is locally non-static.

Keywords: Gravitomagnetism; 1+3 quasi-Maxwell formalism; Sagnac effect; gravitomagnetic clock effect; synchronization gap; local and global staticity; Levi-Civita solution; cosmic strings.

1. Introduction

The general stationary solution of the vacuum Einstein field equations with cylindrical symmetry are the Lewis metrics[1–4]

$$ds^2 = -f(dt^2 + Cd\varphi)^2 + r^{(n^2-1)/2}(dr^2 + dz^2) + \frac{r^2}{f}d\varphi^2 \ ; \tag{1}$$

$$f = ar^{1-n} - \frac{c^2 r^{n+1}}{n^2 a} \ ; \qquad C = \frac{cr^{n+1}}{naf} + b \ , \tag{2}$$

usually interpreted as describing the exterior gravitational field produced by infinitely long rotating cylinders. They divide into two classes: (i) the Weyl class, when all the constants n, a, b, and c are real; (ii) the Lewis class, for n imaginary [implying in turn c real and a and b complex, in order for the line element (1) to be real[3,4]].

In the main article, Ref. 5, we have shown that the Weyl class metrics can be written in the "canonical" form

$$ds^2 = -\frac{r^{4\lambda_{\rm m}}}{\alpha}\left(dt - \frac{j}{\lambda_{\rm m} - 1/4}d\phi\right)^2 + r^{4\lambda_{\rm m}(2\lambda_{\rm m}-1)}(dr^2 + dz^2) + \alpha r^{2(1-2\lambda_{\rm m})}d\phi^2 \ ,$$

depending only on three parameters with a clear physical significance: the Komar mass ($\lambda_{\rm m}$) and angular momentum (j) per unit z-length, plus the parameter α governing the angle deficit. This form allows for a transparent comparison with the Levi-Civita non-rotating cylinder — archetype of the contrast between local and global staticity — which Ref. 5 discusses in detail both on physical and geometrical grounds. Therein its matching to the van Stockum[6] cylinder in star fixed coordinates is also shown. Here we revise some main features of the solution, focusing on its generality [and redundancies of the more usual form (1)-(2)], notable limits, and physical properties, with special attention to those less developed in Ref. 5. Addressing the question posed to us in the discussion following the presentation of this work at MG16,[7] we focus here on its comparison with a non-static (globally and locally) stationary solution, exemplified by the Kerr spacetime.

2. Stationary spacetimes and levels of gravitomagnetism

The line element $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$ of a stationary spacetime can generically be written as

$$ds^2 = -e^{2\Phi}(dt - \mathcal{A}_i dx^i)^2 + h_{ij}dx^i dx^j \ , \tag{3}$$

where $e^{2\Phi} = -g_{00}$, $\Phi \equiv \Phi(x^j)$, $\mathcal{A}_i \equiv \mathcal{A}_i(x^j) = -g_{0i}/g_{00}$, and $h_{ij} \equiv h_{ij}(x^k) = g_{ij} + e^{2\Phi}\mathcal{A}_i\mathcal{A}_j$. Observers of 4-velocity $u^\alpha = e^{-\Phi}\partial_t^\alpha \equiv e^{-\Phi}\delta_0^\alpha$, whose worldlines are tangent to the timelike Killing vector field ∂_t, are *at rest* in the coordinate system of (3). They are dubbed "static" or "laboratory" observers. The quotient of the spacetime by their worldlines yields a 3-D Riemannian manifold Σ with metric h_{ij} (called the "spatial metric"), which measures the spatial distances between neighboring laboratory observers.[8] It is identified in spacetime with the *space projector* with respect to u^α, $h_{\alpha\beta} \equiv u_\alpha u_\beta + g_{\alpha\beta}$. Let $U^\alpha = dx^\alpha/d\tau$ be the 4-velocity of a test point particle in geodesic motion. The space components of the geodesic equation $DU^\alpha/d\tau = 0$ yield[5,8–11]

$$\frac{\tilde{D}\vec{U}}{d\tau} = \gamma\left[\gamma\vec{G} + \vec{U}\times\vec{H}\right] \ ; \qquad \vec{G} = -\tilde{\nabla}\Phi \ ; \qquad \vec{H} = e^{\Phi}\tilde{\nabla}\times\vec{\mathcal{A}} \ , \tag{4}$$

where $\gamma = -U^\alpha u_\alpha$ is the Lorentz factor between U^α and u^α, $\tilde{\nabla}$ denotes covariant differentiation with respect to the spatial metric h_{ij} (i.e., the Levi-Civita connection of Σ, $\tilde{\nabla}_j X^i = X^i{}_{,j} + \Gamma(h)^i_{jk}X^k$, for some spatial vector $\vec{X}$), $\tilde{D}/d\tau \equiv U^i\tilde{\nabla}_i$, so that $\tilde{D}\vec{U}/d\tau$ describes the acceleration of the 3-D curve obtained by projecting the timelike geodesic onto the space manifold Σ, being $\vec{U}$ its tangent vector. The latter is identified in spacetime with the projection of U^α onto Σ: $(\vec{U})^\alpha = h^\alpha{}_\beta U^\beta$ [so its space

components equal those of U^α, $(\vec{U})^i = U^i$]. The spatial vectors $\vec{G}$ and $\vec{H}$ (living on Σ) are dubbed, respectively, "gravitoelectric" and "gravitomagnetic" fields [or, jointly, "gravitoelectromagnetic" (GEM) fields]. These play in Eq. (4) roles analogous to those of the electric ($\vec{E}$) and magnetic ($\vec{B}$) fields in the Lorentz force equation, $DU^i/d\tau = (q/m)[\gamma\vec{E} + \vec{U} \times \vec{B}]^i$, and are identified in spacetime, respectively, with minus the acceleration and twice the vorticity of the laboratory observers:

$$G^\alpha = -\nabla_{\mathbf{u}} u^\alpha \equiv -u^\alpha{}_{;\beta} u^\beta \; ; \qquad H^\alpha = \epsilon^{\alpha\beta\gamma\delta} u_{\gamma;\beta} u_\delta \; . \tag{5}$$

They motivate also dubbing the scalar Φ and the vector $\vec{\mathcal{A}}$ gravitoelectric and gravitomagnetic potentials, respectively.

Other realizations of the analogy arise in the equations of motion for a "gyroscope" (i.e., a spinning pole-dipole particle) in a gravitational field and a magnetic dipole in a electromagnetic field. According to the Mathisson-Papapetrou equations,[12,13] under the Mathisson-Pirani spin condition,[14] the spin vector S^α of a gyroscope of 4-velocity U^α is Fermi-Walker transported along its center of mass worldline, $DS^\alpha/d\tau = S^\mu a_\mu U^\alpha$, where $a^\alpha \equiv DU^\alpha/d\tau$. If the gyroscope's center of mass is at rest in the coordinate system of (3) ($U^\alpha = u^\alpha$) the space part of this equation yields[5]

$$\frac{d\vec{S}}{d\tau} = \frac{1}{2}\vec{S} \times \vec{H} \; , \tag{6}$$

which is analogous to the precession of a magnetic dipole in a magnetic field, $D\vec{S}/d\tau = \vec{\mu} \times \vec{B}$. When the electromagnetic field is non-homogeneous, a force is also exerted on the magnetic dipole, covariantly described by $DP^\alpha/d\tau = B^{\beta\alpha}\mu_\beta$,[14,15] where μ^β is the magnetic dipole moment 4-vector, and $B_{\alpha\beta} = \star F_{\alpha\mu;\beta} U^\mu$ ($F^{\alpha\beta} \equiv$ Faraday tensor, $\star \equiv$ Hodge dual) is the "magnetic tidal tensor" as measured by the particle. A covariant force is likewise exerted on a gyroscope in a gravitational field (the "spin-curvature" force[12–15]), which can be written in the remarkably similar form[14]

$$\frac{DP^\alpha}{d\tau} = -\mathbb{H}^{\beta\alpha} S_\beta; \qquad \mathbb{H}_{\alpha\beta} \equiv \star R_{\alpha\mu\beta\nu} U^\mu U^\nu = \frac{1}{2}\epsilon_{\alpha\mu}{}^{\lambda\tau} R_{\lambda\tau\beta\nu} U^\mu U^\nu. \tag{7}$$

Here $\mathbb{H}_{\alpha\beta}$ is the "gravitomagnetic tidal tensor" (or "magnetic part" of the Riemann tensor) as measured by the particle, playing a role analogous to that of $B_{\alpha\beta}$ in electromagnetism. For a particle at rest in a stationary field in the form (3), it is related to the gravitomagnetic field $\vec{H}$ by the expression[11]

$$\mathbb{H}_{ij} = -\frac{1}{2}\left[\tilde{\nabla}_j H_i + (\vec{G} \cdot \vec{H}) h_{ij} - 2G_j H_i\right] \; . \tag{8}$$

In the linear regime, $\mathbb{H}_{ij} \approx H_{i,j}$, and so one can say that (comparing to $\vec{H}$) $\mathbb{H}_{\alpha\beta}$ is essentially a quantity one order higher in differentiation of $\vec{\mathcal{A}}$.

2.1. *Sagnac effect*

By contrast with classical electromagnetism (where only the curl of the magnetic vector potential, $\nabla\times\vec{A}=\vec{B}$, manifests physically), and more like in quantum theory (where $\vec{A}$ manifests itself in the so-called Aharonov-Bohm effect[16]), in General Relativity there are also gravitational effects governed by the gravitomagnetic vector potential $\vec{\mathcal{A}}$ (or 1-form $\boldsymbol{\mathcal{A}}$). One of them is the Sagnac effect, consisting of the difference in arrival times of light-beams propagating around a spatially closed path in opposite directions. In flat spacetime, where the concept was first introduced (see e.g. Refs. 17, 18), the time difference is originated by the rotation of the apparatus with respect to global inertial frames (thus to the "distant stars"), see Fig. 1 in Ref. 5. In a gravitational field, however, it arises also in apparatuses which are fixed relative to the distant stars (i.e., to *asymptotically* inertial frames), Fig. 1(a) below, signaling frame-dragging.[5,18–22] In both cases the effect can be read off from the spacetime metric (3), encompassing the flat Minkowski metric written in a rotating coordinate system, as well as arbitrary stationary gravitational fields. Along a photon worldline, $ds^2=0$; by (3), this yields two solutions, the future-oriented one being $dt=\mathcal{A}_i dx^i+e^{-\Phi}dl$, where $dl=\sqrt{h_{ij}dx^i dx^j}$ is the spatial distance element. Consider photons constrained to move within a closed loop C in the space manifold Σ; for instance, an optical fiber loop, as depicted in Fig. 1 (a). Using the + (-) sign to denote the anti-clockwise (clockwise) directions, the coordinate time it takes for a full loop is, respectively, $t_\pm=\oint_{\pm C}dt=\oint_C e^{-\Phi}dl\pm\oint_C\mathcal{A}_i dx^i$; therefore, the Sagnac *coordinate* time delay $\Delta t_{\rm S}$ (Δt in the notation of Ref. 5) is

$$\Delta t_{\rm S}\equiv t_+-t_-=2\oint_C\mathcal{A}_i dx^i=2\oint_C\boldsymbol{\mathcal{A}}\,, \tag{9}$$

translating, in the observer's proper time, to $\Delta\tau_{\rm S}=e^{\Phi}\Delta t_{\rm S}$. We can thus cast gravitomagnetism into the three distinct levels in Table 1, corresponding to different orders of differentiation of $\boldsymbol{\mathcal{A}}$ (the first one being $\boldsymbol{\mathcal{A}}$ itself).

2.2. *Synchronization gap*

Another physical process where $\boldsymbol{\mathcal{A}}$ manifests is in the synchronization of the clocks carried by the "laboratory observers" (i.e., tangent to the Killing vector field ∂_t). Consider a curve $x^\alpha(\lambda)$ of tangent $dx^\alpha/d\lambda$ which is spatially closed (i.e., its projection onto Σ yields a closed curve C, so that after each loop it re-intersects the worldline of the original observer). Along $x^\alpha(\lambda)$, the synchronization through Einstein's light signaling procedure[8,21] amounts to the condition that the curve be orthogonal (at every point) to ∂_t^α, that is, $g_{0\beta}dx^\beta/d\lambda=0\Leftrightarrow dt=\mathcal{A}_i dx^i$. This curve will thus re-intersect the worldline of the original observer at a coordinate time $t_f=t_i+\Delta t_{\rm sync}$, where[8]

$$\Delta t_{\rm sync}=\oint_C\mathcal{A}_i dx^i=\oint_C\boldsymbol{\mathcal{A}}\quad(=\Delta t_{\rm S}/2)\,. \tag{10}$$

The observer will then find that his clock is not synchronized with his preceding neighbor's by a time gap (as measured in his proper time) $\Delta\tau_{\rm sync} = e^{\Phi}\Delta t_{\rm sync}$, corresponding, in coordinate time, to an interval $\Delta t_{\rm sync}$ [which is one half the Sagnac time delay $\Delta t_{\rm S}$ along such loop, Eq. (9)]. Only when $\Delta t_{\rm sync} = 0$ the observers are able to fully synchronize their clocks along a closed loop.

2.3. *Gravitomagnetic clock effect*

As is well known, in the field of a spinning body the periods of co- and counter-rotating circular geodesics differ; such an effect has been dubbed[23–26] gravitomagnetic "clock effect". The corresponding angular velocities read (see e.g. Sec. 3.1 of Ref. 5)

$$\Omega_{\rm geo\pm} = \frac{-g_{0\phi,r} \pm \sqrt{g_{0\phi,r}^2 - g_{\phi\phi,r}g_{00,r}}}{g_{\phi\phi,r}} \tag{11}$$

and thus the difference between their periods, $\Delta t_{\rm geo} = 2\pi(\Omega_{\rm geo+}^{-1} + \Omega_{\rm geo-}^{-1}) = -4\pi g_{0\phi,r}/g_{00,r}$, is[5,27]

$$\Delta t_{\rm geo} = \Delta t_{\rm S} + \Delta t_H; \qquad \Delta t_{\rm S} = 4\pi\mathcal{A}_\phi; \qquad \Delta t_H = -2\pi\frac{\star H_{r\phi}}{G_r e^{\Phi}} \ , \tag{12}$$

where $\star H_{jk} \equiv \epsilon_{ijk}H^i$ is the 2-form dual to the gravitomagnetic field $\vec{H}$, such that $\star H_{r\phi} = \sqrt{h}H^z$ in cylindrical coordinates, and $\star H_{r\phi} = -\sqrt{h}H^\theta$ in spherical coordinates. Here h is the determinant of the space metric h_{ij} in (3). Hence, the effect consists of the sum of two contributions of different origin, corresponding to two distinct levels of gravitomagnetism in Table (1): the Sagnac time delay $\Delta t_{\rm S}$ around the circular loop, governed by $\boldsymbol{\mathcal{A}}$, cf. Eq. (9), plus the term Δt_H due to the gravitomagnetic force $\gamma\vec{U}\times\vec{H}$ in Eq. (4) (which has a direct electromagnetic counterpart, see Sec. 3.1 in Ref. 5; cf. also Ref. 26).

The delay (12) corresponds to orbital periods (in coordinate time) as seen by the "laboratory" observers, at rest in the coordinates of (3). Other observers (e.g. rotating with respect to the former) will measure different periods since, from their point of view, the closing of the orbits occurs at different points. An *observer-independent* akin effect[25,28] can however be derived, based on the proper times (τ_+ and τ_-) measured by each orbiting particle between the events where they meet, see Fig. 1 (b). Set a starting meeting point at $\phi_+ = \phi_- = 0$, $t = 0$; the next meeting point is defined by $\phi_+ = 2\pi + \phi_-$. Since $\phi_\pm = \Omega_{\rm geo\pm}t$, the meeting point occurs at a coordinate time $t = 2\pi/(\Omega_{\rm geo+} - \Omega_{\rm geo-})$. Hence,

$$\tau_\pm = \frac{t}{U_\pm^0} = \frac{2\pi(U_\pm^0)^{-1}}{\Omega_{\rm geo+} - \Omega_{\rm geo-}} \ ; \qquad \Delta\tau \equiv \tau_+ - \tau_- = 2\pi\frac{(U_+^0)^{-1} - (U_-^0)^{-1}}{\Omega_{\rm geo+} - \Omega_{\rm geo-}} \ . \tag{13}$$

3. Artificial features of the usual form of the Weyl class metric

In the case of the static Levi-Civita cylinder, which follows from (1)-(2), with n and a real, by making[a] $b = 0 = c$, we have[b] $\Phi = 2\lambda_{\rm m} \ln(r) + const.$, $\mathbf{G} = -2\lambda_{\rm m}/r\mathbf{d}r$, $\boldsymbol{\mathcal{A}} = 0 = \vec{H}$, where $\lambda_{\rm m} = (1-n)/4$. This exactly matches the electromagnetic counterparts for an infinite static charged cylinder, if ones identifies $\lambda_{\rm m}$ with minus the charge per unit z-length. ($\lambda_{\rm m}$ being actually the Komar mass per unit z- length, as we shall see in Sec. 4.3). However, in the general case $b \neq 0 \neq c$, we have

- Φ, $\boldsymbol{\mathcal{A}}$, and $\vec{G}$ complicated [Eqs. (43) of Ref. 5], and very different from the electromagnetic counterparts for a infinitely long spinning charged cylinder (in the inertial rest frame);
- $\vec{H}$ and $\mathbb{H}_{\alpha\beta}$ both *non-zero* [and complicated, Eqs. (43) and (45) of Ref. 5], at odds with the electromagnetic analogue.

These features are somewhat unexpected given the similarities with the electromagnetic analogue in the static case, and given that this metric is known to be *locally static*. The situation resembles more the electromagnetic analogue as seen from a rotating frame. Moreover,

- ∂_t ceases to be time-like for $r^{2n} > a^2n^2/c^2 \Rightarrow$ no observers at rest are possible past this radius

which is, again, reminiscent of a rigidly rotating frame in flat spacetime where, past a certain value of r, the observers would be superluminal. The question then arises, can the metric, in the usual form (1)-(2) given in the literature, be actually written in some trivially rotating coordinate system? We will next show this to be the case.

4. The "canonical" form of the Weyl class metric

4.1. *"Star-fixed" coordinates: the metric with only three parameters*

An analysis of the curvature invariants [cubic and quadratic, Eqs. (39)-(41) and (50) of Ref. 5] reveals that the Weyl class metric (i.e., with n real) is a "purely electric" Petrov type I spacetime.[5,29,30] This means that, at each point, there is an (unique) observer for which the gravitomagnetic tidal tensor vanishes, $\mathbb{H}_{\alpha\beta} = 0$. These observers have 4-velocity of the form $U^\alpha = U^t(\partial_t^\alpha + \Omega\partial_\phi^\alpha)$, with *constant* angular velocity Ω given by

$$\Omega = \frac{c}{n - bc} \qquad \text{or} \qquad \Omega = -\frac{1}{b} \ , \tag{14}$$

the first (second) value yielding a time-like U^α if $a > 0$ ($a < 0$). Thus, by performing a coordinate rotation at constant angular velocity

$$\phi = \varphi - \Omega t \ , \tag{15}$$

[a] We shall see that the condition $c = 0$ is actually not necessary, cf. Eqs. (16)-(18).
[b] Taking $a > 0$, so that t is the temporal coordinate.

one switches to a coordinate system where these observers are at rest, and the metric takes the form

$$ds^2 = -\frac{r^{1-s}}{\alpha}(dt + \bar{C}d\phi)^2 + r^{(s^2-1)/2}(dr^2 + dz^2) + \alpha r^{1+s} d\phi^2 \ , \qquad (16)$$

with, for $\Omega = c/(n - bc)$,

$$s = n \ ; \qquad \alpha = \frac{\bar{C}^2}{ab^2} \ ; \qquad \bar{C} = b\frac{n - bc}{n} = b\frac{s - bc}{s} \ , \qquad (17)$$

and, for $\Omega = -1/b$,

$$s = -n \ ; \qquad \alpha = -ab^2 \ ; \qquad \bar{C} = -b\frac{n - bc}{n} = b\frac{s - bc}{s} \qquad (18)$$

Equation (16) shows that the metric depends only on three effective parameters: α, s, and $\bar{C}$, manifesting a redundancy in the original four parameters [different values of (n, a, b, c) yielding the same values of $(s, \alpha, \bar{C})$ correspond to the same physical solution]. The two values (14) for the angular velocity Ω are two equivalent paths of reaching (16), and manifest a particular case of the redundancy: two sets of parameters (a_1, b_1, c_1, n_1) and (a_2, b_2, c_2, n_2), with $a_1 > 0$ and $a_2 < 0$, such that the values of $(s, \alpha, \bar{C})$ are the same. There is one special case excluded from each of the transformations (14)-(15); namely, $bc = n$ for the first value of Ω, and $b = 0$ for the second; they are however redundant, as both lead to the Levi-Civita line-element[c] (26).

Observe that the Killing vector field ∂_t is, in (16), *everywhere time-like* ($g_{00} < 0$ for all r). Therefore, observers of 4-velocity $u^\alpha = (-g_{00})^{1/2}\partial_t^\alpha$, at rest in the coordinates of (16), exist everywhere (even for arbitrarily large r). As we shall see in Sec. (4.4) below, ∂_t is actually tangent to inertial observers at infinity, hence the reference frame associated to the coordinate system in (16) is asymptotically inertial, and thus fixed to the "distant stars".

4.2. *Symmetry under swap of time and angular coordinates*

In the transformation (15) one assumes, as is usual practice, that φ is an angular coordinate, ranging $[0, 2\pi]$, and t the time coordinate [which in turn implies $\alpha > 0$ and thus $a > 0$ in case (17), and $a < 0$ in case (18)]. Such assumption is however not necessary to reach (16). Indeed, swapping $t \leftrightarrow \varphi$ in (1), again leads to (16), as we shall now show. Substituting, in (1), $\varphi \to t'$, $t \to \varphi'$, the time-like observers measuring vanishing gravitomagnetic tidal tensor have now angular velocity $\Omega' = 1/\Omega$, where Ω is given by (14), and $\Omega' = (n - bc)/c$ yields a time-like U^α if $a > 0$, and likewise $\Omega' = -b$ for $a < 0$. Applying the transformation $\phi' = \varphi' - \Omega' t' = \varphi' - t'/\Omega$ to such line element leads to a primed version of (16),

$$ds^2 = -\frac{r^{1-s}}{\alpha}(dt' + \bar{C}d\phi')^2 + r^{(s^2-1)/2}(dr^2 + dz^2) + \alpha r^{1+s} d\phi'^2 \qquad (19)$$

[c] That it is so for $b = 0$ can be immediately seen by substituting in (16)-(17), yielding (26); likewise, that it is so for $n = bc$ can be seen by substituting $n \to bc$ in the expression for $\bar{C}$ in (18).

with, for $\Omega' = (n - bc)/c$, the identifications $s = n$, $\alpha = c^2/(an^2)$, $\bar{C} = c/n$; and, for $\Omega' = -b$, the identifications $s = -n$, $\alpha = -a$, $\bar{C} = -c/n$. This is a natural result. Equations (1)-(2) are just a general solution of the vacuum Einstein field equations with three commuting Killing vector fields, one time-like and two spacelike,[4,31] having no information on global aspects of the coordinates; it is thus natural that one be free to choose the angular coordinate rendering the axial symmetry.

One must note, however, that the metric in star fixed coordinates (16) does not preserve such symmetry. The coordinate rotation (15), with the identification $\varphi = \varphi + 2\pi$, breaks that symmetry by implicitly choosing ϕ (and φ) as a periodic coordinate, and t non-periodic. Indeed, substituting in (16) $\phi \to t'$, $t \to \phi'$, leads to (19) with $\phi' \leftrightarrow t'$ swapped:

$$ds^2 = -\frac{r^{1-s}}{\alpha}(\bar{C}dt' + d\phi')^2 + r^{(s^2-1)/2}(dr^2 + dz^2) + \alpha r^{1+s}dt'^2 \; ; \qquad (20)$$

forcing now on it the identification $(t', \phi') = (t', \phi' + 2\pi)$ (i.e, taking ϕ' to be periodic), makes it become the Levi-Civita metric in a rotating coordinate system, immediately diagonalizable through the coordinate rotation $\phi'' = \phi' + \bar{C}t$. This occurs because, by overriding the original identifications $(t, \phi) = (t, \phi + 2\pi)$, the geometry was *globally* (albeit not locally) changed. Indeed such transformation (with such identifications) is not a global diffeomorphism, as can be seen e.g. from the fact the ordered pairs $\mathcal{P}_1$: (t, ϕ) and $\mathcal{P}_2$: $(t, \phi + 2\pi)$, which represented the same event in the original metric (16), are mapped into the two different events $\mathcal{P}'_1$: (t', ϕ') and $\mathcal{P}'_2$: $(t' + 2\pi, \phi')$ in the metric (20). The transformation $(t', \phi') = (\phi, t)$, followed by $\phi'' = \phi' + \bar{C}t$, actually amounts to (76) of Ref. 5 which, as shown therein, corresponds to the "famous"[2,4] transformation that takes the Weyl class metric into the static Levi-Civita one.

4.3. *The metric in terms of physical parameters — "canonical" form*

The fact that in Eq. (16) the Killing vector field $\xi^\alpha = \partial_t^\alpha$ is everywhere time-like, tangent to inertial observers at infinity, and appropriately normalized,[5] allows for defining a corresponding Komar integral on simply connected tubes $\mathcal{V}$ of unit z-length parallel to the z-axis, having a physical interpretation of "active" gravitational mass per unit z-length, as discussed in Secs. 5.2.1 and 2.4 of Ref. 5 (cf. also Refs. 32–35). It is given by

$$\lambda_{\rm m} = \frac{1}{8\pi}\int_{\partial\mathcal{V}} \star\mathbf{d\xi} = \frac{1}{8\pi}\int_{\mathcal{S}} (\star d\xi)_{\phi z} d\phi dz = \frac{1-s}{4} \; ,$$

where $\partial\mathcal{V} = \mathcal{S} \cup \mathcal{B}_1 \cup \mathcal{B}_2$ is the tube's boundary, $\mathcal{S}$ its lateral surface, parameterized by $\{\phi, z\}$, and $\mathcal{B}_1$ and $\mathcal{B}_2$ its bases, lying on the planes orthogonal to the z-axis and parameterized by $\{r, \phi\}$, and in the second and third equalities we noticed that $(\star d\xi)_{r\phi} = 0$ and $(\star d\xi)_{\phi z} = 1 - s$. Likewise, the Komar integral associated with the

axial symmetry Killing vector field $\zeta^\alpha = \partial^\alpha_\phi$ has the interpretation of the spacetime's angular momentum per unit z-length,

$$j = -\frac{1}{16\pi}\int_{\partial\mathcal{V}} \star\mathbf{d}\boldsymbol{\zeta} = -\frac{1}{16\pi}\int_{\mathcal{S}} (\star d\zeta)_{\phi z} d\phi dz = \frac{s\bar{C}}{4}$$

where we in the second equality we noticed that $(\star\zeta)_{r\phi} = 0$. It follows that (16) can be re-written as

$$\boxed{ds^2 = -\frac{r^{4\lambda_{\rm m}}}{\alpha}\left(dt - \frac{j}{\lambda_{\rm m} - 1/4}d\phi\right)^2 + r^{4\lambda_{\rm m}(2\lambda_{\rm m}-1)}(dr^2+dz^2) + \alpha r^{2(1-2\lambda_{\rm m})}d\phi^2 \ .} \tag{21}$$

4.4. *Physical properties; gravitomagnetism*

For $\alpha > 0$ [so that t in Eq. (21) is a temporal coordinate], the metric can be put in the form (3), with

$$e^{2\Phi} = \frac{r^{4\lambda_{\rm m}}}{\alpha} \quad \Rightarrow \quad \Phi = 2\lambda_{\rm m}\ln(r) + K \ ; \qquad \boldsymbol{\mathcal{A}} = -\frac{4j}{1-4\lambda_{\rm m}}\mathbf{d}\phi \ ; \tag{22}$$

$$h_{ij}dx^i dx^j = r^{4\lambda_{\rm m}(2\lambda_{\rm m}-1)}(dr^2+dz^2) + \alpha r^{2(1-2\lambda_{\rm m})}d\phi^2 \tag{23}$$

[with $K \equiv -\ln(\alpha)/2$]. The gravitoelectric and gravitomagnetic 1-forms/fields read, cf. Eqs. (4),

$$\mathbf{G} = -\frac{2\lambda_{\rm m}}{r}\mathbf{d}r \ ; \qquad \vec{H} = \mathbf{H} = 0 \ . \tag{24}$$

Thus Φ, $\mathbf{G}$ and $\mathbf{H} = 0$ match their electromagnetic counterparts for a spinning charged cylinder (as viewed from the inertial rest frame, see e.g. Sec. 4 in Ref. 5) identifying the Komar mass with *minus* the charge, $\lambda_{\rm m} \leftrightarrow -\lambda$; the gravitomagnetic potential 1-form $\boldsymbol{\mathcal{A}} = \mathcal{A}_\phi\mathbf{d}\phi$ also resembles the magnetic potential 1-form $\mathbf{A} = \mathfrak{m}\mathbf{d}\phi$. The cylinder's rotation does not manifest in the inertial forces (nor in the tidal forces, as shown in Sec. 5.2.3 of Ref. 5); the only inertial force acting on test particles is the gravitoelectric (Newtonian-like) force $m\vec{G}$, independent of j. Thus, particles dropped from rest or in initial radial motion move along radial straight lines, cf. Eq. (4); and circular geodesics have a *constant* speed given by $v_{\rm geo} = \sqrt{\lambda_{\rm m}/(1/2-\lambda_{\rm m})}$, being thus possible when $0 \leq \lambda_{\rm m} < 1/4$ (it is when $\lambda_{\rm m} > 0$ that $\vec{G}$ is attractive, and they become null for $\lambda_{\rm m} = 1/4$). The vanishing of $\vec{H}$ means also that gyroscopes at rest in the coordinates of (21) do not precess, the components of their spin vector $\vec{S}$ remaining constant, cf. Eq. (6). Since $\vec{G} \overset{r\to\infty}{\to} \vec{0}$, it follows moreover that the reference frame associated to the coordinate system in (21) is *asymptotically* inertial, thus one can take it as the rest frame of the "distant stars" ("star-fixed" frame).

The only surviving gravitomagnetic object is thus the gravitomagnetic 1-form $\boldsymbol{\mathcal{A}}$, corresponding to the first level in Table 1. One of its physical manifestations is the

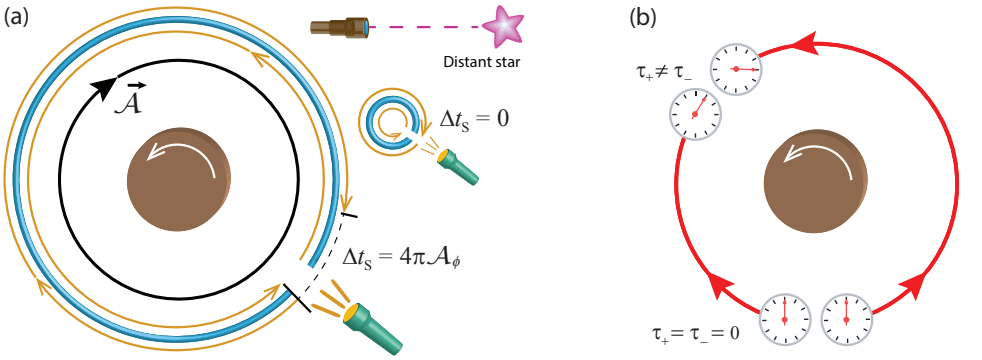


Fig. 1. (a) Sagnac effect around spinning cylinders: a flashlight sends light beams propagating in opposite directions along optical fiber loops fixed with respect to the "distant stars" (i.e., to the asymptotic inertial frame at infinity). In each loop C, the difference in beam arrival times is $\Delta t_S = 2\oint_C \boldsymbol{\mathcal{A}}$; since $\boldsymbol{\mathcal{A}}$ is a *closed* form ($\mathbf{d}\boldsymbol{\mathcal{A}} = 0$), the effect vanishes for all loops not enclosing the cylinder, and has the same value (25) for all loops enclosing it (the co-rotating beam arriving first). (b) Frame independent gravitomagnetic clock effect: a pair of clocks in oppositely rotating circular geodesics around a cylinder; when the cylinder spins ($j \neq 0$) the clocks measure different proper times between the events where they meet, $\tau_+ > \tau_-$.

Sagnac effect: consider, as depicted in Fig. 1(a) optical fiber loops fixed with respect to the distant stars, i.e., at rest in the coordinate system of (21). The difference in arrival times for light beams propagating in opposite directions along any of such loops is given by the circulation of $\boldsymbol{\mathcal{A}}$ along the loop, c.f. Eq. (9). Observe that $\boldsymbol{\mathcal{A}}$ is a closed form, $\mathbf{d}\boldsymbol{\mathcal{A}} = 0$ (since $\mathcal{A}_\phi$ is constant), defined in a space manifold Σ homeomorphic to $\mathbb{R}^3\backslash\{r = 0\}$. By the Stokes theorem, this means[5] that the effect vanishes along any loop which does not enclose the central cylinder (or the axis $r = 0$), such as the small loop in Fig. 1 (a), and has the *same* nonzero value

$$\Delta t_S = 4\pi\mathcal{A}_\phi = -\frac{4\pi j}{1/4 - \lambda_m} \tag{25}$$

along any loop enclosing the cylinder, regardless of its shape; for instance, the large circular loop depicted in Fig. 1(a). It is worth noticing that this mirrors the situation for the Aharonov-Bohm effect around spinning charged cylinders, which is likewise independent of the shape of the paths; the two effects are actually described by formally analogous equations[d].

The apparatus above makes use of a star-fixed reference frame, which is physically realized by aiming telescopes at the distant stars.[36,37] It is possible, however (still based on the Sagnac effect) to detect the cylinder's rotation in a more local way, without the need for setting up a specific frame; only not with a single loop, as along a single loop the effect can always be made to vanish by spinning it with some angular velocity. In particular, for a concentric circular loop, the effect

[d]Re-writing (9) in terms of the half-loop phase delay $\Delta\varphi = (E/\hbar)\Delta t_S/2 = (2\pi E/\hbar)\mathcal{A}_\phi$ and identifying $\{E, \mathcal{A}_\phi\} \leftrightarrow \{q, A_\phi\}$, where $E \equiv$ photon's energy, see Sec. 4.1 in Ref. 5.

vanishes if it has zero angular momentum, i.e., if it comoves with the zero angular momentum observers (ZAMOs) of the same radius. The angular velocity of such observers, Eq. (69) of Ref. 5, is however r-dependent; hence, considering instead a "coil" of optical loops, as depicted in Fig. 4 of Ref. 5, provides a frame-independent (thought) experiment to detect the cylinder's rotation, since it is impossible to make the effect vanish simultaneously in every loop when $j \neq 0 \Leftrightarrow \boldsymbol{\mathcal{A}} \neq 0$.

4.4.1. *Observer-independent gravitomagnetic clock effect*

Another consequence of the vanishing of $\vec{H}$ is that the gravitomagnetic clock effect in Eq. (12) reduces to the Sagnac time delay, $\Delta t_{\rm geo} = \Delta t_{\rm S} = 4\pi \mathcal{A}_\phi$; hence, all that was said above about beams in optical loops around the cylinder, applies as well to pairs of particles in oppositely rotating circular geodesics (the co-rotating geodesics having thus shorter periods). It is however actually possible to detect the cylinder's rotation using only *one pair* of particles (i.e., a pair of clocks), through the difference in the proper times (τ_+ and τ_-) measured by each of them between the events where they meet, see Fig. 1 (b). From Eqs. (13) and (11), with $U^0_\pm = (-g_{00} - 2\Omega_{{\rm geo}\pm} g_{0\phi} - \Omega^2_{{\rm geo}\pm} g_{\phi\phi})^{-1/2}$, we have

$$\Delta\tau \equiv \tau_+ - \tau_- = \frac{8\pi j r^{2\lambda_{\rm m}}}{\sqrt{\alpha(1-4\lambda_{\rm m})(1-2\lambda_{\rm m}) + 8\lambda_{\rm m} j^2 r^{8\lambda_{\rm m}-2}\alpha^{-1}(1/4-\lambda_{\rm m})^{-1}}} \quad (>0)$$

(this result is mentioned in main paper,[5] though without presenting it explicitly). Hence, when $j \neq 0$, the proper times measured by each clock differ when they meet, the co-rotating clock measuring a longer time.

4.5. *Important limits: Levi-Civita static cylinder and cosmic strings*

It is immediate to obtain important limits from the canonical form (21). Taking the limit $j \to 0$ yields the Levi-Civita metric[2–4]

$$ds^2 = -\frac{r^{4\lambda_{\rm m}}}{\alpha} dt^2 + r^{4\lambda_{\rm m}(2\lambda_{\rm m}-1)}(dr^2 + dz^2) + \alpha r^{2(1-2\lambda_{\rm m})} d\phi^2 \ . \tag{26}$$

The inertial fields $\vec{G}$ and $\vec{H} = 0$, as well as the spatial metric h_{ij}, remain the same as in (21) (the same applying to the tidal fields/forces, see Sec. 5.2.3 in Ref. 5). They differ only in the gravitomagnetic potential 1-form $\boldsymbol{\mathcal{A}} = -4j/(1-4\lambda_{\rm m})\mathbf{d}\phi$, governing global physical effects such as the Sagnac effect and synchronization gap (10) in loops around the cylinder, and the gravitomagnetic clock effect, which are all zero for the static metric (26), see Figs. 1-2.

The limit $\lambda_{\rm m} \to 0$ yields

$$ds^2 = -\frac{1}{\alpha}\left[dt + 4j d\phi\right]^2 + dr^2 + dz^2 + \alpha r^2 d\phi^2 \tag{27}$$

which is the metric of a spinning cosmic string[2,38,39] of *Komar* angular momentum per unit length j and angle deficit $2\pi(1-\alpha^{1/2}) \equiv 2\pi\delta$. In this case the spacetime

is locally flat ($R_{\alpha\beta\gamma\delta} = 0$) for $r \neq 0$. All the GEM inertial fields vanish, $\vec{G} = \vec{H} = 0$ (and the same for the tidal fields), thus there are no gravitational forces of any kind. Global gravitational effects however subsist, governed by $\boldsymbol{\mathcal{A}} = -4j\mathbf{d}\phi$ and α. The non-vanishing gravitomagnetic potential 1-form $\boldsymbol{\mathcal{A}}$ means that a Sagnac effect remains, thus the apparatuses manifesting the source's rotation discussed in Sec. 4.4 apply here as well. The same applies to the synchronization of clocks: observers at rest in the coordinates of (27) (which are in this case inertial observers) can synchronize their clocks along closed loops not enclosing the string (i.e., the axis $r = 0$), but are unable to do so for loops enclosing it. As for the gravitomagnetic clock effect, it does not apply here, as circular geodesics do not exist (since there is no gravitational attraction, $\vec{G} = 0$). The angle deficit generates double images of objects located behind the string,[40,41] and a holonomy[41,42] along closed (in spacetime or only spatially) loops around the string. Namely, vectors parallel transported along such loops turn out rotated by an angle $-2\pi\alpha^{1/2}$ (i.e, $2\pi\delta$) about the z-axis when they return to the initial position — an effect which is independent of the shape of the loop and of j; see Sec. 5.2.4 of Ref. 5. One thus can say that the metric (27) possesses two holonomies: a spatial holonomy governed by α, which is the same for spinning or non-spinning strings, plus a synchronization holonomy (Sec. 5.3.3 of Ref. 5) that arises in the spinning case.

4.6. *Summary of "canonical" features*

We argue Eq. (21) to be the most natural, or *canonical*, form for the metric of a Weyl class rotating cylinder for the following reasons:

- the Killing vector field ∂_t is (for $\alpha > 0$) everywhere time-like (i.e., $g_{00} < 0$ for all r), therefore physical observers $u^\alpha = (-g_{00})^{-1/2}\partial_t^\alpha$, at rest in the coordinates of (21), exist everywhere.
- The associated reference frame is *asymptotically* inertial, and thus fixed with respect to the "distant stars" (Sec. 4.4).
- A conserved Komar mass per unit length ($\lambda_{\rm m}$) can be defined from ∂_t which matches its expected value from the gravitational field $\vec{G}$ and potential Φ in Sec. 4.4 (see also Sec. 5.2.1 of Ref. 5), and also that of the Levi-Civita static cylinder (26).
- It is irreducibly given in terms of three parameters with a clear physical interpretation: the Komar mass ($\lambda_{\rm m}$) and angular momentum (j) per unit length, plus the parameter α governing the angle deficit of the spatial metric h_{ij}.
- The GEM fields are strikingly similar to the electromagnetic analogues — the electromagnetic fields of a rotating cylinder as measured in the inertial rest frame (namely $\boldsymbol{\mathcal{A}} = \mathcal{A}_\phi\mathbf{d}\phi$; $\mathcal{A}_\phi \equiv$constant, $\vec{H} = \mathbb{H}_{\alpha\beta} = 0$, and Φ and $G_{,i}$ match the electromagnetic counterparts identifying the Komar mass per unit length $\lambda_{\rm m}$ with the minus charge per unit length λ, cf. Sec. 4 of Ref. 5).

- It is immediate from it to obtain the two important limits: spinning cosmic string ($\lambda_{\rm m} = 0$), and Levi-Civita static solution (evincing that $j = 0$ is the necessary and *sufficient* condition).
- The GEM inertial fields and tidal tensors are the *same* as those of the Levi-Civita static cylinder (just like the electromagnetic forces produced by a charged spinning cylinder are the same as by a static one).
- It is obtained from a simple rigid rotation of coordinates, Eq. (15), which is a well-defined *global* coordinate transformation (Sec. 4.1).
- It makes immediately transparent the locally static but globally stationary nature of the metric (see Sec. 5 below).
- It has a smooth matching to the van Stockum interior solution (corresponding to a cylinder of rigidly rotating dust) *written in star-fixed coordinates* (Sec. 5.4.2 of Ref. 5).

We conclude that the Lewis metric in its usual form (1)-(2) indeed possesses a trivial coordinate rotation [of angular velocity $-\Omega$, equivalently given by either of Eqs. (14)], which has apparently gone unnoticed in the literature, and causes the artificial features listed in Sec. 3. As shown in Sec. 5.4 of Ref. 5, such rotation has a simple interpretation when the solution is matched to the van Stockum interior solution (corresponding to a rigidly rotating cylinder of dust): the coordinate system in (1)-(2) is *rigidly comoving* with the cylinder.

5. Contrast with a locally (and globally) *non-static* solution — the Kerr spacetime

Question by O. Semerák: *you were comparing the results for the (rotating) Weyl class Lewis metric with the static case; how about the comparison with Kerr, which is different because there the vorticity should contribute to the gravitomagnetic field?*

The contrast with the Kerr spacetime is indeed instructive. In what pertains to gravitomagnetism, it fundamentally differs from the Weyl class cylindrical metrics (rotating or non-rotating) in two mains aspects: it is not locally static, and its Riemann tensor is not (except at the equatorial plane) "purely electric".[29]

Staticity.— a spacetime is static[43] within some region *iff* a time-like Killing vector field ξ^α exists which is proportional to the gradient of some (single-valued) function ψ, $\xi_\alpha = \eta\partial_\alpha\psi$. Locally, this condition is equivalent to the integral lines of ξ^α having no vorticity, i.e., being hypersurface orthogonal (globally, however, the vorticity-free condition is not sufficient[43,44]). One can show (Proposition 5.1 in Ref. 5) that, in the GEM framework, local staticity amounts to the existence of a coordinate system where the metric takes the stationary form (3) with $\boldsymbol{\mathcal{A}}$ closed ($\mathbf{d}\boldsymbol{\mathcal{A}} = 0$); and global staticity to $\boldsymbol{\mathcal{A}}$ being moreover an *exact* form (in a globally well defined coordinate system).

The Weyl class Lewis metric (21) is locally static since $\mathbf{d}\boldsymbol{\mathcal{A}} = 0$; but, unless $j = 0 \Rightarrow \boldsymbol{\mathcal{A}} = 0$ (Levi-Civita static cylinder), *not globally* static, since $\boldsymbol{\mathcal{A}} = \mathcal{A}_\phi \mathbf{d}\phi$ is not an exact form. This means that the Killing vector field ∂_t is hypersurface

orthogonal but (unless $j = 0$) such hypersurfaces are not of global simultaneity, see Fig. 2 (a)-(b). In the case of the Kerr spacetime, $\mathbf{d}\boldsymbol{\mathcal{A}} \neq 0$, so it is not globally static; no hypersurface orthogonal time-like Killing vector field exists, the only Killing vector field which is time-like at infinity being ∂_t in Boyer-Lindquist coordinates, whose integral lines are well known to have vorticity. Geometrically, this means that the distribution of hyperplanes orthogonal to ∂_t (i.e., the hyperplanes of *local* simultaneity,[21] or local rest spaces of the "laboratory" observers) is not integrable, see Fig. 2 (c). On top of this, outside the equatorial plane, $R_{\alpha\beta\gamma\delta}$ is not purely electric (see Sec. V.C of Ref. 29), thus no observers exist measuring a vanishing gravitomagnetic tidal tensor $\mathbb{H}_{\alpha\beta}$.

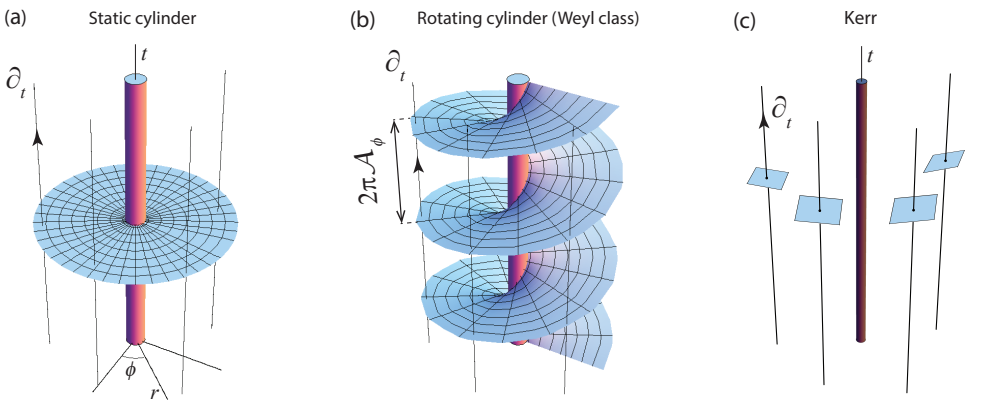


Fig. 2. (a)-(b): In the Weyl class Lewis metrics (21) the Killing vector field ∂_t is hypersurface orthogonal; such hypersurface is of global simultaneity (a plane, in a t, r, ϕ plot) for a non-spinning (Levi-Civita) cylinder, and of local but *non-global* simultaneity (the helicoid $t - \mathcal{A}_\phi \phi = const.$) in the spinning case. (c) In the Kerr spacetime ∂_t is not hypersurface orthogonal, i.e., the distribution of hyperplanes orthogonal to ∂_t (hyperplanes of *local* simultaneity) is not integrable. In (a) observers of worldlines tangent to ∂_t ("laboratory observers") are able to globally synchronize their clocks. In (b) they are unable to synchronize their clocks *around* the cylinder: each 2π turn along the helicoid leads to a different event in time, the jump between turns being the synchronization gap $\Delta t_{\text{sync}} = 2\pi\mathcal{A}_\phi$. In (c) the laboratory observers are (generically) unable to synchronize their clocks along any spatially closed loop, $\Delta t_{\text{sync}} = \oint \boldsymbol{\mathcal{A}} \neq 0$.

Physically, this means that whereas for the Weyl class spinning cylinder (21) only the first level of gravitomagnetism in Table 1 is non-zero, in the Kerr spacetime all the three levels are non-zero. Therein it is thus possible to detect the source's rotation in a more local way (i.e., not needing experiments on loops around the source): in a reference frame fixed to the distant stars, due to the non-zero gravitomagnetic field $\vec{H}$, test particles in geodesic motion will appear to be deflected by a gravitomagnetic (or Coriolis) force $\gamma\vec{U} \times \vec{H}$, cf. Eq. (4), causing e.g. their orbits to precess (Lense-Thirring precession[36]), and gyroscopes will as well be seen to precess, cf. Eq. (6). The non-vanishing $\mathbb{H}_{\alpha\beta}$ means moreover that gyroscopes at rest (or generically moving) will be acted by a force (7).

Table 1. Gravitomagnetic effects present in the Weyl class Lewis metrics and in the Kerr spacetime, as measured in star-fixed reference frames (canonical and Boyer-Lindquist coordinate systems, respectively), split by levels of gravitomagnetism, corresponding to orders of differentiation of $\boldsymbol{\mathcal{A}}$.

Level of Gravitomagnetism		Lewis-Weyl	Kerr
Governing object	Physical effect		
$\vec{\mathcal{A}}$ (gravitomagnetic vector potential)	• Sagnac effect • Synchronization gap	✓ (global effects)	✓ (global *and* local)
$\vec{H}$ (gravitomagnetic field $= e^{\phi}\nabla\times\vec{\mathcal{A}}$)	• gravitomagnetic force $m\gamma\vec{U}\times\vec{H}$ • gyroscope precession $d\vec{S}/d\tau=\vec{S}\times\vec{H}/2$ • local Sagnac effect in light gyroscope	×	✓
$\vec{H}+\vec{\mathcal{A}}$	• Gravitomagnetic "clock" effect	• co-rotating geodesic has *shorter* period	• co-rotating geodesic has *longer* period
$\mathbb{H}_{\alpha\beta}$ (gravitomag. tidal tensor $\sim\partial_i\partial_j\mathcal{A}_k$)	• Force on gyroscope $DP^{\alpha}/d\tau=-\mathbb{H}^{\beta\alpha}S_{\beta}$	×	✓

Even in what pertains to the first level of gravitomagnetism (governed by $\boldsymbol{\mathcal{A}}$), present in both, they substantially differ. The fact that $\mathbf{d}\boldsymbol{\mathcal{A}}=0$ in the Lewis-Weyl metric means that a Sagnac effect (9) arises only on loops enclosing the cylinder (as discussed in Sec. 4.4), and is independent of the shape of the loop; and similarly for the synchronization of clocks: the laboratory observers are able to synchronize their clocks along spatially closed loops that do not enclose the cylinder [in other words, closed in spacetime synchronization curves exist along the helicoid of Fig. 2 (b)]; it is only on loops around the cylinder that a synchronization gap (10) arises, see Fig. 2 (b). In the Kerr spacetime, by contrast, since $\mathbf{d}\boldsymbol{\mathcal{A}}\neq 0$, the Sagnac effect depends on the shape of the loop, and is generically non-zero (regardless of the loop enclosing or not the axis). The laboratory observers are likewise unable to synchronize their clocks around generic closed loops.

Another interesting contrast is in the gravitomagnetic clock effect (12). Around the spinning cylinder (21), since $\vec{H}=0$, it reduces to the Sagnac time-delay (25), and thus the co-rotating geodesic has a shorter period. In the case of the Kerr spacetime, by contrast, the term Δt_H of (12) is not zero,

$$\Delta t_{\rm geo}=\Delta t_{\rm S}+\Delta t_H=4\pi\frac{J}{M};\quad \Delta t_{\rm S}=-\frac{8\pi J}{r-2M}\,(<0);\quad \Delta t_H=\frac{4\pi Jr}{M(r-2M)}\,(>0),$$

and is actually dominant,[27] so it is the other way around: the co-rotating orbit has a longer period, $\Delta t_{\rm geo}=4\pi J/M>0$. The physical interpretation of $\Delta t_H>0$ is that the gravitomagnetic force $\gamma\vec{U}\times\vec{H}$ in Eq. (4) is repulsive (attractive) for co- (counter-) rotating geodesics, see Fig. 1(b) of Ref. 27.

References

1. T. Lewis, Some special solutions to the equations of axially symmetric gravitational fields, *Proc. Roy. Soc. Lond.* **A136**, 179 (1932).
2. M. F. A. da Silva, L. Herrera, F. M. Paiva and N. O. Santos, The parameters of the Lewis metric for the Weyl class, *General Relativity and Gravitation* **27**, 859 (Aug 1995).
3. M. F. A. da Silva, L. Herrera, F. M. Paiva and N. O. Santos, On the parameters of Lewis metric for the Lewis class, *Class. Quant. Grav.* **12**, 111 (1995).
4. J. B. Griffiths and J. Podolsky, *Exact Space-Times in Einstein's General Relativity*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 2009).
5. L. F. O. Costa, J. Natário and N. O. Santos, Gravitomagnetism in the Lewis cylindrical metrics, *Class. Quant. Grav.* **38**, p. 055003 (2021).
6. W. J. van Stockum, IX. The Gravitational Field of a Distribution of Particles Rotating about an Axis of Symmetry, *Proceedings of the Royal Society of Edinburgh* **57**, 135 (1938).
7. Gravitomagnetism in the Lewis cylindrical metrics, *Invited talk at the parallel session "Dragging is never draggy: MAss and CHarge flows in GR", sixteenth Marcel Grossmann Meeting (MG16), July 5-10 2021* (https://youtu.be/ho31IgLNxu8).
8. L. D. Landau and E. M. Lifshitz, *The classical theory of fields; 4rd ed.*, Course of theoretical physics, Vol. 2 (Butterworth-Heinemann, Oxford, UK, 1975), Trans. from the Russian.
9. D. Lynden-Bell and M. Nouri-Zonoz, Classical monopoles: Newton, NUT space, gravomagnetic lensing, and atomic spectra, *Rev. Mod. Phys.* **70**, 427 (Apr 1998).
10. R. T. Jantzen, P. Carini and D. Bini, The Many faces of gravitoelectromagnetism, *Annals Phys.* **215**, 1 (1992).
11. L. F. O. Costa and J. Natário, Gravito-electromagnetic analogies, *General Relativity and Gravitation* **46**, p. 1792 (2014).
12. M. Mathisson, Neue mechanik materieller systemes, *Acta Phys. Polon.* **6**, p. 163 (1937), [Gen. Relativ. Gravit. **42**, 1011 (2010)].
13. A. Papapetrou, Spinning test particles in general relativity. 1., *Proc. Roy. Soc. Lond.* **A209**, 248 (1951).
14. L. F. O. Costa, J. Natário and M. Zilhão, Spacetime dynamics of spinning particles: Exact electromagnetic analogies, *Phys. Rev. D* **93**, p. 104006 (2016).
15. W. G. Dixon, A covariant multipole formalism for extended test bodies in general relativity, *Il Nuovo Cimento* **34**, p. 317 (1964).
16. Y. Aharonov and D. Bohm, Significance of electromagnetic potentials in the quantum theory, *Phys. Rev.* **115**, 485 (Aug 1959).
17. E. J. Post, Sagnac effect, *Rev. Mod. Phys.* **39**, 475 (Apr 1967).
18. E. Kajari, M. Buser, C. Feiler and W. P. Schleich, Rotation in relativity and the propagation of light, in *Proceedings of the International School of Physics "Enrico Fermi", Course CLXVIII*, (IOS Press, Amsterdam, 2009).
19. A. Ashtekar and A. Magnon, The Sagnac effect in general relativity, *J. Math. Phys.* **16**, 341 (1975).
20. A. Tartaglia, General relativistic corrections to the Sagnac effect, *Phys. Rev. D* **58**, p. 064009 (1998).
21. E. Minguzzi, Simultaneity and generalized connections in general relativity, *Class. Quant. Grav.* **20**, 2443 (2003).
22. M. L. Ruggiero, Sagnac Effect, Ring Lasers and Terrestrial Tests of Gravity, *Galaxies* **3**, 84 (2015).

23. J. Cohen and B. Mashhoon, Standard clocks, interferometry, and gravitomagnetism, *Physics Letters A* **181**, 353 (1993).
24. W. B. Bonnor and B. R. Steadman, The gravitomagnetic clock effect, *Classical and Quantum Gravity* **16**, p. 1853 (1999).
25. D. Bini, R. T. Jantzen and B. Mashhoon, Gravitomagnetism and relative observer clock effects, *Class. Quant. Grav.* **18**, 653 (2001).
26. L. Iorio, H. I. M. Lichtenegger and B. Mashhoon, An Alternative derivation of the gravitomagnetic clock effect, *Class. Quant. Grav.* **19**, 39 (2002).
27. L. F. O. Costa and J. Natário, Frame-Dragging: Meaning, Myths, and Misconceptions, *Universe* **7**, p. 388 (2021).
28. A. Tartaglia, Detection of the gravitomagnetic clock effect, *Class. Quant. Grav.* **17**, 783 (2000).
29. L. F. O. Costa, L. Wylleman and J. Natário, Gravitomagnetism and the significance of the curvature scalar invariants, *Phys. Rev. D* **104**, p. 084081 (2021).
30. C. B. G. McIntosh, R. Arianrhod, S. T. Wade and C. Hoenselaers, Electric and magnetic Weyl tensors: Classification and analysis, *Classical and Quantum Gravity* **11**, 1555 (jun 1994).
31. M. A. H. MacCallum and N. O. Santos, Stationary and static cylindrically symmetric Einstein spaces of the Lewis form, *Classical and Quantum Gravity* **15**, p. 1627 (1998).
32. W. B. Bonnor, Solution of Einstein's equations for a line-mass of perfect fluid, *Journal of Physics A: Mathematical and General* **12**, 847 (1979).
33. W. Israel, Line sources in general relativity, *Phys. Rev. D* **15**, 935 (Feb 1977).
34. L. Marder, Gravitational Waves in General Relativity. I. Cylindrical Waves, *Proceedings of the Royal Society of London Series A* **244**, 524 (1958).
35. D. Lynden-Bell and J. Bičák, Komar fluxes of circularly polarized light beams and cylindrical metrics, *Phys. Rev. D* **96**, p. 104053 (2017).
36. I. Ciufolini and J. A. Wheeler, *Gravitation and Inertia* (Princeton Series in Physics, Princeton, NJ, 1995).
37. I. Ciufolini, Dragging of inertial frames, *Nature* **449**, 41 (2007).
38. A. Barros, V. B. Bezerra and C. Romero, Global aspects of gravitomagnetism, *Mod. Phys. Lett.* **A18**, 2673 (2003).
39. B. Jensen and H. H. Soleng, General-relativistic model of a spinning cosmic string, *Phys. Rev. D* **45**, 3528 (May 1992).
40. M. Hindmarsh and T. Kibble, Cosmic strings, *Rept. Prog. Phys.* **58**, 477 (1995).
41. L. H. Ford and A. Vilenkin, A gravitational analogue of the Aharonov-Bohm effect, *Journal of Physics A: Mathematical and General* **14**, 2353 (1981).
42. M. Nouri-Zonoz and A. Parvizi, Gaussian curvature and global effects: Gravitational Aharonov-Bohm effect revisited, *Phys. Rev. D* **88**, p. 023004 (2013).
43. J. Stachel, Globally stationary but locally static space-times: A gravitational analog of the Aharonov-Bohm effect, *Phys. Rev. D* **26**, 1281 (1982).
44. W. B. Bonnor, The rigidly rotating relativistic dust cylinder, *Journal of Physics A: Mathematical and General* **13**, p. 2121 (1980).

Magnetized black holes: The role of rotation, boost, and accretion in twisting the field lines and accelerating particles*

Ondřej Kopáček and Vladimír Karas

Astronomical Institute, Czech Academy of Sciences,
Boční II 1401, CZ-141 00 Prague, Czech Republic
E-mail: kopacek@ig.cas.cz, vladimir.karas@asu.cas.cz
http://astro.cas.cz

Combined influence of rotation of a black hole and ambient magnetic fields creates conditions for powerful astrophysical processes of accretion and outflow of matter which are observed in many systems across the range of masses; from stellar-mass black holes in binary systems to supermassive black holes in active galactic nuclei. We study a simplified model of outflow of electrically charged particles from the inner region of an accretion disk around a spinning (Kerr) black hole immersed in a large-scale magnetic field. In particular, we consider a non-axisymmetric magnetosphere where the field is misaligned with the rotation axis. In this contribution we extend our previous analysis of acceleration of jet-like trajectories of particles escaping from bound circular orbits around a black hole. While we have previously assumed the initial setup of prograde (co-rotating) orbits, here we relax this assumption and we also consider retrograde (counter-rotating) motion. We show that the effect of *counter-rotation may considerably increase the probability of escape from the system, and it allows more efficient acceleration of escaping particles to slightly higher energies compared to the co-rotating disk.*

Keywords: Black holes, magnetosphere, non-axisymmetry, dynamics of ionized matter, acceleration of particles

1. Introduction

In this contribution we further study astrophysically relevant class of escaping (jet-like accelerated) trajectories in a dynamical system of a magnetized rotating black hole. Electrically charged particles following such orbits may escape the attraction of the black hole in a collimated outflow and they may be accelerated to very high energies. In Paper I[1] we adopted a simple axisymmetric model consisting of a rotating black hole in an asymptotically uniform magnetic field aligned with the rotation axis. Initially neutral particles are supposed to follow circular orbits in the equatorial plane (and freely fall below ISCO). At a particular ionization radius, these particles obtain electric charge and their dynamics changes due to the presence of electromagnetic field, and escape from the accretion disk along high-velocity jet-like trajectories becomes possible. However, the efficiency of the acceleration process remains only moderate in the axisymmetric setup.[1]

*Based on an invited contribution at 16th Marcel Grossmann Meeting, Session PT5 "Dragging is never draggy: MAss and CHarge flows in GR" (id. #393), 5–10 July 2021.

In Paper II[2] we relax the assumption of the perfect spin-magnetic alignment and consider a generalized non-axisymmetric model. The effect of non-axisymmetry in electro-vacuum magnetospheres of compact objects is known to introduce dynamically relevant features like magnetic null-points,[3] and breaking the axial symmetry also appears to strongly affect the acceleration process in the current model. In particular, the number of escaping orbits generally increases and acceleration to ultrarelativistic energies becomes possible in the oblique magnetosphere.[2,4] Furthermore,[5] by employing a boost transformation of the system into a frame in translation motion, we explored the electromagnetic aspects of the linear (kick) velocity of the black hole. Refer to enlightening discussions by various authors[6,7] and see also the relevant context of these solutions.[8]

In the above-mentioned previous papers, we numerically studied escaping orbits and we assessed in detail the role of a broad variety of parameters of the model. Nevertheless, the assumption that initially neutral matter co-rotates with the central black hole, was adopted in all previous analyses. Here we relax this assumption and we also consider counter-rotating (retrograde) circular orbits as an initial setup of neutral particles. We compare the both cases (of co-rotating vs. counter-rotating) in terms of the efficiency of the formation and acceleration of the individual trajectories and the emerging outflow.

The role of counter-rotation in processes of accretion and outflow near compact objects has been only partially explored in the literature and various relevant aspects of accretion of counter-rotating matter has been studied. In particular, high-resolution hydrodynamic simulations suggest that counter-rotating structure may largely increase the accretion rate.[9] Moreover, general relativistic magnetohydrodynamic (GRMHD) simulations show that accretion of counter-rotating matter may produce conditions for highly efficient jet launching.[10] We also note that counter-rotating high-velocity outflows are indeed reported in various systems, such as the spiral galaxy NGC 1068.[11]

The paper is organized as follows. In Section 2 we specify the employed model of oblique black hole magnetosphere, we review the equations of motion and describe the charging process of initially neutral matter. Jet-like trajectories of charged particles escaping from counter-rotating orbits are discussed in Section 3 with the focus on the direct comparison with the co-rotating analogues. Results are summarized and briefly discussed in Section 4. As the initial setup of the model investigated in the present paper is analogous to that of Papers I and II, we refer to the more complete introduction and references presented therein.

2. Model specification

We consider a spacetime around a rotating, axially-symmetric, uncharged black hole of mass M and spin a described by the Kerr metric, which may be expressed in Boyer–Lindquist coordinates $x^\mu = (t,\, r,\, \theta,\, \varphi)$ as follows:[12,13]

$$ds^2 = -\frac{\Delta}{\Sigma}\,[dt - a\sin\theta\, d\varphi]^2 + \frac{\sin^2\theta}{\Sigma}\,[(r^2+a^2)d\varphi - a\, dt]^2 + \frac{\Sigma}{\Delta}\, dr^2 + \Sigma d\theta^2, \quad (1)$$

where $\Delta(r) \equiv r^2 - 2Mr + a^2$ and $\Sigma(r,\theta) \equiv r^2 + a^2\cos^2\theta$. Roots of $\Delta(r)$ locate the outer (+) and the inner (-) horizons as: $r_\pm = M \pm \sqrt{M^2 - a^2}$. Geometrized units are used throughout the paper; values of basic constants (gravitational constant G, speed of light c, Boltzmann constant k, and Coulomb constant k_C) therefore equal unity, $G = c = k = k_C = 1$.

The black hole is supposed to be weakly magnetized by an external, asymptotically uniform magnetic field with an asymptotic strength B which has an arbitrary inclination (angle α) with respect to the rotation axis. Vector potential $A_\mu = (A_t, A_r, A_\theta, A_\varphi)$ of a corresponding test-field solution is given as:[14]

$$A_t = \frac{B_z a M r}{\Sigma}\left(1 + \cos^2\theta\right) - B_z a + \frac{B_x a M \sin\theta\cos\theta}{\Sigma}\left(r\cos\psi - a\sin\psi\right), \quad (2)$$

$$A_r = -B_x(r - M)\cos\theta\sin\theta\sin\psi, \quad (3)$$

$$A_\theta = -B_x a(r\sin^2\theta + M\cos^2\theta)\cos\psi - B_x(r^2\cos^2\theta - Mr\cos 2\theta + a^2\cos 2\theta)\sin\psi, \quad (4)$$

$$A_\varphi = B_z\sin^2\theta\left[\frac{1}{2}(r^2 + a^2) - \frac{a^2 M r}{\Sigma}(1 + \cos^2\theta)\right] - B_x\sin\theta\cos\theta\left[\Delta\cos\psi + \frac{(r^2 + a^2)M}{\Sigma}\left(r\cos\psi - a\sin\psi\right)\right], \quad (5)$$

where B_z denotes the component parallel to the rotation axis, while B_x corresponds to the perpendicular component, i.e., $B_z = B\cos\alpha$ and $B_x = B\sin\alpha$. Setting $B_x = 0$ reduces the above vector potential A_μ to the axisymmetric solution[15] employed in Paper I. Azimuthal coordinate ψ of Kerr ingoing coordinates is expressed in Boyer–Lindquist coordinates as follows:

$$\psi = \varphi + \frac{a}{r_+ - r_-}\ln\frac{r - r_+}{r - r_-}. \quad (6)$$

The Hamiltonian $\mathcal{H}$ of a particle of electric charge q and rest mass m in the field A_μ and metric $g^{\mu\nu}$ is defined as:[12]

$$\mathcal{H} = \tfrac{1}{2}g^{\mu\nu}(\pi_\mu - qA_\mu)(\pi_\nu - qA_\nu), \quad (7)$$

where π_μ is the generalized (canonical) momentum. The equations of motion are expressed as:

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \equiv p^\mu = \frac{\partial\mathcal{H}}{\partial\pi_\mu}, \quad \frac{d\pi_\mu}{d\lambda} = -\frac{\partial\mathcal{H}}{\partial x^\mu}, \quad (8)$$

where $\lambda \equiv \tau/m$ is dimensionless affine parameter (τ denotes the proper time). Employing the first equation we obtain the kinematical four-momentum as: $p^\mu = \pi^\mu - qA^\mu$, and the conserved value of the Hamiltonian is therefore given as: $\mathcal{H} = -m^2/2$. System is stationary and the time component of canonical momentum π_t is therefore an integral of motion which equals (negatively taken) energy of the test particle $\pi_t \equiv -E$. In the rest of paper we switch to specific quantities $E/m \to E$,

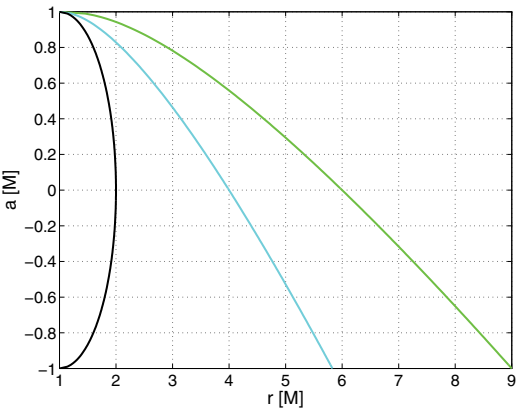


Fig. 1. Position of the black hole's outer horizon (black), marginally bound circular orbit (cyan) and ISCO (green). Negative spin values correspond to the counter-rotating orbits.

$q/m \to q$ which corresponds to setting the rest mass of the particle $m = 1$ in the formulas.

As an initial configuration of electrically neutral particles, we consider circular Keplerian orbits specified by the values of constants of motion, i.e., by specific energy $E_{\rm Kep}$ and specific angular momentum $L_{\rm Kep}$ given as follows:[16]

$$E_{\rm Kep} = \frac{r^2 - 2Mr \pm a\sqrt{Mr}}{r\sqrt{r^2 - 3Mr \pm 2a\sqrt{Mr}}}, \quad L_{\rm Kep} = \frac{\pm\sqrt{M}(r^2 + a^2 \mp 2a\sqrt{Mr})}{\sqrt{r(r^2 - 3Mr \pm 2a\sqrt{Mr})}}, \tag{9}$$

where the upper signs are valid for the prograde (co-rotating) orbits and the lower ones for the retrograde (counter-rotating) orbits. Circular geodesics remain stable only above the innermost stable circular orbit (ISCO):

$$r_{\rm ISCO} = M\left(3 + Z_2 \mp \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)}\right), \tag{10}$$

where $Z_1 \equiv 1 + \left(1 - \frac{a^2}{M^2}\right)^{1/3}\left[\left(1 + \frac{a}{M}\right)^{1/3} + \left(1 - \frac{a}{M}\right)^{1/3}\right]$ and $Z_2 \equiv \sqrt{\frac{3a^2}{M^2} + Z_1^2}$. Positions of the ISCO and marginally bound circular orbit[16] are plotted in Fig. 1.

Below ISCO the geodesics are supposed to turn into freely falling inspirals maintaining the energy and angular momentum corresponding to the ISCO radius, i.e., the particles keep $E = E_{\rm Kep}(r_{\rm ISCO})$ and $L = L_{\rm Kep}(r_{\rm ISCO})$ during their infall while the radial velocity u^r is obtained from the normalization condition $u^\mu u_\mu = -1$.

In our model, we suppose that initially neutral elements undergo a sudden charging process (e.g., due to photoionization) at a given radius r_0 and obtain specific electric charge q. While the change of the rest mass m is considered as negligible, the particle dynamics changes due to the presence of the electromagnetic field.[17]

In particular, the conserved value of energy is modified as:

$$E = E_{\rm Kep} - qA_t, \tag{11}$$

while the values of spatial components of canonical momentum are changed as:

$$\pi_r = \pi_r^0 + qA_r, \quad \pi_\theta = qA_\theta, \quad \pi_\varphi = L_{\rm Kep} + qA_\varphi, \tag{12}$$

where π_r^0 is zero for particles ionized above/at ISCO and for infalling particles with ionization radius $r_0 < r_{\rm ISCO}$ the value is calculated from the normalization condition.

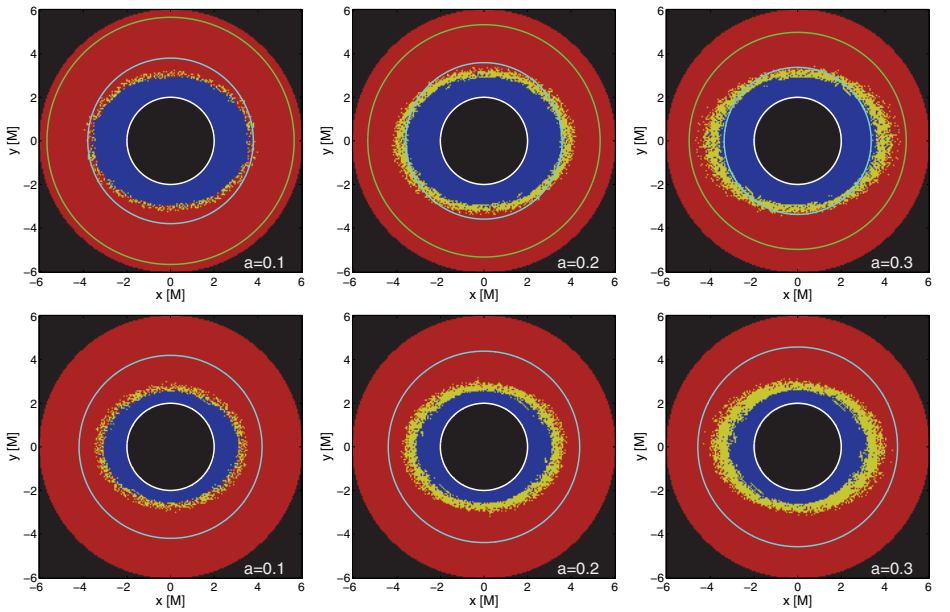


Fig. 2. Comparison of initially co-rotating orbits (upper row) with counter-rotating orbits (lower row). Color-coding: blue for plunging orbits, red for stable ones, and yellow for escaping trajectories. The green circle denotes the ISCO, the cyan circle marks the radius of the marginally bound orbit and the white one shows the ergosphere. The inner black region marks the horizon of the black hole. The parameters of the system are $\alpha = 35^\circ$ and $qB = -5\,M^{-1}$. The magnetic field is inclined in the positive x-direction.

3. Escape zones

The dynamics of initially neutral particles on stable circular orbits (above ISCO) or on a plunging trajectory (below ISCO) changes abruptly as they obtain the electric charge and become affected by the electromagnetic field given by Eqs. (2)-(5). As a result, some stable particles may become plunging, while some plunging trajectories are stabilized by the field. Moreover, it appears that particles may also become unbound and escape to infinity in jet-like trajectories.

We have previously analyzed escaping particles in this model and, in particular, in Paper I[1] we considered the axisymmetric system ($B_x = 0$), while in the Paper II[2] the general non-axisymmetric configuration was studied. We have used the method of effective potential to find the necessary conditions for the escape of particles. It appears that for the parallel orientation of the spin axis z and magnetic field component B_z, only negatively charged particles may escape (while positively charged escape for the anti-parallel orientation). Non-zero spin a and B_z are required for the escape and final Lorentz factor γ is found to be an increasing function of the spin, specific charge $|q|$ and $|B_z|$ while it decreases with the ionization radius r_0.

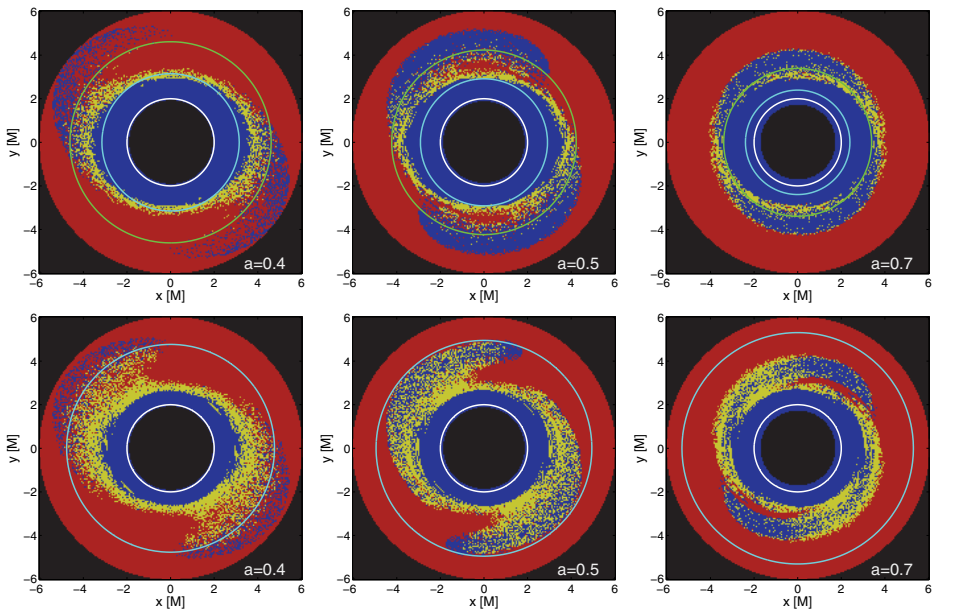


Fig. 3. Comparison of initially co-rotating orbits (upper row) with counter-rotating orbits (lower row). The color-coding and parameters as in Fig. 2.

However, as the above conditions are necessary but not sufficient, we need to investigate the dynamic system numerically in order to determine which particles actually follow the escaping trajectories. Initial locations of escaping particles form *escape zones* whose emergence and evolution with respect to the parameters of the system was discussed in our previous papers.[1,2,4] Efficiency of the acceleration mechanism was quantified by the final Lorentz factor of escaping particles leading to the conclusion that also ultrarelativistic velocities with very high energies may be achieved within the non-axisymmetric model.

In this contribution, we extend the previous discussion and consider a modified setup. In particular, here we suppose that the initially neutral matter may follow counter-rotating geodesics, i.e., unlike the previous papers, here we also consider

the lower signs in Eqs. (9) and (10). The ISCO of counter-rotating orbits is located generally farther from the horizon compared to the co-rotating ones (see Fig. 1). In particular, no stable retrograde circular orbits around rotating black hole are allowed for $r \leq 6\,M$ and the ISCO location grows with the spin up to $r_{\rm ISCO} = 9\,M$ for the maximally rotating black hole with $a = M$.

We assume that below ISCO the initially neutral particle falls freely towards the horizon keeping the energy and angular momentum corresponding to the ISCO radius, while the radial velocity is computed from the normalization condition. At the ionization radius r_0 the particle obtains the electric charge and its dynamics changes due to magnetic field. Three main scenarios may occur then; the particle plunges into the horizon, follows the stable orbit around the black hole or escapes to infinity.

Analogously to our previous studies, we adopt the visual representation of resulting orbits in escape boundary plots in which we assign blue dots to plunging orbits, red dots to stable orbits and yellow dots to escaping ones. In Figs. 2-4 we compare the co-rotating setup (upper rows) with the counter-rotating case (bottom rows) in a series of escape boundary plots for a particular set of parameters ($qB = -5\,M^{-1}$ and $\alpha = 35^\circ$) with varying values of spin. In Fig. 2 we observe that for low spins the both cases do not differ significantly. Corresponding escape zones have comparable shapes and sizes regardless the sense of revolution. However, if we consider moderate values of spin, the situation changes and significant differences between the escape zones in co-rotating and counter-rotating disks arise as shown in Fig. 3. Most importantly, we observe that the number of escaping trajectories is substantially higher for counter-rotating orbits forming considerably larger escape zones. The tendency of more rapid rotation to favorize the escape from counter-rotating orbits is confirmed in Fig. 4, where we notice that high spin allows only narrow escape zones in the co-rotating disk while counter-rotating zones become significantly wider. Also, these plots exhibit clear signatures of a transition from the regular to chaotic system in agreement with previous studies of similarly perturbed black hole systems.[2, 18–20]

Another convenient way to visualize the escape zones is to fix the initial value of the azimuthal angle φ_0 and plot the trajectories in the $r_0 \times a$ plane (ionization radius vs. spin) using the same color-coding of trajectories as in Figs. 2-4. The angle φ is measured anticlockwise from the positive direction of x-axis (which is the direction of the field inclination). In particular, we choose the value $\varphi_0 = 135^\circ$ for which the structure of the escape zone seems especially complex. In Fig. 5 we present the resulting plots for several values of magnetization parameter qB and compare co-rotating escape zones (left column) with the counter-rotating zones (right column). It confirms that counter-rotating orbits are generally more prone to escape regardless the value of qB while the both cases differ most significantly for rapidly rotating black holes.

Previous analysis in Papers I and II shows that spin parameter is an important factor for the acceleration of escaping particles. In particular, it appears that final

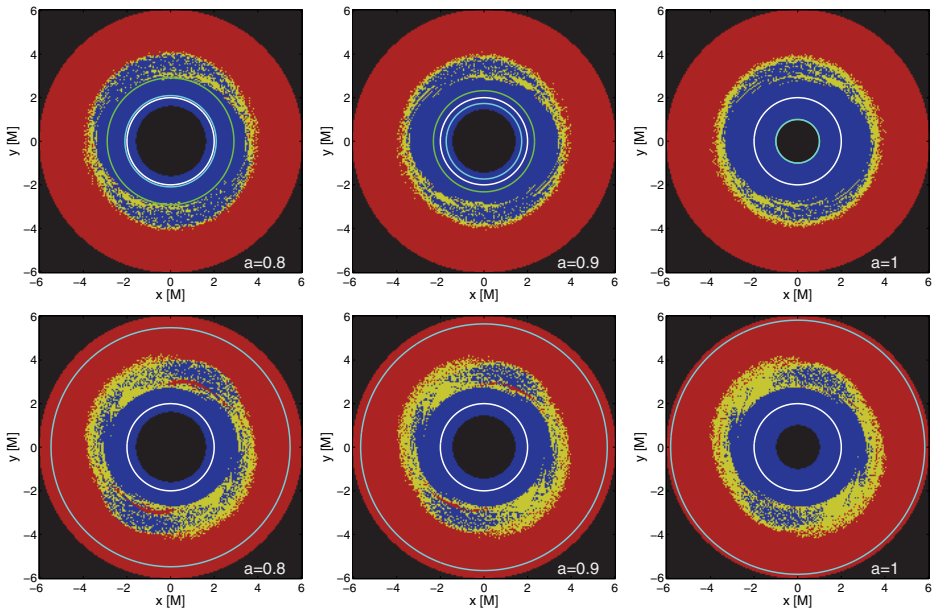


Fig. 4. Comparison of initially co-rotating orbits (upper row) with counter-rotating orbits (lower row). The color-coding and parameters as in Fig. 2.

value of γ for the fixed qB grows with the spin and decreases with the ionization radius r_0. Maximally accelerated particles are thus expected to originate from the inner edge of the escape zone near maximally rotating black hole with spin $a = M$.

In Fig. 5 we observe that counter-rotating escape zones for high spins are considerably wider compared to co-rotating ones and, moreover, we notice that counter-rotating zones are located closer to the horizon. We thus expect that particles escaping from counter-rotating orbits could reach slightly higher energies. This is indeed the case as shown in Fig. 6 where the final values of the Lorentz factor are encoded with the color-scale. For this particular set of parameters ($qB = -20\,M^{-1}$, $\varphi_0 = 135°$) we find that maximally accelerated counter-rotating particle reaches $\gamma_{\max} = 8.1$ being launched from $r_0 = 2.3\,M$ while for the co-rotating disk we find $\gamma_{\max} = 7.3$ for the particle escaping from $r_0 = 2.5\,M$.

4. Conclusions

We have numerically analyzed a simplified model of the outflow of electrically charged particles from the inner region of the accretion disk near a weakly magnetized rotating black hole. Stable circular orbits (freely falling below ISCO) of initially neutral particles were perturbed by the ionization process occuring in the non-axisymmetric magnetosphere. As a result, the escaping jet-like trajectories may be realized for some range of parameters and acceleration to high energies becomes

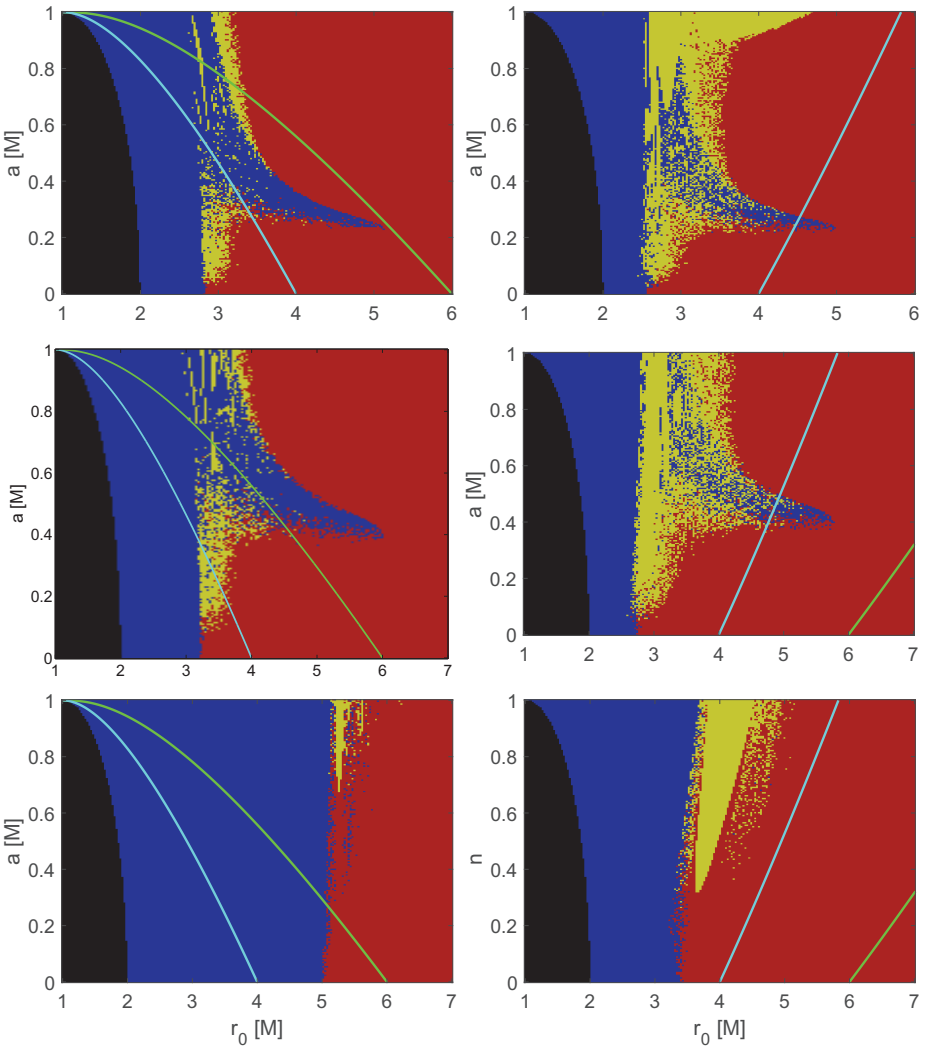


Fig. 5. Comparison of escape zones in co-rotating (left panels) and counter-rotating disks (right panels) for several values of magnetization parameter: $qB = -M^{-1}$ (bottom panels), $qB = -5\,M^{-1}$ (middle panels) and $qB = -10\,M^{-1}$ (upper panels). The inclination of the field is $\alpha = 35^\circ$ and $\varphi_0 = 135^\circ$. The color-coding as in Fig. 2.

possible. While we have previously studied various aspects of given model and investigated the formation of escape zones in detail,[1,2] we have assumed the initial setup of co-rotating accretion disk. In particular, the neutral particles above ISCO were supposed to follow prograde circular orbits and freely fall below ISCO.

In this contribution we have generalized the previous analysis by considering also counter-rotating (retrograde) Keplerian orbits as an initial setup of neutral

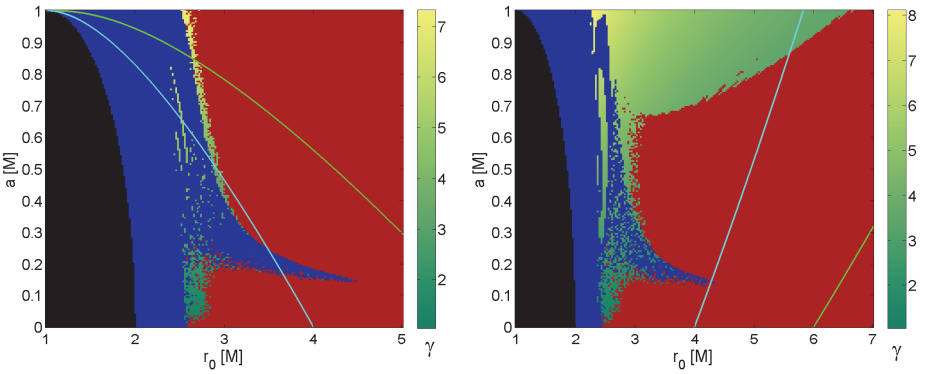


Fig. 6. Final Lorentz factor γ of escaping particles is encoded with the colorscale comparing the co-rotating (left) and the counter-rotating disks (right) while other parameters remain fixed as: $\alpha = 35°$ and $qB = -20\,M^{-1}$.

matter. We have studied the escape zones emerging in the counter-rotating disks which were directly compared with the corresponding co-rotating versions. Based on the analyzed set of escape zones we may conclude that counter-rotation tends to increase the efficiency of the outflow. While its effect becomes negligible for low spins, for moderate to high spin values we observe that counter-rotating disks allow the formation of significantly wider escape zones compared to their co-rotating analogues. The position of the zone is also affected as counter-rotating zones are shifted towards the horizon which makes the acceleration process more efficient. Although the difference is rather small, we observe that counter-rotating particles may be accelerated to higher energies.

Acknowledgements

Authors acknowledge the support from the Inter-Excellence mobility program in science of the Czech Ministry of Education, Youth and Sports (projects ref. 8JCH 1080 and LTC 18058). OK has been supported by the project "Lumina Quaeruntur" (ref. LQ100032102) of the Czech Academy of Sciences. VK thanks the Czech Science Foundation grant "Mass and charge currents in general relativity and astrophysics" (ref. 21-11268S).

References

1. O. Kopáček and V. Karas, Near-horizon Structure of Escape Zones of Electrically Charged Particles around Weakly Magnetized Rotating Black Hole, *The Astrophysical Journal* **853**, p. 53 (January 2018).
2. O. Kopáček and V. Karas, Near-horizon Structure of Escape Zones of Electrically Charged Particles around Weakly Magnetized Rotating Black Hole. II. Acceleration and Escape in the Oblique Magnetosphere, *The Astrophysical Journal* **900**, p. 119 (September 2020).

3. O. Kopáček, T. Tahamtan and V. Karas, Null points in the magnetosphere of a plunging neutron star, *Physical Review D* **98**, p. 084055 (October 2018).
4. V. Karas and O. Kopáček, Near horizon structure of escape zones of electrically charged particles around weakly magnetized rotating black hole: Case of oblique magnetosphere, *Astronomische Nachrichten* **342**, 357 (January 2021).
5. V. Karas, O. Kopáček and D. Kunneriath, Influence of frame-dragging on magnetic null points near rotating black holes, *Classical and Quantum Gravity* **29**, p. 035010 (February 2012).
6. V. Karas, O. Kopáček and D. Kunneriath, Magnetic Neutral Points and Electric Lines of Force in Strong Gravity of a Rotating Black Hole, *International Journal of Astronomy and Astrophysics* **3**, 18 (January 2013).
7. A. Abdujabbarov, N. Dadhich and B. Ahmedov, Electromagnetic field around boosted rotating black hole, *arXiv:1810.08066* (October 2018).
8. E. Gallo and T. Mädler, Comment on "Boosted Kerr black holes in general relativity", *Physical Review D* **101**, p. 028501 (January 2020).
9. S. Dyda, R. V. E. Lovelace, G. V. Ustyugova, M. M. Romanova and A. V. Koldoba, Counter-rotating accretion discs, *Monthly Notices of the Royal Astronomical Society* **446**, 613 (January 2015).
10. S. Koide, D. L. Meier, K. Shibata and T. Kudoh, General Relativistic Simulations of Early Jet Formation in a Rapidly Rotating Black Hole Magnetosphere, *The Astrophysical Journal* **536**, 668 (June 2000).
11. C. M. V. Impellizzeri, J. F. Gallimore, S. A. Baum, M. Elitzur, R. Davies, D. Lutz, R. Maiolino, A. Marconi, R. Nikutta, C. P. O'Dea and E. Sani, Counter-rotation and High-velocity Outflow in the Parsec-scale Molecular Torus of NGC 1068, *The Astrophysical Journal Letters* **884**, p. L28 (October 2019).
12. C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* 2017.
13. R. P. Kerr, Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics, *Physical Review Letters* **11**, 237 (September 1963).
14. J. Bičák and V. Janiš, Magnetic fluxes across black holes, *Monthly Notices of the Royal Astronomical Society* **212**, 899 (February 1985).
15. R. M. Wald, Black hole in a uniform magnetic field, *Physical Review D* **10**, 1680 (September 1974).
16. J. M. Bardeen, W. H. Press and S. A. Teukolsky, Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation, *The Astrophysical Journal* **178**, 347 (December 1972).
17. Z. Stuchlík and M. Kološ, Acceleration of the charged particles due to chaotic scattering in the combined black hole gravitational field and asymptotically uniform magnetic field, *European Physical Journal C* **76**, p. 32 (January 2016).
18. V. Karas and D. Vokrouhlický, Chaotic motion of test particles in the Ernst spacetime, *General Relativity and Gravitation* **24**, 729 (July 1992).
19. P. S. Letelier and W. M. Vieira, Chaos in black holes surrounded by gravitational waves, *Classical and Quantum Gravity* **14**, 1249 (May 1997).
20. V. Witzany, O. Semerák and P. Suková, Free motion around black holes with discs or rings: Between integrability and chaos - IV, *Monthly Notices of the Royal Astronomical Society* **451**, 1770 (August 2015).

Spinning particle: Is Newton-Wigner the only way?

V. Witzany
School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4, Ireland
E-mail: vojtech.witzany at ucd.ie

A rapidly spinning compact object couples to an ambient curved background via the so-called spin-curvature coupling. In expressing this, one has to deal with the ambiguity of the definition of the center of mass of the body. What is worse, in a Hamiltonian formalism, this choice corresponds to an unphysical "parasitic" degree of freedom in the dynamical system. A solution to this is to apply a Hamiltonian constraint on the system and to obtain a set of brackets where the center-of-mass degree of freedom is erased from the algebra. I report on my progress in this procedure in the case of the so-called Tulczyjew-Dixon (or "covariant") supplementary spin condition and in my effort to cover the resulting phase space with canonical coordinates.

Keywords: Equations of motion – Classical relativity – Spinning bodies – Two-body problem – Gravitational waves

1. Introduction: Motion of spinning particle

Notation and convention: I use the $G = c = 1$ geometrized units and a $(-+++)$ metric signature. Greek indices $\mu, \nu, ...$ run from 0 to 3, and Einstein summation convention is assumed.

When the surface of a given body is rotating at relativistic speeds, it will couple to any curved space-time background by the so-called *spin-curvature coupling*. This effect is an important ingredient for the relativistic two-body problem and, in particular, for the modelling of gravitational waves from compact binaries.

To understand how the spin-curvature coupling arises, imagine a compact object such as a neutron star moving in an ambient curved space-time background where, in order to treat the problem step by step, we will for now understand the compact object exclusively as a "test" object. When the variability length, or the curvature scale of the background is much longer than the size of the body, it is meaningful to construct a set of Riemann normal coordinates X^μ with respect to the background geometry centred on the body. The coordinate acceleration of individual elements of the body caused by the background gravitational field then is $\sim \Gamma^\mu{}_{\nu\kappa} U^\nu U^\kappa$, where $\Gamma^\mu{}_{\nu\kappa}$ are Christoffel symbols of the background and U^ν the four-velocity of the body elements, both expressed in the local frame.

Let us now assume that the body is approximately rigidly rotating with respect to the time-foliation by the Riemann normal coordinates $X^0 = \text{const}$.. We can then

formally write $U^i = (\vec{\omega} \times \vec{X})^i$ $(i = 1, 2, 3)$ where $\vec{\omega}$ is some angular velocity vector. Then let us further assume that we are comparing relative accelerations on the constant-time slice $X^0 = 0$ in the Riemann normal frame so that we obtain

$$\delta a^\mu \sim \Gamma^\mu{}_{\nu\kappa} U^\nu U^\kappa \sim R^\mu{}_{0j0} X^j + R^\mu{}_{0ji} X^j (\vec{\omega} \times \vec{X})^i + \dots , \tag{1}$$

where $R_{\mu\nu\kappa\lambda}$ is the Riemann tensor along the wordline of the particle and the three dots denote higher-order terms in the size of the body. Now by averaging over $\vec{X}$, we see that we are able to choose a centre of coordinates such that the first, dipole-type term vanishes. However, the second term cannot be reduced by any coordinate shift. If we then take any system of particles, be it dust or a body held together by some force, we always see that the centre of mass of this system is subject to a residual *spin-curvature force*.

1.1. *Mathisson-Papapetrou-Dixon equations*

Procedures such as the one sketched above were carried out in various degrees of generality by Mathisson,[1] Papapetrou[2] and Dixon[3] to derive the equations of motion of extended bodies in curved space-time

$$\frac{\mathrm{D}P_\mu}{\mathrm{d}\tau} = -\frac{1}{2} R_{\mu\nu\kappa\lambda} \dot{x}^\nu S^{\kappa\lambda} + \dots , \tag{2}$$

$$\frac{\mathrm{D}S^{\mu\nu}}{\mathrm{d}\tau} = 2P^{[\mu} \dot{x}^{\nu]} + \dots , \tag{3}$$

where $x^\nu(\tau)$ is a referential position within the body, P_μ its momentum, and $S^{\kappa\lambda}$ has the general meaning of an angular momentum or "gravitomagnetic dipole" about its own center of mass. The parameter chosen here is proper time τ, but the equations can be parametrized by any other parameter. It should be noted that even though P_μ has the meaning of the overall linear momentum of the system, it may not be strictly true that the word-line is chosen so that $\dot{x}^\nu \propto P^\nu$.[4]

The three dots in the equation above denote higher-order terms corresponding to the interactions of the higher-order multipoles with the background curvature. The ability to discard higher-order terms is possible only for compact objects, for many astrophysical bodies the spin-curvature term is vastly sub-dominant. However, one can also easily show that even for compact objects, the spin-curvature coupling is suppressed to the 1.5 post-Newtonian order on the level of the equations of motion of binaries, and to linear order in the mass ratio. Thus, there is a limit to the applicability of the discussion of the strictly "test-particle" motion and at some point cross-reaction with other "non-test" effects cannot be ignored any more.

1.2. *The ambiguity of x^ν and $S^{\kappa\lambda}$*

As discussed at length in the literature the procedure of finding a center of mass for a rotating body is ambiguous.[4,5] In particular, the result depends on the slicing

(the frame) we choose for the worldtube of the body. This is an issue even in flat space-time, where one is free to transform to another center of mass and angular-momentum tensor by the shifts[6]

$$x'^{\mu} = x^{\mu} + \delta x^{\mu}\,, \quad S'^{\nu\kappa} = S^{\nu\kappa} + P^{\nu}\delta x^{\kappa} - P^{\kappa}\delta x^{\nu}\,, \tag{4}$$

where δx^{μ} is some shift vector. This worldline shifting has a natural covariant extension to curved space-time by using Synge's world functions, but the overall physical content of the ambiguity is the same in both cases.[7]

A way to fix this "gauge freedom" of the worldline and the angular-momentum tensor is to choose a so-called *supplementary spin condition* of the form $S^{\mu\nu}V_{\nu} = 0$ with V_{ν} some time-like vector. The vector V^{μ} then has the meaning of the frame in which the center of mass and angular momentum are measured. When chosen judiciously, the supplementary condition eliminates all the nonphysical degrees of freedom from the system and one is left with only the physical ones to be evolved. Sometimes the choice of V^{ν} is implicit; the vector V^{ν} refers to the quantities it is supposed to fix (such as $x^{\nu}, S^{\kappa\lambda}$). In such cases one often ends up in a system where the non-physical degrees of freedom are not entirely constrained.[4,5,8,9]

2. Hamiltonian formalism

There is a range of reasons to use the Hamiltonian formalism, for the spinning-particle motion, from efficient numerical methods,[9] through perturbation theory,[10] to the use of the results in the so-called Effective-one-body model for relativistic binaries.[7,11] Let us now briefly summarize the Hamiltonians introduced by Witzany *et al.* in a previous work.[9] Consider the three covariant Hamiltonians corresponding to different choice of the frame V^{ν}:

$$V^{\nu} \text{ parallel transported along } x^{\nu}(\tau): \tag{5}$$

$$H_{\mathrm{KS}} = \frac{1}{2m} g^{\mu\nu} P_{\mu} P_{\nu}\,, \tag{6}$$

$$V^{\nu} \propto P^{\nu}: \tag{7}$$

$$H_{\mathrm{TD}} = \frac{1}{2\mathcal{M}} \left(g^{\mu\nu} - \frac{4 S^{\nu\gamma} R^{\mu}{}_{\gamma\kappa\lambda} S^{\kappa\lambda}}{4\mathcal{M}^2 + R_{\chi\eta\omega\xi} S^{\chi\eta} S^{\omega\eta}} \right) P_{\mu} P_{\nu}\,, \tag{8}$$

$$V^{\nu} = \dot{x}^{\nu}: \tag{9}$$

$$H_{\mathrm{MP}} = \frac{1}{2m} g^{\mu\nu} P_{\mu} P_{\nu}\,, \tag{10}$$

where $m = -\dot{x}^{\nu} P_{\nu}$, $\mathcal{M} = \sqrt{-P^{\nu} P_{\nu}}$ are mass parameters with slightly different meanings but all corresponding roughly to the mass of the spinning body.[9,12]

The equations of motion are obtained with the use of the Poisson bracket

$$\{x^\mu, P_\nu\} = \delta^\mu_\nu\,,\ \{x^\mu, S^{\kappa\lambda}\} = 0\,,\ \{x^\mu, x^\nu\} = 0\,, \tag{11}$$

$$\{P_\mu, P_\nu\} = -\frac{1}{2} R_{\mu\nu\kappa\lambda} S^{\kappa\lambda}\,, \tag{12}$$

$$\{S^{\mu\nu}, P_\kappa\} = -\Gamma^\mu{}_{\lambda\kappa} S^{\lambda\nu} - \Gamma^\nu{}_{\lambda\kappa} S^{\mu\lambda}\,, \tag{13}$$

$$\{S^{\mu\nu}, S^{\kappa\lambda}\} = g^{\mu\kappa} S^{\nu\lambda} - g^{\mu\lambda} S^{\nu\kappa} + g^{\nu\lambda} S^{\mu\kappa} - g^{\nu\kappa} S^{\mu\lambda}\,, \tag{14}$$

There is a set of canonical coordinates that reduce the Poisson brackets to the simple canonical form, these were also discussed by Witzany *et al.*[9]

2.1. *Hunting down degrees of freedom*

Amongst the choices for the supplementary conditions discussed above, only the so-called Tulczyjew-Dixon or "covariant" supplementary spin condition $V^\nu \propto P^\nu$ constrains the gauge freedom fully. This can be seen, in particular, by counting the degrees of freedom in phase space within the Hamiltonian formalism. A covariant formalism for a spinning particle parametrized by proper time τ should then not have more than 4 orbital degrees of freedom (corresponding to the 4 space-time coordinates), and a *single* degree of freedom corresponding to the spin sector (recall that each degree of freedom corresponds to a pair of canonically conjugate coordinates). If a larger number degrees of freedom crops up in the system of evolution equations, there are necessarily nonphysical or gauge-type degrees of freedom left in the system.

This being said, the issue with the Hamiltonian $H_{\rm TD}$ presented in equation (8) is the following. The supplementary spin condition $S^{\mu\nu}P_\nu = 0$ is only an *integral of motion* of $H_{\rm TD}$ when used along with the brackets (11)-(14). This means that in a numerical evolution one finds initial data such that $S^{\mu\nu}P_\nu = 0$ initially and then evolves them – the system consequently conserves the supplementary condition up to numerical noise.

Nevertheless, what one *cannot* do is to assume $S^{\mu\nu}P_\nu = 0$ as an identity to reduce the number of degrees of freedom that are evolved numerically while respecting the geometric structure of the problem.[9] That, is the system still keeps an additional gauge degree of freedom in the numerical evolution, even though it stays zero up to numerical error. Is there perhaps a way to obtain just the minimal Hamiltonian system without this redundant degree of freedom?

Efforts of this type have actually been carried out perturbatively with the use of the so-called Newton-Wigner supplementary spin condition $V^\mu \propto P^\mu/\mathcal{M} + \xi^\mu(x^\nu)$, where $\xi^\mu(x^\nu)$ is a fixed time-like vector field.[7,8] As such, this supplementary spin condition is necessarily non-covariant and refers to an additional fixed structure on the background. In other words, the vector $\xi^\mu(x^\nu)$ defines a privileged observer frame and the dynamics are in some sense defined with respect to this privileged frame. This impression is even more strengthened by the fact that the constraint procedure of the algebra (11)-(14) by the Newton-Wigner condition seems to be necessarily

coupled to a constraint procedure on the time parameter. To rephrase this, the Newton-Wigner condition forces one's hand into an explicit 3+1 formalism – and covariance is lost.

2.2. *Dirac constraint procedure*

Let me now sketch the Dirac constraint procedure for the Tulczyjew-Dixon constraint $S^{\mu\nu}P_\nu = 0$ applied to the Poisson bracket (11)-(14). This should serve as a covariant counter-example to the procedures using the Newton-Wigner condition.

It was already shown in the paper by Witzany *et al.*[9] that one can add the Tulczyjew-Dixon Lagrangian-constraint term to the "minimal" Hamiltonian (6), solve for the Lagrange multiplier, and the resulting Hamiltonian gives the correct Mathisson-Papapetrou-Dixon equations for the motion of the body under the Tulczyjew-Dixon supplementary spin condition. The resulting Hamiltonian is actually H_{TD} presented in eq. (8). That is, we already *know* that applying the constraint $S^{\mu\nu}P_\nu = 0$ as a Lagrangian constraint to the system will lead to the correct dynamics, one only needs to figure out the correct *Dirac bracket* corresponding to this constraint.

Let me now briefly summarize the Dirac constraint procedure.[13,14] One starts with a set of n independent constraints $\Phi_a = 0$, $a = 1...n$ with a Poisson bracket $\{\Phi_a, \Phi_b\} = C_{ab}$, where one assumes the matrix C to be invertible, $\exists (C^{-1})^{bc}$: $C_{ab}(C^{-1})^{bc} = \delta_a^c$. Then we define the Dirac bracket $[,]$ for any phase-space functions A, B as

$$[A, B] = \{A, B\} - \{A, \Phi_a\}(C^{-1})^{ab}\{\Phi_b, B\}\,. \tag{15}$$

Now it is easy to see that this bracket fulfils the Jacobi identity and that $[\Phi_c, A] = 0$ for any A when evaluated on the $\Phi_a = 0$ hypersurface in phase space. As a result, the set of constraints $\Phi_a = 0$ can then be used as an *identity* on phase space in the sense that its use will commute with the Dirac bracket. In particular, it can be shown that the system with a given Hamiltonian and a Dirac bracket is equivalent to the system with the same Hamiltonian and a Lagrangian-constraint term.[13,14]

Now to the specific case of the constraint $S^{\mu\nu}P_\nu = 0$. Even though there are 4 components to this equation, not all of them are linearly independent. For instance, independently on whether the constrain is fulfilled or not, it holds that $P_\mu(S^{\mu\nu}P_\nu) = 0$, and $w_\mu(S^{\mu\nu}P_\nu) = 0$, where $w_\mu = \epsilon_{\mu\rho\sigma\chi}S^{\rho\sigma}P^\chi$ and $\epsilon_{\mu\rho\sigma\chi}$ is the Levi-Civita pseudo-tensor. As such, there are only *two* non-trivial constraints on the system hidden in the formula $S^{\mu\nu}P_\nu = 0$. Nevertheless, as already shown partially by Witzany *et al.*,[9] one can proceed in a covariant fashion by defining

$$C^{\mu\nu} = \{S^{\mu\kappa}P_\kappa, S^{\nu\lambda}P_\lambda\} = -\tilde{\mathcal{M}}^2 S^{\mu\nu}\,, \tag{16}$$

$$\tilde{\mathcal{M}}^2 \equiv -P_\lambda P^\lambda + \frac{1}{4}R_{\gamma\rho\sigma\chi}S^{\gamma\rho}S^{\sigma\chi}\,, \tag{17}$$

$$C^\dagger_{\kappa\lambda} = -\frac{1}{\tilde{\mathcal{M}}^2 S^2}S_{\kappa\lambda}\,, \tag{18}$$

where $S^2 \equiv S^{\mu\nu}S_{\mu\nu}/2$. The matrix $C^\dagger_{\kappa\lambda}$ is then a pseudo-inverse of $C^{\mu\nu}$ on the non-trivial sub-spaces of the constraint (specifically, $C^{\mu\nu}C^\dagger_{\nu\lambda}$ projects out the sub-spaces $\sim P_\mu, w_\mu$ discussed above). One can then covariantly define the Dirac bracket using the original Poisson bracket (11)-(14) as

$$[A,B] = \{A,B\} - \{A, S^{\mu\nu}P_\nu\}C^\dagger_{\mu\kappa}\{S^{\kappa\lambda}P_\lambda, B\}\,. \tag{19}$$

The resulting brackets for $P_\mu, x^\nu, S^{\kappa\lambda}$ are now straightforward to compute and they contain various orders of new spin and curvature terms as compared to the original brackets. (The full expressions and simplified bases will be reported on elsewhere.)

2.3. *The zeroth-order problem*

The bracket (19) is essentially what we have been looking for: only a minimal number of degrees of freedom is left in the phase space and one can happily evolve them in the Hamiltonian formalism defined by the Dirac bracket and the minimal Hamiltonian (6). In particular, it is advantageous to define the spin vector

$$s^\lambda \equiv \frac{1}{2\mathcal{M}}\epsilon^{\lambda\gamma\rho\sigma}S_{\gamma\rho}P_\sigma\,, \tag{20}$$

$$\Rightarrow S^{\mu\nu} = \frac{1}{\mathcal{M}}\epsilon^{\mu\nu\kappa\lambda}P_\kappa s_\lambda\,, \tag{21}$$

where the second equality applies on the $S^{\mu\nu}P_\nu = 0$ hypersurface. This vector partially reduces the number of the many dependent components of the tensor $S^{\mu\nu}$, but not fully, since it also holds that $s^\mu P_\mu = 0$.

Now one would like to transform to canonical coordinates so that the bracket (19) reduces to a simple form. One soon realizes that one of the biggest obstacles is the *zeroth-order problem*. That is, even in flat space-time $R_{\mu\nu\kappa\lambda} = 0$ and Minkowski coordinates, the bracket is quite non-trivial. Specifically, it reduces in the pure Minkowski case to

$$[x^\mu, P_\nu] = \delta^\mu_\nu\,,\ [P_\mu, P_\nu] = 0\,,\ [P_\mu, s^\lambda] = 0\,, \tag{22}$$

$$[x^\mu, x^\nu] = \frac{1}{\mathcal{M}^3}\epsilon^{\mu\nu\kappa\lambda}P_\kappa s_\lambda\,, \tag{23}$$

$$[x^\mu, s^\nu] = \frac{P^\mu s^\nu - s^\mu P^\nu}{\mathcal{M}}\,, \tag{24}$$

$$[s^\mu, s^\nu] = -\frac{1}{\mathcal{M}}\epsilon^{\mu\nu\kappa\lambda}P_\kappa s_\lambda\,. \tag{25}$$

There is no obvious and/or elegant way to cover this Dirac algebra with canonical coordinates.

2.4. *Is Newton-Wigner the only way?*

The difficulty with Poisson brackets of spinning systems in flat space-time was already the concern of the works of Newton and Wigner,[15] even though there the

motivation was the quantum analogue of this problem for the commutation relations of operators. Jordan[16] proved that the Newton-Wigner basis of positions and spins is the only basis such that the position "transforms as a position should". (For a review and a classical rederivation, see the recent paper by Schwartz and Giulini.[17])

The results on the Newton-Wigner basis can be rephrased in the context of this work as follows. Consider any physical Dirac-constraint procedure that 1) reduces the unphysical spin degrees of freedom, 2) forces the time parametrization to be coordinate time $t = x^0$ (by using the $P^\mu P_\mu = -\mathcal{M}^2$ on-shell constraint[13,18]), and 3) parametrizes the phase-space in terms of variables $X^i, P_j, \tilde{S}^k$ such that P_j are spatial components of P_μ and such that the final bracket is

$$[X^i, X^j]' = 0\,, \tag{26}$$
$$[X^i, P_j]' = \delta^i_j\,, \tag{27}$$
$$[X^i, \tilde{S}^j]' = 0\,, \tag{28}$$
$$[\tilde{S}^j, P_i]' = 0\,, \tag{29}$$
$$[\tilde{S}^i, \tilde{S}^j]' = \epsilon^{ijk}\tilde{S}^k\,, \tag{30}$$

where ϵ^{ijk} is the permutation symbol. Then (up to singular systems), the variables $X^i, P_j, \tilde{S}^k$ are *necessarily* the Newton-Wigner variables with $\xi^\mu = \delta^\mu_0$.

One conclusion of this theorem is that if one was to cover the algebra (22)-(25) with coordinates X^μ canonically conjugate to P_μ and two other canonically conjugate coordinates on the spin sector of the phase space, then these would necessarily be Newton-Wigner variables when reduced to spatial sections. Even this result would be interesting, since it would "covariantize" the Newton-Wigner prescription. Of course, it is also possible to construct canonical coordinates that do not have $\sim P_\mu$ as canonical variables, but these seem to almost always violate manifest rotational and/or translational invariance of the dynamics.

3. Summary and Outlooks

I have shown that capturing the spin-curvature coupling of a test body on a curved background in a covariant Hamiltonian formalism – while eliminating redundant degrees of freedom – is a non-trivial matter. This matter is also far from closed. In the future, I hope to find a set of "covariant Newton-Wigner variables" for the brackets (22)-(25) and formulate a (possibly implicit) iteration problem to canonicalize the full curved-spacetime brackets (19). These should provide a first, truly general basis for the canonical treatment of spinning test particles in general relativity (but see also the work of Steinhoff in the ADM formalism[19]).

This result should then have obvious pay-offs for the relativistic two-body problem and gravitational-wave modelling. For example, the procedure of Vines *et al.*[7] suggests that the minimally coupled spinning particle (that is, with no additional multipoles) under the Tulczyjew-Dixon condition represents a "test Kerr black hole"

on the curved background, since its effective quadrupole in the Newton-Wigner variables matches the Hansen multipole of a Kerr black hole.[20] Does this apply at all orders? A more general and robust formalism is needed for an answer. The question of the minimally coupled particle is all the more interesting due to the fact that the motion of a spining test particle under the Tulzcyjew-Dixon condition is integrable at $O(S)$ in Kerr space-time,[10,21,22] but apparently non-integrable at $O(S^2)$.[23] This then naturally ties-in to the important question of the integrability and smoothness of gravitational-wave inspirals and our ability to efficiently generate predictions for them.[24,25]

Acknowledgements

I would like to thank to thank Oldřich Semerák for organizing this interesting parallel session. This work was supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 894881.

References

1. M. Mathisson, Neue mechanik materieller systemes, *Acta Phys. Polon.* **6**, 163 (1937).
2. A. Papapetrou, Spinning test-particles in general relativity. I, *Proc. Roy. Soc. Lond.* **A209**, 248 (1951).
3. W. Dixon, A covariant multipole formalism for extended test bodies in general relativity, *Il Nuovo Cimento (1955-1965)* **34**, 317 (1964).
4. L. F. O. Costa and J. Natário, Center of mass, spin supplementary conditions, and the momentum of spinning particles, in *Equations of Motion in Relativistic Gravity*, (Springer, 2015) pp. 215–258.
5. L. F. O. Costa, G. Lukes-Gerakopoulos and O. Semerák, On spinning particles in general relativity: Momentum-velocity relation for the Mathisson-Pirani spin condition, *Phys. Rev.* **D97**, p. 084023 (2018).
6. K. Kyrian and O. Semerák, Spinning test particles in a kerr field–ii, *Mon. Not. Roy. Astron. Soc.* **382**, p. 1922 (2007).
7. J. Vines, D. Kunst, J. Steinhoff and T. Hinderer, Canonical Hamiltonian for an extended test body in curved spacetime: To quadratic order in spin, *Phys. Rev.* **D93**, p. 103008 (2016).
8. E. Barausse, E. Racine and A. Buonanno, Hamiltonian of a spinning test particle in curved spacetime, *Phys. Rev.* **D80**, p. 104025 (2009), [Erratum: Phys. Rev.D85,069904(2012)].
9. V. Witzany, J. Steinhoff and G. Lukes-Gerakopoulos, Hamiltonians and canonical coordinates for spinning particles in curved space-time, *Classical and Quantum Gravity* **36**, p. 075003 (2019).
10. V. Witzany, Hamilton-Jacobi equation for spinning particles near black holes, *Physical Review D* **100**, p. 104030 (2019).
11. E. Barausse and A. Buonanno, Improved effective-one-body Hamiltonian for spinning black-hole binaries, *Phys. Rev. D* **81**, p. 084024 (April 2010).
12. O. Semerák, Spinning test particles in a Kerr field—I, *Monthly Notices of the Royal Astronomical Society* **308**, 863 (1999).
13. P. A. M. Dirac, *Lectures on quantum field theory* (Yeshiva Univ., 1966).

14. A. Hanson, T. Regge and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, 1976).
15. T. D. Newton and E. P. Wigner, Localized States for Elementary Systems, *Rev. Mod. Phys.* **21**, 400 (1949).
16. T. F. Jordan, Simple derivation of the Newton-Wigner position operator, *Journal of Mathematical Physics* **21**, 2028 (1980).
17. P. K. Schwartz and D. Giulini, Classical perspectives on the Newton–Wigner position observable, *International Journal of Geometric Methods in Modern Physics* **17**, p. 2050176 (2020).
18. A. J. Hanson and T. Regge, The Relativistic Spherical Top, *Annals Phys.* **87**, p. 498 (1974).
19. J. Steinhoff, Canonical formulation of spin in general relativity, *Annalen der Physik* **523**, 296 (2011).
20. R. O. Hansen, Multipole moments of stationary space-times, *J. Math. Phys.* **15**, 46 (1974).
21. R. Rüdiger, Conserved quantities of spinning test particles in general relativity. I, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **375**, 185 (1981).
22. R. Rüdiger, Conserved quantities of spinning test particles in general relativity. II, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **385**, 229 (1983).
23. O. Zelenka, G. Lukes-Gerakopoulos, V. Witzany and O. Kopáček, Growth of resonances and chaos for a spinning test particle in the schwarzschild background, *Physical Review D* **101**, p. 024037 (2020).
24. E. E. Flanagan and T. Hinderer, Transient resonances in the inspirals of point particles into black holes, *Physical review letters* **109**, p. 071102 (2012).
25. G. Lukes-Gerakopoulos and V. Witzany, Non-linear effects in EMRI dynamics and their imprints on gravitational waves, *arXiv:2103.06724, to appear in "Handbook of Gravitational Wave Astronomy"* (Eds. C. Bambi, S. Katsanevas, and K. Kokkotas; Springer, Singapore) (2021).

Gravitomagnetic resonance and gravitational waves

Matteo Luca Ruggiero
Politecnico di Torino, Torino - Italy and INFN, Laboratori Nazionali di Legnaro, Legnaro - Italy
E-mail: matteo.ruggiero@polito.it

Antonello Ortolan
INFN, Laboratori Nazionali di Legnaro, Legnaro - Italy

We show that using Fermi coordinates it is possible to describe the gravitational field of a wave using a gravitoelectromagnetic analogy. In particular, we show that using this approach, a new phenomenon, called gravitomagnetic resonance, may appear. We describe it both from classical and quantum viewpoints, and suggest that it could in principle be used as the basis for a new type of gravitational wave detectors.

Keywords: Gravitomagnetism; gravitational waves; gravitomagnetic resonance

1. Introduction

General Relativity (GR) is the best model that we have to understand gravitational interactions, and its predictions were verified with great accuracy during last century, even though we know that there are difficulties to explain, in the general relativistic framework, observations on galactic and cosmological scales, without claiming the existence of *dark components*.[1,2] Remarkably, GR not only predicts corrections to known Newtonian effects such as in the case of perihelion advance, but there are general relativistic effects without Newtonian counterparts: for instance, this is the case of gravitational waves and the so-called gravitomagnetic effects produced by mass currents.

As for gravitational waves, the first indirect evidence of their existence came from the observation of the binary pulsar B1913+16, whose orbital parameters are modified by the emission of gravitational waves.[3,4] It took about 100 years after the publication of Einstein's theory of gravity to obtain, in 2015, the first direct detection of gravitational waves,[5] which was the beginning of gravitational wave astronomy.

It is well known[6] that Einstein equations, in weak-field approximation (small masses, low velocities), can be written in analogy with Maxwell equations for the electromagnetic field, where the mass density and current play the role of the charge density and current, respectively; more in general, both the inertial and curvature effects in the vicinity of a given world-line, can be dealt with using a gravitoelectromagnetic formalism.[7,8] These gravitomagnetic effects are very small if compared to the gravitoelectric ones, originating from mass density and, consequently, it is very difficult to measure them. Nonetheless, there have been various attempts and proposals: we remember the LAGEOS tests around the Earth,[9,10] the subsequent

LARES mission,[11,12] and the recent measurements performed with laser-tracked satellites.[13] A comprehensive analysis of the Lense-Thirring effect in the solar system can be found in Ref. 14. The mission Gravity Probe B[15] was launched to measure the precession of orbiting gyroscopes.[16] There have been other proposals, such as LAGRANGE, which exploit spacecrafts located in the Lagrangian points of the Sun-Earth system,[17] or the use of satellites around the Earth.[18] In addition, we mention the GINGER experiment, which aims to measure gravitomagnetic effects in a terrestrial laboratory by using an array of ring lasers.[19–22]

Recently, the gravitomagnetic effects connected with the passage of a gravitational wave were analyzed:[23] this should not be surprising, since a gravitational wave transports angular momentum. In particular, these effects can be easily understood by using Fermi coordinates on the basis of a gravitoelectromagnetic analogy.[8] Here, we review this approach and suggest how it could be possible to detect the effects due to the magnetic-like part of a plane gravitational wave. The plan of the paper is as follows: in Section 2 we review Fermi coordinates and the definition of local spacetime metric, then we use this approach to study the effect of a plane gravitational wave in Section 3. Conclusions are in Section 4.

2. Local spacetime metric in Fermi coordinates

If we consider the world-line of a given observer, which ideally constitutes our laboratory frame, it is possible to write the expression of the local spacetime metric in its vicinity, using Fermi coordinates. This expression depends both on the background spacetime and on the properties of the world-line. Fermi coordinates in the vicinity of an arbitrary accelerated world-line with rotating tetrads were studied in Refs. 24–26, and the general expression of the line element, up to quadratic displacements $|X^i|$ from the reference world-line, turns out to be

$$ds^2 = -\left[\left(1+\frac{\mathbf{a}\cdot\mathbf{X}}{c^2}\right)^2 - \frac{1}{c^2}\left(\mathbf{\Omega}\wedge\mathbf{X}\right)^2 + R_{0i0j}X^iX^j\right]c^2dT^2 + \\ + \left[\frac{1}{c}\left(\mathbf{\Omega}\wedge\mathbf{X}\right)_i - \frac{4}{3}R_{0jik}X^jX^k\right]cdTdX^i + \left(\delta_{ij} - \frac{1}{3}R_{ikjl}X^kX^l\right)dX^idX^j. \quad (1)$$

Here, $\mathbf{X}$ is the position vector in the Fermi frame. We see that in the line element (1) there are both the gravitational effects, deriving from the curvature tensor, and the inertial effects, due to world-line acceleration $\mathbf{a}$ and the tetrad rotation $\mathbf{\Omega}$.

The metric (1) can be written in terms of the gravitoelectromagntic potentials $(\Phi, \mathbf{A})$ (see Refs. 7, 8), neglecting the terms g_{ij} related to the spatial curvature:

$$ds^2 = -\left(1-2\frac{\Phi}{c^2}\right)c^2dT^2 - \frac{4}{c}(\mathbf{A}\cdot d\mathbf{X})dt + \delta_{ij}dX^idX^j, \quad (2)$$

where

$$\Phi(T,\mathbf{X}) = \Phi^I(\mathbf{X}) + \Phi^C(T,\mathbf{X}), \quad \mathbf{A}(T,\mathbf{X}) = \mathbf{A}^I(\mathbf{X}) + \mathbf{A}^C(T,\mathbf{X}), \quad (3)$$

In particular, in the gravitoelectric potential $\Phi(T, \mathbf{X})$

$$\Phi^I(\mathbf{X}) = -\mathbf{a}\cdot\mathbf{X} - \frac{1}{2}\frac{(\mathbf{a}\cdot\mathbf{X})^2}{c^2} + \frac{1}{2}\left[|\mathbf{\Omega}|^2|\mathbf{X}|^2 - (\mathbf{\Omega}\cdot\mathbf{X})^2\right] \tag{4}$$

is the *inertial* contribution, while

$$\Phi^C(T, \mathbf{X}) = -\frac{1}{2}R_{0i0j}(T)X^iX^j \tag{5}$$

is the *curvature* contribution. As for the gravitomagnetic potential $\mathbf{A}(T, \mathbf{X})$, we may distinguish the *inertial* contribution

$$A^I_i(\mathbf{X}) = -\left(\frac{\mathbf{\Omega}c}{2}\wedge\mathbf{X}\right)_i, \tag{6}$$

and the *curvature* contribution:

$$A^C_i(T, \mathbf{X}) = \frac{1}{3}R_{0jik}(T)X^jX^k. \tag{7}$$

The gravitoelectric and gravitomagnetic fields $\mathbf{E}$ and $\mathbf{B}$ are defined in terms of the potentials by

$$\mathbf{E} = -\nabla\Phi - \frac{1}{c}\frac{\partial}{\partial T}\left(\frac{1}{2}\mathbf{A}\right), \quad \mathbf{B} = \nabla\times\mathbf{A}. \tag{8}$$

which, up to up to linear order in $|X^i|$, can be written as

$$\mathbf{E}^I = \mathbf{a}\left(1 + \frac{\mathbf{a}\cdot\mathbf{X}}{c^2}\right) + \mathbf{\Omega}\wedge(\mathbf{\Omega}\wedge\mathbf{X}), \quad E^C_i(T, \mathbf{X}) = c^2R_{0i0j}(T)X^j. \tag{9}$$

and

$$\mathbf{B}^I = -\mathbf{\Omega}c, \quad B^C_i(T, \mathbf{R}) = -\frac{c^2}{2}\epsilon_{ijk}R^{jk}{}_{0l}(T)X^l. \tag{10}$$

In summary, the gravitoelectricmagnetic fields are written in the form

$$\mathbf{E} = \mathbf{E}^I + \mathbf{E}^C, \quad \mathbf{B} = \mathbf{B}^I + \mathbf{B}^C. \tag{11}$$

In addition, the analogy with electromagnetism can be exploited to describe the motion of free test masses; in particular, the motion of free test masses *relative* to a reference mass, at rest at origin of the Fermi frame, is determined by the geodesics of the metric (1). The latter can be written in the form of a Lorentz-like force equation[7]

$$m\frac{\mathrm{d}^2\mathbf{X}}{\mathrm{d}T^2} = -m\mathbf{E} - 2m\frac{\mathbf{V}}{c}\times\mathbf{B} \tag{12}$$

up to linear order in the particle velocity $\mathbf{V} = \frac{\mathrm{d}\mathbf{X}}{\mathrm{d}T}$ (which is the *relative velocity* with respect to the reference mass). Moreover, the evolution equation of classical spinning particle with spin $\mathbf{S}$ in an external gravitomagnetic field $\mathbf{B}$ is

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T} = \frac{1}{c}\mathbf{B}\times\mathbf{S}, \tag{13}$$

in analogy with the corresponding equation for a charged spinning test particle in a magnetic field.[7]

In the following Section we will apply this formalism to plane gravitational waves.

3. Gravitomagnetic resonance due to the passage of a gravitational wave

Gravitomagnetic effects deriving from the passage of the gravitational wave can be in principle detected by using devices such as the heterodyne antenna (see e.g. Ref. 8) or studying the perturbations of planetary motion.[27] Here, we focus on a different approach, that is based on the fulfilment of a *resonance condition*.[28]

We consider a spinning particle interacting with the gravitatomagnetic field a plane wave. In the Fermi frame, where the spacetime metric is written in the form (1), we consider coordinates T, X, Y, Z with a set of unit vectors $\{\mathbf{u}_X, \mathbf{u}_Y, \mathbf{u}_Z\}$; the direction of propagation of the wave is the X axis. In this case, for a circularly polarized wave, the components of the gravitomagnetic field deriving from the spacetime curvature (and, hence, connected to the passage of the wave) can be written in the form[8]

$$B_X^C = 0, \quad B_Y^C = -\frac{A\omega^2}{2}\left[-\cos(\omega T)\, Y + \sin(\omega T)\, Z\right],$$
$$B_Z^C = -\frac{A\omega^2}{2}\left[\sin(\omega T)\, Y + \cos(\omega T)\, Z\right], \tag{14}$$

where A is the amplitude and ω the frequency of the wave. In order to study the interaction with a spinning particle, we consider a frame clockwise rotating in the YZ plane with the wave frequency ω; then, the corresponding basis vectors are $\mathbf{u}_{X'} = \mathbf{u}_X$, $\mathbf{u}_{Y'}(T) = \cos(\omega T)\,\mathbf{u}_Y - \sin(\omega T)\,\mathbf{u}_Z$ and $\mathbf{u}_{Z'}(T) = \sin(\omega T)\,\mathbf{u}_Y + \cos(\omega T)\,\mathbf{u}_Z$. As a consequence, the gravitomagnetic field is written as

$$\mathbf{B}^C(T) = \frac{A\omega^2}{2}\left[Y\,\mathbf{u}_{Y'}(T) - Z\,\mathbf{u}_{Z'}(T)\right] \tag{15}$$

Notice that $\mathbf{B}^C$ is a *static field* in the rotating frame that we have considered.

Let us consider the spin evolution equation (13); the total gravitomagnetic field is $\mathbf{B} = \mathbf{B}^I + \mathbf{B}^C$, where $\mathbf{B}^I = -\boldsymbol{\Omega} c$ and it is simply proportional to the rotation rate $\boldsymbol{\Omega}$ of the frame. We suppose that $\boldsymbol{\Omega}$ is constant and it is in the direction of propagation of the wave: then, we may write $\mathbf{B}^I = -B^I \mathbf{u}_X$ where $B^I = \Omega c$. As a consequence, the spin evolution equation turns out to be

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T} = \frac{1}{c}\left[\mathbf{B}^C(T) + \mathbf{B}^I\right] \times \mathbf{S}. \tag{16}$$

If we consider the frame co-rotating with $\mathbf{B}^C(T)$, since $\boldsymbol{\omega} = -\omega \mathbf{u}_X$ is the rotation rate, the time derivatives in the two frames are related by

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T} = \left(\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T}\right)_{\mathrm{rot}} + \boldsymbol{\omega} \times \mathbf{S} = \left(\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T}\right)_{\mathrm{rot}} - \omega \mathbf{u}_{X'} \times \mathbf{S}. \tag{17}$$

Then, from Eqs. (16) and (17) we get

$$\left(\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}T}\right)_{\mathrm{rot}} = \left[\Delta\omega\mathbf{u}_{X'} + \frac{1}{c}\mathbf{B}^C\right] \times \mathbf{S} = \frac{1}{c}\mathbf{B}_{eff} \times \mathbf{S}, \tag{18}$$

where we set $\omega - \frac{1}{c}B^I = \omega - \Omega = \Delta\omega$ and $\frac{1}{c}B^C = \omega^*$ (see below).

Then, according to Eq. (18), we may say that the spinning particle undergoes a precession determined by the static effective gravitomagnetic field $\mathbf{B}_{eff} = c\left[\Delta\omega\mathbf{u}_{X'} + \frac{1}{c}\mathbf{B}^C\right]$. Accordingly, we see that when $\Delta\omega \simeq 0$, i.e. in *resonance* condition, the spin precession is around the direction of $\mathbf{B}^C$, which is in any case in the YZ plane, so the precession may flip the spin completely. In summary, the *gravitomagnetic resonance* is obtained when the rotation rate of the frame is equal to the frequency of the gravitational wave. It is important to remember that all precessions are referred to a reference spinning particle,[23] at the origin of the Fermi frame so, in any case, we are talking about a *relative precession.*

Actually, the above description, which is analogous to the classical dynamics of a magnetic moment in a magnetic field, can be translated into quantum terms for a two-level system,[28] taking into account the Hamiltonian description of the interaction of the spin of intrinsic particles in a gravitational field.[7, 29–31] As a consequence, we may introduce a probability transition for spinning particles in the field of a gravitational wave. Let us suppose that $|g>$ and $|e>$ are the two eigenvectors, respectively of the ground and excited states, of the projection of the spinning particle along the X axis. If we suppose that a spin is, at $t = 0$, in the ground state $|g>$, the probability of transition to the excited state $|e>$ at time t is given by Rabi's formula

$$P_{g\to e}(T) = \frac{(\omega^*)^2}{(\omega^*)^2 + \Delta\omega^2}\sin^2\left(\sqrt{(\omega^*)^2 + \Delta\omega^2}\frac{T}{2}\right). \tag{19}$$

Again, at resonance, i.e. when $\Delta\omega = 0$, or $\omega = \Omega$, even a weak gravitational field can reverse the direction of the spin: the probability of transition is equal to 1 *independently of the strength of the gravitomagnetic field*, for $T = \frac{2n+1}{(\omega^*)}\pi$.

Let us add a comment on how the resonance condition could be achieved without requiring the physical rotation of our reference frame. In fact, if we consider charged spinning particles, we may get an equivalent situation by using a true magnetic field, on the basis of Larmor theorem, which states the equivalence between a system of electric charges in a magnetic field and the same system rotating with the Larmor frequency. So a magnetic field can be used to produce the gravitomagnetic field $\mathbf{B}^I$.

4. Conclusions

We have seen that, using Fermi coordinates, it is possible to emphasize the gravitomagnetic effects connected to the passage of a gravitational wave. Current detectors such as LIGO and VIRGO can detect only the interaction of a system of masses

with the electric-like component of the field, so we discussed the possibility of detecting the interaction of a suitable probe with the magnetic-like component of the wave field. In particular, we considered the interaction of the wave with a spinning particle, both using a classical and a quantum approach, and we showed that in analogy with what happens in electromagnetism, a gravitational magnetic resonance phenomenon may appear when the reference frame rotates along the direction of propagation of the wave and the rotation rate is equal to the wave frequency. Actually, since it is not possible to have physical rotations for arbitrary frequencies, we pointed out that an equivalent situation can be obtained by using a true magnetic field, on the basis of the Larmor theorem. As for the detection of this effect, we imagine not to detect the modification of a single spinning particle, rather to consider a great number of identical particles. For instance, the precession induced by the gravitational wave can modify the magnetization of a macroscopical sample which, in turn, can be detected by measuring the differences in the magnetic field produced.

References

1. C. M. Will, Was Einstein Right? A Centenary Assessment, in *General Relativity and Gravitation. A Centennial Perspective*, eds. A. Ashtekar, B. K. Berger, J. Isenberg and M. MacCallum (Cambridge University Press, Cambridge, jul 2015).
2. I. Debono and G. F. Smoot, General Relativity and Cosmology: Unsolved Questions and Future Directions, *Universe* **2**, p. 23 (2016).
3. R. Hulse and J. Taylor, Discovery of a pulsar in a binary system, *Astrophys. J. Lett.* **195**, L51 (1975).
4. J. M. Weisberg and J. H. Taylor, Relativistic binary pulsar B1913+16: Thirty years of observations and analysis, *ASP Conf. Ser.* **328**, p. 25 (2005).
5. B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Observation of gravitational waves from a binary black hole merger, *Physical review letters* **116**, p. 061102 (2016).
6. M. L. Ruggiero and A. Tartaglia, Gravitomagnetic effects, *Nuovo Cim.* **B117**, 743 (2002).
7. B. Mashhoon, Gravitoelectromagnetism: A Brief review, in *The Measurement of Gravitomagnetism: A Challenging Enterprise*, ed. L. Iorio (Nova Science, New York, 2003)
8. M. L. Ruggiero and A. Ortolan, Gravito-electromagnetic approach for the space-time of a plane gravitational wave, *Journal of Physics Communications* **4**, p. 055013 (may 2020).
9. I. Ciufolini and E. C. Pavlis, A confirmation of the general relativistic prediction of the lense–thirring effect, *Nature* **431**, p. 958 (2004).
10. I. Ciufolini, E. C. Pavlis, J. Ries, R. Koenig, G. Sindoni, A. Paolozzi and H. Newmayer, Gravitomagnetism and its measurement with laser ranging to the lageos satellites and grace earth gravity models, in *General Relativity and John Archibald Wheeler*, eds. I. Ciufolini and R. A. Matzner (Springer, 2010) pp. 371–434.
11. I. Ciufolini, A. Paolozzi, E. Pavlis, J. Ries, V. Gurzadyan, R. Koenig, R. Matzner, R. Penrose and G. Sindoni, Testing general relativity and gravitational physics using the lares satellite, *The European Physical Journal Plus* **127**, p. 133 (2012).
12. I. Ciufolini, A. Paolozzi, E. C. Pavlis, R. Koenig, J. Ries, V. Gurzadyan, R. Matzner, R. Penrose, G. Sindoni, C. Paris *et al.*, A test of general relativity using the lares and

lageos satellites and a grace earth gravity model, *The European Physical Journal C* **76**, p. 120 (2016).
13. D. Lucchesi, M. Visco, R. Peron, M. Bassan, G. Pucacco, C. Pardini, L. Anselmo and C. Magnafico, A 1% measurement of the gravitomagnetic field of the earth with laser-tracked satellites, *Universe* **6**, p. 139 (2020).
14. L. Iorio, H. I. M. Lichtenegger, M. L. Ruggiero and C. Corda, Phenomenology of the Lense-Thirring effect in the Solar System, *Astrophys. Space Sci.* **331**, 351 (2011).
15. C. F. Everitt, D. DeBra, B. Parkinson, J. Turneaure, J. Conklin, M. Heifetz, G. Keiser, A. Silbergleit, T. Holmes, J. Kolodziejczak *et al.*, Gravity probe b: Final results of a space experiment to test general relativity, *Physical Review Letters* **106**, p. 221101 (2011).
16. L. I. Schiff, Possible new experimental test of general relativity theory, *Physical Review Letters* **4**, p. 215 (1960).
17. A. Tartaglia, D. Lucchesi, M. L. Ruggiero and P. Valko, How to use the Sun–Earth Lagrange points for fundamental physics and navigation, *Gen. Rel. Grav.* **50**, p. 9 (2018).
18. M. L. Ruggiero and A. Tartaglia, Test of gravitomagnetism with satellites around the earth, *The European Physical Journal Plus* **134**, p. 205 (2019).
19. F. Bosi, G. Cella, A. Di Virgilio, A. Ortolan, A. Porzio, S. Solimeno, M. Cerdonio, J. Zendri, M. Allegrini, J. Belfi *et al.*, Measuring gravitomagnetic effects by a multi-ring-laser gyroscope, *Physical Review D* **84**, p. 122002 (2011).
20. M. L. Ruggiero, Sagnac effect, ring lasers and terrestrial tests of gravity, *Galaxies* **3**, 84 (2015).
21. A. Di Virgilio, M. Allegrini, A. Beghi, J. Belfi, N. Beverini, F. Bosi, B. Bouhadef, M. Calamai, G. Carelli, D. Cuccato *et al.*, A ring lasers array for fundamental physics, *Comptes Rendus Physique* **15**, 866 (2014).
22. A. Tartaglia, A. Di Virgilio, J. Belfi, N. Beverini and M. L. Ruggiero, Testing general relativity by means of ring lasers, *Eur. Phys. J. Plus* **132**, p. 73 (2017).
23. D. Bini, A. Geralico and A. Ortolan, Deviation and precession effects in the field of a weak gravitational wave, *Phys. Rev. D* **95**, p. 104044 (May 2017).
24. W.-T. Ni and M. Zimmermann, Inertial and gravitational effects in the proper reference frame of an accelerated, rotating observer, *Physical Review D - Particles and Fields* **17**, p. 1473 (March 1978).
25. W.-Q. Li and W.-T. Ni, Coupled inertial and gravitational effects in the proper reference frame of an accelerated, rotating observer, *Journal of Mathematical Physics* **20**, 1473 (July 1979).
26. K.-P. Marzlin, Fermi coordinates for weak gravitational fields, *Phys. Rev. D* **50**, 888 (Jul 1994).
27. L. Iorio and M. L. Ruggiero, Perturbations of the orbital elements due to the magnetic-like part of the field of a plane gravitational wave, *IJMPD (to appear)* (2021).
28. M. L. Ruggiero and A. Ortolan, Gravitomagnetic resonance in the field of a gravitational wave, *Phys. Rev. D* **102**, p. 101501 (Nov 2020).
29. B. Mashhoon, Gravitational couplings of intrinsic spin, *Classical and Quantum Gravity* **17**, 2399 (jun 2000).
30. F. W. Hehl and W.-T. Ni, Inertial effects of a dirac particle, *Physical Review D* **42**, p. 2045 (1990).
31. L. Ryder, Relativistic treatment of inertial spin effects, *Journal of Physics A: Mathematical and General* **31**, p. 2465 (1998).

Studying the EPRL spinfoam self-energy

Pietropaolo Frisoni
Department of Physics and Astronomy, Western University, London, Ontario, Canada
E-mail: pfrisoni@uwo.ca

I present some recent progresses in the study of the EPRL self-energy amplitude.[1] New numerical methods allow to analyze how the divergence scales, for which previous works only provided bounds spanning more than 9 orders of magnitude. I discuss the role that the Immirzi parameter plays in the asymptotic behavior, and the dependence of the scaling on some boundary data. Finally, I discuss the dynamical expectation values of some relevant geometric boundary observables.

Keywords: Spinfoam; self energy; infrared divergences; HPC; LQG.

1. Introduction

The spinfoam formalism is an attempt to define the dynamics of loop quantum gravity in an explicitly Lorentz covariant way.[2,3] It defines transition amplitudes for spinnetwork states of the canonical theory in a form of a sum (or equivalently a refinement[4]) over all the possible two-complexes having the chosen spinnetwork as boundary. This is equivalent to a sum over histories of quantum geometries providing in this way a regularised version of the quantum gravity path integral.

The state of the art of the spinfoam approach to LQG is currently the model proposed by Engle, Pereira, Rovelli and Livine (EPRL),[5,6] independently developed by Freidel and Krasnov[7] and extended to arbitrary spinnetwork states.[8,9] I assume that the reader is familiar with the EPRL-FK[a] model, and refer to the original literature[10–13] and existing reviews (e.g.[2,14–16]) for the details on how it encodes the covariant dynamics of Loop Quantum Gravity.

The EPRL model presents infrared divergences, which have a similar structure to the UV divergences in the Feynman expansion of a standard quantum field theory. The presence of divergences require a renormalization procedure. This is an important open direction of investigation in the theory, since their study and understanding is important in the definition of the continuum limit. A number of questions are still open, for example regarding the proper normalisation of the n point functions and similar.[14] Divergences have been the subject of many studies following different investigation strategies: via refining of the 2-complex as proposed in,[17–19] or via a resummation, defined for instance using group field theory/random tensor models as proposed in.[20–23] The properties of these divergences have been studied

[a]From now on I will call it just EPRL for notation convenience.

in the context of the Ponzano-Regge model of 3d quantum gravity and discrete BF theory,[24,25] group field theory[26] and EPRL model: with both Euclidean[27,28] and Lorentzian signature.[29]

Analytic estimates of divergences in the Lorentzian EPRL model[29] and[30] are based on two different methods. In[29] the "self-energy" amplitude is considered, finding a logarithmic divergence $\log K$ as a *lower bound*, where K is an artificial cut-off on the internal $SU(2)$ spins associated with the faces of the diagrams (the need to introduce a cut-off is justified by the fact that the sum on the internal spins is unbounded). The computation is rather involved and relies on the techniques developed for the asymptotic analysis of the vertex amplitude of the model.[31] This approach requires an independent study of each geometrical sector and the logarithmic divergence is obtained by looking at the non-degenerate geometries. On the other hand, in[30] is proposed an algorithm to systematically determine the potential infrared divergence for all spinfoam diagrams. The approach is based on the hypothesis that the dominant contribution to the divergence scaling of the amplitude comes from the uniform scaling of all the spins and that there is no interference between various terms of the sum. This leads to an *upper bound* on the divergence proportional to K^9.

These two bounds are therefore remarkably different, leading a window of possibilities which spans several orders of magnitude, and it is necessary to find methods of investigation that can help clarify the question. Here I report the results of a numerical study[1] regarding the scaling of the self-energy amplitude, which also investigates the dependence on the Immirzi parameter and some other boundary data. In particular, the numerical estimates confirm both the two bounds previously found.

For the numerical calculation of the EPRL vertex amplitudes, the "sl2cfoam-next" library was used.[32] All computations were performed on Compute Canada's Cedar and Graham clusters (www.computecanada.ca). The plots were made with julia[33] and Mathematica.[34] Here I present a summary of the results obtained. The interested reader can find further details and analysis, including explicit formulas for the $SU(2)$ invariant symbols and the booster functions, in the full analysis.[1]

The paper is organized as follows. In section 2 I describe the triangulation used to study the self-energy amplitude. In section 3 the amplitude is derived in the purely $SU(2)$ BF model. This intermediate step allows to compare the numerical results with the analytical ones present in the literature. In section 4 the EPRL amplitude is derived from the BF model. In section 5 I perform the numerical analysis of the EPRL self-energy amplitude. Finally, in section 6, I report the results of the dynamical expectation value of the boundary angle operators.

2. Self-energy diagram

I consider bubbles, studying one of the most elementary diagrams appearing in the self-energy amplitude.[29,30] Since the associated divergence can be viewed as the one

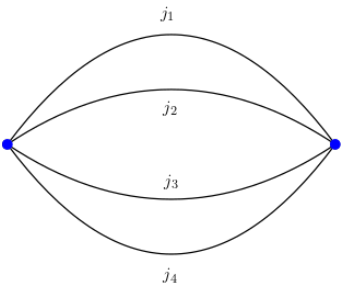


Fig. 1. Boundary graph of the spinfoam associated with the self-energy amplitude. In blue are highlighted the two four valent nodes, corresponding to the two tetrahedra in the dual triangulation. The nodes are joined by all the links. The boundary spins are denoted with j_f, where $f = 1, ..., 4$.

related to a particularly simple triangulation, from a computational point of view it turns out to be the one of the simplest divergence to deal with. The triangulation is formed by two 4-simplices joined by four tetrahedra. The associated two-complex turns out to be composed by two vertices, four edges, six internal faces (one per couple of edges) and four external faces (one per edge). The boundary of the dual triangulation is formed by two tetrahedra joined by all the faces, therefore the boundary graph consists in two four valent nodes connected by all the links (see Fig. 1).

The kinematical Hilbert space of LQG at fixed graph Γ, with L links and N nodes, is:

$$\mathcal{H}_\Gamma = L_2 \left[SU(2)^L / SU(2)^N \right] \tag{1}$$

For the self-energy diagram, we have $L = 4$, $N = 2$ and the corresponding graph[b] Γ is represented in Figure 1. Since the self-energy spinfoam graph is entirely symmetrical and has two nodes, I shall use the $+$ and $-$ symbols as labels for the latter. I use the same symbols to distinguish the corresponding intertwiners $i_\pm$, i.e. the invariant $SU(2)$ tensors associated with each node. The intertwiner space of a 4-valent node is denoted as:

$$\mathcal{I}_4 = Inv \left[V^{j_1} \otimes V^{j_2} \otimes V^{j_3} \otimes V^{j_4} \right] \tag{2}$$

where V^{j_i} is the irreducible representation of spin j_i. I consider the recoupling base (j_1, j_2) for the intertwiner space (2). That is, I fix a pairing of the links at each node and I choose the basis that diagonalises the modulus square of the sum of the $SU(2)$ generators in the pair (j_1, j_2). A basis for the Hilbert space (1) is given by the spin-network states $|j_l, i_\pm\rangle$, where the j_l's are the spin associated with each link of the graph, $l = 1...4$, and the $i_\pm$'s are a basis in the corresponding intertwiners space (2) according to the recoupling scheme. The spin-network states are interpreted

[b] From now on, I will drop the explicit dependence on Γ.

as quantum tetrahedra or, if the node have valence higher than 4, as quantum polyhedra.[c, 35,36] The corresponding discrete geometry, in which the shared faces between the tetrahedra must have the same area but nor necessarily the same shape or orientation, is called *twisted geometry*.[37]

I shall focus on the subspaces $\mathcal{H}_j$ of (1) such that all spins are the same $j_l = j$, for which the basis states are denoted as

$$|j, i_\pm\rangle \equiv \bigotimes_\pm |i_\pm\rangle = |i_+\rangle \otimes |i_-\rangle \tag{3}$$

The explicit dependence on j can be dropped in the notation, since all the spins have the same value.

3. Derivation of the BF spinfoam amplitude

I start deriving the formula for the spinfoam transition amplitude associated to the triangulation described in section 2. The strategy chosen to derive the self-energy EPRL amplitude consists in starting from the corresponding amplitude for the BF model, for which it is possible to compare our numerical computations with the exact estimate of the amplitude's divergence,[30] since in the pure $SU(2)$ model one can isolate the divergent factor. Then, I include the contribution of the booster functions (see[38]), as they encode all the details of the EPRL model. This process serves to facilitate the derivation and, at the same time, make sure that the phases and normalization factors are correct.

3.1. *Numerical scaling of the BF amplitude*

The BF self-energy amplitude is written as:

$$W_{BF}\left(j, S_+, S_-\right) = \sum_{j_{ab}} \prod_{(a,b)} \left(2j_{ab}+1\right) \{15j\}^2 \tag{4}$$

where the algebric notation $\{15j\}^2$ refers to the spinfoam in Figure 2. For the algebraic expression of the $SU(2)$ $\{15j\}$ invariant quantities, I refer to.[1,39] In (4), $(a,b) = (23, 24, 25, 35, 45, 34)$ are the labels which denote the spins attached to the 6 internal faces of the spinfoam, while the dependence on the 4 boundary spins is denoted with $j \equiv j_f$ with $f = 1...4$. In (4), the parameters specifying the boundary tetrahedra are generically denoted with S. The sums over the internal faces (explicitly highlighted in red) are unbounded: this is precisely the way in which the bubble's divergence manifests itself. Strictly speaking, the dimensional factor of the boundary faces should be taken into account in the expression for the amplitude, but this is a constant overall factor that can be ignored in the analysis, since it does not affect the scaling of the divergence. The 4-simplices are labeled in the following

[c] At least to some extent, since the Heisenberg uncertainty principle prevent us from determining the full geometry of the polyhedra. Thus, the interpretation of a spinnetwork state as a sharp polyhedron fails.

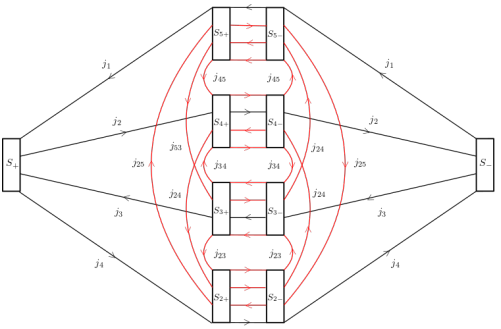


Fig. 2. Spinfoam corresponding to the BF self-energy amplitude. The two $\{15j\}$ symbols are graphically represented in order to highlight the bulk contractions.

way. The subscripts + and − distinguish the tetrahedra belonging to one vertex amplitude from those of the other, while the numeric subscripts distinguish the ones belonging to the same vertex. The two boundary tetrahedra, on the "opposite sides" of the spinfoam, are only labeled with + and − subscripts according the notation introduced in 2. The internal faces, dual to triangles, are labeled by two points indicating the tetrahedra attached to them, while the spins attached to the boundary links are labeled according to Figure 1.

In order to study numerically the scaling behavior of the amplitude, I impose an arbitrary cutoff K on the spin's values of the 6 internal faces. This allows to estimate the divergence scaling without performing an infinite number of sums. By doing so, the amplitude acquires an *artificial* dependence also on K. From now on, I refer to K as the "bulk spins cut-off".

I perform the integration over $SU(2)$ and obtain:

$$W_{BF}\left(j,i;K\right) =(-1)^{2(i_+ + i_-)} \sum_{j_{ab}} \prod_{(a,b)} (2j_{ab}+1) \sum_{i_e} \prod_{e} (2i_e+1) \{15j\}^2 \qquad (5)$$

I denoted with $i \equiv (i_+, i_-)$ the dependence on the intertwiners at the two nodes in the recoupling base (j_1, j_2), and the phase $(-1)^{2(i_+ + i_-)}$ comes from the fact that I changed the orientation of the line represented by the boundary intertwiners, according to our convention for the $\{15j\}$ symbol, following the rules of the $SU(2)$ graphical calculus.[40]

In passing from (4) to (5), I attached an intertwiner i_e with $e = 2, ..., 5$ to each edge. Therefore, each edge carries a boundary spin, three face spins and an intertwiner. Triangular inequalities constrain the intertwiner to assume values in an interval centered on a face spin, implying that, for a fixed value of the spins on the faces, the sums over these intertwiners are bounded. I choose the spinnetwork boundary state with fixed spins and intertwiners and analyze the bubble's divergence with it, since the contraction of the amplitude with less trivial boundary states (for example, coherent states) requires to compute all the possible amplitudes

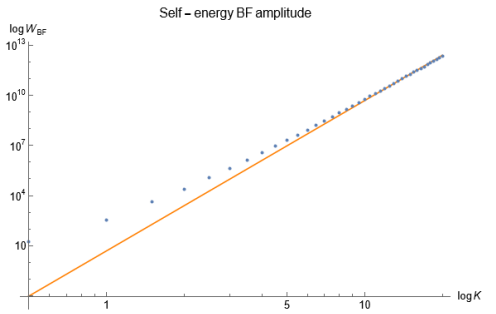


Fig. 3. Logarithmic plot of the BF self-energy amplitude scaling. The points computed with (5) are fitted with the curve $W_{BF} = 4.8 * K^9$, and the latter is shifted so that the last point coincides with the end of the curve. Notice that the divergence scaling is reached for very low values of the bulk spins cutoff K.

obtained by varying independently the intertwinwers i_+, i_-, thus increasing the computational complexity. The computation with other choices of boundary states are considered in.[1]

As previously stated, I only consider equilateral boundary spins configuration, that is, the spins $j \equiv j_f = j_1 ... j_4$ have all the same value[d]. In fixing the boundary state entirely, that is, both the boundary spins and the intertwiners, the amplitude turns out to depend only on 6 parameters, namely the 4 spins $j \equiv j_f$ and the 2 intertwinwers $i \equiv (i_+, i_-)$. To further minimize the computational time, I fix the spins of the boundary faces to their minimum non-trivial value $j = \frac{1}{2}$. Notice that the value of the boundary spins j and intertwiners i is not relevant, since in the BF model the amplitude must be independent of the values of the latter. Computing the amplitude (5) numerically, for different values of K, I verify that it correctly reproduces the divergent scaling derived in,[30] as shown in Figure 3:

$$W_{BF}(j, i, K) \propto K^9. \tag{6}$$

4. From BF to EPRL

In the BF model it is possible to isolate the divergent factor, but the aim is to use the numerical tools to study the EPRL model, where only analytical estimates of the lower[29] and upper bounds[30] on the scaling of the bubble's divergence are available in the literature. The picture is therefore not clear, especially considering that these two limits differ greatly. In fact, while the lower bound is logarithmically divergent in the cutoff K over the $SU(2)$ representation spins, the upper bound turns out to be K^9.

[d]Notice that, by doing so, each boundary intertwiner assumes a range of possible integer values ranging from 0 to $2j + 1$, so the overall phase of the amplitude is always unitary.

The crucial point of the passage from BF to EPRL model lies in the expression of the EPRL vertex amplitude. Since the reader may be unfamiliar with such amplitude form, I briefly underline the essential ideas, referring to the original paper[38] for an accurate description.

4.1. *EPRL vertex amplitude*

The EPRL vertex amplitude is built from the topological $SL(2,\mathbb{C})$ spinfoam vertex amplitude by imposing, weakly, the simplicity constraints. This results in a restriction of the unitary irreducible representations in the principal series.[10,41] In order to evaluate it in its original form, one should perform four group integrals (one of the original five integrals is redundant and has to be removed to guarantee finiteness[42]). Each group integral is, in general, a six dimensional unbounded highly oscillating integral for which numerical integration methods are not easy to implement efficiently. To get around this computational hurdle, an alternative form for the amplitude has been derived, resulting in a superposition of $SU(2)$ $\{15j\}$ symbols weighted by one booster functions $B4$ per edge in the considered vertex.

$$A_v\,(j_f,\,i_e,\gamma) = \lim_{\Delta l\to\infty} \sum_{l_f=j_f}^{j_f+\Delta l} \sum_{k_e} \prod_e (2k_e+1)\, B_4(j_f,l_f;i_e,k_e;\gamma)\{15j\}_{j_f,i_1}(l_f,k_e) \tag{7}$$

For the explicit formulas of the booster functions B_4, I refer to.[1,38,43] In[1] it is also introduced a convenient graphical notation which combines the $SU(2)$ graphical calculus with algebraic formulas to define the quantities involved. Notice that in (7) the dimensional factors attached to the boundary intertwiners i_e and to the boundary spins j_f are neglected, since I assume the latter to be determined by the boundary spinnetwork state, so that they turn out to be constant. In less trivial spinfoams, as in the self-energy case, there are typically more vertex amplitudes glued together. This implies that dimensional factors of bulk spins and intertwiners must be taken into account, as did for the BF self-energy amplitude (5), in which the elementary vertex amplitude is represented by a single $SU(2)$ invariant $\{15j\}$ symbol. The key conceptual step to obtain the expression (7), starting from the integral representation over $SL(2,C)$, is to decompose each $SL(2,C)$ integral h according to the Cartan decomposition:

$$h = ue^{(\frac{r\sigma_3}{2})}v^{-1}, \tag{8}$$

where u and v are $SU(2)$ arbitrary rotations and $r\in[0,\infty)$ is the rapidity parameter of a boost along the z axis. The compact integrals, resulting from the above parameterization, are then evaluated exactly composing the $SU(2)$ invariants in the amplitude (7). This decomposition introduces a summation over a set of auxiliary spins l_f for each face involving the vertex (excluding the gauge fixed one), for a total

of 6 distinct l_f, with lower bound $l_f \geq j_f$, and a set of auxiliary intertwiners k_e for each edge in the vertex (excluding the gauge fixed one) for a total of 4, which can assume all the values compatible with triangular inequalities. Finally, the "Y-map" imposes that the polyhedron shared by two adjacent polytopes lives in the same space-like hyperplane.[16, 44, 45]

The Y-map is present only on one side of each booster function, namely the one reaching out to the next vertex: the group elements joining at the vertex are instead multiplied together without the latter. Because of this, an infinite sum appears on the auxiliary spins l_f and the EPRL model is, in principle, recovered only in the limit in which this sum becomes infinite. The full EPRL amplitude is well defined,[42] as the summations over the l_f are convergent. Nevertheless, in order to perform a numerical evaluation of the amplitude, we need to introduce a homogeneous cut-off [e] Δl on the auxiliary spins l_f. In the following, I will refer to the cutoff Δl as the number of *shell*.

Beside the numerical precision with which the single terms that contribute to the vertex amplitude are computed, notice that this is the only approximation on which this method is based. If the convergence in the sum over the auxiliary spins l_f is sufficiently good, then we obtain a reasonable estimate of the EPRL model. As originally shown in,[38] the largest contributions to the booster functions come from configurations with $l_f = j_f$, namely the minimal admissible values for the l_f spins. The EPRL model defined with the approximation $l_f = j_f$ is usually called "simplified model". Even if the convergence of the amplitude (7) as a function of Δl is assured,[46, 47] it is not possible to have a unique prescription to set the optimal Δl to get an acceptable convergence, since it depends on the details of data such as the face spins j_f and the Immirzi parameter. Furthermore, the convergence strongly depends on the structure of the 2-complex, and there is still no method that allows to estimate the error made in truncating the sum over the auxiliary spins. In section 5.1, I describe a numerical property of the convergence that allows extrapolating the limit $\Delta l \to \infty$ for the self-energy.

4.2. *EPRL self-energy amplitude*

According the procedure described above, starting from the BF amplitude (5), I attach a booster function between the bulk intertwiners, choosing the boundary intertwiners $i \equiv (i_+.i_-)$ as the gauge fixed ones in the two vertex amplitudes which form the self-energy spinfoam, which is the most convenient choice from a computational point of view. Here I consider the case[1] in which the two boundary intertwiners have the same value. When this happens, the fully algebraic compact

[e] The nature of this cut-off is obviously completely different from the K cutoff on the internal faces of self-energy amplitude introduced in section 3.1 for the BF amplitude.

expression of amplitude becomes[f]:

$$W(j,i,\gamma;K,\Delta l)=\sum_{j_{ab}}^{K}\prod_{(a,b)}(2j_{ab}+1)\sum_{i_e}A_v(j,j_{ab},i,i_e,\gamma;\Delta l)^2 \tag{9}$$

where $q = 2,...5$, and the EPRL vertex amplitude is defined in (7). The EPRL self-energy amplitude (9) is the main object of our analysis. The dependence on the Immirzi parameter γ is "hidden" inside the booster functions, which appear in the definition of the single vertex amplitude (7). In the EPRL self-energy amplitude (9) I explicitly emphasized the *artificial* dependence on the two cut-off K and Δl. I shall always consider the two boundary intertinwers $i \equiv (i_+.i_-)$ to have the same value. When this is the case, the asymptotic divergence scaling remains essentially[g] unchanged by modifying the value of the intertwiners. Since it is necessary to choose a specific value of i for the computations, in the following I use $i = (0, 0)$ by default. This completes the derivation of amplitude we considered in,[1] and now I describe the numerical results obtained.

5. Divergence analysis

The amplitude (9) has been studied numerically. Recalling that the analysis must be limited to a small number of parameters' configurations, the goal of the latter consists in answering the following question (for further numerical studies I refer again to,[1] which addresses more questions), for which there are currently no analytical methods of investigation: what is the exact asymptotic scaling of the EPRL divergence, and is there a dependence on the Immirzi parameter?

The logical step is to proceed in the same way as we did with the BF amplitude (5), that is, fix the boundary state entirely to the lowest value $j = 0.5$, and compute the amplitude as a function of the cut-off K on the internal faces. It is reasonable to do so using several different values of the Immirzi parameter γ. However, compared to the purely $SU(2)$ case, the EPRL amplitude (9) has an additional cut-off Δl necessary to truncate the sum on the auxiliary spins l_f in the single vertex amplitudes (7) which constitute self-energy. This fact makes the EPRL analysis much more complicated due to the reasons explained in section 4. In section 5.1 I present a property of the convergence which allows to extrapolate the limit $\Delta l \to \infty$. Since for computations it is necessary to select a specific value of the Immirzi parameter, I first illustrate the divergence scaling extrapolation algorithm by selecting a specific value for the latter. The analysis for different γ values is similar.

The result of the computation of the amplitude (9), for increasing values of Δl and K, is shown in Figure 4. From a simple qualitative analysis we infer that the convergence is faster for a reduced bulk spins cutoff, while it becomes slower

[f] In the following omit the EPRL subscript for the amplitude, since every time I write W (without the BF subscript) I always implicitly refer to the EPRL model.
[g] A part from irrelevant numerical fluctuations.

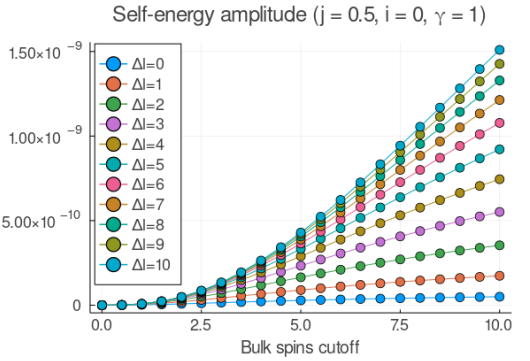


Fig. 4. Divergence of the EPRL self-energy amplitude (9) computed numerically. All boundary spins $j \equiv j_f$, where $f = 1...4$, are equal to $\frac{1}{2}$, while boundary intertwiners i_+, i_- are both set to zero. The plots with i_+, i_- equal to 1 are identical.

as K increases. Strictly speaking, the numerically computed curves turn out to be a *lower bound* to the EPRL divergence, since the latter is recovered only in the limit $\Delta l \to \infty$. Since further increasing numerically the parameter Δl requires exponentially increasing the computation times, it is necessary to introduce an algorithm for extrapolating the above limit.

5.1. *Extrapolation algorithm for the amplitude*

In order to derive the full EPRL amplitude, we study the convergence of (9) in the parameter Δl based on the data of Figure 4, and we extrapolate infinite shell limit based on this trend. In order to do so, we plot the ratios of the differences between adjacent curves of Figure 4, at fixed K, as a function of Δl. That is, we study the function:

$$f(K, \Delta l, \gamma) \equiv \frac{W(K, \Delta l + 2, \gamma) - W(K, \Delta l + 1, \gamma)}{W(K, \Delta l + 1, \gamma) - W(K, \Delta l, \gamma)}. \tag{10}$$

The result is shown in Figure 5. Numerical data shows that the convergence of (9) in the shell parameter Δl, for each bulk spins cut-off K, is such that the function (10) is first a decreasing function of Δl, then it remains constant in the convergence phase. Therefore, it exists an integer N such that:

$$W(K, \Delta l + 1, \gamma) - W(K, \Delta l, \gamma) \approx (c_{K,\gamma})^{\Delta l} \quad \text{for} \quad \Delta l \geq N \tag{11}$$

where the coefficient $c_{K,\gamma}$ is roughly equal to function (10) for the highest number of shells computed numerically. So, if the amplitude has been calculated up to N shells, we approximate the coefficient $c_{K,\gamma}$ as:

$$c_{K,\gamma} \equiv f\left(K, N - 2, \gamma\right). \tag{12}$$

For the extrapolation of the EPRL amplitude we used $N = 10$, even if Figure 5 shows that property (11) becomes evident even for lower values. Once the convergence is

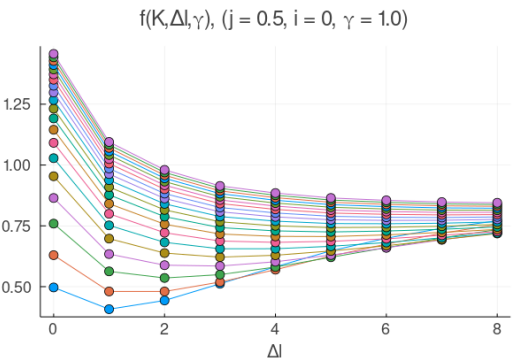


Fig. 5. Plot of the function $f(K, \Delta l, \gamma)$ defined as in (10). I plot the values of K between $\frac{1}{2}$ and 10 in ascending order from bottom to top.

reached, that is, when $\frac{W(K,\Delta l+1,\gamma)}{W(K,\Delta l,\gamma)} \approx 1$, function (10) shows some slight fluctuations by further increasing the shells[h]. This is evident for low bulk spins cut-off values, where convergence in Δl is extremely fast, as shown in Figure 5. Property (11) allows to obtain a good estimate of the EPRL curve by using the equation:

$$W(K,\gamma) \equiv \lim_{\Delta l \to \infty} W(K, \Delta l, \gamma) \approx W(K, N-1, \gamma) + \frac{W(K,N,\gamma) - W(K,N-1,\gamma)}{1 - c_{K,\gamma}} \tag{13}$$

Equation (13) is obtained by using the known limit of the geometric series:

$$\sum_{l=N}^{\infty} (c_{K,\gamma})^l = \frac{(c_{K,\gamma})^N}{1 - c_{K,\gamma}} \approx \frac{W(K,N,\gamma) - W(K,N-1,\gamma)}{1 - c_{K,\gamma}} \tag{14}$$

where in the second passage we used the property (11). In Figure 7 we plot the EPRL amplitude extrapolated with equation (13) along with the curves in Figure 4.

Let's discuss the above extrapolation scheme:

- Extrapolating the limit $\Delta l \to \infty$ of (9) for low values of the bulk spins cut-off K, for which a good convergence in the auxiliary spin sum has already been reached numerically, equation (13) provides a value which essentially coincides with the last computed point. This happens because, despite the fluctuations of function (10) for low K, the difference $W(K,N,\gamma) - W(K,N-1,\gamma)$ is extremely small. Therefore, the only relevant contribution to the extrapolated amplitude comes from the computed one, which is a good approximation of the EPRL model.
- For each bulk spins cut-off value K, the corresponding EPRL amplitude is extrapolated *independently* from the other values the latter. By doing so,

[h] As we shall see, this fluctuations are not relevant.

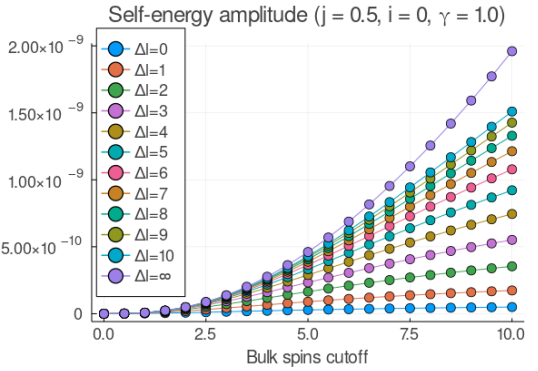


Fig. 6. EPRL asymptotic scaling curve extrapolated with the equation (13), plotted together with those computed numerically, that is, the curves of Figure 4. Notice that for $K \leq 4$, where the convergence is reached numerically, the extrapolated point essentially coincides with the last computed one.

we actually see that the resulting amplitude can be approximated extremely well by a polynomial fit $W(K,\gamma) = a + bK^c$ as a function of the bulk spins cut-off K.

- The specific properties of the convergence in the auxiliary spins sum of the single vertex amplitude (7) are still unexplored. It is possible that property (11) also manifests for spinfoam amplitudes defined on a more elaborate 2-complex with respect to the triangulation described in section 2. If this would be the case, the above extrapolation scheme could allow to obtain a good estimate of the EPRL model by using a negligible amount of computational resources compared to that necessary to reach a good approximation by using solely numerical techniques.
- In order to apply equation (13) it is necessary to know $W(K,N,\gamma)$, $W(K,N-1,\gamma)$, $W(K,N-2,\gamma)$. That is, only three amplitudes must be computed numerically (for N sufficiently high). Despite this, we still opted to compute all the amplitudes $W(K,\Delta l,\gamma)$ for $\Delta l = 0,1...N$. This was done both to test property (11) and, on the other hand, to estimate the effectiveness of the extrapolation with a qualitative comparison between all the numerically calculated amplitudes and the extrapolated one. Furthermore, since the curves in Figure 5 are not exactly constant, that is, they start to increase when $\frac{W(K,\Delta l+1,\gamma)}{W(K,\Delta l,\gamma)} \approx 1$, the extrapolated curve is an approximation, which improves by increasing N in equation (13). Therefore, the higher is the number of shells computed numerically, the better is the approximation represented by the extrapolated amplitude. In[1] we discuss this in more detail and we also provide further numerical confirmation of the effectiveness of this extrapolation algorithm.

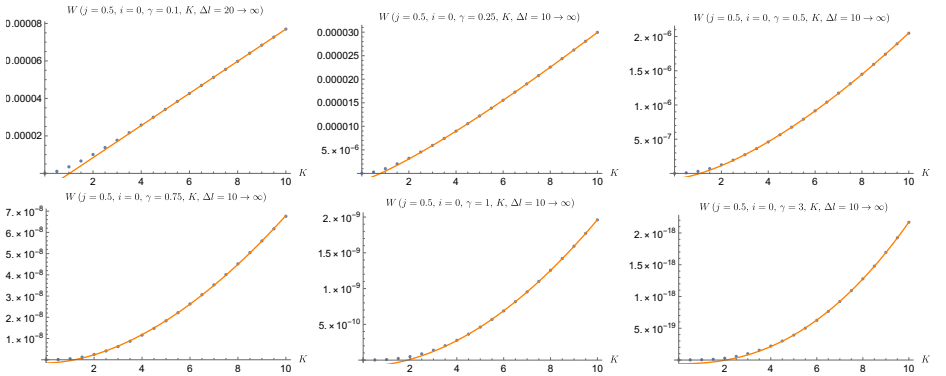


Fig. 7. EPRL asymptotic scaling curve extrapolated with the equation (13), plotted together with those computed numerically, that is, the curves of Figure 4. Notice that for $K \leq 4$, where the convergence is reached numerically, the extrapolated point essentially coincides with the last computed one.

5.2. *Scaling of the divergence and role of the Immirzi parameter*

Using the above scheme in,[1] we study the scaling of the asymptotic divergence of the EPRL self-energy amplitude for 9 different values of the Immirzi parameter. We choose an approximately uniform sampling of γ between 0.1 and 10, arguing that a significant effect of the latter in the scaling should occur in a range which spans two orders of magnitude. We fit the curves with a function:

$$W(K,\gamma) = a + bK^c \tag{15}$$

where a, b, c are real coefficients. The values[i] of $W(K,\gamma)$ are shown in table 1.

Table 1. Fit coefficients table

$W(K,\gamma) = a + bK^c \quad (j = 0.5, i = 0),\ K \in [0,10]$			
Immirzi parameter	a	b	c
$\gamma = 0.1$	$-3.0884 * 10^{-6}$	$6.0304 * 10^{-6}$	1.1
$\gamma = 0.25$	$-1.58614 * 10^{-6}$	$1.9783 * 10^{-6}$	1.2
$\gamma = 0.5$	$-7.5523 * 10^{-8}$	$6.7306 * 10^{-8}$	1.5
$\gamma = 0.75$	$-1.7153 * 10^{-9}$	$1.10779 * 10^{-9}$	1.8
$\gamma = 1$	$-8.6894 * 10^{-11}$	$2.5801 * 10^{-11}$	1.9
$\gamma = 3$	$-6.0026 * 10^{-20}$	$1.1186 * 10^{-20}$	2.3
$\gamma = 5$	$-4.5405 * 10^{-25}$	$1.0642 * 10^{-25}$	2.3
$\gamma = 7.5$	$-2.35075 * 10^{-29}$	$8.25374 * 10^{-30}$	2.3
$\gamma = 10$	$-1.7872 * 10^{-32}$	$9.2770 * 10^{-33}$	2.3

[i]Recall that we neglected all the dimensional factors which, in the amplitude (9), are constant as a function of K.

The analysis of the scaling shows that the amplitude is strongly suppressed as γ increases. Moreover, the divergence turns out to be well fitted by a linear scaling when the Immirzi parameter is small enough, while for $\gamma \geq 1$ the curve is approximated by a quadratic function in the range $K \in [0, 10]$. Unfortunately, the huge computation cost required by increasing the bulk spins cut-off in the divergence analysis prevents us from testing the divergence for $K > 10$. It is possible that the divergence scaling is actually independent of the Immirzi parameter, and that the different asymptotic trends, as γ varies, are only an effect of the fact that the range of the bulk spins cut-off K is too small. On the other hand, it is also possible that the Immirzi parameter plays an effective role in modifying the asymptotic scaling of self-energy amplitude. This deserves future investigations.

In any case, the numerically observed scaling in the range $K \in [0, 10]$, with boundary parameters $j = 0.5, i = 0$ falls within the upper and lower bounds present in the literature. In particular, unlike the BF self-energy divergence (see equation (3)), the role of the destructive interference between the oscillations of the booster functions $B_4(j_f, l_f; i, k)$ and the $\{15j\}$ symbols in the sum over the bulk spins j_{ab} implies that the divergence is considerably dumped. This was not expected, since it has been shown that in the three-dimensional EPRL model, the upper bound provided by the algorithm proposed in[30] provides an excellent estimate of the divergence.

6. Boundary observables

In this section we compute some spinfoam boundary observables. We focus on the normalized dynamical expectation value of geometrical operators, that is:

$$\langle O \rangle = \frac{\langle W|O|\Psi\rangle}{\langle W|\Psi\rangle} \tag{16}$$

where the bra W contains the propagator, namely the dynamics, while the ket Ψ turns out to be the tensor product of the *in* and *out* states of the LQG Hilbert space. With the term "propagator" we refer to the square matrices (they are such since the self-energy triangulation has two boundary tetrahedra) in which the element a, b corresponds to the EPRL self-energy amplitude with $a = i_+, b = i_-$. The observable O therefore contains the dynamic correlations.

The booster function $B_4(j_f, l_f; i, k)$ are interpreted as a quantum tetrahedron being boosted among adjacent frames: the two sets j_f and l_f describe the four areas of the tetrahedron in the two frames connected by a boost, and the two intertwiners i and k describe the quantum intrinsic shape of the tetrahedron.[48] Therefore, the expectation values (16) of geometric operators describing boundary tetrahedra should be equivalent with respect to the $SU(2)$ model. From a numerical point of view it is an excellent check, especially considering that numerical calculations carried out with the EPRL model are still in their primordial stages,[30, 43, 47] and, as far as we know, there are no such numerical computations (ie dynamic expectation value of

geometric operators) with spinfoams more complicated than those consisting of a single vertex amplitude[49] which used these specific techniques.

In[1] we find that the boundary observables are not affected in any way by the presence of the spinfoam divergence in the bulk. Furthermore, while in the divergence analysis we see an important contribution due to the number of shells Δl of the amplitude (9), for the observables there is no trace of a dependence in that sense[j]. In fact, an excellent approximation of the correct geometric value is obtained with the approximation $\Delta l = 0$. The fact that the number of shells is completely irrelevant in the computation of geometric boundary observables allows them to be computed very fast.

6.1. *Angles*

The shape of the tetrahedra in twisted geometry is measured by the angle operator:

$$A_{ab}|i_\pm\rangle = \cos(\theta_{ab})|i_\pm\rangle \tag{17}$$

which is interpreted as the cosine of the external dihedral angle between the faces a and b of the tetrahedron defined on the nodes $\pm$. The spinnetwork basis states diagonalize the dihedral angle θ_{ab} between faces a and b. The equation for measuring the dihedral angle $\cos(\theta_{ab})$ of $|i_\pm\rangle$ in terms of intertwiner spin $i_\pm$ was derived in[49] and it reads:

$$\cos(\theta_{ab}) = \frac{i_+(i_+ + 1) - j_a(j_a + 1) - j_b(j_b + 1)}{2\sqrt{j_a(j_a + 1)j_b(j_b + 1)}}. \tag{18}$$

We consider the expectation value (16) of the angle operator (17) in any of the two (equal) boundary regular tetrahedra of the triangulation using the spinnetwork state. According to the recoupling basis (j_1, j_2), we focus on the angle between faces 1 and 2. The expectation value can be computed as:

$$\frac{\langle W|A_{12}|W\rangle}{\langle W|W\rangle} = \frac{\sum\limits_{i_\pm} [W(j, i, K)]^2 \cos(\theta_{12})}{\sum\limits_{i_\pm} [W(j, i, K)]^2}. \tag{19}$$

Carrying out the numerical computation of (19), we obtain a value that is in agreement with the geometric value of the external angle of a regular tetrahedron up to the tenth significant digit, as shown in Figure 8. This is consistent with the fact that we are looking at the only angle which is completely sharp, while the others turn out to be spread. In fact, according to the Heisenberg uncertainty principle, since different angle operators do not commute, only one of the dihedral angles can be determined. We also analyzed the boundary states of the self-energy amplitude by using the Livine-Speziale coherent intertwiners, which allow to define a superposition of spin-network states peaked on a classical geometry with minimal

[j] At least for the ones that we computed.

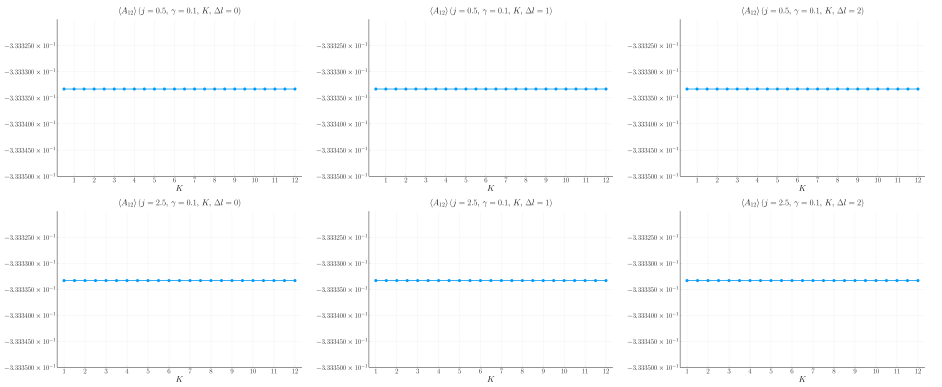


Fig. 8. Dynamic expectation value of the cosine of the external dihedral angle operator (19). We plot the angle for different values of Δl, showing that an excellent agreement with the value $-0.\bar{3}$ is obtained with the approximation $\Delta l = 0$. We used $\gamma = 0.1$.

spread. A wave packet peaked on a classical triangulated geometry, that can be viewed as a coherent state in the Hilbert space (1), is obtained by combining coherent intertwiners at each node. In this case, all the dihedral angles are minimally spread around the classical values, and this is exactly what we verify numerically. The plots obtained are qualitatively identical to those in Figure 8, except that the computed value of the angle is approximately[k] that of the external dihedral angle of a regular tetrahedron. This is due to the fact that the square modulus of the coefficients of the coherent states, in the spin-intertwiner basis, turns out to be a a distribution centered around the value of the intertwiner which determines the semiclassical value of the dihedral angle, and the width of the distribution is much higher as the spins are low. For the computation of the volumes, we refer to.[1]

7. Conclusions

I have presented the application[1] of new computational techniques applied to the study of the infrared divergence represented by the self-energy EPRL amplitude. The divergence scaling obtained falls within the upper and lower bounds in the literature. Particular emphasis was placed on the role of the Immirzi parameter in the asymptotic divergence. I presented the extrapolation method introduced to overcome the computational cost represented by the convergence in the shell parameter Δl, which can potentially be used in other contexts. The analysis of the boundary angles shows that the latter are finite and consistent with classical geometry, despite the divergence in the bulk, and that the approximation represented by the shells does not play any relevant role in the dynamical expectation values of the latter.

[k] Up to the first 2 significant digits.

We thank all the collaborators of this work, in particular Francesco Gozzini, without whom this would not have been possible, as well as Carlo Rovelli for useful discussions and comments. I explicitly thank my supervisor Francesca Vidotto for the constant support and for an accurate review in the drafting of this paper. I also thank Pietro Dona for countless contributions in various stages of.[1] Finally, we also thanks the Compute Canada staff for their help in using the Graham and Cedar clusters. This work is supported by the NSERC Discovery Grant.

References

1. P. Frisoni and Others, Numerical analysis of the EPRL spin foam self-energy, *in preparation* (2021).
2. A. Perez, The Spin-Foam Approach to Quantum Gravity, *Living Reviews in Relativity* **16** (2013).
3. C. Rovelli, D. Colosi, L. Doplicher, W. Fairbairn, L. Modesto and K. Noui, Background independence in a nutshell, *arXiv.org* **gr-qc** (2004).
4. C. Rovelli and M. Smerlak, In quantum gravity, summing is refining, *Class. Quant. Grav.* **29**, p. 55004 (2012).
5. R. Pereira, Lorentzian LQG vertex amplitude, *Class. Quant. Grav.* **25**, p. 85013 (2008).
6. J. Engle, R. Pereira and C. Rovelli, Flipped spinfoam vertex and loop gravity, *arXiv.org* **gr-qc** (2007).
7. L. Freidel, Reconstructing AdS/CFT (2008).
8. Y. Ding, M. Han and C. Rovelli, Generalized Spinfoams (2010).
9. W. Kaminski, M. Kisielowski and J. Lewandowski, Spin-Foams for All Loop Quantum Gravity, *Class. Quant. Grav.* **27**, p. 95006 (2010).
10. J. Engle, E. Livine, R. Pereira and C. Rovelli, LQG vertex with finite Immirzi parameter, *Nucl. Phys.* **B799**, 136 (2008).
11. E. R. Livine and S. Speziale, A new spinfoam vertex for quantum gravity, *Phys. Rev.* **D76**, p. 84028 (2007).
12. E. R. Livine and S. Speziale, Consistently Solving the Simplicity Constraints for Spinfoam Quantum Gravity, *Europhys. Lett.* **81**, p. 50004 (2008).
13. L. Freidel and K. Krasnov, A New Spin Foam Model for 4d Gravity, *Class. Quant. Grav.* **25**, p. 125018 (2008).
14. C. Rovelli and F. Vidotto, *Covariant Loop Quantum Gravity* (Cambridge University Press, Cambridge, 2015).
15. C. Rovelli, Zakopane lectures on loop gravity, *arXiv.org* **gr-qc** (2011).
16. P. Doná and S. Speziale, Introductory lectures to loop quantum gravity, in *TVC 79. Gravitation: théorie et expérience*, eds. A. Bounames and A. Makhlouf (Hermann, 2013).
17. B. Dittrich, The continuum limit of loop quantum gravity - A framework for solving the theory (2016).
18. B. Dittrich, S. Mizera and S. Steinhaus, Decorated tensor network renormalization for lattice gauge theories and spin foam models, *New Journal of Physics* **18**, p. 053009 (May 2016).
19. B. Bahr and S. Steinhaus, Investigation of the spinfoam path integral with quantum cuboid intertwiners, *Physical Review D* **93** (May 2016).
20. V. Bonzom, R. Gurau and V. Rivasseau, Random tensor models in the large N limit: Uncoloring the colored tensor models, *Physical Review D* **85** (Apr 2012).

21. D. Benedetti and R. Gurau, Phase transition in dually weighted colored tensor models, *Nuclear Physics B* **855**, p. 420–437 (Feb 2012).
22. S. Carrozza, D. Oriti and V. Rivasseau, Renormalization of a su(2) tensorial group field theory in three dimensions, *Communications in Mathematical Physics* **330**, p. 581–637 (Mar 2014).
23. J. B. Geloun, T. A. Koslowski, D. Oriti and A. D. Pereira, Functional renormalization group analysis of rank-3 tensorial group field theory: The full quartic invariant truncation, *Physical Review D* **97** (Jun 2018).
24. V. Bonzom and M. Smerlak, Bubble divergences from twisted cohomology (2010).
25. V. Bonzom and M. Smerlak, Bubble divergences: sorting out topology from cell structure (2011).
26. A. Baratin, S. Carrozza, D. Oriti, J. Ryan and M. Smerlak, Melonic phase transition in group field theory, *Letters in Mathematical Physics* **104**, p. 1003–1017 (May 2014).
27. C. Perini, C. Rovelli and S. Speziale, Self-energy and vertex radiative corrections in LQG, *Phys. Lett.* **B682**, 78 (2009).
28. T. Krajewski, J. Magnen, V. Rivasseau, A. Tanasa and P. Vitale, Quantum Corrections in the Group Field Theory Formulation of the EPRL/FK Models, *Phys. Rev.* **D82**, p. 124069 (2010).
29. A. Riello, Self-Energy of the Lorentzian EPRL-FK Spin Foam Model of Quantum Gravity, *Phys. Rev.* **D88**, p. 24011 (2013).
30. P. Donà, Infrared divergences in the EPRL-FK Spin Foam model, *Class. Quant. Grav.* **35**, p. 175019 (2018).
31. J. W. Barrett, R. J. Dowdall, W. J. Fairbairn, F. Hellmann and R. Pereira, Lorentzian spin foam amplitudes: Graphical calculus and asymptotics, *Classical and Quantum Gravity* **27** (2010).
32. F. Gozzini, A high-performance code for eprl spin foam amplitudes (2021), *In preparation.*
33. J. Bezanson, A. Edelman, S. Karpinski and V. B. Shah, Julia: A fresh approach to numerical computing, *SIAM review* **59**, 65 (2017).
34. W. R. Inc., Mathematica, Version 12.2 Champaign, IL, 2020.
35. E. Bianchi and Y. Ding, Lorentzian spinfoam propagator, *Phys. Rev.* **D86**, p. 104040 (2012).
36. E. Bianchi and H. M. Haggard, Discreteness of the volume of space from Bohr-Sommerfeld quantization (2011).
37. L. Freidel and S. Speziale, Twisted geometries: A geometric parametrisation of SU(2) phase space, *Phys. Rev.* **D82**, p. 84040 (2010).
38. S. Speziale, Boosting Wigner's nj-symbols, *Journal of Mathematical Physics* **58** (sep 2017).
39. A. P. Yutsin, I. B. Levinson and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translation, Jerusalem, Israel, 1962).
40. V. D. Aleksandroviic, A. N. Moskalev and K. V. Kel'manoviic, *Quantum theory of angular momentum: Irreducible tensors, spherical harmonics, vector coupling coefficients, 3nj symbols* (World Scientific, 1988).
41. J. Engle, R. Pereira and C. Rovelli, The loop-quantum-gravity vertex-amplitude, *Phys. Rev. Lett.* **99**, p. 161301 (2007).
42. J. Engle and R. Pereira, Regularization and finiteness of the Lorentzian LQG vertices, *Phys. Rev.* **D79**, p. 84034 (2009).
43. P. Donà, M. Fanizza, G. Sarno and S. Speziale, Numerical study of the Lorentzian Engle-Pereira-Rovelli-Livine spin foam amplitude, *Physical Review D* **100** (Mar 2019).

44. E. Bianchi, P. Donà and S. Speziale, Polyhedra in loop quantum gravity, *Physical Review D* **83** (2011).
45. P. Donà, M. Fanizza, G. Sarno and S. Speziale, Su(2) graph invariants, Regge actions and polytopes, *Classical and Quantum Gravity* **35** (2018).
46. P. Donà and G. Sarno, Numerical methods for EPRL spin foam transition amplitudes and Lorentzian recoupling theory, *General Relativity and Gravitation* **50** (2018).
47. P. Donà, M. Fanizza, G. Sarno and S. Speziale, Numerical study of the lorentzian engle-pereira-rovelli-livine spin foam amplitude, *Physical Review D* **100** (2019).
48. P. Dona and S. Speziale, Asymptotics of lowest unitary SL(2,C) invariants on graphs (Jul 2020).
49. F. Gozzini and F. Vidotto, Primordial fluctuations from quantum gravity (2019).

A spin foam framework for the black-to-white hole transition

Farshid Soltani

Department of Physics and Astronomy, The University of Western Ontario, London, Ontario N6A 3K7, Canada
E-mail: fsoltan2@uwo.ca

Black holes formation and evolution have been extensively studied at the classical level. However, not much is known regarding the end of their lives, a phase that requires to consider the quantum nature of the gravitational field. A black-to-white hole transition can capture the physics of this phenomenon, in particular the physics of the residual small black holes at the end of the Hawking evaporation. This work shows how the spin foam formalism is able to describe this non-perturbative phenomenon. A thorough examination of the black hole spacetime region in which quantum effects cannot be neglected indicates that the scenario in which the black hole geometry undergoes a quantum transition in a white hole geometry is natural and conservative. This quantum transition is then studied using the spin foam formalism and the resulting transition amplitude is explicitly computed.

Keywords: Black holes, white holes, black-to-white hole transition, loop quantum gravity, spin foam formalism.

1. Introduction

The exterior region of a black hole is described extraordinarily well by general relativity. Its interior region, on the other hand, is not and it is thus not known what happens inside a black hole. The reason for this breakdown of predictability is the presence of a spacetime singularity in the interior of a black hole: since the quantum nature of the gravitational field cannot be neglected in the vicinity of a spacetime singularity, the classical theory cannot be trusted in this region. Although effective black hole models exhibiting a non-singular interior have been extensively studied in the literature, a non-singular black hole interior consisting of a black hole geometry that undergoes a quantum transition in a white hole geometry was proposed for the first time in Ref. 1.

A different open question concerning black holes is what happens at the end of their life. Working in semiclassical gravity, Hawking[2] famously showed that black holes evaporate and are thus not eternal. However, since the quantum nature of the gravitational field near the horizon can no longer be neglected when the horizon reaches Planckian size (or possibly even before then[1]), the end of the evaporation process of a black hole is a quantum gravity phenomenon. A very natural and conservative scenario for the end of the life of a black hole is the black-to-white hole transition.

The aim of this work is to report in a concise and coherent fashion the results obtained in Refs. 3 and 4. The black hole spacetime region in which quantum effects cannot be neglected is analyzed in detail and it is shown to be actually composed of three physically independent subregions, with one of them being a region surrounding the black hole horizon at the end of the evaporation process. As a consequence, the last stage of the life of a black hole can be studied focusing solely on this region. Independent analyses of the three separate quantum regions consistently point toward a black-to-white hole transition. The last stage of the life of a black hole in this scenario, that is the quantum region where the black hole horizon undergoes a quantum transition in a white hole horizon, is then studied using the spin foam formalism[5,6] (also known as covariant loop quantum gravity) and a concrete spin foam framework for the black-to-white hole transition is developed.

The discussion is here limited to the case of a Schwarzschild black hole. See however Ref. 7 for a generalization of the black-to-white hole geometry to the case of a charged black hole.

2. The three quantum regions of a black hole spacetime

The conformal diagram of the spacetime describing the formation of a black hole by gravitationally collapsed matter and its subsequent evaporation is reported in Fig. 1. The light grey region represents the interior of the collapsing matter, the dashed line represents the apparent horizon of the black hole and the dark grey region represents the spacetime region where the quantum nature of the gravitational field cannot be neglected and where consequently the classical (or semiclassical) theory can no longer be trusted.

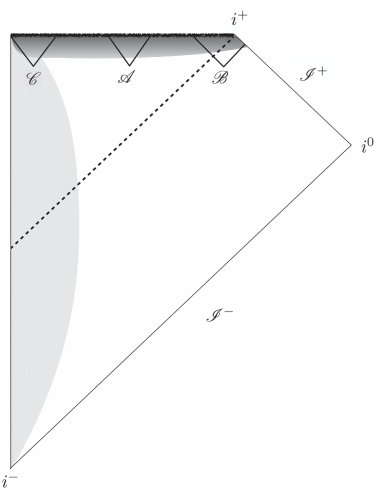


Fig. 1. Conformal diagram of the spacetime describing the formation of a black hole by gravitationally collapsed matter and its subsequent evaporation.

The quantum region can be divided[8] in three different subregions: region $\mathscr{C}$, where the collapsing matter reaches Planckian density; region $\mathscr{B}$, where the horizon reaches Planckian size at the end of the evaporation process; region $\mathscr{A}$, where the curvature reaches Planckian scale but the spacetime events belong neither to region $\mathscr{B}$ nor to region $\mathscr{C}$. In order to prove that these subregions are physically independent from each other, it is sufficient to show that their spatial separation is considerable. In the classical theory the principle of locality assures that two events whose separation is spacelike and significant cannot be causally connected. Furthermore, although the quantum theory may alter the causal structure of spacetime, quantum fluctuations cannot causally connect widely spacelike-separated events.

A rough estimate of the distance $d_{\mathscr{B}C}$ between regions $\mathscr{B}$ and $\mathscr{C}$ is[3]

$$d_{\mathscr{B}C} \sim \ell_{\mathrm{Pl}} \left(\frac{m_0}{m_{\mathrm{Pl}}}\right)^{\frac{10}{3}}, \tag{1}$$

where m_0 is the initial mass of the black hole before the start of the evaporation process and $\ell_{\mathrm{P}l}$ and $m_{\mathrm{P}l}$ are respectively the Planck length and the Planck mass. For a stellar black hole this distance is

$$d_{\mathscr{B}C}\big(m_0 = M_\odot\big) \sim 10^{75} \text{ light years}. \tag{2}$$

Regions $\mathscr{B}$ and $\mathscr{C}$ are thus spacelike separated and considerably distant from each other. It follows that regions $\mathscr{B}$ and $\mathscr{C}$ are physically independent from each other and they can thus be studied separately. Furthermore, since region $\mathscr{A}$ contains spacetime events that are physically independent from region $\mathscr{B}$ as well as spacetime events that are physically independent from region $\mathscr{C}$, the physics of region $\mathscr{A}$ must be independent from the physics of both region $\mathscr{B}$ and region $\mathscr{C}$. However, since region $\mathscr{A}$ contains spacetime events that are causally connected to region $\mathscr{B}$ as well as spacetime events that are causally connected to region $\mathscr{C}$, the physics of both region $\mathscr{B}$ and region $\mathscr{C}$ depend on the physics of region $\mathscr{A}$. These three quantum regions will be now analyzed separately.

2.1. *Region* $\mathscr{A}$

Since the physics of region $\mathscr{A}$ does not depend on the quantum gravity regime of the collapsing matter (region $\mathscr{C}$) or on the last stage of the life of the black hole (region $\mathscr{B}$), it can be studied in the context of an eternal (Schwarzschild) black hole.

The effort to understand the physics of region $\mathscr{A}$ has led to the development of several effective models (see Refs. 9–11 and references therein) that describe the internal region of an eternal black hole using techniques developed in the framework of loop quantum cosmology. Although the specifics of these models are different, they all exhibit a regular interior region where the trapped region of the black hole makes a smooth transition in an anti-trapped region bounded by a future horizon describing the interior of a white hole. This result, besides supporting the conjecture that classical curvature singularities are not a true physical prediction of the theory

but rather an indication that the classical theory can no longer be trusted, is the first evidence suggesting a black-to-white hole transition.

Interestingly, a hint of the same result can be found also at the classical level. It can indeed be shown[12,13] that, using specific coordinate systems, the geodesics in the interior region of a black hole can be naturally continued across the curvature singularity into the interior region of a white hole. The resulting geometry is still singular, but now it is geodesically complete and it can be interpreted as the classical limit of the quantum geometry of the effective models.

2.2. *Region* $\mathscr{C}$

The physics of region $\mathscr{C}$ is independent from the physics of region $\mathscr{B}$ and it can thus be studied neglecting the evaporation process of the black hole. However, it is not independent from the physics of region $\mathscr{A}$ and its analysis must be consistent with the scenario emerging from the investigation of region $\mathscr{A}$.

The study of the classical physics of the collapsing matter is a hard task. The analysis of its quantum gravity regime is even trickier. In Ref. 14 it was hypothesized that, in analogy with the cosmological singularity resolution in loop quantum cosmology,[15] the collapsing matter bounces due to a quantum-gravitational repulsion effect. This possibility, that will be assumed to accurately represent the physics of region $\mathscr{C}$ in the following, is consistent with the physics of region $\mathscr{A}$. A qualitative conformal diagram of the spacetime emerging from this partial analysis (region $\mathscr{B}$ still needs to be discussed) can be found in Fig. 2(a).

This scenario is further corroborated by several independent quantum descriptions of the phenomenon (see e.g. Refs. 16–19). Although these models use different techniques and focus on different aspects of the phenomenon, they all predict that the collapsing matter undergoes a bounce. This is a strong indication of the general validity of the scenario.

2.3. *Region* $\mathscr{B}$

The conformal diagram in Fig. 2(a) represents the black hole spacetime that emerges taking into account the quantum physics of regions $\mathscr{A}$ and $\mathscr{C}$. It is however immediate to see that this spacetime does not represent properly the physics of region $\mathscr{B}$. While the black hole evaporation process takes a finite amount of time to shrink the horizon to the Planck scale, in Fig. 2(a) region $\mathscr{B}$ is reached only asymptotically. The physics of region $\mathscr{B}$ thus need to be properly modified whilst remaining consistent with the scenario emerging from regions $\mathscr{A}$ and $\mathscr{C}$.

A scenario that is often considered for the end of the life of a black hole is the complete evaporation of the black hole. Having spent time investigating regions $\mathscr{A}$ and $\mathscr{C}$ separately, it is now easy to see that a complete evaporation of the black hole, besides being an ad hoc assumption with no foundation in any quantum gravity model, is hardly consistent with the global picture of the spacetime in Fig. 2(a).

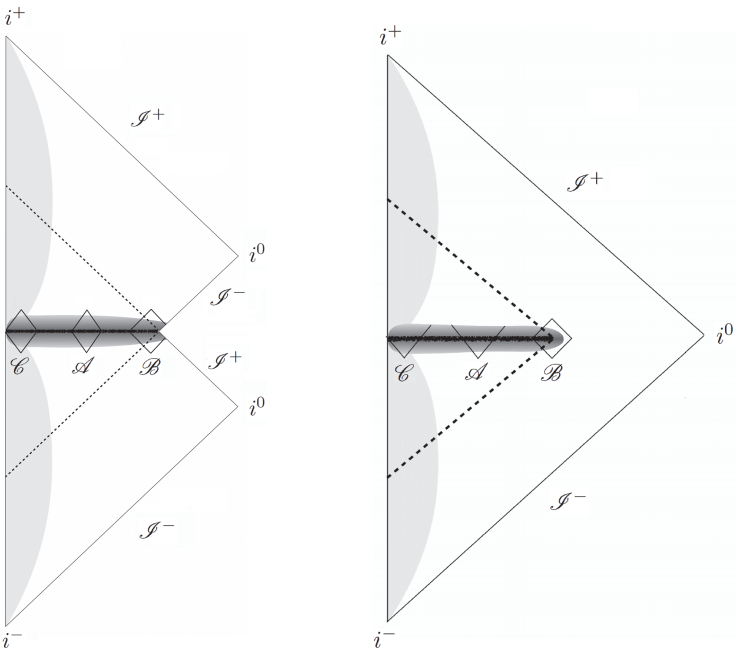


Fig. 2. (a) Qualitative conformal diagram of the black hole spacetime emerging from the analysis of regions $\mathscr{A}$ and $\mathscr{C}$. (b) Conformal diagram of the spacetime describing the black-to-white hole transition.

The most natural and conservative scenario for the end of the life of a black hole consistent with the physics of region $\mathscr{A}$ and region $\mathscr{C}$ is a quantum transition of the black hole horizon in a white hole horizon. The conformal diagram of the spacetime describing the complete black-to-white hole transition can be found in Fig. 2(b). The purpose of this work is to complete the analysis of region $\mathscr{B}$ by investigating its quantum physics using the spin foam approach.

The conformal diagram in Fig. 2(b) describes a physical phenomenon only if there exists a classical metric that satisfies the Einstein field equations and that covers the whole diagram except for the quantum region. This metric exists and it has been explicitly constructed in Refs. 1, 20, 21. This is an extraordinary result from the point of view of the classical theory and it provides yet another strong evidence in favor of the black-to-white hole transition scenario.

3. The black-to-white hole transition

In order to investigate in detail the physics of region $\mathscr{B}$ of the black-to-white hole spacetime it is necessary to specify its boundary $\Sigma := \partial\mathscr{B}$ and to compute the classical geometry that the black-to-white hole metric induces on it. This boundary

geometry represents the outcome of the quantum transition taking place in region $\mathscr{B}$ and it thus uniquely defines the quantum boundary state for the transition. The boundary Σ can be chosen freely as long as it bounds the entirety of the quantum subregion.

As a first approximation, the presence of the Hawking radiation near region $\mathscr{B}$ is neglected. Its inclusion is left for future work. The metric around region $\mathscr{B}$ is thus taken to be the Schwarzschild metric up to quantum corrections from region $\mathscr{A}$. The main quantum correction that the physics of region $\mathscr{A}$ induces on the boundary between regions $\mathscr{A}$ and $\mathscr{B}$ is the absence of the classical singularity. The black hole interior can be foliated with surfaces of topology $S^2 \times \mathbb{R}$. If one angular dimension is suppressed, these surfaces can be seen as long cylinders of different radii and heights. Closer is the singularity, smaller is the radius of the cylinder. In the classical theory the foliation ends at the singularity, where the cylinder has a null radius. In the quantum theory the cylinder radius shrinks (black hole geometry) until it reaches a minimum value r_* (smooth transition from black to white hole geometry) and then starts to increase again (white hole geometry). The presence of this minimum radius r_* in the effective geometry of region $\mathscr{A}$ has a significant impact on the physics of region $\mathscr{B}$.

Having neglected Hawking radiation, the physics of region $\mathscr{B}$ must be invariant under time-reversal. Accordingly, the boundary Σ can be decomposed as $\Sigma = \Sigma^p \cup \Sigma^f$, where the past surface Σ^p and the future surface Σ^f are equal up to time reflection. This means that to completely define Σ it is sufficient to only specify Σ^p. A convenient choice of Σ^p is reported in Fig. 3 and it can be constructed as follows. Let S_+ and S_- be the points in the conformal diagram of Schwarzschild spacetime defined by the ingoing Eddington-Finkelstein coordinates $S_+ = (v_+, r_+)$ and $S_- = (v_-, r_-)$. The point S_+ is taken to be on the surface of

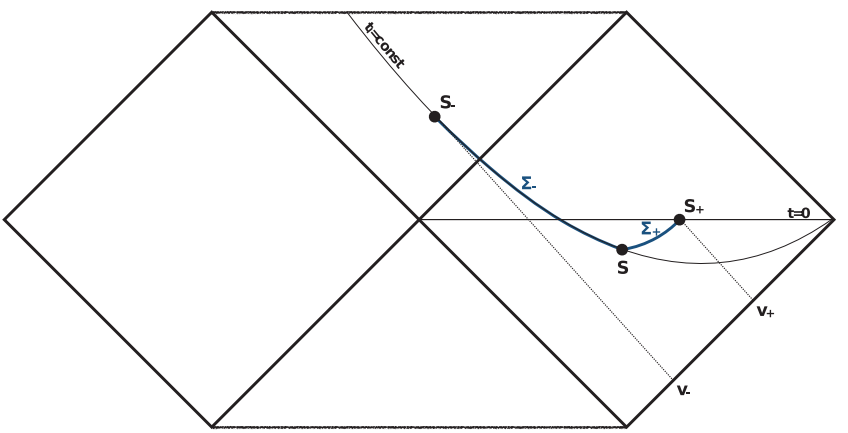


Fig. 3. The past portion $\Sigma^p = \Sigma^p_+ \cup \Sigma^p_-$ of the boundary Σ.

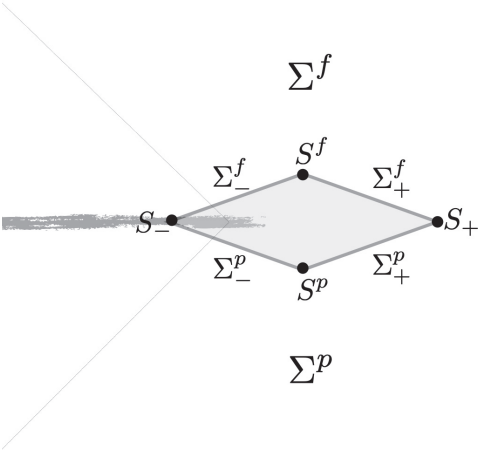


Fig. 4. Portion of the conformal diagram of the black-to-white hole transition depicting region $\mathscr{B}$ and its boundary Σ.

constant Schwarzschild time $t = 0$, thus fixing the value of v_+ in terms of r_+ (or viceversa), and the point S_- is taken to be on the surface of constant Schwarzschild radius $r = r_*$, thus fixing the value of r_- to be the minimum radius r_* characterizing the effective geometry of region $\mathscr{A}$. The values of r_+ and v_- can be chosen freely as long as they define a boundary Σ that bounds the entirety of the quantum subregion. Let then Σ^p_- be the surface of constant Lemaitre time coordinate

$$t_L := t + 2\sqrt{2mr} + 2m \log \left| \frac{\sqrt{r/2m} - 1}{\sqrt{r/2m} + 1} \right| \tag{3}$$

passing by S_-, where m is the mass of the black hole at the moment in which the quantum transition of the horizon takes place, and Σ^p_+ be the surface of equation

$$v - \beta\, r = \text{const}\,, \tag{4}$$

where β is an arbitrary constant in $\mathbb{R}$, passing by S_+. Given the point S^p of their intersection, the past boundary Σ^p is defined to be the union of the portion of the surface Σ^p_- bounded by S_- and S^p and the portion of the surface Σ^p_+ bounded by S_+ and S^p. Requiring the normal to Σ^p to be continuous at S^p uniquely fixes the value of β. The surface Σ^f is defined as the time-reversal of Σ^p. The portion of the conformal diagram of the black-to-white hole transition depicting region $\mathscr{B}$ and its boundary Σ can be found in Fig. 4.

The intrinsic and the extrinsic geometry of Σ^p_+ and Σ^p_- can be straightforwardly computed from their definition. The line element $\mathrm{d}s^2_+$ on Σ^p_+ reads

$$\mathrm{d}s^2_+ = \beta \left(2 - \beta\Big(1 - \frac{2m}{r}\Big)\right) \mathrm{d}r^2 + r^2\, \mathrm{d}\Omega^2 \tag{5}$$

and the line element $\mathrm{d}s^2_-$ on Σ^p_- reads

$$\mathrm{d}s^2_- = \mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2\,, \tag{6}$$

where $\mathrm{d}\Omega^2$ is the line element of the two-sphere. Let $k^\pm_{ij}$ be the extrinsic curvature tensor of $\Sigma^p_\pm$. Then it can be shown that

$$k_+ := k^+_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j = \frac{m\beta^{3/2}(r(3-\beta)+2m\beta)}{\sqrt{r^5(r(2-\beta)+2m\beta)}}\,\mathrm{d}r^2 - \frac{r(1-\beta)+2m\beta}{\sqrt{\beta(2-(1-2m/r)\beta)}}\,\mathrm{d}\Omega^2 \tag{7}$$

and

$$k_- := k^-_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j = \frac{m}{2r^3}\,\mathrm{d}r^2 - \sqrt{2mr}\,\mathrm{d}\Omega^2\,, \tag{8}$$

where x^i, $i = 1, 2, 3$, are coordinates on Σ^p.

The geometry of Σ is thus completely determined by four parameters: the mass m of the black hole at the moment in which the quantum transition of the horizon takes place; the external asymptotic (retarded) time $v = v_+ - v_-$ it takes for the transition to happen; the minimal external radius r_+ for which the classical theory can still be trusted; the minimal internal radius r_- (it is important to stress that r is a temporal coordinate in the interior region of the black hole) reached by the black hole interior in region $\mathscr{A}$.

Each set of data $\big(m,\, v,\, r_\pm\big)$ corresponds to a different outcome of the quantum transition taking place in region $\mathscr{B}$. This means that each set of data $\big(m,\, v,\, r_\pm\big)$ uniquely defines a different quantum boundary state $\Psi\big(m,\, v,\, r_\pm\big)$ for the transition. Given the boundary state Ψ, any sufficiently developed tentative theory of quantum gravity should be able to assign a transition amplitude $W\big(m,\, v,\, r_\pm\big) \equiv W\Big[\Psi\big(m,\, v,\, r_\pm\big)\Big]$ to it. The transition amplitude for the phenomenon can then be used to analyze the physics of region $\mathscr{B}$.

The next section is devoted to the computation of the transition amplitude $W\big(m,\, v,\, r_\pm\big)$ for the black-to-white hole transition using the spin foam approach.

4. Spin foam framework

The spin foam formalism[5,6] is a tentative path integral quantization of general relativity. The current state of the art is the EPRL-KKL spin foam model.[22–24] The theory is regularized and concretely defined in the discrete setting. The boundary Σ of a generic quantum region is discretized by a graph Γ with a finite number of nodes and a boundary Hilbert space $\mathcal{H}_\Gamma$ is assigned to it. The latter is the space of SU(2) spin-networks on Γ. Namely, the boundary Hilbert space is $\mathcal{H}_\Gamma := L^2\big[\mathrm{SU}(2)^L/\mathrm{SU}(2)^N\big]_\Gamma$, where N and L are respectively the total number of nodes and links in Γ. A boundary state is then given by a square integrable function $\psi(\{h_\ell\})$ that is gauge invariant at every node $\mathrm{n} \in \Gamma$. Each $h_\ell \in \mathrm{SU}(2)$ can be seen as the holonomy of the Ashtekar-Barbero connection along the link $\ell \in \Gamma$. The interior of the quantum region is then discretized by a two-complex $\mathcal{C}$ ($\partial\mathcal{C} \equiv \Gamma$) with a finite number of vertices.

Let f, e, v $\in \mathcal{C}$ denote respectively a face, an edge and a vertex in $\mathcal{C}$. To each internal oriented edge $\mathcal{C} \ni \mathrm{e} := (\mathrm{v}, \mathrm{v}')$ having source in v and target in v′ are assigned two $\mathrm{SL}(2,\mathbb{C})$ elements $g_{\mathrm{ve}} = g_{\mathrm{ev}}^{-1}$ and $g_{\mathrm{ev}'} = g_{\mathrm{v}'\mathrm{e}}^{-1}$. The oriented edge $\mathrm{e}^{-1} := (\mathrm{v}', \mathrm{v})$ is the edge e with opposite orientation. To each boundary edge $\mathrm{E} \in \mathcal{C}$, that is an edge linking an internal vertex $\mathrm{v} \in \mathcal{C}$ and a node $\mathrm{n} \in \Gamma$, is assigned an $\mathrm{SL}(2,\mathbb{C})$ element $g_{\mathrm{vn}} = g_{\mathrm{nv}}^{-1}$. A face $\mathrm{f} := (\mathrm{e}_1, \dots, \mathrm{e}_n)$ is the oriented face bounded by the edges $\mathrm{e}_1, \dots, \mathrm{e}_n$. The orientation of the face is given by the ordering of the edges. For easier reading, the latters are written oriented accordingly to the orientation that the face induces on them. An internal face $\mathrm{f} \in \mathcal{B} := \mathcal{C}/\Gamma$ is a face bounded by vertices and internal edges. A boundary face $\mathrm{F} \in \Gamma$ is a face containing a link $\ell \in \Gamma$ in its boundary. Finally, a face $\mathrm{f} \ni \mathrm{e}$ is a face f containing the edge e in its boundary, a face $\mathrm{f} \ni \mathrm{v}$ is a face f containing the vertex v in its boundary and an edge $\mathrm{e} \ni \mathrm{v}$ is an edge e containing the vertex v in its boundary.

Given an arbitrary boundary state $\psi \in \mathcal{H}_\Gamma$, the theory assigns to it the amplitude

$$W_{\mathcal{C}}[\psi] = \int_{\mathrm{SU}(2)} \prod_{\ell\in\Gamma} \mathrm{d}h_\ell \, W_{\mathcal{C}}\big(\{h_\ell\}\big)\, \psi\big(\{h_\ell\}\big)\,, \tag{9}$$

where the two-complex amplitude $W_{\mathcal{C}}(\{h_\ell\})$ can be defined in terms of elementary face amplitudes as

$$W_{\mathcal{C}}\big(\{h_\ell\}\big) = \int_{\mathrm{SL}(2,\mathbb{C})} \Big[\prod_{\mathrm{v}\in\mathcal{C}} \prod_{\mathrm{e}\ni\mathrm{v}}{}' \mathrm{d}g_{\mathrm{ve}}\Big] \Big[\prod_{\mathrm{f}\in\mathcal{B}} A_{\mathrm{f}}\big(\{g_{\mathrm{ve}}\}\big)\Big] \Big[\prod_{\mathrm{F}\in\Gamma} A_{\mathrm{F}}\big(\{g_{\mathrm{ve}}\}, h_{\ell_{\mathrm{F}}}\big)\Big]\,. \tag{10}$$

The product $\prod'_{\mathrm{e}\ni\mathrm{v}}$ stands for the product over all the edges $\mathrm{e} \ni \mathrm{v}$ except one (which can be chosen arbitrarily at each vertex), ℓ_{F} is the unique link in Γ that belongs to the boundary face $\mathrm{F} \in \Gamma$ and the face amplitudes $A_{\mathrm{f}}(\{g_{\mathrm{ve}}\})$ and $A_{\mathrm{F}}(\{g_{\mathrm{ve}}\}, h_{\ell_{\mathrm{F}}})$ are given by[a]

$$A_{\mathrm{f}}\big(\{g_{\mathrm{ve}}\}\big) := \sum_{j_{\mathrm{f}}} d_{j_{\mathrm{f}}} \mathrm{Tr}\Big[D_\gamma^{(j_{\mathrm{f}})}\big(g_{\mathrm{ev}} g_{\mathrm{ve}'}\big) \cdots D_\gamma^{(j_{\mathrm{f}})}\big(g_{\mathrm{e}^{(n)}\mathrm{v}^n} g_{\mathrm{v}^{(n)}\mathrm{e}}\big)\Big]\,, \tag{11}$$

$$A_{\mathrm{F}}\big(\{g_{\mathrm{ve}}\}, h_{\ell_{\mathrm{F}}}\big) := \sum_{j_{\mathrm{F}}} d_{j_{\mathrm{F}}} \mathrm{Tr}\Big[D_\gamma^{(j_{\mathrm{F}})}\big(g_{\mathrm{n}_t\mathrm{v}} g_{\mathrm{ve}'}\big) \cdots D_\gamma^{(j_{\mathrm{F}})}\big(g_{\mathrm{e}^{(n)}\mathrm{v}^n} g_{\mathrm{v}^{(n)}\mathrm{n}_s}\big) D^{(j_{\mathrm{F}})}\big(h_{\ell_{\mathrm{F}}}\big)\Big]\,. \tag{12}$$

The matrix $D^{(j)}$ is the Wigner D-matrix of the d_j-dimensional ($d_j = 2j+1$) representation of SU(2). The matrix $D_\gamma^{(j)}$ is the $d_j \times d_j$ matrix $\big[D_\gamma^{(j)}\big]_{mn}(g) = D^{(\gamma j,j)}_{jm\,jn}(g)$, $g \in \mathrm{SL}(2,\mathbb{C})$, where $D^{(\rho,k)}_{lm\,jn}$ are the matrix elements of the (ρ,k) unitary representation of the principal series of $\mathrm{SL}(2,\mathbb{C})$ in the basis labeled by the eigenfunctions of L^2 and L_z. The nodes n_s and n_t are respectively the source and the

[a]For the sake of simplicity, the orientation of every edge in each face is assumed to be the one induced from the face orientation. Since however this kind of orientation cannot be implemented consistently throughout the whole two-complex, in actual computations the orientation of the edges needs to be taken into account properly.

target of the link ℓ_{F} and γ is the Barbero-Immirzi parameter. The value of the label (n) in $\mathrm{v}^{(n)}$ and $\mathrm{e}^{(n)}$ is fixed for each face by the topology of the two-complex $\mathcal{C}$.

4.1. *Discretization of region $\mathscr{B}$*

In order to exploit the spin foam formalism to analyze the physics of region $\mathscr{B}$, the latter needs to be properly discretized. There is no unique or right way to perform the discretization. However, in order to get simpler and clearer calculations throughout the spin foam analysis it is particularly useful to preserve as much symmetries as possible during the procedure. In this subsection the combinatorial definition of both the cellular decomposition discretizing region $\mathscr{B}$ and its dual two-complex are given. Although the combinatorial definition of these objects is compact and complete, it does not convey as much geometrical insight as graphical representations do. For this reason, the interested reader is strongly encouraged to look at the graphical representations reported in Refs. 3 and 4.

As it can be seen from Fig. 4, the topology of region $\mathscr{B}$ is $S^2 \times [0,1] \times [0,1]$ and the topology of Σ is $S^2 \times S^1$. The geometry of Σ has two symmetries: the spherical SO(3) symmetry and the Z_2 time-reversal symmetry that exchanges p and f. There is also an additional topological Z_2 symmetry that exchanges $+$ and $-$. Since however the geometry of S_+ and S_- is different, this is not a symmetry of the geometry of Σ. Although the topology of Σ is not easy to discretize while preserving its symmetries, the discretization presented in this subsection is able to accomplish this task.

Before introducing the discretization, it is useful to fix some notational conventions. Let a, b, c, d be indices taking values in the set $\{1,2,3,4\}$, t be an index taking values in the set $\{p,f\}$ and ϵ be an index taking values in the set $\{+,-\}$. If the same expression contains several indices a, b, c, d, they are assumed to be all different from each other. The order of two consecutive indices is not important and the exchange of these indices results in the same element. If however two indices are separated by a comma, the exchange of these indices results in a different element.

Let p_a^+ be four points on S_+ and p_a^- be four points on S_-. The three-dimensional triangulation discretizing Σ is then defined by the points p_a^ϵ, the segments s_{ab}^ϵ and $s_{a,b}^t$, the triangles L_a^ϵ and $L_{a,b}^{t\epsilon}$, the tetrahedra $N_a^{t\epsilon}$ and N_{ab}^t, and their boundary relations:

$$\partial s_{ab}^\epsilon = (p_a^\epsilon, p_b^\epsilon); \tag{13}$$

$$\partial s_{a,b}^t = (p_a^-, p_b^+)^t; \tag{14}$$

$$\partial L_a^\epsilon = (s_{bc}^\epsilon, s_{cd}^\epsilon, s_{db}^\epsilon); \tag{15}$$

$$\partial L_{a,b}^{t+} = (s_{cd}^+, s_{a,c}^t, s_{a,d}^t); \tag{16}$$

$$\partial L_{a,b}^{t-} = (s_{cd}^-, s_{c,a}^t, s_{d,a}^t); \tag{17}$$

$$\partial N_a^{t\epsilon} = (L_{a,b}^{t\epsilon}, L_{a,c}^{t\epsilon}, L_{a,d}^{t\epsilon}, L_a^\epsilon); \tag{18}$$

$$\partial N_{ab}^t = (L_{a,b}^{t+}, L_{b,a}^{t+}, L_{c,d}^{t-}, L_{d,c}^{t-}). \tag{19}$$

Besides being a three-dimensional object of its own, this triangulation serves as the boundary of the cellular decomposition discretizing region $\mathscr{B}$. This four-dimensional cellular decomposition is defined by the two-dimensional surfaces $f_{a,b}$, the three-dimensional cells $e^{\epsilon}_{a,b}$, the four-dimensional cells v^{ϵ}_{a} and v_{ab}, and their boundary relations:

$$\partial f_{a,b} = (s^{p}_{a,b}, s^{f}_{a,b}); \tag{20}$$

$$\partial e^{\epsilon}_{a,b} = (L^{p\epsilon}_{a,b}, L^{f\epsilon}_{a,b}, f_{a,c}, f_{a,d}); \tag{21}$$

$$\partial v^{\epsilon}_{a} = (N^{p\epsilon}_{a}, N^{f\epsilon}_{a}, e^{\epsilon}_{a,b}, e^{\epsilon}_{a,c}, e^{\epsilon}_{a,d}); \tag{22}$$

$$\partial v_{ab} = (N^{p}_{ab}, N^{f}_{ab}, e^{+}_{a,b}, e^{+}_{b,a}, e^{-}_{c,d}, e^{-}_{d,c}). \tag{23}$$

Note that this cellular decomposition is not a triangulation.

The spin foam formalism is defined using the discrete object dual to the cellular decomposition. The graph Γ dual to the three-dimensional triangulation of Σ is defined as

Nodes
$$\mathrm{n}^{t\epsilon}_{a} \text{ and } \mathrm{n}^{t}_{ab}; \tag{24}$$

Links
$$\ell^{\epsilon}_{a} = (\mathrm{n}^{p\epsilon}_{a}, \mathrm{n}^{f\epsilon}_{a}); \tag{25}$$

$$\ell^{t+}_{a,b} = (\mathrm{n}^{t}_{ab}, \mathrm{n}^{t+}_{a}); \tag{26}$$

$$\ell^{t-}_{a,b} = (\mathrm{n}^{t}_{cd}, \mathrm{n}^{t-}_{a}). \tag{27}$$

The two-complex $\mathcal{C}$, whose boundary $\partial\mathcal{C}$ is Γ, dual to the four-dimensional cellular decomposition of region $\mathscr{B}$ is defined as

Vertices
$$\mathrm{v}^{\epsilon}_{a} \text{ and } \mathrm{v}_{ab}; \tag{28}$$

Edges
$$\mathrm{E}^{t\epsilon}_{a} = (\mathrm{v}^{\epsilon}_{a}, \mathrm{n}^{t\epsilon}_{a}); \tag{29}$$

$$\mathrm{E}^{t}_{ab} = (\mathrm{v}_{ab}, \mathrm{n}^{t}_{ab}); \tag{30}$$

$$\mathrm{e}^{+}_{a,b} = (\mathrm{v}^{+}_{a}, \mathrm{v}_{ab}); \tag{31}$$

$$\mathrm{e}^{-}_{a,b} = (\mathrm{v}^{-}_{a}, \mathrm{v}_{cd}); \tag{32}$$

Faces
$$\mathrm{F}^{\epsilon}_{a} = \left(\ell^{\epsilon}_{a}, (\mathrm{E}^{f\epsilon}_{a})^{-1}, \mathrm{E}^{p\epsilon}_{a}\right); \tag{33}$$

$$\mathrm{F}^{t+}_{a,b} = \left(\ell^{t+}_{a,b}, (\mathrm{E}^{t+}_{a})^{-1}, \mathrm{e}^{+}_{a,b}, \mathrm{E}^{t}_{ab}\right); \tag{34}$$

$$\mathrm{F}^{t-}_{a,b} = \left(\ell^{t-}_{a,b}, (\mathrm{E}^{t-}_{a})^{-1}, \mathrm{E}^{-}_{a,b}, \mathrm{E}^{t}_{cd}\right); \tag{35}$$

$$\mathrm{f}_{a,b} \overset{c\leq d}{=} \left(\mathrm{e}^{+}_{a,c}, (\mathrm{e}^{-}_{b,d})^{-1}, \mathrm{e}^{-}_{b,c}, (\mathrm{e}^{+}_{a,d})^{-1}\right). \tag{36}$$

The orientation of each element of the two-complex can be easily read from this combinatorial definition.

It is interesting to analyze how many of the topological symmetries of region $\mathscr{B}$ are preserved under this discretization. The spherical SO(3) symmetry of region $\mathscr{B}$ is discretized to a tetrahedral symmetry of the two-complex, which is realized as an even permutation of the indices a, b, c, d. The Z_2 time-reversal symmetry that exchanges the indices p and f and the Z_2 symmetry that exchanges the indices $+$ and $-$ are instead preserved exactly.

4.2. *Discrete geometrical data and boundary state*

The continuous geometry of Σ is approximated by the assignment of discrete geometrical data to the triangulation discretizing Σ (or equivalently to is dual graph Γ). There is once again no unique or right way to do it. Different assignments of discrete geometrical data simply define different approximations of the same continuous geometry. The discrete geometry presented in this subsection preserves the two geometrical symmetries of the continuous geometry of Σ.

To preserve the Z_2 time-reversal symmetry is sufficient to firstly construct the discrete geometrical data for the triangulation discretizing Σ^p and then to define the discrete geometrical data for the triangulation discretizing Σ^f as their time reversal. The sphere S_+ (S_-) in Σ is discretized by the tetrahedron bounded by the four triangles $\{L_a^+\}$ ($\{L_a^-\}$) in the triangulation. In order to preserve as much symmetry as possible in the discretization process of the geometry, the same area $a_+ = \pi r_+^2$ ($a_- = \pi r_-^2$), which is one fourth of the area of the sphere S_+ (S_-), is assigned to each triangle L_a^+ (L_a^-).

The continuous intrinsic geometry of Σ^p is specified by the line element

$$\mathrm{d}s^2 = f^2(r)\,\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2\,, \tag{37}$$

in which $f^2(r) = \beta\left(2 - \beta\,(1 - 2m/r)\right)$ on Σ_+^p and $f^2(r) = 1$ on Σ_-^p. To begin with, this line element is approximated with

$$\mathrm{d}s^2 = \xi^2\,\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega^2\,, \tag{38}$$

where ξ is a constant that is fixed by requiring the volume of Σ^p computed with the line element in Eq. (38) to coincide with the volume of Σ^p computed with the line element in Eq. (37). This condition uniquely fixes the value of ξ in terms of the four parameters $\left(m,\, v,\, r_\pm\right)$ characterizing the continuous geometry of Σ^p. Requiring the discrete geometrical data to preserve the topological tetrahedral symmetry of the triangulation of Σ, the same area A_+ (A_-) must be assigned to each triangle $L_{a,b}^{p+}$ ($L_{a,b}^{p-}$). The line element in Eq. (38) is then used to compute the value of $A_\pm$, which is

$$A_\pm = \pi r_\pm^2 \sqrt{\frac{\xi^2}{18}\Big(1 - 3\,\frac{r_\mp}{r_\pm}\Big)^2 + \frac{2}{3}}\,. \tag{39}$$

The symmetry conditions together with the values of the areas $a_\pm$ and $A_\pm$ completely define the discrete intrinsic geometry of the triangulation of Σ. Analogously, the discrete extrinsic geometry of the triangulation is completely defined by the assignment of the extrinsic angles $\theta_\pm$ between $N_a^{p\pm}$ and $N_a^{f\pm}$ at $L_a^\pm$ and the extrinsic angles $\Theta_\pm$ between $N_a^{p\pm}$ and $N_{a,b}^p$ at $L_{a,b}^{p\pm}$.

Let $n_\mu^{t\epsilon}$ be the normal one-form of Σ_ϵ^t. From the definition of $\theta_\pm$ it follows that

$$\cos(\theta_\pm) = \Big(\, g^{\mu\nu}\, n_\mu^{f\,\pm}\, n_\nu^{p\,\pm}\,\Big)\Big|_{S_\pm}\,, \tag{40}$$

where $g^{\mu\nu}$ is the inverse of the metric tensor defined by the line element in Eq. (37), giving

$$\cos(\theta_+) = \frac{1 + \Big[\big(1 - 2m/r_+\big)\,\beta - 1\Big]^2}{\big|\,\beta\,\big(\,\beta - 2 - 2m\beta/r_+\big)\big|\,\big(1 - 2m/r_+\big)} \tag{41}$$

and

$$\cos(\theta_-) = \frac{1 + 2m/r_-}{1 - 2m/r_-}\,. \tag{42}$$

The angles $\Theta_\pm$ represent a discrete approximation of the continuous extrinsic curvature of $\Sigma_\pm$. A convenient discretization that preserves the symmetries of the triangulation is

$$\Theta_\pm = \frac{1}{12}\int_{\Sigma^p_\pm} \big(k^\pm\big)^i{}_i\,, \tag{43}$$

where $k^\pm_{ij}$ is the extrinsic curvature tensor of $\Sigma^p_\pm$ defined in Eqs. (7) and (8).

The discrete geometry of the triangulation of Σ has thus been explicitly constructed in terms of the four parameters $\big(m,\, v,\, r_\pm\big)$ characterizing the continuous geometry of Σ. From the point of view of the dual graph Γ, the discrete geometry consists in the assignment of an area and an angle to each link of Γ. The area assigned to each link represents the area of the triangle that is dual to the link and the angle assigned to each link represents the extrinsic curvature between the two tetrahedra that share the triangle dual to the link. This geometrical data uniquely specifies an extrinsic coherent state[5] $\Psi_{BW} = \Psi_{BW}\big(a_\pm,\, \theta_\pm,\, A_\pm,\, \Theta_\pm\big) = \Psi_{BW}\big(m,\, v,\, r_\pm\big) \in \mathcal{H}_\Gamma$ that is peaked on the discrete classical geometry defined by $\big(a_\pm,\, \theta_\pm,\, A_\pm,\, \Theta_\pm\big)$. The boundary state $\Psi_{BW} \in \mathcal{H}_\Gamma$ is the quantum state representing the outcome of the transition taking place in region $\mathscr{B}$.

4.3. *Transition amplitude*

Having explicitly constructed the two-complex discretizing region $\mathscr{B}$ and the boundary state describing the outcome of the quantum transition, the spin foam transition amplitude for the black-to-white hole transition can be readily computed using the formulas in Eqs. (9) and (10).

The assignment of group elements to the edges and the links of the two-complex $\mathcal{C}$ is:

$$\ell^\epsilon_a \longleftrightarrow h^\epsilon_a \in \mathrm{SU}(2)\,; \tag{44}$$

$$\ell^{t\epsilon}_{a,b} \longleftrightarrow h^{t\epsilon}_{a,b} \in \mathrm{SU}(2)\,; \tag{45}$$

$$\mathrm{E}^{t\epsilon}_a \longleftrightarrow g^{t\epsilon}_a \in \mathrm{SL}(2,\mathbb{C})\,; \tag{46}$$

$$\mathrm{E}^t_{ab} \longleftrightarrow g^t_{ab} \in \mathrm{SL}(2,\mathbb{C})\,; \tag{47}$$

$$\mathrm{e}^\epsilon_{a,b} \longleftrightarrow g^\epsilon_{a\to b}\,,\; g^\epsilon_{a\leftarrow b} \in \mathrm{SL}(2,\mathbb{C})\,. \tag{48}$$

The group elements $g^{\epsilon}_{a\leftrightarrow b}$ are assigned to the two oriented half-edges of $\mathrm{e}^{\epsilon}_{a,b}$. The element $g^{\epsilon}_{a\to b}$ is assigned to the oriented half-edge with source in the source of $\mathrm{e}^{\epsilon}_{a,b}$ and target in the center of $\mathrm{e}^{\epsilon}_{a,b}$. The element $g^{\epsilon}_{a\leftarrow b}$ is assigned to the oriented half-edge with source in the target of $\mathrm{e}^{\epsilon}_{a,b}$ and target in the center of $\mathrm{e}^{\epsilon}_{a,b}$. Carefully considering the topology and the orientation pattern of the two-complex, the elementary face amplitudes (see Eqs. (11) and (12)) for $\mathcal{C}$ can be computed:

$$A^{\epsilon}_{a}\big(\{g^{t\epsilon}_{a}\},\, h^{\epsilon}_{a}\big) = \sum_{j} d_j \,\mathrm{Tr}\Big[D^{(j)}_{\gamma}\big(\big(g^{f\epsilon}_{a}\big)^{-1}\, g^{p\epsilon}_{a}\big)\, D^{(j)}\big(h^{\epsilon}_{a}\big)\Big], \tag{49}$$

$$\begin{aligned} A^{t+}_{a,b}\big(g^{t+}_{a},\, g^{t}_{ab},\, g^{+}_{a\leftrightarrow b}, h^{t+}_{a,b}\big) = \sum_{j} d_j \,\mathrm{Tr}\Big[& D^{(j)}_{\gamma}\big(\big(g^{t+}_{a}\big)^{-1}\, g^{+}_{a\to b}\big) D^{(j)}_{\gamma}\big(\big(g^{+}_{a\leftarrow b}\big)^{-1}\; g^{t}_{ab}\big) \\ & \times D^{(j)}\big(h^{t+}_{a,b}\big)\Big], \end{aligned} \tag{50}$$

$$\begin{aligned} A^{t-}_{a,b}\big(g^{t-}_{a}, g^{t}_{cd}, g^{-}_{a\leftrightarrow b}, h^{t-}_{a,b}\big) = \sum_{j} d_j \,\mathrm{Tr}\Big[& D^{(j)}_{\gamma}\big(\big(g^{t-}_{a}\big)^{-1}\, g^{-}_{a\to b}\big) D^{(j)}_{\gamma}\big(\big(g^{-}_{a\leftarrow b}\big)^{-1}\; g^{t}_{cd}\big) \\ & \times D^{(j)}\big(h^{t-}_{a,b}\big)\Big], \end{aligned} \tag{51}$$

$$\begin{aligned} A_{a,b}\big(\{g^{+}_{a\leftrightarrow c}\},\, \{g^{-}_{b\leftrightarrow c}\}\big) \overset{c\leq d}{=} \sum_{j} d_j \,\mathrm{Tr}\Big[& D^{(j)}_{\gamma}\big(\big(g^{+}_{a\to d}\big)^{-1}\, g^{+}_{a\to c}\big)\, D^{(j)}_{\gamma}\big(\big(g^{+}_{a\leftarrow c}\big)^{-1}\, g^{-}_{b\leftarrow d}\big) \\ & \times D^{(j)}_{\gamma}\big(\big(g^{-}_{b\to d}\big)^{-1}\, g^{-}_{b\to c}\big)\, D^{(j)}_{\gamma}\big(\big(g^{-}_{b\leftarrow c}\big)^{-1}\, g^{+}_{a\leftarrow d}\big)\Big]. \end{aligned} \tag{52}$$

The two-complex transition amplitude $W_{\mathcal{C}}(\{h_\ell\})$ can then be written as

$$\begin{aligned} W_{\mathcal{C}}\big(\{h^{\epsilon}_{a}\},\, \{h^{t\epsilon}_{a,b}\}\big) = \int_{\mathrm{SL}(2,\mathbb{C})} & \prod_{\epsilon a} \mathrm{d}g^{p\epsilon}_{a} \prod_{a<b} \mathrm{d}g^{p}_{ab} \prod_{\epsilon ab} \mathrm{d}g^{\epsilon}_{a\leftrightarrow b} \\ & \times \prod_{\epsilon a} A^{\epsilon}_{a}\big(\{g^{t\epsilon}_{a}\},\, h^{\epsilon}_{a}\big) \prod_{ab} A_{a,b}\big(\{g^{+}_{a\leftrightarrow c}\},\, \{g^{-}_{b\leftrightarrow c}\}\big) \\ & \times \prod_{tab} A^{t+}_{a,b}\big(g^{t+}_{a},\, g^{t}_{ab},\, g^{+}_{a\leftrightarrow b}, h^{t+}_{a,b}\big) \\ & \times \prod_{tab} A^{t-}_{a,b}\big(g^{t-}_{a}, g^{t}_{cd}, g^{-}_{a\leftrightarrow b}, h^{t-}_{a,b}\big). \end{aligned} \tag{53}$$

Since Eq. (10) dictates to drop one integration per vertex, the integration over the $\{g^{f\epsilon}_{a}\}$ and $\{g^{f}_{ab}\}$ variables has been dropped.

Finally, the black-to-white hole transition amplitude $W\big(m,\, v,\, r_{\pm}\big)$ is

$$\begin{aligned} W\big(m,\, v,\, r_{\pm}\big) = W\big[\Psi_{BW}\big] = \int_{\mathrm{SU}(2)} & \prod_{\epsilon a} \mathrm{d}h^{\epsilon}_{a} \prod_{t\epsilon ab} \mathrm{d}h^{t\epsilon}_{a,b}\; W_{\mathcal{C}}\big(\{h^{\epsilon}_{a}\},\, \{h^{t\epsilon}_{a,b}\}\big) \\ & \times \Psi_{\mathrm{BW}}\big(\{h^{\epsilon}_{a}\},\, \{h^{t\epsilon}_{a,b}\}\big). \end{aligned} \tag{54}$$

This expression contains the whole physics of the phenomenon. Its investigation is currently ongoing and it will be reported elsewhere.

5. Summary and outlook

The quantum region of a black hole spacetime can be divided in three different subregions: region $\mathscr{C}$, where the collapsing matter enters its quantum gravity regime; region $\mathscr{B}$, where the horizon reaches Planckian size at the end of the evaporation process; region $\mathscr{A}$, where the curvature reaches Planckian scale independently from region $\mathscr{B}$ and region $\mathscr{C}$. The principle of locality demands that these regions are independent from each other and that they can subsequently be studied separately. The evidence in favor of the black-to-white hole transition scenario resulting from the separate analysis of the physics of these regions is overwhelming.

Focusing on the black-to-white hole horizon transition in region $\mathscr{B}$, that is the last stage of the life of a black hole in this scenario, the physics of the boundary of the quantum region is completely determined by four parameters: the mass m of the black hole at the moment in which the transition takes place; the external asymptotic (retarded) time v it takes for the transition to happen; the minimal external radius r_+ for which the classical theory can still be trusted; the minimal internal radius r_- reached by the black hole interior in region $\mathscr{A}$.

The spin foam formalism provides a clear framework to study this scenario. However, since the theory is properly defined in the discrete setting, to compute the transition amplitude for the phenomenon the physics of region $\mathscr{B}$ needs to be appropriately discretized. From a practical point of view the discretization needs to be both sufficiently refined to capture the relevant degrees of freedom of the phenomenon and sufficiently simple for the transition amplitude to be explicitly computed and analyzed. In this work a convenient discretization that preserves the symmetries of the continuous geometry is explicitly constructed and the resulting transition amplitude $W\big(m, v, r_\pm\big)$ is computed. Although the two-complex $\mathcal{C}$ discretizing region $\mathscr{B}$ is rather complicated, its high degree of symmetry allows the transition amplitude to be expressed in a remarkably compact way. Due to the severe complexity of spin foam amplitudes, an analytical study of the transition amplitude $W\big(m, v, r_\pm\big)$ for the black-to-white hole transition is not feasible at this point. On the other hand, recent developments[25–28] in the numerical computation of spin foam amplitudes and the high degree of symmetry of the constructed discretization should allow a numerical investigation of the transition amplitude.

In this work the black hole lifetime is assumed to be of the order of the evaporation process (although the constructed framework can describe also smaller timescales). This choice is motivated by the analysis of the black-to-white hole transition performed in Refs. 29 and 30, where, neglecting Hawking radiation, the lifetime of the black hole was estimated to be much longer of the evaporation process, thus proving that the assumption of neglecting Hawking radiation was not justified. However, due to the use of a fairly coarse discretization and of several rough approximations it is unclear whether this result is reliable. Furthermore, recent results[17] seems to support the black hole lifetime of the order m_0^2 heuristically suggested in Ref. 1. The numerical analysis of the transition amplitude $W\big(m, v, r_\pm\big)$

computed in this work may provide an estimation of the black-to-white hole transition timescales and improve the understanding of its phenomenology.[31–34]

References

1. H. M. Haggard and C. Rovelli, Black hole fireworks: Quantum-gravity effects outside the horizon spark black to white hole tunneling, *Physical Review D* **92** (2015).
2. S. W. Hawking, Black hole explosions?, *Nature* **248** (1974).
3. F. D'Ambrosio, M. Christodoulou, P. Martin-Dussaud, C. Rovelli and F. Soltani, End of a black hole's evaporation, *Physical Review D* **103** (2021).
4. F. Soltani, C. Rovelli and P. Martin-Dussaud, End of a black hole's evaporation. II., *Physical Review D* **104** (2021).
5. C. Rovelli and F. Vidotto, *Covariant Loop Quantum Gravity* (Cambridge University Press, 2015).
6. A. Perez, The spin-foam approach to quantum gravity, *Living Reviews in Relativity* **16** (2013).
7. A. Rignon-Bret and C. Rovelli, Black to white transition of a charged black hole (2021).
8. E. Bianchi, M. Christodoulou, F. D'Ambrosio, H. M. Haggard and C. Rovelli, White holes as remnants: A surprising scenario for the end of a black hole, *Classical and Quantum Gravity* **35** (2018).
9. A. Corichi and P. Singh, Loop quantization of the Schwarzschild interior revisited, *Classical and Quantum Gravity* **33** (2016).
10. A. Ashtekar, J. Olmedo and P. Singh, Quantum transfiguration of Kruskal black holes, *Physical Review Letters* **121** (2018).
11. R. Gambini, J. Olmedo and J. Pullin, Spherically symmetric loop quantum gravity: Analysis of improved dynamics, *Classical and Quantum Gravity* **37** (2020).
12. K. Peeters, C. Schweigert and J. W. van Holten, Extended geometry of black holes, *Classical and Quantum Gravity* **12** (1995).
13. F. D'Ambrosio and C. Rovelli, How information crosses Schwarzschild's central singularity, *Classical and Quantum Gravity* **35** (2018).
14. C. Rovelli and F. Vidotto, Planck stars, *International Journal of Modern Physics D* **23** (2014).
15. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the big bang, *Physical Review Letters* **96** (2006).
16. P. Hajicek and C. Kiefer, Singularity avoidance by collapsing shells in quantum gravity, *International Journal of Modern Physics D* **10** (2001).
17. J. G. Kelly, R. Santacruz and E. Wilson-Ewing, Black hole collapse and bounce in effective loop quantum gravity, *Classical and Quantum Gravity* **38** (2020).
18. W. Piechocki and T. Schmitz, Quantum Oppenheimer-Snyder model, *Physical Review D* **102** (2020).
19. J. Münch, Effective quantum dust collapse via surface matching, *Classical and Quantum Gravity* **38** (2021).
20. C. Rovelli and P. Martin-Dussaud, Interior metric and ray-tracing map in the firework black-to-white hole transition, *Classical and Quantum Gravity* **35** (2018).
21. P. Martin-Dussaud and C. Rovelli, Evaporating black-to-white hole, *Classical and Quantum Gravity* **36** (2019).
22. J. Engle, E. Livine, R. Pereira and C. Rovelli, LQG vertex with finite Immirzi parameter, *Nuclear Physics B* **799** (2008).

23. W. Kamiński, M. Kisielowski and J. Lewandowski, Spin-foams for all loop quantum gravity, *Classical and Quantum Gravity* **27** (2010).
24. Y. Ding, M. Han and C. Rovelli, Generalized spinfoams, *Physical Review D* **83** (2011).
25. P. Donà and G. Sarno, Numerical methods for EPRL spin foam transition amplitudes and Lorentzian recoupling theory, *General Relativity and Gravitation* **50** (2018).
26. P. Donà, M. Fanizza, G. Sarno and S. Speziale, Numerical study of the lorentzian Engle-Pereira-Rovelli-Livine spin foam amplitude, *Physical Review D* **100** (2019).
27. F. Gozzini, A high-performance code for EPRL spin foam amplitudes (2021).
28. P. Frisoni, Studying of the EPRL self-energy, in *16th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories*, 2021.
29. M. Christodoulou, C. Rovelli, S. Speziale and I. Vilensky, Planck star tunneling time: An astrophysically relevant observable from background-free quantum gravity, *Physical Review D* **94** (2016).
30. M. Christodoulou and F. D'Ambrosio, Characteristic time scales for the geometry transition of a black hole to a white hole from spinfoams (2018).
31. A. Barrau, C. Rovelli and F. Vidotto, Fast radio bursts and white hole signals, *Physical Review D* **90** (2014).
32. A. Barrau, B. Bolliet, F. Vidotto and C. Weimer, Phenomenology of bouncing black holes in quantum gravity: A closer look, *Journal of Cosmology and Astroparticle Physics* **2016** (2016).
33. A. Barrau, B. Bolliet, M. Schutten and F. Vidotto, Bouncing black holes in quantum gravity and the Fermi gamma-ray excess, *Physics Letters B* **772** (2017).
34. C. Rovelli and F. Vidotto, Small black/white hole stability and dark matter, *Universe* **4** (2018).

Holographic properties of the bulk-to-boundary transmission of information in regions of quantum space

Eugenia Colafranceschi
School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom
E-mail: eugenia.colafranceschi@nottingham.ac.uk

We show that quantum gravity states associated to open spin network graphs implicitly define maps from the bulk to the boundary of the corresponding region of quantum space. Employing random tensor network techniques, we then investigate under which conditions the flow of information from the bulk to the boundary is an isometric map, which is a necessary condition for holography.

Keywords: Holography; Spin Networks; Random Tensor Networks.

1. Introduction

Over the past two decades, quantum information has been playing an increasingly prominent role in quantum gravity.[1] A crucial factor in this is the relation between entanglement and geometry, which has been highlighted by various results in different quantum gravity contexts. As an example, within the AdS/CFT correspondence,[2] which describes quantum gravity in asimptotically anti-de Sitter (AdS) spacetime in terms of a conformal field theory (CFT) living on its boundary, the Ryu-Takayanagi (RT) formula[3,4] relates the entanglement entropy of the boundary CFT in a subregion to the area of a codimension-2 surface in the dual bulk spacetime. The entanglement-geometry relation pointed out by the RT formula is further strengthened by the observation that entanglement (of boundary degrees of freedom) is a necessary condition for the connectivity of the dual spacetime.[5,6] More generally, quantum correlations are assumed to be at the origin of spacetime geometry in all those approaches to quantum gravity that regard spacetime as emerging from fundamental quantum entities.[7] This aspect is particularly emphasised in the approach we adopt here, group field theories[8–10] (GFT), where discrete geometries arise as pattern of entanglement among space quanta. Specifically, the GFT quanta are fundamental simplices dual to spin network vertices, which upon becoming entangled (in their edge degrees of freedom) form spin network graphs dual to simplicial complexes. In this respect, GFT realises what the aforementioned results suggest: the idea of *quantum spacetime as a geometric representation of entanglement*.

Remarkably, a geometric representation of entanglement is the key feature of a quantum information tool, tensor networks[11–13] (TN), which consists in a

decomposition of many-body wave-functions (specifically, of their entanglement structure) into interconnected (one-body) tensors. This feature has been largely exploited in the AdS/CFT correspondence, starting from Ref. 14, which showed that the entanglement renormalization of CFT boundary states carries out a TN decomposition of the latter, with the possibility to interpret the emerging network as (a spatial slice of) AdS spacetime (upon definition of a metric from combinatorial ingredients). Exactly-solvable models of the AdS/CFT correspondence have then been constructed through perfect[15] and random[16] tensors.

However, in these works (and others of similar nature) tensor networks, which *a priori* have nothing to do with gravity, are provided with a spatiotemporal interpretation by defining a metric on the network through combinatorial elements. We make a change of perspective[17] here, in which (suitably generalised) tensor networks are identified, through the match with the GFT formalism, with a quantum gravity language, where the "bodies" of the many-body system are quanta of space, and link-entanglement expresses adjacency relations between them; in other terms, the GFT-generalised tensor networks naturally carry a quantum-geometry interpretation. This identification leads to the possibility of using tensor network techniques in a full quantum gravity context, to study entanglement-based properties of spin network states, which are of interest to a wider part of the quantum gravity community (as they enter other approaches as well, such as loop quantum gravity, spinfoam models and lattice quantum gravity).

On this ground, we look at what is supposed to be a constitutive feature of quantum gravity, originally prompted by the Bekenstein-Hawking area law for the black hole entropy: holography. The holographic principle,[18,19] of which the AdS/CFT correspondence is a realization, argues that the three-dimensional world is the holographic image of data living on a two-dimensional space. Crucially, certain classes of tensor networks turn out to possess holographic features[20] (and have in fact been largely employed in the AdS/CFT context, as recalled earlier). Within the perspective change mentioned above, we thus exploit tensor network methods to highlight properties of spin network states that are related to holography.

We specifically look at the transmission of information from the bulk to the boundary of an open spin network graph (where the bulk is made by the intertwiner degrees of freedom associated to the vertices, and the boundary by spin degrees of freedom on open links of the graph), pointing out that every spin network state can be seen as a map between these two sets. As we shall see, the map then turns out to be an isometry (which is a necessary condition for an holographic bulk-to-boundary transmission of information) if and only if the reduced bulk state has maximum entanglement entropy. Assuming a random distribution of weights associated to the individual spin network vertices, we carry out the entropy calculation via random tensor network techniques, and thereby manage to highlight the properties of spin network graphs that favour in them an holographic behaviour. The results reported here have been obtained as part of the work presented in Ref. 21.

2. Quantum space from spin network entanglement graphs

The GFT models for quantum gravity[8–10] are quantum field theories defined on d copies of a group G, whose fundamental excitations are $(d-1)$-simplices (to be understood as quanta of space), decorated by group variables (one for each face) that capture their geometric properties. In the model of $4D$ quantum gravity, $d = 4$ and the usual choice for the group is $G = SU(2)$; the space quanta are then quantized tetrahedra. In a dual picture, the fundamental $(d-1)$-simplex is represented as a d-valent vertex, with an edge dual to each face and carrying the corresponding geometric data, i.e. a spin network vertex.

As we are going to detail in the following, spatial geometries, in the form of simplicial complexes dual to spin network graphs, arise in GFT from the gluing of fundamental simplices/spin-network vertices, given by entanglement of their edge degrees of freedom.

2.1. *Combinatorial structure of spin network graphs*

We start by introducing some combinatorics notions needed to construct spin network states in the GFT framework.

Let $V = \{1, ..., N\}$ be a set of vertices. The d lines departing from a vertex are distinguished by a "colour" $i = 1, ..., d$, and the i-th line of a vertex v is denoted as e^i_v. In graph theory, a connectivity pattern among N vertices is encoded in a $N \times N$ matrix A, called *adjacency matrix*, such that its element A_{vw} is equal to 1 or 0 if, respectively, vertices v and w are connected or not by a link. To include the information of the colour labelling the links between vertices, we can use a $(N \cdot d) \times (N \cdot d)$ generalised adjacency matrix; we also assume that vertices can be connected only along edges of the same colour. Then:

- e^i_v and e^i_w form a link, denoted by e^i_{vw}, if and only if $A_{(v-1)\cdot d+i,(w-1)\cdot d+i} = 1$
- e^i_v is an open edge if and only if $A_{(v-1)\cdot d+i,(w-1)\cdot d+i} = 0 \ \forall \ w \in V$

The connectivity pattern defined by the adjacency matrix A, also called *graph*, is indicated as γ; the two notions A and γ thus contain the same information. Given the graph γ, we denote as L the set of internal links: $L = \{e^i_{vw} | A_{(v-1)\cdot d+i,(w-1)\cdot d+i} = 1\}$, and as $\partial\gamma$ the set of boundary links: $\partial\gamma = \{e^i_v | A_{(v-1)\cdot d+i,(w-1)\cdot d+i} = 0 \ \forall \ w \in V\}$. We also call E the set of all edges of the graph: $E = L \cup \partial\gamma$.

2.2. *Spin network entanglement graphs*

The d-valent vertices introduced above, dressed with group geometric data, are spin network vertices dual to $(d-1)$-simplices, representing fundamental excitations of the GFT field. In particular, each line e^i_v of a vertex v, corresponding to a face of the dual simplex, is decorated with a group variable g^i_v, and the complete set $g^1_v, ..., g^d_v$ specifies the geometry of the $(d-1)$-simplex up to global action of the group G; the associated Hilbert space is thus given by $\mathcal{H}_v = L^2(G^d/G)$.

A spin network graph (dual to a simplicial complex) can be constructed out of a set $V = \{1, ..., N\}$ of vertices by correlating their edge degrees of freedom (which graphically amounts to gluing their open edges); the corresponding state thus lives in a subspace of the Hilbert space associated to the set of (a priori, open) vertices $\mathcal{H}_V = \bigotimes_{v=1}^{N} \mathcal{H}_v$. In particular, invariance of the state under right action of the group on the variables g_v^i and g_w^i implies that e_v^i and e_w^i are glued together to form a link e_{vw}^i, which carries the element $g_{vw}^i = g_v^i (g_w^i)^{-1}$. The wavefunction ψ_γ associated with a graph γ can thus be obtained by a generic $\psi \in \mathcal{H}_V$ as follows:

$$\psi_\gamma(..., g_{vw}^i = g_v^i (g_w^i)^{-1}, ...) = \int \mathrm{d}h \; \psi(..., g_v^i h, ..., g_w^i h, ...) \tag{1}$$

and pertains to the Hilbert subspace $\mathcal{H}_\gamma = L^2(G^E/G^V) \subseteq \mathcal{H}_V$. Note that, although GFT and LQG associate the same set of states to a given graph γ (namely, the ones of $\mathcal{H}_\gamma$), the nature of such states in the two theories is radically different.[22] In fact, the GFT ones live in a Hilbert space ($\mathcal{H}_V$) which describes a collection of possibly disconnected vertices; the graph structure, and hence the simplicial complex associated with it, is not given *a priori*, but emerges from the particular form of the state (specifically, from its entanglement structure, as explained below).

By the Peter-Weyl theorem, the single-vertex Hilbert space can be decomposed into irreducible representations j of $SU(2)$ (spin network representation):

$$\mathcal{H}_v = \bigoplus_{\vec{j}_v} \left(\mathcal{I}^{\vec{j}_v} \otimes \bigotimes_{i=1}^{d} V^{j_v^i} \right) \tag{2}$$

where $\vec{j}_v = j_v^1, ..., j_v^d$ are the $SU(2)$ representations associated to the edges $e_v^1, ..., e_v^d$ of vertex v, V^j is the space carrying the representation j, and $\mathcal{I}^{\vec{j}} = \mathrm{Inv}_{SU(2)} \left[V^{j^1} \otimes ... \otimes V^{j^d} \right]$ is the space of d-valent intertwiners recoupling spins $j^1, ..., j^d$. A basis state for Eq. (2) is denoted by $|\vec{j}_v \vec{n}_v \iota_v\rangle$, where $\vec{n}_v = n_v^1, ..., n_v^d$ with n_v^i quantum number (spin projection) labelling a basis of $V^{j_v^i}$, and $\iota_v \in \mathcal{I}^{\vec{j}_v}$.

In the spin network basis, gluing vertices corresponds to imposing maximal entanglement between edge spins:

$$|\psi_\gamma\rangle = \left(\bigotimes_{e_{vw}^i \in L} \langle e_{vw}^i | \right) |\psi\rangle \tag{3}$$

where $|e_{vw}^i\rangle$ is a maximally entangled state of spins associated to e_v^i and e_w^i. The entanglement pattern of a state thus constitutes the graph dual to the simplicial complex the state represents. For this reason we refer to states like that in Eq. (3) as *entanglement graphs*.

Denoting by J an assignment of spins to all edges of γ, i.e. $J = \{j_e | e \in E\}$, the Hilbert subspace of $\mathcal{H}_V$ associated to γ can be written as follows:

$$\mathcal{H}_\gamma = \bigoplus_{J} \left(\bigotimes_{v} \mathcal{I}^{\vec{j}_v} \otimes \bigotimes_{e \in \partial\gamma} V^{j_e} \right) \tag{4}$$

It is clear from Eq. (4) that we can identify, for each graph, two different sets of degrees of freedom: intertwiners attached to the vertices, which we refer to as *bulk degrees of freedom*, and spin projections attached to the boundary edges, the *boundary degrees of freedom*. Note that the two sets are not independent, as intertwiners also recouple spins that pertain to the boundary; however, as can be easily seen from Eq. (4), they can be considered independent within each fixed-spin subspace of the direct sum.

2.2.1. *Classes of spin network entanglement graphs as tensor networks*

A tensor network[11–13] is, formally, a set of tensors whose indices are contracted according to a certain patter; when each tensor accounts for the state of a quantum system (say a "particle"), the tensor network is a many-body wavefunction. Its peculiarity (underlying the efficiency of this quantum information tool) is that the network represents the skeleton of quantum correlations (typically, entanglement) among the individual particles. This recalls the GFT graphs, which represent the entanglement structure of a many-vertex state. Let us illustrate the analogy with an example: consider a set of *open* vertices in the factorized state $\bigotimes_v |f_v\rangle \in \mathcal{H}_V$, where

$$|f_v\rangle = \bigoplus_{\vec{j}_v} \sum_{\vec{n}_v \iota_v} (f_v)^{\vec{j}_v}_{\vec{n}_v \iota_v} |\vec{j}_v \vec{n}_v \iota_v\rangle \tag{5}$$

is the generic state of vertex v, and glue them according to a given combinatorial pattern γ:

$$|\phi_\gamma\rangle = \left(\bigotimes_{e^i_{vw} \in L} \langle e^i_{vw}| \right) \bigotimes_v |f_v\rangle \tag{6}$$

In the case where every f_v is picked on specific values $\vec{j}_v$ of the edge spins, the state of Eq. (6) is an ordinary tensor network, specifically a PEPS, having j_e with $e \in E$ as bond dimensions[13]:

$$(\phi_\gamma)^{\{j_{e\in E}\}}_{\{n_{e\in\partial\gamma}\}\{\iota_{v\in V}\}} = (f_1)^{\vec{j}_1}_{\vec{n}_1 \iota_1} \cdots (f_N)^{\vec{j}_N}_{\vec{n}_N \iota_N} \prod_{e^i_{vw} \in L} \delta_{j^i_v j^i_w} \delta_{n^i_v n^i_w} \tag{7}$$

In the more general case where every f_v runs over all possible values of the edge spins, Eq. (6) corresponds to a generalised tensor network, whose bond dimensions are dynamical variables. For a detailed discussion on the analogy between GFT entanglement graphs and tensor networks see Ref. 17.

3. Bulk-to-boundary transmission of information

3.1. *Spin network entanglement graphs as bulk-to-boundary maps*

By considering a bipartition of the degrees of freedom of an entanglement graph (like the one in Eq. (3)) into *bulk* (intertwiners on vertices, collectively indicated as $\dot{\gamma}$)

and *boundary* (spin projections on open edges), we can regard it as a map between these two sets. Such a viewpoint is quite natural when we consider an entanglement graph with fixed spins $J = \{j_e | e \in E\}$, whose Hilbert space factorizes, as can be seen from Eq. (4), into a bulk space $\mathcal{H}_{\dot{\gamma}}(J) = \bigotimes_v \mathcal{I}^{\vec{j}_v}$ and a boundary space $\mathcal{H}_{\partial\gamma}(J_\partial) = \bigotimes_{e\in\partial\gamma} V^{j_e}$, where $J_\partial = \{j_e | e \in \partial\gamma\} \subset J$ is the set of spins on the boundary edges. In particular, to every state $|\psi_\gamma\rangle \in \mathcal{H}_\gamma(J) = \mathcal{H}_{\dot{\gamma}}(J) \otimes \mathcal{H}_{\partial\gamma}(J_\partial)$ we can associate a map $M[\psi_\gamma] : \mathcal{H}_{\dot{\gamma}}(J) \to \mathcal{H}_{\partial\gamma}(J_\partial)$ acting as follows:

$$M[\psi_\gamma]: \quad |\zeta\rangle \in \mathcal{H}_{\dot{\gamma}}(J) \quad \to \quad \langle\zeta|\psi_\gamma\rangle = |\psi_{\partial\gamma}(\zeta)\rangle \in \mathcal{H}_{\partial\gamma}(J_\partial) \tag{8}$$

In the more general case of $|\psi_\gamma\rangle$ spreading over all possible values of the edge spins we can still define an associated bulk-to-boundary map $M[\psi_\gamma]$: by embedding the graph Hilbert space $\mathcal{H}_\gamma$, which is given by Eq. (4), into $\mathcal{H}_{\dot{\gamma}} \otimes \mathcal{H}_{\partial\gamma}$, where

$$\mathcal{H}_{\dot{\gamma}} = \bigoplus_J \mathcal{H}_{\dot{\gamma}}(J), \tag{9}$$

$$\mathcal{H}_{\partial\gamma} = \bigoplus_{J_\partial} \mathcal{H}_{\partial\gamma}(J_\partial), \tag{10}$$

we can regard Eq. (9) as *bulk input space* and Eq. (10) as *boundary output space*, and define the map action similarly to the fixed-spin case.

3.2. *Isometry condition*

A necessary condition for the bulk and boundary information in $|\psi_\gamma\rangle$ to be holographically related is that the map $M[\psi_\gamma]$ is an isometry. As we are going to show, this property can be verified by studying the entanglement entropy of the reduced bulk state. We focus, for simplicity, on the spin-fixed case. Let $\{|\vec{\iota} = \iota_1 ... \iota_N\rangle\}$ be a basis of $\mathcal{H}_{\dot{\gamma}}(J)$; by writing the state $|\psi_\gamma\rangle$ in terms of the corresponding map (denoted with M only, for brevity),

$$|\psi_\gamma\rangle = \sum_{\vec{\iota}} (M \otimes \mathbb{I}) |\vec{\iota}\rangle \otimes |\vec{\iota}\rangle, \tag{11}$$

we obtain the following expression for the (normalised) reduced bulk state:

$$\rho_{\dot{\gamma}} = \mathrm{Tr}_{\partial\gamma} \left(\frac{|\psi_\gamma\rangle\langle\psi_\gamma|}{D_{\dot{\gamma}}} \right) = \frac{1}{D_{\dot{\gamma}}} \sum_{\vec{\iota}\vec{\iota}'} \left(M^\dagger M\right)^*_{\vec{\iota}\vec{\iota}'} |\vec{\iota}\rangle\langle\vec{\iota}'| \tag{12}$$

where $D_{\dot{\gamma}} = \prod_v D_{\vec{j}_v}$ is the dimension of the bulk, with $D_{\vec{j}_v}$ dimension of the intertwiner at vertex v. From Eq. (12) we can easily see that the isometry condition $M^\dagger M = \mathbb{I}$ translates into the requirement that the bulk state is maximally mixed, i.e. $\rho_{\dot{\gamma}} = \frac{\mathbb{I}}{D_{\dot{\gamma}}}$, namely that it has maximum entropy.

4. Investigating holography in spin networks via random tensor networks

We focus on states having the form

$$|\phi_\gamma\rangle = \left(\bigotimes_{e\in L}\langle e|\right)\bigotimes_v |f_v\rangle, \tag{13}$$

with the single-vertex wavefunctions $f_1, ..., f_N$ chosen randomly according to a uniform probability distribution, corresponding to a random tensor network; moreover, we assume that in γ each vertex has at most one boundary edge.

Our goal is to identify the combinatorial and dimensional features of $|\phi_\gamma\rangle$ that make the corresponding bulk-to-boundary map isometric; to accomplish this, we check under which conditions the reduced bulk state has maximum entropy.

4.1. *Rényi-2 entropy of random tensor networks*

We start by illustrating how to compute the average Rényi-2 entropy of a generic region of the entanglement graph in Eq. (13), consisting of a boundary portion $A \in \partial\gamma$ and a bulk portion $\Omega \in \dot{\gamma}$, under the assumption that the random wavefunctions f_v are picked on specific values of the edge spins. Through the replica trick, and in the large-spins regime, the average Rényi-2 entropy of $\rho_{A\cup\Omega}$, where $\rho = |\phi_\gamma\rangle\langle\phi_\gamma|$, is given by

$$\overline{S_2(\rho_{A\cup\Omega})} \approx -\log\left(\frac{\overline{Z_1}}{\overline{Z_0}}\right) \tag{14}$$

with

$$\overline{Z_0} = \mathrm{Tr}\left[\rho_L^{\otimes 2}\bigotimes_v \overline{\rho_v^{\otimes 2}}\right], \quad \overline{Z_1} = \mathrm{Tr}\left[\rho_L^{\otimes 2}\bigotimes_v \overline{\rho_v^{\otimes 2}}\,\mathcal{S}_{A\cup\Omega}\right], \tag{15}$$

where $\rho_L = \bigotimes_{e\in L}|e\rangle\langle e|$ and $\rho_v = |f_v\rangle\langle f_v|$, the average $\overline{\rho_v^{\otimes 2}}$ is computed by randomizing over vertex wavefunctions with a uniform probability measure, and $\mathcal{S}_{A\cup\Omega}$ is the copy-swap operator on $A\cup\Omega$. The characterizing aspects of the calculation are summarized hereafter. The randomization over vertex wavefunctions yields, by the Schur's lemma, a linear combination of the identity and the copy-swap operator for each vertex v; the quantities $\overline{Z_1}$ and $\overline{Z_0}$ thus involve a sum of 2^N terms, and can be written as partition functions for configurations of two-level spins, $\sigma_v = \pm 1$ with $v = 1, ..., N$, controlling the presence ($\sigma = -1$) or absence ($\sigma_v = +1$) of a copy-swap operator for each vertex v in all terms of the resulting sum. Such partition functions also depend on two-level "pinning fields": boundary fields, $\mu_e = \pm 1$ with $e \in \partial\gamma$, each one controlling the presence ($\mu_e = -1$) or absence ($\mu_e = +1$) of a copy-swap operator on the corresponding boundary edge e, potentially contained in $\mathcal{S}_A$; and bulk fields, $\nu_v = \pm 1$ with $v = 1, ..., N$, each one controlling the presence ($\nu_v = -1$)

or absence ($\nu_v = +1$) of a copy-swap operator on the corresponding vertex v, potentially deriving from $\mathcal{S}_\Omega$. As a result, $\overline{Z_0}$ and $\overline{Z_1}$ turn out to be particular values of the partition function

$$\overline{Z}\left(\vec{\mu},\vec{\nu}\right) = \sum_{\vec{\sigma}} e^{-\mathcal{A}[\vec{\mu},\vec{\nu}](\vec{\sigma})} \tag{16}$$

where

$$\begin{aligned}\mathcal{A}[\vec{\mu},\vec{\nu}]\left(\vec{\sigma}\right) = &-\frac{1}{2}\sum_{e^i_{vw}\in L}\left(\sigma_v\sigma_w - 1\right)\log d_{j^i_{vw}} - \frac{1}{2}\sum_{e\in\partial\gamma}\left(\sigma_v\mu_e - 1\right)\log d_{j_e} \\ &- \frac{1}{2}\sum_v(\sigma_v\nu_v - 1)\log D_{\vec{j}_v} + k\end{aligned} \tag{17}$$

is an Ising-type action (k is a constant), having a parametric dependence on the pinning fields $\vec{\mu} = \{\mu_e | e \in \partial\gamma\}$ and $\vec{\nu_v} = \{\nu_v | v \in V\}$; specifically, $\overline{Z_0}$ is given by Eq. (16) with all pinning fields pointing up (i.e. $\mu_e = +1$ for all e, and $\nu_v = +1$ for all v), while $\overline{Z_1}$ corresponds to the case in which the pinning fields within $A \cup \Omega$ are flipped down (i.e. $\mu_e = -1$ if and only if $e \in A$, and $\nu_v = -1$ if and only if $v \in \Omega$). By defining the free energies $F_0 = -\log\overline{Z_0}$ and $F_1 = -\log\overline{Z_1}$, the average Rényi-2 entropy is then given by $\overline{S_2(\rho_{A\cup\Omega})} = F_1 - F_0$.

We consider the large-spins regime (corresponding to a low-temperature regime of the Ising model), in which the partition functions are estimated by the contribution of the configuration $\vec{\sigma}$ that minimizes the action, i.e.

$$F(\vec{\mu},\vec{\nu}) = -\log\overline{Z}(\vec{\mu},\vec{\nu}) \approx \min_{\vec{\sigma}}\mathcal{A}[\vec{\mu},\vec{\nu}](\vec{\sigma}) \tag{18}$$

Note that F_0 is independent of the particular reduction of ρ we want to compute the entropy of; in fact, all pinning fields in $\overline{Z_0}$ take the "neutral value" $+1$. In the low-temperature regime all Ising spins get aligned with them, reaching the all-up configuration that minimizes $\mathcal{A}_0$, which takes the value $\mathcal{A}_{0\min} = k$. Therefore, from Eq. (18), $F_0 = k$. Since what matters is the difference $F_1 - F_0$, the constant k can be set equal to zero, and the calculation of $\overline{S_2(\rho_{A\cup\Omega})}$ thus be reduced to that of F_1 via Eq. (18), with $k = 0$ in the action.

4.2. *Checking the isometry condition for bulk-to-boundary maps*

We exploit the above technique to compute the average Rényi-2 entropy of the reduced bulk state: $\overline{S_2(\rho_{\dot{\gamma}})}$. This requires setting $A = \emptyset$ and $\Omega = \dot{\gamma}$; the partition function $\overline{Z_1}$ is thus given by Eq. (16) with all boundary fields $\vec{\mu}$ equal to $+1$, and all bulk fields $\vec{\nu}$ equal to -1.

4.2.1. *Homogeneous case*

In the homogeneous case, i.e. when all spins on the graph take the same value j, the quantity $\log d_j = \beta$ plays the role of inverse Ising temperature, as can be seen

by expressing the action as $\mathcal{A}(\vec{\sigma}) = \beta H(\vec{\sigma})$, with

$$H(\vec{\sigma}) = -\frac{1}{2}\left[\sum_{e^i_{vw}\in L}(\sigma_v\sigma_w - 1) + \sum_{e\in\partial\gamma}(\sigma_v - 1) + \frac{\log D_j}{\beta}\sum_v(\sigma_v\nu_v - 1)\right], \quad (19)$$

where D_j is the dimension of an intertwiner recoupling d j-spins; we then obtain, for the partition function of Eq. (16), the form $\overline{Z} = \sum_{\vec{\sigma}} e^{-\beta H(\vec{\sigma})}$ (we removed the dependence on $\vec{\mu}$, $\vec{\nu}$ for brevity).

In the range $\beta \gg 1$ we have that $\overline{Z} \approx \exp\{-\beta \min_{\vec{\sigma}} H(\vec{\sigma})\}$, and the free energy is thus given by $F \approx \beta \min_{\vec{\sigma}} H(\vec{\sigma})$. In this range the maximum value of the entropy is obtained when the free energy F_1 is estimated through the configuration with all Ising spins pointing up (in short, $\vec{\sigma} = \uparrow$); in fact,

$$\overline{S_2(\rho_{\dot{\gamma}})} = F_1 - F_0 \approx \beta H_1(\vec{\sigma} = \uparrow) = \log(D_j^N), \quad (20)$$

which is the maximum entropy the reduced bulk state can have, as D_j^N is the bulk dimension in the homogeneous case.

Based on this consideration, we can translate the isometry condition into the requirement of stability of the all-up configuration:

$$H_1(\vec{\sigma}) > H_1(\vec{\sigma} = \uparrow) \quad \forall\vec{\sigma} \quad (21)$$

By inserting Eq. (19) into Eq. (21), and denoting by $X(\vec{\sigma})$ the region that, for a given configuration $\vec{\sigma}$, contains the Ising spins that point down (i.e. $\sigma_v = -1$ if and only if $v \in X(\vec{\sigma})$), the stability condition of Eq. (21) can be written as follows:

$$|\partial X(\vec{\sigma})| \log d_j > |X(\vec{\sigma})| \log D_j \quad \forall\vec{\sigma} \quad (22)$$

We are interested in graphs made of 4-valent vertices; in that case $D_j = d_j = 2j + 1$, and Eq. (22) becomes

$$|\partial X(\vec{\sigma})| > |X(\vec{\sigma})| \quad \forall\vec{\sigma}, \quad (23)$$

which cannot be satisfied for $\vec{\sigma} = \downarrow$, as $X(\vec{\sigma} = \downarrow) = \dot{\gamma}$ and every vertex has at most one boundary link. This implies that no bulk-to-boundary map associated to an homogeneous graph, made of 4-valent vertices and with at most one boundary link for each vertex, can be an isometry.

4.2.2. *Inhomogeneous case*

In the generic, inhomogeneous case, we can still put in evidence the parameter $\beta = \log d_j$, where j is now the average spin, and write the partition function as $\overline{Z} = \sum_{\vec{\sigma}} e^{-\beta H(\vec{\sigma})}$ with

$$H(\vec{\sigma}) = -\frac{1}{2}\sum_{e^i_{vw}\in L}(\sigma_v\sigma_w - 1)\frac{\log d_{j^i_{vw}}}{\beta} - \frac{1}{2}\sum_{e\in\partial\gamma}(\sigma_v - 1)\frac{\log d_{j_e}}{\beta} - \frac{1}{2}\sum_v(\sigma_v\nu_v - 1)\frac{\log D_{\vec{j}_v}}{\beta} \quad (24)$$

Similarly to the homogeneous case, when the partition function $\overline{Z_1}$ is dominated by the all-up configuration, and so $F_1 \approx \beta H_1(\vec{\sigma} = \uparrow)$, the entropy takes its maximum value: $\overline{S_2(\rho_{\hat{\gamma}})} = \log\left(\prod_v D_{\vec{j}_v}\right)$. We can therefore express the isometry condition of the bulk-to-boundary map as stability condition of the $\vec{\sigma} = \uparrow$ configuration; by making use of the $\vec{\sigma}$-dependent region $X(\vec{\sigma})$ previously defined, that takes the following form:

$$\sum_{e\in\partial X(\vec{\sigma})} \log d_{j_e} > \sum_{v\in X(\vec{\sigma})} \log D_{\vec{j}_v} \quad \forall\vec{\sigma} \tag{25}$$

Let us now focus on graphs with 4-valent vertices. By renaming the spins of every vertex with labels a, b, c, d according to their increasing order, i.e. $j^a \leq j^b \leq j^c \leq j^d$, we can write the intertwiner dimension as follows: $D_{j^a j^b j^c j^d} = \min\{j^a + j^d, j^b + j^c\} - \Delta + 1$, where $\Delta = j^d - j^a$ is the difference between the maximum and the minimum spin[a]. Equation (25) then becomes

$$\sum_{e\in\partial X(\vec{\sigma})} \log d_{j_e} > \sum_{v\in X(\vec{\sigma})} \log\left(\min\{j_v^a + j_v^d, j_v^b + j_v^c\} - \Delta_v + 1\right) \quad \forall\vec{\sigma} \tag{26}$$

To highlight the trend of the isometric character of the map as a function of Δ, we consider the scenarios corresponding to the extreme values of Δ:

(i) $\Delta_v = 0\ \forall v$: homogeneous case, the bulk-to-boundary map is not an isometry
(ii) $\Delta_v = j_v^b + j_v^c - 1\ \forall v$: the stability condition becomes

$$\sum_{e\in\partial X(\vec{\sigma})} \log d_{j_e} > |X(\vec{\sigma})| \log 2 \quad \forall\vec{\sigma} \tag{27}$$

which is satisfied for sufficiently high values of the edge spins.

5. Conclusions and outlook

We investigated the holographic features of spin networks by regarding them as emerging from the entanglement of individual vertices, in the group field theory approach to quantum gravity. This enabled us to highlight the entanglement structure of spin network states, and write them in the language of tensor networks. Based on this correspondence, we showed that spin networks naturally define bulk-to-boundary maps, whose properties can efficiently be investigated via tensor network techniques. Specifically, we used the latter to identify which (entanglement-induced) features of a spin network can give rise to an isometric mapping of quantum data from the bulk to the boundary, which is the pre-condition for an holographic behaviour. In the case of graphs made of four-valent vertices (which is the interesting one for 4D quantum gravity) we found that homogeneity of the spin assignment to the spin network prevents the latter to be an isometric bulk-to-boundary map,

[a]Note that $\Delta \leq j^b + j^c$; however, since for $\Delta = j^b + j^c$ the intertwiner degree of freedom is suppressed, in the following we assume $\Delta \leq j^b + j^c - 1$.

while increasing the range of spins attached to each vertex favours the attainment of the isometry condition. The results presented here are part of Ref. 21, which also contains an extension of the above analysis to states involving spin superposition, with randomization diagonal with respect to the sum over spins. Importantly, a full generalization to the case of spin superposition, as well as graph superposition, is crucial for a promotion of our results to the dynamical level, since both features can be produced by the quantum dynamics.

Besides contributing to the understanding of the entanglement origin of holography in quantum gravity, our work is meant to be the starting point of a research direction that exploits quantum information techniques in a quantum gravity framework having a purely information-theoretic characterisation, i.e. the one in which spin networks are regarded as generalised tensor networks in the group field theory formalism for quantum gravity. In fact, we have recently used the techniques presented here to study how the boundary entropy depends on the bulk entanglement, specifically that determining the graph connectivity (internal-link entanglement) and that among intertwiner degrees of freedom[23]; in particular, we have analysed under which conditions the latter gives rise to a black hole-like region, analogously to what has been shown in Ref. 16. These methods and results, combined with the typicality approach to the study of the local behaviour of spin networks,[24,25] should pave the way for an information-theoretic characterization of black hole horizons and thus provide important insights into the structure of quantum black holes.

References

1. X.-L. Qi, Does gravity come from quantum information?, *Nature Phys.* **14**, 984 (2018).
2. J. M. Maldacena, The Large N limit of superconformal field theories and supergravity, *Int. J. Theor. Phys.* **38**, 1113 (1999).
3. S. Ryu and T. Takayanagi, Holographic derivation of entanglement entropy from AdS/CFT, *Phys. Rev. Lett.* **96**, p. 181602 (2006).
4. S. Ryu and T. Takayanagi, Aspects of Holographic Entanglement Entropy, *JHEP* **08**, p. 045 (2006).
5. M. Van Raamsdonk, Comments on quantum gravity and entanglement (7 2009).
6. M. Van Raamsdonk, Building up spacetime with quantum entanglement, *Gen. Rel. Grav.* **42**, 2323 (2010).
7. D. Oriti, Levels of spacetime emergence in quantum gravity (7 2018).
8. D. Oriti, The microscopic dynamics of quantum space as a group field theory, 257 (10 2011).
9. T. Krajewski, Group field theories, *PoS* **QGQGS2011**, p. 005 (2011).
10. S. Carrozza, Flowing in Group Field Theory Space: A Review, *SIGMA* **12**, p. 070 (2016).
11. S.-J. Ran, E. Tirrito, C. Peng, X. Chen, L. Tagliacozzo, G. Su and M. Lewenstein, Lecture Notes of Tensor Network Contractions, *arXiv e-prints* (August 2017).
12. D. Perez-Garcia, F. Verstraete, M. M. Wolf and J. I. Cirac, Matrix Product State Representations, *arXiv e-prints* (August 2006).
13. R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Annals of Physics* **349**, 117 (2014).

14. B. Swingle, Entanglement Renormalization and Holography, *Phys. Rev. D* **86**, p. 065007 (2012).
15. W. Helwig, W. Cui, A. Riera, J. I. Latorre and H.-K. Lo, Absolute Maximal Entanglement and Quantum Secret Sharing, *Phys. Rev. A* **86**, p. 052335 (2012).
16. P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, Holographic duality from random tensor networks, *JHEP* **11**, p. 009 (2016).
17. E. Colafranceschi and D. Oriti, Quantum gravity states, entanglement graphs and second-quantized tensor networks, *JHEP* **07**, p. 052 (2021).
18. G. 't Hooft, Dimensional reduction in quantum gravity, *Conf. Proc. C* **930308**, 284 (1993).
19. L. Susskind, The World as a hologram, *J. Math. Phys.* **36**, 6377 (1995).
20. I. Cirac, D. Perez-Garcia, N. Schuch and F. Verstraete, Matrix Product States and Projected Entangled Pair States: Concepts, Symmetries, and Theorems (11 2020).
21. E. Colafranceschi, G. Chirco and D. Oriti, Holographic maps from quantum gravity states as tensor networks (5 2021).
22. D. Oriti, Group field theory as the 2nd quantization of Loop Quantum Gravity, *Class. Quant. Grav.* **33**, p. 085005 (2016).
23. G. Chirco, E. Colafranceschi and D. Oriti, in preparation.
24. F. Anzà and G. Chirco, Fate of the Hoop Conjecture in Quantum Gravity, *Phys. Rev. Lett.* **119**, p. 231301 (2017).
25. F. Anzà and G. Chirco, Typicality in spin-network states of quantum geometry, *Phys. Rev. D* **94**, p. 084047 (2016).

Minimal length discretization and properties of modified metric tensor and geodesics

Abdel Nasser Tawfik

Egyptian Center for Theoretical Physics (ECTP), Cairo, Egypt
World Laboratory for Cosmology and Particle Physics (WLCAPP), Cairo, Egypt
E-mail: atawfik@cern.ch

Fady T. Farouk, F. Salah Tarabia and Muhammad Maher

Helwan University, Faculty of Science, Physics Department, 11792 Ain Helwan, Cairo, Egypt

We argue that the minimal length discretization generalizing the Heisenberg uncertainty principle, in which the gravitational impacts on the non–commutation relations are thoughtfully taken into account, radically modifies the spacetime geometry. The resulting metric tensor and geodesic equation combine the general relativity terms with additional terms depending on higher–order derivatives. Suggesting solutions for the modified geodesics, for instance, isn't a trivial task. We discuss on the properties of the resulting metric tensor, line element, and geodesic equation.

Keywords: Quantum gravity, Noncommutative geometry, Relativity and gravitation, Line element, Metric tensor, Geodesic equation, Generalized uncertainty principle

1. Introduction

Suggesting a consistent theory for quantum gravity is an ultimate goal of recent research aiming at unifying quantum mechanics (QM) with the theory of general relativity (GR). The quantum nature of gravity is still an open question in physics. It turns to form an enigma occupying the minds of many physicists for many decades, for instance, so–far there is no scientific explanation for the origin of the gravitational fields and how particles behave inside them! This makes the gravity foundamentally different from the other fources of nature. There were various attempts to reconcile some principles of GR in a quantum framework.[1,2] The quantum gravity - on one hand - is conjectured to add new elements to GR and QM. On the other hand, the Einstein field equations combine classical geometry with quantum stress–energy–momentum tensor! For QM, we recall the possible modification in the Heisenberg uncertainty principle due to gravitational effects[a], known as generalized uncertainty princile (GUP).[3,4] The present work introduces some insights on the possible impacts of GUP on the metric tensor and the geodesic equation and their properties.

Couple decades ago, the line element modification and the gravity quantization have been suggested in litrature.[5–7] A review of these models could be found in Refs.[8–10]. An extended class of the metric tensors which are functions of a field of an

[a]They weren't taken into considration in the original Heisenberg uncertainty principle.

internal vectors $y^\alpha(x)$, $g_{\mu\nu}(x^\alpha \cdot y^\alpha(x))$, was also discussed in litrature.[11] Accordingly, the origin of the gravitational field of spin-1 massless particle could be determined.[11] Regarding the equivalence principle of gravitationan and enirtial mass in extended gravity and non–geodesic motion, a machian request was scratched,[12] where a direct coupling between the Ricci curvature scalar and the matter Lagrangian could be fixed in cosmological observations.[12]

The present script aims at tackling the long–standing essential puzzle that the Einstein field equations relate *non–quantized* semi–Riemannian geometry expressed in Ricci and Einstein tensors, which are basically depending on the metric tensor $g_{\mu\nu}(x^\alpha)$, with the *full–quantized* stress–energy-momentum tensor $T_{\mu\nu}$.[13] The approach applied here is the finite minimal length uncertainty obtained from GUP. GUP was inspired by black hole physics, doubly special relativity, and string theory.[3,4] It is worthy highligting that GUP is comparable with the Planck length, where quantum fluctuations in *quasi–quantized* manifold likely emerge.[14–16]

The modifications in the metric tensor and geodesic equation due to GUP shall be introduced. Accordingly, we believe that the present study offers a theoretical framework for the observations that the universe acceleratingly expands. Alongside forces deriving this type of expansion, such as dark energy[16] and cosmological constant,[17] we suggest that it might be also originated to interdependence of GR on QM, especially at the Planck scale.

The present paper is organized as follows. Section 2 reviews the generalized uncertainly principle and minimal length uncertainty. The discretized metric tensor and its properties shall be discussed in section 3. The discretized geodesic equation and its properties shall be introduced in section 4. Section 5 is devoted to the final conclusions.

2. Generalized uncertainly principle and minimal length uncertainty

The Heisenberg uncertainty principle (HUP) is a recipe on how the uncertainties are constrained in the quantum non–commutation relations of length and momentum operators, for instance. This was originally suggested in absence of any gravitational field. Analyzing HUP in gravitational fields comes up with modifications, know as GUP. Such a phenomenological aspect can be originated in black hole physics, doubly special relativity, and string theory.[3,4] Phenomenology because so–far there is no theoretical explanation of the gravitational fields origin making gravity substantially different from the other three forces of nature. A finite minimal length was suggested as,[18]

$$\Delta x\, \Delta p \geq \frac{\hbar}{2}\left[1+\beta(\Delta p)^2\right], \tag{1}$$

where Δx and Δp are length and momentum uncertainties, respectively. $\beta = \beta_0(\ell_p/\hbar)^2 = \beta_0/(m_p c)^2$, is the GUP parameter with $\ell_p = \sqrt{\hbar G/c^3} = 1.977 \times 10^{-16}$ GeV^{-1} the Planck length and $m_p = \sqrt{\hbar c/G} = 1.22 \times 10^{19}$ GeV$/c^2$ the

Planck mass. An upper bound on β_0 is based on astronomical observations, such as, $\beta_0 \lesssim 5.5 \times 10^{60}$ based on recent gravitational waves observations. From Eq. (1), then the minimum length uncertainty reads

$$\Delta x_{\min} \approx \hbar\sqrt{3\beta_0} = \ell_p\sqrt{\beta_0}. \tag{2}$$

The experimental estimations[19] of β_0 range from $\sim 10^{21}$ to $\sim 10^{39}$. Another GUP approach motivated by quantum deformation of the Pioncarre algebra[20,21] suggests that the minimal length associated with is given as

$$\Delta x_{\min} \simeq \frac{\hbar G}{\gamma_0} = \frac{\ell_p^2}{\gamma_0}, \tag{3}$$

where γ_0 is photon's wavelength sent from infinity to be absorbed in a black hole. Its experimental estimations[19] range from $\sim 10^{26}$ to $\sim 10^{34}$.

The UV/IR correspondence can be applied on various phenomena in which short and long distance physics play a characterized role. For example, gravity[22] and the "*deformed*" commutation relations have features of UV/IR correspondence.[23] Therefore, we assume that the minimal length discretization could be analyzed from the UV/IR correspondence.[10] Indeed, Δx rapidly increases (IR) as the Δp increases beyond the Planck scale (UV).[24–26]

Analogous to expression (1), the canonical non–commutation relation of length and momentum quantum operators is given as

$$[\hat{x}_i\,, \hat{p}_j] \geq \delta_{ij} i\,\hbar\left(1 + \beta p^2\right), \tag{4}$$

where $p^2 = g_{ij}p^{0i}\,p^{0j}$ and g_{ij} is Minkowski spacetime metric tensor, for instance $(-,+,+,+)$. Both length and momentum operators, respectively, read

$$\hat{x}_i = \hat{x}_{0i}(1 + \beta p^2), \tag{5}$$

$$\hat{p}_j = \hat{p}_{0j}, \tag{6}$$

in which $\hat{x}_{0i}$ and $\hat{p}_{0j}$ can be derived from the corresponding non–communitation relation,

$$[\hat{x}_{0i}, \hat{p}_{0j}] = \delta_{ij} i\,\hbar. \tag{7}$$

3. Discretized metric tensor and its properties

With the quantum non–commutation operations discussed in section 2 the four–dimensional Minkowskian manifold of the line element can be given as

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu, \tag{8}$$

where μ, ν, and $\lambda = 0, 1, 2, 3$. This could be extended to eight–dimensional spacetime tangent bundle with coordinates $x^A = (x^\mu(\zeta^a), \beta\dot{x}^\mu(\zeta^a))$, where $\dot{x}^\mu = dx^\mu/d\zeta^\mu$.[27] Obviously, the line element reads

$$d\tilde{s}^2 = g_{AB}\,dx^A\,dx^B, \tag{9}$$

where $x^A = x^A(\zeta^\mu)$, $g_{AB} = g_{\mu_\nu} \otimes g_{\mu\nu}$, and the indices A and $B = 0, 1, \ldots, 7$. With some assumptions and approximations, this manifold could be reformulated as an effective four–dimensional spacetime manifold,

$$\tilde{g}_{\mu\nu} = g_{AB} \frac{\partial x^A}{\partial \zeta^\mu} \frac{\partial x^B}{\partial \zeta^\nu} \simeq g_{ab} \Big[\frac{\partial x^a}{\partial \zeta^\mu} \frac{\partial x^b}{\partial \zeta^\nu} + \beta \frac{\partial \dot{x}^a}{\partial \zeta^\mu} \frac{\partial \dot{x}^b}{\partial \zeta^\nu} \Big] \simeq \big(1 + \beta \ddot{x}^\lambda \ddot{x}_\lambda\big)\, g_{\mu\nu}. \quad (10)$$

$\ddot{x}^\mu = \partial \dot{x}^\mu / \partial \zeta^\mu$ is four–dimensional acceleration. The indices a, and $b = 0, 1, \ldots, 7$.

It is now proper to highlight a number of the discretized metric tensor properties. First, we distinguish between flat and curved spacetime:

- For flat spacetime, $g_{\mu\nu} = \eta_{\mu\nu}$ and the modified metric tensor reads

$$\tilde{g}_{\mu\nu} \simeq \eta_{\mu\nu} + \beta \ddot{x}^\lambda \ddot{x}_\lambda \eta_{\mu\nu} \simeq \eta_{\mu\nu} + h_{\mu\nu}, \quad (11)$$

  where $h_{\mu\nu} = \beta \ddot{x}^\lambda \ddot{x}_\lambda \eta_{\mu\nu}$. The quantum fluctuations from which the minimal length uncertainty likely emerges are compassed in $h_{\mu\nu}$. In other words, both GR metric tensor and this modified one are only distinguishable, at finite $h_{\mu\nu}$.
- For curved spacetime, the modified metric tensor reads

$$\tilde{g}_{\mu\nu} \simeq g_{\mu\nu} + \beta \ddot{x}^\lambda \ddot{x}_\lambda g_{\mu\nu}. \quad (12)$$

  Again, $\beta \ddot{x}^\lambda \ddot{x}_\lambda g_{\mu\nu}$ represents the minimal length contributions.

The $\tilde{g}_{\mu\nu}$, Eq. (12), seems to play the same role as that of $g_{\mu\nu}$, namely it turns a covariant tensor into a contravariant tensor and vice versa. $\tilde{g}_{\mu\nu}$ is apparently symmetric in its indices μ and ν as its components are likely equal upon exchanging the indices. The $\tilde{g}_{\mu\nu}$ symmetric properties are invariant upon basis transformation only if its indices are of the same type; variant or covariant. For spacetime represented by a four–dimensional differentiable manifold $\mathcal{M}$, the metric tensor, $\tilde{g}_{\mu\nu}$, Eq. (12), is covariant, second–degree, symmetric on $\mathcal{M}$. Such manifold is Lorentzian. In local coordinates x^μ, the metric can be expressed in dx^μ (one–form gradients of the scalar coordinate fields x^μ), and the coefficients $\tilde{g}_{\mu\nu}$ (16 real–valued functions). If the local coordinates are specified, the metric can be written as a 4×4 symmetric matrix with the coefficients $\tilde{g}_{\mu\nu}$. If the matrix is non–singular having non–vanishing determinant, $\tilde{g}_{\mu\nu}$ is non–degenerate. The Lorentzian signature of $\tilde{g}_{\mu\nu}$ obviously implies that the matrix has one negative and three positive eigenvalues. Under a change of the local coordinates $x^\mu \to x^{\bar{\mu}}$, we obtain

$$\tilde{g}_{\bar{\mu}\bar{\nu}} = \frac{\partial x^\alpha}{\partial x^{\bar{\mu}}} \frac{\partial x^\beta}{\partial x^{\bar{\nu}}}\, \tilde{g}_{\alpha\beta} = \Theta^\alpha_{\bar{\mu}}\, \Theta^\beta_{\bar{\nu}}\, \tilde{g}_{\alpha\beta} \quad (13)$$

If dx^μ are regarded as the components of an infinitesimal coordinate displacement four–vector, the metric determines the invariant square of an infinitesimal line element. Also, it is apparent that the principle of the general covariance is satisfied even in absence of gravitational effects on the modified Minkowski metric. Then,

the modified four–dimensional line element can be expressed as

$$d\tilde{s}^2 = g_{\mu\nu}\left(dx^\mu dx^\nu + \beta^2 d\dot{x}^\mu\, d\dot{x}^\nu\right) = ds^2 + \beta^2 \ddot{x}^\lambda \ddot{x}_\lambda\, ds^2, \tag{14}$$

and $d\tilde{s}^2$ manifests the causal structure of the spacetime. Only timelike four–dimensional line element, $d\tilde{s}^2 < 0$, can physically covered by a massive object. If $d\tilde{s}^2 = 0$ the line element is lightlike (covered by massless photon). On the other hand, for $d\tilde{s}^2 > 0$, the four–dimensional line element is an incremental proper length, i.e., spacelike.

4. Discretized geodesic equation and its properties

The properties of the manifold including geodesics in GR [$\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta}\dot{x}^\mu\dot{x}^\nu = 0$ with $\dot{x}^\mu = \partial x^\mu/\partial s$, Christopher symbols $\Gamma^\mu_{\alpha\beta} = (g^{\mu\gamma}/2)(\partial g_{\gamma\alpha}/\partial x^\beta) + \partial g_{\gamma\beta}/\partial x^\alpha - \partial g_{\alpha\beta}/\partial x^\gamma)$, and s is a scalar parameter of motion such as a proper time] can be generalized. For example, the notion of a "*straight line*" is to be generalized to "*curved spacetime*". The consequences of the length discretization or minimal length uncertainty based on GUP, generalize the world line of a free particle. GR assumes that gravity couldn't be regarded as a force but it should be emerged from "*classical*" curved spacetime geometry and "*quantized*" stress–energy–momentum tensor is causing the "*classical*" spacetime curvature. The approach utilized here is discretized minimal length, which modifies both metric tensor and line element.

As discussed in Ref.[10], using the variational principle and extremization of the path s_{AB}, the geodesic equation could be derived as follows.

- In flat spacetime (Minkowski space)

$$\beta\mathcal{L}\frac{d^2\dot{x}^\mu}{ds^2} - \frac{dx^\mu}{ds} + c = 0. \tag{15}$$

  Even if β, the GUP parameter vanishes, the geodesic equation reduces to that of a straight line, at constant velocity, i.e., both β terms disappear.
- and in curved spacetime

$$\begin{aligned}\frac{d^2x^\eta}{ds^2} + \Gamma^\eta_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} &= \beta\frac{d}{ds}\left(\mathcal{L}\frac{d^3x^\eta}{ds^3}\right)\\ &+ \beta g^{\eta\alpha} g_{\mu\nu,\alpha}\frac{d^2x^\mu}{d^2s}\frac{d^2x^{nu}}{ds^2}\\ &+ \beta\mathcal{L} g^{\eta\alpha} g_{\mu\alpha,\gamma}\left[\frac{dx^\gamma}{ds}\frac{d^2\dot{x}^\mu}{ds^2} + \frac{d}{ds}\left(\frac{dx^\gamma}{ds}\frac{d\dot{x}^\mu}{ds}\right)\right]\\ &+ \beta\mathcal{L} g^{\eta\alpha} g_{\mu\alpha,\gamma,\delta}\frac{dx^\delta}{ds}\frac{dx^\gamma}{ds}\frac{d\dot{x}^\mu}{ds},\end{aligned} \tag{16}$$

where the proper time can be deduced from the Lagrangian, $\tau = \int \mathcal{L}(s, \dot{x}, \ddot{x})\, ds$,

$$\frac{\partial \tau}{\partial s} = \mathcal{L}(s, \dot{x}, \ddot{x}) = -\left[g_{\mu\nu} \left(\frac{dx^\mu}{ds} \frac{dx^\nu}{ds} + \beta \frac{d\dot{x}^\mu}{ds} \frac{d\dot{x}^\nu}{ds} \right) \right]^{1/2},$$

$$\Gamma^\eta_{\mu\nu} = \frac{1}{2} g^{\eta\alpha} \left[g_{\mu\alpha,\nu} - g_{\alpha\nu,\mu} + g_{\mu\nu,\alpha} \right],$$

$$g_{\mu\nu,\alpha} = \frac{\partial g_{\mu\nu}}{\partial x^\alpha},$$

$$g_{\mu\alpha,\gamma,\delta} = \frac{\partial}{\partial x^\delta} \left(\frac{\partial g_{\mu\alpha}}{\partial x^\gamma} \right).$$

The β–terms in Eq. (16) distinguish this expression of discretized geodesics from the GR geodesics,

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0, \tag{17}$$

which could be solved when, for instance, assuring that the proper time reads $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$. Under the condition that

$$g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = \epsilon, \tag{18}$$

where $\epsilon = 0$ for photon and $\epsilon = 1$ for massive particles, the geodesic equation is supplemented. For conserved energy and angular momentum the geodesic equation can be expressed as an ordinary first–order differential equation.[28]

Solving the discretized geodesic equation, Eq. (16), is not a trivial task, because of the finite terms in rhs, which encompass higher–order derivatives.

5. Conclusions

The approach of minimal length uncertainty conserves both Christoffel connection[b] and the equivalence principle of the gravitational and inertial mass. They are not affected. This allows to recover any violations in the equivalence principle, especially as a possible reason of quantum gravity. We highlight that this conclusion contradicts previous studies.[29,30] The reason could be the rigorous procedure implemented in the present study. We have first determined the modified metric tensor. Accordingly, we have obtained the modified line element. Both are key principles of the proposed theory of discretized geometry. We could also derive the corresponding geodesics, in which the appearance of higher–order derivatives substantially distinguishes this expression from the GR geodesics.

The modification to the line element is summarized in the term $\beta^2 \ddot{x}^\lambda \ddot{x}_\lambda$, which highlights the essential contributions added by the acceleration of the test particle.

[b]The metric connection characterizes the affine connection to a manifold provided with the metric. This allows lengths, for instance, to be measured on that manifold.

We conclude that the length discretization ensures that the line element wouldn't only expand but apparently accelerates. The metric tensor is also modified by the term $\beta\ddot{x}^{\lambda}\ddot{x}_{\lambda}$. The properties of the modified metric tensor has been elaborated.

We find that the modified geodesics is not only providing the acceleration of a test particle inside the gravitational field. Higher–order derivatives are also present,[31] namely the jerk, $\dddot{x}^{\mu}$, which in turn is derived from the time derivative of the acceleration. In contract to GR geodesics, solving the modified geodesics isn't a trivial task. The corresponding differential equations are complex and non–linear.

Acknowledgments

AT is very grateful to organizers of sixteenth Marcel Grossmann meeting - MG16 for their kind invitation and financial support!

References

1. C. Rovelli and L. Smolin, Loop Space Representation of Quantum General Relativity, *Nucl. Phys. B* **331**, 80 (1990).
2. J. F. Donoghue, General relativity as an effective field theory: The leading quantum corrections, *Phys. Rev. D* **50**, 3874 (1994).
3. A. N. Tawfik and A. M. Diab, Generalized Uncertainty Principle: Approaches and Applications, *Int. J. Mod. Phys. D* **23**, p. 1430025 (2014).
4. A. N. Tawfik and A. M. Diab, Review on Generalized Uncertainty Principle, *Rept. Prog. Phys.* **78**, p. 126001 (2015).
5. A. Feoli, G. Lambiase, G. Papini and G. Scarpetta, Schwarzschild field with maximal acceleration corrections, *Phys. Lett. A* **263**, 147 (1999).
6. S. Capozziello, A. Feoli, G. Lambiase, G. Papini and G. Scarpetta, Massive scalar particles in a modified Schwarzschild geometry, *Phys. Lett. A* **268**, 247 (2000).
7. V. Bozza, A. Feoli, G. Lambiase, G. Papini and G. Scarpetta, Maximal acceleration effects in Kerr space, *Phys. Lett. A* **283**, 53 (2001).
8. P. O. Hess, M. Schafer and W. Greiner, *Pseudo-complex General Relativity* (Springer, Heidelberg, 2015).
9. P. Hess, Alternatives to Einstein's General Relativity Theory, *Prog. Part. Nucl. Phys.* **114**, p. 103809 (2020).
10. A. N. Tawfik, A. M. Diab, S. Shenawy and E. A. El Dahab, Consequences of minimal length discretization on line element, metric tensor and geodesic equation, *Astron. Nachr.* **342**, 54 (2021).
11. C. G. Oliveira, On the properties of an extended class of metric tensors in relativity, *Int. J. Theor. Phys.* **24**, p. 1081 (1985).
12. I. Licata, C. Corda and E. Benedetto, A machian request for the equivalence principle in extended gravity and nongeodesic motion, *Grav. Cosmol.* **22**, 48 (2016).
13. H. Stephani, D. Krämer, M. MacCallum, C. Hoenselaers and E. Herlt, *Exact solutions of Einstein's field equations; 2nd ed.* (Cambridge Univ. Press, Cambridge, 2003).
14. F. Karolyhazy, Gravitation and quantum mechanics of macroscopic objects, *Nuovo Cimento (Italy) Divided into Nuovo Cimento A and Nuovo Cimento B* **(10), 42A** (1966).
15. A. Tawfik and A. Diab, Emergence of cosmic space and minimal length in quantum gravity: A large class of spacetimes, equations of state, and minimal length approaches, *Indian J. Phys.* **90**, 1095 (2016).

16. A. N. Tawfik and E. A. El Dahab, Review on Dark Energy Models, *Grav. Cosmol.* **25**, 103 (2019).
17. P. Peebles and B. Ratra, The Cosmological Constant and Dark Energy, *Rev. Mod. Phys.* **75**, 559 (2003).
18. A. Kempf, G. Mangano and R. B. Mann, Hilbert space representation of the minimal length uncertainty relation, *Phys. Rev. D* **52**, 1108 (1995).
19. D. Gao and M. Zhan, Constraining the generalized uncertainty principle with cold atoms, *Phys. Rev. A* **94**, p. 013607 (2016).
20. M. Maggiore, A Generalized uncertainty principle in quantum gravity, *Phys. Lett. B* **304**, 65 (1993).
21. M. Maggiore, Quantum groups, gravity and the generalized uncertainty principle, *Phys. Rev. D* **49**, 5182 (1994).
22. M. G. Jackson, Gravity from a modified commutator, *Int. J. Mod. Phys. D* **14**, 2239 (2005).
23. M. Jackson, Gravity from a modified commutator, *International Journal of Modern Physics D* **14**, 2239 (2005).
24. J. M. Maldacena, The Large N limit of superconformal field theories and supergravity, *Int. J. Theor. Phys.* **38**, 1113 (1999).
25. S. Gubser, I. R. Klebanov and A. M. Polyakov, Gauge theory correlators from non-critical string theory, *Phys. Lett. B* **428**, 105 (1998).
26. E. Witten, Anti-de Sitter space and holography, *Adv. Theor. Math. Phys.* **2**, 253 (1998).
27. H. E. Brandt, Causal Domain of Minkowski-spacetime Tangent Bundle, *Found. Phys. Lett.* **13**, 581 (2000).
28. E. Hackmann and C. Lammerzahl, Complete Analytic Solution of the Geodesic Equation in Schwarzschild- (Anti-) de Sitter Spacetimes, *Phys. Rev. Lett.* **100**, p. 171101 (2008).
29. S. Ghosh, Quantum Gravity Effects in Geodesic Motion and Predictions of Equivalence Principle Violation, *Class. Quant. Grav.* **31**, p. 025025 (2014).
30. F. Scardigli and R. Casadio, Is the equivalence principle violated by generalized uncertainty principles and holography in a brane-world?, *Int. J. Mod. Phys. D* **18**, 319 (2009).
31. D. Eager, A.-M. Pendrill and N. Reistad, Beyond velocity and acceleration: Jerk, snap and higher derivatives, *European Journal of Physics* **37**, p. 065008 (oct 2016).

The structure of the multiverse from the entanglement entropy

Samuel Barroso Bellido
Institute of Physics, University of Szczecin
Wielkopolska 15, 70-451 Szczecin, Poland
E-mail: samuel.barroso-bellido@usz.edu.pl

Third quantization of canonical quantum gravity allows us to consider the multiverse as the playground of a field theory of universes. In the scenario where two universes are created by analogy with pair production in quantum field theory, we analyze the entanglement entropy between them and get some conclusions about the structure of the multiverse.

Keywords: Multiverse, Entanglement Entropy, Third Quantization

1. What is Time and Space?

Canonical quantum gravity[1,2] is one of the most natural ways to venture in search of a theory which combines gravity and the quantum world. Even though it is an old topic, there are several fundamental problems one needs to sort out to get satisfactory predictions from it.[3,4] However, the theory has got quite important results we should not ignore.[2,5]

This approach to a quantum theory of gravity predicts the existence of a wave function of the universe, which describes the universe as a whole.[6,7] It is defined by the Wheeler-DeWitt equation

$$\mathcal{H}\Psi = 0, \tag{1}$$

where $\mathcal{H}$ is the Hamiltonian which considers all possible degrees of freedom from fields and gravity, and Ψ is such wave function. To assume a homogeneous and isotropic universe where there is a limited number of fields, is the simplest way to reduce the complexity of the calculations. Then, the Hamiltonian is quantized by replacement of the momenta by their associated operators, and the wave function of the universe is found solving the differential equation (1).

For instance, let us choose the proper factor ordering for the scale factor a, and the parametrization $\alpha = \ln(a)$, to write its associated momentum as[2]

$$p_a^2 = -\frac{1}{a}\frac{\partial}{\partial a}\left(a\frac{\partial}{\partial a}\right) = -\mathrm{e}^{-2\alpha}\frac{\partial^2}{\partial \alpha^2}. \tag{2}$$

A single scalar field is going to be quantized as usual by

$$p_\phi^2 = -\frac{\partial^2}{\partial \phi^2}. \tag{3}$$

Thus, the Wheeler-DeWitt equation (1) has the shape of a wave equation

$$\left[\frac{\partial^2}{\partial\alpha^2} - \frac{\partial^2}{\partial\phi^2} + f(\alpha,\phi,...)\right]\Psi(\alpha,\phi) = 0, \tag{4}$$

where $f(\alpha,\phi,...)$ contains the information of the curvature of the universe, the cosmological constant, and so on.

In this theory, *time* is not present. Usually, the scale factor, or a function like α, is the variable which replaces time, but it is an arbitrary choice that leads to physical and even philosophical problems.[8] Since the scale factor as a function of time is not an isomorphism, it is not a well-behaved parametrization of time. Expanding and contracting phases of a closed universe will be described by the same scale factor at two different times, and so big bang and big crunch will have no differences at all. In a practical sense, our universe has never gone through a phase of contraction,[9] and it looks like it will never be in such phase.[10] Hence, the parametrization could still be valid for realistic purposes. On the contrary, *space* is replaced by the scalar field, the one which has the opposite sign in the hyperbolic differential equation (4).

Third quantization of canonical quantum gravity[11–14] is the next step behind a more essential description. By means of it, our universe is seen as one of many universes of a multiverse, where a full quantum field theory takes place, the wave function of the universe plays now the role of an operator over vectors of a Fock space, and annihilation and creation operators can also be defined.

One might expect the number of universes from this picture to be infinite. All of them interacting under the rules of the field theory. Therefore, pair creation of universes is somewhat natural in this picture,[15] as it is in quantum electrodynamics, and two universes can be seen as particle and antiparticle, born at the same time and space in the multiverse (whether *time* and *space* exist or not), being entangled since then.[16] One of those two universes could be our universe. We need to assume that the moment we know as the origin of our universe is the same as that precise moment when our universe is created. Thereafter, we have not notice much difference from what the classical theory of gravity tells us,[17] so we can say that there has not been any interaction with any other universe, and therefore the trajectory through the multiverse has been unaffected. According to this, the pair has *always* been a pair, and the number of universes of this subsystem of the multiverse, has always been constant since its origin.

When considering a curved background, third quantization suffers from the lack of definition of particle as in quantum field theory in curved spacetimes,[18] and so the number of universes is variable. In general, the number of universes appears to be infinite with a certain distribution of modes,[19–21] which are defined by the proper Fock space. To solve this, we turn to an invariant representation[22,23] in which the number operator remains constant, and so the number of universes.

The quantum state at the birth of the pair[24–26] is going to be denoted by $|00\rangle$, which is the same in the natural description due to the diagonal Hamiltonian and in the invariant representation we would like to work with. The Bogoliubov

transformation between both representations at different time lets us to write the density operator $\rho = |00\rangle_{\text{inv}}\langle 00|$ in terms of the vectors of the diagonal representation, where $|00\rangle_{\text{inv}}$ is the vacuum state in the invariant representation. Finally, the entanglement entropy as the von Neumann entropy of the bipartite system is found tracing out the degrees of freedom of one of the two universes.

2. How is the Multiverse?

First of all, the quantum field theory we are used to, is not in perfect harmony with third quantization picture. By ordinary field theory, we states that fundamental particles are really fundamental, and so, they have no internal structure.

In the third quantization picture, particles are universes where we could be living *in*, where all laws of Physics seem to be satisfied. They have an internal structure. As far as it seems, we have no idea about how this multiverse really affects the internal structure of the universes. It is, for example, how the interactions between universes in the multiverse could change the dynamics of our universe. Some attempts have been done previously,[27] but it is not conclusive since the power spectrum of the cosmic microwave background is only affected by an interacting multiverse in the low multipoles interval.

The multiverse is a *place* where time and space are not well defined. We can imagine a particle and antiparticle distancing one from the other, since our laboratory is a *place* where they can be moving away. A pair of universes could have not that place to move freely. Assuming the existence of such place could be imprudent.

When an electron and a positron are created from a virtual photon, their spins are entangled. At that moment, the entanglement[28] is maximum, and it can only be reduced or removed naturally by decoherence. The entanglement cannot be increased naturally in any case after they start to travel freely. For the case in which there is not environment with which the particles can interact, the entanglement must be constant.

For most of the models where there are critical points along the classical evolution, the entanglement entropy of a pair of universes has been found[25,26] to be high or divergent at the initial singularity (when they are created), and divergent at those critical points of the expansion of each universe. It implies that the quantum entanglement of the pair is high at their origin, and at the moment when the expansion of the universes is maximum or minimum.

If such notion of usual space is present in this kind of multiverse, where the universes move and interact, the rules of entanglement should be different from the ones we know from our experiments. It must allow the possibility for the entanglement to increase naturally, such that at the maximum of the expansion of the universes, the entanglement is high again. Alternatively, the multiverse might not be a *place* like the ordinary one we are used to.

Somehow, if we consider the entanglement to work like in our laboratories, the structure of this multiverse cannot be seen as a laboratory where entangled particles

run away from each other, otherwise the entanglement would be constant or vanishing while time goes by. And perhaps, the problem is not only about the space but also the time. If a time-like variable exists for the multiverse, we cannot tell what it is, or how it is intermingled with the spatial variable.

3. Conclusions

The results of the entanglement entropy of the multiverse, assuming that third quantization is correct, indicate that the multiverse is a different *space* from the one we know from the experience in our universe, or maybe it is the *time* the one which is different, or both at the same time. Here, when we talk about time we mean any notion of time, whether it is the usual time, thermodynamic entropy, or any other physical variable which controls the regular arrow of time.

There could be another possibility where the quantum entanglement works differently in the multiverse than into the universes. In such case, the relationship between both behaviors is unknown. Furthermore, taking for granted that the laws of Physics are satisfied anywhere else could be mistaken. If they hold, we know that the sweet spot where our universe is to let life occurs[29] is highly unlikely. If the rest of the universes exist, and the fundamental constants are different from the ones in our universe, it is expected not to harbor life or, more importantly, avoid immediate recollapse after creation.[30] Indeed, for inflation to occur and let the universe expands, the initial conditions of a universe must also satisfy some requirements, otherwise that universe would recollapse and disappear. There is not much doubt then, that whenever we are able to check the presence of other universes, we would better believe in the possibility that there are different laws of Physics out there.

The manner in which the actual theory of quantum gravity describes this kind of multiverse is in terms of the *internal* degrees of freedom of the universes, like the scale factor or the fields. It could not be the appropriate way to do it, since they have not a good definition out of each universe, and so either space and time.

The emergence of a universe which is exact to ours is probabilistic as long as we are assuming a full quantum behavior of the multiverse. The curvature, the cosmological constant, and the rest of the characteristics of a universe are subject to the probabilistic rules of the multiverse. Therefore, the chance for the existence of a universe like ours cannot be far from null.

There is, of course, the possibility that third quantization picture is not correct at all and the multiverse, as it is described by it, is not realistic. Nevertheless, our results could also be showing that third quantization is unavoidably wrong. If our results are correct, then canonical quantum gravity is correct and the third quantization picture is wrong, or both are wrong. The interpretation of third quantization scenario is also a bit odd, and many improvements can be done behind a more strict field theory of universes.

At present, there seems not to be a way out of our ignorance when multiverse comes in context. We are making so many assumptions when we talk about the

multiverse that the most likely outcome seems to be reasonably incorrect. The remaining work for us is to check whether our theories are, by coincidence, in agreement with the observations or not. The cosmic microwave background is still effectively blind to the effects of the entanglement and other interactions in the multiverse, whatever the kind of multiverse is analyzed.[31–35] So, we may wait for a future discovery of the cosmic neutrino background, or the spectrum of the primordial gravitational waves to confirm that some of the theories fit to the observations. Up to that day and even later, the multiverse will remain an unsolvable mystery.

Acknowledgments

This work was supported by the Polish National Research and Development Center (NCBR) project "UNIWERSYTET 2.0. – STREFA KARIERY", POWR.03.05.00-00-Z064/17-00 (2018-2022).

References

1. B. S. DeWitt, Quantum theory of gravity. I. the canonical theory, *Phys. Rev.* **160**, 1113 (1967).
2. C. Kiefer, *Quantum gravity* (2004).
3. C. J. Isham, Canonical quantum gravity and the problem of time, *NATO Sci. Ser. C* **409**, 157 (1993).
4. C. Kiefer, Conceptual Problems in Quantum Gravity and Quantum Cosmology, *ISRN Math. Phys.* **2013**, p. 509316 (2013).
5. C. Rovelli, Notes for a brief history of quantum gravity, in *9th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories (MG 9)*, 6 2000.
6. J. B. Hartle and S. W. Hawking, Wave function of the universe, *Phys. Rev. D* **28**, 2960 (Dec 1983).
7. A. Vilenkin, Quantum creation of universes, *Phys. Rev. D* **30**, 509 (Jul 1984).
8. L. Mersini-Houghton and R. Vaas (eds.), *The Arrows of Time: A Debate in Cosmology* (Springer, 2012).
9. S. Weinberg (ed.), *Cosmology* (Oxford University Press, 2008).
10. D. Rubin and B. Hayden, Is the expansion of the universe accelerating? All signs point to yes, *The Astrophysical Journal* **833** (2016).
11. M. McGuigan, Third quantization and the wheeler-dewitt equation, *Phys. Rev. D* **38**, 3031 (Nov 1988).
12. A. Hosoya and M. Morikawa, Quantum Field Theory of Universe, *Phys. Rev. D* **39** (1989).
13. S. B. Giddings and A. Strominger, Baby Universes, Third Quantization and the Cosmological Constant, *Nucl. Phys. B* **321**, 481 (1989).
14. A. Strominger and J. R. Ellis, Third quantization [and discussion], *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **329**, 395 (1989).
15. M. McGuigan, Universe creation from the third-quantized vacuum, *Phys. Rev. D* **39**, 2229 (Apr 1989).
16. S. J. R. Pérez, Inter-universal entanglement, in *Open Questions in Cosmology*, ed. G. J. Olmo (IntechOpen, Rijeka, 2012)

17. L. Perivolaropoulos and F. Skara, Challenges for ΛCDM: An update (2021).
18. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics (Cambridge Univ. Press, Cambridge, UK, 2 1984).
19. X. Yingming and L. Liao, Third quantization of a solvable model in quantum cosmology in brans-dicke theory, *Chinese Physics Letters* **8**, 52 (Jan 1991).
20. L. O. Pimentel and C. Mora, Third quantization of brans–dicke cosmology, *Physics Letters A* **280**, 191 (2000).
21. A. Balcerzak and K. Marosek, Emergence of multiverse in third quantized varying constants cosmologies, *The European Physical Journal C* **79** (07 2019).
22. H. R. Lewis and W. B. Riesenfeld, An exact quantum theory of the time-dependent harmonic oscillator and of a charged particle in a time-dependent electromagnetic field, *Journal of Mathematical Physics* **10**, 1458 (1969).
23. S. Robles-Pérez, Invariant vacuum, *Physics Letters B* **774**, 608 (2017).
24. S. Robles-Pérez and P. F. González-Díaz, Quantum state of the multiverse, *Phys. Rev. D* **81**, p. 083529 (Apr 2010).
25. S. Robles-Pérez, A. Balcerzak, M. P. Dąbrowski and M. Krämer, Interuniversal entanglement in a cyclic multiverse, *Phys. Rev. D* **95**, p. 083505 (Apr 2017).
26. A. Balcerzak, S. Barroso-Bellido, M. P. Dąbrowski and S. Robles-Pérez, Entanglement entropy at critical points of classical evolution in oscillatory and exotic singularity multiverse models, *Phys. Rev. D* **103**, p. 043507 (Feb 2021).
27. M. Bouhmadi-López, M. Krämer, J. Morais and S. Robles-Pérez, The interacting multiverse and its effect on the cosmic microwave background, *Journal of Cosmology and Astroparticle Physics* **2019**, 057 (Feb 2019).
28. R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (Jun 2009).
29. *Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity* (Cambridge University Press, 2004).
30. A. Vilenkin, Boundary conditions in quantum cosmology, *Phys. Rev. D* **33**, 3560 (Jun 1986).
31. S. Kanno, Cosmological implications of quantum entanglement in the multiverse, *Physics Letters B* **751**, 316 (2015).
32. W. H. Kinney, Limits on entanglement effects in the string landscape from planck and BICEP/keck data, *Journal of Cosmology and Astroparticle Physics* **2016**, 013 (Nov 2016).
33. E. D. Valentino and L. Mersini-Houghton, Testing predictions of the quantum landscape multiverse 1: The starobinsky inflationary potential, *Journal of Cosmology and Astroparticle Physics* **2017**, 002 (Mar 2017).
34. E. D. Valentino and L. Mersini-Houghton, Testing predictions of the quantum landscape multiverse 2: The exponential inflationary potential, *Journal of Cosmology and Astroparticle Physics* **2017**, 020 (Mar 2017).
35. E. Di Valentino and L. Mersini-Houghton, Testing predictions of the quantum landscape multiverse 3: The hilltop inflationary potential, *Symmetry* **11** (2019).

Effective field theory from relativistic Generalized Uncertainty Principle

Vasil N. Todorinov*, Saurya Das, and Pasquale Bosso
Department of Physics & astronomy, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada
**E-mail: v.todorinov@uleth.ca*
https://www.ulethbridge.ca

Theories of Quantum Gravity predict a minimum measurable length and a corresponding modification of the Heisenberg Uncertainty Principle to the so-called Generalized Uncertainty Principle (GUP). However, this modification is usually formulated in non-relativistic language, making it unclear whether the minimum length is Lorentz invariant. We have formulated a Relativistic Generalized Uncertainty Principle, resulting in a Lorentz invariant minimum measurable length and the resolution of the composition law problem. This proved to be an important step in the formulation of Quantum Field Theory with minimum length. We derived the Lagrangians consistent with the existence of minimal length and describing the behaviour of scalar, spinor, and U(1) gauge fields. We calculated the Feynman rules (propagators and vertices) associated with these Lagrangians. Furthermore, we calculated the Quantum Gravity corrected scattering cross-sections for a lepton-lepton scattering. Finally, we compared our results with current experiments, which allowed us to improve the bounds on the scale at which quantum gravity phenomena will become relevant.

Keywords: Quantum Gravity: Quantum Gravity Phenomenology, Effective Field theory, Generalized Uncertainty Principle

1. Introduction

The quantization of gravity is one of the major outstanding problems of fundamental physics. The family of theories tackling this problem is known as Quantum Gravity (QG). This family is home to a large diversity of ideas and approaches to the problem. The most prevalent members of the family, namely String Theory, Loop Quantum Gravity in addition to models such as Doubly-Special Relativity theories as well as many others, agree on the existence of a minimum measurable length or a similar scale in spacetime. The minimum length is usually considered to be the proportional to the Planck length, $\ell_{Pl} = 10^{-35}m$, and it signifies the scale at which quantum gravity effects would manifest. Currently, we have no access to energy regimes anywhere close to the Planck scale, therefore all theories of QG lack a direct experimental confirmation. The field of Quantum Gravity Phenomenology (QGP) tackles this problem by using the existing theories to make models in order of finding low energy remnants of QG phenomena in already existing experiments.

As we already mentioned, one phenomenon that most theories of QG agree on is the existence of a minimum measurable length, which is in direct contradiction with

the Heisenberg uncertainty principle (HUP), since the latter allows for infinitely small uncertainties in position. The solution is the introduction of a new model which deforms the HUP in order to accommodate for a minimal uncertainty in position. The deformed HUP is known as Generalized Uncertainty Principle (GUP), and was first introduced by Kempf, Mangano, and Mann in 1995.[1] One of the possible position-momentum commutators corresponding to a GUP model has the following form

$$[x_i, p_j] = i\hbar\delta_{ij}(1 + \beta_1 p^2) + i\hbar\beta_2 p_i p_j \,, \tag{1}$$

where $\beta_1 = \beta_1^*(\ell_{\rm Pl}/\hbar)^2$, $\beta_2 = \beta_2^*(\ell_{\rm Pl}/\hbar)^2$, and β_1^*, β_2^* are numerical coefficients. The GUP model has been successful in modeling a wide array of corrections to quantum mechanical phenomena,[1–58] and has provided compelling bound on the minimum length scale, by comparing the result to existing experimental data.

However, the model introduced by Eq. (1) has two shortcomings:

- The position-momentum commutator and the corresponding uncertainty relation are not Lorentz covariant, and therefore the resulting minimum length is frame dependent,
- *Composition Law Problem*: Energies and momenta do not sum up linearly.[59]

The composition law problem arises due to the fact that deformation of the HUP is done through a generalization of the momentum operator. Namely terms of higher order in momentum are added in order to deform the commutator. This leads to a deformation of the Einstein dispersion relation of the form

$$E^2 = p^2c^2 + m^2c^4 \;\rightarrow\; E^2 = p^2c^2 + m^2c^4 + \mathcal{O}(p^4) \,. \tag{2}$$

Consequently one can show that if there are terms of higher than second order in momentum in the dispersion relation, for a composite system moving in the same speed the energies and momenta do not sum up linearly,

$$\vec{p}_3 \neq \vec{p}_2 + \vec{p}_1 \qquad E_3 \neq E_1 + E_2 \,. \tag{3}$$

It is important to note that solutions of this problem have been developed, some examples are.[60,61]

Such problems prevents us to apply the GUP models to Relativistic Quantum Mechanics and Quantum Field Theory, which essentially restricts the energy regimes in which we can search for the remnants of QG effects, by preventing us from using the highest energy experiments currently available, *i.e.* LHC.

In the following sections, we will address these problems by extending the GUP up to second order in Planck length, to relativistic regimes (RGUP), we will explore the deformations of the Lorenz group and symmetries associated with it, and we will obtain Lagrangians allowing for the existence of minimum length for fields of spin-$0, 1/2, 1$.

2. Relativistic Generalized Uncertainty Principle[62]

Inspired by,[63] we will use the most general form of quadratic GUP algebra and extend it to a covariant expression in Minkowski spacetime with the following signature $\{-,+,+,+\}$. The particular form obtained for the commutator is as follows

$$[x^\mu, p^\nu] = i\hbar\,(1 + (\varepsilon - \kappa)\gamma p^\rho p_\rho)\,\eta^{\mu\nu} + i\,\hbar(\beta + 2\varepsilon)\gamma p^\mu p^\nu\,, \tag{4}$$

where $\mu\,, \nu \in \{0, 1, 2, 3\}$. For convenience it is useful to separate the dimensionality of the corrections into a dimensionful parameter γ, with the dimension of inverse momentum, and dimensionless parameters κ, β and ϵ (and also ξ in Eq. (7) below). The scale of the QG effects γ is defined as $\gamma = \frac{1}{(c\,M_{Pl})^2} = \frac{\ell_{\rm Pl}^2}{\hbar^2}$, where M_{Pl} is the Planck mass, and $\ell_{\rm Pl}$ is the Planck length. The minimum length we can obtain from the algebra Eq. (4) is of the order of the Planck length. Important to note is that Eq. (4) reduces in the non-relativistic ($c \to \infty$) limit to the one proposed in.[1] Furthermore, in the non-relativistic ($c \to \infty$), and non-GUP limit ($\gamma \to 0$) limit, Eq. (4) reduces to the standard Heisenberg algebra. We can clearly see from the commutator Eq. (4) that while x^μ and p^ν are the physical position and momentum, they are no longer canonically conjugate variables. Therefore, we introduce two auxiliary 4-vectors, x_0^μ and p_0^ν which are canonically conjugate, such that

$$p_0^\mu = -\,i\hbar\frac{\partial}{\partial x_{0\,\mu}}, \qquad\qquad [x_0^\mu, p_0^\nu] = i\hbar\eta^{\mu\nu}. \tag{5}$$

Then we assume that the physical position and momentum are both functions of the auxiliary ones,

$$x^\mu = x^\mu(x_0, p_0), \quad p^\mu = p^\mu(x_0, p_0)\,. \tag{6}$$

By expanding Eq. (6) in Taylor series, which we truncate up to first order in the QG scale γ, and imposition of commutativity in momentum space $[p_\mu, p_\nu] = 0$, we can write the general expressions of the physical position and momentum are as follows

$$x^\mu = x_0^\mu - \kappa\gamma p_0^\rho p_{0\rho} x_0^\mu + \beta\gamma p_0^\mu p_0^\rho x_{0\rho} + \xi\hbar\gamma p_0^\mu, \tag{7}$$

$$p^\mu = p_0^\mu\,(1 + \varepsilon\gamma p_0^\rho p_{0\rho})\,. \tag{8}$$

Using Eqs. (7) and (8), we can calculate the following expression for the position-position commutator

$$[x^\mu, x^\nu] = i\hbar\gamma\frac{-2\kappa + \beta}{1 + (\varepsilon - \kappa)\gamma p^\rho p_\rho}\,(x^\mu p^\nu - x^\nu p^\mu)\,. \tag{9}$$

Therefore as in,[1] in this case as well we arrive at a non-commutative spacetime.

We also note that the last two terms in Eq. (7) introduce a preferred direction of p_0^μ and therefore break isotropy of spacetime. Since this violates the principles of relativity, we will assume that $\beta = \xi = 0$ from now on.

2.1. *Lorentz and Poincaré algebra*

Using Eqs. (7) and (8), we now construct the generators of the Lorentz group

$$M^{\mu\nu} = p^\mu x^\nu - p^\nu x^\mu = [1 + (\varepsilon - \kappa)\gamma p_0^\rho p_{0\,\rho}]\, \tilde{M}^{\mu\nu}\,, \tag{10}$$

where $\tilde{M}^{\mu\nu} = p_0^\mu x_0^\nu - p_0^\nu x_0^\mu$ is the Lorentz generators constructed from the canonical variables x_0 and p_0. The commutator representative of the Poincaré algebra for the physical operators can then be calculated

$$[x^\mu, M^{\nu\rho}] = i\hbar[1 + (\varepsilon - \kappa)\gamma p^\rho p_\rho]\,(x^\nu \delta^{\mu\rho} - x^\rho \delta^{\mu\nu}) + i\hbar 2(\varepsilon - \kappa)\gamma p^\mu M^{\nu\rho} \tag{11a}$$

$$[p^\mu, M^{\nu\rho}] = i\hbar[1 + (\varepsilon - \kappa)\gamma p^\rho p_\rho]\,(p^\nu \delta^{\mu\rho} - p^\rho \delta^{\mu\nu})\,, \tag{11b}$$

$$[M^{\mu\nu}, M^{\rho\sigma}] = i\hbar\,(1 + (\varepsilon - \kappa)\gamma p^\rho p_\rho)\,(\eta^{\mu\rho} M^{\nu\sigma} - \eta^{\mu\sigma} M^{\nu\rho} - \eta^{\nu\rho} M^{\mu\sigma} + \eta^{\nu\sigma} M^{\mu\rho})\,. \tag{11c}$$

From Eqs. (11a,11b,11c), we can see that on the line $\varepsilon = \kappa$ in the parameter space, one has a non-trivial RGUP with an *unmodified* Poincaré algebra. On this line in parameter space, the RGUP algebra and the non-commutativity of spacetime coming from Eqs. (4) and (9) takes this form:

$$[x^\mu, p^\nu] = i\hbar\,(\eta^{\mu\nu} + 2\kappa\gamma p^\mu p^\nu)\,, \tag{12}$$

$$[x^\mu, x^\nu] = -2i\hbar\kappa\gamma\,(x^\mu p^\nu - x^\nu p^\mu)\,, \tag{13}$$

where $\kappa > 0$. We notice that for any one-dimensional spatial component, Eq. (12) mirrors the one obtained in.[1] Eqs. (12) and (13) have similarities to the ones proposed in.[53]

2.2. *Lorentz invariant minimum length and the Composition Law Problem*

The main reason for extending GUP to relativistic frameworks is to make its minimum uncertainty in position Lorentz invariant, or in other words to make a minimum length frame independent. In relativistic QM, position and momentum are promoted to operators and the Lorentz transformations Λ are represented by a unitary operator $U(p^\nu, M^{\rho\sigma})$, *i.e.* $U^* = U^{-1}$, where p^ν is the generator of translation and $M^{\rho\sigma}$ is the generator of rotations of the Poincaré group. The position and momentum in this case transform as follows

$$x'^\mu = U(p^\nu, M^{\rho\sigma}) x^\mu U^{-1}(p^\nu, M^{\rho\sigma})\,, \tag{14}$$

$$p'^\mu = U(p^\nu, M^{\rho\sigma}) p^\mu U^{-1}(p^\nu, M^{\rho\sigma})\,. \tag{15}$$

The operator $U(p^\nu, M^{\rho\sigma})$ can be expressed from the generators of the Poincaré algebra as

$$U(p^\nu, M^{\rho\sigma}) = \exp\left[i a_\nu p^\nu\right] \exp\left[i \frac{\omega_{\rho\sigma} M^{\rho\sigma}}{2}\right]\,, \tag{16}$$

where the coordinate system is translated by a vector a_μ and is rotated by $\omega_{\rho\sigma}$.

The position-momentum commutator transforms as follows

$$[x'^{\mu}, p'^{\nu}] = U[x^{\mu}, p^{\nu}]U^{-1}\,. \tag{17}$$

Substituting $[x^{\mu}, p^{\nu}]$ for its expression as given in Eq. (4), we get

$$[x'^{\mu}, p'^{\nu}] = U\left\{i\hbar\,(1+(\varepsilon-\kappa)\gamma p^{\rho}p_{\rho})\,\eta^{\mu\nu} + i\,\hbar(\beta+2\varepsilon)\gamma p^{\mu}p^{\nu}\right\}U^{-1} \tag{18a}$$

$$= i\hbar\,\left(1+(\varepsilon-\kappa)\gamma p'^{\rho}p'_{\rho}\right)\eta^{\mu\nu} + i\,\hbar(\beta+2\varepsilon)\gamma U p^{\mu}U^{-1}Up^{\nu}U^{-1} \tag{18b}$$

$$= i\hbar\,\left(1+(\varepsilon-\kappa)\gamma p'^{\rho}p'_{\rho}\right)\eta^{\mu\nu} + i\,\hbar(\beta+2\varepsilon)\gamma p'^{\mu}p'^{\nu}\,. \tag{18c}$$

Therefore, the commutator between position and momentum has the same form in every frame. Then, we can safely conclude that every frame will observe the same minimum measurable length.

For the composition law problem, we need to find the dispersion relation and see if it is quadratic or it has terms of higher order in momentum. As we know, the dispersion relation is connected to the Casimir invariants of the Poincaré algebra. In the unmodified case, there are two Casimir invariants: the momentum squared $p_0^{\mu}p_{0\mu}$; and the physical Pauli-Lubanski vector squared $\tilde{W}^{\mu}\tilde{W}_{\mu}$, where

$$\tilde{W}_{\mu} \equiv \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}\tilde{M}^{\nu\rho}p_0^{\sigma}\,. \tag{19}$$

In our case, both the Poincare algebra and the momentum and Pauli-Lubanski vectors are deformed. However, by taking the commutator between the physical momentum squared $p^{\mu}p_{\mu}$ and all the other generators of the algebra, it can easily be shown that it is a Casimir invariant of the modified Poincaré algebra presented in Eq. (11). This can be repeated for the Pauli-Lubanski vector squared as well.

This shows that our dispersion relation can be written as

$$p_{\mu}p^{\mu} = -m^2c^4\,. \tag{20}$$

As we can see, the dispersion relation is quadratic in the physical momentum which means that energies and momenta sum up linearly as they should. It is important to note that due to the fact that the physical momentum is a polynomial of the auxiliary one, the resulting equations of motion will be of higher than second order.

3. Effective Quantum Field Theory with Minimum Length[64,65]

Having cleared up the frame dependence and the composition law problem for our model, we can proceed to apply it to quantum field theory.

In the general case, the commutation relation we get for the position-position operator depends on the momentum. Therefore, it does not form an algebra and it cannot be represented by a smooth Lie manifold. For that reason we have chosen for further calculations to fix the model by making $\kappa = 0$ and $\epsilon = \gamma_0$. We can redefine the RGUP parameter as follows

$$\gamma = \frac{\gamma_0}{(M_{\mathrm{Pl}}c)^2}\,. \tag{21}$$

In this particular model, the physical position and momentum take the form

$$x^\mu = x_0^\mu \,, \tag{22a}$$
$$p^\mu = p_0^\mu \left(1 + \gamma p_0^\rho p_{0\rho}\right). \tag{22b}$$

From now on, the results will be presented in natural units, *i.e.* $c = \hbar = 1$. With the parameters redefined as such, the position-momentum commutator is

$$[x^\mu, p^\nu] = i\,\left(1 + \gamma p^\rho p_\rho\right)\eta^{\mu\nu} + 2i\,\gamma p^\mu p^\nu \,. \tag{23}$$

The Poincaré group corresponding to the RGUP above is represented by the following algebra

$$[x^\mu, M^{\nu\rho}] = i[1 + \gamma p^\rho p_{\,\rho}]\left(x^\nu \delta^{\mu\rho} - x^\rho \delta^{\mu\nu}\right) + i2\gamma p^\mu M^{\nu\rho} \,, \tag{24a}$$
$$[p^\mu, M^{\nu\rho}] = i[1 + \gamma p^\rho p_{\,\rho}]\left(p^\nu \delta^{\mu\rho} - p^\rho \delta^{\mu\nu}\right) \,, \tag{24b}$$
$$[M^{\mu\nu}, M^{\rho\sigma}] = i\left(1 + \gamma p^\rho p_{\,\rho}\right)\left(\eta^{\mu\rho} M^{\nu\sigma} - \eta^{\mu\sigma} M^{\nu\rho} - \eta^{\nu\rho} M^{\mu\sigma} + \eta^{\nu\sigma} M^{\mu\rho}\right) . \tag{24c}$$

It is important to note that the deformed Poincaré algebra presented above has the same irreducible representations, and therefore describes the same particles as the standard Poincaré algebra. This feature can be seen derived in Appendix B. As we already showed, the physical momentum squared is a Casimir invariant of the modified Poincaré algebra presented above in Eq. (24). Therefore, the Einstein dispersion relation for our case is

$$p^\mu p_\mu = -m^2 \,, \tag{25}$$

where p_μ is the physical momentum, the differential form of which is an equation of motion for any Lagrangian describing a boson field. The Dirac equations *i.e.*, the equations of motion describing the dynamics of fermion fields, can also be easily derived. Starting from the dispersion relation Eq. (25) and expressing it in terms of the auxiliary variables defined in Eq. (22a), we get the modified Klein-Gordon equation in its differential form

$$\left[\partial_\mu \partial^\mu \left(1 + \gamma \partial_\nu \partial^\nu\right)^2 + m\right]\phi = 0 \,. \tag{26}$$

Similarly, the differential form of the Dirac equation is shown to be

$$\left(\tau^\mu p_\mu - m\right)\psi = \left[\tau^\mu p_{0\,\mu}(1 + \gamma p_{0\rho} p_0^\rho) - m\right]\psi = 0 \,, \tag{27}$$

where τ^μ are the Dirac matrices and ψ is a Dirac spinor. We can show that as part of the Lorentz group the generators of the Clifford algebra $Cl_{(3,1)}(\mathbb{R})$, if the Casimir is preserved then the properties of the Clifford algebra will remain unchanged, as well. This is true in our case, thus,

$$\left[i\tau^\mu \partial_\mu (1 + \gamma \partial_\rho \partial^\rho)^2 - m\right]\psi = 0 \,, \tag{28}$$

and its Dirac conjugate are the equations of motion for the RGUP-QED Lagrangian.

4. RGUP Modified Lagrangians

In the previous section, we established the modified equations of motion for the scalar, vector, and spinor fields. Note that the equations of motion are differential equations of higher than second order. Therefore, the Lagrangians corresponding to these equations of motion need to have higher than second derivative. The methodology of working with higher derivative Lagrangians

$$L = L(\phi, \dot{\phi}, \ddot{\phi}, \ldots, \overset{(n)}{\phi}), \tag{29a}$$

$$\mathcal{L} = \mathcal{L}(\phi, \partial_{\mu_1}\phi, \partial_{\mu_1}\partial_{\mu_2}\phi, \ldots, \partial_{\mu_1}\ldots\partial_{\mu_n}\phi), \tag{29b}$$

is given by the Ostrogradsky method.[66–68] The Euler-Lagrange equations for theories with higher derivatives will have the form:

$$\frac{dL}{dq} - \frac{d}{dt}\frac{dL}{d\dot{q}} + \frac{d^2}{dt^2}\frac{dL}{d\ddot{q}} + \ldots + (-1)^n \frac{d^n}{dt^n}\frac{dL}{d(d^n q/dt^n)} = 0, \tag{30}$$

which in the case of fields is

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} + \partial_{\mu_1}\partial_{\mu_2}\frac{\partial\mathcal{L}}{\partial(\partial_{\mu_1}\partial_{\mu_2}\phi)} + \ldots + (-1)^m \partial_{\mu_1}\ldots\partial_{\mu_m}\frac{\partial\mathcal{L}}{\partial(\partial_{\mu_1}\ldots\partial_{\mu_m}\phi)} = 0. \tag{31}$$

The Ostrogradsky method allows us to reconstruct the Lagrangian describing the dynamics of scalar and spinor fields from their equations of motion.

4.1. *Scalar field Lagrangian*

We begin by assuming the most general form of higher derivative Lagrangian, which is one order higher than the equations of motion

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \gamma\left(C_1\,\partial_\mu\partial^\mu\phi\,\partial_\nu\partial^\nu\phi + C_2\,\partial_\mu\phi\,\partial^\mu\partial_\nu\partial^\nu\phi + C_3\,\partial_\nu\partial^\nu\partial^\mu\phi\,\partial_\mu\phi\right) \\ & + \gamma^2\left(C_4\,\partial_\mu\partial^\mu\partial_\nu\phi\,\partial^\nu\partial_\rho\partial^\rho\phi + C_5\,\partial_\mu\partial^\mu\partial_\nu\partial^\nu\phi\,\partial_\rho\partial^\rho\phi + C_6\,\partial_\mu\partial^\mu\partial_\nu\partial^\nu\partial_\rho\phi\,\partial^\rho\phi\right. \\ & \left. + C_7\,\partial_\mu\partial^\mu\phi\,\partial_\nu\partial^\nu\partial_\rho\partial^\rho\phi + C_8\,\partial_\mu\phi\,\partial^\mu\partial_\nu\partial^\nu\partial_\rho\partial^\rho\phi\right) + C_9 m^2\phi^2,\end{aligned} \tag{32}$$

where $\partial_\mu = \partial/\partial x^\mu$. Using the Euler-Lagrange equation prescribed by the Ostrogradsky method Eq. (30), we obtain the following Lagrangian

$$\mathcal{L}_{\phi,\mathbb{R}} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \gamma\,\partial_\nu\partial^\nu\partial^\mu\phi\,\partial_\mu\phi + \frac{\gamma^2}{2}\partial_\mu\phi\,\partial^\mu\partial_\nu\partial^\nu\partial_\rho\partial^\rho\phi, \tag{33}$$

As for the Lagrangian for a complex scalar field ϕ, we generalize Eq. (33) by including additional terms obtaining

$$\begin{aligned}\mathcal{L}_{\phi,\mathbb{C}} = {} & \frac{1}{2}\left(\partial_\mu\phi\right)^\dagger\partial^\mu\phi - \frac{1}{2}m^2\phi^\dagger\phi + \gamma\left[\left(\partial_\nu\partial^\nu\partial^\mu\phi\right)^\dagger\partial_\mu\phi + \partial_\nu\partial^\nu\partial^\mu\phi\left(\partial_\mu\phi\right)^\dagger\right] \\ & + \frac{\gamma^2}{2}\left[\left(\partial_\mu\phi\right)^\dagger\partial^\mu\partial_\nu\partial^\nu\partial_\rho\partial^\rho\phi + \partial_\mu\phi\left(\partial^\mu\partial_\nu\partial^\nu\partial_\rho\partial^\rho\phi\right)^\dagger\right],\end{aligned} \tag{34}$$

such that hermiticity is restored, *i.e.* $\mathcal{L}^\dagger_{\phi,\mathbb{C}} = \mathcal{L}_{\phi,\mathbb{C}}$. Furthermore, it is worth noticing that Eq. (34) is consistent with Eq. (55) in[42] up to a numerical factor.

4.2. *Spinor field Lagrangian*

The first step in the derivation of the spinor field Lagrangian is to assume a general form of the Lagrangian, where the order of derivatives is determined by the order of the equation of motion Eq. (28). Moreover, an assumption is made that different terms will be multiplied by an arbitrary numerical coefficient

$$\mathcal{L}_\psi = \bar{\psi}\,[iC_1\tau^\mu\partial_\mu(1 + C_2\gamma\partial_\rho\partial^\rho) - C_3 m]\,\psi\,. \tag{35}$$

Next step is to prove that it has Eq. (28) as an equation of motion. Applying the Ostrogradsky method, one gets the following equations of motion for the field and its complex conjugate

$$C_1 i\tau^\mu\partial_\mu\psi + C_1C_2\gamma\partial_\rho\partial^\rho\psi - C_3 m\psi = 0, \tag{36}$$

$$C_1 i\tau^\mu\partial_\mu\bar{\psi} + C_1C_2\gamma\partial_\rho\partial^\rho\bar{\psi} - C_3 m\bar{\psi} = 0\,. \tag{37}$$

The equations of motion obtained through the Ostrogradsky method from Eq. (35) need to be identical to Eq. (28), which is obtained from the dispersion relation. Therefore, the Lagrangian corresponding to the QFT spinor with minimum length will be of the form

$$\mathcal{L}_\psi = \bar{\psi}\,[i\tau^\mu\partial_\mu(1 - \gamma\,\partial_\rho\partial^\rho) - m]\,\psi\,. \tag{38}$$

4.3. *$U(1)$ gauge field theory*

The gauge field Lagrangian is obtained in a different way. The Ostrogradsky method is not utilized in this case. The derivation is done by firstly assuming that the vector boson A^μ obeys the ususal $U(1)$ gauge symmetries. For the equations of motion, we assume that they have the same RGUP corrections as the KG equation of the form Eq. (26)

$$\partial_\mu F^{\mu\nu} = \partial_\mu\partial^\mu A^\nu + 2\gamma\partial_\mu\partial^\mu\partial_\rho\partial^\rho A^\nu = 0\,. \tag{39}$$

Defining the standard gauge invariant field strength tensor as

$$F_0^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu, \tag{40}$$

one can express the RGUP modified field strength tensor in terms of the standard one up to first order in γ, as follows:

$$F^{\mu\nu} = F_0^{\mu\nu} + 2\gamma\,\partial_\rho\partial^\rho F_0^{\mu\nu}\,. \tag{41}$$

Then Eq. (39) can be rewritten as

$$\partial_\mu F^{\mu\nu} = \partial_\mu F_0^{\mu\nu} + 2\gamma\partial_\rho\partial^\rho\partial_\mu F_0^{\mu\nu} + \gamma^2\partial_\sigma\partial^\sigma\partial_\rho\partial^\rho\partial_\mu F_0^{\mu\nu}\,. \tag{42}$$

A test of gauge invariance of $F^{\mu\nu}$ is carried through by considering the following gauge transformation of the four-potential

$$A^\mu \to A'^\mu = A^\mu + \partial^\mu\Lambda\,, \tag{43}$$

and the field strength tensor is shown to be gauge invariant. Thus, up to first order in the RGUP parameter γ, the gauge field Lagrangian reads

$$\mathcal{L}_A = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} = -\frac{1}{4}F_0^{\mu\nu}F_{\mu\nu 0} - \frac{\gamma}{2}F_{\mu\nu 0}\partial_\rho\partial^\rho F_0^{\mu\nu}\,. \tag{44}$$

Notice that both the field tensor in Eq. (41) and the gauge field Lagrangian Eq. (44) are invariant under $U(1)$ gauge transformations.

5. Feynman Rules

We use the Lagrangians derived above to formulate a RGUP deformed scalar and spinor electrodynamics. This is achieved by minimally coupling the lepton to the gauge fields. The Feynman rules, consisting of the propagators and vertices for both cases, are calculated using the standard methodology, which can be found in any QFT textbook.

5.1. *Propagators*

The Feynman propagator for the scalar field with a minimum length is the Green's function of the the modified KG equation in Eq. (26)

$$\left[\partial_\mu\partial^\mu\left(1+\gamma\partial_\nu\partial^\nu\right)+(mc)^2\right]G(x-x') = -i\delta(x-x')\,. \tag{45}$$

Expressing the Green's function $G(x-x')$ in terms of its Fourier transform

$$G(x-x') = \int\frac{d^4p_0}{(2\pi)^4}\tilde{G}(p_0)e^{-ip_0\cdot(x-x')}, \tag{46}$$

and substituting it in Eq. (45), we get

$$\int\frac{d^4p_0}{(2\pi)^4}\tilde{G}(p_0)\left[-p_0^2(1-\gamma p_0^2)+(mc)^2\right]e^{-ip_0\cdot(x-x')} = -i\int\frac{d^4p_0}{(2\pi)^4}e^{-ip_0\cdot(x-x')}\,. \tag{47}$$

The Fourier transform of the Feynman propagator is

$$\tilde{G}(p_0) = \frac{-i}{-p_0^2(1+\gamma p_0^2)+(mc)^2}\,, \tag{48}$$

while the propagator itself is

$$G(x-x') = \int\frac{d^4p_0}{(2\pi)^4}\frac{-i}{-p_0^2(1+\gamma p_0^2)+(mc)^2}e^{-ip_0\cdot(x-x')}\,. \tag{49}$$

With the exception of a little trick consisting of multiplying both sides by

$$\left[\tau^\mu p_{0\,\mu}(1-\gamma p_{0\rho}p_0^\rho)+m\right]\,, \tag{50}$$

we find the Dirac field propagator in the same way

$$G(x-x') = \int\frac{d^4p_0}{(2\pi)^4}\frac{-i\left[\tau^\mu p_{0\,\mu}(1+\gamma p_{0\rho}p_0^\rho)+m\right]}{p_0^\mu p_{0\,\mu}(1+\gamma p_{0\rho}p_0^\rho)^2-m^2}e^{-ip_0\cdot(x-x')}\,. \tag{51}$$

Again, the gauge field propagator can be treated in a similar manner. And the Feynman propagator is found to be

$$G(x-x') = \int \frac{d^4q_0}{(2\pi)^4} \frac{-i}{-q_0^2 + 2\gamma q_0^4} e^{-iq_0\cdot(x-x')}\,, \tag{52}$$

where q_0 is the auxiliary four-momentum of the gauge field.

This construction assumes that one can define Dirac delta functions in position space, and the minimal length arises when one tries to localize the fields.

5.2. *Vertices*

The complete set of Feynman rules for the system requires the calculation of the vertices for the charged and gauge fields. Starting from the Lagrangian in Eq. (34), one introduces the minimal coupling as follows

$$\partial_\mu \to D_\mu = \partial_\mu - ieA_\mu\,, \tag{53}$$

where A_μ is the four-potential or the gauge field.

Replacing partials in Eq. (34) by covariant derivatives, the full action of the minimally coupled complex scalar field and the gauge field is obtained. The action reads as follows

$$\begin{aligned}\int \mathcal{L}\, d^4x = \int \left[\mathcal{L}_A + \mathcal{L}_{\phi,\mathbb{C}}\right] d^4x = \int \Bigg\{ & \frac{1}{2}\left(D_\mu\phi\right)^\dagger D^\mu\phi - \frac{1}{2}m^2\phi^\dagger\phi \\ & - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \gamma\left[\left(D_\nu D^\nu D^\mu\phi\right)^\dagger D_\mu\phi + D_\nu D^\nu D^\mu\phi\left(D_\mu\phi\right)^\dagger\right]\Bigg\} \\ & + \frac{\gamma^2}{2}\left[\left(D_\mu\phi\right)^\dagger D^\mu D_\nu D^\nu D_\rho D^\rho\phi + D_\mu\phi\left(D^\mu D_\nu D^\nu D_\rho D^\rho\phi\right)^\dagger\right]\Bigg\}\, d^4x\,.\end{aligned} \tag{54}$$

By expanding the Lagrangian we get a very rich theory

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\left(\partial_\mu\phi\right)^\dagger\partial^\mu\phi - ieA^\mu\left[\phi^\dagger\partial_\mu\phi - \phi\left(\partial_\mu\phi\right)^\dagger\right] - \frac{1}{2}m^2\phi^\dagger\phi + e^2A_\mu A^\mu\phi^\dagger\phi \\ & - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\gamma\Bigg\{\left(\partial_\nu\partial^\nu\partial^\mu\phi\right)^\dagger\partial_\mu\phi + \partial_\nu\partial^\nu\partial^\mu\phi\left(\partial_\mu\phi\right)^\dagger - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\partial_\mu\partial_\nu F^{\mu\nu} \\ & - ie\Bigg\{\left(\partial_\nu\partial^\nu A^\mu\right)\left[\phi^\dagger\partial_\mu\phi - \phi\left(\partial_\mu\phi\right)^\dagger\right] + 2\partial_\nu A^\mu\left[\left(\partial^\nu\phi\right)^\dagger\partial_\mu\phi - \partial^\nu\phi\left(\partial_\mu\phi^\dagger\right)\right] \\ & + A^\mu\left[\left(\partial_\nu\partial^\nu\phi\right)^\dagger\partial_\mu\phi - \partial_\nu\partial^\nu\phi\left(\partial_\mu\phi^\dagger\right)\right]\Bigg\} + e^2\Bigg\{A^\nu(\partial_\nu A_\mu)\left[\left(\partial^\mu\phi\right)^\dagger\phi + \left(\partial^\mu\phi\right)\phi^\dagger\right] \\ & + 2A^\nu A_\mu\left[\left(\partial^\mu\phi\right)^\dagger\partial_\nu\phi + \left(\partial^\mu\phi\right)\left(\partial_\nu\phi\right)^\dagger\right] + A^\nu A_\nu\left[\left(\partial^\mu\phi\right)^\dagger\partial_\mu\phi + \left(\partial^\mu\phi\right)\left(\partial_\mu\phi\right)^\dagger\right] \\ & - 2A^\mu A^\nu\left[\phi^\dagger\partial_\nu\partial_\mu\phi + \phi\left(\partial_\nu\partial_\mu\phi\right)^\dagger\right] + 2A^\mu(\partial_\nu\partial^\nu A_\mu)\phi^\dagger\phi \\ & + A^\mu(\partial^\nu A_\mu)\left[\phi^\dagger\partial_\nu\phi + \phi\left(\partial_\nu\phi\right)^\dagger\right] + A^\mu A_\mu\left[\phi^\dagger\partial_\nu\partial^\nu\phi + \phi\left(\partial_\nu\partial^\nu\phi\right)^\dagger\right]\Bigg\} \\ & + ie^3\Bigg\{A_\mu A_\nu A^\nu\left[\phi\left(\partial^\mu\phi\right)^\dagger - \phi^\dagger\partial^\mu\phi\right] + 2A^\mu A_\mu A^\nu\left[\phi^\dagger\partial_\nu\phi - \phi\left(\partial_\nu\phi\right)^\dagger\right]\Bigg\} \\ & + 2e^4A^\mu A_\mu A^\nu A_\nu\phi^\dagger\phi\Big\} + \mathcal{O}(\gamma^2)\,.\end{aligned} \tag{55}$$

The above expression contains all terms relevant to Feynman diagrams predicted by the usual scalar Quantum Electrodynamics (QED) Lagrangian with the addition of RGUP corrections. By counting the fields in each term, it is easy to see that up to 6-point vertices (2 scalar and 4 gauge) are allowed. This Lagrangian contains 74 Feynman diagrams, with coupling constants classified by powers of the fine structure constant α and the minimum length parameter γ, see Table 1. It is also worth mentioning that in the low energy limit $\gamma \to 0$ we recover the usual Lagrangian for a complex scalar fields.

Table 1. Classification of the Feynman vertices arising from Eq. (55) in terms of the powers of the coupling constants: α the fine structure constant; γ the minimum length coefficient.

	Powers of γ			
Powers of α	$\alpha^{1/2}$	$\alpha^{1/2}\gamma$	$\alpha^{1/2}\gamma^2$	$\alpha^{1/2}\gamma^3$
	α	$\alpha\gamma$	$\alpha\gamma^2$	$\alpha\gamma^3$
	N/A	$\alpha^{3/2}\gamma$	$\alpha^{3/2}\gamma^2$	$\alpha^{3/2}\gamma^3$
	N/A	$\alpha^2\gamma$	$\alpha^2\gamma^2$	$\alpha^2\gamma^3$
	N/A	$\alpha^{5/2}\gamma$	$\alpha^{5/2}\gamma^2$	$\alpha^{5/2}\gamma^3$
	N/A	$\alpha^3\gamma$	$\alpha^3\gamma^2$	$\alpha^3\gamma^3$

Similar procedure is done for the Dirac field Lagrangian the full RGUP-QED action reads

$$\begin{aligned} S = \int \mathcal{L}\, d^4x &= \int \left[\mathcal{L}_A + \mathcal{L}_\psi\right] d^4x \\ &= \int \bar{\psi}\left[i\tau^\mu D_\mu\psi - \gamma\bar{\psi}i\tau^\mu D_\mu D_\nu D^\nu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right] d^4x\,. \end{aligned} \tag{56}$$

We can read off the vertices after we expand the covariant derivatives in the minimally coupled modified Dirac field Lagrangian Eq. (38)

$$\begin{aligned} \mathcal{L}_\psi = {} & i\bar{\psi}\tau^\mu\partial_\mu\psi + i\gamma\bar{\psi}\tau^\mu\partial^\rho\partial_\rho\partial_\mu\psi - m\bar{\psi}\psi \\ & - e\left[\bar{\psi}\tau^\mu A_\mu\psi - 2\gamma\bar{\psi}\tau^\mu\left(\partial_\mu A^\rho\right)\partial_\rho\psi - 2\gamma\bar{\psi}\tau^\mu A^\rho\partial_\mu\partial_\rho\psi - \gamma\bar{\psi}\tau^\mu A_\mu\partial^\rho\partial_\rho\psi\right] \\ & - ie^2\gamma\left[2\bar{\psi}\tau^\mu A_\mu A^\rho\partial_\rho\psi + 2\bar{\psi}\tau^\mu\left(\partial_\mu A^\rho\right)A_\rho\psi - \bar{\psi}\tau^\mu A^\rho A_\rho\partial_\mu\psi\right] \\ & - e^3\gamma\bar{\psi}\tau_\mu A^\mu A^\rho A_\rho\psi\,. \end{aligned} \tag{57}$$

Up to five particle vertices allowed, they include always two fermions and from one to three gauge bosons. Furthermore, the coupling constants for each vertex are a product of powers of the electronic charge e and the RGUP coefficient γ. In fact, the power of the electronic charge determines how many bosons couple to the vertex and the power of γ is 0 for the usual terms and 1 for the RGUP corrections terms. As before we can classify the coupling constants in Table 2.

Table 2. Classification of the coupling constants in Eq. (57) in terms of the powers: α the fine structure constant; γ the minimum length coefficient.

	Powers of γ	
	$\alpha^{1/2}$	$\alpha^{1/2}\gamma$
Powers of α	N/A	$\alpha\gamma$
	N/A	$\alpha^{3/2}\gamma$

6. RGUP Corrections to QED Scattering Amplitudes

The calculations for the scattering amplitudes can be quite cumbersome so we will not present them in full here. If the reader is interested the full calculations are presented in.[64,65,69]

The particular case we will be studying is an electron-muon scattering. We made this choice to avoid mass degeneracies which will complicate our calculations without helping in clarifying the concept. First when calculating the scattering amplitudes the focus will be on the 3-point vertices, containing up to first order the RGUP coefficient γ. This approximation is justified by the fact that the 3-point vertices will have the largest contribution to the scattering amplitudes.

In brief, we begin by calculating the transition amplitudes for a single three legged vertex. For the scalar field we get

$$\begin{aligned} T_{fi} =& -i\int eA^{\mu}\left[\phi_f^{\dagger}\partial_{\mu}\phi_i - \phi_f\partial_{\mu}\phi_i^{\dagger}\right]d^4x \\ & -i\int e\gamma A^{\mu}\left[4\partial_{\nu}\partial^{\nu}\phi^{\dagger}\partial_{\mu}\phi + 4\partial^{\nu}\phi^{\dagger}\partial_{\nu}\partial_{\mu}\phi + \phi^{\dagger}\partial_{\nu}\partial^{\nu}\partial_{\mu}\phi - 4\partial_{\nu}\partial^{\nu}\phi\partial_{\mu}\phi^{\dagger}\right. \\ & \left. -4\partial^{\nu}\phi\partial_{\nu}\partial_{\mu}\phi^{\dagger} - \phi\partial_{\nu}\partial^{\nu}\partial_{\mu}\phi^{\dagger}\right]d^4x\,. \end{aligned} \tag{58a}$$

In the QED case we get for the different components

$$\begin{aligned} T_{fi} =& \, i\int e\bar{\psi}_f\tau^{\mu}A_{\mu}\psi_i\,d^4x + i\int 2e\gamma\partial_{\rho}\bar{\psi}_f\tau^{\rho}A^{\mu}\partial_{\mu}\psi_i\,d^4x\,, \\ & -i\int e\gamma\bar{\psi}\tau^{\mu}A_{\mu}\partial^{\rho}\partial_{\rho}\psi\,d^4x\,. \end{aligned} \tag{59a}$$

The separation in different terms is purely for convenience.

We then couple two Feynman vertices thorough a gauge boson A^{μ}, and isolate the modified invariant amplitude squared

$$\left|\mathfrak{M}_{\text{Scalar}}\right|^2 = 4\pi\alpha\frac{3+\cos\theta}{1-\cos\theta}\left[1+8\gamma E^2(1-\cos\theta)\right]\,, \tag{60a}$$

$$\left|\mathfrak{M}_{\text{spinor}}\right|^2 = 32\pi^2\epsilon_0^2\alpha\left[\frac{t^2+u^2}{s^2} + \frac{1}{2}\gamma(m_e^2+m_{\text{muon}}^2)\frac{tu-u^2}{s^2}\right]\,, \tag{60b}$$

where s, t, u are the Mandelstam variables given by

$$s = (k + k')^2 = (p + p')^2 \approx 2k \cdot k' \approx 2p \cdot p' \,, \tag{61a}$$

$$t = (k - p)^2 = (p' - k')^2 \approx -2k \cdot p \approx -2k' \cdot p' \,, \tag{61b}$$

$$u = (k - p')^2 = (p - k') \approx -2k \cdot p' \approx -2p \cdot k' \,. \tag{61c}$$

From here, it is easy to calculate the RGUP scaled transition rates, from which we can isolate a formula for the differential cross-ectons for both cases. For the scalar fields, the differential cross-section is

$$\left.\frac{d\sigma}{d\Omega}\right|_{CM} = \frac{1}{4\,s}\alpha^2 \left(\frac{3 + \cos\theta}{1 - \cos\theta}\right)^2 \left[1 + 16\gamma E^2(1 - \cos\theta)\right] \,, \tag{62}$$

and for the spinors

$$\left.\frac{d\sigma}{d\Omega}\right|_{CM} = \frac{\alpha^2}{4\,s}\left[\frac{1}{2}\left(1 + \cos^2\theta\right) + \frac{1}{4}\gamma(m_e^2 + m_{\text{muon}}^2)\left(\cos\theta + \cos^2\theta\right)\right] \,. \tag{63}$$

As one can see from the differential cross-section for the scalar fields Eq. (62), the magnitude og QG effects depends on the energy of the scattering squared, and it will be largest for the backscattering. We can see that this is also the case in the spinor case Eq. (63). It is interesting to note that for the spinor case the corrections depend on the mass squared of the particles involved. This suggests that minimum length effects on scattering amplitudes may be measurable for electromagnetic scattering of heavier systems, such as scattering of heavy ions. Applied to existing data from the Xe-Xe scattering observed in the ATLAS experiment, the order of magnitude of the corrections is $\gamma\, m_{\text{Xe}}^2/4 \sim 10^{-34}$.[70] The respective upper bounds on the RGUP parameter are then set to be $\gamma_0 < 10^{34}$.

7. Conclusion

In this work we have extended the Generalized Uncertainty Principle to relativistic framework. This leads to the solution of two major problems: the dependence of the minimum measurable length on the frame of reference; and the composition law problem. This allowed us to extend the reach of GUP from experiments and effects in Quantum Mechanics into Quantum Fired Theory. We achieved that through the formulation of Effective Field Theory with minimum length. Further, we were able to calculate the scattering amplitudes for high-enegy scattering of scalar and spinor particles. Comparing our results to data collecd in the ATLAS experiments, we were able to obtain bounds on the RGUP parameters, *i.e.* $\gamma_0 < 10^{34}$. This bound does not seem impressive on the first glance, however, it is sixteen orders of magnitude improvement compared to previous results from purely quadratic GUP models.

Additionally, we have already been able to obtain RGUP modified action for a spin-2 particle, and through the classical limit connect it to quadratic gravity, allowing us to test our ideas into fields like cosmological inflation.

The results presented here provide a valuable step to extend the reach of Quantum Gravity Phenomenology to the most highly energetic earthbound experiments like LHC. For example, we need to generalize the results here for the gauge group of the Standard Model $U(1)\otimes SU(2)\otimes SU(3)$. Additionally, the gauge field Lagrangian can be used to calculate corrections to phenomena in classical electrodynamics and light propagation, amongst others, like cosmology with weakly self-interacting photons, gravitational waves, and Higgs mechanism and mass.

Appendix A. Ostrogradsky Method and Vacuum Instability

The Ostrogradsky method is a mathematical constructions allowing one to work with higher derivative Lagrangians.[66–68] One of the major issues with it, is the instability of the Hamiltonian, arising when one considers higher than second order derivative. The instability for the particular case presented in here is discussed in the following appendix. We show that the Hamiltonian for a scalar field is positively defined in sub Planckian regimes, thus fixing the applicable domain for our Effective field theory.

We from the equation of motion for the real scalar field, obtained from the modified dispersion relation which corresponds to the Casimir operator of the modified Poincaré group defined in[62]

$$p_0^\rho p_{0\rho}(1+2\alpha\gamma^2 p_0^\sigma p_{0\sigma}) = -(mc)^2\,. \tag{A.1}$$

The Lagrangian for a real scalar field is derived by assuming the most general form of the Lagrangian, and applying the Ostrogradsky method to obtain its equations of motion and comparing to Eq. (A.1). The form of the Lagrangian is

$$\mathcal{L}_{\phi,\mathbb{R}} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \gamma\,\partial_\nu\partial^\nu\partial^\mu\phi\,\partial_\mu\phi\,. \tag{A.2}$$

The generalized coordinates are defined as

$$q_1 \equiv \phi \quad q_2 \equiv \partial_\mu\partial^\mu\phi\,. \tag{A.3}$$

The generalized momenta are then derived as

$$\pi^{\mu\rho}_{\rho} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\partial_\rho\partial^\rho\phi)} = 2\gamma\partial^\mu\phi\,, \tag{A.4a}$$

$$\pi^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} - \partial_\sigma\partial^\sigma\pi^{\mu\rho}_{\rho} = \partial^\mu\phi\,. \tag{A.4b}$$

The Hamiltonian density is then obtained using the Ostrogradsky Legendre transformations

$$\mathcal{H} = \pi^\mu\partial_\mu\phi + \pi^{\mu\rho}_{\rho}\partial_\sigma\partial^\sigma\partial_\mu\phi - \frac{1}{2}\pi^\mu\pi_\mu - \pi^{\mu\rho}_{\rho}\partial^\sigma\partial_\sigma\pi_\mu + \frac{1}{2}m\phi^2 \tag{A.5a}$$

$$= \frac{1}{2}\pi^\mu\pi_\mu + \frac{1}{2}m^2\phi^2\,. \tag{A.5b}$$

We can easily see that the above is positive definite. However, we can also see that the correction terms cancel, the Quantum Gravity corrections are hidden in the dispersion relation of ϕ. The field ϕ is solution of Eq. (A.1) which has four different solutions all containing Quantum Gravity corrections. The calculation above is done in the framework of De Donder Weyl Covariant Hamiltonian Formulation of Field Theory.

Considering the same definition of generalized coordinates presented in Eq. (A.3), one can derive the field momenta outside of the De Donder Weyl formulation as

$$\pi^{\rho}_{\rho} = \frac{\partial \mathcal{L}}{\partial(\partial_{\rho}\partial^{\rho}\dot{\phi})} = 2\gamma\dot{\phi}\,, \tag{A.6a}$$

$$\pi = \frac{\partial \mathcal{L}}{\partial(\dot{\phi})} - \partial_{\sigma}\partial^{\sigma}\pi^{\rho}_{\rho} = \dot{\phi}\,. \tag{A.6b}$$

Performing the Legendre transformation the Hamiltonian density is obtained as

$$\begin{aligned}\mathcal{H} &= \pi\dot{\phi} + \pi^{\rho}_{\rho}\partial_{\sigma}\partial^{\sigma}\dot{\phi} - \frac{1}{2}\pi\pi - \pi^{\rho}_{\rho}\partial^{\sigma}\partial_{\sigma}\pi + \frac{1}{2}m\phi^2 \\ &\quad + \frac{1}{2}\left(\nabla\phi\right)\cdot\left(\nabla\phi\right) + \gamma\left(\nabla\phi\right)\cdot\left(\partial_{\sigma}\partial^{\sigma}\nabla\phi\right) \qquad \text{(A.7)} \\ &= \frac{1}{2}\pi\pi + \frac{1}{2}m\phi^2 + \frac{1}{2}\left(\nabla\phi\right)\cdot\left(\nabla\phi\right) + \gamma\left(\nabla\phi\right)\cdot\left(\partial_{\sigma}\partial^{\sigma}\nabla\phi\right)\,. \qquad \text{(A.8)}\end{aligned}$$

We can not be sure if he Hamiltonian density is positive definite. This is due to the fact that one is not sure of the sign of the term proportional to $\gamma = \frac{\gamma_0}{(M_{\mathrm{Pl}}\,c)^2}$. If one assumes that γ_0 is of the order one, the coefficient is $\gamma \sim 10^{-38}$. Therefore, for energies smaller than the Planck energy $E < E_{Pl}$, the leading order terms in Eq. (A.7) are several orders of magnitude bigger than the quantum gravity corrections, which means that the Hamiltonian density on the whole is positive definite and does *not* have Ostrogradsky instability. Conclusion about the positive definition of the Hamiltonian for energies bigger than the Planck energy $E > E_{Pl}$ cannot be drawn. Making the QFT presented here an effective theory.

Appendix B. Modified Poincaré Algebra Representations

Let us explore the RGUP modified Poincaré algebra. The physical position x_{μ} and momentum p_{μ} are functions of the auxiliary ones represented as

$$x^{\mu} = x^{\mu}_{0} \tag{B.1a}$$

$$p^{\mu} = p^{\mu}_{0}\left(1 + \gamma p^{\rho}_{0}p_{0\rho}\right). \tag{B.1b}$$

While the Lorentz generators are defined as

$$M^{\mu\nu} = p^{\mu}x^{\nu} - p^{\nu}x^{\mu} = \left[1 + \gamma p^{\rho}_{0}p_{0\,\rho}\right]\tilde{M}^{\mu\nu}\,, \tag{B.2}$$

where it is wort mentioning that all the results are truncated to first order in the RGUP parameter γ. The modified Poincaré algebra is then calculated to be

$$[x^\mu, M^{\nu\rho}] = i\hbar[1+\gamma p^\rho p_\rho]\,(x^\nu\delta^{\mu\rho} - x^\rho\delta^{\mu\nu}) + i\hbar 2\gamma p^\mu M^{\nu\rho} \tag{B.3a}$$

$$[p^\mu, M^{\nu\rho}] = i\hbar[1+\gamma p^\rho p_\rho]\,(p^\nu\delta^{\mu\rho} - p^\rho\delta^{\mu\nu})\,, \tag{B.3b}$$

$$[M^{\mu\nu}, M^{\rho\sigma}] = i\hbar\,(1+\gamma p^\rho p_\rho)\,(\eta^{\mu\rho}M^{\nu\sigma} - \eta^{\mu\sigma}M^{\nu\rho} - \eta^{\nu\rho}M^{\mu\sigma} + \eta^{\nu\sigma}M^{\mu\rho})\,. \tag{B.3c}$$

We can show that the physical rotations J_i and boosts K_i are as follows

$$J_i = \frac{1}{2}\varepsilon_{imn}M^{mn} = \frac{1}{2}\,(1+\gamma p^\rho p_\rho)\,\varepsilon_{imn}\tilde{M}^{mn} \tag{B.4a}$$

$$K_i = M_{0i} = (1+\gamma p^\rho p_\rho)\,\tilde{M}_{0i}\,, \tag{B.4b}$$

The algebra for which is

$$[J_i, J_j] = -i\varepsilon_{ijk}\,(1+\gamma p^\rho p_\rho)\,J^k\,, \tag{B.5a}$$

$$[K_i, K_j] = i\varepsilon_{ijk}\,(1+\gamma p^\rho p_\rho)\,J^k\,, \tag{B.5b}$$

$$[J_i, K_j] = i\varepsilon_{ijk}\,(1+\gamma p^\rho p_\rho)\,K^k\,. \tag{B.5c}$$

Once again we can define new set of this time physical operators A_i and B_i

$$A_i = \frac{1}{2}\,(J_i + iK_i) = \frac{1}{2}\,(1+\gamma p^\rho p_\rho)\,\left(\tilde{J}_i + i\tilde{K}_i\right) \tag{B.6a}$$

$$B_i = \frac{1}{2}\,(J_i - iK_i) = \frac{1}{2}\,(1+\gamma p^\rho p_\rho)\,\left(\tilde{J}_i - i\tilde{K}_i\right)\,. \tag{B.6b}$$

Which form the following algebra

$$[A_i, A_j] = -i\varepsilon_{ijk}\,(1+\gamma p^\rho p_\rho)\,A^k\,, \tag{B.7a}$$

$$[B_i, B_j] = -i\varepsilon_{ijk}\,(1+\gamma p^\rho p_\rho)\,B^k\,, \tag{B.7b}$$

$$[A_i, B_j] = 0\,. \tag{B.7c}$$

We notice that once again this algebra is equivalent to $SU(2)\otimes SU(2)$ algebra. The irreducible representation of which are scalar $(0,0)$, spinors/fermions $(0,1/2)$ and $(1/2,0)$, vector bosons/ photons $(1/2,1/2)$, and potentially gravitons $(1,1)$. We have taken two things in consideration: First that the modified Poincaré algebra has both the auxiliary momentum squared $p_{0\,\mu}p_0^\mu$ and the physical momentum squared $p_\mu p^\mu$ as Casimir invariants; And second that those quantities are scalars and therefore modify the structure constants of the algebra without actually changing its physical properties.

References

1. A. Kempf, G. Mangano and R. B. Mann, Hilbert space representation of the minimal length uncertainty relation, *Phys. Rev. D* **52**, 1108 (1995).
2. R. J. Adler and D. I. Santiago, On gravity and the uncertainty principle, *Mod. Phys. Lett. A* **14**, p. 1371 (April 1999).
3. R. J. Adler, P. Chen and D. I. Santiago, The generalized uncertainty principle and black hole remnants, *Gen. Relat. Grav.* **33**, p. 2101–2108 (Dec 2001).

4. A. F. Ali, S. Das and E. C. Vagenas, A proposal for testing Quantum Gravity in the lab, *Phys. Rev. D* **84**, p. 044013 (2011).
5. A. F. Ali and B. Majumder, Towards a cosmology with minimal length and maximal energy, *Class. Quant. Grav.* **31**, p. 215007 (Oct 2014).
6. A. F. Ali, M. Faizal and M. M. Khalil, Short distance physics of the inflationary de sitter universe, *JCAP* **2015**, p. 025–025 (Sep 2015).
7. A. Alonso-Serrano, M. P. Dabrowski and H. Gohar, Generalized uncertainty principle impact onto the black holes information flux and the sparsity of hawking radiation, *Phys. Rev. D* **97** (Feb 2018).
8. D. Amati, M. Ciafaloni and G. Veneziano, Can spacetime be probed below the string size?, *Phys. Lett. B* **216**, 41 (January 1989).
9. G. Amelino-Camelia, Testable scenario for relativity with minimum length, *Phys. Lett. B* **510**, p. 255–263 (Jun 2001).
10. G. Amelino-Camelia, Relativity in spacetimes with short-distance structure governed by an observer-independent (planckian) length scale, *Int. J. Mod. Phys. D* **11**, p. 35–59 (Jan 2002).
11. G. Amelino-Camelia, Quantum-Spacetime phenomenology, *Living Rev. Relativ.* **16**, p. 5 (June 2013).
12. P. Bargueño and E. C. Vagenas, Semiclassical corrections to black hole entropy and the generalized uncertainty principle, *Phys. Lett. B* **742**, p. 15–18 (Mar 2015).
13. C. Bambi and F. R. Urban, Natural extension of the generalised uncertainty principle, *Class. Quant. Grav.* **25**, p. 095006 (September 2007).
14. M. Bawaj and et al., Probing deformed commutators with macroscopic harmonic oscillators, *Nature Communications* **6** (Jun 2015).
15. M. Bojowald and A. Kempf, Generalized uncertainty principles and localization of a particle in discrete space, *Phys. Rev. D* **86**, p. 085017 (December 2011).
16. B. Bolen and M. Cavagliá, (anti-)de sitter black hole thermodynamics and the generalized uncertainty principle, *Gen. Relat. Grav.* **37**, 1255 (July 2005).
17. P. Bosso and S. Das, Generalized uncertainty principle and angular momentum, *Ann. of Phys.* **383**, 416 (2017).
18. P. Bosso, S. Das and R. B. Mann, Potential tests of the Generalized Uncertainty Principle in the advanced LIGO experiment, *Phys. Lett. B* **785**, 498 (2018).
19. P. Bosso, Rigorous Hamiltonian and Lagrangian analysis of classical and quantum theories with minimal length, *Phys. Rev. D* **97**, p. 126010 (2018).
20. P. Bosso and O. Obregón, Minimal length effects on quantum cosmology and quantum black hole models, *Class. Quant. Grav.* **37**, p. 045003 (2020).
21. D. J. Burger and et al, Towards the Raychaudhuri equation beyond general relativity, *Phys. Rev. D* **98** (Jul 2018).
22. P. Bushev and et al, Testing the generalized uncertainty principle with macroscopic mechanical oscillators and pendulums, *Phys. Rev. D* **100**, p. 066020 (2019).
23. S. Capozziello, G. Lambiase and G. Scarpetta, Generalized uncertainty principle from quantum geometry, *Int. J. Theor. Phys.* **39**, 15 (2000).
24. R. Casadio and F. Scardigli, Generalized uncertainty principle, classical mechanics, and general relativity, *Phys. Lett. B* **807**, p. 135558 (Aug 2020).
25. L. N. Chang and et al, On the Minimal Length Uncertainty Relation and the Foundations of String Theory, *Adv. High Energy Phys.* **2011**, p. 493514 (2011).
26. J. L. Cortes and J. Gamboa, Quantum uncertainty in doubly special relativity, *Phys. Rev. D* **71**, p. 065015 (2005).
27. R. N. Costa Filho and et al, Extended uncertainty from first principles, *Phys. Lett. B* **755**, 367 (April 2016).

28. M. P. Dabrowski and F. Wagner, Extended Uncertainty Principle for Rindler and cosmological horizons, *Eur. Phys. J. C* **79**, p. 716 (2019).
29. M. P. Dabrowski and F. Wagner, Asymptotic generalized extended uncertainty principle, *Eur. Phys. J. C* **80**, p. 676 (June 2020).
30. S. Das and E. C. Vagenas, Universality of Quantum Gravity Corrections, *Phys. Rev. Lett.* **101**, p. 221301 (2008).
31. S. Das and E. C. Vagenas, Phenomenological Implications of the Generalized Uncertainty Principle, *Can. J. Phys.* **87**, 233 (2009).
32. S. Das, E. C. Vagenas and A. F. Ali, Discreteness of Space from GUP II: Relativistic Wave Equations, *Phys. Lett. B* **690**, 407 (2010), [Erratum: Phys. Lett. B 692, 342–342 (2010)].
33. S. Das and R. Mann, Planck scale effects on some low energy quantum phenomena, *Phys. Lett. B* **704**, 596 (2011).
34. S. Das, M. P. G. Robbins and M. A. Walton, Generalized Uncertainty Principle Corrections to the Simple Harmonic Oscillator in Phase Space, *Can. J. Phys.* **94**, 139 (2016).
35. S. Das, S. S. Haque and B. Underwood, Constraints and horizons for de sitter with extra dimensions, *Phys. Rev. D* **100** (Aug 2019).
36. A. Das, S. Das and E. C. Vagenas, Discreteness of space from gup in strong gravitational fields, *Phys. Lett. B* **809**, p. 135772 (Oct 2020).
37. A. Garcia-Chung and et al, Propagation of quantum gravity-modified gravitational waves on a classical FLRW spacetime, *Phys. Rev. D* **103**, p. 084053 (2021).
38. S. B. Giddings, Black holes and other clues to the quantum structure of gravity, *Galaxies* **9**, p. 16 (December 2020).
39. B. Hamil, M. Merad and T. Birkandan, Applications of the extended uncertainty principle in AdS and ds spaces, *Eur. Phys. J. Plus* **134**, p. 278 (June 2019).
40. S. Hossenfelder, Interpretation of quantum field theories with a minimal length scale, *Phys. Rev. D* **73**, p. 105013 (2006).
41. S. Hossenfelder, Minimal Length Scale Scenarios for Quantum Gravity, *Living Rev. Rel.* **16**, p. 2 (2013).
42. M. Kober, Gauge Theories under Incorporation of a Generalized Uncertainty Principle, *Phys. Rev. D* **82**, p. 085017 (2010).
43. K. Konishi, G. Paffuti and P. Provero, Minimum physical length and the generalized uncertainty principle in string theory, *Phys. Lett. B* **234**, 276 (1990).
44. M. Maggiore, A generalized uncertainty principle in quantum gravity, *Phys. Lett. B* **304**, 65 (1993).
45. M. Maggiore, Quantum groups, gravity, and the generalized uncertainty principle, *Phys. Rev. D* **49**, 5182 (May 1994).
46. F. Marin *et al.*, Gravitational bar detectors set limits to Planck-scale physics on macroscopic variables, *Nature Phys.* **9**, 71 (2013).
47. C. A. Mead, Observable consequences of Fundamental-Length hypotheses, *Phys. Rev.* **143**, 990 (March 1966).
48. H. Moradpour, S. Aghababaei and A. H. Ziaie, A note on effects of generalized and extended uncertainty principles on jüttner gas, *Symmetry* **13**, p. 213 (January 2021).
49. J. R. Mureika, Extended uncertainty principle black holes, *Phys. Lett. B* **789**, 88 (February 2019).
50. Y. S. Myung, Y.-W. Kim and Y.-J. Park, Black hole thermodynamics with generalized uncertainty principle, *Phys. Lett. B* **645**, p. 393–397 (Feb 2007).
51. M.-I. Park, The generalized uncertainty principle in (A)dS space and the modification of hawking temperature from the minimal length, *Phys. Lett. B* **659**, 698 (January 2008).

52. F. Scardigli, Generalized uncertainty principle in quantum gravity from micro - black hole gedanken experiment, *Phys. Lett. B* **452**, 39 (April 1999).
53. H. S. Snyder, Quantized space-time, *Phys. Rev.* **71**, 38 (Jan 1947).
54. M. Sprenger, P. Nicolini and M. Bleicher, Neutrino oscillations as a novel probe for a minimal length, *Class. Quant. Grav* **28**, p. 235019 (Nov 2011).
55. L. Sriramkumar and S. Shankaranarayanan, Path integral duality and Planck scale corrections to the primordial spectrum in exponential inflation, *JHEP* **12**, p. 050 (2006).
56. D. J. Stargen, S. Shankaranarayanan and S. Das, Polymer quantization and advanced gravitational wave detector, *Phys. Rev. D* **100** (Oct 2019).
57. L. Tedesco, Fine structure constant, domain walls, and generalized uncertainty principle in the universe, *Int. J. Math. and Math. S.* **2011** (2011).
58. B. Wang, C. Long, Z. Long and T. Xu, Solutions of the schrödinger equation under topological defects space-times and generalized uncertainty principle, *Eur. Phys. J. Plus* **131**, p. 378 (2016).
59. S. Hossenfelder, The Soccer-Ball Problem, *SIGMA* **10**, p. 074 (2014).
60. G. Amelino-Camelia, Planck-scale soccer-ball problem: A case of mistaken identity, *Entropy* **19**, p. 400 (2017).
61. M. J. Lake, A Solution to the Soccer Ball Problem for Generalized Uncertainty Relations, *Ukr. J. Phys.* **64**, 1036 (2019).
62. V. Todorinov, P. Bosso and S. Das, Relativistic generalized uncertainty principle, *Ann. Phys.* **405**, 92 (October 2018).
63. C. Quesne and V. Tkachuk, Lorentz-covariant deformed algebra with minimal length, *Czech. J. Phys.* **56**, 1269 (2006).
64. P. Bosso, S. Das and V. Todorinov, Quantum field theory with the generalized uncertainty principle i: Scalar electrodynamics, *Ann. Phys.* **422**, p. 168319 (May 2020).
65. P. Bosso, S. Das and V. Todorinov, Quantum field theory with the generalized uncertainty principle II: Quantum electrodynamics, *Ann. Phys.* **424**, p. 168350 (May 2020).
66. J. Pons, Ostrogradski Theorem for Higher Order Singular Lagrangians, *Lett. Math. Phys.* **17**, p. 181 (1989).
67. R. P. Woodard, Ostrogradsky's theorem on Hamiltonian instability, *Scholarpedia* **10**, p. 32243 (2015).
68. F. de Urries and J. Julve, Ostrogradski formalism for higher derivative scalar field theories, *J. Phys. A* **31**, 6949 (1998).
69. V. Todorinov, Relativistic Generalized Uncertainty Principle and Its Implications (2020).
70. P. Balek, Charged-hadron suppression in Pb+Pb and Xe+Xe collisions measured with the ATLAS detector, *Nucl. Phys. A* **982**, 571 (2019).

Stelle gravity as the limit of quantum gravity with maximal momentum

V. Nenmeli* and S. Shankaranarayanan†

Department of Physics, Indian Institute of Technology Bombay, Mumbai, Maharashtra, 400076, India
**vijay.nenmeli@iitb.ac.in*
†shanki@phy.iitb.ac.in

V. Todorinov* and S. Das†

Theoretical Physics Group and Quantum Alberta, Department of Physics and Astronomy, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, T1K 3M4, Canada
**v.todorinov@uleth.ca*
†saurya.das@uleth.ca

Many quantum gravity theories predict several interesting phenomenological features such as minimal length scales and maximal momenta. Generalized uncertainty principles (GUPs), which are extensions of the standard Heisenberg uncertainty principle, have proven very useful in modelling the effects of such features on physics at sub-Planck energy scales. In this talk, we use a GUP modelling maximal momentum to establish a correspondence between the GUP modified dynamics of a massless spin 2 field and Stelle gravity with suitably constrained parameters. Thus, Stelle gravity can be regarded as the classical manifestation of the imposition of a momentum cutoff at the quantum gravity level. We then study the applications of Stelle gravity to cosmology. Specifically, we analytically show that Stelle gravity, when applied to a homogeneous, isotropic background, leads to inflation with exit. Lastly, using numerical simulations and data from CMB observations, we obtain strong bounds on the GUP parameter. Unlike previous works which fixed only upper bounds for GUP parameters, we show that we can bound the GUP parameter from above and from below.

Keywords: Generalized Uncertainty Principles, Quadratic Gravity, Inflation, Based on [1]

1. Introduction

Generalized Uncertainty Principles (GUPs) are phenomenological tools used to model generic features of Quantum Gravity theories such as minimal measurable length-scales[2] and maximal momentum.[3] Since GUPs have traditionally been formulated in non-covariant settings, they have primarily been used for computing Planck scale corrections to classical and quantum *mechanics*. Studying corrections to quantum *field theory* necessitates the use of a covariant GUP. Recently, Todorinov et al. introduced a class of such GUPs, and their effects on relativistic quantum mechanics were elucidated.[4] Following this, the authors considered a covariant, maximal momentum GUP modeled by the modified commutator (we set $\hbar = c = 1$)

$$[x^\mu, p^\nu] = i\eta^{\mu\nu}(1 - \gamma p^\sigma p_\sigma) - 2i\gamma p^\mu p^\nu \tag{1}$$

and studied the effects of this RGUP on free spin-0,[5] spin-1/2[6] and spin-1 fields. Combining these observations, the corrections to scalar and spinor electrodynamics were also obtained.[5,6] A natural follow-up would be to study the effects of GUPs on spin-2 fields (and, by extension, General Relativity). We would then have a complete lexicon detailing the effects of GUPs on all the "basic" quantum fields. Practically, such a study would also allow us to study deviations from General Relativity (GR) in the strong-field regime.

2. Computing RGUP Corrections — An Overview

From the RGUP (1), it is clear that the position and momentum operators are no longer canonically conjugate. The introduction of a "canonical" 4-momentum p_0^μ satisfying $[x^\mu, p_0^\nu] = i\eta^{\mu\nu}$ simplifies calculations considerably. The "physical" 4-momentum p^μ can be expressed in terms of p_0^μ as[5,6]

$$p^\mu = p_0^\mu(1 - \gamma p_{0\nu} p_0^\nu). \tag{2}$$

We can break down the procedure followed for computing RGUP corrections to QED into stages:

A. Compute the RGUP modified Poincare algebra and show that $p^\mu p_\mu$ is a Casimir operator.
B. Starting with the free field equations for spin-0, spin-1/2 and spin-1, obtain the modified EOM by replacing the canonical 4-momentum by the physical 4-momentum.
C. Using the Ostrogradsky approach, find actions that yield the modified EOM.
D. Minimally couple the matter and gauge field actions to obtain the modified QED action.
E. Compute corrections to cross-sections arising from additional interaction terms and compare with existing high-energy experiments.

When computing corrections to the spin-2 field, our aim is not to evaluate cross-sections but to search for modifications to GR. Hence, the procedure we must follow is slightly different. As with QED, we proceed with finding the GUP-modified EOM for the spin-2 field. After this, however, to connect our results to GR, we search for modified gravity theories *whose EOM, in the linear limit, match the GUP modified spin-2 field equations*. We then claim that this modified gravity theory can be regarded as the classical limit of spin-2 theory with the imposition of a maximal momentum scale.

3. RGUP Corrections to the Spin-2 Field

While the spin-2 field equations can be obtained by simply linearizing the Einstein field equations, the RGUP corrections to the EOM seem more natural if we instead

begin with a field theoretical approach. Such an approach was first outlined by Gupta[7] and popularized by Feynman[8] and can be summarized as follows:

A. The Lagrangian for the *free* spin-2 field must be bilinear in the field, $h_{\mu\nu}$, and its derivatives. The relevant bilinears are easily listed out.
B. Since the gravitational force is long range, the mediating gauge bosons must be massless. Thus, we may assume that the Lagrangian has no mass term and is strictly a function of field derivatives.
C. We can construct a *minimal* list by identifying pairs of bilinears that differ by surface terms. We can thus fix the action up to undetermined coefficients. In particular, we have

$$\mathcal{L} = ah_{\mu\nu,\sigma}h^{\mu\nu,\sigma} + bh_{\mu\nu}{}^{,\nu}h^{\mu\sigma}{}_{,\sigma} + ch_{\mu\nu}{}^{,\nu}h^{\sigma}{}_{\sigma,\mu} + dh_{\mu}{}^{\mu}{}_{,\sigma}h^{\nu}{}_{\nu}{}^{,\sigma} \tag{3}$$

D. *If* matter-gravity interactions were present, we would presumably have an interaction term of the form $\lambda h_{\mu\nu}T^{\mu\nu}$. By adding such a term to the above action and appealing to energy-momentum conservation (i.e. $T^{\nu}_{\mu\nu} = 0$), we get

$$(2a+b)\Box h_{\mu\nu}{}^{\nu} + (b+c)h^{;\nu\sigma}_{\nu\sigma}{}_{,\mu} + (c+2d)\Box h_{\nu}{}^{\nu}{}_{,\mu} = 0 \tag{4}$$

which fixes the coefficients.

Thus, up to an irrelevant scale factor, we see that the Lagrangian for the free spin-2 field is given by

$$\mathcal{L}_{free} = \frac{1}{2}h_{\mu\nu,\sigma}h^{\mu\nu,\sigma} - h_{\mu\nu}{}^{,\nu}h^{\mu\sigma}{}_{,\sigma} + h_{\mu\nu}{}^{,\nu}h^{\sigma}{}_{\sigma,\mu} - \frac{1}{2}h_{\mu}{}^{\mu}{}_{,\sigma}h^{\nu}{}_{\nu}{}^{,\sigma} \tag{5}$$

The EOM are obtained from the resulting action by variation

$$-G^{L}_{\mu\nu} = \Box h_{\mu\nu} - h_{\mu\sigma,\nu}{}^{\sigma} - h_{\nu\sigma,\mu}{}^{\sigma} + h_{\sigma,\mu\nu}{}^{\sigma} + \eta_{\mu\nu}h^{;\sigma\rho}_{\sigma\rho} - \eta_{\mu\nu}\Box h_{\sigma}{}^{\sigma} = 0, \tag{6}$$

where $G^{L}_{\mu\nu}$ is the linearized Einstein tensor. The RGUP modified EOM are obtained from the position space representation of (2) (i.e. $\partial_\mu \to \partial_\mu(1-\gamma\Box)$). We have

$$G^{L}_{\mu\nu} + \gamma\mathcal{G}_{\mu\nu} = 0, \tag{7}$$

where

$$\mathcal{G}_{\mu\nu} = \Box^2 h_{\mu\nu} - \Box h_{\mu\sigma,\nu}{}^{,\sigma} - h_{\nu\sigma\mu}{}^{,\sigma}. + \Box h_{\sigma,\mu\nu}{}^{,\sigma} + \eta_{\mu\nu}h^{;\sigma\rho}_{\sigma\rho} - \eta_{\mu\nu}\Box^2 h^{\sigma}_{\sigma}. \tag{8}$$

4. Stelle Gravity and GUP modified GR

The Einstein-Hilbert action describing the most general quadratic gravity theory (Stelle gravity) is[9]

$$S_{Stelle} = \int d^4x\sqrt{-g}[\frac{1}{2\kappa^2}R - \alpha R_{\mu\nu}R^{\mu\nu} + \beta R^2]. \tag{9}$$

By studying static, localized and isotropic solutions in the linear limit, it was shown[9] that the resultant theory contains two extra gauge bosons with masses $\frac{1}{\sqrt{2\alpha}\kappa}$ and

$\frac{1}{\sqrt{4(3\beta-\alpha)}\kappa}$. where $\kappa^2 = 1/8\pi G$. To avoid tachyonic instabilities, α and β must satisfy

$$3\beta \geq \alpha \geq 0. \tag{10}$$

The EOM for Stelle gravity are

$$\frac{1}{2\kappa^2}G_{\mu\nu} + (\alpha - 2\beta)R_{;\mu\nu} - \alpha\Box R_{\mu\nu} - \left(\frac{\alpha}{2} - 2\beta\right) g_{\mu\nu}\Box R - 2\alpha R^{\rho\sigma}R_{\mu\rho\nu\sigma} \tag{11}$$
$$+2\beta R R_{\mu\nu} + \frac{1}{2}g_{\mu\nu}(\alpha R^{\rho\sigma}R_{\rho\sigma} - \beta R^2) = 0.$$

To first order in $h_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu}$ (where $\eta_{\mu\nu}$ is the Minkowski metric), we have

$$\frac{1}{2\kappa^2}G^L_{\mu\nu} - (\alpha - 2\beta)(\Box h^{\lambda}_{\lambda,\mu\nu} - h^{\lambda,\rho}_{\rho,\lambda\mu\nu}) + \frac{\alpha}{2}(\Box^2 h_{\mu\nu} + \Box h^{\lambda}_{\lambda,\mu\nu} - \Box h^{\lambda}_{\mu,\lambda\nu} - \Box h^{\lambda}_{\nu,\lambda\mu})$$
$$+ \left(\frac{\alpha}{2} - 2\beta\right)\eta_{\mu\nu}(\Box^2 h^{\lambda}_{\lambda} - \Box h^{\lambda\rho}_{,\lambda\rho}) = 0. \tag{12}$$

It is easy to show that (12) match (7) perfectly when $\alpha = 2\beta = \gamma/\kappa^2$. From (10), this means that we must have $\gamma \geq 0$. Thus, we may claim that $\alpha = 2\beta$ Stelle gravity is a modified gravity theory modelling maximal momentum. Note, however, that it is by no means the *unique* modified gravity theory modelling maximal momentum. This is because the effects of curvature terms of order at least three will not appear in the linear regime. Thus, a more precise statement would be that $\alpha = 2\beta$ Stelle gravity is the *minimally modified, metric-only* theory of gravity which models the effects of maximal momentum.

There are two further points worthy of notice. First, it has been shown[10,11] that Stelle Gravity (and thus GUP modified gravity) does not contain malicious ghosts. Second, when $\alpha = 2\beta$, the additional gauge bosons mentioned earlier have *equal* masses. In this sense, GUP modified GR is a *degenerate* theory.

5. Applications to Cosmology

It is well known that quantum corrections to GR can lead to inflation. A natural question to ask is whether we can obtain inflation from GUP modified GR. To answer this, we compute the EOM for *generic* Stelle gravity (with parameters α and β) in an FRW background. We assume a flat background although the results can be extended. In terms of the Hubble parameter $H(t) \equiv \dot{a(t)}/a(t)$, we have

$$H^2 - 4\lambda[\dot{H}^2 - 2H\ddot{H} - 6H^2\dot{H}] = 0 \tag{13}$$

and

$$3H^2 + 2\dot{H} + \lambda\left[18H^2\dot{H} + 9\dot{H}^2 + 12H\ddot{H} + 2\dddot{H}\right] = 0. \tag{14}$$

where the arguments have been suppressed and $\lambda \equiv (3\beta - \alpha)/\kappa^2$. From these equations, it is clear that the *qualitative* dependence of the solutions is *independent* of

the parameter values. Thus, we can qualitatively study the dynamics of *any* Stelle gravity theory by observing the dynamics of one particular case. Moreover, since Stelle gravity with $\alpha = 0$ corresponds to the Starobinsky action[12] which is a well studied inflationary model, it is clear that *any* Stelle gravity theory and in particular GUP modified GR leads to inflation.

6. Numerics

We can also graphically verify the existence of inflation with exit. To do so, we would like to numerically solve the EOM and study the properties of the resulting solutions. While there are apparently two equations, (13) and (14), the two are completely equivalent as they are related via the Bianchi identities. Therefore, it is preferable to solve (13) as it is second order in time as opposed to (14) which is third order. However, the large scale of H during inflation and the small scale of γ make it difficult to solve (13) numerically. We, therefore, define the following dimensionless variables:

$$s = \frac{t}{\sqrt{\gamma}}; \quad b(s) = \ln\left(\frac{H(s)}{H_{inf}}\right); \quad U = \sqrt{\gamma}\, H_{inf} \tag{15}$$

where H_{inf} is a constant denoting the energy scale during inflation. In terms of the new dynamical variable $b(s)$, (13) is transformed to:

$$1 + 2\dot{b}(s)^2 + 4\ddot{b}(s) + 12Ue^{b(s)}\dot{b}(s) = 0 \tag{16}$$

While γ and H_{inf} are respectively small and large, their composition,U, is not too large or too small. Additionally, since $b(s)$ is clearly $\mathcal{O}(1)$ during inflation, solving (16) is quite tractable. Figure 1a depicts the variation of the scale factor $a(s)$ with s, the rescaled cosmic time, for various values of U. The existence of inflection points graphically demonstrates the existence of inflation with exit. Figure 1b shows the variation of the slow roll parameter $\epsilon(s)$ with s for various values of U. The necessary condition for inflation is $\epsilon < 1$. Such behaviour is evident from the graphs, along with a crossover to $\epsilon > 1$, which shows the exit from inflation.

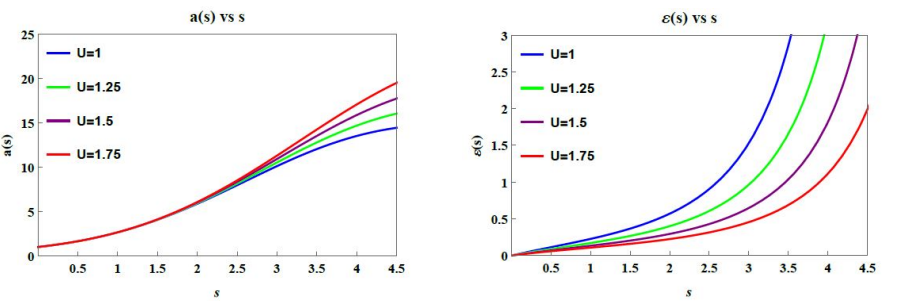


Fig. 1. (a) Variation of a with s. (b) Variation of ϵ with s.

7. Parameter Estimation

The fact that GUP modified GR predicts inflation means that we should able to constrain the GUP parameter γ using the existing constraints on inflation. These constraints stem from theoretical considerations and CMB observations. To do this, we plot the variation of s_0, the exit time (defined as the value of s when $\epsilon = 1$), with U. We see that $s_0 = U$ holds as an order of magnitude relation. This observation can be used to constrain the GUP parameter. Define $t_0 = s_0\sqrt{\gamma}, \gamma_0 = \gamma/l_{Pl}^2$ and $\alpha = t_0/l_{Pl}$. t_0 is simply the exit time is units of cosmic time while α is the exit time in units of Planck time/Planck length (l_{Pl}) and γ_0 is the ratio of two energy scales - the Planck scale and the GUP energy scale. We now have

$$s_0 = \frac{\alpha l_{Pl}}{\sqrt{\gamma}} = \frac{\alpha}{\gamma_0}. \tag{17}$$

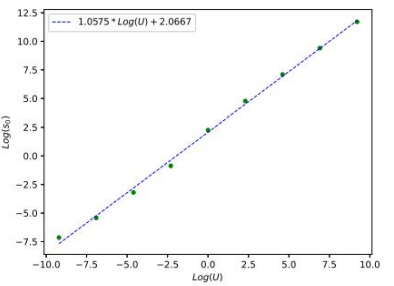


Fig. 2. Variation of s_0 with U.

Since $s_0 \sim U$, we can replace the former with the latter in (17). Simplifying, we get

$$\gamma_0 = \frac{\alpha}{H_{inf} l_{Pl}}. \tag{18}$$

Note that the right hand side is now a function of constants and known parameters. From PLANCK data,[13] we have $H_{inf} \leq 10^{-4} M_{Pl}$. We take $H_{inf} = 10^{-5} M_{Pl}$ which also coincides from the GUT scale. Typically, $\alpha \in (10^9 - 10^{12})$, but we stretch these bounds to $(10^6 - 10^{15})$ to account for errors stemming from the "hand-waving" nature of the argument. Substituting these bounds in (18) and utilizing the additional constraint that the GUP energy scale must exceed the CERN energy scale $\sim 10TeV$, we obtain the following bounds

$$10^{10} \leq \gamma_0 \leq 10^{15} \tag{19}$$

or

$$10^{-28} GeV^{-2} \leq \gamma \leq 10^{-23} GeV^{-2}. \tag{20}$$

In addition to being among the most stringent bounds available (which is expected since, unlike most constraining procedures, the one used above involved inflation and

hence the strong gravity regime), our bounds are double-sided. To our knowledge, this is the first time double-sided bounds have been obtained for GUP parameters.

8. Conclusions and Future Prospects

One expects that the imposition of a maximal momentum scale at the quantum level leads to modifications to GR at the classical level. In this proceeding, we have shown that $\alpha = 2\beta$ Stelle gravity can be regarded as the minimal extension of GR that can model maximal momentum. We have also shown that GUP modified GR leads to inflation and used this feature to obtain some of the best-known bounds on GUP parameters. We reiterate that our bounds possess a unique feature - they are double-sided. Using the GUP-Stelle Gravity correspondence, we can study the effects of maximal momentum scales on gravitational systems by studying these systems using $\alpha = 2\beta$ Stelle gravity. Additionally, we believe that approaches similar to the one outlined in this proceeding can be used to study the effects of other Planck scale phenomena at the classical level.

GUP modified GR is degenerate. It is not clear what the origins or the implications of this degeneracy are. These are currently being looked into. Finally, it is worth pondering over whether we can extend the GUP used above to construct a two-parameter GUP that "corresponds," in the classical limit, to *generic* Stelle gravity.

References

1. V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das, Maximal Momentum GUP leads to quadratic gravity, *Physics Letters* **B821**, 136621 (2021).
2. A. Kempf, G. Mangano and R. B. Mann, Hilbert space representation of the minimal length uncertainty relation, *Physical Review D* **52**, 1108 (1995).
3. P. Pedram, A Higher Order GUP with Minimal Length Uncertainty and Maximal Momentum, *Physics Letters* **B714**, 317 (2012).
4. V. Todorinov, P. Bosso and S. Das, Relativistic Generalized Uncertainty Principle, *Annals of Physics* **405**, 92 (2018).
5. P. Bosso, V. Todorinov and S. Das, Quantum field theory with the generalized uncertainty principle I: Scalar electrodynamics, *Annals of Physics* **422**, 168319 (2020).
6. P. Bosso, V. Todorinov and S. Das, Quantum field theory with the generalized uncertainty principle II: Quantum Electrodynamics, *Annals of Physics* **424**, 168350 (2020).
7. S. N. Gupta, Gravitation and Electromagnetism, *Physical Review D* **96**, 1683 (1954).
8. R. P. Feynman, *The Feynman Lectures on Gravitation*, 1st edn. (CRC Press, 2003)
9. K. S. Stelle, Classical gravity with higher derivatives, *General Relativity and Gravitation* **9**, 353 (1978).
10. E. T. Tomboulis, Unitarity in Higher Derivative Quantum Gravity, *Physical Review Letters* **52**, 1173 (1984).
11. I. Antoniadis and E. T. Tomboulis, Gauge invariance and unitarity in higher-derivative quantum gravity, *Physical Review D* **33**, 2756 (1986).
12. A. A. Starobinsky, A new type of isotropic cosmological models without singularity, *Physics Letters* **B91**, 99 (1980).
13. Y. Akrami and Others, Planck 2018 results. X. Constraints on inflation, *Astronomy and Astrophysics Supplement Series* **641**, A10 (2020).

Baryon asymmetry and minimum length

Saurya Das* and Mitja Fridman†

Theoretical Physics Group and Quantum Alberta, Department of Physics and Astronomy, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, T1K 3M4, Canada
**E-mail: saurya.das@uleth.ca*
†E-mail: fridmanm@uleth.ca

Gaetano Lambiase

Dipartimento di Fisica E.R: Caianiello, Universita di Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano, Salerno, Italy & INFN - Gruppo Collegato di Salerno, Italy
E-mail: lambiase@sa.infn.it

Elias C. Vagenas

Theoretical Physics Group, Department of Physics, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait
E-mail: elias.vagenas@ku.edu.kw

We study Quantum Gravity effects in cosmology, and in particular that of the Generalized Uncertainty Principle on the Friedmann equations. We show that the Quantum Gravity induced variations of the energy density and pressure in the radiation dominated era provide a viable explanation of the observed baryon asymmetry in the Universe.

Keywords: Quantum Gravity, Quantum Gravity Phenomenology, Minimal Length, Cosmology, Baryon Asymmetry.

1. Introduction

So far there are two successful theories which describe the world around us. Quantum theory, which describes interactions on the smallest scales on one hand and General Relativity, which describes gravitational interactions on large scales on the other hand.[1,2] There is no straightforward way to combine these two theories to describe physics in a regime, where both contribute, which is why we need a theory of Quantum Gravity (QG). Several QG theories have been proposed, but there is no direct experimental evidence yet to verify them. However, phenomenologists have a good intuition what fundamental concepts the regime of QG may imply on physics.[3–11] One such concept is the existence of a minimal measurable length, predicted by all or most models of QG. A natural way to incorporate minimal length in our considerations is to modify the Heisenberg uncertainty principle.[12–30] The modified Heisenberg uncertainty principle is called the Generalized Uncertainty Principle (GUP) and in the commutator form it reads as[21]

$$[x_i, p_j] = i\hbar \left(\delta_{ij} - \alpha \left(p\delta_{ij} + \frac{p_i p_j}{p} \right) + \beta \left(p^2 \delta_{ij} + 3 p_i p_j \right) \right) \tag{1}$$

where $\alpha \equiv \alpha_0/(M_P c)$, $\beta \equiv \beta_0/(M_P c)^2$, α_0, β_0 are the dimensionless linear and quadratic GUP parameters and $M_P = \sqrt{\hbar c/G}$ is the Planck mass. This phenomenological model describes a length scale $\alpha_0\,\ell_P$ and $\sqrt{\beta_0}\,\ell_P$ between the electroweak scale, i.e., $\ell_{EW} \approx 10^{-18}\,\mathrm{m}$, and the Planck scale, i.e., $\ell_P = \sqrt{\hbar G/c^3} \approx 10^{-35}\,\mathrm{m}$. We apply the GUP to quantum quantum mechanical and gravitational systems and predict deviations from standard theory, which we hope experiments can detect now or in the future, or provide explanations for observed phenomena, as in the case of this work, namely the observed baryon asymmetry that is observed in the Universe.

The asymmetry between matter and anti-matter has been an unsolved problem in physics for a long time. Standard theories such as quantum theory and general relativity suggest that there is no reason why the asymmetry between matter and anti-matter in the Universe should exist.[31] This means there should be an equal amount of matter as there is anti-matter in the Universe, while observations show a very large asymmetry. To explain this observed asymmetry we have to satisfy the so-called Sakharov conditions[32]:

1) violation of baryon number B,
2) violation of C and CP symmetries and
3) deviation from thermal equilibrium.

If a proposed mechanism is able to satisfy the above three Sakharov conditions, then it may be possible to explain the origin of the observed baryon asymmetry in the Universe. The mechanism we use is an interaction term from supergravity theories which couples the baryon current and space-time to satisfy the first two Sakharov conditions, and the holographic principle with the thermodynamics of horizons to derive the GUP-modified Friedmann equations, which satisfy the third Sakharov condition. Several proposals how GUP modifies the Friedmann equations have been presented so far.[33–36] In this way we provide a viable explanation of the observed baryon symmetry in the Universe using the GUP.

The paper is structured as follows. In section 2 we modify the Bekenstein-Hawking entropy using GUP, then in section 3, by using the holographic principle we modify the Friedmann equations, and in section 4 by using these modified Friedmann equations we attempt to explain the observed baryon symmetry in the Universe. In section 5 we summarize the work with concluding remarks.

2. Modified Bekenstein-Hawking entropy

In order to modify the Friedmann equations, using the holographic principle, we need to modify the Bekenstein-Hawking entropy formula, since the Friedmann equations can be derived from the first law of thermodynamics, which requires the change in entropy. The holographic principle makes it possible to write a d-dimensional theory on a $d-1$ dimensional boundary.[37,38] We can re-write the GUP from Eq. (1) in the usual form[39] (where $\langle p \rangle = 0$)

$$\Delta x \Delta p \gtrsim \left[1 - \alpha \Delta p + 4\beta \Delta p^2\right] \,, \tag{2}$$

where we have set the numerical prefactor to be $\mathcal{O}(1)$.[40–42]

When a particle with energy E gets absorbed by a horizon, the horizon area changes by $\Delta A_{min} \geq 8\pi\ell_P^2 E\,\Delta x$, where Δx is the uncertainty in position of the absorbed particle.[43,44] To obtain the energy of the particle E, we solve the quadratic equation for Δp from Eq. (2) and obtain the expression

$$E = \Delta p \gtrsim \frac{\Delta x + \alpha}{8\beta}\left(1 - \sqrt{1 - \frac{16\beta}{\Delta x^2 + 2\alpha\Delta x + \alpha^2}}\right), \tag{3}$$

where we chose the negative solution, since we want to obtain the smallest change in area of the apparent horizon and this solution reduces to the standard Heisenberg uncertainty for $\alpha, \beta \longrightarrow 0$. The position uncertainty Δx is the diameter of the observable Universe $\Delta x = 2r_S$, where r_S is the Schwarzschild radius. The square of Δx is then related to the area of the apparent cosmic horizon as $\Delta x^2 = A/\pi$. Plugging all of the above in the change in ΔA_{min}, we get

$$\Delta A_{min} \simeq \lambda\frac{\ell_P^2(A + \alpha\sqrt{\pi}A^{1/2})}{\beta}\left(1 - \sqrt{1 - \frac{16\pi\beta}{A + 2\alpha\sqrt{\pi}A^{1/2} + \alpha^2\pi}}\right), \tag{4}$$

where λ is determined by the Bekenstein-Hawking entropy formula $b/\lambda = 2\pi$. Here $b = \Delta S_{min} = \ln 2$ is the minimal increase in entropy, corresponding to one bit of information.[35]

The Bekenstein-Hawking entropy was first introduced for a black hole, but the holographic principle suggests that it can be used for any horizon. Therefore we can write it for a cosmic horizon in the same way[45,46]

$$S = \frac{A}{4G} = \frac{A}{4\ell_P^2}\ . \tag{5}$$

If we introduce some modification to the Heisenberg uncertainty principle or we modify the entropy in any other way, the only variable in this equation, the area, gets modified.[33] The Bekenstein-Hawking entropy then reads $S = \frac{f(A)}{4\ell_P^2}$. Since the first law of thermodynamics includes a differential of entropy, it useful to calculate its derivative

$$\frac{\mathrm{d}S}{\mathrm{d}A} = \frac{f'(A)}{4\ell_P^2}\ . \tag{6}$$

On the other hand, the above derivative can be written using the minimal change in entropy ΔS_{min} and ΔA_{min}

$$\frac{\mathrm{d}S}{\mathrm{d}A} = \frac{\Delta S_{min}}{\Delta A_{min}} = \frac{\beta^*}{8\ell_P^2\left(A + \alpha^* A^{1/2} - \sqrt{A^2 + 2\alpha^* A^{3/2} + (\alpha^{*2} - \beta^*)A}\right)}\ , \tag{7}$$

where $\alpha^* = \sqrt{\pi}\alpha$ and $\beta^* = 16\pi\beta$. By comparing Eqs. (6) and (7) we can read out what $f'(A)$ is

$$f'(A) = \frac{1}{2}\frac{\beta^*}{\left(A + \alpha^* A^{1/2} - \sqrt{A^2 + 2\alpha^* A^{3/2} + (\alpha^{*2} - \beta^*)A}\right)}\ . \tag{8}$$

To obtain the modified Bekenstein-Hawking entropy we must integrate Eq. (6) over area A, from A to ∞, using $f'(A)$ from Eq. (8). The exact GUP-modified Bekenstein-Hawking entropy then reads

$$S = \frac{1}{8\ell_P^2}\left[A\left(1+\sqrt{1+2\alpha^*\frac{1}{A^{1/2}}+(\alpha^{*2}-\beta^*)\frac{1}{A}}\right)\right.$$
$$+\alpha^* A^{1/2}\left(2+\sqrt{1+2\alpha^*\frac{1}{A^{1/2}}+(\alpha^{*2}-\beta^*)\frac{1}{A}}\right)$$
$$\left.-\beta^*\ln\left(1+\frac{A^{1/2}}{\alpha^*}\left(1+\sqrt{1+2\alpha^*\frac{1}{A^{1/2}}+(\alpha^{*2}-\beta^*)\frac{1}{A}}\right)\right)\right]\ , \tag{9}$$

where we obtain the standard result from Eq.(5) for $\alpha^*, \beta^* \longrightarrow 0$.

3. Modified Friedmann Equations

The Friedmann equations describe the evolution of a dynamical Universe and provide the foundations for the standard model of cosmology. They can be derived in different ways. The most common is by solving the Einstein equations in $(n+1)$-dimensional space-time for a FLRW metric

$$\mathrm{d}s^2 = h_{cd}\mathrm{d}x^c\mathrm{d}x^d + \tilde{r}^2\mathrm{d}\Omega_{n-1}^2\ , \tag{10}$$

where $x^c = (t, r)$, $\tilde{r} = a(t)\, r$, $h_{cd} = \mathrm{diag}(-1, a^2/(1-kr^2))$, $\mathrm{d}\Omega_{n-1}$ is the angular part of the $(n-1)$-dimensional sphere, $a = a(t)$ is the scale factor, r is the comoving radius and k is the spatial curvature constant. Indices c and d only take values 0 and 1. The other way to derive the Friedmann equations is using the holographic principle, which takes advantage of the thermodynamics of horizons.

To generalize the Friedmann equations to include minimum length effects we derive the same from the first law of thermodynamics, with the difference of using the modified Bekenstein-Hawking entropy, derived in the previous section. The first law of thermodynamics for the content inside an apparent horizon is given by[33–36]

$$\mathrm{d}E = T\mathrm{d}S + W\mathrm{d}V\ , \tag{11}$$

where E is the energy contained inside the apparent horizon, given by $E = \rho V$, V is the volume of an n-dimensional sphere, given by $V = \Omega_n\, \tilde{r}_A^{\,n}$, with $\Omega_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$ which has an area of $A = n\,\Omega_n\, \tilde{r}_A^{\,n-1}$), T is the Hawking temperature given by[47,48]

$$T = \frac{\kappa}{2\pi} = -\frac{1}{2\pi\tilde{r}_A}\left(1-\frac{\dot{\tilde{r}}_A}{2H\tilde{r}_A}\right)\ , \tag{12}$$

$\mathrm{d}S$ is the differential of the Bekenstein-Hawking entropy given by Eq. (6), and W is the work density which reads as[49]

$$W = -\frac{1}{2}T^{cd}h_{cd} = \frac{1}{2}\left(\rho - p\right)\ . \tag{13}$$

In the above T^{cd} is a projection of the energy-momentum tensor $T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}$ on the $t - r$ subspace. What remains is to differentiate the energy and volume, write everything in terms of $\mathrm{d}A$, and plug the above equations in the first law of thermodynamics from Eq. (11). From here we derive the first Friedmann equation

$$-\frac{8\pi G}{n-1}(\rho + p) = \left(\dot{H} - \frac{k}{a^2}\right) f'(A)\ . \tag{14}$$

By using the continuity equation $\dot{\rho} + nH(\rho + p) = 0$, originating from the conservation of energy $T^{\mu\nu}{}_{;\nu} = 0$, with Eq.(14) and integrating it, one obtains the second Friedmann equation

$$-\frac{8\pi G}{n(n-1)}\rho = \frac{(n\Omega_n)^{\frac{n+1}{n-1}}}{n(n-1)\Omega_n}\int f'(A)\frac{\mathrm{d}A}{A^{\frac{n+1}{n-1}}}\ . \tag{15}$$

By setting $f'(A) = 1$ ($f(A) = A$ in standard theory) we obtain the standard Friedmann equations as we expect. However, to obtain GUP corrections to the Friedmann equations, we plug $f'(A)$ from Eq. (8) in Eqs. (14) and (15). Both of the Friedmann equations have exact solutions in $n = 3$, which read as

$$-4\pi G(\rho + p) = \left(\dot{H} - \frac{k}{a^2}\right) \times \frac{\beta^*}{8\pi}\frac{\left(H^2 + \frac{k}{a^2}\right)}{1 + \frac{\alpha^*}{(4\pi)^{1/2}}\left(H^2 + \frac{k}{a^2}\right)^{1/2} - \sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}\left(H^2 + \frac{k}{a^2}\right)^{1/2} + \frac{(\alpha^{*2} - \beta^*)}{4\pi}\left(H^2 + \frac{k}{a^2}\right)}} \tag{16}$$

and

$$\begin{aligned}\frac{8\pi G}{3}(\rho - \Lambda) = {} & \frac{1}{2}\left(H^2 + \frac{k}{a^2}\right) + \frac{\alpha^*}{3(4\pi)^{1/2}}\left(H^2 + \frac{k}{a^2}\right)^{3/2} + \frac{2\pi(\alpha^{*2} + 2\beta^*)}{3(\alpha^{*2} - \beta^*)^2} \\ & + \left[\frac{1}{3}\left(H^2 + \frac{k}{a^2}\right) + \frac{(4\pi)^{1/2}\alpha^*}{6(\alpha^{*2} - \beta^*)}\left(H^2 + \frac{k}{a^2}\right)^{1/2} - \frac{2\pi(\alpha^{*2} + 2\beta^*)}{3(\alpha^{*2} - \beta^*)^2}\right] \\ & \times \sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}\left(H^2 + \frac{k}{a^2}\right)^{1/2} + \frac{(\alpha^{*2} - \beta^*)}{4\pi}\left(H^2 + \frac{k}{a^2}\right)} \\ & + \frac{2\pi\alpha^*\beta^*}{(\alpha^{*2} - \beta^*)^{5/2}}\ln\left[1 + \frac{(\alpha^{*2} - \beta^*)}{(4\pi)^{1/2}(\alpha^* + \sqrt{\alpha^{*2} - \beta^*})}\left(H^2 + \frac{k}{a^2}\right)^{1/2}\right. \\ & + \frac{\sqrt{\alpha^{*2} - \beta^*}}{\alpha^* + \sqrt{\alpha^{*2} - \beta^*}} \\ & \left. \times \left(\sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}\left(H^2 + \frac{k}{a^2}\right)^{1/2} + \frac{(\alpha^{*2} - \beta^*)}{4\pi}\left(H^2 + \frac{k}{a^2}\right)} - 1\right)\right],\end{aligned} \tag{17}$$

where

$$A = 4\pi\tilde{r}_A^2 = \frac{4\pi}{H^2 + \frac{k}{a^2}} \tag{18}$$

was used. We are interested in the radiation dominated era, since the baryon asymmetry was frozen in at that epoch. In that epoch the small cosmological constant Λ can be ignored, since $\rho \gg \Lambda$ and the curvature constant can be set to $k = 0$, which is consistent with the observations which show a flat Universe. The modified Friedmann equations then get simplified to

$$-4\pi G\,(\rho + p) = \frac{\beta^*\dot{H}}{8\pi}\,\frac{H^2}{1 + \frac{\alpha^*}{(4\pi)^{1/2}}H - \sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}H + \frac{(\alpha^{*2} - \beta^*)}{4\pi}H^2}} \tag{19}$$

and

$$\begin{aligned}\frac{8\pi G}{3}\rho = {} & \frac{1}{2}H^2 + \frac{\alpha^*}{3(4\pi)^{1/2}}H^3 + \frac{2\pi(\alpha^{*2} + 2\beta^*)}{3(\alpha^{*2} - \beta^*)^2} \\ & + \left[\frac{1}{3}H^2 + \frac{(4\pi)^{1/2}\alpha^*}{6(\alpha^{*2} - \beta^*)}H - \frac{2\pi(\alpha^{*2} + 2\beta^*)}{3(\alpha^{*2} - \beta^*)^2}\right] \\ & \times \sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}H + \frac{(\alpha^{*2} - \beta^*)}{4\pi}H^2} \\ & + \frac{2\pi\alpha^*\beta^*}{(\alpha^{*2} - \beta^*)^{5/2}}\ln\left[1 + \frac{(\alpha^{*2} - \beta^*)}{(4\pi)^{1/2}(\alpha^* + \sqrt{\alpha^{*2} - \beta^*})}H\right. \\ & \left. + \frac{\sqrt{\alpha^{*2} - \beta^*}}{\alpha^* + \sqrt{\alpha^{*2} - \beta^*}}\left(\sqrt{1 + \frac{2\alpha^*}{(4\pi)^{1/2}}H + \frac{(\alpha^{*2} - \beta^*)}{4\pi}H^2} - 1\right)\right]. \end{aligned} \tag{20}$$

The Taylor expansions of the above modified Friedmann equations up to first order in α^* and β^* turn out to be

$$-4\pi G\,(\rho + p) = \dot{H}\left(1 + \frac{\alpha^*}{2\sqrt{\pi}}H - \frac{\beta^*}{16\pi}H^2\right) \tag{21}$$

and

$$\frac{8\pi G}{3}\rho = H^2\left(1 + \frac{\alpha^*}{3\sqrt{\pi}}H - \frac{\beta^*}{32\pi}H^2\right). \tag{22}$$

The above simplified equations are used to study the breaking of thermal equilibrium in the radiation dominated era, which satisfies one of the Sakharov conditions.

4. Gravitational Baryogenesis

To provide an explanation for the observed baryon asymmetry in the Universe, all Sakharov conditions must be satisfied. Theories of supergravity introduce an extra interaction term to the action which couples the baryon current and space-time[50]

$$\frac{1}{M_*^2}\int \mathrm{d}^4x\sqrt{-g}\,J^\mu\partial_\mu R\,, \tag{23}$$

and thus violates the C, CP and CPT symmetries. In the above M_* is the cutoff scale characterizing the effective theory.[51] The magnitude of M_* is chosen to be in the regime of the Grand Unified Theories (GUT), where interactions which violate the baryon number exist. The term in Eq. (23) therefore satisfies the first two Sakharov conditions. We can see that the integrand from Eq. (23) reduces to

$$\frac{1}{M_*^2} J^\mu \partial_\mu R = \frac{1}{M_*^2}(n_B - n_{\bar{B}})\dot{R} \ , \tag{24}$$

where n_B ($n_{\bar{B}}$) is the baryon (anti-baryon) number density. In the above equation only the time derivative survives, since the spatial derivative of the Ricci scalar vanishes $\nabla R = 0$ for the FLRW metric from Eq. (10). Because the integrand from Eq. (24) corresponds to the energy density term of the grand canonical ensemble, we can read the chemical potential for baryons and anti-baryons from it as

$$\mu_B = -\mu_{\bar{B}} = -\frac{\dot{R}}{M_*^2} \ . \tag{25}$$

To study the baryon asymmetry, we need a suitable variable, which parametrizes this asymmetry. Following Ref.[52], the baryon asymmetry parameter is defined as

$$\eta \equiv \frac{n_B - n_{\bar{B}}}{n_\gamma} \simeq 7\frac{n_B - n_{\bar{B}}}{s} \ , \tag{26}$$

where the photon number density n_γ and the entropy density s are related by a constant factor $n_\gamma \simeq s/7$ since the epoch of $e^\pm$ annihilation. For relativistic particles, the entropy density turns out to be[52]

$$s = \frac{2\pi^2 g_*}{45} T^3 \tag{27}$$

and the baryon number density[52]

$$n_B - n_{\bar{B}} = \frac{g_b}{6}\mu_B T^2 \ . \tag{28}$$

In the above equations $g_* \approx 106$ is the number of degrees of freedom for particles which contribute to the entropy of the universe and $g_b \sim \mathcal{O}(1)$ is the number of intrinsic degrees of freedom for baryons.[52] Plugging Eqs. (25), (27) and (28) in Eq. (26) the baryon asymmetry parameter reads

$$\eta \simeq -\frac{105\, g_b}{4\pi^2 g_*} \frac{\dot{R}}{M_*^2 T}\bigg|_{T_D} \ . \tag{29}$$

The above parameter is evaluated at some decoupling temperature T_D. This decoupling temperature represents the energy scale at which the baryon violating interactions freeze out and go out of equilibrium. The baryon number is conserved below this temperature for all future epochs.

Next, we use the modified Friedmann equations which turn out to break thermal equilibrium. To see that they indeed break thermal equilibrium, we write the energy density and pressure as their equilibrium values plus some small variations

$$\rho = \rho_0 + \delta\rho \, , \qquad p = p_0 + \delta p \ . \tag{30}$$

If there is no variations $\delta\rho = \delta p = 0$, we have thermal equilibrium and no baryon asymmetry is produced due to the third Sakharov condition. We require some variations $\delta\rho, \delta p \neq 0$ to deviate from the thermal equilibrium. To obtain non-vanishing variations we plug the energy density and pressure from Eq. (30) into the modified Friedman equations from Eqs. (21) and (22) to express $\delta\rho$ and δp in terms of the GUP parameters and the equation of state $p_0 = w\rho_0$. The variations then read

$$\delta\rho = \frac{\alpha^*}{3}\sqrt{\frac{8G}{3}}\rho_0^{3/2} - \frac{\beta^*}{12}G\rho_0^2 \tag{31}$$

and

$$\delta p = \frac{\alpha^*}{6}\sqrt{\frac{8G}{3}}(1+3w)\rho_0^{3/2} - \frac{\beta^*}{12}G(1+2w)\rho_0^2 \,, \tag{32}$$

We can see if there is no GUP corrections $\alpha^*, \beta^* = 0$, we get no variations in the energy density and pressure and thermal equilibrium is preserved.

The trace of the Einstein equations give us the Ricci scalar $R = -8\pi G\, T_g = -8\pi G\,(\rho - 3p)$, where we plug in the energy density and pressure from Eq. (30) with their respective variations from Eqs. (31) and (32). We compute its time derivative, use the continuity equation for $n = 3$ and evaluate the whole expression in the radiation-dominated era with $w = 1/3$, where the baryon violating interactions froze out. The time derivative of the Ricci scalar then reads

$$\dot{R} = -\alpha^* \frac{256}{3}\,\pi^{3/2}\,G^2\rho_0^2 + \beta^*\, 8\left(\frac{8\pi}{3}\right)^{3/2} G^{5/2}\rho_0^{5/2} \,. \tag{33}$$

We can see that the above derivative would also vanish if there would be no GUP corrections because every term is proportional to α^* or β^*. We plug Eq. (33) in Eq. (29) and use the equilibrium energy density for relativistic particles[52]

$$\rho_0 = \frac{\pi g_*}{30}T^4 \,, \tag{34}$$

to obtain the expression for the predicted baryon asymmetry

$$\eta = \alpha_0 \frac{112\,\pi^2 g_* g_b}{45}\left(\frac{T_D}{M_P}\right)^7\left(\frac{M_P}{M_*}\right)^2 - \beta_0\frac{896\sqrt{5}\,\pi^3 g_*^{3/2} g_b}{675}\left(\frac{T_D}{M_P}\right)^9\left(\frac{M_P}{M_*}\right)^2 \,, \tag{35}$$

where we expressed the gravitational constant in terms of the Planck mass, i.e., M_P, as $G = 1/M_P^2$. From the above we can see that the baryon asymmetry is dependent on α_0 and β_0 parameters and the scales M_* and T_D.

To compare Eq. (35) to the observed baryon asymmetry we plug in the suggested values for $M_* = M_P/\sqrt{8\pi}$, the reducep Planck mass and $T_D = M_I \sim 2\times 10^{16}\,\text{GeV}$ which corresponds to the tensor mode fluctuations in the inflationary scale.[53] We then obtain

$$\eta = \alpha_0\, 2.08\times 10^{-15} - \beta_0\, 2.16\times 10^{-19} \,, \tag{36}$$

where we can see the baryon asymmetry vanishes for $\alpha_0, \beta_0 \longrightarrow 0$. If $\alpha_0, \beta_0 \neq 0$ the baryon asymmetry is explained by the existence of a minimal length scale.

We compare the predicted baryon asymmetry from Eq. (36) to the observational bounds $5.7 \times 10^{-11} \lesssim \eta \lesssim 9.9 \times 10^{-11}$.[54] By doing this we can fix the GUP parameters to their actual values for different models:

- Only linear GUP ($\beta_0 = 0$): $2.74 \times 10^4 \lesssim \alpha_0 \lesssim 4.76 \times 10^4$
- Only quadratic GUP ($\alpha_0 = 0$): $-4.58 \times 10^8 \lesssim \beta_0 \lesssim -2.64 \times 10^8$
- Linear and Quadratic GUP ($\beta_0 = -\alpha_0^2$): $1.21 \times 10^4 \lesssim \alpha_0 \lesssim 1.71 \times 10^4$ and $-2.92 \times 10^8 \lesssim \beta_0 \lesssim -1.46 \times 10^8$
- Linear and Quadratic GUP: $\alpha_0 \gtrsim 4.81 \times 10^3$ and $\beta_0 \lesssim -1.48 \times 10^8$

We can see that the GUP parameters should have values $\alpha_0 \approx 10^4$ and $\beta_0 \approx -10^8$, which suggest a new length scale of $\ell_{new} = \alpha_0 \ell_P \sim \sqrt{-\beta_0}\ell_P \approx 10^{-31}$ m.

5. Conclusion

With this work we test the fundamental physics of nature. More specifically, how the minimal length can affect phenomena on large scales such as in this case, the baryon asymmetry in the Universe, which is still a problem in physics up to this day. The mechanism which we propose satisfies all Sakharov conditions and thus offers a viable explanation of the observed baryon asymmetry in the Universe. We used an interaction term which couples the baryon current and space-time to satisfy the first two Sakharov conditions, and using the holographic principle we modified the Bekenstein-Hawking entropy and the Friedmann equations to break thermal equilibrium and thus satisfying the third and last Sakharov condition. The modified Friedmann equations can also be used more broadly in cosmology and can provide a very rich phenomenology.

Our findings suggest that minimal length effects could have been the cause for generating the observed baryon symmetry. Our work does not constrain the GUP parameters to an improved upper bound because, but rather fixes the values of the GUP parameters α_0 and β_0 and thus sets the new length scale to be $\ell_{new} = \alpha_0 \ell_P \sim \sqrt{-\beta_0}\ell_P \approx 10^{-31}$ m.

We noticed that the quadratic GUP parameter turns out to be negative. This is consistent with other cosmological studies using quadratic GUP,[55] non-trivial structures of space-time[8,16,55–57] and a crystal lattice considerations with GUP.[56] This can suggest that space-time can have a granular structure at the Planck scale.[27,58–60]

References

1. A. Zeilinger, Rev. Mod. Phys. **71**, S288-S297 (1999).
2. C. M. Will, Living Rev. Rel. **17**, 4 (2014) [arXiv:1403.7377 [gr-qc]].
3. J. Magueijo, Phys. Rev. D **73**, 124020 (2006) [arXiv:gr-qc/0603073 [gr-qc]].
4. A. Hamma and F. Markopoulou, New J. Phys. **13**, 095006 (2011) [arXiv:1011.5754 [gr-qc]].

5. M. M. Dos Santos, T. Oniga, A. S. Mcleman, M. Caldwell and C. H. T. Wang, J. Plasma Phys. **79**, 437 (2013) [arXiv:1301.0494 [quant-ph]].
6. A. Feller and E. R. Livine, Class. Quant. Grav. **33**, no. 6, 065005 (2016) [arXiv:1509.05297 [gr-qc]].
7. I. Danshita, M. Hanada and M. Tezuka, PTEP **2017**, no. 8, 083I01 (2017) [arXiv:1606.02454 [cond-mat.quant-gas]].
8. L. Buoninfante, G. G. Luciano and L. Petruzziello, Eur. Phys. J. C **79**, no. 8, 663 (2019) [arXiv:1903.01382 [gr-qc]].
9. M. Blasone, G. Lambiase, G. G. Luciano, L. Petruzziello and F. Scardigli, Int. J. Mod. Phys. D **29**, no. 2, 2050011 (2020) [arXiv:1912.00241 [hep-th]].
10. T. W. van de Kamp, R. J. Marshman, S. Bose and A. Mazumdar, Phys. Rev. A **102**, no. 6, 062807 (2020) [arXiv:2006.06931 [quant-ph]].
11. S. A. Haine, New J. Phys. **23**, no. 3, 033020 (2021) [arXiv:1810.10202 [quant-ph]].
12. M. Maggiore, Phys. Lett. B **304**, 65-69 (1993) [arXiv:hep-th/9301067 [hep-th]].
13. M. Maggiore, Phys. Lett. B **319**, 83-86 (1993) [arXiv:hep-th/9309034 [hep-th]].
14. M. Maggiore, Phys. Rev. D **49**, 5182-5187 (1994) [arXiv:hep-th/9305163 [hep-th]].
15. F. Scardigli, Phys. Lett. B **452**, 39-44 (1999) [arXiv:hep-th/9904025 [hep-th]].
16. A. Kempf, G. Mangano and R. B. Mann, Phys. Rev. D **52**, 1108-1118 (1995) [arXiv:hep-th/9412167 [hep-th]].
17. F. Scardigli and R. Casadio, Class. Quant. Grav. **20**, 3915-3926 (2003) [arXiv:hep-th/0307174 [hep-th]].
18. S. Das and E. C. Vagenas, Can. J. Phys. **87**, 233-240 (2009) [arXiv:0901.1768 [hep-th]].
19. A. F. Ali, S. Das and E. C. Vagenas, "The Generalized Uncertainty Principle and Quantum Gravity Phenomenology" [arXiv:1001.2642 [hep-th]].
20. S. Basilakos, S. Das and E. C. Vagenas, JCAP **09**, 027 (2010) [arXiv:1009.0365 [hep-th]].
21. A. F. Ali, S. Das and E. C. Vagenas, Phys. Rev. D **84**, 044013 (2011) [arXiv:1107.3164 [hep-th]].
22. I. Pikovski, M. R. Vanner, M. Aspelmeyer, M. S. Kim and C. Brukner, Nature Phys. **8**, 393-397 (2012) [arXiv:1111.1979 [quant-ph]].
23. F. Scardigli and R. Casadio, Eur. Phys. J. C **75**, no. 9, 425 (2015) [arXiv:1407.0113 [hep-th]].
24. F. Scardigli, G. Lambiase and E. Vagenas, Phys. Lett. B **767**, 242-246 (2017) [arXiv:1611.01469 [hep-th]].
25. S. P. Kumar and M. B. Plenio, Phys. Rev. A **97**, no. 6, 063855 (2018) [arXiv:1708.05659 [quant-ph]].
26. D. Amati, M. Ciafaloni and G. Veneziano, Phys. Lett. B **197**, 81 (1987); D. J. Gross and P. F. Mende, Phys. Lett. B **197**, 129-134 (1987); D. Amati, M. Ciafaloni and G. Veneziano, Phys. Lett. B **216**, 41-47 (1989); K. Konishi, G. Paffuti and P. Provero, Phys. Lett. B **234**, 276-284 (1990); S. Capozziello, G. Lambiase and G. Scarpetta, Int. J. Theor. Phys. **39**, 15-22 (2000) [arXiv:gr-qc/9910017 [gr-qc]].
27. A. Das, S. Das and E. C. Vagenas, Phys. Lett. B **809**, 135772 (2020) [arXiv:2006.05781 [gr-qc]].
28. G. Lambiase and F. Scardigli, Phys. Rev. D **97**, no. 7, 075003 (2018) [arXiv:1709.00637 [hep-th]].
29. A. Iorio, G. Lambiase, P. Pais and F. Scardigli, Phys. Rev. D **101**, no. 10, 105002 (2020) [arXiv:1910.09019 [hep-th]].
30. L. Petruzziello, Class. Quant. Grav. **38**, no. 13, 135005 (2021) [arXiv:2010.05896 [hep-th]]; G. G. Luciano and L. Petruzziello, Eur. Phys. J. Plus **136**, no. 2, 179 (2021).

31. L. Canetti, M. Drewes and M. Shaposhnikov, New J. Phys. **14**, 095012 (2012) [arXiv:1204.4186 [hep-ph]].
32. A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32-35 (1967).
33. R. G. Cai, L. M. Cao and Y. P. Hu, JHEP **08**, 090 (2008) [arXiv:0807.1232 [hep-th]].
34. T. Zhu, J. R. Ren and M. F. Li, Phys. Lett. B **674**, 204-209 (2009) doi:10.1016/j.physletb.2009.03.020 [arXiv:0811.0212 [hep-th]].
35. A. Awad and A. F. Ali, JHEP **06**, 093 (2014) [arXiv:1404.7825 [gr-qc]].
36. S. Giardino and V. Salzano, Eur. Phys. J. C **81**, no. 2, 110 (2021) [arXiv:2006.01580 [gr-qc]].
37. G. 't Hooft, Conf. Proc. C **930308**, 284-296 (1993) [arXiv:gr-qc/9310026 [gr-qc]].
38. L. Susskind, J. Math. Phys. **36**, 6377-6396 (1995) [arXiv:hep-th/9409089 [hep-th]].
39. P. Bosso and S. Das, Annals Phys. **383**, 416-438 (2017) [arXiv:1607.01083 [gr-qc]].
40. A. J. M. Medved and E. C. Vagenas, Phys. Rev. D **70**, 124021 (2004) [arXiv:hep-th/0411022 [hep-th]].
41. G. Amelino-Camelia, M. Arzano and A. Procaccini, Phys. Rev. D **70**, 107501 (2004) [arXiv:gr-qc/0405084 [gr-qc]].
42. L. J. Garay, Int. J. Mod. Phys. A **10**, 145-166 (1995) [arXiv:gr-qc/9403008 [gr-qc]].
43. D. Christodoulou, Phys. Rev. Lett. **25**, 1596-1597 (1970).
44. D. Christodoulou and R. Ruffini, Phys. Rev. D **4**, 3552-3555 (1971).
45. J. D. Bekenstein, Phys. Rev. D **7**, 2333-2346 (1973).
46. S. W. Hawking, Phys. Rev. D **13**, 191-197 (1976).
47. R. G. Cai and S. P. Kim, JHEP **02**, 050 (2005) [arXiv:hep-th/0501055 [hep-th]].
48. R. G. Cai, L. M. Cao and Y. P. Hu, Class. Quant. Grav. **26**, 155018 (2009) [arXiv:0809.1554 [hep-th]].
49. S. A. Hayward, Class. Quant. Grav. **15**, 3147-3162 (1998) [arXiv:gr-qc/9710089 [gr-qc]].
50. H. Davoudiasl, R. Kitano, G. D. Kribs, H. Murayama and P. J. Steinhardt, Phys. Rev. Lett. **93**, 201301 (2004) [arXiv:hep-ph/0403019 [hep-ph]].
51. N. Azhar, A. Jawad and S. Rani, Phys. Dark Univ. **32**, 100815 (2021); H. Davoudiasl, Phys. Rev. D **88**, 095004 (2013) [arXiv:1308.3473 [hep-ph]]; S. D. Odintsov and V. K. Oikonomou, EPL **116**, no. 4, 49001 (2016) [arXiv:1610.02533 [gr-qc]]; S. D. Odintsov and V. K. Oikonomou, Phys. Lett. B **760**, 259-262 (2016) [arXiv:1607.00545 [gr-qc]]; V. K. Oikonomou and E. N. Saridakis, Phys. Rev. D **94**, no. 12, 124005 (2016) [arXiv:1607.08561 [gr-qc]]; S. Bhattacharjee and P. K. Sahoo, Eur. Phys. J. C **80**, no. 3, 289 (2020) [arXiv:2002.11483 [physics.gen-ph]]; E. H. Baffou, M. J. S. Houndjo, D. A. Kanfon and I. G. Salako, Eur. Phys. J. C **79**, no. 2, 112 (2019) [arXiv:1808.01917 [gr-qc]]; M. P. L. P. Ramos and J. Páramos, Phys. Rev. D **96**, no. 10, 104024 (2017) [arXiv:1709.04442 [gr-qc]]; M. Fukushima, S. Mizuno and K. i. Maeda, Phys. Rev. D **93**, no. 10, 103513 (2016) [arXiv:1603.02403 [hep-ph]]; H. M. Sadjadi, Phys. Rev. D **76**, 123507 (2007) [arXiv:0709.0697 [gr-qc]].
52. E. W. Kolb and M. S. Turner, "The Early Universe," Front. Phys. **69**, 1-547 (1990).
53. W. H. Kinney, E. W. Kolb, A. Melchiorri and A. Riotto, Phys. Rev. D **74**, 023502 (2006) [arXiv:astro-ph/0605338 [astro-ph]].
54. J. M. Cline, "Baryogenesis," [arXiv:hep-ph/0609145 [hep-ph]]; A. D. Dolgov, "CP violation in cosmology," [arXiv:hep-ph/0511213 [hep-ph]]; A. Riotto, "Theories of baryogenesis," [arXiv:hep-ph/9807454 [hep-ph]]; A. Riotto and M. Trodden, Ann. Rev. Nucl. Part. Sci. **49**, 35-75 (1999) [arXiv:hep-ph/9901362 [hep-ph]]; M. Yoshimura, J. Korean Phys. Soc. **29**, S236 (1996) [arXiv:hep-ph/9605246 [hep-ph]]; G. Lambiase, S. Mohanty and A. R. Prasanna, Int. J. Mod. Phys. D **22**, 1330030 (2013) [arXiv:1310.8459 [hep-ph]].

55. V. Nenmeli, S. Shankaranarayanan, V. Todorinov and S. Das, Phys. Lett. B **821**, 136621 (2021) [arXiv:2106.04141 [gr-qc]].
56. P. Jizba, H. Kleinert and F. Scardigli, Phys. Rev. D **81**, 084030 (2010) [arXiv:0912.2253 [hep-th]].
57. T. Kanazawa, G. Lambiase, G. Vilasi and A. Yoshioka, Eur. Phys. J. C **79**, no. 2, 95 (2019).
58. A. F. Ali, S. Das and E. C. Vagenas, Phys. Lett. B **678**, 497-499 (2009) [arXiv:0906.5396 [hep-th]].
59. S. Das, E. C. Vagenas and A. F. Ali, Phys. Lett. B **690**, 407-412 (2010) [erratum: Phys. Lett. B **692**, 342-342 (2010)] [arXiv:1005.3368 [hep-th]].
60. S. Deb, S. Das and E. C. Vagenas, Phys. Lett. B **755**, 17-23 (2016) [arXiv:1601.07893 [gr-qc]].

On quantum gravity and quantum gravity phenomenology

Douglas Edmonds
Department of Physics, Penn State Hazleton, Hazleton, PA, 18202, U.S.A.
E-mail: bde12@psu.edu

Djordje Minic* and Tatsu Takeuchi
Department of Physics, Virginia Tech, Blacksburg, VA 24061, U.S.A.
* *Talk given by D.M. at the 16th Marcel Grossmann Meeting*
E-mail: dminic@vt.edu, takeuchi@vt.edu

This article summarizes a new approach to quantum gravity based on the concepts of modular spacetime, Born geometry, and metastring theory and their applications to quantum gravity phenomenology. In particular, we discuss a new understanding of dark matter in terms of metaparticles (zero modes of the metastring) and its relation to dark energy (the curvature of dual spacetime) in view of the actual astronomical observations.

Keywords: Quantum Gravity, Quantum Spacetime, Dark Matter, Dark Energy.

1. Introduction and summary

In this paper we review recent work[1–10] on quantum foundations of quantum mechanics (QM), quantum field theory (QFT), and quantum gravity (in the guise of metastring theory) as well as unique implications for the problems of dark matter and dark energy.[11,12] This generic formulation of quantum gravity implies a radiatively stable positive cosmological constant (dark energy)[11] in the observed classical spacetime, and metaparticle quanta (the zero modes of the metastring) representing the natural quanta of dark matter[12] (correlated to dark energy and visible matter). The logic of our story is very similar to the path that leads from the Minkowski geometry of special relativity via relativistic non-gravitational field theory to a dynamical spacetime of general relativity. In this paper, we start with a hidden geometry in quantum theory (Born geometry) and proceed to its dynamical implementation in quantum gravity (formulated as a metastring theory) with implications for QFT (formulated in a way that takes into account the hidden Born geometry) with implications for the observed world: metaparticles as dark matter quanta, and dark energy emerging from the geometry of the dual spacetime. In some sense, this is a sharpening of the modern approaches to non-perturbative quantum physics,[13] using a simple but crucial insight about a completeness of quantum kinematics of discretized physical systems.[14]

In particular, in Ref. 5 we demonstrated that any quantum theory is endowed with a generic quantum polarization associated with modular spacetime.[4] The generic polarization manifestly realizes quantum non-locality, associated with the

quantum superposition principle, that is consistent with causality, and reveals a novel geometry structure, called Born geometry,[1,2] which unifies symplectic (ω), orthogonal (η), and conformal geometries (H). Born geometry is fundamental for a particular quantum theory that consistently propagates in this geometry – this turns out to be string theory formulated in a generalized-geometric and intrinsically non-commutative, doubled, chiral phase-space-like form (called metastring theory).[3,6–8] The zero modes of the metastring correspond to metaparticles that explicitly realize the geometry of modular spacetime,[9] and as such, could be considered as an explicit prediction of the modular representation of quantum theory. The metaparticles[10] are quanta of a modular generalization of quantum fields, the low energy remnants of metastring fields. From this new viewpoint[1–10] quantum gravity is essentially defined as "gravitization of the quantum," that is, as a theory of a dynamical Born geometry. As such it incorporates the concept of Born reciprocity[15] as a covariant implementation of T-duality, the fundamental relation between short and long distance physics in string theory, as well as the new idea of relative (or observer dependent) locality.[16]

In what follows, we start with a discussion of quantum theory via quantum (modular) spacetime and then comment on QFT in this approach, and then move on to quantum gravity and its phenomenology in the context of dark matter and dark energy and the actual astronomical observations.

2. Quantum theory and quantum spacetime

The fundamental reason for the existence of modular polarizations in quantum theory is as follows.[5] If one imagines that a quantum system is formulated on a lattice, as assumed in the modern (Wilsonian) non-perturbative approaches,[13] then a theorem due to Zak[14] states that a complete set of quantum numbers needed to describe any quantum system would require both quantum numbers associated with the lattice and its inverse. This is easy to see by realizing that non-commuting Hermitian operators, such as coordinates and momenta $[\,\hat{q},\,\hat{p}\,] = i\hbar$, when exponentiated, together with the appropriate lattice spacing a and its inverse $2\pi\hbar/a$, commute, that is,

$$\left[\exp\left(\frac{i}{\hbar}\hat{q}\frac{2\pi\hbar}{a}\right),\exp\left(\frac{i}{\hbar}\hat{p}a\right)\right] \,=\, 0\,. \tag{1}$$

Such unitary observables were labeled as "modular" by Aharonov.[17] These variables are purely quantum in the sense that their formal $\hbar \to 0$ limit is singular. Also, even though their commutators are zero, the associated Poisson brackets are non-zero, as these are unitary (phase) variables. Finally, the classical limit is defined by starting with a modular formulation and defining an appropriate "extensification,"[5] with, in principle, many classical limits. These purely quantum variables also appear in the context of QFT.

Thus we concentrate on a complete set of unitary operators as opposed to a complete set of Hermitian operators. Let us examine the simplest example of

the $\hat{q}$ and $\hat{p}$ operators. The commuting subalgebra of the original non-commuting $[\,\hat{q},\,\hat{p}\,] = i\hbar$ algebra can be completely described by self-dual lattices (endowed with the natural symplectic form (ω) coming from the commutator bracket). These in turn represent a discretization of a (covariant) phase space defined by q and p, and when lifted to the original non-commutative algebra, require extra data associated with the lift that is described by a doubly orthogonal ($O(d,d)$, where d denotes the spacetime dimension) metric η (a symmetric counterpart of the antisymmetric ω associated with $Sp(2d)$ transformations). Finally, in order to define the vacuum state on this self-dual lattice, we need a conformal structure $O(2, 2(d-1))$.[5] This triplet of structures defines Born geometry[1,2] associated with the modular representation of quantum theory.[5] Born geometry captures quantum non-locality that is consistent with causality, given the quantum nature of the unitary operators and the fact that the triple intersection of $Sp(2d)$, $O(d,d)$ and $O(2,2(d-1))$ gives the Lorentz group.[6]

Let us formalize these insights about the hidden quantum spacetime geometry of quantization.[5] We start with the Heisenberg (or Weyl-Heisenberg) group, which is generated, on the level of the corresponding algebra, by the familiar position $\hat{q}^a$ and momentum $\hat{p}_b$ operators: $[\,\hat{q}^a,\,\hat{p}_b\,] = i\hbar\delta^a_b$. It will be convenient to introduce a length scale λ and a momentum scale ϵ, with $\lambda\epsilon = \hbar$. Also, let us introduce the notation $\hat{x}^a \equiv \hat{q}^a/\lambda$, $\hat{\tilde{x}}_a \equiv \hat{p}_a/\epsilon$ so that $[\,\hat{x}^a,\,\hat{\tilde{x}}_b\,] = i\delta^a_b$, and let us define

$$\mathbb{X}^A \equiv (x^a, \tilde{x}_a)^T\,, \quad [\,\hat{\mathbb{X}}^a,\,\hat{\mathbb{X}}^b\,] = i\omega^{AB}\,, \quad \text{with} \quad \frac{1}{2}\omega_{AB}dX^A dX^B = \frac{1}{\hbar}dp_a \wedge dq^a, \quad (2)$$

where $\omega_{AB} = -\omega_{BA}$ is the canonical symplectic form on phase space $\mathcal{P}$. The Heisenberg group $H_{\mathcal{P}}$ is generated by Weyl operators[18] $W_{\mathbb{K}} \equiv e^{2\pi i\omega(\mathbb{K},\mathbb{X})}$, where $\mathbb{K}$ stands for the pair $(\tilde{k}, k)$ and $\omega(\mathbb{K},\mathbb{K}') = k\cdot\tilde{k}' - \tilde{k}\cdot k'$. These form a central extension of the translation algebra, $W_{\mathbb{K}}W_{\mathbb{K}'} = e^{2\pi i\omega(\mathbb{K},\mathbb{K}')}W_{\mathbb{K}+\mathbb{K}'}$. The projection $\pi : H_{\mathcal{P}} \to \mathcal{P}$ (where $\pi : W_{\mathbb{K}} \to \mathbb{K}$) defines a line bundle over $\mathcal{P}$ (in principle a covariant phase space of quantum probes). In this formulation, states are sections of degree one, $W_{\mathbb{K}'}\Phi(\mathbb{K}) = e^{2\pi i\omega(\mathbb{K},\mathbb{K}')}\Phi(\mathbb{K}+\mathbb{K}')$.

Using notions of non-commutative algebra and non-commutative geometry[19] (such as the theorem of Gelfand-Naimark[20]), we can say that a Lagrangian submanifold (a half-dimensional submanifold of phase space upon which the symplectic form pulls back to zero) is a maximally commutative subgroup of the Heisenberg group. If we accept this notion of a Lagrangian, then the quantum regime is very different from the classical regime. In particular the vanishing Poisson bracket $\{f(q), g(p)\} = 0$ requires either f or g to be constant. However, the vanishing commutator $[f(\hat{q}), g(\hat{p})] = 0$ requires only that the functions be commensurately periodic :

$$e^{i\alpha\hat{p}}e^{i\beta\hat{q}} = e^{i\hbar\alpha\beta}e^{i\beta\hat{q}}e^{i\alpha\hat{p}}\,, \qquad \alpha\beta = 2\pi/\hbar\,. \quad (3)$$

Similar considerations led Aharonov to introduce *modular variables* to describe purely quantum phenomena such as interference.[21]

Modular variables are described in great detail in Ref. 21, where one can find detailed references on this subject. The modular variables, denoted $[\,\hat{q}\,]$ and $[\,\hat{p}\,]$, are defined modulo a length scale R (the slit spacing being a natural choice) as

$$[\,\hat{p}\,]_{2\pi\hbar/R} = \hat{p} \bmod (2\pi\hbar/R) \ , \qquad [\,\hat{q}\,]_R = \hat{q} \bmod (R) \ . \tag{4}$$

They play a central role in understanding interference in terms of operators (and not states). The shift operator $e^{iR\hat{p}/\hbar} = e^{iR[\,\hat{p}\,]/\hbar}$ shifts the position of a particle state (say an electron in the double-slit experiment) by distance R and is a function of the modular momenta (see also Ref. 22). These modular variables (the main examples being the Aharonov-Bohm and Aharonov-Casher phases[21]) satisfy non-local operator equations of motion. For example, given the Hamiltonian $\hat{H} = \hat{p}^2/2m + V(\hat{q})$, the Heisenberg equation of motion for the shift operator is

$$e^{-iR\hat{p}/\hbar}\,\frac{d}{dt}\,e^{iR\hat{p}/\hbar} \ = \ -\frac{iR}{\hbar}\left(\frac{V(\hat{q}+R)-V(\hat{q})}{R}\right). \tag{5}$$

Modular variables are fundamentally non-local in a non-classical sense, since we see here that their evolution depends on the value of the potential at distinct locations. Remarkably, thanks to the uncertainty principle, this dynamical non-locality does not lead to a violation of causality.[21] One of the characteristic features of these variables is that they do not have classical analogues; indeed, the limit $\hbar \to 0$ of $[\,p\,]_{h/R}$ is ill-defined. Also, modular variables capture entanglement of continuous q, p variables. When exponentiated (*i.e.*, when understood as Weyl operators) the modular variables naturally commute. In other words, given $[\,\hat{x}^a,\,\hat{\tilde{x}}_b\,] = i\delta^a_b$ we find[5] $[\,e^{2\pi i\hat{x}},\,e^{2\pi i\hat{\tilde{x}}}\,] = 0$. Thus, the quantum algebra of modular variables possesses more commutative directions than the classical Poisson algebra, since the Poisson bracket of modular variables does not vanish, $\{\,e^{2\pi ix},\,e^{2\pi i\tilde{x}}\,\} \neq 0$.

2.1. *Modular spacetime as quantum spacetime and Born geometry*

In this view of quantum theory we have a structure analogous to a Brillouin cell in condensed matter physics. The volume and shape of the cell are given by λ and ϵ (i.e. $\hbar$ and G (or α')). The uncertainty principle is implemented in a subtle way: we can specify a point in a modular cell, but if so, we cannot say *which* cell we are in. This means that there is a more general notion of quantization.[5] Instead of selecting a classical polarization L (the arguments of the wave function, or the arguments of a local quantum field) we choose a *modular polarization*. In terms of the Heisenberg group, all that is happening is that in order to have a commutative algebra, we need only $\omega(\mathbb{K},\mathbb{K}') \in 2\mathbb{Z}$, and $W_{\mathbb{K}}W_{\mathbb{K}'} = e^{2\pi i\omega(\mathbb{K},\mathbb{K}')}W_{\mathbb{K}+\mathbb{K}'} = W_{\mathbb{K}'}W_{\mathbb{K}'}$. This defines a lattice Λ in phase space $\mathcal{P}$. Finally, we specify a "lift" of the lattice from the phase space $\mathcal{P}$ to the Heisenberg group $H_{\mathcal{P}}$.

Maximally commuting subgroups $\hat{\Lambda}$ of the Heisenberg group correspond to lattices that are integral and self-dual with respect to ω.[23] Given W_λ, where $\lambda \in \Lambda$, there is a lift to $\hat{\Lambda}$ which defines "modular polarization," $U_\lambda = \alpha(\lambda)W_\lambda$, where $\alpha(\lambda)$

satisfies the co-cycle condition $\alpha(\lambda)\,\alpha(\mu)\,e^{\pi i\omega(\lambda,\mu)} = \alpha(\lambda+\mu)$, with $\lambda, \alpha \in \Lambda$. One can parametrize a solution to the co-cycle condition by introducing a symmetric bilinear from $\eta(\mathbb{K},\mathbb{K}') = k\cdot\tilde{k}' + \tilde{k}\cdot k'$ and setting $\alpha_\eta(\lambda) \equiv e^{i\frac{\pi}{2}\eta(\lambda,\lambda)}$. Finally, when we choose a classical Lagrangian L there exists a special translation-invariant state that we associate with the vacuum, which we interpret as "empty space." In modular quantization there is no such translation-invariant state because of the lattice structure. The best we can do is to choose a state that minimizes an "energy," which requires the introduction of another symmetric bilinear form that we call, again suggestively, H. This means that we are looking for operators such that

$$[\,\hat{\mathbb{P}}_A,\,\Phi\,] = \frac{i}{2\pi}\partial_A\Phi\ , \quad \text{and} \quad \Phi(\hat{\mathbb{X}}+\lambda) = \Phi(\hat{\mathbb{X}})\ , \tag{6}$$

where the modular observables $\Phi(\hat{\mathbb{X}}+\lambda) = \Phi(\mathbb{X})$ are generated by the lattice observables U_λ with $\lambda \in \Lambda$. Translation invariance would be the condition $\hat{\mathbb{P}}\,|0\rangle = 0$. Since this is not possible, the next natural choice is to minimize the translational energy. Therefore, we pick a positive definite metric H_{AB} on $\mathcal{P}$, and we define[5] $\hat{E}_H \equiv H^{AB}\hat{\mathbb{P}}_A\hat{\mathbb{P}}_B$, and demand that $|0\rangle_H$ be the ground state of $\hat{E}_H$. This is indeed the most natural choice and it shows that we cannot fully disentangle kinematics (i.e., the definition of translation generators) from dynamics. In the Schrödinger case, since the translation generators commute, the vacuum state $\hat{E}\,|0\rangle = 0$ is also the translation-invariant state and it carries no memory of the metric H needed to define the energy. In our context, due to the non-commutativity of translations, the operators $\hat{E}_H$ and $\hat{E}_{H'}$ do not commute. Thus, the vacuum state depends on H, in other words $|0\rangle_H \neq |0\rangle_{H'}$, and it also possesses a non-vanishing zero point energy.

Therefore, modular quantization involves the introduction of three quadratic forms (ω,η,H), called *Born geometry*,[1,2] which underlies the geometry of modular variables. As we will see, in the context of metastring theory, a choice of polarization is a choice of a spacetime within $\mathcal{P}$ but the most general choice is a *modular polarization* that we have discussed above. From the foundational quantum viewpoint, Born geometry (ω,η,H) arises as a parametrization of such quantizations, which results in a notion of quantum spacetime, that we call *modular spacetime*. In particular, a one-dimensional modular line is a two-dimensional torus that is compact and not simply connected. Finally, large spacetimes of canonical general relativity (and its extensions, like string theory) result as a "many-body" phenomenon through a process of tensoring (entangling) unit modular cells, which we refer to as "extensification."[5] Note that the Lorentz group (in d spacetime dimensions) lies at the intersection of the symplectic, neutral and doubly orthogonal groups,[5]

$$\mathrm{O}(1,d-1) \ = \ \mathrm{Sp}(2d)\cap\mathrm{O}(d,d)\cap\mathrm{O}(2,2(d-1))\ , \tag{7}$$

which sheds new light on the origin of quantum theory through compatibility of the causal (Lorentz) structure and non-locality captured by the discreteness of quantum spacetime. Note that relative (observer-dependent) locality[16] is needed to resolve

the apparent contradiction between discreteness of quantum spacetime and Lorentz symmetry.

One can pass from a classical polarization (such as the Schrödinger representation) to a modular polarization via the Zak transform.[14] Note that, there is a connection on the line bundle over phase space that has unit flux through a modular cell. (This is very similar to Integer Quantum Hall effect.) A modular wave function is quasi-periodic

$$\Psi(x+a,\tilde{x}) \;=\; e^{2i\pi a\tilde{x}}\Psi(x,\tilde{x})\;, \quad \Psi(x,\tilde{x}+\tilde{a}) \;=\; \Psi(x,\tilde{x})\;. \tag{8}$$

The quasi-periods correspond to the tails of an Aharonov-Bohm[24] potential attached to a unit flux. In particular, vacuum states must have at least one zero in a cell, which leads to theta functions (the Zak transforms of Gaussians).

2.2. *A comment on quantum field theory and quantum spacetime*

A few general comments about QFT in the modular form are in order, following the general modular formulation of any quantum theory. The modular polarization of QFT reveals new structures and sheds new light on both the short distance (UV) and long distance (IR) physics of quantum fields, and the continuum limit of QFTs which is self-dual with respect to the UV and IR properties (resembling some crucial properties of non-commutative field theories[25]). In particular, the modular representation of QFT introduces dual "electric" and "magnetic" variables, which are non-commuting in general. This extends our results in the context of the $2d$ conformal field theory formulation of string theory in which the non-commutativity of such "electric" and "magnetic" variables has been explicitly demonstrated.[7,8]

The general modular representation can be defined in terms of the Zak transform of a Schrödinger representation (*i.e.*, wave functions). Given a square normalizable wave function $\psi(x)$ belonging to a Hilbert space, one defines the modular representation as the lattice Fourier transform (or Zak transform)

$$\psi_a(x,\tilde{x}) \;\equiv\; \sqrt{a}\sum_n e^{-2\pi i n\tilde{x}}\psi(a(x+n))\;, \tag{9}$$

where $x \equiv q/a$, $\tilde{x} \equiv p/b$, with $ab = 2\pi\hbar$. Note that if $\psi(x)$ is a Gaussian, its Zak transform, the modular $\psi_a(x,\tilde{x})$, is given by the doubly-periodic theta function associated with the lattice. (The inverse Zak transform

$$\Phi(x+n) \;\equiv\; \frac{1}{\sqrt{a}}\int_0^1 d\tilde{x}\, e^{2\pi i n\tilde{x}}\Phi_a(a^{-1}x,\tilde{x})\;, \tag{10}$$

illustrates that the usual Schrödinger representation is truly singular, and thus not generic.)

Now if one second quantizes $\psi(x)$, one naturally ends up with a quantum field operator $\phi(x)$. Similarly, the second quantization of the modular $\psi_a(x,\tilde{x})$ would lead to a modular quantum field operator $\phi(x,\tilde{x})$, $\phi(x) \to \phi(x,\tilde{x})$. The excitations of such modular quantum fields are non-local in general, and will be discussed below

as *metaparticles*. Note that the usual wave functional approach to QFT defined in terms of functionals $\Psi\big[\phi(x)\big]$ should be now defined in terms of wave functionals of modular fields $\Psi\big[\phi(x,\tilde{x})\big]$. However, now we have more freedom in the general modular polarization. The dual momenta p and $\tilde{p}$ (to x and $\tilde{x}$ respectively) lead, via the canonical minimal-coupling prescription, not only to the usual fields ϕ but also to their duals $\tilde{\phi}$ (see below). This procedure defines the modular polarization of QFT in terms of the functional Zak transform of the original wave functional, $\Psi\big[\phi(x)\big] \rightarrow \Psi\big[\phi(x,\tilde{x}),\tilde{\phi}(x,\tilde{x})\big]$. For example, the Gaussian wave functionals with non-trivial kernels (such as the ones found in the context of non-trivial interacting theories like 2+1 and 3+1 dimensional Yang Mills theory[26]) would be mapped into functional theta functions.

The non-perturbative formulation is defined in a symmetric, self-dual way with respect to the double RG flows (as in non-commutative field theory[25]) with full spacetime covariance, and should be important not only for quantum non-locality of QFT, but also in the realms of strong coupling and deep infrared.

3. Quantum gravity and dynamical quantum (modular) spacetime

The unexpected outcome of this new view of the foundations of quantum theory is that this fundamental geometry of quantum theory can be realized in the context of metastring theory, in which this Born geometry (given by ω, η and H) is "gravitized" (i.e. dynamical). At the classical level, metastring theory[1–10] can be thought of as a formulation of string theory in which the target space is doubled in such a way that T-duality acts linearly on the coordinates. This doubling means that momentum and winding modes appear on an equal footing. In this formulation, T-duality exchanges the Lagrangian sub-manifold with its image under $J = \eta^{-1}H$. Classical metastring theory is defined by the action[3,27]

$$\hat{S} = \frac{1}{4\pi}\int_{\Sigma} d^2\sigma\Big(\partial_\tau \mathbb{X}^A(\eta_{AB}+\omega_{AB})(\mathbb{X})\partial_\sigma \mathbb{X}^B - \partial_\sigma \mathbb{X}^A H_{AB}(\mathbb{X})\partial_\sigma \mathbb{X}^B\Big)\,, \qquad (11)$$

where $\mathbb{X}^A$ are dimensionless coordinates on phase space and the fields η, H, ω are all dynamical (i.e., in general dependent on $\mathbb{X}$) phase space fields ($\mathbb{X}^A \equiv (x^a, \tilde{x}_a)^T$). In the context of a flat metastring we have constant η_{AB}, H_{AB}, and ω_{AB} :

$$\eta_{AB} \equiv \begin{pmatrix} 0 & \delta \\ \delta^T & 0 \end{pmatrix}, \quad H_{AB} \equiv \begin{pmatrix} h & 0 \\ 0 & h^{-1} \end{pmatrix}, \quad \omega_{AB} = \begin{pmatrix} 0 & \delta \\ -\delta^T & 0 \end{pmatrix}, \qquad (12)$$

where δ^μ_ν is the d-dimensional identity matrix, and $h_{\mu\nu}$ is the d-dimensional Lorentzian metric. The Polyakov string[28] is obtained in a singular limit of zero ω after integrating over $\tilde{x}$. (For a phase space structure of the canonical string see Ref. 29 and references therein.)

3.1. *Non-commutativity and non-associativity in quantum gravity*

The metastring formulation points to an unexpected fundamental non-commutativity of closed string theory, that we address in what follows. The string

commutation relations[7,8] state $[\mathbb{X}^A(\sigma),\mathbb{X}^B(\sigma')] = 2i\lambda^2\big[\pi\omega^{AB} - \eta^{AB}\theta(\sigma-\sigma')\big]$, where $\theta(\sigma)$ is the staircase distribution, i.e., a solution of $\theta'(\sigma) = 2\pi\delta(\sigma)$; it is odd and quasi-periodic with period 2π. Following standard practice, all indices are raised and lowered using η and η^{-1}. The momentum density operator is given by

$$\mathbb{P}_A(\sigma) \;=\; \frac{1}{2\pi\alpha'}\,\eta_{AB}\,\partial_\sigma\mathbb{X}^B(\sigma)\,, \tag{13}$$

and the previous commutation relation implies that it is conjugate to $\mathbb{X}^A(\sigma)$. The two-form ω appears when one integrates this canonical commutation relation to include the zero-modes, the integration constant being uniquely determined by worldsheet causality. Denoting by $(\,\hat{\mathbb{X}},\hat{\mathbb{P}}\,)$ the zero mode components of the string operators $\mathbb{X}(\sigma)$ and $\mathbb{P}(\sigma)$, we have $\big[\,\hat{\mathbb{P}}_A,\hat{\mathbb{P}}_B\,\big]=0$, $\big[\,\hat{\mathbb{X}}^A,\hat{\mathbb{P}}_B\,\big] = i\hbar\delta^A{}_B$, and $\big[\,\hat{\mathbb{X}}^A,\hat{\mathbb{X}}^B\,\big] = 2\pi i\lambda^2\omega^{AB}$. This is a deformation of the doubled Heisenberg algebra involving the string length λ as a deformation parameter. Note that under a constant B-field transformation $\mathbb{X}=(x^a,\tilde{x}_a)\mapsto(x^a,\tilde{x}_a+B_{ab}x^b)$, the trivial symplectic form $\omega(\mathbb{K},\mathbb{K}')=k\cdot\tilde{k}'-\tilde{k}\cdot k'$ is mapped onto $\omega(\mathbb{K},\mathbb{K}')=k_a\tilde{k}'^a-k'_a\tilde{k}^a-2B_{ab}\tilde{k}^a\tilde{k}'^b$, and the commutators read

$$\big[\,\hat{x}^a,\hat{x}^b\,\big]=0\,,\quad \big[\,\hat{x}^a,\hat{\tilde{x}}_b\,\big]=2\pi i\lambda^2\delta^a{}_b\,,\quad \big[\,\hat{\tilde{x}}_a,\hat{\tilde{x}}_b\,\big]=-4\pi i\lambda^2B_{ab}\,. \tag{14}$$

We see that the effect of the B-field is to render the dual coordinates non-commutative (and that the B-field originates from the symplectic structure ω). This implies a new view of the axion in four spacetime dimensions. The β-transformation on the other hand corresponds to the map $(x^a,\tilde{x}_a)\mapsto(x^a+\beta^{ab}\tilde{x}_b,\tilde{x}_a)$. Equivalently, it has the effect of mapping the symplectic structure to $\omega(\mathbb{K},\mathbb{K}') = k_a\tilde{k}'^a-k'_a\tilde{k}^a+2\beta^{ab}k_ak'_b$, and yields commutation relations

$$\big[\,\hat{x}^a,\hat{x}^b\,\big]=4\pi i\lambda^2\beta^{ab}\,,\qquad \big[\,\hat{x}^a,\hat{\tilde{x}}_b\,\big]=2\pi i\lambda^2\delta^a{}_b\,,\qquad \big[\,\hat{\tilde{x}}_a,\hat{\tilde{x}}_b\,\big]=0\,. \tag{15}$$

Dramatically, the coordinates that are usually thought of as the spacetime coordinates have become themselves non-commutative. Since this is the result of an $O(d,d)$ transformation it can be thought of in similar terms as the B-field; these are related by T-duality. We are familiar with the B-field background because we have, in the non-compact case, a fixed notion of locality in the target-space theory. However, in the non-geometric β-field background, we do not have such a notion of locality but we can access it through T-duality. However, this background does lead to non-commutative field theory, and one can place a bound on the minimal length λ of $O(10\,\mathrm{TeV})$, which is the current high-energy limit for probing the continuum structure of spacetime. (Similarly, this background can be used to argue for an effective minimal-length extension of commutation relations.[30] For astrophysical probes of the minimal length, see Ref. 31.)

Note that the dilaton can be understood as coming from the volume of phase space.[32] In general, for varying B-backgrounds we encounter non-associativity as well,[33] and the proper closure of such non-commutative and non-associative structure is ensured by the equations of motion. Here we remark that fundamental non-associativity can be related to the robustness of the Standard Model (SM) gauge

group.[34] Similarly, fundamental non-commutativity can be related to the underlying non-commutative nature of the SM and its phenomenology.[19,35]

3.2. *Non-perturbative formulation of quantum gravity*

The metastring offers a new view on the fundamental question of a non-perturbative formulation of quantum gravity.[1–10] The worldsheet can be made modular in our formulation, with the doubling of τ and σ, so that $\mathbb{X}(\tau,\sigma)$ can be in general viewed as an infinite-dimensional matrix (the matrix indices coming from the Fourier components of the doubles of τ and σ). Then the corresponding metastring action reads as

$$\int \mathrm{Tr}\Big[\,\partial_\tau \mathbb{X}^A \partial_\sigma \mathbb{X}^B(\omega_{AB}+\eta_{AB}) - \partial_\sigma \mathbb{X}^A H_{AB}\partial_\sigma \mathbb{X}^B\,\Big] d\tau d\sigma\ , \tag{16}$$

where the trace is over the matrix indices. One can associate partonic degrees of freedom with matrix entries. A non-perturbative quantum gravity follows by replacing the σ-derivative with a commutator involving one extra $\mathbb{X}^{26}$:

$$\partial_\sigma \mathbb{X}^A \to \big[\,\mathbb{X}^{26},\,\mathbb{X}^A\big]\ , \tag{17}$$

with $A = 0,1,2,\ldots,25$. The resulting matrix-model form of the above metastring action is

$$\int \mathrm{Tr}\Big(\partial_\tau \mathbb{X}^a\big[\mathbb{X}^b,\mathbb{X}^c\big]\eta_{abc} - H_{ac}\big[\mathbb{X}^a,\mathbb{X}^b\big]\big[\mathbb{X}^c,\mathbb{X}^d\big]H_{bd}\Big) d\tau\ , \tag{18}$$

with $a,b,c = 0,1,2,\ldots,25,26$, where the first term is of a Chern-Simons form and the second of the Yang-Mills form. η_{abc} contains both ω_{AB} and η_{AB}. This defines a non-perturbative quantum gravity viewed as "a gravitization of the quantum."[10]

This formulations invokes the IIB matrix model,[36] which describes N D-instantons (and is by T-duality related to the Matrix model of M-theory[37]). Given our new viewpoint we suggest a new covariant non-commutative matrix-model formulation of string theory as a theory of quantum gravity, by writing in the large N limit $\partial_\sigma \mathbb{X}^C = [\mathbb{X},\mathbb{X}^C]$ (and similarly for $\partial_\tau \mathbb{X}^B$) in terms of commutators of two (one for $\partial_\sigma \mathbb{X}^C$ and one for $\partial_\tau \mathbb{X}^C$) extra $N\times N$ matrix-valued chiral $\mathbb{X}$'s. Notice that, in general, we do not need an overall trace, and so the action can be viewed as a matrix, rendering the entire non-perturbative formulation as purely quantum in the sense of the original matrix formulation of QM:

$$\mathbb{S}_{\mathrm{ncF}} = \frac{1}{4\pi}\big[\mathbb{X}^a,\mathbb{X}^b\big]\big[\mathbb{X}^c,\mathbb{X}^d\big]f_{abcd}\ , \tag{19}$$

where instead of 26 bosonic $\mathbb{X}$ matrices we have 28, with supersymmetry emerging in 10(+2) dimensions from this underlying bosonic formulation. (This is a non-commutative matrix-model formulation of F-theory.) By T-duality, the new covariant M-theory matrix model is

$$\mathbb{S}_{\mathrm{ncM}} = \frac{1}{4\pi}\int_\tau \Big(\partial_\tau \mathbb{X}^i\big[\mathbb{X}^j,\mathbb{X}^k\big]g_{ijk} - \big[\mathbb{X}^i,\mathbb{X}^j\big]\big[\mathbb{X}^k,\mathbb{X}^l\big]h_{ijkl}\Big), \tag{20}$$

with 27 bosonic $\mathbb{X}$ matrices, with supersymmetry emerging in 11 dimensions. In this approach, holography[38] (such as AdS/CFT[39] or dS/CFT[40]), which can be viewed as a "quantum Jarzynski equality on the space of geometrized RG flows,"[41] is emergent in a particular "extensification" of quantum spacetime. The relevant information about ω_{AB}, η_{AB}, and H_{AB} is now contained in the new dynamical backgrounds f_{abcd} in F-theory, and g_{ijk} and h_{ijkl} in M-theory. This offers a new formulation of covariant Matrix theory in the M-theory limit,[42] which is essentially a partonic formulation; strings emerge from partonic constituents in a certain limit. This new matrix formulation is fundamentally bosonic (*i.e*, supersymmetry is emergent only in a specific limit) and thus it is reminiscent of bosonic M-theory.[43] The relevant backgrounds g_{ijk} and h_{ijkl} should be determined by the matrix RG equations. Also, there are lessons here for the new concept of "gravitization of quantum theory" as well as the idea that dynamical Hilbert spaces, or 2-Hilbert spaces (here represented by matrices), are fundamentally needed in quantum gravity.[44]

This matrix-like formulation should be understood as a general non-perturbative formulation of string theory. In this partonic formulation closed strings are collective excitations, in turn constructed from the product of open string fields. The observed classical spacetime emerges as an "extensification"[1–10] in a particular limit, out of the basic building blocks of quantum spacetime. Their remnants can be found in the low energy bi-local quantum fields, with bi-local (metaparticle) quanta, to which we now turn.

4. Quantum gravity, metastrings, metaparticles and dark matter

The above manifestly T-duality covariant formulation of closed strings (i.e. metastring theory) implies intrinsic non-commutativity of zero-modes. It is thus instructive to formulate a particle-like limit of the metastring that we call the *metaparticle*.[7–9] The theory of metaparticles (the low energy remnants of the metastring, and as such, the low energy remnants (predictions) of quantum gravity) is defined by the following world-line action[7–9]

$$S \equiv \int_0^1 d\tau \left[p\dot{x} + \tilde{p}\dot{\tilde{x}} + \lambda^2 p\dot{\tilde{p}} - \frac{N}{2}\left(p^2 + \tilde{p}^2 + m^2\right) + \tilde{N}\left(p\tilde{p} - \mu\right) \right] . \qquad (21)$$

Here the signature $(+,-,\cdots,-)$ and the contraction of indices are implicitly assumed. At the classical level, theory of metaparticles is a worldline theory with the usual reparameterization invariance and two additional features.[9] The first new feature is the presence of an additional local symmetry, which from the string point of view corresponds to the completion of worldsheet diffeomorphism invariance. From the particle worldline point of view, this symmetry is associated with an additional local constraint. The second new feature is the presence of a non-trivial symplectic form on the metaparticle phase space, also motivated by string theory.[7,8] Because of its interpretation as a particle model on Born geometry, associated with the modular representation of quantum theory, the spacetime on which the metaparticle propagates is ambiguous, with different choices related by T-duality. The attractive

feature of this model include worldline causality and unitarity, as well as an explicit mixing of widely separated energy-momentum scales. The metaparticle propagator follows from the world line path integral defined by the above action and it has the following form in momentum space[9]:

$$G(p,\tilde{p};p_i,\tilde{p}_i) \;\sim\; \delta^{(d)}(p-p_i)\,\delta^{(d)}(\tilde{p}-\tilde{p}_i)\,\frac{\delta(p\cdot\tilde{p}-\mu)}{p^2+\tilde{p}^2+m^2-i\varepsilon}\;. \tag{22}$$

The canonical particle propagator is a highly singular $\tilde{p}\to 0$ (and $\mu\to 0$) limit of this expression. This propagator also predicts the following dispersion relation (in a particular gauge[9]) that can be tested in various experiments and with various probes: $E_p^2+(\mu^2/E_p^2)=\vec{p}^{\,2}+m^2$. This formulation is fully compatible with Lorentz covariance, and is a direct consequence of the consistency of quantum theory and a minimal length (and thus Born geometry). In a cosmological context, one can use this dispersion relation to put a bound on μ for the case of neutrinos, which turns out to be close to the energy scale of dark energy.[45] In general, for each particle at energy E there exists a dual particle at energy μ/E. This is in complete analogy of the well-known prediction of antiparticles in the union of special relativity and quantum theory.

Note that the usual particle limit is obtained, at least classically, by taking $\mu\to 0$ and $\tilde{p}\to 0$. Given the form of the above Lagrangian, the metaparticle looks like two particles that are entangled through a Berry-phase-like $p_\mu\,\dot{\tilde{p}}^\mu$ factor. The metaparticle is fundamentally non-local, and thus it should not be associated with effective local field theory. In particular, by looking at the metaparticle constraints $p^2+\tilde{p}^2=m^2$ and $p\tilde{p}=\mu$, we note that the momenta p and $\tilde{p}$ can be, in principle, widely separated. For example, if m is of the order of the Planck energy, and μ of the order of TeV^2 (which can be considered a characteristic particle physics scale), then the p can be of the order of the Planck energy, and the $\tilde{p}$ of the vacuum energy scale. Thus metaparticle theory can naturally relate widely separated scales, which transcends the usual reasoning based on Wilsonian effective field theory (and should be relevant for the naturalness and hierarchy problems).

We can also discuss the background fields that couple to the metaparticle quanta. Following the well-known procedure of introducing background fields in the case of particles, by shifting the canonical momentum by a gauge field we may try to extend the gauging procedure to the metaparticle counterpart. There is a possible ambiguity in this gauging which depends on which configuration variables one decides to work with. If one takes $(x,\tilde{x})$ as configuration variables, one obtains a gauging which could also be motivated by the presence of a "stringy gauge field" in metastring theory,[3] $p_\mu\to p_\mu+A_\mu(x,\tilde{x})$ and $\tilde{p}^\mu\to\tilde{p}^\mu+\tilde{A}^\mu(x,\tilde{x})$. The generic prediction here is the existence of a dual field $\tilde{A}$, which is correlated (entangled), with the original A field. Thus the entire SM would have a dual SM (which we propose, describes the dark matter sector).

We expect that the correct field theoretic description of the metaparticle is in terms of the above general non-commutative (modular) field theory $\Phi(x,\tilde{x})$ limit of

the metastring.[1–10] Such an effective non-commutative field theory is similar in spirit to Ref. 25. Also, we note that the concept of metaparticles might be argued from the compatibility of the quantum spacetime that underlies the generic representations of quantum theory, as discussed in Ref. 5, and thus the metaparticle might be as ubiquitous as the concept of antiparticles which is demanded by the compatibility of relativity and quantum theory. The metaparticles also provide a natural route to the problem of dark matter. To lowest (zeroth) order of the expansion in the non-commutative parameter λ, the effective action for the SM matter Lagrangian (L_m) and their duals (that could be interpreted as the dark matter Lagrangian L_{dm}) takes the form

$$S_{\text{eff}} = -\iint \sqrt{g(x)\tilde{g}(\tilde{x})}\big[L_m(A(x,\tilde{x})) + \tilde{L}_{dm}(\tilde{A}(x,\tilde{x})) + \cdots\big] \,, \tag{23}$$

where we have included the non-dynamical gravitational background. Note that after integrating over the "hidden variable" parameters $\tilde{x}$, we obtain an effective theory of visible and dual (dark) matter in the observed spacetime x :

$$S_{\text{eff}} = -\int \sqrt{g(x)}\,\Big[L_m(A(x)) + \tilde{L}_{dm}(\tilde{A}(x)) + \cdots\Big] \,. \tag{24}$$

Thus, the metaparticle can be understood as a generic message of string theory/quantum gravity for low energy physics. Like their visible particle cousins, dark matter quanta should be detectable through their particular metaparticle correlation/entanglement to visible matter. This is a Berry-phase-like effect that comes from a fully covariant description, and is uniquely different from the usual effective field theory interaction terms between visible and dark matter particles. We will discuss the observable consequences of this view of dark matter in the last section.

5. Quantum gravity, dark matter and dark energy

We now explain how the generalized geometric formulation of string theory discussed above provides for an effective description of dark energy that is consistent with de Sitter spacetime. This is essentially due to the theory's chirally and non-commutatively doubled realization of the target space, and the stringy effective action on the doubled non-commutative spacetime $(x^a, \tilde{x}^a)$:

$$S^{nc}_{\text{eff}} = \iint \text{Tr}\sqrt{g(x,\tilde{x})}\,\big[R(x,\tilde{x}) + L_m(x,\tilde{x}) + \cdots\big] \,, \tag{25}$$

where the ellipses denote higher-order curvature terms induced by string theory, and L_m is the matter Lagrangian. S^{nc}_{eff} clearly expands into numerous terms with different powers of λ, which upon $\tilde{x}$-integration and from the x-space vantage point produce various effective terms. To lowest (zeroth) order of the expansion in the non-commutative parameter λ of S^{nc}_{eff} takes the form:

$$S_d = -\iint \sqrt{-g(x)}\sqrt{-\tilde{g}(\tilde{x})}\big[R(x) + \tilde{R}(\tilde{x})\big] \,, \tag{26}$$

a result which was first obtained almost three decades ago, effectively neglecting ω_{AB} by assuming that $[\hat{x}^a, \hat{\tilde{x}}_b] = 0$.[46] In this leading limit, the $\tilde{x}$-integration in the first term defines the gravitational constant G, and in the second term produces a *positive* cosmological constant $\Lambda > 0$. In particular, we are lead to the following low-energy effective action valid at long distances of the observed accelerated universe (focusing on the relevant 3+1-dimensional spacetime X, of the $+---$ signature):

$$S_{\text{eff}} = \frac{-1}{8\pi G}\int_X \sqrt{-g}\big(\Lambda + \tfrac{1}{2}R + \mathcal{O}(R^2)\big)\ , \tag{27}$$

with Λ the positive cosmological constant (corresponding to the scale of 10^{-3} eV) and the $\mathcal{O}(R^2)$ denote higher order corrections (which are also required by the sigma model of string theory[47]).

It also follows from this construction that the weakness of gravity is determined by the size of the canonically-conjugate dual space, while the smallness of the cosmological constant is given by its curvature. Given this action, we may proceed reinterpreting Ref. 46. Integrate out the dual spacetime coordinates and write the effective action as

$$\overline{S} \sim \tilde{V}\int_X \sqrt{-g(x)}R(x) + \cdots\ , \quad \text{where} \quad \widetilde{V} = \int_{\tilde{X}}\sqrt{-\tilde{g}(\tilde{x})}\ , \tag{28}$$

and then relate the dual spacetime volume to the observed spacetime volume as $\widetilde{V} \sim V^{-1}$ (T-duality). This produces an "intensive" effective action[46] :

$$\overline{S} = \frac{\int_X \sqrt{-g(x)}\,(R(x) + L_m(x))}{\int_X \sqrt{-g(x)}} + \cdots\ . \tag{29}$$

By concentrating on the classical description first (we discuss below quantum corrections and the central role of intrinsic non-commutativity in string theory) we obtain the Einstein equations[46]

$$R_{ab} - \frac{1}{2}Rg_{ab} + T_{ab} + \frac{1}{2}\overline{S}\,g_{ab} = 0\ , \quad \text{with} \quad T_{ab} \stackrel{\text{def}}{=} \frac{\partial L_m}{\partial g^{ab}} - \frac{1}{2}L_m\,g_{ab}\ . \tag{30}$$

We emphasize that our reinterpretation of Ref. 46 does not follow the original presentation and intention. In particular, we directly relate the intensive action to the cosmological constant, $\overline{S} \sim \Lambda$. Note that this new approach to the question of dark energy (viewed as a cosmological constant) in quantum gravity is realized in certain stringy-cosmic-string-like toy models,[48] which can be viewed as illustrative of a generic non-commutative phase of F-theory.[49] In particular, a "see-saw" formula is directly realized in Ref. 48 as $M_\Lambda \sim M^2/M_P$, where M_Λ is the dark energy scale, M_P the Planck scale, and M an intermediate scale coming from the matter sector (such as the Higgs scale).

Note that, in general, to lowest (zeroth) order of the expansion in the non-commutative parameter λ, S^{nc}_{eff} takes the following form (that also includes the matter sector and its dual)[50] :

$$S_d = -\iint \sqrt{g(x)\tilde{g}(\tilde{x})}\,\Big[R(x) + \tilde{R}(\tilde{x}) + L_m(A(x,\tilde{x})) + \tilde{L}_{dm}(\tilde{A}(x,\tilde{x}))\Big]\ . \tag{31}$$

Here, A denotes the usual SM fields, and $\tilde{A}$ denotes their duals. Note that after integrating over the dual spacetime, and after taking into account T-duality, the intensive action now reads

$$\overline{S} = \frac{\int_X \sqrt{-g(x)}\big(R(x) + L_m(x) + \tilde{L}_{dm}(x)\big)}{\int_X \sqrt{-g(x)}} + \cdots \,. \tag{32}$$

The proposal here is that the dual sector (as already indicated in the previous section) should be interpreted as the dark matter sector, which is correlated to the visible sector via the dark energy sector, as discussed in Ref. 12. (The radiative stability of this construction has been discussed in Ref. 50, which also addresses the hierarchy problem.) We emphasize the unity of the description of the entire dark sector based on the properties of the dual spacetime. Note that one can also discuss statistical effects in this view of dark energy and, in particular, provide an explicit formula for a dynamical form of dark energy[51] that can be compared to cosmological observations.

6. Coda: Quantum gravity phenomenology and the real world

In conclusion, we discuss the implications of our new view on dark matter and dark energy in the context of actual astronomical observations.

6.1. *Observational issues*

The above metaparticle-like dark matter quanta are by construction correlated to visible matter and have been discussed in the literature as Modified dark matter (MDM).[12] MDM is, at the moment, a phenomenological model of dark matter inspired by gravitational thermodynamics. For an accelerating Universe with positive cosmological constant Λ, certain phenomenological considerations lead to the emergence of a critical acceleration parameter related to Λ. This "fundamental acceleration" is just the value of Λ expressed as acceleration $\sim cH_0$, where H_0 is the Hubble constant, and thus, is of the order of $10^{-10}\mathrm{m/s}^2$. Appearance of this acceleration scale in the data is an expected manifestation of MDM, and its existence is observationally supported as discussed below. The resulting MDM mass profiles, which are sensitive to Λ, are consistent with observational data at both the galactic and cluster scales. In particular, the same critical acceleration appears both in the galactic and cluster data fits based on MDM.[12] Furthermore, using some robust qualitative arguments, MDM appears to work well on cosmological scales. If the quanta of MDM are metaparticles, this may explain why, so far, dark matter detection experiments have failed to detect dark matter particles. In particular, the natural model for MDM quanta could be provided by the metaparticle realizations of the SM particles, associated with bi-local extensions of all SM fields. Thus the baryonic matter described by the SM fields (the A backgrounds in the above discussion) would have natural cousins (the $\tilde{A}$ backgrounds in the above discussion) in the

dark matter sector, which in turn would be sensitive to the dark energy modeled by the cosmological constant $\Lambda \sim 1/H_0^2$, that is, the curvature of the dual spacetime, which is radiatively stable and related to the Planck energy and the characteristic energy scale of the visible sector $M_\Lambda \sim M^2/M_P$.

The importance of the fundamental acceleration $10^{-10}\mathrm{m/s}^2$ is manifest in the empirically established baryonic Tully-Fisher[53] (BTF), and baryonic Faber-Jackson[54] (BFJ) relations. The BTF relation refers to the observed correlation between the total baryonic mass of spiral galaxies M_{bar} and the rotational velocity in the flat part of the rotation curve v_{flat} of the form $M_{\mathrm{bar}} \sim v_{\mathrm{flat}}^4$. The slope of the BTF relation gives the fundamental acceleration $v_{\mathrm{flat}}^4/(GM_{\mathrm{bar}}) \sim 10^{-10}\mathrm{m/s}^2$. The relation for elliptical galaxies and other pressure supported systems that parallels the BTF for spirals is the BFJ relation between M_{bar} and the line-of-sight velocity dispersion σ of the form $M_{\mathrm{bar}} \sim \sigma^n$. When fit to data of various pressure supported systems, the best fit to the power n was found to be anywhere from about 3 to 5 depending on the scale of the systems considered. However, when the data for globular clusters, elliptical galaxies, galaxy clusters, etc. are all analyzed together, the preferred power is $n = 4$, though the data does show considerable scatter around this line.[52,55,56] Furthermore, the slope of the $n = 4$ BFJ fit again gives the fundamental acceleration $\sigma^4/(GM_{\mathrm{bar}}) \sim 10^{-10}\mathrm{m/s}^2$.

According to the ΛCDM model of cosmology, our Universe started out with almost uniform distributions of cold dark matter (CDM) and baryonic matter. Consequently, a correlation between the total dark matter and total baryonic matter in a region surrounding a galaxy is to be expected in the ΛCDM model. However, given that galaxies are thought to have gone through various phases including starbursts, emission of gasses, and multiple mergers during their evolutionary histories, a dynamic correlation such as the BTF and BFJ relations involving the fundamental acceleration scale is surprising. The question is: what do these correlations imply on the nature of dark matter? What type of dark matter would be able to explain these relations?

6.2. *Analogy with turbulence*

Given that the temperature fluctuations in the Cosmic Microwave Background have been observed to be scale invariant,[57] and assuming that dark matter (whether traditional or metaparticle) is collisionless, then an analogy with Kolmogorov's theory of turbulence may be instructive.[58]

In the so-called inertial range of the turbulent fluid, *i.e.* at length scales L where the viscosity of the fluid can be neglected, the dynamics of vortices is determined solely by the rate of energy dissipation ε, and if one writes $\langle \mathrm{KE} \rangle = \int dk\, E(k)$, where $\langle \mathrm{KE} \rangle$ is the mean turbulent kinetic energy of the flow and $k = 2\pi/L$ is the wave number, simple dimensional analysis yields the famous Kolmogorov scaling relation $E(k) \propto \varepsilon^{2/3} k^{-5/3}$, which has been confirmed experimentally. The length scale η

below which viscous effects become important, called the Kolmogorov microscale, is the boundary of the inertial region. Again, from dimensional analysis, it is argued that $\eta \propto (\nu^3/\varepsilon)^{1/4}$, where ν is the kinematic viscosity. Essentially, the scale η is determined by the competition between ε and ν. Note that this argument assumes that there is no other dimension-ful scale in the problem other than these two parameters.

In the case of structure formation in a statistically self-similar (*i.e.* scale invariant) Universe, the two competing effects would be attraction due to gravity, parametrized by G, and the speed of the expansion of the Universe $H(z) = \dot{a}/a$, where z is the redshift. The scale of the structures that form at redshift z should be determined by G and $H(z)$. Now, consider the size $R_{\rm vir}$ of a virialized structure of total mass M. We expect $\langle v^2 \rangle \sim \sigma^2 \sim GM/R_{\rm vir}$. The gravitational acceleration at the edge of this structure is $a_{\rm vir} = GM/R_{\rm vir}^2 \sim \sigma^4/(GM)$. Thus, the scale of the structure $R_{\rm vir}$ can be characterized by this acceleration scale $a_{\rm vir}$. From dimensional analysis a la Kolmogorov we expect $a_{\rm vir} \sim \sigma^4/(GM) \propto cH(z)$, which is analogous to the $n = 4$ FJ relation except 1) M is the sum of the dark matter mass and the baryonic mass, 2) dark matter halos are not necessarily virialized, 3) the proportionality constant may not be order one, 4) the z dependence of $H(z)$ suggests that $\sigma^4/(GM)$ will depend on the redshift z at which the structures formed, and 5) in actuality, there are many other scales present. Nevertheless, this simple handwaving argument does suggest a possible path toward the BFJ, or a BFJ-like relation. Note also that metaparticle quanta are sensitive to both the UV (G, the Planck scale) and the IR ($H(z)$), suggesting that they may provide the linchpin connecting those scales, and the crucial ingredient in constructing a dark matter model that could realize this scenario.

One intriguing prediction of the above discussion is the possible z-dependence of $\sigma^4/(GM)$. Replacing M by $M_{\rm bar}$ should not erase this dependence. Dimensionally, $\sigma^4/(GM_{\rm bar})$ should scale as $(1 + z)$. Galaxy clusters have formed recently ($z < 1$), galaxies have started to form at around $z \sim 10$, and globular clusters at around $z \sim 10^2$.[59] So if $\sigma^4/(GM_{\rm bar})$ is around $10^{-10}\,{\rm m/s^2}$ for galaxy clusters, it should be around $10^{-9}\,{\rm m/s^2}$ for galaxies, and $10^{-8}\,{\rm m/s^2}$ for globular clusters. In Fig. 1 we plot $\sigma^4/(GM_{\rm bar})$ for a variety of structures. As can be seen, the galaxy cluster data points are clustered around $10^{-10}\,{\rm m/s^2}$, while the galaxy data points are spread out in both directions and go up as far as $\sim 10^{-9}\,{\rm m/s^2}$, and the globular cluster data are spread out even further and go up as far as $\sim 10^{-8}\,{\rm m/s^2}$. One interpretation of this result is that galaxies and globular clusters started out with characteristic accelerations of $O(10^{-9}\,{\rm m/s^2})$ and $O(10^{-8}\,{\rm m/s^2})$, respectively, but have migrated to lower values as they went through various stages of evolution, resulting in distributions centered around $10^{-10}\,{\rm m/s^2}$ (in the log scale). This possibility had not been manifest in previous BFJ analyses which had always plotted σ^4 against $M_{\rm bar}$, demonstrating that a simple change of perspective can open up new vistas.

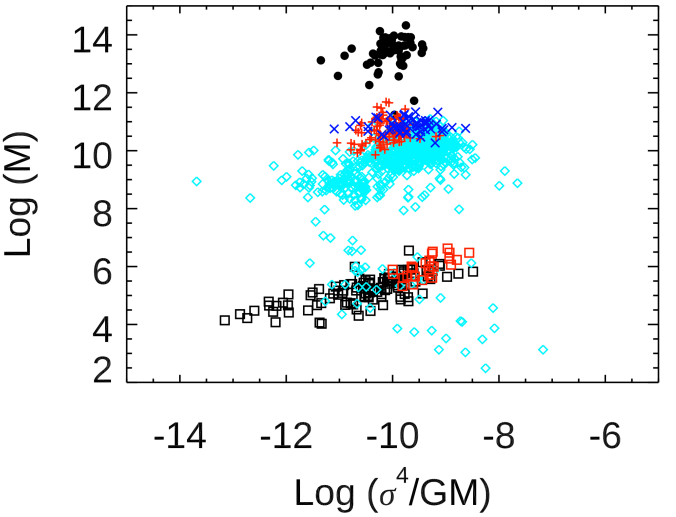


Fig. 1. The virial acceleration for structures spanning about 10 decades of scale $M/M_{\odot} = 10^{4\sim14}$. The data points are: • galaxy clusters,[60] + elliptical galaxies,[61,62] × elliptical galaxies,[63,64] ◇ elliptical, dwarf elliptical, and dwarf spheroidal galaxies,[65] □ Milky Way globular clusters,[66] □ M31 (Andromeda) globular clusters.[67]

6.3. *Closing remarks*

In this paper, we propose that the BFJ relation could be explained by metaparticle-based, or similar MDM models. We are, however, mindful that it could well be explained by more conventional means. Indeed, Kaplinghat and Turner[68] have provided a scenario on how the BTF relation may emerge from ΛCDM, though we are unaware of a similar work on the BFJ relation. MDM models, though suggestive, are lacking in similar concrete scenarios that would connect the models to the BFJ relation; a direction of research we intend to pursue.

This concludes our presentation of the new approach to quantum gravity (based on modular spacetime, Born geometry and metastring theory) and quantum gravity phenomenology in the context of dark matter and dark energy and actual astronomical observations.

Acknowledgements

We are grateful to L. Freidel, R. G. Leigh, J. Kowalski-Glikman, P. Berglund, T. Hübsch, D. C. Dai, D. Stojkovic, J. H. Simonetti, M. Kavic and V. Jejjala for discussions. The research of DM and TT is supported in part by the US Department of Energy (DE-SC0020262). DM also thanks the Julian Schwinger Foundation, and TT the NSF (PHY-1413031) for support.

References

1. L. Freidel, R. G. Leigh and D. Minic, Phys. Lett. B **730**, 302 (2014).
2. L. Freidel, R. G. Leigh and D. Minic, Int. J. Mod. Phys. D **23**, 1442006 (2014).
3. L. Freidel, R. G. Leigh and D. Minic, JHEP **1506**, 006 (2015).
4. L. Freidel, R. G. Leigh and D. Minic, Int. J. Mod. Phys. D **24**, 1544028 (2015).
5. L. Freidel, R. G. Leigh and D. Minic, Phys. Rev. D **94**, 104052 (2016).
6. L. Freidel, R. G. Leigh and D. Minic, J. Phys. Conf. Ser. **804**, 012032 (2017).
7. L. Freidel, R. G. Leigh and D. Minic, JHEP **1709**, 060 (2017).
8. L. Freidel, R. G. Leigh and D. Minic, Phys. Rev. D **96**, 066003 (2017).
9. L. Freidel, J. Kowalski-Glikman, R. G. Leigh and D. Minic, Phys. Rev. D **99**, 066011 (2019).
10. L. Freidel, R. G. Leigh and D. Minic, Int. J. Mod. Phys. A **34**, 1941004 (2019), D. Minic, arXiv:2003.00318 [hep-th].
11. P. Berglund, T. Hubsch and D. Minic, Phys. Lett. B **798**, 134950 (2019); Int. J. Mod. Phys. D **28**, 1902003 (2019); JHEP **1912**, 166 (2019) [JHEP **2019**, 166 (2020)].
12. C. M. Ho, D. Minic and Y. J. Ng, Phys. Lett. B **693**, 567 (2010); Gen. Rel. Grav. **43**, 2567 (2011) [Int. J. Mod. Phys. D **20**, 2887 (2011)]; Phys. Rev. D **85**, 104033 (2012), D. Edmonds, D. Farrah, C. M. Ho, D. Minic, Y. J. Ng and T. Takeuchi, Astrophys. J. **793**, 41 (2014); Int. J. Mod. Phys. A **32**, 1750108 (2017), D. Edmonds, D. Farrah, D. Minic, Y. J. Ng and T. Takeuchi, Int. J. Mod. Phys. D **27**, 1830001 (2017).
13. K. G. Wilson and J. B. Kogut, Phys. Rept. **12**, 75 (1974); K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
14. J. Zak, Phys. Rev. **168**, 686 (1968); Phys. Rev. Lett. **19**, 1385 (1967).
15. M. Born, Nature 136 (1935) 952; Rev. Mod. Phys. 21 (1949) 463.
16. G. Amelino-Camelia, L. Freidel, J. Kowalski-Glikman and L. Smolin, Phys. Rev. D **84**, 084010 (2011).
17. Y. Aharonov, A. Petersen and H. Pendleton, Int. J. Theor. Phys. **2**, 213 (1969).
18. H. Weyl, *The theory of groups and quantum mechanics* (Dover, New York, 1931).
19. A. Connes, *Noncommutative Geometry* (Academic Press, San Diego, 1994); A. Connes and M. Marcolli, *Noncommutative Geometry, Quantum Fields and Motives* (AMS, 2007).
20. I. M. Gelfand, and M. A. Naimark, *Mat. Sbornik* **12** 197 (1943).
21. Y. Aharonov, and D. Rohrlich, *Quantum Paradoxes: Quantum Theory for the Perplexed* (Wiley, New York, 2005).
22. L. N. Chang, Z. Lewis, D. Minic and T. Takeuchi, Mod. Phys. Lett. B **27**, 1350064 (2013); J. Phys. A **46**, 065304 (2013); J. Phys. A **46**, 485306 (2013); Int. J. Mod. Phys. A **29**, 1430006 (2014); AIP Conf. Proc. **1508**, no. 1, 502 (2012), L. N. Chang, D. Minic and T. Takeuchi, J. Phys. Conf. Ser. **1275**, no. 1, 012036 (2019).
23. G. Mackey, *Annals Math.* **55** 101 (1952); *Acta Math.* **99** 265 (1958).
24. Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
25. M. R. Douglas and N. A. Nekrasov, Rev. Mod. Phys. **73**, 977 (2001); R. J. Szabo, Phys. Rept. **378**, 207 (2003); H. Grosse and R. Wulkenhaar, Commun. Math. Phys. **256**, 305 (2005).
26. D. Karabali, C. J. Kim and V. P. Nair, Phys. Lett. B **434**, 103 (1998), R. G. Leigh, D. Minic and A. Yelnikov, Phys. Rev. Lett. **96**, 222001 (2006); Phys. Rev. D **76**, 065018 (2007), L. Freidel, R. G. Leigh and D. Minic, Phys. Lett. B **641**, 105 (2006), L. Freidel, R. G. Leigh, D. Minic and A. Yelnikov, AMS/IP Stud. Adv. Math. **44**, 109 (2008).

27. A. A. Tseytlin, Nucl. Phys. B **350** 395 (1991); Phys. Lett. B **242** 163 (1990), R. Floreanini, and R. Jackiw, Phys. Rev. Lett. **59** 1873 (1987), W. Siegel, Phys. Rev. D **47** 5453 (1993); Phys. Rev. D **48**, 2826 (1993) C. Hull, and B. Zwiebach, JHEP **0909** 099 (2009), O. Hohm, C. Hull and B. Zwiebach, JHEP **1007**, 016 (2010), O. Hohm, D. Lust and B. Zwiebach, Fortsch. Phys. **61**, 926 (2013).
28. J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string* (Cambridge University Press, 1998), M. B. Green, J. H. Schwarz and E. Witten, *Superstring theory: Volume 1* (Cambridge University Press, 1987); A. M. Polyakov, *Gauge Fields and Strings* (Harwood, 1987).
29. D. Minic, J. Polchinski and Z. Yang, Nucl. Phys. B **369**, 324 (1992); J. Polchinski, "What is string theory?", hep-th/9411028.
30. See, for example, L. N. Chang, Z. Lewis, D. Minic and T. Takeuchi, Adv. High Energy Phys. **2011**, 493514 (2011) and references therein.
31. M. Kavic, D. Minic and J. Simonetti, Int. J. Mod. Phys. D **17**, 2495 (2009); Int. J. Mod. Phys. D **27**, 1847007 (2018), J. H. Simonetti, M. Kavic, D. Minic, U. Surani and V. Vijayan, Astrophys. J. **737**, L28 (2011), D. C. Dai, D. Minic and D. Stojkovic, Phys. Rev. D **98**, 124026 (2018), D. C. Dai, D. Minic, D. Stojkovic and Changbo Fu, arXiv:2002.08178 [hep-th], J. H. Simonetti, M. J. Kavic, D. Minic, D. Stojkovic and D. C. Dai, arXiv:2007.12184 [gr-qc].
32. E. Boffo and P. Schupp, JHEP **2001**, 007 (2020)
33. M. Gunaydin and D. Minic, Fortsch. Phys. **61**, 873 (2013); R. J. Szabo, PoS CORFU **2018**, 100 (2019), and references therein.
34. M. Gunaydin and F. Gursey, Phys. Rev. D **9**, 3387 (1974).
35. U. Aydemir, D. Minic and T. Takeuchi, Phys. Lett. B **724**, 301 (2013); U. Aydemir, D. Minic, C. Sun and T. Takeuchi, Phys. Rev. D **91**, 045020 (2015); Int. J. Mod. Phys. A **31**, 1550223 (2016); Mod. Phys. Lett. A **31**, 1650101 (2016); JHEP **1809**, 117 (2018).
36. N. Ishibashi, H. Kawai, Y. Kitazawa, and A. Tsuchiya, Nucl. Phys. **B498** (1997) 467.
37. T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, Phys. Rev. **D55** (1997) 5112.
38. G. 't Hooft, Conf. Proc. C **930308**, 284 (1993); L. Susskind, J. Math. Phys. **36**, 6377 (1995).
39. J. M. Maldacena, Int. J. Theor. Phys. **38**, 1113 (1999) [Adv. Theor. Math. Phys. **2**, 231 (1998)], S. S. Gubser, I. R. Klebanov and A. M. Polyakov, Phys. Lett. B **428**, 105 (1998), E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998).
40. A. Strominger, JHEP **0110**, 034 (2001), E. Witten, hep-th/0106109, V. Balasubramanian, J. de Boer and D. Minic, Class. Quant. Grav. **19**, 5655 (2002) [Annals Phys. **303**, 59 (2003)]; Phys. Rev. D **65**, 123508 (2002), V. Balasubramanian, P. Horava and D. Minic, JHEP **0105**, 043 (2001).
41. D. Minic and M. Pleimling, Phys. Lett. B **700**, 277 (2011), N. Gray, D. Minic and M. Pleimling, Int. J. Mod. Phys. A **28**, 1330009 (2013), V. Balasubramanian, E. G. Gimon, D. Minic and J. Rahmfeld, Phys. Rev. D **63**, 104009 (2001).
42. D. Minic, hep-th/9909022; hep-th/0009131, H. Awata, M. Li, D. Minic and T. Yoneya, JHEP **0102**, 013 (2001), D. Minic and H. C. Tze, Phys. Lett. B **536**, 305 (2002); Phys. Rev. D **68** 061501 (2003); Phys. Lett. B **581**, 111 (2004); V. Jejjala, M. Kavic, and D. Minic, Int. J. Mod. Phys. A **22** 3317 (2007); D. Minic, T. Takeuchi and C. H. Tze [arXiv:2012.06583 [hep-ph]].
43. G. T. Horowitz and L. Susskind, J. Math. Phys. **42**, 3152 (2001). See also F. Englert, H. Nicolai and A. Schellekens, Nucl. Phys. B **274**, 315 (1986); J. J. Atick and E. Witten, Nucl. Phys. B **310**, 291 (1988); R. Penrose, Phil. Trans. Roy. Soc. Lond. A **373**, 20140237 (2015).

44. D. Minic, Phys. Lett. B **442**, 102 (1998), T. Banks and W. Fischler, hep-th/0102077, V. Balasubramanian, J. de Boer and D. Minic, gr-qc/0211003.
45. L. Freidel, J. Kowalski-Glikman, R. G. Leigh and D. Minic, arXiv:2104.00802 [gr-qc].
46. A. A. Tseytlin, Phys. Rev. Lett. **66**, 545 (1991).
47. D. H. Friedan, Ann. Phys. **163** (1985) 318–419; Phys. Rev. Lett. **45**, 1057 (1980).
48. P. Berglund, T. Hübsch, and D. Minic, JHEP **09** (2000) 015; JHEP **02** (2001) 010; JHEP **01** (2001) 041; Phys. Lett. B **512** (2001) 155–160; Phys. Lett. B **534** (2002) 147–154; Phys. Rev. D **67** (2003) 041901.
49. C. Vafa, Nucl. Phys. B **469** no. 3, (1996) 403–415.
50. P. Berglund, T. Hübsch, and D. Minic, LHEP **2021**, 186 (2021); arXiv:2109.01122 [hep-th], V. Jejjala, R. G. Leigh, and D. Minic, Phys. Lett. B **556** (2003) 71–79.
51. V. Jejjala, M. Kavic and D. Minic, Adv. High Energy Phys. **2007**, 21586 (2007), V. Jejjala, M. Kavic, D. Minic and T Takeuchi, arXiv:2011.08852 [hep-th].
52. D. Edmonds, D. Minic and T. Takeuchi, Int. J. Mod. Phys. D **29**, 2043030 (2020); arXiv:2009.12915 [astro-ph.CO].
53. R. B. Tully and Fisher, J. R., Astron. Astrophys. **54** (3): 661-673 (1977).
54. S. M. Faber and R. E. Jackson, Astrophys. J. **204**, 668-683 (1976).
55. R. T. Farouki, S. L. Shapiro and M. J. Duncan, Astrophys. J. **265**, 597 (1983).
56. R. H. Sanders, Astron. Astrophys. **284**, L31 (1994).
57. N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020) [erratum: Astron. Astrophys. **652**, C4 (2021)].
58. U. Frisch, *Turbulence: The Legacy of Kolmogorov* (Cambridge University Press, 1996).
59. G. Bothun, *Modern Cosmological Observations and Problems* (CRC Press, 1st ed., 1998)
60. Y. Y. Zhang, et al., Astron. Astrophys. **526**, 105 (2011).
61. M. Cappellari, et al., Mon. Not. Roy. Astron. Soc. **432**, 1709 (2013).
62. M. Cappellari, et al., Mon. Not. Roy. Astron. Soc. **432**, 1862 (2013).
63. S. Beli, A. B. Newman, and R. S. Ellis, Astrophys. J. **783**, 117 (2014).
64. S. Beli, A. B. Newman, and R. S. Ellis, Astrophys. J. **799**, 206 (2015).
65. J. Dabringhausen and M. Fellhauer, Mon. Not. Roy. Astron. Soc. **460**, 4492 (2016).
66. H. Baumgardt and M. Hilker, Mon. Not. Roy. Astron. Soc. **478**, 1520 (2018).
67. J. Strader, et al., Astron. J. **138**, 547 (2009).
68. M. Kaplinghat and M. Turner, Astrophys. J. Lett. **569**, L19-L22 (2002).

WKB approach to the gravity-matter dynamics: A cosmological implementation

G. Maniccia[1,2,*] and G. Montani[1,3,†]

[1] *Department of Physics, "La Sapienza" University of Rome, P.le A. Moro 5, 00185 Roma, Italy*
[2] *INFN Section of Rome, Piazzale Aldo Moro 2, 00185 Roma, Italy*
[3] *ENEA, C.R. Frascati, Via E. Fermi 45, 00044 Frascati (Roma), Italy*
[*] *E-mail: giulia.maniccia@uniroma1.it*
[†] *E-mail: giovanni.montani@enea.it*

The problem of time emerging in the canonical quantization procedure of gravity signals a necessity to properly define a relational time parameter. Previous approaches, which are here briefly discussed, make use of the dependence of the quantum system on semi-classical gravitational variables in order to define time. We show that such paths, despite the following studies, lead to a non-unitary evolution. We propose a different model for the quantization of the gravity-matter system, where the time parameter is defined via an additional term, i.e. the kinematical action, which acts as a clock for quantum matter. The procedure here used implements a Born-Oppenheimer-like separation of the system, which maintains covariance under the foliation of the gravitational background and keeps the correct classical limit of standard quantum field theory on a fixed background. It is shown with a WKB expansion that quantum gravity corrections to the matter dynamics arise at the next order of expansion, and such contributions are unitary, signaling a striking difference from previous proposals. Applications to a cosmological model are presented and the analogies of the kinematical term with an incoherent dust are briefly discussed.

Keywords: Quantum Gravity; Quantum Field Theory; Born-Oppenheimer approach; Cosmology.

1. Introduction

A suitable quantum description of the gravity-matter system is necessary in order to describe the Universe and its early phases. An important requirement for this purpose is the quantization of the gravitational field, which has opened a debate leading to many approaches and outcomes. The first candidate for this procedure is the Dirac prescription,[1] which leads however to the problem of time[2]: the Universe wave function does not evolve on time, hence a probabilistic interpretation is not clear. For this reason a different clock must be introduced[2,3]: essentially, we have to define a *relational time*[4] in order to describe the quantum system dynamics, adopting a subpart of the system as a clock through carefully chosen boundary conditions.

A first step in this path was proposed in,[5] by separating the system into a semiclassical component and the purely quantum sector due to their different energy scales, in analogy with the Born-Oppenheimer approximation.[6] The external time

was introduced via the dependence of the quantum subsystem on the semiclassical variables and, with a Wentzel-Kramer-Brillouin expansion[7] (WKB) in the Planck constant, the limit of standard quantum field theory on curved background was recovered at the first order, together with a positive-semidefinite probability density under certain conditions (see the original article[5] for such discussion).

Other models,[8–10] which differ for the choice of expansion parameter and initial conditions, have been studied focusing on the next order of expansion. The work[8] in particular dealt with computing quantum gravity induced corrections to the quantum matter dynamics, finding that non-unitary terms emerge. The workaround,[10] proposed to avoid such troublesome feature, cannot be applied in many contexts due to the strict initial requirements (see discussion in[11]). The work[9] followed more closely a Born-Oppenheimer approach on the problem, in order to define a conserved probability density; also in this case, however, it can be shown via the WKB expansion that the Hamiltonian operator describing the system dynamics is not unitary at the next order.[11]

Here we shall present a different implementation of the gravity-matter system, that ensures the standard limit of quantum field theory on curved background, with an expansion parameter proportional to the squared Planck mass (symbolizing the energy scale of gravity). The time parameter will not be linked to the dependence of the quantum sector on the semiclassical variables, but instead arise from an additional Hamiltonian term, i.e. the *kinematical action*,[12] that reinstates covariance of the theory in Arnowitt-Deser-Misner (ADM) variables[13] for assigned foliations. This term, together with the definition of the deformation vector, will allow us to obtain a functional Schrödinger equation for the quantum matter sector and, at the next order, a unitary dynamics including quantum gravity corrections.

An application will be shown for the case of the homogeneous free massless scalar field in a spatially flat Friedmann-Lemaitre-Robertson-Walker background, representing the inflaton field during the de Sitter phase; in this case such quantum gravity corrections indeed modify the Hamiltonian spectrum, with non-vanishing but small contributions due to the perturbative nature of the approach.

An important analogy is also discussed, linking the introduction of the *kinematical action* to the frame fixing procedure using Gaussian coordinates,[14] that emerge as a fluid term (more specifically, an incoherent dust) in the action, and showing how the two procedures are related.

The structure of the paper is the following. In Sec. 2 we will briefly introduce the WKB expansion of the gravity-matter system, discussing the mentioned approaches in literature and their outcomes. In Sec. 3 we propose a model constructing the time parameter from the kinematical action and show that its WKB expansion gives a unitary dynamics for the quantum matter sector at the order next to standard quantum field theory. In Sec. 4 the procedure is applied to a case representing the homogeneous inflaton field, in order to analyze the outcome in a cosmological view. In Sec. 5 the analogy of such construction with the frame-fixing procedure is presented. Sec. 6 contains the conclusive remarks.

2. Main features of the WKB expanded gravity-matter system

The starting point for the study of quantum gravity effects is the quantization of the gravitational field. It is known that, following the Dirac prescription,[1] one finds a time-independent Schrödinger equation which describes no evolution in time: the equation, known as the Wheeler-DeWitt[3] equation, is the essence of the so-called problem of time.[2] Then, in order to overcome such problem, one has to define a relational time parameter,[4] meaning that a selected subsystem must be chosen as a new, different clock to describe the evolution.

Let us briefly recall how this critical result emerges. We start with the foliation of space-time proposed by Arnowitt, Deser and Misner[13]: we identify 3d hypersurfaces immersed in the 4d spacetime environment by using some parametric equations, and derive the induces properties of such hypersurfaces from the environment ones. We label the induced 3d metric with the tensor h_{ij}, and the normal vector field with n^{μ}. Using these variables, one can compute the Hamiltonian formalism of the gravitational field, finding primary and secondary constraints on the wave function of the system.[15] It follows that the wave function depends only on the equivalence class of the induced geometries $\{h_{ij}\}$, and on the matter fields ϕ_a if present. Due to the superHamiltonian constraint, the action of the superHamiltonian on the system wave function must vanish:

$$H\Psi = \left(-\frac{2\hbar^2\kappa}{\sqrt{h}}\nabla_g^2 - \frac{\sqrt{h}\,R^{(3)}}{2\kappa} - \frac{\hbar^2}{2\sqrt{h}}\nabla_m^2 + u(h_{ij},\phi_a) \right)\Psi = 0 \tag{1}$$

where h refers to the determinant of the induced metric, $R^{(3)}$ to its curvature, κ is the Einstein coefficient $\kappa = 8\pi G/c^3$ and $u(h_{ij},\phi_a)$ is the total potential energy of the matter fields. The gradients ∇_g and ∇_m indicate respectively the derivatives with respect to the gravitational and matter variables.

This result is troublesome since it implies the problem of time,[2] where the associated Schrödinger equation states that the Universe wave function does not evolve in time at all. It follows that another time parameter[4] should be implemented in order to describe a non-trivial, meaningful dynamics.

A semiclassical approach can be implemented in this sense. By following the Wentzel-Kramer-Brillouin expansion,[7] we write the system wave function as the complex exponential $\mathrm{e}^{\mathrm{i}S/\hbar}$ and expand it with a chosen parameter, examining the dynamics order by order. The choice of parameter will determine the range of validity: it can be chosen as the Planck constant $\hbar$, or related to the inverse of the Einstein coefficient κ, i.e. proportional to the square of the reduced Planck mass (here labeled m_{Pl}):

$$M \equiv \frac{1}{4c^2\kappa} = \frac{c\,(m_{\mathrm{Pl}})^2}{4\hbar}. \tag{2}$$

Although different parameters can be implemented, these two cases present interesting results and therefore will be discussed in this paper.

The works under examination[5,8] differ from each other in few but fundamental starting hypotheses; let us briefly discuss them in the following.

2.1. *Expansion in* $\hbar$

In the $\hbar \to 0$ expansion,[5] the variables are separated into semiclassical c and quantum ones q, so that Eq. (1) reads

$$\left(-\hbar^2\nabla_c^2 + U_c(c) - \hbar^2\nabla_q^2 + U_q(c,q)\right)\Psi(c,q) = 0, \tag{3}$$

where the operator H_c is obtained neglecting all the quantum variables and U_c, U_q are respectively the semiclassical and quantum sector potentials.

The first assumption states that the action of the quantum Hamiltonian on the system wave function Ψ is small with respect to the semiclassical one:

$$\frac{\hat{H}_q\Psi}{\hat{H}_c\Psi} = \mathcal{O}(\hbar). \tag{4}$$

Secondly, it is assumed that the semiclassical and quantum subspaces are orthogonal; stronger requirements are needed in order to perform the expansion after the quantum mechanical order $\mathcal{O}(\hbar)$ (see discussion in[11]).

These assumptions allow a clean factorization of the action, which is implemented at each order of expansion after the first one:

$$S_n = \sigma_n(c) + \eta_n(c,q), \quad n \geq 1 \tag{5}$$

so that

$$S = S_0 + P + Q \quad \text{with} \quad P(c) = \sum_{n=1}^{\infty}\hbar^n\sigma_n, \quad Q(c,q) = \sum_{n=1}^{\infty}\hbar^n\eta_n, \tag{6}$$

being the term S_0 at lowest order $\mathcal{O}(\hbar^0)$ purely classical. This factorization is reflected in the wave function:

$$\Psi(c,q) = \psi(c)\chi(c,q) = \mathrm{e}^{\mathrm{i}(S_0+P)/\hbar}\,\mathrm{e}^{\mathrm{i}Q/\hbar}, \tag{7}$$

where the background part $\psi(c)$ is assumed to satisfy the semiclassical part of the Wheeler-DeWitt equation:

$$\left(-\hbar^2\nabla_c^2 + U_c\right)\psi(c) = 0. \tag{8}$$

Expanding up to order $\hbar$, a functional Schrödinger equation is recovered for the quantum subsystem χ_q in the background given by the semiclassical variables:

$$\mathrm{i}\hbar\frac{\partial\chi_1}{\partial\tau} = \hat{H}_q\chi_1, \tag{9}$$

with the following definition of time derivative

$$\frac{\partial}{\partial\tau} = 2\,\nabla_c S_0 \cdot \nabla_c\,. \tag{10}$$

This implementation describes a non-trivial dynamics (Eq. (9)) and allows the definition of a positive-semidefinite probability density, with suitable requirements on the background foliation, so that the standard interpretation of the wave function for a small subsystem of the Universe is recovered.

It shall be noted that in this process the semiclassical variables, from which the time parameter is constructed, have been essentially treated as classical ones. At the next order of expansion $\mathcal{O}(\hbar^2)$, not present in the original work, the quantum properties of the background should start to influence the quantum subsystem. By continuing the calculation,[11] it can be seen that some correction terms to the quantum dynamics emerge:

$$\mathrm{i}\hbar\frac{\partial\chi_2}{\partial\tau} = \hat{H}_q\chi_2 - \left(2\mathrm{i}\hbar^2\,\nabla_c\sigma_1\cdot\nabla_c + \hbar^2\,\nabla_c^2\right)\chi_2. \tag{11}$$

It is straightforward to prove that these additional terms, which are attributed to the non-purely classical behavior of the background, are not unitary, thus presenting us a crucial point of discussion.

2.2. *Expansion in $M \propto G$*

The work[8] uses the parameter M defined in Eq. (2) for the expansion. Being it representative of the Planck scale, this choice implies that the lowest order will only see gravity and not the matter sector, describing a vacuum Universe in the limit $M \to \infty$. Eq. (1) becomes then:

$$\left(-\frac{\hbar^2}{2M}\nabla_g^2 + MV(g) + H_m\right)\Psi(g,m) = 0 \tag{12}$$

where H_m contains all the matter terms. The wave function is factorized in a similar matter (see Eqs. (5),(6)) into a gravitational function and a matter one, each expanded with the WKB method:

$$\Psi(g,m) = \psi(g)\chi(m;g) = \mathrm{e}^{\mathrm{i}(MS_0+P)/\hbar}\,\mathrm{e}^{\mathrm{i}Q/\hbar}\,, \tag{13}$$

where the action at the lowest order (pure gravity) is of order M.

Similarly, a time-dependent non-trivial Schrödinger equation (see Eq. (9)) for the matter sector is obtained, defining again time from the dependence on the semiclassical gravitational variables:

$$\frac{\partial}{\partial\tau} = \nabla_g S_0\cdot\nabla_g. \tag{14}$$

The authors also investigate the next order of expansion, finding that the quantum matter dynamics is described by:

$$\mathrm{i}\hbar\frac{\partial\chi_2}{\partial\tau} = \hat{H}_q\chi_2 - \left(\frac{\mathrm{i}\hbar}{M}\nabla_g\sigma_1\cdot\nabla_g + \frac{\hbar^2}{2M}\nabla_g^2\right)\chi_2, \tag{15}$$

presenting non-unitary terms.

Cosmological implications of such a result were presented in.[16] A possible workaround for the issue developed by the authors in,[10] which eliminated the non-Hermitian part of the matter Hamiltonian by suitable redefinitions of the wave functions in the product $\Psi = \psi\chi$, is not applicable to many cases (see discussion in[11]) due to the basic assumptions implemented, thus leaving the problem of how to deal with non-unitarity. We stress that both definitions of time (see Eqs. (10), (14)) make use of the dependence of the quantum subsystem on the background semiclassical variables.

It is essential to cite the work,[9] which applied a procedure more similar to the Born-Oppenheimer method, by subtracting to the initial Eq. (1) its average over the quantum function χ. However, once WKB expanded in the M parameter, with a similar time definition as in Eqs. (10), (14), the total quantum Hamiltonian operator still presents a non-unitary morphology.[11]

3. A different definition of time via the kinematical action

We have seen that the previously mentioned definitions of relational time, using the dependence on the semiclassical variables, give a non-unitary matter Hamiltonian operator due to quantum gravitational corrections (Eqs. (11), (15)). This result seems a direct consequence of such construction, even with different expansion parameters (we refer for this point to Ssec. 4.1 of the mentioned article[11]).

For this reason we here propose a different construction, which implements the kinematical action[12] into the theory as a clock for quantum matter.

3.1. *The kinematical action*

The kinematical action, first introduced in,[12] is an additional action term used to ensure some constraint equations of a quantum system, in analogy with a Lagrangian multiplier. We remind that, in the Hamiltonian formalism, gravity presents both primary and secondary constraints[15]; the components of the deformation vector of the ADM splitting[13]

$$N^\mu = \partial_t y^\mu = N n^\mu + N^i b_i^\mu \tag{16}$$

play the role of the Lagrangian multipliers for the secondary constraints.

When a specific splitting is assigned, however, the components N and N^i still appear in the action of the quantum field theory as multipliers, but their initial meaning as components of the four-vector N^μ is not evident. To recover this, the kinematical action in ADM variables can be inserted:

$$S_k = \int d^4x (p_\mu \partial_t y^\mu - N^\mu p_\mu), \tag{17}$$

where $y^\mu = y^\mu(x^i; x^0)$ are the coordinates defining the parametric equations of the hypersurfaces, and p^μ are the conjugate momenta. Inserting this term into the action, some additional equations of motion appear by variations of y^μ, p_μ and N^μ:

these equations ensure[12] that the momenta p_μ are trivial and recover the physical meaning of the deformation vector N^μ as in Eq. (16).

Additional contributions also emerge in the Hamiltonian formalism to the superHamiltonian and supermomentum functions:

$$\mathcal{H}^k = n^\mu p_\mu \,, \tag{18a}$$

$$\mathcal{H}_i^k = b_i^\mu p_\mu \tag{18b}$$

which are linear in the conjugate momenta, a feature which is crucial in order to construct the time parameter of the theory.

3.2. *Basic hypotheses and conditions*

Let us now start with a Born-Oppenheimer-like separation of the wave function:

$$\Psi(h_a, \phi, y^\mu) = \psi(h_a)\chi(\phi, y^\mu; h_a) \tag{19}$$

where ψ is the slow gravitational sector, and χ the fast quantum one. We are considering the kinematical action as a matter term, regarding its variables as fast. This choice ensures the correct limit at the order of standard quantum theory on curved background.

As stated in,[5] since the matter fields live at an energy scale much lower than the Planck scale, we can reasonably assume that the slow function satisfies the semiclassical Wheeler-DeWitt equation, as in Eq. (8). For the sake of generality, we will not work in the minisuperspace, so that supermomentum constraints are not automatically satisfied. With the same reasoning we assume that the slow function satisfies the gravitational supermomentum constraint:

$$\left[-\frac{\hbar^2}{2M}\left(\nabla_g^2 + g\cdot\nabla_g\right) + MV\right]\psi = 0\,, \tag{20}$$

$$2\mathrm{i}\hbar\, h_i \mathrm{D}\cdot\nabla_g\psi = 0\,. \tag{21}$$

By insertion of the kinematical action, the equations satisfied by the total wave function Ψ are:

$$\left[-\frac{\hbar^2}{2M}\left(\nabla_g^2 + g\cdot\nabla_g\right) + MV - \hbar^2\nabla_m^2 + U_m\right]\Psi = \mathrm{i}\hbar\, n^\mu \frac{\delta}{\delta y^\mu}\Psi\,, \tag{22}$$

$$\left(2h_i\,\mathrm{D}\cdot\nabla_g - \partial_i\phi\cdot\nabla_m\right)\Psi = \mathrm{i}\hbar\, b_i^\mu \frac{\delta}{\delta y^\mu}\Psi\,. \tag{23}$$

We implement M, defined in Eq. (2), as the expansion parameter, so that the perturbative approach will be valid for fields with small associated energy with respect to the Planckian one. Formally, this property can be stated as the smallness of the ratio

$$\frac{\hat{H}\chi(\phi, y^\mu; h_a)}{\hat{H}\psi(h_a)} = \mathcal{O}\left(\frac{1}{M}\right) \tag{24}$$

and the adiabatic approximation also implies the smallness of the variation of the fast wave function with respect to the slow variables:

$$\frac{\delta}{\delta h_a}\chi(\phi, y^\mu; h_a) \simeq \mathcal{O}\left(\frac{1}{M}\right). \tag{25}$$

3.3. *Emerging dynamics for the gravity-matter system*

The model can now be expanded following the WKB method. At the first order $\mathcal{O}(M^1)$, it is clear that only gravitational terms survive, recreating the gravitational Hamilton-Jacobi equation

$$\frac{1}{2}\nabla_g S_0 \cdot \nabla_g S_0 + V = 0 \tag{26}$$

which ensures the classical limit of General Relativity. This is an intrinsic property of the expansion due to the choice of parameter, which selects the Planck scale.

Following the expansion, at $\mathcal{O}(M^0)$ both the gravitational and the matter sector are present. Making use of the gravitational constraints (20), (21), the equation for the matter wave function at this order (here labeled χ_0) is found:

$$\mathrm{i}\hbar\frac{\delta}{\delta\tau}\chi_0 \equiv \mathrm{i}\hbar\int_\Sigma d^3x \left(Nn^\mu + N^i b_i^\mu\right)\frac{\delta}{\delta y^\mu}\chi_0 = \hat{H}\chi_0 \tag{27}$$

where the operator $\hat{H}$ is the matter Hamiltonian up to this order, defined combining both the matter superHamiltonian and supermomentum functions:

$$\hat{H}(\bullet) = \int_\Sigma d^3x \left(N\hat{\mathcal{H}}^m + N^i\hat{\mathcal{H}}_i^m\right)(\bullet). \tag{28}$$

Eq. (27) has been rewritten as a functional Schrödinger equation through an integration over the spatial hypersurfaces and an appropriate definition of the time derivative, using the definition of the deformation vector (16). The striking difference from previous works is that, in this case, time comes from the kinematical action which plays the role of clock in the theory, instead of the dependence from the slow gravitational variables.

The next order of expansion $\mathcal{O}(\frac{1}{M})$ gives for the matter wave function χ, now enclosing the contributions up to this order, the following:

$$\mathrm{i}\hbar\frac{\delta}{\delta\tau}\chi = \hat{H}\chi + \int_\Sigma d^3x \left[N\nabla_g S_0 \cdot (-\mathrm{i}\hbar\nabla_g\chi) - 2N^k h_k \mathrm{D}\cdot\left(\frac{1}{\chi}(-\mathrm{i}\hbar\nabla_g\chi)\right)\chi\right]. \tag{29}$$

It can be shown[11] that the additional terms in Eq. (29) are unitary. This property, together with the smallness of such contributions coming from the initial adiabatic assumption (25), make this model an acceptable one for investigating quantum gravity effects on the matter dynamics.

4. Dynamics of a cosmological scalar field

Let us now consider an homogeneous Universe filled with a free massless homogeneous scalar field, and with a cosmological constant $\Lambda > 0$, a setup similar to.[16] This model reproduces the inflationary scheme during the de Sitter phase, since the kinetic term of the scalar field stands for the kinetic term of the inflaton one, which can become important towards the Planck scale, and the cosmological constant accounts for the almost constant potential during that phase. We now apply the model described in the previous section to this simple case in order to infer the effect of the corrections computed in Eq. (29) to the Hamiltonian spectrum of the scalar matter field.

The background is given by a Robertson-Walker metric:

$$ds^2 = N^2 dt^2 - a(t)^2 \left(dx^2 + dy^2 + dz^2\right) \tag{30}$$

with Ricci scalar:

$$R = \frac{6}{c^2}\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right) \tag{31}$$

where a is the cosmic scale factor. The shift vector N^i (see Eq. (16)) cannot be included in the metric since it would violate isotropy of the model.

Since we are considering the homogeneous case, choosing the ADM foliation such that the normal vector field is $n^\mu = (1, 0, 0, 0)$, it is easy to see that the only surviving contribution of the kinematical conjugate momenta is

$$\hat{p}_0 = \mathrm{i}\hbar \frac{\delta}{\delta T} \tag{32}$$

where the constructed time parameter, here labeled as T, coincides with the synchronous time[17] hence $N = 1$.

Following the steps in Sec. 3, the total wave function of the system will satisfy the gravitational superHamiltonian constraint (20) and the total superHamiltonian constraint (22). The second one, after integration over a proper portion of space, i.e. the fiducial volume V_0, gives:

$$(H_{grav} + H_\phi + H_{kin})\Psi = \left(\frac{1}{2V_0 a^3}\pi_\phi^2 - \frac{2\pi G c^2}{3V_0\, a}\pi_a^2 + \frac{V_0}{8\pi G}\,\Lambda a^3 + p_0\right)\Psi \tag{33}$$

and analogously for the gravitational constraint. The contribution of the supermomentum is canceled due to the condition $N^i = 0$.

In order to proceed with the expansion, the Einstein constant has to be rewritten in terms of the chosen parameter M defined in Eq. (2). Then, the system composed of the two constraints can be expanded order by order as presented in Sec. 3.

The gravitational constraint equation gives solutions for the functions S_0, P_1, P_2 of the gravitational wave function up to order $\mathcal{O}(1/M)$, following the separation in Eq. (13). Implementing these solutions, the total constraint gives at order $\mathcal{O}(M^0)$

the ordinary Schrödinger equation for the scalar field:

$$\mathrm{i}\hbar\frac{\delta}{\delta T}\chi = \hat{H}\chi \tag{34}$$

where we have taken $V_0 = 1$.

The corrections emerge at the next order $\mathcal{O}(1/M)$. After plugging the gravitational solutions into the total constraint equation, we switch to the time parameter τ such that $d\tau = \frac{dT}{a^3}$ and we pass to the Fourier space, so that the dynamical equation for the scalar field becomes:

$$\mathrm{i}\hbar\frac{\delta\tilde{\chi}}{\delta\tau} = -\frac{\hbar^2}{2}\frac{\delta^2\tilde{\chi}}{\delta\phi^2} + \hbar\,\frac{k_a\,(-\tau)^{7/3}}{3(3\Lambda)^{1/6}}\,\tilde{\chi}, \tag{35}$$

where the matter Hamiltonian (given by the standard kinetic term of the free field) is indeed corrected by the additional term related to k_a, which is the eigenvalue of the conjugate momentum π_a of the scale factor a.

The solution for the matter wave function is

$$\tilde{\chi} = e^{-\mathrm{i}\frac{\hbar p^2}{2}\tau + \mathrm{i}\frac{k_a\,(-\tau)^{7/3}}{7(3\Lambda)^{1/6}}}\,, \tag{36}$$

where p is the conjugate momentum with respect to the scalar field ϕ.

The solution (36) can be used to construct a wave packet and infer the magnitude of the spectrum modifications. It is important to stress that, passing from Eq. (33) to Eq. (35), the momentum k_a and hence the cosmic scale factor maintain their quantum nature: for this reason we do not insert the classical limit $a(T)$ of the background metric, but we leave a as an intrinsic quantum variable. We also remind that, due to the initial assumptions (see in particular Eq. (24)), the range of validity of the expansion is $-\frac{1}{M} < k_a < \frac{1}{M}$.

Numerical analysis of the corrected Hamiltonian spectrum has been computed by considering a Gaussian wave packet in a, ϕ constructed with the solution in Eq. (36). The results of the analysis show small corrections as expected from the hypotheses of the theory, since the quantum contributions are considered of order $\mathcal{O}(1/M)$.

5. Analogy with the frame fixing procedure

An interesting observation should be made regarding the procedure we followed. In a later paper,[14] the authors fix the reference frame of a gravity-matter system by inserting an additional term to the action, which resonates with the kinematical action insertion developed years before from one of the authors.[12] We do not show the detailed calculations here, reporting only the key points of such procedure and why it is connected to this work.

The chosen coordinates for their procedure are the Gaussian ones $g^{00} = 1$, $g^{0i} = 0$, implemented with Lagrange multipliers $\mathcal{M}$, $\mathcal{M}_i$ into the action with an

additional term, that is written in a parametrized form as:

$$S^F = \int_\Sigma d^4x \left[-\frac{1}{2}\mathcal{M}\sqrt{-g}(g^{\alpha\beta}\partial_\alpha T\partial_\beta T - 1) + \mathcal{M}_i\sqrt{-g}(g^{\alpha\beta}\partial_\alpha T\partial_\beta X^i) \right], \quad (37)$$

where X^i, T are the Gaussian coordinates and x^α are the new variables ("parameters") associated to the metric $g_{\alpha\beta}$, so that coordinate transformations on the x^α are allowed.

The authors prove that such term (37), corresponding to the choice of reference system, materializes into the theory as a fluid, more precisely as an incoherent dust with energy tensor:

$$T^{\alpha\beta} = \mathcal{M}\, U^\alpha\, U^\beta \qquad \text{with } U^\alpha = g^{\alpha\beta}\partial_\beta T \quad (38)$$

and with positive-definite energy as long as $\mathcal{M} \geq 0$.

Again, going to the order of quantum matter, a Schrödinger equation emerges with time coinciding with the parameter T of the Gaussian frame. This means that the emergent fluid plays the role of clock for quantum matter, in a way which is very similar to the kinematical action implementation.

Starting from the action with the "emergent fluid term", we can implement the same WKB expansion as explained in Sec. 3. We obtain that, in the minisuperspace, such procedure gives at the order $\mathcal{O}(1/M)$ quantum gravity corrections which are isomorphic to those of Eq. (29).

This result has an immediate explanation. In the minisuperspace, one can choose the ADM foliation such that $n^\mu = (1, \vec{0})$, consequently the parametric equations in (17) give $y^0 = T$, $y^i = X^i$; this means that their time derivative becomes

$$\partial_t y^\mu \to \partial_t y^0 = \dot{T} = \frac{\delta T}{\delta t} \quad (39)$$

It is clear that, in the minisuperspace and with a specific foliation, the kinematical action exactly reduces to the "reference fluid" emerging from the Gaussian reference frame fixing of Eq. (37). In this sense, the two implementations are equivalent.

6. Concluding remarks

The semiclassical WKB expansions of the gravity-matter system presented in the works here cited[5,8–10] encounter some problems when the canonical quantization procedure is applied and a time parameter is recovered from the dependence of the quantum system on the semiclassical variables. We have shown that the expansion performed in[5] with parameter $\hbar$ describes a non-unitary quantum matter dynamics at the next order, as seen in Eq. (11). A similar result already emerged in the work,[8] see Eq. (15), which was later modified with an *ad-hoc* procedure[10]; however the basic assumptions of that workaround cannot be satisfied in many cosmological models, thus requiring a different procedure.

Even dealing with an approach more similar to the Born-Oppenheimer one,[9] after the WKB expansion, it seems that non-unitary terms for the matter sector

dynamics do not emerge, but the theory cannot be set by means of a general Hilbert product (see discussion in[11]).

The previous results signal a necessity to define the relational time in a different way, not using the dependence from the gravitational variables, which have an intrinsic quantum component due to the WKB limit. For this reason a different construction of the gravity-matter system has been implemented, where both the WKB expansion and the adiabatic approximation have been used, but with the core difference that time has been constructed from the the kinematical action. The utility of such term, defined in Eq. (17), in quantum field theory on a given background is to recover covariance under ADM foliation, since the deformation vector is assigned due to an *a priori* choice of the reference system.

Thanks to the additional contributions (see Eqs. (18a)-(18b)) coming from the kinematical action, we have recovered a dynamics described by linear constraints in the conjugate momenta (Eqs. (22), (23)) and covariant under choice of an ADM foliation. The momenta p_μ associated to the kinematical variables y^μ, after the quantization procedure, are a strong candidate for the definition of a time parameter.

By performing the WKB expansion in the parameter M defined in (2), we have obtained at the highest order the Hamilton-Jacobi equation (26) for gravity, which is equivalent to the Einstein equations in vacuum.[15] At the next order, with the definition in Eq. (27), we have recovered at the zero-th order a functional Schrödinger equation which corresponds standard quantum field theory. Finally, at order $\mathcal{O}(1/M)$, the corrections arising from quantum gravity to the quantum matter sector dynamics have been computed, showing that the additional terms in Eq. (29) are unitary. Although the obtained corrections are small by construction, they suggest a new investigation tool to evaluate how a non-standard dynamics of quantum field theory behaves, when examining energies small with respect to the Planck scale. We have shown in Sec. 4 that such modifications give a non-zero contribution to the Hamiltonian spectrum of the scalar matter field in a cosmological setting.

We have shown in Sec. 5 the comparison between the kinematical action insertion and the reference frame fixing procedure.[14] Setting the problem in the minisuperspace and choosing a specific ADM foliation, the parametric equations giving the kinematical variables directly correspond to the Gaussian coordinates of the frame fixing procedure. Then, it is clear that the kinematical action in this case is equivalent to the choice of the reference system, that clearly emerges as a fluid in the theory. This analogy provides a starting point for a deeper understanding and future analysis on the presented model.

References

1. P. A. M. Dirac, *Lectures on Quantum Mechanics* (Dover Publications, 2001), first published 1964.
2. C. J. Isham, Canonical quantum gravity and the problem of time, *Integrable Systems, Quantum Groups, and Quantum Field Theories. NATO ASI Series (Series C: Mathematical and Physical Sciences)* **409** (1992).

3. B. S. DeWitt, Quantum theory of gravity. I. the canonical theory, *Phys. Rev.* **160**, 1113 (August 1967).
4. C. Rovelli, Time in quantum gravity: An hypothesis, *Phys. Rev. D* **43**, 442 (January 1991).
5. A. Vilenkin, Interpretation of the wave function of the universe, *Phys. Rev. D* **39**, 1116 (February 1989).
6. D. Prasad Datta, Notes on the Born-Oppenheimer approach in a closed dynamical system, *Classical and Quantum Gravity* **14**, 2825 (October 1997).
7. L. D. Landau and E. M. Lifshitz, *Quantum mechanics: Non-relativistic theory*, Course on Theoretical Physics, Vol. 3, 3 edn. (Pergamon Pr, 1981).
8. C. Kiefer and T. P. Singh, Quantum gravitational corrections to the functional Schrödinger equation, *Phys. Rev. D* **44**, 1067 (August 1991).
9. C. Bertoni, F. Finelli and G. Venturi, The Born-Oppenheimer approach to the matter-gravity system and unitarity, *Classical and Quantum Gravity* **13**, 2375 (September 1996).
10. C. Kiefer and D. Wichmann, Semiclassical approximation of the Wheeler-DeWitt equation: Arbitrary orders and the question of unitarity, *General Relativity and Gravitation* **50**, p. 66 (May 2018).
11. F. Di Gioia, G. Maniccia, G. Montani and J. Niedda, Nonunitarity problem in quantum gravity corrections to quantum field theory with Born-Oppenheimer approximation, *Phys. Rev. D* **103**, p. 103511 (May 2021).
12. K. V. Kuchař, Canonical methods of quantization, in *Oxford Conference on Quantum Gravity* (Clarendon Press, Oxford, January 1980).
13. R. L. Arnowitt, S. Deser and C. W. Misner, Republication of: The dynamics of general relativity, *General Relativity and Gravitation* **40**, 1997 (September 2008).
14. K. V. Kuchař and C. G. Torre, Gaussian reference fluid and interpretation of quantum geometrodynamics, *Phys. Rev. D* **43**, 419 (January 1991).
15. F. Cianfrani, O. M. Lecian, M. Lulli and G. Montani, *Canonical Quantum Gravity* (World Scientific, 2014).
16. D. Brizuela, C. Kiefer and M. Krämer, Quantum-gravitational effects on gauge-invariant scalar and tensor perturbations during inflation: The slow-roll approximation, *Phys. Rev. D* **94**, p. 123527 (December 2016).
17. G. Montani, Canonical quantization of gravity without "frozen formalism", *Nuclear Physics B* **634**, p. 370–392 (Jul 2002).

Natural evidence for fuzzy sphere noncommutative geometry: Super-Chandrasekhar white dwarfs

Surajit Kalita
Department of Physics, Indian Institute of Science, Bangalore-560012, India
E-mail: surajitk@iisc.ac.in

T. R. Govindarajan
The Institute of Mathematical Sciences, Chennai-600113, India
E-mail: trg@imsc.res.in

Banibrata Mukhopadhyay
Department of Physics, Indian Institute of Science, Bangalore-560012, India
E-mail: bm@iisc.ac.in

Noncommutative geometry is one of the quantum gravity theories, which various researchers have been using to describe different physical and astrophysical systems. However, so far, no direct observations can justify its existence, and this theory remains a hypothesis. On the other hand, over the past two decades, more than a dozen overluminous type Ia supernovae have been observed, which indirectly predict that they originate from white dwarfs with super-Chandrasekhar masses $2.1 - 2.8\,\mathrm{M}_{\odot}$. In this article, we discuss that considering white dwarfs as squashed fuzzy spheres, a class of noncommutative geometry, helps in accumulating more mass than the Chandrasekhar mass-limit. The length-scale beyond which the effect of noncommutativity becomes prominent is an emergent phenomenon, which depends only on the inter-electron separations in the white dwarf.

Keywords: Noncommutative geometry, white dwarfs, type Ia supernova, mass-limit.

1. Introduction

Noncommutative geometry, a theory of quantum gravity, has been used for many decades to explain physical systems. Riemann quoted in 1854 that "*... it seems that empirical notions on which the metrical determinations of space are founded, the notion of a solid body and a ray of light cease to be valid for the infinitely small. We are therefore quite at liberty to suppose that the metric relations of space in the infinitely small do not conform to hypotheses of geometry; and we ought in fact to suppose it, if we can thereby obtain a simpler explanation of phenomena ...*". In 1930s, Bronstein, using Heisenberg uncertainty principle and the features of Einstein's general relativity (GR), first argued that gravitational dynamics does not allow to measure arbitrarily small spacetime distances, concluding that the notions of classical Riemmanian geometry should be duly modified. The formalism of a fuzzy sphere, first introduced in 1992,[1] was later modified to show equivalence between noncommutativity (NC) and Landau levels in the presence of magnetic fields.[2] Moreover, NC alters the spacetime metric,[3] which results in shifting the

event horizon and removing the essential singularity at the center of a black hole.[4] etc. Unfortunately, there is no direct way to confirm the existence of NC like most of the other quantum gravity formalism, which compels it to remain a hypothesis.

In astrophysics, a white dwarf (WD) is the end state of a star with mass $\lesssim (10 \pm 2)\,\mathrm{M}_\odot$. Chandrasekhar showed that the maximum mass of a carbon-oxygen non-magnetized and non-rotating WD is about $1.4\,\mathrm{M}_\odot$.[5] Above this mass-limit, they burst out to produce type Ia supernovae (SNe Ia) with nearly similar luminosities. However, observations of various peculiar over-luminous SNe Ia, such as SN 2003fg, SN 2006gz, SN 2007if, SN 2009dc, and many more, suggested that they were originated from super-Chandrasekhar WDs with mass $2.1-2.8\,\mathrm{M}_\odot$.[6,7] Over the years, different models including rotation, magnetic fields, modified gravity, ungravity effects, generalized Heisenberg uncertainty principle, to name a few, have been proposed to explain these massive WDs. However, since these super-Chandrasekhar WDs have not been observed directly, no one can single out the correct theory so far.

This article shows that the squashed fuzzy sphere algebra can quantize the energy dispersion relation, which further alters the equation of state (EoS) of the degenerate electrons present in a WD, resulting in the increase in mass of the WD beyond the Chandrasekhar mass-limit. It also discusses that the length-scale below which NC is prominent depends on the inter-electron separation. Hence, the inference of super-Chandrasekhar WDs can indirectly predicts the existence of NC in high-density regimes.

2. Squashed fuzzy sphere and equation of state

The formalism of a fuzzy sphere is analogous to the angular momentum algebra in ordinary quantum mechanics, where the position coordinates mimic the angular momentum variables. In an N-dimensional irreducible representation of $\mathtt{SU}(2)$ group, the commutation relation for the coordinates of fuzzy sphere, X_i (i=1,2,3), is given by[1]

$$[X_i, X_j] = i\frac{k\hbar}{r}\epsilon_{ijk}X_k, \tag{1}$$

where $k = 2r^2/\hbar\sqrt{N^2-1}$, $\hbar = h/2\pi$, h is the Planck constant and r the radius of the fuzzy sphere.

A squashed fuzzy sphere is a simple geometrical construction obtained from the fuzzy sphere by projecting all the points to an equatorial plane as shown in Figure 1. The equatorial plane can be in any direction, and for illustration, we show the $X_1 - X_2$ plane. Using the above relation, the commutation relation for this squashed fuzzy sphere is given by[2]

$$[X_1, X_2] = \pm i\frac{k\hbar}{r}\sqrt{r^2 - X_1^2 - X_2^2}. \tag{2}$$

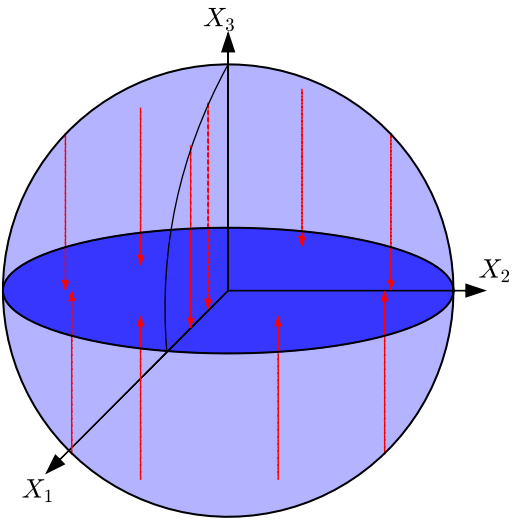


Fig. 1. An illustrative diagram of a squashed fuzzy sphere which is obtained by projecting all the points of the fuzzy sphere on an equatorial plane.

Note that NC vanishes at the surface of the sphere. The square of the quantized energies in a squashed fuzzy sphere, obtained from the Dirac operator, is given by[8]

$$E^2_{l,m} = \frac{2\hbar c^2}{k\sqrt{N^2-1}}\left\{l(l+1) - m(m\pm 1)\right\}, \tag{3}$$

where c is the speed of light and l, m are the quantum numbers with l taking all the integer values from 0 to $N-1$ and m taking all the integer values from $-l$ to l. Moreover, the relations between the Cartesian and spherical polar coordinates are $X_1 = r\sin\theta\cos\phi$ and $X_2 = r\sin\theta\sin\phi$. Therefore, from Equation (2), the squashed fuzzy sphere algebra in spherical polar coordinates is given by

$$[\sin\theta\cos\phi, \sin\theta\sin\phi] = \pm i\frac{k\hbar}{r^2}\cos\theta. \tag{4}$$

A squashed fuzzy sphere is such that it actually provides a NC between its azimuthal and polar coordinates at the surface of the fuzzy sphere. This is because the squashed plane in a fuzzy sphere can be any of its equatorial planes and there is no particular direction for it, which means that the squashed fuzzy sphere has a rotational symmetry about the equatorial plane. There is no NC along the r-direction and an electron with mass m_e does not experience NC along the radial direction. Defining p_r to be the momentum of the electron in r-direction, considering Equation (3), we obtain the total energy dispersion relation in the squashed fuzzy sphere as[8]

$$E^2 = p_r^2c^2 + m_\mathrm{e}^2c^4\left[1 + \left\{l(l+1) - m(m\pm 1)\right\}\frac{2\hbar}{m_\mathrm{e}^2c^2k\sqrt{N^2-1}}\right]. \tag{5}$$

In the large N limit, this expression reduces to[8]

$$E^2 = p_r^2 c^2 + m_{\rm e}^2 c^4 \left(1 + 2\nu \frac{2\hbar}{m_{\rm e}^2 c^2 k}\right), \qquad \nu \in \mathbb{Z}^{0+}. \tag{6}$$

It is evident that $1/k$ behaves as the strength of NC. Note that a similar dispersion relation occurs in the case of a planar NC.[9] We can assume a series of concentric squashed fuzzy spheres, such that at the surface of each spheres, the above relation is valid. Nevertheless, $k \propto r^2$ and hence $1/k$ decreases from center to the surface. Now, applying simple statistical mechanics techniques, we obtain the EoS for the degenerate electrons present in a WD, given by[8,10]

$$\text{pressure} \quad \mathcal{P} = \frac{2\rho^{2/3}}{\xi h \mu_{\rm e}^{2/3} m_{\rm p}^{2/3}} \sum_{\nu=0}^{\nu_{\max}} g_\nu \left\{ p_{\rm F} E_{\rm F} - \left(m_{\rm e}^2 c^3 + 2\nu \frac{2\hbar c}{k}\right) \ln\left(\frac{E_{\rm F} + p_{\rm F} c}{\sqrt{m_{\rm e}^2 c^4 + 2\nu \frac{hc^2}{\pi k}}}\right)\right\}, \tag{7}$$

$$\text{density} \quad \rho = \frac{64 \mu_{\rm e} m_{\rm p} \left(2\nu_{\max} + 1\right)^3 p_{\rm F}^3}{\xi^3 h^3}, \tag{8}$$

where $\xi = kh\left(\rho/\mu_{\rm e} m_{\rm p}\right)^3$, $m_{\rm p}$ is the mass of the proton, $\mu_{\rm e}$ is the mean molecular weight per electron, g_ν is the degeneracy factor, and $p_{\rm F}$ and $E_{\rm F}$ are respectively the Fermi momentum and Fermi energy of the electron gas. ξ needs to be chosen appropriately such that at the low-density limit, EoS is same as the Chandrasekhar's EoS, where NC is almost negligible. Figure 2 shows EoS for the degenerate electrons when they occupy different energy levels. It is evident that if the electrons occupy more levels, it tends to the Chandrasekhar's EoS and NC is not prominent in that density regime.

3. Stellar structure model

To obtain the stellar structure of any compact object in GR, one needs to solve the Tolman-Oppenheimer-Volkoff or TOV equations along-with EoS of the constituent particles. The TOV equations for a non-rotating and non-magnetized star are given by[11]

$$\begin{aligned} \frac{{\rm d}\mathcal{M}}{{\rm d}r} &= 4\pi r^2 \rho, \\ \frac{{\rm d}\mathcal{P}}{{\rm d}r} &= -\frac{G}{r^2}\left(\rho + \frac{\mathcal{P}}{c^2}\right)\left(\mathcal{M} + \frac{4\pi r^3 \mathcal{P}}{c^2}\right)\left(1 - \frac{2G\mathcal{M}}{c^2 r}\right)^{-1}, \end{aligned} \tag{9}$$

where $\mathcal{M}$ is the mass of the star inside a volume of radius r and G is Newton's gravitational constant. Note that NC affects the microscopic physics, while TOV equations describe the pressure and mass balances, which are macroscopic physics. Hence, in a semi-classical approach, TOV equations remain unchanged to that for the classical GR formalism. Figure 3 shows the mass–radius relation for the WDs

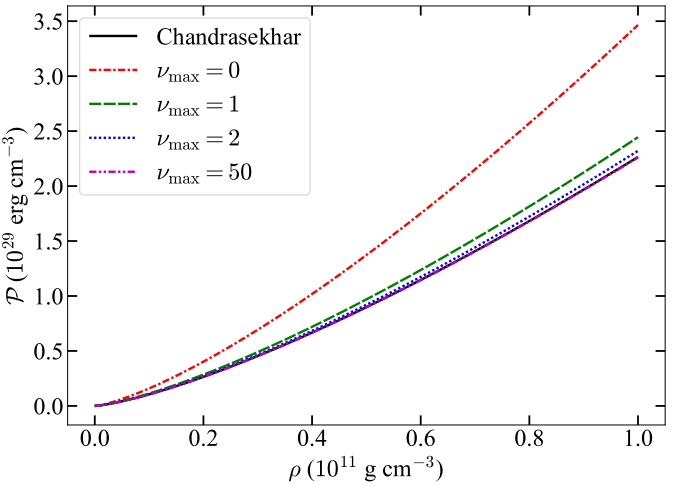


Fig. 2. EoS for degenerate electrons in a squashed fuzzy sphere with various $\nu_{\max}$ along with the Chandrasekhar's EoS. $\nu_{\max} = 0$ means electrons occupy only the ground energy level, $\nu_{\max} = 1$ means they occupy both ground and first energy levels, and so on. Chandrasekhar's EoS corresponds to $\nu_{\max} \to \infty$.

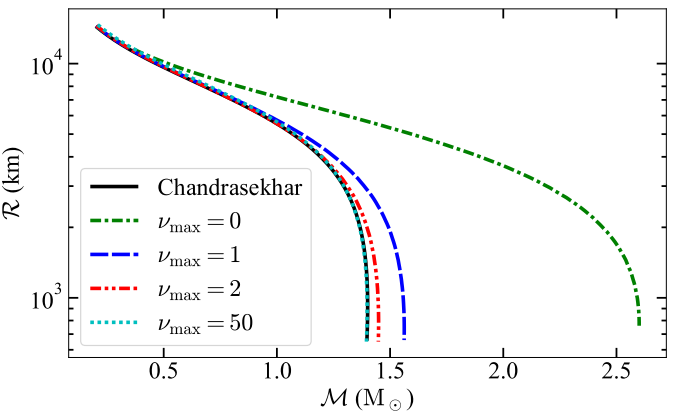


Fig. 3. Mass–radius relations of the WDs in the presence of NC when various energy levels are occupied.

in the presence of NC for various $\nu_{\max}$ along with the Chandrasekhar's mass–radius relation, which corresponds to $\nu_{\max} \to \infty$. It is evident from the figure that a non-rotating and non-magnetized WD can have a limiting mass of about $2.6\,M_\odot$ if all the electrons occupy the ground level. The mass-limit decreases when more energy levels are filled, and finally coincides with the Chandrasekhar's original mass–radius curve when many levels are occupied furnishing a limiting mass of about $1.4\,M_\odot$.

3.1. *Length scale in noncommutative geometry*

It is understood from Equation (6) that NC is prominent if $2\hbar \gtrsim m_e^2 c^2 k$, which implies that the system length-scale $\mathcal{L} \lesssim \mathcal{L}_{\rm eff} = \lambda_e/\sqrt{\pi\xi}$ with λ_e being the Compton wavelength of the electron.[8] In the case of degenerate electrons, $\mathcal{L}$ is the inter-electron separations. It is evident from this expression that for the prominence of NC, $\mathcal{L}$ does not need to be the Planck scale $\mathcal{L}_P$ as opposed to the general belief. Rather it depends upon the system's length-scale. Indeed Salecker and Wigner showed that the new uncertainty in length-scale for a system has to be $\delta \sim \left(\mathcal{L}\mathcal{L}_P^2\right)^{1/3}$, and one needs to consider δ as the quantum measurement of length.[12] Hence, for a WD, one can expect to observe a significant NC effect depending on the inter-electron separation, even though it is far from the Planck scale. In extremely high densities with inter-electron distance much less than the Compton wavelength, one can expect certain features of NC becomes prominent. One such feature is the notion of statistics, which needs localization of particles, gets distorted and can affect physical systems, such as the mass-limit of the WD.

4. Conclusions

We have demonstrated that NC may have a prominent effect in a system whose length-scale is greater than $\mathcal{L}_P$. Whether NC has a significant effect or not depends upon the system's length-scale, and we show that it is an emergent phenomenon. We have used the formalism of a squashed fuzzy sphere where NC is at the surface of the sphere, i.e., between the azimuthal and polar coordinates. We have obtained the energy dispersion relation for this formalism and found that it is very similar to that for the Landau levels in the presence of magnetic fields. Hence, one may assume that NC mimics the internally generated magnetic field. It has a significant effect in a WD if the inter-electron separation is nearly less than one-tenth of the Compton wavelength of electrons. Such prominent NC may provide a super-Chandrasekhar limiting mass $2.6\,M_\odot$ and it can explain the formation of peculiar over-luminous SNe Ia. In the future, if gravitational wave detectors detect these massive WDs directly, it can confirm the existence of NC more firmly.[13]

References

1. J. Madore, The fuzzy sphere, *Classical and Quantum Gravity* **9**, 69 (Jan 1992).
2. S. Andronache and H. C. Steinacker, The squashed fuzzy sphere, fuzzy strings and the Landau problem, *Journal of Physics A Mathematical General* **48**, p. 295401 (Jul 2015).
3. V. P. Nair, *Noncommutative Mechanics, Landau Levels, Twistors, and Yang-Mills Amplitudes*, in *Lecture Notes in Physics, Berlin Springer Verlag*, ed. S. Bellucci 2006, p. 97.
4. R. Kumar and S. G. Ghosh, Accretion onto a noncommutative geometry inspired black hole, *European Physical Journal C* **77**, p. 577 (Sep 2017).
5. S. Chandrasekhar, The Maximum Mass of Ideal White Dwarfs, *ApJ* **74**, p. 81 (July 1931).

6. D. A. Howell, M. Sullivan, P. E. Nugent, R. S. Ellis, A. J. Conley, D. Le Borgne, R. G. Carlberg, J. Guy, D. Balam, S. Basa, D. Fouchez, I. M. Hook, E. Y. Hsiao, J. D. Neill, R. Pain, K. M. Perrett and C. J. Pritchet, The type Ia supernova SNLS-03D3bb from a super-Chandrasekhar-mass white dwarf star, *Nature* **443**, 308 (September 2006).
7. R. A. Scalzo, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, C. Buton, M. Childress, N. Chotard, Y. Copin, H. K. Fakhouri, A. Gal-Yam, E. Gangler, S. Hoyer, M. Kasliwal, S. Loken, P. Nugent, R. Pain, E. Pécontal, R. Pereira, S. Perlmutter, D. Rabinowitz, A. Rau, G. Rigaudier, K. Runge, G. Smadja, C. Tao, R. C. Thomas, B. Weaver and C. Wu, Nearby Supernova Factory Observations of SN 2007if: First Total Mass Measurement of a Super-Chandrasekhar-Mass Progenitor, *ApJ* **713**, 1073 (April 2010).
8. S. Kalita, T. R. Govindarajan and B. Mukhopadhyay, Super-Chandrasekhar limiting mass white dwarfs as emergent phenomena of noncommutative squashed fuzzy spheres, *International Journal of Modern Physics D*, (in press) (November 2021).
9. S. Kalita, B. Mukhopadhyay and T. R. Govindarajan, Significantly super-Chandrasekhar mass-limit of white dwarfs in noncommutative geometry, *International Journal of Modern Physics D* **30**, p. 2150034 (April 2021).
10. D. Lai and S. L. Shapiro, Cold Equation of State in a Strong Magnetic Field: Effects of Inverse beta -Decay, *ApJ* **383**, p. 745 (December 1991).
11. L. Ryder, *Introduction to General Relativity* (Cambridge University Press, 2009).
12. H. Salecker and E. P. Wigner, Quantum Limitations of the Measurement of Space-Time Distances, *Physical Review* **109**, 571 (Jan 1958).
13. S. Kalita, B. Mukhopadhyay, T. Mondal and T. Bulik, Timescales for Detection of Super-Chandrasekhar White Dwarfs by Gravitational-wave Astronomy, *ApJ* **896**, p. 69 (June 2020).

A model of polymer gravitational waves: Theory and some possible observational consequences

Angel Garcia-Chung

Departamento de Física, Universidad Autónoma Metropolitana - Iztapalapa, San Rafael Atlixco 186, Ciudad de México 09340, México
E-mail: alechung@xanum.uam.mx

James B. Mertens

Department of Physics and McDonnell Center for the Space Sciences, Washington University, St. Louis, MO 63130, USA
E-mail: jmertens@wustl.edu

Saeed Rastgoo

Department of Physics and Astronomy, York University 4700 Keele Street, Toronto, Ontario M3J 1P3 Canada
E-mail: srastgoo@yorku.ca

Yaser Tavakoli

School of Astronomy, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5531, Tehran, Iran
E-mail: yaser.tavakoli@guilan.ac.ir

Paulo Vargas Moniz

Departamento de Fisica, Centro de Matematica e Aplicações: CMA-UBI, Universidade da Beira Interior, 6200 Covilhã, Portugal
E-mail: pmoniz@ubi.pt

We propose a polymer quantization scheme to derive the effective propagation of gravitational waves on a classical Friedmann-Lemaitre-Robertson-Walker (FLRW) spacetime. These waves, which may originate from a high energy source, are a consequence of the dynamics of the gravitational field in a linearized low-energy regime. A novel method of deriving the effective Hamiltonian of the system is applied to overcome the challenge of polymer quantizing a time-dependent Hamiltonian. Using such a Hamiltonian, we derive the effective equations of motion and show that (i) the form of the waves is modified, (ii) the speed of the waves depends on their frequencies, and (iii) quantum effects become more apparent as waves traverse longer distances.

Keywords: Quantum gravity, gravitational waves

1. Introduction

Recent discovery of gravitational waves (GWs) and the rapid increase in the sensitivity of GWs observatories has opened up a great opportunity in connecting theory and phenomenology with experiment in precision cosmology, black hole

physics and quantum gravity among other fields. In particular we are now becoming more hopeful about the observation of signatures of quantum gravity in GWs emitted from black hole mergers and high redshift regions of the cosmos.

There have been numerous studies connecting theories of quantum gravity with potential observations regarding the structure of quantum spacetime. In particular, in Loop Quantum Gravity (LQG),[1] there have been studies to understand the consequence of nonpertubative quantization in propagation of Gamma Ray Bursts (GRBs), other matter fields, and GWs on cosmological or black holes spacetimes (for some examples see, Refs. [2–31] and references within).

In this work we consider GWs as effective perturbations propagating on a classical FLRW cosmological spacetime. The effective form of such waves is derived by applying the techniques of polymer quantization[32–36] to the classical perturbations. Such a quantization is a representation of the classical algebra on a Hilbert space that is unitarily inequivalent to the usual Schrödinger representation. In it, operators are regularized and written in a certain exponential form. In such theories, the infinitesimal generators corresponding to some of the operators do not exist on the Hilbert space. As a consequence, the conjugate variables to those operators only admit finite transformations. Thus, the dynamics of the theory leads to the discretization of the spectrum of the conjugate operators (for more details and some examples of polymer quantization applied to particles and path integral formulation of black holes, see Refs. [34, 35, 37–39]).

In the model we present in this paper, the Hamiltonian is time-dependent and directly polymer quantizing it proves to be quite challenging. Hence, we apply a novel method based on the use of extended phase space, to overcome this issue (see Ref. [40]). The Hamiltonian in the extended phase space is rendered time independent by applying a certain canonical transformation and then polymer quantized using some of techniques developed in the literature.[35,41,42] After that, the effective version of this quantum Hamiltonian is made time-dependent again by applying the inverse of the above-mentioned canonical transformation. Finally the system is re-expressed in the usual non-extended phase space. Using this modified Hamiltonian, we derive the effective equations of motion of polymerized GWs and show that i) the form of the waves is modified, ii) the speed of the waves depends on their frequencies, and iii) the quantum effects are amplified by the distance/time the waves travel.

This paper is organized as follows: in Sec. 2, we derive the classical Hamiltonian of perturbations on an FLRW classical background. In Sec. 3, this time-dependent Hamiltonian is turned into a polymer effective time-dependent Hamiltonian by applying a certain method that is inspired by an approach used to deal with time-dependent Harmonic oscillators. Using this Hamiltonian, we then derive the effective equations of motion. In Sec. 4, we study the behavior of the solutions in both nonperturbative and perturbative regimes and show that quantum gravitational effects induce certain imprints on the waveform, frequency, and speed of GWs. Finally, in Sec. 5 we present our concluding remarks.

2. Hamiltonian formalism for GWs

We start with a spacetime manifold $M = \mathbb{T}^3 \times \mathbb{R}$ with a spatial 3-torus topology[a], equipped with coordinates $x^j \in (0, \ell)$ and a temporal coordinate $x^0 \in \mathbb{R}$. The background metric $\mathring{g}_{\mu\nu}$ is then perturbed by a small perturbation $h_{\mu\nu}$ such that the full metric $g_{\mu\nu}$ can be written as

$$g_{\mu\nu} = \mathring{g}_{\mu\nu} + h_{\mu\nu}. \tag{1}$$

GWs are the result of the weak-field approximation to the Einstein field equations for the above metric. As is well-known, a wave traveling along, say, the x^3 direction, can be separated into two polarization scalar modes $h_+(x)$ and $h_\times(x)$ as

$$h_{ij}(x) \;=\; h_+(x)e^+_{ij} + h_\times(x)e^\times_{ij}\,, \tag{2}$$

where

$$e^+ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{and} \qquad e^\times = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{3}$$

We would like to study the dynamics of the perturbations on a homogeneous, isotropic universe described by the FLRW metric

$$\mathring{g}_{\mu\nu}dx^\mu dx^\nu = -N^2(x^0)\, d(x^0)^2 + a^2(x^0)d\mathbf{x}^2, \tag{4}$$

where x^0 is an arbitrary time coordinate, $N(x^0)$ is the lapse function which depends on the choice of x^0, and $d\mathbf{x}^2 = \sum_i^3 d(x^i)^2$ is a unit 3-sphere. To obtain a Hamiltonian which resembles the Hamiltonian of a Harmonic oscillator, we introduce a new field

$$h_\sigma\left(x^0, \mathbf{x}\right) = \frac{a^2\sqrt{\kappa}}{\ell^{3/2}} \sum_{\mathbf{k}} \mathcal{A}_{\sigma,\mathbf{k}}\left(x^0\right) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{5}$$

which together with its conjugate momentum $\mathcal{E}_{\sigma,\mathbf{k}}$ constitute a canonically conjugate pair

$$\left\{\mathcal{A}_{\lambda,\mathbf{k}}, \mathcal{E}_{\lambda',\mathbf{k}'}\right\} = \delta_{\mathbf{k}\mathbf{k}'}\delta_{\lambda\lambda'}. \tag{6}$$

Here ℓ is the result of our quantization on a lattice which corresponds to an upper limit on the momenta involved. The above canonical pair allow us to write the Hamiltonian of the system as

$$H = \frac{N}{2a^3} \sum_{\sigma=+,\times} \sum_{\mathbf{k}} \left[\mathcal{E}^2_{\sigma,\mathbf{k}} + k^2a^4\mathcal{A}^2_{\sigma,\mathbf{k}}\right] =: \sum_{\sigma=+,\times} \sum_{\mathbf{k}} H_{\sigma,\mathbf{k}}\left(x^0\right), \tag{7}$$

where N is the lapse. It is clear that this last equation represents the Hamiltonian of a set of decoupled harmonic oscillators with a time-dependent frequency $\omega^2 = k^2a^4$, due to time dependence of $a(t)$.

At this point, we choose the harmonic time gauge where $N(x^0 = \tau) = a^3(\tau)$ to get rid of the factor a^{-3} in front of Eq. (7). Hence, the Hamiltonian of the

[a]To avoid a discussion of boundary conditions on fields (generated by perturbations), we will assume that the spatial 3-manifold is $\mathbb{T}^3$.

perturbations (for a fixed mode $\mathbf{k}$ and a fixed polarization σ) over the FLRW background in harmonic time becomes

$$H_{\sigma,\mathbf{k}}(\tau) = \frac{1}{2}\left[\mathcal{E}^2_{\sigma,\mathbf{k}} + k^2 a^4 \mathcal{A}^2_{\sigma,\mathbf{k}}\right]. \tag{8}$$

3. Polymer quantization and the effective Hamiltonian

The time-dependence of this Hamiltonian (8) makes deriving its effective polymer corrections quite complicated. This is because its polymer quantization will yield a time-dependent quantum pendulum-type system whose solutions are mathematically difficult to treat. In order to circumvent this issue, we will apply a procedure based on the extended phase space formalism (more details in Ref. [40]). The idea is to first lift the (action of the) system

$$S = \int \left\{ p\frac{dq}{dt} - H(t) \right\} dt, \tag{9}$$

with time-dependent Hamiltonian of the form

$$H(t) = \frac{1}{2m}p^2 + \frac{1}{2}m\omega(t)^2 q^2, \tag{10}$$

to the extended phase space (EPS). In accordance with Dirac's formalism, the system is now described by the extended action

$$S = \int \left\{ p\frac{dq}{d\tau} + p_t\frac{dt}{d\tau} - \lambda\phi \right\} d\tau, \tag{11}$$

where

$$\phi = p_t + H(t) \approx 0, \tag{12}$$

is a first class constraint ensuring the compatibility of the two actions (9) and (11) on the constrained surface $\phi = 0$, λ is a Lagrange multiplier, and p_t is the momentum conjugate to t. This is step (1) in Fig. 1. In step (2) in Fig. 1, we apply a canonical transformation

$$Q = \frac{1}{\rho(t)}q, \tag{13}$$

$$T = \int \frac{1}{\rho^2(t)}\, dt, \tag{14}$$

$$P = \rho(t)p - m\dot{\rho}(t)q, \tag{15}$$

$$P_T = \rho^2(t)p_t + \rho(t)\dot{\rho}(t)\, q\, p - \frac{m}{2}q^2\left[\dot{\rho}^2(t) + \frac{W^2}{\rho^2(t)} - \omega^2(t)\rho^2(t)\right], \tag{16}$$

in the extended phase space which removes the time dependency of the Hamiltonian $H(t)$in ϕ. Here, W is the time-independent frequency of the time-independent

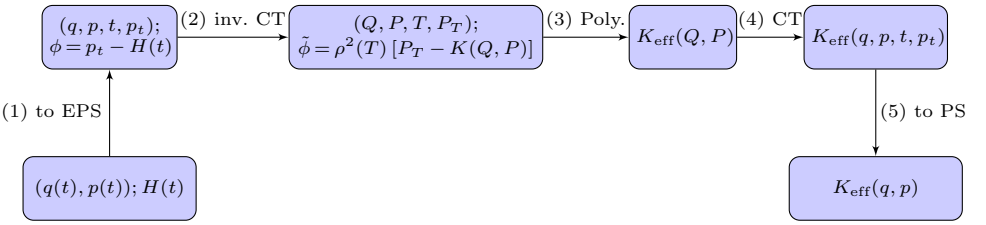


Fig. 1. Schematics of the derivation of a time-dependent effective Hamiltonian constraint. Here "EPS" means extended phase space, "inv. CT" denotes inverse canonical transformation, "Poly." means the process of polymer quantization and getting an effective polymer Hamiltonian from there, "CT" denotes the canonical transformation, and "PS" means the nonextended phase space. The lower row corresponds to the usual phase space, while the upper row corresponds to the extended phase space.

system and ρ is an auxiliary variable to be determined by the specific properties of the system, more precisely by ω and W. After this step the action becomes

$$S = \int \left\{ P\frac{dQ}{d\tau} + P_T\frac{dT}{d\tau} - \lambda\tilde{\phi} \right\} d\tau, \tag{17}$$

where, the first class constraint now reads

$$\tilde{\phi} = \rho^2(T)\,[P_T + K] \approx 0, \tag{18}$$

and the corresponding Hamiltonian K appearing in it is

$$K = \frac{1}{2m}P^2 + \frac{1}{2}mW^2Q^2. \tag{19}$$

Moreover, the auxiliary equation used to fix $\rho(t)$ becomes

$$\ddot{\rho}(t) + \omega^2(t)\rho(t) = \frac{W^2}{\rho^3(t)}. \tag{20}$$

This time-independent harmonic oscillator can now be polymer quantized[35,41,42] as in step (3) of Fig. 1. We then perform the inverse canonical transformations above in step (4) of Fig. 1. Finally, in step (5) of Fig. 1, we apply the inverse of the canonical transformations above and solve the constraint. This yields the polymer effective Hamiltonian on the usual phase space, where now the Hamiltonian is not just effective but also time-dependent.

By applying this method to our Hamiltonian (7), we obtain an effective polymer Hamiltonian with polymer $\mathcal{E}_{\sigma,\mathbf{k}}$ and discrete spectra for $\mathcal{A}_{\sigma,\mathbf{k}}$ as

$$H_{\rm eff}^{(\mathcal{E})} = \sum_{\sigma=+,\times}\sum_{\mathbf{k}\in\mathscr{L}} \left\{ \frac{2}{\mu^2\rho^2}\sin^2\left(\frac{\mu\left(\rho\mathcal{E}_{\sigma,\mathbf{k}} - \dot{\rho}\mathcal{A}_{\sigma,\mathbf{k}}\right)}{2}\right) + \frac{\dot{\rho}\mathcal{A}_{\sigma,\mathbf{k}}\,\mathcal{E}_{\sigma,\mathbf{k}}}{\rho} + \frac{\mathcal{A}_{\sigma,\mathbf{k}}^2}{2}\left[\omega^2 - \frac{\dot{\rho}^2}{\rho^2}\right] \right\}, \tag{21}$$

where μ is called the polymer parameter that sets the scale for which the quantum gravity effects become important, and we have set $\hbar = 1$. The corresponding

equations of motion read

$$\frac{d\mathcal{A}_{\sigma,\mathbf{k}}}{dt} = \frac{1}{\rho}\frac{\sin\left(\mu\left(\rho\mathcal{E}_{\sigma,\mathbf{k}} - \dot{\rho}\mathcal{A}_{\sigma,\mathbf{k}}\right)\right)}{\mu} + \frac{\dot{\rho}}{\rho}\mathcal{A}_{\sigma,\mathbf{k}}, \tag{22}$$

$$\frac{d\mathcal{E}_{\sigma,\mathbf{k}}}{dt} = \frac{\dot{\rho}}{\rho^2}\frac{\sin\left(\mu\left(\rho\mathcal{E}_{\sigma,\mathbf{k}} - \dot{\rho}\mathcal{A}_{\sigma,\mathbf{k}}\right)\right)}{\mu} + \left(\frac{\dot{\rho}}{\rho}\right)^2\mathcal{A}_{\sigma,\mathbf{k}} - \omega^2\mathcal{A}_{\sigma,\mathbf{k}} - \frac{\dot{\rho}}{\rho}\mathcal{E}_{\sigma,\mathbf{k}}. \tag{23}$$

These equations are nonlinear in both $\mathcal{A}_{\sigma,\mathbf{k}}$ and $\mathcal{E}_{\sigma,\mathbf{k}}$, and their $\mu \to 0$ limit matches the classical equations of motion as expected. Also notice that in our case, ρ controls the background geometry such that a time-dependent ρcorresponds to a time-dependent background geometry and a constant ρ corresponds to the flat spacetime.

4. Perturbative and nonperturbative numerical solutions

We will now solve Eqs. (22)–(23) for specific field-space configurations, both perturbatively, and numerically and nonperturbatively for both time-dependent and time-independent backgrounds.

4.1. *Time-independent background*

For a time-independent background, for which $\rho = 1$ and $\dot{\rho} = \ddot{\rho} = 0$, we can obtain full nonpertubative numerical solutions for $\mathcal{A}(t)$ as seen in Fig. 2. We can also obtain a perturbative solution

$$\begin{aligned}\mathcal{A}(t) \simeq\ & \mathcal{E}_I \sin\left[(1-(\mathcal{E}_I k\mu)^2/16)kt\right] \\ & - \frac{\mathcal{E}_I^3 k^2\mu^2}{16}\sin^2\left[(1-(\mathcal{E}_I k\mu)^2/16)kt\right]\cos\left[(1-(\mathcal{E}_I k\mu)^2/16)kt\right]\,.\end{aligned} \tag{24}$$

This solution can exhibits a frequency shift of order μ^2, and a cubic correction term. The cubic term can also be rewritten, and thought of, as an introduction of higher harmonics using angle identities. In observations, the frequency shift may be more important to account for than the excited harmonics. This is because the frequency shift can manifest as a phase shift that has considerable time to develop as the wave traverses cosmological distances. In Fig. 2 we demonstrate this, comparing the perturbative solution to the exact and classical ones for the time-independent case.

We can also analyze the above perturbative solutions and obtain some insight into the speed of propagation of the waves. For that, we note that the dominant contributions to Eq. (24) can be written as

$$\mathcal{A}(t) \simeq \mathcal{E}_I \sin\left[\left(1-\left(\frac{\mathcal{E}_I k\mu}{4}\right)^2\right)kt\right]. \tag{25}$$

Comparing with the classical solution where we identify $ka^2 = \omega_c$, with ω_c being the classical angular speed, we notice that up to first order the polymer angular

speed is

$$\omega_{\mu}^{(\mathcal{E})} \simeq \omega_c \left[1 - k^2 \left(\frac{\mathcal{E}_I \mu}{4}\right)^2\right]. \tag{26}$$

Although this is a perturbative and approximate result and even though we have neglected higher harmonics in Eq. (24), the above equation reveal a curious phenomenon. Noting that $\omega_c = ka^2$ and with the group velocity being

$$v = \frac{d\omega_{\text{poly}}}{d\,(ka^2)} \tag{27}$$

with ω_{poly} being either $\omega_{\nu}^{(\mathcal{A})}$ or $\omega_{\nu}^{(\mathcal{E})}$, we obtain

$$v_{\mu}^{(\mathcal{E})} \simeq 1 - k^2 \left(\frac{\mathcal{E}_I \mu}{4}\right)^2. \tag{28}$$

where $v_{\mu}^{(\mathcal{E})}$ is the velocity of the effective waves. One can see from Eq. (28) that the group velocity of the waves is slower than the speed of light by a factor of $k^2 \left(\frac{\mathcal{E}_I \mu}{4}\right)^2$, and it also depends on the initial momentum $\mathcal{E}_I$ of the waves and the polymer parameter μ due to the factor $k^2 \left(\frac{\mathcal{E}_I \mu}{4}\right)^2$. More importantly, the deviation from the speed of light also depends on the modes k. Hence, waves with larger k (i.e., larger energies) have a lower speed compared to the ones with smaller k and are more affected by the quantum structure of spacetime. Also, notice that this case leads to the violation of Lorentz symmetry as can be seen by squaring both sides of Eq. (26). Of course, due to the sheer smallness of the expected value of μ, and the appearance of their squares in the above expressions, these effects are very small, but a highly energetic phenomenon with a large $\mathcal{E}_I$ may help to amplify it to an extent that future observatories can detect it. We should emphasize that the presence of the violation of the Lorentz symmetry in this case, as seen from the above results, is a consequence of the polymer quantization and, in particular, this model, and is not a direct consequence of LQG.

4.2. *Time-independent background*

For the case of a time-dependent background, we can obtain a solution in one of two ways: directly integrating the EOMs, or using the canonical transformation in Eqs. (13)–(16). In either case, we will need to obtain a solution for ρ by solving Eq. (20). In general, this choice determines whether the mode amplitude will be purely decaying or will contain oscillatory behavior. Here we will seek purely growing solutions for ρ, choosing initial conditions such that oscillatory behavior is minimized; in our case, simply choosing $\rho = 1$ and $\dot{\rho} = 0$ is sufficient. Choosing a different initial amplitude for ρ is in any case equivalent to rescaling of the scale factor a, polymer scale, momentum, and time coordinate. The full nonperturbative

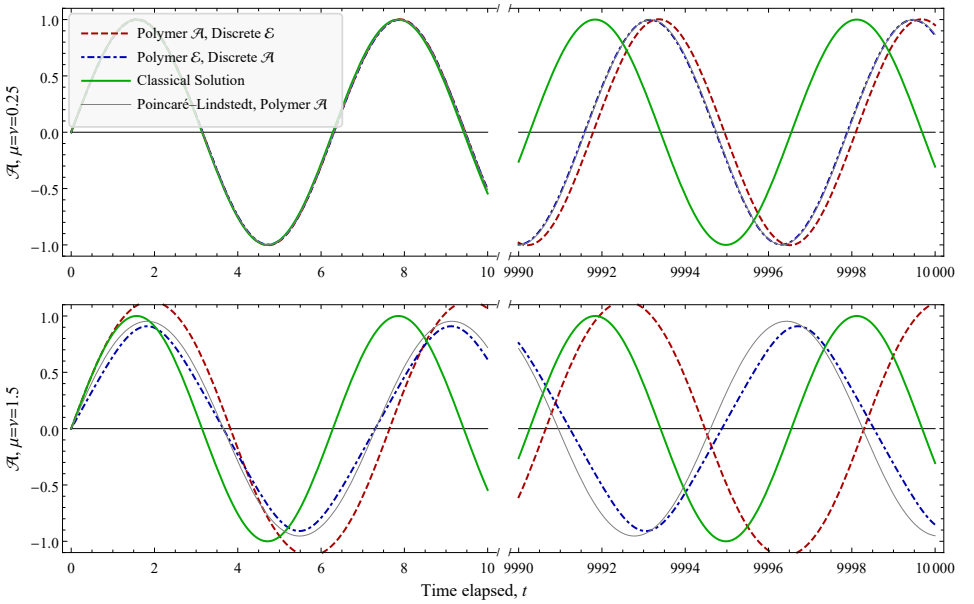


Fig. 2. Time evolution of $\mathcal{A}$ with $\mathcal{A}_I = 0$, $\mathcal{E}_I = 1$, and $k = 1$ for two different choices of μ in the case of a time-independent background spacetime, i.e., $\rho =$ const. Here ν refers to another representation of the polymer quantization in which the momentum $\mathcal{E}$ is discrete, which can be found in our original paper.[31] The solutions are shown at early times, and the axis is broken to show the behavior at a much later time. Solutions can be mapped to different choices of k and $\mathcal{E}_I$, while changing $\mathcal{A}_I = 0$ can be viewed as a phase shift.

solutions are plotted in Fig. 4. We can also obtain a perturbative solution

$$\mathcal{A}(t) \simeq \mathcal{E}_I \rho \sin\left[(1-(\mathcal{E}_I k\mu)^2/16)kT(t)\right] - \frac{\mathcal{E}_I^3 k^2 \mu^2}{16}\rho \sin^2\left[(1-(\mathcal{E}_I k\mu)^2/16)kT(t)\right]\cos\left[(1-(\mathcal{E}_I k\mu)^2/16)kT(t)\right], \tag{29}$$

where

$$T(t) = \int_{t_I}^{t} dt' \frac{1}{\rho(t')^2} \tag{30}$$

For GWs emitted at a time much greater than the characteristic wave time scale, i.e., $t_I \gg k^{-1}$, where T_I is the initial time, and for nonoscillatory solutions of ρ, the second-derivative term is small, and solutions to the auxiliary equations are well approximated by a simple power law, $\rho = 1/a$. In Fig. 3 we show the behavior of ρ for several sets of initial conditions, and for a universe with a cosmological constant with $w = -1$, $a \propto t^{1/3}$, and $t_I = 10^3$ (in units of k^{-1}). In subsequent plots we will use initial conditions that do not result in oscillatory behavior.

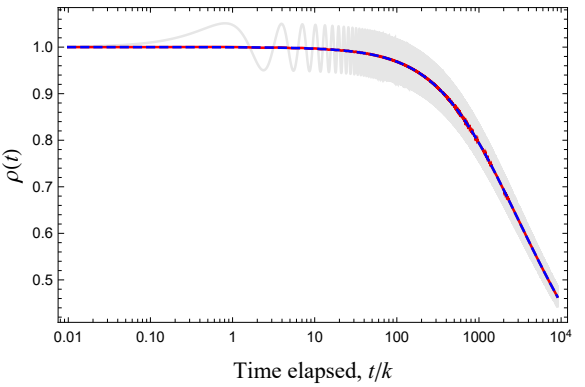


Fig. 3. Evolution of the auxiliary variable $\rho(t)$. The full numerical nonoscillatory solution is shown in solid red, an approximate power-law solution is shown in dashed blue, and a solution with initial conditions that result in oscillatory behavior is shown in light grey.

From the canonical transformation (13)–(15) (or, rather, its inverse), we see that the time-dependent waveform amplitude will pick up an overall factor of ρ relative to the time-independent one, the time coordinate will be altered, and the momentum will be similarly rescaled but will also pick up an additional factor proportional to the wave amplitude. Due to the monotonically decreasing nature of ρ and the smallness of its derivative, this additional factor will be a strongly subdominant contribution. In Fig. 4 we show the final solution for the field $\mathcal{A}(t)$ for this time-dependent background. Somewhat counterintuitively, the frequency is seen to increase at later times; more commonly the frequency is considered to decrease (redshift) with cosmological expansion. This is due to the choice of harmonic slicing we have made, with $N = a^3$ instead of the more commonly used $N = 1$ (synchronous) or $N = a$ (comoving) time coordinate.

5. Discussion and conclusion

In this work we have studied a certain effective form of GWs, considered as quantized perturbations propagating over a classical FLRW spacetime, in order to derive observational signatures to be compared with the results of experiments conducted by GW observatories. We have considered the Hamiltonian of classical gravitational perturbations, a time-dependent Hamiltonian, and have applied the techniques of polymer quantization to it. This polymer quantization was applied to each of the Fourier modes of the GW. A feature of this quantization is that the one-particle Hilbert space is modified and the Lorentz symmetry is no longer present.[38] This modification is "encoded" on the polymer scale μ, which is usually considered to be very small (of the order of the Planck scale). However, our intuition in the present case is that the propagation of the GWs may capture some insights about these modifications despite the small values of the polymer scales.

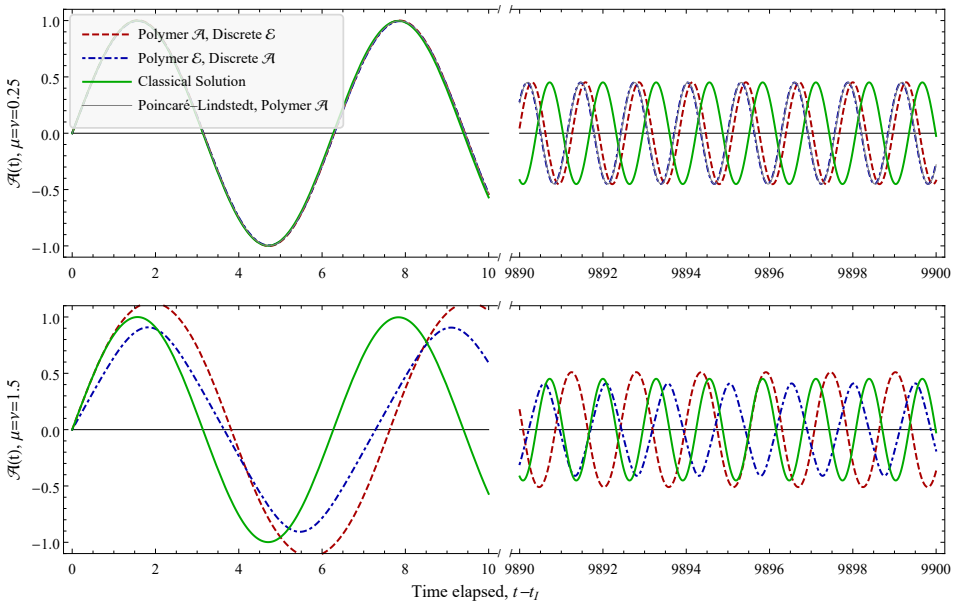


Fig. 4. Time evolution of $\mathcal{A}$ (as in Fig. 2) for two different choices of μ, for the case of a time-dependent background, i.e., $\rho(t)$ as described in the text. The axis is broken to show the behavior at a later time. Again, ν refers to another representation of the polymer quantization in which the momentum $\mathcal{E}$ is discrete, which can be found in our original paper.[31]

After deriving a tine-dependent effective polymer Hamiltonian using a novel approach, we derived both nonperturbative and perturbative analytical expression for the solutions and analyzed them to obtain further insight into the behavior of such waves. As a result, we found the following:

i) The form of the waves is modified. More precisely, there is a phase shift with respect to the classical case. Furthermore, small-amplitude harmonics are excited.
ii) The speed of the waves turns out to be smaller than the speed of light by a factor $k^2\left(\frac{\mathcal{E}_I\mu}{4}\right)^2$. This factor not only depends on the polymer scale μ and the initial momentum of the perturbations $\mathcal{E}_I$, but also on the wave vector k or, equivalently, the frequency of the waves. Thus, the higher-energy waves show a greater deviation from the classical behavior compared to the low-energy waves.
iii) The modifications to the classical behavior due to quantum effects become increasingly visible as the waves travel: the corrections result in an effective phase shift, which can become of order unity when $\mathcal{E}_I\mu^2k^3D_s$ is of order unity for a distance D_s traveled.

In a future work, we will proceed to apply our method to the case where both the background spacetime and the perturbations are effective.

References

1. T. Thiemann, *Modern Canonical Quantum General Relativity*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2007).
2. A. Ashtekar, C. Rovelli and L. Smolin, Gravitons and loops, *Phys. Rev. D* **44**, 1740 (1991).
3. M. Varadarajan, Gravitons from a loop representation of linearized gravity, *Phys. Rev. D* **66**, p. 024017 (2002).
4. L. Freidel and L. Smolin, The linearization of the kodama state, *Class. Quant. Grav.* **21**, 3831 (2004).
5. M. Bojowald and G. M. Hossain, Loop quantum gravity corrections to gravitational wave dispersion, *Phys. Rev.* **D77**, p. 023508 (2008).
6. A. Ashtekar, W. Kaminski and J. Lewandowski, Quantum field theory on a cosmological, quantum space-time, *Phys. Rev.* **D79**, p. 064030 (2009).
7. G. M. Hossain, V. Husain and S. S. Seahra, Background independent quantization and wave propagation, *Phys. Rev. D* **80**, p. 044018 (2009).
8. R. Gambini, J. Pullin and S. Rastgoo, Quantum scalar field in quantum gravity: The vacuum in the spherically symmetric case, *Class.Quant.Grav.* **26**, p. 215011 (2009).
9. J. Mielczarek, T. Cailleteau, J. Grain and A. Barrau, Inflation in loop quantum cosmology: Dynamics and spectrum of gravitational waves, *Phys. Rev. D* **81**, p. 104049 (2010).
10. G. Date and G. M. Hossain, Matter in loop quantum gravity, *SIGMA* **8**, p. 010 (2012).
11. R. Gambini, J. Pullin and S. Rastgoo, Quantum scalar field in quantum gravity: The propagator and lorentz invariance in the spherically symmetric case, *Gen. Rel. Grav.* **43**, 3569 (2011).
12. R. Gambini, S. Rastgoo and J. Pullin, Small lorentz violations in quantum gravity: Do they lead to unacceptably large effects?, *Class. Quant. Grav.* **28**, p. 155005 (2011).
13. P. M. Sa and A. B. Henriques, Gravitational wave generation in loop quantum cosmology, *Phys. Rev. D* **85**, p. 024034 (2012).
14. F. Hinterleitner and S. Major, Towards loop quantization of plane gravitational waves, *Class. Quant. Grav.* **29**, p. 065019 (2012).
15. F. Hinterleitner and S. Major, Plane gravitational waves and loop quantization, *J. Phys. Conf. Ser.* **360**, p. 012030 (2012).
16. D. E. Neville, Plane wave holonomies in quantum gravity. i. A model, *Phys. Rev. D* **92**, p. 044005 (2015).
17. D. E. Neville, Plane wave holonomies in quantum gravity. ii. A sine wave solution, *Phys. Rev. D* **92**, p. 044006 (2015).
18. P. A. Höhn, Canonical linearized regge calculus: Counting lattice gravitons with pachner moves, *Phys. Rev. D* **91**, p. 124034 (2015).
19. M. Arzano and G. Calcagni, What gravity waves are telling about quantum spacetime, *Phys. Rev.* **D93**, p. 124065 (2016) [Addendum: *Phys. Rev. D* **94**, no. 4, 049907 (2016)].
20. S. Rastgoo, Y. Tavakoli and J. C. Fabris, Phenomenology of a massive quantum field in a cosmological quantum spacetime, *Annals Phys.* **415**, p. 168110 (2020).
21. A. Dapor, J. Lewandowski and Y. Tavakoli, Lorentz Symmetry in QFT on Quantum Bianchi I Space-Time, *Phys. Rev.* **D86**, p. 064013 (2012).
22. B. Elizaga Navascués, M. Martín-Benito and G. A. Mena Marugán, Hybrid models in loop quantum cosmology, *Int. J. Mod. Phys. D* **25**, p. 1642007 (2016).
23. Y. Bonder, A. Garcia-Chung and S. Rastgoo, Bounds on the polymer scale from gamma ray bursts, *Phys. Rev. D* **96**, p. 106021 (2017).

24. J. Lewandowski, M. Nouri-Zonoz, A. Parvizi and Y. Tavakoli, Quantum theory of electromagnetic fields in a cosmological quantum spacetime, *Phys. Rev.* **D96**, p. 106007 (2017).
25. F. Hinterleitner, Canonical LQG operators and kinematical states for plane gravitational waves (3 2017).
26. A. Dapor and K. Liegener, Modifications to gravitational wave equation from canonical quantum gravity, *Eur. Phys. J. C* **80**, p. 741 (2020).
27. Y. Tavakoli and J. C. Fabris, Creation of particles in a cyclic universe driven by loop quantum cosmology, *Int. J. Mod. Phys.* **D24**, p. 1550062 (2015).
28. G. Calcagni, Quantum gravity and gravitational-wave astronomy (12 2020).
29. G. Calcagni, S. Kuroyanagi, S. Marsat, M. Sakellariadou, N. Tamanini and G. Tasinato, Gravitational-wave luminosity distance in quantum gravity, *Phys. Lett. B* **798**, p. 135000 (2019).
30. G. Calcagni, S. Kuroyanagi, S. Marsat, M. Sakellariadou, N. Tamanini and G. Tasinato, Quantum gravity and gravitational-wave astronomy, *JCAP* **10**, p. 012 (2019).
31. A. Garcia-Chung, J. B. Mertens, S. Rastgoo, Y. Tavakoli and P. Vargas Moniz, Propagation of quantum gravity-modified gravitational waves on a classical FLRW spacetime, *Phys. Rev. D* **103**, p. 084053 (2021).
32. A. Ashtekar, S. Fairhurst and J. L. Willis, Quantum gravity, shadow states, and quantum mechanics, *Class. Quant. Grav.* **20**, 1031 (2003).
33. A. Corichi, T. Vukasinac and J. A. Zapata, Polymer Quantum Mechanics and its Continuum Limit, *Phys. Rev. D* **76**, p. 044016 (2007).
34. H. A. Morales-Técotl, S. Rastgoo and J. C. Ruelas, Path integral polymer propagator of relativistic and nonrelativistic particles, *Phys. Rev. D* **95**, p. 065026 (2017).
35. H. A. Morales-Técotl, D. H. Orozco-Borunda and S. Rastgoo, Polymer quantization and the saddle point approximation of partition functions, *Phys. Rev. D* **92**, p. 104029 (2015).
36. E. Flores-González, H. A. Morales-Técotl and J. D. Reyes, Propagators in Polymer Quantum Mechanics, *Annals Phys.* **336**, 394 (2013).
37. A. A. García-Chung and H. A. Morales-Técotl, Polymer dirac field propagator: A model, *Physical Review D* **89**, p. 065014 (2014).
38. A. Garcia-Chung and J. D. Vergara, Polymer-fourier quantization of the scalar field revisited, *International Journal of Modern Physics A* **31**, p. 1650166 (2016).
39. H. A. Morales-Técotl, S. Rastgoo and J. C. Ruelas, Effective dynamics of the Schwarzschild black hole interior with inverse triad corrections, *Annals Phys.* **426**, p. 168401 (2021).
40. A. Garcia-Chung, D. G. Ruiz and J. D. Vergara, Dirac's method for time-dependent hamiltonian systems in the extended phase space, *arXiv preprint arXiv:1701.07120* (2017).
41. J. A. Austrich-Olivares, A. Garcia-Chung and J. D. Vergara, Instanton solutions on the polymer harmonic oscillator, *Classical and Quantum Gravity* **34**, p. 115005 (2017).
42. H. A. Morales-Técotl, D. H. Orozco-Borunda and S. Rastgoo, Polymerization, the Problem of Access to the Saddle Point Approximation, and Thermodynamics, in *14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories* (World Scientific, Singapore, 2017).

Primordial power spectrum from a matter-ekpyrotic scenario in loop quantum cosmology[1]

Bao-Fei Li*, Sahil Saini† and Parampreet Singh*

*Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

†Department of Physics, Guru Jambheshwar University of Science & Technology, Hisar 125001, Haryana, India

The general matter bounce scenario, including an Ekpyrotic field to avoid anisotropic instabilities, is studied in a loop quantized isotropic and homogeneous FLRW setting. The matter bounce scenario provides a convenient way to include quantum corrections from the bounce in the perturbations originating in the far past, which also produce a scale invariant power spectrum. LQC provides the right setting for studying quantum corrections in a matter bounce scenario as the bounce in LQC occurs entirely due to quantum geometrical effects without needing any exotic matter fields to avoid the singularity. A detailed exploration of this general matter-Ekpyrotic scenario in spatially flat FLRW spacetime in LQC filled with minimally coupled dust and Ekpyrotic scalar field is studied with the help of numerical simulations. Various features of the background dynamics are shown to be robust under variations in initial conditions and choice of parameters. We use the dressed metric approach for the perturbations and obtain a scale invariant power spectrum for modes exiting the horizon in the dust dominated contracting phase. In contrast to previous studies considering a constant equation of state for the Ekpyrotic field, we found that the magnitude of the power spectrum changes during the evolution. The scale invariant section of the power spectrum also undergoes a rapid increase in its magnitude in the bounce regime, while its scale invariance is unaffected. We argue that apart from increasing the magnitude, the bounce regime may only substantially affect the modes outside the scale invariant regime. However, the spectral index is found to be too close to unity, thus inconsistent with the observational constraints, necessitating further modifications of the model.

Keywords: Loop quantum cosmology; bouncing cosmology; primordial power spectrum.

1. Introduction

Early universe cosmology provides one of the most promising avenues to search for signatures of quantum gravity. Singularity free bouncing cosmologies of LQC provide a convenient setting for this. Inflation being the leading model has been studied widely with LQC backgrounds. Since there is no singularity, it is possible to extend the inflationary scenario further back in time upto the Planck regime of the quantum bounce to include quantum gravity signatures. These studies have led to predictions that are in excellent agreement with standard inflation and CMB observations at small angular scales but are closer to observations at larger scales than standard inflation.[2] The quantum bounce which lies to the past of this inflationary regime is a cataclysmic event which changes the course of the evolution of the universe, thus is expected to contribute significantly to the quantum gravity signatures in

the early universe. In this regard, it's natural to consider alternatives such as the matter-bounce scenario where the perturbations start out in the far past in the contracting branch, hence getting imprinted by the highly quantum phenomena as they pass through the quantum bounce. Due to a duality between inflation and matter-bounce, a scale invariant spectrum of perturbations is obtained in both of them.[3] An Ekpyrotic field is generally also included to guard against a potential BKL instability which may occur due to unchecked growth of anisotropies during the contracting phase.[4] An Ekpyrotic field is a scalar field with a negative-definite potential, thus having an equation of state larger than unity. As the bounce is approached in the contracting phase, the Ekpyrotic field having an equation of state larger than unity is expected to dominate the anisotropies which evolve as a^{-6}, where a denotes the scale factor.[5] Thus combining these two ingredients we obtain the the matter-Ekpyrotic scenario,[6] in which a scale invariant spectrum is produced by matter perturbations which exit the horizon during the contracting phase.

An important ingredient of the matter-Ekpyrotic scenario is a non-singular bounce which allows the scale invariant perturbations to pass through without changing them substantially. This is where LQC provides an advantageous framework. In LQC, the quantum bounce arises purely due to quantum geometry effects without requiring any exotic matter content violating the weak energy conditions.[7] Thus the marriage of the matter-Ekpyrotic scenario with an LQC bouncing background provides a compelling model for obtaining quantum gravity signatures in the early universe. Using effective dynamics of LQC, a nonsingular model with an Ekpyrotic potential was obtained[8,9] but it was found that a viable cyclic model cannot be realized unless another matter field or anisotropies are present.[10] Previous studies of the matter-Ekpyrotic model in LQC have either restricted to a special case of a constant equation of state for the Ekpyrotic field,[11] or when the general case has been considered,[12] the analysis has been limited to qualitative predictions due to difficulties associated with the deformed algebra approach used for perturbations in dealing with ultraviolet modes near the bounce. Thus a detailed study of the general matter-Ekpyrotic scenario in LQC analyzing the impact of the Ekpyrotic phase and the bounce on the perturbations has been lacking so far. Our work aims to fill this gap and provide a basis for further exploration.

2. Background dynamics of the matter-Ekpyrotic scenario in LQC

LQC uses the techniques of non-perturbative canonical quantization of LQG on symmetry reduced cosmological models such as the isotropic and homogeneous FLRW model. The canonical quantization of gravity in LQG is carried out in terms of the Ashtekar-Barbero connection and their conjugate triads instead of metric variables. The discrete quantum dynamics obtained in LQC is found to be well approximated by a continuum effective description in terms of differential equations on a smooth manifold, giving modified versions of Friedmann and Raychaudhuri equations that

include quantum corrections that lead to a singularity free evolution in which the big bang is replaced by a bounce. The effective Hamiltonian describing the dynamics of loop quantized spatially flat FLRW model in terms of symmetry-reduced Ashtekar-Barbero connection and its conjugate triad, namely c and p, is given by[7]

$$\mathcal{H} = -\frac{3v}{8\pi G\gamma^2\lambda^2}\sin^2(\lambda b) + \mathcal{H}_{\rm m}, \tag{1}$$

where $\lambda = \sqrt{\Delta}$. Here $\Delta = 4\sqrt{3}\pi\gamma\ell_{\rm Pl}^2$ is the minimum area eigenvalue in LQG. The matter Hamiltonian $\mathcal{H}_{\rm m}$ is given by

$$\mathcal{H}_{\rm m} = \frac{{p_\phi}^2}{2v} + vU(\phi) + \mathcal{E}_{\rm dust}, \tag{2}$$

and the Ekpyrotic potential $U(\phi)$ is taken to be,

$$U(\phi) = \frac{-2u_o}{e^{-\sqrt{\frac{16\pi}{p}}\phi} + e^{\beta\sqrt{\frac{16\pi}{p}}\phi}}, \tag{3}$$

which has the shape of an asymmetric exponential well. Here u_o, p and β are parameters taking positive values. The potential (3) is negative definite. The width of the potential increases with an increasing p, and the degree of asymmetry is controlled by β.

We obtain the following Hamilton's equations from the above Hamiltonian constraint,

$$\dot{b} = -\frac{3\sin^2(\lambda b)}{2\gamma\lambda^2} - 4\pi G\gamma P, \tag{4}$$

$$\dot{v} = \frac{3\sin(2\lambda b)}{2\gamma\lambda}v, \tag{5}$$

$$\dot{\phi} = \frac{p_\phi}{v}, \qquad \dot{p}_\phi = v\,U_{,\phi}, \tag{6}$$

where $U_{,\phi}$ stands for the differentiation of the potential with respect to the Ekpyrotic field and P is the isotropic pressure $P = -\frac{\partial\mathcal{H}_{\rm m}}{\partial v}$. It can be shown that the above equations lead to a following modified Friedmann equation of the form

$$H^2 = \frac{8\pi G}{3}\rho\left(1 - \frac{\rho}{\rho_c}\right), \tag{7}$$

where $\rho_c = 3/(8\pi G\gamma^2\lambda^2) \approx 0.41\rho_{\rm Pl}$ is called the critical energy density in LQC. The Hubble rate is generically bounded, and vanishes and turns around when the energy density reaches a maximum ρ_c. This is when the bounce occurs. The matter content satisfies the continuity equation $\dot{\rho} + 3H(\rho + P) = 0$, where ρ and P are given respectively by

$$\rho = \frac{\dot{\phi}^2}{2} + U(\phi) + \frac{\mathcal{E}_{\rm dust}}{v}, \quad P = \frac{\dot{\phi}^2}{2} - U(\phi). \tag{8}$$

We use the above set of equations along with suitable initial conditions provided at the bounce to numerically analyze the background dynamics of this model. We

also carry out simulations to study the effect of changes in the potential parameters and the relative proportion of the initial dust energy density compared to the Ekpyrotic field energy density. We find that LQC leads to a non-singular bounce in all cases. Although there can be multiple bounces, we restrict ourselves to the cases with a single bounce to study the evolution of curvature perturbations. As an example, we show here the evolution for the choice of parameters $u_o = 0.75, p = 0.10$ and $\beta = 5.0$ with the following initial conditions:

$$\phi_B = 0, \qquad \mathcal{E}_{dust} = 2.00 \times 10^{-5}, \qquad p_{\phi_B} = 1.50. \tag{9}$$

The volume at bounce is taken to be unity, while we choose $b_B = \pi/2\lambda$ in order to have the correct classical limit on either side of the bounce. The dust energy density is chosen such that the energy density at the bounce is dominated by the Ekpyrotic field. The initial conditions are taken to satisfy the Hamiltonian constraint which has to be satisfied throughout the evolution. The evolution of the energy densities and the equation of state for these initial conditions is shown in Fig. 1. We note that the contracting phase starts in a dust dominated phase where the total energy density is very small and the universe is classical. Eventually it transitions to a phase dominated by the Ekpyrotic field which lasts through the bounce. Thus we find the equation of state at the bounce is $w_B \approx 5.76$. Thus a distinct phase where $w > 1$ is obtained near the bounce, which is very important to avoid the BKL instability as envisaged. The $w > 1$ regime is obtained only when the scalar field is traversing the very bottom of the potential, and is obtained only for a short duration during the scalar field dominated phase near the bounce. After the bounce the equation of state quickly drops below unity and decreases monotonically as the Ekpyrotic field moves away from the bottom of the potential. We did extensive simulations to study the robustness of the qualitative features of the background dynamics obtained above by varying the parameters of the potential and taking different proportions of the initial dust energy density while still keeping the bounce dominated by Ekpyrotic field. We found that the effect of reducing the proportion of the dust energy density

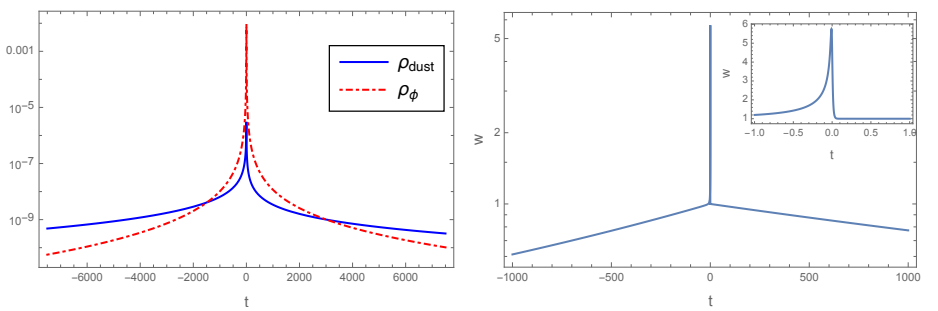


Fig. 1. In the left panel, we see that the energy density of the Ekpyrotic field dominates over the dust energy density in the bounce regime. The right panel shows the behavior of the equation of state when the Ekpyrotic field is dominant and the inset plot depicts the details of the change in the equation of state near the bounce point.

at the bounce has the obvious effect of increasing the duration of Ekpyrotic field dominated regime and the $w > 1$ regime near the bounce. The $w > 1$ regime can also be elongated by widening the potential by increasing p, making it less steep due to which the field takes longer to traverse the bottom of the well. Further, increasing the parameter u_o has effect of increasing both the depth and steepness of the potential. If we increase u_o while keeping the rest of the conditions same, this has the effect of numerically increasing both the kinetic and potential energy of the scalar field at the bounce while keeping the total the same (since the potential is negative-definite, the total energy density is the difference of the numerical values of kinetic and potential energies). Higher kinetic energy means the field climbs out of the well faster, i.e. the duration of the Ekpyrotic field dominated phase is reduced. But a deeper well means that $w > 1$ for most of the duration of this climb. Thus increasing u_o has the effect of decreasing the Ekpyrotic field dominated phase while increasing the $w > 1$ regime. This is illustrated in the following table showing durations of different regimes in the contracting branch (in Planck seconds):

u_o	Ekpyrotic field dominated phase	$w > 1$ regime
1	8.70×10^4	100
0.75	1.50×10^5	65
0.03	3.00×10^5	11
0.008	3.04×10^5	6.5

3. Scalar power spectrum using the dressed metric approach

In this section, we study perturbations around the background spacetime studied in the previous section. As mentioned above, due to the duality with the inflationary epoch,[3] the perturbations in dust matter that exit the horizon during the contracting phase yield a scale invariant spectrum. The dust dominated phase in our background dynamics occurs in the far past, away from the bounce when the universe is classical. Thus we must start perturbations in the far past and then evolve them through the Ekpyrotic phase and the bounce upto the expanding branch.

Since the Ekpyrotic field contributes negligibly to the total energy density in the far past and the equation of state $w \approx 0$, thus we ignore it while considering perturbations in this phase. In this regime the total energy density is far below the Planck density and quantum gravity modifications are negligible. Thus the perturbation equation in terms of the Mukhanov-Sasaki variable also takes its classical

form in this phase,

$$\nu_k'' + \left(k^2 - \frac{z''}{z}\right)\nu_k = 0, \tag{10}$$

where the prime represents the differentiation with respect to the conformal time. And we choose initial conditions such that, the above state corresponds to the Bunch-Davies vacuum in the asymptotic past:

$$\lim_{\eta\to-\infty} \nu_k = \frac{e^{-ik\eta}}{\sqrt{2k}}. \tag{11}$$

The above equations are only valid in the matter dominated phase. As perturbations evolve towards the bounce, the Ekpyrotic field starts dominating and the energy density starts increasing towards the Planck scale. Since the Ekpyrotic field dominated phase overlaps with the bounce regime, we cannot use classical equations to describe the perturbations in the Ekpyrotic field dominated phase. Of the various approaches in LQC that consider perturbations in the quantum spacetime, we work with the dressed metric approach.[13] This avoids the Jeans instability encountered near the bounce in the previous study based on deformed algebra approach.[12] In the dressed metric approach, the quantum spacetime is approximated by a differential manifold with a dressed metric for sharply peaked semi-classical states. The effective description of the quantum perturbations is then provided by a Mukhanov-Sasaki variable with an evolution equation of the form

$$\nu_k'' + \left(k^2 + \Omega^2 - \frac{a''}{a}\right)\nu_k = 0, \tag{12}$$

where Ω^2 can be written in terms of the background variables as

$$\Omega^2 = 3\kappa\frac{p_\phi^2}{a^4} - 18\frac{p_\phi^4}{a^6\pi_a^2} - 12a\frac{p_\phi}{\pi_a}U_{,\phi} + a^2U_{,\phi\phi}. \tag{13}$$

Here $\kappa = 8\pi G$ and π_a is the conjugate momentum of the scale factor. The required background quantities in the above equations are to be calculated using the effective background dynamics described in section 2. The power spectrum is given by

$$\mathcal{P}_\mathcal{R} = \frac{k^3}{2\pi^2}\frac{|\nu_k|^2}{z^2}, \tag{14}$$

with $z = a\dot{\phi}/H$. The modes that exit the horizon in the matter-dominated phase of contraction are scale invariant, and it is these modes that will be observable today. However, numerical simulations are important to study the impact of the Ekpyrotic field dominated phase and the bounce on the spectrum of these perturbations. In particular, the scale invariance of spectrum is to be preserved in order to be consistent with observations. Fig. 2 shows the evolution of scalar power spectrum at different stages.

It is clear from the plots that a scale invariant spectrum is already generated at the end of the dust dominated phase during contraction. However, in contrast to previous studies relying on constant equation of state for the Ekpyrotic field,[11] the

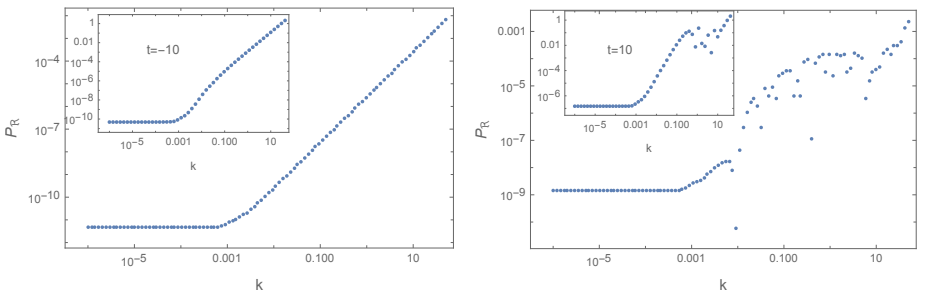


Fig. 2. The power spectra at different times are depicted: in the left panel, the power spectra are evaluated at the transition time from dust dominated to Ekpyrotic field dominated regime in the contracting phase ($t \approx -1.49 \times 10^3$) and at $t = -10$ (in the inset plot); while in the right panel, the power spectra are evaluated at the transition time in the expanding phase ($t \approx 3.02 \times 10^3$) from Ekpyrotic field domination to dust domination and at $t = 10$ (in the inset plot).

magnitude of the scale invariant power spectrum changes over time. Although the scale invariance is preserved through the bounce, a rapid increase in the magnitude of the power spectrum is observed from before the bounce to after the bounce seen by comparing the inset plots. Further, an oscillatory regime reminiscent of inflationary scenarios is seen for $k \geq 10^{-2}$. A comparison of the oscillatory regime at different times as seen in the right panel of Fig. 2 shows that different modes become oscillatory as they exit the horizon in the expanding phase. In contrast to earlier works on the matter-Ekpyrotic scenario in LQC dealing with special cases as well as in contrast to inflationary scenarios, the magnitude of the spectrum in the scale invariant regime gradually decreases as the universe expands. The qualitative behavior of the spectrum remains the same when we vary the proportion of the initial dust energy density relative to the Ekpyrotic field density. However, the range of modes that show scale invariance may get shifted. When the dust energy proportion is decreased, that leads to a shorter duration of dust dominated phase during contraction. Since only those modes that exit during the dust dominated contraction show a scale invariant spectrum, range of modes that show scale invariance is decreased. This may be used to put a constraint on the duration of Ekpyrotic field allowed by observed range of scale invariant modes.

Lastly, in order to compare with observations, we compute the spectral index of the spectrum given by:

$$n_s = 1 + \frac{\mathrm{d} \ln P_{\mathcal{R}}}{\mathrm{d} \ln k}. \tag{15}$$

The spectral index is plotted in Fig. 3. We find that the spectral index is larger than unity in the scale invariant regime and is unfavored by recent CMB observations that put tight constraints on the spectral index: $n_s = 0.9649 \pm 0.0042$ (68%CL).[14] We find that this cannot be ameliorated by either changing the parameters of the scalar field potential or changing the proportion of initial dust density to Ekpyrotic field density. Thus making further modifications of this model necessary.

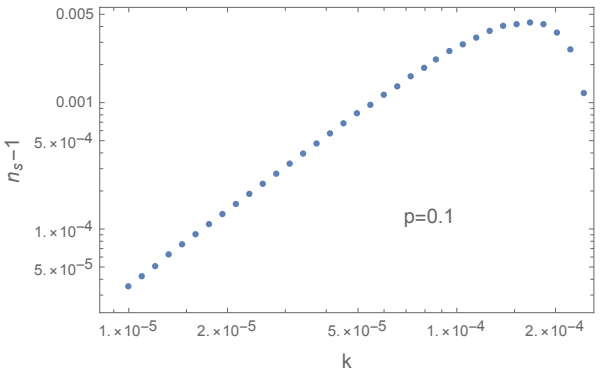


Fig. 3. For the power spectrum displayed in Fig. 2, we evaluate the spectral index n_s at the transition point $t \approx 3.02 \times 10^3$ in the expanding branch and find n_s is larger than unity in the scale invariant regime of the power spectrum.

4. Conclusion

In this paper, we study a general matter-Ekpyrotic scenario in an isotropic and homogeneous FLRW universe in LQC. Using extensive numerical simulations, we establish some general features of the background dynamics and the scalar power spectrum, some of which were not seen in earlier works that either dealt with special cases or were limited to qualitative analysis. A bouncing background cosmology, characteristic of LQC, is obtained where a contracting and an expanding branch connected through a bounce. We find that the background dynamics in each branch is characterized in general by two distinct phases. It starts out in the far past with dust domination and transitions to Ekpyrotic field dominated phase which overlaps with the high energy quantum regime at the bounce. After the bounce there is another transition back to dust dominated universe in the expanding branch. Further we find that the Ekpyrotic field dominated phase generally also contains a phase of ultrastiff equation of state ($w > 1$) leading up to the bounce, which is essential in avoiding the BKL instability. We varied the parameters of the potential and also varied the relative proportion of initial dust density (while keeping it sub-dominant at the bounce), and found the above mentioned features of the background dynamics to hold in general.

We found that the duration of the Ekpyrotic field dominated phase and the $w > 1$ phase can be increased by increasing the width of the Ekpyrotic field potential or by decreasing the relative proportion of initial dust density. In contrast, increasing the depth of the potential has the effect of shortening the Ekpyrotic field dominated phase while increasing the $w > 1$ phase at the same time.

For the study of comiving curvature perturbations, we give our initial conditions in the far past in the dust-dominated classical phase of the contracting branch in order to include the full effect of the quantum bounce on the power spectrum. We employ the classical Mukhanov-Sasaki equation for the dust-dominated contraction

phase as the universe is large and classical in this phase. Since the Ekpyrotic field dominated phase overlaps with the quantum bounce regime, we have employed the dressed metric approach to obtain the effective evolution equation for quantum perturbations on a quantum spacetime in this regime. This is a deliberate choice as previous studies employing the deformed algebra approach have faced difficulty in analyzing ultraviolet modes near the bounce. Using the dressed metric approach for the bounce regime has allowed us to fully explore the scalar power spectrum in the bounce regime.

We found that the perturbation modes exiting the horizon during dust dominated phase have a scale invariant spectrum. These modes then maintain their scale invariance as they pass through the Ekpyrotic field dominated phase, the $w > 1$ phase and the bounce, until they start re-entering the horizon in the expanding branch after which they become oscillatory. However, unlike previous studies considering a constant equation of state for the Ekpyrotic field, the magnitude of the scale invariant modes gradually increases during the Ekpyrotic-field dominated contraction, then rapidly increases during the bounce regime and then gradually decreases as the universe expands. This is also in contrast to inflationary scenarios where the magnitude is frozen after the modes exit the horizon during inflation. Further, the duration of the Ekpyrotic field dominated not only impacts the magnitude of the scale invariant power spectrum, but also impacts the range of comoving modes that lie in the scale invariant part of the spectrum.

We also studied the spectral index of the scale invariant part of the spectrum and found that it is inconsistent with current observational bounds. This is due to the fact that perturbation originating in a purely dust-dominated phase of contraction have an exactly scale invariant spectrum, while the observed spectrum has a slight red tilt. Thus changing potential parameters of the Ekpyrotic field potential has no effect on this feature.

Based on our results, further studies are needed to fully explore the possibilities presented by the matter-Ekpyrotic scenario in LQC. In particular, a lower bound on the minimum number of e-foldings required of the $w > 1$ phase needs to be rigorously established by studying the Ekpyrotic scenario in more general anisotropic spacetimes. Further, a suitable extension of the matter-Ekpyrotic scenario needs to be developed where the perturbations finally exit to a radiation-dominated universe. Bounds on the duration of the Ekpyrotic field dominated phase also needs to be established by employing its impact on the range of scale invariant modes and their magnitude. The issue of the inconsistent spectral index can addressed if instead of sourcing perturbations from a purely dust-dominated phase, we have a matter field with slightly negative equation of state. Although, preliminary studies following these lines have been carried out in the literature in various bouncing cosmologies including those of LQC, but rigorous and extensive studies taking the most general matter-Ekpyrotic scenario need to be done.

References

1. B. F. Li, S. Saini and P. Singh, *Primordial power spectrum from a matter-Ekpyrotic bounce scenario in loop quantum cosmology*, Phys. Rev. D **103**, 066020 (2021).
2. A. Ashtekar, B. Gupt, V. Sreenath, *Cosmic Tango Between the Very Small and the Very Large: Addressing CMB Anomalies Through Loop Quantum Cosmology*, Frontiers in Astronomy & Space Science **8**, 685288 (2021).
3. D. Wands, *Duality invariance of cosmological perturbation spectra*, Phys. Rev. D **60**, 023507 (1999).
4. V. A. Belinsky, I. M. Khalatnikov and E. M. Lifshitz, *Oscillatory approach to a singular point in the relativistic cosmology*, Adv. Phys. **19**, 525 (1970).
5. J. K. Erickson, D. H. Wesley, P. J. Steinhardt and N. Turok, *Kasner and mixmaster behavior in universes with equation of state* $w \geq 1$, Phys. Rev. D **69**, 063514 (2004).
6. Y. F. Cai, D. A. Easson and R. Brandenberger, *Towards a nonsingular bouncing cosmology*, JCAP **1208**, 20 (2012).
7. A. Ashtekar and P. Singh, *Loop Quantum Cosmology: A Status Report*, Class. Quant. Grav. **28**, 213001 (2011).
8. P. Singh, K. Vandersloot, and G. V. Vereshchagin, *Nonsingular bouncing universes in loop quantum cosmology*, Physical Review D **74(4)**, 043510, (2006).
9. M. Bojowald, R. Maartens and P. Singh, *Loop quantum gravity and the cyclic universe,* Phys. Rev. D **70**, 083517 (2004).
10. T. Cailleteau, P. Singh, K. Vandersloot, *Nonsingular Ekpyrotic/cyclic model in loop quantum cosmology*, Phys. Rev. D **80**, 124013, (2009).
11. J. Haro, J. Amoros and L. Salo, *The matter-Ekpyrotic bounce scenario in Loop Quantum Cosmology*, JCAP **09**, 002 (2017).
12. Y. F. Cai and E. Wilson-Ewing, *Non-singular bounce scenarios in loop quantum cosmology and the effective field description.* JCAP **03**, 026 (2014).
13. B. F. Li, P. Singh, A. Wang, *Primordial power spectrum from the dressed metric approach in loop cosmologies*, Phys. Rev. D **100**, 086004 (2020).
14. P. A. R. Ade *et al.* [Planck Collaboration], *Planck 2018 results. X. Constraints on inflation*, Astron. Astrophys. 641, A10 (2020).

The primordial power spectra in modified loop quantum cosmology

Bao-Fei Li*, Javier Olmedo*, Parampreet Singh* and Anzhong Wang†,‡

*Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

†Institute for Theoretical Physics & Cosmology, Zhejiang University of Technology, Hangzhou, 310032, China

‡GCAP-CASPER, Department of Physics, Baylor University, Waco, TX 76798, USA

We review recent developments in the primordial power spectra of two modified loop quantum cosmological models (mLQCs) which originate from the quantization ambiguities while loop quantizing the spatially-flat Friedmann-Lemaître-Robertson-Walker (FLRW) universe. The properties of the background dynamics and the primordial scalar power spectra in two modified models, namely mLQC-I and mLQC-II, are reported. In both models, the inflationary scenario can be naturally extended to the Planck regime when a single scale field is minimally coupled to gravity with an inflationary potential and the big bang singularity is replaced with a quantum bounce. The qualitative difference lies in the behavior of the contracting phase where a quasi de Sitter phase emerges in mLQC-I. When applying the dressed metric approach and the hybrid approach to mLQCs, we find the most distinguishable differences between these models and the standard loop quantum cosmology (LQC) occur in the infrared and intermediate regimes of the power spectra.

Keywords: Loop quantum cosmology; bouncing cosmology; primordial power spectrum.

1. Introduction

Loop quantum cosmology (LQC)[1,2] provides an elegant resolution to the big bang singularity in the standard big bang cosmology by replacing it with a quantum bounce and also naturally extends the inflationary paradigm into the Planck regime. The loop quantization of the spatially flat FLRW spacetime in LQC is based on the homogeneity and isotropy of the spacetime. It has been found that the dynamics of the resulting quantum theory can be well described by the effective dynamics for the states which are sharply peaked around the classical trajectories at late times.[3] However, as is common in any quantum theory, different regularizations of the classical Hamiltonian may give rise to different quantum theory and finally result in distinct physical consequences. So it is important to check the robustness of the theoretical predictions from standard LQC by studying its variants originating from quantization ambiguities. The modified LQC models initially proposed in[4] provide a good platform to investigate the impacts of the quantization prescriptions on the dynamics and the observations.

In standard LQC, when quantizing a spatially flat FLRW universe, the Lorentz term was initially treated in the same way as the Euclidean term since they are

proportional to each other at the classical level due to the homogeneity and isotropy of the background spacetime.[5] A separate treatment of these two terms was first implemented in[4] where two modified LQC models, namely mLQC-I and mLQC-II, were proposed due to different regularizations of the Lorentz term. Later, mLQC-I was rediscovered by computing the expectation value of the Hamiltonian constraint in LQG with complexifier coherent state.[6] The right Friedmann and Raychauduri equations of these two models were first reported in,[7,8] and it was found that after taking into account an inflaton field in mLQCs, the inflationary phase turns out to be an attractor in both models.[8] As a result, the inflationary scenario can also be naturally extended to the Planck regime in both models. In the following, assuming the validity of the effective dynamics in mLQCs, we first summarize the main features of the background dynamics of these two mLQCs and their similarities and differences with/from standard LQC. Then we report the results on the numerical simulations of the power spectra in both mLQCs in the dressed metric and the hybrid approaches.

2. The background dynamics in the modified LQC

Although the big bang singularity is resolved and replaced with a quantum bounce in both mLQCs, the differences in the background dynamics between these models are still distinct. Similar to standard LQC, the evolution of the background dynamics in mLQC-II is symmetric with respect to the bounce when gravity is coupled to a massless scalar field. The maximum energy density in mLQC-II is about $1.73m_{\rm pl}^4$ which is larger than the maximum energy density ($\approx 0.41m_{\rm pl}^4$) in LQC. It turns out that the evolution of the universe in mLQC-II is similar to that in LQC when gravity is coupled to the same matter content. In contrast, in mLQC-I, the contracting and the expanding phases are no longer symmetric with respect to the bounce point which takes place at the maximum energy density ($\approx 0.097m_{\rm pl}^4$). A characteristic quasi de Sitter phase quickly emerges in the contracting phase when the universe is evolved backwards from the bounce point. This de Sitter phase is dominated by a Planck-sized effective cosmological constant and the asymptotic forms of the Friedmann and Raychauduri equations reveal a rescaled Newton's constant in the contracting phase as well.[7] As a result, when filled with an inflaton field, the classical universe can only be recovered in the expanding phase of mLQC-I while in the contracting phase, the universe remains quantum with a Planck-scale Hubble parameter.

Meanwhile, the numerical simulations of the background dynamics shows qualitatively similar behavior of the universe in the expanding phase of mLQCs and standard LQC.[9] The initial conditions of the background dynamics which are relevant to CMB data are those dominated by the kinetic energy of the inflaton field at the bounce. For these initial conditions, three distinct phases can be observed in the expanding phase: the bouncing phase with $w \approx 1$ which lasts only a few number of e-foldings, the transient transition phase where the equation of state quickly

decreases from positive unity to near negative unity and the slow-roll inflationary phase. These generic features of the background dynamics shared by mLQCs and LQC imply the robustness of the preinflationary dynamics in loop quantizations of the spatially flat FLRW universe and the different regularizations of the Lorentz term in the classical Hamiltonian constraint leave imprints only in the contracting phase.

3. The primordial power spectra in the modified LQC

To compute the primordial power spectra in the mLQCs, one can appeal to the techniques developed in standard LQC where several distinct approaches are widely studied. Among them the most popular ones are the dressed metric approach[10–12]and the hybrid approach.[13–16] We have applied both these approaches to mLQCs and numerically obtained the resulting scalar power spectrum.[17,18] Irrespective of specific model and approach, the power spectrum can be generally divided into three distinctive regimes: the infrared (IR) regime, the intermediate regime and the ultraviolet (UV) regime.

The behavior of the IR regime sensitively depends on the initial states, the specific model and the applied approach. For example, when the fourth-order adiabatic states are employed as the initial states in the contracting phase, the magnitude of the power spectra remains almost constant in LQC and mLQC-II in both dressed metric and hybrid approaches. In contrast, when the second-order adiabatic states are used as the initial states, the magnitude of the power spectra keeps increasing with the comoving wavenumber in LQC and mLQC-II in both approaches. In mLQC-I, the power spectrum in the IR regime strongly relies on the approach we use. In the dressed metric approach, due to the emergent de Sitter phase, the preferred choice of the Bunch-Davies (BD) vacuum state yields a Planck-sized magnitude of the power spectrum in the IR regime. While in the hybrid approach, the effective mass changes so drastically as compared with the one in the dressed metric approach that the adiabatic states can be again chosen as the initial states of the scalar perturbations, leading to a power spectrum whose magnitude is comparable to that of LQC and mLQC-II in the IR regime. From our numerical results, we find the relative difference between LQC and mLQC-II in the IR regime turns out to be less than 40% in the dressed metric approach and less than 50% in the hybrid approach. Due to the similar background dynamics between the LQC and mLQC-II, the relative difference between LQC and mLQC-II is much less than that between LQC(mLQC-II) and mLQC-I.

The power spectrum keeps oscillating throughout the intermediate regime in all of the three models with the dressed metric and the hybrid approaches. This oscillatory behavior is believed to be associated with the bouncing regime where adiabatic conditions are violated for the modes whose comoving wavenumbers are close to the characteristic wavenumber in each model and thus particle creations become appreciable.[19] For the adiabatic initial states and in particular the BD

vacuum state in the dressed metric approach of mLQC-I, the magnitude of the power spectrum in this regime is amplified as compared with that in the UV regime. Moreover, the relative difference of the power spectra between different models turns out to be larger than that in the IR regime due to the oscillatory behavior of the power spectra. In particular, the relative difference of the power spectra between LQC and mLQC-II in the intermediate regime can be up to 100% in both approaches and the relative difference between LQC and mLQC-II is also smaller than that between LQC(mLQC-II) and mLQC-I.

Finally, regardless of the initial states, the specific model and approach used to compute the power spectra, we always observe a scale-invariant UV regime whose magnitude agrees with the observations. The relative difference between the magnitudes of the power spectra from three models turns out to be less than 1%. As a result, only the IR and intermediate regimes are mostly affected by the different regularizations of the classical Hamiltonian constraints, there is little impact on the UV regime due to the quantization ambiguities as it should be.

4. Summary

We have studied the background dynamics and the primordial power spectra in mLQCs and compare the results with those from LQC. The background dynamics in mLQC-II is qualitatively similar to that in LQC with a symmetric evolution with respect to the bounce when gravity is minimally coupled to a massless scalar field. An asymmetric evolution is observed only in mLQC-I where the classical limit can only be recovered in the expanding phase and there shows up an emergent de Sitter spacetime in the contracting phase. The difference between the background dynamics also affects the power spectrum in the IR and intermediate regime. From our numerical results, we find the relative difference of the power spectra between LQC and mLQC-II in these two regimes is much smaller than the relative difference between LQC(mLQC-II) and mLQC-I in the dressed metric and hybrid approaches. Moreover, mLQC-I also serves as a good example to explicitly show the differences between the dressed metric and hybrid approaches. In particular, a Planck scale magnitude of the power spectrum in the IR regime of mLQC-I with the dressed metric approach is observed while in the hybrid approach the magnitude of the power spectrum is suppressed in the IR regime of mLQC-I. This difference is completely due to the distinct construction of the Mukhanov-Sasaki equation in each approach which leads to different effective masses. Finally, the power spectrum in the UV regime is found to be consistent with the CMB data in all three models regardless of the approach and the initial states.

References

1. A. Ashtekar and P. Singh, *Loop Quantum Cosmology: A Status Report*, Class. Quant. Grav. **28**, 213001 (2011).

2. I. Agullo and P. Singh, *Loop Quantum Cosmology*, in Loop Quantum Gravity: The First 30 Years, edited by A. Ashtekar and J. Pullin (World Scientific, Singapore, 2017).
3. A. Ashtekar, T. Pawlowski, and P. Singh, *Quantum nature of the Big Bang: Improved dynamics*, Phys. Rev. **D74**, 084003 (2006).
4. J. Yang, Y. Ding and Y. Ma, *Alternative quantization of the Hamiltonian in loop quantum cosmology II: Including the Lorentz term*, Phys. Lett. B**682** (2009) 1.
5. A. Ashtekar, M. Bojowald and J. Lewandowski *Mathematical structure of loop quantum cosmology*, Adv. Theor. Math. Phys. 7:233-268,2003.
6. A. Dapor and K. Liegener, *Cosmological Effective Hamiltonian from full Loop Quantum Gravity Dynamics*, Phys. Lett. B**785** (2018) 506.
7. B.-F. Li, P. Singh, A. Wang, *Towards cosmological dynamics from loop quantum gravity*, Phys. Rev. D**97**, 084029 (2018).
8. B.-F. Li, P. Singh, A. Wang, *Qualitative dynamics and inflationary attractors in loop cosmology*, Phys. Rev. D**98**, 066016 (2018).
9. B.-F. Li, P. Singh, A. Wang, *Genericness of pre-inflationary dynamics and probability of the desired slow-roll inflation in modified loop quantum cosmologies*, Phys. Rev. D**100**, 063513 (2019).
10. I. Agullo, A. Ashtekar and W. Nelson, *Quantum gravity extension of the inflationary scenario*, Phys. Rev. Lett. **109**, 251301 (2012).
11. I. Agullo, A. Ashtekar and W. Nelson, *Extension of the quantum theory of cosmological perturbations to the Planck era*, Phys. Rev. D**87**, 043507 (2013).
12. I. Agullo, A. Ashtekar and W. Nelson, *The pre-inflationary dynamics of loop quantum cosmology: Confronting quantum gravity with observations*, Class. Quantum Grav. **30**, 085014 (2013).
13. L. Castelló Gomar, M. Fernández-Méndez, G. A. Mena Marugan and J. Olmedo, *Cosmological perturbations in hybrid loop quantum cosmology: Mukhanov-Sasaki variables*, Phys. Rev. D**90**, 064015 (2014).
14. M. Fernández-Méndez, G. A. Mena Marugán and J. Olmedo, *Hybrid quantization of an inflationary universe*, Phys. Rev. D**86**, 024003 (2012).
15. M. Fernández-Méndez, G. A. Mena Marugán and J. Olmedo, *Hybrid quantization of an inflationary model: The flat case*, Phys. Rev. D**88**, 044013 (2013).
16. L. Castelló Gomar, M. Martín-Benito, G. A. Mena Marugán, *Gauge-invariant perturbations in hybrid quantum cosmology*, JCAP 1506 (2015) 045.
17. B.-F. Li, P. Singh, A. Wang, *Primordial power spectrum from the dressed metric approach in loop cosmologies*, Phys. Rev. D**100**, 086004 (2020).
18. B.-F. Li, J. Olmedo, P. Singh, and A. Wang, *primordial scalar power spectrum from the hybrid approach in loop cosmologies*, Phys. Rev. D**102**, 126025 (2020).
19. Q. Wu, T. Zhu, and A. Wang, *Nonadiabatic evolution of primordial perturbations and non-Gaussianity in hybrid approach of loop quantum cosmology*, Phys. Rev. D**98**, 103528 (2018).

Primordial perturbations in kinetically dominated regimes of classical and quantum cosmology

B. Elizaga Navascués

JSPS International Research Fellow, Department of Physics, Waseda University, 3-4-1 Okubo, Shinjuku-ku, 169-8555 Tokyo, Japan

R. Jiménez-Llamas and G.A. Mena Marugán*

Instituto de Estructura de la Materia, IEM-CSIC, Serrano 121, 28006 Madrid, Spain
**E-mail: mena@iem.cfmac.csic.es*

There is an increasing interest in very early stages of the Universe in which the energy density of the inflaton could be dominated by its kinetic part. This includes classical inflationary scenarios with deviations from slow-roll regimes that can introduce modifications to the power spectra of the primordial fluctuations. Another example are certain quantum bouncing cosmologies. For instance, this is the typical situation in loop quantum cosmology if quantum corrections may leave observable traces in the power spectra. In models of this type, we discuss the leading-order effects of an inflaton potential on the primordial perturbations. These effects are of two kinds, referred to the case without potential. First, there are changes in the background-dependent mass appearing in the dynamical equations of the perturbations in conformal time. Second, away from conventional slow roll, a Bunch-Davies vacuum may no longer be natural, and possible new choices depend on the potential.

Keywords: Loop Quantum Cosmology, Kinetically Dominated Inflation, Primordial Perturbations

1. Introduction

We live in an era of precision cosmology, in which the technological progress has allowed us to obtain more accurate observations and determine a number of cosmological parameters with a precision of a few percent.[1] This new era opens the possibility of confronting predictions with observations, falsifying cosmological models and gravitational theories that may suggest corrections to general relativity (GR). Although the standard cosmological model, based on GR and an inflationary phase in slow-roll regime, has proved extremely succesful to explain the observations, there exist some anomalies that, at least, point towards the statistical exceptionality of our universe.[2–6] Namely, cosmological data such as those imprinted in the spectrum of anisotropies of the cosmic microwave background (CMB) or the lensing amplitude are not statistically favored within this standard model. These anomalies may be a sign of new physics that modifies in part the paradigms of the model.

A possibility that has attracted a lot of attention lately is the occurrence of kinetically dominated regimes in the very early epochs of the expansion of the Universe, i.e. regimes in which the scalar field that is going to drive the inflationary

mechanism displays a period in its evolution in which its energy density is dominated by the kinetic contribution, rather than by a potential. This kinetic dominance would have taken place before slow roll, and the transition between both regimes may cover a part of the inflationary era. Typically, the situations of interest combine this kinetic dominance with a short lived inflation, because it is then that the frequencies that are most affected by the dynamics away from slow roll can overlap at present with the band that is observed in the CMB fluctuations.[7,8] This change in the early dynamics of the primordial fluctuations can leave traces in their amplitude and modify the predictions about their power and other measurable quantities, mostly for modes that in those early moments have (wavelength) scales similar to those of the physical phenomena that modify the standard picture of slow-roll inflation.[8]

Clearly, this type of modifications of the standard model can happen within GR.[7] Most interesting, they can also appear owing to quantum effects, in a quantum description of cosmology. An example that has been studied in some detail in recent years is the case of loop quantum cosmology (LQC).[9,10] In this formalism, the cosmological background spacetime is quantized adopting the techniques of loop quantum gravity,[11] a quantum version of GR that is based on the use of holonomies and fluxes of densitized triads, and that avoids any dependence of the results on fixed spacetime structures. On top of this quantum version of the cosmological background, matter and geometry fluctuations provide the cosmological perturbations that eventually serve as seeds for the anisotropies of the CMB and the formation of large-scale structures. Different approaches have been suggested to deal with these perturbations.[12–16] Among them, here we will consider only two of them, that share the property of preserving the ultraviolet structure of the dynamical equations for the perturbations in GR. The first one is called hybrid LQC, and is based on a simultaneous quantization of the background and the gauge-invariant degrees of freedom of the perturbations, using loop quantization techniques for the former and Fock techniques for the latter.[17–23] We will focus most of our analysis on this quantization proposal. The other approach that we will mention, but will not discuss in detail in our study, is called the dressed-metric proposal.[24–29] In this case, the gauge-invariant perturbations evolve in a kind of mean-field background that is dressed with quantum corrections, in order to take into account the most relevant effects of its quantization in LQC.

Remarkably, in the case of kinetic dominance, both in GR and in LQC, the changes with respect to the standard evolution of slow-roll inflation in Einstein's theory can be encapsulated in modifications of the time-dependent mass that appears in the dynamical equations of the gauge-invariant perturbations.[30] This modified mass captures the change in the behavior of the background as compared to the standard inflationary case. Essentially, these changes incorporate (at most) two characteristic scales: one is the scale of the onset of inflation away from slow roll; the other is the typical scale of the quantum phenomena, if such corrections are present.[22]

The main purpose of this work is to introduce a formalism capable to deal with the changes introduced by a kinetically dominated era in the predictions of the primordial spectrum of the cosmological perturbations. Employing that in this era the energy density of the inflaton is dominated by the kinetic term, the basic idea is to treat the effects of the inflaton potential just as a modification to the free case of a scalar field with vanishing potential, and include them only at leading order. In the next section, we present the theoretical framework within which we will develop this formalism. Section 3 describes the cosmological system that we will consider for our study. In Sec. 4 we provide the dynamical equations of the gauge invariant perturbations, including their modified mass. We then consider the issue of the choice of vacuum state for the perturbations in Sec. 5. The following sections are devoted to calculate the leading-order correction of the potential to the modified mass and the related change in the selection of vacuum state. Section 6 discusses what types of effects can be expected owing to the presence of a potential. Then, we compute these effects in: (a) the geometric terms that provide the modified mass, (b) the evolution of the background state, (c) the modified mass as a whole, combining the two previous types of effects, and (d) the choice of a vacuum. This is carried out, respectively, in Secs. 7, 8, 9, and 10. Finally, in Sec. 11 we discuss our results, and explain their possible use in future works.

2. The framework

We will consider cosmological perturbations in the framework of LQC and in the classical counterpart that corresponds to GR. We will restrict our discussion to the case of scalar perturbations, because tensor perturbations can be treated in a similar and in fact simpler way, since the inflaton potential directly affects the scalar matter content and hence the corresponding scalar perturbations.[19] On the other hand, vector perturbations are pure gauge in the absence of vector matter fields. The scalar perturbative gauge-invariant degrees of freedom of the geometry and the matter content, when the latter is given by a scalar field, can be described e.g. by the so-called Mukhanov-Sasaki (MS) invariant, which can be straightforwardly related to the comoving curvature perturbations for spatially flat cosmologies.[31–33] The Fourier modes of the MS gauge-invariant field satisfy a second-order differential evolution equation that can be written in the form of that of a harmonic oscillator, except for the addition of a time-dependent mass term to the squared frequency (or wavenorm) contribution. It is not difficult to derive the expression of this mass in GR, using cosmological perturbation theory. Moreover, in the considered approaches to LQC, it is possible to prove that the MS modes satisfy a dynamical equation which parallels that of GR except for the modification of the time-dependent mass. In the specific case of hybrid LQC, this modified mass turns out to be given by the ratio of two expectation values of background geometry operators, quantized with loop methods.[19,23] Those expectations values are calculated employing the inner product characteristic of LQC for homogeneous and isotropic spacetimes, that is

defined using the discrete measure.[9,10] This is the paradigmatic example that we will analyze in full detail in our discussion.

Nonetheless, the study can be extended without too many complications to the case of dressed-metric LQC, case in which the modified MS mass is also provided by expectation values of geometric operators, but the corresponding expression is different from the hybrid one.[30] Similarly, one can consider the case of hybrid geometrodynamics, in which the background geometry is quantized using geometrodynamics[34,35] and the perturbations using a Fock description, as in hybrid LQC. The modified mass is given again in this alternative quantization by a ratio of two expectation values, but in this case the representation of the geometric operators and the quantum integration measure over the background are not those of LQC, but of geometrodynamics.

An exact treatment of the modified mass in any of these quantization approaches would require numerical simulations to compute the evolution of the quantum state of the background, the action on it of the geometric operators, and the expectation value using the corresponding quantum measure. These numerical simulations would be quite intrincate, because the quantum dynamics in the presence of a potential for the inflaton is extremely nontrivial (actually, it is not even guaranteed that there exists a well-defined self-adjoint operator that generates the background evolution when the potential is not zero and generic). Furthermore, the fact that the quantum background state evolves in time implies that (most of) these numerical simulations have to be repeated at each instant of time in the considered epoch. For instance, the evolution can be described in terms of the inflaton, regarded as an internal time, since this is a good clock when the inflaton energy is kinetically dominated. The need to carry out the calculation of the expectation values for each value of interest of the inflaton complicates enormously the numerical challenges. In addition, the time description in terms of the inflaton has to be re-expressed in terms of a conformal time, which is the one that appears in the MS equation for the perturbations. This change of time involves the computation of additional expectation values,[19,23] as we will see in some detail later in our discussion. Besides, of course, all of these complicated numerical calculations need to be repeated if one changes the potential, e.g. if one wants to consider the effect of different families of potentials in the evolution of the primordial perturbations.

To complicate even more the situation, it is known that the choice of the Bunch-Davies state as a natural vacuum for the perturbations[36] is no longer sustainable away from slow roll.[22,37] The vacuum state should be optimally adapted to the background evolution. When this evolution does not reproduce a de Sitter regime, the validity of the Bunch-Davies vacuum is questionable, especially for modes that are able to feel the change of dynamics. To try to deal with this issue, at least in LQC, several proposals have recently been introduced for the choice of a vacuum of the perturbations. For instance, Ashtekar and Gupt have suggested to identify the vacuum as a state with a maximal classical behavior at the end of inflation

among those that, in the genuine quantum regime, have geometric uncertainties compatible with a quantum extension of Penrose's Weyl curvature hypothesis.[38,39] Another interesting proposal, introduced by Martín-de Blas and Olmedo,[20] is to choose a state with a non-oscillating power spectrum at the end of inflation. This proposal was motivated by the finding of states with rapid oscillations superimposed in the spectrum. Typically, an average of these oscillations over phenomenologically realistic frequency bins leads to an enhancement of the power. Recent works have shown that the desirable non-oscillating behavior can be characterized by a suitable asymptotic behavior in the sector of modes with infinitely large wavenumbers.[40,41] From this sector, one would determine the vacuum for all modes by properly extending the asymptotic solution. Moreover, it has been proved that this asymptotic behavior is displayed by positive frequency solutions associated with an asymptotic diagonalization of the Hamiltonian of the perturbations. Namely, by means of that diagonalization one straightforwadly finds a background-dependent frequency that defines the desired solutions, at least in the asymptotic ultraviolet sector. We will focus our discussion on this form of the proposal for the choice of a vacuum for the perturbations, since it can be handled analytically, opposite to the situation found for the alternative form in terms of the absence of oscillations in the spectrum, or for the proposal of Ashtekar and Gupt, which in a way or another require the use of numerical methods to pick out a vacuum solution.

As we have anticipated, the basic idea of our investigation is to approximate the MS mass around its value for a free scalar field and derive the leading-order correction to it owing to the presence of a small but nonzero potential term. This approach to the problem leads to a drastic simplification of the numerical calculations required for a complete study of the evolution of the perturbations, which in any case can be related to the free evolution, that is usually manageable in most of the quantization approaches (sometimes even analytically). Furthermore, as far as kinetic dominance persists in the studied period of time, the analysis can be carried out in a parallel way, if not at once, for any possible inflaton potential.

3. The system

The cosmological system that we will consider in this work is a perturbed flat Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime endowed with a scalar field as matter content. We start with a homogeneous and isotropic FLRW geometry with flat spatial sections and a homogeneous scalar field ϕ provided with a potential $W(\phi)$, which will play the role of an inflaton. In the absence of perturbations, or when their backreaction is negligible, this background is subject to a Hamiltonian constraint of the form[23,42]

$$\hat{\pi}_\phi^2 - \hat{H}_0^{(2)} = 0, \tag{1}$$

where $\hat{\pi}_\phi$ is the momentum operator of the inflaton and $\hat{H}_0^{(2)}$ contains all dependence on the FLRW geometry, namely[a]

$$\hat{H}_0^{(2)} = \left[\hat{H}_0^{(F)}\right]^2 - 2W(\phi)\hat{V}^2, \qquad H_0^{(F)} = 2\sqrt{3\pi G}|\pi_V|V. \tag{2}$$

We have set $\hbar = c = 1$, and G is Newton's constant. The operator $\hat{H}_0^{(F)}$ generates the free evolution, in the absence of potential, and can be interpreted as proportional to the generator of dilations. On the other hand, V is the physical volume of the spatial sections (in principle, we consider them compact, and take the noncompact limit at the end of the calculations about the dynamics of the perturbations[43]). Its momentum is denoted by π_V.

On this background, we introduce perturbations in the geometry and the scalar field, truncating the action of the whole system at quadratic perturbative order, which is the first nontrivial one. As we have said above, we restrict our discussion to scalar perturbations (in the continuous limit[43]) for simplicity. We adopt a canonical set of variables for the background and the perturbations in which the perturbative degrees of freedom correspond to the MS gauge invariant and its momentum, togeteher with a canonical set for the pure gauge sector.[19] The background is still described by suitable variables for the inflaton and the physical volume and their momenta. These variables differ from the original ones by quadratic perturbative corrections[19] but, in order to maintain our notation as simple as possible, we will not introduce new symbols for them and use the ones already introduced.

The resulting system has only one constraint that is not trivial to impose quantum mechanically: the spatial average, or zero-mode, of the Hamiltonian constraint[19] (truncated at quadratic order). We quantize this system and the constraint, and search for physical states by adopting an ansatz of separation of variables. The corresponding wavefunctions depend only on the MS gauge invariant, the background geometry (e.g. the volume), and the inflaton. With our ansatz, the dependence on the MS modes and the FLRW geometry separates in partial wavefunctions, that may all depend on the inflaton, used as a reference variable for the quantum evolution. Moreover, as part of the ansatz, the background state is taken to adopt the expression

$$\chi(V,\phi) = P\left[\exp\left(i\int_{\phi_0}^{\phi} d\tilde{\phi}\,\hat{H}_0(V,\tilde{\phi})\right)\right]\chi_0(V), \tag{3}$$

where $\chi_0(V)$ is the initial state of the FLRW geometry. In addition, the operator $\hat{H}_0$ is the square root of (the positive part of) $\hat{H}_0^{(2)}$. We will call it the background Hamiltonian. We can understand this operator as a(n approximate) generator of the background evolution. Finally, the symbol P denotes path ordering, since the

[a] Actually, we allow some freedom in the factor ordering for the representation of this operator. This will be useful when calculating its square root, to obtain a specially simple operator [see Eq. (14)].

integral in the exponential is defined over the inflaton, and the operator $\hat{H}_0$ depends on it.

Using this ansatz for the physical states, imposing the zero-mode of the Hamiltonian constraint, and assuming that the transitions between different background quantum states mediated by this constraint can be neglected, it is possible to derive a quantum constraint just on the perturbations, that depends on expectation values of geometric operators over the FLRW geometry of the background.[19,23] This constraint on the perturbations implies dynamical equations of the MS type, but with a background-dependent mass that is modified with respect to GR by including quantum effects via expectation values.

4. Modified MS equations

The coefficients of the Fourier expansion (in terms of sines and cosines, for instance) of the MS gauge-invariant field can be treated as time dependent variables $v_{\vec{k},\varepsilon}$, where ε is a dichotomic label that indicates the parity of the mode and $\vec{k}$ is the wavevector, restricted to have a positive first nonvanishing component (in an Euclidean fiducial spatial system of coordinates). We will call k the norm of this wavevector.

In hybrid quantum cosmology, according to the discussion of the previous section, the MS variables satisfy the following dynamical equations:

$$\ddot{v}_{\vec{k},\varepsilon} + \left[k^2 + \frac{\langle \hat{\vartheta}_e^q + (\hat{\vartheta}_o \hat{H}_0)_{sym}\rangle_\chi}{\langle \hat{\vartheta}_e\rangle_\chi}\right] v_{\vec{k},\varepsilon} = 0, \tag{4}$$

which rule their propagation on the FLRW (quantum) background. The dot denotes the derivative in conformal time η. We will call modified mass the term added to k^2 in the factor between square brackets, and denote it by $s(\eta)$. The geometric operators that appear in this modified mass and have not been defined yet are

$$\hat{\vartheta}_e = \frac{3}{2G}\hat{V}^{2/3}, \tag{5}$$

$$\hat{\vartheta}_o = \frac{3}{\pi}\sqrt{\frac{3}{\pi G}}\hat{V}^{2/3}(\hat{H}_0^{(F)})^{-1}\hat{\Lambda}_0^{(F)}(\hat{H}_0^{(F)})^{-1}\hat{V}^{2/3}W'(\phi), \tag{6}$$

$$\begin{aligned}\hat{\vartheta}_e^q = {} & \frac{1}{2\pi}\left[\hat{\frac{1}{V}}\right]^{1/3}\hat{H}_0^{(2)}\left(19 - 18(\hat{H}_0^{(F)})^{-2}\hat{H}_0^{(2)}\right)\left[\hat{\frac{1}{V}}\right]^{1/3} \\ & + \hat{V}^{4/3}\left(\frac{3}{8\pi^2 G}W''(\phi) - \frac{2}{\pi}W(\phi)\right).\end{aligned} \tag{7}$$

The norm of the operator $\hat{\Lambda}_0^{(F)}$ denotes an alternative representation of $\hat{H}_0^{(F)}$ in LQC, obtained with holonomies of double length. Besides, $\left[\hat{\frac{1}{V}}\right]$ is the so-called inverse volume operator of LQC,[9,23] which is not the inverse of the volume operator $\hat{V}$ at scales of the Planck volume or below. On the other hand, the prime denotes the derivative with respect to the inflaton.

The expectation values that appear in the modified mass have to be computed on the quantum state for the background FLRW geometry, of the form (3) according to our ansatz. In hybrid LQC, one uses for this calculation the discrete inner product characteristic of homogeneous and isotropic LQC. If one considers instead hybrid geometrodynamics (i.e. the combination of geometrodynamics and a Fock quantization of the perturbations), both the representation of all the involved geometric operators and the inner product for the expectation values must be those corresponding to a geometrodynamic quantum cosmology. In the case of the dressed-metric approach to LQC, on the other hand, the modified mass is slightly different, although defined again in terms of expectation values of certain geometric operators.[26] We will not give its expression here, but emphasize that can be handled in a similar way as we will explain for hybrid LQC. Finally, in GR, expectation values become evaluation on classical background trajectories, and operators become phase space functions, with certain subtleties that we will comment later in our analysis.

It is worth emphasizing that the modified MS equations described above dictate the evolution of the gauge-invariant scalar perturbations in conformal time. On the other hand, the evolution of the quantum state (3), as well as the explicit dependence on the potential W of the geometric operators, are in principle described in terms of the inflaton ϕ, that in the kinetic dominated regimes is a good reference variable for the quantum background dynamics. In the derivation of the modified MS equation,[19,23] one finds that the relation between the conformal time η and the inflaton as a dynamical parameter is given by the equation

$$\langle \hat{H}_0 \rangle_\chi d\eta = \langle \hat{\vartheta}_e \rangle_\chi d\phi. \tag{8}$$

This relation is state dependent, but it is well defined because $\hat{H}_0$ is a positive operator (a square root, in fact) and the expectation values translate any operator dependence into one on c-numbers (treating the inflaton as such, and realizing that its momentum does not appear in the considered equation).

5. Vacuum state

The fact that the dynamical equations of the (scalar) primordial gauge-invariant perturbations change with respect to those that would dictate the evolution on a de Sitter background, owing to the modifications suffered by the MS mass, calls for a change as well in the choice of vacuum for those perturbations, from the Bunch-Davies state adopted in the standard situations to a new state that is optimally adapted to the background dynamics. As we have commented in the Introduction, several proposals have been put forward for this choice of vacuum state in LQC. Here we will adhere to a prescription that is based on an asymptotic diagonalization of the Hamiltonian of the perturbations in the sector of infinitely large wavenorms, as a way to select a natural positive frequency in that sector, which can then be extended to all MS modes.[40] We have already commented that this prescription

has been shown to select a vacuum with non-oscillating primordial spectrum in the ultraviolet sector.

An important advantage of the asymptotic diagonalization prescription is that it can be handled in an (essentially) analytical way in the asymptotic ultraviolet sector. Let us see this in more detail. Once we are given a (modified) MS mass $s(\eta)$, we can determine a set of positive frequency solutions $\{\mu_k\}$ in the form

$$\mu_k = \frac{1}{\sqrt{-2\Im[h_k]}} e^{i\int_{\eta_0}^{\eta} d\tilde{\eta}\,\Im[h_k]}, \tag{9}$$

where h_k is a complex solution of the Riccati equation

$$\dot{h}_k = k^2 + s + h_k^2. \tag{10}$$

The positivity of the frequency would be guaranteed if and only if $\Im[h_k]$, i.e. the imaginary part of h_k, is negative. In the following, we will call h_k the complex frequency function. Note, by the way, that the above set of mode solutions depends only on the wavenorm k, and not on the specific wavevector $\vec{k}$ associated with it. The asymptotic diagonalization prescription selects complex frequencies with the following asymptotic behavior[41]:

$$\frac{1}{h_k} = \frac{i}{k}\left[1 + 2\sum_{n=0}^{\infty} \frac{\gamma_n}{(2ik)^{n+2}}\right], \tag{11}$$

where the coefficients γ_n are fixed by the recursion relation

$$\gamma_{n+1} = -\dot{\gamma}_n + 4s\left[\gamma_{n-1} + \sum_{l=0}^{n-3} \gamma_l \gamma_{n(l+3)}\right] - \sum_{l=0}^{n-1} \gamma_l \gamma_{n-(l+1)}, \tag{12}$$

starting with the data $\gamma_0 = s$. So, the complex frequency must behave asymptotically as $-ik$, with all subleading terms forming an asymptotic series in inverse powers of k, and with the coefficient γ_0 of the first term equal to the modified MS mass. This asymptotic behavior should characterize the complex frequency and hence the set of mode solutions that determine the vacuum state.

For completeness, recall that, if the solutions were valid until the end of inflation, the corresponding primordial power spectrum would be proportional to

$$P_v(k,\eta) = -\frac{k^3}{4\pi^2 \Im[h_k](\eta)}. \tag{13}$$

If our expressions for the mode solutions $\{\mu_k\}$ are applicable only in an interval that does not contain all the inflationary period, then we have to evolve their data until the end of inflation and compute the primordial spectrum there.

6. Corrections produced by the potential

In order to calculate the leading-order effect of the potential in the regime of kinetic dominance, let us first identify the different types of corrections that the potential

introduces in comparison to the case of a free evolution. First, the potential modifies the expression of the geometric operators that appear in the MS mass. Besides, it affects the evolution of the quantum background state on which the expectation values are computed, both because the generator of the evolution parametrized by the inflaton is different and because the relation between the inflaton and the conformal time that is used in the MS equation changes. Combining all these corrections one obtains the total change in the modified MS mass. This would complete the corrections to the dynamical equations of the perturbations in the kinetic regime, but we also have to consider the initial conditions for these dynamics, that are determined by the vacuum state. We have already explained that the natural choice of a vacuum changes as well with respect to the Bunch Davies state, away of slow roll, and that the new choice depends on the modified MS mass, and hence displays a dependence on the inflaton potential.

In the following sections, we will consider each of these types of corrections separately.

7. Corrections to the geometric operators in the MS mass

Let us consider the expression of the modified MS mass in Eq. (4). The corrections that the presence of a potential introduce in the geometric operators that appear in this expression are easy to calculate, taking into account that they are at most quadratic in this potential (and its derivatives), except for the dependence that may exist via the operator $\hat{H}_0$. In more detail, the operator $\hat{\vartheta}_e$ that gives the denominator is independent of the potential. Meanwhile, at leading order in this potential the operator $\hat{\vartheta}_e^q$ in the numerator contains a free and a linear contribution. Finally, the remaining operator $\hat{\vartheta}_o$, which is multiplied by $\hat{H}_0$ in the numerator, is linear. As a result, at the leading order in the potential that we are considering, we can approximate $\hat{H}_0$ by its free counterpart $\hat{H}_0^{(F)}$ in the numerator. With this replacement and the linear truncation of $\hat{\vartheta}_e^q$ in the potential, the expression of the modified mass is valid (as far as the geometric operators are concerned) at the studied order.

For later discussion, it is helpful to derive the expression of the background Hamiltonian, expanded in the potential. Clearly, the first term in this expansion is the free Hamiltonian $\hat{H}_0^{(F)}$. With a suitable choice of factor ordering, one can show that[30]

$$\hat{H}_0 = \hat{H}_0^{(F)} - W(\phi)(\hat{H}_0^{(F)})^{-1/2}\hat{V}^2(\hat{H}_0^{(F)})^{-1/2} + O(W^2), \tag{14}$$

where the symbol $O(W^2)$ denotes terms that are of quadratic or higher order in the potential and its derivatives.

8. Corrections to the background evolution

In order to calculate the effect of the potential in the quantum evolution of the background state, it is especially convenient to adopt an interaction picture,[44] in

which the dynamics of the free case is isolated and the contribution of the inflaton potential is viewed as an interaction. Approximating, to leading order in the potential, the interaction part of the background Hamiltonian according to Eq. (14) and expanding to the same order the evolution operator associated with this interaction, it is straightforward to see that

$$\chi(V,\phi) = \left[1 - i\int_{\phi_0}^{\phi} d\tilde{\phi}\,\hat{K}(\phi,\tilde{\phi})W(\tilde{\phi}) + O(W^2)\right]\chi^{(F)}(V,\phi), \tag{15}$$

$$\hat{K}(\phi,\tilde{\phi}) = \exp\left[i\hat{H}_0^{(F)}(\phi-\tilde{\phi})\right](\hat{H}_0^{(F)})^{-1/2}\hat{V}^2(\hat{H}_0^{(F)})^{-1/2}\exp\left[-i\hat{H}_0^{(F)}(\phi-\tilde{\phi})\right], \tag{16}$$

where $\chi^{(F)}(V,\phi)$ is the background state evolved with respect to the inflaton with the free Hamiltonian $\hat{H}_0^{(F)}$, namely

$$\chi^{(F)}(V,\phi) = \exp\left[i\hat{H}_0^{(F)}(\phi-\phi_0)\right]\chi_0(V). \tag{17}$$

To express this evolution in terms of the conformal time, we also need to calculate its relation with the inflaton up to the leading-order correction of the potential. Recall that this relation is given by $\langle\hat{H}_0\rangle_\chi d\eta = \langle\hat{\vartheta}_e\rangle_\chi d\phi$. Substituting the approximate expressions of $\hat{H}_0$ and χ that we have obtained, this equation for the conformal time leads first to the free relation

$$\eta^{(F)}(\phi) = \frac{3}{2G}\int_{\phi_0}^{\phi} d\tilde{\phi}\,\frac{\langle\hat{V}^{2/3}\rangle_{\chi^{(F)}}}{\langle\hat{H}_0^{(F)}\rangle_{\chi^{(F)}}}, \tag{18}$$

with inverse that we call $\phi^{(F)}(\eta)$, and next to a leading-order correction to this free term that we denote by $\eta^{(W)}(\phi)$. The expression of this correction is long but manageable. We do not reproduce it here because its explicit form is not relevant for our discussion.

In practice, we are interested in writing the inflaton as a function of the conformal time, rather than the opposite, because we want to obtain the evolution of the background state in terms of this time within our approximations. With this purpose, we can expand the functional time dependence of the inflaton as the free contribution plus corrections of the potential, in the form $\phi(\eta) = \phi^{(F)}(\eta)+\phi^{(W)}(\eta)+O(W^2)$. From our discussion above, it is then possible to show[30] that the leading-order correction to the inflaton dependence is just the negative of $\eta^{(W)}(\phi)$ mapped under the free relation $\phi^{(F)}$:

$$\phi^{(W)}(\eta) = -\phi^{(F)}\left(\eta^{(W)}\big(\phi^{(F)}(\eta)\big)\right). \tag{19}$$

9. Total correction to the MS mass

We can now combine the corrections to the geometric operators and the background state that are necessary to compute the modified MS mass. First, we note that at our leading order we can substitute the free relation between the inflaton and the conformal time in all geometric operators, since the the expression of the latter

was already linear in the potential as far as the numerator of the MS mass is concerned, whereas the operator in the denominator was free of any explicit inflaton dependence.

On the other hand, let us define the freely evolved background state in conformal time, instead of its counterpart in terms of the inflaton:

$$\tilde{\chi}^{(F)}(V,\eta) = \exp\left[i\hat{H}_0^{(F)}\left(\phi^{(F)}(\eta)-\phi_0\right)\right]\chi_0(V). \tag{20}$$

It is then straightforward to see that

$$\chi(V,\phi(\eta)) = \left[1+i\hat{J}^{(W)}(\eta)+O(W^2)\right]\,\tilde{\chi}^{(F)}(V,\eta), \tag{21}$$

$$\hat{J}^{(W)}(\eta) = \hat{H}_0^{(F)}\phi^{(W)}(\eta) - \int_{\phi_0}^{\phi^{(F)}(\eta)} d\tilde{\phi}\,\hat{K}\left(\phi^{(F)}(\eta),\tilde{\phi}\right)W(\tilde{\phi}). \tag{22}$$

With all these ingredients, we get an expansion for the modified MS mass of the form $s(\eta) = s^{(F)}(\eta) + s^{(W)}(\eta) + O(W^2)$. More especifically, the result is

$$s^{(F)} = \frac{G}{3\pi\langle\hat{V}^{2/3}\rangle_{\tilde{\chi}^{(F)}}}\langle\left[\hat{\frac{1}{V}}\right]^{1/3}(\hat{H}_0^{(F)})^2\left[\hat{\frac{1}{V}}\right]^{1/3}\rangle_{\tilde{\chi}^{(F)}} \tag{23}$$

for the free contribution to the mass, and

$$\begin{aligned} s^{(W)} = \frac{1}{\langle\hat{V}^{2/3}\rangle_{\tilde{\chi}^{(F)}}}\Bigg(&\frac{2\sqrt{3G}}{\pi\sqrt{\pi}}W'\langle(\hat{V}^{2/3}(\hat{H}_0^{(F)})^{-1}\hat{\Lambda}_0^{(F)}(\hat{H}_0^{(F)})^{-1}\hat{V}^{2/3}\hat{H}_0^{(F)})_{sym}\rangle_{\tilde{\chi}^{(F)}} \\ &+\Big(\frac{1}{4\pi^2}W''-\frac{4G}{3\pi}W\Big)\langle\hat{V}^{4/3}\rangle_{\tilde{\chi}^{(F)}}+\frac{34G}{3\pi}W\left\langle\left[\hat{\frac{1}{V}}\right]^{1/3}\hat{V}^2\left[\hat{\frac{1}{V}}\right]^{1/3}\right\rangle_{\tilde{\chi}^{(F)}} \\ &+\frac{iG}{3\pi}\left\langle\left[\left[\hat{\frac{1}{V}}\right]^{1/3}(\hat{H}_0^{(F)})^2\left[\hat{\frac{1}{V}}\right]^{1/3},\hat{J}^{(W)}\right]\right\rangle_{\tilde{\chi}^{(F)}} - is^{(F)}\langle[\hat{V}^{2/3},\hat{J}^{(W)}]\rangle_{\tilde{\chi}^{(F)}}\Bigg). \end{aligned} \tag{24}$$

In these equations, we have not displayed explicitly the functional dependence on the conformal time to simplify the expressions.

We can particularize these formulas to hybrid LQC by substituting the definitions of the involved geometric operators, that can be found in Refs.[23,30]. Similarly, the particularization to quantum geometrodynamics is quite straightforward, using the operator representation and integration measure over the volume (or the scale factor) that corresponds to this quantization. Several simplifications occur in this case, because the inverse volume operator is in fact the inverse of the operator that represents the volume, and no distinction has to be made between $|\hat{\Lambda}_0^{(F)}|$ and $\hat{H}_0^{(F)}$.

Finally, let us comment that our formulas provide as well the right result for kinetically dominated regimes in GR if the passage to the classical description is carried out correctly. For this, commutators of operators have to be translated into Poisson brackets of phase space functions using Dirac's rule, any other operator has

to be translated also into its phase space function counterpart, and the expectation values have to be regarded just as the evaluation on the background classical solution. With this correspondence, one recovers the classical relativistic equations.[30]

10. Corrections to the vacuum

As we have already seen, changes in the MS mass affect the complex frequency h_k that determines the positive frequency solution μ_k corresponding to the vacuum state in the asymptotic diagonalization prescription. We recall that this complex frequency satisfies the Ricatti equation $\dot{h}_k = k^2 + s + h_k^2$ in conformal time, with a fixed, specific asymptotic behavior for unboundedly large wavenorms. Similarly to what we have done with other functions in order to calculate their free part and the leading-order correction of the potential, we expand the complex frequency in the form $h_k = h_k^{(F)} + h_k^{(W)} + O(W^2)$. Then, linearizing the Ricatti equation, we obtain a first-order ordinary differential equation for $h_k^{(W)}$ that admits the general solution

$$h_k^{(W)}(\eta) = \left(C_k^{(W)} + \int_{\eta_0}^{\eta} d\tilde{\eta}\; s^{(W)}(\tilde{\eta}) e^{-I_{h_k}^{(F)}(\tilde{\eta})} \right) e^{I_{h_k}^{(F)}(\eta)}, \tag{25}$$

where

$$I_{h_k}^{(F)}(\eta) = 2 \int_{\eta_0}^{\eta} d\tilde{\eta}\, h_k^{(F)}(\tilde{\eta}) \tag{26}$$

and $C_k^{(W)}$ is any (possibly potential-dependent) constant for each given wavenorm k.

Integrating by parts the above expression and calling $d\tau_k = 2h_k^{(F)}(\eta)d\eta$, we obtain the following alternative form of this general solution:

$$h_k^{(W)}(\eta) = -\sum_{n=0}^{\infty} \frac{d^n}{d\tau_k^n} \left(\frac{s^{(W)}(\eta)}{2h_k^{(F)}(\eta)} \right) + D_k^{(W)} e^{I_{h_k}^{(F)}(\eta)}, \tag{27}$$

where $D_k^{(W)}$ is a constant for each k.

We still have to take into account the asymptotic behavior of the function h_k. With this aim, we expand its asymptotic coefficients γ_n around those of the free case in the form $\gamma_n = \gamma_n^{(F)} + \gamma_n^{(W)} + O(W^2)$. From the asymptotics of the total complex frequency h_k and of its counterpart for the free case with vanishing potential, $h_k^{(F)}$, we then obtain that the leading-order correction must have the asymptotic behavior

$$h_k^{(W)} = \sum_{n=0}^{\infty} \frac{\gamma_n^{(W)}}{(2ik)^{n+1}} \left[1 + 2 \sum_{n=0}^{\infty} \frac{\gamma_n^{(F)}}{(2ik)^{n+2}} \right]^{-2}, \tag{28}$$

with $\gamma_0^{(W)} = s^{(W)}$, given by the correction to the MS mass that we have computed in previous sections. The other coefficients $\gamma_n^{(W)}$ for $n \geq 1$ verify a recursion relation that can be derived from the linearization of the known relation for the coefficients γ_n, truncating it at leading order in the potential. For our discussion, the important

point about the asymptotic behavior (28) is that it does not contain any oscillating exponential (in fact, the expression can be reformulated as an asymptotic series in inverse powers of k). This result implies that the constant $D_k^{(W)}$ in Eq. (27) must vanish for the solution selected by our asymptotic condition. Hence, we conclude that (asymptotically at least)

$$h_k^{(W)}(\eta) = -\sum_{n=0}^{\infty} \frac{d^n}{d\tau_k^n} \left(\frac{s^{(W)}(\eta)}{2h_k^{(F)}(\eta)} \right). \tag{29}$$

In this way, we have succeeded in obtaining the expression of the leading-order correction to the complex frequency function, and therefore in getting the ingredients necessary to characterize the vacuum state selected by the asymptotic diagonalization prescription at our order of truncation in the potential and its derivatives.

11. Conclusions

In this work, we have considered kinetically dominated regimes in the very early evolution of the Universe. These regimes are attracting an increasing attention, both within GR and within quantum cosmology, as a mechanism to produce modifications in the CMB power spectra and other related cosmological quantities that can alleviate the possible anomalies that have been claimed in the observations, especially when this period of kinetic dominance in the energy density of the scalar field is accompanied with a later epoch of short lived inflation. In general, intrincate numerical computations are needed to study these kinetically dominated regimes previous to the slow-roll phase.

The situation is particularly complicated in quantum cosmology, where the background is described quantum mechanically. In many practical situations, this is one of the reasons why most of the analyses carried out so far in LQC have concentrated on the consideration of quantum background states that are peaked on *effective* trajectories,[9] described in the purely homogeneous and isotropic case by a corrected Hamiltonian constraint (with respect to that of GR) that incorporates effects of the quantum geometry. One of the effects that these trajectories incorporate is the existence of the renowned quantum bounce that avoids the cosmological singularity.[9, 34, 42] For this kind of states, expectation values are replaced with evaluation on the peak trajectories. However, it is not totally clear if this procedure is valid for any kind of geometric operator, like those that appear in the modified MS mass, which are highly nonlinear in the basic variables in which the peak trajectories have been confirmed. Besides, one cannot analyze in this way background states with a more genuine quantum behavior. Both issues are relevant because the modifications in the MS mass produced by the loop quantization are only significant in the regions with important quantum geometry corrections, that would correspond to the bounce in the case of the commented effective description. So, the interest in performing a fully quantum analysis in those regions is clear. In the situations that are phenomenologically appealing, the evolution is indeed kinetically dominated in those

regions. A formalism like the one that we have presented, which allows a quantum investigation employing approximations that are based on the kinetic dominance, may therefore be of great help. Notice also that, even if one accepts a simplification of the problem by using some kind of effective description for the background, the discussion would need to be carried out separately for each inflaton potential, with a complicated dynamics that would almost certainly require numerical calculations at some step of the analysis, and that would have to be repeated for any different potential that one considers.

Instead of facing any of the aforementioned, involved and time-consuming, numerical calculations, we have developed a formalism that allows us to study the problem starting from the dynamics of a free scalar field including the effects of a potential as corrections, with an applicability restricted to regimes with kinetic dominance and hence away of the period of slow-roll inflation. More especifically, we have analytically studied the leading-order corrections caused by the inflaton potential, focusing our attention on the case of a hybrid quantization in cosmology, which combines a quantization of the FLRW background with a Fock quantization of the perturbations. These corrections of the potential modify the geometric quantities (operators in general) that provide the modified MS mass in the dynamical equations for the perturbations. They also affect the evolution of the background state, both because they modify the generator of its evolution with respect to the inflaton (that can be used in kinetic dominance as an internal reference variable for the dynamics) and because the relation between the inflaton and the conformal time is changed. More subtle is the effect on the vacuum of the perturbations. Away from slow roll, the choice of a Bunch-Davies state does not seem justified anymore. We have adopted a prescription for the choice of a vacuum that rests on an asymptotic diagonalization of the Hamiltonian of the perturbations, and which has been seen to lead to an asymptotically non-oscillating primordial power spectrum. The vacuum selected with this prescription is characterized by a set of positive frequency solutions, determined by a complex function, which we have called the complex frequency. This function satisfies a Ricatti equation and has to fulfill certain asymptotic conditions. Since this Ricatti equation and also the asymptotics depend on the modifed MS mass, the presence of a potential produces corrections in them, and therefore on the complex frequency, altering the choice of vaccum state.

We have succeeded in calculating all of these leading-order corrections, providing in this way all the ingredients needed to study at that order the evolution of the primordial perturbations and the state that fixes their initial conditions. It is worth remarking that our results can be generalized to other prescriptions in quantum cosmology other than a hybrid quantization. For instance, we have commented that the dressed-metric approach to LQC can be studied in a similar way, with the only difference that the expression of the modified MS mass is not the one that we have considered in detail here.[26] Besides, we have also seen that the results are applicable to the case of GR, with certain rules for the translation of our formulas into the

classical case. On the other hand, our results have been derived at leading order in the corrections produced by the potential, but in principle there is no obstruction to extend the analysis to higher orders. The expressions will become much more intrincate, but there is no conceptual obstacle to obtain those corrections, increasing order by order the accuracy of the formalism.

Our treatment opens the possibility of comparing the effects of different potentials without the need of performing explicit calculations for each of them spearately. This is so because the potential appears in our formulas for the corrections in an abstract way, without any requirement about it apart from the validity of the kinetic dominance. It would be very interesting to use this approach and study the different effects that distinct potentials can produce in the primordial power spectra of the perturbations, left as imprints of a kinetically dominated epoch previous to slow roll. Finally, our results make possible the investigation of another very interesting issue, which is extremely relevant to ellucidate whether quantum effects can be detected or not in cosmological observations. This issue is the question of whether the effects of a classical fast-roll period (i.e. an inflationary period that does not satisfy the slow-roll conditions) can be differentiated from the effects of a quantization of the geometry. This question is crucial to answer whether the last of these types of effects has any hope to be observed. Since our formalism is applicable to kinetically dominated regimes both in GR and in the relevant quantum regions of LQC (and other quantum cosmology proposals), it can be used to analyze the effects of these distinct dynamics of the perturbations, studying the differences in the imprints that each of them may eventually leave on the spectra of the primordial perturbations.

Acknowledgments

This work was partially supported by Project. No. MICINN FIS2017-86497-C2-2-P and Project. No. MICINN PID2020-118159GB-C41 from Spain. B.E.N. acknowledges financial support from the Standard Program of JSPS Postdoctoral Fellowships for Research in Japan. The authors are grateful to L. Castelló Gomar, A. García-Quismondo, L. Morala, and S. Prado for discussions.

References

1. B.J.T. Jones, *Precision Cosmology: The First Half Million Years* (Cambridge University Press, Cambridge, England, 2017).
2. E. Di Valentino, A. Melchiorri, and J. Silk, Planck evidence for a closed Universe and a possible crisis for cosmology, *Nature Astron.* **4**, 196 (2019).
3. A. Ashtekar, B. Gupt, D. Jeong, and V. Sreenath, Alleviating the tension in CMB using Planck-scale physics, *Phys. Rev. Lett.* **125**, 051302 (2020).
4. A. Ashtekar, B. Gupt, and V. Sreenath, Cosmic tango between the very small and the very large: Addressing CMB anomalies through Loop Quantum Cosmology, *Front. Astron. Space Sci.* **8**, 685288 (2021).

5. I. Agullo, D. Kranas, and V. Sreenath, Large scale anomalies in the CMB and non-Gaussianity in bouncing cosmologies, *Class. Quantum Grav.* **38**, 065010 (2021).
6. I. Agullo, D. Kranas, and V. Sreenath, Anomalies in the Cosmic Microwave Background and their non-Gaussian origin in Loop Quantum Cosmology, arXiv:2105.12993 (2021).
7. C.R. Contaldi, M. Peloso, L. Kofman, and A.D. Linde, Suppressing the lower multipoles in the CMB anisotropies, *JCAP* **0307**, 002 (2003).
8. I. Agullo and N.A. Morris, Detailed analysis of the predictions of Loop Quantum Cosmology for the primordial power spectra, *Phys. Rev. D* **92**, 124040 (2015).
9. A. Ashtekar and P. Singh, Loop Quantum Cosmology: A status report, *Class. Quantum Grav.* **28**, 213001 (2011).
10. G.A. Mena Marugán, A brief introduction to Loop Quantum Cosmology, *AIP Conf. Proc.* **1130**, 89 (2011).
11. T. Thiemann, *Modern Canonical Quantum General Relativity* (Cambridge University Press, Cambridge, England, 2007).
12. E. Wilson-Ewing, Testing Loop Quantum Cosmology, *Comptes Rendus Phys.* **18**, 207 (2017).
13. M. Bojowald, G. Calcagni, and S. Tsujikawa, Observational constraints on Loop Quantum Cosmology, *Phys. Rev. Lett.* **107**, 211302 (2011).
14. T. Cailleteau, L. Linsefors, and A. Barrau, Anomaly-free perturbations with inverse-volume and holonomy corrections in Loop Quantum Cosmology, *Class. Quantum Grav.* **31**, 125011 (2014).
15. A. Barrau, M. Bojowald, G. Calcagni, J. Grain, and M. Kagan, Anomaly-free cosmological perturbations in effective canonical quantum gravity, *JCAP* **05**, 051 (2015).
16. S. Schander, A. Barrau, B. Bolliet, J. Grain, L. Linsefors, and J. Mielczarek, Primordial scalar power spectrum from the Euclidean Big Bounce, *Phys. Rev. D* **93**, 023531 (2016).
17. M. Fernández-Méndez, G.A. Mena Marugán, and J. Olmedo, Hybrid quantization of an inflationary universe, *Phys. Rev. D* **86**, 024003 (2012).
18. M. Fernández-Méndez, G.A. Mena Marugán, and J. Olmedo, Hybrid quantization of an inflationary model: The flat case, *Phys. Rev. D* **88**, 044013 (2013).
19. L. Castelló Gomar, M. Martín-Benito, and G.A. Mena Marugán, Gauge-invariant perturbations in Hybrid Quantum Cosmology, *JCAP* **06**, 045 (2015).
20. D. Martín de Blas and J. Olmedo, Primordial power spectra for scalar perturbations in Loop Quantum Cosmology, *JCAP* **06**, 029 (2016).
21. L. Castelló Gomar, G.A. Mena Marugán, D. Martín de Blas, and J. Olmedo, Hybrid Loop Quantum Cosmology and predictions for the Cosmic Microwave Background, *Phys. Rev. D* **96**, 103528 (2017).
22. B. Elizaga Navascués, D. Martín de Blas, and G.A. Mena Marugán, The vacuum state of primordial fluctuations in Hybrid Loop Quantum Cosmology, *Universe* **4**, 98 (2018).
23. B. Elizaga Navascués and G.A. Mena Marugán, Hybrid Loop Quantum Cosmology: An overview, *Front. Astron. Space Sci.* **8**, 624824 (2021).
24. I. Agullo, A. Ashtekar, and W. Nelson, A quantum gravity extension of the inflationary scenario, *Phys. Rev. Lett.* **109**, 251301 (2012).
25. I. Agullo, A. Ashtekar, and W. Nelson, Extension of the quantum theory of cosmological perturbations to the Planck era, *Phys. Rev. D* **87**, 043507 (2013).
26. I. Agullo, A. Ashtekar, and W. Nelson, The pre-inflationary dynamics of Loop Quantum Cosmology: Confronting quantum gravity with observations, *Class. Quantum Grav.* **30**, 085014 (2013).

27. I. Agullo, Loop Quantum Cosmology, non-Gaussianity, and CMB power asymmetry, *Phys. Rev. D* **92**, 064038 (2015).
28. I. Agullo, A. Ashtekar, and B. Gupt, Phenomenology with fluctuating quantum geometries in Loop Quantum Cosmology, *Class. Quantum Grav.* **34**, 074003 (2017).
29. I. Agullo, B. Bolliet, and V. Sreenath, Non-Gaussianity in Loop Quantum Cosmology, *Phys. Rev. D* **97**, 066021 (2018).
30. B. Elizaga Navascués, R. Jiménez-Llamas, and G.A. Mena Marugán, Primordial perturbations in kinetically dominated regimes of general relativity and hybrid quantum cosmology, arXiv:2106.05628 (2021).
31. V. Mukhanov, Quantum theory of gauge-invariant cosmological perturbations, *Zh. Eksp. Teor. Fiz.* **94**, 1 (1998) [*Sov. Phys. JETP* **67**, 1297 (1988)].
32. M. Sasaki, Gauge invariant scalar perturbations in the new inflationary universe, *Prog. Theor. Phys.* **70**, 394 (1983).
33. H. Kodama and M. Sasaki, Cosmological perturbation theory, *Prog. Theor. Phys. Suppl.* **78**, 1 (1984).
34. A. Ashtekar, T. Pawłowski, and P. Singh, Quantum nature of the Big Bang: An analytical and numerical investigation, *Phys. Rev. D* **73**, 124038 (2006).
35. J.J. Halliwell, *Introductory lectures to quantum cosmology*, published in *Proceedings of the 1990 Jerusalem Winter School on Quantum Cosmology and Baby Universes*, eds. S. Coleman, J.B. Hartle, T. Piran, and S. Weinberg (World Scientific, Singapore, 1991).
36. T.S. Bunch and P. Davies, Quantum field theory in de Sitter space: Renormalization by point splitting, *Proc. R. Soc. Lond. A* **360**, 117 (1978).
37. B. Elizaga Navascués and G.A. Mena Marugán, Analytical investigation of preinflationary effects in the primordial power spectrum: From General Relativity to Hybrid Loop Quantum Cosmology, arXiv:2104.15002 (2021).
38. A. Ashtekar and B. Gupt, Initial conditions for cosmological perturbations, *Class. Quantum Grav.* **34**, 035004 (2017).
39. A. Ashtekar and B. Gupt, Quantum gravity in the sky: Interplay between fundamental theory and observations, *Class. Quantum Grav.* **34**, 014002 (2017).
40. B. Elizaga Navascués, G.A. Mena Marugán, and T. Thiemann, Hamiltonian diagonalization in Hybrid Quantum Cosmology, *Class. Quantum Grav.* **36**, 18 (2019).
41. B. Elizaga Navascués, G.A. Mena Marugán, and S. Prado, Non-oscillating power spectra in Loop Quantum Cosmology, *Class. Quantum Grav.* **38**, 035001 (2021).
42. A. Ashtekar, T. Pawłowski, and P. Singh, Quantum nature of the Big Bang: Improved dynamics, *Phys. Rev. D* **74**, 084003 (2006).
43. B. Elizaga Navascués and G. A. Mena Marugán, Perturbations in quantum cosmology: The continuum limit in Fourier space, *Phys. Rev. D* **98**, 103522 (2018).
44. A. Galindo and P. Pascual, *Quantum Mechanics II* (Springer, Berlin, 1991).

Revisiting the Hamiltonian formalism of the Ashtekar–Olmedo–Singh black hole model

Alejandro García-Quismondo
Instituto de Estructura de la Materia, IEM-CSIC, Serrano 121, 28006 Madrid, Spain
alejandro.garcia@iem.cfmac.csic.es

Guillermo A. Mena Marugán
Instituto de Estructura de la Materia, IEM-CSIC, Serrano 121, 28006 Madrid, Spain
E-mail: mena@iem.cfmac.csic.es

We reexamine the computation of the equations of motion of the Ashtekar–Olmedo–Singh black hole model in order to establish whether it is possible to construct a Hamiltonian formalism such that parameters that regulate the introduction of quantum geometry effects are suitably treated as true constants of motion. After suggesting that these parameters should capture contributions from two sectors of phase space that had been considered independent in the literature, we proceed to obtain the corresponding dynamical equations and investigate the repercussions of this more general choice. We restrict our discussion exclusively to these dynamical issues. We also analyze whether the proposed procedure can be reconciled with the results of Ashtekar, Olmedo, and Singh, at least in some limit.

Keywords: Loop Quantum Cosmology, Black Holes, AOS Model.

1. Introduction

Almost three years ago, Ashtekar, Olmedo, and Singh proposed a new model to describe black hole spacetimes within the framework of loop quantum cosmology[1–3] (from now on, referred to as the AOS model for short). This model is set apart from prior related investigations[4–39] for a number of reasons, including the fact that it is claimed that its physical predictions neither depend on fiducial structures nor display large quantum corrections in low curvature regions (which count among the problems that previous studies suffered from). By means of the inclusion of quantum geometry effects, which is regulated by two polymerization parameters,[a] the classical central singularity is found to be cured. In this model, it is replaced by a transition surface between a trapped region and an anti-trapped one and the metric, that we will call effective in the sense that it can be treated classically but incorporates quantum effects, is extended to encompass what can be regarded as a white horizon. This effective geometry is shown to be smooth and, in fact, the curvature invariants turn out to admit universal upper bounds that are independent of the mass of the black hole under consideration.[1–3] The authors of the model completed their

[a]Here, the term "polymerization" refers to the name "polymer quantization", which is often used for the quantization of symmetry reduced models with loop techniques.

investigation by formulating an analogous description of the exterior region that can be matched smoothly with that of the black hole interior, resulting in an effective extension of the whole Kruskal spacetime.

One of the defining features of the AOS model is their choice of polymerization parameters. In order to overcome the issues present in earlier works, they proposed a procedure in which the parameters are identified with certain Dirac observables (i.e., quantities that are constant along any given dynamical trajectory but not on the whole phase space), defined by minimal area conditions on the transition surface. Nonetheless, the parameters are treated as constants throughout the Hamiltonian derivation of the equations of motion and only once they have been obtained and solved does the aforementioned identification occur, whereby the parameters are fixed to the value of certain functions of the ADM mass, which is a Dirac observable itself.[b] This was already pointed out by Bodendorfer, Mele, and Münch. In Ref. 40, they showed that an extra phase space dependent factor would appear in the dynamical equations should the parameters be treated as genuine constants of motion from the very beginning. For this, they employed that, given a certain choice of lapse function, the Hamiltonian constraint of the system is essentially given by the difference of two Dirac observables, the on-shell values of which coincide and become proportional to the mass of the black hole. This simple structure was exploited to divide the phase space into two independent sectors (associated with the radial and angular degrees of freedom, respectively), where the dynamics is generated by the previously mentioned Dirac observables, which can be thought of as partial Hamiltonians. Additionally, these partial Hamiltonians were assumed to serve as polymerization parameters,[40] in the sense that each parameter was chosen to be a function of only its associated partial Hamiltonian. As a result, by evaluating the study on shell, one makes contact with the original proposal of Refs. 1–3 to the extent that the on-shell parameters are functions of the mass of the black hole.

Nevertheless, given the fact that the two partial Hamiltonians have coincident values on the constraint surface, there should be no way of telling apart their on-shell contributions. Consequently, one could argue that a more general choice of polymerization parameters would be such that each of them could depend on both partial Hamiltonians, a fact that would imply that the decoupling of phase space sectors that plays a central role in previous analyses would be broken both at the Hamiltonian and dynamical levels. In the following, we will focus on analyzing whether an alternative procedure to perform the Hamiltonian calculation can be constructed starting from this observation. We will ignore for the time being issues concerning the asymptotic behavior of the effective metric and its physical interpretation and quantum covariance, which lie beyond the scope of this work (see Refs. 41, 42, 43, and 44 for details on the criticisms that the model has received). The main objective

[b]The authors of Refs. 1 and 2 presented an argument supporting this procedure that involves an extension of phase space. See Appendix A of Ref. 2 for further details.

of our investigation is to explore the possibility that an alternative dynamical analysis where the polymerization parameters depend on both partial Hamiltonians can be developed and to ascertain whether such an analysis can reconcile the derivation of the solution in Refs. 1–3 with a genuine treatment of the parameters as dynamical constants.

The rest of this article[c] is organized as follows. In Sec. 2, after a brief introduction of the main elements of the AOS model, we discuss the selection of polymerization parameters, including our new proposal. In addition, we address the derivation of the dynamical equations that follow from our choice in Sec. 2.2. In Sec. 3, we delve into certain time redefinitions that lead our equations of motion to adopt a simpler form, paying the price of employing two different time variables instead of one. We also examine whether the existing freedom in the model may allow us to set both times equal. In Sec. 4, we extend this investigation to the asymptotic limit of large masses and study the relation between both time variables for finite values of the mass. Finally, in Sec. 5, we summarize the main ideas of this work and discuss our results. Throughout this paper, we set the speed of light and the reduced Planck constant equal to one.

2. The model

In this section, we briefly introduce the main ingredients of the effective model under consideration,[1–3] followed by a comment on the choice of polymerization parameters in Sec. 2.1 and the derivation of the equations that govern the dynamics of the system in Sec. 2.2. For further details, we refer the reader to the original works.

In this paper, we focus our attention on the interior region of a Schwarzschild black hole, which can be foliated by spatial Cauchy hypersurfaces (namely those characterized by $r = \text{const}$, where r is the Schwarzschild radial coordinate). Furthermore, these spatial hypersurfaces are homogeneous, a property that permits a finite dimensional Hamiltonian description of the system. Indeed, the associated phase space is four dimensional and can be coordinatized by two connection variables b and c which encode the degrees of freedom of the Ashtekar connection, and two triad variables p_b and p_c which encode the degrees of freedom of the densitized triad. These form two canonical pairs with Poisson brackets[1,2]

$$\{b, p_b\} = G\gamma, \qquad \{c, p_c\} = 2G\gamma, \tag{1}$$

where G is the Newtonian gravitational constant and γ is the Immirzi parameter. Given that the first one contains the dynamical information about the radial direction and the second one the dynamical information about the angular directions, we will refer to these pairs and the sectors of phase space that they define as "radial" and "angular", respectively.

[c]The present manuscript is based on a previous publication of ours,[45] although we here provide a restructured and updated view on the subject.

By virtue of the symmetries inherited from general relativity, these variables are bound by a constraint, namely the Hamiltonian of the system, H. With the choice

$$N = \frac{\gamma\delta_b\sqrt{|p_c|}}{\sin\delta_b b}, \tag{2}$$

the product of the lapse function and the Hamiltonian is given by[1,2]

$$NH = \frac{L_o}{G}(O_b - O_c), \tag{3}$$

$$O_b = -\frac{1}{2\gamma}\left(\frac{\sin\delta_b b}{\delta_b} + \frac{\gamma^2\delta_b}{\sin\delta_b b}\right)\frac{p_b}{L_o}, \tag{4}$$

$$O_c = \frac{1}{\gamma}\frac{\sin\delta_c c}{\delta_c}\frac{p_c}{L_o}, \tag{5}$$

where L_o is a fiducial length introduced to avoid divergences when integrating homogeneous quantities over space. In addition, δ_b and δ_c are the polymerization parameters that capture and introduce the effects of quantum geometry in the system. Should we take the limit where these parameters vanish, the classical Hamiltonian would be recovered. Notice that, if no cross-dependence is introduced through the δ-parameters, O_b and O_c only depend on the radial and angular sector, respectively. Therefore, up to multiplicative constants, these two objects would generate the dynamics in their respective sectors of phase space. As such, we refer to them as "partial Hamiltonians", although we emphasize that this interpretation is only valid when the polymerization parameters depend at most on the canonical variables of their respective sectors. In the following subsection, we discuss how to select these parameters.

2.1. *Polymerization parameters as constants of motion*

Let us briefly go over the main idea underlying the discussed identification and handling of the polymerization parameters suggested by the authors of the original model, mentioning an alternative related proposal that was put forward in a later work as well. Afterwards, we argue that, on the light of the dynamical properties of the system, a more general choice could be made, which we adopt for the subsequent analysis.

As already mentioned in the Introduction, one of the main features of the AOS model that set it apart from previous investigations is the way in which one deals with the polymerization parameters. Instead of taking the δ-parameters as constants on the whole phase space or as arbitrary functions of phase space, the authors of Refs. 1 and 2 proposed to tread an intermediate path: the model is based on selecting these parameters in such a way that they are eventually identified as Dirac observables, i.e. constant along any given dynamical trajectory. More concretely, they proposed a procedure where δ_b and δ_c are treated as constants throughout the Hamiltonian derivation of the dynamical equations and fixed to the value of certain Dirac observables only at the very end. These Dirac observables are given

by functions of the mass of the black hole determined by minimal area conditions at the transition surface in the limit of large masses. For an argument supporting the validity of this procedure, we encourage the reader to consult Ref. 2 (and in particular its appendices).

In a later work,[40] this approach was criticized arguing that the attained dynamical equations are not the ones derived from their proposed Hamiltonian if the parameters were to be treated as constants of motion. Instead, an alternative proposal was put forward to use the partial Hamiltonians themselves as polymerization parameters in their respective sectors of phase space. In other words, the authors of Ref. 40 suggested to take each parameter as a function of its associated partial Hamiltonian. This idea is based on the fact that both partial Hamiltonians are Dirac observables in the AOS model. Furthermore, their values coincide on shell (that is, on the phase space surface on which the Hamiltonian constraint vanishes) and are proportional to the mass of the black hole

$$O_b|_{\mathrm{on-shell}} = O_c|_{\mathrm{on-shell}} = m. \tag{6}$$

With this choice, the nontrivial nature of the parameters as functions of phase space is taken into account in the Hamiltonian calculation from the very beginning, leading to contributions coming from the nonvanishing Poisson brackets of δ_b and δ_c which were absent in the original AOS dynamical equations.

Given the attention that the AOS model has drawn in the recent years and the interesting physical results derived from it, it seems to us that it is worth investigating whether these two approaches can be reconciled in some way (and if so, to which extent) in order to gain confidence on the meaningfulness of the physical picture that ensues from it. This is precisely the aim of our work and the intention behind the preliminary computation presented in the next sections. In order to address this question, we follow the ideas of Ref. 40, or rather an extension of them.

Our proposal is based on a simple observation. As we commented in the previous paragraphs, by virtue of the structure of the Hamiltonian constraint, both partial Hamiltonians have the exact same on-shell value. Therefore, we should in principle not be able to tell apart their contributions to the on-shell polymerization parameters. As such, we argue that the most general choice of this form should be

$$\delta_i = f_i(O_b, O_c), \tag{7}$$

where $i = b, c$ and f_i are functions of both O_b and O_c.[d] Moreover, we require that the on-shell value of our parameters coincide with those considered in Refs. 1–3,

[d]From now on, we will use this notation to refer to the radial and angular sectors indistinctly. Thus, letters from the middle of the Latin alphabet $(i, j, ...)$ take values on the set $\{b, c\}$; so that i, p_i, O_i, and δ_i denote the connection variable, the triad variable, the partial Hamiltonian, and the polymerization parameter associated with the radial or angular sector, depending on the value of the label i.

namely

$$\delta_b = \left(\frac{\sqrt{\Delta}}{\sqrt{2\pi}\gamma^2 m}\right)^{1/3} + o\left(m^{-1/3}\right), \quad \delta_c = \frac{1}{2L_o}\left(\frac{\gamma\Delta^2}{4\pi^2 m}\right)^{1/3} + o\left(m^{-1/3}\right), \tag{8}$$

where $o(\cdot)$ denotes the presence of terms that are subdominant with respect to the quantity in parenthesis. The inclusion of these subdominant terms has to do with the fact that the δ-parameters of the AOS model are derived in the limit of large black hole masses, where two relevant solutions to the minimal area conditions that define them in the first place coincide (see Appendix B of Ref. 2 for further details).

Notice that the proposed functional form of these parameters introduces a cross-dependence between the radial and angular sectors of phase space, which breaks the decoupling of these sectors that had been present in the literature until now. In the next subsection, we explore the immediate consequences of this decoupling as regards the derivation of the equations that govern the dynamics of the system.

2.2. *Dynamical equations*

Let us address the derivation of the equations of motion that follow from the effective Hamiltonian of the AOS model if the polymerization parameters δ_i are selected as discussed in Sec. 2.1. Since the time derivative of any phase space function can be obtained by means of its Poisson bracket with the Hamiltonian, it is immediate to see that

$$\partial_t i = s_i \frac{L_o}{G}\left(\{i, O_i\} - \{i, O_j\}\right), \tag{9}$$

where we recall that $i, j = b, c$. Besides, here $j \neq i$, t is the time variable associated with the lapse function N [see Eq. (2)], and s_i is a sign defined by

$$s_b = +1, \qquad s_c = -1. \tag{10}$$

Let us now compute $\{i, O_i\}$ and $\{i, O_j\}$. The first of these brackets is given by

$$\{i, O_j\} = \{i, p_i\}\frac{\partial O_i}{\partial p_i} + \frac{\partial O_i}{\partial \delta_i}\left(\frac{\partial f_i}{\partial O_i}\{i, O_i\} + \frac{\partial f_i}{\partial O_j}\{i, O_j\}\right). \tag{11}$$

In a similar way,

$$\{i, O_j\} = \frac{\partial O_j}{\partial \delta_j}\left(\frac{\partial f_j}{\partial O_i}\{i, O_i\} + \frac{\partial f_j}{\partial O_j}\{i, O_j\}\right). \tag{12}$$

From the form of these two expressions, we realize that Eqs. (11) and (12) constitute a system of equations linear in $\{i, O_i\}$ and $\{i, O_j\}$. Recasting this system in an appropriate manner, we have

$$\begin{pmatrix} 1-\Delta_{ii} & -\Delta_{ij} \\ -\Delta_{ji} & 1-\Delta_{jj} \end{pmatrix}\begin{pmatrix} \{i, O_i\} \\ \{i, O_j\} \end{pmatrix} = \begin{pmatrix} \{i, p_i\}\dfrac{\partial O_i}{\partial p_i} \\ 0 \end{pmatrix}, \tag{13}$$

where we have defined

$$\Delta_{ij} = \frac{\partial O_i}{\partial \delta_i} \frac{\partial f_i}{\partial O_j}. \tag{14}$$

This system can be solved provided that the determinant of the first matrix does not vanish,

$$(1 - \Delta_{ii})(1 - \Delta_{jj}) - \Delta_{ij}\Delta_{ji} \neq 0. \tag{15}$$

From now on, we assume that this invertibility condition holds. Then, we find that the Poisson brackets that enter the equations of motion are given by

$$\begin{pmatrix} \{i, O_i\} \\ \{i, O_j\} \end{pmatrix} = \frac{\{i, p_i\} \dfrac{\partial O_i}{\partial p_i}}{(1 - \Delta_{ii})(1 - \Delta_{jj}) - \Delta_{ij}\Delta_{ji}} \begin{pmatrix} 1 - \Delta_{jj} \\ \Delta_{ji} \end{pmatrix}, \tag{16}$$

which results in

$$\partial_t i = \frac{1 - \Delta_{jj} - \Delta_{ji}}{(1 - \Delta_{ii})(1 - \Delta_{jj}) - \Delta_{ij}\Delta_{ji}} \left[s_i \frac{L_o}{G} \{i, p_i\} \frac{\partial O_i}{\partial p_i} \right], \tag{17}$$

with $j \neq i$. Following the same procedure with the triad variables, their corresponding dynamical equations turn out to be

$$\partial_t p_i = \frac{1 - \Delta_{jj} - \Delta_{ji}}{(1 - \Delta_{ii})(1 - \Delta_{jj}) - \Delta_{ij}\Delta_{ji}} \left[-s_i \frac{L_o}{G} \{i, p_i\} \frac{\partial O_i}{\partial i} \right], \tag{18}$$

where, again, $j \neq i$.

It is interesting to note that, in both cases, the expressions in square brackets are the equations of motion that are attained if the parameters are treated as constant numbers (and, thus, the ones considered in Refs. 1–3). These appear rescaled via a phase space dependent factor

$$C_{ij} = \frac{1 - \Delta_{jj} - \Delta_{ji}}{(1 - \Delta_{ii})(1 - \Delta_{jj}) - \Delta_{ij}\Delta_{ji}}, \tag{19}$$

which is *a priori* different in each sector. Therefore, we conclude that these C-factors that multiply the AOS dynamical equations contain all the information about the nontrivial functional form of the polymerization parameters. Indeed, if we ignore the brackets of δ_i, Δ_{ij} would vanish identically and C_{ij} would reduce to the identity, recovering the original AOS description. Likewise, if we consider that each parameter depends only on its corresponding partial Hamiltonian (reestablishing the radial-angular decoupling), Δ_{ij} vanish if $i \neq j$ and C_{ij} reduces to $(1 - \Delta_{ii})^{-1}$, which reproduces the result obtained in Ref. 40.

In Sec. 3 we discuss how to deal with these phase space dependent factors that result from the coupling of the radial and angular sectors in order to try and find solutions to the dynamics.

3. Local time redefinitions

As we saw in the previous section, the effect of considering polymerization parameters that are functions of both partial Hamiltonians is the appearance of two phase space dependent factors in the equations of motion. Interestingly, the obtained equations turn out to be a rescaling of the ones resulting from constant δ-parameters, and these rescalings, which are different in the radial and angular sectors, are given by the aforementioned factors C_{ij}. The fact that the nontrivial nature of the polymerization parameters affects the equations of motion precisely in this way implies that it can be absorbed through suitable time redefinitions. Indeed, taking these factors to the left-hand side of their respective dynamical equations, we realise that the definitions

$$dt_i = C_{ij}dt, \qquad j \neq i, \tag{20}$$

lead to equations that are identical to those considered by the authors of Refs. 1 and 2. From now on, we often refer to t_b and t_c as the "radial time" and the "angular time", respectively, for reasons that should be obvious. In terms of these newly defined time variables, the equations of motion adopt the same form as in the AOS model, which implies that the equations associated with the radial and angular sectors can be dealt with independently. Those corresponding to the c-sector can be solved in a straightforward manner, leading to[1,2]

$$\tan\frac{\delta_c c(t_c)}{2} = \frac{\gamma L_o \delta_c}{8m} e^{-2t_c}, \tag{21}$$

$$p_c(t_c) = 4m^2\left(e^{2t_c} + \frac{\gamma^2 L_o^2 \delta_c^2}{64m^2} e^{-2t_c}\right), \tag{22}$$

where the integration constants have been fixed so that the effective metric coincides with the standard form of the interior Schwarzschild metric.

Similarly, the equation associated with b can be solved in a straightforward way. We obtain[1,2]

$$\cos\delta_b b(t_b) = b_o \tanh\left(\frac{b_o t_b}{2} + \tanh^{-1}\frac{1}{b_o}\right) = \frac{1 + b_o \tanh\dfrac{b_o t_b}{2}}{1 + b_o^{-1}\tanh\dfrac{b_o t_b}{2}}, \tag{23}$$

where $b_o = \sqrt{1+\gamma^2\delta_b^2}$ and the constant of integration has been fixed so that the horizon lies at $t_b = 0$ (which is equivalent to requiring that b vanishes there). Since the canonical variables are bound by the Hamiltonian constraint, there is no need to solve the fourth differential equation, which is more involved than the others. Indeed, it suffices to solve for p_b in $H = 0$. Since it is convenient not to mix the radial and angular times, we can employ the fact that O_b is a constant of motion. Actually, using Eqs. (21) and (22), we can determine that

$$O_c = \frac{1}{\gamma}\frac{\sin\delta_c c}{\delta_c}\frac{p_c}{L_o} = \frac{2}{\gamma L_o \delta_c}\frac{p_c \tan(\delta_c c/2)}{1+\tan^2(\delta_c c/2)} = m. \tag{24}$$

Given that both partial Hamiltonians coincide on shell (as it follows immediately from $H = 0$), we then get $O_b = m$ as well. Therefore, solving for p_b in $O_b = m$, we obtain

$$p_b(t_b) = -2\gamma L_o m \frac{\sin \delta_b b(t_b)}{\delta_b} \frac{1}{\gamma^2 + \dfrac{\sin^2 \delta_b b(t_b)}{\delta_b^2}}, \tag{25}$$

where we can see that $p_b(0) = 0$, which is consistent with the requirement that the horizon correspond to $t_b = 0$.

So far, we have shown that suitable time redefinitions lead to dynamical equations that are identical to those obtained when δ_i are treated as constants, except for the fact that they are written in terms of two time variables. Hence, the solutions to the dynamics have the same form. Again, this is the case when expressed in two *a priori* different time variables, one per sector of the phase space. Since our objective is to determine whether it is possible to recover the results of the AOS model (at least in some sense), it seems natural to wonder whether we can use the off-shell freedom of our formalism to set both times equal. In the remainder of this section, we answer this question.

To begin with, we can find the relation between the radial and angular times by solving for dt in the definitions of both t_b and t_c and integrating the result of equating both expressions:

$$\int_{t_b}^{0} dt'_b[1 - \Delta_{bb}(t'_b) - \Delta_{bc}(t'_b)] = \int_{t_c}^{0} dt'_c[1 - \Delta_{cc}(t'_c) - \Delta_{cb}(t'_c)], \tag{26}$$

where we have canceled the denominators of C_{bc} and C_{cb}, which are identical [see Eq. (19) to realize that the denominator of these factors is symmetric under the exchange of i and j]. In this equation, $\Delta_{ij}(t_i)$ is obtained by plugging the solutions discussed above into the expression of Δ_{ij}. Moreover, the limits of integration have been chosen in accordance with the fact that the time variables are negative in the interior region of the black hole, as can be understood by examining the classical limit of the solutions to the effective dynamics.[1,2] Notice as well that we have used the existing freedom to make the origins of both time variables coincide.

Once we have fixed this freedom, asking whether t_b can be set equal to t_c is equivalent to asking whether C_{bc} can be chosen equal to C_{cb}. Therefore, the condition to be studied is

$$1 - \frac{\partial O_c}{\partial \delta_c}\frac{\partial f_c}{\partial \mu} = \alpha \left(1 - \frac{\partial O_b}{\partial \delta_b}\frac{\partial f_b}{\partial \mu}\right), \tag{27}$$

where we have called $\mu = (O_b + O_c)/2$, the on-shell value of which is simply given by m, and we have implicitly employed that the dependence of f_b and f_c on O_b and O_c can be equivalently described by their dependence on the difference and the sum of these quantities (divided by 2). In the previous expression, rather than demanding that t_b and t_c coincide, we have actually allowed for the more general case that these two times be proportional to each other, with proportionality factor α, hence

reaching a strict equality in the limit $\alpha \to 1$. Evaluating this expression on shell and assuming that f_i are at least $\mathcal{C}^1$, so that the derivative and the on-shell limit commute, we obtain

$$1 - \frac{\partial f_c}{\partial m}\frac{\partial O_c}{\partial \delta_c} = \alpha\left[1 - \frac{\partial f_b}{\partial m}\frac{\partial O_b}{\partial \delta_b}\right], \tag{28}$$

where $\partial f_i/\partial m$ stands for the on-shell particularization of the partial derivative of f_i with respect to μ, and we have omitted the indication of the rest of on-shell evaluations for the sake of simplicity. In the remainder of this article, whether an expression is evaluated on shell or not will be clear from the context. At this point, the only remaining step is to compute the derivatives of the partial Hamiltonians with respect to the polymerization parameters and evaluate them on shell. A simple calculation shows that

$$\frac{\partial O_b}{\partial \delta_b} = -\frac{1}{2\gamma}\left(1 - \frac{\gamma^2\delta_b^2}{\sin^2\delta_b b}\right)\frac{\delta_b b\cos\delta_b b - \sin\delta_b b}{\delta_b^2}\frac{p_b}{L_o}, \tag{29}$$

$$\frac{\partial O_c}{\partial \delta_c} = \frac{1}{\gamma}\frac{\delta_c c\cos\delta_c c - \sin\delta_c c}{\delta_c^2}\frac{p_c}{L_o}. \tag{30}$$

The fact that these derivatives are functions on phase space appears to be in contrast with the requirement that the on-shell values of t_b and t_c be proportional (or equal), which one would ideally want to hold on the whole phase space. Therefore, we need to find a way to cancel the undesired phase space dependence. Given the fact that $\sin\delta_i i$ can be rewritten in terms of m and p_i on shell [see Eqs. (4) and (5)], every function of the connection variables i can be reexpressed in terms of the mass of the black hole and the triad variables. Hence, the above derivatives can be regarded as nontrivial functions of p_i from the point of view of their dependence on phase space. As a result, we have

$$1 - \frac{\partial f_c}{\partial m}\frac{\partial O_c}{\partial \delta_c}(p_c) = \alpha\left[1 - \frac{\partial f_b}{\partial m}\frac{\partial O_b}{\partial \delta_b}(p_b)\right], \tag{31}$$

where the form of $\partial O_i/\partial\delta_i(p_i)$ is irrelevant for the present argument, except for the fact that they are not trivial functions. Since the phase space contributions of the left-hand side cannot cancel those of the right-hand side, we conclude that the only possibility to achieve the required cancelation is by choosing $\partial f_i/\partial m$ to vanish identically. That is, we can set both time variables equal if and only if the on-shell values of the parameters are independent of the mass. However, this does not make any sense physically speaking. Indeed, we would expect the parameters to vanish in the limit of infinitely large black hole masses, so that the description of general relativity can then be recovered, but not to be constantly zero for any value of the mass.

As a result, we conclude that the answer to the question of whether we can set the radial and angular times equal on shell is in the negative as long as we want physically acceptable polymerization parameters in the sense that they have an appropriate large mass limit. Hence, we realize that the appearance of *two distinct*

time variables is a defining feature of our formalism and cannot be avoided through an astute choice of parameters.

In the next section, we examine whether the proportionality condition (28) can hold in a suitable way in the limit of large black hole masses.

4. The large mass limit

Although we have shown that the radial and angular times cannot be set equal in general, this result does not imply that they may not coincide in some regime. From the point of view of the reconciliation with the results of Refs. 1 and 2, there is a particularly interesting situation, namely that of infinitely large masses. The reason behind this interest is twofold. On the one hand, it would seem especially desirable to achieve some sort of agreement in the large mass limit since the AOS model is adapted to describe very massive black holes to begin with. Indeed, as we commented in Sec. 2.1, their proposed parameters are only valid in that limit, given that they are extracted from the leading order of an asymptotic expansion of the area conditions that define them. On the other hand, astrophysical black holes, some of which may be susceptible of being probed by observations in the near future, are tremendously massive with respect to the Planck mass, which entails the importance of having a deep understanding of the similarities and differences between models in this regime. With these ideas in mind, let us perform an asymptotic expansion of the derivatives of the partial Hamiltonians with respect to the polymerization parameters, which are the functions of phase space that appear both in condition (28) and in the relation between the radial and angular time variables (26).

According to the discussed on-shell values of the polymerization parameters, it is immediate to realize that

$$\frac{\partial f_i}{\partial m} = -\frac{1}{3}\frac{\delta_i}{m} + o\left(\frac{\delta_i}{m}\right). \tag{32}$$

This means that the dominant term goes as $m^{-4/3}$ in the asymptotic limit. Therefore, it is obvious that the proportionality condition (28) will hold for large masses provided that, in that limit,[e]

$$\frac{\partial O_i}{\partial \delta_i} \sim m^{n_i} \quad \text{with} \quad n_i \leq \frac{4}{3}. \tag{33}$$

For simplicity, let us begin by computing the dominant term associated with the angular sector. Given the effective solutions discussed in Sec. 3, we see that the dominant term of the triad variable is given by

$$p_c = 4^{2t_c} m^2 + o(m^2). \tag{34}$$

[e]If any of the two exponents n_i turned out to be equal to 4/3, then an additional condition on the coefficients of the dominant terms must be met for Eq. (28) to hold asymptotically.

An expansion of the trigonometric functions of the connection variable can be obtained in a similar manner. Since our effective solution involves the tangent of $\delta_c c/2$, it is convenient to write

$$\cos\delta_c c = \frac{1-\tan^2(\delta_c c/2)}{1+\tan^2(\delta_c c/2)} = 1 - \frac{\gamma^2 L_o^2}{32 e^{4t_c}} \frac{\delta_c^2}{m^2} + o\left(\frac{\delta_c^2}{m^2}\right). \tag{35}$$

From this expression, we can derive the asymptotic expansions of the remaining functions of the connection variable c. Nevertheless, it is important to bear in mind that, according to the conventions of Refs. 1–3, $b > 0$, $c > 0$, $p_b \leq 0$, and $p_c \geq 0$. Together with the fact that the polymerization parameters are infinitely small in this regime, this implies that every trigonometric function contained in $\partial O_c/\partial\delta_c$ is nonnegative. As a result, a simple calculation reveals that

$$\frac{\partial O_c}{\partial \delta_c} = -\frac{\gamma^2 L_o^2}{48 e^{4t_c}} \frac{\delta_c}{m} + o\left(\frac{\delta_c}{m}\right). \tag{36}$$

This result indicates that the left-hand side of Eq. (28) tends to one as the mass grows infinitely.

An analogous result can be obtained in the radial sector following the same reasoning. As a consequence of the form of the effective solution (23), we get

$$\cos\delta_b b = 1 + C_1 \delta_b^2 + o(\delta_b^2), \tag{37}$$

where

$$C_1 = \gamma^2 \frac{\tanh(t_b/2)}{1+\tanh(t_b/2)}. \tag{38}$$

After a simple computation, we obtain that the relevant function of the connection variable b behaves as follows in the large mass regime:

$$\left(1 - \frac{\gamma^2\delta_b^2}{\sin^2\delta_b b}\right) \frac{\delta_b b\cos\delta_b b - \sin\delta_b b}{\delta_b^2} = -\frac{2}{3}\left(2C_1+\gamma^2\right)\frac{C_1}{\sqrt{-2C_1}}\delta_b + o\left(\delta_b\right). \tag{39}$$

Lastly, the dominant term of the triad variable p_b can be found using Eq. (25) in conjunction with the expansion of $\sin\delta_b b$, which is easily found from that of $\cos\delta_b b$, written above. The outcome is

$$p_b = -2\gamma L_o \frac{\sqrt{-2C_1}}{\gamma^2 - 2C_1} m + o\left(m\right). \tag{40}$$

Thus, the derivative of the radial partial Hamiltonian with respect to δ_b exhibits the asymptotic behavior

$$\frac{\partial O_b}{\partial \delta_b} = -\frac{2}{3} C_1 \frac{\gamma^2 + 2C_1}{\gamma^2 - 2C_1}\delta_b m + o\left(\delta_b m\right). \tag{41}$$

This behavior indicates that $\Delta_{bb} + \Delta_{bc}$ vanishes asymptotically.

Taking into account these results, the proportionality condition reduces at leading asymptotic order to

$$\alpha = 1. \tag{42}$$

Hence, we conclude that, among all the possibilities that the two time variables become proportional in the limit of large black hole masses, the only one that is realized is that where the proportionality factor is the identity.

Using the asymptotic expansions that we have just derived, we can obtain the relation between the radial and angular times for finite masses up to the first subdominant order. Indeed, the implicit relation (26) evaluated on shell reduces to

$$t_b - t_c = \frac{\partial f_b}{\partial m}\int_0^{t_b} dt'_b \frac{\partial O_b}{\partial \delta_b}(t'_b) - \frac{\partial f_c}{\partial m}\int_0^{t_c} dt'_c \frac{\partial O_c}{\partial \delta_c}(t'_c). \tag{43}$$

Substituting the derivatives of the δ-parameters and inserting the asymptotic expansions of the derivatives of the partial Hamiltonians, we find that

$$t_b - t_c = \frac{2}{9}\delta_b^2 \int_0^{t_b} dt'_b C_1(t'_b)\frac{\gamma^2 + 2C_1(t'_b)}{\gamma^2 - 2C_1(t'_b)} + o(\delta_b^2), \tag{44}$$

where we have omitted the angular contribution because it is subdominant with respect to the radial term. Only the latter sector contributes to the difference between the two time variables at the lowest nontrivial order. A trivial integration reveals that

$$t_c = t_b - \frac{1}{9}\gamma^2(-3t_b + 3\sinh t_b + \cosh t_b - 1)\delta_b^2 + o(\delta_b^2). \tag{45}$$

This equation provides the relation between the radial and angular time variables up to first subdominant asymptotic order. The leading term reproduces the result commented before: the two time variables coincide in the limit of infinitely large masses, as made evident by the fact that their difference vanishes when $m \to \infty$. Additionally, the form of the next-to-leading order term allows us to draw yet another parallel with the results of Ref. 40: the difference $t_c - t_b$ vanishes close to the horizon, which is defined by $t_b \to 0$, at least at the considered truncation order.

To summarize, our analysis indicates that the AOS equations of motion are recovered in the limit of large masses, leading to a *partial* reconciliation with the Hamiltonian formalism presented in this work. Indeed, even though our dynamical equations reduce in the large mass limit to those that result from treating the parameters as constants and they can be written in a single time variable (the coincident limit of t_b and t_c), this time is fundamentally different from the one that appears in the AOS model as it was originally formulated. Therefore, the time component of the effective spacetime metric is altered with respect to their case. This constitutes a modification of the effective geometry that must be studied carefully, in order to fully understand the breadth of its effects. This might open the door to an alleviation of some of the criticisms that the model has received,[41–44] which lie beyond the scope of this article. We leave these issues for future studies.

5. Conclusions

In the present work, we have studied whether it is possible to formulate an effective Hamiltonian description of the interior region of a Schwarzschild black hole such

that the polymerization parameters that mediate the inclusion of quantum geometry effects are appropriately treated as Dirac observables throughout the entire Hamiltonian calculation. The idea of identifying these parameters with constants of motion is one of the defining features of the effective solution proposed by Ashtekar, Olmedo, and Singh approximately three years ago.[1,2] In these works, this identification takes place at the very end, once the whole derivation of the Hamiltonian equations has been carried out dealing with these parameters as constant numbers. Only after this derivation, the parameters are fixed equal to the value of certain functions of the mass of the black hole. This procedure was criticized by the authors of a later work,[40] who insisted that the nontrivial nature of the parameters should be taken into account from the beginning, leading to extra terms in the equations that would fail to be accounted for otherwise. They remarked that, given the structure of the Hamiltonian constraint with a certain choice of the lapse function, there are two sectors of phase space (associated with the radial and angular directions, respectively) that are dynamically decoupled provided that the polymerization parameters do not introduce any cross-dependence on these sectors. Bearing this caveat in mind, their dynamics, which is generated by their associated partial Hamiltonians, can be studied independently. Then, they proposed to take each parameter to be a function of its respective partial Hamiltonian, both of which are constants of motion with a common on-shell value, essentially given by the mass of the black hole. This definition led to more complicated equations of motion where two brand new phase space dependent factors appeared. As shown in Ref. 40, these factors can be reabsorbed through time definitions, leading to simpler equations involving two time variables, one per sector of phase space.

We have extended the analysis carried out in Ref. 40 to encompass a more general choice of parameters with the objective of determining whether the two approaches described above can be reconciled in some sense. This reconciliation would give an extra degree of confidence on the very interesting qualitative results of Refs. 1–3. Our proposal is based on the fact that both partial Hamiltonians coincide on shell, which implies that their contributions should not be able to be told apart. Therefore, we have argued in Sec. 2.1 that the most general choice of parameters should capture the contributions from both the radial and angular partial Hamiltonians, breaking the decoupling that was present in previous works in the literature. As in the case of Ref. 40, our choice leads to equations of motion that are modified with respect to those considered in Refs. 1 and 2 by means of a pair of phase space dependent factors, one associated with the radial sector and the other with the angular sector (see Sec. 2.2). These factors are substantially more complicated than those found by the authors of Ref. 40, and can remain well defined in situations where the latter present pathologies. One can see that such previous results would be recovered in the limit where the decoupling is reestablished. On the other hand, as discussed in Sec. 3, appropriate local time redefinitions make the reabsorption of these factors

possible, leading to the dynamical equations considered in Refs. 1 and 2, now written in terms of two different times. These equations can be solved analytically and the result is identical to the one obtained in Refs. 1 and 2 except for the fact that two time variables, namely the radial and angular times, now play a role instead of one. We have investigated whether the off-shell freedom of our Hamiltonian formalism can be exploited in order to set both time variables equal. Upon inspection of the mathematical condition that would guarantee this equality, we have demonstrated that this coincidence cannot be achieved with a physically reasonable choice of polymerization parameters. Therefore, we have proceeded to analyze whether this equality is possible at least in some regime. In the model under consideration, the large mass limit proves to be especially interesting. Indeed, the choice of parameters made by the authors of Refs. 1 and 2 is adapted for the description of very massive black holes in the first place. With this motivation in mind, we have analyzed in Sec. 4 the equality condition for masses much larger than the Planck mass. After performing an asymptotic expansion of the relevant functions of phase space, we have shown that this condition does hold in this regime: both time variables turn out to be equal for infinitely massive black holes. More precisely, using the exact implicit relation between both variables, we have seen that the leading contribution to the difference of both times goes as the square of the radial polymerization parameter, thus vanishing as $m^{-2/3}$ in the limit where the mass of the black hole grows infinitely. For finite masses that are sufficiently large, we find that the differences with respect to the description of Refs. 1 and 2 can be estimated by the size of this term, yielding a quantitative measure of the discrepancies between both formalisms.

Our conclusions imply that the results of Refs. 1–3 can be partially reconciled with a treatment where the polymerization parameters are regarded as proper dynamical constants for black holes with large masses, which are the focus of the analysis of those references. The wording "partially" is key here. Notably, it should not be forgotten that the spacetime geometry is modified by means of time redefinitions. Even if this apparently slight modification does not alter some of the conclusions of Refs. 1–3, it may affect e.g. the behavior of the exterior effective metric at spatial infinity, which has been claimed to be nonstandard (see Refs. 41 and 3 for further details). This issue will constitute the subject of future studies.

Acknowledgments

The authors are grateful to B. Elizaga Navascués for fruitful discussions. This work has been supported by Project. No. MICINN FIS2017-86497-C2-2-P and Project. No. MICINN PID2020-118159GB-C41 from Spain. The projects that gave rise to these results received the support of a fellowship from "la Caixa" Foundation (ID 100010434). The fellowship code is LCF/BQ/DR19/11740028.

References

1. A. Ashtekar, J. Olmedo, and P. Singh, Quantum transfiguration of Kruskal black holes, *Phys. Rev. Lett.* **121**, 241301 (2018).
2. A. Ashtekar, J. Olmedo, and P. Singh, Quantum extension of the Kruskal spacetime, *Phys. Rev. D* **98**, 126003 (2018).
3. A. Ashtekar and J. Olmedo, Properties of a recent quantum extension of the Kruskal geometry, *Int. J. Mod. Phys. D* **29**, 2050076 (2020).
4. A. Ashtekar and M. Bojowald, Black hole evaporation: A paradigm, *Class. Quantum Grav.* **22**, 3349 (2005).
5. A. Ashtekar and M. Bojowald, Quantum geometry and the Schwarzschild singularity, *Class. Quantum Grav.* **23**, 391 (2005).
6. D. Cartin and G. Khanna, Wave functions for the Schwarzschild black hole interior, *Phys. Rev. D* **73** 104009 (2006).
7. L. Modesto, Loop quantum black hole, *Class. Quantum Grav.* **23**, 5587 (2006).
8. M. Bojowald, D. Cartin, and G. Khanna, Lattice refining loop quantum cosmology, anisotropic models and stability, *Phys. Rev. D* **76**, 064018 (2007).
9. C. G. Boehmer and K. Vandersloot, Loop quantum dynamics of Schwarzschild interior, *Phys. Rev. D* **76**, 1004030 (2007).
10. M. Campiglia, R. Gambini, and J. Pullin, Loop quantization of a spherically symmetric midsuperspaces: The interior problem, *AIP Conf. Proc.* **977**, 52 (2008).
11. S. Sabharwal and G. Khanna, Numerical solutions to lattice-refined models in loop quantum cosmology, *Class. Quantum Grav.* **25**, 085009 (2008).
12. D. W. Chiou, Phenomenological loop quantum geometry of the Schwarzschild black hole, *Phys. Rev. D* **78**, 064040 (2008).
13. D. W. Chiou, Phenomenological dynamics of loop quantum cosmology in Kantowski-Sachs spacetime, *Phys. Rev. D* **78**, 044019 (2008).
14. J. Brannlund, S. Kloster, and A. DeBenedictis, The evolution of black holes in the mini-superspace approximation of loop quantum gravity, *Phys. Rev. D* **79**, 084023 (2009).
15. R. Gambini, J. Omedo, and J. Pullin, Quantum black holes in loop quantum gravity, *Class. Quantum Grav.* **31**, 095009 (2014).
16. R. Gambini and J. Pullin, Hawking radiation from a spherical loop quantum gravity black hole, *Class. Quantum Grav.* **31**, 115003 (2014).
17. N. Dadhich, A. Joe, and P. Singh, Emergence of the product of constant curvature spaces in loop quantum cosmology, *Class. Quantum Grav.* **32**, 185006 (2015).
18. H. M. Haggard and C. Rovelli, Quantum-gravity effects outside the horizon spark black to white hole tunneling, *Phys. Rev. D* **92**, 104020 (2015).
19. A. Joe and P. Singh, Kantowski-Sachs spacetime in loop quantum cosmology: Bounds on expansion and shear scalars and viability of quantization prescriptions, *Class. Quantum Grav.* **32**, 015009 (2015).
20. A. Corichi and P. Singh, Loop quantum dynamics of Schwarzschild interior revisited, *Class. Quantum Grav.* **33**, 055006 (2016).
21. M. Campiglia, R. Gambini, J. Olmedo, and J. Pullin, Quantum self-gravitating collapsing matter in a quantum geometry, *Class. Quantum Grav.* **33**, 18LT01 (2016).
22. S. Saini and P. Singh, Geodesic completeness and the lack of strong singularities in effective loop quantum Kantowski-Sachs spacetime, *Class. Quantum Grav.* **33**, 245019 (2016).
23. J. Cortez, W. Cuervo, H. A. Morales-Técotl, and J. C. Ruelas, On effective loop quantum geometry of Schwarzschild interior, *Phys. Rev. D* **95**, 064041 (2017).

24. J. Olmedo, S. Saini, and P. Singh, From black holes to white holes: A quantum gravitational symmetric bounce, *Class. Quantum Grav.* **34**, 225011 (2017).
25. A. Yonika, G. Khanna, and P. Singh, Von-Neumann stability and singularity resolution in loop quantized Schwarzschild black hole, *Class. Quantum Grav.* **35**, 045007 (2018).
26. E. Bianchi, M. Christodoulou, F. D'Ambrosio, H.M. Haggard, and C. Rovelli, White holes as remnants: A surprising scenario for the end of a black hole, *Class. Quantum Grav.* **35**, 225003 (2018).
27. N. Bodendorfer, F. M. Mele, and J. Münch, Effective quantum extended spacetime of polymer Schwarzschild black hole, *Class. Quantum Grav.* **36**, 195015 (2019).
28. E. Alesci, S. Bahrami, and D. Pranzetti, Quantum gravity predictions for black hole interior geometry, *Phys. Lett. B* **797**, 134908 (2019).
29. M. Bouhmadi-López, S. Brahma, C.-Y. Chen, P. Chen, and D.-h. Yeom, A consistent model of non-singular Schwarzschild black hole in loop quantum gravity and its quasinormal modes, *J. Cosmol. Astropart. Phys.* **07**, 066 (2020).
30. M. Bojowald, Black-hole models in loop quantum gravity, *Universe* **6**, 125 (2020).
31. J. Ben Achour, S. Brahma, S. Mukohyama, and J.-P. Uzan, Towards consistent black-to-white hole bounces from matter collapse, *J. Cosmol. Astropart. Phys.* **09**, 020 (2020).
32. R. Gambini, J. Olmedo, and J. Pullin, Spherically symmetric loop quantum gravity: Analysis of improved dynamics, *Class. Quantum Grav.* **37**, 205012 (2020).
33. J. G. Kelly, R. Santacruz, and E. Wilson-Ewing, Effective loop quantum gravity framework for vacuum spherically symmetric spacetimes, *Phys. Rev. D* **102**, 106024 (2020).
34. W.-C. Gan, N. O. Santos, F.-W. Shu, and A. Wang, Towards understanding of loop quantum black holes, *Phys. Rev. D* **102**, 124030 (2020).
35. J. G. Kelly, R. Santacruz, and E. Wilson-Ewing, Black hole collapse and bounce in effective loop quantum gravity, *Class. Quantum Grav.* **38**, 04LT01 (2021).
36. N. Bodendorfer, F. M. Mele, and J. Münch, Mass and horizon Dirac observables in effective models of quantum black-to-white hole transition, *Class. Quantum Grav.* **38**, 095002 (2021).
37. N. Bodendorfer, F. M. Mele, and J. Münch, (b, v)-type variables for black to white hole transitions in effective loop quantum gravity, *Phys. Lett. B* **819**, 136390 (2021).
38. R. G. Daghigh, M. D. Green, and G. Kunstatter, Scalar perturbations and stability of a loop quantum corrected Kruskal black hole, *Phys. Rev. D* **103**, 084031 (2021).
39. J. Münch, Causal structure of a recent loop quantum gravity black hole collapse model, *Phys. Rev. D* **104**, 046019 (2021).
40. N. Bodendorfer, F. M. Mele, and J. Münch, A note on the Hamiltonian as a polymerisation parameter, *Class. Quantum Grav.* **37**, 187001 (2019).
41. M. Bouhmadi-López, S. Brahma, C.-Y. Chen, P. Chen, and D.-h. Yeom, Asymptotic non-flatness of an effective black hole model based on loop quantum gravity, *Phys. Dark Univ.* **30**, 100701 (2020).
42. M. Bojowald, Comment (2) on "Quantum transfiguration of Kruskal black holes", `arXiv:1906.04650`.
43. M. Bojowald, No-go result for covariance in models of loop quantum gravity, *Phys. Rev. D* **102**, 046006 (2020).
44. D. Arruga, J. Ben Achour, and K. Noui, Deformed general relativity and quantum black holes interior, *Universe* **6**, 039 (2020).
45. A. García-Quismondo and G. A. Mena Marugán, *Front. Astron. Space Sci.* **8**, 701723 (2021).

A comparison of different choices of clocks in a reduced phase space quantization in loop quantum cosmology with an inflationary potential using effective techniques

Kristina Giesel[1], Bao-Fei Li[2] and Parampreet Singh[2]

[1] *Institute for Quantum Gravity, Department of Physics, Theoretical Physics III, FAU Erlangen-Nürnberg, Staudtstr. 7, 91058 Erlangen, Germany*
kristina.giesel@gravity.fau.de

[2] *Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA*
baofeili1@lsu.edu, psingh@lsu.edu

We review recent results that apply a reduced phase space quantization of loop quantum cosmology (LQC) for a spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) universe filled with reference fields and an inflaton field in a Starobinsky inflationary potential. All three models that we consider are two-fluid models and they differ by their choice of global clock which are chosen to be either Gaussian dust, Brown-Kuchař dust or a massless Klein-Gordon scalar field. Although two-fluid models are more complicated than models involving the inflaton only, it turns out that some of the technical hurdles in conventional quantum cosmological models can be bypassed in these models. Using the effective dynamics resulting from the reduced phase space quantization we discuss some phenomenological implications of these models including the resolution of the big bang singularity via a quantum bounce and in addition address the question whether different choices of clocks can leave an imprint on the inflationary dynamics.

Keywords: Loop quantum cosmology, reduced phase space quantization, effective techniques.

1. Introduction

Within the last decade in the framework of loop quantum gravity new models have been introduced that apply the technique of reduced phase space quantization to construct the physical Hilbert space and thus the physical sector of the theory.[1–9] This requires to construct Dirac observables at the classical level and derive their corresponding algebra. For this purpose all models have in common that they couple some kind of additional matter to gravity. In the context of the relational formalism, introduced in Refs. 10, 11 and further developed and applied in Refs. 12–17, these additional degrees of freedom are used as reference matter to construct Dirac observables and their corresponding dynamics. The latter is encoded in a so called physical Hamiltonian that becomes a non-vanishing Hamiltonian operator in the physical Hilbert space involved either in a Schrödinger-like or Heisenberg-like equations. For the reason that these reference fields are dynamically coupled to general relativity an interesting question in this context is how a given choice of reference fields influences the physical properties of the model. In full loop quantum gravity

these dynamical equations have a complicated structure and an analysis for individual models as well as a comparison between different models is a non-trivial task. Therefore, the work in Ref. 18 focus on a simplified setting in the framework of loop quantum cosmology where techniques are already available to analyze such questions. The relational formalism has been applied in the context of loop quantum cosmology for instance in Refs. 18–33. The main difference to former models in loop quantum cosmology with reference matter, often also called clocks in cosmology, is that the work in Ref. 18 considers two-fluid models because the clock is coupled in addition to an inflaton. One of the main questions that one is interested in is how the inflationary scenario is affected by the presence of the additional clock degree of freedom and how the imprint of the clock compares for different models. Such questions will be investigated using effective techniques that in LQC have in former work mainly be applied to one-fluid models.

1.1. *Dust and scalar field clock models in loop quantum cosmology*

All three models in the analysis in Ref. 18 are based on models in general relativity coupled to an inflaton field and an additional coupling of 8 and 7 respectively additional dust and scalar fields respectively yielding to a system with second class constraints. After the reduction with respect to the second class constraints one ends up with first class systems with four additional reference fields, that can be used as reference matter for the Hamiltonian and spatial diffeomorphism constraint. The details about the Brown-Kuchař dust model can be found in Ref. 34 and its quantization using LQG techniques has been performed in Ref. 1. For the classical Gaussian dust model we refer the reader to Ref. 35 and its LQG implementation has been discussed in Ref. 36. The four scalar field model in Refs. 5, 6 can be understood as a modification of the model in Ref. 37, that, as shown in Ref. 5 cannot be quantized in the framework of LQG. For flat FLRW spacetimes where the spatial diffeomorphism constraint vanishes identically, the corresponding symmetry reduced models involve one temporal reference field, the clock, only. A reduced phase space quantization for all three symmetry reduced models has been derived in Ref. 18. There it is shown that in all three models the quantum dynamics is encoded in a Schrödinger-like equation in the physical Hilbert space. The explicit form of the physical Hamiltonian differs for the dust and scalar field models where the latter involves a square root. Although the model involve an inflaton with a generic potential due to the fact that clock is coupled additionally and one does not use the inflaton as the clock, as it has for instance be done in the APS-model in Ref. 21, all models possess physical Hamiltonians that are time-independent. This is of advantage for the construction of the physical inner product of the individual models. In this review we will focus on the effective dynamics of these models that were used in Ref. 18 to investigate the above mentioned questions. The quantum dynamics is formulated in the volume representation in Ref. 18 and thus the set of elementary variables in the dynamical equations are $\mathcal{O}_b, \mathcal{O}_V, \mathcal{O}_\varphi, \mathcal{O}_{\pi_\varphi}$, where $\mathcal{O}_f$

denotes the Dirac observable of quantity f. As shown in Ref. 18 the effective physical Hamiltonians of the dust and scalar field model take the form

$$\mathbf{H}_{\text{eff}}^{\text{FLRW, dust}} = -\frac{3\mathcal{O}_V}{8\pi G\lambda^2\gamma^2}\sin^2(\lambda\mathcal{O}_b) + \frac{\mathcal{O}_{\pi_\varphi}^2}{2\mathcal{O}_V} + \mathcal{O}_V U(\mathcal{O}_\varphi) \tag{1}$$

$$\mathbf{H}_{\text{eff}}^{\text{FLRW, scalar}} = \sqrt{-2\mathcal{O}_V\left(-\frac{3\mathcal{O}_V}{8\pi G\lambda^2\gamma^2}\sin^2(\lambda\mathcal{O}_b) + \frac{\mathcal{O}_{\pi_\varphi}^2}{2\mathcal{O}_V} + \mathcal{O}_V U(\mathcal{O}_\varphi)\right)}, \tag{2}$$

where $\lambda = \sqrt{\Delta} = \sqrt{4\sqrt{3}\pi\gamma\ell_p^2}$ denotes the polymerization parameter. Taking into account that the elementary Dirac observables satisfy the following standard Poisson brackets $\{\mathcal{O}_b, \mathcal{O}_V\} = 4\pi G\gamma$ and $\{\mathcal{O}_\varphi, \mathcal{O}_{\pi_\varphi}\} = 1$ with γ denoting the Immirzi parameter and where all remaining Poisson brackets vanish, the system of equations of motion for the Dirac observables in the individual models has been been derived in Ref. 18 and they provide the basis for the results discussed in the next section. Note that although in the case of flat FLRW spacetimes the physical Hamiltonian of the Brown-Kuchař model and the Gaussian dust model agree, these model are still not identical because within the Brown-Kuchař model the dust energy density can be chosen to be either positive or negative and the effect of these two different choices has also been analyzed in Ref. 18.

2. The effect of choosing different clocks on inflation

Similar to the one-fluid models that are obtained via Dirac quantization as for instance the APS-model in Ref. 21, the effective dynamics in the scalar field and dust models can be rewritten in terms of a modified Friedmann equations of the form

$$\mathcal{O}_H^2 = \frac{\dot{\mathcal{O}}_V^2}{9\mathcal{O}_V^2} = \frac{8\pi G}{3}\mathcal{O}_\rho\left(1 - \frac{\mathcal{O}_\rho}{\rho_{\max}}\right) \tag{3}$$

with $\rho_{\max} = \frac{3}{8\pi G\gamma^2\lambda^2}$. The maximal density $\rho_{\max}$ is the same as in Refs. 21–23 but here $\mathcal{O}_\rho$ does not only depend on the inflaton but also on the clock energy density. Furthermore, the temporal coordinate with respect to which the Hubble parameter is determined is given by either the dust and scalar field clock respectively.

In order to analyze how a given choice of clock might affect inflation in Ref. 18 a Starobinsky potential

$$U = \frac{3m^2}{16\pi G}\left(1 - e^{-\sqrt{\frac{16\pi}{3}}\mathcal{O}_\varphi}\right)^2 \quad \text{with} \quad m = 2.44\times 10^{-6} \tag{4}$$

was considered. The choice of initial conditions, set at the bounce, was guided by former results in LQC models[38] and chosen to be $\mathcal{O}_{b_i} = \pi/2\lambda$ and $\mathcal{O}_{V_i} = 10^3$ in Planck units. Fixing further values for $\mathcal{O}_{\varphi_i}, \mathcal{O}_{\pi_\varphi}$ determines $\mathbf{H}_{\text{eff}}^{\text{FLRW, dust}}$ from which one can obtain $\mathcal{O}_{\rho^{\text{clock}}}$. Therefore in Ref. 18 the initial values are parametrized by $(\mathcal{O}_\varphi, \mathcal{O}_{\rho^{\text{clock}}})$. In the case of the dust models a choice of initial conditions given by $\mathcal{O}_{\varphi_i} = -1.45$, $\mathcal{O}_{\rho^{\text{clock}}} = 10^{-8}$ yields to a quantum model with a pre-inflationary

phase, often being present in LQC models after the bounce, and an inflationary phase where the number of inflationary e-foldings for the latter is 63.1. For a similar choice of initial conditions the one-fluid models without an additional clock degree of freedom yield 63.9 inflationary e-foldings, see Ref. 38. This shows that for these choice of initial conditions the dust clock plays only a subdominant role for inflation. This is exactly what one expects from a good clock because such a clock should not have a dominant imprint on the dynamics of the model. A similar analysis for the scalar field model leads to 63.8 inflationary e-foldings being closer to the value obtained in the one-fluid models and showing that the effect of the scalar field clock is weaker compared to the dust clock. However, this is also expected from the fact that the energy density of the scalar field comes with a higher inverse power of the scale factor than the dust contribution and thus decaying faster in the evolution. In a further investigation to better understand the influence of the clock energy density on inflation in Ref. 18 for the same set of initial conditions for the inflaton a varying dust energy density ranging from 10^{-8} up to 10^{-4} was considered. The results show that the number of inflationary e-foldings decrease with increasing dust energy density. The explanation given in Ref. 18 is that this results from the fact that $\mathcal{O}_{\varphi_{\mathrm{on}}}$, the value of scalar field's Dirac observable at the onset of inflation, decreases due to a larger Hubble friction when the dust energy density is increased. For an initial dust energy of $\mathcal{O}_{\rho^{\mathrm{clock}}} = 1.38 \times 10^{-4}$ in Planck units one reaches an upper bound for this set of initial conditions where inflation no longer occurs. The same analysis for the scalar field clock shows that also in this model the number of inflationary e-foldings decreases with higher clock energy density but the effect is less strong here. One sees a significant effect only if $\mathcal{O}_{\rho^{\mathrm{clock}}} \geq 0.001$ in Planck units. As far as the number of pre-inflationary e-foldings is concerned the results from Ref. 18 demonstrate that the increase of the clock energy density has the opposite effect, namely that the number of pre-inflationary e-foldings increase.

3. Conclusions

As a first step towards an investigation how choices of different reference fields affect the physical properties of models the work in Ref. 18 considers different symmetry reduced reference matter models choosing either dust or a massless scalar field as the clock in addition to gravity and the inflaton leading to two-fluid LQC models. For the choice of a Starobinsky potential the fingerprint of the clock on the inflationary dynamics was analyzed in the framework of effective techniques. The analysis shows that initial conditions can be chosen such that the clock has no significant effect on the dynamics. Both models with dust and the scalar field clock respectively show a qualitatively similar behavior and can serve as good clocks. Some difference in the two models are found such as that the model with the scalar field clock has a larger number of inflationary e-foldings compared to the dust model. Furthermore, the scalar field model serves in larger parameter space as a good clock since it is less sensitive to the initial conditions of the clock energy density. Going beyond

the first steps investigated in Ref. 18 and reviewed here requires a more detailed understanding on the physical solutions in these models at the level of the quantum theory. For this purpose the already existing numerical techniques[23,39] need to be generalized to the two-fluid case. Next to the question of singularity resolution at the level of the physical Hilbert space, such kind of generalization will be important because it will also allow to test the validity of effective techniques for two-fluid models.

References

1. K. Giesel and T. Thiemann, Algebraic quantum gravity (AQG). IV. Reduced phase space quantisation of loop quantum gravity, *Class. Quant. Grav.* **27**, p. 175009 (2010).
2. M. Domagala, K. Giesel, W. Kaminski and J. Lewandowski, Gravity quantized: Loop Quantum Gravity with a Scalar Field, *Phys. Rev.* **D82**, p. 104038 (2010).
3. V. Husain and T. Pawlowski, Dust reference frame in quantum cosmology, *Class. Quant. Grav.* **28**, p. 225014 (2011).
4. V. Husain and T. Pawlowski, Time and a physical Hamiltonian for quantum gravity, *Phys. Rev. Lett.* **108**, p. 141301 (2012).
5. K. Giesel and A. Vetter, Reduced loop quantization with four Klein–Gordon scalar fields as reference matter, *Class. Quant. Grav.* **36**, p. 145002 (2019).
6. K. Giesel and A. Oelmann, Comparison Between Dirac and Reduced Quantization in LQG-Models with Klein-Gordon Scalar Fields, *Acta Phys. Polon. Supp.* **10**, 339 (2017).
7. M. Ali, S. M. Hassan and V. Husain, Universe as an oscillator, *Phys. Rev.* **D98**, p. 086002 (2018).
8. M. Han and H. Liu, Effective Dynamics from Coherent State Path Integral of Full Loop Quantum Gravity, *Phys. Rev.* **D101**, p. 046003 (2020).
9. M. Han and H. Liu, Improved ($\bar{\mu}$-Scheme) Effective Dynamics of Full Loop Quantum Gravity (2019).
10. C. Rovelli, What Is Observable in Classical and Quantum Gravity?, *Class. Quant. Grav.* **8**, 297 (1991).
11. C. Rovelli, Partial observables, *Phys. Rev.* **D65**, p. 124013 (2002).
12. A. S. Vytheeswaran, Gauge unfixing in second class constrained systems, *Annals Phys.* **236**, 297 (1994).
13. B. Dittrich, Partial and complete observables for Hamiltonian constrained systems, *Gen. Rel. Grav.* **39**, 1891 (2007).
14. B. Dittrich, Partial and complete observables for canonical general relativity, *Class. Quant. Grav.* **23**, 6155 (2006).
15. T. Thiemann, Reduced phase space quantization and Dirac observables, *Class. Quant. Grav.* **23**, 1163 (2006).
16. J. Pons, D. Salisbury and K. Sundermeyer, Revisiting observables in generally covariant theories in the light of gauge fixing methods, *Phys. Rev. D* **80**, p. 084015 (2009).
17. J. Pons, D. Salisbury and K. Sundermeyer, Observables in classical canonical gravity: Folklore demystified, *J. Phys. Conf. Ser.* **222**, p. 012018 (2010).
18. K. Giesel, B.-F. Li and P. Singh, Towards a reduced phase space quantization in loop quantum cosmology with an inflationary potential, *Phys. Rev. D* **102**, p. 126024 (2020).

19. D. Kaup and A. Vitello, Solvable quantum cosmological model and the importance of quantizing in a special canonical frame, *Phys. Rev. D* **9**, 1648 (1974).
20. W. Blyth and C. Isham, Quantization of a Friedmann Universe Filled with a Scalar Field, *Phys. Rev. D* **11**, 768 (1975).
21. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang, *Phys. Rev. Lett.* **96**, p. 141301 (2006).
22. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang: An Analytical and Numerical Investigation. I., *Phys. Rev.* **D73**, p. 124038 (2006).
23. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang: Improved dynamics, *Phys. Rev.* **D74**, p. 084003 (2006).
24. A. Ashtekar, A. Corichi and P. Singh, Robustness of key features of loop quantum cosmology, *Phys. Rev.* **D77**, p. 024046 (2010).
25. A. Ashtekar, T. Pawlowski, P. Singh and K. Vandersloot, Loop quantum cosmology of k=1 FRW models, *Phys. Rev. D* **75**, p. 024035 (2007).
26. K. Vandersloot, Loop quantum cosmology and the k = - 1 RW model, *Phys. Rev. D* **75**, p. 023523 (2007).
27. E. Bentivegna and T. Pawlowski, Anti-deSitter universe dynamics in LQC, *Phys. Rev. D* **77**, p. 124025 (2008).
28. W. Kaminski and T. Pawlowski, The LQC evolution operator of FRW universe with positive cosmological constant, *Phys. Rev. D* **81**, p. 024014 (2010).
29. T. Pawlowski and A. Ashtekar, Positive cosmological constant in loop quantum cosmology, *Phys. Rev. D* **85**, p. 064001 (2012).
30. T. Pawlowski, R. Pierini and E. Wilson-Ewing, Loop quantum cosmology of a radiation-dominated flat FLRW universe, *Phys. Rev. D* **90**, p. 123538 (2014).
31. J. Mielczarek and W. Piechocki, Observables for FRW model with cosmological constant in the framework of loop cosmology, *Phys. Rev. D* **82**, p. 043529 (2010).
32. F. Amemiya and T. Koike, Gauge-invariant construction of quantum cosmology, *Phys. Rev. D* **80**, p. 103507 (2009).
33. S. Gryb and K. P. Thébault, Bouncing Unitary Cosmology I: Mini-Superspace General Solution, *Class. Quant. Grav.* **36**, p. 035009 (2019).
34. J. D. Brown and K. V. Kuchar, Dust as a standard of space and time in canonical quantum gravity, *Phys. Rev.* **D51**, 5600 (1995).
35. K. V. Kuchar and C. G. Torre, Gaussian reference fluid and interpretation of quantum geometrodynamics, *Phys. Rev.* **D43**, 419 (1991).
36. K. Giesel and T. Thiemann, Scalar Material Reference Systems and Loop Quantum Gravity, *Class. Quant. Grav.* **32**, p. 135015 (2015).
37. K. V. Kuchar and C. G. Torre, The Harmonic gauge in canonical gravity, *Phys. Rev. D* **44**, 3116 (1991).
38. B.-F. Li, P. Singh and A. Wang, Genericness of pre-inflationary dynamics and probability of the desired slow-roll inflation in modified loop quantum cosmologies, *Phys. Rev.* **D100**, p. 063513 (2019).
39. P. Diener, B. Gupt and P. Singh, Numerical simulations of a loop quantum cosmos: Robustness of the quantum bounce and the validity of effective dynamics, *Class. Quant. Grav.* **31**, p. 105015 (2014).

Initial conditions in LQC/mLQCs

Bao-Fei Li*, Parampreet Singh* and Anzhong Wang†,‡

*Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

†Institute for Theoretical Physics & Cosmology, Zhejiang University of Technology, Hangzhou, 310032, China

‡GCAP-CASPER, Department of Physics, Baylor University, Waco, TX 76798, USA

We compare the behavior of the effective masses in the Mukhanov-Sasaki equations of the linear scalar perturbations in the dressed metric and hybrid approaches in the standard loop quantum cosmology (LQC) and the modified LQC models (mLQCs). The effective mass of the Mukhanov-Sasaki equations depend on both the specific model and the approach we study. Analyzing the behavior of the effective masses plays an important role in the choice of the initial states for the scalar perturbations. Based on the properties of the effective masses in the contracting phase, we provide the initial states of the linear perturbations in LQC and mLQCs for the dressed metric and the hybrid approaches.

Keywords: Mukhanov-Sasaki equations; loop quantum cosmology.

1. Introduction

In loop quantum cosmology(LQC), the big bang singularity in the standard cosmology is resolved and replaced with a quantum bounce, and correspondingly the inflationary paradigm, which resolves several big bang puzzles and provides initial seeds for the formation of the large scale structure in the universe, can be extended to the Planck regime by using the effective dynamics.[1,2] Several approaches have been developed in LQC to compute the primordial power spectral produced by the quantum fluctuations which propagate on quantum geometries. However, such effects become negligible soon after the quantum bounce. In particular, at the beginning of inflation, where the energy density has decreased about 12 orders in comparing that at the bounce, $\rho_i/\rho_c \simeq 10^{-12}$, the quantum spacetime can be well approximated by a classical one, whereby the standard relativistic cosmology follows. Two of the most studied approaches are the dressed metric approach[3–5] and the hybrid approach.[6–9] Although these two approaches are formulated in different ways, the resulting equations of motion for the scalar modes in both approaches can be cast into the form

$$\nu_k^{''} + \left(k^2 + m_{\rm eff}\right)\nu_k = 0, \tag{1}$$

here $m_{\rm eff}$ denotes the effective mass which accounts for the differences between two approaches. Similar to the situation in the classical cosmology, the initial condition problem persists in LQC in the form that up to now there is no consensus on whether

the initial states should be imposed in the Planck regime or in the contracting phase where the background dynamics becomes classical again. In this paper, we take the viewpoint to impose the initial states in the deep contracting phase. As a result, we focus on the behavior of the effective mass in the deep contracting phase. Besides, in addition to the standard LQC, we also investigate the behavior of the effective mass in two modified LQC models (mLQCs) in order to study the potential effects of the quantization ambiguities on the observables.

The said mLQCs originate from a separate treatment of the Lorentz term in the classical Hamiltonian constraint in the spatially flat FLRW universe.[10] Two of the mLQCs which attract attention recently are the so-called mLQC-I and mLQC-II. Both the background dynamics and the primordial power spectra of these two mLQCs have been extensively studied in the literature.[11–17] From these studies, we learn that the background dynamics in mLQC-II is qualitatively similar to that in LQC while the contracting phase of mLQC-I has distinctive feature which is characterized by an emergent cosmological constant and a rescaled Newton's constant. When applying the dressed metric and the hybrid approaches to the mLQCs, Planck scale magnitude of the power spectrum is observed in the infrared(IR) regime of mLQC-I with the dressed metric approach while in the hybrid approach, the magnitudes of the power spectra in mLQCs and LQC become comparable. As a result, the modified LQC models also provide good examples to compare the two approaches. In the following, we compare the effective masses in LQC, mLQC-I and mLQC-II, respectively, in the dressed metric and the hybrid approaches, and from the properties of the effective masses in the contracting phase, we seek the appropriate initial states to be imposed during this phase. For more details, we refer readers to Ref. 17.

2. The initial states in LQC/mLQCs in the dressed metric approach

In the dressed metric approach, the quantum fluctuations are described as propagating on a quantum background spacetime whose geometry can be well approximated by a manifold with a dressed metric for the sharply peaked states. Using the effective dynamics, it can be shown that the effective mass in (1) takes the same form as in the classical perturbation theory while the difference originates from two sources.[15] Firstly, the background quantities in $m_{\rm eff}$ should come from the effective dynamics governed by the effective Hamiltonian constraint in LQC. Secondly, the regularization of the conjugate momentum of the scale factor in $m_{\rm eff}$ should be also consistent with the effective Hamiltonian constraint for the background dynamics. With the above two points in mind, we impose the kinetic-energy-dominated initial conditions for the background evolution at the bounce and study the behavior of the effective mass for the chaotic and the Starobinsky potentials. In practice, one can define the comoving Hubble radius as,[17]

$$\lambda_H^2 = -\frac{1}{m_{\rm eff}}. \tag{2}$$

When the wavelength of the relevant modes is larger than λ_H^2, the mode function is oscillating. On the other hand, when the wavelength of the relevant modes are less than λ_H^2, the mode function is exponentially decaying or growing.

From our numerical results, we find the behavior of the comoving Hubble radius in mLQC-II is very similar to that in LQC. Its evolution is almost symmetric with respect to the bounce point. During the deflationary stage in the contracting phase, the comoving Hubble radius keeps increasing until it reaches positive infinity at some time $t = t_i < 0$ where a'' vanishes. After $t = t_i$, it is the intermediate regime where the comoving Hubble radius becomes negative and correspondingly all the modes are oscillating in this regime. Near the bounce, the comoving Hubble radius becomes positive and reaches its minimal right at the bounce, which corresponds to the characteristic wavelength in LQC/mLQC-II. As a result, we find the Bunch-Davies vacuum can be chosen in the intermediate regime of the contracting phase where $\lambda_H^2 < 0$ in LQC and mLQC-II. On the other hand, we find a different behavior of λ_H^2 in the contracting phase of mLQC-I. In particular, λ_H^2 is always positive in the contracting phase and it can be well approximated by $\eta^2/2$ with η denoting the conformal time in the quasi de Sitter regime of the contracting phase. As a result, we can choose the de Sitter state as the initial state in the contracting phase. The above analysis on the behavior of the comoving Hubble radius is robust with respect to the choice of the inflaton potentials and the regularization of the conjugate momentum of the scale factor in the effective mass as long as the bounce is dominated by the kinetic energy of the inflaton field.

3. The initial states in LQC/mLQCs in the hybrid approach

The application of the hybrid approach to the mLQCs is studied in detail in[16] where the equation of motion for the scalar perturbations are given explicitly. Using the effective mass given in the mode equation, we study its properties in the contracting phase when the initial conditions of the background dynamics are chosen the same as in the dressed metric approach. From our numerical results, we find the evolution of the comoving Hubble radius is generally not symmetric with respect to the bounce in any model and it also depends on the specific inflationary potential. Surprisingly, with the chaotic potential, the comoving Hubble radius behaves similarly in all three models. It increases monotonically from some negative value in the deep contracting phase and tends to positive infinity at time $t = \tilde{t}_i < 0$ where $a'' = 0$. Afterwards, it remains negative until the onset of inflation in the expanding phase. This implies the effective mass is always negative near the bounce. As a result, with the chaotic potential, one can choose the BD vacuum in the deep contracting regime or near the bounce in all three models. However, since the comoving Hubble radius becomes infinity at $t = \tilde{t}_i$, choosing BD vacuum at different times, either in the deep contracting regime or near the bounce is expected to give different power spectra in the IR regime due to particle creation effect.

On the other hand, with the Starobinsky potential, we find the comoving Hubble radius has similar behavior in LQC and mLQC-II. It remains negative throughout the contracting phase until the onset of inflation in the expanding phase. As a result, the choice of the BD vacuum can be made either in the deep contracting phase or near the bounce and the resulting power spectra are expected not to be significantly different from one to another. In contrast, in mLQC-I, although the comoving Hubble radius is still negative near the bounce, there exists a finite time interval in the contracting phase where it become positive. In the deep contracting phase, it becomes negative again. So similar to the case of the chaotic potential, imposing the BD vacuum at different times, either in the deep contracting phase or near the bounce is expected to yields different power spectra in the IR regime.

If we compare different approaches for the same model, the most striking difference lies in mLQC-I where the choice of the de Sitter state in the contracting phase in the dressed metric approach yields a Planck scale magnitude of the power spectra in the IR regime, while the choice of the BD vacuum in the deep contracting phase in the hybrid approach highly suppresses the magnitude of the resulting power spectrum in the IR regime. This serves as a concrete example where the different approaches can result in different qualitative behaviors in the power spectra.

4. Summary

We study the behavior of the effective masses in the equation of motion for the scalar mode in LQC and mLQCs in the dressed metric and the hybrid approaches. With the same kinetic-energy-dominated initial conditions for the background dynamics, we find the properties of the comoving Hubble radius in the contracting phase. In the dressed metric approach, the comoving Hubble radius has similar behavior for the chaotic and the Starobinsky potentials. One can choose the BD vacuum in the deep contracting phase of LQC and mLQC-II and the de Sitter state in the contracting phase of mLQC-I. On the other hand, the properties of the comoving Hubble radius in the hybrid approach depends on the specific inflationary potentials. One common feature of the comoving Hubble radius for all the models is that it is negative near the bounce and we find it appropriate to choose the BD vacuum in the hybrid approach for all three models in the deep contracting phase.

We are grateful to Robert Brandenberger and David Wands for valuable comments and helpful discussions. We thank Javier Olmedo and Tao Zhu for various discussions. This work is supported by NSF grant PHY-1454832.

References

1. A. Ashtekar and P. Singh, *Loop Quantum Cosmology: A Status Report*, Class. Quant. Grav. **28**, 213001 (2011).
2. I. Agullo and P. Singh, *Loop Quantum Cosmology*, in Loop Quantum Gravity: The First 30 Years, edited by A. Ashtekar and J. Pullin (World Scientific, Singapore, 2017).

3. I. Agullo, A. Ashtekar and W. Nelson, *Quantum gravity extension of the inflationary scenario*, Phys. Rev. Lett. **109**, 251301 (2012).
4. I. Agullo, A. Ashtekar and W. Nelson, *Extension of the quantum theory of cosmological perturbations to the Planck era*, Phys. Rev. D**87**, 043507 (2013).
5. I. Agullo, A. Ashtekar and W. Nelson, *The pre-inflationary dynamics of loop quantum cosmology: Confronting quantum gravity with observations*, Class. Quantum Grav. **30**, 085014 (2013).
6. L. Castelló Gomar, M. Fernández-Méndez, G. A. Mena Marugan and J. Olmedo, *Cosmological perturbations in hybrid loop quantum cosmology: Mukhanov-Sasaki variables*, Phys. Rev. D**90**, 064015 (2014).
7. M. Fernández-Méndez, G. A. Mena Marugán and J. Olmedo, *Hybrid quantization of an inflationary universe*, Phys. Rev. D**86**, 024003 (2012).
8. M. Fernández-Méndez, G. A. Mena Marugán and J. Olmedo, *Hybrid quantization of an inflationary model: The flat case*, Phys. Rev. D**88**, 044013 (2013).
9. L. Castelló Gomar, M. Martín-Benito, G. A. Mena Marugán, *Gauge-invariant perturbations in hybrid quantum cosmology*, JCAP 1506 (2015) 045.
10. J. Yang, Y. Ding and Y. Ma, *Alternative quantization of the Hamiltonian in loop quantum cosmology II: Including the Lorentz term*, Phys. Lett. B**682** (2009) 1.
11. B.-F. Li, P. Singh, A. Wang, *Towards cosmological dynamics from loop quantum gravity*, Phys. Rev. D**97**, 084029 (2018).
12. B.-F. Li, P. Singh, A. Wang, *Qualitative dynamics and inflationary attractors in loop cosmology*, Phys. Rev. D**98**, 066016 (2018).
13. I. Agullo, *Primordial power spectrum from the Dapor-Liegener model of loop quantum cosmology*, Gen. Relativ. Grav. **50** (2018) 91.
14. B.-F. Li, P. Singh, A. Wang, *Genericness of pre-inflationary dynamics and probability of the desired slow-roll inflation in modified loop quantum cosmologies*, Phys. Rev. D**100**, 063513 (2019).
15. B.-F. Li, P. Singh, A. Wang, *Primordial power spectrum from the dressed metric approach in loop cosmologies*, Phys. Rev. D**100**, 086004 (2020).
16. B.-F. Li, J. Olmedo, P. Singh, and A. Wang, *Primordial scalar power spectrum from the hybrid approach in loop cosmologies*, Phys. Rev. D**102**, 126025 (2020).
17. B.-F. Li, P. Singh, and A. Wang, *Phenomenological implications of modified loop cosmologies: An overview*, Front. Astron. Space Sci. 8 (2021) 701417.

Holonomy corrections in effective midisuperspace models

A. Alonso-Bardaji* and D. Brizuela†

Department of Physics, University of the Basque Country (UPV/EHU), Leioa, 48940, Spain

**E-mail: asier.alonso@ehu.eus*

†E-mail: david.brizuela@ehu.eus

We develop a systematic method to obtain spherically symmetric midisuperspace models with modifications inherited from loop quantum gravity. We obtain a family of effective constraints satisfying Dirac's deformation algebra and show that holonomy corrections can be consistently implemented in the presence of matter with local degrees of freedom.

Keywords: Polymerization; effective theories: loops quantum gravity.

1. Introduction

Holonomy-modified cosmological effective models capture the behavior of the dynamics predicted by loop quantum cosmology.[1,2] Such effective descriptions usually modify the Hamiltonian by hand to include the expected effects from loop quantum gravity. Holonomy corrections are directly related to the spacetime discreteness and replace the initial singularity by a quantum bounce.

We will work with spherically symmetric models of general relativity. The phase-space variables are the components of the symmetry-reduced triad and extrinsic curvature components,[3] with Poisson brackets given by

$$\{E^r(r_a), K_r(r_b)\} = \{E^\varphi(r_a), K_\varphi(r_b)\} = -\delta(r_a - r_b). \tag{1}$$

In addition, we will include the dynamics through a spherically symmetric scalar field:

$$\{\phi(r_a), P_\phi(r_b)\} = \delta(r_a - r_b). \tag{2}$$

In these effective midisuperspaces, the quantized notion of spacetime collides with the continuous diffeomorphism symmetry of general relativity. Hence, we will search for effective theories that explicitly satisfy covariance through Dirac's algebra,

$$\big\{D[N^r_1], D[N^r_2]\big\} = D\big[N^r_1 N^{r\prime}_2 - N^{r\prime}_1 N^r_2\big], \tag{3a}$$

$$\big\{D[N^r], H[N]\big\} = H\big[N^r N'\big], \tag{3b}$$

$$\big\{H[N_1], H\big[N_2\big]\big\} = D\big[\beta E^r (E^\varphi)^{-2}\,(N_1 N_2' - N_1' N_2)\big], \tag{3c}$$

where $H[N] := \int dr\, N\mathcal{H}$ and $D[N^r] := \int dr\, N^r\mathcal{D}$ are, respectively, the infinitesimal generators of normal and tangential hypersurface deformations and β stands for possible quantum modifications.

In terms of the above variables, the classical constraints read,

$$\mathcal{D} = -E^{r\prime}K_r + E^{\varphi}K'_{\varphi} + \phi' P_{\phi}\,, \tag{4}$$

$$\mathcal{H} = -\frac{E^{\varphi}}{2\sqrt{E^r}}\left(1+K_{\varphi}^2\right) - 2\sqrt{E^r}K_rK_{\varphi} + \frac{(E^{r\prime})^2}{8\sqrt{E^r}E^{\varphi}} - \frac{\sqrt{E^r}}{2E^{\varphi 2}}E^{r\prime}E^{\varphi\prime} + \frac{\sqrt{E^r}}{2E^{\varphi}}E^{r\prime\prime} + \frac{P_{\phi}^2}{2\sqrt{E^r}E^{\varphi}} + \frac{(E^r)^{3/2}}{2E^{\varphi}}(\phi')^2 + \sqrt{E^r}E^{\varphi}V(\phi)\,. \tag{5}$$

2. Deformed algebra

Previous studies acknowledge that holonomy corrections cannot be implemented in the presence of minimally coupled matter fields.[4] In fact, we extended that affirmation to the case when no coupling between the radial derivatives of the triads and curvature components are considered.[5]

For that reason, here we suggest a more ambitious approach in which we compute the algebraic relations (3) for the classical diffeomorphism constraint (4) along with a generic Hamiltonian constraint,

$$\begin{aligned}\mathcal{H} = {} & a_0 + (E^{r\prime})^2 a_{rr} + (E^{\varphi\prime})^2 a_{\varphi\varphi} + (\phi')^2 a_{\phi\phi} \\ & + E^{r\prime}E^{\varphi\prime}a_{r\varphi} + E^{r\prime}\phi' a_{r\phi} + E^{\varphi\prime}\phi' a_{\varphi\phi} + E^{r\prime\prime}a_2,\end{aligned} \tag{6}$$

that contains all quadratic combinations of the radial derivatives of E^r, E^{φ}, ϕ. The coefficients are given in terms of certain free functions $a_{ij} = a_{ij}(E^r, E^{\varphi}, \phi, K_r, K_{\varphi}, P_{\phi})$, with $i,j = r,\varphi,\phi$, that depend on all the variables and momenta of the model. Concerning second-order radial derivatives, only $E^{r\prime\prime}$ has been considered, as it also appears in the classical Hamiltonian (5). Finally, $a_0 = a_0(E^r, E^{\varphi}, \phi, K_r, K_{\varphi}, P_{\phi})$ encodes all the different terms that do not contain any radial derivative.

2.1. *The $\{D, H\}$ bracket*

First, we compute the Poisson brackets between the modified Hamiltonian constraint (6) and the classical diffeomorphism constraint (4). A number of anomalies appear, which can be easily solved. This procedure leads to the following most general form of the Hamiltonian constraint

$$\begin{aligned}\mathcal{H} = -\sqrt{E^r}\frac{g}{2}\Big(& E^{\varphi}b_0 + E^{\varphi-1}(E^{r\prime})^2 b_{rr} - E^{\varphi-1}(\phi')^2 b_{\phi\phi} + E^{\varphi-1}E^{r\prime}\phi' b_{r\phi} \\ & + E^{\varphi-2}E^{r\prime}E^{\varphi\prime} - E^{\varphi-1}E^{r\prime\prime}\Big),\end{aligned} \tag{7}$$

which provides a weakly vanishing bracket with the classical diffeomorphism constraint. In this expression, $b_0 = b_0\left(E^r, \phi, K_r/E^{\varphi}, K_{\varphi}, P_{\phi}/E^{\varphi}\right)$ and $b_{ij} = b_{ij}\left(E^r, \phi, K_r/E^{\varphi}, K_{\varphi}, P_{\phi}/E^{\varphi}\right)$, with $i,j = r,\phi$, are arbitrary functions of all possible scalar combinations of the model. In addition, a generic global factor g has been

defined. The anomaly-free bracket between this constraint and the diffeomorphism constraint reads,

$$\big\{D[N^r], H[N]\big\} = H\big[N^r N'\big] + H\big[\mathcal{G}\big], \tag{8}$$

with the smearing $\mathcal{G}$ being a function of N, N^r, g and their derivatives. Nonetheless, for the Hamiltonian to be a density of weight $+1$, it is necessary that the global factor g is a scalar function. This means that all its arguments must be scalar quantities and thus the global factor takes the form $g = g\left(E^r, \phi, K_r/E^\varphi, K_\varphi, P_\phi/E^\varphi\right)$. Under this last condition, the function $\mathcal{G}$ vanishes. Hence, for any scalar global factor g, the bracket between the modified Hamiltonian constraint (7) and the diffeomorphism constraint (4) is given by its classical form (3b).

2.2. *The $\{H, H\}$ bracket*

The requirement of anomaly freedom in the bracket between the modified Hamiltonian constraint with itself, $\{H, H\}$, immediately fixes the form of the Hamiltonian with respect to all the densities (K_r, E^φ, P_ϕ), except as arguments of the global factor g. Hence, let us define the new modified constraint:

$$\begin{aligned}\mathcal{H} = -\sqrt{E^r}\frac{g}{2}\Bigg(& E^\varphi\left(f_0 + \frac{K_r}{E^\varphi}f_1 + \frac{P_\phi}{E^\varphi}h_0 + \frac{P_\phi^2}{{E^\varphi}^2}f_3\right) \\ & + \frac{(E^{r\prime})^2}{E^\varphi}\left(f_2 + \frac{K_r}{E^\varphi}h + \frac{P_\phi}{E^\varphi}h_1 + \frac{P_\phi^2}{{E^\varphi}^2}h_2\right) \\ & + \frac{(\phi')^2}{E^\varphi}f_4 + \frac{E^{r\prime}\phi'}{E^\varphi}\left(h_3 + \frac{P_\phi}{E^\varphi}h_4\right) + \frac{E^{r\prime}E^{\varphi\prime}}{{E^\varphi}^2} - \frac{E^{r\prime\prime}}{E^\varphi}\Bigg),\end{aligned} \tag{9}$$

where $h = h(E^r, \phi, K_\varphi)$, $f_i = f_i(E^r, \phi, K_\varphi)$ and $h_i = h_i(E^r, \phi, K_\varphi)$, for $i = 0, 1, 2, 3, 4$, are free functions of the scalar variables (E^r, ϕ, K_φ). Note that the form of the classical Hamiltonian (5) is recovered for $f_0 \to (1 + K_\varphi^2)/E^r$, $f_1 \to 4K_\varphi$, $f_2 \to -1/(4E^r)$, $f_3 \to -1/E^r$, $f_4 \to -E^r$, $g \to 1$ and for a vanishing value of all the functions h_i and h. With this definition at hand, it is now possible to reduce the requirement of vanishing anomalies to the following system of differential equations:

$$\frac{\partial f_0}{\partial K_\varphi} = 2f_0 h - 2f_1 f_2 - h_0 h_3 + \frac{\partial f_1}{\partial E^r}, \tag{10a}$$

$$\frac{\partial f_1}{\partial K_\varphi} = f_1 h_4 + 4f_3 f_4, \tag{10b}$$

$$\frac{\partial f_2}{\partial K_\varphi} = -h_1 h_3 + \frac{\partial h}{\partial E^r}, \tag{10c}$$

$$\frac{\partial f_4}{\partial K_\varphi} = 2f_4(h - h_4), \tag{10d}$$

$$\frac{\partial f_3}{\partial K_\varphi} = -2f_1 h_2 + 2f_3(h - h_4), \tag{10e}$$

$$\frac{\partial f_1}{\partial \phi} = f_1 h_3 + 2h_0 f_4 \tag{10f}$$

$$\frac{\partial h_0}{\partial K_\varphi} = h_0(2h - h_4) - 2f_1 h_1 - 2f_3 h_3, \tag{10g}$$

$$\frac{\partial h_1}{\partial K_\varphi} = -h_1 h_4 - 2h_2 h_3, \tag{10h}$$

$$\frac{\partial h_2}{\partial K_\varphi} = -2h_2 h_4, \tag{10i}$$

$$\frac{\partial h_3}{\partial K_\varphi} = h_3(h - h_4) - 2h_1 f_4 + \frac{\partial h}{\partial \phi}, \tag{10j}$$

$$\frac{\partial h_4}{\partial K_\varphi} = h_4(h - h_4) - 4h_2 f_4 + \frac{\partial h}{\partial K_\varphi}. \tag{10k}$$

This is a system of eleven first-order partial differential equations for the eleven functions f_i, h_i, and h. The non-linearity of the system due to quadratic combinations of the functions makes the analytical resolution to be very complicated. But it is clear that the possible dependence of the Hamiltonian on the curvature component K_φ is severely restricted, since all the equations except one contain derivatives of the different functions with respect to K_φ.

In summary, whenever the above differential relations are satisfied, the constraint (9) will form a first-class algebra with the classical diffeomorphism constraint, and it will weakly commute with itself: $\{H[N_1], H[N_2]\} \approx 0$. Nonetheless, for H to be the generator of infinitesimal normal hypersurface deformations, one must impose an additional condition on the global factor, that is, $g = \exp(-2\int h\, dK_\varphi)$. In this way, the Poisson bracket of (9) with itself will be of the form (3c) with a deformation function,

$$\beta = \frac{g^2}{4}\left(\frac{\partial f_1}{\partial K_\varphi} + \left(\frac{E^{r\prime}}{E^\varphi}\right)^2 \frac{\partial h}{\partial K_\varphi}\right). \tag{11}$$

3. The vacuum reduction

Before trying a particular solution for the complete anomaly system (10), we proceed to solve the vacuum model. In the absence of matter, f_3, f_4, h_0, h_1, h_2, h_3 and h_4 vanish and we only need to worry about the reduced versions of (10a) and (10c) (Eq. (10b) is an additional restriction that appears only when matter is incorporated).

The general solution for the remaining functions f_0, f_1, f_2 and h can be written in terms of the global factor $g(E^r, K_\varphi)$ and two integration functions $f = f(E^r, K_\varphi)$ and $v = v(E^r)$. More precisely, the form of the modified vacuum Hamiltonian constraint reads,

$$\begin{aligned}\mathcal{H} = -\sqrt{E^r}\Bigg[E^\varphi\left(\frac{1+f^2}{v}\frac{\partial v}{\partial E^r} + \frac{\partial f^2}{\partial E^r}\right) + K_r\frac{\partial f^2}{\partial K_\varphi} - g\frac{E^{r\prime\prime}}{2E^\varphi} \\ + g\frac{E^{r\prime}E^{\varphi\prime}}{2E^{\varphi 2}} - \left(\frac{E^{r\prime}}{2E^\varphi}\right)^2\left(\frac{E^\varphi}{v}\frac{\partial(gv)}{\partial E^r} + K_r\frac{\partial g}{\partial K_\varphi}\right)\Bigg].\end{aligned} \tag{12}$$

In addition, we have been able to obtain a weak Dirac observable of the system,

$$m = \frac{v}{2}\left(1 + f^2 - g\left(\frac{E^{r\prime}}{2E^\varphi}\right)^2\right), \tag{13}$$

which commutes on-shell with the total Hamiltonian $\dot{m} = \{m, H[N] + D[N^r]\} \approx 0$, and represents the mass. In particular, the classical limit of the system is obtained for $g \to 1$, $f \to K_\varphi$ and $v \to \sqrt{E^r}$. In this limit, the above observable is just the usual expression of the Hawking (Schwarzschild) mass.

It is interesting to note that the expression of the effective mass (13) can be used to rewrite the deformation function as follows:

$$\beta = \frac{g^2}{2}\left(\frac{\partial}{\partial K_\varphi}\left[\frac{1}{g}\frac{\partial f^2}{\partial K_\varphi}\right] - \frac{1}{g}\frac{\partial^2 \ln g}{\partial K_\varphi^2}\left(1 + f^2 - \frac{2m}{v}\right)\right), \tag{14}$$

which is free from radial derivatives.

4. An effective Hamiltonian with holonomy corrections

In effective models for cosmological scenarios, holonomy corrections are usually implemented through a polymerization of the connection components. In our current scenario, only K_φ might be subjected to modifications, since the dependence of the Hamiltonian on the other two momenta, K_r and P_ϕ, is completely fixed in (9).

Furthermore, we will search for the Hamiltonian that remains the most similar to the classical constraint. Hence, we fix the matter coefficients $f_3 = -1/E^r$ and $f_4 = -E^r$ (although further couplings with K_φ might be present through the global factor g). In addition, we consider that the scalar field ϕ only appears in a classical potential term $V(\phi)$.

Now, we can obtain a solution to the system (10) in terms of three integration functions $f = f(E^r, K_\varphi)$, $v = v(E^r)$ and $w = w(E^r)$:

$$f_1 = \frac{2}{g}\frac{\partial f^2}{\partial K_\varphi}, \tag{15a}$$

$$f_2 = -\frac{1}{2}\frac{\partial \ln(gv)}{\partial E^r}, \tag{15b}$$

$$h = h_4 = -\frac{1}{2}\frac{\partial \ln g}{\partial K_\varphi}, \tag{15c}$$

$$g = \left(\frac{\partial f}{\partial K_\varphi}\right)^2 \left(1 + \frac{w}{f^2}\right)^{-1}, \tag{15d}$$

$$f_0 = \frac{2}{g}\left(\frac{\partial f^2}{\partial E^r} + \left(1 + f^2\right)\frac{\partial \ln v}{\partial E^r} - V(\phi)\right), \tag{15e}$$

$$h_0 = h_1 = h_2 = h_3 = 0\,. \tag{15f}$$

Notice that when $V(\phi) = 0$, the solutions for f_0, f_1, f_2 and h take exactly the same form as that in the previous section. However, while in the vacuum model the global factor was completely free, when we couple matter g is determined by f and w through relation (15d). In essence, we lost one degree of freedom involving K_φ.

Classically, we have that $f(E^r, K_\varphi) = K_\varphi$. Therefore, in the classical limit $v \to \sqrt{E^r}$ and $w \to 0$. Indeed, we fix these two functions to their classical values in order

to find the polymerized constraint that is closest to (5). In this way, we obtain

$$\begin{aligned}\mathcal{H} = &-\frac{E^\varphi}{2\sqrt{E^r}}\left(1+f^2+4E^r f\frac{\partial f}{\partial E^r}\right)-2\sqrt{E^r}K_r f\frac{\partial f}{\partial K_\varphi}\\ &+\left(\frac{(E^{r\prime})^2}{8\sqrt{E^r}E^\varphi}-\frac{\sqrt{E^r}}{2{E^\varphi}^2}E^{r\prime}E^{\varphi\prime}+\frac{\sqrt{E^r}E^{r\prime\prime}}{2E^\varphi}\right)\left(\frac{\partial f}{\partial K_\varphi}\right)^2\\ &+\left(\frac{P_\phi^2}{2\sqrt{E^r}E^\varphi}+\frac{{E^r}^{3/2}(\phi')^2}{2E^\varphi}\right)\left(\frac{\partial f}{\partial K_\varphi}\right)^2+\sqrt{E^r}E^\varphi V(\phi)\\ &+\frac{\sqrt{E^r}}{2{E^\varphi}^2}E^{r\prime}\left(E^{r\prime}K_r+\phi' P_\phi\right)\frac{\partial f}{\partial K_\varphi}\frac{\partial^2 f}{\partial {K_\varphi}^2}+\frac{\sqrt{E^r}}{2E^\varphi}(E^{r\prime})^2\frac{\partial f}{\partial K_\varphi}\frac{\partial^2 f}{\partial E^r\partial K_\varphi},\end{aligned}\tag{16}$$

which forms a first-class algebra along with the classical diffeomorphism constraint for any function $f = f(E^r, K_\varphi)$. This result allows us to consider holonomy corrections in the presence of a scalar field. In fact, these modifications can be made scale-dependent thanks to the free dependence on E^r. Nonetheless, in order to obtain a K_φ-periodic Hamiltonian constraint, we will restrict ourselves to a scale-invariant function. Setting $f = \sin(\lambda K_\varphi)/\lambda$,

$$\begin{aligned}\mathcal{H}^{(pol)} = &-\frac{E^\varphi}{2\sqrt{E^r}}\left(1+\frac{\sin^2(\lambda K_\varphi)}{\lambda^2}\right)-\sqrt{E^r}K_r\frac{\sin(2\lambda K_\varphi)}{\lambda}\left(1+\left(\frac{\lambda E^{r\prime}}{2E^\varphi}\right)^2\right)\\ &+\left(\frac{(E^{r\prime})^2}{8\sqrt{E^r}E^\varphi}-\frac{\sqrt{E^r}}{2{E^\varphi}^2}E^{r\prime}E^{\varphi\prime}+\frac{\sqrt{E^r}}{2E^\varphi}E^{r\prime\prime}\right.\\ &\qquad\left.+\frac{P_\phi^2}{2\sqrt{E^r}E^\varphi}+\frac{{E^r}^{3/2}(\phi')^2}{2E^\varphi}\right)\cos^2(\lambda K_\varphi)\\ &+\sqrt{E^r}E^\varphi V(\phi)-\frac{\sqrt{E^r}}{4{E^\varphi}^2}E^{r\prime}\phi' P_\phi\lambda\sin(2\lambda K_\varphi).\end{aligned}\tag{17}$$

This constraint satisfies relation (3c) with the deformation function,

$$\beta = \cos^2(\lambda K_\varphi)\left(1+\left(\frac{\lambda E^{r\prime}}{2E^\varphi}\right)^2\right).\tag{18}$$

Following the results in the previous section, we can find the effective mass for this polymerized model:

$$m_{pol} = \frac{\sqrt{E^r}}{2}\left(1+\frac{\sin^2(\lambda K_\varphi)}{\lambda^2}-\left(\frac{E^{r\prime}}{2E^\varphi}\right)^2\cos^2(\lambda K_\varphi)\right).\tag{19}$$

While in vacuum it is a weak observable, the addition of matter makes its evolution non-trivial. Nevertheless, in either case, we can write the deformation as a function of this polymerized mass:

$$\beta = 1+\lambda^2\left(1-\frac{2m_{pol}}{\sqrt{E^r}}\right).\tag{20}$$

In this way, β is covariantly defined. The positiveness requirement for β that can be read from (18) means that the model has a lower bound for the radial component of the triad,

$$\sqrt{E^r} \geq \frac{2\lambda^2 m_{pol}}{1+\lambda^2}. \tag{21}$$

Its saturation corresponds to $\beta = 0$, i.e. the point where the curvature departs most from its classical behavior. In vacuum, m_{pol} should be positive (it is a generalized Schwarzschild mass), and the classical singularity at $E^r = 0$ is inside the forbidden region. In the dynamical case, in order to check whether the system might reach the singularity, one should study the behavior of m_{pol} as E^r approaches zero.

5. Discussion

We have systematically built modified spherically symmetric models coupled to a scalar matter field through a comprehensive study of Dirac's deformation algebra for a generic constraint with quadratic dependence on radial derivatives. The main result of our study is the modified Hamiltonian constraint (9), along with the consistency equations (10) that ensure that the theory remains covariant.

We have found the general vacuum solution in terms of two free functions, which provides the anomaly-free Hamiltonian (12). Moreover, the weak Dirac observable (13) can be interpreted as the mass of the modified system since its classical limit yields the Schwarzschild mass.

Then, we have obtained a particular solution for the full system of anomaly equations and found that a covariant polymerization of the system is possible even under the presence of a scalar field. This solution has led us to the constraint (16), which is parametrized by one free function of E^r and K_φ.

Finally, the polymeric constraint (17) shows a regular behavior for all values of K_φ and forms a first-class algebra with a deformation given by (20). The vanishing of this function corresponds to a minimum positive value of E^r and avoids the classical singularity.

Acknowledgments

AAB receives financial support from the FPI PRE2018-086516 fellowship. This work is funded by Project FIS2017-85076-P (MINECO/AEI/FEDER, UE) and Basque Government Grant No. IT956-16.

References

1. A. Ashtekar and P. Singh, "Loop Quantum Cosmology: A Status Report", *Class. Quant. Grav.* **28**, 213001 (2011).
2. I. Agullo and P. Singh, "Loop Quantum Cosmology" in *Loop quantum gravity: The first 30 years*, edited by A. Ashtekar and J. Pullin (WSP, 2017), pp. 183-240.

3. M. Bojowald and R. Swiderski, "Spherically symmetric quantum geometry: Hamiltonian constraint", *Class. Quant. Grav.* **23**, 2129-2154 (2006).
4. M. Bojowald, S. Brahma, and J. D. Reyes, "Covariance in models of loop quantum gravity: Spherical symmetry", *Phys. Rev. D* **92**, 045043 (2015).
5. A. Alonso-Bardaji and D. Brizuela, "Holonomy and inverse-triad corrections in spherical models coupled to matter", *Eur. Phys. J. C* **81**, 283 (2021).
6. A. Alonso-Bardaji and D. Brizuela, "Anomaly-free deformations of spherical general relativiy coupled to matter", *Phys. Rev. D* **104**, 084064 (2021).

Infrared signatures of quantum bounce in collapsing geometry

Harkirat Singh Sahota* and Kinjalk Lochan†

Department of Physical Sciences, Indian Institute of Science Education & Research (IISER) Mohali, Sector 81 SAS Nagar, Manauli PO 140306 Punjab, India

**E-mail: ph17078@iisermohali.ac.in*

†E-mail: kinjalk@iisermohali.ac.in

We study the radiation profile of the unitarily evolving wave packet constructed for the quantum model of spherically symmetric dust shell collapsing in marginally bound Lemaître-Tolman-Bondi (LTB) model. In this analysis, we consider the quantum model of dust shell collapse in LTB spacetime,[1] where the dust shell collapse to black hole singularity is replaced by a bounce. We identify the observable natural to collapse/expansion character of dust shell, and study the mode decomposition in the quantum model. The incoming/outgoing modes are associated with the eigenfunctions of the Hermitian extension of the operator corresponding to this observable. For the wave packet representing the collapsing and expanding phase of the dust shell, we estimate the contributions of the incoming/outgoing modes. We find that the collapsing and expanding branches do not comprise entirely of incoming and outgoing radiation. The dust shell dynamics is insensitive to the large wavenumber modes as their contribution is negligible. Near the bounce point, the contribution of outgoing (incoming) modes in the collapsing (expanding) branch is substantial and it decreases as the dust shell moves away from the singularity. In the early (later) stage of the collapsing (expanding) phase, the incoming (outgoing) modes dominate. As the dynamics of the dust shell is sensitive to the near-infrared modes of the radiation, the information of the bounce is carried over to infrared modes much before it reaches the observer. In the infrared regime, a flip is observed from largely incoming to largely outgoing radiation as the evolution progresses from collapsing to expanding phase. The information of the short-scale physics is carried over to the longest wavelength in this quantum gravity model.

Keywords: Shell collapse; quantum bounce.

1. Introduction

Ever since the original analysis by Hawking and Unruh,[2,3] the study of particle content of the vacuum and the radiation profile in various geometries have been the frontier research topic in the quantum effects in curved spacetime. These effects are investigated in the context of quantum field theories in the curved spacetime.[4] The key result in this paradigm is that the particle content in the quantum states is not a generally covariant notion, which is the gist of the Hawking-Unruh effect.

The analysis of matter collapse in a consistent quantum gravity model has been addressed in the context of dust shell collapse,[5,6] LTB dust collapse model in canonical quantum gravity[7,8] and loop quantum gravity.[9–11] The radiation profile of dust collapse in the LTB model is studied in[12,13] under the midisuperpace quantization of this model. The Hawking radiation is recovered from the regularized solution of

the Wheeler-DeWitt equation for this model and the non-thermal correction to the spectrum and entropy associated are also computed. In this approach, the Hawking radiation is thought of as the projection of wave functional along the outgoing modes defined in that approach.

There are some issues associated with the aforementioned quantum models. The incoming/outgoing modes defined in this approach are not orthogonal, thus, rendering the notion of incoming/outgoing modes mathematically ill-defined. This issue stems from the fact that the momentum operator in this model is not Hermitian. Apart from that the states in the Kernel of Hamiltonian constraint need not be orthonormal as they are degenerate states with zero eigenvalue.

In this work, we study the radiation profile for the unitarily evolving wave packet constructed for the quantum model for dust shell collapse in a marginally bound LTB geometry.[1] For the collapsing/expanding character of the dust shell, momentum conjugate to areal radius is a good indicator. In the quantum model, we study the mode decomposition viz-a-viz the eigenstates of momentum operator.

2. Minisuperspace Construction of Dust Collapse in Marginal LTB Model

LTB model is an inhomogeneous extension of the FRW model, which is spherically symmetric and sourced by a non-rotational dust of energy density ϵ.[14] The line element for the LTB model is,

$$ds^2 = -c^2 d\tau^2 + \frac{R'^2}{1+2f(\rho)} d\rho^2 + R^2 d\Omega^2, \tag{1}$$

here τ is the dust proper time and $R(\tau, \rho)$ is the areal radius of dust shell with coordinate label ρ at time τ, and $f(\rho)$ is called the energy function. Here, we will restrict ourselves with the marginally bound case of the LTB model for which $f(\rho) = 0$. Einstein's equation for this model is,

$$\frac{F'}{R^2 R'} = \frac{8\pi G\epsilon}{c^2} \quad \text{and} \quad \frac{R\dot{R}^2}{c^2} = F, \tag{2}$$

where $F(\rho)$ is first integral of Einstein's equation and is equal to twice of the Misner-Sharp(MS) mass[15] for LTB spacetime. This model has a curvature singularity when the cloud collapses to a point i.e., at $R = 0$.

Since the equation of motion which dictates the dynamics of the areal radius (2) depends only on R and F, and not on their spatial derivatives, it implies that the different dust shells are dynamically decoupled for a given mass function. We can write an on-shell action which dictates the dynamics of the outermost dust shell. The dynamics of the full dust cloud is then deduced from the action.[1]

$$\mathcal{S} = -\frac{1}{2} \int d\tau R\dot{R}^2 \tag{3}$$

Here R is the areal radius of the outermost dust shell. The Hamiltonian for this model is,

$$H = -\frac{P^2}{2R} \tag{4}$$

Using Brown-Kuchař dust[16,17] as the matter source, the Hamiltonian constraint for the model takes the form

$$\mathcal{H} \equiv p_\tau + H \approx 0. \tag{5}$$

Since the momentum conjugate to scalar field τ appears linearly, the Wheeler-DeWitt equation takes a Schrödinger equation like form,

$$i\hbar \frac{\partial \Psi(R,\tau)}{\partial \tau} = \frac{1}{2} R^{-1+a+b} \frac{d}{dR} R^{-a} \frac{d}{dR} R^{-b} \Psi(R,\tau). \tag{6}$$

The Hamiltonian is the product of areal radius and conjugate momentum, which implies that it does not have a unique quantum counterpart and that the model suffers from operator ordering ambiguity which is parameterized by a and b in equation (6). The eigenvalue of the Hamiltonian is interpreted as the ADM energy. The Hilbert space $L^2(\mathbb{R}^+, R^{1-a-2b}dR)$ is chosen, that makes this Hamiltonian Hermitian. The self adjoint extensions of the Hamiltonian (6) are discussed in.[1] A unitarily evolving wave packet is constructed from positive energy eigenstates,

$$\phi^1_E(R) = \frac{2}{\sqrt{3}} E^{\frac{1}{4}} J_{\frac{2|q|}{3}} \left(\frac{2}{3} \sqrt{2E} R^{\frac{3}{2}} \right). \tag{7}$$

by choosing a Poisson-like distribution

$$A(\sqrt{E}) = \frac{\sqrt{2}\lambda^{\frac{1}{2}(\kappa+1)}}{\sqrt{\Gamma(\kappa+1)}} \sqrt{E}^{\kappa+\frac{1}{2}} e^{-\frac{\lambda}{2}E}, \tag{8}$$

$$\psi(R,\tau) = \int_0^\infty d\sqrt{E} \phi_E(R) e^{iE\tau} A(\sqrt{E}), \tag{9}$$

where the parameters of the distribution satisfy $\kappa \geq 0$ and $\lambda > 0$. The expectation value of the Hamiltonian operator with this choice of distribution is inversely proportional to λ. To simplify the expression, a prescription $\kappa = |1+a|/3$ is adopted[1] and the expression for the wave packet reduces to

$$\psi(R,\tau) = \sqrt{3} \frac{R^{\frac{1}{2}(1+a+|1+a|+2b)}}{\sqrt{\Gamma(\frac{1}{3}|1+a|+1)}} \left(\frac{\frac{\sqrt{2\lambda}}{3}}{\frac{\lambda}{2} - i\tau} \right)^{\frac{1}{3}|1+a|+1} e^{-\frac{2R^3}{9\left(\frac{\lambda}{2} - i\tau\right)}}. \tag{10}$$

This choice comes at the cost that the distribution is now a function of operator ordering parameters. Since the dependence of observables on this parameter is contentious, one cannot be sure if it is a genuine signature of operator ordering or just a dependence on the shape of the distribution. In this analysis, we will focus on the radiation profile of the dust shell and address operator ordering ambiguity elsewhere.[18]

3. Mode Decomposition of Wave Packet

In the earlier midisuperspace models, various prescriptions are used to write incoming and outgoing modes. In,[12] the modes are associated with the asymptotic limit of the solutions of the WDW equation on $\mathcal{I}^+$ written in terms of the Killing time. While in,[13] the modes are associated with the positive/negative frequency solutions of the WDW equation after making transformation from the comoving time to the Killing time. In all these prescriptions, the modes are not orthonormal, making them ill-defined for the Bogoliubov analysis. The problem might lie with the fact that the states belong to the Kernel space of the operator and are degenerate and therefore, orthonormality is not guaranteed. Therefore, we need to identify another observable suitable for the mode decomposition.

Classically, the momentum conjugate to areal radius is given by $P = -R\dot{R}$ and we can associate positive momentum with the collapsing phase and negative momentum with the expanding phase of the dust shell. Thus, the momentum operator is a good choice for an observable that depicts the mode decomposition in the quantum model. The model[1] discussed in the previous section is robust enough for us to accommodate Hermitian Hamiltonian and momentum operator. Although the momentum operator on real half line is not self-adjoint, we will work with the Hermitian extension of the momentum operator. The detailed discussion on the self-adjointness and the Hermiticity of the momentum operator can be found in[18] and we have shown that the states are orthogonal in this case, making them suitable as incoming/outgoing modes.

Using the measure R^2 following the quantum scattering theory in spherical polar coordinates, we have a constraint $1+a+2b = 0$ on operator ordering parameters. For this model, the expectation value of the general observable in a general wave packet is independent of the parameter b.[18] Therefore it appears as a free parameter in the model and the above constraint does not affect the physical content of the model. In this case, the Hermitian extension of the momentum operator is $\hat{P} = -iR^{-1}\partial_R R$. The eigenfunction of the momentum operator with eigenvalue k is given by e^{ikR}/R, where $k \in \mathbb{R}$. The wave packet can be expressed in the form,

$$\psi(R,\tau) = \int_0^\infty dE \int_0^\infty dk \left(\mathcal{A}(k,E) \frac{e^{ikR+i\tau E}}{R} + \mathcal{A}(-k,E) \frac{e^{-ikR+i\tau E}}{R} \right). \tag{11}$$

The contribution of incoming/outgoing radiation in the dust cloud is estimated by computing the projection of wave packet along incoming($u_{k,E} \equiv e^{ikR+iE\tau}/R$) and outgoing modes($u_{-k,E} \equiv e^{-ikR+iE\tau}/R$) .

$$\tilde{\psi}_k(\tau) = \langle u_{k,E}(\tau) | \psi(R,\tau) \rangle = \int dR\, R^2\, \psi(R,\tau) \frac{e^{-ikR-iE\tau}}{R}, \tag{12}$$

$$= \frac{\sqrt{3} e^{-iE\tau}}{\sqrt{\Gamma(\frac{2}{3}|b|+1)}} \left(\frac{\frac{\sqrt{2\lambda}}{3}}{\frac{\lambda}{2} - i\tau} \right)^{\frac{2}{3}|b|+1} \int dR\, e^{-i\,kR} R^{1+|b|} e^{-\frac{2R^3}{9(\frac{\lambda}{2}-i\tau)}}, \tag{13}$$

$$= C(\tau) \int dR\, e^{-i\,kR} R^{1+|b|} e^{-\frac{2R^3}{9(\frac{\lambda}{2}-i\tau)}}. \tag{14}$$

The wavepacket in k space is normalized which implies $|\tilde{\psi}_k(\tau)|^2$ gives the contribution of modes with wavenumber k to the radiation profile at time τ. At $\tau = 0$, equation (14) can be written as

$$\tilde{\psi}_k(0) = C(0)\left(\int dR\ \cos(kR)\, R^{1+|b|} e^{-\frac{4R^3}{9\lambda}} - i\int dR\ \sin(kR)\, R^{1+|b|} e^{-\frac{4R^3}{9\lambda}}\right), \quad (15)$$

and we can see, $|\psi_k(0)|^2 = |\psi_{-k}(0)|^2$. Thus, at the point of classical singularity, the number of incoming modes is equal to the number of outgoing modes for all k. On the other hand, for finite τ, the equation (14) can be cast into the form,

$$|\tilde{\psi}_k(\tau)|^2 = |C(\tau)|^2 \int dR \int dS\ e^{-ik(R-S)} (RS)^{1+|b|} e^{-\frac{2R^3}{9\left(\frac{\lambda}{2}-i\tau\right)} - \frac{2S^3}{9\left(\frac{\lambda}{2}+i\tau\right)}}. \quad (16)$$

Here, $k \to -k$ is same as $\tau \to -\tau$. Thus, the ratio of incoming to outgoing modes $r_k(\tau) = |\tilde{\psi}_k(\tau)|^2/|\tilde{\psi}_{-k}(\tau)|^2$ flips after the bounce happens i.e. $r_k(\tau) = [r_k(-\tau)]^{-1}$, which can be seen in Fig. 1. The difference between the number of incoming and outgoing modes at an instant of time and at a fixed k can be written as,

$$\delta_k(\tau) = |\tilde{\psi}_k(\tau)|^2 - |\tilde{\psi}_{-k}(\tau)|^2, \quad (17)$$

$$= -2i|C(\tau)|^2 \int dR \int dS\ \sin(k(R-S))(RS)^{1+|b|} e^{-\frac{2R^3}{9\left(\frac{\lambda}{2}-i\tau\right)} - \frac{2S^3}{9\left(\frac{\lambda}{2}+i\tau\right)}}. \quad (18)$$

$\delta_k(\tau)$ vanishes when $k \to \infty$, as $\lim_{k\to\infty} \sin(kx) = x\delta(x) = 0$. Thus, we expect the ratio to approach unity for large k, which can be seen in Fig. 1. On the other hand, for $k \to 0$ we have

$$\delta_k(\tau) = \frac{4k|C(\tau)|^2}{9^{(|b|+8)/3}} \Gamma\left[\frac{|b|}{3}+1\right] \Gamma\left[\frac{|b|+2}{3}\right] \left(\frac{\lambda^2}{4}+\tau^2\right)^{\frac{2|b|+5}{6}} \sin\left(\frac{1}{3}\tan^{-1}\left(\frac{-2\tau}{\lambda}\right)\right). \quad (19)$$

In the collapsing branch ($\tau < 0$), this function is positive and therefore, the incoming modes are dominating in this case. In the expanding branch ($\tau > 0$), this function is negative and the outgoing modes dominate in this case.

We have plotted the fraction of modes with wavenumber k and the ratio of the incoming to outgoing modes for fixed time in the Fig. 1 with the parameters specifying a narrow wave packet of unit energy. Early in the collapsing phase ($\tau = -10$), most of the contribution comes from the incoming modes and the outgoing modes contribute a small fraction only in the infrared (small k) regime. As we approach the classical singularity, the contribution of outgoing radiation in the collapsing phase keeps on increasing and becomes equals to the contribution of incoming radiation at the bounce point. After the bounce, initially in the expanding phase, there is significant contribution of incoming radiation. As the expansion progresses, this contribution keeps on decreasing to become negligible and contributes only in the infrared regime.

In the early stage of collapse, the ratio of incoming to outgoing modes starts from unity at $k = 0$ and comes back to one again at a finite wavenumber. In between,

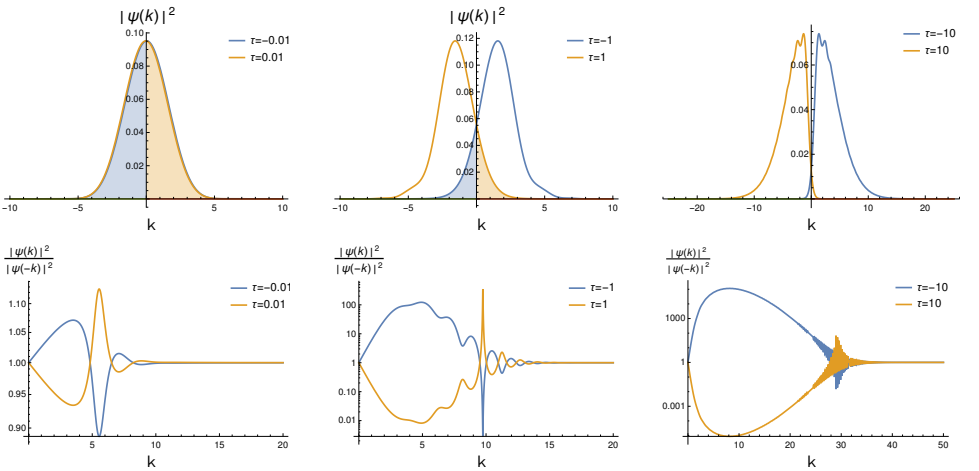


Fig. 1. Fraction of modes $|\tilde{\psi}_k(\tau)|^2$ and ratio of the incoming to outgoing modes vs k at different times. The blue shaded region gives the contribution of outgoing radiation in collapsing phase and the orange shaded region gives the contribution of incoming radiation in expanding phase.

the ratio attains a maximum and oscillates before settling at unity. Since for large wavenumbers, the contribution is minimal, these modes are insensitive to the dynamics of the dust shell. A ratio greater than one implies that the incoming modes are dominating. As we approach the bounce point, the magnitude of the maximum decreases and the ratio goes to unity at shorter wavenumber. At the bounce point, this ratio is one for all wavenumbers. Apart from this, there exists a crossover window of wavenumbers in which the fraction of outgoing modes is greater than incoming modes in the collapsing phase when the dust shell is closer to the classical singularity. This behavior is inverted for the expanding phase - the ratio attains a minimum before settling at unity.

Therefore, we see that the small wavenumber (infrared) modes are most sensitive to the dynamics of the dust cloud. If one observes in the infrared regime, there is an instantaneous flip from largely incoming to largely outgoing radiation when the dust shell goes from collapsing phase to expanding phase. Moreover, the contribution of the outgoing/incoming radiation in the collapsing/expanding regime is coming from the infrared part of the spectrum only. Thus, one should focus on the infrared regime of the dust shell for a signature of quantum bounce.

We have plotted the radiation profile for the wave packets which represents low energy dust shell broadly peaked ($\lambda = 10$) and high energy dust shells sharply peaked ($\lambda = 0.01$) on the classical trajectory in Fig. 2. We see that for the case of sharply peaked wave packet, the major contribution to radiation profile is coming from incoming (outgoing) modes in the collapsing (expanding) regime even near the classical singularity. Whereas, for low energy dust shells, the contribution of incoming (outgoing) modes in expanding (collapsing) branch is significant even far away from singularity.

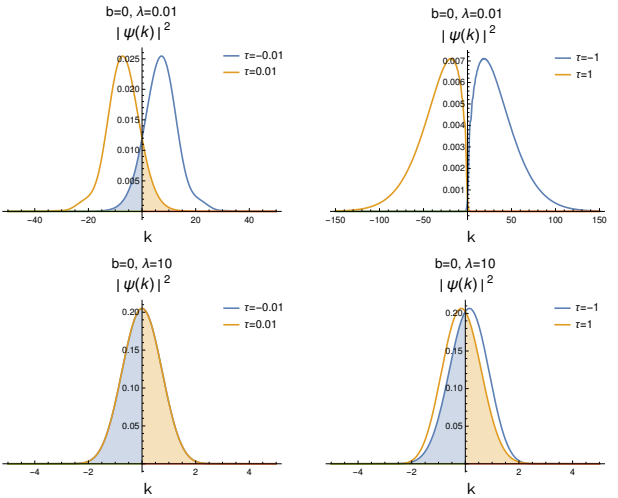


Fig. 2. Fraction of modes $|\tilde{\psi}_k(\tau)|^2$ and the ratio of incoming to outgoing modes for wave at different times. The parameters specify the high energy dust cloud ($\lambda = 0.01$) representing sharper wave packet, and low energy dust cloud($\lambda = 10$) representing broader wave packet.

At the classical singularity, the fraction of infrared modes is directly proportional to the minimal size of the dust shell i.e., the bounce radius.[18]

$$|\tilde{\psi}_{k\to 0}(0)|^2 = \frac{2^{(2|b|-5)/3}3^{2/3}\Gamma\left(\frac{|b|+2}{3}\right)^2\Gamma\left(\frac{2|b|+4}{3}\right)}{\sqrt{2}\pi\Gamma\left(\frac{2|b|+1}{3}\right)\Gamma\left(\frac{2|b|+3}{3}\right)}\bar{R}(0). \tag{20}$$

Thus, the infrared regime of the radiation profile provides a direct estimation of the bounce radius.

4. Conclusions

In this work, we have studied the mode decomposition of wave packet constructed for the Quantum LTB model. We have considered the minisuperspace construction of the dust shell collapse in the LTB model.[1] The classical model of dust shell collapse exhibits black hole singularity which is replaced by bounce from the collapsing phase to expanding phase in the quantum model. We have identified the observable depicting mode decomposition, which is momentum conjugate to areal radius. After identifying the incoming and outgoing modes with the momentum's eigenstates with positive and negative eigenvalues, we have estimated the contribution of the incoming/outgoing modes in the contracting/expanding phase.

We choose the operator ordering parameters for which the Hamiltonian as well as the momentum operator are Hermitian. This is achieved by working with R^2 measure space and choosing the representation which is symmetric with this choice

of inner product. This particular choice puts the constraint $a+2b+1 = 0$ on operator ordering parameters.

We find that at the point of classical singularity or bounce point $\tau = 0$, the number of incoming and outgoing modes is the same. In the contracting branch, apart from the incoming dust, we also have small contribution from the outgoing dust in the infrared regime. As the dust shell continues to move towards the singularity configuration, the contribution of outgoing modes in the infrared regime keeps increasing, culminating in an equal number of incoming and outgoing modes at the point of singularity. This behavior is inverted in the expanding branch. There exists a threshold wavenumber, above which the number of incoming modes is equal to the number of outgoing modes. For the contracting branch, the ratio of incoming to outgoing modes starts from one at $k = 0$, increases to attain a maximum, and then settles back to unity. After the bounce, outgoing modes start to dominate with a small fraction of incoming modes as well. As the dust shell expands, the fraction of the incoming modes keeps decreasing and while the contribution of the outgoing modes keeps increasing. At the later stage of dust shell expansion, most contribution comes from the outgoing modes. Thus, we can conclude that in quantum bounce, incoming radiation is always accompanied by outgoing radiation.

The infrared sector of the wave packet contains significant information about the dynamics of the dust cloud. The major contribution to the incoming/outgoing dust in collapsing/expanding branch comes from the modes with smaller wavenumber. There is a flip from largely incoming to largely outgoing radiation, observed in the infrared regime, as the evolution progresses from the contracting branch to the expanding branch. Therefore, the observer should look at small wavenumber or large wavelength modes to observe if the bounce has happened. Moreover, the fraction of infrared modes near classical singularity is proportional to the bounce radius. Thus, the infrared sector of the process apart from being highly sensitive to the dynamics of the dust cloud is also a direct estimator of the bounce radius, thereby, providing a unique infrared signature of the quantum gravity in the radiation profile of the dust cloud.

Acknowledgments

HSS would like to acknowledge the financial support from University Grants Commission, Government of India, in the form of Junior Research Fellowship (UGC-CSIR JRF/Dec- 2016/503905). Research of KL is partially supported by the Startup Research Grant of SERB, Government of India (SRG/2019/002202).

References

1. C. Kiefer and T. Schmitz, Singularity avoidance for collapsing quantum dust in the Lemaître-Tolman-Bondi model, *Physical Review D* **99**, p. 126010 (June 2019).
2. S. W. Hawking, Particle Creation by Black Holes, *Commun. Math. Phys.* **43**, 199 (1975), [Erratum: Commun. Math. Phys. 46, 206 (1976)].

3. W. G. Unruh, Notes on black-hole evaporation, *Phys. Rev. D* **14**, 870 (Aug 1976).
4. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics (Cambridge Univ. Press, Cambridge, UK, 2 1984).
5. P. Hájíček and C. Kiefer, Singularity Avoidance by Collapsing Shells in Quantum Gravity, *International Journal of Modern Physics D* **10**, 775 (December 2001).
6. D. Malafarina, Classical Collapse to Black Holes and Quantum Bounces: A Review, *Universe* **3**, p. 48 (May 2017).
7. C. Vaz, L. Witten and T. P. Singh, Toward a midisuperspace quantization of LeMaître-Tolman-Bondi collapse models, *Physical Review D* **63**, p. 104020 (April 2001).
8. C. Vaz and L. Witten, Quantum black holes from quantum collapse, *Physical Review D* **64**, p. 084005 (September 2001).
9. M. Bojowald, T. Harada and R. Tibrewala, Lemaitre-Tolman-Bondi collapse from the perspective of loop quantum gravity, *Phys. Rev. D* **78**, p. 064057 (September 2008).
10. M. Bojowald, J. D. Reyes and R. Tibrewala, Nonmarginal Lemaitre-Tolman-Bondi-like models with inverse triad corrections from loop quantum gravity, *Phys. Rev. D* **80**, p. 084002 (October 2009).
11. K. Giesel, B.-F. Li and P. Singh, Non-singular quantum gravitational dynamics of an LTB dust shell model: The role of quantization prescriptions, *arXiv:2107.05797 [gr-qc]* (Aug 2021).
12. C. Vaz, C. Kiefer, T. P. Singh and L. Witten, Quantum general relativity and Hawking radiation, *Physical Review D* **67**, p. 024014 (January 2003).
13. C. Kiefer, J. Müller-Hill, T. P. Singh and C. Vaz, Hawking radiation from the quantum Lemaître-Tolman-Bondi model, *Physical Review D* **75**, p. 124010 (June 2007).
14. A. Krasinski, *Inhomogeneous cosmological models* (Cambridge Univ. Press, Cambridge, UK, 3 2011).
15. L. B. Szabados, Quasi-Local Energy-Momentum and Angular Momentum in General Relativity, *Living Reviews in Relativity* **12**, p. 4 (December 2009).
16. J. D. Brown and K. V. Kuchar, Dust as a Standard of Space and Time in Canonical Quantum Gravity, *Physical Review D* **51**, 5600 (May 1995).
17. H. Maeda, Unitary evolution of the quantum universe with a Brown-Kuchar dust, *Classical and Quantum Gravity* **32**, p. 235023 (December 2015).
18. H. S. Sahota and K. Lochan, Infrared signatures of quantum bounce in collapsing geometry, *arXiv:2110.06247 [gr-qc]* (Oct 2021).

Effective black hole interior and the Raychadhuri equation

Keagan Blanchette

Department of Physics and Astronomy, York University
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada
E-mail: kblanch@yorku.ca

Saurya Das

Theoretical Physics Group and Quantum Alberta, Department of Physics and Astronomy, University of Lethbridge, 4401 University Drive, Lethhbridge, Alberta T1K 3M4, Canada
E-mail: saurya.das@uleth.ca

Samantha Hergott

Department of Physics and Astronomy, York University
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada
E-mail: sherrgs@yorku.ca

Saeed Rastgoo

Department of Physics and Astronomy, York University
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada
E-mail: srastgoo@yorku.ca

We show that loop quantum gravity effects leads to the finiteness of expansion and its rate of change in the effective regime in the interior of the Schwarzschild black hole. As a consequence the singularity is resolved. We find this in line with previous results about curvature scalar and strong curvature singularities in Kantowski-Sachs model which is isometric to Schwarzschild interior.

Keywords: Black hole singularity, loop quantum gravity, expansion, Raychaudhuri equation

1. Introduction

Singularities are well-known predictions of General Relativity (GR). They are recognized as regions that geodesics can reach in finite proper time but cannot be extended beyond them. Such geodesics are called incomplete. This notion can be formulated in terms of the existence of conjugate points using the Raychaudhuri equation.[1] The celebrated Hawking-Penrose singularity theorems prove that under normal assumptions, all spacetime solutions of GR will have incomplete geodesics, and will therefore be singular.[1–3] These objects, however, are in fact predictions beyond the domain of applicability of GR. So there is a consensus among the gravitational physics community that they should be regularized in a full theory of quantum gravity. Although there is no such theory available yet, nevertheless there are a few candidates with rigorous mathematical structure with which we can investigate

the question of singularity resolution. One such candidate is loop quantum gravity (LQG),[4] which is a connection-based canonical framework.

Within LQG, there have been numerous studies of both the interior and the full spacetime of black holes in four and lower dimensions.[5–18] These attempts were originally inspired by loop quantum cosmology (LQC), more precisely a certain quantization of the isotropic Friedmann-Lemaitre-Robertson-Walker (FLRW) model[19,20] which uses a certain type of quantization of the phase space variables called polymer quantization.[21–25] This quantiztion introduces a so called polymer parameter that sets the scale at which the quantum effects become important. There are various schemes of such quantization based on the form of the polymer parameter.

In this paper, we examine the issue of singularity resolution via the LQG-modified Raychaudhuri equation for the interior of the Schwarzschild black hole. By choosing adapted holonomies and fluxes, which are the conjugate variables in LQG, and using polymer quantization, we compute the corresponding expansion of geodesics and derive the effective Raychaudhuri equation. This way we show effective terms introduce a repulsive effect which prevents the formation of conjugate points. This implies that the classical singularity theorems are rendered invalid and the singularity is resolved, at least for the spacetime under consideration. The boundedness of expansion scalar and the resolution of strong curvature singularities have been studied and shown in the context of Kantowski-Sachs model in.[26,27]

This paper is organized as follows. In Sec. 2, we review the classical interior of the Schwarzschild black hole. In Sec. 3, we briefly discuss the classical Raychaudhuri equation and its importance. Then, in Sec. 4.1 we present the behavior of the Raychaudhuri equation in the classical regime. In Sec. 4.2, the effective Raychaudhuri equation for three different schemes of polymer quantization are derived and are compared with the classical behavior. Finally, in Sec. 5 we briefly discuss our results and conclude.

2. Interior of the Schwarzschild black hole

It is well known that the metric of the interior of the Schwarzschild black hole can be obtained by switching the the Schwarzschild coordinates t and r, due to the fact that spacelike and timelike curves switch their causal nature upon crossing the horizon. This yields the metric of the interior as

$$ds^2 = -\left(\frac{2GM}{t} - 1\right)^{-1} dt^2 + \left(\frac{2GM}{t} - 1\right) dr^2 + t^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right). \quad (1)$$

This metric is in fact a special case of a Kantowski-Sachs cosmological spacetime[28]

$$ds^2_{KS} = -N(T)^2 dT^2 + g_{xx}(T)dx^2 + g_{\theta\theta}(T)d\theta^2 + g_{\phi\phi}(T)d\phi^2 \quad (2)$$

One can obtain the Hamiltonian of the interior system in connection variables, one first considers the full Hamiltonian of gravity written in terms of (the curvature) of the $su(2)$ Ashtekar-Barbero connection A^i_a and its conjugate momentum, the

densitized triad $\tilde{E}^i_a$. Using the Kantowski-Sachs symmetry, these variables can be written as[5]

$$A^i_a \tau_i dx^a = \frac{c}{L_0}\tau_3 dx + b\tau_2 d\theta - b\tau_1 \sin\theta d\phi + \tau_3 \cos\theta d\phi, \tag{3}$$

$$\tilde{E}^a_i \tau_i \partial_a = p_c \tau_3 \sin\theta \partial_x + \frac{p_b}{L_0}\tau_2 \sin\theta \partial_\theta - \frac{p_b}{L_0}\tau_1 \partial_\phi, \tag{4}$$

where b, c, p_b, and p_c are functions that only depend on time and $\tau_i = -i\sigma_i/2$ are a $su(2)$ basis satisfying $[\tau_i, \tau_j] = \epsilon_{ij}{}^k \tau_k$, with σ_i being the Pauli matrices. Here L_0 is a fiducial length of a fiducial volume, chosen to restrict the integration limits of the symplectic form so that the integral does not diverge. Substituting these into the full Hamiltonian of gravity written in Ashtekar connection variables, one obtains the symmetry reduced Hamiltonian constraint adapted to this model as[5–7, 10, 16]

$$H = -\frac{N}{2G\gamma^2}\left[\left(b^2 + \gamma^2\right)\frac{p_b}{\sqrt{p_c}} + 2bc\sqrt{p_c}\right]. \tag{5}$$

while the diffeomorphism constraint vanishes identically due to the homogenous nature of the model. Here, γ is the Barbero-Immirzi parameter,[4] and $p_c \geq 0$. The corresponding Poisson brackets of the model become

$$\{c, p_c\} = 2G\gamma, \qquad \{b, p_b\} = G\gamma. \tag{6}$$

The general form of the Kantowski-Sachs metric written in terms of the above variables becomes

$$ds^2 = -N(T)^2 dT^2 + \frac{p_b^2(T)}{L_0^2\, p_c(T)} dx^2 + p_c(T)(d\theta^2 + \sin^2\theta d\phi^2). \tag{7}$$

Comparing this with the standard Schwarzschild interior metric one obtains

$$p_b = 0, \qquad p_c = 4G^2M^2, \qquad \text{on the horizon } t = 2GM, \tag{8}$$

$$p_b \to 0, \qquad p_c \to 0, \qquad \text{at the singularity } t \to 0. \tag{9}$$

3. The Raychaudhuri equation

The celebrated Raychaudhuri equation[1]

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - R_{ab}U^aU^b \tag{10}$$

describes the behavior of geodesics in spacetime purely geometrically and independent of the theory of gravity under consideration. Here, θ is the expansion term describing how geodesics focus or defocus; $\sigma_{ab}\sigma^{ab}$ is the shear which describes how, e.g., a circular configuration of geodesics changes shape into, say, an ellipse; $\omega_{ab}\omega^{ab}$ is the vorticity term; R_{ab} is the Ricci tensor; and U^a is the tangent vector to the geodesics. Note that, due the sign of the expansion, shear, and the Ricci term, they all contribute to focusing, while the vorticity terms leads to defocusing.

In our case, since we consider the model in vacuum, $R_{ab} = 0$. Also, in general in Kantowski-Scahs models, the vorticity term is only nonvanishing if one considers metric perturbations.[28] Hence, $\omega_{ab}\omega^{ab} = 0$ in our model, too. This reduces the Raychaudhuri equation for our analysis to

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab}. \tag{11}$$

To obtain the right hand side of the equation above, we need to consider a congruence of geodesics and derive their expansion and shear. By choosing such a congruence with 4-velocities $U^a = \left(\frac{1}{N}, 0, 0, 0\right)$ we obtain

$$\theta = \frac{\dot{p}_b}{Np_b} + \frac{\dot{p}_c}{2Np_c}, \tag{12}$$

$$\sigma^2 = \frac{2}{3}\left(-\frac{\dot{p}_b}{Np_b} + \frac{\dot{p}_c}{Np_c}\right)^2. \tag{13}$$

4. Classical vs effective Raychaudhuri equation

4.1. *Classical Raychaudhuri equation*

Having obtained the adapted form of the Raychaudhuri equation for our model, we set to find it explicitly. Looking at (11)-(13) we see that we need the solutions to the equations of motion to be able to compute them. In order to facilitate such a derivation, we choose a gauge where the lapse function is

$$N(T) = \frac{\gamma\sqrt{p_c(T)}}{b(T)}, \tag{14}$$

for which the Hamiltonian constraint becomes

$$H = -\frac{1}{2G\gamma}\left[\left(b^2 + \gamma^2\right)\frac{p_b}{b} + 2cp_c\right]. \tag{15}$$

The advantage of this lapse function is that the equations of motion of c, p_c decouple from those of b, p_b. These equations of motion should be solved together with enforcing the vanishing of the Hamiltonian constraint (15) on-shell (i.e., on the constraint surface). Replacing these solutions into (11) one obtains

$$\theta = \pm\frac{1}{2\sqrt{p_c}}\left(\frac{3b}{\gamma} - \frac{\gamma}{b}\right) = \pm\frac{-2t + 3GM}{t^{\frac{3}{2}}\sqrt{(2GM - t)}}, \tag{16}$$

$$\frac{d\theta}{d\tau} = -\frac{1}{2p_c}\left(1 + \frac{9b^2}{2\gamma^2} + \frac{\gamma^2}{2b^2}\right) = \frac{-2t^2 + 8GMt - 9G^2M^2}{(2GM - t)\,t^3}. \tag{17}$$

As expected, the right hand side of $\frac{d\theta}{d\tau}$ is negative (since $p_c > 0$) and both θ and $\frac{d\theta}{d\tau}$ diverge at the singularity in the classical regime. This can be seen from Fig. 1 which reaffirm the existence of a classical singularity at the center of the black hole.

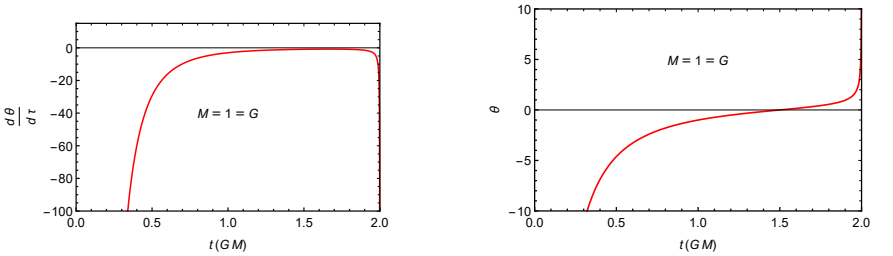


Fig. 1. Left: $\frac{d\theta}{d\tau}$ as a function of the Schwarzschild time t. Right: negative branch of θ as a function of t. Both θ and $\frac{d\theta}{d\tau}$ diverge as we approach $t \to 0$. Note that the divergence at the horizon is due to the choice of Schwarzschild coordinate system.

4.2. *Effective dynamics and Raychaudhuri equation*

The effective behavior of the interior of the Schwarzschild black hole can be deduced from its effective Hamiltonian (constraint). There are various equivalent ways to obtain such an effective Hamiltonian from the classical one.[5–7, 10, 16] It turns out that the easiest way is by replacing

$$b \to \frac{\sin(\mu_b b)}{\mu_b}, \tag{18}$$

$$c \to \frac{\sin(\mu_c c)}{\mu_c}, \tag{19}$$

in the classical Hamiltonian.

The free parameters μ_b, μ_c are the minimum scales associated with the radial and angular directions.[5, 7, 10, 29] If these μ parameters are taken to be constant, the corresponding approach is called the μ_0 scheme. If, however, these parameters depend on the conjugate momenta, the approach is called improved dynamics which itself is divided into various subcategories. In case of the Schwarzschild interior and due to lack of matter content, it is not clear which scheme yields the correct semiclassical limit. Hence, for completeness, in this paper we will study the effective theory in the constant μ scheme, which here we call the $\mathring{\mu}$ scheme, as well as in two of the most common improved schemes, which we denote by $\bar{\mu}$ and $\bar{\mu}'$ schemes.

Replacing (18) and (19) into the classical Hamiltonian (5), one obtains an effective Hamiltonian constraint,

$$H_{\text{eff}}^{(N)} = -\frac{N}{2G\gamma^2}\left[\left(\frac{\sin^2(\mu_b b)}{\mu_b^2} + \gamma^2\right)\frac{p_b}{\sqrt{p_c}} + 2\frac{\sin(\mu_b b)}{\mu_b}\frac{\sin(\mu_c c)}{\mu_c}\sqrt{p_c}\right]. \tag{20}$$

In order to be able to compare the effective results with the classical case, we need to use the same lapse as we did in the classical part. Under (18), this lapse (14) becomes

$$N = \frac{\gamma\mu_b\sqrt{p_c}}{\sin(\mu_b b)}. \tag{21}$$

Using this in (20) we obtain

$$H_{\rm eff} = -\frac{1}{2\gamma G}\left[p_b\left[\frac{\sin(\mu_b b)}{\mu_b} + \gamma^2\frac{\mu_b}{\sin(\mu_b b)}\right] + 2p_c\frac{\sin(\mu_c c)}{\mu_c}\right]. \tag{22}$$

Note that both (20) and (22) reduce to their classical counterparts (5) and (15) respectively, as is expected.

To obtain the effective Raychadhuri equation, we consider three cases:

(1) $\mathring{\mu}$ scheme where $\mu_b = \mathring{\mu}_b$ and $\mu_c = \mathring{\mu}_c$ are constants,
(2) $\bar{\mu}$ scheme where we set

$$\mu_b = \bar{\mu}_b = \sqrt{\frac{\Delta}{p_b}}, \tag{23}$$

$$\mu_c = \bar{\mu}_c = \sqrt{\frac{\Delta}{p_c}}, \tag{24}$$

(3) $\bar{\mu}'$ scheme where we choose

$$\mu_b = \bar{\mu}'_b = \sqrt{\frac{\Delta}{p_c}}, \tag{25}$$

$$\mu_c = \bar{\mu}'_c = \frac{\sqrt{p_c\Delta}}{p_b}. \tag{26}$$

After replacing these (separately for each case) into the effective Hamiltonian constraint (22) and finding their corresponding equations of motion,[30] one replaces the solutions in the Raychadhuri equation (11) to obtain the form of $\frac{d\theta}{d\tau}$. It turns out that for all the three cases above we obtain

$$\begin{aligned}\frac{d\theta}{d\tau} =& \frac{1}{\gamma^2 p_c}\frac{\sin^2(\mu_b b)}{\mu_b^2}\left[\cos(\mu_b b)\cos(\mu_c c) - \frac{\cos^2(\mu_b b)}{4} - 3\cos^2(\mu_c c)\right] \\ &+ \frac{\cos(\mu_b b)}{p_c}\left[\frac{\cos(\mu_b b)}{2} - \cos(\mu_c c) - \frac{\gamma^2}{4}\cos(\mu_b b)\frac{\mu_b^2}{\sin^2(\mu_b b)}\right],\end{aligned} \tag{27}$$

where it is understood that μ's should be substituted for from cases 1–3 suitably for each case.

4.2.1. *$\mathring{\mu}$ scheme*

Let us first consider this case perturbatively in an analytic manner. Replacing $\mu_b = \mathring{\mu}_b$ and $\mu_c = \mathring{\mu}_c$ as constants in (27) and then expanding for small values of μ's up to the second order we obtain

$$\frac{d\theta}{d\tau} \approx -\frac{1}{2p_c}\left(1 + \frac{9b^2}{2\gamma^2} + \frac{\gamma^2}{2b^2}\right) + \mathring{\mu}_b^2\frac{1}{2p_c}\left(\frac{b^4}{\gamma^2} + \frac{\gamma^2}{3}\right) + \mathring{\mu}_c^2\frac{c^2}{2p_c}\left(1 + \frac{5b^2}{\gamma^2}\right). \tag{28}$$

It is seen that the first term on the right-hand side above is the classical expression (17) which is always negative and leads to the divergence of classical expansion

rate at the singularity, i.e., infinite focusing. However, Eq. (28) now involves two additional effective terms proportional to $\mathring{\mu}_b^2$ and $\mathring{\mu}_c^2$, both of which are positive. Furthermore, from the solutions to equations of motion,[30] one can infer that these two terms take over close to where the classical singularity used to be and stop θ and $\frac{d\theta}{d\tau}$ from diverging. This can, in fact, be confirmed by looking at the full nonperturbative behavior of $\frac{d\theta}{d\tau}$ plotted in Fig. 2. There, it is seen that the quantum gravity effects counter the attractive nature of classical terms and turn the curve around such that $\frac{d\theta}{d\tau}$ goes to zero for $t \to 0$.

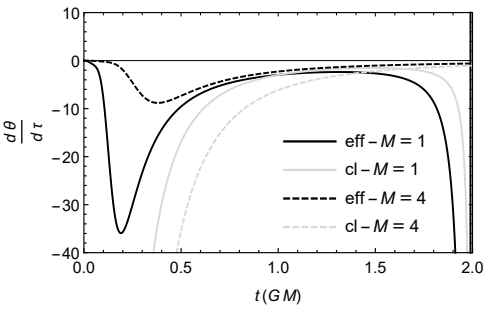


Fig. 2. Plot of $\frac{d\theta}{d\tau}$ as a function of the Schwarzschild time t, for two different masses in classical vs effective regimes. The figure is plotted using $\gamma = 0.5$, $G = 1$, $L_0 = 1$, and $\mathring{\mu}_b = 0.08 = \mathring{\mu}_c$.

4.2.2. $\bar{\mu}$ *scheme*

Similar to the $\mathring{\mu}$ scheme we start by analyzing the perturbative expansion of $\frac{d\theta}{d\tau}$ for this case by replacing (23) and (24) in (27) and expanding it for small $\bar{\mu}$'s up to the lowest correction terms which in this case is Δ (which can be considered as the second order in $\bar{\mu}$ scales). This way we get

$$\frac{d\theta}{d\tau} \approx -\frac{1}{2p_c}\left(1+\frac{9b^2}{2\gamma^2}+\frac{\gamma^2}{2b^2}\right)+\frac{\Delta}{p_c}\left[\frac{1}{6p_b}\left(\frac{3b^4}{\gamma^2}+\gamma^2\right)+\frac{c^2}{2p_c}\left(1+\frac{5b^2}{\gamma^2}\right)\right]. \quad (29)$$

Once again, the first term on the right-hand side is the classical expression of the Raychaudhuri equation (17), which contributes to infinite focusing at the singularity, but all the correction terms are positive and take over close to the position of the classical singularity. This stops $\frac{d\theta}{d\tau}$ from diverging similar to the $\mathring{\mu}$ scheme.

The full nonperturbative form of the modified Raychaudhuri equation in terms of t can also be plotted plotted by substituting the numerical solutions of the equations of motion for the $\bar{\mu}$ in (27). The result is plotted in Fig. 3. We see that, approaching from the horizon to where the classical singularity used to be, an initial bump or bounce in encountered, followed by a more pronounced bounce closer to where the singularity used to be. Once again, the quantum corrections become dominant close to the singularity and turn back the $\frac{d\theta}{d\tau}$ such that at $t \to 0$ no focusing happens at all. Furthermore, from the right plot in Fig. 3, we see that the first bounce in the

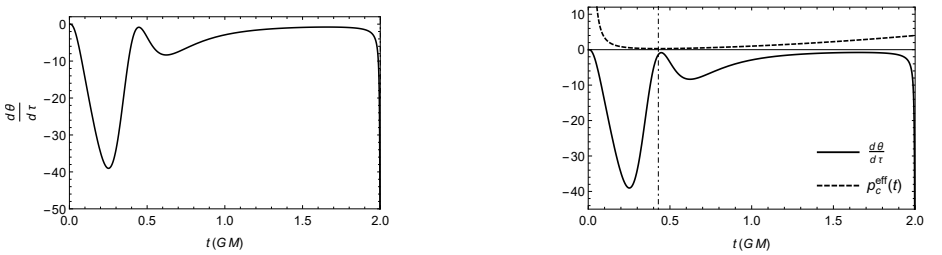


Fig. 3. Left: Raychaudhuri equation in the $\bar{\mu}$ scheme. Right: Raychaudhuri equation vs p_c. The vertical dot-dashed line at $t \approx 0.43GM$ is the position of the bounce of p_c where its minimum $p_c^{\min} = 0.29$ happens in this case. The figure is plotted using $\gamma = 0.5$, $M = 1$, $G = 1$, $L_0 = 1$, and $\Delta = 0.1$.

Raychaudhuri equation happens much earlier than the bounce in p_c for this batch of geodesics.

4.2.3. $\bar{\mu}'$ *scheme*

The perturbative analytical form of $\frac{d\theta}{d\tau}$ for this case, up to the first correction term turns out to be

$$\frac{d\theta}{d\tau} \approx -\frac{1}{2p_c}\left(1 + \frac{9b^2}{2\gamma^2} + \frac{\gamma^2}{2b^2}\right) + \frac{\Delta}{6\gamma^2}\left[\frac{1}{p_c^2}\left(3b^4 + \gamma^4\right) + \frac{3c^2}{p_b^2}\left(5b^2 + \gamma^2\right)\right]. \tag{30}$$

Although this perturbative form of the Raychaudhuri equation is a bit different from previous cases, nevertheless it exhibits the property that the quantum corrections are all positive and take over close to where the classical singularity used to be, and hence once again contribute to defocusing of the geodesics. This case is, however, rather different from the previous two cases since the behavior of some of the canonical variables as a function of the Schwarzschild time t deviates significantly from those cases. In particular, both b and p_c show a kind of damped oscillatory behavior close to the classical singularity,[30] which contributes to a more volatile behavior of the Raychaudhuri equation.

The full nonperturbative Raychaudhuri equation and its close-ups in this case are plotted in Fig. 4. It is seen that in this scheme, the Raychaudhuri equation exhibits a more oscillatory behavior and has various bumps particularly when we get closer to where the singularity used to be. Very close to the classical singularity, its form resembles those of b and p_c, behaving like a damped oscillation.[30]

Two particular features are worth noting in this scheme. First, as we also saw in previous schemes, quantum corrections kick in close to the singularity and dominate the evolution such that the infinite focusing is remedied, hence signaling the resolution of the singularity. Second, this scheme exhibits a nonvanishing value for $\frac{d\theta}{d\tau}$ at, or very close to, the singularity. In Fig. 4 with the particular choice of numerical values of γ, M, G, L_0 and Δ, the value of $\frac{d\theta}{d\tau}$ for $t \to 0$ is approximately -5.5. Hence, although a nonvanishing focusing is not achieved in this case at where

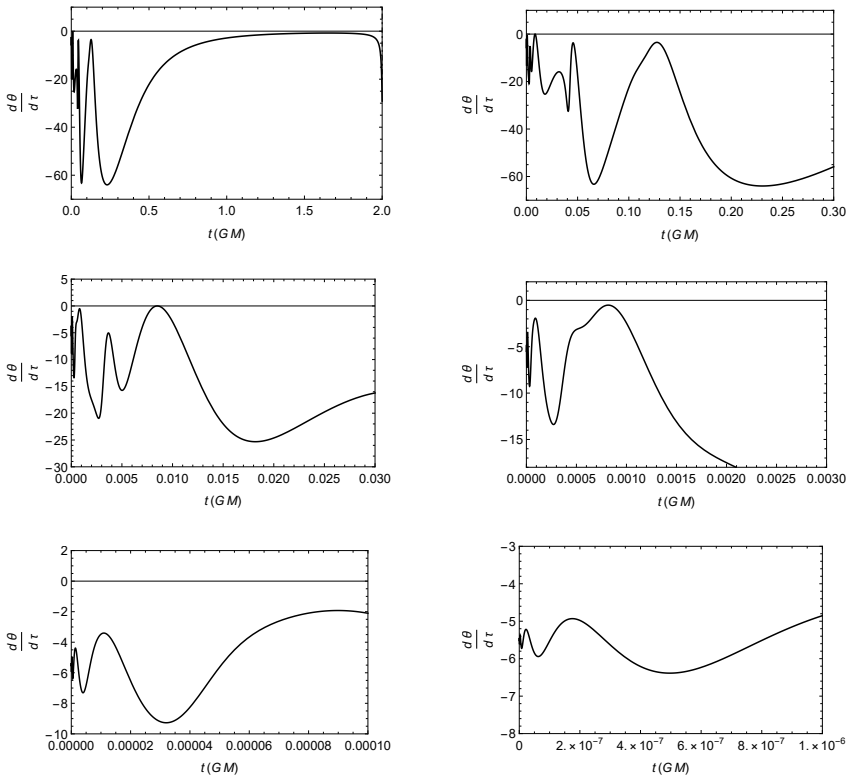


Fig. 4. Raychaudhuri equation in the $\bar{\mu}'$ scheme. The top left figure shows the behavior over the whole $0 \leq t \leq 2GM$ range. Other plots show various close-ups of that plot over smaller ranges of t. The figure is plotted using $\gamma = 0.5$, $M = 1$, $G = 1$, $L_0 = 1$, and $\Delta = 0.1$.

the singularity used to be, nevertheless, there exists a relatively small focusing and certainly $\frac{d\theta}{d\tau}$ remains finite.

5. Discussion and outlook

In this work, we probed the structure of the interior of the Schwarzschild black hole, particularly the region close to the classical singularity, using the effective Raychaudhuri equation. The effective terms in this equation result from considering the effective modifications to the Hamiltonian of the interior due to polymer quantization, which is equivalent to loop quantization of this model. We found out that while the classical rate of expansion $\frac{d\theta}{d\tau}$ diverges for $r \to 0$, the effective terms counter such a divergence close to the singularity and make $\frac{d\theta}{d\tau}$ finite at $r \to 0$. We considered three main schemes of polymer quantization and the results hold in all three. This is a strong indication that LQG points to the resolution of the singularity in the effective regime.

It is also worth noting that very similar behavior has been derived recently for several cases of Generalized Uncertainty Principle (GUP) models.[31,32] In particular, it seems that these cases bare a significant resemblance to $\mathring{\mu}$ and $\bar{\mu}$. This can be taken as a cross-model affirmation that quantum gravity in general does resolve the singularity of the Schwarzschild black hole.

References

1. A. Raychaudhuri, Relativistic cosmology. 1., *Phys. Rev.* **98**, 1123 (1955).
2. R. Penrose, Gravitational collapse and space-time singularities, *Phys. Rev. Lett.* **14**, 57 (1965).
3. S. Hawking and R. Penrose, The Singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. Lond. A* **314**, 529 (1970).
4. T. Thiemann, *Modern Canonical Quantum General Relativity*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2007).
5. A. Ashtekar and M. Bojowald, Quantum geometry and the Schwarzschild singularity, *Class. Quant. Grav.* **23**, 391 (2006).
6. Böhmer, Christian G. and Vandersloot, Kevin, Loop Quantum Dynamics of the Schwarzschild Interior, *Phys. Rev. D* **76**, p. 104030 (2007).
7. A. Corichi and P. Singh, Loop quantization of the Schwarzschild interior revisited, *Class. Quant. Grav.* **33**, p. 055006 (2016).
8. A. Ashtekar, J. Olmedo and P. Singh, Quantum extension of the Kruskal spacetime, *Phys. Rev. D* **98**, p. 126003 (2018).
9. M. Bojowald and S. Brahma, Signature change in loop quantum gravity: Two-dimensional midisuperspace models and dilaton gravity, *Phys. Rev. D* **95**, p. 124014 (2017).
10. D.-W. Chiou, Phenomenological loop quantum geometry of the Schwarzschild black hole, *Phys. Rev. D* **78**, p. 064040 (2008).
11. A. Corichi, A. Karami, S. Rastgoo and T. Vukašinac, Constraint Lie algebra and local physical Hamiltonian for a generic 2D dilatonic model, *Class. Quant. Grav.* **33**, p. 035011 (2016).
12. R. Gambini, J. Olmedo and J. Pullin, Spherically symmetric loop quantum gravity: Analysis of improved dynamics, *Class. Quant. Grav.* **37**, p. 205012 (2020).
13. R. Gambini, J. Pullin and S. Rastgoo, New variables for 1+1 dimensional gravity, *Class. Quant. Grav.* **27**, p. 025002 (2010).
14. S. Rastgoo, A local true Hamiltonian for the CGHS model in new variables (4 2013).
15. A. Corichi, J. Olmedo and S. Rastgoo, Callan-Giddings-Harvey-Strominger vacuum in loop quantum gravity and singularity resolution, *Phys. Rev. D* **94**, p. 084050 (2016).
16. H. A. Morales-Técotl, S. Rastgoo and J. C. Ruelas, Effective dynamics of the Schwarzschild black hole interior with inverse triad corrections (6 2018).
17. J. Olmedo, S. Saini and P. Singh, From black holes to white holes: A quantum gravitational, symmetric bounce, *Class. Quant. Grav.* **34**, p. 225011 (2017).
18. L. Modesto, Black hole interior from loop quantum gravity, *Adv. High Energy Phys.* **2008**, p. 459290 (2008).
19. A. Ashtekar, T. Pawlowski and P. Singh, Quantum nature of the big bang, *Phys. Rev. Lett.* **96**, p. 141301 (2006).
20. A. Ashtekar, T. Pawlowski and P. Singh, Quantum Nature of the Big Bang: An Analytical and Numerical Investigation. I., *Phys. Rev. D* **73**, p. 124038 (2006).

21. A. Ashtekar, S. Fairhurst and J. L. Willis, Quantum gravity, shadow states, and quantum mechanics, *Class. Quant. Grav.* **20**, 1031 (2003).
22. A. Corichi, T. Vukasinac and J. A. Zapata, Polymer Quantum Mechanics and its Continuum Limit, *Phys. Rev. D* **76**, p. 044016 (2007).
23. H. A. Morales-Técotl, S. Rastgoo and J. C. Ruelas, Path integral polymer propagator of relativistic and nonrelativistic particles, *Phys. Rev. D* **95**, p. 065026 (2017).
24. H. A. Morales-Técotl, D. H. Orozco-Borunda and S. Rastgoo, Polymer quantization and the saddle point approximation of partition functions, *Phys. Rev. D* **92**, p. 104029 (2015).
25. E. Flores-González, H. A. Morales-Técotl and J. D. Reyes, Propagators in Polymer Quantum Mechanics, *Annals Phys.* **336**, 394 (2013).
26. A. Joe and P. Singh, Kantowski-Sachs spacetime in loop quantum cosmology: Bounds on expansion and shear scalars and the viability of quantization prescriptions, *Class. Quant. Grav.* **32**, p. 015009 (2015).
27. S. Saini and P. Singh, Geodesic completeness and the lack of strong singularities in effective loop quantum Kantowski–Sachs spacetime, *Class. Quant. Grav.* **33**, p. 245019 (2016).
28. C. Collins, Global structure of the Kantowski-Sachs cosmological models, *J. Math. Phys.* **18**, p. 2116 (1977).
29. D.-W. Chiou, Phenomenological dynamics of loop quantum cosmology in Kantowski-Sachs spacetime, *Phys. Rev. D* **78**, p. 044019 (2008).
30. K. Blanchette, S. Das, S. Hergott and S. Rastgoo, Black hole singularity resolution via the modified Raychaudhuri equation in loop quantum gravity, *Phys. Rev. D* **103**, p. 084038 (2021).
31. P. Bosso, O. Obregón, S. Rastgoo and W. Yupanqui, Deformed algebra and the effective dynamics of the interior of black holes, *Class. Quant. Grav.* **38**, p. 145006 (2021).
32. K. Blanchette, S. Das and S. Rastgoo, Effective GUP-modified Raychaudhuri equation and black hole singularity: Four models, *JHEP* **09**, p. 062 (2021).

Effect of loop quantization prescriptions on the physics of non-singular gravitational collapse

Kristina Giesel*, Bao-Fei Li† and Parampreet Singh†

*Institute for Quantum Gravity, Department of Physics, FAU Erlangen-Nürnberg, Staudtstr. 7, 91058 Erlangen, Germany

†Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

We review recent developments in the physical implications of two loop quantization strategies for the interior of a homogeneous dust cloud. The first is the loop quantization with holonomies and the triads while the second is with holonomies and the gauge covariant fluxes. Although both of the quantization schemes resolve the classical central singularity regardless of the initial conditions, they also lead to a distinct phenomenology. For the first loop quantization, we find that when the dust mass is larger than a threshold value, both black hole and white hole would form and their evolution is symmetric with respect to the bounce point, leading to black hole-white hole twins. In contrast, in the second quantization, the evolution of the outermost dust shell is asymmetric with respect to the bounce point, and as a result the black hole-white hole twins can never form. Even in the situation when both black hole and white hole can form, the mass of the latter is only $2/\pi$ of the mass of the former.

Keywords: Homogeneous dust collapse; loop quantum cosmology.

1. Introduction

The singularity problem in the gravitational collapse results from the breakdown of the classical theory of general relativity (GR) and can thus be resolved by quantum gravity. Understanding the role of quantum gravity in determining the final state of the collapsing stars can help provide insights on the fundamental questions related with cosmic censorship conjecture and black hole evaporation. Because of the quantization ambiguities in quantum theories, it is important to understand if the physical predictions are robust against various quantization prescriptions. In the following, we address this question for a collapsing dust cloud whose interior spacetime is described by the Lemaître-Tolman-Bondi(LTB) metric and study two quantization prescriptions in a loop quantum gravity (LQG) scenario.

To make the problem of a collapsing dust cloud more manageable, we further assume a homogeneous evolution of the dust cloud in the marginally bound case. We then make use of the techniques from LQG to quantize the interior of the dust cloud. The loop quantization of the spacetimes with spherical symmetry dates back to quantization of the interior of the Schwarzschild black hole.[1–6] Later it was extended to include both interior and exterior of the black hole[7–13] as well as the dynamical collapsing spacetime which is filled with a massless scalar or the dust.[14–18] All these studies focussed on the resolution of the central singularity by incorporating quantum geometry effects via holonomies and/or inverse triad

modification but not on comparaing the physics in different quantizations. Further, none of the forementioned models have been directly derived from LQG. Therefore, it is reasonable to ask whether the physical predictions in those models are robust when further modifications from LQG are included in the dynamics. To answer this question, we have studied two distinct quantization prescriptions for the interior of a collapsing dust cloud and compared their physical implications.[19] The first quantization employs holonomy and triads while in the second quantization, in addition to the holonomy corrections, we have also considered the contributions from the gauge-covariant fluxes[20] which have beenr ecently explored in the cosmological setting.[21–23] Our results have shown that although the resolution of the central singularity is robust against different quantization strategies, the resulting dynamics also exhibit distinctive features for each quantization ansatz. In particular, the first quantization results in black hole-white hole twins while the second quantization strategy generally leads to black hole-white hole asymmetry.

2. The classical dust shell model

The classical dust shell model describes the dynamics of the outermost shell of a collapsing dust cloud. It is based on the LTB model which is obtained from a spherically symmetric solution of Einstein's equations in GR with non-rotational dust as a matter source. The metric of LTB spacetime is given by

$$ds^2 = -d\tau^2 + \frac{(R')^2}{1+2f}dx^2 + R^2 d\Omega^2, \tag{1}$$

where R is the areal radius, $f(x)$ is the total energy of a unit mass at x and a prime denotes derivative with respect to the radial coordinate x. Depending on the sign of f, there are three distinct cases: the marginally bound case with $f = 0$, the bound case with $f < 0$ and the unbound case with $f > 0$. In the following, we focus on the marginally bound case. The corresponding Hamiltonian constraint for the LTB dust model is explicitly given in terms of both canonical ADM variables and the Ashtekar variables for a general inhomogeneous dust cloud in.[24] The classical dust shell model can then be obtained from the LTB model by a homogeneous reduction where the energy density of the dust cloud is assumed to depend only on time.[25] Thanks to the spherical symmetry and the homogeneity of the dust cloud interior, one can integrate the classical Hamiltonian constraint of the LTB model along the radial direction as well as the angular part and arrive at the Hamiltonian constraint for the outermost dust shell which can be expressed in terms of a canonical pair consisting of the radial components of the extrinsic curvature and the densitized triad.[19] It turns out that for the homogeneous dust collapse, the interior spacetime is isometric to the spatially-flat Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime. As a result, the classical Hamiltonian constraint of the outermost dust shell can be related with the one for the spatially-flat FLRW spacetime.[19]

For a complete description of thiS collapsing model, in addition to the dust interior, one still needs to specify the exterior spacetime which is glued to the dust interior at the boundary. We choose the exterior spacetime to be the generalized

Vaidya spacetime[26] so that it allows a non-vanishing effective dust pressure after quantum geometrical effects are taken into account.[14] The matching conditions are worked out under the requirements that the first and the second fundamental forms should be continuous across the boundary. In this way, it turns out that in the marginally bound case the Vaidya mass which is the mass of the dust cloud as seen by an outside observer differs from the mass function of the dust cloud by a factor of $1/2G$ with G denoting the Newton's constant.

3. Loop quantization of the dust shell model with holonomy and triad variables

The loop quantization of the dust shell model is based on the classical Hamiltonian constraint of the outermost dust shell which is expressed in terms of the Ashtekar variables.[19] Due to the quantization ambiguities, there are different quantization prescriptions which can result in distinct effective dynamics. Since the interior of a homogeneous dust cloud is isometric to a spatially-flat FLRW spacetime, one can apply techniques developed in loop quantum cosmology (LQC)[27] in a straightforward way. After quantizing the outermost dust shell by employing holonomies and the triads with the $\bar{\mu}$ scheme,[28] we find that the classical central singularity is resolved and replaced with a bounce which takes place at a fixed maximum energy density in the Planck regime. Meanwhile, the collapse of the dust cloud is succeeded by a re-expansion after the bounce point. The formation of the trapped surfaces during the contraction and the expansion of the dust cloud depends solely on the initial dust mass. There also exists a threshold dust mass M^* below which no black hole or white hole can form during the evolution of the dust cloud. On the other hand, when the initial dust mass is larger than M^*, a dynamical black hole can form during the collapse of the dust cloud and correspondingly a white hole can form during the re-expansion of the dust cloud. The evolution of the black hole and the white hole is always symmetric with respect to the bounce point and in particular the mass of the black hole is the same as the mass of the white hole. Moreover, right at the bounce, the trapped surfaces vanishes and the Vaidya mass becomes zero which implies that an asymptotic flat Minkowski spacetime emerges at the bounce point due to the quantum repulsive force. This repulsive force is also reflected by a negative effective pressure near the bounce point when the energy density is in the Planck regime.

4. Loop quantization of the dust shell model with holonomy and gauge covariant flux variables

Apart from the holonomy and triad quantization, we also study the physical consequences of using holonomy and the gauge covariant fluxes[20] which is motivated from the need to go beyond symmetry reduced triads in to obtain an effective Hamiltonian with loop quantum modifications from LQG using suitable coherent states. Its physical implications have been explored for the spatially-flat FLRW model recently.[21–23] This quantization strategy leads to distinct dynamical evolution of the outermost dust shell as compared with the one discussed in the last section.

In particular, although the classical central singularity is resolved and replaced with a quantum bounce which also takes place in the Planck regime, the formation/evolution of the black hole and the white hole is no longer symmetric with respect to the bounce point. Due to quantum gravity effects there exist two characteristic dust masses M_1 and M_2 ($M_1 < M_2$). When the initial dust mass is less than M_1, no trapped surfaces can form during the entire evolution of the dust cloud. On the other hand, when the initial dust mass lies between M_1 and M_2, only a black hole can form during the collapse of the dust cloud. Finally, when the initial dust mass is larger than M_2, both black hole and the white hole can form. In the last case, we find the evolution of the black hole and the white hole is not symmetric with respect to the bounce point and in particular the effective mass of the white hole is only $2/\pi$ of the one for the black hole and the black hole always outlives the white hole in the proper time. This asymmetry is due to the difference in the classical limits of the pre- and post-bounce regimes of the effective dynamics. As a result, black hole-white hole twins do not exist in this quantization strategy.

5. Summary

The gravitational collapse of a homogeneous dust cloud provides a platform to test different loop quantizations and investigate their resulting physical implications. We find even in this simple setting, quantization ambiguities can lead to very distinct phenomenological effects. In particular, we have studied two quantization strategies in this context, the one employing holonomies and triads and the other using holonomies and gauge covariant fluxes. Although in both cases, the central singularity is resolved and replaced with a quantum bounce, there are qualitative differences between two quantization prescriptions. In the former, when the dust mass is larger than the threshold mass, there are black hole-white hole twins. While in the latter, a black hole and white hole twin system is not possible and the the mass of the white hole is only $2/\pi$ of the mass of the black hole because of quantum gravity effects. Further, there can be situations in which only black hole forms during the collapse of the dust cloud.

We wish to thank Jahnvi Verma and Anzhong Wang for discussions. This work is supported by the DFG-NSF grants PHY-1912274 and 425333893, and PHY-1454832 and PHY-2110207.

References

1. A. Ashtekar and M. Bojowald, *Quantum geometry and the Schwarzschild singularity*, Class. Quant. Grav. **23**, 391 (2005).
2. C. G. Bohmer and K. Vandersloot, *Loop quantum dynamics of the Schwarzschild interior*, Phys. Rev. **76**, 104030 (2007).
3. A. Corichi and P. Singh, *Loop quantization of the Schwarzschild interior revisited*, Class. Quant. Grav. **33**, 055006 (2016).
4. J. Olmedo, S. Saini, and P. Singh, *From black holes to white holes: A quantum gravitational, symmetric bounce*, Class. Quant. Grav. **34**, 225011 (2017).

5. A. Ashtekar, J. Olmedo, and P. Singh, *Quantum transfiguration of Kruskal black holes*, Phys. Rev. Lett. **121**, 241301 (2018).
6. A. Ashtekar, J. Olmedo, and P. Singh, *Quantum extension of the Kruskal spacetime*, Phys. Rev. **D98**, 126003 (2018).
7. M. Bojowald and R. Swiderski, *Spherically symmetric quantum geometry: Hamiltonian constraint*, Class. Quant. Grav. **23**, 2129-2154 (2006).
8. R. Gambini and J. Pullin, *Black holes in Loop Quantum Gravity: The complete spacetime*, Phys. Rev. Lett. **101**, 161301 (2008).
9. D.-W. Chiou, W.-T. Ni, and A. Tang, *Loop quantization of spherically symmetric midisuperspaces and loop quantum geometry of the maximally extended Schwarzschild spacetime*, 2013, arXiv:1212.1265.
10. R. Gambini and J. Pullin, *Loop quantization of the Schwarzschild black hole*, Phys. Rev. Lett. **110**, 211301 (2013).
11. R. Gambini, J. Olmedo, and J. Pullin,*Quantum black holes in loop quantum gravity*, Class. Quant. Grav. **31**, 095009 (2014).
12. R. Gambini, J. Olmedo, and J. Pullin, *Spherically symmetric loop quantum gravity: Analysis of improved dynamics*, Class. Quant. Grav. **37**, 205012 (2020).
13. J. G. Kelly, R. Santacruz, and E. Wilson-Ewing, *Effective loop quantum gravity framework for vacuum spherically symmetric spacetimes*, Phys. Rev. **D102**, 106024 (2020).
14. M. Bojowald, R. Goswami, R. Maartens, and P. Singh, *Black hole mass threshold from nonsingular quantum gravitational collapse*, Phys. Rev. Lett. **95**, 091302 (2005).
15. R. Goswami, P. S. Joshi, and P. Singh, *Quantum evaporation of a naked singularity*, Phys. Rev. Lett. **96**, 031302 (2006).
16. R. Gambini, J. Pullin, and S. Rastgoo, *Quantum scalar field in quantum gravity: The vacuum in the spherically symmetric case*, Class. Quant. Grav. **26**, 215011 (2009).
17. C. Bambi, D. Malafarina, and L. Modesto, *Non-singular quantum-inspired gravitational collapse*, Phys. Rev. **D88**, 044009 (2013).
18. F. Benitez, R. Gambini, L. Lehner, S. Liebling, and J. Pullin, *Critical collapse of a scalar field in semiclassical Loop Quantum Gravity*, Phys. Rev. Lett. **124**, 071301 (2020).
19. K. Giesel, B.-F. Li, and P. Singh, *Non-singular quantum gravitational dynamics of an LTB dust shell model: The role of quantization prescriptions*, arXiv:2107.05797.
20. T. Thiemann, *Quantum spin dynamics (QSD): VII. Symplectic structures and continuum lattice formulations of gauge field theories*, Class. Quant. Grav. **18**, 3293-3338 (2001).
21. K. Liegener and P. Singh, *New loop quantum cosmology modifications from gauge-covariant fluxes*, Phys. Rev. **D100**, 124048 (2019).
22. K. Liegener and P. Singh, *Some physical implications of regularization ambiguities in SU(2) gauge-invariant loop quantum cosmology*, Phys. Rev. **D100**, 124049 (2019).
23. K. Liegener and P. Singh, *Gauge-invariant bounce from loop quantum gravity*, Class. Quant. Grav. **37**, 085015 (2020).
24. M. Bojowald, T. Harada, and R. Tibrewala, *Lemaitre-Tolman-Bondi collapse from the perspective of loop quantum gravity*, Phys. Rev. **D78**, 064057 (2008).
25. C. Kiefer and T. Schmitz, *Singularity avoidance for collapsing quantum dust in the Lemaître-Tolman-Bondi model*, Phys. Rev. **D99**, 126010 (2019).
26. C. Barrabès and W. Israel, *Thin shells in general relativity and cosmology: The lightlike limit*, Phys. Rev. **D43**, 1129 (1991).
27. A. Ashtekar and P. Singh, *Loop Quantum Cosmology: A Status Report*, Class. Quant. Grav. **28**, 213001 (2011).
28. A. Ashtekar, T. Pawlowski, and P. Singh, *Quantum nature of the Big Bang: Improved dynamics*, Phys. Rev. **D74**, 084003 (2006).

Summary of the parallel session QG3

Jorge Pullin and Parampreet Singh

Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

We summarize the main results of 19 talks presented at the QG3 session (loop quantum gravity: cosmology and black holes) of the 16th Marcel Grossmann Meeting held online from July 5^{th}-10^{th}, 2021.

Keywords: Loop quantum gravity; cosmology; black holes

The QG3 session was organized in two parts with a total of 19 talks. The first part of the session was held on July 5th (9 talks) focussing on loop quantum gravity effects in black hole spacetimes, and the second part was on July 8th (10 talks) focussing on cosmological implications. A summary of talks in a chronological order is as follows.

We-Cong Gan presented his joint work with Nilton O. Santos, Fu-Wen Shu, and Anzhong Wang on spherically symmetric polymer black holes.[1] Their work is based on a model by Bodendorfer, Mele and Münch and depend on three independent parameters. An exploration of the parameter space was made showing a wealth of physical possibilities including singularity resolution, wormholes and space-times with singularities.

Harkirat Singh Sahota discussed joint work with Kinjalk Lochan[2] about the quantization of the Lemaître–Tolman–Bondi model as a minisuperspace. It leads to a bounce eliminating the singularity and the talk explored how it manifests itself in the infrared regime to an observer looking at the collapse.

Aurélien Barrau discussed joint work with Léonard Ferdinand, Killian Martineau and Cyril Renevey.[3] There have been several proposals discussed in the literature in which after their evaporation black holes tunnel into metastable white holes. They have been considered a dark matter candidate. The talk showed that such models are severely constrained and suggested some possible detection paths.

Parampreet Singh presented his work with Kristina Giesel and Bao-Fei Li on a non-singular collapse of a dust shell model in Lemaître–Tolman–Bondi spacetime.[4] Two quantization prescriptions were studied. One using holonomies and triads, and another using holonomies and gauge-covariant fluxes. While both the prescriptions result in a bounce of the dust shell they lead to different mass gaps at which a trapped surface may form. Unlike in a quantization based on holonomies and triads,

in presence of gauge-covariant fluxes, a formation of white hole in the expanding phase when a black hole forms in the contracting phase is not guaranteed.

Asier Alonso-Bardaji presented work with David Brizuela where they consider holonomy and inverse-triad corrections in spherically symmetric models of loop quantum gravity.[5] It constructs systematically anomaly-free constraints with such corrections. It concludes that holonomy corrections are consistent in the presence of matter fields. Contrary to earlier claims, authors demonstrate a model which provides a family of deformed Hamiltonian constraints with scale-dependent holonomy corrections in the presence of a scalar matter field.

Cong Zhang talked about joint work with Yongge Ma, Shupeng Song and Xiangdong Zhang on the loop quantum gravity quantization of the interior of the Schwarzschild black hole.[6] They study a Dirac observable corresponding to the ADM mass and note that zero is not in its spectrum, suggesting that a remnant is left after the evaporation of the black hole.

Alejandro García Quasimodo talked about work with Guillermo Mena Marugán about some modifications to the Ashtekar–Olmedo–Singh black hole model, in particular the use of polymerization parameters that are Dirac observables.[7] This may modify the asymptotic behavior of the model.

Saeed Rastgoo presented work with Keagan Blanchette, Saurya Das and Samantha Hergott on loop quantum gravity corrections to the Raychaudhuri equation.[8] These corrections imply defocusing of geodesics providing insights on the eliminiation of singularities in this context.

Patrick Fraser analyzed how classical are the Gaussian states in loop quantum cosmology.[9] He showed that contrary to common intuition, such states do not saturate their uncertainty relations and there exist Gaussian states for which the fluctuations are arbitrarily large. However, the usual volume regularization procedure allows to suppress those fluctuations as much as one wishes.

V. Sreenath talked about alleviating tensions in the cosmic microwave background using Planck scale physics in a joint work with Abhay Ashtekar, Brajesh Gupt and Donghui Jeong.[10] The idea is that there are several anomalies of low significance in the agreement of the predictions of the usual Λ-CDM model for the cosmic microwave background. Taken as a whole, however, they imply a tension between theory and data. The talk showed that considering loop quantum cosmology leads to a primordial power spectrum that is scale dependent at large scales. This can potentially alleviate the anomalies.

Sahil Saini discussed a matter-Ekpyrotic bounce model in loop quantum cosmology in a joint work with Bao-Fei Li and Parampreet Singh.[11] Unlike previous attempts in this direction in loop quantum cosmology, Ekpyrotic phase was sourced by a potential motivated from Ekpyrotic scenarios. It was shown that in contrast to previous results where a fixed ultra-stiff equation of state was assumed, the magnitude of power spectrum changes during evolution resulting from an Ekpyrotic potential. It was found that the bouncing regime only leaves imprints on the modes

outside the scale-invariant regime. Further refinements in this direction are necessary since the spectral index shows inconsistency with the observational data.

Kristina Giesel presented a reduced phase space quantization for loop quantum cosmology in presence of inflationary potentials. This was a joint work with Bao-Fei Li and Parampreet Singh.[12] Unlike the case of a massless scalar field which has been addressed rigorously in loop quantum cosmology, singularity resolution at the quantum level in the presence of a potential has remained an open problem because of several conceptual and technical challenges. The idea is to introduce an additional clock degree of freedom to measure time evolution. It was shown that singularity resolution occurs in the presence of potentials at the level of quantum difference equation. Some properties of non-singular solutions were discussed using effective dynamics.

Guillermo Mena Marugán talked about his work with Beatriz Elizaga Navascués and Rafael Jiménez-Llamas on the way potential effects primordial perturbations in kinetic dominated regimes in classical and loop quantum cosmology.[13] Such a situation is of particular interest in loop quantum cosmology where so far kinetic dominated bounce has been studied. Authors found that the choice of Bunch-Davies vacuum may no longer be preferred in presence of such corrections. Further, there are changes in the effective mass in the Mukhanov-Sasaki equation.

Anzhong Wang discussed his joint work with Bao-Fei Li and Parampreet Singh on various subtleties in imposing initial conditions for cosmological perturbations in loop quantum cosmology and its modified versions (mLQC-I and mLQC-II).[14] The talk highlighted differences in choices which are forced by the assumptions in the dressed metric and hybrid approaches. Issues related to when to impose initial conditions – whether at bounce or in the contracting phase, and constraints on possible choices were discussed. It was shown that a consistent choice of the initial conditions depends not only on the choice of a potential but also on the two approaches used in loop quantum cosmology.

Javier Olmedo talked about effects on primordial perturbations in a bouncing model in loop quantum cosmology for a Bianchi-I spacetime in a joint work with Ivan Agullo and V. Sreenath.[15] They considered an evolution where anisotropies are non-vanishing at a bounce dominated by matter energy density. Though anisotropies were diminished in the pre-inflationary branch, effects were found in the scalar and tensor power spectrum. Modification to the angular correlation functions and the way EB and TB correlations may arise was discussed.

Bao-Fei Li discussed features of primordial power spectrum in modified versions of loop quantum cosmology in a joint work with Javier Olmedo, Parampreet Singh and Anzhong Wang.[16] They discussed the way dressed metric and hybrid approaches to perturbations result in different signatures in the primordial power spectrum, especially for mLQC-I model. Differences with loop quantum cosmology in the infra-red and intermediate regimes were also found which show that different loop quantum prescriptions potentially leave distinct signatures in CMB.

Dimitrios Kranas talked about his joint work with Ivan Agullo and V. Sreenath on understanding the origin of various anomalies in the CMB using a bouncing model.[17] They find that that non-Gaussian correlations between CMB modes and super-horizon wavelengths in the power spectrum can potentially explain a power suppression, a dipolar asymmetry, and existence of odd-parity correlations in the power spectrum.

Adriá Delhom i Latorre discussed hos joint work with Gonzalo Olmo and Parampreet Singh.[18] They investigate the existence of a covariant action in $f(R)$ Palatini theories which can reproduce effective equations for modified versions of loop quantum cosmology. It was found that the Lagrangians which result in effective dynamics of LQC, mLQC-I and mLQC-II are part of a three-parameter family of f(R) theories, where two parameters are fixed by initial conditions at the bounce.

Hongguang Liu talked about his work with Muxin Han on a path integral approach based on reduced phase space loop quantum gravity to derive effective dynamical equations in which the graph changes dynamically in the physical time evolution.[19] They were able to derive the so-called improved dynamics from this method found earlier in loop quantum cosmology by Ashtekar, Pawlowski and Singh which results in a singularity resolution. The method also generalizes the path integral formulation used in full LQG to take into account an additional real scalar field.

Acknowledgments

This work was supported in part by grant NSF-PHY-1454832, NSF-PHY-1903799, DFG-NSF-PHY-1912274, NSF-PHY-2110207 and funds of CCT-LSU and the Horace Hearne Institute for Theoretical Physics.

References

1. W. C. Gan, N. O. Santos, F. W. Shu and A. Wang, "Properties of the spherically symmetric polymer black holes," Phys. Rev. D **102**, 124030 (2020).
2. H. S. Sahota, K. Lochan, *To appear.*
3. A. Barrau, L. Ferdinand, K. Martineau and C. Renevey, "Closer look at white hole remnants," Phys. Rev. D **103**, no. 4, 043532 (2021).
4. K. Giesel, B. F. Li and P. Singh, "Non-singular quantum gravitational dynamics of an LTB dust shell model: The role of quantization prescriptions," [arXiv:2107.05797 [gr-qc]].
5. A. Alonso-Bardaji and D. Brizuela, "Anomaly-free deformations of spherical general relativity coupled to matter," [arXiv:2106.07595 [gr-qc]].
6. C. Zhang, Y. Ma, S. Song and X. Zhang, "Loop quantum deparametrized Schwarzschild interior and discrete black hole mass," [arXiv:2107.10579 [gr-qc]].
7. A. García-Quismondo and G. A. Mena Marugán, "Exploring alternatives to the Hamiltonian calculation of the Ashtekar-Olmedo-Singh black hole solution," Front. Astron. Space Sci. **8**, 701723 (2021).

8. K. Blanchette, S. Das, S. Hergott and S. Rastgoo, "Black hole singularity resolution via the modified Raychaudhuri equation in loop quantum gravity," Phys. Rev. D **103**, no. 8, 084038 (2021).
9. P. Fraser, "Taming Fluctuations for Gaussian States in Loop Quantum Cosmology," Phys. Rev. D **103**, no. 8, 086014 (2021).
10. A. Ashtekar, B. Gupt, D. Jeong and V. Sreenath, "Alleviating the Tension in the Cosmic Microwave Background using Planck-Scale Physics," Phys. Rev. Lett. **125**, no. 5, 051302 (2020); A. Ashtekar, B. Gupt and V. Sreenath, "Cosmic Tango Between the Very Small and the Very Large: Addressing CMB Anomalies Through Loop Quantum Cosmology," Front. Astron. Space Sci. **8**, 76 (2021).
11. B. F. Li, S. Saini and P. Singh, "Primordial power spectrum from a matter-Ekpyrotic bounce scenario in loop quantum cosmology," Phys. Rev. D **103**, no. 6, 066020 (2021).
12. K. Giesel, B. F. Li and P. Singh, "Towards a reduced phase space quantization in loop quantum cosmology with an inflationary potential," Phys. Rev. D **102**, no. 12, 126024 (2020).
13. B. Elizaga Navascués, R. Jiménez-Llamas and G. A. Mena Marugán, "Primordial perturbations in kinetically dominated regimes of general relativity and hybrid quantum cosmology," [arXiv:2106.05628 [gr-qc]].
14. B. F. Li, P. Singh and A. Wang, "Phenomenological implications of modified loop cosmologies: An overview," Front. Astron. Space Sci. **8**, 701417 (2021).
15. I. Agullo, J. Olmedo and V. Sreenath, "Predictions for the Cosmic Microwave Background from an Anisotropic Quantum Bounce," Phys. Rev. Lett. **124**, no. 25, 251301 (2020).
16. B. F. Li, P. Singh and A. Wang, "Primordial power spectrum from the dressed metric approach in loop cosmologies," Phys. Rev. D **101**, no. 8, 086004 (2020); B. F. Li, J. Olmedo, P. Singh and A. Wang, "Primordial scalar power spectrum from the hybrid approach in loop cosmologies," Phys. Rev. D **102**, 126025 (2020).
17. I. Agullo, D. Kranas and V. Sreenath, "Anomalies in the CMB from a cosmic bounce," Gen. Rel. Grav. **53**, no. 2, 17 (2021).
18. A. D. i Latorre, G. J. Olmo, P. Singh, *To appear*.
19. M. Han and H. Liu, "Loop quantum gravity on dynamical lattice and improved cosmological effective dynamics with inflaton," Phys. Rev. D **104**, no. 2, 024011 (2021).

Reliable equations of state of viscous strong and electroweak matter

A. Tawfik

Egyptian Center for Theoretical Physics (ECTP), Cairo, Egypt
World Laboratory for Cosmology And Particle Physics (WLCAPP), Cairo, Egypt
E-mail: a.tawfik@cern.ch

For the first time, a reliable estimation for the equations of state (EoS), bulk viscosity, and relaxation time, at temperatures ranging from a few MeV up to TeV or energy density up to 10^{16} GeV/fm^3. This genuine study covers both strong and electroweak epochs of the early Universe. Non–perturbation (up, down, strange, charm, and bottom quark flavor) and perturbative calculations (up, down, strange, charm, bottom, and top quark flavors), are phenomenologically combined, at vanishing baryon–chemical potential. In these results, calculations from Polyakov linear–sigma model (PLSM) of the vacuum and thermal condensations of the gluons and the quarks (up, down, strange, and charm flavors) are also integrated. Furthermore, additional degrees of freedom (photons, neutrinos, charged leptons, electroweak particles, and scalar Higgs boson) are found significant along the entire range of temperatures. As never done before, the present study brings the standard model of elementary particles closer to the standard model for cosmology.

Keywords: Strong and electroweak epochs, Perturbative and non–perturbative calculations, Equations of state, Bulk viscosity, Relaxation time

1. Introduction

Understanding the strong and electroweak epochs of the early Universe, at least, through determining their equations of state (EoS) is impactful for various cosmological studies, such as the nucleosynthesis and the cosmological large–scale structure. Until present date, cosmology is dictated by the standard model of cosmology (SMC). SMC assumes that the cosmic background is isotropically and homogeneously characterized by an ideal fluid. With this regard, we recall that a first proposal that the viscous coefficients are also connected with particle physics was presented Ref.[1,2]. It was also assumed that the influence of viscous coefficients becomes significant first at the end of the lepton era,[3] i.e., during neutrino decoupling epoch or temperature $T \simeq 10^{10}$ K ($\simeq 1$ MeV) which takes place not long after the end of the QCD phase transition from colored quark–gluon plasma (QGP) to colorless hadrons. The QCD phase transition[4–7] takes place at $T \simeq 160$ MeV. Recent studies concluded that the viscous coefficients likely impact early epoches of the Universe.[8]

Taking advantage of recent developments in ultra–relativistic nuclear experiments and non–perturbative lattice QCD simulations, and perturbative calculations, various thermodynamic quantities including pressure p, energy density ρ, bulk viscosity ζ, and relaxation time τ are calculated up to the TeV-scale and a reliable evolution of the early Universe could be achieved. The present study offers

an access to deep epochs of the early Universe with T up to TeV-scale and ρ up to 10^{16} GeV/fm^3.

For Friedmann–Lemaitre–Robertson–Walker (FLRW) metric

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2 \theta d\phi^2 \right) \right], \tag{1}$$

where $a(t)$ is the dimensionless scale factor, $k = \{-1, 0, +1\}$ represents elliptical, flat (Euclidean), and hyperbolic cosmic space, or negative, flat, and positive curvature, respectively. The theory of general relativity (GR) inters the play, when temporal evolution of the line element $s(t)$ can be determined. To this end, GR has to be combined with the matter–energy content of the cosmic geometry,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda_{\mu\nu} = \frac{8\pi}{3} T_{\mu\nu}, \tag{2}$$

where μ, ν run over 0, 1, 2, and 3.

Then, the Einstein field equations in natural units read

$$H(t)^2 = \frac{8\pi}{3} \rho(t) - \frac{k}{a(t)^2} + \frac{\Lambda}{3}, \tag{3}$$

$$\dot{H}(t) + H(t)^2 = -\frac{4\pi}{3} \left[\rho(t) + 3p_{\mathtt{eff}(t)} \right] + \frac{\Lambda}{3}, \tag{4}$$

where $H(t) = \dot{a}(t)/a(t)$ is the Hubble parameter. From (3) and (4), the time evolution of the Hubble parameter can be deduced

$$\dot{H}(t) = -4\pi \left[\rho(t) + p_{\mathtt{eff}}(t) \right] + \frac{k}{a(t)^2}. \tag{5}$$

An EoS relating p with ρ is needed to have a closed system of equations.

The present script is organized as follows. The most reliable non-perturbative and perturbative calculations are reviewed in section 2. Our results on EoS, bulk viscosity and relaxation time for almost the entire SM dof are presented in section 3. The conclusions are outlined in section 4.

2. Most reliable non-perturbative and perturbative calculations

For relaible EoS, various general considerations have been assumed in the recent non-perturbative lattice QCD simulations.[9] The first one is the quark masses. For a dynamical dependence of the quark masse m_q and the lattice spacing a on the guage coupling β, four flavor staggered action with 4 levels of stout smearing has been utilized, where u, d, s, and c quarks are realized as 2+1+1 and 3+1,[9] i.e., except for strange and charm quark masses, $m_{ud} = R \times m_s^{st}(\beta)$, with $m_s^{st}(\beta)$, $1/R = 27.63$ and β is the guage coupling. The c quark mass is given as a function of the gauge coupling; $m_c = C \times m_s^{st}(\beta)$, with $C = 11.85$. Although the degenerate masses of the light quarks, a small isospin asymmetry could also be included, analytically.[9]

The second general consideration is the temperature T, which can primarily be determined as a function of the temporal lattice dimension; $T = (aN_\tau)^{-1}$. Alternatively, varying the gauge coupling β leads to changing T, as well, even if the spacial

and temporal lattice dimensions might not be varied. The gauge coupling can not only allow for varying T, but also it measures the pseudoscalar pion mass m_π and the Wilson–flow based scale ω_0, where $\omega_0 = 0.153 \pm 0.001$ fm and $m_\pi = 712 \pm 5$ MeV. At $T = 0$, ω_0 gives the inverse flow time.[10]

The third general consideration is the lattice QCD computational procedure. After applying a Wilson–flow equation, the clover definition of the topological charge was made in $2 + 1 + 1$ and $3 + 1$ ensembles. To make the computational process more economic, an adaptive step size integration scheme was utilized. The time flow $(8T^2)^{-1}$ was introduced to estimate the finite T of both ensembles, where a variation in the time flow is also allowed. This procedure has greatly contributed to reducing the systematic errors. To control their simulations, it should be checked whether these configurations lead to saturated susceptibility, at large flow times.[9] To determine the topological susceptibility, topological charge was utilized with and without rounding.

The fourth general consideration is the acceptance accuracy. As already outlined for the $2+1+1$ simulations, same configurations have been used also for $3+1$ simulations. Here, the mass ratio of the charm quark and the degenerate lighter quarks (u, d, and s) was taken to be 11.85. As mentioned, for the masses of up– and down–quarks, the physical values are used, while the mass of the strange quark is a function of the guage coupling, $m_s^{\mathtt{st}}(\beta)$. At $T = 0$, simulations were done on 64×32^3 lattice with seven values of the lattice spacing descendingly ranging from 0.15 to 0.06 fm. But at finite T, the same parameters as in $2+1+1$ case were used. The topological charge is measured for every Hybrid Monte–Carlo trajectory. The configurations leading to a topology change are rejected. In other words, configurations are generated, at fixed topology. Quantitatively, this leads to an acceptance probability of about 40% for the coarsest lattice, but higher probabilities for the finest ones.

In the following section we review how various thermodynamic quantities have been determined.

2.1. *Lattice QCD Equation of State in Non-Perturbative Regime*

The inclusion of up–, down–, strange–, and charm–quark in non–perturbative lattice QCD simulations[11,12] assumes that the masses of some quarks are very heavy (partially quenched). Other lattice QCD simulations assume that the four quarks are non–degenerate with unphysical masses.[13,14] The ultimate goal is to carry out non–perturbative simulations with dynamical quarks and physical masses.[15] To this end, $2+1+1$ simulations with staggered action, and 4 levels of stout smearing have been carried out in Ref.[16].

One restriction in the lattice QCD simulations is the universal assumption of thermal equilibrium. This might not fully true, especially because the hadron and parton matter is undergoing phase transition, whether prompt of slows crossover, at critical temperatures, i.e., non–equilibrium due to changes in the underlying

dynamics, symmetry, and degrees of freedom. Another restriction is that the temperature dependence of pressure p, energy density ρ, and entropy s can be deduced from the trace anomaly, at vanishing chemical potential,

$$\frac{I(T)}{T^4} = \frac{\rho - 3p}{T^4}, \tag{6}$$

$$\frac{\rho + p}{T^4} = \frac{s}{T^3}. \tag{7}$$

A third restriction is the common practice to get rid of the temperature independent divergence of the trace anomaly so that its physical value can be evaluated, accurately. A vanishing–T ensemble is subtracted from each finite–T ensemble. As this method doesn't work well at high temperatures, where short lattice spacing and increasing autocorrelation times and computational costs are likely, a renormalization ensemble is generated for each finite–T ensemble, at exactly half of its temperature.[9] The physical trace anomaly in a wide range of temperatures including low T, can be obtained from the subtraction of each half–T from its finite–T ensemble, $[I(T) - I(T/2)]/T^4$. The resulting ρ and s can straightforwardly be estimated from Eqs. (6) and (7).

Up to four quark flavors were included in the non–perturbative lattice QCD simulations. This was achieved by several steps of tree–level corrections.

2.2. *Inclusion of c–Quark in Non–Perturbative Lattice QCD Simulations*

It was concluded[17] that the free energy calculated in next–to–next–to–leading order (NNLO) Hard Thermal Loop (HTL) perturbation theory is in good agreement with $2 + 1$ and $2 + 1 + 1$ non–perturbative lattice QCD simulations, where c–quark is taken massless in the perturbative calculations but assigned the physical mass in the non–perturbative simulations. The mass of c–quark was perturbatively[18] and non–perturbatively[9] estimated. In the perturbative calculations, the effect of heavy quarks is determined to a lower leading order, and accordingly, it was concluded that this refers to $(3 + 1)$ quark flavors pressure normalized to (3) quark flavors pressure p. On the other hand, when comparing p with and without c–quark in both non–perturbative and perturbative calculations, an excellent agreement ($< 3\%$) was obtained.[9] The tree–level correction due to c–quark reads

$$\frac{p^{(2+1+1)}(T)}{p^{(2+1)}(T)} = \frac{SB^{(3)}(T) + F_Q(m_c/T)}{SB^{(3)}(T)}, \tag{8}$$

where SB stand for Stefan–Boltzmann approximation and $F_Q(m_c/T)$ is the free energy density of a free quark with mass $m_c = 1.29$ GeV.

The following section elaborates how the b–quark flavor is included non–perturbative lattice QCD simulations. This was achieved by several steps of tree–level corrections.

2.3. *Inclusion of b–Quark in Non–Perturbative Lattice QCD Simulations*

The success with the inclusion of the c–quark encouraged a recent attempt with the bottom quark, especially that the $2+1+1$ non–perturbative simulations up to $T \lesssim 1$ GeV control the accuracy of the proposed procedure and the perturbative contributions likely dominate, at $T \gtrsim 500$ MeV).[19,20] Recent perturbative calculations were performed up to $\mathcal{O}(g^6 \log g)$.[9] Thus, the inclusion of b–quark and a continuation to higher T look straightforward. A tree–level correction for the b–quark similar to that for the c–quark, Eq. (8), was suggested.

It was concluded that tree–level correction for b–quark works well. When comparing the ratio of the massless 2+1+1 to 2+1+1+1 perturbative pressure with the pressure ratio in the SB limit, an excellently agreement ($< 0.3\%$) is obtained. With a phenomenological approach similar to Eq. (8), reliable non–perturbative lattice QCD simulations become feasible, at 500 MeV $< T <$ 10 GeV. This T–range apparently covers various epochs in the early Universe, where different phase transitions and accordingly different dynamics and degrees of freedom become dominant.

In the following section, we review features of the perturbative calculations up to TeV temperatures.

2.4. *Perturbative Calculations up to TeV Temperature–Scales*

In the previous sections, we have discussed on the various restrictions of the non–perturbative lattice QCD simulations and their possible extensions to a large number of quark flavors and to higher temperatures. The perturbative calculations, on the other hand, allow to cover much higher temperatures and to include more quark flavors. By combining recent perturbation calculations up to a largest leading order with the non–perturbative lattice QCD simulations, relaible EoS for strong and electroweak matter can be deduced.[21,22] It is conjectured that this covers temperatures up to ~ 200 TeV.

2.4.1. *QCD Domain*

In this domain, $0.2 \lesssim T \lesssim 1$ GeV, the gluons and the lightest four quark flavors and the gluons are partonic dof of the strongly interacting matter.[21] The perturbative corrections to the ideal (masslss, noninteracting) Stefan-Boltzmana (SB) EoS can be determined up to different orders of strong coupling constant $\mathcal{O}(g)$. Although, the perturbative contributions up to $\mathcal{O}(g^6 \log g)$ are well known,[19,23,24] where g is expressed as function of N_c the colors dof, N_f the massless quark flavors dof, and μ_f the quark chemical potential, only $\mathcal{O}(g^2)$ terms, at $T = 0$ but $\mu_f \neq 0$, have been precisely, analyzed.

Besides the contributions of gluons up to $\mathcal{O}(g^6 \log g)$ and that of NLO $\mathcal{O}(g^2)$, it was found that these are similar to that of LO $\mathcal{O}(g^0)$, where no $\mathcal{O}(g^6 \log g)$ calculations for finite quark masses are available so far.[18] An alternative procedure was

suggested. This starts with $N_f = 0$ and $N_c = 3$ corresponding to very heavy quark flavors. Then, it calculates the change in the pressure by lowering the quark masses down to their physical values[a]. This procedure suggests that the perturbative calculations are based on the grand–canonical pressure (or free energy).[9] As discussed earlier, the non–perturbative lattice QCD simulations starts with the trace anomaly (also known as the interaction measure) $I(T, \mu_f, \cdots)$, Eq. (6), from which the various thermodynamic quantities can be determined. Based on this renormalization procedure, the ultraviolet divergences are likely removed.

In order to extend the perturbative calculations to the electroweak domain, the so–called *hard modes* should first be removed through either integration with respect to momenta or summation over the Matsubara frequencies, $2\pi T$, gT, g^2T, $\cdots$. Second, the effective mass parameters and the gauge couplings have to be specified and normalized in $\overline{\mathtt{MS}}$–scheme. Third, the connection factors should then be estimated, at changing T and fixed $\Lambda_{\overline{\mathtt{MS}}}$. These can be achieved, when multiplying the non–perturbative lattice results for $N_f = 0$, $N_c = 3$, finite quark masses by the corresponding ones obtained in the SB limit, Eq. (8). The connection factors facilitate the inclusion of heavier quarks, at temperatures greater that strong QCD scale parameter $\Lambda_{QCD} \sim 200$ MeV.

2.4.2. *Electroweak Domain*

As discussed, the perturbative calculations are initiated from the free energy (or thermodynamic pressure), the Higgs potential parameters $v^2(\bar{\mu})$ and $\lambda(\bar{\mu})$ are given as functions of the normalization scale $\bar{\mu}$. Assuming that the bottom and top quarks weakly interact, the free energy, at finite T and $\boldsymbol{\mu}$, reads[18]

$$f(v, T, \boldsymbol{\mu}) = -\frac{1}{2}\nu^2(\bar{\mu})v^2 + \frac{1}{4}\lambda(\bar{\mu})v^4 + \sum_{i}^{n_q+n_g} \pm J_i(m_i(v), T, \mu_i), \tag{9}$$

where n_q (n_g) are the number of quarks (gluons), v is the Higgs expectation value, $m_i(v)$ is the tree–level mass of i–th particle, $\pm$ stand for bosons and fermions, respectively, and $J_i(m_i(v), T, \mu_i)$ counts for the contributions of the physical dof (scalars, vectors, and fermions). The normalization scale $\bar{\mu}$ could be fixed to 100 GeV[b]. Nevertheless, the electroweak free energy, Eq. (9), can be determined, at temperatures far beyond the electroweak scale $\bar{\mu} \sim 100$ GeV.

As the proposed EoS characterizes electroweak matter, a few remarks on the nature of the phase transitions is now in order. In some perturbative calculations, it was concluded that the electroweak EoS seems approaching that of an ideal gas,[18] while in others significant deviations have been obtained.[25] With this regard, it should be considered that the strength of electroweak phase transition is

[a] Decreasing the quark masses increases the thermodynamic pressure.

[b] This is a model–depending assumption. Originally, the electroweak theory has different scales along its wide range of temperatures and chemical potentials.

determined by different SM Lagrangian parameters, which so far aren't determined. The recently estimation of the Higgs mass allowed for a smooth crossover and electroweak baryogenesis.[26, 27]

2.5. *Combining Non-Perturbative and Perturbative EoS*

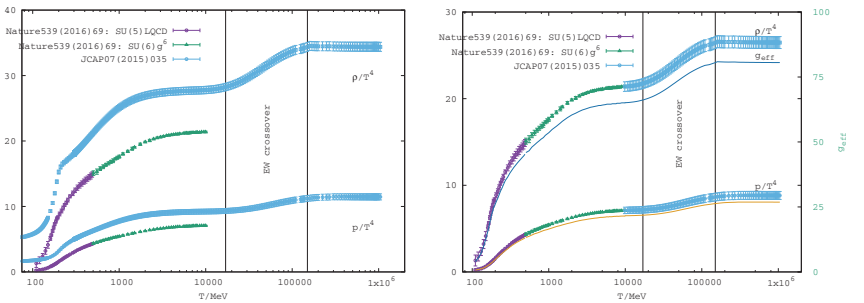


Fig. 1. Left panel: A comparison between non–perturbative lattice QCD simulations[9] and the perturbative calculations[18] for normalized pressure and energy density as functions of temperatures. Right panel: The same as in the left panel but here the perturbative calculation are rescalled (see main text). The vertical lines approximately determine the temperatures, where the electroweak crossover takes place. The curves in right panel present the effective dof $\mathtt{g_{eff}}(T)$ related to pressure and energy density.

As discussed, the proposal to combine reliable non–perturbative with perturbative calculations was already applied in various studies. In the present study, we combine the non–perturbative[9] with the perturbative calculations[18] for different thermodynamic quantities.

The temperature dependence of the normalized pressure and energy density calculated non–perturbatively and perturbatively is illustrated in the left panel of Fig. 1. Here, the temperature covers up to $\sim$ 10 GeV and $\sim$ 1 TeV in non–perturbative and perturbative simulations, respectively. It is obvious that although both types of simulations look quantitatively different, they are qualitatively similar. We find that the perturbative results are significantly larger than the non–perturbative ones.

There are many reasons supporting the assumption that the non–perturbative lattice QCD simulations are more reliable than the perturbative ones, especially at low temperatures. In this range of temperatures, even the vacuum energy change is best evaluated by non–perturbative simulations. On one hand, the non–perturbative lattice QCD simulations are most reliable, at low temperatures, the so–far perturbative calculations are not as accurate even at high temperatures. Thus, we concretely propose that the non–perturbative lattice QCD simulations are most reliable, at $T \gtrsim \Lambda_{\mathtt{QCD}} \approx 200$ MeV up to a few GeV,[28] while the perturbative calculations are the only approach possible, at temperatures up to the TeV–scale. From Fig. 1,

a temperature–indenedent systematic difference of $\sim 20\%$ between both sets of calculations[c] is obtained. Also, we assumed that the systematic difference of $\sim 20\%$ ins the same for both pressure and energy density.

For a contineous temperature–dependence of pressure and energy density, we propose to rescale the perturbative calcuations, at high temperatures.[18] The rescaling factor is phenomenologically adjusted to match the perturbative calculations, at high temperatures, with the non–perturbative lattice QCD simulations, at low temperatures. We assume that a rescaling factor of 0.77 remains constant, at the entire range of temperatures, right panel of Fig. 1. The non–perturbative lattice QCD simulations aren't rescaled, at all. In both thermodynamic quantities, both non–perturbative (T up to ~ 10 GeV) and perturbative ($10\ \texttt{GeV} \lesssim T \lesssim 1$ TeV) calculations are phenomenologically combined. The effective dof $\mathtt{g_{eff}}(T)$ corresponding to pressure, $\mathtt{g_{eff}}(T) = p(T)/p_0$, and to energy density, $\mathtt{g_{eff}}(T) = \rho(T)/\rho_0$, are also depicted in the right panel of Fig. 1, where $\rho_0 = (\pi^2/30)T^4$ is the energy density and p_0 is the pressure for an ideal gas of scalar massless bosons $p_0 = (\pi^2/90)T^4$. The perfect matching of $\mathtt{g_{eff}}(T)$, at low and high temperatures, supports the conclusion that the temperature–independent proposed rescaling, 0.77, seems precise.

3. EoS, bulk viscosity and relaxation time for almost entire SM dof

3.1. *EoS for Strong and Electroweak Matter*

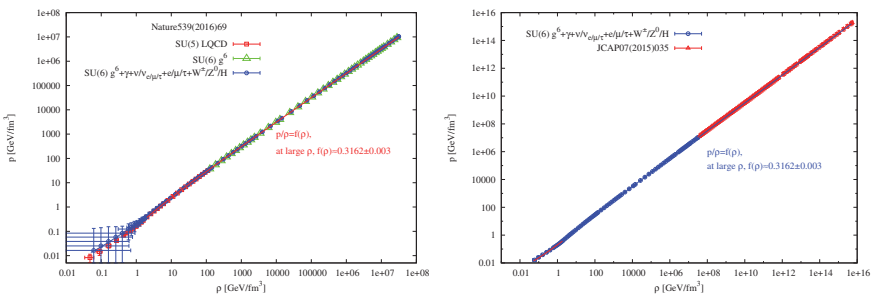


Fig. 2. Left panel: The thermodynamic pressure calculated in $\mathtt{SU(5)}$ non–perturbative and perturbative lattice QCD in dependence on the corresponding energy density (red and green symbols)[9] is confronted to the same calculations extended to other dof; γ, neutrinos, leptons, EW, and Higgs bosons (blue symbols). Right panel: The extended $\mathtt{SU(5)}$ lattice QCD calculations are combined with $\mathtt{SU(6)g^6}$ non–perturnative calculations.[21,29]

Left panel of Fig. 2 shows the EoS deduced from $\mathtt{SU(5)}$ non–perturbative and perturbative lattice QCD simulations (red and green symbols).[9] The thermodynamic contributions from γ, ν, e, μ, τ, ν_e, ν_μ, ν_τ, $W^\pm$, and the Higgs boson H are summed

[c]It was reported in Ref.[9] that this difference reads $7 - 17\%$ and is temperature dependent!

up with the SU(5) results (blue symbols). For each type of the additional particles, a partition function is constructed, from which the temperature dependence of the different thermodynamic quantities, such as the thermodynamic pressure and the energy density, can straightforwardly be derived. It is obvious that the additional dof considerably contribute to both thermodynamic quantities. The quantitative contributions are considrable, for example, within the hadronic phase, the proportionality constant in $p \propto \rho$ increases with the additive dof, i.e., increasing speed of sound squared.[30] In quark–gluon plasma and electroweak phases, more structures are added to the one corresponding EoS based on SU(5) lattice QCD simulations.

The fitting functions illustrated in the bottom panel of Fig. 2 are

$$\texttt{HP}: \qquad p = \alpha_1 + \alpha_2 \rho, \tag{10}$$

$$\texttt{QGP/EW}: \qquad p = \beta_1 + \beta_2 \rho + \beta_3 \rho^{d_2}, \tag{11}$$

$$\texttt{Asymp.}: \qquad p = \gamma \rho, \tag{12}$$

where $\alpha_1 = 0.0034 \pm 0.0023$, $\alpha_2 = 0.1991 \pm 0.0022$, $\beta_1 = 0.0484 \pm 0.0164$, $\beta_2 = 0.3162 \pm 0.0031$, $\beta_3 = -0.21 \pm 0.014$, and $\gamma = 0.3162 \pm 0.003$.

3.2. *Bulk Viscosity for Strong and Electroweak Matter*

It was concluded that the shear viscosity normalized to the entropy density for perturbative gauge QCD likely approaches the lower bound of Anti–de Sitter/Conformal Field Theory (AdS/CFT).[31] A non–perturbative estimation for viscous coefficients, at temperatures several times the QCD scale, has been reported in Ref.[32,33]. This was possible through accumulating a large amount of configurations for the Green function expressed in the Matsubara frequencies and implemented on isotropic $24^2 \times 8$ and $16^2 \times 8$ lattices. The viscous coefficients are determined as slopes of the spectral functions, at vanishing Matsubara frequency. A recent estimation for the temperature dependence of the bulk viscosity of SU(3) gluodynamics was possible with $48^3 \times 16$ lattice QCD simulations.[33] Another estimation is based the retarded Green function defined by the Kramers–Kronkig relation in terms of different thermodynamic quantities.[34,35] In Ref.[35], it was taken into consideration that the bulk viscosity measures the violation of the conformal invariance. This allowed to conclude that QCD at classical level is conformally invariant.

$$\begin{aligned}\zeta = \frac{1}{9\omega_0}\Big[& T\,s\left(\frac{\partial \rho}{\partial p} - 3\right) - 4(\rho - 3p) && \texttt{thermal parts} \\ & + \left(T\frac{\partial}{\partial T} - 2\right)\langle \bar{q} q \rangle(T) + g_g\, G^2(T) && \texttt{thermal q \& g condensates} \\ & + g_f\left(m_\pi^2 f_\pi^2 + m_K^2 f_K^2 + m_D^2 f_D^2 + \cdots\right)\Big]. && \texttt{vacuum q \& g condensates}\end{aligned} \tag{13}$$

where g_g (g_f) are the degeneracy factors for gluons (quarks). The spin polarization multiplied is $g_g = 16$. The color degrees of freedom reads $N_c^2 - 1$, with N_c is the color quantum number. $g_f = 12\, n_f$ with n_f are the degrees of freedom of the quark

flavors. m_D (f_D) are mass (decay constant of D-meson). The scale parameter ω_0 determines the applicability of the parturbation theory. ζ is obtained by using the frequency limit of the spectral density, at vanishing spatial momentum[35,36] and by implementing the various thermodynamic quantities are detailed in Eq. (13).

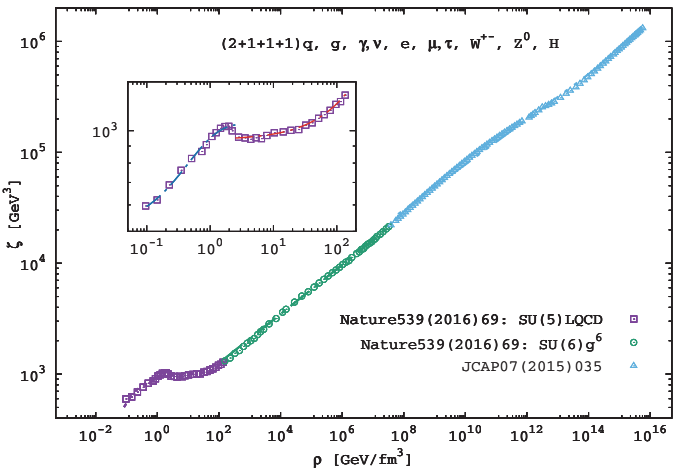


Fig. 3. The bulk viscosity ζ in dependence of the energy density ρ. Both quantities are calculated, at vanishing baryon–chemical potential and given in physical units. The inside box magnifies the region, at low temperature, where $\zeta(\rho)$ is non–monotonic.

By combining the gluons and $(2+1+1+1)$ quark contributions and that of the gauge bosons; the photons, $W^{\pm}$, and Z^0, of the charged leptons; neutrino, electron, muon, and tau, and of the Higgs bosons; scalar Higgs particle,[8] the bulk viscosity is shown in Fig. 3. We conclude that the almost entire SM contributions are very significant. The missing SM–contributions are vacuum and thermal bottom quark condensate, the entire gravitational sector, neutral leptons, and top quark. ζ almost linearly increases with increasing energy–density, so that following three regions of parameterizations (curves) can be distinguished

$$\texttt{Hadron}-\texttt{QGP}:\ \zeta = a_1 + a_2\rho + a_3\rho^{a_4}, \tag{14}$$

$$\texttt{QCD}:\ \zeta = b_1 + b_2\rho^{b_3}, \tag{15}$$

$$\texttt{EW}:\ \zeta = c_1 + c_2\rho^{c_3}. \tag{16}$$

For Hadron–QCD: $a_1 = -9.336 \pm 4.152$, $a_2 = 0.232 \pm 0.003$, $a_3 = 11.962 \pm 4.172$, and $a_4 = 0.087 \pm 0.029$. For QCD: $b_1 = 8.042 \pm 0.056$, $b_2 = 0.301 \pm 0.002$, and $b_3 = 0.945 \pm 0.0001$. For EW: $c_1 = 0.350 \pm 0.065$, $c_2 = 10.019 \pm 0.934$, and $c_3 = 0.929 \pm 8.898 \times 10^{-5}$.

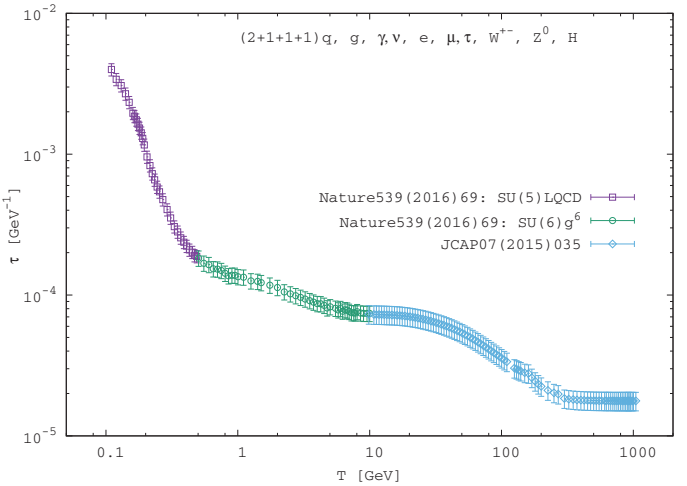


Fig. 4. At vanishing baryon–chemical potential, the relaxation time τ is presented in dependence on the temperature T. The results are based on non–perturbative and perturbative lattice QCD simulations, in which contributions from the quark and gluon condensates and the thermodynamic quantities of gauge bosons, charged leptons, and Higgs bosons (left symbols) are integrated.

3.3. *Relaxation Time for Strong and Electroweak Matter*

The relaxation time τ_f, which involves complicated collision integrals, can be determined as the mean collision time and averages of thermodynamic quantities[37–39]

$$\tau_f(T) = \frac{1}{n_f(T)\,\langle v(T)\rangle \sigma(T)}, \tag{17}$$

where σ is cross section, $\langle v(T)\rangle$ stands for mean relative velocity, and $n_f(T)$ represents the number density. Approaches for $\langle v(T)\rangle$ and σ have been discussed in Ref.[39]. The temperature dependence of the relaxation time $\tau_f(T)$ plays an essential role in bulk viscosity $\zeta(T)$.

Figure 4 presents $\tau(T)$ deduced from the non–perturbative and perturbative QCD simulations[8,9,21,29] (bottom symbols). Contributions from the quark and gluon condensates and the thermodynamic quantities of the gauge bosons, charged leptons, and Higgs bosons are added (left symbols). We notice that τ steadily decreases with increasing T. In the different phases, there are different rates of decreasing τ.

4. Conclusions

Most reliable non–perturbative lattice QCD simulations and perturbative calculations with as much as possible quark flavors with physical masses are phenomenologically combined. The thermodynamic contributions of photons, charged neutrinos, leptons, electroweak particles ($W^{\pm}$ and Z^0 bosons), and the scalar Higgs bosons are also integrated in. This makes the present study pioneering in a) simultaneous

accessing hadron, quark–gluon plasma and electroweak epochs of the early Universe and b) including almost all degrees of freedom of the standard model. With this regard, we emphasize that the only missing SM dof are vacuum and thermal bottom quark condensate, the entire gravitational sector, neutral leptons, and top quark flavor.

We have introduced various thermodynamic quantities, including pressure, energy density, bulk viscosity, and relaxation time, at vanishing net–baryon cosmic matter and temperatures up to the TeV–scale. The main result is that recent non–perturbative lattice QCD simulations and perturbative calculations jointly lead to EoS covering a wide range of temperatures. As never done before, this study makes introduces a framework to combine the standard model of the elementary particles and the standard model for cosmology.

References

1. C. W. Misner, The isotropy of the universe, *The Astrophysical Journal* **151**, p. 431 (1968).
2. Ya. B. Zeldovich and I. D. Novikov, *Relativistic Astrophysics. Vol. 2. The Structure and Evolution of the Universe* 1983.
3. L. Husdal, Viscosity in a lepton-photon universe, *Astrophysics and Space Science* **361**, 1 (2016).
4. A. Tawfik, QCD phase diagram: A Comparison of lattice and hadron resonance gas model calculations, *Phys. Rev. D* **71**, p. 054502 (2005).
5. A. Tawfik, M. Wahba, H. Mansour and T. Harko, Hubble Parameter in QCD Universe for finite Bulk Viscosity, *Annalen Phys.* **522**, 912 (2010).
6. L. Adamczyk *et al.*, Probing parton dynamics of QCD matter with Ω and ϕ production, *Phys. Rev. C* **93**, p. 021903 (2016).
7. E. A. Hakk, A. N. Tawfik, A. Nada and H. Yassin, Cosmic Evolution of Viscous QCD Epoch in Causal Eckart Frame, *Universe* **7**, p. 112 (2021).
8. A. N. Tawfik and I. Mishustin, Equation of State for Cosmological Matter at and beyond QCD and Electroweak Eras, *J. Phys. G* **46**, p. 125201 (2019).
9. S. Borsanyi *et al.*, Calculation of the axion mass based on high-temperature lattice quantum chromodynamics, *Nature* **539**, 69 (2016).
10. S. Borsanyi *et al.*, High-precision scale setting in lattice QCD, *JHEP* **09**, p. 010 (2012).
11. M. Cheng, Charm Quarks and the QCD Equation of State, *PoS* **LATTICE2007**, p. 173 (2007).
12. L. Levkova, Effects of the charm quark on the QCD equation of state, *PoS* **LAT2009**, p. 170 (2009).
13. F. Burger, G. Hotzel, M. Müller-Preussker, E.-M. Ilgenfritz and M. P. Lombardo, Towards thermodynamics with $N_f = 2+1+1$ twisted mass quarks, *PoS* **Lattice2013**, p. 153 (2013).
14. A. Bazavov *et al.*, Update on the 2+1+1 Flavor QCD Equation of State with HISQ, *PoS* **LATTICE2013**, p. 154 (2014).
15. S. Borsanyi, G. Endrodi, Z. Fodor, S. D. Katz, S. Krieg, C. Ratti, C. Schroeder and K. K. Szabo, The QCD equation of state and the effects of the charm, *PoS* **LATTICE2011**, p. 201 (2011).
16. R. Bellwied, S. Borsanyi, Z. Fodor, S. D. Katz, A. Pasztor, C. Ratti and K. K. Szabo, Fluctuations and correlations in high temperature QCD, *Phys. Rev. D* **92**, p. 114505 (2015).

17. J. O. Andersen, L. E. Leganger, M. Strickland and N. Su, NNLO hard-thermal-loop thermodynamics for QCD, *Phys. Lett. B* **696**, 468 (2011).
18. M. Laine and Y. Schroder, Quark mass thresholds in QCD thermodynamics, *Phys. Rev. D* **73**, p. 085009 (2006).
19. K. Kajantie, M. Laine, K. Rummukainen and Y. Schroder, The Pressure of hot QCD up to g6 ln(1/g), *Phys. Rev. D* **67**, p. 105008 (2003).
20. N. Brambilla, X. Garcia i Tormo, J. Soto and A. Vairo, The Logarithmic contribution to the QCD static energy at N**4 LO, *Phys. Lett. B* **647**, 185 (2007).
21. M. Laine and M. Meyer, Standard Model thermodynamics across the electroweak crossover, *JCAP* **07**, p. 035 (2015).
22. M. Laine, G. Nardini and K. Rummukainen, First order thermal phase transition with 126 GeV Higgs mass, *PoS* **LATTICE2013**, p. 104 (2014).
23. A. D. Linde, Infrared Problem in Thermodynamics of the Yang-Mills Gas, *Phys. Lett. B* **96**, 289 (1980).
24. D. J. Gross, R. D. Pisarski and L. G. Yaffe, QCD and Instantons at Finite Temperature, *Rev. Mod. Phys.* **53**, p. 43 (1981).
25. A. Gynther and M. Vepsalainen, Pressure of the standard model at high temperatures, *JHEP* **01**, p. 060 (2006).
26. K. Kajantie, M. Laine, K. Rummukainen and M. E. Shaposhnikov, Is there a hot electroweak phase transition at $m_H \gtrsim m_W$?, *Phys. Rev. Lett.* **77**, 2887 (1996).
27. Y. Aoki, F. Csikor, Z. Fodor and A. Ukawa, The Endpoint of the first order phase transition of the SU(2) gauge Higgs model on a four-dimensional isotropic lattice, *Phys. Rev. D* **60**, p. 013001 (1999).
28. S. J. Brodsky, S. Menke, C. Merino and J. Rathsman, On the behavior of the effective QCD coupling alpha(tau)(s) at low scales, *Phys. Rev. D* **67**, p. 055008 (2003).
29. M. D'Onofrio and K. Rummukainen, Standard model cross-over on the lattice, *Phys. Rev. D* **93**, p. 025003 (2016).
30. A. Nasser Tawfik and H. Magdy, Hadronic Equation of State and Speed of Sound in Thermal and Dense Medium, *Int. J. Mod. Phys. A* **29**, p. 1450152 (2014).
31. P. Kovtun, D. T. Son and A. O. Starinets, Viscosity in strongly interacting quantum field theories from black hole physics, *Phys. Rev. Lett.* **94**, p. 111601 (2005).
32. S. Sakai and A. Nakamura, Lattice calculation of the QGP viscosities: Present results and next project, *PoS* **LATTICE2007**, p. 221 (2007).
33. N. Y. Astrakhantsev, V. V. Braguta and A. Y. Kotov, Temperature dependence of the bulk viscosity within lattice simulation of $su(3)$ gluodynamics, *Phys. Rev. D* **98**, p. 054515 (Sep 2018).
34. D. Kharzeev and K. Tuchin, Bulk viscosity of QCD matter near the critical temperature, *JHEP* **09**, p. 093 (2008).
35. F. Karsch, D. Kharzeev and K. Tuchin, Universal properties of bulk viscosity near the QCD phase transition, *Phys. Lett.* **B663**, 217 (2008).
36. J. Noronha-Hostler, J. Noronha and C. Greiner, Transport Coefficients of Hadronic Matter near T(c), *Phys. Rev. Lett.* **103**, p. 172302 (2009).
37. A. N. Tawfik, A. M. Diab and M. T. Hussein, SU(3) Polyakov linear-sigma model: Conductivity and viscous properties of QCD matter in thermal medium, *Int. J. Mod. Phys.* **A31**, p. 1650175 (2016).
38. A. Tawfik and T. Harko, Quark-Hadron Phase Transitions in Viscous Early Universe, *Phys. Rev.* **D85**, p. 084032 (2012).
39. A. Tawfik, M. Wahba, H. Mansour and T. Harko, Viscous Quark-Gluon Plasma in the Early Universe, *Annalen Phys.* **523**, 194 (2011).

Neutral fermion pair production by Sauter-like magnetic step

T. C. Adorno

Department of Physics, College of Physical Sciences and Technology, Hebei University, Wusidong Road 180, 071002, Baoding, China;
Key Laboratory of High-precision Computation and Application of Quantum Field Theory of Hebei Province, Hebei University, Wusidong Road 180, 071002, Baoding, China;
Department of Physics, Tomsk State University, Lenin Prospekt 36, 634050, Tomsk, Russia
E-mail: adorno@hbu.edu.cn

Zi-Wang He

Department of Physics, College of Physical Sciences and Technology, Hebei University, Wusidong Road 180, 071002, Baoding, China;
Key Laboratory of High-precision Computation and Application of Quantum Field Theory of Hebei Province, Hebei University, Wusidong Road 180, 071002, Baoding, China
E-mail: 1248049347@qq.com

S. P. Gavrilov

Department of Physics, Tomsk State University, Lenin Prospekt 36, 634050, Tomsk, Russia;
Department of General and Experimental Physics, Herzen State Pedagogical University of Russia, Moyka embankment 48, 191186, St. Petersburg, Russia
E-mail: gavrilovsp@herzen.spb.ru

D. M. Gitman

Department of Physics, Tomsk State University, Lenin Prospekt 36, 634050, Tomsk, Russia;
P. N. Lebedev Physical Institute, 53 Leninskiy prospekt, 119991, Moscow, Russia;
Instituto de Física, Universidade de São Paulo, Caixa Postal 66318, CEP 05508-090, São Paulo, S.P., Brazil
E-mail: gitman@if.usp.br

We review our recent results on the creation from the vacuum of neutral fermions with anomalous magnetic moments by a Sauter-like magnetic field. We construct in- and out solutions of the Dirac-Pauli equation with this field and calculate with their help pertinent quantities characterizing the vacuum instability, such as differential mean numbers and flux density of created pairs and and vacuum-to-vacuum transition amplitudes. Special attention is paid to situations where the external field lies in two particular configurations, varying either "gradually" or "sharply" along the inhomogeneity direction. We also estimate critical magnetic field intensities, near which the phenomenon could be observed.

Keywords: Dirac-Pauli equation; quantum electrodynamics; pair production; neutral fermions with anomalous magnetic moment

1. Introduction

The violation of the vacuum stability stimulated by external electromagnetic fields is commonly associated with the possibility of such backgrounds producing work

on virtual pairs of particles and antiparticles. The most well-known examples are electric-like fields, as they produce work on charged particles and are able to tear apart electron/positron pairs from the vacuum if the field amplitudes approach the so-called Schwinger critical value $E_{\rm c} = m^2c^3/e\hbar \approx 1.3 \times 10^{16}\,{\rm V/cm}$.[1] The phenomenon has been a subject of intense investigation since the seminal works of Klein,[2] Sauter,[3,4] Heisenberg and Euler,[5] and Schwinger.[1] An extensive discussion about the origin of the effect, theoretical foundations, and experimental aspects can be found in some reviews and monographs; see e.g.[6–15] and references therein.

Following the above interpretation, one may ask oneself about the possibility that inhomogeneous macroscopic magnetic fields which produce a work on particles with a magnetic moment, may create pairs from the vacuum. The answer to this question is affirmative, provided the particles are neutral and have an anomalous magnetic moment. Bearing in mind, first of all, very strong magnetic fields observed in astrophysics, we can assume that this type of field is practically time-independent and steplike, that is, their gradient has a well-defined sign. At present, there exist two types of particles enjoying the properties mentioned above: the neutron and the neutrino. According to experimental data, neutrons have a magnetic moment $\mu_N \approx -1.04187563(25) \times 10^{-3}\mu_{\rm B}$,[16] where $\mu_{\rm B}$ is the Bohr magneton. As for neutrinos, there is not a general consensus because of the different types of neutrinos, mechanism under which neutrinos acquires magnetic moment, specific models, etc. Presently, experimental constraints range from $\mu_{\nu_\tau} < 3.9 \times 10^{-7}\mu_{\rm B}$ (for the tau neutrino)[17] until $\mu_{\nu_e} < 2.9 \times 10^{-11}\mu_{\rm B}$ (for the electron neutrino).[18] Moreover, stringent constraints obtained from astrophysical observations[19–24] indicate that $\mu_\nu < (2.6 - 4.5) \times 10^{-12}\mu_{\rm B}$ while lower upper bounds, predicted by effective theories above the electroweak scale, suggest that $\mu_\nu < 10^{-14}\mu_{\rm B}$.[25] It is important to point out that for some theories beyond the Standard Model (SM),[26] it was reported that the magnetic moment for the neutrinos lie within the range $\left(10^{-12} - 10^{-14}\right)\mu_{\rm B}$. For a more extensive discussion concerning experimental aspects and theoretical predictions for neutrinos' electromagnetic properties, see e.g. the reviews[27–31] and references therein.

In this work, we review our recent results on the creation of neutral fermions with anomalous magnetic moments from the vacuum by Sauter-like magnetic field.[32] We follow the general formulation developed to describe the effect nonperturbatively, which is based on the canonical quantization of fermion fields with time-independent inhomogeneous external fields.[33–35] The system under consideration is placed in the four-dimensional Minkowski spacetime, parameterized by coordinates $X = (X^\mu\,,\ \mu = 0, i) = (t, \mathbf{r})$, $X^0 = t$, $X^i = \mathbf{r} = (x, y, z)$, $i = 1, 2, 3$, with the metric tensor $\eta_{\mu\nu} = {\rm diag}\,(+1, -1, -1, -1)$. We also employ natural units, in which $\hbar = 1 = c$.

2. Solutions of Dirac-Pauli equation with Sauter-like magnetic step

The Dirac-Pauli (DP) equation[36] for a neutral spin 1/2 particle with the anomalous magnetic moment μ, the mass m interacting with external electromagnetic backgrounds $A^\mu = \left(A^0, \mathbf{A}\right)$ has the form:

$$\left(i\gamma^\mu \partial_\mu - m - \frac{1}{2}\mu\sigma^{\mu\nu}F_{\mu\nu}\right)\psi\left(X\right) = 0\,,$$
$$\sigma^{\mu\nu} = \frac{i}{2}\left[\gamma^\mu, \gamma^\nu\right]_-\,, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu\,. \tag{1}$$

Here $\psi\left(X\right)$ is a bispinor, $\gamma^\mu = \left(\gamma^0, \boldsymbol{\gamma}\right)$ are Dirac matrices, and μ is the algebraic value of the anomalous magnetic moment (e.g., $\mu = -\left|\mu_N\right|$ for a neutron). In what follows we consider external electromagnetic fields of a specific type, corresponding to a time-independent magnetic field oriented along the positive direction of the z-axis, inhomogeneous along the y-direction, $\mathbf{B}\left(\mathbf{r}\right) = \left(0, 0, B_z\left(y\right)\right)$, and homogeneous at remote distances, $B_z\left(\pm\infty\right) = \text{const}$. Moreover, it is assumed that its gradient is always positive $\partial_y B\left(y\right) \geq 0$, $\forall\, y \in \left(-\infty, +\infty\right)$, meaning that $B_z\left(+\infty\right) > B_z\left(-\infty\right)$ and that the field is genuinely a step (or steplike, in short). To study neutral fermion pair production by steplike magnetic fields, we consider that the magnetic field inhomogeneity is given by the analytic function

$$B_z\left(y\right) = \varrho B' \tanh\left(y/\varrho\right)\,, \quad B' > 0\,, \quad \varrho > 0\,, \tag{2}$$

which meets the conditions discussed above and allows solving the DP equation exactly. The amplitude B' and the length scale ϱ describe, respectively, the "slope" of the field with respect to the y-axis and how "rectilinear" it is in the neighborhood of the z-axis. Thus, the larger B' and ϱ, the more "steep" and the more "rectilinear" the pattern of (2) near the z-axis. Because the field (2) resembles a step-like "potential" for charged particles and its gradient has a Sauter profile,[4] we call the field (2) Sauter-like magnetic step. For illustrative purposes, we present the field (2) and its gradient for some values of ϱ and B' in Fig. 1.

In the Schrödinger form, the DP equation (1) reads[32,33,37]

$$i\partial_t\psi\left(X\right) = \hat{H}\psi\left(X\right)\,, \quad \hat{H} = \gamma^0\left(\gamma^3\hat{p}_z + \Sigma_z\hat{\Pi}_z\right)\,. \tag{3}$$

Here $\Sigma_z = i\gamma^1\gamma^2$ and

$$\hat{\Pi}_z = \hat{\pi}_z - \mathbb{I}\mu B_z\left(y\right)\,, \quad \hat{\pi}_z = \Sigma_z\left(\boldsymbol{\gamma}\hat{\mathbf{p}}_\perp + m\right)\,, \tag{4}$$

is an integral-of-motion spin operator, $\left[\hat{\Pi}_z, \hat{H}\right]_- = 0$. The subscript "$\perp$" labels quantities perpendicular to the field, e.g. $\hat{\mathbf{p}}_\perp = \left(\hat{p}_x, \hat{p}_y\right)$ and $\mathbb{I}$ denotes the 4×4 identity matrix. Since the operators $\hat{p}_0$, $\hat{p}_x$, and $\hat{p}_z$ are compatible with the Hamiltonian (and also with $\hat{\Pi}_z$), the DP spinor admits the general form $\psi_n\left(X\right) = \exp\left(-ip_0 t + ip_x x + ip_z z\right)\psi_n\left(y\right)$, where $\psi_n\left(y\right)$ depends exclusively on y

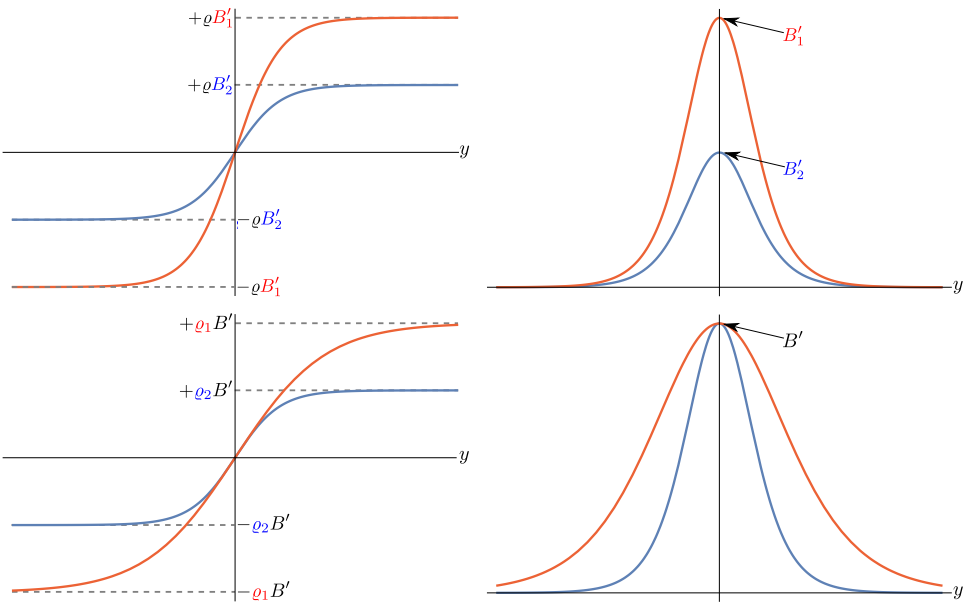


Fig. 1. Sauter-like magnetic steps (2) (pictures on the left) and its gradient (pictures on the right). In the upper panel, $B_1' = 2 \times B_2'$ while in the lower panel, $\varrho_1 = 2 \times \varrho_2$.

and is solution of the eigenvalue equation:

$$\hat{\Pi}_z \psi_n(X) = e^{-ip_0 t + ip_x x + ip_z z} \Pi_z \psi_n(y) \ , \quad \Pi_z \psi_n(y) = s\omega \psi_n(y) \ , \quad s = \pm 1 \, ,$$
$$\Pi_z = \hat{\pi}_z - \mathbb{I}\mu B_z(y) \ , \quad \hat{\pi}_z = \Sigma_z \left(\gamma^1 p_x + \gamma^2 \hat{p}_y + m\right) \ , \quad p_0^2 = \omega^2 + p_z^2 \, . \tag{5}$$

Due to the structure of the external field (2), there is an additional integral-of-motion spin operator

$$\hat{R} = \hat{H}\hat{\Pi}_z^{-1} \left[\mathbb{I} + \left(\hat{p}_z \hat{\Pi}_z^{-1}\right)^2\right]^{-1/2} \ , \tag{6}$$

which is compatible with the Hamiltonian $\left[\hat{R}, \hat{H}\right]_- = 0$ and with all previous operators, $\left[\hat{R}, \hat{\Pi}_z\right]_- = \left[\hat{R}, \hat{p}_0\right]_- = \left[\hat{R}, \hat{p}_x\right]_- = \left[\hat{R}, \hat{p}_z\right]_- = 0$. In particular, the operator (6) implies that $\psi_n(y)$ is a solution of the eigenvalue equation

$$\hat{R}\psi_n(X) = e^{-ip_0 t + ip_x x + ip_z z} R\psi_n(y) \ , \quad R\psi_n(y) = s\psi_n(y) \ ,$$
$$R = \Upsilon \gamma^0 \left(\Sigma_z + \frac{s p_z}{\omega} \gamma^3\right) \ , \quad \Upsilon = \frac{1}{\sqrt{1 + p_z^2/\omega^2}} \, . \tag{7}$$

As a result, the complete set of commuting operators is $\hat{p}_x$, $\hat{p}_z$, $\hat{\Pi}_z$, $\hat{R}$ and the corresponding quantum numbers are $n = (p_x, p_z, \omega, s)$.

The set of equations (5) and (7) are simultaneously satisfied choosing $\psi_n(y)$ in the form

$$\psi_n(y) = (\mathbb{I} + sR)\left[\hat{\pi}_z + \mathbb{I}(\mu B_z(y) + s\omega)\right]\varphi_{n,\chi}(y)\upsilon_\kappa^{(\chi)}, \tag{8}$$

where $\upsilon_\kappa^{(\chi)}$ belongs to a set of four constant spinors, satisfying the eigenvalue equations

$$i\gamma^1\upsilon_\kappa^{(\chi)} = \chi\upsilon_\kappa^{(\chi)}, \quad \gamma^0\gamma^2\upsilon_\kappa^{(\chi)} = \kappa\upsilon_\kappa^{(\chi)}, \quad \chi = \pm 1, \quad \kappa = \pm 1, \tag{9}$$

and the orthonormality conditions $\upsilon_{\kappa'}^{(\chi')\dagger}\upsilon_\kappa^{(\chi)} = \delta_{\chi'\chi}\delta_{\kappa'\kappa}$. As for the scalar functions $\varphi_{n,\chi}(y)$, they are solutions of the second-order ordinary differential equation:

$$\left\{-\frac{d^2}{dy^2} - [s\omega + \mu B_z(y)]^2 + \pi_x^2 + i\mu\chi B_z'(y)\right\}\varphi_{n,\chi}(y) = 0, \quad \pi_x^2 = m^2 + p_x^2. \tag{10}$$

The potential energy of a neutral fermion interacting with the field is $U_s(y) = sU(y)$, where $U(y) = -\mu B_z(y)$. To facilitate subsequent discussions, it is convenient to select a fixed sign for particle magnetic moment. From now on, we choose a fermion with a negative magnetic moment as the main particle, $\mu = -|\mu|$. Because the field (2) increases monotonically with y, the maximum potential energy that may be experienced by the fermion is determined by the magnitude of the "step" $\mathbb{U}$

$$\mathbb{U} \equiv U_{\mathrm{R}} - U_{\mathrm{L}} = 2\varrho|\mu|B' > 0, \tag{11}$$

which is the difference between the asymptotic values $U_{\mathrm{R}} = U(+\infty) = |\mu|\varrho B'$, $U_{\mathrm{L}} = U(-\infty) = -|\mu|\varrho B'$ and is positive, by definition[a]. At remote distances– where the field can be considered homogeneous and no longer accelerates particles– the term proportional to χ in Eq. (10) is absent. Therefore, solutions of Eq. (10) have well-defined "left" ${}_\zeta\varphi_{n,\chi}(y)$ and "right" ${}^\zeta\varphi_{n,\chi}(y)$ asymptotic forms:

$$\begin{aligned} {}_\zeta\varphi_{n,\chi}(y) &= {}_\zeta\mathcal{N}\exp\left(i\zeta\left|p^{\mathrm{L}}\right|y\right), \quad \zeta = \mathrm{sgn}\left(p^{\mathrm{L}}\right), \quad y \to -\infty, \\ {}^\zeta\varphi_{n,\chi}(y) &= {}^\zeta\mathcal{N}\exp\left(i\zeta\left|p^{\mathrm{R}}\right|y\right), \quad \zeta = \mathrm{sgn}\left(p^{\mathrm{R}}\right), \quad y \to +\infty. \end{aligned} \tag{12}$$

Here, ${}_\zeta\mathcal{N}$, ${}^\zeta\mathcal{N}$ are normalization constants, $\left|p^{\mathrm{L/R}}\right|$ are y-components of fermions momenta at remote regions

$$\left|p^{\mathrm{L/R}}\right| = \sqrt{[s\pi_s(\mathrm{L/R})]^2 - \pi_x^2}, \quad \pi_s(\mathrm{L/R}) = \omega - sU_{\mathrm{L/R}}, \tag{13}$$

and $\pi_s(\mathrm{L/R})$ are their transverse kinetic energies at remote areas. Correspondingly, asymptotically-"left" ${}_\zeta\psi_n(X) = \exp(-ip_0t + ip_xx + ip_zz)\,{}_\zeta\psi_n(y)$ and asymptotically-"right" ${}^\zeta\psi_n(X) = \exp(-ip_0t + ip_xx + ip_zz)\,{}^\zeta\psi_n(y)$ sets of DP spinors are solutions of the eigenvalue equations

$$\begin{aligned} \hat{p}_y\,{}_\zeta\psi_n(X) &= \zeta\left|p^{\mathrm{L}}\right|\,{}_\zeta\psi_n(X), \quad \hat{h}_\perp^{\mathrm{kin}}\,{}_\zeta\psi_n(X) = s\pi_s(\mathrm{L})\,{}_\zeta\psi_n(X), \quad y \to -\infty, \\ \hat{p}_y\,{}^\zeta\psi_n(X) &= \zeta\left|p^{\mathrm{R}}\right|\,{}^\zeta\psi_n(X), \quad \hat{h}_\perp^{\mathrm{kin}}\,{}^\zeta\psi_n(X) = s\pi_s(\mathrm{R})\,{}^\zeta\psi_n(X), \quad y \to +\infty, \end{aligned} \tag{14}$$

[a]The labels "L" and "R" mean "asymptotic left region $y \to -\infty$" and "asymptotic right region $y \to +\infty$", respectively.

where $\hat{h}_{\perp}^{\mathrm{kin}} = \hat{\Pi}_z - \mathbb{I}\,|\mu|\,B_z(y)$ is the one-particle transverse kinetic energy operator.

Substituting the field (2) into Eq. (10) and performing a simultaneous change of variables

$$\varphi_{n,\chi}(y) = \xi^{\rho}(1-\xi)^{\sigma} f(\xi)\ , \quad \xi(y) = \frac{1}{2}\left[1+\tanh(y/\varrho)\right]\ , \tag{15}$$

we may convert Eq. (10) to the form of the differential equation for the Gauss Hypergeometric Function[38]

$$\xi(1-\xi) f'' + \left[c - (a+b+1)\,\xi\right] f' - abf = 0\,, \tag{16}$$

provided the parameters ρ, σ, a, b, and c are:

$$\begin{aligned} a &= \frac{1}{2}(1-\chi) - \frac{i\varrho}{2}\left(\mathbb{U} + \left|p^{\mathrm{L}}\right| - \left|p^{\mathrm{R}}\right|\right)\ , \\ b &= \frac{1}{2}(1+\chi) + \frac{i\varrho}{2}\left(\mathbb{U} + \left|p^{\mathrm{R}}\right| - \left|p^{\mathrm{L}}\right|\right)\ , \\ c &= 1 - i\varrho\left|p^{\mathrm{L}}\right|\ , \quad \rho = -\frac{i}{2}\varrho\left|p^{\mathrm{L}}\right|\ , \quad \sigma = \frac{i}{2}\varrho\left|p^{\mathrm{R}}\right|\ . \end{aligned} \tag{17}$$

Among the 24 Hypergeometric functions satisfying Eq. (16),[38] we select those that tend to unity when $y \to \mp\infty$. Solutions meeting this property are proportional to Hypergeometric functions of type $F(a', b'; c'; \xi)$ and $F(a'', b''; c''; 1-\xi)$. For example, a possible set of exact solutions to Eq. (10) behaving asymptotically like Eqs. (12) is

$$\begin{aligned} {}_{\zeta}\varphi_{n,\chi}(y) &= {}_{\zeta}\mathcal{N}\exp\left(i\zeta\left|p^{\mathrm{L}}\right| y\right)\left[1+\exp(2y/\varrho)\right]^{-i\varrho\left(\zeta\left|p^{\mathrm{L}}\right|+\left|p^{\mathrm{R}}\right|\right)/2}\ {}_{\zeta}u(\xi)\ , \\ {}^{\zeta}\varphi_{n,\chi}(y) &= {}^{\zeta}\mathcal{N}\exp\left(i\zeta\left|p^{\mathrm{R}}\right| y\right)\left[1+\exp(-2y/\varrho)\right]^{i\varrho\left(\left|p^{\mathrm{L}}\right|+\zeta\left|p^{\mathrm{R}}\right|\right)/2}\ {}^{\zeta}u(\xi)\ , \end{aligned} \tag{18}$$

where

$$\begin{aligned} &{}_{-}u(\xi) = F(a,b;c;\xi)\ , \quad {}_{+}u(\xi) = F(a+1-c, b+1-c; 2-c; \xi)\ , \\ &{}^{-}u(\xi) = F(a,b;a+b+1-c;1-\xi)\ , \ {}^{+}u(\xi) = F(c-a, c-b; c+1-a-b; 1-\xi)\,. \end{aligned} \tag{19}$$

With the aid of these solutions, we may finally introduce the sets of DP spinors

$$\begin{aligned} {}_{\zeta}\psi_n(X) &= e^{-i(p_0 t - p_x x - p_z z)}\left(\mathbb{I} + sR\right)\left\{\hat{\pi}_z + \mathbb{I}\left[s\omega - |\mu|\,B_z(y)\right]\right\}\ {}_{\zeta}\varphi_{n,\chi}(y)\,\upsilon_{\kappa}^{(\chi)}\,, \\ {}^{\zeta}\psi_n(X) &= e^{-i(p_0 t - p_x x - p_z z)}\left(\mathbb{I} + sR\right)\left\{\hat{\pi}_z + \mathbb{I}\left[s\omega - |\mu|\,B_z(y)\right]\right\}\ {}^{\zeta}\varphi_{n,\chi}(y)\,\upsilon_{\kappa}^{(\chi)}\,, \end{aligned} \tag{20}$$

provided the quantum numbers n obey the conditions

$$\left[s\pi_s(\mathrm{L/R})\right]^2 > \pi_x^2\,. \tag{21}$$

These inequalities ensure the nontriviality of DP spinors with real asymptotic momenta p^{L} and p^{R} in remote areas, fulfilling Eqs. (14).

To calculate the normalization constants ${}_{\zeta}\mathcal{N}$, ${}^{\zeta}\mathcal{N}$, we use the inner product on the timelike surface $y = \mathrm{const.}$,

$$\left(\psi, \psi'\right)_y = \int dt dx dz \psi^{\dagger}(X)\,\gamma^0\gamma^2\psi'(X)\ , \tag{22}$$

after imposing specific normalization conditions[b]. We assume that all processes take place within a macroscopically large space-time box, of volume TV_y, $V_y = L_x L_z$, and impose periodic boundary conditions upon DP spinors in the variables t, x, z at the boundaries. Thus, the integrals in (22) are calculated from $(-T/2, -L_x/2, -L_z/2)$ to $(+T/2, +L_x/2, +L_z/2)$ and the limits $(T, L_x, L_z) \to \infty$ are taken at the end of calculations. Under these conditions, the inner product is y-independent and we may impose the following normalization conditions:

$$\left({}_{\zeta'}\psi_{n'},\ {}_{\zeta}\psi_{n} \right)_y = \zeta \eta_{\mathrm{L}} \delta_{n'n} \delta_{\zeta'\zeta}\,, \quad \left({}^{\zeta'}\psi_{n'},\ {}^{\zeta}\psi_{n} \right)_y = \zeta \eta_{\mathrm{R}} \delta_{n'n} \delta_{\zeta'\zeta}\,, \tag{23}$$

where $\eta_{\mathrm{L/R}} = \operatorname{sgn}\left[\pi_s\left(\mathrm{L/R}\right)\right]$. It should be noted that the time independence of the magnetic field under consideration is an idealization. Physically, it is meaningful to believe that the field inhomogeneity was switched on sufficiently fast before instant t_{in}. By this time, it had time to spread to the whole area under consideration and then acted as a constant field during a large time T. It is supposed that one can ignore effects of its switching on and off. This is a kind of regularization, which could, under certain conditions, be replaced by periodic boundary conditions in t, see Refs.[34,35] for details. Evaluating the inner product (22) for each DP spinor (20) and imposing the normalization conditions (23), we obtain

$$\left|\, {}_{\zeta}\mathcal{N}\right| = \frac{\left[TV_y \Upsilon \left(1 - s\kappa\chi\Upsilon\right)\right]^{-1/2}}{2\sqrt{\left|p^{\mathrm{L}}\right| \left|\pi_s\left(\mathrm{L}\right) - s\chi\zeta \left|p^{\mathrm{L}}\right|\right|}}\,, \quad \left|\, {}^{\zeta}\mathcal{N}\right| = \frac{\left[TV_y \Upsilon \left(1 - s\kappa\chi\Upsilon\right)\right]^{-1/2}}{2\sqrt{\left|p^{\mathrm{R}}\right| \left|\pi_s\left(\mathrm{R}\right) - s\chi\zeta \left|p^{\mathrm{R}}\right|\right|}}\,. \tag{24}$$

Considering that the "left" and "right" sets of DP spinors (20) are orthonormal and complete (with respect to the inner product (22)), we may decompose one set into another with the help of some g-coefficients

$$\begin{aligned} \eta_{\mathrm{L}}\ {}^{\zeta}\psi_n\left(X\right) &= g\left({}_{+}|^{\zeta}\right)\ {}_{+}\psi_n\left(X\right) - g\left({}_{-}|^{\zeta}\right)\ {}_{-}\psi_n\left(X\right)\,, \\ \eta_{\mathrm{R}}\ {}_{\zeta}\psi_n\left(X\right) &= g\left({}^{+}|_{\zeta}\right)\ {}^{+}\psi_n\left(X\right) - g\left({}^{-}|_{\zeta}\right)\ {}^{-}\psi_n\left(X\right)\,, \end{aligned} \tag{25}$$

which, by definition, are inner products between different sets of DP spinors

$$\left({}_{\zeta}\psi_n,\ {}^{\zeta'}\psi_{n'} \right)_y = \delta_{nn'} g\left({}_{\zeta}|^{\zeta'}\right) = \delta_{nn'} g\left({}^{\zeta'}|_{\zeta}\right)^{*}\,, \tag{26}$$

and play an important role in the quantization of DP spinors with steplike magnetic fields. These coefficients link different sets of creation and annihilation operators and contains all the necessary information about vacuum instability, as shall be seen below. Substituting the identities (25) into normalization conditions (23) supply us with two important identities

$$\sum_{\zeta''=\pm} \zeta'' g\left({}^{\zeta'}|_{\zeta''}\right) g\left({}_{\zeta''}|^{\zeta}\right) = \zeta \eta_{\mathrm{L}} \eta_{\mathrm{R}} \delta_{\zeta'\zeta} = \sum_{\zeta''=\pm} \zeta'' g\left({}_{\zeta'}|^{\zeta''}\right) g\left({}^{\zeta''}|_{\zeta}\right)\,, \tag{27}$$

from which we may derive a number of identities, for example $\left|g\left({}_{+}|^{-}\right)\right|^2 = \left|g\left({}_{-}|^{+}\right)\right|^2$, $\left|g\left({}_{+}|^{+}\right)\right|^2 = \left|g\left({}_{-}|^{-}\right)\right|^2$, and $\left|g\left({}_{+}|^{+}\right)\right|^2 - \left|g\left({}_{+}|^{-}\right)\right|^2 = \eta_{\mathrm{L}}\eta_{\mathrm{R}}$.

[b]Note that for $\psi' = \psi$, the inner product (22) divided by T coincides with the definition of the current density accross the y-const. hyperplane.

3. Pair production

Besides providing conditions for the existence of solutions (20), the inequalities (21) imposes certain limitations on the quantum numbers. For critical Sauter-like magnetic steps, whose step magnitudes (11) meet the condition

$$\mathbb{U} > \mathbb{U}_{\rm c} = 2m\,, \tag{28}$$

the whole manifold of quantum numbers divides into five sub-ranges, $\Omega_k\,,\ k = 1,...,5$. Neutral fermion pair production takes place only in a well-defined bounded set of quantum numbers[32–34]

$$\Omega_3 = \{n : U_{\rm L} + \pi_x \leq s\omega \leq U_{\rm R} - \pi_x\,,\ \ \pi_{xz} \leq \mathbb{U}/2\}\ ,\ \ \pi_{xz} = \sqrt{\pi_x^2 + p_z^2}\,, \tag{29}$$

which is conventionally called *Klein zone*, Ω_3. In this subrange, $s\pi_s\,({\rm L}) \geq \pi_x$ and $s\pi_s\,({\rm R}) \leq -\pi_x$, which means that $s\eta_{\rm L} = +1$ and $s\eta_{\rm R} = -1$. As a result, there exist two linearly-independent "left" ${}_{\zeta}\psi_{n_3}\,(X)$ and "right" ${}^{\zeta}\psi_{n_3}\,(X)$ sets of DP spinors with quantum numbers within the Klein zone $n_3 = n \in \Omega_3$.

To quantize the DP field operators using sets of solutions in this subrange, we need to classify them as particle or antiparticle states and as incoming waves (waves traveling toward the "step") or outgoing waves (waves traveling outward the "step") in remote areas. The correct classification demands a careful study of the inner product on y- and t-constant hyperplanes because important quantities to the scattering problem are expressed as surface integrals on such hyperplanes. After a detailed study of these quantities, which was presented in Refs.[34,35] for charged particles and in[33] for neutral fermions, "in"-solutions (incoming waves) and "out"-solutions (outgoing waves) are

$$\text{in-solutions: } {}_{-}\psi_{n_3}\,(X)\ ,\ {}^{-}\psi_{n_3}\,(X)\ ,\ \ \text{out-solutions: } {}_{+}\psi_{n_3}\,(X)\ ,\ {}^{+}\psi_{n_3}\,(X)\ . \tag{30}$$

The above sets of solutions are complete and orthogonal with respect to the inner product on t-constant hyperplane

$$(\psi_n, \psi'_{n'}) = \int_{V_y} dxdz \int_{-K^{(\rm L)}}^{K^{(\rm R)}} dy \psi_n^{\dagger}\,(X)\,\psi'_{n'}\,(X)\ , \tag{31}$$

in which the lower/upper cutoffs $K^{(\rm L/R)}$ are macroscopic but finite parameters of the volume regularization that are situated far beyond the region of a large gradient $\partial_y B_z\,(y)$; see Ref.[34,35] for details. In particular, the inner product (31) between "in" and "out" sets of DP spinors have the form

$$\left({}_{\zeta}\psi_n,\ {}_{\zeta}\psi_{n'}\right) = \left({}^{\zeta}\psi_n,\ {}^{\zeta}\psi_{n'}\right) = \mathcal{M}_n\delta_{nn'}\ ,\ \ \left({}_{\zeta}\psi_n,\ {}^{\zeta}\psi_{n'}\right) = 0\,,\ \ n, n' \in \Omega_3\,. \tag{32}$$

where $\mathcal{M}_n = 2\left|g\left({}_{+}|{}^{-}\right)\right|^2 t^{(\rm L/R)}/T$, $t^{(\rm L/R)} = K^{(\rm L/R)}\left|\pi_s\,({\rm L/R})\,/p^{\rm L/R}\right|$. Because there are two linearly independent sets of spinors (30), the quantization is performed using two distinct "in" and "out" sets of annihilation & creation operators

$$\begin{aligned}&\text{in-set: } {}_{-}b_{n_3}\,(\text{in})\ ,\ {}_{-}b^{\dagger}_{n_3}\,(\text{in})\ ,\ {}^{-}a_{n_3}\,(\text{in})\ ,\ {}^{-}a^{\dagger}_{n_3}\,(\text{in})\ ,\\ &\text{out-set: } {}_{+}b_{n_3}\,(\text{out})\ ,\ {}_{+}b^{\dagger}_{n_3}\,(\text{out})\ ,\ {}^{+}a_{n_3}\,(\text{out})\ ,\ {}^{+}a^{\dagger}_{n_3}\,(\text{out})\ ,\end{aligned} \tag{33}$$

which, in turn, obey the following anticommutation relations

$$\left[\ {}^{-}a_{n_3'}\left(\text{in}\right),\ {}^{-}a_{n_3}^{\dagger}\left(\text{in}\right)\right]_{+}=\left[\ {}_{-}b_{n_3'}\left(\text{in}\right),\ {}_{-}b_{n_3}^{\dagger}\left(\text{in}\right)\right]_{+}=\delta_{n_3'n_3}\,,$$
$$\left[\ {}^{+}a_{n_3'}\left(\text{out}\right),\ {}^{+}a_{n_3}^{\dagger}\left(\text{out}\right)\right]_{+}=\left[\ {}_{+}b_{n_3'}\left(\text{out}\right),\ {}_{+}b_{n_3}^{\dagger}\left(\text{out}\right)\right]_{+}=\delta_{n_3'n_3}\,, \tag{34}$$

and whose annihilation operators (33) annihilate the corresponding vacuum states

$${}_{-}b_{n_3}\left(\text{in}\right)\left|0,\text{in}\right\rangle={}^{-}a_{n_3}\left(\text{in}\right)\left|0,\text{in}\right\rangle=0,\ {}_{+}b_{n_3}\left(\text{out}\right)\left|0,\text{out}\right\rangle={}^{+}a_{n_3}\left(\text{out}\right)\left|0,\text{out}\right\rangle=0. \tag{35}$$

Finally, the quantized DP field operator in the Klein zone reads

$$\begin{aligned}\hat{\Psi}\left(X\right)&=\sum_{n\in\Omega_3}\mathcal{M}_n^{-1/2}\left[\ {}^{-}a_n\left(\text{in}\right)\ {}^{-}\psi_n\left(X\right)+\ {}_{-}b_n^{\dagger}\left(\text{in}\right)\ {}_{-}\psi_n\left(X\right)\right],\\&=\sum_{n\in\Omega_3}\mathcal{M}_n^{-1/2}\left[\ {}^{+}a_n\left(\text{out}\right)\ {}^{+}\psi_n\left(X\right)+\ {}_{+}b_n^{\dagger}\left(\text{out}\right)\ {}_{+}\psi_n\left(X\right)\right].\end{aligned} \tag{36}$$

Using orthogonality relations between DP spinors (23), (32) and the relations given by Eqs. (25), one may establish a linear relation between the "in"-set of creation/annihilation operators in terms of the "out"-set and *vice-versa*. For example, two (out of four) canonical transformations have the following form

$$\begin{aligned}{}_{-}b_n^{\dagger}\left(\text{in}\right)&=-g\left({}^{+}|_{-}\right)^{-1}\ {}^{+}a_n\left(\text{out}\right)+g\left({}_{+}|^{-}\right)^{-1}g\left({}_{-}|^{-}\right)\ {}_{+}b_n^{\dagger}\left(\text{out}\right),\\{}^{+}a_n\left(\text{out}\right)&=-g\left({}_{-}|^{+}\right)^{-1}\ {}_{-}b_n^{\dagger}\left(\text{in}\right)+g\left({}^{-}|_{+}\right)^{-1}g\left({}^{+}|_{+}\right)\ {}^{-}a_n\left(\text{in}\right).\end{aligned} \tag{37}$$

With the aid of the canonical transformations (37), we may finally define vacuum instability quantities, such as the differential mean numbers of "out" particles created from the "in" vacuum,

$$N_n^{\text{cr}}=\left\langle 0,\text{in}\right|\ {}^{+}a_n^{\dagger}\left(\text{out}\right)\ {}^{+}a_n\left(\text{out}\right)\left|\text{in},0\right\rangle=\left|g\left({}_{-}|^{+}\right)\right|^{-2},\quad n\in\Omega_3\,, \tag{38}$$

and the flux density of particles created with a given s,

$$n_s^{\text{cr}}=\frac{1}{V_yT}\sum_{n\in\Omega_3}N_n^{\text{cr}}=\frac{1}{\left(2\pi\right)^3}\int dp_z\int dp_x\int dp_0N_n^{\text{cr}}\,. \tag{39}$$

The total flux density of particles created with both spin polarizations is $n^{\text{cr}}=n_{+1}^{\text{cr}}+n_{-1}^{\text{cr}}$ and the vacuum-vacuum transition probability reads:

$$P_v=\left|\langle 0,\text{out}|0,\text{in}\rangle\right|^2=\exp\left[\sum_{s=\pm1}\sum_{n\in\Omega_3}\ln\left(1-N_n^{\text{cr}}\right)\right]. \tag{40}$$

It should be noted that if the total number of created particles $N^{\text{cr}}=V_yTn^{\text{cr}}$ is small, one may neglect higher-order terms in Eq. (40) to conclude that $P_v\approx1-N^{\text{cr}}$. With the aid of this relation, we may link the total number of neutral fermions created from the vacuum with the imaginary part of an effective action S_{eff} provided it satisfies the Schwinger relation $P_v=\exp\left(-2\text{Im}S_{\text{eff}}\right)$, and it is small, so that $P_v\approx1-2\text{Im}S_{\text{eff}}$. Therefore,

$$\text{Im}S_{\text{eff}}\approx V_yTn^{\text{cr}}/2\,. \tag{41}$$

From these equations, we observe that all the information about pair creation by the external field is enclosed in $g\left({}_{-}|^{+}\right)$. To obtain this coefficient, we may use an appropriate Kummer relation[38] that connects three Gauss Hypergeometric functions appearing in one of the relations given by Eq. (25). After obtaining this coefficient and calculating its absolute square $\left|g\left({}_{-}|^{+}\right)\right|^{-2}$, we finally obtain the differential mean numbers of pairs created from the vacuum:

$$N_n^{\mathrm{cr}} = \frac{\sinh\left(\pi\varrho\left|p^{\mathrm{R}}\right|\right)\sinh\left(\pi\varrho\left|p^{\mathrm{L}}\right|\right)}{\sinh\left[\pi\varrho\left(\mathbb{U}+\left|p^{\mathrm{L}}\right|-\left|p^{\mathrm{R}}\right|\right)/2\right]\sinh\left[\pi\varrho\left(\mathbb{U}+\left|p^{\mathrm{R}}\right|-\left|p^{\mathrm{L}}\right|\right)/2\right]}\,. \tag{42}$$

Note that N_n^{cr} are positive-definite because the difference $\left|\left|p^{\mathrm{L}}\right|-\left|p^{\mathrm{R}}\right|\right|$ bounded in this subrange; $0 \leq \left|\left|p^{\mathrm{L}}\right|-\left|p^{\mathrm{R}}\right|\right| \leq \sqrt{\mathbb{U}\left(\mathbb{U}-2\pi_x\right)}$. The above expression gives the exact distribution of neutral fermions created from the vacuum by the field (2). When summed over the quantum numbers, it provides exact expressions for the flux density of the created particles (39) and the vacuum-vacuum transition probability (40). Lastly, it is noteworthy to discuss some peculiarities associated with the choice of the quantum number s and its impact on the quantization (36). As pointed out in Sec. 2, there are two species of neutral fermions, one with $s = +1$ and another with $s = -1$. In the latter case, the classification differs from the one given by Eq (30), namely ${}_{+}\psi_{n_3}(X)$, ${}^{+}\psi_{n_3}(X)$ are "in"-solutions while ${}_{-}\psi_{n_3}(X)$, ${}^{-}\psi_{n_3}(X)$ are "out"-solutions. Although this classification changes the quantization (36), it does not change the mean numbers (38). This means that the flux density of particles created with $s = -1$ equals the one with $s = +1$, $n_{+1}^{\mathrm{cr}} = n_{-1}^{\mathrm{cr}}$. Therefore, summations over s in Eqs. (39), (40) just produce an extra factor of 2 in final expressions and that is why it is enough selecting s fixed to perform specific calculations; hereafter, we select $s = +1$ for convenience. In what follows, we analyze vacuum instability quantities when the field lies in two special configurations, varying either "gradually" or "sharply" along the inhomogeneity direction.

3.1. *"Gradually"-varying field configuration*

This field configuration corresponds to the case where the amplitude B' is sufficiently large and the field inhomogeneity stretches over a relatively wide region of the space, such that the condition

$$\sqrt{\varrho\mathbb{U}/2} \gg \max\left(1, \frac{m}{\sqrt{|\mu|\,B'}}\right)\,, \tag{43}$$

is satisfied. Accordingly, the arguments of the hyperbolic functions in (42) are large, meaning that the mean numbers of pairs created acquires the following approximate form,

$$N_n^{\mathrm{cr}} \approx e^{-\pi\tau}\,, \quad \tau = \varrho\left(\mathbb{U}-\left|p^{\mathrm{R}}\right|-\left|p^{\mathrm{L}}\right|\right)\,. \tag{44}$$

The above distribution is exponentially small for large values of ω and p_x. Its most significant contribution comes from a finite range of values of quantum numbers

such that the conditions $\min\left(\pi_{+1}^2\left(\mathrm{L}\right),\pi_{+1}^2\left(\mathrm{R}\right)\right)\gg\pi_x^2$ remains valid. In this case, τ admits the following approximation

$$\tau=\frac{\left(\mathbb{U}/2\right)^2}{\left(\mathbb{U}/2\right)^2-\omega^2}\lambda+O\left(\pi_x^4/\left|\pi_{+1}\left(\mathrm{R}\right)\right|^3\right)+O\left(\pi_x^4/\left|\pi_{+1}\left(\mathrm{L}\right)\right|^3\right). \tag{45}$$

Now, we can estimate the flux density of pairs created n^{cr} for a magnetic step evolving gradually along the y-direction according to (43). To this end, it is convenient to transform the original integral over p_0 into an integral over ω through the relation between p_0, ω, and p_z discussed before, $p_0^2=\omega^2+p_z^2$. Performing such a change of variables, the flux density of the particles created by the external field in the configuration (43) has the form

$$\begin{aligned}
&n^{\mathrm{cr}}\approx\frac{4}{\left(2\pi\right)^3}\int_0^{p_z^{\max}}dp_z\int_{-p_x^{\max}}^{p_x^{\max}}dp_x\int_0^{\omega_{\max}^2}d\omega^2\frac{e^{-\pi\tau}}{\sqrt{\omega^2+p_z^2}},\\
&p_z^{\max}=\sqrt{\left(\mathbb{U}/2\right)^2-m^2},\quad p_x^{\max}=\sqrt{\left(\mathbb{U}/2\right)^2-m^2-p_z^2},\quad \omega_{\max}=\mathbb{U}/2-\pi_x.
\end{aligned} \tag{46}$$

The multiplicative factor 4 comes from the summation over s and from the fact that the integrand is symmetric in p_z. To obtain an analytical expression to N^{cr}, we formally extend the integration limits of the last two integrals to infinity. This procedure amounts to incorporating exponentially small contributions to n^{cr} since the differential mean numbers are exponentially small at large p_x and ω. In this case, we may technically interchange the order of the last two integrals in (46) and use the approximation given by Eq. (45) to discover that the flux density of the created particles is approximately given by

$$n^{\mathrm{cr}}\approx\frac{m}{\left(2\pi\right)^3}\mathbb{U}^2\sqrt{b'}e^{-\pi b'}I_{b'},\quad I_{b'}=\int_0^\infty\frac{du}{\left(u+1\right)^{5/2}}\ln\left(\frac{\sqrt{1+u}+\sqrt{1+2u}}{\sqrt{u}}\right)e^{-\pi b'u}, \tag{47}$$

where $b'=m^2/\left|\mu\right|B'$ and $\mathbb{U}=2\varrho\left|\mu\right|B'$. At last, one may use the identity $\ln\left(1-N_n^{\mathrm{cr}}\right)=-\sum_{l=1}^\infty\left(N_n^{\mathrm{cr}}\right)^l/l$ and perform integrations similar to the ones discussed before to discover that the vacuum-vacuum transition probability admits the final form

$$P_v=\exp\left(-\beta V_yTn^{\mathrm{cr}}\right),\quad \beta=\sum_{l=0}^\infty\frac{\epsilon_{l+1}}{\left(l+1\right)^{3/2}}\exp\left(-l\pi b'\right),\quad \epsilon_l=\frac{I_{b'l}}{I_{b'}}, \tag{48}$$

with n^{cr} given by Eq. (47).

It is noteworthy mentioning that relation (40)–which is well-known for strong-field QED with external electromagnetic fields–holds for the case under consideration as well. However, a direct similarity of total quantities for both cases is absent. We see that the flux density of created neutral fermion pairs and the quantity $\ln P_v^{-1}$ are quadratic in the magnitude of the step. This is a consequence of the fact that the number of states with all possible ω and p_z excited by the magnetic-field inhomogeneity is quadratic in the increment of the kinetic momentum. This is also the

reason why the flux density of created pairs and $\ln P_v^{-1}$ per unit of the length are not uniform.

3.2. *"Sharply"-varying field configuration*

A second configuration of interest is when the field (2) "sharply" steeps near the origin. Such a configuration is specified by the conditions:

$$1 \gg \sqrt{\varrho\mathbb{U}/2} \gtrsim \frac{m}{\sqrt{|\mu|\, B'}}\,. \tag{49}$$

The first inequality indicates that the gradient $\partial_y B_z\,(y)$ sharply peaks about the origin, while the second implicates that the Klein zone is relatively small. This configuration is particularly important due to a close analogy to charged pair production by the Klein step, see Ref.[39] for the review. For electric fields whose spatial inhomogeneity meets conditions equivalent to (49), it was demonstrated that the imaginary part of the QED effective action features properties similar to those of continuous phase transitions.[40,41] Recently,[42] we have demonstrated for the inverse-square electric field that this peculiarity also follows from the behavior of total quantities when the Klein zone is relatively small. Because of the condition (49), not only the parameter $\varrho\mathbb{U}/2$ is small but all parameters involving the quantum numbers p_x, p_z, and ω are small as well on account of the inequalities (29). As a result, the arguments of the hyperbolic functions in (42) are small, which means that we may expand the hyperbolic functions in ascending powers and truncate the corresponding series to first-order to demonstrate that the mean numbers admit the approximate form:

$$N_n^{\mathrm{cr}} \approx \frac{4\left|p^{\mathrm{R}}\right|\left|p^{\mathrm{L}}\right|}{\mathbb{U}^2 - \left(\left|p^{\mathrm{L}}\right| - \left|p^{\mathrm{R}}\right|\right)^2}\,. \tag{50}$$

To implement the conditions (49), we conveniently introduce the Keldysh parameter $\gamma = 2m/\mathbb{U}$ and observe that it obeys the condition $1-\gamma^2 \ll 1$ on account of (49). Next, we perform the change of variables

$$\frac{\omega}{m} = \frac{1}{2}\left(1-\gamma^2\right)(1-v)\;,\quad \frac{p_x^2}{m^2} = \left(1-\gamma^2\right) r\,, \tag{51}$$

and expand the asymptotic momenta $\left|p^{\mathrm{L/R}}\right|$ in ascending powers of $1-\gamma^2$ to learn that $\left|p^{\mathrm{R}}\right|/m = \left(1-\gamma^2\right)^{1/2}\sqrt{v-r} + O\left(\left(1-\gamma^2\right)^{3/2}\right)$, $\left|p^{\mathrm{L}}\right|/m = \left(1-\gamma^2\right)^{1/2}\sqrt{2-v-r} + O\left(\left(1-\gamma^2\right)^{3/2}\right)$. Substituting these approximations into (50) we obtain

$$N_n^{\mathrm{cr}} = \left(1-\gamma^2\right)\sqrt{(1-r)^2-(1-v)^2} + O\left(\left(1-\gamma^2\right)^2\right)\,. \tag{52}$$

We now wish to estimate the total number of pairs created from the vacuum by a sharply varying external field. In this case, it is convenient to first integrate over p_z, which is allowed as long as we swap the integration limits indicated in (46), i.e.

$p_z^{\max} = \sqrt{(\mathbb{U}/2)^2 - m^2 - p_x^2}$ and $p_x^{\max} = \sqrt{(\mathbb{U}/2)^2 - m^2}$. Calculating the integral and performing the change of variables proposed in (51), we expand the result in power series of $1-\gamma^2$ to find

$$\int_0^{p_z^{\max}} \frac{dp_z}{\sqrt{\omega^2 + p_z^2}} = -\frac{1}{2}\ln\left(1-\gamma^2\right) + 2\ln 2 + \ln\sqrt{1-r} - \ln\sqrt{(1-v)^2} + O\left(1-\gamma^2\right). \tag{53}$$

The most significant contribution to total quantities in this regime comes from the logarithm $\ln\left(1-\gamma^2\right)$, as $1-\gamma^2 \ll 1$. Neglecting higher-order terms in $1-\gamma^2$, the flux density of the particles created is approximately given by

$$n^{\mathrm{cr}} \approx \frac{\left(1-\gamma^2\right)^{7/2}\left|\ln\left(1-\gamma^2\right)\right| m^3}{(2\pi)^3} \int_0^{r_{\max}} \frac{dr}{\sqrt{r}} \int_{v_{\min}}^{v_{\max}} dv\,(1-v)\sqrt{(1-r)^2 - (1-v)^2}, \tag{54}$$

where $v_{\min} \approx r$ and $v_{\max} \approx r_{\max} \approx 1$. After straightforward integrations, the flux density of the particles created from the vacuum by a sharply varying Sauter-like magnetic step takes the approximate form

$$n^{\mathrm{cr}} \approx \frac{4}{105\pi^3} m^3 \left(1-\gamma^2\right)^{7/2}\left|\ln\left(1-\gamma^2\right)\right|. \tag{55}$$

Due to the smallness of the coefficient $\left(1-\gamma^2\right)^{7/2}$, the total number of neutral fermions pairs created from the vacuum is also small $N^{\mathrm{cr}} = V_y T n^{\mathrm{cr}}$, which means that the vacuum-vacuum transition probability is approximately given by $P_v \approx 1 - N^{\mathrm{cr}}$. We may use this result to link the flux density of pairs created (55) with the imaginary part of the effective action, given by the approximation (41).

4. Concluding remarks

Here we review our recent results on the creation of neutral fermion pairs with anomalous magnetic moments from the vacuum by Sauter-like magnetic field.[32] We show that the problem is technically analogous to the problem of charged-particle creation by an electric step, for which the nonperturbative formulation of strong-field QED exists.[34,35] To employ this formulation, we first find exact solutions of the DP equation with Sauter-like magnetic field with well-defined spin polarization and calculated all quantities characterizing the effect, in particular when the field lies in two specific configurations. When the field varies "gradually" along the inhomogeneity direction, we found that the flux density of created neutral fermion pairs is quadratic in the magnitude of the step $\mathbb{U}$. This feature is particularly different from the case of charged pair production by electric steps, in which the the flux density features a linear dependence on the magnitude of the electric step. The quadratic dependence for neutral fermions derives from the non-cartesian geometry of the parameter space formed by the quantum numbers, and it is inherent to the dynamics of neutral fermions with anomalous magnetic moments in inhomogeneous magnetic

fields. This also explains why the flux density of created pairs per unit of the length are not uniform. In particular, it means that the Schwinger method of the effective action works for the case under consideration only after a suitable parameterization. The second feature worth discussing is the behavior of total quantities when the field "sharply" varies. It is exactly the form of the Klein effect.[39] If we compare the flux density of neutral fermion pairs created with the total number of electron-positron pairs created from the vacuum by inhomogeneous electric fields (given, for example, by Eq. (88) with $d = 4$ in[42]), we observe two major differences: the first is the presence of a logarithmic coefficient $\left|\ln\left(1-\gamma^2\right)\right|$, that can be traced back to the integration over p_z (53) and therefore does not depend on the external field. To our knowledge, this term has no precedents in QED (although a logarithmic coefficient of this type may appear in scalar QED). The second, and more important, is the value of the scaling (or critical) exponent seen in (55). In contrast to QED in $3+1$ dimensions, in which $N^{\mathrm{cr}} \sim \left(1-\gamma^2\right)^3$,[40–42] the total number of neutral fermions pairs created from the vacuum features a larger exponent, $7/2$. Aside from minor numerical differences, this means that the total number (55) has an extra term $\sqrt{1-\gamma^2}\left|\ln\left(1-\gamma^2\right)\right|$, which is always less than unity in the range of values to γ within the interval $0 \leq \gamma < 1$. Formally, this indicates that backreaction effects caused by neutral fermions produced by sharply-evolving inhomogeneous magnetic fields may be significantly smaller compared to QED under equivalent conditions.

The mechanism here described raises the question about the critical magnetic field intensity, near which the phenomenon could be observed. It is possible to estimate such a value based on fermion's mass and its magnetic moment. Since $\max B_z(y) = B_z(+\infty) = \varrho B' \equiv B_{\max}$, the nontriviality of the Klein zone (29) yields the following condition

$$\mathbb{U} = 2\left|\mu\right|\varrho B' > 2m \Rightarrow B_{\max} > B_{\mathrm{cr}}\,, \quad B_{\mathrm{cr}} \equiv \frac{m}{|\mu|} \approx 1.73\times 10^8 \times \left(\frac{m}{1\,\mathrm{eV}}\right)\left(\frac{\mu_{\mathrm{B}}}{|\mu|}\right)\mathrm{G}\,, \tag{56}$$

where $\mu_{\mathrm{B}} = e/2m_e \approx 5.8\times 10^{-9}\,\mathrm{eV/G}$ is the Bohr magneton.[16] For neutrons, whose mass and magnetic moment are $m_N \approx 939.6\times 10^6\,\mathrm{eV}$, $\mu_N \approx -1.042\times 10^{-3}\mu_{\mathrm{B}}$, the critical magnetic field (56) is $B_{\mathrm{cr}} \approx 1.56\times 10^{20}\,\mathrm{G}$. More optimistic values can be estimated for neutrinos because of their light masses and small magnetic moments. For example, considering recent constraints for neutrinos effective magnetic moment $\mu_\nu \approx 2.9\times 10^{-11}\mu_{\mathrm{B}}$[18] and mass $m_\nu \approx 10^{-1}\,\mathrm{eV}$,[30] we find $B_{\mathrm{cr}} \approx 5.97\times 10^{17}\,\mathrm{G}$. Evidently, this value changes considering different values to neutrinos' magnetic moment and mass. Taking, for instance, the experimental estimate to the tau-neutrino magnetic moment $\mu_\tau \approx 3.9\times 10^{-7}\mu_{\mathrm{B}}$[17] and assuming its mass $m_{\nu_\tau} \approx 10^{-1}\,\mathrm{eV}$ we obtain a value to B_{cr} near QED critical field $B_{\mathrm{QED}} = m^2/e \approx 4.4\times 10^{13}\,\mathrm{G}$, namely $B_{\mathrm{cr}} \approx 4.44\times 10^{13}\,\mathrm{G}$. On the other hand, assuming the lower bound found in Ref.[25] $\mu_\nu \approx 10^{-14}\mu_{\mathrm{B}}$ and the same mass $m_\nu \approx 10^{-1}\,\mathrm{eV}$ we obtain a value to B_{cr} orders of magnitude larger than B_{QED}, $B_{\mathrm{cr}} \approx 1.73\times 10^{21}\,\mathrm{G}$. The critical magnetic field surprisingly increases if one considers the magnetic moment predicted by the SM,

$\mu_\nu \approx 3.2 \times 10^{-19} \mu_B \times (m_\nu/1\,\mathrm{eV})$.[27,30] Substituting this value into (56) and considering $m_\nu \approx 1\,\mathrm{eV}$ we find $B_{cr} \approx 5.41 \times 10^{26}\,\mathrm{G}$. Based on these estimates, we believe that neutral fermion pair production may occur in astrophysical enviroments, in particular during a supernova explosion or in the vicinity of magnetars, whose typical order of magnetic field intensities range from $10^{16} - 10^{18}\,\mathrm{G}$ (or up to 10^{20} G), as reported in Refs.[43–46].

Acknowledgments

T.C.A. was supported by the Advanced Talents Development Program of the Hebei University, grant no. 801260201271. Z.-W.H. was supported by the Post-graduate's Innovation Fund Project of Hebei Province, grant No. CXZZSS2021016. The work of S.P. Gavrilov and D.M. Gitman was supported by Russian Science Foundation, grant no. 19-12-00042.

References

1. J. Schwinger, Phys. Rev. **82**, 664 (1951).
2. O. Klein, Z. Phys. **53**, 157 (1929).
3. F. Sauter, Z. Phys. **69**, 742 (1931).
4. F. Sauter, Z. Phys. **73**, 547 (1932).
5. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936).
6. W. Greiner, B. Müller, and J. Rafelski, *Quantum Electrodynamics of Strong Fields* (Springer-Verlag, Berlin, 1985).
7. A. A. Grib, S. G. Mamaev and V. M. Mostepanenko, *Vacuum Quantum Effects in Strong Fields* (Friedmann Laboratory, St. Petersburg, 1994).
8. E. S. Fradkin, D. M. Gitman, and S. M. Shvartsman, *Quantum Electrodynamics with Unstable Vacuum* (Springer-Verlag, Berlin, 1991).
9. G. V. Dunne, *Heisenberg-Euler Effective Lagrangians: Basics and Extensions*, in I. Kogan Memorial Volume, From Fields to Strings: Circumnavigating Theoretical Physics, pp. 445-522, edited by M Shifman, A. Vainshtein, and J. Wheater (World Scientific, Singapore, 2005).
10. R. Ruffini, G. Vereshchagin, and S. Xue, Phys. Rep. **487**, 1 (2010).
11. G. Dunne, Eur. Phys. J. D **55**, 327 (2009).
12. F. Gelis and N. Tanji, Prog. Part. Nucl. Phys. **87**, 1 (2016).
13. A. Di Piazza, C. Müller, K. Z. Hatsagortsyan, and C. H. Keitel, Rev. Mod. Phys. **84**, 1177 (2012).
14. B. M. Hegelich, G. Mourou, and J. Rafelski, Eur. Phys. J. Spec. Top. **223**, 1093 (2014).
15. T. C. Adorno, S. P. Gavrilov, and D. M. Gitman, Int. J. Mod. Phys. **32**, 1750105 (2017).
16. The Nist Reference on Constants, Units, and Uncertainty `https://physics.nist.gov/cuu/Constants/index.html`
17. Schwienhorst, R., et al. (DONUT Collaboration), Phys. Lett. B **513**, 23 (2001).
18. A. Beda et al., Adv. High Energy Phys. **2012**, 350150 (2012).
19. G. G. Raffelt, Phys. Rev. Lett. **64**, 2856 (1990).
20. G. G. Raffelt, Astrophys. J. **365**, 559 (1990)
21. G. G. Raffelt and A. Weiss, Astron. Astrophys. **264**, 536 (1992).
22. V. Castellani and S. Degl'Innocenti, Astrophys. J. **402**, 574 (1993).

23. M. Catelan, J. F. Pacheco, and J. Horvath, Astrophys. J. **461**, 231 (1996).
24. N. Viaux, M. Catelan, P. B. Stetson, G. G. Raffelt, J. Redondo, A. A. R. Valcarce, and A.Weiss, Astron. Astrophys. **558**, A12 (2013).
25. N. F. Bell, V. Cirigliano, M. J. Ramsey-Musolf, P. Vogel, and M. B. Wise, Phys. Rev. Lett. **95**, 151802 (2005).
26. A. Aboubrahim, T. Ibrahim, A. Itani, and P. Nath, Phys. Rev. D **89**, 055009 (2014).
27. C. Giunti and A. Studenikin, Phys. Atom. Nucl. **72**, 2089 (2009).
28. M. Dvornikov, in *Neutrinos: Properties, Sources and Detection*, edited by J. P. Greene (Nova Science Publishers, New York, 2011), p. 23.
29. C. Broggini, C. Giunti, and A. Studenikin, Adv. High Energy Phys. **2012**, 459526 (2012).
30. C. Giunti and A. Studenikin, Rev. Mod. Phys. **87**, 531 (2015).
31. C. Giunti, K. A. Kouzakov, Y-F. Li, A. V. Lokhov, A. I. Studenikin, and S. Zhou, Ann. Phys. **528**, 198 (2016).
32. T. C. Adorno, Z.-W. He, S. P. Gavrilov and D. M. Gitman, arXiv:2109.06053.
33. S. P. Gavrilov and D. M. Gitman, Phys. Rev. D **87**, 125025 (2013).
34. S. P. Gavrilov and D. M. Gitman, Phys. Rev. D **93**, 045002 (2016).
35. S. P. Gavrilov and D. M. Gitman, Eur. Phys. J. C **80**, 820 (2020).
36. W. Pauli, Rev. Mod. Phys. **13**, 203 (1941).
37. V. G. Bagrov and D. M. Gitman, *The Dirac Equation and Its Solutions* (De Gruyter, Berlin, 2014).
38. A. Erdelyi *et al.* (Ed.), *Higher Transcendental Functions* (*Bateman Manuscript Project*), Vol. 1 (McGraw-Hill, New York, 1953).
39. N. Dombey and A. Calogeracos, Phys. Rep. **315,** 41 (1999).
40. H. Gies, G. Torgrimsson, Phys. Rev. Lett. **116**, 090406 (2016).
41. H. Gies, G. Torgrimsson, Phys. Rev. D **95**, 016001 (2017).
42. T. C. Adorno, S. P. Gavrilov and D. M. Gitman, Eur. Phys. J. C **80**, 88 (2020).
43. D. Lai, Rev. Mod. Phys. **73**, 629 (2001).
44. S. Akiyama, J. C. Wheeler, D. L. Meier, and I. Lichtenstadt, Astrophys. J. **584**, 954 (2003).
45. S. Mereghetti, Astron. Astrophys. Rev. **15**, 225 (2008).
46. E. J. Ferrer, V. de la Incera, J. P. Keith, I. Portillo, and P. Springsteen, Phys. Rev. C **82**, 065802 (2010); L. Paulucci, E. J. Ferrer, V. de la Incera, and J. E. Horvath, Phys. Rev. D **83**, 043009 (2011).

On the magnetic field screening in strong crossed electromagnetic field

S. Campion

ICRA, Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy
E-mail: stefano.campion@icranet.org
www.phys.uniroma1.it/fisica

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy

Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy

J. A. Rueda Hernandez

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy,
E-mail: jorge.rueda@icra.it

ICRA, Dipartimento di Fisica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, Roma, 00185, Italy

ICRANet-Ferrara, Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, Ferrara, 44122, Italy

Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Saragat 1, Ferrara, 44122, Italy

INAF, Istituto de Astrofisica e Planetologia Spaziali, Via Fosso del Cavaliere 100, Rome, 00133, Italy

S.-S. Xue

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy,
E-mail: xue@icra.it

ICRA, Dipartimento di Fisica, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, Roma, 00185, Italy

R. Ruffini

International Center for Relativistic Astrophysics Network, Piazza della Repubblica 10, Pescara, 65122, Italy
E-mail: ruffini@icra.it

ICRA, Dipartimento di Fisica, Sapienza Università di Roma, P.le Aldo Moro 5, Rome, 00185, Italy

Université de Nice Sophia Antipolis, CEDEX 2, Grand Château Parc Valrose, Nice, France

ICRANet-Rio, Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, Rio de Janeiro, 22290–180, Brazil

INAF, Viale del Parco Mellini 84, Rome, 00136, Italy

We study the screening of a strong magnetic field operated by an initial huge number of $e^{\pm}$ pairs (we do not discuss here their production mechanism). The background fields configuration is of crossed fields, $\left(\vec{B} = B\,\hat{z},\ \vec{E} = E\,\hat{y}\right)$, with $E/B < 1$. In this system the following series of processes occur: 1) the electric field accelerates the pairs, which radiate high-energy synchrotron photons; 2) these synchrotron photons interact with the background magnetic field via the magnetic pair production process (MPP hereafter), i.e. $\gamma + B \rightarrow e^{+} + e^{-}$, producing additional pairs; 3) the dynamic of all the pairs around the magnetic field lines generates a current that induces a magnetic field oriented in the opposite direction to the background one and then shielding it. We get that, for instance, for an initial number of pairs $N_{\pm,0} = 10^{10}$, an initial magnetic field of 10^{12} G can be reduced by a few percent. The whole screening process described by the steps above, occurs in the short timescales $10^{-21} \leq t \leq 10^{-15}$ s, i.e. the time necessary before the particles acceleration timescale equals the synchrotron cooling timescale. Further developments (as the study of this mechanism in different geometries of the $\vec{E}$ and $\vec{B}$ fields, quantum effects in overcritical fields, other mechanisms for the production, distribution and multiplicity of the $e^{\pm}$ pairs) are necessary in order to apply this model to specific and extreme astrophysical systems (as Black Hole or Neutron Star).

Keywords: Magnetic Pair Production; Magnetic field screening; Shower processes

1. Introduction

For many years the screening of a strong electric field, operated by $e^{\pm}$ pairs created by QED processes, has been studied (e.g. see recently Ref. 1[a]). No analogous work and conclusions have been reached for a magnetic field. The main topic of this paper is to build a simple model to analyze the magnetic field screening (MFS hereafter) process owing to the motion of $e^{\pm}$ pairs in a region filled by magnetic $\vec{B}$ and electric $\vec{E}$ fields, oriented in a crossed configuration.

The basic idea of the screening process is described by the following series steps:

(1) An initial number of $e^{\pm}$ is placed in a region filled by $\vec{E}$ and $\vec{B}$, with $E/B < 1$, $B \leq B_{\rm cr} = m_e^2 c^3/(e\hbar) \approx 4.4 \times 10^{13}$ G and then $E < E_{\rm cr}$. We do not discuss here the creation process of these initial pairs (that could be, for example, the vacuum breakdown).

[a]In this paper it is shown that an electric field as high as $E \sim \alpha_f E_{\rm cr}$, with α_f the fine structure constant and $E_{\rm cr} = m_e^2 c^3/(e\hbar) \approx 1.32 \times 10^{16}$ V/cm the critical field for vacuum polarization, cannot be maintained since the creation of particle showers depletes the field

(2) The initial pairs are accelerated by $\vec{E}$ and emit radiation via the curvature/synchrotron mechanism (or their combination), due to the $\vec{B}$ field. We assume that the electric and magnetic field strength are related by the following relation

$$E(t) = \Upsilon\ B(t), \tag{1}$$

with $0 < \Upsilon \leq 1$, since we are interested in conditions of magnetic dominance.

(3) The photons create a new $e^{\pm}$ pairs via the magnetic pair production process (MPP), $\gamma + B \rightarrow e^- + e^+$.

(4) Also these new pairs are accelerated and radiate synchrotron photons. All the pairs, initial + created by the MPP process, circularize around the magnetic field lines generating a current that gives rise to an induced magnetic field, $\vec{B}_{\rm ind}$, oriented in the opposite direction with respect to the original one, thereby screening it. Because of the creation of these new charged particles and the proportionality between the strength of the fields, also the electric field is screened.

(5) This series of the processes occurs at every time t leading to the development of a particles shower.

This model can find useful application in extreme astrophysical systems, as GRBs. Indeed, following the "*inner engine*" model for GRBs[2,3] within the binary-driven hypernova (BdHN) framework (see, e.g., Refs. 4–6), these type of systems are composed by a rotating back hole, surrounded by the magnetic field and low-density ionized matter. The gravitomagnetic interaction of the rotating BH and the magnetic field induces an electric field which accelerates e^- which emit GeV photons by synchrotron radiation.[2] Then a decrease of the magnetic field leads to the decrease of the optical depth for synchrotron photons, which can reach the threshold to freely escape from the region near the BH and become observable. Therefore, from this brief introduction of the model it is easy to understand that the physical process that we present here could be necessary for the explanation of some emission in extreme astrophysical systems.

2. Model equations

Here we analyze the whole screening process for the specific configuration of perpendicular fields: $\vec{E} = E\ \hat{y}$ for the electric field and $\vec{B} = B\ \hat{z}$ for the magnetic field. In this section we just summarize the fundamental equations that describe the system. The entire set of equations and their derivation can be found in Refs. 7, 8.

In order to follow the entire screening process we have built a set of ordinary differential equations which describe, at the same time: 1) the motion of the particles; 2) the creation of new particles via the energy loss and the MPP processes; 3) the decrease of the magnetic field due to the generation of an induced magnetic field by the particles currents.

The equations of motion of a particle immersed in an EM field [b] read (see, e.g., Refs. 9, 10)

$$\frac{dp^{\alpha}}{d\tau} = m\frac{du^{\alpha}}{d\tau} = \frac{e}{c}\mathcal{F}^{\alpha\beta}u_{\beta}. \tag{2}$$

and for the radiation emitted we use the energy loss formula in quantum regime, $I = |-dE/dt|$, given in Ref. 11.

In this paper they use the parameter $\chi \equiv \varepsilon_*/2\,\varepsilon_e$ to distinguish between emission in quantum ($\chi \gtrsim 1$) and classical ($\chi < 1$) regimes, where $\varepsilon_e = \gamma m_e c^2$ is the electron energy and $\varepsilon_* = \hbar\omega^* = (3e\gamma^2/2m_e c)\sqrt{\left(\vec{E}+\vec{\beta}\times\vec{B}\right)^2 - \left(\vec{\beta}\cdot\vec{E}\right)^2}$ the critical photons energy.

The photons production rate, $\dot{N}_{\gamma}(t,\phi)$ (with ϕ the angle between the particle/photon direction and the magnetic field) can be calculated as the product between the number of pairs present at the specific time t, $N_{\pm}(t,\phi)$ and the rate of photons production via synchrotron emission, $I/\varepsilon_{\gamma}^{e}$ (with ε_{γ}^{e} the energy of synchrotron photons emitted by the $e^{\pm}$). The equation for the evolution of the number of pairs is strictly related to the one for the number of photons. Indeed, the rate of pair production $\dot{N}_{\pm}(t,\phi)$ is given by the product of the photon number N_{γ}, the attenuation coefficient for the MPP process $R_{\mathrm{A}}^{e}(t,\phi)$ times the velocity of light c.

The motion of a particle inside an electromagnetic field can be considered as the combination of an acceleration along the $z-$direction, and of a circular motion in the $x-y$ plane, around the magnetic field lines. This circular motion of the pairs produces coils traversed by the electric current. This current creates an induced magnetic field, from the Biot–Savart law, whose evolution with time is described by the following formula (calculated at the center of the coil):

$$\frac{dB_{\mathrm{z,\ ind}}}{dt} = e\frac{\beta_{\perp}(t)}{R_c(t)^2}\frac{dN_{\pm}}{dt}. \tag{3}$$

with $\beta_{\perp} = (\,\beta_x^2 + \beta_y^2\,)^{1/2}$ the perpendicular component of the particle velocity and $R_c^{-1} = \left|d\vec{\beta}/cdt\right| = \left(e/\gamma mc^2\right)\sqrt{\left(\vec{E}_{\mathrm{tot}}+\vec{\beta}\times\vec{B}_{\mathrm{tot}}\right)^2 - \left(\vec{\beta}\cdot\vec{E}_{\mathrm{tot}}\right)^2}$ the curvature radius for the particle trajectory.

We have introduced the attenuation coefficient for the magnetic pair production and we have called it R_A^e. From now on, we refer to as $\zeta \equiv R_A^e c$ the MPP rate. In Ref. 12, the expression for the pair production rate has been derived in the observer frame at rest, for strong perpendicular electric and magnetic fields ($\vec{E}\cdot\vec{B} = 0$). In order to get the expression of the MPP rate in our frame we have: 1) moved to a new K' frame, where no electric field is present, through a Lorentz transformation; 2) calculated all the necessary quantities (energy, momentum and director cosines of the photons, the component of the magnetic field perpendicular to the propagation direction of the photons, the MPP rate); 3) transforming back the rate to the

[b] We use throughout cgs-Gaussian units, in which the magnetic and electric fields share the same dimensions ($\mathrm{g}^{1/2}\ \mathrm{cm}^{-1/2}\ \mathrm{s}^{-1}$) and a -2 signature so the spacetime metric is $\eta_{\mu\nu} = (1,-1,-1,-1)$.

original frame. The expression for the MPP rate in the laboratory frame (observer at infinity, the frame where we work) assumes the following form (further details on this procedure can be found in Refs. 7, 8):

$$\zeta = 0.23 \frac{\alpha_f c}{\lambda\!\!\!^{-}_c} \frac{B_z}{B_{\rm cr}} \left(1 - \frac{E_y^2}{B_z^2}\right) \frac{\sqrt{\eta_y^2 \left(1 - \frac{E_y^2}{B_z^2}\right) + \left(\eta_x - \frac{E_y}{B_z}\right)^2}}{1 - \frac{E_y}{B_z}\eta_x} \times \exp\left\{-\frac{8}{3}\frac{mc^2}{\varepsilon_\gamma}\frac{B_{\rm cr}}{B_z}\left[\eta_y^2\left(1 - \frac{E_y^2}{B_z^2}\right) + \left(\eta_x - \frac{E_y}{B_z}\right)^2\right]^{-1/2}\right\}, \tag{4}$$

which is valid until the following condition is satisfied

$$\Psi = \frac{3}{4}\frac{e\hbar}{m\,c}\frac{B^2}{B_{\rm cr}}\gamma^2 \sqrt{{\beta_y}^2\left(1 - \frac{E^2}{B^2}\right) + \left(\frac{E}{B} - \beta_x\right)^2}\sqrt{{\eta_y}^2\left(1 - \frac{E^2}{B^2}\right) + \left(\eta_x - \frac{E}{B}\right)^2} \ll 1. \tag{5}$$

In Eq. (4) and Eq. (5) the quantities $\vec{\eta}$ are the photon director cosines, whose expression as a function of the electron velocity, polar and azimuthal angle of their emission in the comoving frame can be found in Ref. 7.

3. Results

We now present the results of the numerical integration of the set of equations described in the previous section. We have selected three values for the parameter Υ = (1/2, 1/10, 1/100). Since the proportionality between the fields is requested at any time, when $B(t)$ changes even $E(t)$ changes accordingly to keep Υ constant. These combination of these effects affect the motion of particles and, consequently, all the related processes that give rise to the screening.

The condition on Ψ in Eq. (5) brings with it three initials conditions for the background magnetic field, the electron Lorentz factor and for the emission direction of the pairs $\left(B_0,\ \gamma_0,\ \vec{\beta}_0\right)$. A further condition is necessary for the director cosines of the photons $\vec{\eta}$. Then, we need to choose the right values for these variables and parameters in order to apply the equation for the rate

We select three emission directions of the particles: 1) along the $\hat{y}$−axis; 2) along the $\hat{z}$−axis; 3) along a direction characterized by polar and azimuth angles, respectively, $\theta = 75°$ and $\phi = 30°$ (hereafter we refer to this direction as "*generic*" or "*G*"). For each direction, we have chosen the initial value of the magnetic field B_0 and, consequently, the maximum value of particles Lorentz factor γ_0, which satisfy Eq. (5). After we have integrated the set of equations that describe the entire screening process considering different number of emitted particles $N_{\pm,0} = 1,\ 10^3,\ 10^6,\ 10^{10}$ and photons $N_{\gamma,0} = 0$; $N_{\gamma,0} = 10^3$, with $N_{\pm,0} = 1$. Each numerical integration starts at $t_0 = 10^{-21}$ s and stops when the Lorentz factor of the pairs becomes 1, i.e. when the particle has lost all of its energy, at $t_f \sim 10^{-17} - 10^{-15}$ s.

In Table 1, Table 2, Table 3 we show the results we have obtained from the integrations of our set of equations describing, for all the considered initial conditions

Table 1. Results for the integrations of the set of equations for the different initial conditions, with $\Upsilon = 1/2$. **Column (1):** initial emission direction for particles; **Column (2):** initial value of the magnetic field strength; **Column (3):** initial value of the Lorentz factor; **Column (4):** percentage variation of the magnetic field (calculated as $\left[\left(B(t_0) - B\left(t_f\right)\right)/B(t_0)\right] \times 100$); **Column (5):** initial number of emitted particles; **Column (6):** final number of created particles; **Column (7):** initial number of emitted photons; **Column (8):** final number of emitted photons.

Direction	$B_0(B_{cr})$	γ_0	$\Delta B(\%)$	$N_{\pm,0}$	$N_{\pm,f}$	$N_{\gamma,0}$	$N_{\gamma,f}$
Generic	0.1	6.48	$-4.7 \times 10^{-12}\%$	1	1	0	3.5
			$-4.7 \times 10^{-9}\%$	10^3	10^3	0	3500
			$-4.7 \times 10^{-6}\%$	10^6	$\gtrsim 10^6$	0	3.5×10^6
			$-4.66 \times 10^{-2}\%$	10^{10}	$\gtrsim 10^{10}$	0	3.5×10^{10}
			$-6.3 \times 10^{-8}\%$	1	1	10^3	1003
	0.3	2.27	$-1.34 \times 10^{-10}\%$	1	1	0	0.45
			$-1.34 \times 10^{-7}\%$	10^3	10^3	0	447
			$-1.34 \times 10^{-4}\%$	10^6	10^6	0	4.48×10^5
			-1.09%	10^{10}	$\gtrsim 10^{10}$	0	4.46×10^9
			$-1.46 \times 10^{-6}\%$	1	1	10^3	1000.4
Along y	0.1	3.66	$-2.25 \times 10^{-10}\%$	1	1	0	2.79
			$-2.25 \times 10^{-7}\%$	10^3	10^3	0	2798
			$-2.2 \times 10^{-4}\%$	10^6	10^6	0	2.8×10^6
			-1.72%	10^{10}	$\gtrsim 10^{10}$	0	2.84×10^{10}
			$-1.7 \times 10^{-6}\%$	1	1	10^{-3}	1002
Along z	0.1	7.098	$-1 \times 10^{-12}\%$	1	1	0	11.3
			$-1 \times 10^{-9}\%$	10^3	10^3	0	1.13×10^4
			$-1 \times 10^{-6}\%$	10^6	10^6	0	1.13×10^7
			-0.01%	10^{10}	$\gtrsim 10^{10}$	0	1.13×10^{11}
			$-1.5 \times 10^{-8}\%$	1	1	10^3	1011
	0.3	2.14	$-1.06 \times 10^{-10}\%$	1	1	0	2.33
			$-1.6 \times 10^{-7}\%$	10^3	10^3	0	2333
			$-1.6 \times 10^{-4}\%$	10^6	10^6	0	2.3×10^6
			-1.27%	10^{10}	$\gtrsim 10^{10}$	0	2.4×10^{10}
			$-1.5 \times 10^{-6}\%$	1	1	10^3	1002

(emission direction, B_0, γ_0, $N_{\pm,0}$, $N_{\gamma,0}$). We report the percentage variation of the magnetic field $\Delta B(\%)$, the final number of created pairs $N_{\pm,f}$ and photons $N_{\gamma,f}$. In Table 1 the results are for $\Upsilon = 1/2$, Table 2 is for $\Upsilon = 1/10$ and Table 3 for $\Upsilon = 1/100$. From Table 1, Table 2 and Table 3, we can get some interesting results. Indeed, we can see that the model (namely the system of equations we have built) is strictly correlated to the initial conditions and, in particular, to the particles emission direction and to the strength of the background magnetic field. A sufficient decrease of $\vec{B}$ is obtained when a huge initial number of particles $N_{\pm,0} \gtrsim 10^{10}$ is injected. For particles emitted along the $\hat{z}-$axis and along the *generic* direction, the stronger the background field, the higher the reduction of the field itself is. Even when they are emitted along the $\hat{y}-$axis, there is a significant reduction of $\vec{B}$, but this

Table 2. Same results of Table 1, but for $\Upsilon = 1/10$.

Direction	$B_0(B_{\rm cr})$	γ_0	$\Delta B(\%)$	$N_{\pm,0}$	$N_{\pm,f}$	$N_{\gamma,0}$	$N_{\gamma,f}$
Generic	0.1	4.18	$-3.2\times10^{-11}\%$	1	1	0	4.63
			$-3.2\times10^{-8}\%$	1000	10^3	0	4633
			$-3.2\times10^{-5}\%$	10^6	10^6	0	4.63×10^6
			-0.308%	10^{10}	$\gtrsim 10^{10}$	0	4.62×10^{10}
			$-4.35\times10^{-7}\%$	1	1	10^3	1004
Along y	0.1	3.71	$-9.14\times10^{-11}\%$	1	1	0	4.45
			$-9.14\times10^{-8}\%$	10^3	10^3	0	4457
			$-9.14\times10^{-5}\%$	10^6	10^6	0	4.46×10^6
			-0.798%	10^{10}	$\gtrsim 10^{10}$	0	4.42×10^{10}
			$-9.49\times10^{-7}\%$	1	1	10^{-3}	1004
Along z	0.1	22.66	0%	1	1	0	64
			0%	10^3	10^3	0	6.4×10^4
			$-2.2\times10^{-14}\%$	10^6	10^6	0	6.4×10^7
			$-4.2\times10^{-11}\%$	10^{10}	10^{10}	0	6.4×10^{11}
			0%	1	1	10^3	1063
	0.3	10.04	$-1.13\times10^{-12}\%$	1	1	0	19.6
			$-1.15\times10^{-9}\%$	10^3	10^3	0	1.96×10^4
			$-1.15\times10^{-6}\%$	10^6	10^6	0	1.96×10^7
			$-1.15\times10^{-2}\%$	10^{10}	$> 10^{10}$	0	1.96×10^{11}
			$-8\times10^{-9}\%$	1	1.0025	10^3	1019
	0.5	4.8	$-1.75\times10^{-11}\%$	1	1	0	10.6
			$-1.75\times10^{-8}\%$	10^3	10^3	0	1.05×10^4
			$-1.75\times10^{-5}\%$	10^6	10^6	0	1.05×10^7
			-0.171%	10^{10}	$> 10^{10}$	0	1.05×10^{11}
			$-8.3\times10^{-8}\%$	1	1.0036	10^3	1010
	0.7	2.63	$-8.3\times10^{-11}\%$	1	1	0	5.8
			$-8.3\times10^{-8}\%$	10^3	10^3	0	5811
			$-8.3\times10^{-5}\%$	10^6	10^6	0	5.81×10^6
			-0.734%	10^{10}	$> 10^{10}$	0	5.93×10^{10}
			$-2.87\times10^{-7}\%$	1	1.0027	10^3	1005

occurs with a lower strength of the field. These characteristics can be understood by looking on how the fundamental ingredients of the model $\left(\vec{B},\ \vec{\beta},\ \vec{\eta},\ N_{\pm}\right)$ enter in the equations and, in particular, looking at the dependence of the rate $\zeta(t)$ in Eq. (4) on $B(t)$ and $\vec{\eta}$.

In Fig. 1 is shown the variation of the parameter Ψ in Eq. (5), for three considered values of Υ and for the three directions of emission (**generic**, $\hat{y}$**-axis** and $\hat{z}$**-axis**), with $B_0 = 0.1\ B_{\rm cr}$ and $N_{\pm,\,0} = 10^6$. It is evident that the condition in Eq. (5) is well satisfied for all the analyzed cases and for all the integration time. In Fig. 2 we present the result for the screening of the magnetic field when an initial number of particles $N_{\pm,0} = 10^{10}$ is emitted along the: 1) *generic* direction with $\gamma_0 = 2.27$; 2) $\hat{z}$-axis with $\gamma_0 = 2.14$, with initial strength of the background magnetic field $B_0 = 0.3\ B_{\rm cr}$ and $\Upsilon = 1/2$.

Table 3. Same results of Table 1, but for $\Upsilon = 1/100$.

Direction	$B_0(B_{cr})$	γ_0	$\Delta B(\%)$	$N_{\pm,0}$	$N_{\pm,f}$	$N_{\gamma,0}$	$N_{\gamma,f}$
Generic	0.1	3.81	$-2.5\times10^{-11}\%$	1	1	0	5.15
			$-2.5\times10^{-8}\%$	10^3	10^3	0	5153
			$-2.5\times10^{-5}\%$	10^6	10^6	0	5.15×10^6
			-0.237%	10^{10}	$\gtrsim 10^{10}$	0	5.14×10^{10}
			$-3.4\times10^{-7}\%$	1	1	10^3	1005
Along y	0.1	3.71	$-8\times10^{-11}\%$	1	1	0	4.91
			$-8\times10^{-8}\%$	10^3	10^3	0	4912
			$-8\times10^{-5}\%$	10^6	10^6	0	4.91×10^6
			-0.711%	10^{10}	$\gtrsim 10^{10}$	0	5.02×10^{10}
			$-8.8\times10^{-7}\%$	1	1	10^{-3}	1004
Along z	0.1	22.66	0%	1	1	0	22.23
			0%	10^3	10^3	0	2.22×10^4
			0%	10^6	10^6	0	2.22×10^7
			0%	10^{10}	10^{10}	0	2.22×10^{11}
			0%	1	1	10^3	1022
	1.1	10.04	$-1.6\times10^{-14}\%$	1	1	0	26.8
			$-7.8\times10^{-12}\%$	10^3	10^3	0	1.49×10^4
			$-7.8\times10^{-9}\%$	10^6	10^6	0	1.77×10^7
			$-7.5\times10^{-5}\%$	10^{10}	$\gtrsim 10^{10}$	0	1.75×10^{11}
			$-2.7\times10^{-11}\%$	1	1.0022	10^3	1026
	2.5	2.13	$-5.7\times10^{-12}\%$	1	1	0	35.3
			$-5.4\times10^{-9}\%$	10^3	10^3	0	3.53×10^4
			$-5.6\times10^{-6}\%$	10^6	$\gtrsim 10^6$	0	3.53×10^7
			$-5.5\times10^{-2}\%$	10^{10}	$> 10^{10}$	0	3.52×10^{11}
			$-1.3\times10^{-8}\%$	1	1.025	10^3	1036

Figure 3, instead, shows the emitted photons energy for $N_{\pm,\,0} = 10^{10}$ particles emitted along the *generic* direction, for the three considered values of $\Upsilon = 1/2,\ 1/10,\ 1/100$ and for $B_0 = 0.1\ B_{cr}$.

4. Conclusions

In this work, we have built a simplified model to study the magnetic field screening process operated by a huge number of $e^{\pm}$ pairs in presence of a strong crossed electromagnetic field. The model we have built is based on two principal assumptions:

(1) We have constructed one-particle equations to describe the particles motion as a fluid. This assumption is justified considering the dynamics of the entire process and the initial conditions (the particles obey to the continuity equation and are forced to follow the same trajectory due to the high strength of the fields).
(2) Due to the proportionality between the magnetic and electric field, also the electric field is screened. This effect can be justified considering that the creation of new charged particles, via the MPP and synchrotron processes, leads to the

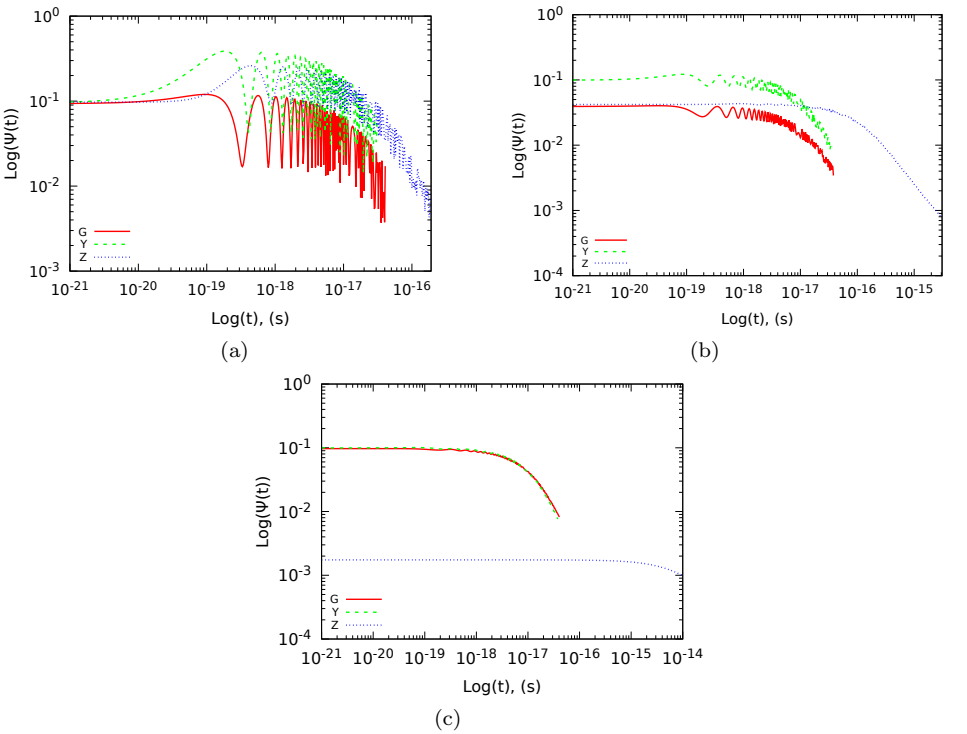


Fig. 1. Variation of the Ψ parameter, calculated for the case ϒ = 1/2 (a), 1/10 (b), 1/100 (c), for the three directions of emission (**generic**, **ŷ-axis** and **ẑ-axis**) and for an initial magnetic field strength of $B_0 = 0.1\ B_{cr}$.

formation of a current which screens the electric field too. We address to future work a more detailed treatment of this process.

We briefly summarized here the results that we got from our research. Further and more complete details can be found in Refs. 7, 8. We have obtained that the screening process occurs and the significance of the process (and, then, of the reduction of the field) depends by

(I) the number of initial pairs injected in the system: the screening is more efficient if one increases $N_{\pm,0}$ from 10^6 to 10^{10}–10^{15};
(II) the initial particles emission direction: the major effect is obtained when the particles are emitted with a higher perpendicular component of their velocity $\beta_\perp$ (namely along the *generic* and ŷ-axis directions);
(III) the parameter ϒ. For particle with lower $\beta_\perp$ (*generic* direction) a decrease of ϒ leads to a stronger screening effect. Instead, for particles with a higher $\beta_\perp$ (namely emitted along the ŷ-axis), the screening increases if ϒ decreases.

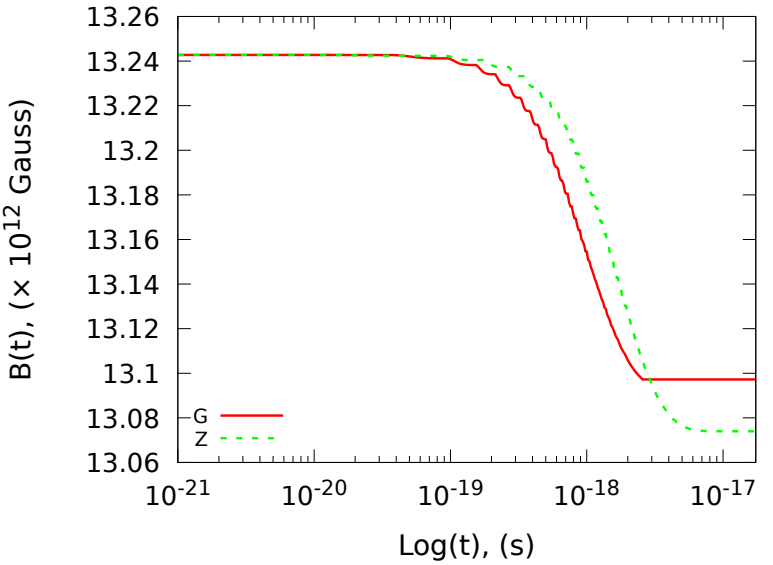


Fig. 2. Magnetic field decrease due to an initial number of particles, $N_{\pm,0} = 10^{10}$, emitted initially along the *generic* direction (with $\gamma_0 = 2.27$) and along the $\hat{z}$-axis (with $\gamma_0 = 2.14$). Here $\Upsilon = 1/2$ and $B_0 = 0.3\ B_{\rm cr}$.

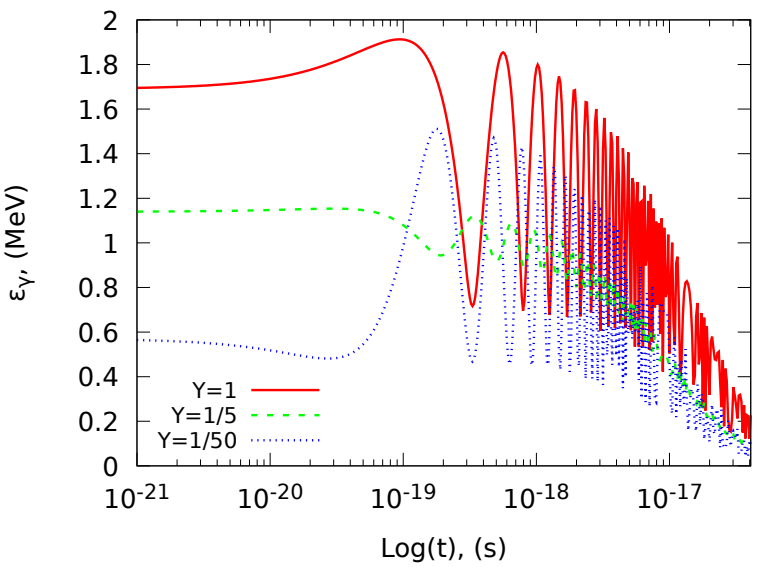


Fig. 3. Emitted photons energy, as a function of time, for $N_{\pm,0} = 10^{10}$ emitted along the *generic* direction, with $B_0 = 0.1\ B_{\rm cr}$, for $\Upsilon = 1/2,\ 1/10,\ 1/100$.

We conclude that, under the studied physical conditions, the screening effect occurs and leads to a reduction of the field up to a few percent (for the specific initial conditions considered in this work). This implies that this process needs to be taken into account in the treatment of particles dynamics in extreme astrophysical systems, as pulsars and gamma-ray bursts. This study has been the first one on this subject and, in order to apply it to real physical systems, some improvements are necessary. Future development of this work concern its extension to different configurations between the electric and magnetic fields, the extension of the treatment to the case of overcritical fields (where the entire process can be no longer treated classically). However, the present approach have helped us to clarify the main physical ingredients and the necessary conditions for the occurrence of this effect.

References

1. A. Fedotov, N. Narozhny, G. Mourou and G. Korn, Limitations on the attainable intensity of high power lasers, *Physical review letters* **105**, p. 080402 (2010).
2. R. Ruffini, R. Moradi, J. A. Rueda, L. Becerra, C. L. Bianco, C. Cherubini, S. Filippi, Y. C. Chen, M. Karlica, N. Sahakyan, Y. Wang and S. S. Xue, On the GeV Emission of the Type I BdHN GRB 130427A, *APJ* **886**, p. 82 (Dec 2019).
3. J. A. Rueda and R. Ruffini, The blackholic quantum, *European Physical Journal C* **80**, p. 300 (April 2020).
4. L. Becerra, C. L. Ellinger, C. L. Fryer, J. A. Rueda and R. Ruffini, SPH Simulations of the Induced Gravitational Collapse Scenario of Long Gamma-Ray Bursts Associated with Supernovae, *APJ* **871**, p. 14 (January 2019).
5. Y. Wang, J. A. Rueda, R. Ruffini, L. Becerra, C. Bianco, L. Becerra, L. Li and M. Karlica, Two Predictions of Supernova: GRB 130427A/SN 2013cq and GRB 180728A/SN 2018fip, *APJ* **874**, p. 39 (March 2019).
6. J. A. Rueda, R. Ruffini, M. Karlica, R. Moradi and Y. Wang, Magnetic Fields and Afterglows of BdHNe: Inferences from GRB 130427A, GRB 160509A, GRB 160625B, GRB 180728A, and GRB 190114C, *APJ* **893**, p. 148 (April 2020).
7. S. Campion, J. Rueda, R. Ruffini and S. Xue, Magnetic field screening in strong crossed electromagnetic fields, *Physics Letters B* **820**, p. 136562 (2021).
8. S. Campion, Neutrino emission via proton-proton interaction and magnetic field screening in GRBs, PhD thesis, ICRANeT & Sapienza University of Rome (Rome, Italy, 2021), p. 191.
9. L. D. Landau and E. M. Lifshitz, *The classical theory of fields* (Pergamon, 1971).
10. J. D. Jackson, *Classical electrodynamics* (American Association of Physics Teachers, 1999).
11. S. R. Kelner, A. Y. Prosekin and F. A. Aharonian, Synchro-curvature radiation of charged particles in the strong curved magnetic fields, *The Astronomical Journal* **149**, p. 33 (2015).
12. J. Daugherty and I. Lerche, On pair production in intense electromagnetic fields occurring in astrophysical situations, *Astrophysics and Space Science* **38**, 437 (1975).

Particle creation by strong fields and quantum anomalies*

José Navarro-Salas

Departamento de Física Teórica and IFIC, Centro Mixto Universidad de Valencia-CSIC, Facultad de Física, Universidad de Valencia, Burjassot-46100, Valencia, Spain
E-mail: jnavarro@ific.uv.es

Particle creation by strong and time-varying backgrounds is a robust prediction of quantum field theory. Another well-established feature of field theory is that classical symmetries do not always extend to the quantized theory. When this occurs, we speak of quantum anomalies. In this contribution we discuss the entwining relationship between both predictions, relating chiral anomalies with an underlying process of particle creation. Within this context, we will also argue that the symmetry under electric-magnetic duality rotations of the source-free Maxwell theory is anomalous. This is a quantum effect, and it can be understood as the generalization of the fermion chiral anomaly to fields of spin one. This implies that the net polarization of photons propagating in a gravitational field could change in time.

Keywords: Chiral anomalies, gravitational particle creation, electromagnetic duality rotations, photon helicity

1. Introduction

Two fundamental predictions of quantum field theory in presence of strong field backgrounds were established in the sixties:

i) The spontaneous creation of particles out of the vacuum by time-varying gravitational fields. This was first discovered in the analysis of quantized fields in an expanding universe[1] and some years later applied to black holes[2] (for reviews see,[3–8] and for a historical perspective see[9,10]). It also offered a better understanding of the pair creation phenomena induced by electric fields, as explored by Heisenberg and Euler[11,12] and also Schwinger.[13]

ii) The breaking of the axial symmetry of Dirac fermions by quantum effects.[14,15] Classical symmetries of field theories may fail to survive quantization, leading to what is known in the literature as quantum anomalies (for a review see[16]). The origin of the anomalies is rooted in the renormalization mechanism need to tame the ultraviolet divergences that affect most models of quantum field theory. Renormalization in presence of an electromagnetic background can spoil the classical conservation law of the axial Noether current J_5^μ of massless charged fermions. It is codified by the Adler-Bell-Jackiw anomaly[14,15] ($\alpha = e^2/(4\pi\hbar)$ is the fine-structure

*Invited talk given at the Sixteenth Marcel Grossmann Meeting (2021), Session: Strong Electromagnetic and Gravitational Field Physics.

constant; we take $c = 1$)

$$\partial_\mu \langle 0|J_5^\mu|0\rangle = \frac{\alpha}{2\pi} F_{\mu\nu}{}^* F^{\mu\nu} \ . \tag{1}$$

The axial anomaly can also be interpreted as a low-energy phenomena in terms of particle production. A time-varying electric field creates both left-handed and right-handed fermions. The net amount of created chirality ΔQ_5 (i.e., the number of right-handed fermions N_R minus the number of left-handed fermions N_L created between t_1 and t_2)) is evaluated according to[a]

$$\Delta Q_5 = \hbar(N_R - N_L) = \int_{t_1}^{t_2} dt \int_{\Sigma_t} d\Sigma_\mu \langle 0|J_5^\mu|0\rangle = \frac{\alpha}{2\pi} \int_{t_1}^{t_2} \int_{\Sigma_t} d^4x F_{\mu\nu}{}^* F^{\mu\nu} \ . \tag{2}$$

It is important to stress that the minimum amount of created (massless) fermions is obtained in the adiabatic limit,[17] and it is just encapsulated in the creation of the chirality accounted by the anomaly. This might be somewhat surprising since the particle number is naively expected to be an adiabatic invariant[b]. The breaking of the adiabatic invariance of the particle number can then be understood as a signal of the chiral anomaly. Furthermore, the existence of the axial anomaly implies that there is necessarily a minimum amount of particle creation, even for an adiabatic process, to account for the creation of chirality. In more pedestrian words, the creation of chirality can be regarded as the "smoking gun" of the full particle creation process. We can summarize these features schematically:

$$\boxed{Quiral \ \ Anomaly \rightarrow Particle \ \ creation}$$

$$\boxed{Adiabatic \ \ Particle \ \ creation \rightarrow Quiral \ \ Anomaly}$$

The above results on the chiral anomaly can be extended to curved spacetime.[19] Expression (1) generalize to

$$\nabla_\mu \langle 0|J_5^\mu|0\rangle = \frac{\alpha}{2\pi} F_{\mu\nu}{}^* F^{\mu\nu} + \frac{\hbar}{192\pi^2} R_{\mu\nu\alpha\beta}{}^* R^{\mu\nu\alpha\beta} \ , \tag{3}$$

where $R_{\mu\nu\alpha\beta}$ is the Riemann tensor and ${}^*R^{\mu\nu\alpha\beta}$ its dual, and ∇_μ is the covariant derivative. We note that the gravitational part of the chiral anomaly persists for all type of fermions, either charged or neutral. Therefore, the chirality of (neutral) fermions fails also to be conserved in curved spacetimes for which

$$\Delta Q_5 = \frac{\hbar}{192\pi^2} \int_{t_1}^{t_2} \int_{\Sigma} d^4x \sqrt{-g} \ R_{\mu\nu\alpha\beta} \ {}^*R^{\mu\nu\alpha\beta} \tag{4}$$

is non-vanishing. This can be heuristically understood as a consequence of the universal character of gravity, as prescribed by Einstein's equivalence principle. If (4) is valid for a type of massless spin-1/2 field it must also be valid for any other type.

[a] We assume that a consistent particle interpretation is available at early and late times.
[b] It can be proved rigorously for a scalar field in an expanding universe.[1,18]

On physical grounds one can also argue that the universality of gravity suggests that these anomalies are not specific of spin 1/2 fermions. Therefore, one could also expect that a somewhat similar anomaly (also associated to an underlying particle production process) will arise for other fields admitting axial-type symmetries.

In this work we will further discuss the entwining relationship between particle creation and quantum anomalies. In particular, we are especially interested in exploring a new scenario in which (spontaneous and stimulated) particle creation can be relevant. The proposed scenario will be suggested by its link with quantum anomalies, as first explored in.[20] Our heuristic and intuitive argument is based on the universality of gravity, which suggest that the chiral anomaly for spin 1/2 fields should also be extended to the electromagnetic field. In the language of particle creation, one would also expect that a chiral gravitational configuration will create photons with different helicities in unequal amounts, in the same way as it happens, according to (4), for spin 1/2 fermions.

2. Trace and Axial anomalies for massless fermions

Free massless Dirac spinors are highly symmetric. In addition to their Poincaré invariance in Minkowski spacetime, they exhibit two extra symmetries: conformal and axial invariance. The conformal (or Weyl symmetry) implies the tracelessness of the stress-energy tensor $T_{\mu\nu}$, while the axial symmetry ($\psi \to e^{i\theta\gamma^5}\psi$) implies the conservation of the axial current $J^\mu_5 = \bar{\psi}\gamma^\mu\gamma^5\psi$. Both symmetries cannot be extended to the quantum theory when the Dirac field is coupled to an electromagnetic background. The conservation of the axial current is broken according to (1). Furthermore, the trace of the stress-energy tensor also acquires a non-vanishing vacuum expectation value[16,21,22] (for a more recent derivation, see[23])

$$\langle T^\mu_\mu \rangle = \frac{\alpha}{6\pi} F_{\mu\nu}F^{\mu\nu} \ , \tag{5}$$

which is usually interpreted in terms of the running of the coupling constant in quantum electrodynamics.[24] In the same way, neither of the two symmetries is preserved when the Dirac field is coupled to a gravitational background. One also finds a trace anomaly

$$\langle T^\mu_\mu \rangle = \frac{\hbar}{2880\pi^2}[aC_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} + bG + c\Box R] \ , \tag{6}$$

where $C_{\mu\nu\rho\sigma}$ is the Weyl tensor and G is the integrand of the Gauss-Bonnet topological invariant. The numerical coefficients are given by $a = -9$, $b = 11/2$, $c = 6$, and they can be obtained by different methods.[22,25] It is interesting to point out that the specific form of the trace anomaly implies that there are no massless fermions created in Friedman-Lemaître-Robertson-Walker (FLRW) universes (in this case the trace anomaly is proportional to G, up to total derivatives, and no further term proportional to R^2 appears[26]). This is fully consistent with earlier results[1] showing the absence of particle creation for fields obeying conformally invariant equations.

As remarked in the introduction one also finds an axial anomaly of the form

$$\nabla_\mu \langle J_5^\mu \rangle = \frac{\hbar}{192\pi^2} R_{\mu\nu\alpha\beta}{}^* R^{\mu\nu\alpha\beta} \ . \tag{7}$$

This anomaly can be interpreted in terms of particle creation induced by a chiral gravitational configuration (see[27] for a cosmological setting). Here we want to remark that (7) is consistent with late-time black hole emission.[2] Stationary Kerr black holes emit fermions with a helicity-dependent angular distribution,[28–30] as one could heuristically expect from the local form of the anomaly (7). Fermions with positive helicity are emitted preferentially along the direction of rotation, while fermions with negative helicity in the opposite direction.[For instance, neutrinos are preferentially emitted in the direction opposite to hole's rotation, while antineutrinos in the direction of rotation]. However, the net contribution when integrated over all angles is zero, in agreement with the vanishing of (4) for the Kerr metric. Nevertheless, in the transient process through the formation of a single Kerr black hole, as for instance the merger of two black holes (as the ones currently observed by LIGO-Virgo), the net contribution is not zero, as it has been evaluated using numerical relativity.[31] [For neutrinos/antineutrinos this means creation of matter-antimatter asymmetry]. Although the net creation of helicity is still small, due to the short duration of the process, it could be more significant in scenarios displaying an accumulative process over long periods of time. In any case, one should always take into account that the net creation of helicity represents only a lower bound of the full particle creation process. More particles could be produced without contributing to the creation of helicity.

The above discussion applies equally to the emission of photons. Right-handed photons are radiated more abundantly in the direction parallel to the axis of rotation of a Kerr black hole, and left-handed photons are emitted more abundantly in the opposite direction. This suggests the existence of an "axial anomaly" for spin-1 fields.

3. Electro-magnetic duality as an axial anomalous symmetry

Can we extend the above considerations to the electromagnetic field? Concerning the conformal symmetry, it is well-known that it is broken by quantum effects induced by the gravitational background. The trace of the renormalized stress-energy tensor can be expressed as

$$\langle T_\mu^\mu \rangle = \frac{-62\hbar}{2880\pi^2} \left(R^{\mu\nu} R_{\mu\nu} - \frac{1}{3} R^2 \right) + \frac{\hbar}{16\pi^2} c \Box R \ , \tag{8}$$

where c is an ambiguous coefficient, which depend on the particular regularization method ($c = -1/10$ for point-plitting and zeta-function regularization, and $c = 1/15$ for dimensional regularization[22]). We note again[26] that, for a FLRW universe, the trace anomaly turns out to be proportional to the integrand of the Gauss-Bonnet invariant G, up to the ambiguous and total derivative term $\Box R$. The absence of

an extra term proportional R^2 is crucial to predict the absence of massless, spin-1 particle creation in an isotropically expanding universe.

The point now is: what is the analog of axial symmetry for the free electromagnetic field? A natural candidate are the well-known electric-magnetic duality transformations, defined by

$$F^{\mu\nu} \to F'^{\mu\nu} = \sin\theta \ {}^*F^{\mu\nu} + \cos\theta \ F^{\mu\nu} \ . \tag{9}$$

The above transformations leave the action

$$S_{Maxwell} = -\frac{1}{4}\int d^4x\sqrt{-g}(\nabla_\mu A_\nu - \nabla_\nu A_\mu)(\nabla^\mu A^\nu - \nabla^\nu A^\mu) \tag{10}$$

invariant, as first proved in Refs.[32,33] in Minkowski space and in Refs.[34,35] in curved space. The associated Noether current can be expressed as

$$j^\mu_D = \frac{1}{2}[A_\nu \ {}^*F^{\mu\nu} - Z_\nu \ F^{\mu\nu}] \ , \tag{11}$$

where the auxiliary field Z_μ is defined as $\nabla_\mu Z_\nu - \nabla_\nu Z_\mu = {}^*F_{\mu\nu}$. j^μ_D is gauge invariant, but it cannot be expressed locally in terms of the field strength, in sharp contrast with the stress-energy tensor. The conserved charge[c]

$$Q_D = \int_{\Sigma_t} d\Sigma_\mu j^\mu_D \tag{12}$$

evaluates the electromagnetic helicity of a classical electromagnetic configuration.[33]

We note that Q_D can be decomposed in two contributions[37]

$$Q_D = Q_m + Q_e \ , \tag{13}$$

where

$$\begin{aligned} Q_m &= \frac{1}{2}\int_{\Sigma_t} d\Sigma_\mu A_\nu \ {}^*F^{\mu\nu} \\ Q_e &= -\frac{1}{2}\int_{\Sigma_t} d\Sigma_\mu Z_\nu \ F^{\mu\nu} \ . \end{aligned} \tag{14}$$

Q_m is the magnetic helicity and Q_e the electric helicity. Neither Q_m or Q_e are conserved quantities, in general.

At the quantum level Q_D is proportional to the difference between the number of right-handed and left-handed photons[33,38]

$$Q_D = \hbar(N_R - N_L) \ . \tag{15}$$

Therefore, this quantity is no longer conserved if the symmetry is afflicted by an anomaly. This issue was worked out in Refs.[20,39–41] from different methods and viewpoints, arguing that the symmetry under electric-magnetic duality rotations

[c] Q_D can be written in terms of a non-local integral involving only electric and magnetic fields.[36]

becomes anomalous in curved spacetime. The result is somewhat parallel to that found for spin-1/2 fields

$$\nabla_\mu \langle j_D^\mu \rangle \propto \hbar R_{\mu\nu\alpha\beta}{}^* R^{\mu\nu\alpha\beta} \ . \tag{16}$$

We want to remark here that previous works in the eighties[42–44] computed the divergence of the Pauli-Lubansky vector $K^\mu \propto A_\nu {}^* F^{\mu\nu}$ in curved space

$$\nabla_\mu \langle A_\nu \ {}^* F^{\mu\nu} \rangle = \frac{1}{2} \langle F_{\mu\nu} {}^* F^{\mu\nu} \rangle = \frac{\hbar}{96\pi^2} R_{\mu\nu\alpha\beta} {}^* R^{\mu\nu\alpha\beta} \ . \tag{17}$$

This result is indeed related to the (classically non-conserved) magnetic helicity, instead of the electromagnetic helicity. We can easily realize this from the fact that the Pauli-Lubansky vector is proportional to the current

$$j_m^\mu \equiv \frac{1}{2} A_\nu \ {}^* F^{\mu\nu} \ . \tag{18}$$

This current gives the magnetic helicity as the integral

$$Q_m(t) = \int_{\Sigma_t} d\Sigma_\mu j_m^\mu \ . \tag{19}$$

j_m^μ is not a Noether current, hence Q_m is not time-independent. The proper Noether current associated to the electro-magnetic duality symmetry involves an extra contribution $j_e^\mu \equiv -\frac{1}{2} Z_\nu F^{\mu\nu}$. This additional term is crucial to produce a classical conserved current

$$j_D^\mu = j_m^\mu + j_e^\mu \ , \tag{20}$$

which can be physically interpreted in terms of the spin-1 axial anomaly if $\nabla_\mu \langle j_D^\mu \rangle$ is different from zero. Since $\nabla_\mu \langle j_e^\mu \rangle = -\langle F^*_{\mu\nu} F^{\mu\nu} \rangle - \frac{1}{2} \langle Z_\nu \nabla_\mu F^{\mu\nu} \rangle$, one gets

$$\nabla_\mu \langle j_D^\mu \rangle = -\frac{1}{2} \langle Z_\nu \nabla_\mu F^{\mu\nu} \rangle \ . \tag{21}$$

It has been argued in[39–41] that the above vacuum expectation value is nonvanishing and that it is proportional, as expected, to $R_{\mu\nu\alpha\beta}{}^* R^{\mu\nu\alpha\beta}$, in parallel with the fermionic case. As stressed above, and in the language of particles, this would imply that the difference in the number of photons with positive and negative helicities, $N_R - N_L$, is not necessarily conserved in curved spacetimes.

Nevertheless, we want to remark that the result (17) contains an indirect signal of the electromagnetic axial anomaly found in.[39–41] If the invariance of the electromagnetic field equations under the duality transformation $F^{\mu\nu} \to F'^{\mu\nu} = \sin\theta \ {}^* F^{\mu\nu} + \cos\theta \ F^{\mu\nu}$ is strictly translated to composite quantum operators one would get

$$\langle F_{\mu\nu} {}^* F^{\mu\nu} \rangle = (\cos^2\theta - \sin^2\theta) \langle F_{\mu\nu} {}^* F^{\mu\nu} \rangle - \sin\theta \cos\theta (\langle F_{\mu\nu} F^{\mu\nu} \rangle - \langle {}^* F_{\mu\nu} {}^* F^{\mu\nu} \rangle) \ . \tag{22}$$

This would force $\langle F_{\mu\nu} {}^* F^{\mu\nu} \rangle = 0 = \langle F_{\mu\nu} F^{\mu\nu} \rangle - \langle {}^* F_{\mu\nu} {}^* F^{\mu\nu} \rangle$. Since neither $\langle F_{\mu\nu} {}^* F^{\mu\nu} \rangle$ or $(\langle F_{\mu\nu} F^{\mu\nu} \rangle - \langle {}^* F_{\mu\nu} {}^* F^{\mu\nu} \rangle)$ are zero (see[45] for a detailed discussion on this), we

should conclude that electro-magnetic duality fails for non-linear vacuum expectation values. This is the underlying reason permitting the result

$$\nabla_\mu \langle j_m^\mu \rangle + \nabla_\mu \langle j_e^\mu \rangle \neq 0 \ . \tag{23}$$

3.1. *Chiral anomalies and gravitational radiation*

We end this section by outlining an very interesting connection between chiral anomalies and gravitation radiation.[31,46] It is well-know that the right-hand side of (16), as for anomalies in gauge theories, is a total divergence. This simple fact suggests to reinterpret the result (16), or the analogous one for massless fermions, in a physically appealing manner. To evaluate the produced quirality induced by particle creation one should evaluate four-dimensional integrals of the form (4). In doing this one gets crucial contributions from the boundary of the spacetime (i.e., null infinity) involving the outgoing flux of gravitational waves. The contribution of the chiral anomaly can then be exactly related to the amount of circular polarization of the outoging gravitation radiation. Following[31,46] one gets the intriguing relation (see[46] for details)

$$\int d^4x R_{\mu\nu\lambda\sigma}{}^\star R^{\mu\nu\lambda\sigma} \propto \int_0^\infty \frac{d\omega\omega^3}{24\pi^3} \sum_{\ell m} [|h_+^{\ell m}(\omega) + i h_\times^{\ell m}(\omega)|^2 - |h_+^{\ell m}(\omega) - i h_\times^{\ell m}(\omega)|^2] \ , \tag{24}$$

where h_+, $h_\times$ are the standard gravitational waves polarization modes. The right-hand-side is related to the difference in the intensity between right and left circularly-polarized gravitational waves reaching future null infinity. This shows that a flux of circularly polarized gravitational waves triggers the spontaneous creation of quanta with net helicity. Note the similarity with the Hawking emission by rotating black holes. The angular momentum of the Kerr black hole triggers the spontaneous creation of quanta with net angular momentum.

4. Final remarks

It is well-known that spontaneous emission induces stimulated emission in presence of bosons. This is also true for gravitational particle creation.[1] One can write the simple and basic result

$$\langle N_i(t) \rangle \equiv \langle \Psi | a_i^\dagger(t) a_i(t) | \Psi \rangle = \langle N_i^0 \rangle + (1 + 2\langle N_i^0 \rangle) |\beta_i(t)|^2 \tag{25}$$

where N_i^0 is the initial number of quanta in mode i contained in the quantum state $|\Psi\rangle$ (the effect is reversed for fermions). This was also considered for black hole radiation in,[47,48] and more recently in[49,50] for non-gaussianities during inflation.

The stimulated counterpart effect is the main difference in the consequences of the axial anomaly for spin-1/2 fermions and photons. In the latter case, the presence of photons in a given mode will trigger the creation of photons of the same mode. It is not easy to evaluate quantitatively the consequences of this effect on

a macroscopic pulse of radiation. But it not difficult to guess that it will change the circular polarization of light-rays, with trajectory $x^\mu = x^\mu(\tau)$, propagating through a gravitational field with non-trivial $R_{\mu\nu\lambda\sigma}{}^\star R^{\mu\nu\lambda\sigma}(x(\tau))$. This is a quantum effect, probably very tiny, to be added to the classical gravitational redshift and the deflection of light rays by massive bodies.[51]

Acknowledgements

This work was supported in part by Spanish Ministerio de Economia, Industria y Competitividad Grants No. FIS2017-84440-C2-1-P(MINECO/FEDER, EU), No. FIS2017-91161-EXP and the project PROMETEO/2020/079 (Generalitat Valenciana).

References

1. L. Parker, *The creation of particles in an expanding universe*, Ph.D. thesis, Harvard University (1966). Dissexpress.umi.com, Publication Number 7331244; Phys. Rev. Lett. **21**, 562 (1968); Phys. Rev. D **183**, 1057 (1969); Phys. Rev. D **3**, 346 (1971).
2. S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
3. N. D. Birrell and P. C. W. Davies, *Quantum fields in curved space*, Cambridge University Press, Cambridge, UK (1982).
4. S. A. Fulling, *Aspects of quantum field theory in curved space-time*, Cambridge University Press, Cambridge, UK (1989).
5. R. M. Wald, *Quantum field theory in curved spacetime and black hole thermodynamics*, Chicago University Press, Chicago, USA (1994).
6. L. Parker and D. J. Toms, *Quantum field theory in curved spacetime: Quantized fields and gravity*, Cambridge University Press, Cambridge, UK (2009).
7. V. P. Frolov and I. D. Novikov, *Black hole physics*, Kluwer Academic Publishers, Dordrecht, The Netherlands (1998).
8. A. Fabbri and J. Navarro-Salas, *Modeling black hole evaporation*, ICP-World Scientific, London, UK (2005).
9. L. Parker, J. Phys. Conf. Ser. **600**, 1, 012001 (2015); arXiv: 1503.00359.
10. L. Parker and J. Navarro-Salas, *Fifty years of cosmological particle creation*, arXiv:1702.07132.
11. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936). English translation: arXiv:physics/0605038.
12. G. V. Dunne, Int. J. Mod. Phys. A **27**, 1260004 (2012); F. Karbstein, *Particles* **3**, 39 (2020).
13. J. Schwinger, Phys. Rev. **82**, 664 (1951).
14. S. L. Adler, Phys. Rev. **82**, 2426 (1969).
15. J. S. Bell and R. Jackiw, Nuovo Cimento A **51**, 47 (1969).
16. R. A. Bertlmann, *Anomalies in quantum field theory*, Oxford University Press, Oxford.
17. P. Beltran-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla, Phys. Rev. D **100**, 085014 (2019).
18. L. Parker, J. Phys. A **45**, 374023 (2012).
19. T. Kimura, Progress. Theor. Phys. **42**, 1191 (1969).
20. I. Agullo, A. del Rio and J. Navarro-Salas, Int. J. Mod. Phys. D **26**, 1742001 (2017).

21. R. J. Crewther, Phys. Rev. Lett. **28**, 1421 (1972).
22. M. J. Duff, Class. Quantum Grav. **11** 1397 (1994).
23. P. Beltran-Palau, A. Ferreiro, J. Navarro-Salas and S. Pla, Phys. Rev. D **101**, 105014 (2020).
24. S. L. Adler, J. C. Collins and A. Duncan, Phys. Rev. D **15**, 1712 (1977).
25. A. Landete, J. Navarro-Salas and F. Torrenti, Phys. Rev. D **88** (2013) 061501; Phys. Rev. D **89** 044030 (2014). J. F. Barbero G., A. Ferreiro, J. Navarro-Salas and E. J. S. Villaseñor, Phys. Rev. D **98**, 025016 (2018).
26. L. Parker, *Aspects of quantum field theory in curved spacetime: Effective action and energy-momentum tensor*, in Recent Developments in Gravitation, Cargese 1978, ed. M. Lévy and S. Deser (Plenum Press, New York), 219-273.
27. G. W. Gibbons, Phys. Lett. B **84**, 431 (1979).
28. Don N. Page, Phys. Rev. D **14**, 3260 (1976).
29. A. Vilenkin, Phys. Rev. D **20**, 1807 (1979).
30. D. A. Leahy and W. G. Unruh, Phys. Rev. D **19**, 3509 (1979).
31. A. del Rio et al., Phys. Rev. Lett. **124**, 211301 (2020).
32. D. M. Lipkin, J. of Math. Phys. **5**, 696 (1964).
33. M. G. Calkin, Am. J. Phys. **33**, 958 (1965)
34. S. Deser, and C. Teitelboim, Phys. Rev. D **13**, 1592 (1976).
35. S. Deser, J. Phys. A **15**, 1053 (1982).
36. J. Bernabeu and J. Navarro-Salas, Symmetry **11** (10), 1191 (2019); arXiv:1910.05041.
37. M. Galaverni and G. S. J. Gabriele, Gen. Relativ. Gravit. **53**, 46 (2021).
38. J. L. Trueba and A. F. Rañada, European Journal of Physics **17**, 141 (1996).
39. I. Agullo, A. del Rio and J. Navarro-Salas, Phys. Rev. Lett. **118**, 111301 (2017).
40. I. Agullo, A. del Rio and J. Navarro-Salas, Phys. Rev. D **98**, 125001 (2018).
41. I. Agullo, A. del Rio and J. Navarro-Salas, Symmetry **10**, 763 (2018).
42. A. D. Dolgov, I. B. Khriplovich and V. I. Zakharov, JETP Lett. **45**, 651 (1987).
43. A. D. Dolgov, I. B. Khriplovich, A. I. Vainshtein and V. I. Zakharov, Nucl. Phys. B **315**, 138 (1989).
44. M. Reuter, Phys. Rev. D **37**, 1456 (1988).
45. I. Agullo, A. Landete and J. Navarro-Salas, Phys. Rev. D **90**, 124067 (2014).
46. A. del Rio, Phys. Rev. D **104**, 065012 (2021).
47. R. M. Wald, Phys. Rev. D **13**, 3176 (1976).
48. J. D. Bekenstein and A. Meisels, Phys. Rev. D **15**, 2775 (1977).
49. I. Agullo and L. Parker, Phys. Rev. D **83**, 063526 (2011); Gen. Rel. Grav. **43**, 2541 (2011).
50. I. Agullo, J. Navarro-Salas and L. Parker, J. Cosmol. Astropart. Phys. **05**, 019 (2012).
51. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, San Francisco, USA (1973).

Constraints on the non-minimal coupling of electromagnetic fields from astrophysical observations

Susmita Jana and S. Shankaranarayanan
Department of Physics, Indian Institute of Technology Bombay, Mumbai, 400076, India
E-mail: susmitajana@iitb.ac.in, shanki@phy.iitb.ac.in

Strong gravity regions, like the neighborhood of black holes or neutron stars, can induce non-minimal couplings between electromagnetic fields(EM) and gravity. In these regions, gravitational fields behave as a non-linear medium in which the electromagnetic fields propagate. For a system of mass M and size R, the surface potential scales as M/R. Pulsar timing array, Double pulsar Shapiro delay, and Event horizon telescope probe that largest surface potentials $[10^{-4} - 10^{-2}]$. With many future experiments, it is possible to constrain the non-minimal coupling between electromagnetic fields and gravity. As a step in this direction, we consider the non-minimal coupling of EM field tensor through Riemann tensor for a dynamical black-hole, described by Sultana-Dyer(SD) metric. The non-minimal coupling leads to modified dispersion relations of photons, which get simplified at $\tilde{E}/\tilde{L} >> 1$ regime, where $\tilde{E}$ and $\tilde{L}$ are two conserved quantities obtained by taking into account the symmetries of the metric. We calculate polarization-dependent photon deflection angle and arrival time from these dispersion relations, which we evaluate considering different astrophysical sources of photons. We compare the analytical results with the current astrophysical observations to constraint the non-minimal coupling parameters to Riemann tensor more stringently.

Keywords: Non-minimal coupling, time-dependent black hole, strong fields; Based on.[1]

1. Introduction

The general theory of relativity has been tested in the weak-field regime with great accuracy from astrophysical observations — gravitational redshift, light deflection, and perihelion precession of planetary orbits.[2] However, it has not been verified well in strong-field regions such as near a black hole or a neutron star and many have proposed higher-order curvature corrections to GR.[3–7] Testing theories of gravity both in the strong and weak sector requires astrophysical observations, electromagnetic wave is an important source of such observations.

While testing GR through properties associated with photons we consider the minimal coupling of EM fields with gravity —Maxwell's action in curved space-time. Theories with minimal coupling are well tested in weak, but not in strong gravity regions.[8,9] Hence, the accuracy of the theories of gravity including the minimal coupling of the electromagnetic field requires investigation. To test GR in the strong-field limit it is essential to understand how does gravity interact with electromagnetic fields in such regions. One possibility is to consider coupling of the electromagnetic field tensor with the curvature of the space-time through Riemann and Ricci tensors, known as non-minimal coupling of electromagnetic fields.[3]

A model describing non-minimal coupling of electromagnetic fields with curved space-time, violating Einstein's equivalence principle, was first introduced in Ref.[3]. After that different models with more generalized actions were proposed in Refs.[4–7]. As a result, they obtained polarization-dependent photon propagation obeying distinct dispersion relations for the individual polarization modes. Whereas, minimally coupled photons follow dispersion relation $p^\mu p_\mu = 0$ irrespective of their polarization mode.

Even though non-minimal coupling in the space-time of static black hole reveals new behavior of EM waves, it is essential to study the same for a time-dependent black hole as well. One example of a time-dependent black hole is the Sultana-Dyer[10] black hole which describes a Schwarzschild black hole immersed in background FLRW space-time. In this talk we have discussed the properties of photons, non-minimally coupled to the curvature produced by a Sultana-Dyer black-hole. We expect separate dispersion relations of photons for each polarization mode, different than what was obtained for a time-independent black hole.

We also inspect the trajectory of non-minimally coupled photon for an individual mode of polarizations in the vicinity of a time-dependent black hole. In order to do so, we calculate the polarization-dependent deflection angle of a photon in Sultana Dyer space-time. We include coupling of EM field tensor with Riemann tensor only in our analysis as mentioned in.[3] From the expression of the polarization-dependent deflection angles, we constraint the value of the coupling-constant present in action. Inclusion of the dynamical black hole instead of a static one improves the bound on the coupling constant up to $\mathcal{O}(10^{10})$ for a solar-mass black hole.

2. General model

Action describing the most general model of non-minimal coupling that consists coupling of electromagnetic field tensor with Riemann, Ricci tensors and Ricci scalar is[7]:

$$\mathcal{S}_{NC} = \int d^4x\sqrt{-g}\left[\frac{R}{\kappa} - \frac{1}{4}\,F_{\mu\nu}F^{\mu\nu} + \chi^{\mu\nu\alpha\beta}\,F_{\mu\nu}F_{\alpha\beta}\right], \tag{1}$$

where,

$$\chi^{\mu\nu\alpha\beta} = \frac{q_1 R}{2}\left(g^{\mu\alpha}g^{\nu\beta} - g^{\mu\beta}g^{\nu\alpha}\right) + \frac{q_2}{2}\left(R^{\mu\alpha}g^{\nu\beta} - R^{\mu\beta}g^{\nu\alpha} + R^{\nu\beta}g^{\mu\alpha} - R^{\nu\alpha}g^{\mu\beta}\right) + \frac{\lambda}{2}R^{\mu\nu\alpha\beta}. \tag{2}$$

As mentioned earlier, the models of non-minimal coupling violate the Einstein equivalence principle(EEP), hence this model also violates EEP while being invariant under CPT and general coordinate transformation. A simpler model from the above action can be obtained by choosing $q_1 = q_2 = 0$.[11]

$$S = \int d^4x\,\sqrt{-g}\,\left(-\frac{1}{4}\,F_{\mu\nu}F^{\mu\nu} + \frac{\lambda}{2}\,R^{\mu\nu\alpha\beta}\,F_{\mu\nu}\,F_{\alpha\beta}\right). \tag{3}$$

Authors in Ref.[11] have studied properties of non-minimally coupled photon obeying this action for a spherically symmetric static black hole described by Schwarzschild metric. They have obtained non-identical modified dispersion relation for each mode of polarization followed by two polarization-dependent photon arrival times. Calculating the difference between arrival time for individual polarization they obtained constraints on the coupling coefficient λ to be $\mathcal{O}(10^{20})$ cm^2 for a Solar mass black hole.

However, in reality, the black holes are non-static — embedded in the background expanding universe. That is why in search of the imprints of a dynamical black hole we have done a similar analysis for a time-dependent black hole with mass $[M_\odot 10^{-10}, M_\odot]$.

Electromagnetic fields non minimally coupled to the curvature of space-time obey the following equation of motion,

$$\nabla_\nu F^{\mu\nu} = 2\lambda \left[R^{\mu\nu\alpha\beta} \left(\nabla_\nu F_{\alpha\beta}\right) + \left(\nabla_\alpha R^\mu_\beta\right) F^{\alpha\beta} \right]. \tag{4}$$

The LHS of the above equation contains dynamical properties of the EM field, whereas the RHS of the EoM depends on the geometry of the space-time. The first term in RHS acts as a metric-dependent momentum term which varies with the component of the electromagnetic field tensor. The second term in RHS is similar to a mass term, in this case, it differs for each value of μ. For a static black hole described by Schwarzschild or Kerr metric the term with $\nabla_\alpha R^\mu{}_\beta$ is zero in Eq. (4) as $R_{\mu\nu} = 0$ for such metrics.

Our goal is to get dispersion relation of photon that depends on photon momentum, hence it is reasonable to define a local inertial frame attached to the non-inertial one and define the photon momentum in the local inertial frame by $p_{(\mu)}$. Momentum in local frame is related to momentum in non-inertial frame (p_ν) by the relation: $p_\nu = e_\nu{}^{(\mu)}\ p_{(\mu)}$, where, $e_\nu{}^{(\mu)}$ is the tetrad connecting the two — inertial and non-inertial frames.

In the case of black-hole space-times we consider, the variations of the curvature along the spatial coordinate ($\mathcal{L}$) and the time ($\mathcal{T}$) are large compared to the wavelength and period of electromagnetic waves, i.e.

$$\mathcal{L} >> \bar{\lambda} = 1/\bar{k}, \quad T >> 1/\omega \tag{5}$$

Under the Eikonal approximation, locally we can treat the EM waves as planar and monochromatic[11]:

$$\nabla_{(\mu)} F^{(\mu)(\nu)} = p_{(\mu)}\, F^{(\mu)(\nu)} \tag{6}$$

Using this relation into Eq. (4) we get the equation that describes the dynamics of the electromagnetic field in the local inertial frame,

$$p_{(\nu)}\, F^{(\mu)(\nu)} = 2\lambda \left[R^{(\mu)(\nu)(\alpha)(\beta)}\, p_{(\nu)} F_{(\alpha)(\beta)} + \left(\nabla_{(\alpha)} R^{(\mu)}{}_{(\beta)}\right) F^{(\alpha)(\beta)} \right]. \tag{7}$$

Setting the space-time index μ to the 3-space index j followed by incorporation of the Bianchi identity in local inertial frame into the above equation leads us to the equation below.

$$\begin{aligned}&[\ p^{(\mu)}p_{(\mu)}\,\delta^{(j)}_{(k)} - p_{(k)}p^{(j)} + 4\lambda\, R^{(j)(\mu)(\nu)}{}_{(k)}\, p_{(\mu)}p_{(\nu)}\\&+ 2\lambda\,\left(\nabla_{(\alpha)}R^{(j)}{}_{(k)}\,p^{(\alpha)} - \nabla_{(k)}R^{(j)}{}_{(\alpha)}p^{(\alpha)}\right)\ ]\,F^{(0)(k)} = 0\end{aligned} \tag{8}$$

Using Eq. (7) into the above equation, the equation of motion of the EM field in the Eikonal approximation is:

$$\begin{aligned}&\Bigg[p^{(\mu)}p_{(\mu)}\,\delta^{(j)}_{(k)} + 4\lambda\left(\frac{p^{(j)}}{p_{(0)}}\epsilon^{(0)}{}_{(k)} + \epsilon^{(j)}{}_{(k)}\right)\\&+ 2\lambda\,\left(\frac{p^{(j)}}{p_{(0)}}\nabla_{(\alpha)}R^{(0)}{}_{(k)}\,p^{(\alpha)} + \nabla_{(\alpha)}R^{(j)}{}_{(k)}p^{(\alpha)}\right)\\&- 2\lambda\,\left(\frac{p^{(j)}}{p_{(0)}}\nabla_{(k)}R^{(0)}{}_{(\alpha)}\,p^{(\alpha)} + \nabla_{(k)}R^{(j)}{}_{(\alpha)}p^{(\alpha)}\right)\Bigg]F^{(0)(k)} = 0,\end{aligned} \tag{9}$$

where, $\epsilon^{(\alpha)}{}_{(\beta)} \equiv R^{(\alpha)(\mu)(\nu)}{}_{(\beta)}\,p_{(\mu)}p_{(\nu)}$ acts as a polarization tensor that generates correction to EoM according to the component of electric field.

From the above equation, one can study the trajectory of non-minimally coupled photons considering different space-time metrics. As mentioned earlier for photons in Schwarzschild and Kerr black hole terms with Ricci tensor in Eq. (9) will vanish and the EoM reduces to,

$$\left[p^{(\mu)}p_{(\mu)}\,\delta^{(j)}_{(k)} + 4\lambda\left(\frac{p^{(j)}}{p_{(0)}}\epsilon^{(0)}{}_{(k)} + \epsilon^{(j)}{}_{(k)}\right)\right]F^{(0)(k)} = 0, \tag{10}$$

Our goal is to study the effects of the curvature coupling on photon trajectory near a time-dependent black-hole and identify the differences with the same due to a static black hole. In literature the analysis has been done only for static black holes and as a result constraint on λ has also been obtained. As a step ahead, We do this analysis for a cosmological black hole described by Sultana Dyer metric.[10]

3. Sultana Dyer metric

In reality, black holes are non-static and the Sultana Dyer metric describes such a time-dependent black hole.[10] Sultana Dyer black hole metric can be interpreted as a Schwarzchild black hole immersed in the background FLRW metric:

$$ds^2 = \left(\frac{\eta}{\eta_0}\right)^4\left[-\left(1-\frac{2GM}{r}\right)\,d\eta^2\ + \frac{4GM}{r}\,d\eta dr + \left(1+\frac{2GM}{r}\right)dr^2\ +\ r^2\,d\Omega^2\right] \tag{11}$$

where η is the conformal time—related to comoving time t as, $dt = a(\eta)d\eta$; $\eta = \eta_0$ at $t = t_0$, where t_0 is current cosmic time. M is mass of the black hole.

Sultana Dyer metric has been obtained by performing conformal transformation $a(t) \propto t^{2/3}$ to the Schwarzschild metric. Thus it corresponds to black-holes in the matter-dominated epoch. The conformal transformation preserves the casual structure also, as a result, the black hole has the horizon at $r_{\rm H} = 2GM/c^2$. Sultana-Dyer space-time is sourced by two non-interacting perfect fluids — null dust and massive dust. At the horizon, $r = r_{\rm H}$ the dust particles become superluminal for time $\eta > 2r_{\rm H}$.[10]

Writing the above equation in non-geometrised unit—$\eta \to c\eta$, $r_{\rm H} \to \frac{r_{\rm H}}{c^2}$ and substituting the following dimensionless variables:

$$\tilde{r} = \frac{c^2 r}{r_{\rm H}}; \quad \tilde{\eta} \equiv H_0\, \eta; \quad \tilde{\eta}_0 \equiv H_0\, \eta_0, \tag{12}$$

in the metric we have

$$ds^2 = \alpha_0^2\ \tilde{\eta}^4 \left[-\left(1 - \frac{1}{\tilde{r}}\right)\ \beta_0^2 d\tilde{\eta}^2 + \frac{2}{\tilde{r}}\beta_0\ d\tilde{\eta} d\tilde{r} + \left(1 + \frac{1}{\tilde{r}}\right) d\tilde{r}^2 + \tilde{r}^2\ d\Omega^2 \right] = \alpha_0^2\ d\tilde{s}^2 \tag{13}$$

where H_0 is Hubble constant,

$$d\tilde{s}^2 = \tilde{\eta}^4 \left[-\left(1 - \frac{1}{\tilde{r}}\right)\ \beta_0^2 d\tilde{\eta}^2 + \frac{2}{\tilde{r}}\beta_0\ d\tilde{\eta} d\tilde{r} + \left(1 + \frac{1}{\tilde{r}}\right) d\tilde{r}^2 + \tilde{r}^2\ d\Omega^2 \right] \tag{14}$$

$$\alpha_0 = \frac{r_{\rm H}}{c^2 \tilde{\eta}_0^2} \qquad \beta_0 = \frac{c^3}{r_{\rm H}\, H_0} \tag{15}$$

From here we will be using the dimensionless line-element $d\tilde{s}$ defined in Eq. (14). This choice is suitable because the affine parameter (τ) and all the physical quantities in this space-time depend on τ.

3.1. *Conserved quantities from symmetries of the metric*

The trajectory of a null object such as a photon is being fixed by its momentum p^μ. In this work, we focus on photons moving in the orbital plane of the black hole. For the metric (13), we define the contravariant photon momentum as,

$$p^0 = \frac{d\tilde{\eta}}{d\tau}, \quad p^1 = \frac{d\tilde{r}}{d\tau}, \quad p^2 = 0, \quad p^3 = \frac{d\phi}{d\tau} \tag{16}$$

Using the metric and the relation $p_\nu = g_{\mu\nu} p^\mu$ we get the covariant form:

$$p_0 = \tilde{\eta}^4 \left(-\beta_0^2\, A \frac{d\tilde{\eta}}{d\tau} + C_1 \frac{d\tilde{r}}{d\tau} \right) := \tilde{E}; \quad p_1 = -\frac{C_1}{A}\tilde{E} + \tilde{\eta}^4 \left(\frac{C_1^2}{A} + B \right) \frac{d\tilde{r}}{d\tau} \tag{17}$$

$$p_3 = \tilde{\eta}^4 C_2 \frac{d\phi}{d\tau} := \tilde{L} \tag{18}$$

Though the metric is time-dependent it is —conformally invariant w.r.t η or $\tilde{\eta}$ and spherically symmetric. Taking into account these two symmetries it can be shown that $\tilde{E}$ and $\tilde{L}$ defined in Eq. (17) are conserved w.r.t the affine parameter τ. where, $A = (1 - \frac{1}{\tilde{r}})$, $B = (1 + \frac{1}{\tilde{r}})$, $C_1 = \frac{\beta_0}{\tilde{r}}$, $C_2 = \tilde{r}^2$.

Our goal is to see effects on the non-minimally coupled photon close to the black hole, and later it will be clear that the effects depend on $\frac{\tilde{E}}{\tilde{L}}$, which is:

$$\frac{\tilde{E}}{\tilde{L}} = \left(\frac{r_{\rm H}}{c^2\, r}\right)^2 \left(-\left(1-\frac{r_{\rm H}}{c^2\, r}\right)\beta_0^2 H_0\, \frac{d\eta}{d\phi} + \frac{\beta_0}{r}\, \frac{dr}{d\phi}\right) \tag{19}$$

Close to the black hole at radius of photon sphere, $r = 3r_{\rm H}/(2c^2)$, it can be shown that for M in the range $[M_\odot 10^{-10}, M_\odot]$ the ratio $\tilde{E}/\tilde{L} >> 1$.

4. Modified dispersion relation

As a result of the non-minimal coupling dispersion relation for photon near Schwarzschild and Sultana Dyer black hole can be obtained from Eq. (10) as[11]:

$$p^2 = \pm\, \frac{6\lambda}{r_{\rm H}^2 \tilde{r}^3}\, p_{(3)}^2. \tag{20}$$

From the above expression, we infer that the photon dispersion relation remains quadratic also as similar to the relation $p^\mu p_\mu = 0$ with $r-$dependent corrections. There $(\pm)$ sign refers to dispersion relation for $(\pm)$ polarization mode. We have verified that for the Kerr metric, describing a time-independent rotating black hole the dispersion relation remains quadratic — even in the inertial frame of reference it gives the same dispersion relation (20).

The dispersion relations for the Sultana Dyer metric will be obtained from Eq. (9). Unlike Schwarzschild and Kerr black holes, the Ricci tensor is non-zero for Sultana-Dyer black hole. Hence, the terms with the Ricci tensor and its derivative will contribute to the dispersion relation. However, it is possible to check that the contribution arising from the Ricci tensor containing terms is much smaller than the other terms in the dispersion relation. Finally, keeping only the relevant terms we obtain the dispersion relations for the local inertial frame in the Sultana-Dyer metric as:

$$-f_1\, p_{(0)}^2 + f_1\, p_{(1)}^2 + f_2\, p_{(3)}^2 + f_3\, \frac{p_{(1)} p_{(3)}^2}{p_{(0)}} = 0 \tag{21}$$

$$-f_4\, p_{(0)}^2 + f_2\, p_{(1)}^2 + f_5\, p_{(3)}^2 + 2f_3\, p_{(1)} p_{(0)} = 0\,. \tag{22}$$

where, $f_n(\tilde{r}, \tilde{\eta})$ depends on the coupling constant $\tilde{\lambda} \equiv \lambda/\alpha_0^2$ and the components of Riemann tensor in local inertial frame for Sulatana-Dyer metric — $R_{(\alpha)(\beta)(\mu)(\nu)}$. These two dispersion relations in the local frame lead to the following relations in the non-inertial frame:

$$\frac{(C_1^3 - C_1)}{B} p_1^3 + (1 - 3C_1^2)\; p_0 p_1^2 + \left(3C_1 B\, p_0^2 + \frac{(f_3 - f_2\, C_1)}{f_1\, C_2} p_3^2\right) p_1 - B^2 p_0^3 + \frac{f_2\, B}{f_1\, C_2}\, p_0 p_3^2 = 0 \tag{23}$$

$$\left(-f_4 C_1^2 - 2f_3 C_1 + f_2\right) p_1^2 + 2\,(f_3 + f_4 C_1)\, B p_0 p_1 - f_4 B^2 p_0^2 + f_5 \frac{B}{C_2} p_3^2 = 0 \tag{24}$$

It is important to note here that the above-mentioned dispersion relations for the Sultana Dyer metric are very different than the dispersion relations obtained for the Schwarzschild metric. For Schwarzschild case the dispersion relations do not contain terms with $p_0 p_1$ or $p_1 p_3^2$. Secondly, in the dispersion relations for Sultana Dyer metric terms with every component of momentum get modified whereas for Schwarzschild metric only the term with p_3 gets modified. Last point, Eq. (23) is cubic and it occurs due to the last term in Eq. (21), while for the Schwarzschild case both the dispersion relations are quadratic.

5. Results

Till now we have discussed how photon dispersion relations get modified due to the presence of non-minimal coupling in Sultana Dyer space-time as compared to Schwarzschild metric. In this part we discuss the effects of these modified dispersion relations on photon trajectory. To see the effects one needs to study the modification to the trajectory of the photon, which can be done through these astrophysical observables — 1) Photon arrival-time at the detector, 2) Photon deflection-angle at the detector. Since we are working with Sultana Dyer black hole — a time-dependent and non-diagonal metric, we prefer deflection-angle over photon arrival-time.

For verifying the existence of non-minimal coupling we have primarily two goals — 1) Put constraint on the non-minimal coupling coefficient, 2) Compare the effects on non-minimally coupled photon trajectory caused by time-independent and time-dependent black hole.

Reaching the above-mentioned goals requires derivation of the the deflection angle of photon for each polarization mode, and then their difference considering time-independent(Schwarzschild) and time-dependent(Sultana Dyer) black hole. For simplicity, we derive the difference in the deflection angle comparison at a radius of photon sphere — $\tilde{r} = 3/2$ as we are interested in the behavior of photons near the black hole.

For Schwarzschild metric at $r = (3/2)r_{\rm H}$ and $\tilde{\eta_0}$ the difference between the deflection angle for (+) and (-) polarization is,

$$\left.\frac{d\phi}{dr}\right|^{+}_{\rm Sch} - \left.\frac{d\phi}{dr}\right|^{+}_{\rm Sch} = \frac{4\beta_0\, b}{9 r_{\rm H}} \left[\left[1 - \frac{4\,\beta_0^2 b^2}{27} \left(1 - \frac{16\tilde{\lambda}}{729} \right) \right]^{-\frac{1}{2}} - \left[1 - \frac{4\,\beta_0^2 b^2}{27} \left(1 + \frac{16\tilde{\lambda}}{729} \right) \right]^{-\frac{1}{2}} \right] . \tag{25}$$

where, $b = \tilde{L}/\tilde{E}$ at fixed $\tilde{r}$ and $\tilde{\eta} = \tilde{\eta}_0$. The LHS of the above equation will be real if and only if,

$$1 - \frac{4\,\beta_0^2 b^2}{27} \left(1 \mp \frac{16\tilde{\lambda}}{729} \right) > 0\,. \tag{26}$$

From this condition, we determine the value of b, provided the values of β_0 and $\tilde{\lambda}$ are known.

Now for Sultana Dyer space-time, the difference between the polarization-dependent deflection angles can be obtained from the cubic (23) and the quadratic dispersion relation (24) separately. For the limit, $\beta_0 >> 1$ and $\tilde{E}/\tilde{L} >> 1$ both Eqs. (23) and (24) get simplified and the cubic dispersion relation becomes entirely independent of the coupling coefficient $\tilde{\lambda}$. This implies the simplified cubic dispersion relation does not hold any imprint of non-minimal coupling. Hence we stress our analysis into the quadratic dispersion relation only. From the simplified quadratic dispersion relation we get the deflection angle:

$$\frac{d\phi}{d\tilde{r}} = \frac{\tilde{L}\left(\frac{C_1^2}{A}+B\right)}{C_2\left(\frac{-\left(2f_4C_1B\tilde{E}\right)\pm\sqrt{\left(2f_4C_1B\tilde{E}\right)^2-4\left(-f_4C_1^2\right)B\tilde{L}^2\left(-f_4B\frac{\tilde{E}^2}{\tilde{L}^2}+\frac{f_5}{C_2}\right)}}{2\left(-f_4C_1^2\right)}+\frac{C_1}{A}\tilde{E}\right)}. \quad (27)$$

The difference between the deflection angle for $(\pm)$ polarizations at fixed $r = (3/2)r_{\rm H}$ and current comoving time $\tilde{\eta_0} = 3$ becomes,

$$\left.\frac{d\phi}{dr}\right|_{SD}^{+} - \left.\frac{d\phi}{dr}\right|_{SD}^{-} \sim \frac{162\,b^2\sqrt{\left(3^6-16\tilde{\lambda}\right)\left(8\tilde{\lambda}+3^6\right)}}{16\,k^3r_{\rm H}^4\left(8\tilde{\lambda}+3^6\right)}\frac{1}{\beta_0^2}. \quad (28)$$

where, $k = H_0/c$. Eqs. (25) and (28) depend on $r_{\rm H}$, β_0, b and $\tilde{\lambda}$. But it is crucial to note that in Schwarzschild space-time the difference between two deflection angles (25) depends on $b\,\beta_0$, whereas in Sultana Dyer space time (28) is proportional to b/β_0. Hence, at this point we stress that photon trajectory gets modified in different manner while travelling near a Schwarzschild or Sultana Dyer black hole. Understanding this fact better requires the value of b and $\tilde{\lambda}$.

From the condition that Eq. (28) will be real if and only if the constraint $\tilde{\lambda} < \frac{729}{16}$ is followed. As $\tilde{\lambda} = \lambda/\alpha_0^2$ we get the constraint of the coupling coefficient λ as: $\lambda < (9/16)r_{\rm H}^2$. This is another important result regarding which we like to stress the following points: First, the coupling constant λ is an intrinsic property of the field and should be independent of the parameters of the model. However, from the above expression, we see that the constraint on λ depends on the mass of the black hole. On the face of it, this might look unphysical. However, it can be pointed that the energy density of the mull dust in Sultana Dyer space-time does not remain positive for all values of η. The condition that the energy density remains positive at the photon radius translates to $\eta < 15r_{\rm H}/4$. Since, we have fixed $\eta = \eta_0$, this leads to the condition on λ. Thus, the above constraint on the coupling parameter λ ensures that the energy density of the electromagnetic field is non-negative. The table below gives the upper limit on λ for black holes with different mass ranges. Finally, it is interesting to compare the bounds on λ compared to obtained in the literature. In Ref.[11], the authors obtained constraints on λ considering photons in Schwarzschild

space-time. Considering signals from radar ranging past the Sun, the authors found $\lambda \sim 1.1 \times 10^{20}$ cm^2, which is about three orders of magnitude more stringent than the one obtained in Ref.[12]. Our analysis shows that $\lambda < 0.563 \times 10^{10}$cm^2 for the Sultana-Dyer black hole of solar-mass. Note that, in the case of signals coming from binary pulsar PSR B1534+12, the bound on $\lambda \sim 0.6 \times 10^{11}$ cm^2.[13]

Black hole mass	$r_{\rm H}$ (cm)	λ (cm^2)
Solar-mass	10^5	0.563×10^{10}
$10^{-3} M_\odot$	10^2	0.563×10^4
$10^{-5} M_\odot$	1	0.563
$10^{-10} M_\odot$	10^{-5}	0.563×10^{-10}

Now using the value of $\tilde{\lambda}$ ranging between $[0, 729/16]$ into condition (26) with the (-) sign we get the constraint $b\beta_0 < 1.8$. Similarly for the (+) sign we get the constraint as $b\beta_0 < 2.598$. However, for the constraint $b\beta_0 < 1.8$ Eq. (26) is valid for both ($\pm$) sign. From this constraint, we have fixed the value of b for a particular value of β_0, where the value of β_0 varies with the mass of the black hole.

Since we have fixed the variables $\tilde{\lambda}$, $r_{\rm H}$, β_0 and b, now we compare Eqs. (25) and (28) by evaluating their values for black holes with mass — $M_\odot$ and $10^{-5} M_\odot$. We evaluate the difference between two deflection angles in Schwarzschild and Sultana Dyer space-time considering two separate cases, 1) $b\beta_0 \sim 1$ and 2) $b\beta_0 \sim 10^{-3}$. Both these cases satisfies the condition $b\beta_0 < 1.8$. All the values has been mentioned in the table below.

Black hole mass	$b\beta_0$	$\Delta\left(\frac{d\phi}{dr}\right)_{\rm Sch}$	$\Delta\left(\frac{d\phi}{dr}\right)_{\rm SD}$
$M_\odot$	1	-1.322×10^{-5}	7.893×10^{-7}
	10^{-3}	-1.032×10^{-14}	7.893×10^{-13}
$10^{-5} M_\odot$	1	-1.322	0.079
	10^{-3}	-1.032^{-9}	7.893×10^{-8}

Inspecting the numbers in the above table we infer one more important key feature: for $b\beta_0 \sim 1$ the modulus of the difference between the deflection angles for Schwarzschild is greater than the same in Sultana Dyer space-time. However, for, $b\beta_0 \sim 10^{-3}$ the difference between the deflection angles for Sultana Dyer black hole is greater. The reason behind this behavior is that the difference between the deflection angle (28) for the Sultana Dyer black hole is proportional to b^2/β_0^2 and the same quantity (25) for the Schwarzschild black hole depends on b and β_0 differently.

As a result, we see the larger the value of b larger the effects due to Schwarzschild, and smaller the value of b the effects are larger for Sultana Dyer black hole. Another crucialfeature is that, the difference between the photon deflection angles for Schwarzschild metric is negative, while for Sultana Dyer it is always positive — that manifests two distinct signatures of two black holes. Similar behavior in the difference of deflection angles has been observed for a black hole of mass $10^{-5}M_{\odot}$ as well.

6. Conclusions and discussions

In this work, we have studied the effects of non-minimal coupling of electromagnetic fields with gravitational field near a cosmological black hole, described by Sultana Dyer metric, and compared them with the effects obtained for a Schwarzschild black hole. To our knowledge, the analysis for non-minimal coupling near a time-dependent black hole has not been explored in literature. For doing the analysis we have applied Eikonal approximation in a local inertial frame assuming that the wavelength and time period of the propagating electromagnetic wave is lesser than the variation of curvature with respect to the space-time. Our study for the Sultana Dyer metric produced a pair of modified dispersion relations which consist of corrections in all components of momentum, while for the Schwarzschild case the modification occurs only in the p_3 component. Besides that, one of the modified dispersion relations for the Sultana Dyer metric is cubic, while for the Schwarzschild metric both the dispersion relations remain quadratic. By calculating photon deflection angle for individual polarization mode in Sultana Dyer space-time we get a bound on λ for a solar mass black hole which is more stringent of order 10^{10} than the bound obtained in Ref.[11]. Through the evaluation of the difference between deflection angles for two polarizations we have found distinct signatures of two black holes — Schwarzschild and Sultana Dyer.

We hope that through observing electromagnetic emission in broad wave-band from different astrophysical events it is plausible to measure the deflection angle of electromagnetic wave and confirm the existence or non-existence of the non-minimal coupling.

For the entire work, we have assumed the time-dependent black hole to be spherically symmetric and non-rotating. To observe the signatures mentioned in this work from astrophysical observation might involve axially symmetric (like Kerr) black hole. Hence, we are investigating the possibilities to extend the analysis for a rotating black-hole space-time.

Acknowledgements

This work is supported by MATRICS-SERB grant. The authors thank the service personnel in India whose untiring work allowed the authors to continue this work during the COVID-19 outbreak.

References

1. S. Jana and S. Shankaranarayanan, Constraints on the non-minimal coupling of electromagnetic field from astrophysical observations (Oct 2021).
2. T. A. Wagner, S. Schlamminger, J. H. Gundlach and E. G. Adelberger, Torsion-balance tests of the weak equivalence principle, *Class. Quant. Grav.* **29**, p. 184002 (2012).
3. A. Prasanna, A new invariant for electromagnetic fields in curved space-time, *Physics Letters A* **37**, 331 (1971).
4. G. W. Horndeski, Conservation of Charge and the Einstein-Maxwell Field Equations, *J. Math. Phys.* **17**, 1980 (1976).
5. H. A. Buchdahl, Theories of gravitation with nonminimal coupling of matter and the gravitational field, *J. Phys. A* **12**, 1037 (1979).
6. G. M. Shore, A Local effective action for photon gravity interactions, *Nucl. Phys. B* **646**, 281 (2002).
7. A. B. Balakin and J. P. S. Lemos, Non-minimal coupling for the gravitational and electromagnetic fields: A General system of equations, *Class. Quant. Grav.* **22**, 1867 (2005).
8. E. Berti *et al.*, Testing General Relativity with Present and Future Astrophysical Observations, *Class. Quant. Grav.* **32**, p. 243001 (2015).
9. D. Psaltis, Probes and Tests of Strong-Field Gravity with Observations in the Electromagnetic Spectrum, *Living Rev. Rel.* **11**, p. 9 (2008).
10. J. Sultana and C. C. Dyer, Cosmological black holes: A black hole in the Einstein- de Sitter universe, *Gen. Rel. Grav.* **37**, 1347 (2005).
11. A. R. Prasanna and S. Mohanty, Constraints on nonminimally coupled curved space electrodynamics from astrophysical observations, *Class. Quant. Grav.* **20**, 3023 (2003).
12. M. L. Bedran and B. Lesche, An example of affine collineation in the Robertson–Walker metric, *Journal of Mathematical Physics* **27**, 2360 (1986).
13. I. H. Stairs, S. E. Thorsett, J. H. Taylor and A. Wolszczan, Studies of the relativistic binary pulsar psr b1534+12: I. Timing analysis, *Astrophys. J.* **581**, 501 (2002).

New partial resummation of the QED effective action

Silvia Pla* and José Navarro-Salas†

Departamento de Fisica Teorica and IFIC, Centro Mixto Universidad de Valencia-CSIC, Facultad de Fisica, Universidad de Valencia, Burjassot-46100, Valencia, Spain

**E-mail: silvia.pla@uv.es*

†E-mail: jnavarro@ific.uv.es

We explain a conjecture which states that the proper-time series expansion of the one-loop effective Lagrangian of quantum electrodynamics can be partially summed in all terms containing the field-strength invariants $\mathcal{F} = \frac{1}{4}F_{\mu\nu}F^{\mu\nu}(x)$, $\mathcal{G} = \frac{1}{4}\tilde{F}_{\mu\nu}F^{\mu\nu}(x)$. This summation is encapsulated in a factor with the same form as the (spacetime-dependent) Heisenberg-Euler Lagrangian density. We also discuss some implications and a possible extension in presence of gravity. We will focus on the scalar field case.

Keywords: One-loop effective Lagrangian; Partial resummation; Heisenberg-Euler Lagrangian; asymptotic expansion.

1. Introduction

In this contribution[a], we give more details on a conjecture proposed in Ref.[2], that claims that it is possible to make a factorization in the proper-time series of the one-loop effective Lagrangian of quantum electrodynamics (QED) that captures all explicit dependence on the field-strength invariants $\mathcal{F}$ and $\mathcal{G}$. The text is organized as follows. First, in Sections 1.1 and 1.2 we review some basics about the one-loop effective Lagrangian and its proper-time series expansion. We focus on the scalar QED effective Lagrangian and also on the corresponding effective Lagrangian in presence of gravity. We emphasize the existence of a modified, asymptotic expansion (also called R-summed expansion) in the gravitational scenario. In Section 2 we explain our conjecture for scalar QED (a sketch for spinor QED is given in Ref.[2]). We also check it up to order $n = 6$ (the last available coefficient in the literature). In Section 3 we explain some physical consequences that can be easily obtained using our results. Finally, in Section 4 we give a brief summary of this contribution. For simplicity, we will focus on the scalar QED case. In Appendices A and B we give the coefficients of the usual and the resummed proper-time series expansions for scalar QED.

1.1. *Heisenberg-Euler Lagrangian and the proper-time series expansion*

The Heisenberg-Euler Lagrangian captures the one-loop quantum corrections to the classical electromagnetic Lagrangian for quantum electrodynamics. It is obtained

[a]Expanded version of the talk given by S.P. in the Sixteenth Marcel Grossmann Meeting (2021).

by integrating out the matter degrees of freedom, keeping the strength of the electromagnetic background constant.[3,4] The intrinsic nonlinearities of the quantum corrections have important implications, such as pair creation from vacuum and vacuum polarization,[5] or light by light scattering or vacuum polarization (see, for example, Ref.[6] and references therein). It was originally derived by Heisenberg and Euler[7] for spinor QED in 1936. A few months later, it was obtained by Weisskopf[8] for scalar QED (see Ref.[9] for a charged spin 1 field). These two theories are regarded as the paradigm of the modern effective theories in Quantum Field Theory. From now on, we will focus on the scalar QED Heisenberg-Euler Lagrangian. It is usually written in terms of the proper time parameter s, but, apart from this, it be written in several ways. By convenience, here we use the following form borrowed from Ref.[10]:

$$\mathcal{L}^{(1)}_{scalar} = -\frac{1}{(4\pi)^2}\int_0^\infty \frac{ds}{s^3} e^{-im^2 s}\left[\det\left(\frac{esF}{\sinh(esF)}\right)\right]^{1/2}, \tag{1}$$

where $F \equiv F^{\mu}_{\ \nu}$. As we have stressed before, this result was obtained for constant electromagnetic backgrounds, but for arbitrary background configurations the form of the effective action is, in general, *unknown*.

However, it is possible to build a general, asymptotic expansion for the one-loop effective Lagrangian in terms of the proper-time parameter. Here we use the expansion given in Ref.[11], obtained via the string-inspired method in the world-line formalism (see Refs.[11–14] and also Refs.[15,16]), namely

$$\mathcal{L}^{(1)}_{scalar} = \int_0^\infty \frac{ds}{s} e^{-im^2 s} g(x; is)\ , \tag{2}$$

with

$$g(x;is) = \frac{1}{(4\pi is)^2}\sum_{n=0}^{\infty}\frac{(-is)^n}{n!}O_n(x) \longrightarrow \left\{\begin{array}{c} O_0 = 1, \quad O_1 = 0, \\ O_2 = -\frac{e^2}{6}F_{\kappa\lambda}F^{\kappa\lambda} \\ O_3 = -\frac{e^2}{20}\partial_\mu F_{\kappa\lambda}\partial^\mu F^{\kappa\lambda} \\ (\cdots) \end{array}\right\} \tag{3}$$

The expression above consist of an expansion in the number of external fields and the number of derivatives, and captures some general behaviour of the unknown formal one-loop effective Lagrangian for arbitrary backgrounds. The coefficients O_n are gauge-invariant and have mass dimension $2n$. They have been obtained up to 12th adiabatic order ($n = 6$), and their length grows substantially with n. In Appendix A can be found the complete list of coefficients. We would like to stress that this expansion presents some clear advantages with respect to other proposals. In particular, it is written in the most compact form possible. Using the Bianchi identity, the antisymmetry of $F_{\mu\nu}$ and also integration by parts, the gauge-invariant coefficients O_n have been simplified to the so-called minimal basis.[17]

1.2. *One loop effective Lagrangian in presence of gravity*

The role of the classical electromagnetic background can be replaced by a gravitational field, which is naturally coupled to quantized matter fields. In other words, it is possible to obtain quantum corrections to the Einstein-Hilbert Lagrangian induced by a quantum scalar field. In general, the form of the effective action is not known, but it is also possible to make an asymptotic expansion of it,[18] as in the previous case.

In this context, the expansion of the one-loop effective Lagrangian is usually obtained in terms of the Schwinger proper-time expansion of the Feynman propagator,[5] closely connected to the heat-kernel expansion and related techniques.[19] Specifically, the one loop effective Lagrangian in terms of the heat-kernel reads

$$\mathcal{L}^{(1)}_{scalar} = \int_0^\infty \frac{ds}{s} e^{-im^2 s} \langle x, s|x, 0\rangle \ . \tag{4}$$

In this expression, $\langle x', s|x, 0\rangle$ is the kernel or "transition amplitude" associated to the Feynman propagator, which obeys a Schrodinger-type equation in terms of the proper time parameter with appropriated boundary conditions, and $\langle x, s|x, 0\rangle$ represents the coincident limit $x' \to x$. In general, the kernel admits a general asymptotic expansion in terms of the proper time parameter, i.e.,

$$\langle x, s|x, 0\rangle = \frac{1}{(4\pi is)^2} \sum_{n=0}^{\infty} (is)^n a_n(x). \tag{5}$$

The coefficients $a_n(x)$ are called DeWitt coefficients. These quantities of mass dimension $2n$ are local, covariant and gauge invariant.[18,20] We recall here that for QED, the expansion $g(x; is)$ [see Eq. (5)] coincides with the heat kernel expansion, up to total derivatives

$$g(x; is) = \langle x, s|x, 0\rangle + total\ derivatives. \tag{6}$$

Although the general form of the gravitational one-loop effective Lagrangian is not known, there is one special case where it can be evaluated exactly[21]: the Static Einstein Universe. In this case, the kernel $\langle x, s|x, 0\rangle$ reads

$$\langle x, s|x, 0\rangle = \frac{e^{-i(\xi - \frac{1}{6})Rs}}{(4\pi is)^2}. \tag{7}$$

where R is the Ricci scalar and ξ is the coupling constant to the scalar field. This is a non-perturbative result, that involves all powers of the proper time parameter s. From this result, Parker and Toms proposed in 1985 a new asymptotic expansion for the kernel,[22] defined as

$$\langle x, s|x, 0\rangle = \frac{e^{-i(\xi - \frac{1}{6})R(x)s}}{(4\pi is)^2} \sum_{n=0}^{\infty} (is)^n \bar{a}_n(x). \tag{8}$$

The new adiabatic expansion in terms of the coefficients $\bar{a}_n(x)$ has an important advantage with respect to the previous one: it does not contain any term that vanishes when $R(x)$ is replaced by zero (see Ref.[23] for detailed proof). In other words, the non-perturbative exponential factor captures the exact dependence on the Ricci scalar in the generic Schwinger-DeWitt asymptotic expansion. This factorization has some important physical consequences, related, for example, to the effective dynamics of the Universe[24] or to the curvature dependence in the running of the gauge coupling constants.[25]

The key idea is that, from a solvable case, it is possible to extract a non-perturbative factor in the general asymptotic expansion of the one-loop effective Lagrangian that completely captures its dependence with the curvature scalar $R(x)$. In this context, a natural question arises: is it possible to find a similar factorization for the QED Lagrangian?

2. The conjecture

Based on the gravitational case, we propose the following conjecture for the electromagnetic case[2]: *The proper-time asymptotic expansion of the QED effective Lagrangian admits an exact resummation in all terms involving the field-strength invariants* $\mathcal{F} = \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ *and* $\mathcal{G} = \frac{1}{4}\tilde{F}_{\mu\nu}F^{\mu\nu}$. *The form of the factor is just the Heisenberg-Euler Lagrangian for QED, where the electric and magnetic fields depend arbitrarily on space-time coordinates*, namely

$$\mathcal{L}^{(1)}_{scalar} = \int_0^\infty \frac{ds}{s} e^{-im^2 s} \left[\det\left(\frac{esF(x)}{\sinh(esF(x))} \right) \right]^{1/2} \bar{g}(x; is) \tag{9}$$

[to compare with (5)], where the new asymptotic expansion $\bar{g}(x; is)$ reads

$$\bar{g}(x; is) = \frac{1}{(4\pi is)^2} \sum_{n=0}^{\infty} \frac{(-is)^n}{n!} \bar{O}_n(x), \tag{10}$$

and where the coefficients $O_n(x)$ do not have terms that vanish when the electromagnetic invariants $\mathcal{F}(x)$ and $\mathcal{G}(x)$ are replaced by zero. The asymptotic expansions $g(x; is)$ and $\bar{g}(x; is)$ are related by

$$g(x; is) = \left[\det\left(\frac{esF(x)}{\sinh(esF(x))} \right) \right]^{1/2} \bar{g}(x; is). \tag{11}$$

In summary, we have proposed a new adiabatic expansion $\bar{g}(x; is)$ that can be built from the previous one via Eq. (11), and we have conjectured that, if the pre-factor is just the Euler Heisenberg Lagrangian, the new coefficients $\bar{O}_n$ associated with the new expansion will not have terms that vanish when the electromagnetic invariants are replaced by zero.

In the last part of the section, we are going to compute the new coefficients and check that, effectively, our conjecture is satisfied. First, we should expand the

Heisenberg-Euler determinant in terms of the proper time,

$$\left[\det\left(\frac{esF(x)}{\sinh(esF(x))}\right)\right]^{1/2} \sim 1 + U_2(x)(-is)^2 + U_4(x)(-is)^4 + U_6(x)(-is)^6 + \cdots, \tag{12}$$

where

$$U_2(x) = \tfrac{e^2}{12}\,\mathrm{Tr}(F^2), \tag{13}$$

$$U_4(x) = \tfrac{e^4}{288}\,\mathrm{Tr}(F^2)^2 + \tfrac{e^4}{360}\,\mathrm{Tr}(F^4), \tag{14}$$

$$U_6(x) = \tfrac{e^6}{10368}\,\mathrm{Tr}(F^2)^3 + \tfrac{e^6}{4320}\,\mathrm{Tr}(F^2)\,\mathrm{Tr}(F^4) + \tfrac{e^6}{5670}\,\mathrm{Tr}(F^6). \tag{15}$$

we note here that the trace $\mathrm{Tr}(F^{2n})$ can be always written in terms of powers of the electromagnetic invariants $\mathcal{F}$ and $\mathcal{G}$. For example, $\mathrm{Tr}(F^2) = -F_{\mu\nu}F^{\mu\nu} = -4\mathcal{F}$. Now, inserting expansions (12), (3) and (10) in Eq. (11) we find

$$O_0 + O_1 + (-is) + \frac{O_2}{2}(-is)^2 ... = (1 + U_2(-is)^2 + ...)\cdot(\bar{O}_0 + \bar{O}_1(-is) + \frac{\bar{O}_2}{2}(-is)^2 + ...) \tag{16}$$

and equating order by order we arrive to $\bar{O}_0 = 1$, $\bar{O}_1 = 0$,

$$\bar{O}_2 = O_2 - 2U_2 = 0, \tag{17}$$

$$\bar{O}_3 = O_3, \tag{18}$$

$$\bar{O}_4 = O_4 - 4!\,U_4, \tag{19}$$

$$\bar{O}_5 = O_4 - 20\,U_2 O_3, \tag{20}$$

$$\bar{O}_6 = O_6 - 6!\,U_6 - 30\,U_2\bar{O}_4. \tag{21}$$

Finally, substituting the values of the coefficients U_n and O_n in the previous equations we arrive at the explicit values of $\bar{O}_n$, which can be found in Appendix B. It is easy to see that our conjecture is satisfied. Comparing the coefficients O_n (Appendix A) with $\bar{O}_n$, one sees that all terms going with $\mathrm{Tr}(F^{2n})$ [or, equivalently, with $\mathcal{F}$ or $\mathcal{G}$] have disappeared.

2.1. *Some comments about the proper-time series expansion*

The asymptotic expansion (proper-time series expansion) that we are using here [see Eqs. (3) and (5)] is an expansion that group terms with the same mass dimension (remember that the coefficients O_n and a_n have mass dimension $2n$). This expansion is closely related with the adiabatic expansion of the scalar field modes, widely used in cosmology [see Ref.[26] for the equivalence in curved spacetime, and[27] for the adiabatic counterpart of the R-summed expansion given in (8)]. In its normal form, it is commonly used to renormalize physical quantities. It is important to recall that this expansion is not a derivative expansion (i.e., an expansion that groups terms with the same number of derivatives; see, for example, Ref.[28]). For a detailed explanation of the differences between both expansions see Ref.[3].

3. Main consequences

Our factorization, allow us to predict or reproduce some interesting physical consequences. First, it is possible to find some exactly solvable electromagnetic backgrounds. For electric and magnetic fields pointing in the $\hat{z}$ direction that depend arbitrarily on the light-cone coordinate $x^+ = (t + z)$, we find that $\bar{O}_0 = 1$ and $\bar{O}_{n>0} = 0$. Therefore, the exact form of the unrenormalized effective Lagrangian should be (in agreement with Refs.[29–31])

$$\mathcal{L}^{(1)}_{scalar} = -\frac{1}{16\pi^2}\int_0^\infty \frac{ds}{s^3} e^{-im^2 s} \frac{e^2 s^2 E(x^+)B(x^+)}{\sinh esE(x^+)\sin esB(x^+)}. \tag{22}$$

For a single electromagnetic plane wave, $\mathcal{F}(x) = 0$ and $\mathcal{G}(x) = 0$, therefore, according to our conjecture, the one-loop effective action trivially vanishes, in agreement with the result obtained in Ref.[5] by other methods. Second, our results seem to be consistent in presence of gravity. [b]. In this case, it is possible to perform a double factorization (the exponential $R(x)$ factorization is ensured[23]),

$$g(x, is) = e^{-is(\xi-\frac{1}{6})R(x)}\left[\det\left(\frac{esF(x)}{\sinh(esF(x))}\right)\right]^{1/2}\tilde{g}(x; is). \tag{23}$$

Finally, the conjecture allows us to make some general predictions regarding the Schwinger's formula for the pair production rate. The factorization suggests that the poles of the imaginary part of the one-loop effective Lagrangian are located at the same points as in the constant electric field case $\tau_n = n\pi/|eE(x)|$.

4. Summary and conclusions

The one-loop (scalar) QED effective-Lagrangian is, in general, unknown. However, it is possible to obtain an adiabatic expansion of it in terms of the functions $g(x; is)$ defined in (2). We have proposed an alternative adiabatic expansion $\bar{g}(x; is)$, which encapsulates in a global pre-factor all its dependence on the electromagnetic invariants $\mathcal{F}(x)$ and $\mathcal{G}(x)$. The form of the non-perturbative factor involved in this partial resummation is just the Heisenberg-Euler Lagrangian for QED, but with the electric and magnetic fields depending arbitrarily on spacetime coordinates. The new expansion does not contain terms that vanish when $\mathcal{F}(x)$ and $\mathcal{G}(x)$ are replaced by zero. This factorization allows us to obtain some exact solutions and seems to be consistent in presence of gravity. In this contribution, we have focused on the scalar case, but our conjecture is also valid for the spin-$\frac{1}{2}$ case. We expect our results to be consistent for non-abelian gauge backgrounds. Most of the computations in this paper have been done with the help of the xAct package of the *Mathematica Software*.[34] This work was supported in part by Spanish Ministerio de Economia, Industria y Competitividad Grants No. FIS2017-84440-C2-1-P(MINECO/FEDER, EU), No. FIS2017-91161-EXP and the project PROMETEO/2020/079 (Generalitat Valenciana). S. P. is supported by a Ph.D. fellowship, Grant No. FPU16/05287.

[b] For $\nabla_\rho F^{\mu\nu} = 0$ the factorization of the Euler-Heisenberg Lagrangian was found in Refs.[32, 33].

Appendix A

Here we give the coefficients O_n of the original expansion $g(x;is)$. $O_0 = 1$,

$$O_1 = 0,\ \ O_2 = -\tfrac{\mathbf{e^2}}{\mathbf{6}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda},\ \ O_3 = -\tfrac{e^2}{20}\partial_\mu F_{\kappa\lambda}\partial^\mu F^{\kappa\lambda}, \tag{A.1}$$

$$\begin{aligned} O_4 = {} & \tfrac{\mathbf{e^4}}{\mathbf{15}}\mathbf{F}_\kappa{}^\mu\mathbf{F}^{\kappa\lambda}\mathbf{F}_\lambda{}^\nu\mathbf{F}_{\mu\nu} + \tfrac{\mathbf{e^4}}{\mathbf{12}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu} \\ & -\tfrac{e^2}{70}\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\nu\partial^\mu F^{\kappa\lambda}, \end{aligned} \tag{A.2}$$

$$\begin{aligned} O_5 = {} & \tfrac{2e^4}{7}F^{\kappa\lambda}F^{\mu\nu}\partial_\lambda F_{\nu\rho}\partial_\mu F_\kappa{}^\rho - \tfrac{4e^4}{63}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\lambda F^{\nu\rho}\partial_\mu F_{\nu\rho} \\ & -\tfrac{e^4}{9}F_\kappa{}^\mu F^{\kappa\lambda}F^{\nu\rho}\partial_\mu\partial_\lambda F_{\nu\rho} - \tfrac{16e^4}{63}F^{\kappa\lambda}F^{\mu\nu}\partial_\mu F_\kappa{}^\rho\partial_\nu F_{\lambda\rho} \\ & +\tfrac{5e^4}{18}F^{\kappa\lambda}F^{\mu\nu}\partial_\rho F_{\mu\nu}\partial^\rho F_{\kappa\lambda} + \tfrac{34e^4}{189}F^{\kappa\lambda}F^{\mu\nu}\partial_\nu F_{\lambda\rho}\partial^\rho F_{\kappa\mu} \\ & +\tfrac{25e^4}{189}F^{\kappa\lambda}F^{\mu\nu}\partial_\rho F_{\lambda\nu}\partial^\rho F_{\kappa\mu} + \tfrac{4e^4}{21}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\rho F_{\mu\nu}\partial^\rho F_\lambda{}^\nu \\ & +\tfrac{\mathbf{e^4}}{\mathbf{12}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda}\partial_\rho\mathbf{F}_{\mu\nu}\partial^\rho\mathbf{F}^{\mu\nu} - \tfrac{e^2}{252}\partial_\rho\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\rho\partial^\nu\partial^\mu F^{\kappa\lambda}, \end{aligned} \tag{A.3}$$

$$\begin{aligned} O_6 = {} & -\tfrac{\mathbf{8}}{\mathbf{63}}\mathbf{F}_\kappa{}^\mu\mathbf{F}^{\kappa\lambda}\mathbf{F}_\lambda{}^\nu\mathbf{F}_\mu{}^\rho\mathbf{F}_\nu{}^\sigma\mathbf{F}_{\rho\sigma} - \tfrac{\mathbf{1}}{\mathbf{6}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda}\mathbf{F}_\mu{}^\rho\mathbf{F}^{\mu\nu}\mathbf{F}_\nu{}^\sigma\mathbf{F}_{\rho\sigma} \\ & -\tfrac{\mathbf{5}}{\mathbf{72}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu}\mathbf{F}_{\rho\sigma}\mathbf{F}^{\rho\sigma} + \tfrac{397}{3465}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\mu F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma} \\ & +\tfrac{1}{63}F^{\kappa\lambda}F^{\mu\nu}\partial_\mu\partial_\kappa F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma} + \tfrac{187}{1260}\partial_\mu F^{\rho\sigma}\partial^\mu F^{\kappa\lambda}\partial_\nu F_{\rho\sigma}\partial^\nu F_{\kappa\lambda} \\ & -\tfrac{1079}{3465}F^{\kappa\lambda}\partial_\mu\partial_\lambda F_{\rho\sigma}\partial_\nu F^{\rho\sigma}\partial^\nu F_\kappa{}^\mu - \tfrac{43}{3465}F^{\kappa\lambda}\partial_\mu F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma}\partial^\nu F_\kappa{}^\mu \\ & -\tfrac{4}{35}F^{\kappa\lambda}\partial_\lambda F^{\rho\sigma}\partial_\nu\partial_\mu F_{\rho\sigma}\partial^\nu F_\kappa{}^\mu - \tfrac{101}{3465}\partial_\lambda F_{\rho\sigma}\partial^\mu F^{\kappa\lambda}\partial_\nu F_\mu{}^\sigma\partial^\rho F_\kappa{}^\nu \\ & +\tfrac{43}{3465}\partial_\mu F_{\nu\sigma}\partial^\mu F^{\kappa\lambda}\partial_\rho F_\lambda{}^\sigma\partial^\rho F_\kappa{}^\nu - \tfrac{2}{21}F^{\kappa\lambda}\partial_\rho F_\mu{}^\sigma\partial^\rho F^{\mu\nu}\partial_\sigma\partial_\nu F_{\kappa\lambda} \\ & -\tfrac{4}{21}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho\partial_\lambda F_{\mu\nu}\partial^\sigma F_\kappa{}^\rho + \tfrac{1486}{3465}\partial_\mu F_{\nu\sigma}\partial^\mu F^{\kappa\lambda}\partial^\rho F_\kappa{}^\nu\partial^\sigma F_{\lambda\rho} \\ & +\tfrac{428}{1155}\partial^\mu F^{\kappa\lambda}\partial^\rho F_\kappa{}^\nu\partial_\sigma F_{\mu\nu}\partial^\sigma F_{\lambda\rho} + \tfrac{3}{35}\partial_\mu F_\kappa{}^\nu\partial^\mu F^{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F_\lambda{}^\rho \\ & +\tfrac{6}{35}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\nu F_{\mu\rho}\partial^\sigma F_\lambda{}^\rho + \tfrac{13}{42}F^{\kappa\lambda}\partial^\rho F^{\mu\nu}\partial_\sigma\partial_\rho F_{\kappa\lambda}\partial^\sigma F_{\mu\nu} \\ & -\tfrac{2}{35}\partial^\mu F^{\kappa\lambda}\partial^\nu F_{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F_\mu{}^\rho + \tfrac{64}{231}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma F_\mu{}^\rho \\ & +\tfrac{118}{693}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma F_\mu{}^\rho - \tfrac{302}{3465}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma F_\mu{}^\rho \\ & +\tfrac{1388}{3465}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma F_\mu{}^\rho + \tfrac{344}{1155}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma F_\mu{}^\rho \\ & +\tfrac{398}{3465}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma F_\mu{}^\rho - \tfrac{62}{165}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\lambda F_{\mu\rho}\partial^\sigma F_\nu{}^\rho \\ & +\tfrac{76}{693}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\mu F_{\lambda\rho}\partial^\sigma F_\nu{}^\rho - \tfrac{326}{1155}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\rho F_{\lambda\mu}\partial^\sigma F_\nu{}^\rho \\ & +\tfrac{1}{40}\partial_\mu F_{\kappa\lambda}\partial^\mu F^{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F^{\nu\rho} + \tfrac{1}{3}F^{\kappa\lambda}\partial^\mu F_{\kappa\lambda}\partial_\sigma\partial_\mu F_{\nu\rho}\partial^\sigma F^{\nu\rho} \\ & -\tfrac{1}{15}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\mu\partial_\lambda F_{\nu\rho}\partial^\sigma F^{\nu\rho} - \tfrac{2}{105}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\mu F_{\nu\rho}\partial^\sigma\partial_\lambda F^{\nu\rho} \\ & +\tfrac{43}{231}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma\partial_\mu F_\kappa{}^\rho - \tfrac{43}{231}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma\partial_\mu F_\kappa{}^\rho \\ & +\tfrac{1}{6}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho F_{\mu\nu}\partial^\sigma\partial^\rho F_{\kappa\lambda} + \tfrac{103}{693}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma\partial^\rho F_{\kappa\mu} \\ & +\tfrac{46}{693}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma\partial^\rho F_{\kappa\mu} + \tfrac{2}{21}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\rho F_{\mu\nu}\partial^\sigma\partial^\rho F_\lambda{}^\nu \\ & +\tfrac{\mathbf{1}}{\mathbf{28}}\mathbf{F}_{\kappa\lambda}\mathbf{F}^{\kappa\lambda}\partial_\sigma\partial_\rho\mathbf{F}_{\mu\nu}\partial^\sigma\partial^\rho\mathbf{F}^{\mu\nu} - \tfrac{1}{924}\partial_\sigma\partial_\rho\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\sigma\partial^\rho\partial^\nu\partial^\mu F^{\kappa\lambda}. \end{aligned} \tag{A.4}$$

The terms that have to disappear when computing the modified expansion $\bar{g}(x;is)$ for our conjecture to be satisfied appeared highlighted.

Appendix B

Here we give the coefficients $\bar{O}_n$ of the $(\mathcal{F},\mathcal{G})-$summed expansion $\bar{g}(x;is)$, $\bar{O}_0=1$,

$$\bar{O}_1=0,\ \ \bar{O}_2=0,\ \ \bar{O}_3=-\tfrac{e^2}{20}\partial_\mu F_{\kappa\lambda}\partial^\mu F^{\kappa\lambda}, \tag{B.1}$$

$$\bar{O}_4=-\tfrac{e^2}{70}\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\nu\partial^\mu F^{\kappa\lambda}, \tag{B.2}$$

$$\begin{aligned}\bar{O}_5=&\tfrac{2e^4}{7}F^{\kappa\lambda}F^{\mu\nu}\partial_\lambda F_{\nu\rho}\partial_\mu F_\kappa{}^\rho-\tfrac{4e^4}{63}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\lambda F^{\nu\rho}\partial_\mu F_{\nu\rho}\\&-\tfrac{e^4}{9}F_\kappa{}^\mu F^{\kappa\lambda}F^{\nu\rho}\partial_\mu\partial_\lambda F_{\nu\rho}-\tfrac{16e^4}{63}F^{\kappa\lambda}F^{\mu\nu}\partial_\mu F_\kappa{}^\rho\partial_\nu F_{\lambda\rho}\\&+\tfrac{5e^4}{18}F^{\kappa\lambda}F^{\mu\nu}\partial_\rho F_{\mu\nu}\partial^\rho F_{\kappa\lambda}+\tfrac{34e^4}{189}F^{\kappa\lambda}F^{\mu\nu}\partial_\nu F_{\lambda\rho}\partial^\rho F_{\kappa\mu}\\&+\tfrac{25e^4}{189}F^{\kappa\lambda}F^{\mu\nu}\partial_\rho F_{\lambda\nu}\partial^\rho F_{\kappa\mu}+\tfrac{4e^4}{21}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\rho F_{\mu\nu}\partial^\rho F_\lambda{}^\nu\\&-\tfrac{e^2}{252}\partial_\rho\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\rho\partial^\nu\partial^\mu F^{\kappa\lambda},\end{aligned} \tag{B.3}$$

$$\begin{aligned}\bar{O}_6=&+\tfrac{397}{3465}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\mu F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma}+\tfrac{1}{63}F^{\kappa\lambda}F^{\mu\nu}\partial_\mu\partial_\kappa F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma}\\&+\tfrac{187}{1260}\partial_\mu F^{\rho\sigma}\partial^\mu F^{\kappa\lambda}\partial_\nu F_{\rho\sigma}\partial^\nu F_{\kappa\lambda}-\tfrac{1079}{3465}F^{\kappa\lambda}\partial_\mu\partial_\lambda F_{\rho\sigma}\partial_\nu F^{\rho\sigma}\partial^\nu F_\kappa{}^\mu\\&-\tfrac{43}{3465}F^{\kappa\lambda}\partial_\mu F^{\rho\sigma}\partial_\nu\partial_\lambda F_{\rho\sigma}\partial^\nu F_\kappa{}^\mu-\tfrac{4}{35}F^{\kappa\lambda}\partial_\lambda F^{\rho\sigma}\partial_\nu\partial_\mu F_{\rho\sigma}\partial^\nu F_\kappa{}^\mu\\&-\tfrac{101}{3465}\partial_\lambda F_{\rho\sigma}\partial^\mu F^{\kappa\lambda}\partial_\nu F_\mu{}^\sigma\partial^\rho F_\kappa{}^\nu+\tfrac{43}{3465}\partial_\mu F_{\nu\sigma}\partial^\mu F^{\kappa\lambda}\partial_\rho F_\lambda{}^\sigma\partial^\rho F_\kappa{}^\nu\\&-\tfrac{2}{21}F^{\kappa\lambda}\partial_\rho F_\mu{}^\sigma\partial^\rho F^{\mu\nu}\partial_\sigma\partial_\nu F_{\kappa\lambda}\tfrac{4}{21}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho\partial_\lambda F_{\mu\nu}\partial^\sigma F_\kappa{}^\rho\\&+\tfrac{1486}{3465}\partial_\mu F_{\nu\sigma}\partial^\mu F^{\kappa\lambda}\partial^\rho F_\kappa{}^\nu\partial^\sigma F_{\lambda\rho}+\tfrac{428}{1155}\partial^\mu F^{\kappa\lambda}\partial^\rho F_\kappa{}^\nu\partial_\sigma F_{\mu\nu}\partial^\sigma F_{\lambda\rho}\\&+\tfrac{3}{35}\partial_\mu F_\kappa{}^\nu\partial^\mu F^{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F_\lambda{}^\rho+\tfrac{6}{35}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\nu F_{\mu\rho}\partial^\sigma F_\lambda{}^\rho\\&+\tfrac{13}{42}F^{\kappa\lambda}\partial^\rho F^{\mu\nu}\partial_\sigma\partial_\rho F_{\kappa\lambda}\partial^\sigma F_{\mu\nu}-\tfrac{2}{35}\partial^\mu F^{\kappa\lambda}\partial^\nu F_{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F_\mu{}^\rho\\&+\tfrac{64}{231}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma F_\mu{}^\rho+\tfrac{118}{693}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma F_\mu{}^\rho\\&-\tfrac{302}{3465}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma F_\mu{}^\rho+\tfrac{1388}{3465}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma F_\mu{}^\rho\\&+\tfrac{344}{1155}F^{\kappa\lambda}\partial_\kappa F^{\mu\nu}\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma F_\mu{}^\rho+\tfrac{398}{3465}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma F_\mu{}^\rho\\&-\tfrac{62}{165}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\lambda F_{\mu\rho}\partial^\sigma F_\nu{}^\rho+\tfrac{76}{693}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\mu F_{\lambda\rho}\partial^\sigma F_\nu{}^\rho\\&-\tfrac{326}{1155}F^{\kappa\lambda}\partial^\nu F_\kappa{}^\mu\partial_\sigma\partial_\rho F_{\lambda\mu}\partial^\sigma F_\nu{}^\rho+\tfrac{1}{40}\partial_\mu F_{\kappa\lambda}\partial^\mu F^{\kappa\lambda}\partial_\sigma F_{\nu\rho}\partial^\sigma F^{\nu\rho}\\&+\tfrac{1}{3}F^{\kappa\lambda}\partial^\mu F_{\kappa\lambda}\partial_\sigma\partial_\mu F_{\nu\rho}\partial^\sigma F^{\nu\rho}-\tfrac{1}{15}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\mu\partial_\lambda F_{\nu\rho}\partial^\sigma F^{\nu\rho}\\&-\tfrac{2}{105}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\mu F_{\nu\rho}\partial^\sigma\partial_\lambda F^{\nu\rho}+\tfrac{43}{231}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\lambda F_{\nu\rho}\partial^\sigma\partial_\mu F_\kappa{}^\rho\\&-\tfrac{43}{231}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma\partial_\mu F_\kappa{}^\rho+\tfrac{1}{6}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho F_{\mu\nu}\partial^\sigma\partial^\rho F_{\kappa\lambda}\\&+\tfrac{103}{693}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\nu F_{\lambda\rho}\partial^\sigma\partial^\rho F_{\kappa\mu}+\tfrac{46}{693}F^{\kappa\lambda}F^{\mu\nu}\partial_\sigma\partial_\rho F_{\lambda\nu}\partial^\sigma\partial^\rho F_{\kappa\mu}\\&+\tfrac{2}{21}F_\kappa{}^\mu F^{\kappa\lambda}\partial_\sigma\partial_\rho F_{\mu\nu}\partial^\sigma\partial^\rho F_\lambda{}^\nu-\tfrac{1}{924}\partial_\sigma\partial_\rho\partial_\nu\partial_\mu F_{\kappa\lambda}\partial^\sigma\partial^\rho\partial^\nu\partial^\mu F^{\kappa\lambda}.\end{aligned} \tag{B.4}$$

References

1. S. Pla, *New partial resummation of the QED effective action*, contribution to the Marcel Grossman Meeting (MG16), `https://indico.icranet.org/event/1/contributions/979/`
2. J. Navarro-Salas and S. Pla, Phys. Rev. D **103**, L081702 (2021).
3. G. D. Dunne, in *From Fields to Strings: Circumnavigating Theoretical Physics*, edited by M. Shifman, A. Vainshtein and J. Wheater (World Scientific, 2005), pp. 445-522; arXiv:hep-th/0406216.
4. M. D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, Cambridge, England, 2014).
5. J. Schwinger, Phys. Rev. **82**, 664 (1951).
6. G. V. Dunne, Int. J. Mod. Phys. A **27**, 1260004 (2012); F. Karbstein, *Particles* **3**, 39 (2020).
7. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936). English translation: arXiv:physics/0605038.
8. V. Weisskopf, The electrodynamics of the vacuum based on the quantum theory of the electron, Kong. Dans. Vid. Selsk. Math-fys. Medd. **14N6** (1936).
9. V. S. Vanyashin and M. V. Terent'ev, Sov. Phys. JETP **21**, 375 (1965).
10. L. Parker and D. J. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Fields and Gravity* (Cambridge University Press, Cambridge, England, 2009).
11. D. Fliegner, P. Haberl, M. G. Schmidt and C. Schubert, Ann. Phys. (N.Y.) **264**, 51 (1998).
12. D. Fliegner, M. G. Schmidt and C. Schubert, Nucl. Instrum. Methods Phys. Res., Sect. A **389**, 374 (1997).
13. D. Fliegner, M. G. Schmidt and C. Schubert, Z. Phys. C **64**, 111 (1994).
14. M. Reuter, M. G. Schmidt, C. Schubert, Ann. Phys. (N.Y.) **259**, 313 (1997).
15. M. G. Schmidt and C. Schubert, Phys. Lett. B **318**, 438 (1993).
16. Z. Bern and D. A. Kosower, Phys. Rev. Lett. **66**, 1669 (1991).
17. U. Müller, *New Computing Techniques in Physics Research IV* (World Scientific, Singapore, 1996), p. 193; DESY-96-154, arXiv:hep-th/9701124.
18. B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
19. D. V. Vassilevich, Phys. Rep. **388**, 279 (2003).
20. P. B. Gilkey, J. Differ. Geom. **10**, 601 (1975).
21. J.D. Bekenstein and L. Parker, Phys. Rev. D **23**, 2850 (1981).
22. L. Parker and D. J. Toms, Phys. Rev. D **31**, 953 (1985).
23. I. Jack and L. Parker, Phys. Rev. D **31**, 2439 (1985).
24. L. Parker and A. Raval, Phys. Rev. D **60**, 063512 (1999); Phys. Rev. D **62**, 083503 (2000); L. Parker and D. A. T. Vanzella, Phys. Rev. D **69**, 104009 (2004); R.R. Caldwell, W. Komp, L. Parker, and D.A.T. Wanzella, Phys. Rev. D **73**, 023513 (2006); E. Di Valentino, E. V. Linder, and A. Melchiorri, Phys. Rev. D **97**, 043528 (2018); T. Markkanen, S. Nurmi, A. Rajantie, and S. Stopyr, J. High Energy Phys. 06 (2018) 040; E. V. Castro, A. Flachi, P. Ribeiro, and V. Vitagliano, Phys. Rev. Lett. **121**, 221601 (2018); O. Czerwinska, Z. Lalak, and L. Nakonieczny, J. High Energy Phys. 11 (2015) 207.
25. E. Calzetta, I. Jack, and L. Parker, Phys. Rev. Lett. **55**, 1241 (1985).
26. A. del Rio and J. Navarro-Salas, Phys. Rev. D **91**, 064031 (2015); P. Beltrán-Palau, A. del Río, S. Nadal-Gisbert and J. Navarro-Salas, Phys. Rev. D **103**, 105002 (2021).
27. A. Ferreiro, J. Navarro-Salas and S. Pla, Phys. Rev. D **101** 105011 (2020).

28. V. P. Gusynin and I. A. Shovkovy, J. Math. Phys. (N.Y.) **40**, 5406 (1999).
29. H. M. Fried and R. P. Woodard, Phys. Lett. B **524**, 233 (2002).
30. T. N. Tomaras, N. C. Tsamis, and R. P. Woodard, Phys. Rev. D **62**, 125005 (2000).
31. A. Ilderton, J. High Energy Phys. 09 (2014) 166.
32. I.G. Avramidi and G. Fucci, Commun. Math. Phys. **291**, 543 (2009).
33. G. Fucci and I.G. Avramidi, On the Gravitationally Induced Schwinger Mechanism, in *9th Conference on Quantum Field Theory under the Influence of External Conditions (QFEXT 09)*, pp. 485-491 [arXiv:0911.1099].
34. J. M. Martín-García, xAct: Efficient tensor computer algebra for the Wolfram language, `https://www.xAct.es`.

Can a detector detect soft photons

Sanved Kolekar

Indian Institute of Astrophysics,
2nd Block, Koramangala, Bangalore 560034, India
E-mail: sanved.kolekar@iiap.res.in
www.iiap.res.in

Jorma Louko

School of Mathematical Sciences, University of Nottingham,
Nottingham NG7 2RD, UK
E-mail: jorma.louko@nottingham.ac.uk

We explore the question whether the presence of soft photons in a system can have direct or indirect consequences in the outcome of suitably defined quantum processes involving the system. We consider an local quantum detector with dual energy levels and coupled to the background gauge invariant charged scalar field in a flat spacetime. The de-excitation rate for downward transitions for such a system on an inertial trajectory is found to depend on the soft charges corresponding to the radial component of the electric field dressing chosen at the asymptotic boundary. While the excitation rate for upward transitions still vanishes as expected. The implications are discussed.

Keywords: Soft photon, gauge invariance, detector

1. Introduction

The BMS symmetries[1] of asymptotically flat spacetimes found by Bondi, Metzner and Sachs in 1962; the gravitational memory effect[2] first proposed by Zeldovich and Polnarev in 1974 and the Weinberg's soft theorems are related to each other through the so called Universal triad relations.[5–9] When acted upon by a BMS transformation corresponding to a BMS diffeomorphism vector at the asymptotic flat boundary or equivalently due to the passage of a gravitational wave at the asymptotic flat boundary, the vacuum state at the boundary - although classically same before and after the transformation, differs quantum mechanically in the content of soft particles and hence the soft charges defined at the boundary. Analogous to the gravitational case, there is a universal triad for the electromagnetic case as well. The electromagnetic memory effect, the $U(1)$ large gauge transformations and the Weinberg's soft photon theorem are also related to one other through analogous type connections. The significance of these universal triad relations were realised by Hawking, Perry and Strominger[3,4] to conjecture soft hair for black holes. In a recent work, a quantum version of the gravitational memory effect was shown to arise due to the BMS symmetries in a Schwarzschild spacetime.[10,11]

In the related context. we explore the question whether the presence of soft photons in a system can have direct or indirect consequences in the outcome of suitably defined quantum processes involving the system.[12] We consider a local quantum system moving on an inertial trajectory in a flat spacetime and linearly couples to a gauge invariant charged scalar field. The gauge invariant form of the field is as defined by Dirac[13] in 1955 to account for the fact that an electric charge is always accompanied by it electric field. We compute the transition rates for such a system when the system is at rest in an inertial frame and the electric field is given a soft photon dressing. Interestingly, we find that the transition rate of the local quantum system shows a de-excitation for downward transitions which depends on the soft charges corresponding to the dressed electric field. This holds true for a wide class of functions that can be chosen for the soft photon dressing of the electric field. We find that the excitation rate for upward transitions of the system vanishes as expected.

2. The local quantum system and the dressing field

We consider a flat spacetime with background electromagnetic and charged scalar fields. We then use the gauge invariant construction of a charged scalar field as defined by Dirac.[13] The 'dressed' field gauge invariant operator is defined by

$$\Phi(x) = e^{iC(x)}\phi(x), \tag{1}$$

where

$$C(x) = \int d^4x'\, f^a(x,x')A_a(x'), \tag{2}$$

Here $A_a(x')$ is the electromagnetic four vector potential and the two-point function $f^a(x,x')$ that specifies the dressing satisfies the four-dimensional Poisson's equation

$$\partial'_a f^a(x,x') = q\,\delta^4(x-x'). \tag{3}$$

Under the gauge transformation, the phase factor $e^{iC(x)}$ transforms as

$$e^{iC(x)} \to e^{iC(x)}\, e^{-iq\Omega(x)}, \tag{4}$$

Thus rendering the $\Phi(x)$ operator to be gauge invariant. The physical significance of the construction in Eq. (1) is that a charge can never be completely described without its electric field and the exponential phase factor in front of $\phi(x)$ precisely does that. The two-point function $f^a(x,x')$ contains the information about the electric field of the charge which can include soft dressing contributions in addition to the usual Coulombic electric field in the rest frame of the charge. We quantize the system by the usual standard procedures and assume the fields $A_a(x)$ and $\phi(x)$ to be free fields at the lowest order in perturbation theory.

The local quantum system is defined to be a pointlike charged two-level system moving on an pointlike trajectory $x^a(\tau)$ and parametrised by the proper time τ.

The monopole moment $m(\tau)$ of the system couples linearly to the gauge invariant scalar field operator Φ with the interaction Hamiltonian being defined as

$$\mathcal{H}_{int} = \lambda\chi(\tau)\left(m(\tau)\Phi(\tau) + m^{\dagger}(\tau)\Phi^{\dagger}(\tau)\right), \tag{5}$$

where λ is a real-valued coupling constant, $\Phi(\tau)$ is the pull-back of $\Phi(x)$ to the detector's worldline, and the real-valued switching function χ specifies how the interaction is switched on and off. One can check that the $\mathcal{H}_{int}$ is gauge invariant by construction as well as Hermitian as required for the conservation of total charge.

The time independent response function for the local quantum system on an inertial trajectory $t(\tau) = \tau$ and $\mathbf{x} = 0$ coupled to the Minkowski vacuum state in linear order of perturbation theory is determined to be

$$F(E) = \int_{-\infty}^{\infty} ds\, e^{-iEs}\, W_{\phi}(s)\, W_{A}(s), \tag{6}$$

where E is the difference between the energies of excited state and the ground state of the local quantum system. Hence $E > 0$ determines upward transitions, that is excitations, while $E < 0$ determines downward transitions, that is de-excitations. $W_{\phi}(s)$ is the pull-back of the usual charged scalar field Wightman function

$$\begin{aligned} W_{\phi}(x'', x') &= \langle 0_{\phi}|\, \phi^{\dagger}(x'')\, \phi(x')\, |0_{\phi}\rangle = \langle 0_{\phi}|\, \phi(x'')\, \phi^{\dagger}(x')\, |0_{\phi}\rangle \\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{1}{2\omega_{\mathbf{k}}}\, e^{-i\omega_{\mathbf{k}}(t''-t')}\, e^{i\mathbf{k}\cdot(\mathbf{x}''-\mathbf{x}')}, \end{aligned} \tag{7}$$

on the detector's worldline and $W_A(s)$ is the pull-back of the electromagnetic dressing two-point function,

$$\begin{aligned} W_{A}(x'', x') = \frac{1}{|\alpha|^2 + |\beta|^2} \Big(& |\alpha|^2\, \langle 0_A|\, e^{-iC(x'')}\, e^{iC(x')}\, |0_A\rangle \\ & + |\beta|^2\, \langle 0_A|\, e^{iC(x'')}\, e^{-iC(x')}\, |0_A\rangle \Big). \end{aligned} \tag{8}$$

where α and β are fixed constants pertaining to the excited energy state of the detector more precisely defined in.[12] From the form of response function in Eq. (6), one can note that for the local quantum system considered here with the gauge invariant operator, there is an extra dependence through the $W_A(s)$ factor than in the corresponding expression of the standard detector[15] without the gauge invariance. It is this extra factor which brings the dependence of the soft photons in the de-excitation rate of the system.

3. The result

We choose the electric field dressing to have a $1/r^2$ radial dependence but allow the magnitude to have arbitrary angular dependence as well, which provides the soft photon dressing to the standard Coulombic field. The following form is chosen[14]

$$\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{x}') = \frac{q}{4\pi\epsilon_0}\, \frac{g(\theta^{A'})}{|\mathbf{x} - \mathbf{x}'|^3} (\mathbf{x} - \mathbf{x}'), \tag{9}$$

where the function $g(\theta^{A'})$ depends only on the angular coordinates and it satisfies the normalisation condition $\int_0^{\pi}\int_0^{2\pi} g(\theta^A)\sin\theta\, d\theta\, d\phi = 4\pi$. The soft charge for the dressed electric field in Eq. (9) can be found to be

$$\begin{aligned} Q_\epsilon &= \int d\Omega\ \epsilon(\theta^A) \lim_{r\to\infty}\left(r^2 E^r\right) \\ &= \int d\Omega\ \epsilon(\theta^A)\ g(\theta^A), \end{aligned} \tag{10}$$

where the integral over $d\Omega$ runs over the transverse angular coordinates, $\epsilon(\theta^A)$ is a weight function and E^r is the radial component of the electric field.

The response function for the dressed field in Eq. (9) is found through the expression in Eq. (6) to be[12]

$$F(E) = \frac{(-E)\,\Theta(-E)}{2\pi} \times \frac{e^{-\tilde{Q}\gamma}\,(-E/\Lambda)^{\tilde{Q}}}{\Gamma(2+\tilde{Q})}. \tag{11}$$

where γ is the Euler-Mascheroni constant, Λ is the ultraviolet cut-off and $\tilde{Q}$ depends on the averaged squares of soft charges through

$$\tilde{Q} := \frac{q^2}{64\pi^4\epsilon_0^2}\left\{\langle Q^2_{\epsilon_1}\rangle + \langle Q^2_{\epsilon_2}\rangle + \langle Q^2_{\epsilon_3}\rangle + \langle Q^2_{\epsilon_4}\rangle\right\}. \tag{12}$$

In the limit of no soft charges, $\tilde{Q}\to 0$, the response function $F(E)$ reduces to the standard expression of the response of a detector without a gauge field coupling.[15]

$$F_{\text{Mink}}(E) = \frac{(-E)\,\Theta(-E)}{2\pi}, \tag{13}$$

In the presence of the soft charges, $\tilde{Q} > 0$, the excitation rate in Eq. (11) vanishes as expected, but the dependence of the de-excitation rate on the energy gap E gets modified through a different power dependence. We conclude that the soft charges, and hence the soft photon dressing to the electric field, do not induce spontaneous excitations in our local quantum system, but they do affect the detector's de-excitation rate. The general features of these results, in particular the vanishing of excitation rate and the dependence of the de-excitation rate on soft charges, were shown to hold for our local quantum system for a wide class of soft photon dressed electric fields defined only by their asymptotic properties.[12]

References

1. H. Bondi, M. G. J. van der Burg and A. Metzner, Proc. Roy. Soc. Lond. **A269**, 21-52 (1962); R. Sachs, Proc. Roy. Soc. Lond. **A270**, 103-126 (1962); R. Sachs, Phys. Rev. **128**, 2851-2864 (1962).
2. B. Zeldovich and A. G. Polnarev, Sov. Astron. Lett. **18**, 17-23 (1974).
3. S. W. Hawking, M. J. Perry and A. Strominger, Phys. Rev. Lett. **116**, 231301 (2016) [arXiv:1601.00921 [hep-th]].
4. S. W. Hawking, M. J. Perry and A. Strominger, JHEP **1705**, 161 (2017) [arXiv:1611.09175 [hep-th]].

5. A. Strominger, JHEP **1407**, 152 (2014) [arXiv:1312.2229].
6. T. He, V. Lysov, P. Mitra and A. Strominger, [arXiv:1401.7026 hep-th].
7. J. Winicour, Class. Quant. Grav. **31**, 205003 (2014) [arXiv:1407.0259].
8. A. Ashtekar, Surveys in Differential Geometry **20**, 99 (2015) [arXiv:1409.1800].
9. A. Strominger and A. Zhiboedov, JHEP **1601**, 086 (2016) [arXiv:1411.5745].
10. S. Kolekar and J. Louko, Phys. Rev. D **96**, 024054 (2017) [arXiv:1703.10619 [hep-th]].
11. S. Kolekar and J. Louko, Phys. Rev. D **97**, 085012 (2018) [arXiv:1709.07355 [hep-th]].
12. S. Kolekar and J. Louko, "Can a local quantum system see soft photons," [arXiv: 2108.04278 [hep-th]].
13. P. A. M. Dirac, Can. J. Phys. **33**, 650 (1955).
14. S. Giddings, Phys. Rev. D **100**, 126001 (2019).
15. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press 1982).

Breaking of the adiabatic invariance in the production of particles by strong fields*

P. Beltrán-Palau†, A. Ferreiro, J. Navarro-Salas and S. Pla

Departamento de Fisica Teorica and IFIC, Centro Mixto Universidad de Valencia-CSIC. Facultad de Fisica, Universidad de Valencia, Burjassot-46100, Valencia, Spain

† *E-mail: pau.beltran@uv.es*

Particles are spontaneously created from the vacuum by time-varying gravitational or electromagnetic backgrounds. It has been proven that the particle number operator in an isotropically expanding universe is an adiabatic invariant. In this work we show that, in some special cases the expected adiabatic invariance of the particle number breaks down in presence of electromagnetic backgrounds. We also show a close relation between this breaking of the adiabatic invariance and the emergence of the axial anomaly.

Keywords: Particle creation, Schwinger effect, adiabatic invariance, axial anomaly.

1. Introduction

The proper understanding of the particle creation phenomena, via Bogoliubov transformations, was pioneered in the analysis of quantized fields in an isotropically expanding universe.[1–3] A fundamental issue in the study of particle creation in an expanding universe was the adiabatic invariance of the number of created particles. The particle number of a quantized field, in the limit of an infinitely slow and smooth expansion of the universe, does not change with time,[4] even if the quantized field is massless. In other words, the density of created particles by the expanding universe approach zero when the Hubble rate $\dot{a}/a$ is each time negligible even if the final amount of expansion $a(t_{final})/a(t_{initial})$ is large. Moreover, pair production can also take place in time-varying electric backgrounds,[5] and it can be regarded as a very important non-perturbative process in quantum field theory.[6]

The main purpose of this work is to generalize the analysis of the adiabatic invariance of the particle number observable in presence of an electromagnetic background. We find that, for massive fields, adiabatic invariance is, as expected, preserved (for more details see[7]). For slowly varying electromagnetic potentials no quanta is being produced, even if the change in A_μ over a long period is very large. However, in some cases and only for massless fields, the particle number is not an adiabatic invariant. We analyze the problem in detail in a two-dimensional scenario, for both scalar and Dirac fields. The main results can be easily translated to four dimensions. As a by-product of our analysis, we point out a connection between the

*Talk given by P. Beltran-Palau at the Sixteenth Marcel Grossmann Meeting (2021).

(anomalous) breaking of the adiabatic invariance of the particle number operator and the emergence of a quantum anomaly in the chiral symmetry. The requirements for the non-conservation of chirality seems to be directly related to those found for the breaking of the adiabaticity in the particle number observable. Concerning the renormalization, it is convenient in this context to use the adiabatic renormalization method[1–3] for a scalar field in a gravitational background, and recently extended for a Dirac field[8,9] and for an electric background.[10–13]

2. Brief introduction to particle creation in an expanding universe

First of all, we will introduce the phenomenon of the spontaneous particle creation in the gravitational context. It was firstly introduced by L. Parker[1,2] in the context of cosmology, and he developed the analysis for a real scalar field in an expanding universe, given by the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric $ds^2 = dt^2 - a^2(t)d\vec{x}^2$. Following the prescription of the Quantum Field Theory in curved spacetimes, the gravitational field can be considered as a classical background and scalar the quantum field is described by the Klein-Gordon equation in curved space-times

$$(g^{\mu\nu}\nabla_\mu\nabla_\nu + m^2 + \xi R)\phi = 0\,, \tag{1}$$

where ξ is the coupling constant of the field with the gravitational background. We expand the quantum field as

$$\phi(t,\vec{x}) = \int \frac{d^3k}{\sqrt{2(2\pi)^3 a(t)^3}} \left(A_{\vec{k}} h_k(t) e^{i\vec{k}\vec{x}} + A^\dagger_{\vec{k}} h^*_k(t) e^{-i\vec{k}\vec{x}} \right)\,, \tag{2}$$

where $\omega_k = \sqrt{m^2 + (k/a)^2}$ and $A^\dagger_{\vec{k}}$ and $A_{\vec{k}}$ are the usual creation and annihilation operators. Introducing it into the Klein-Gordon equation we obtain the following differential equation for the modes

$$\frac{d^2 h_k}{dt^2} + \Omega_k(t)^2 h_k = 0\,, \tag{3}$$

where

$$\Omega_k(t)^2 = \omega_k^2 + \left(6\xi - \frac{3}{4}\right)\left(\frac{\dot{a}}{a}\right)^2 + \left(6\xi - \frac{3}{2}\right)\frac{\ddot{a}}{a}\,. \tag{4}$$

This is the equation of a harmonic oscillator with a time-varying frequency, which depends on the derivatives of the metric. this idea will be usefull to understand the concept of adiabatic invariance.

It is well-known that the notion of particle is not well defined when the universe is in expansion, therefore a useful strategy to analyze the phenomenon is to choose a configuration where the expansion is bounded at early and late times, as shown in Fig. 1. In this case the number of particles is well defined at early and late times, and then we can compare the difference between them.

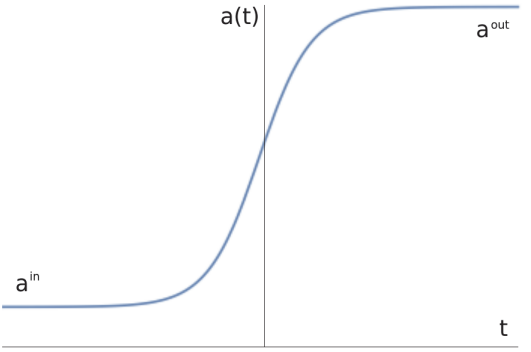


Fig. 1. Scale factor of a bounded expanding universe

We consider that at $t \to -\infty$ the vacuum expectation value of the particle number is zero, i.e., we choose the vacuum state at early times. This is analogous to choose the (normalized) positive frequency solution at early times

$$h_k^{in}(t) = \frac{1}{\sqrt{\omega_{in}}} e^{-i\omega_k^{in} t}\,, \tag{5}$$

where $\omega_k^{in} = \sqrt{m^2 + (k/a^{in})^2}$. However, the expansion of the universe induces a final state which is a combination of positive and negative frequency solutions

$$h_k^{out}(t) = \frac{1}{\sqrt{\omega_{out}}} \left(\alpha_k e^{-i\omega_k^{out} t} + \beta_k e^{i\omega_k^{out} t}\right)\,, \tag{6}$$

where $\omega_k^{in} = \sqrt{m^2 + (k/a^{out})^2}$ and α_k and β_k are the well-known Bogoliubov coefficients. It can be proven that the vacuum expectation value of the particle density at late times is not null, as a consequence of the presence of the negative frequency solution at late times. It is given by

$$\langle N^{out} \rangle = \frac{1}{(2\pi a^{out})^3} \int d^3k |\beta_k|^2\,. \tag{7}$$

Therefore, the beta coefficient gives the spectrum of momenta of the spontaneously created particles.

2.1. *Adiabatic invariance of the particle number*

Let us now introduce the concept of adiabatic invariance. As we have seen above the modes of the field follow an equation of a harmonic oscillator with time-varying frequency. In the model we are considering it would be an oscillator that starts oscillating with an specific frequency ω_k^{in} and decreases slowly its frequency tending asymptotically to ω_k^{out}. The energy of an oscillator of this kind changes in time, and is be given by the expression

$$E_k(t) = \frac{1}{2}(|\dot{h}_k(t)|^2 + \Omega_k(t)^2 |h_k(t)|^2)\,. \tag{8}$$

A main property of this kind of oscillators is that the quantity $E_k(t)/\Omega_k(t)$ is an adiabatic invariant, which means that it remains constant in the limit of an infinitely slow variation of the frequency. This limit is known as adiabatic limit. A straightforward calculation gives us the expressions of the initial and final energy in our particular case of the bounded expansion

$$E_k^{in} = \omega_k^{in} \tag{9}$$

$$E_k^{out} = \omega_k^{out}(1 + 2|\beta_k|^2)\,. \tag{10}$$

By the property mentioned above, the difference of quotients $(E^{in}/\omega_k^{in}) - (E^{out}/\omega_k^{out})$ should tend to 0 in the adiabatic limit. One can see easily that this implies necessarily that $|\beta_k|^2 \to 0$ in the adiabatic limit. This means that in the limit of an infinitely slow expansion of the universe no particles would be created by the effect of the gravity, and that's why we say that the particle number is an adiabatic invariant in this context (for more details on this concept see[4]). But there are other scenarios of particle creation in which this adiabatic invariance of the particle number can be broken, as we will see in the following sections.

3. Breaking of the adiabatic invariance by the electric field in 2D

We know that the phenomenon of particle creation is not a particular property of the gravitational field. For instance, a strong electric background can also produce spontaneously pairs particle-antiparticle from the vacuum. This phenomenon is known as the Schwinger effect.[6] A logical question we can formulate is, does the adiabatic invariance of the particle number remains when the particle production is induced by an electric field? To answer this question, we will firstly analyze the 2-dimensional case and then we will extend it to 4 dimensions. Let us consider a spatially homogeneous and time dependent classical electric field $E(t)$ acting in the unique spatial direction. For convencience we choose a particular form of this background, which is a Sauter-type[14] electric pulse given by

$$E(t) = -\frac{\rho A_0}{2}\cosh^{-2}(\rho t)\,, \tag{11}$$

shown in Fig. 2(a). It can also be described by the potential $A_\mu = (0, -A(t))$, with $E(t) = -\dot{A}(t)$, i.e.

$$A(t) = \frac{1}{2}A_0\left(\tanh(\rho t) + 1\right)\,. \tag{12}$$

This potential, analogously to the scale factor in the gravitational case, is bounded at early and late times, as shown explicitly in Fig. 2(b), starting in 0 and tending to a constant value that we denote as A_0. Note that ρ plays the role of a slowness parameter. The variation of the potential vector is slower when ρ decreases, then the adiabatic limit is equivalent to the limit $\rho \to 0$. As a first approach we will consider the 2-dimensional case for simplicity, and in the next section we will extend the results to the 4-dimensional case.

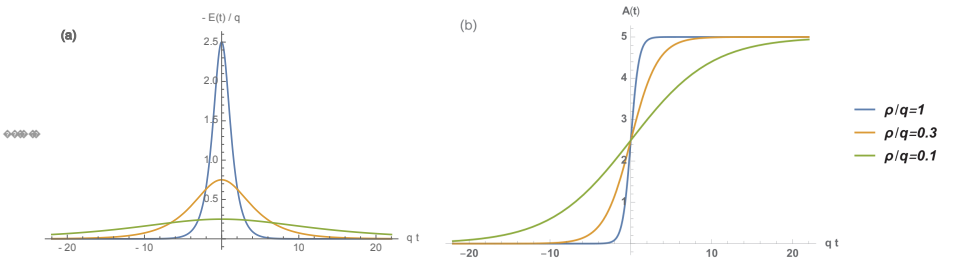


Fig. 2. (a) Sauter-type electric pulse for $A_0 = 5$ and for different values of the slowness parameter ρ. (b) Electric potential associated to the pulse for the same values of ρ. The adiabatic limit corresponds to $\rho \to 0$. The variables have been scaled with q to obtain dimensionless variables.

3.1. *Scalar field*

We will firstly analyze the case of the creation of scalar particles by an electric field. The Klein-Gordon equation for an electric background is given by

$$(D_\mu D^\mu + m^2)\phi = 0\,, \tag{13}$$

where $D_\mu \phi = (\partial_\mu + iqA_\mu)\phi$. We expand the field in terms of the usual annihilation and creation operators as

$$\phi(t,x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2(2\pi)}} [A_k e^{ikx} h_k(t) + B_k^\dagger e^{-ikx} h_{-k}^*(t)]\,. \tag{14}$$

Therefore the differential equation for the modes is given by

$$\ddot{h}_k(t) + \left(m^2 + (k - qA(t))^2\right) h_k(t) = 0\,. \tag{15}$$

Analogously to the gravitational case we locate the vacuum at early times, which is analogous to choose the positive frequency solution, and solving the equation we obtain at late times a combination of positive and negative frequency solutions, in terms of Bogoliubov coefficients. Finally we obtain the expression for the β_k coefficient which gives the spectrum of momenta of the created scalar particles

$$|\beta_k|^2 = \frac{\cosh\left(\pi \frac{\omega_{out} - \omega_{in}}{\rho}\right) + \cosh\left(\pi \sqrt{(\frac{qA_0}{\rho})^2 - 1}\right)}{2\sinh\left(\pi \frac{\omega_{in}}{\rho}\right)\sinh\left(\pi \frac{\omega_{out}}{\rho}\right)}\,, \tag{16}$$

where $\omega_{in} = \sqrt{k^2 + m^2}$ and $\omega_{out} = \sqrt{(k - qA_0)^2 + m^2}$. For conservation of the charge, the same amount of anti-particles will also be created with the same spectrum, but in the opposite direction, i.e., $|\beta_{-k}|^2$.

From this expression one can prove[7] that in the adiabatic limit, $\rho \to 0$, $|\beta_k|^2$ tends to 0 for massive particles, but in the massless case it tends to 1 inside the interval of momenta $k \in (0, qA_0)$, and it is 0 elsewhere. In other words, in the limit of an infinitely slow electric pulse only massless particles would be spontaneously created. Therefore, in the case of massless particles the adiabatic invariance of the particle number is broken.

The density of created particles can be obtained by integrating $|\beta_k|^2$ in all momenta. In Fig. 3 we represent the density of particles for different values of the mass in terms of the slowness parameter ρ. Note that in all the massive cases the production of particles is reduced when the process is more adiabatic, and in the limit when it is infinitely adiabatic ($\rho \to 0$) the number of particles is 0. However, in the massless case the particle density tends to a non-null value, which is given by $\langle N \rangle = \frac{|qA_0|}{\pi}$.

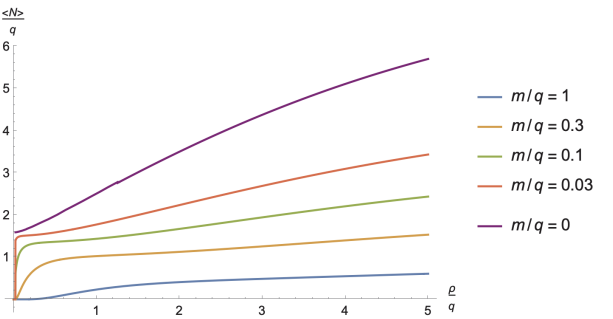


Fig. 3. Density of late-time created scalar particles as a function of the adiabaticity parameter ρ, for $A_0 = 5$ and for different values of the mass.

3.2. *Dirac field*

A Dirac field coupled to an electric background in 2 dimensions is described by the following Dirac equation

$$(i\gamma^\mu D_\mu - m)\psi = 0\,, \tag{17}$$

where γ^μ are the Dirac matrices and $D_\mu \equiv \partial_\mu - iqA_\mu$. We expand the field in terms of the creation and annihilation operators and normalized spinors as follows

$$\psi(t,x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}} \left[B_k e^{ikx} \begin{pmatrix} h_k^I(t) \\ -h_k^{II}(t) \end{pmatrix} + D_k^\dagger e^{-ikx} \begin{pmatrix} h_{-k}^{I*}(t) \\ h_{-k}^{I*}(t) \end{pmatrix} \right] . \tag{18}$$

The field equation is converted into the system of differential equations

$$\dot{h}_k^I - i\,(k + qA)\,h_k^I - imh_k^{II} = 0 \tag{19}$$

$$\dot{h}_k^{II} + i\,(k + qA)\,h_k^{II} - imh_k^I = 0\,. \tag{20}$$

Following a similar procedure as in the scalar case, we impose a positive frequency solution at early times and solve the equation, obtaining at late times also a negative frequency term.[7] From that expression we can obtain the Bogoloiubov coefficient that describes the spectrum of momenta of the created particles

$$|\beta_k|^2 = \frac{\cosh\left(\pi\frac{\omega_{out}-\omega_{in}}{\rho}\right) - \cosh\left(\pi\frac{qA_0}{\rho}\right)}{2\sinh\left(\pi\frac{\omega_{in}}{\rho}\right)\sinh\left(\pi\frac{\omega_{out}}{\rho}\right)}\,. \tag{21}$$

Note that $|\beta_k|^2 \leq 1$ in accordance with the the Pauli's exclusion principle. As for the scalar field, in the adiabatic limit $\rho \to 0$ this expression tends to 0 in the massive case, but in the massless case instead $|\beta_k|^2 = 1$ inside the interval $k \in (0, qA_0)$. Therefore, the adiabatic invariance is also broken for massless Dirac fermions. To clarify this idea, in Fig. 4 we show the density of particles, after integrating in all momenta, in terms of ρ for different masses. Note that, unlike the scalar case, the density of particles in the massless case is the same ($\langle N \rangle = |qA_0|/\pi$) for any value of ρ, i.e., it only depends on the final value of the electric potential, not on its form.

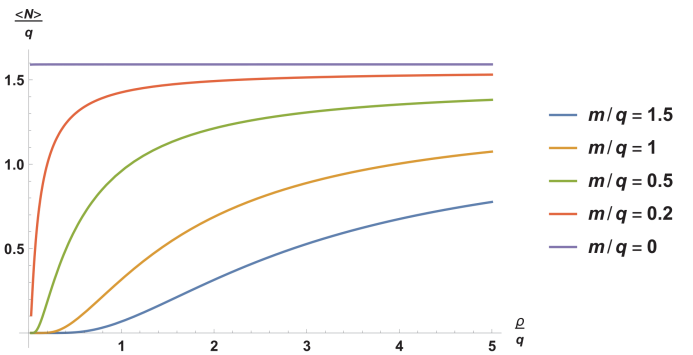


Fig. 4. Density of late-time created Dirac fermions as a function of the adiabaticity parameter ρ, for $A_0 = 5$ and for different values of the mass.

3.3. *Relation with the axial anomaly*

This breaking of the adiabatic invariance is closely related with the axial anomaly, which is also manifested in the creation of massless particles. Let us analyze this relation. It is well-known that the classical axial symmetry of the massless Dirac field is broken when considering it as a quantum field, which is known as axial anomaly. In 2 dimensions this anomaly can be expressed by[15]

$$\partial_\mu \langle j_A^\mu \rangle_{ren} = -\frac{q}{2\pi} \epsilon^{\mu\nu} F_{\mu\nu} = \frac{q\dot{A}}{\pi}\,, \tag{22}$$

where $\langle j_A^\mu \rangle$ is the 4-vector that contains the axial charge and current. Note that its divergence gives a non-vanishing value, unlike in the classical case.

In 2 dimensions it is easy to see the relation of this phenomenon with the creation of particles, since the chirality of a particle is linked to its electric charge and the direction of its movement. For instance, a positive charged particle moving in the positive direction of movement would have right-handed chirality. The created particles by the electric field are produced in pairs with opposite charge and opposite direction of movement, therefore both particles of every pair will have the same

chirality, creating a net chiral charge. In fact, the created pairs generate a electric current, which in 2 dimensions is directly proportional to the chiral charge. This observable can be calculated by using adiabatic renormalization, and the result of the total chiral charge of the created massless Dirac fermions is

$$\langle j^0_A \rangle_{ren} = \frac{qA(t)}{\pi}\,. \tag{23}$$

(for more details on the adiabatic renormalization method for Dirac fields see[10, 12, 13]) Note that this expression coincides with the one obtained by integrating (22), i.e., the chiral charge created by the massless fermions is the one predicted by the axial anomaly. This anomaly remains even if the background changes infinitely slow, just like the creation of massless particles as we have seen in the previous section. Then we can see the close relation between the axial anomaly and the breaking of the adibatic invariance. It is also interesting to remark that for massless scalar fields in 2 dimensions one can define an analogue axial anomaly, and it can be proven to be also compatible with the particle production effect.

4. Generalization to 4 dimensions

In this section we will generalize the previous results to the 4-dimensional case, and will analyze the influence of the introduction of a magnetic background apart from the electric one. We will choose the electric field acting in the direction z. We will not give the explicit details of the calculation but a summary of it, since it is analogue to the one in 2D. In fact one can see that the differential equations for the modes (and therefore the expression of β_k) can be related with the 2D ones by changing k by k_3 and the mass by an effective mass, which depends on the case. This simplifies a lot the analysis, since we know for the previous analysis that the breaking of the adiabatic invariance (i.e. $|\beta_{\vec{k}}| \neq 0$ when $\rho \to 0$) is given when this effective mass (m_{eff}) is 0.

If we consider only an electric background, the effective mass is $m^2_{eff} = m^2 + k_1^2 + k_2^2$, and then for $m = k_1 = k_2 = 0$ the Bogoliubov coefficient would tend to a non-null value when $\rho \to 0$. But since the momentum is a continuous quantum number and the density of particles is obtained by the integration in all momenta, the amount of created particles with this specific momentum is diluted in the infinite space, and the total particle density vanishes in the adiabatic limit. And then the adiabatic invariance is preserved in this case.

But the situation changes when adding a classical magnetic background. We choose for simplicity a constant magnetic field in the same direction as the electric one. In this case the perpendicular momentum k_1 is quantized in the well known Landau levels. For the scalar field the effective mass is $m^2_{eff} = m^2 + (2n+1)|qB|$, where n denote the Landau levels. This quantity never vanishes, and then the adiabatic invariance is preserved. On the other hand, for the Dirac field the effective mass is $m^2_{eff} = m^2 + 2n|qB|$, which vanishes for $m = 0$ and in the ground level

$n = 0$. In this case the particle number is obtained by the summation of all levels, and then the contribution of $n = 0$ is not diluted. Therefore in this case the particle density does not vanish in the adiabatic limit, which means that there is a breaking of the adiabatic invariance.

We can also see in this cases the close relation between this phenomenon and the axial anomaly. On one hand, in 4 dimensions this anomaly appears in the same conditions as the breaking of the adiabatic invariance, which are massless fermions in presence of not perpendicular electric and magnetic background fields. On the other hand, the chiral charge of the created massles fermions is given by

$$\langle j^0_A \rangle_{ren} = -\frac{q^2}{2\pi^2} \int_{-\infty}^{t} dt' \vec{E}(t')\vec{B}\,, \tag{24}$$

which has been obtained by using adiabatic renormalization. This expression coincides with the axial anomaly in 4 dimensions. And finally, the only mode that contributes to the chiral charge created by the axial anomaly is the lower one, $n = 0$, which is also the one that produces the breaking of the adiabatic invariance, as seen before.

5. Summary and conclusions

We have analyzed the adiabatic invariance of the particle density spontaneously created by an electromagnetic background. This property, which is verified in the case of a gravitational background, is broken in some cases when the background is an electromagnetic field. Specifically for massless scalar and Dirac fields in 2 dimensions in presence of an electric background, and for massless Dirac fields in 4 dimensions in presence of non perpendicular electric and magnetic background fields. We have also seen that there is a close relation between the breaking of the adiabatic invariance and the axial anomaly since both phenomena emerge in the same cases. Morover the amount of chiral charge produced by the creation of massless particles that remains in the adiabatic limit coincides with the expected one by the axial anomaly.

Acknowledgements

We thank I.Agullo and A. del Rio for useful comments and discussions. This work was supported by Spanish Ministerio de Economia, Industria y Competitividad Grants No. FIS2017-84440-C2-1-P (MINECO/FEDER, EU), No. FIS2017-91161-EXP and by the project PROMETEO/2020/079 (Generalitat Valenciana). P. B. is supported by the Ministerio de Ciencia, Innovación y Universidades, Ph.D. fellowship, Grant No. FPU17/03712. S. P. is supported by the Ministerio de Ciencia, Innovación y Universidades, Ph.D. fellowship, Grant No. FPU16/05287. A. F. is supported by the Severo Ochoa Ph.D. fellowship, Grant No. SEV-2014-0398-16-1, and the European Social Fund.

References

1. L. Parker, *The creation of particles in an expanding universe*, Ph.D. thesis, Harvard University (1966). Dissexpress.umi.com, Publication Number 7331244; *Phys. Rev. Lett.* **21**, 562 (1968); *Phys. Rev. D* **183**, 1057 (1969); *Phys. Rev. D* **3**, 346 (1971).
2. L. Parker and D. J. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Fields and Gravity*, Cambridge University Press, Cambridge, England (2009).
3. N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge University Press, Cambridge, England (1982).
4. L. Parker, *J. Phys. A* **45**, 374023 (2012).
5. W. Pittrich and H. Gies, *Probing the Quantum Vacuum*, Springer, Heidelberg (2000).
6. J. Schwinger, *Phys. Rev.* **82**, 664 (1951).
7. P. Beltrán, A. Ferreiro, J. Navarro-Salas and S. Pla *Phys. Rev. D* **100**, 085014 (2019).
8. A. Landete, J. Navarro-Salas and F. Torrenti, *Phys. Rev. D* **88**, 061501 (2013).
9. A. del Rio, J. Navarro-Salas and F. Torrenti, *Phys. Rev. D* **90**, 084017 (2014).
10. A. Ferreiro and J. Navarro-Salas, *Phys. Rev. D* **97**, 125012 (2018).
11. A. Ferreiro, J. Navarro-Salas and S. Pla, *Phys. Rev. D* **98**, 045015 (2018).
12. J. F. Barbero G., A. Ferreiro, J. Navarro-Salas and E. J. S. Villaseñor, *Phys. Rev. D* **98**, 025016 (2018).
13. P. Beltrán-Palau, J. Navarro-Salas and S. Pla, *Phys. Rev. D* **99** 105008 (2019).
14. F. Sauter, *Z. Phys.* **69**, 742 (1931).
15. R. A. Bertlmann, *Anomalies in Quantum Field Theory*, Oxford University Press, Oxford (2000).

Dynamics of relativistic electrons in non-uniform magnetic fields and its applications in quantum computing and astrophysics

Srishty Aggarwal* and Banibrata Mukhopadhyay

Department of Physics, Indian Institute of Science, Bengalore, Karnataka 560012, India
**E-mail: srishtya@iisc.ac.in*

We explore the two-dimensional motion of relativistic electrons when they are trapped in magnetic fields having spatial power-law variation. Its impacts include lifting of degeneracy that emerged in the case of the constant magnetic field, special alignment of Landau levels of spin-up and spin-down electrons depending on whether the magnetic field is increasing or decreasing from the centre, splitting of Landau levels of electrons with zero angular momentum from that of positive one and the change in the equation of state of matter. Landau quantization (LQ) in variable magnetic fields has interdisciplinary applications in a variety of disciplines ranging from condensed matter to quantum information. As examples, we discuss the increase in quantum speed of the electron in presence of spatially increasing magnetic field; and the attainment of super Chandrasekhar mass of white dwarfs by taking into account LQ and Lorentz force simultaneously.

Keywords: Landau quantization; Non-uniform magnetic field; Dirac equation; Chandrasekhar-limit; Quantum speed limit

1. Introduction

Landau Quantization (LQ) is the phenomenon of the quantization of the cyclotron orbits of charged particles in presence of magnetic field. LQ has been widely discussed for uniform magnetic fields both in non-relativistic[1] as well as relativistic[2] cases. It leads to many interesting effects like quantum Hall effect, de Has Van Alphen effect, Shubnikov oscillations, modification to the equation of state (EOS) and change in the neutron drip line. The modified EOS was later proved to be useful in explaining the mass of super Chandrasekhar white dwarfs.[3]

However, a complete uniform magnetic field is an idealistic realization. What if it is non-uniform as in astrophysical systems and plasmas? In white dwarfs, neutron stars and even main sequence stars, the magnetic field varies drastically from centre to surface. Plasma can act as a source of both increasing and decreasing magnetic fields depending on its properties. Even in laboratories, fluctuations in the magnetic field are a common sight. This motivates us to initiate the exploration of the dynamics of a relativistic electron in presence of strictly spatially variable magnetic fields.[4]

We show different arrangements of the energy levels of an electron when it is trapped in variable magnetic fields. Further, we explore its applications in the fields of quantum information by showing an increase in quantum speed of an electron in presence of a spatially increasing magnetic field, and astrophysics, in the magnetized white dwarfs.

2. Solution of Dirac equation in presence of magnetic fields

The Dirac equation in the presence of magnetic field for an electron of mass m_e and charge q $(-e)$ is given by

$$i\hbar\frac{\partial\Psi}{\partial t} = \left[c\boldsymbol{\alpha}\cdot\left(-i\hbar\nabla - \frac{q\mathbf{A}}{c}\right) + \beta m_e c^2\right]\Psi, \tag{1}$$

where $\boldsymbol{\alpha}$ and β are Dirac matrices, $\mathbf{A}$ is the vector potential, $\hbar$ is the reduced Planck constant and c is the speed of light. For stationary states, we can write

$$\Psi = e^{-i\frac{Et}{\hbar}}\begin{bmatrix}\chi\\ \phi\end{bmatrix}, \tag{2}$$

where ϕ and χ are 2-component spinors.

Therefore, the decoupled equation for χ obtained from Eq. (1) is

$$(E^2 - m_e^2c^4)\chi = \left[c^2\left(\pi^2 - \frac{q\hbar}{c}\boldsymbol{\sigma}\cdot\mathbf{B}\right)\right]\chi, \tag{3}$$

where $\boldsymbol{\pi} = -i\hbar\nabla - q\mathbf{A}/c$.

We choose a simple power law variation of the magnetic field in compliance with Maxwell's equations as

$$\mathbf{B} = B_0\rho^n\hat{z}, \tag{4}$$

in cylindrical coordinates (ρ, ϕ, z). Such a field profile can take into account the uniformity in magnetic field when $n = 0$ as well as an increase and a decrease in fields with respect to radial coordinate with $n > 0$ and $n < 0$ respectively, satisfying other physics intact. Using a gauge freedom for the vector potential $\mathbf{A}$, we choose

$$\mathbf{A} = B_0\frac{\rho^{n+1}}{n+2}\hat{\phi} = A\hat{\phi}. \tag{5}$$

Let

$$\chi = e^{i\left(m\phi + \frac{p_z}{\hbar}z\right)}R_{\pm}(\rho). \tag{6}$$

Since the electron is confined to a plane perpendicular to the direction of magnetic field (z-direction), $p_z = 0$. By substituting χ from Eq. (6) into Eq. (3), and dividing it by $m_e^2c^2$, Eq. (3) becomes

$$\begin{aligned}\alpha R_{\pm}(\rho) = &-\lambda_e^2\left[\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{m^2}{\rho^2}\right]\tilde{R}_{\pm}\\ &+\left[\left(\frac{kB_0\rho^{n+1}}{n+2}\right)^2 + k\lambda_e\left(-\frac{2m}{n+2}\pm 1\right)B_0\rho^n\right]R_{\pm},\end{aligned} \tag{7}$$

where α is the eigenvalue of the problem and is equal to $\epsilon^2 - 1$, $\epsilon = E/m_ec^2$, $\lambda_e = \hbar/m_ec$ and $k = e/m_ec^2$. Further, choosing $R_{\pm}(\rho) = u_{\pm}(\rho)/\sqrt{\rho}$, the above equation is reduced to

$$\alpha u_{\pm} = \left(-\lambda_e^2\frac{\partial^2}{\partial\rho^2} + V_{eff}\right)u_{\pm}, \tag{8}$$

such that

$$V_{eff} = -\lambda_e^2 \left[\frac{1}{4\rho^2} - \frac{m^2}{\rho^2}\right] + \left(\frac{kB_0\rho^{n+1}}{n+2}\right)^2 + k\lambda_e \left(-\frac{2m}{n+2} \pm 1\right) B_0\rho^n. \qquad (9)$$

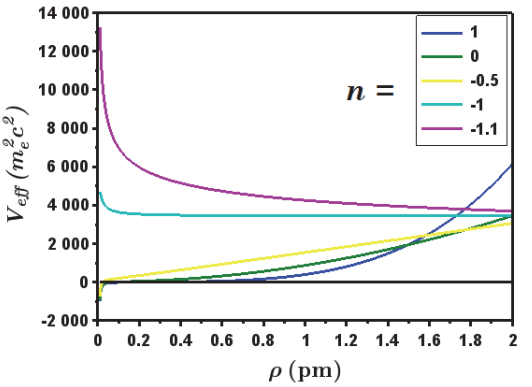


Fig. 1. The variation of effective potential for different n for $B_0 = 10^{15}$ G pm^{-n}. The black horizontal line represents $V_{eff} = 0$. Various potentials at $\rho = 1$ pm from bottom to top represent $n = 1, 0, -0.5, -1$ and -1.1 respectively.

The variation of V_{eff} for different n is shown in Fig. 1. It can be seen that V_{eff} is completely repulsive for n less than or equal to -1, whose solution will depend on the distance from the source (origin of the system) upto which a particle can move. Therefore, the energy eigenvalues for such cases depend upon where we put a hard wall making the system equivalent to confining the electron in a box. However, we do not want to apply any such restrictions on the electron. We, therefore, limit our analysis to the cases with n greater than -1.

We use "Shooting and Matching" method to solve Eq. (8). For computational analysis, we choose ρ and B_0 in the units of picometer (pm) and G pm$^{-\mathrm{n}}$ respectively.

For uniform magnetic field, i.e. $n = 0$, the eigenvalue of level ν is given by

$$\alpha_\nu = 2k\lambda_e B_0 \left(\nu + \frac{|m|}{2} - \frac{m}{2} + \frac{1}{2} \pm \frac{1}{2}\right), \qquad (10)$$

where 'm' is the azimuthal quantum number.[1] One can easily see from Eq. (10) that ground state eigenvalue (α_0) is 0 and all the other energy levels are doubly degenerate. Also, energies are same for $m \geq 0$, which turn out to be

$$E^2 = p_z^2c^2 + m_e^2c^4 \left(1 + 2\nu\frac{B_0}{B_c}\right), \qquad (11)$$

where $B_c = m_e^2c^3/e\hbar$, the Schwinger limit of pair production.

3. Eigenspectra and dispersion relations

3.1. $m = 0$

Figures 2 and 3 show the eigenspectra for different n, obtained from Eq. (7) by taking $m = 0$. On comparison of the eigenlevels without Zeeman effect (yellow-solid lines in the left) in Fig. 2, it can be observed that the spacing between the eigenlevels increases for $n > 0$, remains same if $n = 0$ and decreases for $n < 0$, from ground to higher levels. It owes to the chosen magnetic field variation in the region. If the magnetic field is spatially increasing, it has low strength near the centre and becomes strong near the outer boundary. Hence, the lower levels have lesser spacing, and the gap between consecutive levels increases for higher levels. The opposite is true for a decaying field profile.

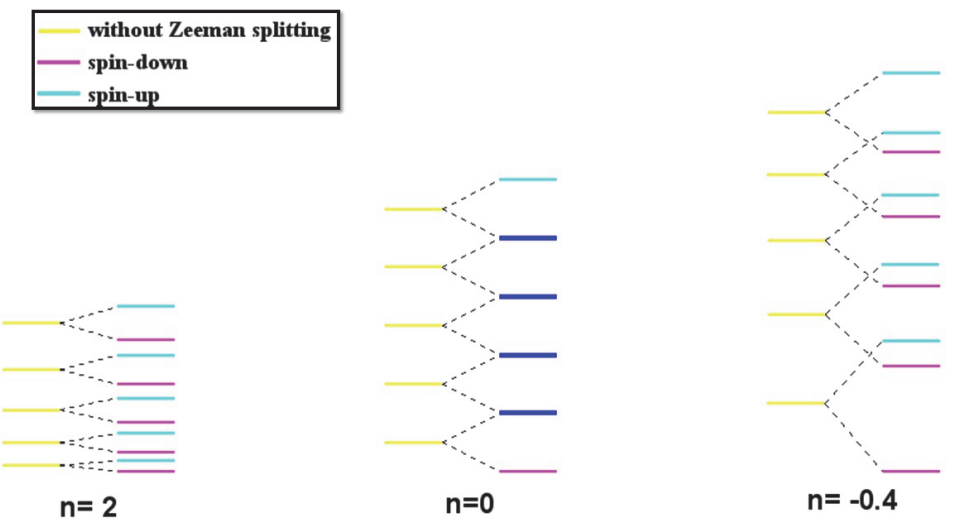


Fig. 2. Comparison of eigenlevels and the splitted states of $+\sigma.B$ and $-\sigma.B$ for $n = 2, 0$ and -0.4.

Breaking of degeneracy that arose in uniform magnetic field takes place for both increasing and decreasing magnetic fields if Zeeman effect is included. However, the alignment of levels for spin-up and spin-down electrons is quite different for these two cases as shown in Fig. 2. While the spin pattern for $n > 0$ is *dududu...*, it becomes *ddudud...* for $n < 0$ until $n = -0.6$, as can be inferred from Fig. 3. Below -0.6, the spin pattern is highly ambiguous, say for $n = -0.8$, it is *dddudud...*, whereas for $n = -0.9$, it changes to *ddddudud....* Here, d and u denote the levels for spin-down and spin-up electrons respectively. It is remarkable that the ground level for the spin-down electron always lies at 0 for all n.

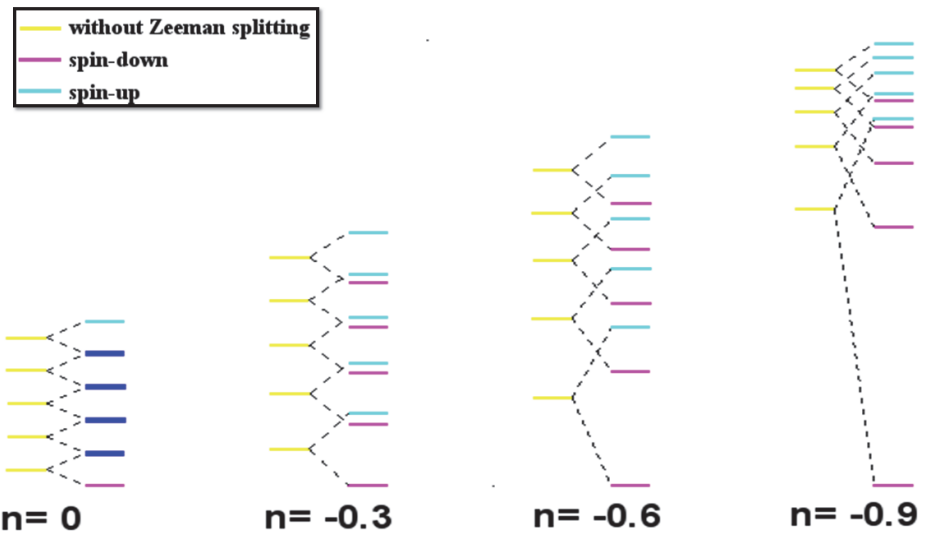


Fig. 3. Comparison of eigenlevels and the splitted states of $+\sigma.B$ and $-\sigma.B$ for $n = 0$ and when n is negative.

One can, thus, obtain the desired spacing between the eigenlevels using non-uniform magnetic fields. For example, we can prepare a system with electrons confined to lower levels in a decaying magnetic field, say with $n = -0.8$ or -0.9. Such a system will then have only spin-down electrons due to the availability of lower levels for the same. Thus, we can achieve a single spin system using such variable magnetic fields. Also, a spin-based transition can be observed in presence of increasing magnetic fields, if we restrict the energy of electron between spin-up ground level and spin-down first excited state energies.

Table 1. The values of the constants of Eq. (12) for various n. Here B_0 in Eq. (12) is chosen in the units of 10^{15} G pm^{-n}

n	C_3	C_5
−0.5	195.66	0.484
−0.4	134.63	0.486
−0.3	97	0.488
−0.2	72.5	0.4934
−0.1	56	0.50
0	44.42	0.50
1	10.95	0.5156
2	5.72	0.50
3	3.965	0.51
4	3.15	0.51

We obtain a general expression for the eigenvalue of a level ν for arbitrary n using trial and error method, given by[4]

$$\alpha_{\nu\pm} = C_3\, B_0^{\frac{2}{n+2}}\, (\nu + C_5)^{\frac{2+2n}{n+2}} \left[1 \pm \frac{C_5}{(\nu + C_5)}\right]. \tag{12}$$

where C_3 and C_5 are constants whose value depends on n. The values of these constants for some n are given in Table 1.

3.2. $m \neq 0$

We compare the eigenvalues between positive m ($m = 1$) and negative m ($m = -1$) with $m = 0$ for different variations in the magnetic fields. When $n = 0$, positive m does not have any impact on the eigenvalues in comparison with $m = 0$, while negative m leads to the increment in energies, as depicted in Fig. 4 as well as can be inferred from Eq. (10). The Fig. 4 shows that for variable magnetic fields, the trend with respect to negative m remains same as that for the uniform one, however, the impact of positive m changes for the varying field. While it leads to an increase in energies for $n > 0$, the energies decrease, if n is negative.

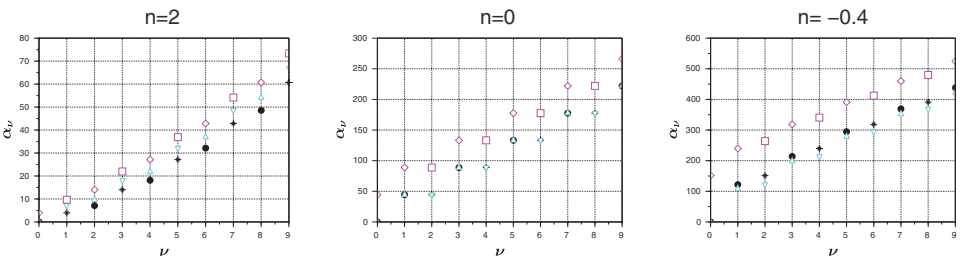


Fig. 4. The variation of eigenvalue with the eigen-index for $m = 0$ with $-\sigma.B$ (black solid circles) and $+\sigma.B$ (black asterisks); $m = +1$ with $-\sigma.B$ (cyan upward triangles) and $+\sigma.B$ (cyan downward triangles); $m = -1$ with $-\sigma.B$ (magenta diamonds) and $+\sigma.B$ (magenta squares), for $n = 2, 0$ and -0.4.

4. Finding ground state energy using variational method

We can find the upper bound of the ground state eigenvalue α_0 using the variational principle,[5] which states that

$$E_{gs} \leq \langle H \rangle, \tag{13}$$

where E_{gs} represents the ground state energy and $\langle H \rangle$ is the expectation value of Hamiltonian.

We choose the ground state wavefunction for both the spins of electron as

$$R_0 = e^{-\frac{KB_0\rho^{n+1}}{\lambda_e(n+2)^2}}. \tag{14}$$

The normalisation constant (N_0) is given by

$$N_0 = \sqrt{\frac{(n+2)}{\Gamma\left(\frac{2}{n+2}\right)}\left(\frac{2KB_0}{\lambda_e(n+2)^2}\right)^{\frac{2}{n+2}}}, \tag{15}$$

where

$$\Gamma\left(\frac{a}{b}\right) = b\int_0^\infty x^{a-1}e^{-x^b}dx.$$

Hence, the normalised wavefunction $\tilde{R}_0 = N_0 R_0$. Therefore,

$$\tilde{R}_0' = -\rho^{n+1}\frac{KB_0}{\lambda_e(n+2)}\tilde{R}_0, \tag{16}$$

and

$$\tilde{R}_0'' = -(n+1)\rho^n\frac{KB_0}{\lambda_e(n+2)}\tilde{R}_0 + \left(\frac{KB_0\rho^{n+1}}{\lambda_e(n+2)}\right)^2\tilde{R}_0. \tag{17}$$

Hence,

$$\alpha_{0_{Th}\pm} = \langle\tilde{R}_0|\,\alpha_\pm\,|\tilde{R}_0\rangle \tag{18}$$

$$= \frac{KB_0\lambda_e(1\pm1)}{\Gamma\left(\frac{2}{n+2}\right)}\left(\frac{2KB_0}{\lambda_e(n+2)^2}\right)^{-\frac{n}{n+2}}. \tag{19}$$

Thus, Eq. (19) provides an upper bound on the ground state energies for spin-down and spin-up electrons. (Eq. 19 can be obtained from Eq. 7).

Table 2. Comparison of analytical and computational α_0 values at different n for $B_0 = 10^{15}$ G pm$^{-\mathrm{n}}$.

n	$\alpha_{0_{Comp}+}$	$\alpha_{0_{Th}+}$
−0.5	225.83	256.70084
−0.4	151.124	162.464
−0.3	105.612	109.563
−0.2	76.716	77.840
−0.1	57.55	57.746
0	44.4	44.418
1	8.702	10.090
2	3.969	5.700
3	2.534	4.427
4	1.907	3.953

Note that $\alpha_{0_{Th}-}$ is zero for all n, which is the exact eigenvalue for the ground state of spin-down electron. For spin-up electron, the above expression provides the exact ground state eigenvalue only for the uniform magnetic field, but deviates for the non-uniform one. The comparison of actual α_0 obtained computationally ($\alpha_{0_{Comp}+}$) and the approximate analytical one obtained in this section ($\alpha_{0_{Th}+}$) for

spin-up electron at $B_0 = 10^{15}$ G pm^{-n} for different n is shown in Table 2. The deviation between the two values increases as one goes away from $n = 0$ in either direction. It is noteworthy that the power-law dependence of $\alpha_{0_{Th}}$ on B_0 in Eq. (19) is $2/(n+2)$ which is same as in Eq. (12), where α for a general ν is obtained using an appropriate ansatz.

5. Applications

In general, LQ in non-uniform magnetic fields can be useful in multiple domains of physics ranging from quantum information to condensed matter systems. Here, we will discuss its applications in quantum information via the increase in quantum speed of an electron and in astrophysics for the super-Chandrasekhar white dwarfs.

5.1. *Increase of quantum speed*

Quantum speed of a particle is defined as the speed of its transition from one energy level to the other. It has a direct influence on the processing speed of quantum information. The electron levels are equivalent to qubits in quantum information. Thus, if the transition between the levels takes place at higher pace, switching between the qubits will be more rapid, thereby, leading to faster processing of quantum information. Therefore, attaining higher quantum speed is one of the major requirements of the researchers in the field of quantum computation.

Assuming that the initial state of the electron is the superposition of the two consecutive states: the ground and the first excited one, given by

$$\Psi(\rho, 0) = \frac{1}{\sqrt{2}} \left[\psi_0(\rho) + \psi_1(\rho)\right], \tag{20}$$

with $m = 0$, $p_z = 0$ and the respective energies E_0 and E_1. For the transition of an electron from ground state to first excited state, the minimum time of evolution of wavefunction (T_{min}) is evaluated using Mandelstam-Tamm bound[6] as

$$T_{min} = \frac{\pi\hbar}{2\Delta H}, \tag{21}$$

where

$$\Delta H = \frac{E_1 - E_0}{2}. \tag{22}$$

The radial displacement of electron in time T_{min} is given by

$$\rho_{disp} = 2 \left| \int_0^\infty \rho D_S(\rho) d\rho \right|, \tag{23}$$

where

$$D_s(\rho) = \psi_0^\dagger \, \rho \, \psi_1. \tag{24}$$

Hence, the quantum speed of electron is

$$\tilde{\mathrm{v}} = \frac{\rho_{disp}}{T_{min}}, \tag{25}$$

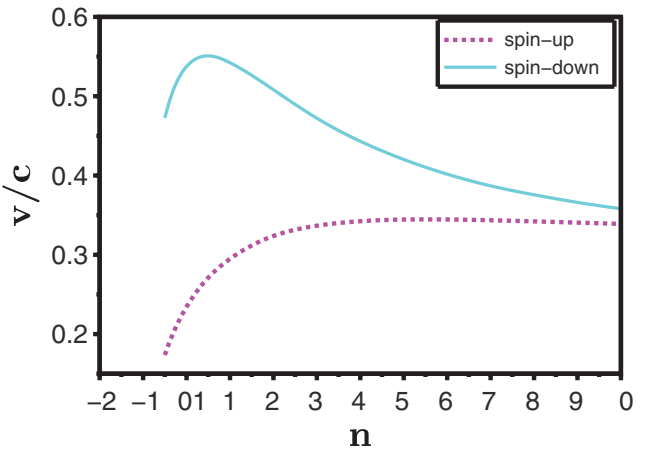


Fig. 5. Variation of quantum speed of spin-up and spin-down electrons for transition from ground state to first excited state with different n at $B_0 = 10^{16}$ G pm^{-n}.

Figure 5 illustrates the variation of the quantum speed of an electron with the change in non-uniformity index of the magnetic field in the relativistic regime. As it can be seen from the figure, the quantum speed of the electron increases with increasing n; reaches the maximum and then begins to decrease. Whereas quantum speed is maximum for $n = 1$ for spin-down electron, n has to be 6 to reach the highest speed for spin-up electron for $B_0 = 10^{16}$ G pm$^{-\mathrm{n}}$. This is associated with the varied alignment of energy levels in non-uniform magnetic fields lifting the degeneracy between $n < 0$ and $n > 0$ as shown in Fig. 2. Therefore, if a spin-up electron is trapped in a magnetic field that is spatially increasing in magnitude even linearly ($n = 1$), then we can achieve a higher speed of transition of the electron as shown in Fig. 5.

5.2. *Super-Chandrasekhar white dwarfs*

Astrophysical bodies like white dwarfs, neutron stars and magnetars are the natural sites to observe non-uniform magnetic fields, wherein, the difference between the central and the surface magnetic fields lies upto $3 - 4$ orders of magnitude or even higher.

The main impact of LQ in stellar physics is to modify the underlying EOS due to the discretization of energy levels in presence of strong magnetic fields. For a constant magnetic field, the energy levels are degenerate and the spacing between the eigenlevels is constant, but for a variable magnetic field case, the degeneracy breaks down and the spacing between levels is non-uniform, as shown in Figs. 2 and 3, which brings a considerable change in EOS in presence of non-uniform magnetic fields.

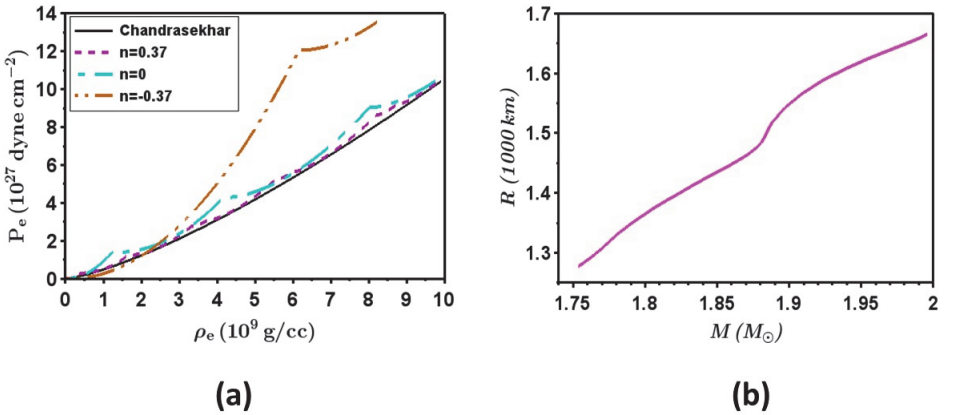


Fig. 6. (a) EOS for different n at $\epsilon_{Fmax} = 17$ for $B_0 = 2\times10^{15}\,\mathrm{G\,pm^{-n}}$. (b) Mass-radius relation of magnetized white dwarf for the magnetic field profile chosen in Eq. (33).

In a non-uniform magnetic field, the electron number density at zero temperature is given by

$$n_e = \frac{1}{(2\pi)^2\lambda_e^3}\left(\sum_{\nu=0}^{\nu=\nu_{m-}}\beta_-(\nu)x_{F-}(\nu) + \sum_{\nu=0}^{\nu=\nu_{m+}}\beta_+(\nu)x_{F+}(\nu)\right), \tag{26}$$

where '+' sign and '−' sign indicate spin-up and spin-down respectively,

$$x_{F\pm}(\nu) = \left[\epsilon_F^2 - (1+\alpha_\pm(\nu)\right]^{\frac{1}{2}} \tag{27}$$

and

$$\beta_\pm = \left(\alpha_\pm(\nu+1) - \alpha_\pm(\nu-1)\right)/2. \tag{28}$$

Note that $\nu_{m_\pm}$ is the largest integer value of ν such that

$$\alpha_{\nu_\pm} \le \epsilon_F^2 - 1. \tag{29}$$

The pressure of an electron gas at zero temperature is given by

$$P_e = \frac{m_e c^2}{(2\pi)^2\lambda_e^3}\left(\sum_{\nu=0}^{\nu=\nu_{m-}}\beta_-(\nu)(1+\alpha_-(\nu))f_2\left[\frac{x_{F-}(\nu)}{(1+\alpha_-(\nu))^{1/2}}\right] + \sum_{\nu=0}^{\nu=\nu_{m+}}\beta_+(\nu)(1+\alpha_+(\nu))f_2\left[\frac{x_{F+}(\nu)}{(1+\alpha_+(\nu))^{1/2}}\right]\right), \tag{30}$$

where

$$f_2(z) = \frac{1}{2}\left(z\sqrt{1+z^2} - ln(z+\sqrt{1+z^2})\right). \tag{31}$$

Figure 6(a) shows how the EOS changes with n at $B_0 = 2\times10^{15}\,\mathrm{G\,pm^{-n}}$. For a given ϵ_{Fmax}, the allowed number of levels decreases with decreasing n. With the

decrease in number of levels, EOS becomes stiffer and softer in high and low density regimes respectively. Although, the spatially increasing magnetic field is not common in stellar objects, we also show the change in EOS of electron in presence of an increasing magnetic field that may come handy for future explorations.

To probe the effect of LQ and Lorentz force simultaneously in white dwarfs, we choose the following sample magnetic field profile.

$$B = \begin{cases} B_0 \hat{z}, & \text{if } \rho < 850 \text{ km}, \\ B_0 \left(\frac{\rho}{1 \text{ km}}\right)^{-0.37} \hat{z}, & \text{if } 850 \text{ km} \leq \rho \leq 900 \text{ km}, \\ B_0 \left(\frac{\rho}{1 \text{ km}}\right)^{-0.99} \hat{z}, & \text{otherwise}, \end{cases} \tag{32}$$

such that the central and the surface magnetic field are 2×10^{15} G and around 10^{12} G respectively. Also, $(\mathbf{B} \cdot \nabla)\mathbf{B} = 0$ and $\nabla \cdot \mathbf{B} = 0$ for the chosen profile. Therefore, the non-rotating white dwarfs will be spherical in shape. Hence, in spherical polar coordinates with $\theta = \pi/2$, the field profile is given by

$$B = \begin{cases} -B_0 \hat{\theta}, & \text{if } r < 850 \text{ km}, \\ -B_0 \left(\frac{r}{1 \text{ km}}\right)^{-0.37} \hat{\theta}, & \text{if } 850 \text{ km} \leq r \leq 900 \text{ km}, \\ -B_0 \left(\frac{r}{1 \text{ km}}\right)^{-0.99} \hat{\theta}, & \text{otherwise}, \end{cases} \tag{33}$$

The mass and radius of a white dwarf can be obtained by solving

$$\frac{d}{dr}\left(P_e + \frac{B^2}{8\pi}\right) = -\frac{GM(r)(\rho_e + \rho_B)}{r^2}, \tag{34}$$

$$\frac{dM(r)}{dr} = 4\pi r^2 (\rho_e + \rho_B), \tag{35}$$

where ρ_B is the magnetic density, $B^2 = \mathbf{B} \cdot \mathbf{B}$, $\rho_e = n_e m_p \mu_e$, m_p is the mass of proton, μ_e is the mean molecular weight per electron and G is Newton's gravitation constant.

The mass-radius relation, shown in Fig. 6(b), depicts the existence of super-Chandrasekhar mass white dwarfs in presence of central strong magnetic field. Although, here, the surface magnetic field obtained is around 10^{12} G, the results are unaffected even if the surface magnetic field of white dwarfs is lower, say around 10^9 G which is detectable.

Here, for $r < 850$ km, since, magnetic field is uniform, EoS would be Landau quantized for the same magnetic field at $B_0 = 2 \times 10^{15}$ G. For $r \geq 850$ km, the field decays to a lower strength so that Chandrasekhar's non-magnetic EoS suffices. Only at the interface around 850 km, non-uniform field based LQ applies in EoS, but in a very tiny region. Hence practically, for the present example, LQ EOS based on non-uniform field does not play an important role in controlling the dynamics of magnetised white dwarf. However, it helps to account for the whole system including the small zone around 850 km that contains the decay of the magnetic field.

6. Summary

LQ in strictly spatially variable magnetic fields is a new venture on its own. It leads to the different alignment of energy levels of an electron as well as the breaking of spin degeneracy, depending on the variation of magnetic field. Hence, it can be useful in the multitude of fields wherein one can attain the desired spacing and alignment of levels through an appropriate non-uniformity in magnetic fields.

We have explored its application in quantum computing through an increase in quantum speed of electron in spatially increasing magnetic field. We also discuss the increase in stiffness and softenss of EOS at high and low densities in presence of non-uniform magnetic fields as compared to the case of uniform magnetic field (as well as Chandrasekhar EOS). This, we, further, use to understand the properties of magnetized white dwarfs in the small zone of rapidly decaying magnetic field.

References

1. L. D. Landau and E. M. Lifshitz, *Quantum mechanics: Non-Relativistic Theory*, 2nd edn. (Elsevier, Amsterdam, the Netherlands, 1965).
2. P. Strange, *Relativistic Quantum Mechanics* (Cambridge University Press, Cambridge, 1998).
3. U. Das and B. Mukhopadhyay, Strongly magnetized cold degenerate electron gas: Mass-radius relation of the magnetized white dwarf, *Phys. Rev. D* **86**, 042001 (2012).
4. S. Aggarwal, B. Mukhopadhyay and G. Gregori, Relativistic landau quantization in non-uniform magnetic field and its applications to white dwarfs and quantum information, *SciPost Physics* (2021), *accepted for publication*, arXiv:2110.09543.
5. D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd edn. (Cambridge University Press, Cambridge, 2017).
6. L. Mandelstam and I. Tamm, The energy–time uncertainty relation in non-relativistic quantum mechanics, *Izv. Akad. Nauk SSSR* **9**, 122 (1945).

Validity of the semiclassical approximation in 1+1 electrodynamics: Numerical solutions to the linear response equation

Ian M. Newsome
Department of Physics, Wake Forest University, Winston-Salem, NC 27109, USA
Email: newsim18@wfu.edu

From previous work,[1] the semiclassical backreaction equation in 1+1 dimensions was solved and a criterion was implemented to assess the validity of the semiclassical approximation in this case. The criterion involves the behavior of solutions to the linear response equation which describes perturbations about solutions to the semiclassical backreaction equation. The linear response equation involves a time integral over a two-point correlation function for the current induced by the quantum field and it is expected that significant growth in this two-point function (and therefore in quantum fluctuations) will result in significant growth in solutions to the linear response equation. It was conjectured for early times that the difference of two nearby solutions to the semiclassical backreaction equation, with similar initial conditions, can act as an approximate solution to the linear response equation. A comparative analysis between the approximate and numerical solutions to the linear response equation, for the critical scale for particle production, will be presented for the case of a massive, quantized spin ½ field in order to determine how robust the approximation method is for representing its solutions.

Keywords: Schwinger effect; semiclassical electrodynamics; linear response; validity.

1. Introduction

The semiclassical approximation has been utilized in a wide range of physical situations where a quantized field evolves in a classical background. The relevant scenario for the following discussion involves the decay of an electric field via the Schwinger effect.[2] The original calculation by Schwinger in 1951 involved a background field calculation in which an electric field E, that is constant in space and time, gives rise to a particle production rate due to vacuum decay. At leading one-loop order the particle production rate is proportional to a factor of $\exp\{-\pi m^2/qE\}$, from which one can define a critical scale for pair production that is $E_{\text{crit}} \sim m^2/q$.

The time evolution of the net electromagnetic field is described by the semiclassical backreaction equation and has been solved for the case of a homogeneous electric field coupled to a massive scalar or a massive spin ½ field in $1{+}1^{3-5}$ and $3{+}1^{5-7}$ dimensions. When particles are produced, they accelerate in reaction to the background electric field, creating an electric current. This current produces a counter electric field which initially damps the original background electric field, and at late times the net electric field will oscillate.

Given that the semiclassical electrodynamics model is an approximation to quantum electrodynamics, it must be asked to what extent is this an accurate

approximation? Concerning the model presented here, the semiclassical backreaction equation involves an expectation value of the current constructed from quantum fields and for this to be a satisfactory approximation to what one would measure, it is necessary that quantum fluctuations be small.

To summarize the previous work which led to the development of the numerical solution to the linear response equation presented in these proceedings, the main goals in [1] had been to study the details of the particle production process when backreaction effects were taken into account and subsequently to estimate the significance of certain types of quantum fluctuations to investigate the validity of the semiclassical approximation for 1+1 electrodynamics. This included solving the necessary backreaction equations for scalar and spin ½ fields coupled to two different classical source profiles, one which led to an asymptotically constant electric field and the other being the Sauter pulse. To wit, a criterion was implemented to assess the validity of the semiclassical approximation which had been previously applied to the process of preheating in models of chaotic inflation,[8] with an earlier version applied to semiclassical gravity.[9] This criterion involves the behavior of solutions to the linear response equation, which can be derived by perturbing the semiclassical backreaction equation about a background field solution.

In general, the linear response equation is a second order integro-differential equation which can be cumbersome to solve numerically. Therefore a method was developed[8] to approximate solutions to the linear response equation which involves the difference between solutions to the semiclassical backreaction equation with similar initial conditions. For early times, this difference is expected to act as an approximate solution to the linear response equation as long as the exact solution is relatively small.

Subsequently, an effort has been made to directly solve for the numerical solutions to the linear response equation for the case of an asymptotically constant background electric field, initially zero, coupled to a spin ½ field. For the purposes of this proceeding, the numerical solutions to the linear response equation are desired in order to determine how robust the method for approximating its solutions will be. Therefore, a comparative analysis between the numerical solutions to the linear response equation and its approximate solutions described previously will be presented in what follows, with special attention given to the critical scale for particle production.

The structure of this paper is as follows: in Sec. 2 a brief review of the system is discussed including quantization of the spin ½ field, the semiclassical backreaction equation, and the renormalization procedure used. Sec. 3 contains a review of relevant material[1] including the criterion for the validity of the semiclassical approximation with the general and specific forms of the linear response equation presented for the case of a spin ½ field coupled to a background electric field. Also, included is a discussion of the method used to approximate solutions to the linear response equation. In Sec. 4 an analysis of the numerical solutions to the linear

response equation is given, with a comparison drawn between its approximate solutions, with an emphasis on the critical scale for significant particle production. Sec. 5 includes a summary of results and discussion thereof.

2. Quantization and Renormalization of a Spin ½ Field

This section describes the model under consideration: a quantized spin ½ field which interacts with a background electric field generated by a prescribed classical source. Analysis is restriced to 1+1 Minkowski spacetime with the assumption that the background electric field is spatially homogeneous in a given reference frame, i.e. $E = E(t)$. Furthermore, the metric signature is chosen to be $(-,+)$ with $c = \hbar = 1$.

The classical action describing a spin ½ field $\psi(t,x)$ coupled to a background electric field is given by

$$S = \int d^2x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + A_\mu J_C^\mu + i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi \right] . \tag{1}$$

Here $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength tensor, m the mass of spin ½ field excitations, $D_\mu = \partial_\mu - iqA_\mu$ the gauge covariant derivative, J_C^μ a classical and conserved external source, and $\bar{\psi} = \psi^\dagger \gamma^0$. The Dirac matrices γ^μ satisfy the following anticommutation relations $\{\gamma^\mu, \gamma^\nu\} = -2\eta_{\mu\nu}$.

Variation of (1) with respect to the vector potential yields the general form of Maxwell's equation

$$-\Box A^\mu + \partial^\mu \partial_\nu A^\nu = J_C^\mu + J_Q^\mu \ . \tag{2}$$

The classical source and background electric field generated by this source is chosen to be

$$J_C = -\frac{qE_0}{(1+qt)^2} \ , \quad E_C = -\int J_C \, dt = E_0 \left(\frac{qt}{1+qt} \right) , \tag{3}$$

for $t \geq 0$ and $J_C = 0$ for $t < 0$. Note, the term J_C in (3) is the spatial component of J_C^μ. Since the classical current is initially zero, and gives rise to an electric field that is initially zero as well, there is no ambiguity in the choice of vacuum state. The source term J_Q^μ induced by the spin ½ field is given by

$$J_Q^\mu = q\bar{\psi}\gamma^\mu \psi \ . \tag{4}$$

Variation of (1) with respect to the field $\bar{\psi}$ yields the Dirac equation

$$(i\,\gamma^\mu D_\mu - m)\,\psi(t,x) = 0 \ . \tag{5}$$

In what follows, the gauge choice

$$A^\mu = (0, A(t)) \ , \tag{6}$$

will be implemented. Expanding the spin ½ field $\psi(t,x)$ in terms of a complete set of modes yields

$$\psi(t,x) = \int_{-\infty}^{\infty} dk \, \left[B_k u_k(t,x) + D_k^\dagger v_k(t,x) \right] , \tag{7}$$

with $B_k, B_k^\dagger, D_k, D_k^\dagger$ the usual creation and annihilation operators obeying the anticommutation relations $\{B_k, B_{k'}\} = \{D_k, D_{k'}^\dagger\} = \delta(k - k')$. Utilizing a particular form for the modes,[10] two independent spinor solutions can be constructed as follows

$$u_k(t,x) = \frac{e^{ikx}}{\sqrt{2\pi}} \begin{pmatrix} h_k^I(t) \\ \\ -h_k^{II}(t) \end{pmatrix}, \quad v_k(t,x) = \frac{e^{-ikx}}{\sqrt{2\pi}} \begin{pmatrix} h_{-k}^{II\,*}(t) \\ \\ h_{-k}^{I\,*}(t) \end{pmatrix}. \tag{8}$$

Using the Weyl representation of the Dirac matrices γ^μ

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{9}$$

the functions $h_k^I(t)$ and $h_k^{II}(t)$ satisfy the mode equations

$$\dot{h}_k^I - i(k - qA)h_k^I - imh_k^{II} = 0\ , \tag{10a}$$

$$\dot{h}_k^{II} + i(k - qA)h_k^{II} - imh_k^I = 0\ . \tag{10b}$$

The time evolution of this system is governed by the semiclassical backreaction equation. This can be obtained by replacing J_Q^μ in (2) with its expectation value $\langle J_Q^\mu \rangle$ and then use the gauge choice (6) with (7) to yield

$$\frac{d^2}{dt^2}A(t) = -\frac{d}{dt}E(t) = J_C + \langle J_Q \rangle\ . \tag{11}$$

Due to the coupling between the classical background electric field and the quantized spin ½ field, charged spin ½ particles will be created. These particles will be accelerated by the background electric field, creating a current which generates a secondary electric field. In the semiclassical approximation, the current created from the accelerated spin ½ particles is given by $\langle J_Q \rangle$. The renormalized expression for $\langle J_Q \rangle$, evaluated in the vacuum state is[1]

$$\langle J_Q \rangle_{\rm ren} = \frac{q}{2\pi}\int_{-\infty}^{\infty} dk\ \left[|h_k^I(t)|^2 - |h_k^{II}(t)|^2 + \frac{k}{\omega} - \frac{q\,m^2}{\omega^3}A(t) \right]. \tag{12}$$

with $\omega^2 = k^2 + m^2$. Here adiabatic regularization was used to eliminate the ultraviolet divergences.[10]

3. Validity Criterion for the Semiclassical Approximation

This section gives a review of relevant material[1] required for the investigation of the solutions to the linear response equation.

Since the current term $\langle J_Q \rangle$, in part, characterizes the quantum particle production process, a natural way to determine the size of quantum fluctuations, compared to other relevant quantities, is to evaluate a two-point correlation function for the current. In general, there are a number of different correlation functions which could be used, but in order to avoid such problems as state-dependent divergences,[11]

incompatible results from various renormalization techniques,[12] or covariance issues,[9] it is useful to proceed with a two-point correlation function which emerges naturally from the semiclassical theory itself, namely $\langle [J_Q(t,x), J_Q(t^{'},x^{'})]\rangle$. This two-point function measures the extent to which the value of the current $\langle J_Q\rangle$ at two separate spacetime points commutes, thereby having the interpretation of characterizing the degree to which quantum fluctuations are introduced into the system.

Perturbing the semiclassical backreaction equation yields the linear response equation which contains this two-point correlation function and describes the time evolution of perturbations about a given semiclassical solution. A criterion for the validity of the semiclassical approximation was originally developed for semiclassical gravity[9] and modified for preheating in chaotic inflation.[8] This criterion was applied to semiclassical electrodynamics[1] and states: the semiclassical approximation will break down if any linearized gauge invariant quantity constructed from solutions to the linear response equation with finite non-singular initial data grows rapidly for some period of time. It is important to note this is a necessary, but not sufficient condition for the validity of the semiclassical approximation.

3.1. *The Linear Response Equation*

The linear response equation for semiclassical electrodynamics is found by perturbing the semiclassical backreaction equation about a given solution. From (11) this becomes

$$\frac{d^2}{dt^2}\delta A(t) = -\frac{d}{dt}\delta E(t) = \delta J_C + \delta\langle J_Q\rangle \ . \tag{13}$$

More specifically, the type of perturbation being performed is one that changes the classical current by altering the value of the classical background electric field amplitude E_0 in (3). Thus the term δJ_C is expressed as

$$\delta J_C = -\frac{q}{(1+qt)^2}\delta E_0 \ . \tag{14}$$

In conjunction with the validity criterion stated previously, it is useful to break up the solutions to the semiclassical backreaction equation into the purely classical and quantum pieces

$$E_C = -\int_{t_0}^{t} J_C(t_1)\, dt_1 \ , \tag{15a}$$

$$E_Q = E - E_C \ . \tag{15b}$$

Solutions to the linear response equation can be partitioned in the same way which allows for the statement of the validity criterion to be modified to state that if δE_Q grows rapidly during some time interval, then the semiclassical approximation is not valid.

For the case of a spin ½ field, the renormalized $\delta\langle J_Q\rangle$ term present in (13) can be expressed as[1]

$$\delta\langle J_Q\rangle_{\rm ren} = -\frac{q^2}{\pi}\delta A(t) + i\int_{-\infty}^{\infty} dx^{'}\int_{-\infty}^{t} dt^{'}\,\langle[J_Q(t,x), J_Q(t^{'},x^{'})]\rangle\,\delta A(t^{'})\ , \tag{16}$$

with the relationship

$$\int_{-\infty}^{\infty} dx^{'}\,\langle[J_Q(t,x), J_Q(t^{'},x^{'})]\rangle = \frac{4iq^2}{\pi}\int_{-\infty}^{\infty} dk\,{\rm Im}\left\{h_k^I(t)h_k^{II}(t)h_k^{I*}(t^{'})h_k^{II*}(t^{'})\right\}. \tag{17}$$

Thus (13) takes the specific form

$$\begin{aligned}\frac{d^2}{dt^2}\delta A(t) = &-\frac{q}{(1+qt)^2}\delta E_0 - \frac{q^2}{\pi}\delta A(t)\\ &-\frac{4q^2}{\pi}\int_{-\infty}^{\infty} dk\int_{-\infty}^{t} dt^{'}\,{\rm Im}\left\{h_k^I(t)h_k^{II}(t)h_k^{I*}(t^{'})h_k^{II*}(t^{'})\right\}\delta A(t^{'})\ .\end{aligned} \tag{18}$$

3.2. *Approximate Solutions to the Linear Response Equation*

A technique that was developed to approximate solutions to the linear response equation for the case of homogeneous perturbations[1] involves solving the semiclassical backreaction equation for two solutions whose initial conditions are similar. It is expected that the difference between these two solutions ΔE is an approximate solution to the linear response equation δE, for early times, as long as $\Delta E \approx \delta E$. If the difference does grow large, then the corresponding solution to the linear response equation should grow significantly as well. This would violate the criterion discussed above and thus signal a breakdown in the semiclassical approximation.

A way to measure the relative growth of two solutions E_1 and E_2 to the semiclassical backreaction equation is with the modified relative difference expression[1]

$$R = \frac{|\Delta E|}{|E_1| + |E_2|}\ , \quad \Delta E = E_2 - E_1\ . \tag{19}$$

From the semiclassical backreaction equation, it is clear the difference ΔE will be a solution of the following equation

$$-\frac{d}{dt}\Delta E = \Delta J_C + \Delta\langle J_Q\rangle\ , \tag{20}$$

with the definitions

$$\Delta J_C = J_{C,2} - J_{C,1}\ , \quad \Delta\langle J_Q\rangle = \langle J_{Q,2}\rangle - \langle J_{Q,1}\rangle\ . \tag{21}$$

For ΔE to be an approximate solution to the linear response equation, i.e. $\Delta E \approx \delta E$, it is clear that $\Delta\langle J_Q\rangle \approx \delta\langle J_Q\rangle$ must hold, since one can set $\Delta J_C = \delta J_C$ for all times.

A natural way to measure the growth of the approximate quantum contribution to the finite difference equation ΔE_Q is with a relative difference R_Q given as

$$R_Q = \frac{|\Delta E_Q|}{|E_{Q,1}| + |E_{Q,2}|} \ , \quad \Delta E_Q = E_{Q,2} - E_{Q,1} \ . \tag{22}$$

This acts as the gauge invariant quantity constructed from solutions to the linear response equation mentioned in the formal statement for the validity criterion. This difference can be compared to the relative difference R_C between corresponding classical solutions, which does not vary in time. Therefore if R_Q grows rapidly for some period of time, this signals a breakdown in the semiclassical approximation.

In Fig. 1, some results[1] are given for the quantity R_Q for a variety of different particle masses[a]. The most significant effect on the behavior of R_Q is the size of the characteristic dimensionless quantity ${}^{qE_0}/_{m^2}$. It is therefore useful to distinguish between three different regimes: (i) ${}^{qE_0}/_{m^2} \gg 1$ in which the mass is relatively small compared to the electric field, resulting a large amount of particle production, (ii) ${}^{qE_0}/_{m^2} \sim 1$ in which the mass is of the same order as the electric field, resulting in less but still significant particle production, and (iii) ${}^{qE_0}/_{m^2} \ll 1$ in which the mass is relatively large compared to the electric field, resulting in very little particle production. The critical case for particle production is defined as $E \sim E_{\text{crit}} = m^2/q$ and therefore the most relevant mass value is $m^2/q^2 = 1$.

As seen in Fig. 1, this particular mass value corresponds to the most rapid growth of R_Q based on the cases considered here, and this implies the most severe breakdown of the semiclassical approximation for this case. However, one can see that R_Q has the least amount of growth in the small mass regime and steadily increases its early time growth as the mass is increased up to the critical case. The conclusion drawn in [1], in the context of the validity criterion put forward, is that the semiclassical approximation is most greatly violated when ${}^{qE_0}/_{m^2} \sim 1$ and becomes more accurate as the particle mass is decreased (or increased since quantum fluctuations are expected to diminish due to lack of particle production in the high mass regime).

4. Numerical Solutions to the Linear Response Equation

In this section, an analysis of the numerical solution to the linear response equation for the critical case in which the characteristic quantity ${}^{qE_0}/_{m^2} = 1$ is presented. A comparison is then drawn between its approximate and numerical solutions in the context of the validity criterion stated previously.

To arrive at numerical solutions to the linear response equation, one can partition the problem into two parts. The first part involves solving the semiclassical backreaction equation, and the second involves using the output from the

[a]This is an aggregate plot built from the results and data presented in [1].

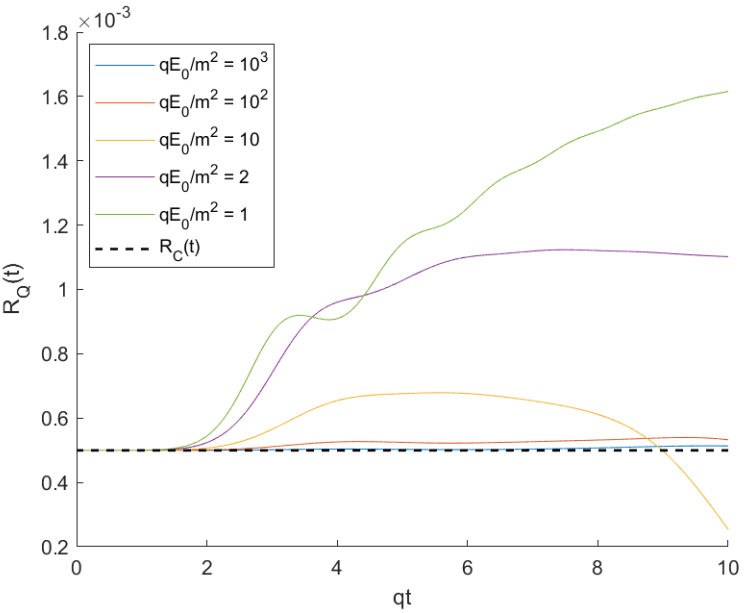


Fig. 1. Results obtained for the quantity R_Q. The values $E_{0,1}/q = 1$ and $E_{0,2}/q = 1 + 10^{-3}$ have been chosen for the solutions which comprise ΔE. The values of the characteristic quantity qE_0/m^2 for each case are shown along with the type of curve for that solution in the legend. Here $R_C(t)$ denotes the relative difference excluding quantum effects.

semiclassical backreaction equation to supply the necessary data for solving the linear response equation. Generating the solutions to the semiclassical backreaction equation involves simultaneously solving a set of three coupled differential equations, namely (10a), (10b), and (11) with (12). The solution output will be the vector potential $A(t)$, and by extension the electric field $E(t)$, and the time dependent spin ½ field modes $h_k^{I,II}(t)$. The quantities needed for solving the linear response equation (18) are the modes $h_k^{I,II}(t)$, which the two-point function in (17) depends on.

4.1. *Numerical Details*

The $|in\rangle$ state of the time dependent modes $h_k^{I,II}(t)$, prior to the classical source being turned on, is defined as positive frequency plane wave solutions to the mode equation. When the source term is turned on at $qt = 0$ and subsequently $A(t) \neq 0$, the mode solutions are altered by the presence of the background field, specifically $k \to k - qA(t)$. This leads to particle production which is measured, in part, by the quantity $\langle J_Q \rangle$. Since the two-point function $\langle [J_Q(t,x), J_Q(t^{'}, x^{'})] \rangle$ depends on these time dependent modes in a non-trivial way, the modification to the mode solutions will induce a non-zero contribution to the two-point function.

It is useful to separate the time dependent modes $h_k^{I,II}(t)$ into their real and imaginary parts

$$h_k^I = \mathrm{Re}\{h_k^I\} + i\,\mathrm{Im}\{h_k^I\}\ , \tag{23a}$$

$$h_k^{II} = \mathrm{Re}\{h_k^{II}\} + i\,\mathrm{Im}\{h_k^{II}\}\ . \tag{23b}$$

From this, the two mode equations defined in (10a) and (10b) now become four mode equations, given as

$$\mathrm{Re}\{\dot{h}_k^I\} + (k - qA)\,\mathrm{Im}\{h_k^I\} + m\,\mathrm{Im}\{h_k^{II}\} = 0\ , \tag{24a}$$

$$\mathrm{Im}\{\dot{h}_k^I\} - (k - qA)\,\mathrm{Re}\{h_k^I\} - m\,\mathrm{Re}\{h_k^{II}\} = 0\ , \tag{24b}$$

$$\mathrm{Re}\{\dot{h}_k^{II}\} - (k - qA)\,\mathrm{Im}\{h_k^{II}\} + m\,\mathrm{Im}\{h_k^I\} = 0\ , \tag{24c}$$

$$\mathrm{Im}\{\dot{h}_k^{II}\} + (k - qA)\,\mathrm{Re}\{h_k^{II}\} - m\,\mathrm{Re}\{h_k^I\} = 0\ . \tag{24d}$$

Furthermore, the quantity $\langle J_Q\rangle$ in (12) can now be expressed in the following way

$$\langle J_Q\rangle_{\mathrm{ren}} = -\frac{q^2}{\pi}A(t) + \frac{q}{2\pi}\int_{-\infty}^{\infty} dk\left[\mathrm{Re}^2\{h_k^I\} + \mathrm{Im}^2\{h_k^I\} - \mathrm{Re}^2\{h_k^{II}\} - \mathrm{Im}^2\{h_k^{II}\} + \frac{k}{\omega}\right]. \tag{25}$$

It is straightforward to now implement a numerical routine which will solve Eqs. (10) and (11).

From (17) it is clear that $\langle[J_Q(t,x), J_Q(t',x')]\rangle$ depends on the modes $h_k^{I,II}(t)$. However, the time integral present in (18) depends on both the current time t and the integration variable t'. In order to make this integral suitable for numerical evaluation, it is convenient to utilize (23a) and (23b) which allows it to be expressed as

$$\begin{aligned}
&\int_{-\infty}^{t} dt'\ \mathrm{Im}\big\{h_k^I(t)h_k^{II}(t)h_k^{I*}(t')h_k^{II*}(t')\big\}\delta A(t') \\
&\quad= \Big(\mathrm{Im}\big\{h_k^I(t)\big\}\mathrm{Im}\big\{h_k^{II}(t)\big\} - \mathrm{Re}\big\{h_k^I(t)\big\}\mathrm{Re}\big\{h_k^{II}(t)\big\}\Big) \\
&\qquad\times\int_{-\infty}^{t} dt'\ \Big(\mathrm{Re}\big\{h_k^{I*}(t')\big\}\mathrm{Im}\big\{h_k^{II*}(t')\big\} + \mathrm{Im}\big\{h_k^{I*}(t')\big\}\mathrm{Re}\big\{h_k^{II*}(t')\big\}\Big)\delta A(t') \\
&\qquad+\Big(\mathrm{Re}\big\{h_k^I(t)\big\}\mathrm{Im}\big\{h_k^{II}(t)\big\} + \mathrm{Im}\big\{h_k^I(t)\big\}\mathrm{Re}\big\{h_k^{II}(t)\big\}\Big) \\
&\qquad\times\int_{-\infty}^{t} dt'\ \Big(\mathrm{Re}\big\{h_k^{I*}(t')\big\}\mathrm{Re}\big\{h_k^{II*}(t')\big\} - \mathrm{Im}\big\{h_k^{I*}(t')\big\}\mathrm{Im}\big\{h_k^{II*}(t')\big\}\Big)\delta A(t')\ .
\end{aligned} \tag{26}$$

Here the inter-dependence of t and t', seen in the product $h_k^I(t)h_k^{II}(t)h_k^{I*}(t')h_k^{II*}(t')$, has been separated out. This allows for the time integral to be computed which then acts as the integrand for the k-integral in (18), which can be computed as well. From this the numerical solutions $\delta A(t)$ and $\delta E(t)$ to the linear response equation can be found.

4.2. *Results and Discussion*

In what follows, numerical results are presented for the critical case of ${}^{qE_0}/_{m^2} = 1$. In Fig. 2, the net electric field $E(t)$ including backreaction is plotted[1] with both the electric field contributions E_C when quantum effects are absent and E_Q generated by only quantum effects, seen in (15b) and (15a). One can see that at early times the net electric field begins to dampen compared to its classical counterpart E_C. This is due to the backreaction of the produced particles, resulting from the coupling of the spin ${}^1/_2$ field to the classical background source, causing an increase in the field E_Q.

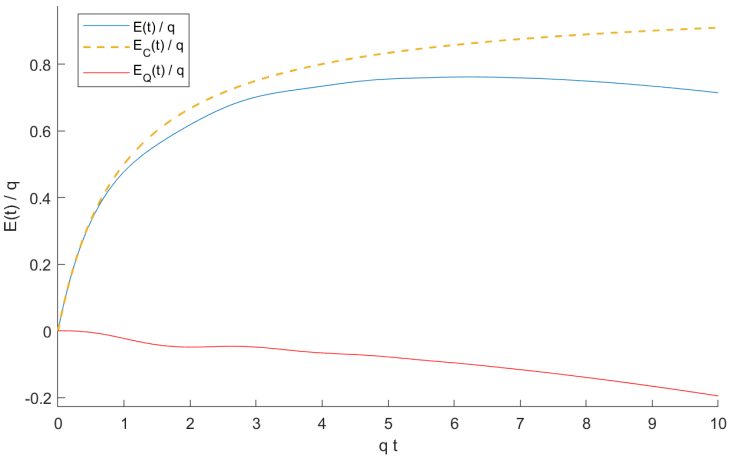


Fig. 2. A plot of the electric field including backreaction. The net electric field $E(t)$ is given by the blue curve, the electric field $E_C(t)$ with no quantum effects present is given by the yellow dashed curve, and the electric field $E_Q(t)$ generated by only quantum effects is given by the red curve.

Numerical solutions to the linear response equation are plotted in Fig. 3. The top plot shows the numerical solution δE to the linear response equation, its approximate solution given by the finite difference between two nearby semiclassical backreaction equation solutions ΔE, and the classical pieces $\delta E_C = \Delta E_C$. The bottom plot shows the purely quantum contributions to both the approximate ΔE_Q and numerical δE_Q solutions to the linear response equation. From (14) the initial value δE_0 was chosen in such a way as to equal the finite difference ΔE at early times, with this value being of order 10^{-3}.

For early times, there appears to be agreement between δE and ΔE, as well as ΔE_Q and δE_Q since the classical contribution dominates due to a lack particle production. However, near the time $qt = 1$ significant deviation between the two begins to occur, as could be measured by $(\delta E - \Delta E)/\delta E$. This deviation and subsequent late time growth in δE is driven by the dependence of $\langle [J_Q(t,x), J_Q(t^{'}, x^{'})] \rangle$

on the modes $h_k^{I,II}(t)$, whose positive frequency plane wave solutions are being altered due to the presence of the classical source term $A(t)$ having been switched on and leading to particle production. Since the critical case ${}^{qE_0}/_{m^2} = 1$ considered is a threshold case for significant particle production, above which even more particle production will occur, the growth of δE due to $\langle [J_Q(t,x), J_Q(t^{'},x^{'})] \rangle$ is expected.

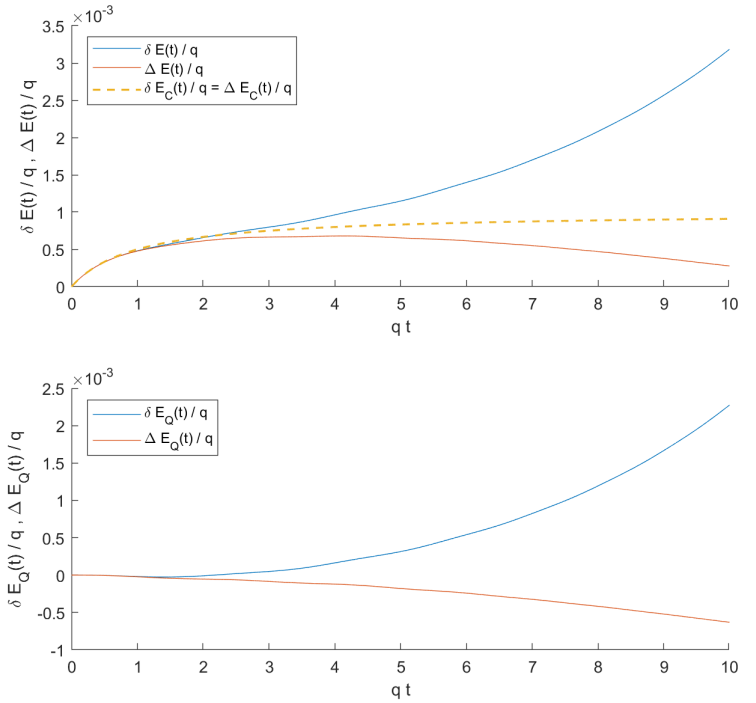


Fig. 3. The top plot shows solutions to the linear response equation, with red and blue curves representing the approximate and numerical solutions ΔE and δE, respectively, with the yellow dashed curve being the classical contribution to the solution. The bottom plot isolates the quantum contribution to the linear response equation, for both the approximate and numerical solutions ΔE_Q and δE_Q. All plots reflect the critical case of ${}^{qE_0}/_{m^2} = 1$ and were done with δE_0 and ΔE of order 10^{-3} for early times.

The time $qt = 1$ is also when one begins to notice substantial deviation between the net and classical electric fields, seen in Fig. 2, due to the presence of a nontrivial $\langle J_Q \rangle$ factor in (11) which will grow as more particles are produced. Furthermore, this time is also when one sees substantial growth of R_Q in Fig. 1 for the critical case. It therefore appears that for early times the ΔE is a good approximation to the linear response equation solution δE, and for late times $\delta E \neq \Delta E$. Comparing to the corresponding R_Q behavior, there is a breakdown in the semiclassical approximation based on the validity criterion implemented here. This was the conclusion that was

arrived at in [9], and it appears the behavior of the solutions δE as compared with the ΔE validates this conclusion.

5. Conclusions

In previous work,[1] a criterion for the validity of the semiclassical approximation was applied to the case of 1+1 semiclassical electrodynamics which involves the behavior of solutions to the linear response equation. The linear response equation depends, in part, on a two-point correlation function $\langle [J_Q(t,x), J_Q(t^{'},x^{'})\rangle$, which characterizes the quantum fluctuations introduced through the current $\langle J_Q\rangle$. Numerical solutions to the linear response equation have been obtained for 1+1 semiclassical electrodynamics using a model of the Schwinger effect in which particle production occurs in the presence of a strong, spatially homogeneous, electric field. The particle production, in the context of the system considered here, is a consequence of the coupling between a quantized spin ½ field and a classical, asymptotically constant, background electric field generated by an external source. Since the linear response equation depends on $\langle [J_Q(t,x), J_Q(t^{'},x^{'})]\rangle$, if there is rapid growth in this correlation function for some period of time, then quantum fluctuations must be significant, and will drive the growth of solutions δE to the linear response equation, resulting in a breakdown of the semiclassical approximation.

A previously developed technique to approximate solutions to the linear response equation involves the difference ΔE between two solutions to the semiclassical backreaction equation with similar initial conditions. The behavior of a quantity R_Q built from this difference of semiclassical backreaction equation solutions gives a measure for the validity of the semiclassical approximation. When R_Q grows large over some time, the corresponding linear response equation solution is expected to do so as well.

An analysis comparing the numerical δE and approximate ΔE solutions to the linear response equation has been conducted. At the critical scale, where ${}^{qE_0}/_{m^2} \sim 1$, it was found that at early times the quantity ΔE approximated the linear response equation solutions δE quite well. It was only after enough time had passed and significant particle production occurred that the numerical and approximate solutions deviated from one another. Therefore the claim that ΔE adequately approximates solutions to the linear response equation for early times is substantiated in this case. This critical scale was also the case for which the quantity R_Q had its largest growth[1] and therefore was the case considered for the purposes of this proceeding. The significant growth in both δE and R_Q signals a breakdown of the semiclassical approximation based on the validity criterion utilized here.

Regarding future work, the numerical solutions to the linear response equation will be used to further investigate the nature of how quantum fluctuations affect these solutions. More specifically, since the linear response equation involves several factors which dictate the behavior of its solutions, one of which is a term characterizing quantum fluctuations introduced through the particle production process,

having the numerical solution will allow one to isolate and investigate its respective contributions in detail.

Acknowledgements

I. M. N. would like thank Silvia Pla for sharing semiclassical backreaction numerical code and data as well as helpful discussions, Eric Grotzke for helpful discussions regarding linear response equation numerical details, Paul R. Anderson for guidance regarding the creation of this manuscript as well as helpful discussions, and Jose Navarro-Salas and Kaitlin Hill for helpful discussions. All numerical work was performed using MATLAB software with special thanks to the Distributed Environment for Academic Computing (DEAC) at Wake Forest University for providing HPC resources which contributed to the research presented in this proceeding.

References

1. S. P. Pla, I. M. Newsome, R. S. Link, P. R. Anderson, and J. Navarro-Salas, Phys. Rev. D **103**, 105003 (2021).
2. J. Schwinger, Phys. Rev. **82**, 664 (1951).
3. Y. Kluger, J. M. Eisenberg, B. Svetitsky, F. Cooper and E. Mottola, Phys. Rev. Lett. **67**, 2427 (1991).
4. Y. Kluger, J. M. Eisenberg, B. Svetitsky, F. Cooper and E. Mottola, Phys. Rev. D **45**, 4659 (1992).
5. Y. Kluger, J. M. Eisenberg, and B. Svetitsky, Int. J. Mod. Phys. E **02**, 333 (1993).
6. N. Tanji, Ann. Phys. (Amsterdam) **324**, 1691 (2009).
7. F. Gelis and N. Tanji, Phys. Rev. D **87**, 125035 (2013).
8. P. R. Anderson, C. Molina-Paris, and D. H. Sanders, Phys. Rev. D **92**, 083522 (2015).
9. P. R. Anderson, C. Molina-Paris, and E. Mottola, Phys. Rev. D **67**, 024026 (2003).
10. A. Ferreiro and J. Navarro-Salas, Phys. Rev. D **97**, 125012 (2018).
11. C.-H. Wu and L. H. Ford, Phys. Rev. D **60**, 104013 (1999).
12. N. G. Phillips and B. L. Hu, Phys. Rev. D **62**, 084017 (2000).

On Kerr black hole perfect MHD processes in Doran coordinates

C. Cherubini

Department of Science and Technology for Humans and the Environment and ICRA, Campus Bio-Medico University of Rome, Rome, 00128 Italy
and International Center for Relativistic Astrophysics Network-ICRANet, 65122 Pescara, Italy
E-mail: c.cherubini@unicampus.it
www.unicampus.it

S. Filippi

Department of Engineering and ICRA, Campus Bio-Medico University of Rome, Rome, 00128 Italy
and International Center for Relativistic Astrophysics Network-ICRANet, 65122 Pescara, Italy
E-mail: s.filippi@unicampus.it
www.unicampus.it

A. Loppini

Department of Engineering, Campus Bio-Medico University of Rome, Rome, 00128 Italy
Center for Life Nano- & Neuro-Science, Istituto Italiano di Tecnologia (IIT), Viale Regina Elena 291, 00161 Roma, Italy
E-mail: a.loppini@unicampus.it
www.unicampus.it

R. Ruffini*, R. Moradi† and Y. Wang‡

ICRANet, Piazza della Repubblica 10, Pescara, I–65122, Italy,
ICRA, Dipartimento di Fisica, Sapienza Universita' di Roma, Rome, I–00185, Italy
and INAF, Rome, 00136, Italy
**E-mail: ruffini@icra.it*
†*E-mail: rahim.moradi@icranet.org*
‡*E-mail: yu.wang@icranet.org*
www.icranet.org

S.S. Xue

ICRANet, Piazza della Repubblica 10, Pescara, I–65122, Italy,
ICRA, Dipartimento di Fisica, Sapienza Universita' di Roma, Rome, I–00185, Italy
E-mail: xue@icra.it
www.icranet.org

Doran horizon penetrating coordinates are adopted to study specific perfect MHD processes around a Kerr black hole, focusing in particular on the physical relevance of selected electrodynamical quantities.

Keywords: Black hole physics, Relativistic MHD, Analytical solutions.

1. Introduction

General Relativity is a non linear theory which couples any physical field to gravitation described in terms of the geometry of space-time. This makes the search of exact or approximate analytical solutions for physical relevant problems an extraordinary task. Nowadays, Numerical Relativity has greatly circumvented such a problem by using the huge computational power available[1] although analytical solutions still remain fundamental to have important physical insights. The $60s$ and the $70s$ represented an important moment of rediscovery of Einstein's Theory. At that time, numerical relativity was almost unknown and Mathematical Physics techniques were widely adopted to find exact or approximate solutions. In particular, in the field of Black Holes Physics, the understanding of the electrodynamics and magnetohydrodynamics (MHD) around the recently discovered Kerr rotating black hole solution[2] was central. Due to the aforementioned difficulty associated to the non-linearity of the equations, approximate techniques of perturbative type were applied to this aim as, for instance, in the classical work by Ruffini and Wilson[3] and the one by Damour[4] (RWD) based on the properties of the geodesics in Kerr background studied by Carter.[5] In particular Ruffini and Wilson used a simplified model to describe possible charge separation processes involving the black hole and its magnetosphere. Such a studies have been recently revisited[6] by using coordinates regular on the horizon found decades later by Doran[8] and are here reviewed. Doran's work in particular generalizes classical Painlevé-Gullstrand coordinates for spherical black holes. These coordinates are horizon penetrating and are naturally associated to regular infalling physical observers. They are usually adopted in Numerical Relativity in union with the *excision* technique which allows to extend the computational domain beyond the black hole event horizon, avoiding to impose problematic boundary conditions for the partial differential equations there. Moreover, Painlevé-Gullstrand type coordinates naturally occur in Analogue Gravity in relation with acoustic black holes.[7] Summarizing, what we discuss here is the revisitation of RWD works by using useful coordinates found almost 25 years later. We will show in particular that these coordinates are the most natural ones to describe plasma physics by comoving with the fluid itself.

2. The RWD solution

RWD started from the Kerr metric in Boyer-Lindquist coordinates

$$\begin{aligned} ds^2 = &-\left(1-\frac{2Mr}{\Sigma}\right)dt^2 \; - \frac{4aMr\sin^2\theta}{\Sigma}dtd\phi + \frac{\Sigma}{\Delta}dr^2 + \Sigma d\theta^2 \\ &+ \left[r^2+a^2+\frac{2Mra^2\sin^2\theta}{\Sigma}\right]\sin^2\theta d\phi^2, \end{aligned} \tag{1}$$

with $\Sigma = r^2 + a^2\cos^2\theta$, $\Delta = r^2 - 2Mr + a^2$, a being the specific angular momentum and M the black hole mass, while the outer event horizon is located at $r_+ = M + \sqrt{M^2 - a^2}$ and Boyer-Lindquist coordinates are singular there. In the test field

approximation (no metric back reaction), neglecting pressure gradients in MHD equations as well as magnetic force terms and imposing a perfect plasma condition $F_{\mu\nu}U^{\nu} = 0$, the fluid must follow the geodesics on Kerr background studied by Carter:

$$U_{\mu;\nu}U^{\nu} = 0\,. \tag{2}$$

In particular in their works, RWD consider $U_{\phi} = 0$ and $U_t = -1$ at infinity and this implies that U_{θ} is a constant of motion. The four-velocity geodesic vector is then:

$$\begin{aligned} U^t &= \frac{\Sigma(r^2+a^2)+2Mra^2\sin^2\theta}{\Sigma\Delta}, \quad U^r = -\frac{[-\Delta U_{\theta}{}^2+2Mr(r^2+a^2)]^{\frac{1}{2}}}{\Sigma} \\ U^{\theta} &= \frac{U_{\theta}}{\Sigma}, \quad U^{\phi} = \frac{2MRa}{\Sigma\Delta}\,. \end{aligned} \tag{3}$$

Finally, requiring overall neutral stationary and axisymmetric configuration, the vector potential A_{μ} is characterized by the A_{ϕ} component only. RWD solving the simplified MHD equations found in such a a first approximation the analytical solution:

$$A_{\phi} = A(\theta_{\infty}) = A_{\phi}(\theta, r) \tag{4}$$

where

$$\theta_{\infty} = \theta - U_{\theta}\,\xi(r) \tag{5}$$

and

$$\xi(r) = \int_r^{\infty} \frac{dr'}{\sqrt{-(r'^2-2Mr'+a^2)U_{\theta}^2+2Mr'(r'^2+a^2)}}\,. \tag{6}$$

From these relations, one case easily reconstruct the entire Maxwell tensor $F_{\mu\nu}$ and obtain the associated four-current J^{μ}. Great simplification occur by choosing $U_{\theta} = 0$ because in such a case it results $A_{\phi} = F(\theta)$ with F a being an arbitrary function. Specifically, in RWD works it has been assumed the simple form $F(\theta) = A_0|\cos\theta|$ where quantity A_0 is a constant. Concerning Maxwell invariants, for this solution they have these properties:[6]

$$\frac{1}{2}F_{\mu\nu}F^{\mu\nu} = (\mathbf{B}^2 - \mathbf{E}^2) \geq 0\,, \qquad \frac{1}{4}F_{\mu\nu}{}^{*}F^{\mu\nu} = \mathbf{E}\cdot\mathbf{B} = 0\,, \tag{7}$$

so there must exist a frame in which the observer associated to the geodesics four-velocity measures a magnetic field only, while the electric one vanishes in consequence of the perfect plasma condition. In standard perfect MHD physics, both charge density and electric field disappear in a locally comoving (and corotating) plasma frame. This frame is in general not easy to be found analytically,[9] but for the RWD solution however, by using Doran coordinates, we will successfully obtain it, as now discussed.

3. Doran coordinates analysis

The transformation from Boyer-Lindquist (BL) coordinates (t, r, θ, ϕ) to Doran Painlevé-Gullstrand-like (DPG) ones (T, R, Θ, Φ) is:

$$T = t - \int^r f(r)dr \ , \qquad R = r \ , \qquad \Theta = \theta \ ,$$
$$\Phi = \phi - \int^r \frac{a}{r^2 + a^2} f(r)dr \, , \quad f(r) = -\frac{\sqrt{(2Mr)(r^2 + a^2)}}{\Delta} \, , \tag{8}$$

so Kerr solution becomes

$$\begin{aligned} ds^2 = &- \left(1 - \frac{2Mr}{\Sigma}\right) dT^2 + 2\sqrt{\frac{2Mr}{r^2 + a^2}} dT dr - \frac{2a(2Mr)}{\Sigma} \sin^2\theta dT d\Phi \\ &+ \sin^2\theta \left[r^2 + a^2 + \frac{a^2(2Mr)}{\Sigma} \sin^2\theta\right] d\Phi^2 - 2a \sin^2\theta \sqrt{\frac{2Mr}{r^2 + a^2}} dr d\Phi \\ &+ \frac{\Sigma}{r^2 + a^2} dr^2 + \Sigma d\theta^2 \, . \end{aligned} \tag{9}$$

In the following, due to some of the coordinates' coincidence, we shall denote R with r and Θ with θ again. Using the coordinates transformation above, the RWD vector potential becomes

$$A_\mu = \left[0, -a\sqrt{\frac{2Mr}{r^2 + a^2}} \frac{F(\theta)}{\Delta}, 0, F(\theta)\right] \, . \tag{10}$$

with $F(\Theta) \equiv F(\theta) = A_0 |\cos\theta|$. Concerning the geodesics, we get

$$U_\mu = \left[-1, -\frac{\sqrt{-\Delta U_\theta^2 + 2Mr(r^2 + a^2)}}{\Delta} + \frac{\sqrt{2Mr(r^2 + a^2)}}{\Delta}, U_\theta, 0\right] \, , \tag{11}$$

which in the case of interest $U_\theta = 0$ case become

$$U_\mu = [-1, 0, 0, 0] \tag{12}$$

$$U^\mu = \left[1, -\frac{\sqrt{(2Mr)(r^2 + a^2)}}{\Sigma}, 0, 0\right] \, . \tag{13}$$

The four-velocity U^μ above represents the $T = const$ normal geodesic observer. Quantity T is the local proper time of observers in free fall along trajectories characterized by constant θ and Φ. The DPG electromagnetic tensor of the $U_\theta = 0$ RWD solution field has the only non vanishing component, undefined on the equatorial plane, given by:

$$F_{\Phi\theta} = A_{\Phi,\theta} \equiv \frac{dF(\theta)}{d\theta} = -A_0 \sin\theta \frac{|\cos\theta|}{\cos\theta} \, . \tag{14}$$

Defining the orthonormal locally Lorentzian frame $e^{(a)}{}_{\mu}$ associated to the DPG normal observer discussed above:

$$e^{(0)}{}_{\mu} = [-1,0,0,0]\ , \quad e^{(1)}{}_{\mu} = \left[\sqrt{\frac{2Mr}{\Sigma}}, \sqrt{\frac{\Sigma}{r^2+a^2}}, 0, -\sqrt{\frac{2Mr}{\Sigma}} a \sin^2\theta\right]$$
$$e^{(2)}{}_{\mu} = \left[0,0,\sqrt{\Sigma},0\right]\ , \quad e^{(3)}{}_{\mu} = \left[0,0,0,\sqrt{r^2+a^2}\sin\theta\right]\ , \tag{15}$$

one can easily show that the Maxwell tensor in the frame $F_{(a)(b)} = e_{(a)}{}^{\mu} e_{(b)}{}^{\mu} F_{\mu\nu}$ has the only non vanishing component (giving a magnetic field only):

$$F_{(3)(4)} = -F_{(4)(3)} \equiv B_{(1)} \equiv B_{\hat{r}} = -\frac{A_0}{\sqrt{(r^2+a^2)\Sigma}} \frac{|\cos\theta|}{\cos\theta}\ . \tag{16}$$

Moreover, by computing the charge density $J^{\mu}U_{\mu} \equiv J_{(0)} = \rho$ measured by the same observer one easily finds that it vanishes everywhere. As anticipated, we have found the natural frame to describe this perfect MHD problem which is the comoving fluid one. Moreover by using the fact that we have horizon penetrating coordinates, we can follow the fields inside the black hole. In particular it is possible to plot the four current lines for this RWD solution given by the numerical integration of the differential equations set $dx^{\alpha}/d\lambda = J^{\alpha}$ with λ parametrising the curves.[6] To note that while in BL coordinates the current lines whirl infinite times around the event horizon without entering inside, in DPG coordinates (we remind that these are naturally associated to comoving observers) the current lines are not whirled and can be continued regularly in the interior of the black hole. Finally, the use of DPG coordinates allows one to obtain elegant expressions for studying the energetics of the RWD solution. Always in the $U_{\theta} = 0$ case, one can easily compute an important quasi-local quantity i.e. the electromagnetic energy stored outside the event horizon through a $T = \ const$ cut in space-time as measured by the DPG normal observer. This quantity is given by

$$E_{\sigma}(U) = \int_{\sigma} T^{(\mathrm{em})}_{\mu\nu} U^{\mu} d\sigma^{\nu}\ . \tag{17}$$

Here σ is a bounded hypersurface which contains a portion of spacetime while U^{μ} represents the normal observer's four-velocity. In the relation above,

$$\mathcal{E} = T^{(\mathrm{em})}_{\mu\nu} U^{\mu} U^{\nu} \equiv \frac{A_0^2}{8\pi(r^2+a^2)\Sigma} \equiv 8\pi\mathcal{F} \geq 0\,, \tag{18}$$

is the local electromagnetic energy density obtained from the $T^{(\mathrm{em})}_{\mu\nu}$ electromagnetic energy-momentum tensor in DPG coordinates, here proportional to the first Maxwell invariant. Assuming the outer boundary of σ being the 2-surface $r = \mathcal{R} = const$, we get the simple relation:

$$E_{(r_+,\mathcal{R})}(U) = \frac{A_0^2}{2a}\left[\arctan\frac{\mathcal{R}}{a} - \arctan\frac{r_+}{a}\right]\ . \tag{19}$$

4. Conclusions

We have shown that the adoption of recent modern tools is extremely useful for revisiting classical studies of the $60s$ and $70s$. We have adopted in particular the most natural object to describe the physics of Ruffini-Wilson and Damour works, represented by Kerr black hole Doran horizon penetrating coordinates with their naturally associated normal observer. We have in particular described plasma physics from the fluid comoving frame and found great simplification for the electrodynamical quantities. The limitation of the analysis here presented however rely on the $U_\theta = 0$ choice which does not allow to obtain possible charge separation processes (charge density is vanishing everywhere in this case in fact), obtained by Ruffini and Wilson by imposing $U_\theta(\theta) = -U_\theta(\pi - \theta) = const$ instead. In order to address such a more complicated problem one should require a further generalization of Doran's work for the $U_\theta \neq 0$ case first. This study is not present in the literature and deserves future studies in order to revisit also the problem of Ruffini and Wilson charge separation.

Acknowledgments

C.C., S.F. and A.L. acknowledge ICRANet and INdAM-GNFM for support.

References

1. T. W. Baumgarte and S. L. Shapiro, *Numerical Relativity: Solving Einstein's Equations on the Computer* (Cambridge University Press 2010)
2. R. P. Kerr, *Phys. Rev. Lett.* **11**, 237 (1963)
3. R. Ruffini and J.R. Wilson, *Phys. Rev. D* **12**, 2959 (1975).
4. T. Damour, *Annals of the New York Academy of Sciences*, **262**, 113 (1975).
5. B. Carter, *Phys. Rev.* **174**, 1559 (1968).
6. C. Cherubini, S. Filippi, A. Loppini, R. Moradi, R. Ruffini, Y. Wang and S-S. Xue, *Phys. Rev. D.*, **97**, 064038 (2018)
7. C. Barcelo, S. Liberati and M. Visser, *Living Rev. Relativ.*, **14**, 3 (2011).
8. C. Doran, *Phys. Rev. D* **61**, 067503 (2000)
9. H. C. Spruit, *arXiv: 1301.5572v3* page 49 (2016).

Tadpole contribution to magnetic photon-graviton conversion

N. Ahmadiniaz

Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, 01328 Dresden, Germany
E-mail: n.ahmadiniaz@hzdr.de

F. Bastianelli

Dipartimento di Fisica ed Astronomia, Università di Bologna, Via Irnerio 46, I-40126 Bologna, Italy and INFN, Sezione di Bologna, Via Irnerio 46, I-40126 Bologna, Italy
E-mail: bastianelli@bo.infn.it

F. Karbstein

Helmholtz-Institut Jena, Fröbelstieg 3, 07743 Jena, Germany and Theoretisch-Physikalisches Institut, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany
E-mail: felix.karbstein@uni-jena.de

C. Schubert

Instituto de Física y Matemáticas Universidad Michoacana de San Nicolás de Hidalgo Edificio C-3, Apdo. Postal 2-82 C.P. 58040, Morelia, Michoacán, México
E-mail: christianschubert137@gmail.com

Photon-graviton conversion in a magnetic field is a process that is usually studied at tree level, but the one-loop corrections due to scalars and spinors have also been calculated. Differently from the tree-level process, at one-loop one finds the amplitude to depend on the photon polarization, leading to dichroism. However, previous calculations overlooked a tadpole contribution of the type that was considered to be vanishing in QED for decades but erroneously so, as shown by H. Gies and one of the authors in 2016. Here we compute this missing diagram in closed form, and show that it does not contribute to dichroism.

Keywords: Photon-graviton; tadpole; Einstein-Maxwell

1. Introduction: Photon-graviton conversion

Einstein-Maxwell theory contains a tree-level vertex for photon-graviton conversion in a constant electromagnetic field:

$$\frac{1}{2}\kappa h_{\mu\nu}\Big(F^{\mu\alpha}f^{\nu}{}_{\alpha}+f^{\mu}{}_{\alpha}\,F^{\nu\alpha}\Big)-\frac{1}{4}\kappa h^{\mu}_{\mu}F^{\alpha\beta}f_{\alpha\beta}. \tag{1}$$

Here $h_{\mu\nu}$ denotes the graviton, $f_{\mu\nu}$ the photon, $F^{\mu\nu}$ the external field, and κ the gravitational coupling constant.

This interaction leads to photon-graviton oscillations similar to the better-known neutrino or photon-axion oscillations[1–11] (see[12] for a recent application to gravitational waves).

In momentum space, this vertex becomes

$$\Gamma^{(\mathrm{tree})}(k,\varepsilon;F) = \epsilon_{\mu\nu}\varepsilon_\alpha \Pi^{\mu\nu,\alpha}_{(\mathrm{tree})}(k;F), \quad \Pi^{\mu\nu,\alpha}_{(\mathrm{tree})}(k;F) = -\frac{i\kappa}{2} C^{\mu\nu,\alpha} \tag{2}$$

with

$$C^{\mu\nu,\alpha} = F^{\mu\alpha}k^\nu + F^{\nu\alpha}k^\mu - \big(F\cdot k\big)^\mu \delta^{\nu\alpha} - \big(F\cdot k\big)^\nu \delta^{\mu\alpha} + \big(F\cdot k\big)^\alpha \delta^{\mu\nu}\,. \tag{3}$$

Since in a constant field the four-momentum is preserved, k^μ here is the four-momentum of the photon as well as of the graviton.

For the photon polarizations, as is customary we will use the Lorentz frame where $\mathbf{E}$ and $\mathbf{B}$ are collinear, and choose the polarization basis $\varepsilon^\alpha_\perp, \varepsilon^\alpha_\parallel$, where $\varepsilon^\alpha_\parallel$ lies in the plane spanned by k and $\mathbf{B}$ or $\mathbf{E}$, and $\varepsilon^\alpha_\perp$ is perpendicular to it. For convenience we construct also the graviton polarization tensor using the same vectors:

$$\varepsilon^{\oplus\mu\nu} = \varepsilon^{\perp\mu}\varepsilon^{\perp\nu} - \varepsilon^{\parallel\mu}\varepsilon^{\parallel\nu}, \quad \varepsilon^{\otimes\,\mu\nu} = \varepsilon^{\perp\mu}\varepsilon^{\parallel\nu} + \varepsilon^{\parallel\mu}\varepsilon^{\perp\nu}\,. \tag{4}$$

From the CP-invariance of Einstein-Maxwell theory one can then derive the following selection rules:

- For a purely magnetic field $\varepsilon^\oplus$ couples only to $\varepsilon^\perp$ and $\varepsilon^\otimes$ only to $\varepsilon^\parallel$.
- For a purely electric field $\varepsilon^\oplus$ couples only to $\varepsilon^\parallel$ and $\varepsilon^\otimes$ only to $\varepsilon^\perp$.

2. One-loop photon-graviton vacuum polarization

In[13,14] the worldline formalism was used along the lines of[15–17] to study the one-loop corrections to this amplitude due to a scalar or spinor loop, see Fig. 1.

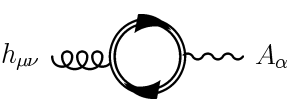


Fig. 1. One-loop correction to the photon-graviton amplitude in a constant field.

Here we employ the usual double-line notation for the full propagator in the external electromagnetic field, Fig. 2.

Fig. 2. Full scalar or spinor propagator in a constant field.

These calculations lead to the same type of two-parameter integrals as for the well-studied case of the one-loop photon vacuum polarisation in a constant field.[18–29]

Both the scalar and the spinor-loop amplitudes are UV divergent, but multiplicatively renormalizable. For example, the scalar-loop contribution in dimensional regularization displays the pole

$$\Pi^{\mu\nu,\alpha}_{\rm scal,div}(k) = \frac{ie^2\kappa}{3(4\pi)^2}\frac{1}{D-4}C^{\mu\nu,\alpha}$$

where $C^{\mu\nu,\alpha}$ is the tree level vertex (3).

For studying the relative importance of the one-loop amplitudes it is useful to normalize them by the tree-level amplitude (the "bar" on Π denotes renormalization)

$$\hat{\Pi}^{Aa}_{\rm scal,spin}(\hat{\omega},\hat{B},\hat{E}) \equiv \frac{\bar{\Pi}^{Aa}_{\rm scal,spin}(\hat{\omega},\hat{B},\hat{E})}{-\frac{i}{2}\kappa C^{Aa}} \tag{5}$$

where $A = \oplus, \otimes$ and $a = \perp, \parallel$, $\hat{\omega} = \frac{\omega}{m}$, $\hat{B} = \frac{eB}{m^2}$, $\hat{E} = \frac{eE}{m^2}$.

It then becomes obvious that the one-loop amplitudes become quantitatively relevant only for field strengths close to the critical ones $\hat{B}, \hat{E} \approx 1$. Macroscopic fields of this magnitude are known to exist only for the magnetic case, so that we will set $\hat{E} = 0$ in the following. In[14] it was shown that for this purely magnetic case the parameter integrals allow for a straightforward numerical evaluation as long as the photon energy $\hat{\omega}$ is below the pair-creation threshold $\hat{\omega}_{\rm cr}$,

$$\begin{aligned}\hat{\omega}^{\oplus\perp}_{\rm cr,scal} = \hat{\omega}^{\otimes\parallel}_{\rm cr,scal} &= 2\sqrt{1+\hat{B}}\,,\\ \hat{\omega}^{\oplus\perp}_{\rm cr,spin} &= 1+\sqrt{1+2\hat{B}}\,,\\ \hat{\omega}^{\otimes\parallel}_{\rm cr,spin} &= 2\,.\end{aligned} \tag{6}$$

There also a number of special cases were studied that allow for more explicit representations:

(1) The case of small B and arbitrary ω leads to single - parameter integrals over Airy functions.
(2) In the large B limit one finds a logarithmic growth in the field strength,

$$\hat{\Pi}^{Aa}_{\rm scal}(\hat{\omega},\hat{B}) \overset{\hat{B}\to\infty}{\sim} -\frac{\alpha}{12\pi}\ln(\hat{B})\,, \tag{7}$$

$$\hat{\Pi}^{Aa}_{\rm spin}(\hat{\omega},\hat{B}) \overset{\hat{B}\to\infty}{\sim} -\frac{\alpha}{3\pi}\ln(\hat{B})\,. \tag{8}$$

(3) In the limit of vanishing photon energy the amplitudes can be related to the corresponding one-loop effective Lagrangians:

$$\hat{\Pi}^{\oplus\perp}_{\rm scal,spin}(\hat{\omega}=0,\hat{B}) = -\frac{2\pi\alpha}{m^4}\Big(\frac{1}{\hat{B}}\frac{\partial}{\partial\hat{B}} + \frac{\partial^2}{\partial\hat{B}^2}\Big)\,\mathcal{L}^{\rm EH}_{\rm scal,spin}(\hat{B})\,, \tag{9}$$

$$\hat{\Pi}^{\otimes\parallel}_{\rm scal,spin}(\hat{\omega}=0,\hat{B}) = -\frac{4\pi\alpha}{m^4}\frac{1}{\hat{B}}\frac{\partial}{\partial\hat{B}}\,\mathcal{L}^{\rm EH}_{\rm scal,spin}(\hat{B})\,. \tag{10}$$

Here $\mathcal{L}^{\rm EH}_{\rm scal,spin}(\hat{B})$ denotes the one-loop effective Lagrangian in a constant magnetic field, obtained for the spinor QED case by Heisenberg and Euler[30] and for scalar QED by Weisskopf[31]:

$$\mathcal{L}^{\rm EH}_{\rm scal}(\hat{B}) = -\frac{m^4}{16\pi^2}\int_0^\infty \frac{d\hat{s}}{\hat{s}^3}\,\mathrm{e}^{-i\hat{s}}\left[\frac{\hat{B}\hat{s}}{\sin(\hat{B}\hat{s})} - \frac{(\hat{B}\hat{s})^2}{6} - 1\right], \tag{11}$$

$$\mathcal{L}^{\rm EH}_{\rm spin}(\hat{B}) = \frac{m^4}{8\pi^2}\int_0^\infty \frac{d\hat{s}}{\hat{s}^3}\,\mathrm{e}^{-i\hat{s}}\left[\frac{\hat{B}\hat{s}}{\tan(\hat{B}\hat{s})} + \frac{(\hat{B}\hat{s})^2}{3} - 1\right]. \tag{12}$$

3. Dichroism

For realistic parameters, the one-loop corrections turn out to be small compared to the tree-level amplitudes. However, there is also a qualitative difference to the tree-level amplitude. As has been emphasized in,[9] the tree level photon-graviton conversion does, contrary to the better known photon-axion case, not lead to a dichroism effect for photon beams. This is because both photon polarization components have equal conversion rates. This symmetry gets broken by the loop corrections. Although this effect is, of course, tiny and hardly measurable in the near future, an exhaustive analysis by Ahlers et al.[32] has shown that it is still the leading contribution to magnetic dichroism in the standard model (including standard gravity)!

4. Quantum electrodynamics in external fields

In vacuum QED, there is, of course, no one-photon amplitude because of Furry's theorem. In the presence of an external field, however, one-photon tadpole diagrams such as the one in Fig. 3 will in general be nonzero.

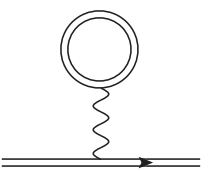


Fig. 3. Full scalar or spinor propagator in a constant field.

If the external field is constant, then the one-photon amplitude Fig. 4 can still be shown to vanish.

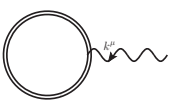


Fig. 4. One photon amplitude in a constant field.

The argument goes as follows:

(1) A constant field emits only photons with zero energy-momentum, thus there is a factor of $\delta(k)$.
(2) Because of gauge invariance, this diagram in a momentum expansion starts with the term linear in momentum.
(3) $\delta(k)k^\mu = 0$.

Since the tadpole vanishes, it has been assumed for decades that also any diagram containing it can be discarded. For example, in the book "Quantum Electrodynamics with Unstable Vacuum" by E.S. Fradkin, D.M. Gitman and S.M. Shvartsman[34] it is stated (on page 225) even more generally that, "*Thus, in the constant and homogeneous external field combined with that of a plane-wave, all the diagrams containing the causal current $\mathcal{J}^\mu(x)$ (i.e., those containing tadpoles having a causal propagator $S^c(x,y)$), are equal to zero.*"

Sometimes also additional arguments have been given; in the book "Effective Lagrangians in Quantum Electrodynamics" by Dittrich and Reuter[33] it is argued that the "handcuff" diagram of Fig. 5 vanishes because of Lorenz invariance.

Fig. 5. Handcuff diagram in a constant field.

However, in 2016 H. Gies and one of the authors[35] noted that such diagrams can give finite values because of the infrared divergence of the connecting photon propagator. In dimensional regularization, the key integral is

$$\int d^D k\, \delta^D(k) \frac{k^\mu k^\nu}{k^2} = \frac{\eta^{\mu\nu}}{D}\,. \tag{13}$$

Applying this integral to the handcuff diagram one finds a non-vanishing result, which can be expressed in the following simple way in terms of the one-loop Euler-Heisenberg Lagrangian[36]

$$\mathcal{L}^{\rm 1PR}_{\rm spin} = \frac{1}{2}\frac{\partial \mathcal{L}^{(EH)}_{\rm spin}}{\partial F^{\mu\nu}}\frac{\partial \mathcal{L}^{(EH)}_{\rm spin}}{\partial F_{\mu\nu}}$$

(the superscript "1PR" stands for "one-particle reducible"). This adds on to the standard diagram for the two-loop Euler-Heisenberg Lagrangian, studied by V.I. Ritus[39] half a century ago:

$$\mathcal{L}^{(EH)2-{\rm loop}}_{\rm spin} = \qquad + \tag{14}$$

Along the same lines, it was found in[37,38] that the one-loop tadpole contribution Fig. 3 to the scalar or spinor propagator in a constant field is also non-vanishing,

and given by

$$S^{\rm 1PR}(p) = \frac{\partial S(p)}{\partial F_{\mu\nu}} \frac{\partial \mathcal{L}^{(EH)}}{\partial F^{\mu\nu}}$$

where $S(p)$ denotes the tree-level propagator in the field.

5. Tadpole contribution to the photon-graviton amplitude

Returning to the photon-graviton amplitude in a constant field, the point of the present talk is that this amplitude, too, has a previously overlooked tadpole contribution, shown in Fig. 6.

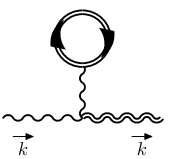


Fig. 6. Tadpole contribution to photon-graviton conversion.

Using the integral (13) it is easy to show that its contribution to the one-loop amplitudes with a scalar or spinor loop can be written as

$$\Gamma^{(\rm tadpole)}_{\rm scal}(k^\alpha, \varepsilon_\beta; \epsilon_{\mu\nu}; F_{\kappa\lambda}) = -i\frac{\alpha}{8\pi}\kappa\Big(\varepsilon\cdot F\cdot\epsilon\cdot k + \varepsilon\cdot\epsilon\cdot F\cdot k\Big) \times \int_0^\infty \frac{dz}{z}\, {\rm e}^{-\frac{m^2}{eB}z}\, \frac{\coth({\rm z}) - 1/{\rm z}}{\sinh(z)}\,, \tag{15}$$

$$\Gamma^{(\rm tadpole)}_{\rm spin}(k^\alpha, \varepsilon_\beta; \epsilon_{\mu\nu}; F_{\kappa\lambda}) = i\frac{\alpha}{4\pi}\kappa\Big(\varepsilon\cdot F\cdot\epsilon\cdot k + \varepsilon\cdot\epsilon\cdot F\cdot k\Big) \times \int_0^\infty \frac{dz}{z}\, {\rm e}^{-\frac{m^2}{eB}z}\, \frac{\coth({\rm z}) - \tanh({\rm z}) - 1/{\rm z}}{\tanh({\rm z})}\,. \tag{16}$$

Expanding out the tadpole in powers of the external field we see that the leading term, which is linear in the field (Fig. 7), is removed by the photon wave function renormalization.

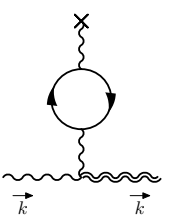


Fig. 7. Expanding the tadpole to lowest order in the field.

This gives the renormalized amplitudes

$$\Gamma^{(\text{tadpole})}_{\text{scal,ren}}(k^\alpha, \varepsilon_\beta; \epsilon_{\mu\nu}; F_{\kappa\lambda}) = -i\frac{\alpha}{8\pi}\kappa\Big(\varepsilon\cdot F\cdot\epsilon\cdot k + \varepsilon\cdot\epsilon\cdot F\cdot k\Big) \times \int_0^\infty \frac{dz}{z}\,\mathrm{e}^{-\frac{m^2}{eB}z}\left[\frac{\coth(\mathrm{z}) - 1/\mathrm{z}}{\sinh(z)} - \frac{1}{3}\right],$$

$$\Gamma^{(\text{tadpole})}_{\text{spin,ren}}(k^\alpha, \varepsilon_\beta; \epsilon_{\mu\nu}; F_{\kappa\lambda}) = i\frac{\alpha}{4\pi}\kappa\Big(\varepsilon\cdot F\cdot\epsilon\cdot k + \varepsilon\cdot\epsilon\cdot F\cdot k\Big) \times \int_0^\infty \frac{dz}{z}\,\mathrm{e}^{-\frac{m^2}{eB}z}\left[\frac{\coth(\mathrm{z}) - \tanh(\mathrm{z}) - 1/\mathrm{z}}{\tanh(\mathrm{z})} + \frac{2}{3}\right].$$

6. Comparison with the main diagram

These amplitudes are of a structure similar to what we got from the main diagram Fig. 1 in the limit of low photon energy ω, eqs. (9), (10).

However, they do not contribute to dichroism since the polarizations are still bound up in the tree-level vertex ($\varepsilon\cdot F\cdot\epsilon\cdot k + \varepsilon\cdot\epsilon\cdot F\cdot k$). Thus the above-mentioned analysis of Ahlers et al.[32] remains unaffected by the presence of the tadpole diagram.

7. Summary and outlook

- We have presented the first example of a non-vanishing diagram in Einstein-Maxwell theory involving a tadpole in a constant field.
- This diagram does not contribute to magnetic dichroism.
- A more quantitative analysis is in progress.
- In the ultra strong-field limit, the tadpoles have been shown even to dominate the (multi-loop) effective action in QED.[40] It would be interesting to extend this analysis to the Einstein-Maxwell case.

References

1. M.E. Gertsenshtein, Sov. Phys. JETP **14** (1962) 84.
2. Ya. B. Zel'dovich and I.D. Novikov, *The structure and evolution of the universe*, Rel. Astrophys. Vol. 2 (1983), Chicago University Press.
3. G.G. Raffelt and L. Stodolsky, Phys. Rev. **D37** (1988) 1237.
4. P. Sikivie, Phys. Rev. Lett. **51** (1983) 1415; erratum-ibid. **52** (1984) 695.
5. D.E. Morris, Phys. Rev. **D34** (1986) 843.
6. H.N. Long, D.V. Soa, and T.A. Tran, Mod. Phys. Lett. **A9** (1994) 3619, astro-ph/9410003.
7. J.C.R. Magueijo, Phys. Rev. **D49** (1994) 671.
8. P. Chen, Phys. Rev. Lett. **74** (1995) 634; erratum-ibid. **74** (1995) 3091.
9. A.N. Cillis and D. Harari, Phys. Rev. **D54** (1996) 4757, astro-ph/9609200.
10. N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. **B429** (1998) 263, hep-ph/9803315; Phys. Rev. **D59** (1999) 086004, hep-ph/9807344.
11. C. Deffayet and J.-P. Uzan, Phys. Rev. **D62** (2000) 063507, hep-ph/0002129.

12. A. Ejlli, D. Ejlli, A.M. Cruise, G. Pisano and H. Grote, Eur. Phys. J. C **79** (2019) 1032, arXiv:1908.00232[gr-qc].
13. F. Bastianelli and C. Schubert, JHEP **0502**, 069 (2005), gr-qc/0412095.
14. F. Bastianelli, U. Nucamendi, C. Schubert, and V.M. Villanueva, JHEP **0711** (2007) 099, arXiv:0710.5572.
15. M.J. Strassler, Nucl. Phys. **B 385** (1992) 145.
16. M. Reuter, M.G. Schmidt, and C. Schubert, Ann. Phys. (N.Y.) 259 (1997) 313, hep-th/9610191.
17. C. Schubert, Nucl. Phys. **B 585** (2000) 407, hep-ph/0001288.
18. J.S. Toll, PhD thesis, Princeton Univ., 1952 (unpublished).
19. A. Minguzzi, Nuovo Cim. **6** (1956) 476.
20. R. Baier and P. Breitenlohner, Acta Phys. Austr. **25** (1967) 212; Nuovo Cim. **47** (1967) 261.
21. Z. Bialynicka-Birula and I. Bialynicka-Birula, Phys. Rev. **D 2** (1970) 2341.
22. S.L. Adler, Ann. Phys. (N.Y.) **67** (1971) 599.
23. I.A. Batalin and A.E. Shabad, Zh. Eksp. Teor. Fiz. **60** (1971) 894 [JETP **33** (1971) 483].
24. V.I. Ritus, Ann. Phys. (N.Y.) **69** (1972) 555.
25. W.-Y. Tsai and T. Erber, Phys. Rev. **D 10** (1974) 492; Phys. Rev. **D 12** (1975) 1132.
26. V.N. Baier, V.M. Katkov, and V.M. Strakhovenko, Zh. Eksp. Teor. Fiz. **68** (1975) 405 [JETP **41** (1975) 198].
27. D.B. Melrose and R.J. Stoneham, Nuov. Cim. **32** (1976) 435.
28. W. Dittrich and H. Gies, *Probing the Quantum Vacuum*, Springer 2000.
29. K. Kohri and S. Yamada, Phys. Rev. **D 65**:043006 (2002), astro-ph/0102225.
30. W. Heisenberg and H. Euler, "Folgerungen aus der Diracschen Theorie des Positrons", Z. Phys. **98** (1936) 714.
31. V. Weisskopf, "Über die Elektrodynamik des Vakuums auf Grund der Quantentheorie des Elektrons", Kong. Dans. Vid. Selsk. Math-fys. Medd. XIV No. 6 (1936), reprinted in *Quantum Electrodynamics*, J. Schwinger (Ed.) (Dover, New York, 1958).
32. M. Ahlers, J. Jaeckel and A. Ringwald, Phys. Rev. D **79** 075017 (2009), arXiv:0812.3150 [hep-ph].
33. W. Dittrich and M. Reuter, *Effective Lagrangians in Quantum Electrodynamics*, Springer 1985.
34. E.S. Fradkin, D.M. Gitman, S.M. Shvartsman, *Quantum Electrodynamics with Unstable Vacuum*, Springer 1991.
35. H. Gies and F. Karbstein, *JHEP* **1703** (2017) 108, arXiv:1612.07251 [hep-th].
36. F. Karbstein, *JHEP* **10** (2017) 075, arXiv:1709.03819 [hep-th].
37. J.P. Edwards and C. Schubert, Nucl. Phys. B **923** (2017) 339, arXiv:1704.00482 [hep-th].
38. N. Ahmadiniaz, F. Bastianelli, O. Corradini, J.P. Edwards and C. Schubert, Nucl. Phys. B **924** (2017) 377, arXiv:1704.05040 [hep-th].
39. V.I. Ritus, Zh. Eksp. Teor. Fiz **69** (1975) 1517 [Sov. Phys. JETP **42** (1975) 774].
40. F. Karbstein, Phys. Rev. Lett. **122** (2019) 21, arXiv:1903.06998 [hep-th].

Correspondence of gamma radiation coming from GRBs and magnetars based on the effects of nonlinear vacuum electrodynamics

Tursynbek Yernazarov[1,*], Medeu Abishev[1] and Yerlan Aimuratov[1,2,†]

[1] *Department of Theoretical and Nuclear Physics*
Faculty of Physics and Technology, al-Farabi Kazakh National University
al-Farabi avenue 71, Almaty, 050040, Kazakhstan
E-mail: medeu.abishev@kaznu.kz
yernazarov_tursynbek@live.kaznu.kz

[2] *Fesenkov Astrophysical Institute*
Observatory 23, Kamenskoye Plateau
Almaty, 050020, Kazakhstan
E-mail: aimuratov@aphi.kz

It appears that studying data from the catalogue of Gamma-Ray Bursts (GRBs) can be used to examine the birefringence phenomenon in the magnetosphere of the magnetars. By analysing the data from the McGill Online Magnetar and HEASARC Fermi Burst Catalogues, in this work we studied the angular distances between GRBs and magnetars in projection, we built their distribution map by the end of 2020, and the relative lag time periods of lights coming from GRBs and magnetars. It is confirmed that there are 29 galactic magnetars and their candidates, while the other two are located out of the Milkyway. The maximum separation angle for GRB and Magnetar projectiles was 3.76 degrees (4U0142+61 and GRB110818860), while minimum angular resolution was 0.54 degrees (SGR 1627-41 and GRB090829672). Currently, we discuss the relationship of GRB light intensity by their lag time as it would come after bending in the magnetosphere of the magnetar.

Keywords: Neutron stars; magnetars; light bending; gamma-ray bursts; birefringence.

1. Introduction

Magnetars emit a variety of electromagnetic waves ranging from gamma-ray to radio spectrum. Their bursts last from millisecond to a month.[1] Magnetar's bursts are generated by the powerful magnetic field inside of it. By taking into account the probability of boosting of core circular magnetic field covered by forceful reaction in a protoneutron star coming after a core-collapse supernova the magnetar model was created.[2] Currently there were detected 34 objects that are considered to be magnetars: 10% of them might belong to young neutron star populations.

To the date nearly thousand research articles were published on the topic of magnetars.[1] The first observational reviews on magnetars was carried out by Rea and Esposito, 2011.[3] The leading theoretical review was done by Turolla in 2015.[4] Also, both of observational and theoretical reviews were done by a group of researchers.[5,6]

*Doctoral student
†Postdoctoral associate

Historically magnetars were first presented by covering the name "Soft gamma repeaters" (SGR) and "Anomalous X-ray pulsars" (AXP).[1] The primary exposure of magnetar was done by the interplanetary space probes Venera 11 and 12 by detecting the burst activity.[7] The short-term bursts and outbursts, also continuous emissions were described by the sources of the impulsive magnetic field decays. The essential relation of SGRs to AXP was studied. It is believed that future SGRs burst activity might be detected from AXPs.[1]

It was also expected that magnetars might be prolific glitchers, described as unusual increase of rotation spin of a neutron star and they are observed basically in young neutron stars.[8,9] The first estimation of an SGR spin-down was published and the SGR 1806-20 was confirmed of a magnetar model by its sign and magnitude. Also for SGR 1900+14, as the SGR 1806-20, measurements verified that all SGRs are magnetar models.[10]

Contrary possibility of SGRs to be defined as a magnetar model was reviewed by using Einstein's observatory. They discovered a new unusual celestial bright X-ray source located at the center of the magnetar.[11] Then the new unusual celestial bright X-ray source's pulsation was estimated with the period of 3.5 s. (later it was concluded as a second harmonic of a 7-s fundamental).[11] The SGR bursts from two AXPs were explored and this confirmed Thompson and Duncans' (1996) expectations. Futhermore, the SGR bursts are one of the basic features of AXPs and, it is believed that, there is not any difference between them.[12,13]

The first catalogue of observed magnetar models and their complete characteristics were made by Olausen and Kaspi, in 2014.[14] Height scale of the known magnetar was detected to be about 20-30 pc (parsec). This height scale confirmed the magnetars to be young. Majority of magnetars from the catalog were discovered by their X-ray bursts by the sensitive all-sky monitors such as Burst Alert Telescope (BAT) onboard of Swift Gamma-Ray Burst Mission[15] and Gamma-Ray Burst Monitor (GBM)[16] onboard of Fermi Gamma-Ray Space Telescope. As well as magnetars X-ray burst activities they make X-ray pulsation in the period of range of 2-12 s.[1] Primary exploration of magnetars' glitch was carried out by Kaspi et al., 2000. One of the magnetar candidate had an unusual 6-hr period in X-ray with a different pulse and was explored in the source of the supernova remnant RCW 103, 1E 161348-5055. However, this object was considered not to be the magnetar because of their period of the bursts.[17] Another object was expected to be a magnetar by its bursts coupling with one more large X-ray flux outburst. But these bursts had long periods and this might be defined by a fall-back disc that interrupts the primarily fast-rotating neutron star.[18]

Some magnetar characteristics was detected from low-B radio pulsars from the source of SGR0418+5729.[19] The high-B radio pulsars are expected to be the magnetar by the confirmation of the exciting X-ray and soft gamma-rays are bursted by a magnetic field of neutron stars.[20,21] The relationship of magnetars to high-magnetic-field radio pulsars is expected by the research of radio pulsation from magnetars. But the magnetars' radio characters are unlike the common radio pulsars.[22,23]

Gravitational effect to the light bending is crucial in astrophysics. A weak gravitational bending effect causes gravitational lensing phenomenon, also a strong gravitational effect takes place near a compact objects, such as black holes or neutron stars.The gravitational effect of light bending near pulsars is studied by defining the angle of deflection and the flux coming to observer in a symmetrical metric of Schwarcshild.[24]

Non-linear electrodynamic bending angle based on the theory of Heisenberg-Euler and gravitational bending of a ray passing through the magnetosphere of magnetars are defined.[25] Due to this work the bending angle of light passing through the magnetars is about as a radian or 60 degree. Also the gravitational effect on the bending angle on the surface of a magnetar with the radius of about 3 km is up to 4 degree. Thus the polarisation does not effect on the gravitational bending.These non-linear electrodynamics and the gravitational bending are observable by the method of lensing. These non-linear electrodynamics and the gravitational bending are observable by the method of lensing. Additionally, it is proposed that, if the GRB comes from the nearest sources from the magnetars the scattering og light is negligible. Thus the falling of detected emission intensity will not be as deep as in the case of extra galaxies.

It appears that studying the data from the catalogue of Gamma-Ray Bursts (GRBs) can be used to study the birefringence phenomenon in the magnetosphere of the magnetars. By analysing the data from the McGill Online Magnetar and HEASARC Fermi Burst Catalogues, in this work we studied the angular distances between the nearest GRBs and magnetars in projection, built their distribution map as detected by 2020, and the relative lag time periods of lights coming from GRBs and magnetars. Currently, we discuss the relationship of GRB light intensity by their lag time as it would come after bending by the magnetosphere.

2. Methods

The effects of nonlinear electrodynamics of vacuum on the light propagation in dipole magnetic field causes the light to split into two normal modes and they can be detected in time interval of Δt.[26] It is said that the two modes of light coming to the observer can be observed in different years one by another. Initially, we should observe the detected GRBs and magnetars by their spatial distributions.

All the calculations were done in the Phyton programming platform. We downloaded the fits catalogue as a fits file from the website of HEASARC catalogue search (heasarc.gsfc.nasa.gov) in November 2020. By that time there were about 2915 gamma-ray bursts detected by Fermi GBM satellite telescope. Firstly, we defined short(less than 2 seconds) and long(more than 2 seconds) period bursts by the T90(time duration) table. It is confirmed that the long GRB bursts corresponds to core-collapse supernova stars, while the short bursts are the merges of neutron stars.[27] The number of short GRBs were 478 and the long GRBs were 2437 events.

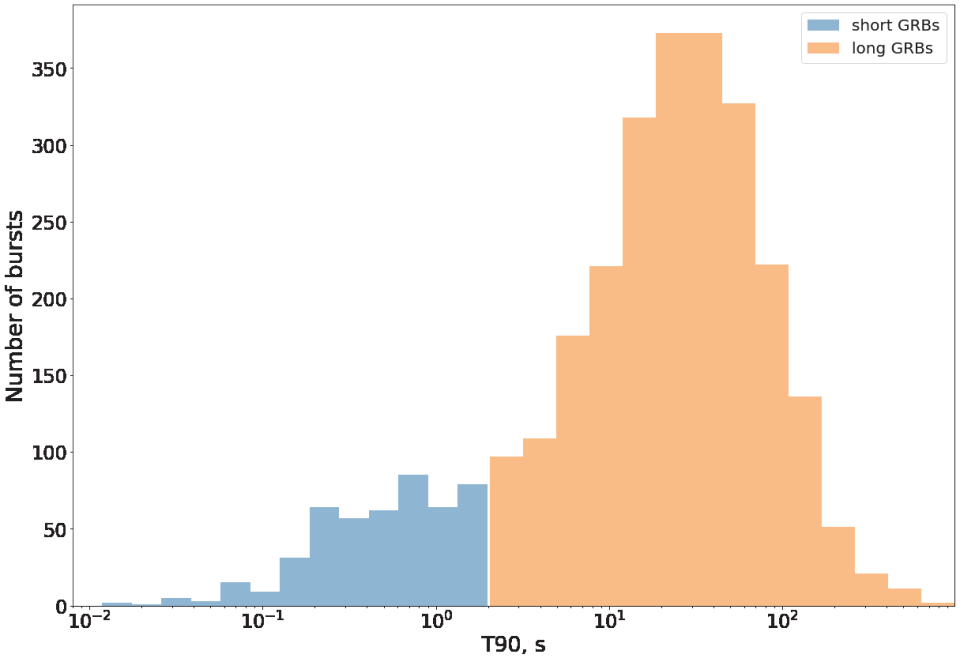


Fig. 1. The histogram of the GRBs by their periods.

In Fig. 1 we build the logarithmic histogram of the number of the bursts by their periods.[28] The minimum period of the short bursts lasted up to 0.008 s, while the maximum period was 1.984 s. Furthermore, the minimum period of the long bursts lasted up to 2.045 s, while the maximum period was 984.084 s. The majority of bursts lay between 10 and 100 s. But other short period bursts have the maximum number corresponds approximately to a second. This results are comparable with the previous works although in size by the order of magnitude.[28] As can be learned from their red shifts the bursts come from different parts of the universe. Let us define their spatial distribution.

Firstly, we took spatial coordinates from catalogue and changed them from the ecliptic into galactic (l,b) values.

We used Aitoff-Hammer projection in order to make the distribution of 2915 bursts on Fig. 2. Marked as red(long) and blue(short) dots these events isotropically distributed, which proves theis extragalactic origin. In previous works the spatial distribution were observed as an object of the definite goals. We use the spatial distribution to study the delay of light within the distribution of the observed magnetars and their candidates.

Nowadays there are 29 galactic and 2 extragalactic magnetars observed; among them there are 26 known magnetars and 5 magnetar candidates.[14] Their distribution

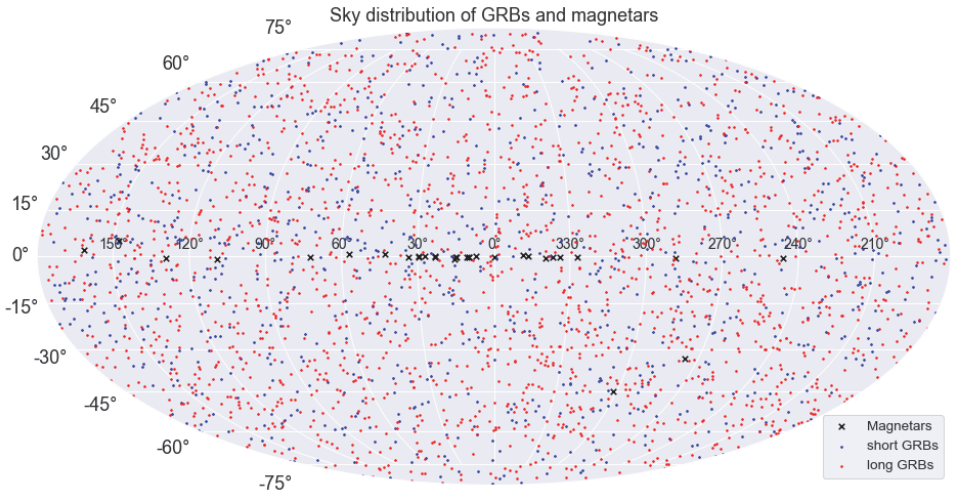


Fig. 2. The sky distribution of GRBs from HEASARC Catalogue as for November, 2020 observed by Fermi's Gamma-ray Burst Monitor (Fermi-GBM).[29]

in the Aitoff-Hammer map are made by using the McGill Online Magnetar catalog and indicated as black crosses in Fig. 2. As for GRBs we changed their coordinates into the galactic coordinates. Indeed, we could present both of the GRB and magnetars in the same map. Most of the magnetars are accumulated along the galactic plane. The other two can be seen out of the Galaxy of Milky-way.

When the light coming from the GRB passes through the magnetosphere of the magnetar it declines to some angle and the light splits into two modes.[30] This phenomena requires first to find out the relation of the spatial coordinates of the GRBs and the magnetars.

In order to illustrate the effect of the light bending we present its schematic view in Fig. 3. The light passing through the magnetars inclines to some θ angle.

If the light coming directly to the observer passes l distance, while the delayed distance is defined as: $\Delta l = (1 - \mathrm{ME} \cdot cos\alpha) \cdot \mathrm{GE}$. Here, ME is a distance between a magnetar and an observer. GE is a distance between a GRB and the Earth.

The next step is a calculation of the distance between the GRBs and the magnetars in projection. Here we take a magnetar, then we define the distance between this magnetar and the GRBs in order to find the closest pair.

Thus it is a projection distance: angular separation between the magnetar and the GRBs. Firstly, we define the difference of right ascensions of the magnetars and the GRBs, we do the same, as well as for the declination. Secondly, we define the distance between a magnetar and GRBs by the simple formula for triangle: $\Delta\ \mathrm{D} = \sqrt{RA^2 + Dec^2}$.

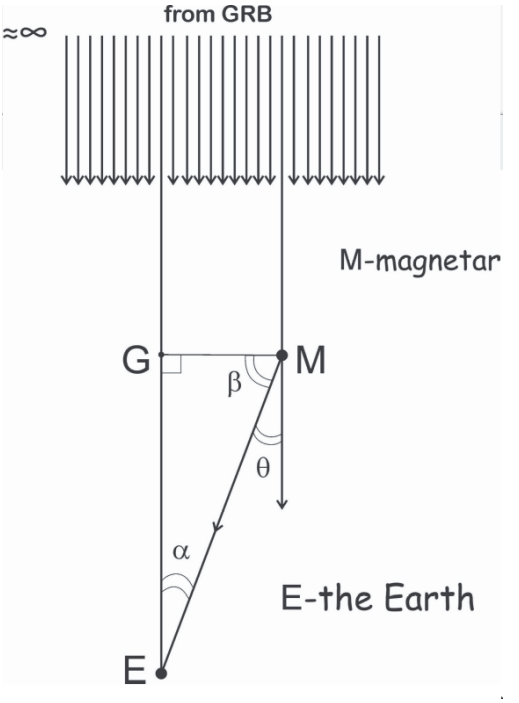


Fig. 3. Light propagation scheme.

We calculated angular distance corresponded to every 31 magnetars. Hence, every magnetars in the map surrounded by 2915 GRBs. The minimum of the angular distance corresponded to the magnetar is defined as the nearest GRB to the magnetar. As in Table 1, α is an angular projection distance for magnetars and GRBs. This results help to define the lag time of light coming to the detector by passing through the magnetar.

Delay of the GRB coming to the Earth can be defined by the distance between the magnetar and the detector. This data can be used from the McGill Online Magnetar Catalog website. We changed the units of parsec to light years and then we calculated the delayed time of light of GRB by dividing the delayed distance by the speed of light. Hopefully, we present the result in Figure 6. Here nan is a unknown data of the distance to a given magnetar. We do not use them in the future. Minimum of lag time corresponds to the magnetar of SGR 0418+5729, while the longest time is for the magnetar of CXOU J010043.1-721134. Futhermore, the errors of the distance to a magnetar should be useful for the calculation of delayed time period of the light.

Table 1. Minimum distance in projection between magnetars and GRBs.

GRB	Magnetar	α	Lagtime, ly
GRB080905570	CXOU J010043.1-721134	2.92347	264.873
GRB110818860	4U 0142+61	3.76798	25.3813
GRB130304658	SGR 0418+5729	1.08932	1.1789
GRB110319628	SGR 0501+4516	1.96316	3.82867
GRB181026540	SGR 0526-66	2.54773	172.803
GRB141102536	1E 1048.1-5937	0.656687	1.92799
GRB170803415	1E 1547.0-5408	3.19981	22.8822
GRB140619490	PSR J1622-4950	2.59277	30.0502
GRB090829672	SGR 1627-41	0.547211	1.63625
GRB190630257	CXOU J164710.2-455216	1.51366	4.43859
GRB120711446	1RXS J170849.0-400910	2.95812	16.5147
GRB140511095	CXOU J171405.7-381031	3.98594	104.138
GRB160819852	SGR J1745-2900	1.05186	4.56176
GRB160819852	SGR 1806-20	1.93754	16.2229
GRB160819852	XTE J1810-197	2.64306	12.1438
GRB160819852	Swift J1818.0-1607	2.13983	10.9169
GRB140627401	Swift J1822.3-1606	1.90291	2.87785
GRB150811849	SGR 1833-0832	1.92805	nan
GRB150811849	Swift J1834.9-0846	1.69774	6.01329
GRB171010792	1E 1841-045	1.49363	9.41957
GRB150705588	3XMM J185246.6+003317	2.4124	20.5231
GRB140723067	SGR 1900+14	1.25032	9.70704
GRB140723067	SGR 1935+2154	3.18331	nan
GRB140723067	1E 2259+586	2.92617	13.6084
GRB120224898	SGR 0755-2933 #	2.38008	nan
GRB171120556	SGR 1801-23 #	1.93446	nan
GRB180517309	SGR 1808-20 #	1.70049	nan
GRB151118554	AX J1818.8-1559 #	1.95134	nan
GRB130404877	AX J1845.0-0258 #	1.94219	15.9261
GRB170511477	SGR 2013+34 #	3.29051	47.3196
GRB151212064	PSR J1846-0258 ##	2.2824	15.5248

3. Summary

By building the distribution of magnetars on galactic map, it is confirmed that there are 29 galactic magnetars and their candidates, while the other two are located out of the Milkyway. The number of GRBs we studied is 2915. The maximum separation angle for GRB and Magnetar projectiles was 3.76 degrees (4U0142+61 and GRB110818860), while minimum angular resolution was 0.54 degrees (SGR1627-41 and GRB090829672). There are five candidates of lag time period for GRBs (GRB130304658, GRB110319628, GRB141102536, GRB090829672, GRB140627401) and corresponding magnetars(SGR 0418+5729, SGR 0501+4516, 1E 1048.1-5937, SGR 1627-41, Swift J1822.3-1606) we expect to study in the future works. We discuss the relationship of GRB light intensity by their lag time as it

would come after bending by the magnetosphere. The further steps will include modelling of the front emission from GRBs passing through the magnetosphere of the given magnetar.

References

1. V. M. Kaspi and A. M. Beloborodov, Magnetars, *Annual Review of Astronomy and Astrophysics* **55**, 261 (August 2017).
2. R. C. Duncan and C. Thompson, Formation of Very Strongly Magnetized Neutron Stars: Implications for Gamma-Ray Bursts, *Astrophysical Journal Letters* **392**, p. L9 (June 1992).
3. N. Rea and P. Esposito, Magnetar outbursts: An observational review, *Astrophysics and Space Science Proceedings* **21**, p. 247 (January 2011).
4. R. Turolla, S. Zane and A. L. Watts, Magnetars: The physics behind observations. A review, *Reports on Progress in Physics* **78**, p. 116901 (November 2015).
5. P. M. Woods and C. Thompson, *Soft gamma repeaters and anomalous X-ray pulsars: Magnetar candidates*, in *Compact stellar X-ray sources*, 2006, pp. 547–586.
6. S. Mereghetti, J. A. Pons and A. Melatos, Magnetars: Properties, Origin and Evolution, *Space Science Reviews* **191**, 315 (October 2015).
7. E. P. Mazets, S. V. Golenetskij and Y. A. Guryan, Soft gamma-ray bursts from the source B1900+14, *Soviet Astronomy Letters* **5**, p. 343 (December 1979).
8. C. Thompson and R. C. Duncan, The soft gamma repeaters as very strongly magnetized neutron stars - I. Radiative mechanism for outbursts, *Monthly Notices of the Royal Astronomical Society* **275**, 255 (July 1995).
9. C. Thompson and R. C. Duncan, The Soft Gamma Repeaters as Very Strongly Magnetized Neutron Stars. II. Quiescent Neutrino, X-Ray, and Alfven Wave Emission, *Astrophysical Journal* **473**, p. 322 (December 1996).
10. C. Kouveliotou, S. Dieters, T. Strohmayer, J. van Paradijs, G. J. Fishman, C. A. Meegan, K. Hurley, J. Kommers, I. Smith, D. Frail and T. Murakami, An X-ray pulsar with a superstrong magnetic field in the soft γ-ray repeater SGR1806 - 20, *Nature* **393**, 235 (May 1998).
11. G. G. Fahlman and P. C. Gregory, An X-ray pulsar in SNR G109.1-1.0, *Nature* **293**, 202 (September 1981).
12. F. P. Gavriil, V. M. Kaspi and P. M. Woods, Magnetar-like X-ray bursts from an anomalous X-ray pulsar, *Nature* **419**, 142 (September 2002).
13. V. M. Kaspi, F. P. Gavriil, P. M. Woods, J. B. Jensen, M. S. E. Roberts and D. Chakrabarty, A Major Soft Gamma Repeater-like Outburst and Rotation Glitch in the No-longer-so-anomalous X-Ray Pulsar 1E 2259+586, *The Astrophysical Journal* **588**, L93 (May 2003).
14. S. A. Olausen and V. M. Kaspi, The McGill Magnetar Catalog, *The Astrophysical Journal* **212**, p. 6 (May 2014).
15. N. Gehrels and Swift, The Swift Gamma-Ray Burst Mission, in *American Astronomical Society Meeting Abstracts*, American Astronomical Society Meeting Abstracts Vol. 205, December 2004.
16. C. Meegan, G. Lichti, P. N. Bhat, E. Bissaldi, M. S. Briggs, V. Connaughton, R. Diehl, G. Fishman, J. Greiner, A. S. Hoover, A. J. van der Horst, A. von Kienlin, R. M. Kippen, C. Kouveliotou, S. McBreen, W. S. Paciesas, R. Preece, H. Steinle, M. S. Wallace, R. B. Wilson and C. Wilson-Hodge, The Fermi Gamma-ray Burst Monitor, *The Astrophysical Journal* **702**, 791 (September 2009).

17. A. De Luca, P. A. Caraveo, S. Mereghetti, A. Tiengo and G. F. Bignami, A Long-Period, Violently Variable X-ray Source in a Young Supernova Remnant, *Science* **313**, 814 (August 2006).
18. A. D'Aì, P. A. Evans, D. N. Burrows, N. P. M. Kuin, D. A. Kann, S. Campana, A. Maselli, P. Romano, G. Cusumano, V. La Parola, S. D. Barthelmy, A. P. Beardmore, S. B. Cenko, M. De Pasquale, N. Gehrels, J. Greiner, J. A. Kennea, S. Klose, A. Melandri, J. A. Nousek, J. P. Osborne, D. M. Palmer, B. Sbarufatti, P. Schady, M. H. Siegel, G. Tagliaferri, R. Yates and S. Zane, Evidence for the magnetar nature of 1E 161348-5055 in RCW 103, *Monthly Notices of the Royal Astronomical Society* **463**, 2394 (December 2016).
19. N. Rea, P. Esposito, R. Turolla, G. L. Israel, S. Zane, L. Stella, S. Mereghetti, A. Tiengo, D. Götz, E. Göğüş and C. Kouveliotou, A Low-Magnetic-Field Soft Gamma Repeater, *Science* **330**, p. 944 (November 2010).
20. V. M. Kaspi and M. A. McLaughlin, Chandra X-Ray Detection of the High Magnetic Field Radio Pulsar PSR J1718-3718, *The Astrophysical Journal* **618**, L41 (January 2005).
21. C. Y. Ng, V. M. Kaspi, R. Dib, S. A. Olausen, P. Scholz, T. Güver, F. Özel, F. P. Gavriil and P. M. Woods, Chandra and RXTE Observations of 1E 1547.0-5408: Comparing the 2008 and 2009 Outbursts, *The Astrophysical Journal* **729**, p. 131 (March 2011).
22. F. Camilo, S. M. Ransom, J. P. Halpern, J. Reynolds, D. J. Helfand, N. Zimmerman and J. Sarkissian, Transient pulsed radio emission from a magnetar, *Nature* **442**, 892 (August 2006).
23. F. Camilo, S. M. Ransom, J. P. Halpern and J. Reynolds, 1E 1547.0-5408: A Radio-emitting Magnetar with a Rotation Period of 2 Seconds, *The Astrophysical Journal* **666**, L93 (September 2007).
24. A. M. Beloborodov, Gravitational Bending of Light Near Compact Objects, *The Astrophysical Journal* **566**, L85 (February 2002).
25. V. I. Denisov, I. P. Denisova and S. I. Svertilov, Nonlinear Electrodynamic Effect of Ray Bending in the Magnetic-Dipole Field, *Physics - Doklady* **46**, 705 (October 2001).
26. M. Abishev, Y. Aimuratov, Y. Aldabergenov, N. Beissen, Z. Bakytzhan and M. Takibayeva, Some astrophysical effects of nonlinear vacuum electrodynamics in the magnetosphere of a pulsar, *Astroparticle Physics* **73**, 8 (January 2016).
27. S. I. Shirokov, A. A. Raikov and Y. V. Baryshev, Spatial Distribution of Gamma-Ray Burst Sources, *Astrophysics* **60**, 484 (December 2017).
28. C. Kouveliotou, C. A. Meegan, G. J. Fishman, N. P. Bhat, M. S. Briggs, T. M. Koshut, W. S. Paciesas and G. N. Pendleton, Identification of Two Classes of Gamma-Ray Bursts, *Astrophysical Journal Letters* **413**, p. L101 (August 1993).
29. A. von Kienlin, C. A. Meegan, W. S. Paciesas, P. N. Bhat, E. Bissaldi, M. S. Briggs, E. Burns, W. H. Cleveland, M. H. Gibby, M. M. Giles, A. Goldstein, R. Hamburg, C. M. Hui, D. Kocevski, B. Mailyan, C. Malacaria, S. Poolakkil, R. D. Preece, O. J. Roberts, P. Veres and C. A. Wilson-Hodge, The Fourth Fermi-GBM Gamma-Ray Burst Catalog: A Decade of Data, *The Astrophysical Journal* **893**, p. 46 (April 2020).
30. M. E. Abishev, S. Toktarbay, N. A. Beissen, F. B. Belissarova, M. K. Khassanov, A. S. Kudussov and A. Z. Abylayeva, Effects of non-linear electrodynamics of vacuum in the magnetic quadrupole field of a pulsar, *Monthly Notices of the Royal Astronomical Society* **481**, 36 (November 2018).

Absorption of massless scalar waves by electrically charged regular black holes

Marco A. A. de Paula[1,*], Luiz C. S. Leite[1,2,†], and Luís C. B. Crispino[1,‡]

[1] *Programa de Pós-Graduação em Física, Universidade Federal do Pará, 66075-110, Belém, Pará, Brazil*

[2] *Campus Altamira, Instituto Federal do Pará, 68377-630, Altamira, Pará, Brazil*

[*] *E-mail: marco.paula@icen.ufpa.br*

[†] *E-mail: luiz.leite@ifpa.edu.br*

[‡] *E-mail: crispino@ufpa.br*

We study the absorption cross section of a massless test scalar field, for arbitrary frequency values, in the background of an Ayón-Beato-García electrically charged regular black hole spacetime. We show that Ayón-Beato-García regular black holes can mimic the absorption properties of Reissner-Nordström black holes in the whole frequency range, for small-to-moderate values of the normalized electric charge.

Keywords: Regular black holes; scalar field; absorption cross section.

1. Introduction

General relativity (GR) is a geometric theory of gravity that over more than one hundred years has successfully passed several experimental tests.[1–3] The observational evidences of GR[4–6] confirm the existence of black holes (BHs). These objects are characterized by a one-way membrane, called the event horizon, and in standard GR they are normally associated to a curvature singularity. Since the known laws of physics break down at such singularities, the standard GR is not able to explain the physics of a BH singular core.

An attempt to avoid the curvature singularity issues are the so-called regular BH (RBH) spacetimes,[7] i.e., singularity-free black hole geometries. The first exact charged RBH solution was proposed in 1998 by Eloy Ayón-Beato and Alberto García (ABG).[8] This solution was based on the minimal coupling between gravity and a nonlinear electrodynamics model, and describes a static and electrically charged RBH. The nonlinear electrodynamics theory can be seen as a possible generalization of Maxwell's linear electrodynamics for strong electromagnetic fields.[9]

In real astrophysical scenarios BHs are surrounded by distributions of matter.[10] Among the bosonic and fermionic distributions of matter, the simplest one is the scalar field, corresponding to spinless particles.[11] It is possible to study how BHs absorb and scatter matter fields as an attempt to better comprehend how BHs interact with their surroundings.[12–17] We present a selection of results for the massless scalar absorption by ABG RBHs, focusing in the possibility of RBHs mimic the absorption properties of standard BHs. More details on this work are presented in a previously published paper (see Ref. 18).

2. Ayón-Beato-García RBH Spacetime

The line element of the ABG solution is given by[8]

$$ds^2 = f(r)dt^2 - f(r)^{-1}dr^2 - r^2\left(d\theta^2 + \sin^2\theta\, d\varphi^2\right), \tag{1}$$

with

$$f(r) = f^{\rm ABG}(r) \equiv 1 - \frac{2Mr^2}{(r^2+Q^2)^{3/2}} + \frac{Q^2r^2}{(r^2+Q^2)^2}, \tag{2}$$

where M and Q are the mass and the electric charge of the BH, respectively. The asymptotic limits of the metric function are given by

$$\lim_{r\to 0} f^{\rm ABG}(r) \approx \ f^{\text{de Sitter}}(r) = 1 - \frac{1}{3}\left[\frac{3\left(2|Q|M - Q^2\right)}{Q^4}\right] r^2 \tag{3}$$

and

$$\lim_{r\to \infty} f^{\rm ABG}(r) \approx \ f^{\rm RN}(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}. \tag{4}$$

The ABG solution describes RBHs when $|Q| \leq Q_{\rm ext} \approx 0.6341M$. For $|Q| < Q_{\rm ext}$ we have a Cauchy, r_-, and an event horizon, r_+, given by

$$r_\pm = |Q|\sqrt{\left(\pm\frac{\sqrt{-\frac{9(12s^2-1)}{\sqrt{u(s)}} - 12s^2 - \frac{u(s)}{6} + \frac{9}{2}}}{2\sqrt{6}s} + \frac{\sqrt{u(s)}}{12s} + \frac{1}{4s}\right)^2 - 1}, \tag{5}$$

where $s \equiv |Q|/(2M)$ and

$$u(s) \equiv 6\left(-\frac{4(11s^2-3)s}{\sqrt[3]{g(s)}} + s\sqrt[3]{g(s)} - 4s^2 + \frac{3}{2}\right), \tag{6}$$

$$g(s) \equiv 4(74s^3 + 3\sqrt{3}\sqrt{400s^6 - 112s^4 + 47s^2 - 4} + 9s). \tag{7}$$

On the other hand, $Q = Q_{\rm ext}$ corresponds to the so-called extreme ABG RBH, for which we have only one horizon; while $Q > Q_{\rm ext}$ corresponds to a horizonless solution. The regularity of the solution at the core of the geometry is evidenced by the behavior of the scalar invariants presented in Ref. 8, which are all regular there.

3. Absorption of Massless Test Scalar Fields

3.1. *Massless scalar field*

The massless and chargeless test scalar field Φ obeys the Klein-Gordon equation:

$$\nabla_\mu\nabla^\mu\Phi = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\right)\Phi = 0. \tag{8}$$

We can decompose Φ as

$$\Phi = \sum C_{\omega l} \left(\frac{\psi_{\omega l}(r)}{r} P_l(\cos\theta) e^{-i\omega t} \right), \tag{9}$$

where $C_{\omega l}$ are constant coefficients. The indexes ω and l denote the frequency and the angular momentum of the wave, respectively, $\psi_{\omega l}(r)$ is the radial function and P_l are the Legendre Polynomials.

By inserting Eq. (9) into Eq. (8) we may find the following radial equation

$$\frac{d^2}{dr_\star^2}\psi_{\omega l} + \left[\omega^2 - V_{\rm eff}(r)\right]\psi_{\omega l} = 0, \tag{10}$$

where $\frac{dr_\star}{dr} \equiv \frac{1}{f(r)}$ is the tortoise coordinate and the effective potential $V_{\rm eff}(r)$ reads

$$V_{\rm eff}(r) \equiv f(r)\left[\frac{1}{r}\frac{df(r)}{dr} + \frac{l(l+1)}{r^2}\right]. \tag{11}$$

In Fig. 1 we present the effective potential in the background of ABG RBHs. We note that the peak of the effective potential increases as we consider higher values of the normalized charge $\alpha \equiv Q/Q_{\rm ext}$, as well as higher values of l; and for the asymptotic limits, we have $\lim_{r_\star \to \pm\infty} V_{\rm eff}(r_\star) = 0$.

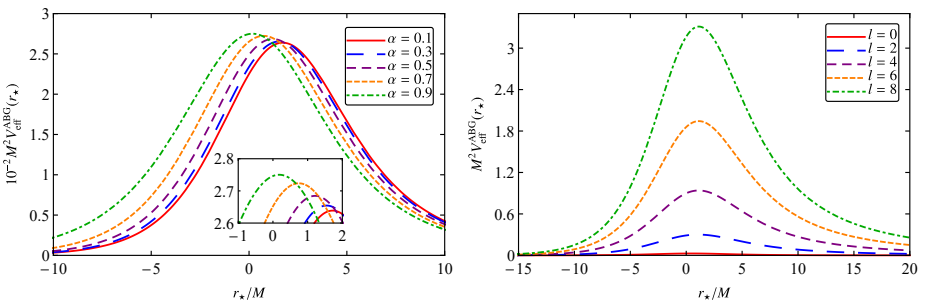


Fig. 1. The effective potential of scalar waves in the background of the ABG RBH, as a function of $r_\star$: (i) with $l = 0$, for different choices of the normalized charge α (left panel); and (ii) for different choices of l, with $\alpha = 0.8$ (right panel). The lower inset in the left panel emphasizes the peaks of the effective potential.

The appropriate boundary conditions are given by plane waves incoming from the infinite null past (the so-called *in modes*), so that the solutions of Eq. (10) reads

$$\psi_{\omega l} \sim \begin{cases} T_{\omega l} e^{-i\omega r_\star}, & r \to r_+ \ (r_\star \to -\infty), \\ e^{-i\omega r_\star} + R_{\omega l} e^{i\omega r_\star}, & r \to \infty \ (r_\star \to \infty), \end{cases} \tag{12}$$

where $T_{\omega l}$ and $R_{\omega l}$ are complex coefficients. By using the conservation of the flux, one can show that the transmission $|T_{\omega l}|^2$ and reflection $|R_{\omega l}|^2$ coefficients satisfy

$$|R_{\omega l}|^2 + |T_{\omega l}|^2 = 1. \tag{13}$$

3.2. *Absorption cross section*

By using the so-called partial-wave method, we may write the total absorption cross section (ACS) of the massless scalar wave as a sum of partial-waves contributions, given by[19]

$$\sigma = \sum_{l=0}^{\infty} \sigma_l, \tag{14}$$

where the partial ACS, σ_l, reads

$$\sigma_l = \frac{\pi}{\omega^2}(2l+1)|T_{\omega l}|^2. \tag{15}$$

There are analytic approximations for the ACS of massless test scalar fields by BHs in the low- and high-frequency regimes. In the low-frequency regime, for stationary BH geometries, the ACS tends to the surface area of the BH event horizon, namely[20,21]

$$\mathrm{A} = 4\pi r_+^2. \tag{16}$$

On the other hand, in the high-frequency regime, the absorption of a massless scalar field is governed by the geometric cross section (GCS) of null geodesics, i.e., $\sigma_{\mathrm{gcs}} = \pi b_c^2$, where b_c is the critical impact parameter. By using the so-called sinc approximation, it is possible to obtain an improvement for the high-frequency approximation, which can be expressed as[22]

$$\sigma_{\mathrm{hf}} = \sigma_{\mathrm{gcs}}\left[1 - 8\pi b_c \Lambda e^{-\pi b_c \Lambda}\mathrm{sinc}\left(2\pi b_c \omega\right)\right], \tag{17}$$

where $\mathrm{sinc}(x) \equiv \sin(x)/x$, and Λ is the Lyapunov exponent.[23]

4. Results

4.1. *Main features*

The numerical method employed to obtain the ACS consists in integrating numerically Eq. (10) from very close to the BH event horizon r_+, up to some radial position very far from the BH, with the boundary conditions in these regions given by Eq. (12).

In Fig. 2 we show the total ACS, normalized by the BH area, for distinct ABG RBHs. We note that, as $\omega \to 0$, the ratio σ/A tends to the unity. Hence, at the low-frequency limit, the numerical result for the ACS tends to the BH area, as expected.

A comparison between the horizon area of ABG and RN BHs is presented in Fig. 3. We see that for fixed values of α, ABG and RN BHs have very similar BH areas.

In Fig. 4 we compare the total ACS obtained numerically with the analytic approximations for the high-frequency regime (see Subsec. 3.2). As we can see, in this frequency regime, the total ACS obtained numerically oscillates around the

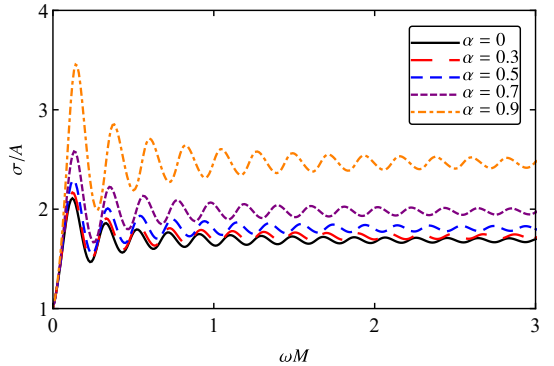


Fig. 2. The total ACS of the scalar wave for distinct ABG RBHs (divided by the BH area), as a function of ωM. We also exhibit, for comparison, the results for the Schwarzschild case ($\alpha = 0$).

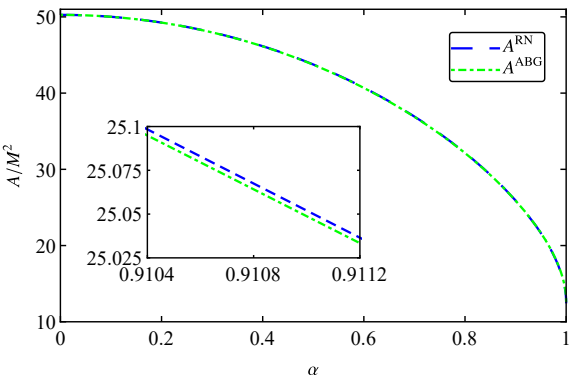


Fig. 3. The surface area of the BH event horizon of ABG and RN BHs, as a function of the normalized charge. The central inset shows the areas of ABG and RN BHs for α between 0.9104 and 0.9116, emphasizing the high similarity between the corresponding areas.

corresponding GCS and the oscillatory profile is very well reproduced by the sinc approximation. These results, combined with the ones for the low-frequency regime (cf. Fig. 2), show that our numerical results are consistent with the analytic approximations available in the literature. We also see that the total ACS of the ABG RBH diminishes as we increase α, so that the wave is more absorbed in the background of small-charge ABG RBH spacetimes. This is in accordance with the fact that the potential barrier raises as we increase α (see Fig. 1).

Examples of partial ACSs in the background of different ABG RBHs are shown in Fig. 5. We observe that the peak of the partial-wave modes diminishes as we consider higher values of the normalized electric charge and the mode $l = 0$ provides the main contributions for the ACS in the low-frequency regime.

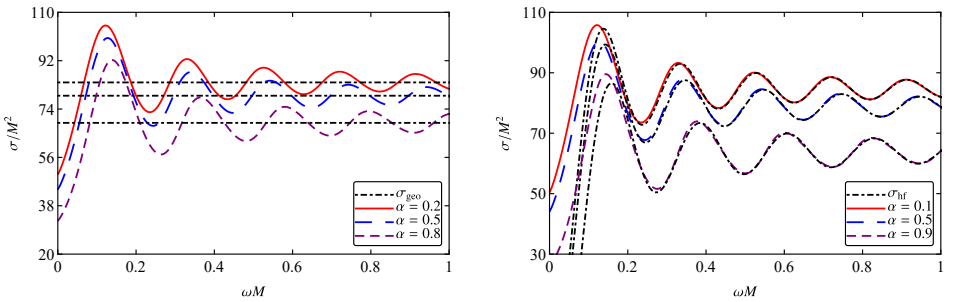


Fig. 4. Comparison of the total ACS (as a function of ωM) of the ABG RBH obtained numerically with: (i) the corresponding GCSs (left panel) and (ii) the corresponding sinc approximation (right panel), considering distinct values of α.

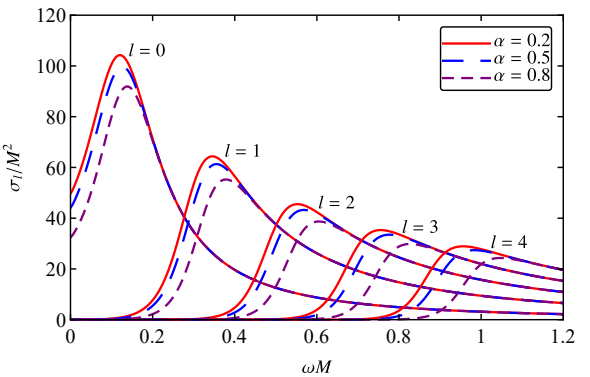


Fig. 5. Partial ACSs of the scalar wave considering different ABG RBHs (distinct values of α), as functions of ωM.

In Fig. 6 we compare the absorption results for ABG and RN BHs, considering two values of α. The total ACS of the ABG RBH is typically larger than the corresponding RN case. However, for small values of α, the total ACS of ABG and RN BH solutions can be very similar for arbitrary values of the frequency. Notice also that the zero-frequency regime of the total ACSs of ABG and RN BHs are very similar, for the same α value. This is in agreement with the behavior of their areas which are very similar for the same choices of α (cf. Fig. 3).

4.2. *RBHs as standard BHs mimickers*

We can consider the values of $(\alpha^{\rm ABG}, \alpha^{\rm RN})$ for which the corresponding GCSs coincide in order to find situations in which the absorption results of ABG and RN BHs are very similar. In Fig. 7 we display such values, which can be found up to $(\alpha^{\rm ABG}, \alpha^{\rm RN}) = (1, 0.9161)$.

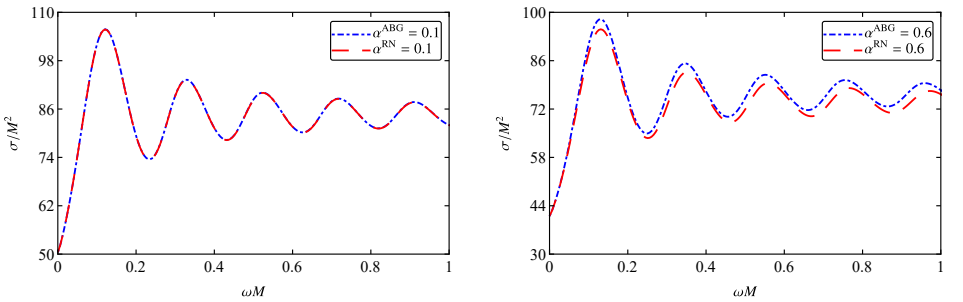


Fig. 6. The total ACSs of ABG and RN BHs as a function of ωM, for $\alpha = 0.1$ (left panel) and for $\alpha = 0.6$ (right panel).

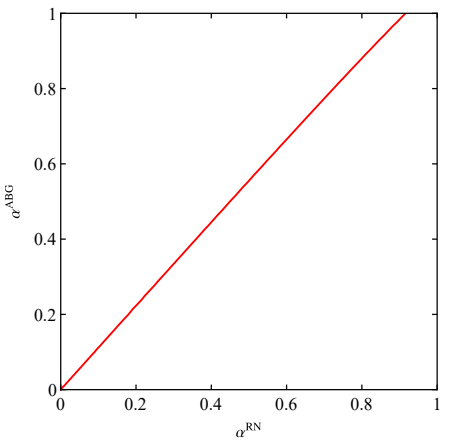


Fig. 7. The values of the normalized charges for which the GCSs of ABG and RN BHs coincide.

In Fig. 8 we exhibit the total ACSs for two pairs of $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}})$, for which the GCSs coincide. For low-to-moderate α values, the total ACSs of the ABG and RN BHs are basically indistinguishable along the whole frequency regime. This contrasts with the absorption results for massless test scalar fields in the Bardeen case,[24] for which the ACSs of Bardeen and RN BH solutions can be very similar only in the high-frequency regime.

We can define a function to estimate the percentual deviation between the BH area of each pair $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}})$, namely

$$\mathrm{A}_{\mathrm{dif}} \equiv \frac{\mathrm{A}^{\mathrm{RN}} - \mathrm{A}^{\mathrm{ABG}}}{\mathrm{A}^{\mathrm{ABG}}} 100\%. \tag{18}$$

The smaller the function $\mathrm{A}_{\mathrm{dif}}$ is, the more similar are the ABG and RN BHs areas. In Table 1 we present the corresponding BH areas, as well as $\mathrm{A}_{\mathrm{dif}}$, for several pairs $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}})$ displayed in Fig. 7. As we can see, the function $\mathrm{A}_{\mathrm{dif}}$ increases with

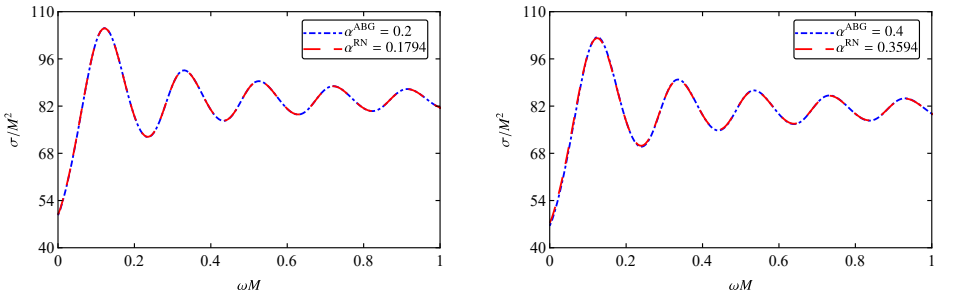


Fig. 8. The total ACS for the pairs $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}}) = (0.2, 0.1794)$ (left panel) and $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}}) = (0.4, 0.3594)$ (right panel). In both cases, the corresponding GCSs coincide.

Table 1. The BH areas and $\mathrm{A_{dif}}$ for several pairs $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}})$ shown in Fig. 7.

α^{ABG}	α^{RN}	$\mathrm{A}^{\mathrm{ABG}}/M^2$	$\mathrm{A}^{\mathrm{RN}}/M^2$	$\mathrm{A_{dif}}$
0.1516	0.1365	49.6831	49.7961	0.2274%
0.182	0.1629	49.425	49.5963	0.3453%
0.2027	0.1815	49.2219	49.4341	0.4309%
0.2852	0.2548	48.189	48.6201	0.8946%
0.318	0.2855	47.6772	48.1951	1.0863%
0.3417	0.3054	47.2708	47.8927	1.3156%
0.4003	0.3608	46.1309	46.9368	1.7469%
0.4707	0.4226	44.4976	45.6667	2.6273%
0.5003	0.4489	43.7207	45.0586	3.0601%

the increase of α. We also note that for an ABG RBH with $\alpha \lesssim 0.5$, we have that $\mathrm{A_{dif}} \lesssim 3\%$. Therefore, pairs $(\alpha^{\mathrm{ABG}}, \alpha^{\mathrm{RN}})$ that satisfy $\sigma^{\mathrm{ABG}}_{\mathrm{gcs}} = \sigma^{\mathrm{RN}}_{\mathrm{gcs}}$ correspond to very similar BH areas, as long as we consider only low-to-moderate values of α, implying that the ACSs of ABG and RN BHs are very similar also at the low-frequency regime. This also contrasts with the absorption results for massless test scalar field in the Bardeen case.[24]

5. Final Remarks

We revisited the scalar absorption of massless scalar fields by the first electrically charged ABG RBH solution,[18] presenting a selection of additional results. The height of potential barrier (related to the propagation of massless scalar waves in the background of the ABG RBH) enhances as we increase l or α. The ACSs obtained with our numerical method tend to the corresponding BH area in the low-frequency regime, as expected,[20] and, in this limit, the mode $l = 0$ provides the

main contribution to the ACS. We also obtained that the BH areas of ABG and RN BHs are very similar for certain values of α.

The total ACS of the ABG RBH, in the mid-to-high frequency limit, oscillates around the corresponding GCS and the sinc approximation provides excellent results for ACS in this regime. Therefore, considering the results in the low- and high-frequency regime, we conclude that our numerical results for the ACS agree very well with the analytic approximations. We also noted that the ABG RBH total ACS diminishes as we increase α, what is in accordance with the fact that $V_{\text{eff}}(r)$ increases as we increase α. The ABG RBH total ACS is typically larger than the RN one, for the same choice of α, but it is possible to find situations for which the total ACSs of ABG and RN BHs are very similar in the whole frequency range, mainly for small values of α. This similarity shows that it is possible that electrically charged RBHs mimic standard BHs, in what concerns the absorption of massless test scalar fields.

Acknowledgments

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001, from Brazil, for partial financial support. This research has also received funding from the European Union's Horizon 2020 research and innovation programme under the H2020-MSCA-RISE-2017 Grant No. FunFiCO-777740.

References

1. C. M. Will, The Confrontation between General Relativity and Experiment, Living Rev. Relativ. **17**, 4 (2014).
2. L. C. B. Crispino and D. J. Kennefick, A hundred years of the first experimental test of general relativity, Nat. Phys. **115**, 416 (2019).
3. L. C. B. Crispino and S. Paolantonio, The first attempts to measure light deflection by the Sun, Nat. Astron. **4**, 6 (2020).
4. R. Narayan and J. E. McClintock, Observational Evidence for Black Holes, arXiv:1312.6698 [astro-ph.HE].
5. B. P. Abbott *et al.* [LIGO Scientific Collaboration and Virgo Collaboration], Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).
6. The Event Horizon Telescope Collaboration *et al.*, First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, Atrophys. J. **875**, L1 (2019).
7. J. Bardeen, Non-singular General Relativistic Gravitational Collapse, in *Proceedings of the International Conference GR5* (Tbilisi, Georgia, U.S.S.R., 1968), p. 174.
8. E. Ayón-Beato and A. García, Regular Black Hole in General Relativity Coupled to Nonlinear Electrodynamics, Phys. Rev. Lett. **80**, 5056 (1998).
9. M. Born and L. Infeld, Foundations of the new field theory, Proc. R. Soc. A **144**, 425 (1934).
10. R. Narayan, Black holes in astrophysics, New J. Phys. **7**, 199 (2005).

11. W. Greiner and J. Reinhardt, *Field Quantization* (Springer, Berlim, 1996).
12. J. A. Futterman, F. A. Handler, and R. A. Matzner, *Scattering from Black Holes* (Cambridge University Press, Cambridge, England, 1988).
13. L. C. B. Crispino, S. R. Dolan, and E. S. Oliveira, Scattering of massless scalar waves by Reissner-Nordström black holes, Phys. Rev. D **79**, 0640022 (2009).
14. E. S. Oliveira, L. C. B. Crispino, and A. Higuchi, Equality between gravitational and electromagnetic absorption cross sections of extreme Reissner-Nordstrom black holes, Phys. Rev. D **84**, 084048 (2011).
15. L. C. B. Crispino, S. R. Dolan, A. Higuchi, and E. S. de Oliveira, Inferring black hole charge from backscattered electromagnetic radiation, Phys. Rev. D **90**, 064027 (2014).
16. L. C. B. Crispino, S. R. Dolan, A. Higuchi, and E. S. de Oliveira, Scattering from charged black holes and supergravity, Phys. Rev. D **92**, 084056 (2015).
17. C. L. Benone and L. C. B. Crispino, Superradiance in static black hole spacetimes, Phys. Rev. D **93**, 024028 (2016).
18. M. A. A. Paula, L. C. S. Leite, and L. C. B. Crispino, Electrically charged black holes in linear and non-linear electrodynamics: Geodesic analysis and scalar absorption, Phys. Rev. D **102**, 104033 (2020).
19. W. Unruh, Absorption cross section of small black holes, Phys. Rev. D **14**, 3251 (1976).
20. S. R. Das, G. Gibbons, and S. D. Mathur, Universality of Low Energy Absorption Cross Sections for Black Holes, Phys. Rev. Lett. **78**, 417 (1997).
21. A. Higuchi, Low-frequency Scalar absorption cross sections for stationary black holes, Classical Quantum Gravity **18**, L139 (2001); **19**, 599(A) (2002).
22. Y. Décanini, G. Esposito-Farèse, and A. Folacci, Universality of high-energy absorption cross sections for black holes, Phys. Rev. D **83**, 044032 (2011).
23. V. Cardoso, A. S. Miranda, E. Berti, H. Witek, and V. T. Zanchin, Geodesic stability, Lyapunov exponents, and quasinormal modes, Phys. Rev. D **79**, 064016 (2009).
24. C. F. B. Macedo and L. C. B. Crispino, Absorption of planar massless scalar waves by Bardeen regular black holes, Phys. Rev. D **90**, 064001 (2014).

Modeling Type Ia supernovae with explosions in white dwarfs near and below the Chandrasekhar mass

Friedrich K. Röpke

Zentrum für Astronomie der Universität Heidelberg, Institut für Theoretische Astrophysik and Heidelberger Institut für Theoretische Studien Heidelberg, Germany
E-mail: friedrich.roepke@h-its.org

Florian Lach and Sabrina Gronow

Zentrum für Astronomie der Universität Heidelberg, Astronomisches Recheninstitut and Heidelberger Institut für Theoretische Studien Heidelberg, Germany

Stuart A. Sim and Fionntan P. Callan

Astrophysics Research Center, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK

Christine E. Collins

Astrophysics Research Center, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK and
GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany

The progenitor evolution and the explosion mechanism of Type Ia supernovae remain unexplained. Nonetheless, substantial progress has been made over the past years with three-dimensional hydrodynamic simulations of different scenarios. Here, we review some recent work pertaining to the leading paradigms of modeling: thermonuclear explosions of white dwarf stars near and below the Chandrasekhar mass limit. We discuss implications of the different explosion channels and their predictions of observables.

Keywords: Type Ia supernovae, thermonuclear explosions, white dwarf stars, numerical simulations

1. Introduction

Despite substantial progress in theoretical modeling and numerical simulations over the past years,[1] our understanding of the physical mechanism of Type Ia supernovae remains incomplete. This has two main reasons. (i) The progenitor systems from which these explosions arise have not been identified, and therefore the initial conditions for the explosion simulations are uncertain. (ii) Modeling the explosion stage itself is a severe multi-scale multi-physics challenge and relies on assumptions and approximations. Some of these approximations could be mitigated with multi-dimensional hydrodynamical simulations. They form a cornerstone of a consistent modeling pipeline that follows a progenitor model over explosion and nucleosynthesis to the formation of observables. By avoiding tunable parameters, such a modeling pipeline facilitates a direct comparison of model predictions with astronomical data.

This allows for conclusions to be drawn on the validity of the assumed progenitor scenarios. In the following, we describe the application of this modeling pipeline to two different explosion scenarios.

2. Explosion models

Ignoring the fascinating but complex and still enigmatic evolution of progenitor systems of Type Ia supernovae, the main question to simulations is how to set up the state of the white dwarf at the onset of explosion. The two fundamental choices, a configuration close to the limit of stability, the Chandrasekhar mass, and a white dwarf below that mass limit, imply different explosion scenarios.[2] The compact structure of a near-Chandrasekhar mass object causes high densities of the material ahead of the thermonuclear burning front[3] if it propagates as a supersonic detonation. The products of such an explosion, almost exclusively iron group elements, are inconsistent with observations of Type Ia supernova. To produce the required intermediate mass elements detected in their spectra, burning has to start out as a subsonic deflagration in a white dwarf close to the Chandrasekhar mass. After some time of pre-expansion of the star, the burning front may turn into a supersonic detonation. In a sub-Chandrasekhar mass white dwarf, in contrast, the densities are lower and allow for the required intermediate-mass elements to be produced in a detonation.

For both scenarios, the actual ignition of the burning remains uncertain and is difficult to resolve in multidimensional hydrodynamic simulations.[4] Therefore, simulations often start out with an assumption on the triggering of the explosive burning.

3. Near-Chandrasekhar mass explosions

Sets of simulations have been carried out to test the impact of initial parameters on the outcome of explosions in near-Chandrasekhar mass white dwarf stars. Testing the ignition configuration[5] revealed that the number and spatial distribution of ignition sparks is the most important parameter for the strength of the deflagration. Few and asymmetrically distributed sparks lead to an incomplete disruption of the white dwarf. With very many ignition kernels (that are less likely to be realized in nature[4,6]), a complete unbinding of the star becomes possible, but the mass of ^{56}Ni produced is too low to explain the brighness of normal Type Ia supernovae. A detonation may form later and enhance the thermonuclear burning,[7] but here we restrict our discussion to cases where the flame propagation remains subsonic throughout.

Pure deflagrations in near-Chandrasekhar mass white dwarfs have been discussed as a model for the subclass of Type Iax supernovae.[8] An open question, however, remains: Can deflagrations in Chandrasekhar-mass white dwarfs cover the entire range of objects in this class, including the very faint events? To explore this,

we have carried out an extended systematic study of three-dimensional hydrodynamic explosion simulations[9] varying the distance of single-spark ignitions from the stellar center, but also other parameters such as the central density of the white dwarf at the onset of explosion, its metallicity, its carbon mass fraction, and its rotation state. This suite of models shows that it it well possible to decrease the ^{56}Ni production and thus the brightness of the modeled events to values that would match the faintest members of the Type Iax supernova class. However, inconsistencies were discovered, too. The faint events evolve to quickly in brightness. All models fall onto a strong correlation between the produced ^{56}Ni mass and the total eject mass. This correlation does not match observations and none of the initial parameters was able to perturb it significantly. For the brighter models, however, reasonable matches with observations were found. Previous claims of chemically layered ejecta structures[10] based on the "abundance tomography" method contradict the picture of Type Iax supernovae originating from deflagrations in Chandrasekhar-mass white dwarfs. Because of the intrinsic instabilities of subsonic flame propagation, such a scenario would predict well-mixed ejecta. Recent forward-modeling,[11] however, finds that the predictions of such models may still be consistent with observations.

Improvements in explosion modeling and – in particular – in the treatment of non local thermodynamic equilibrium (NLTE) effects in the radiation transfer calculations are needed to settle the question of whether deflagrations in Chandrasekhar-mass white dwarfs can explain at least the brighter Type Iax supernovae. Given the failure to model the faint events in this framework, it seems possible that not all members of the observationally-defined class of Type Iax supernovae pertain to the same physical explosion mechanism.

4. Sub-Chandrasekhar mass explosions

Explosions of white dwarf stars below the Chandrasekhar mass are an appealing model because they seem to reproduce important observational trends.[12] The question, however, is how such inert objects trigger a detonation. A classical model is that of double detonations: A helium shell is accreted on top of a carbon-oxygen white dwarf. Once massive enough, it triggers a shell detonation that initiates a secondary detonation of the carbon-oxygen core. If the helium shell is not too massive, its products do not strongly impact the observables and the match with data from normal Type Ia supernovae improves.[13–16]

We have recently explored the mechanism of triggering of the secondary core detonation and the impact of the shell detonation products on predicted observables in an extended sequence of three-dimensional hydrodynamic simulations.[17–19] This study identifies different possibilities for the core detonation initiation depending on the mass of the helium shell and the carbon-oxygen core. Although a reasonable match is obtained in the predicted observables with observational data, some shortcomings remain. These include too red spectra and too wide variations of the lightcurve width-luminosity relation with viewing angle. Some of these deficiencies

can be attributed to approximations in the treatment of NLTE effects in radiation transport,[16] but the mismatches may also call into question the explosion model itself.

5. Imprints on nucleosynthesis yields

Apart from comparing to optical observables, another approach to discriminate between and assess the validity of Type Ia supernova explosion models is by their imprints on the nucleosynthesis yields.[20] An important difference between near- and sub-Chandrasekhar mass explosion models is the production of manganese.[21] In explosive carbon burning, it can only be produced in (super-)solar ratio to iron if the densities are sufficiently high to allow for normal freeze-out from nuclear statistical equilibrium. This is the case for explosive burning in the cores of Chandrasekhar-mass white dwarfs. Alpha-rich freezeout, as occurring at lower densities in explosions of sub-Chandrasekhar mass objects, destroys the mother nucleus of ^{55}Mn, ^{55}Co, by proton captures. This produces additional ^{56}Ni at the expense of manganese. Therefore, it was concluded that a substantial fraction of Type Ia supernovae has to originate from the Chandrasekhar-mass explosion channel so that these objects can drive the manganese-over-iron trend in galactic chemical evolution towards the solar value. Our new double-detonation sub-Chandrasekhar mass explosion models,[19,20] however, show that additional manganese can be produced in the helium shell detonation. This lowers the fraction of Chandrasekhar-mass models needed to explain the galactic chemical evolution of manganese.

6. Conclusions

Three-dimensional hydrodynamic simulations help to avoid tunable parameters in the modeling of different explosion scenarios for Type Ia supernovae. The optical observables derived from such models via nucleosynthesis postprocessing[22] and radiative transfer calculations can be exposed directly to observational data. For the time being, however, the discriminative power of this approach is insufficient to identify a valid model for normal Type Ia supernovae. All considered scenarios have some advantages and some shortcomings. The reason may simply be that the correct scenario has not yet been found. Sub-Chandrasekhar mass explosions are a promising model, but in the double detonation mechanism they still fail to match some important observational properties of Type Ia supernovae. Similar explosions can, however, also be triggered by mergers of two white dwarfs.[23–26]

A similar situation is encountered with deflagrations in near-Chandrasekhar mass white dwarf stars. While this model looks promising for explaining brighter members of the Type Iax supernova class, its fainter end cannot be reproduced.

To ultimately settle the question of the origin of Type Ia supernovae, constant improvement is required in the explosion modeling as well as in the treatment of radiative transfer predicting the optical observables. Alternative observables that

may help to discriminate between models include the nucleosynthesis yields discussed here, but also spectropolarimetry data,[27–29] the search for surviving companion stars in the double degenerate progenitor model[30] and imprints of different explosion scenarios on the forming supernova remnants.[31,32]

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 138713538 – SFB 881 ("The Milky Way System", subproject A10), by the ChETEC COST Action (CA16117), and by the National Science Foundation under Grant No. OISE-1927130 (IReNA). FL and FKR acknowledge support by the Klaus Tschira Foundation. FPC acknowledges an STFC studentship and SAS acknowledges funding from STFC Grant Ref: ST/P000312/1. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer JUWELS[33] at Jülich Supercomputing Centre (JSC). Part of this work was performed using the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. DiRAC is part of the National e-Infrastructure.

References

1. W. Hillebrandt, M. Kromer, F. K. Röpke and A. J. Ruiter, Towards an understanding of type ia supernovae from a synthesis of theory and observations, *Frontiers of Physics* **8**, 116 (April 2013).
2. I. R. Seitenzahl and D. M. Townsley, *Nucleosynthesis in Thermonuclear Supernovae*, in *Handbook of Supernovae*, eds. A. W. Alsabti and P. Murdin 2017, p. 1955.
3. F. K. Röpke, Combustion in thermonuclear supernova explosions, in *Handbook of Supernovae*, eds. A. Alsabti and P. Murdin (Springer, March 2017) pp. 1185–1209.
4. A. Nonaka, A. J. Aspden, M. Zingale, A. S. Almgren, J. B. Bell and S. E. Woosley, High-resolution simulations of convection preceding ignition in Type Ia supernovae using adaptive mesh refinement, *ApJ* **745**, p. 73 (January 2012).
5. M. Fink, M. Kromer, I. R. Seitenzahl, F. Ciaraldi-Schoolmann, F. K. Röpke, S. A. Sim, R. Pakmor, A. J. Ruiter and W. Hillebrandt, Three-dimensional pure deflagration models with nucleosynthesis and synthetic observables for Type Ia supernovae, *MNRAS* **438**, 1762 (February 2014).
6. C. Byrohl, R. Fisher and D. Townsley, The intrinsic stochasticity of the ^{56}ni distribution of single-degenerate near-Chandrasekhar-mass SN Ia, *ApJ* **878**, p. 67 (June 2019).
7. I. R. Seitenzahl, F. Ciaraldi-Schoolmann, F. K. Röpke, M. Fink, W. Hillebrandt, M. Kromer, R. Pakmor, A. J. Ruiter, S. A. Sim and S. Taubenberger, Three-dimensional delayed-detonation models with nucleosynthesis for Type Ia supernovae, *MNRAS* **429**, 1156 (February 2013).

8. S. W. Jha, *Type Iax Supernovae*, in *Handbook of Supernovae*, eds. A. W. Alsabti and P. Murdin 2017, p. 375.
9. F. Lach, F. P. Callan, D. Bubeck, F. K. Roepke, S. A. Sim, M. Schrauth, S. T. Ohlmann and M. Kromer, Type iax supernovae from deflagrations in Chandrasekhar-mass white dwarfs, *arXiv e-prints*, arXiv:2109.02926 (September 2021).
10. B. Barna, T. Szalai, W. E. Kerzendorf, M. Kromer, S. A. Sim, M. R. Magee and B. Leibundgut, Type Iax supernovae as a few-parameter family, *MNRAS* **480**, 3609 (November 2018).
11. M. R. Magee, J. H. Gillanders, K. Maguire, S. A. Sim and F. P. Callan, An analysis of the spectroscopic signatures of layering in the ejecta of type Iax supernovae, *arXiv e-prints*, arXiv:2110.12294 (October 2021).
12. S. A. Sim, F. K. Röpke, W. Hillebrandt, M. Kromer, R. Pakmor, M. Fink, A. J. Ruiter and I. R. Seitenzahl, Detonations in sub-Chandrasekhar-mass C+O white dwarfs, *ApJ* **714**, L52 (May 2010).
13. M. Fink, F. K. Röpke, W. Hillebrandt, I. R. Seitenzahl, S. A. Sim and M. Kromer, Double-detonation sub-Chandrasekhar supernovae: can minimum helium shell masses detonate the core?, *A&A* **514**, p. A53 (May 2010).
14. M. Kromer, S. A. Sim, M. Fink, F. K. Röpke, I. R. Seitenzahl and W. Hillebrandt, Double-detonation Sub-Chandrasekhar Supernovae: Synthetic Observables for Minimum Helium Shell Mass Models, *ApJ* **719**, 1067 (August 2010).
15. D. M. Townsley, B. J. Miles, K. J. Shen and D. Kasen, Double Detonations with Thin, Modestly Enriched Helium Layers can Make Normal Type Ia Supernovae, *ApJ* **878**, p. L38 (June 2019).
16. K. J. Shen, S. Blondin, D. Kasen, L. Dessart, D. M. Townsley, S. Boos and D. J. Hillier, Non-local Thermodynamic Equilibrium Radiative Transfer Simulations of Sub-Chandrasekhar-mass White Dwarf Detonations, *ApJ* **909**, p. L18 (March 2021).
17. S. Gronow, C. Collins, S. T. Ohlmann, R. Pakmor, M. Kromer, I. R. Seitenzahl, S. A. Sim and F. K. Röpke, SNe Ia from double detonations: Impact of core-shell mixing on the carbon ignition mechanism, *A&A* **635**, p. A169 (March 2020).
18. S. Gronow, C. E. Collins, S. A. Sim and F. K. Röpke, Double detonations of sub-M_{Ch} CO white dwarfs: Variations in Type Ia supernovae due to different core and He shell masses, *A&A* **649**, p. A155 (May 2021).
19. S. Gronow, B. Côté, F. Lach, I. R. Seitenzahl, C. E. Collins, S. A. Sim and F. K. Roepke, Metallicity-dependent nucleosynthetic yields of Type Ia supernovae originating from double detonations of sub-M_{Ch} white dwarfs, *arXiv e-prints*, arXiv:2103.14050 (March 2021), submitted to *A&A*.
20. F. Lach, F. K. Roepke, I. R. Seitenzahl, B. Coté, S. Gronow and A. J. Ruiter, Nucleosynthesis imprints from different Type Ia Supernova explosion scenarios and implications for galactic chemical evolution, *A&A* **644**, p. A118 (December 2020).
21. I. R. Seitenzahl, G. Cescutti, F. K. Röpke, A. J. Ruiter and R. Pakmor, Solar abundance of manganese: A case for near chandrasekhar-mass type ia supernova progenitors, *A&A* **559**, p. L5 (November 2013).
22. I. R. Seitenzahl, F. K. Röpke, M. Fink and R. Pakmor, Nucleosynthesis in thermonuclear supernovae with tracers: Convergence and variable mass particles, *MNRAS* **407**, 2297 (October 2010).
23. R. Pakmor, M. Kromer, F. K. Röpke, S. A. Sim, A. J. Ruiter and W. Hillebrandt, Sub-luminous type ia supernovae from the mergers of equal-mass white dwarfs with mass $\sim 0.9 m_{\odot}$, *Nature* **463**, 61 (January 2010).
24. R. Pakmor, M. Kromer, S. Taubenberger, S. A. Sim, F. K. Röpke and W. Hillebrandt, Normal Type Ia supernovae from violent mergers of white dwarf binaries, *ApJ* **747**, p. L10 (March 2012).

25. R. Pakmor, M. Kromer, S. Taubenberger and V. Springel, Helium-ignited violent mergers as a unified model for normal and rapidly declining type Ia supernovae, *ApJ* **770**, p. L8 (June 2013).
26. K. J. Shen, S. J. Boos, D. M. Townsley and D. Kasen, Multi-Dimensional Radiative Transfer Calculations of Double Detonations of Sub-Chandrasekhar-Mass White Dwarfs, *arXiv e-prints*, arXiv:2108.12435 (August 2021).
27. M. Bulla, S. A. Sim, R. Pakmor, M. Kromer, S. Taubenberger, F. K. Röpke, W. Hillebrandt and I. R. Seitenzahl, Type Ia supernovae from violent mergers of carbon-oxygen white dwarfs: Polarization signatures, *MNRAS* **455**, 1060 (January 2016).
28. M. Bulla, S. A. Sim, M. Kromer, I. R. Seitenzahl, M. Fink, F. Ciaraldi-Schoolmann, F. K. Röpke, W. Hillebrandt, R. Pakmor, A. J. Ruiter and S. Taubenberger, Predicting polarization signatures for double-detonation and delayed-detonation models of Type Ia supernovae, *MNRAS* **462**, 1039 (October 2016).
29. M. Bulla, Z. W. Liu, F. K. Röpke, S. A. Sim, M. Fink, M. Kromer, R. Pakmor and I. R. Seitenzahl, White dwarf deflagrations for Type Iax supernovae: Polarisation signatures from the explosion and companion interaction, *A&A* **635**, p. A179 (March 2020).
30. K. J. Shen, D. Boubert, B. T. Gänsicke, S. W. Jha, J. E. Andrews, L. Chomiuk, R. J. Foley, M. Fraser, M. Gromadzki, J. Guillochon, M. M. Kotze, K. Maguire, M. R. Siebert, N. Smith, J. Strader, C. Badenes, W. E. Kerzendorf, D. Koester, M. Kromer, B. Miles, R. Pakmor, J. Schwab, O. Toloza, S. Toonen, D. M. Townsley and B. J. Williams, Three Hypervelocity White Dwarfs in Gaia DR2: Evidence for Dynamically Driven Double-degenerate Double-detonation Type Ia Supernovae, *ApJ* **865**, p. 15 (September 2018).
31. G. Ferrand, D. C. Warren, M. Ono, S. Nagataki, F. K. Röpke and I. R. Seitenzahl, From Supernova to Supernova Remnant: The Three-dimensional Imprint of a Thermonuclear Explosion, *ApJ* **877**, p. 136 (Jun 2019).
32. G. Ferrand, D. C. Warren, M. Ono, S. Nagataki, F. K. Röpke, I. R. Seitenzahl, F. Lach, H. Iwasaki and T. Sato, From Supernova to Supernova Remnant: Comparison of Thermonuclear Explosion Models, *ApJ* **906**, p. 93 (January 2021).
33. Jülich Supercomputing Centre. (2019). JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre. *Journal of Large-Scale Research Facilities* **5**, A171.

Type Ia supernovae and their explosive nucleosynthesis: Constraints on progenitors

Shing-Chi Leung
Department of Mathematics and Physics, SUNY Polytechnic Institute, 100 Seymour Road, Utica, New York 13502, USA
TAPIR, Mailcode 350-17, California Institute of Technology, Pasadena, CA 91125, USA
E-mail: leungs@sunypoly.edu
https://sunypoly.edu/faculty-and-staff/shing-chi-leung.html

Ken'ichi Nomoto
Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
E-mail: nomoto@astron.s.u-tokyo.ac.jp

What the progenitors of Type Ia supernovae (SNe Ia) are, whether they are near-Chandrasekhar mass or sub-Chandrasekhar mass white dwarfs, has been the matter of debate for decades. Various observational hints are supporting both models as the main progenitors. In this paper, we review the explosion physics and the chemical abundance patterns of SNe Ia from these two classes of progenitors. We will discuss how the observational data of SNe Ia, their remnants, the Milky Way Galaxy, and galactic clusters can help us to determine the essential features where numerical models of SNe Ia need to match.

Keywords: Supernova; Hydrodynamics; Nucleosynthesis; Supernova Remnant; Galactic Chemical Evolution

1. Introduction

Type Ia supernovae (SNe Ia) are well-understood as the thermonuclear explosions of carbon-oxygen white dwarfs (CO WDs).[1–3] They produce the majority of iron-peak elements in the galaxy, in particular ^{55}Mn. Their light curves can be standardized for measuring distance in the cosmological scale.[4,5] Understanding their progenitors, the explosion mechanisms and their obseravables are important for understanding the Universe in the larger scale.[6,7] In this review paper, we will explore possible progenitors of SNe Ia, whether they are the explosions of near-Chandrasekhar mass (Ch-mass) WDs or sub-Chandrasekhar mass (subCh-mass) WDs. In Table 1 we tabulate the important features to contrast between the Ch-mass and subCh-mass WDs.

The rise of the two classes of models comes from the diversity of observed SNe Ia. In the literature, a number of explosion models have been proposed to explain the normal and peculiar SNe Ia. For the Ch-mass WD, representative models include the pure turbulent deflagration model (PTD),[8–15] PTD with deflagration-detonation transition,[16–22] gravitationally confined detonation model[23–27] and pulsation reverse

detonation models.[28,29] The subCh-mass WD models include the double-detonation model,[30–38] violent merger model[39–42] and WD head-on collision model.[43–45] On top of these, unconventional models such as magnetized WDs,[46] super-Chandrasekhar mass WDs,[47] differentially rotating WDs[48,49] and interaction with dark matter gravity[50,51] have been proposed to explain some unusual SNe Ia.

Table 1. Comparing essential features of Ch-mass and subCh-mass WDs.

	unit	Ch-mass WD	subCh-mass WD
mass	$M_\odot$	$1.30 - \geqslant 1.38$	0.9 – 1.2
central density	g cm^{-3}	$10^9 - 10^{10}$	10^7 – 10^8
composition		^{12}C+^{16}O+^{22}Ne	core: ^{12}C+^{16}O+^{22}Ne envelope (env): ^{4}He
reaction		subsonic deflagration	supersonic detonation
first site		(near-)center	off-center (He-env)

The study of SNe Ia as explosions of (sub)Ch-mass WDs is often linked to the open question about the progenitors of SNe Ia: the single degenerate (SD) vs. the double degenerate (DD) scenario. The SD scenario means that the primary WD develops its nuclear runaway by mass accretion from its companion star, which can be a slightly evolved main-sequence, a red-giant, or a He-star.[49,52] The DD scenario means that the primary WD triggers the runaway by dynamical interaction with its companion WD.

We remind that the question on whether SNe Ia develop from Ch-mass WDs is not equivalent to arguing SNe Ia mainly develop in the SD scenario. For example, in the SD scenario, when the WD explodes as an SN Ia depends on the mass accretion rate from its companion star and the WD initial mass (see the left panel of Figure 1). A WD having (1) a high mass accretion rate above $\sim 10^{-9}$ $M_\odot$ yr^{-1} or (2) having a low mass accretion rate and a high initial mass > 1.1 $M_\odot$ is likely to develop nuclear runaway in the Ch-mass limit. Otherwise, the WD is more likely to explode as a subCh-mass WD.[53] Similar features have been seen also for WDs in the DD scenario.

To understand why the C-deflagration is associated with the Chandrasekhar mass WD, we show in the right panel of Figure 1 the relative pressure change of the CO-rich matter as a function of the matter density. During the thermonuclear runaway, ^{12}C and ^{16}O burn to form iron-peak elements peaked at ^{56}Ni, releasing an amount of $\sim 10^{18}$ erg g^{-1}. When the density is high ($\sim 10^9$ g cm^{-3}), the electron degeneracy pressure dominates the matter pressure, and the overall pressure becomes insensitive to its temperature. As a result, the relative pressure jump decreases as the matter becomes more degenerate. Without an abrupt pressure jump, the nuclear runaway in the Ch-mass WD may not spontaneously trigger a shock wave and hence no detonation may form. The hot matter may ignite ^{12}C in the nearby cold matter only by thermal conduction.

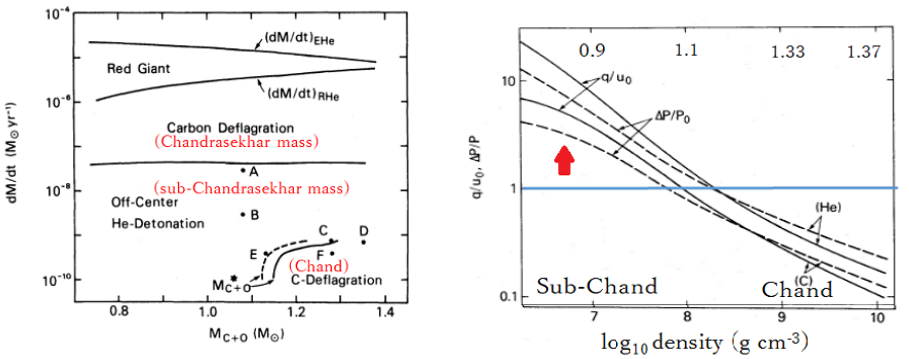


Fig. 1. (left panel) The final fate of the WD in the SD scenario with the mass accretion rate and the initial CO WD mass as parameters (derived and edited from Ref. 53). (right panel) The relative pressure change $\Delta P/P_0$ and relative internal energy change q/u_0 before and after nuclear runaway as a function of the matter density for the He-rich (solid line) and CO-rich (dashed line) matter (Ref. 30). The numbers on the top corresponds to the mass of the WD when the density corresponds to the central density of the WD. The red arrow indicates the relative pressure change of the CO-rich matter.

Unlike the detonation, the subsonic deflagration is subject to hydrodynamical instabilities such as the Rayleigh-Taylor (RT), Kelvin-Helmholtz (KH) and Landau-Derrieus[54–56] instabilities. The analytic model suggests that the buoyancy force can drive the early flame away from the center.[57] In Figure 2 we plot the electron fraction Y_e profile of a canonical PTD model where the deflagration has quenched after the expansion of the WD. The Y_e profile is a useful scalar for tracking how the fluid elements move inside the star. We observe the elongated "mushroom" shape as features of the RT-instabilities and the spiral along and inside the "mushrooms" as features of the KH-instabilities.

However, a WD may not naturally explode if there is only a slow subsonic nuclear flame because the WD expands and quenches the flame before the whole WD is burnt.[8,12] To alleviate this issue, a deflagration-detonation transition[16] and a flame-acceleration scheme[9,58,59] have been proposed for assisting nuclear burning to spread around the entire WD before the WD expands.

2. Typical Type Ia Supernova Explosion

Both the Ch-mass and subCh-mass WD models have their individual strengths and concerns, despite both of them can reproduce the observed features of normal SNe Ia,[60–62] including the Philip's relation.[37,63] For example, the Ch-mass model can produce Mn with an amount consistent with the solar abundance,[64] while the subCh-mass models do not produce a significant amount of Mn. But the DDT mechanism remains a matter of debate whether or not the turbulence is sufficient to pre-condition the CO rich matter.[57,65–69]

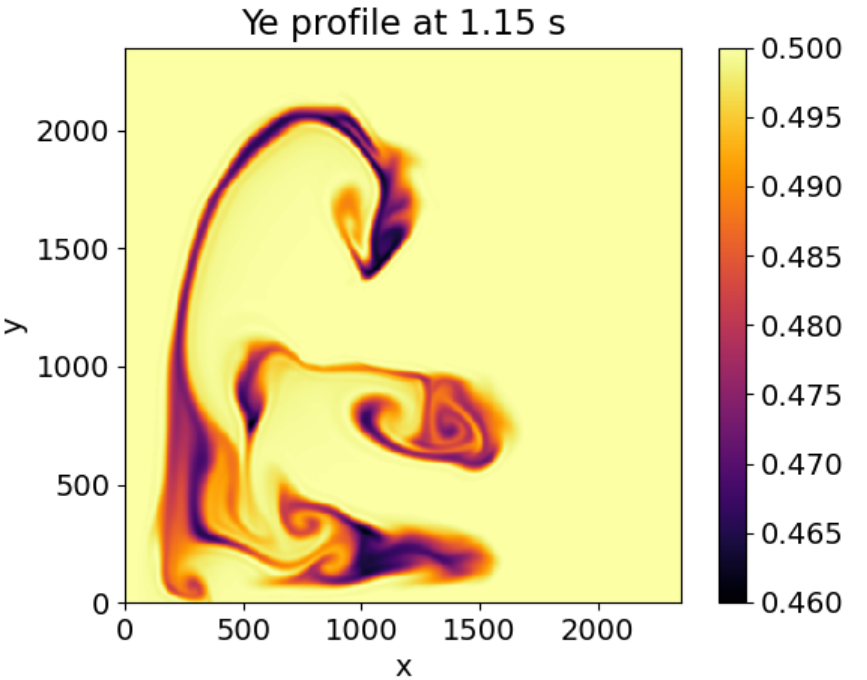


Fig. 2. A snapshot of the electron fraction Y_e profile in a typical PTD model demonstrating simultaneously the Rayleigh-Taylor and Kelvin-Helmholtz instabilities due to interaction of turbulent fluid motion with the deflagration front.

2.1. *Typical Explosion Mechanism of Ch-mass and subCh-mass Models*

We now examine the typical explosion mechanism in both the Ch-mass and subCh-mass WDs. Even though we have described a number of explosion mechanisms in the previous section, in general they are only different by the progenitor or the initial explosion kernel. The underlying mechanism, namely the deflagration and detonation, remains unchanged. Here we examine how the WD explodes accordingly.

In Figures 3 and 4 we plot the temperature profiles of the representative Ch-mass WD explosion using the PTD model with DDT for a WD of 1.37 $M_\odot$, metallicity $Z = 0.02$ and a $c3$ deflagration kernel[22] based on two-dimensional simulations.[70] The WD is burnt by subsonic flame for around 1 s, consuming about $\sim 30\%$ of the CO-rich matter in mass. After that, DDT is assumed to take place and the remaining matter is burnt within ~ 0.1 s. Eventually, the WD undergoes homologous expansion which quenches both deflagration and detonation.

In Figures 5 and 6 we plot similar profiles to Figure 3-4 but for the subCh-mass model with the initial mass 1.10 $M_\odot$, $Z = 0.02$ and a single He-detonation bubble.[35] In the first 1 s, the detonation burns the He-rich matter along the envelope.

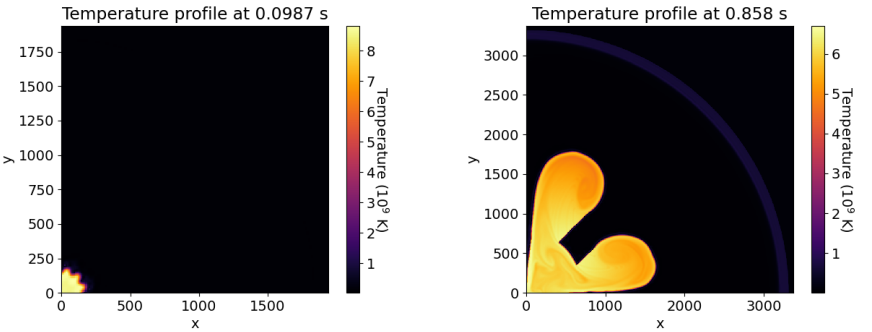


Fig. 3. (left panel) The initial temperature profile of the quadrant cross-section in a typical Ch-mass model using the PTD model with DDT for an initial mass $M = 1.37\ M_{\odot}$, metallicity $Z = 0.02$, and a "three-finger" initial flame kernel.[22] (top right panel) Same as the top left panel when the DDT is assumed to be triggered.

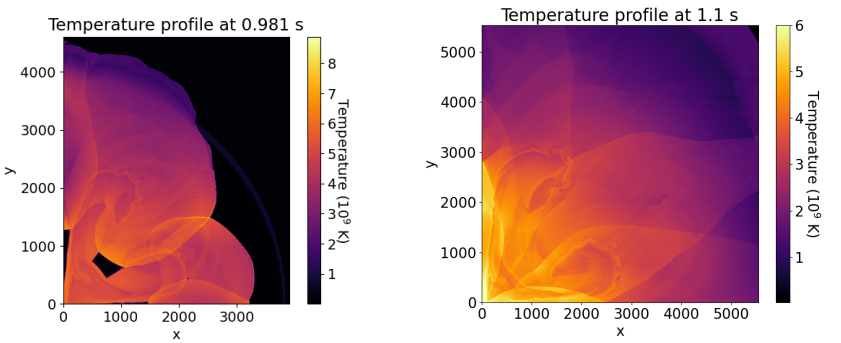


Fig. 4. (left panel) Same as Figure 3 but during the detonation phase. (right panel) Near complete disruption of the WD.

The detonation strength increases during the collision, which creates a shock that penetrates into the CO-core. This creates the C-detonation which later disrupts the entire WD.

2.2. *General Thermodynamical Features*

Typical multi-dimensional SN Ia simulations solve the Eulerian hydrodynamics equations with a simplified nuclear reaction network. To obtain the detailed chemical

Table 2. Major isotopes of iron-peak elements and their corresponding electron fraction.

Isotope	^{54}Fe	^{55}Mn	^{55}Fe	^{55}Fe	^{56}Fe	^{56}Co	^{56}Ni	^{57}Fe	^{58}Ni	^{60}Ni
Ye	0.481	0.454	0.472	0.490	0.464	0.482	0.500	0.456	0.483	0.467

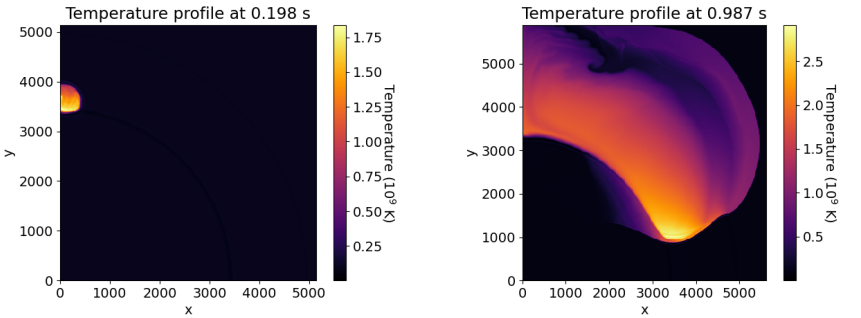


Fig. 5. (left panel) The initial temperature profile of a typical subCh-mass model using the double detonation model with the initial mass $M = 1.10\ M_\odot$, $Z = 0.02$, and a "single bubble" initial detonation kernel.[35] (right panel) Same as the top left panel but during the amplification of the He detonation.

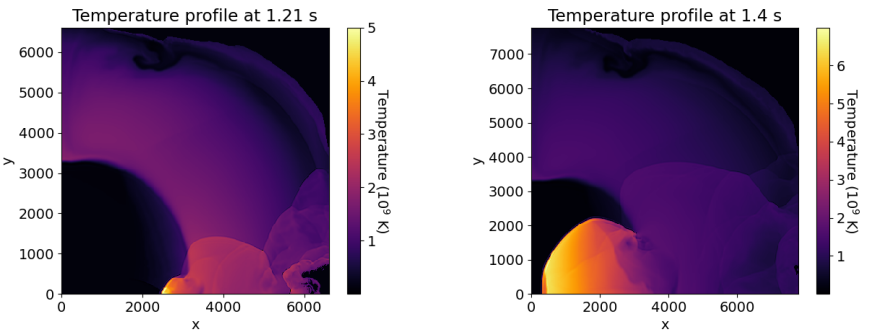


Fig. 6. (left panel) Same as Figure 5 but during the onset of the C-detonation. (right panel) Same as the left panel but during the C-detonation phase.

features of the explosion, a passive tracer particle scheme[71–74] is necessary. This scheme allocates a number of Lagrangian tracers to follow the fluid motion. The notation "passive" means that the tracers do not affect the fluid motion; they only record the thermodynamical condition along their trajectories.

The tracer particles record $(\rho(t), T(t))$ as a Lagrangian fluid packet along its path for reconstructing the exact chemical abundances. For SNe Ia, the trajectory is less convoluted that its peak density and temperature $(\rho_{\rm peak}, T_{\rm peak})$ can characterize the typical nucleosynthesis features inside the tracer. We make numerical experiments to show how various nucleosynthesis quantities depend on the parameters $(\rho_{\rm peak}, T_{\rm peak})$ parameter space.

We assume that the tracers start from given $(\rho_{\rm peak}, T_{\rm peak})$ and then adiabatically expand. The expansion timescale is chosen according to the typical explosion energy 10^{51} erg. The nuclear reactions are computed using the 495-isotope network.[75]

In Figures 7 and 8 we plot the final mean atomic number $\bar{A}$, $Y_{\rm e}$, asymptotic mass fraction of ^{55}Mn and ^{56}Fe for tracers under different initial conditions. The region is

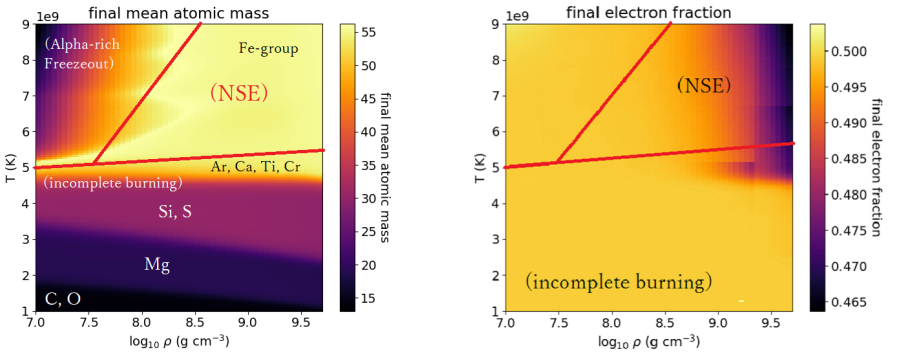


Fig. 7. (left panel) The final mean atomic mass number $\bar{A}$ of the tracer particles starting from different $\rho_{\rm peak}$ and $T_{\rm peak}$ (in units of 10^9 K). (right panel) Same as the left panel, but for the final electron fraction $Y_{\rm e}$ of the tracer.

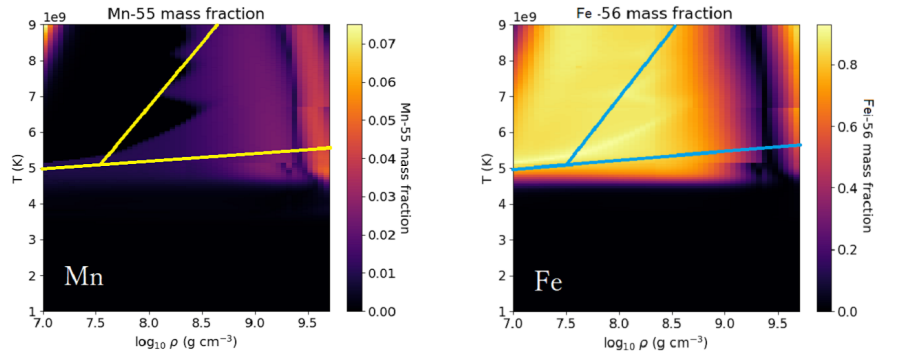


Fig. 8. (left panel) Same as Figure 7, but for the final mass fraction of stable ^{55}Mn. (right panel) Same as the left panel but for the final mass fraction of stable ^{56}Fe.

divided into three regions.[76,77] The low-$\rho_{\rm peak}$ region corresponds to the incomplete Si-burning regime, where the nuclear reaction terminates before reaching Fe-group elements, such as Si, S, Ar and so on. The high-$T_{\rm peak}$ (in units of 10^9 K) and low-$\rho_{\rm peak}$ region corresponds to the α-rich freezeout regime. As the name suggests, the nuclear reaction is confined to be along the α-chain from ^{12}C to ^{56}Ni. The high-$T_{\rm peak}$ and high-$\rho_{\rm peak}$ region corresponds to the nuclear statistical equilibrium (NSE) regime. This regime plays an important role in the Ch-mass WD as it allows isotopes away from the α-chain to form through weak interaction (electron capture).

As the $Y_{\rm e}$-profile indicates, the NSE zone is also the region where matter with $Y_{\rm e} < 0.5$ can be formed. The low $Y_{\rm e}$ environment is vital for forming the parents of ^{55}Mn (see Table 2 for the representative $Y_{\rm e}$ for the major neutron-rich isotopes of iron-peak elements). The ^{55}Mn profile also shows that the NSE zone is the primary site for generating a significant amount of stable ^{55}Mn after decay. On the other hand, ^{56}Fe is mostly formed in the α-rich freezeout and NSE ($Y_{\rm e} \approx 0.5$) regions.

2.3. *Thermodynamical Trajectories of SN Ia Models*

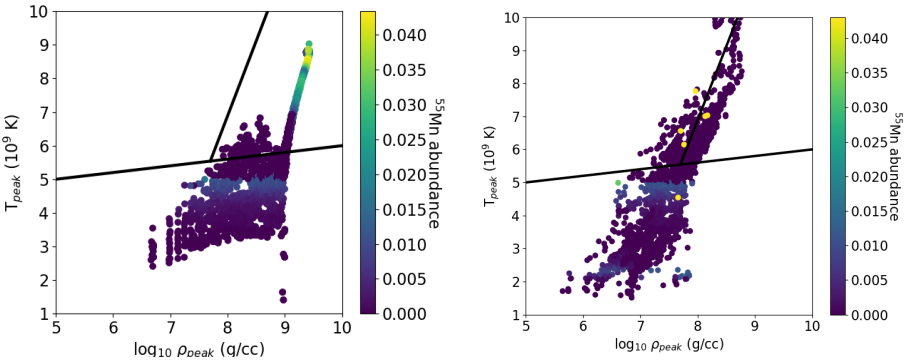


Fig. 9. (left panel) The thermodynamical trajectories of tracer particles of the Ch-mass model with the colour being the asymptotic ^{55}Mn abundance. Same as the left panel but for the subCh-mass model.

Having explored which thermodynamical parameter space is responsible for iron-peak elements, we show in Figure 9 the thermodynamics trajectories of tracers obtained from the typical Ch-mass and subCh-mass models. The chemical abundance of each tracer is directly computed according to its individual (ρ, T) time evolution.

The Ch-mass model (left panel) has two distinctive parts: the high density thin tail and the thick body at low density. At high density ($\rho_{\rm peak} > 10^9$ g cm^{-3}), the tracers are in the NSE regime and have a significantly higher ^{55}Mn and low fluctuations in $T_{\rm peak}$ for the same $\rho_{\rm peak}$. These are the tracers burnt by the subsonic deflagration. The absence of shock ensures that nuclear burning does not generate strong acoustic waves. On the other hand, the majority of tracers burnt by the detonation undergo incomplete Si-burning. The aspherical explosion allows tracers with the same initial mass coordinate to be burnt at a range of time. This leads to a wide temperature range for the same $\rho_{\rm peak}$. There is also a narrow band of tracers for $7 < \log_{10} \rho_{\rm peak} < 9$ and $T_{\rm peak} \approx 5 \times 10^9$ K also responsible for synthesizing a small fraction of ^{55}Mn.

The subCh-mass model (right panel) has a uniform structure where the $T_{\rm peak}$ scales with $\rho_{\rm peak}$ with some fluctuations. Only a small part of tracers reaches the NSE regime but their density is not high enough for the ^{55}Mn synthesis. There is also a narrow band of tracers containing ^{55}Mn by the synthesis of ^{55}Co. In general the global ^{55}Mn in the subCh-mass model is lower than that of the Ch-mass model.

2.4. *Typical Nucleosynthesis in Ch-mass and subCh-mass Models*

Now we have examined the thermodynamical differences between the Ch-mass and subCh-mass WDs. In Figure 10 we compare the qualitative differences in the nucleosynthesis pattern.

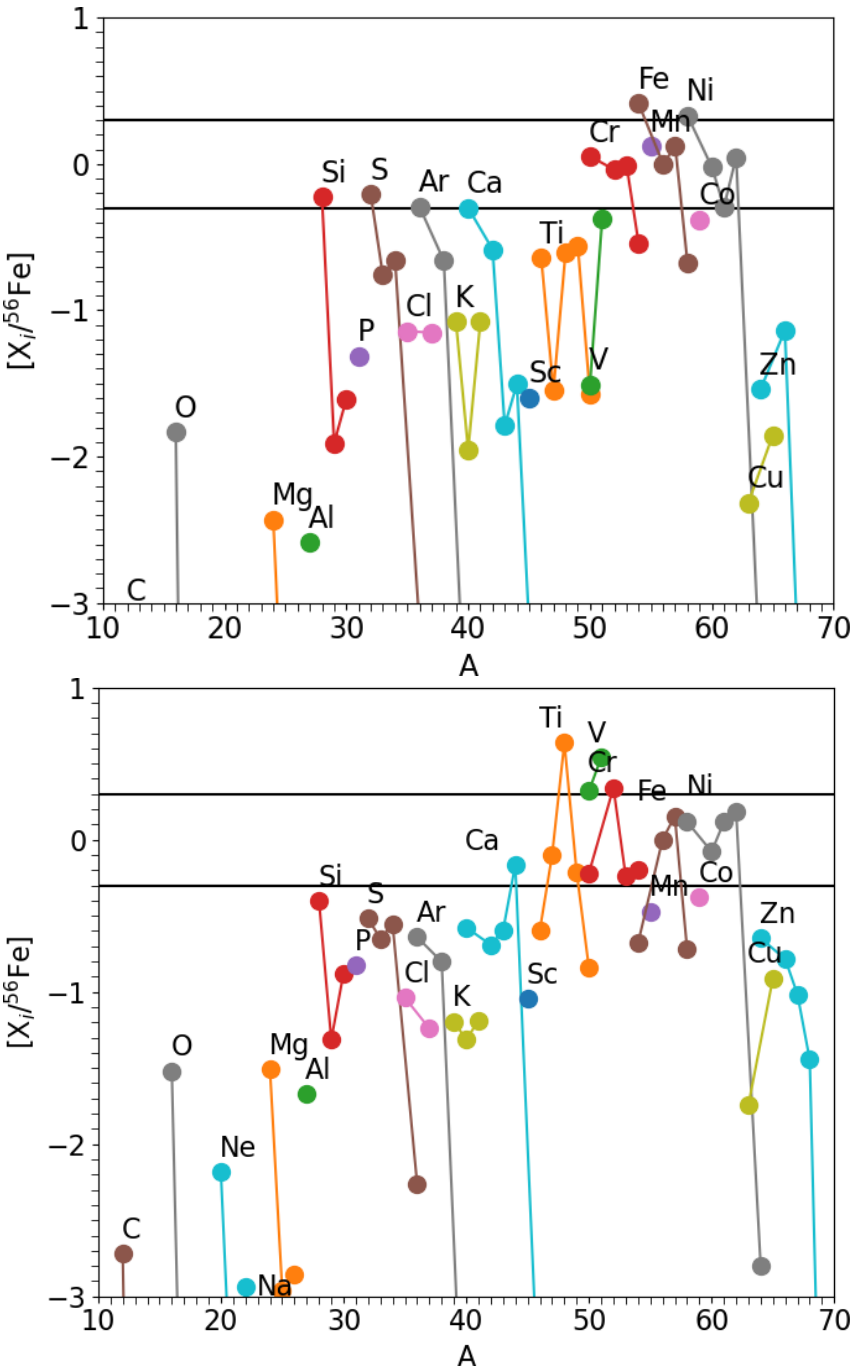


Fig. 10. (top panel) The final chemical abundance pattern of the typical Ch-mass WD[22] assuming the aspherical explosion. (bottom panel) Same as the top panel but for the typical subCh-mass WD.[35] $[X_i/^{56}\mathrm{Fe}] = \log_{10}[(X_i/^{56}\mathrm{Fe})\ /\ (X_i/^{56}\mathrm{Fe})_{\odot}]$. The two horizontal lines correspond to 50% and 200% of the solar value.

Both Ch-mass and subCh-mass WDs share some common features. They are responsible for the production of intermediate mass elements (IMEs) from Si to Ca, and the iron-peak elements from Ti to Ni. Odd number elements of IMEs are underproduced in SNe Ia. Some individual features allow us to distinguish the two models. (1) The aspherical explosion of the subCh-mass model can lead to signatures of strong Ti, V and Cr. (2) Mn is well-produced in the Ch-mass model but not in the subCh-mass model.

3. Applications of Nucleosynthesis

We have surveyed the major differences of the nucleosynthetic signature between the Ch-mass and subCh-mass WDs. Comparisons with observational data allow us to understand the progenitors of observed SNe Ia, which directly constrains the modeling. We can compare the optical signatures directly (i.e., light curves and spectra) by matching the radiative transfer model with SN Ia data.[41,60] One can also extract the chemical abundances from the spectra, and compare with nucleosynthetic results.[62,78] We shall focus on the latter method here.

3.1. *Supernova Remnant Sagittarius A East*

Within thousand years after the SN explosion, the shock-heated gas remains observable in the X-ray band, where the spectra reveal the metal composition inside the ejecta. Such a technique has been applied to the study of galactic supernova remnants (SNRs) including Tycho,[80] Kepler[81] and N103B.[82]

In Ref. 79 the SNR in Sagittarius A (Sgr A) East (G0.0+0.0) is observed based on the X-ray data taken by the *Chandra* telescope. The observed abundance ratios relative to Fe (with respect to the solar ratios) [Xi/Fe] are shown in Figure 11. The SNR features sub-solar intermediate mass elements (IMEs) and slightly super-solar iron-peak elements (Cr, Mn, and Ni).

The sub-solar IMEs exclude the possibility of associating a core-collapse SN as the origin of this remnant. On the left panel, the abundances of two distinctive classes of models, the subCh-mass and Ch-mass DDT models are plotted. The model uncertainties are shown by the shaded area. The subCh-mass models clearly overproduce the IME. Among the Ch-mass DDT models, the model that produces enough Mn and Ni overproduces Cr and the IME. There is a model whose Cr and Ni are consistent with the data points and IMEs are marginal, but its Mn is too small.

The Ch-mass PTD model (i.e., no DDT) with the initial central density of $\sim 5 \times 10^9$ g cm^{-3} is shown to be compatible with the data (right panel of Figure 11). [Such a high central density is realized in the rotating WD model.[83]] Note that this Ch-mass PTD model can well-explain the observed features of SNe Iax. Thus this object is the first identified SN Iax in the Milky Way Galaxy observed as SNR. This example also shows how the abundance guides us to identify the explosion mechanism.

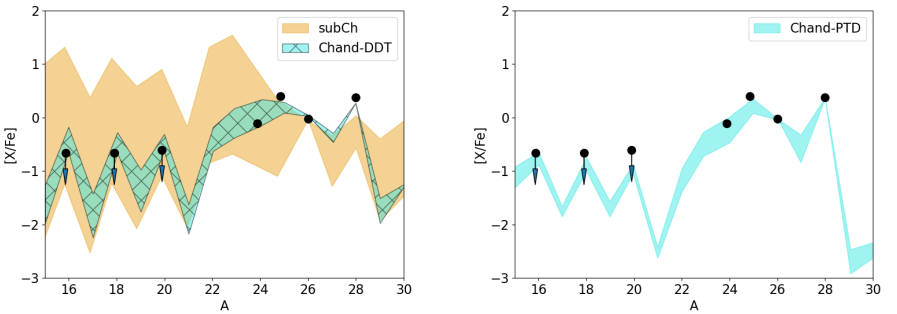


Fig. 11. (left panel) The chemical abundance pattern of the supernova remnant (SNR) Sagittarius A (from Ref. 79) for the data points compared with those of the subCh-mass[35] and Ch-mass DDT[22] models shown by the shaded regions. (right panel) Same as the left panel but for the Ch-mass PTD[15] models.

3.2. *Supernova Remnant 3C 397*

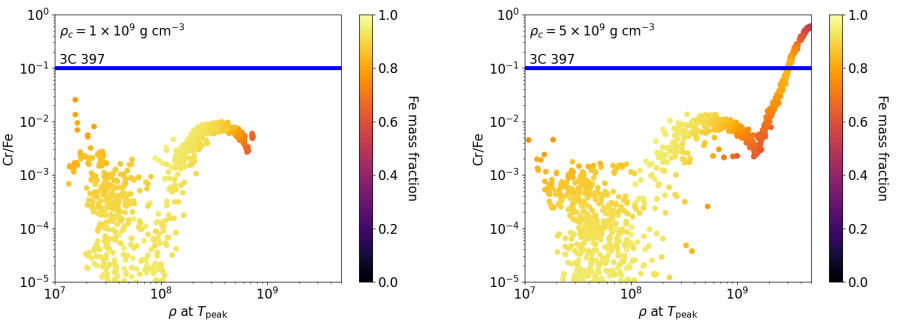


Fig. 12. (left panel) The Cr/Fe distribution of the tracers taken from the Ch-mass model with the initial central density 1×10^9 g cm^{-3}.[22] The color represents the tracer Fe mass fraction. The horizontal line is the measured value in SNR 3C 397 from Ref. 84. (right panel) Same as the left panel but for the model with the initial central density 5×10^9 g cm^{-3}.

The SNR 3C 397 is a nearby object (8 kpc) on the galactic plane. Its close distance allows astronomers to extract the spectra from individual parts similar to Sgr A. This object features a high Mn/Ni ratio, which is a key evidence of the Ch-mass explosion.[85]

In a recent observation using the *XMM-Newton* telescope, the spectra from the South and West hot blobs are measured, which give the constraint on the Cr/Fe mass ratio $\sim 0.106 \pm^{0.011}_{0.009}$.[84] The high value is used to distinguish the explosion progenitor shown in Figure 12. By comparing the tracer in different Ch-mass models, it becomes clear that the low-mass model ($\rho_c = 1 \times 10^9$ g cm^{-3}) does not have tracers reaching the observed high value. Meanwhile the high density tail in the high-mass model

($\rho_c = 5 \times 10^9$ g cm^{-3}) has tracers crossing the expected value. This provides a strong indication that this object is the explosion of the high-mass Ch-mass WD. This also demonstrates how a precise measurement of element abundance ratios can guide us to select the potential progenitor.

3.3. *Milky Way Galaxy*

In the last two sections we have shown how the SNR abundance determines its progenitor and the explosion mechanism. While there is no distinctive SNR showing chemical abundances exclusively for subCh-mass WD models, it is possible that a large sample size is needed to understand the distribution of each model. To understand the SN Ia explosion globally, we need the chemical abundances from a larger system, for example, the Milky Way Galaxy. The elements ejected by supernovae become the building block of the next-generation stars.[6] The surface abundance of stars in the solar neighbourhood may thus indicate how much each element is ejected by generations of SNe Ia.

In Ref. 86 the galactic chemical evolution model is computed with supernova abundance patterns taken from literatures. The Mn/Fe evolution is plotted in Figure 13. Two contrasting classes of models are shown, one assuming the pure Ch-mass WD explosion, and the other two assuming pure subCh-mass WD. To reproduce the trend as well as the magnitude of the data, a non-negligible fraction of the Ch-mass WD is necessary.

We remark that the supernova history can be strongly dependent on the galaxy evolution history. Some galaxies (e.g., Sculptor dwarf spheroidal galaxy) have a low Mn/Fe ratio that indicates the dominance of the subCh-mass WD explosion in their evolution histories.[90,91] Meanwhile, some early rise of [Mn/Fe] in this subclass of galaxies can be a result of the Ch-mass SN Iax explosion.[92]

3.4. *Perseus Galactic Cluster*

The Milky Way Galaxy can provide a detailed reference in how generations of stars contributes to the cosmic metal enrichment. However, large N-body simulations suggest that each galaxy is unique in their evolution history. To understand how each supernova model contributes in the cosmic scale, data from an even larger system is important to average out the statistical fluctuations of individual galaxies.

In Ref. 93 the X-ray spectra of the Perseus Cluster is studied by the *Hitomi* telescope. The highly resolved spectral lines provide the abundance measurement with uncertainties down to $\sim 10\%$. The high precision can distinguish supernova models and mechanisms explicitly. The fitting using SN Ia and CCSN models from literature is shown in Table 3. The best-fit model is found to be the scenario assuming pure Ch-mass WD explosion. If the fraction of the Ch-mass WD is relaxed as a model parameter, the expected Ch-mass WD still contributes about $10 - 40\%$ of the SN Ia population, depending on the exact CCSN models.

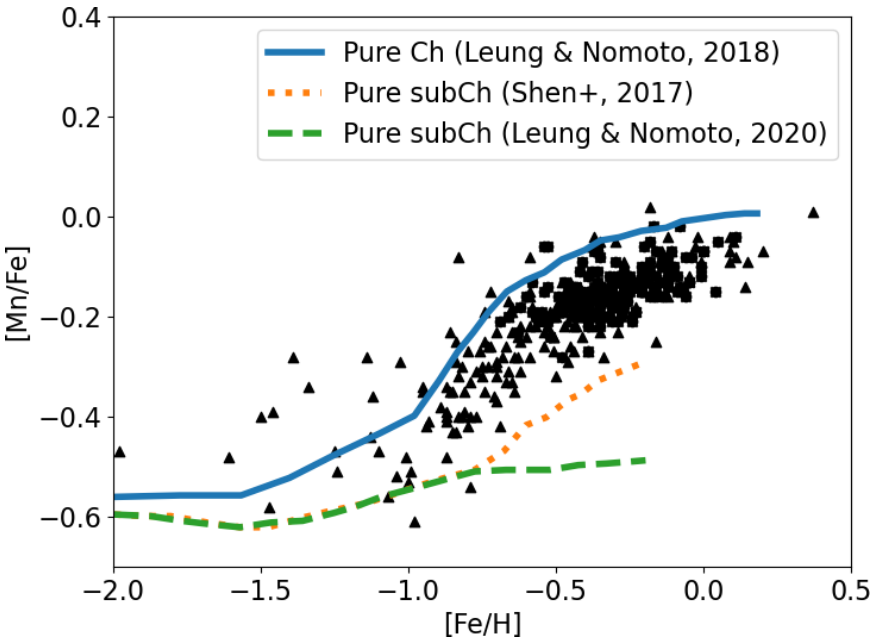


Fig. 13. The [Mn/Fe] against metallicity [Fe/H] for the galactic chemical evolution models taken from Ref. 86. Solid lines come from theoretical models assuming pure Ch-mass and subCh-mass explosion history. Data points are the stellar abundances from the solar neighbourhood.[87–89]

Table 3. Models assuming different stellar and supernova models and their corresponding (Ch-mass) SN Ia rates (data taken from [Ref. 93]).

Model	$f_{\rm Ia}$	$f_{\rm Chand}$	χ^2
pure Ch-mass[Ref. 22]+ CCSN[Ref. 94]	0.21 ± 0.02	N/A	11.78
Ch-mass[Ref. 95] + subCh-mass[Ref. 32] + CCSN[94]	0.25 ± 0.06	0.36 ± 0.14	23.96
Ch-mass[Ref. 95] + subCh-mass[Ref. 32] + CCSN[96]	0.38 ± 0.06	0.09 ± 0.09	15.73

4. Conclusion

In this review article we have presented the physical background about the Ch-mass and subCh-mass WD models as the SN Ia explosion progenitors. We discussed the differences in their explosion mechanisms and their associated nucleosynthetic signatures. We have also demonstrated how the chemical abundances of SNRs, Milky Way Galaxy, and galactic clusters can help us distinguish (1) the individual SN explosion scenario and (2) the relative importance of each explosion model.

Nucleosynthesis will remain an important subject in the future supernova study thanks to observational projects such as XRISM (X-Ray Imaging and Spectroscopy Mission). Given the power of resolving spectral lines as its predecessor *Hitomi*, we can anticipate that the high quality spectral data, and hence the precise chemical

abundance measurements, will shed light on supernova models to an unprecedented accuracy.

Acknowledgments

S.C.L. thanks the session chairpersons Pilar-Ruiz Lapuente and Robert Fisher for the invition to the introductory talk in the Marcel Grossmann 16 Meeting. S.C.L acknowledges support by NASA grants HST-AR-15021.001-A and 80NSSC18K1017. K.N. has been supported by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and JSPS KAKENHI Grant Numbers JP17K05382, JP20K04024, and JP21H04499.

References

1. W. D. Arnett, A Possible Model of Supernovae: Detonation of ^{12}C, *Astrophysics and Space Science* **5**, 180 (October 1969).
2. W. Hillebrandt and J. C. Niemeyer, Type IA Supernova Explosion Models, *Annual Review of Astronomy and Astrophysics* **38**, 191 (January 2000).
3. K. Nomoto and S.-C. Leung, *Thermonuclear Explosions of Chandrasekhar Mass White Dwarfs*, in *Handbook of Supernovae*, eds. A. W. Alsabti and P. Murdin 2017, p. 1275.
4. A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff and J. Tonry, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, *Astronomical Journal* **116**, 1009 (September 1998).
5. S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch and T. S. C. Project, Measurements of Ω and Λ from 42 High-Redshift Supernovae, *Astrophysical Journal* **517**, 565 (June 1999).
6. F. Matteucci, *The chemical evolution of the Galaxy* 2001.
7. C. Kobayashi, A. I. Karakas and H. Umeda, The evolution of isotope ratios in the Milky Way Galaxy, *Monthly Notices of the Royal Astronomical Society* **414**, 3231 (July 2011).
8. K. Nomoto, D. Sugimoto and S. Neo, Carbon Deflagration Supernova, an Alternative to Carbon Detonation, *Astrophysics and Space Science* **39**, p. L37 (February 1976).
9. K. Nomoto, F. K. Thielemann and K. Yokoi, Accreting white dwarf models for type I supern. III. Carbon deflagration supernovae., *Astrophysical Journal* **286**, 644 (November 1984).
10. E. Livne, Numerical Simulations of the Convective Flame in White Dwarfs, *Astrophysical Journal Letter* **406**, p. L17 (March 1993).
11. M. Reinecke, W. Hillebrandt and J. C. Niemeyer, Thermonuclear explosions of Chandrasekhar-mass C+O white dwarfs, *Astronomy and Astrophysics* **347**, 739 (July 1999).

12. M. Reinecke, W. Hillebrandt and J. C. Niemeyer, Refined numerical models for multidimensional type Ia supernova simulations, *Astronomy and Astrophysics* **386**, 936 (May 2002).
13. F. K. Röpke, W. Hillebrandt, W. Schmidt, J. C. Niemeyer, S. I. Blinnikov and P. A. Mazzali, A Three-Dimensional Deflagration Model for Type Ia Supernovae Compared with Observations, *Astrophysical Journal* **668**, 1132 (October 2007).
14. H. Ma, S. E. Woosley, C. M. Malone, A. Almgren and J. Bell, Carbon Deflagration in Type Ia Supernova. I. Centrally Ignited Models, *Astrophysical Journal* **771**, p. 58 (July 2013).
15. S.-C. Leung and K. Nomoto, Explosive Nucleosynthesis in Near-Chandrasekhar Mass White Dwarf Models for Type Iax Supernovae: Dependence on Model Parameters, *Astrophysical Journal* **900**, p. 54 (September 2020).
16. A. M. Khokhlov, Delayed detonation model for type IA supernovae, *Astronomy and Astrophysics* **245**, 114 (May 1991).
17. H. Yamaoka, K. Nomoto, T. Shigeyama and F.-K. Thielemann, Late Detonation Models for the Type IA Supernovae SN 1991T and SN 1990N, *Astrophysical Journal Letter* **393**, p. L55 (July 1992).
18. K. Iwamoto, F. Brachwitz, K. Nomoto, N. Kishimoto, H. Umeda, W. R. Hix and F.-K. Thielemann, Nucleosynthesis in Chandrasekhar Mass Models for Type IA Supernovae and Constraints on Progenitor Systems and Burning-Front Propagation, *Astrophysical Journal Supplementary* **125**, 439 (December 1999).
19. I. Golombek and J. C. Niemeyer, A model for multidimensional delayed detonations in SN Ia explosions, *Astronomy and Astrophysics* **438**, 611 (August 2005).
20. F. K. Röpke and J. C. Niemeyer, Delayed detonations in full-star models of type Ia supernova explosions, *Astronomy and Astrophysics* **464**, 683 (March 2007).
21. M. Fink, M. Kromer, I. R. Seitenzahl, F. Ciaraldi-Schoolmann, F. K. Röpke, S. A. Sim, R. Pakmor, A. J. Ruiter and W. Hillebrandt, Three-dimensional pure deflagration models with nucleosynthesis and synthetic observables for Type Ia supernovae, *Monthly Notices of the Royal Astronomical Society* **438**, 1762 (February 2014).
22. S.-C. Leung and K. Nomoto, Explosive Nucleosynthesis in Near-Chandrasekhar-mass White Dwarf Models for Type Ia Supernovae: Dependence on Model Parameters, *Astrophysical Journal* **861**, p. 143 (July 2018).
23. T. Plewa, A. C. Calder and D. Q. Lamb, Type Ia Supernova Explosion: Gravitationally Confined Detonation, *Astrophysical Journal Letter* **612**, L37 (September 2004).
24. C. A. Meakin, I. Seitenzahl, D. Townsley, I. Jordan, George C., J. Truran and D. Lamb, Study of the Detonation Phase in the Gravitationally Confined Detonation Model of Type Ia Supernovae, *Astrophysical Journal* **693**, 1188 (March 2009).
25. I. Jordan, G. C., C. Graziani, R. T. Fisher, D. M. Townsley, C. Meakin, K. Weide, L. B. Reid, J. Norris, R. Hudson and D. Q. Lamb, The Detonation Mechanism of the Pulsationally Assisted Gravitationally Confined Detonation Model of Type Ia Supernovae, *Astrophysical Journal* **759**, p. 53 (November 2012).
26. D. García-Senz, R. M. Cabezón, I. Domínguez and F. K. Thielemann, Type Ia Supernovae: Can Coriolis Force Break the Symmetry of the Gravitational Confined Detonation Explosion Mechanism?, *Astrophysical Journal* **819**, p. 132 (March 2016).
27. I. R. Seitenzahl, M. Kromer, S. T. Ohlmann, F. Ciaraldi-Schoolmann, K. Marquardt, M. Fink, W. Hillebrandt, R. Pakmor, F. K. Röpke, A. J. Ruiter, S. A. Sim and S. Taubenberger, Three-dimensional simulations of gravitationally confined detonations compared to observations of SN 1991T, *Astronomy and Astrophysics* **592**, p. A57 (July 2016).

28. E. Bravo, D. García-Senz, R. M. Cabezon and I. Domnguez, Pulsating Reverse Detonation Models of Type Ia Supernovae. I. Detonation Ignition, Astrophysical Journal 695, 1244 (April 2009).
29. E. Bravo, D. García-Senz, R. M. Cabezón and I. Domínguez, Pulsating Reverse Detonation Models of Type Ia Supernovae. II. Explosion, *Astrophysical Journal* **695**, 1257 (April 2009).
30. K. Nomoto, Accreting white dwarf models for type I supernovae. II. Off-center detonation supernovae, *Astrophysical Journal* **257**, 780 (June 1982).
31. M. Fink, F. K. Röpke, W. Hillebrandt, I. R. Seitenzahl, S. A. Sim and M. Kromer, Double-detonation sub-Chandrasekhar supernovae: Can minimum helium shell masses detonate the core?, *Astronomy and Astrophysics* **514**, p. A53 (May 2010).
32. K. J. Shen, D. Kasen, B. J. Miles and D. M. Townsley, Sub-Chandrasekhar-mass White Dwarf Detonations Revisited, *Astrophysical Journal* **854**, p. 52 (February 2018).
33. A. Tanikawa, K. Nomoto and N. Nakasato, Three-dimensional Simulation of Double Detonations in the Double-degenerate Model for Type Ia Supernovae and Interaction of Ejecta with a Surviving White Dwarf Companion, *Astrophysical Journal* **868**, p. 90 (December 2018).
34. A. Tanikawa, K. Nomoto, N. Nakasato and K. Maeda, Double-detonation Models for Type Ia Supernovae: Trigger of Detonation in Companion White Dwarfs and Signatures of Companions' Stripped-off Materials, *Astrophysical Journal* **885**, p. 103 (November 2019).
35. S.-C. Leung and K. Nomoto, Explosive Nucleosynthesis in Sub-Chandrasekhar-mass White Dwarf Models for Type Ia Supernovae: Dependence on Model Parameters, *Astrophysical Journal* **888**, p. 80 (January 2020).
36. A. Polin, P. Nugent and D. Kasen, Nebular Models of Sub-Chandrasekhar Mass Type Ia Supernovae: Clues to the Origin of Ca-rich Transients, *Astrophysical Journal* **906**, p. 65 (January 2021).
37. K. J. Shen, S. Blondin, D. Kasen, L. Dessart, D. M. Townsley, S. Boos and D. J. Hillier, Non-local Thermodynamic Equilibrium Radiative Transfer Simulations of Sub-Chandrasekhar-mass White Dwarf Detonations, *Astrophysical Journal Letter* **909**, p. L18 (March 2021).
38. S. Gronow, C. E. Collins, S. A. Sim and F. K. Röpke, Double detonations of sub-M_{Ch} CO white dwarfs: Variations in Type Ia supernovae due to different core and He shell masses, *Astronomy and Astrophysics* **649**, p. A155 (May 2021).
39. R. Pakmor, S. Hachinger, F. K. Röpke and W. Hillebrandt, Violent mergers of nearly equal-mass white dwarf as progenitors of subluminous Type Ia supernovae, *Astronomy and Astrophysics* **528**, p. A117 (April 2011).
40. R. Pakmor, M. Kromer, S. Taubenberger, S. A. Sim, F. K. Röpke and W. Hillebrandt, Normal Type Ia Supernovae from Violent Mergers of White Dwarf Binaries, *Astrophysical Journal Letter* **747**, p. L10 (March 2012).
41. M. Kromer, R. Pakmor, S. Taubenberger, G. Pignata, M. Fink, F. K. Röpke, I. R. Seitenzahl, S. A. Sim and W. Hillebrandt, SN 2010lp—A Type Ia Supernova from a Violent Merger of Two Carbon-Oxygen White Dwarfs, *Astrophysical Journal Letter* **778**, p. L18 (November 2013).
42. A. Tanikawa, N. Nakasato, Y. Sato, K. Nomoto, K. Maeda and I. Hachisu, Hydrodynamical Evolution of Merging Carbon-Oxygen White Dwarfs: Their Pre-supernova Structure and Observational Counterparts, *Astrophysical Journal* **807**, p. 40 (July 2015).
43. D. García-Senz, R. M. Cabezón, A. Arcones, A. Relaño and F. K. Thielemann, High-resolution simulations of the head-on collision of white dwarfs, *Monthly Notices of the Royal Astronomical Society* **436**, 3413 (December 2013).

44. D. Kushnir, B. Katz, S. Dong, E. Livne and R. Fernández, Head-on Collisions of White Dwarfs in Triple Systems Could Explain Type Ia Supernovae, *Astrophysical Journal Letter* **778**, p. L37 (December 2013).
45. O. Papish and H. B. Perets, Supernovae from Direct Collisions of White Dwarfs and the Role of Helium Shell Ignition, *Astrophysical Journal* **822**, p. 19 (May 2016).
46. U. Das and B. Mukhopadhyay, Maximum mass of stable magnetized highly super-Chandrasekhar white dwarfs: Stable solutions with varying magnetic fields, *Journal of Cosmology and Astroparticle Physics* **2014**, p. 050 (June 2014).
47. Y. Kamiya, M. Tanaka, K. Nomoto, S. I. Blinnikov, E. I. Sorokina and T. Suzuki, Super-Chandrasekhar-mass Light Curve Models for the Highly Luminous Type Ia Supernova 2009dc, *Astrophysical Journal* **756**, p. 191 (September 2012).
48. S. C. Yoon and N. Langer, On the evolution of rapidly rotating massive white dwarfs towards supernovae or collapses, *Astronomy and Astrophysics* **435**, 967 (June 2005).
49. I. Hachisu, M. Kato, H. Saio and K. Nomoto, A Single Degenerate Progenitor Model for Type Ia Supernovae Highly Exceeding the Chandrasekhar Mass Limit, *Astrophysical Journal* **744**, p. 69 (January 2012).
50. S. C. Leung, M. C. Chu and L. M. Lin, Dark Matter Admixed Type Ia Supernovae, *Astrophysical Journal* **812**, p. 110 (October 2015).
51. H.-S. Chan, M.-c. Chu, S.-C. Leung and L.-M. Lin, Delayed Detonation Thermonuclear Supernovae with an Extended Dark Matter Component, *Astrophysical Journal* **914**, p. 138 (June 2021).
52. K. Nomoto and S.-C. Leung, Single Degenerate Models for Type Ia Supernovae: Progenitor's Evolution and Nucleosynthesis Yields, *Space Science Review* **214**, p. 67 (June 2018).
53. K. Nomoto, Accreting white dwarf models for type I supernovae. I - Presupernova evolution and triggering mechanisms, *Astrophysical Journal* **253**, 798 (February 1982).
54. J. C. Niemeyer and W. Hillebrandt, Turbulent Nuclear Flames in Type IA Supernovae, *Astrophysical Journal* **452**, p. 769 (October 1995).
55. J. B. Bell, M. S. Day, C. A. Rendleman, S. E. Woosley and M. Zingale, Direct Numerical Simulations of Type Ia Supernovae Flames. I. The Landau-Darrieus Instability, *Astrophysical Journal* **606**, 1029 (May 2004).
56. J. B. Bell, M. S. Day, C. A. Rendleman, S. E. Woosley and M. Zingale, Direct Numerical Simulations of Type Ia Supernovae Flames. II. The Rayleigh-Taylor Instability, *Astrophysical Journal* **608**, 883 (June 2004).
57. R. Fisher and K. Jumper, Single-degenerate Type Ia Supernovae Are Preferentially Overluminous, *Astrophysical Journal* **805**, p. 150 (June 2015).
58. J. C. Niemeyer and W. Hillebrandt, Microscopic Instabilities of Nuclear Flames in Type IA Supernovae, *Astrophysical Journal* **452**, p. 779 (October 1995).
59. S. E. Woosley, Neutron-rich Nucleosynthesis in Carbon Deflagration Supernovae, *Astrophysical Journal* **476**, 801 (February 1997).
60. O. Graur, D. Zurek, M. M. Shara, A. G. Riess, I. R. Seitenzahl and A. Rest, Late-time Photometry of Type Ia Supernova SN 2012cg Reveals the Radioactive Decay of 57 Co, *Astrophysical Journal* **819**, p. 31 (March 2016).
61. B. J. Shappee, K. Z. Stanek, C. S. Kochanek and P. M. Garnavich, Whimper of a Bang: Documenting the Final Days of the Nearby Type Ia Supernova 2011fe, *Astrophysical Journal* **841**, p. 48 (May 2017).
62. K. Mori, M. A. Famiano, T. Kajino, T. Suzuki, P. M. Garnavich, G. J. Mathews, R. Diehl, S.-C. Leung and K. Nomoto, Nucleosynthesis Constraints on the Explosion Mechanism for Type Ia Supernovae, *Astrophysical Journal* **863**, p. 176 (August 2018).
63. D. Kasen, F. K. Röpke and S. E. Woosley, The diversity of type Ia supernovae from broken symmetries, *Nature* **460**, 869 (August 2009).

64. I. R. Seitenzahl, G. Cescutti, F. K. Röpke, A. J. Ruiter and R. Pakmor, Solar abundance of manganese: A case for near Chandrasekhar-mass Type Ia supernova progenitors, *Astronomy and Astrophysics* **559**, p. L5 (November 2013).
65. A. M. Lisewski, W. Hillebrandt and S. E. Woosley, Constraints on the Delayed Transition to Detonation in Type IA Supernovae, *Astrophysical Journal* **538**, 831 (August 2000).
66. F. K. Röpke, Flame-driven Deflagration-to-Detonation Transitions in Type Ia Supernovae?, *Astrophysical Journal* **668**, 1103 (October 2007).
67. S. E. Woosley, A. R. Kerstein, V. Sankaran, A. J. Aspden and F. K. Röpke, Type Ia Supernovae: Calculations of Turbulent Flames Using the Linear Eddy Model, *Astrophysical Journal* **704**, 255 (October 2009).
68. D. Fenn, T. Plewa and A. Gawryszczak, No double detonations but core carbon ignitions in high-resolution, grid-based simulations of binary white dwarf mergers, *Monthly Notices of the Royal Astronomical Society* **462**, 2486 (November 2016).
69. E. Brooker, T. Plewa and D. Fenn, Type Ia supernovae deflagration-to-detonation transition explosions powered by the Zel'dovich reactivity gradient mechanism, *Monthly Notices of the Royal Astronomical Society* **501**, L23 (January 2021).
70. S. C. Leung, M. C. Chu and L. M. Lin, A new hydrodynamics code for Type Ia supernovae, *Monthly Notices of the Royal Astronomical Society* **454**, 1238 (December 2015).
71. S. Nagataki, T. M. Shimizu and K. Sato, Matter Mixing from Axisymmetric Supernova Explosion, *Astrophysical Journal* **495**, 413 (March 1998).
72. C. Travaglio, W. Hillebrandt, M. Reinecke and F. K. Thielemann, Nucleosynthesis in multi-dimensional SN Ia explosions, *Astronomy and Astrophysics* **425**, 1029 (October 2004).
73. I. R. Seitenzahl, F. K. Röpke, M. Fink and R. Pakmor, Nucleosynthesis in thermonuclear supernovae with tracers: Convergence and variable mass particles, *Monthly Notices of the Royal Astronomical Society* **407**, 2297 (October 2010).
74. D. M. Townsley, B. J. Miles, F. X. Timmes, A. C. Calder and E. F. Brown, A Tracer Method for Computing Type Ia Supernova Yields: Burning Model Calibration, Reconstruction of Thickened Flames, and Verification for Planar Detonations, *Astrophysical Journal Supplementary* **225**, p. 3 (July 2016).
75. F. X. Timmes, Integration of Nuclear Reaction Networks for Stellar Hydrodynamics, *Astrophysical Journal Supplementary* **124**, 241 (September 1999).
76. F. K. Thielemann, K. Nomoto, and K. Yokoi, Explosive nucleosynthesis in carbon deflagration models of Type I supernovae, *Astronomy and Astrophysics* **158**, 17 (April 1986).
77. F. Lach, F. K. Röpke, I. R. Seitenzahl, B. Coté, S. Gronow and A. J. Ruiter, Nucleosynthesis imprints from different Type Ia supernova explosion scenarios and implications for galactic chemical evolution, *Astronomy and Astrophysics* **644**, p. A118 (December 2020).
78. S.-C. Leung, R. Diehl, K. Nomoto and T. Siegert, Exploration of Aspherical Ejecta Properties in Type Ia Supernovae: Progenitor Dependence and Applications to Progenitor Classification, *Astrophysical Journal* **909**, p. 152 (March 2021).
79. P. Zhou, S.-C. Leung, Z. Li, K. Nomoto, J. Vink and Y. Chen, Chemical Abundances in Sgr A East: Evidence for a Type Iax Supernova Remnant, *Astrophysical Journal* **908**, p. 31 (February 2021).
80. S. Park, C. Badenes, K. Mori, R. Kaida, E. Bravo, A. Schenck, K. A. Eriksen, J. P. Hughes, P. O. Slane, D. N. Burrows and J.-J. Lee, A Super-solar Metallicity for the Progenitor of Kepler's Supernova, *Astrophysical Journal Letter* **767**, p. L10 (April 2013).

81. H. Yamaguchi, J. P. Hughes, C. Badenes, E. Bravo, I. R. Seitenzahl, H. Martínez-Rodríguez, S. Park and R. Petre, The Origin of the Iron-rich Knot in Tycho's Supernova Remnant, *Astrophysical Journal* **834**, p. 124 (January 2017).
82. H. Yamaguchi, F. Acero, C.-J. Li and Y.-H. Chu, Discovery of Double-ring Structure in the Supernova Remnant N103B: Evidence for Bipolar Winds from a Type Ia Supernova Progenitor, *Astrophysical Journal Letter* **910**, p. L24 (April 2021).
83. O. G. Benvenuto, J. A. Panei, K. Nomoto, H. Kitamura and I. Hachisu, Final Evolution and Delayed Explosions of Spinning White Dwarfs in Single Degenerate Models for Type Ia Supernovae, *Astrophysical Journal Letter* **809**, p. L6 (August 2015).
84. Y. Ohshiro, H. Yamaguchi, S.-C. Leung, K. Nomoto, T. Sato, T. Tanaka, H. Okon, R. Fisher, R. Petre and B. J. Williams, Discovery of a Highly Neutronized Ejecta Clump in the Type Ia Supernova Remnant 3C 397, *Astrophysical Journal Letter* **913**, p. L34 (June 2021).
85. H. Yamaguchi, C. Badenes, A. R. Foster, E. Bravo, B. J. Williams, K. Maeda, M. Nobukawa, K. A. Eriksen, N. S. Brickhouse, R. Petre and K. Koyama, A Chandrasekhar Mass Progenitor for the Type Ia Supernova Remnant 3C 397 from the Enhanced Abundances of Nickel and Manganese, *Astrophysical Journal Letter* **801**, p. L31 (March 2015).
86. C. Kobayashi, S.-C. Leung and K. Nomoto, New Type Ia Supernova Yields and the Manganese and Nickel Problems in the Milky Way and Dwarf Spheroidal Galaxies, *Astrophysical Journal* **895**, p. 138 (June 2020).
87. B. E. Reddy, J. Tomkin, D. L. Lambert and C. Allende Prieto, The chemical compositions of Galactic disc F and G dwarfs, *Monthly Notices of the Royal Astronomical Society* **340**, 304 (March 2003).
88. B. E. Reddy, D. L. Lambert and C. Allende Prieto, Elemental abundance survey of the Galactic thick disc, *Monthly Notices of the Royal Astronomical Society* **367**, 1329 (April 2006).
89. S. Feltzing, M. Fohlman and T. Bensby, Manganese trends in a sample of thin and thick disk stars. The origin of Mn, *Astronomy and Astrophysics* **467**, 665 (May 2007).
90. E. N. Kirby, J. L. Xie, R. Guo, M. A. C. de los Reyes, M. Bergemann, M. Kovalev, K. J. Shen, A. L. Piro and A. McWilliam, Evidence for Sub-Chandrasekhar Type Ia Supernovae from Stellar Abundances in Dwarf Galaxies, *Astrophysical Journal* **881**, p. 45 (August 2019).
91. M. A. C. de los Reyes, E. N. Kirby, I. R. Seitenzahl and K. J. Shen, Manganese Indicates a Transition from Sub- to Near-Chandrasekhar Type Ia Supernovae in Dwarf Galaxies, *Astrophysical Journal* **891**, p. 85 (March 2020).
92. C. Kobayashi, K. Nomoto and I. Hachisu, Subclasses of Type Ia Supernovae as the Origin of [α/Fe] Ratios in Dwarf Spheroidal Galaxies, *Astrophysical Journal Letter* **804**, p. L24 (May 2015).
93. A. Simionescu, S. Nakashima, H. Yamaguchi, K. Matsushita, F. Mernier, N. Werner, T. Tamura, K. Nomoto, J. de Plaa, S. C. Leung, A. Bamba, E. Bulbul, M. E. Eckart, Y. Ezoe, A. C. Fabian, Y. Fukazawa, L. Gu, Y. Ichinohe, M. N. Ishigaki, J. S. Kaastra, C. Kilbourne, T. Kitayama, M. Leutenegger, M. Loewenstein, Y. Maeda, E. D. Miller, R. F. Mushotzky, H. Noda, C. Pinto, F. S. Porter, S. Safi-Harb, K. Sato, T. Takahashi, S. Ueda and S. Zha, Constraints on the chemical enrichment history of the Perseus Cluster of galaxies from high-resolution X-ray spectroscopy, *Monthly Notices of the Royal Astronomical Society* **483**, 1701 (February 2019).

94. K. Nomoto, C. Kobayashi and N. Tominaga, Nucleosynthesis in Stars and the Chemical Enrichment of Galaxies, *Annual Review of Astronomy and Astrophysics* **51**, 457 (August 2013).
95. I. R. Seitenzahl, F. Ciaraldi-Schoolmann, F. K. Röpke, M. Fink, W. Hillebrandt, M. Kromer, R. Pakmor, A. J. Ruiter, S. A. Sim and S. Taubenberger, Three-dimensional delayed-detonation models with nucleosynthesis for Type Ia supernovae, *Monthly Notices of the Royal Astronomical Society* **429**, 1156 (February 2013).
96. T. Sukhbold, T. Ertl, S. E. Woosley, J. M. Brown and H. T. Janka, Core-collapse Supernovae from 9 to 120 Solar Masses Based on Neutrino-powered Explosions, *Astrophysical Journal* **821**, p. 38 (April 2016).

Charged polarized white dwarfs with finite temperature as a possible source of type Ia supernovae

Sílvia P. Nunes

Departamento de Física, Instituto Tecnológico de Aeronáutica, 12228-900 São José dos Campos, São Paulo, Brazil
E-mail: silviapn@ita.br

José D. V. Arbañil

Departamento de Ciencias, Universidad Privada del Norte, Avenida el Sol 461 San Juan de Lurigancho, 15434 Lima, Peru

Facultad de Ciencias Físicas, Universidad Nacional Mayor de San Marcos, Avenida Venezuela s/n Cercado de Lima, 15081 Lima, Peru

Manuel Malheiro

Departamento de Física, Instituto Tecnológico de Aeronáutica, 12228-900 São José dos Campos, São Paulo, Brazil

We investigate the structure of polarized charged white dwarfs with finite temperature as a possible type Ia supernovae source. The white dwarf is modeled considering an isothermal core with central temperature 10^8 [K] and an envelope where the temperature distribution depends on the mass density. Regarding the fluid, we assume that it is composed of nucleons and electrons. The structure of the polarized charged white dwarfs is obtained by solving the Einstein-Maxwell equations with charge densities represented by two Gaussians, forming an electric dipole layer at the stellar surface. We obtain larger and more massive white dwarfs when polarized charge and the Gaussians width are increased. We find that to appreciate effects in the white dwarf's structure, the electric polarized charge must be in the order of 5.0×10^{20}[C]. We obtain a maximum white dwarf mass of around $2.4M_{\odot}$ for a polarized charge of 1.5×10^{21}[C]. This mass result can indicate that polarized charged white dwarfs are possible progenitors of superluminous type Ia supernovae. Furthermore, the mass-central density curves we obtain are very similar to the ones reached recently for ultra-magnetized white dwarfs.

Keywords: White dwarfs; charged white dwarfs; type Ia supernovae explosions

Introduction

The mechanisms that increase the mass of a white dwarf (WD) have been extensively studied since supermassive WDs were associated with superluminous Type Ia supernovae (2003fg,[1] 2009dc,[2] 2006gz[3] and 2007if[4]). In what concerns the structure of WDs, in literature, there are works consider magnetic field, rotation, and electric field to explain the increase of such mass. To present considerable effects in the WD structure, recent works[5,6] present magnetic fields are the order of 9×10^{13}[G] and electric field of around 10^{16} [V/cm] .

We investigate charged white dwarfs made up by a non-degenerate fluid with isothermal core with central temperature $T_c = 10^8$[K] and envelope with temperature distribution.[7] Moreover, we assume a charge polarization[8] in the surface of WDs. We obtain very massive white dwarfs which can be the progenitors of Type Ia Supernovae. We note that the polarization generate notable effect in the structure of WDs and it decreases the electric field outside the star.

1. Structure equilibrium equations and boundary conditions

The line element considered to describe a charged WD, in Schwarzschild-like coordinates, is given by

$$ds^2 = -e^{\nu}dt^2 + e^{\lambda}dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2, \tag{1}$$

where ν and λ depend on the radial coordinate only. The inclusion of the electromagnetic tensor in the energy-momentum tensor leads to the nonzero Maxwell-Einstein equations:

$$\frac{dq}{dr} = 4\pi\rho_e r^2 e^{\lambda/2}, \tag{2}$$

$$\frac{dm}{dr} = 4\pi\varepsilon^2 + \frac{q}{r}\frac{dq}{dr}, \tag{3}$$

$$\frac{dP}{dr} = -(P+\varepsilon)\left[4\pi r p + \frac{m}{r^2} - \frac{q^2}{r^3}\right]e^{\lambda} + \frac{q}{4\pi r^4}\frac{dq}{dr}, \tag{4}$$

$$\frac{d\nu}{dr} = -\frac{2}{(p+\varepsilon)}\left[\frac{dP}{dr} - \frac{q}{4\pi r^4}\frac{dq}{dr}\right], \tag{5}$$

with q and m representing respectively the electric charge and the mass within the radius r. The potential metric e^{λ} is described as

$$e^{\lambda} = \left[1 - \frac{2m}{r} + \frac{q^2}{r^2}\right]^{-1}. \tag{6}$$

Due to temperature, the white dwarfs are considered to have a partial degeneracy, thus the contribution of electrons in the EoS is related to Fermi-Dirac integrals. Once the temperature distribution is defined,[7] these integrals are solved by using an adaptive quadrature method.[9] Next, we proceed to integrate Eqs. (2)-(5) by using the fourth-order Runge-Kutta method for the uncharged star, i.e., $Q = 0$. For this aim, in the center $r = 0$, it is considered the boundary conditions:

$$m(0) = 0, \;\; \varepsilon(0) = \varepsilon_0, \text{ and } \nu(0) = \nu_0, \tag{7}$$

and the star's surface is determined at $P(r = R^+) = 0$. This first integration enables us to find the values of the total star mass M and total radius R^+. The location of the second and negative Gaussian at R^- is determined through the relation $R^- = R^+ + \Delta R$, where $\Delta R = 4b$ and $b = 5$[km]. The value of b employed is the

smallest necessary to obtain considerable effects on the structure of the white dwarf. Once found the polarized charge, by using the charge density, we integrate again Eqs. (2)-(5) to obtain the charged white dwarf structure. The boundary conditions in Eq. (7) are used together with $q(0) = 0$ and $\rho_e(0) = \rho_{e0}$.

2. Equation of state and charge density profile

Regarding the stellar fluid, we consider that the equation of state is made of by nucleons, electrons, and photons.[10,11] Due to our consideration of very high central temperatures ($T_c = 10^8$ [K]), we neglect the lattice interactions. Thus, the pressure P and energy density ε is depicted by

$$P = P_e + P_N + P_\gamma, \tag{8}$$

$$\varepsilon = \varepsilon_e + \varepsilon_N + \varepsilon_\gamma, \tag{9}$$

where the subindexes e, N and γ represent the electrons, nucleons, and photons contributions, respectively.

We assume that electrons and protons are susceptible to a centrifugal force responsible for accelerating heavy particles away from light ones. Consequently, polarization forms in the stellar envelope. Inspired in Negreiros work,[8] we consider the polarization to be defined with the inclusion of two Gaussians in the charge density:

$$\rho_e = e^{-\lambda/2}\left[k^+ \exp\left(-\frac{(r-R^+)^2}{b^2}\right) + k^- \exp\left(-\frac{(r-R^-)^2}{b^2}\right)\right], \tag{10}$$

where r represents the radial coordinate, $b = 5$ [km] is the Gaussian width, and with k^+ and k_- being constants that depend on the global charge neutrality. The R^+ is the radius for the uncharged configuration. Moreover, $R^- = R^+ + \Delta R$, being $\Delta R = 4b$ a fixed value for all the central energy densities.

To obtain k, we solve the integration by considering a total zero charge

$$Q_t = \int_0^\infty e^{\lambda/2} 4\pi r^2 \rho_e dr = 0. \tag{11}$$

By using Eq. (10) in this last relation, we find

$$k^+ = \frac{Q}{8\pi}\left(\frac{\sqrt{\pi}b\left(R^+\right)^2}{2} + \frac{\sqrt{\pi}b^3}{4}\right)^{-1}, \tag{12}$$

$$k^- = \frac{-Q}{8\pi}\left(\frac{\sqrt{\pi}b\left(R^-\right)^2}{2} + \frac{\sqrt{\pi}b^3}{4}\right)^{-1}. \tag{13}$$

With the intent to better illustrate the stellar charge profile, we present the charge density as a function of radial coordinate in Fig. 1.

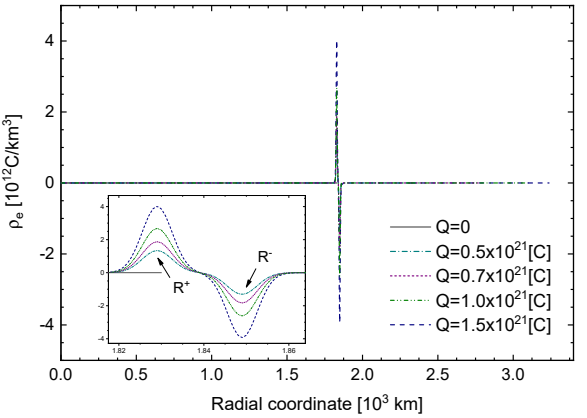


Fig. 1. The charge density as a function of the radial coordinate for a star with central energy density $\varepsilon_c = 10^{10}[\mathrm{g\ cm}^{-3}]$ and several polarized charge values.

3. White dwarf configurations

Consequently to the polarized charge, we find a very high electric field in the stellar envelope and no electric field outside the star (in contrast to previous charge white dwarfs studies[6]). In our calculations, considering a star with central density $\varepsilon_c = 5 \times 10^{10}$ [g/cm^3], we find maximum internal electric fields of around 10^{16} [V/cm]. Under these conditions, the radius is increased in 77% and the mass 63% for $Q = 1.5 \times 10^{21}$[C]. This electric field induces an increase in the internal star pressure.

The pressure in the stellar interior enhancing due to the charge density was already noticed in previous charge white dwarfs studies.[6] The pressure profile found by Carvalho and collaborators increases with charge and decreases due to self-gravitation. In our work, the neutralization of charge is the main responsible to decrease the stellar charge pressure. Thus, compared to Carvalho's work,[6] the charge density we propose affects more the mass than the radius.

In Fig. 2 we show the mass as a function of central energy density. This figure is important to analyze the relation between polarized charge white dwarfs and Type Ia Supernovae. We can note that for the uncharged sequence, the mass increases with central energy density until it reaches a maximum mass (pink dot) and decreases. This behavior is different for sequences of charged stars. These curves increase mass with central energy density (the same pattern observed for ultra-magnetized white dwarfs[5]). Another important observation in this figure is related to the maximum mass obtained.

For the maximum polarized charge we find maximum masses around $\approx 2.4 M_\odot$ for 1.5×10^{21}[C]. This increase in mass is obtained for the first time for charged white dwarfs with no electric field outside the star, which is the main novelty of this work.

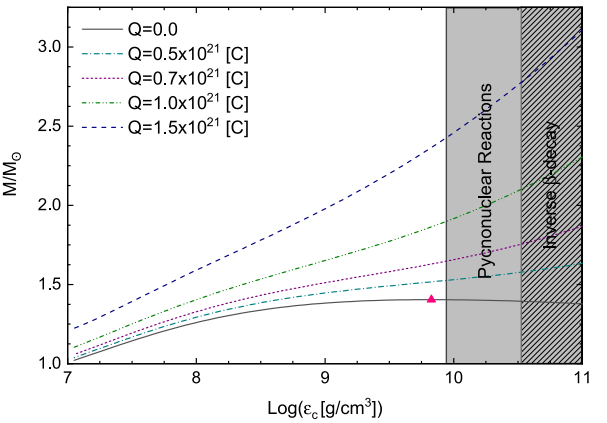


Fig. 2. The mass as a function of the energy density for several sequences of constant polarized charge and $T_c = 10^8$[K].

4. Conclusion

In this work, we investigated the structure of charged hot white dwarfs to associate them with superluminous type Ia supernovae. In what concerns to the stellar fluid, we consider a central temperature of $T_c = 10^8$[K] and a temperature distribution in the envelope.[11] Besides, due to the high central temperature considered, we neglected the particles' interactions. We used Maxwell-Einstein equations to find the structure of these charged polarized white dwarfs. We found that more massive and bigger white dwarfs are found with our predictions.

Our calculations enable the finding of maximum masses around $2.4M_{\odot}$ for stable charge white dwarfs configurations. Albeit we reach very high electric fields to obtain these results ($\approx 10^{16}$ [V/cm]), the electric field outside the star is zero. For the first time, isolated charged white dwarfs are obtained with a zero electric field outside.

Acknowledgments

We would like to thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant No. 2013/26258 − 4. SPN thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant No. 140863/2017−6 and CAPES for the financial support. JDVA would like to thank the Universidad Privada del Norte and Universidad Nacional Mayor de San Marcos for funding - RR Nº 005753-2021-R/UNMSM under the project number B21131781. MM is grateful to CAPES and CNPq financial support.

References

1. D. A. Howell, M. Sullivan, P. E. Nugent, R. S. Ellis, A. J. Conley, D. Le Borgne, R. G. Carlberg, J. Guy, D. Balam, S. Basa, D. Fouchez, I. M. Hook, E. Y. Hsiao, J. D. Neill, R. Pain, K. M. Perrett and C. J. Pritchet, The type Ia supernova SNLS-03D3bb from a super-Chandrasekhar-mass white dwarf star, *Nature* **443**, 308 (September 2006).
2. S. Taubenberger, S. Benetti, M. Childress, R. Pakmor, S. Hachinger, P. A. Mazzali, V. Stanishev, N. Elias-Rosa, I. Agnoletto, F. Bufano, M. Ergon, A. Harutyunyan, C. Inserra, E. Kankare, M. Kromer, H. Navasardyan, J. Nicolas, A. Pastorello, E. Prosperi, F. Salgado, J. Sollerman, M. Stritzinger, M. Turatto, S. Valenti and W. Hillebrandt, High luminosity, slow ejecta and persistent carbon lines: SN 2009dc challenges thermonuclear explosion scenarios, *MNRAS* **412**, 2735 (April 2011).
3. M. Hicken, P. M. Garnavich, J. L. Prieto, S. Blondin, D. L. DePoy, R. P. Kirshner and J. Parrent, The Luminous and Carbon-rich Supernova 2006gz: A Double Degenerate Merger?, *ApJ* **669**, L17 (November 2007).
4. R. A. Scalzo, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, C. Buton, M. Childress, N. Chotard, Y. Copin, H. K. Fakhouri, A. Gal-Yam, E. Gangler, S. Hoyer, M. Kasliwal, S. Loken, P. Nugent, R. Pain, E. Pécontal, R. Pereira, S. Perlmutter, D. Rabinowitz, A. Rau, G. Rigaudier, K. Runge, G. Smadja, C. Tao, R. C. Thomas, B. Weaver and C. Wu, Nearby Supernova Factory Observations of SN 2007if: First Total Mass Measurement of a Super-Chandrasekhar-Mass Progenitor, *ApJ* **713**, 1073 (April 2010).
5. E. Otoniel, B. Franzon, G. A. Carvalho, M. Malheiro, S. Schramm and F. Weber, Strongly Magnetized White Dwarfs and Their Instability Due to Nuclear Processes, *ApJ* **879**, p. 46 (July 2019).
6. G. A. Carvalho, J. D. V. Arbañil, R. M. Marinho and M. Malheiro, White dwarfs with a surface electrical charge distribution: Equilibrium and stability, *European Physical Journal C* **78**, p. 411 (May 2018).
7. A. L. Kritcher, D. C. Swift, T. Döppner, B. Bachmann, L. X. Benedict, G. W. Collins, J. L. DuBois, F. Elsner, G. Fontaine, J. A. Gaffney, S. Hamel, A. Lazicki, W. R. Johnson, N. Kostinski, D. Kraus, M. J. MacDonald, B. Maddox, M. E. Martin, P. Neumayer, A. Nikroo, J. Nilsen, B. A. Remington, D. Saumon, P. A. Sterne, W. Sweet, A. A. Correa, H. D. Whitley, R. W. Falcone and S. H. Glenzer, A measurement of the equation of state of carbon envelopes of white dwarfs, *Nature* **584**, 51 (August 2020).
8. R. Picanço Negreiros, I. N. Mishustin, S. Schramm and F. Weber, Properties of bare strange stars associated with surface electric fields, *Phys. Rev. D* **82**, p. 103010 (November 2010).
9. K. A. Boshkayev, J. A. Rueda, B. A. Zhami, Z. A. Kalymova and G. S. Balgymbekov, Equilibrium structure of white dwarfs at finite temperatures, in *IJMPC*, 2016.
10. F. X. Timmes and D. Arnett, The Accuracy, Consistency, and Speed of Five Equations of State for Stellar Hydrodynamics, *ApJS* **125**, 277 (November 1999).
11. S. P. Nunes, J. D. V. Arbañil and M. Malheiro, The structure and stability of massive hot white dwarfs, *To appear in the Astrophysical Journal*, p. arXiv:2108.08238 (August 2021).

CTCV J2056-3014 and other fast-spinning white dwarfs

C. V. Rodrigues

Divisão de Astrofísica, Instituto Nacional de Pesquisas Espaciais
12227-010, São José dos Campos, SP, Brazil
E-mail: claudia.rodrigues@inpe.br
www.inpe.br

R. Lopes de Oliveira

Departamento de Física, Universidade Federal de Sergipe
Av. Marechal Rondon, S/N, 49100-000 São Cristóvão, SE, Brazil
E-mail: raimundo.lopes@academico.ufs.br
and
Observatório Nacional,
Rua Gal. José Cristino 77, 20921-400, Rio de Janeiro, RJ, Brazil

A. Bruch

Laboratório Nacional de Astrofísica
Rua Estados Unidos, 154, CEP 37504-364, Itajubá, MG, Brazil
E-mail: albert@lna.br

A. S. Oliveira

IP&D, Universidade do Vale do Paraíba, 12244-000, São José dos Campos, SP, Brazil
E-mail: alexandre@univap.br

K. Mukai

CRESST II and X-ray Astrophysics Laboratory, NASA Goddard Space Flight Center
Greenbelt, MD 20771, USA
E-mail: koji.mukai-1@nasa.gov
and
Center for Space Science and Technology, University of Maryland
Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

This contribution summarises our recent determination of the spin period of the magnetic white dwarf in CTCV J2056-3014, a cataclysmic variable binary system. Its X-ray and optical emission comes from its magnetic accretion column and is modulated with a 29.6 s period, due to the WD rotation. We briefly discuss this object in the context of other fast-spinning white dwarfs.

Keywords: White dwarfs; Cataclysmic variable stars; X-rays: stars.

1. Introduction

White dwarfs (WDs) are the evolutionary fate of most stars. WDs are composed of degenerated matter and, together with neutrons stars and stellar black holes, constitute the group of compact stellar objects with strong gravitational fields in

their neighborhood. The maximum mass that a WD can reach is around 1.4 $M_\odot$, the Chandrasekhar limit. The exact limit depends on many assumptions as the rotation, the magnetic field strength, and the adopted equation of state, which is related to the physical conditions and processes taking place in the stellar interior. Due to the compactness of their degenerated interior, WDs can rotate very rapidly, reaching spin periods as small as a few tens of seconds. WDs also show the largest values of magnetic moments among stellar objects. As stars in general, they are found isolated or in binary systems.

In isolated WDs, high values of mass, spin, and magnetic field seem to appear together. WD 0316-849 is one example of such an extreme WD. It has a magnetic field (B) of around 500 MG, a spin (P_{spin}) of 725 s, and a mass (M_{WD}) equal to 0.86 $M_\odot$(see Ref. 1 and references therein). ZTF J190132.9+145808.7 has been recently associated with a very massive WD, $M_{WD} > 1.3$ $M_\odot$. Its P_{spin} is 416.4 s and its surface magnetic field is estimated to be in the range 600 – 900 MG.[2] Those WDs are probably originated by mergers - see, for instance, the discussion in Ref. 3. In particular, mergers can produce remnants with high angular momenta.

CVs are binary systems in which a WD accretes matter from the secondary star, a low-mass on the main sequence or slightly evolved that loses material by Roche lobe overflow. The magnetic CVs are those systems in which the WD magnetic field is strong enough to play a role in the dynamics of the accretion flow. If the WD magnetic field is not strong enough to synchronize the WD spin with the orbital period, the system is classified as an intermediate polar.

The optical and X-ray emission of intermediate polars is caused by mass accretion onto the magnetic WD and modulates with the WD spin. It is caused by accretion onto a magnetic WD. This process produces a stand-off shock near the WD surface where gravitational energy is converted to thermal energy. This enhances the gas density and heats it to keV temperatures producing the X-ray emission and the high excitation optical spectrum. The footprint of the accretion structure covers a small area of the WD surface. Therefore, the WD rotation causes a modulation of the emission as seen by the observer.

In this contribution, we outline our recent finding that CTCV J2056-3014, an intermediate polar, harbors a very fast-spinning WD. We also discuss this result in the context of similar objects.

2. CTCV J2056-3014

The optical spectrum of CTCV J2056-3014[4,5] is characterized by a very blue continuum and strong emission lines, including the Bowen complex, indicative of high excitation temperatures. The radial velocity variation of the emission lines indicates an orbital period of 1.76 h.[4] The first indications that CTCV J2056-3014 could be an intermediate polar were the possible association with the X-ray source 1RXS J205652.1−301433 and a photometric modulation of around 15 min.[4] However, this periodic variability was not confirmed by further observations.[6]

Recently, XMM-Newton X-ray observations of CTCV J2056-3014 have revealed a coherent modulation at 29.6098 ± 0.0014 s, also present in the optical emission.[7] This modulation and the X-ray spectrum of CTCV J2056-3014 confirm that the object is an intermediate polar, i.e., an accreting magnetic WD in an asynchronous CV. The fit of its X-ray spectrum provides an accretion mass rate, $\dot{M}$, of about 6×10^{-12} $M_{\odot}yr^{-1}$ (Ref. 7).

Most intermediate polars have X-ray luminosities, L_X, of $\sim 10^{33}$ erg s^{-1}. The luminosity of CTCV J2056-3014 is around 100 times smaller (10^{31} erg s^{-1}); a property shared with a very small group of CVs named low-luminosity intermediate polars (LLIPs).[8] It is not clear if the number of LLIPs is small because they are rare objects or because their X-ray faintness hampers their discovery in X-ray surveys. If the true number of LLIPs is high and we are missing them because of observational biases, this class of objects is probably a relevant contributor to the Galactic X-ray background.[8]

3. The fastest spinning WDs

In this section, we enumerate the bona fide fastest spinning WDs, which are all located in compact binary systems. However, for completeness, we also mention some objects that could be WDs with a rapid spin. A summary of selected characteristics of individual objects is shown in Table 1.

There is a debate in the literature if some soft gamma-ray repeaters (SGR) and anomalous X-ray pulsars (AXP), the so-called magnetars, could contain a WD instead of a neutron star.[9,10] For example, the spectral energy distribution from the infrared to the gamma-rays of 4U 0142+61 (P_{spin}= 8.62 s) can be fitted in a scenario of a fast, massive, and magnetic WD.[3] If some magnetars were indeed WDs, they would put strong constraints on the physics of the WD interiors, to cite only one consequence of this hypothesis. In spite of the importance of this subject, those objects are not considered here.

WZ Sge is a short orbital period cataclysmic variable that shows dwarf-nova outbursts with a very long recurrence time (a few decades). The system has intermittent periodical signals at 28.87 s and 28.96 s. One of these periods could be the WD spin period, but it is not a settled question (see Ref. 11 and references therein). We do not consider this system as a confirmed fast-spinning WD, but we list its properties in Table 1.

Another fast-spinning compact object is RX J0648.0-4418, with a P_{spin} of 13.2 s. This object is in a relatively wide binary system, with an orbital period of 1.55 d. However, it is debated if the compact object is a WD or a neutron star.[7,12] In some sense, this object is a link between fast-spinning WDs and magnetars. As for WZ Sge, we do not include it in the list of bona fide fast-spinning WDs, but its properties can be found in Table 1.

AE Aqr was the first discovered fast and magnetic WD, and its peculiar properties prompted many studies.[13] Its WD spins at a 33 s rate. The system is the

prototype of a propeller: a system in which the centrifugal force of a magnetic fast-rotating object prevents the accretion from a mass reservoir to reach the object's surface. This is one of the possible accretion-flow configurations in cataclysmic variables. It depends on the WD magnetic field, WD spin, and accretion rate.[14] However, even if part of the material is propelled out of the WD magnetosphere, some material can reach the WD surface. This is probably the case for AE Aqr,[15,16] since its X-ray luminosity is small but not null: $\dot{M} \approx 10^{31}$ erg s^{-1}.[15] The radius of the companion star is twice as large as is expected for its mass, indicating a very inflated object.

V1460 Her is an intermediate polar with a WD that spins at 38.9 s.[17] The companion star is less massive than that of AE Aqr, but it is also significantly inflated. Contrary to AE Aqr, there is an accretion disk in V1460 Her, clearly seen in optical spectra.

At the time of this conference, CTCV J2056-3014 was known as the fastest-spinning WD. However, a faster WD has been recently discovered: LAMOST J024048.51+195226.9, with a spin period of 24.9328 $\pm$ 0.0038 s.[18] The system shows strong evidence of an outflow with properties consistent with those expected for a propeller.[19]

4. Discussion and conclusions

Two possible mechanisms for spinning up WDs are mergers and accretion.

The isolated WDs with large values of M_{WD}, B, and P_{spin} are probably produced by mergers. The majority of the most massive WDs have high values of magnetic fields,[1,2,24–26] which indicates that these two properties may have a common origin. Simulations show that double-degenerated mergers can produce WDs with high and stable magnetic fields.[27] From an observational perspective, there is also evidence for mergers. The color-magnitude distribution of a *Gaia* sample of nearby WDs ($d \lesssim 100$ pc) indicates the presence of two populations of WDs, which can be explained if some objects are formed by a merger.[28] Considering objects within 20 pc, the fraction of WDs in binary systems is considerably smaller than the fraction of binary systems composed of two main-sequence objects. This can be explained if a portion of the isolated WDs is formed via mergers.[29] Although mergers can play an important role in producing extreme WDs, the bona fide fastest-rotating WDs are all located in binary systems, specifically cataclysmic variables.

In cataclysmic variables, the rapid WD spin may be the result of the mass transfer. Specifically, the accretion transforms the orbital angular momentum into the rotational angular momentum of the WD. To reach the observed high P_{spin}, high values of $\dot{M}$ are necessary. However, a high mass-transfer rate should produce a high X-ray luminosity, which is not observed in the fast-spinning WDs (see Table 1). Therefore, the current high spin is explained by a previous phase of high mass transfer rate from a secondary star in a thermal time-scale phase. In fact, the

Table 1. Selected properties of the confirmed fastest spinning WDs and related objects.

Object	P_{spin} (s)	Orbital period (h)	M_{sec} ($M_\odot$)	M_{WD} ($M_\odot$)	L_X (erg s^{-1})	Confirmed	Comments	References
LAMOST J024048.51+195226.9	24.9328	7.33	-	-	-	Yes	Propeller	18, 19
CTCV J2056-3014	29.6098	1.76	-	0.56 – 1.38	10^{-31}	Yes	Below the period gap	4, 7, 20
AE Aqr	33	9.88	0.37 ± 0.04	0.63 ± 0.05	10^{-31}	Yes	Inflated secondary; Propeller	21, 21
V1460 Her	38.9	4.99	0.295 ± 0.004	0.869 ± 0.006	-	Yes	Inflated secondary	17
RX J0648.0-4418	13.2	37.2	1.50 ± 0.05	1.28 ± 0.05	10^{-32}	No	WD or neutron star?	22
WZ Sge	28.87 or 28.96	1.36	0.078 – 0.13	0.88 – 1.53	10^{-30}	No	Period is not firmly associated with spin	11, 23

secondary stars in AE Aqr[13] and V1460 Her[17] are twice as large as is expected for their masses, which reinforces the idea of a star slightly out of equilibrium. If the same is true for CTCV J2056-3014 and LAMOST J024048.51+195226.9 cannot be said, because their secondary stars have not yet been directed observed. Along its evolution, a CV evolves to shorter orbital periods. Consequently, the component separation as well as the radius of the secondary star shrink. In particular, the mass of the secondary, M_{sec}, is expected to decrease together with the orbital period. Moreover, above the observed period gap in the orbital period distribution of CVs, the angular-momentum transfer occurs by magnetic braking and produces larger $\dot{M}$ in comparison with objects below the period gap, which should have the angular momentum transfer sustained by gravitational radiation. Interestingly, CTCV J2056-3014 is the only object among the fast-spinning WD below the period gap.

The WDs in CTCV J2056-3014 and AE Aqr are certainly magnetic.[7] The understanding of the origin of the modulation in V1460 Her and LAMOST J024048.51+195226.9 would benefit from X-ray observations, which could reveal the presence of a magnetic WD.

Until recently, AE Aqr occupied an isolated position as the fastest-spinning confirmed WD. In the last years, other fast-spinning WDs have been discovered, all of them in CVs. Although most of them are similar to AE Aqr in the context of CV evolution, CTCV J2056-3014 - the only system with an orbital period below the CV period gap - seems to be a more evolved object. Those new results can help us understanding the physics behind the rapid spin of WDs in CVs and how the WD spin evolves along the CV evolution.

Acknowledgements

The authors thank the organizers for the invitation to present this work at this conference. CVR and RLO acknowledge CNPq – *Conselho Nacional de Desenvolvimento Científico e Tecnológico*, through Grants 303444/2018-5 and 312705/2020-4, respectively.

References

1. S. Bagnulo and J. D. Landstreet, New insight into the magnetism of degenerate stars from the analysis of a volume-limited sample of white dwarfs, *MNRAS* **507**, 5902 (November 2021).
2. I. Caiazzo, K. B. Burdge, J. Fuller, J. Heyl, S. R. Kulkarni, T. A. Prince, H. B. Richer, J. Schwab, I. Andreoni, E. C. Bellm, A. Drake, D. A. Duev, M. J. Graham, G. Helou, A. A. Mahabal, F. J. Masci, R. Smith and M. T. Soumagnac, A highly magnetized and rapidly rotating white dwarf as small as the Moon, *Nature* **595**, 39 (June 2021).
3. S. V. Borges, C. V. Rodrigues, J. G. Coelho, M. Malheiro and M. Castro, A Magnetic White Dwarf Accretion Model for the Anomalous X-Ray Pulsar 4U 0142+61, *The Astrophysical Journal* **895**, p. 26 (May 2020).

4. T. Augusteijn, C. Tappert, T. Dall and J. Maza, Cataclysmic variables from the Calán-Tololo Survey - II. Spectroscopic periods, *MNRAS* **405**, 621 (June 2010).
5. A. S. Oliveira, C. V. Rodrigues, D. Cieslinski, F. J. Jablonski, K. M. G. Silva, L. A. Almeida, A. Rodríguez-Ardila and M. S. Palhares, Exploratory Spectroscopy of Magnetic Cataclysmic Variables Candidates and Other Variable Objects, *The Astronomical Journal* **153**, p. 144 (April 2017).
6. A. Bruch, Photometry of some more neglected bright cataclysmic variables and candidates, *New Astronomy* **58**, 53 (January 2018).
7. R. Lopes de Oliveira, A. Bruch, C. V. Rodrigues, A. S. Oliveira and K. Mukai, CTCV J2056-3014: An X-Ray-faint Intermediate Polar Harboring an Extremely Fast-spinning White Dwarf, *The Astrophysical Journal Letters* **898**, p. L40 (August 2020).
8. M. L. Pretorius and K. Mukai, Constraints on the space density of intermediate polars from the Swift-BAT survey, *MNRAS* **442**, 2580 (August 2014).
9. M. Malheiro, J. A. Rueda and R. Ruffini, SGRs and AXPs as Rotation-Powered Massive White Dwarfs, *Publications of the Astronomical Society of Japan* **64**, p. 56 (June 2012).
10. J. G. Coelho and M. Malheiro, Magnetic dipole moment of soft gamma-ray repeaters and anomalous X-ray pulsars described as massive and magnetic white dwarfs, *Publications of the Astronomical Society of Japan* **66**, p. 14 (February 2014).
11. A. A. Nucita, E. Kuulkers, F. De Paolis, K. Mukai, G. Ingrosso and B. M. T. Maiolo, XMM-Newton and Swift observations of WZ Sagittae: Spectral and timing analysis, *Astronomy & Astrophysics* **566**, p. A121 (June 2014).
12. S. Mereghetti, F. Pintore, T. Rauch, N. La Palombara, P. Esposito, S. Geier, I. Pelisoli, M. Rigoselli, V. Schaffenroth and A. Tiengo, New X-ray observations of the hot subdwarf binary HD 49798/RX J0648.0-4418, *MNRAS* **504**, 920 (June 2021).
13. P. J. Meintjes, A. Odendaal and H. van Heerden, AE Aquarii: A Short Review, *Acta Polytechnica CTU Proceedings* **2**, 86 (January 2015).
14. A. J. Norton, O. W. Butters, T. L. Parker and G. A. Wynn, The Accretion Flows and Evolution of Magnetic Cataclysmic Variables, *The Astrophysical Journal* **672**, 524 (January 2008).
15. T. Kitaguchi, H. An, A. M. Beloborodov, E. V. Gotthelf, T. Hayashi, V. M. Kaspi, V. R. Rana, S. E. Boggs, F. E. Christensen, W. W. Craig, C. J. Hailey, F. A. Harrison, D. Stern and W. W. Zhang, NuSTAR and Swift Observations of the Fast Rotating Magnetized White Dwarf AE Aquarii, *The Astrophysical Journal* **782**, p. 3 (February 2014).
16. C. Rodrigues, K. da Silva, G. Luna, J. Coelho, I. Lima, J. Costa and J. de Araujo, The accretion column of AE Aqr, in *The X-ray Universe 2017*, eds. J.-U. Ness and S. Migliari (ESA, October 2017).
17. R. P. Ashley, T. R. Marsh, E. Breedt, B. T. Gänsicke, A. F. Pala, O. Toloza, P. Chote, J. R. Thorstensen and M. R. Burleigh, V1460 Her: A fast spinning white dwarf accreting from an evolved donor star, *MNRAS* **499**, 149 (November 2020).
18. I. Pelisoli, T. R. Marsh, V. S. Dhillon, E. Breedt, A. J. Brown, M. J. Dyer, M. J. Green, P. Kerry, S. P. Littlefair, S. G. Parsons, D. I. Sahman and J. F. Wild, Found: A rapidly spinning white dwarf in LAMOST J024048.51+195226.9, *arXiv e-prints*, p. arXiv:2108.11396 (August 2021).
19. P. Garnavich, C. Littlefield, R. M. Wagner, J. van Roestel, A. D. Jaodand, P. Szkody and J. R. Thorstensen, Confirmation of a Second Propeller: A High-inclination Twin of AE Aquarii, *The Astrophysical Journal* **917**, p. 22 (August 2021).
20. E. Otoniel, J. G. Coelho, S. P. Nunes, M. Malheiro and F. Weber, Mass limits of the extremely fast-spinning white dwarf CTCV J2056-3014, *arXiv e-prints*, p. arXiv:2010.12441 (October 2020).

21. J. Echevarría, R. C. Smith, R. Costero, S. Zharikov and R. Michel, High-dispersion absorption-line spectroscopy of AE Aqr, *MNRAS* **387**, 1563 (July 2008).
22. S. Mereghetti, A. Tiengo, P. Esposito, N. La Palombara, G. L. Israel and L. Stella, An Ultramassive, Fast-Spinning White Dwarf in a Peculiar Binary System, *Science* **325**, p. 1222 (September 2009).
23. D. Steeghs, S. B. Howell, C. Knigge, B. T. Gänsicke, E. M. Sion and W. F. Welsh, Dynamical Constraints on the Component Masses of the Cataclysmic Variable WZ Sagittae, *The Astrophysical Journal* **667**, 442 (September 2007).
24. M. Należyty and J. Madej, A catalogue of isolated massive white dwarfs. Mass distribution of massive star, *Astronomy & Astrophysics* **420**, 507 (June 2004).
25. L. Ferrario, D. de Martino and B. T. Gänsicke, Magnetic White Dwarfs, *Space Science Review* **191**, 111 (October 2015).
26. L. Ferrario, D. Wickramasinghe and A. Kawka, Magnetic fields in isolated and interacting white dwarfs, *Advances in Space Research* **66**, 1025 (September 2020).
27. E. García-Berro, P. Lorén-Aguilar, G. Aznar-Siguán, S. Torres, J. Camacho, L. G. Althaus, A. H. Córsico, B. Külebi and J. Isern, Double Degenerate Mergers as Progenitors of High-field Magnetic White Dwarfs, *The Astrophysical Journal* **749**, p. 25 (April 2012).
28. M. Kilic, N. C. Hambly, P. Bergeron, C. Genest-Beaulieu and N. Rowell, Gaia reveals evidence for merged white dwarfs, *MNRAS* **479**, L113 (September 2018).
29. S. Toonen, M. Hollands, B. T. Gänsicke and T. Boekholt, The binarity of the local white dwarf population, *Astronomy & Astrophysics* **602**, p. A16 (June 2017).

Gravitational waves from fast-spinning white dwarfs

M. F. Sousa* and J. C. N. de Araujo†

Divisão de Astrofísica, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP/12227-010, Brazil
**E-mail: manoel.sousa@inpe.br*
†E-mail: jcarlos.dearaujo@inpe.br
www.inpe.br

J. G. Coelho

Departamento de Física, Universidade Tecnológica Federal do Paraná, Medianeira, PR/85884-000, Brazil
E-mail: jazielcoelho@utfpr.edu.br

We discuss some aspects of Sousa et al. [1, 2] concerning two mechanisms of gravitational wave (GW) emission in fast-spinning white dwarfs (WDs): accretion of matter and magnetic deformation. In both cases, the GW emission is generated by an asymmetry around the rotation axis of the star. However, in the first case, the asymmetry is due to the amount of accreted matter in the magnetic poles, while in the second case it is due to the intense magnetic field. We have estimated the GW amplitude and luminosity for three binary systems that have a fast-spinning magnetized WD, namely, AE Aquarii, AR Scorpii and RX J0648.0-4418. In addition, we applied the magnetic deformation mechanism for SGRs/AXPs described as WD pulsars. We found that, for the first mechanism, the systems AE Aquarii and RX J0648.0-4418 can be observed by the space detectors BBO and DECIGO if they have an amount of accreted mass of $\delta m \geq 10^{-5} M_{\odot}$. For the second mechanism, the three systems studied require that the WD has a magnetic field above $\sim 10^9$ G to emit GWs that can be detected by BBO. Furthermore, we found that some SGRs/AXPs as WD pulsars can be detected by BBO and DECIGO, whereas SGRs/AXPs as highly magnetized neutron stars are far below the sensitivity curves of these detectors.

Keywords: Gravitational Waves; White Dwarfs; Magnetic Field; Rapid Rotation.

1. Introduction

Over the last years, the astrophysical community's interest in highly magnetized white dwarfs (HMWDs) has increased. Recent results of the Sloan Digital Sky Survey (SDSS) have confirmed these white dwarfs (WDs) with surface magnetic fields ranging from 10^6 G up to 10^9 G [see e.g. Refs. 3–5]. Besides their high magnetic fields, most of them have been shown to be massive, and responsible for the high-mass peak at 1 $M_{\odot}$ of the WD mass distribution.[3,6,7]

Typically, WDs have rotation periods of days or even years. However, recently, a WD pulsar was discovered, called AR Scorpii. This star emits from X-ray to radio wavelengths, pulsing in brightness with a period of 1.97 min.[8] Moreover, other sources have been proposed as candidates of WD pulsars. Specific examples are AE Aquarii with a short rotation period of $P = 33.08$ s[9] and RX J0648.0-4418

(RX J0648, hereafter) that is a massive WD with $M = 1.28 M_{\odot}$ and with a very fast spin period of $P = 13.2$ s, that belongs to the binary system HD 49798/RX J0648.0-4418.[10] Nevertheless, it is worth mentioning that the nature of RX J0648 is unclear, meaning it is not yet clearly known whether this star is a WD or a neutron star.[11,12]

Recently, from XMM-Newton observations, the authors in Ref. 13 reported that CTCV J2056-3014 is a X-ray-faint intermediate polar harboring an extremely fast-spinning WD with a coherent pulsation of 29.6 s, thus being the fastest confirmed spin in a WD [see also Ref. 14]. Other fast-spinning WDs have been observed more recently. As examples we can cite: V1460 Her, which is an eclipsing cataclysmic variable, with a overluminous K5-type donor star and a WD that rotates with a period of 38.9 s[15] and ZTF J190132.9+145808.7, which is a highly magnetized and rapidly rotating white dwarf, featuring a magnetic field with strengths between 600 MG and 900 MG on its surface, and a stellar radius that is only slightly larger than the radius of the Moon.[16] This WD has a rotation period of 6.94 min which is considered short as this star is an isolated WD.

Notwithstanding, several studies of magnetized and fast-rotating WDs have been done. In particular, we can highlight one involving WD pulsars in an alternative description for Soft Gamma Repeaters (SGRs) and Anomalous X-Ray Pulsars (AXPs) [see e.g. Refs. 17–20]. From this perspective, the process of energy emission released by dipole radiation in a WD can be explained by a canonical spin-powered pulsar model, since they share quite similar aspects.[18,21]

On the other hand, LIGO and Virgo detectors have recently made direct observations of gravitational waves (GWs).[22,23] All these GW detections are within a frequency band ranging from 10 Hz to 1000 Hz, which is the operating band of LIGO and Virgo. Nevertheless, as is well known, there are proposed missions for lower frequencies, such as LISA,[24,25] whose frequency band is of $(10^{-4} - 0.01)$ Hz, BBO[26,27] and DECIGO[28,29] in the frequency band ranging from 0.01 Hz to 10 Hz.

The generation of continuous GWs in different possibilities has already been proposed.[30–34] More recently, in Refs. 1, 2, 35, it has been suggested that rotating magnetized WDs can emit continuous GWs with amplitudes possibly detected by upcoming GW detectors such as LISA, DECIGO and BBO. Here we revisit our two works [Refs. 1, 2], where we investigate two mechanisms of gravitational radiation emission in fast-rotating magnetized WD: matter accretion and magnetic deformation. In both cases, the emission in GW is produced by the asymmetry around the star's rotation axis.

2. Gravitational emission mechanisms

WDs might generate gravitational radiation whether they are not perfectly symmetric around their rotation axis. The huge dipole magnetic field that can make the star become oblate[36] and accretion of matter are two examples where this asymmetry can occur.

2.1. *Accretion of matter*

The GW emission is shown here for the case of a WD accreting matter via the magnetic poles which do not coincide with the rotation axis of the star. In this scenario, the system's secondary star transfers matter to the WD through an accretion column and accumulates an amount of mass on the magnetic poles.

Thus, we consider a rigid object rotating about a non-major axis (x_1, x_2, x_3) and which has a deformity about one of the major axes (x_1, x_2, x_3), where are positioned the main moments of inertia I_1, I_2 e I_3, respectively.

With this configuration and doing $I_1 = I_2$, the gravitational amplitude and luminosity are given respectively by[37,38]

$$h_{0_{ac}} = \frac{4G}{c^4}\frac{(I_1 - I_3)\omega^2}{r}\sin^2\theta, \tag{1}$$

$$L_{GW_{ac}} = -\frac{2}{5}\frac{G}{c^5}(I_1 - I_3)^2\omega^6\sin^2\theta(16\sin^2\theta + \cos^2\theta), \tag{2}$$

where, ω is the angular velocity, θ is the angle between the rotation and magnetic dipole axes and r is the distance to the emitting source.

Now, to determine the moments of inertia I_1 and I_3, we consider that the object has an amount of mass accumulated on the x_3' axis. We reduce this system to a large sphere with two smaller spheres of matter on the x_3' axis: one at each of the poles of the larger sphere. This would be equivalent to a WD accreting matter by the two magnetic poles. Therefore, it follows immediately that

$$\begin{aligned} I_1 &= \frac{2}{5}MR^2 + 2\delta m\ R^2, \\ I_3 &= \frac{2}{5}MR^2 + 2\frac{2}{5}\delta m\ a^2, \end{aligned} \tag{3}$$

where M is the mass of the star, R is the radius of the star, δm is the amount of mass accumulated on one magnetic pole and a its radius.

Considering that $R \gg a$ and by substituting these last expressions into Eqs. (1) and (2), one obtains

$$h_{0_{ac}} = \frac{8G}{c^4}\frac{\delta m\ R^2\omega^2}{r}\sin^2\theta, \tag{4}$$

and

$$L_{GW_{ac}} = -\frac{8}{5}\frac{G}{c^5}\delta m^2R^4\omega^6\sin^2\theta(16\sin^2\theta + \cos^2\theta). \tag{5}$$

Thereby, we find expressions for the gravitational luminosity and the GW amplitude for the case of a WD accumulating mass, which depends on the accreted mass, the distance to the source, the radius of the star and how fast it is rotating.

2.2. *Magnetic deformation*

This section deals with the deformation of the WD induced by its own huge magnetic field. Due to the combination of magnetic field and rotation, a WD can become triaxial, presenting therefore a triaxial moment of inertia. In order to investigate the effect arising from the magnetic stress on the equilibrium of stars, let us introduce the equatorial ellipticity, defined as[37,38]

$$\epsilon = \frac{I_1 - I_2}{I_3}. \tag{6}$$

where I_1, I_2 and I_3 are main moments of inertia with respect to the (x, y, z) axes, respectively.

If the star rotates around the $z-$axis, then it will emit monochromatic GWs with a frequency twice the rotation frequency, f_{rot}, with amplitude given by[37,38]

$$h_{0_{df}} = \frac{16\pi^2 G}{c^4}\frac{I_3 f_{rot}^2}{r}\,\epsilon, \tag{7}$$

and luminosity as follows:

$$L_{GW_{df}} = -\frac{32}{5}\frac{G}{c^5} I_3^2 \epsilon^2 \omega_{rot}^6. \tag{8}$$

On the other hand, recall that the ellipticity of magnetic origin can be written as[39,40]

$$\epsilon = \kappa \frac{B_s^2 R^4}{GM^2}, \tag{9}$$

where, B_s is the magnetic field strength on the star's surface and κ is the distortion parameter, which depends on the magnetic field configuration and equation of state (EoS) of the star.

Now, substituting this last equation into Eqs. (7) and (8) and considering $I_3 = 2MR^2/5$, one immediately obtains that

$$h_{df} = \frac{32\pi^2}{5c^4}\frac{R^6 f_{rot}^2}{rM}\kappa B_s^2, \tag{10}$$

$$L_{GW_{df}} = -\frac{2^{13}\pi^6}{5^3 c^5}\frac{R^{12} f_{rot}^6}{GM^2}\kappa^2 B_s^4. \tag{11}$$

Note that the two equations just above depend on the rotation frequency and the magnetic field strength.

In contrast, the GW amplitude can also be written as a function of the variation of the star's rotation frequency $\dot{f}_{rot}$. In this case, we must consider that a part of the spindown luminosity is converted into GWs. Thus, we can infer an efficiency, η_{df}, for the variation of the rotation frequency as $\dot{\bar{f}}_{rot} = \eta_{df}\dot{f}_{rot}$, such that $\dot{\bar{f}}_{rot}$ can be

interpreted as the part of $\dot{f}_{rot}$ related to the GW brake. Hence, the GW amplitude can be written as follows

$$h_0^{sd} = \left(\eta_{df} \, \frac{5}{2} \frac{G}{c^3} \frac{I_3 \dot{f}_{rot}}{r^2 f_{rot}} \right)^{1/2} . \tag{12}$$

3. Gravitational waves from rapid rotation white dwarfs

3.1. *Accretion of matter*

Considering the scenario of an amount of mass accumulated on the magnetic poles, we apply Eq. (4) for the three binary systems that have fast-spinning WDs: AE Aqr, AR Sco and RX J0648. The parameters for the systems are listed in Table 1.

From Eq. (4), one notes that the amplitude depends on the amount of mass accumulated; however, it is not easy to predict how much matter may have been accreted to WD and how much has been dispersed on its surface. Thus, for this analysis, we assign four values for the mountain of matter for the analyzed systems: $\delta m = (10^{-3} M_\odot,\ 10^{-4} M_\odot,\ 10^{-5} M_\odot,\ 10^{-6} M_\odot)$ [see Refs. 41–43 for details about accretion in WDs]. Besides that, we consider that the angle between the magnetic and rotation axes is $\theta = 30°$.

Table 1. Parameters of 3 binary systems: Period (P), spindown ($\dot{P}$), WD radius (R) and distance to Earth (r).

Systems	P (s)	$\dot{P}$ (10^{-15} s/s)	R (10^8 cm)	r (pc)
AE Aqr[a]	33.08	56.4	7.0	100
AR Sco[b]	118.2	392	7.1	116
RX J0648[c]	13.18	6.0	3.0	650

Note: [a]see 44; [b]see 34; [c]see 45.

Assuming these values for δm and the parameters listed in Table 1, we obtain $h_{0_{ac}}$ for the three systems, which are shown in Figure 1.

With the amplitude estimations, we compare them with the sensitivity curves of the gravitational wave space detectors. This outcome is shown in Figure 2 where we have the sensitivity curve for LISA, BBO and DECIGO with a signal-to-noise ratio (SNR) of 8 and an integration time of $T = 1$ yr. We notice from this figure that AE Aqr and RX J0648 are good candidates to be detected by BBO and DECIGO if they have an accumulated mass of $\delta m \geq 10^{-5} M_\odot$. For the AR Sco system, the gravitational radiation emitted by this process would hardly be able to be detected by the three space instruments. This system would need to have a very high mass mountain of $\sim 10^{-3} M_\odot$ to be above, for example, the sensitivity curve of the BBO detector.

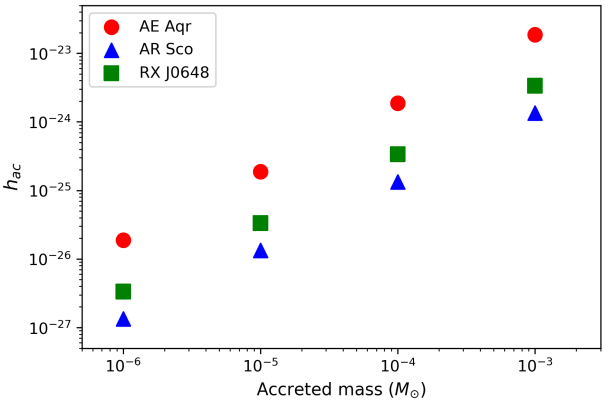


Fig. 1. GW amplitude as a function of accreted mass to AE Aqr, AR Sco and RX J0648.

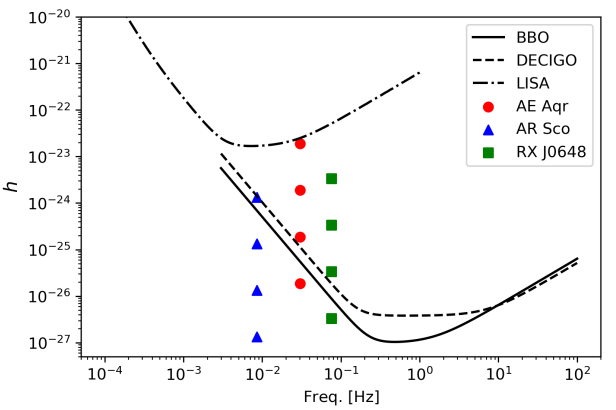


Fig. 2. GW amplitude for AE Aqr, AR Sco and RX J0648 for different values of mass ($10^{-3}M_\odot$, $10^{-4}M_\odot$, $10^{-5}M_\odot$, $10^{-6}M_\odot$, from top to bottom) and the sensitivity curves for LISA, BBO and DECIGO for $SNR = 8$ and integration time of $T = 1$ yr.

Now, we consider the efficiency of this mechanism with respect to the rotational energy rate lost by the systems. To do this, we consider the efficiency of the process ($\eta_{acr} = L_{GW_{acr}}/L_{sd}$) for the four δm's considered above, i.e., how much of the spin-down luminosity is converted to gravitational luminosity for every δm (see Table 2). We find that the contribution of gravitational luminosity to the spindown luminosity is irrelevant, since, for the four values of δm adopted, the efficiency $\eta_{acr} \ll 1$. Thereby, the contribution of gravitational luminosity to the spindown luminosity is negligible when we consider this mechanism.

Table 2. The efficiency of the mechanism of GWs due to the amount of mass accumulated at the WD magnetic poles for different values of δm.

AE Aquarii		**AR Scorpii**		**RX J0648**	
δm	η_{acr}	δm	η_{acr}	δm	η_{acr}
$(M_\odot)$	$(L_{GW_{acr}}/L_{sd})$	$(M_\odot)$	$(L_{GW_{acr}}/L_{sd})$	$(M_\odot)$	$(L_{GW_{acr}}/L_{sd})$
10^{-3}	1.02×10^{-2}	10^{-3}	3.41×10^{-5}	10^{-3}	0.175
10^{-4}	1.02×10^{-4}	10^{-4}	3.41×10^{-7}	10^{-4}	1.75×10^{-3}
10^{-5}	1.02×10^{-6}	10^{-5}	3.41×10^{-9}	10^{-5}	1.75×10^{-5}
10^{-6}	1.02×10^{-8}	10^{-6}	3.41×10^{-11}	10^{-6}	1.75×10^{-7}

3.2. *Magnetic deformation*

Here, we consider the generation of GWs due to the deformation of the WD structure of the same binary systems (AE Aqr, AR Sco and RX J0648) caused by their own intense magnetic field. For this, we use Eq. (12) to calculate the GW amplitude as a function of the efficiency $\eta_{df} = L_{GW_{def}}/L_{sd}$. The GW amplitudes are shown in Figure 3 as a function of η_{df}, where we use the parameters of Table 1 for all 3 systems.

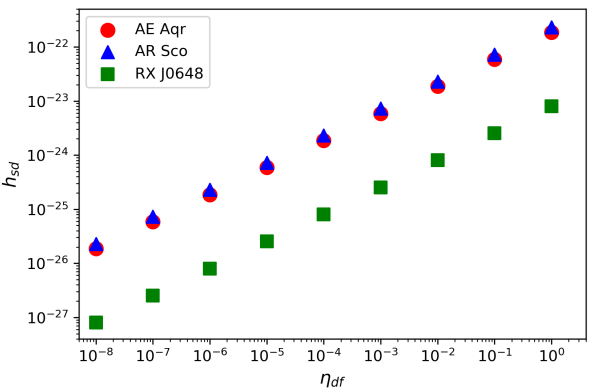


Fig. 3. GW amplitude for different values of efficiency ($\eta_{df} = L_{GW_{def}}/L_{sd}$) to AE Aqr, AR Sco and RX J0648.

From Figure 4, we plot the GW amplitudes inferred in Figure 3, together with sensitivity curves for LISA, BBO and DECIGO for one year of integration time and $SNR = 8$. It is worth noting that all three systems are detectable by BBO and DECIGO as long as AE Aqr has an efficiency $\eta_{df} \geq 10^{-6}$, AR Sco an efficiency $\eta_{df} \geq 10^{-4}$ and RX J0648 an efficiency $\eta_{df} \geq 10^{-5}$. Thus, even if the GWs have a small contribution to the spindown of these systems, they can emit GWs with amplitudes that can be detected by the space antennas.

Nevertheless, it is interesting to know what the value of the magnetic field needed to produce these detectable amplitudes. Thus, we calculate the magnetic

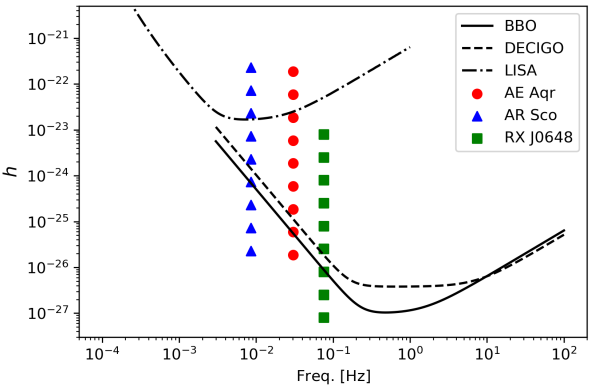


Fig. 4. GW amplitudes as presented in Figure 3 compared to the sensitivity curves of LISA, BBO and DECIGO for $SNR = 8$ and integration time of $T = 1$ year. Here, the efficiency values (1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7} and 10^{-8}) are displayed from top to bottom.

field strength so that these sources can be detected by BBO, which is the most sensitive instrument of the three considered in the present work. To do so, we use Eq. (10) together with the minimum efficiency for which each system is detectable by this instrument and we adopted $\kappa \simeq 10$ [see Ref. 46]. Table 3 shows theses results. Notice that the systems must have WDs with high magnetic fields, around $(10^9 - 10^{10})$ G, which are about two orders of magnitude larger than the canonical model of WD pulsars.

In addition, we can further calculate the GW amplitude by considering the upper limit values of the magnetic field strength, B_{dip}, inferred from the canonical model of WD pulsars. Table 4 presents the results of this study. Notice that the amplitudes of the GW shown in this Table is very small to be observed by the space detectors, since they are well below the sensitivities of these detectors. In other words, the space detectors will not be able to detect these sources when considering the upper limit of the magnetic field strength.

Table 3. Minimum efficiency for the sources to be detected by the BBO detector along with the required magnetic field strength.

Minimum efficiency detected by BBO			
Systems	η_{df}	h_{def}	B (G)
AE Aqr	10^{-6}	1.9×10^{-25}	2.8×10^{9}
AR Sco	10^{-4}	2.3×10^{-24}	3.6×10^{10}
RX J0648	10^{-5}	2.5×10^{-26}	1.6×10^{10}

3.3. *GWs from SGRs/AXPs as fast-spinning WDs*

An alternative model has been proposed for SGRs/AXPs considering they are fast-rotating and magnetized WDs [see e.g. Refs. 17–20 for further details]. From this

Table 4. Elipticity (ϵ), GW amplitude (h_{def}), GW luminosity ($L_{GW_{def}}$) and efficiency of the mechanism (η_{df}) for the upper limit of magnetic field (B_{dip}) of each system.

SYSTEMS	B_{dip} (G)	ϵ	h_{def}	$L_{GW_{def}}$ (erg/s)	η_{df}
AE Aqr	5.0×10^{7}	5.1×10^{-9}	6.2×10^{-29}	2.13×10^{21}	1.1×10^{-13}
AR Sco	5.0×10^{8}	5.3×10^{-7}	4.6×10^{-28}	1.25×10^{22}	4.02×10^{-12}
RX J0648	1.0×10^{8}	2.8×10^{-10}	9.5×10^{-31}	1.33×10^{20}	1.4×10^{-14}

perspective, a canonical spin-powered pulsar model can explain the process of energy emission released by dipole radiation in a WD, since they share quite similar aspects.[21] In addition, these sources could also be candidates for GW emission, since the high magnetic field can deform the star in a non-symmetrical way, thus generating a variation in the quadrupolar moment of the star.

Therefore, we consider in this section that SGRs/AXPs are fast-spinning and magnetized WDs which emit GWs due to the deformation caused by their own intense magnetic field. Thus, using Eq. 10 and adopting $\kappa \simeq 10$, we calculate the GW amplitude for the 23 confirmed SGRs/AXPs[47] [a], considering these objects as a very massive WD. To do so, we assume three values of mass and their corresponding radius, namely, $M_{WD} = 1.4 M_{\odot}$ ($R_{WD} = 1.0 \times 10^{8}$ cm), 1.2 $M_{\odot}$ ($R_{WD} = 6.0 \times 10^{8}$ cm) and 1.0 $M_{\odot}$ ($R_{WD} = 7.5 \times 10^{8}$ cm) [see Ref. 48 for further details about the mass-radius relation].

After estimating the amplitude, we placed them on the sensitivity curves of the BBO and DECIGO detectors. Figure 5 shows theses results, such that the GW amplitude is presented as a function of frequency for some SGRs/AXPs. In this Figure, the bullets stands for $M_{WD} = 1.2 M_{\odot}$ and the vertical bars, that crosses the bullets, stands for $1.0 M_{\odot} \leq M_{WD} \leq 1.4 M_{\odot}$, from top to bottom.

Notice that some SGRs/AXPs produce GWs with amplitudes that can be detected by BBO and DECIGO. Some of them, for example 1E 1547.0-5408 and SGR 1806-20, could well be detected for the entire mass range considered, while others would be detected depending on how massive they are.

SGRs/AXPs described as WDs generate GW amplitudes much larger than SGRs/AXPs described in the magnetar model where they are neutron stars [see e.g. Refs. 49, 50 for details about magnetar model]. This is because WDs have moments of inertia four orders of magnitude greater than a neutron star. Consequently, the GW amplitudes generated by these sources as neutron stars are far below the sensitivity curves of BBO and DECIGO [see Fig. 6]. Therefore, if these space based instruments detect continuous GWs from these sources, this would corroborate the model of fast spinning and magnetic WDs.

[a]For information about the SGRs/AXPs, we refer the reader to the McGill University's online catalog available at: `http://www.physics.mcgill.ca/~pulsar/magnetar/main.html`

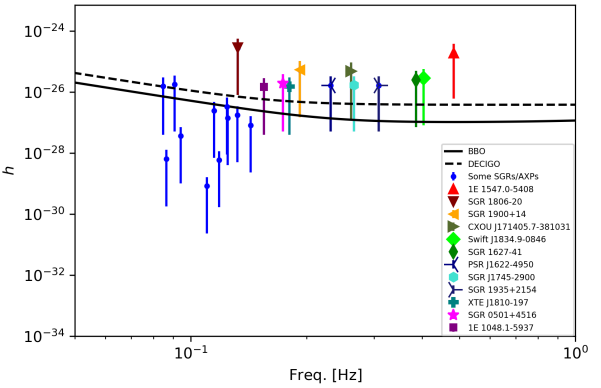


Fig. 5. GW amplitude as a function of frequency for SGRs/AXPs as fast-spinning and magnetized WDs for masses in the interval $1.0M_{\odot} \leq M_{WD} \leq 1.4M_{\odot}$, represented by the vertical bars, from top to bottom. The bullets stand for $M_{WD} = 1.2M_{\odot}$. Also plotted the sensitivity curves for BBO and DECIGO for $SNR = 8$ and integration time $T = 1$ year.

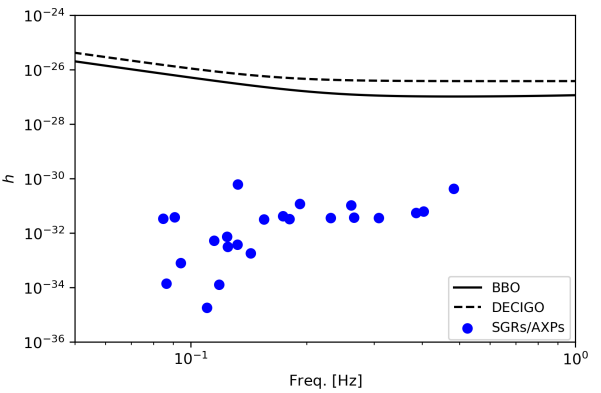


Fig. 6. GW amplitude as a function of frequency for SGRs/AXPs as NSs. Also plotted the sensitivity curves for BBO and DECIGO for $SNR = 8$ and integration time of $T = 1$ year. We consider a NS of $M = 1.4M_{\odot}$, radius $R = 10$ km.

4. Final remarks

We investigate two mechanisms - accretion of matter and magnetic deformation - for the production of gravitational waves in fast-spinning WDs. These uncommon WDs have high rotation and huge magnetic fields. Also, these stars are considered in an alternative model to describe SGRs and AXPs, where they are characterized as rotation-powered WD pulsars.

Then, we firstly study the following three binary systems: AE Aqr, AR Sco and RX J0648. For the accretion of matter mechanism, our results show that the AE Aqr and RX J0648 are good candidates for BBO and DECIGO if they have an

amount of mass accumulated of $\delta m \geq 10^{-5} M_{\odot}$, considering 1 year of integration time and $SNR = 8$. In contrast, AR Sco is unlikely to be detected because it is needed a large amount of mass accumulated in the magnetic pole of this WD.

Now, regarding the magnetic deformation mechanism, to emit gravitational radiation with amplitudes that are detectable by BBO, for example, the three binary systems studied require that the WD has a magnetic field above $\sim 10^9$ G. Nevertheless, these WDs are inferred to have magnetic fields with intensity around two orders of magnitude smaller. Moreover, we also conclude that gravitational radiation has an irrelevant contribution to the spindown luminosity of these systems for both mechanism.

Still, taking into account the magnetic deformation mechanism, we investigate the SGRs/AXPs as rotation-powered WD pulsars assigning a mass range $1.0 M_{\odot} \leq M_{WD} \leq 1.4 M_{\odot}$ for these objects. We conclude that a possible detection of continuous GWs coming from SGRs/AXPs would be a good indication that could corroborate the WD pulsar model, as for the neutron stars description, they are far below the BBO and DECIGO sensitivity curves.

Acknowledgements

M.F.S. thanks CAPES for the financial support. J.C.N.A. thanks FAPESP (2013/26258-4) and CNPq (308367/2019-7) for partial financial support. J.G.C. is likewise grateful to the support of CNPq (421265/2018-3 and 305369/2018-0), FAPESP Project No. 2015/15897-1, and NAPI (Fenômenos Extremos do Universo) of Fundação de Apoio à Ciência, Tecnologia e Inovação do Paraná.

References

1. M. F. Sousa, J. G. Coelho and J. C. de Araujo, Gravitational waves from fast-spinning white dwarfs, *Monthly Notices of the Royal Astronomical Society* **492**, 5949 (2020).
2. M. F. Sousa, J. G. Coelho and J. C. de Araujo, Gravitational waves from SGRs and AXPs as fast-spinning white dwarfs, *Monthly Notices of the Royal Astronomical Society* **498**, 4426 (2020).
3. B. Külebi, S. Jordan, F. Euchner, B. T. Gänsicke and H. Hirsch, Analysis of hydrogen-rich magnetic white dwarfs detected in the Sloan Digital Sky Survey, *Astronomy and Astrophysics* **506**, 1341 (November 2009).
4. S. O. Kepler, I. Pelisoli, S. Jordan, S. J. Kleinman, D. Koester, B. Külebi, V. Peçanha, B. G. Castanheira, A. Nitta, J. E. S. Costa, D. E. Winget, A. Kanaan and L. Fraga, Magnetic white dwarf stars in the Sloan Digital Sky Survey, *Monthly Notices of the RAS* **429**, 2934 (March 2013).
5. S. O. Kepler, I. Pelisoli, D. Koester, G. Ourique, S. J. Kleinman, A. D. Romero, A. Nitta, D. J. Eisenstein, J. E. S. Costa, B. Külebi, S. Jordan, P. Dufour, P. Giommi and A. Rebassa-Mansergas, New white dwarf stars in the Sloan Digital Sky Survey Data Release 10, *Monthly Notices of the RAS* **446**, 4078 (February 2015).
6. G. D. Schmidt, P. Bergeron, J. Liebert and R. A. Saffer, Two ultramassive white dwarfs found among candidates for magnetic fields, *Astrophysical Journal* **394**, 603 (August 1992).

7. B. Külebi, S. Jordan, E. Nelan, U. Bastian and M. Altmann, Constraints on the origin of the massive, hot, and rapidly rotating magnetic white dwarf RE J 0317-853 from an HST parallax measurement, *Astronomy and Astrophysics* **524**, p. A36 (December 2010).
8. T. R. Marsh, B. T. Gänsicke, S. Hümmerich, F.-J. Hambsch, K. Bernhard, C. Lloyd, E. Breedt, E. R. Stanway, D. T. Steeghs, S. G. Parsons, O. Toloza, M. R. Schreiber, P. G. Jonker, J. van Roestel, T. Kupfer, A. F. Pala, V. S. Dhillon, L. K. Hardy, S. P. Littlefair, A. Aungwerojwit, S. Arjyotha, D. Koester, J. J. Bochinski, C. A. Haswell, P. Frank and P. J. Wheatley, A radio-pulsing white dwarf binary star, *Nature* **537**, 374 (September 2016).
9. Y. Terada, T. Hayashi, M. Ishida, K. Mukai, T. Dotani, S. Okada, R. Nakamura, S. Naik, A. Bamba and K. Makishima, Suzaku Discovery of Hard X-Ray Pulsations from a Rotating Magnetized White Dwarf, AE Aquarii, *Publications of the ASJ* **60**, 387 (April 2008).
10. S. Mereghetti, A. Tiengo, P. Esposito, N. La Palombara, G. L. Israel and L. Stella, An Ultramassive, Fast-Spinning White Dwarf in a Peculiar Binary System, *Science* **325**, p. 1222 (September 2009).
11. S. Mereghetti, F. Pintore, P. Esposito, N. La Palombara, A. Tiengo, G. L. Israel and L. Stella, Discovery of spin-up in the x-ray pulsar companion of the hot subdwarf HD 49798, *Monthly Notices of the Royal Astronomical Society* **458**, 3523 (2016).
12. S. Popov, S. Mereghetti, S. Blinnikov, A. Kuranov and L. Yungelson, A young contracting white dwarf in the peculiar binary HD 49798/RX J0648. 0-4418?, *Monthly Notices of the Royal Astronomical Society* **474**, 2750 (2017).
13. R. Lopes de Oliveira, A. Bruch, C. V. Rodrigues, A. S. Oliveira and K. Mukai, CTCV J2056-3014: An X-Ray-faint Intermediate Polar Harboring an Extremely Fast-spinning White Dwarf, *The Astrophysical Journal Letters* **898**, p. L40 (August 2020).
14. E. Otoniel, J. G. Coelho, M. Malheiro and F. Weber, Mass limits of the extremely fast-spinning white dwarf CTCV J2056-3014, *arXiv e-prints*, p. arXiv:2010.12441 (October 2020).
15. R. P. Ashley, T. R. Marsh, E. Breedt, B. T. Gänsicke, A. F. Pala, O. Toloza, P. Chote, J. R. Thorstensen and M. R. Burleigh, V1460 Her: A fast spinning white dwarf accreting from an evolved donor star, *Monthly Notices of the Royal Astronomical Society* **499**, 149 (September 2020).
16. I. Caiazzo, K. B. Burdge, J. Fuller, J. Heyl, S. Kulkarni, T. A. Prince, H. B. Richer, J. Schwab, I. Andreoni, E. C. Bellm *et al.*, A highly magnetized and rapidly rotating white dwarf as small as the moon, *Nature* **595**, 39 (2021).
17. M. Malheiro, J. A. Rueda and R. Ruffini, SGRs and AXPs as rotation-powered massive white dwarfs, *Publications of the Astronomical Society of Japan* **64**, p. 56 (2012).
18. J. G. Coelho and M. Malheiro, Magnetic dipole moment of soft gamma-ray repeaters and anomalous x-ray pulsars described as massive and magnetic white dwarfs, *Publications of the Astronomical Society of Japan* **66**, p. 14 (2014).
19. R. V. Lobato, M. Malheiro and J. G. Coelho, Magnetars and white dwarf pulsars, *International Journal of Modern Physics D* **25**, p. 1641025 (July 2016).
20. B. Mukhopadhyay and A. R. Rao, Soft gamma-ray repeaters and anomalous X-ray pulsars as highly magnetized white dwarfs, *Journal of Cosmology and Astroparticle Physics* **5**, p. 007 (May 2016).
21. V. V. Usov, Gamma-radiation generation by rotating magnetic white dwarfs, *Pisma v Astronomicheskii Zhurnal* **14**, 606 (1988).
22. B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Gwtc-1: A gravitational-wave transient catalog

of compact binary mergers observed by ligo and virgo during the first and second observing runs, *Physical Review X* **9**, p. 031040 (2019).
23. R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run, *Physical Review X* **11**, p. 021053 (2021).
24. P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi *et al.*, Laser interferometer space antenna, *arXiv preprint arXiv:1702.00786* (2017).
25. T. Robson, N. J. Cornish and C. Liug, The construction and use of LISA sensitivity curves, *Classical and Quantum Gravity* **36**, p. 105011 (2019).
26. G. M. Harry, P. Fritschel, D. A. Shaddock, W. Folkner and E. S. Phinney, Laser interferometry for the big bang observer, *Classical and Quantum Gravity* **23**, p. 4887 (2006).
27. K. Yagi and N. Seto, Detector configuration of DECIGO/BBO and identification of cosmological neutron-star binaries, *Physical Review D* **83**, p. 044011 (2011).
28. S. Kawamura *et al.*, The japanese space gravitational wave antenna—DECIGO, *Classical and Quantum Gravity* **23**, p. S125 (2006).
29. K. Yagi and N. Seto, Erratum: Detector configuration of DECIGO/BBO and identification of cosmological neutron-star binaries [phys. rev. d 83, 044011 (2011)], *Physical Review D* **95**, p. 109901 (2017).
30. S. Bonazzola and E. Gourgoulhon, Gravitational waves from pulsars: Emission by the magnetic-field-induced distortion, *Astronomy and Astrophysics* **312**, 675 (August 1996).
31. J. C. N. de Araujo, J. G. Coelho and C. A. Costa, Gravitational wave emission by the high braking index pulsar PSR J1640-4631, *Journal of Cosmology and Astroparticle Physics* **2016**, p. 023 (Jul 2016).
32. J. C. N. de Araujo, J. G. Coelho and C. A. Costa, Gravitational Waves from Pulsars and Their Braking Indices: The Role of a Time Dependent Magnetic Ellipticity, *Astrophysical Journal* **831**, p. 35 (Nov 2016).
33. J. C. N. de Araujo, J. G. Coelho and C. A. Costa, Gravitational waves from pulsars in the context of magnetic ellipticity, *European Physical Journal C* **77**, p. 350 (May 2017).
34. B. Franzon and S. Schramm, Ar scorpii and possible gravitational wave radiation from pulsar white dwarfs, *Monthly Notices of the Royal Astronomical Society* **467**, 4484 (2017).
35. S. Kalita and B. Mukhopadhyay, Continuous gravitational wave from magnetized white dwarfs and neutron stars: Possible missions for LISA, DECIGO, BBO, ET detectors, *Monthly Notices of the Royal Astronomical Society*, p. 2346 (Oct 2019).
36. S. Chandrasekhar and E. Fermi, Problems of gravitational stability in the presence of a magnetic field, *The Astrophysical Journal* **118**, 116 (1953).
37. S. L. Shapiro and S. A. Teukolsky, *Black holes, white dwarfs and neutron stars: The physics of compact objects* (John Wiley & Sons, 1983).
38. M. Maggiore, *Gravitational waves: volume 1: Theory and experiments* (OUP Oxford, 2008).
39. K. Konno, T. Obata and Y. Kojima, Flattening modulus of a neutron star by rotation and magnetic field, *Astronomy and Astrophysics* **356**, 234 (April 2000).
40. T. Regimbau and J. A. de Freitas Pacheco, Gravitational wave background from magnetars, *Astronomy and Astrophysics* **447**, 1 (February 2006).

41. W. F. Welsh, K. Horne and R. Gomer, Doppler signatures of Hα flares in AE aquarii, *Monthly Notices of the Royal Astronomical Society* **298**, 285 (1998).
42. B. Warner, *Cataclysmic Variable Stars* (Cambridge University Press, 2003).
43. C. Hellier, *Cataclysmic Variable Stars—How and why they vary* (Springer Science & Business Media, 2001).
44. C.-S. Choi and I. Yi, On the rapid spin-down and low-luminosity pulsed emission from AE aquarii, *The Astrophysical Journal* **538**, 862 (2000).
45. S. Mereghetti, N. La Palombara, A. Tiengo, F. Pizzolato, P. Esposito, P. Woudt, G. Israel and L. Stella, X-ray and optical observations of the unique binary system hd 49798/rx j0648. 0-4418, *The Astrophysical Journal* **737**, p. 51 (2011).
46. V. C. A. Ferraro, On the Equilibrium of Magnetic Stars, *Astrophysical Journal* **119**, p. 407 (March 1954).
47. S. Olausen and V. Kaspi, The mcgill magnetar catalog, *The Astrophysical Journal Supplement Series* **212**, p. 6 (2014).
48. K. Boshkayev, L. Izzo, J. A. Rueda and R. Ruffini, Sgr 0418+5729, swift j1822.3-1606, and 1e 2259+586 as massive fast rotating highly magnetized white dwarfs, *Astronomy and Astrophysics* **555** (05 2013).
49. R. C. Duncan and C. Thompson, Formation of very strongly magnetized neutron stars-implications for gamma-ray bursts, *The Astrophysical Journal* **392**, L9 (1992).
50. V. M. Kaspi and A. M. Beloborodov, Magnetars, *Annual Review of Astronomy and Astrophysics* **55**, 261 (2017).

Highly magnetized white dwarfs: Implications and current status

B. Mukhopadhyay[1,*], M. Bhattacharya[2], A. J. Hackett[3], S. Kalita[1], D. Karinkuzhi[1] and C. A. Tout[3]

[1] *Department of Physics, Indian Institute of Science, Bangalore 560012, India*
[*] *E-mail: bm@iisc.ac.in*
[2] *Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA*
[3] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

Over the last decade or so, we have been developing the possible existence of highly magnetized white dwarfs with analytical stellar structure models. While the primary aim was to explain the nature of the peculiar overluminous type Ia supernovae, later on, these magnetized stars were found to have even wider ranging implications including those for white dwarf pulsars, soft gamma-ray repeaters and anomalous X-ray pulsars, as well as gravitational radiation. In particular, we have explored in detail the mass-radius relations for these magnetized stars and showed that they can be significantly different from the Chandrasekhar predictions which essentially leads to a new super-Chandrasekhar mass-limit. Recently, using the stellar evolution code STARS, we have successfully modelled their formation and cooling evolution directly from the magnetized main sequence progenitor stars. Here we briefly discuss all these findings and conclude with their current status in the scientific community.

Keywords: Conduction; equation of state; magnetic fields; magnetohydrodynamics; opacity; radiative transfer; white dwarfs.

1. Introduction

More than a dozen overluminous Type Ia supernovae (SNe Ia) have already been observed[1,2] which strongly suggest the existence of massive progenitors with masses $M > 2M_{\odot}$. Although the evolutions of accreting or rapidly differentially rotating white dwarf (WD) binaries have been used to explain such progenitors,[3,4] these models are unable to explain masses up to $2.8M_{\odot}$ that are inferred from the observations. Highly magnetised super-Chandrasekhar WDs (B-WDs) have been recently proposed as the possible progenitors of these peculiar overluminous SNe Ia. In addition to SNe Ia, B-WDs are also considered as promising candidates for soft gamma-ray repeaters (SGRs) and anomalous X-ray pulsars (AXPs) at lower magnetic fields than neutron star (NS) based magnetar, satisfying their ultraviolet luminosity cut-off.[13]

It has been shown that strong magnetic fields can modify the equation of state (EoS) of electron degenerate matter and yield super-Chandrasekhar WDs,[6–8] irrespective of the rotation rate. Indeed, the observational data from the Sloan Digital Sky Survey (SDSS) suggest that magnetized WDs tend to have larger masses than

their non-magnetic counterparts, even though they span the same effective temperature range.[9,10] Inspired by above findings, the effect of strong magnetic fields on the stellar structure, for various field configurations, has been explored for both Newtonian[6] and general relativistic formalisms.[8,11,12]

Magnetized WDs have many important implications apart from their link to peculiar SNe Ia and hence their other properties are worth exploring.[13–15] Recent works[16–19] have shown that B-WDs can be too dim to detect directly, and have also explored[20,21] the ability of rotating B-WDs to generate gravitational radiation which can be detected by the space-based gravitational wave detectors. Furthermore, other additional physics such as modified gravity,[22,23] ungravity effect,[24] effects of net charge,[25,26] lepton number violation[27] and anisotropic pressure[28] have also been explored to show the possible existence of super-Chandrasekhar WDs. Here we discuss the broad implications of these magnetized stars as well as their current status.

2. Origin and evolution of strong magnetic fields

It has been well known that purely poloidally or toroidally dominated fields are both structurally unstable.[29,30] However, it also has been shown that magnetized WDs with toroidally dominated mixed field configuration (along with small poloidal component) are one of the most plausible cases[31] and have approximately spherical shape.[8] Although the surface fields can be observationally inferred, the interior field cannot be directly constrained. Nevertheless, there is a sufficient evidence that the stars exhibit dipolar fields in their outer regions and hence are expected to have stronger interior (toroidally dominated) fields than at the surface. Numerical simulations have indeed shown that central fields of B-WDs can be several orders of magnitude higher than the surface field.[8,14,32]

The evolution of the poloidal and toroidal magnetic field components along with the angular momentum has been modelled recently with the Cambridge stellar evolution code STARS,[32] using advection-diffusion equations coupled to the structural and compositional equations of stars. They have shown that the magnetic field is likely to be dipolar, decaying as an inverse square law for most of the star. Their results also suggest that at the end of main sequence, the star may have toroidally dominated magnetic fields. The left panel of Figure 1 shows the evolution of toroidal field in the stellar interior as a function of radius at the end of main sequence, while the right panel shows the field as a function of the mass coordinate at various times after the helium exhaustion in the core, during the asymptotic giant phase.

Large-scale magnetic fields can be governed in the degenerate core of B-WDs even during the late stages of stellar evolution,[32] and very high fields can develop in these stars based on the conservation of magnetic flux, besides from the dynamo mechanism. Hence, strong fields inside magnetized WDs can also be of fossil origin.[33] This can be understood as follows. While the mass of the WD increases due to accretion, magnetic field is advected into its interior. Consequently, the gravitational

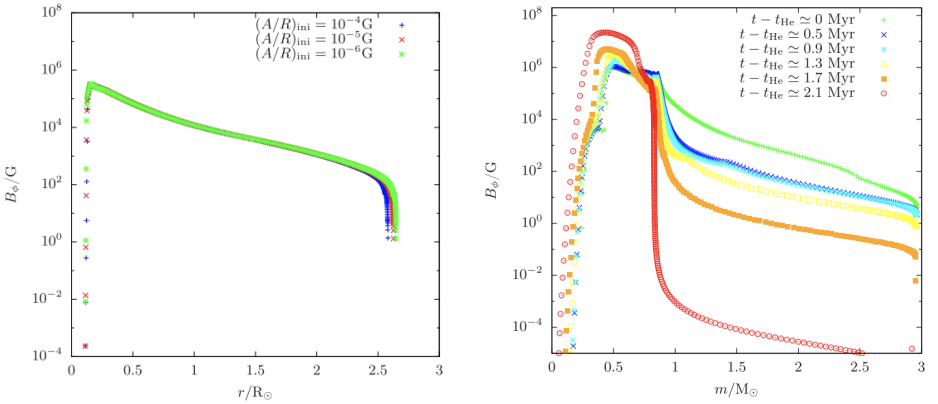


Fig. 1. *Left panel:* Toroidal field in the interior of the star as a function of the radius at the end of the main sequence. *Right panel:* Toroidal field inside the star as a function of the mass coordinate at various times after the helium exhaustion in the core, during the asymptotic giant phase.

power dominates over the degeneracy pressure leading to the contraction of the star. The initial seed magnetic field is then amplified as the total magnetic flux remains conserved. For magnetic field $B \sim 10^8$ G in a star of size $R \sim 10^6$ km, the resultant flux will be $\sim 10^{20}$ G km^2. From flux freezing, for a 1000 km size B-WD, the magnetic field can then grow up to $\sim 10^{14}$ G. Once the field increases, the total outward force further builds up to balance the inward gravitational force and the whole cycle is repeated multiple times. Therefore, the magnetic fields of highly magnetized WDs are likely to be fossil remnants from their main-sequence progenitor stars.

Repeated episodes of accretion and spin-down have also been suggested as a plausible mechanism that can lead to a highly magnetized WD.[14] Here, the entire evolution of the B-WD can be classified in two phases: accretion-powered and rotation-powered. The accretion-powered phase is governed by three conservation laws: linear and angular momenta conservation and conservation of magnetic flux, around the stellar surface, given by

$$\begin{aligned} l\Omega(t)^2 R(t) &= \frac{GM(t)}{R(t)^2}, \\ I(t)\Omega(t) &= \text{constant}, \\ B_s(t)R(t)^2 &= \text{constant}, \end{aligned} \tag{1}$$

where l accounts for the dominance of gravitational force over the centrifugal force, hence $l > 1$, I is the stellar moment of inertia and Ω is the angular velocity of the star that includes contribution acquired due to accretion as well. Solving these equations simultaneously gives the time evolution of radius, magnetic field and angular velocity during the accretion phase. Accretion discontinues when

$$-\frac{GM}{R^2} = \frac{1}{\rho}\frac{d}{dr}\left(\frac{B^2}{8\pi}\right)_{r=R} \sim -\frac{B_s^2}{8\pi R\rho}, \tag{2}$$

where ρ is the density of the inner disk edge. If the magnetic field is dipolar, $\dot{\Omega} \propto \Omega^3$ for a fixed magnetic field. Generalizing it to $\dot{\Omega} = k\Omega^n$ with constant k giving for the spin-powered phase, we obtain

$$\Omega = [\Omega_0^{1-n} - k(1-n)(t-t_0)]^{1/1-n}, \tag{3}$$

$$\text{and } B_s = \sqrt{\frac{5c^3 I k \Omega^{n-m}}{R^6 \sin^2\alpha}}. \tag{4}$$

Here Ω_0 is the initial angular velocity for the spin-powered phase (once accretion stops) at time $t = t_0$. The value of k is fixed such that B_s can be constrained at $t = t_0$, which is known from the field evolution in the preceding accretion-powered phase. Here $n = m = 3$ corresponds to the dipole field configuration, therefore m represents the deviation from dipolar field, especially for $n = 3$. Figure 2 shows the sample evolutions of angular velocity and magnetic field as functions of stellar mass. In both cases, initially larger Ω with accretion drops significantly during the spin-powered phase, followed by a phase of its increasing trend. At the end of the evolution, the star can be left either as a super-Chandrasekhar WD and/or an SGR/AXP candidate with a higher spin frequency.

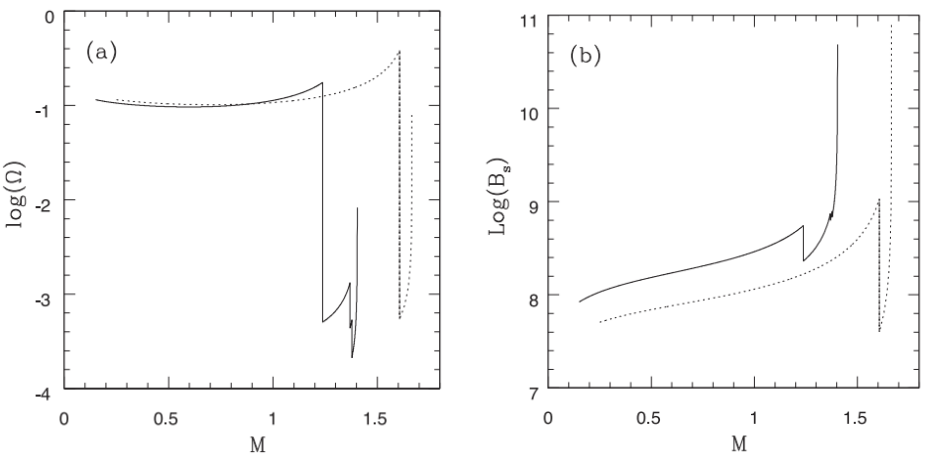


Fig. 2. Time evolution of angular velocity (left) and magnetic field (right) as functions of mass. The solid curves correspond to the case with $n = 3$, $m = 2.7$, $\rho = 0.05\,\mathrm{g\,cm^{-3}}$, $l = 1.5$ and dotted curves correspond to the case with $n = 3$, $m = 2$, $\rho = 0.1\,\mathrm{g\,cm^{-3}}$, $l = 2.5$. The other parameters are fixed with $k = 10^{-14}$, $\dot{M} = 10^{-8}\,M_\odot \mathrm{yr}^{-1}$, $\alpha = 10^\circ$ and $R = 10^4$ km at $t = 0$.

3. Rotating magnetized white dwarfs

Although, in nature, WDs are expected to consist of mixed fixed geometry, here we consider toroidally dominated magnetic field cases as they ensure the stability of these stars. It has been shown that toroidally dominated (and purely toroidal) field not only makes the star (slightly) prolate but also increases its equatorial

radius.[8,34,35] Figure 3 shows specific cases for toroidal field configuration combined with rotation. In the left panel, as the angular frequency is small, it does not affect the star considerably and results in a marginally prolate star. In contrast, the right panel, due to high angular velocity, exhibits that the low density region is affected more due to rotation than the high density region, resulting in an oblate shaped WD. From the magnetic field strength isocontours shown in the center panel, we can see that the surface magnetic field can decrease up to $\sim 10^9$ G even if the central field is $\sim 10^{14}$ G. For both cases, the magnetic to gravitational energies ratio (ME/GE) as well as kinetic to gravitational energies ratio (KE/GE) are chosen to be $\lesssim 0.1$ to maintain stable equilibrium.[36–38]

We have also considered differentially rotating B-WDs.[8] The angular velocity profile in the XNS code is specified as[39,40]

$$F(\Omega) = A^2(\Omega_c - \Omega) = \frac{R^2(\Omega - \omega)}{\alpha^2 - R^2(\Omega - \omega)^2}, \tag{5}$$

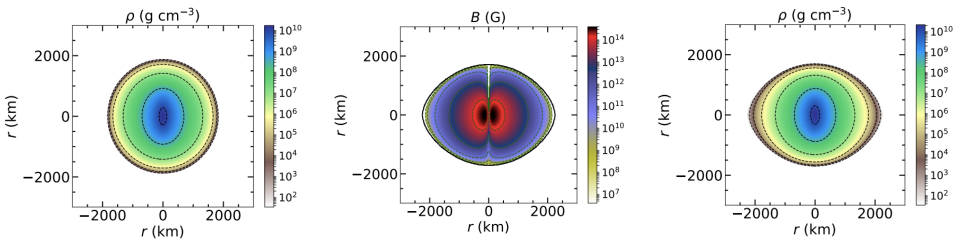


Fig. 3. Density and magnetic field strength isocontours for uniformly rotating B-WDs with toroidal magnetic field are shown in the left and center panels, respectively. The parameters used for both are $\Omega \sim 0.0628$ rad/s, $B_{max} \sim 2.7 \times 10^{14}$ G, ME/GE ~ 0.1, KE/GE $\sim 3.6 \times 10^{-6}$. The density isocontours for the $\Omega \sim 3.6537$ rad/s, $B_{max} \sim 2.7 \times 10^{14}$ G, ME/GE ~ 0.1, KE/GE $\sim 1.3 \times 10^{-2}$ case is shown in the right panel.

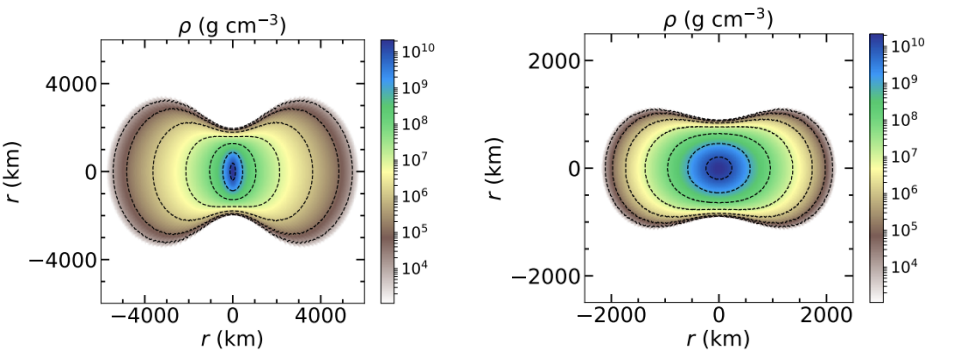


Fig. 4. Density isocontours for differentially rotating magnetized WD is shown. For the left panel, we have toroidal field with $\Omega \sim 0.62$ rad/s, $\Omega_c \sim 10.15$ rad/s, $B_{max} \sim 3.2 \times 10^{14}$ G, ME/GE ~ 0.14, KE/GE ~ 0.1. For the right panel, we use poloidal field with $\Omega \sim 2.74$ rad/s, $\Omega_c \sim 20.30$ rad/s, $B_{max} \sim 3.9 \times 10^{11}$ G, ME/GE $\sim 7.8 \times 10^{-8}$, KE/GE ~ 0.14.

where A is a constant that indicates the extent of differential rotation, $R = \psi^2 r\sin\theta$, $\omega = -\beta^\phi$, Ω_c is the angular velocity at the center and ω is the angular velocity at radius r. Figure 4 shows the density isocontours of differentially rotating B-WDs for toroidal field with surface angular velocity $\Omega \sim 0.62\,\mathrm{rad/s}$, $\Omega_c \sim 10.15\,\mathrm{rad/s}$, $B_{max} \sim 3.2 \times 10^{14}\,\mathrm{G}$, ME/GE ~ 0.14, KE/GE ~ 0.1 (left panel) and poloidal field with $\Omega \sim 2.74\,\mathrm{rad/s}$, $\Omega_c \sim 20.30\,\mathrm{rad/s}$, $B_{max} \sim 3.9 \times 10^{11}\,\mathrm{G}$, ME/GE $\sim 7.8 \times 10^{-8}$, KE/GE ~ 0.14 (right panel). It can be seen that "polar hollow" structure can form with differential rotation regardless of the specific geometry of the magnetic field.

4. Non-rotating magnetized WDs with finite temperature

Apart from increasing the limiting mass of WDs, strong magnetic fields can also influence the thermal properties such as luminosity, temperature gradient and cooling rate of the star.[16,18,19] In order to model such a WD, the total pressure inside the star is modelled by including the contributions from the degenerate electron gas, ideal gas and magnetic pressures. The interface is defined to be the radius where the contributions from electron degenerate core and outer ideal gas pressures are equal. The presence of strong fields gives rise to additional pressure $P_B = B^2/8\pi$ and density $\rho_B = B^2/8\pi c^2$ inside the magnetized WDs.[41]

Assuming the B-WD to be approximately spherical, the model equations for magnetostatic equilibrium, photon diffusion and mass conservation can be written within a Newtonian framework as

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}r}(P_{\mathrm{deg}} + P_{\mathrm{ig}} + P_B) &= -\frac{Gm(r)}{r^2}(\rho + \rho_B),\\
\frac{\mathrm{d}T}{\mathrm{d}r} &= -\max\left[\frac{3}{4ac}\frac{\kappa\rho}{T^3}\frac{L_r}{4\pi r^2}, \left(1 - \frac{1}{\gamma}\right)\frac{T}{P}\frac{\mathrm{d}P}{\mathrm{d}r}\right],\\
\frac{\mathrm{d}m}{\mathrm{d}r} &= 4\pi r^2(\rho + \rho_B).
\end{aligned} \tag{6}$$

Here we have ignored the magnetic tension terms for radially varying B. In these equations, P_{deg} and $P_{\mathrm{ig}} = \rho kT/\mu m_p$ are the electron degeneracy pressure and hence the ideal gas pressure respectively, ρ is the matter density, T is the temperature, $m(r)$ is the mass enclosed within radius r, κ is the radiative opacity, L_r is the luminosity at radius r, and γ is the adiabatic index of the gas.

The opacity in the surface layers of non-magnetised WD is approximated with the Kramers' formula, $\kappa = \kappa_0 \rho T^{-3.5}$, where $\kappa_0 = 4.34 \times 10^{24} Z(1 + X)\ \mathrm{cm^2\ g^{-1}}$, and X and Z are the mass fractions of hydrogen and heavy elements (other than hydrogen and helium) in the stellar interior respectively. To capture the radial variation of the field magnitude within the B-WD, we adopt a profile used extensively to model magnetized NSs and B-WDs,[11,42] given by

$$B\left(\frac{\rho}{\rho_0}\right) = B_{\mathrm{s}} + B_0\left[1 - \exp\left(-\eta\left(\frac{\rho}{\rho_0}\right)^{\gamma}\right)\right], \tag{7}$$

where B_s is the surface magnetic field, B_0 is a fiducial magnetic field, $\eta = 0.8$ and $\gamma = 0.9$ are dimensionless parameters along with $\rho_0 = 10^9\,\mathrm{g/cm^3}$ determining how the field decays from the core to the surface. The radial luminosity can be assumed to be constant so that $L_r = L$ as there is no hydrogen burning or other nuclear fusion reactions taking place within the WD core. We solve the differential equations by providing the surface density, mass and surface temperature as the boundary conditions.

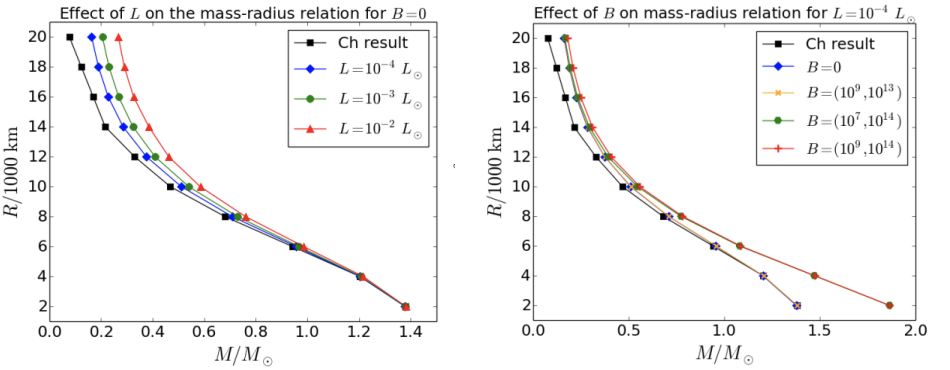


Fig. 5. *Left panel:* The effect of L on the mass–radius relation of non-magnetised WDs is shown for $L = 10^{-4}\ L_\odot$ (blue diamonds), $L = 10^{-3}\ L_\odot$ (green circles) and $L = 10^{-2}\ L_\odot$ (red triangles), along with the Chandrasekhar result (black squares). *Right panel:* The effect of field strength on the mass–radius relation of B-WDs is shown for $B = (B_s, B_0) = (0, 0)$ (blue diamonds), $B = (10^9, 10^{13})$ G (orange crosses), $B = (10^7, 10^{14})$ G (green circles) and $B = (10^9, 10^{14})$ G (red pluses), along with the Chandrasekhar result (black squares) for $L = 10^{-4}\ L_\odot$.

For strong magnetic fields, variation of radiative opacity with B can be modelled as $\kappa = \kappa_B \approx 5.5 \times 10^{31} \rho T^{-1.5} B^{-2} \mathrm{cm^2\ g^{-1}}$.[43,44] The field dependent Potekhin's opacity is used instead of Kramers' opacity if $B/10^{12}\ \mathrm{G} \geq T/10^6\ \mathrm{K}$, which is valid for the strong B cases that we consider here. The left panel of Figure 5 shows the effect of luminosity on the mass-radius relation for non-magnetized WDs compared to Chandrasekhar's results.[45] Although the increase in L leads to progressively higher masses for larger WDs, Chandrasekhar mass limit is retained irrespective of the luminosity. The right panel of Figure 5 shows the effect of magnetic field on the mass-radius relation for B-WDs with $L = 10^{-4}\,L_\odot$ and compares them with the non-magnetic Chandrasekhar results. It can be seen that the magnetic field affects the mass-radius relation in a manner analogous to increasing L, by shifting the curve towards higher masses for WDs with larger radii. The mass-radius curves for $B_0 \lesssim 10^{13}\,\mathrm{G}$ practically overlap with each other and retain the Chandrasekhar mass limit. However, for strong central fields with $B_0 \sim 10^{14}\,\mathrm{G}$, super-Chandrasekhar WDs are obtained with masses as high as $\sim 1.9\,M_\odot$.

In order to ensure structural stability of a B-WD, an increase in magnetic energy density has to be compensated by a corresponding decrease in the thermal energy

and hence the luminosity. This effect is especially prominent for B-WDs with larger radii where the magnetic, thermal and gravitational energies are comparable with each other. We find that a slight decrease in the luminosity for $R \gtrsim 12000\,\mathrm{km}$ WDs leads to masses that are similar to their non-magnetic counterparts. However, the smaller radii B-WDs require a substantial drop in their luminosity (well outside the observable range) and still do not really achieve masses that are similar to the non-magnetized WDs.

5. Effect of cooling evolution and field dissipation

Magnetic fields inside a WD undergo decay by Ohmic dissipation and Hall drift processes with timescales given by[46, 47]

$$t_{\rm Ohm} = (7 \times 10^{10}\ \mathrm{yr})\, \rho_{c,6}^{1/3} R_4^{1/2} (\rho_{\rm avg}/\rho_c), \tag{8}$$

$$t_{\rm Hall} = (5 \times 10^{10}\,\mathrm{yr})\ l_8^2 B_{0,14}^{-1} T_{c,7}^2 \rho_{\rm c,10}, \tag{9}$$

where $\rho_{\rm c,n} = \rho_c/10^n\,\mathrm{g\,cm^{-3}}$, $R_4 = R/10^4\,\mathrm{km}$, $T_{c,7} = T_c/10^7\,\mathrm{K}$, $B_{0,14} = B_0/10^{14}\,\mathrm{G}$ and $l = l_8 \times 10^8\,\mathrm{cm}$ is characteristic length scale of the flux loops through WD outer core. Ohmic decay is the dominant field dissipation process for $B \lesssim 10^{12}$ G, while for $10^{12} \leq B/\mathrm{G} \leq 10^{14}$ the decay occurs via Hall drift and for $B \gtrsim 10^{14}$ G, the principal decay mechanism is likely to be ambipolar diffusion.[46] The magnetic field decay in magnetars with surface fields between 10^{14} and 10^{16} G can be solved using

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -B\left(\frac{1}{t_{\rm Ohm}} + \frac{1}{t_{\rm Amb}} + \frac{1}{t_{\rm Hall}}\right), \tag{10}$$

where $t_{\rm Amb}$ denotes the ambipolar diffusion time scale.

On the other hand, the thermal energy of WDs is radiated away gradually over time in the observed luminosity from the surface layers as the star evolves. The rate at which thermal energy of ions is transported to surface and radiated depends on the specific heat, given by

$$L = -\frac{d}{dt}\int c_v dT = (2 \times 10^6\ \mathrm{erg/s})\, \frac{A m_\mu}{M_\odot}\left(\frac{T}{K}\right)^{7/2}, \tag{11}$$

where $c_v \approx 3k_B/2$ is the specific heat at constant volume. Given an initial L and temperature T_0 at time t_0, final temperature after cooling is given by $(T/\mathrm{K})^{-5/2} - (T_0/\mathrm{K})^{-5/2} = 2.406 \times 10^{-34}\,\tau/\mathrm{s}$, where $\tau = t - t_0$ is the WD age.

The left panel of Figure 6 shows the effect of B-WD evolution on their mass-radius relations including both magnetic field decay and thermal cooling effects. The luminosities are varied with field strength such that the masses can match those obtained for the non-magnetized WDs. For $B = (B_s, B_0) = (10^9, 10^{14})\,\mathrm{G}$, although the maximum mass $\sim 1.9\,M_\odot$ shown in the track at small radius turns out to be much larger than the Chandrasekhar limit, we find that it is lowered considerably

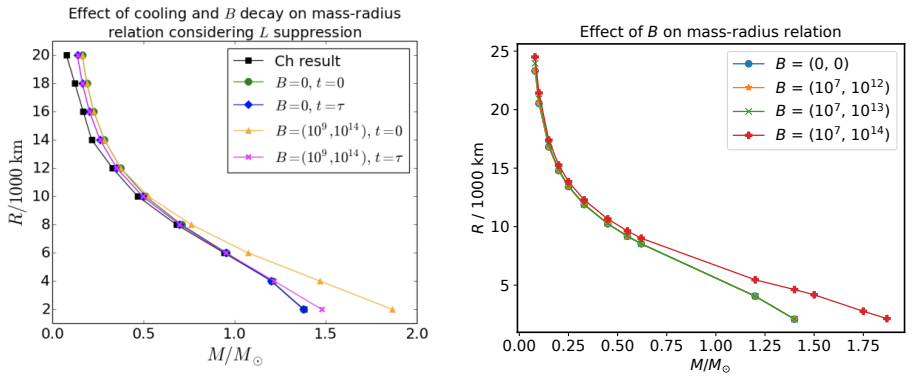


Fig. 6. *Left panel:* The effect of magnetic field on the WD luminosity set to match with the non-magnetised mass–radius relation for the analytical model. The results are shown for $B = (0, 0)$ at initial time $t = 0$ (green circles), $B = (0, 0)$ at $t = 10$ Gyr (blue diamonds), $B = (10^9, 10^{14})$ G at $t = 0$ (orange triangles) and $B = (10^9, 10^{14})$ G at $t = 10$ Gyr (magenta crosses). *Right panel:* STARS results to show the effect of magnetic field on the mass–radius relation of highly magnetized WDs for $B = (0, 0)$ (blue circles), $B = (10^7, 10^{12})$ G (orange stars), $B = (10^7, 10^{13})$ G (green crosses) and $B = (10^7, 10^{14})$ G (red pluses).

to $\sim 1.5\,M_\odot$ primarily as a result of magnetic field decay and also thermal cooling over $t = 10$ Gyr.

Further, we use the STARS stellar evolution code to qualitatively investigate the B-WD mass-radius relationship at different field strengths, with the objective of numerically validating our semi-analytical models. In the right panel of Figure 6, we show the mass-radius relations obtained from STARS. We obtain results that are in good agreement with our analytical formalism and the magnitude of B_0 dictates the shape of the mass-radius curve. In validation of our analytical approach, we have found that the limiting mass $\sim 1.8703\,M_\odot$ obtained with the STARS numerical models is in very good agreement with $M \approx 1.87\,M_\odot$ that is inferred from the semi-analytical calculations for WDs with strong fields $B = (10^{6-9}, 10^{14})$ G. We argue that the young super-Chandrasekhar B-WDs only sustain their large masses up to $\sim 10^5 - 10^6$ yr, and this essentially explains their apparent scarcity even without the difficulty of detection owing to their suppressed luminosities.

6. Continuous gravitational waves from magnetized white dwarfs

The question then remains, how to detect B-WDs directly. Continuous gravitational waves can be among the alternate ways to detect super-Chandrasekhar WD candidates directly. If these B-WDs are rotating with certain angular frequency, then they can efficiently emit gravitational radiation, provided that their magnetic field and rotation axes are not aligned,[48] and these gravitational waves can be detected by upcoming instruments such as LISA, BBO, DECIGO, Einstein Telescope, etc. The dimensionless amplitudes of the two polarizations of the gravitational wave

(GW) at a time t are given by[48,49]

$$h_+ = h_0 \sin\chi \left[\frac{1}{2}\cos i \sin i \cos\chi \cos\Omega t - \frac{1+\cos^2 i}{2}\sin\chi\cos 2\Omega t\right],$$
$$h_\times = h_0 \sin\chi \left[\frac{1}{2}\sin i \cos\chi \sin\Omega t - \cos i \sin\chi \sin 2\Omega t\right], \quad (12)$$

with $h_0 = (-6G/c^4)Q_{z'z'}(\Omega^2/d)$, where $Q_{z'z'}$ is the quadrupole moment of the distorted star, χ is the angle between the rotation axis z' and the body's third principal axis z, i is the angle between the rotation axis of the object and our line of sight. The left panel of Figure 7 shows a schematic diagram of a pulsar with z' being the rotational axis and z the magnetic field axis, where the angle between these two axes is χ. The GW amplitude is

$$h_0 = \frac{4G}{c^4}\frac{\Omega^2 \epsilon I_{xx}}{d}, \quad (13)$$

where $\epsilon = (I_{zz} - I_{xx})/I_{xx}$ is the ellipticity of the body and I_{xx}, I_{yy}, I_{zz} are the principal moments of inertia. Note, we have used the XNS code[50] to simulate the underlying axisymmetric equilibrium configuration of B-WDs in general relativity. Moreover, we assume the distance between the WD and the detector to be 100 pc.

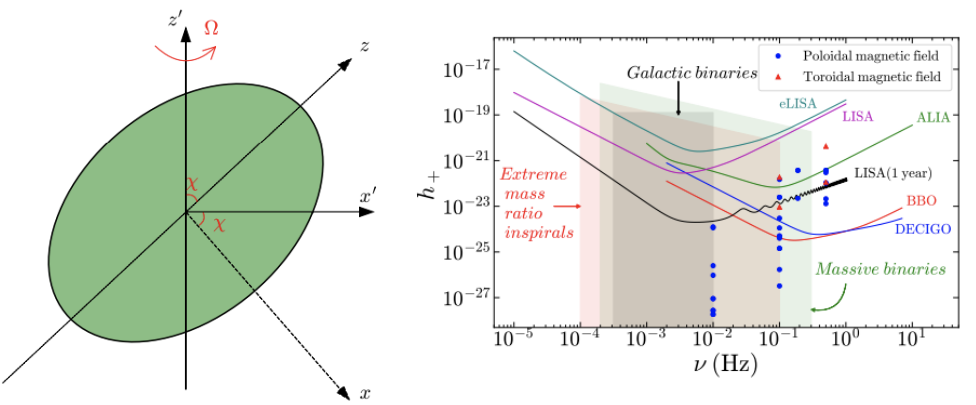


Fig. 7. *Left panel:* Schematic diagram of a B-WD with z' being the rotational axis and z the magnetic axis. *Right panel:* The dimensionless GW amplitudes for WDs are shown as functions of frequency, along with the sensitivity curves of various detectors. Optimum i is chosen for χ at $t = 0$.

Since a pulsating WD can emit both dipole and gravitational radiations simultaneously, it is associated with both dipole and quadrupolar luminosities. The dipole luminosity for an axisymmetric WD is given by[51]

$$L_D = \frac{B_p^2 R_p^6 \Omega^4}{2c^3}\sin^2\chi\, F(x_0), \quad (14)$$

where $x_0 = R_0\Omega/c$, B_p is the magnetic field strength at the pole, R_p is radius of the pole and R_0 is the average WD radius. The function $F(x_0)$ is defined as

$$F(x_0) = \frac{x_0^4}{5(x_0^6 - 3x_0^4 + 36)} + \frac{1}{3(x_0^2 + 1)}. \tag{15}$$

Similarly, the quadrupolar GW luminosity is given by[49]

$$L_{GW} = \frac{2G}{5c^5}(I_{zz} - I_{xx})^2\Omega^6 \sin^2\chi\,(1 + 15\sin^2\chi). \tag{16}$$

It should be noted that this formula is valid if χ is very small. The total luminosity is due to both dipole and gravitational radiations. Therefore, the changes in Ω and χ with time are dependent on both L_D and L_{GW}. The variations of Ω and χ with respect to time are given by[51]

$$\frac{d(\Omega I_{z'z'})}{dt} = -\frac{2G}{5c^5}(I_{zz} - I_{xx})^2\Omega^5 \sin^2\chi\,(1 + 15\sin^2\chi) - \frac{B_p^2 R_p^6 \Omega^3}{2c^3}\sin^2\chi\,F(x_0), \tag{17}$$

$$I_{z'z'}\frac{d\chi}{dt} = -\frac{12G}{5c^5}(I_{zz} - I_{xx})^2\Omega^4 \sin^3\chi\cos\chi - \frac{B_p^2 R_p^6 \Omega^2}{2c^3}\sin\chi\cos\chi\,F(x_0), \tag{18}$$

where $I_{z'z'}$ is the moment of inertia about z'-axis. Equations (17) and (18) need to be solved simultaneously to obtain the timescale over which a WD can radiate.

The right panel of Figure 7 shows the dimensionless GW amplitudes for the WDs as functions of their frequencies, along with the sensitivity curves of various detectors. It can be seen that the isolated WDs may not be detected directly by LISA, but can be detected after integrating the signal to noise ratio S/N for 1 year. As WDs are larger in size compared to NS, they cannot rotate as fast as NS and hence ground-based GW detectors such as LIGO, Virgo and KAGRA are not expected to detect the isolated WDs. These isolated WDs are also free from the noise due to the galactic binaries as well as from the extreme mass ratio inspirals (EMRIs).

7. Summary and conclusions

We have shown that highly magnetized, stable WDs, namely B-WDs, have a variety of implications, including enigmatic peculiar over-luminous SNe Ia. Numerical simulations utilizing the stellar evolution code STARS indicate that the central field in strongly magnetized B-WDs can be toroidally dominated whereas the surface fields are more of dipole nature. The mass $\sim 1.87\,M_\odot$ obtained for these B-WDs from the STARS numerical models, as of now, is totally consistent with the estimates from the analytical stellar structure models for strong fields $B = (10^{6-9}, 10^{14})$ G. These young super-Chandrasekhar B-WDs have suppressed luminosities and are difficult to observe or rare, due to their decaying magnetic fields. We have found that these WDs may not remain super-Chandrasekhar for long i.e. beyond $\sim 10^{5-6}$ year due to

decaying field primarily, and indicate rapidly losing pulsar nature and/or low luminosity. However, these stars can be very promising candidates for GW detectors such as LISA (with 1 year integrated S/N) and also for Einstein Telescope and future DECIGO/BBO missions. Therefore, appropriate missions in GW astronomy and otherwise, e.g. radio astronomy, should be planned in order to probe them in the future.

Acknowledgments

B.M. acknowledges a partial support by a project of Department of Science and Technology (DST-SERB) with research Grant No. DSTO/PPH/BMP/1946 (EMR/2017/001226). M.B. acknowledges support from Graduate Continuing Fellowship at the University of Texas, Austin. A.J.H. thanks the Science and Technology Facilities Council (STFC) and the Cambridge Commonwealth, European & International Trust for his doctoral funding. C.A.T. thanks Churchill College for his fellowship.

References

1. D. A. Howell, et al., Nature, 443, 308 (2006).
2. R. A. Scalzo, et al., ApJ, 713, 1073 (2010).
3. I. Hachisu, Ap&SS, 61, 479 (1986).
4. S.-C. Yoon, N. Langer, A&A, 419, 623 (2004).
5. B. Mukhopadhyay, A. R. Rao, JCAP, 05, 007 (2016).
6. U. Das, B. Mukhopadhyay, Phys. Rev. D, 86, 042001 (2012).
7. U. Das, B. Mukhopadhyay, Phys. Rev. Lett., 110, 071102 (2013).
8. S. Subramanian, B. Mukhopadhyay, MNRAS, 454, 752 (2015).
9. K. M. Vanlandingham, et al., AJ, 130, 734 (2005).
10. L. Ferrario, D. Martino, B. Gaensicke, Space Sci. Rev., 191, 111 (2015).
11. U. Das, B. Mukhopadhyay, JCAP, 06, 050 (2014).
12. U. Das, B. Mukhopadhyay, JCAP, 05, 016 (2015).
13. B. Mukhopadhyay, A. R. Rao, JCAP, 05, 007 (2016).
14. B. Mukhopadhyay, A. R. Rao, T. S. Bhatia, MNRAS, 472, 3564 (2017).
15. B. Mukhopadhyay, U. Das, A. R. Rao, S. Subramanian, M. Bhattacharya, S. Mukerjee, T. S. Bhatia, J. Sutradhar, in P.-E. Tremblay, B. Gansicke, T. Marsh, eds, Astronomical Society of the Pacific Conference Series, Vol. 509, 20th European White Dwarf Workshop. Astron. Soc. Pac., San Francisco, p. 401 (2017).
16. M. Bhattacharya, B. Mukhopadhyay, S. Mukerjee, MNRAS, 477, 2705 (2018).
17. M. Bhattacharya, B. Mukhopadhyay, S. Mukerjee, Proceedings of the 21st European Workshop on White Dwarfs, arXiv:1810.07836 (2018).
18. A. Gupta, B. Mukhopadhyay, C. A. Tout, MNRAS, 496, 894 (2020).
19. M. Bhattacharya, A. J. Hackett, A. Gupta, C. A. Tout, B. Mukhopadhyay, arXiv:2106.09736 (2021).
20. S. Kalita, B. Mukhopadhyay, MNRAS, 490, 2692 (2019).
21. S. Kalita, B. Mukhopadhyay, T. Mondal, T. Bulik, ApJ, 896, 69 (2020).
22. S. Banerjee, S. Shankar, T. P. Singh, J. Cosmol. Astropart. Phys., 2017, 004 (2017).
23. B. E. Eslam Panah, H. Liu, Phys. Rev. D, 99, 104074 (2019).

24. O. Bertolami, H. Mariji, Phys. Rev. D, 93, 104046 (2016).
25. H. Liu, X. Zhang, D. Wen, Phys. Rev. D, 89, 104043 (2014).
26. G. A. Carvalho, J. D. V. Arbanil, R. M. Marinho, M. Malheiro, Eur. Phys. J. C, 78, 411 (2018).
27. V. B. Belyaev, P. Ricci, F. Simkovic, J. Adam, M. Tater, E. Truhlik, Nucl. Phys. A, 937, 17 (2015).
28. L. Herrera, W. Barreto, Phys. Rev. D, 87, 087303 (2013).
29. P. Markey, R. J. Tayler, MNRAS, 163, 77 (1973).
30. R. J. Tayler, MNRAS, 161, 365 (1973).
31. D. T. Wickramasinghe, C. A. Tout, L. Ferrario, MNRAS, 437, 675 (2013).
32. L. G. Quentin, C. A.Tout, MNRAS, 477, 2298 (2018).
33. U. Das, B. Mukhopadhyay, A. R. Rao, ApJ, 767, L14 (2013).
34. K. Ioka, M. Sasaki, ApJ, 600, 296 (2004).
35. J. Frieben, L. Rezzolla, MNRAS, 427, 3406 (2012).
36. S. Chandrasekhar, E. Fermi, ApJ, 118, 116 (1953).
37. H. Komatsu, Y. Eriguchi, I. Hachisu, MNRAS, 237, 355 (1989).
38. J. Braithwaite, MNRAS, 397, 763 (2009).
39. N. Stergioulas, Living Reviews in Relativity, 6, 3 (2003).
40. N. Bucciantini, L. Del Zanna, A&A, 528, A101 (2011).
41. M. Sinha, B. Mukhopadhyay, A. Sedrakian, Nucl. Phys. A, 898, 43 (2013).
42. D. Bandyopadhyay, S. Chakrabarty, S. Pal, Phys. Rev. Lett, 79, 2176 (1997).
43. A. Y. Potekhin, D. G. Yakovlev, A&A, 374, 213 (2001).
44. J. Ventura, A. Y. Potekhin, The Neutron Star - Black Hole Connection, Kluwer, Dordrecht (2001).
45. S. Chandrasekhar, MNRAS, 95, 207 (1935).
46. J. S. Heyl, S. R. Kulkarni, ApJ, 506, 61 (1998).
47. A. Cumming, MNRAS, 333, 589 (2002).
48. S. Bonazzola, E. Gourgoulhon, A&A, 312, 675 (1996).
49. M. Zimmermann, E. Szedenits Jr., Phys. Rev. D., 20, 351 (1979).
50. A. G. Pili, N. Bucciantini, L. Del Zanna, MNRAS, 439, 3541 (2014).
51. A. Melatos, MNRAS, 313, 217 (2000).

Electron captures and stability of white dwarfs

N. Chamel* and L. Perot

Institut d'Astronomie et d'Astrophysique, Université Libre de Bruxelles, CP-226, 1050 Brussels, Belgium
**E-mail: nicolas.chamel@ulb.be*

A. F. Fantina

Grand Accélérateur National d'Ions Lourds (GANIL), CEA/DRF - CNRS/IN2P3, Boulevard Henri Becquerel, 14076 Caen, France

Institut d'Astronomie et d'Astrophysique, CP-226, Boulevard du Triomphe, Université Libre de Bruxelles, 1050 Brussels, Belgium

D. Chatterjee and S. Ghosh

Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune University Campus, Pune, 411007, India

J. Novak and M. Oertel

LUTH, Observatoire de Paris, PSL Research University, CNRS, Université Paris Diderot, Sorbonne Paris Cité, 5 place Jules Janssen, 92195 Meudon, France

Electron captures by atomic nuclei in dense matter are among the most important processes governing the late evolution of stars, limiting in particular the stability of white dwarfs. Despite considerable progress in the determination of the equation of state of dense Coulomb plasmas, the threshold electron Fermi energies are still generally estimated from the corresponding Q values in vacuum. Moreover, most studies have focused on nonmagnetized matter. However, some white dwarfs are endowed with magnetic fields reaching 10^9 G. Even more extreme magnetic fields might exist in super Chandrasekhar white dwarfs, the progenitors of overluminous type Ia supernovae like SN 2006gz and SN 2009dc. The roles of the dense stellar medium and magnetic fields on the onset of electron captures and on the structure of white dwarfs are briefly reviewed. New analytical formulas are derived to evaluate the threshold density for the onset of electron captures for arbitrary magnetic fields. Their influence on the structure of white dwarfs is illustrated by simple analytical formulas and numerical calculations.

Keywords: White dwarf; Chandrasekhar limit; Electron capture; Magnetic field.

1. Introduction

In 1926, the British physicist Ralph Fowler showed that the energy and the pressure of matter in the dense core of a white dwarf remain finite at zero temperature due to quantum mechanics. Considering that electrons are no longer bound to nuclei under such extreme conditions, he derived the equation of state of a degenerate electron Fermi gas.[1] Another British physicist, Edmund Clifton Stoner, calculated the

structure of white dwarfs with uniform density in 1929.[2] The same year, Wilhelm Anderson, a physicist at the University of Tartu in Estonia, first stressed the importance of taking into account the relativistic motion of electrons.[3] He showed that the mass of a white dwarf tends to some finite value as the electron concentration increases, namely $M_{\rm And} \sim 0.69 M_\odot$ assuming an electron fraction $y_e = 0.4$ (with $M_\odot$ the mass of the Sun). Improving Anderson's approximate treatment, Stoner[4] found $M_{\rm St} \simeq 1.10 M_\odot$. Soon afterwards, the young Indian physicist Subrahmanyan Chandrasekhar solved the hydrostatic equilibrium equations for an ultrarelativistic electron Fermi gas using the theory of polytropes.[5] His numerical result, $M_{\rm Ch} \simeq 0.91 M_\odot$ (for $y_e = 0.4$), thus differed by less than 20% from that obtained earlier by Stoner, a remarkably close agreement as pointed out by Chandrasekhar himself. A year later, the Russian physicist Lev Landau showed that any degenerate star has a maximum mass, which can be expressed in terms of the fundamental constants as

$$M_{\rm L} = 3.1 \frac{m_{\rm P}^3}{m^2} y_e^2 \,, \tag{1}$$

where we have introduced the Planck mass

$$m_{\rm P} = \sqrt{\frac{\hbar c}{G}} \,, \tag{2}$$

($\hbar$ is the Planck-Dirac constant, c is the speed of light, G is the constant of gravitation) and m is the average mass per nucleon. Incidentally, such a scaling was already apparent in Stoner's analysis of a uniform density star. Combining his equations (17) and (19b) leads to

$$M_{\rm St} = \frac{15\sqrt{5\pi}}{16} \frac{m_{\rm P}^3}{m^2} y_e^2 \,. \tag{3}$$

Landau however did not believe in the physical reality of this limit and even invoked some violation of the laws of quantum mechanics inside massive stars to prevent them from collapsing (see Ref. 6 for a historical perspective on Landau's contribution). As early as 1928, the process by which atoms are crushed at high densities was actually discussed by the Russian physicist Yakov Frenkel.[7] More importantly, he calculated the equation of state of an electron Fermi gas for an arbitrary degree of relativistic motion as well as the correction due to electrostatic interactions between electrons and atomic nuclei. He also studied the conditions for which incompressible "superdense stars" can exist and derived indirectly the mass limit. Correcting for an error in his expression for the gravitational energy[a] and neglecting the electrostatic correction lead to the same result as the one published by Stoner in 1930. Frenkel's pioneer work remained unnoticed during several decades, and is still not very well-known today.[8]

[a] The factor 5/3 in his key equation (19a) should read 3/5; this error goes back to his incorrect expression (18) for the gravitational energy.

The analyses of Anderson and Stoner showed that the maximum mass of white dwarfs is only reached asymptotically when the electron concentration tends to infinity and the stellar radius goes to zero. This conclusion was later confirmed by the numerical calculations of Chandrasekhar for more realistic density profiles.[9] However, soon after the discovery of the neutron by James Chadwick in 1932, it was realized that matter becomes predominantly composed of neutrons at high densities.[10,11] In December 1933, during a meeting of the American Physical Society at Stanford, Wilhelm Baade and Fritz Zwicky predicted the existence of *neutron stars* formed from the catastrophic gravitational collapse of stars during supernova explosions.[12] Baade and Zwicky were apparently unaware of the studies about white dwarfs. The connection was first made by Landau[13] and Gamow.[14] At a conference in Paris in 1939, Chandrasekhar also pointed out[15]:

> If the degenerate core attains sufficiently high densities, the protons and electrons will combine to form neutrons. This would cause a sudden diminution of pressure resulting in the collapse of the star to a neutron core.

In 1956, the French physicist Evry Schatzman[16] showed that the central density of white dwarfs is limited by the onset of electron captures by nuclei, implying that the radius of the most massive white dwarfs remains *finite*. Detailed calculations of the structure of white dwarfs taking into account matter neutronization were performed at the end of the 1950s and at the beginning of the 1960s.[17–19] Electron captures are now known to play a key role in the late stages of stellar evolution (see Ref. 20 for a recent review).

Although the electrostatic correction to the equation of state of an electron Fermi gas was calculated long ago[7] (see Ref. 21 and references therein for the latest developments on the equation of state of dense Coulomb plasmas), the threshold electron Fermi energy μ_e for the onset of electron captures by nuclei is still generally estimated from the corresponding Q value in *vacuum*. Moreover, the presence of magnetic fields is usually ignored. However, a significant fraction of white dwarfs have been found[22] to have magnetic fields up to 10^9 G, and potentially much stronger fields may exist in their core.[23] Moreover, it has been recently proposed that very massive so-called super-Chandrasekhar white dwarfs (with a mass $M \gtrsim 2M_\odot$) endowed with extremely strong magnetic fields could be the progenitors of overluminous type Ia supernovae like SN 2006gz and SN 2009dc[24–26] (see also Refs. 27–30). The existence of such stars was actually first studied much earlier by Shul'man,[31] who found that the maximum mass of degenerate stars could be increased by two orders of magnitude if the magnetic field is strongly quantizing.[32,33] However, the stability of such super Chandrasekhar white dwarfs will still be limited by electron captures.[34,35] Detailed calculations of the global structure of these stars taking these processes into account have been carried out in Refs. 36, 37.

In this paper, we review our recent studies of the role of electron-ion interactions and magnetic fields in the onset of electron captures in cold white-dwarf cores.[36,38]

We also present more accurate and more general formulas for the threshold density and pressure, applicable not only to nonmagnetic and strongly magnetized white dwarfs but also to stars with intermediate magnetic field strengths. The impact of electron captures on the global structure of white dwarfs is discussed and new numerical results are presented.

2. Core of white dwarfs with magnetic fields

The core of a white dwarf consists of a dense Coulomb plasma of nuclei in a charge compensating background of relativistic electrons. Apart from carbon and oxygen (the primary ashes of helium burning), the core may contain other nuclei like helium,[39–41] neon and magnesium,[42] or even iron.[43,44] Iron white dwarfs could be formed from the explosive ignition of electron degenerate oxygen-neon-magnesium cores,[45] or from failed detonation supernovae.[46] For simplicity, we assume that the stellar core is made of only one type of nuclei with charge number Z and mass number A (see, e.g. Ref. 38 for the treatment of mixtures). We further suppose that the star has sufficiently cooled down such that thermal effects can be neglected.

Whereas nuclei with number density n_N exert a negligible pressure $P_N = 0$, they contribute to the mass density

$$\rho = n_N M'(A, Z)\,, \tag{4}$$

where $M'(A, Z)$ denotes the ion mass including the rest mass of Z electrons and can be obtained from the experimental *atomic* mass $M(A, Z)$ by subtracting out the binding energy of the atomic electrons (see Eq. (A4) of Ref. 47). In principle, the presence of a magnetic field can have some effect on nuclei.[48,49] However, the change of nuclear masses is negligible even for the strongest magnetic fields of order 10^{15} G expected in super Chandrasekhar white dwarfs.[29,36,50]

To a very good approximation, electrons can be treated as an ideal Fermi gas. In the presence of a magnetic field, the electron motion perpendicular to the field is quantized into Landau-Rabi levels.[51,52] Quantization effects on the equation of state are significant when B exceeds $B_{\rm rel}$, where

$$B_{\rm rel} \equiv \frac{m_e^2 c^3}{e\hbar} \approx 4.4 \times 10^{13}\ {\rm G}\,, \tag{5}$$

where m_e is the electron mass and e is the elementary electric charge. The expressions for the energy density $\mathcal{E}_e$ and pressure P_e for arbitrary magnetic field strength can be found in Ref. 53.

The main correction to the ideal electron Fermi gas arises from the electron-ion interactions, as first shown by Frenkel[7] (see e.g. Ref. 53 for a discussion of higher-order corrections). For pointlike ions embedded in a uniform electron gas with number density $n_e = Z n_N$ (from electric charge neutrality), the corresponding energy density is given by (see e.g. Chap. 2 of Ref. 53)

$$\mathcal{E}_L = C_M \left(\frac{4\pi}{3}\right)^{1/3} e^2 n_e^{4/3} Z^{2/3}\,, \tag{6}$$

where C_M is the Madelung constant. The contribution to the pressure is thus given by

$$P_L = n_e^2 \frac{d(\mathcal{E}_L/n_e)}{dn_e} = \frac{\mathcal{E}_L}{3}\,. \tag{7}$$

The pressure of the Coulomb plasma finally reads $P = P_e + P_L$.

In the following, we shall consider that ions are arranged in a body-centered cubic lattice since this configuration leads to the lowest energy.[53] In this case, the Madelung constant is given by[54] $C_M = -0.895929255682$. According to the Bohr-van Leeuwen theorem,[55] the electrostatic corrections (6) and (7) are independent of the magnetic field apart from a negligibly small contribution due to quantum zero-point motion of ions about their equilibrium position.[56]

3. Onset of electron captures by nuclei in dense environments

The onset of electron captures by nuclei (A, Z) is formally determined by the same condition irrespective of the magnetic field strength by requiring the constancy of the Gibbs free energy per nucleon at fixed temperature and pressure.[38] The threshold electron Fermi energy is found to first order in the fine-structure constant $\alpha = e^2/(\hbar c)$ from the condition:

$$\gamma_e + C_M \left(\frac{4\pi n_e}{3}\right)^{1/3} \alpha \lambda_e F(Z) = \gamma_e^\beta(A, Z)\,, \tag{8}$$

$$F(Z) \equiv Z^{5/3} - (Z-1)^{5/3} + \frac{1}{3} Z^{2/3}\,, \tag{9}$$

$$\gamma_e^\beta(A, Z) \equiv -\frac{Q_{\rm EC}(A, Z)}{m_e c^2} + 1\,, \tag{10}$$

where $\gamma_e \equiv \mu_e/(m_e c^2)$, $\lambda_e = \hbar/(m_e c)$ is the electron Compton wavelength, and we have introduced the Q-value (in vacuum) associated with electron capture by nuclei (A, Z):

$$Q_{\rm EC}(A, Z) = M'(A, Z)c^2 - M'(A, Z-1)c^2\,. \tag{11}$$

These Q-values can be obtained from the tabulated Q-values of β decay by the following relation:

$$Q_{\rm EC}(A, Z) = -Q_\beta(A, Z-1)\,. \tag{12}$$

In principle, the daughter nucleus may be in an excited state. However, such a transition would occur at a higher density.

In the absence of magnetic fields, the threshold condition (8) can be solved analytically.[57] Recalling that the electron Fermi energy is given by

$$\mu_e = m_e c^2 \sqrt{1 + x_r^2}\,, \tag{13}$$

where $x_r = \lambda_e k_e$ and $k_e = (3\pi^2 n_e)^{1/3}$ is the electron Fermi wave number, the solution reads

$$x_r^\beta = \gamma_e^\beta \left\{ \sqrt{1 - \left[1 - \tilde{F}(Z)^2\right]/(\gamma_e^\beta)^2} - \tilde{F}(Z) \right\} \left[1 - \tilde{F}(Z)^2\right]^{-1}, \tag{14}$$

with

$$\tilde{F}(Z) \equiv C_M \left(\frac{4}{9\pi}\right)^{1/3} \alpha F(Z)\,. \tag{15}$$

The pressure $P_\beta(A, Z)$ at the onset of electron captures is given by

$$P_\beta(A,Z) = \frac{m_e c^2}{8\pi^2 \lambda_e^3} \left[x_r^\beta \left(\frac{2}{3}(x_r^\beta)^2 - 1 \right) \sqrt{1 + (x_r^\beta)^2} + \ln(x_r^\beta + \sqrt{1 + (x_r^\beta)^2}) \right] + \frac{C_M \alpha}{3} \left(\frac{4}{243\pi^7}\right)^{1/3} (x_r^\beta)^4 \frac{m_e c^2}{\lambda_e^3} Z^{2/3}\,. \tag{16}$$

The corresponding average mass density is found from Eq. (4) and is given by

$$\rho_\beta(A,Z) = \frac{M'(A,Z)}{Z} \frac{(x_r^\beta)^3}{3\pi^2 \lambda_e^3}\,. \tag{17}$$

Numerical results are summarized in Table 1 for nuclei expected to be found in the core of white dwarfs, using data from the 2020 Atomic Mass Evaluation.[58] Fundamental constants were taken from NIST CODATA 2018[b]. We have not considered helium since this element is expected to undergo pycnonuclear fusion in white-dwarf cores before capturing electrons.

Table 1. Dimensionless threshold electron Fermi energy $\gamma_e = \mu_e/(m_e c^2)$, mass density ρ_β and pressure P_β for the onset of electron captures by the given nuclei in unmagnetized white dwarfs.

	^{16}O	^{12}C	^{20}Ne	^{24}Mg	^{56}Fe
γ_e	22.0	27.8	15.2	12.2	8.74
ρ_β [g cm^{-3}]	2.06×10^{10}	4.16×10^{10}	6.81×10^{9}	3.51×10^{9}	1.37×10^{9}
P_β [dyn cm^{-2}]	2.73×10^{28}	6.99×10^{28}	6.21×10^{27}	2.56×10^{27}	6.46×10^{26}

In the presence of a magnetic field, the threshold condition (8) must be solved using the following relation between n_e and γ_e:

$$n_e = \frac{2B_\star}{(2\pi)^2 \lambda_e^3} \sum_{\nu=0}^{\nu_{\max}} g_\nu x_e(\nu)\,, \tag{18}$$

$$x_e(\nu) = \sqrt{\gamma_e^2 - 1 - 2\nu B_\star}\,, \tag{19}$$

[b] https://physics.nist.gov/cuu/Constants/index.html

where we have introduced $B_\star \equiv B/B_{\rm rel}$, and the degeneracy $g_\nu = 1$ for $\nu = 0$ and $g_\nu = 2$ for $\nu \geq 1$. The index $\nu_{\rm max}$ is the highest integer for which $\gamma_e^2 - 1 - 2\nu_{\rm max} B_\star \geq 0$, i.e.

$$\nu_{\rm max} = \left[\frac{\gamma_e^2 - 1}{2B_\star} \right], \tag{20}$$

where [.] denotes the integer part.

In the weakly quantizing regime meaning that many Landau-Rabi levels are populated ($\nu_{\rm max} \gg 1$), analytical solutions can be found. Remarking that the magnetic field enters explicitly in (8) only through the small electrostatic correction, the threshold electron Fermi energy is still approximately given by the solution in the absence of magnetic fields, namely $\gamma_e \approx \sqrt{1 + (x_r^\beta)^2}$ with x_r^β given by (14). Substituting in Eqs. (18) and (19), using the expansions (41) obtained in Ref. 59 leads to the following estimate for the density marking the onset of electron captures:

$$\rho_\beta(A,Z) \approx \frac{M'(A,Z)}{2\pi^2 Z \lambda_e^3} \left[\frac{2}{3} \left(\gamma_e^2 - 1 \right)^{3/2} + (2B_\star)^{3/2} \zeta \left(\frac{-1}{2}, \left\{ \frac{\gamma_e^2 - 1}{2B_\star} \right\} \right) + \frac{B_\star^2}{6\sqrt{\gamma_e^2 - 1}} \right], \tag{21}$$

where $\zeta(z,q)$ is the Hurwitz zeta function defined by

$$\zeta(z,q) = \sum_{\nu=0}^{+\infty} \frac{1}{(\nu+q)^z} \tag{22}$$

for $\Re(z) > 1$ and by analytic continuation to other $z \neq 1$ (excluding poles $\nu + q = 0$), and $\{.\}$ in the argument denotes the fractional part. The first term in Eq. (21) represents the threshold density in the absence of magnetic field. The second term accounts for magnetic oscillations while the last term is a higher-order correction. The expression for the associated pressure is more involved. Using Eqs. (41), (43) and (44) of Ref. 59 yields[c]

$$\begin{aligned} P_\beta(A,Z) \approx \frac{m_e c^2}{4\pi^2 \lambda_e^3} \Bigg\{ & \frac{1}{2} \left(1 - 2B_\star + \frac{2B_\star^2}{3} \right) \log \left(\frac{\gamma_e + \sqrt{2B_\star + \gamma_e^2 - 1}}{1 + \sqrt{2B_\star}} \right) \\ & - \frac{1}{2} \left(\gamma_e \sqrt{2B_\star + \gamma_e^2 - 1} - \sqrt{2B_\star} \right) + \frac{1}{3} \left(\gamma_e \sqrt{2B_\star + \gamma_e^2 - 1}^3 - \sqrt{2B_\star}^3 \right) \\ & + B_\star \left(\operatorname{arccosh} \gamma_e - \gamma_e \sqrt{\gamma_e^2 - 1} \right) - (2B_\star)^{5/2} \int_0^{+\infty} \frac{\tilde{\zeta}_3(-1/2, q+1)}{\sqrt{1 + 2B_\star q}} dq \end{aligned}$$

[c]In the notations of Ref. 59, the electron contribution to the pressure can be directly obtained from the grand potential density by $P_e = -\omega_0^{(\rm mon)} - \omega_0^{(\rm osc)}$. The total pressure is found by adding the electrostatic correction (7) using the expansion for the electron density.

$$+\frac{2}{3}\frac{(2B_\star)^{5/2}}{\gamma_e}\zeta\left(\frac{-3}{2},\left\{\frac{\gamma_e^2-1}{2B_\star}\right\}\right)+\frac{2}{15}\frac{(2B_\star)^{7/2}}{\gamma_e^3}\zeta\left(\frac{-5}{2},\left\{\frac{\gamma_e^2-1}{2B_\star}\right\}\right)$$
$$+\frac{1}{240}\left(\frac{B_\star}{\gamma_e}\right)^4+4B_\star^2\int_0^1\zeta\left(\frac{-1}{2},q\right)\zeta\left(\frac{1}{2},q+\frac{1}{2B_\star}\right)dq$$
$$+\frac{2}{3}\left(\frac{2}{3\pi}\right)^{1/3}C_M\alpha Z^{2/3}\left[\frac{2}{3}\left(\gamma_e^2-1\right)^{3/2}\right.$$
$$\left.+(2B_\star)^{3/2}\zeta\left(\frac{-1}{2},\left\{\frac{\gamma_e^2-1}{2B_\star}\right\}\right)+\frac{B_\star^2}{6\sqrt{\gamma_e^2-1}}\right]^{4/3}\Bigg\}, \tag{23}$$

with

$$\tilde{\zeta}_3(z,q)=\zeta(z,q)-\frac{1}{z-1}q^{-z+1}-\frac{1}{2}q^{-z}-\frac{z}{12}q^{-z-1}\,. \tag{24}$$

An analytical solution also exists in the strongly quantizing regime whereby electrons are all confined to the lowest Landau-Rabi level ($\nu_{\max}=0$). Introducing

$$\bar{F}(Z,B_\star)\equiv\frac{1}{3}C_M\alpha F(Z)\left(\frac{2B_\star}{3\pi}\right)^{1/3}<0\,, \tag{25}$$

$$\upsilon\equiv\frac{\gamma_e^\beta}{2|\bar{F}(Z,B_\star)|^{3/2}}\,, \tag{26}$$

the solutions are given by the following formulas[60]:

$$\gamma_e=\begin{cases}8|\bar{F}(Z,B_\star)|^{3/2}\cosh^3\left(\frac{1}{3}\operatorname{arccosh}\upsilon\right) & \text{if } \upsilon\geq 1\,,\\ 8|\bar{F}(Z,B_\star)|^{3/2}\cos^3\left(\frac{1}{3}\arccos\upsilon\right) & \text{if } 0\leq\upsilon<1\,.\end{cases} \tag{27}$$

The threshold pressure and density are respectively given by:

$$P_\beta(A,Z,B_\star)=\frac{B_\star m_e c^2}{4\pi^2\lambda_e^3}\left[\gamma_e\sqrt{\gamma_e^2-1}-\ln\left(\sqrt{\gamma_e^2-1}+\gamma_e\right)\right.$$
$$\left.+\frac{C_M\alpha}{3}\left(\frac{16B_\star Z^2}{3\pi}\right)^{1/3}\left(\gamma_e^2-1\right)^{2/3}\right], \tag{28}$$

$$\rho_\beta(A,Z,B_\star)=\frac{B_\star}{2\pi^2\lambda_e^3}\frac{M'(A,Z)}{Z}\sqrt{\gamma_e^2-1}\,. \tag{29}$$

Let us recall that Eq. (27) is only valid if all electrons lie in the lowest Landau-Rabi level, i.e. if their Fermi energy does not exceed $\gamma_e=\sqrt{1+2B_\star}\approx\sqrt{2B_\star}$ or equivalently if their density does not exceed $n_e=B_\star^{3/2}/(\sqrt{2}\pi^2\lambda_e^3)$ using Eqs. (18) and (19).

This condition translates into a lower bound for the magnetic field $B_\star \geq B_{\star 1}$. To find $B_{\star 1}$, we substitute the above expressions for γ_e and n_e in Eq. (8). This leads to

$$B_{\star 1} = \frac{(\gamma_e^\beta)^2}{2}\left[1 + \frac{C_M \alpha}{(3\pi)^{1/3}} F(Z)\right]^{-2}. \tag{30}$$

Values for some nuclei are summarized in Table 2. For magnetic field strength $B_\star \lesssim B_{\star 1}$, the threshold condition (8) must be solved numerically.

Table 2. Magnetic field strength in units of $B_{\rm rel}$ above which electrons are confined to the lowest Rabi level when captured by the given nuclei.

^{16}O	^{12}C	^{20}Ne	^{24}Mg	^{56}Fe
240	384	115	74.1	37.8

Full results for different nuclei expected to be found in white-dwarf cores are shown in Figs. 1 and 2. Changes in the occupation of Rabi levels as the magnetic field is increased lead to typical oscillations of the threshold density ρ_β: the onset of electron captures can thus be shifted to either lower or higher density as compared to unmagnetized matter. The lowest value is reached for $B_\star = B_{\star 1}$ and lies about 25% below the threshold density in the absence of magnetic fields irrespective of the composition (this can be easily shown by taking the ratio of Eqs. (29) and (17) using (30) and ignoring all terms in α). In the strongly quantizing regime, ρ_β is unbound from above and increases almost linearly with $B_\star$.

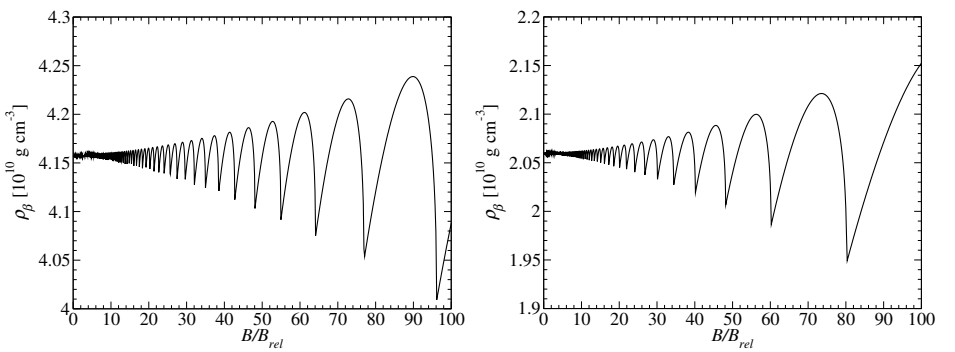


Fig. 1. Threshold density for the onset of electron captures by nuclei as a function of the magnetic field strength in units of the characteristic field (5). (a) Results for ^{12}C. (b) Results for ^{16}O.

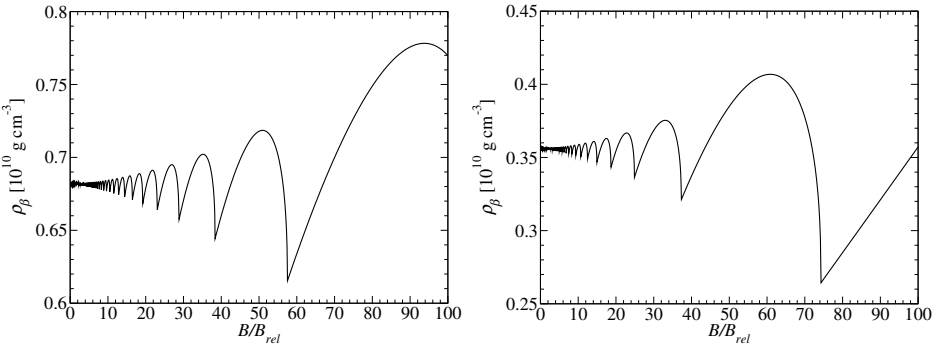


Fig. 2. Same as Fig. 1 for (a) ^{20}Ne and (b) ^{24}Mg.

4. Global structure of white dwarfs

The core of a white dwarf is expected to be surrounded by a helium mantle and an hydrogen envelope. Their contribution to the mass of the star cannot exceed $\sim 1\%$ to avoid a thermonuclear runaway. For simplicity, we will ignore these layers here.

In the model originally considered by Chandrasekhar,[5] the interior of a white dwarf was described using the equation of state of an ideal electron Fermi gas. The composition, assumed to be uniform, was only included throughout the charge neutrality condition. In the ultrarelativistic limit, the resulting equation of state reduces to the polytropic form $P \approx K_0 \rho^{4/3}$ with

$$K_0 = \frac{\hbar c (3\pi^2)^{1/3}}{4} \left(\frac{Z}{M'(A,Z)} \right)^{4/3} . \tag{31}$$

The mass and radius of a white dwarf with central density ρ_c can be deduced from the theory of polytropes[61]

$$M_{\rm Ch} = 4\pi \left(\frac{K_0}{\pi G} \right)^{3/2} \xi_1^2 |\theta'(\xi_1)| \, , \tag{32}$$

$$R_{\rm Ch} = \sqrt{\frac{K_0}{\pi G}} \frac{\xi_1}{\rho_c^{1/3}} \tag{33}$$

respectively, where $\xi_1 \simeq 6.89685$ and $\xi_1^2 |\theta'(\xi_1)| \simeq 2.01824$. Substituting the expression for K_0 thus yields

$$M_{\rm Ch} = \frac{\sqrt{3\pi}}{2} \xi_1^2 |\theta'(\xi_1)| \frac{m_{\rm P}^3}{m^2} y_e^2 \simeq 3.09798 \frac{m_{\rm P}^3}{m^2} y_e^2 \, , \tag{34}$$

$$R_{\rm Ch} = \frac{\sqrt{3\pi}}{2} \xi_1 \lambda_e \frac{m_{\rm P}}{m} \frac{y_e}{x_{r,c}} \simeq 10.5866 \lambda_e \frac{m_{\rm P}}{m} \frac{y_e}{x_{r,c}} \, , \tag{35}$$

where

$$m \equiv \frac{M'(A,Z)}{A}\,, \tag{36}$$

and $x_{r,c}$ denotes the relativity parameter at the density ρ_c. Equation (34) coincides with the maximum mass because the ultrarelativistic regime is only valid at high densities such that the relativity parameter in the central core of the star satisfies $x_{r,c} \gg 1$. In the limit $x_{r,c} \to +\infty$, the radius $R_{\rm Ch}$ vanishes.

The electrostatic corrections (6) and (7) can be easily taken into account by simply renormalizing the constant K_0 as

$$K = K_0\left[1 + \alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3} Z^{2/3}\right]. \tag{37}$$

The mass and radius thus become[d]

$$M = \left[1 + \alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3} Z^{2/3}\right]^{3/2} M_{\rm Ch} < M_{\rm Ch}\,, \tag{38}$$

$$R = R_{\rm Ch}\left[1 + \alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3} Z^{2/3}\right]^{1/2} < R_{\rm Ch}\,. \tag{39}$$

This shows how the electron-ion interactions lower both the mass and the radius. It can be seen that the radius of the most massive white dwarfs still vanishes in the limit $x_{r,c} \to +\infty$. However, as discussed earlier, nuclei in the white dwarf core will capture electrons as soon as the central density ρ_c exceeds the threshold density ρ_β. The daughter nuclei are generally unstable against a second electron capture. The transition is accompanied by a discontinuous increase of density given by

$$\frac{\Delta\rho}{\rho_\beta} = \frac{Z}{Z-2}\frac{M'(A,Z-2)}{M'(A,Z)}\left[1 + \alpha C_M\left(\frac{4}{9\pi}\right)^{1/3}\left(Z^{2/3} - (Z-2)^{2/3}\right)\frac{\sqrt{1+x_r^2}}{x_r}\right] - 1\,, \tag{40}$$

where x_r is given by Eq. (14). The density jump is about 50% for carbon, 33% for oxygen, 25% for neon, 20% for magnesium, and 8.2% for iron.

Because these reactions occur at the same pressure, the adiabatic index defined by

$$\Gamma = \frac{d\log P}{d\log\rho} \tag{41}$$

therefore vanishes thus making the star unstable. In the limit of ultrarelativistic electrons, the average threshold density and pressure reduce to

$$\rho_\beta(A,Z) \approx \frac{M'(A,Z)}{Z}\frac{\mu_e^\beta(A,Z)^3}{3\pi^2(\hbar c)^3}\left[1 + \alpha C_M\left(\frac{4}{9\pi}\right)^{1/3} F(Z)\right]^{-3}, \tag{42}$$

[d]Such kind of scaling was briefly mentioned in Ref. 19 through an effective mean molecular weight per electron.

$$P_\beta(A,Z) \approx \frac{\mu_e^\beta(A,Z)^4}{12\pi^2(\hbar c)^3}\left[1+4C_M\alpha Z^{2/3}\left(\frac{4}{243\pi}\right)^{1/3}\right]\left[1+\alpha C_M\left(\frac{4}{9\pi}\right)^{1/3}F(Z)\right]^{-4}. \tag{43}$$

Substituting Eq. (42) in (39) using (35) leads to the following lower bound for the radius of a white dwarf:

$$R_{\min} = \frac{\sqrt{3\pi}}{2}\xi_1\lambda_e\frac{m_{\rm P}}{m}\frac{y_e}{\gamma_e^\beta(A,Z)}\left[1+\alpha C_M\left(\frac{4}{9\pi}\right)^{1/3}F(Z)\right] \times\left[1+\alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3}Z^{2/3}\right]^{1/2}. \tag{44}$$

The factors in square brackets account for the electron-ion interactions: the first arises from the shift in the threshold density for the onset of electron captures whereas the second is due to the correction to the equation of state. Both factors lead to a reduction of the radius. Results of Eq. (44) are summarized in Table 3.

Table 3. Minimum radius (in km) of nonmagnetic stable white dwarfs.

^{12}C	^{16}O	^{20}Ne	^{24}Mg	^{56}Fe
958	1209	1743	2168	2791

The mass remains unchanged since it does not depend on the density and is still given by Eq. (38). However, departure from the polytropic equation of state $P = K\rho^{4/3}$ and the fact that the central density ρ_c is finite induce a slight reduction of the maximum mass. Following the perturbative approach described in Chapter 6 of Ref. 62 and based on the minimization of some approximation for the total energy of the star, we find

$$\frac{\delta M}{M} \approx -\frac{k_3}{k_2}\frac{3\pi}{(\gamma_e^\beta)^2}\left(\frac{3}{2\xi_1^4|\theta'(\xi_1)|^2}\right)^{1/3}\left[1+\alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3}Z^{2/3}\right]^{-1} \times\left[1+\alpha C_M\left(\frac{4}{9\pi}\right)^{1/3}F(Z)\right]^2 \tag{45}$$

where $k_2 = 0.639001$ and $k_3 = 0.519723$. Results are summarized in Table 4. The influence of electron captures and electron-ion interactions are all the more important that matter contains heavier elements: whereas the overall reduction of the Chandrasekhar mass $M_{\rm Ch}$ amounts to 3% for carbon, it reaches 13% for iron.

The stability of a white dwarf can be further limited by general relativity, as first shown by Kaplan.[63] The critical density above which the stellar core becomes unstable can be estimated from the minimization of the total energy. Including the electrostatic correction in Eq.(6.10.28) of Ref. 62 we find

$$\rho_{\rm GR} = \frac{16k_3(k_2)^2}{(3\pi^2)^{2/3}k_4(k_1)^2}\frac{M'(A,Z)^2}{Z^2\lambda_e^3 m_e}\left[1+\alpha\frac{8C_M}{6}\left(\frac{4}{9\pi}\right)^{1/3}Z^{2/3}\right]^{-2}\ {\rm g\ cm^{-3}}, \tag{46}$$

Table 4. Maximum mass (in solar units) of nonmagnetic white dwarfs for an ultra-relativistic electron Fermi gas (first line), with electrostatic correction (second line) and electron captures (third line).

^{12}C	^{16}O	^{20}Ne	^{24}Mg	^{56}Fe
1.456	1.457	1.457	1.458	1.259
1.424	1.418	1.411	1.406	1.184
1.413	1.401	1.377	1.353	1.095

where $k_1 = 1.75579$ and $k_4 = 0.918294$. Because $C_M < 0$, electron-ion interactions thus make the star more stable. We have also solved the Tolman-Oppenheimer-Volkoff[64,65] equations for calculating the whole sequence of white dwarfs in full general relativity. In this way, we have determined more accurately the central (mass-energy) density $\rho_{\rm GR}$ at which $dM/d\rho_{\rm GR} = 0$ marking the onset of instability. The hydrostatic equilibrium equations were integrated from the center of the star up to the point where electrons start to bind to nuclei at the density $\rho_{\rm eip} = ZM'(A,Z)/a_0^3$ with a_0 the Bohr radius. Results are collected in Table 5. The values are systematically lower than those estimated from Eq. (46). Comparing Tables 1 and 5 shows that the central density is limited by general relativity rather than electron captures in carbon white dwarfs. Their mass-density relation is plotted in Fig. 3. The minimum radii and maximum masses of white dwarfs with different composition are indicated in Tables 6 and 7 respectively (note that the mass and radius of carbon white dwarfs at the onset of electron captures are respectively $1.484 M_\odot$ and 855 km). For comparison, results obtained by solving numerically the Newtonian hydrostatic equilibrium equations are also given. Examining Tables 3 and 6 reveals that the polytropic approximation is more reliable for estimating the maximum mass than the minimum radius.

Table 5. Highest density $\rho_{\rm GR}$ (in g cm^{-3}) in nonmagnetic stable white dwarfs in general relativity, as calculated by minimization of the approximate total energy (first line) and by solving the hydrostatic equilibrium equations (second line).

^{16}O	^{12}C	^{20}Ne	^{24}Mg	^{56}Fe
2.73×10^{10}	2.72×10^{10}	2.75×10^{10}	2.76×10^{10}	3.31×10^{10}
2.39×10^{10}	2.40×10^{10}	2.42×10^{10}	2.44×10^{10}	2.93×10^{10}

The influence of a very strong magnetic field on the structure of white dwarfs was studied in Ref. 24 within the theory of polytropes making use of the fact that $P \propto \rho^2$ in the strongly quantizing regime. However, it was soon realized that this assumption is unrealistic and that the influence of the magnetic field itself on the stellar structure cannot be ignored.[27,34,66–68] Moreover, the magnetic-field configuration inside the

Table 6. Minimum radius (in km) of nonmagnetic stable white dwarfs in Newtonian theory (first line) and general relativity (second line).

^{12}C	^{16}O	^{20}Ne	^{24}Mg	^{56}Fe
857.7	1055	1447	1736	2082
1010	1052	1444	1732	2079

Table 7. Maximum mass (in solar units) of nonmagnetic stable white dwarfs in Newtonian theory (first line) and general relativity (second line).

^{12}C	^{16}O	^{20}Ne	^{24}Mg	^{56}Fe
1.414	1.403	1.382	1.363	1.118
1.383	1.378	1.366	1.350	1.111

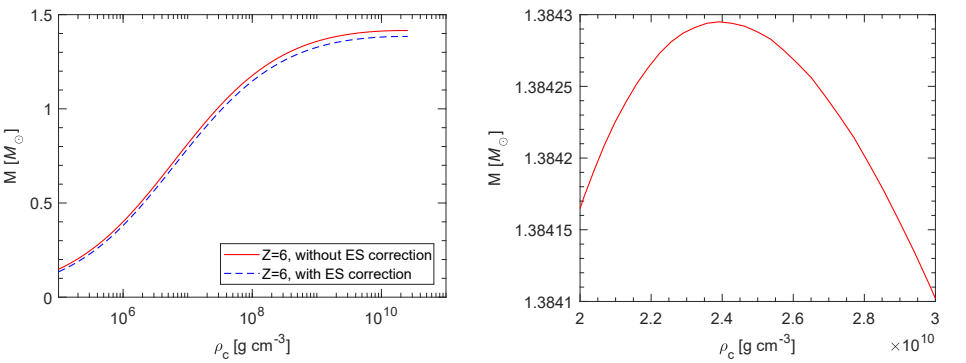


Fig. 3. Gravitational mass in solar masses versus central (mass-energy) density in g cm^{-3} of carbon white dwarfs in general relativity: (a) comparing results with and without electrostatic corrections, (b) closer view of results with electrostatic corrections around the critical point.

star should be calculated from Maxwell's equations consistently with the stellar equilibrium equations. For all these reasons, the structure of magnetic white dwarfs is not easily amenable to a simple analytical treatment.

To study the influence of strong magnetic fields and the role of electron captures in white dwarfs, we have thus computed fully self-consistently numerical solutions of the Einstein-Maxwell equations using the LORENE library,[69] suitably extended to allow for magnetic-field dependent equations of state and magnetization effects.[70] We have found that purely poloidal magnetic fields of order 10^{14} G lead to super Chandrasekhar white dwarfs with a mass $\sim 2M_\odot$. Although such magnetic fields

have been found to have a rather small influence on the equation of state, they induce extreme stellar deformations with the most massive white dwarfs adopting a donut-like shape.[36] Similarly to nonmagnetic white dwarfs, electron captures limit the maximum mass and the minimum radius of magnetic white dwarfs, as summarized in Table 8. However, the distortion of the star induced by magnetic fields tends to lower their density. For the most extreme configurations, the maximum density thus lies below the threshold density for the onset of electron captures. The mass-radius relations for magnesium and neon white dwarfs are plotted in Fig. 4.

Table 8. Maximum mass (in solar units) of magnetic white dwarfs, as determined by extreme deformations or electron captures (values in parentheses) for different magnetic moments $\mathcal{M}$ in A m^2.

$\mathcal{M}$	^{12}C	^{16}O	^{20}Ne	^{24}Mg
10^{33}	1.41 (1.41)	1.41 (1.40)	1.40 (1.38)	1.39 (1.38)
5×10^{33}	1.60 (1.50)	1.59 (1.46)	1.59 (1.41)	1.58 (1.36)
10^{34}	2.00 (1.86)	1.97 (1.67)	1.97 (1.51)	1.96 (1.45)
2×10^{34}	1.99 (1.99)	1.96 (1.96)	1.97 (1.95)	1.95 (1.73)
3×10^{34}	1.96 (1.96)	1.94 (1.94)	1.98 (1.98)	1.97 (1.97)

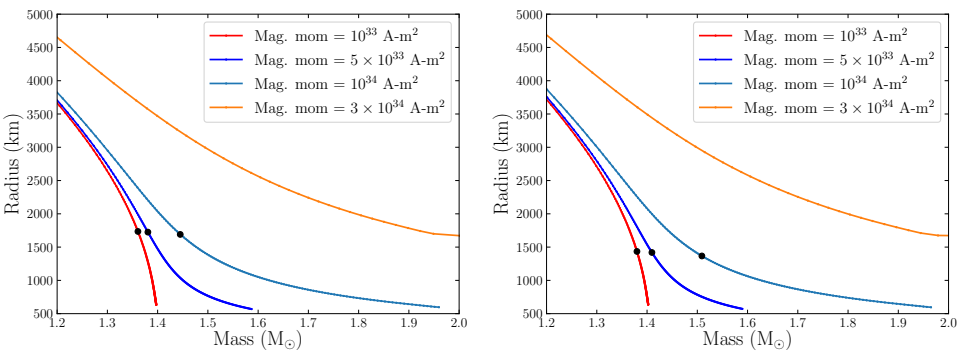


Fig. 4. Radius (in km) versus gravitational mass (in solar units) of white dwarfs for different magnetic moments: (a) for ^{24}Mg, (b) for ^{20}Ne. The onset of electron captures is marked by a filled circle.

The global stability criterion $\partial M/\partial \rho > 0$ still remains valid for magnetic white dwarfs provided the derivatives are evaluated for a fixed magnetic moment $\mathcal{M}$. As shown in Fig. 5, this instability is entirely removed by the presence of strong magnetic fields. In other words, the magnetic field makes the star more stable. In particular, the maximum mass of magnetic white dwarfs made of carbon is not limited by general relativity as their nonmagnetic relatives but by electron captures.

For magnetic white dwarfs to be the progenitors of overluminous type Ia supernova, i.e. to have masses $\sim 2M_\odot$, their magnetic field must be strong enough. Their minimum observable polar magnetic fields are indicated in Table 9 for different compositions. In all cases, the magnetic dipole moment is 3×10^{34} A m^2.

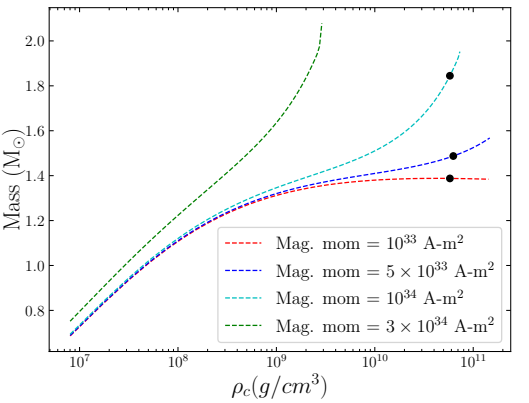


Fig. 5. Gravitational mass (in solar masses) versus central (mass-energy) density in g cm^{-3} of carbon white dwarfs in general relativity with given magnetic moments. The onset of electron captures is marked by a filled circle.

Table 9. Polar magnetic field (in 10^{13} G) of white dwarfs with $\mathcal{M} = 3 \times 10^{34}$ A m^2.

^{12}C	^{16}O	^{20}Ne	^{24}Mg
2.73	2.55	2.89	2.85

5. Conclusions

The stability of white dwarfs is limited by the onset of electron captures by nuclei in their core. We have presented very accurate analytical formulas for calculating the threshold electron Fermi energy, mass density and pressure in cold dense Coulomb plasmas with magnetic fields, taking into account Landau-Rabi quantization of electron motion. In particular, the density exhibits typical quantum oscillations associated with the filling of energy levels. As a consequence, electron captures in magnetic white dwarfs may occur at a lower or higher density than in their non-magnetic relatives depending on the magnetic field strength. The lowest possible density is found to lie about 25% below its value in the absence of magnetic field essentially independently of the composition. On the contrary, the density is not

limited from above and increases almost linearly with the magnetic field strength in the strongly quantizing regime.

Taking into account electron-ion interactions using the polytropic approximation $P \approx K\rho^{4/3}$ with K given by Eq. (37), we have explicitly shown how electron captures alter the maximum mass of nonmagnetic white dwarfs and set a lower limit to their radius. We have also solved numerically the hydrostatic equilibrium equations using the full equation of state, both in Newtonian theory and in general relativity. We have found that electron captures reduce the maximum mass of white dwarfs by 3-13% compared to the Chandrasekhar model.

Solving the Einstein-Maxwell equations taking into account the magnetization of dense matter, we have found that white dwarfs with purely poloidal magnetic fields can be significantly more massive than their nonmagnetic relatives. We have also shown that the presence of strong magnetic fields makes the star more stable. The maximum mass of magnetic white dwarfs (including those made of carbon) is thus solely limited by electron captures. In turn, the large stellar deformations induced by the strongest magnetic fields lower the stellar density below the electron capture threshold. White dwarfs with polar magnetic fields exceeding 10^{13} G could thus be massive enough to explain overluminuous type Ia supernova independently of the composition of their core. On the other hand, a purely poloidal magnetic-field configuration is unstable. The magnetic field in super Chandrasekhar white dwarfs is expected to have both poloidal and toroidal components. The question as to whether such stars can exist still remains open.

Although the present study was focused on white dwarfs, it may be also of interest for neutron stars as electron captures by nuclei constituting the outer crust of magnetars may explain their persistent luminosity and outbursts.[71] The general analytical formulas for the threshold density and pressure could thus be applied to determine the precise locations of these reactions.

Acknowledgments

The work of N.C. was financially supported by F.R.S.-FNRS under Grant No. IISN 4.4502.19. L. P. is a FRIA grantee of F.R.S.-FNRS. This work was also supported by COST CA16214 and the CNRS International Research Project "Origine des éléments lourds dans l'univers: Astres Compacts et Nucléosynthèse (ACNu)".

References

1. R. H. Fowler, On dense matter, *MNRAS* **87**, 114 (December 1926).
2. E. C. Stoner, V. the limiting density in white dwarf stars, *Philos. Mag.* **7**, 63 (1929).
3. W. Anderson, Über die Grenzdichte der Materie und der Energie, *Z. Phys.* **56**, 851 (November 1929).
4. E. C. Stoner, Lxxxvii. the equilibrium of dense stars, *Philos. Mag.* **9**, 944 (1930).
5. S. Chandrasekhar, The Maximum Mass of Ideal White Dwarfs, *ApJ* **74**, p. 81 (July 1931).

6. D. G. Yakovlev, P. Haensel, G. Baym and C. Pethick, Lev Landau and the concept of neutron stars, *Physics Uspekhi* **56**, 289 (March 2013).
7. J. Frenkel, Zur wellenmechanischen Theorie der metallischen Leitfähigkeit, *Z. Phys.* **47**, 819 (November 1928).
8. D. G. Yakovlev, The article by Ya I Frenkel' on 'binding forces' and the theory of white dwarfs, *Physics Uspekhi* **37**, 609 (June 1994).
9. S. Chandrasekhar, The highly collapsed configurations of a stellar mass (Second paper), *MNRAS* **95**, 207 (January 1935).
10. T. E. Sterne, The equilibrium theory of the abundance of the elements: A statistical investigation of assemblies in equilibrium in which transmutations occur, *MNRAS* **93**, p. 736 (June 1933).
11. F. Hund, Materie unter sehr hohen Drucken und Temperaturen, *Ergebnisse der exakten Naturwissenschaften* **15**, p. 189 (January 1936).
12. W. Baade and F. Zwicky, Minutes of the stanford meeting, December 15-16, 1933, *Phys. Rev.* **45**, 130 (Jan 1934).
13. L. Landau, Origin of Stellar Energy, *Nature* **141**, 333 (February 1938).
14. G. Gamow, Physical Possibilities of Stellar Evolution, *Phys. Rev.* **55**, 718 (April 1939).
15. S. Chandrasekhar, The White Dwarfs and Their Importance for Theories of Stellar Evolution, in *Conférences du Collège de France, Colloque International d'Astrophysique III, 17-23 Juillet 1939*, 1941.
16. E. Schatzman, Influence of the nucleon-electron equilibrium on the internal structure of white dwarfs, *Astronomiceskij Zhurnal* **33**, p. 800 (1956).
17. E. Schatzman, *White Dwarfs* (North-Holland Publishing Company, 1958).
18. B. K. Harrison, M. Wakano and J. A. Wheeler, Matter-energy at high density: End point of thermonuclear evolution, in *Onzième Conseil de Physique Solvay, Stoops, Brussels*, 1958.
19. T. Hamada and E. E. Salpeter, Models for Zero-Temperature Stars., *ApJ* **134**, p. 683 (November 1961).
20. K. Langanke, G. Martínez-Pinedo and R. G. T. Zegers, Electron capture in stars, *Rep. Prog. Phys.* **84**, p. 066301 (June 2021).
21. A. S. Jermyn, J. Schwab, E. Bauer, F. X. Timmes and A. Y. Potekhin, Skye: A Differentiable Equation of State, *ApJ* **913**, p. 72 (May 2021).
22. L. Ferrario, D. de Martino and B. T. Gänsicke, Magnetic White Dwarfs, *Space Sci. Rev.* **191**, 111 (October 2015).
23. K. Fujisawa, S. Yoshida and Y. Eriguchi, Axisymmetric and stationary structures of magnetized barotropic stars with extremely strong magnetic fields deep inside, *MNRAS* **422**, 434 (May 2012).
24. U. Das and B. Mukhopadhyay, New Mass Limit for White Dwarfs: Super-Chandrasekhar Type Ia Supernova as a New Standard Candle, *Phys. Rev. Lett.* **110**, p. 071102 (February 2013).
25. U. Das and B. Mukhopadhyay, GRMHD formulation of highly super-Chandrasekhar magnetized white dwarfs: Stable configurations of non-spherical white dwarfs, *J. Cosmolol. Astropart. P.* **2015**, p. 016 (May 2015).
26. S. Subramanian and B. Mukhopadhyay, GRMHD formulation of highly super-Chandrasekhar rotating magnetized white dwarfs: Stable configurations of non-spherical white dwarfs, *MNRAS* **454**, 752 (November 2015).
27. P. Bera and D. Bhattacharya, Mass-radius relation of strongly magnetized white dwarfs: Nearly independent of Landau quantization, *MNRAS* **445**, 3951 (December 2014).

28. B. Franzon and S. Schramm, Effects of strong magnetic fields and rotation on white dwarf structure, *Phys. Rev. D* **92**, p. 083006 (October 2015).
29. P. Bera and D. Bhattacharya, Mass-radius relation of strongly magnetized white dwarfs: Dependence on field geometry, GR effects and electrostatic corrections to the EOS, *MNRAS* **456**, 3375 (March 2016).
30. P. Bera and D. Bhattacharya, A perturbation study of axisymmetric strongly magnetic degenerate stars: The case of super-Chandrasekhar white dwarfs, *MNRAS* **465**, 4026 (March 2017).
31. G. A. Shul'Man, Degenerate equilibrium stellar configurations with a frozen-in magnetic field, *Sov. Astron.* **20**, p. 689 (December 1976).
32. G. A. Shulman, Equilibrium Configurations of Degenerate Superstars in the Quantum Limit of a Frozen-In Magnetic Field, *Sov. Astron.* **33**, p. 393 (August 1989).
33. G. A. Shulman, Equilibrium and Stability of Degenerate Stars with a Frozen-In Quantizing Magnetic Field, *Sov. Astron.* **36**, p. 398 (August 1992).
34. N. Chamel, A. F. Fantina and P. J. Davis, Stability of super-Chandrasekhar magnetic white dwarfs, *Phys. Rev. D* **88**, p. 081301 (October 2013).
35. N. Chamel, E. Molter, A. F. Fantina and D. P. Arteaga, Maximum strength of the magnetic field in the core of the most massive white dwarfs, *Phys. Rev. D* **90**, p. 043002 (August 2014).
36. D. Chatterjee, A. F. Fantina, N. Chamel, J. Novak and M. Oertel, On the maximum mass of magnetized white dwarfs, *MNRAS* **469**, 95 (July 2017).
37. E. Otoniel, B. Franzon, G. A. Carvalho, M. Malheiro, S. Schramm and F. Weber, Strongly Magnetized White Dwarfs and Their Instability Due to Nuclear Processes, *ApJ* **879**, p. 46 (July 2019).
38. N. Chamel and A. F. Fantina, Electron capture instability in magnetic and nonmagnetic white dwarfs, *Phys. Rev. D* **92**, p. 023008 (July 2015).
39. G. Nelemans and T. M. Tauris, Formation of undermassive single white dwarfs and the influence of planets on late stellar evolution, *A&A* **335**, L85 (July 1998).
40. J. Liebert, P. Bergeron, D. Eisenstein, H. C. Harris, S. J. Kleinman, A. Nitta and J. Krzesinski, A Helium White Dwarf of Extremely Low Mass, *ApJ* **606**, L147 (May 2004).
41. O. G. Benvenuto and M. A. De Vito, The formation of helium white dwarfs in close binary systems - II, *MNRAS* **362**, 891 (September 2005).
42. K. Nomoto, Evolution of 8-10 solar mass stars toward electron capture supernovae. I - Formation of electron-degenerate O + Ne + Mg cores, *ApJ* **277**, 791 (February 1984).
43. J. A. Panei, L. G. Althaus and O. G. Benvenuto, Mass-radius relations for white dwarf stars of different internal compositions, *A&A* **353**, 970 (January 2000).
44. S. Catalán, I. Ribas, J. Isern and E. García-Berro, WD0433+270: An old Hyades stream member or an Fe-core white dwarf?, *A&A* **477**, 901 (January 2008).
45. J. Isern, R. Canal and J. Labay, The Outcome of Explosive Ignition of ONeMg Cores: Supernovae, Neutron Stars, or "Iron" White Dwarfs?, *ApJ* **372**, p. L83 (May 1991).
46. G. C. Jordan, H. B. Perets, R. T. Fisher and D. R. van Rossum, Failed-detonation Supernovae: Subluminous Low-velocity Ia Supernovae and their Kicked Remnant White Dwarfs with Iron-rich Cores, *ApJ* **761**, p. L23 (December 2012).
47. D. Lunney, J. M. Pearson and C. Thibault, Recent trends in the determination of nuclear masses, *Rev. Mod. Phys.* **75**, 1021 (August 2003).
48. D. Peña Arteaga, M. Grasso, E. Khan and P. Ring, Nuclear structure in strong magnetic fields: Nuclei in the crust of a magnetar, *Phys. Rev. C* **84**, p. 045806 (October 2011).

49. M. Stein, J. Maruhn, A. Sedrakian and P. G. Reinhard, Carbon-oxygen-neon mass nuclei in superstrong magnetic fields, *Phys. Rev. C* **94**, p. 035802 (September 2016).
50. S. Subramanian and B. Mukhopadhyay, GRMHD formulation of highly super-Chandrasekhar rotating magnetized white dwarfs: Stable configurations of non-spherical white dwarfs, *MNRAS* **454**, 752 (November 2015).
51. I. I. Rabi, Das freie Elektron im homogenen Magnetfeld nach der Diracschen Theorie, *Z. Phys.* **49**, 507 (July 1928).
52. L. Landau, Diamagnetismus der Metalle, *Z. Phys.* **64**, 629 (September 1930).
53. P. Haensel, A. Y. Potekhin and D. G. Yakovlev, *Neutron Stars 1: Equation of state and structure* (Springer, 2007).
54. D. A. Baiko, A. Y. Potekhin and D. G. Yakovlev, Thermodynamic functions of harmonic Coulomb crystals, *Phys. Rev. E* **64**, p. 057402 (November 2001).
55. J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities* (Oxford University Press, London, 1932).
56. D. A. Baiko, Coulomb crystals in the magnetic field, *Phys. Rev. E* **80**, p. 046405 (October 2009).
57. N. Chamel and A. F. Fantina, Binary and ternary ionic compounds in the outer crust of a cold nonaccreting neutron star, *Phys. Rev. C* **94**, p. 065802 (Dec 2016).
58. M. Wang, W. J. Huang, F. G. Kondev, G. Audi and S. Naimi, The AME 2020 atomic mass evaluation (II). Tables, graphs and references, *Chin. Phys. C* **45**, p. 030003 (March 2021).
59. C. O. Dib and O. Espinosa, The magnetized electron gas in terms of Hurwitz zeta functions, *Nucl. Phys. B* **612**, 492 (October 2001).
60. N. Chamel and Z. K. Stoyanov, Analytical determination of the structure of the outer crust of a cold nonaccreted neutron star: Extension to strongly quantizing magnetic fields, *Phys. Rev. C* **101**, p. 065802 (June 2020).
61. S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Dover, 1957).
62. S. L. Shapiro and S. A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects* (John Wiley& Sons, 1983).
63. S. A. Kaplan, Sverkhplotnyè Zvezdy, *Naukovy Zapiski* **15**, 109 (1949).
64. R. C. Tolman, Static Solutions of Einstein's Field Equations for Spheres of Fluid, *Phys. Rev.* **55**, 364 (February 1939).
65. J. R. Oppenheimer and G. M. Volkoff, On Massive Neutron Cores, *Phys. Rev.* **55**, 374 (February 1939).
66. R. Nityananda and S. Konar, Strong constraints on magnetized white dwarfs surpassing the Chandrasekhar mass limit, *Phys. Rev. D* **89**, p. 103017 (May 2014).
67. J. G. Coelho, R. M. Marinho, M. Malheiro, R. Negreiros, D. L. Cáceres, J. A. Rueda and R. Ruffini, Dynamical Instability of White Dwarfs and Breaking of Spherical Symmetry Under the Presence of Extreme Magnetic Fields, *ApJ* **794**, p. 86 (October 2014).
68. D. L. Cáceres, J. A. Rueda and R. Ruffini, On the stability of ultra-magnetized white dwarfs, *J. Korean Phys. Soc.* **65**, 846 (September 2014).
69. E. Gourgoulhon, P. Grandclément, J.-A. Marck, J. Novak and K. Taniguchi, LORENE: Spectral methods differential equations solver (August 2016).
70. D. Chatterjee, T. Elghozi, J. Novak and M. Oertel, Consistent neutron star models with magnetic-field-dependent equations of state, *MNRAS* **447**, 3785 (March 2015).
71. N. Chamel, A. F. Fantina, L. Suleiman, J.-L. Zdunik and P. Haensel, Heating in Magnetar Crusts from Electron Captures, *Universe* **7**, p. 193 (June 2021).

Massive hot white dwarfs: Consequences of finite temperature in the structure and on the onset of instabilities

Sílvia P. Nunes

Departamento de Física, Instituto Tecnológico de Aeronáutica, 12228-900 São José dos Campos, São Paulo, Brazil
E-mail: silviapn@ita.br

José D. V. Arbañil

Departamento de Ciencias, Universidad Privada del Norte, Avenida el Sol 461 San Juan de Lurigancho, 15434 Lima, Peru

Facultad de Ciencias Físicas, Universidad Nacional Mayor de San Marcos, Avenida Venezuela s/n Cercado de Lima, 15081 Lima, Peru

Manuel Malheiro

Departamento de Física, Instituto Tecnológico de Aeronáutica, 12228-900 São José dos Campos, São Paulo, Brazil

In this work, we investigate the observable white dwarfs with high-surface gravity in the framework of general relativity. We consider the stellar fluid composed of nucleons and electrons confined in a Wigner-Seitz cell surrounded by free photons. Besides, we implement a temperature depending on the mass density with the presence of an isothermal core. The impact of temperature on the equilibrium and stability of white dwarfs is observed. We compare our results with massive white dwarfs estimated from the Extreme Ultraviolet Explorer Survey and Sloan Digital Sky Survey. We find that the high-surface gravity white dwarfs are well described by our curves with higher central temperatures. Our results suggest that these hot massive stars detected are within the range of white dwarfs with more radial stability. Moreover, we note the radial instability is attained before the pycnonuclear reaction for $T_c \geq 1.0 \times 10^8$[K].

Keywords: White dwarfs; charged white dwarfs; type Ia supernovae explosions

Introduction

Observable white dwarfs' structure has been reported in works using the Extreme Ultraviolet Explorer Survey (EUVE) and Sloan Digital Sky Survey (SDSS).[1–4] Some of these stars have a great surface gravity ($\log(g/g_\odot) \geq 4.4$) and a high effective temperature ($T_{\rm eff} \gtrsim 10^4$[K]).[1,2] Due to their high surface gravity, it is important to investigate their instabilities against small radial perturbation and pycnonuclear reactions in general relativity scope considering temperature.

The studies concerning instabilities in white dwarfs[5,6] consider a completely degenerate star. To compare results with observable white dwarfs, we include temperature in the model. Thus, following recent white dwarfs models with finite temperature,[7,8] we analyze the structure and stability of hot white dwarf' and compare

the results with some observable data. In this article, we detail some conclusions of the previous work.[9]

We study the structure and stability of hot massive white dwarfs and we compare our results with those ones reported by EUVE and SDSS. In addition, we made an instability analysis against the pycnonuclear reactions and radial perturbations. We analyze the stellar fluid by considering nucleons and electrons confined in a Wigner-Seitz cell with surrounding free photons. For this case, the pressure and energy density are respectively considered as following

$$P = P_{\rm e} + P_{\rm N} + P_{\gamma} + P_{L}, \tag{1}$$

$$\varepsilon = \varepsilon_{\rm e} + \varepsilon_{\rm N} + \varepsilon_{\gamma} + \varepsilon_{L}, \tag{2}$$

with the subscripts e, N, and γ representing the respective contributions of electrons, nucleons, and photons. Due to the partially degeneracy, the electron contributions are obtained with an adaptive quadrature method.[8]

1. The stellar structure equations and boundary conditions

We consider the unperturbed white dwarf is made of by a perfect fluid, whose energy-momentum tensor is expressed of the form

$$T_{\mu\beta} = (P + \varepsilon)u_{\mu}u_{\beta} + Pg_{\mu\beta}. \tag{3}$$

The background metric employed to describe the spherically symmetric white dwarf, in Schwarzschild-like coordinates, is given by

$$ds^2 = -e^{\nu}dt^2 + e^{\lambda}dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2, \tag{4}$$

where the functions ν and λ dependent on the radial coordinate only. The nonzero Einstein field equation components, for the line element assumed, lead to obtain the stellar structure equations:

$$\frac{dm}{dr} = 4\pi\varepsilon^2, \tag{5}$$

$$\frac{dP}{dr} = -(P + \varepsilon)\left[4\pi rp + \frac{m}{r^2}\right]e^{\lambda}, \tag{6}$$

$$\frac{d\nu}{dr} = -\frac{2}{(p + \varepsilon)}\frac{dP}{dr}, \tag{7}$$

being m the mass inside the radius r. The potential metric e^{λ} is described as

$$e^{\lambda} = \left[1 - \frac{2m}{r}\right]^{-1}. \tag{8}$$

Since we are considering a temperature distribution T dependent on mass density ρ with the presence of an isothermal core, we assume that the parameters T and ρ are related by the equality

$$T/\rho^{2/3} = \text{constant}. \tag{9}$$

The process starts integrating the stellar structure equations from the center ($r = 0$) until the surface of the star ($r = R$). At the center of the star are considered the initial conditions:

$$m(0) = 0, \;\; \varepsilon(0) = \varepsilon_0, \text{ and } \nu(0) = \nu_0. \tag{10}$$

The surface of the star is determined at $P(r = R) = 0$.

2. Results

Due to our analysis be restricted to very massive WD, we can use a simple equation for the temperature distribution as Eq. (9). To ensure this affirmation, we compare the mass and radius obtained with the temperature distribution in Eq. (9) and the one found by Kritcher and collaborators[10] which considers a more complete processes as radiation and convection. To analyze massive WDs, we use the temperature distribution stated Eq. (9). By using this relation, we obtain similar masses to those reported by Kritcher's and collaborators[10] for stars near the maximum-mass limit. This is due to the more contribution of the WD mass coming from the degenerate core, which has a constant temperature. The envelope with a different temperature distribution should only modify the stellar radius. In this case, the percentage between the difference obtained by Eq. (9) and Kritcher's envelope, for $g \geq 10^9 [\text{cm s}^{-2}]$, is around 2%.

The surface gravity as a function of radial coordinate is plotted in Fig. 1 for several central temperatures and $M = 1.37 M_\odot$. From the panel, we note that the surface gravity increases with the radial coordinate until it reaches the maximum gravity point, hereafter, gravity decays monotonically with the increment of the radial coordinate.

The change of surface's gravity and total radius of the star with the increment of the central temperature is also observed in Fig. 1. For larger temperature, greater total stellar radius and lower surface's gravity are found. In this sense, WDs with smaller surface gravity -than those ones predicted with the cold catalyzed matter- are expected to have high central temperatures.

The mass against the total radius is shown in Fig. 2 for different values of central temperature. The full triangles in pink over the curves indicate the maximum mass points. The blue dots represent the threshold instability due to pycnonuclear reactions. In addition, in figure, some observational results obtained from the catalogs Nalezyty[2] in green diamond, and in Vennes[1] in brown squares. In all curves of the panel, we note that the mass grows with the surface's gravity until reaches the maximum mass point, thereafter, the mass decreases with the grows of gravity.

In Fig. 2, we observe that the analysis of the pycnonuclear reaction has to be considered for central temperatures below $T_c < 10^8 [\text{K}]$. For central temperatures above $10^8 [\text{K}]$, the secular stability point is reached before the pycnonuclear reaction. It is important to mention that the secular stability point is marked by the maximum mass point.

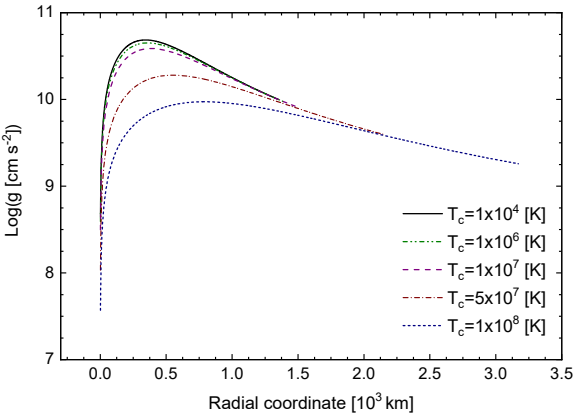


Fig. 1. Surface gravity as a function of radial coordinate for several central temperatures with a fixed $M = 1.37 M_{\odot}$.

Vennes,[1] and Nalezyty[2] have reported some very massive white dwarfs with very high surface gravity. In Fig. 2, these white dwarfs are matching with the curves of very high central temperatures and are very close to the instabilities thresholds. Indeed, one of them, with a mass of $M = 1.41 M_{\odot}$ can have a central temperature higher than 10^8[K]. From this, we could understand that these massive white dwarfs could be treated differently. Using the surface gravity and effective temperature reported by Vennes, we could investigate these stars by employing the values of effective temperature and surface gravity. Moreover, considering the general relativity effects, we found that the masses of these stars decrease, thus highlighting the importance of general relativity in the study of massive white dwarfs.

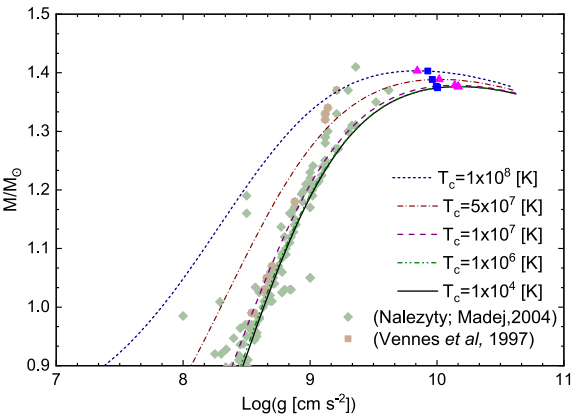


Fig. 2. Mass as a function of surface gravity for several central temperatures. The green diamonds[2] and the brown squares[1] represents observational data.

3. Conclusion

In this work, we investigate the structure and stability of hot white dwarfs. We regard the stellar fluid as composed of nucleons and electrons confined in a Wigner-Seitz cell, with free photons. We consider the temperature profile depending on the mass density with the presence of an isothermal core. We use the stellar structure equations to analyze the equilibrium configurations of white dwarfs and the stability against the pycnonuclear reactions and small radial perturbations.

Comparing our results with observational data, we found that the high masses white dwarfs provided by Vennes[1] could be described by our model. Morevoer, we dicussed about the importance of considering central temperature and general relativity in the study of massive white dwarfs.

Acknowledgments

We would like to thank Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Grant No. 2013/26258 − 4. SPN thanks Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant No. 140863/2017−6 and CAPES for the financial support. JDVA would like to thank the Universidad Privada del Norte and Universidad Nacional Mayor de San Marcos for funding - RR Nº 005753-2021-R/UNMSM under the project number B21131781. MM is grateful to CAPES and CNPq financial support.

References

1. S. Vennes, P. A. Thejll, R. Génova Galvan and J. Dupuis, Hot White Dwarfs in the Extreme Ultraviolet Explorer Survey. II. Mass Distribution, Space Density, and Population Age, *ApJ* **480**, 714 (May 1997).
2. M. Należyty and J. Madej, A catalogue of isolated massive white dwarfs. Mass distribution of massive star, *Astron. Astrophys.* **420**, 507 (June 2004).
3. D. Koester and S. O. Kepler, Carbon-rich (DQ) white dwarfs in the Sloan Digital Sky Survey, *Astron. Astrophys.* **628**, p. A102 (August 2019).
4. J. Madej, M. Należyty and L. G. Althaus, Mass distribution of DA white dwarfs in the First Data Release of the Sloan Digital Sky Survey, *Astron. Astrophys.* **419**, L5 (May 2004).
5. G. Chanmugam, Radial oscillations of zero-temperature white dwarfs and neutron stars below nuclear densities, *ApJ* **217**, 799 (November 1977).
6. E. Otoniel, B. Franzon, G. A. Carvalho, M. Malheiro, S. Schramm and F. Weber, Strongly Magnetized White Dwarfs and Their Instability Due to Nuclear Processes, *ApJ* **879**, p. 46 (July 2019).
7. S. M. de Carvalho, M. Rotondo, J. A. Rueda and R. Ruffini, Relativistic Feynman-Metropolis-Teller treatment at finite temperatures, *Phys. Rev. C* **89**, p. 015801 (January 2014).
8. K. A. Boshkayev, J. A. Rueda, B. A. Zhami, Z. A. Kalymova and G. S. Balgymbekov, Equilibrium structure of white dwarfs at finite temperatures, in *IJMPC*, 2016.

9. S. P. Nunes, J. D. V. Arbañil and M. Malheiro, The structure and stability of massive hot white dwarfs, *To appear in the Astrophysical Journal*, p. arXiv:2108.08238 (August 2021).
10. A. L. Kritcher, D. C. Swift, T. Döppner, B. Bachmann, L. X. Benedict, G. W. Collins, J. L. DuBois, F. Elsner, G. Fontaine, J. A. Gaffney, S. Hamel, A. Lazicki, W. R. Johnson, N. Kostinski, D. Kraus, M. J. MacDonald, B. Maddox, M. E. Martin, P. Neumayer, A. Nikroo, J. Nilsen, B. A. Remington, D. Saumon, P. A. Sterne, W. Sweet, A. A. Correa, H. D. Whitley, R. W. Falcone and S. H. Glenzer, A measurement of the equation of state of carbon envelopes of white dwarfs, *Nature* **584**, 51 (August 2020).

A study of the infrared emission of SGR/AXPs in a disk scenario and its implications for their origin

Sarah Villanova Borges

Department of Physics, University of Wisconsin-Milwaukee,
1900 E. Kenwood Boulevard, Milwaukee, WI 53211, USA
E-mail: svborges@uwm.edu

Soft Gamma-Ray Repeaters and Anomalous X-ray Pulsars (SGR/AXPs) are widely accepted to be magnetars, isolated neutron star (NS) with a huge decaying magnetic field. They can emit from radio up to hard X-rays, being infrared (IR) one of the most interesting and poorly studied energy ranges. The origin of this IR emission is still controversial, since some authors argue in favor of a magnetospheric emission, whereas others point out to a disk. In this work, we revisit the IR emission of SGR/AXPs in a disk perspective. We find that all SGR/AXPs with IR emission in the K-band are consistent with the presence of an irradiated disk. Even though this finding does not rule out any of the proposed natures, it is a piece of new evidence to confirm the presence of disk around this class. If the disk origin where unambiguously confirmed, this would have important impacts on understanding the origin of SGR/AXPs.

Keywords: Soft Gamma-Ray Repeaters; Anomalous X-ray Pulsars; infrared emission.

1. Introduction

The soft gamma-repeaters/anomalous X-ray pulsars (SGR/AXPs) are observationally characterized by a quiescent soft X-ray with luminosity in the range 10^{29} – 10^{35} erg.s^{-1}, period of 2 – 12 s, and spin-down of 10^{-15} to 10^{-10} s.s^{-1} see, e.g.,[24]. Some SGR/AXPs also present emission in other energy ranges, such as radio, optical, infrared (IR), and hard X-rays, as well as soft gamma-ray flares events. SGR/AXPs are widely thought to be magnetars, isolated neutron stars (NSs) emitting by the decay of their ultra-strong magnetic field, in the range of $10^{13} - 10^{15}$ G.[6] There is also alternative models, such as the NS accreting scenario,[36] the white dwarf (WD) pulsar model,[20] and the accreting WD model.[1]

The near-infrared (NIR) emission was confirmed or proposed for a total of 13 SGR/AXPs[15,24,31] and its origin is still debated. For the AXP 4U 0142+61 , the IR emission is well modeled by an irradiated disk, reinforced by an unconfirmed silicate feature around 9.5 μm.[39,40] Nonetheless, the lack of mid-IR emission during the 2007 flare of 1E 1048.1-5937 seems to rule out the presence of a disk in this object.[38] Thus, some authors still argues in favor of an IR magnetospheric emission.[13]

Two classical observational approaches can confirm magnetospheric origin in the IR emission of SGR/AXPs: search for polarization and pulsed emission. High polarization and pulsed fraction are expected for a magnetospheric emission. However, a low pulsed fraction is consistent with both models. Moreover, low polarization

discards magnetospheric synchrotron and curvature emission but cannot distinguish between a cyclotron emission and a disk. In this sense, an extensive study of IR polarization and periodic emission can be inconclusive.

For that reason, we suggest a distinct method to determine the IR origin of SGR/AXPs. We propose the search for a correlation between the luminosity in IR and X-rays that can be confronted with the expected correlation for a disk and magnetospheric models. In this sense, the disk could represent the best path to confirm the origin of SGR/AXPs. For instance, if the leftover of a supernova event formed the disk, it probably contains iron, silicon, oxygen, helium, and traces of hydrogen.[17] In this sense, an IR spectroscopy of 4U 0142+61, the magnetar with the brightest quiescent emission, could give essential clues to the origin of the class.

This proceeding aims to study the IR emission of SGR/AXPs and is organized as follows. In Section 2, we present the disk model we have used to fit the data. In section 3, we present the correlation between IR and X-rays. In Section 4, we discuss the consequences of the results for the origin of SGR/AXPs, slow rotating NSs, and the period of Fast Radio Bursts (FRBs). In Section 5, we summarize our findings.

2. An irradiated disk model for isolated neutrons stars

To study the expected correlation between IR and X-rays for an irradiated disk emission, we are going to consider the same irradiated disk model used to fit 4U 0142+61.[39] The emitted flux of a debris disk in a given frequency ν can be expressed by[8,27]:

$$F_{disk} = \frac{2\cos(i)}{D^2} \times \int_{R_{in}}^{R_{out}} B_\nu(T_{irr}(R)) R dR; \tag{1}$$

where, i is the inclination of the disk in respect to the observer, D is the distance to the observer, R is the distance from each infinitesimal part of the disk to the NS, $T_{irr}(R)$ is the debris disk temperature, and B_ν is the emitted spectrum of each element of the disk. Meanwhile T_{irr} is given by[37]:

$$T_{irr}(R) = 25,280 \text{ K } f^{2/7} \left(\frac{L_X}{L_E}\right)^{2/7} \left(\frac{R_\odot}{R}\right)^{3/7}, \tag{2}$$

where L_E is the Eddington luminosity, $R_\odot$ is the radius of the Sun and the f parameter related with the albedo. Since we expect all SGR/AXPs to have similar temperature and composition, f can be treated as a constant.

With those equations in hand, we can estimate the expected correlation between IR and X-rays luminosity in IR. Calling $x = \frac{h\nu}{kT(R)}$, we have:

$$R \propto x^{7/3} \left(\frac{k}{h\nu}\right)^{7/3} \left(f\frac{L_X}{L_E}\right)^{2/3}; \tag{3}$$

$$dR \propto x^{4/3} \left(\frac{k}{h\nu}\right)^{7/3} \left(f\frac{L_X}{L_E}\right)^{2/3} dx; \tag{4}$$

Considering $T_{out} \sim 500$ K and $T_{in} \sim 1500$ K, which is consistent with the temperature range for 4U 0142+61 and also for disk in isolated WDs,[11,39] the flux equation can be approximated to:

$$F_{Ks} \propto \frac{4\pi h \cos i \nu^3}{c^2 d^2} \left((L_X)^{4/3} \left(\frac{k}{h\nu} \right)^{14/3} \right) \times \int_{T=500K}^{T=1500K} \frac{x^{11/3} dx}{e^x - 1}, \tag{5}$$

where all parameters are constant, except for F_{Ks}, $\cos i$, and L_x. If we define the monochromatic luminosity in the *Ks-band* such as:

$$\begin{aligned} L_{Ks} &= 4\pi D^2 \int_0^\infty F_{Ks} \delta(\nu - \nu_{Ks}) d\nu \\ &= 4\pi D^2 F_{Ks} \nu_{Ks}, \end{aligned} \tag{6}$$

we find the correlation:

$$L_{Ks} \propto (\cos i) L_{XR}^{1.33}.$$

Since $0 < \cos i < 1$, we expect, at most, one order of magnitude change for a given object. Thus, for a correlation ranging in several orders of magnitude the expected correlation between the monochromatic luminosity in the Ks-band and the bolometric X-rays luminosity is can be simplified to:

$$L_{Ks} \propto L_{XR}^{1.33}. \tag{7}$$

It is important to highlight that this correlation considers that the disk properties (albedo, inner and outer temperatures of the dusty region) are similar among all objects. Moreover, this correlation discards the contribution of two different mechanisms in the Ks wavelength (such as a non-thermal component along with the disk).

3. A correlation between infrared and X-rays

The search for correlations is a valuable tool to understand the emission mechanism on different energy ranges. For SGR/AXPs, there are a few preliminary studies of the correlation between IR and X-rays. For some outbursts events, an enhancement of the IR emission was proposed.[10,28,30] In others, no correlation has been seen.[29,32,38] Nonetheless, those previous studies were limited since they have only examined the behavior of a single object per time along with the event and did not take into account the hard X-rays.

In this sense, we decide to revisit the possible correlation between IR and X-rays as a tool to confirm the irradiated disk origin. As discussed in sec. 2, if the disk is the main emission mechanics for IR, it is expected a correlation between the X-rays and IR luminosities. To search for such correlation, we chose the IR k_s-band since it

is the one with most observations in the magnetar class. For the bolometric X-rays luminosity, we have considered the sum of the luminosity of the 0.5-10 keV (L_s) and 15 - 60 keV (L_H) ranges. All IR and X-rays data are simultaneous or from close dates.

We have used observation from K_s, K' and K bands, considering that the luminosity in all bands is the same (justified since their central wavelengths are extremely close). Nonetheless, we have used the corrected zero-point magnitude for each specific filter. Moreover, not all objects have simultaneous Soft and Hard X-rays observation. When only one is available, the following equation was used to estimate the other[7]:

$$\frac{F_{hard}}{F_{soft}} = (0.59 \pm 0.07) \times (\dot{P}/10^{-11})^{0.51\pm0.05} \tag{8}$$

where $\dot{P}$ is the spin-down.

We first check the correlation using both Spearman ($\rho = 0.87$) and Pearson ($p = 0.93$) correlation coefficients, which points out to a strong correlation between the data. To find the best-fit from the data and its goodness, we consider a linear, logarithmic trend and R^2 defined by:

$$R^2 = \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \tag{9}$$

where y_i is the data, $\hat{y}_i$ is the fitted result, and $\bar{y}$ is the mean of the data. We use a 1σ error for all fits. Figure 1 shows the best correlation between the luminosity in the K_s-band (L_{Ks}) and X-rays (L_{XR}). The best fit is ($R^2 = 0.87$): $L_{IR} \propto L_S^{1.16\pm0.09}$, considering a linear fit without the error-bars (using the python class "stats.linregress") and $L_{IR} \propto L_S^{1.27\pm0.09}$ considering the error-bars (using the python class "scipy.odr.ODR"). We have used quiescence and outburst events, and we do not see any discrepancy in results.

For the fit with error bars, the expected slope of an irradiated disk is within the 1σ range, which favors the disk interpretation. Moreover, Fig. 1 considers quiescence/outbursts, and confirmed/candidates indistinctly, which argues in favor of: (i) the same mechanism is possibly responsible for the IR emission in quiescence and outbursts for the majority of objects; (2) candidate emission is possibly correct.

We have also compared the best fit with the IR emission of some few isolated NSs with detected emission in IR: (i) The rotation powered pulsars (RPPs) B0531+21, B0833-45, and B1509-58[12]; (ii) the XDIN RXJ0806.4-4123, and the Central Compact Object (CCO) 1E 161348-5055. We have also included the data from the SGR 1935+2154 in the F140W band. We have not used any of those points in the fit since they are not SGR/AXPs and/or the IR observation is not in the K-band. Even though they were not used in the fit, all but one are in the 90% prediction bound, showing that it is likely that the exact mechanism is responsible for the IR emission in magnetars and other pulsars with IR. This is interesting evidence in favor of a disk: fits using the magnetospheric curvature emission model can reproduce the IR

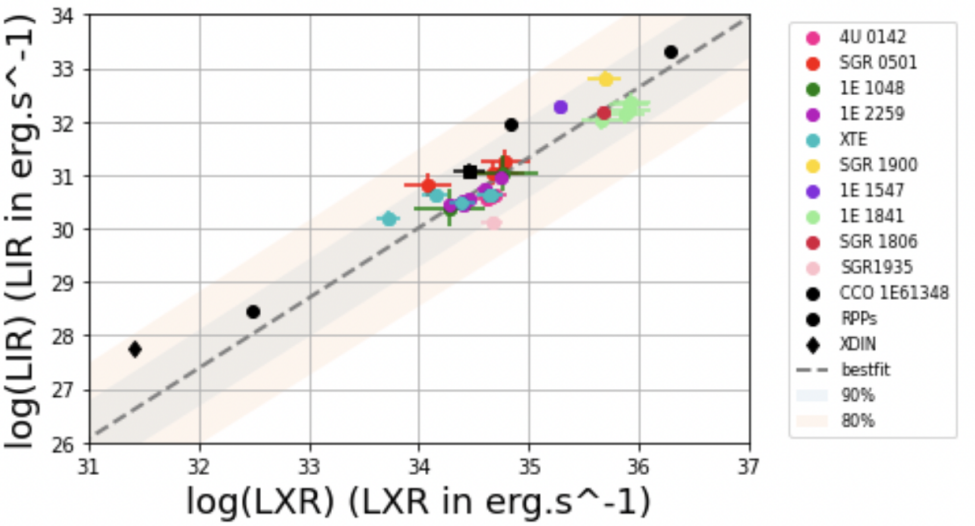


Fig. 1. Best-fit (dotted line) correlation between the luminosity in the Ks band and X-rays. The best-fit was estimated using only magnetars (colored points). The black squares, diamonds, and circles (1E 161348-5055, B0531+21, RXJ0806.4-4123, B0833-45, and B1509-58, respectively) were not used in the best-fit estimate and are only represented for comparison.

luminosity of magnetars using $B \sim 10^{14}$ G.[42] Nonetheless, the highest IR luminosity belongs to Crab Pulsar (B0531+21), a young pulsar with a dipole magnetic field of "only" 7e+12 G.

4. Importance of the disk on magnetars

4.1. *Origin of magnetars*

One of the biggest unanswered questions about magnetars is what are their progenitors. One of the first proposed models relates to massive main-sequence progenitors. A supernova explosion with energy higher than 10^{52} erg/s would enable the formation of such high magnetic fields, since part of the energy would be transferred to the ejecta via the strong magnetic coupling with the proto-neutron star.[35] This high-mass origin is corroborated by the presence of the magnetar CXOU J164710.2 in Westerlund 1, a young star cluster, with 4 Myr, which puts a lower limit in the mass of the progenitor on 40 $M_{\odot}$.[23]

Nonetheless, this massive origin has encountered some challenges. For instance, the explosion of the associated SNR for magnetars is $< 10^{52}$ erg/s, which put a higher limit on the mass of the progenitor in 20 $M_{\odot}$.[43] For those small progenitor masses, the formation of high magnetic fields would have to invoke the amplification of a fossil magnetic field.

For this main-sequence scenario, the presence of the disk can be challenging to explain. 1-D simulations have shown that fallback disk in NSs are formed only

on very fine-tuned SN events.[26] Nonetheless, if this is this case, an IR spectroscopy (with the to-be-launched James Webb Space Telescope) of 4U 0142+61could confirm the SN origin, since the leftover of a supernova event is expected to contain iron, silicon, oxygen, helium, and traces of hydrogen.[17]

Another possible mechanism, that have been long proposed, is the merger scenario. The merger of compact objects can generate isolated NSs with high magnetic fields and WD+WD, WD+NS, and NS+NS mergers have already been proposed as progenitors of magnetars.[16,18,21] This origin mechanism had gained a lot of attention in the last couple of years for being able to explain the occurrence of Gamma-Ray Bursts[22] and Fast Radio Bursts (FRBs) in a magnetar context.[21]

This merger origin would also be detectable in a IR spectroscopy. For instance, for a CO-WD/CO-WD merger, the disk is formed by the remains of the secondary and contains carbon and oxygen, with traces of other elements such as neon, magnesium, silicon, and nickel.[19] Moreover, the AIC event is expected to produce unusual neutron-rich isotopes.[5]

4.2. *Period of 1E 161348-5055*

The CCO 1E 161348-5055 is one of the most singular NSs: with a period of 6.67 hours, this is the slowest isolated NS known.[4] In the last decade, this curious object gained much attention when magnetar-like bursts activities were detected,[3] which raises a canon explanation evoking an ultra-strong magnetic field (in the order of $\sim 10^{15}$ G) and a fallback to explain how this object's large period came to happen.[9,33,41]

This way, proving that the IR emission of this object is consistent with a disk gives an observational evidence for the disk models explaining period of 1E 161348-5055.

4.3. *Periodicity in Fast Radio Bursts*

Since the discovery of FRB-like events in the SGR 1935+2154,[14] models invoking magnetars have been the focal point in understanding FRBs. In this sense, the presence of the disk in magnetars can emerge as an explanation for long periods in FRB events, such as the 16.35 days observed in the FRB 180916.J0158+65.[2]

In an isolated NS, the most plausible reason for such long period is precession. There are some models that were raised in the last years to explain such motion, with some invoking a disk, such as a forced precession caused by fallback disk,[34] or Lense-Thirring precession of the disk. Thus, proving the presence of the disk gives an extra piece of evidence for those models.

5. Conclusion

We searched for a correlation between IR and X-rays for magnetars and the result is consisted with an irradiated disk. We have also found that the results are consistent

the IR emission of some other classes of isolated NSs: RPP, XDINs and CCOs, which suggest a similar IR emission mechanism for all isolated NSs and favors a disk origin for IR. If the presence of such a disk was confirmed, its composition could help determine the origin of magnetars, as well as the long period of 1E 161348-5055 and the two-week-long period of the FRB 180916.J0158+65.

References

1. Borges S. V., Rodrigues C. V., Coelho J. G., Malheiro M., Castro M., 2020, Astrophysical Journal, 895, 26.
2. Chime/Frb Collaboration, Amiri, M., Andersen, B. C. et al., 2020, Nature, 582, 351.
3. D'A'ı, A., Evans, P. A., Burrows, D. N. et al., 2016, MNRAS, 463, 239D'Aı, A., Evans, P. A., Burrows, D. N. et al., 2016, MNRAS, 463, 2394.
4. De Luca, A., Caraveo, P. A., Mereghetti, S., Tiengo, A. & Bignami, G. F. 2006, Science, 313, 814.
5. Dessart, L., Burrows, A., Ott, C. D. et al., 2006, ApJ, 644, 1063.
6. Duncan, R. C. & Thompson, C. 1992, ApJ Lett., 392, L9.
7. Enoto, T., Shibata, S., Kitaguchi, T. et al., 2017, ApJS, 231, 8.
8. Frank, J., King, A. & Raine, D. J. 2002, Accretion Power in Astrophysics: Third Edition (Cambridge University Press), 398.
9. Ho, W. C. G. & Andersson, N. 2017, MNRAS, 464, L65.
10. Israel, G., Covino, S., Mignani, R. et al., 2005, A&A, 438, L1.
11. Jura, M., Farihi, J. & Zuckerman, B. 2007, ApJ, 663, 1285.
12. Kaplan, D. L. & Moon, D.-S. 2006, ApJ, 644, 1056.
13. Kaspi, V. M. & Beloborodov, A. M. 2017, ARA&A, 55, 261.
14. Kirsten, F., Snelders, M. P., Jenkins, M. et al., 2021, Nature Astronomy, 5, 414.
15. Levan, A., Kouveliotou, C. & Fruchter, A. 2018, ApJ, 854, 161.
16. Levan, A. J., Wynn, G. A., Chapman, R. et al., 2006, MNRAS, 368, L1.
17. Lin, D. N. C., Woosley, S. E. & Bodenheimer, P. H. 1991, Nature, 353, 827.
18. Liu, X. 2020, International Journal of Astronomy and Astrophysics, 10, 28.
19. Loren-Aguilar, P., Isern, J. & Garcıa-Berro, E. 2009, A&A, 500, 11933.
20. Malheiro, M., Rueda, J. A. & Ruffini, R. 2012, PASJ, 64, 56.
21. Margalit, B., Berger, E. & Metzger, B. D. 2019, ApJ, 886, 110.
22. Metzger, B. D., Giannios, D., Thompson, T. A., Bucciantini, N. & Quataert, E. 2011, MNRAS, 413, 2031.
23. Muno, M. P., Clark, J. S., Crowther, P. A. et al., 2006, ApJ Lett., 636, L41.
24. Olausen, S. A. & Kaspi, V. M. 2014, ApJS, 212, 6.
25. Ouyed, R., Leahy, D. & Niebergal, B. 2011, MNRAS, 415, 1590.
26. Perna, R., Duffell, P., Cantiello, M. & MacFadyen, A. I. 2014, ApJ, 781, 119.
27. Perna, R., Hernquist, L. & Narayan, R. 2000, ApJ, 541, 344.
28. Rea, N., Testa, V., Israel, G. L. et al., 2004, A&A, 425, L5.
29. Tam, C. R., Gavriil, F. P., Dib, R. et al., 2008, ApJ, 677, 503.
30. Tam, C. R., Kaspi, V. M., van Kerkwijk, M. H. & Durant, M. 2004, ApJ Lett., 617, L53.
31. Testa, V., Mignani, R. P., Hummel, W., Rea, N. & Israel, G. L.2018, MNRAS, 473, 3180.
32. Testa, V., Rea, N., Mignani, R. P. et al., 2008, A&A, 482, 607.
33. Tong, H., Wang, W., Liu, X. W. & Xu, R. X. 2016, ApJ, 833, 265.

34. Tong, H., Wang, W. & Wang, H.-G. 2020, Research in Astronomy and Astrophysics, 20, 142.
35. Turolla, R., Zane, S. & Watts, A. L. 2015, Reports on Progressin Physics, 78, 116901.
36. van Paradijs, J., Taam, R. E. & van den Heuvel, E. P. J. 1995, A&A, 299, L41.
37. Vrtilek, S. D., Raymond, J. C., Garcia, M. R. et al., 1990, A&A, 235, 162.
38. Wang, Z., Bassa, C., Kaspi, V. M., Bryant, J. J. & Morrell, N. 2008a, ApJ, 679, 1443.
39. Wang, Z., Chakrabarty, D. & Kaplan, D. L. 2006, Nature, 440, 772.
40. Wang, Z., Chakrabarty, D. & Kaplan, D. L. 2008b, in American Institute of Physics Conference Series, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More.
41. Xu, K. & Li, X.-D. 2019, ApJ, 877, 138.
42. Zane, S., Nobili, L. & Turolla, R. 2011, Astrophysics and Space Science Proceedings, 21, 329.
43. Zhou, P., Vink, J., Safi-Harb, S. & Miceli, M. 2019, A&A, 629, A51.

Particle acceleration and high energy emission in the white dwarf binaries AE Aquarii and AR Scorpii

P.J. Meintjes[1,*], S.T. Madzime[1,] Q. Kaplan[1], H.J van Heerden[1], K.K Singh[1,2], D.A.H. Buckley[1,3], P.A. Woudt[4] and R. Fender[5]

[1] *Department of Physics, University of the Free State, Bloemfontein, 9301, South Africa*
[*] *E-mail: MeintjPJ@ufs.ac.za*

[2] *Astrophysical Sciences Division, Bhabha Atomic Research Center, Mumbai, 400085, India*
E-mail: kksastro@barc.gov.in

[3] *South African Astronomical Observatory (SAAO), Observatory, 7925, Cape Town, South Africa,*
E-mail: dibnob@saao.ac.za

[4] *Department of Astronomy, University of Cape Town (UCT), Rondebosh, Cape Town, 7700 South Africa,*
E-mail: WoudtPA@ast.uct.ac.za

[5] *Department of Physics, University of Oxford, Oxford, OX1 3PU, United Kingdom*
E-mail: Rob.Fender@physics.ox.ac.uk

Here we present results from an in-depth search for pulsed emission from both close binary systems AE Aquarii (AE Aqr) and AR Scorpii (AR Sco) in radio and gamma-ray energies. Both systems were observed recently with the MeerKAT telescope, and combined with this, we utilized the combined 10 year Pass 8 Fermi-LAT dataset to search for pulsed gamma-ray emission from both white dwarfs in these systems. Pulsed emission was detected in MeerKAT data from both these close binary systems at a period that is at, or close to, the spin period of the white dwarf. The search for pulsed gamma-ray emission revealed pulsed emission at the spin period of the white dwarf of AE Aqr after selecting data sets with duration of 2 weeks that show excess emission above the 2 σ significance level. Braking these two-week sets up in 10 minute intervals and stacking the power spectra revealed pulsed emission at both the spin ($P* = 33.08$ s) and its associated first harmonic ($P_1 = 16.54$ s). A full 10 year analysis of the AR Sco data revealed pulsed emission at the spin period/beat period of the white dwarf, albeit at a lower significance level. Several control analyses were performed to verify the authenticity of the emission in both radio and gamma-rays, which will be discussed in the main text. The results of this study definitely reveal that both white dwarfs in these systems contain a particle accelerator that accelerates charged particles to high energies resulting in associated non-thermal radio and gamma-ray emission.

Keywords: Cataclysmic variable, white dwarf, pulsar, close binary system, particle acceleration, synchrotron radiation, gamma-rays

1. Introduction

Emission of Very High Energy (VHE) gamma-rays from the novalike variable AE Aqr had been reported in the late 1980's early 1990's by two independent groups (e.g. 1–3). These reports showed that AE Aqr display sporadic aperiodic burst-like gamma-ray emission with associated gamma-ray pulsations at, or close to, the spin period of the white dwarf. On occasions the first harmonic of the spin period ($P_1 = 16.54\,\mathrm{s}$) was also detected.

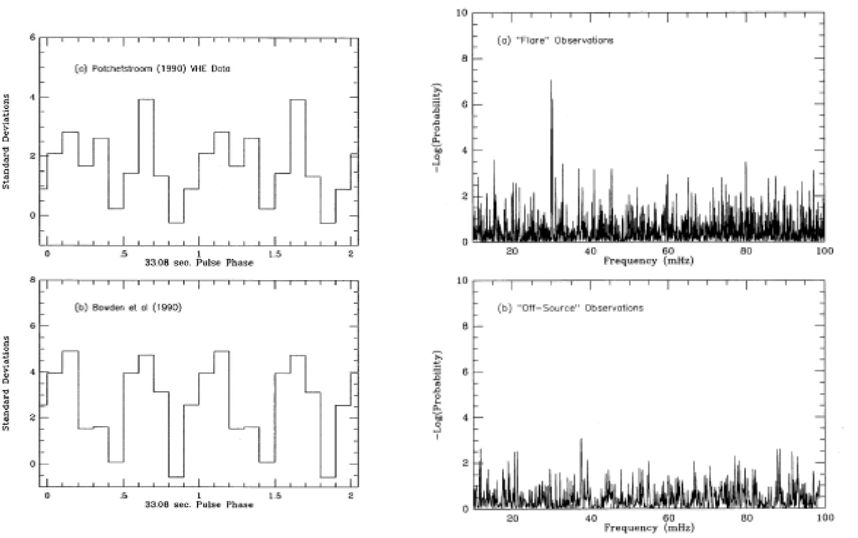


Fig. 1. (a) Pulsed VHE gamma-ray emission detected from AE Aqr at both the 33.08 s and 16.54 s during two contemporaneous burst-like VHE gamma-ray events observed by two independent groups. (b) The top panel represents an incoherently stacked power spectrum of VHE observations made simultaneous with optical flares and show clear indications of pulsed emission at the spin period ($P_* = 33.08\,\mathrm{S}$) of the white dwarf while the control off-source stacked power spectrum of observations made simultaneously with those above displays a white noise profile without any noticeable pulsed signal at or close to the spin period of the white dwarf of AE Aqr.

These reports have never been confirmed by independent follow-up studies (e.g. 4). However, the cadence in these follow-up studies was significantly poorer than the initial studies. The potential of white dwarfs as possible particle accelerators and gamma-ray sources came under the spotlight again with the discovery of the unique close binary AR Sco. The majority of AR Sco's Spectral Energy Distribution (SED) is the result of non-thermal emission since there is no indication of mass transfer and accretion in this system (e.g. 5). A follow-up study (6) revealed high levels of linear polarization in the optical light, a clear indication of a strong synchrotron component. It was shown (see e.g. 7) that the radio to X-ray SED of AR Sco can be explained by two component synchrotron spectrum produced by a pulsar-like process in the magnetosphere of the white dwarf (optical to X-ray), combined with synchrotron emission in the pumped magnetosphere of the secondary

star (radio to infrared). See also (8) for a detailed discussion of the properties of AR Sco.

The evidence of particle acceleration and accompanied non-thermal emission, combined with the availability of the new Fermi-LAT Pass 8 data and the MeerKAT radio telescope array in South Africa, stimulated a new radio and gamma-ray study of these two sources. The paper is constructed as follows: In the following section the results of the radio observations of these two sources will be reported. This will be followed by the results of our Fermi-LAT analysis, after which we will present some conclusions and future prospects.

2. MeerKAT Observations and Results of AE Aqr and AR Sco

Both AE Aqr and AR Sco were observed by the SKA precursor telescope MeerKAT on 25 February 2019 (AE Aqr) and 15-16 June 2020 (AR Sco) respectively. The AE Aqr observation was performed under a ThunderKAT project (e.g.9) while AR Sco was observed under the Director's Discretionary Time (DDT) proposal (DDT-20200615-DB-01).

The L-band (0.9-1.7 GHz) light curves of these observations (see Fig. 2a,b) are presented below. The gaps in the light curves of AE Aqr represent periods when calibration sources were observed. For AR Sco only the last short light curve has been included for illustration purposes since these results are the main focus of another forthcoming publication, i.e Buckley et al. 2021 (in preparation). Both L-band light curves display a significant level of variability from both these sources. Since both these sources are not driven by the direct accretion of mass onto the surface of the white dwarf the variability in the radio band is most probably the result of the interaction between the white dwarf magnetosphere and the propeller ejected matter in the case of AE Aqr, or the secondary star's magnetosphere in the case of AR Sco.

A search for periodicity was conducted for both these sources (see Fig. 3a,b). Here we present the first detection of pulsed emission in the radio band from AE Aqr (see Fig. 3a). Similarly, Fig. 3b shows significant pulsed emission from AR Sco, probably on the beat period between the spin and orbital periods (see 5).

The L-Band spectrum of both these sources reveal a power-law, which may indicate non-thermal synchrotron emission. The exact mechanism driving these pulsations are currently under investigation. The radio flux however displays a non-thermal origin with clear power-law spectral profile in the L-band (see Fig. 4a,b), which may signify a non-thermal synchrotron emission process.

The average L-band flux is also presented on the SED for both AE qr and AR Sco (see Fig. 5a,b).

3. Fermi-LAT gamma-ray emission form AE Aqr and AR Sco

A search was conducted for steady and pulsed gamma-ray emission from both AE Aqr and AR Sco using the upgraded Fermi-LAT Pass 8 dataset. The Pass 8 dataset

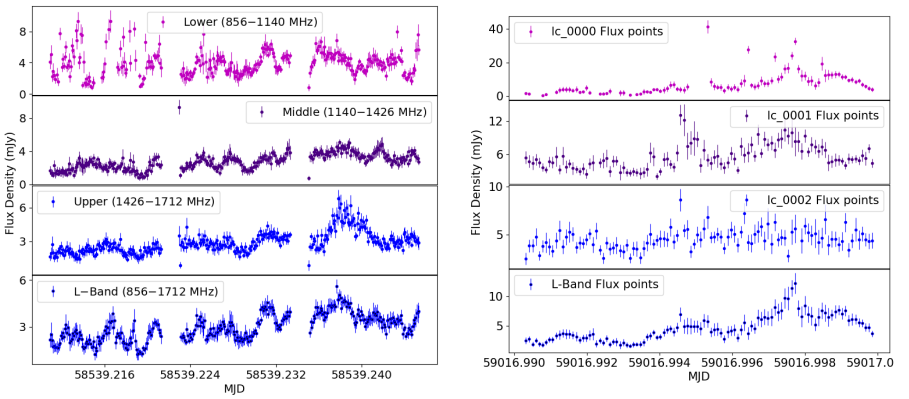


Fig. 2. (a) MeerKAT L-band observation of AE Aqr on 25 February 2019. The light curve was broken up into the lower, middle and upper L-band, with the last light curve representing the total L-Band light curve. The data was binned in 8 second intervals. (b) L-band radio observation of AR Sco performed on 16-17 June 2020. The light curve was broken up in lower, middle and upper L-band, with the last light curve representing the total L-band. The data was binned in 8 second intervals.

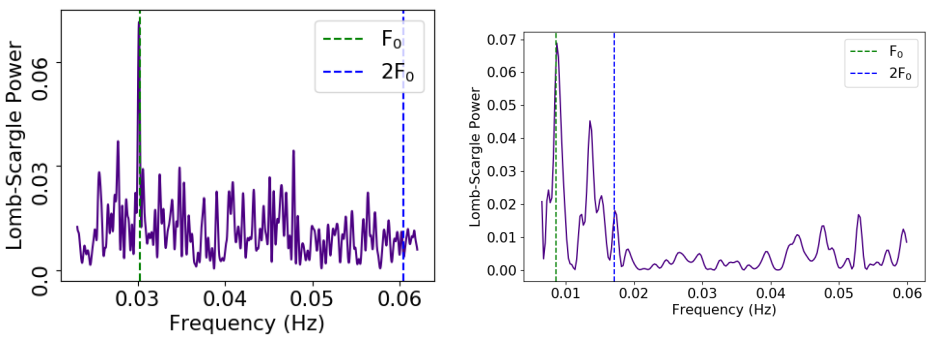


Fig. 3. (a) MeerKAT L-band power spectrum of AE Aqr on 25 February 2019. The power spectrum clearly shows pulsed emission at the spin period of the white dwarf $\nu_\circ = 30.23\,\mathrm{mHz}$. This is the first report of periodic emission from the white dwarf in AE Aqr in radio. (b) MeerKAT total L-band power spectrum of the short AR Sco observation displayed in Fig. 2b. The power spectrum shows strong pulsed emission close to a beat period between the white dwarf spin period and the orbital period.

is an upgraded version of the earlier versions and allowed for better background estimate, gamma-ray source location and energy determination to name a few.

By utilizing the full 10 year Pass 8 dataset with standard Fermi-tools revealed no steady emission from both these sources. However, when a 10 year light curve was determined for AE Aqr there were sections in the light curve that revealed gamma-ray emission above the 2 sigma significance level. Focusing on these data sections, braking it up in 10 minute sections, each section was searched for period-

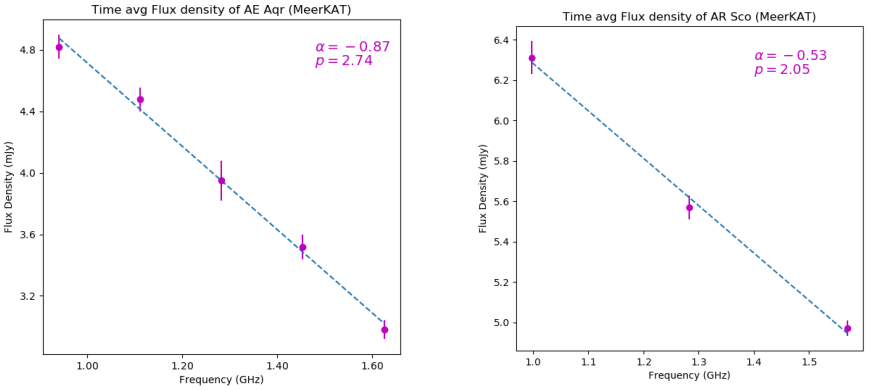


Fig. 4. (a) MeerKAT L-band radio power-law spectrum of AE Aqr. (b) MeerKAT-L-band radio power-law spectrum of AR Sco .

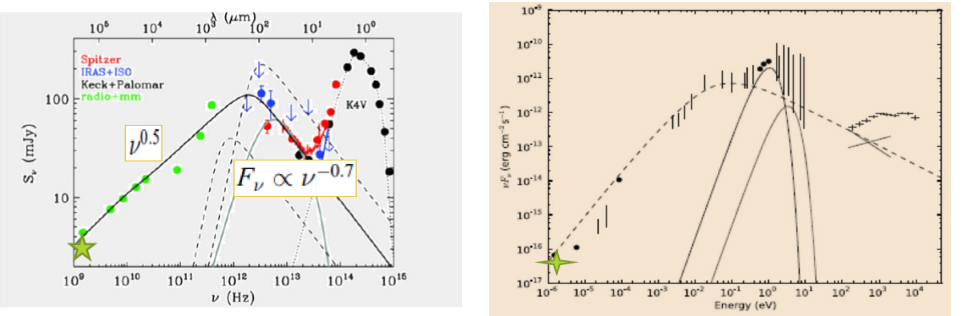


Fig. 5. (a) SED of AE Aqr showing the average MeerKAT L-band flux density (green star) on the SED. Adapted form (10). (b) SED of AR Sco showing the average L-band radio flux density (green cross). Adapted form (11).

icity. The periodograms of these sections all had the same frequency resolution and were then added (stacked) incoherently. These stacked power spectra reveal clear indications of pulsations at both 33.08 s and 16.54 s periods (see Fig. 6a for a few examples). To test whether these pulsations are not spurious we selected regions in the sky consecutively further away from AE Aqr's position in the sky $3^\circ - 15^\circ$ and produced the same power spectra for the same periods. The pulsed power distribution of these were determined and the results displayed in Fig. 6b. Clearly visible is that the power at the 33.08 s and 16.54 s periods deviates from the theoretical white noise distribution for the power spectra produced for data corresponding to AE Aqr's position in the sky, while the distribution of power conforms to a white noise distribution for data sections further away from the source. Fermi-LAT data sections that coincide with optical flares observed with the UFS-Boyden 1.5 m telescope (see Fig. 7a) were selected for periodic analysis. Individual power spectra

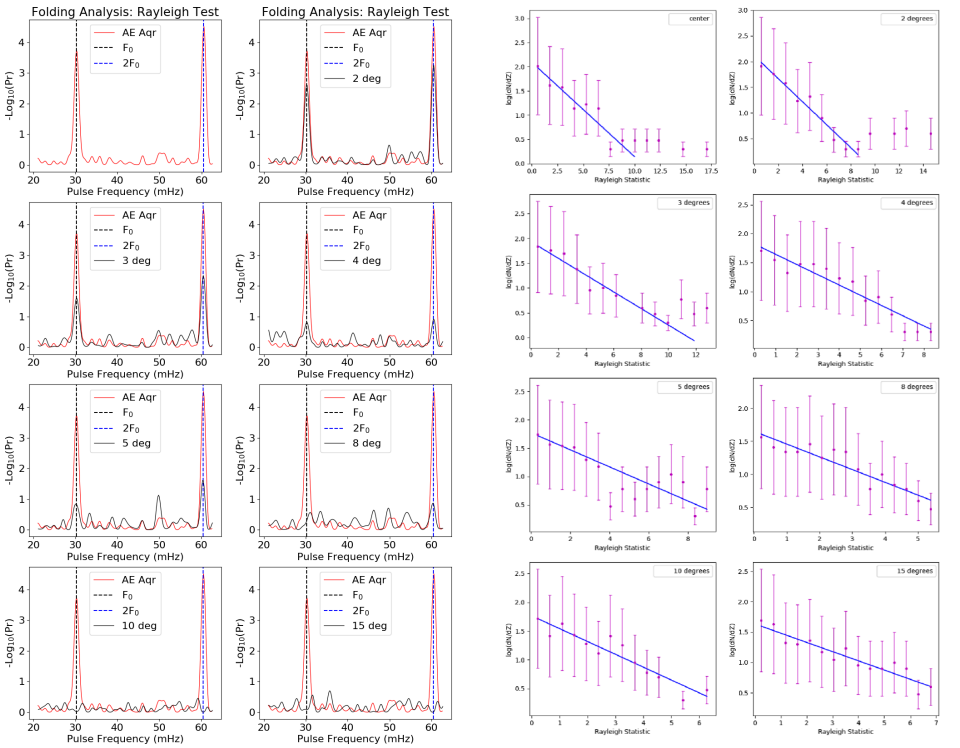


Fig. 6. (a) Example of stacked Fermi power spectra of AE Aqr of epochs showing gamma-ray emission above the 2 sigma significance level. (b) Control analysis showing how the power in the power spectrum return to white noise further away from the source-illustrating that the pulsed power is confined to AE Aqr's position in the sky.

corresponding to flares have been established with the same resolution to enable incoherent stacking (see Fig. 7b). It is clear from Fig. 7b that the gamma-ray periodograms that coincide with enhanced optical activity show clear pulsed emission at a period that corresponds to the spin period of the white dwarf. This result confirms the previously reported pulsed behaviour of VHE emission during optical flares (1, 2).

The unbinned and binned gamma-ray spectra associated with data sections that display a significance above 1σ were produced (see Fig. 8a). Both spectra display a clear power-law profile with photon index $\Gamma \sim 2$. Also shown in Fig. 8a is that the gamma-ray emission seems to be more significant in the lower energy bins. The spectra associated with data sections showing emission above 1 σ and 2 σ respectively reveal power-law spectra that clearly indicate that AE Aqr may be of interest for more sensitive Air Cherenkov detectors like CTA. It is obvious that when the data sections are selected which reveal gamma-ray emission above the 2σ

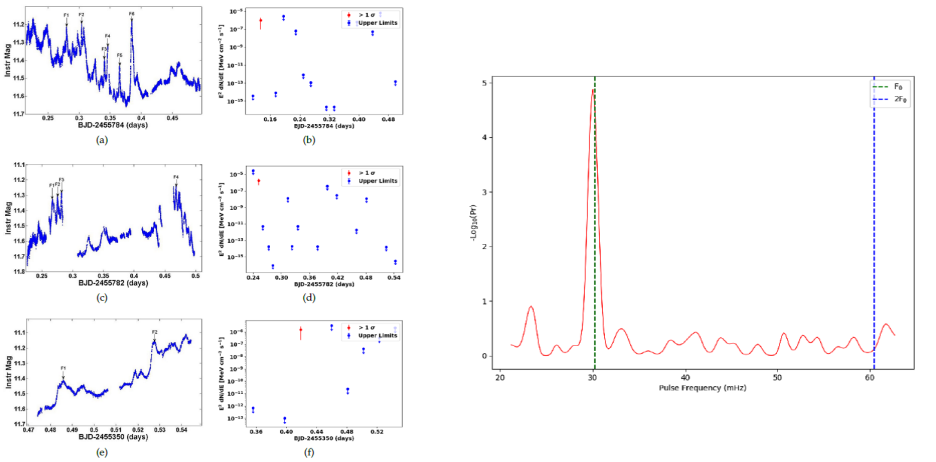


Fig. 7. (a) Optical photometry performed using the UFS-Boyden 1.5 m telescope outside Bloemfontein (South Africa). The gamma-ray data that coincided with optical flares are indicated on the RHS of Fig. 8a. (b) The stacked Fermi-LAT power spectrum of data sections overlapping with optical flares showing clear pulsed emission at or close to the 33.08 s spin period of the white dwarf.

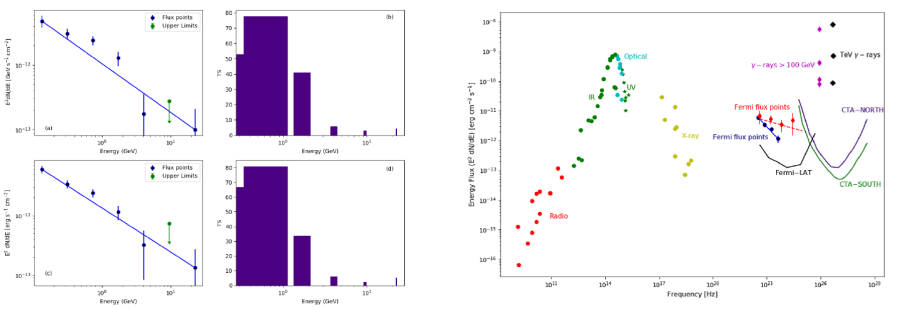


Fig. 8. (a) Gamma-ray spectra for both binned and unbinned analysis, Clearly showing is that the lower energy bins contribute most of the flux. (b) The gamma-ray spectra for quiescence and flare sections of the Fermi-LAT data. It can be seen that the flare sections of the data display spectral hardening which may indicate that AE Aqr may be a source that is of interest for the future Cherenkov Telescope Array (CTA).

level, spectral hardening is observed (see Fig. 8b (red spectrum)), which may result in detectable flux levels during periods of enhanced activity.

The total 10 year Fermi-LAT data of AR Sco was searched for steady gamma-ray emission using the standard Fermi-LAT science tools. No significant steady gamma-ray emission could be distinguished above the 2σ Fermi-LAT threshold (see Fig. 9a). A search for pulsed emission was conducted using the full 10 year Fermi-LAT data. The power spectrum (Fig. 9b) reveals pulsed emission at or close to the spin period of the white dwarf. The period range has been limited to a narrow region around

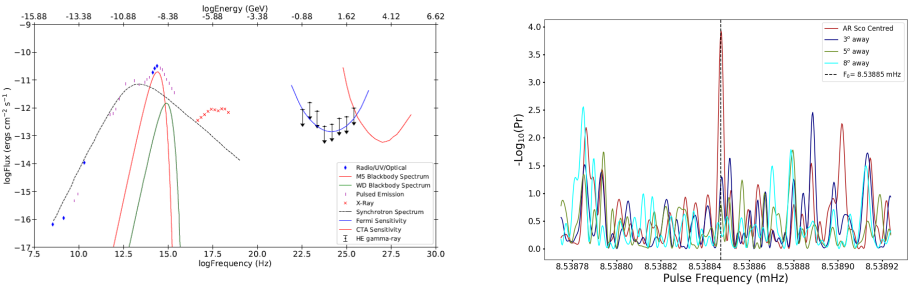


Fig. 9. (a) Gamma-ray spectrum combined with the multi-wavelength SED for AR Sco showing 2 sigma Fermi-LAT upperlimits. (b) The 10 year Fermi-LAT gamma-ray power spectrum revealing pulsations close to the spin period of the white dwarf. Control power spectra at several spatial separations from $3^\circ - 8^\circ$ away from the location of AR Sco does not reveal similar pulsed emission.

the spin period of the white dwarf. To test the validity of these results regions in the sky further away from the position of AR Sco have been selected and the same periodic analysis has been conducted on those sections. These control power spectra correspond to regions in the sky separated from the position of AR Sco between 3° - 8° are also included in Fig. 9b and show no significant pulsed emission at or close to the spin period of the white dwarf. This may support the notion that low-level pulsed gamma-ray activity may in fact be associated with AR Sco.

4. Discussion

The results from the MeerKAT observations and our analyses of the Fermi-LAT data reveal that both AE Aqr and AR Sco contain a particle accelerator of some sort. For example, the MeerKAT data reveal pulsed emission from AE Aqr at the spin period of the white dwarf. This is the first report of pulsed radio emission from AE Aqr and it makes the white dwarf in this system probably one of the fastest white dwarf radio pulsars. Similarly AR Sco is showing pulsed radio emission at the beat period between the spin and orbital period. This may be the result of the white dwarf's magnetosphere pumping the extended magnetosphere of the secondary star, resulting in particle acceleration and non-thermal synchrotron emission. The Fermi-LAT analysis of both these sources reveal pulsed emission. AE Aqr reveals pulsed emission in gamma-rays that resembles reported emission during flare-like events by two independent groups. A period search on data sections displaying flux above the 2 σ level in both binned and unbinned analyses show clear indications of pulsed emission at or close to the white dwarf spin period and its associated first harmonic. A full 10 year periodic power spectrum of AR Sco reveals pulsed power at or close to the spin period of the white dwarf. For both these sources a control analysis has been performed that validates the authenticity of our detection.

Both these sources are not powered by mass accretion and the majority of the emission is in some or other way related to the spin-down power of the white dwarf.

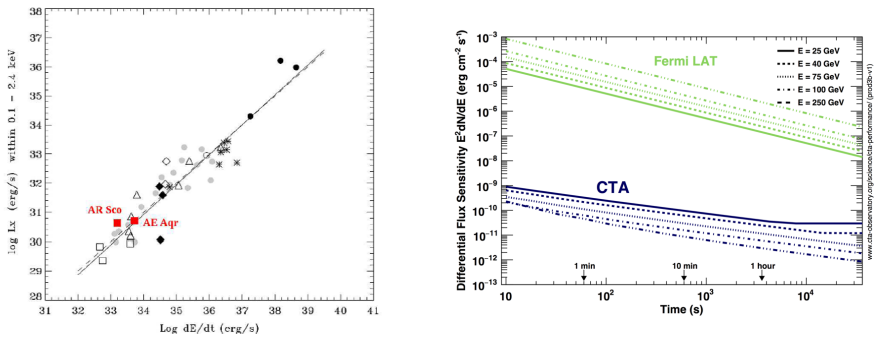


Fig. 10. (a) The X-ray power vs spin-down power of both AE Aqr and AR Sco showing the familiar ratio of X-ray luminosity to spin-down power characteristic of spin-powered neutron stars. Adapted from (12). (b) The most recent 2019 detection threshold comparison between Fermi-LAT and CTA for various observing periods. (**https://www.cta-observatory.org/the-low-end-why-cta-interested-low-energy-gamma-rays**).

Both these sources conform to the well known relation found in rotation powered neutron stars (Fig. 10a), namely that $L_X \sim 10^{-3} L_{s-d}$ (12), meaning that the X-ray emission is not driven by the accretion of mass onto the surface but is driven through a plethora of processes by the spin-down power reservoir. This places both these sources in a unique class. The results that we have presented also show that both these sources provide an opportunity for detection with a new generation of sophisticated ground-based Cherenkov observatories like the Cherenkov Telescope Array (CTA). For example, Fig. 10b shows the latest comparison between the low-energy sensitivity of Fermi-LAT and CTA, which differs slightly form the earlier comparison (13). It is clear that the CTA array will reach the relevant flux levels at energies $\epsilon_\gamma \leq 50\,\mathrm{GeV}$ after observing times of $t_{obs} \sim$ few $\times\, 10^4\,\mathrm{s}$. However, it may be possible to extract a periodic signal after stacking power spectra of shorter observations with similar resolution. The confirmed detection of pulsed gamma-ray emission from both these sources may in fact establish fast rotating highly magnetized white dwarfs in close binaries as a new class of gamma-ray source.

Acknowledgements

The authors thank the organisers of the conference for the invitation to present these results. The authors also acknowledge the Fermi-LAT and MeerKAT collaborations for the gamma-ray and radio data being analysed and presented in this study.

References

1. P.J. Meintjes, et al. 1992, AE Aquarii: An emitter of pulsed gamma-ray emission resembling optical emission during flares, *Astrophysical Journal*, **401**, 325
2. P.J. Meintjes et al. 1994, Simultaneous optical and TeV Gamma-Ray observations of the cataclysmic variable AE Aquarii, *Astrophysical Journal*, **434**, 292

3. C.C.G. Bowden et al. 1992, Sporadic pulsed 350 GeV gamma-ray emission detected from AE Aqr. *Astroparticle Physics* **1**, 47
4. J. Aleksic et al. 2014, MAGIC search for pulsed VHE gamma-ray emission from AE Aquarii in a multiwavelelength context, *Astronomy and Astrophysics*, **568** A109
5. T.R. Marsh et al. 2016, A radio puling white dwarf binary star, *Nature*, **537**, 374
6. D.A.H. Buckley et al. 2017, Polarimetric evidence of a white dwarf pulsar in the binary system AR Scorpii, *Nature Astronomy*, **1**, 0029
7. K.K. Singh, P.J. Meintjes, Q. Kaplan, F.A. Ramamonjisoa & S. Sahayanathan 2020, Modelling the broadband emission from the white dwarf binary system AR Scorpii, *Astroparticle Physics*, 123, 102488
8. K.K. Singh, P.J. Meintjes, K.K. Yadav 2021, Properties of white dwarf in the binary system AR Scorpii and its observed features, *Modern Physics, Lett. A*, 36, 2150096
9. R. Fender, P.A. Woudt et al. 2017, ThundeKAT: The MeerKAT large survey project for image-plane radio transients, In Proceedings of "MeerKAT Science: On the Pathway to the SKA, 25-27 May, 2016, Stellenbosch, South Africa", *Proceedings of Science*, (PoS(MeerKAT2016(013))
10. G, Dubus et al. 2007, Spitzer space telescope observations of the magnetic cataclysmic variable AE Aquarii, *Astrophysical Journal*, **663**, 516
11. J. Takata et al. 2018, A nonthermal pulsed x-ray emission of AR Scorpii, *Astrophysical Journal*, **853**, 106
12. W. Becker and J. Trumpher 1997, The x-ray luminosity of rotation powered neutron stars, *Astronomy and Astrophysics*, **326**, 682
13. S. Funk, J.A. Hinton and the CTA Consortium 2013, Comparison of Fermi-LAT and CTA in the region between 10 to 100 GeV, *Astroparticle Physics*, **43**, 348-355

Study the effects of anisotropy on the highly magnetized white dwarfs

Debabrata Deb
Department of Physics, Indian Institute of Science, Bangalore 560012, India
E-mail: debabratadeb@iisc.ac.in

Banibrata Mukhopadhyay
Department of Physics, Indian Institute of Science, Bangalore 560012, India
E-mail: bm@iisc.ac.in

Fridolin Weber
Department of Physics, San Diego State University, San Diego, CA 92182, USA
Center for Astrophysics and Space Sciences, University of California at San Diego, La Jolla, CA 92093, USA
E-mail: fweber@sdsu.edu

The equilibrium configuration of white dwarfs composed of anisotropic fluid distribution in the presence of a strong magnetic field is investigated in this work. By considering a functional form of the anisotropic stress and magnetic field profile, some physical properties of magnetized white dwarfs, such as mass, radius, density, radial and tangential pressures, are derived; their dependency on the anisotropy and central magnetic field is also explored. We show that the orientations of the magnetic field along the radial direction or orthogonal to the radial direction influence the stellar structure and physical properties of white dwarfs significantly. Importantly, we show that ignoring anisotropy governed by the fluid due to its high density in the presence of a strong magnetic field would destabilize the star. Through this work, we can explain the highly massive progenitor for peculiar over-luminous type Ia supernovae, and low massive progenitor for under-luminous type Ia supernovae, which poses a question of considering 1.4 solar mass white dwarf to be related to the standard candle.

Keywords: Magnetic fields; white dwarfs; super-Chandrasekhar white dwarf; massive stars

1. Introduction

In a recent article, we have shown that the presence of strong magnetic field, orientations of magnetic field and the combined anisotropy due to both the matter and magnetic field exhibit notable influence on the microscopic and thermodynamic properties of compact stars, viz., neutron stars and strange quark stars. We have also very briefly discussed in their study (hereinafter Paper I) that same effects can be observed for magnetized and anisotropic white dwarfs (WDs) which further may leads to possible violation of Chandrasekhar critical mass limit. Importantly Paper I shows that to achieve hydrostatic stable configuration for the highly magnetized compact stars, it is necessary to consider local pressure anisotropy due to matter besides the anisotropy due to magnetic field. Following Paper I, in the present

work we assume two type of magnetic field orientations, such as *radial orientation (RO)* when the magnetic field is directed toward the radial direction and *transverse orientation (TO)* when field is oriented perpendicular to the radial direction (say along θ or ϕ directions). In this work, we investigate in detail the effects of magnetic field strength, their orientations and anisotropy on the highly magnetized WDs and their thermodynamic parameters.

WDs achieve stable stellar configuration via equilibrium of the forces when the inward gravitational pull is counterbalanced by the outward repulsive electron degeneracy pressure. During the accretion of mass from the companion stars when WDs surpass their critical mass limit, immediately a violent explosion triggers, which completely disrupt WDs and releases huge amount of energy of the order $\sim 10^{51}$ erg, known as type Ia supernova (SNIa). The critical mass limit was introduced by Chandrasekhar in his famous works[2,3] where he predicted that the critical mass limit for non-magnetized and non-rotating carbon-oxygen (C-O) WD is 1.4 $M_{\odot}$ which is known as "Chandrasekhar mass limit" (CML) for WDs.

The standard and specific set of relations among colour, intrinsic luminosity and light-curve width[4,5] make SNIa a perfect candidate for "standard candle" which helps astronomers to measure accurately the distance of the host galaxy. Behind the measurement of cosmological parameters, the disproportional relation between distance and luminosity of SNIa is used as a key technique. In fact, SNIa as a "standard candle" had pivotal role to confirm the accelerated expansion of universe.[6,7] However, the idea of "standard candle" faces a major setback with the recent series of observation of several peculiar over-luminous type Ia supernovae (SNeIa) such as *SN 2003cv*,[8] *SN2007fg*,[9] *SN2009if*[10] and *SN 2013dc*,[11] which are best explained through WDs as a progenitor having mass beyond CML. Howell and collaborators[8] predicted that to explain *SN 2003cv* it requires to power by the radioactive decay of $\sim 1.3M_{\odot}$ of ^{56}Ni which further predicts $\sim 2.1M_{\odot}$ WD as a progenitor, known as *super-Chandrasekhar progenitor WD* (SCPWD). Later researchers have predicted further massive SCPWD against the overluminous SN 2009if of mass as high as $\sim 2.8M_{\odot}$.

To explain these unexpected overluminous SNeIa, researchers suggested two possible conjectures such as (i) explosion of rapidly rotating WDs[8,12,13] and (ii) merger of two massive WDs.[14,15] Chen and Li[22] showed that the differentially rotating and accreting WD can not gain mass more than $1.7\,M_{\odot}$. On the other hand, further investigations showed that through the double degenerate scenario, it is not possible to explain such a highly massive 2.8 $M_{\odot}$ SCPWD.[16–21] The results from numerical simulation of massive WD merger suggest that due to off-center carbon burning accreting WD collapses into neutron star.[23,24] Hence, it is evident through the above discussion that the double degenerate scenario conjectures never leads to a SNIa which can explain the massive SCPWDs.

Introducing magnetic model, Mukhopadhyay and collaborators[25,26] are able to explain SCPWDs where they considered Landau levels due to high magnetic field in a plane perpendicular to the magnetic field axis and they[27] are also able

to suitably explain $\sim$ 2.6 $M_{\odot}$ highly massive SCPWD. In another study, Das and Mukhopadhyay[28] showed that the average magnetic field and related magnetic pressure are significantly low compared to the matter pressure when inside magnetized WDs (hereinafter B-WDs) the strong magnetic field is tangled in a length scale larger than quantum length scale, which however leads to the Landau quantization to matter. The mass-radius relation for B-WDs deviates significantly beyond CML when mass accretion continues as the effective pressure counterbalances the inward gravitational pull.[25,29] Das and Mukhopadhyay[30] further predicated that it requires a time scale of 2×10^7 years to turn into a super-Chandrasekhar B-WD (SCBWD) for a 0.2 $M_{\odot}$ accreting WD having surface magnetic field (B_s) 10^9 G. In fact for the last several years, Mukhopadhyay and his group have explored various aspects of SCBWDs through different sophisticated models by considering variation of magnetic field and geometry, Landau quantization, a pure general relativistic effect, differential rotation, deviation from the spherical symmetric stellar configuration and thermal luminosity, etc.[31–38] They also explored the techniques of the possible direct detection of SCBWDs via continuous gravitational wave astronomy by the various upcoming detectors, such as BBO, DECIGO and LISA. The idea of SCBWDs has been further studied and supported by different independent researcher works.[41–47]

Bowers and Liang[48] raised a strong argument against the oversimplified assumption of the isotropic perfect fluid as the constituent matter distribution of compact stars and proposed a modified Tolman-Oppenheimer-Volkoff (TOV) equation,[49,50] which includes another force raised due to pressure anisotropy, besides the hydrodynamic and gravitational force. They also explored notable effect of local anisotropy due to matter on the different physical properties of the stellar objects, such as density, surface redshift, total mass and radius. Letelier[51] and Bayin[52] pointed out that the possible reason behind this local anisotropy is the presence and mixture of two (or more) fluid within stellar interior. On the other hand, the breaking of spatial rotational symmetry $\mathcal{O}(3)$ due to spontaneous creation of magnetic field instigates pressure anisotropy in the stellar interior.[53–55] Interestingly, during the cooling process of WDs at low temperature, they undergoes a first-order phase transition,[56–58] which is another possible reason for the presence of anisotropy inside the dense plasma. Stevenson[59] introduced a phase diagram through which they predicted that carbon and oxygen remains immiscible when WDs begin to freeze. Later, Garcia and collaborators[60] further explored and confirmed this idea. Nag and Chakrabarty[61] also supported the idea of first-order phase transition during the cooling of WDs through the Bose condensation. In fact, they proposed that "*crystalline normal crustal matter*" is present at the massive WDs core, which is also Bose condensed and similar as the ultra-dense structure of the neutron stars. So, based on the above discussion we are convinced to consider local anisotropy of matter inside WDs.

In this context, it is worth mentioning that though there are several articles[32,33,39,40,62–71] which study non-spherical magnetized compact stars in general

relativity based on the perturbative approach to solve the the Einstein-Maxwell field equations or the numerical techniques (such as Lorene and XNS codes) by considering poloidal or toroidal or mixed field geometries, but they did not include the effects of anisotropy. Although, a recent work[72] has explored the effect of anisotropy on WDs in scalar-tensor (ST) theories of gravity but it did not consider magnetic field. The idea of magnetic field orientations viz. RO and TO was originally coined by Chu et al.[73] In the present study, we explore the combined effects of magnetic field strength, its orientations and anisotropy on the WDs to have a more generalized perspective of the WD stellar interior.

This paper is organized as follows: We discuss the basic mathematical formalism in Section 2. Our results and combined effects of the magnetic field, its orientations and anisotropy on the physical properties of WDs are discussed in Section 3. Finally, we summarise our work and present some concluding remarks in Section 4.

2. Basic mathematical formulation

Mukhopadhyay and group[32,33,39] have shown that the magnetized compact stars which are toroidally dominated approximately retains their spherical symmetric stellar structure. Hence in this work we describe B-WDs by the spherically symmetric spacetime metric given by

$$ds^2 = e^{\nu(r)}dt^2 - e^{\lambda(r)}dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{1}$$

where we assume that the metric potentials ν and λ are functions of radial coordinate r. We choose $c = G = 1$, where G is Newton's gravitational constant and c is the speed of light.

Now neglecting the effect of the electric field by assuming that inside the B-WDs there are no macroscopic charges, the energy-momentum tensor, which describes anisotropic and magnetized matter distribution, reads

$$\begin{aligned} T_m^{\mu\nu} &= (\rho + p_t)u^\mu u^\nu - p_t g^{\mu\nu} + (p_r - p_t)\, v^\mu v^\nu \\ &\quad + \mathcal{M}B\left(g^{\mu\nu} - u^\mu u^\nu + \frac{B^\mu B^\nu}{B^2}\right) + \frac{B^2}{4\pi}\left(u^\mu u^\nu - \frac{1}{2}g^{\mu\nu}\right) - \frac{B^\mu B^\nu}{4\pi}, \end{aligned} \tag{2}$$

where $u^\mu = \delta^\mu_0 e^{-\nu(r)/2}$ is a timelike 4-velocity unit vector of fluid and $v^\mu = \delta^\mu_1 e^{-\lambda(r)/2}$ is a spacelike unit vector acting along the radial direction. It is noteworthy that these orthogonal vectors satisfy $u^\mu u_\mu = -v^\mu v_\mu = 1$ and $u^\mu v_\mu = 0$. Here, ρ, p_r and p_t are the density, radial pressure and tangential pressure of the matter inside B-WDs. $g_{\mu\nu}$ denotes metric tensor and $\mathcal{M}$ represents magnetization per unit volume. B is the magnetic field strength given by $B^\mu B_\mu = -B^2$. On the other hand, we do not consider the associated magnetization effect since magnetic pressure is one order higher in magnitude compared to the magnetization which has already been predicted by a few articles.[53,74] So in this work the presented numerical results do not include effect of magnetization.

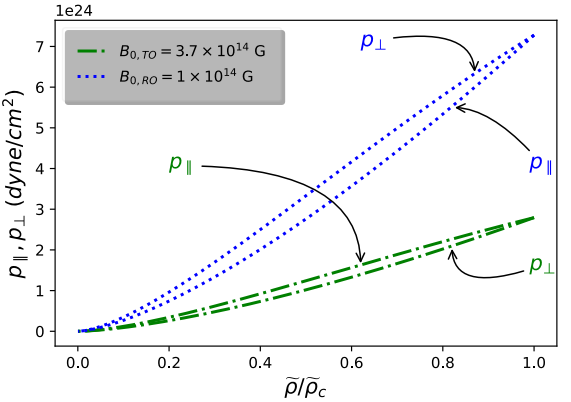


Fig. 1. Variation of parallel pressure ($p_\parallel$) and transverse pressure ($p_\perp$) with the system density ($\widetilde{\rho}$) normalized by the central system density ($\widetilde{\rho}_c$) for a 1.3 $M_\odot$ magnetized WD candidate.

Now using Eqs. (1) and (2) the conservation of energy-momentum tensor, i.e., $\nabla_\mu T^{\mu\nu} = 0$, leads to the essential stellar structure equations which describe static and anisotropic B-WDs as follows:

$$\frac{\mathrm{d}m}{\mathrm{d}r} = 4\pi\widetilde{\rho}r^2, \tag{3}$$

$$\begin{cases} \dfrac{\mathrm{d}p_\parallel}{\mathrm{d}r} = -\left(\widetilde{\rho} + p_\parallel\right)\dfrac{4\pi r^3 p_\parallel + m}{r\left(r-2m\right)} + \dfrac{2}{r}\Delta, & \text{for RO,} \\ \dfrac{\mathrm{d}p_\perp}{\mathrm{d}r} = -\left(\widetilde{\rho} + p_\perp\right)\dfrac{4\pi r^3 p_\perp + m}{r\left(r-2m\right)} + \dfrac{2}{r}\Delta, & \text{for TO.} \end{cases} \tag{4}$$

where $\widetilde{\rho}$ is the system density defined as $\widetilde{\rho} = \rho + \frac{B^2}{8\pi}$, $p_\parallel$ and $p_\perp$ are the system pressures, which take the form based on the magnetic field orientations as follows:

$$p_\parallel = \begin{cases} p_r - \dfrac{B^2}{8\pi}, & \text{for RO,} \\ p_t - \dfrac{B^2}{8\pi}, & \text{for TO.} \end{cases} \tag{5}$$

and (6)

$$p_\perp = \begin{cases} p_t + \dfrac{B^2}{8\pi}, & \text{for RO,} \\ p_r + \dfrac{B^2}{8\pi}, & \text{for TO.} \end{cases} \tag{7}$$

In Eq. (4), Δ denotes effective anisotropy of the stars and based on the choice of magnetic field orientations it takes the forms as $p_t - p_r + \frac{B^2}{4\pi}$ and $p_t - p_r - \frac{B^2}{8\pi}$ due to RO and TO, respectively. One may get the standard form of TOV equations[48, 75] for non-magnetized and anisotropic starts for $B = 0$.

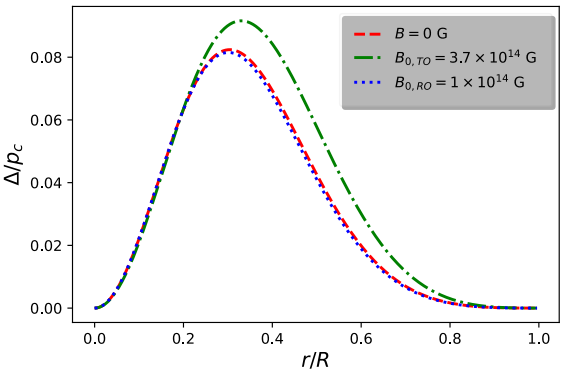


Fig. 2. Variation of anisotropy (Δ) normalized by matter central pressure (p_c) with the radial coordinate r/R for a 1.3 $M_\odot$ magnetized WD candidate.

Bowers and Liang proposed a generalized functional form for the anisotropic stress which has been widely used in different studies. However, since for the present work we require a functional form for anisotropy which includes both the effects of magnetic field strength and magnetic field orientation, we have suitably modified the Bowers-Liang anisotropic form which reads

$$\Delta = \begin{cases} \kappa \dfrac{(\widetilde{\rho} + p_{\parallel})\,(\widetilde{\rho} + 3\,p_{\parallel})}{(r-2m)} r^3, & \text{for RO,} \\ \kappa \dfrac{(\widetilde{\rho} + p_{\perp})\,(\widetilde{\rho} + 3\,p_{\perp})}{(r-2m)} r^3, & \text{for TO,} \end{cases} \tag{8}$$

where κ is a constant parameter that controls the amount of anisotropy within the star and we use κ maintaining its permissible range, given by $\left[-\frac{2}{3}, \frac{2}{3}\right]$.[76] Now, one may notice that if matter anisotropy is not considered, the anisotropic stress become $\sim \mid B^2 \mid$, which has a finite, non-zero value at the stellar center. Since at the center, the hydrodynamic and gravitational forces are zero, non-zero anisotropic force at the center should trigger instability due to nonequilibrium of the forces. We show this issue can easily be overcome considering combined anisotropy due to both the matter and magnetic field.

Further, to describe matter distribution of WDs we consider equation of state (EoS) proposed by Chandrasekhar which reads

$$\begin{aligned} p_r &= \frac{\pi m_e^4 c^5}{3h^3}\left[x\left(2x^2-3\right)\sqrt{x^2+1} + 3\mathrm{sinh}^{-1}x\right], \\ \rho &= \frac{8\pi \mu_e m_H \left(m_e c\right)^3}{3h^3} x^3, \end{aligned} \tag{9}$$

where m_H and m_e denote mass of hydrogen atom and electron, respectively, $\mu_e = 2$ is the mean molecular weight., p_F is the Fermi momentum and $x = p_F/m_e c$. Note that for clarity here we introduce the speed of light c and Planck constant h in their

appropriate places. For the present case of C-O WDs we have $\mu_e = 2$. We consider the exterior spacetime is defined by the well-known Schwarzschild exterior spacetime, given by

$$ds^2 = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{1}{\left(1 - \frac{2M}{r}\right)} dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{10}$$

where M represents total mass of the star.

Both the fossil field of the progenitor star[31,33,77] and dynamo effect[78] ensure stronger magnetic field at the center (B_c), which decreases gradually to reach its minimum value at the surface (B_s). Hence, in this work we consider a density dependent functional form for the magnetic field,[31,36,79,80] which can suitably mimic the spatial dependence of magnetic field strength, given by

$$B(\rho) = B_s + B_0 \left[1 - \exp\left\{-\eta \left(\frac{\rho}{\rho_0}\right)^{\gamma}\right\}\right], \tag{11}$$

where the η and γ are the dimensionless constants which controls how fast magnetic field strength decays from the center to the surface. In this work we chose $\rho_0 = 10^9\,\mathrm{g/cm^3}$, $\gamma = 0.9$ and $\eta = 0.2$.[36] We also consider $B_s = 10^9$ G which is consistent with the Sloan Digital Sky Survey.[81]

3. Result and discussion

In this study, we explore that besides magnetic field strength, the magnetic field orientations (such as RO and TO) and combined anisotropy due to matter and field significantly influence the mass-radius curve and physical parameters of B-WDs. Note that to present some basic properties of thermodynamics and anisotropy we consider a $1.3\,M_\odot$ B-WD candidate having surface magnetic field 10^9 G in Figs. 1 and 2. Further, our choices of B_0 for RO and TO are different as they exhibit an asymmetric effect on the stellar mass. In the figures, we put RO and TO in the subscript of B_0 to specify the respective magnetic field orientations. It is also noteworthy that our choice of B_0 is also motivated by the highly massive SCPWDs of peculiar over-luminous SNeIa.

In Fig. 1 we feature combined effects of magnetic field strength and its orientations on the system parallel pressure $p_\parallel$ and transverse pressure $p_\perp$. We find that for TO, as B_0 increases, the slope of the system pressures decreases gradually, whereas, for RO with the increasing B_0, the system pressures become gradually stiffened. This phenomenon is because, as B_0 increases in the TO case, the stellar radius increases, which reduces gravitational force and consequently decreases the hydrodynamic force. However, for increasing B_0 in the RO case, the stellar radius of B-WDs decreases, consequently increasing outward hydrodynamic force to support the increased gravitation force. Importantly, note that at the center the effective anisotropy is zero since $p_\parallel = p_\perp$ at $r = 0$, which reflects consistency of TOV equations. In Fig. 2 we show the profile for system anisotropy Δ which also shows zero anisotropy at the center.

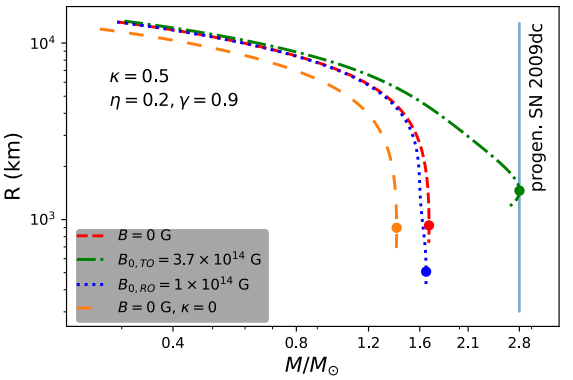


Fig. 3. Variation of the total mass in the units of solar mass ($M/M_\odot$) with the total radius (R) of the stars for varying B_0 and $\kappa = 0.5$. Solid circles represent maximum possible mass for the stars.

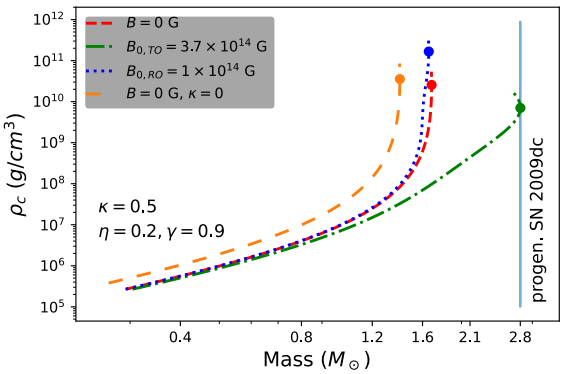


Fig. 4. Variation of $M/M_\odot$ as a function of the matter central density (ρ_c). Solid circles represent maximum possible mass for the stars.

Further, one may note that the maximum magnitude of anisotropy increases with increasing B_0 for the TO case, whereas it decreases for the RO case. We find that for $B_0 = 3.7 \times 10^{14}$ G the maximum magnitude of Δ is $\sim 92\%$ lower compared to the central matter pressure p_c which is insufficient to drive the spherical symmetric stellar structure of B-WDs to non-spherical symmetry.

The total mass and radius (M-R) curve for B-WDs due to different B_0 for both RO and TO is shown in Fig. 3. We find the maximum mass ($M_{\rm max}$) and the corresponding radius ($R_{\rm Mmax}$) in TO due to $B_0 = 3.7 \times 10^{14}$ G are 2.8 $M_\odot$ and 1457.24 km, respectively, whereas in RO due to $B_{0,\rm TO} = 10^{14}$ G they are 1.66 $M_\odot$ and 507.10 km, respectively. In TO due to $B_{0,\rm TO} = 3.7 \times 10^{14}$ G, $M_{\rm max}$ and $R_{\rm Mmax}$ increase compared to their non-magnetized anisotropic case of $\sim$66% and $\sim$57%, respectively, whereas in RO due to $B_{0,\rm RO} = 10^{14}$ G they decrease to $\sim$2% and $\sim$45%, respectively. Only considering local anisotropy due to matter and not

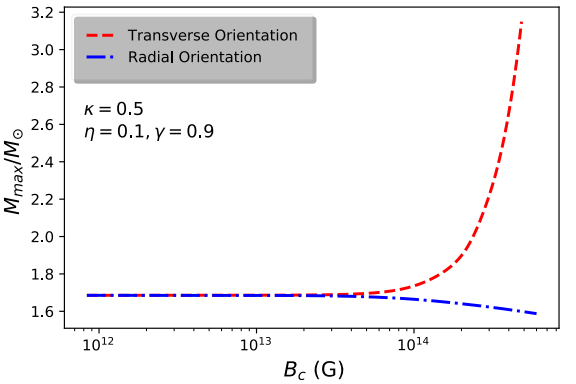


Fig. 5. Variation of $M/M_\odot$ with the central magnetic field (B_c).

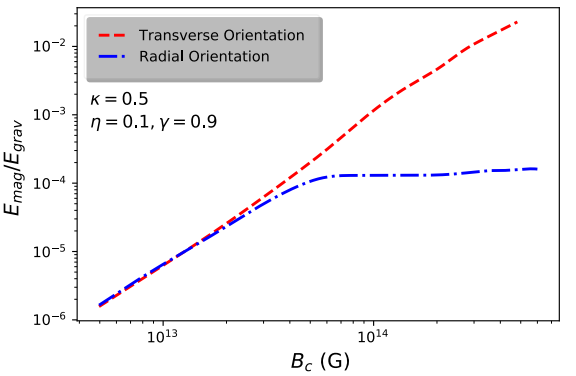


Fig. 6. Variation of the ratio of the magnetic energy ($E_{\rm mag}$) to the gravitational energy ($E_{\rm grav}$) with the central magnetic field (B_c).

considering magnetic field through our model one may have 1.81 $M_\odot$ WD having radius 959.63 km which are ∼29% and ∼7%, respectively, higher compared to their values for WD CML. Harrison et al.[82] predicted that for the stability of a spherically symmetric stellar model the M-R relation should be consistent with $\mathrm{d}M/\mathrm{d}\rho_{\rm c} > 0$ up to $M_{\rm max}$. It has been shown in Fig. 4, which confirms stable B-WDs upto maximum limiting mass.

In Fig. 5 we explore the effects of magnetic field orientations on the mass of B-WDs. The effects of magnetic field strength on $M_{\rm max}$ for the TO and RO cases are asymmetric which motivates us to chose different B_0 for different magnetic field orientations. One may find that for a specific set of constant parameter such as $\eta = 0.1$, $\gamma = 0.9$ and $\kappa = 0.5$ the asymmetry of maximum mass for the TO and RO cases due to $B_c = 4 \times 10^{14}$ G is ∼65% which confirms significant effect of

Table 1. Physical parameters of WDs with $\kappa = 0.5$, $\eta = 0.2$ and $\gamma = 0.9$ for different B_0.

Orientation of magnetic field	B_0 (Gauss)	Maximum mass ($M_\odot$)	Corresponding predicted radius (km)	Central magnetic field B_c (Gauss)	Central density $\tilde{\rho}_c$ (gm/cm^3)	Central pressure $\tilde{p}_c$ ($dyne/cm^2$)	$\frac{E_{\rm mag}}{E_{\rm grav}}$
Transverse Orientation	3.7×10^{14}	2.80	1457.24	2.538×10^{14}	7.040×10^{9}	2.809×10^{28}	1.405×10^{-2}
No magnetic Field	-	1.69	926.91	-	2.560×10^{10}	9.865×10^{28}	-
Radial Orientation	10^{14}	1.66	507.10	10^{14}	1.665×10^{11}	1.152×10^{30}	7.227×10^{-5}

magnetic field orientations on B-WDs. Fig. 5 features that for TO (RO) the increase (decrease) of $M_{\rm max}$ is rapid (slow) with the increase of central magnetic field B_c.

Chandrasekhar and Fermi[83] predicted that for magnetized spherical symmetric stable compact stars it is not sufficient that they are consistent with the inequality $\Gamma > 4/3$ as the strong magnetic field may lead to stellar instability. They further predicted that for the dynamical stability of the magnetized compact stars the necessary condition is $E_{\rm mag}/|E_{\rm grav}| << 1$, where $E_{\rm mag}$ and $|E_{\rm grav}|$ denote the magnetic and gravitational potential energies, respectively. In Fig. 6 we present that for both TO and RO due to different B_0, $|E_{\rm grav}|$ overpowers $E_{\rm mag}$ considerably to achieve dynamical stability of B-WDs.

For the better understanding of the readers, we present the obtained results for the different cases in Table 1. We present the variation of $M_{\rm max}$, R, B_c, system central density ($\tilde{\rho}_c$), system central pressure ($\tilde{p_c}$) and $E_{\rm mag}/\mid E_{\rm grav} \mid$ with different B_0 for both TO and RO and non-magnetized cases in Table 1. Note that the presented results in Table. 1 are based on the chosen set of constants such as $\eta = 0.2$, $\gamma = 0.9$ and $\kappa = 0.5$. Clearly, for TO with the increase of B_0, B-WDs become massive, larger in size and less dense compact object, whereas for RO as B_0 increases B-WDs become gradually less massive, smaller in size and highly dense compact object compared to their non-magnetized case.

4. Conclusion

In the present work we explore the combined effects due to magnetic field strength, orientations of magnetic field and effective anisotropy raised due to matter and field on B-WDs. Although anisotropic WDs were studied by Chowdhury and Sarkar,[72] their study was based on the non-magnetised case. On the other hand, Chu et al.[73] explored the effect of magnetic field and their orientation on total mass of compact stars. However, they ignored important effect of anisotropy and did not show effects of magnetic field orientations on the different physical parameters of compact stars. In this work for the first time we have studied the combined effects of magnetic field strength, their orientations and anisotropy on WDs in the strong magnetic field. In conclusion, the two major results of our study on B-WDs are as follows:

(i) We show that to achieve magneto-hydrostatic stability at the stellar center of B-WDs, the effective anisotropy due to matter and magnetic field should be considered.

(ii) With the choice of appropriate set of constant parameters such as κ and B_0 our model can appropriately explain massive progenitors of peculiar over luminous SNeIa as high as 2.8 $M_\odot$, which further raise question regarding the idea of 1.4 $M_\odot$ WDs as the "standard candle".

Acknowledgments

Research of DD is funded by the C.V. Raman Postdoctoral Fellowship (Reg. No. R(IA)CVR-PDF/2020/222) from the Department of Physics, Indian Institute of Science. BM acknowledges partial support by a project of the Department of Science and Technology (DST), India, with Grant No. DSTO/PPH/BMP/1946 (EMR/2017/001226). All computations were performed in open source softwares, and the authors are sincerely thankful to the open source community.

References

1. D. Deb, B. Mukhopadhyay and F. Weber, *accepted for publication in ApJ, arXiv:2108.12436* (2021); Paper I.
2. S. Chandrasekhar, *Astrophys. J.* **74**, 81 (1931).
3. S. Chandrasekhar, *Mon. Not. R. Astron. Soc.* **95**, 207 (1935).
4. M. M. Phillips, *Astrophys. J. Lett.* **413**, L105 (1993).
5. G. Goldhaber, D. E. Groom, A. Kim, G. Aldering, P. Astier, A. Conley, S. E. Deustua, et al. *Astrophys. J.* **558**, 359 (2001).
6. A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, et al., *Astron. J.* **116**, 1009 (1998).
7. S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, et al., *Astrophys. J.* **517**, 565 (1999).
8. D. A. Howell, M. Sullivan, P. E. Nugent, R. S. Ellis, A. J. Conley, D. Le Borgne, R. G. Carlberg, et al., *Nature* **443**, 308 (2006).
9. R. A. Scalzo, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, et al., *Astrophys. J.* **713**, 1073 (2010).
10. Taubenberger S., Benetti S., Childress M., Pakmor R., Hachinger S., Mazzali P. A., Stanishev V., et al., *Mon. Not. R. Astron. Soc.* **412**, 2735 (2011).
11. Y. Cao, J. Johansson, P. E. Nugent, A. Goobar, J. Nordin, S. R. Kulkarni, S. B. Cenko, et al., Astrophys. J. **823**, 147 (2016)
12. K. Boshkayev, L. Izzo, J. A. Rueda Hernandez and R. Ruffini, *Astron. Astrophys.* **555**, A151 (2013).
13. D. Branch, *Nature* **443**, 283 (2006).
14. I. Iben and A. V. Tutukov, *Astrophys. J., Suppl.* **54**, 335 (1984).
15. M. Hicken, P. M. Garnavich, J. L. Prieto, S. Blondin, D. L. DePoy, R. P. Kirshner and J. Parrent, *Astrophys. J. Lett.* **669**, L17 (2007).
16. K. Nomoto, *Astrophys. J.* **277**, 791 (1984).
17. H. Saio and K. Nomoto, *Astron. Astrophys.* **150**, L21 (1985).
18. K. Nomoto, *Astrophys. J.* **322**, 206 (1987).
19. K. Nomoto and Y. Kondo, *Astrophys. J. Lett.* **367**, L19 (1991).
20. F. X. Timmes, S. E. Woosley and R. E. Taam, *Astrophys. J.* **420**, 348 (1994).
21. H. Saio and K. Nomoto, *Astrophys. J.* **500**, 388 (1998).
22. W.-C. Chen and X.-D. Li, *Astrophys. J.* **702**, 686 (2009).

23. H. Saio and K. Nomoto, *Astrophys. J.* **615**, 444 (2004).
24. R. G. Martin, C. A. Tout and P. Lesaffre, *Mon. Not. R. Astron. Soc.* **373**, 263 (2006).
25. U. Das and B. Mukhopadhyay, *Int. J. Mod. Phys. D* **21**, 1242001 (2012a).
26. A. Kundu and B. Mukhopadhyay, *Mod. Phys. Lett. A* **27**, 1250084 (2012).
27. U. Das and B. Mukhopadhyay, *Phys. Rev. Lett.* **110**, 071102 (2013a).
28. U. Das and B. Mukhopadhyay, *Mod. Phys. Lett. A* **29**, 1450035 (2014a).
29. U. Das and B. Mukhopadhyay, *Phys. Rev. D* **86**, 042001 (2012).
30. U. Das and B. Mukhopadhyay, *A. R. Rao, Astrophys. J. Lett.* **767**, L14 (2013).
31. U. Das and B. Mukhopadhyay, *J. Cosmol. Astropart. Phys.* **2014**, 050 (2014).
32. U. Das and B. Mukhopadhyay, *J. Cosmol. Astropart. Phys.* **2015**, 016 (2015a).
33. S. Subramanian and B. Mukhopadhyay, *Mon. Not. R. Astron. Soc.* **454**, 752 (2015).
34. B. Mukhopadhyay and A. R. Rao, *J. Cosmol. Astropart. Phys.* **2016**, 007 (2016).
35. B. Mukhopadhyay, A. R. Rao and T. S. Bhatia, *Mon. Not. R. Astron. Soc.* **472**, 3564 (2017).
36. M. Bhattacharya and B. Mukhopadhyay, S. Mukerjee, *Mon. Not. R. Astron. Soc.* **477**, 2705 (2018).
37. A. Gupta, B. Mukhopadhyay and C. A. Tout, *Mon. Not. R. Astron. Soc.* **496**, 894 (2020).
38. B. Mukhopadhyay, A. Sarkar and C. A. Tout, *Mon. Not. R. Astron. Soc.* **500**, 763 (2021).
39. S. Kalita and B. Mukhopadhyay, *Mon. Not. R. Astron. Soc.* **490**, 2692 (2019).
40. S. Kalita and B. Mukhopadhyay, *T. Mondal, T. Bulik, Astrophys. J.* **896**, 69 (2020).
41. P. Federbush, T. Luo and J. Smoller, *Arch. Ration. Mech. Anal.* **215**, 611 (2015).
42. B. Franzon and S. Schramm, *Phys. Rev. D* **92**, 083006 (2015).
43. B. Franzon and S. Schramm, *Mon. Not. R. Astron. Soc.* **467**, 4484 (2017).
44. M. Moussa, *Ann. Phys.* **420**, 168263 (2020).
45. H. Shah and K. Sebastian, *Astrophys. J.* **843**, 131 (2017).
46. H. Sotani and T. Tatsumi, *Mon. Not. R. Astron. Soc.* **467**, 1249 (2017).
47. S. K. Roy, S. Mukhopadhyay, J. Lahiri and D. N. Basu, *Phys. Rev. D* **100**, 063008 (2019).
48. R. L. Bowers and E. P. T. Liang, *Astrophys. J.* **188**, 657 (1974).
49. R. C. Tolman, *Phys. Rev.* **55**, 364 (1939).
50. J. R. Oppenheimer and G. M. Volkoff, *Phys. Rev.* **55**, 374 (1939).
51. P. S. Letelier, *Phys. Rev. D* **22**, 807 (1980).
52. S. S. Bayin, *Phys. Rev. D* **26**, 1262 (1982).
53. E. J. Ferrer, V. de La Incera, J. P. Keith, I. Portillo and P. L. Springsteen, *Phys. Rev. C* **82**, 065802 (2010).
54. A. A. Isayev and J. Yang, *Phys. Rev. C* **84**, 065802 (2011).
55. A. A. Isayev and J. Yang, *Phys. Lett. B* **707**, 163 (2012).
56. D. A. Kirzhnits, *Zh. Eksp. Teor. Fiz.* **38**, 503 (1960); *Sov. Phys.—JETP* **11**, 365 (1960).
57. A. A. Abrikosov, *Zh. Eksp. Teor. Fiz.* **39**, 1798 (1960).
58. E. E. Salpeter, *Astrophys. J.* **134**, 669 (1961).
59. D. J. Stevenson, *J. Phys. Colloques* **41**, C2_61 (1980).
60. E. Garcia-Berro, M. Hernanz, R. Mochkovitch and J. Isern, *Astron. Astrophys.* **193**, 141 (1988).
61. N. Nag and S. Chakrabarty, arXiv:astro-ph/0008477 (2000).
62. M. Bocquet, S. Bonazzola, E. Gourgoulhon and J. Novak, *Astron. Astrophys.* **301**, 757 (1995).
63. K. Konno, T. Obata and Y. Kojima, *Astron. Astrophys.* **352**, 211 (1999).

64. C. Y. Cardall, M. Prakash and J. M. Lattimer, *Astrophys. J.* **554**, 322 (2001).
65. K. Ioka and M. Sasaki, *Astrophys. J.* **600**, 296 (2004).
66. A. Oron, *Phys. Rev. D* **66**, 023006 (2002).
67. K. Kiuchi and K. Kotake, *Mon. Not. R. Astron. Soc.* **385**, 1327 (2008).
68. K. Kiuchi, K. Kotake and S. Yoshida, *Astrophys. J.* **698**, 541 (2009).
69. N. Yasutake, K. Kiuchi and K. Kotake, *Mon. Not. R. Astron. Soc.* **401**, 2101 (2010).
70. J. Frieben and L. Rezzolla, *Mon. Not. R. Astron. Soc.* **427**, 3406 (2012).
71. S. Yoshida, K. Kiuchi and M. Shibata, *Phys. Rev. D* **86**, 044012 (2012).
72. S. Chowdhury and T. Sarkar, *Astrophys. J.* **884**, 95 (2019).
73. P.-C. Chu, L.-W. Chen and X. Wang, *Phys. Rev. D* **90**, 063013 (2014).
74. M. Sinha, B. Mukhopadhyay and A. Sedrakian, *Nucl. Phys. A* **898**, 43 (2013).
75. L. Herrera and W. Barreto, *Phys. Rev. D* **88**, 084022 (2013).
76. H. O. Silva, C. F. B. Macedo, E. Berti and L. C. B. Crispino, *Class. Quantum Gravity* **32**, 145008 (2015).
77. K. Fujisawa, S. Yoshida and Y. Eriguchi, *Mon. Not. R. Astron. Soc.* **422**, 434 (2012).
78. A. T. Potter, C. A. Tout, *Mon. Not. R. Astron. Soc.* **402**, 1072 (2010).
79. D. Bandyopadhyay, S. Chakrabarty and Pal S., *Phys. Rev. Lett.* 79 (1997) 2176.
80. D. Bandyopadhyay, S. Pal and S. Chakrabarty, *J. Phys. G* **24**, 1647 (1998).
81. G. D. Schmidt, H. C. Harris, J. Liebert, D. J. Eisenstein, S. F. Anderson, J. Brinkmann, P. B. Hall, et al., *Astrophys. J.* **595**, 1101 (2003).
82. B. K. Harrison, K. S. Thorne, M. Wakano and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse, Chicago: University of Chicago Press*, (1965).
83. S. Chandrasekhar and E. Fermi, *Astrophys. J.* **118**, 116 (1953).

LIST OF PARTICIPANTS

Abbate, Federico	MPI for Radio astronomy	Germany
Abbsi, Kamran	NUST	Pakistan
Abdollahi, Pooya	Isfahan University of Technology	Iran
Abe, Katsuya	Nagoya University	Japan
Abebe, Amare	North-West University	South Africa
Abele, Hartmut	TU Wien	Austria
Abishev, Medeu	Al Farabi Kazakh National Univ.	Kazakhstan
Abreu de Paula, Marco Aurélio	Federal University of Pará	Brazil
Acebron, Ana	Università degli Studi di Milano	Italy
Acharya, Sandeep Kumar	University of Manchester	UK
Addazi, Andrea	Laboratori Nazionali del Gran Sasso	Italy
Adorno, Tiago	Hebei University	China
Afrin, Misba	Jamia Millia Islamia	India
Afshordi, Niayesh	University of Waterloo	Canada
Aggarwal, Srishty	Indian Institute of Science	India
Aghaei, Maryam	Sejong Uiversity & UNIST	South Korea
Aghanim, Nabila	Institut d'Astrophysique Spatiale	France
Aguayo-Ortiz, Alejandro	UNAM	Mexico
Aguilar Ruiz, Edilberto	Instituto de Astronomia, UNAM	Mexico
Ahlgren, Björn	RIKEN iTHEMS and KTH	Japan
Ahmadvand, Moslem	IPM	Iran
Ahumada, Tomas	University of Maryland	USA
Aimuratov, Yerlan	Fesenkov Astrophysical Institute	Kazakhstan
Ajith, Siddarth	University of Virginia	USA
Akhmedov, Emil	MIPT and ITEP	Russia
Alcubierre, Miguel	Inst. de Ciencias Nucleares	Mexico
Alekseev, Georgy	Steklov Mathematical Institute RAS	Russia
Alestas, George	University of Ioannina	Greece
Ali-Haïmoud, Yacine	New York University	USA
Almeida, Ana	CAUP	Portugal
Almheiri, Ahmed	Institute for Advanced Study	USA
Alonso-Bardaji, Asier	UPV/EHU	Spain
Aloy, Miguel ángel	University of Valencia	Spain
Alstein, Paul	Freudenthal Institute	Netherlands
Amati, Lorenzo	INAF - OAS Bologna	Italy
Ambrus, Victor E.	Goethe University	Germany
Amelino-Camelia, Giovanni	Università Federico II di Napoli	Italy
Amendola, Luca	University of Heidelberg	Germany
Anagnostou, Oliver	University of Melbourne	Australia

Anda, Marcelo	Universidad San Francisco de Quito	Ecuador
Anderson, Paul	Wake Forest University	USA
Ando, Masaki	Univ. of Tokyo	Japan
Andrusenko, Svetlana	Bauman Moscow State Technical Univ.	Russia
Ansoldi, Stefano	University of Udine	Italy
Antypas, Dionysios	Johannes Gutenberg-Universität Mainz	Germany
Anzuini, Filippo	The University of Melbourne	Australia
Arcodia, Riccardo	MPI for Extraterrestrial Physics	Germany
Argüelles, Carlos Raúl	Universidad Nacional de La Plata	Argentina
Arkhangelskaja, Irene	National Research Nuclear University	Russia
Arminjon, Mayeul	Univ. Grenoble Alpes & CNRS	France
Arrechea, Julio	Institute of Astrophysics of Andalusia	Spain
Ashtekar, Abhay	Inst. for Gravitation & the Cosmos	USA
Aspelmeyer, Markus	University of Vienna	Austria
Atri, Pikky	ASTRON	Netherlands
Atteia, Jean-Luc	IRAP - CNRS/UPS/CNES	France
Audagnotto, Giulio	Max Planck Institut	Italy
Avgoustidis, Tasos	University of Nottingham	UK
Aviles, Alejandro	Conacyt	Mexico
Ayuso, Ismael	Instituto de Astrofísica de Portugal	Portugal
Babaie, Narges	University of Tehran	Iran
Badaracco, Francesca	UCLouvain	Belgium
Badia, Javier	IACE	Argentina
Bagheri, Mohammad	University of Tehran	Iran
Balbus, Steven	Oxford University	UK
Bamba, Kazuharu	Fukushima University	Japan
Banerjee, Shreya	Friedrich Alexander University	Germany
Barba González, David	Universidad de Salamanca	Spain
Barbato, Felicia	GSSI	Italy
Barca, Gabriele	La Sapienza University of Rome	Italy
Barrau, Aurélien	LPSC CNRS	France
Barreiro, Tiago	IACE	Portugal
Barroso Bellido, Samuel	University of Szczecin	Poland
Bartosch Caminha, Gabriel	Max Planck Institute for Astrophysics	Germany
Baryshev, Andrey	Kapteyn Astronomical Institute	Netherlands
Başaran öz, Işil	Istanbul University	Turkey
Basquens, Marc	Universidad Carlos III de Madrid	Spain
Basu, Kaustuv	Universität Bonn	Germany
Batrakov, Alexander	MPI for Radio Astronomy	Germany
Battista, Emmanuele	Institute for Theoretical Physics, KIT	Germany
Battistelli, Elia Stefano	Sapienza University of Rome	Italy
Bavarsad, Ehsan	University of Kashan	Iran

Beckwith, Andrew	Chongqing University, physics	China
Beesham, Aroonkumar	University of Zululand	South Africa
Begue, Damien	Bar Ilan University	Israel
Beheshtipour, Banafsheh	AEI	Germany
Bejger, Michał	Nicolaus Copernicus Astronomical Ctr	Poland
Belczynski, Krzysztof	The Polish Academy of Sciences	Poland
Belfiglio, Alessio	University of Camerino	Italy
Belkhadria, Zakaria	Unica - INFN	Italy
Belli, Pierluigi	INFN Roma Tor Vergata	Italy
Bellini, Gianpaolo	University of Milano and INFN	Italy
Bellorin, Jorge	Universidad de Antofagasta	Chile
Bellucci, Stefano	INFN-LNF	Italy
Belonenko, Aleksei	Lomonosov Moscow State University	Russia
Beltrán Palau, Pau	Universitat de València - IFIC (CSIC)	Spain
Beniamini, Paz	Caltech	USA
Benisty, David	University of Cambridge	UK
Berezhiani, Zurab	Univ. L'Aquila and LNGS (Italy)	Italy
Bergamini, Pietro	INAF-OAS	Italy
Bernabei, Rita	University and INFN Roma Tor Vergata	Italy
Bernardini, Elisa	Unipd	Italy
Bernardini, Maria Grazia	INAF - OAB	Italy
Bernardini, Paolo	Universitá del Salento & INFN, Lecce	Italy
Bernuzzi, Sebastiano	FSU Jena	Germany
Berti, Alessio	MPI for Physics	Germany
Berti, Maria	SISSA	Italy
Bhandari, Shivani	CSIRO astronomy and space science	Australia
Bhattacharya, Swastik	BITS Pilani Hyderabad	India
Bhattacharya, Mukul	Virginia Tech, Blacksburg	USA
Bhattacharyya, Soham	AEI Hannover	Germany
Bianchi, Eugenio	Penn State	USA
Bianco, Carlo Luciano	ICRANet	Italy
Bicak, Jiri	Charles University	Czechia
Bihain, Gabriel	Albert Einstein Institute	Germany
Bij, Akanksha	Queen's University	Canada
Bilal, Muhammad	School of Mathematical Sciences	Pakistan
Bini, Donato	CNR	Italy
Biondi, Riccardo	INFN - LNGS	Italy
Birrer, Simon	Stanford University	USA
Bishop, Nigel	Rhodes University	South Africa
Bissaldi, Elisabetta	Politecnico and INFN Bari	Italy
Blair, David	University of Western Australia	Australia
Blanchet, Luc	Institut d'Astrophysique de Paris	France

Blandford, Roger	KIPAC	USA
Blas, Diego	King's College London	UK
Blasi, Simone	VUB	Belgium
Bluemlein, Johannes	DESY	Germany
Bluhm, Robert	Colby College	USA
Bobrick, Alexey	Lund University	Sweden
Boccioli, Luca	University of Notre Dame	USA
Bodmann, Benno	UFRGS	Brazil
Bogush, Igor	Lomonosov Moscow State University	Russia
Bolliet, Boris	Columbia University	USA
Bonavera, Laura	University of Oviedo	Spain
Bondani, Stefano	Università degli studi dell'Insubria	Italy
Bonnaire, Tony	Institut d'Astrophysique Spatiale	France
Bonvin, Camille	University of Geneva	Switzerland
Boos, Jens	William & Mary	USA
Bora, Jyatsnasree	Dibrugarh University	India
Borghi, Nicola	University of Bologna	Italy
Borsevici, Victor	Free International Univ. of Moldova	Moldova
Bose, Nirban	IIT Bombay	India
Boshkayev, Kuantay	Kazakh National University	Kazakhstan
Bošnjak, Željka	Zagreb University - FER	Croatia
Botta, Gioele	University of Warsaw	Italy
Boublil, Shachar Shon	University of Western Australia	Australia
Boudet, Simon	Trento University	Italy
Bozza, Valerio	Università di Salerno, Italy	Italy
Branchesi, Marica	Gran Sasso Science Institute/INFN	Italy
Brandt, William	Penn State University	USA
Brewer, William Dean	Freie Universität Berlin	Germany
Briscese, Fabio	SUSTech	China
Broderick, Avery	Perimeter Institute	Canada
Bromm, Volker	University of Texas at Austin	USA
Bulbul, Esra	MPI for Extraterrestrial Physics	Germany
Buldgen, Gaël	University of Geneva	Switzerland
Burgess, J. Michael	MPE	Germany
Burigana, Carlo	INAF-Istituto di Radioastronomia	Italy
Burns, Eric	Louisiana State University	USA
Burtscher, Annegret	Radboud University Nijmegen	Netherlands
Cabral, Francisco	University of Lisbon	Portugal
Cabrera Rosas, Omar de Jesús	CINVESTAV	Mexico
Caccianiga, Barbara	INFN	Italy
Cai, Yi-Fu	Univ. of Science and Technology of China	China
Calderón, Rodrigo	Université de Montpellier	France

Caleb, Manisha	University of Manchester	UK
Camci, Ugur	Roger Williams University	USA
Cameron, Andrew	Swinburne University of Technology	Australia
Caminata, Alessio	INFN-Genova	Italy
Campion, Stefano	ICRANeT	Italy
Canay, Ezgi	Istanbul Technical University	Turkey
Cantiello, Michele	INAF OAAb	Italy
Capone, Antonio	University La Sapienza and INFN	Italy
Caracciolo, Vincenzo	University of Rome "Tor Vergata"	Italy
Carella, Elisabetta	Università degli Studi di Milano	Italy
Carelli, Giorgio	INFN sezione di Pisa	Italy
Carlin, Julian	The University of Melbourne	Australia
Carmona, Jose Manuel	CAPA/University of Zaragoza	Spain
Carosi, Alessandro	University of Geneva - DPNC	Switzerland
Carvalho, Ana	IA/FCUL	Portugal
Cederbaum, Carla	Tuebingen University	Germany
Célérier, Marie-Noëlle	LUTH Observatoire de Paris	France
Cerruti, Matteo	Université de Paris - APC	France
Cerutti, Benoît	CNRS & Univ. Grenoble Alpes	France
Cervantes-Cota, Jorge L.	ININ (Mexico)	Mexico
Chael, Andrew	Princeton University	USA
Chakraborty, Indranil	Indian Institute of Technology Kharagpur	India
Chakraborty, Sumanta	Indian Ass. for the Cultivation of Science	India
Chakraborty, Srija	Scuola Normale Superiore	Italy
Chamel, Nicolas	Université Libre de Bruxelles	Belgium
Chandra, Koustav	Indian Institute of Technology Bombay	India
Chandran, S. Mahesh	IIT Bombay	India
Chasovnikov, Aristarh	Lomonosov Moscow State University	Russia
Chatziioannou, Katerina	Caltech	USA
Chauvineau, Bertrand	UCA/OCA/Lagrange laboratory	France
Chavanis, Pierre-Henri	Lab. de Physique Theorique de Toulouse	France
Chaves-Montero, Jonas	DIPC	Spain
Chawla, Pragya	McGill University	Canada
Chen, Liang	IHEP	China
Chen, Geoff Chih-Fan	UCLA	USA
Cherubini, Christian	University Campus Bio-Medico of Rome	Italy
Chluba, Jens	JBCA	UK
Chojnacki, Jan	University of Warsaw	Poland
Choudhury, Tirthankar Roy	Natl. Ctr for Radio Astrophys., Pune	India
Christiansen, Øyvind	University of Oslo	Norway
Ciufolini, Ignazio	UNIVERSITY OF SALENTO	Italy
Cleaver, Gerald	Baylor University	USA

Colafranceschi, Eugenia	University of Nottingham	UK
Conceição, Miguel	FCUL; IACE	Portugal
Cong, Wan	University of Waterloo	Canada
Coppi, Bruno	MIT	USA
Cordier, Bertrand	CEA-Saclay	France
Correa, Miguel	University of Notre Dame	USA
Correia, José Ricardo	Universidade do Porto	Portugal
Costa, Filipe	CAMGSD - Universidade de Lisboa	Portugal
Cotsakis, Spiros	Univ. Aegean, Greece	Greece
Crinquand, Benjamin	IPA Grenoble	France
Crispino, Luis	Universidade Federal do Pará	Brazil
Criswell, Alexander	University of Minnesota	USA
Crosta, Mariateresa	INAF-OATo	Italy
Curceanu, Catalina Oana	INFN-LNF	Italy
D'Amico, Giacomo	University of Bergen	Norway
D'Angelo, Annalisa	Universitá di Roma Tor Vergata	Italy
D'Avanzo, Paolo	INAF - OAB	Italy
da Fonseca, Vitor	Inst. of Astrophysics and Space Science	Portugal
da Silva, Antonio	IACE	Portugal
Dafermos, Mihalis	Princeton University	UK
Dahal, Pravin	Macquarie University	Australia
Dainotti, Maria	NAOJ	Japan
Darling, Jeremy	University of Colorado, Boulder	USA
Das, Saurya	University of Lethbridge	Canada
Das, Siddhant	Ludwig-Maximilians-Universitat	Germany
Datta, Sudeb Ranjan	Indian Institute of Science	India
Dayal, Pratika	Kapteyn Astronomical Institute	Netherlands
de Amorim, Erik	Universität zu Köln	Germany
de Cruz, Javier	University of Barcelona	Spain
De Felice Proia, Giuseppina	University of Rome "Tor Vergata"	Italy
De Felice Proia, Valeria	IIS Federico Caffè	Italy
de Mink, Selma	MPI for Astrophysics	Germany
De Mitri, Ivan	GSSI and INFN	Italy
De Paolis, Francesco	University of Salento	Italy
De Rosa, Alessandra	IAPS/INAF	Italy
De Simone, Biagio	Università degli Studi di Salerno	Italy
de Souza Campos, Lissa	University of Pavia	Italy
Deb, Debabrata	Indian Institute of Science	India
Decelle, Aurélien	Universidad Complutense de Madrid	Spain
Del águila, Juan Carlos	I.P.N.	Mexico
Delabrouille, Jacques	CNRS	France
Delband, Masoumeh	Shiraz University	Iran

Delgado, Jorge	University of Aveiro	Portugal
Delhom I Latorre, Adrià	University of Valencia	Spain
DeMille, David	University of Chicago	USA
Deng, Heling	Arizona State University	USA
Dereli-Bégué, Hüsne	Bar-Ilan University	Israel
Dergachev, Vladimir	AEI Hannover	USA
Dev, Bhupal	Washington Univ. in St. Louis	USA
Dhanuka, Ankit	IISER Mohali	India
Di Clemente, Francesco	INFN Ferrara	Italy
Di Giovanni, Fabrizio	Universitat de Valencia	Spain
Di Mauro, Marco	Università di Salerno	Italy
di Niccolo, Cinzia	ICRANet Secretariat	Italy
Di Palma, Irene	Sapienza University of Rome	Italy
Di Valentino, Eleonora	Durham University	UK
Di Virgilio, Angela D. V.	INFN-Pisa	Italy
Díaz Jiménez, Bogar	Universidad Carlos III de Madrid	Spain
Dichiara, Simone	UMD/NASA-GSFC	USA
Dickau, Jonathan	Independent Researcher	USA
Dimakis, Nikolaos	Sichuan University	China
Ding, Chikun	Huaihua University	China
Dorozsmai, Andras	University of Birmingham	UK
dos Anjos, Rita de Cassia	UFPR	Brazil
Douspis, Marian	IAS	France
Drago, Alessandro	University of Ferrara	Italy
Dreissen, Laura	Physikalisch-Technische Bundesanstalt	Germany
Drew, Amelia	University of Cambridge	UK
Droguett, Byron	Universidad de antofagasta	Chile
Dror, Jeff	UC Santa Cruz	USA
Drossel, Barbara	Technische Universität Darmstadt	Germany
Du, Yuchen	University of Virginia	USA
Duarte Lima Junior, Haroldo	Universidade Federal do Pará	Brazil
Ducobu, Ludovic	University of Mons (PhD student)	Belgium
Dumin, Yurii	Sternberg Astronomical Institute	Russia
Dunn, Liam	University of Melbourne	Australia
Dunsby, Peter	University of Cape Town	South Africa
Dupree, Andrea	Harvard & Smithsonian	USA
Durkan, Leanne	University College Dublin	Ireland
Dutta Chowdhury, Dhruba	Yale University	USA
Dutta Roy, Poulami	IIT Kharagpur	India
Dzuba, Vladimir	UNSW Sydney	Australia
Ebrahimi Khuzani, M.	IUT	Iran
Eder, Konstantin	Friedrich-Alexander-Universität	Germany

Efstathiou, George	University of Cambridge	UK
Eiroa, Ernesto F.	IAFE (CONICET-UBA)	Argentina
Elizalde, Emilio	ICE-CSIC and IEEC	Spain
Ellis, David	University of Göttingen	Germany
Elor, Gilly	MITP	Germany
Engel, Kristi	HAWC, University of Maryland	USA
Er, Xinzhong	SWIFAR, Yunnan University	China
Escrivà, Albert	University of Barcelona	Spain
Espinosa-Portalés, Llorenç	Instituto de Física Teórica UAM-CSIC	Spain
Estellés, Héctor	Universitat de les Illes Balears	Spain
Etebar, Sina	Isfahan University of Technology	Iran
Evangelista, Yuri	INAF IAPS	Italy
Famiano, Michael	Western Michigan University	USA
Fanizza, Giuseppe	IA - Lisbon	Portugal
Faraji, Shokoufe	ZARM	Germany
Faraoni, Valerio	Bishop's University	Canada
Farzan, Yasaman	IPM	Iran
Feng, Hua	Tsinghua University	China
Fernandez Uria, Sara	University of the Basque Country	Spain
Feroci, Marco	INAF/IAPS	Italy
Ferreiro, Antonio	University of Valencia/IFIC	Spain
Ferro, Lisa	University of Ferrara	Italy
Fesik, Liudmila	Albert Einstein Institute	Germany
Fialkov, Anastasia	Cambridge University	UK
Field, Scott	University of Massachusetts	USA
Fier, Jared	Baylor Univsersity	USA
Figueroa-Aguirre, Griselda	IAFE (CONICET-UBA)	Argentina
Filippi, Simonetta	Campus Bio-Medico University	Italy
Finster, Felix	Universität Regensburg	Germany
Fishbach, Maya	CIERA/Northwestern University	USA
Fisher, Robert	University of Massachusetts	USA
Flambaum, Victor	University of New South Wales	Australia
Foffa, Stefano	Geneva University	Switzerland
Foley, Ryan	UC Santa Cruz	USA
Fragos, Tassos	University of Geneva	Switzerland
Fraija, Nissim	Instituto de Astronomia, UNAM	Mexico
Franchino-Viñas, Sebastián	Universität Heidelberg	Germany
Franco, Antonio	Università del Salento	Italy
Fraser, Patrick	University of Toronto	Canada
Frati, Maria	Senato della Repubblica	Italy
Frederiks, Dmitry	Ioffe Institute, Saint-Petersburg	Russia
Freedman, Wendy	University of Chicago	USA

Freire, Paulo	MPI for Radio Astronomy	Germany
Fridman, Mitja	University of Lethbridge	Canada
Frisoni, Pietropaolo	Western University	Canada
Frontera, Filippo	Univ. di Ferrara e INAF-OAS Bologna	Italy
Frost, Torben	University of Bremen	Germany
Fruck, Christian	Max-Planck-Institute for Physics	Germany
Frusciante, Noemi	IA/FCUL	Portugal
Fryer, Chris	Los Alamos National Laboratory	USA
Furey, Nichol	Humboldt University	Germany
Gaijan, Mrunali	University of Milan	Italy
Galindo-Uribarri, Salvador	ININ	Mexico
Gallego Cadavid, Alexander	Universidad de Valparaíso	Chile
Gallegos, Omar	Cinvestav-IPN	Mexico
Galtsov, Dmitry	Lomonosov Moscow State University	Russia
Galván, Antonio	Institute of Astronomy, UNAM	Mexico
Gan, Wencong	Baylor University	China
Gangopadhyay, Mayukh	Center For Theoretical Physics, JMI	India
Gao, Dongfeng	IAPMST	China
Gao, Feng	Hamburger Sternwarte	Germany
Garcia-Bellido, Juan	Universidad Autonoma de Madrid	Spain
García-Peláez, David	Universidad Autónoma Metropolitana	Mexico
García-Quismondo, Alejandro	IEM-CSIC	Spain
Gargano, Fabio	INFN	Italy
Garofalo, David	Kennesaw State University	USA
Gasparyan, Sargis	ICRANet-Armenia	Armenia
Gautam, Tasha	MPI For Radioastronomy	Germany
Gautam, Shalabh	IUCAA	India
Gelles, Zachary	Harvard University	USA
Gendre, Bruce	University of Western Australia	Australia
Genzel, Reinhard	MPI for Extraterrestrial Physics	Germany
George, Sandip	Univ. Medical Ctr Groningen	Netherlands
Gervalle, Romain	Institut Denis Poisson	France
Ghezelbash, Masoud	University of Saskatchewan	Canada
Gholizadeh Siahmazgi, S.	Wake Forest University	USA
Gholizadeh, Kasra	—	Iran
Ghosh, Subham	Indian Institute of Science	India
Ghosh, Sushant	University of KwaZulu-Natal	India
Giacomazzo, Bruno	University of Milano-Bicocca	Italy
Giacomelli, Umberto	GSSI - INFN - Univ. di Pisa	Italy
Giannakopoulos, Thanasis	Instituto Superior Técnico, Lisbon	Portugal
Giesel, Kristina	FAU Erlangen-Nürnberg	Germany
Gill, Ramandeep	The Open University of Israel	Israel

Gilman, Daniel	University of Toronto	Canada
Ginat, Barry	Technion - Israel Institute of Technology	Israel
Giovannelli, Franco	INAF-IAPS	Italy
Giovannetti, Eleonora	Università di Roma “La Sapienza”	Italy
Giri, Priyanka	INFN Pisa	Italy
Gitman, Dmitry	P.N. Lebedev Physical Institute	Russia
Gittins, Fabian	University of Southampton	UK
Glushikhina, Maria	Space Research Institute (IKI RAN)	Russia
Gogoi, Dhruba Jyoti	Dibrugarh University	India
Goicoechea, Luis J.	Universidad de Cantabria	Spain
Gomes de Oliveira, Fernanda	LSQM	Germany
Gomes, Pierre Alexandre	Universidade do Estado de Santa Catarina	Brazil
Gómez-Valent, Adrià	Institut für Theoretische Physik	Germany
Goncalves, Tiago	University of Lisbon	Portugal
Good, Michael	Nazarbayev University	Kazakhstan
Gorbatsievich, Alexander	Belarusian State University	Belarus
Gorla, Paolo	INFN-LNGS	Italy
Gottlieb, Ore	Tel Aviv University/CIERA	Israel
Goulart Coelho, Jaziel	UTFPR	Brazil
Gouttenoire, Yann	Tel Aviv University	France
Gralla, Samuel	University of Arizona	USA
Granot, Jonathan	The Open University of Israel	Israel
Grant, Alexander	University of Virginia	USA
Gray, Finnian	Perimeter Institute	Canada
Gregoris, Daniele	Jiangsu University	China
Grillo, Claudio	University of Milan	Italy
Gudmundsson, Jon	Stockholm University	Sweden
Guerrero, Merce	UCM	Spain
Guetta, Dafne	ORT Braude	Israel
Guffanti, Daniele	Johannes Gutenberg-Univ. Mainz	Germany
Guglielmi, Diego	IIS Federico Caffè	Italy
Guinan, Edward	Villanova University	USA
Guo, Minyong	Beijing Normal University	China
Gushima, Yuya	Fukuoka university	Japan
Gwak, Bogeun	Dongguk University	South Korea
Ha, Yuan K.	Temple University	USA
Hackmann, Eva	ZARM, University of Bremen	Germany
Hackstein, Jan	ZARM, University of Bremen	Germany
Haemmerlé, Lionel	Université de Genève	Switzerland
Hafizi, Mimoza	University of Tirana	Albania
Haghani, Zahra	Damghan University	Iran
Hajela, Aprajita	Northwestern University	USA

Hajian, Kamal	HWK institute for advanced study	Iran
Halim, Odysse	INFN Trieste	Italy
Halla, Mourad	ZARM, University of Bremen	Germany
Halzen, Francis	University of Wisconsin- Madison	USA
Hamada, Yu	KEK	Japan
Hamed, Nima	Isfahan University of Technology	Iran
Hammad, Fayçal	Bishop's University	Canada
Hamolli, Lindita	University of Tirana	Albania
Hamuy, Mario	AURA	Chile
Hanauske, Matthias	Goethe University Frankfurt	Germany
Harms, Jan	Gran Sasso Science Institute	Italy
Harper, Graham	University of Colorado Boulder	USA
Harriott, Tina	Mount Saint Vincent University	Canada
Hart, Luke	JBCA, University of Manchester	UK
Harte, Abraham	Dublin City University	Ireland
Hartmann, Angelo	University of Insubria	Italy
Harutyunyan, Gevorg	ICRANet-Armenia	Armenia
Haxton, Wick	University of California, Berkeley	USA
Hayama, Kazuhiro	University of Texas	Japan
Hayes, Fergus	University of Glasgow	UK
He, Hong-Jian	Shanghai Jiao Tong University	China
Heimersheim, Stefan	University of Cambridge	UK
Heinzel, Gerhard	Max Planck Institut Hannover	Germany
Held, Aaron	Imperial College London	UK
Hendi, Seyed Hossein	Shiraz University	Iran
Herdeiro, Carlos	Aveiro University, Portugal	Portugal
Herman, Nicolas	University of Namur	Belgium
Hernanz, Margarita	ICE (CSIC) & IEEC	Spain
Herrera Aguilar, Alfredo	Universidad Autónoma de México	Mexico
Herrmann, Friedrich	Karlsruhe Institute of Technology	Germany
Hidalgo, Carlos	ICF-UNAM	Mexico
Hill, Colin	Columbia Univ./Flatiron Inst.	USA
Himwich, Elizabeth	Harvard University	USA
Hinton, Jim	MPI Kernphysik, Heidelberg	Germany
Hiramatsu, Takashi	Rikkyo Univ.	Japan
Hoang, Vu	University of Texas at San Antonio	USA
Holz, Daniel	University of Chicago	USA
Hoque, SK Jahanur	Charles University	Czechia
Hu, Huanchen	MPI for Radio Astronomy	Germany
Huang, Yong-Feng	Nanjing University	China
Hughes, Scott	MIT	USA
Husa, Sascha	University of the Balearic Islands	Spain

Ianni, Aldo	INFN Gran Sasso Laboratory	Italy
Ibata, Rodrigo	Obs. Astronomique de Strasbourg	France
Ilyas, Amara	Univ. of Science and Technology of China	China
Imtiaz, Batool	Univ. of Science and Technology of China	China
Inomata, Keisuke	University of Chicago, KICP	USA
Islam, Tousif	University of Massachusetts	USA
Israyelyan, Davit	ICRANet-Armenia	Armenia
Ito, Hirotaka	RIKEN	Japan
Iyonaga, Aya	Rikkyo University	Japan
Izzo, Luca	University of Copenhagen	Denmark
Jain, Parul	Indian Institute of Technology	India
Jana, Susmita	IIT BOMBAY	India
Jani, Karan	Vanderbilt University	USA
Jantzen, Robert Theodore	Villanova University	USA
Jawad, Abdul	COMSATS University Islamabad	Pakistan
Jensko, Erik	University College London	UK
Jetzer, Philippe	University of Zurich	Switzerland
Jha, Saurabh	Rutgers University	USA
Jiang, Nan	University of Virginia	USA
Johnson, Joseph P	IIT Bombay	India
Johnson, Michael	Harvard & Smithsonian	USA
Jones, Michael	University of Colorado	USA
Jordana-Mitjans, Nuria	University of Bath	UK
Joshi, Jagdish	Nanjing University China	China
Joudaki, Shahab	University of Waterloo	Canada
Joyce, Meridith	Space Telescope Science Institute	USA
Jullo, Eric	Aix-Marseille Univ/LAM	France
Kaczmarek, Jane	NRC	Canada
Kalinani, Jay Vijay	University of Padova	Italy
Kalita, Surajit	Indian Institute of Science, Bangalore	India
Kamionkowski, Marc	Johns Hopkins University	USA
Kanatchikov, Igor	Quantum Information Center	Poland
Kang, Sunghyun	Sogang University	South Korea
Karas, Vladimír	Astronomical Institute, CAS	Czechia
Kargaltsev, Oleg	The George Washington University	USA
Karinkuzhi, Drisya	Indian Institute of Science	India
Kaspi, Victoria	McGill University	Canada
Kawamura, Seiji	Nagoya University	Japan
Kazanas, Demosthenes	NASA/Goddard Space Flight Center	USA
Kazaras, Demetre	Duke University	USA
Kedia, Atul	University of Notre Dame	USA
Keeley, Ryan	Korea Astr. Space Sci. Inst.	South Korea

Keitel, David	Universitat de les Illes Balears	Spain
Kerin, Alex	University of Melbourne	Australia
Kerr, Roy P.	Canterbury University	New Zealand
Kersting, Magdalena	University of Oslo	Norway
Kerzendorf, Wolfgang	Michigan State University	USA
Khadka, Narayan	Kansas State University	USA
Kharuk, Ivan	INR RAS & MIPT	Russia
Khetan, Nandita	Gran Sasso Science Institute	Italy
Khlopunov, Mikhail	Lomonosov Moscow State University	Russia
Kiessling, Michael	Rutgers University	USA
Kim, Vitaliy	Fesenkov Astrophysical institute	Kazakhstan
Kim, Sang Pyo	Kunsan National University	South Korea
Kim, Sung-Won	Ewha Womans University	South Korea
King, Steven	Physikalisch-Technische Bundesanstalt	Germany
Kirillov, Alexandr	Bauman Moscow State Technical Univ.	Russia
Kisielowski, Marcin	National Centre for Nuclear Research	Poland
Klainerman, Sergiu	Princeton University	USA
Klaoudatou, Ifigeneia	University of the Aegean	Greece
Klinkhamer, Frans R.	Karlsruhe Institute of Technology	Germany
Kobialko, Kirill	Moscow State University	Russia
Kohri, Kaz	KEK	Japan
Kolanowski, Maciej	University of Warsaw	Poland
Kole, Merlin	University of Geneva	Switzerland
Kolekar, Sanved	DAE - Ctr for Excellence in Basic Sci.	India
Komarov, Stanislav	Belarusian State University	Belarus
Kondratyev, Ilya	Space Research Institute RAS	Russia
Konkowski, Deborah	U S Naval Academy	USA
Kontou, Eleni-Alexandra	University of Amsterdam	Netherlands
Konysbayev, Talgar	Al-Farabi Kazakh National Univ.	Kazakhstan
Koo, Hanwool	KASI/UST	South Korea
Korzyński, Mikołaj	Center for Theoretical Physics	Poland
Kosakowski, Daniel	University of Massachusetts	USA
Koshelev, Alexey	universidade da beira interior	Portugal
Kovalev, Yuri	Lebedev Physical Institute	Russia
Kowalski-Glikman, Jerzy	University of Wroclaw	Poland
Kramer, Michael	MPI fuer Radioastronomie	Germany
Kranas, Dimitrios	Louisiana State University	USA
Krichevskiy, Daniil	Bauman Moscow State Technical Univ.	Russia
Krizmanic, John	UMBC/CRESST/NASA/GSFC	USA
Krut, Andreas	ICRANet	Germany
Kubiznak, David	Perimeter Institute	Canada
Kühnel, Florian	LMU Munich	Germany

Kukihara, Moe	Fukuoka University	Japan
Kulitskii, Aleksadr	Lomomosov Moscow state university	Russia
Kumar, Shailesh	Indian Inst. of Information Technology	India
Kunz, Jutta	University of Oldenburg	Germany
Kurmanov, Ergali	Al-Farabi Kazakh National Univ.	Kazakhstan
Kuruvilla, Joseph	Universite Paris-Saclay	France
Kushwaha, Ashu	Indian Institute of Technology Bombay	India
Kutluk, Emine Seyma	METU	Turkey
Kwapisz, Jan	University of Warsaw	Poland
L'Huillier, Benjamin	Sejong University	South Korea
La Mura, Giovanni	LIFEP	Portugal
La Placa, Riccardo	INAF - Osservatorio astronomico di Roma	Italy
Lacchini, Mattia	Università dell'Insubria	Italy
Laemmerzahl, Claus	University of Bremen	Germany
Laguna, Pablo	University of Texas at Austin	USA
Lampe, Bodo	Univ. Hamburg	Germany
Langer, Norbert	Uni Bpnn/MPIfR Bonn	Germany
Lapola, Marcelo	Instituto Tecnológico de Aeronáutica	Brazil
Lappicy, Phillipo	University of Sao Paulo	Brazil
Lapponi, Alessio	UNICAM, University of Camerino.	Italy
Lasky, Paul	Monash University	Australia
Lau, Shu Yan	University of Virginia	USA
Lazzaro, Claudia	Università degli Studi di Padova	Italy
Lebed, Andrei	University of Arizona	USA
Lecian, Orchidea Maria	Sapienza University of Rome	Italy
Ledvinka, Tomáš	Charles University	Czechia
Lee, Yi Shuen Christine	The University of Melbourne	Australia
Lee, Yi Shuen Christine Lee	The University of Melbourne	Australia
LeFloch, Philippe G.	Sorbonne University	France
Lehoucq, Léonard	ENS Paris-Saclay	France
Leibundgut, Bruno	European Southern Observatory	Germany
Leigh, Nathan	Universidad de Concepcion	Chile
Lenart, Aleksander	Jagiellonian University	Poland
Lentz, Erik	Pacific Northwest National Lab	USA
Leung, James	University of Sydney	Australia
Leung, Calvin	MIT	USA
Leung, Shing-Chi	TAPIR, Caltech	USA
Levine, Delina	University of Maryland	USA
Lewandowski, Jerzy	Uniwersytet Warszawski	Poland
Lewicki, Marek	University of Warsaw	Poland
Li, Ang	Xiamen University	China
Li, Benliang	Southwest Jiaotong University	China

Li, Di	National Astronomical Observatories	China
Li, Xiang	Purple Mountain Observatory, CAS	China
Li, Xiangdong	Nanjing University	China
Li, Liang	ICRANet	Italy
Li, Dian-Heng	886910871458	Taiwan
Li, BaoFei	Louisiana state university	USA
Li, Dongzi	California Institute of Technology	USA
Lim, Eugene	King's College London	UK
Lim, Hyun	Los Alamos National Laboratory	USA
Lin, Shin-Ted	Sichuan University	China
Ling, Eric	Rutgers University	USA
Lipunov, Vladimir	Lomonosov Moscow State University, SAI	Russia
Liška, Marek	Charles University	Czechia
List, Meike	DLR	Germany
Liu, Jianglai	SJTU	China
Liu, Lang	Institute of Theoretical Physics, CAS	China
Liu, Ruoyu	Nanjing University	China
Liu, Zike	Nanjing University	China
Liu, Hongguang	FAU Erlangen-Nürnberg	Germany
Lizarraga, Joanes	University of the Basque Country	Spain
Lobanov, Andrei	MPIfR Bonn	Germany
Lobato, Ronaldo	Universidad de los Andes	Brazil
Lobo, Francisco	University of Lisbon	Portugal
Lochan, Kinjalk	IISER Mohali	India
Lockhart, Will	University of Arizona	USA
Longo, Francesco	University of Trieste and INFN, Trieste	Italy
Lopez Eiguren, Asier	Tufts University	USA
López Pérez, Samantha	ICMF	Cuba
López-Caniego, Marcos	Aurora Technology for ESA	Spain
Lopez-Monsalvo, Cesar Simon	Universidad Autonoma Metropolitana	Mexico
Loppini, Alessandro	Università Campus Bio-Medico di Roma	Italy
Lorimer, Duncan	WVU	USA
Lovell, Mark	University of Iceland	Iceland
Lu, Fangjun	IHEP, CAS	China
Lucca, Matteo	ULB	Belgium
Lucchesi, David	INAF/IAPS Tor Vergata	Italy
Lundblad, Nathan	Bates College	USA
Luo, Wentao	USTC	China
Luongo, Orlando	University of Camerino, Physics Division	Italy
Lupsasca, Alex	Princeton University	USA
Lusso, Elisabeta	Università degli Studi di Firenze	Italy
Lutovinov, Alexander	Space Research Institute (IKI)	Russia

Lynch, Morgan	Technion - Israel Institute of Technology	Israel
Lyubarsky, Yuri	Ben-Gurion University	Israel
Macaulay, Ed	University of East Anglia	UK
Mach, Patryk	Jagiellonian University	Poland
Macias, Alfredo	Universidad Autonoma Metropolitana	Mexico
MacLaurin, Colin	University of Queensland	Australia
Maffei, Bruno	IAS, Université Paris-Saclay	France
Magalhães, Renan	Universidade Federal do Pará	Brazil
Mahmoudi, Somayye	Shiraz University, Iran	Iran
Maiorano, Michele	Università del Salento	Italy
Maitra, Chandreyee	MPI for Extraterrestrial Physics	Germany
Majid, Amal	University of the Punjab	Pakistan
Malafarina, Daniele	Nazarbayev University	Kazakhstan
Maldacena, Juan	Institute for Advanced Study	USA
Málek, Tomáš	Institute of Mathematics, CAS	Czechia
Malheiro, Manuel	Instituto Tecnológico de Aeronáutica	Brazil
Malik, Sunil	IIT Bombay and University of Delhi	India
Maltsev, Kiril	University of Heidelberg / HITS	Germany
Malybayev, Algis	Al-Farabi Kazakh National Univ.	Kazakhstan
Maniccia, Giulia	Sapienza University	Italy
Manreza-Paret, Daryel	Universidad de La Habana	Cuba
Marakis, Vassilios	Universität Hildesheim	Germany
Marcantognini, Nicola	University of Camerino	Italy
Margalef, Juan	PSU	USA
Margutti, Raffaella	UC Berkeley	USA
Marrocchesi, Pier Simone	University of Siena and INFN Pisa	Italy
Marsh, David	King's College London	UK
Martins, Carlos	CAUP	Portugal
Martire, Gianni	Applied Physics	USA
Masi, Silvia	Sapienza Università di Roma	Italy
Mastrototaro, Leonardo	Università degli Studi di Salerno	Italy
Mateu-Lucena, Maria Teresa	University of the Balearic Islands	Spain
Mathews, Grant	University of Notre	USA
Matos, Tonatiuh	IPN	Mexico
Mauro, Lorenza	Roma Tre	Italy
Mavromatos, Nikolaos	King's College London	UK
Mayerson, Daniel	IPhT, CEA Saclay	France
McCabe, Christopher	King's College London	UK
McGrew, William	NIST-Boulder	USA
McLaughlin, Andrew	North Carolina Central University	USA
Meda, Paolo	University of Genoa - INFN	Italy
Meintjes, Pieter	University of the Free State	South Africa

Melatos, Andrew	University of Melbourne	Australia
Memmen, Jan-Menno	ZARM	Germany
Mena Marugán, Guillermo A.	IEM, CSIC	Spain
Mendes, Melissa	McGill University	Canada
Menéndez-Pidal, Lucía	University of Nottingham	UK
Meng, Xinhe	Nankai University	China
Meng, Yanzhi	Nanjing University	China
Merafina, Marco	University of Rome La Sapienza	Italy
Merloni, Andrea	MPE	Germany
Messina, Andrea	Sapienza University of Rome & INFN	Italy
Metzger, Brian	Columbia University	USA
Meyers, Patrick	University of Melbourne	Australia
Meynet, Georges	Geneva University	Switzerland
Miceli, Davide	University of Udine & INFN Trieste	Italy
Michilli, Daniele	McGill University	Canada
Michimura, Yuta	University of Tokyo	Japan
Micko, Jakob	Institut Laue-Langevin, TU Wien	Austria
Miedema, Pieter	ASFYON	Netherlands
Mikhailov, Evgeny	Lomonosov Moscow State University	Russia
Miller-Jones, James	ICRAR	Australia
Millhouse, Meg	University of Melbourne	Australia
Mimoso, José Pedro	IACE	Portugal
Minic, Djordje	Virginia Tech	USA
Mirabel, Felix	IAFE & CEA	Argentina
Mirzoyan, Razmik	Max-Planck-Institute for Physics	Germany
Mishra, Tanmaya	University of Florida, Gainesville	USA
Mitsou, Vasiliki	IFIC - Univ. of Valencia and CSIC	Spain
Mitsui, Tadao	Tohoku University	Japan
Modafferi, Luana Michela	Universitat de les Illes Balears	Spain
Mohammadi, Rohoollah	National Science and Technology Museum	Iran
Moiseenko, Sergey	Space Research Institute	Russia
Moita, Miguel	University of Ferrara	Italy
Momtaz, Aidin	Isfahan University of Technology	Iran
Mondal, Tushar	Indian Institute of Science	India
Mooley, Kunal	Caltech	USA
Moradi, Rahim	ICRANet	Italy
Morales, Gibran	IA-UNAM	Mexico
Moreno Pulido, Cristian	University of Barcelona	Spain
Moresco, Michele	University of Bologna	Italy
Mori, Masaki	Ritsumeikan University	Japan
Moriya, Takashi	National Astronomical Observatory	Japan
Moriyama, Kotaro	MIT	USA

Most, Elias Roland	Princeton University	USA
Mueller, Bernhard	Monash University	Australia
Mukherjee, Dipayan	IISER Mohali	India
Mukherjee, Suvodip	University of Amsterdam	Netherlands
Mukhopadhyay, Banibrata	Indian Institute of Science	India
Mumtaz, Saadia	University of the Punjab	Pakistan
Müntinga, Hauke	German Aerospace Center (DLR)	Germany
Murk, Sebastian	Macquarie University, Sydney	Australia
Murshed, Adib	Qavloical Authority	Romania
Myrvold, Wayne	The University of Western Ontario	Canada
Myrzakul, Aizhan	Nazarbayev University	Kazakhstan
Nadal Gisbert, Sergi	University of València-IFIC (CSIC)	Spain
Naddeo, Adele	INFN, Sezione di Napoli	Italy
Naidoo, M	Rhodes University	South Africa
Nakas, Theodoros	University of Ioannina	Greece
Nalewajko, Krzysztof	Nicolaus Copernicus Astronomical Ctr	Poland
Namdar, Mohammadhossein	Isfahan University Of Technology	Iran
Nandra, Kirpal	MPI for extraterrestrial Physics	Germany
Naqvi, Syed	Jagiellonian Universtiy	Poland
Narayan, Ramesh	Harvard University	USA
Narita, Makoto	National Institute of Technology	Japan
Natale, Elisabetta	ICRANet	Italy
Natarajan, Priya	Yale University	USA
Naumov, Dmitry	JINR	Russia
Nava-Callejas, Martin Javier	Instituto de Astronomia - UNAM	Mexico
Navarro-Salas, Jose	University of Valencia-IFIC (CSIC)	Spain
Nenmeli, Vijay	Indian Institute of Technology, Bombay	India
Newsome, Ian	Wake Forest University	USA
Ng, Y. Jack	University of North Carolina	USA
Ni, Wei-Tou	APMST	China
Nickel, Otmar	Bundesdeutsche Arbeitsgemeinschaft	Germany
Nicolas, Nora	IP2I	France
Nielson, Via	University of Michigan	USA
Nimmo, Kenzie	ASTRON/University of Amsterdam	Netherlands
Nomoto, Ken'ichi	The University of Tokyo	Japan
Noui, Karim	IDP Tours & APC Paris	France
Novak, Jan	Technical University of Liberec	Czechia
Novikov, Dmitry	Astro Space Center, Moscow	Russia
Nucita, Achille	University of Salento & INFN	Italy
Nunes, Sílvia	ITA	Brazil
Ó Colgáin, Eoin	Sogang University	South Korea
O'Connor, Brendan	George Washington University	USA

O'Neill, Nicholas	University of Melbourne	Australia
Oates, Samantha	University of Birmingham	UK
Okolow, Andrzej	University of Warsaw	Poland
Olejak, Aleksandra	Nicolaus Copernicus Astronomical Ctr	Poland
Olmedo, Javier	Universidad de Granada	Spain
Olmo, Gonzalo J.	University of Valencia	Spain
Ortolan, Antonello	INFN - LNL	Italy
Osborn, Nicole	Purdue University	USA
Ossowski, Maciej	University of Warsaw	Poland
Özsoy, Ogan	CEICFP	Czechia
Pace, Francesco	Università di Bologna	Italy
Page, Dany	UNAM	Mexico
Pagliara, Giuseppe	University of Ferrara (IT)	Italy
Pagliaro, Gianluca	MPI/AEI	Germany
Pai, Archana	Indian Institute of Technology Bombay	India
Palagiano, Cosimo	Sapienza università di Roma	Italy
Pallavicini, Marco	Università di Genova and INFN Genova	Italy
Palumbo, Daniel	Harvard & Smithsonian	USA
Pannarale, Francesco	Sapienza University of Rome	Italy
Panotopoulos, Grigorios	Universidade de Lisboa	Portugal
Papa, Maria Alessandra	MPI for Gravitational Physics	Germany
Parsotan, Tyler	Oregon State University	USA
Pashentseva, Maria	Lomonosov Moscow State University	Russia
Paulo Bessa Brito, João	Universidade federal do Pará	Brazil
Paynter, James	The University of Melbourne	Australia
Pe'er, Asaf	Bar Ilan University	Israel
Peik, Ekkehard	PTB	Germany
Peng, Wenxi	Institute of High Energy Physics, CAS	China
Pereira, Jonas	Nicolaus Copernicus Astronomical Ctr	Poland
Perez Martinez, Aurora Maria	Universidad de Salamanca	Spain
Perez-Garcia, M. Angeles	University de Salamanca	Spain
Perlick, Volker	ZARM, University of Bremen	Germany
Perlman, Eric	Florida Institute of Technology	USA
Pernas, David	Saturno Interconsult	USA
Perrodin, Delphine	INAF - OAC	Italy
Philipp, Dennis	ZARM, University of Bremen	Germany
Phurailatpam, Hemantakumar	The Chinese University of Hong Kong	India
Pian, Elena	INAF-OAS Bologna	Italy
Piedipalumbo, Ester	University of Naples Federico II	Italy
Pietroni, Silvia	UNISA - Università di Salerno	Italy
Pinfold, James	University of Alberta	Canada
Piran, Tsvi	The Hebrew University	Israel

Pisani, Alice	Princeton University	USA
Pitschmann, Mario	TU Wien	Austria
Pittelli, Antonio	University of Surrey	Italy
Pla Garcia, Silvia	University of Valencia - IFIC	Spain
Planas Llompart, Maria de Lluc	University of the Balearic Islands	Spain
Plavin, Alexander	Lebedev Physical Institute	Russia
Plefka, Jan	Humboldt University Berlin	Germany
Pober, Jonathan	Brown University	USA
Pogosian, Levon	Simon Fraser University	Canada
Pohlig, Michael	Karlsruhe Inst. of Technology	Germany
Polarski, David	Université de Montpellier	France
Polin, Abigail	Carnegie Observatories & Caltech	USA
Poludnenko, Alexei	University of Connecticut	USA
Pombo, Alexandre	Aveiro University	Portugal
Pompa, Tiziana	Liceo Scientifico Galileo Galilei	Italy
Ponti, Gabriele	INAF OAB	Italy
Popkova, Anastasia	University of Western Australia	Australia
Poplawski, Nikodem	University of New Haven	USA
Possenti, Andrea	INAF-OAC	Italy
Pourtsidou, Alkistis	QMUL	UK
Pradler, Josef	Austrian Academy of Sciences	Austria
Prakapenia, Mikalai	ICRANet-Minsk	Belarus
Predehl, Peter	MPI für extraterrestrische Physik	Germany
Pugliese, Daniela	RCTPA	Czechia
Pullin, Jorge	Louisiana State University	USA
Punsly, Brian	ICRANet	USA
Qadir, Asghar	Government college University	Pakistan
Quevedo, Hernando	UNAM	Mexico
Quintero Angulo, Gretel	Universidad de La Habana	Cuba
Quirola, Jonathan	Universidad Catolica de Chile	Chile
Qvarfort, Sofia	Imperial College London	UK
Rafiei Karkevandi, Davood	Isfahan University of Technology	Iran
Rah, Maria	Natl. Astron. Observatories of China	Iran
Rahman, Mostafizur	IIT	India
Rahmani, Elnaz	University of Tehran	Iran
Ramírez Romero, Cupatitzio	Universidad Autónoma de Puebla	Mexico
Ramirez-Baca, Pedro Isaac	Universidad Autonoma de San Luis	Mexico
Ranucci, Gioacchino	INAF – sezione di Milano	Italy
Rastegar Nia, Fatemeh	Alzahra University and ICRANet	Iran
Rastgoo, Saeed	York University	Canada
Rätzel, Dennis	Humboldt Universität zu Berlin	Germany
Ravenni, Andrea	Jodrell Bank Centre for Astrophys.	UK

Re, Alessandra Carlotta	Universita' degli Studi & INFN, Milano	Italy
Reboucas, Marcelo	CBPF	Brazil
Recchia, Sarah	University of Torino, INFN Torino	Italy
Reiber, Thomas	Hildesheim University	Germany
Remazeilles, Mathieu	Jodrell Bank Centre for Astrophys.	UK
Ren, Xin	Univ. of Science and Technology of China	China
Renevey, Cyril	LPSC CNRS	France
Riahi, Rashid	ICRANet Isfahan center, Isfahan	Iran
Ribeiro, Ana Rita	Universidade de Lisboa	Portugal
Ribeiro, Deivid	Columbia University	USA
Ricci, Marco	INFN - Laboratori Nazionali Frascati	Italy
Ricciardi, Lorenzo	University of Strathclyde	Italy
Rick Perche, Tales	Perimeter Institute	Canada
Riess, Adam	Johns Hopkins University	USA
Rincon Ramirez, Monica	Pennsylvania State University	USA
Ripperda, Bart	Princeton Univ. /Flatiron Institute	USA
Robbins, Matthew	University of Waterloo	Canada
Robson, Charles	Tampere University	Finland
Rodrigues De Lima, Rafael	Univ. do Estado de Santa Catarina	Brazil
Rodrigues, Claudia	INPE	Brazil
Rodrigues, Manuel	ONERA	France
Rodriguez-Fernandez, Manuel	Universidad de Guadalajara	Mexico
Roelofs, Freek	Harvard & Smithsonian	USA
Romagnolo, Amedeo	Nicolaus Copernicus Astronomical Ctr	Poland
Romano, Antonio Enea	UDEA/CERN	Colombia
Röpke, Friedrich	Heidelberg University/HITS	Germany
Rosa, João	University of Tartu	Portugal
Rosati, Piero	University of Ferrara	Italy
Rosati, Giacomo	IFT Wroclaw University	Poland
Rossi, Andrea	Istituto Nazionale di Astrofisica	Italy
Rossi, Nicola	INFN LNGS	Italy
Rota, Paolo	Università degli Studi di Salerno	Italy
Roulet, Esteban	CONICET	Argentina
Rovelli, Carlo	AMU University	France
Ruan, Manqi	IHEP	China
Rubiera-Garcia, Diego	Complutense University of Madrid	Spain
Rudenko, Valentin	Sternberg Astronomical Institute	Russia
Rueda Hernandez, Jorge	ICRANet	Italy
Ruep, Maximilian Heinz	University of York, UK	UK
Ruffini, Remo	ICRANet	Italy
Ruggiero, Matteo Luca	Politecnico di Torino and INFN, LNL	Italy
Ruiz Velasco, Edna Loredana	MPI for Nuclear Physics	Germany

Ruiz-Lapuente, Pilar	ICCUB U. Barcelona and IFF (CSIC)	Spain
Rybak, Ivan	IACE	Portugal
Saab, Tarek	University of Florida	USA
Saadatmand, Homa	Tehran University	Iran
Sabido, Miguel	Universidad de Guanajuato	Mexico
Saffer, Alexander	University of Virginia	USA
Safronova, Marianna	University of Delaware	USA
Sahakyan, Narek	ICRANet-Armenia	Armenia
Sahlmann, Hanno	FAU Erlangen-Nürnberg	Germany
Sahoo, Pradyumn	BITS-Pilani, Hyderabad Campus	India
Sahoo, Parbati	University of KwaZulu-Natal	South Africa
Sahota, Harkirat Singh	IISER Mohali	India
Saini, Sahil	Guru Jambheshwar University	India
Sajadmanesh, Sina	EPFL	Switzerland
Sakellariadou, Mairi	King's College London	UK
Saleem, Rabia	COMSATS University Islamabad	Pakistan
Salemi, Francesco	University of Trento and INFN	Italy
Salisbury, Donald	Austin College	USA
Salmon, Sébastien	Université de Genève	Switzerland
Salvati, Laura	IAS, Paris Saclay - CNRS	France
Sanchez, Ariel G.	MPI for Extraterrestrial Physics	Germany
Sanchis-Gual, Nicolas	Universidade de Aveiro	Portugal
Sangwan, Archana	Indian Institute Of Technology	India
Sangwan, Archana	Indian Institute of Technology	India
Santa, Camilo	PhD student	Colombia
Santacruz, Robert	University of New Brunswick	Canada
Santos Pereira, Osvaldo Luiz	Universidade Federal do Rio de Janeiro	Brazil
Santos, Nuno	Técnico Lisboa & Univ. de Aveiro	Portugal
Sapio, Feliciana	Istituto Nazionale di Astrofisica	Italy
Saridakis, Emmanuel	National Observatory of Athens	Greece
Sarkar, Shilpa	Aryabhatta Research Institute	India
Sarkar, Sarben	King's College London	UK
Sartini, Francesco	ENS de Lyon	France
Sarwar, Rafia	Institute of Space Technology	Pakistan
Sato, Yuki	Nagoya University	Japan
Sato, Yuri	Aoyama Gakuin University	Japan
Saurabh, Saurabh	University of Delhi	India
Savelova, Elena	Bauman Moscow State Technical Univ.	Russia
Schaan, Emmanuel	Lawrence Berkeley National Lab	USA
Schiavone, Tiziano	University of Pisa & INFN Pisa	Italy
Scholz, Paul	University of Toronto	Canada
Schubert, Christian	Universidad Michoacana	Mexico

Schüssler, Fabian	IRFU/CEA Paris-Saclay	France
Scott, Susan	The Australian National University	Australia
Scott, Kameron	Baylor University	USA
Sedmik, René	TU Wien	Austria
Seeger, Robert	FAU Erlangen-Nürnberg	Germany
Seglar-Arroyo, Monica	L.A.P.P., Université Savoie Mont-Blanc	France
Sellers, Luke	Univ. of California Santa Barbara	USA
Semerak, Oldrich	Charles University	Czechia
Semrén, Philip	Umeå University	Sweden
Sengo, Ivo	University of Aveiro	Portugal
Serbenta, Julius	Center for Theoretical Physics, PAS	Poland
Serenelli, Aldo	Institute of Space Sciences (ICE, CSIC)	Spain
Serra, Francesco	Scuola Normale Superiore, Pisa	Italy
Seto, Osamu	Hokkaido University	Japan
Shafieloo, Arman	Korea Astron. and Space Sci. Inst.	South Korea
Shahidi, Shahab	Damghan University	Iran
Shakeri, Soroush	ICRANet-Isfahan	Iran
Sharif, Muhammad	University of the Punjab	Pakistan
Sharifian, Mohammad	ICRANet-Isfahan	Iran
Shears, Jeremy	University of Manchester	UK
Shen, Rong-Feng	Sun Yat-Sen University	China
Shen, Ken	UC Berkeley	USA
Shkerin, Andrey	University of Minnesota	USA
Shu, Fu-Wen	Nanchang University	China
Sidharth, Gowtham	vellore institute of technology	India
Sieroka, Norman	University of Bremen	Germany
Sigismondi, Costantino	ICRA Sapienza and ICRANET Pescara	Italy
Silk, Joe	IAP	France
Silva, Hilberto	Inst. Astrofísica e Ciências do Espaço	Portugal
Silvestri, Alessandra	Leiden University	Netherlands
Simão, José Diogo	TPI Jena	Germany
Simpson, Alex	Victoria University of Wellington	New Zealand
Singh, Chandra Bahadur	South-Western Inst. for Astron. Research	China
Singh, Parampreet	Louisiana State University	USA
Skordis, Constantinos	CEICO, Institute of Physics	Czechia
Slagter, Reinoud	ASFYON	Netherlands
Smartt, Stephen	Queen's University Belfast	UK
Smirnov, Oleg	JINR, Dubna	Russia
Sobouti, Yousef	Inst. for Adv. Studies in Basic Sci.	Iran
Soffitta, Paolo	INAF-IAPS	Italy
Sokolenko, Anastasia	Austrian Academy of Sciences	Austria
Solà Peracaula, Joan	Universitat de Barcelona	Spain

Soldateschi, Jacopo	Università degli Studi di Firenze	Italy
Solís López, Jordi	Cinvestav	Mexico
Solon, Mikhail	UCLA	USA
Soltani, Farshid	Western U.	Canada
Sousa, Manoel Felipe	Natl. Inst. for Space Research	Brazil
Spengler, Felix	Universität Tübingen	Germany
Spruit, Hendrik	MPI for Astrophysics	Germany
Sravan Kumar, Korumilli	Tokyo Institute of Technology	Japan
Sreenath, V	Natl. Inst. of Technology Karnataka	India
Srivastava, Manu	IIT Bombay	India
Stadnik, Yevgeny	Kavli IPMU, University of Tokyo	Japan
Staicova, Denitsa	INRNE, BAS	Bulgaria
Starobinsky, Alexei	Landau Institute for Theoretical Physics	Russia
Stathopoulos, Stamatios Ilias	National and Kapodistrian University	Greece
Steele, John	UNSW	Australia
Stefano, Scopel	Sogang University	Italy
Stephan, Preiβ	Hildesheim University	Germany
Stephenson, Gary	Seculine Consulting	USA
Stepniczka, Michael	University of Virginia	USA
Strang, Lucy	University of Melbourne	Australia
Stratta, Giulia	INAF/OAS-Bologna	Italy
Straub, Odele	MPI for Extraterrestrial Physics	Germany
Stubbings, Rod	AAVSO	Australia
Stuchlik, Zdenek	Insitute of Physics, Silesian University	Czechia
Suárez Fontanella, Duvier	ICIMAF	Cuba
Suarez Gonzalez, Lismary	ICIMAF	Cuba
Sullivan, Raelyn	The University of British Columbia	Canada
Sun, Ling	The Australian National University	Australia
Suresh, Sandeep	TU Wien	India
Suresh, Jishnu	Institute for Cosmic Ray Research	Japan
Sutton, Patrick	Cardiff University	UK
Suvikranth, Gera	IIT KHARAGPUR	India
Suyu, Sherry	MPI for Astrophysics	Germany
Svarc, Robert	Charles University	Czechia
Svertilov, Sergey	Lomonosov Moscow State University	Russia
Swami, Himanshu	IISER Mohali	India
Szczepanczyk, Marek	University of Florida	USA
't Hooft, Gerard	Utrecht University	Netherlands
T U, Jeevitha	Vellore Inst. of Technology	India
Tahir, Noraiz	Universita Del Salento	Italy
Tahura, Shammi	University of Virginia	USA
Tahvildar-Zadeh, A. Shadi	Rutgers University	USA

Takamoto, Masao	RIKEN	Japan
Takata, Jumpei	Huazhong University	China
Tam, Pak-Hin Thomas	Sun Yat-sen University	China
Tamar, Aditya	Early Career Independent Researcher	India
Tamburini, Matteo	MPI for Nuclear Physics	Germany
Tan, Abraham	National Cheng Kung University	Taiwan
Tanikawa, Ataru	The University of Tokyo	Japan
Tanimura, Hideki	L'Institut d'Astrophysique Spatiale (IAS)	France
Tao, Lian	Institute of High Energy Physics, CAS	China
Tartaglia, Angelo	INAF	Italy
Tashiro, Makoto	Saitama University	Japan
Tauber, Jan	ESA	Netherlands
Tawfik, Abdel Nasser	Egyptian Center for Theoretical Physics	Egypt
Tchekhovskoy, Alexander	Northwestern University	USA
Tedesco, Antonio	University of Salerno	Italy
Teixeira, Elsa	University of Sheffield	UK
Tenorio, Rodrigo	Universitat de les Illes Balears	Spain
Testera, Gemma	INFN Genova (Italy)	Italy
Thomas, Cameron	University of Sheffield	UK
Tiurina, Nataly	Lomonosov Moscow State University	Russia
Tizchang, Seddigheh	Inst. for Research in Fundamental Sci.	Iran
Tluczykont, Martin	University of Hamburg	Germany
Tobar, Michael	The University of Western Australia	Australia
Toda, Yo	Hokkaido University	Japan
Todorinov, Vasil	University of Lethbridge	Canada
Tohuvavohu, Aaron	University of Toronto	Canada
Tokareva, Anna	University of Jyvaskyla	Finland
Toktarbay, Saken	Al-Farabi Kazakh National Univ.	Kazakhstan
Tominaga, Nozomu	NAOJ	Japan
Toprak, Ebru	Rutgers University	USA
Toropina, Olga	Space Research Institute (IKI)	Russia
Torres Sanchez, Victor A.	University of Trento	Italy
Trani, Alessandro	The University of Tokyo	Japan
Tristram, Matthieu	IJClab, CNRS, France	France
Troja, Eleonora	Arizona State University	USA
Trova, Audrey	ZARM, University of Bremen	Germany
Tsujikawa, Shinji	Waseda University	Japan
Tsunetoe, Yuh	Kyoto University	Japan
Tsupko, Oleg	Space Research Institute	Russia
Turundaevskiy, Andrey	SINP MSU	Russia
Turzynski, Krzysztof	University of Warsaw	Poland
Ulmer, Stefan	RIKEN (JP)	Japan

Urena-Lopez, Luis	Universidad de Guanajuato	Mexico
Ursi, Alessandro	INAF-IAPS	Italy
Uzun, Nezihe	Université Claude Bernard Lyon 1	France
V H, Satheeshkumar	UNIRIO	Brazil
V, Gayathri	University of Florida	USA
Vallejo Peña, Sergio Andrés	Universidad De Antioquia	Colombia
van der Walt, Petrus	Rhodes University	South Africa
van Eerten, Hendrik	University of Bath	UK
van Leeuwen, Joeri	ASTRON/U. Amsterdam	Netherlands
van Loon, Jacco	Keele University	UK
van Putten, Maurice	Sejong university	South Korea
Vardanyan, Vazgen	ICRANet-Armenia	Armenia
Vargas, Andrés	The university of Melbourne	Australia
Varo, Valle	Universidad Carlos III de Madrid	Spain
Veitch, John	University of Glasgow	UK
Venkatraman Krishnan, Vivek	MPI for Radio Astronomy	Germany
Vera Garfias, Gabriela Isabel	Universidad Autónoma Metropolitana	Mexico
Verde, Licia	ICREA & ICC Universitat de Barcelona	Spain
Veres, Peter	University of Alabama in Huntsville	USA
Vereshchagin, Gregory	ICRANet	Italy
Vermeulen, Sander	Cardiff University	UK
Verrecchia, Francesco	INAF/OAR, SSDC-ASI	Italy
Veske, Doğa	Columbia University	USA
Vicentini, Silvia	University of Trento	Italy
Vilalta, Ricardo	University of Houston	USA
Villanova Borges, Sarah	University of Wisconsin Milwaukee	USA
Villante, Francesco	University of L'Aquila	Italy
Villaseñor, Eduardo J S	Universidad Carlos III de Madrid	Spain
Vincent, Frederic	CNRS/Paris Observatory/LESIA	France
Virgilli, Enrico	INAF OAS Bologna	Italy
Visser, Matt	Victoria University of Wellington	New Zealand
Vladykina, Polina	Bachelor	Russia
Volkov, Mikhail	University of Tours, France	France
Vollmann, Wolfgang	BAV, AAVSO	Austria
Vorobyeva, Alexandra	RUDN University	Russia
Vyas, Mukesh Kumar	Bar Ilan University, Ramat Gan, Israel	Israel
Wadiasingh, Zorawar	NASA GSFC	USA
Wagner, Jenny	ITP Heidelberg University	Germany
Wallace, John	Bar-Ilan University	Israel
Wands, David	University of Portsmouth	UK
Wang, Fayin	Nanjing University	China
Wang, Xiang-Yu	Nanjing University	China

Wang, Yong-Qiang	Lanzhou University	China
Wang, Yusa	IHEP, CAS, China	China
Wang, Zhiwei	Jilin University	China
Wang, Yu	ICRANet	Italy
Wang, Anzhong	Baylor University	USA
Wardell, Barry	University College Dublin	Ireland
Warren, Don	RIKEN	Japan
Watson, Alan	UNAM	Mexico
Watts, Anna	University of Amsterdam	Netherlands
Weissenborn, Sven	University of Hildesheim	Germany
Weltman, Amanda	University of Cape Town	South Africa
Westerweck, Julian	AEI Hannover	Germany
White, Nicholas	George Washington University	USA
Wicker, Raphaël	IAS	France
Wieland, Wolfgang	IQOQI	Austria
Wielgus, Maciek	Black Hole Initiative at Harvard Univ.	USA
Williams, Jeff	Brandon University	Canada
Williams, Liliya	University of Minnesota	USA
Witzany, Vojtech	University College Dublin	Ireland
Woerner, Lisa	DLR-QT	Germany
Wolz, Anna	University of Virginia	USA
Wondrak, Michael Florian	Goethe-Universität Frankfurt	Germany
Wright, Michael	The Archive Trust for Research	UK
Wu, An-Ming	National Space Program Office	Taiwan
Xavier, Sérgio	Federal University of Pará	Brazil
Xavier, Semin	IIT Bombay	India
Xiong, Shaolin	Institute of High Energy Physics	China
Xu, Fan	Nanjing University	China
Xue, She-Sheng	ICRANet, Sapienza University of Rome	Italy
Yagi, Kent	University of Virginia	USA
Yamamoto, Kohei	Kumamoto University	Japan
Yamazaki, Ryo	Aoyama Gakuin University	Japan
Yan, Sheng-Feng	INFN (Milano), USTC	China
Yan, Zhenyu	Nanjing University	China
Yang, Jun	Nanjing University	China
Yang, Yuanpei	South-Western Inst. for Astron. Research	China
Yasmin, Aqsa	NUST	Pakistan
Ye, Claire	CIERA/Northwestern University	USA
Yefremov, Alexander	RUDN University	Russia
Yeh, Chih-Han	Physikalisch-Technische Bundesanstalt	Germany
Yernazarov, Tursynbek	Al-Farabi Kazakh National Univ.	Kazakhstan
Yildirim, Akin	MPI for Astrophysics	Germany

Yokokura, Yuki	RIKEN	Japan
Yoon, Myungseok	Sangmyung University	South Korea
Young, Samuel	University of Pennsylvania	USA
Yu, Hang	California Institute of Technology	USA
Yuan, Qiang	Purple Mountain Observatory	China
Yuan, Weimin	National Astronomical Observatories	China
Yue, Qian	Tsinghua University	China
Yunis, Rafael Ignacio	ICRANet Pescara	Italy
Zaccagnini, Fabio	University of Rome La Sapienza	Italy
Zahariade, George	Institut de Fisica d'Altes Energies	Spain
Zakharov, Alexander	ITEP, Moscow	Russia
Zamani, Saboura	Golestan University, Iran	Iran
Zaslavskii, Oleg	Kharkov V.N. Karazin Natl. Univ.	Ukraine
Zavatarelli, Sandra	INFN - Sezione di Genova	Italy
Zen Vasconcellos, Cesar A.	ICRANet/UFRGS	Brazil
Zhan, Mingsheng	Wuhan Inst. of Physics and Mathematics	China
Zhang, Binbin	Nanjing University	China
Zhang, Guoqiang	Nanjing University	China
Zhang, Shuang-Nan	Institute of High Energy Physics	China
Zhang, Xuefeng	Sun Yat-sen University	China
Zhang, Yuanhao	AEI Hannover	Germany
Zhang, Cong	University of Warsaw	Poland
Zhang, Bing	University of Nevada, Las Vegas	USA
Zheng, Jiaming	Shanghai Jiao-tong University	China
Zhirkov, Kirill	Moscow State University	Russia
Zhitnitsky, Ariel	University of British Columbia	Canada
Zhou, Hao	Shanghai Jiao Tong University	China
Zhou, Ning	Shanghai Jiao Tong University	China
Zhu, Jinping	Peking University	China
Zhu, Tao	Zhejiang University of Technology	China
Zhu, Sylvia	DESY	Germany
Ziegler, Joshua	University of Texas at Austin	USA
Zuccon, Paolo	Trento University and INFN TIFPA	Italy